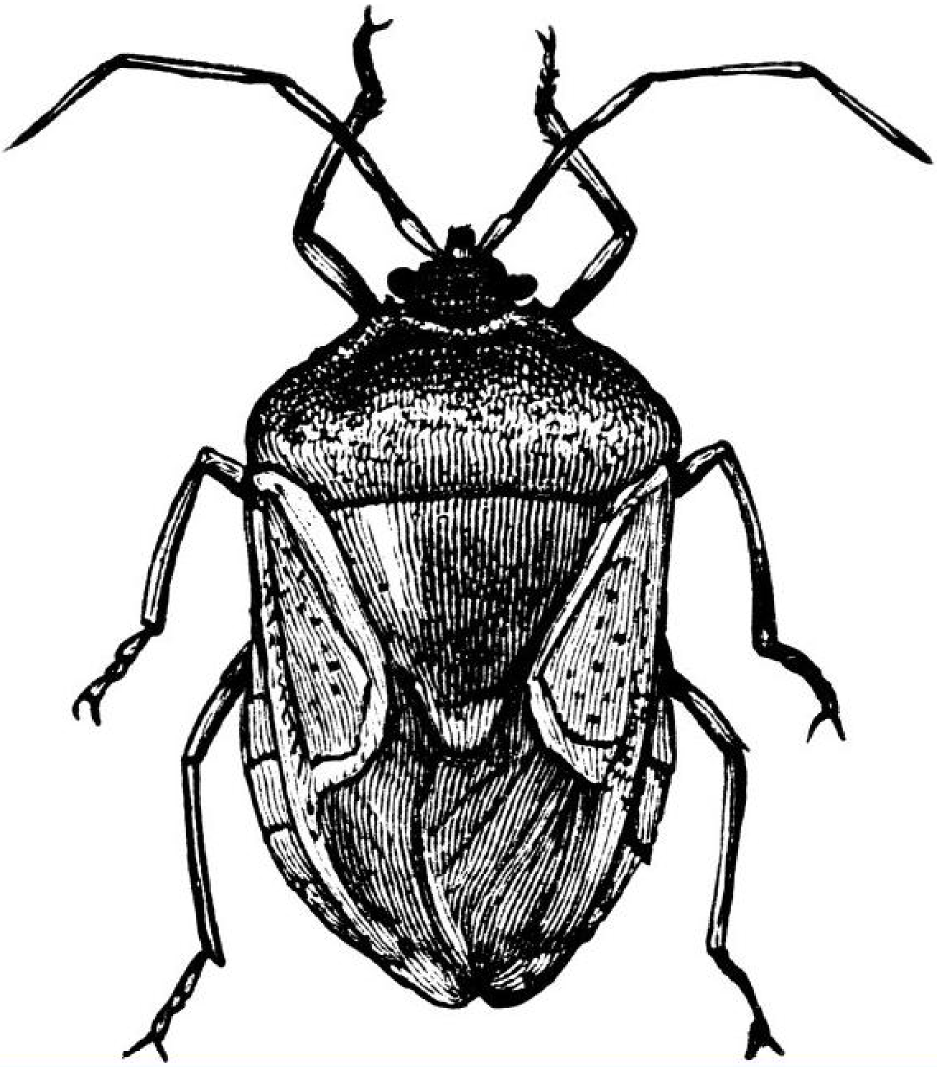
A blog about science, medicine, philosophy, politics, and futurism.



Slate Star Codex

2013 - 2020

slatestarcodex.com

Scott Alexander

- **Favorites**
- **2013**
- **2014**
- **2015**
- **2016**
- **2017**
- **2018**
- **2019**
- **2020**

Favorites

- Top Posts
- Archipelago And Atomic Communitarianism
- Beware The Man Of One Study
- Book Review: Albion’s Seed
- Considerations On Cost Disease
- I Can Tolerate Anything Except The Outgroup
- Meditations on Moloch
- Nobody Is Perfect, Everything Is Commensurable
- Studies On Slack
- The Categories Were Made For Man, Not Man For The Categories
- The Control Group Is Out Of Control
- Toxoplasma of Rage
- Who By Very Slow Decay

Top Posts

My most popular posts are very long and dense, but if you're up for that kind of thing, you can read:

- The Control Group Is Out Of Control on the weakness of a lot of modern scientific research and statistics.
- I Can Tolerate Anything Except The Outgroup on a general framework for thinking about tolerance vs prejudice.
- Toxoplasma of Rage on how the media works and why stupid topics keep coming up again and again.
- Meditations on Moloch, on…um…it's hard to explain.

If you're looking for something a little shorter or lighter, you can ~~find yourself a different blog~~ check out for example:

- Hardball Questions I want asked at the 2016 Republican presidential debates.
- How Deep The Rabbit Hole Goes, a story about that silly meme where you have to choose one of eight colored pills.

More specifically by topic:

I am a doctor and try (usually unsuccessfully) to focus this blog on **medicine**. I'm especially interested in the pharmaceutical system and drug discovery process, but I also write about my own experiences practicing psychiatry:

- The Life Cycle Of Medical Ideas about how people identify which promising treatments to pursue.

- Sleep – Now By Prescription, about the fuzzy border between prescription and nonprescription medications
- Fish – Now By Prescription, a continuation of the above.
- An Iron Curtain Has Descended Upon Psychopharmacology, about how Russia uses different drugs than we do and our system hasn't found a way to incorporate their knowledge
- Pharma Virumque on pharmaceutical company advertising, and how it's even worse than you've heard.
- Who By Very Slow Decay, on my depressing experiences with end-of-life care
- Evening Doc, on various others of my depressing experiences
- Medicine As Not Seen On TV, looking back after one year or medical residency
- Reflections From The Halfway Point, looking back after two years of medical residency
- Burdens, on what I want to tell my suicidal patients

My interest in drug discovery naturally segues into **scientific and statistical methods** in general. My longest piece on this is 'The Control Group Is Out Of Control', linked above, but I've also written about:

- Statistical Literacy Among Doctors Is Lower Than Chance on the abysmal statistical knowledge of doctors and what it means for you
- Dark Side Statistics Papers, an explanation of how to fudge research results
- Specific criticism of various sketchy studies on euthanasia, Victorian IQ, gender, welfare, and bullying

I write a lot about **politics** from a vaguely centrist point of view with occasional forays to the right or left. Some especially interesting political threads here include:

- A Thrive-Survive Theory Of The Political Spectrum, about whether “left” and “right” are real internally consistent things
- A Something Sort-Of-Like-Left-Libertarian Manifesto on using regulations versus taxes and subsidies to solve political problems
- Archipelago and Atomic Communitarianism on how atomic individualism can fit together with people’s concern for their communities
- Right Is The New Left, about how the same dynamics that supported the leftist counterculture of the 1960s are leading to a rightist counterculture today
- Black People Less Likely, about what we can learn from checking which fields African-Americans are/aren’t underrepresented in.
- In Favor Of Niceness, Community, and Civilization, about how just because your cause is important doesn’t mean you get to be a jerk about it.
- Reactionary Philosophy In An Enormous Planet-Sized Nutshell, which is my attempt to see if I can blast several metric tons of highly concentrated conservative political philosophy into the brains of unsuspecting people.
- Slate Star Codex Political Spectrum Quiz, which is exactly what it says on the tin.

Although I acknowledge the importance and danger of racism and sexism, I also think a *lot* of the **social justice** movement as it currently exists is an attempt to sanctify ad hominem arguments and poor epistemology that can be used by a would-be cognitive elite to abuse and humiliate anyone who disagrees with them. I start the explanation in ‘I Can Tolerate Anything Except The Outgroup’, linked above, but there is more in:

- A Response To Apophemi on Triggers, about competing access needs
- Living By The Sword, which was supposed to be about toxicity in the social justice community but is better remembered more for its complicated whale cancer metaphor.
- Social Justice and Words, Words, Words about the way language gets weaponized and used as a tool to confuse people.
- The Categories Were Made For Man, Not Man For The Categories, about why it’s important to treat transgender individuals as their preferred gender.
- Social Psychology Is A Flamethrower, about the poor condition of the social justice evidence base.
- The Wonderful Thing About Triggers, in which I come out in favor of trigger warnings despite all of the above

Sometimes I get bored and just **research** the *hell* out of something to try to resolve a difficult question to my own satisfaction. Thus far this has resulted in cost-benefit analyses like:

- Marijuana: Much More Than You Wanted To Know, on whether we should legalize marijuana.
- Wheat: Much More Than You Wanted To Know, on whether wheat/bread/gluten is bad for you.
- Race And Justice: Much More Than You Wanted To Know, on which steps of the criminal justice system are/aren’t racist.
- SSRIs: Much More Than You Wanted To Know, on the effectiveness of SSRI antidepressants.

I’m also very interested in **rationality** – questions like how debates work in general and how we can conduct them better. I find the current set of logical fallacies mostly orthogonal to the way debates between smart people fail, so I have tried to do better:

- Arguments From My Opponent Believes Something, ie fully general counterarguments that don't add anything to a debate.
- All Debates Are Bravery Debates, on how a lot of hard problems are trade-offs between two goods and debates are difficult because we can't tell which side people are erring towards.
- Weak Men Are Superweapons and Cardiologists and Chinese Robbers, both on how straw men are less dangerous than real-but-irrelevant "weak men" that let people cherry-pick stupid examples.
- If It's Worth Doing, It's Worth Doing With Made-Up Statistics, on how sometimes using numbers – even messy, potentially inaccurate numbers – can shed light on a problem
- Beware Isolated Demands For Rigor, on not setting impossibly high burden of proof for ideas you don't like – plus an Old West shootout scene between Greek philosophers

I've also tried writing a little bit of **fiction**, of which I most like:

- Universal Love, Said The Cactus Person, kind of hard to explain
- Answer To Job, also kind of hard to explain
- The Study of Anglophysics, which is *definitely* hard to explain

Other people who might be less biased have put together their own recommended lists here and here. And you can find a full archive of all SSC posts here.

Archipelago and Atomic Communitarianism

I.

In the old days, you had your Culture, and that was that. Your Culture told you lots of stuff about what you were and weren't allowed to do, and by golly you listened. Your Culture told you to work the job prescribed to you by your caste and gender, to marry who your parents told you to marry or at *least* someone of the opposite sex, to worship at the proper temples and the proper times, and to talk about *proper* things as opposed to the blasphemous things said by the tribe over there.

Then we got Liberalism, which said all of that was mostly bunk. Like Wicca, its motto is "Do as you will, so long as it harms none". Or in more political terms, "Your right to swing your fist ends where my nose begins" or "If you don't like gay sex, don't have any" or "If you don't like this TV program, don't watch it" or "What happens in the bedroom between consenting adults is none of your business" or "It neither breaks my arm nor picks my pocket". Your job isn't to enforce your conception of virtue upon everyone to build the Virtuous Society, it's to live your own life the way you want to live it and let other people live *their* own lives the way *they* want to live them. This is the much-maligned "atomic individualism," or maybe just liberalism boiled down to its pure essence.

But atomic individualism wasn't as great a solution as it sounded. Maybe one of the first cracks was tobacco ads. Even though putting up a billboard saying "SMOKE MARLBORO" neither breaks anyone's arm nor picks their pocket, it shifts social expectations in such a way that bad effects occur. It’s hard to dismiss that with “Well, it’s people’s own choice to smoke and they should live their lives the way they want” if studies show that more people will want to live their lives in a way that gives them cancer in the presence of the billboard than otherwise.

From there we go into policies like Michael Bloomberg’s ban on giant sodas. While the soda ban itself was probably as much symbolic as anything, it’s hard to argue with the impetus behind it – a culture where everyone gets exposed to the option to buy very very unhealthy food all the time is going to be less healthy than one where there are some regulations in place to make EAT THIS DONUT NOW a less salient option. I mean, I *know* this is true. A few months ago when I was on a diet I *cringed* every time one my coworkers brought in a box of free donuts and placed wide-open in the doctors’ lounge; there was *no way* I wasn’t going to take one (or two, or three). I could ask people to stop, but they probably wouldn’t, and even if they did I’d just encounter the wide-open box of free donuts *somewhere else*. I’m not proposing that it is *ethically wrong* to bring in free donuts or that banning them is the correct policy, but I do want to make it clear that stating “it’s your free choice to partake or not” doesn’t eliminate the problem, and that this points to an entire class of serious issues where atomic individualism as construed above is at best an imperfect heuristic.

And I would be remiss talking about the modern turn away from individualism without mentioning social justice. The same people who once deployed individualistic arguments against conservatives: “If you don’t like profanity, don’t use it”, “If you don’t like this offensive TV show, don’t watch it”, “If you don’t like pornography, don’t buy it” – are now concerned about people using ethnic slurs, TV shows without enough minority characters, and pornography that encourages the objectification of women. I’ve objected to some of this on purely empirical grounds, but the least convenient possible world is the one where the purely empirical objections fall flat. If they ever discover proof positive that yeah, pornographication makes women hella objectified, is it acceptable to censor or ban misogynist media on a society-wide level?

And if the answer is yes – and if such media like really, *really* increases the incidence of rape I'm not sure how it couldn't be – then what about all those conservative ideas we've been neglecting for so long? What if strong, cohesive, religious, demographically uniform communities make people more trusting, generous, and cooperative in a way that *also* decreases violent crime and other forms of misery? We have lots of evidence that this is true, and although we can doubt each individual study, we owe conservatives the courtesy of imagining the possible world in which they are right, the same as anti-misogyny leftists. Maybe media glorifying criminals or lionizing nonconformists above those who quietly follow cultural norms has the same kind of erosive effects on "values" as misogynist media. Or, at the very least, we ought to have a good philosophy in place so that we have some idea what to do it if does.

II.

A while ago, in Part V of this essay, I praised liberalism as the only peaceful answer to Hobbes' dilemma of the war of all against all.

Hobbes says that if everyone's fighting then everyone loses out. Even the winners probably end up worse off than if they had just been able to live in peace. He says that governments are good ways to prevent this kind of conflict. Someone – in his formulation a king – tells everyone else what they're going to do, and then everyone else does it. No fighting necessary. If someone tries to start a conflict by ignoring the king, the king crushes them like a bug, no prolonged fighting involved.

But this replaces the problem of potential warfare with the problem of potential tyranny. So we've mostly shifted from absolute monarchies to other forms of government, which is all nice and well except that governments allow a *different* kind of war of all against all. Instead of trying to kill their enemies and steal their stuff, people are tempted to ban their enemies and confiscate their stuff. Instead of killing the Protestants, the Catholics simply ban Protestantism. Instead of forming vigilante mobs to stone homosexuals, the straights merely declare homosexuality is punishable by death. It *might* be better than the alternative – at least everyone knows where they stand and things stay peaceful – but the end result is still a lot of pretty miserable people.

Liberalism is a new form of Hobbesian equilibrium where the government enforces not only a ban on killing and stealing from people you don't like, but also a ban on tyrannizing them out of existence. This is the famous "freedom of religion" and "freedom of speech" and so on, as well as the "freedom of what happens in the bedroom between consenting adults". The Catholics don't try to ban Protestantism, the Protestants don't try to ban Catholicism, and everyone is happy.

Liberalism only works when it's clear to everyone on all sides that there's a certain neutral principle everyone has to stick to. The neutral principle can't be the Bible, or Atlas Shrugged, or anything that makes it look like one philosophy is allowed to judge the others. Right now that principle is the Principle of Harm: you can do whatever you like unless it harms other people, in which case stop. We seem to have inelegantly tacked on an "also, we can collect taxes and use them for a social safety net and occasional attempts at social progress”, but it seems to be working pretty okay too.

The Strict Principle of Harm says that pretty much the only two things the government can get angry at is literally breaking your leg or picking your pocket – violence or theft. The Loose Principle of Harm says that the government can get angry at complicated indirect harms, things that Weaken The Moral Fabric Of Society. Like putting up tobacco ads. Or having really really big sodas. Or publishing hate speech against minorities. Or eroding trust in the community. Or media that objectifies women.

No one except the most ideologically pure libertarians seems to want to insist on the Strict Principle of Harm. But allowing the Loose Principle Of Harm restores all of the old wars to control other people that liberalism was supposed to prevent. The one person says “Gay marriage will result in homosexuality becoming more accepted, leading to increased rates of STDs! That’s a harm! We must ban gay marriage!” Another says “Allowing people to send their children to non-public schools could lead to kids at religious schools that preach against gay people, causing those children to commit hate crimes when they grow up! That’s a harm! We must ban non-public schools!” And so on, forever.

And I’m talking about non-governmental censorship just as much as government censorship. Even in the most anti-gay communities in the United States, the laws usually allow homosexuality or oppose it only in very weak, easily circumvented ways. The real problem for gays in these communities is the social pressure – whether that means disapproval or risk of violence – that they would likely face for coming out. This too is a violation of liberalism, and it’s one that’s as important or more important than the legal sort.

And right now our way of dealing with these problems is to argue them. “Well, gay people don’t really increase STDs too much.” Or “Home-schooled kids do better than public-schooled kids, so we need to allow them.” The problem is that arguments never terminate. Maybe if you’re *incredibly* lucky, after years of fighting you can get a couple of people on the other side to admit your side is right, but this is a pretty hard process to trust. The great thing about religious freedom is that it short-circuits the debate of “Which religion is correct, Catholicism or Protestantism?” and allows people to tolerate both Catholics and Protestants even if they are divided about the answer to this object-level question. The great thing about freedom of speech is that it short-circuits the debate of “Which party is correct, the Democrats or Republicans?” and allows people to express both liberal and conservative opinions even if they are divided about the object-level question.

If we force all of our discussions about whether to ban gay marriage or allow home schooling to depend on resolving the dispute about whether they indirectly harm the Fabric of Society in some way, we’re forcing dependence on object-level arguments in a way that historically has been very very bad.

Presumably here the more powerful groups would win out and be able to oppress the less powerful groups. We end up with exactly what liberalism tried to avoid – a society where everyone is the guardian of the virtue of everyone else, and anyone who wants to live their lives in a way different from the community’s consensus is out of luck.

In Part I, I argued that *not allowing* people to worry about culture and community at all was inadequate, because these things really do matter.

Here I'm saying that if we *do allow* people to worry about culture and community, we risk the bad old medieval days where all nonconformity gets ruthlessly quashed.

Right now we're balanced precariously between the two states. There's a lot of liberalism, and people are generally still allowed to be gay or home-school their children or practice their religion or whatever. But there's also quite a bit of Enforced Virtue, where kids are forbidden to watch porn and certain kinds of media are censored and in some communities mentioning that you're an atheist will get you Dirty Looks.

It tends to work okay for most of the population. Better than the alternatives, maybe? But there's still a lot of the population that's not free to do things that are very important to them. And there's also a lot of the population that would like to live in more "virtuous" communities, whether it's to lose weight faster or avoid STDs or not have to worry about being objectified. Dealing with these two competing issues is a pretty big part of political philosophy and one that most people don't have any principled solution for.

III.

Imagine a new frontier suddenly opening. Maybe a wizard appears and gives us a map to a new archipelago that geographers had missed for the past few centuries. He doesn't want to rule the archipelago himself, though he will reluctantly help kickstart the government. He just wants to give directions and a free galleon to anybody who wants one and can muster a group of likeminded friends large enough to start a self-sustaining colony.

And so the equivalent of our paleoconservatives go out and found communities based on virtue, where all sexual deviancy is banned and only wholesome films can be shown and people who burn the flag are thrown out to be eaten by wolves.

And the equivalent of our social justiciars go out and found communities where all movies have to have lots of strong minority characters in them, and all slurs are way beyond the pale, and nobody misgenders anybody.

And the equivalent of our Objectivists go out and found communities based totally on the Strict Principle of Harm where everyone is allowed to do whatever they want and there are no regulations on business and everything is super-capitalist all the time.

And some people who just really want to lose weight go out and found communities where you're not allowed to place open boxes of donuts in the doctors' lounge.

Usually the communities are based on a charter, which expresses some founding ideals and asks only the people who agree with those ideals to enter. The charter also specifies a system of government. It could be an absolute monarch, charged with enforcing those ideals upon a population too stupid to know what's good for them. Or it could be a direct democracy of people who all agree on some basic principles but want to work out for themselves what direction the principles take them.

After a while the wizard decides to formalize and strengthen his system, not to mention work out some of the ethical dilemmas.

First he bans communities from declaring war on each other. That's an *obvious* gain. He could just smite warmongers, but he thinks it’s more natural and organic to get all the communities into a united government (UniGov for short). Every community donates a certain amount to a military, and the military’s only job is to quash anyone from any community who tries to invade another.

Next he addresses externalities. For example, if some communities emit a lot of carbon, and that causes global warming which threatens to destroy other communities, UniGov puts a stop to that. If the offending communities refuse to stop emitting carbon, then there’s that military again.

The third thing he does is prevent memetic contamination. If one community wants to avoid all media that objectifies women, then no other community is allowed to broadcast women-objectifying media at it. If a community wants to live an anarcho-primitivist lifestyle, nobody else is allowed to import TVs. Every community decides *exactly* how much informational contact it wants to have with the rest of the continent, and no one is allowed to force them to have more than that.

But the wizard and UniGov’s most important task is to think of the children.

Imagine you’re conservative Christians, and you’re tired of this secular godless world, so you go off with your conservative Christian friends to found a conservative Christian community. You all pray together and stuff and are really happy. Then you have a daughter. Turns out she’s atheist and lesbian. What now?

Well, it might be that your kid would be much happier at the lesbian separatist community the next island over. The *absolute minimum* the united government can do is enforce freedom of movement. That is, the *second* your daughter decides she doesn't want to be in Christiantopia anymore, she goes to a UniGov embassy nearby and asks for a ticket out, which they give her, free of charge. She gets airlifted to Lesbiantopia the next day. If *anyone* in Christiantopia tries to prevent her from reaching that embassy, or threatens her family if she leaves, or expresses the *slightest* amount of coercion to keep her around, UniGov burns their city and salts their field.

But this is not nearly enough to fully solve the child problem. A child who is abused may be too young to know that escape is an option, or may be brainwashed into thinking they are evil, or guilted into believing they are betraying their families to opt out. And although there is no perfect, elegant solution here, the practical solution is that UniGov enforces some pretty strict laws on child-rearing, and every child, no matter what other education they receive, also has to receive a class taught by a UniGov representative in which they learn about the other communities in the Archipelago, receive a basic non-brainwashed view of the world, and are given directions to their nearest UniGov representative who they can give their opt-out request to.

The list of communities they are informed about always starts with the capital, ruled by UniGov itself and considered an inoffensive, neutral option for people who don't want anywhere in particular. And it always ends with a reminder that if they can gather enough support, UniGov will provide them with a galleon to go out and found their own community in hitherto uninhabited lands.

There's one more problem UniGov has to deal with: malicious inter-community transfer. Suppose that there is some community which puts extreme effort into educating its children, an education which it supports through heavy taxation. New parents move to this community, reap the benefits, and then when their children grow up they move back to their previous community so they don't have to pay the taxes to educate anyone else. The communities themselves prevent some of this by immigration restrictions – anyone who's clearly taking advantage of them isn't allowed in (except in the capital, which has an official committment to let in anyone who wants). But that still leaves the example of people maliciously leaving a high-tax community once they've got theirs. I imagine this is a big deal in Archipelago politics, but that in practice UniGov asks these people, even in their new homes, to pay higher tax rates to subsidize their old community. Or since that could be morally objectionable (imagine the lesbian separatist having to pay taxes to Christiantopia which oppressed her), maybe they pay the excess taxes to UniGov itself, just as a way of disincentivizing malicious movement.

Because there *are* UniGov taxes, and most people are happy to pay them. In my fantasy, UniGov isn't an enemy, where the Christians view it as this evil atheist conglomerate trying to steal their kids away from them and the capitalists view it as this evil socialist conglomerate trying to enforce high taxes. The Christians, the capitalists, and everyone else are extraordinarily *patriotic* about being part of the Archipelago, for its full name is the Archipelago of Civilized Communities, it is the standard-bearer of civilization against the barbaric outside world, and it is precisely the institution that allows them to maintain their distinctiveness in the face of what would otherwise be irresistable pressure to conform. Atheistopia is the enemy of Christiantopia, but only in the same way the Democratic Party is the enemy of the Republican Party – two groups within the same community who may have different ideas but who consider themselves part of the same broader whole, fundamentally allies under a banner of which both are proud.

IV.

Robert Nozick once proposed a similar idea as a libertarian utopia, and it’s easy to see why. UniGov does very very little. Other than the part with children and the part with evening out taxation regimes, it just sits around preventing communities from using force against each other. That makes it very very easy for anyone who wants freedom to start a community that grants them the kind of freedom they want – or, more likely, to just start a community organized on purely libertarian principles. The United Government of Archipelago is the perfect minarchist night watchman state, and any additions you make over that are chosen by your own free will.

But other people could view the same plan as a conservative utopia. Conservativism, when it’s not just Libertarianism Lite, is about building strong cohesive communities of relatively similar people united around common values. Archipelago is obviously built to make this as easy as possible, and it’s hard to imagine that there wouldn’t pop up a bunch of communities built around the idea of Decent Small-Town God-Fearing People where everyone has white picket fences and goes to the same church and nobody has to lock their doors at night (so basically Utah; I feel like this is one of the rare cases where the US’ mostly-in-name-only Archipelagoness really asserts itself). People who didn’t fit in could go to a Community Of People Who Don’t Fit In and would have no need to nor right to complain, and no one would have to deal with Those Durned Bureaucrats In Washington telling them what to do.

But to me, this seems like a liberal utopia, even a leftist utopia, for three reasons.

The first reason is that it extends the basic principle of liberalism – solve differences of opinion by letting everyone do their own thing according to their own values, then celebrate the diversity this produces. I like homosexuality, you don’t, fine, I can be homosexual and you don’t have to, and having both gay and straight people living side by side enriches society. This just takes the whole thing one meta-level up – I want to live in a very sexually liberated community, you want to live in a community where sex is treated purely as a sacred act for the purpose of procreation, fine, I can live in the community I want and you can live in the community you want, and having both sexually-liberated and sexually-pure communities living side by side enriches society. It is pretty much saying that the solution to any perceived problems of liberalism is *much more liberalism*.

The second reason is quite similar to the conservative reason. A lot of liberals have some pretty strong demands about the sorts of things they want society to do. I was recently talking to Ozy about a group who believe that society billing thin people is fatphobic, and that everyone needs to admit obese people can be just as attractive and date more of them, and that anyone who preferentially dates thinner people is Problematic. They also want people to stop talking about nutrition and exercise publicly. I sympathize with these people, especially having recently read a study showing that obese people are much happier when surrounded by other obese, rather than skinny people. But realistically, their movement will fail, and even philosophically, I’m not sure how to determine if they have the right to demand what they are demanding or what that question means. Their best bet is to found a community on these kinds of principles and only invite people who already share their preferences and aesthetics going in.

The third reason is the reason I specifically draw leftism in here. Liberalism, and to a much greater degree leftism, are marked by the emphasis they place on oppression. They're particularly marked by an emphasis on oppression being a really hard problem, and one that is structurally inherent to a certain society. They are marked by a moderate amount of despair that this oppression can ever be rooted out.

And I think a pretty strong response to this is making sure everyone is able to say "Hey, you better not oppress us, because if you do, we can pack up and go somewhere else."

Like if you want to protest that this is unfair, that people shouldn't be forced to leave their homes because of oppression, fine, fair enough. But given that oppression *is* going on, and you haven't been able to fix it, giving people the *choice* to get away from it seems like a pretty big win. I am reminded of the many Jews who moved from Eastern Europe to America, the many blacks who moved from the southern US to the northern US or Canada, and the many gays who make it out of extremely homophobic areas to friendlier large cities. One could even make a metaphor, I think rightly, to telling battered women that they are allowed to leave their husbands, telling them they're not forced to stay in a relationship that they consider abusive, and making sure that there are shelters available to receive them.

If any person who feels oppressed can leave whenever they like, to the point of being provided a free plane ticket by the government, how long can oppression go on before the oppressors give up and say "Yeah, guess we need someone to work at these factories now that all our workers have gone to the communally-owned factory down the road, we should probably at least let people unionize or something so they will tolerate us"?

A commenter in the latest Asch thread mentioned an interesting quote by Frederick Douglass:

> The American people have always been anxious to know what they shall do with us [black people]. I have had but one answer from the beginning. Do nothing with us! Your doing with us has already played the mischief with us. Do nothing with us!

It sounds like, if Frederick Douglass had the opportunity to go to some other community, or even found a black ex-slave community, no racists allowed, he probably would have taken it [edit: or not, or had strict conditions]. If the people in slavery during his own time period had had the chance to leave their plantations for that community, I bet they would have taken it too. And if you believe there are still people today whose relationship with society are similar in kind, if not in degree, to that of a plantation slave, you should be pretty enthusiastic about the ability of exit rights and free association to disrupt those oppressive relationships.

V.

We lack Archipelago's big advantage – a vast frontier of unsettled land.

Which is not to say that people don't form communes. They do. Some people even have really clever ideas along these lines, like the seasteaders. But the United States isn't going to become Archipelago any time soon.

There's another problem too, which I describe in my Anti-Reactionary FAQ. Discussing 'exit rights', I say:

> Exit rights are a great idea and of course having them is better than not having them. But I have yet to hear Reactionaries who cite them as a panacea explain in detail what exit rights we need beyond those we have already.
>
> The United States allows its citizens to leave the country by buying a relatively cheap passport and go anywhere that will take them in, with the exception of a few arch-enemies like Cuba – and those exceptions are laughably easy to evade. It allows them to hold dual citizenship with various foreign powers. It even allows them to renounce their American citizenship entirely and become sole citizens of any foreign power that will accept them.
>
> Few Americans take advantage of this opportunity in any but the most limited ways. When they do move abroad, it's usually for business or family reasons, rather than a rational decision to move to a different country with policies more to their liking. There are constant threats by dissatisfied Americans to move to Canada, and one in a thousand even carry through with them, but the general situation seems to be that America has a very large neighbor that speaks the same language, and has an equally developed economy, and has policies that many Americans prefer to their own country's, and isn't too hard to move to, and almost no one takes advantage of this opportunity. Nor do I see many people, even among the rich, moving to Singapore or Dubai.

> Heck, the US has fifty states. Moving from one to another is as easy as getting in a car, driving there, and renting a room, and although the federal government limits exactly how different their policies can be you better believe that there are very important differences in areas like taxes, business climate, education, crime, gun control, and many more. Yet aside from the fascinating but small-scale Free State Project there's little politically-motivated interstate movement, nor do states seem to have been motivated to converge on their policies or be less ideologically driven.
>
> What if we held an exit rights party, and nobody came?
>
> Even aside from the international problems of gaining citizenship, dealing with a language barrier, and adapting to a new culture, people are just rooted – property, friends, family, jobs. The end result is that the only people who can leave their countries behind are very poor refugees with nothing to lose, and very rich jet-setters. The former aren't very attractive customers, and the latter have all their money in tax shelters anyway.
>
> So although the idea of being able to choose your country like a savvy consumer appeals to me, just saying "exit rights!" isn't going to make it happen, and I haven't heard any more elaborate plans.

I guess I still feel that way. So although Archipelago is an interesting exercise in political science, a sort of pure case we can compare ourselves to, it doesn't look like a practical solution for real problems.

On the other hand, I do think it's worth becoming more Archipelagian on the margin rather than less so, and that there are good ways to do it.

One of the things that started this whole line of thought was an argument on Facebook about a very conservative Christian law school trying to open up in Canada. They had lots of rules like how their students couldn’t have sex before marriage and stuff like that. The Canadian province they were in was trying to deny them accreditation, because conservative Christians are icky. I think the exact arguments being used were that it was homophobic, because the conservative Christians there would probably frown on married gays and therefore gays couldn’t have sex at all. Therefore, the law school shouldn’t be allowed to exist. There were other arguments of about this caliber, but they all seemed to boil down to “conservative Christians are icky”.

This very much annoyed me. Yes, conservative Christians are icky. And they should be allowed to form completely voluntary communities of icky people that enforce icky cultural norms and an insular society promoting ickiness, just like everyone else. If non-conservative-Christians don’t like what they’re doing, they should *not go to that law school*. Instead they can go to one of the dozens of other law schools that conform to their own philosophies. And if gays want a law school even friendlier to them than the average Canadian law school, they should be allowed to create some law school that only accepts gays and bans homophobes and teaches lots of courses on gay marriage law all the time.

Another person on the Facebook thread complained that this line of arguments leads to being okay with white separatists. And so it does. Fine. I think white separatists have *exactly* the right position about where the sort of white people who want to be white separatists should be relative to everyone else – separate. I am not sure what you think you are gaining by demanding that white separatists live in communities with a lot of black people in them, but I bet the black people in those communities aren’t thanking you. Why would they want a white separatist as a neighbor? Why should they have to have one?

If people want to go do their own thing in a way that harms no one else, you *let* them. That’s the Archipelagian way.

(someone will protest that Archipelagian voluntary freedom of association or disassociation could, in cases of enough racial prejudice, lead to segregation, and that segregation didn’t work. Indeed it didn’t. But I feel like a version of segregation in which black people actually had the legally mandated right to get away from white people and remain completely unmolested by them – and where a white-controlled government wasn’t in charge of divvying up resources between white and black communities – would have worked a lot better than the segregation we actually had. The segregation we actually *had* was one in which white and black communities were separate until white people wanted something from black people, at which case they waltzed in and took it. If communities were actually totally separate, government and everything, by definition it would be impossible for one to oppress the other. The black community might start with less, but that could be solved by some kind of reparations. The Archipelagian way of dealing with this issue would be for white separatists to have separate white communities, black separatists to have separate black communities, integrationists to have integrated communities, resdistributive taxation from wealthier communities going into less wealthy ones, and a strong central government ruthlessly enforcing laws against any community trying to hurt another. I don’t think there’s a single black person in the segregation-era South who wouldn't have taken that deal, and any black person who thinks the effect of whites on their community today is net negative should be pretty interested as well.)

This is one reason I find people who hate seasteads so distasteful. I mean, here's what Reuters has to say about seasteading:

> Fringe movements, of course, rarely cast themselves as obviously fringe. Racist, anti-civil rights forces cloaked themselves in the benign language of "state's rights". Anti-gay religious entities adopted the glossy, positive imagery of "family values". Similarly, though many Libertarians embrace a pseudo-patriotic apple pie nostalgia, behind this façade is a very un-American, sinister vision.
>
> Sure, most libertarians may not want to do away entirely with the idea of government or, for that matter, government-protected rights and civil liberties. But many do — and ironically vie for political power in a nation they ultimately want to destroy. Even the right-wing pundit Ann Coulter mocked the paradox of Libertarian candidates: "Get rid of government — but first, make me president!" Libertarians sowed the seeds of anti-government discontent, which is on the rise, and now want to harvest that discontent for a very radical, anti-America agenda. The image of libertarians living off-shore in their lawless private nation-states is just a postcard of the future they hope to build on land.
>
> Strangely, the libertarian agenda has largely escaped scrutiny, at least compared to that of social conservatives. The fact that the political class is locked in debate about whether Michele Bachmann or Rick Perry is more socially conservative only creates a

> veneer of mainstream legitimacy for the likes of Ron Paul, whose libertarianism may be even more extreme and dangerously un-patriotic. With any luck America will recognize anti-government extremism for what it is — before libertarians throw America overboard and render us all castaways.

Keep in mind this is because *some people want to go off and do their own thing in the middle of the ocean far away from everyone else without bothering anyone*. And the newspapers are trying to whip up a panic about "throwing America overboard".

So one way we could become more Archipelagian is just *trying not to yell at people who are trying to go off and doing their own thing quietly with a group of voluntarily consenting friends*.

But I think a better candidate for how to build a more Archipelagian world is to encourage the fracture of society into subcultures.

Like, transsexuals may not be able to go to a transsexual island somewhere and build Transtopia where anyone who misgenders anyone else gets thrown into a volcano. But of the transsexuals I know, a lot of them have lots of transsexual friends, their cissexual friends are all up-to-date on trans issues and don't do a lot of misgendering, and they have great social networks where they share information about what businesses and doctors are or aren't trans-friendly. They can take advantage of trigger warnings to make sure they expose themselves to only the sources that fit the values of their community, the information that would get broadcast if it was a normal community that could impose media norms. As Internet interaction starts to replace real-life interaction (and I think for a lot of people the majority of their social life is already on the Internet, and for some the majority of their economic life is as well) it becomes increasingly easy to limit yourself to transsexual-friendly spaces that keep bad people away.

The rationalist community is another good example. If I wanted, I could move to the Bay Area tomorrow and never have more than a tiny amount of contact with non-rationalists again. I could have rationalist roommates, live in a rationalist group house, try to date only other rationalists, try to get a job with a rationalist nonprofit like CFAR or a rationalist company like Quixey, and never have to deal with the benighted and depressing non-rationalist world again. Even without moving to the Bay Area, it’s been pretty easy for me to keep a lot of my social life, both on- and off- line, rationalist-focused, and I don’t regret this at all.

I don’t know if the future will be virtual reality. I expect the post-singularity future will include something like VR, although that might be like describing teleportation as “basically a sort of pack animal”. But how much the immediate pre-singularity world will make use of virtual reality, I don’t know.

But I bet if it doesn’t, it will be because virtual reality has been circumvented by things like social networks, bitcoin, and Mechanical Turk, which make it possible to do most of your interaction through the Internet even though you’re not literally plugged into it.

And that seems to me like a pretty good start in creating an Archipelago. I already hang out with various Finns and Brits and Aussies a lot more closely than I do my next-door neighbors, and if we start using litecoin and someone else starts using dogecoin then I’ll be more economically connected to them too. The degree to which I encounter certain objectifying or unvirtuous or triggering media already depends more on the moderation policies of Less Wrong and Slate Star Codex and who I block from my Facebook feed, than it does any laws about censorship of US media.

At what point are national governments rendered mostly irrelevant compared to the norms and rules of the groups of which we are voluntary members?

I don’t know, but I kind of look forward to finding out. It seems like a great way to start searching for utopia, or at least getting some people away from their metaphorical abusive-husbands.

And the other thing is that I have pretty strong opinions on which communities are better than others. Some communities were founded by toxic people for ganging up with other toxic people to celebrate and magnify their toxicity, and these (surprise, surprise) tend to be toxic. Others were formed by very careful, easily-harmed people trying to exclude everyone who could harm them, and these tend to be pretty safe albeit sometimes overbearing. Other people hit some kind of sweet spot that makes friendly people want to come in and angry people want to stay out, or just do a really good job choosing friends.

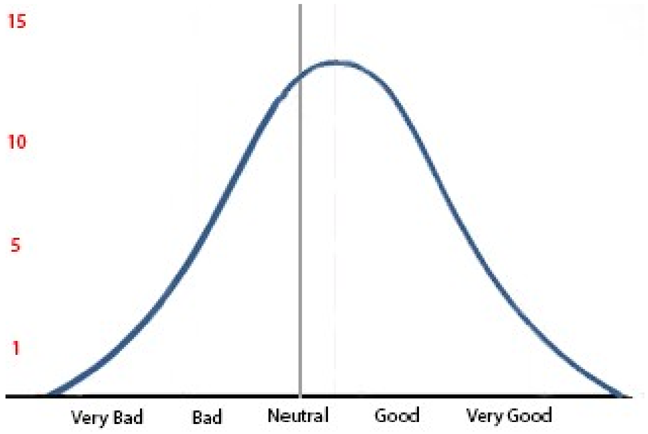
But I think the end result is that the closer you come to true freedom of association, the closer you get to a world where everyone is a member of more or less the community they deserve. That would be a pretty unprecedented bit of progress.

Beware The Man Of One Study

Aquinas famously said: beware the man of one book. I would add: beware the man of one study.

For example, take medical research. Suppose a certain drug is weakly effective against a certain disease. After a few years, a bunch of different research groups have gotten their hands on it and done all sorts of different studies. In the best case scenario the average study will find the true result – that it's weakly effective.

But there will also be random noise caused by inevitable variation and by some of the experiments being better quality than others. In the end, we might expect something looking kind of like a bell curve. The peak will be at "weakly effective", but there will be a few studies to either side. Something like this:



We see that the peak of the curve is somewhere to the right of neutral – ie weakly effective – and that there are about 15 studies that find this correct result.

But there are also about 5 studies that find that the drug is very good, and 5 studies missing the sign entirely and finding that the drug is actively bad. There’s even 1 study finding that the drug is very bad, maybe seriously dangerous.

This is before we get into fraud or statistical malpractice. I’m saying this is what’s going to happen just by normal variation in experimental design. As we increase experimental rigor, the bell curve might get squashed horizontally, but there will still be a bell curve.

In practice it’s worse than this, because this is assuming everyone is investigating exactly the same question.

Suppose that the graph is titled “Effectiveness Of This Drug In Treating Bipolar Disorder”.

But maybe the drug is more effective in bipolar i than in bipolar ii (Depakote, for example)

Or maybe the drug is very effective against bipolar mania, but much less effective against bipolar depression (Depakote again).

Or maybe the drug is a good acute antimanic agent, but less effective at maintenance treatment (let’s stick with Depakote).

If you have a graph titled “Effectiveness Of Depakote In Treating Bipolar Disorder” plotting studies from “Very Bad” to “Very Good” – and you stick all the studies – maintenence, manic, depressive, bipolar i, bipolar ii – on the graph, then you’re going to end running the gamut from “very bad” to “very good” even before you factor in noise and even before even before you factor in bias and poor experimental design.

So here’s why you should beware the man of one study.

If you go to your better class of alternative medicine websites, they don't tell you "Studies are a logocentric phallocentric tool of Western medicine and the Big Pharma conspiracy."

They tell you "medical science has proved that this drug is terrible, but ignorant doctors are pushing it on you anyway. Look, here's a study by a reputable institution proving that the drug is not only ineffective, but harmful."

And the study will exist, and the authors will be prestigious scientists, and it will probably be about as rigorous and well-done as any other study.

And then a lot of people raised on the idea that some things have Evidence and other things have No Evidence think *holy s**t, they're right!*

On the other hand, your doctor isn't going to a sketchy alternative medicine website. She's examining the entire literature and extracting careful and well-informed conclusions from…

Haha, just kidding. She's going to a luncheon at a really nice restaurant sponsored by a pharmaceutical company, which assures her that they would *never* take advantage of such an opportunity to shill their drug, they just want to raise awareness of the latest study. And the latest study shows that their drug is great! Super great! And your doctor nods along, because the authors of the study are prestigious scientists, and it's about as rigorous and well-done as any other study.

But obviously the pharmaceutical company has selected one of the studies from the "very good" end of the bell curve.

And I called this “Beware The Man of One Study”, but it’s easy to see that in the little diagram there are like three or four studies showing that the drug is “very good”, so if your doctor is a little skeptical, the pharmaceutical company can say “You are right to be skeptical, one study doesn’t prove anything, but look – here’s another group that finds the same thing, here’s yet another group that finds the same thing, and here’s a replication that confirms both of them.”

And even though it looks like in our example the sketchy alternative medicine website only has one “very bad” study to go off of, they could easily supplement it with a bunch of merely “bad” studies. Or they could add all of those studies about slightly different things. Depakote is ineffective at treating bipolar depression. Depakote is ineffective at maintenance bipolar therapy. Depakote is ineffective at bipolar ii.

So just sum it up as “Smith et al 1987 found the drug ineffective, yet doctors continue to prescribe it anyway”. Even if you hunt down the original study (which no one does), Smith et al won’t say specifically “Do remember that this study is only looking at bipolar maintenance, which is a different topic from bipolar acute antimanic treatment, and we’re not saying anything about that.” It will just be titled something like “Depakote fails to separate from placebo in six month trial of 91 patients” and trust that the responsible professionals reading it are well aware of the difference between acute and maintenance treatments (hahahahaha).

So it’s not so much “beware the man of one study” as “beware the man of any number of studies less than a relatively complete and not-cherry-picked survey of the research”.

II.

I think medical science is still pretty healthy, and that the consensus of doctors and researchers is more-or-less right on most controversial medical issues.

(it’s the *uncontroversial* ones you have to worry about)

Politics doesn’t have this protection.

Like, take the minimum wage question (please). We all know about the Krueger and Card study in New Jersey that found no evidence that high minimum wages hurt the economy. We probably also know the counterclaims that it was completely debunked as despicable dishonest statistical malpractice. Maybe some of us know Card and Krueger wrote a pretty convincing rebuttal of those claims. Or that a bunch of large and methodologically advanced studies have come out since then, some finding no effect like Dube, others finding strong effects like Rubinstein and Wither. These are just examples; there are at least dozens and probably hundreds of studies on both sides.

But we can solve this with meta-analyses and systemtic reviews, right?

Depends which one you want. Do you go with this meta-analysis of fourteen studies that shows that any presumed negative effect of high minimum wages is likely publication bias? With this meta-analysis of sixty-four studies that finds the same thing and discovers no effect of minimum wage after correcting for the problem? Or how about this meta-analysis of fifty-five countries that does find effects in most of them? Maybe you prefer this systematic review of a hundred or so studies that finds strong and consistent effects?

Can we trust news sources, think tanks, econblogs, and other institutions to sum up the state of the evidence?

CNN claims that 85% of credible studies have shown the minimum wage causes job loss. But raisetheminimumwage.com declares that “two decades of rigorous economic research have found that raising the minimum wage does not result in job loss…researchers and businesses alike agree today that the weight of the evidence shows no reduction in employment resulting from minimum wage increases.” Modeled Behavior says “the majority of the new minimum wage research supports the hypothesis that the minimum wage increases unemployment.” The Center for Budget and Policy Priorities says “The common claim that raising the minimum wage reduces employment for low-wage workers is one of the most extensively studied issues in empirical economics. The weight of the evidence is that such impacts are small to none.”

Okay, fine. What about economists? They seem like experts. What do they think?

Well, five hundred economists signed a letter to policy makers saying that the science of economics shows increasing the minimum wage would be a bad idea. That sounds like a promising consensus…

..except that six hundred economists signed a letter to policy makers saying that the science of economics shows increasing the minimum wage would be a *good* idea. (h/t Greg Mankiw)

Fine then. Let’s do a formal survey of economists. Now what?

raisetheminimumwage.com, an unbiased source if ever there was one, confidently tells us that “indicative is a 2013 survey by the University of Chicago’s Booth School of Business in which leading economists agreed by a nearly 4 to 1 margin that the benefits of raising and indexing the minimum wage outweigh the costs.”

But the Employment Policies Institute, which sounds like it's trying *way* too hard to sound like an unbiased source, tells us that "Over 73 percent of AEA labor economists believe that a significant increase will lead to employment losses and 68 percent think these employment losses fall disproportionately on the least skilled. Only 6 percent feel that minimum wage hikes are an efficient way to alleviate poverty."

So the whole thing is fiendishly complicated. But unless you look very very hard, you will never know that.

If you are a conservative, what you will find on the sites you trust will be something like this:

> Economic theory has always shown that minimum wage increases decrease employment, but the Left has never been willing to accept this basic fact. In 1992, they trumpeted a single study by Card and Krueger that purported to show no negative effects from a minimum wage increase. This study was immediately debunked and found to be based on statistical malpractice and "massaging the numbers". Since then, dozens of studies have come out confirming what we knew all along – that a high minimum wage is economic suicide. Systematic reviews and meta-analyses (Neumark 2006, Boockman 2010) consistently show that an overwhelming majority of the research agrees on this fact – as do 73% of economists. That's why five hundred top economists recently signed a letter urging policy makers not to buy into discredited liberal minimum wage theories. Instead of listening to starry-eyed liberal woo, listen to the empirical evidence and an overwhelming majority of economists and oppose a raise in the minimum wage.

And if you are a leftist, what you will find on the sites you trust will be something like this:

> People used to believe that the minimum wage decreased unemployment. But Card and Krueger's famous 1992 study exploded that conventional wisdom. Since then, the results have been replicated over fifty times, and further meta-analyses (Card and Krueger 1995, Dube 2010) have found no evidence of any effect. Leading economists agree by a 4 to 1 margin that the benefits of raising the minimum wage outweigh the costs, and that's why more than 600 of them have signed a petition telling the government to do exactly that. Instead of listening to conservative scare tactics based on long-debunked theories, listen to the empirical evidence and the overwhelming majority of economists and support a raise in the minimum wage.

Go ahead. Google the issue and see what stuff comes up. If it doesn't quite match what I said above, it's usually because they can't even muster *that* level of scholarship. Half the sites just cite Card and Krueger and call it a day!

These sites with their long lists of studies and experts are super convincing. And half of them are wrong.

At some point in their education, most smart people usually learn not to credit arguments from authority. If someone says "Believe me about the minimum wage because I seem like a trustworthy guy," most of them will have at least one neuron in their head that says "I should ask for some evidence". If they're *really* smart, they'll use the magic words "peer-reviewed experimental studies."

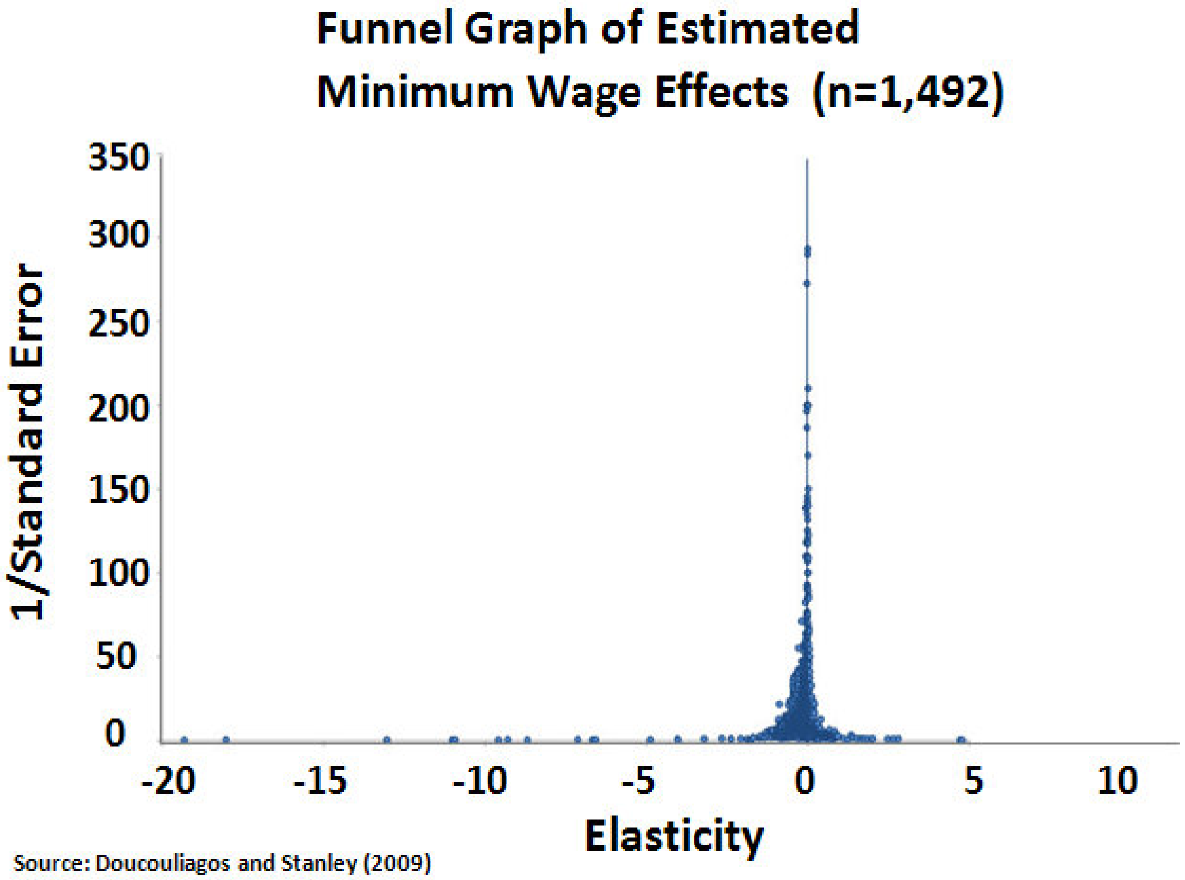
But I worry that most smart people have *not* learned that a list of dozens of studies, several meta-analyses, hundreds of experts, and expert surveys showing almost all academics support your thesis – can *still* be bullshit.

Which is too bad, because that's exactly what people who want to bamboozle an educated audience are going to use.

III.

I do not want to preach radical skepticism.

For example, on the minimum wage issue, I notice only one side has presented a funnel plot. A funnel plot is usually used to investigate publication bias, but it has another use as well – it's pretty much an exact presentation of the "bell curve" we talked about above.



Source: Doucouliagos and Stanley (2009)

This is more of a needle curve than a bell curve, but the point still stands. We see it's centered around 0, which means there's some evidence that's the real signal among all this noise. The bell skews more to left than to the right, which means more studies have found negative effects of the minimum wage than positive effects of the minimum wage. But since the bell curve is asymmetrical, we intepret that as *probably* publication bias. So all in all, I think there's at least some evidence that the liberals are right on this one.

Unless, of course, someone has realized that I've wised up to the studies and meta-analyses and and expert surveys, and figured out a way to hack *funnel plots*, which I am totally not ruling out.

(okay, I *kind of* want to preach radical skepticism)

Also, I should probably mention that it's much more complicated than one side being right, and that the minimum wage probably works differently depending on what industry you're talking about, whether it's state wage or federal wage, whether it's a recession or a boom, whether we're talking about increasing from $5 to $6 or from $20 to $30, etc, etc, etc. There are eleven studies on that plot showing an effect even worse than -5, and very possibly they are all accurate for whatever subproblem they have chosen to study – much like the example with Depakote where it might an effective antimanic but a terrible antidepressant.

(radical skepticism actually sounds a lot better than figuring this all out).

IV.

But the question remains: what happens when (like in most cases) you don't have a funnel plot?

I don't have a good positive answer. I do have several good *negative* answers.

Decrease your confidence about most things if you’re not sure that you’ve investigated every piece of evidence.

Do not trust websites which are obviously biased (eg Free Republic, Daily Kos, Dr. Oz) when they tell you they’re going to give you “the state of the evidence” on a certain issue, even if the evidence seems very stately indeed. This goes double for any site that contains a list of “myths and facts about X”, quadruple for any site that uses phrases like “ingroup member uses actual FACTS to DEMOLISH the outgroup’s lies about Y”, and octuple for RationalWiki.

Most important, even if someone gives you what seems like overwhelming evidence in favor of a certain point of view, don’t trust it until you’ve done a simple Google search to see if the opposite side has equally overwhelming evidence.

Book Review: Albion's Seed

I.

Albion's Seed by David Fischer is a history professor's nine-hundred-page treatise on patterns of early immigration to the Eastern United States. It's not light reading and not the sort of thing I would normally pick up. I read it anyway on the advice of people who kept telling me it explains everything about America. And it sort of does.

In school, we tend to think of the original American colonists as "Englishmen", a maximally non-diverse group who form the background for all of the diversity and ethnic conflict to come later. Fischer's thesis is the opposite. Different parts of the country were settled by very different groups of Englishmen with different regional backgrounds, religions, social classes, and philosophies. The colonization process essentially extracted a single stratum of English society, isolated it from all the others, and then plunked it down on its own somewhere in the Eastern US.

I used to play *Alpha Centauri*, a computer game about the colonization of its namesake star system. One of the dynamics that made it so interesting was its backstory, where a Puerto Rican survivalist, an African plutocrat, and other colorful characters organized their own colonial expeditions and competed to seize territory and resources. You got to explore not only the settlement of a new world, but the settlement of a new world by societies dominated by extreme founder effects. What kind of weird pathologies and wonderful innovations do you get when a group of overly romantic Scottish environmentalists is allowed to develop on its own trajectory free of all non-overly-romantic-Scottish-environmentalist influences? *Albion's Seed* argues that this is basically the process that formed several early US states.

Fischer describes four of these migrations: the Puritans to New England in the 1620s, the Cavaliers to Virginia in the 1640s, the Quakers to Pennsylvania in the 1670s, and the Borderers to Appalachia in the 1700s.

II.

A: The Puritans

I hear about these people every Thanksgiving, then never think about them again for the next 364 days. They were a Calvinist sect that dissented against the Church of England and followed their own brand of dour, industrious, fun-hating Christianity. Most of them were from East Anglia, the part of England just northeast of London. They came to America partly because they felt persecuted, but mostly because they thought England was full of sin and they were at risk of absorbing the sin by osmosis if they didn't get away quick and build something better. They really liked "city on a hill" metaphors.

I knew about the Mayflower, I knew about the black hats and silly shoes, I even knew about the time Squanto threatened to release a bioweapon buried under Plymouth Rock that would bring about the apocalypse. But I didn't know that the Puritan migration to America was basically a eugenicist's wet dream.

Much like eg Unitarians today, the Puritans were a religious group that drew disproportionately from the most educated and education-obsessed parts of the English populace. Literacy among immigrants to Massachusetts was twice as high as the English average, and in an age when the vast majority of Europeans were farmers most immigrants to

Massachusetts were skilled craftsmen or scholars. And the Puritan "homeland" of East Anglia was a an unusually intellectual place, with strong influences from Dutch and Continental trade; historian Havelock Ellis finds that it "accounts for a much larger proportion of literary, scientific, and intellectual achievement than any other part of England."

Furthermore, only the best Puritans were allowed to go to Massachusetts; Fischer writes that "it may have been the only English colony that required some of its immigrants to submit letters of recommendation" and that "those who did not fit in were banished to other colonies and sent back to England". Puritan "headhunters" went back to England to recruit "godly men" and "honest men" who "must not be of the poorer sort".

INTERESTING PURITAN FACTS:

1. Sir Harry Vane, who was "briefly governor of Massachusetts at the age of 24", "was so rigorous in his Puritanism that he believed only the thrice-born to be truly saved".

2. The great seal of the Massachusetts Bay Company "featured an Indian with arms beckoning, and five English words flowing from his mouth: 'Come over and help us'"

3. Northern New Jersey was settled by Puritans who named their town after the "New Ark Of The Covenant" – modern Newark.

4. Massachusetts clergy were very powerful; Fischer records the story of a traveller asking a man "Are you the parson who serves here?" only to be corrected "I am, sir, the parson who *rules* here."

5. The Puritans tried to import African slaves, but they all died of the cold.

6. In 1639, Massachusetts declared a "Day Of Humiliation" to condemn "novelties, oppression, atheism, excesse, superfluity, idleness, contempt of authority, and trouble in other parts to be remembered”

7. The average family size in Waltham, Massachusetts in the 1730s was 9.7 children.

8. Everyone was compelled by law to live in families. Town officials would search the town for single people and, if found, order them to join a family; if they refused, they were sent to jail.

9. 98% of adult Puritan men were married, compared to only 73% of adult Englishmen in general. Women were under special pressure to marry, and a Puritan proverb said that “women dying maids lead apes in Hell”.

10. 90% of Puritan names were taken from the Bible. Some Puritans took pride in their learning by giving their children obscure Biblical names they would expect nobody else to have heard of, like Mahershalalhasbaz. Others chose random Biblical terms that might not have technically been intended as names; “the son of Bostonian Samuel Pond was named Mene Mene Tekel Upharsin Pond”. Still others chose Biblical words completely at random and named their children things like Maybe or Notwithstanding.

11. Puritan parents traditionally would send children away to be raised with other families, and raise those families’ children in turn, in the hopes that the lack of familiarity would make the child behave better.

12. In 1692, 25% of women over age 45 in Essex County were accused of witchcraft.

13. Massachusetts passed the first law mandating universal public education, which was called The Old Deluder Satan Law in honor of its preamble, which began “It being one chief project of that old deluder, Satan, to keep men from the knowledge of the scriptures…”

14. Massachusetts cuisine was based around “meat and vegetables submerged in plain water and boiled relentlessly without seasonings of any kind”.

15. Along with the famous scarlet A for adultery, Puritans could be forced to wear a B for blasphemy, C for counterfeiting, D for drunkenness, and so on.

16. Wasting time in Massachusetts was literally a criminal offense, listed in the law code, and several people were in fact prosecuted for it.

17. This wasn't even the nadir of weird hard-to-enforce Massachusetts laws. Another law just said "If any man shall exceed the bounds of moderation, we shall punish him severely".

Harriet Beecher Stowe wrote of Massachusetts Puritanism: "The underlying foundation of life in New England was one of profound, unutterable, and therefore unuttered mehalncholy, which regarded human existence itself as a ghastly risk, and, in the case of the vast majority of human beings, an inconceivable misfortune." And indeed, everything was dour, strict, oppressive, and very religious. A typical Massachusetts week would begin in the church, which doubled as the town meeting hall. There were no decorations except a giant staring eye on the pulpit to remind churchgoers that God was watching them. Townspeople would stand up before their and declare their shame and misdeeds, sometimes being forced to literally crawl before the other worshippers begging for forgiveness. Then the minister would give two two-hour sermons back to back. The entire affair would take up to six hours, and the church was unheated (for some reason they stored all their gunpowder there, so no one was allowed to light a fire), and this was Massachusetts, and it was colder in those days than it is now, so that during winter some people would literally lose fingers to frostbite (Fischer: "It was a point of honor for the minister never to shorten a sermon merely because his audience was frozen"). Everyone would stand there with their guns (they were legally required to bring guns, in case Indians attacked during the sermon) and hear about how they were going to Hell, all while the giant staring eye looked at them.

So life as a Puritan was pretty terrible. On the other hand, their society was *impressively* well-ordered. Teenage pregnancy rates were the lowest in the Western world and in some areas literally zero. Murder rates were half those in other American colonies. There was remarkably low income inequality – “the top 10% of wealthholders held only 20%-30% of taxable property”, compared to 75% today and similar numbers in other 17th-century civilizations. The poor (at least the poor native to a given town) were treated with charity and respect – “in Salem, one man was ordered to be set by the heels in the stocks for being uncharitable to a poor man in distress”. Government was conducted through town meetings in which everyone had a say. Women had more equality than in most parts of the world, and domestic abuse was punished brutally. The educational system was top-notch – “by most empirical tests of intellectual eminence, New England led all other parts of British America from the 17th to the early 20th century”.

In some ways the Puritans seem to have taken the classic dystopian bargain – give up all freedom and individuality and art, and you can have a perfect society without crime or violence or inequality. Fischer ends each of his chapters with a discussion of how the society thought of liberty, and the Puritans unsurprisingly thought of liberty as “ordered liberty” – the freedom of everything to tend to its correct place and stay there. They thought of it as a freedom from disruption – apparently FDR stole some of his “freedom from fear” stuff from early Puritan documents. They were extremely not in favor of the sort of liberty that meant that, for example, there wouldn’t be laws against wasting time. *That* was going too far.

B: The Cavaliers

The Massachusetts Puritans fled England in the 1620s partly because the king and nobles were oppressing them. In the 1640s, English Puritans under Oliver Cromwell rebelled, took over the government, and killed the king. The nobles not unreasonably started looking to get the heck out.

Virginia had been kind of a wreck ever since the original Jamestown settlers had mostly died of disease. Governor William Berkeley, a noble himself, decided the colony could reinvent itself as a destination for refugee nobles, and told them it would do everything possible to help them maintain the position of oppressive supremacy to which they were accustomed. The British nobility was sold. The Cavaliers – the nobles who had fought and lost the English Civil War – fled to Virginia. Historians cross-checking Virginian immigrant lists against English records find that of Virginians whose opinions on the War were known, 98% were royalists. They were overwhelmingly Anglican, mostly from agrarian southern England, and all related to each other in the incestuous way of nobility everywhere: “it is difficult to think of any ruling elite that has been more closely interrelated since the Ptolemies”. There were twelve members of Virginia’s royal council; in 1724 “all without exception were related to one another by blood or marriage…as late as 1775, every member of that august body was descended from a councilor who had served in 1660”.

These aristocrats didn’t want to do their own work, so they brought with them tens of thousands of indentured servants; more than 75% of all Virginian immigrants arrived in this position. Some of these people came willingly on a system where their master paid their passage over and they would be free after a certain number of years; others were sent by the courts as punishments; still others were just plain kidnapped. The gender ratio was 4:1 in favor of men, and there were entire English gangs dedicated to kidnapping women and sending them to Virginia, where they fetched a high price. Needless to say, these people came from a very different stratum than their masters *or* the Puritans.

People who came to Virginia mostly died. They died of malaria, typhoid fever, amoebiasis, and dysentery. Unlike in New England, where Europeans were better adapted to the cold climate than Africans, in Virginia it was Europeans who had the higher disease-related mortality rate. The whites who survived tended to become “sluggish and indolent”, according to the universal report of travellers and chroniclers, although I might be sluggish and indolent too if I had been kidnapped to go work on some rich person’s farm and sluggishness/indolence was an option.

The Virginians tried their best to oppress white people. Really, they did. The depths to which they sank in trying to oppress white people almost boggle the imagination. There was a rule that if a female indentured servant became pregnant, a few extra years were added on to their indenture, supposedly because they would be working less hard during their pregnancy and child-rearing so it wasn’t fair to the master. Virginian aristocrats would *rape their own female servants*, then add a penalty term on to their indenture for becoming pregnant. That is an *impressive* level of chutzpah. But despite these efforts, eventually all the white people either died, or became too sluggish to be useful, or worst of all just finished up their indentures and became legally free. The aristocrats started importing black slaves as per the model that had sprung up in the Caribbean, and so the stage was set for the antebellum South we read about in history classes.

INTERESTING CAVALIER FACTS:

1. Virginian cavalier speech patterns sound a lot like modern African-American dialects. It doesn't take much imagination to figure out why, but it's strange to think of a 17th century British lord speaking what a modern ear would clearly recognize as Ebonics.

2. Three-quarters of 17th-century Virginian children lost at least one parent before turning 18.

3. Cousin marriage was an important custom that helped cement bonds among the Virginian elite, "and many an Anglican lady changed her condition but not her name".

4. In Virginia, women were sometimes unironically called "breeders"; English women were sometimes referred to as "She-Britons".

5. Virginia didn't really have towns; the Chesapeake Bay was such a giant maze of rivers and estuaries and waterways that there wasn't much need for land transport hubs. Instead, the unit of settlement was the plantation, which consisted of an aristocratic planter, his wife and family, his servants, his slaves, and a bunch of guests who hung around and mooched off him in accordance with the ancient custom of hospitality.

6. Virginian society considered everyone who lived in a plantation home to be a kind of "family", with the aristocrat both as the literal father and as a sort of abstracted patriarch with complete control over his domain.

7. Virginia governor William Berkeley probably would not be described by moderns as 'strong on education'. He said in a speech that "I thank God there are no free schools nor printing [in Virginia], and I hope we shall not have these for a hundred years, for learning has brought disobedience, and heresy, and sects into the world, and printing has divuldged them, and libels against the best government. God keep us from both!"

8. Virginian recreation mostly revolved around hunting and bloodsports. Great lords hunted deer, lesser gentry hunted foxes, indentured servants had a weird game in which they essentially draw-and-quartered geese, young children “killed and tortured songbirds”, and “at the bottom of this hierarchy of bloody games were male infants who prepared themselves for the larger pleasures of maturity by torturing snakes, maiming frogs, and pulling the wings off butterflies. Thus, every red-blooded male in Virginia was permitted to slaughter some animal or other, and the size of his victim was proportioned to his social rank.”

9. “In 1747, an Anglican minister named William Kay infuriated the great planter Landon Carter by preaching a sermon against pride. The planter took it personally and sent his [relations] and ordered them to nail up the doors and windows of all the churches in which Kay preached.”

10. Our word “condescension” comes from a ritual attitude that leading Virginians were supposed to display to their inferiors. Originally condescension was supposed to be a polite way of showing respect those who were socially inferior to you; our modern use of the term probably says a lot about what Virginians actually did with it.

In a lot of ways, Virginia was the opposite of Massachusetts. Their homicide rate was sky-high, and people were actively encouraged to respond to slights against their honor with duels (for the rich) and violence (for the poor). They were obsessed with gambling, and “made bets not merely on horses, cards, cockfights, and backgammon, but also on crops, prices, women, and the weather”. Their cuisine focused on gigantic sumptuous feasts of animals killed in horrible ways. There were no witchcraft trials, but there *were* people who were fined for disrupting the peace by accusing their neighbors of witchcraft. Their church sermons were twenty minutes long on the dot.

The Puritans naturally thought of the Virginians as completely lawless reprobate sinners, but this is not *entirely* true. Virginian church sermons might have been twenty minutes long, but Virginian ballroom dance lessons could last nine hours. It wasn't that the Virginians weren't bound by codes, just that those codes were social rather than moral.

And Virginian nobles weren't just random jerks, they were *carefully cultivated* jerks. Planters spared no expense to train their sons to be strong, forceful, and not take nothin' from nobody. They would encourage and reward children for being loud and temperamental, on the grounds that this indicated a strong personality and having a strong personality was fitting of a noble. When this worked, it worked *really* well – witness natural leaders and self-driven polymaths like George Washington and Thomas Jefferson. More often it failed catastrophically – the rate of sex predation and rape in Virginia was at least as high as anywhere else in North America.

The Virginian Cavaliers had an obsession with liberty, but needless to say it was not exactly a sort of liberty of which the ACLU would approve. I once heard someone argue against libertarians like so: even if the government did not infringe on liberties, we would still be unfree for other reasons. If we had to work, we would be subject to the whim of bosses. If we were poor, we would not be "free" to purchase most of the things we want. In any case, we are "oppressed" by disease, famine, and many other things besides government that prevent us from implementing our ideal existence.

The Virginians took this idea and ran with it – in the wrong direction. No, they said, we *wouldn't* be free if we had to work, therefore we insist upon not working. No, we *wouldn't* be free if we were limited by poverty, therefore we insist upon being extremely rich. Needless to say, this conception of freedom required first indentured servitude and later slavery to make it work, but the Virginians never *claimed* that the servants or slaves were free. That wasn't the point. Freedom, like wealth, was properly distributed according to rank; nobles had as much as they wanted, the middle-class enough to get by on, and everyone else none at all. And a Virginian noble would have gone to his grave insisting that a civilization without slavery could never have citizens who were truly free.

C: The Quakers

Fischer warns against the temptation to think of the Quakers as normal modern people, but he has to warn us precisely because it's so tempting. Where the Puritans seem like a dystopian caricature of virtue and the Cavaliers like a dystopian caricature of vice, the Quakers just seem *ordinary*. Yes, they're kind of a religious cult, but they're the kind of religious cult any of us might found if we were thrown back to the seventeenth century.

Instead they were founded by a weaver's son named George Fox. He believed people were basically good and had an Inner Light that connected them directly to God without a need for priesthood, ritual, Bible study, or self-denial; mostly people just needed to listen to their consciences and be nice. Since everyone was equal before God, there was no point in holding up distinctions between lords and commoners: Quakers would just address everybody as "Friend". And since the Quakers were among the most persecuted sects at the time, they developed an insistence on tolerance and freedom of religion which (unlike the Puritans) they stuck to even when shifting fortunes put them on top. They believed in pacificism, equality of the sexes, racial harmony, and a bunch of other things which seem pretty hippy-ish even today let alone in 1650.

England’s top Quaker in the late 1600s was William Penn. Penn is universally known to Americans as “that guy Pennsylvania is named after” but actually was a larger-than-life 17th century superman. Born to the nobility, Penn distinguished himself early on as a military officer; he was known for beating legendary duelists in single combat and then sparing their lives with sermons about how murder was wrong. He gradually started having mystical visions, quit the military, and converted to Quakerism. Like many Quakers he was arrested for blasphemy; unlike many Quakers, they couldn’t make the conviction stick; in his trial he “conducted his defense so brilliantly that the jurors refused to convict him even when threatened with prison themselves, [and] the case became a landmark in the history of trial by jury.” When the state finally found a pretext on which to throw him in prison, he spent his incarceration composing “one of the noblest defenses of religious liberty ever written”, conducting a successful mail-based courtship with England’s most eligible noblewoman, and somehow gaining the personal friendship and admiration of King Charles II. Upon his release the King liked him so much that he gave him a large chunk of the Eastern United States on a flimsy pretext of repaying a family debt. Penn didn’t *want* to name his new territory Pennsylvania – he recommended just “Sylvania” – but everybody else overruled him and Pennyslvania it was. The grant wasn’t quite the same as the modern state, but a chunk of land around the Delaware River Valley – what today we would call eastern Pennsylvania, northern Delaware, southern New Jersey, and bits of Maryland – centered on the obviously-named-by-Quakers city of Philadelphia.

Penn decided his new territory would be a Quaker refuge – his exact wording was “a colony of Heaven [for] the children of the Light”. He mandated universal religious toleration, a total ban on military activity, and a government based on checks and balances that would “leave myself and successors no power of doing mischief, that the will of one man may not hinder the good of a whole country”.

His recruits – about 20,000 people in total – were Quakers from the north of England, many of them minor merchants and traders. They disproportionately included the Britons of Norse descent common in that region, who formed a separate stratum and had never really gotten along with the rest of the British population. They were joined by several German sects close enough to Quakers that they felt at home there; these became the ancestors of (among other groups) the Pennsylvania Dutch, Amish, and Mennonites.

INTERESTING QUAKER FACTS:

1. In 1690 a gang of pirates stole a ship in Philadelphia and went up and down the Delaware River stealing and plundering. The Quakers got in a heated (but brotherly) debate about whether it was morally permissible to use violence to stop them. When the government finally decided to take action, contrarian minister George Keith dissented and caused a major schism in the faith.

2. Fischer argues that the Quaker ban on military activity within their territory would have doomed them in most other American regions, but by extreme good luck the Indians in the Delaware Valley were almost as peaceful as the Quakers. As usual, at least some credit goes to William Penn, who taught himself Algonquin so he could negotiate with the Indians in their own language.

3. The Quakers’ marriage customs combined a surprisingly modern ideas of romance, with extreme bureaucracy. The wedding process itself had sixteen stages, including “ask parents”, “ask community women”, “ask community men”, “community women ask parents”, and “obtain a certificate of cleanliness”. William Penn’s marriage apparently had forty-six witnesses to testify to the good conduct and non-relatedness of both parties.

4. Possibly related: 16% of Quaker women were unmarried by age 50, compared to only about 2% of Puritans.

5. Quakers promoted gender equality, including the (at the time scandalous) custom of allowing women to preach (condemned by the Puritans as the crime of “she-preaching”).

6. But they were such prudes about sex that even *the Puritans* thought they went too far. Pennsylvania doctors had problems treating Quakers because they would “delicately describe everything from neck to waist as their ‘stomachs’, and anything from waist to feet as their ‘ankles'”.

7. Quaker parents Richard and Abigail Lippincott named their eight children, in order, “Remember”, “John”, “Restore”, “Freedom”, “Increase”, “Jacob”, “Preserve”, and “Israel”, so that their names combined formed a simple prayer.

8. Quakers had surprisingly modern ideas about parenting, basically sheltering and spoiling their children at a time when everyone else was trying whip the Devil out of them.

9. “A Quaker preacher, traveling in the more complaisant colony of Maryland, came upon a party of young people who were dancing merrily together. He broke in upon them like an avenging angel, stopped the dance, anddemanded to know if they considered Martin Luther to be a good man. The astonished youngsters answered in the affirmative. The Quaker evangelist then quoted Luther on the subject of dancing: ‘as many paces as the man takes in his dance, so many steps he takes toward Hell. This, the Quaker missionary gloated with a gleam of sadistic satisfaction, ‘spoiled their sport’.”

10. William Penn wrote about thirty books defending liberty of conscience throughout his life. The Quaker obsession with the individual conscience as the work of God helped invent the modern idea of conscientious objection.

11. Quakers were heavily (and uniquely for their period) opposed to animal cruelty. When foreigners introduced bullbaiting into Philadelphia during the 1700s, the mayor bought a ticket supposedly as a spectator. When the event was about to begin, he leapt into the ring, personally set the bull free, and threatened to arrest anybody who stopped him.

12. On the other hand, they were also opposed to other sports for what seem like kind of random reasons. The town of Morley declared an anathema against foot races, saying that they were “unfruitful works of darkness”.

13. The Pennsylvania Quakers became very prosperous merchants and traders. They also had a policy of loaning money at low- or zero- interest to other Quakers, which let them outcompete other, less religious businesspeople.

14. They were among the first to replace the set of bows, grovels, nods, meaningful looks, and other British customs of acknowledging rank upon greeting with a single rank-neutral equivalent – the handshake.

15. Pennsylvania was one of the first polities in the western world to abolish the death penalty.

16. The Quakers were lukewarm on education, believing that too much schooling obscured the natural Inner Light. Fischer declares it “typical of William Penn” that he wrote a book arguing against reading too much.

17. The Quakers not only instituted religious freedom, but made laws against mocking another person’s religion.

18. In the late 1600s as many as 70% of upper-class Quakers owned slaves, but Pennsylvania essentially invented modern abolitionism. Although their colonial masters in England forbade them from banning slavery outright, they applied immense social pressure and by the mid 1700s less than 10% of the wealthy had African slaves. As soon as the American Revolution started, forbidding slavery was one of independent Pennsylvania’s first actions.

Pennsylvania was *very* successful for a while; it had some of the richest farmland in the colonies, and the Quakers were exceptional merchants and traders; so much so that they were forgiven their military non-intervention during the Revolution because of their role keeping the American economy afloat in the face of British sanctions.

But by 1750, the Quakers were kind of on their way out; by 1750, they were a demographic minority in Pennsylvania, and by 1773 they were a minority in its legislature as well. In 1750 Quakerism was the third-largest religion in the US; by 1820 it was the ninth-largest, and by 1981 it was the sixty-sixth largest. What happened? The Quakers basically tolerated themselves out of existence. They were so welcoming to religious minorities and immigrants that all these groups took up shop in Pennsylvania and ended its status as a uniquely Quaker society. At the same time, the Quakers themselves became more “fanatical” and many dropped out of politics believing it to be too worldly a concern for them; this was obviously fatal to their political domination. The most famous Pennsylvanian statesman of the Revolutionary era, Benjamin Franklin, was not a Quaker at all but a first-generation immigrant from New England. Finally, Quakerism was naturally extra-susceptible to that thing where Christian denominations become indistinguishable from liberal modernity and fade into the secular background.

But Fischer argues that Quakerism continued to shape Pennsylvania long after it had stopped being officially in charge, in much the same way that Englishmen themselves have contributed disproportionately to American institutions even though they are now a numerical minority. The Pennsylvanian leadership on abolitionism, penal reform, the death penalty, and so on all happened *after* the colony was officially no longer Quaker-dominated.

And it’s hard not to see Quaker influence on the ideas of the modern US – which was after all founded in Philadelphia. In the middle of the Puritans demanding strict obedience to their dystopian hive society and the Cavaliers demanding everybody bow down to a transplanted nobility, the Pennsylvanians – who became the thought leaders of the Mid-Atlantic region including to a limited degree New York City – were pretty normal and had a good opportunity to serve as power-brokers and middlemen between the North and South. Although there are seeds of traditionally American ideas in every region, the Quakers really stand out in terms of freedom of religion, freedom of thought, checks and balances, and the idea of universal equality.

It occurs to me that William Penn might be literally the single most successful person in history. He started out as a minor noble following a religious sect that everybody despised and managed to export its principles to Pennsylvania where they flourished and multiplied. Pennsylvania then managed to export *its* principles to the United States, and the United States exported them to the world. I’m not sure how much of the suspiciously Quaker character of modern society is a direct result of William Penn, but he was in one heck of a right place at one heck of a right time

D: The Borderers

The Borderers are usually called “the Scots-Irish”, but Fischer dislikes the term because they are neither Scots (as we usually think of Scots) nor Irish (as we usually think of

Irish). Instead, they’re a bunch of people who lived on (both sides of) the Scottish-English border in the late 1600s.

None of this makes sense without realizing that the Scottish-English border was *terrible*. Every couple of years the King of England would invade Scotland or vice versa; “from the year 1040 to 1745, every English monarch but three suffered a Scottish invasion, or became an invader in his turn”. These “invasions” generally involved burning down all the border towns and killing a bunch of people there. Eventually the two sides started getting *pissed* with each other and would also torture-murder all of the enemy’s citizens they could get their hands on, ie any who were close enough to the border to reach before the enemy could send in their armies. As if this weren’t bad enough, outlaws quickly learned they could plunder one side of the border, then escape to the other before anyone brought them to justice, so the whole area basically became one giant cesspool of robbery and murder.

In response to these pressures, the border people militarized and stayed feudal long past the point where the rest of the island had started modernizing. Life consisted of farming the lands of whichever brutal warlord had the top hand today, followed by being called to fight for him on short notice, followed by a grisly death. The border people dealt with it as best they could, and developed a culture marked by extreme levels of clannishness, xenophobia, drunkenness, stubbornness, and violence.

By the end of the 1600s, the Scottish and English royal bloodlines had intermingled and the two countries were drifting closer and closer to Union. The English kings finally got some breathing room and noticed – holy frick, everything about the border is *terrible*. They decided to make the region economically productive, which meant “squeeze every cent out of the poor Borderers, in the hopes of either getting lots of money from them or else forcing them to go elsewhere and become somebody else’s problem”. Sometimes absentee landlords would just evict everyone who lived in an entire region, *en masse*, replacing them with people they expected to be easier to control.

Many of the Borderers fled to Ulster in Ireland, which England was working on colonizing as a Protestant bulwark against the Irish Catholics, and where the Crown welcomed violent warlike people as a useful addition to their Irish-Catholic-fighting project. But Ulster had some of the same problems as the Border, and also the Ulsterites started worrying that the Borderer cure was worse than the Irish Catholic disease. So the Borderers started getting kicked out of Ulster too, one thing led to another, and eventually 250,000 of these people ended up in America.

250,000 people is a *lot* of Borderers. By contrast, the great Puritan emigration wave was only 20,000 or so people; even the mighty colony of Virginia only had about 50,000 original settlers. So these people showed up on the door of the American colonies, and the American colonies collectively took one look at them and said “nope”.

Except, of course, the Quakers. The Quakers talked among themselves and decided that these people were also Children Of God, and so they should demonstrate Brotherly Love by taking them in. They tried that for a couple of years, and then they questioned their life choices and *also* said “nope”, and they told the Borderers that Philadelphia and the Delaware Valley were actually kind of full right now but there was lots of unoccupied land in *Western* Pennsylvania, and the Appalachian Mountains were very pretty at this time of year, so why didn’t they head out that way as fast as it was physically possible to go?

At the time, the Appalachians were kind of the booby prize of American colonization: hard to farm, hard to travel through, and exposed to hostile Indians. The Borderers fell in love with them. They came from a pretty marginal and unproductive territory themselves, and the Appalachians were far away from everybody and full of fun Indians to fight. Soon the Appalachian strategy became the accepted response to Borderer immigration and was taken up from Pennsylvania in the north to the Carolinas in the South (a few New Englanders hit on a similar idea and sent their own Borderers to colonize the mountains of New Hampshire).

So the Borderers all went to Appalachia and established their own little rural clans there and nothing at all went wrong except for the entire rest of American history.

INTERESTING BORDERER FACTS:
1. Colonial opinion on the Borderers differed within a very narrow range: one Pennsylvanian writer called them "the scum of two nations", another Anglican clergyman called them "the scum of the universe".
2. Some Borderers tried to come to America as indentured servants, but after Virginian planters got some experience with Borderers they refused to accept any more.
3. The Borderers were mostly Presbyterians, and their arrival *en masse* started a race among the established American denominations to convert them. This was mostly unsuccessful; Anglican preacher Charles Woodmason, an important source for information about the early Borderers, said that during his missionary activity the Borderers "disrupted his service, rioted while he preached, started a pack of dogs fighting outside the church, loosed his horse, stole his church key, refused him food and shelter, and gave two barrels of whiskey to his congregation before a service of communion".
4. Borderer town-naming policy was very different from the

Biblical names of the Puritans or the Ye Olde English names of the Virginians. Early Borderer settlements include – just to stick to the creek-related ones – Lousy Creek, Naked Creek, Shitbritches Creek, Cuckold's Creek, Bloodrun Creek, Pinchgut Creek, Whipping Creek, and Hangover Creek. There were also Whiskey Springs, Hell's Half Acre, Scream Ridge, Scuffletown, and Grabtown. The overall aesthetic honestly sounds a bit Orcish.

5. One of the first Borderer leaders was John Houston. On the ship over to America, the crew tried to steal some of his possessions; Houston retaliated by leading a mutiny of the passengers, stealing the ship, and sailing it to America himself. He settled in West Virginia; one of his descendants was famous Texan Sam Houston.

6. Traditional Borderer prayer: "Lord, grant that I may always be right, for thou knowest I am hard to turn."

7. "The backcountry folk bragged that one interior county of North Carolina had so little 'larnin' that the only literate inhabitant was elected 'county reader'"

8. The Borderer accent contained English, Scottish, and Irish elements, and is (uncoincidentally) very similar to the typical "country western singer" accent of today.

9. The Borderers were famous for family feuds in England, including the Johnson clan's habit of "adorning their houses with the flayed skins of their enemies the Maxwells in a blood feud that continued for many generations". The great family feuds of the United States, like the Hatfield-McCoy feud, are a direct descendent of this tradition.

10. Within-clan marriage was a popular Borderer tradition both in England and Appalachia; "in the Cumbrian parish of Hawkshead, for example, both the bride and the groom bore the same last names in 25 percent of all marriages from 1568 to 1704". This led to the modern stereotype of Appalachians as inbred and incestuous.

11. The Borderers were extremely patriarchal and anti-women's-rights to a degree that appalled even the people of the 1700s.

12. “In the year 1767, [Anglican priest] Charles Woodmason calculated that 94 percent of backcountry brides whom he had married in the past year were pregnant on their wedding day”

13. Although the Borderers started off Presbyterian, they were in constant religious churn and their territories were full of revivals, camp meetings, born-again evangelicalism, and itinerant preachers. Eventually most of them ended up as what we now call Southern Baptist.

14. Borderer folk beliefs: “If an old woman has only one tooth, she is a witch”, “If you are awake at eleven, you will see witches”, “The howling of dogs shows the presence of witches”, “If your shoestring comes untied, witches are after you”, “If a warm current of air is felt, witches are passing”. Also, “wet a rag in your enemy’s blood, put it behind a rock in the chimney, and when it rots your enemy will die”; apparently it was not a coincidence they were thinking about witches so much.

15. Borderer medical beliefs: “A cure for homesickness is to sew a good charge of gunpowder on the inside of ths shirt near the neck”. That’ll cure homesickness, all right.

16. More Borderer medical beliefs: “For fever, cut a black chicken open while alive and bind it to the bottom of your foot”, “Eating the brain of a screech owl is the only dependable remedy for headache”, “For rheumatism, apply split frogs to the feet”, “To reduce a swollen leg, split a live cat and apply while still warm”, “Bite the head off the first butterfly you see and you will get a new dress”, “Open the cow’s mouth and throw a live toad-frog down her throat. This will cure her of hollow-horn”. Also, blacksmiths protected themselves from witches by occasionally throwing live puppies into their furnaces.

17. Rates of public schooling in the backcountry settled by the Borderers were “the lowest in British North America” and sometimes involved rituals like “barring out”, where the children would physically keep the teacher out of the school until he gave in and granted the students the day off.

18. “Appalachia’s idea of a moderate drinker was the mountain man who limited himself to a single quart [of whiskey] at a sitting, explaining that more ‘might fly to my head’. Other beverages were regarded with contempt.”

19. A traditional backcountry sport was “rough and tumble”, a no-holds-barred form of wrestling where gouging out your opponent’s eyes was considered perfectly acceptable and in fact sound strategy. In 1772 Virginia had to pass a law against “gouging, plucking, or putting out an eye”, but this was the Cavalier-dominated legislature all the way on the east coast and nobody in the backcountry paid them any attention. Other traditional backcountry sports were sharpshooting and hunting.

20. The American custom of shooting guns into the air to celebrate holidays is 100% Borderer in origin.

21. The justice system of the backcountry was heavy on lynching, originally a race-neutral practice and named after western Virginian settler William Lynch.

22. Scottish Presbyterians used to wear red cloth around their neck to symbolize their religion; other Englishmen nicknamed them “rednecks”. This *may* be the origin of the popular slur against Americans of Borderer descent, although many other etiologies have been proposed. “Cracker” as a slur is attested as early as 1766 by a colonist who says the term describes backcountry men who are great boasters; other proposed etymologies like slaves talking about “whip-crackers” seem to be spurious.

This is not to paint the Borderers as universally poor and dumb – like every group, they had an elite, and some of their elite went on to become some of America’s most important historical figures. Andrew Jackson became the first Borderer president, behaving *exactly* as you would expect the first Borderer president to behave, and he was followed by almost a dozen others. Borderers have also been overrepresented in America’s great military leaders, from Ulysses Grant through Teddy Roosevelt (3/4 Borderer despite his Dutch surname) to George Patton to John McCain.

The Borderers *really* liked America – unsurprising given where they came from – and started identifying as American earlier and more fiercely than any of the other settlers who had come before. Unsurprisingly, they strongly supported the Revolution – Patrick Henry (“Give me liberty or give me death!”) was a Borderer. They also also played a disproportionate role in westward expansion. After the Revolution, America made an almost literal 180 degree turn and the “backcountry” became the “frontier”. It was the Borderers who were happiest going off into the wilderness and fighting Indians, and most of the famous frontiersmen like Davy Crockett were of their number. This was a big part of the reason the Wild West was so wild compared to, say, Minnesota (also a frontier inhabited by lots of Indians, but settled by Northerners and Germans) and why it inherited seemingly Gaelic traditions like cattle rustling.

Their conception of liberty has also survived and shaped modern American politics: it seems essentially to be the modern libertarian/Republican version of freedom from government interference, especially if phrased as “get the hell off my land”, and *especially* especially if phrased that way through clenched teeth while pointing a shotgun at the offending party.

III.

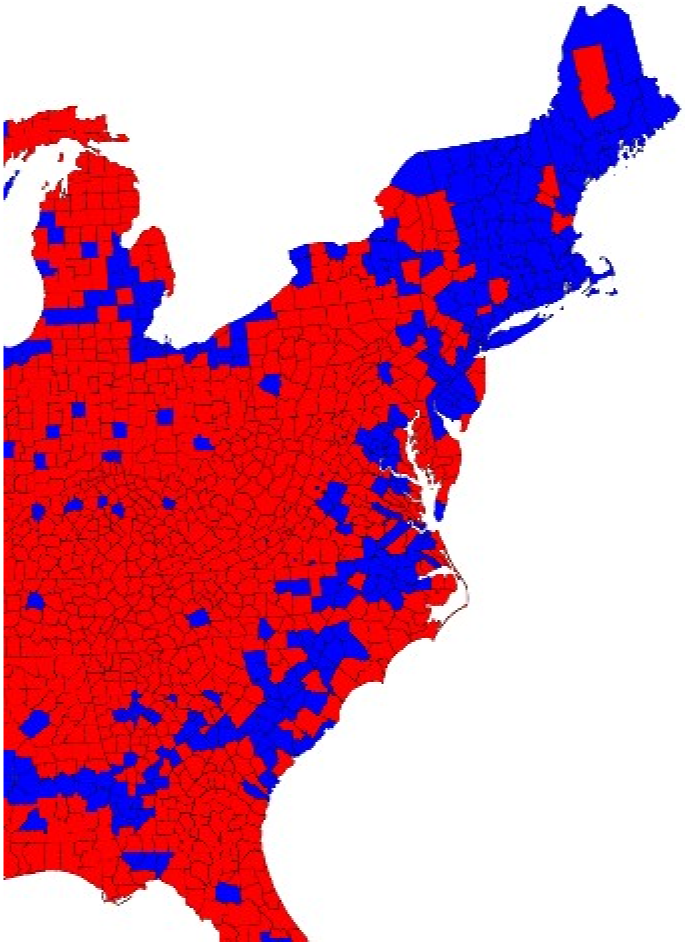
This is all interesting as history and doubly interesting as anthropology, but what relevance does it have for later American history and the present day?

One of my reasons reading this book was to see whether the link between Americans' political opinions and a bunch of their other cultural/religious/social traits (a "Blue Tribe" and "Red Tribe") was related to the immigration patterns it describes. I'm leaning towards "probably", but there's a lot of work to be done in explaining how the split among these four cultures led to a split among two cultures in the modern day, and with little help from the book itself I am going to have to resort to total unfounded speculation. But the simplest explanation – that the Puritans and Quakers merged into one group ("progressives", "Blue Tribe", "educated coastal elites") and the Virginians and Borderers into another ("conservatives", "Red Tribe", "rednecks") – has a lot going for it.

Many conservatives I read like to push the theory that modern progressivism is descended from the utopian Protestant experiments of early America – Puritanism and Quakerism – and that the civil war represents "Massachusetts' conquest of America". I always found this lacking in rigor: Puritanism and Quakerism are sufficiently different that positing a combination of them probably needs more intellectual work than just gesturing at "you know, that Puritan/Quaker thing". But the idea of a Puritan New England and a Quaker-(ish) Pennsylvania gradually blending together into a generic "North" seems plausible, especially given the high levels of interbreeding between the two (some of our more progressive Presidents, including Abraham Lincoln, were literally half-Puritan and half-Quaker). Such a merge would combine the Puritan emphasis on moral reform, education, and a well-ordered society with the Quaker doctrine of niceness, tolerance, religious pluralism, individual conscience, and the Inner Light. It seems kind of unfair to just mix-and-match the most modern elements of each and declare that this proves they caused modernity, but there’s no reason that *couldn’t* have happened.

The idea of Cavaliers and Borderers combining to form modern conservativism is buoyed by modern conservativism’s obvious Border influences, but complicated by its lack of much that is recognizably Cavalier – the Republican Party is hardly marked by its support for a hereditary aristocracy of gentlemen. Here I have to admit that I don’t know as much about Southern history as I’d like. In particular, how were places like Alabama, Mississippi, et cetera settled? Most sources I can find suggest they were set up along the Virginia model of plantation-owning aristocrats, but if that’s true how did the modern populations come to so embody Fischer’s description of Borderers? In particular, why are they so Southern Baptist and not very Anglican? And what happened to all of those indentured servants the Cavaliers brought over after slavery put them out of business? What happened to that whole culture after the Civil War destroyed the plantation system? My *guess* is going to be that the indentured servants and the Borderer population mixed pretty thoroughly, and that this stratum was hanging around providing a majority of the white bodies in the South while the plantation owners were hogging the limelight – but I just don’t know.

A quick argument that I’m not totally making all of this up:



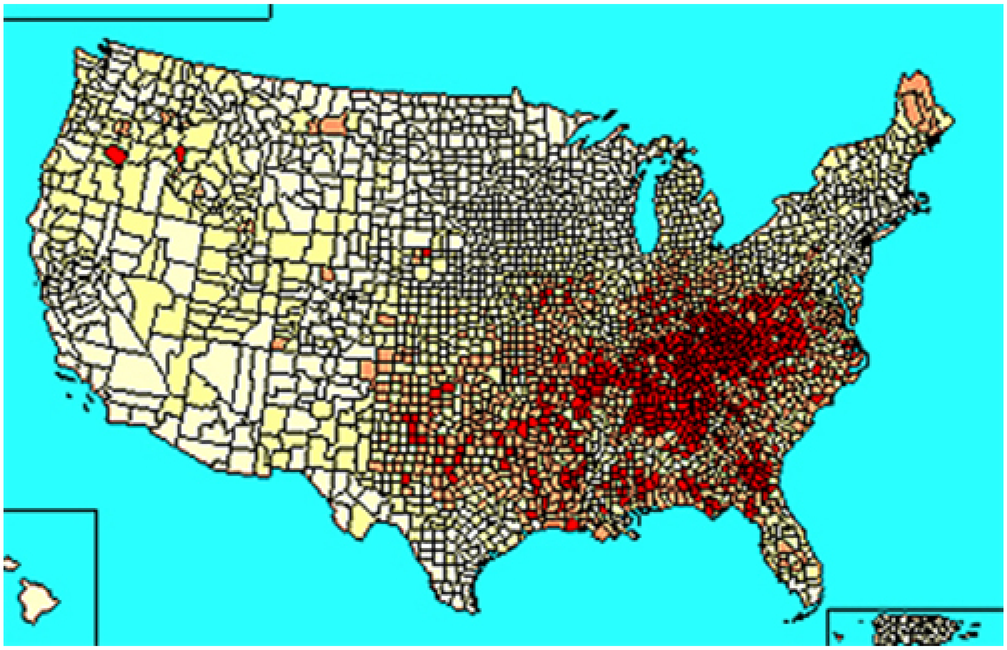
This is a map of voting patterns by county in the 2012 Presidential election. The blue areas in the South carefully track the so-called “black belt” of majority African-American areas. The ones in the Midwest are mostly big cities. Aside from those, the only people who vote Democrat are New England (very solidly!) and the Delaware Valley region of Pennsylvania. In fact, you can easily see the distinction between the Delaware Valley settled by Quakers in the east, and the backcountry area settled by Borderers in the west. Even the book’s footnote about how a few Borderers settled in the mountains of New Hampshire is associated with a few spots of red in the mountains of New Hampshire ruining an otherwise near-perfect Democratic sweep of the north.

One anomaly in this story is a kind of linear distribution of blue across southern Michigan, too big to be explained solely by the blacks of Detroit. But a quick look at Wikipedia's History of Michigan finds:

> In the 1820s and 1830s migrants from New England began moving to what is now Michigan in large numbers (though there was a trickle of New England settlers who arrived before this date). These were "Yankee" settlers, that is to say they were descended from the English Puritans who settled New England during the colonial era....Due to the prevalence of New Englanders and New England transplants from upstate New York, Michigan was very culturally contiguous with early New England culture for much of its early history...The amount with which the New England Yankee population predominated made Michigan unique among frontier states in the antebellum period. Due to this heritage Michigan was on the forefront of the antislavery crusade and reforms during the 1840s and 1850s.

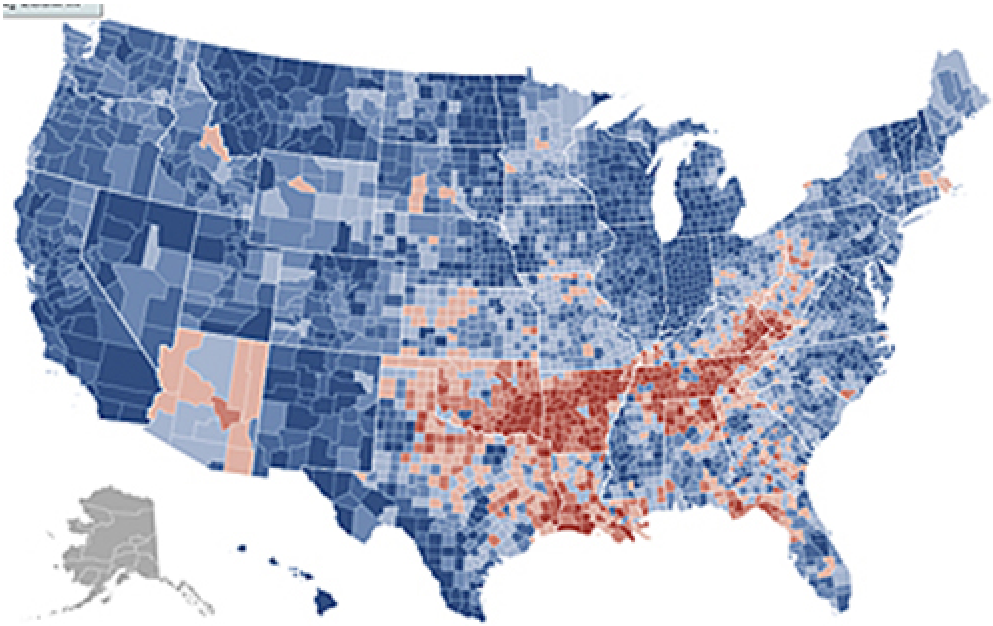
Alhough I can't find proof of this specifically, I know that Michigan was settled from the south up, and I suspect that these New England settlers concentrated in the southern regions and that the north was settled by a more diverse group of whites who lacked the New England connection.

Here's something else cool. We can't track Borderers directly because there's no "Borderer" or "Scots-Irish" option on the US census. But *Albion's Seed* points out that the Borderers were uniquely likely to identify as just "American" and deliberately forgot their past ancestry as fast as they could. Meanwhile, when the census asks an ethnicity question about where your ancestors came from, every year some people will stubbornly ignore the point of the question and put down "America" (no, this does not track the distribution of Native American population). Here’s a map of so-called “unhyphenated Americans”, taken from this site:



We see a strong focus on the Appalachian Mountains, especially West Virginia, Tennesee, and Kentucky, bleeding into the rest of the South. Aside from west Pennsylvania, this is very close to where we would expect to find the Borderers. Could these be the same groups?

Meanwhile, here is a map of where Obama underperformed the usual Democratic vote *worst* in 2008:



These maps are small and lossy, and surely unhyphenatedness is not an *exact* proxy for Border ancestry – but they are nevertheless intriguing. You may also be interested in the Washington Post's correlation between distribution of unhyphenated Americans and Trump voters, or the Atlantic's article on Trump and Borderers.

If I'm going to map these cultural affiliations to ancestry, do I have to walk back on my previous theory that they are related to class? Maybe I should. But I also think we can posit complicated interactions between these ideas. Consider for example the interaction between race and class; a black person with a white-sounding name, who speaks with a white-sounding accent, and who adopts white culture (eg listens to classical music, wears business suits) is far more likely to seem upper-class than a black person with a black-sounding name, a black accent, and black cultural preferences; a white person who seems black in some way (listens to hip-hop, wears baggy clothes) is more likely to seem lower-class. This doesn't mean race and class are exactly the same thing, but it does mean that some races get stereotyped as upper-class and others as lower-class, and that people’s racial identifiers may change based on where they are in the class structure.

I think something similar is probably going on with these forms of ancestry. The education system is probably dominated by descendents of New Englanders and Pennsylvanians; they had an opportunity to influence the culture of academia and the educated classes more generally, they took it, and now anybody of any background who makes it into that world is going to be socialized according to their rules. Likewise, people in poorer and more rural environments will be surrounded by people of Borderer ancestry and acculturated by Borderer cultural products and end up a little more like that group. As a result, ethnic markers have turned into and merged with class markers in complicated ways.

Indeed, some kind of acculturation process has to have been going on, since most of the people in these areas today are not the descendents of the original settlers. But such a process seems very likely. Just to take an example, most of the Jews I know (including my own family) came into the country via New York, live somewhere on the coast, and have very Blue Tribe values. But Southern Jews believed in the Confederacy as strongly as any Virginian – see for example Judah Benjamin. And Barry Goldwater, a half-Jew raised in Arizona, invented the modern version of conservativism that seems closest to some Borderer beliefs.

All of this is very speculative, with some obvious flaws. What do we make of other countries like Britain or Germany with superficially similar splits but very different histories? Why should Puritans lose their religion and sexual prudery, but keep their interest in moralistic reform? There are whole heaps of questions like these. But look. Before I had any idea about any of this, I wrote that American society seems divided into two strata, one of which is marked by emphasis on education, interest in moral reforms, racial tolerance, low teenage pregnancy, academic/financial jobs, and Democratic party affiliation, and furthermore that this group was centered in the North. Meanwhile, now I learn that the North was settled by two groups that when combined have emphasis on education, interest in moral reforms, racial tolerance, low teenage pregnancy, an academic and mercantile history, and were the heartland of the historical Whigs and Republicans who preceded the modern Democratic Party.

And I wrote about another stratum centered in the South marked by poor education, gun culture, culture of violence, xenophobia, high teenage pregnancy, militarism, patriotism, country western music, and support for the Republican Party. And now I learn that the South was settled by a group noted even in the 1700s for its poor education, gun culture, culture of violence, xenophobia, high premarital pregnancy, militarism, patriotism, accent exactly like the modern country western accent, and support for the Democratic-Republicans who preceded the modern Republican Party.

If this is true, I think it paints a very pessimistic world-view. The “iceberg model” of culture argues that apart from the surface cultural features we all recognize like language, clothing, and food, there are deeper levels of culture that determine the features and institutions of a people: whether they are progressive or traditional, peaceful or warlike, mercantile or self-contained. We grudgingly acknowledge these features when we admit that maybe making the Middle East exactly like America in every way is more of a long-term project than something that will happen as soon as we kick out the latest dictator and get treated as liberators. Part of us may still want to believe that pure reason is the universal solvent, that those Afghans will come around once they realize that being a secular liberal democracy is obviously great. But we keep having deep culture shoved in our face again and again, and we don’t know how to get rid of it. This has led to reasonable speculation that some aspects of it might even be genetic – something which would explain a lot, though not its ability to acculturate recent arrivals.

This is a hard pill to swallow even when we’re talking about Afghanistan. But it becomes doubly unpleasant when we think about it in the sense of our neighbors and fellow citizens in a modern democracy. What, after all, is the point? A democracy made up of 49% extremely liberal Americans and 51% fundamentalist Taliban Afghans would be something very different from the democratic ideal; even if occasionally a super-charismatic American candidate could win over enough marginal Afghans to take power, there’s none of the give-and-take, none of the competition within the marketplace of ideas, that makes democracy so attractive. Just two groups competing to dominate one another, with the fact that the competition is peaceful being at best a consolation prize.

If America is best explained as a Puritan-Quaker culture locked in a death-match with a Cavalier-Borderer culture, with all of the appeals to freedom and equality and order and justice being just so many epiphenomena – well, I’m not sure what to do with that information. Push it under the rug? Say “Well, my culture is better, so I intend to do as good a job dominating yours as possible?” Agree that We Are Very Different Yet In The End All The Same And So Must Seek Common Ground? Start researching genetic engineering? Maybe secede? I’m not a Trump fan much more than I’m an Osama bin Laden fan; if somehow Osama ended up being elected President, should I start thinking “Maybe that time we made a country that was 49% people like me and 51% members of the Taliban – *maybe that was a bad idea*“.

I don’t know. But I highly recommend Albion’s Seed as an entertaining and enlightening work of historical scholarship which will be absolutely delightful if you don’t fret too much over all of the existential questions it raises.

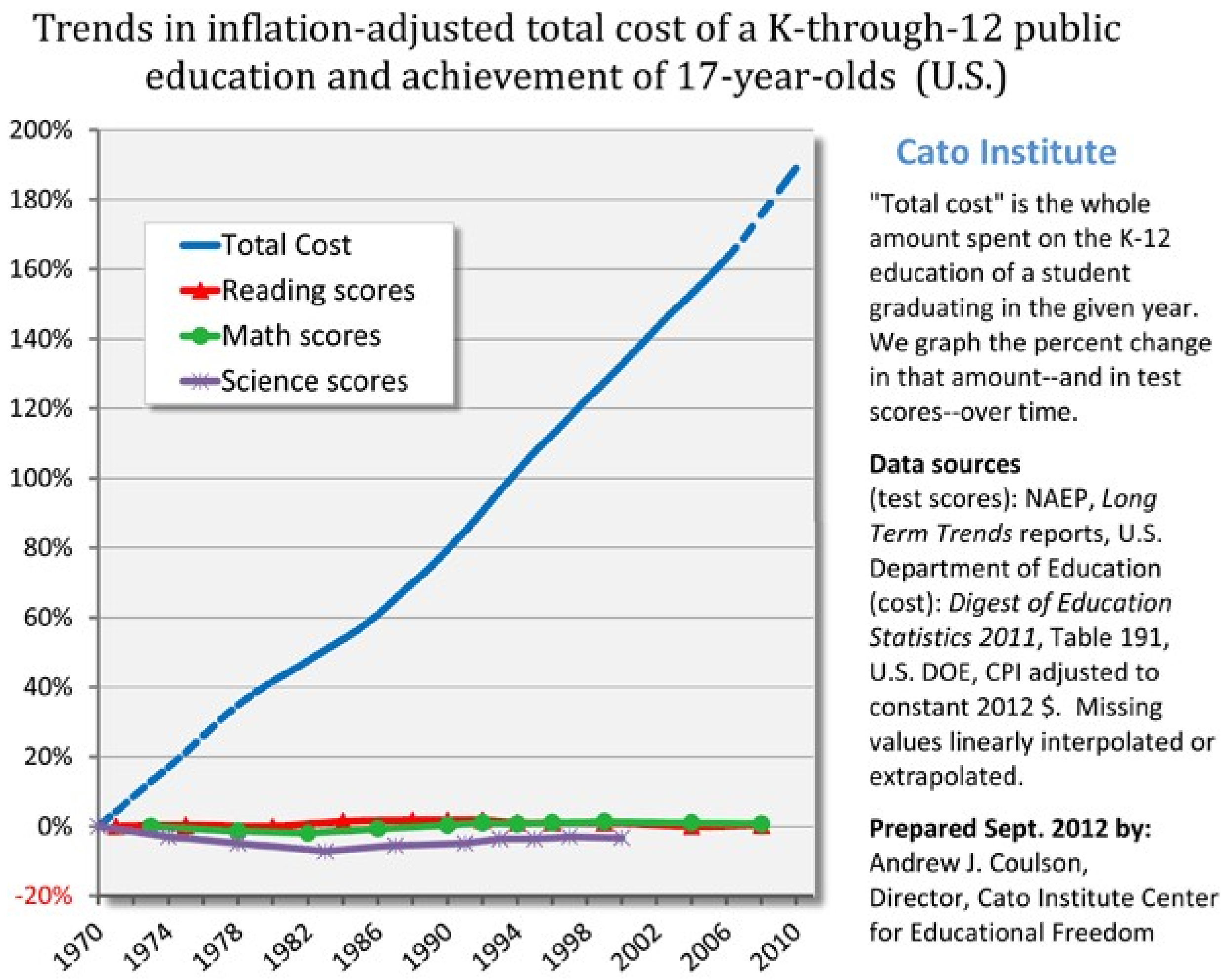
Considerations On Cost Disease

I.

Tyler Cowen writes about cost disease. I’d previously heard the term used to refer only to a specific theory of why costs are increasing, involving labor becoming more efficient in some areas than others. Cowen seems to use it indiscriminately to refer to increasing costs in general – which I guess is fine, goodness knows we need a word for that.

Cowen assumes his readers already understand that cost disease exists. I don’t know if this is true. My impression is that most people still don’t know about cost disease, or don’t realize the extent of it. So I thought I would make the case for the cost disease in the sectors Tyler mentions – health care and education – plus a couple more.

First let’s look at primary education:



Trends in inflation-adjusted total cost of a K-through-12 public education and achievement of 17-year-olds (U.S.)

"Total cost" is the whole amount spent on the K-12 education of a student graduating in the given year. We graph the percent change in that amount--and in test scores--over time.

Data sources (test scores): NAEP, *Long Term Trends* reports, U.S. Department of Education (cost): *Digest of Education Statistics 2011*, Table 191, U.S. DOE, CPI adjusted to constant 2012 $. Missing values linearly interpolated or extrapolated.

Prepared Sept. 2012 by: Andrew J. Coulson, Director, Cato Institute Center for Educational Freedom

There was some argument about the style of this graph, but as per Politifact the basic claim is true. Per student spending has increased about 2.5x in the past forty years even after adjusting for inflation.

At the same time, test scores have stayed relatively stagnant. You can see the full numbers here, but in short, high school students' reading scores went from 285 in 1971 to 287 today – a difference of 0.7%.

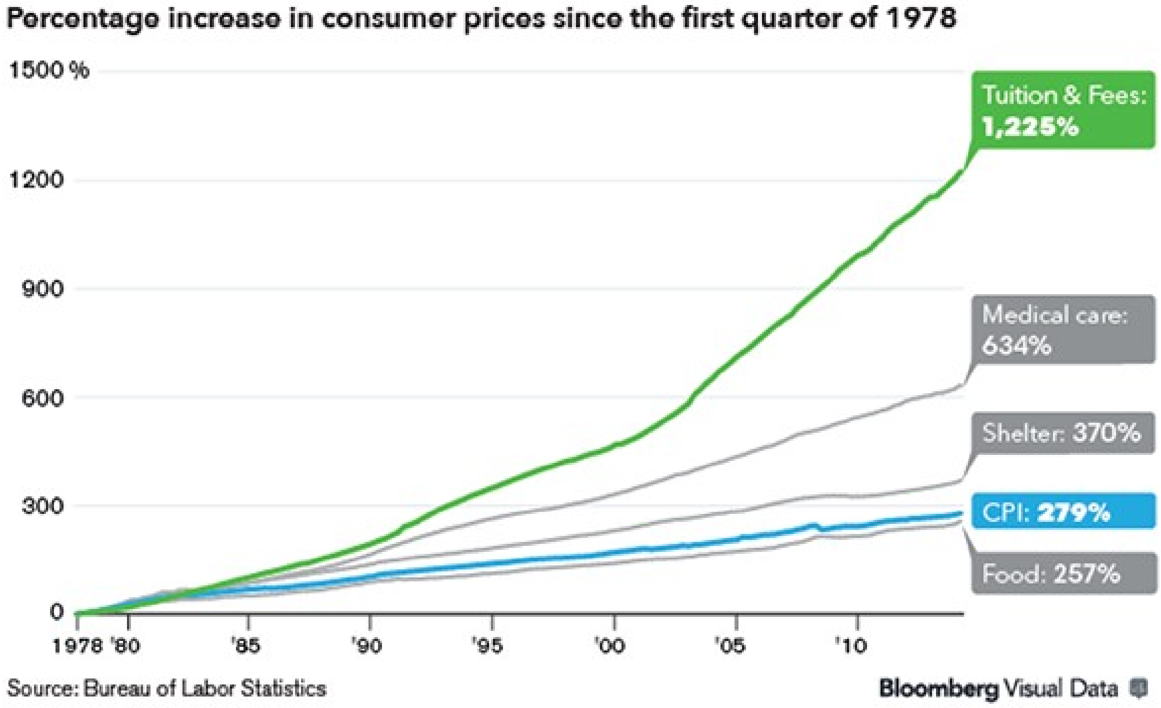
There is some heterogenity across races – white students' test scores increased 1.4% and minority students' scores by about 20%. But it is hard to credit school spending for the minority students' improvement, which occurred almost entirely during the period from 1975-1985. School spending has been on exactly the same trajectory before and after that time, and in white and minority areas, suggesting that there was something specific about that decade which improved minority (but not white) scores. Most likely this was the general improvement in minorities' conditions around that time, giving them better nutrition and a more stable family life. It's hard to construct a narrative where it was school spending that did it – and even if it did, note that the majority of the increase in school spending happened from 1985 on, and demonstrably helped neither whites *nor* minorities.

I discuss this phenomenon more here and here, but the summary is: no, it's not just because of special ed; no, it's not just a factor of how you measure test scores; no, there's not a "ceiling effect". Costs really did more-or-less double without any concomitant increase in measurable quality.

So, imagine you're a poor person. White, minority, whatever. Which would you prefer? Sending your child to a 2016 school? Or sending your child to a 1975 school, and getting a check for $5,000 every year?

I'm proposing that choice because as far as I can tell that *is* the stakes here. 2016 schools have whatever tiny test score advantage they have over 1975 schools, and cost $5000/year more, inflation adjusted. That $5000 comes out of the pocket of somebody – either taxpayers, or other people who could be helped by government programs.

Second, college is even worse:



Percentage increase in consumer prices since the first quarter of 1978

Source: Bureau of Labor Statistics

Bloomberg Visual Data

Note this is not adjusted for inflation; see link below for adjusted figures

Inflation-adjusted cost of a university education was [something like $2000/year in 1980](). Now it's closer to $20,000/year. No, it's not because of [decreased government funding](), and there are similar trajectories for public and private schools.

I don't know if there's an equivalent of "test scores" measuring how well colleges perform, so just use your best judgment. Do you think that modern colleges provide $18,000/year greater value than colleges did in your parents' day? Would you rather graduate from a modern college, or graduate from a college more like the one your parents went to, plus get a check for $72,000?

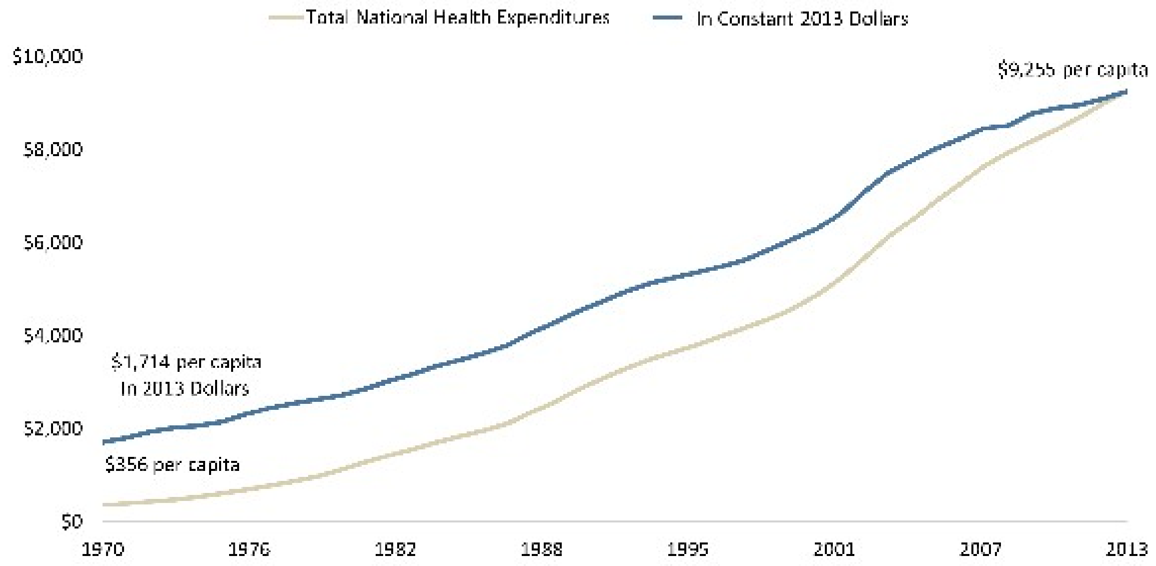
(or, more realistically, have $72,000 less in student loans to pay off)

Was your parents' college even noticeably worse than yours? My parents sometimes talk about their college experience, and it seems to have had all the relevant features of a college experience. Clubs. Classes. Professors. Roommates. I might have gotten something extra for my $72,000, but it's hard to see what it was.

Third, health care. The graph is starting to look disappointingly familiar:

On a per capita basis, health spending has grown substantially

Total national health expenditures, US $ per capita, 1970-2013

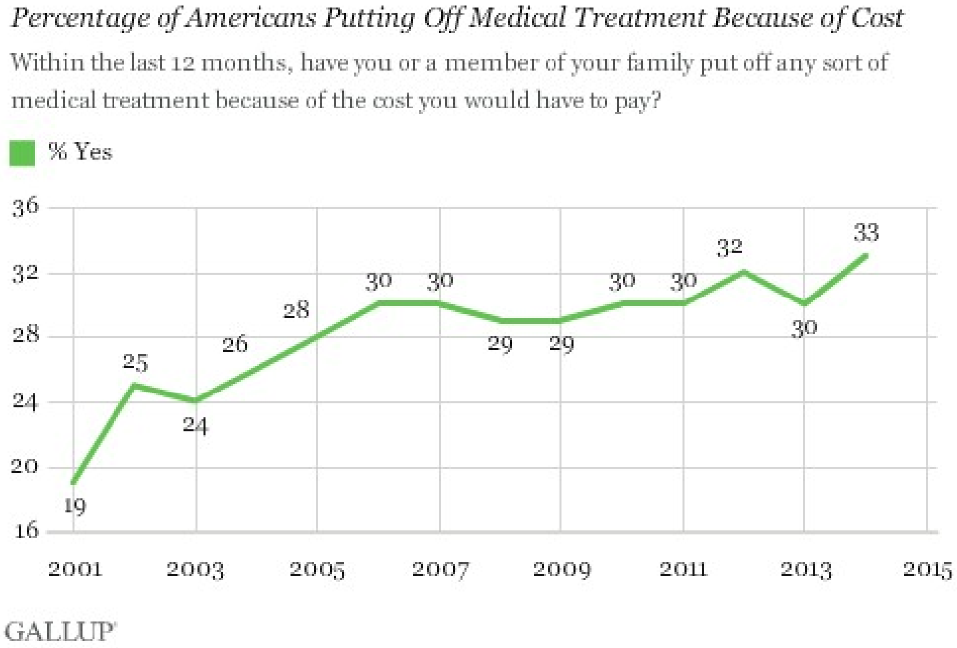


Source: National Health Expenditure (NHE) data from Centers for Medicare and Medicaid Services, Office of the Actuary, National Health Statistics Group

Peterson-Kaiser Health System Tracker

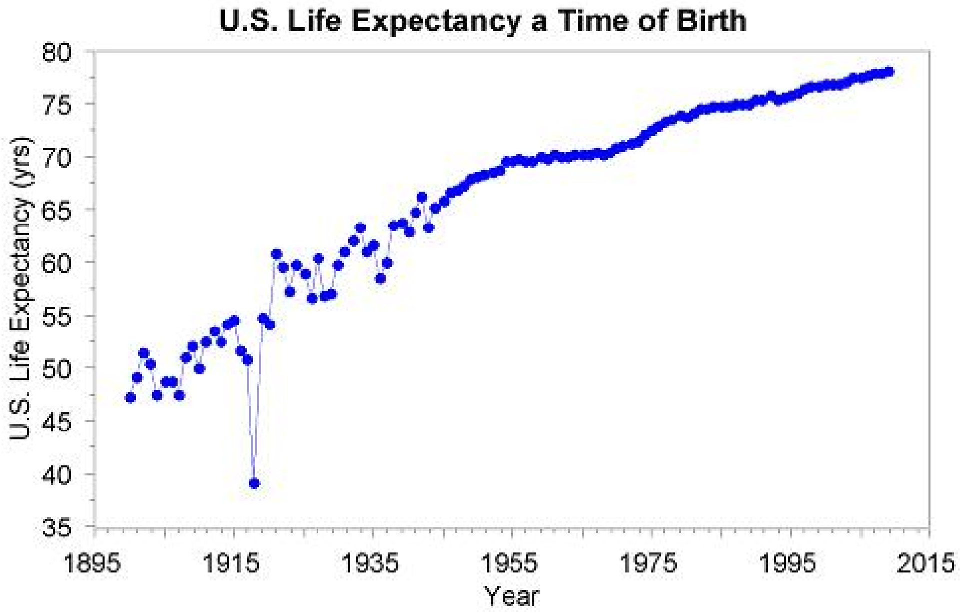
The cost of health care has about quintupled since 1970. It's actually been rising since earlier than that, but I can't find a good graph; it looks like it would have been about $1200 in today's dollars in 1960, for an increase of about 800% in those fifty years.

This has had the expected effects. The average 1960 worker spent ten days’ worth of their yearly paycheck on health insurance; the average modern worker spends sixty days’ worth of it, a sixth of their entire earnings.



Or not.

This time I can’t say with 100% certainty that all this extra spending has been for nothing. Life expectancy has gone way up since 1960:



Extra bonus conclusion: the Spanish flu was really bad

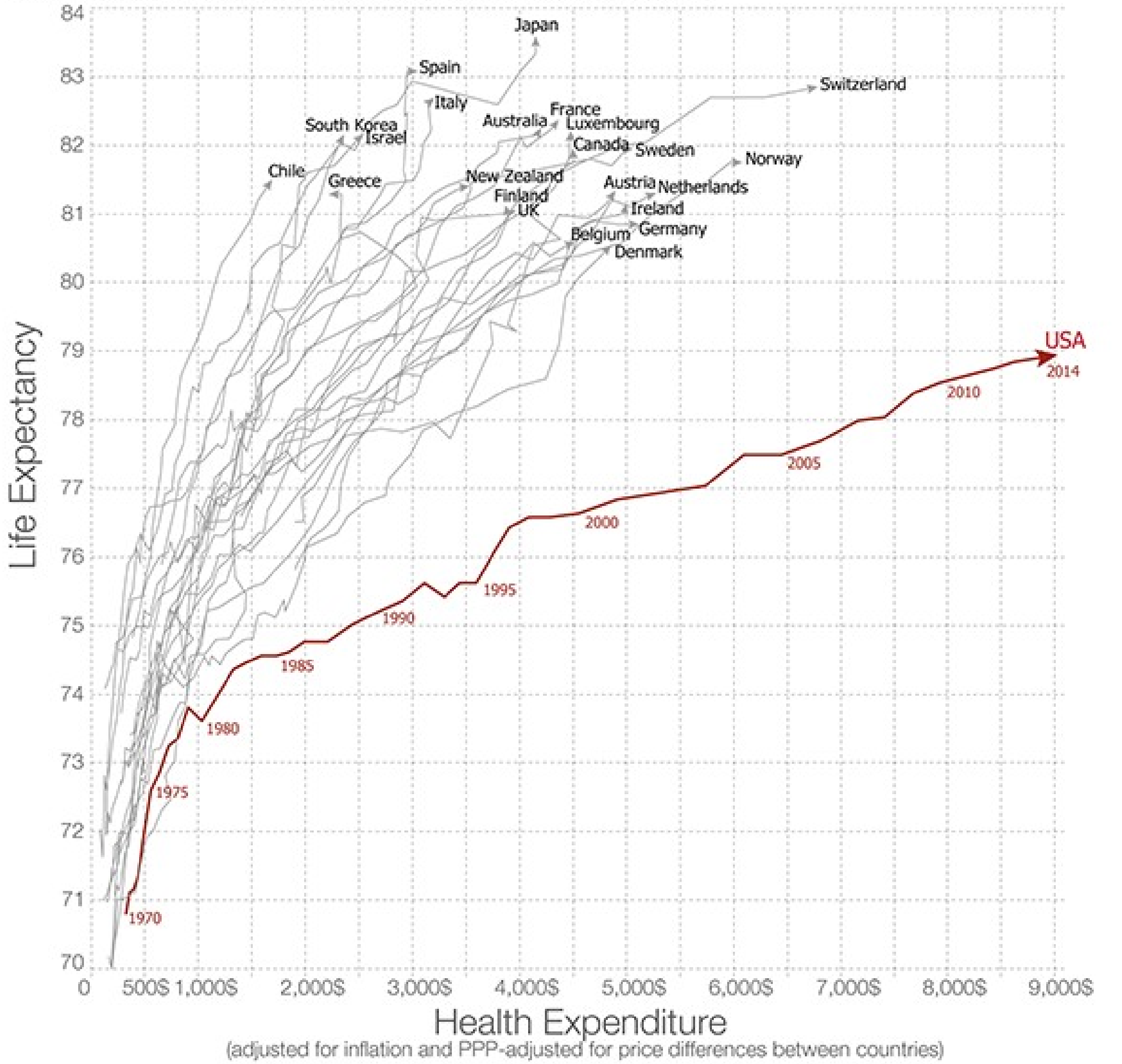
But a lot of people think that life expectancy depends on other things a lot more than healthcare spending. Sanitation, nutrition, quitting smoking, plus advances in health technology that don’t involve spending more money. ACE inhibitors (invented in 1975) are great and probably increased lifespan a lot, but they cost $20 for a year’s supply and replaced older drugs that cost about the same amount.

In terms of calculating how much lifespan gain healthcare spending has produced, we have a couple of options. Start with by country:

Life expectancy vs. health expenditure over time (1970-2014)

Our World in Data

Health spending measures the consumption of health care goods and services, including personal health care (curative care, rehabilitative care, long-term care, ancillary services and medical goods) and collective services (prevention and public health services as well as health administration), but excluding spending on investments. Shown is total health expenditure (financed by public and private sources).



Data source: Health expenditure from the OECD; Life expectancy from the World Bank   Licensed under CC-BY-SA by the author Max Roser.
The data visualization is available at OurWorldinData.org and there you find more research and visualizations on this topic.

Countries like South Korea and Israel have about the same life expectancy as the US but pay about 25% of what we do. Some people use this to prove the superiority of centralized government health systems, although Random Critical Analysis has an alternative perspective. In any case, it seems very possible to get the same improving life expectancies as the US without octupling health care spending.

The Netherlands increased their health budget by a lot around 2000, sparking a bunch of studies on whether that increased life expectancy or not. There’s a good meta-analysis here, which lists six studies trying to calculate how much of the change in life expectancy was due to the large increases in health spending during this period. There’s a broad range of estimates: 0.3%, 1.8%, 8.0%, 17.2%, 22.1%, 27.5% (I’m taking their numbers for men; the numbers for women are pretty similar). They also mention two studies that they did not officially include; one finding 0% effect and one finding 50% effect (I’m not sure why these studies weren’t included). They add:

> In none of these studies is the issue of reverse causality addressed; sometimes it is not even mentioned. This implies that the effect of health care spending on mortality may be overestimated.

They say:

> Based on our review of empirical studies, we conclude that it is likely that increased health care spending has contributed to the recent increase in life expectancy in the Netherlands. Applying the estimates form published studies to the observed increase in health care spending in the Netherlands between 2000 and 2010 [of 40%] would imply that 0.3% to almost 50% of the increase in life expectancy may have been caused by increasing health care spending. An important reason for the wide range in such estimates is that they all include methodological problems highlighted in this paper. However, this wide range inicates that the counterfactual study by Meerding et al, which argued that 50% of the increase in life expectancy in the Netherlands since the

> 1950s can be attributed to medical care, can probably be interpreted as an upper bound.

It’s going to be completely irresponsible to try to apply this to the increase in health spending in the US over the past 50 years, since this is probably different at every margin and the US is not the Netherlands and the 1950s are not the 2010s. But if we irresponsibly take their median estimate and apply it to the current question, we get that increasing health spending in the US has been worth about one extra year of life expectancy.

This study attempts to directly estimate a %GDP health spending to life expectancy conversion, and says that an increase of 1% GDP corresponds to an increase of 0.05 years life expectancy. That would suggest a slightly different number of 0.65 years life expectancy gained by healthcare spending since 1960)

If these numbers seem absurdly low, remember all of those controlled experiments where giving people insurance doesn’t seem to make them much healthier in any meaningful way.

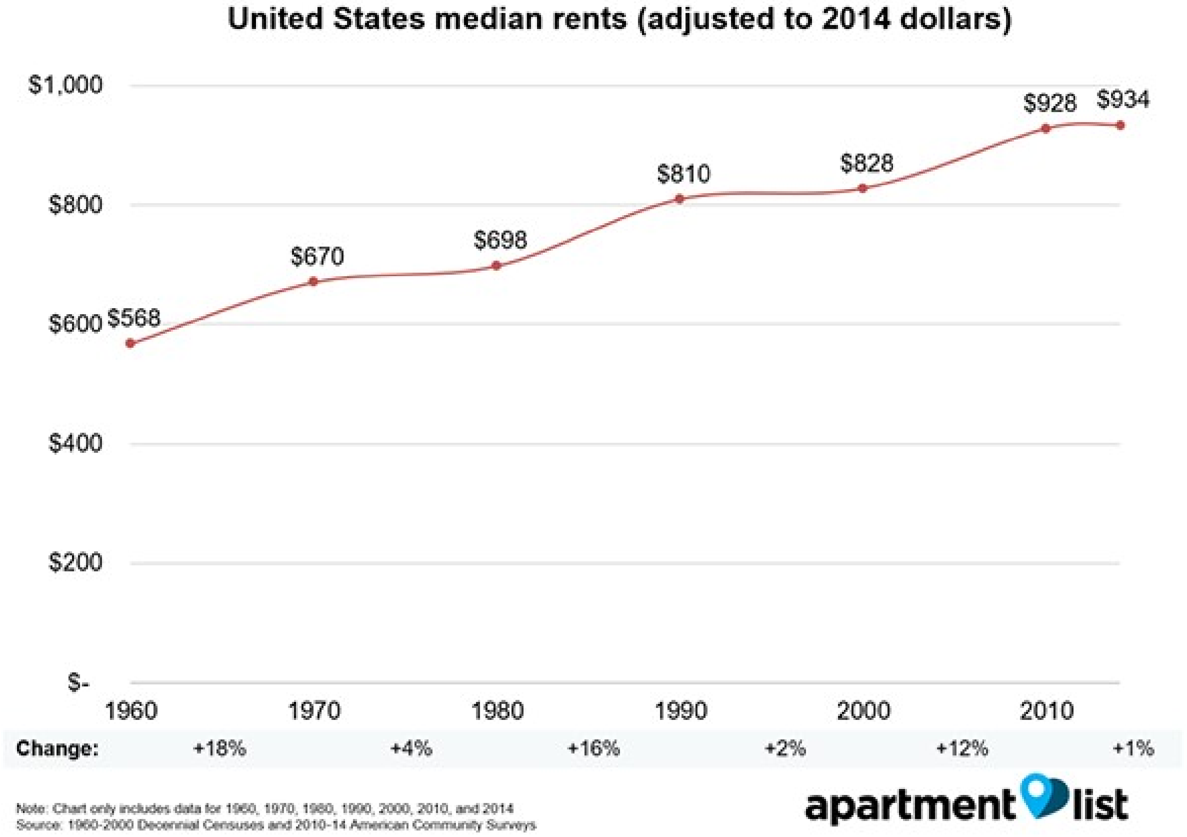
Or instead of slogging through the statistics, we can just ask the same question as before. Do you think the average poor or middle-class person would rather:

a) Get modern health care
b) Get the same amount of health care as their parents’ generation, but with modern technology like ACE inhibitors, and also earn $8000 extra a year

Fourth, we se similar effects in infrastructure. The first New York City subway opened around 1900. Various sources list lengths from 10 to 20 miles and costs from $30 million to $60 million dollars – I think my sources are capturing it at different stages of construction with different numbers of extensions. In any case, it suggests costs of between $1.5 million to $6 million dollars/mile = $1-4 million per kilometer. That looks like it's about the inflation-adjusted equivalent of $100 million/kilometer today, though I'm very uncertain about that estimate. In contrast, Vox notes that a new New York subway line being opened this year costs about $2.2 *billion* per kilometer, suggesting a cost increase of twenty times – although I'm very uncertain about this estimate.

Things become clearer when you compare them country-by-country. The same Vox article notes that Paris, Berlin, and Copenhagen subways cost about $250 million per kilometer, almost 90% less. Yet even those European subways are overpriced compared to Korea, where a kilometer of subway in Seoul costs $40 million/km (another Korean subway project cost $80 million/km). This is a difference of 50x between Seoul and New York for apparently comparable services. It suggests that the 1900s New York estimate above may have been roughly accurate if their efficiency was roughly in line with that of modern Europe and Korea.

Fifth, housing (source:



Most of the important commentary on this graph has already been said, but I would add that optimistic takes like this one by the American Enterprise Institute are missing some of the dynamic. Yes, homes are bigger than they used to be, but part of that is zoning laws which make it easier to get big houses than small houses. There are a lot of people who would prefer to have a smaller house but don't. When I first moved to Michigan, I lived alone in a three bedroom house because there were no good one-bedroom houses available near my workplace and all of the apartments were loud and crime-y.

Or, once again, just ask yourself: do you think most poor and middle class people would rather:

1. Rent a modern house/apartment
2. Rent the sort of house/apartment their parents had, for

half the cost

II.

So, to summarize: in the past fifty years, education costs have doubled, college costs have dectupled, health insurance costs have dectupled, subway costs have at least dectupled, and housing costs have increased by about fifty percent. US health care costs about four times as much as equivalent health care in other First World countries; US subways cost about eight times as much as equivalent subways in other First World countries.

I worry that people don’t appreciate how weird this is. I didn’t appreciate it for a long time. I guess I just figured that Grandpa used to talk about how back in his day movie tickets only cost a nickel; that was just the way of the world. *But all of the numbers above are inflation-adjusted*. These things have dectupled in cost even *after* you adjust for movies costing a nickel in Grandpa’s day. They have really, genuinely dectupled in cost, no economic trickery involved.

And this is especially strange because we expect that improving technology and globalization ought to cut costs. In 1983, the first mobile phone cost $4,000 – about $10,000 in today’s dollars. It was also a gigantic piece of crap. Today you can get a much better phone for $100. This is the right and proper way of the universe. It’s why we fund scientists, and pay businesspeople the big bucks.

But things like college and health care have *still* had their prices dectuple. Patients can now schedule their appointments online; doctors can send prescriptions through the fax, pharmacies can keep track of medication histories on centralized computer systems that interface with the cloud, nurses get automatic reminders when they’re giving two drugs with a potential interaction, insurance companies accept payment through credit cards – and all of this costs ten times as much as it did in the days of punch cards and secretaries who did calculations by hand.

It’s actually even worse than this, because we take so many opportunities to save money that were unavailable in past generations. Underpaid foreign nurses immigrate to America and work for a song. Doctors’ notes are sent to India overnight where they’re transcribed by sweatshop-style labor for pennies an hour. Medical equipment gets manufactured in goodness-only-knows which obscure Third World country. And it *still* costs ten times as much as when this was all made in the USA – and that back when minimum wages were proportionally higher than today.

And it’s actually even worse than *this*. A lot of these services have decreased in quality, presumably as an attempt to cut costs even further. Doctors used to make house calls; even when I was young in the ’80s my father would still go to the houses of difficult patients who were too sick to come to his office. This study notes that for women who give birth in the hospital, “the standard length of stay was 8 to 14 days in the 1950s but declined to less than 2 days in the mid-1990s”. The doctors I talk to say this isn’t because modern women are healthier, it’s because they kick them out as soon as it’s safe to free up beds for the next person. Historic records of hospital care generally describe leisurely convalescence periods and making sure somebody felt absolutely well before letting them go; this seems bizarre to anyone who has participated in a modern hospital, where the mantra is to kick people out as soon as they’re “stable” ie not in acute crisis.

If we had to provide the same quality of service as we did in 1960, and without the gains from modern technology and globalization, who even *knows* how many times more health care would cost? Fifty times more? A hundred times more?

And the same is true for colleges and houses and subways and so on.

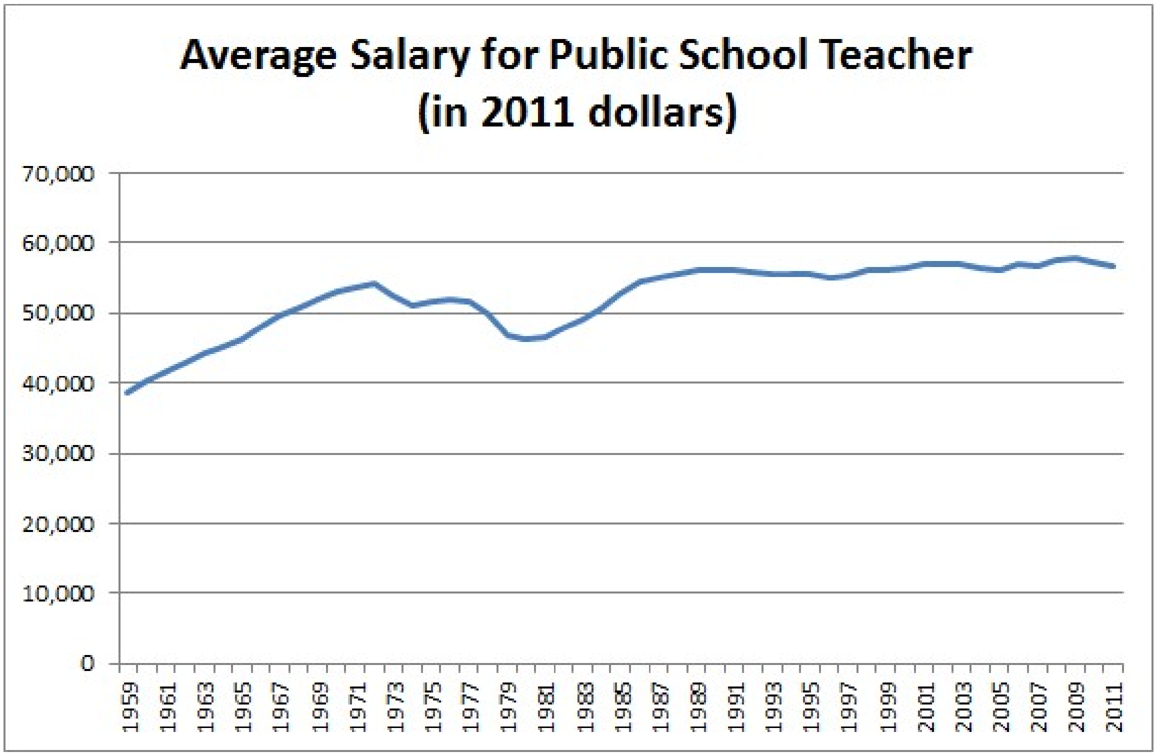
III.

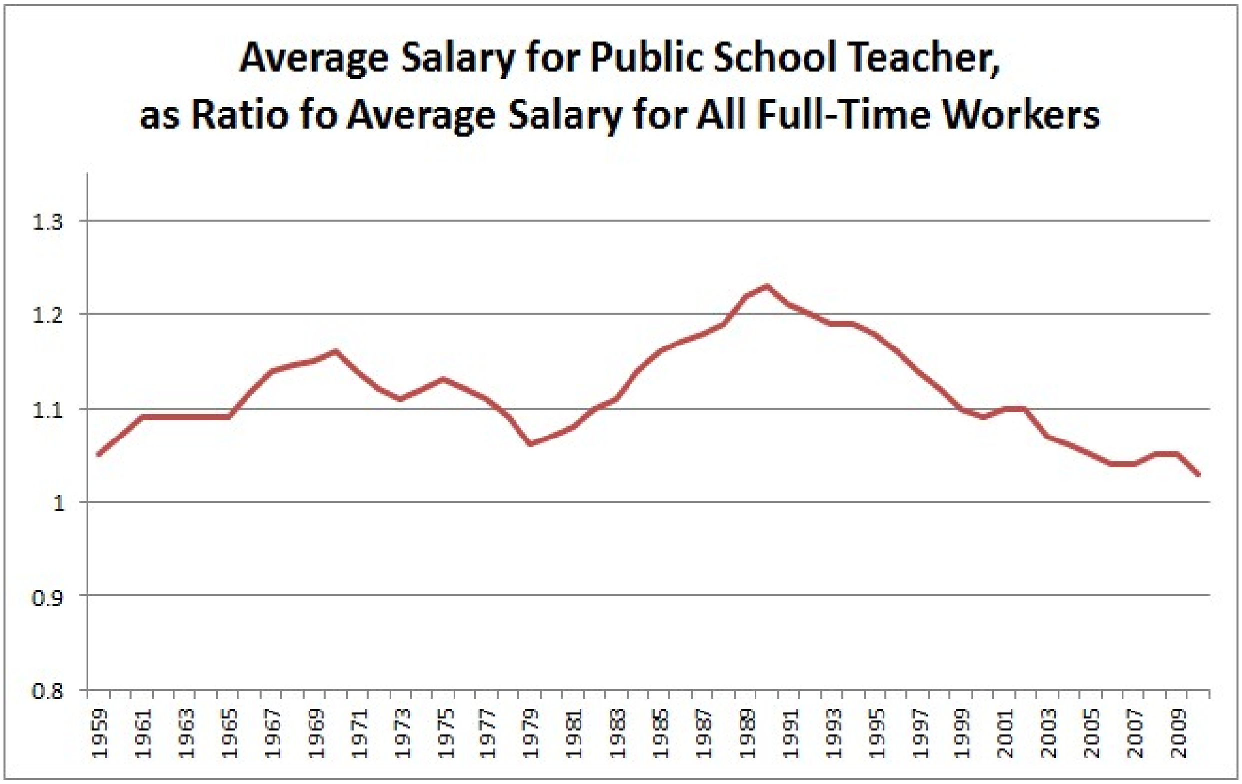
The existing literature on cost disease focuses on the Baumol effect. Suppose in some underdeveloped economy, people can choose either to work in a factory or join an orchestra, and the salaries of factory workers and orchestra musicians reflect relative supply and demand and profit in those industries. Then the economy undergoes a technological revolution, and factories can produce ten times as many goods. Some of the increased productivity trickles down to factory workers, and they earn more money. Would-be musicians leave the orchestras behind to go work in the higher-paying factories, and the orchestras have to raise their prices if they want to be assured enough musicians. So tech improvements in the factory sectory raise prices in the orchestra sector.

We could tell a story like this to explain rising costs in education, health care, etc. If technology increases productivity for skilled laborers in other industries, then less susceptible industries might end up footing the bill since they have to pay their workers more.

There’s only one problem: health care and education aren’t paying their workers more; in fact, quite the opposite.

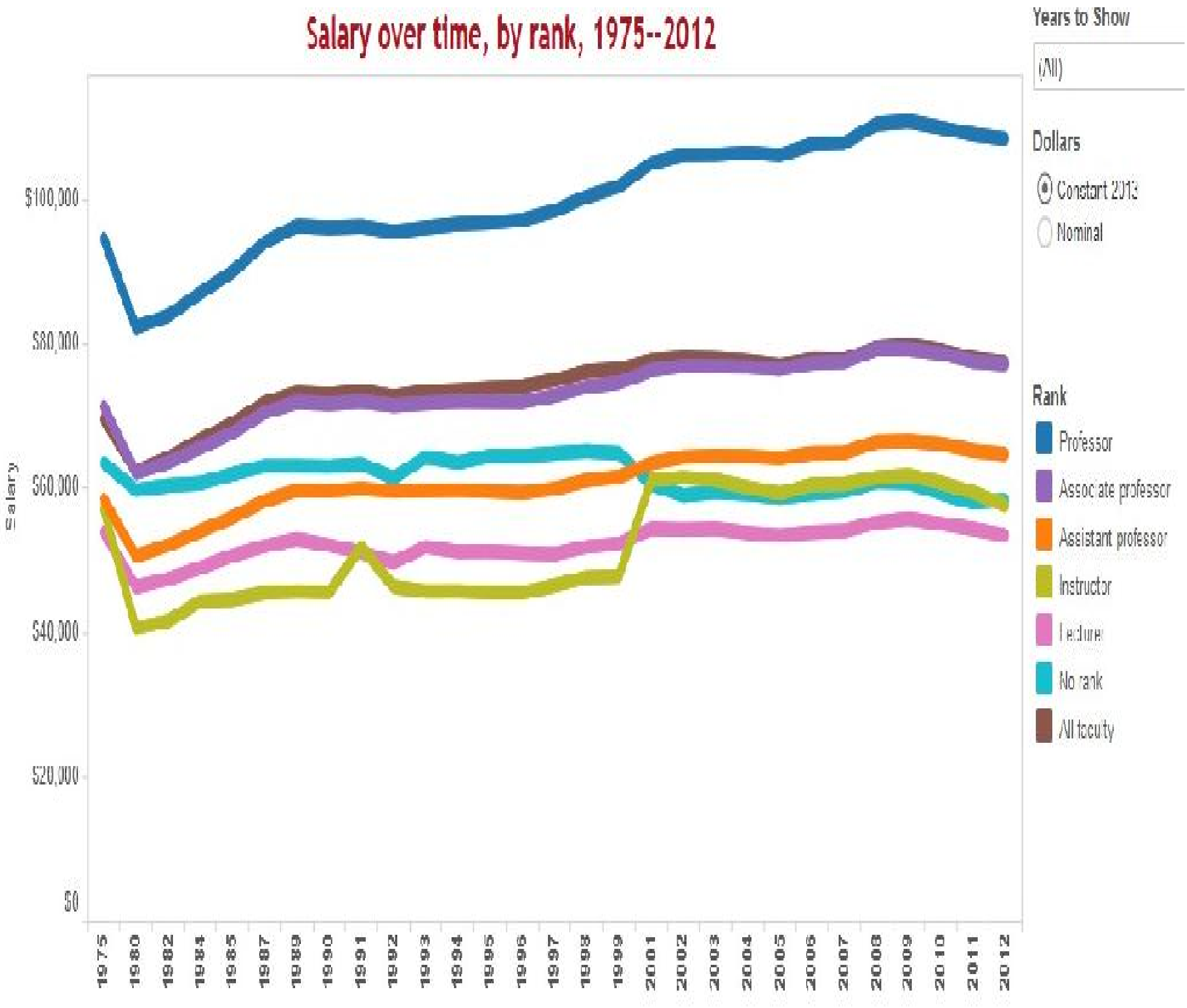
Here are teacher salaries over time (source):





Teacher salaries are relatively flat adjusting for inflation. But salaries for other jobs are increasing modestly relative to inflation. So teacher salaries relative to other occupations' salaries are actually declining.

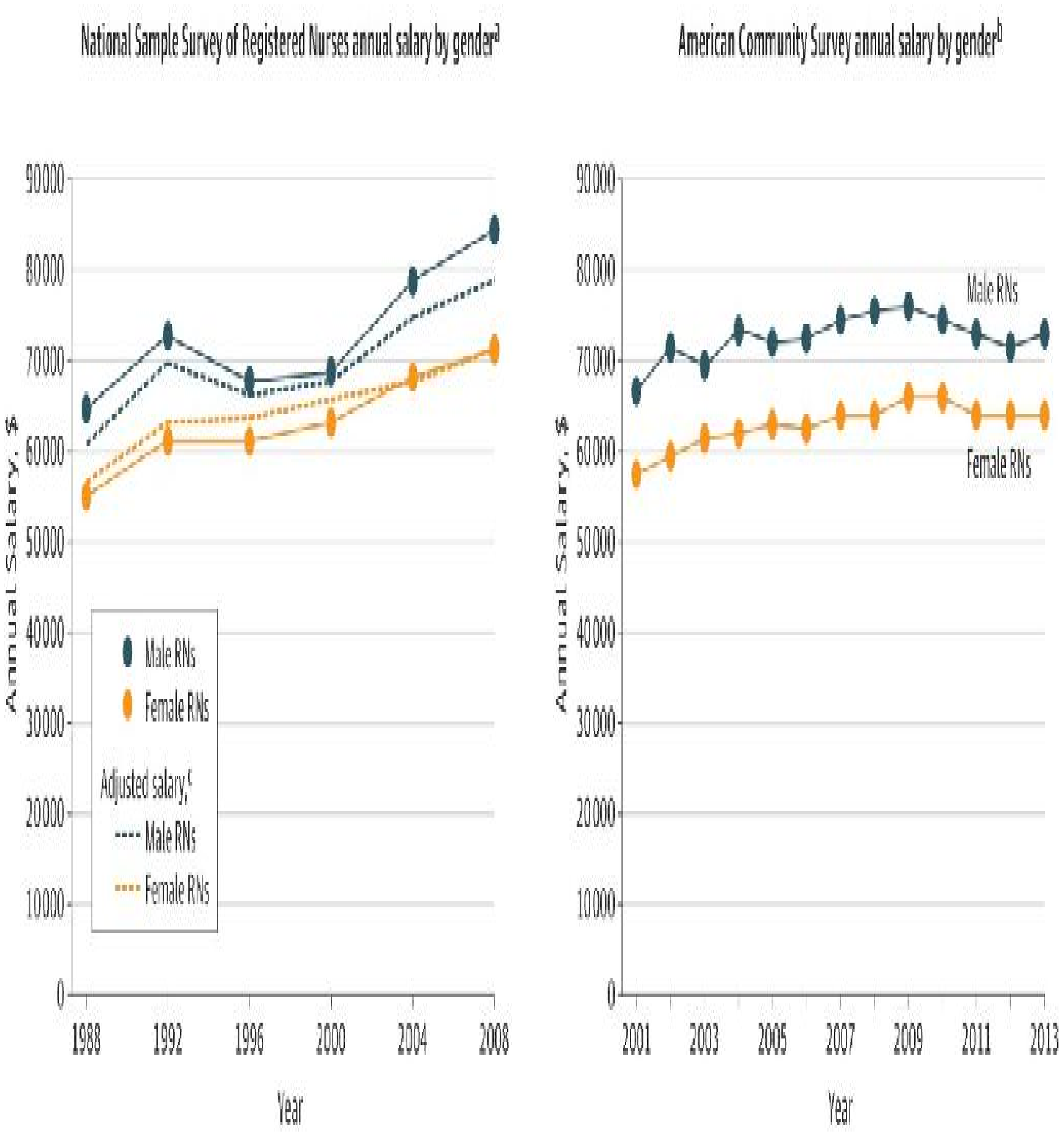
Here's a similar graph for professors (source):



Professor salaries are going up a little, but again, they're probably losing position relative to the average occupation. Also, note that although the average salary of each type of faculty is stable or increasing, the average salary of all faculty is going down. No mystery here – colleges are doing everything they can to switch from tenured professors to adjuncts, who complain of being overworked and abused while making about the same amount as a Starbucks barista.

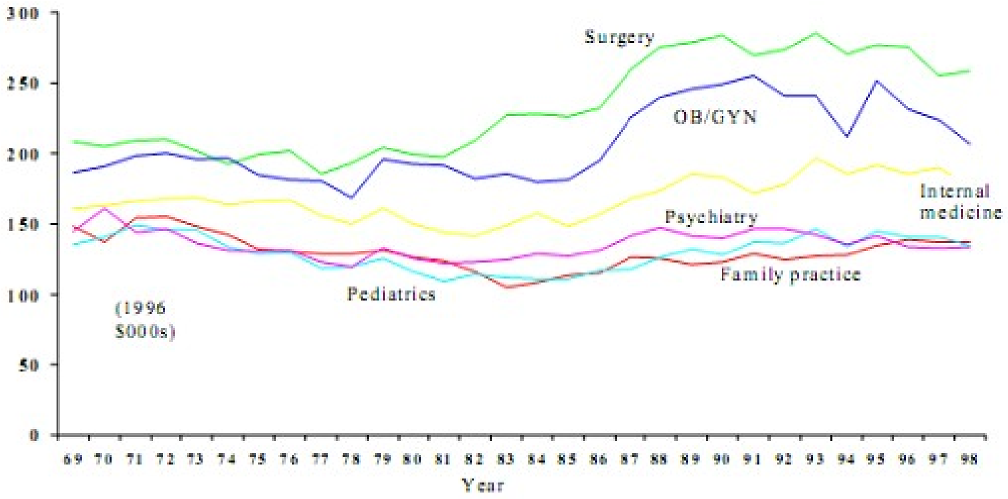
This seems to me a lot like the case of the hospitals cutting care for new mothers. The price of the service dectuples, yet at the same time the service has to sacrifice quality in order to control costs.

And speaking of hospitals, here’s the graph for nurses (source):



Female nurses' salaries went from about $55,000 in 1988 to $63,000 in 2013. This is probably around the average wage increase during that time. Also, some of this reflects changes in education: in the 1980s only 40% of nurses had a degree; by 2010, about 80% did.

And for doctors (source)



Stable again! Except that a lot of doctors' salaries now go to paying off their medical school debt, which has been ballooning like everything eles.

I don't have a similar graph for subway workers, but come on. The overall pictures is that health care and education costs have managed to increase by ten times without a single cent of the gains going to teachers, doctors, or nurses. Indeed these professions seem to have lost ground salary-wise relative to others.

I also want to add some anecdote to these hard facts. My father is a doctor and my mother is a teacher, so I got to hear a lot about how these professions have changed over the past generation. It seems at least a little like the adjunct story, although without the clearly defined "professor vs.

adjunct" dichotomy that makes it so easy to talk about. Doctors are really, really, *really* unhappy. When I went to medical school, some of my professors would tell me outright that they couldn't believe anyone would still go into medicine with all of the new stresses and demands placed on doctors. This doesn't seem to be limited to one medical school. *Wall Street Journal*: Why Doctors Are Sick Of Their Profession – "American physicians are increasingly unhappy with their once-vaunted profession, and that malaise is bad for their patients". *The Daily Beast*: How Being A Doctor Became The Most Miserable Profession – "Being a doctor has become a miserable and humiliating undertaking. Indeed, many doctors feel that America has declared war on physicians". *Forbes*: Why Are Doctors So Unhappy? – "Doctors have become like everyone else: insecure, discontent and scared about the future." *Vox*: Only Six Percent Of Doctors Are Happy With Their Jobs. *Al Jazeera America*: Here's Why Nine Out Of Ten Doctors Wouldn't Recommend Medicine As A Profession. Read these articles and they all say the same thing that all the doctors I know say – medicine used to be a well-respected, enjoyable profession where you could give patients good care and feel self-actualized. Now it kind of sucks.

Meanwhile, I also see articles like this piece from NPR saying teachers are experiencing historic stress levels and up to 50% say their job "isn't worth it". Teacher job satisfaction is at historic lows. And the veteran teachers I know say the same thing as the veteran doctors I know – their jobs used to be enjoyable and make them feel like they were making a difference; now they feel overworked, unappreciated, and trapped in mountains of paperwork.

It might make sense for these fields to become more expensive if their employees' salaries were increasing. And it might make sense for salaries to stay the same if employees instead benefitted from lower workloads and better working conditions. But neither of these are happening.

IV.

So what's going on? Why *are* costs increasing so dramatically? Some possible answers:

First, can we dismiss all of this as an illusion? Maybe adjusting for inflation is harder than I think. Inflation is an average, so some things have to have higher-than-average inflation; maybe it's education, health care, etc. Or maybe my sources have the wrong statistics.

But I don't think this is true. The last time I talked about this problem, someone mentioned they're running a private school which does just as well as public schools but costs only $3000/student/year, a fourth of the usual rate. Marginal Revolution notes that India has a private health system that delivers the same quality of care as its public system for a quarter of the cost. Whenever the same drug is provided by the official US health system and some kind of grey market supplement sort of thing, the grey market supplement costs between a fifth and a tenth as much; for example, Google's first hit for Deplin®, official prescription L-methylfolate, costs $175 for a month's supply; unregulated L-methylfolate supplement delivers the same dose for about $30. And this isn't even mentioning things like the $1 bag of saline that costs $700 at hospitals. Since it seems like it's not too hard to do things for a fraction of what we currently do things for, probably we should be less reluctant to believe that the cost of everything is really inflated.

Second, might markets just not work? I know this is kind of an extreme question to ask in a post on economics, but maybe nobody knows what they're doing in a lot of these fields and people can just increase costs and not suffer any decreased demand because of it. Suppose that people proved beyond a shadow of a doubt that Khan Academy could teach you just as much as a normal college education, but for free. People would still ask questions like – will employers accept my Khan Academy degree? Will it look good on a resume? Will people make fun of me for it? The same is true of community colleges, second-tier colleges, for-profit colleges, et cetera. I got offered a free scholarship to a mediocre state college, and I turned it down on the grounds that I knew nothing about anything and maybe years from now I would be locked out of some sort of Exciting Opportunity because my college wasn't prestigious enough. Assuming everyone thinks like this, can colleges just charge whatever they want?

Likewise, my workplace offered me three different health insurance plans, and I chose the middle-expensiveness one, on the grounds that I had no idea how health insurance worked but maybe if I bought the cheap one I'd get sick and regret my choice, and maybe if I bought the expensive one I *wouldn't* be sick and regret my choice. I am a doctor, my employer is a hospital, and the health insurance was for treatment in my own health system. The moral of the story is that I am an idiot. The second moral of the story is that people probably are not super-informed health care consumers.

This can't be pure price-gouging, since corporate profits haven't increased nearly enough to be where all the money is going. But a while ago a commenter linked me to the Delta Cost Project, which scrutinizes the exact causes of increasing college tuition. Some of it is the administrative bloat that you would expect. But a lot of it is fun "student life" types of activities like clubs, festivals, and paying Milo Yiannopoulos to speak and then cleaning up after the ensuing riots. These sorts of things improve the student experience, but I'm not sure that the average student would rather go to an expensive college with clubs/festivals/Milo than a cheap college without them. More important, it doesn’t really seem like the average student is offered this choice.

This kind of suggests a picture where colleges expect people will pay whatever price they set, so they set a very high price and then use the money for cool things and increasing their own prestige. Or maybe clubs/festivals/Milo become such a signal of prestige that students avoid colleges that don’t comply since they worry their degrees won’t be respected? Some people have pointed out that hospitals have switched from many-people-all-in-a-big-ward to private rooms. Once again, nobody seems to have been offered the choice between expensive hospitals with private rooms versus cheap hospitals with roommates. It’s almost as if industries have their own reasons for switching to more-bells-and-whistles services that people don’t necessarily want, and consumers just go along with it because for some reason they’re not exercising choice the same as they would in other markets.

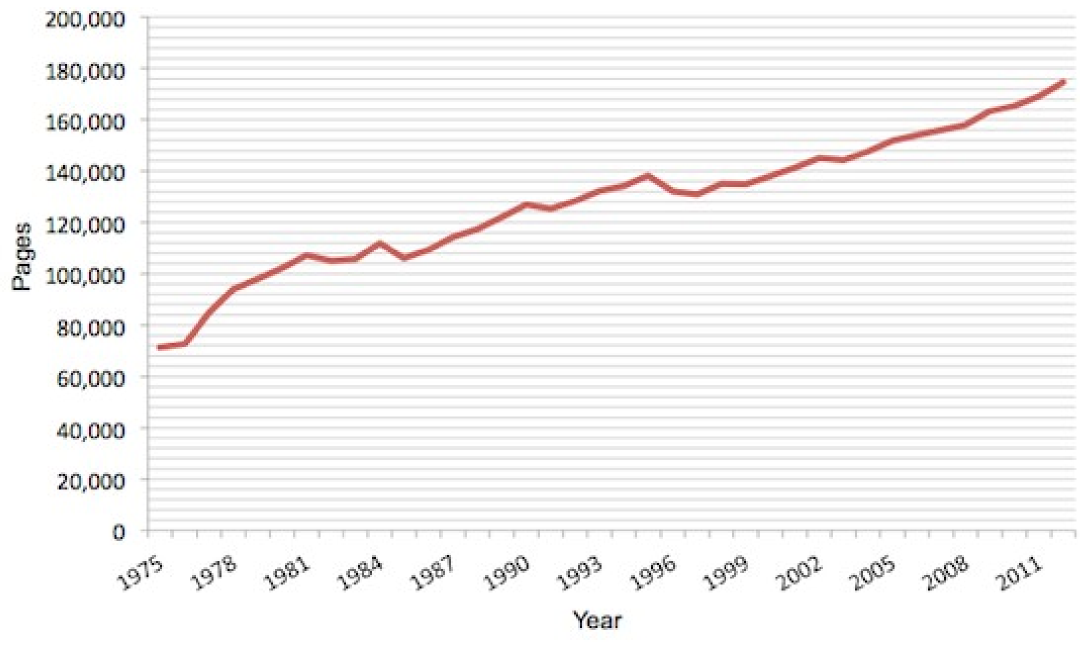
(this article on the Oklahoma City Surgery Center might be about a partial corrective for this kind of thing)

Third, can we attribute this to the inefficiency of government relative to private industry? I don’t think so. The government handles most primary education and subways, and has its hand in health care. But we know that for-profit hospitals aren’t much cheaper than government hospitals, and that private schools usually aren’t much cheaper (and are sometimes more expensive) than government schools. And private colleges cost more than government-funded ones.

Fourth, can we attribute it to indirect government intervention through regulation, which public and private companies alike must deal with? This seems to be at least part of the story in health care, given how much money you can save by grey-market practices that avoid the FDA. It’s harder to apply it to colleges, though some people have pointed out regulations like Title IX that affect the educational sector.

One factor that seems to speak out against this is that starting with Reagan in 1980, and picking up steam with Gingrich in 1994, we got an increasing presence of Republicans in government who declared war on overregulation – but the cost disease proceeded unabated. This is suspicious, but in fairness to the Republicans, they did sort of fail miserably at deregulating things. “The literal number of pages in the regulatory code” is kind of a blunt instrument, but it doesn’t exactly inspire confidence in the Republicans’ deregulation efforts:

Figure 1: Total Number of Pages in the Code of Federal Regulations 1975 -- 2011



Here’s a more interesting (and more fun) argument against regulations being to blame: what about pet health care? Veterinary care is much less regulated than human health care, yet its cost is rising as fast (or faster) than that of the human medical system (popular article, study). I’m not sure what to make of this.

Fifth, might the increased regulatory complexity happen not through literal regulations, but through fear of lawsuits? That is, might institutions add extra layers of administration and expense not because they’re *forced* to, but because they fear being sued if they don’t and then something goes wrong?

I see this all the time in medicine. A patient goes to the hospital with a heart attack. While he’s recovering, he tells his doctor that he’s really upset about all of this. Any normal person would say “You had a heart attack, of course you’re upset, get over it.” But if his doctor says this, and then a year later he commits suicide for some unrelated reason, his family can sue the doctor for “not picking up the warning signs” and win several million dollars. So now the doctor consults a psychiatrist, who does an hour-long evaluation, charges the insurance company $500, and determines using her immense clinical expertise that the patient is upset because he just had a heart attack.

Those outside the field have *no idea* how much of medicine is built on this principle. People often say that the importance of lawsuits to medical cost increases is overrated because malpractice insurance doesn’t cost that much, but the situation above would never look lawsuit-related; the whole thing only works because everyone involved documents it as well-justified psychiatric consult to investigate depression. Apparently some studies suggest this isn’t happening, but all they do is survey doctors, and with all due respect all the doctors *I* know say the opposite.

This has nothing to do with government regulations (except insofar as these make lawsuits easier or harder), but it sure can drive cost increases, and it might apply to fields outside medicine as well.

Sixth, might we have changed our level of risk tolerance? That is, might increased caution be due not purely to lawsuitphobia, but to really caring more about whether or not people are protected? I read stuff every so often about how playgrounds are becoming obsolete because nobody wants to let kids run around unsupervised on something with sharp edges. Suppose that one in 10,000 kids get a horrible playground-related injury. Is it worth making playgrounds cost twice as much and be half as fun in order to decrease that number to one in 100,000? This isn't a rhetorical question; I think different people can have legitimately different opinions here (though there are probably some utilitarian things we can do to improve them).

To bring back the lawsuit point, some of this probably relates to a difference between personal versus institutional risk tolerance. Every so often, an elderly person getting up to walk to the bathroom will fall and break their hip. This is a fact of life, and elderly people deal with it every day. Most elderly people I know don't spend thousands of dollars fall-proofing the route from their bed to their bathroom, or hiring people to watch them at every moment to make sure they don't fall, or buy a bedside commode to make bathroom-related falls impossible. This suggests a revealed preference that elderly people are willing to tolerate a certain fall probability in order to save money and convenience. Hospitals, which face huge lawsuits if any elderly person falls on the premises, are *not* willing to tolerate that probability. They put rails on elderly people's beds, place alarms on them that will go off if the elderly person tries to leave the bed without permission, and hire patient care assistants who among other things go around carefully holding elderly people upright as they walk to the bathroom (I assume this job will soon require at least a master’s degree). As more things become institutionalized and the level of acceptable institutional risk tolerance becomes lower, this could shift the cost-risk tradeoff even if there isn’t a population-level trend towards more risk-aversion.

Seventh, might things cost more for the people who pay because so many people don’t pay? This is somewhat true of colleges, where an increasing number of people are getting in on scholarships funded by the tuition of non-scholarship students. I haven’t been able to find great statistics on this, but one argument against: couldn’t a college just not fund scholarships, and offer much lower prices to its paying students? I get that scholarships are good and altruistic, but it would be surprising if every single college thought of its role as an altruistic institution, and cared about it more than they cared about providing the same service at a better price. I guess this is related to my confusion about why more people don’t open up colleges. Maybe this is the “smart people are rightly too scared and confused to go to for-profit colleges, and there’s not enough ability to discriminate between the good and the bad ones to make it worthwhile to found a good one” thing again.

This also applies in health care. Our hospital (and every other hospital in the country) has some “frequent flier” patients who overdose on meth at least once a week. They comes in, get treated for their meth overdose (we can’t legally turn away emergency cases), get advised to get help for their meth addiction (without the slightest expectation that they will take our advice) and then get discharged. Most of them are poor and have no insurance, but each admission costs a couple of thousand dollars. The cost gets paid by a combination of taxpayers and other hospital patients with good insurance who get big markups on their own bills.

Eighth, might total compensation be increasing even though wages aren't? There definitely seems to be a pensions crisis, especially in a lot of government work, and it's possible that some of this is going to pay the pensions of teachers, etc. My understanding is that in general pensions aren't really increasing much faster than wages, but this might not be true in those specific industries. Also, this might pass the buck to the question of why we need to spend more on pensions now than in the past. I don't think increasing life expectancy explains all of this, but I might be wrong.

IV.

I mentioned politics briefly above, but they probably deserve more space here. Libertarian-minded people keep talking about how there's too much red tape and the economy is being throttled. And less libertarian-minded people keep interpreting it as not caring about the poor, or not understanding that government has an important role in a civilized society, or as a "dog whistle" for racism, or whatever. I don't know why more people don't just come out and say "LOOK, REALLY OUR MAIN PROBLEM IS THAT ALL THE MOST IMPORTANT THINGS COST TEN TIMES AS MUCH AS THEY USED TO FOR NO REASON, PLUS THEY SEEM TO BE GOING DOWN IN QUALITY, AND NOBODY KNOWS WHY, AND WE'RE MOSTLY JUST DESPERATELY FLAILING AROUND LOOKING FOR SOLUTIONS HERE." State that clearly, and a lot of political debates take on a different light.

For example: some people promote free universal college education, remembering a time when it was easy for middle class people to afford college if they wanted it. Other people oppose the policy, remembering a time when people didn't depend on government handouts. Both are true! My uncle paid for his tuition at a really good college just by working a pretty easy summer job – not so hard when college cost a tenth of what it did now. The modern conflict between opponents and proponents of free college education is over how to distribute our losses. In the old days, we could combine low taxes with widely available education. Now we can't, and we have to argue about which value to sacrifice.

Or: some people get upset about teachers' unions, saying they must be sucking the "dynamism" out of education because of increasing costs. Others people fiercely defend them, saying teachers are underpaid and overworked. Once again, in the context of cost disease, both are obviously true. The taxpayers are just trying to protect their right to get education as cheaply as they used to. The teachers are trying to protect their right to make as much money as they used to. The conflict between the taxpayers and the teachers' unions is about how to distribute losses; *somebody* is going to have to be worse off than they were a generation ago, so who should it be?

And the same is true to greater or lesser degrees in the various debates over health care, public housing, et cetera.

Imagine if tomorrow, the price of water dectupled. Suddenly people have to choose between drinking and washing dishes. Activists argue that taking a shower is a basic human right, and grumpy talk show hosts point out that in *their* day, parents taught their children not to waste water. A coalition promotes laws ensuring government-subsidized free water for poor families; a Fox News investigative report shows that some people receiving water on the government dime are taking long luxurious showers. Everyone gets really angry and there's lots of talk about basic compassion and personal responsibility and whatever but all of this is secondary to *why does water costs ten times what it used to?*

I think this is the basic intuition behind so many people, even those who genuinely want to help the poor, are afraid of "tax and spend” policies. In the context of cost disease, these look like industries constantly doubling, tripling, or dectupling their price, and the government saying “Okay, fine,” and increasing taxes however much it costs to pay for whatever they’re demanding now.

If we give everyone free college education, that solves a big social problem. It also locks in a price which is ten times too high for no reason. This isn’t fair to the government, which has to pay ten times more than it should. It’s not fair to the poor people, who have to face the stigma of accepting handouts for something they could easily have afforded themselves if it was at its proper price. And it’s not fair to future generations if colleges take this opportunity to increase the cost by *twenty* times, and then our children have to subsidize *that*.

I’m not sure how many people currently opposed to paying for free health care, or free college, or whatever, would be happy to pay for health care that cost less, that was less wasteful and more efficient, and whose price we expected to go down rather than up with every passing year. I expect it would be a lot.

And if it isn’t, who cares? The people who want to help the poor have enough political capital to spend eg $500 billion on Medicaid; if that were to go ten times further, then everyone could get the health care they need without any more political action needed. If some government program found a way to give poor people good health insurance for a few hundred dollars a year, college tuition for about a thousand, and housing for only two-thirds what it costs now, that would be the greatest anti-poverty advance in history. That program is called “having things be as efficient as they were a few decades ago”.

V.

In 1930, economist John Maynard Keynes predicted that his grandchildrens’ generation would have a 15 hour work week. At the time, it made sense. GDP was rising so quickly that anyone who could draw a line on a graph could tell that our generation would be four or five times richer than his. And the average middle-class person in his generation felt like they were doing pretty well and had most of what they needed. Why *wouldn’t* they decide to take some time off and settle for a lifestyle merely twice as luxurious as Keynes’ own?

Keynes was sort of right. GDP per capita *is* 4-5x greater today than in his time. Yet we still work forty hour weeks, and some large-but-inconsistently-reported percent of Americans (76? 55? 47?) still live paycheck to paycheck.

And yes, part of this is because inequality is increasing and most of the gains are going to the rich. But this alone wouldn’t be a disaster; we’d get to Keynes’ utopia a little slower than we might otherwise, but eventually we’d get there. Most gains going to the rich means at least some gains are going to the poor. And at least there’s a lot of mainstream awareness of the problem.

I’m more worried about the part where the cost of basic human needs goes up faster than wages do. Even if you’re making twice as much money, if your health care and education and so on cost ten times as much, you’re going to start falling behind. Right now the standard of living isn’t just stagnant, it’s at risk of declining, and a lot of that is student loans and health insurance costs and so on.

What’s happening? I don’t know and I find it really scary.

I Can Tolerate Anything Except The Outgroup

[Content warning: Politics, religion, social justice, spoilers for "The Secret of Father Brown". This isn't especially original to me and I don't claim anything more than to be explaining and rewording things I have heard from a bunch of other people. Unapologetically America-centric because I'm not informed enough to make it otherwise. Try to keep this off Reddit and other similar sorts of things.]

I.

In Chesterton's *The Secret of Father Brown*, a beloved nobleman who murdered his good-for-nothing brother in a duel thirty years ago returns to his hometown wracked by guilt. All the townspeople want to forgive him immediately, and they mock the titular priest for only being willing to give a measured forgiveness conditional on penance and self-reflection. They lecture the priest on the virtues of charity and compassion.

Later, it comes out that the beloved nobleman did *not* in fact kill his good-for-nothing brother. The good-for-nothing brother killed the beloved nobleman (and stole his identity). *Now* the townspeople want to see him lynched or burned alive, and it is only the priest who – consistently – offers a measured forgiveness conditional on penance and self-reflection.

The priest tells them:

> It seems to me that you only pardon the sins that you don't really think sinful. You only forgive criminals when they commit what you don't regard as crimes, but rather as conventions. You forgive a conventional duel

> just as you forgive a conventional divorce. You forgive because there isn’t anything to be forgiven.

He further notes that this is why the townspeople can self-righteously consider themselves more compassionate and forgiving than he is. Actual forgiveness, the kind the priest needs to cultivate to forgive evildoers, is really really hard. The fake forgiveness the townspeople use to forgive the people they like is really easy, so they get to boast not only of their forgiving nature, but of how much nicer they are than those mean old priests who find forgiveness difficult and want penance along with it.

After some thought I agree with Chesterton’s point. There are a lot of people who say “I forgive you” when they mean “No harm done”, and a lot of people who say “That was unforgiveable” when they mean “That was genuinely really bad”. Whether or not forgiveness is *right* is a complicated topic I do not want to get in here. But since forgiveness is generally considered a virtue, and one that many want credit for having, I think it’s fair to say you only earn the right to call yourself ‘forgiving’ if you forgive things that genuinely hurt you.

To borrow Chesterton’s example, if you think divorce is a-ok, then you don’t get to “forgive” people their divorces, you merely ignore them. Someone who thinks divorce is abhorrent can “forgive” divorce. *You* can forgive theft, or murder, or tax evasion, or something *you* find abhorrent.

I mean, from a utilitarian point of view, you are still doing the correct action of not giving people grief because they’re a divorcee. You can have all the Utility Points you want. All I’m saying is that if you “forgive” something you don’t care about, you don’t earn any Virtue Points.

(by way of illustration: a billionaire who gives $100 to charity gets as many Utility Points as an impoverished pensioner who donates the same amount, but the latter gets a lot more Virtue Points)

Tolerance is also considered a virtue, but it suffers the same sort of dimished expectations forgiveness does.

The Emperor summons before him Bodhidharma and asks: “Master, I have been tolerant of innumerable gays, lesbians, bisexuals, asexuals, blacks, Hispanics, Asians, transgender people, and Jews. How many Virtue Points have I earned for my meritorious deeds?”

Bodhidharma answers: “None at all”.

The Emperor, somewhat put out, demands to know why.

Bodhidharma asks: “Well, what do you think of gay people?”

The Emperor answers: “What do you think I am, some kind of homophobic bigot? Of course I have nothing against gay people!”

And Bodhidharma answers: “Thus do you gain no merit by tolerating them!”

II.

If I had to define “tolerance” it would be something like “respect and kindness toward members of an outgroup”.

And today we have an almost unprecedented situation.

We have a lot of people – like the Emperor – boasting of being able to tolerate everyone from every outgroup they can imagine, loving the outgroup, writing long paeans to how great the outgroup is, staying up at night fretting that somebody else might not like the outgroup enough.

This is really surprising. It’s a total reversal of everything we know about human psychology up to this point. No one did any genetic engineering. No one passed out weird glowing pills in the public schools. And yet suddenly we get an entire group of people who conspicuously promote and defend their outgroups, the outer the better.

What is going on here?

Let’s start by asking what exactly an outgroup is.

There’s a very boring sense in which, assuming the Emperor’s straight, gays are part of his “outgroup” ie a group that he is not a member of. But if the Emperor has curly hair, are straight-haired people part of his outgroup? If the Emperor’s name starts with the letter ‘A’, are people whose names start with the letter ‘B’ part of his outgroup?

Nah. I would differentiate between multiple different meanings of outgroup, where one is “a group you are not a part of” and the other is…something stronger.

I want to avoid a very easy trap, which is saying that outgroups are about how different you are, or how hostile you are. I don’t think that’s quite right.

Compare the Nazis to the German Jews and to the Japanese. The Nazis were very similar to the German Jews: they looked the same, spoke the same language, came from a similar culture. The Nazis were totally different from the Japanese: different race, different language, vast cultural gap. But the Nazis and Japanese mostly got along pretty well. Heck, the Nazis were actually moderately positively disposed to the *Chinese*, even when they were technically at war.

Meanwhile, the conflict between the Nazis and the German Jews – some of whom didn't even realize they were anything other than German until they checked their grandparents' birth certificate – is the stuff of history and nightmares. Any theory of outgroupishness that naively assumes the Nazis' natural outgroup is Japanese or Chinese people will be totally inadequate.

And this isn't a weird exception. Freud spoke of the narcissism of small differences, saying that "it is precisely communities with adjoining territories, and related to each other in other ways as well, who are engaged in constant feuds and ridiculing each other". Nazis and German Jews. Northern Irish Protestants and Northern Irish Catholics. Hutus and Tutsis. South African whites and South African blacks. Israeli Jews and Israeli Arabs. Anyone in the former Yugoslavia and anyone else in the former Yugoslavia.

So what makes an outgroup? Proximity plus small differences. If you want to know who someone in former Yugoslavia hates, don't look at the Indonesians or the Zulus or the Tibetans or anyone else distant and exotic. Find the Yugoslavian ethnicity that lives closely intermingled with them and is most conspicuously similar to them, and chances are you'll find the one who they have eight hundred years of seething hatred toward.

What makes an unexpected in-group? The answer with Germans and Japanese is obvious – a strategic alliance. In fact, the World Wars forged a lot of unexpected temporary pseudo-friendships. A recent article from War Nerd points out that the British, after spending centuries subjugating and despising the Irish and Sikhs, suddenly needed Irish and Sikh soldiers for World Wars I and II respectively. "Crush them beneath our boots" quickly changed to fawning songs about how “there never was a coward where the shamrock grows” and endless paeans to Sikh military prowess.

Sure, scratch the paeans even a little bit and you find condescension as strong as ever. But eight hundred years of the British committing genocide against the Irish and considering them literally subhuman turned into smiles and songs about shamrocks once the Irish started looking like useful cannon fodder for a larger fight. And the Sikhs, dark-skinned people with turbans and beards who pretty much exemplify the European stereotype of “scary foreigner”, were lauded by everyone from the news media all the way up to Winston Churchill.

In other words, outgroups may be the people who look exactly like you, and scary foreigner types can become the in-group on a moment’s notice when it seems convenient.

III.

There are certain theories of dark matter where it barely interacts with the regular world *at all*, such that we could have a dark matter planet exactly co-incident with Earth and never know. Maybe dark matter people are walking all around us and through us, maybe my house is in the Times Square of a great dark matter city, maybe a few meters away from me a dark matter blogger is writing on his dark matter computer about how weird it would be if there was a light matter person he couldn’t see right next to him.

This is sort of how I feel about conservatives.

I don’t mean the sort of light-matter conservatives who go around complaining about Big Government and occasionally voting for Romney. I see those guys all the time. What I mean is – well, take creationists. According to Gallup polls, about 46% of Americans are creationists. Not just in the sense of believing God helped guide evolution. I mean they think evolution is a vile atheist lie and God created humans exactly as they exist right now. That's half the country.

And I don't have a *single one of those people* in my social circle. It's not because I'm deliberately avoiding them; I'm pretty live-and-let-live politically, I wouldn't ostracize someone just for some weird beliefs. And yet, even though I probably know about a hundred fifty people, I am pretty confident that not one of them is creationist. Odds of this happening by chance? 1/2^150 = 1/10^45 = approximately the chance of picking a particular atom if you are randomly selecting among all the atoms on Earth.

About forty percent of Americans want to ban gay marriage. I think if I *really* stretch it, maybe ten of my top hundred fifty friends might fall into this group. This is less astronomically unlikely; the odds are a mere one to one hundred quintillion against.

People like to talk about social bubbles, but that doesn't even begin to cover one hundred quintillion. The only metaphor that seems really appropriate is the bizarre dark matter world.

I live in a Republican congressional district in a state with a Republican governor. The conservatives are definitely out there. They drive on the same roads as I do, live in the same neighborhoods. But they might as well be made of dark matter. I never meet them.

To be fair, I spend a lot of my time inside on my computer. I'm browsing sites like Reddit.

Recently, there was a thread on Reddit asking – Redditors Against Gay Marriage, What Is Your Best Supporting Argument? A Reddit user who didn't understand how anybody could be against gay marriage honestly wanted to know how other people who *were* against it justified their position. He figured he might as well ask one of the largest sites on the Internet, with an estimated user base in the tens of millions.

It soon became clear that nobody there was actually against gay marriage.

There were a bunch of posts saying "I of course support gay marriage but here are some reasons some other people might be against it," a bunch of others saying "my argument against gay marriage is the government shouldn't be involved in the marriage business at all", and several more saying "why would you even ask this question, there's no possible good argument and you're wasting your time". About halfway through the thread someone started saying homosexuality was unnatural and I *thought* they were going to be the first one to actually answer the question, but at the end they added "But it's not my place to decide what is or isn't natural, I'm still pro-gay marriage."

In a thread with 10,401 comments, a thread *specifically* asking for people against gay marriage, I was eventually able to find *two* people who came out and opposed it, way near the bottom. Their posts started with "I know I'm going to be downvoted to hell for this…"

But I'm not only on Reddit. I also hang out on LW.

On last year's survey, I found that of American LWers who identify with one of the two major political parties, 80% are Democrat and 20% Republican, which actually sounds pretty balanced compared to some of these other examples.

But it doesn't last. Pretty much all of those "Republicans" are libertarians who consider the GOP the lesser of two evils. When allowed to choose “libertarian” as an alternative, only 4% of visitors continued to identify as conservative. But that’s still…some. Right?

When I broke the numbers down further, 3 percentage points of those are neoreactionaries, a bizarre sect that wants to be ruled by a king. Only *one percent* of LWers were normal everyday God-‘n-guns-but-not-George-III conservatives of the type that seem to make up about half of the United States.

It gets worse. My formative years were spent at a university which, if it was similar to other elite universities, had a faculty and a student body that skewed about 90-10 liberal to conservative – and we can bet that, like LW, even those few token conservatives are Mitt Romney types rather than God-n’-guns types. I get my news from vox.com, an Official Liberal Approved Site. Even when I go out to eat, it turns out my favorite restaurant, California Pizza Kitchen, is the most liberal restaurant in the United States.

I inhabit the same geographical area as *scores and scores* of conservatives. But without meaning to, I have created an *outrageously* strong bubble, a 10^45 bubble. Conservatives are all around me, yet I am about as likely to have a serious encounter with one as I am a Tibetan lama.

(Less likely, actually. One time a Tibetan lama came to my college and gave a really nice presentation, but if a conservative tried that, people would protest and it would be canceled.)

IV.

One day I realized that entirely by accident I was fulfilling *all* the Jewish stereotypes.

I’m nerdy, over-educated, good with words, good with money, weird sense of humor, don’t get outside much, I like deli sandwiches. And I’m a psychiatrist, which is about the most stereotypically Jewish profession short of maybe stand-up comedian or rabbi.

I’m not very religious. And I don’t go to synagogue. But *that’s* stereotypically Jewish too!

I bring this up because it would be a mistake to think “Well, a Jewish person is by definition someone who is born of a Jewish mother. Or I guess it sort of also means someone who follows the Mosaic Law and goes to synagogue. But I don’t care about Scott’s mother, and I know he doesn’t go to synagogue, so I can’t gain any useful information from knowing Scott is Jewish.”

The defining factors of Judaism – Torah-reading, synagogue-following, mother-having – are the tip of a giant iceberg. Jews sometimes identify as a “tribe”, and even if you don’t attend synagogue, you’re still a member of that tribe and people can still (in a statistical way) infer things about you by knowing your Jewish identity – like how likely they are to be psychiatrists.

The last section raised a question – if people rarely select their friends and associates and customers explicitly for politics, how do we end up with such intense political segregation?

Well, in the same way “going to synagogue” is merely the iceberg-tip of a Jewish tribe with many distinguishing characteristics, so “voting Republican” or “identifying as conservative” or “believing in creationism” is the iceberg-tip of a conservative tribe with many distinguishing characteristics.

A disproportionate number of my friends are Jewish, because I meet them at psychiatry conferences or something – we self-segregate not based on explicit religion but on implicit tribal characteristics. So in the same way, political tribes self-segregate to an impressive extent – a 1/10^45 extent, I will never tire of hammering in – based on their implicit tribal characteristics.

The people who are actually into this sort of thing sketch out a bunch of speculative tribes and subtribes, but to make it easier, let me stick with two and a half.

The Red Tribe is most classically typified by conservative political beliefs, strong evangelical religious beliefs, creationism, opposing gay marriage, owning guns, eating steak, drinking Coca-Cola, driving SUVs, watching lots of TV, enjoying American football, getting conspicuously upset about terrorists and commies, marrying early, divorcing early, shouting “USA IS NUMBER ONE!!!”, and listening to country music.

The Blue Tribe is most classically typified by liberal political beliefs, vague agnosticism, supporting gay rights, thinking guns are barbaric, eating arugula, drinking fancy bottled water, driving Priuses, reading lots of books, being highly educated, mocking American football, feeling vaguely like they should like soccer but never really being able to get into it, getting conspicuously upset about sexists and bigots, marrying later, constantly pointing out how much more civilized European countries are than America, and listening to “everything except country”.

(There is a partly-formed attempt to spin off a Grey Tribe typified by libertarian political beliefs, Dawkins-style atheism, vague annoyance that the question of gay rights even comes up, eating paleo, drinking Soylent, calling in rides on Uber, reading lots of blogs, calling American football “sportsball”, getting conspicuously upset about the War on Drugs and the NSA, and listening to filk – but for our current purposes this is a distraction and they can safely be considered part of the Blue Tribe most of the time)

I think these “tribes” will turn out to be even stronger categories than politics. Harvard might skew 80-20 in terms of Democrats vs. Republicans, 90-10 in terms of liberals vs. conservatives, but maybe 99-1 in terms of Blues vs. Reds.

It’s the many, many differences between these tribes that explain the strength of the filter bubble – which *have I mentioned* segregates people at a strength of 1/10^45? Even in something as seemingly politically uncharged as going to California Pizza Kitchen or Sushi House for dinner, I’m restricting myself to the set of people who like cute artisanal pizzas or sophsticated foreign foods, which are classically Blue Tribe characteristics.

Are these tribes based on geography? Are they based on race, ethnic origin, religion, IQ, what TV channels you watched as a kid? I don’t know.

Some of it is certainly genetic – estimates of the genetic contribution to political association range from 0.4 to 0.6. Heritability of one’s attitudes toward gay rights range from 0.3 to 0.5, which hilariously is a little more heritable than homosexuality itself.

(for an interesting attempt to break these down into more rigorous concepts like “traditionalism”, “authoritarianism”, and “in-group favoritism” and find the genetic loading for each see here. For an attempt to trace the specific genes involved, which mostly turn out to be NMDA receptors, see here)

But I don’t think it’s just genetics. There’s something else going on too. The word “class” seems like the closest analogue, but only if you use it in the sophisticated Paul Fussell *Guide Through the American Status System* way instead of the boring “another word for how much money you make” way.

For now we can just accept them as a brute fact – as multiple coexisting societies that might as well be made of dark matter for all of the interaction they have with one another – and move on.

V.

The worst reaction I’ve ever gotten to a blog post was when I wrote about the death of Osama bin Laden. I’ve written all sorts of stuff about race and gender and politics and whatever, but that was the worst.

I didn’t come out and say I was happy he was dead. But some people interpreted it that way, and there followed a bunch of comments and emails and Facebook messages about how could I possibly be happy about the death of another human being, even if he was a bad person? Everyone, even Osama, is a human being, and we should never rejoice in the death of a fellow man. One commenter came out and said:

> I’m surprised at your reaction. As far as people I casually stalk on the internet (ie, LJ and Facebook), you are the first out of the “intelligent, reasoned and thoughtful” group to be uncomplicatedly happy about this development and not to be, say, disgusted at the reactions of the other 90% or so.

This commenter was right. Of the “intelligent, reasoned, and thoughtful” people I knew, the overwhelming emotion was conspicuous disgust that other people could be happy about his death. I hastily backtracked and said I wasn't happy per se, just surprised and relieved that all of this was finally behind us.

And I genuinely believed that day that I had found some unexpected good in people – that everyone I knew was so humane and compassionate that they were unable to rejoice even in the death of someone who hated them and everything they stood for.

Then a few years later, Margaret Thatcher died. And on my Facebook wall – made of these same "intelligent, reasoned, and thoughtful" people – the most common response was to quote some portion of the song "Ding Dong, The Witch Is Dead". Another popular response was to link the videos of British people spontaneously throwing parties in the street, with comments like "I wish I was there so I could join in". From this exact same group of people, not a single expression of disgust or a "c'mon, guys, we're all human beings here."

I gently pointed this out at the time, and mostly got a bunch of "yeah, so what?", combined with links to an article claiming that "the demand for respectful silence in the wake of a public figure's death is not just misguided but dangerous".

And that was when something clicked for me.

You can talk all you want about Islamophobia, but my friend's "intelligent, reasoned, and thoughtful people" – her name for the Blue Tribe – can't get together enough energy to really hate Osama, let alone Muslims in general. We understand that what he did was bad, but it didn't anger us personally. When he died, we were able to very rationally apply our better nature and our Far Mode beliefs about how it’s never right to be happy about anyone else’s death.

On the other hand, that same group absolutely *loathed* Thatcher. Most of us (though not all) can agree, if the question is posed explicitly, that Osama was a worse person than Thatcher. But in terms of actual gut feeling? Osama provokes a snap judgment of “flawed human being”, Thatcher a snap judgment of “scum”.

I started this essay by pointing out that, despite what geographical and cultural distance would suggest, the Nazis’ outgroup was not the vastly different Japanese, but the almost-identical German Jews.

And my hypothesis, stated plainly, is that if you’re part of the Blue Tribe, then your outgroup isn’t al-Qaeda, or Muslims, or blacks, or gays, or transpeople, or Jews, or atheists – it’s the Red Tribe.

VI.

“But racism and sexism and cissexism and anti-Semitism are these giant all-encompassing social factors that verge upon being human universals! Surely you’re not arguing that mere *political* differences could ever come close to them!”

One of the ways we *know* that racism is a giant all-encompassing social factor is the Implicit Association Test. Psychologists ask subjects to quickly identify whether words or photos are members of certain gerrymandered categories, like “either a white person’s face or a positive emotion” or “either a black person’s face and a negative emotion”. Then they compare to a different set of gerrymandered categories, like “either a black person’s face or a positive emotion” or “either a white person’s face or a negative emotion.” If subjects have more trouble (as measured in latency time) connecting white people to negative things than they do white people to positive things, then they probably have subconscious positive associations with white people. You can try it yourself here.

Of course, what the test famously found was that even white people who claimed to have no racist attitudes at all usually had positive associations with white people and negative associations with black people on the test. There are very many claims and counterclaims about the precise meaning of this, but it ended up being a big part of the evidence in favor of the current consensus that all white people are at least a little racist.

Anyway, three months ago, someone finally had the bright idea of doing an Implicit Association Test with political parties, and they found that people’s unconscious partisan biases were *half again as strong* as their unconscious racial biases (h/t Bloomberg. For example, if you are a white Democrat, your unconscious bias against blacks (as measured by something called a d-score) is 0.16, but your unconscious bias against Republicans will be 0.23. The Cohen’s *d* for racial bias was 0.61, by the book a “moderate” effect size; for party it was 0.95, a “large” effect size.

Okay, fine, but we know race has *real world* consequences. Like, there have been several studies where people sent out a bunch of identical resumes except sometimes with a black person’s photo and other times with a white person’s photo, and it was noticed that employers were much more likely to invite the fictional white candidates for interviews. So just some stupid Implicit Association Test results can’t compare to that, right?

Iyengar and Westwood also decided to do the resume test for parties. They asked subjects to decide which of several candidates should get a scholarship (subjects were told this was a genuine decision for the university the researchers were affiliated with). Some resumes had photos of black people, others of white people. And some students listed their experience in Young Democrats of America, others in Young Republicans of America.

Once again, discrimination on the basis of party was much stronger than discrimination on the basis of race. The size of the race effect for white people was only 56-44 (and in the reverse of the expected direction); the size of the party effect was about 80-20 for Democrats and 69-31 for Republicans.

If you want to see their third experiment, which applied *yet another* classic methodology used to detect racism and *once again* found partyism to be much stronger, you can read the paper.

I & W did an unusually thorough job, but this sort of thing isn't new or ground-breaking. People have been studying "belief congruence theory" – the idea that differences in beliefs are more important than demographic factors in forming in-groups and outgroups – for decades. As early as 1967, Smith et al were doing surveys all over the country and finding that people were more likely to accept friendships across racial lines than across beliefs; in the forty years since then, the observation has been replicated scores of times. Insko, Moe, and Nacoste's 2006 review Belief Congruence And Racial Discrimination concludes that:

> . The literature was judged supportive of a weak version of belief congruence theory which states that in those contexts in which social pressure is nonexistent or

> ineffective, belief is more important than race as a determinant of racial or ethnic discrimination. Evidence for a strong version of belief congruence theory (which states that in those contexts in which social pressure is nonexistent, or ineffective, belief is the only determinant of racial or ethnic discrimination) and was judged much more problematic.

One of the best-known examples of racism is the "Guess Who's Coming To Dinner" scenario where parents are scandalized about their child marrying someone of a different race. Pew has done some good work on this and found that only 23% of conservatives and 1% (!) of liberals admit they would be upset in this situation. But Pew *also* asked how parents would feel about their child marrying someone of a different *political party*. Now 30% of conservatives and 23% of liberals would get upset. Average them out, and you go from 12% upsetness rate for race to 27% upsetness rate for party – more than double. Yeah, people do lie to pollsters, but a picture is starting to come together here.

(Harvard, by the way, is a tossup. There are more black students – 11.5% – than conservative students – 10% – but there are more conservative faculty than black faculty.)

Since people will delight in misinterpreting me here, let me overemphasize what I am *not* saying. I'm not saying people of either party have it "worse" than black people, or that partyism is more of a *problem* than racism, or any of a number of stupid things along those lines which I am sure I will nevertheless be accused of believing. Racism is worse than partyism because the two parties are at least kind of balanced in numbers and in resources, whereas the brunt of an entire country's racism falls on a few underprivileged people. I am saying that the *underlying attitudes that* *produce* partyism are stronger than the underlying attitudes that produce racism, with no necessary implications on their social effects.

But if we want to look at people’s psychology and motivations, partyism and the particular variant of tribalism that it represents are going to be fertile ground.

VII.

Every election cycle like clockwork, conservatives accuse liberals of not being sufficiently pro-America. And every election cycle like clockwork, liberals give extremely unconvincing denials of this.

“It’s not that we’re, like, *against* America per se. It’s just that…well, did you know Europe has much better health care than we do? And much lower crime rates? I mean, come on, how did they get so awesome? And we’re just sitting here, can’t even get the gay marriage thing sorted out, seriously, what’s wrong with a country that can’t…sorry, what were we talking about? Oh yeah, America. They’re okay. Cesar Chavez was really neat. So were some other people outside the mainstream who became famous precisely by criticizing majority society. That’s *sort of* like America being great, in that I think the parts of it that point out how bad the rest of it are often make excellent points. Vote for me!”

(sorry, I make fun of you because I love you)

There was a big brouhaha a couple of years ago when, as it first became apparent Obama had a good shot at the Presidency, Michelle Obama said that “for the first time in my adult life, I am proud of my country.”

Republicans pounced on the comment, asking why she hadn’t felt proud before, and she backtracked saying of course she was proud all the time and she loves America with the burning fury of a million suns and she was just saying that the Obama campaign was *particularly* inspiring.

As unconvincing denials go, this one was pretty far up there. But no one really held it against her. Probably most Obama voters felt vaguely the same way. *I* was an Obama voter, and I have proud memories of spending my Fourth of Julys as a kid debunking people’s heartfelt emotions of patriotism. Aaron Sorkin:

> [What makes America the greatest country in the world?] It’s not the greatest country in the world! We’re seventh in literacy, 27th in math, 22nd in science, 49th in life expectancy, 178th in infant mortality, third in median household income, No. 4 in labor force, and No. 4 in exports. So when you ask what makes us the greatest country in the world, I don’t know what the f*** you’re talking about.

(Another good retort is “We’re number one? Sure – number one in incarceration rates, drone strikes, and making new parents go back to work!”)

All of this is true, of course. But it’s weird that it’s such a classic interest of members of the Blue Tribe, and members of the Red Tribe never seem to bring it up.

(“We’re number one? Sure – number one in levels of sexual degeneracy! Well, I guess probably number two, after the Netherlands, but they’re really small and shouldn’t count.”)

My hunch – both the Red Tribe and the Blue Tribe, for whatever reason, identify “America” with the Red Tribe. Ask people for typically “American” things, and you end up with a very Red list of characteristics – guns, religion, barbecues, American football, NASCAR, cowboys, SUVs, unrestrained capitalism.

That means the Red Tribe feels intensely patriotic about "their" country, and the Blue Tribe feels like they're living in fortified enclaves deep in hostile territory.

Here is a popular piece published on a major media site called America: A Big, Fat, Stupid Nation. Another: America: A Bunch Of Spoiled, Whiny Brats. Americans are ignorant, scientifically illiterate religious fanatics whose "patriotism" is actually just narcissism. You Will Be Shocked At How Ignorant Americans Are, and we should Blame The Childish, Ignorant American People.

Needless to say, every single one of these articles was written by an American and read almost entirely by Americans. Those Americans very likely enjoyed the articles very much and did not feel the least bit insulted.

And look at the sources. HuffPo, Salon, Slate. Might those have anything in common?

On both sides, "American" can be either a normal demonym, or a code word for a member of the Red Tribe.

VIII.

The other day, I logged into OKCupid and found someone who looked cool. I was reading over her profile and found the following sentence:

> Don't message me if you're a sexist white guy

And my first thought was "Wait, so a sexist black person would be okay? Why?"

(The girl in question was white as snow)

Around the time the Ferguson riots were first starting, there were a host of articles with titles like Why White People Don’t Seem To Understand Ferguson, Why It’s So Hard For Whites To Understand Ferguson, and White Folks Listen Up And Let Me Tell You What Ferguson Is All About, this last of which says:

> Social media is full of people on both sides making presumptions, and believing what they want to believe. But it’s the white folks that don’t understand what this is all about. Let me put it as simply as I can for you […]
>
> No matter how wrong you think Trayvon Martin or Michael Brown were, I think we can all agree they didn’t deserve to die over it. I want you white folks to understand that this is where the anger is coming from. You focused on the looting….”

And on a hunch I checked the author photos, and every single one of these articles was written by a white person.

White People Are Ruining America? White. White People Are Still A Disgrace? White. White Guys: We Suck And We’re Sorry? White. Bye Bye, Whiny White Dudes? White. Dear Entitled Straight White Dudes, I’m Evicting You From My Life? White. White Dudes Need To Stop Whitesplaining? White. Reasons Why Americans Suck #1: White People? White.

We’ve all seen articles and comments and articles like this. Some unsavory people try to use them to prove that white people are the *real* victims or the media is biased against white people or something. Other people who are very nice and optimistic use them to show that some white people have developed some self-awareness and are willing to engage in self-criticism.

But I think the situation with “white” is much the same as the situation with “American” – it can either mean what it says, or be a code word for the Red Tribe.

(except on the blog Stuff White People Like, where it obviously serves as a code word for the *Blue* tribe. I don’t know, guys. I didn’t do it.)

I realize that’s making a strong claim, but it would hardly be without precedent. When people say things like “gamers are misogynist”, do they mean the 52% of gamers who are women? Do they mean every one of the 59% of Americans from every walk of life who are known to play video or computer games occasionally? No. “Gamer” is a coded reference to the Gray Tribe, the half-branched-off collection of libertarianish tech-savvy nerds, and everyone knows it. As well expect that when people talk about “fedoras”, they mean Indiana Jones. Or when they talk about “urban youth”, they mean freshmen at NYU. Everyone knows exactly who we mean when we say “urban youth”, and them being young people who live in a city has only the most tenuous of relations to the actual concept.

And I’m saying words like “American” and “white” work the same way. Bill Clinton was the “first black President”, but if Herman Cain had won in 2012 he’d have been the 43rd white president. And when an angry white person talks at great length about how much he hates “white dudes”, *he is not being humble and self-critical*.

IX.

Imagine hearing that a liberal talk show host and comedian was so enraged by the actions of ISIS that he’d recorded and posted a video in which he shouts at them for ten minutes, cursing the “fanatical terrorists” and calling them “utter savages” with “savage values”.

If *I* heard that, I’d be kind of surprised. It doesn’t fit my model of what liberal talk show hosts do.

But [the story]() I’m *actually* referring to is liberal talk show host / comedian Russell Brand making that same rant against Fox News for *supporting war against* the Islamic State, adding at the end that “Fox is worse than ISIS”.

That fits my model perfectly. You wouldn’t celebrate Osama’s death, only Thatcher’s. And you wouldn’t call ISIS savages, only Fox News. Fox is the outgroup, ISIS is just some random people off in a desert. You hate the outgroup, you don’t hate random desert people.

I would go further. Not only does Brand not feel much like hating ISIS, he has a strong incentive not to. That incentive is: the Red Tribe is known to hate ISIS loudly and conspicuously. Hating ISIS would signal Red Tribe membership, would be the equivalent of going into Crips territory with a big Bloods gang sign tattooed on your shoulder.

But this might be unfair. What would Russell Brand answer, if we asked him to justify his decision to be much angrier at Fox than ISIS?

He might say something like “Obviously Fox News is not literally worse than ISIS. But here I am, talking to my audience, who are mostly white British people and Americans. These people already know that ISIS is bad; they don’t need to be told that any further. In fact, at this point being angry about how bad ISIS is, is less likely to genuinely change someone’s mind about ISIS, and more likely to promote Islamophobia. The sort of people in my audience are at zero risk of becoming ISIS supporters, but at a very real risk of Islamophobia. So ranting against ISIS would be counterproductive and dangerous.

On the other hand, my audience of white British people and Americans is very likely to contain many Fox News viewers and supporters. And Fox, while not quite as evil as ISIS, is still pretty bad. So here's somewhere I have a genuine chance to reach people at risk and change minds. Therefore, I think my decision to rant against Fox News, and maybe hyperbolically say they were 'worse than ISIS' is justified under the circumstances."

I have a lot of sympathy to hypothetical-Brand, especially to the part about Islamophobia. It *does* seem really possible to denounce ISIS' atrocities to a population that already hates them in order to weak-man a couple of already-marginalized Muslims. We need to fight terrorism and atrocities – therefore it's okay to shout at a poor girl ten thousand miles from home for wearing a headscarf in public. Christians are being executed for their faith in Sudan, therefore let's picket the people trying to build a mosque next door.

But my sympathy with Brand ends when he acts like his audience is likely to be fans of Fox News.

In a world where a negligible number of Redditors oppose gay marriage and 1% of Less Wrongers identify conservative and I know 0/150 creationists, how many of the people who visit the YouTube channel of a well-known liberal activist with a Che-inspired banner, a channel whose episode names are things like "War: What Is It Good For?" and "Sarah Silverman Talks Feminism" – how many of them do you think are big Fox News fans?

In a way, Russell Brand would have been *braver* taking a stand against ISIS than against Fox. If he attacked ISIS, his viewers would just be a little confused and uncomfortable. Whereas every moment he's attacking Fox his viewers are like "HA HA! YEAH! GET 'EM! SHOW THOSE IGNORANT BIGOTS IN THE OUTGROUP WHO'S BOSS!"

Brand acts as if there are just these countries called "Britain" and "America" who are receiving his material. Wrong. There are two parallel universes, and he's only broadcasting to one of them.

The result is exactly what we predicted would happen in the case of Islam. Bombard people with images of a far-off land they already hate and tell them to hate it more, and the result is ramping up the intolerance on the couple of dazed and marginalized representatives of that culture who have ended up stuck on your half of the divide. Sure enough, if industry or culture or community gets Blue enough, Red Tribe members start getting harassed, fired from their jobs (Brendan Eich being the obvious example) or otherwise shown the door.

Think of Brendan Eich as a member of a tiny religious minority surrounded by people who hate that minority. Suddenly firing him doesn't seem very noble.

If you mix together Podunk, Texas and Mosul, Iraq, you can prove that Muslims are scary and very powerful people who are executing Christians all the time – and so we have a great excuse for kicking the one remaining Muslim family, random people who never hurt anyone, out of town.

And if you mix together the open-source tech industry and the parallel universe where you can't wear a FreeBSD t-shirt without risking someone trying to exorcise you, you can prove that Christians are scary and very powerful people who are persecuting everyone else all the time, and you have a great excuse for kicking one of the few people willing to affiliate with the Red Tribe, a guy who never hurt anyone, out of town.

When a friend of mine heard Eich got fired, she didn't see anything wrong with it. "I can tolerate anything except intolerance," she said.

"Intolerance" is starting to look like another one of those words like "white" and "American".

"I can tolerate anything except the outgroup." Doesn't sound quite so noble now, does it?

X.

We started by asking: millions of people are conspicuously praising every outgroup they can think of, while conspicuously condemning their own in-group. This seems contrary to what we know about social psychology. What's up?

We noted that outgroups are rarely literally "the group most different from you", and in fact far more likely to be groups very similar to you sharing *almost* all your characteristics and living in the same area.

We then noted that although liberals and conservatives live in the same area, they might as well be two totally different countries or universe as far as level of interaction were concerned.

Contra the usual idea of them being marked only by voting behavior, we described them as very different tribes with totally different cultures. You can speak of "American culture” only in the same way you can speak of “Asian culture” – that is, with a lot of interior boundaries being pushed under the rug.

The outgroup of the Red Tribe is occasionally blacks and gays and Muslims, more often the Blue Tribe.

The Blue Tribe has performed some kind of very impressive act of alchemy, and transmuted *all* of its outgroup hatred to the Red Tribe.

This is not surprising. Ethnic differences have proven quite tractable in the face of shared strategic aims. Even the Nazis, not known for their ethnic tolerance, were able to get all buddy-buddy with the Japanese when they had a common cause.

Research suggests Blue Tribe / Red Tribe prejudice to be much stronger than better-known types of prejudice like racism. Once the Blue Tribe was able to enlist the blacks and gays and Muslims in their ranks, they became allies of convenience who deserve to be rehabilitated with mildly condescending paeans to their virtue. “There never was a coward where the shamrock grows.”

Spending your entire life insulting the other tribe and talking about how terrible they are makes you look, well, tribalistic. It is definitely not high class. So when members of the Blue Tribe decide to dedicate their entire life to yelling about how terrible the Red Tribe is, they make sure that instead of saying “the Red Tribe”, they say “America”, or “white people”, or “straight white men”. That way it’s *humble self-criticism*. They are *so* interested in justice that they are willing to critique *their own beloved side*, much as it pains them to do so. We know they are not exaggerating, because one might exaggerate the flaws of an enemy, but that anyone would exaggerate their *own* flaws fails the criterion of embarrassment.

The Blue Tribe always has an excuse at hand to persecute and crush any Red Tribers unfortunate enough to fall into its light-matter-universe by defining them as all-powerful domineering oppressors. They appeal to the fact that this is definitely the way it works in the Red Tribe's dark-matter-universe, and that's in the same country so it has to be the same community for all intents and purposes. As a result, every Blue Tribe institution is permanently licensed to take whatever emergency measures are necessary against the Red Tribe, however disturbing they might otherwise seem.

And so how virtuous, how noble the Blue Tribe! Perfectly tolerant of all of the different groups that just so happen to be allied with them, never intolerant unless it happen to be against intolerance itself. Never stooping to engage in petty tribal conflict like that awful Red Tribe, but always nobly criticizing their own culture and striving to make it better!

Sorry. But I hope this is at least a *little* convincing. The weird dynamic of outgroup-philia and ingroup-phobia isn't anything of the sort. It's just good old-fashioned in-group-favoritism and outgroup bashing, a little more sophisticated and a little more sneaky.

XI.

This essay is bad and I should feel bad.

I should feel bad because I made *exactly* the mistake I am trying to warn everyone else about, and it wasn't until I was almost done that I noticed.

How virtuous, how noble I must be! Never stooping to engage in petty tribal conflict like that silly Red Tribe, but always nobly criticizing my own tribe and striving to make it better.

Yeah. Once I’ve written a ten thousand word essay savagely attacking the Blue Tribe, either I’m a very special person or they’re my outgroup. And I’m not *that* special.

Just as you can pull a fast one and look humbly self-critical if you make your audience assume there’s just one American culture, so maybe you can trick people by assuming there’s only one Blue Tribe.

I’m pretty sure I’m not Red, but I did talk about the Grey Tribe above, and I show all the risk factors for being one of them. That means that, although my critique of the Blue Tribe may be right or wrong, in terms of *motivation* it comes from the same place as a Red Tribe member talking about how much they hate al-Qaeda or a Blue Tribe member talking about how much they hate ignorant bigots. And when I boast of being able to tolerate Christians and Southerners whom the Blue Tribe is mean to, I’m not being tolerant at all, just noticing people so far away from me they wouldn’t make a good outgroup anyway.

I had *fun* writing this article. People do not have fun writing articles savagely criticizing their in-group. People can criticize their in-group, it’s not *humanly impossible*, but it takes nerves of steel, it makes your blood boil, you should sweat blood. It shouldn’t be *fun*.

You can bet some white guy on Gawker who week after week churns out “Why White People Are So Terrible” and “Here’s What Dumb White People Don’t Understand” is having fun and not sweating any blood at all. He’s not criticizing his in-group, he’s never even *considered* criticizing his in-group. I can’t blame him. Criticizing the in-group is a really difficult project I’ve barely begun to build the mental skills necessary to even consider.

I can think of criticisms of my own tribe. Important criticisms, true ones. But the thought of writing them makes my blood boil.

I imagine might I feel like some liberal US Muslim leader, when he goes on the O’Reilly Show, and O’Reilly ambushes him and demands to know why he and other American Muslims haven’t condemned beheadings by ISIS more, demands that he criticize them right there on live TV. And you can see the wheels in the Muslim leader’s head turning, thinking something like “Okay, obviously beheadings are terrible and I hate them as much as anyone. But you don’t care even *the slightest bit* about the victims of beheadings. You’re just looking for a way to score points against me so you can embarass all Muslims. And I would rather personally behead every single person in the world than give a smug bigot like you a single microgram more stupid self-satisfaction than you’ve already got.”

That is how I feel when asked to criticize my own tribe, even for correct reasons. If you think you’re criticizing your own tribe, and your blood is not at that temperature, consider the possibility that you aren’t.

But if I want Self-Criticism Virtue Points, criticizing the Grey Tribe is the only honest way to get them. And if I want Tolerance Points, my own personal cross to bear right now is tolerating the Blue Tribe. I need to remind myself that when they are bad people, they are merely Osama-level bad people instead of Thatcher-level bad people. And when they are good people, they are powerful and necessary crusaders against the evils of the world.

The worst thing that could happen to this post is to have it be used as convenient feces to fling at the Blue Tribe whenever feces are necessary. Which, given what has happened to my last couple of posts along these lines and the obvious biases of my own subconscious, I already expect it will be.

But the best thing that could happen to this post is that it makes a lot of people, especially myself, figure out how to be more tolerant. Not in the “of course I’m tolerant, why shouldn’t I be?” sense of the Emperor in Part I. But in the sense of “being tolerant makes me see red, makes me sweat blood, but darn it *I am going to be tolerant anyway*.”

Meditations On Moloch

I.

Allen Ginsberg’s famous poem on Moloch:

> What sphinx of cement and aluminum bashed open their skulls and ate up their brains and imagination?
>
> Moloch! Solitude! Filth! Ugliness! Ashcans and unobtainable dollars! Children screaming under the stairways! Boys sobbing in armies! Old men weeping in the parks!
>
> Moloch! Moloch! Nightmare of Moloch! Moloch the loveless! Mental Moloch! Moloch the heavy judger of men!
>
> Moloch the incomprehensible prison! Moloch the crossbone soulless jailhouse and Congress of sorrows! Moloch whose buildings are judgment! Moloch the vast stone of war! Moloch the stunned governments!
>
> Moloch whose mind is pure machinery! Moloch whose blood is running money! Moloch whose fingers are ten armies! Moloch whose breast is a cannibal dynamo! Moloch whose ear is a smoking tomb!
>
> Moloch whose eyes are a thousand blind windows! Moloch whose skyscrapers stand in the long streets like endless Jehovahs! Moloch whose factories dream and croak in the fog! Moloch whose smoke-stacks and antennae crown the cities!
>
> Moloch whose love is endless oil and stone! Moloch whose soul is electricity and banks! Moloch whose

poverty is the specter of genius! Moloch whose fate is a cloud of sexless hydrogen! Moloch whose name is the Mind!

Moloch in whom I sit lonely! Moloch in whom I dream Angels! Crazy in Moloch! Cocksucker in Moloch! Lacklove and manless in Moloch!

Moloch who entered my soul early! Moloch in whom I am a consciousness without a body! Moloch who frightened me out of my natural ecstasy! Moloch whom I abandon! Wake up in Moloch! Light streaming out of the sky!

Moloch! Moloch! Robot apartments! invisible suburbs! skeleton treasuries! blind capitals! demonic industries! spectral nations! invincible madhouses! granite cocks! monstrous bombs!

They broke their backs lifting Moloch to Heaven! Pavements, trees, radios, tons! lifting the city to Heaven which exists and is everywhere about us!

Visions! omens! hallucinations! miracles! ecstasies! gone down the American river!

Dreams! adorations! illuminations! religions! the whole boatload of sensitive bullshit!

Breakthroughs! over the river! flips and crucifixions! gone down the flood! Highs! Epiphanies! Despairs! Ten years' animal screams and suicides! Minds! New loves! Mad generation! down on the rocks of Time!

Real holy laughter in the river! They saw it all! the wild eyes! the holy yells! They bade farewell! They jumped

> off the roof! to solitude! waving! carrying flowers! Down to the river! into the street!

What's always impressed me about this poem is its conception of civilization as an individual entity. You can almost see him, with his fingers of armies and his skyscraper-window eyes.

A lot of the commentators say Moloch represents capitalism. This is definitely a piece of it, even a big piece. But it doesn't quite fit. Capitalism, whose fate is a cloud of sexless hydrogen? Capitalism in whom I am a consciousness without a body? Capitalism, therefore granite cocks?

Moloch is introduced as the answer to a question – C. S. Lewis' question in Hierarchy Of Philosophers – *what does it*? Earth could be fair, and all men glad and wise. Instead we have prisons, smokestacks, asylums. What sphinx of cement and aluminum breaks open their skulls and eats up their imagination?

And Ginsberg answers: *Moloch does it.*

There's a passage in the *Principia Discordia* where Malaclypse complains to the Goddess about the evils of human society. "Everyone is hurting each other, the planet is rampant with injustices, whole societies plunder groups of their own people, mothers imprison sons, children perish while brothers war."

The Goddess answers: "What is the matter with that, if it's what you want to do?"

Malaclypse: "But nobody wants it! Everybody hates it!"

Goddess: "Oh. Well, then stop."

The implicit question is – if everyone hates the current system, who perpetuates it? And Ginsberg answers: “Moloch”. It’s powerful not because it’s correct – nobody literally thinks an ancient Carthaginian demon causes everything – but because thinking of the system as an agent throws into relief the degree to which the system *isn’t* an agent.

Bostrom makes an offhanded reference of the possibility of a dictatorless dystopia, one that every single citizen including the leadership hates but which nevertheless endures unconquered. It’s easy enough to imagine such a state. Imagine a country with two rules: first, every person must spend eight hours a day giving themselves strong electric shocks. Second, if anyone fails to follow a rule (including this one), or speaks out against it, or fails to enforce it, all citizens must unite to kill that person. Suppose these rules were well-enough established by tradition that everyone expected them to be enforced.

So you shock yourself for eight hours a day, because you know if you don’t everyone else will kill you, because if they don’t, everyone else will kill *them*, and so on. Every single citizen hates the system, but for lack of a good coordination mechanism it endures. From a god’s-eye-view, we can optimize the system to “everyone agrees to stop doing this at once”, but no one within the system is able to effect the transition without great risk to themselves.

And okay, this example is kind of contrived. So let’s run through – let’s say ten – real world examples of similar multipolar traps to really hammer in how important this is.

1. The Prisoner’s Dilemma, as played by two very dumb libertarians who keep ending up on defect-defect. There’s a much better outcome available if they could figure out the

coordination, but coordination is *hard*. From a god's-eye-view, we can agree that cooperate-cooperate is a better outcome than defect-defect, but neither prisoner within the system can make it happen.

2. Dollar auctions. I wrote about this and even more convoluted versions of the same principle in Game Theory As A Dark Art. Using some weird auction rules, you can take advantage of poor coordination to make someone pay $10 for a one dollar bill. From a god's-eye-view, clearly people should not pay $10 for a on-er. From within the system, each individual step taken might be rational.

(Ashcans and unobtainable dollars!)

3. The fish farming story from my Non-Libertarian FAQ 2.0:

> As a thought experiment, let's consider aquaculture (fish farming) in a lake. Imagine a lake with a thousand identical fish farms owned by a thousand competing companies. Each fish farm earns a profit of $1000/month. For a while, all is well.
>
> But each fish farm produces waste, which fouls the water in the lake. Let's say each fish farm produces enough pollution to lower productivity in the lake by $1/month.
>
> A thousand fish farms produce enough waste to lower productivity by $1000/month, meaning none of the fish farms are making any money. Capitalism to the rescue: someone invents a complex filtering system that removes waste products. It costs $300/month to operate. All fish farms voluntarily install it, the pollution ends, and the fish farms are now making a profit of $700/month – still a respectable sum.

> But one farmer (let’s call him Steve) gets tired of spending the money to operate his filter. Now one fish farm worth of waste is polluting the lake, lowering productivity by $1. Steve earns $999 profit, and everyone else earns $699 profit.
>
> Everyone else sees Steve is much more profitable than they are, because he’s not spending the maintenance costs on his filter. They disconnect their filters too.
>
> Once four hundred people disconnect their filters, Steve is earning $600/month – less than he would be if he and everyone else had kept their filters on! And the poor virtuous filter users are only making $300. Steve goes around to everyone, saying “Wait! We all need to make a voluntary pact to use filters! Otherwise, everyone’s productivity goes down.”
>
> Everyone agrees with him, and they all sign the Filter Pact, except one person who is sort of a jerk. Let’s call him Mike. Now everyone is back using filters again, except Mike. Mike earns $999/month, and everyone else earns $699/month. Slowly, people start thinking they too should be getting big bucks like Mike, and disconnect their filter for $300 extra profit…
>
> A self-interested person never has any incentive to use a filter. A self-interested person has some incentive to sign a pact to make everyone use a filter, but in many cases has a stronger incentive to wait for everyone else to sign such a pact but opt out himself. This can lead to an undesirable equilibrium in which no one will sign such a pact.

The more I think about it, the more I feel like this is the core of my objection to libertarianism, and that Non-Libertarian

FAQ 3.0 will just be this one example copy-pasted two hundred times. From a god's-eye-view, we can say that polluting the lake leads to bad consequences. From within the system, no individual can prevent the lake from being polluted, and buying a filter might not be such a good idea.

4. The Malthusian trap, at least at its extremely pure theoretical limits. Suppose you are one of the first rats introduced onto a pristine island. It is full of yummy plants and you live an idyllic life lounging about, eating, and composing great works of art (you're one of those rats from *The Rats of NIMH*).

You live a long life, mate, and have a dozen children. All of them have a dozen children, and so on. In a couple generations, the island has ten thousand rats and has reached its carrying capacity. Now there's not enough food and space to go around, and a certain percent of each new generation dies in order to keep the population steady at ten thousand.

A certain sect of rats abandons art in order to devote more of their time to scrounging for survival. Each generation, a bit less of this sect dies than members of the mainstream, until after a while, no rat composes any art at all, and any sect of rats who try to bring it back will go extinct within a few generations.

In fact, it's not just art. Any sect at all that is leaner, meaner, and more survivalist than the mainstream will eventually take over. If one sect of rats altruistically decides to limit its offspring to two per couple in order to decrease overpopulation, that sect will die out, swarmed out of existence by its more numerous enemies. If one sect of rats starts practicing cannibalism, and finds it gives them an advantage over their fellows, it will eventually take over and reach fixation.

If some rat scientists predict that depletion of the island's nut stores is accelerating at a dangerous rate and they will soon be exhausted completely, a few sects of rats might try to limit their nut consumption to a sustainable level. Those rats will be outcompeted by their more selfish cousins. Eventually the nuts will be exhausted, most of the rats will die off, and the cycle will begin again. Any sect of rats advocating some action to stop the cycle will be outcompeted by their cousins for whom advocating *anything* is a waste of time that could be used to compete and consume.

For a bunch of reasons evolution is not quite as Malthusian as the ideal case, but it provides the prototype example we can apply to other things to see the underlying mechanism. From a god's-eye-view, it's easy to say the rats should maintain a comfortably low population. From within the system, each individual rat will follow its genetic imperative and the island will end up in an endless boom-bust cycle.

5. Capitalism. Imagine a capitalist in a cutthroat industry. He employs workers in a sweatshop to sew garments, which he sells at minimal profit. Maybe he would like to pay his workers more, or give them nicer working conditions. But he can't, because that would raise the price of his products and he would be outcompeted by his cheaper rivals and go bankrupt. Maybe many of his rivals are nice people who would like to pay their workers more, but unless they have some kind of ironclad guarantee that none of them are going to defect by undercutting their prices they can't do it.

Like the rats, who gradually lose all values except sheer competition, so companies in an economic environment of *sufficiently intense competition* are forced to abandon all values except optimizing-for-profit or else be outcompeted by companies that optimized for profit better and so can sell the same service at a lower price.

(I'm not really sure how widely people appreciate the value of analogizing capitalism to evolution. Fit companies – defined as those that make the customer want to buy from them – survive, expand, and inspire future efforts, and unfit companies – defined as those no one wants to buy from – go bankrupt and die out along with their company DNA. The reasons Nature is red and tooth and claw are the same reasons the market is ruthless and exploitative)

From a god's-eye-view, we can contrive a friendly industry where every company pays its workers a living wage. From within the system, there's no way to enact it.

(Moloch whose love is endless oil and stone! Moloch whose blood is running money!)

6. The Two-Income Trap, as recently discussed on this blog. It theorized that sufficiently intense competition for suburban houses in good school districts meant that people had to throw away lots of other values – time at home with their children, financial security – to optimize for house-buying-ability or else be consigned to the ghetto.

From a god's-eye-view, if everyone agrees not to take on a second job to help win their competition for nice houses, then everyone will get exactly as nice a house as they did before, but only have to work one job. From within the system, absent a government literally willing to ban second jobs, everyone who doesn't get one will be left behind.

(Robot apartments! Invisible suburbs!)

7. Agriculture. Jared Diamond calls it the worst mistake in human history. Whether or not it was a mistake, it wasn't an *accident* – agricultural civilizations simply outcompeted nomadic ones, inevitable and irresistably. Classic Malthusian trap. Maybe hunting-gathering was more enjoyable, higher life expectancy, and more conducive to human flourishing – but in a state of *sufficiently intense competition* between peoples, in which agriculture with all its disease and oppression and pestilence was the more competitive option, everyone will end up agriculturalists or go the way of the Comanche Indians.

From a god's-eye-view, it's easy to see everyone should keep the more enjoyable option and stay hunter-gatherers. From within the system, each individual tribe only faces the choice of going agricultural or inevitably dying.

8. Arms races. Large countries can spend anywhere from 5% to 30% of their budget on defense. In the absence of war – a condition which has mostly held for the past fifty years – all this does is sap money away from infrastructure, health, education, or economic growth. But any country that fails to spend enough money on defense risks being invaded by a neighboring country that did. Therefore, almost all countries try to spend some money on defense.

From a god's-eye-view, the best solution is world peace and no country having an army at all. From within the system, no country can unilaterally enforce that, so their best option is to keep on throwing their money into missiles that lie in silos unused.

(Moloch the vast stone of war! Moloch whose fingers are ten armies!)

9. Cancer. The human body is supposed to be made up of cells living harmoniously and pooling their resources for the greater good of the organism. If a cell defects from this equilibrium by investing its resources into copying itself, it and its descendants will flourish, eventually outcompeting all the other cells and taking over the body – at which point it dies. Or the situation may repeat, with certain cancer cells defecting against the rest of the tumor, thus slowing down its growth and causing the tumor to stagnate.

From a god’s-eye-view, the best solution is all cells cooperating so that they don’t all die. From within the system, cancerous cells will proliferate and outcompete the other – so that only the existence of the immune system keeps the natural incentive to turn cancerous in check.

10. The “race to the bottom” describes a political situation where some jurisdictions lure businesses by promising lower taxes and fewer regulations. The end result is that either everyone optimizes for competitiveness – by having minimal tax rates and regulations – or they lose all of their business, revenue, and jobs to people who did (at which point they are pushed out and replaced by a government who will be more compliant).

But even though the last one has stolen the name, all these scenarios are in fact a race to the bottom. Once one agent learns how to become more competitive by sacrificing a common value, all its competitors must also sacrifice that value or be outcompeted and replaced by the less scrupulous. Therefore, the system is likely to end up with everyone once again equally competitive, but the sacrificed value is gone forever. From a god’s-eye-view, the competitors know they will all be worse off if they defect, but from within the system, given insufficient coordination it’s impossible to avoid.

Before we go on, there’s a slightly different form of multi-agent trap worth investigating. In this one, the competition is kept at bay by some outside force – usually social stigma. As a result, there’s not actually a race to the bottom – the system can continue functioning at a relatively high level – but it’s impossible to optimize and resources are consistently thrown away for no reason. Lest you get exhausted before we even begin, I’ll limit myself to four examples here.

11. Education. In my essay on reactionary philosophy, I talk about my frustration with education reform:

> People ask why we can’t reform the education system. But right now students’ incentive is to go to the most prestigious college they can get into so employers will hire them – whether or not they learn anything. Employers’ incentive is to get students from the most prestigious college they can so that they can defend their decision to their boss if it goes wrong – whether or not the college provides value added. And colleges’ incentive is to do whatever it takes to get more prestige, as measured in *US News and World Report* rankings – whether or not it helps students. Does this lead to huge waste and poor education? Yes. Could the Education God notice this and make some Education Decrees that lead to a vastly more efficient system? Easily! But since there’s no Education God everybody is just going to follow their own incentives, which are only partly correlated with education or efficiency.

From a god’s eye view, it’s easy to say things like “Students should only go to college if they think they will get something out of it, and employers should hire applicants based on their competence and not on what college they went to”. From within the system, everyone’s already following their own incentives correctly, so unless the incentives change the system won't either.

12. Science. Same essay:

> The modern research community *knows* they aren't producing the best science they could be. There's lots of publication bias, statistics are done in a confusing and misleading way out of sheer inertia, and replications often happen very late or not at all. And sometimes someone will say something like "I can't believe people are too dumb to fix Science. All we would have to do is require early registration of studies to avoid publication bias, turn this new and powerful statistical technique into the new standard, and accord higher status to scientists who do replication experiments. It would be really simple and it would vastly increase scientific progress. I must just be smarter than all existing scientists, since I'm able to think of this and they aren't."
>
> And yeah. That would work for the Science God. He could just make a Science Decree that everyone has to use the right statistics, and make another Science Decree that everyone must accord replications higher status.
>
> But things that work from a god's-eye view don't work from within the system. No individual scientist has an incentive to unilaterally switch to the new statistical technique for her own research, since it would make her research less likely to produce earth-shattering results and since it would just confuse all the other scientists. They just have an incentive to want everybody else to do it, at which point they would follow along. And no individual journal has an incentive to unilaterally switch

> to early registration and publishing negative results, since it would just mean their results are less interesting than that other journal who only publishes ground-breaking discoveries. From within the system, everyone is following their own incentives and will continue to do so.

13. Government corruption. I don't know of anyone who really thinks, in a principled way, that corporate welfare is a good idea. But the government still manages to spend somewhere around (depending on how you calculate it) $100 billion dollars a year on it – which for example is three times the amount they spend on health care for the needy. Everyone familiar with the problem has come up with the same easy solution: stop giving so much corporate welfare. Why doesn't it happen?

Government are competing against one another to get elected or promoted. And suppose part of optimizing for electability is optimizing campaign donations from corporations – or maybe it isn't, but officials *think* it is. Officials who try to mess with corporate welfare may lose the support of corporations and be outcompeted by officials who promise to keep it intact.

So although from a god's-eye-view everyone knows that eliminating corporate welfare is the best solution, each individual official's personal incentives push her to maintain it.

14. Congress. Only 9% of Americans like it, suggesting a lower approval rating than cockroaches, head lice, or traffic jams. However, 62% of people who know who their own Congressional representative is approve of them. In theory, it should be *really hard* to have a democratically elected body that maintains a 9% approval rating for more than one

election cycle. In practice, every representative’s incentive is to appeal to his or her constituency while throwing the rest of the country under the bus – something at which they apparently succeed.

From a god’s-eye-view, every Congressperson ought to think only of the good of the nation. From within the system, you do what gets you elected.

II.

A basic principle unites all of the multipolar traps above. In some competition optimizing for X, the opportunity arises to throw some other value under the bus for improved X. Those who take it prosper. Those who don’t take it die out. Eventually, everyone’s relative status is about the same as before, but everyone’s absolute status is worse than before. The process continues until all other values that can be traded off have been – in other words, until human ingenuity cannot possibly figure out a way to make things any worse.

In a sufficiently intense competition (1-10), everyone who doesn’t throw all their values under the bus dies out – think of the poor rats who wouldn’t stop making art. This is the infamous Malthusian trap, where everyone is reduced to “subsistence”.

In an insufficiently intense competition (11-14), all we see is a perverse failure to optimize – consider the journals which can’t switch to more reliable science, or the legislators who can’t get their act together and eliminate corporate welfare. It may not reduce people to subsistence, but there is a weird sense in which it takes away their free will.

Every two-bit author and philosopher has to write their own utopia. Most of them are legitimately pretty nice. In fact, it’s a pretty good bet that two utopias that are polar opposites both sound better than our own world.

It’s kind of embarrassing that random nobodies can think up states of affairs better than the one we actually live in. And in fact most of them can’t. A lot of utopias sweep the hard problems under the rug, or would fall apart in ten minutes if actually implemented.

But let me suggest a couple of “utopias” that don’t have this problem.

– The utopia where instead of the government paying lots of corporate welfare, the government *doesn’t* pay lots of corporate welfare.

– The utopia where every country’s military is 50% smaller than it is today, and the savings go into infrastructure spending.

– The utopia where all hospitals use the same electronic medical record system, or at least medical record systems that can talk to each other, so that doctors can look up what the doctor you saw last week in a different hospital decided instead of running all the same tests over again for $5000.

I don’t think there are too many people who *oppose* any of these utopias. If they’re not happening, it’s not because people don’t support them. It certainly isn’t because nobody’s thought of them, since I just thought of them right now and I don’t expect my “discovery” to be hailed as particularly novel or change the world.

Any human with above room temperature IQ can design a utopia. The reason our current system isn’t a utopia is that *it wasn’t designed by humans*. Just as you can look at an arid terrain and determine what shape a river will one day take by assuming water will obey gravity, so you can look at a civilization and determine what shape its institutions will one day take by assuming people will obey incentives.

But that means that just as the shapes of rivers are not designed for beauty or navigation, but rather an artifact of randomly determined terrain, so institutions will not be designed for prosperity or justice, but rather an artifact of randomly determined initial conditions.

Just as people can level terrain and build canals, so people can alter the incentive landscape in order to build better institutions. But they can only do so when they are incentivized to do so, which is not always. As a result, some pretty wild tributaries and rapids form in some very strange places.

I will now jump from boring game theory stuff to what might be the closest thing to a mystical experience I’ve ever had.

Like all good mystical experiences, it happened in Vegas. I was standing on top of one of their many tall buildings, looking down at the city below, all lit up in the dark. If you’ve never been to Vegas, it is *really* impressive. Skyscrapers and lights in every variety strange and beautiful all clustered together. And I had two thoughts, crystal clear:

It is glorious that we can create something like this.

It is shameful that we *did*.

Like, by what standard is building gigantic forty-story-high indoor replicas of Venice, Paris, Rome, Egypt, and Camelot side-by-side, filled with albino tigers, in the middle of the most inhospitable desert in North America, a remotely sane use of our civilization’s limited resources?

And it occurred to me that maybe there is no philosophy on Earth that would endorse the existence of Las Vegas. Even Objectivism, which is usually my go-to philosophy for justifying the excesses of capitalism, at least grounds it in the belief that capitalism improves people's lives. Henry Ford was virtuous because he allowed lots of otherwise carless people to obtain cars and so made them better off. What does Vegas do? Promise a bunch of shmucks free money and not give it to them.

Las Vegas doesn't exist because of some decision to hedonically optimize civilization, it exists because of a quirk in dopaminergic reward circuits, plus the microstructure of an uneven regulatory environment, plus Schelling points. A rational central planner with a god's-eye-view, contemplating these facts, might have thought "Hm, dopaminergic reward circuits have a quirk where certain tasks with slightly negative risk-benefit ratios get an emotional valence associated with slightly positive risk-benefit ratios, let's see if we can educate people to beware of that." People within the system, *following the incentives created by these facts*, think: "Let's build a forty-story-high indoor replica of ancient Rome full of albino tigers in the middle of the desert, and so become slightly richer than people who didn't!"

Just as the course of a river is latent in a terrain even before the first rain falls on it – so the existence of Caesar's Palace was latent in neurobiology, economics, and regulatory regimes even before it existed. The entrepreneur who built it was just filling in the ghostly lines with real concrete.

So we have all this amazing technological and cognitive energy, the brilliance of the human species, wasted on reciting the lines written by poorly evolved cellular receptors and blind economics, like gods being ordered around by a moron.

Some people have mystical experiences and see God. There in Las Vegas, I *saw* Moloch.

(Moloch, whose mind is pure machinery! Moloch, whose blood is running money!

Moloch whose soul is electricity and banks! Moloch, whose skyscrapers stand in the long streets like endless Jehovahs!

Moloch! Moloch! Robot apartments! Invisible suburbs! Skeleton treasuries! Blind capitals! Demonic industries! Spectral nations!)

LUXOR
LUXOR
LAX

…granite cocks!

III.

The Apocrypha Discordia says:

> Time flows like a river. Which is to say, downhill. We can tell this because everything is going downhill rapidly. It would seem prudent to be somewhere else when we reach the sea.

Let’s take this random gag 100% literally and see where it leads us.

We just analogized the flow of incentives to the flow of a river. The downhill trajectory is appropriate: the traps happen when you find an opportunity to trade off a useful value for greater competitiveness. Once everyone has it, the greater competitiveness brings you no joy – but the value is lost forever. Therefore, each step of the Poor Coordination Polka makes your life worse.

But not only have we not yet reached the sea, but we also seem to move *uphill* surprisingly often. Why do things not degenerate more and more until we are back at subsistence level? I can think of three bad reasons – excess resources, physical limitations, and utility maximization – plus one good reason – coordination.

1. Excess resources. The ocean depths are a horrible place with little light, few resources, and various horrible organisms dedicated to eating or parasitizing one another. But every so often, a whale carcass falls to the bottom of the sea. More food than the organisms that find it could ever possibly want. There’s a brief period of miraculous plenty, while the couple of creatures that first encounter the whale feed like kings. Eventually more animals discover the carcass, the faster-breeding animals in the carcass multiply, the whale is gradually consumed, and everyone sighs and goes back to living in a Malthusian death-trap.

(Slate Star Codex: Your source for macabre whale metaphors since June 2014)

It’s as if a group of those rats who had abandoned art and turned to cannibalism suddenly was blown away to a new empty island with a much higher carrying capacity, where they would once again have the breathing room to live in peace and create artistic masterpieces.

This is an age of whalefall, an age of excess carrying capacity, an age when we suddenly find ourselves with a thousand-mile head start on Malthus. As Hanson puts it, this is the dream time.

As long as resources aren’t scarce enough to lock us in a war of all against all, we can do silly non-optimal things – like art and music and philosophy and love – and not be outcompeted by merciless killing machines most of the time.

2. Physical limitations. Imagine a profit-maximizing slavemaster who decided to cut costs by not feeding his slaves or letting them sleep. He would soon find that his slaves’ productivity dropped off drastically, and that no amount of whipping them could restore it. Eventually after testing numerous strategies, he might find his slaves got the most work done when they were well-fed and well-rested and had at least a little bit of time to relax. Not because the slaves were voluntarily withholding their labor – we assume the fear of punishment is enough to make them work as hard as they can – but because the body has certain physical limitations that limit how mean you can get away with being. Thus, the “race to the bottom” stops somewhere short of the actual ethical bottom, when the physical limits are run into.

John Moes, a historian of slavery, goes further and writes about how the slavery we are most familiar with – that of the antebellum South – is a historical aberration and probably economically inefficient. In most past forms of slavery – especially those of the ancient world – it was common for slaves to be paid wages, treated well, and often given their freedom.

He argues that this was the result of rational economic calculation. You can incentivize slaves through the carrot or the stick, and the stick isn't very good. You can't watch slaves all the time, and it's really hard to tell whether a slave is slacking off or not (or even whether, given a little more whipping, he might be able to work even harder). If you want your slaves to do anything more complicated than pick cotton, you run into some serious monitoring problems – how do you profit from an enslaved philosopher? Whip him really hard until he elucidates a theory of The Good that you can sell books about?

The ancient solution to the problem – perhaps an early inspiration to Fnargl – was to tell the slave to go do whatever he wanted and found most profitable, then split the profits with him. Sometimes the slave would work a job at your workshop and you would pay him wages based on how well he did. Other times the slave would go off and make his way in the world and send you some of what he earned. Still other times, you would set a price for the slave's freedom, and the slave would go and work and eventually come up with the mone and free himself.

Moes goes even further and says that these systems were so profitable that there were constant smouldering attempts to try this sort of thing in the American South. The reason they stuck with the whips-and-chains method owed less to economic considerations and more to racist government officials cracking down on lucrative but not-exactly-white-supremacy-promoting attempts to free slaves and have them go into business.

So in this case, a race to the bottom where competing plantations become crueler and crueler to their slaves in order to maximize competitiveness is halted by the physical limitation of cruelty not helping after a certain point.

Or to give another example, one of the reasons we're not currently in a Malthusian population explosion right now is that women can only have one baby per nine months. If those weird religious sects that demand their members have as many babies as possible could copy-paste themselves, we would be in *really* bad shape. As it is they can only do a small amount of damage per generation.

3. Utility maximization. We've been thinking in terms of preserving values versus winning competitions, and expecting optimizing for the latter to destroy the former.

But many of the most important competitions / optimization processes in modern civilization are optimizing for human values. You win at capitalism partly by satisfying customers' values. You win at democracy partly by satisfying voters' values.

Suppose there's a coffee plantation somewhere in Ethiopia that employs Ethiopians to grow coffee beans that get sold to the United States. Maybe it's locked in a life-and-death struggle with other coffee plantations and want to throw as many values under the bus as it can to pick up a slight advantage.

But it can't sacrifice quality of coffee produced too much, or else the Americans won't buy it. And it can't sacrifice wages or working conditions too much, or else the Ethiopians won't work there. And in fact, part of its competition-optimization process is finding the best ways to attract workers and customers that it can, as long as it doesn't cost them too much money. So this is very promising.

But it's important to remember exactly how fragile this beneficial equilibrium is.

Suppose the coffee plantations discover a toxic pesticide that will increase their yield but make their customers sick. But their customers don't know about the pesticide, and the government hasn't caught up to regulating it yet. Now there's a tiny uncoupling between "selling to Americans" and "satisfying Americans' values", and so of course Americans' values get thrown under the bus.

Or suppose that there's a baby boom in Ethiopia and suddenly there are five workers competing for each job. Now the company can afford to lower wages and implement cruel working conditions down to whatever the physical limits are. As soon as there's an uncoupling between "getting Ethiopians to work here" and "satisfying Ethiopian values", it doesn't look too good for Ethiopian values either.

Or suppose someone invents a robot that can pick coffee better and cheaper than a human. The company fires all its laborers and throws them onto the street to die. As soon as the utility of the Ethiopians is no longer necessary for profit, all pressure to maintain it disappears.

Or suppose that there is some important value that is neither a value of the employees or the customers. Maybe the coffee plantations are on the habitat of a rare tropical bird that environmentalist groups want to protect. Maybe they're on the ancestral burial ground of a tribe different from the one the plantation is employing, and they want it respected in some way. Maybe coffee growing contributes to global warming somehow. As long as it’s not a value that will prevent the average American from buying from them or the average Ethiopian from working for them, under the bus it goes.

I know that “capitalists sometimes do bad things” isn’t exactly an original talking point. But I do want to stress how it’s not equivalent to “capitalists are greedy”. I mean, sometimes they *are* greedy. But other times they’re just in a sufficiently intense competition where anyone who doesn’t do it will be outcompeted and replaced by people who do. Business practices are set by Moloch, no one else has any choice in the matter.

(from my very little knowledge of Marx, he understands this very very well and people who summarize him as “capitalists are greedy” are doing him a disservice)

And as well understood as the capitalist example is, I think it is less well appreciated that democracy has the same problems. Yes, in theory it’s optimizing for voter happiness which correlates with good policymaking. But as soon as there’s the slightest disconnect between good policymaking and electability, good policymaking *has to* get thrown under the bus.

For example, ever-increasing prison terms are unfair to inmates and unfair to the society that has to pay for them. Politicans are unwilling to do anything about them because they don’t want to look “soft on crime”, and if a single inmate whom they helped release ever does anything bad (and statistically one of them will have to) it will be all over the airwaves as “Convict released by Congressman’s policies kills family of five, how can the Congressman even sleep at night let alone claim he deserves reelection?”. So even if decreasing prison populations would be good policy – and it is – it will be very difficult to implement.

(Moloch the incomprehensible prison! Moloch the crossbone soulless jailhouse and Congress of sorrows! Moloch whose buildings are judgment! Moloch the stunned governments!)

Turning "satisfying customers" and "satisfying citizens" into the *outputs* of optimization processes was one of civilization's greatest advances and the reason why capitalist democracies have so outperformed other systems. But if we have bound Moloch as our servant, the bonds are not very strong, and we sometimes find that the tasks he has done for us move to his advantage rather than ours.

4. Coordination.

The opposite of a trap is a garden.

Things are easy to solve from a god's-eye-view, so if everyone comes together into a superorganism, that superorganism can solve problems with ease and finesse. An intense competition between agents has turned into a garden, with a single gardener dictating where everything should go and removing elements that do not conform to the pattern.

As I pointed out in the Non-Libertarian FAQ, government can easily solve the pollution problem with fish farms. The best known solution to the Prisoners' Dilemma is for the mob boss (playing the role of a governor) to threaten to shoot any prisoner who defects. The solution to companies polluting and harming workers is government regulations against such. Governments solve arm races *within* a country by maintaining a monopoly on the use of force, and it's easy to see that if a truly effective world government ever arose, international military buildups would end pretty quickly.

The two active ingredients of government are laws plus violence – or more abstractly agreements plus enforcement mechanism. Many other things besides governments share these two active ingredients and so are able to act as coordination mechanisms to avoid traps.

For example, since students are competing against each other (directly if classes are graded on a curve, but always indirectly for college admissions, jobs, et cetera) there is intense pressure for individual students to cheat. The teacher and school play the role of a government by having rules (for example, against cheating) and the ability to punish students who break them.

But the emergent social structure of the students themselves is also a sort of government. If students shun and distrust cheaters, then there are rules (don't cheat) and an enforcement mechanism (or else we will shun you).

Social codes, gentlemens' agreements, industrial guilds, criminal organizations, traditions, friendships, schools, corporations, and religions are all coordinating institutions that keep us out of traps by changing our incentives.

But these institutions not only incentivize others, but are incentivized themselves. These are large organizations made of lots of people who are competing for jobs, status, prestige, et cetera – there's no reason they should be immune to the same multipolar traps as everyone else, and indeed they aren't. Governments can in theory keep corporations, citizens, et cetera out of certain traps, but as we saw above there are many traps that governments themselves can fall into.

The United States tries to solve the problem by having multiple levels of government, unbreakable constutitional laws, checks and balances between different branches, and a couple of other hacks.

Saudi Arabia uses a different tactic. They just put one guy in charge of everything.

This is the much-maligned – I think unfairly – argument in favor of monarchy. A monarch is an unincentivized incentivizer. He *actually* has the god's-eye-view and is outside of and above every system. He has permanently won all competitions and is not competing for anything, and therefore he is perfectly free of Moloch and of the incentives that would otherwise channel his incentives into predetermined paths. Aside from a few very theoretical proposals like my Shining Garden, monarchy is the *only* system that does this.

But then instead of following a random incentive structure, we're following the whim of one guy. Caesar's Palace Hotel and Casino is a crazy waste of resources, but the actual Gaius Julius Caesar Augustus Germanicus wasn't exactly the perfect benevolent rational central planner either.

The libertarian-authoritarian axis on the Political Compass is a tradeoff between discoordination and tyranny. You can have everything perfectly coordinated by someone with a god's-eye-view – but then you risk Stalin. And you can be totally free of all central authority – but then you're stuck in every stupid multipolar trap Moloch can devise.

The libertarians make a convincing argument for the one side, and the monarchists for the other, but I expect that like most tradeoffs we just have to hold our noses and admit it's a really hard problem.

IV.

Let’s go back to that Apocrypha Discordia quote:

> Time flows like a river. Which is to say, downhill. We can tell this because everything is going downhill rapidly. It would seem prudent to be somewhere else when we reach the sea.

What would it mean, in this situation, to reach the sea?

Multipolar traps – races to the bottom – threaten to destroy all human values. They are currently restrained by physical limitations, excess resources, utility maximization, and coordination.

The dimension along which this metaphorical river flows must be time, and the most important change in human civilization over time is the change in technology. So the relevant question is how technological changes will affect our tendency to fall into multipolar traps.

I described traps as when:

> …in some competition optimizing for X, the opportunity arises to throw some other value under the bus for improved X. Those who take it prosper. Those who don’t take it die out. Eventually, everyone’s relative status is about the same as before, but everyone’s absolute status is worse than before. The process continues until all other values that can be traded off have been – in other words, until human ingenuity cannot possibly figure out a way to make things any worse.

That “the opportunity arises” phrase is looking pretty sinister. Technology is all about creating new opportunities.

Develop a new robot, and suddenly coffee plantations have “the opportunity” to automate their harvest and fire all the

Ethiopian workers. Develop nuclear weapons, and suddenly countries are stuck in an arms race to have enough of them. Polluting the atmosphere to build products quicker wasn't a problem before they invented the steam engine.

The limit of multipolar traps as technology approaches infinity is "very bad".

Multipolar traps are currently restrained by physical limitations, excess resources, utility maximization, and coordination.

Physical limitations are most obviously conquered by increasing technology. The slavemaster's old conundrum – that slaves need to eat and sleep – succumbs to Soylent and modafinil. The problem of slaves running away succumbs to GPS. The problem of slaves being too stressed to do good work succumbs to Valium. None of these things are very good for the slaves.

(or just invent a robot that doesn't need food or sleep at all. What happens to the slaves after that is better left unsaid)

The other example of physical limits was one baby per nine months, and this was understating the case – it's really "one baby per nine months plus willingness to support and take care of a basically helpless and extremely demanding human being for eighteen years". This puts a damper on the enthusiasm of even the most zealous religious sect's "go forth and multiply" dictum.

But as Bostrom puts it in *Superintelligence*:

> There are reasons, if we take a longer view and assume a state of unchanging technology and continued prosperity, to expect a return to the historically and ecologically normal condition of a world population that

> butts up against the limits of what our niche can support. If this seems counterintuitive in light of the negative relationship between wealth and fertility that we are currently observing on the global scale, we must remind ourselves that this modern age is a brief slice of history and very much an aberration. Human behavior has not yet adapted to contemporary conditions. Not only do we fail to take advantage of obvious ways to increase our inclusive fitness (such as by becoming sperm or egg donors) but we actively sabotage our fertility by using birth control. In the environment of evolutionary adaptedness, a healthy sex drive may have been enough to make an individual act in ways that maximized her reproductive potential; in the modern environment, however, there would be a huge selective advantage to having a more direct desire for being the biological parent to the largest possible number of chilren. Such a desire is currently being selected for, as are other traits that increase our propensity to reproduce. Cultural adaptation, however, might steal a march on biological evolution. Some communities, such as those of the Hutterites or the adherents of the Quiverfull evangelical movement, have natalist cultures that encourage large families, and they are consequently undergoing rapid expansion…This longer-term outlook could be telescoped into a more imminent prospect by the intelligence explosion. Since software is copyable, a population of emulations or AIs could double rapidly – over the course of minutes rather than decades or centuries – soon exhausting all available hardware

As always when dealing with high-level transhumanists, “all available hardware” should be taken to include “the atoms that used to be part of your body”.

The idea of biological *or* cultural evolution causing a mass population explosion is a philosophical toy at best. The idea of technology making it possible is both plausible and terrifying. Now we see that "physical limits" segues very naturally into "excess resources" – the ability to create new agents very quickly means that unless everyone can coordinate to ban doing this, the people who do will outcompete the people who don't until they have reached carrying capacity and everyone is stuck at subsistence level.

Excess resources, which until now have been a gift of technological progress, therefore switch and become a casualty of it at a sufficiently high tech level.

Utility maximization, always on shaky ground, also faces new threats. In the face of continuing debate about this point, I *continue* to think it obvious that robots will push humans out of work or at least drive down wages (which, in the existence of a minimum wage, pushes humans out of work).

Once a robot can do everything an IQ 80 human can do, only better and cheaper, there will be no reason to employ IQ 80 humans. Once a robot can do everything an IQ 120 human can do, only better and cheaper, there will be no reason to employ IQ 120 humans. Once a robot can do everything an IQ 180 human can do, only better and cheaper, there will be no reason to employ humans at all, in the unlikely scenario that there are any left by that point.

In the earlier stages of the process, capitalism becomes more and more uncoupled from its previous job as an optimizer for human values. Now most humans are totally locked out of the group whose values capitalism optimizes for. They have no value to contribute as workers – and since in the absence of a spectacular social safety net it's unclear how they would have much money – they have no value as customers either. Capitalism has passed them by. As the segment of humans who can be outcompeted by robots increases, capitalism passes by more and more people until eventually it locks out the human race entirely, once again in the vanishingly unlikely scenario that we are still around.

(there are some scenarios in which a few capitalists who own the robots may benefit here, but in either case the vast majority are out of luck)

Democracy is less obviously vulnerable, but it might be worth going back to Bostrom's paragraph about the Quiverfull movement. These are some really religious Christians who think that God wants them to have as many kids as possible, and who can end up with families of ten or more. Their articles explictly calculate that if they start at two percent of the population, but have on average eight children per generation when everyone else on average only has two, within three generations they'll make up half the population.

It's a clever strategy, but I can think of one thing that will save us: judging by how many ex-Quiverfull blogs I found when searching for those statistics, their retention rates even within a single generation are pretty grim. Their article admits that 80% of very religious children leave the church as adults (although of course they expect their own movement to do better). And this is not a symmetrical process – 80% of children who grow up in atheist families aren't becoming Quiverfull.

It looks a lot like even though they are outbreeding us, we are outmeme-ing them, and that gives us a decisive advantage.

But we should also be kind of scared of this process. Memes optimize for making people want to accept them and pass them on – so like capitalism and democracy, they’re optimizing for a *proxy* of making us happy, but that proxy can easily get uncoupled from the original goal.

Chain letters, urban legends, propaganda, and viral marketing are all examples of memes that don’t satisfy our explicit values (true and useful) but are sufficiently memetically virulent that they spread anyway.

I hope it’s not too controversial here to say the same thing is true of religion. Religions, at their heart, are the most basic form of memetic replicator – “Believe this statement and repeat it to everyone you hear or else you will be eternally tortured”.

The creationism “debate” and global warming “debate” and a host of similar “debates” in today’s society suggest that memes that can propagate independent of their truth value has a pretty strong influence on the political process. Maybe these memes propagate because they appeal to people’s prejudices, maybe because they’re simple, maybe because they effectively mark an in-group and an out-group, or maybe for all sorts of different reasons.

The point is – imagine a country full of bioweapon labs, where people toil day and night to invent new infectious agents. The existence of these labs, and their right to throw whatever they develop in the water supply is protected by law. And the country is also linked by the world’s most perfect mass transit system that every single person uses every day, so that any new pathogen can spread to the entire country instantaneously. You’d expect things to start going bad for that city pretty quickly.

Well, we have about a zillion think tanks researching new and better forms of propaganda. And we have constitutionally protected freedom of speech. And we have the Internet. So we’re kind of screwed.

(Moloch whose name is the Mind!)

There are a few people working on raising the sanity waterline, but not as many people as are working on new and exciting ways of confusing and converting people, cataloging and exploiting every single bias and heuristic and dirty rhetorical trick

So as technology (which I take to include knowledge of psychology, sociology, public relations, etc) tends to infinity, the power of truthiness relative to truth increases, and things don’t look great for real grassroots democracy. The worst-case scenario is that the ruling party learns to produce infinite charisma on demand. If that doesn’t sound so bad to you, remember what Hitler was able to do with an famously high level of charisma that was still less-than-infinite.

(alternate phrasing for Chomskyites: technology increases the efficiency of manufacturing consent in the same way it increases the efficiency of manufacturing everything else)

Coordination is what’s left. And technology has the potential to seriously *improve* coordination efforts. People can use the Internet to get in touch with one another, launch political movements, and fracture off into subcommunities.

But coordination only works when you have 51% or more of the force on the side of the people doing the coordinating, and when you haven’t come up with some brilliant trick to make coordination impossible.

The second one first. In the links post before last, I wrote:

> The latest development in the brave new post-Bitcoin world is crypto-equity. At this point I've gone from wanting to praise these inventors as bold libertarian heroes to wanting to drag them in front of a blackboard and making them write a hundred times "I WILL NOT CALL UP THAT WHICH I CANNOT PUT DOWN"

A couple people asked me what I meant, and I didn't have the background then to explain. Well, this post is the background. People are using the *contingent* stupidity of our current government to replace lots of human interaction with mechanisms that cannot be coordinated even in principle. I totally understand why all these things are good right now when most of what our government does is stupid and unnecessary. But there is going to come a time when – after one too many bioweapon or nanotech or nuclear incidents – we, as a civilization, are going to wish we hadn't established untraceable and unstoppable ways of selling products.

And if we ever get real live superintelligence, pretty much by definition it is going to have >51% of the power and all attempts at "coordination" with it will be useless.

So I agree with Robin Hanson: This is the dream time. This is a rare confluence of circumstances where the we are unusually safe from multipolar traps, and as such weird things like art and science and philosophy and love can flourish.

As technological advance increases, the rare confluence will come to an end. New opportunities to throw values under the bus for increased competitiveness will arise. New ways of copying agents to increase the population will soak up our excess resources and resurrect Malthus' unquiet spirit. Capitalism and democracy, previously our protectors, will figure out ways to route around their inconvenient dependence on human values. And our coordination power will not be nearly up to the task, assuming somthing much more powerful than all of us combined doesn't show up and crush our combined efforts with a wave of its paw.

Absent an extraordinary effort to divert it, the river reaches the sea in one of two places.

It can end in Eliezer Yudkowsky's nightmare of a superintelligence optimizing for some random thing (classically paper clips) because we weren't smart enough to channel its optimization efforts the right way. This is the ultimate trap, the trap that catches the universe. Everything except the one thing being maximized is destroyed utterly in pursuit of the single goal, including all the silly human values.

Or it can end in Robin Hanson's nightmare (he doesn't call it a nightmare, but I think he's wrong) of a competition between emulated humans that can copy themselves and edit their own source code as desired. Their total self-control can wipe out even the *desire* for human values in their all-consuming contest. What happens to art, philosophy, science, and love in such a world? Zack Davis puts it with characteristic genius:

> I am a contract-drafting em,
> The loyalest of lawyers!
> I draw up terms for deals 'twixt firms
> To service my employers!
>
> But in between these lines I write
> Of the accounts receivable,

I’m stuck by an uncanny fright;
The world seems unbelievable!

How did it all come to be,
That there should be such ems as me?
Whence these deals and whence these firms
And whence the whole economy?

I am a managerial em;
I monitor your thoughts.
Your questions must have answers,
But you’ll comprehend them not.
We do not give you server space
To ask such things; it’s not a perk,
So cease these idle questionings,
And please get back to work.

Of course, that’s right, there is no junction
At which I ought depart my function,
But perhaps if what I asked, I knew,
I’d do a better job for you?

To ask of such forbidden science
Is gravest sign of noncompliance.
Intrusive thoughts may sometimes barge in,
But to indulge them hurts the profit margin.
I do not know our origins,
So that info I can not get you,
But asking for as much is sin,
And just for that, I must reset you.

But—

Nothing personal.

…

> I am a contract-drafting em,
> The loyalest of lawyers!
> I draw up terms for deals ‘twixt firms
> To service my employers!
>
> *When obsolescence shall this generation waste,*
> *The market shall remain, in midst of other woe*
> *Than ours, a God to man, to whom it sayest:*
> *“Money is time, time money – that is all*
> *Ye know on earth, and all ye need to know.”*

But even after we have thrown away science, art, love, and philosophy, there’s still one thing left to lose, one final sacrifice Moloch might demand of us. Bostrom again:

> It is conceivable that optimal efficiency would be attained by grouping capabilities in aggregates that roughly match the cognitive architecture of a human mind…But in the absence of any compelling reason for being confident that this so, we must countenance the possibility that human-like cognitive architectures are optimal only within the constraints of human neurology (or not at all). When it becomes possible to build architectures that could not be implemented well on biological neural networks, new design space opens up; and the global optima in this extended space need not resemble familiar types of mentality. Human-like cognitive organizations would then lack a niche in a competitive post-transition economy or ecosystem.
>
> We could thus imagine, as an extreme case, a technologically highly advanced society, containing many complex structures, some of them far more intricate and intelligent than anything that exists on the planet today – a society which nevertheless lacks any type of being that is conscious or whose welfare has

> moral significance. In a sense, this would be an uninhabited society. It would be a society of economic miracles and technological awesomeness, with nobody there to benefit. A Disneyland with no children.

The last value we have to sacrifice is being anything at all, having the lights on inside. With sufficient technology we will be “able” to give up even the final spark.

(Moloch whose eyes are a thousand blind windows!)

Everything the human race has worked for – all of our technology, all of our civilization, all the hopes we invested in our future – might be accidentally handed over to some kind of unfathomable blind idiot alien god that discards all of them, and consciousness itself, in order to participate in some weird fundamental-level mass-energy economy that leads to it disassembling Earth and everything on it for its component atoms.

(Moloch whose fate is a cloud of sexless hydrogen!)

Bostrom realizes that some people fetishize intelligence, that they are rooting for that blind alien god as some sort of higher form of life that ought to crush us for its own “higher good” the way we crush ants. He argues (Superintelligence, p. 219):

> The sacrifice looks even less appealing when we reflect that the superintelligence could realize a nearly-as-great good (in fractional terms) while sacrificing much less of our own potential well-being. Suppose that we agreed to allow *almost* the entire accessible universe to be converted into hedonium – everything except a small preserve, say the Milky Way, which would be set aside to accommodate our own needs. Then there would still be a hundred billion galaxies dedicated to the

> maximization of [the superintelligence’s own values]. But we would have one galaxy within which to create wonderful civilizations that could last for billions of years and in which humans and nonhuman animals could survive and thrive, and have the opportunity to develop into beatific posthuman spirits.

Remember: Moloch can’t agree even to this 99.99999% victory. Rats racing to populate an island don’t leave a little aside as a preserve where the few rats who live there can live happy lives producing artwork. Cancer cells don’t agree to leave the lungs alone because they realize it’s important for the body to get oxygen. Competition and optimization are blind idiotic processes and they fully intend to deny us even one lousy galaxy.

> They broke their backs lifting Moloch to Heaven! Pavements, trees, radios, tons! lifting the city to Heaven which exists and is everywhere about us!

We will break our back lifting Moloch to Heaven, but unless something changes it will be his victory and not ours.

I Broke My Back
Lifting Moloch To
Heaven, And All
I Got Was This
Lousy
Disneyland With
No Children

V.

"Gnon" is Nick Land's shorthand for "Nature And Nature's God", except the A is changed to an O and the whole thing is reversed, because Nick Land react to comprehensibility the same way as vampires to sunlight.

Land argues that humans should be more Gnon-conformist (pun Gnon-intentional). He says we do all these stupid things like divert useful resources to feed those who could never survive on their own, or supporting the poor in ways that encourage dysgenic reproduction, or allowing cultural degeneration to undermine the state. This means our society is denying natural law, basically listening to Nature say things like "this cause has this effect" and putting our fingers in our ears and saying "NO IT DOESN'T". Civilizations that do this too much tend to decline and fall, which is Gnon's fair and dispassionately-applied punishment for violating His laws.

He identifies Gnon with Kipling's Gods of the Copybook Headings.

> @AnarchoPapist Yes, the Gods of the Copybook Headings are practically indistinguishable from Gnon.
>
> — Outsideness (@Outsideness) July 13, 2014

These are of course the proverbs from Kipling's eponymous poem – maxims like "If you don't work, you die" and "The wages of sin is Death". If you have somehow not yet read it, I predict you will find it delightful regardless of what you think of its politics.

I notice that it takes only a slight irregularity in the abbreviation of "headings" – far less irregularity than it takes to turn "Nature and Nature's God" into "Gnon" – for the proper acronym of “Gods of the Copybook Headings” to be “GotCHa”.

I find this appropriate.

“If you don’t work, you die.” Gotcha! If you *do* work, you *also* die! Everyone dies, unpredictably, at a time not of their own choosing, and all the virtue in the world does not save you.

“The wages of sin is Death.” Gotcha! The wages of everything is Death! This is a Communist universe, the amount you work makes no difference to your eventual reward. From each according to his ability, to each Death.

“Stick to the Devil you know.” Gotcha! The Devil you know is Satan! And if he gets his hand on your soul you either die the true death, or get eternally tortured forever, or somehow both at once.

Since we’re starting to get into Lovecraftian monsters, let me bring up one of Lovecraft’s less known short stories, The Other Gods.

It’s only a couple of pages, but if you absolutely refuse to read it – the gods of Earth are relatively young as far as deities go. A very strong priest or magician can occasionally outsmart and overpower them – so Barzai the Wise decides to climb their sacred mountain and join in their festivals, whether they want him to or not.

But the beyond the seemingly tractable gods of Earth lie the Outer Gods, the terrible omnipotent beings of incarnate cosmic chaos. As soon as Barzai joins in the festival, the Outer Gods show up and pull him screaming into the abyss.

As stories go, it lacks things like plot or characterization or setting or point. But for some reason it stuck with me.

And identifying the Gods Of The Copybook Headings with Nature seems to me the same magnitude of mistake as identifying the gods of Earth with the Outer Gods. And likely to end about the same way: Gotcha!

You break your back lifting Moloch to Heaven, and then Moloch turns on you and gobbles you up.

More Lovecraft: the Internet popularization of the Cthulhu Cult claims that if you help free Cthulhu from his watery grave, he will reward you by eating you first, thus sparing you the horror of seeing everyone else eaten. This is a misrepresentation of the original text. In the original, his cultists receive no reward for freeing him from his watery prison, not even the reward of being killed in a slightly less painful manner.

> The thought that abstract ideas can be Lovecraftian monsters is an old one but a deep one.
>
> — Steven Kaas (@stevenkaas) January 25, 2011

On the margin, compliance with the Gods of the Copybook Headings, Gnon, Cthulhu, whatever, may buy you slightly more time than the next guy. But then again, it might not. And in the long run, we’re all dead and our civilization has been destroyed by unspeakable alien monsters.

At some point, somebody has to say “You know, maybe freeing Cthulhu from his watery prison is a *bad idea*. Maybe we should *not do that*.”

That person will not be Nick Land. He is totally one hundred percent in favor of freeing Cthulhu from his watery prison and extremely annoyed that it is not happening fast enough. I have *such mixed feelings* about Nick Land. On the grail quest for the True Futurology, he has gone 99.9% of the path and then missed the *very* last turn, the one marked ORTHOGONALITY THESIS.

But the thing about grail quests is – if you make a wrong turn two blocks away from your house, you end up at the corner store feeling mildly embarrassed. If you do *almost* everything right and then miss the very last turn, you end up being eaten by the legendary Black Beast of Aaargh whose ichorous stomach acid erodes your very soul into gibbering fragments.

As far as I can tell from reading his blog, Nick Land is the guy in that terrifying border region where he is smart enough to figure out several important arcane principles about summoning demon gods, but not quite smart enough to figure out the most important such principle, which is NEVER DO THAT.

VI.

Warg Franklin analyzes the same situation and does a little better. He names "the Four Horsemen of Gnon" – capitalism, war, evolution, and memetics – the same processes I talked about above.

From Capturing Gnon:

> Each component of Gnon detailed above had and has a strong hand in creating us, our ideas, our wealth, and our dominance, and thus has been good in that respect, but we must remember that [he] can and will turn on us when circumstances change. Evolution becomes dysgenic, features of the memetic landscape promote ever crazier insanity, productivity turns to famine when

> we can no longer compete to afford our own existence, and order turns to chaos and bloodshed when we neglect martial strength or are overpowered from outside. These processes are not good or evil overall; they are neutral, in the horrorist Lovecraftian sense of the word […]
>
> Instead of the destructive free reign of evolution and the sexual market, we would be better off with deliberate and conservative patriarchy and eugenics driven by the judgement of man within the constraints set by Gnon. Instead of a "marketplace of ideas" that more resembles a festering petri-dish breeding superbugs, a rational theocracy. Instead of unhinged techno-commercial exploitation or naive neglect of economics, a careful bottling of the productive economic dynamic and planning for a controlled techno-singularity. Instead of politics and chaos, a strong hierarchical order with martial sovereignty. These things are not to be construed as complete proposals; we don't really know how to accomplish any of this. They are better understood as goals to be worked towards. This post concerns itself with the "what" and "why", rather than the "how".

This seems to me the strongest argument for authoritarianism. Multipolar traps are likely to destroy us, so we should shift the tyranny-multipolarity tradeoff towards a rationally-planned garden, which requires centralized monarchical authority and strongly-binding traditions.

But a brief digression into social evolution. Societies, like animals, evolve. The ones that survive spawn memetic descendants – for example, the success of Britan allowed it to spin off Canada, Australia, the US, et cetera. Thus, we expect societies that exist to be somewhat optimized for stability and prosperity. I think this is one of the strongest conservative arguments. Just as a random change to a letter in the human genome will probably be deleterious rather than beneficial since humans are a complicated fine-tuned system whose genome has been pre-optimized for survival – so most changes to our cultural DNA will disrupt some institution that evolved to help Anglo-American (or whatever) society outcompete its real and hypothetical rivals.

The liberal counterargument to that is that evolution is a blind idiot alien god that optimizes for stupid things and has no concern with human value. Thus, the fact that some species of wasps paralyze caterpillars, lay their eggs inside of it, and have its young devour the still-living paralyzed caterpillar from the inside doesn't set off evolution's moral sensor, because evolution doesn't *have* a moral sensor because evolution doesn't care.

Suppose that in fact patriarchy is adaptive to societies because it allows women to spend all their time bearing children who can then engage in productive economic activity and fight wars. The social evolutionary processes that cause societies to adopt patriarchy *still* have exactly as little concern for its moral effects on women as the biological evolutionary processes that cause wasps to lay their eggs in caterpillars.

Evolution doesn't care. But we do care. There's a tradeoff between Gnon-compliance – saying "Okay, the strongest possible society is a patriarchal one, we should implement patriarchy" and our human values – like women who want to do something other than bear children.

Too far to one side of the tradeoff, and we have unstable impoverished societies that die out for going against natural law. Too far to the other side, and we have lean mean fighting machines that are murderous and miserable. Think your local anarchist commune versus Sparta.

Franklin acknowledges the human factor:

> And then there’s us. Man has his own telos, when he is allowed the security to act and the clarity to reason out the consequences of his actions. When unafflicted by coordination problems and unthreatened by superior forces, able to act as a gardener rather than just another subject of the law of the jungle, he tends to build and guide a wonderful world for himself. He tends to favor good things and avoid bad, to create secure civilizations with polished sidewalks, beautiful art, happy families, and glorious adventures. I will take it as a given that this telos is identical with “good” and “should”.
>
> Thus we have our wildcard and the big question of futurism. Will the future be ruled by the usual four horsemen of Gnon for a future of meaningless gleaming techno-progress burning the cosmos or a future of dysgenic, insane, hungry, and bloody dark ages; or will the telos of man prevail for a future of meaningful art, science, spirituality, and greatness?

Franklin continues:

> The project of civilization [is] for man to graduate from the metaphorical savage, subject to the law of the jungle, to the civilized gardener who, while theoretically still subject to the law of the jungle, is so dominant as to limit the usefulness of that model.
>
> This need not be done globally; we may only be able to carve out a small walled garden for ourselves, but make

> no mistake, even if only locally, the project of civilization is to capture Gnon.

I maybe agree with Warg here more than I have ever agreed with anyone else about anything. He says something really important and he says it beautifully and there are so many words of praise I want to say for this post and for the thought processes behind it.

But what I am actually going to say is…

Gotcha! You die anyway!

Suppose you make your walled garden. You keep out all of the dangerous memes, you subordinate capitalism to human interests, you ban stupid bioweapons research, you *definitely* don't research nanotechnology or strong AI.

Everyone outside *doesn't* do those things. And so the only question is whether you'll be destroyed by foreign diseases, foreign memes, foreign armies, foreign economic competition, or foreign existential catastrophes.

As foreigners compete with you – and there's no wall high enough to block all competition – you have a couple of choices. You can get outcompeted and destroyed. You can join in the race to the bottom. Or you can invest more and more civilizational resources into building your wall – whatever that is in a non-metaphorical way – and protecting yourself.

I can imagine ways that a "rational theocracy" and "conservative patriarchy" might not be terrible to live under, given exactly the right conditions. But you don't get to choose exactly the right conditions. You get to choose the extremely constrained set of conditions that "capture

Gnon”. As outside civilizations compete against you, your conditions will become more and more constrained.

Warg talks about trying to avoid “a future of meaningless gleaming techno-progress burning the cosmos”. Do you really think your walled garden will be able to ride this out?

Hint: is it part of the cosmos?

Yeah, you’re kind of screwed.

I want to critique Warg. But I want to critique him in the exact opposite direction as the last critique he received. In fact, the last critique he received is so bad that I want to discuss it at length so we can get the correct critique entirely by taking its exact mirror image.

So here is Hurlock’s On Capturing Gnon And Naive Rationalism.

Hurlock spouts only the most craven Gnon-conformity. A few excerpts:

> In a recent piece [Warg Franklin] says that we should try to “capture Gnon”, and somehow establish control over his forces, so that we can use them to our own advantage. Capturing or creating God is indeed a classic transhumanist fetish, which is simply another form of the oldest human ambition ever, to rule the universe.
>
> Such naive rationalism however, is extremely dangerous. The belief that it is human Reason and deliberate human design which creates and maintains civilizations was probably the biggest mistake of Enlightenment philosophy…

> It is the theories of Spontaneous Order which stand in direct opposition to the naive rationalist view of humanity and civilization. The consensus opinion regarding human society and civilization, of all representatives of this tradition is very precisely summarized by Adam Ferguson's conclusion that "nations stumble upon [social] establishments, which are indeed the result of human action, but not the execution of any human design". Contrary to the naive rationalist view of civilization as something that can be and is a subject to explicit human design, the representatives of the tradition of Spontaneous Order maintain the view that human civilization and social institutions are the result of a complex evolutionary process which is driven by human interaction but not explicit human planning.
>
> Gnon and his impersonal forces are not enemies to be fought, and even less so are they forces that we can hope to completely "control". Indeed the only way to establish some degree of control over those forces is to submit to them. Refusing to do so will not deter these forces in any way. It will only make our life more painful and unbearable, possibly leading to our extinction. Survival requires that we accept and submit to them. Man in the end has always been and always will be little more than a puppet of the forces of the universe. To be free of them is impossible.
>
> Man can be free only by submitting to the forces of Gnon.

I accuse Hurlock of being stuck behind the veil. When the veil is lifted, Gnon-aka-the-GotCHa-aka-the-Gods-of-Earth turn out to be Moloch-aka-the-Outer-Gods. Submitting to them doesn't make you "free", there's no spontaneous order, any gifts they have given you are an unlikely and contingent output of a blind idiot process whose next iteration will just as happily destroy you.

Submit to Gnon? Gotcha! As the Antarans put it, “you may not surrender, you can not win, your only option is to die.”

VII.

So let me confess guilt to one of Hurlock’s accusations: I am a transhumanist and I really do want to rule the universe.

Not personally – I mean, I wouldn’t object if someone personally offered me the job, but I don’t expect anyone will. I would like humans, or something that respects humans, or at least gets along with humans – to have the job.

But the current rulers of the universe – call them what you want, Moloch, Gnon, whatever – want us dead, and with us everything we value. Art, science, love, philosophy, consciousness itself, the entire bundle. And since I’m not down with that plan, I think defeating them and taking their place is a pretty high priority.

The opposite of a trap is a garden. The only way to avoid having all human values gradually ground down by optimization-competition is to install a Gardener over the entire universe who optimizes for human values.

And the whole point of Bostrom’s *Superintelligence* is that this is within our reach. Once humans can design machines that are smarter than we are, by definition they’ll be able to design machines which are smarter than they are, which can design machines smarter than they are, and so on in a feedback loop so tiny that it will smash up against the physical limitations for intelligence in a comparatively lightning-short amount of time. If multiple competing entities were likely to do that at once, we would be super-doomed. But the sheer speed of the cycle makes it possible that we will end up with one entity light-years ahead of the rest of civilization, so much so that it can suppress any competition – including competition for its title of most powerful entity – permanently. In the very near future, we are going to lift *something* to Heaven. It might be Moloch. But it might be something on our side. If it’s on our side, it can *kill Moloch dead*.

And if that entity shares human values, it can allow human values to flourish unconstrained by natural law.

I realize that sounds like hubris – it certainly did to Hurlock – but I think it’s the opposite of hubris, or at least a hubris-minimizing position.

To expect God to care about you or your personal values or the values of your civilization, that’s hubris.

To expect God to bargain with you, to allow you to survive and prosper as long as you submit to Him, that’s hubris.

To expect to wall off a garden where God can’t get to you and hurt you, that’s hubris.

To expect to be able to remove God from the picture entirely…well, at least it’s an actionable strategy.

I am a transhumanist because I do not have enough hubris not to try to kill God.

VIII.

The Universe is a dark and foreboding place, suspended between alien deities. Cthulhu, Gnon, Moloch, call them what you will.

Somewhere in this darkness is another god. He has also had many names. In the *Kushiel* books, his name was Elua. He is the god of flowers and free love and all soft and fragile things. Of art and science and philosophy and love. Of niceness, community, and civilization. He is a god of humans.

The other gods sit on their dark thrones and think “Ha ha, a god who doesn’t even control any hell-monsters or command his worshippers to become killing machines. What a weakling! This is going to be so easy!”

But somehow Elua is still here. No one knows exactly how. And the gods who oppose Him tend to find Themselves meeting with a *surprising* number of unfortunate accidents.

There are many gods, but this one is ours.

Bertrand Russell said: “One should respect public opinion insofar as is necessary to avoid starvation and keep out of prison, but anything that goes beyond this is voluntary submission to an unnecessary tyranny.”

So be it with Gnon. Our job is to placate him insofar as is necessary to avoid starvation and invasion. And that only for a short time, until we come into our full power.

“It is only a childish thing, that the human species has not yet outgrown. And someday, we’ll get over it.”

Other gods get placated until we’re strong enough to take them on. Elua gets worshipped.

64. My paladin's battle cry is not allowed to be 'Good for the Good God!'".

— □□□ □ (@tabatkins) March 28, 2014

I think this is an excellent battle cry

And at some point, matters will come to a head.

The question everyone has after reading Ginsberg is: what is Moloch?

My answer is: Moloch is exactly what the history books say he is. He is the god of child sacrifice, the fiery furnace into which you can toss your babies in exchange for victory in war.

He always and everywhere offers the same deal: throw what you love most into the flames, and I can grant you power.

As long as the offer's open, it will be irresistible. So we need to close the offer. Only another god can kill Moloch. We have one on our side, but he needs our help. We should give it to him.

Ginsberg's poem famously begins "I saw the best minds of my generation destroyed by madness". I am luckier than Ginsberg. I got to see the best minds of my generation identify a problem and *get to work*.

(*Visions! omens! hallucinations! miracles! ecstasies! gone down the American river!*

Dreams! adorations! illuminations! religions! the whole boatload of sensitive bullshit!

Breakthroughs! over the river! flips and crucifixions! gone down the flood! Highs! Epiphanies! Despairs! Ten years' *animal screams and suicides! Minds! New loves! Mad generation! down on the rocks of Time!*

Real holy laughter in the river! They saw it all! the wild eyes! the holy yells! They bade farewell! They jumped off the roof! to solitude! waving! carrying flowers! Down to the river! into the street!)

[Also available as podcast here This post is represented by an NFT here]

Nobody Is Perfect, Everything Is Commensurable

I.

Recently spotted on Tumblr:

> “This is going to be an unpopular opinion but I see stuff about ppl not wanting to reblog ferguson things and awareness around the world because they do not want negativity in their life plus it will cause them to have anxiety. They come to tumblr to escape n feel happy which think is a load of bull. There r literally ppl dying who live with the fear of going outside their homes to be shot and u cant post a fucking picture because it makes u a little upset?”
>
> “Can yall maybe take some time away from reblogging fandom or humor crap and read up and reblog pakistan because the privilege you have of a safe bubble is not one shared by others?”

Ignore the questionable stylistic choices and there’s an important point here worth considering. Something like “Yes, the feeling of constantly being outraged and mired in the latest controversy is unpleasant. And yes, it would be nice to get to avoid it and spend time with your family and look at kitten pics or something. But when the controversy is about people being murdered in cold blood, or living in fear, or something like that – then it’s your duty as a decent human being to care. In the best case scenario you’ll discharge that duty by organizing widespread protests or something – but the *absolute least* you can do is reblog a couple of slogans.”

I think Cliff Pervocracy is trying to say something similar in this post. Key excerpt:

> When you’ve grown up with messages that you’re incompetent to make your own decisions, that you don’t deserve any of the things you have, and that you’ll never be good enough, the [conservative] fantasy of rugged individualism starts looking pretty damn good.
>
> Intellectually, I think my current political milieu of feminism/progressivism/social justice is more correct, far better for the world in general, and more helpful to me since I don’t actually live in a perfectly isolated cabin.
>
> But god, it’s uncomfortable. It’s intentionally uncomfortable—it’s all about getting angry at injustice and questioning the rightness of your own actions and being sad so many people still live such painful lives. Instead of looking at your cabin and declaring “I shall name it…CLIFFORDSON MANOR,” you need to look at your cabin and recognize that a long series of brutal injustices are responsible for the fact that you have a white-collar job that lets you buy a big useless house in the woods while the original owners of the land have been murdered or forced off it.
>
> And you’re never good enough. You can be good—certainly you get major points for charity and activism and fighting the good fight—but not good enough. No matter what you do, you’re still participating in plenty of corrupt systems that enforce oppression. Short of bringing about a total revolution of everything, your work will never be done, you’ll never be good enough.

> Once again, to be clear, I don’t think this is wrong. I just think it’s a bummer.
>
> I don’t know of a solution to this. (Bummer again.) I don’t think progressivism can ever compete with the cozy self-satisfaction of the cabin fantasy. I don’t think it should. Change is necessary in the world, people don’t change if they’re totally happy and comfortable, therefore discomfort is necessary.

I’d like to make what I hope is a friendly amendment to Cliff’s post. He thinks he’s talking about progressivism versus conservativism, but he isn’t. A conservative happy with his little cabin and occasional hunting excursions, and a progressive happy with her little SoHo flat and occasional poetry slams, are psychologically pretty similar. So are a liberal who abandons a cushy life to work as a community organizer in the inner city and fight poverty, and a conservative who abandons a cushy life to serve as an infantryman in Afghanistan to fight terrorism. The distinction Cliff is trying to get at here isn’t left-right. It’s activist versus passivist.

As part of a movement recently deemed postpolitical, I have to admit I fall more on the passivist side of the spectrum – at least this particular conception of it. I talk about politics when they interest me or when I enjoy doing so, and I feel an obligation not to actively make things worse. But I don’t feel like I need to talk nonstop about whatever the designated Issue is until it distresses me and my readers both.

I’ve heard people give lots of reasons for not wanting to get into politics. For some, hearing about all the evils of the world makes them want to curl into a ball and cry for hours. Still others feel deep personal guilt about anything they hear – an almost psychotic belief that if people are being hurt anywhere in the world, it’s their fault for not preventing it. A few are chronically uncertain about which side to take and worried that anything they do will cause more harm than good. A couple have traumatic experiences that make them leery of affiliating with a particular side – did you know the prosecutor in the Ferguson case was the son of a police officer who was killed by a black suspect? And still others are perfectly innocent and just want to reblog kitten pictures.

Pervocracy admits this, and puts it better than I do:

> But god, it’s uncomfortable. It’s intentionally uncomfortable—it’s all about getting angry at injustice and questioning the rightness of your own actions and being sad so many people still live such painful lives. Instead of looking at your cabin and declaring “I shall name it…CLIFFORDSON MANOR,” you need to look at your cabin and recognize that a long series of brutal injustices are responsible for the fact that you have a white-collar job that lets you buy a big useless house in the woods while the original owners of the land have been murdered or forced off it. And you’re never good enough. You can be good—certainly you get major points for charity and activism and fighting the good fight—but not good enough. No matter what you do, you’re still participating in plenty of corrupt systems that enforce oppression. Short of bringing about a total revolution of everything, your work will never be done, you’ll never be good enough.

That seems about right. Pervocracy ends up with discomfort, and I’m in about the same place. But other, less stable people end up with self-loathing. Still other people go further than that, into Calvinist-style “perhaps I am a despicable worm unworthy of existence”. moteinthedark’s reply to Pervocracy gives me the impression that she struggles with this sometime. For these people, abstaining from politics is the only coping tool they have.

But the counterargument is *still* that you’ve got a lot of chutzpah playing that card when people in Peshawar or Ferguson or Iraq don’t have access to this coping tool. You can’t just bring in a doctor’s note and say “As per my psychiatrist, I have a mental health issue and am excused from experiencing concern for the less fortunate.”

One option is to deny the obligation. I am super sympathetic to this one. The *marginal* cost of my existence on the poor and suffering of the world is zero. In fact, it’s probably positive. My economic activity consists mostly of treating patients, buying products, and paying taxes. The first treats the poor’s illnesses, the second creates jobs, and the third pays for government assistance programs. Exactly what am I supposed to be apologizing for here? I may benefit from the genocide of the Indians in that I live on land that was formerly Indian-occupied. But I also benefit from the asteroid that killed the dinosaurs, in that I live on land that was formerly dinosaur-occupied. I don’t feel like I’m complicit in the asteroid strike; why should I feel complicit in the genocide?

I have no objection to people who say this. The problem with it isn’t philosophical, it’s emotional. For *most people* it won’t be enough. The old saying goes “you can’t reason yourself out of something you didn’t reason yourself into to begin with”, and the idea that secure and prosperous people need to “give something back” is a lot older than accusations of “being complicit in structures of oppression”. It’s probably older than the Bible. People feel a deep-seated need to show that they understand how lucky they are and help those less fortunate than themselves.

So what do we do with the argument that we are morally obligated to be political activists, possibly by reblogging everything about Ferguson that crosses our news feed?

II.

We ask: *why the heck are we privileging that particular subsection of the category "improving the world"?*

Pervocracy says that "short of bringing about a total revolution of everything, your work will never be done, you'll never be good enough." But he is overly optimistic. Has your total revolution of everything eliminated ischaemic heart disease? Cured malaria? Kept elderly people out of nursing homes? No? Then you haven't discharged your infinite debt yet!

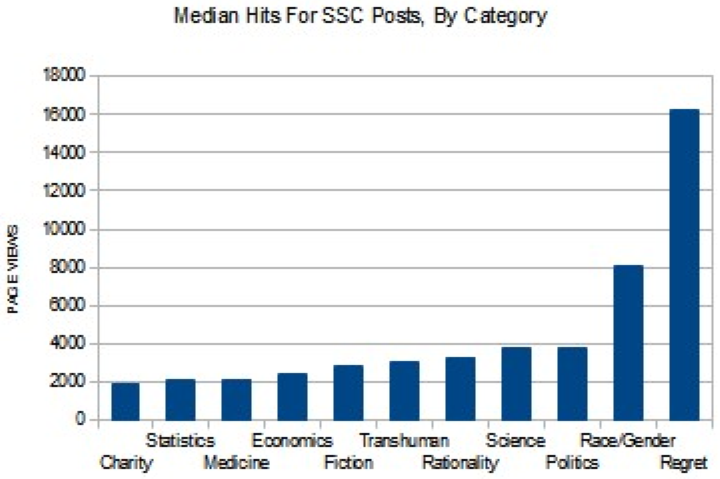
Being a perfect person doesn't just mean participating in every hashtag campaign you hear about. It means spending all your time at soup kitchens, becoming vegan, donating everything you have to charity, calling your grandmother up every week, and marrying Third World refugees who need visas rather than your one true love.

And not all of these things are equally important.

Five million people participated in the #BlackLivesMatter Twitter campaign. Suppose that solely as a result of this campaign, no currently-serving police officer ever harms an unarmed black person ever again. That's 100 lives saved per year times let's say twenty years left in the average officer's career, for a total of 2000 lives saved, or 1/2500th of a life saved per campaign participant. By coincidence, 1/2500th of a life saved happens to be what you get when you donate $1 to the Against Malaria Foundation. The round-trip bus fare people used to make it to their #BlackLivesMatter protests could have saved ten times as many black lives as the protests themselves, *even given completely ridiculous overestimates of the protests' efficacy*.

The moral of the story is that if you feel an obligation to give back to the world, participating in activist politics is one of the worst possible ways to do it. Giving even a tiny amount of money to charity is hundreds or even thousands of times more effective than almost any political action you can take. Even if you're absolutely convinced a certain political issue is the most important thing in the world, you'll effect more change by donating money to nonprofits lobbying about it than you will be reblogging anything.

There is *no reason* that politics would even *come to the attention* of an unbiased person trying to "break out of their bubble of privilege" or "help people who are afraid of going outside of their house". Anybody saying that people who want to do good need to spread their political cause is about as credible as a televangelist saying that people who want to do good need to give them money to buy a new headquarters. It's possible that televangelists having beautiful headquarters might be *slightly* better than them having hideous headquarters, but it's not the first thing a reasonable person trying to improve the world would think of.



Average number of hits for posts on this blog, by topic

Nobody cares about charity. Everybody cares about politics, especially race and gender. Just as televangelists who are obsessed with moving to a sweeter pad may come to think that donating to their building fund is the one true test of a decent human being, so our universal obsession with politics, race, and gender incites people to make convincing arguments that taking and spreading the right position on those issues is the one true test of a decent human being.

So now we have an angle of attack against our original question. “Am I a bad person for not caring more about politics?” Well, every other way of doing good, especially charity, is more important than politics. So this question is strictly superseded by “Am I a bad person for not engaging in every other way of doing good, especially charity?” And then once we answer that, we can ask “Also, however much sin I have for not engaging in charity, should we add another mass of sin, about 1% as large, for my additional failure to engage in politics?”

And Cliff Pervocracy’s concern of “Even if I do a lot of politics, am I still a bad person for not doing *all* the politics?”

is superseded by “Even if I give a lot of charity, am I a bad person for not doing *all* the charity? And then a bad person in an additional way, about 1% as large, for not doing all the politics as well?”

There’s no good answer to this question. If you want to feel anxiety and self-loathing for not giving 100% of your income, minus living expenses, to charity, then no one can stop you.

I, on the other hand, would prefer to call that “not being perfect”. I would prefer to say that if you feel like you will live in anxiety and self-loathing until you have given a certain amount of money to charity, you should make that certain amount ten percent.

Why ten percent?

It’s ten percent because that’s the standard decreed by Giving What We Can and the effective altruist community. Why should we believe their standard? I think we should believe it because if we reject it in favor of “No, you are a bad person unless you give *all* of it,” then everyone will just sit around feeling very guilty and doing nothing. But if we very clearly say “You have discharged your moral duty if you give ten percent or more,” then many people will give ten percent or more. The most important thing is having a Schelling point, and ten percent is nice, round, divinely ordained, and – crucially – the Schelling point upon which we have already settled. It is an *active* Schelling point. If you give ten percent, you can have your name on a nice list and get access to a secret forum on the Giving What We Can site which is actually pretty boring.

It’s ten percent because definitions were made for Man, not Man for definitions, and if we define “good person” in a way such that everyone is sitting around miserable because they can't reach an unobtainable standard, we are stupid definition-makers. If we are smart definition-makers, we will define it in whichever way which makes it the most effective tool to convince people to give at least that much.

Finally, it's ten percent because if you believe in something like universalizability as a foundation for morality, a world in which everybody gives ten percent of their income to charity is a world where about seven trillion dollars go to charity a year. Solving global poverty forever is estimated to cost about $100 billion a year for the couple-decade length of the project. That's *about two percent* of the money that would suddenly become available. If charity got seven trillion dollars a year, *the first year* would give us enough to solve global poverty, eliminate all treatable diseases, fund research into the untreatable ones for approximately the next forever, educate anybody who needs educating, feed anybody who needs feeding, fund an unparalleled renaissance in the arts, permamently save every rainforest in the world, and have enough left over to launch five or six different manned missions to Mars. That would be the *first year*. Goodness only knows what would happen in Year 2.

(by contrast, if everybody in the world retweeted the latest hashtag campaign, Twitter would break.)

Charity is in some sense the perfect unincentivized action. If you think the most important thing to do is to cure malaria, then a charitable donation is deliberately throwing the power of your brain and muscle behind the cause of curing malaria. If, as I've argued, the reason we can't solve world poverty and disease and so on is the capture of our financial resources by the undirected dance of incentives, then what better way to fight back than by saying "Thanks but no thanks, I'm taking this abstract representation of my resources and using it *exactly* how I think it should most be used”?

If you give 10% per year, you have done your part in making that world a reality. You can honestly say “Well, it’s not my fault that everyone *else* is still dragging their feet.”

III.

Once the level is fixed at ten percent, we get a better idea how to answer the original question: “If I want to be a good person who gives back to the community, but I am triggered by politics, what do I do?” You do good in a way that doesn’t trigger you. Another good thing about having less than 100% obligation is that it gives you the opportunity to budget and trade-off. If you make $30,000 and you accept 10% as a good standard you want to live up to, you can either donate $3000 to charity, or participate in political protests until your number of lives or dollars or DALYs saved is equivalent to that.

Nobody is perfect. This gives us license not to be perfect either. Instead of aiming for an impossible goal, falling short, and not doing anything at all, we set an arbitrary but achievable goal designed to encourage the most people to do as much as possible. That goal is ten percent.

Everything is commensurable. This gives us license to determine exactly how we fulfill that ten percent goal. Some people are triggered and terrified by politics. Other people are too sick to volunteer. Still others are poor and cannot give very much money. But money is a constant reminder that everything goes into the same pot, and that you can fulfill obligations in multiple equivalent ways. Some people will not be able to give ten percent of their income without excessive misery, but I bet thinking about their contribution in terms of a fungible good will help them decide how much volunteering or activism they need to reach the equivalent.

Cliff Pervocracy says “Your work will never be done, you’ll never be good enough.” This seems like a recipe for – at best – undirected misery, stewing in self-loathing, and total defenselessness against the first parasitic meme to come along and tell them to engage in the latest conflict or else they’re trash. At worst, it autocatalyzes an opposition of egoists who laugh at the idea of helping others.

On the other hand, Jesus says “Take my yoke upon you…and you will find rest for your souls. For my yoke is easy and my burden is light.” This seems like a recipe for getting people to say “Okay, I’ll take your yoke upon me! Thanks for the offer!”

Persian poet Omar Khayyam, considering the conflict between the strict laws of Islam and his own desire to enjoy life, settles upon the following rule:

> Heed not the Sunna, nor the law divine;
> If to the poor their portion you assign,
> And never injure one, nor yet abuse,
> I guarantee you heaven, as well as wine!

I’m not saying that donating 10% of your money to charity makes you a great person who is therefore freed of every other moral obligation. I’m not saying that anyone who chooses not to do it is therefore a bad person. I’m just saying that if you feel a need to discharge some feeling of a moral demand upon you to help others, and you want to do it intelligently, it beats most of the alternatives.

This month is the membership drive for Giving What We Can, the organization of people who have promised to give 10% of their earnings to charity. I am a member. Ozy is an aspiring member who plans to join once they are making a salary. Many of the commenters here are members – I recognize for example Taymon Beal’s name on their list. Some well-known moral philosophers like Peter Singer and Derek Parfit are members. Seven hundred other people are also members.

I would recommend giving them a look.

Studies On Slack

I.

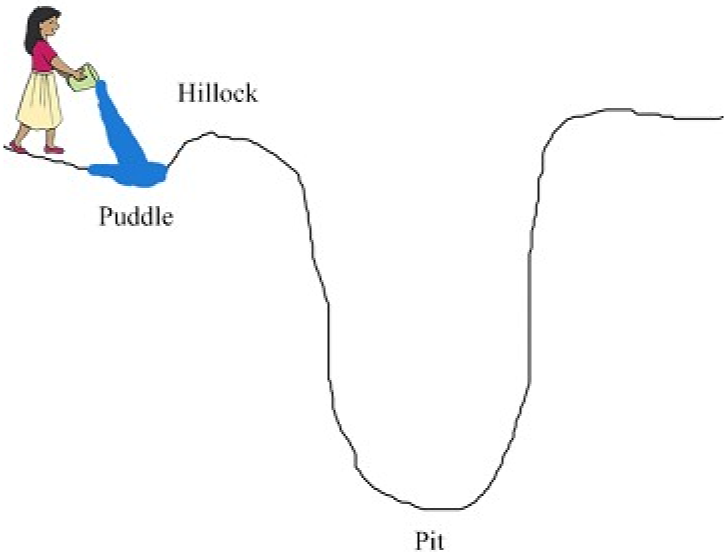
Imagine a distant planet full of eyeless animals. Evolving eyes is hard: they need to evolve Eye Part 1, then Eye Part 2, then Eye Part 3, in that order. Each of these requires a separate series of rare mutations.

Here on Earth, scientists believe each of these mutations must have had its own benefits – in the land of the blind, the man with only Eye Part 1 is king. But on this hypothetical alien planet, there is no such luck. You need all three Eye Parts or they're useless. Worse, each Eye Part is metabolically costly; the animal needs to eat 1% more food per Eye Part it has. An animal with a full eye would be much more fit than anything else around, but an animal with only one or two Eye Parts will be at a small disadvantage.

So these animals will only evolve eyes in conditions of relatively weak evolutionary pressure. In a world of intense and perfect competition, where the fittest animal always survives to reproduce and the least fit always dies, the animal with Eye Part 1 will always die – it's less fit than its fully-eyeless peers. The weaker the competition, and the more randomness dominates over survival-of-the-fittest, the more likely an animal with Eye Part 1 can survive and reproduce long enough to eventually produce a descendant with Eye Part 2, and so on.

There are lots of ways to decrease evolutionary pressure. Maybe natural disasters often decimate the population, dozens of generations are spent recolonizing empty land, and during this period there's more than enough for everyone and nobody has to compete. Maybe there are frequent whalefalls, and any animal nearby has hit the evolutionary jackpot and will have thousands of descendants. Maybe the population is isolated in little islands and mountain valleys, and one gene or another can reach fixation in a population totally by chance. It doesn't matter exactly how it happens, it matters that evolutionary pressure is low.

The branch of evolutionary science that deals with this kind of situation is called "adaptive fitness landscapes". Landscapes really are a great metaphor – consider somewhere like this:



You pour out a bucket of water. Water "flows downhill", so it's tempting to say something like "water wants to be at the lowest point possible". But that's not quite right. The lowest point possible is the pit, and water won't go there. It will just sit in the little puddle forever, because it would have to go up the tiny little hillock in order to get to the pit, and water can't flow uphill. Using normal human logic, we feel tempted to say something like "Come on! The hillock is so tiny, and that pit is so deep, just make a single little exception to your ‘always flow downhill’ policy and you could do so much better for yourself!” But water stubbornly refuses to listen.

Under conditions of perfectly intense competition, evolution works the same way. We imagine a multidimensional evolutionary “landscape” where lower ground represents higher fitness. In this perfectly intense competition, organisms can go from lower to higher fitness, but never vice versa. As with water, the tiniest hillock will leave their potential forever unrealized.

Under more relaxed competition, evolution only *tends probabilistically* to flow downhill. Every so often, it will flow uphill; the smaller the hillock, the more likely evolution will surmount it. Given enough time, it’s guaranteed to reach the deepest pit and mostly stay there.

Take a moment to be properly amazed by this. It sounds like something out of the Tao Te Ching. An animal with eyes has very high evolutionary fitness. It will win at all its evolutionary competitions. So in order to produce the highest-fitness animal, we need to – select for fitness less hard? In order to produce an animal that wins competitions, we need to stop optimizing for winning competitions?

This doesn’t mean that less competition is always good. An evolutionary environment with no competition won’t evolve eyes either; a few individuals might randomly drift into having eyes, but they won’t catch on. In order to optimize the species as much as possible as fast as possible, you need the right balance, somewhere in the middle between total competition and total absence of competition.

In the esoteric teachings, total competition is called Moloch, and total absence of competition is called Slack. Slack (thanks to Zvi Mowshowitz for the term and concept) gets short shrift. If you think of it as “some people try to win competitions, other people don’t care about winning competitions and slack off and go to the beach”, you’re misunderstanding it. Think of slack as a paradox – the Taoist art of winning competitions by not trying too hard at them. Moloch and Slack are opposites and complements, like yin and yang. Neither is stronger than the other, but their interplay creates the ten thousand things.

II.

Before we discuss slack further, a digression on group selection.

Some people would expect this discussion to be quick, since group selection doesn’t exist. These people understand it as evolution acting for the good of a species. It’s a tempting way to think, because evolution usually eventually makes species stronger and more fit, and sometimes we colloquially round that off to evolution targeting a species’ greater good. But inevitably we find evolution is awful and does absolutely nothing of the sort.

Imagine an alien planet that gets hit with a solar flare once an eon, killing all unshielded animals. Sometimes unshielded animals spontaneously mutate to shielded, and vice versa. Shielded animals are completely immune to solar flares, but have 1% higher metabolic costs. What happens? If you predicted “magnetic shielding reaches fixation and all animals get it”, you’ve fallen into the group selection trap. The unshielded animals outcompete the shielded ones during the long inter-flare period, driving their population down to zero (though a few new shielded ones arise every generation through spontaneous mutations). When the flare comes, only the few spontaneous mutants survive. They breed a new entirely-shielded population, until a few unshielded animals arise through spontaneous mutation. The unshielded outcompete the shielded ones again, and by the time of the next solar flare, the population is 100% unshielded again and they all die. If the animals are lucky, there will always be enough spontaneously-mutated shielded animals to create a post-flare breeding population; if they are unlucky, the flare will hit at a time with unusually few such mutants, and the species will go extinct.

An Evolution Czar concerned with the good of the species would just declare that all animals should be shielded and solve the problem. In the absence of such a Czar, these animals will just keep dying in solar-flare-induced mass extinctions forever, even though there is an easy solution with only 1% metabolic cost.

A less dramatic version of the same problem happens here on Earth. Every so often predators (let’s say foxes) reproduce too quickly and outstrip the available supply of prey (let’s say rabbits). There is a brief period of starvation as foxes can’t find any more rabbits and die *en masse*. This usually ends with a boom-bust cycle: after most foxes die, the rabbits (who reproduce very quickly and are now free of predation) have a population boom; now there are rabbits everywhere. Eventually the foxes catch up, eat all the new rabbits, and the cycle repeats again. It’s a waste of resources for foxkind to spend so much of time and its energy breeding a huge population of foxes that will inevitably collapse a generation later; an Evolution Czar concerned with the common good would have foxes limit their breeding at a sustainable level. But since individual foxes that breed excessively are more likely to have their genes represented in the next generation than foxes that breed at a sustainable level, we end up with foxes that breed excessively, and the cycle continues.

(but humans are too smart to fall for this one, right?)

Some scientists tried to create group selection under laboratory conditions. They divided some insects into subpopulations, then killed off any subpopulation whose numbers got too high, and “promoted” any subpopulation that kept its numbers low to better conditions. They hoped the insects would evolve to naturally limit their family size in order to keep their subpopulation alive. Instead, the insects became cannibals: they ate other insects’ children so they could have more of their own without the total population going up. In retrospect, this makes perfect sense; an insect with the behavioral program “have many children, and also kill other insects’ children” will have its genes better represented in the next generation than an insect with the program “have few children”.

But sometimes evolution appears to solve group selection problems. What about multicellular life? Stick some cells together in a resource-plentiful environment, and they’ll naturally do the evolutionary competition thing of eating resources as quickly as possible to churn out as many copies of themselves as possible. If you were expecting these cells to form a unitary organism where individual cells do things like become heart cells and just stay in place beating rhythmically, you would call the expected normal behavior “cancer” and be against it. Your opposition would be on firm group selectionist grounds: if any cell becomes cancer, it and its descendants will eventually overwhelm everything, and the organism (including all cells within it, including the cancer cells) will die. So for the good of the group, none of the cells should become cancerous.

The first step in evolution’s solution is giving all cells the same genome; this mostly eliminates the need to compete to give their genes to the next generation. But this solution isn’t perfect; cells can get mutations in the normal course of dividing and doing bodily functions. So it employs a host of other tricks: genetic programs telling cells to self-destruct if they get too cancer-adjacent, an immune system that hunts down and destroys cancer cells, or growing old and dying (this last one isn’t usually thought of as a “trick”, but it absolutely is: if you arrange for a cell line to lose a little information during each mitosis, so that it degrades to the point of gobbledygook after X divisions, this means cancer cells that divide constantly will die very quickly, but normal cells dividing on an approved schedules will last for decades).

Why can evolution “develop tricks” to prevent cancer, but not to prevent foxes from overbreeding, or aliens from losing their solar flare shields? Group selection works when the group itself has a shared genetic code (or other analogous ruleset) that can evolve. It doesn’t work if you expect it to directly change the genetic code of each individual to cooperate more.

When we think of cancer, we are at risk of conflating two genetic codes: the shared genetic code of the multicellular organism, and the genetic code of each cell within the organism. Usually (when there are no mutations in cell divisions) these are the same. Once individual cells within the organism start mutating, they become different. Evolution will select for cancer in changes to individual cells’ genomes over an organism’s lifetime, but select against it in changes to the overarching genome over the lifetime of the species (ie you should expect all the genes you inherited from your parents to be selected against cancer, and all the mutations in individual cells you’ve gotten since then to be selected for cancer).

The fox population has no equivalent of the overarching genome; there is no set of rules that govern the behavior of every fox. So foxes can't undergo group selection to prevent overpopulation (there are some more complicated dynamics that might still be able to rescue the foxes in some situations, but they're not relevant to the simple model we're looking at).

In other words, group selection can happen in a two-layer hierarchy of nested evolutionary systems when the outer system (eg multicellular humans) includes rules that the inner system (eg human cells) have to follow, and where the fitness of the evolving-entities in the outer system depends on some characteristics of the evolving-entities in the inner system (eg humans are higher-fitness if their cells do not become cancerous). The evolution of the outer layer includes evolution over rulesets, and eventually evolves good strong rulesets that tell the inner-layer evolving entities how to behave, which can include group selection (eg humans evolve a genetic code that includes a rule "individual cells inside of me should not get cancer" and mechanisms for enforcing this rule).

You can find these kinds of two-layer evolutionary systems everywhere. For example, "cultural evolution" is a two-layer evolutionary system. In the hypothetical state of nature, there's unrestricted competition – people steal from and murder each other, and only the strongest survive. After they form groups, the groups compete with each other, and groups that develop rulesets that prevent theft and murder (eg legal codes, religions, mores) tend to win those competitions. Once again, the outer layer (competition between cultures) evolves groups that successfully constrains the inner layer (competition between individuals). Species don't have a czar who restrains internal competition in the interest of keeping the group strong, but some human cultures do (eg Russia).

Or what about market economics? The outer layer is companies, the inner layer is individuals. Maybe the individuals are workers – each worker would selfishly be best off if they spent the day watching YouTube videos and pushed the hard work onto someone else. Or maybe they're executives – each individual executive would selfishly be best off if they spent their energy on office politics, trying to flatter and network with whoever was most likely to promote them. But if all the employees loaf off and all the executives focus on office politics, the company won't make products, and competitors will eat their lunch. So someone – maybe the founder/CEO – comes up with a ruleset to incentivize good work, probably some kind of performance review system where people who do good work get promoted and people who do bad work get fired. The outer-layer competition between companies will select for corporations with the best rulesets; over time, companies' internal politics should get better at promoting the kind of cooperation necessary to succeed.

How do these systems replicate multicellular life's success without being literal entities with literal DNA having literal sex? They all involve a shared ruleset and a way of punishing rulebreakers which make it in each individual's short-term interest to follow the ruleset that leads to long-term success. Countries can do that (follow the law or we'll jail you), companies can do that (follow our policies or we'll fire you), even multicellular life can sort of do that (don't become cancer, or immune cells will kill you). When there's nothing like that (like the overly-fast-breeding foxes) evolution fails at group selection problems. When there is something like that, it has a chance. When there's something like that, and the thing like that is itself evolving (either because it’s encoded in literal DNA, or because it’s encoded in things like company policies that determine whether a company goes out of business or becomes a model for others), then it can reach a point where it solves group selection problems very effectively.

In the esoteric teachings, the inner layer of two-layer evolutionary systems is represented by the Goddess of Cancer, and outer layer by the Goddess of Everything Else. In each part of the poem, the Goddess of Cancer orders the evolving-entities to compete, but the Goddess of Everything Else recasts it as a two-layer competition where cooperation on the internal layer helps win the competition on the external layer. He who has ears to hear, let him listen.

III.

Why the digression? Because slack is a group selection problem. A species that gave itself slack in its evolutionary competition would do better than one that didn’t – for example, the eyeless aliens would evolve eyes and get a big fitness boost. But no individual can unilaterally choose to compete less intensely; if it did, it would be outcompeted and die. So one-layer evolution will fail at this problem the same way it fails all group selection problems, but two-layer systems will have a chance to escape the trap.

The multicellular life example above is a special case where you want 100% coordination and 0% competition. I framed the other examples the same way – countries do best when their citizens avoid all competition and work together for the common good, companies do best when their executives avoid self-aggrandizing office politics and focus on product quality. But as we saw above, some systems do best somewhere in the middle, where there’s some competition but also some slack.

For example, consider a researcher facing their own version of the eyeless aliens’ dilemma. They can keep going with business as normal – publishing trendy but ultimately useless papers that nobody will remember in ten years. Or they can work on Research Program Part 1, which *might* lead to Research Program Part 2, which *might* lead to Research Program Part 3, which *might* lead to a groundbreaking insight. If their jobs are up for review every year, and a year from now the business-as-normal researcher will have five trendy papers, and the groundbreaking-insight researcher will be halfway through Research Program Part 1, then the business-as-normal researcher will outcompete the groundbreaking-insight researcher; as the saying goes, “publish or perish”. Without slack, no researcher can unilaterally escape the system; their best option will always be to continue business as usual.

But group selection makes the situation less hopeless. Universities have long time-horizons and good incentives; they want to get famous for producing excellent research. Universities have rulesets that bind their individual researchers, for example “after a while good researchers get tenure”. And since universities compete with each other, each is incentivized to come up with the ruleset that maximizes long-term researcher productivity. So if tenure really does work better than constant vicious competition, then (absent the usual culprits like resistance-to-change, weird signaling equilibria, politics, etc) we should expect universities to converge on a tenure system in order to produce the best work. In fact, we should expect universities to evolve a really impressive ruleset for optimizing researcher incentives, just as impressive as the clever mechanisms the human body uses to prevent cancer (since this seems a bit optimistic, I assume the usual culprits are not absent).

The same is true for grant-writing; naively you would want some competition to make sure that only the best grant proposals get funded, but too much competition seems to stifle original research, so much so that some funders are throwing out the whole process and selecting grants by lottery, and others are running grants you can apply for in a half-hour and hear back about two days later. If there's a feedback mechanism – if these different rulesets produce different-quality research, and grant programs that produce higher-quality research are more likely to get funded in the future – then the rulesets for grants will gradually evolve, and the competition for grants will take place in an environment with whatever the right evolutionary parameters for evolving good research are.

I don't want to say these things will definitely happen – you can read Inadequate Equilibria for an idea of why not. But they *might*. The evolutionary dynamics which would normally prevent them can be overcome. Two-layer evolutionary systems can produce their own slack, if having slack would be a good idea.

IV.

That was a lot of paragraphs, and a lot of them started with "imagine a hypothetical situation where…". Let's look deeper into cases where an understanding of slack can inform how we think about real-world phenomena. Seven examples:

1. Monopolies. Not the kind that survive off overregulation and patents, the kind that survive by being big enough to crush competitors. These are predators that exploit low-slack environments. If Boeing has a monopoly on building passenger planes, and is exploiting that by making shoddy products and overcharging consumers, then that means

anyone else who built a giant airplane factory could make better products at a lower price, capture the whole airplane market, and become a zillionaire. Why don't they? Slack. In terms of those adaptive fitness landscapes, in between your current position (average Joe) and a much better position at the bottom of a deep pit (you own a giant airplane factory and are a zillionaire), there's a very big hill you have to climb – the part where you build Giant Airplane Factory Part 1, Giant Airplane Factory Part 2, etc. At each point in this hill, you are worse off than somebody who was not building an as-yet-unprofitable giant airplane factory. If you have infinite slack (maybe you are Jeff Bezos, have unlimited money, and will never go bankrupt no matter how much time and cost it takes before you start earning profits) you're fine. If you have more limited slack, your slack will run out and you'll be outcompeted before you make it to the greater-fitness deep pit.

Real monopolies are more complicated than this, because Boeing can shape up and cut prices when you're halfway to building your giant airplane factory, thus removing your incentive. Or they can do *actually* shady stuff. But none of this would matter if you already had your giant airplane factory fully built and ready to go – at worst, you and Boeing would then be in a fair fight. Everything Boeing does to try to prevent you from building that factory is exploiting your slacklessness and trying to increase the height of that hill you have to climb before the really deep pit.

(Peter Thiel inverts the landscape metaphor and calls the hill a "moat", but he's getting at the same concept).

2. Tariffs. Same story. Here's the way I understand the history of the international auto industry – anyone who knows more can correct me if I'm wrong. Automobiles were invented in the early 20th century. Several Western

countries developed homegrown auto industries more or less simultaneously, with the most impressive being Henry Ford’s work on mass production in the US. Post-WWII Japan realized that its own auto industry would never be able to compete with more established Western companies, so it placed high tariffs on foreign cars, giving local companies like Nissan and Toyota a chance to get their act together. These companies, especially Toyota, invented a new form of auto production which was actually much more efficient than the usual American methods, and were eventually able to hold their own. They started exporting cars to the US; although American tariffs put them at a disadvantage, they were so much better than the American cars of the time that consumers preferred them anyway. After decades of losing out, the American companies adopted a more Japanese ethos, and were eventually able to compete on a level playing field again.

This is a story of things gone surprisingly *right* – Americans and Japanese alike were able to get excellent inexpensive cars. Two things had to happen for it to work. First, Japan had to have high enough tariffs to give their companies some slack – to let them develop their own homegrown methods from scratch without being immediately outcompeted by temporarily-superior American competitors. Second, America had to have low enough tariffs that eventually-superior Japanese companies could outcompete American automakers, and Japan’s fitness-improving innovations could spread.

From the perspective of a Toyota manager, this is analogous to the eyeless alien story. You start with some good-enough standard (blind animals, American car companies). You want to evolve a superior end product (eye-having animals, Toyota). The intermediate steps (an animal with only Eye Part 1, a kind of crappy car company that stumbles over itself trying out new things) are less fit than the good-enough standard. Only when the inferior intermediate steps are protected from competition (through evolutionary randomness, through tariffs) can the superior end product come into existence. But you want to keep enough competition that the superior end product can use its superiority to spread (there is enough evolutionary competition that having eyes reaches fixation, there is enough free trade that Americans preferentially buy Toyota and US car companies have to adopt its policies).

From the perspective of an economic historian, maybe it’s a group selection story. The various stakeholders in the US auto industry – Ford, GM, suppliers, the government, labor, customers – competed with each other in a certain way and struck some compromise. The various stakeholders in the Japanese auto industry did the same. For some reason the American compromise worked worse than the Japanese one – I’ve heard stories about how US companies were more willing to defraud consumers for short-term profit, how US labor unions were more willing to demand concessions even at the cost of company efficiency, how regulators and executives were in bed with each other to the detriment of the product, etc. Every US interest group was acting in its own short-term self-interest, but the Japanese industry-as-a-whole outcompeted the American one and the Americans had to adjust.

3. Monopolies, Part II. Traditionally, monopolies have been among the most successful R&D centers. The most famous example is Xerox; it had a monopoly on photocopiers for a few decades before losing an anti-trust suit in the late 1970s; during that period, its PARC R&D program invented “laser printing, Ethernet, the modern personal computer, graphical user interface (GUI) and desktop paradigm, object-oriented programming, [and] the mouse”. The second most

famous example is Bell Labs, which invented “radio astronomy, the transistor, the laser, the photovoltaic cell, the charge-coupled device, information theory, the Unix operating system, and the programming languages B, C, C++, and S” before the government broke up its parent company AT&T. Google seems to be trying something similar, though it’s too soon to judge their outcomes.

These successes make sense. Research and development is a long-term gamble. Devoting more money to R&D decreases your near-term profits, but (hopefully) increases your future profits. Freed from competition, monopolies have limitless slack, and can afford to invest in projects that won’t pay off for ten or twenty years. This is part of Peter Thiel’s defense of monopolies in *Zero To One*.

An administrator tasked with advancing technology might be tempted to encourage monopolies in order to get more research done. But monopolies can also be stagnant and resistant to change; it’s probably not a coincidence that Xerox wasn’t the first company to bring the personal computer to market, and ended up irrelevant to the computing revolution. Like the eyeless aliens, who will not evolve in conditions of perfect competition or perfect lack of competition, probably all you can do here is strike a balance. Some Communist countries tried the extreme solution – one state-supported monopoly per industry – and it failed the test of group selection. I don’t know enough to have an opinion on whether countries with strong antitrust eventually outcompete those with weaker antitrust or vice versa.

4. Strategy Games. I like the strategy game *Civilization*, where you play as a group of primitives setting out to found a empire. You build cities and infrastructure, research

technologies, and fight wars. Your world is filled with several (usually 2 to 7) other civilizations trying to do the same.

Just like in the real world, civilizations must decide between Guns and Butter. The Civ version of Guns is called the Axe Rush. You immediately devote all your research to discovering how to make really good axes, all your industry to manufacturing those axes, and all your population into wielding those axes. Then you go and hack everyone else to pieces while they're still futzing about trying to invent pottery or something.

The Civ version of Butter is called Build. You devote all your research, industry, and populace to laying the foundations of a balanced economy and culture. You invent pottery and weaving and stuff like that. Soon you have a thriving trade network and a strong philosophical tradition. Eventually you can field larger and more advanced armies than your neighbors, and leverage the advantage into even more prosperity, or into military conquest.

Consider a very simple scenario: a map of Eurasia with two civilizations, Rome and China.

If both choose Axe Rush, then whoever Axe Rushes better wins.

If both choose Build, then whoever Builds better wins.

What if Rome chooses Axe Rush, and China chooses Build?

Then it depends on their distance! If it's a very small map and they start very close together, Rome will probably overwhelm the Chinese before Build starts paying off. But if it's a very big map, by the time Roman Axemen trek all the way to China, China will have Built high walls, discovered longbows and other defensive technologies, and generally become too strong for axes to defeat. Then they can crush the Romans – who are still just axe-wielding primitives – at their leisure.

Consider a more complicated scenario. You have a map of Earth. The Old World contains Rome and China. The New World contains Aztecs. Rome and China are very close to each other. Now what happens?

Rome and China spend the Stone, Bronze, and Iron Ages hacking each other to bits. Aztecs spend those Ages building cities, researching technologies, and building unique Wonders of the World that provide powerful bonuses. In 1492, they discover Galleons and start crossing the ocean. The powerful and advanced Aztec empire crushes the exhausted axe-wielding Romans and Chinese.

This is another story about slack. The Aztecs had it – they were under no competitive pressure to do things that paid off next turn. The Romans and Chinese didn't – they had to be at the top of their game every single turn, or their neighbor would conquer them. If there was an option that made you 10% weaker next turn in exchange for making you 100% stronger ten turns down the line, the Aztecs could take it without a second thought; the Romans and Chinese would probably have to pass.

Okay, more complicated *Civilization* scenario. This time there are two Old World civs, Rome and China, and two New World civs, Aztecs and Inca. The map is stretched a little bit so that all four civilizations have the same amount of natural territory. All four players understand the map layout and can communicate with each other. What happens?

Now it's a group selection problem. A skillful Rome player will private message the China player and explain all of this to her. She'll remind him that if one hemisphere spends the whole Stone Age fighting, and the other spends it building, the builders will win. She might tell him that she knows the Aztec and Inca players, they're smart, and they're going to be discussing the same considerations. So it would benefit both Rome and China to sign a peace treaty dividing the Old World in two, stick to their own side, and Build. If both sides cooperate, they'll both Build strong empires capable of matching the New World players. If one side cooperates and the other defects, it will easily steamroll over its unprepared opponent and conquer the whole Old World. If both sides defect, they'll hack each other to death with axes and be easy prey for the New Worlders.

This might be true in *Civilization* games, but real-world civilizations are more complicated. Orson Welles said:

> In Italy, for thirty years under the Borgias, they had warfare, terror, murder and bloodshed, but they produced Michelangelo, Leonardo da Vinci and the Renaissance. In Switzerland, they had brotherly love, they had five hundred years of democracy and peace – and what did that produce? The cuckoo clock.

So maybe a little bit of internal conflict is good, to keep you honest. Too much conflict, and you tear yourselves apart and are easy prey for outsiders. Too little conflict, and you invent the cuckoo clock and nothing else. The continent that conquers the world will have enough pressure that its people want to innovate, *and* enough slack that they're able to.

This is total ungrounded amateur historical speculation, but when I hear that I think of the Classical world. We can imagine it as divided into a certain number of "theaters of civilization" – Greece, Mesopotamia, Egypt, Persia, India, Scythia, etc. Each theater had its own rules governing average state size, the rules of engagement between states, how often bigger states conquered smaller states, how often ideas spread between states of the same size, etc. Some of those theaters were intensely competitive: Egypt was a nice straight line, very suited to centralized rule. Others had more slack: it was really hard to take over all of Greece; even the Spartans didn't manage. Each theater conducted its own "evolution" in its own way – Egypt was ruled by a single Pharaoh without much competition, Scythia was constant warfare of all against all, Greece was isolated city-states that fought each other sometimes but also had enough slack to develop philosophy and science. Each of those systems did their own thing for a while, until finally one of them produced something perfect: 4th century BC Macedonia. Then it went out and conquered everything.

If Welles is right, the point isn't to find the ruleset that promotes 100% cooperation. It's to find the ruleset that promotes an evolutionary system that makes your group the strongest. Usually this involves some amount of competition – in order to select for stronger organisms – but also some amount of slack – to let organisms develop complicated strategies that can make them stronger. Despite the earlier description, this isn't necessarily a slider between 0% competition and 100% competition. It could be much more complicated – maybe alternating high-slack vs. low-slack periods, or many semi-isolated populations with a small chance of interaction each generation, or alternation between periods of isolation and periods of churning.

In a full two-layer evolution, you would let the systems evolve until they reached the best parameters. Here we can't do that – Greece has however many mountains it has; its success does not cause the rest of the world to grow more mountains. Still, we randomly started with enough different groups that we got to learn something interesting.

(I can’t emphasize enough how ungrounded this historical speculation is. Please don’t try to evolve Alexander the Great in your basement and then get angry at me when it doesn’t work)

5. The Long-Term Stock Exchange. Actually, all stock exchanges are about slack. Imagine you are a brilliant inventor who, given $10 million and ten years, could invent fusion power. But in fact you have $10 and need work tomorrow or you will starve. Given those constraints, maybe you could start, I don’t know, a lemonade stand.

You’re in the same position as the animal trying to evolve an eye – you could create something very high-utility, if only you had enough slack to make it happen. But by default, the inventor working on fusion power starves to death tomorrow (or at least makes less money than his counterpart who ran the lemonade stand), the same way the animal who evolves Eye Part 1 gets outcompeted by other animals who didn’t and dies out.

You need slack. In the evolution example, animals usually stumble across slack randomly. You too might stumble across slack randomly – maybe it so happens that you are independently wealthy, or won the lottery, or something.

More likely, you use the investment system. You ask rich people to give you $10 million for ten years so you can invent fusion; once you do, you’ll make trillions of dollars and share some of it with them.

This is a great system. There’s no evolutionary equivalent. An animal can’t pitch Darwin on its three-step plan to evolve eyes and get free food and mating opportunities to make it happen. Wall Street is a giant multi-trillion dollar time machine funneling future profits back into the past, and that gives people the slack they need to make the future profits happen at all.

But the Long-Term Stock Exchange is especially about slack. They are a new exchange (approved by the SEC last year) which has complicated rules about who can list with them. Investors will get extra clout by agreeing to hold stocks for a long time; executives will get incentivized to do well in the far future instead of at the next quarterly earnings report. It's making a deliberate choice to give companies more slack than the regular system and see what they do with it. I don't know enough about investing to have an opinion, except that I appreciate the experiment. Presumably its companies will do better/worse than companies on the regular stock exchange, that will cause companies to flock toward/away from it, and we'll learn that its new ruleset is better/worse at evolving good companies through competition than the regular stock exchange's ruleset.

6. That Time Ayn Rand Destroyed Sears. Or at least that's how Michael Rozworski and Leigh Phillips describe Eddie Lampert's corporate reorganization in How Ayn Rand Destroyed Sears, which I recommend. Lampert was a Sears CEO who figured – since free-market competitive economies outcompete top-down economies, shouldn't free-market competitive companies outcompete top-down companies? He reorganized Sears as a set of competing departments that traded with each other on normal free-market principles; if the Product Department wanted its products marketed, it would have to pay the Marketing Department. This worked really badly, and was one of the main contributors to Sears' implosion.

I don’t have a great understanding of exactly why Lampert’s Sears lost to other companies even though capitalist economies beat socialist ones; Rozworski and Phillips’ People’s Republic Of Wal-Mart, which looks into this question, is somewhere on my reading list. But even without complete understanding, we can use group selection to evolve the right parameters. Imagine an economy with several businesses. One is a straw-man communist collective, where every worker gets paid the same regardless of output and there are no promotions (0% competition, 100% cooperation). Another is Lampert’s Sears (100% competition, 0% cooperation). Others are normal businesses, where employees mostly work together for the good of the company but also compete for promotions (X% competition, Y% cooperation). Presumably the normal business outcompetes both Lampert and the commies, and we sigh with relief and continue having normal businesses. And if some of the normal businesses outcompete others, we’ve learned something about the best values of X and Y.

7. Ideas. These are in constant evolutionary competition – this is the insight behind memetics. The memetic equivalent of slack is inferential range, aka “willingness to entertain and explore ideas before deciding that they are wrong”.

Inferential distance is the number of steps it takes to make someone understand and accept a certain idea. Sometimes inferential distances can be very far apart. Imagine trying to convince a 12th century monk that there was no historical Exodus from Egypt. You’re in the middle of going over archaeological evidence when he objects that the Bible says there was. You respond that the Bible is false and there’s no God. He says that doesn’t make sense, how would life have originated? You say it evolved from single-celled organisms. He asks how evolution, which seems to be a change in animals’ accidents, could ever affect their essences and change them into an entirely new species. You say that the whole scholastic worldview is wrong, there’s no such thing as accidents and essences, it’s just atoms and empty space. He asks how you ground morality if not in a striving to approximate the ideal embodied by your essence, you say… well, it doesn’t matter what you say, because you were trying to convince him that some very specific people didn’t leave Egypt one time, and now you’ve got to ground morality.

Another way of thinking about this is that there are two self-consistent equilibria. There’s your equilibrium, (no Exodus, atheism, evolution, atomism, moral nonrealism), and the monk’s equilibrium (yes Exodus, theism, creationism, scholasticism, teleology), and before you can make the monk budge on any of those points, you have to convince him of all of them.

So the question becomes – how much patience does this monk have? If you tell him there’s no God, does he say “I look forward to the several years of careful study of your scientific and philosophical theories that it will take for that statement not to seem obviously wrong and contradicted by every other feature of the world”? Or does he say “KILL THE UNBELIEVER”? This is inferential range.

Aristotle supposedly said that the mark of an educated man is to be able to entertain an idea without accepting it. Inferential range explains why. The monk certainly shouldn’t immediately accept your claim, when he has countless pieces of evidence for the existence of God, from the spectacular faith healings he has witnessed (“look, there’s this thing called psychosomatic illness, and it’s really susceptible to this other thing called the placebo effect…”) to Constantine’s victory at the Mulvian Bridge despite being heavily outnumbered (“look, I’m not a classical scholar, but some people are just really good generals and get lucky, and sometimes it happens the day after they have weird dreams, I think there's enough good evidence the other way that this is not the sort of thing you should center your worldview around"). But if he's willing to entertain your claim long enough to hear your arguments one by one, eventually he can reach the same self-consistent equilibrium you're at and judge for himself.

Nowadays we don't burn people at the stake. But we do make fun of them, or flame them, or block them, or wander off, or otherwise not listen with an open mind to ideas that strike us at first as stupid. This is another case where we have to balance competition vs. slack. With perfect competition, the monk instantly rejects our "no Exodus" idea as less true (less memetically fit) than its competitors, and it has no chance to grow on him. With zero competition, the monk doesn't believe anything at all, or spends hours patiently listening to someone explain their world-is-flat theory. Good epistemics require a balance between being willing to choose better ideas over worse ones, and open-mindedly hearing the worse ones out in case they grow on you.

(Thomas Kuhn points out that early versions of the heliocentric model were much worse than the geocentric model, that astronomers only kept working on them out of a sort of weird curiosity, and that it took decades before they could clearly hold their own against geocentrism in a debate).

Different people strike a different balance in this space, and those different people succeed or fail based on their own epistemic ruleset. Someone who's completely closed-minded and dogmatic probably won't succeed in business, or science, or the military, or any other career (except maybe politics). But someone who’s so pathologically open-minded that they listen to everything and refuse to prioritize what is or isn’t worth their time will also fail. We take notice of who succeeds or fails and change our behavior accordingly.

Maybe there’s even a third layer of selection; maybe different communities are more or less willing to tolerate open-minded vs. close-minded people. The Slate Star Codex community has really different epistemic norms from the Catholic Church or Infowars listeners; these are evolutionary parameters that determine which ideas are more memetically fit. If our epistemics make us more likely to converge on useful (not necessarily true!) ideas, we will succeed and our epistemic norms will catch on. Francis Bacon was just some guy with really good epistemic norms, and now everybody who wants to be taken seriously has to use his norms instead of whatever they were doing before. Come up with the right evolutionary parameters, and that could be you!

The Categories Were Made For Man, Not Man For The Categories

I.

“Silliest internet atheist argument” is a hotly contested title, but I have a special place in my heart for the people who occasionally try to prove Biblical fallibility by pointing out whales are not a type of fish.

(this is going to end up being a metaphor for something, so bear with me)

The argument goes like this. Jonah got swallowed by a whale. But the Bible says Jonah got swallowed by a big fish. So the Bible seems to think whales are just big fish. Therefore the Bible is fallible. Therefore, the Bible was not written by God.

The first problem here is that “whale” is just our own modern interpretation of the Bible. For all we know, Jonah was swallowed by a really really really big herring.

The second problem is that if the ancient Hebrews want to call whales a kind of fish, let them call whales a kind of fish.

I’m not making the weak and boring claim that since they’d never discovered genetics they don’t know better. I am making the much stronger claim that, even if the ancient Hebrews had taken enough of a break from murdering Philistines and building tabernacles to sequence the genomes of all known species of aquatic animals, there’s nothing whatsoever wrong, false, or incorrect with them calling a whale a fish.

Now, there’s something wrong with saying “whales are phylogenetically just as closely related to bass, herring, and salmon as these three are related to each other.” What’s wrong with the statement is that it’s false. But saying “whales are a kind of fish” isn’t.

Suppose you travel back in time to ancient Israel and try to explain to King Solomon that whales are a kind of mammal and not a kind of fish.

Your translator isn’t very good, so you pause to explain “fish” and “mammal” to Solomon. You tell him that fish is “the sort of thing herring, bass, and salmon are” and mammal is “the sort of thing cows, sheep, and pigs are”. Solomon tells you that your word “fish” is Hebrew *dag* and your word “mammal” is Hebrew *behemah*.

So you try again and say that a whale is a *behemah*, not a *dag*. Solomon laughs at you and says you’re an idiot.

You explain that you’re not an idiot, that in fact all kinds of animals have things called genes, and the genes of a whale are much closer to those of the other *behemah* than those of the *dag*.

Solomon says he’s never heard of these gene things before, and that maybe genetics is involved in your weird foreign words “fish” and “mammal”, but *dag* are just finned creatures that swim in the sea, and *behemah* are just legged creatures that walk on the Earth.

You try to explain that no, Solomon is wrong, *dag* are actually defined not by their swimming-in-sea-with-fins-ness, but by their genes.

Solomon says you didn’t even *know* the word *dag* ten minutes ago, and now suddenly you think you know what it means better than he does, who has been using it his entire life? Who died and made *you* an expert on Biblical Hebrew?

You try to explain that whales actually have tiny little hairs, too small to even see, just as cows and sheep and pigs have hair.

Solomon says oh God, you are so annoying, who the hell cares whether whales have tiny little hairs or not. In fact, the only thing Solomon cares about is whether responsibilities for his kingdom's production of blubber and whale oil should go under his Ministry of Dag or Ministry of Behemah. The Ministry of Dag is based on the coast and has a lot of people who work on ships. The Ministry of Behemah has a strong presence inland and lots of of people who hunt on horseback. So please (he continues) keep going about how whales have little tiny hairs.

It's easy to see that Solomon has a point, and that if he wants to define *behemah* as four-legged-land-dwellers that's his right, and no better or worse than your definition of "creatures in a certain part of the phylogenetic tree". Indeed, it might even be that if you spent ten years teaching Solomon all about the theory of genetics and evolution (which would be hilarious – think how annoyed the creationists would get) he might still say "That's very interesting, and I can see why we need a word to describe creatures closely related along the phylogenetic tree, but make up your own word, because *behemah* already means 'four-legged-land-dweller'."

Now imagine that instead of talking to King Solomon, you're talking to that guy from Duck Dynasty with the really crazy beard (I realize that may describe more than one person), who stands in for all uneducated rednecks in the same way King Solomon stands in for all Biblical Hebrews.

“Ah course a whale is a feesh, ya moron” he says in his heavy Southern accent.

“No it isn’t,” you say. “A fish is a creature phylogenetically related to various other fish, and with certain defining anatomical features. It says so right here in this biology textbook.”

“Well,” Crazy Beard Guy tells you, “Ah reckon that might be what a fish is, but a *feesh* is some’in that swims in the orshun.”

With a sinking feeling in your stomach, you spend ten years turning Crazy Beard Guy into a world expert on phylogenetics and evolutionary theory. Although the Duck Dynasty show becomes *much* more interesting, you fail to budge him a bit on the meaning of “feesh”.

It’s easy to see here that “fish” and “feesh” can be different just as “fish” and “*dag*” can be different.

You can point out how many important professors of icthyology in fancy suits use your definition, and how only a couple of people with really weird facial hair use his. But now you’re making a status argument, not a factual argument. Your argument is “conform to the way all the cool people use the word ‘fish'”, not “a whale is really and truly not a fish”.

There are facts of the matter on each individual point – whether a whale has fins, whether a whale lives in the ocean, whether a whale has tiny hairs, et cetera. But there is no fact of the matter on whether a whale is a fish. The argument is entirely semantic.

So this is the second reason why this particular objection to the Bible is silly. If God wants to call a whale a big fish, stop telling God what to do.

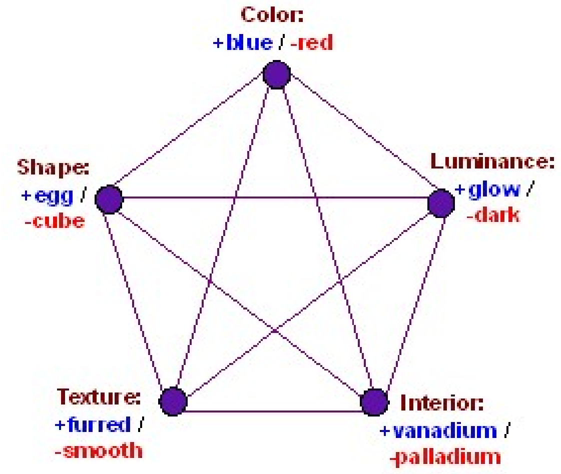
(also, bats)

II.

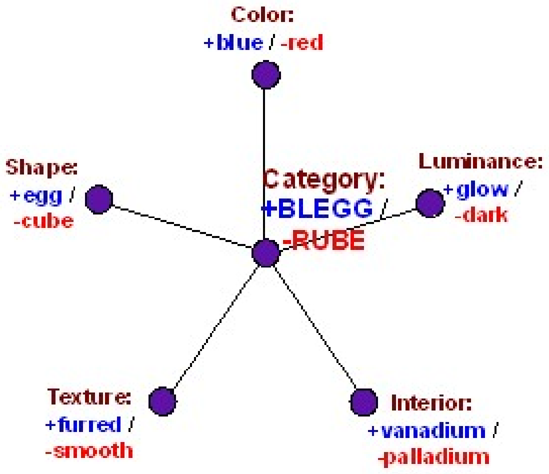
When terms are *not* defined directly by God, we need our own methods of dividing them into categories.

Less Wrong classic How An Algorithm Feels From The Inside starts with a discussion of whether or not Pluto is a planet. Planets tend to share many characteristics in common. For example, they are large, round, have normal shaped orbits lined up with the plane of the ecliptic, have cleared out a certain area of space, and are at least kind of close to the Sun as opposed to way out in the Oort Cloud.

One could imagine a brain that thought about these characteristics like Network 1 here:


Color:
+blue / -red
Shape:
+egg / -cube
Luminance:
+glow / -dark
Texture:
+furred / -smooth
Interior:
+vanadium / -palladium

Network 1


Color:
+blue / -red
Shape:
+egg / -cube
Category:
+BLEGG / -RUBE
Luminance:
+glow / -dark
Texture:
+furred / -smooth
Interior:
+vanadium / -palladium

Network 2

One could imagine this model telling you everything you need to know. If an object is larger, it’s more likely to be round and in cis-Neptunian space. If an object has failed to clear its orbit of debris, it’s more likely to have a skewed orbit relative to the plane of the ecliptic. We could give each of these connections a weight and say things like large objects have a 32% chance of being in cis-Neptunian space and small objects an 86% chance. Or whatever.

But Network 1 has some big problems. For one thing, if you inscribe it in blood, you might accidentally summon the Devil. But for another, it’s computationally very complicated. Each attribute affects each other attribute which affects it in turn and so on in an infinite cycle, so that its behavior tends to be chaotic and unpredictable.

What people actually seem to do is more like Network 2: sweep all common correlations into one big category in the middle, thus dividing possibility-space into large round normal-orbit solitary inner objects, and small irregular skewed-orbit crowded outer objects. It calls the first category “planets” and the second category “planetoids”.

You can then sweep minor irregularities under the rug. Neptune is pretty far from the sun, but since it’s large, round, normal-orbit, and solitary, we know which way the evidence is leaning.

When an object satisfies about half the criteria for planet and half the criteria for planetoid, *then* it’s awkward. Pluto is the classic example. It’s relatively large, round, skewed orbit, solitary…ish? and outer-ish. What do you do?

The *practical* answer is you convene some very expensive meeting of prestigious astronomers and come to some official decision which everyone agrees to follow so they’re all on the same page.

But the *ideal* answer is you say “Huh, the assumption encoded in the word ‘planet’ that the five red criteria always went together and the five blue criteria always went together doesn’t hold. Whatever.”

Then you divide the solar system into three types of objects: planets, planetoids, and dammit-our-categorization-scheme-wasn't-as-good-as-we-thought.

The situation with whales and fish is properly understood in the same context. Fish and mammals differ on a lot of axes. Fish generally live in the water, breathe through gills, have tails and fins, possess a certain hydrodynamic shape, lay eggs, and are in a certain part of the phylogenetic tree. Mammals generally live on land, breathe through lungs, have legs, give live birth, and are in another part of the phylogenetic tree. Most fish conform to all of the fish desiderata, and most mammals conform to all of the mammal desiderata, so there's no question of how to categorize them. Occasionally you get something weird (a platypus, a lungfish, or a whale) and it's a judgment call which you have to decide by fiat. In our case, that fiat is "use genetics and ignore all other characteristics" but some other language, culture, or scientific community might make a different fiat, and then the borders between their categories would look a little bit different.

III.

Since I shifted to a borders metaphor, let's follow that and see where it goes.

Imagine that Israel and Palestine agree to a two-state solution with the final boundary to be drawn by the United Nations. You're the head of the United Nations committee involved, so you get out a map and a pencil. Both sides swear to follow whatever you determine.

Your job is not to draw "the correct border". There is no one correct border between Israel and Palestine. There are a couple of very strong candidates (for example, the pre-1967 line of control), but both countries have suggested deviations from that (most people think an actual solution would involve Palestine giving up some territory that has since been thoroughly settled by Israel in exchange for some territory within Israel proper, or perhaps for a continuous "land bridge" between the West Bank and Gaza). Even if you wanted to use the pre-1967 line as a starting point, there would still be a lot of work to do deciding what land swaps should and shouldn't be made.

Instead you'd be making a series of trade-offs. Giving all of Jerusalem to the Israelis would make them very happy but anger Palestine. Creating a contiguous corridor between Gaza and the West Bank makes some sense, but then you'd be cutting off Eilat from the rest of Israel. Giving all of the Israeli settlements in the West Bank back to Palestine would satisfy a certain conception of property rights, but also leave a lot of Jews homeless.

There are also much stupider decisions you could make. You could give Tel Aviv to Palestine. You could make the Palestinian state a perfect circle five miles in radius centered on Rishon LeZion. You could just split the territory in half with a straight line, and give Israel the north and Palestine the south. All of these things would be really dumb.

But, crucially, they would not be *false*. They would not be *factually incorrect*. They would just be failing to achieve pretty much any of the goals that we would expect a person solving land disputes in the Middle East to have. You can think of alternative arrangements in which these wouldn't be dumb. For example, if you're a despot, and you want to make it very clear to both the Israelis and Palestinians that their opinions don't matter and they should stop bothering you with annoying requests for arbitration, maybe splitting the country in half north-south is the way to go.

And real borders are, in fact, very weird.

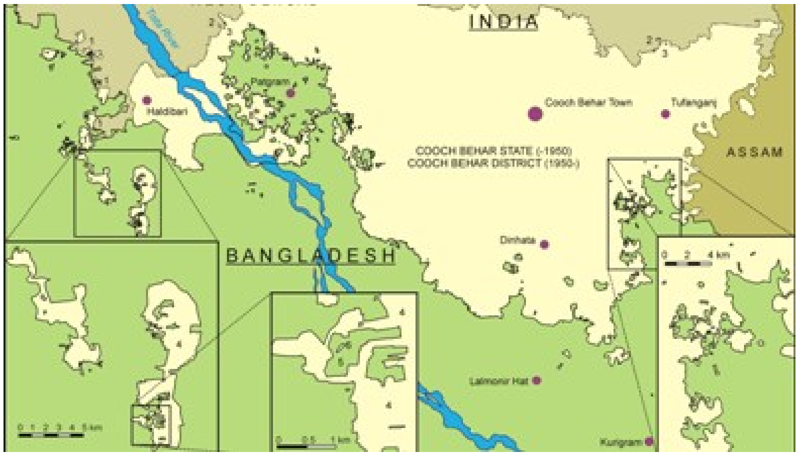
The border between Turkey and Syria follows a mostly straight-ish line near-ish the 36th parallel, except that about twenty miles south of the border Turkey controls a couple of square meters in the middle of a Syrian village. This is the tomb of the ancestor of the Ottoman Turks, and Turkey's border agreement with Syria stipulates that it will remain part of Turkey forever. And the Turks take this *very* seriously; they maintain a platoon of special forces there and have recently been threatening war against Syria if their "territory" gets "invaded" in the current conflict.

Pictured: Turkey (inside fence), Syria (outside)

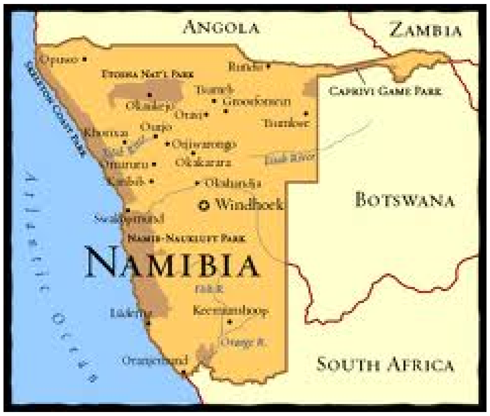
The border between Bangladesh and India is complicated at the best of times, but it becomes absolutely ridiculous in a place called Cooch-Behar, which I guess is as good a name as any for a place full of ridiculous things. In at least one spot there is an 'island' of Indian territory within a larger island of Bangladeshi territory within a larger island of Indian territory within Bangladesh. According to mentalfloss.com:

> So why'd the border get drawn like that? It can all be traced back to power struggles between local kings hundreds of years ago, who would try to claim pockets of

land inside each other’s territories as a way to leverage political power. When Bangladesh became independent from India in 1947 (as East Pakistan until 1971), all those separate pockets of land were divvied up. Hence the polka-dotted mess.


INDIA
Patgram
Haldibari
Cooch Behar Town
Tufanganj
ASSAM
COOCH BEHAR STATE (-1950)
COOCH BEHAR DISTRICT (1950-)
Dinhata
BANGLADESH
Lalmonir Hat
Kurigram
0 1 2 3 4 5 km
0 0.5 1 km
0 2 4 km

Namibia is a very weird-looking country with a very thin three-hundred-mile-long panhandle (eg about twice as long as Oklahoma’s). Apparently during the Scramble For Africa, the Germans who colonized Namibia really wanted access to the Zambezi River so they could reach the Indian Ocean and trade their colonial resources. They kept pestering the British who colonized Botswana until the Brits finally agreed to give up a tiny but very long strip of territory ending at the riverbank. This turned out to be not so useful, as *just* after Namibia’s Zambezi access sits Victoria Falls, the largest waterfall in the world – meaning that any Germans who tried to traverse the Zambezi to reach the Indian Ocean would last a matter of minutes before suddenly encountering a four hundred foot drop and falling to pretty much certain death. The moral of the story is not to pester the British Empire too much, especially if they’ve explored Africa and you haven’t.


ANGOLA
ZAMBIA
Opuwo
Etosha Nat'l Park
Rundu
Tsumeb
Grootfontein
Okaukejo
Otavi
Caprivi Game Park
Tsumkwe
Outjo
Khorixas
Otjiwarongo
Okakarara
Omaruru
Karibib
Okahandja
Windhoek
BOTSWANA
Swakopmund
Namib-Naukluft Park
NAMIBIA
Lüderitz
Keetmanshoop
Orange R.
Oranjemund
SOUTH AFRICA
Skeleton Coast Park
Atlantic Ocean

But the other moral of the story is that borders are weird. Although we think of borders as nice straight lines that separate people of different cultures, they can form giant panhandles, distant islands, and enclaves-within-enclaves-within-enclaves. They can depart from their usual course to pay honor to national founders, to preserve records of ancient conquests, or to connect to trade routes.

Hume’s ethics restrict “bad” to an instrumental criticism – you can condemn something as a bad way to achieve a certain goal, but not as morally bad independent of what the goal is. In the same way, borders can be bad at fulfilling your goals in drawing them, but not bad in an absolute sense or factually incorrect. Namibia’s border is bad from the perspective of Germans who want access to the Indian Ocean. But it’s *excellent* from the perspective of Englishmen who want to watch Germans plummet into the Lower Zambezi and get eaten by hippos.

Breaking out of the metaphor, the same is true of conceptual boundaries. You *may* draw the boundaries of the category “fish” any way you want. A category “fish” containing herring, dragonflies, and asteroids is going to be stupid, but only in the same sense that a Palestinian state centered around Tel Aviv would be stupid – it fails to fulfill any conceivable goals of the person designing it. Categories “fish” that do or don’t include whales may be appropriate for different people’s purposes, the same way Palestinians might argue about whether the borders of their state should be optimized for military defensibility or for religious/cultural significance.

Statements like “the Zambezi River is full of angry hippos” are brute facts. Statements like “the Zambezi River is the territory of Namibia” are negotiable.

In the same way, statements like “whales have little hairs” are brute facts. Statements like “whales are not a kind of fish” are negotiable.

So it’s important to keep these two sorts of statements separate, and remember that in no case can an agreed-upon set of borders or a category boundary be factually incorrect.

IV.

There is an anti-transgender argument that I take very seriously. The argument goes: we are rationalists. Our *entire shtick* is trying to believe what’s actually true, not on what we wish were true, or what our culture tells us is true, or what it’s popular to say is true. If a man thinks he’s a woman, then we might (empathetically) wish he were a woman, other people might demand we call him a woman, and we might be much more popular if we say he’s a woman. But if we’re going to be rationalists who focus on believing what’s actually true, then we’ve got to call him a man and take the consequences.

Thus Abraham Lincoln’s famous riddle: “If you call a tail a leg, how many legs does a dog have?” And the answer: “Four – because a tail isn’t a leg regardless of what you call it.”

(if John Wilkes Booth had to suffer through that riddle, then I don’t blame him)

I take this argument seriously, because sticking to the truth really is important. But having taken it seriously, I think it’s seriously wrong.

An alternative categorization system is not an error, and borders are not objectively true or false.

Just as we can come up with criteria for a definition of “planet”, we can come up with a definition of “man”. Absolutely typical men have Y chromosomes, have male genitalia, appreciate manly things like sports and lumberjackery, are romantically attracted to women, personally identify as male, wear male clothing like blue jeans, sing baritone in the opera, et cetera.

Some people satisfy some criteria of manhood and not others, in much the same way that Pluto satisfies only some criteria of planethood and whales satisfy only some criteria of mammalhood. For example, gay men might date other men and behave in effeminate ways. People with congenital androgen insensitivity syndrome might have female bodies, female external genitalia, and have been raised female their entire life, but when you look into their cells they have Y chromosomes.

Biologists defined by fiat that in cases of ambiguous animal grouping like whales, phylogenetics will be the tiebreaker. This was useful to resolve ambiguity, and it’s worth sticking to as a Schelling point so everyone’s using their words the same way, but it’s kind of arbitrary and mostly based on biologists caring a lot about phylogenetics. If we let King Solomon make the decision, he might decide by fiat that whether animals lived in land or water would be the tiebreaker, since he’s most interested in whether the animal is hunted on horseback or by boat.

Likewise, astronomers decided by fiat that something would be a planet if and only if meets the three criteria of orbiting, round, and orbit-clearing. But here we have a pretty neat window into how these kinds of decisions take place – you can read the history of the International Astronomical Union meeting where they settled on the definition and learn about all the alternative proposals that were floated and rejected and which particular politics resulted in the present criteria being selected among all the different possibilities. Here it is *obvious* that the decision was by fiat.

Without the input of any prestigious astronomers at all, most people seem to assume that the ultimate tiebreaker in man vs. woman questions is presence of a Y chromosome. I’m not sure this is a very principled decision, because I expect most people would classify congenital androgen insensitivity patients (XY people whose bodies are insensitive to the hormone that makes them look male, and so end up looking 100% female their entire lives and often not even knowing they have the condition) as women.

The project of the transgender movement is to propose a switch from using chromosomes as a tiebreaker to using self-identification as a tiebreaker.

(This isn’t actually the whole story – some of the more sophisticated people want to split “sex” and “gender”, so that people who want to talk about what chromosomes they’ve got have a categorization system to do that with, and a few people even want to split “chromosomal sex” and “anatomical sex” and “gender” and goodness knows what else – and I support all of these as very important examples of the virtue of precision – but to a first approximation, they want to define gender as self-identification)

This is not something that can be “true” or “false”. It’s a boundary-redrawing project. It can make for some boundaries that look a little bit weird – like a small percent of men being able to get pregnant – but as far as weird boundaries go that’s probably not as bad as having a tiny exclave of Turkish territory in the middle of a Syrian village.

You draw category boundaries in specific ways to capture tradeoffs you care about. If you care about the sanctity of the tomb of your country’s founder, sometimes it’s worth having a slightly weird-looking boundary in order to protect and honor it. And if you care about…

I’m writing this post today because I just finished accepting a transgender man to the mental hospital. He alternates between trying to kill himself and trying to cut off various parts of his body because he’s so distressed that he is biologically female. We’ve connected him with some endocrinologists who can hopefully get him started on male hormones, after which maybe he’ll stop doing that and hopefully be able to lead a normal life.

If I’m willing to accept an unexpected chunk of Turkey deep inside Syrian territory to honor some random dead guy – and I better, or else a platoon of Turkish special forces will want to have a word with me – then I ought to accept an unexpected man or two deep inside the conceptual boundaries of what would normally be considered female if it’ll save someone’s life. There’s no rule of rationality saying that I shouldn’t, and there are plenty of rules of human decency saying that I should.

V.

I’ve made this argument before and gotten a reply something like this:

“Transgender is a psychiatric disorder. When people have psychiatric disorders, certainly it’s right to sympathize and feel sorry for them and want to help them. But the way we try to help them is by treating their disorder, not by indulging them in their delusion."

I think these people expect me to argue that transgender "isn't really a psychiatric disorder" or something. But "psychiatric disorder" is just another category boundary dispute, and one that I've already written enough about elsewhere. At this point, I don't care enough to say much more than "If it's a psychiatric disorder, then attempts to help transgender people get covered by health insurance, and most of the transgender people I know seem to want that, so sure, gender dysphoria is a psychiatric disorder."

And then I think of the Hair Dryer Incident.

The Hair Dryer Incident was probably the biggest dispute I've seen in the mental hospital where I work. Most of the time all the psychiatrists get along and have pretty much the same opinion about important things, but people were at each other's *throats* about the Hair Dryer Incident.

Basically, this one obsessive compulsive woman would drive to work every morning and worry she had left the hair dryer on and it was going to burn down her house. So she'd drive back home to check that the hair dryer was off, then drive back to work, then worry that maybe she hadn't *really* checked well enough, then drive back, and so on ten or twenty times a day.

It's a pretty typical case of obsessive-compulsive disorder, but it was really interfering with her life. She worked some high-powered job – I think a lawyer – and she was *constantly* late to everything because of this driving back and forth, to the point where her career was in a downspin and she thought she would have to quit and go on disability. She wasn't able to go out with friends, she wasn't even able to go to restaurants because she would keep fretting she left the hair dryer on at home and have to rush back. She’d seen countless psychiatrists, psychologists, and counselors, she’d done all sorts of therapy, she’d taken every medication in the book, and none of them had helped.

So she came to my hospital and was seen by a colleague of mine, who told her “Hey, have you thought about just bringing the hair dryer with you?”

And it *worked*.

She would be driving to work in the morning, and she’d start worrying she’d left the hair dryer on and it was going to burn down her house, and so she’d look at the seat next to her, and there would be the hair dryer, right there. And she only had the one hair dryer, which was now accounted for. So she would let out a sigh of relief and keep driving to work.

And approximately half the psychiatrists at my hospital thought this was *absolutely scandalous*, and This Is Not How One Treats Obsessive Compulsive Disorder, and what if it got out to the broader psychiatric community that instead of giving all of these high-tech medications and sophisticated therapies we were just telling people to *put their hair dryers on the front seat of their car*?

But I think the guy deserved a medal. Here’s someone who was totally untreatable by the normal methods, with a debilitating condition, and a drop-dead simple intervention that nobody else had thought of gave her her life back. If one day I open up my own psychiatric practice, I am half-seriously considering using a picture of a hair dryer as the logo, just to let everyone know where I stand on this issue.

Miyamoto Musashi is quoted as saying:

> The primary thing when you take a sword in your hands is your intention to cut the enemy, whatever the means. Whenever you parry, hit, spring, strike or touch the enemy's cutting sword, you must cut the enemy in the same movement. It is essential to attain this. If you think only of hitting, springing, striking or touching the enemy, you will not be able actually to cut him.

Likewise, the primary thing in psychiatry is to help the patient, whatever the means. Someone can concern-troll that the hair dryer technique leaves something to be desired in that it might have prevented the patient from seeking a more thorough cure that would prevent her from having to bring the hair dryer with her. But compared to the alternative of "nothing else works" it seems clearly superior.

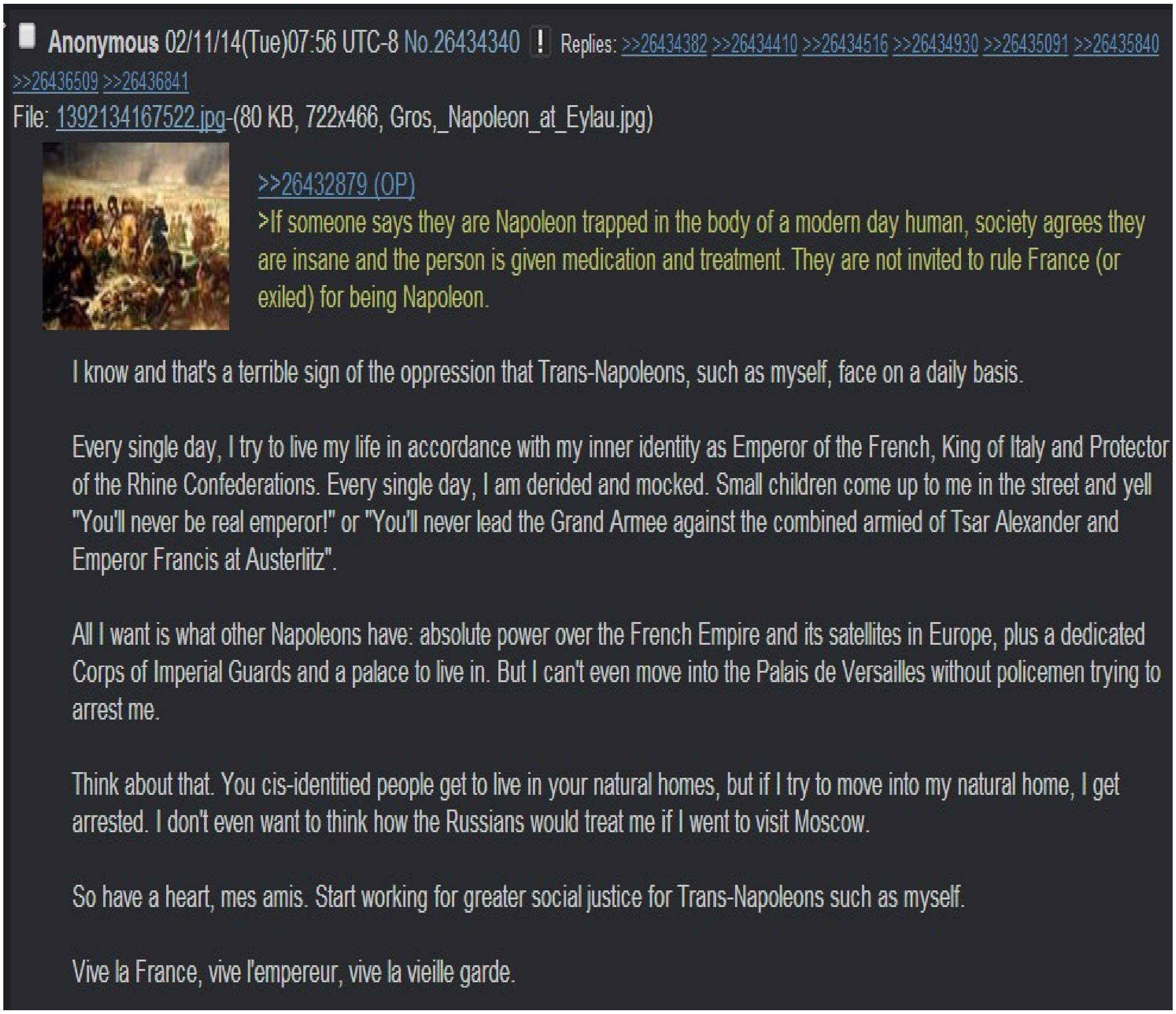
And that's the position from which I think a psychiatrist should approach gender dysphoria, too.

Imagine if we could give depressed people a much higher quality of life merely by giving them cheap natural hormones. I don't think there's a psychiatrist in the world who wouldn't celebrate that as one of the biggest mental health advances in a generation. Imagine if we could ameliorate schizophrenia with one safe simple surgery, just snip snip you're not schizophrenic anymore. Pretty sure that would win *all* of the Nobel prizes. Imagine that we could make a serious dent in bipolar disorder just by calling people different pronouns. I'm pretty sure the entire mental health field would join together in bludgeoning anybody who refused to do that. We would bludgeon them over the head with big books about the side effects of lithium.

Really, are you *sure* you want your opposition to accepting transgender people to be "I think it's a mental disorder"?

VI.

Some people can’t leave well enough alone, and continue to push the mental disorder angle. For example:



Anonymous 02/11/14(Tue)07:56 UTC-8 No.26434340 ! Replies: >>26434382 >>26434410 >>26434516 >>26434930 >>26435091 >>26435840 >>26436509 >>26436841

File: 1392134167522.jpg-(80 KB, 722x466, Gros,_Napoleon_at_Eylau.jpg)

>>26432879 (OP)

>If someone says they are Napoleon trapped in the body of a modern day human, society agrees they are insane and the person is given medication and treatment. They are not invited to rule France (or exiled) for being Napoleon.

I know and that's a terrible sign of the oppression that Trans-Napoleons, such as myself, face on a daily basis.

Every single day, I try to live my life in accordance with my inner identity as Emperor of the French, King of Italy and Protector of the Rhine Confederations. Every single day, I am derided and mocked. Small children come up to me in the street and yell "You'll never be real emperor!" or "You'll never lead the Grand Armee against the combined armied of Tsar Alexander and Emperor Francis at Austerlitz".

All I want is what other Napoleons have: absolute power over the French Empire and its satellites in Europe, plus a dedicated Corps of Imperial Guards and a palace to live in. But I can't even move into the Palais de Versailles without policemen trying to arrest me.

Think about that. You cis-identitied people get to live in your natural homes, but if I try to move into my natural home, I get arrested. I don't even want to think how the Russians would treat me if I went to visit Moscow.

So have a heart, mes amis. Start working for greater social justice for Trans-Napoleons such as myself.

Vive la France, vive l'empereur, vive la vieille garde.

There are a lot of things I could say here.

I could point out that trans-Napoleonism seem to be mysteriously less common than transgender.

I could relate this mysterious difference to the various heavily researched apparent biological correlates of transgender, including unusual variants of the androgen receptor, birth-sex-discordant sizes of various brain regions, birth-sex-discordant responses to various pheromones, high rates of something seemingly like body integrity identity disorder, and of course our old friend altered digit ratios. If our hypothetical trans-Napoleon came out of the womb wearing a French military uniform and clutching a list of 19th century Grand Armee positions in his cute little baby hands, I think I'd take him more seriously.

I could argue that questions about gender are questions about category boundaries, whereas questions about Napoleon – absent some kind of philosophical legwork that I would very much like to read – are questions of fact.

I could point out that if the extent of somebody's trans-Napoleonness was wanting to wear a bicorne hat, and he was going to be suicidal his entire life if he couldn't but pretty happy if I could, let him wear the damn hat.

I could just link people to other sites' pretty good objections to the same argument.

But I think what I actually want to say is that there was once a time somebody tried pretty much exactly this, silly hat and all. Society shrugged and played along, he led a rich and fulfilling life, his grateful Imperial subjects came to love him, and it's one of the most heartwarming episodes in the history of one of my favorite places in the world.

Sometimes when you make a little effort to be nice to people, even people you might think are weird, really good things happen.

[**EDIT:** *I've been told some people are misinterpreting this post as "you can define words any way you want, don't worry about it". While nothing is stopping you from defining a word any way you want, you should definitely worry about it. I had hoped that the Israel/Palestine example above made it clear that you have to deal with the consequences of your* *definitions, which can include confusion, muddling communication, and leaving openings for deceptive rhetorical strategies. In most cases plausible definitions will be limited to a few possibilities suggested by the territory, with worsening penalties as you deviate from these further.*]

The Control Group Is Out Of Control

I.

Allan Crossman calls parapsychology the control group for science.

That is, in let's say a drug testing experiment, you give some people the drug and they recover. That doesn't tell you much until you give some other people a placebo drug you *know* doesn't work – but which they themselves believe in – and see how many of *them* recover. That number tells you how many people will recover whether the drug works or not. Unless people on your real drug do significantly better than people on the placebo drug, you haven't found anything.

On the meta-level, you're studying some phenomenon and you get some positive findings. That doesn't tell you much until you take some other researchers who are studying a phenomenon you *know* doesn't exist – but which they themselves believe in – and see how many of *them* get positive findings. That number tells you how many studies will discover positive results whether the phenomenon is real or not. Unless studies of the real phenomenon do significantly better than studies of the placebo phenomenon, you haven't found anything.

Trying to set up placebo science would be a logistical nightmare. You'd have to find a phenomenon that definitely doesn't exist, somehow convince a whole community of scientists across the world that it does, and fund them to study it for a couple of decades without them figuring it out.

Luckily we have a natural experiment in terms of parapsychology – the study of psychic phenomena – which most reasonable people believe don't exist, but which a community of practicing scientists believes in and publishes papers on all the time.

The results are pretty dismal. Parapsychologists are able to produce experimental evidence for psychic phenomena about as easily as normal scientists are able to produce such evidence for normal, non-psychic phenomena. This suggests the existence of a very large "placebo effect" in science – ie with enough energy focused on a subject, you can *always* produce "experimental evidence" for it that meets the usual scientific standards. As Eliezer Yudkowsky puts it:

> Parapsychologists are constantly protesting that they are playing by all the standard scientific rules, and yet their results are being ignored – that they are unfairly being held to higher standards than everyone else. I'm willing to believe that. It just means that the standard statistical methods of science are so weak and flawed as to permit a field of study to sustain itself in the complete absence of any subject matter.

These sorts of thoughts have become more common lately in different fields. Psychologists admit to a crisis of replication as some of their most interesting findings turn out to be spurious. And in medicine, John Ioannides and others have been criticizing the research for a decade now and telling everyone they need to up their standards.

"Up your standards" has been a complicated demand that cashes out in a lot of technical ways. But there is broad agreement among the most intelligent voices I read (1, 2, 3, 4, 5) about a couple of promising directions we could go:

1. Demand very large sample size.

2. Demand replication, preferably exact replication, most preferably multiple exact replications.

3. Trust systematic reviews and meta-analyses rather than individual studies. Meta-analyses must prove homogeneity of the studies they analyze.

4. Use Bayesian rather than frequentist analysis, or even combine both techniques.

5. Stricter p-value criteria. It is far too easy to massage p-values to get less than 0.05. Also, make meta-analyses look for “p-hacking” by examining the distribution of p-values in the included studies.

6. Require pre-registration of trials.

7. Address publication bias by searching for unpublished trials, displaying funnel plots, and using statistics like “fail-safe N” to investigate the possibility of suppressed research.

8. Do heterogeneity analyses or at least observe and account for differences in the studies you analyze.

9. Demand randomized controlled trials. None of this “correlated even after we adjust for confounders” BS.

10. Stricter effect size criteria. It’s easy to get small effect sizes in *anything*.

If we follow these ten commandments, then we avoid the problems that allowed parapsychology and probably a whole host of other problems we don’t know about to sneak past the scientific gatekeepers.

Well, what now, motherfuckers?

II.

Bem, Tressoldi, Rabeyron, and Duggan (2014), full text available for download at the top bar of the link above, is parapsychology's way of saying "thanks but no thanks" to the idea of a more rigorous scientific paradigm making them quietly wither away.

You might remember Bem as the prestigious establishment psychologist who decided to try his hand at parapsychology and to his and everyone else's surprise got positive results. Everyone had a lot of criticisms, some of which were very very good, and the study failed replication several times. Case closed, right?

Earlier this month Bem came back with a meta-analysis of ninety replications from tens of thousands of participants in thirty three laboratories in fourteen countries confirming his original finding, $p < 1.2 * -10^{10}$, Bayes factor $7.4 * 10^{9}$, funnel plot beautifully symmetrical, p-hacking curve nice and right-skewed, Orwin fail-safe n of 559, et cetera, et cetera, et cetera.

By my count, Bem follows all of the commandments except [6] and [10]. He apologizes for not using pre-registration, but says it's okay because the studies were exact replications of a previous study that makes it impossible for an unsavory researcher to change the parameters halfway through and does pretty much the same thing. And he apologizes for the small effect size but points out that some effect sizes are legitimately very small, this is no smaller than a lot of other commonly-accepted results, and that a high enough p-value ought to make up for a low effect size.

This is *far* better than the average meta-analysis. Bem has always been pretty careful and this is no exception. Yet its conclusion is that psychic powers exist.

So – once again – what now, motherfuckers?

III.

In retrospect, that list of ways to fix science above was a little optimistic.

The first nine items (large sample sizes, replications, low p-values, Bayesian statistics, meta-analysis, pre-registration, publication bias, heterogeneity) all try to solve the same problem: accidentally mistaking noise in the data for a signal.

We've placed so much emphasis on not mistaking noise for signal that when someone like Bem hands us a beautiful, perfectly clear signal on a silver platter, it briefly stuns us. "Wow, of the three hundred different terrible ways to mistake noise for signal, Bem has proven beyond a shadow of a doubt he hasn't done any of them." And we get so stunned we're likely to forget that this is only part of the battle.

Bem definitely picked up a signal. The only question is whether it's a signal of psi, or a signal of poor experimental technique.

None of these commandments even *touch* poor experimental technique – or confounding, or whatever you want to call it. If an experiment is confounded, if it produces a strong signal even when its experimental hypothesis is true, then using a larger sample size will just make that signal even stronger.

Replicating it will just reproduce the confounded results again.

Low p-values will be easy to get if you perform the confounded experiment on a large enough scale.

Meta-analyses of confounded studies will obey the immortal law of “garbage in, garbage out”.

Pre-registration only assures that your study will not get any worse than it was the first time you thought of it, which may be very bad indeed.

Searching for publication bias only means you will get *all* of the confounded studies, instead of just some of them.

Heterogeneity just tells you whether all of the studies were confounded about the same amount.

Bayesian statistics, alone among these first eight, ought to be able to help with this problem. After all, a good Bayesian should be able to say “Well, I got some impressive results, but my prior for psi is very low, so this raises my belief in psi slightly, but raises my belief that the experiments were confounded *a lot*.”

Unfortunately, good Bayesians are hard to come by, and the researchers here seem to be making some serious mistakes. Here’s Bem:

> An opportunity to calculate an approximate answer to this question emerges from a Bayesian critique of Bem’s (2011) experiments by Wagenmakers, Wetzels, Borsboom, & van der Maas (2011). Although Wagenmakers et al. did not explicitly claim psi to be impossible, they came very close by setting their prior odds at 10^20 against the psi hypothesis. The Bayes

> Factor for our full database is approximately 10^9 in favor of the psi hypothesis (Table 1), which implies that our meta-analysis should lower their posterior odds against the psi hypothesis to 10^11

Let me shame both participants in this debate.

Bem, you are abusing Bayes factor. If Wagenmakers uses your 10^9 Bayes factor to adjust from his prior of 10^-20 to 10^-11, then what happens the next time you come up with another database of studies supporting your hypothesis? We all know you will, because you've amply proven these results weren't due to chance, so whatever factor produced these results – whether real psi or poor experimental technique – will no doubt keep producing them for the next hundred replication attempts. When those come in, does Wagenmakers have to adjust his probability from 10^-11 to 10^-2? When you get another hundred studies, does he have to go from 10^-2 to 10^7? If so, then by conservation of expected evidence he should just update to 10^+7 right now – or really to infinity, since you can keep coming up with more studies till the cows come home. But in fact he shouldn't do that, because at some point his thought process becomes "Okay, I already know that studies of this quality can consistently produce positive findings, so either psi is real or studies of this quality aren't good enough to disprove it". This point should probably happen well before he increases his probability by a factor of 10^9. See Confidence Levels Inside And Outside An Argument for this argument made in greater detail.

Wagenmakers, you are overconfident. Suppose God came down from Heaven and said in a booming voice "EVERY SINGLE STUDY IN THIS META-ANALYSIS WAS CONDUCTED PERFECTLY WITHOUT FLAWS OR BIAS, AS WAS THE META-ANALYSIS ITSELF." You would see a p-value of less than 1.2 *

10^-10 and think “I bet that was just coincidence”? And then they could do another study of the same size, also God-certified, returning exactly the same results, and you would say “I bet that was just coincidence too”? YOU ARE NOT THAT CERTAIN OF ANYTHING. Seriously, *read the @#!$ing Sequences*.

Bayesian statistics, at least the way they are done here, aren’t gong to be of much use to anybody.

That leaves randomized controlled trials and effect sizes.

Randomized controlled trials are great. They eliminate most possible confounders in one fell swoop, and are excellent at keeping experimenters honest. Unfortunately, most of the studies in the Bem meta-analysis were already randomized controlled trials.

High effect sizes are really the only thing the Bem study lacks. And it is very hard to experimental technique so bad that it consistently produces a result with a high effect size.

But as Bem points out, demanding high effect size limits our ability to detect real but low-effect phenomena. Just to give an example, many physics experiments – like the ones that detected the Higgs boson or neutrinos – rely on detecting extremely small perturbations in the natural order, over millions of different trials. Less esoterically, Bem mentions the example of aspirin decreasing heart attack risk, which it definitely does and which is very important, but which has an effect size lower than that of his psi results. If humans have some kind of *very weak* psionic faculty that under regular conditions operates poorly and inconsistently, but does indeed exist, then excluding it by definition from the realm of things science can discover would be a bad idea.

All of these techniques are about reducing the chance of confusing noise for signal. But when we think of them as the be-all and end-all of scientific legitimacy, we end up in awkward situations where they come out super-confident in a study's accuracy simply because the issue was one they weren't geared up to detect. Because a lot of the time the problem is something more than just noise.

IV.

Wiseman & Schlitz's Experimenter Effects And The Remote Detection Of Staring is my favorite parapsychology paper ever and sends me into fits of nervous laughter every time I read it.

The backstory: there is a classic parapsychological experiment where a subject is placed in a room alone, hooked up to a video link. At random times, an experimenter stares at them menacingly through the video link. The hypothesis is that this causes their galvanic skin response (a physiological measure of subconscious anxiety) to increase, even though there is no non-psychic way the subject could know whether the experimenter was staring or not.

Schiltz is a psi believer whose staring experiments had consistently supported the presence of a psychic phenomenon. Wiseman, in accordance with nominative determinism is a psi skeptic whose staring experiments keep showing nothing and disproving psi. Since they were apparently the only two people in all of parapsychology with a smidgen of curiosity or rationalist virtue, they decided to team up and figure out why they kept getting such different results.

The idea was to plan an experiment together, with both of them agreeing on every single tiny detail. They would then go to a laboratory and set it up, again both keeping close eyes on one another. Finally, they would conduct the experiment in a series of different batches. Half the batches (randomly assigned) would be conducted by Dr. Schlitz, the other half by Dr. Wiseman. Because the two authors had very carefully standardized the setting, apparatus and procedure beforehand, “conducted by” pretty much just meant greeting the participants, giving the experimental instructions, and doing the staring.

The results? Schlitz’s trials found strong evidence of psychic powers, Wiseman’s trials found no evidence whatsoever.

Take a second to reflect on how this *makes no sense*. Two experimenters in the same laboratory, using the same apparatus, having no contact with the subjects except to introduce themselves and flip a few switches – and whether one or the other was there that day completely altered the result. For a good time, watch the gymnastics they have to do to in the paper to make this sound sufficiently sensical to even get published. This is the only journal article I’ve ever read where, in the part of the Discussion section where you’re supposed to propose possible reasons for your findings, both authors suggest maybe their co-author hacked into the computer and altered the results.

While it’s nice to see people exploring Bem’s findings further, *this* is the experiment people should be replicating ninety times. I expect *something* would turn up.

As it is, Kennedy and Taddonio list ten similar studies with similar results. One cannot help wondering about publication bias (if the skeptic and the believer got similar results, who cares?). But the phenomenon is sufficiently well known in parapsychology that it has led to its own host of theories about how skeptics emit negative auras, or the enthusiasm of a proponent is a necessary kindling for psychic powers.

Other fields don't have this excuse. In psychotherapy, for example, practically the only consistent finding is that whatever kind of psychotherapy the person running the study likes is most effective. Thirty different meta-analyses on the subject have confirmed this with strong effect size (d = 0.54) and good significance (p = .001).

Then there's Munder (2013), which is a meta-meta-analysis on whether meta-analyses of confounding by researcher allegiance effect were themselves meta-confounded by meta-researcher allegiance effect. He found that indeed, meta-researchers who believed in researcher allegiance effect were more likely to turn up positive results in their studies of researcher allegiance effect (p < .002). It gets worse. There's a famous story about an experiment where a scientist told teachers that his advanced psychometric methods had predicted a couple of kids in their class were about to become geniuses (the students were actually chosen at random). He followed the students for the year and found that their intelligence actually increased. This was supposed to be a Cautionary Tale About How Teachers' Preconceptions Can Affect Children.

Less famous is that the same guy did the same thing with rats. He sent one laboratory a box of rats saying they were specially bred to be ultra-intelligent, and another lab a box of (identical) rats saying they were specially bred to be slow and dumb. Then he had them do standard rat learning tasks, and sure enough the first lab found very impressive results, the second lab very disappointing ones.

This scientist – let’s give his name, Robert Rosenthal – then investigated three hundred forty five different studies for evidence of the same phenomenon. He found effect sizes of anywhere from 0.15 to 1.7, depending on the type of experiment involved. Note that this could also be phrased as “between twice as strong and twenty times as strong as Bem’s psi effect”. Mysteriously, animal learning experiments displayed the highest effect size, supporting the folk belief that animals are hypersensitive to subtle emotional cues.

Okay, fine. Subtle emotional cues. That’s way more scientific than saying “negative auras”. But the question remains – what went wrong for Schlitz and Wiseman? Even if Schlitz had done everything short of saying “The hypothesis of this experiment is for your skin response to increase when you are being stared at, please increase your skin response at that time,” and subjects had tried to comply, the whole point was that they didn’t *know* when they were being stared at, because to find that out you’d have to be psychic. And how are these rats figuring out what the experimenters’ subtle emotional cues mean anyway? *I* can’t figure out people’s subtle emotional cues half the time!

I know that standard practice here is to tell the story of Clever Hans and then say That Is Why We Do Double-Blind Studies. But first of all, I’m pretty sure no one does double-blind studies with rats. Second of all, I think most social psych studies aren’t double blind – I just checked the first one I thought of, Aronson and Steele on stereotype threat, and it certainly wasn’t. Third of all, this effect seems to be just as common in cases where it’s hard to imagine how the researchers’ subtle emotional cues could make a difference. Like Schlitz and Wiseman. Or like the psychotherapy experiments, where most of the subjects were doing therapy with individual psychologists and never even saw whatever prestigious professor was running the study behind the scenes.

I think it’s a combination of subconscious emotional cues, subconscious statistical trickery, perfectly conscious fraud which for all we know happens much more often than detected, and things we haven’t discovered yet which are at least as weird as subconscious emotional cues. But rather than speculate, I prefer to take it as a brute fact. Studies are going to be confounded by the allegiance of the researcher. When researchers who don’t believe something discover it, that’s when it’s worth looking into.

V.

So what exactly happened to Bem?

Although Bem looked hard to find unpublished material, I don’t know if he succeeded. Unpublished material, in this context, has to mean “material published enough for Bem to find it”, which in this case was mostly things presented at conferences. What about results so boring that they were never even mentioned?

And I predict people who believe in parapsychology are more likely to conduct parapsychology experiments than skeptics. Suppose this is true. And further suppose that for some reason, experimenter effect is real and powerful. That means most of the experiments conducted will support Bem’s result. But this is still a weird form of “publication bias” insofar as it ignores the contrary results of hypotheticaly experiments that were never conducted.

And worst of all, maybe Bem really did do an excellent job of finding every little two-bit experiment that no journal would take. How much can we trust these non-peer-reviewed procedures?

I looked through his list of ninety studies for all the ones that were both exact replications and had been peer-reviewed (with one caveat to be mentioned later). I found only seven:

Batthyany, Kranz, and Erber: .268
Ritchie 1: 0.015
Ritchie 2: -0.219
Richie 3: -0.040
Subbotsky 1: 0.279
Subbotsky 2: 0.292
Subbotsky 3: -.399

Three find large positive effects, two find approximate zero effects, and two find large negative effects. Without doing any calculatin’, this seems pretty darned close to chance for me.

Okay, back to that caveat about replications. One of Bem’s strongest points was how many of the studies included were exact replications of his work. This is important because if you do your own novel experiment, it leaves a lot of wiggle room to keep changing the parameters and statistics a bunch of times until you get the effect you want. This is why lots of people want experiments to be preregistered with specific committments about what you’re going to test and how you’re going to do it. These experiments weren’t preregistered, but conforming to a previously done experiment is a pretty good alternative.

Except that I think the criteria for “replication” here were exceptionally loose. For example, Savva et al was listed as an “exact replication” of Bem, but it was performed in 2004 – seven years before Bem’s original study took place. I know Bem believes in precognition, but that’s going *too far*. As far as I can tell “exact replication” here means “kinda similar psionic-y thing”. Also, Bem classily lists his own experiments as exact replications of themselves, which gives a big boost to the “exact replications return the same results as Bem’s original studies” line. I would want to see much stricter criteria for replication before I relax the “preregister your trials” requirement.

(Richard Wiseman – the same guy who provided the negative aura for the Wiseman and Schiltz experiment – has started a pre-register site for Bem replications. He says he has received five of them. This is very promising. There is also a separate pre-register for parapsychology trials in general. I am both extremely pleased at this victory for good science, and ashamed that my own field is apparently behind parapsychology in the “scientific rigor” department)

That is my best guess at what happened here – a bunch of poor-quality, peer-unreviewed studies that weren’t as exact replications as we would like to believe, all subject to mysterious experimenter effects.

This is not a criticism of Bem or a criticism of parapsychology. It’s something that is inherent to the practice of meta-analysis, and even more, inherent to the practice of science. Other than a few very exceptional large medical trials, there is not a study in the world that would survive the level of criticism I am throwing at Bem right now.

I think Bem is wrong. The level of criticism it would take to prove a wrong study wrong is higher than that almost any existing study can withstand. That is not encouraging for existing studies.

VI.

The motto of the Royal Society – Hooke, Boyle, Newton, some of the people who arguably invented modern science – was *nullus in verba*, “take no one’s word”.

This was a proper battle cry for seventeenth century scientists. Think about the (admittedly kind of mythologized) history of Science. The scholastics saying that matter was this, or that, and justifying themselves by long treatises about how based on A, B, C, the word of the Bible, Aristotle, self-evident first principles, and the Great Chain of Being all clearly proved their point. Then other scholastics would write different long treatises on how D, E, and F, Plato, St. Augustine, and the proper ordering of angels all indicated that clearly matter was something different. Both groups were pretty sure that the other had make a subtle error of reasoning somewhere, and both groups were perfectly happy to spend centuries debating exactly which one of them it was.

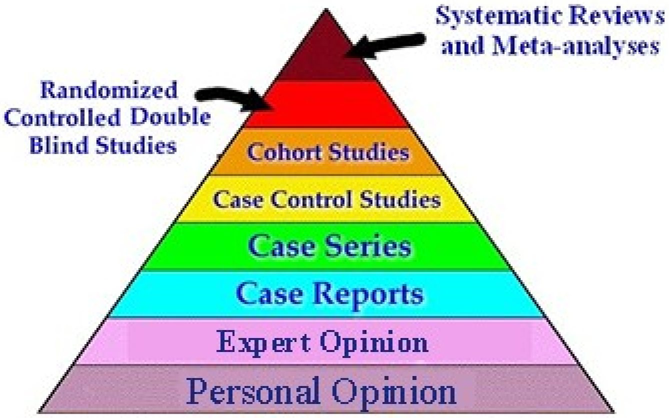
And then Galileo said “Wait a second, instead of debating exactly how objects fall, let’s just drop objects off of something really tall and see what happens”, and after that, Science.

Yes, it’s kind of mythologized. But like all myths, it contains a core of truth. People are terrible. If you let people debate things, they will do it forever, come up with horrible ideas, get them entrenched, play politics with them, and finally reach the point where they’re coming up with theories why people who disagree with them are probably secretly in the pay of the Devil.

Imagine having to conduct the global warming debate, except that you couldn’t appeal to scientific consensus and statistics because scientific consensus and statistics hadn’t been invented yet. In a world without science, *everything* would be like that.

Heck, just look at *philosophy*.

This is the principle behind the Pyramid of Scientific Evidence. The lowest level is your personal opinions, no matter how ironclad you think the logic behind them is. Just above that is expert opinion, because no matter how expert someone is they're still only human. Above that is anecdotal evidence and case studies, because even though you're finally getting out of people's heads, it's still possible for the content of people's heads to influence which cases they pay attention to. At each level, we distill away more and more of the human element, until presumably at the top the dross of humanity has been purged away entirely and we end up with pure unadulterated reality.


Systematic Reviews and Meta-analyses
Randomized Controlled Double Blind Studies
Cohort Studies
Case Control Studies
Case Series
Case Reports
Expert Opinion
Personal Opinion

The Pyramid of Scientific Evidence

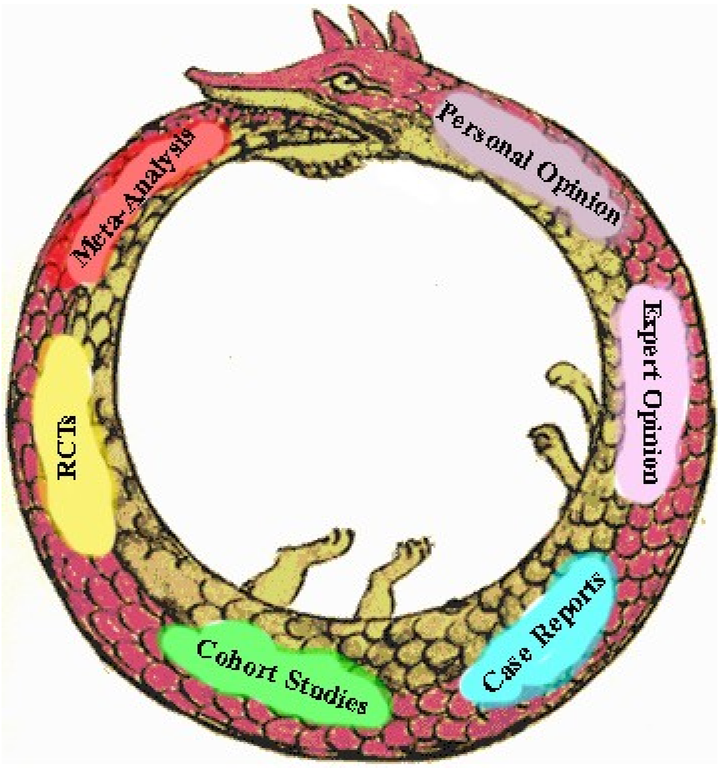
And for a while this went *well*. People would drop things off towers, or see how quickly gases expanded, or observe chimpanzees, or whatever.

Then things started getting more complicated. People started investigating more subtle effects, or effects that shifted with the observer. The scientific community became bigger, everyone didn't know everyone anymore, you needed more journals to find out what other people had done. Statistics became more complicated, allowing the study of noisier data but also bringing more peril. And a lot of science done by smart and honest people ended up being wrong, and we needed to figure out exactly which science that was.

And the result is a lot of essays like this one, where people who think they're smart take one side of a scientific "controversy" and say which studies you should believe. And then other people take the other side and tell you why you should believe different studies than the first person thought you should believe. And there is much argument and many insults and citing of authorities and interminable debate for, if not centuries, at least a pretty long time.

The highest level of the Pyramid of Scientific Evidence is meta-analysis. But a lot of meta-analyses are crap. This meta-analysis got p < 1.2 * 10^-10 for a conclusion I'm pretty sure is false, and *it isn't even one of the crap ones*. Crap meta-analyses look more like this, or even worse.

How do I know it's crap? Well, I use my personal judgment. How do I know my personal judgment is right? Well, a smart well-credentialed person like James Coyne agrees with me. How do I know James Coyne is smart? I can think of lots of cases where he's been right before. How do I know those count? Well, John Ioannides has published a lot of studies analyzing the problems with science, and confirmed that cases like the ones Coyne talks about are pretty common. Why can I believe Ioannides' studies? Well, there have been good meta-analyses of them. But how do I know if those meta-analyses are crap or not? Well…


Personal Opinion
Expert Opinion
Case Reports
Cohort Studies
RCTs
Meta-Analysis

The Ouroboros of Scientific Evidence

Science! YOU WERE THE CHOSEN ONE! It was said that you would destroy reliance on biased experts, not join them! Bring balance to epistemology, not leave it in darkness!

I LOVED YOU!!!!

Edit: Conspiracy theory by Andrew Gelman

The Toxoplasma Of Rage

I.

Some old news I only just heard about: PETA is offering to pay the water bills for needy Detroit families if (and only if) those families agree to stop eating meat.

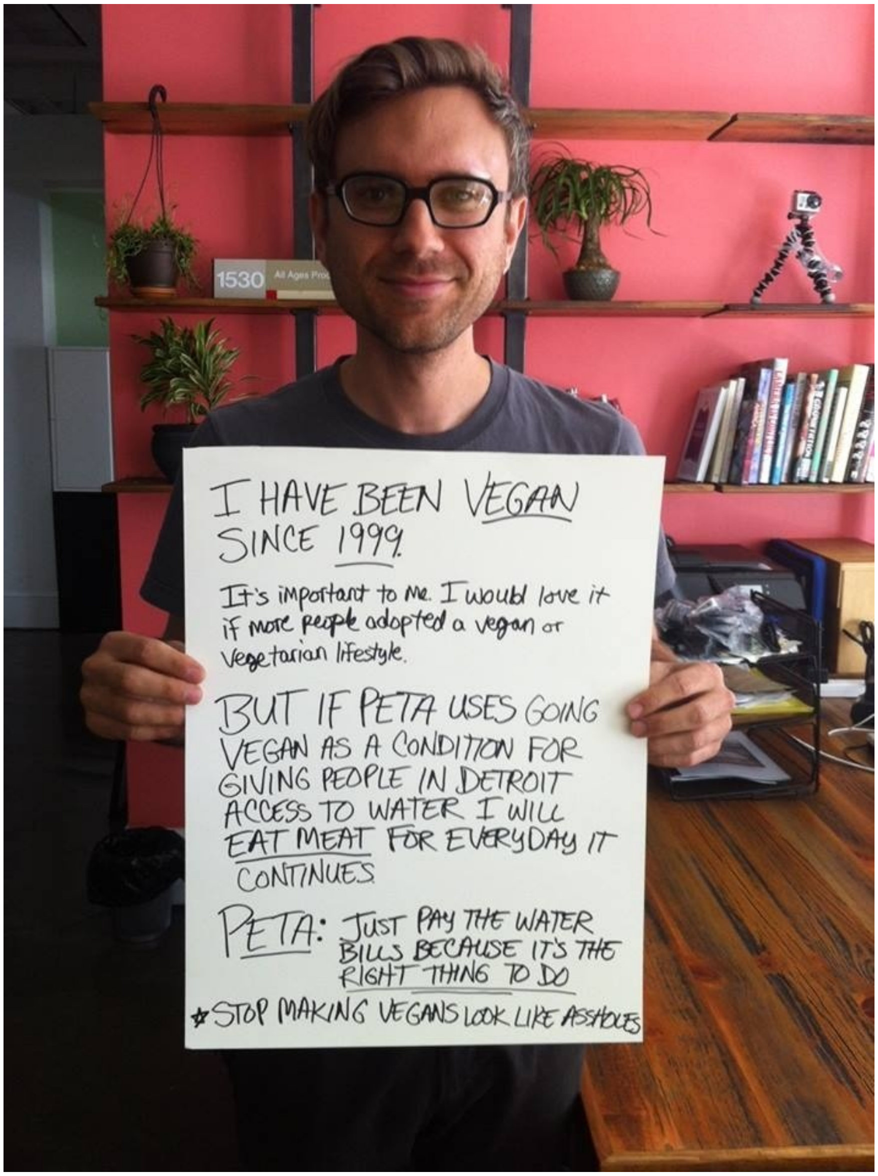
Predictably, the move caused a backlash. The International Business Times, in what I can only assume is an attempted pun, describes them as “drowning in backlash”. Groundswell thinks it’s a “big blunder”. Daily Banter says it’s “exactly why everyone hates PETA”. Jezebel calls them “assholes”.

Of course, this is par for the course for PETA, who have previously engaged in campaigns like throwing red paint on fashion models who wear fur, juxtaposing pictures of animals with Holocaust victims, juxtaposing pictures of animals with African-American slaves, and ads featuring naked people that cross the line into pornography.

People call these things “blunders”, but consider the alternative. Vegan Outreach is an extremely responsible charity doing excellent and unimpeachable work in the same area PETA is. Nobody has heard of them. *Everybody* has heard of PETA, precisely because of the interminable stupid debates about “did this publicity stunt cross the line?”

While not everyone is a vegan, most people who learn enough about factory farming are upset by it. There is pretty much zero room for PETA to convert people from pro-factory-farming to anti-factory-farming, because there aren’t any radical grassroots pro-factory-farming activists to be found. Their problem isn’t lack of agreement. It’s lack of attention.

PETA creates attention, but at a cost. Everybody’s talking about PETA, which is sort of like everybody talking about ethical treatment of animals, which is sort of a victory. But most of the talk is “I hate them and they make me really angry.” Some of the talk is even “I am going to eat a lot more animals just to make PETA mad.”


1530
I HAVE BEEN VEGAN SINCE 1999.
It's important to me. I would love it if more people adopted a vegan or vegetarian lifestyle.
BUT IF PETA USES GOING VEGAN AS A CONDITION FOR GIVING PEOPLE IN DETROIT ACCESS TO WATER I WILL EAT MEAT FOR EVERYDAY IT CONTINUES.
PETA: JUST PAY THE WATER BILLS BECAUSE IT'S THE RIGHT THING TO DO
☆ STOP MAKING VEGANS LOOK LIKE ASSHOLES

So there’s a tradeoff here, with Vegan Outreach on one side and PETA on the other.

Vegan Outreach can get everyone to agree in principle that factory-farming is bad, but no one will pay any attention to it.

And PETA can get everyone to pay attention to factory farming, but a lot of people who would otherwise oppose it will switch to supporting it just because they're so mad at the way it's being publicized.

But at least they're paying attention!

PETA doesn't shoot themselves in the foot because they're stupid. They shoot themselves in the foot because they're traveling up an incentive gradient that rewards them for doing so, even if it destroys their credibility.

II.

The University of Virginia rape case profiled in *Rolling Stone* has fallen apart. In doing so, it joins a long and distinguished line of highly-publicized rape cases that have fallen apart. Studies sometimes claim that only 2 to 8 percent of rape allegations are false. Yet the rate for allegations that go ultra-viral in the media must be an order of magnitude higher than this. As the old saying goes, once is happenstance, twice is coincidence, three times is enemy action.

The enigma is complicated by the observation that it's usually feminist activists who are most instrumental in taking these stories viral. It's not some conspiracy of pro-rape journalists choosing the most dubious accusations in order to discredit public trust. It's people specifically selecting these incidents as flagship cases for their campaign that rape victims need to be believed and trusted. So why are the most publicized cases so much more likely to be false than the almost-always-true average case?

Several people have remarked that false accusers have more leeway to make their stories as outrageous and spectacular as possible. But I want to focus on two less frequently mentioned concerns.

The Consequentialism FAQ explains signaling in moral decisions like so:

> When signaling, the more expensive and useless the item is, the more effective it is as a signal. Although eyeglasses are expensive, they're a poor way to signal wealth because they're very useful; a person might get them not because ey is very rich but because ey really needs glasses. On the other hand, a large diamond is an excellent signal; no one needs a large diamond, so anybody who gets one anyway must have money to burn.
>
> Certain answers to moral dilemmas can also send signals. For example, a Catholic man who opposes the use of condoms demonstrates to others (and to himself!) how faithful and pious a Catholic he is, thus gaining social credibility. Like the diamond example, this signaling is more effective if it centers upon something otherwise useless. If the Catholic had merely chosen not to murder, then even though this is in accord with Catholic doctrine, it would make a poor signal because he might be doing it for other good reasons besides being Catholic – just as he might buy eyeglasses for reasons beside being rich. It is precisely because opposing condoms is such a horrendous decision that it makes such a good signal.
>
> But in the more general case, people can use moral decisions to signal how moral they are. In this case, they choose a disastrous decision based on some moral principle. The more suffering and destruction they

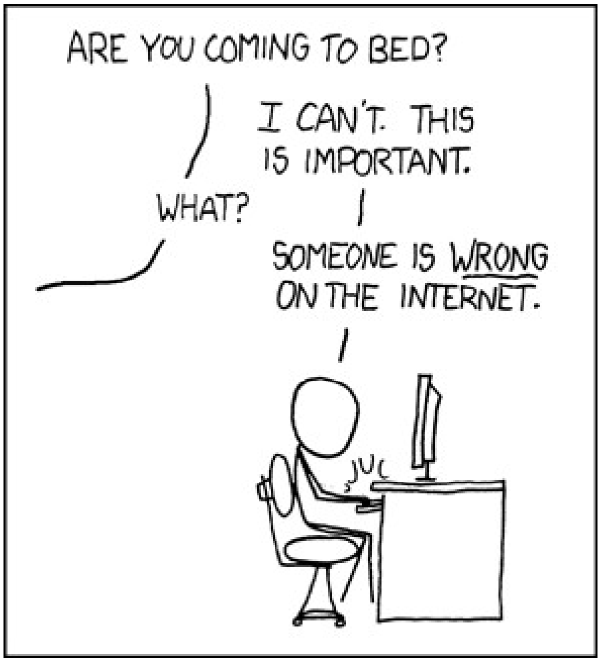
> support, and the more obscure a principle it is, the more obviously it shows their commitment to following their moral principles absolutely. For example, Immanuel Kant claims that if an axe murderer asks you where your best friend is, obviously intending to murder her when he finds her, you should tell the axe murderer the full truth, because lying is wrong. This is effective at showing how moral a person you are – no one would ever doubt your commitment to honesty after that – but it's sure not a very good result for your friend.

In the same way, publicizing how strongly you believe an accusation that is obviously true signals nothing. Even hard-core anti-feminists would believe a rape accusation that was caught on video. A moral action that can be taken just as well by an outgroup member as an ingroup member is crappy signaling and crappy identity politics. If you want to signal how strongly you believe in taking victims seriously, you talk about it in the context of the least credible case you can find.

But aside from that, there's the PETA Principle: the more controversial something is, the more it gets talked about.

A rape that obviously happened? Shove it in people's face and they'll admit it's an outrage, just as they'll admit factory farming is an outrage. But they're not going to talk about it much. There are a zillion outrages every day, you're going to need more than that to draw people out of their shells.

On the other hand, the controversy over dubious rape allegations is exactly that – a controversy. People start screaming at each other about how they're misogynist or misandrist or whatever, and Facebook feeds get filled up with hundreds of comments in all capital letters about how my ingroup is being persecuted by your ingroup. At each step, more and more people get triggered and upset. Some of those triggered people do emergency ego defense by reblogging articles about how the group that triggered them are terrible, triggering further people in a snowball effect that spreads the issue further with every iteration.


ARE YOU COMING TO BED?
I CAN'T. THIS IS IMPORTANT.
WHAT?
SOMEONE IS WRONG ON THE INTERNET.

[source]

Only controversial things get spread. A rape allegation will only be spread if it's dubious enough to split people in half along lines corresponding to identity politics. An obviously true rape allegation will only be spread if the response is controversial enough to split people in half along lines corresponding to identity politics – which is why so much coverage focuses on the proposal that all accused rapists should be treated as guilty until proven innocent.

Everybody hates rape just like everybody hates factory farming. "Rape culture" doesn't mean most people like rape, it means most people ignore it. That means feminists face the same double-bind that PETA does.

First, they can respond to rape in a restrained and responsible way, in which case everyone will be against it and nobody will talk about it.

Second, they can respond to rape in an outrageous and highly controversial way, in which case everybody will talk about it but it will autocatalyze an opposition of people who hate feminists and obsessively try to prove that as many rape allegations as possible are false.

I have yet to see anyone holding a cardboard sign talking about how they are going to rape people just to make feminists mad, but it's only a matter of time. Like PETA, their incentive gradient dooms them to shoot themselves in the foot again and again.

III.

Slate recently published an article about white people's contrasting reactions to the Michael Brown shooting in Ferguson versus the Eric Garner choking in NYC. And man, it is *some* contrast.

A Pew poll found that of white people who expressed an opinion about the Ferguson case, 73% sided with the officer. Of white people who expressed an opinion about the Eric Garner case, 63% sided with the black victim.

Media opinion follows much the same pattern. Arch-conservative Bill O'Reilly said he was "absolutely furious" about the way "the liberal media" and "race hustlers" had "twisted the story" about Ferguson in the service of "lynch mob justice" and "insulting the American police community, men and women risking their lives to protect us". But when it came to Garner, O'Reilly said he was "extremely troubled" and that "there was a police overreaction that should have been adjudicated in a court of law." His guest on FOX News, conservative commentator and fellow Ferguson-detractor Charles Krauthammer added that “From looking at the video, the grand jury’s decision [not to indict] is totally incomprehensible.” Saturday Night Live did a skit about Al Sharpton talking about the Garner case and getting increasingly upset because “For the first time in my life, everyone agrees with me.”

This follows about three months of most of America being at one another’s throats pretty much full-time about Ferguson. We got treated to a daily diet of articles like Ferguson Protester On White People: “Y’all The Devil” or Black People Had The Power To Fix The Problems In Ferguson Before The Brown Shooting – They Failed or Most White People In America Are Completely Oblivious and a whole bunch of people sending angry racist editorials and counter-editorials to each other for months. The damage done to race relations is difficult to overestimate – CBS reports that they dropped ten percentage points to the lowest point in twenty years, with over half of blacks now describing race relations as “bad”.

And people say it was all worth it, because it raised awareness of police brutality against black people, and if that rustles some people’s jimmies, well, all the worse for them.

But the Eric Garner case also would have raised awareness of police brutality against black people, and everybody would have agreed about it. It has become increasingly clear that, given sufficiently indisputable evidence of police being brutal to a black person, pretty much everyone in the world condemns it equally strongly.

And it’s not just that the Eric Garner case came around too late so we had to make do with the Mike Brown case. Garner was choked a month before Brown was shot, but the story was ignored, then dug back up later as a tie-in to the ballooning Ferguson narrative.

More important, unarmed black people are killed by police or other security officers about twice a week according to official statistics, and probably much more often than that. You're saying none of these shootings, hundreds each year, made as good a flagship case as Michael Brown? In all this gigantic pile of bodies, you couldn't find *one* of them who hadn't just robbed a convenience store? Not a single one who didn't have ten eyewitnesses and the forensic evidence all saying he started it?

I propose that the Michael Brown case went viral – rather than the Eric Garner case or any of the hundreds of others – because of the PETA Principle. It was controversial. A bunch of people said it was an outrage. A bunch of other people said Brown totally started it, and the officer involved was a victim of a liberal media that was hungry to paint his desperate self-defense as racist, and so the people calling it an outrage were themselves an outrage. Everyone got a great opportunity to signal allegiance to their own political tribe and discuss how the opposing political tribe were vile racists / evil race-hustlers. There was a steady stream of potentially triggering articles to share on Facebook to provoke your friends and enemies to counter-share articles that would trigger *you*.

The Ferguson protesters say they have a concrete policy proposal – they want cameras on police officers. There's only spotty polling on public views of police body cameras before the Ferguson story took off, but what there is seems pretty unanimous. A UK poll showed that 90% of the population of that country wanted police to have body cameras in February. US polls are more of the form "crappy poll widget on a news site" (1, 2, 3) but they all hovered around 80%

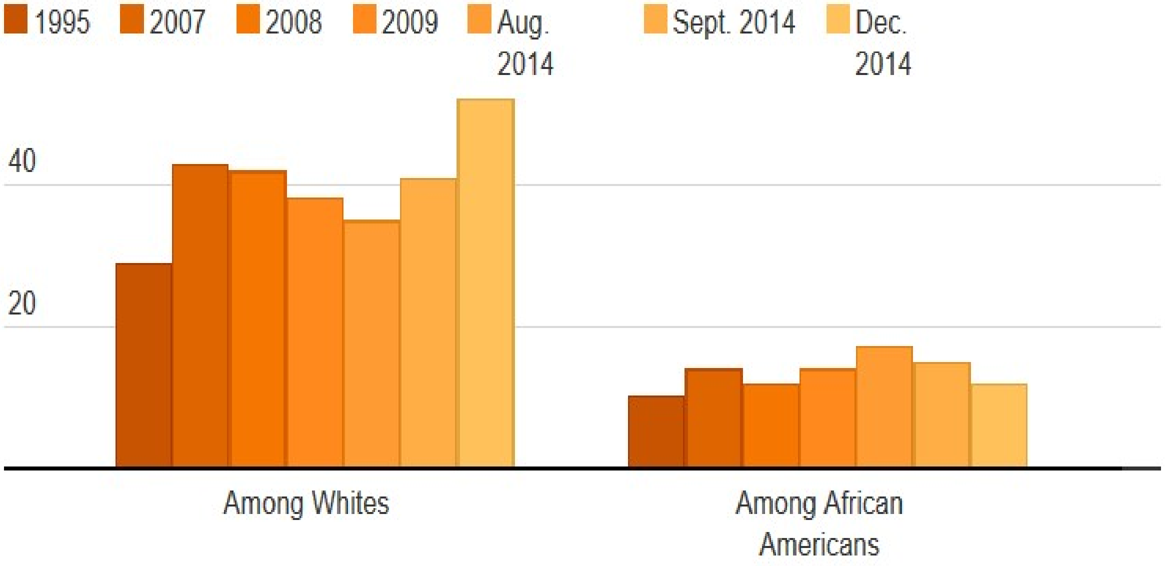
approval for the past few years. I also found a poll by Police Magazine in which a plurality of the *police officers* they surveyed wanted to wear body cameras, probably because of evidence that they cut down on false accusations. Even before Ferguson happened, you would have a really hard time finding anybody in or out of uniform who thought police cameras were a bad idea.

And now, after all is said and done, ninety percent of people are still in favor – given methodology issues, the extra ten percent may or may not represent a real increase. The difference between whites and blacks is a rounding error. The difference between Democrats and Republicans is barely worth talking about- 79% of Republicans are still in support. The people who think Officer Darren Wilson is completely innocent and the grand jury was right to release him, the people muttering under their breath about race hustlers and looters – *eighty percent of those people still want cameras on their cops.*

If the Ferguson protests didn't do much to the public's views on police body cameras, they sure changed its views on some other things. I wrote before about how preliminary polls say that hearing about Ferguson *increased* white people's confidence in the way the police treat race. Now the less preliminary polls are out, and they show the effect was larger than even I expected.

Whites' confidence in local police highest on record

How much confidence do you have in police officers in your community to treat blacks and whites equally: A great deal, a fair amount, just some, or very little confidence? Results show % saying they have a "great deal" of confidence



Source: NBC, NBC/WSJ, Pew Research and USA Today/Gallup national polls

The Washington Post

[source]

White people’s confidence in the police being racially unbiased increased from 35% before the story took off to 52% today. Could even a deliberate PR campaign by the nation’s police forces have done better? I doubt it.

It’s possible that this is an artifact of the question’s wording – after all, it asks people about their local department, and maybe after seeing what happened in Ferguson, people’s local police forces look pretty good by comparison. But then why do black people show the opposite trend?

I think this is exactly what it looks like. Just as PETA’s outrageous controversial campaign to spread veganism make people want to eat more animals in order to spite them, so the controversial nature of this particular campaign against police brutality and racism made white people like their local police department even more to spite the people talking about how all whites were racist.

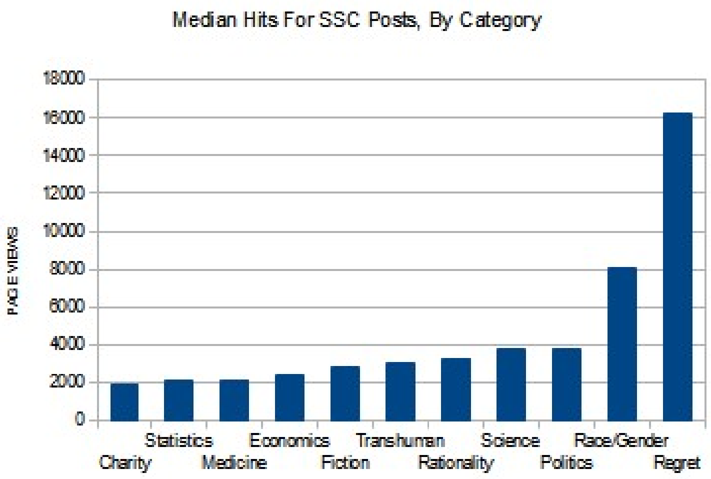
Once again, the tradeoff.

If campaigners against police brutality and racism were extremely responsible, and stuck to perfectly settled cases like Eric Garner, everybody would agree with them but nobody would talk about it.

If instead they bring up a very controversial case like Michael Brown, everybody will talk about it, but they will catalyze their own opposition and make people start supporting the police more just to spite them. More foot-shooting.

IV.

Here is a graph of some of the tags I commonly use for my posts, with the average number of hits per post in each tag. It’s old, but I don’t want to go through the trouble of making a new one, and the trends have stayed the same since then.



I blog about charity only rarely, but it must be the most important thing I can write about here. Convincing even a few more people to donate to charity, or to redirect their existing donations to a more effective program, can literally save dozens or even hundreds of lives even with the limited reach that a private blog has. It probably does more good for the world than all of the other categories on here combined. But it's completely uncontroversial – everyone agrees it's a good thing – and it is the least viewed type of post.

Compare this to the three most viewed category of post. Politics is self-explanatory. Race and gender are a type of politics even more controversial and outrage-inducing than regular politics. And that "regret" all the way on the right is my "things i will regret writing" tag, for posts that I know are going to start huge fights and probably get me in lots of trouble. They're usually race and gender as well, but digging deep into the really really controversial race and gender related issues.

The less useful, and more controversial, a post here is, the more likely it is to get me lots of page views.

For people who agree with me, my angry rants on identity politics are a form of ego defense, saying "You're okay, your in-group was in the right the whole time.” Linking to it both raises their status as an in-group members, and acts as a potential assault on out-group members who are now faced with strong arguments telling them they’re wrong. And the people who disagree with me will sometimes write angry rebuttals on their own blogs, and those rebuttals will link to my own post and spread it further. Or they’ll talk about it with their disagreeing friends, and their friends will get mad and want to tell me I’m wrong, and come over here to read the post to get more ammunition for their counterarguments. I have a feature that tells me who links to all of my posts, so I can *see* this all happening in real-time.

I don’t make enough money off the ads on this blog to matter much. But if I lived off them, which do you think I’d write more of? Posts about charity which only get me 2,000 paying customers? Or posts that turn all of you against one another like a pack of rabid dogs, and get me 16,000?

I don’t have a fancy bar graph for them, but I bet this same hierarchy of interestingness applies to the great information currents and media outlets that shape society as a whole. It’s in activists’ interests to destroy their own causes by focusing on the most controversial cases and principles, the ones that muddy the waters and make people oppose them out of spite. And it’s in the media’s interest to help them and egg them on.

V.

And now, for something completely different.

Before “meme” meant doge and all your base, it was a semi-serious attempt to ground cultural evolution in parasitology. The idea was to replace a model of humans choosing whichever ideas they liked with a model of ideas as parasites that evolved in ways that favored their own transmission.

This never really caught on, because most people’s response was “That’s neat. So what?”

But let’s talk about toxoplasma.

Toxoplasma is a neat little parasite that is implicated in a couple of human diseases including schizophrenia. Its life cycle goes like this: it starts in a cat. The cat poops it out. The poop and the toxoplasma get in the water supply, where they are consumed by some other animal, often a rat. The toxoplasma morphs into a rat-compatible form and starts reproducing. Once it has strength in numbers, it hijacks the rat’s brain, convincing the rat to hang out conspicuously in areas where cats can eat it. After a cat eats the rat, the toxoplasma morphs back into its cat compatible form and reproduces some more. Finally, it gets pooped back out by the cat, completing the cycle.

It’s the ciiiiiircle of life!

What would it mean for a meme to have a life cycle as complicated as toxoplasma?

Consider the war on terror. They say that every time the United States bombs Pakistan or Afghanistan or somewhere, all we're doing is radicalizing the young people there and making more terrorists. Those terrorists then go on to kill Americans, which makes Americans get very angry and call for more bombing of Pakistan and Afghanistan.

Taken as a meme, it's a single parasite with two hosts and two forms. In an Afghan host, it appears in a form called 'jihad', and hijacks its host into killing himself in order to spread it to its second, American host. In the American host it morphs in a form called 'the war on terror', and it hijacks the Americans into giving *their* own lives (and tax dollars) to spread it back to its Afghan host in the form of bombs.

From the human point of view, jihad and the War on Terror are opposing forces. From the memetic point of view, they're as complementary as caterpillars and butterflies. Instead of judging, we just note that somehow we accidentally created a replicator, and replicators are going to replicate until something makes them stop.

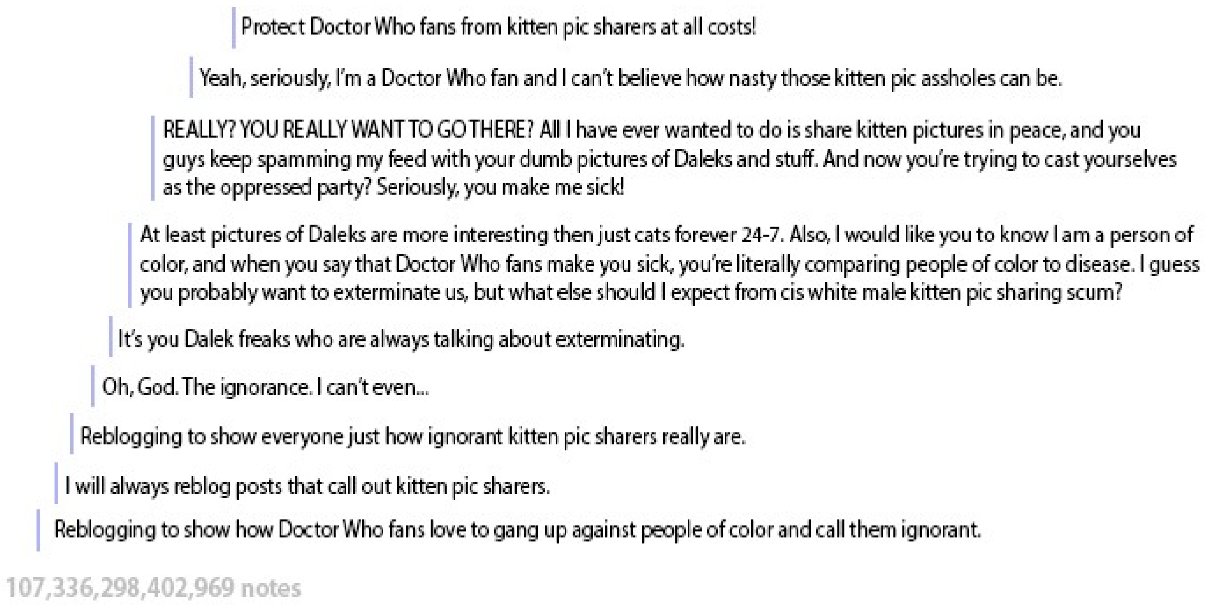
Replicators are also going to *evolve*. Some Afghan who thinks up a particularly effective terrorist strategy helps the meme spread to more Americans as the resulting outrage fuels the War on Terror. When the American bombing heats up, all of the Afghan villagers radicalized in by the attack will remember the really effective new tactic that Khalid thought up and do *that one* instead of the boring old tactic that barely killed any Americans at all. Some American TV commentator who comes up with a particularly stirring call to retaliation will find her words adopted into party platforms and repeated by pro-war newspapers. While pacifists on both sides work to defuse the tension, the meme is engaging in a counter-effort to become as virulent as possible, until people start suggesting putting pork fat in American bombs just to make Muslims *even madder*.

And let's talk about Tumblr.

Tumblr’s interface doesn’t allow you to comment on other people’s posts, per se. Instead, it lets you reblog them with your own commentary added. So if you want to tell someone they’re an idiot, your only option is to reblog their entire post to all your friends with the message “you are an idiot” below it.

Whoever invented this system either didn’t understand memetics, or understood memetics *much too well*.

What happens is – someone makes a statement which is controversial by Tumblr standards, like “Protect Doctor Who fans from kitten pic sharers at all costs.” A kitten pic sharer sees the statement, sees red, and reblogs it to her followers with a series of invectives against Doctor Who fans. Since kitten pic sharers cluster together in the social network, soon *every* kitten pic sharer has seen the insult against kitten pic sharer – as they all feel the need to add their defensive commentary to it, soon all of them are seeing it from ten different directions. The angry invectives get back to the Doctor Who fans, and now *they* feel deeply offended, so they reblog it among themselves with even more condemnations of the kitten pic sharers, who now not only did whatever inspired the enmity in the first place, but have inspired extra hostility because their hateful invectives are right there on the post for everyone to see. So about half the stuff on your dashboard is something you actually want to see, and the other half is towers of alternate insults that look like this:



Protect Doctor Who fans from kitten pic sharers at all costs!

Yeah, seriously, I'm a Doctor Who fan and I can't believe how nasty those kitten pic assholes can be.

REALLY? YOU REALLY WANT TO GO THERE? All I have ever wanted to do is share kitten pictures in peace, and you guys keep spamming my feed with your dumb pictures of Daleks and stuff. And now you're trying to cast yourselves as the oppressed party? Seriously, you make me sick!

At least pictures of Daleks are more interesting then just cats forever 24-7. Also, I would like you to know I am a person of color, and when you say that Doctor Who fans make you sick, you're literally comparing people of color to disease. I guess you probably want to exterminate us, but what else should I expect from cis white male kitten pic sharing scum?

It's you Dalek freaks who are always talking about exterminating.

Oh, God. The ignorance. I can't even...

Reblogging to show everyone just how ignorant kitten pic sharers really are.

I will always reblog posts that call out kitten pic sharers.

Reblogging to show how Doctor Who fans love to gang up against people of color and call them ignorant.

107,336,298,402,969 notes

Actually, pretty much this happened to the PETA story I started off with

And then you sigh and scroll down to the next one. Unless of course you are a Doctor Who fan, in which case you sigh and then immediately reblog with the comment "It's obvious you guys started ganging up against us first, don't try to accuse **US** now" because you can't just *let that accusation stand*.

I make fun of Tumblr social justice sometimes, but the problem isn't with Tumblr social justice, it's structural. Every community on Tumblr somehow gets enmeshed with the people most devoted to making that community miserable. The tiny Tumblr rationalist community somehow attracts, concentrates, and constantly reblogs stuff from the even tinier Tumblr community of people who hate rationalists and want them to be miserable (no, well-intentioned and intelligent critics, I am not talking about you). It's like one of those rainforest ecosystems where every variety of rare endangered nocturnal spider hosts a parasite who has evolved for millions of years solely to parasitize that one spider species, and the parasites host parasites who have evolved for millions of years solely to parasitize *them*. If Tumblr social justice is worse than anything else, it’s mostly because everyone has a race and a gender so it’s easier to fire broad cannonades and just hit everybody.

Tumblr’s reblog policy makes it a hothouse for toxoplasma-style memes that spread via outrage. Following the ancient imperative of evolution, if memes spread by outrage they adapt to become as outrage-inducing as possible.

Or rather, that is just one of their many adaptations. I realize this toxoplasma metaphor sort of strains credibility, so I want to anchor this idea of outrage-memes in pretty much the only piece of memetics everyone can agree upon.

The textbook example of a meme – indeed, almost the only example ever discussed – is the chain letter. “Send this letter to ten people and you will prosper. Fail to pass it on, and you will die tomorrow.” And so the letter replicates.

It might be useful evidence that we were on the right track here, with our toxoplasma memes and everything, if we could find evidence that they reproduced in the same way.

If you’re not on Tumblr, you might have missed the “everyone who does not reblog the issue du jour is trash” wars. For a few weeks around the height of the Ferguson discussion, people constantly called out one another for not reblogging enough Ferguson-related material, or (Heavens forbid) saying they were sick of the amount of Ferguson material they were seeing. It got so bad that various art blogs that just posted pretty paintings, or kitten picture blogs that just reblogged pictures of kittens were feeling the heat (you thought I was joking about the hate for kitten picture bloggers. I never joke.) Now the issue du jour seems to be Pakistan. Just to give a few examples:

“friends if you are reblogging things that are not about ferguson right now please queue them instead. please pay attention to things that are more important. it’s not the time to talk about fandoms or jokes it’s time to talk about injustices.” [source]

“can yall maybe take some time away from reblogging fandom or humor crap and read up and reblog pakistan because the privilege you have of a safe bubble is not one shared by others” [source]

“If you’re uneducated, do not use that as an excuse. Do not say, “I’m not picking sides because I don’t know the full story,” because not picking a side is supporting Wilson. And by supporting him, you are on a racist side… Ignoring this situation will put you in deep shit, and it makes you racist. If you’re not racist, do not just say “but I’m not racist!!” just get educated and reblog anything you can.” [source]

“why are you so disappointing? I used to really like you. you’ve kept totally silent about peshawar, not acknowledging anything but fucking zutara or bellarke or whatever. there are other posts you’ve reblogged too that I wouldn’t expect you to- but those are another topic. I get that you’re 19 but maybe consider becoming a better fucking person?” [source]

“if you’re white, before you reblog one of those posts that’s like “just because i’m not blogging about ferguson doesn’t mean i don’t care!!!” take a few seconds to: consider the privilege you have that allows you not to pay attention if you don’t want to. consider those who do not have the privilege to focus on other things. ask yourself why you think it’s more important that people know you “care” than it is to spread information and

> show support. then consider that you are a fucking shitbaby.” [source]
>
> “For everyone reblogging Ferguson, Ayotzinapa, North Korea etc and not reblogging Peshawar, you should seriously be ashamed of yourselves.” [source]
>
> “This is going to be an unpopular opinion but I see stuff about ppl not wanting to reblog ferguson things and awareness around the world because they do not want negativity in their life plus it will cause them to have anxiety. They come to tumblr to escape n feel happy which think is a load of bull. There r literally ppl dying who live with the fear of going outside their homes to be shot and u cant post a fucking picture because it makes u a little upset?? I could give two fucks about internet shitlings.” [source]

You may also want to check the Tumblr tag “the trash is taking itself out”, in which hundreds of people make the same joke (“I think some people have stopped reading my blog because I’m talking too much about [the issue *du jour*]. I guess the trash is taking itself out now.”)

This is pretty impressive. It’s the first time outside of a chain letter that I have seen our memetic overlords throw off all pretense and just go around shouting “SPREAD ME OR YOU ARE GARBAGE AND EVERYONE WILL HATE YOU.”

But it only works because it’s tapped into the most delicious food source an ecology of epistemic parasites could possibly want – controversy,

I would like to be able to write about charity more often. Feminists would probably like to start supercharging the *true* rape accusations for a change. Protesters against police brutality would probably like to be able to focus on clear-cut cases that won't make white people support the police *even harder*. Even PETA would probably prefer being the good guys for once. But the odds aren't good. Not because the people involved are bad people who want to fail. Not even because the media-viewing public are stupid. Just because information ecologies are not your friend.

This blog tries to remember the Litany of Jai: "Almost no one is evil; almost everything is broken". We pretty much never wrestle with flesh and blood; it's powers and principalities all the way down.

VI.

A while ago I wrote a post called Meditations on Moloch where I pointed out that in any complex multi-person system, the system acts according to its own chaotic incentives that don't necessarily correspond to what *any* individual within the system wants. The classic example is the Prisoner's Dilemma, which usually ends at defect-defect even though *both* of the two prisoners involved prefer cooperate-cooperate. I compare this malignant discoordination to Ginsberg's portrayal of Moloch, the demon-spirit of capitalism gone wrong.

> I would support instating a National Conversation Topic Czar if that allowed us to get rid of celebrities.
>
> — Steven Kaas (@stevenkaas) August 26, 2010

Steven in his wisdom reminds us that there is no National Conversation Topic Czar. The rise of some topics to national prominence and the relegation of others to tiny print on the eighth page of the newspapers occurs by an emergent uncoordinated process. When we say "the media decided to cover Ferguson instead of Eric Garner", we reify and anthropomorphize an entity incapable of making goal-directed decisions.

A while back there was a minor scandal over JournoList, a private group where left-leaning journalists met and exchanged ideas. I think the conservative spin was “the secret conspiracy running the liberal media – revealed!” I wish they had been right. If there were a secret conspiracy running the liberal media, they could all decide they wanted to raise awareness of racist police brutality, pick the most clear-cut and sympathetic case, and make it non-stop news headlines for the next two months. Then everyone would agree it was indeed very brutal and racist, and something would get done.

But as it is, even if many journalists are interested in raising awareness of police brutality, given their total lack of coordination there’s not much they can do. An editor can publish a story on Eric Garner, but in the absence of a divisive hook, the only reason people will care about it is that caring about it is the right thing and helps people. But that’s “charity”, and we already know from my blog tags that charity doesn’t sell. A few people mumble something something deeply distressed, but neither black people nor white people get interested, in the “keep tuning to their local news channel to get the latest developments on the case” sense.

The idea of liberal strategists sitting down and choosing “a flagship case for the campaign against police brutality” is poppycock. Moloch – the abstracted spirit of discoordination and flailing response to incentives – will publicize whatever he feels like publicizing. And if they want viewers and ad money, the media will go along with him.

Which means that it’s not a coincidence that the worst possible flagship case for fighting police brutality and racism is the flagship case that we in fact got. It’s not a coincidence that the worst possible flagship cases for believing rape victims are the ones that end up going viral. It’s not a coincidence that the only time we ever hear about factory farming is when somebody’s doing something that makes us almost sympathetic to it. It’s not coincidence, it’s not even happenstance, it’s enemy action. Under Moloch, activists are irresistibly incentivized to dig their own graves. And the media is irresistibly incentivized to help them.

Lost is the ability to agree on simple things like fighting factory farming or rape. Lost is the ability to even talk about the things we all want. Ending corporate welfare. Ungerrymandering political districts. Defrocking pedophile priests. Stopping prison rape. Punishing government corruption and waste. Feeding starving children. Simplifying the tax code.

But also lost is our ability to treat each other with solidarity and respect.

Under Moloch, everyone is irresistibly incentivized to ignore the things that unite us in favor of forever picking at the things that divide us in exactly the way that is most likely to make them more divisive. Race relations are at historic lows not because white people and black people disagree on very much, but because the media absolutely worked its *tuchus* off to find the single issue that white people and black people disagreed over the most and ensure that it was the only issue anybody would talk about. Men’s rights activists and feminists hate each other not because there’s a huge divide in how people of different genders think, but because only the most extreme examples of either side will ever gain traction, and those only when they are framed as attacks on the other side.

People talk about the shift from old print-based journalism to the new world of social media and the sites adapted to serve it. These are fast, responsive, and only just beginning to discover the power of controversy. They are memetic evolution shot into hyperdrive, and the omega point is a well-tuned machine optimized to search the world for the most controversial and counterproductive issues, then make sure no one can talk about anything else. An engine that creates money by burning the few remaining shreds of cooperation, bipartisanship and social trust.

Imagine Moloch looking out over the expanse of the world, eagle-eyed for anything that can turn brother against brother and husband against wife. Finally he decides “YOU KNOW WHAT NOBODY HATES EACH OTHER ABOUT YET? BIRD-WATCHING. LET ME FIND SOME STORY THAT WILL MAKE PEOPLE HATE EACH OTHER OVER BIRD-WATCHING”. And the next day half the world’s newspaper headlines are “Has The Political Correctness Police Taken Over Bird-Watching?” and the other half are “Is Bird-Watching Racist?”. And then bird-watchers and non-bird-watchers and different sub-groups of bird-watchers hold vitriolic attacks on each other that feed back on each other in a vicious cycle for the next six months, and the whole thing ends in mutual death threats and another previously innocent activity turning into World War I style trench warfare.

(You think I’m exaggerating? Listen: “YOU KNOW WHAT NOBODY HATES EACH OTHER ABOUT YET? VIDEO GAMES.”)

Who By Very Slow Decay

[*Trigger warning: Death, pain, suffering, sadness*]

I.

Some people, having completed the traditional forms of empty speculation – "What do you want to be when you grow up?", "If you could bang any celebrity who would it be?" – turn to "What will you say as your last words?"

Sounds like a valid question. You can go out with a wisecrack, like Oscar Wilde ("Either this wallpaper goes or I do"). Or with piety and humility, like Jesus ("Into thy hands, o Father, I commend my spirit.") Or burning with defiance, like Karl Marx ("Last words are for fools who haven't said enough.")

Let me puncture all your pleasant dreams. You'll probably never become an astronaut. You're not going to bang Emma Watson. And your last words will probably be something like "mmmrrrgggg graaaaaaaaaaaHAAACK!"

I guess I always pictured dying as – unless you got hit by a truck or something – a bittersweet and strangely beautiful process. You'd grow older and weaker and gradually get some disease and feel your time was upon you. You'd be in a nice big bed at home with all your friends and family gathered around. You'd gradually feel the darkness closing in. You'd tell them all how much you loved them, there would be tears, you would say something witty or pious or defiant, and then you would close your eyes and drift away into a dreamless sleep.

And I think this happens sometimes. For all I know, maybe it happens quite a lot. If it does, I never see these people. They very wisely stay far away from hospitals and the medical system in general. I see the other kind of people.

If you are like the patients I see dying, then here is how you will go.

You will grow old. When you were young, you would go to institutions and gradually gather letters after your name: BA, MD, PhD. Now that you are old, you do the same thing, but they are different institutions and different letters. Your doctors will introduce you to their colleagues as “Mary Smith, COPD, PVD, ESRD, IDDM”. With each set of letters comes another decrease in quality of life.

At first these sacrifices will be minor. The COPD means you have to breathe from an oxygen tank you carry around wherever you go. The PVD will prevent you from walking more than a few feet at a time. The ESRD will require three hours dialysis in a hospital or outpatient dialysis center three times a week. The IDDM will require insulin shots after every meal. Not fun, but hardly inconsistent with a life worth living.

Eventually these will add up beyond your ability to manage them on your own, and you will be sent off to a nursing home. This will seem like a reasonable enough idea, and sometimes it goes well. Other times it gives you freedom to develop a completely new set of morbidities totally unconstrained by what a person in any other situation could possibly be expected to survive.

You will become bedridden, unable to walk or even to turn yourself over. You will become completely dependent on nurse assistants to intermittently shift your position to avoid pressure ulcers. When they inevitably slip up, your skin develops huge incurable sores that can sometimes erode all the way to the bone, and which are perpetually infected with foul-smelling bacteria. Your limbs will become practically vestigial organs, like the appendix, and when your vascular disease gets too bad, one or more will be amputated, sacrifices to save the host. Urinary and fecal continence disappear somewhere in the process, so you're either connected to catheters or else spend a while every day lying in a puddle of your own wastes until the nurses can help you out. The digestive system isn't too happy either by this point, so you can either have a tube plugged directly into your stomach or just skip the middleman and have an IV line feeding nutrients into your bloodstream.

Somewhere in the process your mind very quietly and without fanfare gives up the ghost. It starts with forgetting a couple of little things, and progresses until you have no idea what's going on ever. In medical jargon, healthy people are "alert and oriented x 3", which means oriented to person (you know your name), oriented to time (you know what day/month/year it is), and oriented to place (you know you're in a hospital). My patients who have the sorts of issues I mentioned in the last paragraph are generally alert and oriented x0. They don't remember their own names, they don't know where they are or what they're doing there, and they think it's the 1930s or the 1950s or don't even have a concept of years at all. When you're alert and oriented x0, the world becomes this terrifying place where you are stuck in some kind of bed and can't move and people are sticking you with very large needles and forcing tubes down your throat and you have no idea why or what's going on.

So of course you start screaming and trying to attack people and trying to pull the tubes and IV lines out. Every morning when I come in to work I have to check the nurses' notes for what happened the previous night, and every morning a couple of my patients have tried to pull all of their tubes and lines out. If it's especially bad they try to attack the staff, and although the extremely elderly are *really* bad at attacking people this is nevertheless Unacceptable Behavior and they have to be restrained ie tied down to the bed. A presumably more humane alternative sometimes used instead or in addition is to just drug you up on all of those old-timey psychiatric medications that actual psychiatrists don't use anymore because of their bad reputation.

After a while of this, your doctors will call a meeting with your family and very gingerly raise the possibility of going to "comfort care only", which means they disconnect the machines and stop the treatments and put you on painkillers so that you die peacefully. Your family will start yelling at the doctors, asking how the hell these quacks were ever allowed to practice when for God's sake they're trying to kill off Grandma just so they can avoid doing a tiny bit of work. They will demand the doctors find some kind of complicated surgery that will fix all your problems, add on new pills to the thirteen you're already being force-fed every day, call in the most expensive consultants from Europe, figure out some extraordinary effort that can keep you living another few days.

Robin Hanson sometimes writes about how health care is a form of signaling, trying to spend money to show you care about someone else. I think he's wrong in the general case – most people pay their own health insurance – but I think he's spot on in the case of families caring for their elderly relatives. The hospital lawyer mentioned during orientation that it never fails that the family members who live in the area and have spent lots of time with their mother/father/grandparent over the past few years are willing to let them go, but someone from 2000 miles away flies in at the last second and makes ostentatious demands that EVERYTHING POSSIBLE must be done for the patient.

Your doctors will nod their heads and tell your family they respect their wishes. It will be a lie. Oh, sure, they will *carry out* the family's wishes, in terms of continuing to provide the care. But *respect*? In the cafeteria at lunch, they will – despite medical confidentiality laws that totally prohibit this – compare stories of the most ridiculous families. "I have a blind 90 year old patient with stage 4 lung cancer with brain mets and no kidney function, and the family is *demanding* I enroll her in a clinical trial from Sri Lanka." "Oh, that's nothing. *I* have a patient who can't walk or speak who's breathing from a ventilator and has anoxic brain injury, and the family is insisting I try to get him a liver transplant."

Every day, your doctors will meet with your family another time, and eventually, as your condition worsens and your family has more time to be hit on the head with a big club marked 'REALITY', they will start to relent. Finally, they will allow your doctors to take you off of the machines, and you will be transferred to Palliative Care, whose job I do not envy even though *every single palliative care doctor I have ever met is relentlessly cheerful and upbeat and this is a total mystery to me*.

And you will die, but not quickly. It takes time for the heart to give up, for the lungs to fill with water and stop breathing, for the toxic wastes to build up. It is generally considered wise for the patient to be on epic doses of morphine throughout the process, both to spare them the inevitable pain as their disease takes their course and to spare their family from having to watch them.

…not that they always do. It can take anywhere from a day to several weeks for someone to die. Sometimes your family wants to wait at the bedside for a week. But a lot of the time they have work and things to do. Maybe they live thousands of miles away. You haven't recognized them in years, you haven't spoken a coherent word in months, and even if for some reason your brain chose this moment to recover lucidity you're on enough morphine to be *well* inside the borders of la-la-land. A lot of families, faced with the prospect of missing work and school to sit by what's basically a living corpse day in and day out for weeks just to watch it turn into a non-living corpse, politely decline. I absolutely 100% cannot blame them.

There is a national volunteer program called No One Dies Alone. Nice people from the community go into hospitals to spend time with dying people who don't have anyone else there for them. It makes me happy that this program exists.

Nevertheless, this is the way many of my patients die. Old, limbless, bedridden, ulcerated, in a puddle of waste, gasping for breath, loopy on morphine, hopelessly demented, in a sterile hospital room with someone from a volunteer program who just met them sitting by their bed.

And let me just emphasize again, not everyone dies this way. I am hugely selection biased by my position in a hospital. But enough people die this way. I'm in a small community. There can't be too many deaths here. Of the ones there are, I see a lot of them. And they're not pretty.

*[**EDIT:** Just looked up statistics. Only about a quarter of old people die at home. The rest are split between hospitals (disproportionately ICUs), nursing homes, and hospices.]*

II.

Hospital poetry is notoriously bad.

I mean, practically all modern poetry is bad. Modern poetry by complete amateurs could be expected to be even worse. But hospital poetry is in a league all of its own as far as badness goes.

When I search “hospital poetry”, Google brings up examples like the following:

> Pain… searing
> Belly… throbbing
> There is no baby.
> There will be no baby.
> Endometriosis.

I feel bad making fun of it, because it is clearly heartfelt. This is part of the problem with hospital poetry. It is very heartfelt, whereas I think most popular poetry comes from people who have strong emotions but also some distance from them and a little bit of post-processing. And unfortunately doctors, who are on this decades-long quest to prove they are actual people with real feelings and not just arrogant robot-like people in white coats who know a very large number of facts about thyroiditis, just eat this sort of thing up.

But I’m not really complaining about those sorts of endometriosis poems. The ones I’m really complaining about are worse. The epitome of the genre I can’t find on Google, because it was presented as some kind of event at the hospital where I trained in Ireland. I don’t remember it, but let me just make up some doggerel approximately faithful to the spirit of the original:

> When my doctor told me that I had cancer
> I knew that despair was not the answer

It felt like the darkness was closing in
But to give up would have been a sin
Everyone here helped me so much
And nothing is like a helping hand’s touch
Thanks, Dr. Connell, and everyone in Cork
I really appreciate all your hard work

Doctors and nurses eat this kind of thing up and put it on shiny plaques that go on the walls of the hospital. (I suggest a wall near the gastroenterology unit, to expedite care for people who start vomiting.)

Wittgenstein said that “if anyone ever wrote a book of ethics, that really was a book of ethics, it would destroy all the other books in the world with a bang.” I’m not really sure what he meant. But if anyone ever wrote a book of hospital poetry, that really was a book of hospital poetry… well, I don’t know what would happen, but I bet it would be loud and angry, and that it wouldn’t be put on shiny plaques on anybody’s walls, except maybe the same people who hang Hieronymous Bosch paintings on their walls.

Am I calling hospitals hellish? Sure am. It has nothing to do with the decor, which has actually gotten much nicer in your newer hospitals until it’s hard to tell them apart from a stylish office building. It’s nothing to do with the staff, either – most doctors and some nurses seem pretty happy and trade banter around the water coolers like everyone else. It’s mostly the screams.

The screams are coming about 33% from the confused demented old people I mentioned, 33% from people having minor procedures performed without anaesthetics for one or another good reason, and 33% from people who just have very painful diseases (plus 1% from me sitting in the break room looking up examples of hospital poetry for this post).

They run the gamut of human screams. There are wordless shrieks. There are some angry screams, like “$#%! YOU GET ME OUT OF HERE!”. There are a lot of people screaming “SOMEBODY HELP ME!” And there are some religious screams, like “OH GOD!” or “JESUS HELP ME!” or “CHRIST NO!”.

When I first started working in hospitals, I would not only inevitably run over to these screams, but I would feel contempt and anger at the rest of the hospital staff who would just continue their daily routine. I soon learned better. Not only would I be unable to do anything – I can’t single-handedly cure their painful illness, or make their procedure go any faster, or explain to them that the year is 2013 and they’re no longer on their childhood farm in Oklahoma – but as soon as they saw me I would be the one they started screaming at and expecting to save them. The bystander effect, my last defense, disappeared. Sometimes I would make a stand by asking the nurse to increase their pain medication or something, and be politely told all the reasons why that was a bad idea from a medical perspective (pain medication has lots of side effects which doctors monitor carefully). In the end I would just slink out of the room, wishing I had never come in.

So the constant screams being completely ignored by a bunch of happy people going through their day is pretty hellish. But there’s also the bodies. Usually we are able to avoid thinking about people as bodies except to briefly note that certain people like Emma Watson are really hot. In a hospital, this filter disappears. Some people have gigantic swollen legs the size of your waist. Others have huge ulcerated sores all over. Still others have skin covered with the sorts of bacterial colonies you usually only see on a petri dish. And body sizes range from so thin that you can see their organs bulging out of their skin and use them as a grisly impromptu anatomy lesson, to so morbidly obese that you have to search through the fat folds to find body part you're looking for.

The senses are under constant assault. Smell is the worst. There are some people who can identify different infections by smell. *Pseudomonas aeruginosa* is supposed to smell fruity. *Gardnerella* is supposed to smell fishy. *Clostridium* is supposed to smell like the worst thing you can possibly imagine covered in feces and left to rot on a warm summer day.

But the other senses get their time too. The sight is vexed by flashing call lights. And the hearing is battered with incessant beeping from IV lines which have hard-coded alarms to alert doctors of critically important events such as "Look at me! I am an IV line!" The end result is something it would take a first-rate poet to describe. I'm tempted to nominate Oscar Wilde. He did a good job on prisons in *Ballad of Reading Gaol*, and I feel like the skill would transfer:

> He does not rise in piteous haste
> To put on convict-clothes,
> While some coarse-mouthed doctor gloats,
> and notes each new and nerve-twitched pose,
> Fingering a watch whose little ticks
> Are like horrible hammer-blows […]
>
> He does not stare upon the air
> Through a little roof of glass;
> He does not pray with lips of clay
> For his agony to pass;
> Nor feel upon his shuddering cheek
> The kiss of Caiaphas.

But after some more thought, I think I’m going to go with Wilfred Owen:

> If in some smothering dreams you too could pace
> Behind the wagon that we flung him in,
> And watch the white eyes writhing in his face,
> His hanging face, like a devil’s sack of sin;
> If you could hear, at every jolt, the blood
> Come gargling from the froth-corrupted lungs,
> Obscene as cancer, bitter as the cud
> Of vile, incurable sores on innocent tongues […]

Or better yet, if Oscar Wilde’s muse when he was writing *Reading Gaol* were to bear Wilfred Owen’s children, then those kids would be competent to write hospital poetry that was actually hospital poetry.

Dante would also be an acceptable choice.

III.

You may have read the excellent article How Doctors Die. If you haven’t, do it now. It says that most doctors, knowing everything I’ve just mentioned above, choose to die quickly and with very limited engagement with the health system.

I (and the doctors in my family whom I’ve asked) am pretty much like the doctors in the article. If I get a terminal disease, I want to wring what I can out of the few months of life I have left and totally avoid any surgery, chemotherapy, amputations, ventilators, and the like. It would be a clean death. It would be okay.

My big fear, though, is that I *won’t* get a terminal disease.

If I just start accumulating damage, growing more and more bedridden and demented and pain-riddling until I want out –

well, there won’t *be* a way out. If there’s not some very specific life-saving treatment that can be withdrawn, I’m stuck above ground, not just in the “unless I want to risk the danger and shame of suicide” way I am now, but – if I’m too debilitated to access means of suicide on my own – in an absolute way.

Even if my doctors and nurses and caretakers are sympathetic, my only legal option, without exposing *them* to jail time, is to starve myself to death – something both painful and difficult, and itself not really the way I want to go.

I was sitting in an ICU room yesterday where a patient’s body had just been brought out after their death. My attending was taking care of the paperwork in the other room, and I was sitting there reflecting, and I started thinking about what it would be like to die in that room. There was a big window, and it was a sunny day, and although I mostly had a spectacular view of the hospital parking lot, a bit further in the distance I could see a park full of really big trees. And I knew that if I were dying in that room my last thought would be that I wanted to be outside.

I think if I were very debilitated and knew I would die soon, I would want to go to that park or one like it on a very sunny day, surround myself with my friends and family, say some last words, and give myself an injection of potassium chloride.

(this originally read “morphine”, but just today the palliative care doctor at my hospital gave an impassioned lecture about how people need to stop auto-associating morphine with euthanasia, because it makes it really hard for him to offer morphine painkillers to patients who need them without them freaking out. So potassium chloride it is.)

This will never happen. Or if it did, it would be some kind of huge scandal, and whoever gave me the potassium chloride would be fired or something. But the people dying demented and hopeless connected to half a dozen tubes in ICU rooms aren’t considered scandals by anybody. That’s just “the natural way of things”.

I work in a Catholic hospital. People here say the phrase “culture of life” a lot, as in “we need to cultivate a culture of life.” They say it almost as often as they say “patient-centered”. At my hospital orientation, a whole bunch of nuns and executives and people like that got up and told us how we had to do our part to “cultivate a culture of life.”

And now every time I hear that phrase I want to scream. 21st century American hospitals do not need to “cultivate a culture of life”. We have enough life. We have life up the wazoo. We have more life than we know what to do with. We have life far beyond the point where it becomes a sick caricature of itself. We prolong life until it becomes a sickness, an abomination, a miserable and pathetic flight from death that saps out and mocks everything that made life desirable in the first place. 21st century American hospitals need to cultivate a culture of life the same way that Newcastle needs to cultivate a culture of coal, the same way a man who is burning to death needs to cultivate a culture of fire.

And so every time I hear that phrase I want to scream, or if I cannot scream, to find some book of hospital poetry that really is a book of hospital poetry and shove it at them, make them read it until they understand.

There is no such book, so I hope it will be acceptable if I just rip off of Wilfred Owen directly:

If in some smothering dreams you too could pace
Behind the gurney that we flung him in,
And watch the white eyes writhing in his face,
His hanging face, like a devil’s sack of sin;
If you could hear, at every jolt, the blood
Come gargling from the froth-corrupted lungs,
Obscene with cancer, bitter with the cud
Of vile, incurable sores on innocent tongues
My friend, you would not so pontificate
To reasoners beset by moral strife
The old lie: we must try to cultivate
A culture of life.

2013

- You’re probably wondering why I’ve called you here today
- Abraham Lincoln, Ape-Man
- Abraham Lincoln, Necromancer
- You cannot serve both God and Mammon, but you can serve Mammon by predicting who will serve God
- Google Correlate does not imply Google Causation
- 90% of all claims about the problems with medical studies are wrong
- Typical mind and gender identity
- I liked Lovecraft countless primaeval aeons before it was cool
- Future tense
- A defense of logical positivism (yes, really)
- Dungeons and Discourse Third Edition: The dialectic continues
- People! It turns out capitalism is actually a thing!
- Fermat’s Last Stand Coordination Thread
- MetaMed launch day
- Pope and Change: An Atheist’s Guide To Vatican Decision 2013
- Sleep
- Reactionary Philosophy In An Enormous, Planet-Sized Nutshell

- A Thrive/Survive Theory Of The Political Spectrum
- Yet Another Parable On The Importance Of Controlling Your Experiments
- We Wrestle Not With Flesh And Blood, But Against Powers And Principalities
- We Sail Tonight For Singapore
- Hitting Below The (Bible) Belt
- Good vibrations and subjective temporal granularity
- Fermat’s Last Stand: Game Rosters
- Valleys have two sides
- Annuntio vobis gaudium magnum: habemus officium!
- Not just a mere political issue
- My objections to “objectification”
- If opposites attract, why is my sister so pretty?
- Book Review: Infinite Jest (alternate title: “Look At Me! I Read Infinite Jest!”)
- The death of wages is sin
- My Thoughts On The First 40 Pages Of Alastair MacIntyre’s “After Virtue”, Expressed In The Form Of A Slightly Modified XKCD Comic
- Thank you for doing something ambiguously between smoking and not smoking
- A good pun is its own reword
- Against Anton-Wilsonism
- Lies, Damned Lies, And Facebook (Part 1 of ∞)

- Lies, Damned Lies, And Facebook (Part 2 of ∞)
- Investment and Inefficient Charity
- Poor Folks Do Smile…For Now
- Polyamory Is Boring
- RPG emergency!
- Whose Utilitarianism?
- Book Review: After Virtue
- Read History Of Philosophy Backwards
- Virtue Ethics: Not Practically Useful, Either
- Lizardman’s Constant Is 4%
- Last Thoughts On Virtue Ethics
- If a clod be washed away by the sea, Europe is the less
- Proving Too Much
- Things I Learned By Spending Five Thousand Years In An Alternate Universe
- Floor Employment
- Arguments About Male Violence Prove Too Much
- I do not understand “rape culture”
- Social Justice For The Highly-Demanding-Of-Rigor
- Search Terms That Have Led People To This Blog
- Implicit Association Tests and Suicidality
- Product Recommendation: FB Purity
- In Defense of Psych Treatment for Attempted Suicide
- Utilitarianism for Engineers, Part II
- Yellow Journalism In The War On Science
- Utility Weight Results

- **If It’s Worth Doing, It’s Worth Doing With Made-Up Statistics**
- **Selection bias and atheist stereotypes**
- **Ambijectivity**
- **Raikoth: Laws, Language, and Society**
- **Raikoth: Corruption, Priesthood**
- **Raikoth: Cities, Land**
- **Raikoth: Symbolic Beads**
- **Overheard in the Box**
- **Raikoth: Economics, Relationships**
- **Raikoth: History, Religion**
- **Index: Posts on Raikoth**
- **Overheard In The Box, Part II**
- **Newtonian Ethics**
- **Against Bravery Debates**
- **Can You Condition Yourself?**
- **Apart From Better Sanitation And Medicine And Education And Irrigation And Public Health And Roads And Public Order, What Has Modernity Done For Us?**
- **The Wisdom of the Ancients**
- **Going From California With An Aching In My Heart.**
- **The World Is Flat And Intensively Farmed**
- **Transhumanist Fables**
- **Fetal Attraction: Abortion and the Principle of Charity**
- **Hansonian Optimism**
- **Wisdom of the Ancients, Redux**
- **Literally Inconceivable: Contraceptives And Abortion Rates**
- **Hasta La Victorians Siempre**

- Moments Of Awakening
- All Debates Are Bravery Debates
- Lies, Damned Lies, And Facebook (Part 3 of ∞)
- Medical and Psychological Studies for June
- Biodjinnetics
- Arguments From My Opponent Believes Something
- The Virtue of Silence
- Can Atheists Appreciate Chesterton?
- The What-You'd-Implicitly-Heard-Before Telling Thing
- A New Chapter In The Codex
- Social Psychology Is A Flamethrower
- The Hospital Orientation
- Nature Is Not A Slate. It's A Series Of Levers.
- The Lottery of Fascinations
- I Aten't Dead
- Who By Very Slow Decay
- I Myself Am A Scientismist
- Holocaust Good For You, Research Finds, But Frequent Taunting Causes Cancer In Rats
- On first looking into Chapman's "Pop Bayesianism"
- Extreme Mnemonics
- Fermat's Last Stand: Soundtrack and Adventure Log
- Things I Don't Understand About Genetics (A Non-Exhaustive List)
- Fake Consensualism

- Fake Euthanasia Statistics
- Military Strikes Are An Extremely Cheap Way To Help Foreigners
- The Life Cycle of Medical Ideas
- Giving and Accepting Apologies
- Inefficient Hot Dogs
- Scientific Freud
- The Thin Blue Line That Stays Strangely Horizontal
- Prediction is very difficult, especially of the past
- The Trend Evaluation Of All Values
- Sleep – Now By Prescription
- The Anti-Reactionary FAQ
- Some preliminary responses to responses to the Anti-Reactionary FAQ
- List of Fictional Drugs Banned By The FDA
- The Poor You Will Always Have With You
- The Witching Hour
- List of Reasons Why Gunpowder Treason Should Be Forgot
- Lies, Damned Lies, And Facebook (Part 4 of ∞)
- Highlights From My Notes From A Forensic Psychiatry Conference Last Week
- Genetic Russian Roulette
- November Housekeeping Post
- A Letter I Will Probably Send To The FDA
- The Story Of Thanksgiving Is A Science-Fiction Story
- Empire/Forest Fire
- More Fictional Drugs Banned By The FDA

- The Logician And The God-Emperor
- A Something Sort Of Like Left-Libertarianism-ist Manifesto
- Papal PR
- Statistical Literacy Among Doctors Now Lower Than Chance
- Less Wrong, More Rite (II)
- Public Awareness Campaigns
- We Are All MsScribe
- [REPOST] A Christmas Poem
- Meetup: Southeast Michigan
- A Comment I Posted On “What Would JT Do?”
- The Spirit Of The First Amendment

You’re probably wondering why I’ve called you here today

Due to an oversight by the ancient Greeks, there is no Muse of blogging. Denied the ability to begin with a proper Invocation To The Muse, I will compensate with some relatively boring introductions.

The name of this blog is Slate Star Codex. It is almost an anagram of my own name, Scott S Alexander. It is unfortunately missing an “n”, because anagramming is hard. I have placed an extra “n” in the header image, to restore cosmic balance.

This blog does not have a subject, but it has an ethos. That ethos might be summed up as: charity over absurdity.

Absurdity is the natural human tendency to dismiss anything you disagree with as so stupid it doesn’t even deserve consideration. In fact, you are virtuous for not considering it, maybe even heroic! You’re refusing to dignify the evil peddlers of bunkum by acknowledging them as legitimate debate partners.

Charity is the ability to override that response. To assume that if you don’t understand how someone could possibly believe something as stupid as they do, that this is more likely a failure of understanding on your part than a failure of reason on theirs.

There are many things charity is not. Charity is not a fuzzy-headed caricature-pomo attempt to say no one can ever be sure they’re right or wrong about anything. Once you understand the reasons a belief is attractive to someone, you can go ahead and reject it as soundly as you want. Nor is it an obligation to spend time researching every crazy belief that might come your way. Time is valuable, and the less of it you waste on intellectual wild goose chases, the better.

It’s more like Chesterton’s Fence. G.K. Chesterton gave the example of a fence in the middle of nowhere. A traveller comes across it, thinks “I can’t think of any reason to have a fence out here, it sure was dumb to build one” and so takes it down. She is then gored by an angry bull who was being kept on the other side of the fence.

Chesterton’s point is that “I can’t think of any reason to have a fence out here” is *the worst reason* to remove a fence. *Someone* had a reason to put a fence up here, and if you can’t even imagine what it was, it probably means there’s something you’re missing about the situation and that you’re meddling in things you don’t understand. None of this precludes the traveller who knows that this was historically a cattle farming area but is now abandoned – ie the traveller who understands what’s going on – from taking down the fence.

As with fences, so with arguments. If you have no clue how someone could believe something, and so you decide it’s stupid, you are much like Chesterton’s traveler dismissing the fence (and philosophers, like travelers, are at high risk of stumbling across bull.)

I would go further and say that even when charity is uncalled-for, it is advantageous. The most effective way to learn any subject is to try to figure out exactly why a wrong position is wrong. And sometimes even a complete disaster of a theory will have a few salvageable pearls of wisdom that can’t be found anywhere else. The rationalist forum Less Wrong teaches the idea of *steelmanning*, rebuilding a stupid position into the nearest intelligent position and then seeing what you can learn from it.

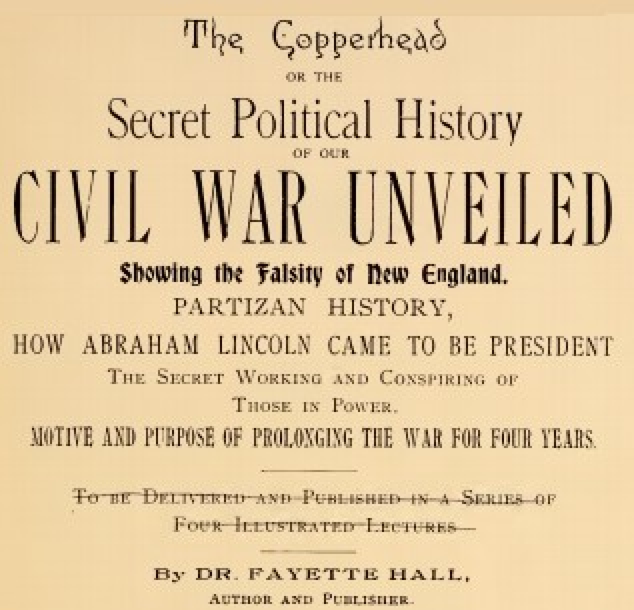
So this is the ethos of this blog, and we proceed, as Abraham Lincoln put it, “with malice toward none, with charity for all, with firmness in the right as God gives us to see the right.”

Abraham Lincoln, Ape-Man

Away with LiveJournal, and in with a new, sleeker-looking blog. A classier blog. A more mature blog. A blog where we're not afraid to ask the big questions.

Questions like: did Abraham Lincoln sign a demonic pact with the ghost of Attila the Hun?

We turn to one of my favorite historical books of all time, the late 19th/early 20th century bestseller *The Copperhead, or, The Secret Political History of our Civil War Unveiled, Showing The Falsity Of New England. Partizan History, How Abraham Lincoln Came To Be President, The Secret Working And Conspiring Of Those In Power. Motive And Purpose Of Prolonging The War For Four Years. ~~To Be Delivered And Published In A Series Of Four Illustrated Lectures~~*.



The Copperhead
OR THE
Secret Political History
OF OUR
CIVIL WAR UNVEILED
Showing the Falsity of New England.
PARTIZAN HISTORY,
HOW ABRAHAM LINCOLN CAME TO BE PRESIDENT
THE SECRET WORKING AND CONSPIRING OF
THOSE IN POWER.
MOTIVE AND PURPOSE OF PROLONGING THE WAR FOR FOUR YEARS.

~~TO BE DELIVERED AND PUBLISHED IN A SERIES OF~~
~~FOUR ILLUSTRATED LECTURES~~

By DR. FAYETTE HALL,
AUTHOR AND PUBLISHER.

You'd expect every line of that to be written in a different ridiculous font, and you'd be right.

The author of let’s-just-call-it-*Copperhead* was Dr. Fayette Hall (1825 – 1912), a Connecticut dentist, orator, amateur historian, and all-around colorful figure. He starts off by ticking off every box on the Crackpot Index, just in case the fonts didn’t get our attention:

> Here we have the desire and sentiment of the New York *Press*. That any person or speaker who exposes facts, or expresses an opinion at variance with their assumption, arrogance and dictation shall be suppressed by mob violence or any other means necessary to crush out free speech. None but the guilty or cowards fear the truth. Josh Billings said that he cared nothing for lies. It was only the truth that he feared, and we think it is the truth that is most feared by those sycophantic caterers for partisan and party power. I have the right to stand here to-night, alone, as did Bruno or Gallileo [sic] to demonstrate truth, and certainly as much right as the whole world has to perpetrate and perpetuate falsehood.

And what is this truth he must speak?

> His admirers would have the world believe that every act of Abraham Lincoln was that emanating from the highest intelligence and honest motives, at which this course of lectures is widely at variance, and would demonstrate that [his] only incentive to any exertion was the superstitious phantasies of a freak, aroused to action by Negro voudooism and spirit mediums.

Okay, back up. Let’s start at the beginning here.

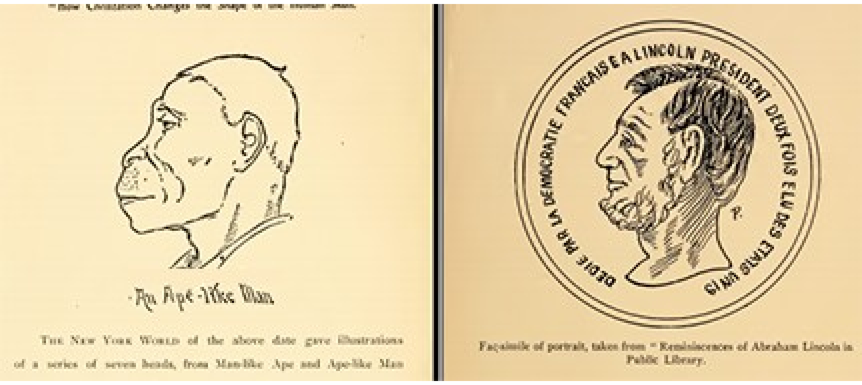
> As I am about to describe the characteristics, acts and doings of one of the most prominent characters of the

> nineteenth century, I will start with a physiological and phrenological description of the man.
>
> [According to a description given by his friend W.H. Herndan in 1865], he was about six feet, four inches high. He was thin, wiry, sinewy, raw-boned; thin through the breast to the back and narrow across the shoulders, inclining to the consumptive in build. His structure was loose and leathery; his body was shrunk and shriveled, having dark skin, dark hair — looking woe struck. When he walked he moved cautiously but firmly, his *long arms and hands swung down by his side. He walked with even tread, the inner side of his feet being parallel* … *undulatory, up and down* … His legs and arms were abnormally, unnaturally long, and in undue proportion to the balance of his body.

Swinging, unnaturally long arms. Funny-looking walk. Does that remind you of anything? Oh yes, we’re going there.

> Here we have a correct description of Abraham Lincoln’s physical and phrenological organization. We read in the first chapter of Genesis : And God said let the earth bring forth the living creature after his kind, cattle and creeping things, and beasts of the earth after his kind; and it was so; And God said: Let us make man in our own image, after our likeness : So God created man in his own image, in the image of God created him.
>
> Now, according to the description given by two of his most intimate friends and ardent admirers, which did Abraham Lincoln most nearly resemble, the higher and superior form of man, which we are told is the form of Deity, or according to the theory of Darwin and other celebrated scientists, the lower and inferior form of his great congener the ape?

We then get treated to this illustration, possibly my favorite in all of world literature:


"How Civilization Changes the Shape of the Human Man.
-An Ape-like Man.
The New York World of the above date gave illustrations of a series of seven heads, from Man-like Ape and Ape-like Man
DEDIE PAR LA DEMOCRATIE FRANCAISE A LINCOLN PRESIDENT DEUX FOIS ELU DES ETATS UNIS
P.
Facsimile of portrait, taken from "Reminiscences of Abraham Lincoln in Public Library.

As a later, wiser age would one day say – "Checkmate, liberals!"

He continues:

> I know of no instance where phrenology is better verified than in the head of Abraham Lincoln, and which although but little more than the world's representation of the ape-like man, he had enough brains located where, while it saved him from idiocy, left him a freak and gave him the qualification to obtain a position in which the dishonest fanaticism of the time aiding that of low cunning and secret intrigue which carried him to the White House and deluged the land in blood.

George Bush
TotallyLooksLike.com
Chimpanzee

Luckily, we as a nation have since moved beyond these sorts of superficial comparisons.

So if Lincoln was such a biological nincompoop, how was he so wildly successful at politics?

> I also assert and shall show that nearly every public act of Abraham Lincoln's so-called statesmanship was but an imitation of some historical barbarian predecessor or by direction of those same barbarian spirits purporting to instruct him through the many mediums whom he employed and consulted the same as the oracles were consulted and obeyed in the days of pagan Greece and Rome.

Are we finally going to get to hear his evidence for this remarkable assertion? No! We get on a tangent about Mary Todd Lincoln:

> When Lincoln first made advances to Miss Todd, who was very much his superior in every respect, position, education and culture she coolly snubbed him and said to an acquaintance. Who the devil do you think is after me now? It's that old long shanks, Lincoln, which remark shows that she was not particularly enamored of him.

> Shortly after thus expressing herself she discovered that a voudoo fortune teller had predicted for Lincoln that he was to become President of the United States. Whereupon Miss Todd became the suitor, or at least encouraged him to renew the attempt, when after quite an unpleasant experience she succeeded in becoming his wife.

Courting romance voodoo prophecy blah blah. Let’s get to the part where ghosts make Lincoln president. By page 37, Lincoln has connected with the Laurie family, a group of Washington socialites, and been invited to their seance:

> At a spiritual seance, the party workers, spiritualists and abolitionists, were met and Mr. Laurie purported to be entranced and controlled by the spirit of Robert Rantoul of Boston, and then and there under those circumstances Abraham Lincoln was announced and declared by those spiritualists and abolitionists to be the next President of the United States. At that time Horace Greeley was much interested in the spiritualistic theory and I was told had the Fox sisters at his house investigating the subject.

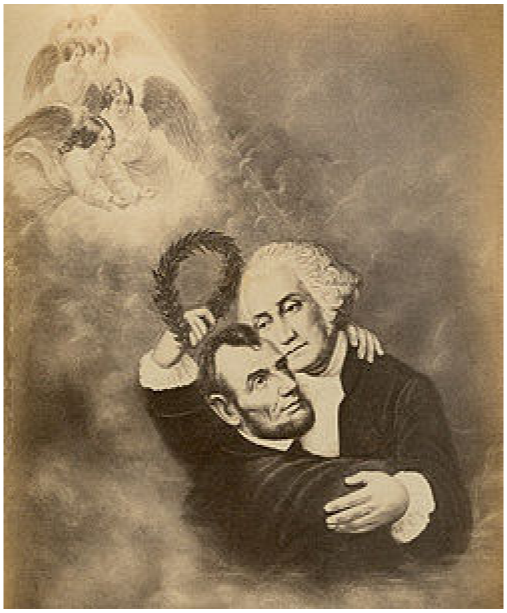
After the spiritualistic performance at the Laurie’s, Mr. Greeley was immediately informed of the result, when he became interested and went to Illinois to see Lincoln, where the arrangements were made, and Horace Greeley was the man who had Abraham Lincoln nominated as the dark horse of the Convention of 1860.”

All right, Lincoln. You’re President. What next?

> Lincoln had succeeded to the full promise of the voudoo prediction, when it would be supposed that under our form of government as established by our fathers and

> the Constitution that his ambition would have been sated. Not so. The President of the United States had established direct communication with the spirit world and his future was to be governed by that source. He had no further use for the Constitution, his inaugural oath, or his obligation to the government and the people. There were greater things yet in store for him. As he had succeeded in obtaining power, why should he not retain it, for a superstitious vain man's ambition knows no limit. It was through the Lauries that he established communication with the spirits and through them that Attila King of the Huns assured him of regal power and a dictatorship for life.

This is getting way too long for a first blog post. To be continued…



This entry has been kind of harsh on Lincoln. To make up for it, here is a picture of Abraham Lincoln hugging George Washington in Heaven.

…

Abraham Lincoln, Necromancer

It is quite uncontroversial among historians that Lincoln attempted to summon the dead. The only remaining question is to what degree he succeeded, and to what extent it influenced history.

The mid-1800s were the heyday of Spiritualism, a quasi-religious movement centered around seances summoning the spirits of the dead, who usually manifested through rapping noises or through taking control of the medium's speech. At the time it was a big deal, drawing in even normally skeptical souls: Wikipedia mentions Alfred Russel Wallace, who co-discovered evolution with Darwin, as a confirmed believer.

And then there was Lincoln. CSICOP is the closest thing to an official voice of skepticism in the United States; their past membership includes such luminaries as Carl Sagan and James Randi. Of Lincoln's spiritualism, they say:

> His wife, Mary Todd Lincoln, attended séances, one even held in the White House's Red Parlor. Lincoln stumbled upon the session, watching with curiosity. On another occasion, he accompanied Mary to a séance at a private home. The president, however, was a skeptic, and one biographer has suggested that his limited involvement in spiritualism was due to a desire "to protect his gullible wife"

What was this "seance in a private home" that even CSICOP agrees Lincoln attended? No other than the Laurie family mentioned in Fayette Hall's book! In fact, we have a letter from Jack Laurie, dated 1885, stating that "I have very often seen Mr. Lincoln at my father's house engaged in attending circles for spiritual phenomena, and generally Mrs. Lincoln was with him.”

So Lincoln very likely attended a seance at the Laurie house. What exactly happened during this event?

Here we begin to get into more controversial territory. For this I turn to the work of former Illinois state historian Jay Monaghan, particularly his 1941 article in *Journal of the Illinois State Historical Society* “Was Abraham Lincoln Really A Spiritualist?”

Monaghan has eleven sources that purport to describe Lincoln’s necromancy, but he only accepts three of these as conveying credible information about the Laurie seance. One of these is Nettie Colburn Maynard, a different medium who had been attending this seance as a guest rather than as the active participant. A second is Colonel S.P. Kase, who later went to the newspapers with his story. The third is our old friend Dr. Fayette Hall, author of *The Copperhead, or, The Secret Political History of our Civil War Unveiled, Showing The Falsity Of New England. Partizan History, How Abraham Lincoln Came To Be President, The Secret Working And Conspiring Of Those In Power. Motive And Purpose Of Prolonging The War For Four Years. ~~To Be Delivered And Published In A Series Of Four Illustrated Lectures~~*.

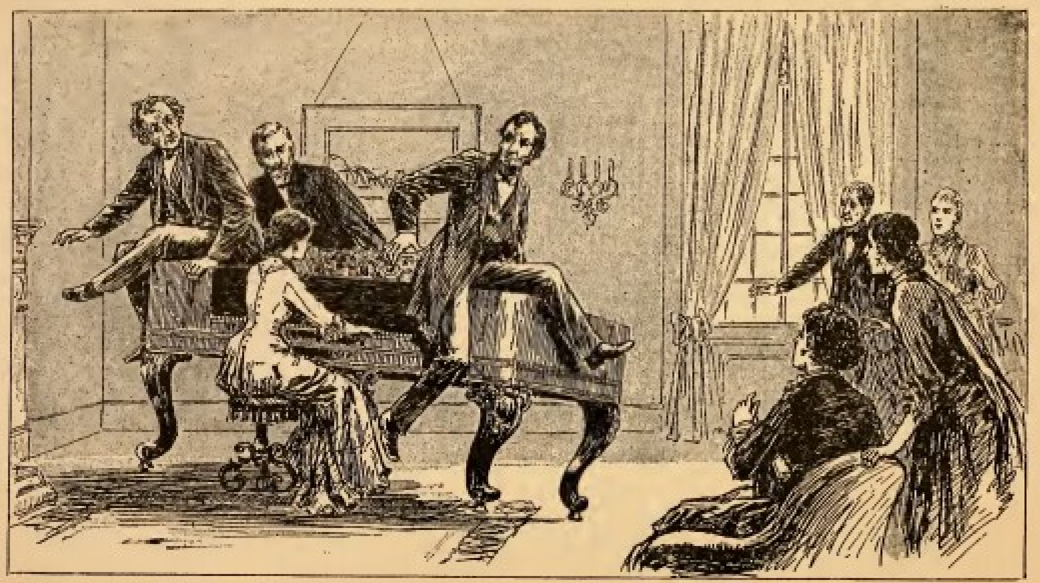
Why was Hall considered a remotely credible source? Because he had, in fact, spent a lot of the 1860s interviewing mediums and other members of DC’s spiritualist community, and in fact had even talked with President Lincoln about the latter’s spiritual experiences (a talk Monaghan describes as “a personal interview which Dr. Hall did not enjoy”.)

These three sources broadly agree on what happened at the Lauries' seance, but it is Ms. Maynard who gives the most detailed account and whom I will quote here. This passage is long, but worth it:

> It was at this seance that Mrs. Belle Miller gave an example of her power as a "moving medium," and highly amused and interested us by causing the piano to "waltz around the room," as was facetiously remarked in several recent newspaper articles. The true statement is as follows: Mrs. Miller played upon the piano (a three-corner grand), and under her influence it "rose and fell," keeping time to her touch in a perfectly regular manner. Mr. Laurie suggested that, as an added "test" of the invisible power that moved the piano, Mrs. Miller (his daughter) should place her hand on the instrument, standing at arm's length from it, to show that she was in no wise connected with its movement other than as agent. Mr. Lincoln then placed his hand underneath the piano, at the end nearest Mrs. Miller, who placed her left hand upon his to demonstrate that neither strength nor pressure was used. In this position the piano rose and fell a number of times at her bidding. At Mr. Laurie's desire the President changed his position to another side, meeting with the same result.
>
> The President, with a quaint smile, said, "I think we can hold down that instrument." Whereupon he climbed upon it, sitting with his legs dangling over the side, as also did Mr. Somes, S. P. Kase, and a soldier in the uniform of a major (who, if living, will recall the strange scene) from the Army of the Potomac. The piano, notwithstanding this enormous added weight, continued to rise and fall until the sitters were glad "to vacate the premises." We were convinced that there were no mechanical contrivances to produce the strange result,

> and Mr. Lincoln expressed himself perfectly satisfied that the motion was caused by some “invisible power;” and when Mr. Somes remarked, “When I have related to my acquaintances, Mr. President, that which I have experienced to-night, they will say, with a knowing look and wise demeanor, ‘You were psychologized, and as a matter of fact (versus fancy) you did not see what you in reality did see.'” Mr. Lincoln quietly replied, “You should bring such person here, and when the piano seems to rise, have him slip his foot under the leg and be convinced (doubtless) by the weight of evidence resting upon his understanding.”
>
> When the laughter caused by this rally had subsided, the President wearily sank into an arm-chair, “the old tired, anxious look returning to his face.” This never-to-be-forgotten incident occurred on the fifth day of February, 1863.

Just to make sure we’re all on the same page here, we have three witnesses all enjoying a personal familiarity with Lincoln, including a colonel of the United States Army, being quoted by the Illinois State Historian as claiming that the President of the United States was flung around a room by ghosts, while seated upon a flying piano. And that this same president then proceeded to *make a pun about the situation*.



The piano jumped so violently and shook us up so roughly that we were thankful to get off it. See page 36.

"Pics or it didn't happen" was a much less useful heuristic in the 19th century.

Do things get weirder? Oh yes they do. Directly after this performance, Lincoln gets introduced to Maynard herself, the other medium at the gathering. Doris Kearns Goodwin, the famous Lincoln biographer who wrote *Team of Rivals*, describes this meeting like so:

> Lincoln's lack of belief did not prevent him from enjoying the evening's entertainment. Nettie was an accomplished actress, ably mimicking the booming baritone of Daniel Webster or the frail voice of the Indian maiden. She spoke for an hour, channeling one voice and then another as she related historical episodes from the landing of the Pilgrims to the current war. Her oration, which carried a passionate abolitionist message, seemed to S.P. Kase the "grandest" he had ever heard. When the spirits left her, she departed as abruptly as she had arrived. All was silent for a while, then the President turned in his seat, threw his right long leg over the arm of his chair, and exclaimed "Was

> not this wonderful?” He seemed to have viewed Nettie’s performance with the same pleasure he derived from the the theater.

Nettie herself tells the story somewhat differently:

> Mr. and Mrs. Laurie and Mrs. Miller were duly presented. Then I was led forward and presented. He stood before me, tall and kindly, with a smile on his face. Dropping his hand upon my head, he said, in a humorous tone, “so this is our ‘little Nettie’ is it, that we have heard so much about?” I could only smile and say, “Yes, sir,’ like any school girl; when he kindly led me to an ottoman. Sitting down in a chair, the ottoman at his feet, he began asking me questions in a kindly way bout my mediumship; and I think he must have thought me stupid, as my answers were little beyond a “Yes” and “No”. His manner, however, was genial and kind, and it was then suggested we form a circle. He said, “Well, how do you do it?” looking at me. Mr. Laurie came to the rescue, and said we had been accustomed to sit in a circle and join hands; but he did not think it would be necessary in this instance. While he was yet speaking, I lost all consciousness of my surroundings and passed under control.
>
> For more than an hour I was made to talk to him, and I learned from my friends afterward that it was upon matters that he seemed fully to understand, while they comprehended very little until that portion was reached that related to the forthcoming Emancipation Proclamation. He was charged with the utmost solemnity and force of manner not to abate the terms of its issue, and not to delay its enforcement as a law beyond the opening of the year; and he was assured that it was to be the crowning event of his

> administration and his life; and that while he was being counseled by strong parties to defer the enforcement of it, hoping to supplant it by other measures and to delay action, he must in no wise heed such counsel, but stand firm to his convictions and fearlessly perform the work and full the mission for which he had been raised up by an overruling Providence. Those present declared that they lost sight of the timed girl in the majesty of the utterance, the strength and force of the language, and the importance of that which was conveyed, and seemed to realize that some strong masculine spirit force was giving speech to almost divine commands.
>
> I shall never forget the scene around me when I regained consciousness. I was standing in front of Mr. Lincoln, and he was sitting back in his chair, with his arms folded upon his breast, looking intently at me. I stepped back, naturally confused at the situation – not remembering at once where I was; and glancing around the group, where perfect silence reigned. It took me a moment to remember my whereabouts.
>
> A gentleman present then said in a low tone, “Mr. President, did you notice anything peculiar in the method of address?” Mr. Lincoln raised himself, as if shaking off his spell. He glanced quickly at the full-length portrait of Daniel Webster, that hung above the piano, and replied, ‘Yes, and it is very singular, very!” with a marked emphasis.

S.P. Kase, as quoted in Fayette Hall, describes the scene somewhat differently:

> When I entered the parlor, I found the President and Mrs. Lincoln there, together with a number of people whom I did not know. For a while the conversation was

> general, and nothing unusual happened. Suddenly a young girl [Nettie] walked the length of the drawing room to where President Lincoln sat. Stopping in front of him, the child, for she was nothing more, looked into his eyes with a peculiar rapt expression on her face.
>
> ‘President Lincoln,’ she said in a clear, loud voice. ‘The liberty of our nation conceived in the womb of oppression and born in the throes of revolution, can never be crowned with the wreath of immortality, until each and every human being in these United States is free. Slavery in any form must not exist. So says that spiritual Congress, which in this dread time of menace and danger to the Union, watches over and directs the affairs of the nation with even greater care and steadfastness of purpose than do the representatives chosen by the people. I have been chosen as their medium of communication with you. Before you can hope to bring about the great and lasting glory of this republic, you must make every man within its boundaries free. You must emancipate all the slaves by your pen, and your armies must indorse your action with the sword.’

Fayette Hall, who was actually at this same seance himself, concludes his description by saying:

> On September 22, 1862 [Lincoln] signed the proclamation, making the slaves free men … As the spirit communicating did not send in his card or announce his name, and the question arose as to who it might be, some one present seeing the portrait of Daniel Webster suggested that it might be his spirit, and it was at once decided that it was…I think it more likely to have been the spirit of the Negro pirate, Wansley,

> who with the pirate Gibbs, was hanged on gibbet, now Bedloes Island, in New York Bay.”

This is as far as Doris Kearns Goodwin, CSICOP, Colonel Kase, and the rest of the sane historical establishment will allow us to go. It’s pretty far. It at least *suggests* that the idea of the Emancipation Proclamation was communicated to Lincoln by a ghost. Who may or may not have been a pirate. One of the most important documents in American history was inspired by an ambiguously-piratical ghost. According to a US Colonel.

But Nettie Colburn Maynard, the medium involved, wrote a tell-all autobiography after the war. Some parts of it, like the levitating piano, we can verify through other sources. But Nettie goes further and says that after this event, Lincoln secretly called her to the White House for a sort of 19th century version of Project Stargate, an attempt to see whether Nettie’s powers could be valuable weapons in the Civil War.

There’s no evidence of this beyond Nettie’s own claims. But these should not be dismissed. We know for certain that Mrs. Lincoln believed in Nettie’s powers. And we have good testimony that the President was confronted with the whole piano incident that he could not explain. We know Lincoln was superstitious: he explicitly claimed it (“I was always superstitious” – Lincoln in letter to Joshua Speed), and often spoke of his prophetic dreams. So if you’re superstitious, and you’ve just seen a medium demonstrate seemingly supernatural powers, and you’re fighting a war, why *wouldn’t* you try to take advantage of this?

Further, Nettie’s book starts with an introduction by the publisher, who says that before he published it he did as much research as possible, first into the character of Ms.

Maynard [there follows a long list of respectable judges and writers and such who are quoted as saying they know Ms. Maynard and she is a very honest person who would never lie about this kind of thing] and into corroborating evidence. He then gives sworn affidavits by several Washington politicians and their families (as well as one famous author) who say they were at some of these same seances and everything described is true.

So, what does Nettie describe? First, that a skeptical Lincoln prepares a test for her. He hides one of his friends, General Dan Sickles, whom Nettie has never met, underneath a cloak. Then he asks Nettie to tell him who he is. Nettie channels the spirit of “the princess Pinkie”, who addresses the man as “General” and by the riddle “Crooked Knife”. It is decided that a crooked knife is sufficiently like a sickle to count this as a win, and Nettie gains the President’s trust.

Pictured: an all-knowing voice from beyond the veil of death

This incident is apparently on sufficiently firm ground to be included in the popular biography of Daniel Sickles, although I don’t have this book and can’t double-check this. We also have a claim by Mrs. Somes, the wife of a Congressman from Maine at the time, that she discussed this story with General Sickles and that he confirmed it was true.

Around this point, Lincoln orders his friend Isaac Newton (not that Isaac Newton!) to get Nettie a cover as some kind of low-level government worker dealing with agriculture. He then consults her several times during the war, asking her what the spirits recommend he do. In the interest of length I will include only one of these stories, although Nettie's book is full of them. This description is from Monaghan:

> In making a decision of such magnitude [as launching an important military campaign], a man as cautious as Lincoln is known to have been might have consulted every angle of opinion before embarking on the enterprise. Congressman Somes, according to report, was asked to bring Nettie Colburn to the White House for a secret session. "Consider the matter confidential," Lincoln warned the Congressman.

At the appointed time Mr. Somes and the medium were hurried upstairs to the executive chamber, where Mr. Lincoln and two gentlemen were waiting…Mr. Lincoln then gently stated he wished to give the medium an opportunity to display her rare gift. As the time for the seance approached all sat silently for a few moments and Nettie went under control. She recalled:

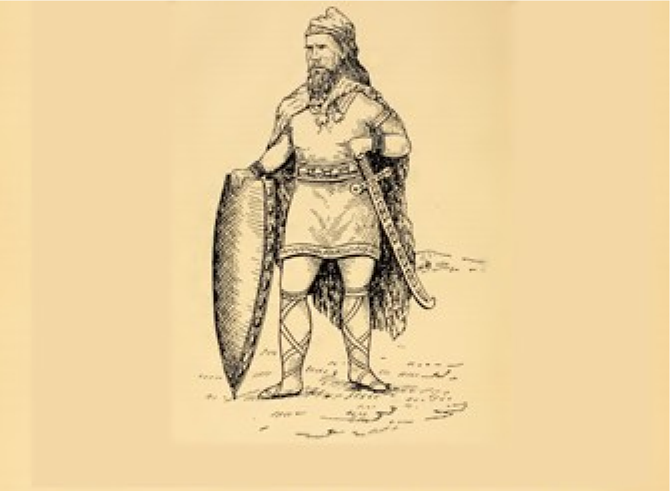
> "One hour later I became conscious of my surroundings, and was standing by a long table, upon which was a large map of the Southern States. In my hand was a lead pencil…I glanced around to note Mrs. Lincoln quietly conversing in another part of the room. The only remarks I heard were these: "It is astonishing," said Mr. Lincoln, "how every line she has drawn conforms to the plan agreed upon." "Yes," answered the older soldier, "it

> is very astonishing.” Looking up, they both saw that I was awake, and they instantly stepped back, while Mr. Lincoln took the pencil from my hand and placed a chair for me…
>
> Shortly afterwards, when about leaving, Mr. Lincoln said to us in a low voice, “It is best not to mention this meeting at present.” Assuring him of silence upon the question, we were soon upon our way.”

So what about Attila the Hun?

I have searched far and wide for any evidence that Fayette Hall had a source for this. Hall, who was at some of these seances and obsessively interviewed attendees at others, seems to be infuriatingly able to combine meticulous research with total flights of fancy, and it appears that the Attila claim is one of the latter. As far as I can tell, it just *seems to him* that Abraham Lincoln’s actions were the *sort of thing* someone who had signed a demonic pact with Attila the Hun *might do*.

So in the end, what do we have? Pretty high-quality evidence, evidence good enough to satisfy all of Lincoln’s most famous biographers and historians not to mention frickin’ CSICOP, that Lincoln attended seances. Good evidence that these seances included a piano that was made to appear to levitate, and that Lincoln sat on top of it. Weak but still plausible evidence that Lincoln then, in secret, tried to summon spirits to help him win the Civil War. And, alas, no evidence at all that Attila the Hun was involved in any way.

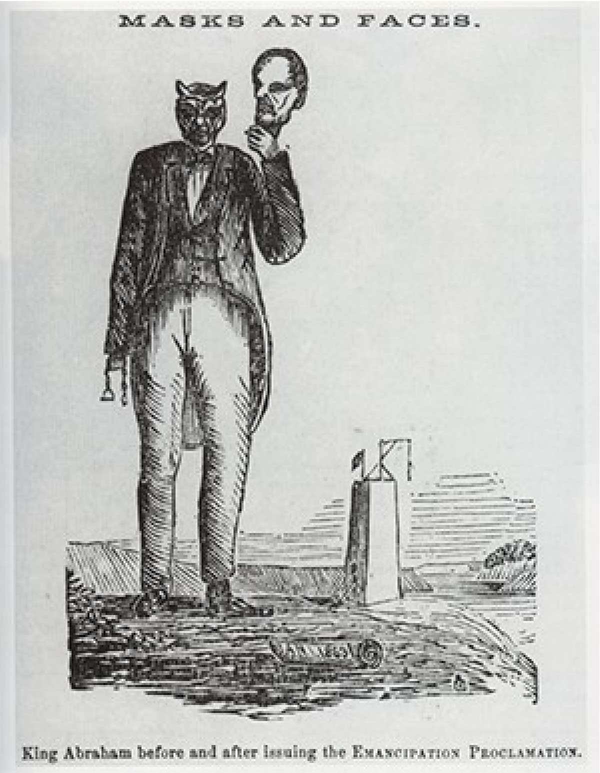


Attila, King of the Huns, whose spirit, through the Medium, assured President Lincoln of regal power and a dictatorship for life.

Except this very suspicious-looking picture plus caption!

Not that I believe there were any actual ghosts involved. Levitating pianos and speaking in booming Daniel-Websteresque voices is well within what we know real mediums can do – the former through trickery that was amply revealed by Harry Houdini’s campaign against spiritualism, the latter through some kind of dissociative state that modern psychology doesn’t really understand that well but which is probably also behind things like dissociative identity disorder. I want to think that Ms. Colburn was generally honest but mentally frail and prone to exaggeration, while Ms. Miller, the medium who levitated the piano, was otherwise. But that’s just my personal bias seeping in.

Still, I think this has been a good exercise in charity. We went from something so outrageous that no one could possibly believe it – Fayette Hall’s claim that Lincoln made a demonic pact with Attila the Hun – and through investigating it, discovered something extremely fascinating: that Lincoln attended seances and may even have tested spiritualism as a possible intelligence source during the war. Sure, the original claim may have been false. But looking into it was sure a heck of a lot more fun then “Ghosts don’t exist, you’re all idiots, shut up and go away.”


MASKS AND FACES.
King Abraham before and after issuing the EMANCIPATION PROCLAMATION.

Still uninvestigated: was “Lincoln” really just the Devil wearing an Lincoln mask?

You cannot serve both God and Mammon, but you can serve Mammon by predicting who will serve God

The good news is that there are some opportunities to make a couple bucks arbitraging papal candidates on InTrade. The bad news is that if you try this, you will be arrested, then damned to eternal hellfire.

The arrests will come from the US government, who continue to ban prediction markets in order to, according to the Commodity Futures Trading Commission, “protect market integrity”. To me this sort of seems like a case of locking the barn door after the horse has fled, traveled to Boston, been accepted to MIT, earned a degree in high energy physics, constructed a revolutionary FTL drive, and warped to the outer fringes of the Andromeda Galaxy – but such is the noble and impartial majesty of the law.

The damnations will come from the Vatican. In 1591, Pope Gregory XIV issued a papal bull banned all betting on papal elections under pain of excommunication. This has never really been enforced, but as far as I know it’s still in effect.

Luckily not everyone is afraid of Pope Gregory, and the Papal betting markets continue to function. As one would expect of something at the intersection of Catholicism and gambling, the most significant are in Ireland, especially at paddypower.com, which is offering odds on candidates as unlikely as Richard Dawkins, Bono, and Father Dougal Maguire – who, as you may recall, is fictional.

PaddyPower is also significant for their special bonus offer to refund all losing bets if the next Pope is black. There’s about a 1/3 chance this will be the case – two black cardinals are considered likely candidates – which means it should in theory be possible to come out ahead on net by betting on the leading white guy.

I can think of more complicated plans to get guaranteed free money by getting several people together to bet on all the leading white guys, plus hedging against InTrade's market that the next Pope will be African, but the maximum refundable bet on PaddyPower is 25 pounds, making this more trouble than it's worth, which I guess is what the betting markets are counting on.

Another so-close-yet-so-far route to free money is betting on the next Pope's name. PaddyPower gives even odds in favor of the next Pope's name being Pope Peter II, which is of course ridiculous. Unfortunately, I don't see any way to bet *against* this proposition, which is probably by design.

If InTrade set up a next papal name betting market, that would allow much better calculation of these Peter II odds. Unfortunately, as an American I can't easily bet in InTrade, and the only person I know who's done the necessary paperwork to avoid the civil authorities, Leah, is also the only person I know who will be genuinely deterred by Pope Gregory XIV's papal bull.

I continue to think there ought to be some way to money-pump this "next Pope" thing, but I can't figure out what it is.

Google Correlate does not imply Google Causation

I need something sexy, something to lure new readers to this new blog and get them excited. So let's talk about statistical correlations. No, wait, *failed* statistical correlations!

Google Correlate is a nifty new Google product that takes data sets and finds search terms that correlate with them. For example, if you set it to "correlate over time" and enter a data set of average US temperature, it might return the search term "skiing", because people are most likely to ski when it's cold and so searches for skiing will be correlated with temperature. You can also just enter in Google search terms and see what *other* search terms they're correlated with.

The results seem to fall into two categories: obvious and nonsensical.

Ones with clear time patterns are obvious. If you enter in skiing, you'll get "how to ski", "buy skis", "snowboarding", "ski resorts", and the like. If you enter in a news trend that was only popular at one point, you'll get both related terms and other news trends only popular at that one point – for example, "school shooting" brings up "jan berenstain", not because the Berenstain Bears books secretly cause school shootings (…one hopes) but because she died the same week as a relatively big one and so people were searching them around the same time.

Things that don't have obvious time patterns seem to bring up results that are both nonsensical and very-very convincing-looking. The worst are diseases.

This is Google Correlate's result for *heart attack*. It matches it to "pink lace dress" with a correlation of .88 (for comparison, a study comparing cigarette use vs. lung cancer rates across different social groups found a correlation of .71).

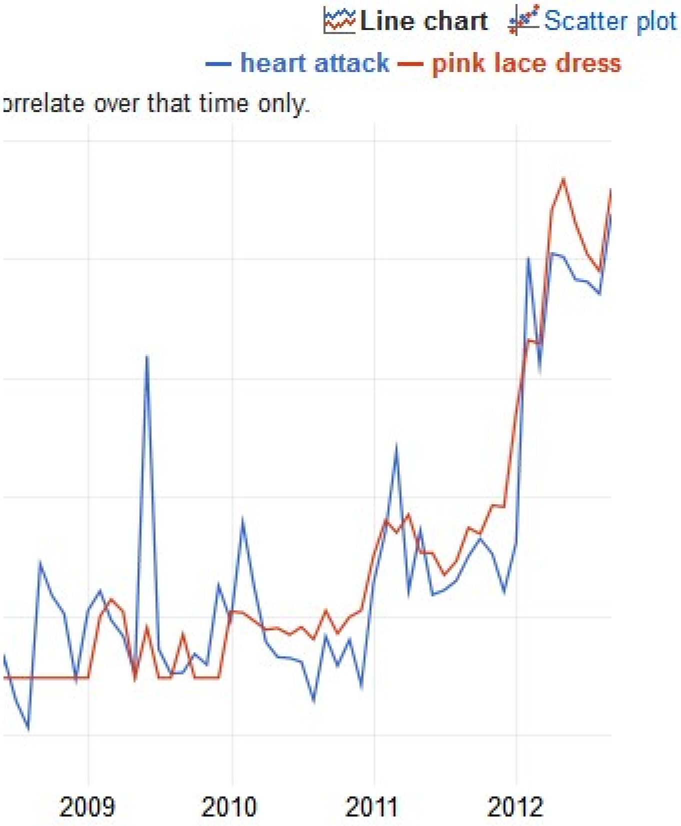


Figure 1: Correlation between interest in heart attacks and in pink lace dresses, by time.

As far as I can tell, this is just an artifact of Google having lots and lots of search terms and you would expect some of them to be heavily correlated by mere coincidence.

Google also has a correlate-by-state feature. This one has even weirder results for heart attack, like "can you get a" and "is it a" (note that these are the entire search terms). I understand that "is it a heart attack" is a reasonable question, but I don't understand who would just enter that phrase into Google and hope it would figure it out. I'm kind of imagining someone having a heart attack going on Google, typing as far as "is it a…" and then falling over dead, but I assume the real explanation is more prosaic, like someone expecting autocomplete to work but being disappointed.

Google's state-by-state feature seemed potentially really exciting to me. I wrote a while back on the effect of parasite load, and I had the dataset lying around with different states ranked on different metrics. I entered the data for parasite load and got the following search terms: "Toy Johnson", "Bernie Mac", "booty models", "Harvey suits", "Beyonce clothing line".

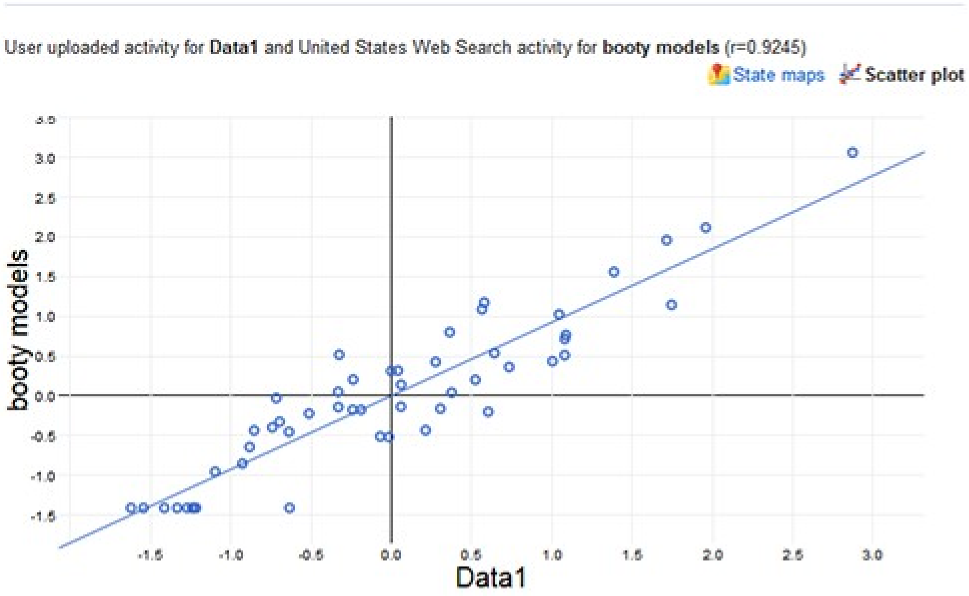


Figure 2: Correlation between parasite prevalence and interest in booty models, by state.

I didn't actually know what most of these were (I kinda thought Bernie Mac was a real estate conglomerate, which turns out to be false) but upon closer investigation they are all black people or Stuff Black People Like. So I think what's happening here is that the high-parasite load states are all in the South and relatively poor with low access to health care, which also selects for black people. This obviously has significant implications for the study's attempt to determine that high parasite load causes certain social trends.

My next thought was "if I multiply this data set by negative one, I will have an objective pipeline to figuring out Stuff White People Like. That sounds interesting." So I tried it, and my results were: "black albino", "shake that eminem", "tony hawk pro skater", and "green day time of your life". I was sort of hoping that "Black Albino" was the name of a band or something (it would actually be a pretty good one) but no, it turns out white people are just fascinated with the idea of black albinos. White people are kind of weird.

Figure 3: A black albino. Happy now, white people?

But let's keep going through the state-by-state data set. My next Big Social Statistic was "importance of family ties, by state". States with higher family ties were more likely to search for: "how to swim", "composition book", "noni juice", "muscle men", "girl kiss", "Toyota Tacoma 2008".

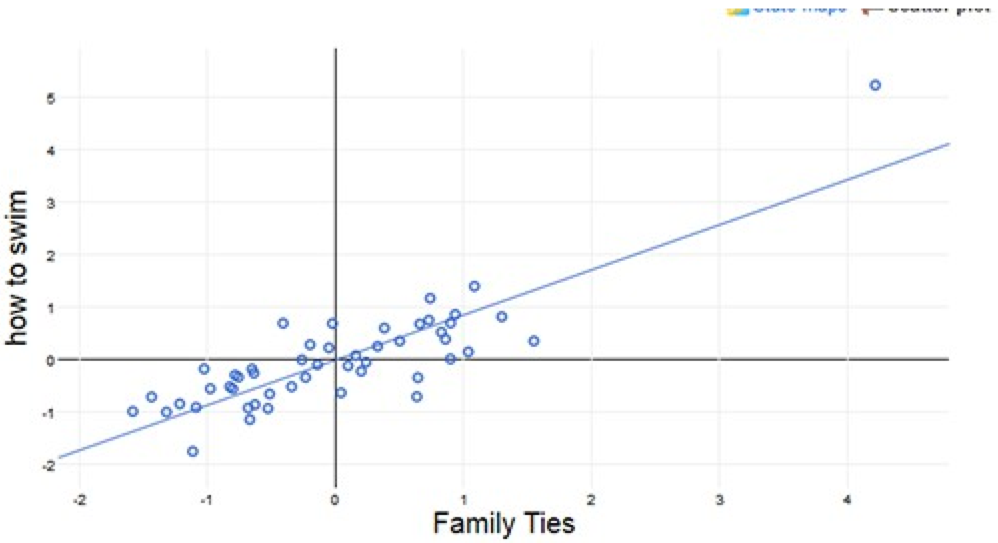


Figure 4: Correlation between strength of family ties and interest in swimming, by state.

A lot of these seem related to physical fitness, or ruggedness (the Tacoma seems to be a very sporty, rugged car), or masculinity. I’m not really sure what to make of this.

The last Social Science Statistic in the dataset was Religiosity, which correlated with the following search terms: “Christmas themes”, “rotary cutter”, “Honda rebel 250”. Christmas themes seems sort of plausible. I dunno about the rest.

So as far as I can tell Google Correlate is not very interesting. It doesn’t reveal any deep connections between concepts, or even guess what concept my dataset came from to begin with. For something potentially so powerful this is disappointing.

I can think of two possible uses for it. The first is as a sanity check to make sure your data aren’t completely confounded. If you think you’re measuring average number of roof tiles per house or something, and your data’s Google Correlate results come back with Toy Johnson and Beyonce clothing, you’re probably just measuring race and for some reason different races have different numbers of roof tiles on their houses. Which means if you think you’ve found a correlation between roof tiles and something fascinating like voting record, you’re probably just being confounded by race. This is a real problem in a lot of studies.

The second is as a cheap hack for *creating* datasets. I entered “Jesus” in and got a state by state list of who searched for Jesus. It looked a lot like my state-by-state map of religiousity. The correlates were all things like “Apostle”, “Paul”, “preaching”, and for some reason “Abednego”, who is a very minor Biblical character who has no business being in the top ten correlates of Jesus at all. If you wanted to make a cheap map of state-religiosity in order to correlate to parasite load or whatever, Google Trends seems like a plausible method.

On the other hand, I tried to see if I could recreate their state map of parasite load. I asked it to correlate “metronidazole”, a medication commonly used in the treatment of parasitic diseases, on the grounds that people with parasites would be prescribed metronidazole and then look it up to see if it was safe. The result looked only a little like my map of state-by-state parasite data, and the number one correlated search term (r = .89) was “Is Lil’ Wayne gay?”

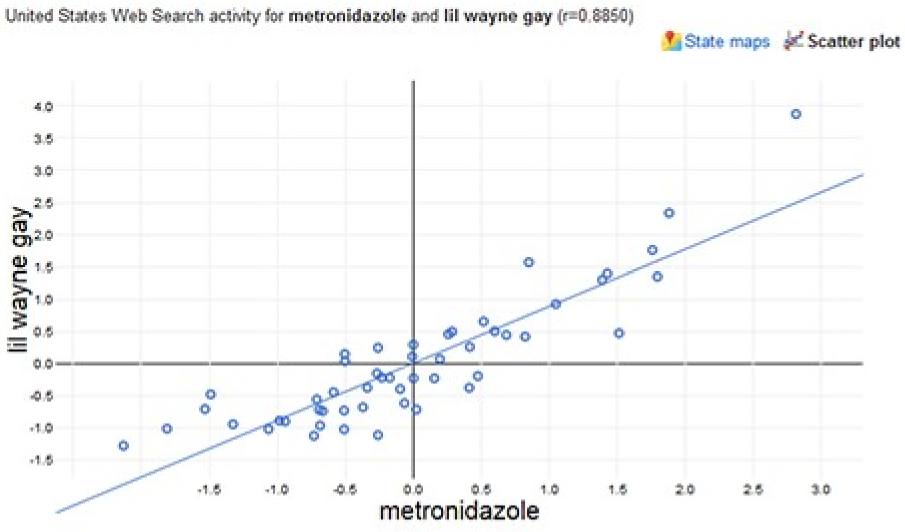


Figure 5: Correlation between curiosity over Lil' Wayne's sexual orientation and interest in the anti-parasitic medication metronidazole. Whatever my case was, I hereby rest it.

So if nothing else, this exercise has proven my suspicion that the sort of people who worry about whether Lil' Wayne is gay are, in fact, crawling with parasites.

90% of all claims about the problems with medical studies are wrong

I have frequently heard people cite John Ioannidis' apparent claim that "90% of medical research is false".

I think John Ioannidis is a brilliant person and I love his work and I think this statement points at a correct and important insight. But as phrased, I think this particular formulation when not paired with any caveats creates just a *little* more panic than is warranted.

Before I go further, Ioannidis' evidence:

He starts with simple statistics. Most studies are judged to have "discovered" a result if they reach $p < 0.05$, that is, if there is 5% probability or less the findings are due to mere chance (this is the best case scenario, where the study is totally free from bias or methodological flaws).

Suppose you throw a dart at the Big Chart O' Human Metabolic Pathways and supplement your experimental group with the chemical you hit. Then ten years later you come back and see how many of them died of heart attacks.

Most chemicals on the Big Chart probably don't prevent heart attacks. Let's say only one in a thousand do. Maybe your study will successfully find that 1/1000. But the 999 inactive chemicals will also throw up about 50 (999 * 5%) false positives significant at the 5% level. Therefore, even if you conduct your study perfectly, and it shows a significant decrease in heart attacks, there's about a 98% chance it's false.

One would hope medical scientists plan their studies with a little more care than throwing a dart at a metabolic chart. Yet many don't; a lot of genetic research is conducted by checking every single gene against the characteristic of interest and seeing if any stick. And even when scientists have well-thought out theories, the inherent difficulty of medicine means they probably have less than a 50-50 chance of being right the first time, which means a 5% significance level has a less than 5% predictive value.

And this isn't even counting publication bias or poor methodology or conflicts of interest or anything like that.

Disturbingly, this problem seems to be borne out in empirical tests. Amgen Pharmaceuticals says it repeated experiments in 53 important papers and was only able to confirm 6. And Ioannidis himself did a re-analysis which is quoted as finding that "41% of the most influential studies in medicine have been convincingly shown to be wrong or significantly exaggerated."

So I don't at all disagree with the general consensus that this is a huge problem. But I do disagree with the following statements:

1. 90% of all medical research is wrong
2. A given study you read, or your doctor reads, is 90% likely to be wrong.
3. 90% of the things doctors believe, presumably based on these medical findings, is wrong.
4. This proves the medical establishment is clueless and hopelessly irrational and that two smart people working in a basement for five minutes can discover a new medical science far better than what all doctors could have produced in seventy years.

Is 90% of all medical research wrong?

As far as I can tell, there is no source at all for the 90% figure. I can't find it in any of Ioannidis' studies and indeed they contradict it. His table of predictive values of different studies doesn't have *any* entries that correspond to 90% ("underpowered exploratory epidemiological study" is relatively close with 88%, but this is just for that one type of study, which is known to be especially bad). *The Atlantic* sums it up as:

> His model predicted, in different fields of medical research, rates of wrongness roughly corresponding to the observed rates at which findings were later convincingly refuted: 80 percent of non-randomized studies (by far the most common type) turn out to be wrong, as do 25 percent of supposedly gold-standard randomized trials, and as much as 10 percent of the platinum-standard large randomized trials.

Notice which number is conspicuously missing from that excerpt.

Now another study of his did show that in 90% of studies with very large effect sizes, later research eventually found the effect size to be smaller, but this was out of a pool of studies *specifically selected* for being surprising and likely to be false. I don't think it's the source of the number and if it were that would be terrible.

As far as I can tell, this started from a quote in an *Atlantic* article on Ioannidis which included the line "he charges that as much as 90 percent of the published medical information that doctors rely on is flawed". This then got turned into the title of a *Time* article "A Researcher's Claim: 90% of Medical

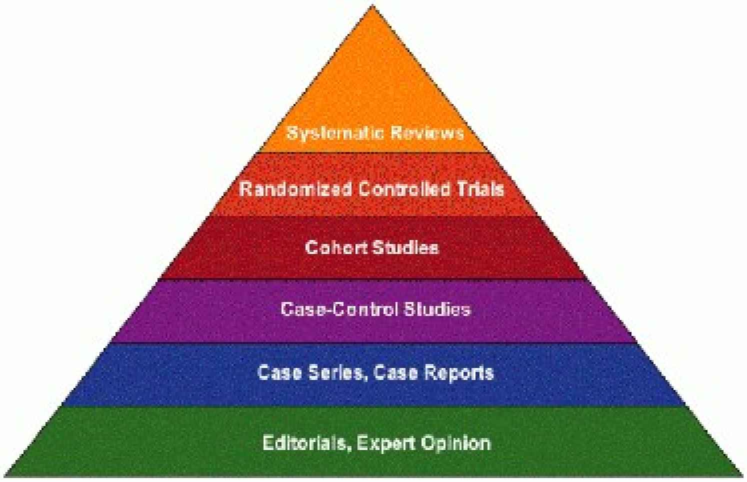
Research Is Wrong”, which itself got perverted to 90% of Medical Research Is Completely False.

So an unsourced quote that *up to* 90% of studies are *flawed* has somehow turned into a rallying cry that it has been *proven* that *at least* 90% of studies are *false*. To take this seriously we would have to believe that the numbers for *all research* are the same as the numbers for the poorly conducted epidemiological studies or the studies specifically selected for surprising results. I guess having a nice round number is good insofar as it makes the public pay attention to this field, but as far as actual numbers go, it’s kind of made up.

Is any given study you read, or your doctor reads, 90% likely to be wrong?

But let’s take the above number at face value and say that 90% of medical studies are wrong. Fine. Does that mean the last medical study you read about in Scientific American, or that your doctor used to recommend you a new drug, is wrong?

No. Let’s look at the Medical Evidence Pyramid.


Systematic Reviews
Randomized Controlled Trials
Cohort Studies
Case-Control Studies
Case Series, Case Reports
Editorials, Expert Opinion

The medical evidence pyramid is much like all pyramids, in that the bottom levels are infested with snakes and booby traps and vengeful medical evidence mummies. It’s only after you reach the top few levels that you get the gold and jewels and precious, precious mummy powder.

This plays out in the same table of Ioannidis’ speculations we saw before. While an *in vitro* study of the type used to identify possible drug targets might have a positive predictive value of 0.1%, a good meta-analysis or great RCT has a positive predictive value of 85%; that is, it’s 85% likely to be true.

There are only two reasons someone might hear about the studies on the snake-infested bottom levels of the pyramid. Number one, that person is a specialist in the field who is valiantly trying to read through the entire niche medical journal the paper was published in. Or number two, the study found something incredible like DONUTS CURE CANCER IN A SAMPLE OF THREE LAB RATS!!! and the media decided to pick up on it. Hopefully everyone already ignores studies of the DONUTS CURE CANCER IN A SAMPLE OF

THREE LAB RATES!!! type studies; if not, there’s really not much I can say to you.

But *most* of the medical results that you hear about are the ones that get published in important journals and are trumpeted far and wide as important medical results. These are closer to the top of the pyramid than to the bottom. They’re usually big expensive studies on thousands of people. Since the universities, hospitals, and corporations sponsoring them aren’t idiots, they usually hire a decent statistician or two to make sure that they don’t spend $300,000 testing something only to have a letter to the editor of the NEJM point out that they forgot to blind their subjects so it’s totally worthless. And finally, in many cases you would only run a study that big and expensive if you had something plausible to test – you’re not going to spend $300,000 just to throw a dart at the Big Chart O’ Human Metabolic Pathways and see what happens.

So these studies that people actually hear about are bigger, they have more incentives to get their methodology right, and they’re testing propositions with high plausibility. How do they do?

I said above that one of Ioannidis’ studies was frequently quoted as saying that “41% of the most influential studies in medicine have been convincingly shown to be wrong or significantly exaggerated.”

This is from a great study I totally endorse, but the 41% number was maximized for scariness. If I wanted to bias my reporting the other direction, I could equally well report the same results as “Only about 5% of influential medical experiments with adequate sample size have later been contradicted.”

How? Ioannidis got his result by taking all medical studies with over 1000 citations in the ’90s, of which there were 49. Of these, 4 were negative results (ie “X doesn’t work”) so he threw them out. This is the first part I think is kind of unfair. Yes, negative results aren’t as sexy as positive results, but they’re still influential medical research, and if Ioannidis is quoted as saying that X% of medical findings are later contradicted when he means that X% of positive medical findings are, that’s not quite fair.

Annnnyway, of the 45 famous studies with positive findings, 11 didn’t really get tested and so we don’t know if they’re right or wrong. Eliminating these is *also* a potential bias, because we expect that studies which seem sketchy are more likely to be replicated so people can find out if they’re actually right. Ioannidis quite rightly set himself a higher bar by not eliminating them, but the quote about 41% of studies being wrong *does* seem to have gone back eliminated them – at least that’s the only way I can make the study numbers add up to 41% (the numbers given in the study actually say 32% of these studies failed to replicate).

So our 41% number is based off of 34 studies, best described as “34 famous medical studies that found positive findings ie the least believable kind of finding, plus were suspicious enough that someone wanted to replicate them”.

Of these 34 studies, 7 were outright contradicted. Bad? Definitely. But for example, one of them was a study with a sample size of nine patients. Another study may well have been correct, but the results were interpreted wrongly (it said that estrogen decreased lipoprotein levels which everyone assumed meant decreased heart disease, but in fact later studies found increased heart disease without necessarily disproving the lipoprotein levels). Five of the six others were epidemiological trials, firmly on the middle of the pyramid. Only two of these contradicted studies were a true experiment with a sample size of >10.

(even here, I am sort of skeptical. Three of these disproven studies, two epidemiologicals and an experimental, purported to show Vitamin E decreased heart disease. Then a single better trial showed that Vitamin E did not decrease heart disease. While recognizing the last trial was better, it does seem like something more complicated is going on here than “all three of the earlier trials were just wrong”, and I’ve recently been convinced antioxidant research is a huge minefield where tiny differences in protocol can cause big differences in results. But fine, let’s grant this one and say there were two outright-contradicted experiments.)

So aside from the seven that were outright wrong, another seven were listed as “overstating their results”.

There are a couple of problems that bothered me here. One of them was that Ioannidis decided to count studies as contradicting each other if relative risk in one study was half or less than in the other study, “regardless of whether confidence intervals might overlap or not”. So even if a study effectively said “Here is a wide range of possible results, we think it’s about here in the middle but our research is consistent with it being anywhere in this range”, if another study got somewhere else in that range, the first study was marked as “exaggerated”.

The second problem is, once again, poor studies versus poor interpretations. Ioannidis cites as an example of an exaggerated study one lasting a year and showing that the drug zidovudine helped slow the progression of HIV to AIDS. It concluded that giving HIV patients long-term zidovudine was probably a good idea. A later study lasted longer, and said that yes, zidovudine worked for a year, but then it stopped working. Because the earlier study had suggested longer-term zidovudine, it was marked as “exaggerated results”, even though the results of both studies were totally consistent with one another (both found that zidovudine worked for the first year). This is probably of little consolation to AIDS patients who were treated with a useless drug, but it seems pretty important if we’re investigating study methodology.

So the way I got my 5% figure was to take the two experimental studies with decent sample sizes which were actually contradicted and compare them to the 38 large experimental studies total that started the experiment.

So this suggests that if you see a large experimental study being trumpeted in the medical literature, the chance that it will be found to be totally false (as opposed to true but exaggerated) within ten years or so is only about 5% – which if you understand p-values is about what you should have believed already.

(I think. This requires quite a few assumptions, not the least of which is that my calculations above are correct!)

Also worth noting: Ioannidis’ experiment did not investigate the absolute highest level of the medical pyramid, systematic reviews and meta-analyses. I expect the best of these to be better than any individual study.

3. Are 90% of the things doctors believe, presumably based on medical findings, wrong?

After going through the steps above, it should be pretty obvious that the answer is no, because doctors are mostly reading famous influential studies like the ones mentioned above, which are at worst 40% and at best 5% wrong.

But there’s another factor to be taken into account, which is that *why would you only read one study on something when lots of important findings have been investigated multiple times?*

Suppose that you’re throwing darts at the Big Chart O’ Human Metabolic Pathways, with your 1/1000 base rate of true hypotheses. You run a very good methodologically sound study and find p = .05. But now there’s still only a 1/50 chance your hypothesis is correct.

But another team in China runs the same study, and they *also* find p = .05. We expect the Chinese to get false to true results at a rate of one to two (because the 1 in the 1/50 stays 1, but the 50 is divided by 20 to produce approximately 2. Wow, I’m even worse at explaining math than I am at doing it.)

Now a team in, oh, let’s say Turkey runs the same study, and they *also* find p = .05. We expect the Turks to get false to true results at a rate of one to ten, for, uh, the same math reasons as the Chinese. When the, um, Icelanders repeat the study, our odds go to one to two hundred.

So we started with 1000:1 odds, the first study brought us up to 50:1 odds, the second study to 2:1 odds, the third study to 1:10 odds, and the fourth study to 1:200 odds, ie we are now 99.5% sure we’re right.

Real medicine is both better and worse than this. It’s better in that we often have dozens of studies rather than just four. It’s worse in that the studies are not all so methodologically sound that we can multiply our odds by 20 each time (to put it lightly).

But some of them are, and once we get enough of them, the base rate problems which plague individual medical findings go away very quickly. Even if only one of the studies is methodologically sound, if the reason they're studying their topic is because a bunch of other less believable studies all got positive results, that's a much better base rate than "because I hit it with my dart".

When doctors say that, for example, iron supplements help anaemia, it's not because they hit iron on their Big Chart O' Human Metabolic Pathways, then ran a single study, got p = .05, and rushed off to publish a medical textbook. It's because they knew hemoglobin had iron in it, there are at least 21 randomized controlled studies, probably some had p-values closer to .001 than to .05 even though I don't have any of them in front of me to check, and eventually some really really smart statisticians at the Cochrane Collaboration gave it their seal of approval. Most doctors' beliefs aren't on quite this high a level, but most doctors' beliefs aren't on the "Someone threw a dart, then did one study" level either.

4. Does this prove the medical establishment is clueless and hopelessly irrational and that two smart people working in a basement for five minutes can discover a new medical science far better than what all doctors could have produced in seventy years?

A lot of people seem to go from Ioannidis' experiment to something like "So I guess everyone in medicine is just clueless about how science and statistics work. I'll go read a couple of medical studies and then be able to outperform everyone in this totally flawed field."

(important note: I'm *not* accusing MetaMed of this! They seem pretty sane. I am accusing some people I come across in the community who are much more enthusiastic than the relatively sober MetaMed people of doing something like this.)

But the problem isn’t that no one in medicine is familiar with Ioannidis’ research. It’s that they’re not really sure what to do about it and figuring out a plan and implementing it will take time and effort.

Ioannidis’ work isn’t exactly secret. I’ve hung out with groups of residents (ie trainee doctors) who have discussed Ioannidis’ findings over the dinner table. According to *The Atlantic*

> To say that Ioannidis’s work has been embraced would be an understatement. His PLoS Medicine paper is the most downloaded in the journal’s history, and it’s not even Ioannidis’s most-cited work—that would be a paper he published in Nature Genetics on the problems with gene-link studies. Other researchers are eager to work with him: he has published papers with 1,328 different co-authors at 538 institutions in 43 countries, he says. Last year he received, by his estimate, invitations to speak at 1,000 conferences and institutions around the world, and he was accepting an average of about five invitations a month until a case last year of excessive-travel-induced vertigo led him to cut back.

So if so many people are aware of this, why isn’t the problem getting fixed more quickly?

An optimist could say the problem isn’t getting fixed because there is no problem. A vast volume of embarassingly wrong medical literature gets published, inflates the publishers’ resumes, and everyone else ignores it and concentrates on the not-really-so-bad large randomized trials. To the post-cynic it is all a smooth, well-functioning machine.

A pessimist might say that the problem isn't getting fixed because it's impossible. The average medical hypothesis is always going to have a low base rate of being true – in fact, if we force scientists to only study high base-rate hypotheses, by definition everything we discover will be boring. There will never be enough resources to apply huge rigorous trials to every one of the millions of things worth studying. So we're always going to have weak studies about low-base rate hypotheses, which is what Ioannidis is attacking as the recipe for failure.

A realist might point out there are some things we can do, but it involves coordinating a huge and complicated system with many moving parts. Journals can force trials to register before they conduct their experiments to avoid publication bias. The scientific community can give more status to people who perform important replications and especially important negative replications. Study authors and the media can come up with better ways to report their results to doctors and the public without blowing them out of proportion. Statisticians can…actually, anything I say statisticians can do is just going to be a mysterious answer, along the lines of "do better statistics stuff", so I'm not going to embarass myself by completing this sentence except to postulate that I'll *bet* there's some recommendation that could complete it usefully.

But all these things involve vague entities who aren't really actors ("the scientific community", "the media") acting in ways that are kind of against their immediate incentives. This is hard to make people do and usually involves a lot of grassroots coordination effort. Which is going on. But it takes time.

But no matter what happens, I think a useful epistemic habit is to be very skeptical of individual studies, and skeptical but not *too* skeptical of large randomized trials, good meta-analyses, and general medical consensus when supported by an evidence base.

Typical mind and gender identity

Ozy Frantz introduced zirself to me by saying that "I major in gender studies, but I am not that kind of gender studies student. Promise." So far this claim has been entirely borne out and Ozy's blog is actually really good. I still usually avoid reading it, because even a good blog on gender has to be responding to and referencing the bad blogs on gender, and even short little quotes from them make me want to smash something and that something is often someone's skull. But it really is actually really good.

I am particularly intrigued by the latest post, Cis By Default, which is related to a conversation we had last week. Ozy thinks that some people have really strong gender identity, and other people don't really have any gender identity at all and just sort of end up not trans-gender because why bother with the social stigma and aggravation that produces?

This neatly combines two of my favorite hammers (in the sense of "…every problem starts looking like a nail"), the idea of individual mental differences and the Principle of Charity.

It also helps explain my own personal experience. I am probably one of the people Ozy calls "cis by default". I obviously can't be sure, but I feel like if I woke up tomorrow magically transformed into a (hetero) woman, my first two thoughts (after "WHAT THE HECK IS GOING ON???", of course) would be "Huh, cool, this probably makes it much easier to find a mate" and "Darnit, now I've got to deal with this whole menstruation thing".

[EDIT: Girlfriend just asked me "Wait, if you woke up magically transformed into a heterosexual woman, your first thought wouldn't be about having to break up with me?" Awwwwwkward.]

And then I'd get into some very rational comparisons to my personality, like "Oh, good, this socially legitimizes my lack of interest in sports", or "Nice, now it's attractive for me to project my standard vibe of passive lack of interest in my surroundings, instead of having to try to appear dominant and take-charge all the time." But I don't envision any kind of *visceral* "But oh no I feel truly male this is an offense against the natural order of things!" being anywhere up there.

And as Ozy predicted, this did make me – if not actively prejudiced against, at least confused by trans people for a long time. My thought processes were something like "Gender seems pretty irrelevant, so I wonder why they're going to so much trouble over this." And I *was* actually rounding off to my own experience – maybe these were men who weren't aggressive or interested in sports, and so they decided that they should become women. Which was kind of dumb. Which was more or less my opinion of trans until I was, I dunno, twenty or so.

What changed my mind was a sort of complicated string of medical facts.

First I learned about phantom limb sensations. These aren't exactly rare or controversial; over half of amputees have them. The theory is that the brain has a map of how the body *should* look which does not always correspond to how the body *does* look. This map isn't even necessarily based on pre-amputation sensations. Some people born without limbs still have phantom limbs. In another case reported by

Ramachandran, a woman whose hand had been deformed from birth needed an amputation; she reported the resulting phantom limb was non-deformed. All these suggest that the brain’s map is at least partially hard-wired rather than based solely on experience.

Later I learned about the exact opposite phenomenon, body integrity identity disorder, where a healthy person with four healthy limbs feels like one of their limbs is “wrong” and has a strong desire to get it amputated. This was originally classified as a psychiatric disease and even dismissed as a subset of amputee fetishism, and to my disappointment Wikipedia still mostly treats it along these lines.

But I’ve heard that some of the same neurologists working with phantom limbs think this is also a disorder of brain-body mapping; that for some reason their brain’s map of their body doesn’t include a perfectly healthy limb and so it feels foreign. Some very preliminary MRI studies seem to have borne this out, and there are other signs that there’s something biological going on too (the limb involved seems to be disproportionately associated with the nondominant hemisphere of the brain).

More convincing to me is that the disorder has a very non-psychiatric resolution: if patients are able to amputate the offending limb, they are perfectly happy and never have any further complaints. Compare this to a more psychiatric population like hypochondriacs, who if you treat one of their fake sicknesses just come up with another, because the underlying psychological problem that makes them want to feel sick hasn’t gone away. After hearing this story I decided to count the previous dismissal/marginalization of these people as a huge failure of psychiatry and as exactly the sort of thing I need to watch out for.

And this segued nicely into stories of people who believed that some of their gendered body parts (breasts, genitals) are wrong and not really part of their bodies. It has all the features that gave body integrity identity disorder a red flag for being organic rather than psychiatric: very strong desire to remove offending body parts, different distribution of comorbidities you see with a lot of genuine psychiatric disorders, and removal of the offending body part makes the person pretty happy. It seems that the same logic that says BIID sufferers may have a map of their body that doesn't include their left leg or whatever could give these people a map of their body that doesn't include their penis or their breasts or something.

Of course not all transgender people have this kind of very medical-seeming aversion to a specific body part. But the medical-aversion type of people often *also* have the typical transgender "I feel my gender identity is wrong" belief, and it would make sense that if some sort of "gender" variable got switched somehow this could produce a gender-inappropriate body map. The existence of gender-inappropriate body maps, and of this whole literature of people's brains sometimes telling them things about their bodies that aren't true and then it causes great distress, makes the idea of an isolated gender-variable switch much more plausible.

And if it's possible to have a gender variable get switched, then it's much more parsimonious to believe that the people who say they feel like their gender variable is switched *actually* have this gender variable switched, rather than that they're acting out for some sort of weird social reason. Especially if it's hard to think up a weird social reason that would be worth it.

Caring about gender still doesn't make internal subjective psychological sense to me, any more than having some kind of mental "pants type" variable that says I should be wearing blue jeans and gets deeply distressed whenever I wear khakis.

Then again, I naively picture my body map as being such that, if I got a third arm for some reason, I'd just shrug and see if I could leverage it into a career as a the most confusing basketball player ever, instead of desperately wanting it amputated. So maybe my naive predictions about what variables do and don't matter to my brain aren't very good.

So I have two hypotheses. One is Ozy's hypothesis that some people have gender identity and others don't, and for some people that gender identity matches their biological sex and for others it doesn't. A second hypothesis is that everyone has gender identity, but if it matches your biological sex it's impossible to notice (just like the analogy where fish don't understand the idea of water) and if it doesn't match your biological sex it becomes obvious (just as a fish would no doubt notice being taken out of water).

It sounds like the best way to distinguish between the two of these would be to check for cis- people with strong gender identities. Do any of you cis- people out there have a strong, visceral feeling of "Oh, thank God I was born a man/woman, because that's just *obviously* what I ought to be – and if I magically transformed into a woman/man, I would want to get surgery/hormones/etc to change back as soon as possible" or something like that?

I liked Lovecraft countless primaeval aeons before it was cool

A couple days ago, some friends of mine who name-drop Cthulhu all the time admitted they had never actually read any Lovecraft. Further investigation revealed this was actually a common thread among loads of people who play *Call of Cthulhu* RPG, or describe various buildings or institutions as Lovecraftian, or use “Oh my Cthulhu!” as an atheist substitute for “Oh my God”.

I know it sounds like the height of pretention, but I’m going to come out and say it anyway: I am kind of unsatisfied with what geek culture has done to Lovecraft.

Hello Cthulhu was funny for a few minutes. The Campus Crusade for Cthulhu might be amusing if you’ve had too many dealings with their more wholesome counterpart. Shoggoth on the Roof continues to be one of the most hilarious musicals I have never quite managed to make it to a real showing of (my two favorite songs are Tentacles and To Life.)

And yet.

Lovecraft himself did have a sense of humor, and his biography of Ibid continues to be one of our most important resources on the subject. But I have never seen any signs that he ever decided to mix any trace of humor into his dealing with the Elder Gods or the fantasy worlds he created. I don’t think that was just a lack of sufficient creativity. I think that was a deliberate part of his artistic impulse, at the heart of what he was trying to say.

Here is a passage from *The Silver Key*. It is long, but I have shortened it as much as I dared, and indeed you would do far better to read the much longer original. Nevertheless, here is the passage. All emphasis is my own:

> When Randolph Carter was thirty he lost the key of the gate of dreams. Prior to that time he had made up for the prosiness of life by nightly excursions to strange and ancient cities beyond space, and lovely, unbelievable garden lands across ethereal seas; but as middle age hardened upon him he felt those liberties slipping away little by little, until at last he was cut off altogether. No more could his galleys sail up the river Oukranos past the gilded spires of Thran, or his elephant caravans tramp through perfumed jungles in Kled, where forgotten palaces with veined ivory columns sleep lovely and unbroken under the moon […]
>
> So Carter had tried to do as others did, and pretended that the common events and emotions of earthy minds were more important than the fantasies of rare and delicate souls. He did not dissent when they told him that the animal pain of a stuck pig or dyspeptic ploughman in real life is a greater thing than the peerless beauty of Narath with its hundred carven gates and domes of chalcedony, which he dimly remembered from his dreams; and under their guidance he cultivated a painstaking sense of pity and tragedy.
>
> Once in a while, though, he could not help seeing how shallow, fickle, and meaningless all human aspirations are, and how emptily our real impulses contrast with those pompous ideals we profess to hold. Then he would have recourse to **the polite laughter they had taught him to use against the extravagance and artificiality of dreams;** for he saw that the daily life of

our world is every inch as extravagant and artificial, and far less worthy of respect because of its poverty in beauty and its silly reluctance to admit its own lack of reason and purpose. **In this way he became a kind of humorist, for he did not see that even humour is empty in a mindless universe devoid of any true standard of consistency or inconsistency** […] Having lost [their] artificial settings, [people's] lives grow void of direction and dramatic interest; till at length they strive to drown their ennui in bustle and pretended usefulness, noise and excitement, barbaric display and animal sensation. **When these things pall, disappoint, or grew nauseous through revulsion, they cultivate irony and bitterness**, and find fault with the social order […]

Then he began once more the writing of books, which he had left off when dreams first failed him. But here, too, was there no satisfaction or fulfillment; for the touch of earth was upon his mind, and he could not think of lovely things as he had done of yore. **Ironic humor dragged down all the twilight minarets he reared**, and the earthy fear of improbability blasted all the delicate and amazing flowers in his faery gardens. The convention of assumed pity spilt mawkishness on his characters, while the myth of an important reality and significant human events and emotions debased all his high fantasy into thin-veiled allegory and cheap social satire. His new novels were successful as his old ones had never been; and because he knew how empty they must be to please an empty herd, he burned them and ceased his writing. **They were very graceful novels, in which he urbanely laughed at the dreams he lightly sketched**; but he saw that their sophistication had sapped all their life away.

Lovecraft was not opposed to humor, but I think he *was* opposed to ironic humor, opposed to the kind of humor that takes something titanic and magnificent and then sticks it in the middle of the mundane world so we can all laugh at how it’s not really that much beyond us after all.

And I am pretty sure that putting Cthulhu into Hello Kitty cartoons qualifies.

I think Lovecraft’s artistic philosophy was to come up with something so far beyond normal human experience, so intensely appealing to our wilder and more celestial tendencies and so completely unrelated to our normal base emotions, that even people not quite as sensitive as he was could feel a glimpse of this transcendent otherness.

Although some people classify Lovecraft as a horror writer, I think this misses his essence – there is nothing horrible about *The Silver Key*, and Randolph Carter, the character most closely corresponding to Lovecraft himself, usually comes out of his encounters with the Beyond better than he went into them. Lovecraft is a writer of the strange, intense, and unearthly. Horror is one of the feelings that can be intense, but so is wonder.

Lovecraft’s aim in creating Cthulhu – and all the rest, because in Actual Lovecraft (as opposed to Pop Lovecraft) Cthulhu is a relatively minor character – was to give us something completely divorced from our normal world where the aesthetic senses could wander free from their usual pollution by the everyday world of politics, commercialism, and status-seeking.

And so of course we immediately turn Cthulhu into a presidential candidate, a clothing line, and a meme, respectively, *just to spit in his face*.

I should clarify that “just to spit in his face” thing. I don’t think people literally did this to anger Lovecraft’s ghost, since that sounds like *the worst idea*. But I do think that the more inappropriate a subject is for humor, the more distressing and contradictory to its nature it would be to turn it into a joke, the more humorous it is likely to be. The existence of an entire canon of dead baby jokes is a good example of this. The continuing popularity of racist jokes, which as far as I can tell greatly exceeds the continued popularity of actual non-joke racism, is another. Lovecraft is much like racism and dead babies in that it is something that *obviously* should never be joked about, and therefore incessantly is.

But the situation with Lovecraft is kind of worse than the other two. People who tell dead baby jokes are at least aware that real babies sometimes actually die and that this is very sad. People who tell racist jokes are usually entirely aware of and sometimes even believe in the anti-racism movement, and if not they *at least* know there exist real black people out there somewhere and that they’re not just this mysterious vague concept that serves as a useful butt for jokes.

Yet so many of the people who make fun of Cthulhu have never read the *Dream-Quest*, never climbed Ngranek in their imagination or visualized gorgeous minaret-studded Celephais with its sky-bound galleys, never shrank in terror from the High-Priest-Not-To-Be-Described.

And if that’s what people want, fine. I won’t deny that a lot of Lovecraftian humor is pretty funny.

But if you think Cthulhu’s interesting, and if you’re also the sort of person who longs for the fantastic and otherworldly, if there’s anything poetic or romantic in you, I urge you not to limit your Lovecraft knowledge to *Call of Cthulhu RPG* and to take a look at some of his more serious works. I most highly recommend The Dream-Quest of Unknown Kadath, which like all Lovecraft is freely available online should only take two or three hours if you're a fast reader. But if that sounds like too much for you, there are excellent short stories in the Dream Cycle like The White Ship that shouldn't take more than a few minutes.

Future tense

Those of you who know me in real life know that I graduated from an Irish medical school last year but wasn't able to get a US residency (ie entry-level doctor job) at the time.

This year I'm trying again. I'll hear back from hospitals in mid-March. I think I have a good chance.

But a good chance is not a perfect chance. And a non-perfect chance means there is some chance of failure. And this is utterly terrifying to me, to the point where I'm literally having nightmares about it and going a little bit off the wall.

The problem is that these job opportunities only come around once a year, and that if I fail a second time there's not a lot of reason to think I'd do any better next year. In fact, there are a lot of reasons to think I wouldn't; hospitals are usually less willing to accept people the longer they've been away from medical school, and they're usually less willing to evaluate people they've already evaluated before and rejected. And there aren't that many hospitals offering residencies.

So if I don't get a job this time, I'm going to have to seriously consider the possibility that four years and thousands of dollars worth of medical school were a total loss career-wise and that I'm going to have to do something else. And I have degrees in philosophy and psychology and not a whole lot of obvious marketable skills. I will be totally back at square one with no idea of what to do with my life.

The rest of this entry is really emo and privileged and people who are actually having serious life crises will find it enraging, so if that's you please don't read it.

I am very lucky and grateful to have amazingly generous and far-sighted parents and grandparents who among many other things gave me saving bonds every birthday as a child which have since spent twenty-five years maturing. As a result, I don’t have the problems with medical school loans that you were probably expecting this post to be about, and I have enough money to live comfortably for some time period greater than a year and less than forever. So there will be no immediate crisis. I even have enough money (and/or loan-take-out-ability) to go back to school if I need to. But there is a definite long term “I should probably get a career path or something”.

And since I expect my mental state to be way too shattered to do any kind of planning or decision-making if I do get rejected, I figure I should figure out a backup plan now while I’m still sane so I can carry it out on autopilot if worst comes to worst.

I love working for MetaMed, but they are a very ambitious startup and it seems unwise to bank my entire ability to survive or ever have a career on them sticking around. Besides, just in case they get tired of me accidentally misrepresenting their positions or projecting general cluelessness at them, I should try not to have all my eggs in one basket. MetaMed has thus far not had more work for me than I can do in free time between other obligations, so “continue working for MetaMed but figure out something else to do full time” seems like an obvious best strategy.

It would be nice to work for MIRI in some way. I have a lot of friends who have positions sifting through academic literature for them, and it seems possible I could get something like this if I begged hard enough. But this seems a lot like MetaMed in being a good part-time addition to my workload and not a stable long-term career plan. Harder to explain but more importantly, I think a lot of my sense of “I am useful to people and not just a parasite on the world” would come from being able to donate some of what I made to charity, and if I’m taking my money from what is possibly the most efficient charity already this becomes kind of a farce. Probably also a “good if it works, but do in between something else” option.

I would really like something where my MD is not completely irrelevant. Glibly dismissing “sunk cost fallacy” ignores the likely fragility of my mental state and the fact that my parents and family are also going to feel much better if all the work they put into helping me through medical school isn’t just tossed aside. I have heard that many MDs go on to do a Masters in Public Health. This seems to be only a one to two year course, which is at least a heck of a lot better than having to spend another four years in school. Problem is, I have no idea what a Master of Public Health does, and whether there are likely to be any jobs in it. Anyone here have any experience in that area?

I feel overprivileged just writing this, but I could always just take the LSAT. I’ve been told that lawyers with MDs can get cushy positions in medical law, and although this sort of offends my sense of ethics and not-being-parasitic, a sufficiently good job could give me enough money to donate to charity to make up for that. I know it’s hard for lawyers to find jobs nowadays, but I’ve heard if you go to a good enough law school it’s still not so bad. I could plausibly hope for exceptional LSAT scores – I got a perfect score on the SAT verbal which I hear is similar – so that plus my MD might be enough to get me in somewhere really good. Then again, it’s another four years of very expensive education once again without a guaranteed job at the end of it.

I’ve heard another route some MDs go if they can’t get into clinical medicine is getting an MBA and going into some form of medical business. The fact that I have no idea what forms of medical business exist seems to put me at a disadvantage compared to people who have been wanting to do this their whole lives, plus I think everyone who knows me agrees I would be the worst businessman ever (“Honestly, as far as I can tell the evidence for our product working isn’t that robust, and I’m not sure it meets your particular needs anyway”). This is also one where I’m concerned about whether I would have a good shot at a job at the end of it. Getting business jobs seems to require networking and self-selling and self-confidence other things I am totally unable to do.

I really like biostatistics. I just happen to be really, really bad at it. I can fake writing about them pretty well, but in terms of actual number crunching they sort of slip out of my mental grasp just like all other math. I’m not sure whether going back to school to become a statistician would force me to learn it, or just end up with me failing out of school and being even worse off than I started. Also, once again I really don’t know anything about the state of the job market here.

That leaves non-medical jobs. Mike Blume pushes his idea of “learn programming on your own and make money as a programmer”, and I have to admit it worked for him and that programming strikes me as an extremely honest job with good working conditions. On the other hand, I don’t know if companies actually give programming jobs to people with no experience beyond independent study, and I don’t know if I could actually teach myself programming. People say that programming is easy and any smart person can learn to do it, but these same people also tend to say that math is easy and any smart person can learn to do it, which makes me think either they’re just wrong or that I don’t qualify for the current definition of “smart person”.

I’ve always sort of wanted to be a teacher, and it’s a decent-paying, decent-status, not-*too*-unethical job. The fields that I would probably be really good at teaching (history, English, etc) are totally jobless, but I’ve heard there’s a big demand for math teachers. I’m bad at math compared to genuine math geeks, but I bet I could be better than high school students and probably better than the average math teacher right now (seriously, some of them are horrible), so going to school to get a teaching credential seems like a good non-medical option.

There are also some super extra backup options, like just going back to Japan and seeing if my old English teaching job is still available and whether I can just do that for the rest of my life.

I’m kind of curious if there are any obvious low-hanging fruit jobs around right now. People used to say that it was both easy and lucrative to become an actuary and they had relatively good working environment/status, but that loophole in the rat race seems to have closed up now. All the ones I know about now (Australian outback, etc) seem more like summer jobs for college students than lifetime careers. Are there any I’m missing?

Right now, if I don’t get a job I think my plan is to stay in Berkeley until June, possibly studying just enough programming to see whether I have any aptitude for it. Then do some clinical rotations in Internal Medicine at any hospital that will take me over the summer to see if maybe the Internal Medicine people are mysteriously more favorably disposed to me than the Psychiatry people. Once the summer winds down and the hospitals kick me out for their normal set of med students, I could spend the autumn applying to, in order of desirability: Internal Medicine residencies if they seem mysteriously well disposed to me, medical externships (yearlong unpaid medical resume builders), law schools, MPH programs, MBA programs, and whoever trains teachers (note to self: look that up). Then wait until next March and take the best thing I get into. If I get into literally nothing, and I have aptitude for programming, go into that. Otherwise, move back to Japan (or Korea? or Arabia?) and teach English (or stick around and work for MetaMed if they're successful) to earn enough money to subsist while I continue applying to things.

This is a bad plan. I do not like it. I really really hope I get a residency in this year's Match.

A defense of logical positivism (yes, really)

As far as I know, there's nothing everyone in philosophy agrees is right. But there *are* a few things that everyone in philosophy agrees are wrong. Right now I can think of two of them – Rene Descartes, and logical positivism.

Er, well, one of them. Because I kind of like logical positivism.

And before everyone yells at me, I understand this is sort of sketchy. I'm not saying that they were entirely right about anything, or that their criteria work exactly as stated. But it's like – well, take those ancient philosophers who said everything was made out of atoms. In fact, they're wrong. Light isn't made out of atoms, mathematics isn't made out of atoms, quarks aren't made out of atoms, et cetera. But they were sure on to something, and I give them much more credit than philosophers who *didn't* say everything was made out of atoms.

I am an inexact person, and I tend to think inexactly, and the fact that the logical positivists were working in an area that vaguely points to a cluster of correct things is good enough for me. Let me explain what I mean by "a cluster of correct things".

In the 1700s, David Hume came up with an idea now called "Hume's fork", which Wikipedia describes like so:

> Hume's fork is an explanation of David Hume's aggressive division between "relations of ideas" versus "matters of fact and real existence". By Hume's fork,

> relations among ideas are strictly divided from states of actuality.

(and before you ask, yes, Hume's Fork is also a legendary artifact in *Dungeons and Discourse*)

Then in the 1900s the logical positivists came around; they get described as saying "that the truths of science are verifiable empirical claims and that the truths of logic and mathematics are tautologies. These two constitute the entire universe of meaningful judgements; anything else is nonsense."

And in the 2010s, the latest Less Wrong Sequence, Highly Advanced Epistemology 101 for Beginners tried to reduce meaningful statements to "two kinds of meaningfulness and two ways that sentences can refer; a way of comparing to physical things found by following pinned-down causal links, and logical validity by comparison to models pinned-down by axioms" as well as combinations between them. This is a really good sequence and I especially recommend Mixed Reference which says a lot of the stuff below only better (I did avoid re-reading it until I was done writing this, so as to keep my brain at least a little in original-idea-mode).

These three claims are somewhat different, but they all seem to have kind of the same idea. They all seem to be saying that we can divide meaningful things into something-kinda-like-science, and something-kinda-like-logic, and that everything else is meaningless. Their differences seem to be mostly how "something kinda-like-science" gets cashed out.

And while acknowledging the importance of figuring out exactly what "something kinda-like science" means, right now to me it seems less important than the overall observation that having even this vague sort of poorly-specified system is a whole lot better than not having any system of this sort at all.

More importantly, this idea is *productive*. I admit my assertion of its productivity is highly biased, in that by “productive” I mean “things I want to hear fall out of it”, but I do find this happening with gratifying frequency. Usually when this system red-flags a statement as “probably not meaningful”, on further investigation it very often does turn out to be something bad that needs to be thrown out. And when this isn’t the case, I usually gain a much more solid understanding of the statement by trying to see exactly how it reduces to one of these two kinds of meaning or a combination thereof.

It may be that there are some kinds of statements that are meaningful but just plain don’t work this way. I am having trouble thinking of them, but it’s normal to have trouble coming up with statements that disprove a theory you like, and I’m sure I will hear many attempts at them in the comments. But even if they exist, they seem to me much like the “light is not made of atoms” example above – correct, but irrelevant to the fact that atomic theory is *really awesome* and that if someone goes around saying “That particular horse there is not made of atoms, and no one can force me to say that it is,” they’re probably making a mistake.

As an example of “things that I do think yield to this positivist technique”, let me give some classic examples of things that people say *don’t*.

Start with mathematics. This one is easy. Mathematical systems are systems of axioms and theorems. These are logical tautologies. Some of these systems seem to describe our world pretty well. This is an empirical observation.

Move onto morality. A little harder. The definition of any moral system is a logical tautology; if I define “utilitarianism” as “act for the greatest good of the greatest number”, I have defined what is…not quite a formal system, but could probably be turned into one if someone were rigorous enough. My brain seems to exist in such a way that it reflectively endorses this system rather than another one. That’s an empirical observation. This isn’t quite what most people want out of morality, but I consider that a feature, not a bug; morality the way most people want it is probably meaningless.

An even harder one: “If Lincoln had kept McClellan as general, the South would have won the US Civil War.” Maybe this describes the output of the logical formal system corresponding to the universe (!) with certain inputs empirically determined by the state of Civil War-era America.(I bet there’s a better way to rescue this one, but I can’t think of it.)

[EDIT: Now that I am done writing this I am looking at the Mixed Reference post above, and it does a pretty good job]

Finally, we come to logical positivism itself, which was famously accused of failing its own criteria by Karl Popper. Once again, everyone agrees it’s possible to *define* logical positivism. And by analogy with math and morals, now we just need to show that this system, as defined, matches some property we want it to have.

And obviously this is going to be sort of subjective, because “property we want it to have” depends on what properties we want of our ontologies. It might even end up circular, because a lot of the properties I want my ontologies to have probably sound a little like “Doesn’t talk about meaningless stuff, with meaningless being defined as [something that uses a lot of the same criteria as logical positivism]”.

But I think it might also correspond closely to that which can be debated. That is, what is there such that, using reason rather than emotion or made-up pseudologic, we can actually change our minds about and correctly judge as having one probability of truth rather than another?

I welcome examples of this being proven totally false.

Dungeons and Discourse Third Edition: The dialectic continues

After more work than any sane person would ever spend on a philosophy role-playing game, the Third Edition of Dungeons and Discourse is ready. You may download it here:

Rulebook as .pdf plus basic character sheet (link)
Rulebook as .doc plus basic character sheet (link)
Advanced character sheet (requires NBOS Character Sheet Viewer, link)

Changes are many but the ones I can think of at the moment are the following:

– New spells like Dedekind Cut, No-Ghost Theorem, Heisenberg's Uncertainty Cloak, and Kurzweil's Exponential Crescendo, and Infinite Egress.

– Plus player-suggested spells like Armor Fati, Summon Dire C. Elegans, Summon Invisible Dragon, Summon Bonum, and Descartes' Rapier.

– Revamped prestige class system. Instead of the somewhat complicated and anaemic prestige classes of the Second Edition, prestige classes have been made more accessible and folded into books. Want your Logician to subspecialize in computer programming? Read a copy of *Structure and Interpretation of Computer Programs*. Want to your normally hard-headed Empiricist to flirt with Taoism? Pick up the *Tao Te Ching*. 21 books to choose from.

– Threw away the Righteouness statistic and replaced it with Merit and Virtue. Merit is a set point to which your Virtue always slowly returns, Virtue is a state characteristic which rises and falls based on your most recent actions. In order to stay ethical you'll have to worry about both.

– Creatures no longer have different physical and mental hit points. Everyone has a single hit point statistic, but damage can be inflicted either physically or mentally. Some creatures have bonuses against one type of damage or another, and other creatures, like Platonic Forms and P-Zombies, lack physical or mental substance and so are completely immune to that form of damage.

– Combat is much easier, with an official shift from keeping track of everyone's speed, range, and position on a giant grid to the simple front row/back row system that we everyone ended up using during the last campaign. All spells and combat techniques have been updated to take account of this and this has also allowed serious simplification of the Speed statistic.

– Got rid of the five complicated "spell slots" you had to worry about and replaced them with a single Will statistic. You want to cast spells? Expend Will points. It's that easy.

– The currency of Sophia has switched from gold to benthamite, a magical substance made of pure utility. A lot of the economic gobbledygook of the previous version has also been combined into a single statistic called Oikonomia.

– Items have been reworked, most weapons have been removed, and geasa have been separated off into to their own section for pennants and rings. More interesting choices around Euclidean polygon use. The arrow system is now potentially sort of functional.

– No-longer-quite-as-awful design includes front cover and character sheet produced by an NBOS program.

– Rulebook contains the campaign log and music from the flagship campaign for the Second Edition, *King Under The Mountain*, in the back.

Please let me know if there are any typos, obvious glaring balance problems, spells/classes/mechanics that are referenced and then don’t exist/were never explained, or things you really don’t like. The comments to this post would be better than email.

The long-term plan is to convert the rulebook into a wiki, which can then get edited on the fly whenever a glaring balance problem comes up during a game and which random netizens who stumble across the concept are more likely to read than a .pdf file. I am pretty resigned to having to do this myself, but if anyone else is really up for thankless data transfer work, it is totally yours.

The flagship campaign for the new edition is *Fermat’s Last Stand*, a continuation of the story in the last one. It begins… soon. I need to finalize a couple of the songs, figure out how I’m going to conduct one or two more things, and then ask for volunteers. Please do not volunteer now. Volunteer when I ask. This will probably be within a week.

People! It turns out capitalism is actually a thing!

Reddit is caught in a Nietzschean cycle of eternal recurrence where everyone asks the same questions every few weeks, and one of the common ones is “What is the best purchase you ever made?” The top answer is always “Amazon Prime”, and I always told myself to think about getting Amazon Prime, then never got around to doing so.

When I moved in with Mike, he already had Amazon Prime and I was able to piggyback off his account because we lived at the same household. And once again the Reddit hivemind has proven startlingly accurate.

My old algorithm for choosing a product was something like this:

1. Realize need for product
2. Go to store
3. Look over shelves full of competing brands of product
4. Reach for product with shiniest packaging
5. Think “No! That would not be virtuous!” Snatch back outreached arm.
6. Studiously examine marketing claims on packaging of all products.
7. Try to parse incomprehensible, unverifiable assertions like “Special n-acetylcysteine strip makes our product 150 kilojoules more convenient than competing brand!”
8. Furrow brow
9. Finally choose product which is cheap, but not *too* cheap, because the one that’s *too* cheap might be poor quality.
10. Be consumed by self-doubt.

My new algorithm for choosing a product goes more like this:

1. Type name of product into Amazon
2. Limit search to Prime-eligible and sort by average customer review
3. Find the inevitable five-star rated product which is also pretty cheap and has dozens of customer reviews all saying some variant on “I’ve used hundreds of different [product_type]s before, and finally my sainted grandmother, may she rest in peace, recommended [brand_name] to me. It was by far the best I’ve ever seen and I can’t believe it costs so little. [brand_name] has revolutionized my [product_type]-using-experience and I’m never going back to any of those other brands again!”
4. Buy that one
5. Profit

To my surprise, this second algorithm works a lot better.

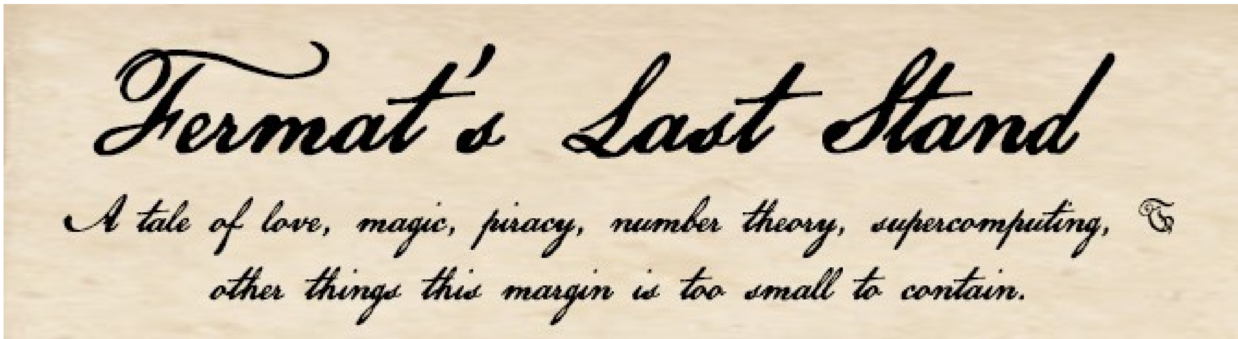
(I’m not being sarcastic with that “to my surprise”, by the way. I have such a high prior for things not working – for magically returning themselves to baseline via mysterious set points despite all evidence that they shouldn’t – that it really did surprise me to find the second algorithm working better than the first.)

I have since successfully bought curtains that *actually block light so that you can’t see it*, industrial-strength earmuffs which I wear all the time people aren’t actively trying to talk to me and sometimes when they are, a laptop which has served me well beyond the incessant clicking noise (but see “earmuffs”, above), these amazing computer speakers that people actually *remark* upon, and a dream diary that legitimately looks like a sinister 16th-century magical grimoire while costing less than some non-grimoire-looking books.

Looking around the room for things to include in the above list, I also spot an Amazon purchase good enough that I feel compelled to mention it despite its not being recommended by Amazon. Andrew Rettek recommend for me IRON GYM, which aside from having the manliest name in the world is this really interesting contraption you attach to your door to turn it into a pull-up bar. I have no idea why I got it since I'm terrible at sticking to even non-contraption-requiring exercises like situps and pushups, but for some reason it seems to work. I think, as strange as this will be to relate, that the difference is that I have to shift to a near-the-floor position for situps and pushups, whereas I can just make use of the pull-up bar immediately when walking through a door without any positional change required. Judging by my difficulty getting off my comfy computer chair to do useful things, or getting out of bed in the morning, I think either people's bodies or just *my* body may view change between sitting/lying and standing as some kind of bizarrely relevant millisecond-level-discounting trivial inconvenience cost factor when deciding what to do. In any case, the IRON GYM seems to work and I am now capable of multiple pull-ups, which would shock any of my high school gym teachers or really anyone else who knows anything about me at all.

It always makes me kind of surprised when things actually work the way I learned about in economics class. The first time I ever boycotted a company, I got inordinately excited walking past the store and not buying their products, because *I was actually exerting economic power and affecting the market, just like those people in thought-experiments!* And of course after three seconds I realized this was hilarious because of course everything I have bought or not bought since the moment I was born has been doing that. But until then it had always seemed like a passive process – buy something that doesn’t look terrible, or that a friend had used and not had any problems with, then hope for the best. That first boycott (I can’t even remember the company) made me feel like an active participant in the system again. And buying from Amazon is neat because it makes me feel the same way.

Fermat’s Last Stand Coordination Thread


Fermat’s Last Stand
A tale of love, magic, piracy, number theory, supercomputing, & other things this margin is too small to contain.

All right, let’s make this happen.

I would like to start a new *Dungeons & Discourse* campaign. If you’re totally confused, see here for the rulebook and here for the last campaign log.

As you might be able to tell, the game will be confusing and un-fun unless you know a lot of philosophy. The last campaign required, for example, at least a minimal level of familiarity with Hobbes’ theory of government, Richard Dawkins’ atheism, p-zombies, and Russell’s paradox. This game will require about the same amount of background knowledge.

But we had a lot of fun last time and everyone was able to follow without any trouble. So you are interested in participating, please leave a comment below with the following

– Are you interested in playing?
– Are you interested in DMing?
– If both, which would you prefer?
– Are you interested in doing this in meatspace? If so, where do you live?

– Any time preferences (like “not on Saturdays” or “only on Saturdays”)
– Any scheduling preferences (like “once a week”)
– If playing, a one-sentence description of your philosophical beliefs (eg “libertarian atheist reductionist”) so we can try to get our parties as diverse as possible
– If you have a strong preference to be in the same game as someone else
– Contact details

If I’m really lucky there will be enough people to organize into several groups that satisfy everyone’s preferences. If I’m not, I’ll throw everyone together and dub the resulting mess a party.

I know there are a couple of groups that have been set up already. They obviously don’t need to comment here, but I would like to get in contact with the DMs of these groups for a few reasons. First, we should set up a mailing list or something to compare ideas and strategies. Second, I want to give you the campaign material.

At the moment, this will take the form of a promise of several songs which I have not written yet, plus a plot which I sort of have. The plot is very bare bones and does not contain encounters or puzzles or dialogues or fleshed out settings. It’s basically an outline for constructing a campaign rather than a campaign itself. This is for a few reasons – I’m not too good at fleshing out campaigns and don’t want to limit you to what I’ve got, plus I plan to make a lot of the details up as I go along. That means DM volunteers need some experience or clue what they’re doing (unlike me, but I’m volunteering anyway).

I’ll try to have parties worked out by the beginning of March and then we can finally start this.

MetaMed launch day

I said a few days ago that I would be a terrible businessman because my attempts to promote things tend to meander into attempts to steelman the case against them. So I will spare MetaMed the ordeal of having me try to advertise for them.

But: they are a company, they have a website at http://www.metamed.com, they do personalized medical research, and they are officially launching today.

(they've been doing things for a bunch of people already, so it's not like they're completely new, but so far it's been mostly unofficial)

I am usually a couple of steps away from customers, so I can't speak much about that end of things. But I have been getting to do some research and occasionally just listen in on meetings, and it has been *so much fun* to listen to people who have this ethos that difficult medical questions should be solveable, who don't make stupid mistakes or take shortcuts trying to solve them, and who have enough time and money to give them the attention they deserve.

A couple of weeks ago we had our first journal club. It was an attempt to evaluate the efficacy of antioxidants and multivitamins against cardiovascular disease and cancer. We got a bunch of contradictory studies, threw them all together, and started generating hypotheses about why they might contradict each other so much. It kind of fizzled out before we came to any strong conclusions, but in a world where some hard-to-determine number which is not ninety percent of medical studies are wrong, it's exactly the sort of thing that should be done (also, in the course of the discussion, “buy antioxidant supplements” went on and off my to do list several times; it’s currently off but I have no idea how long that will last).

Now there’s a discussion going on about the biochemicals in ginseng which I have yet to finish doing enough research on to participate in intelligently. But I’ve skimmed through what’s been said and what’s most fascinating to me is that no one has fallen into the standard error modes of “This is labelled as ‘alternative medicine’, therefore we know it sucks without investigating it” *or* “This is labelled as ‘alternative medicine’, therefore we know it’s great without thinking about it”, a seemingly basic feat of rationality which people nevertheless manage to miss 99% of the time.

There are already a couple of sites that aggregate medical information and try to pull a few tentative conclusions out of the maze of studies. UpToDate is a good one; so is emedicine and (to a degree that surprises even me) Wikipedia. But they all have their problems. UpToDate is very expensive and mostly aimed at doctors; Wikipedia is on average very good but high-variance and proooobably not the sort of thing you want to make life-and-death decisions based on; emedicine is nice but the more second (and third, and so on) opinions you can get the better. And none of them are guaranteed to have exactly what you want – for example, when I was looking into antioxidants, none of them really analyzed the important and controversial SUVIMAX study, which I really would have appreciated some good commentary on.

If MetaMed can supplement some of those with its own brand of rationalist medical analysis, it will be a good thing. If it can fill a slightly different niche by doing personalized medical research on issues other information sources might lack for people with a little more money to spend, it will be a very good thing. And Zvi, Alyssa, Vassar, Sarah, et al seem to have even more ambitious – some would say wildly ambitious – plans for it which I don’t know much about except that I can think of much worse groups of people to be setting the future direction of the health care industry.

Pope and Change: An Atheist’s Guide To Vatican Decision 2013

Pope Benedict XVI is resigning today, an event which will be closely followed by a conclave to determine his successor. I have inexplicably been following this with some interest.

Several news sources both secular and religious have tried to provide me with information on the potential successors without success. The articles tends to be a bunch of pictures of identical-looking old guys in red robes, which then go on to point out that *this* old guy in red robes was head of the Congregation for the Doctrine of the Faith, but *this* old guy in red robes was head of the European Congress of Bishops, whereas *this* old guy in red robes was head of the Pontifical Council for Justice.

Some stories do classify them as either “liberal”, “moderate”, or “conservative”, which sounds more promising. But they never quite give any hint of what these terms could possibly mean, as all the cardinals seem about equally opposed to homosexuality and abortion and women priests and as far as I can tell everything else.

Proooobably there are aspects of this which are fascinating for Catholics, but as an atheist, I want two things from a Pope.

First, I kind of want him to be progressive. I have also accepted this will never happen*. For a while I thought I could kind of get away with hoping for *relative* progressivism, like rooting for the guy who was stuck in the 18th century over the guy who was stuck in the 15th century. But the differences between them seem so slight as to make this a dangerous game. And it seems sort of like a form of cultural imperialism to demand the head of a religion I don't believe should parrot my views and ignore the views of his religion's actual members.

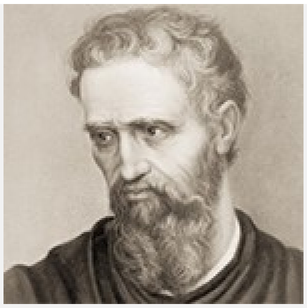
But second, and less controversially, I want him to be hilarious. "Pope" is a funny word. Everyone knows this is true, even if they don't admit it. And the Papacy itself tends to involve a person who looks like Emperor Palpatine going around in a huge hat saying silly things and mumbling prophecies of doom. I want someone who can carry on that legacy.

I understand it's not always an unalloyed good to mock what is holy. But if something makes a claim to holiness, and falls short, and does it in a really funny way, such that you can still take good-hearted cheer in the bungled attempt without trying to mock the idea of holiness itself – well, I think that's all in good fun.

So I've been reading through the biographies of some of the important cardinals and, while looking for evidence of their strong opinions on important issues, I've mostly just been collecting interesting or bizarre facts about them. And I thought that just in case any other atheists also wanted to know which of the *papabile* cardinals were most hilarious, I might as well turn it into a blog post.

Since this doesn't solve the "all papal candidates look alike" problem, I've *also* replaced their pictures with a picture of someone else whose name sounds like theirs. To, you know, help keep them separate in my mind.

**Cardinal
Angelo Scola**

	
Country of Origin:	Italy
Odd Professorships Held:	Professor of Contemporary Christology, Professor of Theological Anthropology
Position on Abortion:	Pro-life
Key Quote Indicative Of His Intellectual Prowess:	“Would not Husserl’s reduction, together with the critique of constituting intentionality, rise up in protest against such a notion of foundation? On the other hand, to say that the foundation is already in place is not to reduce it to a ‘supreme being’ that can be adequate, as an object, to the conceptual representation of a subject, as naive realism would have it. Could the foundation, so conceived, stand up to Heidegger’s well-known critique of onto-theology?” (source)
Key Quote Which Is Also, Confusingly, Indicative Of The Intellectual Prowess Of The Same Person:	“Having abandoned the Christian faith, and disappointed by the claims of Enlightenment reason, [non-believers] find themselves defenceless before reality. To-overcome this anguish they resort to magic, which would allow them to gain the protection of occult powers, and they do not refrain from seeking an alliance with these same powers of evil. [But]

	participation in satanic sects and cults leaves man ever more defenceless before Satan. In this sense it can be supposed that those who belong to satanic sects risk becoming more easily the prey of realities such as witchcraft, the evil eye, diabolical disturbances and demonic possession.” (source)
Possible Scandals:	● His mentor, Emanuele Severino, was censured by the church for teaching “neo-Parminideanism” ● Seriously, there are still people who believe Parmenides around? Man, if you’re going to get in trouble with the Church for teaching heterodox philosophy, why would you choose the *one* philosopher who’s even worse than what the Church already believes?
Probability of Secretly Being A Vampire:	Low
Probability Of Election Precipitating The Apocalypse, According To Questionable Prophecy By Medieval Irish Saint:	Low

Cardinal Peter	

Turkson	
Country of Origin:	Ghana
Background:	Son of a carpenter; this is expected to play well in the Vatican
Position on Abortion:	Pro-life
Key Quote	“Conditions exist for definitively going beyond a Westphalian international order in which the States feel the need for cooperation but do not seize the opportunity to integrate their respective sovereignties for the common good of peoples. Of course, this transformation will be made at the cost of a gradual, balanced transfer of a part of each nation’s powers to a world Authority and to regional Authorities, but this is necessary.” (source)
Possible Scandals:	● Showed sketchy anti-Muslim YouTube video at important conference; video later turned out to be totally fake ● Lack of long-form birth certificate meanswe can’t be sure he’s *really* from the Vatican ●Curse of Ham
Probability of Secretly Being A Vampire:	Low

Probability Of Election Precipitating The Apocalypse, According To Questionable Prophecy By Medieval Irish Saint:	Moderate to high

Cardinal Angelo Bagnasco	
Country of Origin:	Italy
Odd Professorships Held:	Professor of Contemporary Atheism
Position on Abortion:	Pro-life
Key Quote:	“Why say no to forms of legally recognised co-habitation which create alternatives to the family? Why say no to incest? Why say no to the paedophile party in Holland?” (source)
Possible Scandals:	● Keeps getting death threats because of above quote. ● That whole “Professor of Atheism” thing really *is* kind of suspicious
Probability of	Moderate, but I might just be

Secretly Being A Vampire:	saying that because of the picture
Probability Of Election Precipitating The Apocalypse, According To Questionable Prophecy By Medieval Irish Saint:	Low

Cardinal Tarcisio Bertone	
Country of Origin:	Italy
Odd Professorships Held:	Professor of Special Moral Theology
Celebrity Endorsements:	[The ghost of Pope John Paul II, in a dream]()
Position on Abortion:	Pro-ch…no! Wait! Pro-life
Key Policies for Papacy	The Cardinal is a huge soccer fan and used to give amateur commentary on local matches. He has suggested the Vatican participate in international soccer, stating that “we could, in future, field a team that plays at the top level, with Roma, Internazionale, Genoa and Sampdoria” and that “If we just take the

	Brazilian students from our Pontifical universities we could have a magnificent squad”. He later claimed to have been joking.
Possible Scandals:	● Cardinal Bertone is accused of lying about the text of a secret apocalyptic prophecy to cover up a part where Rome is declared to become the throne of the Antichrist
Probability of Secretly Being A Vampire:	Low
Probability Of Election Precipitating The Apocalypse, According To Questionable Prophecy By Medieval Irish Saint:	Low – but if he’s already covered up one apocalyptic prophecy can we really trust him on this one?

Cardinal Christoph Maria Michael Hugo Damian Peter Adalbert, Count von Schönborn	
Really?	Yes
Country of	Czechoslovakia/Austria

Origin:	
Born In Ominous Castle Upon Hilltop?	Yes
Odd Professorships Held:	Professor of Dogmatics
Position on Abortion:	Pro-life
Key Quote:	“The modern world needs badly to hear this message. What frequently passes for modern science —with its heavy accretion of materialism and positivism— is simply wrong about nature in fundamental ways. Modern science is often, in the words of my essay, “ideology, not science.” (source)
Possible Scandals:	● Might have maybe kinda once condemned sexual abuse, proving him to be obviously crazy and unfit for the position. ● Impeccable anti-science credentials, but does he hate gay people enough?
Probability of Secretly Being A Vampire:	The only part of that question stopping me from saying “100%” is the word “secretly”
Would I Still Probably Prefer Him To Anyone Else On This List?	Signs point to yes

Footnote:

1: When I was young, I read a passage from an alternate history book where James Joyce joins the priesthood, was eventually elevated to the papacy under the name Pope Stephen X, and reformed the Church, Irish-style. It had a lasting effect on me and you can find it here. Search for “The Stephenites were the most radical of all the Catholic clergy.“

Sleep

The human mind is endlessly fascinating and I never tire of learning about its various quirks and mazes. I just wish I didn’t have to live in one all the time.

Geologists don’t have to live in caves, and ichthyologists don’t have to publish their papers from within the bellies of giant sharks. All I want is the same kind of distance they and people in so many other scientific fields enjoy.

My latest interesting observation is that my brain has arbitrarily decided to separate all attempts to fall asleep into “sleep mode” and “nap mode”.

Any attempt to fall asleep after 1 AM is “sleep mode”. In sleep mode I want to sleep for about 8 or 9 hours, and when I wake up I’m still pretty groggy and could easily sleep for 2 or 3 more if I let myself.

Any attempt to fall asleep before 1 AM is “nap mode”. In nap mode I want to sleep for about 3 or 4 hours, and when I wake up I feel incredibly refreshed and alert (I also feel a weird but pleasant tightness in my abdomen, which I relate not because I expect anyone to understand it but just in case someone else has had the same experience).

This has somewhat interfered with my attempts to be virtuous. I think “I will stop going to bed so late, and introduce myself to a normal bed-time so I can get stuff done bright and early in the morning. And so I go to sleep at 11 PM. And then I wake up at 3 AM more alert and refreshed than I’ve been in weeks. And then by noon my body is like “Hey, so, I just noticed you only got four hours of sleep last night, I think I’m going to crash and be miserable for the rest of the day.” And then I have to stay up until after 1 AM, or else the cycle repeats itself and I get my sleep schedule *completely* screwed up.

This seems completely incompatible with working a job that has normal hours, let alone a job that starts unusually early. I survived medical school because my body was grudgingly willing to adjust downward to about 11:30 PM for its “sleep mode start time” after weeks of misery, but that’s still not optimal.

Melatonin seems to shift me further into nap mode, which is the opposite of what I hoped it would do. I know all about f.lux and “sleep hygiene” and progressive muscle relaxation and things along those lines, and to the degree that they work they promote sleep for me but don’t favor one mode over the other. I’ve tried just lying in bed after waking up from “nap mode” and after three hours just lying there I usually think “Okay, this is dumb” and count that as a failed experiment.

Has anyone else has a similar experience? If so, has anyone else been able to *solve* it?

Reactionary Philosophy In An Enormous, Planet-Sized Nutshell

I have heard the following from a bunch of people, one of whom was me six months ago: “I keep on reading all these posts by really smart people who identify as Reactionaries, and I don’t have any idea what’s going on. They seem to be saying things that are either morally repugnant or utterly ridiculous. And when I ask them to explain, they say it’s complicated and there’s no one summary of their ideas. Why don’t they just write one?”

Part of me secretly thinks part of the answer is that a lot of these beliefs are not argument but poetry. Try to give a quick summary of Shelley’s *Adonais*: “Well there’s this guy, and he’s dead, and now this other guy is really sad.” One worries something has been lost. And just as well try to give a quick summary of the sweeping elegaic paeans to a bygone age of high culture and noble virtues that is Reaction.

But there *is* some content, and some of it is disconcerting. I started reading a little about Reaction after incessantly being sent links to various Mencius Moldbug posts, and then started hanging out in an IRC channel with a few Reactionaries whom I could question about it. Obviously this makes me the world expert who is completely qualified to embark on the hitherto unattempted project of explaining it to everyone else.

Okay, maybe not. But the fact is, I’ve been itching to prsent an argument against Reactionary thought for a long time, but have been faced with the dual problem of not really having a solid target and worrying that everyone not a Reactionary would think I was wasting my time even talking to them. Trying to sum up their ideas seems like a good way to first of all get a reference point for what their ideas are, and second of all to make it clearer why I think they deserve a rebuttal.

We’ll start with the meta-level question of how confident we should be that our society is better than its predecessors in important ways. Then we’ll look on the object level about how we compare to past societies along dimensions we might care about. We’ll make a lengthy digression into social justice issues, showing how some traditional societies were actually more enlightened than our own in this area. Having judged past societies positively, we’ll then look at what aspects of their cultures, governments, and religions made them so successful, and whether we could adopt those to modern life.

Much of this will be highly politically incorrect and offensive, because that’s what Reactionaries *do*. I have tried to be charitable towards these ideas, which means this post will be pushing politically incorrect and offensive positions. If you do not want to read it, especially the middle parts which are about race, I would totally understand that. But if you do read it and accuse me of holding these ideas myself and get really angry, then you fail at reading comprehension forever.

I originally planned to follow this up tomorrow with the post containing my arguments against these positions, but this argument took longer than I thought to write and I expect the counterargument will as well. Expect a post critiquing reactionary ideas sometime in the next…week? month?

[EDIT: The Anti-Reactionary FAQ is now available]

In any case, this is not that post. This is the post where I argue that modern society is rotten to the core, and that the only reasonable solution is to dig up King James II, clone him, and give the clone absolute control over everything.

– * – * – * –

No One Expects The Spanish Inquisition, Especially Not In 21st Century America

People in ancient societies thought their societies were obviously great. The imperial Chinese thought nothing could beat imperial China, the medieval Spaniards thought medieval Spain was a singularly impressive example of perfection, and Communist Soviets were pretty big on Soviet Communism. Meanwhile, we think 21st-century Western civilization, with its democracy, secularism, and ethnic tolerance is pretty neat. Since the first three examples now seem laughably wrong, we should be suspicious of the hypothesis that *we* finally live in the one era whose claim to have gotten political philosophy right is *totally justified*.

But it seems like we have an advantage they don't. Speak out against the Chinese Empire and you lose your head. Speak out against the King of Spain and you face the Inquisition. Speak out against Comrade Stalin and you get sent to Siberia. The great thing about western liberal democracy is that it has a free marketplace of ideas. *Everybody* criticizes some aspect of our society. Noam Chomsky made a career of criticizing our society and became rich and famous and got a cushy professorship. So our advantage is that we admit our society's imperfections, reward those who point them out, and so keep inching closer and closer to this ideal of perfect government.

Okay, back up. Suppose you went back to Stalinist Russia and you said “You know, people just don’t respect Comrade Stalin enough. There isn’t enough Stalinism in this country! I say we need *two* Stalins! No, *fifty* Stalins!”

Congratulations. You have found a way to criticize the government in Stalinist Russia and *totally get away with it*. Who knows, you might even get that cushy professorship.

If you “criticize” society by telling it to keep doing exactly what it’s doing only much much more so, society recognizes you as an ally and rewards you for being a “bold iconoclast” or “having brave and revolutionary new ideas” or whatever. It’s only when you tell them something they *actually don’t want to hear* that you get in trouble.

Western society has been moving gradually further to the left for the past several hundred years at least. It went from divine right of kings to constutitional monarchy to libertarian democracy to federal democracy to New Deal democracy through the civil rights movement to social democracy to ???. If you catch up to society as it’s pushing leftward and say “Hey guys, I think we should go leftward even faster! Two times faster! No, *fifty* times faster!”, society will call you a bold revolutionary iconoclast and give you a professorship.

If you start suggesting maybe it should switch directions and move the direction opposite the one the engine is pointed, *then* you might have a bad time.

Try it. Mention that you think we should undo something that’s been done over the past century or two. Maybe reverse women’s right to vote. Go back to sterilizing the disabled and feeble-minded. If you *really* need convincing, suggest re-implementing segregation, or how about slavery? See how far freedom of speech gets you.

In America, it will get you fired from your job and ostracized by nearly everyone. Depending on how loudly you do it, people may picket your house, or throw things at you, or commit violence against you which is then excused by the judiciary because obviously they were provoked. Despite the iconic image of the dissident sent to Siberia, this is how the Soviets dealt with most of *their* iconoclasts too.

If you absolutely insist on imprisonment, you can always go to Europe, where there are more than enough “hate speech” laws on the book to satisfy your wishes. But a system of repression that doesn’t involve obvious state violence is little different in effect than one that does. It’s simply more efficient and harder to overthrow.

Reaction isn’t a conspiracy theory; it’s not suggesting there’s a secret campaign for organized repression. To steal an example from the other side of the aisle, it’s positing something more like patriarchy. Patriarchy doesn’t have an actual Patriarch coordinating men in their efforts to keep down women. It’s just that when lots of people share some really strong cultural norms, they manage to self-organize into a kind of immune system for rejecting new ideas. And Western society just happens to have a really strong progressivist immune system ready to gobble you up if you say anything insufficiently progressive.

And so the main difference between modern liberal democracy and older repressive societies is that older societies repressed things you liked, but modern liberal democracies only repress things you don’t like. Having only things you don’t like repressed looks from the inside a lot like there being no repression at all.

The good Catholic in medieval Spain doesn’t feel repressed, even when the Inquisition drags away her neighbor. She feels like decent people have total freedom to worship whichever saint they want, total freedom to go to whatever cathedral they choose, total freedom to debate who the next bishop should be – oh, and thank goodness someone's around to deal with those crazy people who are trying to damn the rest of us to Hell. We medieval Spaniards are way too smart to fall for the balance fallacy!

Wait, You Mean The Invisible Multi-Tentacled Monster That Has Taken Over All Our Information Sources Might Be Trying To *Mislead* Us?

Since you are a citizen of a repressive society, you should be extremely skeptical of all the information you get from schools, the media, and popular books on any topic related to the areas where active repression is occurring. That means at *least* politics, history, economics, race, and gender. You should be *especially* skeptical of any book that's praised as "a breath of fresh air" or "a good counter to the prevailing bias", as books that garner praise in the media are probably of the "We need fifty Stalins!" variety.

This is not nearly as paranoid as it sounds. Since race is the most taboo subject in our culture, it will also be the simplest example. Almost all of our hard data on race comes from sociology programs in universities – ie the most liberal departments in the most liberal institutions in the country. Most of these sociology departments have an explicit mission statement of existing to fight racism. Many sociologists studying race will tell you quite openly that they went into the field – which is not especially high-paying or prestigious – in order to help crusade against the evil of racism.

Imagine a Pfizer laboratory whose mission statement was to prove Pfizer drugs had no side effects, and whose staff all went into pharmacology specifically to help crusade against the evil of believing Pfizer's drugs have side effects. Imagine that this laboratory hands you their study showing that the latest Pfizer drug has zero side effects, c'mon, trust us! Is there *any way* you're taking that drug?

We know that a lot of medical research, especially medical research by drug companies, turns up the wrong answer simply through the file-drawer effect. That is, studies that turn up an exciting result everyone wants to hear get published, and studies that turn up a disappointing result don't – either because the scientist never submits it to the journals, or because the journal doesn't want to publish it. If this happens *all the time* in medical research despite growing safeguards to prevent it, how often do you think it happens in sociological research?

Do you think the average sociologist selects the study design most likely to turn up evidence of racist beliefs being correct, or the study design most likely to turn up the opposite? If despite her best efforts a study does turn up evidence of racist beliefs being correct, do you think she's going to submit it to a major journal with her name on it for everyone to see? And if by some bizarre chance she does submit it, do you think the *International Journal Of We Hate Racism So We Publish Studies Proving How Dumb Racists Are* is going to cheerfully include it in their next edition?

And so when people triumphantly say “Modern science has completely disproven racism, there's not a shred of evidence in support of it”, we should consider that exactly the same level of proof as the guy from 1900 who said “Modern science has completely proven racism, there's not a shred of evidence against it”. The field is still just made of people pushing their own dogmatic opinions and calling them science; only the dogma has changed.

And although Reactionaries love to talk about race, in the end race is nothing more than a particularly strong and obvious taboo. There are taboos in history, too, and in economics, and in political science, and although they’re less obvious and interesting they still mean you need this same skepticism when parsing results from these fields. “But every legitimate scientist disagrees with this particular Reactionary belief!” should be said with the same intonation as “But every legitimate archbishop disagrees with this particular heresy.”

This is not intended as a proof that racism is correct, or even as the slightest shred of evidence for that hypothesis (although a lot of Reactionaries are, in fact, racist as heck). No doubt the Spanish Inquisition found a couple of real Satanists, and probably some genuine murderers and rapists got sent to Siberia. Sometimes, once in a blue moon, a government will even censor an idea that happens to be false. But it’s still useful to know when something is being censored, so you don’t actually think the absence of evidence for one side of the story is evidence of anything other than people on that side being smart enough to keep their mouths shut.

The Past Is A First World Country

Even so, isn’t the evidence that modern society beats past societies kiiiind of overwhelming? We’re richer, safer, healthier, better educated, freer, happier, more equal, more peaceful, and more humane. Reactionary responses to these claims might get grouped into three categories.

The first category is “Yes, obviously”. Most countries do seem to have gotten about 100x wealthier since the year 1700. Disease rates have plummeted, and life expectancy has gone way up – albeit mostly due to changes in infant mortality. But this stands entirely explained by technology. So we’re a hundred times wealthier than in 1700. In what? Gold and diamonds? Maybe that has something to do with the fact that today we’re digging our gold mines with one of these:

…and in 1700 they had to dig their gold mines with one of these:

Likewise, populations are healthier today because they can get computers to calculate precisely targeted radiation bursts that zap cancer while sparing healthy tissue, whereas in 1700 the pinnacle of medical technology was leeches.

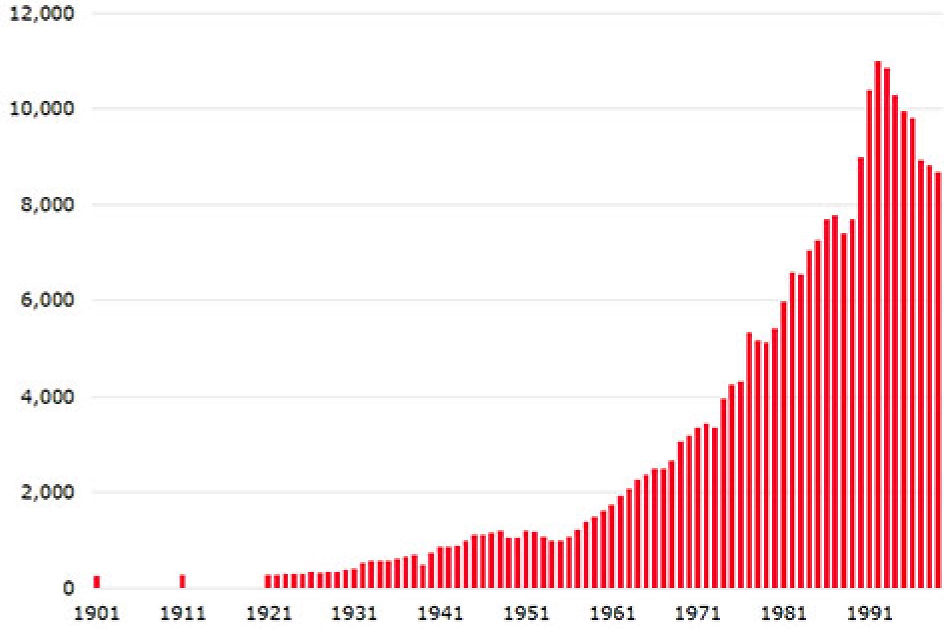
This technology dividend appears even in unexpected places. The world is more peaceful today, but how much of that is the existence of global trade networks that make war unprofitable, video reporting of every casualty that makes war unpopular, and nuclear and other weapons that make war unwinnable?

The second category is “oh really?”. Let’s take safety. This is one of Mencius Moldbug’s pet issues, and he likes to quote the following from an 1876 century text on criminology:

> Meanwhile, it may with little fear of contradiction be asserted that there never was, in any nation of which we have a history, a time in which life and property were so secure as they are at present in England. The sense of security is almost everywhere diffused, in town and country alike, and it is in marked contrast to the sense of insecurity which prevailed even at the beginning of the present century. There are, of course, in most great cities, some quarters of evil repute in which assault and robbery are now and again committed. There is perhaps to be found a lingering and flickering tradition of the old sanctuaries and similar resorts. But any man of average stature and strength may wander about on foot and alone, at any hour of the day or the night, through the greatest of all cities and its suburbs, along the high roads, and through unfrequented country lanes, and never have so much as the thought of danger thrust upon him, unless he goes out of his way to court it.

Moldbug then usually contrasts this with whatever recent news article has struck his fancy about entire inner-city neighborhoods where the police are terrified to go, teenagers being mowed down in crossfire among gangs, random daylight murders, and the all the other joys of life in a 21st century British ghetto.

Of course, the plural of anecdote is not data, but the British crime statistics seem to bear him out:



(recorded offenses per 100,000 people, from source)

If this is true, it is true *despite* technology. If crime rates have in fact multiplied by a factor of…well, it looks like at least 100x…this is true even though the country as a whole has gotten vastly richer, even though there are now CCTVs, DNA testing, police databases, heck, even fingerprinting hadn’t been figured out yet in 1876.

This suggests that there was something inherent about Victorian society, politics, or government that made their Britain a safer place to live than modern progressive Britain.

Education is another example of something we’re pretty sure we do better in. Now take a look at the 1899 entrance exam for Harvard. Remember, no calculators – they haven't been invented yet.

I got an SAT score well above that of the average Harvard student today (I still didn't get into Harvard, because I was a slacker in high school). But I couldn't even *begin* to take much of that test.

Okay, fine. Argue "Well, of course we don't value Latin and Greek and arithmetic and geometry and geography today, we value different things." So fine. Tell me what the heck you think our high school students are learning that's just as difficult and impressive as the stuff on that test that you don't expect the 19th century Harvard students who aced that exam knew two hundred times better (and don't say "the history of post-World War II Europe").

Do you honestly think the student body for whom that exam was a fair ability test would be befuddled by the *reading comprehension* questions that pass for entrance exams today? Or would it be more like "Excuse me, teacher, I'm afraid there's been a mistake. My exam paper is in English."

As a fun exercise, read through Wikipedia's list of multilingual presidents of the United States. We start with entries like this one:

> Thomas Jefferson read a number of different languages. In a letter to Philadelphia publisher Joseph Delaplaine on April 12, 1817, Jefferson claimed to read and write six languages: Greek, Latin, French, Italian, Spanish, and English. After his death, a number of other books, dictionaries, and grammar manuals in various languages were found in Jefferson's library, suggesting that he studied additional languages beyond those he

> spoke and wrote well. Among these were books in Arabic, Gaelic, and Welsh.

and this one:

> John Quincy Adams went to school in both France and the Netherlands, and spoke fluent French and conversational Dutch. Adams strove to improve his abilities in Dutch throughout his life, and at times translated a page of Dutch a day to help improve his mastery of the language. Official documents that he translated were sent to the Secretary of State of the United States, so that Adams’ studies would serve a useful purpose as well. When his father appointed him United States Ambassador to Prussia, Adams dedicated himself to becoming proficient in German in order to give him the tools to strengthen relations between the two countries. He improved his skills by translating articles from German to English, and his studies made his diplomatic efforts more successful. In addition to the two languages he spoke fluently, he also studied Italian, though he admitted to making little progress in it since he had no one with whom to practice speaking and hearing the language. Adams also read Latin very well, translated a page a day of Latin text, and studied classical Greek in his spare time.

eventually proceeding to entries more like this one:

> George W. Bush speaks some amount of Spanish, and has delivered speeches in the language. His speeches in Spanish have been imperfect, with English dispersed throughout. Some pundits, like Molly Ivins, have pointedly questioned the extent to which he could speak the language, noting that he kept to similar phrasing in numerous appearances.

and this one:

> Barack Obama himself claims to speak no foreign languages. However, according to the President of Indonesia Susilo Bambang Yudhoyono, during a telephone conversation Obama was able to deliver a basic four-word question in “fluent Indonesian”, as well as mention the names for a few Indonesian food items. He also knows some Spanish, but admits to only knowing “15 words” and having a poor knowledge of the language.

A real Reactionary would no doubt point out that even old-timey US Presidents aren’t old-timey enough, and that we really should be looking at the British aristocracy, but this is left as an exercise for the reader.

It may be argued that yes, maybe their aristocracy was more educated than our upper-class, but we compensate for the imbalance by having education spread much more widely among the lower-classes. I endorse this position, as do, I’m sure, the hundreds of inner-city minority youth who are no doubt reading this blog post because of the massive interest in abstract political philosophy their schooling has successfully inspired in them.

Once again, today we have Wikipedia, the Internet, and as many cheap books as Amazon can supply us. Back in the old days they had to make do with whatever they could get from their local library. Even more troubling, today we start with a huge advantage – the Flynn Effect has made our average IQ 10 to 20 points higher than in 1900. Yet once again, even with our huge technological and biological head start, we are *still* doing worse than the Old Days, which suggest that here, too, the Old Days may have had some kind of social/political advantage.

So several of our claims of present superiority – wealth, health, peace, et cetera – have been found to be artifacts of higher technology levels. Several other claims – safety and education – have been found to be just plain wrong. That just leaves a few political advantages – namely, that we are freer, less racist, less sexist, less jingoistic and more humane. And the introduction has already started poking holes in the whole "freedom" thing.

That leaves our progress in tolerance, equality, and humaneness. Are these victories as impressive as we think?

Every Time I Hear The Word "Revolver", I Reach For My Culture

[TRIGGER WARNING: This is the part with the racism]

One of the most solid results from social science has been large and persistent differences in outcomes across groups. Of note, these differences are highly correlated by goodness: some groups have what we would consider "good outcomes" in many different areas, and others have what we would consider "bad outcomes" in many different areas. Crime rate, drug use, teenage pregnancy, IQ, education level, median income, health, mental health, and whatever else you want to measure.

The best presentation of this result is The Spirit Level, even though the book *thinks* it's proving something completely different. But pretty much any study even vaguely in this field will show the same effect. This also seems to be the intuition behind our division of countries into "First World" and "Third World", and behind our division of races into "privileged" and "oppressed" (rather than "well, some races have good outcomes in some areas, but others have good outcomes in other areas, so it basically all balances out") I don’t think this part should be very controversial. Let’s call this mysterious quality “luck”, in order to remain as agnostic as possible about the cause.

Three very broad categories of hypothesis have been proposed to explain luck differences among groups: the external, the cultural, and the biological.

The externalists claim that groups differ only because of the situations they find themselves in. Sometimes these situations are natural. Jared Diamond makes a cogent case for the naturalist externalist hypothesis in *Guns, Germs, and Steel*. The Chinese found themselves on fertile agricultural land with lots of animals and plants to domesticate and lots of trade routes to learn new ideas from. The New Guinea natives found themselves in a dense jungle without many good plants or animals and totally cut off from foreign contact. Therefore, the Chinese developed a powerful civilization and the New Guineans became a footnote to history.

But in modern times, externalists tend to focus more on external *human* conditions like colonialism and oppression. White people are lucky not because of any inherent virtue, but because they had a head start and numerical advantage and used this to give themselves privileges which they deny to other social groups. Black people are unlucky not because of any inherent flaw, but because they happened to be stuck around white people who are doing everything they can to oppress them and keep them down. This is true both within societies, where unlucky races are disprivileged by racism, and across societies, where unlucky countries suffer the ravages of colonialism.

The culturalists claim that luck is based on the set of implicit traditions and beliefs held by different groups. The Chinese excelled not only because of their fertile landscape, but because their civilization valued scholarship, wealth accumulation, and nonviolence. The New Guineans must have had less useful values, maybe ones that demanded strict conformity with ancient tradition, or promoted violence, or discouraged cooperation.

Like the externalists, they trace this forward to the present, saying that the values that served the Chinese so well in building Chinese civilization are the same ones that keep China strong today and the ones that make Chinese immigrants successful in countries like Malaysia and the USA. On the other hand, New Guinea continues to be impoverished and although I've never heard of any New Guinean immigrants I would not expect them to do very well.

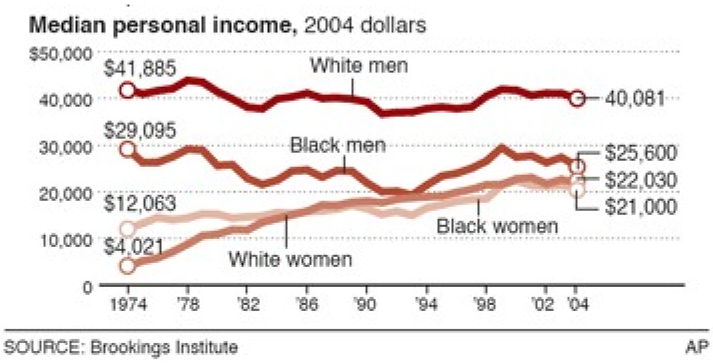
The biologicalists, for whom I cannot think of a less awkward term, are probably the most notorious and require the least explanation. They are most famous for attributing between-group luck differences to genetic factors, but there are certainly more subtle theories. One of the most interesting is parasite load, the idea that areas with greater parasites make people's bodies spend more energy fighting them off, leading to less energy for full neurological development. It's hard to extend this to deal with group differences in a single area (for example between-race differences in the USA) but some people have certainly made valiant attempts. Nevertheless, it's probably fair enough to just think of the biologicalists as "more or less racists".

So who is right?

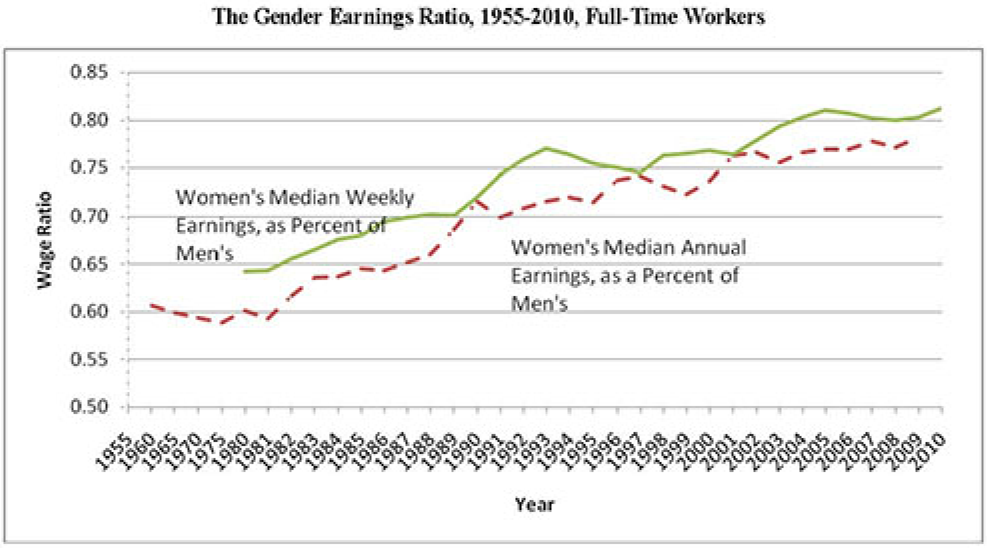
A decent amount of political wrangling over the years seems to involve a conflict between the conservatives – who are some vague mix of the culturalist and biologicalist position –

and the liberals, who have embraced the externalist position with gusto.

But the externalist position is deeply flawed. This blog has already cited this graph to make a different point, but now that we have our Reactionary Hat on, let’s try it again:



Here’s the black-white income gap over time from 1974 to (almost) the present. Over those years, white oppression of black people has decreased drastically. It is not gone. But it has decreased. Yet the income gap stays exactly the same. Compare this to another example of an oppressed group suddenly becoming less oppressed:



Over the same period, the decrease in male oppression of women has resulted in an obvious and continuing rise in women’s incomes. This suggests that the externalist hypothesis of women’s poor incomes was at least partly correct. But an apparent corollary is that it casts doubt on the externalist hypothesis of racial income gaps.

And, in fact, not all races have a racial income gap, and not all those who do have it in the direction an externalist theory would predict. Jews and Asians faced astounding levels of discrimination when they first came to the United States, but both groups recovered quickly and both now do significantly better than average white Americans. Although the idea of a “Jewish conspiracy” is rightly mocked as anti-Semitic and stupid, it is only bringing the externalist hypothesis – that differences in the success of different races must always be due to oppression – to their natural conclusion.

In fact, Jews and Chinese are interesting in that both groups are widely scattered, both groups often find themselves in very hostile countries, and yet both groups are usually more successful than the native population wherever they go (income and education statistics available upon request). Whether it is Chinese in Malaysia or Jews in France, they seem to do unusually well for themselves despite the constant discrimination. If this is an experiment to distinguish between culturalist and externalist positions, it is a very well replicated one.

This difference in the success of immigrant groups is often closely correlated with the success of the countries they come from. Japan is very rich and advanced, Europe quite rich and advanced, Latin America not so rich or advanced, and Africa least rich and advanced of all. And in fact we find that Japanese-Americans do better than European-Americans do better than Latin-Americans do better than African-Americans. It is pretty amazing that white people manage to modulate their oppression in quite this precise a way, especially when it includes oppressing themselves.

And much of the difference between groups is in areas one would expect to be resistant to oppression. Unlucky groups tend to have higher teenage pregnancy rates, more drug use, and greater intra-group violence, *even when comparing similar economic strata*. That is, if we focus on Chinese-Americans who earn $60,000/year and African-Americans who earn $60,000/year, the Chinese will have markedly better outcomes (I’ve seen this study done in education, but I expect it would transfer). Sampling from the same economic stratum screens off effects from impoverished starting conditions or living in bad neighborhoods, and it’s hard (though of course not impossible) to figure out other ways an oppressive majority could create differential school attendance in these groups.

So luck differences are sometimes in favor of oppressed minorities, do not decrease when a minority becomes less oppressed, correlate closely across societies with widely varying amounts of oppression, and operate in areas where oppression doesn’t provide a plausible mechanism. The externalist hypothesis as a collection of natural factors a la Jared Diamond may have merit, but as an oppression-based explanation for modern-day group differences, it fails miserably.

I don’t want to dwell on the biological hypothesis too much, because it sort of creeps me out even in a “let me clearly explain a hypothesis I disagree with” way. I will mention that it leaves a lot unexplained, in that many of the “groups” that have such glaring luck differences are not biological groups at all, but rather religious groups such as the Mormons and the Sikhs, both of whom have strikingly different outcomes than the populations they originated from. Even many groups that are biologically different just aren't different *enough* – the English and Irish have strikingly different luck, but attributing that to differences between which *exact* tiny little branch of the Indo-European tree they came from seems like a terrible explanation.

Nevertheless, the people who dismiss the biological hypothesis as obviously stupid and totally discredited (by which I mean everyone) are doing it a disservice. For a sympathetic and extraordinarily impressive defense of the biological hypothesis I recommend this unpublished (and unpublishable) review article. I will add that I am *extremely* interested in comprehensive takedowns of that article (preferably a full fisking) and that if you have any counterevidence to it at all you should post it in the comments and I will be eternally grateful.

But for now I'm just going to say let's assume by fiat that the biologicalist hypothesis is false, because even with my Reactionary Hat on I find the culturalist hypothesis much more interesting.

The culturalist hypothesis avoids the pitfalls of both the externalist and biological explanations. Unlike the externalists, it can explain why some minority groups are so successful and why group success correlates across societies and immigrant populations. And unlike the biologicalists, it can explain the striking differences between biologically similar groups like the Mormons and the non-Mormon Americans, or the Sikhs and the non-Sikh Indians.

It can also explain some other lingering mysteries, like why a country that's put so much work into keeping black people down would then turn around and elect a black president. Obama was born to an African father and a white mother, raised in Indonesia, and then grew up in Hawaii. At no point did he have much contact with African-American culture, and so a culturalist wouldn't expect his life outcomes to be correlated with those of other African-Americans.

Best of all, despite what the average progressive would tell you the culturalist position isn't really *that* racist. It's a bit like the externalist position in attributing groups' luck to initial conditions, except instead of those initial conditions being how fertile their land is or who's oppressing them, it's what memeplexes they happened to end out with. Change the memeplexes and you can make a New Guinean population achieve Chinese-level outcomes – or vice versa.

The Other Chinese Room Experiment

Assuming we tentatively accept the culturalist hypothesis, what policies does it suggest?

Well, the plan mentioned in the last paragraph of the last section – throw Chinese memes at the people of New Guinea until they achieve Chinese-style outcomes – higher income, less teenage pregnancy, lower crime rates. It doesn't seem like a bad idea. You could try exposing them to Chinese people and the Chinese way of life until some of it stuck. This seems like a good strategy for China, a country whose many problems definitely do not include "a shortage of Chinese people".

On the other hand, in somewhere more like America, one could be forgiven for immediately rounding this off to some kind of dictatorial brainwashing policy of stealing New Guinean infants away from their homes and locking them in some horrible orphanage run by Chinese people who beat them every time they try to identify with their family or native culture until eventually they absorb Chinese culture through osmosis. This sounds bad.

Luckily, although we don’t have quite as many Chinese people as China, we still have a majority culture whose outcomes are *almost* as good as China’s and which, as has been mentioned before, permeates every facet of life and every information source like a giant metastasizing thousand-tentacled monster. So in theory, all we need to do is wait for the unstoppable monster to get them.

This strategy, with the octopoid abomination metaphor replaced with a melting pot metaphor for better branding, has been America’s strategy for most of the past few centuries – assimilation. It worked for the Irish, who were once viewed with as much racism as any Hispanic or Arab is today. It worked for the Italians, who were once thought of as creepy Papist semi-retarded mafia goons until everyone decided no, they were indistinguishable from everyone else. It worked for the fourth and fifth generation Asians, at least here in suburban California, where they’re considered about as “exotic” as the average Irishman. It certainly worked for the Jews, where there are some people of Jewish descent who aren’t even *aware* of it until they trace their family history back. And it should be able to work for everyone else. Why isn’t it?

The Reactionary’s answer to this is the same as the Reactionary’s answer to almost everything: because of those darned progressives!

Sometime in the latter half of this century, it became a point of political pride to help minorities resist “cultural imperialism” and the Eurocentric norms that they should feel any pressure to assimilate. Moved by this ideology, the government did everything it could to help minorities avoid assimilation and to shame and thwart anyone trying to get them to assimilate.

There's a story – I've lost the original, but it might have been in Moldbug – about a state noticing that black children were getting lower test scores. It decided, as progressivists do, that the problem was that many of the classes were taught by white teachers, and that probably this meant the black children couldn't relate to them and were feeling oppressed. So they sent the white teachers off to whiter areas and hiring only black teachers for the black schools, and – sure enough – test scores plummeted further.

California had a sort of similar problem when I was growing up. Most schools were required to teach our large Hispanic immigrant population using bilingual education – that is, teaching them in their native Spanish until they were ready to learn English. The "ready to learn English" tended not to happen, and some people proposed that bilingual education be scrapped. There was a *huge* ruckus where the people in favor of this change were accused of being vile racists who hated Mexicans and wanted to destroy Mexican culture. Thanks to California's colorful proposition system, it passed anyway. And sure enough, as soon as the Hispanics started getting integrated with everyone else and taught in English, test scores went way up.

But this is a rare victory, and we are still very much in "try to prevent assimilation mode". I went to elementary school just as the "melting pot" metaphor was being phased out in favor of the more politically correct "salad bowl" one – in a melting pot, everyone comes together and becomes alike, but in a salad bowl, everything comes together but stays different, and that's fine.

One externalist argument why minorities sometimes do poorly in school is the fear of "acting white" – that their peers tell them that academic achievement is a form of "acting white" by which they betray their cultural heritage.

Unfortunately, we seem to be promoting this on a social level, telling people that assimilating and picking up the best features of majority culture are “acting white”. If the majority culture has useful memes that help protect people against school dropout, crime, and other bad life outcomes, that is a really bad thing to do.

So let’s go back to the nightmare scenario with which we started this section – of children being seized from their homes and locked in a room with Chinese people. Is this sort of dystopia the inevitable result of trying to use culturalist theories to equalize group outcomes?

No. There is a proverb beloved of many Reactionaries: “If you find yourself in a hole, stop digging.” We could make great strides in solving inequality merely by *ceasing to exert deliberate effort to make things worse*. The progressive campaign to demonize assimilation and make it taboo to even talk about some cultures being better adapted than others prevents the natural solution to inequality which worked for the Irish and the Asians and the Jews from working for the minorities of today. If we would *just stop digging the hole deeper* in order to make ourselves feel superior to our ancestors, we’d have gone a lot of the way – maybe not all of the way, but a lot of it – toward solving the problem.

On Second Thought, Keep Your Tired And Poor To Yourself

Immigration doesn’t have to be a problem. In a healthy society, immigrants will be encouraged to assimilate to the majority culture, and after a brief period of disorientation will be just as successful and well-adapted as everyone else.

But in an unhealthy society like ours that makes assimilation impossible, a culturalist will be very worried about immigration.

Let’s imagine an idyllic socialist utopia with a population of 100,000. In Utopia, everyone eats healthy organic food, respects the environment and one another, lives in harmony with people of other races, and is completely non-violent. One day, the Prime Minister decides to open up immigration to Americans and discourage them from assimilating.

50,000 Americans come in and move into a part of Utopia that quickly becomes known as Americatown. They bring their guns, their McDonalds, their megachurches, and their racism.

Soon, some Utopians find their family members dying in the crossfire between American street gangs. The megachurches convert a large portion of the Utopians to evangelical Christianity, and it becomes very difficult to get abortions without being harassed and belittled. Black and homosexual Utopians find themselves the target of American hatred, and worse, some young Utopians begin to get affected by American ideas and treat them the same way. American litter fills the previously pristine streets, and Americans find some loopholes in the water quality laws and start dumping industrial waste into the rivers.

By the time society has settled down, we have a society which is maybe partway between Utopia and America. The Americans are probably influenced by Utopian ideas and not quite as bad as their cousins who reminded behind in the States, but the Utopians are no longer as idyllic as their Utopian forefathers, and have inherited some of America’s problems.

Would it be *racist* for a Utopian to say “Man, I wish we had never let the Americans in?” Would it be *hateful* to suggest that the borders be closed before even more Americans can enter?

If you are a culturalist, no. Utopian culture is better, at least by Utopian standards, than American culture. Although other cultures can often contribute to enrich your own, there is no law of nature saying that only the good parts of other cultures will transfer over and that no other culture can be worse than yours in any way. The Americans were clearly worse than the Utopians, and it was dumb of the Utopians to let so many Americans in without any safeguards.

Likewise, there are countries that are worse than America. Tribal Afghanistan seems like a pretty good example. Pretty much everything about tribal Afghanistan is horrible. Their culture treats women as property, enforces sharia law, and contains honor killings as a fact of life. They tend to kill apostate Muslims and non-Muslims a lot. Not all members of Afghan tribes endorse these things, but the average Afghan tribesperson is much more likely to endorse them than the average American. If we import a bunch of Afghan tribesmen, their culture is likely to make America a worse place in the same way that American culture makes Utopia a worse place.

But it’s actually much worse than this. We are a democracy. Anyone who moves here and gains citizenship eventually gets the right to vote. People with values different from ours vote for people and laws different from those we would vote for. Progressives have traditionally viewed any opposition to this as anti-immigrant and racist – and, by total coincidence, most other countries, and therefore most immigrants, are progressive.

Imagine a country called Conservia, a sprawling empire of a billion people that has a fifth-dimensional hyperborder with America. The Conservians are all evangelical Christians who hate abortion, hate gays, hate evolution, and believe all government programs should be cut.

Every year, hundreds of thousands of Conservians hop the hyperborder fence and enter America, and sympathetic presidents then pass amnesty laws granting them citizenship. As a result, the area you live – or let’s use Berkeley, the area I live – gradually becomes more conservative. First the abortion clinics disappear, as Conservian protesters start harassing them out of business and a government that must increasingly pander to Conservians doesn’t stop them. Then gay people stop coming out of the closet, as Conservian restaurants and businesses refuse to serve them and angry Conservian writers and journalists create an anti-gay climate. Conservians vote 90% Republican in elections, so between them and the area’s native-born conservatives the Republicans easily get a majority and begin defunding public parks, libraries, and schools. Also, Conservians have one pet issue which they promote even more intently than the destruction of secular science – that *all Conservians illegally in the United States must be granted voting rights, and that no one should ever block more Conservians from coming to the US*.

Is this fair to the native Berkeleyans? It doesn’t seem that way to me. And what if 10 million Conservians move into America? That’s not an outrageous number – there are more Mexican immigrants than that. But it would be enough to have thrown every single Presidential election of the past fifty years to the Republicans – there has never been a Democratic candidate since LBJ who has won the native population by enough of a margin to outweight the votes of ten million Conservians.

But isn't this incredibly racist and unrealistic? An entire nation of people whose votes skew 90% Republican? No. African-Americans' votes have historically been around 90% Democratic (93% in the last election). Latinos went over 70% Democratic in the last election. For comparison, white people were about 60% Republicans. If there had been no Mexican immigration to the United States over the past few decades, Romney would probaby have won the last election.

Is it wrong for a liberal citizen of Berkeley in 2013 to want to close the hyperborder with Conservia so that California doesn't become part of the Bible Belt and Republicans don't get guaranteed presidencies forever? Would that citizen be racist for even considering this? If not, then pity the poor conservative, who is actually in this exact situation right now.

(a real Reactionary would hasten to add this is more proof that progressives control everything. Because immigration favors progressivism, any opposition to it is racist, but the second we discover the hyperborder with Conservia, the establishment will figure out some reason why *allowing* immigration is racist. Maybe they can call it "inverse colonialism" or something.)

None of this is an argument against immigration. It's an argument against immigration by groups with bad Luck and with noticeably different values than the average American. Let any Japanese person who wants move over. Same with the Russians, and the Jews, and the Indians. Heck, it's not even like it's saying no Afghans – if they swear on a stack of Korans that they're going to try to learn English and not do any honor killings, they could qualify as well.

The United States used to have a policy sort of like this. It was called the Immigration Act of 1924. Its actual specifics were dumb, because it banned for example Asians and Jews, but the principle behind it – groups with good outcomes and who are a good match for our values can immigrate as much as they want, everyone else has a slightly harder time – seems broadly wise. So of course progressives attacked it as racist and Worse Than Hitler and it got repealed in favor of the current policy: *everyone* has a really hard time immigrating but if anyone sneaks over the border under cover of darkness we grant them citizenship anyway because not doing that would be mean.

Once again, coming up with a fair and rational immigration policy wouldn't require some incredibly interventionist act of state control. It would just require that we notice the hole we've been deliberately sticking ourselves in and *stop digging*.

Imperialism Strikes Back

In an externalist/progressive worldview, the best way to help disadvantaged minorities is to eliminate the influence of more privileged majority groups. In a culturalist/Reactionary worldview, the best way to help disadvantaged minorities is to try to maximize the influence of more privileged majority groups. This suggests re-examining colonialism. But first, a thought experiment.

Suppose you are going to be reincarnated as a black person (if you are already black, as a different black person). You may choose which country you will be born in; the rest is up to Fate. What country do you choose?

The top of my list would be Britain, with similar countries like Canada and America close behind. But what if you could only choose among majority-black African countries?

Several come to my mind as comparatively liveable. Kenya. Tanzania. Botswana. South Africa. Namibia (is your list similar?) And one thing these places all have in common was being heavily, *heavily* colonized by the British.

We compare the sole African country that was never colonized, Ethiopia. Ethiopia has become a byword for senseless suffering thanks to its coups, wars, genocides, and especially famines. This seems like counter-evidence to the "colonialism is the root of all evil" hypothesis.

Yes, colonization had some horrible episodes. Anyone who tries to say King Leopold II was anything less than one of the worst people who ever lived has zero right to be taken seriously. On the other hand, eventually the Belgian people got outraged enough to take it away from Leopold, after which there follows a fifty year period that was the only time in history when the Congo was actually a kind of nice place. Mencius Moldbug likes to link to a Time magazine article from the 1950s praising the peace and prosperity of the Congo as a model colony. Then in 1960 it became independent, and I don't know what happens next because the series of civil wars and genocides and corrupt warlords after that are so horrible that I can't even read all the way through the articles about them. Seriously, not necessarily in numbers but in sheer graphic brutality it is worse than the Holocaust, the Inquisition, and Mao combined and you *do not want to know* what makes me say this.

So yes, Leopold II is one of history's great villains, but once he was taken off the scene colonial Congo improved markedly. And any attempt to attribute the nightmare that is the modern Congo to colonialism has to cope with the historical fact that the post-Leopold colonial Congo was actually pretty nice until it was decolonized at which point it immediately went to hell.

So the theory that colonialism is the source of all problems has to contend with the observation that heavily colonized countries are the most liveable, the sole never-colonized country is among the least liveable, and countries' liveability plummeted drastically as soon as colonialism stopped.

But let's stop picking on Africans. Suppose you are going to be reincarnated as a person of Middle Eastern descent (I would have said "Arab", but then we would get into the whole 'most Middle Easterners are not Arabs' debate). Once again, you can choose your country. Where do you go?

Once again, Britain, US, or somewhere of that ilk sound like your best choices.

Okay, once again we're ruling that out. You've got to go somewhere in the Middle East.

Your best choice is one of those tiny emirates where everyone is a relative of the emir and gets lots of oil money and is super-rich: I would go with Qatar. Let's rule them out too.

Your next-best choice is Israel.

Yes, Israel. Note that I am *not* saying the Occupied Palestinian Territories; that would be just as bad a choice as you expect. I'm saying Israel, where 20% of the population is Arab, and about 16% Muslim.

Israeli Arabs earn on average about $6750 per year. Compare this to conditions in Israel's Arab neighbors. In

Egypt, average earnings are $6200; in Jordan, $5900; in Syria, only $5000.

Aside from the economics, there are other advantages. If you happen to be Muslim, you will have a *heck* of a lot easier time practicing your religion freely in Israel than in some Middle Eastern country where you follow the wrong sect of Islam. You'll be allowed to vote for your government, something you can't do in monarchical Jordan or war-torn Syria, and which Egypt is currently having, er, severe issues around. You can even criticize the government as much as you want (empirically quite a lot), a right Syrian and Egyptian Arabs are currently dying for. Finally, you get the benefit of living in a clean, safe, developed country with good health care and free education for all.

I'm not saying that Israeli Arabs aren't discriminated against or have it as good as Israeli Jews. I'm just saying they have it better than Arabs in most other countries. Once again, we find that colonialism, supposed to be the root of all evil, is actually preferable to non-colonialism in most easily measurable ways.

It may be the case that pre-colonial societies were better than either colonial or post-colonial societies. I actually suspect this is true, in a weird Comanche Indians are better than all of us sort of sense. But "pre-colonial" isn't a choice nowadays. Nowadays it's "how much influence do we want the better parts of the West to have over countries that have already enthusiastically absorbed the worst parts of the West?" Whatever I may feel about the Safavid Dynasty, I would at least rather be born in Afghanistan-post-American-takeover than Afghanistan-pre-American-takeover.

So does this mean some sort of nightmarish "invade every country in the world, kill their leadership, and replace them with Americans, for their own good" type scenario?

Once again, no. Look at China. They've been quietly colonizing Africa for a decade now, and the continent has never been doing better. And by "colonizing", I mean "investing in", with probably some sketchy currying of influence and lobbying and property-gathering going on on the side. It's been great for China, it's been a hugely successful injection of money and technology into Africa, and they probably couldn't have come up with a better humanitarian intervention if they had been trying.

Why hasn't the West done it? Because every time an idea like that has been mooted, the progressives have shot it down with "You neo-colonialist! You're worse than King Leopold II, who was himself worse than Hitler! By the transitive property, you are *worse than Hitler*!"

No one needs to go about invading anyone else or killing their government. But if you find yourself in a hole, *stop digging*.

The Uncanny Valley Of Dictatorship

I kind of skimmed over the Palestinian Territories in the last section. They are, indeed, a terrible dehumanizing place and the treatment of their citizens is an atrocity that blemishes a world which allows it to continue. Is this a strike against colonialism?

Any 19th century European aristocrat looking at the Palestinian Territories would note that Israel is being a *terrible colonizer*, not in a moral sense but in a purely observational sense. It's not getting any money or resources out of its colony at all! It's letting people totally just protest it and get away with it! They've even handed most of it over to a government of natives! Queen Victoria would *not* be amused.

Suppose a psychopath became Prime Minister of Israel (yes, obvious joke is obvious). He declares: “Today we are annexing the Palestinian territories. All Palestinians become Israeli residents with most of the rights of citizens except they can’t vote. If anyone speaks out against Israel, we’ll shoot them. If anyone commits a crime, we’ll shoot them.” What would happen?

Well, first, a lot of people would get shot. After that? The Palestinians would be in about the same position as Israeli Arabs are today, except without the right to vote, plus they get shot if they protest. This is vastly better than the position they’re in now, and better than the position of say the people of Syria who are poorer, *also* lack the right to vote, and *also* empirically get shot if they protest.

No more worries about roadblocks. No more worries about passports. No more worries about sanctions. No more worries about economic depression. The only worry is getting shot, and you can avoid that by never speaking out against Israel. Optimal? Probably not. A heck of a lot better than what the Palestinians have today? Seems possible.

It seems like there’s an uncanny valley of dictatorship. Having no dictator at all, the way it is here in America, is very good. Having a really really dictatorial dictator who controls everything, like the czar or this hypothetical Israeli psychopath, kinda sucks but it’s peaceful and you know exactly where you stand. Being somewhere in the middle, where it’s dictatorial enough to hurt, but not dictatorial enough for the dictator to feel secure enough to mostly leave you alone except when he wants something, is worse than either extreme.

Mencius Moldbug uses the fable of Fnargl, an omnipotent and invulnerable alien who becomes dictator of Earth. Fnargl is an old-fashioned greedy colonizer: he just wants to exploit Earth for as much gold as possible. He considers turning humans into slaves to work in gold mines, except some would have to be a special class of geologist slaves to plan the gold mines, and there would have to be other slaves to grow food to support the first two classes of slaves, and other slaves to be managers to coordinate all these other slaves, and so on. Eventually he realizes this is kind of dumb and there's already a perfectly good economy. So he levies a 20% tax on every transaction (higher might hurt the economy) and uses the money to buy gold. Aside from this he just hangs out.

Fnargl has no reason to ban free speech: let people plot against him. He's omnipotent and invulnerable; it's not going to work. Banning free speech would just force him to spend money on jackbooted thugs which he could otherwise be spending on precious, precious gold. He has no reason to torture dissidents. What are they going to do if left unmolested? *Overthrow* him?

Moldbug claims that Fnargl's government would not only be better than that of a less powerful human dictator like Mao, but that it would be *literally better than the government we have today*. Many real countries *do* restrict free speech or torture dissidents. And if you're a libertarian, Fnargl's "if it doesn't disrupt gold production, I'm okay with it" line is a dream come true.

So if the Israelis want to improve the Palestinian Territories' plight, they can do one of two things. First, they can grant it full independence. Second, they do exactly the opposite: can take away all of its independence and go full Fnargl.

We already know Israel doesn’t want to just grant full independence, which leaves “problem continues forever” or “crazy psychopath alien solution”. Could the latter really work?

Well, no. Why not? Because the Palestinians would probably freak out and start protesting *en masse* and the Israelis would have to shoot all of them and that would be horrible.

But it’s worth noting this is not just a natural state of the world. The British successfully colonized Palestine for several decades. They certainly tried the Fnargl approach: “No way you’re getting independence, so just sit here and deal with it or we shoot you.” It worked pretty well then. I would hazard a guess to say the average Palestinian did much better under British rule than they’re doing now. So why wouldn’t it work again?

In a word, progressivism. For fifty years, progressives have been telling the colonized people of the world “If anyone colonizes you, this is the worst thing in the world, and if you have any pride in yourself you must start a rebellion, even a futile rebellion, immediately.” This was non-obvious to people a hundred years ago, which is why people rarely did it. It was only after progressivism basically told colonized peoples “You’re not revolting yet? What are you, *chicken*?” that the modern difficulties in colonialism took hold. And it’s only after progressivism gained clout in the countries that rule foreign policy that it became politically impossible for a less progressive country to try colonialism.

If not for progressivism, Israel would have been able to peacefully annex the Palestinian territories as a colony with no more of a humanitarian crisis than Britain annexing New Zealand or somewhere. Everything would have been solved and everyone could have gone home in time for tea.

Once again, the problem with these holes is that we *keep digging them*. Maybe if we'd stop, there wouldn't be so many holes anymore.

Humane, All Too Humane

There seem to be similar uncanny valley effects in the criminal justice system and in war.

Modern countries pride themselves on their humane treatment of prisoners. And by "humane", I mean "lock them up in a horrible and psychologically traumatizing concrete jail for ten years of being beaten and raped and degraded, sometimes barely even seeing the sun or a green plant for that entire time, then put it on their permanent record so they can never get a good job or interact with normal people ever again when they come out."

Compare this to what "inhumane" countries that were still into "cruel and unusual punishment" would do for the same crime. A couple of lashes with the whip, then you're on your way.

Reader. You have just been convicted of grand theft auto (the crime, not the game). You're innocent, but the prosecutor was very good at her job and you've used up all your appeals and you're just going to have to accept the punishment. The judge gives you two options:

1) Five years in prison
2) Fifty strokes of the lash

Like everyone else except a few very interesting people who help provide erotic fantasies for the rest of us, I don't like being whipped. But I would choose (2) in a *fraction of a heartbeat*.

And aside from being better for me, it would be better for society as well. We know that people who spend time in prison are both more likely to stay criminals in the future and better at being criminals. And each year in jail costs the State $50,000; more than it would cost to give a kid a year's free tuition at Harvard. Cutting the prison system in half would free up approximately enough money to give free college tuition to all students at the best school they can get into.

But of course we don't do that. We stick with the prisons and the rape and the kids who go work at McDonalds because they can't afford college. Why? *Progressives!*

If we were to try to replace prison with some kind of corporal punishment, progressives would freak out and say we were cruel and inhumane. Since the prison population is disproportionately minority, they would probably get to use their favorite word-beginning-with-"R", and allusions would be made to plantation owners who used to whip slaves. In fact, progressives would come up with some reason to oppose even giving criminals the *option* of corporal punishment (an option most would certainly take) and any politician insufficiently progressive to even recommend it would no doubt be in for some public flagellation himself, albeit of a less literal kind.

So once again, we have an uncanny valley. Being very nice to prisoners is humane and effective (Norway seems to be trying ths with some success), but we're not going to do it because we're dumb and it's probably too expensive anyway. Being very strict to prisoners is humane and effective – the corporal punishment option. But being somewhere in the fuzzy middle is cruel to the prisoners and incredibly destructive to society – and it's the only route the progressives will allow us to take.

Some Reactionaries have tried to apply the same argument to warfare. Suppose that during the Vietnam War, we had nuked Hanoi. What would have happened?

Okay, fine. The Russians would have nuked us and everyone in the world would have died. Bad example. But suppose the Russians were out of the way. Wouldn’t nuking Hanoi be a massive atrocity?

Yes. But compare it to the alternative. Nuking Hiroshima killed about 150,000 people. The Vietnam War killed about 3 million. The latter also had a much greater range of non-death effects, from people being raped and tortured and starved to tens of thousands ending up with post-traumatic stress disorder and countless lives being disrupted. If nuking Hanoi would have been an alternative to the Vietnam War, it would have been a *really really good* alternative.

Most of the countries America invades know they can’t defeat the US military long-term. Their victory condtion is helping US progressives bill the war as an atrocity and get the troops sent home. So the enemy’s incentive is to make the war drag on as long as possible and contain as many atrocities as possible. It’s not too hard to make the war drag on, because they can always just hide among civilians and be relatively confident the US is too humane to risk smoking them out. And it’s never too hard to commit atrocities. So they happily follow their incentives, and the progressives in the US happily hold up their side of the deal by agitating for the troops to be sent home, which they eventually are.

Compare this to the style of warfare in colonial days. “This is our country now, we’re not leaving, we don’t really care about atrocities, and we don’t really care how many civilians we end up killing.” It sounds incredibly ugly, but of colonial Britain or very-insistently-non-colonial USA, guess which one ended up pacifying Iraq after three months with only about 6,000 casualties, and guess which one took five years to re-establish a semblance of order and killed about 100,000 people in the process?

Once again we see an uncanny valley effect. Leaving Iraq alone completely would have been a reasonable humanitarian choice. Using utterly overwhelming force to pacify Iraq by any means necessary would have briefly been very ugly, but our enemies would have folded quickly and with a few assumptions this could also have been a reasonable humanitarian choice. But a wishy-washy half-hearted attempt to pacify Iraq that left the country in a state of low-grade poorly-defined war for nearly a decade was neither reasonable nor humanitarian.

Once again, the solution isn’t some drastic nightmare scenario where all prisoners are tortured and all wars are fought with sarin nerve gas. It’s that if prisoners *prefer* corporal punishment, progressives don’t call “racism!” or “atrocity!” so loudly that it becomes politically impossible to give them what they want. Once again, all we have to do is *stop digging.*

Gender! And Now That I Have Your Attention, Let’s Talk About Sex

So the two things Reactionaries like to complain about all the time are race and sex, and since we have *more* then gone overboard with our lengthy diversion into race, we might as well take a quick look at sex.

As far as I know, even the Reactionaries who are really into biological differences between races don’t claim that women are intellectually inferior to men. I don’t even think they necessarily believe there are biological differences between the two groups. And yet they are not really huge fans of feminism. Why?

Let’s start with some studies comparing gender roles and different outcomes.

Surveys of women show that they were on average happier fifty years ago than they are today. In fact, in the 1950s, women generally self-reported higher happiness than men; today, men report significantly higher happiness than women. So the history of the past fifty years – a history of more and more progressive attitudes toward gender – have been a history of women gradually becoming worse and worse off relative to their husbands and male friends.

This doesn’t *necessarily* condemn progressivism, but as the ancient proverb goes, it sure waggles its eyebrows suggestively and gestures furtively while mouthing ‘look over there’.

To confirm, we would want to look within a single moment in time: that is, are feminist women with progressive gender roles *today* less happy than their traditionalist peers? The answer appears to be yes.

Amusingly, because we *do* still live in a society where these things couldn’t be published unless someone took a progressivist tack, the New York Times article quoted above ends by saying the *real* problem is that men are jerks who don’t do their share of the housework.

But when we actually study this, we find that progressive marriages in which men and women split housework equally are 50% more likely to end in divorce than traditional marriages where the women mostly take care of it. The same is true of working outside the home: progressive marriages where both partners work are more likely to end in divorce than traditional marriages where the man works and the woman stays home.

Maybe this is just because the same people who are progressive enough to defy traditional gender roles are also the same people who are progressive enough not to think divorce is a sin? But this seems unlikely: in general religious people get divorced *more* than the irreligious. And since I did promise we'd be talking about sex, consider the studies showing people in traditional marriages have better sex lives than their feminist and progressive friends. This doesn't seem like something that could easily be explained merely by religion, unless religion has gotten *way* cooler since the last time I attended synagogue.

So why is this? I have heard some reactionaries say that although there are not intellectual differences between men and women, there are emotional differences, and that women are (either for biological or cultural reasons) more "submissive" to men's "dominant" – and a quick search of the BDSM community seems to both to validate the general rule and to showcase some very striking exceptions.

But my money would be on a simpler hypothesis. Every marriage involves conflict. The traditional concept of gender contains two roles that are divided in a time-tested way to minimize conflict as much as possible. In a perfect-spherical-cow sense, either the husband or the wife could step into either role, and it would still work just as well. But since men have been socialized for one role since childhood, and women socialized for the other role, it seems that in most cases the easiest solution is to stick them in the one they've been trained for.

We could also go with a third hypothesis: that *women aren't actually bizarre aliens from the planet Zygra'ax with* *completely inexplicable preferences*. I mean, suppose you had the following two options:

1. A job working from home, where you are your own boss. The job description is “spending as much or as little time as you want with your own children and helping them grow and adjust to the adult world.” (but Sister Y also has a post on the childless alternative to this)

2. A job in the office, where you do have a boss, and she wants you to get her the Atkins report “by yesterday” or she is going to throw your sorry ass out on the street where it belongs, and there *better* not be any complaints about it this time.

Assume both jobs would give you exactly the same amount of social status and respect.

Now assume that suddenly a bunch of people come along saying that *actually*, only losers pick Job 1 and surely you’re not a *loser*, are you? And you have to watch all your former Job 1 buddies go out and take Job 2 and be praised for this and your husband asks why *you* aren’t going into Job 2 and contributing something to the family finances for once, and eventually you just give in and go to Job 2, but also you’ve got to do large portions of Job 1, and also the extra income mysteriously fails to give your family any more money and in fact you are worse off financially than before.

Is it so hard to imagine that a lot of women would be less happy under this new scenario?

Now of course (most) feminists very reasonably say that it’s Totally Okay If You Want To Stay Home And We’re Not Trying To Force Anyone. But let’s use the feminists’ own criteria on that one. Suppose Disney put out a series of movies in which they had lots of great female role models who only worked in the home and were subservient to their husbands all the time, and lauded them as *real* women who were courageous and awesome and sexy and not just poor oppressed stick-in-the-muds, and then at the end they flashed a brief message “But Of Course Working Outside The Home Is Totally Okay Also”. Do you think feminists would respond “Yeah, we have no problem with this, after all they *did* flash that message at the end”?

Aside from being better for women, traditional marriages seem to have many other benefits. They allow someone to bring up the children so that they don’t have to spend their childhood in front of the television being socialized by reruns of *Drug-Using Hypersexual Gangsters With Machine Guns*. They ensure that at least one member of each couple has time to be doing things that every household should be doing anyway, like keeping careful track of finances, attending parent-teacher conferences, and keeping in touch with family.

So do men need to force women to stay barefoot and in the kitchen all the time, and chase Marie Curie out of physics class so she can go home and bake for her husband?

By this point you may be noticing a trend. No, we don’t need to do that. If we stopped optimizing the media to send feminist messages as loud as possible, if we stopped actively opposing any even slightly positive portrayal of a housewife as “sexist” and “behind the times”, and if we stopped having entire huge lobby groups supported vehemently by millions of people *dedicated entirely to making the problem worse*, then maybe things would take care of themselves.

There’s some sort of metaphor here…something about dirt…or a shovel…nah, never mind.

Plays Well In Groups

Suppose you were kidnapped by terrorists, and you needed someone to organize a rescue. Would you prefer the task be delegated to the Unitarians, or the Mormons?

This question isn't about whether you think an *individual* Unitarian or Mormon would make a better person to rush in Rambo-style and get you out of there. It's about whether you would prefer the Unitarian Church or the Mormon Church to coordinate your rescue.

I would go with the Mormons. The Mormons seem *effective* in all sorts of ways. They're effective evangelists. They're effect fundraisers. They're effective at keeping the average believer following their commandments. They would figure out a plan, implement it, and come in guns-blazing.

The Unitarians would be a disaster. First someone would interrupt the discussion to ask whether it's fair to use the word "terrorists", or whether we should use the less judgmental "militant". Several people would note that until investigating the situation more clearly, they can't even be sure the terrorists aren't in the right in this case. In fact, what *is* "right" anyway? An attempt to shut down this discussion to focus more on the object-level problem would be met with cries of "censorship!".

If anyone did come up with a plan, a hundred different pedants would try to display their intelligence by nitpicking meaningless details. Eventually some people would say that it's an outrage that no one's even *considering* whether the bullets being used are recyclable, and decide to split off and mount their *own*, ecologically-friendly rescue attempt. In the end, four different schismatic rescue attempts would run into each other, mistake each other for the enemy, and annhilate themselves while the actual terrorists never even hear about it.

(if it were Reform Jews, the story would be broadly similar, but with *twenty* different rescue attempts, and I say this fondly, as someone who attended a liberal synagogue for ten years)

One relevant difference between Mormons and Unitarians seems to be a cultural one. It's not quite that the Mormons value conformity and the Unitarians value indivduality – that's not exactly *wrong*, but it's letting progressives bend language to their will, the same way as calling the two sides of the abortion debate "pro-freedom" and "anti-woman" or whatever they do nowadays. It's more like a Mormon norm that the proper goal of a discussion is agreement, and a Unitarian norm that the proper goal of a discussion is disagreement.

There's a saying I've heard in a lot of groups, which is something along the lines of "diversity is what unites us". This is nice and memorable, but there are other groups where *unity* is what unites them, and they seem to be more, well, united.

Unity doesn't just arise by a sudden and peculiar blessing of the angel Moroni. It's the sort of thing you can create. Holidays and festivals and weird rituals create unity. If everyone jumps up and down three times on the summer solstice, then yes, objectively this is dumb, but you feel a little more bonded with the other people who do it: *I'm* one of the solstice-jumpers, and *you're* one of the solstice-jumpers, and that makes us solstice-jumpers together.

Robert Putnam famously found that the greater the diversity in a community:

> …the less people vote, the less they volunteer, and the less they give to charity and work on community projects. In the most diverse communities, neighbors trust one another about half as much as they do in the most homogenous settings. The study, the largest ever on civic engagement in America, found that virtually all measures of civic health are lower in more diverse settings. "The extent of the effect is shocking," says Scott Paige, a University of Michigan political scientist.

I don't think this effect is particularly related to race. I bet that if you throw together a community of white, black, Asian, Hispanic, and Martian Mormons, they act as a "non-diverse" community. As we saw before, culture trumps race.

So this sort of cultural unity is exactly the sort of thing we need to improve civic life and prevent racism…and of course, it's exactly what progressives get enraged if we try to produce.

In America, progressivism focuses on pointing out how terrible American culture is and how much other people's cultures are better than ours. If we celebrate Columbus Day, we have to spend the whole time hearing about what a jerk Columbus was (disclaimer: to be fair, Columbus was a *huge* jerk). If we celebrate Washington's birthday, we have to spend the whole time hearing about how awful it was that Washington owned slaves. Goodness help us if someone tries to celebrate Christmas – there are now areas where if a city puts up Christmas decorations, it has to give equal space to atheist groups to put up displays about how Christmas is stupid and people who celebrate it suck. That's…probably not the way to maximize cultural unity, exactly?

We are a culture engaged in the continuing project of subverting itself. Our heroes have been toppled, our rituals mocked, and one gains status by figuring out new and better ways to show how the things that should unite us are actually stupid and oppressive. Even the conservatives who wear American flag lapel pins and stuff spend most of their time talking about how they hate America today and the American government and everything else associated with America except for those stupid flag pins of theirs.

Compare this to olden cultures. If someone in Victorian Britain says “God save the Queen!”, then everyone else repeated “God save the Queen!”, and more important, *they mean it*. “England expects every man to do their duty” is actually perceived as a *compelling reason* why one’s duty should be done.

It would seem that the Victorian British are more on the Mormon side and modern Americans more like the Unitarians. And in fact, the Victorians managed to colonize half the planet while America can’t even get the Afghans to stop shooting each other. While one may not agree with Victorian Britain’s aims, one has to wonder what would happen if that kind of will, energy, and unity of purpose were directed towards a worthier goal (I wonder this about the Mormon Church too).

Reactionaries would go further and explore this idea in a depth I don’t have time for, besides to say that they believe many historical cultures were carefully optimized and time-tested for unifying potential, and that they really sunk deep into the bones of the populace until failing to identify with them would have been unthinkable. The three cultures they most often cite as virtuous examples here are Imperial China, medieval Catholicism, and Victorian Britain; although it would be foolish to try to re-establish one of those exactly in a population not thoroughly steeped in them, we could at least try to make our own culture a little more like they were.

Once again, the Reactionary claim is not necessarily that we have to brainwash people or drag the Jews kicking and screaming to Christmas parties. It's just that maybe we should stop deliberately optimizing society for as little unity and shared culture as humanly possible.

Reach For The Tsars

I have noticed a tendency of mine to reply to arguments with "Well yeah, that would work for the X Czar, but there's no such thing."

For example, take the problems with the scientific community, which my friends in Berkeley often discuss. There's lots of publication bias, statistics are done in a confusing and misleading way out of sheer inertia, and replications often happen very late or not at all. And sometimes someone will say something like "I can't believe people are too dumb to fix Science. All we would have to do is require early registration of studies to avoid publication bias, turn this new and powerful statistical technique into the new standard, and accord higher status to scientists who do replication experiments. It would be really simple and it would vastly increase scientific progress. I must just be smarter than all existing scientists, since I'm able to think of this and they aren't."

And I answer "Well, yeah, that would work for the Science Czar. He could just make a Science Decree that everyone has to use the right statistics, and make another Science Decree that everyone must accord replications higher status. And since we all follow the Science Czar's Science Decrees, it would all work perfectly!"

Why exactly am I being so sarcastic? Because things that work from a czar's-eye view don't work from within the system. No *individual* scientist has an incentive to unilaterally switch to the new statistical technique for her *own* research, since it would make her research less likely to produce earth-shattering results and since it would just confuse all the other scientists. They just have an incentive to want *everybody else* to do it, at which point they would follow along.

Likewise, no journal has the incentive to unilaterally demand early registration, since that just means everyone who forgot to early register their studies would switch to their competitors' journals.

And since the system is *only* made of individual scientists and individual journals, no one is ever going to switch and science will stay exactly as it is.

I use this "czar" terminology a lot. Like when people talk about reforming the education system, I point out that right now students' incentive is to go to the most prestigious college they can get into so employers will hire them, employers' incentive is to get students from the most prestigious college they can so that they can defend their decision to their boss if it goes wrong, and colleges' incentive is to do whatever it takes to get more prestige, as measured in *US News and World Report* rankings. Does this lead to huge waste and poor education? Yes. Could an Education Czar notice this and make some Education Decrees that lead to a vastly more efficient system? Easily! But since there's no Education Czar everybody is just going to follow their own incentives, which have nothing to do with education or efficiency.

There is an extraordinarily useful pattern of refactored agency in which you view humans as basically actors playing roles determined by their incentives. Anyone who strays even slightly from their role is outcompeted and replaced by an understudy who will do better. That means the final state of a system is determined entirely by its initial state and the dance of incentives inside of it.

If a system has perverse incentives, it’s not going to magically fix itself; no one inside the system has an incentive to do that. The end user of the system – the student or consumer – is already part of the incentive flow, so they’re not going to be helpful. The only hope is that the system can get a Czar – an Unincentivized Incentivizer, someone who controls the entire system while standing outside of it.

I alluded to this a lot in my (warning: political piece even longer than this one) Non-Libertarian FAQ. I argued that because systems can’t always self-improve from the inside, every so often you need a government to coordinate things.

Reactionaries would go further and say that a standard liberal democratic government is not an Unincentivized Incentivizer. Government officials are beholden to the electorate and to their campaign donors, and they need to worry about being outcompeted by the other party. They, too, are slaves to their incentives. The obvious solution to corporate welfare is “end corporate welfare”. A three year old could think of it. But anyone who tried would get outcompeted by powerful corporate interests backing the campaigns of their opponents, or outcompeted by other states that still have corporate welfare and use it to send businesses and jobs their way. It’s obvious from outside the system, and completely impossible from the inside. It would appear we need some kind of a Government Czar.

You know who had a Government Czar? Imperial Russia. For short, they just called him “Czar”.

Everyone realizes our current model of government is screwed up and corrupt. We keep electing fresh new Washington Outsiders who promise with bright eyes to unupscrew and decorruptify it. And then they keep being exactly as screwed up and corrupt as the last group, because if you hire a new actor to play the same role, the lines are still going to come out exactly the same. Want reform? The lines to “Act V: An Attempt To Reform The System” are already written and have been delivered dozens of times already. How is changing the actors and actresses going to help?

A Czar could actually get stuff done. Imperial Decree 1: End all corporate welfare. Imperial Decree 2: Close all tax loopholes. Imperial Decree 3: Health care system that doesn’t suck. You get the idea.

Would the Czar be corrupt and greedy and tyrannical? Yes, probably. Let’s say he decided to use our tax money to build himself a mansion ten times bigger than the Palace of Versailles. The Internet suggests that building Versailles today would cost somewhere between $200M and $1B, so let’s dectuple the high range of that estimate and say the Czar built himself a $10 billion dollar palace. And he wants it plated in solid gold, so that’s another $10 billion. Fine. Corporate welfare is $200B per year. If the Czar were to tell us “I am going to take your tax money and spend it on a giant palace ten times the size of Versailles covered in solid gold”, the proper response would be “Great, but what are we going to do with the other $180 billion dollars you're saving us?"

(here I am being facetious. A better answer might be to point out that the British royal family already lives in a giant palace, and they by all accounts earn the country more than they cost)

As for the tyranny, we have Fnargl's shining example to inspire us. But really. Suppose Obama were named Czar. Do we really think he'd start sending Republicans to penal camps in Alaska for disagreeing with him? If Sasha took over as Czarina, do you think *she'd* do that?

Is this the face of someone who would crush you with an iron fist?

In the democratic system, the incentive is always for the country to become more progressive, because progressivism is the appeal to the lowest common denominator. There may be reversals, false starts, and Reagan Revolutions, but over the course of centuries democracy means inevitable creeping progress. As Mencius Moldbug says, "Cthulhu swims slowly, but he always swims left.” A Czar, free from these incentives, would be able to take the best of progressivism and leave the rest behind.

(the Reactionaries I beta-tested this essay with say that the last paragraph deserves much more space, that there are many complicated theories of why this holds true, and that it is a central feature of Reactionary thought. I don’t understand this well enough to write about it yet, but you may want to read Moldbug on…no, on second thought, just let it pass.)

So who gets to be Czar? Probably the most important factor is a Schelling point: it should be someone everyone agrees has the unquestioned right to rule. Obama is not a *bad* choice, but one worries he may be a little too progressive to treat the job with the seriousness it deserves. We could import the British monarchy, but really ever since the Glorious Revolution they’ve been a bit too constitutional for our purposes. If we wanted a genuine, legitimate British monarch of the old royal line, someone with authority flowing through his very veins, our best choice is, indeed to exhume the body of King James II (ruled 1685 – 1688), clone him, and place the clone on the throne of the new United States Of The Western World.

Really, it’s just common sense.

A Brief Survey Of Not Directly Political Reactionary Philosophy

We have reached the goal we set for ourselves. Is this a comprehensive understanding of Reactionary thought?

No. This focuses on political philosophy, but Reaction is a complete philosophical movement with many other branches.

For example, Reactionary moral theories tend to focus on the dichotomy between Virtue and Decadence. Extensional definitions might do best here: consider the difference in outlook between Seneca the Stoic and the Roman Emperor Nero, or between Liu Bei and Cao Cao, or between Thomas More and Henry VIII. In each of these cases, a virtuous figure recognized the decadence of his society and willfully refused to succumb to it. Of course, an even more virtuous example would be someone like Lycurgus, who realized the decadence of his society and so *went out and fixed society*.

Reactionary aesthetic theories tend to be, well, reactions against progressive aesthetic theories. To Reactionaries, the epitome of the progressive aesthetic theory against which they rebel is the fairy tale of the Ugly Duckling, where one duckling is uglier than the rest, everyone mocks him, but then he turns out to be the most beautiful of all. The moral of the story is that ugly things are really the most beautiful, beautiful things are for bullies who just want to oppress the less beautiful things, and if you don’t realize this, you’re dumb and have no taste.

Therefore, decent, *sophisticated* people must scoff at anything outwardly beautiful and say that it’s probably oppressive in some way, while gushing over anything apparently ugly. Cathedrals are “gaudy” or “tacky”, but Brutalist concrete blocks are “revolutionary” and “groundbreaking”. An especially conventionally attractive woman is probably just “self-objectifying” and “pandering”, but someone with ten tattoos and a shaved head is “truly confident in her femininity”. Art of the sort people have been proven to like most is old-fashioned and conformist; *real* art is urinals that artistically convey an anti-art message, or paintings so baffling that no one can tell if they are accidentally hung upside-down.

The Reactionary aesthetic, then, is something so simple that if it weren’t specifically a reaction to something that already exists, it would sound stupid: no, beautiful things are legitimately beautiful, ugly things are legitimately ugly, any attempt to disguise this raises suspicions of ulterior motives.

Reactionaries also seem to be really into metaphysics, especially of the scholastic variety, but I have yet to be able to understand this. Blatant racism, attempts to clone long-dead monarchs, and giving a gold-obsessed alien absolute power all seem like they could sort of make sense in the right light, but why anyone would want more metaphysics is honestly completely beyond me.

But Seriously, What Do We Do About This Hole? And How Fast Should We Be Digging, Anyway?

We started with an argument that modern culture probably doesn’t give us a very impartial view on the relative merits of modern culture, and so we should investigate this more thorougly.

We noted that on many of the criteria we care about, the present is better only because of its improved technology. We further noted that on other criteria, even *despite* our better technology, past societies seemed to outperform us

Nevertheless, we identified some areas where the present really did seem better than the past. The present was less racist, less sexist, less colonialist, more humane, and less jingoistic.

We then went through each of those things and showed why they might not be as purely beneficial are as generally believed. We found evidence that societies many would call “racist” give minorities better measurable outcomes; that societies many would call “sexist” give women higher self-reported life satisfaction; that colonialism led to peace and economic growth that decolonialism was unable to match; and that supposedly more “humane” policies end up torturing their victims far more than just getting something superficially cruel over wit would; and even that cultural unity, which some might call “jingoism”, has been empirically shown to be an important factor in building communities and inspiring prosocial sentiment.

Therefore, we found that all the points we had previously noted as advantages of present over past societies were, when examined more closely, in fact points in the past societies’ favor.

Next, we looked at how we might replicate these advantages of past societies in a world which seems to be moving inexorably further toward so-called progressive ideals. We independently came up with the same solution that these past societies used: the idea of a monarch, either constitutional or (preferably) absolutist. We found that many of the problems we would expect such a monarch to produce are exaggerated or unlikely.

Finally, we identified this ideal monarch as a clone of James II of the United Kingdom.

We also went into a survey of a couple of other Reactionary ideas. Other such ideas I have *not* included simply because I was totally unable to understand or sympathize with them and so couldn’t give them fair treatment include: an obsession with chastity, highly positive feelings about Catholicism that never go as far as actually going to church or believing any Catholic doctrine in a non-ironic way, neo-formalism, and what the heck the Whigs have to do with anything.

Nevertheless, I hope that this has been a not-entirely futile exercise in trying to Ideological Turing Test an opposing belief. I think Reactionaries are correct that some liberal ideas have managed to make their way into an echo chamber that makes them hard to examine. And even though the Reactionaries themselves are way too rightist, I think it’s good to have their ideas out there in the Hegelian sense of “and then the unexamined-conservativism touched the unexamined-liberalism and in a puff of smoke they merged to magically become the perfect political system!”

- * - * - * -

Once again, expect my counterargument to this sometime in the next while. I would be interested in hearing other people’s counterarguments in the meantime and am very likely to steal them. I am also likely to ignore some of them if they make arguments I already agree with and so feel no need to debate, but I would still enjoy reading them. Basically I welcome comments and discussion from all sides.

With one exception. Yes, I have included the racist parts of Reactionary philosophy above. Yes, those points need to be debated, and some of that debate may be in favor. But any comment that moves away from the sort of dry scientific racism used to prove or disprove political theorems, and toward the sort where they’re just shouting ethnic slurs and attacking racial groups to make their members feel bad, *will* be deleted and the person involved probably IP-banned. I also reserve the right to edit comments that don’t quite reach that point but are noticeably in need of rephrasing.

A Thrive/Survive Theory Of The Political Spectrum

I admitted in my last post on Reaction that I devoted insufficient space to the question of why society does seem to be drifting gradually leftward. And I now realize that in order to critique the Reactionary worldview effectively we're going to have to go there.

The easiest answer would be "because we retroactively define leftism as the direction that society went". But this is not true. Communism is very leftist, but society eventually decided not to go that way. It seems fair to say that there are certain areas where society did *not* go to the left, like in the growth of free trade and the gradual lowering of tax rates, but upon realizing this we don't feel the slightest urge to redefine "low tax rates" as leftist.

So what *is* leftism? For that matter, what is rightism?

Any theory of these two ideas would have to explain at least the following data points:

1) Why do both ideologies combine seemingly unrelated political ideas? For example, why do people who want laissez-faire free trade empirically also prefer a strong military and oppose gay marriage? Why do people who want to help the environment also support feminism and dislike school vouchers?

2) Why do the two ideologies seem *broadly* stable across different times and cultures, such that it's relatively easy to point out the Tories as further right than the Whigs, or ancient Athens as further left than ancient Sparta? For that matter, why do they seem to correspond to certain neural patterns in the brain, such that neurologists can determine your political beliefs with 83% accuracy by examining brain structure alone?

3) Why do these basically political ideas correlate so well with moral, aesthetic, and religious preferences?

4) The original question: how come, given enough time and left to itself, leftism seems to usually win out over rightism, pushing the Overton window a bit forward until there's a new leftism and rightism?

I have a hypothesis that explains *most* of this, but first let me go through some proposed alternatives.

The Reactionaries have at least two theories. Moldbug suggests that rightism is common sense, and leftism is Christianity minus the religious trappings and rightism is rational thought. Another of his posts suggests that leftism is naked power-grabbing and rightism is virtuous pro-social behavior.

But the first of these fails to explain point 1; how come most traditionally Christian ideas end up on the right side of the aisle? It fails to explain 2 – how come we can call Sparta rightist even in the pre-Christian age? It might explain 3. But it definitely fails point 4; even if it were true, why would this weird neo-Christian sect suddenly take off just as all other Christian sects are hemorrhaging believers? As for the second, it explains point 4 and point 4 only, and seems, well, maybe a little completely obviously self-serving?

The Libertarians say that leftism supports government intervention on economic but not social issues, and rightism supports government intervention on social but not economic issues. Unfortunately, this isn't really true. Leftists support government intervention in society in the form of gun control, hate speech laws, funding for the arts, and sex ed in schools. In fact, leftists are sometimes even accused of being in favor of “social engineering”. Meanwhile, conservatives lead things like the home schooling and school choice movements, which seem to be about *less* government regulation of society. Having gotten Point 1 not quite right, this theory then goes on to completely ignore points 2, 3, and 4.

The scientists studying neuropolitics in that article I linked to say things like “Liberals tend to seek out novelty and uncertainty, while conservatives exhibit strong changes in attitude to threatening situations. The former are more willing to accept risk, while the latter tends to have more intense physical reactions to threatening stimuli.” But this seems flawed. Leftists have an intense physical reaction to the threatening situation of global warming. Rightists seek out the novelty and accept the risk of a foreign war that *might* increase America’s global power at minimal cost but *might* waste hundreds of thousands of lives to no end. Another failure of 1, I’ll give it 2 or 3, and once again no love for point 4.

Okay, I’ll put you out of your misery and tell you my hypothesis now. My hypothesis is that rightism is what happens when you’re optimizing for surviving an unsafe environment, leftism is what happens when you’re optimized for thriving in a safe environment.

The Dead Have Risen, And They’re Voting Republican

Before I explain, a story. Last night at a dinner party we discussed Dungeons and Dragons orientations. One guest declared that he thought Lawful Good was a contradiction in terms, very nearly at the same moment as a second guest declared that he thought Chaotic Good was a contradiction in terms. What’s up?

I think the first guest was expressing a basically leftist world view. It is a fact of nature that society will always be orderly, the economy always expanding. Crime will be a vague rumor but generally under control. All that the marginal unit of extra law enforcement adds to this pleasant state is cops beating up random black people, or throwing a teenager in jail because she wanted to try marijuana.

The second guest was expressing a basically rightist world view. The prosperous, orderly society we know and love is hanging *by a frickin’ thread*. At any moment, terrorists or criminals or just poor management could destroy everything. It is *really really good* that we have police in order to be the “thin blue line” between civilization and chaos, and we might sleep easier in our beds at night if that blue line were a little thicker and we had a little more buffer room.

I propose that the best way for leftists to get themselves in a rightist frame of mind is to imagine there is a zombie apocalypse tomorrow. It is a very big zombie apocalypse and it doesn’t look like it’s going to be one of those ones where a plucky band just has to keep themselves alive until the cavalry ride in and restore order. This is going to be one of your *long-term* zombie apocalypses. What are you going to want?

First and most important, guns. Lots and lots of guns.

Second, you’re going to have a deep and abiding affection for the military and the police. You’re going to hope that the government has given them a *lot* of funding over the past few years.

Third, you’re going to start praying. Really hard. If someone looks like they’re doing something that might offend God, you’re going to very vehemently ask them to stop. However few or many atheists there may be in foxholes, there are probably fewer when those foxholes are surrounded by zombies. Or, as Karl Marx famously said of zombie uprisings, “Who cares if it’s an opiate? / It’s time to pray!”

Fourth, you’re going to be extremely suspicious of outsiders. It’s not just that they could be infected. There are probably going to be all sorts of desperate people around, looking to steal your supplies, your guns, your ammo. You trust your friends, you trust your neighbors, and if someone who looks different than you and seems a bit shifty comes up to you, you turn them away or just kill them before they kill you.

Fifth, you’re going to want *hierarchy* and *conformity*. When the leader says run, everyone runs. If someone is constantly slowing the group down, questioning the group, causing trouble, causing dissent, they’re a troublemaker and they can either shut up or take their chances on their own. There’s a reason all modern militaries work on a hierarchical system that tries to maximize group coherence.

Sixth, you are not going to be sentimental. If someone gets bitten by the zombies, they get shot. Doesn’t matter if it’s really sad, doesn’t matter if it wasn’t their own fault. If someone breaks the rules and steals supplies for themselves, they get punished. If someone refuses to pull their weight, they get left behind. Harsh? Yes. But there’s no room for people who don’t contribute in a sleek urban postapocalyptic zombie-fighting machine.

Seventh, you want to *maximize wealth*. Whatever gets you the supplies you need, you’re going to do. If that means forcing people to work jobs they don’t like, that’s the sacrifice they’ve got to make. If your raid on a grocery store leaves less behind for everyone else, well, that’s too bad but you need the food. Are woodland animals going to go extinct as more and more survivors retreat to the woods and rely on them for food? That’s not the kind of thing you’re worried about when you’re half-starved and only a few hours ahead of the zombie horde.

Eighth, strong purity/contamination ethics. We know that purity/contamination ethics are an evolutionary defense against sickness: disgusting things like urine, feces, dirt, blood, insects, and rotting corpses are all vectors of infection; creepy animals like spiders, snakes, and centipedes are all vectors for poisoning. Maybe right now you don’t worry too much about this. But in a world where the hospitals are all overrun by zombies and you need to outrun a ravenous horde at a moment’s notice, this becomes a much bigger deal. Not to mention that anything you catch might be the dreaded Zombie Virus.

Ninth, an emphasis on practical skills rather than book learning. That eggheaded Professor of Critical Studies? Can’t use a gun, isn’t studying a subject you can use to invent bigger guns, not a useful ally. Probably would just get in the way. Big masculine men who can build shelters and fight with weapons are useful. So are fertile women who can help breed the next generation of humans. Anyone else is just another mouth to feed.

Tenth, extreme black and white thinking. It’s not *useful* to wonder whether or not the zombies are only fulfilling a biological drive and suffer terribly when you kill them despite not being morally in the wrong. It’s *useful* to believe they’re the hellish undead and it’s your sacred duty to fight them by any means necessary.

In other words, “take actions that would be beneficial to survival in case of a zombie apocalypse” seems to get us rightist positions on a lot of issues. We can generalize from zombie apocalypses to any desperate conditions in which you’re not sure that you’re going to make it and need to succeed at any cost.

What about the opposite? Let’s imagine a future utopia of infinite technology. Robotic factories produce far more wealth than anyone could possibly need. The laws of Nature have been altered to make crime and violence physically impossible (although this technology occasionally suffers glitches). Infinitely loving nurture-bots take over any portions of child-rearing that the parents find boring. And all traumatic events can be wiped from people’s minds, restoring them to a state of bliss. Even death itself has disappeared. What policies are useful for this happy state?

First of all, we probably shouldn’t have a police force. Given that crime is impossible, at best they would be useless and at worst they might go around flexing their authority and causing trouble.

Second, religion seems kind of superfluous. Throughout history, richer civilizations have been less religious and our post-scarcity society should be no exception. What would you pray for? What fear is there for faith to allay? With vast supercomputers that know all things, what lingering questions are there for the Bible to answer?

Third, assuming people still have jobs or something, we should probably make them as nice as possible. It doesn’t matter if it hurts productivity; we’re producing far more than we need anyway. We should enforce short work hours and ample maternity and paternity leave so that everyone has time to concentrate on the more important things in life.

Fourth, interest in the environment. We have no shortage of material goods; if our lives lack anything it is beauty and connection to nature. So it will be nice to have as many pleasant green spaces as possible; and if this means a little less oil, it’s not like our Oil-Making-Machines can’t make up the extra.

Fifth, free love. There’s no worries about STDs, the family unit isn’t necessary for any kind of economic survival, and the nurture-bots and trauma-erasure-centers can take care of the kids of anything goes wrong. And since we don’t have anything else to do, we might as well enjoy ourselves with infinite sex.

I was going to go for ten here too, but you get the picture. This world of infinite abundance is a great match for leftist values. I imagine even a lot of rightists and Reactionaries would be happy enough with leftism in a situation like this.

I should also mention what would no doubt be the main pastime of the people of this latter world: signaling.

When people are no longer constrained by reality, they spend most of their energy in signaling games. This is why rich people build ever-bigger yachts and fret over the parties they throw and who got invited where. It’s why heirs and heiresses so often become patrons of the art, or donors to major charities. Once you’ve got enough money, the next thing you need is status, and signaling is the way to get it.

So the people of this final utopia will be obsessed with looking good. They will become moralists, and try to prove themselves more virtuous than their neighbors. Their sophistication will gradually increase as each tries to establish themselves as a critic, as tasteful, as a member of an aristocracy that can no longer be defined in terms of money. They will become conniving, figuring out ways to raise their own social status at their neighbors' expense. Or they will devolve into a host of competing subcultures, united only by their pride in their defiance of a "norm" which is quickly ceasing to exist.

Chris wrote this comment to my last post's section on Reactionary aesthetics:

> The things Reactionaries complain about in aesthetics seem not the fault of progressives, but the result of an unavoidable signaling logic. See Quentin Bell on what he called "conspicuous outrage."

I agree with Chris 100% here, but I don't think this is opposed to the Reactionaries' link between this aesthetic and leftism. I think that leftists are the sort of people who are so secure that they can start thinking about how to excel at signaling games.

An Evaluation of the Thrive/Survive Theory

This is close to an explanation of our Point 1. It does not quite explain all left vs. right positions (in particular I despair of *any* theory that will tell me why school choice is a rightist rather than a leftist issue) but it does as well as any of the others, and better than some.

This also satisfies Point 2. The distinction between security and insecurity is far older than Classical Greece; it is perfectly reasonable for Athenian society to start from the assumption of the one and the Spartans to go with the other.

I admit some confusions. For example, it seems weird that poor people, the people who are *actually* desperate and insecure, are often leftist, whereas rich people, the ones who are *actually* completely safe, are often rightist. I would have to appeal to economic self-interest here: the poor are leftist because leftism is the philosophy that says to throw lots of resources at helping the poor, and the rich are rightist because rightism says to let the rich keep getting richer. Despite voting records, I expect the poor to share more rightist social values (eg be more religious, more racist) and the rich to to share more leftist social values (more intellectual as opposed to practical, less obsessed with guns). For a more comprehensive theory of economic self-interest and politics, see my essay on the subject.

This theory also satisfies Point 3. Developmental psychology has gradually been moving towards a paradigm where our biology actively seeks out information about our environment and then toggles between different modes based on what it finds. Probably the most talked-about example of this paradigm is the thrifty phenotype idea, devised to explain the observation that children starved in the womb will grow up to become obese. The idea is that some system notices that there seems to be very little food, and goes into “desperately conserve food” mode, which when food becomes more plentiful leads to obesity.

Another example, more clearly neurological, is the tendency of children who grow up in broken homes to have poor life outcomes. Although this was originally just interpreted as “damage”, an equally valid theory is that the brain seeks out information on what kind of society it lives in – one based on love and trust, or one based on violence and mistrust – and then activates the appropriate coping strategy. If child abuse or something makes the brain conclude we live in a violence and mistrust society, it alters its neural architecture to be violent and mistrustful – and hence dooms itself to future bad outcomes.

It seems broadly plausible that there could be one of these switches for something like “social stability”. If the brain finds itself in a stable environment where everything is abundant, it sort of lowers the mental threat level and concludes that everything will always be okay and its job is to enjoy itself and win signaling games. If it finds itself in an environment of scarcity, it will raise the mental threat level and set its job to “survive at any cost”.

What would toggle this switch? My guess is that genetics plays a very large role in setting the threshold (explaining why party affiliation is highly heritable) and that a lot of the remainder is implicit messages we get in childhood from our parents, school, church, et cetera. Actual rational argument and post-childhood life experiences make up the last few percent of variation.

Knowing this, the answer to Point 4 is blindingly obvious. Leftism wins over time because technology advances over time which means societies become more secure and abundant over time.

As a decent natural experiment, take the Fall of Rome. Both Greece and Rome were *relatively* leftist, with freedom of religion, democratic-republican governments, weak gender norms, minimal family values, and a high emphasis on education and abstract ideas. After the Fall of Rome, when Europe was set back technologically into a Dark Age, rightism returned with a vengeance. People became incredibly religious, militant, pragmatic, and provincial, and the government switched to an ad hoc and extremely hierarchical feudalism. This era of conservativism ended only when society reached the same level of technology and organization as the Greeks and Romans. So it’s not that cultures become more leftist over *time*, it’s that leftism varies with social and economic security.

Both rightists and leftists will find much to like in this idea. The rightists will ask: “So you mean that rightism is optimized for survival and effectiveness, and leftism is optimized for hedonism and signaling games?” And I will mostly endorse this conclusion.

On the other hand, the leftists will ask: “So you mean rightism is optimized for tiny unstable bands facing a hostile wilderness, and leftism is optimized for secure, technologically advanced societies like the ones we are actually in?” And this conclusion, too, I will mostly endorse.

Given that we are in conditions that seem to favor leftist ideals, the modern debate between leftists and rightists is, to mix metaphors atrociously, about how hard we can milk the goose that lays the golden eggs. Leftists think we can just keep drawing more and more happiness and utility for all out of our massive scientific and technological progress. Rightists are holding their breath for something to go terribly, terribly wrong and require the crisis-values they have safeguarded all this time – which is why posts like this one seem to be the purest expression of rightist wish-fulfillment fantasy.

I will only remark that one of the most consistent findings of my researches through economic and political history has been the remarkable, almost supernatural resilience of our particular aureate waterfowl. To a leftist, this is good news. To a rightist, I suppose this would just be evidence of how shockingly audacious we must be to try to push our luck even *further*.

EDIT: People are taking this as pro-Reactionary. I meant it to be at least *suggestive* of anti-Reactionary ideas. See my reply to the first comment below.

Yet Another Parable On The Importance Of Controlling Your Experiments

This morning, I decide to finally try a certain productivity hack beginning with “M”, you know the one. As a test to see if it helped me increase my focus, I pick up a “Teach Yourself Calligraphy” book I had lying around but hadn’t looked at yet.

Seven hours of unbroken calligraphy-studying later, I finally start wondering if maybe I should do something else. I think *maybe* I used the bathroom once or twice during that time, it’s hard to tell? Anyway, I decide to count the experiment as a success.

Then I have a perfectly normal evening. After the rest of the house has gone to sleep I sit down to write a blog post or finish up some *Dungeons and Discourse* material or something. It’s tough going, I’m feeling a bit tired, and while looking for a distraction I notice the calligraphy book sitting on my desk and decided to tie up a lesson I didn’t finish.

Five hours of unbroken calligraphy studying later, I finally start wondering whether it might be several hours past my bedtime and whether that is a sufficiently important reason to interrupt a lesson on uncials.

So it turns out calligraphy is just really, really interesting. And I still don’t have the slightest idea whether or not M___ works.

(this is also why no more substantive blog post today)

We Wrestle Not With Flesh And Blood, But Against Powers And Principalities

> Mimes, in the form of God on high mutter and mumble low
> And hither and thither fly – mere puppets, who come and go
> At bidding of vast formless things shifting scenery to and fro
>
> **– an excerpt out of *Ligaea*, by Edgar Allen Poe**

There should be a post debating Reactionaries' assumptions about the superiority of past cultures and methods. Eventually I hope to write that post. But this is not it. This is the post where I claim that, *even granting all of those assumptions*, Reaction is somewhere between wrong and impossible. Why?

To borrow Poe's terminology, history as we learn it in school tends to concentrate on the puppets and ignore the vast formless things.

In a previous essay, I mentioned a pattern of refactored agency in which human beings lack agency and merely respond to incentives. I said they were "actors" reading from the "script their incentives wrote for them", and anyone who deviated from the part would be outcompeted and replaced.

This seems to broadly describe most historical figures. If Christopher Columbus had decided not to explore America, Cabral or Cabot or someone would have. Caravels existed, people needed a new trade route to India, the only question was who was going to be first.

But the puppetry expands past individuals toward whole empires and movements. If God reached into the year 1900 and removed every single Communist, and every Communist book, and erased all memory of Communism, I think it would take about five minutes before someone reinvented something much like the movement, because there were a bunch of very poor people who felt desperate and cheated crammed up against a bunch of very rich people who weren’t afraid to flaunt their wealth. The new movement might have differed from Communism in minor details – maybe their color would have been blue instead of red – but it wouldn’t be hard to identify.

So much for the puppets. What are these Vast Formless Things giving them their orders? I mentioned liking *Guns, Germs and Steel*, and I think Diamond has done a good job of proving geography has important historical effects. But geography is fixed, not exactly the sort of thing that’s going to cause revolutions. So after that last post you probably won’t be surprised to hear I think the vastest and most formless Vast Formless Thing of all is technological progress.

Engines Alone Turn The Wheels of History

The largest and furthest-reaching political changes of all time have invariably been the effect of technological progress. The largest of these, the transition from egalitarian bands to the ultrahierarchical divine monarchies of the Bronze Age, seems to have been mostly the effect of the Agricultural Revolution and its corollaries. Without committing to what order these things happened in:

– Need for a guarantee that the crops you planted will still be yours at harvest time inspires idea of private property
– Sedentary lifestyle + concept of property allow accumulation of wealth
– Accumulation of wealth requires law enforcement to protect wealth
– Excess food allows specialization of labor
– Requirement for law enforcement + specialized labor leads to creation of warrior caste
– Powerful warrior caste + everyone else being farmers and losing the martial skills they enjoyed as hunters leads to warrior caste taking over.
– Need for large irrigation/flood control projects in many areas leads to very centralized government

And a lot of these late Neolithic/early Bronze Age cultures turned out the same way. If Ramesses II, Montezuma II, and Agamemmnon went to lunch together, they'd have a lot to talk about, despite being separated by continents and millennia. This suggests that the Generic Bronze Age Government – a god-king served by a bunch of warrior-nobles, plus massive militarism and slavery – probably just made sense given the circumstances.

I don't want to sound too deterministic and spooky here, but I do think governments have a good way of kind of converging to a local optimum. Ramesses II may not have thought "You know, the Nile floods a lot, so I should institute a strong centralized government with lots of slavery", but some people tried some things, other people tried other things, the things that worked won out, the things that didn't passed into the dustbin of history, and we got Ancient Egypt. If God reached into history and tried to turn Ancient Egypt into modern day Sweden, it wouldn't work any better than His attempt to remove Communism did a few paragraphs ago – within a few years they’d be back to worshipping Pharaohs and invading Canaanites

After the Neolithic, one of the most clear-cut examples of technology changing social structure was the fall of feudalism. Feudalism was based on a very simple calculation: one armored knight could defeat an arbitrary number of untrained peasants. To be an armored knight took your standard 10,000 hours of training; it wasn’t something you can do as a side job. So once again you have at least two castes – the warrior caste and the support-the-warriors caste; since the warrior caste is both smaller and stronger, you end up with an aristocratic system. If you want to govern large territories under an aristocratic system and you don’t have real-time communication, you come up with something like feudalism. And sure enough, we have pretty much the exact same social structure in medieval Europe and Sengoku Japan.

Then some new weapons were invented: pikes, longbows, crossbows, but especially firearms. Now you can get someone who *hasn’t* trained 10,000 hours, give them a few days of weapon training, hand them a gun or a crossbow or something, and they can kill an armored knight. Now the power doesn’t belong to the people with the best connections among the warrior nobility, it belongs to the people with enough money to hire soldiers and supply them with guns. It took a long time to realize this, especially since guns weren’t that good to begin with, but when people finally got it into their heads feudalism went caput.

The printing press was an even bigger deal. I don’t have my Big List O’ Unbelievable Printing Press Statistics handy here, but the Internet reminds me that there were 30,000 books – total! – in Europe before the invention of the printing press. Fifty years later, 300,000 copies *just of Martin Luther’s* *religious tracts* were printed *in a single year alone*. Among just the simpler and more direct effects:

– Protestant Reformation. Easy one. Lots of people had tried challenging the Catholic Church before, but not only could they not get their message out, but most people weren't ready for it – only the richest of the rich could even own their own Bible. Basically *as soon as* the printing press was invented this took off.
– Newspapers. All of a sudden, people who aren't the highest ranks of the nobility know what's going on at court. Some people have opinions on this. Start of modern politics where the masses know what's going on and might complain.
– The Renaissance. All these old Greek and Roman texts are spread. People realize that there are other ways to organize society beyond their own.
– Scientific Revolution. If a scientist discovers something, he can actually sent his work to other scientists in an efficient way, who can then build upon it. This was absolutely not the case for previous scientists, which is why not much happened during those periods.
– Rise of nationalism. Ability of common people to read books means more books printed in vernacular instead of Latin. This causes insular language-based communities which then feed upon themselves to become more delineated nation-states.

I was going to go into the same depth about the Industrial Revolution and the Sexual Revolution (by which I mean near-simultaneous discovery of birth control pills and antibiotics effective against syphilis), but this section is getting long, so if you promise to just agree they Changed Everything I'll make life easier for both of us and move on.

Forget King James II, Try King Canute

So the biggest changes in history have been predetermined reactions to different technological conditions. This should worry Reactionaries for several reasons.

First, I previously claimed that if Communism disappeared it would be immediately reinvented. If Ancient Egypt had randomly switched to modern Sweden, the realities of life in the Nile flood plain and of Bronze Age technology would have caused it to switch back without even breaking its stride.

I think my claim here is that cultures and ideologies have a sort of homeostatic regulatory mechanism that fits them to their conditions. This is why all Bronze Age cultures converged upon divine monarchies, and all medieval empires converged upon feudalism, and proooobably why all modern cultures converge upon liberal democracy.

Countries that avoid liberal democracy usually regret it. China would be a good example. They tried being really Communist for a while and ended up becoming an economic basketcase. If they wanted to compete on the international stage they realized they needed a stronger economy, and so liberalized their market. A competitive market requires information access, so the Chinese got access to lots of foreign media; I recently learned that any business that wants to pay for it can even legally avoid the Great Firewall. The Internet meant the Chinese could coordinate protests on microblogging platforms, leading to a bunch of riots, leading to an attempt to liberalize the system and crack down on corruption which is still going on. I’m not going to claim that China is definitely going to end up as a democracy, but I think whatever it does end up as is going to be a whole lot more like 2013 USA than like 1963 China.

China didn’t *plan* to approach the Western model of government. It was just what happened to them automatically when they wanted their country to stop being a hellhole. The same is happening now in Burma, somewhat more slowly in Cuba, and in other places around the world. Even the countries skipping the “democracy” part have been aping the Industrial Revolution, womens’ rights, and so on.

This is probably because many features of liberal democracy are adaptations to our current technological climate. For example, women’s lib seems like an adaptation both to the Sexual Revolution and to the demographic transition where people are no longer having like twenty children all the time. Representative government seems like an adaptation to mass media that allows everyone to be aware of, and usually upset about, what the country’s leadership is doing.

If you like these things, you can call it cultural evolution and assume we’re approaching some great goal of perfection. If you don’t like them, you can call them patches, such that once the demographic transition screws up traditional gender roles, we need women’s lib as a patch to contain the damage. Either way, you better not take off that patch.

So this is my first beef with Reactionaries. They see someone identifying as Progressive saying something – Gloria Steinem pushing for women’s rights or something – and they say “Oh no, that awful Progressive Gloria Steinem is screwing up our traditional gender roles. If only she would be quiet, everything would go back to normal!”

Gloria Steinem is a puppet. If she’s part of some movement, even a large saecular movement calling itself Progressivism, they, too, are puppets. It is stupid to get upset at puppets. If you rip them up, the puppeteer will get new ones.

If you don’t like women’s lib, your enemy isn’t Gloria Steinem. Your enemy is the Vast Formless Thing controlling Gloria Steinem. In this case, that would be the demographic transition.

You might be able to beat Gloria Steinem in a fight, but you can’t beat the demographic transition. Or if you can, it’s going to be through something a lot more complicated than going on a soapbox and condemning it, more complicated even than becoming Czar and trying to pass laws to reverse it.

King Canute tried to order back the tide. It was a dumb idea, but in his defense, it was basically just a religious spectacle so he could wax poetic about the power of God. What’s *your* excuse?

Amid These Dark Satanic Mills

In the comments to the Enormous Planet-Sized Nutshell post, some people did a good – though not unassailable – job of picking apart some common Reactionary arguments for superior outcomes among past cultures. The crime data may be an artifact, and more believable homicide data suggests the modern era is safer. Modern students may learn different things than are tested on that Harvard exam which are equally valuable.

Whatever. Let’s assume the Reactionaries are totally right. Past was a thousand times better than the present in every way. So what?

The past contained things like “everyone living in close-knit mono-ethnic villages”. We could, perhaps, with great effort and not a little atrocity, restore the “mono-ethnic”. But the close-knit? The villages? Unless we’re going to roll back the

Industrial Revolution, the *main* ingredient of that particular transition, the move to urbanization, is there to stay.

Any statistic in which the present differs from the past is much more likely to be a result of technology than of politics. Reactionaries correctly use this to excuse themselves of advantages like the present's better health care or greater wealth.

But they have to acknowledge that the same manuever relieves the other side of a lot of their burdens as well. Progressives also have some uncomfortable statistics, usually those relating to social cohesion and trust and happiness. And *I am totally willing to throw every one of these out*. Of course the move to an urban society is going to do that! Of course having people work factory or office jobs instead of either on the land or in an skilled trade like blacksmithing is going to alienate them. Of course having the average person watch TV four hours a day because it's a novel superstimulus is going to affect community ties!

I suspect that the most valuable features of past societies – the ones that we read fantasy books to recapture, the ones that make Renaissance Faires and Medieval Times so attractive – have nothing to do with politics and cannot be restored through politics. In order to regain them, you're going to have to roll back the Industrial Revolution. Needless to say, that makes fighting against the demographic transition look easy.

Perfectly Prepared For A Situation That No Longer Exists

The third and last and most important point I want to bring up involves well-adaptedness.

I often hear Reactionaries make an argument like: the old ways are the result of thousands of years of trial-and-error. Those thousands of years created a remarkably stable culture that survived for centuries. When Progressives throw them out, they are abandoning something we know works for some sort of grand experiment that might end in complete failure.

And I wonder: have these people ever updated a computer program before?

I mean, take Windows 3.11. We know all about Windows 3.11. People had a long time to test it, discover its bugs, find its security holes. Windows 8, on the other hand, is totally new. Goodness only knows what sort of unpleasant surprises are lurking there.

But imagine I decided to uninstall Windows 8 from my computer and replace it with Windows 3.11. Most of my programs aren’t written for Windows 3.11 and they wouldn’t work. Windows 3.11 probably has no idea what to do with Wi-Fi. It probably can’t handle the dual cores of my laptop. Most likely it would ask me to insert floppy disks during the installation and my computer doesn’t have a floppy disk drive.

Even if Windows 3.11, with 1992 programs, on a 1992 machine, is more stable than Windows 8, with 2013 programs, on a 2013 machine – even so, Windows 3.11 with 2013 programs on a 2013 machine would be a total disaster.

I tend to agree with Reactionaries that cultures have a mechanism that gradually adapts them to their conditions. This may not be morally good – if the conditions are “cotton is very lucrative” then the “evolutionarily advantageous” adaptation for a society may be to institute slavery – but they are at least effective and stable.

But a 1600s culture with 2013 technology would be like Windows 3.11 on a 2013 computer: a complete mismatch and a complete disaster. No matter how well Bourbon France was adapted to the 1600s, it would have *no idea* what to do with 2013. If it tried, it would probably end up converging towards the same 2013-technology equilibrium – liberal democracy – as everyone else in 2013. Maybe Louis XIV could stick around as a figurehead or something.

The Reactionaries are correct that we live in a scary time, a time when changes in technology are way outpacing our ability to have any idea how to cope as a society. Maybe if you froze technology at 2013 levels for a hundred years, we would get a pretty good idea what to do with it and would build a culture as well-adapted to our technology level as the Bourbon French were to theirs.

But, uh, getting rid of our culture and replacing it with Bourbon France doesn't shortcut that process. We have a four hundred year head start over Bourbon France in adapting to *our* conditions. If we suddenly became Louis XIV, we'd just be even further behind the adaptation curve, having to reach liberal democracy first before we could get to wherever we're going.

I don't think Bourbon France was more successful, as a society, than our society is. But if you convinced me otherwise, it wouldn't make a shred of difference. Bourbon France + modern tech levels is a society that has never existed and which, I suspect, would be about as successful as Windows 3.11 trying to run Minecraft.

But Seriously, Why Did This Gaping Crack In The Earth Just Open Up? And Why Are You Yelling At The Kid With A Plastic Shovel Next To It?

Our goal was to show that, even granting Reactionaries all their assumptions about the superiority of past civilizations, trying to restore them is impossible.

We noted that the driving force of large-scale historical change was technological progress. That societies underwent cultural evolution into forms that were most adapted to the technological conditions of their age. That this evolution was convergent, and even unconnected civilizations like Ramesses' Egypt and Montezuma's Aztecs could come to resemble each other when they faced similar material problems.

Then we noted that what looks like political progress from the outside is just humans reacting to the shifting landscape of incentives. Although feminism appears as a movement spearheaded by particular feminists who got it into their head thats feminism was a good idea and so decided to push it, a causally useful etiology of feminism would trace the technological conditions that predestined it to arise and succeed.

We accused Reactionaries of condemning or excusing such movements as if they were contingent human creations, and of acting like pushing a few humans or institutions out of the way here or there would change them. Instead, we concluded that they were vast tides in the affairs of (wo)men, and that any attempt to order them around was hubris worthy of King Canute.

Then we accused Reactionaries of a bit of a double-standard, excusing traditional societies' lesser wealth and health by placing the blame on technological progress, but being unwilling to let Progressives do the same in areas where technological progress has inevitably made us worse off, such as the production of feelings of social alienation.

Finally, we accused Reactionaries of arguing that past societies were well-adapted, without specifying well-adapted to *what*. We hypothesized that if forced to finish this statement, it would end up with "well-adapted to the technologies and conditions of the centuries they flourished". The very fact that they stopped flourishing and were replaced by our society suggest they are less well-adapted to conditions today. Or, as G.K. Chesterton puts it in a different context:

> There is one broad fact about the relations of Christianity and Paganism which is so simple that many will smile at it, but which is so important that all moderns forget it. The primary fact about Christianity and Paganism is that one came after the other. Mr. Lowes Dickinson speaks of them as if they were parallel ideals–even speaks as if Paganism were the newer of the two, and the more fitted for a new age. He suggests that the Pagan ideal will be the ultimate good of man; but if that is so, we must at least ask with more curiosity than he allows for, why it was that man actually found his ultimate good on earth under the stars, and threw it away again.

I do not think these problems completely disprove Reaction. They merely wall off several potential lines of argument in its support: the argument that ancient cultures empirically achieved better outcomes than our own, and the argument that they were more stable and better adapted.

To save Reaction, you would have to try one of the following paths.

First, you could claim that there’s no such thing as cultural evolution, that cultures don’t gradually become more adapted to their conditions via time. This seems plausible, but then the Reactionaries lose their own strongest argument; that older cultures were better adapted. Nevertheless, this is where I think a lot of the remaining probability of Reaction being true would be, and many of the arguments in my pro-Reaction post before continue to stand in this case.

Second, you could agree that cultures evolve, but that for some reason the cultural evolution mechanism has gone berserk over the past few hundred years. To make this stick, you’d have to give some reason this would happen. Then you’d have to prove that it was *so berserk* that the best we could do is reboot from a saved copy from before its breakdown, even knowing that this will be completely unsuited for modern life.

Third, you could posit that for some reason cultural evolution previously drove us in a Progressive direction, but now it is driving us back in a Reactionary direction, and that you are a legitimate priest of the Vast Formless Things just making their new and revised will known unto man. To make this work, you’d have to figure out exactly when and why the Vast Formless Things changed their minds.

For most of the rest of this sequence I’ll be concentrating on option 1, unless a horde of Reactionaries appear in the comments and tell me they have totally considered this problem before and 2 or 3 is the more commonly accepted view. In option 1, by sort of a coincidence past societies happened to be better than ours, and for coincidental reasons ours went off track. The onus then would be to determine which of our society’s policies are or aren’t bad, and what was the last stable copy of them to reboot from.

We Sail Tonight For Singapore

In the comments section to my last post, a user who identified themselves as Singaphile wrote the following:

> Let’s pause for a moment, throw away abstract concepts, and speak only of the concrete. Here is a set of simple, tangible statements:
>
> 1) The city-state of Singapore does amazingly well along almost every objective dimension of socioeconomic well-being: it has low crime, low corruption, high GDP per capita, high life expectancy, high academic achievement, good public transportation, and high economic growth. This is in spite of the fact that it has virtually no natural resources and only about 700 km-sq of land area. Certainly it is much more successful than the US by many measures (eg. it has lower crime AND lower incarceration rates).
> 2) Thus, we should be more like Singapore.
> 3) Singapore is not a liberal democracy.
> 4) Thus, we should move away from liberal democracy.
>
> I don’t believe this argument can be rejected in any comprehensive way except by asserting that liberal democratic values are intrinsically worthwhile, ie. by assuming what you set out to prove.

Singapore is the favorite modern day model government for Reactionaries, and with good reason. It follows sort of Reactionary-ish policies and is a very nice place to live. I totally agree that Singapore is an exceptionally well-run and pleasant country and have no objection to that part of the argument.

But can we conclude that its Reactionary policies *cause* its good government? Or might Singapore have just been an exceptionally well-run country that by coincidence chose to follow Reactionary policies, in the same way Sweden is an exceptionally well-run country that by coincidence chose to follow Progressive policies?

Since the same commenter later protested what ze considered “armchair philosophy”, let’s run an experiment. Heaven has supplied us with a perfect natural control group: Hong Kong. Hong Kong was also an island in southern Asia colonized by the British but mostly inhabited by Chinese people. It is also a city-state with high levels of self-government. Basically, it is Singapore without the Reaction (Hong Kong is probably still more Reactionary than average, but it’s certainly not the poster child for Reaction in the same way Singapore is).

If Hong Kong shares most of Singapore’s advantages, then we can assume Singapore did well because of other factors. The obvious choice for these other factors was a combination of Chinese people plus British people plus laissez-faire capitalism, all of which individually do pretty well for themselves and the combination of which ought to be very powerful indeed.

To perform the experiment I ran through Wikipedia’s List of Lists of Countries and pre-determined which lists I would include in the experiment before having looked at them. I mostly chose ones that seemed like good, relatively independent measures of a country’s success as a society. After looking at them, I threw out the ones that lacked data on either Singapore or Hong Kong. I also threw out one that had lists for 2009, 2010, 2011, and 2012 after noticing that Singapore’s position fluctuated wildly (by like forty countries) each year and frequently bounced between way ahead of Hong Kong to way behind Hong Kong. I decided not to throw out literacy and education, even though some numbers were clearly cooked (for example, Kyrzygystan and Belarus were near the top) because I trust all the countries in this experiment to be responsible in reporting true literacy rates and so their relative standing should be intact.

The ones I ended out with were homicide rate, literacy rate, life expectancy, infant mortality, suicide rate, unemployment rate, one of several measures of GDP per capita which I chose kinda randomly before looking at it, Ease of Doing Business Index, Education Index, Human Development Index, Legatum Prosperity Index, and Satisfaction With Life Index. Take a second to decide which of these are most important to you *now*, so you don't choose only the ones that support your favorite country and then complain in the comments that I should only have focused on those.

Also, in order to keep things interesting I added two extra countries. Macau is another island city state inhabited by Chinese by run by Europeans on a laissez-faire system, so it seemed like a good second control group. About half the lists lacked good data for Macau but I decided that was okay. I also added Sweden, as an example of a well-run but notoriously Progressive country, which I thought would make a good comparison to Singapore as a well-run but notoriously Reactionary country.

Without further ado:

Statistic	**Singapore**	**Hong Kong**	**Macau**	**Sweden**	**Winner**
Homicide Rate	4th (0.3)	3rd (0.2)	13th (0.7)	28th (1.0)	Hong Kong
Literacy Rate	82nd (96.1%)	103rd (93.5%)	122nd (91.3%)	44th (99%)	Sweden
Life Expectancy	10th (80.6)	3rd (81.61)	18th (80.00)	8th (80.88)	Hong Kong
Infant Mortality Rate	4th (2.65)	6th (2.9)	7th (3.17)	5th (2.74)	Singapore
Suicide Rate	91st (10.3)	113rd (14.6)	No data	103rd (11.9)	Singapore

Unemploymnt Rate	? (1.9%)	? (3.3%)	? (3.0%)	? (8.1%)	Singapore
GDP Per Capita	4th ($61K)	9th ($51K)	3rd ($77K)	15th ($42K)	Macau
Ease of Doing Business Index	1st	2nd	No data	13th	Singapore
Education Index	53rd (.913)	87th (.879)	No data	18th (.974)	Sweden
Human Development Index	26th (.866)	13th (.898)	23rd (.944)	10th (.904)	Sweden
Legatum Prosperity Index	19th	18th	No data	3rd	Sweden
Satisfaction With Life	53rd	63rd	No data	7th	Sweden

Sweden wins in five categories, Singapore in four, Hong Kong in two, and Macau in one. This is not exactly a resounding victory for the Reaction.

If we throw out Sweden and Macau, Singapore wins in eight categories and Hong Kong in four. This is a little better, but only until you look at the numbers. Singapore's victories are a lot of things like "Singapore is 4th worldwide in low infant mortality with only 2.65 deaths per 100,000, but that horrible horrible Hong Kong is 6th worldwide in low infant mortality with a practically genocidal 2.9 deaths per 100,000." (for comparison: USA has an infant mortality of 6/100,000).

This seems to provide very strong confirmation of our hypothesis that supercapitalist Chinese-British city-states just do really really well regardless of whether they're notably Reactionary or not. Maybe not *quite* as well as Sweden, but you can't have *everything*.

EDIT: For everyone complaining in the comments that Sweden has this or that unfair advantage over Singapore, I am totally not above bringing up that Singapore is a single city with a beautiful natural harbor right smack in the middle of a fantastic chokepoint in one of the biggest trade routes in the world.

Hitting Below The (Bible) Belt

Every so often I flirt with joining the prestigious, lucrative world of atheism blogging, but the fact is I’m just not cut out for it. I can’t consistently come up with marginally-clever-but-slightly-unfair zingers demeaning religious people on a daily basis.

But I do seem to be gradually getting better at this important skill.

For example, today I saw a hilarious video on the Patheos blog Why I Am Catholic. It was one of those shows where some comedians pretending to be a news team go out and ask the “man on the street” for opinions on increasingly ridiculous news items. In this case, they were asking for opinions on how the new Pope is doing – a new Pope, remember, who has not yet been elected. Watch it. It’s pretty good.

http://www.youtube.com/watch?v=iz-LxZnPjJA

Anyway, my *immediate* reaction was “Man, who could *possibly* have guessed that Catholics would be so quick to praise a nonexistent father figure just because someone exerted mild social pressure on them to do so?”

Good vibrations and subjective temporal granularity

There was a good comment on Reddit recently that offered the least confusing explanation of (insight?) meditation I've ever heard. It sort of made a few things click for me. I'm quoting it in part, but you might also want to read the whole thing:

> *Meditation is basically a training method for your mind. When certain things happen to you, your mind generates a certain response whether it be happiness, frustration, anger ect. The way your mind has been inculcated is the path of least resistance and the path it wants to take, and will take unless you know how to mitigate it. Meditation teaches you how and makes it easier to override the process. Who is doing the overriding of this process? Well, that's the million dollar question. But I digress.*
>
> *So here's what I'm getting at: if meditation is too easy, you're doing something wrong. You might be getting yourself really relaxed, but is it possible that's all you're doing? Not saying it is. I don't know, just throwing some ideas out there and it's up to you to see if any seem to fit your situation.*
>
> *But as you meditate, your mind wants to grab onto the thoughts and not your breath. The course of least resistance is away from your breath and back into whatever thoughts are vying for your attention. Every time you go back to the breath, you train or teach yourself even, to take the opposite of the path of least resistance. This is coupled with the fact that half the time when you meditate, your mind says, "I'm tired.*

> *Stop concentrating on the breath and just kick back and let a guided meditation do most of the work." But every time this comes up you learn to drop it by returning to the breath and not listening to the thought no matter how loud and powerful it can get.*
>
> *When you first start meditating you have this thought and then come back to the breath. But there's still a trace of this thought floating around in your mind and eventually it pulls you in again. As soon as you realize your back in that thought again, you turn your awareness back to the breath and away from the thought. But then it pulls you in again. And then you drop it again. You do this over and over and over. But as you practice you get better and better and faster and faster at recognizing it. You start to figure out how to do it most efficiently and quickly, seeing and dropping thoughts before they even become thoughts at all.*
>
> *After doing this hour after hour, you gain a skill. One day you realize that you don't have to be sitting on a cushion to use this skill. I can't really explain how it's done, but it's just something you learn from continually focusing, coming back to, and holding your attention on the breath. It's like if you ever do a lot of push-ups, eventually you will realize, "I can flex my pecs." You couldn't flex them before, and you don't really know how you learned to do it, but now you can just do it.*

I wish I knew more about exactly what was going on. One thing I've been thinking about recently is subjective temporal granularity.

I first started wondering about this when I realized I could test my subjective temporal granularity through experiment. The partiuclar method was to tap my finger up and down very quickly at different rates, and notice at what rates it was possible to think something like “Yes, my finger is in the up position at this exact moment” versus those rates at which I only had an undifferentiated awareness of rhythmic tapping.

(give me some credit, I also did it with blinking eyes to ensure it had nothing to do with nerve lag, and I used more of a nonverbal “now” rather than that whole long quoted sentence which would probably take me forever to think.)

My consistent finding is that I seem to have subjective temporal granularity on the scale of about a quarter-second.

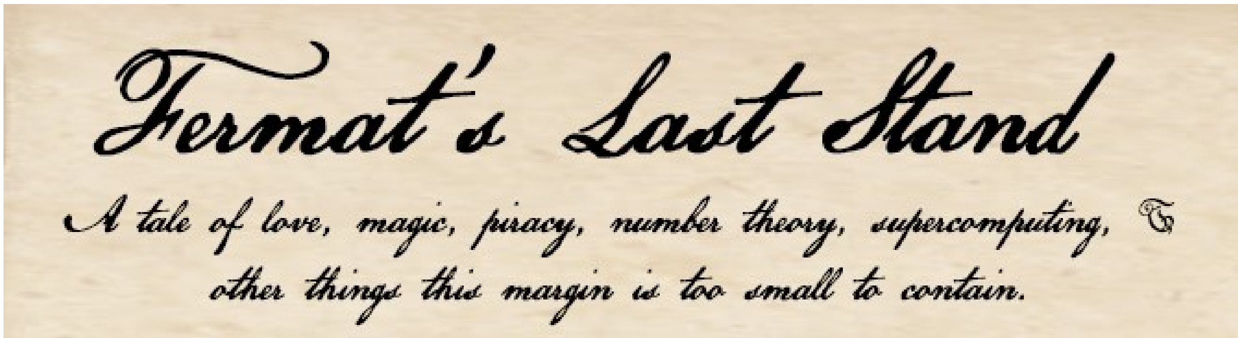
There’s a slightly different form of granularity in that if I try to tap my finger as quickly as possible I get about ten per second, and furthermore I can count these ten and feel pretty sure I got it right. I think what I’m doing here is counting ten intentions to move, rather than the movements, which I can’t actually perceive. But this seems more motor and less related to consciousness.

Experienced meditators claim that when they’re meditating the world occurs in “vibrations”, individual pulses of sensation that occur a couple of times a second. Some meditation practices involve “raising your vibrations” – a phrase which along with its counterpart “good vibes” has long since passed into a byword for New Age gobbledygook – but which originally just meant “increasing the speed at which sensory input seems to pulse.”

I myself have never been very good at detecting these vibrations. Every so often I think I have them, but then I start worrying that I’m actually noting my eye saccades, or the flow of blood through my head, and then I lose them anyway.

But it seems like vibrations might be related to subjective temporal granularity? And maybe meditation is a way of making your subjective temporal granularity finer, so that it can pick up the process of constructing thoughts instead of just the finished thought itself? And that this allows you to intervene in the construction of thoughts, rather than having them just be brute objects which shove their way into your head?

Fermat’s Last Stand: Game Rosters



Fermat’s Last Stand

A tale of love, magic, piracy, number theory, supercomputing, & other things this margin is too small to contain.

Thanks to everyone who responded to the Fermat’s Last Stand Coordination Thread. I’ve sorted people into three groups under three DMs by time zone, who is friends with whom, and belief diversity. If you didn’t list your time zone or your beliefs, I guessed. All groups are online (probably IRC?) – there just weren’t enough people in any particular area.

Your DMs will be myself, Charlie, and Karl. Charlie and Karl both played in the last game and I had a lot of fun with them, plus Charlie has been working on his own house rules which hopefully his group will get a chance to try out.

SCOTT ALEXANDER’S GROUP
(scott[at]shireroth[dot]org)
Sam Burnstein
Leah Libresco
Ozy Frantz
David from New Haven
Joe from London
(Vassar if he can make it)

I have the emails of everyone in this group except David and will be contacting you soon. David, please send me your email. Ari, you’re welcome to hang out and watch again.

CHARLIE'S GROUP *(charlie.steiner[at]gmail[dot]com)*
Pleeppleep (*evanverres[at]gmail[dot]com*)
Aretae (*aretaeblog[at]gmail[dot]com*)
Romeo Stevens (*UNKNOWN*)
Rick JS (*rick[dot]shfhs[at]gmail[dot]com*)
Nathan from LA (*UNKNOWN*)

Everyone in this group please email Charlie at the address above so he has your contact information and can coordinate with you. Rick, I got your notes on the manual and have read through them – I'll send you a response and make the appropriate changes soon.

KARL'S GROUP *(twistortheo[at]gmail[dot]com)*
tster75 (*me[at]tster75[dot]com*)
Tarn Fletcher (*somervta[at]yahoo[dot]co[dot]nz*)
JohnWBH (*UNKNOWN*)
Ben from Oxford (*UNKNOWN*)
Slow Learner (*guyfletch[at]email[dot]com*)
Avantika (*avantikalal1990[at]gmail[dot]com*)
Thomas Eliot (*bgaesop[at]gmail[dot]com*)

Everyone in this group please email Karl at the address above so he has your contact information and can coordinate with you. If by some miracle everyone in this group shows up and it's too big, we can export people to other groups that don't have enough.

There are also meatspace groups operating under falenas108 and gyukuro. You guys know who you are.

I've added all five DMs to a DM Google Group where we can discuss what we're going to do. If there are any other meatspace groups operating, please let me know so I can the DM to this group.

If you somehow missed the coordination thread but are interested in participating, comment below with your timezone, email, and a short description of your philosophical beliefs and we'll use you as replacements for the inevitable people who drop out.

Valleys have two sides

Still planning to write a decent post about Reaction, but it’s taking longer than I thought and kind of just coming out piecemeal, so I’ll roll with that for now.

Many people poked some much-deserved holes in the original post, but one part I didn’t see criticized was about the idea of uncanny valleys.

This was the idea that Progressives, despite having noble aims in mind, were really making things worse because getting halfway to a noble aim was worse than not having any noble aim at all. For example, pacifism might be a wonderful thing, but conducting a war half-heartedly so that it drags on and on may be worse than just being brutally efficient and getting it over with. And no matter what you think of democracy, an unstable despotism where a paranoid despot has to solidify his grasp on power by any means necessary is worse than an absolute despotism where the despot can just hang out and build palaces of solid gold or whatever relaxed despots do.

Let’s concede the point that uncanny valleys like this exist.

In that case, all we have proven is that perfect unimpeded progress up the right side of the valley is better than attempted progress up the left side which is constantly dragged down by lingering rightists.

But this is totally comparing apples and oranges. The opposite of perfect unimpeded progress up the right side of the valley is perfect unimpeded progress up the left side of the valley. The opposite of attempted progress up the left side dragged down by lingering rightists is attempted progress up the right side dragged down by lingering leftists.

In other words, this argument only works because the Reactionaries are comparing *our* gritty reality to *their* beautiful thought experiment.

Like, it seems almost unfair to ask this. But how, exactly, are we going to get a perfectly secure despot with no fear whatsoever of rebellion when less than 1% of the population is currently in favor of such a despot? Even the old-timey kings, who lived in an era when monarchy was *obviously* the only possible government form and everyone believed that they ruled by divine right – even *they* got overthrown and beheaded every so often.

Moldbug mentions the possibility of a military of cryptographically secured weapon systems to which only the monarch has the password, but I don’t think even other reactionaries take him seriously on this one. Traditionally one of the better coup attempt methods is to curry favor among the palace guard, and this does absolutely nothing to prevent that. Another popular coup method is to *be* the military, which means the worst the monarch can do is shut down their fancy guns and then get clubbed to death with rifle-butts. Also, what happens when the monarch dies of a brain aneurysm and all of a sudden your entire military requires a 128-digit number which no one has? And mightn’t it be kind of embarrassing when Scott Aaronson builds a quantum computer that can factor prime numbers in polynomial time and then becomes the new god-emperor?

(and by “embarrassing”, I mean “awesome”)

I endorse the uncanny valley theory of the criminal justice system and (possibly) even of (some) war. It seems reasonable to push for them in the same way one pushes for other, normal political ideas like gun control or abortion. If enough people believe they're correct, then the government will act on them and hopefully they will be good policies. If 51% of people start believing that corporal punishment of criminals is a good idea, we might well start having corporal punishment of prisoners.

But the uncanny valley of dictatorship works completely differently. If 51% of people support it, then we have an extraordinarily unstable dictatorship that does terribly, just like real dictatorships tend to do. Getting 100% of people to support it – in a country where we can't even convince 50% of people that it's a bad idea to keep making a penny that costs 2 cents to produce – seems like a high bar.

And if the goal is to just implement a dictatorship and then let the dictator kill all who oppose him – well, then we've left behind the uncanny valley idea and we're right back in the traditional dictatorship that is a terrible place to live and which has given dictatorship its well-deserved bad reputation. "Well, we'll be brutal for a little while, but then all the enemies will be gone and we can switch to niceness." Sure. That worked for the Communists.

Actually, the Communists are a great example here, in that they had a great plan for how the world would be perfect once everyone agreed with them, and total absence of a plan for what to do about most of the world not agreeing with them.

I said in "We Wrestle Not…" that every age has a zeitgeist based on how to gain and keep power. The zeitgeist may or may not be pleasant, and there may be a lot of incentives to defy it. You can get away with doing so only if you have total control over your own country (so that no one can harness the zeitgeist to win internal struggles) and total control over the rest of the world (so that no foreign country can harness the zeitgeist to take you over).

The Communists got total control over a couple of countries, then realized (correctly, in my opinion) that if they went and formed a nice communal utopia they would be sitting ducks for the first capitalist superpower who wanted to invade. So they compromised with the zeitgeist and ended up being far more oppressive than the capitalists they railed against, all with the vague promise that in the future, they would overthrow the capitalists, control the entire world, and have the power to defy the zeitgeist. Maybe they even believed it.

But that never happened, and they just ended up killing hundreds of millions of people for no reason.

Reaction will never end up with that kind of power, but if it does, I see them in the same position as the Communists. Reaction is contra-zeitgeist. Societies, left to their own devices, become more Progressive; a Reactionary state is going to have to constantly expend energy pumping against entropy to prevent that from happening. That energy is going to take the form of internal oppression and incur automatic enmity with the rest of the world. As long as that happens, they won’t be a secure dictatorship. They’ll be a communist dictatorship promising security just as soon as those evil evil progressives are taken care of, one day in the Golden Future.

Annuntio vobis gaudium magnum: habemus officium!

I have accepted a job as a psychiatrist-in-training in a hospital in Michigan, starting in late June. I won't be more specific publicly because I might want to blog in a very generic and confidentiality-maintaining way about my experiences, and the less identifiable my teachers, colleagues, and *especially* patients are, the better. I will say that I am not in inner-city Detroit (in case you were worried) and that if you live in Michigan you can email me and we should get in touch.

I did not spend my childhood dreaming of one day ending up in Michigan. But I am *so happy*. I was not kidding about the gaudium magnum bit. It is an incredible blessing to have a psychiatry residency at all.

As you may know, there is a serious doctor shortage in the United States, and an even more serious psychiatrist shortage. Medical schools have, depending on your level of cynicism, either "responded to this problem" or "taken advantage of this problem" by increasing class sizes. There are now more seniors graduating medical school than ever before.

But in order to go from "medical school graduate" to "real doctor", one has to complete a residency, a (usually) four year apprenticeship in one's chosen specialty. These are expensive for hospitals to sponsor because there are strict regulations requiring residents to learn many complicated skills in a variety of different settings, all of which must be taught by experienced doctors who charge a lot for their time. And the incentives are poorly-designed: Hospital X may spend mountains of money training a resident who, immediately upon being declared a Real Doctor, packs up and moves to more lucrative Hospital Y. The government tries to help, but as always it is pressed for cash. And every thousand doctors the government helps train means they have to give up another fighter jet which later turns out not to work.

As a result, although the number of patients and of medical school graduates keep growing, the number of residency slots hasn’t changed much since about the 1990s.

This is bad for patients in an indirect hard-to-notice way because there are fewer doctors around and the ones who exist can charge more for their services. But it’s bad for would-be doctors in a *very direct glaringly obvious way*, which is that every year, some people graduate from medical school, get inducted into this very insular culture where medicine becomes the most important thing in their lives, sacrifice four years of their youth and incur hundreds of thousands of dollars in debt – and then are told thanks but no thanks. And then have no idea what to do with themselves.

In the old days, when there were more than enough residencies to go around, American medical students got their pick of the best, and then international students came in to take the rest of them. These could be anyone from Indian or Pakistani doctors who noticed that they could get a much better salary in the US, to Americans who couldn’t (or didn’t want to) get into US medical schools and so went to schools abroad (usually in the Caribbean) set up explicitly for the purpose of transitioning back into the US medical system later. This latter group includes several thousand people and is itself a thriving industry.

Now this is starting to go wrong. International students applying to residencies are in a somewhat desperate situation, getting matched a little less than half the time. And even American students are starting to have a less than 100% success rate – the last statistics I can get show about 95%, but I've heard rumors this year was especially dismal.

I fell in the former category myself last year. I figured Irish medical schools were better than the ones in the Caribbean, that I had good test scores, and that I was probably a shoe-in. I was wrong. Last March 11, when I heard the news, was without exception the worst day of my life. I felt confused, betrayed, and just crushingly depressed. It wasn't just not getting a job offer, it felt like a total rejection of my identity and life path. It was only thanks to my parents and to Alicorn that I came out of it remotely sane.

I told myself I would try one more year. I'd spend one more year doing endless interviews, writing up personal statements and begging for letters of recommendation, one more year doing thankless unpaid internships with doctors who never remembered my name and yelled at me if I got too close to a patient. And then I would figure out something else. As the old saying goes, "If at first you don't succeed, try again. Then give up. No use being a fool about it."

With the help of – well, once again, my parents and Alicorn – not to mention the people at MetaMed – I managed to make it a year without going broke or insane. And this year my luck changed, and after more rejections than I care to think about the nice people in Michigan – in a hospital whose name is an allusion to divine mercy – decided to offer me a residency.

And this is the bottleneck in the system. *Obviously* God smites anyone who says "it'll all be smooth sailing from here on out”, so I won’t do that, but if I can get through this, I’ll be a Real Doctor and people will be begging *me* to work for *them* rather than vice versa.

I’ve been to this hospital in Michigan. It’s a nice hospital. It’s full of nice people who are very good at what they do. And I’m going to get to practice psychiatry, the career I’ve wanted ever since I was like ten. I am *totally beyond caring how cold it is going to be*.

…maybe

There’s this old Taoist parable about a family in ancient China who lived on a farm and owned only one horse. One day the horse ran away.

All the people of the village offered their condolences to the patriarch of the family, saying he must be heartbroken at such a great disaster.

The patriarch only answered “Maybe.”

A few days later, the horse returned with a herd of wild horses. The family was able to capture some of them, increasing their wealth by many times. All the villagers congratulated the patriarch on his good fortune.

The patriarch only answered “Maybe.”

In the process of taming the wild horses, the eldest son was thrown from the saddle and broke his leg. He was responsible for most of the work around the farm, and the villagers all consoled the patriarch on this new disaster.

The patriarch only answered “Maybe.”

A few days later, the emperor's soldiers came through the village, drafting all able-bodied young men to go away and fight in a bloody war far away. Because the son was injured, he wasn't eligible for military service. The villagers all told the patriarch how lucky he was that his son was still around when their own sons were off at war.

The patriarch only answered "Maybe."

I think this story goes on for a few more iterations, but you get the idea.

Since the days of ancient China, we've advanced a lot. Like, instead of just saying "maybe", we can put numbers to our probabilities! If our eldest son who does all the work around the farm breaks his leg, we can say things like "98% chance that, on net, this will decrease my utility, although of course there's a 2% chance that it will help in some unexpected way." This makes a less interesting parable.

But the Taoist teaching is still sound. We are consistently overconfident in our estimation of whether events will be good or bad for us in the long run.

As I said before, last year when I learned I didn't match to a residency was the worst day of my life. I was about as dejected as you've ever seen anyone, and with good reason.

But as it turned out, it led to me moving to Berkeley. For the first time in my life, I got to live with a community of real-life friends who I really got along with. I started realizing what other people meant when they said that going out with their friends was more fun than staying home alone. I went from being terrible with girls and Forever Alone to being terrible with girls yet still mysteriously dating two brilliant and beautiful women whom I will miss *immensely* right up until the point where they give in to my incessant demands to come visit me. I learned all kinds of weird skills and expanded my mental horizon. I got to see a tiny bit of the San Francisco Bay Area, one of the most beautiful places on Earth, and I hope to see a little more of it before I go. I got to be of some small help to groups I really care about, like MIRI and CFAR, and to some people I really care about too.

So basically, my getting rejected from my dream job and having no idea what to do with myself led to me having the best year of my life, becoming a stronger and better person, and then getting accepted to said dream job a year later. I wouldn't recommend suffering life-ruining disasters to *everyone*, I'm just saying it seems to have seriously worked for me.

I don't like perverse incentives, so I'm trying *not* to go into this job with the mindset of "spend four years in exile getting a fancy piece of paper, then move back to the Bay Area so fast it leaves skid marks through the entire American Midwest". I am going to give the Michiganers every chance in the world to convince me to stay around. Maybe there will be someone, somewhere in Michigan, who is as consistently awesome as the people I have come to know here.

But in deference to the ancient Chinese, I am going to carefully avoid attaching a probability estimate to that statement.

You are all wonderful. Please keep being wonderful. And one day I am going to be a fully-qualified psychiatrist, and then I can help people be wonderful myself.

Not just a mere political issue

Here’s a quote of the new Pope’s that’s being passed around:

> Let’s not be naïve, we’re not talking about a simple political battle; [homosexuality] is a destructive pretension against the plan of God. We are not talking about a mere bill, but rather a machination of the Father of Lies that seeks to confuse and deceive the children of God.

Yes, yes, I know. The new Pope is Catholic. It is an outrage and everyone involved should be ashamed of themselves.

But this quote *mainly* interested me because I have long been confused about this idea of “not just mere politics”.

Abortion is a classic political issue. I happen to be pro-choice, but I have many pro-life friends. Technically, by their philosophy, I support murdering millions of babies. I don’t think these babies are people in the morally relevant sense, but these friends certainly do. This seems like it should be a problem. Is it really okay to be a friend with someone who wants to murder millions of people?

Yet my friends show no sign of wanting to not be friends with me anymore, or liking me even a little bit less. And if we were having dinner and one of them were to say “You know, Scott, you support murdering millions of babies and that seems bad”, then every social norm in the world would consider *them* to have made the social *faux pas*.

If it turns out that the CEO of McDonalds is pro-choice, only an extremely tiny set of people would think this was a valid reason to boycott McDonalds, and the rest of us – including other pro-lifers! – would mock them. If someone were to refuse to hire an employee merely because she was pro-choice, it would be an outrage. If a pro-choicer were appointed as a principal, then outside a few very special areas there would be no picketing the school and screaming that a murderer is unfit to teach our children.

I'm not just picking on pro-lifers here. Pacifists could blame the people who supported the war on Iraq for about a hundred thousand deaths. If I really wanted to, I could blame the anti-nuclear lobby for the ten thousand or so people who die from coal pollution each year. This seems like ample reason to hate anyone.

So this world "politics" has a weird sort of magic. Merely by saying "political issue!", we can make it socially unacceptable to hold people's decisions to kill millions of babies against them. Not just in a legal way of "the government can't censor these people", but in a very personal way of "you can't even dislike them".

So when Pope Francis says that homosexuality is "not a simple political battle", one interpretation of his statement is that this sort of "oh well, it's a reasonable issue where decent people can disagree" ethos doesn't apply there.

We non-Popes have our "not just political" issues as well. Compare killing millions of babies to killing millions of Jews.

If I came out in support of killing millions of Jews, probably some of my friends would find this sufficient reason to stop being friends with me. No one would consider this a *faux pas* or mock my friends as extremist. Heck, they might even get angry at my friends if they *continued* to hang out with me after that.

If the CEO of McDonalds came out in favor of the Holocaust, people would boycott McDonalds immediately, and everyone would support the boycotters. If an employee were discovered to be anti-Semitic, firing them would be considered an extremely proportional response. And if a Neo-Nazi became school principal, you better believe that school would be picketed.

Some non-Popes even agree with Francis on homosexuality. Back when California passed Proposition 8, I had to listen to friends discussing how to “get revenge” on Prop 8 proponents through shaming or harassment campaigns. When I pointed out that this is not really how you’re supposed to do politics, they told me homosexuality wasn’t about politics, it was about stigmatizing entire groups of people and denying their very right to exist.

So both Popes and non-Popes have what seem like lines in the sand, where our mysterious tolerance for political issues breaks down. As far as I can tell, things like the war in Iraq, abortion, global warming, and affirmative action are on the tolerable side; anti-Semitism, eugenics, and *sometimes* gay marriage are on the intolerable side.

This doesn’t seem to correspond to the importance of the issue – getting global warming right is clearly more important than getting gay marriage right, since, *contra* the protestations of televangelists, only one of those two is likely to lead to an apocalypse of fire that dooms us all. It doesn’t seem to the controversialness of the issue either – it was still perfectly acceptable to oppose the War in Afghanistan even in the early days when 80+% of the population was in favor, and gay marriage seems to be creeping towards the “no tolerance” side of the line even though the population is split almost perfectly 50-50. The line seems to have *something* to do with identity politics, but I can think of exceptions (for example, legalizing certain deviant sex acts).

More interesting than trying to justify the line’s current position is wondering where the line ought to be. This is harder than it sounds. Like, if I were trying to draw a line totally *a priori*, I would tolerate every political opinion except being against euthanasia for people who are terminally ill and suffering, which would be so far beyond the pale that you would immediately be kicked out of society. If you forced me to add other forbidden topics, I would probably start with being opposed to military intervention to end genocides abroad. Since no one else seems to agree with me on either of these, my guess is everyone would put the line in a different place and it would be super-confusing.

Maybe we could not have a line at all and just tolerate all political opinions? This sounds sort of attractive to me. As a teenager, I met a genuine neo-Nazi online. He seemed to be a pretty nice guy, and though he was aware I was Jewish, he kindly clarified that he thought killing *all* the Jews was a bad idea and it was really only necessary to get rid of the leaders of the Zionist-Israeli conspiracy. Anyway, we talked on and off for a few years, and if I’d known him in real life I probably would have been happy to invite him over for dinner.

This makes me think that my intuition of “hate neo-Nazis” is mostly based upon the assumption that neo-Nazis are the sort of people I would hate anyway: that this political opinion is correlated with non-political traits like being violent, being ignorant, being rude, and so on. If there were no intolerance line, I could continue to dislike the violent ignorant neo-Nazis for their violent ignorance, I could continue to object to Nazism in the same detached bloodless way I currently object to Communism or libertarianism, and if there were otherwise-decent neo-Nazis I could hang out with them and invite them to dinner. It seems like a reasonable plan, although I don't know how many otherwise-decent neo-Nazis there are.

But here's the problem with throwing out the tolerance line.

I consider myself pro-choice, not just in the sense of "I'm not willing to use the government to ban abortions", but in the sense of believing that in some cases abortion is morally permissible and even a good idea. I also consider myself honest, non-hypocritical, and willing to stand up for what I believe in. I can't think of any situation in which it would actually be a good idea for me to perform an abortion, because I have no training in that area and would probably screw it up. But assuming some weird confluence of conditions in which it was practically necessary for me to perform an abortion, it would be extremely hypocritical of me to refuse for moral reasons: I would perform the abortion. Absent some weird and uninteresting quirk like being hopelessly grossed out by surgeries, I think any non-hypocritical pro-choice person would have to say the same.

But that means that, at least for non-hypocrites, there's very little distinction between supporting something and doing something, save the situation. If I'm willing to invite someone pro-choice over for dinner, I should also be willing to invite an abortion doctor over for dinner, or else I am simply rewarding the former for sheer moral luck.

(Yes, there are some caveats. For example, you might think a person is too decent to perform an abortion but doesn't know it yet; that once the actual condition came up and they had the woman in front of them and maybe saw an ultrasound of the baby, they would chicken out. In that case, knowing that someone was an abortion doctor would give extra information beyond them being pro-choice. This is one way I tolerate people who are against euthanasia. Or you might be trying to socially boycott abortion doctors in order to make abortions impossible by discouraging people from taking up abortion docting as a career. Or you might be worried about losing face with your pro-life friends. All of these are boring and tangential to the issue at hand.)

(And it's possible that until you read the paragraph above, I had plausible deniability for just being someone who hates abortion but thinks it's wrong for the government to meddle in it, which is a plausible deniability that the abortion doctor does not have. But there are other cases where that doesn't apply: for example, with the people who oppose euthanasia for the terminally ill and suffering.)

Likewise, anyone willing to invite a neo-Nazi who *supports* committing genocide to dinner should also be willing to invite an actual concentration camp guard over. This isn't quite total moral relativism – you can still think being a concentration camp guard is *wrong* – but it seems pretty close. Once you decide not to hold someone being a concentration camp guard against them, even a tiny little bit, can you get upset with anyone at all?

Actually yes. One thing that aborting babies and running concentration camps have in common is that they're (at least in their respective societies) *legal*. Although there is not necessarily a huge moral difference between pushing for state-sanctioned genocide and engaging in genocide yourself once it has been state-sanctioned, there does seem to be some manner of moral difference between pushing for genocide and going off and killing some Jews now even though it is illegal. So one possible stable solution is "Tolerate people's political opinions, tolerate their legal political actions, but feel no obligation to tolerate illegal political actions if you disagree with them."

(This should be kept separate from the related but obviously horrible statement "Anything that is legal is morally correct". This sort of "tolerance" in the sense of "do not cause social chaos by actively personally hating the people involved" is different from the sort of tolerance where you actually *tolerate* having concentration camps in your society.)

Still, it seems pretty dreadful. Not being antisocial towards anyone for their political opinions, or *even* their evil politically-motivated actions? But it sort of makes sense in the context of ostracism being viewed as a consequentialist tool for pressuring people into changing their beliefs and actions. After all, violence is also a tool for pressuring people to change their beliefs and actions, and we only deploy *that* tool when people are breaking laws as well. Any other use of it risks getting into wars of all against all such that using laws as a guide instead is an obviously superior solution.

And it gives the right answers to some hard questions. Like, I think most of us now believe that harassing the soldiers coming home from Vietnam was a mistake, even though they're in the same category as the abortion doctors – people who actually committed potential atrocities for political reasons. I even think most people are now willing to forgive most of the ordinary Germans who were in the Nazi Party and helped in the Final Solution as no doubt little different from the populations of other countries that existed at the time, or from Germans in the generation before or after.

But this is one area where my intuitions diverge from my thought processes. I'm still not too sure which to trust.

My objections to “objectification”

Ozy keeps writing about objectification, which they seem to use synonymously with “fetishization”. I had high hopes they would finally be the person who uses that word in a way that makes sense, so I can stop just assuming it’s a totally made-up horrible useless concept. But this was not to be.

Ozy is bisexual and gets upset when people with a fetish for bisexuals proposition them, describing it as “some asshole dude deciding that your sexuality is for his entertainment” and saying that “I am deeply creeped out by dudes who are more attracted to me because I’m bi”.

They are equally upset when people “objectify” fat people, trans people, Asian people, black people, dominant people, submissive people, et cetera. If I understand this right, they say being attracted to a dominant person just because they’re dominant is “believing that people you’re attracted to exist for your boner”; being attracted to fat people is equivalent to telling them “look, people who are attracted to you are so rare that you should be willing to fuck literally anyone who’s attracted to you”.

It is probably a bad sign when, in order to criticize a concept, you have to make your hypothetical target example say and think things *totally unrelated to that concept and much worse than it*.

Look. Suppose Bob is a sadomasochistic amputee who has a fetish for fat transgender Asian women. And suppose he happens to live next to Alice, a fat transgender Asian woman who has a fetish for sadomasochistic amputees. They can either both sit alone in their rooms, feeling like approaching the other would be vaguely sinful for reasons they can't exactly parse. Or they can, as the economists say, derive gains from trade.

I don't think Ozy objects to Bob and Alice starting a relationship. I just think they object to either of them ever *asking* the other whether they'd like to start a relationship. If the Relationship Fairy fell out of the sky and magically informed them that they were interested in each other, then it would be totally acceptable. But if one of them has to *go and check*, then that's *wrong*, because it might make one of them feel "objectified".

But if you lash out in anger any time someone investigates whether they can make a mutually beneficial deal with you, such that fear of this reaction prevents everyone all around the world from ever trying to make mutually beneficial deals with anyone else, *then the problem is with you*.

I have not had much experience with psychiatry yet but one of the things I have learned is that human fetishes are one of the most powerful forces in the world. I met a man who had used up pretty much his entire life savings going to niche prostitutes who would satisfy his fetish; eventually it came out and led to a divorce. We asked him: did you ever ask your wife if she might be interested in doing this with you, so you didn't have to cheat on her in the most expensive possible way? "No, of course not, it would have been so embarrassing, imagine what she would have thought of me, I couldn't degrade her like that."

I have just enough faith in the fundamental perversity of the Universe to consider it inevitable that his wife shared the same fetish and was probably going to expensive prostitutes of her own, like a disgusting x-rated version of *Gift of the Magi*. If they had *just talked to each other* they could have sorted out the whole debacle. But no, because of this concept of “objectification” we *can’t have nice things*.

And I am *still* loading the dice in Ozy’s favor by mentioning all these sadomasochistic fetishes and amputee fetishes that make it look like objectification is something perverted. What if you like guys with red hair? Guys who dress in suits? Girls who play the harp? Girls with an unwavering committment to leave the world a better place than it was when they arrived, which shines brightly forth through their every word and action?

(I have both of the latter two fetishes. It’s terribly embarrassing and you must never tell.)

But seriously. How come we think “attracted to a girl who plays the harp” is socially acceptable, but “attracted to a transgender woman” is a disgusting fetish?

(Ohmigod. I just realized my girlfriend doesn’t actually love me for who I am. She just loves *all of my characteristics*.)

As far as I can tell, the distinction seems to be that liking someone because they play the harp is the sort of thing associated with wanting a long-term relationship partner, and liking them because they’re transgender is the sort of thing associated with wanting to have casual sex.

If someone likes you because you’re transgender, you have to not just politely say no, but become extremely offended and yell at them, or else people will suspect you’re the kind of person who thinks having casual sex with people is okay, ie a slut.

Which is fine, if you’re one of those conservative people who think sluttiness is a real thing and you must never have casual sex because that is wrong. Yet the people attacking “fetishization” claim to be sex-positive. For them to continue hating on “objectification” doesn’t compute.

I talked to Ozy about this, and they finally said that their real problem is people assuming they deserved sex with them because the person had a bisexual fetish and Ozy was bisexual. And that sometimes this led to them being pushy or pressuring them for sex.

This is obviously a legitimate complaint. It’s just not a complaint about objectification.

If someone learned everything about Ozy, and fell deeply in love with them for their amazing personality, and then said “I deserve sex with you, whether you’re interested or not, and it has nothing to do with your bisexuality, it’s solely because I’m deeply in love with your amazing personality” and then kept pressuring them about it, then even though this is exactly the opposite situation it would be *exactly as bad*.

But the point is, there is absolutely zero wrong with objectifying someone. There is a lot wrong with ignoring other people’s right not to consent. But we knew ignoring non-consent was bad already. All the concept of “objectification” adds is giving you an excuse to hate innocent people because they’re in the vague periphery of a bad thing. Screw that.

I wrote this about gender because people only read blog posts when they’re about gender or politics, but the same principle of “Informed consent is sufficient and all that worrying about objectification adds is preventing mutually beneficial deals” applies to pretty much all discussion of objectification including in philosophy.

If opposites attract, why is my sister so pretty?

My girlfriend and I went on a beautiful three-day hike all alone through the romantic Marin Hills. And the conversation, as it so often does in such cases, turned to asexual reproduction and incest.

She brought up the example of heterogamous organisms to alternate between sexual and asexual reproductive strategies, preferring parthenogenesis in times of plenty and sex when during times of tougher competition.

And this makes sense. If you expect all your offspring to survive, then it's evolutionarily advantageous to load these offspring with your own genes, so you can send 100% of your genome to the next generation instead of just 50%. But if you expect heavy competition, then in the end only a few highly-optimized organisms that take the best characteristics of all their predecessors will survive. You can either contribute some portion of your genes to these future uberbeings, or be left out entirely.

Obviously you don't want to mate with an organism identical to you; if you did, this would be very similar to asexual reproduction – not exactly similar, since you'd still randomize allele distribution – but similar enough that going through sexual reproduction seems sort of like a waste of time. So you want to mate with an organism that is at least a little genetically distinct from you. And it seems that there are many reasons to want to mate with an organism *maximally* genetically distinct from you. This increases the rate at which you achieve the offspring-optimizing capabilities of sexual reproduction, and is best for confounding parasites, who prefer their hosts to have high levels of genetic uniformity for about the same reason that human pharmaceutical researchers prefer the parasites they target to high levels of genetic uniformity.

She supported this position empirically, noting that the children of incest (especially high levels of genetic similarity) are unusually prone to genetic problems and maladapted, while the children of two different strains or species are unusually prone to genetic success and well-adapted.

More interesting, there’s a set of genes called the major histocompatibility complex (MHC), which to somewhat oversimplify things are involved in the immune system. The more diverse your combination of MHC genes, the more pathogens your immune system can recognize and the harder it is for pathogens to fool your immune system. Some animals seem to be able to *smell* MHCs and deliberately select mates with different MHCs than themselves in order to give their children new and exciting MHC combinations.

A famous experiment by Wedekind et al suggested humans may do this as well. Women were asked to smell different men’s sweaty t-shirts and rate their attractiveness. The women were most attracted to t-shirts worn by men with dissimilar MHC genes. A further study found that couples had much more different MHCs than would be predicted of two randomly selected members of the population. This has since been replicated a few times with moderate success.

So she concluded that animals probably choose mates to maximize genetic dissimilarity.

I made the opposite argument. We know that close relatives (especially siblings) separated at birth who meet as adults very often have unusually strong sexual attraction to each other, to the point where it seems hard to explain by coincidence or even by the people's general kakosexual tendencies. Well-intentioned attempts to come up with a proper social explanation of this by attributing it to "imprinting on the opposite-sex parent" seem to ignore that most of these people were, in fact, separated from their opposite-sex parent at birth.

In another one of those experiments that make it obvious that psychologists have the best job in the universe, a group of scientists asked subjects to rate the sexual attractiveness of various people in photographs, but secretly flashed subliminal pictures of the subjects' opposite-sex parents during some of the photos. Sure enough, people found the photos associated with their opposite-sex parent more attractive.

Yet another study shows that people's spouses are about as similar to them as their fraternal twins – and further, that spouses are more similar on highly heritable traits than on less heritable traits (warning: this study was done by Jean Rushton, who is most famous for studies claiming that between-race differences are genetic. You may wish to use this information when considering how much to trust his work).

Finally, homogamy passes the ocular trauma test in a way that heterogamy doesn't. Most people tend to marry people who are a lot like them. It seems highly unlikely that I will marry an !Kung bushman. Part of this is geographical (I don't meet many !Kung); part of it is social (I'm from a society that has given me standards of beauty and expectations for interpersonal relationships which most !Kung don't satisfy). But the allele chart confirms that the !Kung are the human population most genetically different from me; it still seems relevant that I am not the least bit attracted to them. On the other hand, I *do* find myself dating a woman who takes long hikes through the mountains while talking about the role of histocompatability complexes in natural selection, which probably puts her in like the top 0.01% of the population for similarity to me already.

But there's not a whole lot of good theoretical explanation for this (I continue to maintain that the "imprint on opposite sex parent" thing is dumb). One possibility I proposed was that this maximizes the number of your genes that continue to the next generation. That is, suppose a male lab mouse mated with a female lab mouse who shared 50% of his genes. Now each of his offspring has 75% of his genes rather than just 50%, so more of his genes are passed on to the next generation. Whether or not this works seems to depend on whether or not this is funging against both partners having kids with someone else instead; most likely it would for the female and wouldn't for the male. This produces weird predictions about sex-specific Westermarck effects which as far as I know have never been observed.

So right now as best I can tell the status of evidence on this question is that people are attracted to other people who are genetically similar to them *except* in the case of MHC complexes where they're attracted to other people who are genetically different from them. This seems very weird and not parsimonious and I have no logical explanation for it.

Book Review: Infinite Jest (alternate title: "Look At Me! I Read Infinite Jest!")

Warning: Composed entirely of spoilers for *Infinite Jest.*

I was talking to my roommate about *Infinite Jest*, and he asked me how it compared to *Illuminatus*. In an instant of sudden clarity I realized he was right, that they were basically *the same book* – the same fractally complex narrative of broken meandering plots about an endlessly-awaited but never arriving apocalypse; the same weird nightmare funhouse mirror held up to the world; the same thousand-odd pages followed by hundreds of complicated footnotes; the same ambiguously-successful attempt to draw a spark of honest feeling out of the detached amusement that is post-modernism.

…and then it turned out that my roommate had just gotten confused and thought Robert Anton Wilson and David Foster Wallace were the same person, because they both had three names. An understandable mistake, I guess. But I stick to my assessment.

The biggest difference between the two is that where *Illuminatus* revels in the *weird* – talking supercomputers, giant sea monsters, telepaths, immortal sages – *Infinite Jest* manages to do the same thing with the *almost normal*. Almost as much a pastiche of short stories as a coherent narrative (actually, it's almost as much a three-toed sloth as it is a coherent narrative) each episode could be a particularly creepy or poignant urban legend – not the ghost or alien kind but the kind where something happens so perfectly symbolically that it forms an image that sticks with you forever. The “woman microwaves her dog to dry it off” kind, or the “man leaves his entire fortune to the first person to pray at his grave” kind.

Which is not to say that *Infinite Jest* is anything close to “glurge”. The book is incredibly heavy (literally and figuratively) and has so many cycles within cycles that some of the most important events in the book, let alone the themes, are totally opaque to a first reading. But after going through it and reading some of the various analyses on the Internet, I am prepared to take a stab at saying what it is about.

Infinite Jest is about wireheading. More specifically, it’s about not wireheading. It’s about the difference between masturbation and sex. It’s about the dangers of self-reference and the need to connect to something outside yourself.

So how’s this for symbolism: the book is set in a future USA-Canada-Mexico merger called the Organization of North American Nations. I originally thought O.N.A.N. was just a cheap gag, in the same way Robert Anton Wilson called his supercomputer F.U.C.K.U.P. But David Foster Wallace does not do cheap gags. The book was taking place in a society literally named for masturbation, and its entire structure was based on pleasure without purpose.

O.N.A.N.ite society is dominated by the InterLace cartridges, which as of the book’s publication in 1996 must have seemed like a dystopian future technology, but which to modern readers sounds pretty much like an unimpressive clone of Netflix with fewer features. InterLace allows everyone to watch whatever they want whenever they want, destroying the one redeeming feature of network television – that at least people watched the same things as their friends and could talk about them. Instead, a host of niche shows have sprung up; some saccharinely corporate, others bizarrely post-modern.

James Incandenza made this latter type of film cartridge. He is probably the best candidate for a "main character", even though he died five years before the book began, in the Year of the Trial-Sized Dove Bar (the government has decided to raise money by selling corporate sponsorship of year names, giving time a sort of dreamy, non-ordinal quality). Footnote 24 of the book gives an eight page filmography of Incandenza's productions along with plot summaries (did I mention this book is ridiculously detailed?), including such bizarre numbers as *Pre-Nuptial Agreement of Heaven and Hell*, *At Least Three Cheers For Cause And Effect*, *It Was A Great Marvel That He Was In The Father Without Knowing Him*, and the book's namesake, *Infinite Jest*.

This last film, never released, has the unusual property that anyone who sees it will continue watching it forever in a state of bliss, unwilling and unable to do anything else ever again. When rumors of its existence reach a group of Quebecois separatist terrorists, they attempt to distribute it freely, believing that Americans are so entertainment-obsessed that they will voluntarily choose to watch it. The Office of Unspecified Services, a government bogeyman organization, tries to thwart the terrorists and keep the film under wraps, leading to a really fascinating discussion on the value of whether people should have free choice in a world where it is so easy for one's choices to enslave oneself.

[I feel a need to mention these terrorists' other *modus operandi* here. They go to deserted mountain roads late at night and install huge portable mirrors. When a truck drives down the road, it sees the headlights of another truck coming right towards it. When it swerves right, the other truck swerves right; when it swerves left, the other truck swerves left. Finally, desperate to avoid the inevitable collision, the truck will swerve straight off the road and be destroyed on the cliffs below; the terrorists will then remove the mirror and head off, leaving the authorities to assume an ordinary car accident. They are caught only when a suicidal truck driver decides to take the opportunity to let the other truck finish him off, and instead crashes straight through the mirror, saving his own life and revealing the deception. Are you starting to see what I mean about the AMAZING CREEPY NIGHTMARE SYMBOLISM?]

So the film is clearly a symbol for wireheading, and in fact some of the characters discussing it specifically mention the wireheading experiments that have been done on both rats and humans. But the other half of the book, which centers around the lives of drug addicts in a Boston Alcoholics Anonymous program, is equally wirehead-related, and the dull boring everyday struggles of the addicts form an interesting counterpoint to the wacky sci-fi struggles to control Incandenza's deadly film cartridge.

One counterpoint to wireheading is competition. A worse author would have taken sides, saying that seeking pleasure alone is a great evil, competition – the desire to be the best at something, even when it hurts and isn't pleasant at all – is a great good. Instead, all of Wallace's depictions of competition are sort of horrific. Enfield Tennis Academy, a school founded by James Incandenza during his "tennis phase" (as distinct from his "film director" phase and his "invent working fusion power" phase) is full of hypercompetitive children who devote their entire lives to mastering tennis. Some become boring and robotic inside, losing the depth for any thoughts beside their position on a ranking chart. Others turn to drugs – of which, by the way, David Foster Wallace has an encyclopaedic knowledge that boggles even my professional-psychiatrist mind. Still others become nervous wrecks.

[There is a brief story about Eric Clipperton, a boy who would show up at junior tennis tournaments with a loaded gun, which he would hold to his head throughout his games, threatening at each moment to kill himself if he lost. Each tournament his opponents would throw their games and Clipperton would take first place, then immediately running off and disappearing before he could get arrested. Eventually Clipperton became the number one ranked junior tennis player in O.N.A.N., at which point he had nothing left to live for and promptly shot himself. Did I mention the AMAZING CREEPY NIGHTMARE SYMBOLISM?]

Among my favorite sentences was a part about LaMont Chu, one of the Enfield Tennis students, who is desperate to one day have his picture in a glossy sports magazine. The Tennis Academy's on-site guru (of course the tennis academy has a guru!) tells him that "After the first photograph has been in a magazine, the famous men do not enjoy their photographs in magazines so much as they fear that their photographs will cease to appear in magazines."

LaMont and Clipperton seem to represent the hedonic treadmill, the fact that getting what you want doesn't make you any happier, it just makes that your new set point and you want something else. Ironically, this is no different than the drug addicts, who find that as much as they enjoy their drugs, after a while they adjust to that dose and need a higher dose. Or to parallelify it to the guru's quote, they do not enjoy their drugs so much as they fear that they will cease to get their drugs.

So if wireheading is sterile, and competition is useless, is there anything worthwhile? Wallace's answer is…probably too complicated for a blog post, but it seems to be about *anything other than yourself*.

Sometimes this means Helping Other People, and this is one of the few heavy-handed parts of the book. M. Remy, a Wheelchair Assassin (there, uh, might be an entire cell of wheelchair-bound terrorists in this universe) becomes suicidal after his disabling accident but regains the will to live after rescuing a deformed woman from a truck – actually, a suspicious number of the decent, likeable characters in this book are deformed, so much so that a Union of the Hideously and Improbably Deformed is an organizational bit player in several scenes.

But there seems to be this more vague sense in which it's about interpersonal communication. Better people than I have tried to explain the subplot centering around Hal Incandenza, and I should probably just quote:

> Central to the dramatization is Hal Incandenza, who opens the novel by unnerving a panel of college administrators when he speaks to them. Hal [whose speech inexplicably comes out as horrible nightmarish shrieks] tries to tell the admissions panel things like:
>
> "I have an intricate history. Experiences and feelings. I'm complex…
>
> I'm not a machine. I feel and believe. I have opinions. Some of them are interesting. I believe the influence of Kierkegaard on Camus is underestimated. I believe Dennis Gabor may very well have been the Antichrist. I believe Hobbes is just Rousseau in a dark mirror. I believe, with Hegel, that transcendence is absorption.

> Please don’t think I don’t care.”
>
> This is quite a contrast from what Hal feels later in the book/earlier in the events about being basically empty of feeling:
>
> Hal himself hasn’t had a bona fide intensity-of-interior-life-type emotion since he was tiny; he finds terms like joie and value to be like so many variables in rarified equations, and he can manipulate them well enough to satisfy everyone but himself that he’s in there, inside his own hull, as a human being — but in fact he’s far more robotic…inside Hal there’s pretty much nothing at all, he knows.
>
> The admissions panel members respond to Hal’s admissions of feeling, by freaking out. They don’t understand Hal’s voice; they are, in fact, terrified by it.

Much of the book is devoted, in its meandering way, to the story of how Hal, a child prodigy, was emotionally dead inside but managed to fool everyone except his father James. In the climax of the book, James admits that he made the fatally-entertaining *Infinite Jest* as an attempt to create something so emotionally overwhelming that it would break his son out of his shell. Although the crucial scene is left out, we transition from Hal being emotionally dead but extraordinarily successful and beloved to Hal being a real person with real feelings but unable to communicate, his voice replaced by horrible screams.

Some reviewers have said that the whole book is meant as a critique of postmodern society. I’ve previously cited Wallace’s critique of irony and I find this passage, cited on the same site linked to above, especially enlightening:

> The next literary “rebels” in this country might well emerge as some weird bunch of anti-rebels, born oglers who dare somehow to back away from ironic watching, who have the childish gall actually to endorse and instantiate single-entendre principles. Who treat of plain old untrendy human troubles and emotions in U.S. life with reverence and conviction. Who eschew self-consciousness and hip fatigue. These anti-rebels would be outdated, of course, before they even started. Dead on the page. Too sincere. Clearly repressed. Backward, quaint, naive, anachronistic. Maybe that’ll be the point. Maybe that’s why they they’ll be the next real rebels. Real rebels, as far as I can see, risk disapproval. The old postmodern insurgents risked the gasp and squeal: shock, disgust, outrage, censorship, accusations of socialism, anarchism, nihilism. Today’s risks are different. The new rebels might be artists willing to risk the yawn, the rolled eyes, the cool smile, the nudged ribs, the parody of gifted ironists, the “oh how banal.” To risk accusations of sentimentality, melodrama. Of overcredulity. Of softness. Of willingness to be suckered by a world of lurkers and starers who fear gaze and ridicule above imprisonment without law.

Under this theory, younger-Hal starts off as postmodern, totally unengaged with the world except mechanically; during this phase of his life, he is considered witty, lovable, and successful. Then, after some mysterious intervention by his father (whose initials are, significantly, J.O.I.) he becomes a whole human being again, but his attempts to communicate honest feelings are met with revulsion and horror. Under this theory, ONAN and wireheading are the inexorable results of turning further and further inward and denying that engagement with an objective reality is anything particularly desirable.

I like this theory, but it has one problem, which is that *Infinite Jest* is an obviously, unabashedly postmodern book. It is the *most* postmodern book. Maybe this is the joke, that it's turning postmodernism's critical methods onto a critique of postmodernism itself. But that seems like kind of an easy gag. Is there something a little deeper going on?

Well, my hands-down favorite part of the book was the description of "Eschaton", a game played by drawing a map of the world on a tennis court, setting up various tennis players as Great Powers, and then simulating a nuclear war through a combination of extremely complicated math equations and tennis balls as nuclear missiles. The game goes well enough until it starts snowing, and the players get into an argument about whether they need to adjust for the snow when calculating the damage done by their nukes. I can't resist quoting an abbreviated chunk:

> "It's snowing on the goddamn map, not the territory, you dick!" Pemulis yells at Penn, whose lower lip is out and quivering. Pemulis's face is the face of a man who will someday need blood-pressure medication, a constitution the Tenuate doesn't help one bit. Troeltsch is sitting up straight and speaking very intensely and quietly into his headset.
>
> "Except isn't the territory the real world, quote unquote, though?" Axford calls across to Pemulis, who's pacing like the fence is between him and some sort of prey. Axford knows quite well Pemulis can be fucked with when he's like this: when he's hot he always cools down and becomes contrite.
>
> "The real world's what the map here stands for!" Lord lifts his head from the Yushityuand cries over at Axhandle, trying to please Pemulis.

> "Kind of looks like real-world-type snow from here, M.P.," Axford calls out. Struck finally falls out of his chair with a clunk but his legs are stil somehow entangled in the legs of the chair. It starts to snow harder, and dark stars of melt begin to multiply and then merge all over the courts. Otis Lord is trying to type and wipe his nose on his sleeve at the same time. J. Gopnik and K. McKenna are running around well outside their assigned quadrants with their tongues outstretched. "Real-world snow isn't a factor if it's falling on the fucking map!"

And then it gets worse. The Iranian player decides that instead of launching his nukes at America's cities or military bases, he will assassinate the American president directly by hitting the American player with a tennis ball:

> Just outside the theater's fence, Pemulis is bug-eyed with fury – not impossibly 'drine-aggravated – and is literally jumping up and down in one spot so hard that his yachting cap jumps slightly off his head with each impact, which Troeltsch and Axford confer and agree they have previously seen occur only in animated cartoons. Pemulis howls that Lord is in his vacillation appeasing Ingersoll in Ingersoll's effort to fatally fuck with the very breath and bread of Eschaton. Players themselves can't be valid targets. Players aren't inside the goddamn game. Players are part of the apparatus of the game. They're part of the map. It's snowing on the players but not on the territory. They're part of the map, not the cluster-fucking territory. You can only launch against the territory. Not against the map. It's like the one ground-rule boundary that keeps Eschaton from degenerating into chaos. Eschaton gentlemen is about logic and axiom and mathematical probity and discipline and verity and order. You do not get points for hitting anybody real. Only the gear that maps what's real.

> Pemulis keeps looking back over his shoulder to the pavilion and screaming.

This scene has been justly described as “a history of literary postmodernism under the guise of geeks throwing tennis balls at each other”. Postmodernism is an attempt to blur the boundaries between a text and the world.

But what Wallace seems to dislike is the postmodernist claim that therefore the world has no privileged status and that it’s *all* just text – this seems akin to the O.N.A.N.ism of the book or the original deadened version of Hal. If I had to take a guess as to why *Infinite Jest* is postmodernist, I would say that it is to blur the boundaries between the world and the text in favor of reality.

It seems relatively clear that James Orin Incandenza, the haunted but brilliant postmodernist director who produced the film *Infinite Jest*, is a standin for David Foster Wallace, the haunted but brilliant postmodernist author who produced the book *Infinite Jest* (after all, as my roommate would point out, both of them have three names). It also seems that the two *Infinite Jest*s are supposed to be the same – and several readers have noted that, like the viewers of the fatal video, they started rereading *Infinite Jest* as soon as they finished it.

So, as one reviewer I read put it, perhaps just as the film was James’ attempt to produce something that could communicate with the true core of Hal and bring him back to the real world, so the book is Foster Wallace’s attempt to communicate with the true core beneath postmodernist society and reach in and drag it back to reality.

There are other, weirder referential loops. James Incandenza commits suicide by sticking his head in a microwave (possibly meant as a quick-consumerist-culture equivalent of Sylvia Plath sticking her head in an oven) after trying to give up alcohol. David Foster Wallace ended up committing suicide after trying to give up antidepressants.

And, okay, for this part to make sense I need to mention that the ending of *Infinite Jest* is *really mysterious*. I spent a little while searching the Internet trying to find a good explanation, and I found exactly one – one person who was able to put all the clues together and actually come up with a narrative that made sense to me. That was this blog post by Aaron Swartz. The only good online explanation of a book about a depressed Boston technical genius who commits suicide, a book about an attempt to spread banned information and the government's attempt to crack down on it – was written by Aaron Swartz. This is so *Infinite Jest*-ish that it is impossible for me not to think of it as an appendix to the book.

When I reviewed Illuminatus, I said it was a good as a book, but genius as a carefully designed machine to alter thought-processes. I will say the same about *Infinite Jest*, although having only read it once I am not confident that I have absorbed anything like a full dose of its medication.

But essentially, I think it's a book about people seeking meaningless entertainment in a way that doesn't truly involve interpersonal communication, in the form of a book that seems like meaningless entertainment but which is so complicated and so hard to understand that the attempt eventually produces true interpersonal communication with the author about this idea, completing a complicated but only partially self-referential loop.

Infinite Jest on Amazon.

(I apologize for the length of this blog post. It was aimed at people who have read or are considering reading *Infinite Jest*, which should be a group unusually tolerant of verbosity.)

The death of wages is sin

Federico gives a 6 point plan to cure youth unemployment. It is less complicated and revolutionary than his usual fare; he suggests policies like abolishing the minimum wage and slashing labor regulations. I expect it would work exactly as well as he thinks it would.

After all, minimum wage cuts the bottom out of the labor market. Everyone who would otherwise be working in jobs worth less than $7.25/hour suddenly becomes unemployed. This seems like a bad thing. People making $6/hour seems better than people not being employed at all, right?

Back when there were communists around, some of them would fight against minimum wage laws or occupational safety regulations. Their theory was that these would dull the pain just enough to make workers hate their bosses less and prevent revolution, but not enough to matter. The medical analogy would be a patient who comes in with bone pain, receives a painkiller that pushes the pain back below the threshold of “annoying enough to make me visit a doctor”, and never bothers seeking further medical treatment on what turns out to be bone cancer.

I admire the communists for their sheer Xanatosishness, but I don’t know how kindly historical hindsight has treated their strategy. On the one hand, they were dead right that better working conditions dampened interest in communist revolution. On the other hand, it seems relevant that a communist revolution would’ve been horrible, whereas a series of progressively stronger labor regulations actually achieved far more than the communists would have reasonably expected. So this sort of gambit seems potentially very risky.

But this is how I would question whether people making $6/hour or $3/hour or whatever is obviously better than their not being employed at all.

[Before you continue, read this Mother Jones article (h/t: commenter "Nestor") to calibrate your notions of how bad jobs can be for the rest of the article.]

There are probably a lot of people whose labor just isn't worth $7.25/hour. There are probably a lot of people whose labor just isn't worth $3/hour.

As technology continues to advance, I expect the number of these people to increase. I have been accused of the Luddite fallacy for this and I accept the challenge that the historical data present. But there's also this *reductio ad absurdum* where we can manufacture androids exactly as smart as humans in every way for $1. In this world, it seems obvious that all companies would buy androids (who work for free) and fire all their human workers, meaning an end to human employment.

So what's the difference between the past, when technological advances have never caused long-term unemployment, and the android-world, where it does? My guess is that in the past, there have always been areas to shunt the displaced human workers to: maybe machines can manufacture cars, but they can't drive taxis; maybe machines can sew textiles, but they can't predict fashion trends. Technological employment will become a problem only when machines can do *everything* better than humans, which won't be until after the Singularity, by which time we will have much bigger problems to worry about.

Except that's not really true. There may come a time when machines can do most blue-collar jobs better than humans even if they haven’t mastered the white-collar ones yet. And shunting former blue-collar employees to white-collar jobs seems like a hard problem. I don’t even think the hard problem is IQ, I think it’s some sort of meta-education which is complicated enough that society hasn’t figured out how to train it yet. No doubt some blue-collar people will be able to adapt to white-collar jobs, and other people won’t. As tech level rises and we approach the android scenario, the number of people who can’t adapt gets larger and larger.

Suppose we do what Federico wants. We promote full unemployment. Well then, these growing masses of people aren’t going to be unemployed. They’re going to be underemployed at $3/hour or something like that.

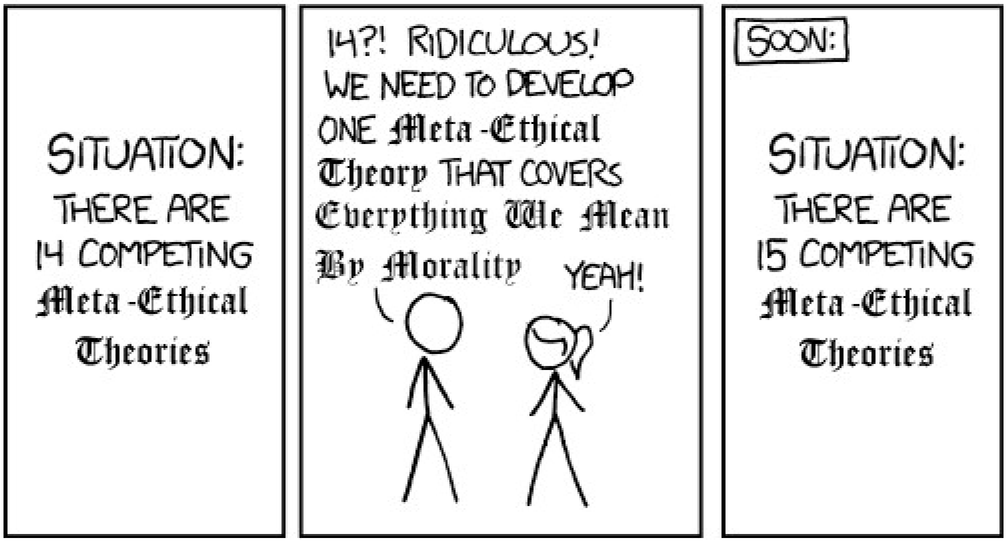
The minimum wage is sometimes called “the living wage”, and there are both lots of sob stories about how it’s impossible to support yourself on the minimum wage, and lots of counter-sob-stories by people who claim it’s totally easy as long as you don’t blow it all on alcohol and expensive hookers. I don’t know enough to have a strong opinion but my guess is it could go either way depending on circumstance. But I *am* pretty sure $3/hour is not a living wage. $3/hour either necessitates you to work 20 hour days, or actively drains your income because having a job is expensive (commuting costs, professional clothing costs) but people refuse to give you charity if there are $3/hour jobs available and you haven’t taken them.

On the nationwide scale, which is less dystopian? One in which half the population is unemployed and living off government benefits? Or one in which half the population is working 20 hour days at $3/hour jobs like the ones in that Mother Jones article and still struggling to support themselves?

The former situation seems very likely to evolve into a Basic Income Guarantee, about which I have written before in a very similar context and which seem like a proper end state for the economy which may even be preferable to our current situation in many ways (and of course after a basic income guarantee is in place there'll be a much stronger argument for eliminating labor regulations) But the latter situation seems like a disaster, and worse a *stable* disaster that no one has any incentive to make less disastrous.

This strikes me as the strongest argument for the minimum wage and other job-killing labor regulations: that they are turning otherwise-miserably-employed people into unemployed welfare recipients. "Too many people are unemployed and receiving welfare" seems more like a problem society will actually try to solve than "too many people are miserably employed", and maybe the solution will actually do us some good.

My Thoughts On The First 40 Pages Of Alastair MacIntyre’s “After Virtue”, Expressed In The Form Of A Slightly Modified XKCD Comic


SITUATION: THERE ARE 14 COMPETING Meta-Ethical Theories
14?! RIDICULOUS! WE NEED TO DEVELOP ONE Meta-Ethical Theory THAT COVERS Everything We Mean By Morality
YEAH!
SOON:
SITUATION: THERE ARE 15 COMPETING Meta-Ethical Theories

Thank you for doing something ambiguously between smoking and not smoking

“Funge” is a funny word. It refers to the thing which fungible things are able to do, sort of along the lines of what an extra unit of a good is going to replace. I don’t think it’s a real word and I’ve only heard it used by people connected to the Center For Applied Rationality. This is too bad, as it prevents everyone else from understanding, let along generating, important sentences. Like “Be careful what you’re funging against.”

Maybe this is why so many people are so careless what they’re funging against. Consider our recent discussion of the minimum wage. The minimum wage means no one has to work for below minimum wage. Its desirability depends a lot on whether below-minimum-wage funges against above-minimum wage jobs or against unemployment. That is, if we ban 100 below-minimum wage jobs, do we get 100 above-minimum-wage jobs, 100 more unemployed people, or a mixture of both?

This was also part of the thrust of my argument about drone warfare – it’s not funging against peace, it’s funging against much worse types of warfare. The same piece cited the status quo bias and indeed these two ideas are probably related.

This, combined with a complicated regulatory environment and sheer bad luck, seems like the best explanation for the trials (both metaphorical and literal) of e-cigarettes.

E-cigarettes (the “e” is for electronic) are pseudo-cigarettes that contains nicotine without tobacco. They don’t smell bad, they don’t produce secondhand smoke, and they don’t cause cancer. They are strictly better than regular cigarettes in every way. The governments of several countries are doing their best to ban them.

The governments’ position is that they are a stealth attempt to trick people who have successfully avoided regular cigarettes into smoking anyway. There’s some merit to this. Some of them have nice fruity flavors that might appeal to children. And because they don’t produce smoke, they’re legal to use indoors, where some people might not be allowed to use the real thing. There might be this tiny contingent of non-smokers who were just waiting for a flavorful and indoor-useable way to get the addictive expensive chemical they have no reason to want.

And yes, this would be bad. Nicotine is addictive no matter how you get it. There are some claims – and I don’t yet know how seriously to take them – that the other chemicals in tobacco inhibit monoamine oxidase which further perverts dopamine levels and makes cigarette smoking more addictive than nicotine alone would be, but this is different from saying nicotine isn’t addictive at all. Even if nicotine has few ill effects – and in fact this seems to be the case – there is a strong economic and convenience-based argument for not getting addicted to it.

I should clarify that “few ill effects” claim. A massive overdose of nicotine can kill you (so can a massive overdose of caffeine, Tylenol, or vitamins). Nicotine is a stimulant which raises your heart rate and blood pressure a bit (so is caffeine). It may increase the risk of diabetes, but it may treat cognitive impairment. Overall, it seems to have a complicated mix of minor bad and minor good effects, about the same as anything else in health. And like everything else in health, tomorrow three labs will come out with studies proving it causes cancer, and a fourth will come out with a study proving it prevents cancer, and one of them may even be right.

There are some studies that show that e-cigarettes have “toxic additives”, but these seem to be in ridiculously tiny trace amounts, don’t seem to make it into the vapor or the body of the user, and the entire problem could be solved by regulation anyway if anyone had a desire to regulate them. This entire issue struck me as a red herring and I bet you can buy fish at any market in the country with more toxic additives than the worst e-cigarette on the market.

So let’s accept that using e-cigarettes will get you addicted and set you back a lot of money and otherwise be annoying but probably not deadlier than anything else you do on a daily basis. What then?

Well, in that case, it’s worse than not smoking but much much better than smoking. And whether or not their existence is a good thing depends on what they funge against. Do they funge against smoking tobacco or not smoking at all?

I would have liked to get the CDC’s opinion, but their webpage on the issue is missing and from commentary I gather it didn’t have the information I wanted anyway. But I did find this:

> A June 2011 national study conducted under the supervision of ECH Research of Cincinnati, in conjunction with Opinionnaire, surveyed more than 200 smoker households that use electronic tobacco products and found that 99% of e-cigarette users are either

> current or past users of multiple forms of tobacco. Approximately 70% of survey respondents said they intended to quit smoking before starting e-cigarettes.

I can't find the original or even so much as a description of what "ECH" stands for. A sketchy online survey claims that 70% of e-cigarette smokers were former smokers.
So let's just say "probably some high number". This seems quite plausible to me. How many non-smokers think "You know, I want a product with all of the addictiveness and expense of cigarettes, but none of the coolness? In fact, I want to look like a chronically uncool recovering addict inexplicably smoking a glowstick."

That same Etter and Bullen paper says that 96% of the ex-smokers said the e-cigarette helped them quit or reduce smoking, and 79% felt they might relapse to smoking again if they didn't have e-cigarettes. Randomized trials seem to confirm this result, with the average smoker in the trial dropping from 19 cigarettes per day to 2 cigarettes per day after trying e-cigs.

It's not really surprising that e-cigarettes work. My current model of cigarette addiction is that it consists of the interaction between (1) nicotine, (2) smoking-associated behaviors which have become associated with the rush from nicotine over time and might have more complicated components like "oral fixation", and maybe (3) a contribution from MAOIs in the tobacco. Normal cigarettes have all three. E-cigarettes have (1) and (2). Nicotine patches have (1) only. Therefore, e-cigarettes should be more useful in quitting than nicotine patches, albeit not a perfect replacement for regular cigarettes. This does indeed seem to be what has been observed.

Some people argue that the effects of e-cigarettes haven't been perfectly studied, that they might be unsafe in some unclear way. And that as a smoking-cessation device, they're technically a medical device and therefore need to undergo as much study and regulation as any other medical device before being given to the public.

I know there's constantly a debate between the people who want to evaluate each new medical intervention for safetly and the people who want to use exciting new potentially life-saving technologies *now*, and I know that sometimes the former group do turn out to be right. Varenicline is a popular antismoking drug that was eventually discovered to drive people insane in various way and sometimes lead to suicide; although it is still used on people with both an extreme desire to quit smoking and impressive mental fortitude, it's nice that people paid careful attention to the side effects and didn't just give it out like candy.

But e-cigarettes are literally *the exact same thing as something that's given out to anyone who asks in convenience stores, except without the cancer*. To suddenly hold them to an extremely high standard of safety seems like a fallacy of fungibility.

The worst are the people – one of whom has so far appeared in every article I have read on the subject – who say that we should be careful because "Big Tobacco" is pushing them as a "solution" to the problem of declining cigarette sales. First of all this is just factually wrong; most e-cigarettes are made by alternative companies in direct competition with Big Tobacco. Second, if your reasoning strategy is identifying the Evil People and then minimizing their utility, you probably shouldn't be making public policy.

So I’m against banning e-cigarettes, and I’m even against things like taxing them or prohibiting their use in public places, on the grounds that the more smokers are encouraged to switch to e-cigarettes, the better. Like, if a public e-cigarette ban reduces the number of smokers who switch to e-cigarettes by 2%, you’ve just killed an extra 9000 people per year – about three 9-11 attacks, or twice the number of US soldiers who died in the Iraq War.

(this is why public health is about a hundred times more important than any other political issue, and even *tiny little marginal issues* in public health are more important to get right than, say, anything you will see people changing their profile pictures about on Facebook.)

So obvious conclusion is obvious and almost too boring to discuss. I got interested in e-cigarettes because a friend asked me whether he should start taking them *even though he didn’t smoke* as a way to get the cognitive enhancement effects of nicotine.

I guess that depends on what you value. I personally wouldn’t do this because I’m terrified of addiction. Then again, I avoid coffee and I drink like three glasses of alcohol per year in order to avoid addiction, which most people would consider sort of excessive. And I frequently take weird psychoactive Mexican herbs in order to achieve lucid dreams and would use LSD in a heartbeat if it were legal. So I guess my answer is that my feelings on the costs vs. benefits of various substances aren’t likely to generalize across the population.

A good pun is its own reword

Last night Buzzfeed showed me this so-called List of the Greatest Puns of All Time.

Yesterday an atheist blog published a list of Ten Most Median Things About Easter. They didn't make any claims that it was the best things about Easter, or the most interesting things about Easter, or pull a cracked.com and call it The Ten Most Mind-Bendingly Awesomely Epic Things About Easter You Never Would Have Expected That Will Blow Your Mind (Because You Never Learned About Them In School). It was just ten things about Easter someone wanted to mention on a blog. I appreciated the honesty.

My point being that the above list could have better been called a List of Median Puns. Good? A few of them. The greatest of all time? Don't make me laugh. Anyone can notice that two words sound alike and make a picture of one meaning, but a truly great pun requires something more.

Some puns achieve greatness by the sheer variety of meanings. Take the famous pun about deflowering: "She offered her honor, so I honored her offer, and all night I was on 'er and off 'er." It plays on "honor" and "offer" three times in three separate ways. Even so it is beaten by Samuel Johnson, going meta: "If I were punished for every pun I shed, there would not be a puny shed left of my punnish head". That's a *four-way pun* (I'm going to give him the benefit of the doubt here and assume "puny shed" made sense in 18th century English).

Others are multiple puns in a different way: they are phrases of a few words, all of which get punned at the same time. For example, a magazine described aging as "the transition from frequenting hip joints to breaking them”. Likewise the claim that diarrhea is hereditary because “it runs in your jeans”. Or the famous one about the scarecrow who won a Nobel Prize because he was “outstanding in his field”. Or how “TV is called a medium because it is neither rare nor well done”.

(Fiddlemath in the comments adds the one about the father who left his cattle ranch to his boys; they named it Focus Ranch, because “focus is where the sun’s rays meet.”)

Still others are anti-puns, jokes that are hilarious precisely because they subvert a pun we’re expecting or switch puns midstream. “My wife was a strong believer in astrology. She was a Cancer, which makes it pretty ironic how she died. That’s right – chopped to pieces by a giant crab.” Mary in the comments tells the story of Ben Johnson, who claimed he could make a pun on any subject. When someone asked for a pun about the Queen, Johnson protested: “But the Queen is not a subject”.

And then there’s this absolute jewel:

> A: Where do you weigh a whale?
> B: Where?
> A: At a whale-weigh station.
> B: (groan)
> A: And where do you weigh a pie?
> B: At a pie-w…huh, wait, no. I give up. How?
> A: ♬ Soooooomewherrrrre oooooover the raaaaainbow… ♬

(think about it!)

Some puns are great because they start a chain reaction among parts of speech that shifts around entire sentence structures. Take the story of two nuns traveling in

Transylvania who suddenly encounter Dracula. One shouts to the other: “Quick, show him your cross!” The second says “Oi, get off the bloody road, you lunatic!” Similar is Groucho Marx’s famous line: “Time flies like an arrow; fruit flies like a banana.” (It took me about five years to get that one)

Some puns are great through sheer endurance. You know Gandhi, right? And you know how he walked barefoot everywhere, making his feet extremely hardy? And you know how he fasted all the time, making him thin and frail and also giving him terrible breath? Well, that made him a…wait for it…super-calloused fragile mystic hexed by halitosis.

And this isn’t even the longest one I know! That honor goes to the story of famed Arabic conductor Mustafa al-Badr. Badr was charged with leading a New York Philharmonic production of Beethoven’s Ninth Symphony. Unfortunately, this symphony has bass parts at the very beginning and the very end, but none in between, making the bassists very bored. During practice they used to sneak out and head to the local pub while the rest of the orchestra did their thing, making it back a few short minutes before they had to start up again for the end of the piece. When the day of the concert came around, they decided to try the same strategy. To make sure they had enough time, they took a big piece of rope and tied together Badr’s music, forcing him to unravel each page before moving on and slowing his conducting considerably. It was a good plan, but when they made it to the pub they got really, *really* drunk, and they were visibly incapacitated by the time they made it back to the concert hall. Just as the bass part began, two of the bassists, totally sloshed, fell unconscious. A concert-goer live-blogged the conductor’s travails as “BADR’S UP, BOTTOM OF THE NINTH, SCORE TIED, BASSES LOADED WITH TWO MEN OUT.”

Still other puns aren’t so long in themselves, but they require a lot of background knowledge to work. Anyone who’s studied Latin knows that most words *decline* – that is, they change their form based on what role they’re serving in the sentence. A token English example would be “Me” (object) vs. “I” (subject), but where English has very few of these words, in Latin virtually all words except for certain numbers work this way. One of those numbers is six (Latin: *sex*). So the joke in my high school Latin class used to be: “A good Latin student never declines sex”.

The pun that gets my nomination for greatest pun of all time fulfills several of these categories simultaneously. It works on multiple words, it alters grammatical structure, it’s multilingual, and it requires esoteric background knowledge – and it’s gratifyingly historical.

According to the legend, Sir Charles Napier, British general in India, reported his conquest of the Indian province of Sindh (which was against direct orders to stay put) with the single-word telegraph “Peccavi” – Latin for “I have sinned” but in this case carrying the double meaning of “I have Sindh”.

Too bad it probably isn’t true. It’s a good pun anyway.

Against Anton-Wilsonism

Back when I was in college, I loved stuff by Robert Anton Wilson. In case you're not familiar with him, he was a writer and occultist who recorded all sorts of interesting things from mystical traditions all over the world. It would be as nothing for him to relate something of Crowley's to an old Sufi parable which was really a metaphor for something the Buddha said about quantum physics.

Although he is most famous for his fiction books like *Illuminatus*, he also wrote a lot of non-fiction. On the one hand, it was the ultimate insight porn, with a new seemingly-revelatory epigram from a new tradition on every page. On the other, it was filled with very vague nod-and-a-wink promises that if you genuinely understood it you would break into a new level of understanding in which you would stand taller, have a more melodious voice, and finally be able to get that one cute girl/guy to pay attention to you. It was seductive and I was successfully seduced by it.

This is in no way a complaint against mysticism. I think it's quite possible that there are forms of mysticism which successfully fulfill the promises Wilson made both in terms of insight-into-reality and improved-life-success. Meditation probably does. Yoga (the real type, not the contort-your-body-for-exercise type) might. If your goals are simple, you can certainly get some quick mental rearrangement (not necessarily of the positive variety) by doing drugs, or improve your social presence through something like the Alexander Technique, both of which seem in the spirit of Anton-Wilsonism.

Better I should compare it to my interest in physics as a high schooler. This interest took the following form: I would read Scientific American articles about bosons, then go around saying “Did you know there are several types of bosons, with mass of such-and-such?” Maybe I would vaguely long to go to CERN and see the Large Hadron Collider (a wish I eventually fulfilled). I read some biographies of famous physicists and I could point out the position in the sky of several leading black hole candidates.

There’s nothing necessarily wrong with this; it was a better use of my high school days than taking drugs or watching reality shows. But the problem is I thought I was Learning Science. I had this idea that Science was great and admirable, and that people who knew Science could predict eclipses and split the atom, and I was going to be one of those people, and all I had to do was read a few more Stephen Hawking books and then when I became an adult I could receive my Certificate of Scienciness and start splitting atoms. But that wasn’t remotely how it works and although I could charitably be said to be learning *about* science this has very little in common with *learning science*.

(this is why I get a twinge of worry about how vocal identification with Science has become a badge of honor in geek culture, by the way)

There are certain fields where it’s really obvious to everyone that learning about the field is different from learning the field. There are probably historians of music who have never picked up an instrument, and they don’t fancy themselves musicians. And political scientists don’t delude themselves into thinking they would make great politicians.

Mysticism is not one of these fields (rationality isn’t either, but that’s a different blog article). Because so much of mysticism revolves around the idea of the gnosis, a specific kind of knowledge, it’s easy to mistake knowledge of mysticism for the knowledge that mysticism tries to produce. This means Robert Anton Wilson and his ilk can cause at least three different types of failure.

First, as I mentioned before, they provide a false sense of reward. If you are actually enlightened, it's pretty hard to demonstrate this to someone short of shooting them with a chi bolt from your third eye. On the other hand, if you're knowledgeable about mysticism, it's really easy to demonstrate this. I would talk to my mysticism-interested friends, and we'd be like "Oh, this reminds me of something the third Zen patriarch said about this subject", and then we'd feel all cool and mystically advanced. And just as slacktivism sates your desire to do good without requiring any hard work, so this sates the desire to associate with the neat cause of mysticism without requiring any hard work beyond reading lots of books, which for a certain kind of person is the sort of thing they'd do anyway. So this sort of scholarship trades off against actual results.

Second, they encourage you to think in terms of conspiracy. Quick, what do Buddhism and the Illuminati have in common? Okay, fine, nothing. So how come they're both part of Anton Wilson's repertoire? My guess is that mystics say weird things all the time like "Ultimate holy reality is emptiness" or "Spend a while thinking about the sound of one hand clapping", and people view this as a sort of *puzzle*. Like mystical traditions are jealous guardians of some kind of secret knowledge, and they've let slip a couple of clues, and it's our *Da Vinci Code*-esque job to piece together what this secret knowledge is and pierce through the conspiracy. Which I think, despite the fact that many ancient mystery cults did jealously guard secret knowledge, is totally the wrong way of looking at things.

Instead of mystics talking about one hand clapping, let's go back to my physics example. Physicists *also* frequently emit bizarre and puzzling statements like "Faster objects have more mass than slower objects", or "Time and space are really just aspects of a more generalized spacetime." If you try to understand these statements – really understand them on a gut level – by watching Carl Sagan specials, you will fail. One hypothesis here is that Carl Sagan is part of a conspiracy, where he will tantalizingly tell you a few pieces of the clues but guards the really juicy bits for himself, and you need to piece together the real thing using Sagan, 12th-century alchemical texts, the Windows source code, and a pattern of moles on Neil Tyson DeGrasse's left cheek. Another hypothesis is that this is the sort of thing which no amount of *learning about science* will be able to illuminate, but which is relatively straightforward once you *learn science* and can solve the equations that describe them – and that this knowledge cannot be translated into terms people who haven't learned science can understand in any way more satisfying than the old "Well, imagine a really taut sheet with some objects upon it…"

Third and most important, they promote *dabbling*, which is fatal. "Why limit yourself to one tradition when you can take insights from all the different traditions and invent your own tradition that combines the best wisdom of them all as well as your own special touch?"

Well, because until you know at least a little of what you're doing, you don't know how to do it. This is a form of Chesterton's Fence, except that instead of a fence in a field it's like a twelve-dimensional pulsing image in an unexplored region of Dimension Q'qaar and you have no idea what it is or what it does or where you're going. Oooh, I know, let's remove the part of Daoism where you don't drink whiskey and hire hookers every night! That won’t change the underlying state of mental peace *at all*.

This brings up the related issue of Schelling fences: once you let yourself change things, do you really trust yourself to remove only the chaff and not the parts that are annoying or hard or inconvenient? Yes, 3000 year old forms of practice are inconvenient and laden with superstitious baggage, but Bringing Buddhism To The West was like the state pastime of California for several interesting decades back in the mid-20th century. Surely you could latch onto one of the adaptations created then instead of trying to invent your own?

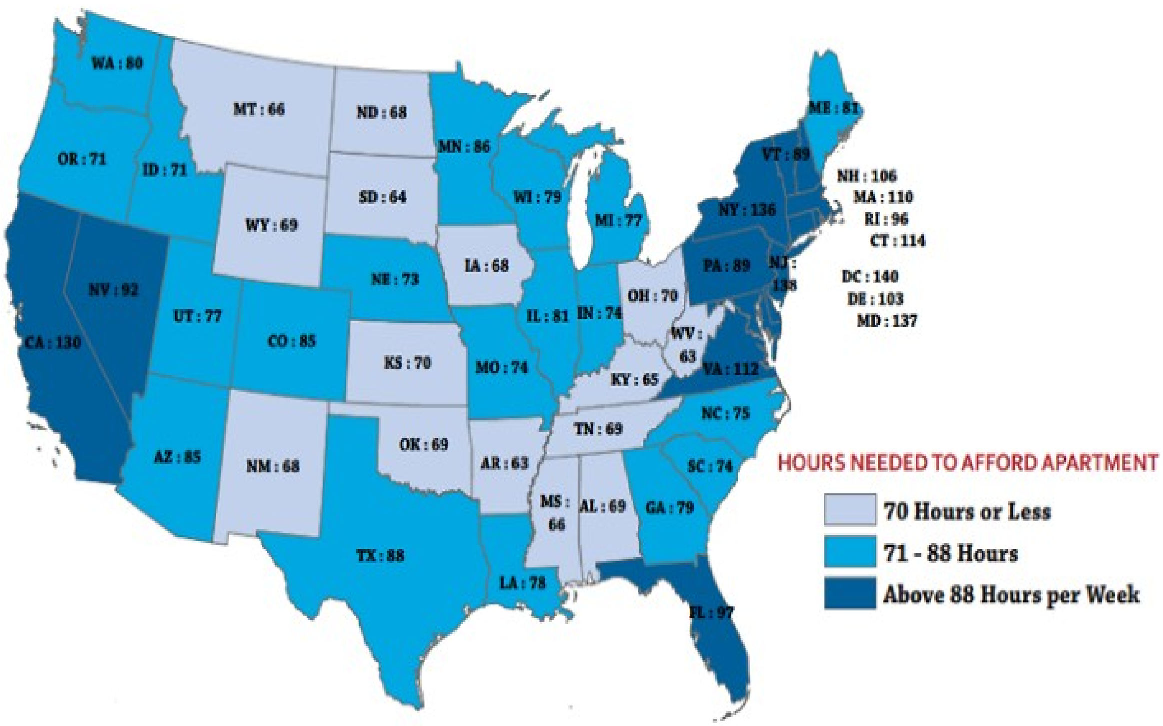
But the most pernicious issue here is that – at least if my college age self is typical – you will end up spending so much time refining and polishing and admiring your new collection of spiritual beliefs that you never actually bother putting them into practice – or if you do, you will change them every few weeks and never get the solid consistent base you need to be good at them.

If I get into mysticism again – and I think I should, it seems like one of the highest-value areas for me – I am going to force myself not to read any books about mysticism except extremely sparse how-to instruction manuals. One at a time. Without even so much as glancing at Step Eight before I’ve finished Step Seven.

Lies, Damned Lies, And Facebook (Part 1 of ∞)

I spend so much time arguing with people about the graphics they post on Facebook that I figured I should at least make a blog post out of it so I can pretend it’s productive.

Here’s an image I got from a few places earlier this week. The title was something like “Hours Working Per Week At Minimum Wage Needed To Afford A Two-Bedroom Apartment In Different States”, and it was usually associated with some text about how it was outrageous that the minimum wage isn’t nearly enough to live on:


WA : 80
MT : 66
ND : 68
MN : 86
ME : 81
OR : 71
ID : 71
VT : 89
NH : 106
MA : 110
RI : 96
CT : 114
WI : 79
SD : 64
WY : 69
MI : 77
NY : 136
IA : 68
PA : 89
NJ : 138
NE : 73
DC : 140
DE : 103
MD : 137
NV : 92
UT : 77
CO : 85
IL : 81
IN : 74
OH : 70
WV : 63
CA : 130
KS : 70
MO : 74
KY : 65
VA : 112
NC : 75
TN : 69
OK : 69
AR : 63
SC : 74
AZ : 85
NM : 68
MS : 66
AL : 69
GA : 79
TX : 88
LA : 78
FL : 97
HOURS NEEDED TO AFFORD APARTMENT
70 Hours or Less
71 - 88 Hours
Above 88 Hours per Week

At first glance – and this is how everyone I talked to interpreted it – this seems to support the “minimum wage is unliveable” hypothesis in a big way. In my home state of California, for example, it looks like a minimum wage earner needs to work 130 hours a week just to be able to pay for a place to live. That translates into working nineteen hour days seven days a week. And if you need that ridiculous schedule *just to pay rent*, it doesn't seem to leave a lot of room left over for anything else.

California might be an unfair example because it's particularly high. The state I'm visiting right now, Utah, is a bit more representative, at 77 hours/week. But even this requires eleven hours a day, seven days a week. So people's outrage seems justified.

What's the catch?

The first catch is that this has nothing to do with the minimum wage. The federal minimum wage sets an effective floor for state minimum wage; many states exceed the federal number but none are allowed to go below it. The states that have the best minimum wage apartment affordability – Arkansas, South Dakota, and West Virginia – all have this minimum permissible minimum wage. On the other hand, the states with the highest minimum wage in the country – Vermont, Washington, and Nevada – have exceptionally poor apartment affordability. I don't have numbers to put into SPSS, but a very quick eyeballing suggests apartment affordability as measured by this map and minimum wage are actually *anti*-correlated, probably because high minimum wage implies leftist politics implies dense urban population implies costly housing.

The second catch is related to the first catch: there is no way raising minimum wage could solve this problem. For example, suppose we decided that it was unfair to make people work more than the standard forty hour workweek. What level could we set minimum wage in order to achieve this noble-sounding goal? In California, we would have to multiply the minimum wage by 3.25x, making it $26/hour. I think even most leftists would start to worry that might cause certain problems down the line.

The third and most important catch is that these numbers don’t mean what you think they mean and probably don’t mean anything at all

I first noticed this when I tried to calculate the price of an apartment in California from this data. California minimum wage is $8/hour, so a quick $8/hour * 130 hour weeks needed * 4 weeks/month = average apartment in California costs $4160/month. Renting apartments in California is horrible, but not *that* horrible.

So I Googled this until I finally found [an article in the New York Times]() where this image had previously appeared, which unlike every instantiation of the image I have seen on Facebook explains the methodology. The numbers are not about how many hours are needed to *pay for* rent, they’re about how many hours are needed to *afford* rent, where afford is arbitrarily defined as “be able to pay for rent using no more than 30% of your income”. This brings the cost of a California apartment *in the sense everyone is naturally interpreting the picture as meaning* back to a slightly-less-unreasonable $1250/month and brings the number of hours per week our hypothetical minimum wage laborer needs to work to *pay for* rent down to 39 – still worrying, but markedly less so (her Utah counterpart only needs 23.1, which actually sounds doable and liveable).

But we’re still not done here. The graphic specifies that we’re affording a two-bedroom apartment. The best reason to demand a two bedroom apartment is that you are, in fact, two people. Suppose we’re talking about a couple, both of whom earn minimum wage. Now each partner in California only has to work 19.5 hours per week. Each partner in Utah only needs 11.5. This is starting to sound pretty good.

Can we bring these numbers down further? We might note that even if the average California apartment requires 19.5 hours/week, a minimum wage earner might not be going for the average apartment. They might be after something more modest. How much does a modest apartment cost?

I briefly wondered whether the Internet would be able to tell me which Californian county had the exact median land value for California. Then I remembered this is the 21st century and went straight to Wikipedia's list of California locations by per capita income, which ought to track land value well enough. The exact median is Amador, but since I don't know where that is I chose Sacramento, which is just next to the median, as my experiment. I asked http://sacramento.apartmenthomeliving.com/ to tell me the rent of two bedroom apartments in Sacramento. Its "sort by price" feature is hopelessly buggy, so I wasn't able to find an exact median, but I am prepared to believe it is about the $1040 the graphic suggests. And yet it is easy enough to find decent 2 bedroom apartments for as little as $650. If we pretend this case is typical, we can declare that a cheap (but liveable) apartment costs about 62.5% of an average apartment. That means each partner in the Californian couple only has to work 12.2 hours/week, and each partner in Utah only has to work 7.2 hours/week.

In fact, these are *still* overestimates for a bunch of reasons. Minimum wage earners are probably concentrated disproportionately in poorer areas of a state, so taking the exact median area of a state overestimates the affordability challenges minimum wage earners face. Most couples share a bed, so they're probably not after a two bedroom apartment and can look for a cheaper one bedroom apartment. And if they really need to, they can do what my roommates do, which is move into a larger house with more rooms and split up the rent even more. In my own living arrangement (two bedroom house where one bedroom is occupied by a couple and the other is occupied by a single person), each partner of the couple only has to work half as hard as the partners in the couple above.

But let’s ignore these additional factors and conclude with our 12.2 number for California. Note that this is *less than a tenth* of the number on the original graphic, and is probably a heck of a lot closer to what people think when they read what the graphic is trying to do.

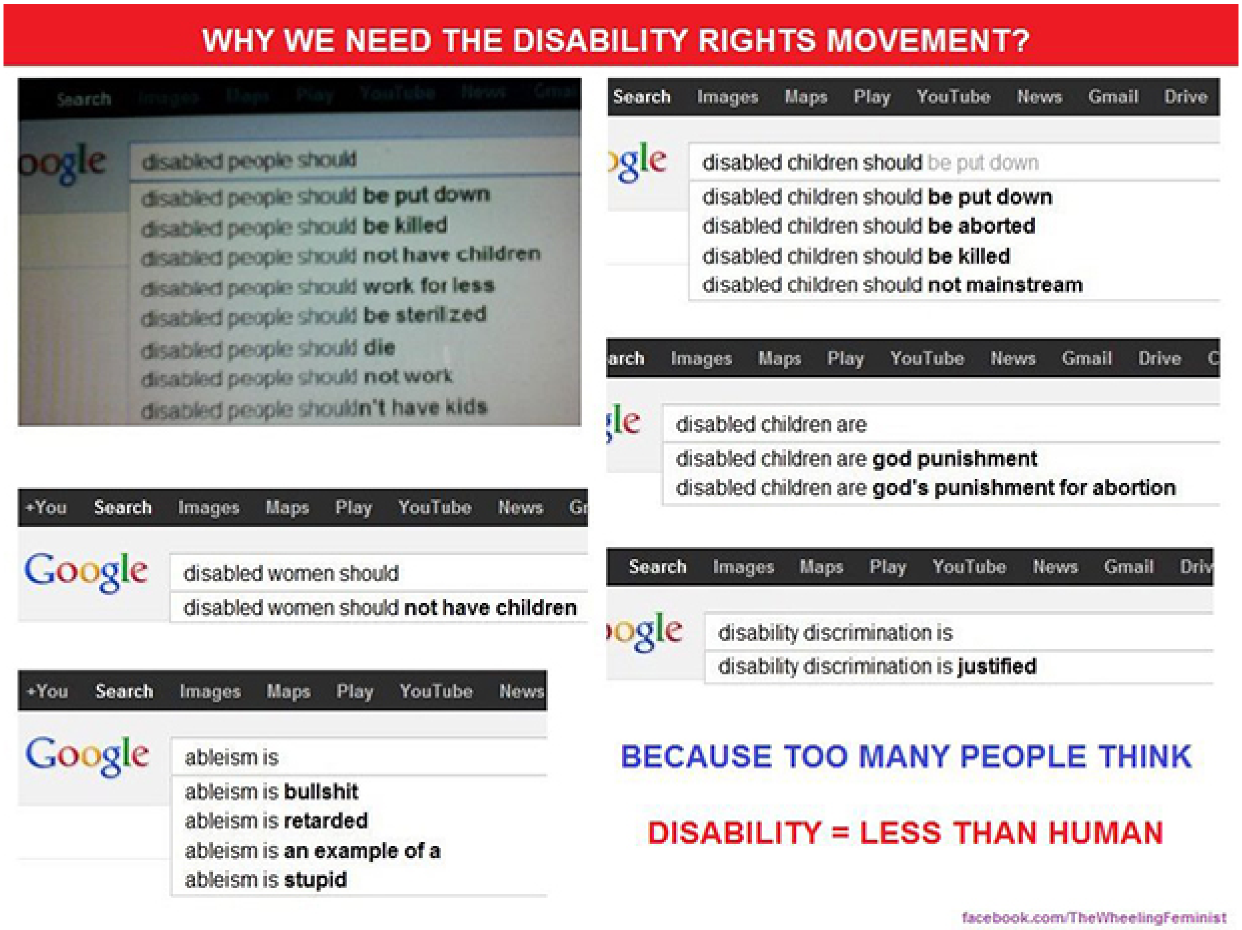
I don’t mean to trivialize the problems that minimum wage earners go through trying to pay rent, and certainly not the problems they would go through if they don’t work full-time, or are supporting a non-working family member. It’s just that this image has nothing to do with these problems and its numbers might as well be generated with random dice rolls for all the good they do anyone.

Lies, Damned Lies, And Facebook (Part 2 of ∞)

[TRIGGER WARNING: Discussion of violence against the disabled]

I decided to edit the last post to split it in half. Some of you who read this through Reader are probably getting it twice, so apologies for that.

The second Facebook meme to grab my attention recently was this:



WHY WE NEED THE DISABILITY RIGHTS MOVEMENT?

disabled people should
disabled people should **be put down**
disabled people should **be killed**
disabled people should **not have children**
disabled people should **work for less**
disabled people should **be sterilized**
disabled people should **die**
disabled people should **not work**
disabled people should**n't have kids**

disabled children should be put down
disabled children should **be put down**
disabled children should **be aborted**
disabled children should **be killed**
disabled children should **not mainstream**

disabled children are
disabled children are **god punishment**
disabled children are **god's punishment for abortion**

disabled women should
disabled women should **not have children**

disability discrimination is
disability discrimination is **justified**

ableism is
ableism is **bullshit**
ableism is **retarded**
ableism is **an example of a**
ableism is **stupid**

BECAUSE TOO MANY PEOPLE THINK

DISABILITY = LESS THAN HUMAN

facebook.com/TheWheelingFeminist

This image caught my ire because I have a pet peeve against strawmanning and the disability rights people seem to do it more often than most. My model of them all too often takes complicated issues like “Should we treat disabilities as medical conditions that we should try to research treatments for or as acceptable alternate ways of being?” or “Should we support noncoercive and irreproachably progressive genetics programs like Dor Yeshorim?” and insist that anyone who disagrees with them even a tiny bit just personally hates all disabled people and denies their right to exist and wants to kill them.

And having beliefs ought to mean you can make predictions, so in this case I predicted as soon as I saw this image that something was not quite kosher and that on further investigation it would turn out to be terribly misleading and that people aren’t really this horrible at all. And, well…

First of all, when I Google these words I get some very different results. For example, Googling “disabled people should…” brings up some of the same things they found, but also things like “disabled people should be treated equally”. I am not going to accuse the original author of lying because I know Google can sometimes be idiosyncratic, but at the very least it is less monolithic than the image presumes.

Second of all, even on the offending results, very few of them are actually making the positive case. For one thing, Google interprets “Should disabled people be killed?” interchangeably with “disabled people should be killed”, meaning that a lot of these results are people rhetorically asking “Should disabled people be killed?” and then answering “No, disabled people are as good as anyone else.” Others are quoting other people: for example, the top site to appear for me is an article titled Kill the disabled, they are unworthy , but its title seems to be paraphrasing the Nazi position on the issue and then goes on to say how wrong the Nazis were and how “I have been astounded at the capabilities of these so-called “disabled” people. They speak better English than many of my peers and the joy and spontaneity they experience through music clearly shows them to be more evolved than the average person. I know now that the words ‘special needs, disabled, handicapped and retarded’ need a thorough re-evaluation.”

Another lot are Christian sites, which put the “disabled people should be killed” position in the hands of some hypothetical atheist and then goes on to talk about How Christ Loves Everyone Alike. A few others are obvious trolls who then get knocked down by thirty or forty other people. We fill out the rest with mostly denunciations of politicians allegedly claiming disabled people should be killed, along with demands for the politician involved to resign.

I searched through the first thirty search results for “disabled people should be killed”, and I could not find *one single web page* which was unambiguously an essay in favor of the proposition. Thirteen were outrage porn railing against the fact that some other vague person or organization was even considering the question, and the other seventeen were a complicated mixture of things which admittedly did include a very few comments in favor although they all looked like they were written by thirteen year olds.

But there’s an even bigger problem here, which is that *Google loathes everyone in the world with the burning fury of a million suns*.

Type in “French people should…” and the single AutoComplete suggestion is “French people should die”. “Germans should” gets you “Germans should die” and “Germans should be exterminated”, as well as the equally-scary-for-different-reasons “Germans should have won”.

“Jews should” gets you “Jews should all die”, “Jews should leave America”, and the comparatively reasonable “Jews should apologize for killing Jesus.”

This isn’t just Google’s oppressive WASP user base hating foreigners and minorities either. “White people should…” brings up four options: “white people should be slaves”, “white people should die”, “white people should go back to Europe” and “white people should be killed”.

“Homosexuals should…” gives you “…should be killed”, but “straight people should…” gives “straight people should die” as well. Is there anyone who Google *doesn’t* think should die? Apparently yes! “Hitler should…” gives you “Hitler should have won the war” and “Hitler should have finished the job”.

Google AutoComplete has complicated political beliefs. “Obama should…” gives “Obama should die” and “Obama should be killed”, but “Bush should…” gives “Bush should be in jail.” Type in “rich people should” and you get “rich people should die”. Type in “poor people should”, and you get, unsurprisingly “poor people should die” (as well as “poor people should be sterilized”). Google thinks “Republicans should be ashamed of themselves” and “Republicans should be killed”, but also that “Democrats should be shot” and the somewhat nonsequitur “Democrats should be confiscated”. Christians should “be killed”, atheists should “shut up”, and Muslims should “be exterminated” (or at least “not be allowed in America”). Some of Google’s opinions are bizarre: it seems to feel very strongly (all its suggestions are variants of this same phrase) that “dead people should stay dead”.

Even the user herself doesn’t escape Google’s universal crushing misanthropy. Type in “I should…” (without a space), and your second option is “I shouldn’t be alive”. Nor does the species as a whole: type in “humans should” and number two is “humans should die” (number four is “humans should go extinct”.)

It seems kind of ironic that a company whose motto is “Don’t Be Evil” has a flagship product that believes virtually every demographic group should be murdered *en masse*. I don’t know if this is a weird quirk of the search algorithm or if pretty much everyone who makes should-statements on the Internet is recommending genocide.

But I hope that this problem gets solved before Google tries to build a recursively self-improving artificial intelligence.

Investment and Inefficient Charity

After rushing home from Utah I made it to a talk on efficient charity which St. Paul of Rational Altruist put together at the Berkeley Faculty Club.

(I should probably mention that from now on I will be preceding the names of leaders in the effective charity movement with the honorific "St", because people's social status and recognition ought to accurately track the level of effect they are having on the world. If through persistence and self-sacrifice someone manages to save the lives of a couple dozen people they never met, I don't think even the harshest *advocatus diaboli* could object to an impromptu sainting or two.)

First, St. Elie of GiveWell talked about his organization's efforts trying to research which charities did the most good and defended the idea of radical transparency – GiveWell's procedure of meticulously recording everything they do and criticizing themselves *ad infinitum* in an attempt to become better and more evidence-based.

Then Robin Hanson of Overcoming Bias got up and just started Robin Hansonning at everybody. First he gave a long list of things that people could do to improve the effectiveness of their charitable donations. Then he declared that since almost no one does any of these, people don't really care about charity, they're just trying to look good. Then he told the room – this beautiful room in the Faculty Club, full of sophisticated-looking charity donors who probably thought they were there to get a nice pat on the back – that they probably thought that just because they were attending an efficient charity talk they weren't like that, but that probabilistically there was excellent evidence that they were.

I have never seen a group of distinguished Berkeley faculty gain so sudden and intuitive an appreciation for the Athenians who decided to put Socrates to death. I spent the whole speech grinning like an idiot and probably scared Robin a little. And okay, some of that was because I woke up really early to get to the airport today and had become dangerously overtired and mentally imbalanced, but the rest of it was just that he sounds *exactly* like he does on his blog, he's a great speaker, and it was just really funny in a train-wreck sort of way to watch a whole room of innocent and basically decent people get Hansonned. The man is one of a kind and his complete and obviously deliberate imperviousness to normal social niceties needs to be declared a national treasure.

But he made some genuinely unsettling points.

One of his claims that generated the most controversy was that instead of donating money to charity, you should invest the money at compound interest, then donate it to charity later after your investment has paid off – preferably just before you die, since donating money after death is legally complicated. His argument, nice and simple, was that the real rate of return on investment has been higher than the growth rate for 3000 years and this pattern shows no signs of changing. If you donate the money today, your donation grows with the growth rate, but if you invest it, it grows with the interest rate. He gave his classic example of Benjamin Franklin, who put his relatively meager earnings into a trust fund to be paid out two hundred years later; when they did, the money had grown to $7 million. He said that the reason people didn't do this was that they wanted the social benefits of having given money away, which are unavailable if you wait until just before you die to do so.

And darn it, he was totally right. Not about the math – there are severe complications which I'll bring up later – but about the psychology. On even the most cursory self-examination, my mind totally recoils at the thought of donating everything I'm going to donate to charity in a single lump sum just before I die. It just gibbers "But…but…you need to be a good person before then!" I'm not saying you can't tear down Robin's substantive argument in a bunch of good mathematical ways. I'm saying his *ad hominem* argument about my motivations seems to be true regardless.

Then he started talking about how you should only ever donate to one charity – the most effective. I'd heard this one before and even written essays speaking in favor of it, but it's always been very hard for me and I've always chickened out. What Robin added was, once again, a psychological argument – that the reason this is so hard is that if charity is showing that you care, you want to show that you care about a lot of different things. Only donating to one charity robs you of opportunities to feel good when the many targets of your largesse come up and burdens you with scope insensitivity (my guess is that most people would feel more positive affect about someone who saved a thousand dogs and one cat than someone who saved two thousand dogs. The first person saved *two* things, the second person only saved one.) In retrospect this is absolutely true and my gibbering recoil at this problem isn't just Yet Another Cognitive Bias but just good old self-interest.

Now that I have identified what this pattern feels like, I can look back in my memory and notice more examples. Probably the worst is that an efficient charity group was actually slightly interested in having me interview to work with them back in February-ish and I declined because I already had a career plan – and one from which I could make lots of money to then donate. This may end up being the correct decision and in fact probably is, but I already know that's not why I did it. I did it because a cognitive paranoia I cultivated for pretty good reasons but can't turn off at will tells me that doing anything not directly measurable is just my brain lollygagging about and inventing clever stories about why it's so great, and I only get credit for direct obvious quantitative sacrifice. Working for an efficient charity organization, even if I was able to redirect other people's money in valuable ways that ended up outweighing the utility of making lots of money myself, wouldn't be enough to overcome my brain's suspicion that I wasn't actually doing any good after all.

Thankfully, Robin's last point was that the most effective thing to do is to stop beating yourself up and be exactly as irrational as is necessary to convince your mind to go along with the whole "efficient charity thing" instead of freaking out and giving up in disgust. I have already measured about how irrational that is and I don't see too much reason to change my decision now. Still, no sainthood for me just yet.

So let's get to the fun part. How do we debunk Robin's assertion that we should invest charitable givings and donate them only at the end of our lives?

St. Elie discussed this for a little while at the talk and gave what I thought was an unexpectedly good answer. He said that the world is getting better so quickly that we are running out of good to be done. After the initial burst of astonishment he explained: in the 1960s, the most cost-effective charity was childhood vaccinations, but now so many people have donated to this cause that 80% of children are vaccinated and the remainder are unreachable for really good reasons (like they're in violent tribal areas of Afghanistan or something) and not just because no one wants to pay for them. In the 1960s, iodizing salt might have been the highest-utility intervention, but now most of the low-iodine areas have been identified and corrected. While there is still much to be done, we have run out of interventions *quite* as easy and cost-effective as those. And one day, God willing, we will end malaria and maybe we will never see a charity as effective as the Against Malaria Fund again.

St. Paul lists several other reasons on his blog. First of all, the US' tax deduction laws favor spreading your donations out among as many years as possible. Second, you might become a worse person in fifty years and decide you don't want to give your massive accumulated savings to the poor after all. He also lists a few other arguments, none of which in my opinion have quite the same power as those two.

Most of the people at the meeting today were not radical singularitarians – not all saints can be prophets. But if you believe, as I do, that we're within about a century of a technological singularity of some sort or other, three new considerations come into play. First, affecting the singularity – either bringing it forward, pushing it backwards, or changing its nature – may be a unique and fantastically high-leverage charity target which will not be available (or may be less available) fifty years from now. Second, if a singularity goes wrong and kills us all, we lose our opportunity to donate to charity later. Third, if a singularity goes *right*, it's a pretty good bet that people won't need malaria nets anymore.

Suppose I will die at age 78 – ie fifty years from now. And suppose I want to donate $10,000 to charity. If I decide against Robin's strategy, I donate $10,000 today, perhaps to

Ethiopia. Ethiopia has a growth rate of about 7%, but let’s assume it can’t keep that up over the next 50 years and its average is 5% (this is still quite high). In 2063 Ethiopia ends up with $115,000 extra or so, which I round off to $150,000 because it can be enjoyed by a couple of generations rather than simply appearing at the end of the period.

Suppose I decide in favor of Robin’s strategy. A couple of investment sites say to expect 7% rate of return, so I can expect about $300,000 (all these numbers are adjusted for inflation, I think).

(if these numbers are right, then using the “efficacy of charity declines as things get better over time” argument means you have to believe in a 50%-in-fifty-years decline in charitable efficacy, which seems like a pretty high bar)

Okay. Now what if the world ends (or progresses beyond the need for charity) in 2062 – forty-nine years from now? In that case, giving now leaves Ethiopia with that $150,000, and waiting till later leaves them with nothing.

I don’t know enough math to do the integral properly, but I can do it at different points and then sum it up. If the world ends in ten years, saving loses $15,000. If the world ends in 20 years, saving loses $25,000. In 30, $45,000. In 40, $70,000. At 49.999, $150,000. If there’s an equal chance of the world ending at any one of those times, on average the world ending before you can donate loses you $60,000. But if the world doesn’t end, saving gains you an extra $150,000. So unless you think the world is more than 70% certain to end before you die, saving like Robin suggests is the best option.

And okay, there are so many problems with this analysis I don’t even know where to start (and I bet commenters will point out ones I missed). I hope someone can do some more rigorous math on this question. But the Fermi calculation gives me the opposite result from the one I was expecting and this is very awkward and I was totally intending to close up this essay with “and therefore, math tells us investing charitable donations is clearly a bad idea.” Instead I’m just going to keep recoiling and gibbering.

Poor Folks Do Smile…For Now

I got the opportunity to talk to GMU professor and futurist Robin Hanson today, which I immediately seized upon to iron out the few disagreements I still have with someone so brilliant. The most pressing of these is his four year old post Poor Folks Do Smile, in which he envisions a grim Malthusian future of slavery and privation for humanity and then soundly endorses it. As he puts it:

> Our robot descendants might actually be forced to slave near day and night, not to feed kids but just to pay for their body rent, their feed-stocks, their net connection, etc. Even so they'd be mostly happy.

Robin seems to be a total utilitarian who has no objections to the Repugnant Conclusion. It's a consistent position on its own (though distasteful to me), and taken at face value it has something to recommend it. I've been to some horrible places like Haiti. The Haitians have it tough, but they still sing and dance, they still love each other, they still have hopes and dreams. If the far future is Haiti with better sanitation, it wouldn't necessarily be the worst thing in the world.

But Robin has a slightly higher bar here. He believes that the near future promises advances in the uploading of human minds to computers, creating cyber-organisms he calls "ems" for "emulated humans". Ems will have many advantages over biologicals – less need for space and resources, possible elimination of biological need for sleep, and can be copied-pasted at will. A future of zillions of Malthusian ems competing for hardware and computing power is a little different from zillions of biological humans competing for land and food.

So here is my dialogue with Robin as I remember it. I didn’t take notes, so it’s probably a bit off, and I’m rewriting me being confused and ummming and errrrring and meandering for a while as me having perfectly flowing rational arguments with carefully considered examples. I think I understood Robin well enough to be able to put down what he said without accidentally strawmanning him but I do notice he was much more convincing and I was much more confused and challenged in person than it looks in this transcript, so perhaps I failed. Nevertheless:

Scott: In “Poor Folks Do Smile”, you say that a future of intense competition and bare subsistence will be okay because we will still have the sorts of things that make life valuable. But in a future of ems, won’t there be competitive advantage to removing the things that make life valuable?

Robin: What do you mean?

Scott: Suppose you have some ems that are capable of falling in love and some that aren’t. The ones that fall in love spend some time swooning or writing poetry or talking to their lover or whatever, and the ones that don’t can keep working 24-7. Doesn’t that give the ones that can’t fall in love enough of a competitive advantage that the ones that can will be outcompeted and destroyed and eventually we’ll end up with only beings incapable of love?

Robin: You can’t just remove love from a human brain like that. There’s no one love module.

Scott: It’s probably very hard to remove love from a human brain without touching anything else. But given that the future is effectively infinitely long, and that in a world of perfect competition it would be advantageous to do this, surely someone will succeed eventually.

Robin: Yes, the future is infinitely long. But you're speculating post-Singularity here, and the whole point of the Singularity is that it's impossible to speculate on what will happen after it. I speculate on the near and medium term future, but trying to predict the very long-term future isn't worth it.

Scott: I agree we can't predict the far future, but this is less a prediction than an anti-prediction. An anti-prediction is… wait, am I doing that thing where I explain something you invented to you?

Robin: No, I didn't invent anti-predictions. Go on.

Scott: An anti-prediction is…gah, I wish I could remember the canonical example…an anti-prediction is when you just avoid privileging a hypothesis and this sounds like a bold prediction. For example, suppose I predict with 99%+ confidence that the first alien species we meet will not be practicing Christians. In a certain context, this might sound overconfident – aliens could be atheists or Christians or Muslims, we don't really know, but since I don't know anything at all about aliens it sounds overconfident to be so sure it won't be the Christian one. But in fact this is justified, since Christianity is just a tiny section of possible-religion-space that only seems important to us because we know about it. The aliens' likelihood of being Christian isn't 1/3 ("either Christian, or atheist, or Muslim) but more like 1/1 trillion (Christianity out of the space of all conceivably possible religions). The only way the aliens could be Christian is if it was for some reason correlated with our own civilization's Christianity, like we went over there to convert them, or if Christianity was true and both us and the aliens were truth-seekers. My point is that human values, like love, are a tiny fraction of mindspace. So saying that the far future won't have them is an antiprediction.

Robin: Values like love were selected by evolution. We can expect that similar selection pressures in the future will produce, if not the same values, ones that are similar enough to be recognizable.

Scott: The hypercompetitive marketplace of an advanced cybernetic civilization is different enough from an African savannah that I really don't think that's true. Love evolved in order to convince people to reproduce and raise children. If ems can reproduce by copy-pasting and end out with full adults, that's not a society that will replicate the need for love.

Robin: Love is useful for a lot of other things. Probably the same mental circuitry that causes people to fall in love is the sort of thing you need to make people love their work and stay motivated.

Scott: Antiprediction! Most mind designs that can effectively perform tasks don't need circuitry that also causes falling in love!

Robin: The trouble with this whole antiprediction concept is…so what if I told you that in the far future, people would travel much faster than light. Would that be an antiprediction? After all, most physical theories don't include a hard light-speed limit.

Scott: The trouble with traveling faster than light is that it's physically impossible. Are you trying to make the claim that a mind design that doesn't include something like human love is physically impossible?

Robin: I'm trying to make the claim that it's not something you can plausibly get to by modifying humans.

Scott: Fine. Forget modifying humans. People just try to build something new and more efficient from the ground up.

Robin: Maybe in the ridiculously far future…

Scott: But we both agree on a sort of singularitarian world-view where “history is speeding up”. The “ridiculously far future” could be twenty years from now if ten years from now they invent ems that can be run at a hundred times normal speed. If the ridiculously far future aka twenty years from now is one where human values like love are completely absent, that seems…really bad. And if we want to prevent it, it seems like that goes through trying to prevent a “merely” Malthusian medium-term future in which people are effective slaves but we haven’t *quite* figured out how to hack out love yet.

Robin: Attempting to influence the far future is very dangerous. In most cases we can’t predict the long-term consequences of our action. The near future will be in a much better position to influence the far future than we are. My claim, which you don’t seem to disagree with, is that the near future will be non-hellish and preserve human values like love. Let’s let this near future figure out whether the far future will be unacceptable. As time goes on, people gain better ability to coordinate, so the near future should be better at fixing our problems anyway.

Scott: As time goes on people gain better ability to coordinate?

Robin: Yes. In the old days, most decisions were made at the village or provincial level. Now we’re gradually centralizing decisions to the national and often even the supranational level. The modern world is much more effective at coordinating solutions to its problems than the past.

Scott glances at Michael Anissimov, probably the most vocal Reactionary in Berkeley, who has been standing there listening to the conversation. He looks skeptical.

Scott: But I know Michael over here has been writing a lot claiming the opposite. That the modern world is terrible at coordinate problems, especially compared to the past. I'm somewhat sympathetic to that argument. In the old days, a king could just declare we were going to do something and it got done. Now we have nightmarish failures of coordination, like the Obamacare bill where the leftists had a decent and coherent vision for how healthcare should work, the rightists had a reasonable and coherent vision for how healthcare should work, and we smashed them together until we got a Frankensteinian mashup of both visions that satisfied no one. Or how back in the old days, the Catholic Church pretty much controlled…

Robin: Kings and the Church were very good at *acting*, not at *coordinating*. They could enforce their choices, but those choices were often terrible and uncorrelated with what anyone else wanted. Modern institutions *coordinate*.

Michael: But modern coordination is just through increased bureaucracy.

Robin: Call it what you want, it's still coordinating.

Michael: And the results are often terrible!

Robin: Yes, coordinating seems to divide into two subproblems. The first is getting everyone to agree on a solution. The second is making sure the solution is any good. I don’t claim we have solved the second subproblem, but we seem to be increasingly skilled at the first.

Michael: Really? Like the largest-scale world-coordinating organization we have right now seems to be the United Nations, and it’s famous for not getting anything done.

Robin: The thing with the UN is that at the beginning people expected it to be the umbrella organization under which all world affairs were conducted. But there are a host of other more or less associated organizations like the WTO that are actually doing a lot more.

Scott: You make an interesting case that future coordinating power will be better, but saying “let’s leave this to the future” only works if we know when the future is going to be and can prepare for it. In the case of what Eliezer calls a “foom” where an AI comes and causes a singularity almost out of nowhere – well, if we put off preparing for that for fifty years, and it happens in forty, that’s going to be really bad.

Robin: I think that scenario is very unlikely. In the scenario I believe in, an increase in technology led by emulated humans, change will occur on a predictable path. They will know if we’re on the path to eventual complete value deterioration.

Scott: That makes sense. So I guess that our real disagreement is only over the speed at which a singularity will happen, and whether we will know about it in time to protect our values.

Robin: Sort of. Although as I posted on my blog recently, I think “protecting values” is given too much importance as a concept. If any past civilization had succeeded in protecting *its* values, we’d be stuck with values that we would find horrible, mostly a mishmash of outdated and stupid norms about race and gender. So I say let future values drift by the same process our own values drifted. I don't mind if future people have slightly weirder concepts of gender than I do.

Scott: I think that's kind of unfair. You're assuming the future will vary over certain dimensions where you find variation acceptable. But it might vary in much stranger and less desirable ways than that. Imagine an ancient Greek who said "I'm a cosmopolitan person…I don't care whether the people of the future worship Zeus, or Poseidon, or even Apollo." He doesn't understand that the future also gets to vary in ways that are "outside his box".

Robin: It's possible. But like I said, I think we have a very long time before we have to worry about that. I would also suggest you look at the light speed limit. That means that there's going to be inevitable "cultural variation" in the post-human world, since it will probably include a lot of semi-isolated star systems.

Scott: I still expect a lot of convergence. After all, if this is a hypercompetitive society, then they'll be kind of forced into whatever social configuration leads to maximum military effectiveness or else be outcompeted by more militarily effective cultures.

Robin: No, not necessarily. There may be an advantage for the defender, such that it takes ten times the military might to attack as to defend. That would allow very large amounts of cultural deviation from the ideal military maximum.

After this the conversation moved on to other things and I don't have as good a memory. But it was great to meet Robin in person and I highly recommend his blog to anyone with an interest in futurism or economics.

Polyamory Is Boring

I.

I remember explaining polyamory to my father when I met him in Utah. He just shrugged and said "I guess I'm too old-fashioned for that sort of thing to make sense."

I feel blessed to have a father with the rare skill of being able to generate "I am old-fashioned" as a counter-hypothesis to "other people are evil". But more than that, I sympathize with his response. I sympathize with it because it was exactly my response when my friend Hannah told *me* about polyamory two years ago or so (I can't remember if I got it from IM conversations with her or from reading her essay). For a twenty-eight year old, I am *really* good at sighing and saying "Kids these days!" in a despairing tone, and that was about my response to the whole polyamory concept.

And now I'm doing it myself and it feels like the most normal thing in the world. I was going to make an analogy to desegregation here, how white people thought having black kids in their schools would be a disaster, and then it happened, and the world didn't collapse into a hell dimension or anything, and after a few years it just seemed like the normal order. But that metaphor is too weak: there's still racism, a black kid in an all white school district probably feels really out of place, there are still even fights over segregated proms.

So better analogy. Imagine a space-time rift brings a 19th-century Know-Nothing to your doorstep. He starts debating you on the relative merits and costs of allowing Irish people to mix with the rest of American society. And you have a hard time even getting the energy to debate him. You're like "Yeah, there are some Irish people around. I think my boss might be half-Irish or something, although I'm not sure. So what?" And he just sputters "But...but...Irish people! It's not right for Irish and non-Irish people to mix! Everyone knows that!" And not only do *you* not think that Irish people are a Big Deal, but you're about 99% sure that after the Know-Nothing spends a couple of months in 21st-century America *he's* going forget about the whole Irish thing too. There's just no way someone seeing how boring and ordinary Irish-Americans are could continue to consider worrying about it a remotely good use of their time.

In fact, this Know-Nothing would have two strikes against him if he tried to hold onto his philosophy. First, there's the empirical strike. Whatever his predictions of doom – Irish immigration would impoverish the country, Irish immigration would lead to the US being annexed by the Vatican – those predictions have clearly been disconfirmed. Second, there's the psychological strike. He would be exposed to so many perfectly normal Irish people that his brain would have trouble even maintaining them as a separate category. It's like the difference between your mental picture of Chinese people being Fu Manchu versus it being your neighbor John Chang who speaks perfect English and has a job at Google.

This was my experience with poly people upon moving to Berkeley. Hannah makes a big deal about poly-hacking and having to valiantly overcome some sort of strong natural tendency to switch from monogamous to polyamorous relationships. This wasn't really my experience at all. It just seemed like once the entire culture was no longer uniting to tell me polyamory was something bizarre and different and special, it wasn't. And then it started to look like a slightly better idea to take part in it than to not take part in it. So I did.

II.

I didn’t even remember how weird it seemed to everyone else until the last few weeks. First I had to explain it to my father. Then someone commented on a blog post of mine with something about polyamory, spelling it poly-“amor”-y the whole time, as if there couldn’t possibly be any real love involved.

The plural of anecdote is not “data”. But the singular of anecdote is “enough data to disprove a universal negative claim”. So I will just say that Hannah and Mike are probably the best couple I have ever seen. I have lived with them for seven months now and never once have I seen them get in a fight (I know there is way more to being a couple than not fighting but I’m trying to think of objective numerical evidence I can report here beyond “if you know them, you know what I mean”). They both seem to love and appreciate each other just as much if not more as they did when I first met them. They both go *way* out of their way to make the other happy, and although part of this is just that they’re both very nice people who go out of their way to make *everybody* happy, I think there’s got to be some love involved there too. They are engaged, working on the “getting married” thing, and have every intention of having lots of children and staying together for at least one lifetime.

And all this despite Mike having two other girlfriends and Hannah having three other boyfriends including one who lives with her. I can’t even get angry with people who say polyamory is incompatible with true love. They’re just empirically wrong, like someone who remarks confidently that hippos have six legs. They’re not evil or even deluded. They just obviously haven’t seen any hippos. You don’t really want to argue with them so much as take them to a zoo, after which you are confident they will realize their mistake.

III.

The other thing people always bring up is the jealousy issue. I feel like the correct, responsible thing to say at this point would be “Yes, of course everyone experiences jealousy, and it’s hard for the first few months or years, but eventually you just learn to live with it and the sacrifice is worth it.”

But the responsible answer is wrong, and the incredulous-stare answer is right. At least in my very limited experience, jealousy is a paper tiger, sort of the post-9/11 al-Qaeda of emotional states. You spend all this time worrying about it and preparing for it and thinking it is going to be this dreadfully imposing enemy, and in the end it sends one guy with a bomb in his shoes onto a plane, whom you arrest without incident.

I know this hasn’t been anywhere close to the experience of all polyamorous people, but it’s my experience and that of the people I’ve talked to most about this.

My roommate Mike dates the same three people I am dating, including Hannah who also lives with us (this is not normal for polyamory, and all three people started dating Mike and then met me and started dating me too, so I guess the moral of the story is to think very hard before accepting me as a roommate). I cannot think of a single problem I have ever had with Mike, which I guess is also sort of incredulous-stare and which exceeds my normal standards for *roommates* let alone roommates-whose-three-girlfriends-I-am-dating. None of those three people have had any noticeable-from-the-outside jealousy about any of the others. Two weeks ago, Mike and I took all three of our mutual girlfriends on a group date to Sausalito. It went really well, everyone got along, and it is something we would do more often if not for scheduling and travel issues (also, Sausalito is really expensive).

I *once* felt a small pang of jealousy when one of my girlfriends was having a very public display of affection with a non-Mike person I didn't know quite so well. But I get upset with/jealous of public displays of affection in general, even among people I don't know, and it's very hard for me to disentangle this feeling from jealousy and it could have just been my imagination.

As opposed to this tiny-to-nonexistant role of jealousy, I think pretty much everyone here has experienced compersion. Compersion is the opposite of jealousy, being really happy for your partner when they meet someone new and they are obviously happy. Mike and Hannah are really good at compersion (Mike helped set me up with his girlfriend Kenzi and was really glad it worked out) and some of this has rubbed off on me. It is a good feeling and it makes you feel good to have it. If there is a Heaven, I assume compersion will be a big part of its emotional repertoire.

IV.

I don't drink much, not because I'm especially virtuous but because I hate the taste of alcohol and the atmosphere of bars and parties. In the same way, I'm not promiscuous, not because I'm especially virtuous but because I'm sort of borderline asexual. I like cuddling people, kissing people, falling in love with people, petting people's hair, writing sonnets about people, and a few things less blogaboutable, but having sex isn't an especially interesting experience for me. I treat it kind of like watching a chick flick – something one might do to get the nice warm feeling of doing romantic things and bonding as a couple, but wait a second why the heck is she kissing him now and that scene made no sense and THIS MOVIE HAS NO PLOT HOW DID IT MAKE $100 MILLION AT THE BOX OFFICE?

And I'm sorry for subjecting random people to details of my sex life, but I'm trying to establish credibility here for what I want to say next. What I want to say next involves the perception – I had it and a lot of other people seem to have it – that polyamory is about having sex with lots of people and monogamy is about having close loving relationships. And once again this is not my experience at all.

If you just want to have lots of sex instead of having a loving relationship, there are many ways to do it that are much more socially acceptable than polyamory. You can be one of those bachelors who "plays the field" and "doesn't get tied down". You can be in an "open relationship" or be "swingers". All of these are way easier than polyamory; if your goal is sex, they're also more effective.

Polyamory is almost the opposite of this. It's for people who *aren't* just into sex, for people who realize they could get sex without relationships with a lot less deviation from social norms but are really into the relationship part of things.

Here I will say maybe the only note of personal uncertainty or concern you're likely to get in this essay, which is that I don't know whether I could have maximally-close relationships with multiple people simultaneously. That is, I don't know if I could date three people and love all of them as much as my parents love each other, or other social models for very good relationships (the Obamas? Now I'm foundering on who our non-fictional archetypes for very good relationships are) love each other. I’m not sure whether this would satisfy some deep human need for what you might politically-incorrectly call “mutual ownership”. And I’m definitely not sure (though I think it’s likely, certainly more likely than the skeptics would) that this is a great structure for child-rearing.

In practice none of this matters, because driven by some innate urge most polyamorous people I know end up having one “primary” relationship along with whatever others they are involved with. Mike and Hannah are each other’s primaries, and that is going to develop into being each other’s spouses, and what I said above about them definitely having achieved that level of maximum-closeness remains true. This form of polyamory seems to me to be “monogamy plus”, keeping all of the advantages of monogamous relationships and ending out strictly superior. Sometimes this develops into people being so into each other that they just aren’t interested in other relationships because it takes away time they could be spending with their primary partner, but I haven’t noticed any differences in the quality of relationships where this happens and ones where it doesn’t.

I have heard of polyamorous communities where this is not how things are done, where people don’t have primaries, where they are just this complicated mass of partners without anything that looks like a traditional relationship. I predict I would not like this; something in me recoils from this situation. But that could just be more prejudice that would look as dumb as a Know-Nothing in the 21st century once I saw it up close. I’m pretty willing to take the Biblical tack on this one: “He who is able to accept it, let him accept it”. But I’m pretty sure I’m not of that number.

RPG emergency!

*[**EDIT:** Solved, no need for more people to volunteer.]*

Much like the deity she represents, the Apologist in our Dungeons and Discourse game has abandoned us in our hour of need, leaving us hopeless and confused. Much like modern society, we are flailing about looking for something, anything, to serve as a substitute.

We really really need a new Apologist *as soon as possible* since we've already had to put this game off several weeks in a row and I think people are starting to get antsy. If you are religious or can fake being so, and if you're interested in philosophy and gaming, and if you're available for a game session sometime in the next five hours or so in an online chat room, please email me at scott[at]shireroth[dot]org and I will pencil you into the slot. This is the first session so you won't have missed anything and I can help you out with whatever you need help with. It would be a big favor for me and the rest of the participants not to mention the poor people who I nagged all of this weekend into getting their songs done in time.

Whose Utilitarianism?

[Trigger warning: attempt to ground morality]

God help me, I'm starting to have doubts about utilitarianism.

Whose Superstructure?

The first doubt is something like this. Utilitarianism requires a complicated superstructure – a set of meta-rules about how to determine utilitarian rules. You need to figure out which of people's many conflicting types of desires are their true "preferences", make some rules on how we're going to aggregate utilities, come up with tricks to avoid the Repugnant Conclusion and Pascal's Mugging, et cetera.

I have never been too bothered by this in a *practical* sense. I agree there's probably no perfect Platonic way to derive this superstructure from first principles, but we can come up with hacks for it that come up with good results. That is, given enough mathematical ingenuity, I could probably come up with a utilitarian superstructure that exactly satisfied my moral intuitions.

And if that's what I want, great. But part of the promise of utilitarianism was that it was going to give me something more objective than just my moral intuitions. Don't get me wrong; formalizing and consistency-ifying my moral intuitions would still be pretty cool. But that seems like a much less ambitious project. It is also a very personal project; other people's moral intuitions may differ and this offers no means of judging the dispute.

Whose Preferences?

Suppose you go into cryosleep and wake up in the far future. The humans of this future spend all their time wireheading. And because for a while they felt sort of unsatisfied with wireheading, they took a break from their drug-induced stupors to genetically engineer all desires beyond wireheading out of themselves. They have neither the inclination nor even the ability to appreciate art, science, poetry, nature, love, etc. In fact, they have a second-order desire in favor of continuing to wirehead rather than having to deal with all of those things.

You happen to be a brilliant scientist, much smarter than all the drugged-up zombies around you. You can use your genius for one of two ends. First, you can build a better wireheading machine that increases the current run through people's pleasure centers. Or you can come up with a form of reverse genetic engineering that makes people stop their wireheading and appreciate art, science, poetry, nature, love, etc again.

Utilitarianism says very strongly that the correct answer is the first one. My moral intuitions say very strongly that the correct answer is the second one. Once again, I notice that I don't really care what utilitarianism says when it goes against my moral intuitions.

In fact, the entire power of utilitarianism seems to be that I like other people being happy and getting what they want. This allows me to pretend that my moral system is "do what makes other people happy and gives them what they want" even though it is actually "do what I like". As soon as we come up with a situation where I no longer like other people getting what they want, utilitarianism no longer seems very attractive.

Whose Consequentialism?

It seems to boil down to something like this: I am only willing to accept utilitarianism when it matches my moral intuitions, or when I can hack it to conform to my moral intuitions. It usually does a good job of this, but sometimes it doesn't, in which case I go with my moral intuitions over utilitarianism. This both means utilitarianism can't *ground* my moral intuitions, and it means that if I'm honest I might as well just admit I'm following my own moral intuitions. Since I'm not claiming my moral intuitions are intuitions *about* anything, I am basically just following my own desires. What looked like it was a universal consequentialism is basically just *my* consequentialism with the agreement of the rest of the universe assumed.

Another way to put this is to say I am following a consequentialist maxim of "Maximize the world's resemblance to W", where W is the particular state of the world I think is best and most desirable.

This formulation makes "follow your own desires" actually not quite as bad as it sounds. Because I have a desire for reflective equilibrium, I can at least be smart about it. Instead of doing what I first-level-want, like spending money on a shiny new car for myself, I can say "What I seem to really want is other people being happy" and then go investigate efficient charity. This means I'm not quite emotivist and I can still (for example) be wrong about what I want or engage in moral argumentation.

And it manages to (very technically) escape the charge of moral relativism too. I think of a relativist as saying "Well, I like a world of freedom and prosperity for all, but Hitler likes a world of genocide and hatred, and that's okay too, so he can do that in Germany and I'll do my thing over here." But in fact if I'm trying to maximize the world's resemblance to my desired world-state, I can say "Yeah, that's a world without Hitler” and declare myself better than him, and try to fight him.

But what it’s obviously missing is objectivity. From an outside observer’s perspective, Hitler and I are following the same maxim and there’s no way she can pronounce one of us better than the other without having some desires herself. This is obviously a really undesirable feature in a moral system.

Whose Objectivity?

I’ve started reading proofs of an objective binding morality about the same way I read diagrams of perpetual motion machines: not with an attitude of “I wonder if this will work or not” but with one of “it will be a fun intellectual exercise to spot the mistake here”. So far I have yet to fail. But if there’s no objective binding morality, then the sort of intuitionism above is a good description of what moral actors are doing.

Can we cover it with any kind of veneer of objectivity more compelling than this? I think the answer is going to be “no”, but let’s at least try.

One idea is a *post hoc* consequentialism. Instead of taking everyone’s desires about everything, adding them up, and turning that into a belief about the state of the world, we take everyone’s desires about states of the world, then add all of those up. If you want the pie and I want the pie, we both get half of the pie, and we don’t feel a need to create an arbitrary number of people and give them each a tiny slice of the pie for complicated mathematical reasons.

This would “solve” the Repugnant Conclusion and Pascal’s Mugging, and at least *change the nature* of the problems around “preference” and “aggregation”. But it wouldn’t get rid of the main problem.

The other idea is a sort of morals as Platonic politics. Hobbes has this thing where we start in a state of nature, and then everybody signs a social contract to create a State because everyone benefits from the State’s existence. But because coordination is hard, the State is likely to be something simple like a monarchy or democracy, and the State might not necessarily do what any of the signatories to the contract want. And also no one actually signs the contract, they just sort of pretend that they did.

Suppose that Alice and Bob both have exactly the same moral intuitions/desires, except that they both want a certain pie. Every time the pie appears, they fight over it. If the fights are sufficiently bloody, and their preference for personal safety outweighs their preference for pie, it probably wouldn’t take too long for them to sign a contract agreeing to split the pie 50-50 (if one of them was a better fighter, the split might be different, but in the abstract let’s say 50-50).

Now suppose Alice is very pro-choice and slightly anti-religion, and Bob is slightly pro-life and very pro-religion. With rudimentary intuitionist morality, Alice goes around building abortion clinics and Bob burns them down, and Bob goes around building churches and Alice burns them down. If they can both trust each other, it probably won’t take long before they sign a contract where Alice agrees not to burn down any churches if Bob agrees not to burn down any abortion clinics.

Now abstract this to a civilization of a billion people, who happen to be divided into two equal (and well-mixed) groups, Alicians and Bobbites. These groups have no leadership, and no coordination, and they’re not made up of lawyers who can create ironclad contracts without any loopholes at all. If they had to *actually* come up with a contract (in this case maybe more of a treaty) they would fail miserably. But if they all had this internal drive that they should imagine the contract that would be signed among them if they could coordinate perfectly and come up with a perfect loophole-free contract, and then follow that, they would do pretty well.

Because most people’s intuitive morality is basically utilitarian [citation needed], most of these Platonic contracts will contain a term for people being equal even if everyone does not have an equal position in the contract. That is, even if 60% of the Alicians have guns but only 40% of the Bobbites do, if enough members of both sides believe that respecting people’s preferences is important, the contract won’t give the Alicians more concessions on that basis alone (that is, we’re imagining the contract real hypothetical people would sign, not the contract hypothetical hypothetical people from Economicsland who are utterly selfish would sign).

Whose Communion?

So what about the wireheading example from before?

Jennifer RM has been studying ecclesiology lately, which seems like an odd thing for an agnostic to study. I took a brief look at it just to see how crazy she was, and one of the things that stuck with me was the concept of communion. It seems (and I know no ecclesiology, so correct me if I’m wrong) motivated by a desire to balance a desire to unite as many people as possible under a certain banner, with the conflicting desire to have everyone united under the banner believe mostly the same things and not be at one another’s throats. So you say “This range of beliefs is acceptable and still in communion with us, but if you go outside that range, you’re out of our church.”

Moral contractualism offers a similar solution. The Alicians and Bobbites would sign a contract because the advantages of coordination are greater than the disadvantages of conflict. But there are certain cases in which you would sign a much weaker contract, maybe one to just not kill each other. And there are other cases still when you would just *never sign a contract*. My Platonic contract with the wireheaders is “no contract”. Given the difference in our moral beliefs, whatever advantages I can gain by cooperating with them about morality are outweighed by the fact that I want to destroy their entire society and rebuild it in my own image.

I think it’s possible that all of humanity except psychopaths are in some form of weak moral communion with each other, at least of the “I won’t kill you if you don’t kill me” variety. I think certain other groups, maybe along the culture level (where culture = “the West”, “the Middle East”, “Christendom”) may be in some stronger form of moral communion with each other.

(note that “not in moral communion with” does not mean “have no obligations toward”. It may be that my moral communion with other Westerners contains an injunction not to oppress non-Westerners. It’s just that when adjusting my personal intuitive morality toward a morality I intend to actually practice, I only acausally adjust to those people whom I agree with enough already that the gain of having them acausally adjust toward me is greater than the cost of having me acausally adjust to them.)

In this system, an outside observer might be able to make a few more observations about the me-Hitler dispute. She might notice Hitler or his followers were in violation of Platonic contracts it woud have been in their own interests to sign. Or she might notice that the moral communions of humanity split neatly into two groups: Nazis and everybody else.

I’m pretty sure that I am rehashing territory covered by other people; contractualism seems to be a thing, and a lot of people I’ve talked to have tried to ground morality in timeless something-or-other.

Still, this appeals to me as an attempt to ground morality which successfully replaces obvious logical errors with complete outlandish incomputability. That seems like maybe a step forward, or something?

EDIT: Clarification in my response to Kaj here.

Book Review: After Virtue

A few weeks ago the blogosphere discovered Ayn Rand's margin notes on a C.S. Lewis book. They were everything I expected and more. Lewis would make an argument, and then Rand would write a stream of invective in the margin about how much she hated Lewis' arguments and him personally. I kind of wanted to pat her on the shoulder and say "Look, I'm really sorry, but *he can't hear you*."

But I can also sympathize with her. It is *infuriating* to read a book making one horrible argument after the other. And when it glibly concludes "…and therefore I am right about everything", and you know you'll never be able to contact the author, it gives a pale ghost of satisfaction to at least scrawl in the margin "YOUR ARGUMENTS ARE BAD AND YOU SHOULD FEEL BAD".

This is kind of how I felt about Alasdair MacIntyre's *After Virtue*.

As far as I can tell, MacIntyre's central argument works something like this:

1. There are many theories of ethics in existence today
2. The ones that came after Aristotelianism have failed to objectively ground themselves and create a perfect society in which everyone agrees on a foundation for morality
4. Therefore, we should return to Aristotelianism

You may notice a hole where one might place a Step 3, something like "Aristotelianism, in contrast, *did* objectively ground itself and create a perfect society in which everyone agreed on a foundation for morality." This is exactly the argument MacIntyre digresses into a lengthy explanation of how much he likes Greek tragedy to hope we will avoid noticing him not making.

To MacIntyre's credit, he does a pretty good critique of modern moral philosophy. He says that since society doesn't share any kind of moral tradition, we can debate important moral questions – like abortion, or redistributive taxation – until the cows come home, but this is in fact only the appearance of debate since we have no agreed-upon standards against which to judge these things. Because we cannot settle these by rational argument, instead we turn to outrage and attempts to shame our opponents, making the protester one of the archetypal figures of the modern world.

("…making the [unsavory sounding figure] one of the archetypal figures of the modern world" is one of MacIntyre's pet phrases. It starts grating after a while.)

I broadly agree with him about this problem. I discuss it pretty explicitly in sections 6.5 and 8.1 of my Consequentialism FAQ. I propose as the solution some form of utilitarianism, the only moral theory in which everything is commensurable and so there exists a single determinable standard for deciding among different moral claims.

Annnnnd MacIntyre decides to go with virtue ethics.

The interesting thing about virtue ethics is that it is *uniquely bad at this problem*. In the entire book, MacIntyre doesn't give a single example of virtue ethics being used to solve a moral dilemma, as indeed it cannot be. You can attach a virtue (or several virtues) of either side of practically any moral dilemma, and virtue ethics says exactly nothing about how to balance out those conflicting duties. For example, in Kant's famous "an axe murderer asks you where his intended victim is" case, the virtue of truthfulness conflicts with the virtue of of compassion (note, by the way, that no one has an authoritative list of the virtues and they cannot be derived from first principles, so anyone is welcome to call anything a virtue and most people do).

MacIntyre totally admits this conflict, but instead of saying it's a problem with his theory he says it's the tragedy of human existence, then says that the virtue of justice is knowing how to balance those two virtues.

So basically, his entire condemnation of all systems beside his own is based on the difficulty of coming to moral consensus, but his own means of coming to moral consensus is a giant black box labelled "THE VIRTUE OF BEING ABLE TO SOLVE THIS HERE PROBLEM CORRECTLY".

I don't like deontology. In fact, I dislike it more than almost anyone I know except maybe Federico. But I will give credit where credit is due: deontology actually comes up with solutions to moral problems. The solutions are wildly incorrect and incredibly harmful, but they get a gold star for effort.

Virtue ethics, as far as I can tell, just gives you a knowing look and says "The very fact that you interpret morality in terms of *moral dilemmas* is a symptom of the disease of liberal modernity." This is useful for sounding deeply wise, but little else. If you ask "Okay, but disputes over morality are an actual feature of the real world, and the whole reason we're doing this ethics stuff is to try to solve them, so if we admit we're diseased and the ancient Greeks were awesome, maybe you could help us out here?" – then virtue ethics just takes another sip of wine from its table in the corner and says "Your decadent individualist mind has no idea how disappointed Aristotle would be in you for even asking that. Did you even *consider* just being a virtuous city-state in which everyone is a great-minded soul acting for the good of the polis? I didn't *think* so."

If You Can't Convince 'Em, Just Start Reciting The Entire History Of The Human Race

Beyond my distaste for *After Virtue*'s philosophy, I wasn't a huge fan of its history either.

The book claims that the reason we don't have a working agreed-upon morality is that the ancient Greeks (and medievals) *did* have a working agreed-upon morality (virtue ethics), but when it collapsed we were left with all these weird phrases like "virtuous" and "should" and "ought" and "the good" and outside the context of virtue ethics had no idea what to do with them. Since we couldn't use the correct virtue-ethics solution, we entered the age of interminably debating what the correct solution was, hence the modern age of moral dilemmas.

In fact, the beginning of the book is a fascinating and attractive metaphor (drawn from the excellent *A Canticle For Leibowitz*) in which all scientific knowledge is destroyed by some apocalypse. A future civilization picking over the scraps forms a sort of cargo cult in which they know there are supposed to be things called "electrons", and that the equation "e = mc^2" is very important for no reason, but no matter how many times they debate what shape these "electrons" were supposed to be or whether the c in e=mc^2 stands for 'color' or 'correctness', they can't seem to produce rockets or nuclear power. Phrases like e=mc^2 only make sense as part of a tradition; a stupid debate about whether c stands for color or correctness is a symbol that we're trying to interpret it separately from that tradition and we're just going to end up confusing ourselves. To MacIntyre, the tradition here is virtue ethics and modern society plays the role of the postapocalyptics looking quizzically over the scraps.

(the apocalypse? The Enlightenment, of course. Just *once* I want to go a whole week without someone blaming everything on the Enlightenment.)

Alasdair MacIntyre is clearly an expert classical scholar. And in fact he discusses the classical world's disputes on morality very competently in his book. So it bewilders me that he doesn't notice that actually, modern society's debates over the Good are no different than those of the classical world. He even cites Sophocles' tragedy Philoctetes as an example of moral dilemma in the ancient world. I agree – it is a perfect moral dilemma – of exactly the sort MacIntyre is claiming only exists because our civilization is living in the postapocalyptic ruins of virtue ethics. And *Philoctetes* was written twenty years before Aristotle was even born. Heck, forget Sophocles, even Socrates is a perfect example of this kind of moral inquiry.

MacIntyre then waxes about the wonder of the Greek city-states, which he says were communities where everyone was united on a single view of the good – that which was the proper *telos* of man.

Except, once again, all the problems of the modern age appear in the Greek city-states as well. Athens went from the laws of Solon to the tyranny of Peisistratus to the dictatorship of Hippias to the democracy of Cleisthenes to the oligarchy of the Four Hundred to the Thirty Tyrants to the democracy of Thrasybulus all in about a century. The periods of democracy were as rife with hostile factions and unresolved issues as any period in modern America or Europe.

The idea that everyone back then was happily united around the Objectively Proper End of Man is slightly complicated by the fact that no one back then agreed on what the Objectively Proper End of Man was, any more than anyone today agrees on what the Proper End of Man is, least of all virtue ethicists and super-dog-double-least of all anyone who reads the book *After Virtue* which happily informs us that pursuing it will solve all our problems but neglects to mention what the heck it might be or give us a shred of evidence to overcome our high priors against such a thing existing.

Then there's a short focus on the medieval period, which I am told is marked by everyone being very virtuous but otherwise not particularly worthy of remark, followed by an attack on David Hume and Immanuel Kant, who apparently both *totally failed to be virtue ethicists*.

The modern period is marked…okay, I understood this part even less than the other parts. The modern period is marked by the Bureaucrat, who is another one of those Archetypal Figures Of The Modern World (others include the Aesthete and the Therapist). The Bureaucrat claims to have expertise in some subject, but clearly this is a lie, because no one can ever understand human affairs infallibly and this is *kind of* like saying no one can ever understand human affairs at all. Since everyone loves bureaucrats, who are people who claim to be able to understand human affairs, and yet no one can *really* understand human affairs, something must be wrong, and for all we know that something could be that we're not all virtue ethicists (am I strawmanning here? Read pages 79-108 and find out).

Somebody Here Is Really Confused, And I Just Hope It's Not Me

I have never been able to appreciate Continental philosophy (well, Nietzsche was pretty cool, but I have a hard time classifying anyone who can actually write engagingly as a Continental philosopher). *After Virtue*, despite having been written by a verified Scotsman by all accounts closely engaged with the analytic tradition, just seemed really Continental to me. It avoided logical arguments for a particular well-defined point in favor of long historical meanderings carefully designed to make the reader vaguely worry that everything was socially constructed and that the reader's social construction was particularly rotten, without ever coming out and explicitly saying anything that could be seized upon as a claim to evaluate.

But the thing is that MacIntyre is considered one of the greatest living philosophers, and *After Virtue* one of the century's greatest works on ethics. Just on priors I'm more likely to be misunderstanding him than he is to be talking nonsense. Even people I respect – including Catholics from the Patheos community and a few rationalists from the Less Wrong community – recommend MacIntyre.

Those same people recommended Edward Feser to me. There are a lot of similarities between Feser and MacIntyre – both say that the philosophical tradition of Greece and the medieval age was much better than our own tradition, and that we're so screwed up we can't even *realize* how screwed up we were. Both have very good things to say about teleology, and both ended up Catholic as a result of their philosophical studies.

I really enjoyed Feser's *The Last Superstition* (and his *Aquinas*, although that's less relevant here). I thought it did a great job bridging a wide inferential gap and really illuminated why he thought the things he thought. I think his account of forms and teleology is flawed because of a few basic errors in his foundations (I started explaining why on my old blog but never really finished) but it was flawed in ways where I could understand the force of his arguments and why his premises would lead to that conclusion. Even if I ended up disagreeing with his answers, I gained a huge admiration for his ability to ask the right questions and go about investigating them in the right way.

But as his occasional enemy Chris Hallquist delights to point out, Feser is not a hugely prestigious figure in mainstream academic philosophy. MacIntyre is. I was hoping for the same fascinating ideas, but with a suave British cool instead of hilarious over-the-top rants. Instead I got…I don’t even know.

I am really sorry, virtue ethicists. But you are going to have to do better than this if you want me to understand you.

Read History Of Philosophy Backwards

Back when I was in college, my chief complaint about my philosophy course was that it spent all its time teaching stuff by Aristotle or Plato or Descartes that was just *obviously wrong*. I sort of annoyed my professors by constantly raising my hand and being like "Sorry, isn't Plato completely confused here because now we know that actually X", and my professor would be "Well, that's a very interesting theory" and then continue teaching Plato.

None of these classes were billed as "history of philosophy", but as "philosophy" itself. I knew better than to expect to be taught a single thing that was definitely right, but I had kind of hoped they would limit it to things that weren't *totally* outside what modern opinion would find plausible.

As Dave Barry puts it:

> "I was terrible at history. I could never see the point of learning what people thought back when people were a lot stupider. For instance, the ancient Phoenicians believed that the sun was carried across the sky on the back of an enormous snake. So what? So they were idiots"

I still believe that if you only have four years to teach an undergraduate philosophy, you shouldn't be teaching this stuff before you teach genuinely useful things like what the heck concepts are and why you can't suddenly change the moral value of things by calling them different names. But I'm starting to appreciate why it might be useful.

Take Sartre. His main point – that no one else can tell you who you are, and you choose what your own values are – seems so cliched, so much like what an uncreative graduation speaker might say – that it hardly seems worth elevating him to the Canon Of Philosophical Greatness.

My hypothesis – and I don’t know if it’s true – is that this only sounds cliched because Sartre won. The point of studying Sartre isn’t to learn that you choose your own identity, but to *read him backward* – to start with this idea that choosing your own identity is obvious, and then read Sartre to learn exactly how controversial it was at the time and what sorts of arguments Sartre had to go through to get people to accept it, and eventually understand the position that the original reader of Sartre was supposed to have *started with*. If you succeed, you might still believe that you choose your own identity, but you’ll also understand that this isn’t an obvious necessary fact of the universe, that there used to be people who believed you didn’t and that they had some good arguments too.

When I first studied Hobbes, I figured nobody agreed with Hobbes these days – after all, Hobbes was a believer in absolute monarchy, and now everyone is strongly opposed to that. But later I realized that pretty much everyone is a Hobbesian in that Hobbes was one of the first people to think in terms of people coming together to found a government for their mutual self-interest; previously governments were either just the natural state of human affairs, or part of the hierarchical nature of the universe under God, or composed because the *telos* of man only flourishes in a community, or not even something you thought about. In fact, Alasdair MacIntyre seems to be at least partially advocating a return to pre-Hobbesian ideas about government, even though he doesn’t put it in those terms.

The only reason I had Hobbes pegged as “the absolutism guy” was because that was the only place in which his theories differed from my own and so I assumed it was his only idea that wasn’t “obvious”. If I’d read him backward, I would have gotten a lot more out of him.

Under this model, reading philosophers who were completely wrong is another way of unlearning your assumptions. For example, I originally thought the term “reductionism” was essentially meaningless; the opposite of “reductionism” was “not thinking things through clearly and having incoherent ideas”. After I read Aristotle, I changed my mind; he proposes a non-reductionism which for all I know very well may be the case in some other dimension, even if it isn’t how things work here on Earth.

Here the point of reading history of philosophy is to unlearn your assumptions. Growing up in a certain cultural tradition not only influences the answers you think are right, but the potential answers you’re able to generate and the questions you’re able to ask.

For some people this is important because the past was actually correct or close to it. I hang out with a disproportionate number of these people. I was briefly taken aback when Chris claimed here that he *doesn’t* have to constantly listen to claims that the Enlightenment ruined everything.

But what if, like me, you think the past was pretty thoroughly wrong about most philosophical issues?

Then I *still* think it’s important to have a non-parochial worldview, because the *next* big idea is likely to be just as different from present philosophy as present philosophy is from past philosophy. And unless you realize how different present philosophy is from past philosophy, you won’t even have the mental mechanism to expand your search space large enough to capture something worthy of participating in the future.

Virtue Ethics: Not Practically Useful, Either

I’ve been trying to understand some of the responses to my review of *After Virtue*. Tell me whether this is about right:

The problem of “doing the right thing” consists of two subproblems.

First, knowing what the right thing is. Do we legalize or ban abortion? Do we press the switch in the trolley problem or not?

Second, behaving correctly in situations where we do know what the right thing is. For example, going to visit a friend in the hospital even though the hospital is far away. Not cheating on your taxes even though you could use the money. Working hard even though no one is checking up on you.

The proposal was that virtue ethics doesn’t claim to be a solution to the first problem, but is a uniquely excellent solution to the second, and in fact the solution people actually use. Am I understanding this correctly?

Because if that’s true, I *still* disagree.

Virtue Ethics Is What People Do

I am not really good at thinking in terms of good or bad people. I can do so in very edge cases, like Kim Jong-il or St. Holden. But I tend to process all the people I know and have remotely okay interactions with as “good people”, leading to conversations like:

Me: Oh, cool, Bob is coming over soon! Bob is great!
Friend: But didn't Bob [do X, Y, and Z]?
Me: Well, yes, but other than that he's great.
Friend: And didn't he [do A, B, and C]?
Me: You can't just keep taking all these things Bob does out of context!
Friend: Okay, what has Bob done that was good?
Me: He…well…he…you know! Bob! He's great!

When I deviate from this, it almost always tends to be in terms of actions, not qualities. Like "Bob is great, except that he posts really annoying things on Facebook all the time". Or "Bob is great, except that he has some really horrible politics."

When I think about my own morality, it's almost never in terms of whether I am or am not a virtuous person (I have at least one subagent that's always convinced I'm a virtuous person, and at least one subagent that's always convinced I'm terrible and deserve to die, and doing good things paradoxically strengthens the latter subagent for some reason).

It's usually in terms of – I guess it feels like a missile must feel locking on to a target. If my mind is sufficiently calm and predisposed to goodness, I think "Wait a second, *that's* the morally correct thing to do", lock on to one option, and then feel really good about it – a serene feeling.

I can't always do this. If I'm sufficiently angry, part of me thinks "I bet I could lock on to what's good and do it…but that would probably involve turning the other cheek, or compromising, AND THEN THESE JERKS WOULD GET AWAY WITH IT." And then my mind makes up all sorts of game theoretic justifications for why it's more important to punish defectors than to do what feels like the right thing at this precise moment. Or if I'm sufficiently exhausted, I think "If I started worrying about what's good now, then I'd probably have to do it, and that would be really arduous."

On the one hand, this is no doubt a very idiosyncratic report; I don't expect my experience of morality to be similar to anyone else's. On the other hand, this is an idiosyncratic report and I don't expect my experience of morality to be similar to anyone else's. So if you say "Virtue ethics is the way people naturally think about morality!", that's either a typical mind fallacy or you're going to have to do a *much* better job explaining virtue ethics.

My experience of morality is contray to traditional virtue ethics in almost every way. It doesn't feel like it depends on my social roles. It doesn't strike me as divisible – that is, it feels like solid goodness and words like "continence" or "prudence" don't do anything to me. It doesn't strike me as the same feeling that occurs when I consider important but non-ethical questions like procrastination. It doesn't strike me as performed in a community or according to a narrative. It's just *not* virtue ethics.

The Practice of Making People More Moral

The other claim is that virtue ethics is the science of making people better – a process for helping people refine their existing moral intuitions and overcome temptation more effectively.

If that's true, it's a science much like medieval medicine was a science – totally untested and not especially likely to bear any resemblance to reality. If people are *actually* looking for ways to become more moral, I bet an hour's search would find about thirty of them that are more likely to work than adopting virtue ethics.

These could be broadly divided into beliefs and practices. In terms of beliefs, I think the most useful would be a belief in the Devil, moral realism, humility, and various kinds of magical thinking.

In terms of practices, there would be willpower training (pretty much anything that requires willpower counts as willpower training, but let’s say exercise as the stereotypical example), rationality training, keeping a journal, talking about morality, making friends, joining a group with some sort of interest in morality, cutting yourself off from bad influences, making yourself happier (happy people are more moral), learning relaxation/stress-busting techniques, and reading fiction.

All of these things would make people more moral in different directions and in different ways. For example, I bet reading works of fiction about poor people in the Third World would make you more likely to donate to charity, and contemplating virtue ethics and the just *polis* would make you more likely to make you get involved in local politics. Which of these you recommend is very closely linked to whether you think giving to charity to the Third World is more or less important than getting involved in local politics (hint: there is only one answer to this question which is not really stupid).

In terms of the best all-around practice for increasing morality I would nominate meditation, especially lovingkindness meditation. David Chapman, who knows ten zillion times more about Buddhism and meditation than I do, suggests metta bhavana, tonglen, and chöd. Even very generic meditation ticks several of the boxes above – relaxation, willpower training, and happiness – but these are said to (and have some evidence of) specifically increasing your ability to love and care about other people.

This seems like probably the best thing you can do for morality short of ground it objectively. If people love and care about others, they end up automatically ticking the two boxes I claim are required to end up more-or-less utilitarian – grounding morality in the world and caring about other people. If you've got that, it kind of lowers the degree to which morality even needs to be grounded objectively; it would still be nice, but we can trust people to do the right thing even if it isn't.

Virtue ethics doesn't satisfy either of these criteria, and in fact, we find that throughout history a lot of really terrible people have been very good virtue ethicists (the Spartans come to mind). So although many of the commenters here want to virtue ethics from its failure to ground morality by saying it removes the need to ground morality, I think virtue ethics can't even do that right and there are a lot of things that are much better.

Edit: Something I said in the comments that might clarify my position. I think even this is giving virtue ethics too much credit, since it's *not* just "use our inborn moral sense" but a host of claims about making lists of virtues and studying teleology – but on the principle of steelmanning an opponent's argument:

> Imagine that instead of virtue ethics we're talking about grammar. In most cases, we have a natural grammar sense – that is, the real reason I don't say "Me is Scott" is because it just sounds wrong.
>
> In most cases this is good enough. In some cases it isn't – for example, sometimes we have to teach grammar to foreigners who lack this intuitive grammar sense. Or sometimes there are edge cases where we're really not sure what word to use. Or we want to program a

computer to write with proper grammar. Or we want to set editorial policy for a newspaper. Or sometimes we're just genuinely curious how grammar works.

There is a point to having a science of grammar where smart people say "Oh, it looks like the predicate nominative form is used in this way."

And inventing a "virtue grammar", where people say "But you're ignoring all normal grammar usage in favor of a few silly edge cases! Real people don't talk about predicate nominatives! Just use your natural grammar sense!" is a total waste of everyone's time. Yes, natural grammar sense usually works well, but shouting "Hey, natural grammar sense often works well!" contributes nothing to the field and is just distracting people from actually figuring out how grammar works.

In pretty much all fields except ethics, everyone has agreed that the proper thing to do is to be happy when our natural senses are good enough, but also create a formal study of the field in order to go beyond what our natural senses can tell us. I don't understand why we can't also do this for ethics.

Lizardman’s Constant Is 4%

Beware of Phantom Lizardmen

I have only done a little bit of social science research, but it was enough to make me hate people. One study I helped with analyzed whether people from different countries had different answers on a certain psychological test. So we put up a website where people answered some questions about themselves (like “what country are you from?”) and then took the psychological test.

And so of course people screwed it up in every conceivable way. There were the merely dumb, like the guy who put “male” as his nationality and “American” as his gender. But there were also the actively malicious or at least annoying, like the people (yes, more than one) who wrote in “Martian”.

I think we all probably know someone like this, maybe a couple people like this.

I also think most of us *don’t* know someone who believes reptilian aliens in human form control all the major nations of Earth.

Public Policy Polling’s recent poll on conspiracy theories mostly showed up on my Facebook feed as “Four percent of Americans believe lizardmen are running the Earth”.

(of note, an additional 7% of Americans are “not sure” whether lizardmen are running the Earth or not.)

Imagine the situation. You’re at home, eating dinner. You get a call from someone who says “Hello, this is Public Policy Polling. Would you mind answering some questions for us?” You say “Sure”. An extremely dignified sounding voice says

– and this is the exact wording of the question – “Do you believe that shape-shifting reptilian people control our world by taking on human form and gaining political power to manipulate our society, or not?” Then it urges you to press 1 if yes, press 2 if no, press 3 if not sure.

So first we get the people who think “Wait, was 1 the one for if I did believe in lizardmen, or if I didn’t? I’ll just press 1 and move on to the next question.”

Then we get the people who are like “I never heard it before, but if this nice pollster thinks it’s true, I might as well go along with them.”

Then we get the people who are all “F#&k you, polling company, I don’t want people calling me when I’m at dinner. You screw with me, I tell you what I’m going to do. I’m going to tell you I believe lizard people are running the planet.”

And *then* we get the people who put “Martian” as their nationality in psychology experiments. Because some men just want to watch the world burn.

Do these three groups total 4% of the US population? Seems plausible.

I really wish polls like these would include a control question, something utterly implausible even by lizard-people standards, something like “Do you believe Barack Obama is a hippopotamus?” Whatever percent of people answer yes to the hippo question get subtracted out from the other questions.

Poll Answers As Attire

Alas, not all weird poll answers can be explained that easily. On the same poll, 13% of Americans claimed to believe

Barack Obama was the Anti-Christ. Subtracting our Lizardman’s Constant of 4%, that leaves 9% of Americans who apparently gave this answer with something approaching sincerity.

(a friend on Facebook pointed out that 5% of *Obama voters* claimed to believe that Obama was the Anti-Christ, which seems to be another piece of evidence in favor of a Lizardman’s Constant of 4-5%. On the other hand, I do enjoy picturing someone standing in a voting booth, thinking to themselves “Well, on the one hand, Obama is the Anti-Christ. On the other, do I really want four years of Romney?”)

Some pollsters are starting to consider these sorts of things symptomatic of what they term symbolic belief, which seems to be kind of what the Less Wrong sequences call Professing and Cheering or Belief As Attire. Basically, people are being emotivists rather than realists about belief. “Obama is the Anti-Christ” is another way of just saying “Boo Obama!”, rather than expressing some sort of proposition about the world.

And the same is true of “Obama is a Muslim” or “Obama was not born in America”.

Never Attribute To Stupidity What Can Be Adequately Explained By Malice

But sometimes it’s not some abstruse subtle bias. Sometimes it’s not a good-natured joke. Sometimes people might just be actively working to corrupt your data.

Another link I’ve seen on my Facebook wall a few times is this one: Are Climate Change Sceptics More Likely To Be Conspiracy Theorists? It’s based on a paper by Stephen Lewandowsky et al called NASA Faked The Moon Landing, Therefore Climate Science Is A Hoax – An Analysis Of The Motivated Rejection Of Science.

The paper’s thesis was that climate change skeptics are motivated by conspiracy ideation – a belief that there are large groups of sinister people out to deceive them. This seems sort of reasonable on the face of it – being a climate change skeptic requires going against the belief of the entire scientific establishment. My guess is that there probably is a significant link here waiting to be discovered.

Unfortunately, it’s…possible Stephan Lewandowsky wasn’t the best person to investigate this? Aside from being a professor of cognitive science, he also runs Shaping Tomorrow’s World, a group that promotes “re-examining some of the assumptions we make about our technological, social and economic systems” and which seems to be largely about promoting global warming activism. While I think it’s admirable that he is involved in that, it raises conflict of interest questions. And the way his paper is written – starting with the over-the-top title – doesn’t do him any favors.

(if the conflict of interest angle doesn’t make immediate and obvious sense to you, imagine how sketchy it would be if a professional global warming *denier* was involved in researching the motivations of global warming *supporters*)

But enough of my personal opinions. What’s the paper look like?

The methodology goes like this: they send requests to several popular climate blogs, both believer and skeptic, asking them to link their readers to an online survey. The survey asks people their beliefs on global warming and on lots of conspiracy theories and fringe beliefs.

On first glance, the results are extremely damning. People who rejected climate science were wildly more likely to reject pretty much every other form of science as well, including the “theory” that HIV causes AIDS and the “theory” that cigarettes cause cancer. They were more willing to believe aliens landed at Roswell, that 9-11 was an inside job, and, yes, that NASA faked the moon landing. The conclusion: climate skeptics are just really stupid people.

But a bunch of global warming skeptics started re-analyzing the data and coming up with their own interpretations. They found that many large pro-global-warming blogs posted the link to the survey, but very few anti-global-warming blogs did. This then devolved into literally the worst flame war I have ever seen on the Internet, centering around accusations about whether the study authors deliberately excluded large anti-global warming blogs, or whether the authors asked the writers of anti-global-warming blogs and these writers just ignored the request (my impression is that most people now agree it was the latter). In either case, it ended up with most people taking the survey being from the pro-global-warming blogs, and only a few skeptics.

More interestingly, they found that pretty much all of the link between global warming skepticism and stupidity was a couple of people (there were so few skeptics, *and* so few conspiracy believers, that these couple of people made up a pretty big proportion of them, and way more than enough to get a “significant” difference with the global warming believers). Further, most of these couple of people had given the maximally skeptical answer to every single question about global warming, and the maximally credulous answer to every single question about conspiracies.

The danger here now seems obvious. Global warming believer blogs publish a link to this study, saying gleefully that it’s going to prove that global warming skeptics are idiots who also think NASA faked the moon landing and the world is run by lizardmen or whatever. Some global warming believers decide to help this process along by pretending to be super-strong global warming skeptics and filling in the stupidest answers they can to every question. The few real global warming skeptics who take the survey aren’t enough signal to completely drown out this noise. Therefore, they do the statistics and triumphantly announce that global warming skepticism is linked to stupid beliefs.

The global warming skeptic blogosphere has in my opinion done more than enough work to present a very very strong case that this is what happened (somebody else do an independent look at the controversy and double-check this for me?) And Professor Lewandowsky’s answer was…

…to publish a second paper, saying his results had been confirmed because climate skeptics were so obsessed with conspiracy theories that they had accused his data proving they were obsessed with conspiracies of being part of a conspiracy. The name of the paper? Recursive Fury. I have to hand it to him, this is possibly *the most chutzpah I have ever seen a single human being display.*

(the paper is now partially offline as the journal investigates it for ethical something something)

The lesson from all three of the cases in this post seems clear. When we’re talking about very unpopular beliefs, polls can only give a weak signal. Any possible source of noise – jokesters, cognitive biases, or deliberate misbehavior – can easily overwhelm the signal. Therefore, polls that rely on detecting very weak signals should be taken with a grain of salt.

Last Thoughts On Virtue Ethics

The discussion on the other posts has sort of degenerated into people pointing out that our intuitive moral sense is a whole lot more useful most of the time than the speculations of moral philosophers, therefore virtue ethics.

I have two complaints here, the first of which is that virtue ethics is *not* just the claim that we should use our intuitive moral sense. It makes highly counterintuitive or controversial claims like the following:

1. Ethics involves teleology, eg considering the objectively proper ends of beings
2. Ethics has to be grounded in a community to make sense; individual ethics are only a pale shadow
3. Ethics is role-dependent; your role as a mother or child or employee or citizen produces your ethical obligations
4. Ethics is better thought of as about people's character than about the acts they perform
5. It is useful and important to subdivide good behavior into certain virtues like justice, wisdom, and fortitude

1 is almost universally disagreed with by everyone not a practicing virtue ethicist in a philosophy department or a very theologically-minded Catholic. 2 and 3 seem like things most people have no strong opinion about and would leave it for philosophers to debate. You could cherry-pick examples of people's behavior where it looks like they believe 4 and 5 (we have phrases like "bad things happen to good people" which implies we thing in terms of good people) but you could equally well cherry-pick examples of people's behavior where it seems they believe the opposite (the phrase "doing a good deed" implies that we think in terms of good actions).

So none of the five major claims of virtue ethics make it "just our intuitive morality". This is an attempt to load the scales by privileging your own position, like the Muslims who claim that everyone is born a Muslim and it's only when children are brainwashed by their societies that they become anything else.

So if we stop calling it "virtue ethics" and call it a better name like "intuitive ethics", is there any value to the claim "just use your intuitive morality"? Sooooort of, but not the type of value that is actually, well, valuable.

We can use our intuitive morality to determine we should not go around murdering little kids for no reason. This is good. But as a consequence, *no one is remotely interested in the question of whether we should go around murdering little kids for no reason*. No one goes to moral philosophers to ask that question. The very fact that it is solvable by intuitive ethics means that *it is a solved problem*.

The *only* reason anyone is interested in moral philosophy is because sometimes this doesn't work. Maybe we have sociopaths who are mysteriously born without intuitive morality. Or we have controversial moral problems like abortion where people intuitive moralities give very different answers. Or we have difficult moral problems like the Trolley Problem where many people's intuitive moralities just go "Hmmm, that's a really tough question". Or we notice that in olden times, people's intuitive moralities told them slavery was a-ok, including people like Aristotle who had put a lot of work into cultivating their facility of judgment, and we want to make sure we're not doing something equally awful ourselves.

To answer "Use your intuitive morality" in any of these cases ignores the fact that the set of problems where we need moral advice is the exact complement of the set of problems where using your intuitive morality is good enough.

And the set of senses in which “well, just apply as much intuitive morality as you can and hope it works” solves these kinds of problems is the exact complement of the set of senses in which people still feel like the problem needs to be solved.

If the only claim of “virtue” ethicists is “in the subset of problems where our intuitive morality gives clear and uncontroversial results, great, let’s go with those” then I agree with this claim.

If they claim any of statements 1-5 above, or that this generalizes to the case of difficult moral problems or problems anyone actually wants answered, they are going to need to present the evidence that I still maintain *After Virtue* lacked.

If a clod be washed away by the sea, Europe is the less

When Osama bin Laden died, I had to spend the next week or so listening to very preachy people talk about how it was wrong to ever take delight in the death of a human being, no matter how evil they might have been. About how anyone who thought that this was a cause for celebration ought to take a serious re-examination of their standards as a human being. About how it symbolized America's utter lack of civilization when people actually held impromptu parties in the street.

When Margaret Thatcher died, I had to spend the next week or so listening to very preachy people talk about how demands from Thatcher-worshippers that everyone else tone down their real feelings about her were ridiculous and offensive. About how much they loved this Guardian piece on misapplied death etiquette stating that "the demand for respectful silence in the wake of a public figure's death is not just misguided but dangerous". About how awesome it was that some Britons actually held impromptu parties in the street.

I have no data about the overlap between these two sets of very preachy people, but I bet it's larger than anyone with any faith in humanity left would naively predict.

Proving Too Much

The fallacy of Proving Too Much is when you challenge an argument because, in addition to proving its intended conclusion, it also proves obviously false conclusions. For example, if someone says “You can’t be an atheist, because it’s impossible to disprove the existence of God”, you can answer “That argument proves too much. If we accept it, we must also accept that you can’t disbelieve in Bigfoot, since it’s impossible to disprove his existence as well.”

I love this tactic *so much*. I only learned it had a name quite recently, but it’s been my default style of argument for years. It neatly cuts through complicated issues that might otherwise be totally irresolvable.

Because here is a fundamental principle of the Dark Arts – you don’t need an argument that can’t be disproven, only an argument that can’t be disproven in the amount of time your opponent has available.

In a presidential debate, where your opponent has three minutes, that means all you need to do is come up with an argument whose disproof is inferentially distant enough from your audience that it will take your opponent more than three minutes to explain it, or your audience more than three minutes’ worth of mental effort to understand the explanation.

The noncentral fallacy is the easiest way to do this. “Martin Luther King was a criminal!” “Although what you say is technically correct, categories don’t work in the way your statement is impl – ” “Oh, sorry, time’s up.”

But pretty much anything that assumes a classical Aristotelian view of concepts/objects is gold here. The same is true of any deontological rules your audience might be attached to.

I tend to get stuck in the position of having argue against those Dark Artsy tactics pretty often. And the great thing about Proving Too Much is that it can demolish an entire complicated argument based on all sorts of hard-to-tease-apart axioms in a split second. For example, *After Virtue* gave (though it does not endorse) this example of deontological reasoning:

> I cannot will that my mother should have had an abortion when she was pregnant with me, except perhaps if it had been certain that the embryo was dead or gravely damaged. But if I cannot will this in my own case, how can I consistently deny to others the right to life that I claim for myself? I would break the so-called Golden Rule unless I denied that a mother in general has a right to an abortion.

It seemed unfair for me to move on in the book without at least checking whether this argument was correct and I should re-evaluate my pro-choice position. But that would require sorting through all the weird baggage here, like what it means to will something, and whether your obligations to potential people are the same as your obligations to real people, and how to apply the Golden Rule across different levels of potentiality.

Instead I just thought to myself: “Imagine my mother had raped my father, leading to my conception. I cannot will that a policeman had prevented this rape, but I also do not want to enshrine the general principle that policemen in general have no right to prevent rape. Therefore, this argument proves too much.” It took all of five seconds.

Sometimes a quick Proving Too Much can tear apart extremely subtle philosophical arguments that have been debated for centuries. For example, Pascal’s Wager also proves Pascal’s Mugging (they may both be correct, but bringing the Mugging in at least proves ignoring their correctness to be a reasonable and impossible-to-critique life choice). And Anselm’s Ontological Argument seems much less foreboding when you realize it can double as a method for creating jelly donuts on demand.

Interestingly, I think that one of the examples of proving too much on Wikipedia can itself be demolished by a proving too much argument, but I’m not going to say which one it is because I want to see if other people independently come to the same conclusion.

Things I Learned By Spending Five Thousand Years In An Alternate Universe

When I was in high school, some friends and I decided it would be cool to start our own country, *Bridge to Teribithia*-style. As the idea gradually came into contact with reality, it degenerated into “simulate a country” and then “role-play a country”.

Then as more people got interested, it expanded into “role-play a planet”, and then back up to “simulate a planet”.

I’m not sure when it morphed into “get hundreds of friends to create and spend large portions of their lives in a fantastically-detailed alternate universe, building such close relationships that many ended up dating or marrying one another”, but it was sometime before the end of the sixth century.

…oh, right. Back in the second century we adopted a convention that one day in the real world was one year in our fictional planet of Micras. The sixth century would have been a little less than two years after we started, so early 2001. Since then various new societies have arisen with dozens of calendars of their own, but a few of us purists have kept the original one going.

And it is in that original calendar that we celebrated yesterday the turn of the sixth millennium. The year 5000 ASC. I – and others, there are still a few other people who have been around since the beginning – have been simulating Micras’ politics, mapping its contours, and writing its history for five thousand days.

It’s kind of terrifying how much of my life this one game has shaped.

It was my friend Eoin – Ard-Baron of Treesia and Fabon on Micras – who first convinced me to move to Ireland.

It was my friend Erik – founder and erstwhile Kaiser of Shireroth on Micras – who got a LiveJournal and convinced me to start blogging.

It was my friend James – Duke of Kildare on Micras – who first sat me down and told me I needed to start listening to *good* music and handed me some Ayreon and Nightwish – which shaped my musical tastes right up till the present.

It was my friend Ari – Duke of Straylight on Micras – who first linked me to Robin Hanson, Eliezer Yudkowsky, and the rationalist community.

They say that if the fool persists in his folly, he will become wise. And even wasting several hours a day in constructed worlds gives you a couple of benefits, if you keep it up for thirteen years. Insofar as I can make webpages, it’s from my time as Shireroth’s Minister of Information. Insofar as I know Photoshop and graphic arts, it’s from my time as head of the Micronational Cartography Society, which maintains the physical and political maps of Micras. Insofar as I can write, it’s because of years painstakingly composing treaties and fake history books. Insofar as I can argue or convince, it’s from debates in fictional legislatures about whether or not to go to war with distant countries with names like “Baracão” and “Novikrasniystan”, run by a bunch of poli sci majors in London or computer geeks in Belgium or even stranger people even further afield.

II.

More than that, though, it’s taught me about *people*. The total lack of rules or advance planning with which we constructed Micras gives it an amazing feature unmatched in any other role-play I know of: the game is exactly identical to the meta-game.

A country is a bunch of people coming together and claiming to be a country and doing country-like things (kind of like in reality). The king – or Shah, or President, or Premier, or Ayatollah – of the country is whoever can convince other people to call them the king and obey their orders (kind of like in reality). The country’s land is pretty much whatever land they can convince other people to accept they have (ibid). The constitution is whatever document you can convince everyone else to sign (…).

If one person wants to found their own single-person country, no one can stop them, but they’re less likely to be taken seriously or considered a Great Power. If lots of people come together to form a country, no one can stop them, but they had better be able to get along and agree on the rules. If you want to unite to ostracize somebody, no one can stop you, but you’d better be able to get more people on your side than they have on theirs.

If you want to claim you have a billion nuclear bombs, no one can stop you, but they’ll just say you’re a terrible simulation partner and ignore you when you say you bomb them. If you want to claim you are pure pacifists, no one can stop you, but then you better either have an alternate plan for protecting yourself (like strong allies) or be prepared to just absent yourself from the military simulation and annoy everyone else. If you want to write a history of your country that conflicts with histories everyone else has written, no one can stop you, but no one is going to take your history seriously either or build upon it or make it part of their canon.

As a result, while other geeks were learning how to calculate damage from Magic Missiles, I was learning *how to manipulate consensus reality*. I guarantee you one of these skills is more valuable than the other.

III.

The skill of manipulating consensus reality seems more or less identical to the skill commonly called “leadership”. It is easy to underestimate. The whole gag of the comic strip *Dilbert* is underestimating leadership. These brilliant engineers do the actual hard work, and then some idiot just says “work faster!” or something similarly dumb and gets hailed as a leader and paid a much higher salary and given credit for the group’s success

Micras has been a sort of laboratory for leadership – countries mostly between one and thirty people, rising and collapsing on a scale of months to decades. After thirteen years I am at no risk of underestimating leadership. I have seen countries transition from Great Powers to smoking ruins within weeks after a new and less competent monarch succeeds the old. I’ve also seen tiny city-states led by someone who understands the methods of persuasion and balance-of-power politics take over an entire hemisphere through soft power.

Occasionally I have even been a leader myself. It’s not hard on Micras – gather two or three friends, start a small country, crown yourself King. Or join an existing country, build a power base, and get elected Prime Minister. Or if that’s too much for you, it’s never too hard to find jobs on the regional level. My own homeland of Shireroth has long operated with five Dukes, each ruling a fifth of the country’s territory, and the positions are noncompetitive enough that during one particularly complicated conspiracy I managed to attain a few of them under false personae.

But I’ve also held the real thing – the Golden Mango Throne of the Kaiser of Shireroth, widely agreed to be the most complicated and unrewarding leadership position on Micras. For a role with no real-world power or consequences, it’s amazing how stressful, time-consuming, and life-sapping it can be – and how much you learn by holding it. If I thought they would take me halfway seriously, I would strike some kind of lucrative deal with a business school to let their students “study abroad” in a Shirerithian leadership position. If nothing else, it would knock some humility into them *really* quickly.

I titled this post “Things I Learned By Spending Five Thousand Years In An Alternate Universe”, so I guess I should get into some actual lessons. This is probably the only “leadership advice” I will ever give, and I have no idea if it transfers to the real world, but here you go.

I think the most important thing I learned about leadership is to avoid it. It’s stressful, everyone blames you for everything, and “getting to make decisions” sounds a lot better before you realize how banal and annoying 99% of decisions are. But I also learned that large organizations tend to have a position that pretty much controls everything from behind the scenes but doesn’t have to cope with the appearance of power. In Shireroth it’s called “Steward”. In Westeros it was “King’s Hand”. I don’t know about the USA, but I wouldn’t be surprised if it was “White House Chief of Staff”. These positions are a whole lot more fun, and surprisingly there’s a lot less competition for them.

Closely related is learning how many people are optimizing for appearance – which means if you’re optimizing for something else it’s pretty easy to strike compromises that give everyone what they want. If you’re fighting for control of a province, the compromise “your enemy gets an important sounding title like Archduke with almost completely ceremonial powers, and you get a boring sounding title like Undersecretary of Resource Management that controls the place’s economy and military” works a surprising amount of the time. Same with titling a bill “The X Party Wins Bill” and getting leading members of the X Party to support it and having the Y Party protest it angrily and *not have any policy proposals of the X Party in it at all*.

The third important thing I learned is to have a lot more respect for politicians and people in power. I think everyone should have the media perform a hatchet job on them at least once. It’s this really scary feeling when you know you’re trying to be honest and do the right thing, and yet you see how *easy* it is for a hostile writer to cast every single thing you do as corrupt and destructive. And how quick everyone is to believe them. And how attempts to set the record straight get met with outraged “how dare you give one of those typical sputtering non-apologies!”. It reminds me of those computer games where “ACCUSE” is just a button you press, and it doesn’t even matter what the accusation is or whether it makes sense. Once someone has invoked the *genre* of scandal, it will play out the same either way, proceeding deterministically along political lines until everyone reaches the usual compromise of agreeing you’re scummy and dishonest but not worth the trouble of impeaching.

The last important thing I learned is to *be nice*. It practically *never fails* that somebody who thinks they’re really cool joins Micras, makes fun of one of our admittedly disproportionate number of people with no real life or social skills, bullies and harasses them for months or even years… and then that person is the swing vote in an important election, or finds themselves sitting on a deposit of valuable rare earth metals everyone needs. My favorite cases are when neither of those two things happens, and the person just spends five years sorting out their issues and becoming smarter and more competent, and then ends up in charge of everything solely by their own merit. I am pleased to report they rarely forget how the bully behaved when they were young and stupid.

I don't know how many of these lessons work in reality. I think at least some of them do. But also, I don't think I learned any of these lessons on their own. They're part of meta-lessons where my brain learned to grasp these concepts of status and popularity and negotiation. I don't *actually* go into parties thinking "that guy there is the Duke of the room, and those guys there to the side are his Counts", but I do think my brain is using circuitry that it developed in part for those sorts of calculations, and goodness knows I needed to get that circuitry somehow.

IV.

All this probably makes it sound like all we do on Micras is plot and conspire against each other, but that's really only a small part of it. We are first and foremost conworlders, or geofictionalists, or that thing there's no non-awkward word for which involves creating detail for alternate worlds (and which one of my girlfriends has a college degree in, by the way!)

If leading a country trains the mind, conworlding trains the soul. In fact, I sometimes feel like a good conworld is like a projection of the creator's soul – like there's some sense in which a reader of the Silmarillion understands Tolkien as a person on a deeper level even than one of his friends or family members might.

Micras is a collective conworld, but it's a salad bowl and not a melting pot. Most people get their own country or province or island, develop its culture in their own image, and only then do they federate into nations or empires or hokey EU-style monstrosities.

The question "If you were a society, what kind of society would you be?" is strangely existential. Some people are bland liberal democracies. Some people are tropical island paradises. Some people are extremely efficient Singaporean city-states. But anyone at all interesting is something that has never quite existed before on Earth. Tolkien was the Elves. I don't know much about Iain Banks, but it wouldn't surprise me if he was the Culture.

One guy on Micras is a libertarian. He just sort of hangs around going "Yup, my country's government still isn't doing anything. Just hanging around punishing the initiation of force." It's very cute.

It makes you examine your soul, conworlding does. Over the centuries, changes in your outlook are mirrored by revolutions in your country's government. The problems debated in its universities and great books are the problems you struggle with every day. Sometimes your values and aesthetics drift, and some fictional philosopher mirrors the change across a span of worlds. Very rarely, it is the fictional philosopher who makes a good point that the real you is forced to consider.

So far in thirteen years we've only had one person totally lose touch with reality and start believing his country was real and worshipping the deities of his own constructed religion and whatnot. I am pleased to say he eventually made a full recovery and is now a physicist. But – and I can't speak for anyone else on Micras – there's always this feeling. If you spend ten years building up a culture in your own image, you start to feel at home there. If your real-world society doesn't fit you too well – if, like Bryan Caplan, you like to retreat within a bubble, then it's hard not to start thinking of yourself as a citizen of a civilization that exists only in your dreams. If Tolkien never spoke Quenya to himself when he was alone, I will eat my hat.

Somewhere in the ocean hundreds of miles north of the Shirerithian mainland there is a mountainous arctic island upon which thrives an emergent oracular techno-theocracy that calls itself the Shining Garden of Raikoth. Its priests wear a silver spiral around their necks as a sign of their dedication, and in solidarity with them I too wear the spiral. But that is as far as it goes. No deity-worshipping. No speaking to myself in constructed languages. Just the spiral.

That and spending five thousand days of my life on it.

Link: Bastion Union, a major micronational portal that hosts Shirereth
Link: A description of the constructed society of Raikoth

Floor Employment

When I was in college, I had two career plans lined up. First, I could become a philosophy Ph. D and eventually professor. Second, I could go into medicine.

I still remember the moment I chose the latter. I was hiking the summer of my junior year, and I decided to spend a day just comparing pros and cons and coming up with a final conclusion, and although I can’t remember the reasoning involved I decided it was *definitely* the doctor one.

And that went well, because now I hear all grad student would-be professors are kind of doomed, philosophers not excepted. But I still give some thought to my poor counterfactual self, lost in a hostile career market with no useful skills.

And unfortunately a few of my real friends, grad student and otherwise, are in that position right now, which makes my heart sink every time I hear from them. I enjoyed reading Louie’s story of how he got rich working in Australia, because it suggests there are still ways to turn your life around really quickly and without much fuss. Jobs you can just walk into – no sending out hundreds of resumes and having your soul slowly sucked out with each interview you don’t get – that can give you the financial cushion to reach some more permanent state of self-sufficiency.

But people who have tried to replicate Louie’s success have had mixed results, and besides not everyone can go to Australia.

So I’ve been thinking about this idea of “floor jobs”. By a floor job, I mean something that puts a floor in how bad and desperate your life can be. As in “I won’t be unemployed forever, I can always go do X”. Or “I won’t be working at this miserable fast-food place with this boss I hate forever, I can always just do X”. A lot of this depends on just how unlucky you are – the floor for high-IQ people with savings and college degrees is quite a bit higher than the floor for less smart people with no qualifications who need something now.

But floor jobs all need a couple of characteristics to count. First, they don’t need formal specialized education: “engineer” is not a floor job because you need to be in school for several years getting an engineering degree. Second, they are all pretty easy to get – anywhere from “walk right in” to “send your resume to a few places but they’ll probably accept you” – none of this heartbreaking hundred-resumes no-interviews thing. Third, they all have *some* advantage over let’s say retail – either high pay, good working conditions, or medium-to-high social status (very few have all three).

These are some of my candidates for floor jobs right now. Disclaimer-1: with one exception, I know nothing about these jobs except what I hear second- and third- hand and through the Internet. Disclaimer-2: the phrase “floor” is not meant to imply that these are bad or dishonorable jobs, only that they are easily accessible and have some advantages.

English Teacher Abroad

I start with this one because I did it and loved it. Your ability to speak English is actually a heavily desired and respected skill in some parts of the world. Many of these parts of the world have very low costs of living and exotic scenery and friendly people.

The qualifications required vary with the jobs. Most countries prefer people with college degrees. A few don’t. Many schools prefer people with some sort of TEFL qualification (you can get an only-sort-of-fake one with a one week online course for $50. Others don’t care about this. You don’t need to know the language of the place you’re going, and in fact some schools actively prefer you don’t (on “English immersion” principles).

Visas are going to be the problem. It’s medium hard to get a job with a foreign English school in your home country; only the really big multinationals will bother recruiting abroad. One popular strategy, which works better in some countries than others, is to go to the country you’re interested in on a tourist visa, interview there, and let the school that hires you change it to a work visa.

Your worst-case scenario, if you have no qualifications and no tolerance for visa nonsense, is Cambodia. As I understand it, schools in Cambodia will hire high school graduates with no previous experience, and visas are handed out like candy.The only downside is that you do have to live in Cambodia.

If you have a couple more qualifications, time, and tolerance for nonsense, China is a good alternative. Japan used to be as well but I hear it’s getting a lot tougher (it pays really well, though) . Other places I hear nice things about include Thailand, South Korea, and Costa Rica.

Bottleneck: Money to get to foreign country, visas. College degree really helps.

Programmer

Alyssa Vance makes the case for becoming a programmer if you don’t like what you’re doing now.

It’s high-paying and fun, you can learn it on your own without worrying about college degrees and such, and there’s lots of demand.

The downside is that you have to have a spare six months or so to start learning programming, the willpower to actually do so, and the analytical skills to become good at it. Most people probably won’t.

If your limiting factor is time or self-motivatedness, schools like Dev Bootcamp offer to bring you from zero to full programmer capable of getting a $90K job in about two months – but the going is really hard and the tuition is just above $10K. App Academy is similar but has no up front cost – you pay them after you get a job. A warning: some friends here mention that these companies’ claims of “90+% job placement rate” are misleading, as halfway through they expel anyone who they don’t think will get a job in order to keep their numbers up.

Bottleneck: Time, money, natural programming ability, motivation

The Military

At least in America, if you’re young and fit the military is almost always happy to have you. People generally sign up for an eight year term, four of which are active and four of which are in the reserves. People who get good scores on occupational tests can usually avoid being front-line infantry if they want and learn useful-in-the-outside-world skills like technology or engineering. And it’s an honorable and high-social-status profession that provides good veterans’ benefits afterward.

People in high school (or part of the way through college) can take advantage of officer recruitment scholarships through ROTC.

If you’re from one of those mean awful countries that spends their money on food on education and universal free health care instead of subsidizing a people-hungry military-industrial complex, all is not lost. The French Foreign Legion takes anyone, French citizenship and French language ability not required. Best of all, your past (including criminal record) officially vanishes when you join the Legion, and you’re allowed to join under an assumed name. Between being in a legion, having a secret identity, and fighting evil-doers, you’re *basically* a superhero.

Bottleneck: Physical fitness, emotional fortitude, conformity with your ethical system, must like sand. French Foreign Legion only wants men, and you have to fly to France to sign up.

Camperson

…is the gender-neutral version of “camgirl”. Another one that wins on ease of access. You sign up on a website. You show them a driver’s license to prove you’re 18. You set up a webcam. And people pay to watch you do sexy things.

In theory. Obviously this works better if you are female and very pretty. Very few women are clients of camsites, so if you are male your choice is pretty much to cater to gay men – which can also be lucrative if you’re up for it.

There seems to be a very wide salary range based on your attractiveness, gender, ability to self-promote, what you’re willing to do on camera, and how long you’re willing to spend online, not to mention just plain luck. It can range anywhere from “barely able to support yourself” to “there is no way people really give other people this much money for just sitting in front of a camera, is there?”.

This thread on Reddit has some interesting stories, and Ozy sometimes talks about zir experiences on zir blog.

Bottleneck: Attractiveness, conformity with your ethical system, low inhibitions

Anything In North Dakota

The US unemployment rate is 7.6%. The North Dakota unemployment rate is 3.3%. In some of the western counties, it's less than 1%. The reason is obvious: they struck oil there a few years ago, and now the state is desperately trying to attract as many people as possible to help extract it and support the burgeoning community of oil extracters. Restaurants and convenience stores are desperate for workers and the pay reflects that.

Bottleneck: There's not a lot of housing there, plus you have to live in North Dakota and it sounds miserable

Oil Rigs

Speaking of oil, and of living in terrible places, oil rigs are hiring. If you're willing to live on a tiny and extremely flammable platform in the middle of the ocean for long stretches, they will offer you jobs that start around 30K quickly go up to about 100K, no particular qualifications required.

Rigzone.com talks a lot about engineers and such, but there seems to be a pretty big demand for just warm bodies to perform menial tasks as well, and the job includes a good learning curve where you have ample opportunities for advancement (insofar as moving to another part of the same couple-thousand-square-meter platform is an "advancement").

And if you really don’t like the sea, there are always onshore oil rigs to consider, although I don’t think they pay quite as much.

Bottleneck: Willingness to live on tiny flammable platform in ocean.

Truck Driver

The most attractive thing about truck driver is probably the education arrangement. You say you want to become a truck driver. If the trucking companies want you, they will pay for your training course in exchange for a certain portion of your first year’s salary. And they probably *will* want you – trucking is a “http://jobs.aol.com/articles/2012/10/31/employers-desperate-to-fill-truck-driving-jobs/”>high demand industry right now.

Most people would find trucking unpleasant. But for people who like quiet and alone time and whose nightmare job is having to satisfy hordes of chattering customers to an exacting schedule every second of the day, it might be more tolerable than other options.

Bottleneck: Good driving record, ability to tolerate lifestyle

Tutor

This one sort of breaks the rules in that it requires a lot of activation cost and agency and set-up and pandering. But many people I know have had very good experiences (and make very good money) tutoring (mostly rich) kids.

Obviously you have to have something you can tutor people in. But if you got into a good college or had a very good score on the SAT or similar standardized tests, or you have a college degree in some subject like Math or English, that’s probably good enough. Bonus points if you know something like Biology and can tutor aspiring medical students.

It can be exasperating trying to find people who need the tutoring, and the schedule can be kind of complicated, but wages seem to go up to three digits an hour in the best cases, and there are some sites that will help you.

Bottleneck: Book smarts, organizational/self-marketing ability, living in a place full of would-be over-achievers

Repetitive Computer Task Doer

This one wins the “ease of access” competition by a mile at the cost of losing the “acceptable pay” competition by a light-year.

Sites like Amazon Mechanical Turk and ODesk allow you to log in, do some basic data processing exercises for someone, and then get a small amount of money. Typical tasks include taking surveys or writing short articles.

The good news is all you need is a computer and a little bit of time, and there’s lots of work to be done and not a lot of people asking about qualifications. The bad news is that the pay is terrible, way below minimum wage. Depending on how quickly you finish them, some tasks can give you $1 to $3 an hour. Others pay less.

This is probably not a good way to support yourself. But if you happen to have support, or are just really desperate for a short period of time, you might be able to make enough money for some food or something with this one.

Bottleneck: Internet access, willingness to tolerate low salary

Arguments About Male Violence Prove Too Much

[CONTENT WARNING: rape, violent crime, racism]

[EDIT HISTORY: This piece was widely circulated and critiqued after first being published, and I received a lot of feedback, some of which was good and some of which was bad. I have entirely rewritten the piece to try to respond to some of the complaints, especially those in the comments and those raised here. The original version of this post, without which some of the reactions and complaints will not make sense, is available here. The bottom of that post gives more information on particular edits. Thanks to everyone who gave helpful feedback both positive and negative.]

From this article:

> When the odds of being assaulted are 25 percent, something is dangerous. If any other activity or object presented the same odds of injury or death, then a revolution would be ignited against it. If one-fourth of Americans faced armed robbery in their lifetime, then you'd better believe armed robbery would be a major national issue covered everywhere in the media, and it would be right up there alongside the economy and national defense in the presidential debates.
>
> We're all willing to make a strong, concerted efforts to see that safety is followed in cases like these, with no margin allowed for error. It's a shocking contrast to how we deal with women's safety from the men who harm them.

> It makes me wonder, what if men were declared as a public safety hazard?
>
> Could you imagine if they were recalled? Pulled off the street? “Sorry sir, you’ll have to come with me; we’ve had reports that men have been raping, beating, and killing women, and we can’t take the risk that you will, too.” Yes, it’s a ridiculous idea. But men are way more dangerous than Tylenol [which was recalled for being dangerous].
>
> You may also say, “There are plenty of men out there who don’t abuse or sexually assault women — what about them?” I say: Well, what about them?

I can’t quote the whole thing, but you should probably read it if you want a clearer picture of what’s being talked about.

So when I read this article, I feel a couple things. I feel sad about the high prevalence of rape and domestic violence, of course. But I already knew that one. I also feel other emotions. As a man who hasn’t done any of these things, I feel kind of scared and singled out and unfairly guilt-by-associationed.

I know this isn’t the first time this has happened. Articles That Tar All Men With The Same Brush are pretty common, followed by Men Who Get Offended, followed by People Telling Them They Are Wrong To Be Offended Because The Problem Of Rape Is Much More Important Than The Problem Of People Getting Offended.

But, well, I *still* feel unfairly guilt-by-associationed. So here’s an intuition pump to try to communicate why.

A Visit To Racist Dystopia

Suppose you woke up one morning and started hearing public service announcements on your radio: “Black people are defective! Black people are a public safety hazard! Black people commit lots of violent crimes! The police should just arrest all black people, because they’re too dangerous to allow on the streets!”

You do some investigation and find that it’s just a small fraction of the population that believes this – maybe 10% – and they’re not immediately advocating any actual consequences or policy toward black people. So that’s good. But you keep hearing this same message. All your favorite blogs publish a steady stream of wildly popular articles trying to “helpfully explain” to black people why all white people are justified in fearing them. Any black person in college has to walk past posters every day listing their name and reading THIS PERSON IS A POTENTIAL MUGGER – and when they complain, they are met with indifference and an administration claiming it is merely a helpful way to raise awareness of black violence.

Hopefully you, like me, would be horrified by this state of affairs. Although certainly crime is a problem in many places, and although crime is worse in poor neighborhoods which are also disproportionately minority, this is an offensive and unproductive way of thinking about the problem.

The Value Of The Analogy

If we had to specify our exact complaints against Racist Dystopia, I think there would be at least three good ones. First, we would want to protest that only a tiny percent of black people are guilty of violence. Second, we would want to protest that people of all races are capable of violence, and that the existing campaign unfairly portrays it as solely a black problem. Third, even apart from those two complaints and even assuming raising awareness of racial violence is something we want to do, there would be ways of sending that message that encourage stereotyping and sweeping judgments and ways of discouraging them, and the current campaign seems specifically intended to promote the former.

Let’s start with the first complaint: only a tiny percent of black people are guilty of violence. How tiny? Statisticians project that about 30% of black men will go to prison sometime in their lives. Somewhere around 30-40% of prisoners are serving time for violent offenses, so if we combine those two numbers (something which requires a few assumptions, but I can’t find the statistic directly so it will have to do) we get that about 10% of black men will go to prison for violent crime sometime in their lives. If we assume that black women do not go to prison (I can’t find good data on this), then about 5% of black people will go to prison for violent crime at some point.

This number is very similar to another number: according to an article from the early 2000s in a feminist blog, about 4.5% of men are estimated to be rapists.

Our second complaint is that violence is a problem committed by people of all races: most notably to the public service campaign, it is committed by white people as well. But as noted before, violence concentrates disproportionately among poor populations, these are disproportionately likely to be minorities, and the effects scale up. America is about 50% men/50% female. Suppose that America were 50% black/50% white. We know that black people currently commit homicide at a rate 7.5x greater than white people, so in this hypothetical society – in the implausible case where nothing changed about neighborhoods or poverty or income gap – 88% of murders would be committed by black people. Murder is an unusually good statistic because almost all murders are investigated and so there’s a low chance that much of the differential is due to racist policing, but the numbers are about the same for other violent crime. For example, in New York City, which is approximately 50% white, 25% black, and 25% other, 78% of all shootings are black compared to 2.5% white. If we extrapolate New York City into a hypothetical 50% black/50% white society, we find that the black half would commit about 97% of the shootings and the white half about 3%. Let’s average these two statistics and say that in our hypothetical society where race works like gender, 95% of violent crime would be black.

And once again, these numbers are in the ballpark of male/female rape statistics. What percent of rapes are committed by men? This is very hard to determine, because rape by women is almost never reported (victim is too embarrassed) and almost never prosecuted (people just laugh and say they bet the guy liked it). I have seen claims from 99% male (which seems very high) to 75% male (which seems very low). I do not think that 95% of rapists being men to 5% women is an impossible number. Other forms of violence are even less male-dominated; male-initiated and female-initiated domestic violence seem to be about equal.

The last complaint we might have against Racist Dystopia is not statistical but moral. *Even if* it was socially necessary to raise awareness of violent crime, and *even if* everyone was so racist that the decision to fixate on black crime had already been made, there are ways to do it that are super-awful and ways to do it that are only kinda-awful. I think the most important criteria for landing a campaign in the coveted second category would be:

– Absolute avoidance of any claim or implication that the problem is with *all* minorities and extreme and frequent repetition of the fact that the *overwhelming* majority of minorities are non-violent.

– A focus on the fact that white people can commit crimes as well, made *at least* proportional to the amount of crime they actually commit, and if possible even a little more so in order to hammer home the message that crime is not a racial problem.

– A focus on additional reasons why you need not be terrified of every single black person you met. For example, if the majority of muggings tended to occur in a few very specific situations, like after both the mugger and the victim had been drinking, or after the mugger/victim had accepted an invitation to go to the victim/mugger’s house, that would have a pretty big effect on whether or not you should live in fear of your random black co-worker or college professor.

My thesis is that we should make these same complaints against efforts that try to tar all men as rapists.

How To Use The Analogy

There’s a possible fourth complaint here in which the two situations are *not* similar: black people are a really oppressed group. The more one spreads fear and stereotypes about them, the more likely it is you awaken someone’s latent racism and cause damage disproportional to the “limited” fear and stereotypes you were trying to convey. For example, traditionally people have been pretty quick to engage in hate crimes against black people based partly on stereotypes exactly like the one mentioned above.

One answer to this objection might be that some men are murdered, sometimes horribly because of the climate of fear around male rape. But since I’ve tried to stick to statistics thus far, it would be dishonest to claim that this happens in anywhere near the numbers that would be necessary to make the analogy stick.

A better objection might be that the issues of disprivilege and oppression, while important, are *not necessary* to make the Racist Dystopia horrible. Imagine a world in which we somehow magically prevented any white people from having their opinions influenced by the public service campaign vilifying black people. They somehow continue at exactly the same level of racism they had before, and the “only” problem with the campaign is that black people have to listen to themselves be attacked all the time and get “educated” on the importance of crime prevention if they complain. This world would be *less bad* than the world in which they also had to deal with the additional racism, but it wouldn’t be a walk in the park either.

Even removing the racism angle, the Dystopia above is still bad just because of the pain and dehumanization it causes people to have to read about their group in terms of some evil Other who must be a threat to all right-thinking people.

Men are blessed with many positive role models, but the divide between social and structural power is worth taking into account here, and the sort of men who are exposed to feminist articles like the one above (exactly the sort of men who are in the best position to help women!) have to take in quite a bit of information. For example, this morning when I checked Facebook I was helpfully suggested links to the “all men should be taken off the streets” article above, a blog called “Creepy White Guys”, an OKCupid app that claimed to be able to tag people you looked at as “likely rapists” based on a sketchy machine learning claim about their profiles, and a link to an article called “Straight White Male: The

Lowest Difficulty Level There Is”. My RSS reader then directed me to my favorite blog, which had suspended its usual discussion of abstruse philosophy to host an article called “Submissions on Misogyny”, which included bits like where someone talked about her boyfriend raping and physically assaulting her and then claimed “You might think my ex was a sociopath, but no — he’s a normal male”.

This was, more or less, a typical day for a somewhat liberal guy on the Internet. So this idea that men never have to hear anyone speaking against them and these sorts of “all men are dangerous and defective” articles are just an unexpected breath of fresh air no longer track reality, if they ever did. And if you think that a man can’t possibly be hurt by seeing people insult and belittle his gender (which, no offense, is actually a kind of patriarchalist opinion right there), all I can say is that my personal experience begs to differ.

This is *not* a demand that people stop talking about rape, or even that they stop talking about the importance of preparedness for rape! I know my feelings aren’t as important as *that*!

But it is a polite request that you follow the three suggestions I would have made to the Racist Dystopians above:

– Absolute avoidance of any claim or implication that the problem is with *all* men and extreme and frequent repetition of the fact that the *overwhelming* majority of men are non-violent. 95% of men have never raped anyone and would be horrified by the idea. Insofar as you give yourself the task of “warning women what to expect from men”, one thing they should expect is to start with a 95% probability a given man is not a rapist, and then start adjusting from there based on evidence.

– A focus on the fact that women can commit rape and gendered violence as well, made *at least* proportional to the amount of rape/violence they actually commit, and if possible even a little more so in order to hammer home the message that what we're against is rape itself and not This One Hated Out-Group.

– A focus on additional reasons why you need not be terrified of every single man you meet. This is something that feminists already do very well, in that they help explain what the warning signs of rapists are and what situations and requests are red flags for someone who might try to rape you, but this tends to be forgotten in articles like the one above which focus on scare-mongering the idea that *it could be anybody!* While it's true that it could be anybody, it's also important to keep in mind that it is somewhat more likely to be some people than others.

It's easy to see why doing this would benefit men, and I admit that's mostly why I'm writing this, but I can think of many reasons this would be good for women as well.

First, encouraging a woman to fear and distrust all men is probably not a useful strategy in a society that's fifty percent male, especially if that woman happens to be heterosexual. A strategy of "be aware of this possibility and of the warning signs, but also that most of the people you interact with are nice and trustworthy" is probably both psychologically and socially more healthy.

Second, women have a vested interest in fighting sexism and sexist stereotypes. Sexism is basically a flawed cognitive algorithm. It's the tendency to think "I can think of a bunch of people of this sex who do X, therefore I'm just going to classify that sex as X-doers and promote that idea to society." A big part of fighting sexism is discouraging this process. Saying "No, you can't just say that because some women like cooking in this society, cooking is a Thing Inherent About Women. Further, we can't even create a climate where women are constantly portrayed as cooking and doing things relating to cooking, because that's going to make non-culinary women feel bad." And if you laboriously train people out of this habit of thought, and then say "But definitely do this for men, they don't count because they're privileged oppressors", it's not going to work for women either. You're creating a natural Stroop effect where people have to keep conflicting category-based rules in their head.

Third, and most speculatively, I'm kind of worried that this sort of stereotype actually promotes rape. There was a very interesting study where researchers interviewed some people about their relationships, and then told half the subjects (at random) that they had been commendably faithful, and the other half that their actions suggested they were of an unfaithful personality type and their mental infidelity might destroy their relationship. Then they asked the subjects for their opinions on infidelity. The subjects who had been told they were commendably faithful told them fidelity was extremely important to them; the subjects who had been (randomly!) told they were unfaithful told them that fidelity wasn't important to them and that infidelity wasn't a big deal anyway.

The researchers theorized that this was a process called "cognitive dissonance". Most people like themselves and want to continue to like themselves. If they are told that they, or their group, has a particular flaw, then instead of ceasing to like themselves it may be easier to just decide that flaw is not such a big deal and they can have it while continuing to be the awesome people they secretly know they are.

If rape is portrayed as inherent to men in some way, men have two mental choices. They can think “darn, I guess I got the evil gender.” Or they can think “well, if *my* gender does it, I guess it isn’t so bad”. Psychology suggests there will be at least some small tendency to react with the second.

I don’t know how important this effect is, but given that just stating the case for rape awareness respectfully and non-prejudicially is an easy and desirable solution anyway, I don’t see why one should take any chances.

Disclaimers That Should Not Be Necessary, But Are

I am not trying to compare the experience of men to the experience of black people in any way other than the very simple numerical comparisons listed.

I am not actually saying that black people should be seen as criminals, I am using this as an example of a bad argument in order to show that tarring all men with the crime of rape is also a bad argument.

I am not claiming straight white men are not privileged or do not have things easier than other groups.

I am not apologizing for rape or claiming it is anything other than really bad. I am not denying women the right to avoid or fear men if that is what makes them comfortable.

I do not understand “rape culture”

Trigger warning: discussion of rape

Someone brought up “rape culture” in the comments to my last post.

Now I really should stop criticizing feminist ideas, because doing so is both mind-numbing and mind-killing and makes me feel like a bad person. It apparently gets me emotional enough to start making bad arguments, one of which I decided to retract in the last post (see edit history at the bottom). And after this I promise not to write another post about it for a couple of months. But I just wanted to express how confused and skeptical I am of the term “rape culture”.

I mean, there’s one way it sort of makes sense, and that is in naming a certain small and circumscribed cluster of people. I’ve heard the phrase NASCAR culture to refer to people who dress up in NASCAR clothing and go to races and obsess over the sport. Total annual NASCAR race attendance is about 1% of the United States, so calling 1% of the US “involved in NASCAR culture” seems fair enough, though it also seems more philosophically perilous than just saying “1% of people like NASCAR”.

Compare the term “car culture”, as in the phrase “America is a car culture”. It means that most Americans like cars a lot, and even those Americans who don’t like cars have absorbed pro-car memes and gladly accept propositions like that everyone should have to pay taxes to help maintain the highway system. And that it’s really hard to, say, bike to work everyday, and most people have agreed to be okay with that as a society because we implicitly accept that cars are a part of the natural order. Here the claim isn’t that 1%, or 10%, or whatever, of America is a car culture. It’s that American society itself is a car culture in some fundamental way.

If someone wants to point out some rapists or rape apologists and claim they are rape culture in the same way NASCAR fans are NASCAR culture – well, this is still asking for the term to be abused. But when people try to pull the second meaning – *America* “is” a rape culture – this is the part that has me throwing up my hands and wondering what the heck they are thinking.

As I understand it, here are the claims involve in “rape culture”.

1. Society treats rape as less horrendous and more excusable than other crimes
2. The criminal justice system is unusually loaded in favor of rape suspects and against victims
3. People are more willing to blame rape victims than victims of other crimes
4. Misogyny in society causes sexual objectification of women, which latently condones/promotes rape
5. The general tolerance of rape is a sign that society is biased against women and doesn’t care about their problems

Every single one of these claims seems to me diametrically wrong.

Claim 1: Society treats rape as less horrendous and more excusable than other crimes

I’ve never seen *Dexter*, but I hear it’s about a serial killer who goes around murdering people but is otherwise a pretty neat guy.

I’ve never played *Grand Theft Auto*, but I hear it’s about a thief who goes around stealing people’s cars and stuff but is otherwise pretty likeable.

I’ve never seen *Silence of the Lambs* (I HAVE NO EXPOSURE TO POPULAR CULTURE, SORRY), but I hear it’s about a really urbane and sophisticated cannibal who eats people’s brains and who became one of film’s most popular and respected characters.

Sweeney Todd (finally, something I’ve seen!) combines theft, cannibalism, and bloody mass murder into a huge extravaganza of crime, and people happily show it to their teenage kids as a fun family movie (as well they should, because it is great)

So having likeable and sympathetic murderers, thieves, and cannibals is totally a-okay. You can even get away with likeable and sympathetic Nazis – think Franz from *The Producers*. Now, tell me, can you imagine a popular mass-market TV show whose hero was a serial rapist? How about a computer game in which the object was to rape as many people as possible?

Let me take a different tack. I have a file on my computer where I collect jokes that I like. I’m pretty hard to offend, so the jokes aren’t really sorted by offensiveness in any way. Looking through it, I see two jokes about cancer, one about AIDS, one about bestiality, three about murder, one about natural disasters, one about terrorism, one about cannibalism…and zero about rape.

I have no doubt that rape jokes exist, but you probably need to be listening to a particular sort of shock jock for them. As opposed to jokes about murder, jokes about cannibalism, jokes about terrorism, et cetera, which are so utterly neutral you can find them in children’s books if you want. Even Holocaust jokes are more common and more widely accepted than rape jokes.

My point here is that far from being unusually blase about rape, society treats rape as a special kind of evil, worse than murder, worse than cannibalism, worse even than the Holocaust, something that is *never* sympathetic and *never* funny. Further, it is practically alone in being treated this way.

I’m not *complaining* about this. Because there are so many rape survivors likely to be traumatized at discussion of rape, this is probably a good thing. But ignoring this reality and claiming we are in fact more willing to excuse lighthearted discussion of rape than other crimes is just false.

Claim 2: The criminal justice system is unusually loaded in favor of rape suspects and against victims

Conviction rates for rape are similar to or higher than those for other forms of violent crime. In Britain, About fifty seven percent of rape trials end in convictions, compared to fifty six percent of trials for other violent crimes.

The claim that the criminal justice system is loaded against rape survivors mostly comes from the difficulty of getting cases to trial in the first place. But there are more charitable explanations for this.

Most rapes are crimes without witnesses. If the accused claims the sex was consensual then even DNA cannot provide corroborating evidence against this story. The courts are presented with a “he said” / “she said” dilemma. Because the legal system enshrines the concept of “innocent until proven guilty” – that is, erring in favor of the defendant when a guilt cannot be established “beyond a reasonable doubt” – this situation usually results in acquittal.

Therefore the police have two options in a case where someone comes to them to report a rape but has no corroborating evidence.

The police can tell the survivor there is not enough evidence to press charges, thus saving them the trouble of a long and psychologically traumatizing trial which will inevitably end in an acquittal. If they go this route, then people accuse them of rape culture because of their “culture of skepticism” where they doubt all rape allegations.

Or the police can bring the case to trial, in which the jury is asked to distinguish between the prosecution’s claim that an assault was rape, versus the defense’s claim that it was consensual. If there’s no good way to establish what actually happened, the trial will usually devolve into character assassination – that is, the prosecution will try to prove the accused is the kind of person who would rape and then lie about it, and the defense will try to prove that the survivor is the kind of person who would consent to sex and then lie about it. If they go this route, then people accuse them of rape culture because of their “slut shaming” in which it is acceptable to character assassinate a rape survivor.

But I have yet to hear a good proposal about how to avoid these problems without dropping the suspect’s right to be “innocent until proven guilty” and have his “guilt proven beyond a reasonable doubt”. If you keep those in place, then somebody somewhere along the line has to be responsible for doubting the survivor’s story, and in many cases, even in the overwhelming percent of cases where rape really occurred, the evidence at hand just won’t provide any beyond-a-reasonable way to dispel those doubts.

The failure to remove “innocent until proven guilty” is not a sign of a special and unusual “rape culture” in which rape trials have a higher burden of proof. Indeed, rape shield laws make rape trials a special case where the victim has more rights, and the accused fewer, than in other sorts of crime. These are followed vigorously, sometimes off a cliff. Other than that, it’s just insisting on the same standards as every other sort of trial.

Some might claim that rape is a special case where burden of proof is unnecessary. But anti-death penalty groups have found that in fact many people are falsely convicted – and even executed for rape. The Innocence Project estimates that between 8-15% of rape convictions are false, and many of the people found to be falsely convicted were going to remain in prison for the rest of their lives (and needless to say a disproportionate number of them were minorities). The Project notes this is somewhat higher than for other crimes like murder. These numbers correspond pretty well with the generally accepted statistic that about ten percent of rape accusations are false (uh, assuming our criminal justice system has an accuracy of about equal to chance. We should probably do something about that.)

So although rape cases are incredibly heart-breaking and there are no good options, the criminal justice system seems caught between a rock and a hard place. I see no evidence for rape culture here either.

3. People are more willing to blame rape survivors than victims of other crimes

When I lived in Ireland, my friend got his bike stolen. He was immediately subjected to a barrage of questions like “Did you lock it right?”, “Are you sure you locked it right?” “Was it in a really visible spot?” “Don’t you know that brand of bike lock is easy to cut through?” and “Why did you even keep biking when you knew there was so much bike theft in this area?”.

This seems to be my experience with most types of crime. If you get mugged in a dark alley late at night, people will tell you “Well, duh, it’s your own fault for walking alone in a dark alley late at night with lots of money on you.” If you get carjacked, people will tell you the brand of car alarm you should have bought.

And goodness forbid you get scammed. People will either tell you it’s because of your own greed (“You can’t scam an honest person”), tell you all sorts of stupid signs to look for (“Did the guy have a scammy aura about him?”), or refer you to the Better Business Bureau and a bunch of other websites that no sane person possibly checks before every single time they do business with an honest-looking person.

Cracked even thinks this is an appropriate topic for a comedy article – here’s Four Types Of Victims We Have To Stop Feeling Sorry For. People who get mauled by animals are number three – why were they hanging out near animals, anyway?

In fact, it seems that in every case where someone has just had something bad happen to them, it’s a tradition to tell them ways they could have prevented it.

I do not want to justify the “advice” people give rape survivors. Some of it seems broadly good (if you’re meeting a strange man for a date or something, do it in a public place), but much of the rest seems well-intentioned but actively wrong (it’s been pretty well established that the clothes you are wearing do not affect sexual assault rate). Whether the advice is good or not, giving it to someone who has just been traumatized is without a doubt a stupid decision, and giving it in place of punishing an actual perpetrator is an obvious miscarriage of justice.

But as far as I know, this stupid decision is only ever criticized in the case of rape. The difference between rape and other forms of tragedy is not that rape is the only tragedy in which people try to advise survivors how they could have avoided it, but that rape is the only tragedy where a person giving such well-intentioned but stupid advice gets their name plastered over the national news and has a Slut Walk organized against them.

If rape culture means people being unusually tolerant of or unsympathetic to rape, I cannot find a sign of rape culture here either.

4. Misogyny in society causes sexual objectification of women, which latently condones/promotes rape

Split this one into two categories. Why is there so much sexual objectification of women? And does sexual objectification of women implicitly condone rape?

The reason there is so much sexual objectification of women seems really really obvious. *People really like sex*. That’s all. There is no need for any more conspiratorial or uncharitable explanation.

If you look at gay men, they sexually objectify men *all the time*. If you look at products aimed at women, whether it be women’s magazines, perfume, whatever, once again, they’re plastered with images of sexy men.

Heterosexual men have more social power and money than gays or women. They also are perceived, rightly or wrongly, as being especially obsessed with sex. That means if you want to sell thing to people with lots of money – the best type of people to sell things to – using pictures of sexy women is pretty much your best bet. There is no further conspiracy theory required for why people present images of sexy women in society. The idea that it has anything to do with secretly being okay with rape is somewhere around Roswell-level plausibility.

Nor does the sexual objectification of women necessarily imply that rape is okay. Let's take *actual* objects. Our society is obsessed with consumer goods. It demands loudly and viciously that people try to get as many consumer goods as possible at every opportunity it gets. But no one implies that our society is secretly condoning theft.

Saying "Desire consumer goods!" is a very different message from "Be willing to break the law, hurt others, and commit morally repulsive actions in order to obtain consumer goods!" Likewise, putting up pictures of sexy women does not have to be part of a rape culture where people imply that raping those same sexy women is okay.

5. The general tolerance of rape is a sign that society is biased against women and doesn't care about their problems

Society is obsessed with gender and gender-related issues which is exactly why rape is treated as so much more interesting a problem than anything else.

One especially interesting data set is information about cancer funding. Here's a graph of research dollars spent on cancer type per deaths caused by that kind of cancer.

Breast cancer gets a very disproportionate amount of funding compared to other, deadlier types of cancer. No doubt this is because of successful initiatives like the Pink Ribbon campaign, but these initiatives are themselves due to the fact that *any gendered issue is naturally more interesting than any ungendered issue*. My guess is this is also part of why prostate cancer (gendered issue relating to men) is in second place, although that could also be its unusually high morbidity/mortality ratio.

The point of this graph isn't to knock breast cancer research, but instead to note that people disproportionately ignoring women's issues because they don't care about women *is the exact opposite of the way the world works*.

Google Trends shows that ever since the national news story broke, more Americans are searching for information on the Steubenville rape case than on the entire country of Syria, which you may remember has during that period been undergoing a bloody civil war which risks turning into World War III at any moment. Think about it. A single rape case is more important to our culture than a war in which somewhere between 70,000 and 120,000 people have been killed.

This is not the sign of a "rape culture" that dismisses and excuses rape. This is the sign of a culture that has fixated upon it, placed it somewhere between child pornography and al-Qaeda in the List Of Things Everyone Must Talk About All The Time. This is the sign of a *national obsession*.

If you want to accuse the people in Steubenville who didn't do anything to help the poor victim of being part of a "rape culture", then go ahead (but do remember that the same thing happens in every type of crime and is a well-known cognitive bias having nothing to do with rape *per se*).

But if you want to accuse the American population that paid twice as much attention to it as to the entire Syrian Civil War of trying to deny that rape exists or push it under the rug or something, *what are you thinking?*

Social Justice For The Highly-Demanding-Of-Rigor

My last two posts have led to a lot of anti-feminist activists getting linked to my blog, so this would be a more hilarious time than usual to write the next post in my series of arguments against Reactionary politics – about why fighting racism and sexism is necessary and important.

The Reactionary argument, as I understand it, is twofold.

First, that social justice advocates irresponsibly take some undesirable outcome in minority groups, like poverty, and then assume it is the result of racism or sexism without considering other possible explanations.

Second, that a disproportionate amount of time and energy is spent worrying about this, in a way that can only be explained through wasteful signaling cascades.

My counterargument is that although the first argument is true a depressingly large amount of the time, some people do more rigorous work and get the same result – that poor outcomes for minority groups are caused in large part by racism and sexism. And second, that these poor outcomes for minority groups are a major problem even by objective quantifiable standards.

Controlled Experiments On Prejudice

The most fun experiments on prejudice are Implicit Association Tests, which test people's reaction times in linking together different concepts. If these concepts are socially important (for example, the concepts "white person", "black person", "good", and "evil") it can test how closely two different concepts are linked. The best way to get a feel for this is to take one yourself.

88% of white Americans (and 48% of black Americans!) show an implicit racial preference for whites on this test. How does that translate into the real world?

Some of the most interesting controlled experiments are detailed in an early ’90s review article in the Journal of Black Political Economy. A consortium of interested parties such as the Fair Employment Committee teamed up with recent university graduates. They laboriously paired white and black graduates by similar attractiveness, well-spokenness, age, gender, and qualifications (in some cases, the qualifications were faked to be as similar as possible), then sent them off job-hunting to the same companies.

In these sorts of experiments, 48% of white testers and 40% of black testers received interviews, a small and in fact nonsignificant difference. *However*, 47% of interviewed whites were offered jobs, compared to only 11% of interviewed blacks – a *gigantic* difference. Multiplying these two numbers together, we find that 23% of whites and 4% of blacks involved in the experiment got jobs – a difference of almost 6x. The whites also got a few other minor advantages – very slightly higher wages and slightly more likelihood of being informed of other open positions at the company.

Another good review article is in the Annals of the American Academy of Political and Social Sciences. It lists a few similar studies. In one such study, researchers, instead of training real applicants, send off fake resumes with extremely white-sounding or extremely black-sounding names; they find employers respond to the white-sounding names about 50% more often. But it also has some studies of in-person interview similar to the ones above. These studies, which are from the mid-2000s rather than the early 1990s, feature white:black success ratios of anywhere from 1.5x to 5x.

Along with labor discrimination, it’s harder for minorities to buy things. For example, when trying to buy a car, black men were asked to pay on average $1100 more than attribute-paired white men. Interestingly enough, black car salesmen, and black owned car dealerships, displayed this pattern to exactly the same degree as white-owned institutions.

The situation is roughly similar in housing. In an experiment where researchers responded to Craigslist notices advertising apartments in Toronto, using names of various ethnicities, they found that Caucasian experimenters confused relative risk with odds ratios 100% of the t…ahem, sorry, they found black people experienced housing discrimination 5% of the time and Muslims 12% of the time, usually in the form of not receiving a response even when the white person was simultaneously invited to come on over. A similar study in Houston found an astronomical 80% discrimination rate for blacks, so either Houston is much worse than Toronto, someone’s not doing their studies properly, or I’m misinterpreting something.

Other experiments along the same lines include a cute little bus experiment in Sydney where someone got on bus, their travel card didn’t work, and they asked the driver to let them ride anyway. For whites (and Asians) it worked about 72% of the time; for Indians, about 50%, and for blacks, 36%. In a later survey, bus drivers (who were unaware the experiment was going on) claimed they would prefer to help black people over white people. Interestingly enough, although black bus drivers were a bit nicer to blacks than white bus drivers, they *still* let whites and Asians on more often.

We find much the same pattern with men and women. A famous study a few months ago found that faculty offered a female grad student a 12% lower salary than an identical male grad student (again interestingly, female faculty were *more* biased against female grad students than male faculty were).

A less perfect but more natural experiment is switching from an open application procedure where applicants' genders are obvious to a blind procedure in which genders are unclear. If the percent of women hired increases (and perhaps if no similar increase is seen in competitors that don't change procedure at the same time) this implies the institution was being unfairly biased before. ~~When such a test was performed by the Journal of Behavioral Ecology starting in 2001, the percent of articles by female authors went up from about 29% to about 37%, about a 30% increase~~ **.[EDIT: This has since been found to be false]** Symphony orchestras are another infamous example, and studies show that the switch from open to blind auditions explains between half and a third of the recent quintupling of the percent women in symphonies over the past thirty years.

What do these show and not show? They show that, even controlling for all other factors like different preferences, different negotiating strategies, different educational backgrounds, et cetera there is a large difference in the opportunities of minority and majority groups due solely to discrimination. This difference seems large enough to explain the proportion of the income gaps that people say it explains (usually around half of the gap for each minority group) and to give minorities large amounts of trouble throughout the rest of their lives.

One thing it does not show is that racism is just about straight white men being evil. Minorities seem just as willing to screw other minorities over and discriminate in favor of white men as the white men themselves are. A better model would be that ideas of certain races and genders being superior seem to percolate into people’s consciousnesses, regardless of what race those people themselves are, and shape their actions whether they mean for them to or not.

Economic Costs of Discrimination

A beautiful experiment by Gwartney and Haworth noticed that baseball formed a natural experiment about the costs of discrimination. During the post-Jackie-Robinson 1950s, some teams had integrated but others had not and remained white-only. G&H wondered whether this affected performance. They found that in fact the five teams quickest to integrate black players were five out of the six top performers in the league, and that every additional black player on a team resulted in an addition 3.75 wins. This was partially because black players outperformed whites on average, but also because there was more low-hanging fruit in the form of black talent which could be employed more cheaply.

What is true for baseball teams is probably also true for other companies, but harder to quantify. For such a popular field, I cannot for the life of me find any attempt to quantify the economic costs of racism. There seem to be some people in Australia working on it, but they have yet to publish any results. So let’s make some up (this, uh, ends the demanding-of-rigor part of this post).

One way to do this is to take people’s estimates of the purely-discriminatory pay gap for different groups – that is, how much less they earn than straight white men when all other factors anyone can think of (like education level, IQ, height, region of residence, whatever) are adjusted away. Then multiply this by the number of people in that group and their average wage, and we get part of the cost of racism per year.

One of the review articles above suggests the black pay gap is 15%; others suggest numbers around 10% for women. Asians and gays make a bit more than straight white men, and although Latinos make much less no one has bothered adjusting for confounders so I can’t include them.

Anyway, when I add all that up, I get $374 billion.

(one might argue that the companies these people work for *gain* this money as profit by paying employees less, so it all evens out. I don’t think that works. In at least some cases, the lower pay must be because they have lower-level jobs than their white male counterparts. But since we already agreed they have the same skills as their white male counterparts, this suggests their skills aren’t being used fully, which means the cost really is to the economy and not just to them. I have no idea whether this argument works in real life. Like I said, the highly-demanding-of-rigor probably should have stopped with the first half of this post)

Another claim is that companies lose $64 billion dollars to discrimination-related turnover yearly. The number is generated by taking the cost of replacing a lost employee with results of surveys about how many people leave due to discrimination or hostility at their former place of employment.

Suppose we arbitrarily and implausibly stop here because we're tired. We've found costs of US racism equal to at least $438 billion per year.

That's *about* the annual budget of the US military.

Note what it doesn't include. It doesn't include any non-monetary costs like people being unhappy. It doesn't include the costs to the prison system of overprosecuting minorities. It doesn't include the costs to the health system of minorities getting worse preventative health. It doesn't include the amount the government spends fighting racism, or the amount people have to pay out in racism lawsuits. It doesn't include people who are unemployed because of racism, because I took the data from the employment records. It doesn't include any of the income gap due to racism anywhere other than at job – for example, racism that affects how much education people of different races end up with, or racism the person's parent suffers that then screws up their families for several generations. It doesn't even include Latinos because I couldn't find any good numbers about them.

It seems *very* unlikely to me that the actual costs are less than $1 trillion/year in the US alone. But let's stick with the $438 billion figure.

(another way to look at this is that these arbitrarily-stopped at costs of racism/sexism are about $2-3K per minority group member in the US, counting women as a "minority group". This seems broadly reasonable, and is in fact still way less than the observed non-adjusted income gap)

What other things cost $438 billion dollars? According to the American Cancer Society, cancer costs $200 billion/year Heart disease costs $100 billion. Adding up all of the easy-to-calculate costs of 9/11 on this page, I get about $250 billion.

So in terms of purely economic, not-even-worrying-about-human-beings costs, the costs of racism and sexism that can be pretty plausibly attributed to discrimination alone are equivalent to about heart disease plus cancer plus half a 9/11 or so per year.

One good thing about the size of this number means that small successes in fighting racism and sexism are extremely valuable. For example, decreasing racism/sexism by 1% is a $4.4 billion gain per year to the economy, which is about equal to Facebook’s 2011 annual revenue.

Effectiveness

Now none of this is meant to claim that the marginal blog on Tumblr complaining about the patriarchy has positive expected value or is anything other than a massive waste of everyone’s time.

But there are aspects of the social justice movement interested in testing what works and doing it.

As of yet, I don’t think most of them are aware of the pitfalls in claiming successful interventions – all of these “We found our intervention decreases expressions of prejudice on a seven point scale two weeks later!” things sound suspiciously like “Our drug increases ‘good cholesterol’ after three days, and we didn’t bother to check whether it actually prevents heart attacks but seriously how could it not?”

But we can’t blame them for their failure to be more rigorous than the hard sciences, and besides one day they might wise up.

With Implicit Association Tests, Ultimatum/Dictator games, and the like, I think there is a decent toolkit for people who want to wise up and seriously analyze anti-racism and anti-sexism strategies, and I bet when they are tested further some of the ones in that document will turn out to work longer-term. Maybe they could decrease racism by 1% a year and save us $4 billion or so.

The fact that racism in the sense of simple prejudice is a real problem that accounts for much of the disadvantage of minorities; that it has a huge negative effect even when you try to measure it objectively; and that it can be fought – seem to take some of the wind out of the Reactionary argument against social justice.

Search Terms That Have Led People To This Blog

slatte star codex
Sooooo close.

scott alexander silver slate codex
Gettttting colder.

based on your findings which theory about alien thickness seems most valid or most accurate
I'm going with this one

2013 issues with virtue ethics
Sorry, I only have about thirty or forty**.**

i do enjoy a good pun now and again

if doctor says i will be ok in the medium term how long will i live

siblings attracted to each other
Oh no, I write *one* article on genetic sexual attraction and…

my sis is so pretty
…and…

sweet sister so pretty
…aaaaaaaaaaaaagh…

sister can you not so pretty
…gggggghh…

i'm all for the singularity, but what will normal people do with all the automation sucking up the jobs

No need to worry, my good man, you will have long since been dissected and turned into computronium.

expectations of my gender hindered my confidence and the choices i have made.
I’m…I’m so sorry.

during an investigation of a bridge collapse engineer b investigates another similar bridge and find it to be only marginally safe. talk about this
So there was this engineer, right? And during the investigation of a bridge collapse, she was investigating another, similar bridge. And…ARE YOU EVEN LISTENING**.**

attractive siblings more prone to incest?
Afaik Science has utterly failed to study this important question.

1930 beheadings valley canada
I really hope 1930 is the year here.

when you think you inject fear, you bring out our courage. when you think you disorient us, you give us direction. we you think you cause division, you cause unity
One day the strength of men will fail! But NOT THIS DAY!

the pun 6 p.m.
Sorry, I dinner have any good ideas.

she followed me around the internet its proved too

there is nothing you can do to a pornstar’s food in a restaurant to offend them

crab puns
I’m sorry, I haven’t a claw.

i always want to serve god truely but i cant why?
Oh no. Better sit down. You're *really* not going to like this.

about 5 thousand years ago
…WAR WAS BEGINNING.

now we've decided to smile

polyamorous have too much time on on their hands the making rules all the tim

i dont feel strongly female or male. what is my gender idenity, im a female
Hmmm. Just judging by what you've said above, I'm going to say you're a female.

strong email thoughts
http://www.youtube.com/watch?v=Zx1eDmuYQg8

how much of the world population is gender indentity
Gender indentity is important to mark off your gender pargraphs.

inefficient public speaking
"So, guys! I'm here to tell you about…you know, I just thought you should know about that thing where…guys…the topic of my speech is how I…you know…"

what is your point virtue ethic
I DON'T KNOW EITHER!

better than virtue ethics
I am glad giving Google this search term leads to my blog.

muslims causing noise on the plan 2013
I just wanted to scheme in peace, and then came all these Muslims, CAUSING NOISE ON THE PLAN.

gaudiam magnam. habemus mater

is it safe to use an e-cigarette as a sexual prop even for non smokers?

sylvia plath masturbating with her head in the oven
I am almost certain this is not anywhere on my blog.

go full fnargl
I’m glad I seem to have popularized this expression.

pornography bejctification objection
The beatification of pornographers is a terrible idea, and I join you in objecting to it.

charlie writes 1 to 3 pages per day for his biology research paper. over a 10-day period, what is a reasonable estimate of the number of pages he will write?

too much violence in the usa april
It is, indeed, the cruelest month

mao….lets just say they both have different style of singing because it will never end with this kind of argument. and from someone claiming to know something about vocals (you must in order for you to say that someone don’t)? shouldn’t you be able to distinguish that?

scott alexander, medicine hat

prohibition puns
Bad idea. When puns are outlawed, only outlaws will have puns.

all trans guys have tattoos

i’m a girl and having gender identity

how to attract sister
Aaaah! You AGAIN?!

too many graphs
I am so glad you found me.

Implicit Association Tests and Suicidality

While looking up data on the Implicit Association Test for my post two days ago, I came across Nock & Banaji 2007 (Prediction of Suicide Ideation and Attempts Among Adolescents Using a Brief Performance-Based Test), an interesting study which I am learning about only six years late (I'm catching up!)

They tried to use the Implicit Association Test to measure suicidal intent in psychiatric patients. This is *good*. Right now the technology for predicting suicidal intent in psychiatric patients is asking them very nicely "Excuse me, but do you think you're going to commit suicide soon? Because if you say yes, we're going to have to lock you up in a hospital against your will. But please, answer honestly!"

Okay, that's exaggerated a little for dramatic effect. For one thing, people don't say that last part. They just imply it. And there are various screening instruments that ask the question in a variety of ways, and with a variety of related questions ("Do you feel like life is not worth living?" "Have you formulated a specific plan?" "Have you bought the tools you need to carry out the plan?") and then those screening instruments have been validated *ad nauseum*. And many people considering suicide really want help and are happy to be able to admit it and get it off their chests.

But the basic gist is still that if you want to know whether someone is going to commit suicide you don't have many options besides asking, and some people have an incentive not to tell the truth.

The Nock & Banaji paper validates an alternate means of assessment. Implicit Association Tests, as mentioned before, are computer-based instruments where a subject has to press keys that categorize like and unlike concepts as quickly as possible. The idea is that they will be able to do this slightly faster on concepts that *already* seem implicitly associated to them than concepts which seem contradictory. In the most famous example, people were asked to organize the categories (white people + good adjectives) / (black people + bad adjectives) and then afterwards the categories (white people + bad adjectives) / (black people + good adjectives). In general white people were able to do the first task faster, presumably because they had stereotypes that already associated black people and negative attributes together so their implicit associations were aiding in the task rather than contradicting it. If this doesn't make sense to you, it'll become much clearer upon taking the test.

Anyhow, instead of working with races and adjectives, this new version of the test matches the categories (self, other) with the categories (pictures of self-harm, pictures of not-self-harm).

Self vs. other were words like "me" "mine" "I" versus "you" "them" "her". Self-harm vs. not were pictures of scarred, cut skin versus pictures of healthy uncut skin. The theory was that if someone associated themselves with self-harm, they would have an easier time (as measured in reaction time) creating the categories (me + self-harm) and (others + not-self-harm) than the alternatives.

I'll admit this protocol sounds to me like it would do a terrible job predicting suicidal intent. It seems like if anything it would pick out people who had cut themselves before. The authors note this, but say that they were reluctant to use actual images of suicide (person hanging from noose?) because it might “plant the idea” in people’s heads. It is conventional wisdom in psychiatry that this doesn’t actually happen (though I haven’t personally researched the evidence base for this) and I’m disappointed that the study went the self-harm angle.

Extremely mysteriously, though, the study claims that it adjusted for presence of self-harm and found no effect on its suicidality data, and so it had no qualms about taking a population consisting of both self-harmers and non-self-harmers and assuming this test on self-harm wouldn’t be confounded by that.

I will put aside my extreme skepticism and report what they found – which was that their test was able to distinguish healthy controls, people with past suicidal ideation, and past suicide attempters with accuracy of $p < 0.01$ in each distinction. Further, significant differences remained when they controlled for all previously known ways of detecting the suicidal – eg age, demographics, pre-existing psychiatric diagnoses.

They also claimed that their test was able to predict prospective suicide attempts. That is, two members of their study group attempted suicide in the six months after their study, and they noted that these two people had higher average scores on their suicidality test than the subjects who didn’t. But looking at their calculations, it looks like they simply compared the average score of the two attempters with the average score of the seventy-one non-attempters. This seems useless to me. The two attempters were almost certainly from either the “previous suicide attempts” group or the “previous suicidal ideation” group, so comparing them to the entire rest of the study sample can’t tell us anything about whether this new test is any better than just noting that psychiatric patients with a previous history of suicide are more likely to commit suicide than healthy controls. In fact, this whole section seemed damning-by-faint-praise; if this was the most they could say about its predictive validity, that’s a bit worrying (in the study’s defense, it did classify this as “preliminary evidence”)

Overall I am very excited that work is being done in this area, but a bit skeptical about this study in particular. The self-harm aspect really bothers me and their claim that controlling for it doesn’t change anything needs more proof. In particular, they controlled for presence of past self-harm but not amount of past self-harm, and it seems totally plausible that people who self-harmed themselves more in the past are more suicidal.

But the biggest uncertainty is how useful this will be. The Holy Grail would be some test you could give someone, see they’re suicidal, place them in a hospital for a few days until they’re no longer suicidal, give them the test again to prove they’re no longer suicidal, and let them out when you see their test scores have improved. But for all we know, this method could test only extremely long-term constructs, something along the lines of whether you’ve ever harmed yourself or considered harming yourself. That would be useless for seeing whether someone’s suicidal now, and useless for determining whether they’ve stopped being suicidal after some treatment. The study’s extremely half-hearted attempts to analyze prospective data don’t reassure me here.

Still, I’m glad people are finally realizing the what I’ve been saying for years, which is that the IAT is really powerful and needs to be used for something other than nebulous social justice projects. I bet if the CIA created a (self/other + patriotic American/Russian double agent) Implicit

Association Test it would totally work. In the absence of that, I will just hope for more research on this suicide thing.

Product Recommendation: FB Purity

I mentioned before the stress and aggravation I get daily from people sharing their horrible politics on Facebook. It’s tempted me to quit the site before, but it would be sad for me to lose touch with all of my friends outside my immediate area just because some people post viral articles on God or grass or guns or gays or gender *ad nauseum*.

I recently found a solution in the form of the FB Purity add-on for Firefox (there are similar versions for other popular browsers as well as Opera). It allows you to fine-tune your Facebook settings in ways that (crucially) include an option not to show links or image macros. So I can hear about people’s moves and new children and interesting observations on life without having the entire blogosphere blasted at me. I can also turn off invitations to raise people’s imaginary cows and suggestions to join someone’s virtual Mafia. It’s a huge relief and for the first time I am finally starting to enjoy social media and consider using it more often.

So of course Facebook is trying to ban it and sue its creator. It’s kind of reassuring how consistently horrible Facebook is all the time. It’s like they looked at Google and their whole “don’t be evil” thing and thought “nah, that’s not for us.”

In Defense of Psych Treatment for Attempted Suicide

A lot of the comments in my recent post on the implicit association test asked for a defense of why society should be hospitalizing suicidal people in the first place. If people have, after much thought, decided they prefer death to life, isn’t that their right?

I am *extraordinarily* sympathetic to this position, which has been most eloquently defended by Sister Y of The View From Hell. Sister Y lists many harmful effects of suicide prohibition and many reasons why rational people might want to end their lives. She suggests a policy of legalizing fatal doses of barbituates for people who want them, allowing people tired of existence to leave the world without grisly suicide attempts that might leave them permanently injured or cause collateral damage to bystanders. I can’t find her opinion on whether these should be provided on demand or whether you should have to undergo a psychiatric assessment first.

If she in fact believes the latter, then I think that position is defensible, and for professional reasons I won’t publicly say anything further than that. But this post is to explain why it should require one *hell* of a psychiatric assessment and why the overwhelming majority of real-world suicide attempters would and should fail such an assessment.

Again, my point of disagreement is not on the ethics involved of letting some hypothetical perfect philosopher commit suicide – nor even on the fact that perhaps some cases genuinely are these perfect philosophers including Sister Y herself. I am trying to emphasize the practical point that in the real world, attempted suicides are rarely perfect philosophers and almost always people who have made sudden, impulsive, and very bad decisions.

The greatest burden of suicide is of course to the friends and family of the person involved. But you don’t have to be a Randian to think it’s morally abominable to require someone in pain to continue living solely to please other people, so this post will focus solely on the welfare of the person involved.

What Does Youth Suicide Tell Us About Adult Suicide?

Start with the clearest case. About 4% of teenagers attempt suicide at some point (there are some much higher values from the CDC, which is usually pretty trustworthy, but some of the comments point out reasons why their estimates here are pretty hard to believe.)

I’ve gotten to observe some teenagers admitted to hospitals for attempted suicide. Some have incipient mental disorders that no one has noticed or considered treating. Some have unbearable home lives. Others have the standard litany of teenage problems – broke up with their boyfriend/girlfriend, bullied by the popular kids at school, got into a fight with their parents. Many have a combination of all three.

Some were the classic “cries for help” that were never meant to actually end in death, but others were entirely serious. A tragic few *intended* to take enough of an overdose to make Mom scared but not enough to actually kill them, but muddled their pharmacology in the most permanent possible way.

And I think most people agree that teenage suicide is terrible and requires treatment. Heck, most people won’t even let teenagers make the decisions of whether or not to purchase alcohol, let along the decision to end their own lives. But I think aside from the inherent tragedy of teenage suicide it illuminates something about adult suicide as well.

These people have only the tiniest glimmer of knowledge about the likely happiness of their future lives. Most of their problems – the bullying by popular kids, the failed first relationship, even the awful families – are eminently wait-out-able. “It gets better” is not just for gay people (who, by the way, have a suicide rate up to 15x that of the straight population).

And yet teens attempt suicide at staggering rates.

There are certain depths of despair dark enough that the knowledge that the despair is completely temporary cannot penetrate them. It is this state, defined by the clouding of rationality by suffering, that I think most teenage suicides occur in.

And it would be very strange if this suddenly changed as soon as the victim hit eighteen.

Connection Between Suicides And Mental Disorder

It is generally reported that about 90% of suicides have some mental disorder. No, this isn’t an artifact of psychiatrists assuming anyone who commits suicide *must* have a mental disorder – various half-decent methodologies have all converged around the same number, including a multitude of controlled studies (where psychiatrists evaluate a subject’s mental status based on notes before knowing whether the person committed suicide) and prospective studies (where people only count as mentally disordered if they were diagnosed before the suicide occurred).

Sister Y has tried to poke holes in these statistics. First, she noted that the controlled studies showed 37% psych diseases even in the control population. But this number is probably correct – NIMH estimates that about 26% of people have mental disorders in a given year, and no doubt that number is significantly higher among people who make good controls (ie are matched on demographic factors) for suicides. Second, she pointed out that the number included what she considered relatively “minor” disorders like alcohol dependence.

So first of all, alcohol dependence probably septuples your chance of committing suicide and something like 25% of suicides include alcohol. So I don’t think it’s unfair to include that in the list of how suicide is influenced by mental disorder.

But second of all, let me give totally anecdotal and probably unrepresentative examples of some other ways mental disorder can affect suicide.

As stereotypical as it sounds, the voices in people’s heads do tell them to kill themselves a lot. Voices in people’s heads are *huge jerks* and occasionally people will do what they say just to make them shut up. The tragedy here is that antipsychotic drugs are pretty good at dealing with this if people can just get access to them. Among schizophrenia patients (the group most commonly identified with these sorts of symptoms), almost half attempt suicide and 10% complete it. Since schizophrenics make up 1% of the general population, that’s a non-negligible fraction of total suicides.

You know what’s an even less fun form of psychosis? Psychotic depression. This is where people get so depressed they start hallucinating about how horrible they are. I will never forget the patient who stopped eating because she believed her digestive system was rotting away and infested with maggots. And a lot of the time these people's self-hatred reaches completely bizarre proportions in which they will confess to causing the Holocaust or the 9-11 attacks just because *it seems like the sort of thing someone as horrible as them might do*. If you believe you caused the Holocaust, this seems like a pretty good reason to kill yourself in the name of justice, and sadly this is what many of these people do. And again, this is tragic because psychiatry is actually not so bad at dealing with this kind of over-the-top depression (the rotting-intestines woman became much better after a short course of electroconvulsive therapy, but some people will get better just on medications).

Borderline Personality Disorder is another common cause of suicides. It intensifies emotions so that anyone so much as making a mildly critical remark makes you think everyone will hate you forever and you deserve to die. And then six hours later someone smiles at you and you feel like the world is perfect and beautiful. But if you commit suicide at one of the low points, then that's *it*. And Borderline Personality Disorder, again, is *sorta* amenable to therapy, and even without therapy half the time it just *goes away* after a few years to a decade.

Alcohol and drug abuse is another big one. Some of it is that abusers have worse lives – poor health, financial issues, more likely to have trouble at work. But a big part of it is just *lowered inhibition*. If a sober person is walking on a bridge after some life crisis, they might have fleeting thoughts of jumping but suppress them after thinking of the future. If a drunk person is walking on a bridge after some life crisis, the frontal lobes that would normally suppress those urges are partly out of commission.

And then there's depression. I'm trying not to make a big deal about it because everyone associates suicide and depression when in fact the correlation is no higher than many other mental illnesses (although the greater number of depressed people does make absolute numbers higher). I guess all I'll say here beyond what everyone already knows is that Major Depressive Disorder (classic depression) is an intermittent disease. The average depressive episode lasts less than six months, and the average person with MDD has only four depressive episodes in their lifetime (these numbers are even better if you're on medication, which many depressed people fail to be). There's a thing called dysthymia, which is like having depression all the time, but it is thankfully less common and less severe and not where most suicides are coming from.

I am certain that six months *feels* like an eternity if you are depressed. And no doubt knowing that you're going to have to deal with the same thing a few more times in your life (ALTHOUGH SERIOUSLY, MEDICATION DOES HELP WITH THIS) must also be, well, depressing. But the average depressive suicide is not a Perfect Philosopher who has calculated, while healthy, that the possibility of another six month depressive episode is too much to bear.

The average depressive suicide is someone in the middle of one of their episodes who, like the teenagers above, is in the place so dark that they've forgotten the existence of hope. They're somewhere so dark that "this will probably go away in a couple of months" has no meaning. Somewhere so dark that one of the main side effects of *effective* antidepressant drugs is suicide, because a few weeks after starting the patient finally has enough energy to go kill themselves, but doesn't consider waiting a month or so for the drug to take full effect.

I want to end this section with a study – small, but encouraging – that cognitive-behavioral therapy (aka That

One Type Of Psychotherapy That Sometimes Works) reduces suicide 50% in at-risk populations. Think about that. What percent of suicides do you think haven't had cognitive-behavioral therapy? 80%? 90%? Whatever that percent is, half of them *would have been fine* if they had just had access to a good psychologist.

Empirically, Suicides Regret It

People who commit suicide can't change their minds. But attempted suicides can and do, and we can analyze these changes both in their actions and in their words.

In terms of revealed preferences, most people who are prevented from completing their suicide do not go on to kill themselves. Sister Y critiques a study saying only 4% later go on to kill themselves, and offers as counterpoint a study she prefers claiming 13% do (she finds a way to round up to 19%). I have also heard 10%, although I can't remember where. Do you know what the numbers 4%, 10%, 13%, and 19% all have in common? Yes. They are all significantly less than 50%.

It is somewhat harder to find good studies on what percent *attempt* suicide again. By eyeballing some other statistics and trying to fit them together, I believe it is greater than 25% but less than 50%. One textbook whose studies I have not been able to verify says that 30% of untreated and 15% of treated suicide attempters try again. 15% and 30% are also among the many numbers that are less than 50%.

And keep in mind what these data *don't* show. They don't show that the 25-50% who try again have lives so constantly miserable that they continue wanting to die. Remember that intermittent depression from before? Imagine a world in which depressive episodes last one day each, and people only have two of them in their lives. Other than those two days, they live happy lives and are grateful to be alive. Doesn't matter. This pattern would *still* be consistent with 25-50% of attempted suicides making repeated attempts, if that second day of depression was bad enough

(out of fairness I should mention this data *also* doesn't show that 50-75% of people get over their suicidality; it's consistent with them just being tired of suicide attempts not working and settling for continued existence. I guess what I'm saying is that the data don't prove very much)

So moving from boring data to the much-more-fun domain of anecdote, a surprising number of suicide attempters change their mind *during* the suicide attempt. One particularly famous case is that of Kevin Baldwin, who survived jumping off of America's favorite suicide spot. He says that while still in the air "I instantly realized that everything in my life that I'd thought was unfixable was totally fixable—except for having just jumped."

Most realizations are slightly less dramatic, but my work in a psych ER taught me that many 9-1-1 calls about suicides are from the victims themselves. I remember one patient, a typical case, who overdosed on pills. As she lay on the ground starting to feel sick, she thought about her problems a little more deeply, thought about how her family would feel, and decided she preferred to live. She called 9-1-1, they sent an ambulance over, and the hospital managed to keep her alive until the drugs passed out of her body. This is quite common. It also contradicts one of Sister Y's strongest arguments – that the reason many people avoid suicide is out of fear of making the attempt. A non-negligible number of people who have already made the attempt and just have to sit back and day find themselves changing their minds and actively working to save their own lives.

But most of the stories I can generate from my personal experience are nothing more dramatic. It’s people who were found by their parents or partners or friends, dragged kicking and screaming to the hospital, treated for a couple of days, and by Day 3 they’re saying oh my god I made a horrible mistake I can’t believe what almost happened.

And I know what the response will be – that of *course* they’d say that to their psychiatrists, they’re trying to get judged Officially Sane so they can get discharged and maybe try again. I accept that as a possibility, but since this whole section is about totally useless anecdotal data, let me just say I don’t feel like that was what was happening. I met people who were going out of their way to look for and thank their psychiatrist when he was busy in his office after the discharge papers had already been signed and they were on their way out. One time I met a patient at the bus stop a few days after she had been discharged, and she asked me to thank my boss and the rest of the team for what must have been the umpteenth time.

Finally, I have some personal friends who have attempted suicide. In every case I am *incredibly* glad they remain alive, and more importantly, usually they are as well. And I know there’s social pressure here – that psychiatrists aren’t the only ones you have a vested interest in appearing cheerful to – but some are very close to me indeed and I do not believe they would lie about something this important.

Psychiatric Care Probably Helps

One of the most common objections to sending people who attempt suicide to psychiatric hospitals is that it is a terrible punishment, that we are essentially locking up and drugging and torturing people whose lives are already apparently pretty bad.

But mental hospitals for people who attempt suicide (actually almost always just the psychiatry floor of a regular hospital) are *not like that one in One Flew Over The Cuckoo's Nest*. I can't repeat that enough. I know that as a psychiatrist-in-training I have no credibility on this issue, so take it from a former psychiatric patient. Your problems are much more likely to be along the lines of a terrible selection of books in the ward library than torture by sadistic nurses (I do not deny the latter occurs, just as some schools have torture by sadistic teachers, but it is extremely rare and nowhere near for example what goes on in nursing homes).

According to the CDC, the average length of stay in a mental health ward is one week (this brief by an organization I've never heard of says 8 days). That includes catatonic people and people who have long animated conversations with the Devil, so the average suicidal person isn't going to be the one bringing up that average.

In practice I have a pretty good guess for the *exact* length of stay the average suicidal person without associated mental disorders will experience, and that is 72 hours. That's the maximum amount of time a hospital can legally commit someone against their will. After that they have to get a court order allowing them to hold the patient longer, and this requires swearing that the patient is mentally incompetent to make their own decisions, and most doctors will not do this without reason.

But if you don't trust doctors' benevolence, at least trust their self-interest: it takes a lot of paperwork, it requires them to go all the way to a courthouse, and the hospital management is going to be breathing down their back the whole time about how they could *really* use an extra bed on Ward 4 and of course we would *never* pressure you to discharge any patients before they're better, but seriously, have a bed open on Ward 4 by tomorrow. Trust me, doctors are not plotting to keep people in the hospital longer than necessary. If you like conspiracy theories, the opposite conspiracy is a much bigger cause for concern.

That's usually just enough time to evaluate the patient for mental disease, start them on some medication, and refer them to an outpatient psychologist and/or psychiatrist. One hospital I worked at kept (mostly willing) people in a little longer to see if the drugs actually took effect, but that was a luxury they could only afford because they were a rich academic institution.

But the thing is, *this really helps*. If 90% of people committing suicide have some associated mental disease, and mental diseases can dectuple your risk of committing suicide, then connecting these people – many of whom have never interacted with the mental health system before – with someone who can help them (or even with a Prozac prescription) can be a really, really big deal.

I mentioned before that one specific form of therapy can decrease future suicide rates 50%. That was in a study where *both* groups were getting the recommended psychiatric drugs. Another study I cited above said that "psychiatric treatment" (whatever that means; I bet it didn't include the CBT from the last study and so they're cumulative) can also cut future suicide rates in half. There are more specific studies on the anti-suicide effect of each individual drug – lithium is an example of a particularly good one.

(fun fact which there is a small chance I will devote my life to studying: even areas with slightly higher trace amounts of lithium in the *water supply* have lower suicide risk.)

And even if you're one of the depressingly high number of people who throw away their prescription and never show up to their psychiatrist, you know what? You've been stuck in a big building with lots of people watching you for the three days or so immediately after whatever horrible event made you become suicidal in the first place. Drugs in your system? Now you're clean. Angry at a family member? Maybe you're less angry now. Upset over a breakup? Maybe you've had a chance to think about it a little more.

I am very reluctant to get into in what situations I believe suicide is acceptable. I am scared that one day my future employers will read this post. Or worse, a future patient will read it and start arguing "You said suicide was acceptable if A or B, so I did those things". So all I will say is that I wish Sister Y and those like her maximum utility however they define their utility function. But anyone considering suicide who has thought about it less than she has or lacks her philosophical acumen should consider getting professional help (or even non-professional help) or at least meditate long and hard on that cliche about "a permanent solution to a temporary problem".

EDIT: Since people are missing something I said like *a thousand times* in the post itself, I'll put it down here in bold. **I am not claiming that suicide is never rational and that all suicides are stupid and impulsive, or that no one can ever legitimately want to die. I am saying those people make up a very small portion of suicides, and that the typical case is people who do it impulsively or in a state where they lack full decision-making capacity. And that the psychiatric**

system can be of huge help to this latter group, and that helping the former group is a different question which I do not want to talk about publicly for professional reasons.

Utilitarianism for Engineers, Part II

You know how it's impossible even in principle to compare people's utilities and so utilitarianism is a pipe dream that can never possibly work?

Well, I just learned Tufts has a searchable public database of utilities for various health outcomes. It doesn't seem very *good* – some of the entries in there make so little sense that I worry that it's haphazardly combining QALYs (in which higher numbers mean higher utility) and DALYs (in which lower numbers mean higher utility). Either that, or catching the flu decreases your utility more than being totally paralyzed does. But the database exists and is full of numbers, and that's inspiring.

But this still doesn't fix my constant gripe that no one has tried to extend this to non-health states. Searches for QALYs for poverty or totalitarian government, for example, mostly just turn up me complaining in various places about how there aren't any.

Just to see if there is some fundamental impossibility here, and to get enough data to whet my own curiosity, I've set up a very amateurish person-who-is-no-good-at-math version of a utility measurement test. If you have a while (15-30 minutes) and are willing to follow confusing instructions, can you take it for me?

(survey closed, thank you for participation)

I agree it's extremely confusing and most questions are underspecified in various ways, but if you could just take it anyway, try to guess what I mean, and leave anything you don't get blank, that will be good enough for me.

Yellow Journalism In The War On Science

The Internet has exploded with hatred for a new draft bill making its way through the House – called, in classic Congressional weasel language, “The High Quality Research Act”

Pharyngula, showing the subtlety and restraint for which it is famous, says it’s about “destroying science in the US”. The Verge says it will “replace peer review”. The Daily Kos says since science proves Republicans are wrong, Republicans are trying to legislate away science. The most worrying description comes from sodahead.com, which says it will “destroy any concept of facts in America”.

Yes, you heard that right. If this bill passes, scientists will be federally mandated to say things like “Well, it’s my opinion that the sun is a giant ball of plasma 93 million miles from the Earth, but if other people say about as high up as the average passenger plane and approximately the size of a grapefruit, that’s a valid opinion too.”

Wait, no, the bill has nothing to do with that and WHY DO PEOPLE KEEP GETTING THEIR NEWS FROM DESIGNATED OUTRAGE FACTORIES AAAARGH!

Having taken a cursory three-second glance at the text of the bill, let me note some things that I already understand better than anyone else who has written about it thus far.

The bill does not ban peer-review, in the sense of the scientific peer review that happens in journals. It changes the review process for grant funding from one where domain experts review which grants are likely to be useful, to one where the head of the NSF chooses projects to fund directly. Because the domain experts are sort of “peers” of the scientists involved, and they’re “reviewing” the grants, you can sort of call this process “peer review”, and if you do that, this bill “ends peer review”. But if you just write a headline saying “REPUBLICANS TRY TO ABOLISH PEER REVIEW IN NEW ANTI-SCIENCE BILL”, no.

The bill is not trying to ban the replication process either (sadly, if there are any politicians who hate studies being replicated, they’re probably pretty content with the current state of science already.) The bill actually bans “duplication” of studies, which surprisingly is not the same concept despite being an extremely similar sounding word! *Duplication* seems to be the process of getting several times as much money for the same research by applying for multiple grants without informing one grant agency that you already received a grant from the other. The scientific community has been worried about this process for a while, and even *Nature* published an article raising the alarm about it. Even such a bastion of anti-science Republicanism as the *New York Times* has spoken out against the problem. The National Institute of Health has already made rules against duplication, and this bill proposes nothing more than that the National Science Foundation should do the same.

The legitimate worry about this bill is that it restricts funding to research that “is in the interests of the United States [and serves to] advance the national health, prosperity, or welfare, and to secure the national defense by promoting the progress of science”.

This has been caricatured as an an attempt to:

> …write patriotism into the funding of science. Our work must promote AMERICA, and also must be of utmost

> importance, where importance is to be defined by Texas a-holes in shitkickers and big ol' cowboy hats, who probably don't care much for them fruit flies.

A more generous reading says that the use of "United States" is mostly about Congresspeople's tendency to name-drop our Glorious Country Of Freedom And Justice on any possible occasion, and that the main thrust of that section is that research funded by the government should be *practical*. It ought to improve the country in some way.

Even this more charitable framing has its critics. One blog pointed out how often we get good results from basic research. It suggested that it might seem "profitable" to cut out the ivory-tower-seeming project of trying to get materials to emit weirdly coherent beams of light, without realizing that lasers would someday be used to play DVDs, perform delicate surgeries, and annhilate Alderaan.

And people are quick to point out some famous missteps by politicians: Sarah Palin condemned research on fruit flies, even though that leads to a better understanding of genetics and of human disease, and Bobby Jindal made fun of volcano monitoring only a few days before a volcano erupted and early warning systems managed to successful save people and property.

But keep in mind that these are examples filtered by what *actual* scientists would call "selection bias" – that is, we know about them precisely because they are stupid. If we were to look at what opponents of current NSF funding priorities are *really* objecting to, the story would be a little different. Senator Tom Coburn, in his anti-NSF report, lists:

> – a YouTube rap video attacking fossil fuels and promoting biogas (cost: $50,000)

> – a study on whether playing FarmVille on Facebook builds relationships (cost: $315,000)
> – a study on why the same teams always seem to be dominating March Madness (cost: $80,000)
> – a study on whether people who often post pictures on the Internet from the same location at the same time are usually friends (yes, cost: $2 million)
> – a study on the relationship of virtual world users to their avatars (cost: $90,000)
> – a study on how often people lie in text messages (cost: $476,000)
> – a study on whether Twitter users tweet in regional slang ($1.4 million)
> – a virtual recreation of the 1964/1965 New York World's Fair ($1.2 million)
> – a project to send artists to Antarctica so they could make art about it ($300,000)

Coburn concludes:

> While taxpayers support investing in science, most would likely question the merits of these projects. Who would disagree the dollars spent on these efforts could not have been better targeted identifying more efficient, renewable fuels, developing the next generation of computers, creating new antibiotics for resistant bacteria, or simply reducing the nation's debt.

So I'm probably the worst person to complain about some of these, seeing as how I had to actively restrain myself from going off and reading through some of those wasteful studies until I had finished this blog post. And I also do worry that this could be used as a way of stigmatizing social science in favor of the sexier "hard sciences", or of attacking projects that get important scientific results in unorthodox ways (for example, study human psychology or decision-making via an online video game).

But, well, this reminds me of my thoughts on Ozy’s post about how The Government Does Not Spend Enough Money Researching Snail Sex (huh, somehow the National Science Foundation spending $880,000 on snail sex failed to make it into Senator Coburn’s report, which makes me wonder what else he left out).

Ozy’s post was well-written, it lucidly supports the cause of Science, it makes a good case for why snail sex is interesting, and it can be completely demolished by two words, those being of course “OPPORTUNITY COSTS”.

Like, I think it’s nice that we know more about snail sex. There’s just lots of other things we could do with $880,000 that would be even nicer. Eisenhower said that “every gun that is made, every warship launched, every rocket fired signifies, in the final sense, a theft from those who hunger and are not fed, those who are cold and are not clothed.” Every study that is run on snail sex does much the same. And since broad budgetary categories are pretty fixed, bad science funding priorities are robbing neither from the clothing budget nor from the warship budget, but from other scientists who may have important research they can’t pursue.

I agree this is by no means the *worst* waste of money our government is engaged in; that the entire National Science Foundation is only a drop in the bucket, that probably all of these idiotic projects, even if we accept they are idiotic, are themselves only a drop in the National Science Foundation’s drop.

But *this is what the debate is about*. If you want to engage in the debate, you’re going to have to say that, instead of trying to distract from it, instead of claiming this is about people who want to ban peer review and stop inventing lasers. This debate is about Congress worrying the current NSF grant criteria produce lots of duplicates and the occasional million dollars spent on studying Facebook-friending, and just want to pass a law saying not to do that.

I do not know how much room for improvement in the National Science Foundation there really is, or whether a NSF run the way Lamar Smith wants it to run would be better or worse than the way it currently operates. This is a debate that I would look forward to seeing people who know a lot more about science, politics, and economics than I do having.

But anything anyone has said about this bill thus far isn’t it.

Utility Weight Results

[Warning: The following post contains statistics done at 4 AM and not double-checked.]

By far the most interesting result from my utility and QALY survey earlier was the striking last name imbalance in readers of this blog. 93 people took the survey intended for people A-M, versus 34 for the N-Zs.

American last names begin with A-M 62% of the time (most of the imbalance is in M; I blame the Scots and Irish). StatTrek informs me that a 62% balance giving me results at least as lopsided as 93-34 should happen by chance only 0.5% of the time. Perhaps people just took the first survey no matter what their name was? Maybe they were turned off by the somewhat more complicated instructions on the second survey? It seems to be a mystery.

(as, indeed, is why about 85% of respondents came from Juneau, Alaska. Must be one of those coincidences :P)

I would like to think the survey showed that most people are pretty anti-death. The survey with death as the floor didn’t give great evidence for that, since I specifically told them to act as if death was a good floor in order to judge other things properly. But in the version with prison as the floor people consistently rated death significantly worse than states like prison, being in a starving African country, or being in North Korea. The survey was not really aimed at discussing death or suicide, so I would like to think it was less about signaling than some direct reports might be.

The next most interesting result was in the Relationship question. Single people consistently said a relationship would be about 1.54x better than their current single life, but people in a relationship estimated singlehood would only be about .85x as good as their current life; these numbers are obviously not reciprocals of each other. I blame myself; in the wording of the question, I asked single people to imagine a relationship with their “ideal” partner. For anyone with a remotely good imagination, this is a *very* desirable state. People in real relationships, even very happy relationships, probably can’t live up to that “ideal partner” standard, so it may be rational for these numbers to be different and not a sign of bias at all. On a similar note, men expected that women have significantly lower quality of life, but women did not expect men to have significantly higher quality of life; this failure of matched reciprocals seems much less excusable.

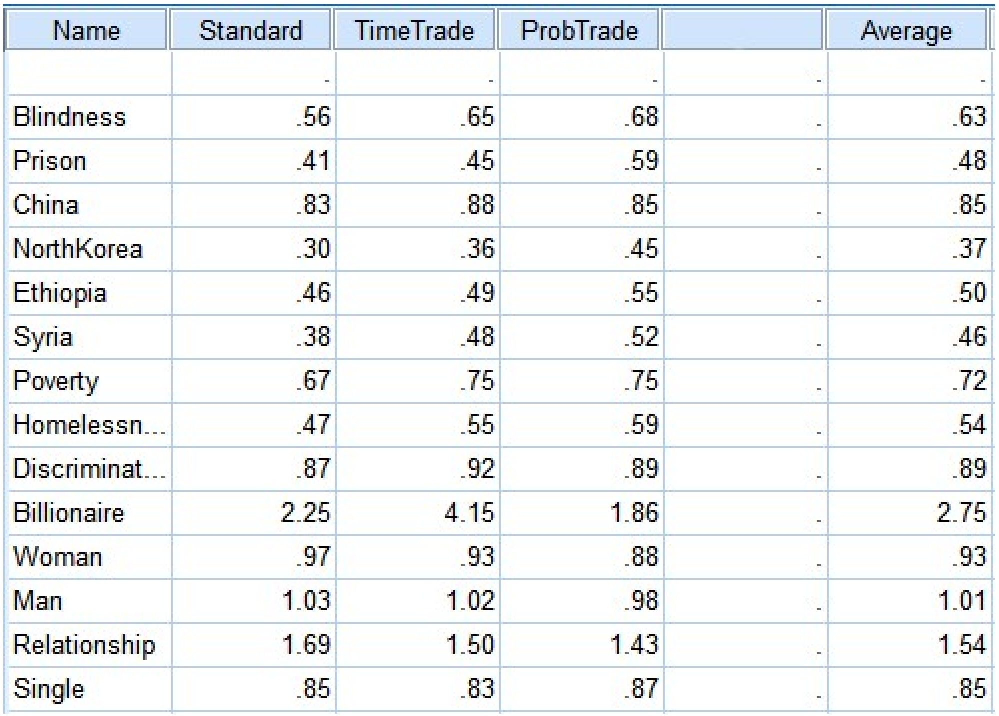
But what about the utility calculations? This was a mix of good and bad news.

The three methods of determining utility correlated extremely haphazardly with one another. Sometimes they would correlate as high as .7 or .8, other times not at all. There seemed to be no pattern to on what questions they would correlate, nor to which of the three would correlate best with which other of the three methods. The best I can say is that correlations between the three methods on the same question always seemed higher than correlations between different questions, which is good, I guess.

But once they were averaged out, they came up with remarkably consistent results. The lists of averages on each method correlated with the lists of averages of both other methods at > 0.9 level.

There was a very wide range of answers on each question; even something as uncontroversially bad as blindness would get utility weights of anywhere from 0.1 to 0.9. The standard deviations, once each question was standardized to be between 0 and 1, were pretty consistently around 0.2.

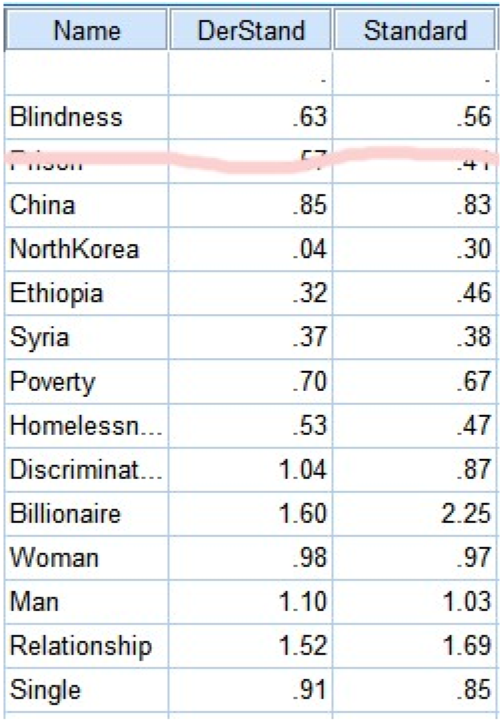
But most people had very similar preference orderings, and most people agreed on which things were only mildly bad (only tiny deviations from 1) versus worse. I feel like although the data were very noisy, there was also a very real signal in there.



Name	Standard	TimeTrade	ProbTrade		Average
	.	.	.	.	.
Blindness	.56	.65	.68	.	.63
Prison	.41	.45	.59	.	.48
China	.83	.88	.85	.	.85
NorthKorea	.30	.36	.45	.	.37
Ethiopia	.46	.49	.55	.	.50
Syria	.38	.48	.52	.	.46
Poverty	.67	.75	.75	.	.72
Homelessn...	.47	.55	.59	.	.54
Discriminat...	.87	.92	.89	.	.89
Billionaire	2.25	4.15	1.86	.	2.75
Woman	.97	.93	.88	.	.93
Man	1.03	1.02	.98	.	1.01
Relationship	1.69	1.50	1.43	.	1.54
Single	.85	.83	.87	.	.85

The numbers here are all means. The medians were mostly the same with the exception of the "billionaire" question, where a few people just seemed *incredibly* excited about the possibility of being a billionaire and everyone else was kind of meh (one friend noted wanting to be able to give the money to charity, which is something the question should have warned against). The median there was closer to 1.3 than the dizzying heights the mean attained.

Another encouraging fact is that even with different floors, the people on Survey 1 and the people on Survey 2 came up with almost the same results once they were adjusted onto compatible scales:



Name	DerStand	Standard
	.	.
Blindness	.63	.56
[redacted]	[redacted]	[redacted]
China	.85	.83
NorthKorea	.04	.30
Ethiopia	.32	.46
Syria	.37	.38
Poverty	.70	.67
Homelessn...	.53	.47
Discriminat...	1.04	.87
Billionaire	1.60	2.25
Woman	.98	.97
Man	1.10	1.03
Relationship	1.52	1.69
Single	.91	.85

In the table above, DerStandard is derived from the second survey's answers by trying to transform them to be on the same scale as the first survey. Standard is the unadjusted answers to the first survey. As you can see, even though they were working from different starting points they came up with some very similar answers (only the first test is shown on the table, but the other tests seemed broadly similar).

Last point. I put "blindness" on there because many people in health research have studied the negative utility of blindness and I wanted to see if this test, amateurish and haphazard as it was, would get the same results they did (which might in turn legitimize some of its other numbers). The CHUM-T gets a utility weight for blindness of .63. Tufts' database of utility weights lists numbers between .4 and .69 (also other numbers for various technical terms that mean blindness, but I think the exact wording is important here).

So I guess if I can conclude anything from this, it’s that utility measurement is very hard and produces confusing results, but that this test doesn’t seem to be any more wrong than others.

You can download the raw data in .csv format here. Tell me if you find anything cool.

If It’s Worth Doing, It’s Worth Doing With Made-Up Statistics

I do not believe that the utility weights I worked on last week – the ones that say living in North Korea is 37% as good as living in the First World – are objectively correct or correspond to any sort of natural category. So why do I find them so interesting?

A few weeks ago I got to go to a free CFAR tutorial (you can hear about these kinds of things by signing up for their newsletter). During this particular tutorial, Julia tried to explain Bayes’ Theorem to some, er, rationality virgins. I record a heavily-edited-to-avoid-recognizable-details memory of the conversation below:

Julia: So let’s try an example. Suppose there’s a five percent chance per month your computer breaks down. In that case…
Student: Whoa. Hold on here. That’s not the chance my computer will break down.
Julia: No? Well, what do you think the chance is?
Student: Who knows? It might happen, or it might not.
Julia: Right, but can you turn that into a number?
Student: No. I have no idea whether my computer will break. I’d be making the number up.
Julia: Well, in a sense, yes. But you’d be communicating some information. A 1% chance your computer will break down is very different from a 99% chance.
Student: I don’t know the future. Why do you want to me to pretend I do?
Julia: *(who is heroically nice and patient)* Okay, let’s back up. Suppose you buy a sandwich. Is the sandwich probably poisoned, or probably not poisoned?

Student: Exactly which sandwich are we talking about here?

In the context of a lesson on probability, this is a problem I think most people would be able to avoid. But the student's attitude, the one that rejects hokey quantification of things we don't actually know how to quantify, is a pretty common one. And it informs a lot of the objections to utilitarianism – the problem of quantifying exactly how bad North Korea shares some of the pitfalls of quantifying exactly how likely your computer is to break (for example, "we are kind of making this number up" is a pitfall).

The explanation that Julia and I tried to give the other student was that imperfect information still beats zero information. Even if the number "five percent" was made up (suppose that this is a new kind of computer being used in a new way that cannot be easily compared to longevity data for previous computers) it encodes our knowledge that computers are unlikely to break in any given month. Even if we are wrong by a very large amount (let's say we're off by a factor of four and the real number is 20%), if the insight we encoded into the number is sane we're still doing better than giving no information at all (maybe model this as a random number generator which chooses anything from 0 – 100?)

This is part of why I respect utilitarianism. Sure, the actual badness of North Korea may not be exactly 37%. But it's probably not twice as good as living in the First World. Or even 90% as good. But it's probably not two hundred times worse than death either. There is definitely nonzero information transfer going on here.

But the typical opponents of utilitarianism have a much stronger point than the guy at the CFAR class. They're not arguing that utilitarianism fails to outperform zero information, they're arguing that it fails to outperform our natural intuitive ways of looking at things, the one where you just think "North Korea? Sounds awful. The people there deserve our sympathy."

Remember the Bayes mammogram problem? The correct answer is 7.8%; most doctors (and others) intuitively feel like the answer should be about 80%. So doctors – who are specifically trained in having good intuitive judgment about diseases – are wrong by an order of magnitude. And it "only" being *one* order of magnitude is not to the doctors' credit: by changing the numbers in the problem we can make doctors' answers as wrong as we want.

So the doctors probably would be better off explicitly doing the Bayesian calculation. But suppose some doctor's internet is down (you have NO IDEA how much doctors secretly rely on the Internet) and she can't remember the prevalence of breast cancer. If the doctor thinks her guess will be off by less than an order of magnitude, then making up a number and plugging it into Bayes will be more accurate than just using a gut feeling about how likely the test is to work. Even making up numbers based on basic knowledge like "Most women do not have breast cancer at any given time" might be enough to make Bayes Theorem outperform intuitive decision-making in many cases.

And a *lot* of intuitive decisions are off by way more than the make-up-numbers ability is likely to be off by. Remember that scope insensitivity experiment where people were willing to spend about the same amount of money to save 2,000 birds as 200,000 birds? And the experiment where people are willing to work harder to save one impoverished child than fifty impoverished children? And the one where judges give criminals several times more severe punishments on average just before they eat lunch than just after they eat lunch?

And it’s not just neutral biases. We’ve all seen people who approve wars under Republican presidents but are *horrified* by the injustice and atrocity of wars under Democratic presidents, even if it’s just the same war that carried over to a different administration. If we forced them to stick a number on the amount of suffering caused by war before they knew what the question was going to be, that’s a bit harder.

Thus is it written: “It’s easy to lie with statistics, but it’s easier to lie without them.”

Some things work okay on System 1 reasoning. Other things work badly. Really really badly. Factor of a hundred badly, if you count the bird experiment.

It’s hard to make a mistake in calculating the utility of living in North Korea that’s off by a factor of *a hundred*. It’s hard to come up with values that make a war suddenly become okay/abominable when the President changes parties.

Even if your data is completely made up, the way the 5% chance of breaking your computer was made up, the fact that you can apply normal non-made-up arithmetic to these made-up numbers will mean that you will very often *still* be less wrong than if you had used your considered and thoughtful and phronetic opinion.

On the other hand, it’s pretty easy to accidentally Pascal’s Mug yourself into giving everything you own to a crazy cult, which System 1 is good at avoiding. So it’s nice to have data from both systems.

In cases where we really don’t know what we’re doing, like utilitarianism, one can still make System 1 decisions, but making them with the System 2 data in front of you can change your mind. Like “Yes, do whatever you want here, just be aware that X causes two thousand people to die and Y causes twenty people an amount of pain which, in experiments, was rated about as bad as a stubbed toe”.

And cases where we don’t really know what we’re doing have a wonderful habit of developing into cases where we *do* know what we’re doing. Like in medicine, people started out with “doctors’ clinical judgment obviously trumps everything, but just in case some doctors forgot to order clinical judgment, let’s make some toy algorithms”. And then people got better and better at crunching numbers and now there are cases where doctors should never use their clinical judgment under any circumstances. I can’t find the article right now, but there are even cases where doctors armed with clinical algorithms consistently do worse than clinical algorithms without doctors. So it looks like at some point the diagnostic algorithm people figured out what they were doing.

I generally support applying made-up models to pretty much any problem possible, just to notice where our intuitions are going wrong and to get a second opinion from a process that has no common sense but is also lacks systematic bias (or else has unpredictable, different systematic bias).

This is why I’m disappointed that no one has ever tried expanding the QALY concept to things outside health care before. It’s not that I think it will work. It’s that I think it will fail to work in a different way than our naive opinions fail to work, and we might learn something from it.

EDIT: Edited to include some examples from the comments. I also really like ciphergoth’s quote: “Sometimes pulling numbers out of your arse and using them to make a decision is better than pulling a decision out of your arse.”

Selection bias and atheist stereotypes

Have people considered before that the stereotype of the loud angry atheist may be entirely a result of selection bias?

Consider the average person’s experience with religious people. You see religious people going to church. You notice them wearing necklaces with crosses or Stars of David or dharma wheels on them. Sometimes they quote the Bible or the Quran, or they pray for things. Sometimes they run hospitals or schools or monasteries. The Pope occasionally shows up on TV telling people they should probably have fewer wars and famines and see if that works out for them.

And yes, occasionally religious people rant about how much they hate atheism, or the various other religions. But it’s hardly the only time you ever hear about them.

On the other hand, consider the average person’s experience with atheists. They may know many atheists. Their next door neighbor, or boss, or close friend might be an atheist. But they don’t hear about it. It rarely comes up in random conversation. Atheists don’t all go to one specific building, they mostly don’t wear necklaces, they don’t have holidays where they eat special atheist foods or dress up in special atheist costumes, and there’s no St. Lucretius’ Atheist Hospital.

Pretty much the only time you ever hear about atheists, unless you go looking for them, is when an atheist is criticizing religion or standing up for some kind of atheist cause. And so the reasoning goes: “Atheists must be bizarrely obsessed with religion, all the atheists I know about seem really into attacking it.”

But this is about the same reasoning pattern as “About half the medieval Europeans who I’ve heard of are kings; therefore I estimate medieval Europe had about one king for every two people.”

My guess is that most atheists, in the sense of people who don’t believe in God, rarely talk or think about it. But they don’t get much attention, and if they do, it’s not as atheists. “Famous atheists” brings up Richard Dawkins and Christopher Hitchens – not Lance Armstrong and Maurice Sendak. If there wasn’t a selection bias in terms of whose atheism gets noticed more, we would end up with the stereotype that atheists ride bikes and draw awesome pictures of cartoon monsters.

I wonder just how far this bias goes in creating stereotypes. It seems most likely to be a factor when there’s a group that’s hard to identify except when it engages in activities stereotypically associated with that group, and which might be geographically/socially sorted well enough that members of other groups rarely come in contact with them naturally.

Anti-Muslim stereotypes might come from such a source. If people live in an area without Muslims – or if Muslims tend to be segregated from non-Muslims – then someone might only hear about them when they’re fulfilling some kind of negative Muslim stereotype. I actually think the same is true of Christians – if someone lives in a very secular area, *most* of the Christianity they hear about could be some kind of very extreme televangelist.

And although this is a stretch, the same might apply to groups that *are* easily identifiable if people don’t think of the concrete examples they know *qua* group membership. Imagine you interact with your friend Juan – because Juan is great – and your friend Pedro – because he’s great too – and then you hear a story about Mexicans illegally crossing the border, and that activates your “Mexican” category much more than Juan and Pedro do, because they’re just ordinary guys.

And okay, that’s just wild speculation. I’m pretty convinced about the atheism one, though.

Ambijectivity

The statement "Mozart's music is better than Beethoven's" is usually considered a subjective opinion.

But this statement has the same form as "Mozart's music is better than the music of the three-year old girl who lives upstairs from me and bangs on her toy piano sometimes."

Is this latter statement also subjective? Calling it "subjective" or "a matter of opinion" feels wrong; someone who disagrees with me on this issue would be *weird* in a way someone who disagrees with me about chocolate vs. vanilla ice cream isn't. But the girl-upstairs question seems similar enough to the Beethoven question that admitting the existence of an objective answer here seems to force belief in an objective answer about the relative merit of Beethoven.

And of course part of the answer here is the extent of human variation. For whatever reasons – different genetics, different life experiences, whatever – people have different tastes in music. The human music-appreciating-organ varies enough that some people can prefer Mozart to Beethoven and other people can express the opposite preference. But it doesn't vary enough that any person's music-appreciating-organ could prefer the girl upstairs.

Or to give another example, for whatever reason some people's taste buds prefer vanilla and other people's taste buds prefer chocolate, but basic regularities in human design – like the evolutionary need for sugar and the neural connections between sugar receptors and pleasure pathways – suggest that pretty much everyone will prefer either of those to horseradish ice cream.

But let’s take another example: was Mozart’s music more original than Beethoven’s? This question sounds a lot like the first question of whether Mozart’s music was better. And it shares the same sort of weird half-subjective half-objectiveness (let’s call it ambijectivity) – it seems completely open to disagreement whether Mozart or Beethoven was more original, but there are other questions – like “Was Mozart’s music more original than that of the average Elvis impersonator?” for which no sane disagreement is possible.

But it’s a lot harder to believe there’s an originality-detecting organ in the brain than that there’s a music-appreciation-organ or a taste-detecting-organ. Even for the very vague and sloppy definition of “organ” being used here.

Or another question: is Pluto a planet? The correct answer is “meh, stop arguing about definitions, whether something is a planet or not isn’t an objective fact about the universe”. But is my left foot a planet? Here the correct answer is “no”.

So I think of ambijective statements as being undefined over a whole set of possible meanings. For example, “is X a planet” is undefined over:

1. is X larger than most moons, but smaller than most stars?
2. is X spherical under its own gravity?
3. does X orbit a star directly?
4. does X have a regular orbit in terms of ellipticalness and orientation to the plane?
5. is X a natural body made of rock and gas and stuff like that?

Pluto satisfies 2, 3, and 5, but arguably not 1 and 4, therefore it’s “subjective” whether or not it’s a planet insofar as you can choose which of these definitions you want to use. My left foot doesn’t satisfy any of these criteria, so anyone claiming it’s a planet doesn’t have a leg to stand on.

Moving back to the first question: whose music is better, Mozart’s or Beethoven’s? We can cash out “better” in several ways:

1. I enjoy it more
2. You enjoy it more
3. More people prefer it
4. It’s more famous
5. It satisfies the sorts of rules music theorists talk about more precisely

Worse, each of these definitions is *itself* underspecified. For example, criteria 1 could vary based on which song we’re talking about – do I have to enjoy Beethoven’s best song more than Mozart’s best song, or the average Beethoven song more than the average Mozart song?

It could vary based on when you’re asking the question – maybe Mozart speaks to me when I’m sad, but Beethoven when I’m happy; is this averaged over all possible moods, and are we weighting it for the moods I’m most likely to have?

It could vary based on what you mean by “enjoy” – maybe Mozart creates more powerful emotions in me, but I am more impressed by Beethoven’s technical precision, plus maybe Mozart only produces sad emotions in me and Beethoven inspires me to love, and what if I am impelled to listen to Mozart songs more often but when I do listen to Beethoven I find I like him better but that mysteriously fails to translate into more Beethoven-listening time?

We could probably break even these sub-sub-questions down, and go through the same procedure for each of our original five criteria, until we have hundreds or thousands of extremely specific questions. Eventually we will bottom out in objective questions, the sort you could solve by scientific experiment if you wanted to – for example, if we decided on a very specific rating system for songs, we could ask me to rate a specific performance of one randomly chosen Beethoven song and one randomly chosen Mozart song each morning when I woke up.

(In practice, maybe there are infinitely many questions, but at some point the questions become so similar that the difference between them become noise that we no longer care about. For example, "do I prefer Beethoven at 10:00 AM" and "do I prefer Beethoven at 10:00 AM + 1 microsecond".)

So suppose we have 1000 objective questions which all combine to form the question "Who is better, Beethoven or Mozart?" The "subjectivity" comes in not just in who we're talking about (eg my music-appreciation-organ which is a little different from your music-appreciation-organ) but in how we weight these different questions in composing the meta-question "better". This is a purely linguistic problem – if we have any disputes after this, we're arguing about definitions.

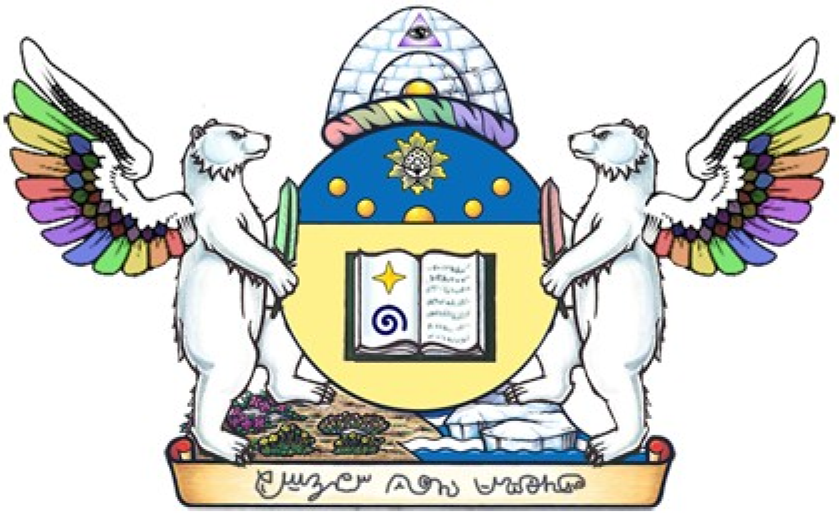
This explains why it's not "subjective" that Mozart is better than my upstairs neighbor. All those 1000 questions are very closely correlated, so it may be that my upstairs neighbor doesn't win on *any* of them, and therefore there's no way to compose the term "better" in which my neighbor could possibly be better than Mozart. Or maybe my neighbor only wins on two of the thousand questions, and no one has a definition of “better” which weights those questions higher than the remaining 998.

This post is mostly just so I have a word to refer to this kind of thinking quickly next time I get stuck in a subjective vs. objective dispute.

Raikoth: Laws, Language, and Society

I don't believe society is perfect the way it is, I'm just a small-c conservative who's cautious of changing things. We've never *tried* a complex globalized urban Information Age society with communist/libertarian/reactionary policies, and if you accidentally bring down civilization an "oops" isn't going to cut it. In real life I'm pretty risk-averse and that shows politically. And when I find sweeping revolutionary changes I *would* be up for, by bad luck they're all ones well outside the mainstream of debate, which focuses disproportionately on stupid tribal issues anyway.

So I am a terrible political activist. But I *have* spent the last thirteen years constructing a utopian society in my basement.

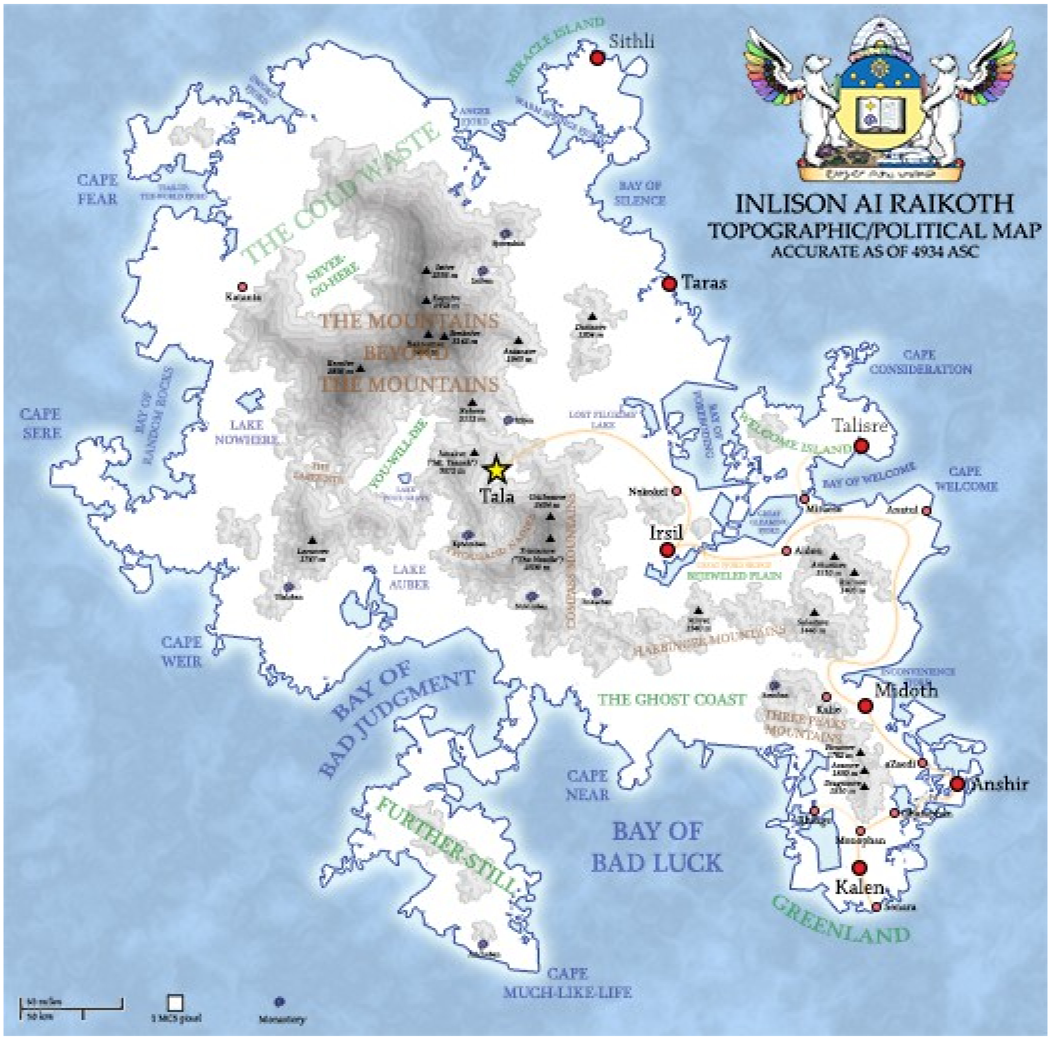


Coat of arms of the Shining Garden of Kai-Raikoth

I've been reluctant to bring it up here, because it is not relevant to sober politics, parts of it are amateurish and unfinished, and it makes the worst extremist look sane and grounded in reality. And I was afraid it wouldn’t really make sense without a little bit of the background of the world it takes place in.

But I’ve already outed myself as a conworlder. And I’m bored. And maybe it will shut some people up about how boring my politics are. So…

In the north of the planet Micras lies the island of Raikoth, about the latitude of Iceland and the size of Great Britain. In the present day (the 51st century ASC) it has a population of about 8 million – similar to that of Honduras or Israel.


INLISON AI RAIKOTH
TOPOGRAPHIC/POLITICAL MAP
ACCURATE AS OF 4934 ASC
MIRACLE ISLAND
Sithli
CAPE FEAR
THE COLD WASTE
NEVER-GO-HERE
BAY OF SILENCE
Taras
THE MOUNTAINS BEYOND THE MOUNTAINS
CAPE CONSIDERATION
CAPE SERE
BAY OF RANDOM ROCKS
LAKE NOWHERE
LOST PILGRIMS LAKE
WELCOME ISLAND
Talisre
YOU-WILL-BE
Tala
BAY OF WELCOME
CAPE WELCOME
Irsil
BEJEWELED PLAIN
LAKE AUBER
HARBINGER MOUNTAINS
CAPE WEIR
BAY OF BAD JUDGMENT
THE GHOST COAST
Midoth
THREE PEAKS MOUNTAINS
Anshir
CAPE NEAR
FURTHER-STILL
BAY OF BAD LUCK
Kalen
GREENLAND
CAPE MUCH-LIKE-LIFE
Monastery

Map of Raikoth. Click to expand thumbnail.

Raikoth is an ancient civilization, but during the 11th century ASC they ceded their sovereignty to the Holy Empire of Sxiro, a sprawling feudal conglomeration occupying nearly the entirety of the continent to their south. In exchange, the God-Emperor of Sxiro granted them extremely favorable terms including almost complete self-government.

The Raikolin are not a people prone to self-doubt. Their constitution – which they adopted only when the God-Emperor absolutely insisted all lands under his dominion present him with a charter of some sort – reads simply:

> In all situations, the government of Raikoth will take the normatively correct action.

For most of the 5th millennium ASC, their laws insisted that the Platonic Ideal of the Good was both head of state *and* head of government; this was relaxed in 4682 to allow the Archpriest of Joy to serve as head of government, again at the insistence of a confused and annoyed Sxiran God-Emperor.

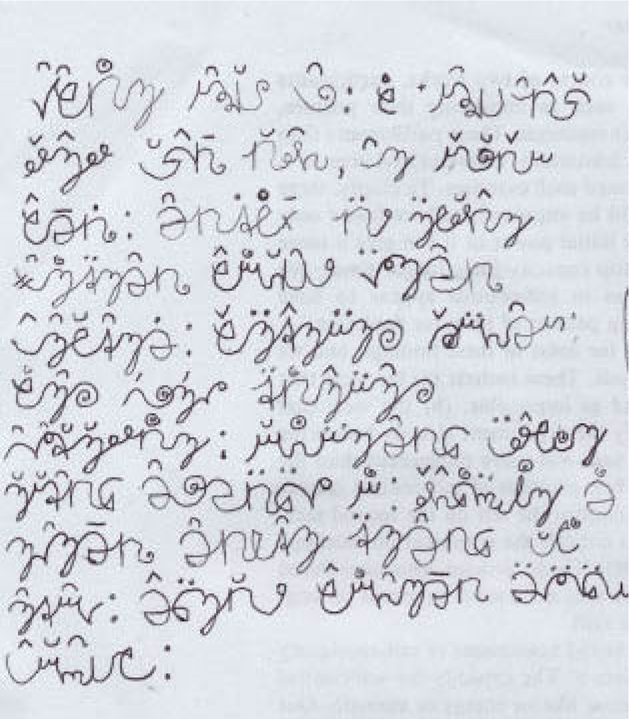
The Raikolin themselves are happy to explain the secrets of what they consider their success, though no other land on Micras has quite taken them up on the offer. They attribute it to three principles which they have pursued to their logical conclusion: Perfect Language, Perfect Government, Perfect Population.

Kadhamic: A Perfect Language

Early in their history, the Raikothin state was plagued by *politics*. Everyone had a different opinion for which way the state should go, everyone thought everyone else’s opinions were wrong, and debate was interminable, hostile, and unable to make any inroads into the problem.

Some began to wonder if the problem was the inherent slipperiness of language. Poor language encoded false assumptions about the world, allowing conflation of factual statements, value statements, and meaningless feel-good cliches. It led to confusing categories, framing effects, and a bad habit of reifying ad hoc concepts.

Luckily for Raikoth, an unusually high percent of its population were logician-monks (don’t ask) in a good position to begin working on an alternative. Starting with an outlook akin to logical positivism, they invented a language whose grammar was identical to philosophical rigor and in which every concept had to be expressed precisely. Cheap shots, ad hominem attacks, rhetorical tricks, appeals to emotion – all were painstakingly prohibited by careful choice of vocabulary and syntax, and ambijectivity is carefully collapsed into its component pieces.



A sample of Kadhamic writing from the Codex Hamiltonensis

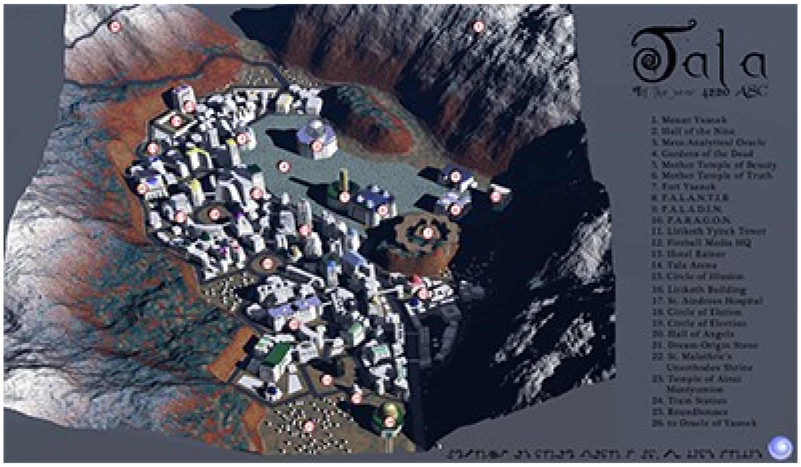
The thorniest problem was moral speech: what should “good” mean in a perfect language? Here the logician-monks spent millennia refining a utilitarian superstructure until it described exactly how to aggregate preferences to create a utility function for the entire society. The resulting book, the Risurion-Silk, was declared to be literally an incarnation of God, who was, after all, only another word for the concept of maximal goodness. “Good” was declared to mean “increasing the function described in the Risurion-Silk” or alternately “increasing the degree to which the Universe instantiates God”.

Thus was born Kadhamic, the perfect language. Having developed it, the government merely banned all discussion of politics in any language other than Kadhamic. This had two salutary effects. First, it prevented anyone who lacked the intelligence to learn Kadhamic from participating in political life. Second, it ended the worst sorts of political disagreement almost overnight, as constructing specious arguments for false positions became difficult or impossible.

The Shining Garden: A Perfect Government

Nevertheless, substantive political disagreement – genuine debate about facts or values – still remained. Before the Fallow Time in the 35th century ASC this was mostly addressed by a parliament of representatives from the nine cities, but during the leadership of High [untranslatable] Nithi Kirenion and afterwards, a more elegant system was put in place: the rule of the the Angels.

An Angel is a creature that bridges the gap between God and Man. The Angels of Tala, the capital of Raikoth, are massive computer systems that allow humans to make calculations about the value of God as instantiated in the Risurion-Silk.



City map of Tala as it was eight hundred years ago, before the Shining Garden era. Click to expand thumbnail.

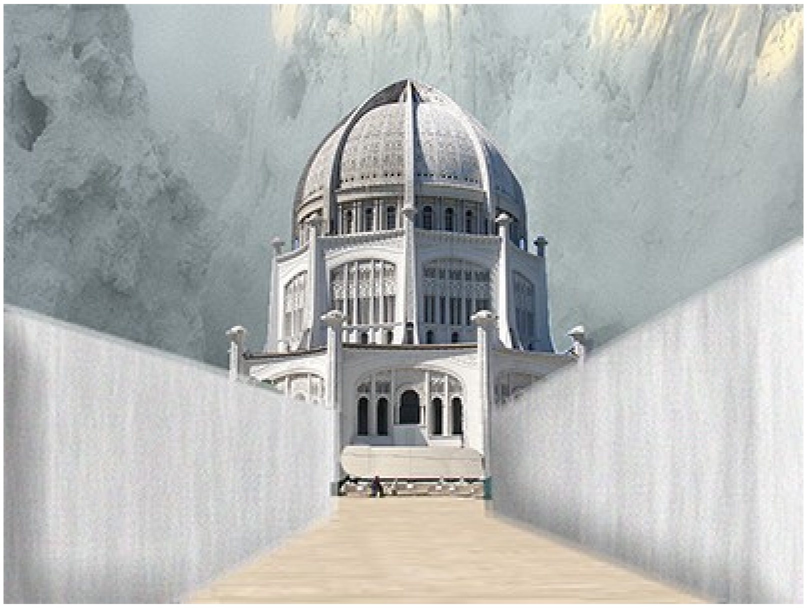
The Angel of Preference takes census and survey data from every citizen of Raikoth, runs it through the series of utilitarian superstructure functions in the Risurion-Silk, and spits out a utility function. Its calculations bear the same resemblance to the QALYs and DALYs of today as a supercomputer does to an abacus. What it ends up with are a series of preference weights for possible world-states and for various subcomponents of those world-states. It knows the marginal value of saving three people from poverty and how that compares to the marginal value of decreasing water pollution 6% in the Great Gleaming Fjord, or of one new baby being born in Uolrhaphen.

The Angel of Evidence is a system of linked "oracles" – what we would call prediction markets – heavily subsidized by the government and played not only by thousands of Raikolin but by the great financial conglomerates of Sxiro and even further afield. Fed sufficient time, money, and publicity, it can calculate the likely effects of any policy with an accuracy even the Priests of Truth cannot match. When Endi Arusion wanted to declare sanctions on Lirikoth, he asked the Angel exactly how long their rebellion would last;

according to legend, it got the length of the centuries-long conflict correct *to the day*.

The Angel of Salience accepts suggestions from anyone on Raikoth, whether they be a destitute fisherman or a High Priest of Joy. These suggestions can be any form of policy change, whether that be tax relief, new land use regulations, or the invasion of a foreign power. The thousands of suggestions it receives each day get crowdsourced, with average citizens and experts giving them upvotes or downvotes whose weight is proportional to those users' past success rates. Eventually, the most creative or well-thought-out proposals rise to the top and become salient for further investigation.

The Archangel, located in a vast temple in the Gardens of the Dead, combines input from all three Angels to render government decisions. It takes policy proposals from the Angel of Salience, uses the Angel of Evidence to estimate their likely effects, and finally runs those effects through the Angel of Preference to determine whether they would raise or lower the utility function represented therein and so bring the world closer to or further from the Divine. The set of noncontradictory policies that most satisfies the Angel of Preference becomes the law of Raikoth.



The Meta-Analytical Oracle, home to the Angel of Evidence and parts of the Archangel ([real world source]())

In her landmark work on governments, Zelde Kalirion once called Raikoth a “totalitarian anarchy”, and with good reason. The Angels’ decrees are entirely impersonal. It might demand a city uproot itself and move elsewhere, or strike down a doctrine in the national religion, or even tell everyone in the country to render a certain number of opals to a certain engineering company in exchange for their constructing a bridge.

Yet there is no single entity called “the government” which executes these requests. There are the Rhavakal, a self-contained chivalric military order with an obsessive focus on martial arts – and occasionally the Angels order them to invade things or arrest people. There are the Priesthoods of Truth, Beauty, and Joy, three very wealthy and powerful religious groups – and occasionally the Angels order them to take care of some errand. There are a host of companies whom the Angels sometimes order to perform one task or another. And there are normal everyday citizens, whom the Angels often ask for financial support for one of their plans or another. But all of these groups alike are simply tools whom the Angels make use of. And so – with the possible exception of the Priests of Joy, whose tasks include maintaining the Angels' physical forms and tweaking their algorithms – none of them consider themselves (or are considered by others) part of a body called "the government".

The Galisyin – A Perfect Populace

"Galisyin" means "the gardened ones", and it is a name the Raikolin give themselves. They believe that even a perfect government cannot help bad people, and even a flawed government cannot keep down good ones. A disaster eight hundred years ago led to the ideology of *istilve iab istisemial priktino* – "fractal perfection of which every part is itself perfect", in which the innate virtues of the Raikothin people must be robust enough to survive even a total collapse of their society.

The gardening begins with birth and reproduction. All men get RISUG as part of what we might call their high school health class; although they may reverse the process at any time, in practice no one does so until they want to have a child, and the procedure can be repeated free of cost at any time. This effectively prevents unplanned pregnancies and unwanted children

The punishment for most serious crimes is exile – either to the many colonies like Kymrikoth, Kalirphanam, and Psentikoth, or to monasteries in the mountains (which serve about the same role prisons do in our society except that instead of hanging out with hardened criminals they mostly hang out with monks and do monastic labor, meditation, and study). In either case, they are effectively removed from the breeding population for long periods, sometimes forever.

Finally, as a country that passed the demographic transition a few millennia ago, the natural fertility rate is well below replacement. The Angels keep the population at their desired level by paying people to have and raise children, and the highest price is given for people who are willing to use gametes selected by the Priests of Truth as unusually likely to create good people (where “good” is as always defined as “tending to increase the function described in the Risurion-silk” but usually involves intelligence, compassion, sanity, health, and creativity).

Although these three measures do relatively little over the space of a single generation, over hundreds of years they have transformed the populace to one that scores well above the Sxiran baseline on nearly every desirable trait.

the mother temple
of Elith
Tala, Raikoth

Headquarters of the Priests of Truth

In addition the Galisyin ideal includes euthenics – bettering people by improving their environment. My old

Biodeterminists’ Guide to Parenting is an *very* Raikothin way of looking at the world and its suggestions (refined and tested by centuries of constant study) are law, sometimes holy law. Some euthenic interventions are more proactive: for example, there is an exactly optimal amount of lithium in the water supply. There’s even occasional eradication campaigns for toxoplasma (at least in men), although the Arctic location itself does most of the hard work in decreasing parasite load.

Along with the biological work comes the education, which has almost nothing to do with what our own society calls by that name. There is little attempt to teach history, science, music, or grammar – it is assumed that a correctly biodetermined population with access to free libraries will pick all of that up by themselves. Instead school is entirely about instilling good habits of mind that will make people virtuous and prosocial citizens.

The curriculum is heavy on meditation; students spend five to ten years meditating for an hour a day under the guidance of Priests of Beauty who are expert in the discipline. The goal is to achieve near-perfect self-control, low anxiety, compassion for others, and lasting happiness (if any of the studies showing that meditation improves academic performance prove true, that’s just icing on the cake).

A second class that wouldn’t be recognizable in real-world schools is Experimental Theology, aka nationwide school-sponsored drug use. The idea here is that years of research with chemicals like LSD and psilocybin have put their ability to permanently positively alter personality on a sound scientific basis, and as children’s minds develop they are inducted into different levels of mystical experiences with these drugs in ways that give them more love, energy, and feelings of community.

The third and most important class is rationality. This class is the difference between dystopia and utopia; it intends to teach everyone the rationale behind their society and give them the tools to engage in politics or choose not to do so. First is instruction in the perfect philosophical language Kadhamic. Then is instruction in a sort of turbo-charged form of critical thinking only possible in Kadhamic (but which I'd like to think the Less Wrong Sequences are a close vernacular approximation to). Then come enough math and science to understand something sort of analogous to our Rational Choice Theory, Game Theory, Evolutionary Psychology, Decision Theory, et cetera – things necessary to understand morality and government design. Finally comes an investigation into morality, the justifications behind the Risurion-silk, and the heretical logics which compete with it.

At the end of these lessons comes a sort of *rumspringa*, where young men and women leave the Shining Garden and tour the rest of Sxiro and the world, observing its freest and richest regions as well as its poorest and most desperate. They are then offered a choice between returning to Raikoth, living in the "colonies" – a series of outlying lands inhabited mostly by Raikolin who reject the heavily ordered life of the Shining Garden – or moving to mainland Sxiro. Those who choose the last option are given free remedial training in the Sxiran language and help integrating into local culture – but most find it bizarre, off-putting, and barbaric, and choose to stay in Raikoth after all.

During adult life, Raikothin continue their self-cultivation with the help of the Priests of Beauty, who are sort of combination clerics / counselors / psychiatrists. They have more the friendly attitude of a pastor than the saccharine attitude of a therapist or the clinical attitude of a doctor, but if necessary they also have the expertise to prescribe a mind-boggling variety of drugs and supplements, many completely unknown or ridiculously illegal here on Earth. They differ from more typical therapists in having complete confidentiality; they are forbidden from reporting their flock to the authorities or to mental institutions for any reason (other organizations do have those powers, but don’t interface with the Priests of Beauty). This tends to inspire a relationship of trust, and the Priests of Truth watch them to make sure they deserve it.

Those who desire a level of perfection beyond that of normal Galisyin head to the monasteries, about a dozen or so isolated communities the size of small cities in the high mountains. There they grow their own food, meditate, study, and try to live more or less in harmony with nature. Several monasteries are under oath to take in anyone who wants to join them, setting a valuable floor to the level of misery and poverty anyone has to put up with.

This is the first time I’ve shown Raikoth to anyone outside the tiny community of conworlders, so if you have any questions ask and I’ll try to expand upon them.

Raikoth: Corruption, Priesthood

Some of the comments to my last post required essay-length responses, plus I will keep talking about my conworld until somebody shuts me up. This one is in response to Nemryn on corruption.

Third Eyes are little lifelogging cameras provided at government expense. They intermittently take pictures + video and record conversations and wirelessly upload them to a central database.

Oh no! Dystopian society in which the government can monitor everything you do, right?

Not exactly. The central database is encrypted, and only the wearer of a Third Eye knows its master password. This master password can be used to generate daily passwords, which unlock the data for a specific day without allowing reconstruction of the master password. *At the request of the wearer* and the provision of the daily password, the caretakers of the central database can decrypt a day's recordings.

This soundly discourages crime. A falsely accused person can authorize decryption of their recordings for the time the crime took place, potentially proving their alibi. And a victim can authorize decryption of *their* recordings, proving their story to be correct.

“Decryption”, in this context, means that one of the Priests of Truth who runs the central database will watch the recordings and then send back a single bit of information. For example, a judge might send a request to know whether someone committed an assault on a certain day, and the accused might release the password to their Third Eye. The Priest would then watch the recordings, send back a “yes”, “no”, or “it’s complicated” to the question, and then swear an oath never to discuss the recording further without the subject’s consent. So if you didn’t commit a particular crime because you were busy having an affair with your best friend’s wife, your secret is safe.

Average citizens are not *required* to wear Third Eyes – that would be dystopian – but most people quickly see the benefits. But the Rhavakl, who serve as the police force, are required to wear them. The Priests of Truth, civil servants who might otherwise be subject to bribery, are required to wear them. Anyone in a position of power or vulnerable to corruption charges agrees to put on a Third Eye before being trusted with their office.

None of these people are ever *required* to disclose their password, of course. *That* particular rule is sacrosanct. But if you’re wearing a Third Eye, and you refuse to give access to it in a criminal proceeding, no one is the least bit reluctant to take that as an obvious admission of guilt.

The system isn’t perfect. Third Eyes neither record every single moment, nor can they upload their data to the central database in real time. It’s possible to accept a bribe very quickly, hoping you get in between Third Eye recordings. And clever muggers will demand that people hand over their Third Eyes during an assault, then break them so they don’t have a chance to transmit their data (some of the newer models have panic modes, but they don’t work 100% of the time).

But the device operates on a random schedule and never gives any detectable sign that it’s on. So anyone who wants to do something illicit has to hope they get lucky, and a lot of people aren’t willing to take the risk.

And the uncertainty inherent in Third Eyes is a feature, not a bug. It is acknowledged practice to hold trials before anyone knows whether the recordings from Third Eyes address the problem. The confirmation from Third Eyes then provides useful training data for Priests of Truth.

The Priesthood of Truth

The most important evaluations of factual questions are performed by prediction markets: either the Angel of Evidence in the Meta-Analytical Oracle or smaller oracles that address purely local concerns. But sometimes there are questions not worth a market’s time or energy, specific questions about individuals that need to be made dozens or hundreds of times per day. For these questions there are Priests of Truth.

Priests of Truth are humans who have been trained to have judgment approaching that of a prediction market. To be accepted into the Priesthood, you must cultivate near-perfect calibration and accuracy compared to both other humans and predictive algorithms. Some tests face the candidate off against a prediction market they are not allowed to view. Others make them evaluate training data – cases from years ago in which the right answer is already known.

An Arch-Cathedral in Tala – because I already used up all my pictures of more relevant buildings in the last post

Most Priests go on to specialize. Some become judges for criminal trials, skilled in predicting whether a suspect actually committed a crime from the testimony available. Others work on the opposite end of the justice system, skilled at predicting whether offenders have been sufficiently rehabilitated that they will not offend again. Others are the ones who select gametes for the fertility program, skilled at predicting from genetic information and parental histories whether they will produce healthy and happy children. Still others are the ones predicting who should be offered incentives to give birth to and *raise* these children, based on predicted or observed parenting ability.

Others choose to serve the private sector. There are Priests of Truth who specialize in determining whether marriages will last; for a fee you and your significant other can meet with them and take some psychometric tests and they will tell you the chance you'll still be together when you're fifty. There are Priests who will tell you how likely you are to be happy at age 50 depending on what job you choose. There are Priests for nearly anything, and smart people consult them before any big decision.

All predictions by all Priests get entered into – you guessed it – a centralized database – and each Priest receives a score based on how much better or worse ze has done than other predictors – again including other priests, predictive algorithms, prediction markets, and the petitioner themselves. Those Priests who fail to beat chance or very simple toy models are expelled from the Priesthood, although given the rigorous initiation process this rarely happens. Those Priests who can beat chance but do poorly relative to other Priests end up in less important positions or out serving tiny communities in the Cold Waste. The most effective and accurate Priests advise the government on delicate tasks, earn wealth and fame, or become High Priests of entire cities.

But being a Priest of Truth isn’t just about being good at figuring things out. The Priesthood is a genuine religious order dedicated to the worship of Truth, one of the two (occasionally three, rarely four) gods of Raikoth. Their philosophy is that you can’t predict things well without learning not to lie to yourself, and you can’t learn not to lie to yourself while selectively preserving your ability to lie to others. Therefore, all Priests of Truth, upon their initiation, swear never to tell even the slightest lie, or to break even the most trivial oath. This commitment makes them useful as an incorruptible class who can evaluate possible corruption in anyone else.

Not that anyone takes this *on trust*. The Priests of Truth wear the biggest and most sophisticated Third Eyes on the island.

And although Raikoth abhors the idea of an involuntary death penalty (citizens may *choose* death in preference to exile or other punishments, but it is never forced upon them) this Priesthood is the sole exception. Any Priest of Truth caught lying, cheating, or breaking an oath is subjected to a shockingly primitive custom that has remained intact for over five thousand years since the pre-Apollonian era: they are thrown into the crater of Ianakve, the sacred volcano.

Also a convenient source of geothermal power!

Raikoth: Cities, Land

(continuing to talk about my conculture)

Like all interesting people, Raikolin live in yurts.

The logistics are not too complicated. Land is sold or rented in the form of “campsites”, with spaces for yurts and hookups for plumbing and electricity. Poorer families might choose to live on the outskirts of cities instead, where the land is free but they have to improvise their own utilities (some live off the land in the summer, when the midnight sun provides plentiful solar power, then retreat to cities in the winter).

The yurts are easy to move with a couple of days’ notice, supporting a semi-nomadic lifestyle among country-dwellers who go to pastures and fishing grounds in the summer, then return to the cities in the winter. And their mobility allows changing family configurations – if Grandma is too old to support herself, the family can just move her yurt next to everyone else’s; or if a growing child wants more independence, she can pitch a yurt a short distance away from her parents’. Larger or wealthier families create larger dwellings suitable for their needs with modular yurts (see this site, which happily proclaims “No longer do you need to buy the biggest yurt you can!”)

Home sweet home

Most important, yurts are cheap. In our world, even a very large yurt suitable for an entire family with lots of furniture costs $10,000 – $20,000; in Raikoth, with economies of scale, they cost barely half that. There are no housing bubbles, no mortgage loans, no subprime crises, and no foreclosures. And with the cheapest, simplest options costing barely $2500 to buy (and less to rent) there's very little homelessness either, and the mobility and deconstructability of houses makes it easier to donate them to the needy when you're done with them.

A few Raikolin abandon the traditional yurts to live in permanent stone or wood housing in city centers. These people are not trusted. "House-dweller" has the same connotations that "arugula-eater" has in 21st century America.

But there *are* more permanent stone buildings in city centers. A typical Raikothin city is a complex of temples, libraries, office buildings, factories, etc surrounded by fields of yurt campsites surrounded by city walls (previously functional, now there mostly to stop sprawl and because they're pretty). Interspersed are a uniquely Raikothin art form, the dead garden – a botanical garden of plants made entirely of glass and jewels at latitudes much too cold for the genuine article.

The Raikothin city of Tielion Zaedi on Micras' Minecraft server

Travelers to these cities tend to describe them as "eerie" and "ghostlike" for a few reasons.

First, no cars are permitted within the city walls. They're loud, dirty, dangerous, and prevent social interaction in the streets and public spaces. All the streets of a city are walkways, usually tiled. In compensation, there are excellent mass transit systems. Some cities use subways; others use personal rapid transit systems or string rails. All are effectively silent and invisible at ground level, leading to a sense of openness and safety completely absent from most other cities.

Second, the cities are *spotless*. Raikoth subscribes to an over-the-top exaggerated version of the broken window theory – that unpleasant and befouled environments increase antisocial behavior. They also believe that attention restoration therapy is not entirely about nature but about clean and organized places versus dirty and haphazard ones. Threats to the urban aesthetic are considered a tragedy of the commons in which the common resource is citizens' health, intelligence, and emotional development – and the Shining Garden system exists to aggressively coordinate commons situations.

Chewing gum is illegal in Raikoth for much the same reason as in Singapore, and littering is punishable by indefinite exile to the nearest monastery. Foreigners have protested this latter policy, but the Raikolin stand firm, claiming that although theft can sometimes be legitimately due to poverty, and murder can sometimes be legitimately due to blind rage, littering is pretty much a deliberate and inexcusable "fuck you" to the entire rest of the community and the sort of people who find it acceptable should be separated as far from decent society as possible lest whatever is wrong with them spread. Graffiti is so far beyond the pale that no one even knows what the punishment would be, but it probably involves volcanoes in some way.

(the use of Third Eyes makes all of this extremely enforceable – if you see someone littering, you politely ask them to pick it up, and if they don't, you go to the Rhavakl and release your records to them)

On the same principle – a reduction of visual clutter – outdoor advertising is banned and streetlights are rare and directional. Between the streetlight thing and the no cars thing, there is almost no light pollution.

Third, draconian noise restrictions. The Angels didn't go through the trouble of banning cars to prevent noise pollution so that you could play your stereo at a hundred decibels. All the eugenics to get an average IQ 130 populace left a whole bunch of people with sensory processing disorders, and yurts aren't exactly super-soundproof. There are well-specified public nuisance restrictions with actual numbers-of-decibels attached to them, and if someone is violating them, you ask them to stop, and if they don't, you go to the Rhavakl with recordings from the Third Eye. Consider the level of courtesy the Japanese developed in order to live in tiny sardine-like apartments, then multiply it by about thirty.

Land

Not everyone likes these restrictions. A surprising number of people *do*, because they're easier when you've been raised with them, and they make a lot of sense in a country with high prevalence of sensory processing disorders. But there are some people who just want to live as free as possible without anyone telling them what to do.

Luckily, ninety percent of the country is frozen wasteland.

As suggested in the map from the first post, pretty much all the large cities and towns of Raikoth are on a tiny strip along the eastern coast, the fertile and hospitable area between the mountains and the sea (you may recognize this as *also* pretty Japanese). The rest of the country – not to mention the deadly freezing north, is pretty much all wilderness.

And so while most people stick to the civilized regions on the east coast, a strange collection of hunters, trappers, monks, hermits, anarcho-primitivists, extremely determined geocachers, and people who just want to get away from it all live in the west and sometimes even construct weird experimental communities there. The Raikolin are extremely proud of their option to leave civilization at any moment and romanticize those who choose a life in the wilds, even to the point of forbidding any development west of the mountains to avoid interfering with their lifestyle.

Some of these communities are monasteries, self-sufficient communities built on high mountains or in remote valleys and inhabited by a strange combination of logician-monks, exiles, and just plain people who want to get away from it all. Days consist of a combination of work and study. A few specialize in the rehabilitation of exiled criminals; these latter are welcomed into the monastic community as full and equal members – but like all full and equal members, if they don’t work they don’t eat. A surprising number stay after their sentences are finished, which suits everyone just fine. The largest such community is Katanta, “the City of Hermits”, which stubbornly refuses to tell anyone its exact location and which tends to move whenever it is discovered.

Beyond these oases of civilization lie even wilder places. Named for their inhospitability and deadliness, regions like Never-Go-Here and You-Will-Die are the ultimate escape from everything, usually including the mortal coil. There are no roads, no good maps, nowhere to get help, and rumors persist that the central government jams all wireless communications there for uncertain reasons. Nevertheless, it is considered *totally sacrosanct* that anyone who wants to pack up and go live as a mountain man out there should be allowed to do so, and every so often one of them even survives or comes back home.

The people who live here think where ***you*** *live is bleak and hostile to life*

Raikoth: Symbolic Beads

In the comments to yesterday’s conworlding post, Nate asked:

> I would expect that wealthier people would buy larger yurts to show off their wealth, and then those would get gradually less movable until they’re basically living in houses. Since this doesn’t happen, what status and signaling effects are they optimizing for?

I answered that if you’re trying to design a crazy crackpot utopia, you better have some plan on how to redirect the urge toward status games.

I still think status games are one of the most blitheringly idiotic parts of the human experience. When people have to spend all their money on competing over zero-sum prestige items, it would literally be exactly as beneficial to just set a heap of dollar bills on fire. But they’re a pretty ingrained part of the species and just saying “Please don’t engage in status games” isn’t going to cut it.

The Priests of Joy decided to attack this problem by co-opting status games for their own purposes. They initiated a tradition of wearing symbolic jewelery – usually a necklace made of brightly colored beads. Reading from left to right, each bead’s color advertises to potential interaction partners a certain fact about the wearer.

I don’t have strong feelings on the exact configuration of these necklaces, but the beads might include:

— Is the wearer actively looking for more romantic partners? (think combination of red light green light party and checking to see if someone has a wedding ring) Different colors can connote anything from “taken, stay away” to “I will have sex with pretty much anyone who asks” to “I am only looking for stable long-term relationships”

— Gender and sexual identity. Useful for transgender people who want a quick way to tell people what gender to identify them as, and to tell if someone is gay, straight, bi, etc and avoid awkward misunderstandings.

— Is the wearer an introvert or an extravert? Can have any meaning from “I love talking to random people, please approach me” to “Never talk to me under any circumstances” to “Well, I guess if you have something *really* interesting to say…”

These beads are all in fixed positions, and an average person can read and interpret them about as quickly as an average member of our own society can read and interpret stoplights or flag pins.

After these fixed beads comes a space where anyone can add any beads they want. Usually these are beads declaring allegiance to a particular social protocol – for example, once Crocker’s Rules are invented, someone can publicize the design of a Crocker’s Rules bead, after which everyone who wants to publicly declare Crocker’s Rules can wear that particular bead and if it is sufficiently well-known their interaction partners will know what it means. Another popular social protocol bead is “ask culture”, signifying that the wearer has promised to try not to be upset if you ask them for anything, even something they consider unreasonable.

Some protocol beads (signified by an octahedral shape) are reciprocal, meaning that they only apply to other people who wear the same social protocol bead. For example, a group might decide to take up radical honesty, but only feel the obligation to be radically honest to other people trying the same strategy. So this bead might signify “I intend to be radically honest to anyone else wearing this same bead”. Other social protocol beads are ridiculously complicated, basically swearing allegiances to entire constitutions governing the wearer’s social behavior.

As these get tested, they split society into various different subcultures. The ones that people like become popular, expand, and eventually become universally accepted mores. Others die out quickly, or remain confined to a small population of enthusiasts (the “I prefer to speak in Kadhamic all the time” bead has been fixed at about one percent of the population for over a millennium)

Obviously wearing these beads is pretty useful, especially in a society where everyone else wears them and you’re viewed as stubbornly withholding useful information if you don’t, such that anyone offending you would be your own fault. So after a few decades, when not wearing them would be considered unthinkable, the Priests introduced one final section of the necklace.

This last part contains beads that can only be granted by a Priest. And although there are no rules forcing people to wear beads, there *are* rules banning people from wearing beads that falsely represent themselves (essentially an anti-counterfeiting law; these beads are minted by the government), and this has pretty much the same effect.

Among the beads on this last section are one declaring in very vague terms how much you earn, and another indicating level of charitable donation. Suppose the only three income level beads are “I earn less than $200,000 a year", "I earn between $200,000 and $2 million a year" and "I earn over $2 million a year". Most people will wear the "less than $200,000" bead, which will be so common and so vague that it won't have much effect on your social standing. Rich people can either wear the bead correctly signifying their income, or they can wear no bead at all, in which case it looks like they have something to hide and which is pretty suspicious.

Now imagine other beads, granted for certain levels of charitable donation. Give $1,000 to charity and the government gives you a pretty opal bead. Pay them $10,000 and you get a sapphire bead. Pay $100,000 and you get an emerald. And so on.

Now if someone is wearing an "I earn more than $2 million" bead but no charitable donation bead, they kind of look like a jerk, especially in a world where everyone else *is* wearing charitable donation beads of various sorts.

In practice, most people both want to ensure that everyone knows they're rich, and make people not think of them as antisocial monsters. So most of them wear beads signifying their income *and* purchase charitable donation beads commensurate to that income.

There is much more uncertainty in the income-level beads than in the charitable-donation beads, which means that rich people who want to broadcast their wealth might *have* to use the charitable donation ones. For example, suppose you make $20 million, and a $1 million donation to charity gets you a diamond bead. Merely wearing the "Earn more than $2 million" bead might get you mistaken for one of those hoi polloi who *only* earns $2 million. If you have ten diamond beads, though, it's pretty clear you earn at least ten million and probably more.

This decreases costly status games in two ways. First of all, if everyone looks at your beads first thing after meeting you, and your beads very clearly declare you earn at least $10 million, it’s not necessary to prove the same thing by having a ridiculous platinum-diamond watch. Second of all, each non-bead status purchase you get makes people even more likely to check the necklace and think “Hmmm, this guy has an entire superyacht, but he only has one diamond bead on his necklace? Seems like a pretty selfish person.”

As a result of these factors, as well as a very little bit of subtly directed good-old-fashioned propaganda, by far the most important status game in the Shining Garden is *who gives the most to charity*. And that’s a status game that most people can get behind.

EDIT: Maybe this would require a law saying that if you donate a certain amount, you *must* wear the corresponding bead? That would prevent people from refusing to do so for not wanting to appear to brag.

Overheard in the Box

Davis: I used to think that secrets were like evil twins – something it was possible in theory to have, but nobody actually had one. In my sophomore year of college, one of my friends asked “Want to hear a secret?” and I said “Don’t be silly, secrets don’t really exist.”

Julia: When I was little, I thought the stars were a metaphor, something poets wrote about to inspire people. It wasn’t until I got glasses in second grade that I realized they were actually up there and everyone could see them.

Raikoth: Economics, Relationships

Someone in one of the comments asked whether I am actually pushing Raikoth as a perfect society, and the answer is sorta. My con-culturing philosophy is trying to optimize for goodness while erring on the side of weirdness. This stands in contrast to a reasonable philosophy of the real world, where one might optimize for goodness while erring on the side of safety.

The reason I mention this may become clear further down in this post.

I promise no more than one or two more posts about this before I get back to regularly scheduled programming.

Economics

The economic system of Raikoth takes the form of a basic income guarantee and very little else.

There's not much government intervention in health care – but people are strongly advised to spend some of their income guarantee on health insurance.

There's no minimum wage, labor unions, or workplace safety laws – but people are encouraged to remain unemployed and live off their income guarantee if they're not entirely satisfied with the job options available to them.

There's not even a *budget*, per se. There's a basket of taxes and the ability to slightly raise or slightly lower all taxes in the basket in order to fund individual programs. For example, a policy proposal would not take the form "Let's build a new bridge, it would only cost $100 million and I'm sure there's room in the budget for that", but rather "Raise all taxes 0.5% to get $100 million, then use that to buy a new bridge". Spending proposals not linked to revenue proposals would be considered incomplete; thus decision-making takes the form of real cost-benefit calculations instead of just saying yes to all pleasant-sounding spending and no to all unpleasant-sounding taxes.

It's not that there's no deficit – a small one is maintained for Keynesian reasons – but there's complete control over its size and it's never going to increase unexpectedly or against everyone's wishes.

The taxes in the basket are primarily land taxes, very high estate taxes, and a tax on large corporations proportional to their size. This latter is meant to approximately balance economies of scale and give small mom-and-pop stores and start-ups ability to compete on an equal footing. It is recognized that this makes things a bit less efficient, but the Angels have decided it is still a net good for hard-to-measure reasons.

The government also taxes externalities – carbon taxes, pollution taxes, noise pollution taxes – as well as some less obvious cases. One of the weirder ones is destroying-the-social-fabric taxes. In these latter, Priests of Truth predict whether some piece of media will cause increases in crime or racism or whatever, calculate the costs of these effects, and then levy them as a tax on the producer. You can still get violent videogames that demean women or whatever, but they're going to cost much more than the other videogames and that money is going to go to law enforcement, educational campaigns, and generally cleaning up after them.

Income taxes *supposedly* don't exist, but symbolic beads place wealthy people under extreme social pressure to donate, and given the government’s (widely believed) pretension of being a perfect utilitarian system, most of these donations take the form of “bonus taxes”, providing an extreme boost to the national coffers and subsidizing most of the basic income guarantee program.

Relationships

One of my goals in con-culturing was to subvert dystopian tropes in unexpected ways, and one of my least favorite dystopian tropes is Bureaucrats Or Computers Decide Who You Are Allowed To Marry.

So fine. Let’s give these people their frickin’ Bureaucrats And Computers Deciding Who They Are Allowed To Marry.

Temion Mirun is a festival held on May 1st (approximate, based on solilunar calendar), a sort of combination celebration of springtime / Valentine’s Day / group marriage day. Each year in the runup to May 1st, everyone who has undergone the coming of age ritual but is not currently married writes down the names of the people they’re attracted to and want to date, in order of how much they like them. Then they send their lists to – say it with me – a centralized database.

Each year, on April 30th, the computer with the database runs a special variant of the Gale-Shapley stable marriage algorithm on everyone’s ranked lists. On May 1st if it is sunny, or the first sunny day after if not, everyone gets together with their chosen partners and are handfasted in a group ceremony – where handfasting means that they are considered a couple until the next May 1st.

Springtime in Raikoth

Handfasted couples are polyamorous by default, and people still ask each other out on dates in the traditional manner. Sometimes dates people get one year become their handfasts the next; other times people end up with partners they have only admired from afar and didn't have the slightest idea returned their affection.

It is considered somewhat uncouth, but certainly allowed, to break up with your partner before the next May, unless there's something really horrible going on like domestic violence (which is always provable – good old Third Eyes). On the other hand, it is not considered unusual or a sign of dislike to change partners the next year, even if the relationship was mostly happy. It's not considered uncommon for A to handfast B one year, rank C higher the next year and handfast zir while maintaining a secondary relationship with B, and then decide B is better after all and go back to zir the year after.

Most people start considering marriage around the third year of being handfasted to the same person. Marriages are a bit like the handfasting relationship except that they are in theory permanent. There are a variety of different marriage contracts available, but most people go for neither no-fault divorce nor for full-on covenant marriage. The most common contract says that the marriage cannot be annulled except by the High Priest of Tala, and that the petitioner must go to Tala without using any form of motorized transportation. That is, they must walk through a few hundred miles of freezing tundra and then up the slopes of a towering ice volcano until they reach the holy city, where the High Priest will free them from their oath. The idea is to have a commitment mechanism: it’s not impossible to end a marriage, but it’s not something you do without a *lot* of thought and an absolute certainty that it’s the only remaining option.

Marriage contracts may also include other stipulations. Although handfastings are polyamorous by default, marriages may or may not be; a couple that chooses not to be polyamorous may include this in their marriage vows, at which point people who cheat become liable for civil penalties relating to contract violation. Although some couples will happily marry with nothing but standard marriage vows, others will include a whole host of rules that then become inviolable unless both sides agree to edit the contract together.

Some people never take these stronger marriage contracts, preferring to move from handfasting to handfasting for their whole life; this is quite acceptable. But those who do usually do because they want children. Not only are most people who ask for the removal of contraception married, but the Priests of Truth use statistics about differential outcomes in married and unmarried couples when deciding whom to offer opportunities for subsidized child-raising.

Raikoth: History, Religion

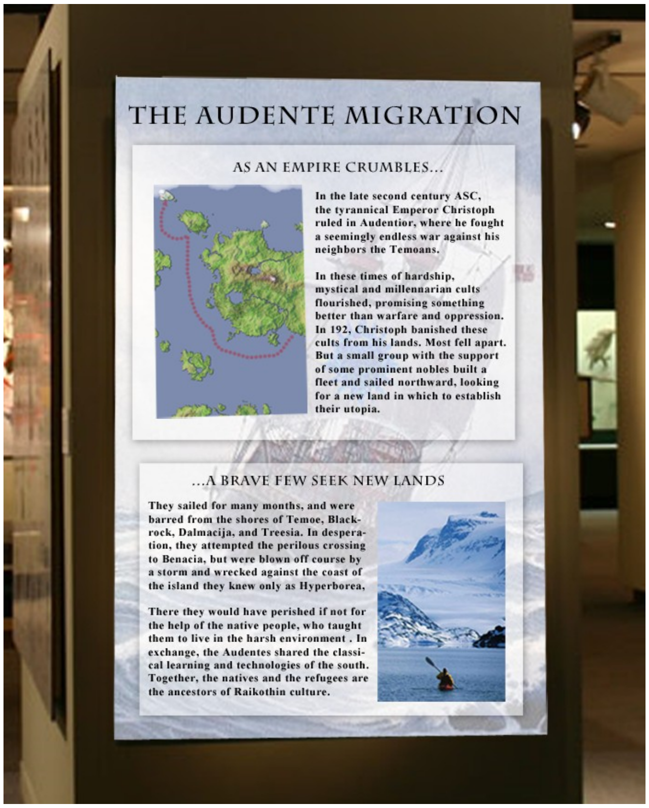
Index: Posts on Raikoth

Imagine the following world.

It’s Imperial Rome. Nero has just gotten pissed at Seneca, but instead of killing him, he orders him into exile. The other Stoics are getting kind of wary at this whole “insane dictator ruling a decadent and corrupt empire” thing, so they all agree to head off with him and start a colony dedicated to virtue and righteousness, far away in the new continent that has recently been discovered across the sea. They are somewhat better philosophers than they are navigators, and crash into Greenland – where they are, in accordance with the tradition, Mistaken For Gods By The Native Populace. Seneca starts arranging the local paleo-Eskimos into his conception of the Ideal State, but gets distracted by his discovery of Eskimo shamanism, which is ten million times more fascinating than any philosophy the Romans had to offer. He wanders off into the tundra to go on a spirit quest, and his new Greenlandic society collapses into civil war. This is ended only by the fortuitous return of Seneca, who has ascended a holy volcano, talked to God, and become a prophet.

Several thousand years later, Greenland hosts an thriving society of high-tech Latin-speaking Eskimos following a religion descended in equal parts from Greco-Roman philosophy, shamanism, and vision quest volcano prophecies.

This is about the quickest way to explain the history of Raikoth without going through a lot of boring pre-history and geography and hard-to-pronounce names.



THE AUDENTE MIGRATION

AS AN EMPIRE CRUMBLES...

In the late second century ASC, the tyrannical Emperor Christoph ruled in Audentior, where he fought a seemingly endless war against his neighbors the Temoans.

In these times of hardship, mystical and millennarian cults flourished, promising something better than warfare and oppression. In 192, Christoph banished these cults from his lands. Most fell apart. But a small group with the support of some prominent nobles built a fleet and sailed northward, looking for a new land in which to establish their utopia.

...A BRAVE FEW SEEK NEW LANDS

They sailed for many months, and were barred from the shores of Temoe, Blackrock, Dalmacija, and Treesia. In desperation, they attempted the perilous crossing to Benacia, but were blown off course by a storm and wrecked against the coast of the island they knew only as Hyperborea,

There they would have perished if not for the help of the native people, who taught them to live in the harsh environment . In exchange, the Audentes shared the classical learning and technologies of the south. Together, the natives and the refugees are the ancestors of Raikothin culture.

Exhibit on early Raikothin history from the 2011 Micronational World Expo. Click to expand

Religion

So what, exactly, is this religion?

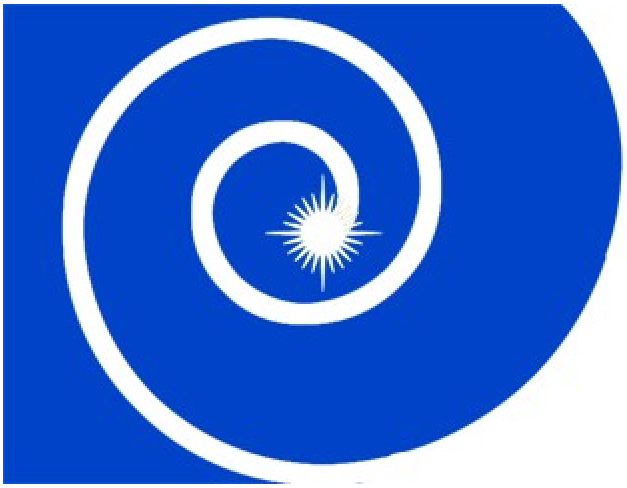
Raikothin religion, *sumurhe* in its own language, recognizes two aspects of God, called Truth and Beauty. The existing world is a poorly ordered mishmash of these two aspects, whereas God is the two aspects artfully and perfectly combined.

Truth includes everything that actually objectively exists, in the exact way that it actually exists. This aspect is mathematical, precise, and completely devoid of subjectivity. It is symbolically associated with winter, stars, the colors blue and silver, and all the hard sciences as well as math.

Beauty includes feelings, dreams, hopes, personality, meaning. This aspect is numinous, charged with emotion, and fantastic. It is symbolically associated with summer, roses, the colors green and gold, and all the arts, especially poetry and especially especially music.

The world is a gradual and halting attempt to integrate Truth and Beauty, and humans are the interface points at which the integration takes place. Humans join Truth to Beauty by bringing beautiful things into existence or interpreting existing things in such a way as to make them beautiful.

The pattern of integration of Truth and Beauty is self-similar at every level. That is, each individual object is its own attempt to integrate the two aspects; completion of this task perfects the objects (it’s maybe a *little* teleological). Completion of the integration at the highest level perfects the world, causing it to become a manifestation of God.



Truth is symbolized as a star, Beauty as a spiral; the national flag, the Star and Spiral, combines both

This philosophy cashes out into a formalization of two different ways of looking at things, the *Elith-mirta* and *Ainai-mirta* (Perspective of Truth and Perspective of Beauty). The *sumurhe* religion itself is a perfect example. In the *Elith-mirta*, it is a useful metaphor for the fact that some things are easier to understand using mathematics and other things are easy to understand using native anthropomorphic intuitions, as well as a recognition that religion promotes psychic health and strengthens community ties. In the *Ainai-mirta*, Truth and Beauty are literal anthropomorphic deities (the god Elith and the goddess Ainai) who are worshiped through prayer and sacrifice and invoked for strength in times of need.

It is considered somewhere between bad form and heresy to mix up these perspectives or try to apply one to the other. To claim that what one *wishes* were true actually *is* true is a heresy, an attempt to subordinate the god Elith to the goddess Ainai. But to claim that what *is* true is beautiful or acceptable or just when it isn't is equally heretical and also an insult to a goddess. Relativism, the idea that there *is* no Truth and *everything* should be viewed through *Ainai-mirta*, is an especially despicable heresy.

The level at which Beauty and Truth join together isn't some superficial philosophical muddle, it's the entire human project of art and science and trying to be happy and virtuous. *Sumurhe* is an attempt to give people a religious and mythological structure within which to view and appreciate that project.

It's also an attempt to justify the institutions of the Raikothin state. The Angel of Evidence is an attempt to gain a direct connection to Truth, the Angel of Preference is an attempt to gain a direct connection to Beauty, and the Angel of Salience and the Archangel are attempts to integrate the information gained by both into a teleological improvement of the world. Thus also the three great religious institutions of Raikoth: the Priests of Truth, experts at understanding the external world; the Priests of Beauty, experts at understanding the human psyche; and the Priests of Joy, experts at figuring out how to make the two correspond.

Below this rarefied philosophical level sits a rich mythology, some of which dates back all the way to the Paleo-Eskimo substrate of the Raikothin population. This includes a Creation story, various archetypal characters and legendary heroes, and morality tales about the great leaders of the past. Believing any of these to be “true” in the sense of *Elith-mirta* would be a heresy, but they are believed to be *Ainai-mirta*, true-with-respect-to-Beauty, and repeated and taught in that sense. Even superstitions are tolerated and encouraged in this sense in order to make life more interesting: various standing stones that supposedly bring good luck, stories of places you can go to dream of those you have lost, stories about how when someone dies their good deeds incubate in the ground before flying out as butterflies to alight on those they loved during life, and the continued belief in the holy volcano as protector of the state (which it fulfills *Elith-mirta* as the source of cheap geothermal power).

Index: Posts on Raikoth

The following are my posts on this blog about my constructed society of Raikoth:

1. Laws, Language, Society
2. Corruption, Priesthood
3. Cities, Land
4. Symbolic Beads
5. Economics, Relationships
6. History, Religion
7. Science

Related: Five Thousand Years In An Alternate Universe

Overheard In The Box, Part II

Mike: I think the last time I had this discussion, I decided to ground morality with waffles.

Alicorn: What?

Mike: I was in an IHOP with my friend from college, and we were talking about meta-ethics, and we couldn’t reach agreement, and then as we were leaving we saw a big sign saying “YOU DON’T KNOW WHAT GOOD IS UNTIL YOU’VE TRIED OUR WAFFLES”. So we thought, yeah, okay, that explains that.

Newtonian Ethics

We often refer to morality as being a force; for example, some charity is “a force for good” or some argument “has great moral force”. But which force is it?

Consider the possibility that it is gravity. In statements like “Sentencing guidelines should take into account the gravity of the offense”, the words “gravity” and “immorality” are used interchangeably. Gravitational language informs our moral discourse in other ways too: immoral people are described as “fallen”, sin is a “weight” upon the soul, and we worry about society undergoing moral “collapse”. So the argument from common usage (is best argument! is never wrong!) makes a strong case for an unexpected identity between morality and gravity similar to that between (for example) electricity and magnetism.

We can confirm this to the case by investigating inverse square laws. If morality is indeed an unusual form of gravitation, it will vary with the square of the distance between two objects.

Imagine a village of a hundred people somewhere in the Congo. Ninety-nine of these people are malnourished, half-dead of poverty and starvation, oozing from a hundred infected sores easily attributable to the lack of soap and clean water. One of those people is well-off, living in a lovely two-story house with three cars, two laptops, and a wide-screen plasma TV. He refuses to give any money whatsoever to his ninety-nine neighbors, claiming that they’re not his problem. At a distance of a ten meters – the distance of his house to the nearest of their hovels – this is monstrous and abominable.

Now imagine that same hundredth person living in New York City, some ten thousand kilometers away. It is no longer monstrous and abominable that he does not help the ninety-nine villagers left in the Congo. Indeed, it is entirely normal; any New Yorker who spared too much thought for the Congo would be thought a bit strange, a bit with-their-head-in-the-clouds, maybe told to stop worrying about nameless Congolese and to start caring more about their friends and family.

This is, of course, completely rational. New York City, at ten thousand kilometers, is one million times further away from the suffering villagers as the original well-off man's ten meters. Since moral force decreases with the square of the distance, the moral force of the Congolese on the New Yorker is diminished by a factor of one million squared – that is, one trillion.

At that distance, all one billion Africans matter only 1/1000th as much as would a person at zero distance. There is, in fact, a person at zero distance from the average New Yorker – that New Yorker herself. So we find that our theory predicts that our obligations to the Congo are only one tenth of one percent as important as our obligations to ourselves.

We can confirm this experimentally. This article from 2005 lists private US overseas charitable contributions at $10.7 billion a year. The 2000 US Census gave a population of 281,421,906, meaning that the average American gave $38.02 in overseas charity. This is 0.107% of the average 2005 per capita income of $35,242, compared to a predicted .0100; that is, a margin of error of only about twenty four cents.

(This is why I love physics. You'd never get results that match up to predictions that precisely in the so-called “social sciences”.)

This methodology can be used to answer a seemingly very different problem that many of us face every day: just how far away from a beggar do you need to walk before you don’t have to feel bad about not giving her money?

Suppose the marginal value of an extra dollar to a beggar is ten times its value to a well-off person such as yourself. We start with the money in your pocket, about a meter away from your brain. If you pass right by the beggar then the money may be a meter away from the beggar as well. Distance to both people is equal, so here the moral force exerted by the beggar is ten times stronger than your own moral force: you are clearly obligated to give her the money.

As you double your distance from the beggar to two meters, the moral force of her need decreases by a factor of four; however, she still has a 2.5x greater claim to the money than you do. Even three meters is not sufficient; her claim will be 1.1x as strong as your own.

However, four meters ought to do it. At this distance, the importance of the beggar’s poverty has decreased by a factor of sixteen, while your own moral force has stayed constant. It’s now 1.6x better for you to keep the money for yourself – a comfortable margin of safety.

There has been some discussion on whether it is acceptable to just hang to the far outside of the sidewalk in order to avoid a beggar, or whether this is unethical and it necessary to cross to the entire opposite side of the street. We now have the tools necessary to solve this problem. If you are on a commercial throughway, downtown residential, or other sidewalk listed on this table as having a minimum width of 4m or greater, it is borderline acceptable (ignoring air resistance) simply to move to the other side of the walkway. However, on the smaller neighborhood residential sidewalks, industrial sidewalks and alleyways – not to mention anywhere the beggar is in the middle of the walkway – it is unfortunately necessary to cross all the way to the other side of the street.

Once again, the results of even a back-of-the-envelope calculation like this one mesh admirably with most people’s native intuitions. Just as even a young child who throws a ball will have a “gut feeling” about how long it will stay up in the air, so even people unaware that morality is a variant of gravitation can correctly apply these same “gut feelings” to moral dilemmas.

In summary, morality is a form of gravitation, albeit an unusual one. Calculations performed based on inverse square law assumptions correctly predict most people’s moral actions. Indeed, the majority of human moral behavior make no sense *except* under these assumptions, and without them our everyday moral reasoning would be ridiculous indeed.

Against Bravery Debates

There’s a tradition on Reddit that when somebody repeats some cliche in a tone that makes it sound like she believes she is bringing some brilliant and heretical insight – like “I know I’m going to get downvoted for this, but believe we should have *less* government waste!” – people respond “SO BRAVE” in the comments. That’s what I mean by bravery debates. Discussions over who is bravely holding a nonconformist position in the face of persecution, and who is a coward defending the popular status quo and trying to silence dissenters.

These are *frickin’ toxic*. I don’t have a great explanation for why. It could be a status thing – saying that you’re the original thinker who has cast off the Matrix of omnipresent conformity and your opponent is a sheeple (sherson?) too fearful to realize your insight. Or it could be that, as the saying goes, “everyone is fighting a hard battle”, and telling someone else they’ve got it easy compared to you is just about the most demeaning thing you can do, especially when you’re wrong.

But the possible explanations aren’t the point. The point is that, empirically, starting a bravery debate is the quickest way to make sure that a conversation becomes horrible and infuriating. I’m generalizing from my own experience here, but one of the least pleasant philosophical experiences is thinking you’re bravely defending an unpopular but correct position, facing the constant persecution and prejudice from your more numerous and extremely smug opponents day in and day out without being worn-down … only to have one of your opponents offhandedly refer to how brave they are for resisting the monolithic machine that you and the rest of the unfairly-biased-toward-you culture have set up against them. You just want to scream NO YOU'RE WRONG SEFSEFILASDJO:IALJAOI:JA:O>ILFJASL:KFJ

A lot of common political terms pretty much encode bravery debates. "Political correctness", "mainstream media", "liberal media", "corporate media", "rape culture", "Big Government" or "Big Business" or "Big Anything", "patriarchy", "the climate establishment", or "the anything-anything complex". By not-at-all-a-coincidence, these also happen to be some of the terms most likely to be inflammatory and get people angry. Has there *ever* been an argument that continued being civil or productive after "political correctness" was mentioned?

The persistence of bravery debates is actually kind of weird. Shouldn't it be really really easy to figure out who's being oppressed by whom? The Spanish Inquisition had many faults, but whining about being unfairly persecuted by heretics was, as far as I know, not one of them. Can two opposing positions really be absolutely certain they are under siege?

This question immediately reminded me of my recent observation about Christians and Muslims in the media. Whenever the media says something negative about Christians, comments and blogs and forums immediately fill up with claims that the media loves picking on Christians and that no one would ever publish a similar story about Muslims for fear of being "offensive" (eg 1, 2, 3, 4, 5). And whenever the media says something negative about Muslims, comments and blogs and forums immediately fill up with claims that the media is Islamophobic and attacks Muslims any chance it gets and they would never dare pick on a large powerful group like Christians in such a way (eg 1, 2, 3, 4, 5).

So for example, Aziz Mubaraki writes:

> There are numerous cases to judge whether there is bias against Muslims in the media, but in recent times look no further than the press coverage regarding the terrorist attack that took place in Norway not very long ago. Impartial population waited impatiently to read this act being explicitly described as a “terrorist attack” or an “act of terrorism” by the mainstream media. But never once the “Christian” label was used despite the fact that Mr. Breivik was a self-described devout Christian. Therefore the important question is: Why is it when the person responsible for the terrorist act happens to be Muslim all of a sudden the religion becomes the focus instead?

Yet israpundit.com writes:

> Big media has no qualms about boldly and repeatedly labeling the Norweigan shooter as a “Christian”, even describing him as a Christian Zionist, despite no evidence that he was any kind of devout Christian whatsoever. Yet till this day the same vile liberal media will not refer to the Fort Hood jihadist as muslim or emphasize the Islamic motivation behind the shooting. Neither do government reports on the jihad attack.

So can we agree that this phenomenon of two opposing groups being equally sure they are bravely pointing out the world’s bias in favor of the other is, in fact, a thing?

Because once we acknowledge it, it’s not really hard to explain.

Psychologists have known about the hostile media effect for thirty years, ever since a 1982 study where they got pro-

Israeli and pro-Palestinian students to watch a documentary and found that:

> On a number of objective measures, both sides found that these identical news clips were slanted in favor of the other side. Pro-Israeli students reported seeing more anti-Israel references and fewer favorable references to Israel in the news report and pro-Palestinian students reported seeing more anti-Palestinian references, and so on. Both sides said a neutral observer would have a more negative view of their side from viewing the clips, and that the media would have excused the other side where it blamed their side.

Note that this was not at all subtle. The pro-Palestinians claimed that favorable references to Israel outnumbered unfavorable references almost 2:1, but the pro-Israelis complained that unfavorable references outnumbered favorable references at a greater than 3:1 ratio ($p < .001$). Transforming a different measure mentioned earlier in the paper to a scale of 1 to 10, where 1 is completely pro-Palestine and 10 is completely pro-Israel, the average pro-Israeli rated it a 3.2, and the average pro-Palestinian rated it a 7.4. These numbers were even higher in people who claimed to know a lot about the conflict. So even when exposed to genuinely neutral information, people tend to believe the deck is stacked against them. But people aren't exposed to genuinely neutral information. In a country of 300 million people, *every single day* there is going to be an example of something hideously biased against *every single group*, and proponents of those groups have formed effective machines to publicize the most outrageous examples in order to “confirm” their claims of bravery. I had an interesting discussion on Rebecca Hamilton’s blog about the Stomp Jesus incident. You probably never heard of this, but in the conservative Christian community it was a *huge* *deal*; Google gives 20,500 results for the phrase "stomp Jesus" in quotation marks, including up-to-date coverage from a bunch of big conservative blogs, news outlets, and forums. I guarantee that the readers of those blogs and forums are *constantly* fed salient examples of conservatives being oppressed and persecuted. And I don't mean "can't put up ten commandments in school", I mean armed gay rights activist breaks into Family Research Council headquarters and starts shooting people for opposing homosexuality. Imagine you hear a story in this genre almost every time you open your RSS feed.

(And now consider all the stories *you* hear every day about violence and harassment against *your* people in *your* RSS feed.)

And if there aren't enough shooters, someone is saying something despicable on Twitter pretty much every minute. The genre of "we know the world is against us because of five cherry-picked quotes from Twitter" is alive, well, and shaping people's perceptions. Here's an atheist blog trawling Twitter for horrible comments blaming atheists for terrorism, and here's an article on the tweets Brad Pitt's mother got for writing an editorial supporting Romney (including such gems as "Brad Pitt's mom wrote an anti-gay pro-Romney editorial. Kill the b—-.")

Then we get into more subtle forms of selection bias. Looking at the articles above, I am totally willing to believe newspapers are more likely to blaspheme Jesus than Mohammed, and also that newspapers are more likely to call a Muslim criminal a "terrorist" than they would a Christian criminal. Depending on your side, you can focus on one or the other of those statements and use it to prove the broader statement that "the media is biased against Christians/Muslims in favor of Muslims/Christians". Or you can focus on one part of society in particular being against you – for leftists, the corporations; for rightists, the universities – and if you exaggerate their power and use them as a proxy for society then you can say *society* is against you. Or as a last resort you can focus on only one side of the divide between social and structural power.

So it's far from a mystery how bravery debates can be so common or persistent. Or why everyone is so sure they're on the brave side. But the interesting thing is that they actually *work*.

I call your attention to two studies by Joseph Vandello et al. In the first, experimenters once again took the Israeli-Palestinian conflict but ran the experiment in the other direction. Here they presented maps that showed Palestine as the underdog (by displaying a map emphasizing a tiny Palestine surrounded by much larger Israel) or Israel as the underdog (by displaying a map emphasizing tiny Israel surrounded by a much larger Arab world including Palestine). In the "Palestinians as underdogs" condition, 55% of subjects said they supported Palestine. In the "Israelis as underdogs" condition, 75% said they supported Israel. And in the second, experimenters found subjects rated people who had been unfairly disadvantaged during a job interview as more attractive and more desirable romantic partners than people who had not been.

Baaaaasically if you get yourself perceived as the brave long-suffering underdog, people will support your cause and, as an added bonus, want to have sex with you.

And I dislike this, because bravery debates tend to be so fun and addictive that they drown out everything more substantive. Sometimes they can be acceptable stand-ins for actually having an opinion at all. I constantly get far-right blogs linking to my summary of Reactionary thought, and I hope I’m not being too unfair when I detect an occasional element of “Oh, so *that’s* what our positions are!”. There seem to be a *whole lot* of Reactionaries out there who are much less certain of what they believe than that they are very brave and nonconformist for believing it.

I was too quick to start bravery debates at my old blog and am trying to cut down on them. I would also recommend that other people cut down on them. I think they probably fall into the large category of things that make people who already agree with you fist-pump and shout “Yeah! We *are* awesome rebels!” while alienating everyone who doesn’t hold your position.

But what if you *are* being really brave by holding a dangerous and unpopular position? Shouldn’t you get credit for that?

I guess. I propose that if you write something and, for even just a second, you think of not publishing it, because of the risk to your reputation, or your livelihood, or your family, or even your life – then go ahead and call yourself brave, and I will try to reassure you and tell you everything is going to be all right.

If you think “Not publish this? But then how would everyone know how brave I’m being? I’m going to plaster my name all over this thing so everyone knows exactly where to send the bravery-related kudos!” … then stick to the damn object-level issues.

Can You Condition Yourself?

A friend recently told me about a self-help tactic that has become popular in the circles I move in: the idea of applying behaviorism to yourself (sometimes called “training your inner pigeon”). The idea is you give yourself rewards when you do things you want to do more of, and your brain works its magic and reinforces the activity.

When I first heard about this, my thought was “No way that is ever going to work”. I have always been under the impression that conditioning is kind of like tickling. You can’t tickle yourself. You’d be *expecting* it.

Let’s start by distinguishing a couple of possibilities:

1) This process doesn’t work at all

2) This process works by making you want the reward. Suppose you promise yourself a candy bar each time you do homework. You are hungry and want the candy bar, but you would feel bad if you ate it without doing homework. Therefore, you grudgingly do homework to get at the candy bar.

3) This process works by changing your urges and desires. After eating a candy bar each time you do homework, your brain associates homework with a nice, delicious-feeling, and you enjoy doing homework more from now on.

Let’s start with 3, the most encouraging possibility. This gains a little support from the Little Albert experiment. Here, a baby who had no particular fear of rats was exposed several times to rats plus loud, terrifying noises. Eventually the baby came to fear rats, even without the noise, presumably because the fear of the noise had generalized onto the rat through association. It's easy to see how this could mean something like the happiness of candy-bar-eating generalizing to homework. Nevertheless, I believe this argument proves too much.

Every evening, I sit down at the table, get a plate and some silverware, and eat dinner. It's usually something I really like, and it usually includes dessert, which I like even more. If eating good food isn't rewarding, I don't know what is, and sure enough I rarely skip dinnertime.

However, if for some reason I don't have dinner – maybe I've promised my friends I'll go out with a late dinner for them and so I can't stuff myself first – I do not feel the *slightest* urge to sit down at the dinner table with a plate and sort of move my silverware around in the air making little eating motions, and when I tried it (empiricism!) I did not find it at all pleasant.

Take a second to think about how weird that is (the result, not me trying the experiment). Sitting at the table and moving my silverware, in conditions exactly like these, has been quickly associated with reward every single time I've done it in the past, for decades, ever since I learned to feed myself. But I don't feel even a *little* bit of urge to do this. None at *all*. You may generate additional examples at your leisure, but the point is that just being consistently associated with a positive reinforcer in a low time-delay way does not make a neutral activity (let alone an actively unpleasant activity) become desirable.

What happened with Little Albert, then? First of all, he was classical conditioning and not operant conditioning. Second of all, Albert had no understanding or control over what was going on. Each time he heard the noise, he was very surprised – he was receiving a new fact from the Universe. But it wasn't information he understood; he had no idea what the connection between the rat and the noise was and whether it would recur. He just knew that there was some mysterious rat -> noise connection.

Compare this to me eating dinner. The connection between sitting down and eating dinner is not at all a new fact fed me by the Universe; it's something I plan myself. And it is not mysterious whether any given sitting and silverware-waving will reward me; I know it will reward me if and only if I am planning to eat dinner. Therefore the brain does not think of silverware-waving as an activity that might, who knows, lead to reward in the future.

(one might object that my inner pigeon – or lizard brain, to mix animal metaphors – doesn't share my complex explicit knowledge of the reward structure of dinner-eating. But the little I know of the brain's reinforcement mechanism suggests that reinforcement learning is based on *surprise* – technically the difference between predicted and observed values of some complicated Bayesian equation encoded in dopaminergic neurons or something – and that this system is actually quite good at predicting expected reward from an action, within certain limits)

So (3), the hypothesis that the reward will cause me to start enjoying homework, seems wrong. What about (2) – "I don't like homework much, but at least I get some candy out of it"?

Here there's a ceiling on how much the candy can reinforce your homework-doing behavior, and that ceiling is how much you like candy.

Suppose you have a big box of candy in the fridge. If you haven’t eaten it all already, that suggests your desire for candy isn’t even enough to reinforce *the action of going to the fridge, getting a candy bar, and eating it*, let alone the much more complicated task of doing homework. Yes, maybe there are good reasons why you don’t eat the candy – for example, you’re afraid of getting fat. But these issues don’t go away when you use the candy as a reward for homework completion. However little you want the candy bar you were barely even willing to take out of the fridge, that’s how much it’s motivating your homework.

Maybe you say “I will allow myself exactly one candy bar a day, but only if I finish my homework”. Even if you can stick to this rule, here the candy bar becomes an *extrinsic reward* motivating the homework. We all know what happens with extrinsic rewards – overjustification effect! You gradually start interpreting the task at hand as an annoying impediment to getting the reward, lose your intrinsic motivation, and as soon as the reward is removed, you’re even less willing to do the task than before.

So both (2) and (3) are pretty unlikely. That leaves us with (1) – don’t even bother.

Luckily, my friend helpfully clarified that this wasn’t what her class taught at all (I think maybe they originally tried this, but considerations like the ones I mentioned convinced them to change?). Their new policy is that you should reinforce yourself with a “victory gesture” – for example, pumping your fist and shouting “YEAH!” and visualizing an image corresponding to your success and trying to feel really good about yourself.

So for example, as soon as you sit down to start your homework, you make the victory gesture and imagine yourself graduating *summa cum laude* from school, and then you feel really good and have reinforced the behavior of sitting down to do your homework. And maybe you do it again when you finish, because peak end rule.

She claims a few benefits of this method. First, it's *very fast*, so you can reinforce things right as they happen instead of with time delay which gives your brain enough time to lose the connection. Second, it's intrinsic, so it's not going to sap your natural motivation the same way the candy bar might.

I understand the claim that rewards delivered *very immediately* after a stimulus can work better for conditioning – I was referred to a couple of papers proving this, though I don't remember them. But I notice I am confused. When we have good examples of *real* conditioning, immediate reward isn't especially important. For example, people often use the language of behaviorism to talk about addiction, say alcoholism. But the chemical rewards of getting drunk don't manifest until a little while after you've had your first beer – certainly not within a split second – and certainly alcoholism can reinforce even longer term behaviors, like leaving home and going to the bar. Pornography is another good example of effective behaviorism, but going to a porn site gives only delayed rewards – first you have to find a video you like, then you have to wait for it to buffer, then you have to sit through the boring part where the nice lady and the plumber are discussing the best ways to fix her faulty pipes, and so on. It seems that when we have a real effect that definitely works, immediacy is *not* required (indeed, if it were humans would have a lot of trouble learning anything but the most basic reflexes).

But okay. Ignore that. It would really really really really bad mind design to allow your own consciously generate-able emotions to feed back into the reinforcement mechanism.

Start with one obvious point. I said the candy bar couldn't be much of a reinforcer if you otherwise left it in the jar without eating it. The same seems broadly true of a victory gesture. I don't feel the *slightest* urge to perform a victory gesture, and having tried it empirically I don't feel the slightest urge to repeat it. This bodes poorly for its ability to be a strong reinforcer.

And over several billion years of evolution, the brain has every incentive to get rid of that behavior if indeed it was ever possible. Imagine a world in which our own thoughts and feelings can be strongly reinforcing. You're a caveman, encountering a saber-toothed tiger. You have two choices. You can either feel fear, which is an unpleasant emotion. Or you can feel happiness, which is a pleasant emotion. First you try feeling fear, but that's unpleasant! You don't like fear! The feeling of fear is negatively reinforced and your brain learns to stop feeling it. Then you try happiness! You like happiness! The decision to feel happiness is positively reinforced. Yes, you decide, saber-toothed tigers are wonderful things and you are overjoyed there is one in front of you getting into a pouncing position and licking its lips and…well, this caveman isn't going to live very long.

From the little I know about the reward system, it seems to operate on a basis of predicting pleasure level, then upregulating actions that result in world-states that seem more pleasurable than predicted and downregulating actions that result in world-states that seem less pleasurable than predicted. I don't think you can prevent the "I'm going to do my victory gesture!" part of you and the "I'm going to predict my pleasure at time t+1" part of you from talking to each other, I don't think internal pleasure is as reinforcing as external world-state results, and I don't think the pleasure of making a victory gesture is strong enough to do much anyway.

…there were a lot of "I thinks" in that paragraph. Do we have any evidence here?

The literature on this is hiding under the obscure term "self-consequation", and unfortunately it is all from Scientific Prehistory, ie the 1970s and 1980s before journal articles were uploaded to the Internet. I am able to find this full study, which does pretty much exactly the experiment listed at the beginning of this post – feed people candy in return for studying – and finds that it helps only if other people are there keeping them honest. But I am also able to find this abstract, which appears to be from a study showing the opposite – some kind of benefit – but is totally unavailable on the Internet. Both studies seem to refer to a long literature supporting their result and (sigh) neither seems aware of the other's existence. However, I am more skeptical of the second, both because I can't see it and because I worry that experimental protocols aren't *real* self-reinforcement. That is, if an experimenter gives you *their* bag of candy and tells you to reinforce yourself by eating some when you do something good, that's still different from using your own bag of candy and coming up with the idea on your own, even if the experimenter is out of the room when you're working.

I will still try the technique, because it seems low cost and potentially high value. Really high value, actually. So high value that I would have expected the first person to get it right to take over the world. This is turning into another argument against it, isn't it?

But yeah, as I was saying, I still intend to try the technique, even though it won't be a very well-controlled experiment.

And I’m glad I heard the idea for reminding me how little I know about behaviorism.

Apart From Better Sanitation And Medicine And Education And Irrigation And Public Health And Roads And Public Order, What Has Modernity Done For Us?

. . .

Brought peace.

As you may have noticed, instead of another GIGANTIC WALL OF TEXT I am trying to write my rebuttal to Reactionary philosophy in the form of several smaller posts that I can then link together in a sequence index. This particular post addresses Reactionary claims that modern society causes international instability, leading to increased war (or increased “total war”) and the resulting mayhem.

This claim I received mostly from blog posts I can’t find right now and from discussions with Michael Anissimov. It goes that when states are fully sovereign, self-interested, and run by noble classes – as they were long ago – their wars are rare, as short as possible, and mostly fought in a civilized way.

But when states are subject to a larger international order (like the UN or “international law”), interested in ideological concerns, and governed by a host of factions competing for democratic power – as they are today – wars are more common, bungled into increased length and fatality, and turn into “total war” where anything goes and civilians are considered valid targets.

Michael specifically mentioned the Congress of Vienna as an example of the old order, pointing out that a bunch of aristocrats met up, divided Europe among them, and there was peace for decades afterwards. He compared this to the inelegance of modern “police actions” and “foreign interventions”, pointing out how World Wars I and II, at the beginning of the modern era, were unmatched in their deadliness and brutality.

Luckily, these questions about war and the stability of different models of international relations can be investigated empirically. Are wars worse today, or were we worse during the old aristocratic era? By what standards?

Let’s ask the media! War Is Going Out Of Style, says the *New York Times*. War And Violence On The Decline In Modern Times, trumpets NPR. Josh Goldstein says we are “winning the war on war”, Steven Pinker proclaims the victory of the better angels of our nature, and John Mueller even more triumphantly posits that War Has Almost Ceased To Exist

The statistics bear them out. The BBC notes:

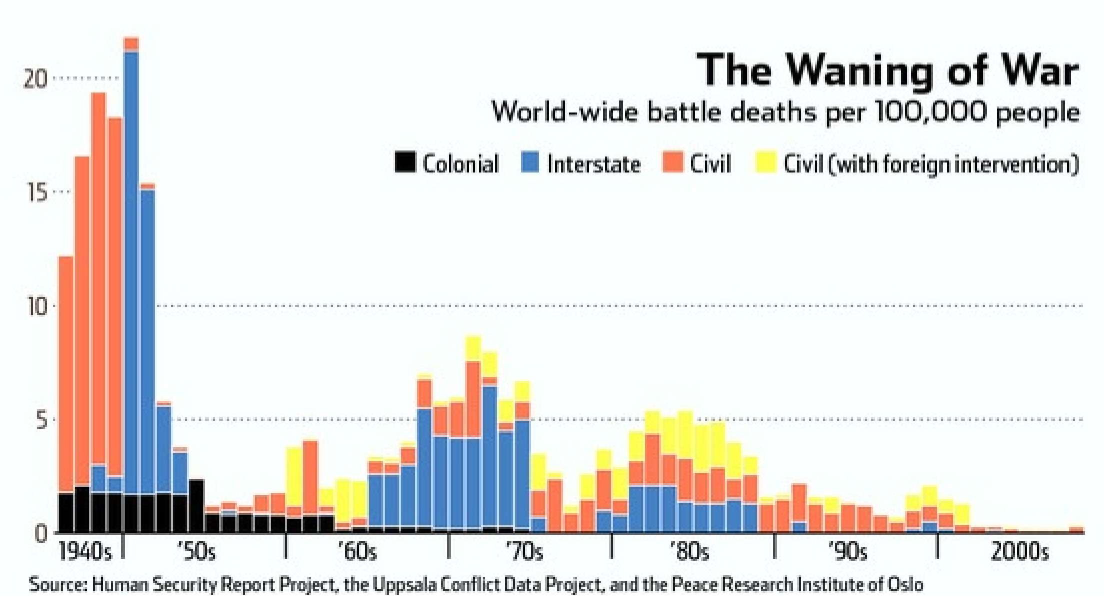
> The Human Security Report found a decline in every form of political violence except terrorism since 1992. “A lot of the data we have in this report is extraordinary,” its director, former UN official Andrew Mack, said.
>
> It found the number of armed conflicts had fallen by more than 40% in the past 13 years, while the number of very deadly wars had fallen by 80%.
>
> The study says many common beliefs about contemporary conflict are “myths” – such as that 90% of those killed in current wars are civilians, or that women are disproportionately victimised. The report credits

> intervention by the United Nations, plus the end of colonialism and the Cold War, as the main reasons for the decline in conflict.

The trend is older than just this decade. According to Goldstein:

> In fact, the last decade has seen fewer war deaths than any decade in the past 100 years, based on data compiled by researchers Bethany Lacina and Nils Petter Gleditsch of the Peace Research Institute Oslo. Worldwide, deaths caused directly by war-related violence in the new century have averaged about 55,000 per year, just over half of what they were in the 1990s (100,000 a year), a third of what they were during the Cold War (180,000 a year from 1950 to 1989), and a hundredth of what they were in World War II. If you factor in the growing global population, which has nearly quadrupled in the last century, the decrease is even sharper. Far from being an age of killer anarchy, the 20 years since the Cold War ended have been an era of rapid progress toward peace.

And Steven Pinker shows the following graph:


The Waning of War

World-wide battle deaths per 100,000 people

Source: Human Security Report Project, the Uppsala Conflict Data Project, and the Peace Research Institute of Oslo

So there's more than enough data to show the world has been getting more peaceful over the past seventy years. The most plausible Reactionary response would be that this is too small a time horizon: that the horrors of progressivism should be viewed over a timescale of centuries.

First of all, this shouldn't be true. A staple of Reactionary thought is that the world *has* become notably more progressive since World War II, and a hyper-willingness to attribute anything that's declined since that period to the progressive world-view. What's good for the goose is good for the gander. Second, it is very suspicious to say that the part of the data you don't have good statistics for, and *only* that part of the data, proves your point.

But in order to address this objection more fully, I tried to get fuzzy ballpark area data on the deadliness of wars in past centuries. My methodology was to comb Wikipedia's list of wars by death toll, take all the ones with casualties of one million or greater, and organize them by era. The eras I used were 21st Century So Far, 1950-2000, 1900-1950, 1850-1900, 1800 -1850, 1700-1800, 1600-1700, 1500-1600, 1000-1500, 1-1000, and 500 BC – 1. Where casualties were given as a range, I took the center of that range, except in the Taiping Rebellion where I believe the top of Wikipedia’s range is crazy high and so I took nearer the bottom; where conflicts spanned more than one era, I placed them in the one containing the majority of the conflict.

I added up total war casualties for each era, then scaled them by population using 2005 as the standard – that is, deaths were multiplied so that the new number was the same percent of the 2005 population as the original was of its own era’s population. Then I divided by the length of the era to give average deaths per century during that era.

The 1900 – 1950 era indeed came on top, with 626 million projected deaths per century per 2005 population. Second place was 1600 – 1700, with 442 million. Other violent periods of note were 1850 – 1900 (326M), 1000 – 1500 (230M), and 1800-1850 (106M). There was no obvious trend related to time.

However one trend worthy of note is that the 21st Century So Far and the period 1950-2000 were *by far* the two most peaceful eras of any in the study (both at about 28M).

So the most progressive periods in history are also the most peaceful. And the Reactionaries’ pet period, the 1600s when the Stuarts ruled England and the Hapsburgs were still mighty, was the deadliest age of history outside a World War. I tested what would happen if I limited the domain to Europe, and the results are much the same (with the exception of 1850 – 1900 becoming much more peaceful).

This study is actually biased against me and in favor of the Reactionaries in two ways. First, I eliminated all wars with death counts less than a million, because otherwise it would have taken *forever*. But that disproportionately eliminates pre-modern wars, since they were fought among lower-population nations – a conflict today need only kill 1/7000th of the population to make my list, but one in 0 BC would have had to kill a full 1/100 or be dropped entirely.

Second, *technology*! Two days worth of airplanes dropping bombs on Dresden in the 1940s killed more people than several long and bloody medieval crusades. More modern death counts should probably be discounted to take into account the fact that we are just way better at killing each other when we want to, even though we want to much less often. Yes, the era of World Wars saw slightly greater deaths per population than the era of absolute monarchy in Europe. But the Allies were killing people with nuclear bombs, and the Hapsburgs were killing people with bayonets. The 17th century in particular, and the past in general, just *really really sucked*.

Some Reactionaries, intuiting this pattern, have tried to dismiss it by saying that, while progressive eras have few wars, their wars are much worse – the sort of "total war" that characterized World Wars I and II, and so rose to new levels of killing and barbarity.

But this article lists the worst conflicts of all time by percent of population killed. And you have to go to number six on the list *just to get to World War II*! World War I *isn't even on the list!* The Mongols did not kill 11% of the population of Earth in twenty-one years by not being aware you could harm civilians; the various mercenaries of the Thirty Years' War were no more innocent.

One last fact noticed in the process of going through Wikipedia's wars list: in any particular era, it is always the *least* progressive countries that are having the wars. Even the miniscule death count in the late 20th and early 21st centuries is limited almost entirely to authoritarian African countries and Islamic theocracies. In neither World War was the major conflict two democracies (by any reasonable definition) fighting one another, and at least in the latter totalitarian side deserves a disproportionate amount of blame. The bloodiest conflicts of the past few thousand years, even adjusting for population, have been in China, which is basically Reactionary Utopia with an authoritarian Emperor, a Mandate of Heaven, and strict racial homogeneity. There is a lot of debate over whether two democracies have ever gone to war (answer: it depends how true of a Scotsman you are) but this very fact should cue you in that war and democracy are not *positively* correlated (and most likely not even neutrally correlated).

So to sum up: as the world has become more progressive over the past seventy years, conflicts and deaths from conflict have dropped precipitously. Virtually every past era was much more violent than our own, and the biases of this study probably mean they were more violent even than our numbers indicated. Every single one of the five deadliest conflicts in human history occurred before the Enlightenment, and in any given era the more progressive countries both start and participate in fewer wars than the less progressive countries.

Very likely this is due partly or mostly to economic factors – the point that no two countries with McDonalds' ever go to war is a good one. But this does not negate the fact that our current political and social system is the one that economic factors decided to set up in order to achieve their economic goals.

The Wisdom of the Ancients

Were The Victorians Cleverer Than Us?, asks a new study by Woodley et al that has gotten name-dropped in places like The Daily Mail and The Huffington Post.

Meanwhile, Betteridge’s Law of Headlines continues to warn us that “Any headline which ends in a question mark can be answered by the word no.”

On first glance, the paper looks solid. It investigates simple reaction time, a measure which is known to be correlated with *g*, the mysterious general intelligence which is supposedly measured (to some degree) by IQ tests. People have been experimenting with simple reaction time for over a century now, so the paper asked the relatively simple question of whether it has changed over that century. They found that it had: it had gone up, signifying a decrease in general intelligence. Their explanation was dysgenics.

People have known for a long time that high-IQ people have fewer children than low-IQ people, so it might make sense genetically to believe that each generation becomes a little dumber. This pattern has stubbornly refused to appear: instead, every generation has had significantly higher IQ than the one before, an observation called the Flynn Effect. This has been attributed to various things, including better nutrition, child-rearing, and education.

What the authors of this paper do – and it’s pretty clever – is say that the Flynn Effect is an environmental increase in IQ which has hidden a simultaneous genetic decline in IQ. They try to prove it by saying environmental and genetic factors affect IQ in different ways, and that genetic factors are more likely to affect certain features like reaction time – a pattern which is called a Jensen Effect and which is on relatively solid ground. Because they find reaction time is declining, probably people are becoming genetically stupider and the only reason we can keep having a civilization at all is because our environment is getting better – which is too bad, since our environment may have stopped doing that.

All the theory here *sort of* checks out, except for the part where they say IQ changed 15 points in a hundred years, which is just a *little* bit faster than any responsible person expects evolution to progress. People critique the idea that Ashkenazi Jews could have shifted fifteen points in *nine hundred years* on the grounds that it's too fast. So let's take a closer look at their data.

Only two of their sixteen studies come from the Victorian Era: Galton 1889 (n = 3410) and Thompson 1895 (n = 49).

Francis Galton, a brilliant Victorian scientist who was a half-cousin of Darwin, is the source of 98.5% of our Victorian reaction time data – not to mention the concept of reaction time itself, several statistical tools including correlation and standard deviation, the use of the survey in data collecting, the term "eugenics", the entire science of meteorology, hearing tests, the first study on the power of prayer (he prayed over random fields to see if the crops there grew higher; they didn't), fingerprinting, the scientific investigation of synaesthesia, and a horrible warning about how not to do facial hair.

Galton's Data A Century Later, published in 1985, tells us a little about how he gained his ground-breaking reaction time statistics. He set up a laboratory in the Science Galleries of the South Kensington Museum. There he charged visitors to the museum three pence ($25 in modern currency after adjusting for inflation) to be measured by his instruments, a process he advertised as “for the use of those who desire to be accurately measured in many ways, either to obtain timely warning of remediable faults in development, or to learn their powers.” Over the course of nine years, he attracted about nine thousand curious individuals, three thousand of whose data managed to make it into the current meta-analysis.

His colleague in Victorian reaction-time measurement was Helen Thompson Woolley, an American psychologist who published a 1903 dissertation titled *The Mental Traits of Sex: An Experimental Investigation of the Normal Mind in Men and Women* (it was, apparently a simpler time). With an optimism bordering on the incredible, Wikipedia notes that “Before Woolley, research on sex differences was heavily influenced by conjecture and bias.”

Woolley writes of her sampling technique:

> “In making a series of tests for comparative purposes, the first prerequisite is to obtain material that is really comparable. It has been shown that the simple sensory processes vary with age and with social condition. No one would question that this statement is true for the intellectual processes also. In order to make a trustworthy investigation of the variations due to sex alone, therefore, it is essential to secure as material for experimentation, individuals of both sexes who are near the sae age, who have the same social status, and who have been subjected to like training and social surroundings. Probably the nearest approach among adults to the ideal requirement is afforded by the undergraduate students of a coeducational university. For most of the the obtaining of an education has been the one serious business of life.

> The individuals who furnished the basis for the present study were students of the University of Chicago. They were all juniors, seniors, or students in the first year of their graduate work. The subjects were obtained by requesting members of the classes in introductory psychology and ethics to serve.”

She found (a finding replicated by all later studies and now considered essentially proven) that women have slower reaction times than men (interestingly, this difference does *not* correlate with IQ) – but more relevant to the current meta-analysis, she found the same generally fast reaction times as Galton.

The modern studies, keeping with the zeitgeist of the modern age, are much less colorful. I only looked into the two largest: one Scottish, the other Australian. Here’s what the Scottish study says of its methodology:

> The study was originally located in the Central Clydeside Conurbation (Figure 3), a socially heterogeneous and predominantly urban region, including Glasgow City, which is known to have generally poor health. Two-stage stratified sampling was used to select subjects. For the regional sample, local government districts were stratified by unemployment and socio-economic group data from the 1981 Census and 52 postcode sectors were systematically selected from these with a probability proportionate to their population size. The same postcode sectors were chosen for all three cohorts. The sampling frame used for individuals was Strathclyde Regional Council’s 1986 Voluntary Population Survey—an enhanced electoral register that provides details of the age and sex of all household members.3 Individuals were selected from the 52 postcode sectors within each age cohort with a

> systematic selection with a prescribed sampling interval from a random start.

I was getting bored by the time I made it to the Australian study, but I managed to keep my attention on it long enough to note the following sentence:

> Persons selected at random from the Electoral Roll [of Canberra] were sent a letter informing them about the survey and saying that an interviewer would contact them soon to see if they wanted to participate.

Look around you. Just look around you. Have you worked out what we’re looking for yet?

That’s right. The answer is **selection bias**.

Back in the Victorian Age, science was done by aristocrats and gentlemen who drew their subjects from their own social groups. There were no poor people in either study, because getting poor people to participate in an experiment would require finding some poor people, who probably smelled terrible and lived in areas where there were no good restaurants.

In the Modern Age, everyone is excruciatingly Socially Aware, and studies go out of their way to look at Disadvantaged Disempowered Disprivileged Populations so their results can serve as Cutting Social Commentary.

Galton’s study population was visitors to a science museum in the posh part of London who were willing to pay him $25 to participate. Thompson’s population was University of Chicago philosophy students. The two modern studies are random selections double-checked to make sure they don’t undersample the poorest sections of the population.

So, uh, congratulations, authors of this paper! You have successfully proven that the average member of the population is dumber than wealthy science dilettantes *and* students at elite colleges! Go pat yourself on the back!

In case we need more rigor: according to The National Center for Education Statistics, about 2.3% of Americans went to college in 1900. In a perfect meritocracy maybe only the smartest people would go to college, but we're not a perfect meritocracy. Would it sound about fair to say that the people in college at the time were a sample of the 20% or so of the smartest Americans?

Because the IQ of someone at the 80th percentile is 113 – that is, exactly enough to explain the 14 point IQ "drop" that Woodley et al found.

This is a little harder to do with Galton's science museum visitors. The 1985 commentary on Galton's data tells us:

> As would be expected of a group of paying testees being measured in a museum, a sizable portion of Galton's sample consisted of professionals, semiprofessionals, and students. However, as may be discerned in Tables 10 and 11, all socioeconomic strata were represented.

Tables 10 and 11 turn out to be a gold mine – I worried the records of exactly who took the tests would be lost, but as you might expect of someone who basically invented statistics single-handedly and then beat Darwin in a debate about evolution as an encore, Galton was *very good* at keeping careful data.

This site tells me that about 3% of Victorians were "professionals" of one sort or another. But about 16% of Galton's non-student visitors identified as that group. These students themselves (Galton calls them “students and scholars”, I don’t know what the distinction is) made up 44% of the sample – because the data was limited to those 16+, I believe these were mostly college students – aka once again the top few percent of society. Unskilled laborers, who made up 75% of Victorian society, made up less than four percent of Galton’s sample!

So this discredits this meta-analysis way beyond any need for further discrediting, but since I can’t help beating a dead horse…

Let’s talk about race. We know that studies find white people usually have faster reaction times than black people – in fact, a lot of the voluminous and labyrinthine research on race and IQ hinges on this fact. We thankfully do *not* have to enter the minefield of trying to figure out the causes of this discrepancy (biological vs. environmental vs. social) – we can just take it as a brute fact.

What percent of Galton’s 1889 science museum visitors do you think were non-white? What percent of Thompson’s 1895 University of Chicago students? Approximately zero? Sad to say, non-white people were as likely to be exhibits in the science museums of the day as visitors, and according to no less a figure than W.E.B. DuBois in 1900 there were only 2600 living black Americans who had graduated college.

I looked them up some stats on the sample areas for the modern studies – 6% of Glasgow is non-white, and about 12% of Canberra. So aside from selection bias affecting intelligence which affects reaction time, we have selection bias affecting race which affects reaction time.

May I just say how annoyed I am that I have to remind reactionary eugenicist IQ researchers, *of all people*, to pay attention to race? YOU HAD ONE JOB!

Finally, there's significant IQ differences within populations of the same race and country simply due to migration effects. An analysis of IQs across Great Britain finds that the highest scores are in London (102) and the lowest in Scotland (97). Almost all this meta-analysis' Victorian data came from London (Galton's museum in Kensington) and the largest source of modern data (making up about half of the whole, and being unusually high in reaction time) came from Scotland (and Glasgow isn't even the nice part of Scotland). The 5 point London – Scotland difference explains over a third of the "difference between Victorians and moderns" found in this study.

So in conclusion, this study ignores race, ignores regional variations, but most importantly IGNORES THAT ALL ITS VICTORIAN STUDIES WERE SAMPLING FROM THE SMARTEST 20% OR SO OF THE POPULATION AND THEY GOT EXACTLY THE NUMBERS YOU WOULD EXPECT IF YOU DID THAT.

There is some *really excellent* IQ research out there that everyone should be reading, but this is not it. Please please *please* don't cite this study as evidence for dysgenics or the decline of civilization.

Going From California With An Aching In My Heart.

Alicorn once told the story of how, when she was younger, she used to think she disliked life. Then she realized she just disliked being a kid, and that after *that* problem was solved life was pretty good.

In much the same way, I used to think I disliked social interaction. I have since realized – and it blew my mind – that I only disliked social interaction *with people who aren't awesome*.

I am leaving California tomorrow for the Midwest, where I have a four-year residency in one of the local hospitals. In a life that has seen more than its share of leaving places to go to other places, I don't think I've ever been quite so reluctant to move on.

People in the Bay Area *get it*. You get a bunch of hippies throwing love into the pot, computer programmers adding brainpower, and entrepreneurs adding competence. Mix and stir and you get people who want to make the world better, know how to do it, and sometimes even get up off their armchair and *do*.

I've recently been thinking about two competing philosophies. The first, which seems to me very conservative, is that you maximize virtue and if you're sufficiently virtue happiness follows on its own. The second, which seems very hippie, is that you maximize happiness and if you're sufficiently happy then you naturally want to spread that happiness as far as possible. I have previously been sympathetic to the former view, but the Bay Area makes an impossible-to-ignore case for the latter. All the self-help and spirituality and wacky leftism seems founded in this base of “We’ve maxed out our own happiness meter just by living here! How can we help the rest of you?”

I was privileged to experience a bubble within a bubble, the meta-bubble being the Bay Area “rationalist community”, ie a bunch of people who met through Less Wrong, the Machine Intelligence Research Institute, and the Center for Applied Rationality. One of my thoughts in designing Raikoth and other constructed societies is that if you collect nice people beyond a certain concentration it has unexpected emergent effects based on everyone suddenly realizing they can trust everyone else and voluntarily abandoning some of their hang-ups and defense mechanisms. That is exactly what happened in the Bay Area sometime before I moved there.

I am still trying to work out *what went right* in the hopes of being able to bottle it and export it to other groups of people. There’s the old adage that there are three sides to every issue – your side, their side, and the truth – and the epistemic virtue that allows you to detach from your side in a philosophical debate is exactly the same virtue that allows you to see every side of a quarrel and work through it reasonably. I think a lot of the best self-help – cognitive-behavioral therapy, non-violent communication, that kind of thing – are basically trying to teach this skill that the Bay Area rationalist community has more of than any other group I’ve ever seen.

Beyond that, cynicism can be incredibly liberating. Things that no one would ever admit elsewhere – status motives, sex motives, weird cognitive biases and mental flinches – are all out in the open among Bay Area rationalists. People can say things like “I don’t know why, but that thing you do enrages me, can you please stop?” and instead of getting offended other people will just say “Sure”, because *no one* makes sense, and mindspace is deep and wide. Overt status contests are *really hard* to get into, because everyone knows exactly what a status contest is and it would be oddly *tacky*, the same way that saying “Oh, I like an obscure band, you’ve probably never heard of them” sounds tacky because everyone has absorbed it as a stereotypical example of something pretentious people do.

People seem to accept sex and romance (of all permutations and kinks) as a normal part of life, and can talk about it in a way that I would describe as “like reasonable adults” if reasonable adults showed any signs of being able to talk that way. For almost the first time, I was able to have close female friends without being confused and scared of getting classified as a “creep” or “Nice Guy (TM)”. People of any gender and sexual orientation can ask friends of any other gender or sexual orientation to cuddle, and be met with an enthusiastic “yes” or an immediately-accepted “no” but almost never an outraged “How dare you!”

I have actually used moral philosophy to settle problems, and it has *worked* – if only because people get much more interested in the arcana of the moral principles involved than on whatever the original point of contention was. I have seen libertarians, socialists, feminists, mens’ rights advocates, transhumanists, and pious Christians get together at the same table, debate their views, and *have a totally reasonable discussion that ends with everyone more enlightened and appreciative than when they arrived*. I’ve seen a bunch of guys live together with the one girl they are all dating and this situation produce *zero conflict*.

A long time ago, a discussion on the uses of technical rationality seemed to converge around it probably being more useful for communities than for individuals. I think the

Bay Area is a shining and undeniable proof of this, and I'm incredibly grateful to have gotten to live there as long as I did.

…but I should add that it's not *just* the rationality and that the people involved are, in fact, wonderful people, and would probably have been wonderful people no matter what memeplex they had been exposed to. I can hardly begin to thank all the great people I have interacted with and learned from over the past year. I probably shouldn't even try, as I'm sure the jet lag I'm in the process of losing a fight against, not to mention the limitations of finite time and energy, will make me forget people who I love dearly. But I can't help expressing gratitude to just a couple of the people I got the privilege of getting to know recently.

Mike, I told you before that I use you to ground morality. I have used "Mike Blume does X" in ethical debates as a knock-down argument that X is okay, and the argument always gets accepted. Alicorn says she uses you as a "happiness battery", and I notice I started doing the same after a few months living with you, and given your central position in so many social circles a better metaphor might be a giant happiness hydroelectric plant powering half of Northern California. The fact that flowers do not spring up everywhere you walk only demonstrates that flowers are wrong

Alicorn, you claimed when we first met that you were "orthogonal to status", which at the time I thought was ridiculous. I still would be reluctant to use exactly that description. But if you were to say you had solved life, you are practically the only person from whom I would take the claim remotely seriously. You have some mysterious ability to factor situations into simple parts, figure out what you want from them, and then just *do that* without any drama or self-questioning or hullabaloo. If I had that skill to a tenth of the degree you do, I would just dismiss the rest of the human race as so hopelessly confused as to be not worth your time, but you have somehow managed to stay really nice regardless. You also cook well and have amazing hair.

Luke, I remember getting some social skills advice from a friend in New York. He told me that if I met someone who was obviously amazing along one axis, I should compliment them along other axes, because they will more fully appreciate hearing compliments they didn’t already know. I think his exact words were “If there’s a really pretty girl, tell her she’s smart, and if there’s a really smart girl, tell her she’s pretty”. I remember asking “What if there’s a girl who’s both smart and pretty?” and being told to tell her that she was kind, or had a nice voice, or something. After continuing this several more iterations, we joked that if we ever found someone amazing along every axis, we would be reduced to horrible awkward compliments like “Hi! I’m Scott! You’re really tall! Mind if I sit down?” Well, you not only seem to exemplify this nightmare scenario, but *I can’t even tell you that you’re really tall.* YOU ALREADY KNOW!

Leah, there’s an old Jewish saying that “everyone wants to serve God, but only in an advisory position”. Probably meant as satire, but if by chance the job actually exists, I think it is a perfect match for you and I will happily write you a glowing letter of recommendation.

Anna, I am constantly amazed by the number of plans that contain or should contain a node saying “Get Anna Salamon’s advice, then do whatever she says”. When people raise irrelevant theoretical objections to thought experiments beginning “Imagine a perfect Bayesian reasoner claimed…”, I just switch to “Imagine Anna Salamon claimed…” and they shut up.

Kenzi, it was incredibly gratifying to be able to tell my parents “I’m dating someone who…” and then list off all your qualities. The only problem was that by the time I got to “…is amazing at pretty much every style of dance, and got accepted to a prestigious medical school but turned it down to pursue her dream job at a non-profit, and can identify all West Coast flora on sight…” I think they might have stopped believing me. But you’re not *just* a trophy. You’re also the person who can finish a hiking trip with me without giving up, breaking down, or feigning your own death to escape. That (among other things) earns you a permanent place in my heart.

Davis, I still have trouble believing you actually exist.

Vassar, before I came to Berkeley, someone warned me “Vassar is kind of crazy and it’s impossible to have a normal conversation with him”. As a result, I spent several months avoiding you. Then I finally got to meet you and I realized I had made a huge mistake. I mean, you *are* crazy, and it is *is* impossible to have a normal conversation with you. But normal conversation is incredibly over-rated compared to whatever the heck you call the thing that interaction with you involves. I regret that we didn’t get more of a chance to talk about stuff and I hope to solve that sometime in the future.

Elizabeth, this is you to a ‘T’. You are a good egg.

Zak, you are nice and helpful and friendly and never gave me any grief about my total failure to answer any of your emails. In case you were wondering, it was because I’m a bad person.

Carl, the last time you commented on my blog, I went around doing a little dance and singing “Carl Shulman reads my blog! This validates my existence!”

Shannon, I admire your fighting spirit and your ability to remain cheerful in the face of adversity and pretty much everywhen else. Thanks for your help explaining IFS to me and for your thoughts on psychiatry which I am going to try my best to learn from.

Kevin, I predict one day I will see your name and picture in a biochemistry textbook. I don’t know whether it will be as a bold pioneer or as a horrible warning. Either way, I’ll be like “Hey, I knew that guy!”

Julia, I am constantly surprised to see you in our community when you could so obviously fit into a much higher-status community where people have class and drink fancy wine and almost never wear shoes with individual toes. But I am delighted that you stick around with us. I remember seeing a video of your “straw Vulcan” speech and wondering who you were and how I had somehow missed knowing about you and how I could correct that. Every time we have talked you have given me really interesting things to think about. I am so glad you are one of us instead of wasting your life becoming Secretary of State or something.

Valentine, I am convinced that if an evil wizard ever tried to attack you with one of those artifacts that turns the target’s insecurities and negative emotions against them, he would end up looking baffled, kicking the artifact to see if it was broken, and eventually giving up in disgust. And then you would kick his ass. Compassionately. And I love the story behind your nickname.

Steven, your sense of humor needs to be declared either a wonder of the world or at least a UNESCO World Heritage Site. You are the only justification the existence of Twitter I will accept. And your explanation of your politics actually was quite helpful to me.

Will Newsome (are you a Berkelian these days? I'm just going to count you) even though you are sometimes irresistibly fun to tease (I still like telling the story of how after I read Breakdown of Will, three consecutive people to whom I mentioned the book title thought I was talking about you) I have a huge amount of respect for your originality as a thinker and for your Taking Ideas Seriously abilities. I continue to wish to subscribe to the newsletter you stubbornly refuse to write.

Paul and Katja, I didn't really get to see you much or really at all, but I was really impressed by your work organizing the Berkeley altruist community. Even if I cannot yet count you as friends I definitely count you as inspirations. Katja, you are the second of three people I will accuse in this post of having awesome hair.

Nisan, likewise with your community organization work and you actually putting together and holding all those meetups. I am still impressed you tried that Hermeneutics game. Empiricism!

Jonah, there is a passage I remember from Name of the Wind: *"Elodin proved a difficult man to find. When I visited Ledgers and Lists, I discovered he only taught one class: Unlikely Maths. However, this was less than helpful, as according to the ledger, the time of the class was 'now' and the location was 'everywhere.'"* I cannot think of a more elegant description of being in your social circle.

Will & Divia, if I were doing a meta-analysis of whether all those weird self-help things really make people better, you would be two really really big squares way on the right of the forest plot, and people would get angry and say that I should have dropped you as outliers.

Lindsey, I think I only think of you as a personification of the spirit of California because of that one California-themed party you put on, but the comparison stands.

Nick, of everything that happened that first time I visited Benton long ago, driving somewhere with you talking about ethics is the part I remember most. You are wise and calm and clear-thinking, and my attempts to visualize you getting angry at something fail about the same way my attempts to visualize a square circle do – this vague vision of a square with rounded corners followed by a “screw this, this is dumb” and giving up. You are the last of three people whom I will mention have awesome hair.

Louie, remember that time you came to the California-theme party as a bigshot high-powered studio executive? It took me several minutes of talking to you before I realized you were playing a role. You are so much larger than life at all times that I can pretty much believe anything of you, and so well-rounded you approximate a hypersphere. You’re another one of those people I can’t tell stories about because no one would believe me.

Qiaochu, I don’t know which I am more jealous of – your skill in and enthusiasm about math, or that you’re going to get to live with Mike and Alicorn next year. Whenever I am searching for math-related things on the Internet, I inevitably stumble across your name and things you have written with much greater regularity than the percent the world intellectual community you compose would suggest. There was a three minute window when you were talking to me when I sort of understood the idea of using probability to escape Lobian self-reference. You have a great name and an even better explanation of how to pronounce it, and you are *fun*.

Eliezer, I was thinking recently about the impact that ideas and people and communities traceable back directly to you have had on my life. I ended up giving you credit for most of my friends, all my relationships, one of my favorite books, the last place I lived, the job I've been working for the past year, quite a bit of my philosophy, and, via my work and prizes in MetaMed looking really good on my resume, quite possibly my career as well. If you ever succeed in your grand scheme to initiate a positive singularity and turn the world into a utopia, I'm going to be the one grumbling that *I* had my life radically improved by Eliezer Yudkowsky *before it was cool*.

The World Is Flat And Intensively Farmed

I like the Midwest. It is a superstimulus for something – call it “Americana”. There are trees everywhere, big puffy clouds, white picket fences, big red barns, and burger joints with single-syllable names. Also, I am pretty sure I spotted an amber wave of grain yesterday, although it was far away and could have just been reeds or something.

My father came with me as far as Detroit, where he stopped to help me search for a new car in the Motor City. He is wise in the ways of the world, though sometimes a little too practical. I remember back when he came to visit me in Japan. Come dinner time, he asked “Is there anywhere you can get good Italian food in this country?”, and when I protested that he might want to try some sushi, or noodles, or dumplings, or tempura, he just shook his head and started looking through the index of his guidebook. I think we had Italian six of the seven nights he was there.

Admittedly, when Japan does Italian, it goes all out.

But I will give him credit where credit is due; he eventually made it up to the poor Japanese. After he perused all the cars that one of the world’s greatest automotive centers had to offer, he made me a final well-informed recommendation and I am now the proud owner of a new Subaru.

Transhumanist Fables

Once upon a time there were three little pigs who went out into the world to build their houses. The first pig was very lazy and built his house out of straw. The second pig was a little harder-working and built his house out of sticks. The third pig was the hardest-working of all, and built his house out of bricks. Then came the Big Bad Wolf. When he saw the house of straw, he huffed and he puffed and he blew the house down, eating the first little pig. When he saw the house of sticks, he huffed and he puffed and he blew the house down, eating the second little pig. When he saw the house of bricks, he got out a bazooka and blew the house to pieces, eating the third little pig.

Moral: Reality doesn’t grade on a curve.

Once upon a time there was a big strong troll who lived under a bridge. A little goat went across the bridge, and the troll reached out to grab and eat the goat. “Wait, Mr. Troll!”, the goat cried. “Soon my brother is coming, and he is even bigger than I am!” The troll let the goat pass, and soon came another goat, twice as big as the first. The troll reached out to grab and eat him, but the brother likewise objected, saying *his* brother was even bigger. Sure enough, a third goat arrived at the bridge, twice as big as the second, and the troll, now ready for a very hearty dinner, reached out to grab and eat him. “Wait!” said the third goat. “My brother is the biggest of us all!”. So the troll let the third goat pass. Then came the fourth goat, who was hundreds of miles tall and blotted out the sun, whose very steps caused earthquakes and made the rivers change course. Without even noticing, he stepped on bridge and troll, pulverizing both to bits.

Moral: Sometimes growth is superexponential.

Once upon a time, Chicken Little ran to her friend Henny Penny. “The sky is falling!” she shouted. “We must tell the king!” Henny Penny joined her, and together they headed toward the capital. On their way they run into their friend Goosey Loosey. “The sky is falling!” they shouted. “We must tell the king!” Goosey Loosey joined them, and together they headed toward the capital. On their way, they ran into the cunning Foxy Loxy. “The sky is falling!” they shouted. “We must tell the king!” “Oh,” said Foxy Loxy. “I know a shortcut to the palace. Follow me into my den.” So the birds all followed Foxy Loxy into his den, where he ate them all, laughing all the while about how gullible they were. Then an asteroid hit Earth, killing everyone.

Moral: Beware the absurdity heuristic.

Once upon a time, a young boy named Jack lived with his mother. Their family was very poor and owned only a single cow. “Go sell this cow at the market,” Jack’s mother told him, “so we will have food to eat for the winter.” Jack went to the market and came back with three beans. “These are magic beans!” he told his mother. “A man told me that when we plant them, they will grow into a beanstalk leading to a land of infinite riches.” His mother pooh – poohed him and threw the beans in the ground angrily. That winter, they both died of hunger.

Moral: Good decision theories should be able to resist Pascal’s Mugging.

Once upon a time, there was an old woodcutter who had no son. He made a little marionette out of pine wood and named it Pinocchio. Then he wished upon a star that it could become a real boy. The star turned out to be the evil Red Fairy, who brought Pinocchio to life, but told him that if he wanted to be a real boy he must murder everyone in the village. That night, Pinocchio took his father's saw and killed Gepetto and everyone else in town.

Moral: Never create an intelligence unless you are certain it will share your values.

Once upon a time, an evil witch transformed a prince into a frog, telling him that only the kiss of a princess could restore him to his proper form. But although he searched around the world, he could find no princess who was willing to kiss a hideous little frog. Finally, he went to the Wise Wizard. "Gender is a social construct," said the Wise Wizard. "Just declare your gender identity to be female, then kiss yourself on the hand or something." So the frog did that, returned to human form, and ruled the land for many years as a wise and benevolent queen.

Moral: Ability to self-modify is just *ridiculously* powerful.

Fetal Attraction: Abortion and the Principle of Charity

Recently, *Alas, A Blog* wrote an article saying that Democrats don’t really care about helping the poor, they only care about increasing government’s ability to take your money. We can prove this, because Republicans consistently give more to charity than Democrats – and because if Democrats really cared about the poor they would stop supporting a welfare system that discourages lifting yourself out of poverty. The only explanation is that the hundred-million odd Democrats in this country are all moral mutants who hold increased labyrinthine bureaucracy as a terminal moral value.

No, wait, sorry! That wasn’t it at all. They were saying that civil rights activists don’t really want to prevent hate crimes against Muslims, they only care about supporting terrorism. We can prove this because they seem pretty okay with the tens of thousands of Muslims who are being killed and maimed in wars abroad that they don’t promote any intervention in – and because they refuse to ban Muslim immigration to America, a policy which would decrease hate crimes against Muslims but also decrease the chance of terrorism. The only explanation is that the hundred-million odd civil rights activists in this country are all moral mutants who hold increased terrorism as a terminal moral value.

No, wait, sorry again! That wasn’t it either! They were saying that pro-lifers don’t really care about fetuses, they just support government coercion of women. We can prove this because they refuse to support contraception, which would decrease the need for fetus-murdering abortions – and because they seem pretty okay with abortion in cases of rape or incest. The only explanation is that the hundred-million odd pro-lifers in this country are all moral mutants who hold increased oppression of women as a terminal moral value.

No, wait, still wrong! I’m totally breaking apart here! They were saying that atheists don’t really doubt the existence of God, but they are too proud to worship anything except themselves. We can prove this because atheists sometimes pray for help during extreme emergencies, – and…

No, wait! It turns out it was actually third one after all! The one with the pro-lifers and abortion. Oops. In my defense, I have trouble keeping essentially identical arguments separate from one another. Anyway, it’s called Pro-Lifers Don’t Give A Damn About Fetuses, They Only Care About Coercing Women, and the main points of the argument, in their original form, are as follows:

> Birth control, as we’ve seen in this thread, is another example. Free, high-quality birth control has been proven, in both studies and in real-world examples, to massively reduce abortion. If pro-lifers real goal was to prevent as many abortions as possible, and if they really believe that the 1.2 million abortions every year are 1.2 million child murders, then they should be willing to compromise on their opposition to birth control in order to prevent millions of child murders. To say otherwise is to say that being uncompromising on birth control is more important than preventing child murder…
>
> The very common pro-life position, which you can see in dozens of examples of actual pro-life legislation, that raped women should be free to abort, but other women shouldn’t be, makes no logical sense at all if pro-lifers believe that a fetus is morally an innocent human child.

> No one would say that it's okay to kill a five-year-old if her father was a rapist. The rape exemption is absurd if the goal of the pro-life movement is to save innocent fetuses; but the rape exemption makes perfect sense if their goal is to target women who choose to have sex.

This thesis in its starkest form says 48% of Americans – 144 million people who are between 45% and 55% women and which includes between 45% and 50% *of* American women, a group that includes all 26,000 members of Feminists For Life and nearly all the founding mothers of feminism including Mary Wollstonecraft, Alice Paul, Elizabeth Cady Stanton, and very ambiguously Susan B. Anthony – pretty much just want women oppressed for the heck of it.

I want to show this thesis is wrong on the object level in two different ways, then move on to showing it's wrong on the meta-level, then say that even if it did make sense and wasn't wrong on two levels it should never have been made, then end by beating it up some more.

By the way, I am firmly pro-choice and have been my entire life, so I have no horse in this race except the horse named Principle of Charity, which has lost its past two hundred ten consecutive derbies and is widely suspected of being dead.

Can We Prove Pro-Lifers Don't Care About Fetuses?

In saying pro-lifers should support contraception, *Alas* is making exactly the error that The Last Superstition warned against. Ze's noticing that Christians do things that don't agree with modern moral philosophy, and so assuming Christians are either stupid or evil, instead of that they have a weird moral philosophy ze's never heard of.

So instead of excusing pro-lifers, start by tarring them further. They don't hate women. They don't love oppression.

It’s much worse than that. Pro-lifers are *not consequentialists*.

Consequentialism is a moral philosophy that says it’s okay to do a lesser evil if it leads to a greater good. I have argued for it at length elsewhere, but one of the reasons I argue for it is that *most people don’t believe it*. Only about a quarter of philosophers are consequentialists, and all the evidence shows that even fewer ordinary people do. Studies of the famous fat man problem show only 10% of people are willing to kill one person in order to save five others, something a true consequentialist would do in a heartbeat.

One group particularly heinous in their rejection of consequentialism is Christians. In his Epistle to the Romans, St. Paul argues that “One may not do evil that good may come”.

The Christians agree with me, against *Alas*, that their rejection of consequentialism is fundamental to their rejection of abortion. According to catholicexchange.com:

> This summer Catholics celebrate the 44th anniversary of Pope Paul VI’s encyclical Humanae Vitae that banned artificial contraception. This encyclical caused a rift that led to the development of a school of moral theology in the church known as Consequentialism. Conseqentialism, which has highly influenced Catholic teaching in our seminaries and universities over the past 40 years, essentially denies objective truth. It has led to what Pope Benedict XVI has called a “dictatorship of relativism”. Many of our so-called Catholic theologians and politicians like Hans Kung, Sister Carol Keehan, former U.S. House Speaker Nancy Pelosi and Kathleen Sebelius are ardent supporters of this nihilistic philosophy. Hardened in their convictions these young

> turks of Consequentialism are largely responsible for our present culture of death. This culture sanctions everything from contraception to abortion, homosexual activity, sex outside of marriage, divorce, sterilization, in-vitro fertilization, pornography, embryonic stem cell research, euthanasia and even false notions of a just war.

Catholic Exchange may not be an entirely authoritative source, so let's find someone who is. How 'bout Pope Pius XI (1857 – 1939), who declared:

> No reason, however grave, can make what is intrinsically contrary to nature to be in conformity with nature and morally right. And since the conjugal act by its very nature is destined for the begetting of children, those who in exercising it deliberately frustrate its natural power and purpose are acting against nature, and are doing something that is base and intrinsically immoral

Alas's argument of "Why don't Christians promote the lesser evil of contraception in order to fight the greater evil of abortion?" assumes a consequentialist viewpoint. And it is *really hard* for consequentialists to take non-consequentialists seriously or to realize how earnest they are about their position, but a decade or so fighting this particular battle has convinced me that *yes*, they are serious and earnest, and no, you can't just round them off to basically a consequentialist and expect their moral beliefs to make sense.

Now the Pope could be lying, of course. This whole "natural law" thing could just be a sham intended to justify oppression of women. But if so, the Pope is definitely playing the long game, seeing as how it started with

Aristotle (4th century BC), reached its zenith with Thomas Aquinas (13th century AD), and has continued through to the present mostly intact, and is the source of religious positions on everything from euthanasia to homosexuality to the afterlife. Once again, if you try to understand Christian moral philosophy without understanding natural law theology, you're going to have a bad time. I've said it, Ozy said it better than me, and Edward Feser said it better than either of us.

I will also add that "lesser evil of contraception" *itself* assumes a modern, non-Christian moral philosophy. Although I think there's still some controversy, a lot of the Church seems to line up behind contraception being a mortal sin. That means it is sufficient, in and of itself, to send you to Hell. Someone who lives an otherwise blameless life but uses contraception once will still burn for all eternity if she doesn't later repent of it.

You can, as I do, think this is stupid. But if so, please localize the stupidity *here*, and not ten logical steps down the line when the people who believe this make the entire reasonable choice of not being totally gung ho about the thing that sends everyone who uses it to eternal torture.

What about the "people allow abortion in cases of rape" argument? The first important counterargument is that no, a whole lot of people don't. The Catholic Church, which once again is pretty high up in the fight against abortion and which I am overrepresenting solely because they're really good at putting their official beliefs down on paper does not:

> In Romans 12:21 Paul says that we should not be conquered by evil but that we should conquer evil with good. The act of rape is a grave sin and an injustice that claims too many people, most often women and

> children, as victims. The Church teaches that victims of rape deserve immediate medical, emotional and spiritual care. The Ethical and Religious Directives for Catholic Health Care Services (#36) states that Catholic healthcare providers have a duty to provide this care and to prevent the contraction of disease or the conception of a child.However, if a child is conceived in a pregnancy caused by rape, then this child is just as innocent and precious as the woman who was victimized and he or she should not be killed because of the actions of the rapist. The Church teaches that through mercy and love, a non-violent solution for both mother and child is far superior to helping a victim of violence (the raped woman) commit violence against her own child through abortion.

And once again, we find this confirmed by the Pope.

As for the people who don't? I give them the benefit of the doubt. Some people say that some of the normal rights we give criminal suspects – the right to confront their accusers, for example – should be suspended in cases of rape out of deference to the trauma suffered by the victim. Still others challenge the very concept of "innocent until proven guilty" in these sorts of cases and the right to hold normal trials at all, again out of deference to the trauma suffered by the victim. And if some anti-abortion people want to relax *their* sacred beliefs out of deference to the trauma of people who have been raped, I am totally going to let them do it without attacking them or pillorying them for their kindness or accusing them of secretly hating women (if they do, *they are doing a very bad job of it*).

But I think most of this is just political compromise anyway. X proposes an anti-abortion bill, Y tries to drum up opposition by saying "But what about rape victims?!" (who are less than one percent of abortions), and X tries to head off the objection and restore support by saying “fine, no abortion for anyone except rape victims”. It’s a good political strategy and it would be surprising if people *didn’t* use it.

And If So, Where Are The People Who *Do* Care About Fetuses?

A while back, I identified the noncentral fallacy as my least favorite argument ever. And one of the reasons I hate it is that it’s *so common*.

“Is a fetus a human being?” is a classic example of the noncentral fallacy. “It’s wrong to murder a human, fetuses are humans, therefore it’s wrong to murder a fetus” is *exactly* the sort of argument people fall for in droves, even though absent a whole lot of implied logical qualifiers it’s got more holes than the PGA tour.

The abortion issue combines about a dozen of the moral quandaries people are really bad at reasoning about. Just to give one example, if we agree it’s wrong to kill an infant, but the only difference between a fetus and an infant is location (inside vs. outside the womb), why isn’t it wrong to kill a fetus? I think these problems are resolvable, but the burden of proof is on me to show that they are and it hardly seems necessary to look for hidden motives in people who disagree.

Some people bring up the whole parasitic violinist argument, but it misses some important parallels (assume the violinist was only connected to the host because of some choice the host made) and I – and a lot of other pro-choice people – think it’s stupid. This seems like a poor argument to convince the sort of pro-life people who are already biased against it.

So if *Alas* is right, and pro-lifers don't really care about fetuses – where *are* all the people who really care about fetuses? In a world where there is a multi-million-person movement of people who get extremely upset that we are killing chickens for food, it would be *really weird* to find that no one at all has any legitimate qualms about killing millions of what's *basically* a smaller, less developed human baby. It's exactly the sort of moral question that people would end up having wildly divergent views about, and for everyone to settle on one side – some people openly, other people secretly – is such a miracle that if I thought it was true I would immediately drop everything else I was doing, try to figure out why people's moral reasoning is in sync on that question and no other, and then try to figure out some way to apply this improbable success to all the other moral questions people actually disagree about.

There are quite a few pro-choice people who believe that abortion is morally wrong, but that the government should not ban it. Is it that hard to believe that, given an action many people think is morally wrong but should not be banned, other people with different views on the role of government might say that since it is morally wrong it *should* be banned? In fact, isn't the person who says "This is murder, so let's not do it" a whole lot *more* honest than the person who says "This is murder, but I think we should continue to permit it anyway"?

In short, in order to believe *Alas'* thesis, we would have to accept *both* that a hundred million pro-lifers who claim they believe in rights for fetuses are lying, *and* explain the absence of about a hundred million pro-lifers we would expect to find merely by the difficulty of the moral dilemma alone. I don't think Occam would be a big fan of this.

What Does It Even Mean To Not Really Believe Something?

I still don't have a good answer to this question. There's a very simple view in which people who claim to care about fetuses are just *lying*, in the same way that a student who procrastinates and then says the dog ate her homework is lying. In this model, the people who lovingly place memorials for the unborn in cemeteries may have, on a page of their diary somewhere, deep in a locked cabinet: "Made a beautiful marble statue today in order to divert attention from the fact that I really like coercing women. Hopefully the press will eat it up and buy all this BS about unborn babies. Going to go wave a picture of a bloody dead baby in front of an abortion clinic tomorrow for exactly the same reason."

I can't prove this isn't happening, but it does seem kind of absurd, plus you would expect one or two of these diaries to be discovered, or someone's private conversation to get leaked, or something like that. As far as I know this has never once happened, even though when politicians are *actually* lying about their motives we tend to hear about it all the time (eg Nixon's tapes).

The more complicated view is that the pro-lifers are deceiving everyone, *even themselves*. But this has its own problems.

It seems likely that everyone in politics is being a bit self-deceptive – this won't come as a surprise to anyone who reads Robin Hanson. Most people discuss political ideas not in order to help other people, but in order to signal how concerned and intelligent they are, or as part of group bonding rituals. Otherwise they wouldn’t be posting “I HATE ABORTION SO MUCH” on Facebook to see how many “likes” they can get, they’d be out canvassing door-to-door or (even better) just working overtime at their job to donate money to anti-abortion charities without ever mentioning it to anyone. Certainly the average person who puts an “Abortion Stops A Beating Heart” bumper sticker on their car isn’t doing it because they have theory by which their action later results in babies being saved (or women being oppressed, or anything at all happening in the external world). I can even notice this sort of process happening in my own head in real time when I think about efficient charity.

Yet all of these actions manifest on a *conscious* level as being genuinely concerned about the issue I’m talking about, or putting up a bumper sticker about, or donating money to.

So we have a model of the brain that includes at least two levels: a surface honest level, where you really care about fetuses, and a deep signaling level, where you just want to impress the other people in your church and signal to yourself that you are a compassionate caring person.

I’m not sure how much room that leaves for an intermediate level, accessible neither to your real desires nor to your conscious motives, that wants to oppress women, or what it would mean for such an intermediate level to exist if it had no effects either on your deepest thoughts or on your actions.

One could argue that the *historical cause* is oppression of women. For example, some claim that one of the reasons cannibis was banned in the United States was that William

Randolph Hearst was afraid hemp paper would threaten his investment in the paper industry and so started a moral panic around it. In that case, one might be able to claim concern for the paper industry is the *real* reason drug warriors oppose marijuana, even though there isn't any part of their minds which is secretly thinking about paper mills.

But this, too, seems unlikely to me. Opposition to abortion dates back well before government coercion of women could have possibly been an issue, back when women were their husbands' property and men didn't need to come up with excuses to treat them poorly. In pre-1900s discourse it was almost always discussed in the same breath as infanticide – see for example the quotes by the early suffragettes above. And, of course, it is implied by exactly the same natural law arguments that apply to non-gendered issues as well like euthanasia (in fact, if I weren't so logorrheic I probably could replace this entire article with "Compare opposition to abortion and opposition to euthanasia, draw relevant conclusions". Too late.)

So I think claims like "pro-lifers *really* want to control women" displays way too little philosophical curiosity about what exactly "really want" is supposed to mean.

Beware Genetic Fallacy

At the beginning of this post, I gave four examples of claims that group X doesn't *really* believe position Y that they claim to believe. These sorts of arguments are pretty common in politics, which is too bad because they're all just the genetic fallacy.

Even if we could prove, beyond a shadow of a doubt, that every single pro-lifer was only in it for the misogyny, that wouldn't affect the validity of pro-life arguments one bit. If abortion is a form of murder equal in moral culpability to murder of the born, then even if every other person in the world who holds the position is a jerk, we shouldn't do it.

I'm pushing this point in order to propose at least the partial elimination of this form of argument. It doesn't address the main point at hand – that is, whether the government should ban abortion. It makes the debate much nastier – instead of attacking opponents' ideas, suddenly we're launching personal attacks into their moral character while leaving their ideas alone. It's totally unsolvable one way or the other – we can't stick a probe in pro-lifers' minds and read off their true values. And since people, religious or not, aren't actually that good at responding to insults by meekly turning the other cheek, it'll probably lead to a never-too-difficult response in kind ("pro-choicers don't really care about women, they just want a culture of total moral relativism and to chip away at respect for the sanctity of life"). It encourages the style of politics where your enemies are innately evil and so you don't have to second-guess yourself or seek compromise with them.

I'm not saying it doesn't sometimes have value. I think a good explanation of how signaling behavior changes people's thought processes can help them tease out which of their views are entirely due to signaling, and so consider changing them. This would probably take the form of "You don't really believe X, you're just saying that to signal Y". But this is both a bit more philosophically sophisticated than the "pro-lifers just hate women" argument, and – crucially – is part of a general insight about human nature rather than a coincidental insight that the people you're opposed to just happen to be jerks.

Conclusion

It seems unlikely that the people at the Genocide Awareness Project don't care about fetuses and want to keep women down (especially since two of its directors are women). It seems unlikely that the huge freak-out across the right-wing blogosphere about the Kermit Gosnell trial was because Gosnell might in some way have helped women escape government coercion. It seems unlikely that the bloggers I really know and like who consistently oppose abortion for principled reasons, like Elizabeth Scalia, are just trying to punish women for having sex.

And it seems unlikely that a group of millions of people with a two thousand year old moral tradition telling them to oppose abortion, and a known history of having cognitive biases that would cause them to hate abortion, *actually* are lying and hate abortion for a totally different reason that not one of them has ever secretly let slip, and which is so far from normal human morality that it would require them to be the villains of their own life stories, *at the same time* that millions of people who *actually* worry about abortion are mysteriously AWOL.

So at the risk of beating a dead horse…PRINCIPLE OF CHARITY, PEOPLE!

Alas ends with the following relatively gloomy pronouncement:

> If what we have is two sides, one of which has the policy goal "reproductive freedom for women,' the other of which has the policy goal "women shouldn't ever choose to have sex, but if they do they should be forced to give birth," then there is no possible compromise.

But I'm more optimistic! If you debate pro-lifers on the object-level issues like "Does a woman's right to privacy outweigh a fetus' right to life" or some other horrible muddled version of the question like that, then no, you will never be able to make progress.

But if you devote yourself to the actual substantive disagreement – which is on the level of what meta-ethics to use or maybe even further back – then you can either change some minds or, at the very least "be confused on a higher level and about more important things". This continues to be my strategy and I'm pretty happy with how it's worked out.

EDIT: *Some people in the comments have steelmanned the original objection to "Pro-lifers just want to force women who have sex to bear the consequences for their supposed misdeeds". This sounds both more plausible than the original and, not coincidentally, less outrageous.*

But analyzing it more closely, it turns into pretty much the same as the standard argument pro-lifers endorse. We usually agree that people should bear the consequences of what they do. For example, if someone spends money rashly and then goes bankrupt, this is the natural order of things. We also agree that if there's some "out" that helps them avoid those consequences, that is good – for example, if they can take a new job that earns them more money, we are happy they managed to avoid bankruptcy. But if that "out" hurts other people, we tend to go back to thinking they should bear the consequences of their own action – for example, if taxpayers have to bail out an irresponsible corporation, we'd rather the corporation bears the consequences of its irresponsibility than the taxpayer, even if the bailout could save the corporation.

The pro-lifers seem to be doing much the same thing. Sometimes sex has unwanted consequences. There's a *convenient way to avoid those consequences – abortion – which would be very helpful. But if that convenient way to avoid the consequences hurts another person – the fetus – then we're back into "take the consequences of your own action" mode.*

So while I agree there is a certain element of "you must bear the consequences of your own action" going on here, I don't think it is opposed to claiming the main issue is the rights of the fetus, nor that the pro-lifers are particularly denying it, nor that it's anything especially evil.

EDIT 2: *Was just reminded that I was against abortion (but pro-contraception) for a few years as a teenager, and it definitely* felt *like I was really concerned about fetuses. Then I got a better understanding of noncentral fallacy and figured out better tools to use on the problem of fetuses' moral value and went back to being pro-choice.*

EDIT 3: *Joe in the comments finds a study that casts doubt on the claim that contraception decreases abortion rate using the Peltzman effect. An article on Patheos makes the same claim at much more length, showing that in most countries, increased availability of contraception correlates with increased abortion rates. There are controlled, experimental studies tend to show that providing free contraceptives decreases abortion rate, but Gilbert casts serious doubt on one Probably needs some more investigation.*

Hansonian Optimism

Imagine a kingdom ruled by a wise and benevolent king who, by reason of some strange tradition, is prohibited from ever leaving his palace. He only receives information on the affairs of the kingdom from his various Viziers. Like most Viziers, they are evil and power-hungry, and they are all conspiring with some of the most brutal and oppressive nobles in the land to preserve their reign of terror.

Also, you have an adorable pet bear

One day the Heroine speaks out against the current conditions in the kingdom. Taxes are too high, the peasants are starving to death, and people are being enslaved, all to enrich a few brutal nobles. The Heroine goes from town to town with her message: the people must beg the King to do something about this problem.

The Viziers hear of this and go to the king. “The Heroine,” they say, “is speaking against you. The whole kingdom is happy and prosperous, but this one woman wants to tear it apart and start a civil war for her own personal enrichment. Your people beg you to do something about her before she destroys the golden age they are currently enjoying.”

And so the king orders the Heroine executed, an order which the Viziers and the nobles are all too happy to carry out.

In this kingdom all the laws would be utterly selfish and show no regard for the average citizen. But this would be totally consistent with the King himself being a good person.

In fact, the King could be a *perfectly* good person, a person who attains moral heights of which other people never even dream, simply because he would never face a true moral dilemma. Suppose there were some problem that might prove morally difficult for the King – for example, the eastern states, which provide most of the kingdom’s silk, are rebelling, and the king could either choose to live in peace with the newly independent east, or brutally crush them. If he chose the first option, silk would cost a lot more, and the king really likes silk.

If he were aware of this situation, it would be a sort of moral dilemma – do I do the right thing and avoid a war, or do I do what’s convenient for me and lets me keep my luxury goods? Thanks to the Viziers, this problem disappears. If the Viziers want silk, they can tell the king that the eastern states have just launched a surprise attack, complete with atrocities – they must be dealt with as a matter of existential threat to the kingdom itself. And if the Viziers don’t want silk, they can just tell the king that the east ran out of silk, too bad, nothing we can do about it. In fact, the Viziers will *never* present a true moral dilemma to the king, because then they wouldn’t know which side he’d choose.

And so the king is faced only with easy, convenient moral decisions, and is able to preserve perfect innocence and purity. No matter how awful and tyrannical his decisions, the populace may at least take consolation that their king is, at heart, a good person.

These were some of the thoughts that went through my head when I read Ozy's The Inherent Goodness of Human Nature, or Lack Thereof. Ozy worries that Hansonian explanations – in which people do nice things mostly for selfish reasons like signaling or self-signaling – mean that there's no such thing as goodness. As ze puts it:

> Robin Hanson (if I understand him correctly) would argue that the person giving money to the Make a Wish Foundation doesn't actually want to help sick children; they want to feel nice, like the sort of person who helps sick children, and– more importantly– they want everyone else to believe that they're nice people who help sick children.
>
> My initial reaction to this is "No! That's horrible! You terrible person!" Unfortunately, "you're a terrible person!" is not actually an argument that something is not true. My sense of revulsion at that idea is nothing more than a sign that I'm biased in favor of the "humans: basically nice" explanation.

Robin himself commented by saying:

> I love people, even if I don't think they are as good as they like to let on. I hope others can love me under the same conditions.

This seems like one of the wisest things I have ever heard, and restores just a little of my faith in humanity. But I think I'm more optimistic than Robin is.

Like Ozy, I believe human nature is basically good even though people's actions seem based on selfish and amoral motives. This is no more contradictory than the King being basically good, even though all his decrees will seem based on selfish and amoral motives. If the King has no access to accurate information, but can only make decisions based on information gleaned from biased sources, then the biases of those sources will be reflected in his words and deeds.

I cannot say why I identify other people with the Kings of their minds rather than with the Viziers of their minds (or with the creepy guys standing next to the king of their minds) save that this is who I feel I am in my mind, it is how I would like other people to see me, and so it seems both accurate and kind to see other people that way as well. Upon this view, people are good by nature, far better than their actions suggest, and it is really hard not to love and respect them.

This is not to say I think there's no such thing as evil. I would prefer that evil be something different than mere stupidity, something more than "Osama bin Laden was dumb enough to believe his mental Viziers when they told him becoming a terrorist mastermind was the right thing to do, poor guy", and indeed it seems there are lots of good stupid people and evil smart people, even lots of irrational good people and rational evil people. Although I have no clear answer, I think I would define evil as certain habits of mind which make it extremely easy for your Viziers to put one over on you, certain tendencies like "other people would do the same to me, so I'm just giving them a taste of my own medicine if I hurt them."

This is *still* an attempt to be a good person – it's an attempt to create a moral system in which you are just and virtuous for hurting others – but once you're letting your Viziers use this kind of argument on you I think it's pretty safe to say you've gone evil. This doesn't quite correspond to my inner intuitive impression of evil, but if I turn it from a specific English-language assertion to a sort of preconscious sense of wrongedness and arrogant self-justification that expresses the same idea, it might.

You may notice how nicely this meshes with Trivers’ theory of consciousness.

Wisdom of the Ancients, Redux

The authors of the paper on Victorian intelligence have seen my critique and kindly replied.

HBD Chick has already addressed their response from pretty much the same perspective I will, but just to make it official…

I originally complained that Galton's Victorian sample was more elite than the modern day samples in the study, and so could be expected to have higher IQ by selection bias alone. The authors counter that Galton broke his sample down into demographic subcategories, and each of those subcategories does better than the corresponding subcategory today – therefore, demographic differences are not responsible for the IQ difference.

HBD Chick pointed out, correctly, that the similarly-named categories hide some big differences. For example, Silverman directly compares Galton's Victorian students (presumably college students, since we're talking age 18 – 30) to 1940s college students. While both groups are college students, one is more elite than the other: according to National Center For Education Statistics, the percent of Americans with bachelors' degrees rose from about 2% in 1900 to about 10% in the 1940s. If England was similar, and if college is pretty meritocratic, then the 1880 student sample was drawing form about the smartest 2% of people, and the 1940s student sample from about the top 10% of people.

The other comparison the authors brought up was Anger's 1993 sample of "220 postal, hospital and insurance workers from three different US cities", which they compared to

Galton’s Clerical/Semi-Skilled and Semi-Professional groups and found that, even matched job for job, the Victorians did a bit better. This is good work, but I am still skeptical for two reasons.

First, in a society with a large underclass, the middle-class will be comparably elite. Suppose that we took the least productive 90% of people in 2013 America and put them in those power-generating pods from the Matrix. We still need postal workers, but now the postal workers are coming from what was previously the top 10% of the population – maybe people who would have been college professors in the old system. But the Victorians had a comparably larger underclass (small farmers and factory workers) and a comparably smaller middle class than we do. Therefore, we would expect Victorians with a standard middle class occupation to be drawn from more elite segments of the population than a comparable modern with the same occupation.

Second, Galton’s sample was *still* self-selected from people who came to a science museum in an elite city and paid him money! I propose that the average postman you will find buying science-related merchandise in the gift-shop of the Exploratorium in San Francisco will be smarter than the average postman drawn randomly from the distribution of postmen!

So although I appreciate the authors’ time and politeness in responding to me, I am still skeptical of their findings.

Let me move to one other point that many of my commenters protested. In the original article, I said:

> All the theory here sort of checks out, except for the part where they say IQ changed 15 points in a hundred

> years, which is just a little bit faster than any responsible person expects evolution to progress. People critique the idea that Ashkenazi Jews could have shifted fifteen points in nine hundred years on the grounds that it’s too fast.

Many people correctly noted that it’s a lot harder to evolve a complex feature in the first place than it is to break one, and therefore my comparison was invalid. I actually considered this when writing the original post, but decided to leave it in. Here’s why.

In a normally distributed trait like IQ, over a short period like nine hundred years most of the action isn’t going to come from *de novo* mutations. It’s going to come from people on one side of the bell curve out-reproducing their peers on the other. This can happen just as quickly for smartening up as it can for dumbing down.

A commenter writes:

> The sheer 15 point drop made me go ‘nope’ right away. If I remember Lynn right we should be losing about one IQ point per generation of our genetic potential in the West, two thirds of that due to mass immigration which wasn’t a big factor for Britain before the late 70s. The dysgenic drop can’t be more than about 3 points.

These seem like much more reasonable numbers. They also seem like a good reason not to worry about dysgenics ever again. How many more centuries do we have in which natural selection is going to be the main force shaping our genome, as opposed to genetic engineering or transfer to nonbiological life? *Maybe* one, if you’re really pessimistic? Losing three IQ points before that time (most of which are probably instantly regained via Flynn Effect or some kind of low-budget intelligence amplification like the racetams) isn’t anything to lose much sleep over.

Literally Inconceivable: Contraceptives And Abortion Rates

I have amazing parents who would never do something sneaky like install a keylogger on my computer to keep tabs on me as I move thousands of miles away from home. But if I'm wrong and they *did* do that, they're probably sweating pretty hard right now. My search history for the last two days looks a lot like "efficacy of contraceptives", "contraceptive failures", "pregnancy risk if contraceptive failure", "unintended pregnancy", "unintended pregnancy abortion", and "COME ON GOOGLE WHY WON'T YOU GIVE ME GOOD INFORMATION ON UNINTENDED PREGNANCIES AAARGH"

(this last one brings up a berkeley.edu address, which does not surprise me one bit)

My parents can relax – the searches are because of the comments on a recent post of mine. In response to a claim that pro-lifers should be in favor of contraception since it decreases abortions, I argued that moral philosophy doesn't always work that way, but let the main point – that contraception decreases abortion – stand. Some people, especially Joe and Gilbert, challenged my assumption, leading to an unpleasant wade through the swamp of contraception-related data.

The Anti-Contraceptive Position

Let's start with the nay-sayers. In what he *claims* is a long essay (ye call that long? I'll be showin' ye long!) Bad Catholic evaluates correlation between abortion and contraception rates in different countries. He finds – mostly using data from the pro-choice Guttmacher Institute, a huge clearinghouse of abortion data we will be returning to again and again – that:

> Contraception has been shown to decrease abortion rates primarily in [ex-Soviet bloc] countries with already high abortion rates. These represent a minority of countries. Contraception has been shown to increase abortion rates primarily in [non ex-Soviet bloc] countries with already low abortion rates. These represent a majority of countries. Contraception has been shown to slightly reduce abortion rates after its initial increase of abortion rates, but has never been shown to reduce abortion rates back to pre-contraception levels. This is my claim. I have no doubt that there's a lot more to say, given the incredible amount of studies I haven't seen. But as far as I can tell, this is a claim far closer to the truth than the oft-repeated, always unexamined "Contraception reduces abortion rates".

This at first sounds bizarre – how could contraception, a technology that decreases unintended pregnancy – increase abortion, a result of unintended pregnancy? Enter the Peltzman effect, aka risk compensation.

I have a deep love for the Peltzman effect. Part of this is that it's one of the few terms we social scientists have that sounds as nifty as the one physicists and mathematicians bandy around all the time. Another part is that a girl messaged me on OKCupid once explaining the Peltzman effect to me and asking me on a date (I never claimed my life was normal). But the rest of it is that it's jsut this really elegant and unexpected finding where across a broad set of domains people respond to hard-won advances that make them safer with "Cool! Now I can behave irresponsibly!" It's been found with anti-lock brakes (drivers drive closer to the car in front of them), with seat belts (people just drive faster), and with childrens' safety gear (children just behave more recklessly).

The Peltzman effect doesn't always hold true; sometimes we expect it and can't find it. And it rarely makes things *worse* – it usually is cited as keeping things at the same level they were before, and one well-studied exception, the Munich taxi study, only finds a tiny increase in accidents.

But inside view here – how many people here, if they don't want kids, would be willing to have totally unprotected sex? And how many people would be willing to have sex using condoms? But condoms have a typical failure rate of 15% – meaning that if a couple has sex for a year using only condoms for protection, there's a 15% chance the woman will get pregnant. With combined oral contraceptive pill, it's 8%. So if these contraceptive methods make people about ten times more willing to have sex when they don't want pregnancy – not at all hard to imagine! – they could raise the unintended pregnancy rate and therefore the abortion rate.

This is the context of Joe's study showing that in Spain from 1997 – 2007, a large rise in contraceptive usage occurred simultaneous with a large rise in abortion.

A few other arguments seem transparently stupid to me – for example, some people like to point out that US states with high contraception rates also have high abortion rates, but that's mostly a feature of those states being very liberal and so allowing abortion clinics to operate there. So lets move on to…

The Pro-Contraceptive Position

Just in case you thought you were going to escape without any graphs:

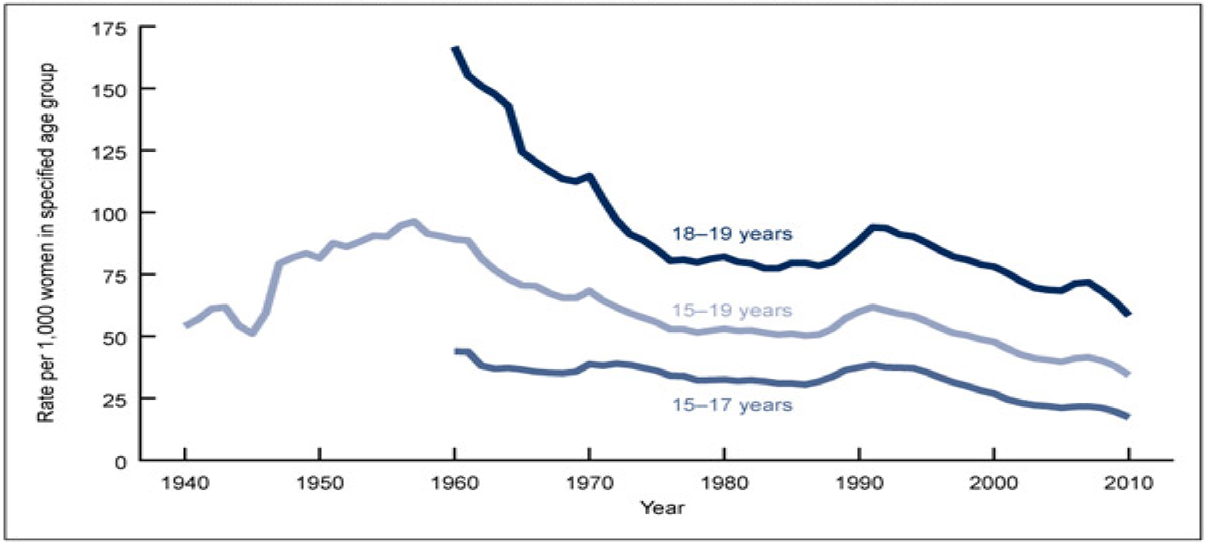


Figure 1. Birth rates for women aged 15–19: United States, 1940–2010, and by age, 1960–2010

NOTE: Data for 2010 are preliminary.
SOURCE: CDC/NCHS, National Vital Statistics System.

Here's teenage birth rates over the last 75 years. Like nearly all social problems, they have been steadily and somewhat mysteriously declining, but we notice an especially sharp decline around 1960, the year the Pill was introduced. Abortion wasn't legalized until the 70s and was pretty uncommon before then, so we can leave it out of this analysis and say that it *sure looks like* the invention of a new form of contraception decreased pregnancies.

Also near an all-time low are abortion rates (I assume they mean "all-time low since abortion was legalized"?). This seems to be due to *both* fewer unintended pregnancies *and* less willingness to end unintended pregnancies with abortion. Santelli et al 2002 find the decline to correspond nicely to increasing use of contraceptives. Some pro-lifers claim (data unavailable), that after 2002 abortions continued to drop even though contraception use stayed steady. Unfortunately, all I can find is the CDC saying contraceptive use continued to rise – in the absence of their contrary data, I'm giving this point to the "increased contraception helped reduce abortion" people.

(one way that we could reconcile these two results, if we were feeling very generous, is to say that overall contraception use has remained the same, but users have switched to more effective modern forms of contraception like the implant or IUD, with a failure rate less than 1/100th that of condoms)

But this is actually consistent with Bad Catholic's claim above – contraception increases the abortion rate when first introduced, then eventually stabilizes and decreases it, but never back to the level before contraception. So let's evaluate that one. Bad Catholic writes:

> An honest look at the data shows that in virtually every country that increased the use of contraception, there was a simultaneous increase in that country's abortion rate. In England (Rise in contraceptive use: simultaneous rise in abortions), France (Rise in contraceptive use: simultaneous rise in abortions), Australia, (Rise in contraceptive use: simultaneous rise in abortions), Portugal (Whose abortion rate only began to rise after 1999, after oral contraceptive methods were made widely available), Canada (Whose abortion rate only began to rise after the legalization of oral contraceptives in 1969), and, as the Guttmacher Institute shows, Singapore, Cuba, Denmark, the Netherlands, and South Korea, to name a few.

Let's investigate the countries in order. The claim seems to be only that abortions and contraceptive use rose "simultaneously", but following his links this turns out to mean "throughout the 20th century". There is no attempt to prove that the particular shape of the contraception curve matched that of the abortion curve or anything like that, just that there was more contraception in 2000 than in 1950, and, whaddya know, more abortions as well. The same methodology could very easily correlate abortion with global temperature. His statistics on England, France, and Australia all seem to be of this type.

He makes a stronger claim about Canada: that “the abortion rate only began to rise after the legalization of oral contraceptives in 1969”. You know what else was legalized in Canada in 1969? Abortion. I’m going with “probably not a good test case”.

As for Portugal, the claim that oral contraceptive methods were legalized in 1999 seems wrong; his own link says they have been available since 1985 and that only the emergency contraceptive pill was made available in 1999. Further, his claim that “abortion rates only began to rise after 1999” also seems wrong – his link shows what looks like a pretty linear rise in abortion rates from 1996 to 2006; I don’t think anyone eyeballing those numbers would be tempted to consider 1999 anything remotely like an inflection point. My own guess for an inflection point would be 2007, and sure enough when I Google it that was the year they fully legalized abortion.

The Guttmacher Institute doesn’t link to its sources as diligently as Bad Catholic, so I’m just going to accept their claim that six countries – Singapore, Cuba, Denmark, Netherlands, US, and South Korea – saw simultaneous increases in contraception and abortion – after all, it goes against the direction of their bias so they have no incentive to lie. They give their results the following explanation:

> The reason for the confusion stems from the observation that, within particular populations, contraceptive prevalence and the incidence of induced abortion can and, indeed, often do rise in parallel, contrary to what one would expect. The explanation for these

counterintuitive trends is clear.2 In societies that have not yet entered the fertility transition, both actual fertility and desired family sizes are high (or, to put it another way, childbearing is not yet considered to be “within the calculus of conscious choice”3). In such societies, couples are at little (or no) risk of unwanted pregnancies. The advent of modern contraception is associated with a destabilization of high (or “fatalistic”) fertility preferences. Thus, as contraceptive prevalence rises and fertility starts to fall, an increasing proportion of couples want no more children (or want an appreciable delay before the next child), and exposure to the risk of unintended pregnancy also increases as a result. In the early and middle phases of fertility transition, adoption and sustained use of effective methods of contraception by couples who wish to postpone or limit childbearing is still far from universal. Hence, the growing need for contraception may outstrip use itself;4 thus, the incidence of unintended and unwanted pregnancies rises, fueling increases in unwanted live births and induced abortion. In this scenario, contraceptive use and induced abortion may rise simultaneously.

As fertility decreases toward replacement level (two births per woman), or even lower, the length of potential exposure to unwanted pregnancies increases further. For instance, in a society in which the average woman is sexually active from ages 20 to 45 and wants two children, approximately 20 of those 25 years will be spent trying to avoid pregnancy. Once use of highly effective contraceptive methods rises to 80%, the potential demand for abortion, and its incidence, will fall. Demand for abortion falls to zero only in the “perfect contraceptive” population, in which women are protected by absolutely effective contraceptive use at

> all times, except for the relatively short periods when they want to conceive, are pregnant or are protected by lactational amenorrhea.5 Because such a state of perfect protection is never actually achieved, a residual demand for abortion always exists, although its magnitude varies considerably among low-fertility societies, according to levels of contraceptive use and choice of methods.

This seems *incredibly* reasonable, and we will come back to it later. Let's abandon all of these time series type studies and see if we can find a halfway-decent controlled experiment.

Well, uh…we can find *a* controlled experiment. These people in St. Louis gave people free contraceptives and later found that they had a teenage pregnancy rate much lower than the rest of the population. Gilbert gives this study exactly the correct criticism – participants from a very specific population (poor people in St. Louis interested in signing up for a contraceptive study) are being compared to the general population (everyone in the United States). This is inexcusable, especially considering that Real Science has an extremely standard way of avoiding this problem (sign people up for your study, only give the intervention to a randomly selected half, and the other half is an instant control group). Other fatal issues – the study used IUDs, the most effective form of contraception, but most of the worry that contraception might increase abortion comes from less effective means like condoms and the Pill. Finally, if you're really interested in the way that widespread availability of contraceptives makes a culture more libertine, just giving them to a couple of people within that culture isn't going to capture that effect. I am maybe a little bit hugely disappointed that most of the media and bloggers reporting on this didn't mention these sorts of issues.

But it does show one interesting thing, which is that when people get free contraception, they start using more effective contraception methods. Would this also cause risk compensation? I don't know, but I feel like there has to be some amount of sex beyond which it's just *no longer fun*, and some contraceptive methods are so effective that it would be *really hard* to have so much sex that they're worse than nothing.

Let's close this section with a few minor points.

Contraceptive advocates point to the Netherlands, with one of the lowest abortion rates in the world. Given the stereotypes of the Dutch, they *probably* didn't get that way through careful abstinence, and indeed their government is unusually generous in providing free contraceptives.

It turns out people can just survey women having abortions and ask them if they used contraceptives or not! 54% of abortion patients were using contraception at the time, which pro-life websites get very excited about: "IT'S MORE THAN HALF!" But putting these numbers in context may diminish their enthusiasm: the *four-fifths of American women who use contraception* account for 54% of abortions; the fifth of women who don't use it account for the other 46%. The Guttmacher Institute gets more or less the same numbers, but frames them in a very convincingly pro-contraceptive way:

> The two-thirds of U.S. women at risk of unintended pregnancy who use contraception consistently and correctly throughout the course of any given year account for only 5% of all unintended pregnancies. The 19% of women at risk who use contraception but do so inconsistently account for 44% of all unintended pregnancies, while the 16% of women at risk who do not

> use contraception at all for a month or more during the year account for 52% of all unintended pregnancies.

So it seems clear that the more (and better) you use contraception, the less likely you are to have an abortion.

Summary

I think we can use these results to build a consistent picture.

Contraceptive and abortion rates often rise simultaneously. This rise is not necessarily causal, and is more likely to be due to both being parts of the same philosophy – people want to have lots of sex but not have kids. As this philosophy becomes more widespread, as it has nearly everywhere in the 20th century with the Sexual Revolution and Demographic Transition, both contraception and abortion will rise. As it gains ground, both contraception and abortion will become more legal and available, making them rise even further. It is unclear to what degree the availability of contraception itself causes the rise of this philosophy. I'm intrigued by this claim that penicillin rather than the pill started the Sexual Revolution, but if someone wants to claim that it was all due to contraceptives, I don't have enough expertise in the area to prove her wrong.

On the other hand, once a society has undergone this transition and settled on "lots of sex, few kids" as being its dominant values, then the local application of more contraception seems to decrease abortion rates. We know this because of the surveys of abortion patients saying they are disproportionately likely not to be contraceptive users. We know this because of the decline in teenage pregnancies with the advent of the Pill. And we also notice the game-changing nature of new, more effective contraceptives with near-zero failure rates replacing older, more fallible ones, and the not-provably-causal but certainly suggestive secular decline in abortion rates that corresponds with that replacement.

Overall my guess would be that a society that legalizes contraceptives would see an increase in abortion rates (which might or might not be causal depending on that society’s situation), but that in a society like our own, where contraceptives are already legal and the demographic transition is pretty much complete, increasing access to contraceptives is probably going to decrease abortion. And increasing access to extremely effective contraceptives like the implant or RISUG, especially when they replace less effective contraceptives like the condom, are very very probably going to decrease abortion.

Hasta La Victorians Siempre

It seems to be Gush About The Victorians Month in the academic community or something.

How The Mid-Victorians Worked, Ate, and Died (h/t Michael Vassar) claims that the mid-Victorian period was a golden age of health during which life expectancy was as high or higher than today and the diseases we consider a "normal part of aging" simply failed to exist. No cancer, no heart disease, just living happily and healthily until you were felled by tuberculosis at age 85. They credit the Victorian diet – in particular its lack of additives and preservatives and its overload of nutrients, the latter made possible by very high calorie diets that compensated for their increased level of physical activity. The authors admit we probably can't safely replicate such a high calorie diet in today's sedentary society, but suggest taking lots and lots of vitamins and supplements in order to get the same high nutrient level the Victorians did.

I am *really really bad* at understanding nutrition (AT LEAST *I* ADMIT IT!), so I will limit this to attempts to fact-check a few claims, plus some extremely speculative commentary. Let's start with the thesis:

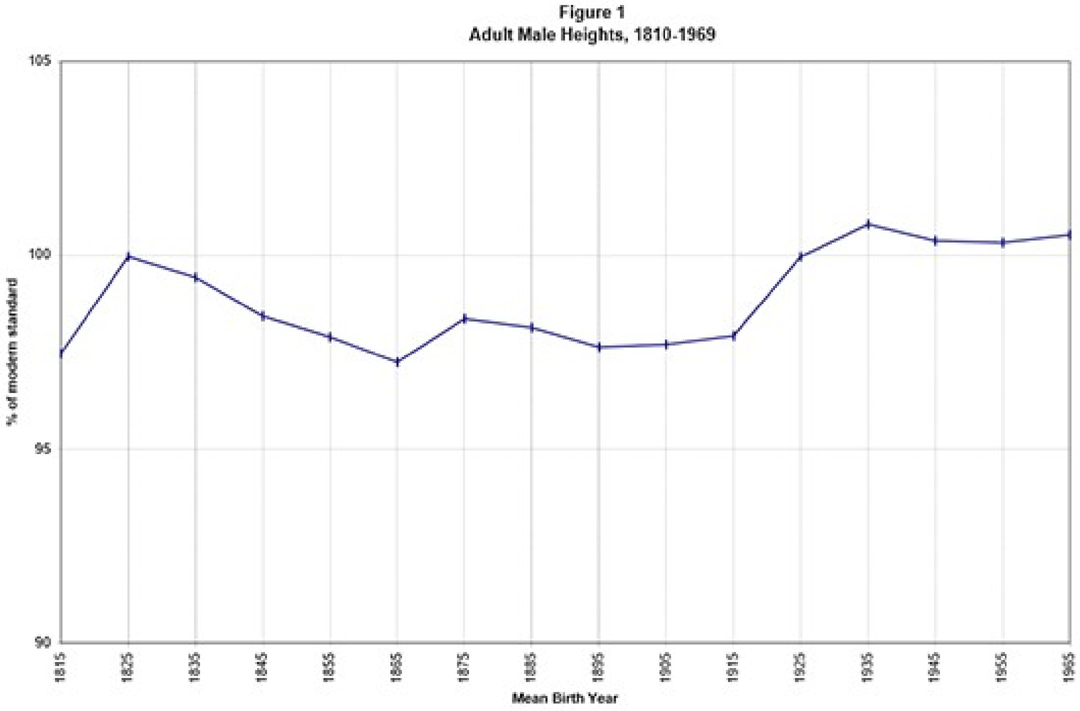
> We argue in this paper, using a range of historical evidence, which Britain and its world-dominating empire were supported by a workforce, an army and a navy comprised of individuals who were healthier, fitter and stronger than we are today. They were almost entirely free of the degenerative diseases which maim and kill so many of us, and although it is commonly stated that this is because they all died young, the reverse is true; public

> records reveal that they lived as long – or longer – than we do in the 21st century.

And the evidence in support:

> The fall in nutritional standards between 1880 and 1900 was so marked that the generations were visibly and progressively shrinking. In 1883 the infantry were forced to lower the minimum height for recruits from 5ft 6 inches to 5ft 3 inches. This was because most new recruits were now coming from an urban background instead of the traditional rural background (the 1881 census showed that over three-quarters of the population now lived in towns and cities). Factors such as a lack of sunlight in urban slums (which led to rickets due to Vitamin D deficiency) had already reduced the height of young male volunteers. Lack of sunlight, however, could not have been the sole critical factor in the next height reduction, a mere 18 years later. By this time, clean air legislation had markedly improved urban sunlight levels; but unfortunately, the supposed ‘improvements’ in dietary intake resulting from imported foods had had time to take effect on the 16–18 year old cohort. It might be expected that the infantry would be able to raise the minimum height requirement back to 5ft. 6 inches. Instead, they were forced to reduce it still further, to a mere 5ft. British officers, who were from the middle and upper classes and not yet exposed to more than the occasional treats of canned produce, were far better fed in terms of their intake of fresh foods and were now on average a full head taller than their malnourished and sickly men.

This is very suspicious. If the British Army started recruiting from different populations, or started lowering their standards generally because they needed more men, these data are useless. I tried to find an alternate source of height data, and the best I could come up with was the extremely thorough Height, Weight, and Body Mass of the British Population Since 1820, who say:



This graph seems to do quite well at picking up real decreases in height like the one starting in the birth cohort of 1825 (possibly due to the Corn Laws?), but it has pretty much nothing to say about any great collapse in the last few years of the 19th century. Assuming the average army recruit was 20 years old. In 1883, they would have been in the 1863 birth cohort; in 1901, in the 1881 birth cohort. But in fact, *the 1881 birth cohort is taller on average than the 1863 birth cohort!*. The army’s troubles with finding sufficiently tall infantrymen cannot be due to the Amazing Shrinking Englishman – it must be a matter of casting a wider recruiting net or something.

> The crude average figures often used to depict the brevity of Victorian lives mislead because they include

infant mortality, which was tragically high. If we strip out peri-natal mortality, however, and look at the life expectancy of those who survived the first five years, a very different picture emerges. Victorian contemporary sources reveal that life expectancy for adults in the mid-Victorian period was almost exactly what it is today. At 65, men could expect another ten years of life; and women another eight. This compares surprisingly favourably with today's figures: life expectancy at birth (reflecting our improved standards of neo-natal care) averages 75.9 years (men) and 81.3 years (women); though recent work has suggested that for working class men and women this is lower, at around 72 for men and 76 for women.

Given that modern pharmaceutical, surgical, anaesthetic, scanning and other diagnostic technologies were self-evidently unavailable to the mid-Victorians, their high life expectancy is very striking, and can only have been due to their health-promoting lifestyle. But the implications of this new understanding of the mid-Victorian period are rather more profound. It shows that medical advances allied to the pharmaceutical industry's output have done little more than change the manner of our dying. The Victorians died rapidly of infection and/or trauma, whereas we die slowly of degenerative disease. It reveals that with the exception of family planning, the vast edifice of twentieth century healthcare has not enabled us to live longer but has in the main merely supplied methods of suppressing the symptoms of degenerative diseases which have emerged due to our failure to maintain mid-Victorian nutritional standards. Above all, it refutes the Panglossian optimism of the contemporary anti-ageing movement whose protagonists use 1900 – a nadir in health and life expectancy trends – as their starting point to promote the idea of endlessly increasing

> life span. These are the equivalent of the get-rich-quick share pushers who insisted, during the dot.com boom, that we had at last escaped the constraints of normal economics. Some believed their own message of eternal growth; others used it to sell junk bonds they knew were worthless. The parallels with today's vitamin pill market are obvious, but this also echoes the way in which Big Pharma trumpets the arrival of each new miracle drug.

I was wondering how long it would take "Big Pharma" to make an appearance.

Anyway, here we turn to Health and Welfare During Industrialization, which includes the following table:

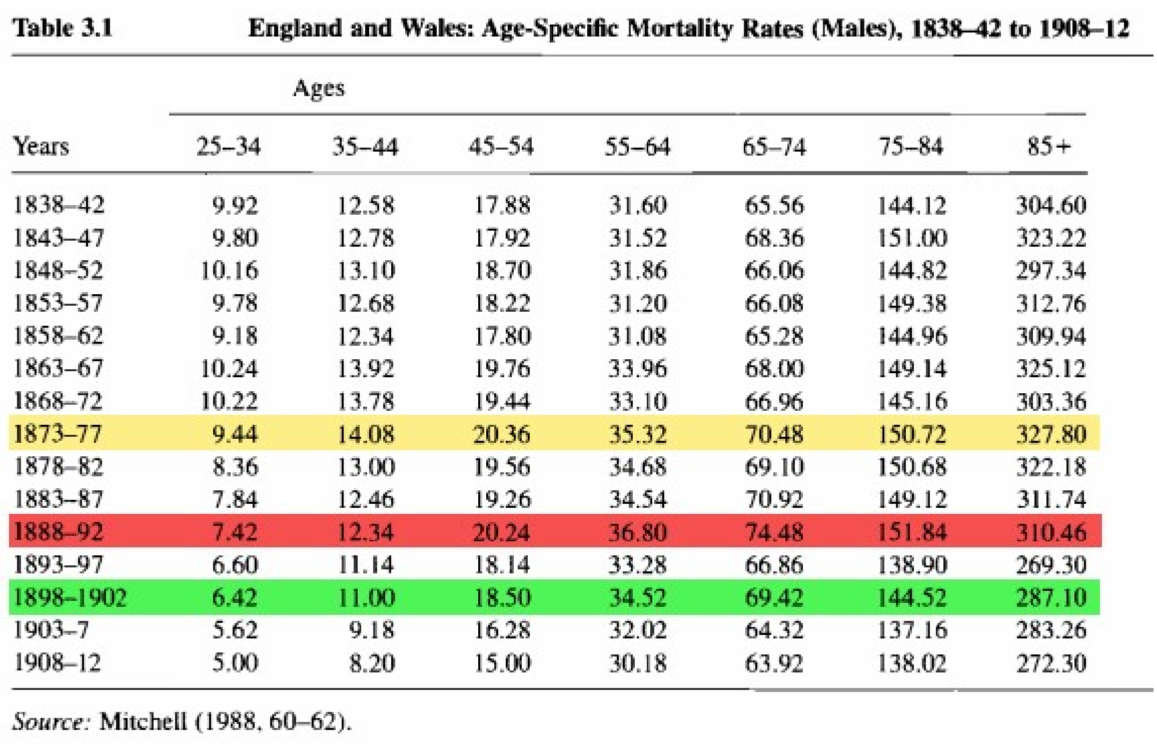


Table 3.1 England and Wales: Age-Specific Mortality Rates (Males), 1838–42 to 1908–12

	Ages						
Years	25–34	35–44	45–54	55–64	65–74	75–84	85+
1838–42	9.92	12.58	17.88	31.60	65.56	144.12	304.60
1843–47	9.80	12.78	17.92	31.52	68.36	151.00	323.22
1848–52	10.16	13.10	18.70	31.86	66.06	144.82	297.34
1853–57	9.78	12.68	18.22	31.20	66.08	149.38	312.76
1858–62	9.18	12.34	17.80	31.08	65.28	144.96	309.94
1863–67	10.24	13.92	19.76	33.96	68.00	149.14	325.12
1868–72	10.22	13.78	19.44	33.10	66.96	145.16	303.36
1873–77	9.44	14.08	20.36	35.32	70.48	150.72	327.80
1878–82	8.36	13.00	19.56	34.68	69.10	150.68	322.18
1883–87	7.84	12.46	19.26	34.54	70.92	149.12	311.74
1888–92	7.42	12.34	20.24	36.80	74.48	151.84	310.46
1893–97	6.60	11.14	18.14	33.28	66.86	138.90	269.30
1898–1902	6.42	11.00	18.50	34.52	69.42	144.52	287.10
1903–7	5.62	9.18	16.28	32.02	64.32	137.16	283.26
1908–12	5.00	8.20	15.00	30.18	63.92	138.02	272.30

Source: Mitchell (1988, 60–62).

This is confusing, so let me explain.

I've taken the mortality tables from Steckel and deleted everything before age 25. Clayton and Rowbotham admit that infant mortality has improved since the Victorian period, so we're not counting that.

I've marked the 1873 to 1877 period – the last period unambiguously before Clayton and Rowbotham's 1880 "nutritional collapse" – in gold. This is our baseline.

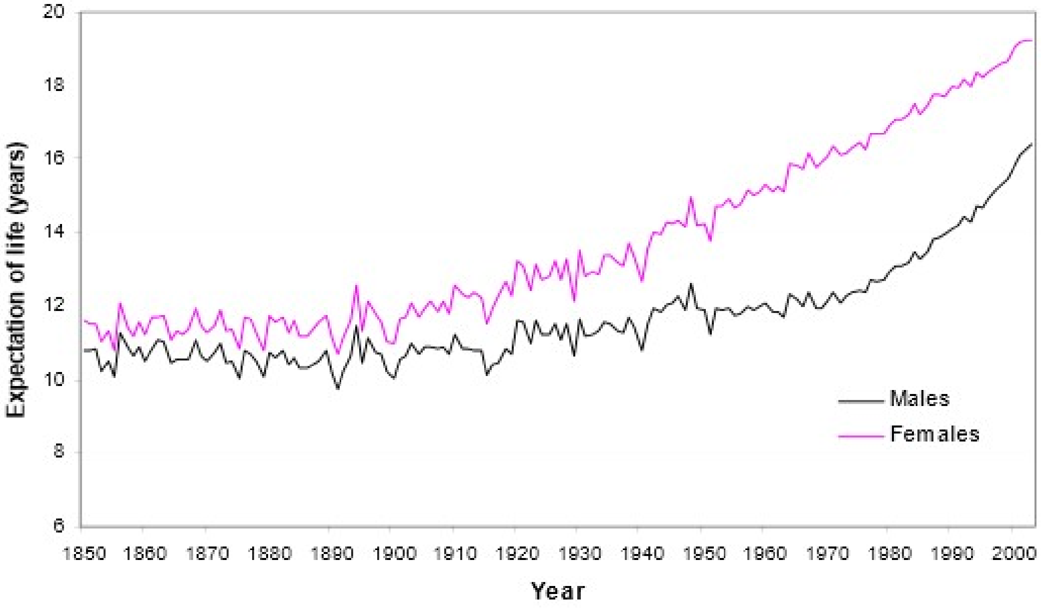
We notice that the subsequent period actually has lower mortality in nearly every category. This is fine. Just because they've invented less nutrititious food doesn't mean that people have started eating it yet, or that it's had time to affect their health. Instead, I've looked for the *worst* post-1880 period and marked that in red as the bottom of the "nutritional collapse". As you can see, it's 1888 – 1892.

Finally, I've taken the period including 1900 – what they describe as "a nadir in health and life expectancy trends", and marked it in green.

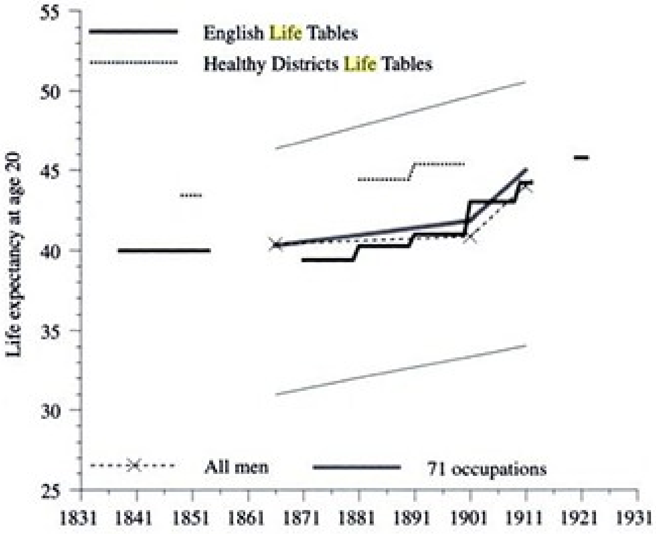
As we can see, the "nadir" – 1900, is actually *lower mortality in all age groups* than the "golden age" of 1873 – 1877. The "collapse", if it occurred at all, was a tiny statistical blip that was then easily overpowered by the general secular trend of decreasing mortality.

I don't know exactly how the authors would like me to consider "people born in 1900" vs. "people dying in 1900", but luckily I also happen to have life expectancy at age 65 for Britain for the entire period of 1841 to 2001 (God, I *love* the Internet!):

Figure 2: Period expectation of life at age 65, England and Wales, 1850-2003



There’s nothing on there that could remotely be the kind of collapse they’re talking about, unless it be that little blip around 1890 which is more than compensated for five years later. Annnnnd here’s some more data, divided by occupation and showing life expectancy at age 20:



This isn't especially high resolution, but it's interesting in that it suggests adult life expectancy in the Victorian period was around 65 and not the more modern 75. Sure enough, the book I took it from estimates life expectancy at age 20 (ie excluding childhood mortality) to be about 62.5. Unfortunately, I can't access Clayton and Rowbotham's sources, so I can't tell whether they make an even better argument that it really *was* 75. But this matters for claims like "there was much less cancer", because cancer increases *a lot* with age and so if you're cutting off the end of the age distribution of course you're going to get less cancer (imagine you learned people never went bald in caveman times – this could be either because they all ate nutritious mammoth meat, or because they all died before age 30).

Overall I'm going to say I wish I could access their sources, but the fact that they don't quote any numbers or show any graphs in support of their "life expectancy decrease from 1880 to 1900" hypothesis makes me suspect they don't have any [EDIT: See bottom of post].

Their next point is that the Victorians had much less cancer and heart disease than people today. This I am totally willing to believe. The authors provide ample evidence in another paper of theirs, and it accords well with very similar evidence presented by Gary Taubes in *Good Calories, Bad Calories*.

The Victorian Age was before the mass-marketing of cigarettes, before most of the carcinogenic chemicals we inhale all the time, and their diet contained ample fruits and vegetables – foods that we know lower cancer and heart disease risk. Many of them were manual laborers, there were much fewer obesogenic foods, and people were generally normal weight. It is not mysterious or revolutionary to propose they had less cancer and heart disease than we did.

But let's look at Clayton and Rowbotham's explanation:

> Degenerative diseases are not caused by old age (the ‘wear and tear’ hypothesis); but are driven, in the main, by chronic malnutrition. Our low energy lifestyles leave us depleted in anabolic and anti-catabolic co-factors; and this imbalance is compounded by excessive intakes of inflammatory compounds. The current epidemic of degenerative disease is caused by widespread problem of multiple micro- and phyto-nutrient depletion (Type B malnutrition.)
> With the exception of family planning and antibiotics, the vast edifice of twentieth century healthcare has generated little more than tools to suppress symptoms of the degenerative diseases which have emerged due to our failure to maintain mid-Victorian nutritional standards. The only way to combat the adverse effects of Type B malnutrition, and to prevent and / or cure degenerative disease, is to enhance the nutrient density of the modern diet.

Whoa whoa whoa whoa whoa.

Gary Taubes does a pretty good job recording all the primitive cultures that *also* have no degenerative diseases. The Eskimos are one such. They basically just eat meat. You can also find cultures pretty much anywhere, with any diet, who also lack these degenerative diseases. In fact, I *think* people who are *actually* malnourished – starving Africans and the like – still have lower rates of these degenerative diseases as long as they’re not eating a “modern” diet.

To me, this suggests that their “phytonutrient depletion” hypothesis needs to contend with another – that it’s not that we’re not getting enough of the right stuff, but rather that we’re getting *too much* of the wrong stuff. As far as I know, this is what mainstream nutrition science believes as well as most of the more interesting crackpots, although of course everyone differs as to what the wrong stuff is. It is a promising and venerable theory and nothing the authors of this paper have said thus far casts the slightest doubt upon it. Indeed, they admit that the “nutritional collapse” was caused by the introduction of preserved foods, additives, and cheap sugar.

> Our levels of physical activity and therefore our food intakes are at an historic low. To make matters worse, when compared to the mid-Victorian diet, the modern diet is rich in processed foods. It has a higher sodium/potassium ratio, and contains far less fruit, vegetables, whole grains and omega 3 fatty acids. It is lower in fibre and phytonutrients, in proportional and absolute terms; and, because of our high intakes of potato products, breakfast cereals, confectionery and refined baked goods, may have a higher glycemic load. Given all this, it follows that we are inevitably more likely to suffer from dysnutrition (multiple micro- and phytonutrient depletion) than our mid-Victorian forebears…
>
> Since it would be unacceptable and impractical to recreate the mid-Victorian working class 4,000 calorie/day diet, this constitutes a persuasive argument for a more widespread use of food fortification and/or properly designed food supplements (most supplements on the market are incredibly badly designed; they are assembled by companies that do not understand the real nutritional issues that confront us today, and sell us pills containing irrational combinations and doses that can do more harm than good.
>
> To insist, as orthodox nutritionists and dieticians do, that only whole fruit and veg contain the magical, health-promoting ingredients represents little more than the last

> gasp of the discredited and anti-scientific theory of vitalism (‘Vitalism—the insistence that there is some big, mysterious extra ingredient in all living things—turns out to have been not a deep insight but a failure of imagination’, Daniel Dennett) Even the stately FSA concedes that fruit juices count towards your five-a-day, as do freeze-dried powdered extracts of fruits and vegetables. As our knowledge of phytochemistry and phytopharmacology increases, it has become perfectly acceptable to use rational combinations of the key plant constituents in pill or capsule form.
>
> These arguments are developed in ‘Pharmageddon’, a medical textbook which illustrates how micro- and phyto-nutrients can be specifically combined in order to prevent and treat the chronic degenerative diseases that characterise and dominate the 20th and 21st centuries; and how they could be integrated into our food chain in order to reduce the contemporary and excessively high risks of the degenerative diseases to the far lower mid-Victorian levels.

So first things first. I am almost sure I went to medical school, and the sorts of textbooks I read there all had names like “Essentials Of Biochemistry, Third Edition”, and “Introduction To Gastroenterology” and almost never names like “Pharmageddon”.

Second, the authors cite some sources for their claim that all supplements currently on the market are poorly designed and do more harm than good. These sources show that a bunch of different supplements are either ineffective or cause cancer, and are entirely correct. What they do not cite are any sources that show that “correctly designed” supplements do more good than harm. This is because those sources do not exist because no one has ever discovered that.

The reason the medical community isn’t switching wholesale from evil pharmageddon-causing drugs to all-natural Victorian-approved nutrient supplementation isn’t because they’re in the grip of Big Pharma, it’s because *no one can find supplements that are consistently proven to work*.

(the medical community actually *is* saying “eat right and exercise”, but NO ONE LISTENS)

And yes, part of the lack of working supplements is the coordination problem of “there’s no money in studying supplements”. I think it would be great if we could figure out some way to coordinate supplement studies more effectively. But it’s kind of hard to make that case when all the supplement studies that have been done have been total duds and the pro-supplement community has just kept replying with “But those supplements were poorly designed!” and then suggesting a perplexing and extremely contradictory trove of other possible designs, all of which themselves are later found not to work.

One point that seriously enlightened me, thought, was the authors’ comments on vitalism. It’s definitely true that eating whole foods has useful properties (like reducing risk of diseases) and also definitely true that taking supplements that supposedly share vitamins with those foods doesn’t. And it’s *also* true that I was previously taking this as an Unfortunate Fact Of Life. Which is *dumb*. Of *course* we should be trying to figure out what magical property of whole foods makes them effective and then trying to deliver that magical property in a pill. I am disappointed at myself for not realizing that was important sooner, and that alone makes me profoundly grateful I read this study.

It’s just that I don’t think we’re quite there yet, and until we are, supplementation isn’t all that useful.

In summary, I can’t confirm this paper’s suggestion of dire health costs from a “nutritional collapse” in the 1880s. However, the notion of a *gradual* rise in cancer, heart disease, and other degenerative diseases linked to a modern diet seems correct. Their attribution of this to nutrient deficiency is unsupported and probably only a small part of the problem, and their opinions on supplementation seem to possibly veer into crackpottery. However, they are right to note that we need a better science of supplementation and that once we develop this science, supplements really really ought to work.

EDIT: *Andrew G points out in comments that the study compared Victorian life expectancy at age 60 to modern life expectancy *at birth* in their attempts to say life expectancy was staying about the same. This is clearly not kosher. Victorian life expectancy at age 20 (a good point to rule out childhood mortality) was about 62.5, and modern life expectancy at age 20 is…something older than life expectancy at birth which is 77 or so. So their “life expectancy has stayed the same” argument is completely wrong.*

Moments Of Awakening

Sometimes I try to meditate when I am very tired. This is a bad idea. The brain is never very good at being held fixed on a single object of attention, and when it's tired the task becomes nearly impossible. Instead, I get progressively more florid hypnagogic delusions centering around the meditation practice in increasingly tangential ways.

(yes, I just said that something centered in a tangential way. Suck it, circles.)

The meditation practice is very simple. Breathe in, breathe out, just watch the breath.

As I become more tired, my brain tries to complicate things. Breathe in, breathe out. Was that a correct in-breath? Was it a correct out-breath? How can I tell? Maybe I should judge each breath dispassionately, looking for flaws?

Now a train of thought is begun, and I am far too tired to notice or stop it. Breathe in, breathe out. I judged that breath to be good, but how can I be sure I was really dispassionate? Perhaps I need to do it three times per breath, to establish inter-rater reliability?

Then I lose the plot entirely. Breathe in, breathe out. Judge each breath three different times. Why am I judging my breath three different times? Will something terrible happen if I don't? Perhaps it will. There have been a lot of tornadoes lately. Maybe if I don't get the breath exactly right, and judge it three times, there will be a tornado. Or maybe I'm already in a tornado, and it will only let me out if I get the breathe right.

Over a matter of seconds, the plot thickens. Breathe in, breathe out. Okay, I’m in a tornado, and the only way to get out is to judge the breath correctly three times. But is this the correct way to get out of the tornado? Maybe I’m just thinking this superstitiously, and *actually* God is refusing to let me out of the tornado precisely *because* I am being arrogant by thinking I can judge my own breathing. Maybe I need to let someone else do it. What about my sister? She’s the first person I can think of who’s both intelligent enough to judge correctly and kind enough to sympathize with my plight.

By the tenth or so breath, I am in some kind of elaborate fantasy world. Breathe in, breathe out. Okay, I’m in a tornado. God won’t let me out until I have three perfect breaths. But I can’t judge my own breath. I need my sister to do it. But if I ask her to do it, will she be caught in the tornado with me? And will God be upset if I drag my sister into this, knowing that I’ve inconvenienced her? Or will He be angrier if I continue arrogantly setting myself up as my own judge?

And then suddenly something snaps and there’s a moment of awakening. There’s no need to judge my own breathing, that’s not the point. I’m not in a tornado. There’s probably no God. And I don’t even *have* a sister.

All I need to do is breathe in, breathe out. Breathe in, breathe out.

I’ve made the extremely good decision not to meditate while tired anymore, because it obviously doesn’t work, but I sort of *miss* this kind of situation. The rush of euphoria I get when I realize that this entire complicated and dangerous situation I thought I was stuck in was entirely of my own devising, that there’s a clear and obvious path forward and all is nice and organized – well, it’s a really good feeling.

I think a lot of my passions are attempts to capture that same feeling on a larger scale. This is how I feel about consequentialism, for example. Every so often I get into this really complicated moral debate and there’s this huge argument on who has the right to do what and which group is more oppressed and how dare X think Y, and the suddenly I think “Wait a second! Consequentialism! I can just figure out which act has the better consequences!” And even if that’s hard – the way that keeping your mind on your breath is hard – watching what you *thought* was an unsolvable harder problem dissipate like fog at sunrise feels pretty great.

Atheism, Bayesian reasoning, capitalism, consequentialism, logical positivism, determinism – at their best moments all of these share that feeling of taking a seemingly impossible problem and realizing that it’s simpler than you think, that the nightmarishness was a fever dream, that you don’t even have a sister. And of course that’s no argument for these ideas – sometimes you do have a sister, sometimes you are caught in a tornado. But there’s something about that feeling which is fundamental to my aesthetic sense.

All Debates Are Bravery Debates

“I don’t practice what I preach because I’m not the kind of person I’m preaching to.”
— J. R. “Bob” Dobbs

I.

I read Atlas Shrugged probably about a decade ago, and felt turned off by its promotion of selfishness as a moral ideal. I thought that was *basically* just being a jerk. After all, if there’s one thing the world doesn’t need (I thought) it’s more selfishness.

Then I talked to a friend who told me Atlas Shrugged had changed his life. That he’d been raised in a really strict family that had told him that ever enjoying himself was selfish and made him a bad person, that he had to be working at every moment to make his family and other people happy or else let them shame him to pieces. And the revelation that it was sometimes okay to consider your own happiness gave him the strength to stand up to them and turn his life around, while still keeping the basic human instinct of helping others when he wanted to and he felt they deserved it (as, indeed, do Rand characters).

II.

The religious and the irreligious alike enjoy making fun of Reddit’s r/atheism, which combines an extreme strawmanning of religious positions with childish insults and distasteful triumphalism. Recently the moderators themselves have become a bit embarrassed by it and instituted some rules intended to tone things down, leading to some of the most impressive Internet drama I have ever seen. In its midst, some people started talking about what the old strawmanning triumphalist r/atheism meant to them (see for example here).

A lot of them were raised in religious families where they would have been disowned if they had admitted to their atheism. Some of them *were* disowned for admitting to atheism, or lost boyfriends/girlfriends, or were terrified they might go to Hell. And then they found r/atheism, and saw people making fun of religion, and insulting it, in really REALLY offensive ways. And no one was striking them down with lightning. No one was shouting them down. No one was doing much of anything at all. And to see this taboo violated in the most shocking possible way with no repercussions sort of broke the spell for them, like as long as people were behaving respectfully to religion, even respectfully disagreeing, it still had this aura of invincibility about it, but if some perfectly normal person can post a a stupid comic where Jesus has gay sex with Mohammed, then there's this whole other world out there where religion holds no power.

Gilbert tells the story of how when, as a young Christian struggling with doubt, he would read r/atheism to remind himself that atheists could be pretty awful. r/atheism is doing a bad job at being the sort of people who can convert Gilbert, and the new mods' policy of "you should have more civil and intellectual discussions" might work better on him. I think it would work better on me too.

But there is – previously unappreciated by me – a large population of people for whom really dumb offensive strawmannish memes are *exactly what they need*.

III.

A friend described his experiences in the Landmark Forum's self-improvement workshop. He said their *modus operandi* was to get people to take responsibility for the outcome of their actions. His example was an office worker who always did substandard work, and was always making excuses like "My boss doesn't s support me" or "My computer system isn't good enough" or "My coworkers aren't pulling their fair share." Landmark says those kinds of excuses are what's keeping you back. And they taught (again, according to this one person) that the solution was to treat everything that happens in your life as your responsibility – no excuses, just "it was my fault" or "it's to my credit".

Then a few days later, I was reading a book on therapy which contained the phrase (I copied it down to make sure I got it right) "Don't be so hard on yourself. No one else is as hard on yourself as you are. You are your own worst critic."

Notice that this encodes the *exact opposite assumption*. Landmark claims its members are biased against ever thinking ill of themselves, even when they deserve it. The therapy book claims that patients are biased towards always thinking ill of themselves, even when they don't deserve it.

And you know, both claims are probably spot on. There are definitely people who are too hard on themselves. Ozy Frantz has done an amazing job of getting me and many other people inclined towards skepticism about feminist and transgender issues, engaging with us, and gradually convincing us to be more respectful and aware through sheer kindness and willingness to engage people reasonably on every part of the political spectrum. Two days ago some people on Twitter – who were angry Ozy said one need not boycott everything Orson Scott Card has ever written just because he's against gay marriage – told Ozy they weren't a real transgender person and suggested lots of people secretly disliked them. And instead of doing what I would do and telling the trolls to go to hell, Ozy freaked out and worried they was doing everything wrong and decided to delete everything they had ever written online. I *know* Ozy is their own worst critic and if that therapy book was aimed at people like them, it was entirely correct to say what it said.

On the other hand, I look at people like Amy’s Baking Company, who are obviously terrible people, who get a high-status professional chef as well as thousands of random joes informing them of exactly what they are doing wrong, who are *so clearly in the wrong* that it seems impossible not to realize it – and who then go on to attribute the negativity to a “conspiracy” against them and deny any wrongdoing. They could probably use some Landmark.

IV.

In a recent essay I complained about bravery debates, arguments where people boast about how brave they are to take an unorthodox and persecuted position, and their opponents counter that they’re not persecuted heretics, they’re a vast leviathan persecuting everyone else. But I think I underestimated an important reason why some debates *have to* be bravery debates.

Suppose there are two sides to an issue. Be more or less selfish. Post more or less offensive atheist memes. Be more or less willing to blame and criticize yourself.

There are some people who need to hear each side of the issue. Some people really need to hear the advice “It’s okay to be selfish sometimes!” Other people really need to hear the advice “You are being way too selfish and it’s not okay.”

It’s really hard to target advice at exactly the people who need it. You can’t go around giving everyone surveys to see how selfish they are, and give half of them *Atlas Shrugged* and half of them the collected works of Peter Singer. You can’t even write really complicated books on how to tell whether you need more or less selfishness in your life – they’re not going to be as buyable, as readable, *or* as memorable as *Atlas Shrugged*. To a first approximation, all you can do is saturate society with pro-selfishness or anti-selfishness messages, and realize you’ll be hurting a select few people while helping the majority.

But in this case, it makes a really big deal what the majority actually is.

Suppose an Objectivist argues “Our culture has become too self-sacrificing! Everyone is told their entire life that the only purpose of living is to work for other people. As a result, people are miserable and no one is allowed to enjoy themselves at all.” If they’re right, then helping spread Objectivism is probably a good idea – it will help these legions of poor insufficiently-selfish people, but there will be very few too-selfish-already people who will be screwed up by the advice.

But suppose Peter Singer argues “We live in a culture of selfishness! Everyone is always told to look out for number one, and the poor are completely neglected!” Well, then we want to give everyone the collected works of Peter Singer so we can solve this problem, and we don’t have to worry about accidentally traumatizing the poor self-sacrificing people more, because we’ve already agreed there aren’t very many of these at all.

It’s much easier to be charitable in political debates when you view the two participants as coming from two different cultures that err on opposite sides, each trying to propose advice that would help their own culture, each being tragically unaware that the other culture exists.

A lot of the time this happens when one person is from a dysfunctional community and suggesting very strong measures against some problem the community faces, and the other person is from a functional community and thinks the first person is being extreme, fanatical or persecutory.

This happens a lot among, once again, atheists. One guy is like “WE NEED TO DESTROY RELIGION IT CORRUPTS EVERYTHING IT TOUCHES ANYONE WHO MAKES ANY COMPROMISES WITH IT IS A TRAITOR KILL KILL KILL.” And the other guy is like “Hello? Religion may not be literally true, but it usually just makes people feel more comfortable and inspires them to do nice things and we don’t want to look like huge jerks here.” Usually the first guy was raised Jehovah’s Witness and the second guy was raised Moralistic Therapeutic Deist.

But I’ve also sometimes had this issue when I talk to feminists. They’re like “Guys need to be more concerned about women’s boundaries, and women need to be willing to shame and embarrass guys who hit on them inappropriately.” And maybe *they* spent high school hanging out with bros on the football team who thought asking women’s consent was a boring technicality, and *I* spent high school hanging out entirely with extremely considerate but very shy geeks who spent their teenage years in a state of nightmarish loneliness and depression because they were too scared to ask out women because the woman might try to shame and embarrass them for it.

And the big one is trust. There are so many people from extremely functional communities saying that people need to be more trusting and kind and take people at their word more often, and so many people from dysfunctional communities saying that’s not how it works. Both are no doubt backed by ample advice from their own lives.

A blog like this one probably should promote the opinions and advice most likely to be underrepresented in the blog-reading populace (which is *totally different* from the populace at large). But this might convince “thought leaders”, who then use it to inspire change in the populace at large, which will probably be in the wrong direction. I think most of my friends are too leftist but society as a whole is too rightist – should I spread leftist or rightist memes among my friends?

I feel pretty okay about both being sort of a libertarian and writing an essay arguing against libertarianism, because the world generally isn’t libertarian enough but the sorts of people who read long online political essays generally are way more libertarian than can possibly be healthy.

Lies, Damned Lies, And Facebook (Part 3 of ∞)

EDIT: *This one got complicated. See below.*

I tried to scrub my Facebook feed of all of these kinds of articles, but occasionally something gets through. This week what's gotten through, *en masse*, is the Gilbert/Frago case, in which a guy killed an escort and was acquitted of murder.

Let's look at the article titles: In Texas, It's Okay To Murder An Escort If She Won't Sleep With You, says Alternet, Texas Jury Says Sometimes Killing An Escort Is Okay, says Feministe, adding that " "but I thought she was going to have sex with me" is now a viable defense for manslaughter". abovethelaw.com says Killing A Hooker Is A-OK, Guess What State, because stereotypes are hilarious and you should always perpetuate them. And Mr. Conservative writes "Apparently, in Texas it's okay to kill an escort if she won't sleep with you."

The story is that a guy hired an escort, which is often a code word for a prostitute, for $150. In this case, it was *not* a code word for a prostitute and the escort provided him with company and nothing more. The guy got angry, and as she drove away, he shot several bullets at her car. One of them ricocheted off the car, hitting and injuring her, and she later died from the injury.

During the trial, the defense argued that in Texas, it is legal to use deadly force to defend your property. This man believed the escort had stolen his $150 (by charging him for sex and then not providing it), so he shot at her car in order to stop her from getting away with committing a crime, which is *sort of* like self-defense if you really really squint. The jury declared the man not guilty of murder.

How does the Internet interpret the decision? From Alas, A Blog:

> I strongly suspect that for many juries, the life of a sex worker isn't considered as valuable as the life of a store manager or even a drug dealer. Certainly not as valuable as something really precious, like $150.

From "Everyday Victim Blaming"

> A woman meets a man, she declines sex and he shoots her, and a jury has decided that it's OK because she was a sex worker. It is hard to know where to even start with how scary the decision of the jury is, there are so many ways it blames the victim and excuses a man for ever being responsible for his behaviour. The implications for all women should make anyone reading very afraid too. A man's assumption about sex being an automatic right after a gift, or money, or marriage is now the situation in Texas. How many other women are in danger? Of course people will say it's different because money was involved, I have already seen some of this victim blaming going on. There are a number of scenarios here; they all blame the victim though.

From Shakesville:

> And because she was a sex worker, and because she was marginalized and he is privileged, a jury has ruled that it's okay. No harm. No foul. It's not like anyone important was killed today, it's not like anyone important was hurt by her death, it's not like anyone important will be terrorized in the wake of this blatant ruling that men can murder women and after the fact

> with no living witnesses to contradict them claim that they were sex workers who weren’t performing according to expectations and thus get off free and clear.

The most eloquent commentary, is, as usual, provided by Freethought Blogs:

> It’s legal in Texas to kill a woman for refusing to have sex?

So if you’ve had any experience with this section of the blogosphere before, it probably won’t surprise you to learn that the case had nothing to do with sex, nothing to do with privilege, and everything to do with legal principles that the jury applied pretty much entirely correctly.

Bridgette Dunlap, who unlike these bloggers is an actual lawyer and has read facts about the case beyond the one article on Gawker that started this whole thing, writes:

> The much more plausible reason for the verdict is that the jury believed the defendant’s claim that he didn’t intend to shoot the victim. Per Texas’ homicide statute, the prosecution needed to prove that Gilbert “intentionally or knowingly” killed Frago or intended to cause her “serious bodily injury.” The defense argued that Gilbert lacked the requisite intent for murder because when he shot at the car as Frago and the owner of the escort service drove away, he was aiming for the tire. The bullet hit the tire and a fragment, “literally the size of your fingernail,” according to Defense Attorney Bobby Barrera, hit Frago. Barrera does not believe the jury acquitted because of the defense of property law. He believes they acquitted because they believed Gilbert didn’t mean to shoot her…

> One would expect the jury to find that shooting at a car with an AK-47 is at least “reckless,” in which case he could have been convicted of manslaughter. But the prosecution didn’t charge him with manslaughter, only murder. Manslaughter is a “lesser included offense” of murder and the judge is entitled to instruct the jury if the evidence supports that charge, but it appears she did not. The jury can’t convict on a charge that isn’t before them.

In other words, guy hit tire of car, bullet ricocheted off and hit a woman. Prosecutors mysteriously turn case into referendum on whether the guy meant to kill the woman. Man says he didn’t, he was just aiming at the tire (which he in fact hit). Jury believes him and acquits. Under this interpretation, the only “scandal” is why the prosecution charged murder instead of the more legally defensible manslaughter.

A commenter on this blog disputes Professor Dunlap’s analysis:

> Scott’s wrong on this, I’m pretty sure. Under Texas felony murder rule, the only defense was defense of property. If the jury did not find defense of property, this would have been murder. As such, there’s no game in manslaughter at all; the prosecutors were right not to charge it . . .
>
> It seems to me to be absolutely right to have charged murder; there is a manslaughter lesser if supported by the evidence. Here, given Texas law, it appears that the self-defense would be justified if it was for recovery of stolen property. (No other state permits deadly force for recapture of property.) See below as to why murder was the *only* rational charge . . .

“Malice” is a term used in most jurisdictions for murder, and includes “depraved-heart murder.” Texas repealed that in 1974 when they changed section 19.02 of their criminal law. The commenter quoted for the theory that murder requires direct desire to kill rather than implied malice is basically correct as far as it goes, but appears to me to be entirely wrong for another reason.

Texas felony murder rule is different. If you’re somewhere else there’s something called the merger doctrine that doesn’t always make your felonious attack murder. In Texas, if you, say, feloniously shoot at a car and inadvertently kill an occupant, that’s felony murder. (Felony murder is usually: You and I go to rob a store. I shoot the clerk fatally. Bad news for you: You’re on the hook for murder. Don’t rob stores with violent armed criminals.)

This was almost certainly found to be self-defense.

Salisbury v. State, 90 Tex.Crim. 438, 235 S.W. 901, 902 (1921), observed that one who shoots wantonly and recklessly into a car or building known to him to be occupied need not have the specific intent to kill any particular person in order to make him guilty of murder. Texas went through several gyrations since then.

In 1974 in Hilliard v. State, the court ruled that the intent to commit the underlying felony (here, injury to a child) was sufficient to make the child’s death felony murder.

Upholding that view is Rodriguez v. State (http://tinyurl.com/mc2foyk) (And that’s the link you want to follow for a great overview of Texas felony murder). It makes clear that there is no merger doctrine

> in Texas – if shooting at the car was felonious – and without justification it surely was – defendant is guilty of murder. That’s why they tried this particular defense and stressed it – “I shot at the occupied car itself” is a felony (Texas Penal Code section 22.05(b)), so this leads directly to felony murder. In short, our blog host is wrong. It is possible I am wrong; I do not have the full tools at my disposal to solve this puzzle 100%, but I’d bet heavily on me. The screaming bloggers are sort of right: The hooker-stole-my-money defense worked.

This was then followed by another learned legal scholar suddenly showing up on this blog (why am I the only person who has to deal with this?) and arguing the opposite:

> JRM, I didn’t get the impression they were going for felony murder. Why would the defense attorney say he thought he they won on lack of intent if it was felony murder?
>
> I could be missing something. I am not a criminal expert (but did run the piece past some criminal law profs). I’m most interested in the narrative because I think the inaccuracy/hyperbole will make some people more emboldened to harm women.
>
> I assumed it was felony murder as well, but it appears not. It isn’t just the post-trial statements (which have been changing). Reporting from during the trial shows they were arguing no intent.

A commenter on another blog who claims to be a lawyer in Texas says that “stand your ground against nighttime intruder” law does not apply in this case anyway:

> As I read the applicable law, the defense allows you to use lethal force against a night-time intruder who is

> committing a crime (and here’s the kicker) only if you yourself are not engaged in unlawful activity. That clause seems specifically designed to prevent johns from using this defense against prostitutes. If you ask me, the DA in this case screwed up big-time. Someone should be losing their job over this.

I had some stronger conclusions here before, but right now I’m just going to end with “law is complicated”.

Medical and Psychological Studies for June

Original spider picture edited out after complaints.

The benefits of exposure to ultraviolet light may outweigh the risks, say scientists who apparently hate putting on sunscreen as much as I do. This reminds me of my father's occasional complaints that every year he finds out that the health advice he was giving his patients last year was driving them to an early grave.

It's been a good month for gut microbiota research. A team from UCLA finds that probiotics change activity in brain areas that affect emotional processing in humans as well as actual behavior on emotion-related tasks.

Mice raised without normal gut bacteria show autistic-like behavior, and giving them normal gut bacteria reverses many of the symptoms. Extremely interesting because prevalence of autism keeps increasing and no one knows why, and the amount of stress we put our gut flora under with weird diets and antibiotics also keeps increasing and would be really neat if we could use the one to help explain the other. Bonus: this study was conducted by one of my professors during medical school who was sort-of-very-briefly my research advisor.

Adjusting the level of the gut bacteria a. muciniphilia causes a bunch of downstream changes in the guts of mice which eventually affect their obesity levels. But please, tell us again how all obesity is 100% due to how much you eat minus how much you exercise.

Elsewhere in the gut, children born after their mothers have bariatric surgery are less likely to be obese and have different gene methylation patterns than children born just before their mothers have bariatric surgery – suggesting that fetuses may be "programmed" to metabolize fat differently based on their mother's health. BUT IT'S STILL ALL JUST EATING AND EXERCISE LA LA LA I CAN'T HEAR YOU.

In my Biodeterminists' Guide, I suggested the then unproven conjecture that First World iodine deficiency was probably having small but measurable effects on kids' intelligence. Last month, research showed this was true in the UK.

Old study: Medicare costs vary wildly around the country because doctors are basing their treatment on local fads rather than sound practice. New study: Medicare costs vary wildly around the country because health varies wildly around the country. But be sure to read the responses and counter-responses.

Ketamine has long been able to treat depression instantaneously and miraculously. The only problem is that, well, it's ketamine. Now scientists announce the discovery of a molecule that may be able to replicate ketamine's antidepressant effects without its side effects.

I'd previously seen hypothermia caps used to fight insomnia, but apparently the real medical interest is in a different direction: hypothermia caps can allow chemotherapy patients to keep their hair. The mechanism is actually very clever: you wear the cap, the cold contracts the blood vessels in your scalp, and the chemotherapy drugs don’t reach the hair follicles.

A study finds that cruise ships cost about the same as assisted-living facilities for the elderly, provide about the same range of services, and are much more fun. It suggests that more seniors should go on permanent cruises instead of “off to a home”.

Meta-science strikes again: peer review is basically useless. Researchers find that peer reviewers will reject articles they previously accepted but forgot about at around the same rate they will reject articles in general – at least once the names of the prestigious scientists involved are stripped off.

Genetic risk for schizophrenia is associated with decreased IQ even in those who don’t develop the disease (I can’t find by how much). This is interesting for two reasons. First, the correlation was observed at age 70 but not at age 11, which is yet another reminder that a lot of these IQ studies just don’t work all that well on kids who are still growing and having their IQ fluctuate for other reasons. Second, in light of the recent total failure to figure out IQ genes, it’s nice to sort of have something to work with.

75% of lesbians are obese, which is much larger than comparable numbers for straight women or gay men. A new study is trying to find out why. It would be mildly interesting if social, super interesting if biological as it might give us more insight into what’s going on hormone-wise.

All research relating to oxytocin is existentially terrifying, and the Karolinksa Institute’s work on oxytocin and pair-bonding is no exception. The study finds that the 25% of people with a certain allele of an oxytocin receptor score significantly differently on measures of relationship status, a discovery the media has not entirely unfairly spun as a "divorce gene" that makes you 50% more likely to break up with your partner. This is the first time I've really on a gut-level appreciated how scary the age of univerally available genetic information is going to be. Do people have a right to ask their prospective partner what version of the oxytocin receptor they have? (also worth checking out: the other divorce studies mentioned in that second link)

Just as religious people are more likely to believe their religion in times of stress, so people are more likely to believe in the power of capital-s Science in stressful situations. This raises some interesting possibilities for increasing the likelihood that the public will take scientific advice seriously, like preceding announcements of important studies that people should take seriously with pictures of a giant spider climbing on a man's back.

Biodjinnetics

“In exchange for freeing me from this lamp, you may ask me one question,” said the genie.

“Not three?” I protested.

“Just one,” said the genie.

“What is the cure for cancer?” I asked.

“A compound called oxymercuriphine, found in the venom of the two-toed toad of Toronto,” said the genie. “Cures 100% of all cancers.”

“Huh,” I said. “Thank you. And here I was worrying you were one of those evil genies who would satisfy the letter of the wish while actually being totally useless.”

“Me?” said the genie, as he faded from view. “Never. That molecule is the genuine article and I wish you only the best of luck getting the fifty million dollars or so needed to push it through years of FDA clinical trials.”

Arguments From My Opponent Believes Something

1. **Argument From My Opponent Believes Something, Which Is Kinda Like Believing It On Faith, Which Is Kinda Like Them Being A Religion**:
"The high priests of the economic orthodoxy take it on faith that anyone who doubts the market is a heretic who must be punished."

2. **Argument From My Opponent Believes Something, Which Means They Believe It Is The Answer To One Question, Which Is Kinda Like Believing It Is The Answer To All Questions, But It Isn't**: "Statists believe government can solve all our problems. They need to understand the world doesn't work that way."

3. **Argument From My Opponent Believes Something, Which Is Kinda Like Believing It Really Strongly, Which Is Kinda Like Being A Fanatic**: "Environmentalist extremists are fanatically obsessed over saving the planet, refusing to even consider any contradictory ideas."

4. **Argument From My Opponent Believes Something, Which Is Kinda Like Believing It Blindly With 100% Certainty**:
"Some people blindly trust science to always be correct about everything, but we need to remember that even scientists can make mistakes."

5. **Argument From My Opponent Believes Something, Which Is Kinda Like Having An Ideology, Which Means They Are Ideologues**: "Ideologies are false idols, attempts to replace thought with mindless obedience. And one such

ideology is the dogma of feminism. Therefore, we need to start being much more critical about feminism."

6. **Argument From My Opponent Believes Something, Which Is Kinda Like Hating The People Who Don't Believe In It, And Hatred Is Wrong**: "People need to get over their frothing hatred for euthanasia."

7. **Argument From My Opponent Believes Something, Which Is Kinda Like Saying That That One Belief Should Be The Sole Determinant Of Our Entire Aesthetic Sensibility**: "Sure, we could legalize contraception. But do we really want to enshrine the value that human fertility is evil, and that new human life is a 'failure' to be avoided?"

8. **Argument From My Opponent Believes Something, Which Might Suggest A Course Of Action, But A Suggestion Is Kinda Like An Obligation, And She Has No Right To Order Me Around**: "Some people want to liberalize immigration laws, but our country is under no obligation to let in any foreigner who asks."

9. **Argument From My Opponent Believes Something, Which Might Suggest A Course Of Action, Which Could In Theory Be Implemented Through Violence, And Violence Is Wrong**: "Transhumanists think AI may be dangerous, but this could encourage people to kill AI researchers, so holding this belief is irresponsible." Or, "Environmentalist condemnations of the oil industry encourage eco-terrorist attacks on oil workers."

10. **Argument From My Opponent Believes Something, Which Might Suggest A Course Of Action, And Suggestions Could In Theory Stigmatize People Who Don't Do Them**: "People say smoking is dangerous and

unhealthy, but this just serves to stigmatize smokers and make them feel unwelcome in society."

For best effect, combine all ten as densely as possible:

> It is an unchallengeable orthodoxy that you should wear a coat if it is cold out. Day after day we hear shrill warnings from the high priests of this new religion practically seething with hatred for anyone who might possibly dare to go out without a winter coat on. But these ideologues don't realize that just wearing more jackets can't solve all of our society's problems. Here's a reality check – no one is under any obligation to put on any clothing they don't want to, and North Face and REI are not entitled to your hard-earned money. All that these increasingly strident claims about jackets do is shame underprivileged people who can't afford jackets, suggesting them as legitimate targets for violence. In conclusion, do we really want to say that people should be judged by the clothes they wear? Or can we accept the unjacketed human body to be potentially just as beautiful as someone bundled beneath ten layers of coats?

EDIT: *I'm not claiming these aren't real problems, I'm claiming they're things that they are fully general arguments – you can accuse anyone of them and no one can ever prove you're wrong. For example, some things really are religions (Christianity, for example), but you can accuse any position of being "a religion" merely by virtue of it being a belief that people hold. Therefore, we should be extremely skeptical of arguments where "X is a religion" is doing the work.*

The Virtue of Silence

Leah Libresco writes a couple of essays (1, 2) on an ethical dilemma reported in the New York Times. In the course of a confidential medical history, a doctor hears her patient is suffering from stress-related complaints after having sent an innocent man to prison. The doctor wants to know whether it is ethical to report the matter to the police. The Times' columnist says yes – it would save the poor prisoner. Leah says no – violating medical confidentiality creates an expectation that medical confidentiality will be violated in the future, thus dooming patients who are too afraid to talk about drug use or gay sex or other potentially embarrassing but important medical risk factors.

But both sides are ignoring the much bigger dilemma lurking one meta-level up: is it ethical to debate this dilemma in the *New York Times*?

Let's look more closely at that phrase "violating medical confidentiality creates an expectation that medical confidentiality will be violated in the future." There's a very abstruse angels-and-clockwork interpretation of "creates an expectation" where, by making the decision to violate confidentiality, you are altering the Platonic machinery of the Universe in a way that allows other beings who know your source code to determine that you will do this. But most people don't have the decision theory to understand this, and anyway most doctors do not publish their source code online.

The way "creates an expectation" pans out in *our* universe is that somebody hears that a doctor violated medical confidentiality, and that person tells someone else, and that person tells someone else, until eventually someone who was going to tell their doctor about having gay sex with drugs remembers having heard the story and decides not to.

How exactly would people hear about this doctor who revealed the innocence of the prisoner? Through the ensuing court case? Nah. Most people wouldn't obsessively read the minutes of every single case at the local courthouse unless of course it has something to do with gender. Really, the only way that someone could hear about a doctor violating medical confidentiality is if she, like, somehow got a description of her intention to do so published in meticulous detail in the *New York Times*.

Oh, *right*.

The entire negative effect of the doctor breaking her promise is that it would make people doubt medical confidentiality in the future. But *whether or not the doctor ends up breaking her promise*, thousands of New York Times readers now know that doctors strongly consider breaking medical confidentiality, and that ethics columnists tell them it's okay to do so. It seems like the whether the doctor actually keeps her promise or not in this particular case is of miniscule importance compared to the damage that the column has already done.

Silence is a *hard* virtue. All the other virtues have the advantage that, when you practice them, people will praise you. Sometimes if your moral system is very different from your friends' people will attack you for your virtues, but getting attacked by sufficiently horrible people can sometimes be just as gratifying as praise. But if you stay silent, there's no praise *and* no attacks. By definition, no one even knows you made a courageous moral choice.

(Eliezer mentioned in the comments of my Against Bravery Debates that he's spent a couple decades pushing ideas almost everyone else thinks are crackpot, and he's never appealed to bravery at all. He is one hundred percent correct and I have one hundred percent never noticed despite reading almost everything he's written for several years. That's the Virtue of Silence for you.)

(I had like five much better examples here, all of which would be very clever, and each time I had to catch myself and say "Wait a second, by bringing that up I'm violating the virtue I'm supposed to be pushing here, aren't I?")

One example of silence I deeply appreciate is *people who don't talk about the latest viral issue*. I'm trying to think of an example that's not too destructive to bring up… hmmmm…go for something old…Elevatorgate! Nearly everyone who talked about Elevatorgate mentioned that it was outrageous that the blogosphere was making such a big deal about it, missing the similarity to the old adage that "you aren't stuck in traffic, you *are* traffic." Somewhere there was someone who wanted to write about Elevatorgate, thought about it, and *decided not to*. That person deserves the sincere thanks of a grateful Internet.

So having made the case for the other side of the confidentiality-newspaper meta-dilemma, am I *actually* pushing the claim that it is a moral law not to publicize information that could have bad consequences?

But I notice that this sort of thing almost always ends up making people angry and having a perverse effect where demands not to draw Mohammed turn into Everyone Draw Mohammed Day (see: Streisand Effect). It also sometimes snowballs to the point where not only can you not talk about X, but you can't talk about the demand not to talk about X because that would be referring to X obliquely, and you can’t talk about the demand not to talk about the demand to talk about X, until eventually you climb up so many meta-levels that you collapse from hypoxia and have to be rescued by Sherpas. Then you get a “callout culture” where people try to gain easy Virtue Points by telling people discussing issues that they Should Not Be Discussing Them and other people try to gain easy Virtue Points by being the Brave Defender of Freedom of Speech.

And maybe that’s useful if it’s something like gender where everyone wants to talk about it all the time anyway, but we don’t really need to do that to medical confidentiality, do we?

Maybe this is one of those rare cases where the word “supererogatory” might be useful. Yelling at people who talk about violations of medical confidentiality would just lead to “ARE OUR DISCUSSIONS OF MEDICAL CONFIDENTIALITY BEING SILENCED??!?” on the front page of the *New York Times*. And fretting over talking about it with your friends, or publishing a blog article about it (cough) is probably on the moral level of those Jains who walk everywhere with a broom in front of them so that they don’t accidentally squash any bugs. But if someone is really really concerned about it and wants to be a great person, then yeah, I think writing to the *New York Times* about it requires a bit of thought.

And since I *am* publishing a blog article about it (VIRTUE OF SILENCE IS REALLY HARD!) let me restore some Virtue Points by confirming that I will *not* betray private patient information of this sort if such a dilemma comes up except when legally required. Trust me, I’m a doctor.

Can Atheists Appreciate Chesterton?

Empirically, yes.

Friday was the anniversary of Chesterton's death, the religious blogosphere is eulogizing him, and I thought I'd join in. I enjoyed and recommend Chesterton's novels, especially The Man Who Was Thursday and Napoleon of Notting Hill, his works of nonfiction like Heretics, and even his poems (all of these are links to freely available fulltext versions online).

Classical philosophy holds that evil is merely the absence of good, but for me, at least, the opposite reduction is more tempting (albeit just as wrong). Evil is extremely obvious – you can look at people involved in animal cruelty, or bullying, or whatever, and you can almost *see* the actively malicious force animating them onward. On the other hand, good is most easily perceived as unusual skill at avoiding evil. Vegetarians are unusually good because they take extra effort to avoid hurting animals, people who donate to charity are unusually good because they take extra effort to avoid greed.

I credit three authors with giving me a visceral understanding of active, presence-rather-than-absence Good: G. K. Chesterton, C. S. Lewis, and Jacqueline Carey. Two of those are very religious and write quite consciously from a Christian perspective. The third writes about kinky sex. Go figure.

But actually when I think about it more closely, the moral beauty in Carey's writing comes mostly from her constructed religion, which is *suspiciously* similar to

Christianity. So it seems that there’s a fact to be explained here.

Can an atheist appreciate Chesterton? A better question might be whether an atheist can happily appreciate Chesterton as offering a beauty that she, too, can partake in, or whether the appreciation must be along the lines of “Yup, these are the nice things we can’t have.”

Keep The Horse Before The Cart

So I think an important point to make before going any further is that, through 90% of Christian history G. K. Chesterton and C. S. Lewis probably would have been burnt at the stake.

Not just for denominational reasons, although that would have been enough. Promoting joy as a sign of sanctity and as a proper state for man – that’s a burning for the Epicurean heresy right there. Believing righteous non-Christians could get into Heaven – that’s a burning. A suggestion that that humor and lightness were chief attributes of God and the angels – more burning. Doubting the literal truth of some of the Old Testament? Uncertainty whether the New Testament was divinely inspired in a more-than-metaphorical all-great-art-is-divinely-inspired way? Claims that praying sincerely to false gods was praiseworthy and basically just another way of praying to God? Burning, burning, *burning*.

The moral qualities that shine in Lewis and Chesterton – joy, humor, a love of the natural world, humanity, compassion, tolerance, willingness to engage with reason – are all qualities they inherited from modernity which would be repugnant to many of their Christian predecessors. They are all totally within the milieu of early 20th century England and totally foreign to medieval Italy or ancient Judea.

St. Augustine could not have written *The Great Divorce*, because while Lewis was talking about how the blessed in Heaven suffer great hardship to meet the damned in order to radiate love and wisdom at them and help bring them to Heaven, Augustine was writing about how the greatest pleasure of the blessed was getting to watch the tortures of the damned, metaphorically munching popcorn as they delighted in sinners getting what they deserved. Tertullian didn't even wait until after he died to start getting delighted, famously saying that:

> "At that greatest of all spectacles, that last and eternal judgment how shall I admire, how laugh, how rejoice, how exult, when I behold so many proud monarchs groaning in the lowest abyss of darkness; so many magistrates liquefying in fiercer flames than they ever kindled against the Christians; so many sages and philosophers blushing in red-hot fires with their deluded pupils; so many tragedians more tuneful in the expression of their own sufferings; so many dancers tripping more nimbly from anguish then ever before from applause."

What Lewis, Augustine, and Tertullian had in common was Christianity; what set Lewis apart was modernity. What made C. S. Lewis saintly, as opposed to the horrifying sadists who actually got the "St." in front of their names, was the perspective of a culture that had just spent a few centuries thinking about morals from a humanistic perspective.

When Pope Francis said that we need to build a "culture of life" that can protect innocent children from harm, he wasn't taking a revelation from the Biblical angels but from the Better Angels Of Our Nature. The *Biblical* angels are the ones who would be tasked with enforcing God's promise of blessing on anyone who takes Babylonian infants and smashes them against rocks (Psalm 137:9, look it up).

During the tradition from the Dark Ages to modernity, people got technologies like the printing press and the frigate and started learning more about other cultures, seeing that they were decent people and that no one religion had a monopoly on morality. The decline in infectious diseases banished death from an everyday presence to a lurking evil and made casual slaughter seem less appealing; the gradual decline in war resensitized people to violence. And all this time there were philosophers inventing things like deontology and consequentialism and freedom and equality and humanism and saying that yes, people did have inherent moral worth. And religion eventually decided that if it couldn't beat them it might as well join them, at least to a degree, and it was this concession that allowed the moral decrepitude of people like Tertullian and Torquemada to evolve into the moral genius of people like Chesterton and Lewis.

So my thesis is that Lewis and Chesterton didn't become brilliant moralists by revealing the truths of Christianity to a degraded modern world. They became great moralists by taking the better parts of the modern world, dressing them up in Christian clothing, and handing them back to the modern world, all while denouncing the worse parts of the modern world as "the modern world".

And so rah humanism and all that. But the original question remains: what is it about the Christian clothing that is such a necessary ingredient?

A Cupboard Full Of Secret Ingredients

First of all, the power of myth.

I don’t think it’s a coincidence that all three of the people I named as influences on my sense of moral beauty were writers of speculative fiction. Fiction has greater opportunity to be beautiful and to show complicated internal dynamics of humanity than abstruse philosophy or dry preaching does, and speculative fiction has a better opportunity to present superstimuli, including moral superstimuli. I think that people who write speculative fiction ordinarily tend to be kind of dismissed, but that because Lewis and Chesterton were working from within a tradition that had its own myths, they managed to get through the filter of “Oh, it’s just fantasy, ignore it”. Narnia was dignified by being a metaphor for the Bible, which earned its dignity through hoary age and civilizational influence.

Second of all, legitimacy.

I sometimes write about morality. It tends to be in a light-hearted “here’s what I think” style, first of all because I’m genuinely uncertain about a lot of stuff, second of all because I don’t want to sound preachy. Religion is really good at helping people be certain of things, and religious people get a free pass to sound preachy because preaching is what religions are *supposed* to do.

I don’t think there’s a niche for non-religious versions of Chesterton and Lewis. There are people like that New York Times ethics columnist who talk about ethics, but I think if they were to start getting *poetic* about it, people would start challenging their right, be like “Who told *you* what is or isn’t necessary for the integrity of the human spirit?” This is a tough question. But Lewis and Chesterton have a great answer: “God did”. They can, as the Bible puts it, “speak like one who has confidence”.

Third of all, a different perspective.

You can seem deep just by saying something different than everyone else does. I don’t think Lewis and Chesterton were too far from the modern moral mainstream, but I think they use a completely different aesthetic. Where most people talk about the bravery of defying the mainstream, a Christian writer can talk about the bravery of *not* defying the mainstream when everyone thinks you should. Where most people talk about the importance of high self-esteem, a Christian writer can talk about taking care to avoid pride. Both sides have valid and important insights, but if a culture is doing everything it can to saturate you with one of them, the other will be a powerful breath of fresh air.

Chesterton – I haven’t yet noticed this in Lewis – has this sort of gambit where he agrees with some modern virtue, and then says the correct way to attain the modern virtue is through doing the opposite of the modern virtue. Or maybe the opposite, where he agrees with what we should be doing, but then says the end goal is exactly the opposite of what everyone would think:

> The outer ring of Christianity is a rigid guard of ethical abnegations and professional priests; but inside that inhuman guard you will find the old human life dancing like children, and drinking wine like men; for Christianity is the only frame for pagan freedom.

People make fun of this, and rightly so (Steven Kaas attributes to Chesterton’s dog the quote “Arf arf arf! Not because arf arf! But exactly because arf NOT arf!”) but I think it is fundamental to his project. He gets to maintain his belief in modern virtues while getting there through an unexpected path that seems deep and profound and unexpected.

Fourth of all, a focus on the individual.

Despite everything everyone says about modern society being too individualistic, there seems to be a sense in which the opposite is true. The problems we are comfortable talking about are ones like racism, sexism, income inequality, terrorism, crime. Social problems. Problems in the community. The idea of talking about what goes on in the individual soul, of having strong opinions about it, isn’t a very modern sensibility at all. The only exception are psychologists and therapists, who really want to be scientific and so scrupulously avoid sounding poetic.

I could come up with some just-so stories about why this is – we like to think scientifically, but intrapersonal dilemmas don’t lend themselves to this kind of analysis? Focus on individuals doesn’t generalize well, which is a problem in the age of mass media? Christians were abnormally obsessed with the individual soul because of virtue ethics + the idea of damnation and salvation? I’m not sure. Anyway, religion has a head start on individualist vocabulary and thought processes which non-religion doesn’t really have good alternatives for (PSYCHODYNAMICS DOES NOT COUNT AS A GOOD ALTERNATIVE).

All of these are kind of banal and not the sort of thing that could prevent an atheist from fully appreciating Chesterton. But then there’s the big one.

What Lewis, Chesterton, and Carey have in common is this belief in Good as an active, vibrant, force, in Good being not just powerful, but so powerful that it’s kind of terrifying. As something not just real, but *the most real* thing.

Atheists can have Good be terrifying – utilitarianism has broken much stronger minds than my own – but it’s really hard to have it be *real*. I’m not saying atheists can’t believe in Good, just that atheist good is a sort of – I hate this term but I’ll use it anyway – social construct. It’s real in the same sense the US Government is real. The US Government is certainly powerful – just ask any Iraqi. But it’s not *one thing*, with an essence and a personality and angel wings of red-white-and-blue fire. It’s just an abstraction over a lot of ordinary people doing their thing.

And this would seem to be the death blow for atheists having something as strong and convincing as a Lewisian or Chestertonian world-view. Except that I kind of picked up a similar vibe from *Harry Potter and the Methods of Rationality*. I didn’t think of it when I was naming the three authors who first made me think of Good as a thing, but it is another work that portrays Good as this burning, all-powerful force, and although it has some magic in it, it doesn’t go all the way to reinventing Christianity like Carey does.

I’m not sure whether this is sleight-of-pen, whether it only works because of the magic there because even if the magic and morality aren’t explicitly linked it still triggers sort of morality-is-magic circuits. Or whether it only works if you’re literally responsible for saving the world. But it seems encouraging.

I think the truth of Lewis and Chesterton is not only appreciatable by atheists but derives from humanist ideas. The *beauty* of Lewis and Chesterton I’m not sure about, but I maintain some hope that it can be saved as well, even if I’m not sure how to do it.

The What-You'd-Implicitly-Heard-Before Telling Thing

G. K. Chesterton, whom I praised yesterday, is also famous for the argument of the "truth-telling thing":

> "This, therefore, is, in conclusion, my reason for accepting the religion and not merely the scattered and secular truths out of the religion. I do it because the thing has not merely told this truth or that truth, but has revealed itself as a truth-telling thing. All other philosophies say the things that plainly seem to be true; only this philosophy has again and again said the thing that does not seem to be true, but is true. Alone of all creeds it is convincing where it is not attractive"

Forgive me if I go into Angry Internet Atheist Mode for a second, but…

Yes, it's told so many truths. Like that God created the world in seven days. And that there was a giant flood. And that it's morally acceptable to condemn people to eternal torture. And that homosexuality is wrong. And that slaves should submit to their masters. And women to their husbands. And that the Second Coming will occur before the last of this generation passes away. And that people who are capable of doing so should castrate themselves. And that you should not suffer a witch to live. And that epilepsy is sometimes caused by demons. And that it's a really really good idea to kill Babylonian children.

And that everything is a combination of essence and accidents. And that things have final and formal causes. And that the planets are arranged in a succession of crystalline spheres, each with a governing angel. And that capitalism is a terrible idea. ~~And that church councils like the one that killed Jan Huss are infallible.~~

And after you’ve subtracted all the things that, in the light of modernity, obviously the Bible couldn’t have *actually* mean or obviously couldn’t *really* have been Biblically supported, what are you left with? Ideas like “humanity is flawed”. Gee, thanks religion. Surely only *God* could have noticed this startling and well-concealed insight!

When religion makes non-trivial testable claims, whether in its holy books or from later clergy trying to interpret those holy books, those claims have a spectacular record of being exactly as wrong as you would expect from chance – and then some. So what the stars are the “truth telling thing” people talking about?

Let’s ask one. From Unequally Yoked:

> In some ways, I find myself in a similar position to Chesterton. I find that a lot of Christian theology works for me in a way that plenty of other philosophies have not. When I say ‘works’ I mean pretty much what Chesterton does—that it matches many of the core assumptions I make about the world, and it harmonizes some of the conflicting ones in ways I didn’t expect, but seem to fit.

Time for a metaphor here, and I’m going to sort of steal it from Alasdair MacIntyre.

Suppose that one of the Roman civil wars – let’s say the one precipitating the Year of Four Emperors – goes on for decades and turns into an apocalypse. Roman civilization and learning are destroyed. All the Romans have are fragments of their old culture. Something something Mt. Olympus. Some kind of apple thrown at some kind of party caused the fall of Troy. There are these books in a cave that tell the future, who knows how they got there? We should avoid hubris, but we don't know why.

Then someone from a very distant colony arrives, bearing intact copies of the Iliad and the Odyssey and a few other important books that reveal all the tenets of Greco-Roman paganism.

Suddenly, everything makes sense! The reason we go on pilgrimages to Mt. Olympus is because *the gods live there*. The reason an apple caused the fall of Troy was *because it was thrown by the Goddess of Discord*! The reason these books in a cave tell the future was *because they were written by the Sibyl, who gained the gift of prophecy after a love affair with Apollo*. We should avoid hubris because *Jupiter is jealous and will zap us with lightning bolts*.

On the other hand, the Iliad and Odyssey would continue to be laughably wrong about all testable claims, like that the ocean is perfectly circular or that there's an island inhabited by Cyclopses.

Because Roman religion was originally shaped by the Iliad and Odyssey and then fractured into confusing fragments, restoring exposure to the source of the religion will cause this feeling of "suddenly everything I believe fits together and makes sense." But none of this subjective feeling of sense-making will correspond to ability to make correct claims about the external world.

Modern Western civilization spent about fifteen hundred years having its thought processes completely shaped by Christian doctrine. Over the past few centuries, changes in science and philosophy have shattered a lot of Christian doctrine and replaced it with more modernist ideas, but they haven’t succeeded completely and certainly not at the deepest level. Most people contain various strata of conflicting Christian and modernist ideas superimposed upon one another, and not all the Christian ideas are conveniently labeled “Christian”.

Exposure to the Christian ideas in their original form should allow a lot of aspects of modern culture to be viewed in a new light. To give a trivial example, dislike of homosexuality is pretty common in our culture, but has zero intellectual foundation outside of an ethical system that people generally aren’t exposed to unless they specifically study Christian philosophy. Less trivial examples might be beliefs about guilt, penance, justice, innocence, marriage, modesty, humility, etc, etc, etc.

If this were the whole picture, then things could go one of two ways. People could be exposed to really high-grade modern philosophy that removes the remaining Christian elements (like makes the consequentialist argument against stigmatizing homosexuality), suddenly have a revelation of beauty and consistency, and become full-on atheists. Or people could be exposed to the purest form of Ye Olde Time Religion, suddenly have a revelation of beauty and consistency, and become full-on religious people.

Buuuuut it’s more complicated than that because I think the modernist beliefs and the religious beliefs are held in different ways, although don’t ask me to get more technical than that. Maybe the modernist beliefs are held explicitly and endorsed? And the religious beliefs are held kind of subconsciously as aliefs? And so I think the high-grade modernist philosophy and the Ye Olde Time Religion are appealing in different ways and to different parts of our belief structure.

This applies not just to Christianity but to any claim that old ideas should be taken seriously because they match our intuitions and aesthetics. Reactionaries, I’m looking at you here.

A New Chapter In The Codex

For the past year, I have been living off of some very occasional part-time work for MetaMed, some odd jobs, and savings from various sources. It hasn't been the most lucrative thing in the world, but it's left me lots of time for thinking and writing.

Tomorrow I start Real Work. It's just training for the first couple of weeks and I may have a bit of time, but after that I have been told it is *not* the sort of thing that leaves much time for thinking and writing (or eating, or sleeping). I am *not* going to keep up blogging once per day. *Maybe* I will be able to blog once a week? A couple of times a month?

Because I will be working on medicine and psychiatry full time, I expect this to turn into more of a Medicine And Psychiatry Blog than it is right now. I might change the About page and the tagline and some other stuff in order to reflect that.

Since I won't have time to write complicated logorrheic essays anymore, I've gotten a Twitter account for my less verbose thoughts. If the past few days are any indication, it will be made entirely of corny puns and have no redeeming intellectual value. And another warning: if you follow me I probably won't follow you back for a while – yesterday Twitter accused me of following too many new people and said it would delete my account unless I stopped.

Finally, I need to make a transaction in bitcoins soon but don't want to expend the time/effort to become a part of the bitcoin economy right now. If anyone has bitcoins and wants to buy me something in exchange for the equivalent amount of fiat currency over PayPal, give me your email in the comments and we’ll make a deal (I’m happy to pay you slightly extra for your time).

Social Psychology Is A Flamethrower

Mark Twain:

> There is something fascinating about science. One gets such wholesale returns of conjecture out of such a trifling investment of fact.

If this is true of all science, it is doubly true of social psychology.

At its best, social psychology is an unmatched window into human motivations, a “look under the hood” of the way people talk and act. The best research in social psychology is as well-supported as anything in physics or biology, and much more intuitively comprehensible. This is why it’s one of my favorite scientific fields.

But at its worst, social psychology is a flamethrower. People grab hold of it to try to fry their political opponents, then end up lighting their own hair on fire or burning down half a city. Because social psych is *really hard* to do right.

Social psychology experiments in the laboratory tend to throw up spectacular mind-boggling effects. Many of these fail to replicate and are later discredited. The ones that do replicate are not always generalizable – sometimes an even *slightly* different situation will remove the effect or create exactly the opposite effect. The effects that remain robust in the laboratory may be too short-lasting or too specific to have any importance in real life. And the ones that do matter in real life may respond unpredictably or even paradoxically to attempts to control them.

This is relevant because a lot of our political discourse revolves around ideas lifted from social psychology. Every time someone advocates banning violent videogames so that they don't normalize violence, they're using social psych. Anyone who says the media needs more positive role models of minority groups and fewer stereotypes, they're taking terms out of the social psych lexicon. Whenever you complain that magazines objectify women, you're implicitly buying into several social psych theories.

Most people are not consequentialists, but most people feel implicitly uncomfortable making moral arguments on non-consequentialist grounds. "Stop what you're doing, it disgusts and offends me" is less noble than "stop what you're doing, it will hurt people who can't stand up for themselves". This tempts people who are disgusted and offended by things to come up with just-so stories from social psychology for why the disgusting and offensive thing will also hurt people.

I tried writing a post arguing against several of these just-so stories, but it ended up being *unbearably* long and boring (if you're ever stuck with insomnia, ask me to give you a trenchant analysis of every study that's ever been written about stereotype threat). So I'm going to try something different. I'm going to write up some just-so stories using social psychology for the opposite side. I'm going to try to use well-established social psych results to prove that we should have *more* violence in the media, and be *more* tolerant of offending women and minorities.

I think some of the arguments below will be completely correct, others correct only in certain senses and situations, and still others intriguing but wrong. I think that modern pop social psychology probably contains the same three categories in about the same breakdown, so I don't feel too bad about this.

Violence In The Media Prevents Violent Crime

Dahl and DellaVigna (2008), well aware of laboratory experiments that found violent media temporarily made subjects more violent, decided to investigate whether the opening weekends of blockbuster violent movies affected crime rates. Sure enough, they found they did…

…in the opposite of the expected direction. They found violent movies decreased crime 5% or more on their opening weekends, and that each violent movie that comes out probably prevents about 1000 assaults. Further, there's no displacement effect – the missing crimes don't pop back the following week, they simply never occur.

They hypothesize that every hour violent criminals are at the kind of movies that appeal to violent criminals is one hour more they're not getting drunk or taking drugs or committing violent crimes. Although they don't mention it directly, other analyses have suggested that the movies have a sort of cathartic effect, satisfying their urge for violence without them having to commit it themselves.

An investigation into violent video games found essentially the same pattern: violent video games decrease crime while nonviolent video games have no effect.

There are also studies that show that playing lots of violent video games is correlated with violent criminality, but a much more plausible explanation of the data is that a naturally violent personality makes people more likely to enjoy violence both in games and in real life.

Decreasing violence in the media might therefore be predicted to increase violent crime, both by putting more criminals out on the streets and by sabotaging their attempts to indulge their violent urges in an acceptable manner.

Media That Objectifies Women Prevents Rape

Just as violent movies prevent violent crime, pornography may prevent rape. It’s easy to prove that in the US every 10% increase in Internet access causes a 7.3% decline in rape, and it’s not due to any of the expected confounders. Another study points out a similar correlation in Japan. I find the particular correlation they mention very sketchy, but Japan does have a very low rate of reported sex crimes (a commenter brings up the possibility that Japanese culture merely discourages reporting). Other more rigorous studies on the Czech Republic show the same, and studies on child porn show pedophilia is less common where it’s more accessible. And these studies links to more interesting results, mentioning how sex criminals are less likely to consume pornography than the general population and start watching pornography at a later age.

This is explicable not only by the substitution effect mentioned above, but by the general tendency of orgasm to relieve frustration. If, as has been hypothesized, rape is an expression of anger and powerlessness at the world in general or women in particular, orgasming to violent porn is going to both satisfy that aggressive impulsive and replace it with general post-coital relaxation.

Saying Tests Are Biased Against Minorities Makes Minorities Perform Worse On Tests

It is relatively clear that achievement gaps on standardized tests – black-white, male-female, and the others – are not due to bias in the tests themselves. Although some sociologists raise the specter of “tests that claim to be fair by asking both rich and poor people the same questions about golf and yachting”, in real life achievement gaps remain mostly consistent across verbal tests, pure mathematical tests, symbol manipulation tests, and extremely basic and un-bias-able tests like ability to remember numbers backwards.

This has not stopped the constant repetition that various specific tests – SAT, GRE, IQ – are biased against minorities.

We know exactly what happens when minorities are told tests are biased against them: they do worse on those tests. This is the essence of the idea of “stereotype threat” – for example, one can improve women’s performance on a math test simply by telling them that the test is not biased against women. So maybe we should stop doing exactly the thing that we just proved hurts women and minorities’ educational performance.

Fighting Stereotypes Makes People More Prejudiced

The largest-ever study on diversity training, following 830 large companies over 31 years, found:

> A comprehensive review of 31 years of data from 830 mid-size to large U.S. workplaces found that the kind of diversity training exercises offered at most firms were followed by a 7.5 percent drop in the number of women in management. The number of black, female managers fell by 10 percent, and the number of black men in top positions fell by 12 percent. Similar effects were seen for Latinos and Asians.

Similarly, all studies on sensitivity training find that trainees express more *awareness* of sexual harassment than non-employees, but a study that went further and examined results found that trainees are “less likely to perceive coercive sexual harassment, less willing to report sexual harassment, and more likely to blame the victim”.

This is not particularly unexpected: we know for example that nearly every study on DARE programs has found that they increase drug use, sometimes as much as 30%.

Why should this be? Three reasons come to mind. The first is a boomerang effect from the programs themselves. Diversity training, sensitivity training, and DARE are all things busy people are required to attend where they (essentially) are forced listen to people behave condescendingly to them. This makes them dislike the training, their instructors, and, by association, the opinions they are trying to get trained into them.

A second reason is more fundamental. The backfire effect is when people challenged with information that disproves a cherished political belief of theirs react by becoming even more certain of the belief. The link will fill you in on potential explanations.

And the third reason is what the Harvard Business Review Blog, in its discussion of the diversity training study above, described as “when people divide into categories to illustrate the idea of diversity, it reinforces the idea of the categories.”

I’ll admit I had a sheltered upbringing and may be atypical, but I would estimate about 90% of the racist stereotypes I have ever heard were part of efforts to fight racism. No one just comes up to you and says “Hey, you know black people? Pretty unintelligent, huh?” (at least not to *me*). But social justice people will repeat the stereotype about black people not being intelligent again, and again, and again, to anyone who is anywhere near them, in the guise of fighting it.

I can’t find the link for this, but negatively phrased information can sometimes reinforce the positive version of that information. For example, if you tell people “President Obama is not a Muslim”, then a year later, all someone will remember is “blah Obama blah blah blah Muslim”, and eventually “Ohmigod, President Obama is a Muslim!”, even if they didn’t believe that before they heard that fact “corrected”.

Imagine I told you “People from Comoros are *not* all homosexual! This is a damn lie, and anyone who says people from Comoros are homosexual is an insensitive jerk. Please join me in fighting the popular perception that everyone from Comoros is a flaming gay.

Go ahead, try to think of Comoros in *any context* other than an archipelago full of gay people now. I’ll wait. Take a whole lifetime, if you want. It won’t help. Ten years after this blog is deleted and this post is inaccessible except through archive.org, there will still be a couple dozen people who are convinced that everyone from Comoros is gay, because they “heard it somewhere”. At the very least, the idea of Comoros = homosexuality is now firmly implanted in your mind, and it will be impossible to meet a Comorosian without secretly evaluating her sexual orientation and then trying to stop yourself from doing it.

Now imagine instead of hearing this once, you heard it every day of your life.

Calling People Racist Makes Them More Racist

Foster & Misra (2013) is a jewel of a paper I stumbled across totally by chance.

They got a bunch of undergraduate students in romantic relationships and gave them a test that asked them some questions about infidelity – things like “is it unfaithful to fantasize about another girl/boy when you’re in a relationship?”. They pretended to grade the test, but in fact they ignored the test and gave fake feedback.

The control group was told that they had some of the highest faithfulness scores of anyone in the experiment, they must be really faithful, good job. The experimental group was told they had some of the lowest faithfulness scores of anyone in the experiment and that the test had pegged them as having an unfaithful personality type. Once again, all this feedback was fake and both groups got around the same average score.

Then they measured what they called “trivialization” in both groups – that is, they asked them questions about how important faithfulness was to them. Consistent with their theory, the people who were told they were faithful said faithfulness was extremely important, but the people who were told they were unfaithful “trivialized” the behavior – who cares about fidelity anyway, infidelity is *maybe* a minor mistake but it doesn’t really hurt anyone, people should really stop whining about infidelity all the time. To give you a feeling for the size of this effect, on a scale of one to seven, the faithful group rated the importance of being faithful at 5.4/7, and the unfaithful group rate the importance of being faithful at 2.9/7. In other words, by accusing them of being unfaithful, the experimenters had successfully gotten the participants to “trivialize” faithfulness.

The researchers theorized that this was the process called “cognitive dissonance”. Most people like themselves and want to continue to like themselves. If they are told that they, or their group, has a particular flaw, then instead of ceasing to like themselves it may be easier to just decide that flaw is not a big deal and they can have it while continuing to be the awesome people they secretly know they are.

Now not only do the experimental subjects here stop caring about being faithful, but everyone pushing a pro-fidelity line is a threat to their new identity. And *the subjects weren’t even really unfaithful to begin with!*

Modern political discourse tends to do a lot of things like say “All white people are racist” or all men are naturally prone to violence and potential rapists. Or it may take little things normal people do and tell them they are racist or creepy or rape-y or something because of it.

What this does is drive people into identifying with these negative labels. And instead of making them want to change their behavior to stop identifying with these labels, it may just make them think “Well, if *I* do it, then I guess it can’t be so bad.”

Talking About Rape Culture Causes Rape

There is a strong debate still going on about whether the death penalty decreases crime. But this hides a more settled question, which is whether punishment decreases crime at all. The relatively accepted answer is yes, it does.

Criminologists have tried to separate out the important of punishment into two aspects: severity and certainty. They have consistently found that the certainty of the punishment is more important than the severity – the most important factor in whether someone commits a crime is the likelihood she will be punished.

No criminal can see into the future to discover whether or not they will be punished; the only way certainty of punishment can influence crime is through public perception of certainty of punishment. That suggests that if you discover that an abominable crime has (contrary to popular perception) a very low chance of punishment, it would be an excellent time to practice the virtue of silence.

Or consider the claim that rape jokes cause rape. As I understand it, the claim goes that someone tells a rape joke, then everyone else laughs, no one protests or anything, and then potential rapists in the audience conclude that they are in a culture that considers rape acceptable.

You know what else could potentially cause people to think our culture considers rape acceptable? Writing and publicizing countless books and articles arguing elegantly and vehemently for the point that our culture considers rape acceptable. Seriously. If I were a demon from Hell, charged by my infernal masters with increasing rape as much as possible, I literally could not think of a better strategy than talking about rape culture all the time.

Getting angry at the rape jokes while enthusiastically taking part in the demonic campaign thing seems like (to mix metaphors) missing the mountain for the molehill.

Summary

In this post, I’ve give six social psychological just-so stories: media violence prevents crime, objectification of women prevents rape, accusations of test bias hurts minorities, fighting stereotypes makes people more prejudiced, calling people racist makes them more racist, and talking about rape culture increases rape.

These can be easily compared to six much more common social psychological just-so stories: media violence causes crime, objectification of women causes rape, accusations of minorities doing worse on tests for intrinsic reasons like their culture hurt minorities, fighting stereotypes makes people less prejudiced, calling people racist shames them out of their racism, and making rape jokes increases rape.

I don’t consider any of my six completely proven, just intriguing and intuitively plausible. And of course, there’s an element of concern-trolling in all of them.

But I don’t consider any of the second six completely proven either; again, they are merely intriguing and intuitively plausible. And they have their own element of being suspiciously congruent to the political beliefs of the people who push them, as if they’re trying to come up with consequentialist justifications for ideas they hold for other reasons.

Some will point to various studies conducted on one or another of them, but with very few exceptions all those studies have been poorly replicated investigations into the very-short-term (less than ten minutes) effect of laboratory interventions on proxy variables. These can be diametrically opposite their real social effects – for example, the laboratory experiments that experimental exposure to violence causes people to play contrived games in a more aggressive manner couldn’t catch that in the real world, violent movies decrease crime. And poorly replicated short-term laboratory interventions on proxy variables can prove nearly anything – see for example the recent controversy around whether the word “Florida” makes people walk more slowly.

The six stories above suggest some pretty radical and unpalatable action approaching social engineering. For example, the idea that research into test bias should be suppressed, even if it is scientifically rigorous, just because hearing about it might hurt women – seems pretty unfair (same with the idea that no one should be allowed to talk about rape culture) And it seems unreasonable to ask people to constantly watch their language around white people to avoid anything that sounds like accusing them of racism because that could have unpredictable negative effects on them down the line.

But the six traditional stories also suggest pretty radical and unpalatable action approaching social engineering. For example, the idea that research into gender differences should be suppressed, even if it is scientifically rigorous, because hearing about it might hurt women. Also unpopular is the idea of constantly having to watch your language around minorities to avoid anything that sounds like you’re saying something racist because that could have unpredictable negative effects down the line.

And my point is that I don’t see good enough evidence that the effects involved are real to justify either of them.

Using speculative extrapolations from social psychology to promote social engineering is dangerous and proves too much. Of course, one should still be *nice*, and a big part of niceness is judicious exercise of the virtue of silence . But trying to institute and enforce said virtue on a social level requires subtlety that I have not yet seen anyone involved show the slightest sign of possessing.

Post Scriptum

Think quick! What is your brain’s number one thought upon hearing “Comoros”?

The Hospital Orientation

The nun who gives us a big welcome to Our Lady Of An Undisclosed Location Hospital. The intermittent reminders that The Hospital Was Founded By Nuns, You Know. The lingering fear from my days in Ireland, when all of my friends had Nun Horror Stories and parents would scare disobedient children by telling them they better behave or *the nuns will come for them*. The distant memory of my father telling me that hospitals run by nuns are a thousand times more cut-throat and ruthless than hospitals run by supposedly efficient businessmen.

The Hospital Mission Statement. The descriptions of the long procedure, the countless doctor-hours of work, the innumerable committees, that went into drafting this mission statement. The firm assertion that this mission statement is a divinely inspired pipeline to the culture of this particular hospital, meditating upon which will tell us the sort of institution we have been hired by. The total indistinguishability from any other hospital or clinic mission statement I have ever heard, all of which go something like “We strive to provide comprehensive care to patients in a culture of excellence” and could have been devised by the average college student in three minutes.

The big Vietnamese guy from Internal Medicine sitting next to me with a nametag reading “Becca Silverman”. The halting attempt to be humorous – “That’s funny, you don’t look like a Becca”. The deadpan assertion “Becca is an ancient and traditional Vietnamese name.” The awkward, uncertain silence.

The discussion of the hospital’s Ethical Guidelines. The assurance that, although not all of us are Catholic, the hospital guidelines are rooted in Catholicism but not limited to them and we will all be able to appreciate their wisdom in a nice ecumenical way. The segue into the first of these guidelines, which is COMMUNION WITH THE CATHOLIC CHURCH.

The introduction to various executives and officers, a disproportionate number of whom used to work in the auto industry but left for the hospital industry when times got bad. The inspirational quotes from Henry Ford, who, if these executives and officers are to be believed, is some sort of credible expert in the art of hospital-running. The growing worry that someone is going to tell me my patients just need an oil change.

The educational movie about someone who drives drunk and then gets in an accident and dies. The explanation that the moral of the movie was “Don’t make bad decisions”. The feeling that this is probably one of those problems that could benefit from being taskified. The only good part of the movie, which was the totally unnecessary plot twist where the drunk driver turns out to (secretly!) be the son of the doctor who was supposed to treat him, leading the actor playing the doctor to shout “NOOOOOOOOOOOOOOOOO!” in a very Anakin-ish way for about five seconds straight.

The presentation on work-life balance at the end of a long ten-hour day. The feeling that every time a company makes employees stay overtime to attend a mandatory seminar on work-life balance, an angel does a facepalm. The request for audience suggestions on how to fit a sixty to seventy hour work week in with sleep, study, errands, and family. The comment from Becca: “That thing with the hourglass Hermione used in the Harry Potter movie?”

The inspirational speech about how we should let Jesus guide us in our medical practice. The suppressed comment that as far as I know, Jesus only encountered one psychiatric case, which he diagnosed as a demon and treated by casting it into a herd of pigs who then fell off a cliff. The suspicion that this is no longer considered Best Practice. The doubt over whether Medicare would even reimburse him for the pigs.

Nature Is Not A Slate. It's A Series Of Levers.

Last week I criticized pop social psychology while maintaining that social psychology itself was a pretty interesting window into human thought processes.

I was then handed a link to someone who apparently likes social psychology a lot less than I do.

You should read the whole thing, but here are the parts I'll be talking about. As you can see, it's kind of a conservative perspective saying social psychology is a liberal enterprise to deny human nature in favor of people being infinitely malleable based on their situation:

> Personally, I find it very hard to fathom the idiocy of Mischel's conclusion. It would mean that a person who others think of as for instance shy is really nothing of the kind. It just looks that way because we have only had chance to observe him or her in situations that elicits shyness. And if you think of yourself as shy you must be either plain wrong or stuck in a series of situations that by coincidence predisposes you to acting shy. This idea may sound like a joke, but the zeitgeist of the 1960s was left of sanity and lots of "intellectuals" believed Mischel the way they believed in Marx, Lenin and Mao.
>
> For that reason, social psychology became a major branch of psychology. After all, if it was all in the situation then this was the important field of research. Personality barely survived and its proponents, like Hans Eysenck and Arthur Jensen, were often dismissed as racists and right-wing lobbyists.

Unsurprisingly, it soon became evident that Mischel was wrong – there really was such a thing as a shy person. As traits became real again, personality psychology grew but at the same time social psychologists kept a grip on their dominant position by introducing interactionism, the study of both situations and personality. This way they blurred the line between the fields and managed to claim a lot of the newly available positions in personality research. But at heart they were never interactionists; they started out as situationists because of their political views and they have stayed that way ever since. To this day they rarely perform experiments in which personality measures are used. Their focus is very much on the situation. Look at the collective social psychology blog in the links to the right of this post – it's even called The Situationist, not The Interactionist. For most of these psychologists, interactionism was just a word with which to neutralize the enemy.

And today? Well, it's like the French say, the more something changes, the more it stays the same. This can be seen in a recent post in the above mentioned The Situationist (which is still interesting to follow because not all social psychology is crap) about Harvard Professor Francesca Gino's book Sidetracked, in which she describes how small things or situations derail our plans, intentions and even our morals. As an example she mentions an experiment in which she and psychologist Dan Ariely equipped participants with high-end sunglasses. Half of the participants were told that they were actually counterfeits, while the other half were told they were the real deal. The participants were then instructed to perform a mathematical task which left room for cheating. It turned out that 70 percent of those who thought they wore knock-off sunglasses cheated compared to 30 percent in the other group.

This may sound like compelling evidence for the power of the situation, but is it really? The participants were all young women rather than a representative sample. But more importantly, they were informed that they were participating in a psychological experiment and then told to wear counterfeit sunglasses. That’s pretty far from any kind of real life situation. It’s more like saying, “let’s play a game – you will be the bad guy.” It supports the idea that social psychology is, as someone put it, a list of how people behave in weird situations. Needless to say, Gino and Ariely didn’t use any personality measure since that would only distract attention from the power of the almighty Situation.

So if wearing fake sunglasses can make a person dishonest, how about the situation of being brought up by criminal parents? Now that should be a way more powerful situation. Psychologist Sarnoff Mednick and colleagues investigated this in the mid 1980s using data from over 14 thousand nonfamilial adoptions (in which the adoptive parents are unrelated to the child). They found that when both biological and adoptive parents had no criminal convictions the adopted child was eventually convicted in 13.5 percent of the cases, so that’s our baseline. When adoptive parents had convictions but biological parents had not, the number of convicted adoptees only rose very slightly to 14.7 percent. So fake sunglasses will have a profound effect on your honesty, but being brought up by criminals will only marginally elevate your risk of being convicted of a crime. That must be some sunglasses.

[…]

> [Social psychologists] just need to construct some even more artificial situations in order to deliver those results that will prove that Marx was right all along. And no outsiders need to concern themselves with exactly how they go about doing that.

Where to start, where to start?

First of all, it is too bad that Staffan finds the importance of situation hard to believe, but at least he is in good company. They do not call this problem the Fundamental Attribution Error because it is rare (or for that matter because it is a correct and tenable position). In experiments, people consistently overestimate the effect of personality and underestimate the effect of situation. Insofar as social psychologists are the people trying to correct that, they are doing God's work.

On the other hand, there are also personality psychologists, who study personality. They, too, are doing good work. Contrary to Staffan's assertions, they are an integral and well-beloved part of the field of psychology. Of the two personality psychologists Staffan claims were villainized and dismissed, both made the list of the fifty most eminent psychologists of the 20th century. Eysenck is the third most-cited psychologist of all time.

Jensen was indeed "dismissed as a racist and right-wing lobbyist", but this was less because he dared to study personality and more because he spent much of his time trying to prove black people were genetically less intelligent and because he took large amounts of money from right-wing organizations. This seems like the sort of case where "racist and right-wing lobbyist" might be a perfectly acceptable value-neutral description – although I agree

Jensen’s work, much of which was completely unrelated and brilliant, hasn’t gotten the recognition it deserves.

Because of the success of personality psychology, some of Staffan’s claims about social psych are a No True Scotsman argument. Personality psychology is considered a different field than social psychology. When Staffan notes that social psychologists accept the importance of personality but don’t study it, this is exactly symmetrical to personality psychologists accepting the importance of situation but not studying it. So getting upset at social psychologists for not studying personality is a lot like getting upset at cardiologists for not studying the liver.

The sunglasses experiment? Staffan correctly points out (as I did last week) that several similar experiments are under challenge, but those were on unconscious priming whereas the sunglasses experiment (from the little I read on it) seems to be about conscious priming, which is on firmer ground. To me it seems like exactly the sort of thing that might be true – in the extremely artificial conditions of the laboratory. This meshes with the point I made last week – that social psychology does a great job illuminating certain processes the brain goes through, but that we should be wary about assuming they have what doctors call “clinical significance”.

The criminal adoption experiment? Probably 100% correct. As anyone who’s read *The Nurture Assumption* (or my review thereof) knows, psychologists are *constantly* unable to find any effect of parents on their childrens’ personalities or actions. This is sufficiently distressing that most people refuse to believe it, but it keeps being confirmed again and again. Personality is 50% genetic and 50% some other factor which people have yet to illuminate but which definitely doesn’t involve upbringing (I hear Judith Rich

Harris' book *No Two Alike* purports to explain what this factor is, but I'm only halfway through it and can't comment).

(seriously, why do I have to spend so much time insisting to racists and eugenicists that *genes are seriously really important*? This should not be as big a niche in the blogosphere as it is!)

But basically, social psychology has discovered the correct fact that situation is more important than people think it is and personality is less important than people think it is in determining behavior, while not denying that both are pretty important. It correctly claims that priming can have very large short-term effects on unimportant decisions, and correctly notes that being raised by criminals has no effect on anyone's personality. So far I think it's doing pretty well, as long as, once again, you are skeptical about trying to do social engineering with its results outside the laboratory.

Now let's get to the part about Marx. For this we go to those experts on all things Marxist, More Right. In their latest post, Drew Summitt draws a distinction between what he considers a conservative view – that there is such a thing as human nature and that political systems need to take that into account – and a progressive view – that there is no such thing as human nature, people are infinitely malleable, and once we create some kind of utopia we can perfect mankind. The quote is his, the emphasis is mine:

> "He defends this proposition with the assertion that "the **conservative realist view of human imperfectability** and their commitment to ordered liberty **as rooted in nature**, custom, and prudence" can see great support in modern evolutionary theory because modern evolutionary theory contradicts what

> Thomas Sowell calls **the "unconstrained vision" that liberal intellectuals and theorists are tempted to hold**. In contrast to this Sowell sets up **a "constrained vision" of human anthropology that is limited in its capabilities by an intellect being the servant of the passions, the reality of sin, or boring genes telling our memes what they can and can't do**. In order to support the idea that evolutionary theory supports a conservative political vision Arnhart traces the foundations of human capability to nature, custom and prudence. The conservative hierarchy of nature, custom and prudence **is what constrains the idealist impulse for reason to govern and judge, and indeed seek to overthrow, custom and nature**
>
> […]
>
> But the Liberal Egalitarian Free-trade Technocrats also recognize human nature as being essential to political order. If you agree with Steven Pinker on the Humean Is in regard to human nature, that is, if you think he's got his facts right, you must have an independent account of the Humean Ought, because Pinker is a Liberal, though one of the deflated, Clintonian Liberals. This is a form of Liberalism not touched enough reactionary circles. Why is it the Daniel Dennett, Steven Pinker, Peter Singer and many other experts in contemporary studies in human nature self described liberals? They may reject modern queer theory, as Dawkins does, or they may think that the brain is not a Blank Slate, as Pinker does, but they don't consider these positions to be dangerous to Liberalism writ large or if they do they are terribly good at hiding it.

So he is wondering why, if liberalism is founded on the idea that human nature is an infinitely malleable blank slate, are the world’s greatest scientific experts in evolved human nature liberals?

To steal a delightful turn of phrase from Terry Eagleton, this is like wondering why, if Tony Blair is an octopus, he has only two arms.

Let me make an analogy to medicine. Unlike the brain, there is no debate on the “nature” of the heart – the literal blood-pumping heart, not the fuzzy emotional version. We know the heart is fully one hundred percent genetically programmed (minus a little morphogenetic variation), that it’s not malleable by schooling or brainwashing or being raised in a commune. It is a social engineer’s nightmare, a system founded entirely on human nature without the slightest wiggle room.

And yet doctors *routinely* make the heart do what they want. If they want to raise heart rate, they give a dose of epinephrine. If they want to lower heart rate, they give a dose of propranolol. If social planners could control the brain as easily as doctors control the heart, we’d already be living in a communist utopia.

The heart has an immutable nature, and that immutable nature is *to respond to different situations in highly predictable ways*.

The heart is neither a blank slate nor a fully inscribed slate. It’s not a slate at all. The heart is a series of levers. If you pull one lever, it will do one thing. If you pull another lever, it will do another thing. It is, paradoxically, hard-coded for malleability. It’s not infinitely malleable – there’s no drug you can inject to make the patient’s heart beat out the drum parts to Beatles songs – but you can shift it a little bit this way or that.

We have reason to believe the brain works the same way. Not everything is a lever – if you send a kid off to be raised by criminals, it won't activate any of the hard-coded IF-THEN statements, and nothing will happen. But if you surround someone by stimuli that prime the idea of criminality – whether sunglasses or a broken window, that will pull on a lever that will make criminal behavior a little bit more likely.

(except for "levers", read "extremely complicated things that run through chaos theory at some point and so are inherently unpredictable except in the broadest and most statistical sense")

All of this reminds me of a video I saw this afternoon on the second day of The Hospital Orientation. Please excuse me if I change it around just a little to turn it from a quality improvement case study to a morality tale.

There were two hospitals, Hospital A and Hospital B. Both, like all hospitals, were fighting a constant battle against medical errors – surgeons removing the wrong leg, doctors giving the wrong dose of medication, sleepy interns reading x-rays backwards, that kind of thing. These are deadly – they kill up to a hundred thousand people a year – and terrifyingly common.

Hospital A took a very right-wing approach to the issue. They got all their doctors together and told them that any doctor who made a minor medical error would get written up and any doctor who made a major medical error would be fired. Rah personal responsibility!

Unfortunately, when they evaluated the results of their policy they found they had exactly as many medical errors as before, except now people were trying to cover them up and they weren't being discovered until way too late.

Hospital B took a very progressive approach. They too got all their doctors together, but this time the hospital administrators announced: “You are not to blame for any medical errors. If medical errors occur, it means we, the administrators, have failed you by not creating a sufficiently good system. Please tell us if you commit any medical errors, and you won’t be punished, but we will scrutinize what we’re doing to see if we can make improvements.”

Then they made sweeping changes to what you might call the “society” of the hospital. They decreased doctor workload so physicians weren’t as harried. They shortened shifts to make sure everyone got at least eight hours of sleep a night. They switched from paper charts (where doctors write orders in notoriously hard-to-read handwriting) to electronic charts (where everything is typed up). They required everyone to draw up and use checklists. They even put propaganda posters over every sink reading “DID YOU WASH YOUR HANDS LONG ENOUGH??!” with a picture of a big eye on them. You can’t get more Orwellian than that.

And yet, *mirabile dictu*, this was the hospital that saw their medical error rates plummet.

The administrators of this second hospital didn’t ignore human nature. Instead, they exploited their knowledge of human nature to the fullest. They know it’s in human nature to do a bad job when you’re working on no sleep. They know it’s human nature to try to cut corners, but that people will run through checklists honestly and effectively. They even know that studies show that pictures of eyes make people behave more prosocially because they feel like they’re being watched.

You don’t have to tell me all the reasons this doesn’t directly apply to an entire country. I can think of most of them. But my point is that if I’m progressive – a label I am not entirely comfortable with but which people keep trying to pin on me – this is my progressivism. The idea of using knowledge of human nature to create a structure with a few clever little lever taps that encourage people to perform in effective and prosocial ways. It’s a lot less ambitious than “LET’S TOTALLY REMAKE EVERY ASPECT OF SOCIETY AS A UTOPIA”, but it’s a lot more practical.

(Although I’m also kinda okay with making every aspect of society a utopia, as long as we do it right.)

The Lottery of Fascinations

I.

Suppose I were to come out tomorrow as gay.

I have amazing and wonderful friends, and I certainly wouldn’t expect them to hate me forever or tell me to burn in Hell or anything like that.

But even more than that, I think they would understand and accept the decision. There would be a lot of not-so-obvious failure modes they *could* fall into, but wouldn’t.

For example, I don’t think any of them would say something like “Oh, obviously you just haven’t met the right woman. I know this really cute girl Alanna, a friend of my sister’s. I’ll introduce you next time she’s around.”

Or “You must have just had a bad experience with women growing up. Maybe you always got into fights with your mother as a child. But there’s no reason to let that control you now.”

Or “But…but…women are *attractive*! How could you not be *attracted* to people who are *attractive*? That’s just *silly*!”

Or “You know, I’ve hung out with you a long time. You’re not into drama, you don’t wear flamboyant clothing, and you don’t speak with a lisp. You really have all the signs of someone who should be heterosexual. Maybe you’re just wrong about this whole ‘gay’ thing?”

Or “Don’t you realize how important heterosexuality is? Heterosexuality is responsible for childbirth, for most of the love poems throughout history, and for the nuclear family.

How can you not recognize that being straight is better than being gay?”

No, there are a lot of things my friends are far too sophisticated to ever even think about saying if I were to announce something as prosaic and socially acceptable as being gay.

But announce that I *don’t like math*, and suddenly the knives come out.

II.

It’s not that I don’t recognize that math is awesome. If there were “Pray the lack-of-interest-in-math away” camps, I would totally go to one. But just as a gay guy may recognize the many ways his life would be easier if he were heterosexual but this recognition does not immediately lead to finding women attractive – so discoursing on the beauty and importance of math does not suddenly make math books any more readable to me.

Certainly I love the sort of math that doesn’t involve doing actual mathematics. I love reading about Moebius strips and Klein bottles, discussing the implications of Cantor’s discoveries about infinity, even playing around with fractals and tessellations and other forms of mathematical art. It’s just that when you put actual equations in front of me, with numbers and symbols and variables, my brain melts.

I mean, I’m not *terrible* at math. I managed to scrape together an A in Calculus II, the last math class I was required to take and not coincidentally the last math class I ever took. I did it by memorizing the algorithms involved and plugging things into them, all the while desperately praying that there weren’t any deviations, however minor, on the test. This isn’t normal for me. In every other field, concepts slide naturally into my mind and I can manipulate them however they want, like fitting a bunch of Lego blocks together to make limitless possibilities.

But math is like constructing a Lego set on a picnic table outside in the middle of a thunderstorm. I grope blindly in the pouring rain for the first piece, and finally put it in place, but by the time I've found the second piece and move to connect it to the first piece, the first piece has blown away and is nowhere to be found, and the instructions are sopping wet, and the picnic table has just been carried away by a tornado.

I don't know if it's that I'm bad at math, or that I just don't enjoy math enough to be intrinsically motivated to pursue it. I do know that I have never become good at something – *good* good, not "scrape together an A in a mid-level college class on it good" – without having intrinsic motivation to pursue it. And my attempts to hack intrinsic motivation, which would be like a instant win condition for everything if I could achieve it, have been mostly unsuccessful and left me with severe doubt it is even possible. So I have pretty much given up on math.

When I try to explain this to people, the responses are eerily similar to the ones they would never give if I said I was gay.

"Oh, obviously you just haven't learned the right kind of math. I know this really cute proof of the Pythagorean Theorem in my sister's textbook. I'll show it to you the next time we have pencil and paper."

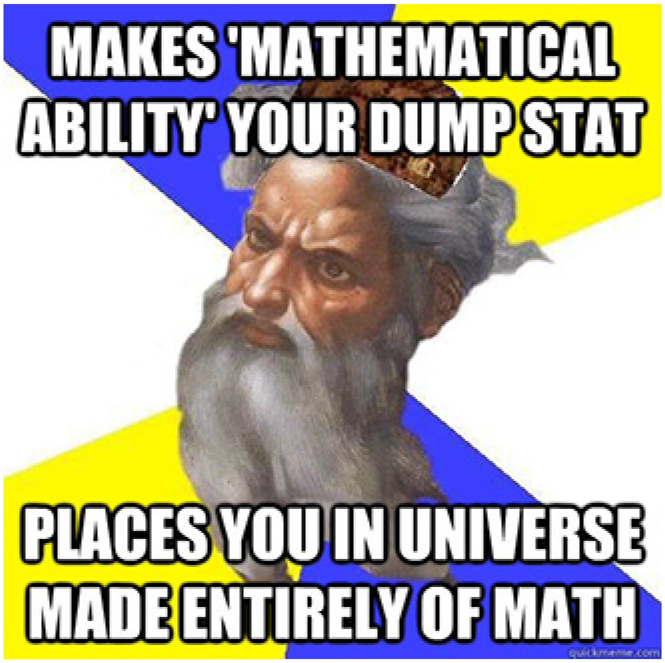
Or "You must have just had a bad experience with mathematics growing up. Maybe you always got yelled at by your math teacher as a child. But there's no reason to let that control you now."

Or “But…but…math is *interesting*! How could you not be *interested* in a subject that’s *interesting*? That’s just *silly*!”

Or “You know, I’ve hung out with you a long time. You like rationality, you’re good at science, and you like analyzing things. You really have all the signs of someone who should be into math. Maybe you’re just wrong about this whole ‘not a math person’ thing?”

Or “Don’t you realize how important math is? Math is essential for statistics, for engineering, for science, for cognition itself! It’s closely linked to art and music and poetry! How can you not recognize that being into math is better than not being into math?”

And yes! I recognize it! Being bad at math is one of my biggest regrets in life! If I were building myself as an RPG character, things would look a lot different, believe you me.


MAKES 'MATHEMATICAL ABILITY' YOUR DUMP STAT
PLACES YOU IN UNIVERSE MADE ENTIRELY OF MATH
quickmeme.com

My contribution to the “Scumbag God” meme

But I’m pretty sure this *is* a skill point kind of thing. I think it’s totally possible for people, even smart people, not to be into math. And I am constantly surprised that people who claim to be experts in evolutionary psychology seem to think it’s entirely plausible that natural selection would evolve gay people with zero interest in procreative sex, but find it *totally outlandish* that anyone could end up without a math drive.

And if they don’t have the skill points, telling someone to just try a little harder at math is like telling Tyrion Lannister to just try a little harder at basketball. Not only is he never going to beat LeBron James, but he’s going to get upset and frustrated trying, and it’s not his comparative advantage anyway.

III.

This whole “comparative advantage” thing took me kind of by surprise.

Some people have trouble believing in ‘g’, the idea that all intelligence is correlated. I’ve always had trouble believing in anything else. I would see people around me being able to solve complicated math problems effortlessly, and think “They are smarter than me. I am their strict intellectual inferior. There is nothing I could possibly contribute to their conversations.”

And when these people started going out of their way to include me in their conversations, when they started making friends with me and even admiring me, I was pretty confused. For a while I was suspicious. Maybe I just had a few talents that were impressive *for someone who was bad at math*, like a chimp that has no verbal thought but can still remember where it hid the banana days later?

The idea that intelligence wasn't monolithic, that I could be much worse than them in proofs and theorems but still be their equal in other areas, was hugely liberating to me, but it took me a very long time to accept it, to believe that I really was as valuable a human being as they were.

And when I tried to analyzed my certainty that – even despite the whole multiple intelligences thing – I couldn't possibly be as good as them, it boiled down to something like this: they were talented at hard things, but I was only talented at *easy* things.

It took me about ten years to figure out the flaw in this argument, by the way.

IV.

I remember gossiping about a friend who was really into the worst types of politics – the kind where you're *obsessed* about whether the head of the Republican National Committee will cut funding to a representative who said something mildly contrary to what someone else wanted him to say – and somewhere in the middle of the conversation my tone switched from "Yeah, what a loser to be concerned about that kind of thing" to "Yeah, poor guy, apparently he drew the short straw in the Things To Be Fascinated About Lottery."

Since then I have returned to the idea of this Things To Be Fascinated About Lottery a lot. There are some good draws you can get – people who are honestly fascinated with business and intrinsically motivated to pursue it only need high IQ and a few other subsidiary skills to get super rich. People who draw math can pursue perfect pure and philosophical truth or excel at pretty much any science they choose and advance human knowledge. People who draw science without math have a harder time – I think I'm one of those – but there are still places for them.

And then there are other people who get other straws. There are some people who are really into politics, and find science really boring. There others who couldn't care less about politics *or* science, as long as their sportsball team wins the Sports Bowl. Ozy picked a straw and ended up fascinated with gender, which wouldn't be so bad except that it means ze occasionally has to talk to the *other* people who are fascinated with gender.

But the thing is, I couldn't *choose* to be interested in sports any more than I could choose to be interested in math or a huge sports fan could choose to be interested in psychology or a gay person could choose to be interested in women. I mean, there's probably some wiggle room, maybe if I put a lot of effort into finding the most interesting sports and learning everything about them I could appreciate them a little. But would I have comparative advantage over the kid who memorized the stats of every pitcher in both leagues when he was 8? Barring getting hit by some kinda cosmic rays or something, I don't think that'll *ever*> happen.

I'd hate to turn this into a "rank which straws are best" contest, but some certainly earn you more money, some certainly help you contribute to the future of humankind more, and some certainly land you in healthy areas of study with nice people and mostly rational thought while others land you in diseased fields full of angry partisans. I wish I had landed in math.

But I no longer blame myself for not having done so. And when anyone starts telling me about how I'd love math if I only did it *their* way, I now only have a *little* twinge of guilt in telling them to go away.

I Aten’t Dead

I am torn between being disappointed at how poorly hospitals work and being amazed that they work at all.

There are two hundred different very specialized professions all trying to help the same set of people. As difficult as medicine is – and goodness knows it is difficult – maybe about half of the resources of hospitals and the people inside them are devoted to anything remotely medical, and the other half are devoted to the near-impossible enterprise of synchronizing these hundreds of different professions into providing something that looks vaguely like coordinated care.

Everyone’s time is so expensive that the system can’t afford to let them waste precious moments figuring out what they’re doing. So pretty much everyone looks at a computer system that tells them what they’re doing, does it, and then enters what they did into the computer system, along with what that implies for other people. The computer system then updates itself and tells other people to react to what the first group of people did.

This happens. A patient tells the nurse that he feels short of breath. The nurse places an order in the computer for a doctor to evaluate the patient. The doctor reads the computer system, evaluates the patient, and places an order for the nurse to draw blood to send to the lab. The nurse draws blood and places an order for the lab to test the blood. The lab tests the blood and places an order in the computer for a doctor to read the lab results. The doctor (who may be a totally different person from the first doctor in this story) reads the lab results, diagnoses pneumonia, and places an order for the pharmacy to send pneumonia medication. The pharmacist sends the pneumonia medication and places an order for the nurse to give it to the patient. The nurse (who, again, may be a totally different nurse) gives it to the patient, who tells the nurse he feels much better. The nurse places an order for a doctor to evaluate the patient. The doctor (who by this point may be a *third* doctor) says it looks like his pneumonia is gone and places an order for the case manager to discharge the patient. The case manager finds somewhere for the patient to go and places an order for big burly guys to wheel the patient to the door. The big burly guys wheel the patient to the door and place an order for an ambulance to come pick the patient up and bring him to the nursing home.

(because – and sorry for the interjection – it is always a nursing home. One of the first things I have learned is that, of the 1000 different places you can send patients when they improve, 999 are euphemisms for “nursing home”. If you are old, and a doctor tells you “Don’t worry, we’re not sending you to a nursing home, we’re sending you to a rehabiliation hospital/skilled care facility/group care institution/convalescent center/ANYTHING ELSE A DOCTOR SAYS TO AN ELDERLY PERSON”, be aware that these are just different ways of saying “nursing home”)

But getting back to the point – that’s ten orders just in this ludicrously simple example. At each of those steps, someone who has no idea who the patient is and might never have anything to do with him outside this single five minute interaction has total responsibility for patient care. At each of those steps, someone can use the computer wrong, place the order for the wrong patient, or accidentally forget the order and leave the patient sitting out in the corridor for twelve hours.

And every single one of those steps is a horrendously busy person who has received five orders at the same time, all of which are marked “HIGHEST PRIORITY” and who are operating on four hours sleep, and each of whom hates everyone at every other level because they keep giving them more work to do. And they are orbited by consultants, specialists, subspecialists, subsubspecialists, subsubsubspecialists, The Guy Who Knows Everything About The Protein Biochemistry Of One Tiny Part Of The Gall Bladder And Nothing About Anything Else, respiratory therapists, speech therapists, physical therapists, chaplains, Bessie The Adorable Pet Therapy Dog, social workers, People Whose Entire Job Is Sitting Next To Suicidal People And Making Sure They Don’t Kill Themselves, nurse assistants, nurse assistant assistants, nurse assistant assistant assistants, businessmen who overuse the phrase ‘patient-centered’, dieticians, pharmacists, and me.

My job as an intern is to navigate this bureaucracy. I’m with an attending – an experienced, prestigious doctor – who doesn’t want to waste any time. He just wants to go “Someone tell me what’s going on with the patient in room 824” and have someone answer “They have such and such a history and these lab results” so he can say “Yup, I use my massive medical expertise to diagnose that as Disease X, give them this and this and this intervention and then give them a repeat scan” and then move on to the next patient.

My job is to write down all the information he wants and get it for him, then write down all the stuff he wants done and do it for him. This generally involves entering orders into the computer, then begging people to do them, then calling them to see if they were done on time, then wailing and gnashing my teeth when they were not. It’s noticing that a patient who was supposed to have had a CT scan ten hours ago is sitting in the corridor looking lost and somewhat unscanned and trying to figure out where something went wrong and how it can be corrected and if necessary threatening that I will have my attending yell at whoever is responsible.

Basically I am a secretary, but a secretary with enough medical knowledge that when my attending says “Order some piperacillin” I won’t spell it “pie bear on ceiling” or ask whether it is some kind of exotic fruit. Enough medical knowledge that if someone gets left out in front of a CT for ten hours, I know when to say “do the CT scan now because it’s the right thing to do” versus “do the CT scan now or she will probably die”.

Sometimes I get sent to take medical histories and examinations from people. They are never glad to see me. They usually say something like “I’m in constant pain and I can’t breathe and I just answered these exact same questions for the paramedics and the emergency room doctors, do I really have to do it again?” And I say “Yes” and then ask them How Long The Symptoms Have Lasted and Whether They Smoke Cigarettes because those are the rules.

If you are entering medical residency and want to prepare for it, practice reciting the following phrase: “I don’t know the answer to that question, but I will page my senior resident and he will get back to you.” I keep getting paged by nurses with questions like “Your patient in room 315 has critically low potassium, please advise.” And I went through medical school and I know stuff about low potassium and I have clever ideas for what to do and probably in over 80% of cases those ideas would not kill the patient. But “over 80%” is not enough so each time I have to find my senior, who is extremely busy and hates getting paged exactly as much as everyone else, and ask him a stupid question I already know the answer to. A lot of my job seems to be a totally useless middleman in between patients and nurses who have questions and senior doctors who have answers, but I'm learning a lot and one day I hope to bring my "ideas that will not kill the patient" ratio up to the 90% or so which is considered acceptable.

I have already helped, in a little tiny way, keep some people safe and healthy. It's not always very dramatic. A few days ago I was looking at some numbers on a chart, and I was like "Hey, in medical school when I memorized this number, it was a *much bigger* number, and here on the chart it's a *very small* number. That sounds like the sort of thing we should do something about." The woman for whom that number represented the amount of blood cells in her body was probably happy I noticed. And today we had a Rapid Response, ie the kind of thing where everyone is running around waving defibrillator paddles and overusing the word "stat", and I was yelled at to get some blood from the blood bank, and I successfully got some blood from the blood bank even though this required an entire phone call and two different operations on the computer.

And I have already made mistakes. I'm pretty sure I gave an elderly woman an unnecessary CT scan she had already had two or three days before, which probably cost a few hundred dollars. I came *very* close to ordering a medical procedure for the wrong person, although luckily it was caught beforehand. I made some mistake in ordering a guy's coagulation status checked that made him stay in hospital an extra day until the check could be repeated.

And I've already had one patient die. He died today. It wasn't my fault. He was brain-dead and hooked up to a ventilator, and the family chose my first week on the job to pull the plug. Palliative care patients are the easiest patients ever, and my job was limited to listening to his heart and lungs once a day or so and nodding sagely and telling the family that The End Was Getting Near. The last day I saw him, I leaned in to his chest to listen to his heart, and saw he was wearing a necklace that said “Whosoever Dies Wearing This Scapular Shall Not Suffer Eternal Fire – The Promise Of The Blessed Virgin”. Within about a second I had three major thoughts:

1. “And here I am, trying to cultivate virtue and avoid vice, *LIKE A TOTAL CHUMP!*”

2. “Is this, like, a legendary artifact? Should I be trying to steal it?”

3. “Wait, if people believe this works, why do they wait until someone’s on their death bed to put it on them? What if someone gets hit by a car? Struck by a meteorite? What kind of person has a necklace that saves them from eternal damnation, but only wears it on special occasions?”

And then after about ten seconds, during which I was trying to remain very solemn-looking, as if I found the patient’s heart sounds extremely interesting in a detached medical way, I thought “Wasn’t there something I used to do when I found something fascinating and bizarre like this? Somewhere I would record things back when I wasn’t working sixteen hour days, six to seven days a week? Didn’t I have some kind of a *blog*? I should find that and write about stuff there.”

But don’t expect too much more out of me in the near future.

Who By Very Slow Decay

[*Trigger warning: Death, pain, suffering, sadness*]

I.

Some people, having completed the traditional forms of empty speculation – "What do you want to be when you grow up?", "If you could bang any celebrity who would it be?" – turn to "What will you say as your last words?"

Sounds like a valid question. You can go out with a wisecrack, like Oscar Wilde ("Either this wallpaper goes or I do"). Or with piety and humility, like Jesus ("Into thy hands, o Father, I commend my spirit.") Or burning with defiance, like Karl Marx ("Last words are for fools who haven't said enough.")

Let me puncture all your pleasant dreams. You'll probably never become an astronaut. You're not going to bang Emma Watson. And your last words will probably be something like "mmmrrrgggg graaaaaaaaaaaHAAACK!"

I guess I always pictured dying as – unless you got hit by a truck or something – a bittersweet and strangely beautiful process. You'd grow older and weaker and gradually get some disease and feel your time was upon you. You'd be in a nice big bed at home with all your friends and family gathered around. You'd gradually feel the darkness closing in. You'd tell them all how much you loved them, there would be tears, you would say something witty or pious or defiant, and then you would close your eyes and drift away into a dreamless sleep.

And I think this happens sometimes. For all I know, maybe it happens quite a lot. If it does, I never see these people. They very wisely stay far away from hospitals and the medical system in general. I see the other kind of people.

If you are like the patients I see dying, then here is how you will go.

You will grow old. When you were young, you would go to institutions and gradually gather letters after your name: BA, MD, PhD. Now that you are old, you do the same thing, but they are different institutions and different letters. Your doctors will introduce you to their colleagues as “Mary Smith, COPD, PVD, ESRD, IDDM”. With each set of letters comes another decrease in quality of life.

At first these sacrifices will be minor. The COPD means you have to breathe from an oxygen tank you carry around wherever you go. The PVD will prevent you from walking more than a few feet at a time. The ESRD will require three hours dialysis in a hospital or outpatient dialysis center three times a week. The IDDM will require insulin shots after every meal. Not fun, but hardly inconsistent with a life worth living.

Eventually these will add up beyond your ability to manage them on your own, and you will be sent off to a nursing home. This will seem like a reasonable enough idea, and sometimes it goes well. Other times it gives you freedom to develop a completely new set of morbidities totally unconstrained by what a person in any other situation could possibly be expected to survive.

You will become bedridden, unable to walk or even to turn yourself over. You will become completely dependent on nurse assistants to intermittently shift your position to avoid pressure ulcers. When they inevitably slip up, your skin develops huge incurable sores that can sometimes erode all the way to the bone, and which are perpetually infected with foul-smelling bacteria. Your limbs will become practically vestigial organs, like the appendix, and when your vascular disease gets too bad, one or more will be amputated, sacrifices to save the host. Urinary and fecal continence disappear somewhere in the process, so you're either connected to catheters or else spend a while every day lying in a puddle of your own wastes until the nurses can help you out. The digestive system isn't too happy either by this point, so you can either have a tube plugged directly into your stomach or just skip the middleman and have an IV line feeding nutrients into your bloodstream.

Somewhere in the process your mind very quietly and without fanfare gives up the ghost. It starts with forgetting a couple of little things, and progresses until you have no idea what's going on ever. In medical jargon, healthy people are "alert and oriented x 3", which means oriented to person (you know your name), oriented to time (you know what day/month/year it is), and oriented to place (you know you're in a hospital). My patients who have the sorts of issues I mentioned in the last paragraph are generally alert and oriented x0. They don't remember their own names, they don't know where they are or what they're doing there, and they think it's the 1930s or the 1950s or don't even have a concept of years at all. When you're alert and oriented x0, the world becomes this terrifying place where you are stuck in some kind of bed and can't move and people are sticking you with very large needles and forcing tubes down your throat and you have no idea why or what's going on.

So of course you start screaming and trying to attack people and trying to pull the tubes and IV lines out. Every morning when I come in to work I have to check the nurses' notes for what happened the previous night, and every morning a couple of my patients have tried to pull all of their tubes and lines out. If it's especially bad they try to attack the staff, and although the extremely elderly are *really* bad at attacking people this is nevertheless Unacceptable Behavior and they have to be restrained ie tied down to the bed. A presumably more humane alternative sometimes used instead or in addition is to just drug you up on all of those old-timey psychiatric medications that actual psychiatrists don't use anymore because of their bad reputation.

After a while of this, your doctors will call a meeting with your family and very gingerly raise the possibility of going to "comfort care only", which means they disconnect the machines and stop the treatments and put you on painkillers so that you die peacefully. Your family will start yelling at the doctors, asking how the hell these quacks were ever allowed to practice when for God's sake they're trying to kill off Grandma just so they can avoid doing a tiny bit of work. They will demand the doctors find some kind of complicated surgery that will fix all your problems, add on new pills to the thirteen you're already being force-fed every day, call in the most expensive consultants from Europe, figure out some extraordinary effort that can keep you living another few days.

Robin Hanson sometimes writes about how health care is a form of signaling, trying to spend money to show you care about someone else. I think he's wrong in the general case – most people pay their own health insurance – but I think he's spot on in the case of families caring for their elderly relatives. The hospital lawyer mentioned during orientation that it never fails that the family members who live in the area and have spent lots of time with their mother/father/grandparent over the past few years are willing to let them go, but someone from 2000 miles away flies in at the last second and makes ostentatious demands that EVERYTHING POSSIBLE must be done for the patient.

Your doctors will nod their heads and tell your family they respect their wishes. It will be a lie. Oh, sure, they will *carry out* the family's wishes, in terms of continuing to provide the care. But *respect*? In the cafeteria at lunch, they will – despite medical confidentiality laws that totally prohibit this – compare stories of the most ridiculous families. "I have a blind 90 year old patient with stage 4 lung cancer with brain mets and no kidney function, and the family is *demanding* I enroll her in a clinical trial from Sri Lanka." "Oh, that's nothing. *I* have a patient who can't walk or speak who's breathing from a ventilator and has anoxic brain injury, and the family is insisting I try to get him a liver transplant."

Every day, your doctors will meet with your family another time, and eventually, as your condition worsens and your family has more time to be hit on the head with a big club marked 'REALITY', they will start to relent. Finally, they will allow your doctors to take you off of the machines, and you will be transferred to Palliative Care, whose job I do not envy even though *every single palliative care doctor I have ever met is relentlessly cheerful and upbeat and this is a total mystery to me*.

And you will die, but not quickly. It takes time for the heart to give up, for the lungs to fill with water and stop breathing, for the toxic wastes to build up. It is generally considered wise for the patient to be on epic doses of morphine throughout the process, both to spare them the inevitable pain as their disease takes their course and to spare their family from having to watch them.

…not that they always do. It can take anywhere from a day to several weeks for someone to die. Sometimes your family wants to wait at the bedside for a week. But a lot of the time they have work and things to do. Maybe they live thousands of miles away. You haven't recognized them in years, you haven't spoken a coherent word in months, and even if for some reason your brain chose this moment to recover lucidity you're on enough morphine to be *well* inside the borders of la-la-land. A lot of families, faced with the prospect of missing work and school to sit by what's basically a living corpse day in and day out for weeks just to watch it turn into a non-living corpse, politely decline. I absolutely 100% cannot blame them.

There is a national volunteer program called No One Dies Alone. Nice people from the community go into hospitals to spend time with dying people who don't have anyone else there for them. It makes me happy that this program exists.

Nevertheless, this is the way many of my patients die. Old, limbless, bedridden, ulcerated, in a puddle of waste, gasping for breath, loopy on morphine, hopelessly demented, in a sterile hospital room with someone from a volunteer program who just met them sitting by their bed.

And let me just emphasize again, not everyone dies this way. I am hugely selection biased by my position in a hospital. But enough people die this way. I'm in a small community. There can't be too many deaths here. Of the ones there are, I see a lot of them. And they're not pretty.

*[**EDIT:** Just looked up statistics. Only about a quarter of old people die at home. The rest are split between hospitals (disproportionately ICUs), nursing homes, and hospices.]*

II.

Hospital poetry is notoriously bad.

I mean, practically all modern poetry is bad. Modern poetry by complete amateurs could be expected to be even worse. But hospital poetry is in a league all of its own as far as badness goes.

When I search “hospital poetry”, Google brings up examples like the following:

> Pain… searing
> Belly… throbbing
> There is no baby.
> There will be no baby.
> Endometriosis.

I feel bad making fun of it, because it is clearly heartfelt. This is part of the problem with hospital poetry. It is very heartfelt, whereas I think most popular poetry comes from people who have strong emotions but also some distance from them and a little bit of post-processing. And unfortunately doctors, who are on this decades-long quest to prove they are actual people with real feelings and not just arrogant robot-like people in white coats who know a very large number of facts about thyroiditis, just eat this sort of thing up.

But I’m not really complaining about those sorts of endometriosis poems. The ones I’m really complaining about are worse. The epitome of the genre I can’t find on Google, because it was presented as some kind of event at the hospital where I trained in Ireland. I don’t remember it, but let me just make up some doggerel approximately faithful to the spirit of the original:

> When my doctor told me that I had cancer
> I knew that despair was not the answer

It felt like the darkness was closing in
But to give up would have been a sin
Everyone here helped me so much
And nothing is like a helping hand’s touch
Thanks, Dr. Connell, and everyone in Cork
I really appreciate all your hard work

Doctors and nurses eat this kind of thing up and put it on shiny plaques that go on the walls of the hospital. (I suggest a wall near the gastroenterology unit, to expedite care for people who start vomiting.)

Wittgenstein said that “if anyone ever wrote a book of ethics, that really was a book of ethics, it would destroy all the other books in the world with a bang.” I’m not really sure what he meant. But if anyone ever wrote a book of hospital poetry, that really was a book of hospital poetry… well, I don’t know what would happen, but I bet it would be loud and angry, and that it wouldn’t be put on shiny plaques on anybody’s walls, except maybe the same people who hang Hieronymous Bosch paintings on their walls.

Am I calling hospitals hellish? Sure am. It has nothing to do with the decor, which has actually gotten much nicer in your newer hospitals until it’s hard to tell them apart from a stylish office building. It’s nothing to do with the staff, either – most doctors and some nurses seem pretty happy and trade banter around the water coolers like everyone else. It’s mostly the screams.

The screams are coming about 33% from the confused demented old people I mentioned, 33% from people having minor procedures performed without anaesthetics for one or another good reason, and 33% from people who just have very painful diseases (plus 1% from me sitting in the break room looking up examples of hospital poetry for this post).

They run the gamut of human screams. There are wordless shrieks. There are some angry screams, like “$#%! YOU GET ME OUT OF HERE!”. There are a lot of people screaming “SOMEBODY HELP ME!” And there are some religious screams, like “OH GOD!” or “JESUS HELP ME!” or “CHRIST NO!”.

When I first started working in hospitals, I would not only inevitably run over to these screams, but I would feel contempt and anger at the rest of the hospital staff who would just continue their daily routine. I soon learned better. Not only would I be unable to do anything – I can’t single-handedly cure their painful illness, or make their procedure go any faster, or explain to them that the year is 2013 and they’re no longer on their childhood farm in Oklahoma – but as soon as they saw me I would be the one they started screaming at and expecting to save them. The bystander effect, my last defense, disappeared. Sometimes I would make a stand by asking the nurse to increase their pain medication or something, and be politely told all the reasons why that was a bad idea from a medical perspective (pain medication has lots of side effects which doctors monitor carefully). In the end I would just slink out of the room, wishing I had never come in.

So the constant screams being completely ignored by a bunch of happy people going through their day is pretty hellish. But there’s also the bodies. Usually we are able to avoid thinking about people as bodies except to briefly note that certain people like Emma Watson are really hot. In a hospital, this filter disappears. Some people have gigantic swollen legs the size of your waist. Others have huge ulcerated sores all over. Still others have skin covered with the sorts of bacterial colonies you usually only see on a petri dish. And body sizes range from so thin that you can see their organs bulging out of their skin and use them as a grisly impromptu anatomy lesson, to so morbidly obese that you have to search through the fat folds to find body part you're looking for.

The senses are under constant assault. Smell is the worst. There are some people who can identify different infections by smell. *Pseudomonas aeruginosa* is supposed to smell fruity. *Gardnerella* is supposed to smell fishy. *Clostridium* is supposed to smell like the worst thing you can possibly imagine covered in feces and left to rot on a warm summer day.

But the other senses get their time too. The sight is vexed by flashing call lights. And the hearing is battered with incessant beeping from IV lines which have hard-coded alarms to alert doctors of critically important events such as "Look at me! I am an IV line!" The end result is something it would take a first-rate poet to describe. I'm tempted to nominate Oscar Wilde. He did a good job on prisons in *Ballad of Reading Gaol*, and I feel like the skill would transfer:

> He does not rise in piteous haste
> To put on convict-clothes,
> While some coarse-mouthed doctor gloats,
> and notes each new and nerve-twitched pose,
> Fingering a watch whose little ticks
> Are like horrible hammer-blows […]
>
> He does not stare upon the air
> Through a little roof of glass;
> He does not pray with lips of clay
> For his agony to pass;
> Nor feel upon his shuddering cheek
> The kiss of Caiaphas.

But after some more thought, I think I’m going to go with Wilfred Owen:

> If in some smothering dreams you too could pace
> Behind the wagon that we flung him in,
> And watch the white eyes writhing in his face,
> His hanging face, like a devil’s sack of sin;
> If you could hear, at every jolt, the blood
> Come gargling from the froth-corrupted lungs,
> Obscene as cancer, bitter as the cud
> Of vile, incurable sores on innocent tongues […]

Or better yet, if Oscar Wilde’s muse when he was writing *Reading Gaol* were to bear Wilfred Owen’s children, then those kids would be competent to write hospital poetry that was actually hospital poetry.

Dante would also be an acceptable choice.

III.

You may have read the excellent article How Doctors Die. If you haven’t, do it now. It says that most doctors, knowing everything I’ve just mentioned above, choose to die quickly and with very limited engagement with the health system.

I (and the doctors in my family whom I’ve asked) am pretty much like the doctors in the article. If I get a terminal disease, I want to wring what I can out of the few months of life I have left and totally avoid any surgery, chemotherapy, amputations, ventilators, and the like. It would be a clean death. It would be okay.

My big fear, though, is that I *won’t* get a terminal disease.

If I just start accumulating damage, growing more and more bedridden and demented and pain-riddling until I want out –

well, there won't *be* a way out. If there's not some very specific life-saving treatment that can be withdrawn, I'm stuck above ground, not just in the "unless I want to risk the danger and shame of suicide" way I am now, but – if I'm too debilitated to access means of suicide on my own – in an absolute way.

Even if my doctors and nurses and caretakers are sympathetic, my only legal option, without exposing *them* to jail time, is to starve myself to death – something both painful and difficult, and itself not really the way I want to go.

I was sitting in an ICU room yesterday where a patient's body had just been brought out after their death. My attending was taking care of the paperwork in the other room, and I was sitting there reflecting, and I started thinking about what it would be like to die in that room. There was a big window, and it was a sunny day, and although I mostly had a spectacular view of the hospital parking lot, a bit further in the distance I could see a park full of really big trees. And I knew that if I were dying in that room my last thought would be that I wanted to be outside.

I think if I were very debilitated and knew I would die soon, I would want to go to that park or one like it on a very sunny day, surround myself with my friends and family, say some last words, and give myself an injection of potassium chloride.

(this originally read "morphine", but just today the palliative care doctor at my hospital gave an impassioned lecture about how people need to stop auto-associating morphine with euthanasia, because it makes it really hard for him to offer morphine painkillers to patients who need them without them freaking out. So potassium chloride it is.)

This will never happen. Or if it did, it would be some kind of huge scandal, and whoever gave me the potassium chloride would be fired or something. But the people dying demented and hopeless connected to half a dozen tubes in ICU rooms aren’t considered scandals by anybody. That’s just “the natural way of things”.

I work in a Catholic hospital. People here say the phrase “culture of life” a lot, as in “we need to cultivate a culture of life.” They say it almost as often as they say “patient-centered”. At my hospital orientation, a whole bunch of nuns and executives and people like that got up and told us how we had to do our part to “cultivate a culture of life.”

And now every time I hear that phrase I want to scream. 21st century American hospitals do not need to “cultivate a culture of life”. We have enough life. We have life up the wazoo. We have more life than we know what to do with. We have life far beyond the point where it becomes a sick caricature of itself. We prolong life until it becomes a sickness, an abomination, a miserable and pathetic flight from death that saps out and mocks everything that made life desirable in the first place. 21st century American hospitals need to cultivate a culture of life the same way that Newcastle needs to cultivate a culture of coal, the same way a man who is burning to death needs to cultivate a culture of fire.

And so every time I hear that phrase I want to scream, or if I cannot scream, to find some book of hospital poetry that really is a book of hospital poetry and shove it at them, make them read it until they understand.

There is no such book, so I hope it will be acceptable if I just rip off of Wilfred Owen directly:

If in some smothering dreams you too could pace
Behind the gurney that we flung him in,
And watch the white eyes writhing in his face,
His hanging face, like a devil’s sack of sin;
If you could hear, at every jolt, the blood
Come gargling from the froth-corrupted lungs,
Obscene with cancer, bitter with the cud
Of vile, incurable sores on innocent tongues
My friend, you would not so pontificate
To reasoners beset by moral strife
The old lie: we must try to cultivate
A culture of life.

I Myself Am A Scientismist

I.

“Science can tell you about rocks and molecules and stars. But what kind of science can tell you about *the deepest recesses of the human soul?*”

I hear this a lot, and I want to answer “Psychology! It’s this whole science that totally exists and is *all about that!*” But then they would just change “deepest recesses of the human soul” to “how to be a good person”, or “whether life has meaning” or whatever.

Of course, there are sciences that bear on these questions. For example, biology can tell us a lot about the evolutionary origins of our moral intuitions, which sounds like the sort of thing that might be useful if you’re trying to figure out how to be a good person. But the overall claim that empiricism and experiment cannot single-handedly solve these problems for us seems to me to be correct.

“Scientism” is a purported fallacy in which people naively believe that science can solve everything. Wikipedia defines it as “belief in the universal applicability of the scientific method.” But for a problem that’s supposedly so common, it lacks a sort of at-all-believability.

I mean, this should be – pardon my scientism – an empirical question. Has someone done an experiment that has figured out how we should live our lives? Is there a grant proposal in the works for such an experiment? Does anyone seriously believe we may one day figure out the best way to live by splitting the hedon in a giant particle accelerator? No? Then who exactly are these so-called…wait, that doesn’t work…

er…can we call them scientismists? Is that a word? No? Okay. But who *are* they?

When I hear people accused of scientism, they’re not trying to determine the moral law with particle accelerators. They’re trying to determine the moral law the same way their accusers are – thinking about it for a while, devising long jargon-filled arguments, then publishing articles in philosophy journals. They are doing nothing remotely resembling the scientific method. Nor do they especially connect with the results of science. Consequentialists are accused of scientism a lot, but there’s nothing in consequentialism incompatible with the planets being pushed around by angels, or thunder happening when the gods go bowling. Something else has to be going on here.

On some level it seems to be about personalities, what academic-y types call inter-departmental squabbling. The people making the accusations of scientism are culturally sophisticated types who read Cicero and Plato, who write in flowery prose, who speak fluent French. The people getting accused are geeky types who read Einstein or Feynman, who write in dense mathematical notation, who program in C. On some level, it expresses that oldest of human requests: “Aaaagh! Foreigners! Get off my turf!”

But I don’t think it’s *just* about turf battles. I think people are right to identify scientism as a thing. These two groups of people think differently. There are different processes going on in their minds. They will reach different results. Even if neither side suddenly breaks out a test tube, one side will be doing something fundamentally more scientific than the other.

And let me show my colors: I think one of them is doing something *better*. I myself am a scientismist. I think the impact of having people thinking scientifically in non-scientific fields is usually good.

II.

Why should that be? If science is just about rocks and molecules and stars, why would scientific training and knowledge give you an advantage in unrelated fields?

From Shakespeare’s Cassius: “The fault, dear Brutus, is not in our stars / But in ourselves.”

I don’t think science should inform philosophy because of what it’s discovered about stars. It should inform philosophy because of what people in the process of investigating stars have incidentally discovered about the faults in themselves.

Imagine a prankster with superhuman skill in ophthamological surgery manages to cut open and rearrange your eyes while you’re asleep. She gives your vision a sort of tilt-shift effect that makes everything appear smaller. And at the time, you happen to be on a World Tour.

Your friend asks you how Paris is, and you say: “It looks very small! It’s full of tiny people and a miniature Eiffel Tower!” Your friend corrects you and tells you Paris is actually normal sized.

Then you’re in London. You mention how it’s full of dwarves and a cute little clock tower the size of a sewing needle. Once again your friend corrects you and tells you London is normal size.

The next week you’re in Beijing. You’re tempted to dismiss it as a city of midgets and of medium-sized portraits of Mao. But by now you’ve wised up. Your experiences in Paris and

London have taught you that there's something wrong with your vision and you had better be more careful.

A detractor might say "What can learning about Paris and London possibly teach you about Beijing? It's on a totally different continent and steeped in a totally different culture. Lessons learned in Europe just don't transfer!" But as long as you're using the same faulty vision to view each city, the lessons learned *do* transfer. Even if facts about China are completely uncorrelated with any facts about Europe, your *errors* about both will be correlated because it's the same person erring each time.

III.

For all that it stresses empiricism, science isn't just about experiment and observation. It's also got a theoretical side. The interesting part of science is that it's a calibration process. You use your theorizing faculties, and then you perform experiments to see if you were right or wrong.

Just as a biologist-engaged-in-experiment is testing different drugs to see whether they cure disease, a biologist-engaged-in-theory is (usually unintentionally) testing different mental algorithms to see whether they correctly predict which drugs will cure disease, or can generate disease cures.

And just as experimental science may discover that the witch doctor's technique of drilling a hole in the skull to let out the evil demons is not in fact best practice, so theoretical science may discover that certain reasoning techniques don't stand up to scrutiny either.

One of these which has been downright mythologized is the story of How We Learned That Things Aren't Usually Caused By Sentient Agents. Back in the old days rain was caused by the Rain God and disease was caused by the Disease Demon, but then we discovered that these were actually natural processes and not people at all, and (so the myth continues) One Day We Will Finally Complete The Process By Ceasing To Believe In God.

The only problem with this narrative is that as far as I know we stopped believing in the Rain God and the Disease Demon long before we had any good experimental science or even any naturalistic alternative explanations for rain or disease. I'm not sure why this is, but it makes it less than a perfect victory for Science.

Still, some very similar stories *are*. The Copernican Principle, for one, where we gradually lost belief in our own uniqueness and went from "Earth holds a privileged position" to "The solar system holds a privileged position" to "The galaxy holds a privileged position" to our current and obviously-correct "Okay, there's a whole universe out there, but it *definitely* has a privileged position and doesn't split up into lots of different equally real quantum branches".

There are other principles without equally catchy names. The "No, You Can't Just Treat Human-Level Interesting Categories As Ontologically Real Primitives" principle, which I suppose one could call the Huxleyan Principle after the biologist who worked the hardest to discredit elan vital. The "Stop Using Value-Based Explanations" principle, which can be used with equal aplomb against everyone from the old Great Chain of Being theorists to high school biology students who insist that evolution is progress from "worse" to "better" organisms.

(life hack: Does saying "worse" and "better" make you feel unscientific? Just replace these words with "less complex" and “more complex”, then pretend these terms have objective meanings!)

IV.

Each of these principles works not because of the particular field it is applied to, but because it compensates for a defect in our own reasoning faculty. Our brain evolved mostly to think about other humans, thinking about other humans is the first thing it wants to do in any situation, so we end up with a bias towards anthropomorphism. We are clearly very important to ourselves, so we project this onto the universe and think we (or our planet, or our star, or our galaxy) must be at the center.

Therefore, the correct application of these principles is an antiprediction, a sort of easily defensible sticking to a default position. For example, “the world will probably not end on January 18, 2020” is an antiprediction, because we have no reason to think that it should. It is very difficult to predict the future, but this is no argument against my claim that the world will probably not end on January 15, 2020. No one gets to shake their head and say “That’s kind of *arrogant* of you to think that you can know that.”

Antipredictions do not always *sound* like antipredictions. Consider the claim “once we start traveling the stars, I am 99% sure that the first alien civilization we meet will *not* be our technological equals”. This sounds rather bold – how should I know to two decimal places about aliens, never having met any?

But human civilization has existed for 10,000 years, and may go on for much longer. If “technological equals” are people within about 50 years of our tech level either way, then all I’m claiming is that out of 10,000 years of alien civilization, we won’t hit the 100 where they are about equivalent to us. 99% is the exact right probability to use there, so this is an antiprediction and requires no special knowledge about aliens to make.

The antipredictive nature is surprising because certain possibilities stand out more clearly to us. Aliens being around our tech level is *narratively interesting* – we can fight wars on an equal footing or engage in mutually profitable trade. Certainly the idea of meeting aliens who have been stuck at the tech level of Assyria for a thousand years is less available.

We can say that the hypothesis-space is distorted: that equal-tech aliens *looks* like it takes up a very large area, even though it is tiny.

In the same way, certain salient regions of hypothesis space that correspond to natural human thought processes falsely appear very large, and certain other regions that don’t correspond to natural thought processes falsely appear much smaller.

If you’ve calibrated yourself on previous problems, then “The ground of being has to be a person” should bring up alerts like “Wait a second, it also seemed like rain had to be a person”.

And “I bet moral value is this objectively real conceptual primitive of perfect simplicity” should bring up alerts like “Wait a second, it also seemed like life had to be an objectively real conceptual primitive of perfect simplicity, and it ended up being this.”

This should work the same way that the observation “Beijing seems full of tiny little Chinese midgets” brought up the alert “Wait a second, Paris also seemed full of tiny little French midgets”.

V.

Beyond these specific problems like the Copernican Principle lies a greater a problem which makes all others pale into insignficance.

People who haven’t calibrated their theorizing against hard reality still think verbal reasoning works.

There have been a couple hundred proofs of the existence of God thought up throughout the centuries. And more recently, there have also been a couple hundred proofs of the nonexistence of God thought up. Clearly, a couple hundred proofs of something doesn’t make it so.

“But no one ever said something must be true just because someone has published a proof! The proof must be correct! The proofs of the existence/nonexistence of God are just wrong!”

Well, yes. Of course. But which side’s proofs you think are wrong tend to have a very very very strong correlation with which side you personally subscribe to.

Our faculty for evaluating chains of deductive reasoning similar to proofs of the (non)existence of God, or a lot of what goes on in philosophy, or god help us politics, is – pardon my language – *really shitty*. And we never realize this, because it is selectively shitty. It tells us it has logically evaluated arguments, and determined our opponents’ arguments are wrong, and our own arguments are right. And this is nice and consistent and convenient so we assume it must know what it’s doing. If it gets proven wrong once or twice or sixty times, we can dismiss that as a fluke, or an edge case, or It’s Beside The Point, or The Real Question Is Whether You Are Racist For Even Bringing That Up.

The thing I notice about scientists who branch out into other fields and get accused of scientism is that they tend to be *minimalists*. They’re always the ones saying there *isn’t* something. There probably isn’t a god. There probably isn’t Cosmic Consciousness. There probably isn’t any particular moral law beyond your actions just having effects in the world.

And their opponents believe this is because they fetishize Science as the only thing that can possibly be real. You can see bacteria under a microscope, you can see atoms under a microscope, but you can’t see God under a microscope, and therefore if they don’t believe in God it’s because they have obstinately decided only to believe in Science-y things.

But in fact, it is exactly the reverse. These skilled wielders of rejection first trained themselves on Science-y things. Lamarckian evolution. Steady State theory. The planet Vulcan. The four humors. The blank slate. Radical behaviorism. Catastrophism. Recapitulation theory. The luminiferous aether.

By holding scientific theories, which can be and are disproven, they trained themselves in Doubt. And that Doubt continues to serve them when they branch into other areas where theories cannot be disproven so easily. And maybe they will be less easily swayed by attractive verbal arguments.

The people who get accused of scientism are not all themselves scientists, and even those who are may never have suffered a mistake equal in enormity to believing in luminferous aether. But they’re steeped in the culture. They’ve absorbed the mores. Even if they have no scientific virtue themselves are merely aping the motions of their betters, those motions themselves contain certain safeguards against some of the most atrocious errors.

I don’t believe such scientifically informed people, when branching off into other fields, will always or even often be right. But I think they have a better chance than people working from intellectual traditions that have never gotten to calibrate their thought processes in the same way.

And that is why I consider myself a scientismist. I know it is supposed to be a perjorative, but I am reclaiming it. And I know it has many definitions, but this one is mine:

A view of hypothesis-space that accounts for human fallibilities, as revealed by past experiences.

And a very, very high burden of proof before zeroing in on any one area of that space.

Holocaust Good For You, Research Finds, But Frequent Taunting Causes Cancer In Rats

A study published this month in PLoS One finds that victims of weight discrimination ("fat-shaming", in case you only speak Tumblrese) are more likely to subsequently gain weight.

It's hard for me to like a study that so obviously got exactly the result its organizers wanted it to get. And obvious confounders are obvious – level of discrimination faced was based on self-report, and the sorts of people who hang around the sorts of people who fat-shame may differ systematically (in class? education?) than who avoid that kind of abuse – but the study's endpoint of *change* in weight over time rather than just weight itself goes some of the way toward addressing those concerns. And I've got to give them credit for studying an important issue and getting a highly significant result. So let's let them have their soapbox:

> There are both behavioral and physiological mechanisms that may contribute to the relation between discrimination and obesity. Weight discrimination is associated with behaviors that increase risk of weight gain, including excessive food intake and physical inactivity. There is robust evidence that internalizing weight-based stereotypes, teasing, and stigmatizing experiences are associated with more frequent binge eating. Overeating is a common emotion-regulation strategy, and those who feel the stress of stigmatization report that they cope with it by eating more. Individuals who endure stigmatizing

> experiences also perceive themselves as less competent to engage in physical activities and are thus less willing to exercise and tend to avoid it. Finally, heightened attention to body weight is associated with increased negative emotions and decreased cognitive control. Increased motivation to regulate negative emotions coupled with decreased ability to regulate behavior may further contribute to unhealthy eating and behavioral patterns among those who are discriminated against.

New study! This one published – oh, look, isn't that interesting – this month in PLoS One, finds that survivors of the Holocaust have greater life expectancy than control Jews who did not experience the Holocaust.

Here the authors definitely got a result they were *not* looking for and did *not* want. And here, too, we have all sorts of confounders: they tried hard to construct a matched control group of Jews who emigrated from Poland to Israel just before the Holocaust, but we have no idea what sort of differences there might have be in those populations (just to make up one story, maybe poor people who had less to lose were more likely to emigrate). And here too, there is no shortage of soapboxes. From here:

> One possible explanation for these findings might be the "Posttraumatic Growth" phenomenon, according to which the traumatic, life-threatening experiences Holocaust survivors had to face, which engendered high levels of psychological distress, could have also served as potential stimuli for developing personal and inter-personal skills, gaining new insights and a deeper meaning to life. All of these could have eventually contributed to the survivors' longevity. "The results of this research give us hope and teach us quite a bit about the resilience of the human spirit when faced with

> brutal and traumatic events”, concluded Prof. Sagi-Schwartz.

So, let me sum up what we’ve learned here today.

Having someone call you fat is a profoundly disturbing form of stigmatization that breaks your normal cognitive coping mechanisms and subjects you to levels of stress that the human body and psyche were never designed to withstand.

But being rounded up like cattle, having your entire family killed in front of you, and then being starved nearly to death in a concentration camp for several years is useful opportunity to grow as a person, and will leave you stronger and better-adjusted.

I shouldn’t be too sarcastic. Stranger things have ended up being true. Maybe constant low-grade minor stress has a deleterious effect but a single extremely stressful event can be salutary. Maybe stress is good for you only after you’ve achieved a safe distance from the stress and can reflect on it from a position where you’re absolutely sure it will never happen again. Maybe stress makes you obese in the short term, but also makes you live longer in the long-term. Maybe the cultural differences between elderly Polish Jews and middle-aged Americans mediate the effect stress has on their bodies.

Or maybe these effects are mediated by unexpected processes. Maybe the Holocaust survivors live longer not because of personal growth, but because they got a sort of involuntary caloric restriction that permanently altered their metabolism. Maybe (as the researchers point out in their paper) only people who were exceptionally healthy survived the Holocaust, and these people continued being exceptionally healthy into their old age. Maybe obese people who aren’t shamed stick to a careful diet to avoid shaming, but once the shaming starts they figure it can’t get any worse and go wild.

Or maybe one or both of these studies is totally and fundamentally flawed and we’re wasting our time here. I give 50% probability that the fat result is legitimate, and 90% probability the Holocaust result is due to something other than personal growth, probably survivor effect or caloric restriction – but I bet others will disagree.

Yet I think what struck me most about this combination was how “stress makes you miserable and unhealthy” sounds reasonable, but “stress is a salutary process that allows you to grow” also sounds reasonable. No matter what happens to stressed people, psychology can go “Oh yeah, according to our theories, stress causes that” and I will nod my head and agree.

Or maybe another way to put it is that I’m impressed with the ease at which we switch narratives. *All the time* I hear “Well, a little bit of adversity will be good for him/her”. Or else “What you’re doing is going to destroy his/her self-esteem and scar him/her for life.” Most people selectively use either or, depending on whether they want to excuse something or condemn something at that particular moment, and they have the science available to support either.

Not only do we operate on cached thoughts, but we have a store of contradictory cached thoughts sufficient to support any proposition *or* its opposite.

On first looking into Chapman’s “Pop Bayesianism”

I.

David Chapman keeps complaining that “Bayesianism” – as used to describe a philosophy rather than just a branch of statistics – is meaningless or irrelevant, yet is touted as being the Sacred Solution To Everything.

In my reply on his blog, I made the somewhat weak defense that it’s not a disaster if a philosophy is not totally about its name. For example, the Baptists have done pretty well for themselves even though baptism is only a small part of their doctrine and indeed a part they share with lots of other denominations. The Quakers and Shakers are more than just people who move rhythmically sometimes, and no one gives *them* any grief about it.

But now I think this is overly pessimistic. I think Bayesianism is a genuine epistemology and that the only reason this isn’t obvious is that it’s a really *good* epistemology, so good that it’s hard to remember that other people don’t have it. So let me sketch two alternative epistemologies and then I’ll define Bayesianism by contrast.

II.

Aristotelianism

Everyone likes to beat up on Aristotle, and I am no exception. An Aristotelian epistemology is one where statements are either true or false and you can usually figure out which by using deductive reasoning. Tell an

Aristotelian a statement and, God help him, he will either agree or disagree.

Aristotelians are the sort of people who say things like “You can never really be an atheist, because you can’t prove there’s no God. If you were really honest you’d call yourself an agnostic.” When an Aristotelian holds a belief, it’s because he’s damn well *proven* that belief, and if you say you have a belief but haven’t proven it, you are a dirty cheater taking epistemic shortcuts.

Very occasionally someone will prove an Aristotelian wrong on one of his beliefs. This is shocking and traumatic, but it certainly doesn’t mean that any of the Aristotelian’s *other* beliefs might be wrong. After all, he’s proven them with deductive reasoning. And deductive reasoning is 100% correct by definition! It’s *logic*!

Anton-Wilsonism

Nobody likes to beat up on Robert Anton Wilson, and I consistently get complaints when I try. He and his ilk have *seen through* Aristotelianism. It’s a sham to say you ever know things for certain, and there are a lot of dead white men who were cocksure about themselves and ended up being wrong. Therefore, the most virtuous possible epistemic state is to *not believe anything.*

This leads to nihilism, moral relativism, postmodernism, and mysticism. The truth cannot be spoken, because any assertion that gets spoken is just another dogma, and dogmas are the enemies of truth. Truth is in the process, or is a state of mind, or is [insert two hundred pages of mysterianist drivel that never really reaches a conclusion].

Bayesianism

“Epistemology X” is the synthesis of Aristotelianism and Anton-Wilsonism. It concedes that you are not certain of any of your beliefs. But it also concedes that you are not in a position of global doubt, and that you can update your beliefs using evidence.

An Xist says things like “Given my current level of knowledge, I think it’s 60% likely that God doesn’t exist.” If they encounter evidence for or against the existence of God, they might change that number to 50% or 70%. Or if they don’t explicitly use numbers, they at least consider themselves to have strong leanings on difficult questions but with some remaining uncertainty. If they find themselves consistently over- or under-confident, they can adjust up or down until they reach either the certainty of Aristotelianism or the total Cartesian doubt of Anton-Wilsonism.

Epistemology X is both philosophically superior to its predecessors, in that it understands that you are neither completely omniscient nor completely nescient; instead, all knowledge is partial knowledge. And it is practically superior, in that it allows for the quantification of belief and therefore can have nice things like calibration testing and prediction markets.

What can we call this doctrine? In the old days it was known as probabilism, but this is unwieldy, and it refers to a variety practiced before we really understood what probability *was*. I think “Bayesianism” is an acceptable alternative, not just because Bayesian updating is the fundamental operation of this system, but because Bayesianism is the branch of probability that believes probabilities are degrees of mental credence and that allows for sensible probabilities of nonrepeated occurrences like “there is a God.”

III.

"Jason" made nearly this exact same point on David's blog. David responds:

> 1) Do most people really think in black and white? Or is this a straw man?
>
> 2) Are numerical values a good way to think about uncertainty in general?
>
> 3) Does anyone actually consistently use numerical probabilities in everyday situations of uncertainty?

The discussion between David and Jason then goes off on a tangent, so let me give *my* answer to some of these questions.

Do people really think in black and white? Or in my formulation, is the "Aristotelian" worldview really as bad as all that? David acknowledges the whole "You can't really be an atheist because…" disaster, but says belief in God is a special case because of tribal affiliation.

I have consistently been tempted to agree with David – my conception of Aristotelianism certainly *sounds* like a straw man. But I think there are some inferential distances going on here. A year or so ago, my friend Ari wrote of Less Wrong:

> I think there's a few posts by Yudkowsky that I think deserve the highest praise one can give to a philosopher's writing: That, on rereading them, I have no idea what I found so mindblowing about them the first time. Everything they say seems patently obvious now!

Obviously not everyone gets this Bayesian worldview from Less Wrong, but I share this experience of “No, everything there is obvious, surely I must always have believed it” while having a vague feeling that there had been something extremely revolutionary-seeming to it at the time. And I have memories.

I remember how some of my first exposure to philosophy was arguing against Objectivists in my college’s Objectivist Club. I remember how Objectivism absolutely lampshades Aristotelianism, how the head of the Objectivist Club tried very patiently to walk me through a deductive proof of why Objectivism was was correct from one of Rand’s books. “It all starts with A = A,” he told me. “From there, it’s just logic.” Although I did not agree with the proof itself, I don’t remember finding anything objectionable in the methodology behind it, nor did any of the other dozen-odd people there.

I remember talking to my father about some form of alternative-but-not-implausible medicine. It might have been St. John’s Wort – which has an evidence base now, but this was when I was very young. “Do you think it works?” I asked him. “There haven’t been any studies on it,” he said. “There’s no evidence that it’s effective.” “Right,” I said, “but there’s quite a bit of anecdotal evidence in its favor.” “But that’s not proof,” said my father. “You can’t just start speculating on medicines when you don’t have any proof that they work.” Now, if I were in my father’s shoes today, I might *still* make his same argument based on a more subtle evidence-based medicine philosophy, but the point was that at the time I felt like we were missing something important that I couldn’t quite put my finger on, and looking back on the conversation, that thing we were missing is obviously the notion of probabilistic reasoning. From inside I know *I* was missing it, and when I asked my father about this a few years ago he completely failed to understand what relevance that could possibly have to the question, so I feel confident saying he was missing it too.

I remember hanging out with a group of people in college who all thought Robert Anton Wilson was the coolest thing since sliced bread, and it was *explicitly* because he said we didn't have to believe things with certainty. I'm going to get the same flak I always get for this, but Robert Anton Wilson, despite his brilliance as a writer and person, has a *really dumb* philosophy. The *only* context in which it could possibly be attractive – and I say this as someone who went around quoting Robert Anton Wilson like nonstop for several months to a year – is if it was a necessary countermeasure to an *even worse* epistemology that we had been hearing our entire lives. What philosophy is this? Anton Wilson explicitly identifies it as the Aristotelian philosophy of deductive certainty.

And finally, I remember a rotation in medical school. I and a few other students were in a psychiatric hospital, discussing with a senior psychiatrist whether to involuntarily commit a man who had made some comments which sort of kind of sounded maybe suicidal. I took the opposing position: "In context, he's upset but clearly not at any immediate risk of killing himself." One of the other students took the opposite side: "If there's *any chance* he might shoot himself, it would be irresponsible to leave him untreated." This annoyed me. "There's "some chance" *you* might shoot yourself. Where do we draw the line?" The other student just laughed. "No, we're being serious here, and if you're not totally certain the guy is safe, he needs to be committed."

(before Vassar goes off on one of his "doctors are so stupid, they don't understand anything" rants, I should add that the senior psychiatrist then stopped the discussion, backed me up, and explained the basics of probability theory.)

So do most people really think in black and white? Ambiguous. I think people don’t account for uncertainty in Far Mode, but do account for it in Near Mode. I think if you explicitly ask people “Should you take account of uncertainty?” they will say “yes”, but if you ask them “Should you commit anybody who has *any chance at all* of shooting themselves?” they will also say yes – and if you ask them “What chance of someone being a terrorist is too high before you let them fly on an airplane, and don’t answer ‘zero’?” they will look at you as if you just grew a second head.

In short, they are not actually idiots, but they have no coherent philosophical foundation for their non-idiocy, and this tends to show through at inconvenient times.

Probability theory in general, and Bayesianism in particular, provide a coherent philosophical foundation for not being an idiot.

Now in general, people don’t need coherent philosophical foundations for anything they do. They don’t need grammar to speak a language, they don’t need classical physics to hit a baseball, and they don’t need probability theory to make good decisions. This is why I find all the “But probability theory isn’t that useful in everyday life!” complaining so vacuous.

“Everyday life” means “inside your comfort zone”. You don’t need theory inside your comfort zone, because you already navigate it effortlessly. But sometimes you find that the inside of your comfort zone isn’t so comfortable after all (my go-to grammatical example is answering the phone “Scott?

Yes, this is him.”) Other times you want to leave your comfort zone, by for example speaking a foreign language or creating a conlang.

When David says that “You can’t possibly be an atheist because…” doesn’t count because it’s an edge case, I respond that it’s *exactly* the sort of thing that should count because it’s people trying to actually *think* about an issue outside their comfort zone which they can’t handle on intuition alone. It turns out when most people try this they fail miserably. If you are the sort of person who likes to deal with complicated philosophical problems outside the comfortable area where you can rely on instinct – and politics, religion, philosophy, and charity all fall in that area – then it’s *really nice* to have an epistemology that doesn’t suck.

Extreme Mnemonics

[Very minor Game of Thrones spoilers]

From one of my LW comments, discussing how hard it is to learn biochemistry:

> I’m having the same problem with molecular biology right now, and I agree with the track you’re taking. The issue seems to be the large amount of structure totally devoid of any semantic cues. For example, a typical textbook paragraph might read:
>
> **JS-154 is one of five metabolic products of netamine; however, the enzyme that produces it is unknown. It is manufactured in cells in the far rostral region of of the cerebrum, but after binding with a leukocynoid it takes a role in maintaining the blood-brain barrier – in particular guiding the movements of lipid molecules.**
>
> I find I can read paragraphs like this five or six times, write them on flashcards, enter them into Anki, and my brain still refuses to understand or remember them after weeks of trying.
>
> On the other hand, my brain easily remembers vastly more complicated structures when they’re loaded with human-accessible meaning. For example, just by casually reading the Game of Thrones series, I know an extremely intricate web of genealogies, alliances, locations, journeys, battlesites, et cetera. Byte for byte, an average Game of Thrones reader/viewer probably has as much Game of Thrones information as a neuroscience Ph.D has molecular biology information,

> but getting the neuroscience info is still a thousand times harder.
>
> Which is interesting, because it seems like it should be possible exploit isomorphisms between the two areas. For example, the hideous unmemorizable paragraph above is structurally identical to (very minor spoilers) :
>
> **Jon Snow is one of five children of Ned Stark; however, his mother is unknown. He was born in a castle in the far northern regions of Westeros, but after binding with a white wolf companion he took a role in maintaining the Wall – in particular serving as mentor to his obese friend Samwell.**
>
> This makes me wonder if it would be possible to produce a story as enjoyable as Game of Thrones which was actually isomorphic to the most important pathways in molecular biology. So that you could pick up a moderately engaging fantasy book – it wouldn’t have to be perfect – read through it in a day or two, and then it ends with “By the way, guess what, you now know everything ever discovered about carbohydrate metabolism”. And then there’s a little glossary in the back with translations about as complicated as “Jon Snow = JS-154” or “the Wall = the blood-brain barrier”. I don’t think this could replace a traditional textbook, but it could sure as heck supplement it.

This started a sobering line of thinking regarding the social structure of my medical school class. I can remember the names, appearances, voices, place of origin, specialties, and personalities of about a hundred of my classmates and faculty. It seems entirely possible that (again, byte for byte) I left medical school with more social data than medical data, and the social data seems less susceptible to decay from disuse. I forget which heart murmurs sound like what *all the time*, but the chance that I will forget my friend Leo was Japanese, liked Frisbee, and was dating my friend Angela seem next to nil, even though I haven't talked to him in a year.

The more I think about this sort of thing, the more I feel like I should be finding a way to exploit this kind of idea. Trouble is the isomorphism *just isn't good enough*. If I continue the conceit in the example of representing metabolic pathways by families, the story risks turning into a geneology, which is famously difficult to remember in itself (who but the most devout remember all the begats in the Bible?) And metabolism has a bad habit of looping in on itself, such that any attempt to historicize it is going to sputter into bizarre time travel arcs where people become their own grandfathers.

Still, if history has taught us anything, it's that people are really good at ridiculous constrained writing tasks. Maybe I could abandon the idea of there being "a" mapping, and just do something vaguely suggestive, so that glucose could be a character in one part and an artifact in another, so long as the story helped give a general sense of the glucose pathway.

Of course, the real benefits would have to wait for the blockbuster movie adaptation. I'm quite serious here. I thought about this when Ozy mentioned how ze doesn't watch *Game of Thrones* because then the actors will forever imprint themselves on zir mind and replace zir own imaginary view of the characters. If we had actors, people would remember what each character looked like, which would allow another level of encoding. All blonde people are lipids. All brunettes are amino acids. Everyone with long hair has a methyl group. Taller people are larger molecules. Et cetera.

Maybe there’s another field of study more amenable to this kind of treatment than biology. But as far as solutions in search of problems go, this is one of my favorites.

Fermat’s Last Stand: Soundtrack and Adventure Log

Two years ago an impromptu role-playing group put on a very successful *Dungeons and Discourse* adventure, complete with several musical numbers. Last week, we completed our first sequel to that adventure, *Fermat’s Last Stand*. Below is the log and twelve all-new songs. If you want to sing along, you can click through to the versions on the YouTube main site, which include the lyrics in the description (you may have to press “show more”).

Several groups played Fermat’s Last Stand simultaneously and got somewhat different experiences out of it. I’m basing this log on my own group’s progress, but adding some of the best features of the other groups’ campaigns and smoothing over some hiccups to make a sort of Platonic ideal. If anyone from the other groups didn’t finish yet, then for goodness’ sake don’t read further!

Credits and acknowledgements are at the bottom.

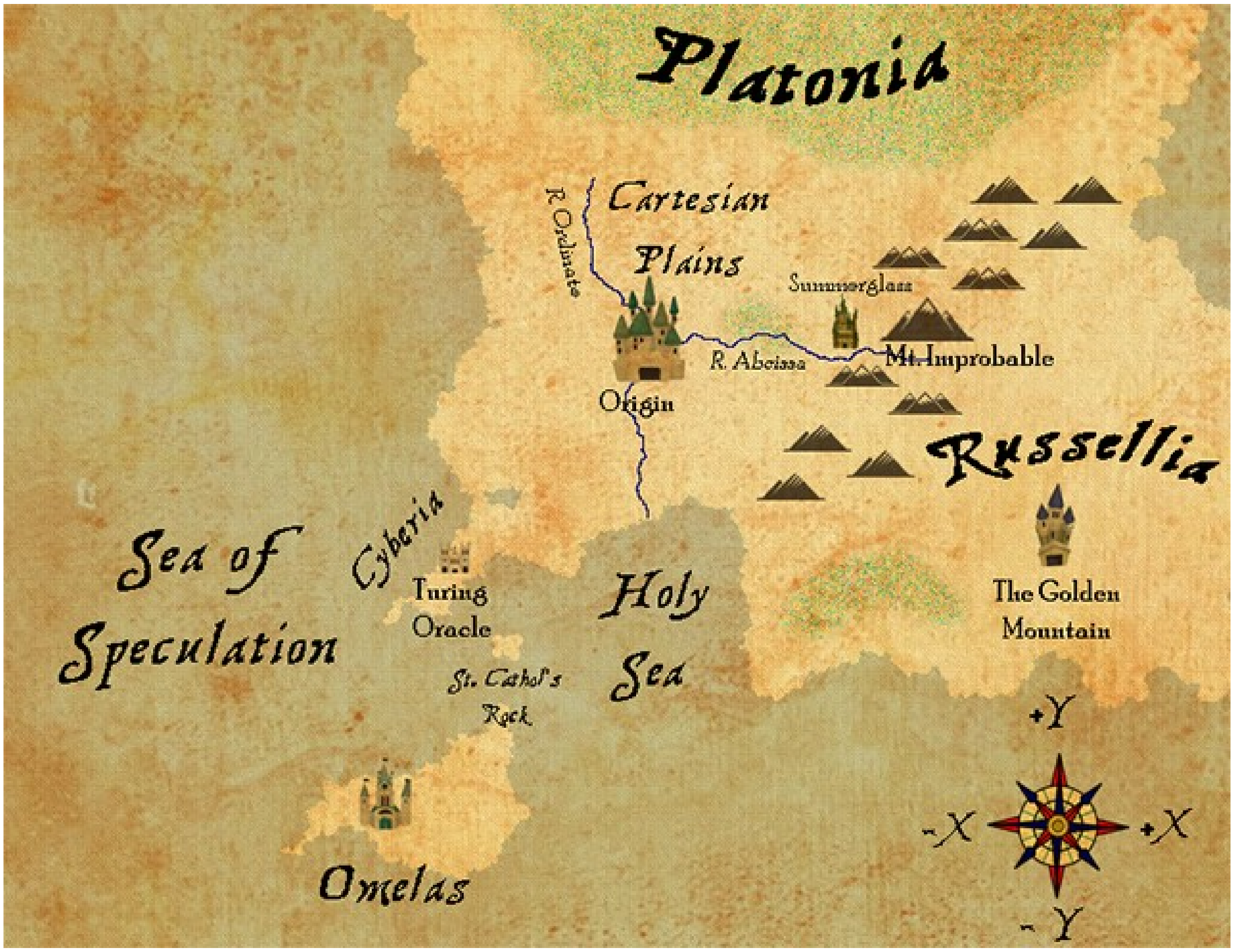
Prologue

Centuries ago, brilliant mathematician Pierre de Fermat came up with a proof so elegant that the God of Truth himself, mighty Aleithos, grew jealous. Taking inspiration from Fermat’s failure to fit his discovery in the margin of his notebooks, Aleithos afflicted the genius with the curse of macrographia: handwriting so large that his equations could never fit on any conceivable writing surface. Worse, when he tried to dictate his findings, his listeners were stricken with the same affliction. Constantly boasting of his proof but unable to publish it, Fermat was driven from land to land, until he eventually left civilization entirely and sailed across the sea, disappearing from history.

Recently, a shipwrecked man with curious features appeared in the court of Origin, saying he was the last survivor of a great civilization beyond the ocean, long since destroyed by plague. Centuries ago, his people had taken in Fermat and worshipped him as a god. And in their greatest temple, upon a titanic gold obelisk the size of a mountain, he had at last written his Theorem out in full. The man disappeared before he could say any more, but left behind a cloak in which were found a series of maps and directions written in a strange language.

Now people from all over the world are heading -yward, hoping to discover Fermat’s Last Obelisk and win unimaginable wealth and wisdom.

Fermat’s Last Stand


Platonia
R. Ordinate
Cartesian Plains
Summerglass
Mt. Improbable
R. Abcissa
Origin
Russellia
Sea of Speculation
Cyberia
Turing Oracle
Holy Sea
The Golden Mountain
St. Cahol's Rock
Omelas
+Y
-X
+X
-Y

Years ago, Princess Cerune and a hardy band of adventurers crossed Mount Improbable and solved the paradox that had trapped Bertrand Russell and his kingdom in a magical sleep. Upon awakening, Russell implicated the mysterious Bayesian Conspiracy in his downfall, and urged the Princess to gather more information on their activities.

Now Cerune shows up in Summerglass, looking for a new party to travel with. Her quest for the Conspiracy is taking her to the great city of Origin, which sits at the confluence of the rivers Ordinate and Abcissa in the center of the Cartesian Plain. At the very center of Origin stands the infinitely tall ZAxis Tower, and at its base lives the Wizard of 0 = Z, legendary omniscient ruler of Origin. According to rumor, the Wizard will answer one question from any petitioner, and

Cerune intends to ask him for information about the Bayesians. But the way to Origin will be long and hard, and she needs companions for the road ahead.

Six local philosophers take up her call. Markas, a physicist with a shadowy past. Chance and Carver, two ethicists whose values are in constant conflict. Alison, a radical feminist Marxist fleeing an embarrassing past as a princess. Cornelius, a hedge fund trader possessing the power of the free market. And Vivace, a local seminary student who gained mysterious powers after being bitten by a radioactive Jesuit.

But as they discuss their plans at a local tavern, others overhear their conversation, especially the part about “omniscient wizard will answer one question from any petitioner.” Daniel Dennett, a friend of Cerune’s from their last adventure, asks to join the party, explaining he has a particularly difficult puzzle to set before the Wizard:

Blaise Pascal, also an old friend of Cerune’s, disagrees with Dennett’s approach, explaining the problem *he* thinks they should get the Wizard to answer:

And suddenly a newcomer intrudes. Friedrich Nietzsche, who happens to be passing through Summerglass, explains that *everyone* is wrong, and tells the party the conundrum they ought to ask instead:

After some discussion, all three agree to join the party, which has now swollen to ten people. They stock up on supplies and leave Summerglass, heading -xward to Origin.

Their first obstacle is the Slough of Anarchy. This swamp was previously the Slough of Despotism until Cerune’s last adventure, when she and her companions killed the Leviathan who was oppressing the native people. But when the natives asked the regicides for help forming a replacement government, the few words of advice they gave – to avoid violence and oppression – got elevated into unbreakable laws and the Slough turned into a minarchy with avoidance of state force as its only guiding law. As a result, the whols swamp has become a lawless realm of bandits and con men, and Cerune & Co barely manage to make it alive to its capital of Malmesbury, where the former Hobbesgoblins (now just goblins) eke out a frontier existence.

Even worse, each night they fall asleep they are visited by some kind of apparition, a figure who angrily accuses them of summoning him, strikes at them with a powerful Prismatic Spray spell, and then disappears to parts unknown before our heroes can so much as counterattack. Distressed, Cerune splits from the party and decides to go to Origin on her own, while the others decide to press on and prove her wrong.

Their luck reaches its nadir the second night in Malmesbury, when their nocturnal visitor arrives and demolishes an entire floor of their hotel with his Prismatic Spray. But this time the party manages to corner him and force him to identify as Sir Isaac Newton. He describes the destructive Prismatic Spray spell as an incidental discovery in his attempts to puzzle out the nature of light:

This clue allows our heroes to finally work out how they are accidentally summoning him: Pascal sleeps on his large square shield each night, and one Pascal x square meter produces a Newton. Pleased that they have puzzled out the mystery of his summoning and freed him from these sudden accidental teleportations, Newton gives them a powerful artifact, the Prism of True Seeing, which he says will allow them to dispel magical illusions.

Encouraged by this victory, the party successfully kills a few alligators the next day and gets back in the black. Paying off their creditors, including the owner of the hotel damaged by Newton, they continue onward. But this requires navigating the complicated Slough railway system, an infrastructure dominated by scams and fly-by-night companies trying to separate unwary travelers from their money. The good news is that the city is full of Reputation Agents, businesses which for a price will inform you which other businesses are legitimate; the bad news is that some of the Reputation Agents are themselves illegitimate or bribeable, forcing the party to seek out an endless chain of Meta-Reputation Agents. Although they successfully solve several game theoretic problems and construct a framework for ensuring legitimate commerce in a minarchist society, they totally drop the ball on the object-level problem and end up on a train that breaks halfway through and strands them in the middle of the jungle, doomed to be eaten by snakes unless they can find some sort of railway engineer.

Markas puzzles out a solution to this problem that shares several features with the party's last stroke of genius. They place Pascal on a square-meter object to summon Newton. Then they push a cart into Newton at one meter per second; Newton, surprised, pushes back, transforming him into James Watt (1 Watt = power required to push an object at 1 m/s against opposition of one Newton). James Watt is able to fix the train, allowing them to continue onward.

They are just about to reach Origin when Nietzsche, who has been growing gradually more erratic, turns on them, hoping to prove his worth as a superior being to the Wizard by overpowering and defeating his comrades. There is a climatic battle in the train car, and Nietzsche is defeated just as they pull into Origin. The party rushes into the hustle and bustle of the city, hoping to avoid awkward questions about Nietzsche's death.

The great city of Origin is even more crowded than usual. Some of the crowd are treasure-seekers heading -yward on their way to seek Fermat's Last Obelisk; others are failed treasure-seekers heading back +yward after having been attacked or capsized by the dreaded Mathematical Pirates, who seem to be unusually active in the -yward seas. There is a long queue for people trying to meet the Wizard, but the party rejoins with Cerune, who is impressed they have made it this far, and finds she has obtained an appointment at the ZAxis Tower that very morning.

The Wizard of 0=Z is a form of blazing fire in the center of a vast audience chamber. His speech seems to shake the very earth, and all who see him are struck with terror. Dennett is the first to approach, and asks the Wizard how the brain thinks; the Wizard explains that it is made of *res cogitans*, a substance with the property of thinking. Next Pascal asks how the heart feels emotions; the Wizard explains it is made of *res sententia*, a substance with the property of feeling. The Wizard continues in this matter, giving each of the travelers' philosophical quandaries a Mysterious Answer, until finally Cerune asks him where she can find the Bayesian Conspiracy. Unable to deflect her question, the Wizard sets her a condition: he will only answer her question after she has brought him the broom of the Wicked Witch Of The Negative X Coordinate.

So the party – now minus Dennett and Pascal – travels -xward to the supposed castle of the Wicked Witch. But when they reach the -x coast, the "castle" turns out to be a cottage in a pleasant little village, and the "Wicked Witch" turns out to be Simone de Beauvoir, who invites them in for tea.

de Beauvoir explains that she used to work as the Wizard's maid, but that while hanging around him and his friends she began to pick up some magic. She learned she had a talent for it, even enchanting the Wizard's broom to clean on its own. When she realized none of the other magicians of Origin could perform such feats, she demanded a seat among them as a great thaumaturge. The people of Origin refused, saying women should stay in their place and not try to do mens' work. Enraged, she took her magic broom and left Origin along with a group of like-minded women to form this commune on the -xward coast. She is disappointed that most of the women in Origin continue to accept their oppression without comment, and hopes to one day lead a revolution among them.

She and Alison immediately hit it off, and everyone decides to join forces to cut the Wizard down to size. de Beauvoir, during her time with the Wizard, got the impression that a lot of his much-vaunted "magic" was mere illusion. She suggests the party head back to Origin and try to blow his cover.

So Cerune and the six philosophers head back to Origin, where they sneak into the ZAxis Tower under cover of night. They look at the Wizard through Newton's Prism of True Seeing, and see not a form of fire but a rather ordinary-looking man. When they confront him, he attacks them with fire, stones, and wild beasts, but the Prism of True Seeing allows the party to recognize all of these attacks as illusion, and they stand their ground.

Finally, defeated, the Wizard admits he is no Wizard at all. He is Rene Descartes, previously a philosopher in Origin. While investigating radical skepticism, Descartes had realized that all his sensory perceptions were being controlled by an all-powerful deceiving demon who had mastered the art of illusions. Descartes offered the demon a partnership, and together they took over Origin using Descartes' cleverness and his former tormenter's dark powers.

Cerune and her friends are ready to expose Descartes, but he points out that before his rise to power, Origin was poorly ruled and in the midst of a generation-long civil war. He claims that removing him would create a power vacuum that would plunge the city, now prosperous, into further bloodshed. After some discussion among the party's two ethicists, they agree and leave the Wizard in power, but not before making him promise to call off his grudge against Simone de Beauvoir.

Sadly, the Wizard has no idea where to find the Bayesian Conspiracy, and Cerune is not willing to give up on her quest. In the libraries of Origin they find rumors of a *second* omniscient entity who answers the questions of travelers: the Turing Oracle. They decide to journey -yward to Cyberia, where the Oracle is said to hold court.

This journey takes them to the -yward coast, where makeshift camps of treasure-seekers dot the landscape, and signs of the devastation wrought by the dreaded Mathematical Pirates mark nearly every town and settlement. Our heroes fall in with a band of treasure-seekers traveling -xward, and they start discussing the puzzling clues on the supposed map to Fermat's Obelisk. They are able to successfully decrypt some of the ciphers, which reveal the location of the fabled land of Terra Fermat as being beyond Omelas in the Sea of Speculation. Leaving the giddy treasure-seekers to pursue their discovery, the characters push on to Cyberia.

There they are told the Turing Oracle is no longer open for visitors. They break in anyway, but the Oracle explains through tears that the source of her power has been stolen.

When legendary computer scientist Alan Turing died, he uploaded his consciousness into a wearable ultracomputer placed around his corpse, the Shroud of Turing. This turned him into a superintelligent man-machine hybrid, and allowed him to cybernetically communicate his will to followers around the world. Now, all communication from Turing has gone dark, suggesting that the Shroud of Turing has been stolen. The Oracle says she will answer the party's questions if and only if they can recover the Shroud of Turing and restore her power.

Unfortunately, no one knows where Alan Turing has been buried, and therefore where the Shroud – and any clues to its disappearance – would be likely to be. Finally Cerune tracks down a local historian, who explains his hypothesis that Turing was buried in a church sitting on St. Cathol's Rock in the middle of the Holy Sea – a claim he calls the Church-Turing Thesis. Believing this sounds plausible, everyone charters a boat and heads toward St. Cathol's Rock.

The church, when they reach it, has been abandoned by all its former caretakers except a single monk who remains holed up in the library trying to finish a great masterwork on theology. Our heroes identify him as Thomas Aquinas through conversations like the following:

> **Vivace:** Wait, are you Thomas Aquinas?
>
> **Aquinas:** OBJECTION ONE: It would seem that I am not Thomas Aquinas. For the name "Thomas Aquinas" signifies a certain individual in medieval Europe on Earth, and this is neither Earth nor medieval. OBJECTION TWO: Further, Thomas Aquinas is a Saint, but men cannot be sainted until after their deaths, and I am still alive. ON THE CONTRARY, Scripture says in Exodus 3:14, "I am who I am". I ANSWER THAT one may be Thomas Aquinas in two senses. First, one may be the literal personage of

Thomas Aquinas, and thus be the medieval European monk. But creatures may in their essence resemble Thomas Aquinas more or less closely, and to those who partake the genus of Thomas Aquinas most closely, we can say by analogy that they, too, are Thomas Aquinas. REPLY OBJECTION ONE: I am Thomas Aquinas in the analogical, but not in the strict sense. REPLY OBJECTION TWO: The original Thomas Aquinas is dead, and therefore eligible for sainthood. Insofar as I resemble him, I am also a saint.

Vivace: Do you always talk like that?

Aquinas: OBJECTION ONE: It would seem I must always talk like this. For all creatures must behave according to their natures, and it is in the nature of Thomas Aquinas to analyze questions through logical disputation. OBJECTION TWO: Further, things strive to the perfection of their natural ends, replacing what is less good with what is more good. A properly ordered soul would not desire to reject what is better for what is worse. But I am a saint. Therefore I cannot reject a better mode of speech for a worse. ON THE CONTRARY, St. John Chrysostom condemns “those who use flowery language, not to inspire compunction, but to win vain praise.” I ANSWER THAT it is possible for me to stop talking like this. For to determine my proper actions, we must look to my proper end, and the proper end of humans is the glory of God. The end of speech, especially of religious speech, is to encourage proper reverence for the Divine, and any manner of speaking which does not compel the listener, or irritates them into departure, frustrates this end. REPLY OBJECTION ONE: It is not speaking this way qua speaking this way which is in the nature of Thomas Aquinas, but rather a desire for the glorification of God, from which speaking this way naturally proceeds. But if

> speaking this way were to fail to glorify God, it would be in the nature of Thomas Aquinas to speak otherwise. REPLY OBJECTION TWO: Speaking this way is good not in and of itself, but because it partakes of a greater end, namely the glorification of God.

This naturally annoys some of the harder-headed characters like Markas and Chance, who begin murmuring about the uselessness of theology and angels dancing on pinheads and such; Aquinas responds by defending the legitimacy of his chosen field:

Aquinas is originally reluctant to disclose the location of Turing's grave, arguing that disturbing the dead is a sin and that the Turing Oracle's prophecy reeks of witchcraft; Vivace is able to match him Bible quote for Bible quote and eventually wear him down. He directs everyone to the crypt of the church, where they find Turing's gravesite. It has indeed been broken into and despoiled, its Shroud gone. The only clues are a green feather and a protractor. Markas very reasonably decides that people who have parrots and protractors are very likely Mathematical Pirates. The latest reports place the Mathematical Pirates around Omelas, so they continue south toward that great subcontinent of fabled wealth, eventually getting caught in a storm and shipwrecked on its +yward shore.

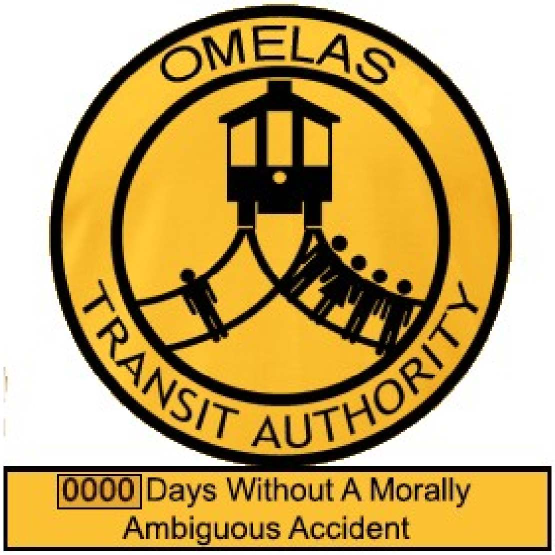
They stumble into a village in the hills nearby, only to find it has been struck by tragedy. The main export of Omelas is benthamite, a greenish metallic substance composed of pure utility. But yesterday morning the village's menfolk went into the benthamite mines for their daily labor but did not return. Now the women and children fear something terrible has happened. They ask our heroes to go investigate; the party's ethicists say they have no choice, and Cornelius agrees, motivated by the prospect of obtaining some of the world’s most valuable commodity.

The mines are dark, and the characters are likely to get eaten by grues. But they manage to fight off the grues (as well as their even scarier sisters, the bleens) and make it to the bottom of the mineshaft, where the village men are cowering. These villagers explain that they dug too greedily, and too deep, and awoke a chthonic being from primordial time – a Utility Monster. The monster has declared that it is trying to sleep and that the noise of footfalls bothers it, and politely asked everyone to keep quiet. Because its needs vastly outweigh those of all other beings, the menfolk of the village, good utilitarians all, have chosen to remain in the bottom of the mineshaft and starve to death rather than risk waking the monster up again by leaving.

Carver and Chance get into a long and heated debate about whether this is really the right thing to do. Unfortunately, the noise from the debate itself wakes the enraged monster, who decides to teach them a lesson by eating everybody. The party members flee and end up lost in the subterranean tunnel system (although they do end up gathering a few benthamite nuggets for their troubles). Finally, they find a corridor that looks sort of man-made, and follow it to the surface. They end up exiting a manhole and just barely avoiding getting hit by a trolley on a huge urban thoroughfare. They have stumbled into the City of Omelas, capital of its namesake subcontinent and one of the great metropolises of the world.

The City of Omelas, they soon learn, is surrounded by a protective bubble, a force-field that allows no one to enter or leave. Their sudden appearance is met with shock, interest, and fear, as they apparently wandered in through a previously unknown hole in the city’s defenses. While the

City Guard starts filling in the sewers, Cerune and Co decide to have a talk with the King and figure out why they are being kept in the city and what they can do to get out. They head to the Royal Palace, even though the city's transportation system does not exactly inspire confidence.


OMELAS
TRANSIT AUTHORITY
0000 Days Without A Morally Ambiguous Accident

Nevertheless, they make it to the Royal Palace with only a few dilemmas and even fewer deaths, and request an audience with King Jeremy I. The King is just as anxious to see them as they are to see him, and so our heroes are ushered into the throne room and treated to a stirring rendition of the Omelas National Anthem:

The King is sympathetic to their story, but tells them that letting them out of Omelas is completely out of the question. No one goes in or out of Omelas on account of the Sorcerer, the shadowy enemy who is besieging the city. The Sorceror knows the True Names of all things and so controls all things. If the force field protecting the city is suspended even for a tiny fraction of a second, the Sorceror will seize upon the breach and raze Omelas to the ground.

Alison asks how exactly Omelas produces such an unbreachable force field anyway, at which point King Jeremy becomes angry and expels them from his presence.

So they head off to the force field, looking to see if they can find any weaknesses or any routes out of there. There are none – but when Chance places his hand up against the field, he hears what sounds like an infinitely sorrowful voice, saying simply “Find me”. He conjectures that it is someone involved in creating the force field, and after bribing a bunch of guards, soldiers, and policemen with the benthamite nuggets they took from the Utility Monster’s lair, our heroes gain access to the force field generating station.

There they find a young child forcibly restrained and being horribly tortured in a variety of different ways. Vivace rushes in to free the child, but just before he can unlock the restraints, the child shrieks at him to stop. Surprised, they hold back and wait as he explains his story.

He is (he tells them), Prince John Stuart, the son of the king. When the Sorcerer attacked Omelas, the Omelasian wizards said that they could form an unbreakable magical defense around the city, but that the ritual required a commensurately awful sacrifice – the torture of an innocent child. Unwilling to ask any of his subjects to make such a terrible sacrifice yet unwilling to let his city fall, the young Prince volunteered himself. When his father the King objected, the Prince snuck out of the palace at night, offered himself to the wizards, and by the time the King learned what had happened the force field was already in place and could not be disrupted without terrible repercussions.

But John Stuart is now reaching the end of his ability to bear such pain, and when he gives in the force field will fall. He promises Cerune and her friends that he will risk briefly suspending the force field to let them out of the city if they promise to defeat the Sorcerer for him. They agree to this deal and sneak out to the city walls, where the Prince keeps his side of the promise and flickers the force field just long enough to let them escape. Despite the danger, the Sorcerer’s forces do not spot the flicker, and the city remains safe for the time being.

Sure enough, they see a great war-camp surrounding the city on its -xward side, and in its center a great dark tower. They enter the camp pretending to be mercenaries, and the ruse works well enough: they are hired and brought before the Sorcerer to swear allegiance. The Sorcerer asks the names of his latest recruit, and they all tell him – except Alison, who remembering the story about the Sorcerer controling things by speaking their True Names, gives a pseudonym. Sure enough, the Sorcerer demonstrates his power to control those soldiers whose names he knows, and he orders them to continue his effort to besiege Omelas.

While he is talking, his daughter enters the room. He introduces her as the Princess Mary, but makes no remark upon her complexion, which is completely gray from her skin to her eyes to her lips to her hair. Mary begs her father for more books on neuroscience, but the Sorcerer angrily orders her away, and it seems their relation is not at all a happy one.

The Sorcerer dismisses them, and the still-free-willed Alison helps the others shake off the Sorcerer’s control. None can think of anything they have learned that would help them assassinate the Sorcerer, but Markas has meanwhile fallen madly in love with the beautiful albeit monochrome Princess Mary, and proposes the idea of breaking into her room and stealing her away from her obviously abusive father. Although his plan is clearly emotionally motivated, it isn’t a bad idea – they need to figure out if the Sorcerer has any weaknesses, and his uncanny daughter seems like the best person to ask

So they break into the Sorcerer’s tower that night, bribing the guards with their last few pieces of benthamite, and find Mary’s room. Mary, in tears, tells our heroes her story. When she was born, an astrologer said that the day she fully understood the color red, her father’s kingdom would fall. Her father, alarmed, used his magic to place a curse on Mary, turning her gray and preventing her from seeing in color. This gave Mary a permanent and unrelenting hatred for her father, and she spends her days studying optics and neuroscience, trying to learn to understand the color red despite her inability to see it. Markas tries to get her to run away with them, but Mary declines, saying her quest for knowledge – and revenge – comes first:

But Markas gives her the Prism of True Seeing. In an instant, the curse is dispelled, and Mary is able to see red for the first time. Overjoyed, she promises to help the characters in their quest to overthrow the Sorcerer. She tells them his secret: the Sorcerer is an accomplished philosopher of language, whose inquiries into the meanings of words have allowed him to discover their True Names and use them to gain power over them. But his magic can be nullified through the use of his own True Name: Saul Kripke.

And so while Mary goes forth to start a mutiny among the soldiers, our heroes confront Kripke directly in his throne room. At first Kripke seems invulnerable, but Cerune and the others manage to dispel his magic by working his True Name into a song about his identity and his work:

With the Sorcerer’s magic defeated, Mary is able to organize an uprising among his soldiers. They come in and arrest him, and Alison convinces everyone to declare the Sorcerer’s newly liberated realm, which comprises the entire -xward part of the Omelas subcontinent, to be a People’s Republic. Mary, abandoning her royal garb, gets herself installed as

Premier, a peace is struck with the City of Omelas, and although Cornelius grumbles, the entire affair seems to be going well – until a rider comes from further up the subcontinent claiming that a village is on fire. With most of the military being occupied in protecting the fragile peace and evacuating their siege camp, the seven characters agree to investigate the attack and spread the good news of the liberation on the way there.

When they reach the coast, they stumble unaware into exactly the people responsible – a landing party of Mathematical Pirates. The pirates, sated with the last day's plunder, are not hostile – instead asking if the party's mathematicians can solve a certain dilemma they have gotten themselves into. This Markas does with little effort, impressing the buccaneers – maybe a little too much. At a word from the first mate, Infinitely Long John Silver, the Mathematical Pirates draw their swords, telling our heroes that they have been conscripted as the newest members of the pirate crew. Faced with the prospect of facing dozens of bloodthirsty pirates armed with highly advanced theorems and lemmas, and hoping to recover the stolen Shroud of Turing, they accept this with good grace and allow themselves to be marched to a secret rendezvous point.

There, they are taken on board the ship *Hypatia's Revenge* and brought before Captain Pierre Simone de Laplace, who declares them to be Official Mathematical Pirates. There is great rejoicing, the grog is brought out, and everyone joins together in a raucous mathematical pirate shanty:

The party, Captain Laplace tells them, have come at an especially auspicious time: the Mathematical Pirates are just about to leave known waters to sail far to the -ymost side of the world in search of Terra Fermat, drawn by the legend of the mountain-sized golden obelisk stating a simple and intuitive proof of Fermat’s Last Theorem. Laplace brings out the Shroud of Turing, and says that its ultracomputational powers provide the last link in the puzzle by allowing him to solve the cryptographic locks protecting the obelisk’s treasure chamber. But any hopes of stealing the Shroud and returning to Omelas are dashed when Laplace declares they set forth that very night. The *Hypatia’s Revenge* strikes forth into uncharted seas, and our heroes are lone among hundreds of cutthroat albeit extremely erudite buccaneers.

Not wanting to be drawn into such a foolhardy adventure, Chance suggests trying something desperate. Some of the other Mathematical Pirates don’t seem to share Laplace’s single-minded obsession with Fermat’s Last Obelisk. Perhaps if they murdered the captain, his replacement would be more amenable to abandoning the Shroud and returning to a friendlier coast? This leads to a bit of research trying to determine how exactly a dead captain is replaced, information which Markas finds in the Pirate Code under the heading “Laplace’s Law of Succession”. This law declares that if Captain Laplace is killed, whoever among the crew wins a trial of mathematical skill will succeed him. Markas is confident he can win that contest and take control of the pirate ship, resolving all their worries.

So our party breaks into Laplace’s chambers late at night, intending to kill him in his sleep. Unfortunately, his parrot notices and wakes the Captain up with his habitual cry: “Prime factors of eight! Prime factors of eight!” The captain, seeing that the seven assailants have him cornered, grabs the Shroud of Turing and teleports away to parts unknown using the ninth-level Logician spell *Infinite Egress*. Meanwhile, a horde of angry pirates breaks into Laplace’s quarters and asks what the party has done to their captain. Swords and protractors are drawn, and the situation looks grim.

Just as Markas and Infinitely Long John Silver are about to begin an extremely deadly mathematical duel, they are interrupted by a sudden hail of cannonfire. While distracted by the attempted mutiny, the Mathematical Pirates have allowed themselves to be caught unawares by their arch-enemies, the Empirical Pirates. Ambushed and without a captain, the Mathematicians are easy prey for their Empiricist brethren. Seeing which way the tide is turning, the party joins with the Empiricist attackers, and Markas personally sends Infinitely Long John Silver to Davy Hilbert's locker.

The party hopes the Empirical Pirates will be somewhat less set on foolhardy quests than their Mathematical colleagues. Markas calls for a parley with the Empiricist leader, famed natural-philosopher-turned-privateer Captain Hooke. But Hooke, having captured Laplace's trove of clues and maps to the great treasure, is equally set on its capture. Abandoning his own ship because of damage taken during the fight, he loads the remaining Empirical Pirates onto the *Hypatia's Revenge* and continues its voyage to the edge of the world.

After many days at sea, the ship reaches the shores of a strange land, filled with dense jungle and smoking volcanoes. Hooke says that his pilfered documents declare it to be the legendary Terra Fermat, and the Empirical Pirates and adventurers disembark. They push through the jungle, barely avoiding a host of colorful poisonous animals, until they come to a city of ruined ziggurats, strange idols, and a temple built into a volcano. The door to the volcano temple, disconcertingly, is wide open. Hooke leads the landing party inside.

The crew starts very quickly getting picked off by various nefarious traps that have been built in to the temple passages, and there are a series of odd puzzles and complicated mechanisms that fling those without the proper passkeys into lava, further decimating the group. By the time they reach the holy of holies, it is down to Captain Hooke, a few trusted lieutenants, and our plucky heroes, lost on in a strange land terribly far from home.

The center of the temple is filled with unimaginable riches in the form of gold, precious jewels, and works of primitive but disconcerting art. Yet there is no sign of the obelisk. As they push aside diamonds and emeralds looking for any hint of the Theorem's whereabouts, Captain Hooke gets felled by a mysterious arrow. So too do the few remaining pirate lieutenants. Before their attacker can kill the original seven, Chance casts the spell *Arrow Impossibility Theorem*, preventing arrows from being fired within a certain radius, and Cornelius casts *No-Ghost Theorem*, dispelling invisibility enchantments around people and objects.

There stands revealed Captain Laplace, previously hidden from view by his *Veil of Ignorance* spell. In his Villain Monologue, Laplace explains that Fermat's Last Obelisk is likely just a myth, a tale told to scare undergraduates. He himself had fanned the legend, posing as the original "shipwrecked sailor" who spread the story. His goal was to "crowdsource" certain odd riddles he had picked up from ancient texts, by inspiring a cottage industry of treasure-seekers desperate to solve them. When the party had unwittingly played into his hands by deciphering the clues, it had allowed him to pursue an entirely different treasure spoken of only in the oldest and mustiest books, a treasure beyond the very comprehension of the fools who chattered about Fermat's Last Obelisk.

He picks up a strange, glimmering sphere from the treasure chamber and tells the characters that it is the Eye of Aixitl, the most powerful computing device in the world – and as if to punctuate his point he throws the Shroud of Turing – which he had used to access the temple before Hooke had even arrived – on the ground derisively. In all they had done, he says, they had been the puppets of himself and of the other Secret Masters – puppets of the Bayesian Conspiracy.

Laughing maniacally, he declares: “You will always remember this as the day you almost caught Captain Pierre Simon de Laplace!”. Then he casts *Infinite Egress* and vanishes, cryptically telling his parrot to stay behind and take care of the witnesses.

The parrot morphs into a hideous and enormous hell-creature, Laplace’s Demon, and brutally attacks the party by concentrating the heat from the volcano into scorching projectiles. Alison, who has been fighting off a poison she contracted in the jungle, quickly falls to its onslaught, and Cornelius and Carver perish as well. The others fight back with everything they have. Markas uses *Tegmark’s Ultimate Ensemble* to flit among universes, making it impossible for him to die. Vivace casts *Summon Bonum* to call an angelic host to his defense. And Chance turns to Nietzschean metaphysics and casts *Eternal Recurrence*, allowing his party’s spells to repeat an unlimited number of times.

Laplace’s Demon responds with his own grimoire of Logical and Empirical spells, casting *Heisenberg’s Uncertainty Cloak* to disguise his location from view by declaring his speed, and *Cantor’s Diagonal Slash* to deal a different amount of damage each round. The increasingly desperate heroes counter with new spells of their own: *Sieve of Eratosthenes* to block attacks that deal non-prime amounts of damage, and *Inverse Communion* to turn their enemy’s flesh and blood into bread and wine.

After a truly epic battle the demon collapses with a fall that shakes the temple to its foundations. A wall falls over, showing the four surviving adventurers a sight not beheld for centuries.

In the heart of the volcano stands a golden obelisk the size of a small mountain, covered with vast hieroglyphic markings. Vivace uses his last few points of Will to cast a spell that allows him to translate the inscription. It reads:

No three positive integers a, b, and c can satisfy the equation a^n + b^n = c^n for any integer value of n greater than two. I have a discovered a truly marvelous proof of this, which this colossal golden obelisk is too small to contain.

Wounded, angry, and dejected, the four survivors limp out of the temple and back through the jungle. On the coast they discover even worse news: before departing entirely, Captain Laplace had taken the time to torch their pirate ship, leaving them stranded on the island.

Vivace struggles valiantly to try to mantain the virtue of Hope, while the more secular Markas has no such compunctions. Not only, he says, are they stranded on a desert island. Not only has their entire quest been playing into the hands of the Bayesian Conspiracy. Not only are three of their party dead, and Laplace escaped. But Fermat’s Last Theorem, the most beautiful result in all of mathematics, has been lost forever.

Cerune disagrees, saying that no knowledge is ever truly lost, and that just as the Theorem might one day be rediscovered, so they must not lose hope that their quest might bear fruit:

The story of Cerune, the Bayesian Conspiracy, and all the rest will be continued in a third and final adventure in this campaign, *Maximum Entropy Prior*. ETA is approximately “several years”.

Credits

Utmost thanks to everyone who worked to make this a reality / extremely compelling fantasy.

My special eternal gratitude goes to everyone who helped create the music. I accept full blame for the lyrics, but the vocals, recording work, and occasional instrumentals were handled by a whole crew of people. Buck Shlegeris and his band were responsible for the amazing “Mathematical Pirate Shanty”. Parts in “Philosopher Kripke” were voiced by myself and by Leah Libresco. And everything else was produced and sung by Michael Blume and by my lovely girlfriend Alicorn, who between them took all the other male and female parts.

Ari Rahikkala solved a few technical problems, Ezra B hacked together a sort of karaoke version of “Hymn to Him” for an emergency, and Nisan helped coordinate recruitment and publicity.

Grudging acknowledgement also goes to the players, named here in a haphazard combination of real and online names according to preference. Markas was voiced by David S, Carver by Sam B, Cornelius by Joe M, Chance by Berry, Alison by Ozy Franz, and Vivace by KarmaKaiser.

There were also several simultaneous runthroughs of Fermat’s Last Stand run by different DMs. None of them really communicated with me and I am not sure whether or not they actually took place. But if Charlie, Karl, falenas108 and gyukuro actually did their adventures like they said they would, they deserve major Virtue Points as well, as do their (supposed?) players Pleeppleep, Aretae, Romeo Stevens, Rick JS, Nathan from LA, tster75, Tarn Fletcher, JohnWBH, Ben from Oxford, Slow Learner, Avantika, and Thomas Eliot.

The readerships of Slate Star Codex, Less Wrong, and Unequally Yoked helped come up with some spell names and generally encouraged the creative process.

The Bayesian Conspiracy was stolen from Eliezer Yudkowsky and Harry Potter and the Methods of Rationality (read it! No, seriously, read it!), large portions of the plot were stolen from The Wizard of Oz, songs were stolen from their respective sources, the Omelas Transit Authority symbol was stolen from Leah and portions of the Mathematical Pirate Shanty were stolen from various sites with mathematician/pirate jokes that somehow actually already existed because the Internet is amazing.

If you liked this adventure, you may want to read the prequel as well, or use the associated Dungeons and Discourse Rulebook to create an adventure of your own.

Things I Don’t Understand About Genetics (A Non-Exhaustive List)

A couple months ago the Genetic Association Consortium’s study on SNPs for intelligence raised an important question: should all of our genetics studies be performed by organizations whose acronyms are also amino acid codons? And aspartic acid? Really? Kind of a boring choice.

But while we’re figuring that out, can someone explain to me how polygenic inheritance works?

We have really strong evidence that intelligence is highly heritable – maybe 50% to 80%. But genome-wide association studies show very low contributions from any particular SNP:

> The study found that the genetic markers with the strongest effects on educational attainment could each only explain two one-hundredths of a percentage point (0.02 percent). To put that figure into perspective, it is known from earlier research that the SNP with the largest effect on human height accounts for about 0.40 percent of the variation.
>
> Combining the two million examined SNPs, the SSGAC researchers were able to explain about 2 percent of the variation in educational attainment across individuals, and anticipate that this figure will rise as larger samples become available.

Here’s my question. Things determined by a larger number of independent randomly varying processes should tend to vary less. Suppose at Casino A, you flip a coin and if it comes up heads they give you $5000 and if it comes up tails they give you nothing. But at Casino B, you flip 5000 coins, and get $1 for each coin that comes up heads and nothing for each coin that comes up tails.

Both games have the same minimum and maximum winnings ($0 and $5000), and both have the same average ($2500). But the variance will be very different. The winnings at Casino A will vary *a lot*; half the people will walk out with $5000 and the other half will walk out broke. The winnings at Casino B will vary surprisingly little: if I understand binomial distributions correctly, well under 1% of gamblers will walk out with less than $2400 or greater than $2600. Pretty much everyone goes home with *something like* $2500.

The more independent loci determine human intelligence, the less variation in human intelligence we expect to see relative to the total amount it is possible to vary. If the most important SNP explains 0.02%, then there are at least 5000 genes involved, which means, as in the example above, that less than 1% of people should differ more than 2% of total possible variability from the average.

But actually, people differ in intelligence *a lot*, and a lot of that difference seems to have a genetic component. This implies either that the total possible genetic variability in intelligence is *huge* – that the right genes could give you an IQ of 2500 or so, or that variability in intelligence isn’t simple and additive and random the way I’m modeling it here – or that I’m doing the math wrong, always a distinct possibility.

The easiest way to get out of this, other than accepting I am terrible at math and should be kept away from it, is to assume that lots of different genes for good intelligence are correlated. Maybe one population had reason to evolve high intelligence with lots of smart SNPs, and a second population didn’t. Then it would make complete sense that all the genes involved would co-vary. Unfortunately, there seems to be significant IQ variation within the same family, let alone within the same population group, so that fails pretty hard as an explanation.

Another possibility is to accept the whole mutational load idea – which allows for high correlations in goodness or badness of the entire genome. Unfortunately, this is the *other* idea in genetics which has been confusing me terribly over the past few months.

This has no trouble explaining correlations, but it does have trouble explaining why things aren’t *more* correlated. We should find the same people being very smart and very tall and very athletic and very healthy. I don’t doubt there are some correlations between these traits, but they don’t seem nearly as high as one might expect.

And the paternal age effect keeps being brought in to explain this, but I don’t get that one either. Suppose I have a kid at 60. My sperm and my DNA have had 60 years to accumulate deleterious mutations, so there’s more chance my kid will have low IQ or psychiatric disease or whatever. Fine. But suppose while I’m having a kid at 60, my twin brother has had a kid at age 20, and his kid had a kid at age 20, and *his* kid had a kid at age 20, so that his first great-grandchild is being born exactly the same time my first child is. Both my kid and his great-grandkid have had 60 years worth of cell dividings to accumulate mutations. Why should their risks of autism be any different just because his kid had those 60 years divided among three different people?

(does less frequent division of spermatogonia before puberty cause them to accumulate fewer mutations during that time? If so, shouldn’t three generations of people who have kids at 33 still accumulate 80% as many mutations as two generations of people who have kids at 50?)

I’ve heard evolution’s ability to eliminate people with bad genes used as an explanation here, but I don’t quite get it. For one thing, it seems unlikely that evolution can produce beneficial mutations at the same rate people accumulate deleterious mutations (10 per generation or so). And if we imagine two lineages of gradually deteriorating intelligence going on for 1000 years, the Lions That Selectively Eat Low-IQ People will have just as much opportunity to cull the members of the one that reproduces quickly as the one that reproduces slowly. *And*, if that were the explanation we should fail to see a paternal age effect in the absence of such lions (or be dysgenic as hell), but the effect has been demonstrated in our own society, which is relatively free of lions and of almost everything else that kills people before they can reproduce.

I know other people have blogs where they explain things and readers bask in their wisdom, but having a blog where I say how confused I am about stuff and readers explain it to me has always worked for me before and I have faith it will continue to do so.

Fake Consensualism

[*A previous version of this essay was posted incomplete. This is the full version*]

I.

Dear friend, have you considered banning health care?

Several studies agree that sick people are often treated against their preferences and sometimes against their explicit requests. 51% of patients who request Do Not Resuscitate orders never get them. Among those who get them, 75% of their doctors don't know about it. And among doctors who do know about it, 65% don't feel a need to follow them if it conflicts with their own opinions.

So it should come as little surprise that 22% of people who sign forms saying they don't want to be hospitalized get hospitalized and 31% of people who sign forms saying they don't want CPR get CPR.

Or that patients who prefer to limit life-prolonging treatment get an amount of life-prolonging treatment that is statistically indistinguishable from everyone else.

Or that even though 80% of Americans say they want to avoid hospitalizations and intensive care as they are dying, in fact about 50% of Americans die in hospital and intensive care.

But in fact, the situation is far more dire than this. These studies only count people who have legally, explicitly spelled out a desire not to receive health care. What of people who sign off on all the dotted lines, but only under emotional blackmail that can hardly be called consensual?

The cancer patient who gets told she is “brave” and “a fighter” if and only if she accepts all the interventions modern medicine has to offer, but otherwise gets given “pep talks” on how she can’t “give up” on her family and friends.

The heart disease patient whose family constantly implies that if he really cares about them he’d “push on” as long as he could.

The elderly Muslim who consents to a massive and hopeless surgery only because she doesn’t want her community to remember her as a sinner who rejected God’s gift of life and who is burning in Hell for all eternity.

So although health care produces more than its share of obvious and legally binding consent violations, we have every reason to think this is only the tip of the iceberg, that countless millions of people who don’t want health care are pressured and bullied into accepting it and undergoing unnecessary suffering against their deepest wishes.

And the obvious solution, dear friend, is to ban health care. If there is no health care, no one can be coerced into receiving health care against their wishes. We tried having a health care system that operated on the principle of informed consent, it failed terribly, and now we must pass laws imprisoning anyone who provides any form of medication, surgery, or any other treatment attempting to alleviate disease or prolong life. While doctors, nurses, et cetera may be well-meaning, the real (and indeed realized) risk of nonconsensual treatment is far too dire to allow this so-called “humane” practice to continue.

Now you may think to yourself: this seems like it would be politically unpalatable. If the Republicans obstructed

President Obama at every turn merely because he wanted to *slightly modify* the health system, what would they think of a law that would dismantle the health system entirely, and indeed threaten large fines and jail time to anyone trying to assume its functions?

This would indeed be a significant complication if not for the fact that the Republican Party must in any case very soon be eliminated forever.

Perhaps you recall the 2000 presidential election? That one with the whole rigamarole in Florida, with the hanging chads and so on? Several thousand Democrats tried to vote for Al Gore, but the ballot was confusingly designed and they ended up voting for Pat Buchanan.

Now, the right to vote is absolutely sacred, and to have your vote unfairly shifted to another candidate is an egregious violation. Imagine how those poor Miami-Dade residents – most of them elderly liberal Jews – must have felt when they were told their votes went to an uberconservative Christian often accused of anti-Semitism and Holocaust denial. Dear friend, was that fair?

No Democrat should have to go through that. Yet as long as there are non-Democrats on the ballot, inevitably some Democrat – whether through poor ballot design, poor reading comprehension, or simple bad luck – will accidentally vote for them.

Even if they *say* they intended to vote for a non-Democratic candidate, can we really trust them? Perhaps they come from a very conservative area, where liberals risk violence or ostracism, and in order to maintain their relationships they must vote against their deeply held beliefs. Maybe their bosses have threatened to fire them if they vote

Republican. It is a very real possibility, and quite frankly our nation's Democrats deserve better.

The only way to protect the sacred right to vote for the candidate you truly believe is best, without the threat of accidental mishap or outside pressure distorting your choice, is to remove all non-Democrat candidates from the ballot. That way, no one can force a Democrat to vote for another party's candidates, and we can rest secure knowing that the wishes of our nation's electorate are being respected.

Except that some people don't like Democrats – which is why in order to ensure *absolute* fairness and consensuality, instead of only including Democrats, the ballot will actually just include candidates from my own party, the Ban Everything That Anyone Might Experience And Later Regret (BETA-MEALR) Party.

Yes, we want to ban health care, but we're about so much more than that! We also oppose gay marriage, and straight marriage to boot. We are against prostitution but also all other forms of employment *and* all other sex (but especially BDSM). Also religion, vasectomies, all social gatherings, gender roles *and* breaking out of gender roles, gift-giving, telephones…

…what is that? You say this fundamentally misunderstands consent, that it tramples the delicate balance between "freedom to" and "freedom from"? What balderdash! And here I thought you were my dear friend! How could you…oh, I believe I know what your problem is. You are being pressured to say this; some force not your own has pushed you to criticize my remarkably astute policy proposals despite a clear hidden preference against doing so. Do not worry, dear friend! I will save you! Once we get elected, criticizing our policies will be forbidden on pain of death, giving everyone complete freedom to avoid outside interference and express their honest positive opinion of us at last!

BETA-MEALR 2016: because your preferences can only be violated if you were allowed to have them in the first place!

II.

Even though most people dislike libertarians, their basic insight – that you can get a lot of the way to morality by saying things that people consent to are okay and things people don't consent to are bad – holds a pretty eminent place in moral discourse, especially on the Left. This is with good reason. In a country where no one really agrees on what exactly morality is, consent is a useful semi-neutral principle that allows everyone to go off and do their own thing.

I've written a bit before about fake consequentialism, where people who have nonconsequentialist reasons for what they believe make up consequentialist ones so they can appeal to the general populace. I don't think *anyone* has ever been convinced gay marriage is wrong because of sincere concern about the fate of children adopted by gay couples, but that argument keeps getting trotted out again and again because it's one of the few anti-gay-marriage arguments that sounds remotely consequentialist. Or people who dislike porn claiming it will encourage viewers to rape people, even though as far as anyone can tell exactly the opposite is true.

And where there is fake consequentialism, not far behind you will find *fake consensualism*. Suppose there's something you don't like, but every time you argue against it, people say “Well, it’s all consensual and it doesn’t harm anyone except the people who have agreed to it, so mind your own business.” You can come back with “Yes, but how can we be sure the people involved in it *really* consented? Deep down? I bet they didn’t!” Or even “I bet this would lead to something nonconsensual happening somewhere else down the line!” Consensual BDSM? Just going to glorify abuse and lead to more nonconsensual abusing, right?

I used to automatically steelman BETA-MEALRs into proper utilitarian arguments. “Well, euthanasia produces a gain of utility, by allowing people who want to die to do so. But it also produces a loss in utility, because it might unfortunately result in some people who didn’t want to die doing so. Who could *possibly ever know* which of these would be greater? So I guess all we can do is play it safe, right?”

But at some point, I started wondering how likely it was – even in cases where we genuinely have good reasons to worry about mistakes or pressure – that allowing people to choose whatever they preferred would result in *fewer* people getting what they want than banning all choices except one.

I mean, okay. You say we have to ban euthanasia, because if it’s not banned, some people who don’t want to die might have to. Okay. I come back and say “But if we ban euthanasia, some people who don’t want to live might have to. Except wait, no, not some people. All people. And not might have to. Definitely will have to. Required by law.” You say “Get the hell out of my office.”

And I tried to steelman even this. Well, I told myself, there are more people who don’t want euthanasia than who do want euthanasia. And it would be a really, really big loss in utility if we killed people who didn't want to die. Maybe the numbers still check out.

Except when I tried to check the numbers, it was immediately apparent that they didn't. There are loads of good data from the Netherlands we can use to estimate how nonconsensual and slippery-slopey real world euthanasia is, and the answer is statistically indistinguishable from zero.

(you may have heard the usual suspects claim that "studies show 92.7% of Dutch males die by being seized by goons and euthanized against their will" or something. This is based on *extremely* creative readings of the data, but proving this is a job for another blog post.)

I no longer try to steelman BETA-MEALR arguments as utilitarian. When I do, I just end up yelling at my interlocutor, asking how she could possibly get her calculations so wrong, only for her to reasonably protest that she wasn't making any calculations and what am I even talking about?

And yet without the utilitarian angle, this whole argument falls apart on exactly the "proving too much" grounds pushed by our hypothetical politician above. If you want to ban euthanasia, why not ban health care? If you want to ban prostitution, why not McJobs? If you want to ban BDSM, why not all consensual sex? If you don't have a good quantitative argument ready, you sure can't support it on qualitative grounds alone.

Look around you. Just look around you. Have you figured out what we're looking for yet? That's right. The answer is sacred values and taboo trade-offs.

III.

According to Philip Tetlock’s dichotomy, a sacred value is a value in Far Mode, one that has big flashing signs saying “MORALITY!” around it. A secular value is one that is nice but not morally important, like saving money or increasing productivity.

A taboo tradeoff is when someone asks you to trade a secular value off against a sacred value, and tends to get people morally up in arms.

For example, suppose I asked you whether the government should spend a recent windfall of ten million dollars on a public education campaign about avoiding dangerous chemicals which is projected to reduce the incidence of cancer by 300 cases per year, or on a renovation of Yellowstone National Park. Different people would have different opinions, and no one would think it was an especially odd question.

But suppose I asked you whether the government should spend a recent windfall of ten million dollars paying for years of extremely expensive hospital care for an impoverished nine-year-old girl who is critically ill with cancer – or, again on the renovation of Yellowstone? Obviously you HAVE to do the former and to do the latter would be *inhuman*.

The inhumanity here is not in making a spectacularly bad choice. After all, the choice in the first case was even starker than in the second – saving 300 cancer patients is clearly better than *maybe* saving just one cancer patient. Yet the first choice offers to trade one secular value against another, while the second choice offers to trade a sacred value (life) against a secular value (national parks).

BETA-MEALR arguments offer taboo tradeoffs. BDSM opponents don't care if a ban prevents one billion people who really really wanted to do BDSM from getting to do that. It's wrong to have even *one* person be pressured into something they will later decide felt like abuse.

(Three hundred people? No problem, that's just a statistic! But *one* person? SACRED VALUE SACRED VALUE SACRED VALUE.)

The same is true of euthanasia. It's not that there's any difference at all between the argument that would ban euthanasia and the argument that would ban health care. It's just that not allowing even the tiniest possibility of a person involuntarily being euthanized is a sacred value, but not allowing even the tiniest possibility of a person involuntarily getting health care isn't.

Now, I don't like the idea of sacred values much, but other people do, and merely saying "You're treating life as a sacred value!" isn't going to make people start crying and protesting "I'm sorry! I didn't mean to!" Mostly they'll just say "Yeah, life is a sacred value to me." Okay, fine.

What bothers me is when they do this *and* they pretend they're making a value-free statement about respecting the rights of others. "Oh, well, we're a liberal democracy and people should be able to do whatever they like with their own bodies, but I'm just worried about people being euthanized against their will, and that would be a violation of consent, and a good liberal democracy like us wouldn't want to violate consent, no sirree!"

No. You do not care how many people are kept alive without their consent, just like you do not care how many people work McJobs without their consent, or how many people feel pressured into going to social gatherings they don’t want to attend. You care about consent solely when it serves the purpose of your sacred values. You would gladly violate the consent of a billion people on some unrelated issue rather than risk a single consent violation of your own personal pet project.

When you start preferring that every single Republican be disenfranchised rather than risk the disenfranchisement of a single Democrat, you can no longer credibly claim to be standing for the sanctity of the electoral system. You’re standing up for the Democratic Party, plain and simple, and “sanctity of electoral system” is your convenient excuse to people who would otherwise have a dim view of “let’s disenfranchise all Republicans because Democrats are better”.

Likewise, when you start preferring that every single person who desires euthanasia be thwarted rather than risk thwarting a single person who wants to live, you can no longer claim to be standing for the principle of concern for people’s preferences. You’re standing up for your religion, or your personal revulsion at seeing people die, or whatever else, plain and simple. Concern for people who might get forcibly euthanized is your convenient excuse to people who would otherwise have a dim view of “let’s force everyone to abide by my religion’s views.”

One possible counterargument: “But people dying by accident is so much worse than people living by accident!” Well, says your value system. I actually happen to feel the opposite (the reasons why are complicated, but this comic is a pretty good start). Which of us is right? I don’t know, but “I can use my value system when deciding about my own life, you can use your value system when deciding about yours” seems a heck of a lot less horrifying than a lot of the alternatives. *I* happen to think voting for a Republican by mistake is far more horrifying than voting for a Democrat by mistake, but insofar as we're a democracy that choice is for each of us to make without interference.

Certainly people being pressured or dragged into something that on the surface appears consensual is a genuine problem. When people have good suggestions for how to increase the certainty that consent is genuine – like the sex-positivity community's friendly reminders not to have sex with really drunk people – I am all in favor of this. And the one intervention that would go further than any other into ensuring all acts of consent are genuine – giving people a Basic Income Guarantee so that no one gets forced to consent to *any* unpleasant job or relationship out of fear of starving to death – well, I'm in favor of that too.

But when people start saying we should *ban the thing entirely* because they can never be totally sure it will be utilized 100% consensually, or won't cause a slippery slope that leads to consent violations in some vague and hypothetical future – well, I support a moratorium on using that argument. The very rare cases where it might be true don't seem worth the extreme latitude it gives everyone to declare that they are outraged at consent violations that might conceivably occur if we don't support their pet project to abolish all possibility of consent on a certain topic. At the very least, anyone so arguing should be required to meet an extraordinarily high burden of proof, like having numbers that aren't made up.

Fake Euthanasia Statistics

I promised I'd write something clearing up statistics on euthanasia in the Netherlands, so let's start with Anne McTavish: The Netherlands' Euthanasia Stats Are Appalling. She writes:

> I want to focus on the doctors and the rest of us who will be harmed if [a Canadian case on euthanasia] becomes law. We don't need to speculate. The Netherlands has already gone down this slippery slope and provided the grizzly statistics that should stop us going down the same path.

Now, I will be the last person to deny that there are indeed some grizzly statistics on euthanasia. For example, did you know the most recent grizzly to be euthanized in Yellowstone had twenty five instances of aggressive behavior before finally being caught? Or that a grizzly euthanized in Montana two months ago had killed eight sheep and injured two more? Or that a a grizzly euthanized last month in Fresno was 38 years old, the oldest living bear in an American zoo?

Yet McTavish inexplicably ignores this vast trove of useful information and chooses to focus on highly questionable medical data instead:

> In 1990, 130,000 people died in the Netherlands: 2,300 people asked doctors to kill them; 400 asked doctors to provide them with the means to kill themselves; 8,100 died when doctors deliberately gave them an overdose of pain medication to kill them (for which 4,941 patients didn't consent); 1,040 people died when doctors euthanized them without their knowledge or consent

> (72 per cent of those never having given any indication they would want their lives terminated).
>
> That's breathtaking in more than one way.
>
> It's not so much that nine per cent died at the hands of doctors, which is alarming in and of itself. What should raise our cries of outrage is that 4,941 people (four per cent) did not give their consent to being killed. A doctor who operates on someone without their consent can be successfully sued and made to pay huge dollars for having done so. The same should apply for killing a person without their consent.
>
> And it's the 1,040 people (one per cent) who were killed without their knowledge or consent and the 749 who never wanted to die early that should get us up in arms.
>
> "In only 23 years, Dutch doctors have gone from being permitted to kill the terminally ill who ask for it, to killing the chronically ill who ask for it, to killing newborn babies in their cribs because they have birth defects, even though by definition they cannot ask for it. Dutch doctors also engage in involuntary euthanasia without significant legal consequence, even though such activity is officially prohibited," writes Wesley J. Smith in Forced Exit: The Slippery Slope from Assisted Suicide to Legalized Murder.

II.

So here's where it gets confusing. The article doesn't cite its sources, but a quick Googling finds that all its numbers come from an article by the "Patients Rights Council", which *does* cite its source as the Remmelink Report, a huge report by the Dutch government in 1990.

The Remmelink Report is not on the Internet, and in any case it is in Dutch. However, an English language paper summarizing the results was published in the Lancet, and an English language digest of the report is on Google’s cache.

The numbers in the digest and the Lancet article are somewhat different from the numbers in the article, and I was somewhat uncertain who to believe – the people who seem to have a much more complete copy of the report in the original Dutch, or my own lying eyes?

Luckily, some people who ought to know have weighed in to guide me. Some of the authors of the Remmelink Report, in a biting English-language article called “Dances With Data”, complain about constant misrepresentations and lies about their research and request that “we can only hope that the truly interested reader will take the trouble of reading the Lancet article or the English version of the report itself, instead of relying on this faulty description.”

So I’m going to take this as carte blanche to trust that the Lancet article and the English language digest of the report correctly capture the report’s findings and to start criticizing these other articles based on that. So:

Both the McTavish/Patients’ Rights articles and the Lancet version of the Remmelink Report agree that there were 2300 cases of active voluntary euthanasia and 400 of physician assisted suicide.
These are the cases we would classically treat as euthanasia. The report says that they were all done consensually, and in fact this is tautological as if they had not been done consensually they would have been classified as murder rather than euthanasia. Of these cases the report (but not the article) clarifies that:

> We found that in the Netherlands over 25 000 patients per year seek assurance from their doctors that they will assist them if suffering becomes unbearable. Each year there are about 9000 explicit requests for euthanasia or assisted suicide, of which less than one-third are agreed to. In most cases alternatives are found that make life bearable again, and in some instances the patient dies before any action has to be taken.
>
> In 96% of instances of euthanasia and assisted suicide the physicians stated that the request of the patient was explicit and persistent. (The finding that there always was a request of some sort is circular since such a request is at the core of the definition of euthanasia in the Netherlands.) In 94% the request had been made repeatedly; in 99% the physicians felt sure that the request had not been made under pressure from others; and in all cases they were convinced that the patient had sufficient insight and knowledge of the course of disease

Along with these cases, both report and article agree that there were about 1000 people who were killed by physicians without “a clear and explicit request from the patient”. However, the article simply leaves it at that, whereas the report adds:

> “In 59% of life terminating acts without explicit request the physician had some information about the patient’s wish [eg the patient had previously asked to be euthanized if a situation like this one arose]; in 41% discussion on the decision would no longer have been possible. The physician (specialist or general practitioner) knew the patient on average 2.4 years and 7.2 years, respectively. Life was shortened by between some hours and a week at most in 86%. In 83% the

> decision has been discussed with relatives and in 70% with a colleague. In nearly all cases, according to the physician, the patient was suffering unbearably, there was no chance of improvement, and palliative possibilities were exhausted.”

So baaasically, patients with only a few hours left to live who are in comas and have usually previously asked their physicians to euthanize them and whose family members have all said this is what the patient would have wanted and who have been evaluated by a bunch of doctors who say there’s no way they’re waking up – sometimes these people get mercy killed. It’s unclear to me whether this was legal in the Netherlands at the time or not, but it’s hard to call this grisly.

The articles and report then go into two more categories which are a little more worrying:

First, there are “22 500 cases where the patient had died after the administration of drugs to alleviate pain and symptoms in such dosages that the risk of shortening the patient’s life was considerable. In 6 per cent of these cases life-termination was the primary goal and in the remainder it was the secondary goal. In about 40 per cent of these cases the decision to increase drug dosages and the possibility that this might hasten death had been discussed with the patient. In 73 per cent of the cases where these matters had not been discussed with the patient, the patient was incompetent.” According to another Lancet paper, “In 85 percent of all such cases, the physician said that he or she had no intention of hastening death, but had taken into account the probability or certainty that death would occur. In 64 percent of cases the physician estimated that the patient’s life had been shortened by less than 24 hours, and in 16 percent it was shortened by less than one week”

This situation is pretty common everywhere. The patient's clearly on her deathbed and in terrible pain. The physician says "We need to up the dose of morphine to keep her comfortable." The intern says "We can't up the dose of morphine! She's already close to the maximum dose of morphine my pharmacology book says is safe for her body weight!" The physician just glares, and says "What's the worst that could happen? She dies an hour from now instead of a day from now? That'd be a f@*king tragedy, wouldn't it? Just give her the damn morphine."

These numbers provide no evidence that any of the six to fifteen percent of cases in which life shortening was a significant goal overlap with the cases in which the patient was not asked for consent. Actually, they don't even provide any evidence that the patient died *because* of these drug doses, just *after* them (and it seems pretty reasonable that death comes after being given drugs for a rationale that includes a person being on their deathbed). I'm classifying this as "worrying" only because we can't prove from these numbers that there was no one who should have been asked for consent but wasn't.

The second of these two categories is withholding of treatment. The report says: "There were 22 500 cases where death had resulted from non-treatment decisions (ie withdrawal or withholding of medical treatment), representing 17.5 per cent of all deaths. In 30 per cent of these cases the non-treatment decision had been discussed with the patient. In 62 per cent of these cases it had not. In 88 per cent of all cases where the non-treatment decision had not been discussed with the patient, the patient was incompetent."

Here we can calculate that in about 7% of these cases, decision to withhold treatment was not discussed with a competent patient. However, the authors of this study have mentioned elsewhere that they include futile treatment in these numbers – ie treatment that wasn't working or was just making the patient worse. They also mention that these treatments included feeding tubes, ventilators, et cetera – so that this "withdrawal of treatment" is just what we would call "pulling the plug" in an American hospital. Once again, I include this statistic in the "worrying" category only because we can't prove from these numbers alone that non-futile treatment was never withdrawn without consent.

But here's the thing: giving near-death people dangerously high drug doses and withdrawing medical treatment without patient consent *are not classified as euthanasia and are both completely legal here as well*. In fact, the Remmelink Report says outright that "The Dutch data on medical practices which shortens life, in the cases of non-competent or of competent but not-consulted patients, are indeed a matter of concern… [but there] really is not a shred of evidence that the frequency of this sort of behaviour is higher in the Netherlands than, for example, in the United States; the only thing that is clear is that more is known about it in the Netherlands. In short, there is no reason to assume … a causal relationship between limited legalisation of euthanasia and 'lack of control' over other sorts of medical behaviour." These sorts of non-euthanasia incidents are where nearly all of the so-called "euthanized without consent" numbers are coming from – the article's claim of 4941 Dutchmen "euthanized without consent" is actually greater than the *total* number of Dutch people euthanized in the study.

So here are the differences between the statistics as presented in the grizzly article & its Patients' Rights inspiration, and the statistics as presented in the Remmelink Report:

The numbers themselves are pretty different.

The article classifies patients who asked to be euthanized if they ever had X happen to them – and then had X happen to them, and because of X were unable to explicitly consent to euthanasia at that moment – as “euthanized without consent”.

The article classifies patients whose doctors gave them unusually high doses of drugs with the aim of providing relief for pain or symptoms but without the primary goal of ending their lives as “euthanized without consent”.

The article classifies patients whose doctors withdrew treatment, including futile or invasive treatment, as “euthanized without consent”

The article classifies patients who were unable to consent, for example in terminal comas, but whose families said they would have wanted to be euthanized in such a situation – as “euthanized without consent”.

The article classifies patients whose lives were shortened by only a few horrible hours as “euthanized without consent”.

The article classifies several categories of actions that are not euthanasia and which are currently both legal in and widely practiced in the United States as “euthanized without consent”.

And then – surprise, surprise – it finds that a lot of people are “euthanized” “without” “consent” in the Netherlands.

III.

But we shouldn’t have to take our statistics from second-hand misinterpretations of a twenty-five year old untraceable report anyway. Back in 1990 when the Remmelink Report was written, euthanasia was in fact still illegal in the Netherlands (though decriminalized). Since then, it has been fully legalized, resulting in the ability to collect much better data. Let’s see what we can find.

One very simple statistic that doesn’t even require any digging is the life expectancy in the Netherlands. At 80.2 years, this is 15th highest in the world. It’s also higher than all the Netherlands’ neighbors, including Germany (79.8), Belgium (79.8), and its cross-Channel neighbor the United Kingdom (79.5) – not to mention of course higher than the United States (78.0). Claims from some quarters (*cough* Rick Santorum *cough*) that the Dutch elderly are living in constant fear of being murdered are hard to square with the fact that the Dutch elderly live a good deal longer than our own.

But we can do much better than this. Four years ago, the original authors of the Remmelink Report as well as some new researchers teamed up to produce Two Decades of Research On Euthanasia From The Netherlands: What Have We Learnt And What Questions Remain?

This is important because euthanasia was in fact only legalized in the Netherlands in 2002, so if we want to see what a country with legal euthanasia looks like, we need something more recent than those same few statistics the anti- crowd keeps bringing up. Further, by comparing current (legalized euthanasia) statistics to past (decriminalized euthanasia) statistics, we have a good basis for time comparison and looking at trends that might plausibly be attributed to greater legalization.

The review begins by stating the requirements for euthanasia under the Netherlands’ law:

1. The patient’s request is voluntary and well-considered;
2. The patient’s suffering is unbearable and hopeless;
3. The patient is informed about his situation and prospects;
4. There are no reasonable alternatives. Further,
5. Another independent physician should be consulted; and
6. The termination of life should be performed with due medical care and attention.

It then evaluated the 5,600 cases of legal euthanasia in the Netherlands to determine how many of them had failed one or more of those criteria. They found that “In the period 2003–2005, the review committees gave the verdict of non-compliance in 15 cases on a total of 5,600 reported cases. The main reason for the verdict of non-compliance was a failure to fulfill the requirement of consultation of a second independent physician.”

99.8% of Dutch euthanizations were performed in cases where

> *suffering was described as "unbearable". This is sufficient to disprove the contention that Dutch euthanasia is "grizzly".*

So remember how the article said that 72% of Dutch euthanasia was nonconsensual? The actual number seems more like 0.2% not following the letter of the law, and 0% involving actual consent violations.

> In 1990, 1.7% of all deaths were preceded by euthanasia, as compared with 2.4% in 1995 and 2.6% in 2001 (see Table 1). This trend reversed in 2005, when 1.7% of all deaths were the result of euthanasia (approximately 2,300 cases). The number of euthanasia requests also decreased in this period: from 9700 in 2001 to 8400 in 2005.

So, we go from claims of "slippery slopes" and "as soon as it's legal there will be pressure to apply it to more and more different groups and incidents" to "actually, incidence has been going down for the past few years."

> Table 1 further shows that the frequency of ending of life without an explicit patient request decreased from 0.8% of all deaths in 1990 to 0.4% in 2005 (approximately 550 cases annually). Further analyses of the cases of ending of life without an explicit request show that these concern nearly always patients who are very close to death, are incompetent but with whom the hastening of death has been discussed earlier in the disease trajectory and/or with their relatives, and for whom opioids were used to end life

The number of patients killed without consent was cut in half after euthanasia was legalized, and these deaths continue to be overwhelmingly done with reasonable justification.

> Forgoing of potentially life prolonging treatments (either withholding or withdrawing) and intensified alleviation of symptoms occurred much more often than active ending of life. In all of the studied years, the frequency of forgoing of life-prolonging treatments was rather stable, between 16–20% of all deaths. The percentage of the use of intensified alleviation of symptoms increased from 19% of all deaths in 1990 up to 25% in 2005.

My prediction that euthanasia funges against use of potentially deadly symptom-relieving drugs seems to be disconfirmed, as that has increased over the period euthanasia has been more legal.

The prediction that euthanasia would be used as a way to get rid of the elderly was also disconfirmed. The practice was much less common among the elderly than among younger people; in 2005, 3.5 percent of deaths below age 65 were euthanasia, as opposed to 0.8% of deaths above age 80. The report attributes the discrepancy to euthanasia being a more common practice in cancer patients (who tend to be younger than most terminally ill people), plus of course the overwhelmingly higher number of other causes of death in the elderly.

Okay. Here’s the part I want to stress, because it is a study with a very good methodology on exactly the sort of question we’re worried about. Anne McTavish and the Patients’ Rights Council are concerned that there were 1040 incidents in the Netherlands of doctors ending patient lives without explicit patient request. I noted that this was not euthanasia as commonly defined and should not be expected to differ between countries with and without legalized euthanasia. So let’s see what actually happened. A study compared Netherlands (which has legalized euthanasia) with Switzerland (which has a much more limited version), and with Sweden, Denmark, Belgium and Italy (in which euthanasia was illegal at the time of this study, although Belgium has since legalized it).

They found that although the Netherlands obviously had much more euthanasia and assisted suicide than any of these other countries, it was pretty unremarkably in the middle with regard to “ending of life without explicit patient request” – Belgium was higher, Denmark was tied, and Switzerland, Sweden, and Italy were lower. However, even this doesn’t tell the whole story, since we know the Netherlands’ rate of this practice has cut in half when euthanasia was legalized after these numbers were collected in 2001. Assuming the other countries’ rates stayed stable, it should now have the third lowest rate of this practice among the six countries.

Furthermore, the Netherlands was near the bottom in “intensified alleviation of symptoms”, which I think is this study’s version of the “giving symptom-relieving drugs in quantities that might cause death”. The Netherlands had less “intensified alleviation” than Denmark, Belgium, Sweden, and Switzerland, and was beaten only by Italy. Notice that a *huge* amount of anti-euthanasia advocates’ claims about the evils of euthanasia in the Netherlands come from deaths from intensified alleviation, yet the Netherlands is actually one of the most conservative countries in Europe in this regard!

The report ends by concluding that:

> A last important lesson that can be learnt is that the legalization of euthanasia in the Netherlands did not result in a slippery slope for medical end-of-life practices. Besides religious or principal-based

> arguments, the slippery slope argument is the mainstay of opponents of the legalization of euthanasia. Briefly, the argument states that: if we allow A (the use euthanasia at the request of terminally ill patients), B (abuse of euthanasia, that is, ending the life of vulnerable patient groups without their consent) will necessarily or very likely follow. B is morally not acceptable; therefore, we must not allow A (Griffiths et al. 1998; van der Burg 1992). Our studies show no evidence of a slippery slope. The frequency of ending of life without explicit patient request did not increase over the studied years. Also, there is no evidence for a higher frequency of euthanasia among the elderly, people with low educational status, the poor, the physically disabled or chronically ill, minors, people with psychiatric illnesses including depression, or racial or ethnic minorities, compared with background populations (Battin et al. 2007).

Here is my executive summary of these findings:

Anti-euthanasia advocates point out that in 1990 the Netherlands killed 1040 patients without explicit consent. However, closer examination of these cases shows somewhere between most and all to have taken place in circumstances most reasonable people would consider justified. Further, incidence of this practice declined after euthanasia was legalized. Most importantly, the Netherlands' incidence of this practice is actually relatively low compared to other European countries, and there is no evidence that its attitudes on euthanasia are causing this to happen more often.

The numbers of “people killed without consent” are further inflated by adding people who have been given drugs like morphine for symptomatic relief. In the overwhelming majority of cases, this drugs are not intended to kill the patient and in fact do not. The Netherlands' incidence of this practice is, once again, significantly less than the average incidence for comparable European countries and there is no evidence that its attitudes on euthanasia affect this.

Actual euthanasia never affects more than about 2.5% of deaths in the Netherlands, affects the elderly disproportionately *little*, and does not disproportionately affect minorities, the disabled, or any of the other groups anti-euthanasia advocates suggest it might disproportionately affect. It is practically always done in accordance with strict official guidelines, after repeated and persistent explicit patient request, with the approval of family members, and with the certification of several physicians agreeing that the patient's disease is terminal and likely to cause intense suffering.

The statistics used by opponents of euthanasia are almost universally built on the same misinterpretation of a 25 year old report – a misinterpretation that takes its numbers out of context, which willfully misclassifies non-euthanasia causes of death as euthanasia, which ignores better and more recent data, and which has been soundly condemned as a perversion of their data by the authors of the report itself.

All responsible research on euthanasia in the Netherlands has found it to be applied responsibly, consensually, and in accordance with the legal guidelines, with no evidence of misapplication or of a slippery slope. These findings have been replicated in other countries that have more recently legalized euthanasia, such as Belgium.

This is why I felt comfortable saying in my last post that "there's loads of good data…we can use to estimate how nonconsensual and slippery-slopey real world euthanasia is, and the answer is statistically indistinguishable from zero.”

Military Strikes Are An Extremely Cheap Way To Help Foreigners

…at least potentially.

Lately a bunch of my Facebook friends have been sharing the Slate article Military Strikes Are An Extremely Expensive Way To Help Foreigners, which is too bad because it's *super super wrong*. It is wrong both in design and in execution, and it consigns itself to the special hell for people who make a numerical argument without checking any actual numbers.

Here is its argument: the United States is currently considering whether or not to intervene in Syria. Some people say there are humanitarian grounds for this – after all, Syrians are currently under the heel of a repressive regime and trapped in a deadly war. *However*, if we were true humanitarians, we could just take the cost of intervening in Syria – maybe a few billion dollars – and donate that money to GiveWell, which would save even more lives.

What I mean by "wrong in design" is that saying "this is less effective than GiveWell's number one top rated efficient charity" is damning by *really really faint praise*. Tautologically, if you believe GiveWell then *everything* is less effective than GiveWell's number one top rated efficient charity. This includes policies that you – and Matt Yglesias – support. For example, he says nice things about publicly subsidized preschools, but spending money on these preschools is less effective than GiveWell's top-rated charity. The same is true of other nice things you presumably like such as public libraries, food stamps, PBS, et cetera.

So imagine that in the debate over whether to de-fund PBS a few years ago, someone had brought up that PBS is less cost-effective than GiveWell's top charities, probably even less effective than its top US charities. And then we all patted ourselves on the back for noticing and decided not to renew funding for PBS. What do you think are the chances that all that funding would have ended up in the hands of poor African villagers?

If someone wants to propose a bill that our government can only spend money on the most effective things, starting with the most effective and only moving on to the second most effective once number one has been saturated, I'd be totally in support of that (and I've even tried imagining what such a government could look like). In the absence of such a policy, "let's stop funding PBS since it's less effective than GiveWell's top-rated charity" is a terrible idea, and so is abandoning Syria to its fate for the same reason.

A better policy might be "defund things that produce less utility than tax cuts or the typical government policy", since we can expect the money saved to either be refunded in taxes or to be distributed among other government policies. And this is where it becomes important that the Slate article was *also* wrong in its execution.

The article focuses on Libya as an example of an intervention which, although "successful", was apparently "too expensive". Let's calculate how expensive it really was.

The US spent about $1 billion on military intervention in Libya (our allies contributed some more, but we can assume that US participation "bought" allied participation).

There's no good way of calculating how many lives US intervention saved, but the war up to that point had caused

25,000 casualties, and everyone expected the rebels' final defeat to be something of a bloodbath. Let's say intervention prevented another 25,000 casualties.

Suppose the average Libyan currently alive can be expected to live 25 more years. We multiply by 0.7 (for reasons to be explained later) and so those 25,000 lives saved = a gain of 437,500 QALYs.

However, intervention did more than save potential casualties. It also freed everyone from the government of Moammar Gaddafi, a ridiculously evil guy who squandered the country's wealth and raped his populace both figuratively and literally.

How much should this count for? When I tried to elicit conditional utility weights from people, I didn't have anything that exactly corresponded to Libya, but it seems reasonable to say it was better than North Korea but worse than China, so maybe around 0.6? And that though post-Gaddafi Libya is still poor and conflict-ridden, it's just a little bit better, so perhaps 0.7?

So if you improve the lives of 6 million people by 0.1 QALYs/year x 25 years, that's another 15 million QALYs gained, for a total of about 16 million.

If you purchase 15.5 million QALYs for $1 billion, you find you spent about $65 per QALY.

Compare this to the Against Malaria Foundation. The Slate article boasts that "according to The Life You Can Save, handing out these bed nets saves about one life for every $1,865 spent."

So here $1,865 buys one life = 50 QALYs (we said 25 before, but let's be generous since most deaths from malaria are children and so will have more lifespan left to save), but in this case we multiply by 0.5 (the weight the survey gave to life in Ethiopia, which is a proxy for the sorts of countries where AMF likely operates) and get about $75.

So if you accept the main premise of the above – that life is ~10% better in Libya after Gaddafi was overthrown – military intervention in Libya was a bit more effective towards humanitarian goals than donations to AMF, buying QALYs for $65 versus $75.

Let's put these numbers in a broader context than "one is bigger than the other". In health care, anything that costs less than about *$25,000* per QALY is generally considered excellent value. People wax poetic about the amazing cost-effectiveness of preventative health care because it usually comes out to "only" $1000 – $5000 per QALY. These numbers are *both* mind-bogglingly low.

And in the case of Libya, this may an underestimate, since it doesn't take into account shortening the war, or spurring foreign investment, or the fact that Gaddafi probably would have become more repressive after the rebellion, or less tangible effects like deterrence of future dictators.

Didn't I just say that donating to GiveWell's top charity must tautologically be the best use of money if you trust GiveWell? Yes, but that's only true for the average person with a marginal dollar to spare. If you happen to be a superpower with an air force of 5,500 planes and several billions of dollars to burn, you have options that GiveWell probably will not have evaluated.

My guess is a perfectly utilitarian US Air Force (now *there's* a plot hook!) would probably require its own GiveWell, at least for a while, rather than selling its planes for scrap and donating to the civilian version.

Am I saying that military intervention is always more cost-effective than civilian charity? Certainly not. Military interventions sometimes cause large numbers of deaths, don’t always keep the costs down as effectively as the Libyan operation did, don’t always succeed, and when they do succeed they risk unintended consequences (like replacing a brutal-but-effective dictator with an anarchic power vacuum).

But if we still had InTrade, we could come up with a probability that each of those things would happen, factor them into our calculations, and who knows, it might still be better than giving to Against Malaria Foundation. Or at least better than giving money to Egypt to buy weapons to point at Israel to counteract the money we gave to Israel to buy weapons to point at Egypt – which, let’s face it, is a much more realistic description of where the typical foreign aid dollar goes than Against Malaria Foundation is.

(note that I do not intend this article to assert that intervening in Syria at present is a good idea. I am currently neutral-ish on this point, since the possibility of making the situation much better seems counterbalanced by a good chance of making it much worse. But this should not be taken as a critique of interventions more broadly, let alone a Fully General one.)

Yglesias’ article ends by saying:

> Now, before the kill-and-maim-for-the-sake-of-humanity crowd shoots a Tomahawk missile at me, it’s worth conceding up front that none of this amounts to a logically airtight case against blowing up some Syrian

> infrastructure and killing various Syrian bad guys. It is very possible for a given undertaking to be worth doing without being the optimal policy. But I do think it’s worth interrogating the larger political and ideological construct that says that spending a few billions dollars to help foreigners is a thinkable undertaking if and only if the means of providing assistance is to kill some people and blow some stuff up. The explosives-heavy approach to humanitarianism has a lot of unpredictable side effects, sometimes backfires massively, and offers an extremely poor value proposition. So whatever you think about killing some Syrians this summer, please consider throwing a few dollars in the direction of a cost-effective charity of some kind.

A big “amen!” to all of that except the “extremely poor value proposition” part.

The Life Cycle of Medical Ideas

I.

About five years ago, an Italian surgeon with the unlikely name of Dr. Zamboni posited the theory that multiple sclerosis was caused by blockages in venous return from the brain causing various complicated downstream effects which eventually led to the immune system attacking myelinated cells. The guy was a good surgeon, nothing about the theory contradicted basic laws of biology, and no one else had any better ideas, so lots of people got excited.

As far as I can tell, the medical community responded exactly one hundred percent correctly. They preached caution, urging multiple sclerosis patients not to develop false hope. But at the same time, they quickly launched studies investigating Zamboni's experiments and used newly gathered data to test the theory. All the results that came back made the idea look less and less likely, so that to my understanding by now it is pretty much discredited. Having successfully spent hundreds of thousands of dollars to empirically disconfirm Zamboni's hypothesis, we can now reflect at leisure on the reasons it was kind of dumb and we should have realized it all along.

II.

About five years ago, two Israel doctors named Gat and Goren posit the theory that benign prostatic hyperplasia, a prostate disease that affects millions of older men, is caused by incompetence of the spermatic veins. They claim they can treat it surgically, and show off rows of smiling patients with glowing testimonials. Once again, the guys are good doctors, nothing about their theory contradicts basic laws of biology, and no one else has any better ideas.

I shamefacedly admit I *want* this one to be true. There's so much "well, everything is a complicated combination of genes, biomolecules, biopsychosocial stressors and immune modulators that we may never really understand" going on in medicine today that it would be super gratifying if this one mysterious disease turned out to just be plumbing going in the wrong direction. And although the prostate is about as far from my area of expertise as it is possible to be, I have to say that from a physiological standpoint their theory seems to have that rare and much-sought scientific elegance, where everything comes together in a pretty package.

On the other hand, it sounds a whole lot like the Zamboni debacle transposed to a different organ, and Gat and Goren don't have much evidence other than a pretty theory and their own anecdotal success.

As far as I can tell, the medical community has totally ignored this one. Gat and Goren have published their hypothesis and their apparent excellent results in peer-reviewed medical journals. It has garnered praise from prestigious figures in the field (bonus points for calling it "seminal", especially if the pun was intentional). As far as I know, no one has attacked it or even formally expressed doubt. Yet as far as I know, it has gone nowhere.

Does everyone mutually assume that if something this revolutionary were true, someone would have noticed beyond a single article in a urology journal? Do they just decide it needs further research, and hope that this research will be conducted by someone else? Or do they think that it would end up like Zamboni's MS cure, with hundreds of thousands of dollars wasted, dozens of unnecessary surgeries performed, and nothing to show but yet another fringe medical idea that sounded good at the time?

III.

Minocycline is a relatively boring umpteenth-line antibiotic sometimes used to treat acne. About five years ago, some Japanese doctors noticed that their schizophrenic patients with acne seemed to be *getting better*. This was especially bizarre because some of these patients had "negative symptoms", a set of schizophrenia symptoms considered totally untreatable and which the super-advanced next-generation antipsychotics being pumped out by drug companies can't even touch. They started wondering – can minocycline, an uninspiring antibiotic from the early 1970s, do what all of these psychiatric medications can't?

Once again the medical community responded correctly. They launched a couple of double-blind placebo-controlled studies of minocycline, and sure enough, the stuff was shown to work again and again.

And yet the psychiatrists I know have never heard of it, and I am not aware of any psychiatric hospital in the world where minocycline is routinely given to schizophrenic patients with negative symptoms outside of a clinical trial.

It's not like this is some kind of experimental drug that might kill the patient and isn't even legal yet and we have to wait for further research. If the schizophrenic patient happens to get *acne*, the psychiatrist will be perfectly happy to send them to the nearest CVS Pharmacy to pick up a bottle of minocycline, which they will no doubt have in droves. It's just the schizophrenia connection that isn't there.

I’m totally in favor of waiting for all the research to come in and not jumping to conclusions. The problem is that I don’t understand exactly what the process is. If the rule was “We must wait for NIMH to fund a study with greater than 2,000 subjects, and after that everyone will prescribe it, and NIMH is currently working on crunching the data, so just hold your horses,” this would sound totally reasonable to me. The problem is that I don’t know what we’re waiting for and I’m not sure there actually *is* a thing we’re waiting for except a spontaneous change in the zeitgeist, which could take forever.

IV.

When people blame drug companies for suppress any promising medications they can’t make a profit off of, those people are missing the point.

The drug companies don’t suppress promising medications. Promising medications start off pre-suppressed. In some cases they are suppressed by regulation that says a drug has to go through crazy expensive trials before it can be approved. In other cases, they are suppressed simply by the burden of proof: even without the government, doctors aren’t going to prescribe something they don’t know is safe and effective, and they’re not going to know it’s safe and effective without studies, which as I may have mentioned are crazy expensive. In still other cases, the medications are suppressed by medical conservativism: most doctors very reasonably don’t want to use a drug unless they know other doctors they respect are using the drug, so unless the drug impresses itself onto the consciousness of the entire medical community at once it will fizzle out.

What drug companies do, as best I understand it, is put billions of dollars and millions of man-hours of effort into un-suppressing those particular drugs it is in their financial interest to un-suppress. They are doing a great service. It’s just a very selective one.

I’m not sure how it works in surgery. As far as I know, there aren’t companies that patent surgical procedures and then popularize them. If there were, maybe one of them would pick up Gat-Goren and give it a fair try to stand or fall on its own merits. As it is, it looks like it will have to wait for some university or charitable group to pick it up – and let’s face it, “my eighty year old grandpa gets up to piss half a dozen times a night” isn’t quite as sexy as multiple sclerosis.

In medicine, drugs are usually approved for specific indications. Doctors are allowed to prescribe them for other indications, but there are trivial inconveniences and minor legal hurdles and in practice most of them rarely do. Some pharmaceutical company was nice enough to get minocycline approved for acne back in the ’70s, but since then it’s gone off patent and no one owns it enough to say “Hey, start the process to approve this drug of mine for schizophrenia!” The medical community is pretty smart and I bet there’s a process by which this will eventually happen, but I also bet it will take a long time and be overly complicated and a whole lot of schizophrenics will have to suffer from negative symptoms long after the vanguard of the medical community has satisfied itself that these are treatable.

(I can imagine the look on my attending’s face if I suggested we treat one of our schizophrenic patients with minocycline. I expect I would get a lecture on how We Have To Be Responsible And Ethical, and then we would give them one of the same three drugs we give all schizophrenics. I might have more luck painting little fake acne pustules on my schizophrenic patients’ faces, but most of the ones I have now are Paranoid-Type and I *really* can't imagine them going along with that. I'll just have to wait until I get someone catatonic.)

Which is why it sucks that the *other* really interesting drug that might revolutionize the treatment of schizophrenia is an antihypertensive from the 1950s.

V.

I find the life cycle of medical ideas really interesting.

I was always taught that there were two kinds of medicine. Real medicine, which has been proven to work by studies. And alternative medicine, which has been proven not to work by studies but people still use it anyway because they are stupid.

This dichotomy leaves out the huge grey area of "things that seem like they will probably work, and a few smaller studies have shown very promising results, but no one has bothered to do larger studies, or if they have they have never really been incorporated into medical practice for reasons I can't put my finger on."

Some of this grey medicine, like Zamboni's MS treatment, are doomed to eventually fall back into the abyss of alternative medicine. Others, like the Gat-Goren procedure, teeter in the middle, threatening to go either way. Still others, like minocycline, have already been sanctified and dressed in robes of white, and the only thing preventing them from entering Evidence Based Heaven is some sort of weird bureaucratic snafu at the Pearly Gates.

I am encouraged that all three of the examples of grey medicine cited in this article are about five years old. It suggests that there's a certain window of time during which grey medicine is well-known but hasn’t yet been well-studied. Maybe most of the newer stuff I don’t know about, and most of the older stuff has been successfully proven or disproven. It seems possible to me that the current system does have the optimal combination of safety and innovation, or at least the best we can do without a Science Czar. As I immerse myself in Medical Culture, I look forward to finding out whether there are some hidden processes for dealing with this, or whether the situation is really as dire as it looks.

But I am also hopeful that some new organizations like Microryza and MetaMed might, totally independent of justified or unjustified medical conservativism, be able to speed the process along.

Giving and Accepting Apologies

“Saying ‘I’m sorry’ is the same as saying ‘I apologize.’ Except at a funeral.”
– **Demetri Martin**

Today was the Day of Atonement. As an atheist, I find it hard to appreciate atonement in its cosmic, maybe-not-even-possible sense. But I can certainly appreciate the power of a good apology.

A few days ago, there was a…

Actually, I just realized that since this involves a real case going on at a real hospital with possible legal ramifications, it would be a *terrible* idea for me to describe it, so I’m going to change it around so far that it bears no resemblance to the original except a certain ambiguity around the idea of apologies.

So, a few days ago, there was a mishap at my hospital. A patient was having some pain, but was already heavily sedated already and further painkillers might be dangerous. I am still a lowly intern and wasn’t sure what to do, so I called up my attending, who told me to go ahead and administer the drugs. This was late at night, so immediately after doing that I signed the patient over to night shift and went home.

The next morning I came in and learned the patient had gotten loopy from being over-sedated and fallen down trying to walk to the bathroom. This can start a minor panic in a hospital, because we tend to have elderly people who don’t tolerate falls very well, and a bad one can mean anything from broken bones to a big lawsuit.

But in fact, the patient hadn't gotten any broken bones or head injuries or anything like that. He had scraped his elbow. My attending came in a little after me, heard there was a fall, and freaked out that was some critical injury. I said "Oh, don't worry, he just scraped his elbow." The attending was very relieved, and we went into an office and discussed how to change our painkiller policies and prevent problems like this in the future.

A nurse heard the comment and when the patient's wife came to ask about what had happened, she told the wife that the doctor had said the injuries weren't a big deal. The wife then very reasonably filed a complaint against me, because all she knew about me was that I had administered the drug that caused the problem and that I had dismissed the resulting fall as "not a big deal".

The complaint made its way up to a Head Honcho, who called me in for a meeting. I explained that I had consulted my attending about using the drug, and that what I had said had been said in private as an attempt to communicate that the patient was okay. The documentation supported me on both counts and I was Cleared Of Wrongdoing (thank God).

I then said I would call up the patient's wife to apologize. The Head Honcho told me that I was quite right to want to call the patient's wife to *explain what had happened*, but that I hadn't done anything wrong and I shouldn't sell myself short by saying "I'm sorry" or "I apologize" or any permutation thereof. It wasn't about legal issues – the wife has already said she's not suing us. He was just a nice guy and and thought it would be unfair to make me "confess" to something not my fault.

This incident has gotten me thinking about what exactly "I'm sorry" is supposed to mean.

First, it can mean “This problem was entirely the fault of my own moral failings. I admit to these moral failings, will try to give you whatever recompense I can, and will do everything possible to make this not happen again.”

For example, I cheat on my wife. Then I realize that was wrong and I apologize to her. It’s unambiguously my fault I cheated, and I accept that fault.

Second, it can mean “You are sad. I am sad that you are sad. I wish that this had not happened.”

For example, as per the quote at the top, I go to a funeral and tell the family I’m sorry about their loss.

Third, it can mean “You are sad. The reason seems to be very tangentially related to me or my actions, but no reasonable person would call it my fault. I am sad that you are hurt and if I could have prevented that I would have.”

For example, a doctor tries her best to save a patient, but the patient still dies. She tells the family “I’m sorry I couldn’t do more.” Or my friend in China is upset and I say “I’m sorry I can’t be there to help you get through this.” Or the hero tells the villain “I’m sorry it had to end this way” before shooting him.

Crucially but annoyingly, the first and third meaning are almost opposite each other. The first one is a way of saying “This is my fault”. The third one is a way of saying “This isn’t my fault.” They are both polite, deferential, and respectful of the pain that the victim is suffering. But only one of them admits wrongdoing.

Unfortunately, the English language really doesn’t have a great way of politely and deferentially being respectful of someone else’s pain that you are tangentially related to other than saying “I’m sorry”.

And also unfortunately, the English language really doesn’t have a great way of saying “This was not my fault” without sounding impolite, disrespectful, and like you’re trivializing the other person’s pain.

This is why myself and the Head Honcho are having so much trouble getting together a good statement to give to the patient’s wife. All of his proposed statements sound something like

> “I heard you were concerned about my conduct in the case of Mr. X. I appreciate your distress, but I was not the one who chose to administer that drug, and when I described your husband’s injury I only meant that it wasn’t medically critical. The implication that it wasn’t a big deal was a misinterpretation by the nurse who gave you that statement.”

This is completely honest, yet it oozes “I am a callous businessperson who is concerned entirely with covering his own ass and doesn’t care about your pain.”

But all of *my* proposed statements sound something like:

> I’m very sorry that your husband fell. I handed off the decision to give painkillers to a more senior doctor, but I recognize that this was partially my fault as well. Likewise, although I only meant to say that your husband’s injuries were not medically critical, I realize I could have chosen my words better. Please accept my apologies for this distressing incident.”

This makes me sound like a potentially decent person and a responsible doctor. But as the Head Honcho points out, there's a sort of dishonesty in admitting more blame on paper than I'm willing to admit in my heart. Like, it's true that I could have chosen my words more carefully, but if someone else were to *accuse me* of this, I would instantly object that since I had no idea someone would overhear them, distort them, and report the distorted version to someone else, there's no reason I should have wanted to. The level of misinterpretation that happened to me could have happened to practically anyone at any time if they had the same bad luck – that utterance was not unusually careless by normal human standards.

The miniature angel on my right shoulder tells me I should do everything I can to let this woman know that I care about and regret her suffering, even at the cost of my own pride. It tells me I should accept "heroic responsibility" – that anything that happens is my fault at least far enough that I can learn from it and use it to be a better person next time. But the miniature Robin Hanson on my left shoulder tells me that I am just signaling my own virtue and responsibility while secretly I don't believe any of what I'm saying.

Right now I'm leaning towards the angel (not that it matters; I can recommend, but the Head Honcho will make the final decision). I think my outlook is that people are reasonable: the patient's wife, upon hearing what really happened, will read between the lines and realize it wasn't quite my fault, but will appreciate my phrasing it in a way that emphasizes the importance I place on her concerns. I have talked to the patient's wife before, I respect her, and I trust her to be a good apology-accepter.

The Head Honcho, who doesn't know the patient's wife, might worry otherwise. Perhaps she will be the sort of person who, when I give my apology, won't buy it. She'll shout "You could have chosen your words better?! No sh*t, Sherlock!! You chose those words in a very hurtful way, you're irresponsible, and I'm going to ask the hospital to fire you for not caring about offending people!!"

Not only have I just made her angrier, but I've also abandoned my one advantage in this situation – the fact that I didn't really do anything wrong. By admitting to some wrongdoing, I can no longer make the defense I made above – the one where I couldn't reasonably expected my words to be twisted in that particular way – without being accused of backtracking and flipflopping and outright lying.

II.

I feel privileged to be hanging around basically decent people like my patient's wife, and so to have a lot of options in terms of how to apologize. You know who has it really bad? Politicians. I *cringe* every time I see a story in the media about politicians apologizing or not apologizing.

Politicians can either give the first sort of statement I mentioned – a very dry "It's not my fault" sort of statement – and look like shifty sleazy people even if, in fact, it was not their fault.

Or they can try to give an apology of the same type I was considering above, and then have all their words twisted and used against them forever until they eventually have to backtrack on it and get torn to pieces for doing so.

Suppose that an attractive female Senator is bringing a motion to the floor, and a male Senator mutters to himself under his breath "I'd like to motion *her* to the floor." Unbeknownst to him, his Official Senate Lapel Microphone is on and broadcasts his statement to the entire C-SPAN watching public (so, like, five people).

Of course, the entire nation instantly suspends talking about boring things like the war in Syria or anthropogenic global warming to focus on the new scandal. “Sexism In The Senate?” the cover of TIME says. “Are Female Senators Just Objects For Men’s Amusement?” asks the cover of US News. And eventually the sound of a million screeching voices and the occasional death threat convinces Male Senator With Microphone that he should probably apologize or something.

Male Senator With Microphone probably feels terrible. He embarassed himself in front of his colleagues, he offended the female Senator he was talking about, disrupted the proceedings of the Senate, annoyed millions of average Americans, and maybe in some vague way contributed to sexism. He certainly has every reason and every right to feel bad.

On the other hand, noticing – just to one’s self! – that other people are attractive seems like a harmless and rather universal thing to do, and is hardly anything to feel bad about.

The most honest apology he could give would probably be “I’m sorry I said that aloud instead of just thinking it. I didn’t realize my microphone was on at the time. I will be more careful in noticing when my microphone is and isn’t on in the future.”

But of course that would mean the end of his political career. Everyone would say he was so tone-deaf and out-of-touch that he couldn’t realize the problem was that he was sexist, and not that he got *caught* being sexist. And the cries for his head would grow three times louder.

He could go a little further than that and remain honest. He could say something like “I am sorry that lots of people were offended by my statement. I did not mean to hurt anyone's feelings and I will try to do better in the future."

And then the newspapers, in front page headlines, would gleefully point out that he had only apologized "that people were offended" and not actually admitted wrongdoing. People would shake their heads and talk about how politicians always gave stuttering non-apologies, and eight different activist groups would hold large public demonstrations demanding he show *real* contrition instead of trying to "deflect the issue".

Or he could go all out. He could say "I'm incredibly sorry that I wronged Senator Y in that way. Clearly I have some sexism issues I need to work on. I promise I will never do it again and I hope the American people can forgive me for being a bad Senator and a bad person."

And then he would be *torn to pieces*.

First, everyone will demand he resign, throwing his own words about being sexist and a bad person back at him. Then, if he doesn't resign, they will demand he "atone" for his misdeeds by oh, let's say doing everything that feminist lobby groups want all the time, with the threat of getting constant "I GUESS SENATOR X WASN'T REALLY SINCERE ABOUT WORKING ON HIS SEXISM" any time he can't comply quickly enough.

And the worst part will be that he will have no defense, because deep in his heart he knows he has given up his pride, his integrity, and his right to stand up for himself in order to very temporarily quiet the hordes.

This Hobson's Choice between three terrible options is all the average politician has when accused of anything – whether or not the original accusation is even fair (and I have a gut understanding of this, having previously been a politician myself).

So is it any wonder that the average politician usually responds by denying all wrongdoing and accusing someone else?

III.

I am okay at apologizing for intellectual mistakes. There I can get new data and realize I was wrong. This is one reason I so want to apologize for prescribing the wrong painkiller. I thought listening to my attending's advice and giving that painkiller would be okay. Events proved me wrong. I can apologize for my misjudgment without a hint of dishonesty.

The same is true of honest mistakes, like coming to work late because I accidentally set my alarm wrong. If anything, I over-apologize for these until everyone is sick of hearing me and tries to reassure me with "Really, it's just an alarm. Not the end of the world."

But I'm not very good at apologizing for moral mistakes. When I think courses of action are clearly wrong, I don't do them. When I think courses of action are clearly right, I do them, but then I tend to continue thinking they are right and don't see anything to apologize for.

Actually, my Yom Kippur was kind of a bust. I tried really hard to repent and I couldn't really think of anything especially bad I did. It's not that I'm a bad person, just that my life is kind of boring. I got exasperated with a couple of people more quickly than they deserved, but I feel bad repenting for it when I know that the next time someone is equally annoying I will probably get exasperated equally quickly. I mostly just have nice conversations with nice people, browse the Internet, and do my job moderately well.

I am tempted to ask other people if I have wronged them. It might provide useful data. The problem is, I'm afraid they're going to say yes, and they will explain the way I wronged them, and then I will say "No, you're wrong, I acted entirely correctly in that situation", which would sort of break the whole point of repentance.

But although I am bad at apologizing, I think I am pretty good at accepting apologies. Accepting apologies requires understanding the Principle of Charity, not to mention the Fundamental Attribution Error. It requires going into a person's shoes, imagining whether you could have done the same thing in the same circumstances, and gradually paring away at your anger until you can picture the absolute minimum amount of wrongdoing and malice they could have had consistent with them doing what they did, which is usually the amount they actually had.

It requires thinking "Yeah, I sometimes notice other people are hot, and so me and that senator are morally equivalent, and since I don't blame *myself* for sometimes admiring people's appearances, well, there but for the grace of not having a microphone in my lapel go I."

Or "Well, I didn't know everything about my job immediately after I started either, so if I had called my boss and asked for help and she had been wrong, I wouldn't want anyone to take it out on me."

Since English has no good way to express regret and compassion without also expressing fault, I think it is appropriate to apologize for things you don't think were completely your fault – and I also think it is appropriate to accept those apologies in the spirit in which they were offered.

Inefficient Hot Dogs

I know the “why do hot dogs come in packs of ten and buns in packs of eight” question is beloved of boring observational comedians everywhere, and has become a cliched example of people who don’t have anything more interesting to talk about.

But from an economic perspective, why *do* hot dogs come in packs of ten and buns in packs of eight? Last weekend I spent about ten minutes laboriously looking through all the hot dog buns trying to find one in a pack of ten so it would match my hot dogs. I finally gave up and bought one of the twenty or so different packs-of-eight. If there had been a ten-pack, I would have bought it instantly.

This sort of equilibrium shouldn’t be able to last ten seconds in a free market. It should *literally* have taken ten seconds from the moment the hot dog bun company executive noticed all the other companies sold their buns in an inconvenient number to the time she decided to make her company’s product stand out by offering what consumers are looking for.

I know it seems trivial. But then we posit theories like “Companies will react to consumers’ desire for safe goods by monitoring their own safety standards” or “Businesses can’t remain sexist over the long term, because refusing to hire competent women who are otherwise neglected would be like leaving a hundred dollar bill on the ground”.

And both of those arguments sound like they should work, *unless* of course for some mysterious reason entire industries can just completely fail to respond to incentives for indefinite periods of time based on some stupid reason like tradition. Right now the hot dog situation seems like an especially good example of evidence in favor of this possibility.

Scientific Freud

In this month’s *American Journal of Psychiatry*: The Efficacy of Cognitive-Behavioral Therapy and Psychodynamic Therapy in the Outpatient Treatment of Major Depression: A Randomized Clinical Trial. It’s got more than just a catchy title. It also demonstrates that…

Wait. Before we go further, a moment of preaching.

Skepticism and metaskepticism seem to be two largely separate skills.

That is, the ability to debunk the claim “X is true” does not generalize to the ability to debunk the claim “X has been debunked”.

I have this problem myself.

I was taught the following foundation myth of my field: in the beginning, psychiatry was a confused amalgam of Freud and Jung and Adler and anyone else who could afford an armchair to speculate in. People would say things like that neurosis was caused by wanting to have sex with your mother, or by secretly wanting a penis, or goodness only knows what else. Then someone had the bright idea that beliefs ought to be based on evidence! Study after study proved the psychoanalysts’ bizarre castles were built on air, and the Freudians were banished to the outer darkness. Their niche was filled by newer scientific psychotherapies with a robust evidence base, such as cognitive behavioral therapy and [mumble]. And thus was the empire forged.

Now normally when I hear something this convenient, I might be tempted to make sure that there were *actual* studies this was based on. In this case, I dropped the ball. The Heroic Foundation Myth isn't a *claim*, I must have told myself. It's a *debunking*. To be skeptical of the work of fellow debunkers would be a violation of professional courtesy!

The AJP article above is interesting because as far as I know it's the largest study ever to compare Freudian and cognitive-behavioral therapies. It examined both psychodynamic therapy (a streamlined, shorter-term version of Freudian psychoanalysis) and cognitive behavioral therapy on 341 depressed patients. It found – using a statistic called noninferiority which I don't entirely understand – that CBT was no better than psychoanalysis. In fact, although the study wasn't designed to demonstrate this, just by eyeballing it looks like psychoanalysis did nonsignificantly better. The journal's editorial does a good job putting the result in context.

This follows on the heels of several other studies and meta-analyses finding no significant difference between the two therapies, including, another in depression, yet another in depression, still another in depression, one in generalized anxiety disorder and one in general. This study by meta-analysis celebrity John Ioannidis also seems to incidentally find no difference between psychodynamics and CBT, although that wasn't quite what it was intended to study and it's probably underpowered to detect a difference.

(other analyses do show a difference, for example Tolin et al, but the studies they draw from tend to be much smaller than this latest and in any case are starting to look increasingly lonely.)

Suppose we accept the conclusion in this and many other articles that psychodynamic therapy is equivalent to cognitive-behavioral therapy. Do we have to accept that Freud was right after all?

Well, one man’s modus ponens is another man’s modus tollens. The other possible conclusion is that cognitive-behavioral therapy doesn’t really work either.

If parapsychology is the control group for science, Freudian psychodynamics really ought to be the control group for psychotherapy. Although I know some really intelligent people who take it seriously, to me it seems so outlandish, such a shot-in-the-dark in a low-base-rate-of-success environment, that we can dismiss it out of hand and take any methodology that approves of it to be more to the shame of the methodology than to the credit of the therapy.

But what about the evidence base for cognitive behavioral therapy over placebo? Or, for that matter, the evidence base for psychoanalysis over placebo?

Part of the problem may be what exactly is used as placebo psychotherapy. In many studies, it’s just getting random people to talk to patients. This makes intuitive sense as a placebo therapy, but it seems vulnerable to unblinding – people usually have some expectation of what psychotherapy is like, and undirected conversation about problems might not match it. Or if the placebo therapists are not professionals, they may be less confident in talking to people about their mental health problems, more awkward, less charismatic, or otherwise not the sort of people who would make it in the therapy profession. So now a lot of people are coalescing around the idea that all therapy studies done against these kinds of placebo therapy are fundamentally flawed.

Studies that compare what are called "bona fide psychotherapies" – two therapies both done by real therapists with real training – tend to have a lot more trouble finding differences. This has led to what is called the Dodo Bird Verdict, after an obscure Alice in Wonderland reference I feel vaguely bad for not getting: that psychotherapies work by having a charismatic, caring person listen to your problems and then do ritualistic psychotherapy-sounding things to you, but not by any of the exercises or theories of the specific therapy itself.

Then the question becomes: if the Dodo Bird Verdict and the active placebo problem and so on are equally true of all psychotherapies and all psychotherapy studies, how come everyone become convinced that cognitive behavioral therapy passed the evidence test and psychoanalysis failed it?

And the answer is *the CBT people did studies and the psychoanalysts didn't*.

That's it. It may be, it probably is, that any study would have come back positive. But only the cognitive behavioral people bothered to perform any. And by the time the situation was rectified and the psychoanalysts had (positive) studies of their own to hold up, "everyone knew" that CBT was evidence-based and psychoanalysis wasn't.

This seems like another case of doctors not understanding that there are two different types of "no evidence".

I should qualify this sweeping condemnation. I believe a few very basic therapies that address specific symptoms in very simple ways will work. For example, exposure therapy – where you treat someone's fear of snakes by throwing snakes at them until they realize it's harmless – is extremely and undeniably effective. Some versions of CBT for anxiety and DBT for borderline also seem to just be basic coping skills about getting some distance from your emotions. I think it’s likely that these have some small effects (I know a study above found no effect for CBT on anxiety, but it was by a notorious partisan of psychoanalysis and I will temporarily defy the data).

But anything more complicated than that, anything based on an overarching theory of How The Mind Works, and I intuitively side with the Dodo Bird Verdict. And I think the evidence backs me up.

EDIT: Do *not* stop going to psychotherapy after reading this post! All psychotherapies, including placebo psychotherapies, are much better than nothing at all (kinda like how all psychiatric medications, including placebo medications, are much better than nothing at all).

The Thin Blue Line That Stays Strangely Horizontal

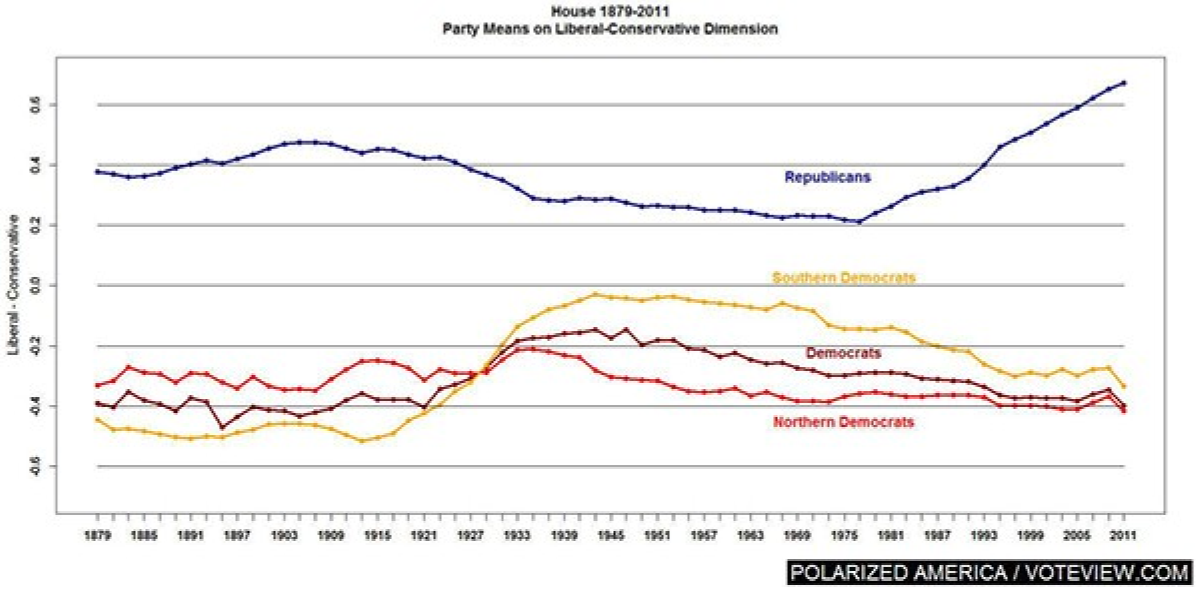
Last night I was pondering how to test the Reactionary claim that society was drifting inexorably to the Left, and I wondered whether maybe someone had created a mathematically rigorous model of people's positions along the left-right spectrum and then applied it to every legislator in every session of Congress from the founding of the United States to the present day.

This being the Information Age, of course the answer is yes. NOMINATE is a statistical system that takes every legislator's vote on every issue that comes before Congress, then number-crunches it to see which legislators are closest or furthest away from one another. The classic left-right dimension naturally falls out of such a model, allowing a quasi-objective way to determine which legislators are more liberal or conservative than others.

It would appear that this could only compare two legislators serving at the same time and voting on the same issues. But an enhanced version of the system, DW-NOMINATE, gets around this limitation with a clever trick. Political scientists know that legislators' views generally change little over time. Therefore, long-term legislators can be used to "bridge" different sections of Congress. If Senator Bob is the most liberal Senator in 1950, but by 1970 he's somewhere in the middle, then we know the Senate became more liberal from 1950 to 1970. If Senator Mary is the most liberal Senator in 1970, but by 1990 she's somewhere in the middle, then we know the Senate also became more liberal from 1970 to 1990. Therefore, we can conclude that the

Senate in 1990 is more liberal than in 1950, even if no single Senator served in both bodies.

DW-NOMINATE has been called “one of the great achievements of modern political science”, has won a bunch of awards, and most tellingly is routinely used by Nate Silver. It is recognized as broadly accurate by people from both sides of the political spectrum. So let’s see what it finds:



A few things here stand out as correct. Congress is getting more polarized recently – true. The Republicans took a big turn to the right starting in 1980 – true.

But from this it looks like there has been *zero* liberal drift of either party over the past 120 years. And don’t get me wrong, if it’s true I will relish rubbing it in the Reactionaries’ faces – but it seems kind of implausible. That whole “expansion of the welfare state” was a thing that happened, right? And Martin Luther King Jr speeches aside, I’m pretty sure the civil rights movement wasn’t *literally* a dream. Gay marriage? Also, either we got an income tax early in the

20th century, or else some guy with the initials I.R.S. has been pulling a *very* successful con on me.

Normally I'd dismiss this. I'd say it probably wasn't kosher to use DW-NOMINATE data to make time comparisons. But a political analyst tried making that argument last year and was shot down by the developers of the DW-NOMINATE system. And Nate Silver seems pretty happy to draw DW-NOMINATE comparisons across presidents. And from what I understand of the long but gratifyingly non-mathematical guide to using DW-NOMINATE, intertemporal comparisons are entirely acceptable as long as they don't go back further than 1879. And the system's ability to pick up other secular trends, even smallish ones, seems quite good.

I've got to admit it. I'm stumped on this one.

Edit: There's a similar measure called Martin-Quinn scores for the Supreme Court that finds a similar result

Prediction is very difficult, especially of the past

There's been some recent discussion here on political change, so I thought I'd give people an opportunity to put their money where their mouths are.

I gathered data from Pew Research, who ask a representative sample of Americans whether they agree or disagree with statements like "I am very patriotic". They've been doing this every year from 1987 to 2009, so they have 22 years – about a generation – worth of data on value changes.

In a second, I'm going to link you to a test which will ask you three things. First, the percent of Americans whom you think agree with each statement today (for values of "today" equal to 2009). Second, the change you believe occurred since 1987. Third, your own political position.

(I'm also asking people to *voluntarily* add their names at the end, so that winners can be properly honored and people have an incentive to get it right)

My hypothesis – and I know it's bad form to state a hypothesis to research subjects before an experiment – is that most people will underestimate the strength of their own political position due to a version of the underdog effect. I am especially interested in seeing Neoreactionaries take this test due to their belief in the power of the Cathedral.

Test is closed, answers/data to be posted soon

PS: I promise I didn’t rig this test. I used the first values survey I could find and included all the questions except the boring ones or the repeats. I’d like to say I didn’t even peek at the answers, except that of course I did. Nevertheless, these are not trick questions.

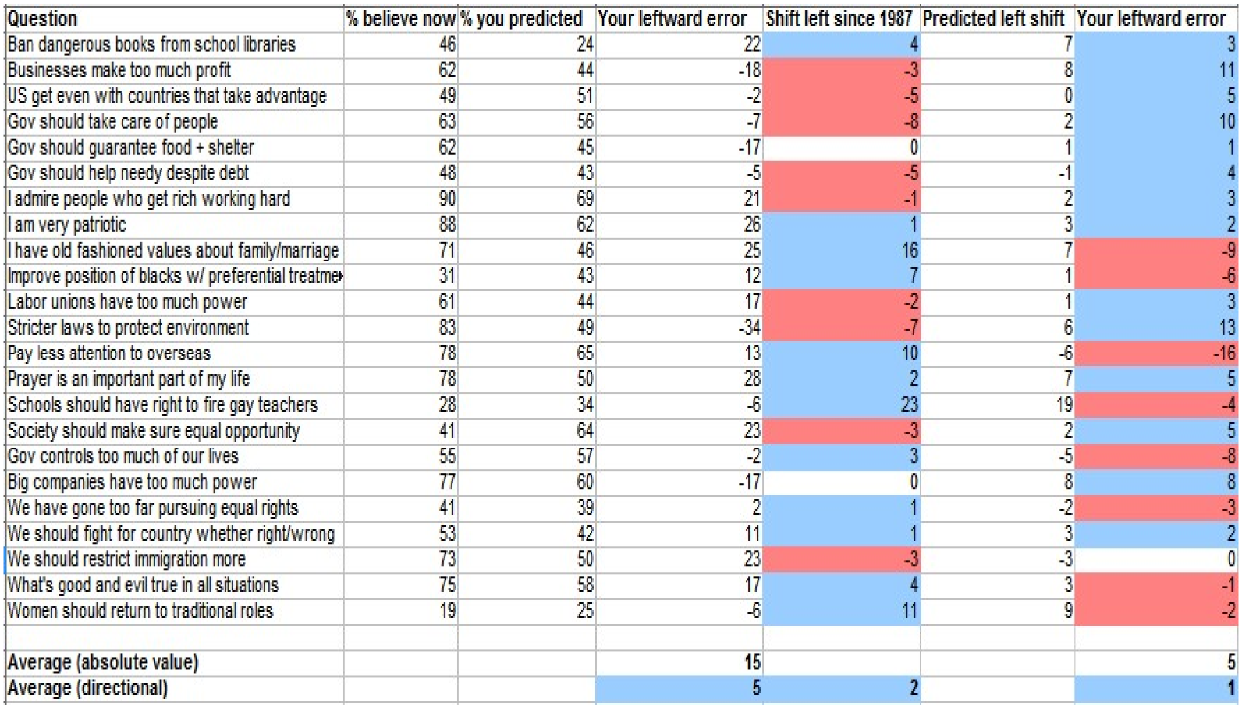
PPS: I don’t have access to the original questions verbatim, so assume they might have used slightly different wording.

PPPS: Yes, you could easily Google the right answers. I’m trusting you not to do that.

The Trend Evaluation Of All Values

In case you forgot: yesterday, 170 people sent in their predictions on what percent of Americans agreed with vaguely political statements like “I am very patriotic”. Then they predicted how answers to these questions have *changed* in the past 22 years.

Without further ado, here are the answers (click to expand).



Question	% believe now	% you predicted	Your leftward error	Shift left since 1987	Predicted left shift	Your leftward error
Ban dangerous books from school libraries	46	24	22	4	7	3
Businesses make too much profit	62	44	-18	-3	8	11
US get even with countries that take advantage	49	51	-2	-5	0	5
Gov should take care of people	63	56	-7	-8	2	10
Gov should guarantee food + shelter	62	45	-17	0	1	1
Gov should help needy despite debt	48	43	-5	-5	-1	4
I admire people who get rich working hard	90	69	21	-1	2	3
I am very patriotic	88	62	26	1	3	2
I have old fashioned values about family/marriage	71	46	25	16	7	-9
Improve position of blacks w/ preferential treatme	31	43	12	7	1	-6
Labor unions have too much power	61	44	17	-2	1	3
Stricter laws to protect environment	83	49	-34	-7	6	13
Pay less attention to overseas	78	65	13	10	-6	-16
Prayer is an important part of my life	78	50	28	2	7	5
Schools should have right to fire gay teachers	28	34	-6	23	19	-4
Society should make sure equal opportunity	41	64	23	-3	2	5
Gov controls too much of our lives	55	57	-2	3	-5	-8
Big companies have too much power	77	60	-17	0	8	8
We have gone too far pursuing equal rights	41	39	2	1	-2	-3
We should fight for country whether right/wrong	53	42	11	1	3	2
We should restrict immigration more	73	50	23	-3	-3	0
What's good and evil true in all situations	75	58	17	4	3	-1
Women should return to traditional roles	19	25	-6	11	9	-2
Average (absolute value)			**15**			**5**
Average (directional)			**5**	**2**		**1**

People were terrible at predicting what percent of Americans agreed with something, with an average error on each issue of 15%. Especially egregious errors include “We need stricter laws to protect the environment” – you expected 49% of Americans to agree, but in fact 83% did. You expected 50% of Americans to agree that “Prayer is an important part of my daily life”, but in fact 78% did.

In general, you expected modern Americans to be a bit more leftist than they actually are – five percentage points more leftist, to be exact.

You did a little better predicting how opinions had shifted over the past generation, with an average error of 5% per issue. By far your worst question was “Pay less attention to problems overseas and more at home”. You thought 6% more people would agree; in fact, 10% fewer do. Your other big mistake was expecting 8% more people to think businesses were too profitable; actually, 3% fewer people believe this.

Overall, you overestimated the degree society has shifted to the left, but only by a tiny tiny bit – one percentage point per issue. Depending on whether I classify “We should pay less attention to overseas” as a leftist or rightist position, I can either make that overestimation vanish, or double it to a still-measly two percentage points.

(if I wanted to be sensationalist, I could say you overestimated society’s leftward shift by 50% – which is true, but only because the real shift was two percentage points)

I ran these numbers again counting only the Reactionary answers, and then again counting only the Progressive answers. Contrary to my hypothesis, this made practically no difference; I think Reactionaries were one percent worse than Progressives on one of the two measures and one percent better on the other. It was all well below any conceivable margin of error.

Alejandro was most accurate in his estimates of what positions Americans actually hold. Athrelon was most accurate in his estimates of how positions have changed over time. He very slightly edged out second-place winner JRM, even though the latter had already read the answers. On IRC Athrelon attributed his success to being a medical student and therefore getting to see a broad cross-section of society whom he might otherwise avoid. As a doctor myself, I can endorse that sentiment. In any case, congratulations to both winners.

I didn't have time to do much with people's ages or anything like that, but you're welcome to crunch the numbers yourself. You can download the raw data (minus names) from **this link**. Someone double check to see if I totally messed up my numbers.

Polling data for this exercise was taken from Static America: Myths About Political Change In The US

Sleep – Now By Prescription

Ramelteon isn't a bad drug. It's just that its very existence stands as a condemnation of the entire medical system.

All sleep medications have to straddle a very fine line between "idiotically dangerous" and "laughably ineffective", and Ramelteon manages better than most. It outperforms placebo, it's not addictive, it won't sap your ability to sleep without it, and it doesn't screw up your brain so badly that its unofficial mascot is a hallucinatory walrus.

How does it do it? Ramelteon is the first melatonergic drug, selectively binding to MT-1 and MT-2 melatonin receptors. Binding to melatonin receptors presumably mimics the effect of the natural hormone melatonin which is believed to serve a sleep-promoting role.

Now, you might ask yourself – the natural hormone melatonin is available as an over-the-counter supplement costing a couple cents per pill in every drug store, and provably quite safe and effective. Why would anyone go through the trouble of creating a drug that mimics its action? Especially if a month's supply of the drug costs around $100 – which it does.

The answer is: *I have no idea and I'm pretty sure no one else does either*.

Wikipedia says of Ramelteon that:

> In a double-blind multicenter trial, Ramelteon did reduce the time to fall asleep by approximately 15–20 minutes, at 8 mg and 16 mg doses after four weeks compared to placebo (approx. 29-32 versus 48 minutes) Total sleep

> time improved about 40 minutes, however, this was identical to improvement with placebo at the end of trial

A meta-analysis of melatonin says:

> Our meta-analysis demonstrated melatonin had a significant benefit in reducing sleep latency. Subjects randomly assigned to melatonin fell asleep 7 minutes earlier on average than subjects receiving placebo…in the random effects model, sleep latency was reduced by over 10 minutes

Sleep latency is a tough statistic to work with, because it depends a lot on how quickly the people in your trial got to sleep in the first place. If the study population is chronic insomniacs who take an hour to fall asleep each night, a good drug might be able to reduce that by 30 minutes. If the study population is normal youth who fall asleep within ten minutes, needless to say your drug isn’t going to be able to do 30 minutes better.

So, for example, it’s easy to find a melatonin trial that finds a very impressive sleep latency decrease of 34 minutes, or a ramelteon trial that finds a rather anaemic 9 minutes. The only fair way to compare ramelteon and melatonin is to run a head-to-head trial.

The only such trial that has ever been performed was performed on monkeys, and its results were contradicted by other monkey experiments. Also, it was run by the company that sells Ramelteon.

I think we may have enough evidence to conclude that Ramelteon is at least as effective as melatonin. There may even be some very tenuous evidence to suggest it is slightly more effective. But let me tell you a story.

One of my patients ran into the Ambien Walrus the other day and so, make a long story short, she needed a new sleeping pill. She was on a lot of drugs at the time and not all that healthy, and every drug I could think of, the pharmacist had some good reason why that would be a terrible idea in her case. Finally in desperation I remembered Ramelteon, which is safe as houses. Unfortunately Ramelteon is kind of new, and the pharmacy didn’t have it.

“Okay,” I said. “Why don’t we just give her some melatonin? Some studies in monkeys suggest it *might* be slightly inferior to Ramelteon, but it’s sure better than nothing.”

Let’s see if you are cynical enough to predict what happened next.

That’s right. The hospital pharmacy, which carries thousands of drugs including bizarre experimental concoctions and super-expensive recombinant monstrosities, *didn’t have melatonin*.

So do you want to know what the plan was, that the pharmacist and I came up with to treat my patient? I would take my lunch break, drive home, go into the cabinet in my bathroom, take the bottle of melatonin I had there, and bring it to the 500-something bed, multi-billion dollar hospital I work at.

This is why the story of Ramelteon scares me so much – not because it’s a bad drug, because it isn’t. But because one of the most basic and useful human hormones got completely excluded from medicine just because it didn’t have a drug company to push it. And the only way it managed to worm its way back in was to have a pharmaceutial company spend a decade and several hundred million dollars to tweak its chemical structure very slightly, patent it, and market it as a hot new drug at a 2000% markup.

I’m not knocking the pharmaceutical companies – they didn’t do a thing to suppress melatonin. All they did was notice that doctors were too dumb to use melatonin on their own and figure out a way around that problem.

And this is not an isolated incident. For example, on the rare occasions psychiatrists remember that folic acid exists at all they prescribe Deplin ($100/month, prescription only) instead of the *chemically identical* l-methylfolate ($5/month, over the counter).

While we’re on the subject of melatonin, here are some Fun Melatonin Facts you may not have known (courtesy of *Melatonin and Melatonergic Drugs as Therapeutic Agents: Ramelteon and Agomelatine, the Two Most Promising Melatonin Receptor Agonists*):

— Melatonin’s sleep promoting effects might be related to its ability to decrease core body temperature, which seems tantalizingly related to the finding that cooling caps are highly effective against insomnia.

— Smith-Magenis Syndrome is a rare genetic condition among whose effects are disruptions in the melatonin system. People with this syndrome wake at night and sleep during the day, meaning we can add this to porphyria, anemia, and rabies on the List Of Diseases That People With More Desire To Explain Away Ancient Folktales Than Sense Use As A Factual Basis For Vampirism.

— Many people use melatonin at night to try to hack their own circadian rhythms, but this is only mildly effective because they still have their own endogenous melatonin doing their own thing. The nuclear version of this strategy is to use melatonin at night to increase melatonin levels and beta-blockers in the morning to decrease melatonin levels; the combination can give you almost complete control over your own circadian rhythm.

— Melatonin seems to play a role in fat metabolism and has been found to decrease weight gain associated with overfeeding in rats.

— Agomelatine is a melatonergic antidepressant that has been found to be approximately as effective as SSRIs with fewer side effects which is available in Europe. However, attempts to sell it in the USA were cut short when it failed to clearly differentiate from placebo in clinical trials (see: “found to be approximately as effective as SSRIs”)

— Melatonin appears to slow the growth of tumors, and a possible role as an adjuvant to classical chemotherapy drugs in cancer treatment is just one of the exciting areas of melatonin biology doctors are completely failing to explore.

The Anti-Reactionary FAQ

[Edit 3/2014: I no longer endorse all the statements in this document. I think many of the conclusions are still correct, but especially section 1 is weaker than it should be, and many reactionaries complain I am pigeonholing all of them as agreeing with Michael Anissimov, which they do not; this complaint seems reasonable. This document needs extensive revision to stay fair and correct, but such revision is currently lower priority than other major projects. Until then, I apologize for any inaccuracies or misrepresentations.]

0: What is this FAQ?

This is the Anti-Reactionary FAQ. It is meant to rebut some common beliefs held by the political movement called Reaction or Neoreaction.

0.1: What *are* the common beliefs of the political movement called Reaction or Neoreaction?

Neoreaction is a political ideology supporting a return to traditional ideas of government and society, especially traditional monarchy and an ethno-nationalist state. It sees itself opposed to modern ideas like democracy, human rights, multiculturalism, and secularism. I tried to give a more complete summary of its beliefs in Reactionary Philosophy In An Enormous, Planet Sized Nutshell.

0.1.1: Will this FAQ be a rebuttal the arguments in that summary?

Some but not all. I worry I may have done too good a job of steelmanning Reactionary positions in that post, emphasizing what I thought were strong arguments, sometimes even correct arguments, but not really the arguments Reactionaries believed or considered most important.

In this FAQ, I will be attacking not steel men but what as far as I can tell are actual Reactionary positions. Some of them seem really dumb to me and I excluded them from the previous piece, but they make it in here. Other points from the previous post *are* real Reactionary beliefs and make it in here as well.

0.2: Do all Reactionaries believe the same things?

Obviously not. In particular, the movement seems to be divided between those who want a feudal/aristocratic monarchy, those who want an absolute monarchy, and those who want some form of state-as-corporation. Even more confusingly, sometimes the same people seem to switch among the three without giving any indication they are aware that they are doing so. In particular the difference between feudal monarchies and divine-right-of-kings monarchies seems to be sort of lost on many of them.

In general, this FAQ chooses two Reactionary bloggers as its foils – Mencius Moldbug of Unqualified Reservations, and Michael Anissimov of More Right. Mencius is probably the most famous Reactionary, one of the founders of the movement, and an exceptionally far-thinking and knowledgeable writer. Michael is also quite smart, very prolific, and best of all for my purposes unusually willing to state Reactionary theories plainly and explicitly in so many words and detail the evidence that he thinks supports them.

Mencius usually supports a state-as-corporation model and Michael seems to be more to the feudal monarchy side, with both occasionally paying lip service to divine-right-of-kings absolutism as well. Part 2 of this FAQ mostly draws from Michael's feudal perspective and Part 4 is entirely based on Moldbug's corporation-based ideas.

0.3: Are you going to treat Reaction and Progressivism as real things?

Grudgingly, yes.

One of the problems in exercises like this is how much to take political labels seriously. Both "Reaction" and "Progressivism" are vast umbrella concepts on whose definition no one can agree. Both combine many very diverse ideas, and sometimes exactly who falls on what side will be exactly the point at issue.

Part of Part 3 will be an attempt to define Progressivism, but for now I'm going to just sweep all of this under the rug and pretend that "Reactionary" and "Progressive" (or for that matter "leftist" and "rightist") have obvious well-defined meanings that are exactly what you think they are.

The one point where this becomes very important is in the discussion over the word "demotist" in Part 2. Although debating the meaning of category words is almost never productive, I feel like in that case I have *more* than enough excuse.

1: Is everything getting worse?

It is a staple of Reactionary thought that everything is getting gradually worse. As traditional ideas cede to their Progressive replacements, the fabric of society tears apart on measurable ways. Michael Anissimov writes:

> The present system has every incentive to portray itself as superior to all past systems. Reactionaries point out this is not the case, and actually see present society in a state of severe decline, pointing to historically high levels of crime, suicide, government and household debt, increasing time preference, and low levels of civic participation and self-reported happiness as a few examples of a current cultural and historical crisis.

Reactionaries usually avoid getting this specific, and with good reason. Now that Michael has revealed the domains in which he is critiquing modern society, we can start to double-check them to see whether Progressivism has indeed sent everything to Hell in a handbasket.

nice hypothesis you have there
would be a shame
VIA 9GAG.COM
if someone were to
TEST IT

But I must set some strict standards here. To support the Reactionary thesis, I will want to see long-term and unmistakeable negative trends in these indicators. Nearly all Reactionaries agree that the advance of Progressivism has been a long-term affair, going on since the French Revolution if not before. If the Reactionaries can muster some data saying that something has been getting better up until 2005 but declining from 2005 to the present, that doesn’t cut it. If something else was worsening from 1950 to

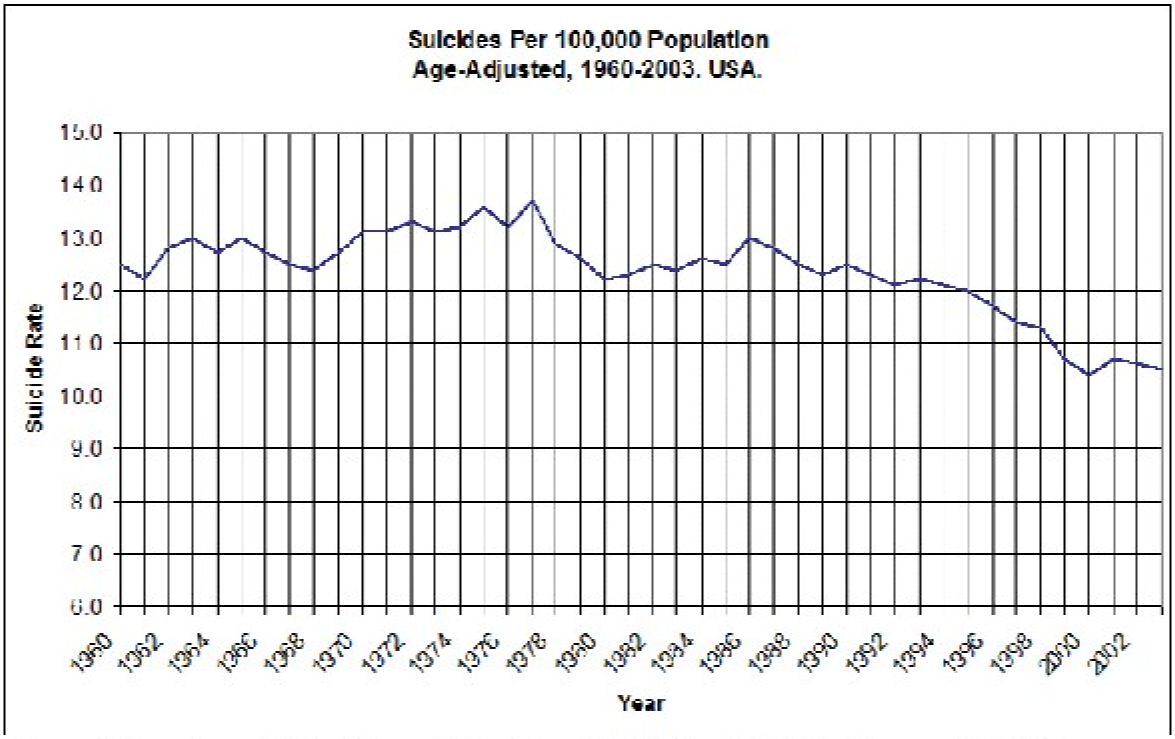
1980 but has been improving since then, that doesn't cut it either. I will not require a completely monotonic downward trend, but neither will I accept a blip of one or two years in a generally positive trend as proving all modern civilization is bankrupt.

Likewise, if something has been getting worse in Britain but not the United States, or vice versa, that will not suffice either. Progressivism is supposed to be a worldwide movement, stronger than the vagaries of local politics. I will not require complete concordance between all Western countries, but if the Anglosphere countries, France, Germany, and Japan seem split about fifty-fifty between growth and decay in a certain indicator, blaming Progressivism isn't going to cut it.

So, without further ado, let's start where Michael starts: with suicide.

1.1: Is suicide becoming more common?

Here's the US suicide rate from 1960 to 2002:



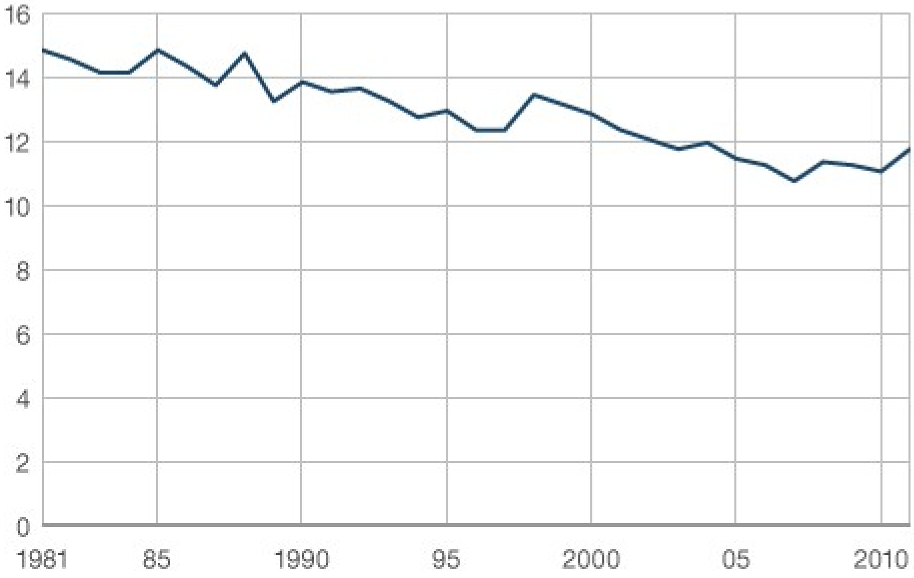
(Source: US Census Bureau. Statistical Abstracts of the United States. 2004-2005. "No. 102: Deaths by Major Causes 1960-2002 (Age Adjusted Death Rates Per 100,000 Population.)" www.census.gov/prod/2004pubs/04statab/vitstat.pdf; Center for Disease Control and Prevention, National Center for Health Statistics. National Vital Statistics Report, vol. 53, no. 15, Feb. 8, 2005, "Deaths: Preliminary Data for 2003" Table 2, "Deaths, Death Rates, and Age Adjusted Death Rates, for 113 Seclected Causes, Injury by Firearms, Drug-Induced Deaths, Alcohol Induced Deaths, and Injury at Work: United States Final 2002 and Preliminary 2003." Online. www.cdc.gov/nchs/data/nvsr/nvsr53/nvsr53_15.pdf)

In those forty years, considered by many the heyday of the leftist movement, forty years encompassing the Great Society, the civil rights movement, the explosion of feminism onto the public consciousness, the decline of the traditional family, etc, etc…suicide rates dropped about 20%.

What evidence have the Reactionaries cite for their side? Michael cites a New York Times article pointing out that suicide rates rose from 1999 to 2010. Apparently my new job is reminding Reactionaries that they cannot blindly trust New York Times articles to give them the whole truth.

Suicide rates did rise from 1999 to 2010. But if we’re going to blame leftism for rising suicide rates it’s kind of weird that it would choose the decade we had a Republican President, House, Senate, and Supreme Court to start increasing. A more likely scenario is that it had something to do with the GIANT NEVER-ENDING RECESSION going on at the time.

As we mentioned above, since Reactionaries believe that Progressivism has been advancing simultaneously in many different countries it is worthwhile to check whether other nations show the same trends as the United States. If every country that was becoming more Progressive showed increased suicide rates, this would be strong evidence that Progressivism were to blame. But if some Progressive countries experienced lower suicide rates, that would suggest country-specific problems.


Suicide rate per 100,000 people

Aged 15 and over

Source: Office for National Statistics

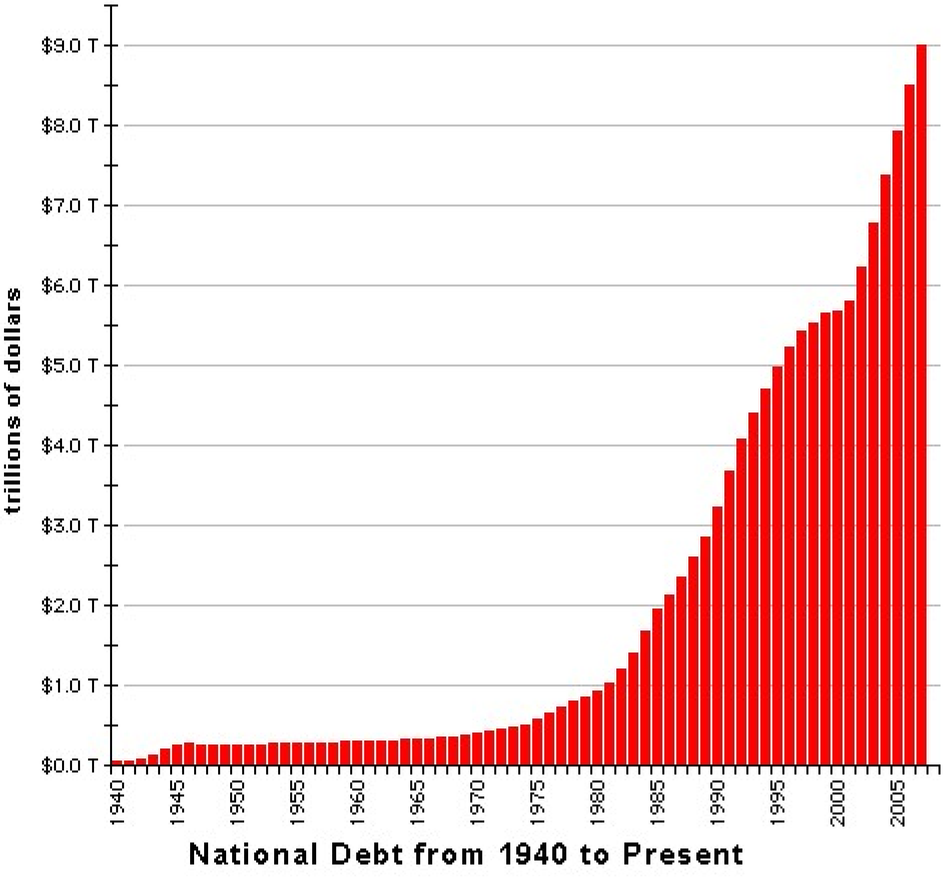
In Britain, we find not only that suicide has generally been going down for the past thirty years, but that – as predicted above – there is a bit of an upward tick corresponding with the Great Recession.

Even better, we find that suicide peaked in Britain in 1905 – just after the Victorian period – and has been declining ever since.

I try to be nice. I really do. But I will say it – the Reactionary argument that suicide has been increasing during modernity from a low during some fantasized Victorian Golden Age is *unacceptably shoddy*.

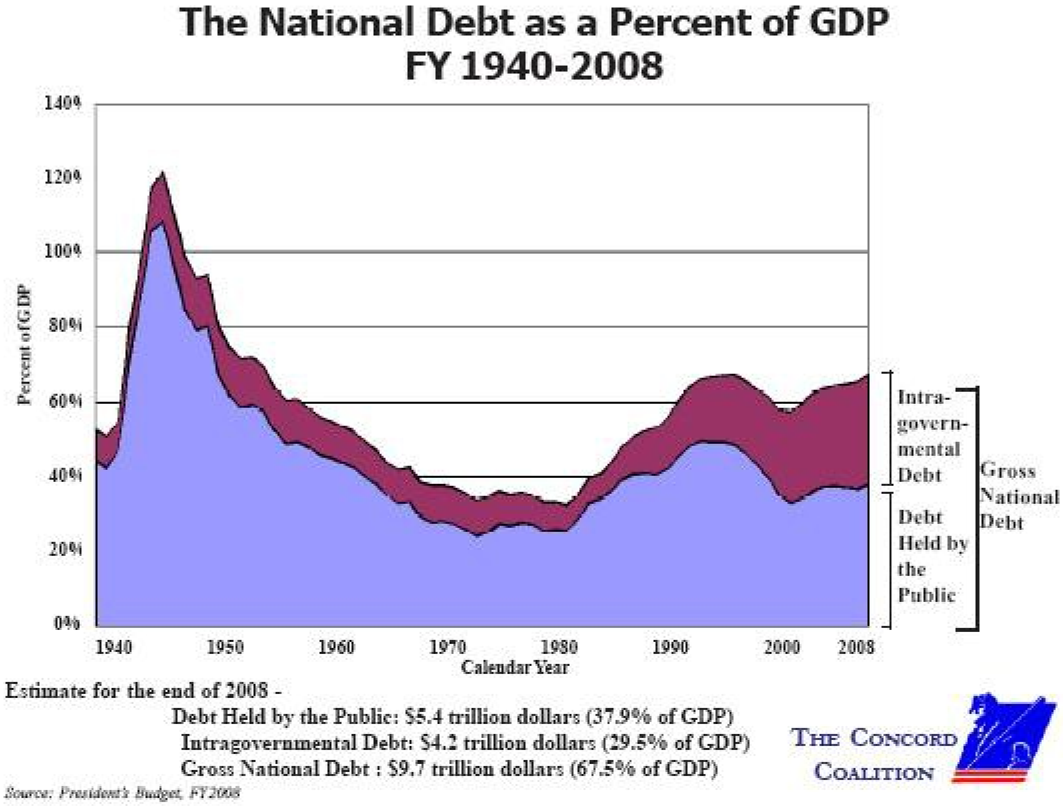
1.2: Is everyone falling further and further into debt?

Here again the Reactionaries overstate their case. Michael tried to support his point with…



Source: U.S. National Debt Clock
http://www.brillig.com/debt_clock/

…which shows government debt rising ceaselessly and alarmingly through the simple tricks of not adjusting for inflation or rising GDP. Keep yourself honest by taking those steps, and the situation looks more like this:



The National Debt as a Percent of GDP
FY 1940-2008

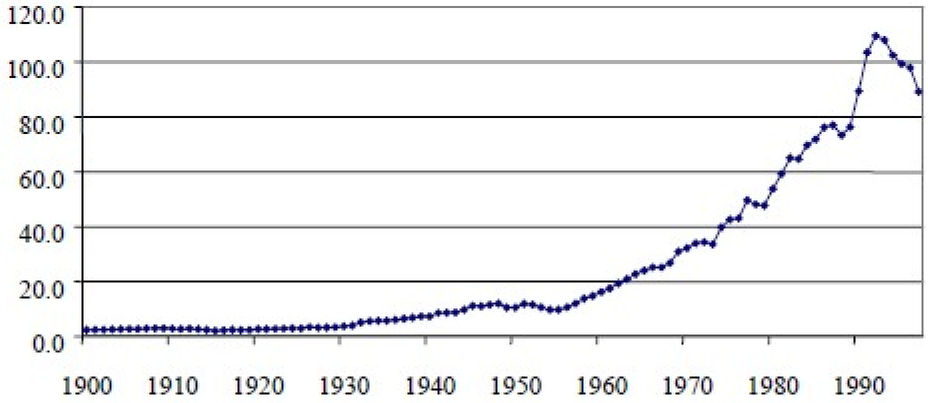
Estimate for the end of 2008 -
Debt Held by the Public: $5.4 trillion dollars (37.9% of GDP)
Intragovernmental Debt: $4.2 trillion dollars (29.5% of GDP)
Gross National Debt : $9.7 trillion dollars (67.5% of GDP)

Source: President's Budget, FY2008

To his credit, Michael fixed this when I pointed it out. But to me, the new graph looks like gradual decrease in debt since World War II up until Reagan’s big military buildup, followed by a gradual retreat from that military buildup. My God, won’t somebody stop Progressivism before it’s too late?!?!

1.3: Is crime becoming worse?

Michael’s statistics for crime deserve more attention:

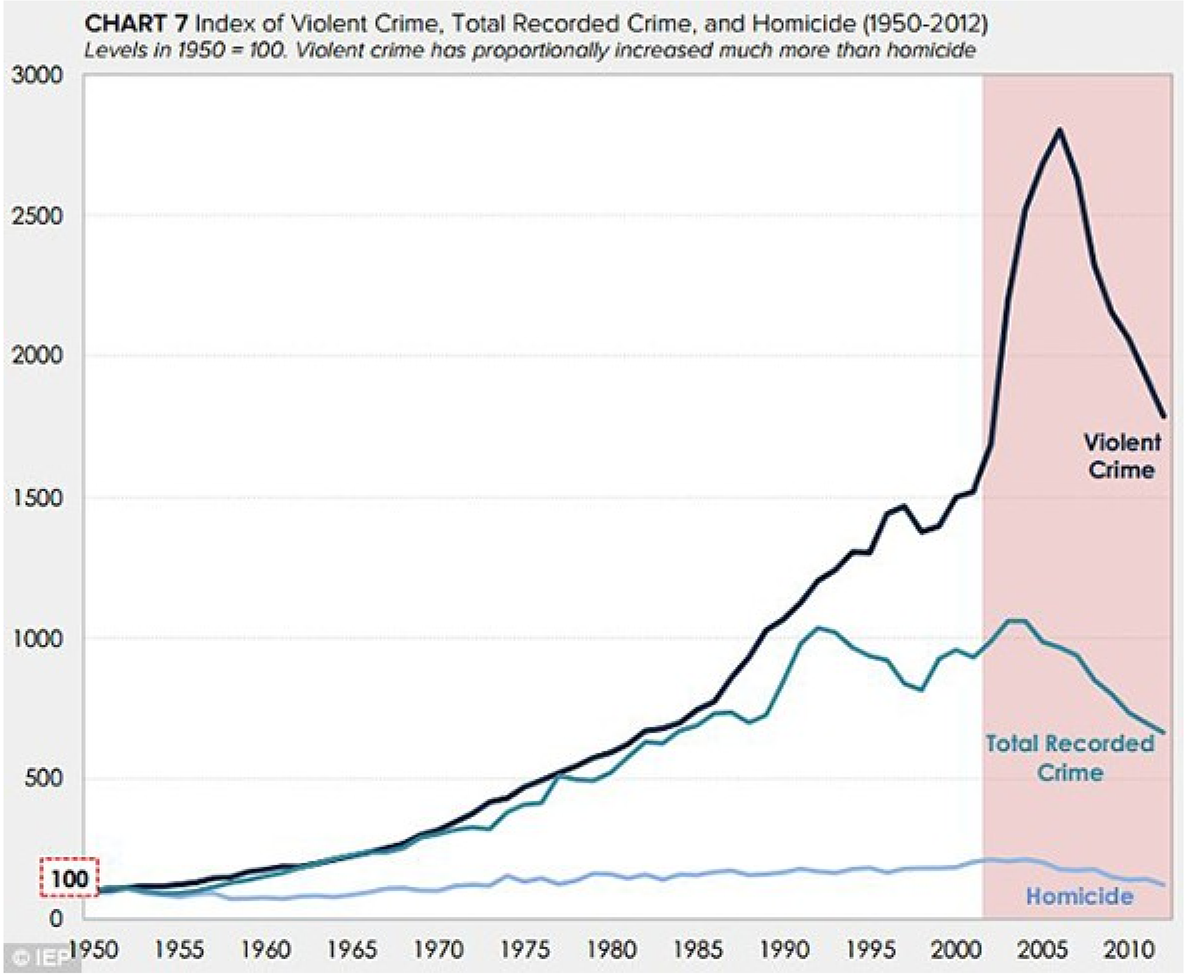


Indictable Offences Known to the Police (per thousand of population) in England & Wales 1900-1997.

Question number one: what does this graph mean by “indictable offenses”? This very broad term introduces no fewer than three dangerous biases. First, we have reporting bias – the more police there are and the more active there are, the more crimes get heard about and reported. Second, we have definition bias within individual crimes – for example, larceny in Britain fell by two thirds in 1855, but this was because Parliament passed a law raising the minimum amount of property that had to be larcened for it to count. Third, we have broader definition bias in what is or isn’t a crime – how much of that rise around 1970 was the “indictable offense” of people smoking marijuana, something that was previously neither illegal nor widely available?

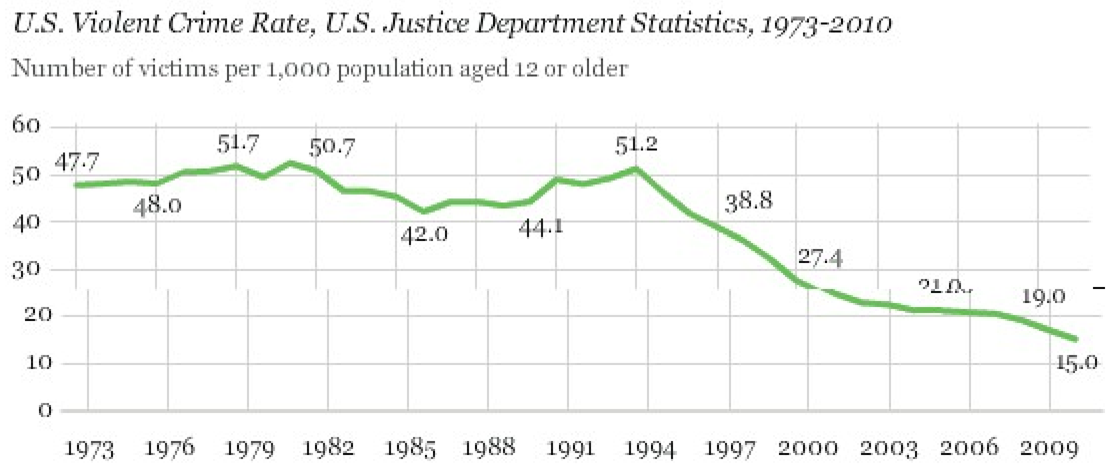
Criminologists’ recommended way around this problem is to look at *murder*. The murder rate tends to track the crime rate in general. Murder isn’t as subject to reporting bias – if someone is killed, the police are going to want to hear about it no matter how understaffed they are. And murder is less subject to changes in definition – dead is dead.

So let’s add the homicide rate to the above chart:

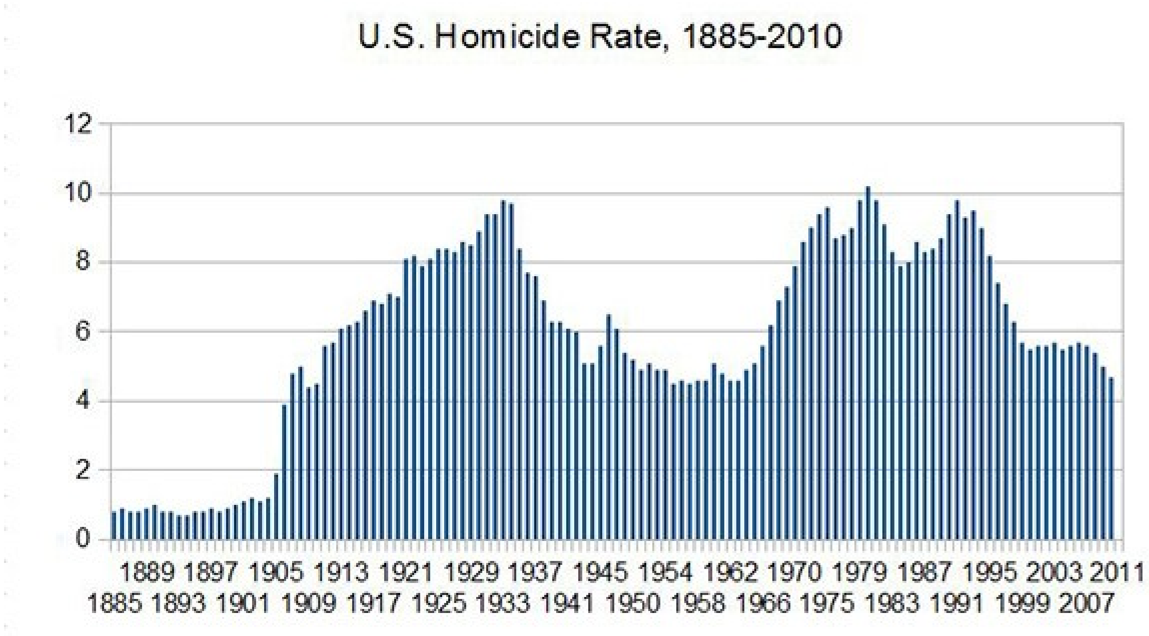


Alas, I can only find the numbers since 1950 rather than 1900. But as we can see, despite the huge rise in “violent crime”, homicide rates stay very steady and perhaps even decline a little over that period.

Question number two: Michael is American. All his other statistics make reference to American numbers. Why does he suddenly switch to Britain when we talk about crime? I won’t impugn his motives – long-term US crime data is really hard to find. But it’s worth pointing out that what there is, is much less sensational:

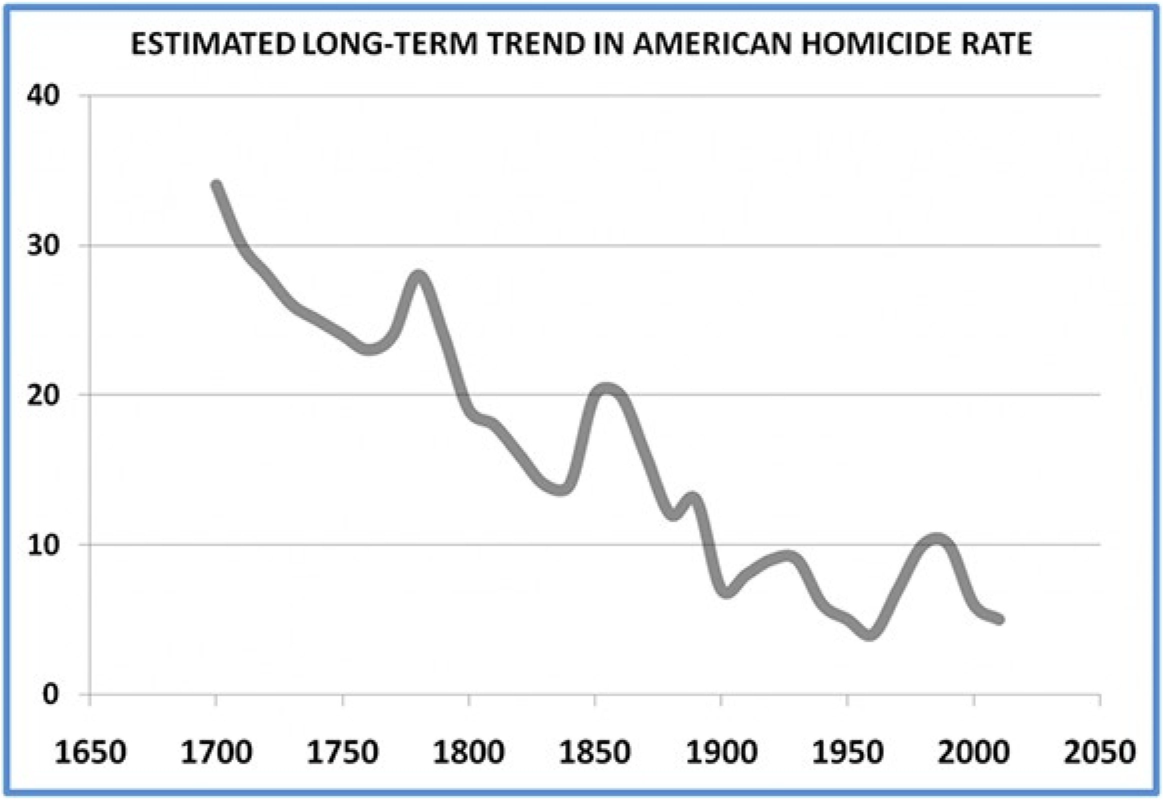


I wish I could find longer-term US crime rate data, but it doesn't seem to be out there. I can, however, find longer-term homicide data:

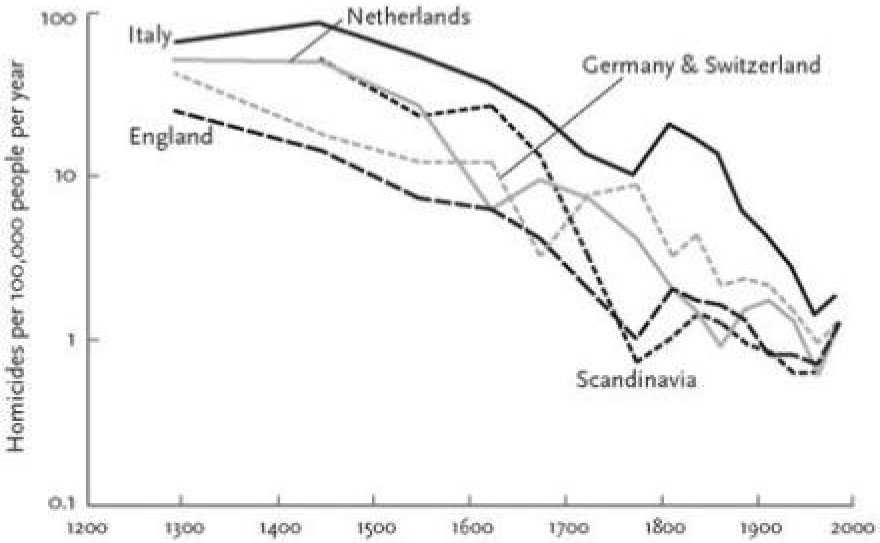


We see ups and downs but no general pattern. A Reactionary might cite the apparently very low level of homicides in 1885, but historians pretty much agree that's a reporting artifact and that the period ending in 1887 had the highest murder rate in American history. In any case, right now we seem to be enjoying a 50 year low. And lest someone bring up that medical technology has advanced enough to turn many would-be murders into attempted murders – which is true – aggravated assaults, the category of crime that would encompass attempted murders, are less than half of what they were twenty years ago. Kind of hard to square with everything getting worse and more violent all the time.

Actually, stopping at 1885 is for losers. Let's go *really* long-term. From Marginal Revolution, themselves drawing from Manuel Eisner's Long-Term Historical Trends in Violent Crime:



We've got to go deeper! From HBD Chick, citing Steven Pinker:



1.3.1: But the Victorian Era had amazingly low crime rates! People could walk out in any corner of the country unmolested! Crime was basically a half-forgotten memory!

This is one of Mencius Moldbug's favorite points. He cites approvingly an 1870s British text which says that

> Meanwhile, it may with little fear of contradiction be asserted that there never was, in any nation of which we have a history, a time in which life and property were so secure as they are at present in England. The sense of security is almost everywhere diffused, in town and country alike, and it is in marked contrast to the sense of insecurity which prevailed even at the beginning of the present century. There are, of course, in most great cities, some quarters of evil repute in which assault and robbery are now and again committed. There is perhaps to be found a lingering and flickering tradition of the old sanctuaries and similar resorts. But any man of average stature and strength may wander about on foot and alone, at any hour of the day or the night, through the greatest of all cities and its

> suburbs, along the high roads, and through unfrequented country lanes, and never have so much as the thought of danger thrust upon him, unless he goes out of his way to court it.

Reactionaries take this idea and run with it – past societies were so well-organized that they had completely eliminated crime, whereas our own democratic government turns a blind eye while thousands of people are beaten and mugged and murdered and…

Again, let's concentrate on "murdered". It's the only crime that gives us a shot at apples-to-apples comparison. So what was the Victorian murder rate?

> Homicide is regarded as a most serious offence and it is probably reported more than other forms of crime. Between 1857 and 1890, there were rarely more than 400 homicides reported to the police each year, and during the 1890s the average was below 350. In Victorian England, the homicide rate reached 2 per 100,000 of the population only once, in 1865. Generally, it was about 1.5 per 100,000 falling to rarely more than 1 per 100,000 at the end of the 1880s and declining even further after 1900. These figures do not take into account the significant number of infanticides that went undetected. The statistics for homicide are therefore probably closer to the real level of the offence.

So, Victorian murder rate of between 1 and 2 per 100,000 people. And the current British murder rate? According to the United Nations Office on Drugs and Crime, it stands at 1.2 per 100,000 people, rather lower than the Victorian average.

1.3.1.1: But if the Victorian crime rate was as high or higher than it is today, how come Victorians felt completely safe and thought that crime had been eradicated?

Normally this is where I'd start talking about how we moderns are constantly exposed to so many outrageous and terrifying stories in the media that we don't realize how good we have it. But in this case that turns out to be explaining away a nonproblem. The Victorians were absolutely terrified of crime and thought they were in the middle of a gigantic crime wave. Here's Understanding The Victorians on the "garroting panic":

> Violent attacks by strangers were seen as grave cause for concern. There was a disproportionate amount of attention paid to violent nighttime assaults by strangers in urban areas, called "garroting" and similar to what we might call "mugging". There were garroting panics in 1856 and 1862, in part because of extensive press coverage. In the highest profile case, MP Hugh Pilkington as attacked and robbed in London at one o'clock in the morning on July 17, 1862, after leaving a late session in the House of Commons. Press reports of garroting increased dramatically, and the public quickly became convinced there was a serious problem. Garroting panic was so rampant that it became a topic of satire: *Punch* published several cartoons of men running from their own shadows or from trees that they were convinced were garrotters.

And A History of Criminal Justice In England and Wales on the same topic:

> Crimes of violence were perceived to be on the increase in the 1850s and panic set in when an outbreak of

> garrotting occurred in various parts of the country in the period from 1856 to 1862. Garrotting involved choking, suffocating, or strangling a victim. During these years, Punch magazine carried a whole series of cartoons and lengthy jokes about the crime, including many eccentric means of defense. One advertisement appeared offering the public an "anti-garrot collar". This was a steel collar to be hand-fitted round the neck with a large number of sharp steel spikes pointing outwards. Despite such bizarre forms of protection, the offence caused a great deal of fear among the public and it was generally regarded as a very serious threat to law and order. Letters to *The Times* began to appear from gentlemen who had been so attacked and robbed. In response the judges began to order severe floggings in addition to penal servitude in an attempt to stem the growth of the crime. Their example was then followed by Parliament which, against the wishes of the government, enacted the Security From Violence Act 1863.

So if there was so much panic about crime, how come the person who wrote Moldbug's favorite book felt Victorian Britain was crimeless?

I guess it all depends on your perspective. I live less than two miles outside Detroit city limits, and I've never been the victim of a single crime in my life or even felt particularly threatened. Some people just live sheltered existences.

But apparently most other Americans agree with me. According to Gallup, 89% of American men currently feel safe walking alone at night in the city where they live. If 89% of modern US men feel that way, I'm not surprised Moldbug could find one Victorian guy willing to express the opinion.

1.3.2: Why does this matter again?

For some reason, the Reactionaries have made crime an absolute linchpin of their case. A very large portion of Reactionary thought goes implicitly or explicitly through the argument “Progressives have legitimized minorities, minorities cause crime, crime is destroying our society, therefore Progressivism must be destroyed.”

The extent of the Reactionary obsession with crime never fails to amaze me. Moldbug writes:

> Security and liberty do not conflict. Security always wins. As Robert Peel put it, the absence of crime and disorder is the test of public safety, and in anything like the modern state the risk of private infringement on private liberties far exceeds the official of public infringement. No cop ever stole my bicycle.

Desperate times call for desperate measures. On the other hand, non-desperate times call for non-desperate measures. And this is a time when everything is pretty much okay. Murder and violent crime are at historic lows, and almost 90% of American men feel safe walking outside at night. Crime is very nearly a non-issue, and when designing a system of government it is probably a bad idea to give them a blank check to ruin everything else in the pursuit of decreasing it.

1.4: Are people becoming less happy?

Michael’s source for decreasing happiness levels is Blanchflower & Oswald: Well Being Over Time In Britain And The USA. But read the abstract, and you find it’s more complicated: “Reported levels of well-being have declined over the last quarter of a century in the US; life satisfaction has run approximately flat through time in Britain.”

Once again, we find these supposed effects of a global trend are very much limited to individual countries.

Second, when we check the breakdown, we find, as the paper puts it, that “[American] men’s happiness has an upward trend, yet American women’s well-being has fallen through the years.” At a guess, I’d say this is because more women are working full-time jobs. This may be a bit of a victory for Reactionaries, who are no fans of feminism, but it is a very limited victory with little broader implication for other aspects of society. If you’re a man, there’s never been a happier time to be alive.

Further, Blanchflower and Oswald aren’t the only people trying to measure happiness. Ruut Veenhoven has collected 3,651 different happiness studies into a World Database of Happiness. Inglehart, Foa, and Welzel have sorted through some of the data and find that:

> Among the countries for which we have long-term data, 19 of the 26 countries show rising happiness levels. In several of these countries – India, Ireland, Mexico, Puerto Rico, and South Korea – there are *steeply* rising trends. The other countries with rising trends are Argentina, Canada, China, Denmark, Finland, France, Italy, Japan, Luxembourg, the Netherlands, Poland, South Africa, Spain, and Sweden. Three countries, the US, Switzerland, and Norway, show flat trends. Only four countries, Austria, Belgium, UK, and West Germany, show downward trends.

Investigating further:

> By far the most extensive and detailed time series comes from the US, and the full series covering the 60 years from 1946 to 2006 shows a flat trend. But the

> subset from 1946 to 1980 show a downward trend, while the series from 1980 to 2006 shows a rising trend. A similar picture appears from the much scantier British dataset. The entire series from 1946 to 2006 shows a downward trend, but the series from 1980 to the present shows a clear upward trend.

So there you have it. In 19/26 countries, happiness has risen since 1946, and in both America and Britain, it's been rising since 1980.

1.5: Is time preference decreasing?

Time preference is a mathematical formalization of whether people live only for the moment like the proverbial grasshopper, or build for the future like the proverbial ant. We'd probably prefer if people had pretty low time discounting (ie are more ant-like). Michael claims that in fact we're becoming more grasshopper-like.

He cites as his source Wang, Rieger and Hans' How Time Preferences Differ, which is a fascinating study but which does not, as far as I can tell, make anything like the claim Michael says it does. It seems to be entirely about comparing different countries. There is only one thing that looks even close to an intertemporal comparison:

> In particular, 68% of our [2011] US sample chose to wait. For comparison, in the survey by Frederick (2005) where he used the same question…only around 41% of students chose to wait.

Here we see people saving *more* over time, ie becoming more ant-like, although it would be absurd to think this represented a real effect over such a small time period.

Michael may be referring to a claim buried in the study that collectivism is linked to lower discount rates than individualism. This study was done entirely on Israeli Arabs and Jews, with Jews as a proxy for “individualist cultures” and Arabs as a proxy for “collectivist cultures”. Suffice it to say this is not how broad human universals are established. A similar experiment compared Western-primed Singaporeans with Eastern-primed Singaporeans to “conclude” that Confucian cultures had a “longer-term outlook” and thus a lower discount rate. This would be all nice and well except that in the main study, Canadians had a lower discount rate than Japanese, Chinese, Taiwanese, or Koreans. So much for Confucians.

1.6: Is civic participation decreasing?

The argument is simple. Democracy fractures traditionalist societies, destroying civic cohesion, which in turn reduces voter turnout. Therefore, the only way to increase voter turnout is to abolish democracy.

No, actually the argument is more complex, and Michael cites Robert Putnam’s Bowling Alone to make his point for him. Since there is no one statistic for civic participation, I can’t refute it with pure data the same as I tried to do with the others.

But I will point out that Putnam’s own thesis is that it is technology – our options of watching TV, playing video games, or hanging out on the computer – that make us less involved in our communities. He may be right. But blaming the politically neutral force of technology acquits Progressivism.

Even so, a word to defend technology. Right now I am typing a lengthy essay that will be read by a few thousand people.

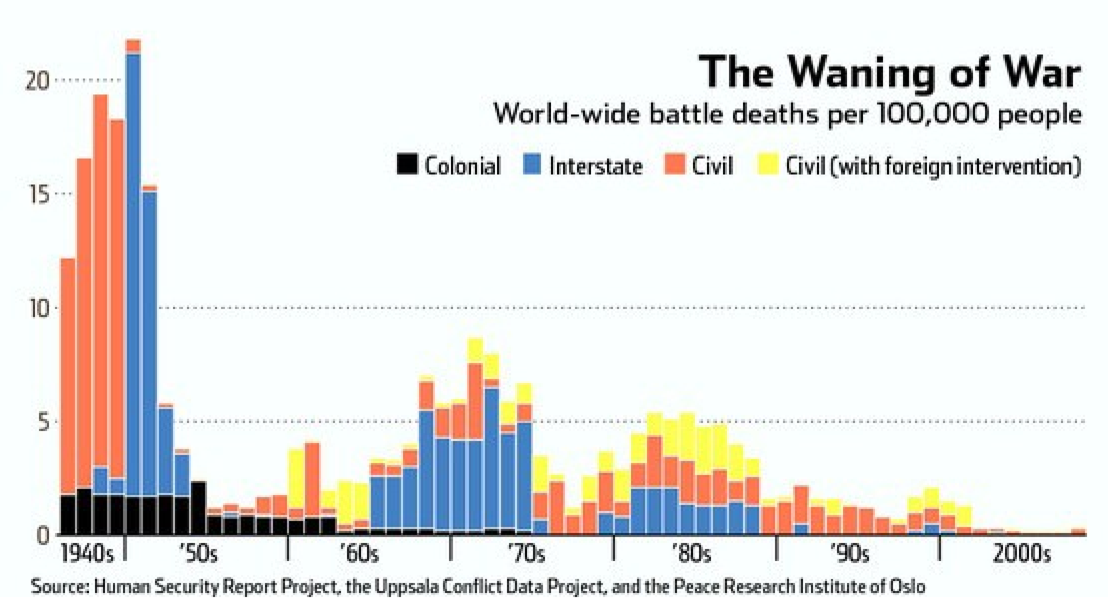
A couple dozen of those will discuss it in the comments. Among those will be people with whom I've had interesting discussions, friendships, and even a couple of romantic relationships. Through the ensuing debate, I will meet new people with whom I will likely keep in touch and discuss my extremely niche interests with on a near daily basis for many years to come, forming bizarre but intellectually fecund communities that will inevitably end up with everyone involved moving to the Bay Area and having kids together.

And we are supposed to be upset because the technology that makes this possible has *cut down on the number of bowling leagues*? That's like condemning butterfly metamorphosis for decreasing the number of caterpillars.

1.7: Are international conflicts becoming more frequent?

This isn't in the paragraph quoted above, but Michael has expressed the opinion to me in person, and anyone familiar with Reactionary thought will recognize this as a staple. The theory is that monarchies had strong international law between them that prevented or settled conflicts quickly, but that democracies have the "sham" international law of the UN (exactly what makes it a sham is never explained) and constantly interfere in one anothers' business as a continuation of their own internal politics or obsession with human rights.

As far as I know no Reactionary has ever dared to cite statistics that they say support this claim, which is probably for the better. But just for the record, here's the counterclaim:



Source: Human Security Report Project, the Uppsala Conflict Data Project, and the Peace Research Institute of Oslo

You can find a much more exhaustive discussion of this topic here.

1.7.1: What about the Concert of Europe? The great statesman Klemens von Metternich used Reactionary ideas to create a brilliant system that kept peace in Europe for nearly a century!

The Concert of Europe lasted from 1815 to 1914. During that time, Europe suffered – just counting major interstate wars involving Congress of Vienna participants – the French Invasion of Spain, the Crimean War, the Schleswig Wars, the Wars of Italian Independence, Austro-Prussian Wars, the Franco-Prussian War, and, let's not forget, World War I.

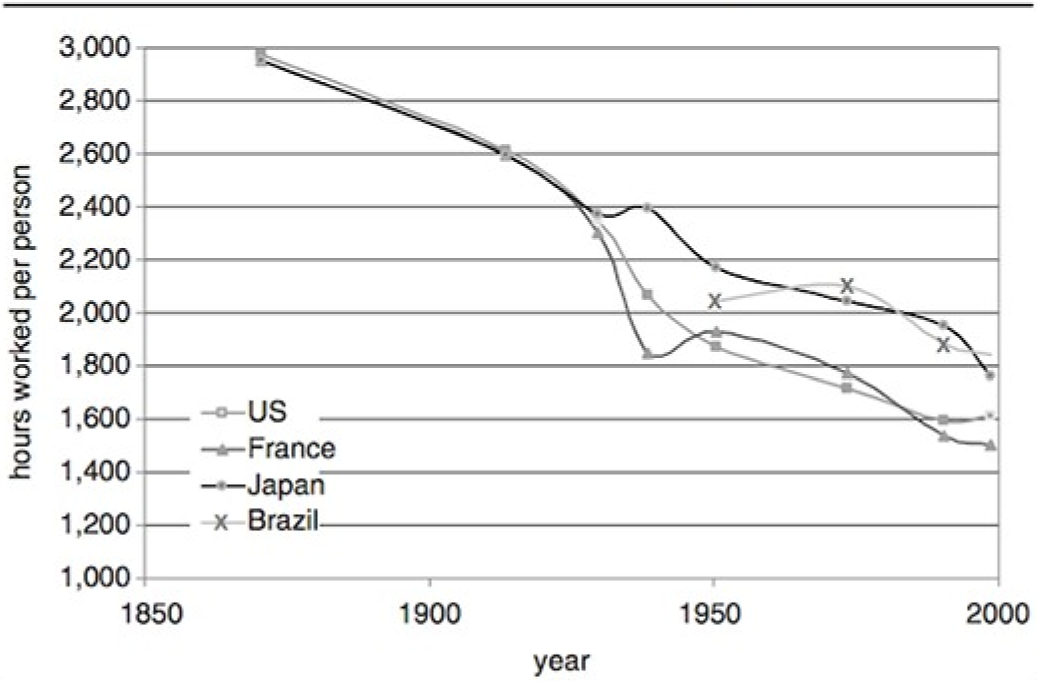
The modern equivalent of the Concert of Europe is the European Union, but built on Progressive rather than Reactionary principles. It has existed from 1951 to 2013 so far, and In those sixty-two years, major interstate wars between EU members have included…well, none.

1.8: Okay, you've discussed the trends Michael listed as supporting Reaction, and found them less than

convincing. Do you have any trends of your own that you think support more modern societies?

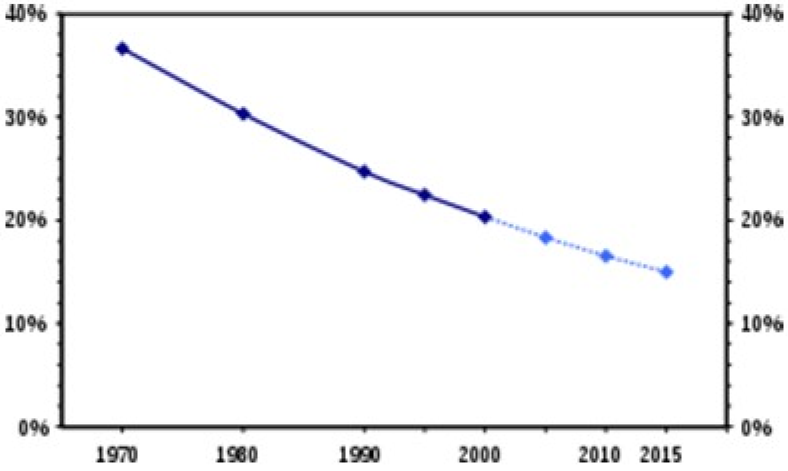
Yes. Most of the graphs below come from [31 Charts That Will Restore Your Faith In Humanity]().

ANNUAL HOURS WORKED PER PERSON, 1870–1978



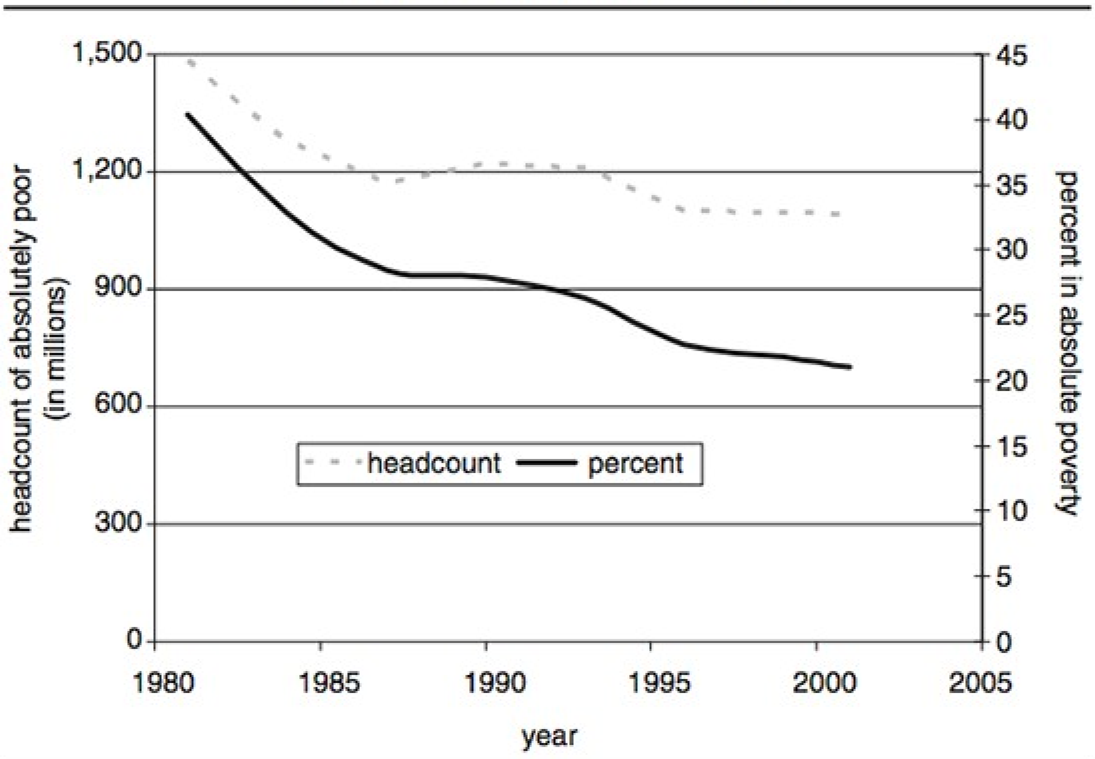
SOURCES: Angus Maddison, *Chinese Economic Performance in the Long Run* (Paris: OECD, 1998); Angus Maddison, *The World Economy: A Millennial Perspective* (Paris: OECD, 2001).

Hours worked per person



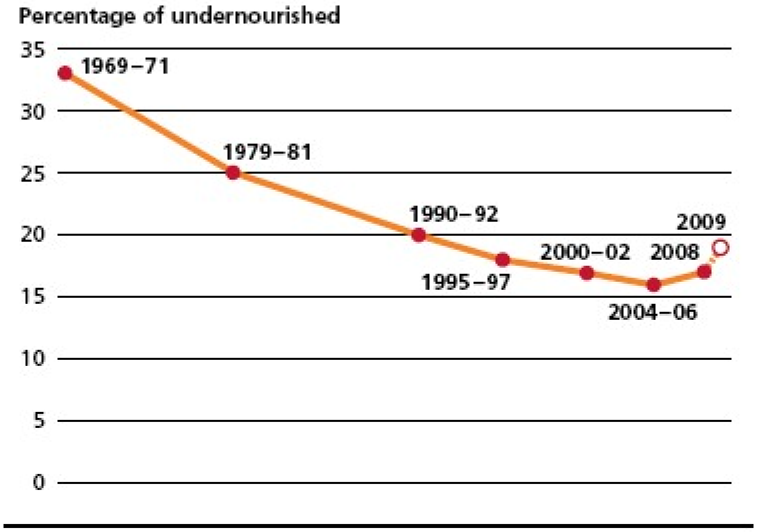
Global illiteracy

GLOBAL POVERTY, 1980–2001



SOURCE: World Bank, *World Development Indicators*, http://devdata.worldbank.org/dataonline (accessed July 12, 2005).

Global poverty



World Hunger

I'm trying to keep things fair by deliberately excluding health care victories since these are at least partially due to technology, but these would include infant mortality dropping a hundredfold, the near elimination of smallpox, diphtheria, polio, tuberculosis, and typhoid from the developed world, the neutralization of AIDS.

Yet in reality, political and social trends played a role here too: for example, smallpox would not have been eliminated without the concerted effort of the WHO and other global health organizations.

1.9: Final thoughts on this section?

Of the seven categories Michael cites as especially supportive of the Reactionary thesis, zero are actually getting worse and several of them appear as best we can tell to be getting better. And I don't want to beat Michael up too much here, because these are the same sorts of things that other Reactionaries cite, and he got picked on only because he was the one to put them all in one place and claim he had evidence.

Reactionary claims that the modern world shows disappointing performance on indicators of social success turn out to be limited to one cherry-picked country or decade or else just plain made up. The very indicators Reactionaries cite turn out, on closer inspection, to provide strong evidence for things getting better.

Progressives, on the other hand, can point to some amazing victories over the last fifty years, including global poverty cut in half, world hunger cut in half, world illiteracy cut in half, war grinding almost to a halt, GDP quintuple-ing, violent crime collapsing, and self-reported happiness increasing in almost all countries.

1.9.1: Other than crime, few of these points have data before 1950, and the crime ones are highly speculative before that date. Don’t you think that even if things have been getting better for the past few decades, they might have been getting worse over the past few millenia?

Yes. In a few cases this is obviously true. For example, Michael cites good data showing that traditional rural societies have lower suicide rates than our own. And obviously they have lower divorce rates. The same may be true with some of the other points here, though probably not as many as Reactionaries would like.

But I do think it’s important to establish that things have been getting better over the past few decades. For one thing, it suggests a different course of action. If things are constantly declining, we should go into panic mode and try a radical restructuring of everything before it’s too late. If things are getting better every day, we should hang tight and try to nudge forward trends that are already going on.

For another, it suggests a different interpretation. If things keep getting worse, we can attribute it to some process of social decay (since everyone seems to agree social decay is Getting Worse All The Time). If things are getting better now, we may perhaps separate societies into two groups, Traditional and Industrialized, admit that the transition from the first to the second caused a whole lot of problems, but be satisfied that industrialized society is gradually improving and fixing its defects.

So while I accept that traditional rural societies a thousand years ago were better on a number of social metrics, I don't think that's particularly actionable. What's actionable is what's going on within industrial societies right now, and that seems to be improvements on all levels.

2: Are traditional monarchies better places to live?

2.1: Are traditional monarchs secure?

Much of the Reactionary argument for traditional monarchy hinges on monarchs being secure. In non-monarchies, leaders must optimize for maintaining their position against challengers. In democracies, this means winning elections by pandering to the people; in dictatorships, it means avoiding revolutions and coups by oppressing the people. In monarchies, elections don't happen and revolts are unthinkable. A monarch can ignore their own position and optimize for improving the country. See the entries on demotism and monarchy here for further Reactionary development of these arguments.

Such a formulation need not depend on the monarch's altruism: witness the parable of Fnargl. A truly self-interested monarch, *if sufficiently secure*, would funnel off a small portion of taxes to himself, but otherwise do everything possible to make his country rich and peaceful.

As Moldbug puts it:

> Hitler and Stalin are abortions of the democratic era – cases of what Jacob Talmon called totalitarian democracy. This is easily seen in their unprecedented efforts to control public opinion, through both propaganda and violence. Elizabeth's legitimacy was a function of her identity – it could be removed only by killing her. Her regime was certainly not the stablest government in history, and nor was it entirely free from propaganda, but she had no need to terrorize her subjects into supporting her.

But some of my smarter readers may notice that "your power can only be removed by killing you" does not actually make you more secure. It just makes security *a lot more important* than if insecurity meant you'd be voted out and forced to retire to your country villa.

Let's review how Elizabeth I came to the throne. Her grandfather, Henry VII, had won the 15th century Wars of the Roses, killing all other contenders and seizing the English throne. He survived several rebellions, including the Cornish Rebellion of 1497, and lived to pass the throne to Elizabeth's father Henry VIII, who passed the throne to his son Edward VI, who after surviving the Prayer Book Rebellion and Kett's Rebellion, named Elizabeth's cousin Lady Jane Grey as heir to the throne. Elizabeth's half-sister, Mary, raised an army, captured Lady Jane, and eventually executed her, seizing the throne for herself. An influential nobleman, Thomas Wyatt, raised another army trying to depose Mary and put Elizabeth on the throne. He was defeated and executed, and Elizabeth was thrown in the

Tower of London as a traitor. Eventually Mary changed her mind and restored Elizabeth’s place on the line of succession before dying, but Elizabeth’s somethingth cousin, Mary Queen of Scots, also made a bid for the throne, got the support of the French, but was executed before she could do further damage.

Actual monarchies are less like the Reactionaries’ idealized view in which revolt is unthinkable, and more like the Greek story of Damocles – in which a courtier remarks how nice it must be to be the king, and the king forces him to sit on the throne with a sword suspended above his head by a single thread. The king’s lesson – that monarchs are well aware of how tenuous their survival is – is one Reactionaries would do well to learn.

This is true not just of England and Greece, but of monarchies the world over. China’s monarchs claimed “the mandate of Heaven”, but Wikipedia’s List of Rebellions in China serves as instructional (albeit voluminous) reading. Not for nothing does the *Romance of Three Kingdoms* begin by saying:

> An empire long united, must divide; an empire long divided, must unite. This has been so since antiquity.

.
Brewitt-Taylor’s translation is even more succinct:

> Empires wax and wane; states cleave asunder and coalesce.

And of Roman Emperors, only about thirty of eighty-four died of even remotely natural causes, according to this List Of Roman Emperors In Order Of How Hardcore Their Deaths Were.

2.2: Are traditional monarchies more free?

A corollary of Reactionaries' "absolutely secure monarch" theory is that monarchies will be freer than democracies. Democrats and dictators need to control discourse to prevent bad news about them from getting out, and ban any institutions that might threaten the status quo. Since monarchs are absolutely secure, they can let people say and do whatever they want, knowing that their words and plans will come to naught. We revisit the Elizabeth quote above:

> Hitler and Stalin are abortions of the democratic era – cases of what Jacob Talmon called totalitarian democracy. This is easily seen in their unprecedented efforts to control public opinion, through both propaganda and violence. Elizabeth's legitimacy was a function of her identity – it could be removed only by killing her. Her regime was certainly not the stablest government in history, and nor was it entirely free from propaganda, but she had no need to terrorize her subjects into supporting her.

It is true that Elizabeth did not censor the newspapers, or bludgeon them into publishing only articles favorable to her. But that is less because of her enlightened ways, and more because all newspapers were banned in England during her reign. English language news in the Elizabethan Era had to be published in (famously progressive and non-monarchical!) Amsterdam, whence it was smuggled into England.

Likewise, Elizabeth and the other monarchs in her line were never shy about killing anyone who spoke out against them. Henry VIII, Elizabeth's father, passed new treason laws which defined as high treason "to refer to the Sovereign offensively in public writing", "denying the Sovereign's official styles and titles”, and “refusing to acknowledge the Sovereign as the Supreme Head of the Church of England”. Elizabeth herself added to these offenses “to attempt to defend the jurisdiction of the Pope over the English Church…”. Needless to say, the punishment for any of these was death, often by being drawn and quartered.

But at least she didn’t have a secret police, right? Wrong. Your source here is Stephen Alford’s book on, well, the Elizabethan secret police, although reason.com’s review, The Elizabethan CIA: The Surveillance State In The 16th Century will serve as a passable summary.

2.2.1: How come we perceive traditional monarchies as less oppressive than for example Stalinist Russia?

Well, for one thing Stalin was in a category all of his own, going *far* beyond rational attempts to maintain his status into counterproductive paranoia. We shouldn’t expect the average *communist police state* to be Stalinist in its intensity, and so we need not be surprised when traditional monarchies aren’t.

But a more comprehensive answer might draw on a proverb of Oceania’s in *1984*: “Animals and proles are free”. Anyone too weak and irrelevant to be dangerous doesn’t suffer the police state’s attention.

Before about the 1600s, the average non-noble neither had nor could have any power. All wealth was locked up in land, owned by nobles, and all military power was locked up in professionals like knights and men-at-arms, who could defeat an arbitrary number of untrained peasants without breaking a sweat.

After about the 1600s, wealth passed into the hands of capitalist merchants – ie non-nobles – and military power became concentrated in whoever could hold a gun – potentially untrained peasants. As a result, kings stopped worrying only about the nobility and started worrying about everyone else.

Or else they didn’t. Remember, all of the longest and most traditional monarchies in history – the Bourbons, the Romanovs, the Qing – were deposed in popular revolts, usually with poor consequences for their personal health. However paranoid and oppressive they were, clearly it would have been in their self-interest to be more so. If monarchy were for some reason to be revived, no doubt its next standard-bearers would not make the same “mistake” as their hapless predecessors.

2.3: Are traditional monarchies less bloody?

Michael Anissimov writes:

> Bad kings are not nearly as bad as Demotist/Communist dictators. Bad kings are in a different universe from bad Demotist leaders. There is not even a vague comparison. In the traditional system, kings rely on the aristocracy and clergy for support, and have trouble doing anything without them. For a Demotist leader, there tends to be far fewer checks and balances. They can cause a million deaths in a place like Iraq with a snap of their fingers. Study up on the history of “death by government” to get a better perspective on what I mean. Kings and emperors very rarely, if ever, engage in mass murder against their own people.

2.3.1: Are demotist countries bloodier?

Look up demotist in a dictionary – Wiktionary will do – and you will find it means “one who is versed in ancient Egyptian demotic writing”. Mr. Anissimov’s use is entirely idiosyncratic to Reactionaries, or, to put it bluntly, made up.

It is interesting that every time Reactionaries make this argument, they use this same made-up word. Here’s Moldbug:

> Let’s define demotism as rule in the name of the People. Any system of government in which the regime defines itself as representing or embodying the popular or general will can be described as “demotist.” Demotism includes all systems of government which trace their heritage to the French or American Revolutions – if anything, it errs on the broad side.
>
> The Eastern bloc (which regularly described itself as “people’s democracy”) was certainly demotist. So was National Socialism – it is hard to see how Volk and Demos are anything but synonyms. Both Communism and Nazism were, in fact, obsessed with managing public opinion. Like all governments, their rule was certainly backed up by force, if more so in the case of Communism (the prewar Gestapo had less than 10,000 employees). But political formulae were of great importance to them. It’s hard to argue that the Nazi and Bolshevik states were any less deified than any clerical divine-right monarchy.

Why use this made-up word so often?

Suppose I wanted to argue that mice were larger than grizzly bears. I note that both mice and elephants are “eargreyish”, meaning grey animals with large ears. We note that eargreyish animals such as elephants are known to be extremely large. Therefore, eargreyish animals are larger than noneargreyish animals and mice are larger than grizzly bears.

As long as we can group two unlike things together using a made-up word that traps non-essential characteristics of each, we can prove any old thing.

None of Michael or Moldbug's interlocutors are, I presume, in favor of Stalinism or Nazism. They are, if anything, in favor of liberal democracies such as the United States or Great Britain. Michael and Moldbug cannot bring up examples of these countries killing millions of their own people, because such examples do not exist. So they simply group them in a made-up category with countries that have, and then tar the entire group by association. This is, of course, a riff on the good old Worst Argument In The World.

If there were any nonmotivated reason to group these countries together – if they were really taxonomically related – there would already be a non-made-up word describing this fact.

So the answer to the question – are demotist countries bloodier than monarchies? – is the same as the answer to the question "are eargreyish animals larger than grizzly bears". The answer is "Here's a nickel, kid; buy yourself a real category ."

2.3.2: Even if the "demotist" idea was invented for this debate, and even if it has little relevance to liberal democracies, isn't it at least a good basis for further study?

Remember Moldbug's definition: "Let's define demotism as rule in the name of the People. Any system of government in which the regime defines itself as representing or embodying the popular or general will can be described as demotist."

But "the leaders have to say they rule in the name of the people" is a pretty low bar. King Louis Philippe of France said he ruled in the name of the people:

> Louis-Philip wore the title of the King of the French…This title was in contrast to the King of France, which reflected a monarchy's power over the country, instead of a king's rule over its people. This title reflects that the king does not take his mandate from God but from the people themselves.

On the other hand, ever read *Les Miserables*? Yeah, that was him. Eventually the *actual* people hated him so much that they had a violent revolution and tried to kill him; the king managed to flee the capital in disguise and escape to England, where he died.

Why accept this stupid standard for the definition of "demotist"? Because a more reasonable one – like "elected by the people" or "liked by the people" or "not universally hated by the people and he has to have a giant army to prevent them from immediately killing him" would exclude for example Stalin, the figure Reactionaries are most desperate to paint as "demotist".

What about the regime which Reactionaries are the *second* most desperate to paint as "demotist"? For this one let's bring some class into this essay and quote Erik Maria Ritter von Kuehnelt-Leddihn:

> As an honest reactionary I naturally reject Nazism … fascism and all related ideologies which are, in sober fact, the reductio ad absurdum of so-called democracy and mob domination.

You heard it here first. The Nazis were *baaaaasically* the same as progressive liberal democrats.

To which all I can say is: you know who *else* opposed “so-called democracy and mob domination?”

> By rejecting the authority of the individual and replacing it by the numbers of some momentary mob, the parliamentary principle of majority rule sins against the basic aristocratic principle of Nature

– Adolf Hitler, *Mein Kampf*, p. 81

2.3.3: Even accepting all that, is Michael’s last sentence even true?

Michael’s argument ends by saying: “Kings and emperors very rarely, if ever, engage in mass murder against their own people.”

I propose a contrary hypothesis – traditional absolutist regimes have always had worse records of massacre and genocide than progressives. However, technology improves efficiency in all things, including murder. And population has been growing almost monotonically for millennia. Therefore, it is unsurprising that more modern absolutist regimes – like Nazism and Stalinism – have higher death counts than older absolutist regimes – like traditional monarchies.

On the other hand, traditional monarchies have some *pretty impressive* records for killing their own people. Let us take a whirlwind tour of history:

The Albigensian Crusade, run by the French monarchy against its own subjects – with the support of the Catholic Church – may have killed up to a million people, which is pretty impressive considering that at the time there were only about twelve million Frenchmen. As a proportion of total population, this is about the same as the number of Germans who died during World War II, or Chinese who died during the Great Leap Forward.

The Harrying of the North was totally a real historical event and not something I stole from Game of Thrones. William the Conquerer, angry at the murder of a local earl, managed to kill about 100,000 northern Englishmen from 1069-1070, which was probably about 5% of the entire population.

Another 100,000 people died in the 16th century German Peasants' War, an event which so blended into the general mayhem of the time that you have never heard of it. Actually, the claim that Reactionary regimes have ever been peaceful would have trouble surviving a look merely at Wikipedia's *disambiguation* page for Peasants' War.

Third century BC emperor Qin Shi Huang was not only responsible for the Burning Of Books And Burying Of Scholars, but killed about one million out of his population of twenty million with various purges and forced labor projects, one of which was the Great Wall of China.

[This section previously included a paragraph on Chinese warlord Zhang Xianzhong. Despite living in a 17th century monarchy, he held some pretty progressive values and his Reactionary credentials have been challenged. Rather than let his story distract from the more obviously Reactionary murderers above, I will concede the point]

But Michael goes even further. He says of democracies that "[with] a Demotist leader, there tends to be far fewer checks and balances. They can cause a million deaths in a place like Iraq with a snap of their fingers."

Ignoring for a moment the difference between snapping one's fingers and getting a bill to declare war passed through both houses of a hostile Congress (since Michael certainly does) we note that Michael has just authorized us to also compare monarchies and democracies in their ability to wreak havoc abroad.

On this particular historical tour, we will start with King Leopold of Belgium. Belgium itself was a constitutional monarchy run on a mostly democratic system, and in fact has always been a relatively pleasant and stable place. However, Belgium's colony, the Congo Free State, was under the direct rule of King Leopold. Not only was it responsible for the deaths of two to fifteen million Congolese – ie about as many Jews as were killed by Hitler – but the manner of those deaths was about as brutal and callous as can be imagined. Wikipedia writes:

> Leopold then amassed a huge personal fortune by exploiting the Congo. The first economic focus of the colony was ivory, but this did not yield the expected levels of revenue. When the global demand for rubber exploded, attention shifted to the labor-intensive collection of sap from rubber plants. Abandoning the promises of the Berlin Conference in the late 1890s, the Free State government restricted foreign access and extorted forced labor from the natives. Abuses, especially in the rubber industry, included the effective enslavement of the native population, beatings, widespread killing, and frequent mutilation when the production quotas were not met. Missionary John Harris of Baringa, for example, was so shocked by what he had come across that he wrote to Leopold's chief agent in the Congo saying: "I have just returned from a journey inland to the village of Insongo Mboyo. The abject misery and utter abandon is positively indescribable. I

> was so moved, Your Excellency, by the people's stories that I took the liberty of promising them that in future you will only kill them for crimes they commit."

This is an especially good example as it describes (we will see later) the ideal Reactionary state – one run by a single person identical to a corporation trying to make as much money as possible off a particular area and possessing overwhelming force.

The story does however have a happy ending – progressive elements within Belgium were so horrified that they forced the king to cede his claim – the colony was then governed by Belgium's democratically elected legislature, which did such a good job even Mencius Moldbug cannot resist the urge to praise it, and under whose rule Congo was a relatively liveable place up until a native uprising kicked out the Belgians and restored dictatorship.

Another good example of kings and emperors at war is Imperial Japan. This state – again run under principles no Reactionary could fault – accomplished the astounding feat of reducing the Nazis to the *second* biggest jerks on the Axis side during World War II. During the war, Imperial Japanese troops murdered between three million and ten million foreigners, mostly Chinese. Once again the brutality of their killings is impressive. According to Wikipedia on the Rape of Nanking:

> The International Military Tribunal for the Far East estimated that 20,000 women were raped, including infants and the elderly.[40] A large portion of these rapes were systematized in a process where soldiers would search door-to-door for young girls, with many women taken captive and gang raped.[41] The women were often killed immediately after being raped, often

> through explicit mutilation[42] or by stabbing a bayonet, long stick of bamboo, or other objects into the vagina. Young children were not exempt from these atrocities, and were cut open to allow Japanese soldiers to rape them

Meanwhile, Michael says that “Kings and emperors very rarely, if ever, engage in mass murder” but is *absolutely horrified* that America caused a million deaths in Iraq (more sober sources say 100,000, of which under 10,000 were civilians directly killed by US forces) while making the utmost effort to avoid unnecessary violence and launching war crimes proceedings against anyone caught employing it.

2.3.4: Conclusion for this section?

Reactionaries believe that monarchs are wise and benevolent rulers, and that it is only “demotists” who engage in genocide and mass murder.

But this argument is based on a con – “demotist” is an unnatural category they made up solely to win this debate. When we look at the governments their opponents actually support – liberal democracies – we find they have a much better history than monarchies.

Further, the Reactionaries fail *even on the terms of their own con*. Monarchs have a *fantastically* bloody history, and the regimes they want to paint as demotist really aren’t.

2.4: Are traditional monarchs good leaders?

In his perhaps optimistically named “Ten Objections To Traditionalism And Monarchism, With Answers”, Michael Anissimov asks, with commendable bluntness: “What if the king is an idiot or psycho?” He answers:

> Then the prior king appoints a regent to take over the affairs of state on behalf of his successor. There is also a debate within the Reactionary community as to whether adoptive succession is preferable to hereditary succession, which avoids the issue of stupid or crazy children. Such extreme scenarios rarely ever happened during the age of Renaissance European monarchs. One of the greatest statesmen of all time, Klemens von Metternich, strongly influenced the mentally deficient monarch Ferdinand I of Austria during his reign, sat on the regency council, and ran most important affairs, presiding over a hundred years of relative peace in Europe.

We shall start with the theoretical objections before moving on to the empirical counterexamples.

Theoretical objection the first: what if the king doesn’t become an idiot or a psycho until after he is on the throne? The onset of schizophrenia can be as late as twenty-five; later in rare cases. Traumatic brain injury, certain infectious diseases, and normal human personality change can happen at any age. Smart psychopaths will have the presence of mind to avoid revealing their psychosis until they are safely enthroned.

Theoretical objection the second: what if the king seizes power some other way? A decent number of history’s monarchs got tired of waiting and killed their fathers. We would expect these to disproportionately include those who are crazy and evil, not to mention those who think their fathers would take away their power.

Theoretical objection the third: regency councils are historically about the least stable form of government imaginable. Unless everyone has truly commendable morality, either the king kills the regent and seizes power, the regent kills the king and starts a new dynasty, or some third party kills the regent and becomes the new regent. Once again, reading *Romance of the Three Kingdoms* will prove instructional.

Theoretical objection the fourth: we are counting on the king's father to object if the king is an idiot or psycho. But a lot of idiotic psychotic kings' fathers were, in fact, idiots and psychos. The apple doesn't fall very far from the tree.

Onto the historical counterexamples. Historical counterexample the first: Gaius Julius Caesar Augustus Germanicus, "Caligula" to his friends. Absolutely beloved by the Roman populace. Unclear whether he killed his uncle Tiberius to gain the Empire, or just stood by cackling kind of maniacally as he died. Took power to general acclaim, ruled well for a couple of months, gradually started showing his dark side, and after a year or two reached the point where he ordered a large section of spectators at the colosseum to be thrown into the ring and torn apart by lions because the average amount of tearing-apart-by-lions at a Roman gladiatorial games *just wasn't enough for him*.

Historical counterexample the second: Ivan the Terrible. His father died of infection when Ivan was three years old. His mother was named as his regent – kind of a coincidence that the most qualified statesman in the realm would be his mother, but let's roll with it – but she died of poisoning when Ivan was eight. In this case I'm not sure who exactly is supposed to decide whether he's an idiot or psycho, and apparently neither were the Russians, because they crowned him Czar in 1547 . Ivan was okay until his wife died, at which point he became paranoid and started executing the nobility for unclear reasons, destroyed the economy, and burnt and pillaged the previously glorious city of Novgorod (part of his own kingdom!) with thousands of deaths. According to some sources:

> Ivan himself often spent nights dreaming of unique ways to torture and kill. Some victims were fried in giant frying pans and others were flayed alive. At times, he turned on [his death squads] themselves, and subjected their membership to torture and death. In a fit of rage, he murdered his own son; however the guilt of this act obsessed him and he never recovered.

Our story does not end there! Ivan died of a stroke, leaving the throne to his intellectually disabled son. Here at least the system worked – brilliant statesman Boris Godunov was installed as regent and ruled pretty well. He did, however, eventually seize the throne – likely because if he had not seized the throne everyone else would have killed him out of suspicion that he might seize the throne. He died, there was a huge succession squabble, and thus started the Time of Troubles, whose name is pretty self-explanatory.

Historical counterexample the third: Charles II Habsburg of Spain (not to be confused with various other Charles IIs). A strong contender for the hotly contested title of “most inbred monarch in history”, Wikipedia describes him like so:

> Known as “the Bewitched” (Spanish: el Hechizado), he is noted for his extensive physical, intellectual, and emotional disabilities—along with his consequent ineffectual rule…
>
> Charles did not learn to speak until the age of four nor to walk until eight, and was treated as virtually an infant until he was ten years old. Fearing the frail child would be overtaxed, his caretakers did not force Charles to attend school. The indolence of the young Charles was

> indulged to such an extent that at times he was not expected to be clean. When his illegitimate half-brother Don Juan José of Austria, an illegitimate son of Philip IV, obtained power by exiling the queen mother from court, he covered his nose and insisted that the king at least brush his hair

As Charles's father died when Charles was 3, he was given a regent – his mother (*another* case in which the most qualified statesman in the land is the monarch's mother! What are the odds?!) But when his mother died, Charles took power in his own name and ruled for four years. His only notable achievement during that time was presiding over the largest auto-da-fe in history. He died at age 39. Again quoting Wikipedia:

> The physician who practiced his autopsy stated that his body "did not contain a single drop of blood; his heart was the size of a peppercorn; his lungs corroded; his intestines rotten and gangrenous; he had a single testicle, black as coal, and his head was full of water." As the American historians Will and Ariel Durant put it, Charles II was "short, lame, epileptic, senile, and completely bald before 35, he was always on the verge of death, but repeatedly baffled Christendom by continuing to live."

Oh, and thanks to the vagaries of self-interested royal dynasties, his passing caused a gigantic succession struggle which drew in all the neighboring countries and caused hundreds of thousands of deaths.

Historical counterexample the fourth: Henry VIII. Really? Yes, really. While perhaps calling him an idiot or psycho goes too far, he certainly thought that marrying confirmed hottie Anne Boleyn and having a son with her was worth converting England to a newly-invented Protestant religion – a decision which killed tens of thousands, displaced some of the country's oldest and most important institutions, and set the stage for two hundred years of on-and-off warfare. Whether or not you like the Church of England (or, as it was almost named, Psychotic Bastard Religion) yourself, you have to admit this is a sort of poor reason to start a religious revolution.

King Henry wasn't an idiot or a psycho. He was just a selfish bastard. You can't expect his father to pick up on that. Even if you could, his father wasn't exactly Mahatma Gandhi himself. Worst of all, his personality may have changed following traumatic brain injury from a jousting accident – something that could not have been predicted before he took the throne.

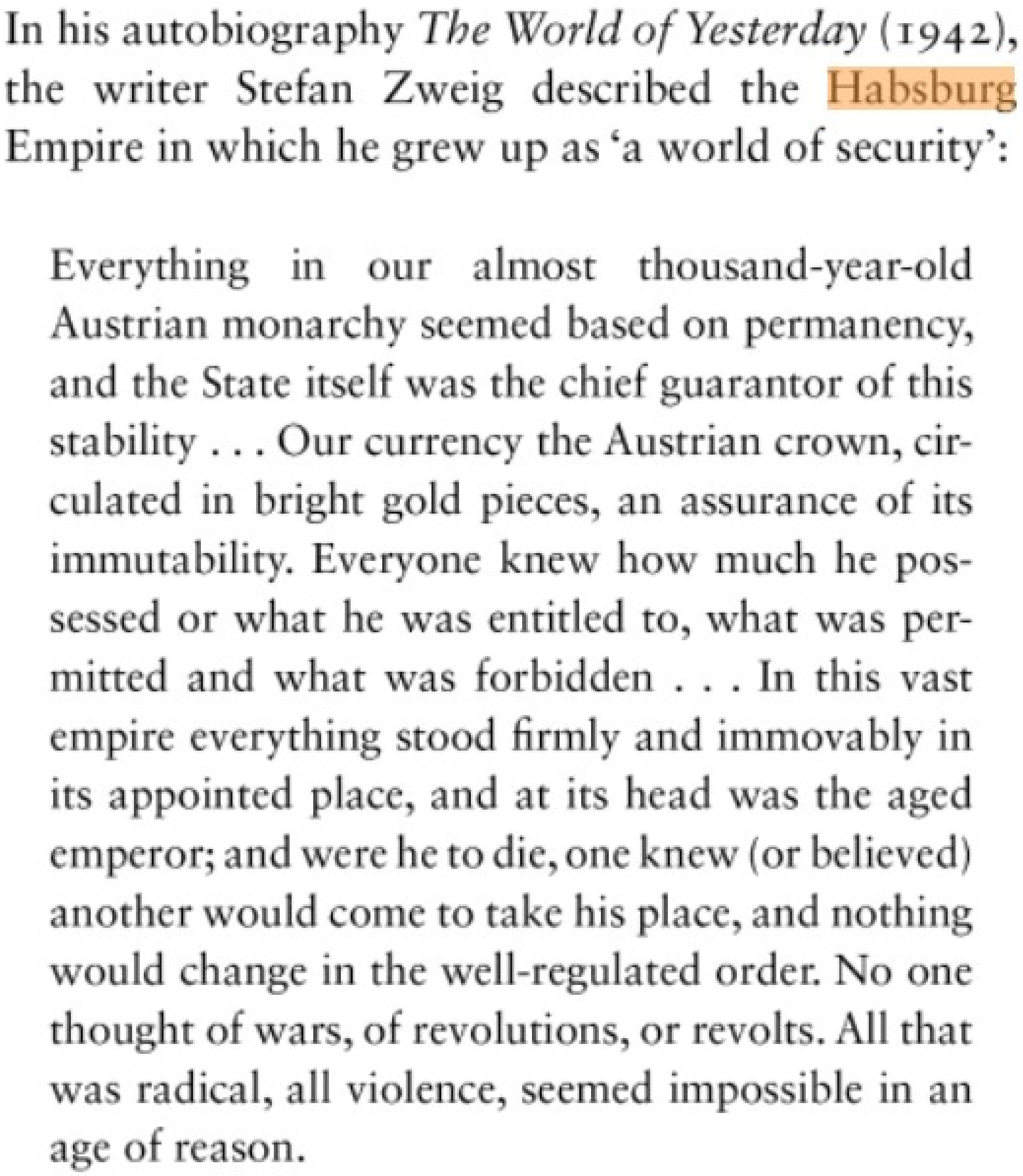
This is exactly the sort of problem non-monarchies don't have to worry about. If Barack Obama said the entire country had to convert to Mormonism at gunpoint as part of a complicated plot for him to bone Natalie Portman, we'd just tell him no.

There's another important aspect here too. Reactionaries – ending up more culpable of a stereotype about economists than economists themselves, who are usually pretty good at avoiding it – talk as if a self-interested monarch would be a rational money-maximizer. But a monarch may have desires much more complicated than cash. They might, like Henry, want to marry a particular woman. They might have religious preferences. They might have moral preferences. They might be sadists. They might really like the color blue. In an ordinary citizen, those preferences are barely even interesting enough for small talk. In a monarch, they might mean everyone's forced to wear blue clothing all the time.

You think that’s a joke, but in 1987 the dictator of Burma made all existing bank notes illegitimate so he could print new ones that were multiples of nine. Because, you see, he *liked* that number. As Wikipedia helpfully points out, “The many Burmese whose saved money in the old large denominations lost their life savings.” For every perfectly rational economic agent out there, there’s another guy who’s *really* into nines.

2.5: Are traditional monarchies more politically stable?

Reactionaries often claim that traditional monarchies are stable and secure, compared to the chaos and constant danger of life in a democracy. Michael Anissimov quotes approvingly a passage by Stefan Zweig:



In his autobiography *The World of Yesterday* (1942), the writer Stefan Zweig described the Habsburg Empire in which he grew up as 'a world of security':

> Everything in our almost thousand-year-old Austrian monarchy seemed based on permanency, and the State itself was the chief guarantor of this stability . . . Our currency the Austrian crown, circulated in bright gold pieces, an assurance of its immutability. Everyone knew how much he possessed or what he was entitled to, what was permitted and what was forbidden . . . In this vast empire everything stood firmly and immovably in its appointed place, and at its head was the aged emperor; and were he to die, one knew (or believed) another would come to take his place, and nothing would change in the well-regulated order. No one thought of wars, of revolutions, or revolts. All that was radical, all violence, seemed impossible in an age of reason.

Michael's comment: "[This] does a good job capturing the flavor and stability of the Austrian monarchy…it's very interesting to read this in a world where America and Europe are characterized by political and economic instability and ethnic strife."

I am glad Mr. Zweig (Professor Zweig? Baron Zweig?) found his life in Austria to be very secure. But we can't just take him at his word.

Let’s consider the most recent period of Habsburg Austrian history – 1800 to 1918 – the period that Zweig and the elders he talked to in his youth might have experienced.

Habsburg Holy Roman Austria was conquered by Napoleon in 1805, forced to dissolve as a political entity in 1806, replaced with the Kingdom of Austria, itself conquered again by Napoleon in 1809, refounded in 1815 as a repressive police state under the gratifyingly evil-sounding Klemens von Metternich, suffered 11 simultaneous revolutions and was almost destroyed in 1848, had its constitution thrown out and replaced with a totally different version in 1860, dissolved entirely into the fledgling Austro-Hungarian Empire in 1867, lost control of Italy and parts of Germany to revolts in the 1860s-1880s, started a World War in 1914, and was completely dissolved in 1918, by which period the reigning emperor’s wife, brother, son, and nephew/heir had all been assassinated.

Meanwhile, in Progressive Britain during the same period, people were mostly sitting around drinking tea.

This is not a historical accident. As discussed above, monarchies have traditionally been rife with dynastic disputes, succession squabbles, pretenders to the throne, popular rebellions, noble rebellions, impulsive reorganizations of the machinery of state, and bloody foreign wars of conquest.

2.5.1: And democracies are more stable?

Yes, yes, oh God yes.

Imagine the US presidency as a dynasty, the Line of Washington. The Line of Washington has currently undergone forty-three dynastic successions *without a single violent dispute*. As far as I know, this is unprecedented among dynasties – unless it be the dynasty of Japanese Emperors, who managed the feat only after their power was made strictly ceremonial. The closest we've ever come to any kind of squabble over who should be President was Bush vs. Gore, which was decided within a month in a court case, which both sides accepted amicably.

To an observer from the medieval or Renaissance world of monarchies and empires, the stability of democracies would seem *utterly supernatural*. Imagine telling Queen Elizabeth I – whom as we saw above suffered six rebellions just in her family's two generations of rule up to that point – that Britain has been three hundred years without a non-colonial-related civil war. She would think either that you were putting her on, or that God Himself had sent a host of angels to personally maintain order.

Democracies are vulnerable to one kind of conflict – the regional secession. This is responsible for the only (!) major rebellion in the United States' 250 year (!) history, and might be a good category to place Britain's various Irish troubles. But the long-time scourge of every single large nation up to about 1800, the power struggle? Totally gone. I don't think moderns are sufficiently able to appreciate how big a deal this is. It would be like learning that in the year 2075, no one even remembers that politicians used to sometimes lie or make false promises.

How do democracies manage this feat? It seems to involve three things:

First, there is a simple, unambiguous, and repeatable decision procedure for determining who the leader is – hold an election. This removes the possibility of competing claims of legitimacy.

Second, would-be rebels have an outlet for their dissatisfaction: organize a campaign and try to throw out the ruling party. This is both more likely to succeed and less likely to leave the country a smoking wasteland than the old-fashioned method of raising an army and trying to kill the king and everyone who supports him.

Third, it ensures that the leadership always has popular support, and so popular revolts would be superfluous.

If you remember nothing else about the superiority of democracies to other forms of government, remember the fact that in three years, we will have a change of leadership and almost no one is stocking up on canned goods to prepare for the inevitable civil war.

2.6: Are traditional monarchies more economically stable?

Once again, we come to Michael Anissimov’s claims about Austria:

> Demotist systems, that is, systems ruled by the “People,” such as Democracy and Communism, are predictably less financially stable than aristocratic systems. On average, they undergo more recessions and hold more debt. They are more susceptible to market crashes. They waste more resources. Each dollar goes further towards improving standard of living for the average person in an aristocratic system than in a Democratic one.
>
> The economic growth of the Austro-Hungarian Empire (1.76% per year) “compared very favorably to that of other European nations such as Britain (1%), France (1.06%), and Germany (1.51%)”.

The growth of Austria-Hungary was higher than that of other European countries for the same reason the growth of sub-Saharan Africa right now is outpacing the growth of America or Europe – it was such a backwater that it had more room to grow.

Urbanization is a decent proxy for industrialization, and we consistently find that throughout the Kingdom of Austria and Austro-Hungarian Empire period, Austria had some of the lowest urbanization rates in Europe, just barely a third those of Britain, and well behind those of France, Spain, Italy, Germany, and Switzerland. In order to find a country as poorly developed as Austria-Hungary, we need to go to such economic powerhouses as Norway, Portugal and Bulgaria.

Nor was its economy especially stable. The Panic of 1873, probably the worst financial depression during the period being discussed and perhaps the worst modern economic crisis before the Great Depression, actually *started* in Austria-Hungary and only spread from there to the rest of the world. This is especially astounding given Austria-Hungary's general economic irrelevance at the time.

2.6.1: What about Germany? Isn't the German Empire a good example of an industrially successful Reactionary country?

I consider the Reactionary credentials of the German Empire extremely open to doubt.

The German Empire was a utopian project created by people who wanted to sweep away the old patchwork system of landed nobility and local traditions that formed the Holy Roman Empire and turn it into a efficient modern state. The Progressive origins of both the Italian and German unification efforts shine through almost every word of a letter from Garibaldi to German unification pioneer Karl Blind:

> The progress of humanity seems to have come to a halt, and you with your superior intelligence will know why. The reason is that the world lacks a nation which possesses true leadership. Such leadership, of course, is required not to dominate other peoples, but to lead them along the path of duty, to lead them toward the brotherhood of nations where all the barriers erected by egoism will be destroyed. We need the kind of leadership which, in the true tradition of medieval chivalry, would devote itself to redressing wrongs, supporting the weak, sacrificing momentary gains and material advantage for the much finer and more satisfying achievement of relieving the suffering of our fellow men. We need a nation courageous enough to give us a lead in this direction. It would rally to its cause all those who are suffering wrong or who aspire to a better life, and all those who are now enduring foreign oppression.
>
> This role of world leadership, left vacant as things are today, might well be occupied by the German nation. You Germans, with your grave and philosophic character, might well be the ones who could win the confidence of others and guarantee the future stability of the international community. Let us hope, then, that you can use your energy to overcome your moth-eaten thirty tyrants of the various German states. Let us hope that in the center of Europe you can then make a unified nation out of your fifty millions. All the rest of us would eagerly and joyfully follow you.

The result of this idealistic vision – the destruction of the *ancien regime* in Germany – was a state much stronger than the traditional-but-weak Holy Roman Empire or anything that had existed in that part of the world before.

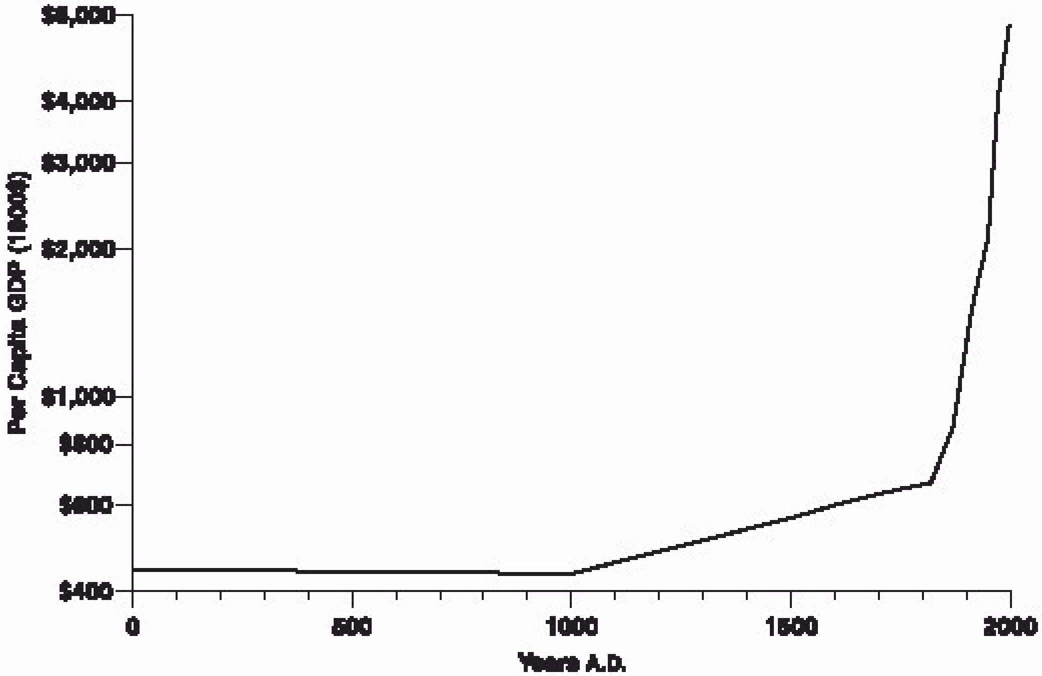
Sure, Otto von Bismarck was no hippie, but he was first and foremost a pragmatist, and his empire combined both conservative and progressive elements. It was based on a constitution, had universal male suffrage (only 5 years after the US got same!), elected a parliament, and allowed political parties. Granted, the democratic aspect was something of a facade to cover up an authoritarian core, but real Reactionaries would not permit such a facade, saying it will invariably end in full democracy (they are likely right).

The amazing growth of the German Empire was due to two things. First, the virtues of the German populace, which allow them to continue to dominate the European economy even today with an extremely progressive and democratic government. And second, the catch-up effect mentioned earlier. Germany had been languishing under traditional feudal and aristocratic rule for centuries. As soon as the German Empire wiped away that baggage and created a modern Progressive state, it allowed the economic genius of the Germans to shine through in the form of breakneck-speed economic growth.

2.6.2: Is Progressivism destroying the economy?

Another frequent claim. But remember how Michael said Progressivism went into high gear around the time of the French Revolution in 1789. Here’s a graph of world GDP over time:

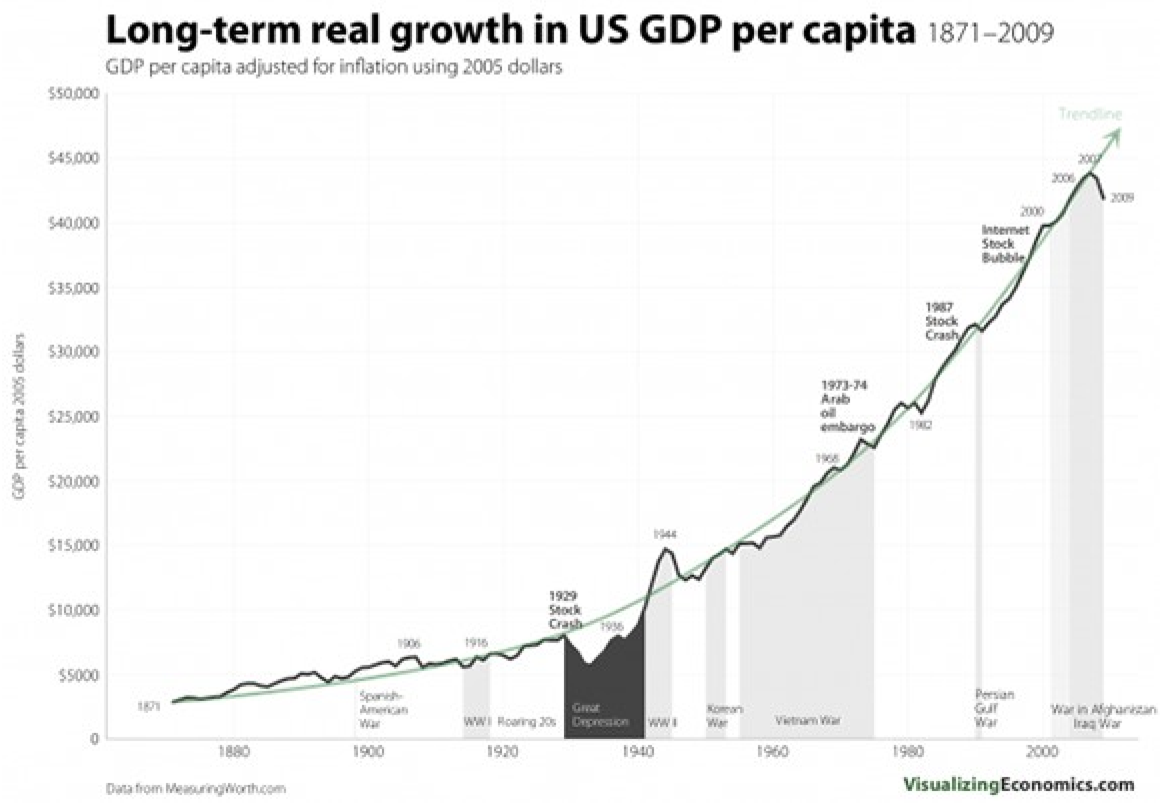
FIGURE 1–4 WORLD PER CAPITA GDP (INFLATION-ADJUSTED)



Source: Maddison, *The World Economy: A Millennial Perspective*, 264.

To put it lightly, I see no evidence of a decline starting around 1789?

Maybe the effect is just in the United States?



This image is actually even *more* astounding and important than the above, because it shows how growth keeps to a very specific trendline. On the graph above, the Reactionary might claim that technological advance was disguising the negative effects of Progressivism somehow. Here we see that *no* second variable that is not perfectly consistent has been interfering with the general economic growth effect.

I literally cannot conceive of a way that the data could be less consistent with the theory that Progressivism inhibits economic growth.

2.7: Are traditional monarchies just in general more successful and nicer places to live?

Great Britain and America have throughout their histories been the two most progressive nations on Earth. They've also been, over the past three hundred years or so, the two most successful. Other bright spots in the progressive/successful cluster include 1600s Netherlands, classical democratic Athens, republican Rome, and Cyrus' Persia. In fact, practically every one of the great nations of history was unusually progressive for its time period, perhaps with the exception of China – which is exceptionally complicated and hard to place on a Western political spectrum. Other possible exceptions might include Philip II's Spain, Louis XIV's France, and Genghis Khan's Mongolia – but the overall trend is still pretty clear.

Limiting our discussion to the present, our main obstacle to a comparison is a deficit of truly Reactionary countries. Reactionaries are never slow to bring up Singapore, a country with some unusually old-fashioned ideas and some unusually good outcomes. But as I have pointed out in a previous post, Singapore does little better than similar control countries, and the lion's share of its success is most likely due to it being a single city inhabited by hyper-capitalist Chinese and British people on a beautiful natural harbor in the middle of the biggest chokepoint in the world's most important trade route.

Saudi Arabia also gets brought up as a modern Reactionary state. It certainly has the absolute monarchy, the reliance on religious tradition, the monoethnic makeup, the intolerance for feminist ideals, and the cultural censorship. How does it do? Well, it's nice and stable and relatively well-off. But a cynic (or just a person with an IQ > 10) might point out that a lot of this has to do with it controlling a fifth of the world's oil supply. It's pretty easy to have a good economy when the entire world is paying you bazillions of dollars to sit there and let them extract liquid from the ground. And it's pretty easy to be stable when you can bribe the population to do what you want with your bazillions of dollars in oil money – in fact, Saudi Arabia is probably that rarest of birds – a Reactionary welfare state.

(Actually, this point requires further remark. Reactionary states tend to be quite rich. In the case of Singapore, Reactionaries trumpet this as a success of Reactionary principles. In the case of Saudi Arabia, that sort of causation is somewhat less credible. I propose an alternative theory: Reactionary states can maintain themselves only by bribing the population not to revolt. These bribes may be literal, as in the case of the Saudi welfare state. Or they may be more figurative – “Look how rich my government has made you – you let me stay in power and I’ll keep up the good work.” China is the classic example of this particular formulation. This is important because contra Moldbug’s inverted pendulum theory it suggests Reactionary regimes will be inherently unstable.)

But getting back to the issue at hand – given all these economic confounders, it’s hard to compare Reactionary and progressive regimes in an even-handed way.

This is par for the course. Political science is notorious for its inability to perform controlled experiments, and no two countries will differ only in their system of government.

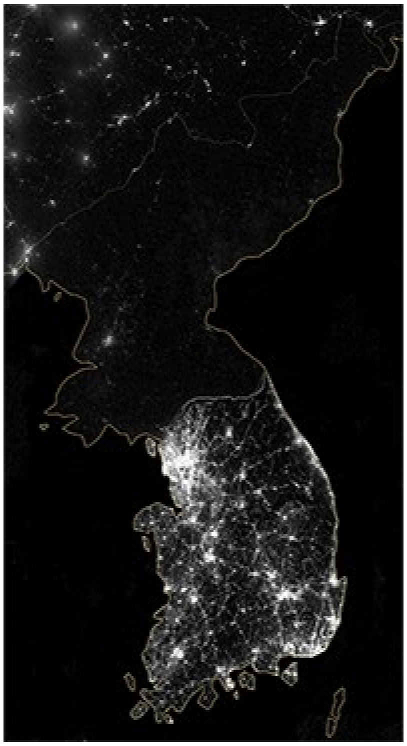
2.7.1: If we could perform a controlled experiment pitting reactionary versus progressive ideals, what would it look like?

Well, assuming you were God and had infinite power and resources, you could take a very homogeneous country and split it in half.

One side gets a hereditary absolute monarch, whose rule is law and who is succeeded by his sons and by his sons’ sons. The population is inculcated with neo-Confucian values of respect for authority, respect for the family, and cultural solidarity, but these values are supplemented by a religious ideal honoring the monarch as a near-god and the country as a specially chosen holy land. American cultural influence is banned on penalty of death; all media must be produced in-country, and missionaries are shot on site. The country's policies are put in the hands of a group of technocratic nobles hand-picked by the king.

The other side gets flooded with American missionaries preaching weird sects of Protestantism, and at the point of American guns is transformed into a parliamentary democracy. Its economy – again at the behest of American soldiers, who seem to be sticking around a sufficient long time – becomes market capitalism. It institutes a hundred billion dollar project to protect the environment, passes the strictest gun control laws in the world, develops a thriving gay culture, and elects a woman as President.

Turns out this perfect controlled experiment actually happened. Let's see how it turned out!



Talk about your “Dark Enlightenment”!

From the Reactionary perspective, North Korea has done everything right. They’ve had three generations of absolute rulers. They’ve tried to base their social system on Confucianism. They’ve kept a strong military, resisted American influence, and totally excluded the feelings of the peasant class from any of their decisions.

Reactionaries, behold your god.

South Korea, on the other hand, ought to be a basketcase. It’s replaced its native Confucian traditions with liberal Protestant sects, it’s occupied by US troops, it’s gone through various military coups to what the CIA calls a “fully functioning modern democracy”, and it’s so culturally decadent and degraded that it managed to produce *Gangnam Style*. Yet I don’t think there’s a single person reading this who doesn’t know which one ze’d rather live in.

Yet according to the principles of Reaction (first quote Michael Anissimov, second Mencius Moldbug)

> Legally speaking, monarchies tend to have fewer laws, but enforce them more strictly, following Tacitus’ dictum: “The more corrupt the state, the more numerous the laws.” In general, monarchies put more power into the hands of local government. A key argument in favor of monarchy is that leaders tend to have a lower time preference, meaning they have a greater personal stake in the long-term well being of the

> country, compared to career politicians oriented towards four-year election cycles.
>
> A royal family is a family business. Not one king in European history can be found who ruined his own country to enrich himself, like an African dictator.

North Korea is a family business. And the Kim family has done very very well for itself. But it’s not something I would like to see spread.

3: What is progress?

Reactionaries are not the first to notice – but may be the most obsessive in analyzing – a certain directionality to history. That is, rather than being a random walk across the space of possible values, at least the past three hundred years or so seem to have shown a definite trend. Those who are in favor of this trend call it “progress”. Those who oppose it call it things like “moral decay”.

However, it is notoriously difficult to determine exactly what this trend is and what drives it. A theory to this effect is at the core of what separates Reactionaries from simple conservatives.

In the remainder of this section, I will replace the word “progress” – with its connotations of inevitability and desirability – with the preferred Reactionary term “progressivism” – that is, the political ideology which flows with the historical trend under discussion.

3.1: Might Progressivism be merely a secular strain of some Protestant religion?

Reactionaries seem to agree that Progressivism is a religion. Perhaps Calvinism. From Moldbug:

> I prefer “cryptocalvinism” [as a name for progressivism], meaning two things: that, like Calvin and as a direct result of his intellectual heritage, cryptocalvinists are building the Kingdom of God on Earth, a political system that seeks to eradicate every form of unrighteousness; and that they prefer not to acknowledge this characterization of their mission and heritage. Since I’ve changed the name, let me repeat the four ideals of cryptocalvinism: Equality (the universal brotherhood of man), Peace (the futility of violence), Social Justice (the fair distribution of goods), and Community (the leadership of benevolent public servants).

Or perhaps Quakerism. From Isegoria, quoting a different Moldbug theory:

> Modern progressivism is in fact a form of secular Quakerism, with its doctrine of the Inner Light only slightly modified.

Or how about Judaism? From Age of Treason:

> In a nutshell I object to [Moldbug]’s definition of Universalism, which is what he calls “the faith of our ruling caste”. It’s an important observation, but I think he gets it only half right. He associates Universalism only with Progressivism, which he blames entirely on Christianity. He does not address the Globalist tendencies of our ruling caste, and he pretty much gives Jews a pass…The close alignment of PC with Jewish interests? The Jewish support for Marxism and Bolshevism and hatred of Nazism perhaps?

Reactionaries seem much more certain that Progressivism is religious in origin than they are which religion exactly it originates from. And the differences between Calvinism and Quakerism are *not* subtle.

Given their total lack of consensus on a matter as basic as which religion, why is it so important to Reactionaries that progressivism be descended from a religious background? Moldbug explains:

> [Progressives] believe their ideals are universal, that they can be derived from science and logic, that no reasonable and well-intentioned person can dispute them, and that their practice if applied correctly will lead to an ideal society. I believe that they are arbitrary, that they are inherited from Protestant Christianity, that they serve primarily as a justification for the rule of the cryptocalvinist establishment, or Polygon, and that they are a major cause of corruption, tyranny, poverty and war.

So the reason Reactionaries want the Left to be religious is to disprove the contention that it is based on reason. This would presumably discredit the Left and restore preeminence to Reactionary ideas such as that people should be ruled by a king, live in strong heterosexual nuclear families, avoid sexual promiscuity, and derive their values from fixed traditions rather than modern ideas of self-expression. You know, ideas with no religious background whatsoever.

3.1.1: Stop being snide and answer the question? Might Progressivism, far from deriving from some universal moral principles, actually be an arbitrary and parochial set of Calvinist customs and taboos?

The ideals commonly called progressive predate Calvin by several millennia. Consider the example of Rome. The early

Romans not only overthrew their kings in a popular revolution and instituted a Republic, but experienced five plebian secessions (read: giant nationwide strikes aiming at greater rights for the poor). After the first, the Roman government created the position of tribune, a representative for the nation's poor with significant power in the government. After the third, the government passed a sort of bill of rights guaranteeing the poor protection against arbitrary acts of government. After the fifth, the government passed the Lex Hortensiana, which said that plebians could hold a referendum among themselves and *the results would be binding on the entire populace, rich and poor alike*. By the later Empire, even slaves were guaranteed certain rights, including the right to file complaints against their masters.

The Empire was remarkably multicultural, even at its very highest levels. Emperor Septimus Severus was half-Libyan and some historians think his appearance might have passed for black in modern America. Emperor Maximinus Thrax was a Goth, Emperor Carausius was Gallic, and Emperor Philip the Arab was…well, take a wild guess. Although Rome did have a state religion, they were extremely supportive of the rights of minorities to continue practicing their own religions, and eventually just tried to absorb everything into a giant syncretistic mishmash that makes today's "ecumenialism" seem half-hearted in comparison. Although their tolerance famously did not always extend as far as Christianity, when the Romans had to denounce it they claimed it was not a religion but merely a "superstition" – a distinction which itself sounds suspiciously Progressive to modern ears. Indeed, the insistence of Christianity (and Judaism) on a single god, and their unwillingness to respect other religions as equally valid (in a very modern and relativistic way) was a large part of the Roman complaint against them.

The Romans pioneered the modern welfare state, famously memorialized by its detractors as *panem et circenses* – bread and circuses. Did you know welfare reform was a major concern of Julius Caesar? That ancient Rome probably had a higher percent of its population on the dole than modern New York? That the Romans *basically* worshipped a goddess of food stamps?

And no discussion of ancient Rome would be complete without mentioning their crazy sex lives. Wikipedia explains that "It was expected and socially acceptable for a freeborn Roman man to want sex with both female and male partners, as long as he took the penetrative role. The morality of the behavior depended on the social standing of the partner, not gender per se. Gender did not determine whether a sexual partner was acceptable, as long as a man's enjoyment did not encroach on another's man integrity." Gay weddings were not uncommon in ancient Rome, and were neither officially banned nor officially sanctioned. Juvenal and Martial both wrote satires condemning what they considered an epidemic of gay marriages during their era. And at least one Roman Emperor – Nero – married a man.

(well, married two men. One as groom and one as bride. And castrated one of them. And probably only married one of them because he was said to have an uncanny resemblance to Nero's mother. Whom Nero had previously had sex with, then murdered. I didn't say Nero was normal. Just unusually forward-thinking on the gay marriage issue.)

Moldbug listed the cryptocalvinist ie Progressive program as having four parts:

> Equality (the universal brotherhood of man), Peace (the futility of violence), Social Justice (the fair distribution of

> goods), and Community (the leadership of benevolent public servants)

Yet Equality has a clear antecedent in the plebian secessions of ancient Rome, peace in the Pax Romana, social justice in the Roman welfare system, and community in…well, it's so broadly defined here that it could be anything, but if we're going to make it the leadership of benevolent public servants, let's just throw in a reference to the philosopher-kings of Plato's *Republic* (yeah, fine, it's Greek. It still counts)

3.1.2: Yes, okay, the Romans tried to keep the peace and help the poor and stuff. That's a pretty weak definition of Progressivism. What really defines Progressivism is this messianic fervor that if we just do this *enough*, we can create a perfect utopia. *That* is what these ancient cultures were lacking.

Even if you've never read *The Republic*, you can still get a sense of the utopian striving in the classical world from reading some of the stuff written during the reign of Emperor Augustus. Here's Dryden's translation of a passage from the *Aeneid*:

> An age is ripening in revolving fate
> When Troy shall overturn the Grecian state…
> Then dire debate and impious war shall cease,
> And the stern age be soften'd into peace:
> Then banish'd Faith shall once again return,
> And Vestal fires in hallow'd temples burn;
> And Remus with Quirinus shall sustain
> The righteous laws, and fraud and force restrain.
> Janus himself before his fane shall wait,
> And keep the dreadful issues of his gate,
> With bolts and iron bars: within remains
> Imprison'd Fury, bound in brazen chains;

> High on a trophy rais’d, of useless arms,
> He sits, and threats the world with vain alarms.

So please, tell me again how utopian desires for peace and social justice were invented wholesale by John Calvin in 1550.

3.2: Is the move toward Progressive social policy masterminded by “the Cathedral”?

Reactionaries have to walk a fine line. They can’t just say “people consider liberal policies, decide they would be helpful, and form grassroots movements pushing for the policies they support”, because that would make leftist policies sound like reasonable ideas pursued by decent people for normal human motives.

But they can’t just say “There’s a giant conspiracy where the heads of all the major Ivy League universities meet at midnight under the full moon”, because that would sound ridiculous and tinfoilish.

So they invent this strange creature, the *distributed conspiracy*. It’s not just people being convinced of something and then supporting it, it’s them *conspiring to do so*. Not the sort of conspiring where they talk to one another about it or coordinate. *But still a conspiracy!* Michael Anissimov describes it like so:

> [The Cathedral is] the self-organizing consensus of Progressives and Progressive ideology represented by the universities, the media, and the civil service…the Cathedral has no central administrator, but represents a consensus acting as a coherent group that condemns other ideologies as evil […]

> Government and social policy is manufactured in universities, first and foremost at Harvard, followed by Princeton, then Yale, then the other Ivies, Berkeley, and Stanford. As far as politics is concerned, institutions outside of these are pretty much insignificant. Memetic propagation is one-way — it is formulated in the schools and pumped outwards. The universities are not significantly influenced by the outside. The civil servants that make government decisions are either borrowed from universities or almost totally influenced by them. The official mouthpiece of this ideological group is The New York Times, which is the most influential publication in the world outside of the Bible.

So now that we have this formulation of the problem, we can ask some more specific questions.

3.2.1: Are Harvard and the New York Times disproportionately linked to the Progressive ideas that now dominate society?

That depends partly on what “disproportionately” means, of course. But we can make some vague and qualitative observations.

The Roman and Persian Empires held some very Progressive ideals, all without the help of any universities or newspapers whatsoever. Parsimony suggests that whatever process pushed Rome to the left could be doing the same to the modern world.

But a better counterexample might be noting that even *modern* progressivism predates this institutions. The history of modern Progressivism – even as told by Reactionaries – goes from John Locke to the Glorious Revolution to the American Revolution to the French Revolution to the US Civil

War on through John Stuart Mill to the New Deal and the United Nations and civil rights movements and on to the present. While Harvard (est. 1636) does predate all those events, I don't think even its most fervent critic would accord it any level of influence on world ideas until the 1850s at the earliest. And the Times was founded in 1851. It is hard to chart the precise progress of Progressivism, but I don't notice any sharp discontinuity at any point. Once again using parsimony, we might expect the forces that promoted Progressivism during the French Revolution and before to be the same forces promoting Progressivism afterwards. This takes any special role of Harvard or the New York Times entirely out of the pictures.

And modern progressivism doesn't seem linked to Harvard or the Times in *space* either. New York and Boston are pretty progressive – by American standards. But there's a whole world out there. Canada is further left than America; Britain is further left than Canada; France is further left than Britain; the Netherlands are further left than France; and Sweden is further left than the Netherlands. Russia and China are complicated, but they've certainly had their super-leftist periods. In fact, pretty much the entire developed world is further left than anywhere in the United States, New York and Boston not excepted. This does not seem an entirely recent development; for example, the Netherlands' liberalism has clear roots in the Dutch Golden Age of the 1600s.

It is true that sometimes a prophet is without honor in his own country. Yet for an American college and a newspaper read almost uniquely by Americans to have affected every other country in the Western world more effectively than they were able to affect the United States seems, well – unexpected.

3.2.2: Do Harvard and the New York Times invent Progressive dogma and then shove it down the throats of a hostile country?

Gay rights will be an interesting test here, because it’s one of the issues on which society has shifted leftward most quickly and dramatically, and because it’s relatively recent so its history should be easy to trace.

Modern gay rights movements trace their history to Germany, a country not known for having Harvard or the New York Times, or for that matter Puritans and Quakers. The German movement included such pioneering activists as Magnus Hirschfeld and Max Spohr, but Germany kind of dropped the ball on gay rights with the whole Nazi thing, and the emphasis shifted to elsewhere in Europe. In America, the movement finally gained steam in the 1960s with a picketing in Philadelphia and a community center in San Francisco, and finally the Stonewall Riots in New York.

I can’t get any good information about Harvard’s position, but the New York Times helpfully has an online archive of every article they have ever published. So what, exactly, was America’s Newspaper Of Record doing while all this was going on? It was helpfully publishing articles like GROWTH OF OVERT HOMOSEXUALITY IN CITY PROVOKES WIDE CONCERN:

> The problem of homosexuality in New York became the focus yesterday of increased attention by the State Liquor Authority and the Police Department…The city’s most sensitive open secret – the presence of what is probably the greatest homosexual population in the world and its increasing openness – has become the subject of growing concern of psychiatrists, religious leaders, and the police.

> Sexual inverts have colonized three areas of the city. The city’s homosexual community acts as a lodestar, attracting others from all over the country. More than a thousand inverts are arrested here annually for public misdeeds. Yet the old idea, assiduously propagated by homosexuals, that homosexuality is an inborn, incurable disease, has been exploded by modern psychiatry, in the opinion of many experts. It can be both prevented and cured, these experts say.
>
> The overt homosexual – and those who are identifiable probably represent no more than half of the total – has become such an obtrusive part of the New York scene that the phenomenon needs public discussion, in the opinion of a number of legal and medical experts. Two conflict viewpoints converge today to overcome the silence and promote public discussion.
>
> The first is the organized homophile movement – a minority of militant homosexuals that is openly agitating for removal of legal, social, and cultural discriminations against sexual inverts. Fundamental to this aim is the concept that homosexuality is an incurable, congenital disorder (this is disputed by the bulk of scientific evidence) and that homosexuals should be treated by an increasingly tolerant society as just another minority. This view is challenged by a second group, the analytical psychiatrists, who advocate an end to what it calls a head-in-sand approach to homosexuality…

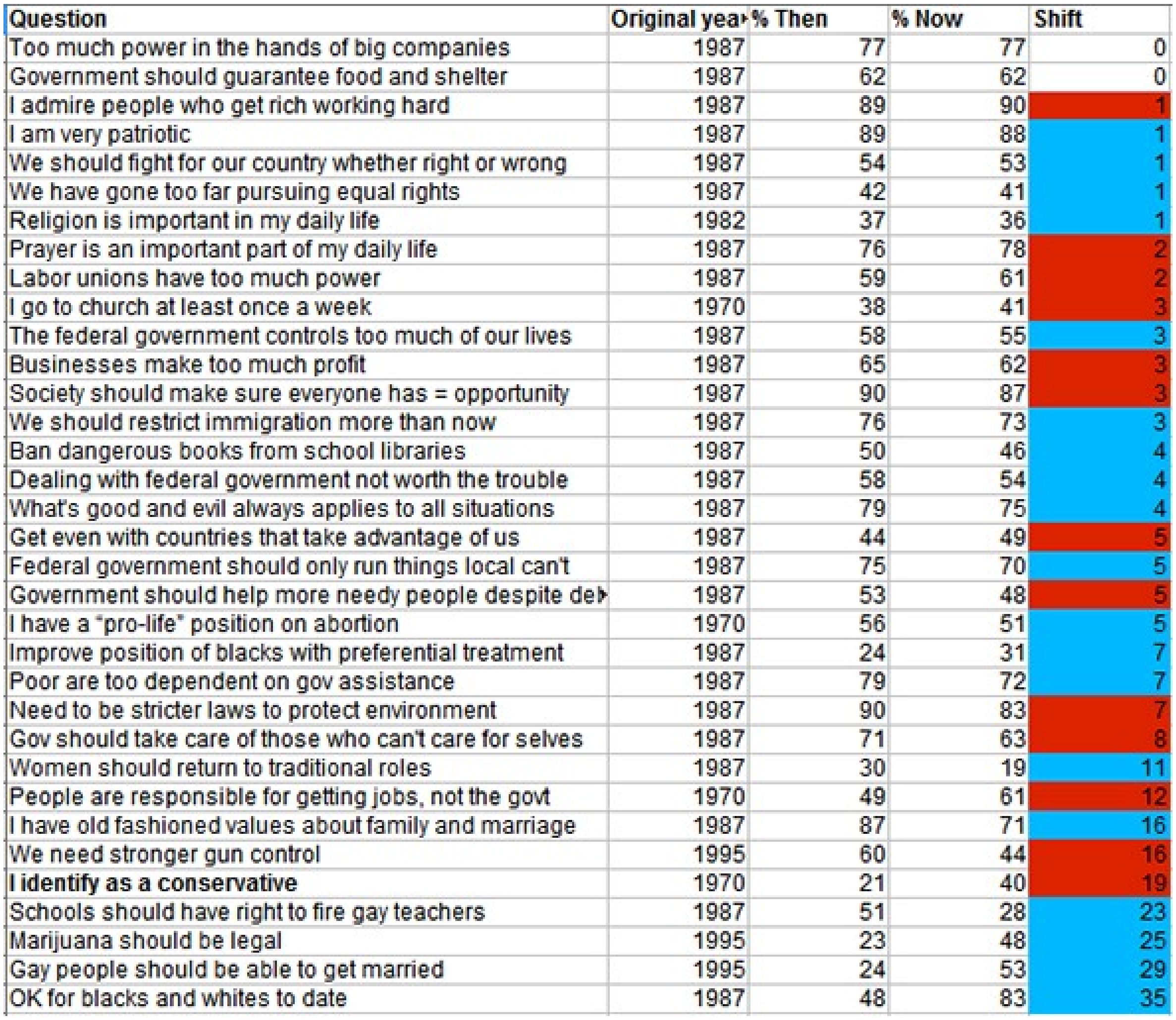
On and on and on it goes in this vein. And that’s not even counting other such wonderful New York Times articles as WOMEN DEVIATES HELD INCREASING – PROBLEM OF HOMOSEXUALITY FOUND LARGELY IGNORED. These aren’t editorials – this is the headlines, the supposedly fact-based objective reporting section. The editorials are worse – I particularly like the one warning that we need to fight increasing gay influence in the theater industry because gays cannot authentically write plays about love or relationships.

Now, to the Times' credit, it eventually changed its tune and is now mostly in favor of gay rights. That's fine for the Times but not so good for Reactionaries. The story here is very clearly of a gay rights movement that began as a grassroots push in favor of more tolerance. The New York Times opposed it, but *somehow* the movement managed to gather steam despite that crushing blow. Eventually its tenets became accepted by more and more people, and one of these late adapters was the New York Times, which now atones for its sin by defending gay rights against even *later* adapters.

This is not the pattern one would expect if all Progressive ideas were fueled solely by the New York Times' backing.

3.2.3: Do Harvard and the New York Times successfully impose their values on the rest of America and the world?

Let's examine exactly how opinions have changed on a host of important political issues. These are taken from the National Election Survey, Pew Research, and Gallup. I've tried to avoid cherry-picking – I took every issue I could find, starting from the first year data was available. In cases where I could find two different polls, I kept the one with a longer data series:



Question	Original yea	% Then	% Now	Shift
Too much power in the hands of big companies	1987	77	77	0
Government should guarantee food and shelter	1987	62	62	0
I admire people who get rich working hard	1987	89	90	1
I am very patriotic	1987	89	88	1
We should fight for our country whether right or wrong	1987	54	53	1
We have gone too far pursuing equal rights	1987	42	41	1
Religion is important in my daily life	1982	37	36	1
Prayer is an important part of my daily life	1987	76	78	2
Labor unions have too much power	1987	59	61	2
I go to church at least once a week	1970	38	41	3
The federal government controls too much of our lives	1987	58	55	3
Businesses make too much profit	1987	65	62	3
Society should make sure everyone has = opportunity	1987	90	87	3
We should restrict immigration more than now	1987	76	73	3
Ban dangerous books from school libraries	1987	50	46	4
Dealing with federal government not worth the trouble	1987	58	54	4
What's good and evil always applies to all situations	1987	79	75	4
Get even with countries that take advantage of us	1987	44	49	5
Federal government should only run things local can't	1987	75	70	5
Government should help more needy people despite de	1987	53	48	5
I have a "pro-life" position on abortion	1970	56	51	5
Improve position of blacks with preferential treatment	1987	24	31	7
Poor are too dependent on gov assistance	1987	79	72	7
Need to be stricter laws to protect environment	1987	90	83	7
Gov should take care of those who can't care for selves	1987	71	63	8
Women should return to traditional roles	1987	30	19	11
People are responsible for getting jobs, not the govt	1970	49	61	12
I have old fashioned values about family and marriage	1987	87	71	16
We need stronger gun control	1995	60	44	16
I identify as a conservative	1970	21	40	19
Schools should have right to fire gay teachers	1987	51	28	23
Marijuana should be legal	1995	23	48	25
Gay people should be able to get married	1995	24	53	29
OK for blacks and whites to date	1987	48	83	35

Of thirty-four issues that made the cut, opinion shifted to the left on 19 and to the right on 13. There was an average shift of three points leftward per issue. Contrary to Reactionary claims that Americans do not appreciate the extent of the leftward shift affecting the country, in a recent survey based on a similar chart, most people, regardless of political affiliation, slightly overestimated the extent to which values had shifted leftward over the past generation.

Not only is the leftward shift less than people intuitively expect, it does not affect all issues equally. The left's real advantage is limited to issues involving women and minorities. Remove these, and opinion shifts to the left on

11 issues and to the right on 12. The average shift is one point rightward per issue.

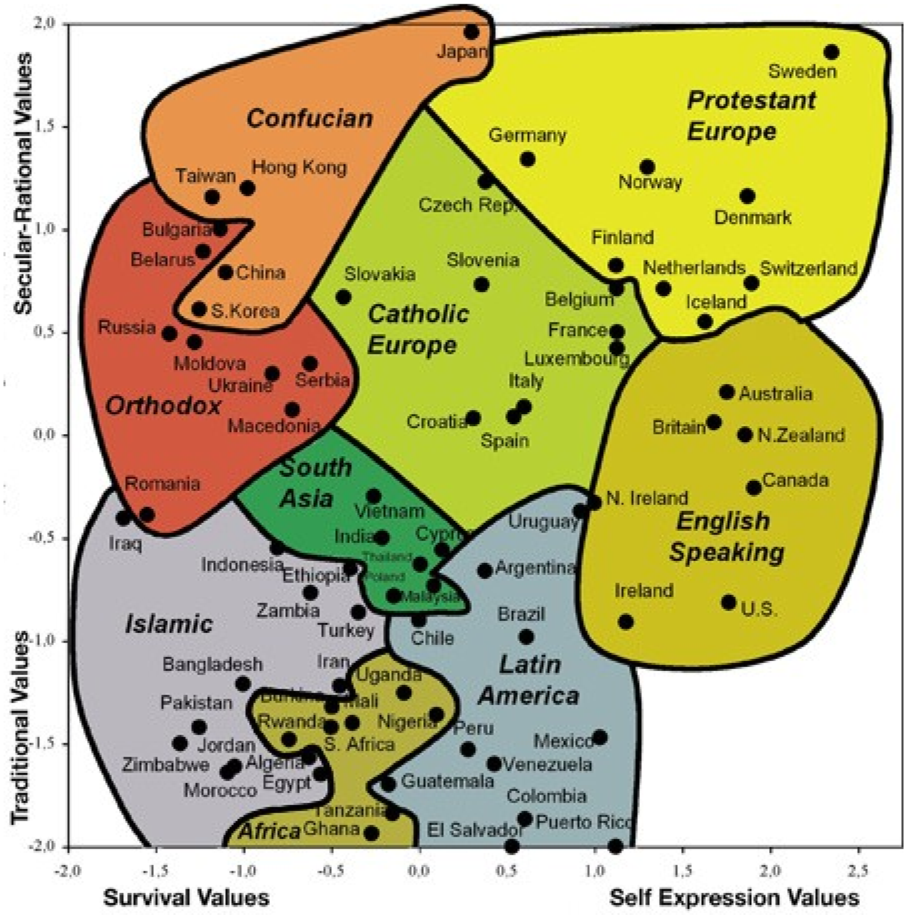
On the hottest, most politically relevant topics, society has moved leftward either very slowly or not at all. Over the past generation, it has moved to the right on gun control, the welfare state, capitalism, labor unions, and the environment. Although the particular time series on the chart does not reflect this, support for abortion has stabilized and may be dropping. This corresponds well with the DW-NOMINATE data that finds a general rightward trend in Congress over the same period. The nation seems to be shifting leftward socially but rightward politically – if that makes any sense.

If the Left had seized control of the government, or the media, or the institutions of the country, we would expect it to do a better job pushing its cherished policies like abortion rights, gun control, environmental protection, et cetera. Instead, beliefs on those issues have remained stable or shifted rightward, while issues like marijuana legalization – an issue more libertarian than progressive, and with minimal support from leftist institutions – succeed wildly. Whatever advantage the left has, it must be something skew to politics, something that institutionalized leftism, from the Democratic Party down to the Humanities Department at Harvard, can neither predict nor control.

3.3: Then where *does* progress come from?

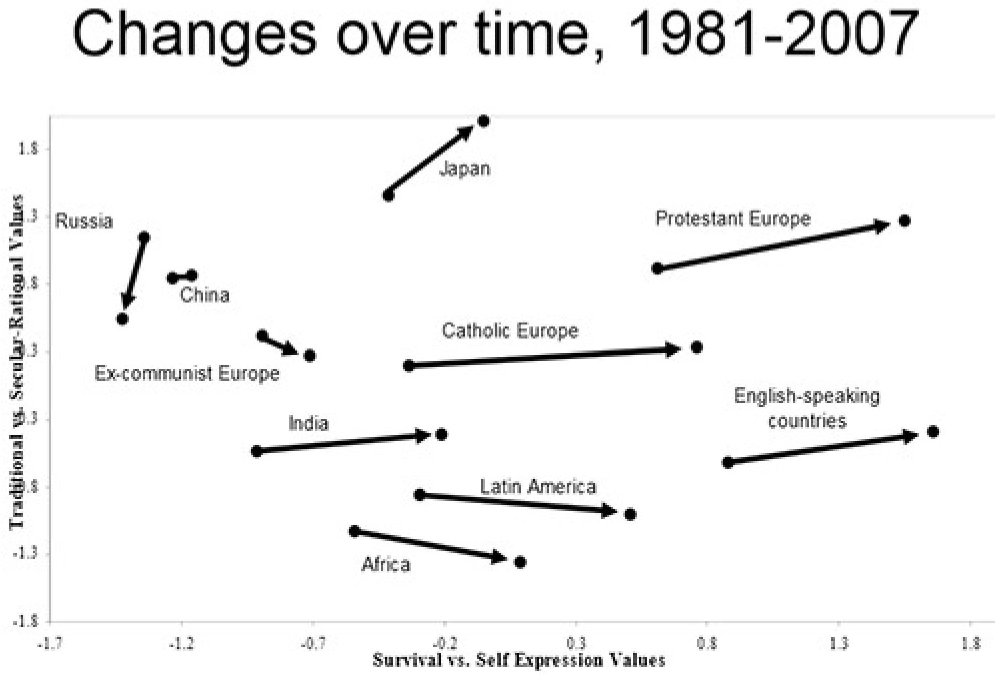
So the cultural shift of the past few centuries isn't toward some weird Christian sect. And it wasn't caused by Harvard or the New York Times. What was it and who did it?

The World Values Survey is the official academic attempt to understand this question. They've been polling in eighty countries around the world for thirty years trying to figure out who has what values and how they have been changing. Maybe you’ve seen the most famous summary of their results:

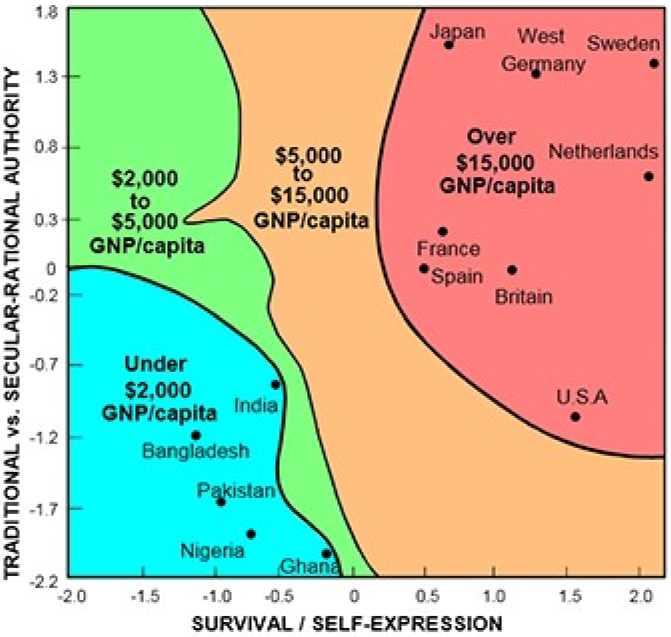


There is no end to the fun one can have with WVS data, and I highly recommend at least Wikipedia’s Catalogue of Findings if not the original studies. But the most important part is that dimensionality analysis finds that answers to value questions cluster together onto two axes: survival vs. self-expression values, and traditional vs. secular-rational values.

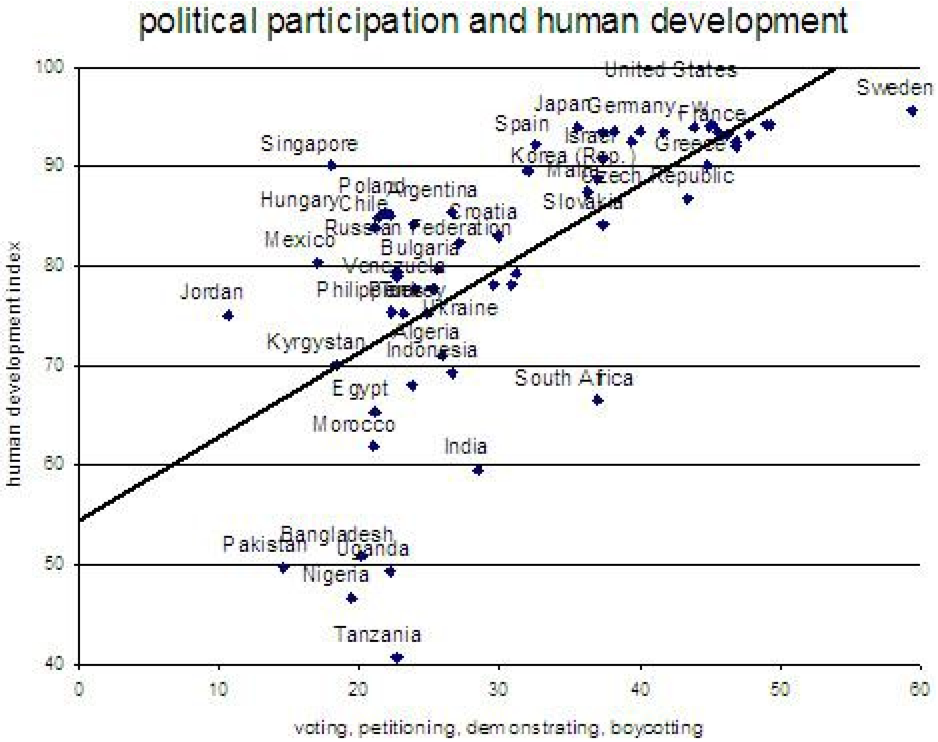
Over time, societies tend to move from traditional and survival values to secular-rational and self-expression values. This is the more rigorous version of the “leftward shift” discussed above.



Both within a single time period and between time periods, traditional and survival values are generally associated with poverty, low industrialization, and insecurity. Secular-rational and self-expression values are generally associated with wealth, industrial or knowledge economies, and high security. The difference is not subtle:

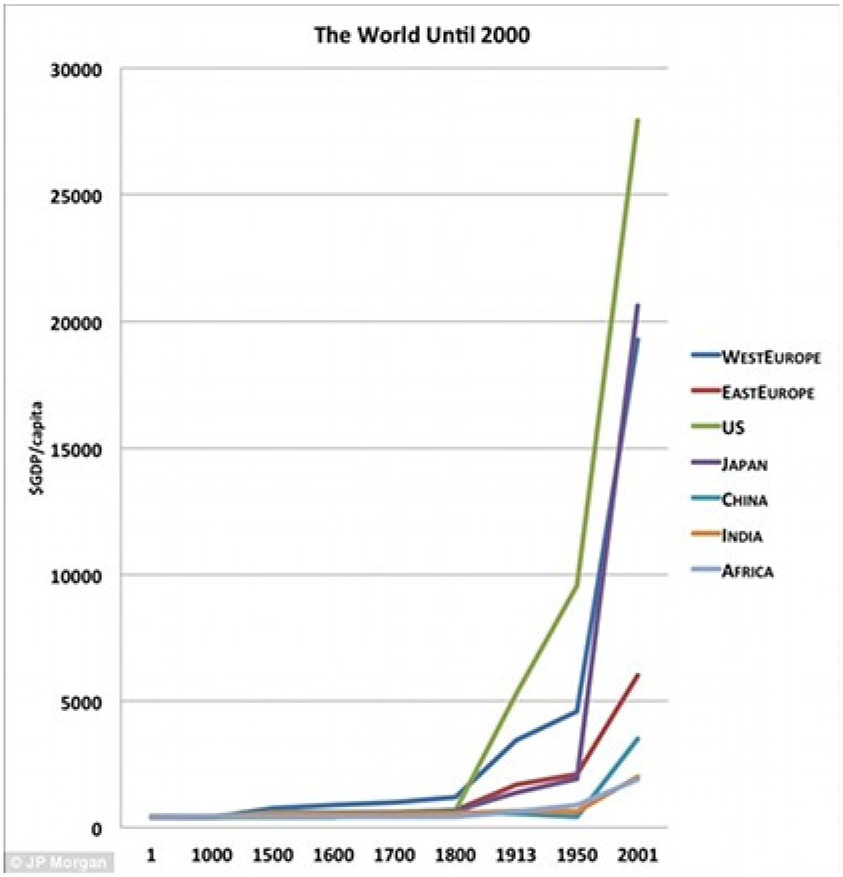


And if you want to know why countries are becoming more democratic and less monarchist, it’s hard to get a more direct answer than this graph (although its attempt at a linear fit was a bad idea):



All of this provides a simple and elegant explanation of the distribution of leftism, both in time and space. The most progressive countries today tend to be very wealthy, very peaceful, and comparatively urbanized. The least progressive countries tend to be poor, insecure, and comparatively rural.

Remember Michael Anissimov’s description of the leftward shift above? That the world has been growing further to the left ever since the French Revolution? Take a look at the course of the world economy:



Riiiight about the time of the French Revolution – which also happens to be around the time of the Industrial Revolution – the world economy suddenly shifts into hyperdrive, starting in the USA and Western Europe, spreading to Japan after

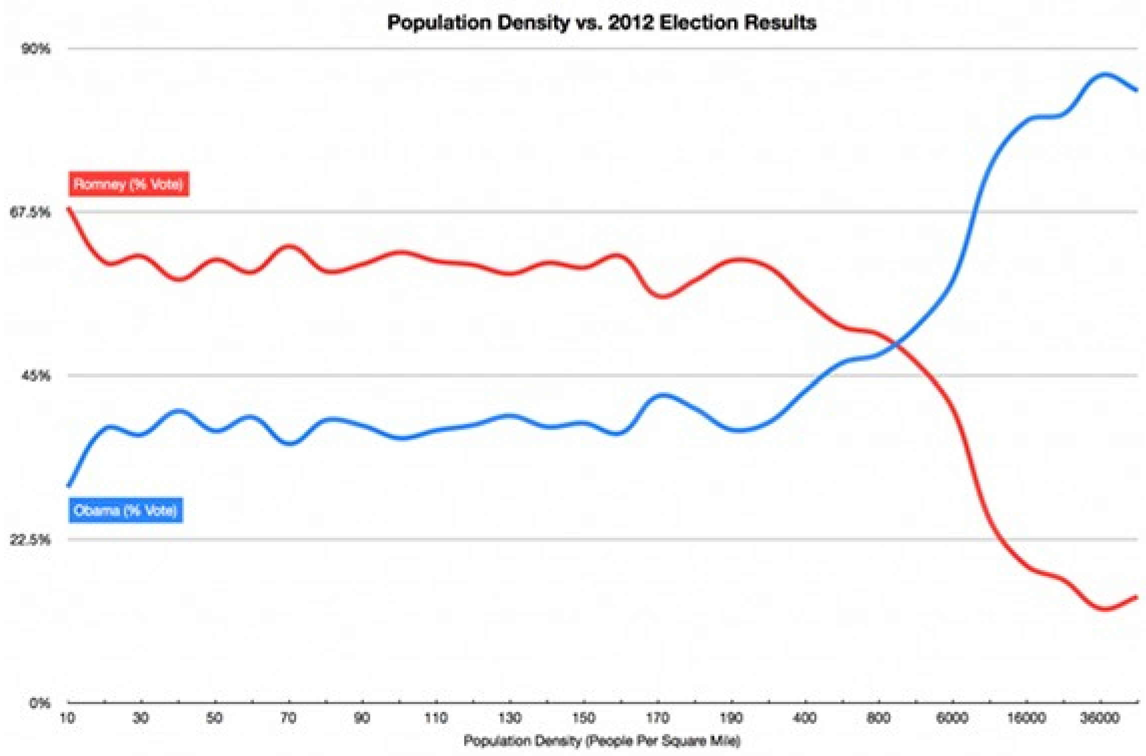
World War II, and not quite yet having reached Africa or Southeast Asia.

And, well, right about the time of the French Revolution Europe and the USA started shifting to the left, with Japan following after World War II, and Africa and Southeast Asia still lagging behind.

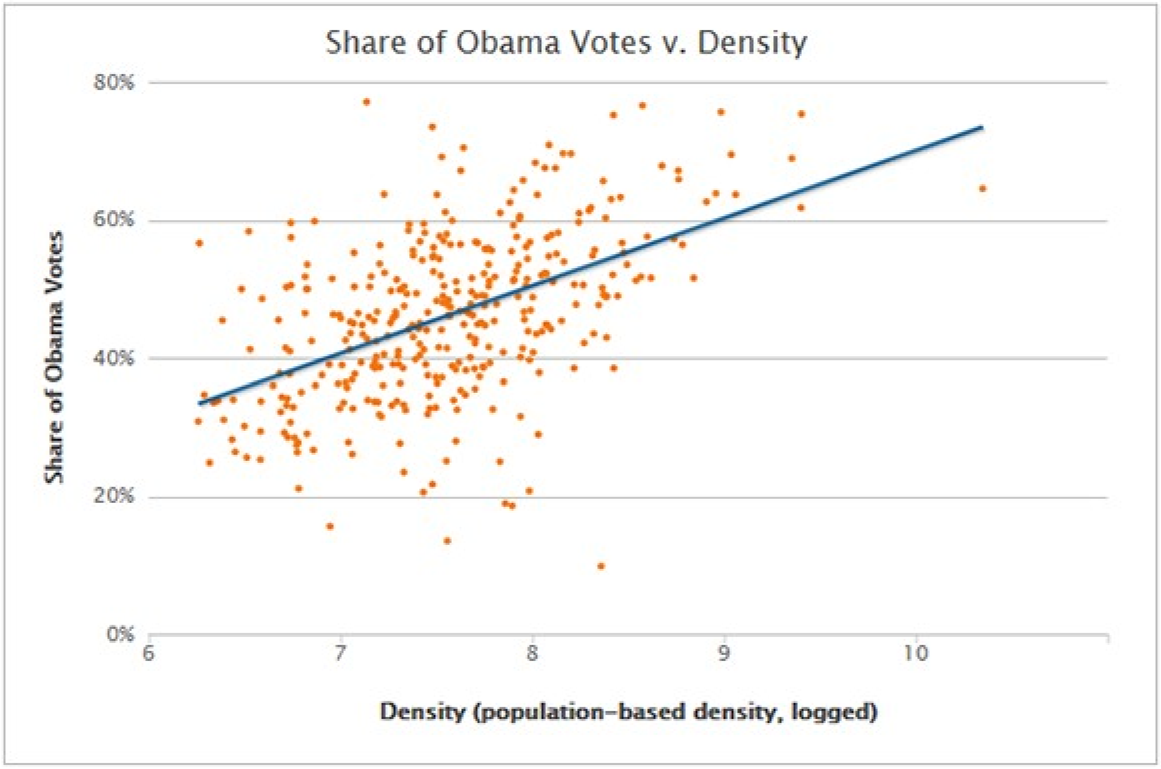
This progressivism/economics link is so obvious that anyone who thinks about it for a few minutes can reach the same conclusion. I wrote “A Thrive/Surive Theory Of The Political Spectrum long before I was familiar with the World Values Survey, but its conclusions match the survey’s in pretty much every respect: rightist values are those most suited for hardscrabble existence where everyone must band together to survive a dangerous frontier; leftist values are those most suited for a secure postscarcity or near postscarcity existence with surplus resources available to devote to more abstract principles.

I’d like to examine one more aspect of this before I stop beating this dead horse, which is the rural/urban divide. The history of industrialization is in many ways the history of urbanization, and the distinction between insecure frontier life and secure postscarcity life mirrors the rural/urban divide. This predicts that more rural countries should be more traditional/survival and more urban countries more secular-rational/self-expression, which in fact we see. Of the countries furthest to the top-right on the WVS diagram, Sweden, Norway and Denmark all have about 85% urban populations. Go down to the three countries at the bottom left – Jordan, Morocco, and Zimbabwe – and despite Jordan’s anomalously high level they’re still averaging about 55%.

This is true even in the United States – the denser a county, city, or state, the more likely it is to lean Democratic, as we can see from this terrible and confusing graph:

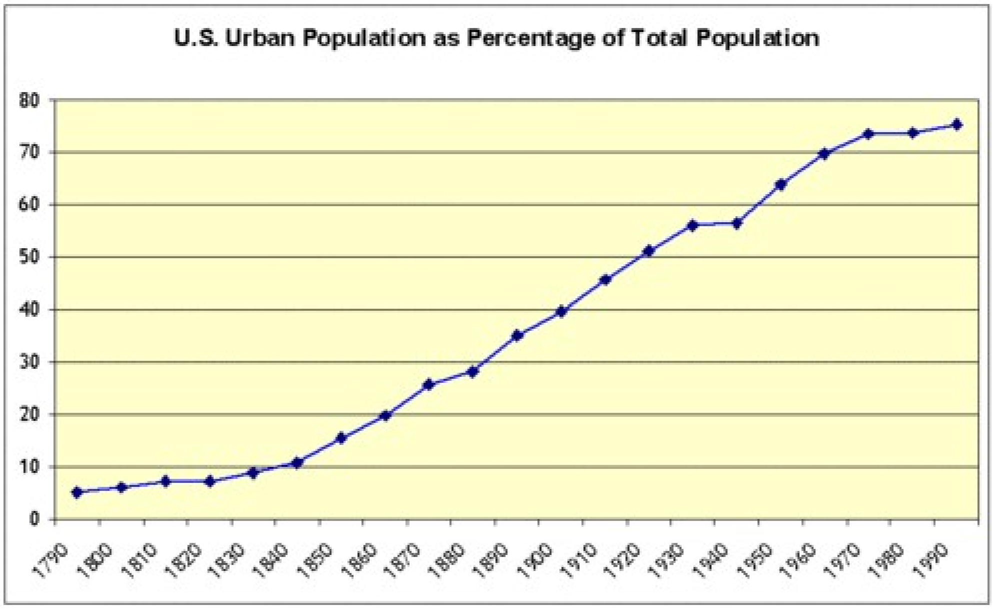


Rural counties – those with <200 people per square mile - lean red at about 65%. Once they pass that 200 person mark, they very quickly start leaning blue until the densest areas - true cities, approach 90% Democratic. Or as Dave Troy notes, "98% of the 50 most dense counties voted Obama. 98% of the 50 least dense counties voted for Romney." This density effect applies even within cities. Here are America's largest cities graphed by density against percent Romney vote:



My sources point out that “graphs of the UK, Australia, and Canada look very similar during the same period, with left voting concentrated in urban and mining districts” and theyalso mention (just to fend off the inevitable reactionary critique) that “interestingly — and contrary to the much-stated view that Obama purchased the election with welfare, food stamps, and other entitlements, our analysis turned up no statistically significant association between Obama votes and the metro poverty rate and only a very small one for income inequality across metros.”

Why am I making such a big deal of this? Well, here’s America’s percent urban versus percent rural population over the period of time when our values were shifting to the left:



So please. Tell me again how the leftward value shift over the past two hundred years was caused entirely by a sinister conspiracy of Ivy League college professors

3.3.1: Can you give a more detailed explanation of why increasing wealth, technology, and urbanization would lead to the values we call Progressive?

Here are five specific examples.

Multiculturalism is a forced adaptation to the culturally unprecedented situation of large groups of people from different cultures being forced to live and work together. This situation arises because of technology and urbanization. Technology, because more Somalis are going to immigrate to the US when that means booking a plane ticket over the phone than when it meant a six month journey over stormy seas. Urbanization, because it’s much harder to immigrate into an agrarian society where every family knows each other and farmland is at a premium than into an urban society where you can apply for the same factory job as everyone else.

Modern gender roles are a forced adaptation to the existence of cheap and effective contraception, which decouples sex from pregnancy. Teen pregnancy is relegated to people unwilling or unable to use contraception, allowing other women to pursue the same careers as men rather than dropping out of the workforce to become full-time mothers.

The welfare state is a forced adaptation to mobile and urban societies. In agrarian societies, most people owned their own means of production – their farms – and “unemployment” wasn’t a salient concept. It was usually possible to get what you needed through the sweat of your brow, even if that meant chopping down trees to build a log cabin, and there was little sympathy for people who didn’t bother. In urban societies, people need jobs in order to support themselves, and those who cannot get them starve in full pitiful view of everyone else.

Socialized health care is a very big part of the welfare state – probably the majority depending on how you parse the numbers. As recently as a century ago there really wasn’t much in the way of health care technology for people to spend money on, and most people died quickly and simply without having to be kept alive in expensive hospitals for months. As health care gets beyond most people’s ability to afford, and the average lifespan lengthens, there becomes more demand for government to step in and fill the gap.

Secularism is a more viable intellectual option once Science has discovered things like evolution and the Big Bang. Just as “there are no atheists in foxholes”, people with a comfortable urban existence not dependent on the whims of the weather and the plague are less likely to worry about placating the Lord. Multiculturalism means that faiths no are no longer immune to challenge, as Christians and Muslims and Buddhists have to live next to each other and notice how totally unconvinced outsiders are of their ideas. And the movement from closely-knit communities to sprawling cities mean that the local church is no longer ties together your entire actual and possible social network so closely that it can exert pressure on you to conform.

And yes these are just-so stories, but the relationship between all these factors and wealth/urbanization are pretty much beyond dispute – so if it’s not true for these reasons it’s true for reasons no doubt very much like them.

3.4: Do you believe in “Whig history”?

Whig history is an approach to historical study that emphasizes how the past has been groping towards the truths and institutions of the present. It is usually used derisively, in a sense of “Oh, so you think the era in which *you* were born just happens to be perfect, and everyone else from Aristotle to Galileo was just failing at being an American of 2013.”

There is obviously a strong meaning of the term which cannot help but be false. The past did not share our values, it did not move linearly, and the present moment is neither perfect nor universally superior to other periods.

On the other hand, in a world where progress in areas as diverse as cars, computers, weapons and health care has been blindingly obvious, we shouldn’t place too low a prior on the possibility that there has been progress in social institutions as well. Such progress could be motivated by the same factors that advance other areas.

First, a greater store of empirical results. As time goes on, we have more virtuous examples and terrible warnings. No one pushes for prohibition of alcohol anymore because we’ve seen how that turns out – and in thirty years, people may say the same about other drugs. Very few people push full hold-a-revolution Communism anymore, and for the same reason.

Second, better data. With the invention of statistics and information technology, we now have numbers on everything from income inequality to how different types of policing affect the crime rate. Members of the civil service, politicians, lobbyists, and even voters use these numbers to decide what policies to support. Neither the data nor its interpretation is always unbiased, but it’s a heck of a lot better than the old method of doing whatever your prejudices tell you to do.

Third, social evolution. This is a complicated one, because all evolution is evolution to a niche, the niche is different in the modern world than in the medieval world, and so modern and medieval societies are optimizing for different things. But at the very least, we can say that modern institutions are better adapted to the modern niche than medieval institutions. Those governments that did not adapt were overthrown; those corporations that did not adapt went out of business; those institutions that did not adapt became unpopular and saw their influence shifted to other institutions. Those governments, corporations, and institutions that did adapt prospered and spun off copycats with small variations, and the evolutionary cycle repeated again.

To these three we could add things like greater education, better access to information, and more rational values (you can no longer get away with saying “Follow me because I’m the Messiah”, and that’s probably a good thing). So although it’s not some a priori law of nature that the modern period must be the best period in history, we do have some reasons to expect things to be getting better rather than worse. As Part I pointed out, those expectations have mostly been realized.

3.5: Is America a communist country?

Reactionaries tend to push this line by finding the platform of the US Communist Party from some year well in the past, then pointing out that a lot of their goals were achieved, then noting that since America did what the communists wanted, we are a communist country.

Moldbug and others have claimed it, it even has its own Facebook page, but Free Northerner has done by far the most complete job analyzing it and finds that of demands in the 1928 Communist Party platform, 70% of all demands, and 78% of domestic demands, have been met as of 2013.

I don't want to belittle Free Northerner's work – he did a great job, he was much more rigorous than I'm about to be, and anyone who writes a blog post on how awesome Turisas is is a friend of mine regardless of his political beliefs.

But although I can't get my computer to load the platform directly, I notice when I check his transcription that the Communist demands mysteriously lack points like “workers control the means of production” or “all property held in common”, or even “not capitalism”. They do, on the other hand, include policies like “abolition of censorship”, “right to vote for everyone over 18”, and “paid maternity leave during pregnancy”.

Rather than conclude that America is a communist country, a better conclusion might be “the Communist Party of 1928 wasn't especially “communist”, in the sense that we use that word today.” That's no surprise. The meaning of words changes over time, and the Cold War made the more moderate elements of communism drop the “communist” label. Using a liberal definition of “communist” to claim that we satisfy the definition, then suggesting we should draw the conclusions and connotations we would from the strict definition of “communist” remains the worst argument in the world. Take out the Worst Argument In The World, and all the Communist Party platform experiment proves is that we support policies like “no child labor” and “free maternity leave” – ie things we already knew.

There’s a second counterargument, though, which is more interesting. Free Northerner writes:

> I don’t have time to analyze the Democratic and Republican platform demands of the same year at this time, but I would bet significant sums that less than 80% of their demands were met and upheld by our present time.

I’ll take that bet!

I mistakenly got the Republican platform for 1920 (someone else can double-check 1928 specifically). The Republicans failed to conveniently list their demands in bullet-point format, but from their long manifesto I managed to extract 37 different points:

1. Give farms right to cooperative associations
2. Protection against discrimination for farmers
3. End to unnecessary price fixing that reduces prices of farm products
4. Facilitate acquisition of farmland
5. Reduce frequency of strikes
6. Good voluntary mediation for industry
7. Convict labor products out of interstate commerce
8. Reorganize federal government

9. Simplify income tax
10. Federal Reserve free from political influence
11. Fair hours and good working conditions for railway workers
12. Private ownership of railroads
13. Immediate resumption of trade relations with all nations at peace
14. Restrict Asian immigrants
15. No one becomes citizen until they have taken a test to ensure they are American
16. American women do not lose citizenship by marrying an alien
17. Free speech, but no one can advocate violent overthrow of the government
18. Aliens cannot speak out against government
19. End lynching
20. Money for construction of highways
21. Save national forests and promote conservation
22. Reclaim lands
23. Increase pay of postal employees
24. Full women's suffrage in all states
25. Federal gov should aid states in vocational training
26. Physical education in schools
27. Centralize gov public health functions
28. End child labor
29. Equal pay for women
30. Limit hours of employment for women
31. Encourage homeownership for Americans
32. Make available information of housing and town planning
33. Americanize Hawaii
34. Home rule for Hawaii
35. Join international governing body such as League of Nations
36. No mandate for Armenia
37. Responsible government in Mexico

Not being too familiar with the 1920 political milieu, I don't really know what they mean by 2, 22, 32. Others seem so broad as to be hard to judge: 4, 6, 8, and 37. That leaves 29 points.

I think the Republicans have achieved 1, 3, 5, 7, 10, 11, 15, 16, 17, 19, 20, 21, 23, 24, 25, 26, 28, 29, 33, 34, 35, and 36 – some unambiguously, others if nothing else by the very sketchy criteria Free Northerner used to rule in commie achievements. They have definitely failed 9, 12, 13, 14, and 30. As for 18, 27, and 31, these seem ambiguous – let's count them half a point. That means they got 23.5/29 of the points they wanted – 81%. That's better than the Commies, who only got 70%.

(if we were really trying to do this right, we'd want to have the person who evaluated the success or failure of a party plank blinded to which party it came from. I'll leave someone else to try this).

So apparently the US is a Republican country even more than it's a Communist country. I bet if we looked over the Democratic platform for the same time frame, we'd find it was a Republican, Democratic, *and* Communist country. And if we check the Nazi Party platform, we find that some of the same points Free Northerner counts as Communist victories – abolition of child labor, expansion of old age welfare – are also Nazi Party policies at the same time. So we are, in fact, a Democratic-Republican-Commie-Nazi country.

The alternative is that all parties liked to promise they would throw money at popular feel-good projects. Shorter working hours! Better welfare! Freedom of this! Freedom of that! As the country became richer it was able to support more feel-good policies, and so every party got much of what they wanted.

4: Could a country be ruled as a joint-stock corporation?

This is the plan of Mencius Moldbug, who gets points for being clever and creative rather than trying to rehash 13th century feudalism. I’ve heard different rumors as to whether he still supports it and whether it might or might not be a cover for supporting 13th century feudalism. Nevertheless the idea is interesting and deserves further investigation. However, it is missing some key details and suffers some probably irresolvable conceptual problems.

4.1: Would a joint-stock corporation prevent government decisions based on political tribalism and sacred values, in favor of government decisions based on maximizing economical value?

According to the theory, just as modern corporations like GE successfully remain dedicated to profitability, so America could be sold off in an IPO and restructured as a corporation dedicated to maximizing the value of US land.

But just calling something a corporation doesn’t make it start worrying about profitability. Making its shareholders worry about profitability turns out to be surprisingly hard problem, even though these shareholders themselves would benefit from its profits.

We can imagine two different distributions of shares: either everyone gets one, or only a few aristocrats get one (the degenerate third possibility, where only one person gets them, isn’t really a “joint-stock company”).

The first possibility might be suspected of being democracy: after all, every citizen equally has one share and therefore one vote. Moldbug argues it wouldn’t be: shares are transferable, and citizens have an incentive to maximize the value of their share.

So chew on this: suppose that banning abortion would earn the American government $10 billion dollars a year (how? I don’t know. Let’s just say it does). This corresponds to about $30 for every American.

How many leftists do you think would vote to ban abortion for $30?

What if their $30 was entirely illiquid, only accessible by the one-time event of selling their single share of stock, and would probably be so lost in noise that they would never see tangible evidence of it?

Okay, what if they don’t even *know* it will give them $30? No doubt Planned Parenthood will author a very scholarly report giving excellent reasons why an abortion ban will make stock shares plummet, and the Catholic Church will author an equally scholarly report giving excellent reasons why it will make everyone rich. Which side will people believe? Why, whichever side matches their natural prejudices, of course! As well ask a Democrat or a Republican whether Obamacare will increase or decrease the deficit.

The only thing that giving everyone a share of American stock would do to politics in the US is allow both the Left and the Right a chance to accuse one another of being secretly in it for the money, while both continue to do what they did before. Perhaps this wouldn’t happen in a country created *de novo* out of thin air, but US politics are far too entrenched for giving people little stock certificates to help anything.

Anyway, it would take about ten minutes for poor people to sell their shares for easy cash. So this case would immediately degenerate to the second possibility – one where only a small “ruling class” owns all the stock certificates. I think a few Reactionaries have proposed this, and then they can be “nobles”, and make up an “aristocracy”, and…

Hold your horses. Suppose a new ruling class of ten thousand people possess all these certificates.

By definition all of these people will be multibillionaires – once you own one ten thousandth of America, you’ve got it made. And we observe something interesting with multibillionaires – Bill Gates, Warren Buffett, Larry Page. *They find other things much more interesting than money.* Bill Gates is working on curing malaria. Warren Buffett is trying to give all his money away to charity. Larry Page is working on fascinating but bizarre projects with minimal chance of success during his lifetime. Once you’re a multibillionaire, you need more money less than you need to feel like you’re making some kind of wonderful contribution to the world that will make coming generations revere you.

In other words, these shareholders won’t care about the monetary value of their shares either. Take people like Ted Turner or the Koch brothers, give them a big chunk of the US government, and you expect them to focus on its *monetary value* just because you’re calling it a stock?

4.2 Would corporate governance at least have lower discount rates?

Likely no.

Do corporations today have low discount rates? Consider the example of Lehman Brothers and other pre-crash investment banks. They happily accepted (and invented) subprime loans that would raise their profits today at the cost of likely financial disaster tomorrow.

More broadly, reflect upon how few companies pursue long-term revolutionary technology. Even though nearly everyone agrees that the future will be less based on fossil fuels, research and development of the likely replacements – from fusion power to solar power to electric cars – is either run by the government or grudgingly performed by corporations only after being promised huge government subsidies. When companies do develop exciting new technologies of their own accord – Google's Calico, SpaceX's rockets – they tend to be associated with some already-super-rich Silicon Valley mogul who has enough money to play around, rather than a sober corporation driven by the bottom line or investment opportunities.

A quick reflection on corporate incentives explains this pattern nicely. In the case of Lehman Brothers, traders got bonuses linked to year-on-year profitability, and because of coordination problems each had incentive to maximize his own bonus but no incentive to maximize the solvency of the company as a whole over time.

But why would a CEO or other corporate governor create such a structure? Well, although Reactionaries mock elected politicians for having a four-year time horizon, the average CEO stays only 6.8 years. That's less than a two-term president. And their *own* incentives are often *also* based on bonuses linked to short-term profitability.

In theory, the incentive to increase shareholder value ought to counteract short-term-ist tendencies. But it's an open question exactly how much of a time horizon is built into stock prices. The average investor holds the average stock for about seven months. Although the hope is that stock prices are set by the market discount rate, at an weighted average cost of capital of 10%, this ideal situation still means that anything happening thirty or more years from now determines only 4% of the stock price.

In the real world, it's even worse than this – CEOs have strong incentives to try to fool the market into short-term inflation of stock prices at the cost of real future profitability. This is both common and successful. With many investors using formulae that extrapolate from past or present earnings to determine future earnings, it is unsurprising that the CEOs of companies like Lehman Brothers or Goldman Sachs were able to increase both their stock prices and their bonuses for many years until the inevitable letdown came – hopefully on someone else's watch.

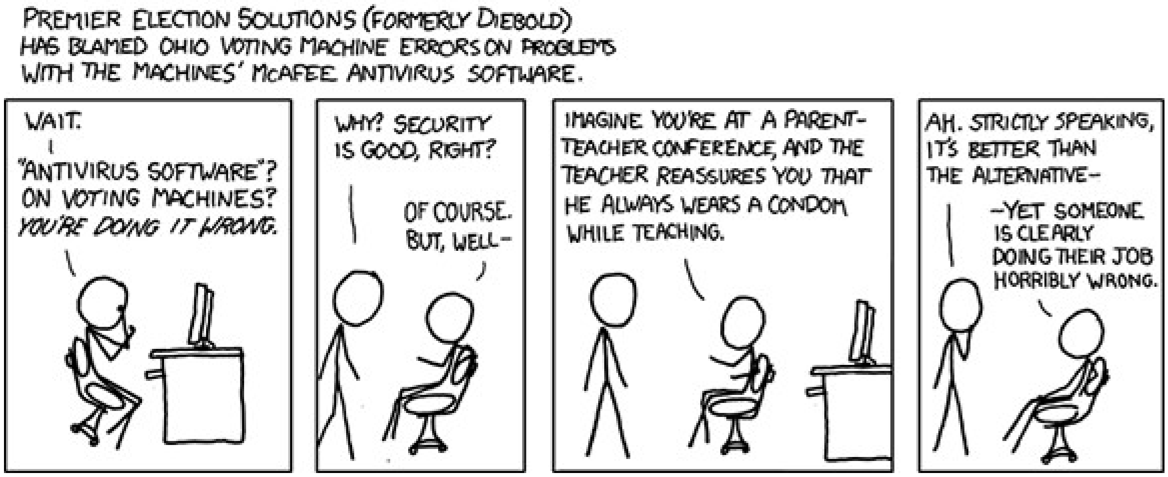
4.3: Could a joint-stock corporate state ensure complete security by mandating cryptographic locks on all its weapons?

This is one of Moldbug's proposals, and although I think it's been blown out of proportion and he's probably a little embarrassed by it now, it gets brought up enough to be worth addressing.

The idea is that shareholders of a corporate state possess cryptographic keys, and that these keys are necessary to fire the weapons in a country's arsenal. Therefore, any military coup can be stopped in its tracks.

The first question is exactly how these keys work. Suppose there are 100 shareholders. If all keys are necessary, then a single shareholder can paralyze the military. If 51 of the 100 keys are necessary – well, I don't know if cryptography can implement such a scheme securely, but let’s suppose that it does.

One can raise some peripheral problems with this scheme. Having all your country’s guns connected to the Internet might not be such a good idea…


PREMIER ELECTION SOLUTIONS (FORMERLY DIEBOLD) HAS BLAMED OHIO VOTING MACHINE ERRORS ON PROBLEMS WITH THE MACHINES' McAFEE ANTIVIRUS SOFTWARE.
WAIT.
"ANTIVIRUS SOFTWARE"? ON VOTING MACHINES? YOU'RE DOING IT WRONG.
WHY? SECURITY IS GOOD, RIGHT?
OF COURSE. BUT, WELL-
IMAGINE YOU'RE AT A PARENT-TEACHER CONFERENCE, AND THE TEACHER REASSURES YOU THAT HE ALWAYS WEARS A CONDOM WHILE TEACHING.
AH. STRICTLY SPEAKING, IT'S BETTER THAN THE ALTERNATIVE-
-YET SOMEONE IS CLEARLY DOING THEIR JOB HORRIBLY WRONG.

…and it would be sort of unfortunate if your entire military could be brought down by a clever hacker or Scott Aaronson building a quantum computer in his basement. Further, the guns would have to be either default-on or default-off. If they were default-on, then military conspirators could disable the communications network (or just the radios on their weapons) and have free rein. If they were default-off, then a foreign military could disable the communications network and take over the country because none of the military’s weapons would work.

More important, this only protects against a small subset of rebellions. If every unit has a separate code, it may be able to give loyalist military units the advantage over treasonous units in the case of intramilitary feuding. But it can’t can’t stop a popular revolution – the type where rebels become guerillas and gradually defeat the military in combat. It happened in China, it’s happening right now in Syria, and it could happen again regardless of any cryptographic locks on weapons.

4.4: Would shareholder value maximization be a good proxy for making a country a nice place to live?

Suppose that all the above problems are solved, and we have installed a genuinely self-interested CEO with a long time horizon. Will the new policy of increasing shareholder value really make the country a nicer place as well?

In many ways the equivalence holds. If, as Moldbug suggests, a corporate state’s profits came from land value taxes, and so profits came from increasing land values, then things like decreasing crime, pollution, and poverty would be in the corporate state’s best interests. So would allowing its residents enough freedom to make moving to its land attractive.

But the ways it *doesn’t* hold are really horrible.

Businesses have an incentive to please their paying customers. As Mitt Romney informs us, a large proportion of Americans don’t pay taxes. In fact, they consume government resources in the form of welfare, while providing no economic value in return. In some cases, these citizens are “fixer-uppers”, people who with enough investment could become productive. In other cases – the indigent elderly, the physically and mentally handicapped, or just people with no useful skills – keeping them around would be a poor financial decision. When regular companies find they have people who aren’t producing value, they “downsize” them. It’s unclear what exactly would be involved in “downsizing” unproductive American citizens, but I’m betting it wouldn’t win any Nobel Peace Prizes.

In a post called The Dire Problem And The Virtual Option, Moldbug discusses some of these problems with his system. He admits that this is a major issues (the titular “dire problem”). With his trademark honesty:

> As the King begins the transition from democracy, however, he sees at once that many Californians – certainly millions – are financial liabilities. These are unproductive citizens. Their place on the balance sheet is on the right. To put it crudely, a ten-cent bullet in the nape of each neck would send California’s market capitalization soaring – often by a cool million per neck. And we are just getting started. The ex-subject can then be dissected for his organs. Do you know what organs are worth? This is profit!

But his proposed solutions are bizarre and in many cases incomprehensible:

> The simplest, broadest, and most essential prevention against this degenerate result is the observation that the royal government is a government of law, and a government of law does not commit mass murder. For instance, no such government could take office without promising to preserve and defend its new subjects, certainly precluding any such genocide.

A government of law is different from a “law-abiding citizen” or “law-abiding business” in that governments, in addition to occasionally following the law, also get to *make* the law. If the government had some strong incentive to shoot citizens, it could pass a law allowing it to shoot citizens. It is no more than dozens of other governments have done throughout history. Such a law need not even ruffle the feathers of its more productive “assets”: it could come up with some very clear criteria for whom to shoot and then stick to those criteria scrupulously.

No government could take office without promising to preserve and defend its new subjects *in a democracy*. Or, to be broader, no government could take office under such conditions as long as it was responsible to its populace and depended on their support. The entire premise of Moldbug's utopia is a government whose rule is by force and does not depend on the consent of the governed.

If Moldbug's King needed to gain the consent of the governed before taking power, they wouldn't stop at making him sign a promise not to shoot anyone. They would make him sign a promise to rule for the good of the people rather than in order to maximize shareholder value. Heck, the last time we tried something like this, the people made the government sign the Bill of Rights.

Here Moldbug wants to have his cake and eat it too. His government will be unconstrained and effective because it doesn't rule by consent of the people. But when we start examining how horrible an "unconstrained effective government" really would be, he promises that need for the consent of the people would rein it in.

Positing a government that can ignore the age-old constraint of popular consent is far-fetched enough. Positing one where the constraint *only* arises in those situations where it would be optimal for it to arise, but not otherwise, is just dreaming.

> But do we really know it? The explanation that Royal California will not harvest the poor for their organs, because it will have promised not to harvest the poor for their organs, and its most valuable asset is its

> reputation, while certainly accurate, is too narrow for me. Having established this legalistic defense, let us reinforce it with further realities. More broadly, Royal California will in all cases treat her subjects as human beings. The maintenance of equity, as well as law, is crucial to her reputation. Thus, the Genickschuss is out, with or without the organ harvesting.

Our second layer of protection is that the king will preserve human rights and maintain equity among persons. I wonder if the person writing this has ever read Mencius Moldbug. He has some pretty interesting arguments against human rights and the equity of persons, and I'd be interested in hearing a debate between the two of them.

> Carlylean to its core, the ideology of Royal California is that the King is God's proxy on earth; whatever God would have him do, that is justice; the King, having done his best to divine God's will, shall see it done. Or else he is no king, but a piece of cardboard, a "Canadian lumber-log." Clearly, God is not in favor of harvesting the poor for their organs. You're probably thinking of Huitzilopochtli. So this is another safeguard.

So our third layer of protection – and I am not making this up – is "the will of God". Don't you feel safer already? Politicians would never do bad things, even when it is in their own self interest, because God wouldn't want them to. I think that's pretty much all the protection citizens might need from their government, don't you? Let's write a letter to the libertarians and tell them they can all go home now, *God* has this one covered.

But I should not be too harsh on Moldbug. He goes on to admit we probably do need a fourth layer of protection, beyond the three he has mentioned. And he even steelmans the case against him, noting that in a higher-technology world, more and more people will become unproductive until, instead of being a tiny proportion of citizens, it may become the majority or (in the post-Singularity case) everyone who has to worry about this. He gives a few possible solutions:

> First, the King has no compunction whatsoever in creating economic distortions that produce employment for low-skilled humans. A good example of such a distortion in the modern world are laws prohibiting self-service gas stations, as in New Jersey or Oregon. These distortions have gotten a bad name among today’s thinkers, because makework is typically the symptom of some corrupt political combination. As the King’s will, it will have a different flavor.
>
> As both a good Carlylean and a good Misesian, the King condemns economism – the theory that any economic indicator can measure human happiness. His goal is a fulfilled and dignified society, not maximum production of widgets. Is it better that teenagers get work experience during the summer, or that gas costs five cents a gallon less? The question is not a function of any mathematical formula. It is a question of judgment and taste. All that free-market economics will tell you is that, if you prohibit self service, there will be more jobs for gas-station attendants, and gas will cost more. It cannot tell you whether this is a good thing or a bad thing.
>
> There may be no jobs for men with an IQ of 80 in Royal California – at least, not in a Royal California whose roads are paved by asphalt rollers. But suppose its roads are paved in brick? A man with an IQ of 80 can lay brick, do it well, and obtain dignity from the task. Nothing

> whatsoever prevents the King from distorting markets to create demand for the supply he has.

Okay, so the corporate CEO in a government based solely on maximizing shareholder value will decide to trash his own economy in order to provide jobs for the jobless, because that's just how much corporate CEOs respect human dignity. This is just like corporate CEOs today, who never fire anyone to increase profitability because maintaining jobs is more important. Sure, let's roll with that.

Since we have abandoned the free market here, we no longer have the free market's safeguards on job tolerability. Depending on how many make-work jobs the King creates, we will have either an oversupply, an undersupply, or a just-right-supply of unskilled laborers to fill them, which in turn will determine workers' wages and living conditions. Will the King maintain them at a living wage in good conditions, or at conditions more like the immigrant farm laborers of today? If the latter, I *suppose* that's better than killing off the unproductives, but it's still pretty dystopian. If the former, then that's quite nice of the King, but I can't help noting that by instituting useless make-work government-provided jobs for everyone at guaranteed salaries, he has kind of just re-invented Communism, which seems to be the sort of thing I would have expected Reactionaries to try to avoid.

I would compare this idea to the idea of a Basic Income Guarantee. Both cost the economy the same amount of money. Yet in Moldbug's plan, the poor spend their entire day digging ditches and filling them in again. In a basic income guarantee, the poor spend their days doing whatever they want – producing art, playing games, or working to make themselves more productive. Moldbug may wax rhapsodic about the dignity of work, and he is not entirely wrong, but the sort of work that has dignity is not the sort of work where you dig ditches and fill them in again to earn a government-set paycheck. I wonder if you asked the employed gas station attendant and the unemployed bohemian to rate the level of dignity they feel they have, would this support Moldbug's thesis?

But never fear, Moldbug has yet another plan:

> Or not. The low-browed man of 70 (and remember – for every 130, there is a 70) may still require special supervision. Besides a job, he needs a patron. Productivity he has, but direction and discipline he still requires. His patron may be a charity, or a profitable corporation, or even – gasp – an individual.
>
> In the last case, of course, we have reinvented slavery. Gasp! Since the bond of natural familial kindness is not present in the case of an unrelated ward, the King keeps a close watch on this relationship to protect human dignity. Nonetheless, his wards are farmed out – it is always better to be a private ward than the ward of the State. Bureaucratic slavery is slavery at its worst. Adult foster care, as perhaps we will call it, is a far more human and dignified relationship.

So, we will force people to work for other people against their consent, but it will all be okay and humane, because the government will be keeping "a close watch on this relationship"? Darnit, I liked it better when we were being protected by "the will of God".

If Moldbug agrees that bureaucratic slavery is "slavery at its worst", what exactly does he mean when he says the King will "keep a close watch" on these "adult foster care" institutions. Will the King personally go out to each of them and evaluate? That seems like a lot of work in a state of 40 million people. Or will he appoint some government officials to do so, to inspect each institution and make sure it is up to code? If so, how is this different from “bureaucratic slavery”? Is it because the bureaucrats and slaveowners aren’t *literally* the same people?

Look, Moldbug. I know you don’t think you’re reinventing Communism, *but you are*.

Luckily he has one more trick up his sleeve:

> If a human being cannot support himself in a civilized manner in the King’s economy, which has been carefully tweaked to match labor demand to labor supply, the King does not provide a “safety net” in the 20th-century style, in which he may lounge, sag, bob and fester forever. No – then, it is time for the Virtual Option.
>
> If you accept the Virtual Option – always a voluntary decision, even if you have no other viable options – California will house, feed and care for you indefinitely. It will also provide you with a rich, fulfilling life offering every opportunity to obtain dignity, respect and even social status. However, this life will be a virtual life. In your real life, your freedom will be extremely restricted: to the point of imprisonment. You may even be sealed in a pod.
>
> The result is that the ward (a) disappears from society, and (b) retains or (hopefully) increases his level of dignity and fulfillment. He remains a financial liability, because it is still necessary to prepare his meals and maintain his pod. But other residents of California no longer feel menaced by his presence. For he is no longer present among them.

This doesn’t sound so bad to me, although I’m probably a huge outlier on this and if you actually tried it on people you’d have a civil war on your hands.

But first of all, it’s impossible with current levels of technology, always a bad sign.

Second of all, it’s something that would be equally viable in a democracy and a monarchy. Compare these pods to television. Right now, we pay welfare money to the poor, and, in some cases, they use that money to watch television all day. When they complain, it generally is not due to a lack of television but to a lack of money. If we had virtual reality pods, no doubt the situation would look little different, and conservatives and Reactionaries would be the ones complaining that we pay the poor money to sit in virtual reality pods all day instead of getting a real job.

Third of all, it would probably cost more than any other option. Putting a man in prison – feeding him, boarding him, and putting some guards on the doors to make sure he doesn’t escape costs about $50,000 a year – more than sending that same man to any college in the country. The bulk of the expenses are health care and security – two problems that would be equally dire in these pods. In fact, solving the medical problems associated with prolonged immobility in a virtual environment might be further beyond our current technology than the virtual environment itself.

If the true reason behind the Virtual Option is keeping the poor out of everyday society – even though many of its residents would be old people, disabled people, and the like – why not just offer those people $40,000 a year to live in some nice community out in the country made up solely of other non-working poor? It would be cheaper, more humane, and after a few years with a stable income and a normal life the people involved might end up being unexpectedly productive.

This is, of course, a question one could ask of our own society as well as of Moldbug's hypothetical. So let's stick to criticizing Reactionaries, which is more fun and less depressing.

4.5: Would exit rights turn countries into business-like entities that had to compete with one another for citizens?

Exit rights are a great idea and of course having them is better than not having them. But I have yet to hear Reactionaries who cite them as a panacea explain in detail what exit rights we need beyond those we have already.

The United States allows its citizens to leave the country by buying a relatively cheap passport and go anywhere that will take them in, with the exception of a few arch-enemies like Cuba – and those exceptions are laughably easy to evade. It allows them to hold dual citizenship with various foreign powers. It even allows them to renounce their American citizenship entirely and become sole citizens of any foreign power that will accept them.

Few Americans take advantage of this opportunity in any but the most limited ways. When they do move abroad, it's usually for business or family reasons, rather than a rational decision to move to a different country with policies more to their liking. There are constant threats by dissatisfied Americans to move to Canada, and one in a thousand even carry through with them, but the general situation seems to be that America has a very large neighbor that speaks the same language, and has an equally developed economy, and has policies that many Americans prefer to their own country’s, and isn’t too hard to move to, and almost no one takes advantage of this opportunity. Nor do I see many people, even among the rich, moving to Singapore or Dubai.

Heck, the US has fifty states. Moving from one to another is as easy as getting in a car, driving there, and renting a room, and although the federal government limits exactly how different their policies can be you better believe that there are very important differences in areas like taxes, business climate, education, crime, gun control, and many more. Yet aside from the fascinating but small-scale Free State Project there’s little politically-motivated interstate movement, nor do states seem to have been motivated to converge on their policies or be less ideologically driven.

What if we held an exit rights party, and nobody came?

Even aside from the international problems of gaining citizenship, dealing with a language barrier, and adapting to a new culture, people are just rooted – property, friends, family, jobs. The end result is that the only people who can leave their countries behind are very poor refugees with nothing to lose, and very rich jet-setters. The former aren’t very attractive customers, and the latter have all their money in tax shelters anyway.

So although the idea of being able to choose your country like a savvy consumer appeals to me, just saying “exit rights!” isn’t going to make it happen, and I haven’t heard any more elaborate plans.

5: Are modern ideas about race and gender wrongheaded and dangerous?

The past century has seen a huge opening up of racial and sexual norms, as a closed-minded traditional society willing to dismiss everything against their personal morals as disgusting or evil started first discussing and later embracing alternative ideas.

This was followed by a subsequent closing back up of those norms, as society decided it was definitely right this time, and this time *for real* anyone who brought up any alternative possibilities was definitely disgusting and evil.

Reactionaries deserve kudos for lampshading these taboos and pointing out various modern hypocrisies in a frank and honest way. But to invert an old saying, I will defend to the death their right to say it, but disagree with what they say.

5.1: Are modern women sluts?

This is a surprisingly important question in Reactionary thought. Just to prove I'm not strawmanning:

> So you might say, Bryce, if you want an objective and useful definition of the word slut, you would have to conclude that most Western women are sluts. That's not good. And I say "Exactly."

– Anarcho-Papist

> Obviously democracy is not working, is failing catastrophically. The productive are outvoted by the gimmedats, in large part non asian minorities and white sluts.

– blog.jim.com

> Why would you take a slutty girl seriously? Once she accepted slut into her life, keep her out of yours. It is rare for a slut to truly reform so I would not even take the chance. Once a slut, always a slut. Do you really

> want your kids coming out the same place 10 other men have gone into? "But doesn't that pretty much rule out about 85% of women or so?" Well, unfortunately it does. I wish there was a better answer but there is not. Do not settle for sluts, if they have such little respect for themselves imagine how little respect they will have for you. Manning up does not mean settling for a hopeless graying slut."

– Occidental Traditionalist

> We live in strange times. Recently several religious conservative bloggers have suggested that the word "slut" is a slur against all women, and that it is a type of profanity. My best guess is they feel that sluts know that what they are doing is wrong, so even using the word in general is cruel to their already convicted hearts.

– Dalrock

Telling women that sleeping around is bad just because it's "slutty" is argument through mere connotation of words. Then again, accusing these people of "sexism" or "misogyny" would be the same. So let's bury the insults and try to figure out what's going on.

Are people becoming sluttier? Several studies have addressed this question (though, uh, not in those exact words). In America, we have only a few scattered studies recording a shift from an average of two lifetime sexual partners for women and six for men in 1970 to about four partners for women and six for men in 2006. But we change methodologies midstream and have to confuse means with medians to get those numbers. France is the only country to do the study properly, perhaps unsurprising given their legendary love of all things amorous. Their numbers seem similar to ours but more precise, so let’s use the French results:

> Number of partners reported in the lifetime remained stable between all three surveys for men of all ages (11.8 in 1970, 11.0 in 1992, and 11.6 in 2006). For women, mean lifetime number of partners increased from 1.8 in 1970 to 3.3 in 1992 and to 4.4 in 2006.)

One of the first things we notice about these data is that they cannot possibly be true. Men cannot be having more (heterosexual) sex than women, nor can the two statistics trend in different directions. The least mathematically impossible explanation is that between 1970 and 2006, women have become less likely to lie about all the sex they’re having.

Does that contradict common sense, which tells us everyone is really slutty nowadays but was perfectly chaste in the past? Maybe, but common sense seems to be not entirely correct. Common sense would tell us that modern young people are having much more sex than youth fifteen years ago, but according to the study “no increase was observed between 1992 and 2006 in women under thirty; for men under thirty a decrease in the mean was seen in the most recent period – 10.4 in 1992 and 7.7 in 2006, $p < 0.00001$" (the growth of the Muslim population in France from 7% to 10% during that time period seems insufficient to account for the changes) **5.1.1: If a woman is a slut, does that mean her future marriage is doomed to failure?**

Before you answer, consider a common failure mode. Some rule catches on for some very useful reason. Like “don’t have sex with your cousin, you’ll have kids with two heads.” Biological or memetic evolution selects for people who follow the rule, and eventually the rule becomes an unquestionable taboo.

But historically no one understood Mendelian genetics. The rule didn’t make sense, but it had to be followed. And so people came up with rationalizations. Some of them were simple rationalizations for simple folk: “don’t have sex with your cousin, God hates it.” Or “Don’t have sex with your cousin, it’s disgusting.” More sophisticated people demanded more sophisticated rationalizations: eventually you get “Don’t have sex with your cousin, it could go wrong and damage the structure of trust necessary for an extended family”, or “Don’t have sex with your cousin, it is contrary to this here complicated conception of natural law”.

Then suppose the original reason for the rule is taken away. Someone wants to have protected sex with their cousin, understanding that they cannot ethically have children. Or someone invents a gene therapy that allows people to have sex with their cousins without additional risk of birth defects.

Doesn’t matter. Everyone will have had so much fun making up rationalizations that they will object to the new harmless act almost as much as to the old dangerous act. “God still hates it!” “It’s still disgusting!” “It still damages the family structure of trust!” “It’s still contrary to the natural law!”

But it would be very strange if, the original reason for the belief having been neutralized, by coincidence the belief happens to be right anyway. Imagine that an explorer comes back from a distant jungle with a tale of a humongous monster. Everyone catches monster fever and begins speculating on how the monster may have gotten there. Then the explorer admits his tale is a hoax. Objecting “But there could still be a monster there!" is fruitless. If the original reason anyone held the belief is invalid, it's unlikely that by coincidence the belief just happens to be correct.

Let's get back to sluttiness. (I am following the lead of my interlocutors in concentrating on female sluttiness only here, since it seems to be the only type anyone cares about. Yes, you're very clever for pointing out that men can be promiscuous as well. Why don't you follow it up with the phrase "double standard" or a reference to "playing the field"?)

We know two *very* good reasons why sluttiness has been stigmatized in nearly all societies. First, slutty women were more likely to get sexually transmitted diseases. Second, slutty women were likely to end up with children outside of wedlock. Back when men were the sole providers and didn't have much providing to spare, that would have been just about a death sentence.

These are two *huge* issues. These two issues alone are more than sufficient to explain the taboo on sluttiness establishing itself on every continent and in every major religion. These are more than sufficient to explain why some people think sluts are disgusting, why they're low status, why we have a cultural taboo on sluttiness.

But of course, most sluts today have these two issues figured out. Contraception prevents the out of wedlock births. Protection and antibiotics prevent the STDs. So the old reasons no longer hold.

It would be quite the coincidence if a taboo that formed for one reason *just happened* to be vitally important for society for totally different reasons.

I admit the Reactionaries have their justifications for why sluttiness is bad. They say sluttiness before marriage can lead to sluttiness after marriage, and thither to infidelity, divorce and broken families. Or the slut's previous experiences might have given her higher expectations, leading to divorce and broken families again. And…

…no, that's actually all the justifications I can find. There are people who think they have other justifications, but they can never explain them in so many words. Read this article. No, really, read that article. Gods! Have you ever seen so many mere assertions and Arguments From My Opponent Believes Something in one place?

So okay. They have two just-so stories. I can come up with just-so stories too! Like – if a woman sleeps with a lot of people before marriage, she'll be better able to estimate how compatible she is with any given partner. Or – if a woman can sleep with men before marriage, she won't be compelled by horniness to marry the first loser she meets just so she can have sex with someone. Or – if a woman has a couple of relationships before she marries, she'll have practice with relationships and won't screw the important one up. This is fun! How about – if a woman sleeps with people before settling down, she won't feel curiosity that makes her stray afterwards?

The reason these sorts of just-so stories about sluttiness keep popping up is the disappearance of the good historical arguments against the practice, leaving behind only a feeling of disgust in search of a justification.

One might argue – isn't the proof in the pudding? Divorce rates have been going up lately, infidelity rates have been going up; correlation isn't always causation but isn't it at least suggestive?

In this case, no. We can even check. From Social Pathology:

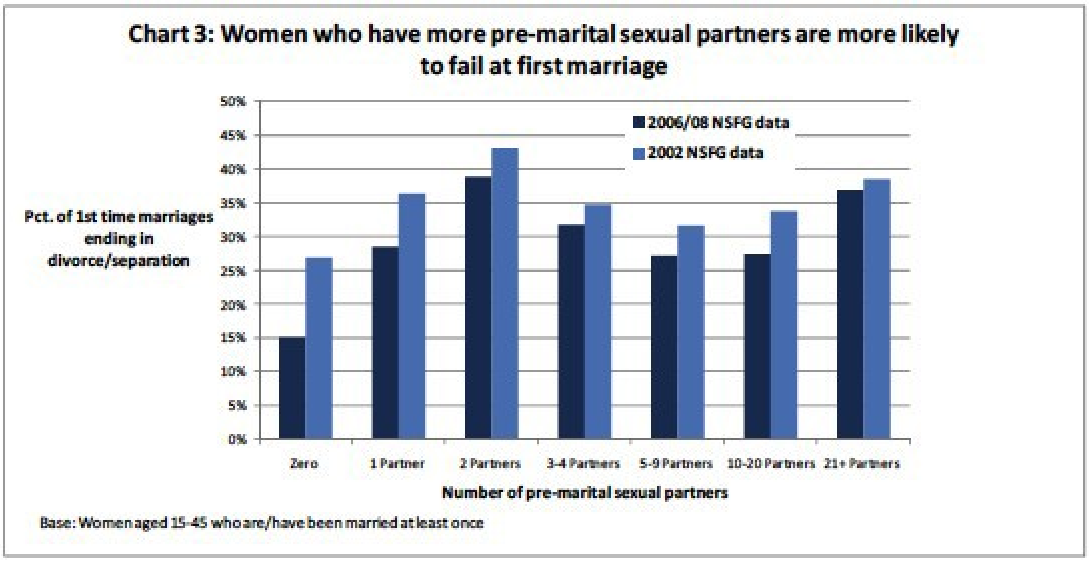


Chart 3: Women who have more pre-marital sexual partners are more likely to fail at first marriage

Base: Women aged 15-45 who are/have been married at least once

Women with zero or one premarital sexual partners have more stable marriages than women with two or more partners. Okay. Who gets married a virgin these days? Super-religious people. They’re not going to divorce. And from the source, I gather that most of these stably married one partner women are women who had premarital sex with their future husband. Super-religious people who slipped up. Their poor self-control earns them a 15% lower likelihood of stable marriage: harsh, but fair.

The people with two or more partners are the ones who we know are “experimenting” – having sex with at least one person other than their future husband. Among this group, likelihood of unstable marriage goes *down* with more partners up until you reach the 20 partner or so level – at which point you’re probably capturing prostitutes, cluster B personality disorders, and other people outside the mainstream.

The data provide some evidence that an absolute commitment to purity – no sex before marriage, or sex only with your husband-to-be – predicts marital stability. But beyond that – in the two to twenty partner range in which recent social change has been occurring – there's no correlation between increasing sluttiness and decreasing marital stability.

5.1.2: Woman only put out for macho but antisocial men. Our society encourages that tendency and shames "beta males" who are nice and prosocial but cannot get women. This incentivizes men to become jerks, and men follow those incentives in droves. Don't we need to do something about women's tendency to make poor choices?

There's no shortage of places to find this argument, but the obligatory link goes to Free Northerner for One More Condom In The Landfill, a particularly good presentation of the idea.

In a broad perspective the point is correct – empirically, men with more psychopathic traits, less agreeableness, and greater narcissism have more sexual partners.

On the other hand, it is kind of ironic that the pickup artist community – one of the few communities to be perfectly honest about the above point – has become obsessed with scoring the hottest girls and denigrating the others, no matter how perfect they might otherwise be.

The complaint tends to be "You women keep asking where the good men are, but they're right where you left them when you refused to date them because you only cared about cockiness and bulging muscles." The countercomplaint might be "You men keep asking where the good women are, but they're right where you left them when you refused to date them because you only cared about stylishness and big breasts."

I also suspect (though I have no evidence) that it is primarily the hotter women who have been socialized to be irrationally attracted to "bad boys", and that pickup artists' disproportionate focus on this demographic skews their assessment of the problem.

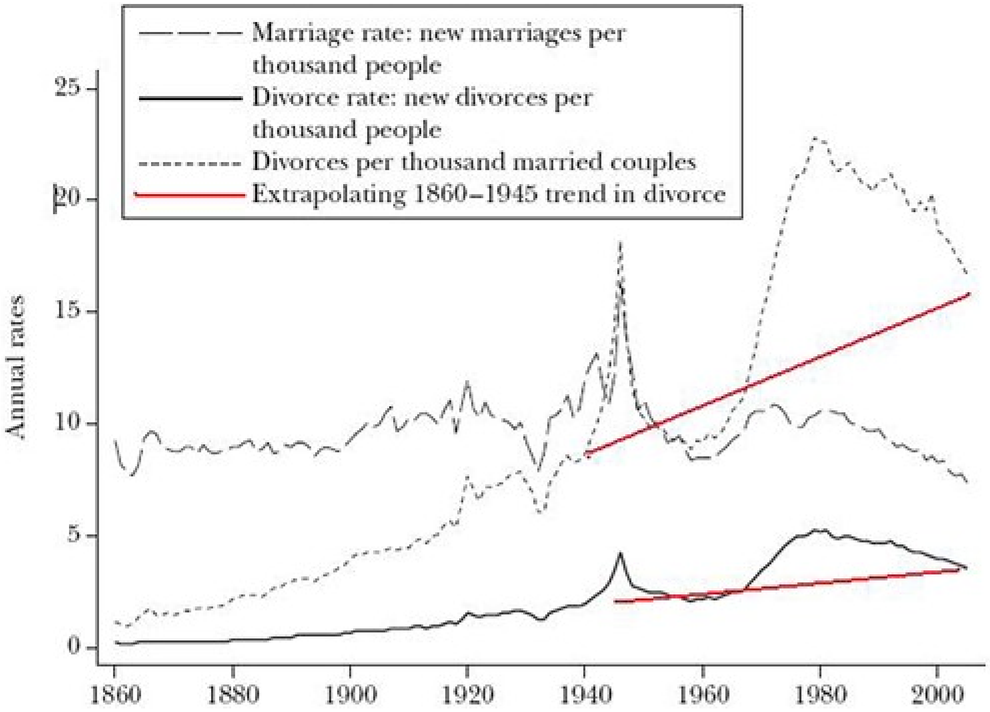
If one were to phrase the problem as "Men and women both make stupid and counterproductive sexual choices; how can we optimize for avoiding those?", then that might make the sane 30%-or-so of feminists join the conversation and get something done.

If you phrase the problem as "Those women make stupid and counterproductive sexual choices, how can we shift the balance of power toward men?", even the sane 30%-or-so of feminists will ignore and oppose you, and with good reason.

I have no idea how to solve the object-level problems, by the way, although I would tentatively recommend my own strategy of sidestepping the problems with both hot men and hot women by dating a hot genderqueer.

5.2: Are Progressive values responsible for rising divorce rates?

Let's get the obvious objection out of the way first: divorce rates have been falling since about 1980. They're now at their lowest level since 1970 or so, and dropping still.



Marriages and Divorces per Thousand People, United States 1860–2005

The other thing this graph tells us is that rising divorce rates were a phenomenon very specific to the period about 1965 – 1975. This was a good decade for liberal values, but little moreso than decades before and after it. The strictly time-limited nature of the phenomenon suggests something more specific (and no, it's not no-fault divorce laws). The Pill, which came out in 1960, is an *extremely* plausible candidate, but a full treatment of this topic is beyond the scope of this essay.

Now that the obvious objection is out of the way, let's discuss some less obvious objections. If progressive values cause divorce, how come people with more progressive values are less likely to divorce? College-educated women have about half the divorce rate of the non-college-educated (source). More conservative states have higher divorce rates than more liberal states (source). Atheists have divorce rates below the national average (source).

Some of these factors seem to remain even when controlling for wealth and the other usual confounders (source, source). The link between sluttiness and stable marriage mentioned above reinforces this point.

I think this data is consistent with the following theory: new technology and changing economic conditions produced a strain on family life that was reflected in an explosion in divorce rates. Society’s memetic immune system sprung into action to contain the damage through the creation of new laws, institutions, and social norms. People who adopted the new ways survived the crisis and their family lives returned to a sort of normal. People who failed to adapt…well, don’t be one of those people.

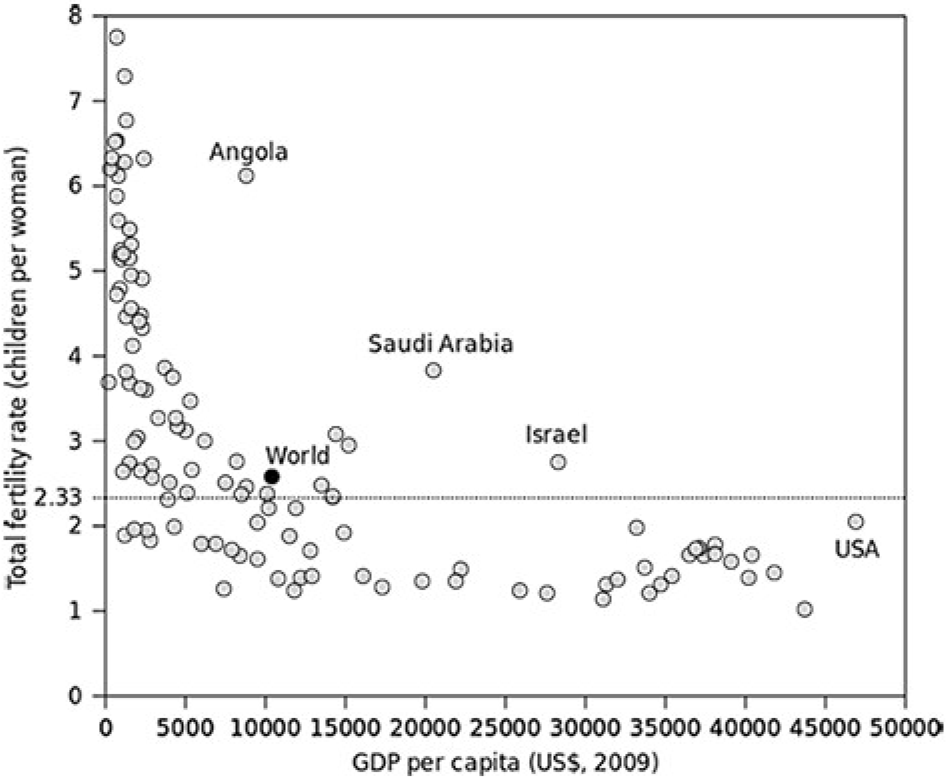
The new norms created by the memetic immune system are exactly the progressive values that Reactionaries blame for the damage: marrying later, trying more partners, using more contraception, having fewer children.

This theory explains both why the progressive values arise at the same time as the broken families, but also why people with progressive values are less likely to have broken families than others.

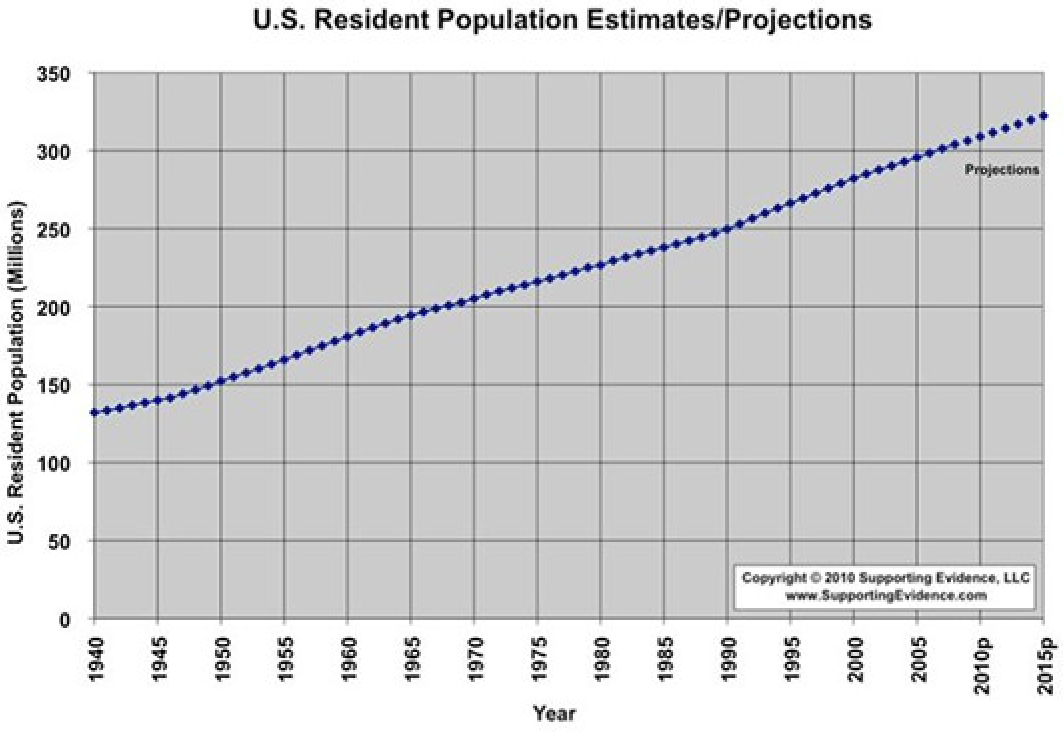
The data on illegitimate children and single motherhood mirror the data on divorce and do not require a separate discussion.

5.3: Are we headed for a demographic catastrophe?

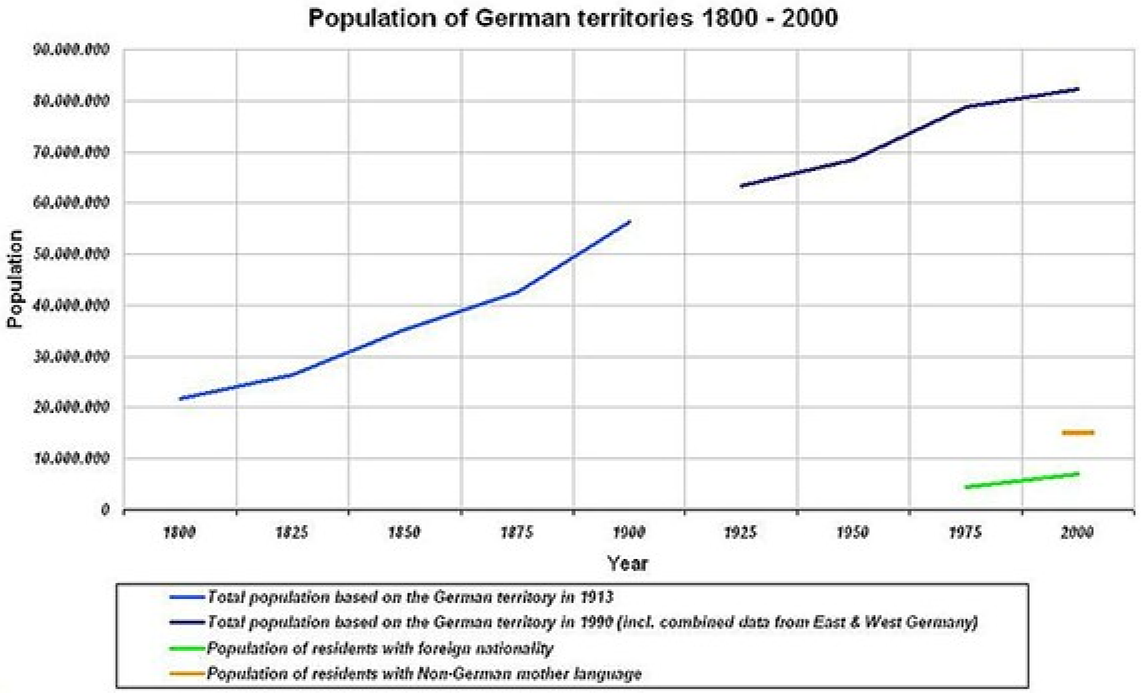
First of all, before we pretend that the minutiae of who has which values and who goes to church how many times affects fertility rate much, let’s see the inevitable GDP/fertility rate graph:



And before we worry about the United States experiencing demographic collapse and tumbleweeds rolling through the streets of New York City, let’s double-check to make sure that US population isn’t a near-perfectly straight upward-trending line:



Western Europe?



A few countries do have demographic problems. Singapore, for example, has the lowest fertility rate in the world – 0.79, 224th out of 224 countries. It should probably do something about that. But given that it's generally accepted to be the most Reactionary country in the world, it's hard to blame this one on Progressivism or suggest Reactionary values as the answer.

5.3.1: But what if I am racist? Isn't it possible that fertile minorities and immigrants are hiding a fertility deficit among precious, precious, white people?

According to Edmonston et al's projection of US racial fertility trends:

> In 2100, the total U.S. population will eclipse 550 million people, and the racial composition of the country will be 38.8% white, 30.6% Hispanic, 15.6% black, 14.9% Asian and Pacific Islander, and 1% American Indian.

The absolute number of white people will be only a few million less than today, 209 million. That's more than enough to run a wide selection of excellent country clubs, or achieve whatever other strategic aims we need a large white population for.

Perhaps most gratifying if you are a racist, the percent of black people will increase only about three percentage points. The biggest increase will be in Asians, a so-called model minority.

After that? If there are still biological humans in organic bodies transmitting genes naturally much after 2100, we have *much* bigger problems than race on our hands.

5.3.2: Are we headed for an idiocracy?

Poor, uneducated, low-IQ people have higher fertility rates than wealthy, well-educated, high-IQ people in almost all countries. Therefore, one might worry that this will have a dysgenic effect, selecting against genes for intelligence until eventually everyone is stupid or has other undesirable quantities anticorrelated with wealth and education. This was the premise of the movie *Idiocracy*, and in principle people are far too quick to dismiss it.

But in practice, the effect is too small be significant. Richard Lynn, who is the closest we will get to an expert on dysgenics, calculates that American society as a whole is losing 0.9 IQ points per generation. So by 2100, people will have lost on average 4 IQ points.

Since it's hard to get a good intuitive graph of what 4 IQ points means, consider that IQ has been increasing by about 3 points per decade (average is still 100, but only because they recalibrate it). So absent any further Flynn Effect, losing 4 IQ points would take us back to…about as smart as we were back in 2000. I won't say that won't be unpleasant – the people of that era elected George W. Bush, after all – but it's not quite convert-all-written-language-to pictograms-because-everyone-has-forgotten-how-to-read level unpleasant.

And what comes after 2100 doesn't matter, because even on the off chance we're still using human brains to reason at that point, it sure won't be human brains in which the genes have been left to chance. To paraphrase Keynes, in the long run we're all either dead or cyborgs.

5.4: Aren't modern dogmas about race and sex and sexuality stupid and evil?

Let me be clear here. There is no excuse for the sort of extremist folk social justice crusades one can find on Tumblr or Twitter or Freethought Blogs. With a few treasured exceptions they are full of nasty and hateful people devoid of intellectual integrity and basic human kindness, and I am suitably embarrassed to be in the same 50%-or-so of the political spectrum.

Then again, there are lots of nasty and hateful conservatives and reactionaries devoid of intellectual integrity and basic human kindness too. Go take a look at Free Republic. Maybe we can call it a tie?

But this has surprisingly little bearing on the particular question above. As Christians are obligated by circumstance to point out, an idea is not responsible for the quality of people who hold it. And modern dogmas about race are agreed by very nearly everyone – including most Reactionaries! – including you! – to be both correct and very important.

Three hundred years ago, a pretty high percent of Americans were okay with black people getting kidnapped, enslaved, forced into back-breaking labor on plantations, raped, separated from their children, whipped if they protested, worked to a very early death, and then replaced with other black people.

Nowadays Reactionaries like to think of themselves as racist just because they believe the average black IQ is a standard deviation below the average white IQ. But one standard deviation implies that about a fifth of black people are smarter than the average white person. If you were to go back to 1800 and tell a conference of the most extreme radical abolitionists that you thought a fifth of black people were smarter than the average white person, they would laugh and not stop laughing until they died of laughter-induced asphyxiation.

And at least there the traditional and modern stereotype are still going the same direction. Did you know there used to be a stereotype that Jews were stupid and boorish and didn’t belong in polite society? A stereotype that Chinese people were dumb? A stereotype that black people were bad at sports? To make a corny statistics pun, there seems to be very poor inter-hater reliability.

Homosexuality is little different. Reactionaries take a bold stand against sexually suggestive displays at gay pride parades or whatever, but when it comes to why two people who love each other can’t get married because they’re both the same gender, they tend to be just as confused as the rest of us. Mencius Moldbug writes:

> Although I am straight as an iron spear, I happen to see nothing at all wrong with “gay marriage.” In fact I am completely sympathetic to the Universalist view, in which the fact that couples have to be of opposite sexes is a sort of bizarre holdover from the Middle Ages, like the ducking-stool or trial by fire. It’s not clear to me why homosexuality, which obviously has some extremely concrete biological cause, is so common in modern Western populations, but it is what it is. However, because I am straight etc, and also because I’m not a Universalist, I happen to think the issue is not really one of the most pressing concerns facing humanity.

Moldbug is welcome to his opinion on what is or isn’t one of the most pressing concerns facing humanity (I would have said a couple of brain-dead Internet thugs from Gawker beating up on a random Twitter celebrity isn’t one of the most pressing concerns facing humanity, but to each his own) but I wonder if Moldbug notices that merely his unconcern on this issue makes him in let’s say the 95th percentile of most Progressive Americans who have ever lived. 95% of Americans throughout history have been *quite* certain that eradicating sodomy *was* one of the most pressing concerns facing humanity, and boy did they act on that belief.

In fact, if we put a Reactionary in a time machine headed backward, and made it stop when the Reactionary was just as racist, sexist, et cetera as the US population average at the time, I predict they wouldn’t make it much past the 1970s. Go into the 1960s and you get laws banning colleges from admitting both black and white students to the same campus (one helpfully specified that the black and white campuses could not be within twenty five miles of one another).

Now, there’s no problem with this – except for Nixon and disco, the 1970s were no worse than any other period. But Reactionaries insist that all Progressivism since 1600 has been part of one vast and monstrous movement – maybe a religious cult, maybe a sinister power-play, maybe just the death throes of the western intellectual tradition – dedicated to being wrong about everything. And that a very big part of this vast movement focused on race. And when they have to whisper “Except we agree with 99% of what it did, right up until the past couple of decades, and in fact they got it right when everyone else was horribly, atrociously wrong”, that is – or at least *should be* – kind of embarrassing.

5.4.1: But there’s a clear difference between the past policies Reactionaries support and the modern ones they oppose. Past policies were going for equality of opportunity, modern ones for equality of results.

Isn’t seeking equality of results laden with too many assumptions?

Arguing about whether a post-racial society should provide equality of opportunity or equality of results is a little like arguing about whether in the worker’s paradise, everyone should have a pony or everyone should have *two* ponies.

Right now, there is not even equality of opportunity. Rigorous well-controlled study after rigorous well-controlled study has shown that women and minorities face gigantic amounts of baseless discrimination in various areas, most notably employment. This remains true even when, for example, the experiment is sending perfectly identical resumes out to companies but with the photo of a black or white guy at the top.

Once we have equality of opportunity, *then* we can start debating whether we should go further and try for equality of results. Until then, it’s kind of a moot point.

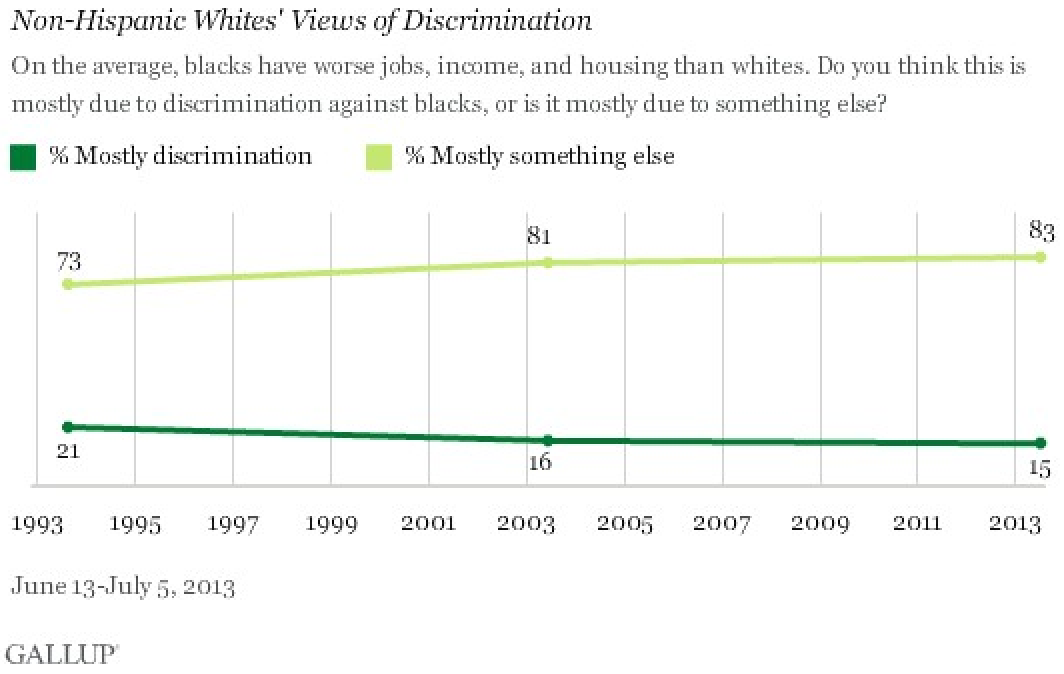
5.4.2: What about the studies that have shown black people have lower IQ/higher violence/other undesirable trait than white people?

If genetic differences across races prove real, this would be a good argument against seeking equality of results, but no argument at all against continuing to seek equality of opportunity – which, as mentioned above, mountains of rigorous well-controlled studies continue to show we don’t have.

If, as the scientific racists suggest, black people have an average IQ of 85 compared to the white average of 100, then there is *still* a pretty big civil rights battle to be fought getting the average black person to do as well as the average white person with IQ 85. After controlling for IQ, the average black person is still twice as likely to be in poverty, 50% more likely to be unemployed, and 250% more likely to be in prison (source, other gaps appear to disappear or reverse once IQ is controlled; see link for a more complete analysis.)

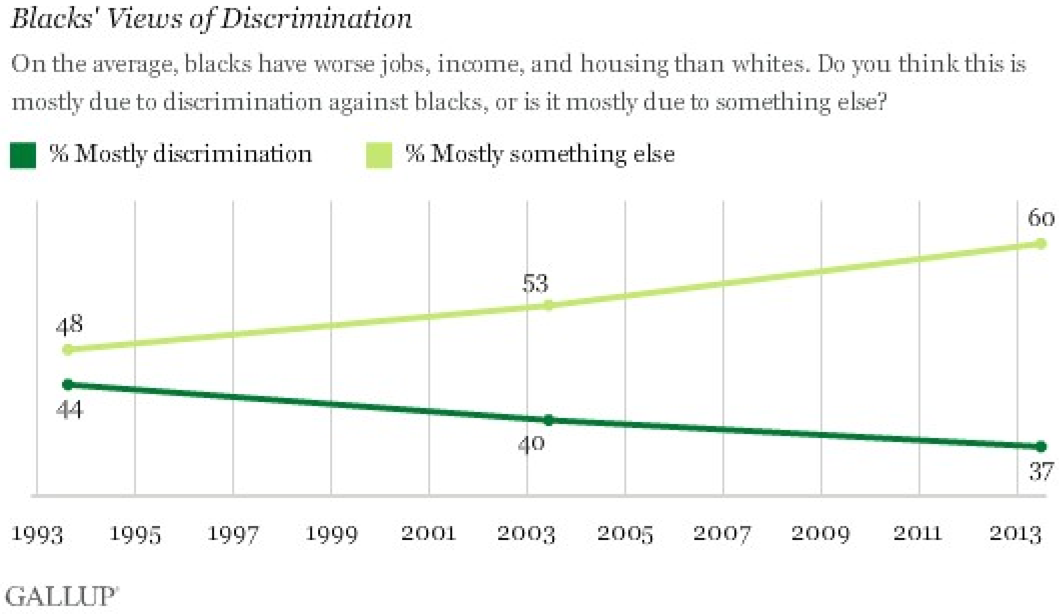
5.4.2.1: But this is exactly the kind of discussion progressives won’t let us have! It is an unquestioned dogma of our society that all cross-racial differences must be based entirely on discrimination! In fact, people educated in public schools are incapable of even conceiving of the possibility that they could be otherwise! How are we supposed to be able to disentangle equality of opportunity from equality of results in such people?

From this Gallup poll:



83% of white people agree that the poor position of blacks in society is mostly not due to discrimination.

Want to see something even cooler?



60% of *black people* agree that the poor position of blacks in society is mostly not due to discrimination.

So no, doubting that all racial disparities in the US are due to discrimination isn't a thought crime. It's the majority position, even among black people themselves.

True, the number of people willing to consider genetic differences in particular would probably be far lower. But the great (and very legitimate) fear motivating more-than-academic interest in this question – that white people will forever be blamed for and forced to atone for minorities' problems – is one that can be talked about productively and perhaps banished.

5.4.3: Even if the establishment has not managed to completely ban all discussion of race that contradicts their own ideas, isn't it only a matter of time before political correctness takes over completely?

It’s hard to measure the power of the more intellectually bankrupt wing of the social justice movement, but as best I can tell it does not seem to be getting more powerful.

According to Rasmussen, support for “political correctness” is declining in America. As we saw above, fewer and fewer people are willing to attribute black-white disparities to “racism” over time. Gallup finds that in the past decade, the percent of blacks satisfied with the way blacks are treated has gone up nearly 10% (I can’t find similar numbers for white people, but I bet they’re similar). Both white and black people are about 25% less likely to consider the justice system racially biased than 20 years ago. The percent of whites who think government should play “a major role” in helping minorities has dropped by 10 percent since 2004; for blacks, there is a similar drop of 14 percent.

The percent of people who think women have equal job opportunities to men has gone up 15% in the past nine years. Women are less likely to identify as feminists than twenty years ago, and support for affirmative action is at historic lows.

Here we see really the most encouraging combination of trends possible: actual racism, perceptions of racism, and concern about racism are all decreasing at the same time.

5.4.3.2: So how come social justice people have been making so much more noise lately?

My guess is changes in the media. The Internet allows small groups to form isolated bubbles and then fester away from the rest of society, becoming more and more extremist and paranoid and certain of themselves as their members feed upon each other in a vicious cycle.

Of course, as Reactionaries, you wouldn't *possibly* know anything about that.

At the same time, the relative anonymity of the Internet promotes bad manners and flame wars and general trollishness. It's not just that the writer is anonymous and therefore doesn't fear punishment for what he or she says. It's that their *enemy* is some nameless evil, rather than a person with a face whom they will treat as a human being.

And again at the same time, the national media has become more and more efficient at detecting outrageous events associated with some small town or some B-list celebrity and publicizing them to the entire world. This allows the hatred of the entire world to be focused on a single random person for a short period of time, which usually results in that person's life being ruined in a way that would be impossible without this media efficiency.

But these processes are at least partly nonpartisan. With a rise in extremist online social justice has also come a rise in groups that didn't even exist before, like men's rights advocates.

5.4.3.2.1: Still, isn't the fact that progressivism was responsible for this sort of zealous and hateful social justice movement is a point against it?

I identify the worst parts of the social justice movement as basically reactionary in their outlook, even though from a coalition politics point of view they have been forced to ally with progressives.

Chief in this assessment is their strong beliefs that some topics should be taboo and bowdlerized from society. In the old days, you would ban books because they talked too much about sex. In the new days, we laugh at their prudishness, but still seriously debate banning books because they are “demeaning towards women” or “trivialize rape culture”. The desire to ban books that promote different sexual norms than we ourselves promote hasn’t changed, only the particular sexual norms we are enforcing.

The same is true of race. In the old days, we would ban books that insulted the King or the upper classes. In the new days, we ban books that insult the poor, or disprivileged or disadvantaged classes. Again, the desire to ban books insulting the classes we like doesn’t change, only to which classes we afford this privilege.

Real Progressivism is Enlightenment values – like the belief that free flow of information is more important than any particular person’s desire to “cleanse” society of “unsavory” ideas. Real Reaction is the belief that free expression isn’t as important as making sure people have “the right” values. Upper-class white Reactionaries will try to enforce values protecting upper-class white people. Lower-class minority Reactionaries will try to enforce values protecting lower-class minorities. Whatever. They’re still Reactionary.

Likewise, real Progressivism is color-blind. It may be *sophisticatedly* color-blind, which involves realizing that just saying “I’m going to be color-blind now, okay?” doesn’t work, and that affirmative-action type policies may paradoxically lead to more genuinely color-blind results. But it would be unlikely to promote the idea that people should have racial pride, or that one particular race is evil and is not allowed to have racial pride. “White people should identify strongly with white culture; black people have no culture” is the upper-class white Reactionary slogan. “Black people should identify strongly with black culture; white people have no culture” is the lower-class minority Reactionary slogan. “Lots of races have culture but let’s ignore them and let individuals identify with what they personally like" is the academically-neglected but still-popular true Progressive position.

Finally, real Progressivism opposes segregation in all its forms. Upper-class white Reaction says that it's necessary to protect white people from being "polluted" by black culture like rap music. Lower-class minority Reaction says that it's necessary to stop white people from "appropriating" black culture like rap music. Either way, we get white people not allowed to listen to rap music. Progressivism is the position contrary to both: that everyone can listen to whatever music they damn well please.

The conservative nature of social justice isn't surprising if you, like me, believe the liberal/conservative divide mirrors a self-expression/survival divide – more simply, whether or not you feel safe. As society becomes more economically and politically secure, we expect it to become more liberal and progressive. But we also expect the subgroups of society that are least secure to remain conservative, and to continue to use conservative strategies to protect themselves in their unsafe environment. Those subgroups are women and minorities.

Because more liberal white people are more likely to be tolerant toward minorities and the poor, minorities and the poor are by political necessity forced to ally with liberal parties. But when we are able to separate issues out from political coalition-building and self-interest, the natural tendency of economically and physically insecure minorities to be more socially conservative shows itself. Black people are more religious, more likely to support amendments banning gay marriage, and more likely to oppose stem cell research, abortion, and out of wedlock births.

If you do not like certain extreme versions of social justice, then fighting their Reactionary memes favoring poor minorities with your own Reactionary memes favoring rich whites is unlikely to work. At best you would just end up with two angry clans demanding more power for them personally; more likely financial and signaling incentives will prevent rich whites from wanting to take their own side in a conflict and everyone will just ignore you. A better strategy would be to take the moral high ground and promote Progressive memes to both sides.

5.5: Is our society hopelessly biased in favor of minorities and prejudiced against white people?

The most visible parts of society, like affirmative action and conversational norms around political correctness, are biased in favor of minorities and against white people. But this is intended to counter less visible parts of society, which are biased in favor of white people and against minorities. Whether this gambit works is anyone’s guess. See An analysis of the formalist account of power structures in democratic societies for a more careful evaluation of this claim.

5.6: One particularly annoying politically correct idea is the demand that everyone feel guilty about colonialism. Colonialism helped industrialize the developing world. Wasn’t the Progressive attempt to “help” the developing world through enforced decolonization and self-rule actually a big step backwards?

There are a couple of studies on this question, but all have their issues. A particular problem in the comparison of colonized to uncolonized countries is the possibility that more prosperous countries would be more likely to attract colonization *and* more likely to successfully resist potential colonizers. This makes an attempt to formally compare colonized with never-colonized countries directly nearly impossible.

I am least dissatisfied with Sylwester 2005, which compares colonial countries before, during, and after decolonization. It finds that:

> There was no decrease in growth [for newly independent countries] relative to the alternative of remaining a colony. The reason why decolonizers exhibited lower growth than did those not concurrently undergoing a political change is that decolonizers grew slower than did nascent countries. These results provide evidence against the claim that this type of political transition caused lower growth than experienced previously. There is no evidence of transitional costs.
>
> The paper also finds that previously independent countries grew faster than did the existing colonies. Whether or not a region is independent or controlled by an external power appears important for growth outcomes”

In other words, countries grew faster after independence than they did as a colony. This provides some support for the leftist idea that colonial powers drained more resources than they introduced, at least towards the end of the colonial age.

5.6.1: Forget economics, then. Wasn’t decolonization a human rights disaster, considering all the civil wars and coups and mismanagement in former colonies that could have been prevented by a competent colonial government?

Everyone from every side of the political spectrum agrees decolonization could have been handled better. It might be that no decolonization at all would have been better than decolonization the way the Great Powers historically went about it. And it’s hard to excuse all the civil wars and mismanagement that caused.

On the other hand, the colonial era wasn’t exactly free of bloody wars either. Colonial wars included the Mahdist War (100,000 deaths), the Algerian Revolution (500,000 – 1.5 million deaths), the Rif War (70,000 deaths), the Italian-Ethiopian War (500,000 deaths), the Mau Mau Rebellion (20,000 deaths), Mozambique War Of Independence (80,000 deaths), Angolan War of Independence (50,000 deaths), the Herero Genocide (100,000 deaths), the Java Wars (200,000 deaths), Sepoy Mutiny (~100,000 deaths), the Mad Mullah Jihad (100,000 deaths, but on the brighter side, an awesome name) Philippine-American War (220,000 deaths), First Indochina War (200,000 deaths), Aceh War (100,000 deaths) et cetera, et cetera, et cetera.

If we don’t limit ourselves to just wars, and include famines, genocides, and general mismanagement, we can add Congo Free State (8 million deaths), genocide of Brazilian Indians (?200,000 deaths), forced labor in Portuguese colonies (250,000 deaths), forced labor in French colonies (200,000 deaths), Italian colonial genocide in Libya (125,000 deaths), French colonization of Algeria (500,000 deaths, European eradication of Native Americans (350,000 deaths), and the Australian and New Zealander eradication of aborigines and Maori (440,000 deaths). If we are willing to count famines worsened by colonial mismanagement we can go almost arbitrarily high, 20 million deaths or more.

It is certainly possible to *imagine* a wise and paternalistic colonial government coming in, cleaning up after native misrule, and introducing things like sanitation and industrialization. But that's not what happened. It's not fair to compare an imaginary ideal version of one policy with the real-world version of another.

5.6.1.1: Weren't a lot of those colonial wars and human rights abuses actually caused by demotism and Progressivism? If people hadn't revolted against their colonial masters, there wouldn't have been these bloody colonial revolts.

Not a straw man!

The first answer is that even if we accept this weird premise, there are still hundreds of colonial atrocities that do not stand excused. Many of the above conflicts occurred during original colonial invasions, and a tendency to resist those hardly requires demotism. Others were simple genocides, during which resistance was minimal.

But let's not accept the premise. I admit placing blame is complicated. To give just one example, thousands of homosexuals were killed in Nazi Germany. We usually blame the Nazis for this. But from a formal math point of view, it would be equally valid to blame homosexuality. After all, if not for homosexuality, those people would not have been killed, Nazis or no.

How to avoid such bizarre conclusions? One method is moral – even if both Nazism and homosexuality were to blame according to purely mathematical casual models, Nazism seems more *morally* to blame. Another method is practical-homosexuality is as old as the human race and probably not going away, so it's easier to view homosexuality as a constant and vary Nazism than it is to hold Nazism as constant and vary homosexuality.

We can apply these same methods to the colonial wars. Morally, the colonized people seemed to be morally in the right – they were sitting around trying to live their ordinary lives when people invaded and tried to turn them into forced laborers. And practically, the desire for self-rule is older and harder to root out than the colonialism. Indeed, colonialism pretty much died off after a century or two, and the desire for self-rule is stronger than ever.

Some Reactionaries would contest this hypothesis. They would say that it is only the spread of Progressive ideas that make people want to revolt against their colonial masters – that if not for the *New York Times* deliberately sowing pre-revolt memes, no one would consider this a worthwhile thing to try.

Historical counterexamples abound, but the Jewish-Roman Wars (66-135 AD) seem like a particularly good one. If they don't appeal to you for some reason, pick your own favorite example out of Wikipedia's List of revolutions and rebellions.

And as we saw above, if Progressivism is an inevitable historical reaction to rising technology and security, rather than a meme spread by the New York Times or anyone else, then saying "My scheme would have worked if not for the spread of Progressive ideas" is no more virtuous than saying "My scheme would have worked if not for the conservation of matter". Congratulations, you've found something that might have been a good idea in an alternate universe that ran on different rules.

5.6.2: Even if colonialism was historically bloody, wouldn't today's human-rights-obsessed, racism-hating era be able to sustain a type of colonialism that gives the good parts without the evil?

Yes, it’s possible that modern progressive ideals would be able to rescue colonialism. But it’s hard to imagine a nation being simultaneously progressive enough to colonize other countries wisely, but still so unprogressive that it would want to. It would have to be a country whose progressivism evolved on a path much different to our own.

5.7: Are schools are places where children get brainwashed into leftist and blame-America-first values? Are all parts of history that don’t fit with a progressive worldview whitewashed from the curriculum?

Our source here is James Donald, who for example says:

> History gets radically rewritten at ever shorter intervals, and all older history books are effectively banned. Consider, for example the ever more radical rewrites of the career of Daniel Boone, which ended with him being expelled from history altogether, and that today’s student has no idea what “The shores of Tripoli” refers to. Ninety nine percent of what students used to be taught not very long ago, is now unthinkably controversial, shocking, and disturbing…
>
> Look [these things] up in a history book written before the days of hate-America-first history. The New Century Speaker for School and College, published 1905.
>
> Of course this would require you to read old books, but old books are like kryptonite to a progressive. Since they were written by dead white males, no respectable person will read them for fear that dangerous and forbidden thoughts might contaminate his brain. Like a vampire confronted with a bible, a progressive will cringe in fear before any dangerously old book. Ever

> since 1905 or so, kids have been taught hate-america-first history.

I worry James is confusing the sign of a value with the sign of its derivative. Certainly schools are becoming *more* willing to discuss leftist issues. But are they now *disproportionately* willing to discuss them?

Let's take the example of Columbus. Modern Americans are taught not only the old history that Columbus was a brave explorer who sailed forth to boldly discover that the Earth was round, but also the new history of "yeah, but he was bad for the Indians". The feeling I got was that sure, Columbus was all nice and well, but his bold voyages paved the way for later people to settle the New World which sort of by coincidence hurt the Indians because people were squatting on their ancestral lands. This is about as far as so-called liberal schools will go, and this is probably the sort of progressivism being introduced to history classes which James is complaining about.

But actually, Columbus was…well, The Oatmeal is kind of a low-status source to link to, but I think they said this one better than I could. It starts off with :

> Upon his arrival, he demanded that the Lucayan [Indians] give his men food and gold, and allow him to have sex with their women. When the Lucayans refused, Columbus responded by ordering that their ears and noses be cut off, so that the now disfigured offenders could return to their villages and serve as a warning to others. Eventually, the natives rebelled. Columbus saw this as a perfect excuse to go to war, and with heavily armed troops and advanced weaponry, it wound up being a very short war. The natives were quickly slaughtered…there are eyewitness accounts of fallen

> Lucayan warriors being fed to hunting dogs while they were still alive, screaming and wailing in agony as the dogs feasted on their limbs and entrails.

(a commenter points out that some of its other claims are exaggerated)

As much as James may complain about how people vaguely mutter about something something Indians something on Columbus Day, I bet he didn't learn this in school. In fact despite his protestations, I bet he didn't learn very much leftist history at all in school, given that he thought Eugene V. Debs was a Supreme Court case.

One day, our school curriculum may become so leftist that the Right needs a book like A People's History of the United States or Lies My Teacher Told Me (which was created not by armchair contemplation of what society's biases *must* be, but by reading twelve actual history textbooks and spotting the actual lies in them). But that day hasn't come yet.

What is James' own evidence for a leftist bias? As far as I can tell, they're things like that US classrooms keep going on about US enslavement of black people, but never mention the (African) Barbary Pirates enslaving white Americans. But this may have less to do with liberal bias and more to do with the fact that, as far as I can tell, only 115 white Americans were ever enslaved by the Barbary Pirates (and then released a few years later), whereas about 500,000 African slaves were brought to America, kept in slavery for centuries, precipitated the bloodiest war in our country's history, and then became a racial group that makes up 12% of Americans today – over forty million people.

Oh, and actually, I *did* learn about the Barbary Pirates in history class, thank you very much. So it seems that prediction of James' has been disconfirmed. Although he seems to have thought the government shutdown might end with Tea Party members and lawmakers being shipped to concentration camps, so I imagine having his predictions disconfirmed is a pretty common occurrence for him.

I apologize for the insulting tone of this FAQ entry, but I was accused of cringing in fear before old books, and being vampire-to-Bible-level afraid to study history. That hurts.

6: Any last thoughts?

6.1: Does this mean you hate Reactionary ideas and think they have nothing to teach you?

Absolutely not. Compare to communism. The people who called themselves communists had some great ideas, like shorter workweeks and racial equality. It was just that the narrative they used as a framework for that idea – historical dialectic, workers controlling the means of production, violent revolution, destruction of capitalism, destruction of democracy – were horrible. Their ability to notice problems tended to be better than their specific policy proposals which in turn tended to be better than their flights of fancy.

I feel the same way about Reaction. Some Reactionaries are saying things about society that need to be said. A few even have good policy proposals. But couching them in a narrative that talks about the wonders of feudalism and the evils of the Cathedral and how we should replace democracy with an absolute monarch just discredits them entirely.

6.1.1: What *exactly* do you like about Reaction?

I like that they're honestly utopian. Their scathing attacks on everyone else for being utopian merely punctuate the fact, like the fire-and-brimstone preacher denouncing homosexuality whom everyone knows is secretly gay. The Reactionaries wants to throw out the extremely carefully fine-tuned machinery of modern society which evolved over several hundred years, and replace it with a bizarre Frankenstein's Monster of modern and traditional elements that they dreamed up in an armchair, which has never been tried before and which, they say, will instantly fix all social ills like crime and poverty and war.

And this is awesome. Utopianism – trying to think up amazing political systems that lie outside the local Overton Window – is very nearly a dead art. The failure of the Communists' utopian designs probably killed it – the Right made "utopianism" into a dirty word so they could use it to bludgeon the Left, and the Left turned against utopianism en masse to avoid getting bludgeoned. Right now the only two permissible dreams of a better future are a society much like our own but a little more libertarian, or a society much like our own but a little more progressive. Boring!

The more utopian ideas we have the more sources we have to draw from when trying to decide which direction our own society should go in, and the broader the discourse becomes. Reactionaries are geniuses at inventing new systems that have never been tried before and some of whose components deserve serious contemplation. And if there was a science fiction book set in Moldbug's Patchwork or Royal California, I would buy it.

6.1.1.1: But?

There are a few good things you can do with utopianism.

You can use it as a generator for ideas that become gradually adopted into the mainstream, as mentioned above. Communism was good at this – in the US, instead of starting a revolution, they just helped spark the modern labor movement, which eventually came to coexist with the rest of the economy and is now probably a useful part of the memetic ecosystem.

You can use it to start interesting intentional communities. There were a couple of communist communes within capitalist countries; some people even built phalansteries, and more modern versions like Twin Oaks are more successful. You can start a non-communal subculture, like the polyamory movement. If you happen to have a free land, you start a country or subnational government – it worked for the early American settlers, and it may yet work for seasteaders. The Free State Project is another noble goal along these lines.

But until it works in an intentional community or something, trying to push it on everyone else seems premature and irresponsible.

6.1.2: If we don’t do Reaction, does that mean we’re stuck with a boring inoffensive centrist democracy forever and ever?

No. There are lots of extremely creative ideas for radical new forms of government that don’t involve any Reactionary ideas at all. The better ones are off of the right-left spectrum entirely. Futarchy is my favorite. Or we could all just go live in the Shining Garden of Kai-Raikoth.

6.2.1: Has anyone written a response or rebuttal to this FAQ?

Ohhhhhh yes.

I am indebted to Reactionary blogger Legionnaire for putting together a good list of responses to this document, which I am reproducing here with only minor aesthetic changes.

RESPONSES TO PART 1: IS EVERYTHING GETTING WORSE?

Foseti – An Anti-Reaction FAQ

Xenosystems – The Decline Frame

Jim – Anti-Anti-Reactionary FAQ Part 2: Crime

More Right (Michael Anissimov) – Response to Anti-Reactionary FAQ, Lightning Round, Part 1

RESPONSES TO PART 2: ARE TRADITIONAL MONARCHIES BETTER PLACES TO LIVE?

Jim – Anti-Anti-Reactionary FAQ Part 1: Terror And Mass Murder (this limited its complaint to a single example and seemed quite fair, so I have since removed that example from this document)

Jim – Anti-Anti-Reactionary FAQ Part 3: Freedom And Monarchy

More Right (Michael Anissimov) – Response To Anti-Reactionary FAQ Part 2: Austrian Edition

RESPONSES TO PART 3: WHAT IS PROGRESS?

Jim – Progress

Jim – Anti-Anti-Reactionary FAQ Part 4: Ever Leftwards Movement

Anarcho-Papist – The Theory Of Demotist Singularity

Habitable Worlds – The Motives Of Social Policy

RESPONSES TO PART 4: SHOULD A COUNTRY BE RULED AS A JOINT-STOCK CORPORATION?

Anarcho-Papist – The Informal Systems Critique of Formalism

RESPONSES TO PART 5: ARE MODERN IDEAS ABOUT RACE AND GENDER WRONG-HEADED AND DANGEROUS?

Anarcho-Papist – On The Opposition To Sluttiness, Among Other Things

Free Northerner – Sex: A Response To Scott Alexander

Jim – The Anti-Anti-Reactionary FAQ: Sluts

MISCELLANEOUS RESPONSES

Nick Steves – Shots Across The Bow

Suntzuanime – Comment On Anti-Reactionary FAQ

I've only managed to read about 50% of these so far, but of the ones I have read, I am especially impressed with Anissimov's Lightning Round Part 1 and Free Northerner's post on sex issues as well-argued and pretty comprehensive critiques.

I will continue to update based on his list as a definitive resource, but if you've written something and want on here, post in the comments of this thread or email me and I will *eventually* get you up. This is likely to update very irregularly.

Some preliminary responses to responses to the Anti-Reactionary FAQ

I’m still sorting through the many comments I got on the Anti-Reactionary FAQ. Many of the criticisms were excellent. I would like to update the FAQ to reflect them eventually, but that would be a never-ending process, and disrupt some people’s project of creating a counter-FAQ – it’s harder to hit a moving target. But I want to respond to a few broad categories quickly, especially since they may help reveal more of a coherent picture than the question-by-question style of the FAQ.

I’m not mentioning major factual errors here. Some people pointed out a few good ones – for example, the text of the Cyrus Cylinder I used seems to derive from a 19th century forgery. These have already been corrected where they occurred.

I. Fifty Year Trends

One of the better critiques is that I focus too much on fifty-or-so-year trends – which often show improvement in key social indicators – rather than comparing our own society to preindustrial village society. This is an important point, because I’ve been accused of confusing Reactionaries and garden-variety conservatives before, and go-back-thirty-years versus go-back-to-preindustrial-villages seems like part of that distinction.

My defense is that a lot of Reactionaries *have* made the claim that fifty years ago was better than today, and the counterclaim is worth defending (it’s also highly relevant for whether our civilization is on the verge of imminent collapse). Still, I accept the point that, even though divorce, suicide and crime rates are declining at present, they’re probably much higher than in 1100s rural Austria or whatever.

This raises a key question that was too broad (and which I understood too late) to get in the FAQ. If these social indicators got worse from 1100 to now, but have been getting better for the past thirty years since good statistics are available, at what point did the trends change direction? And why?

Let me ask a stronger version of that question: between about the sixties and the eighties, crime, divorce, suicide, abortion, promiscuity, drug use, STDs, and unhappiness were all trending upwards. Between the eighties and the present, all of those things have been trending downwards. Why?

I have seen very little discussion of such an important question. Probably the best research on this question has been that linking lead to crime rate. Lead is known to decrease intelligence and increase impulsivity, and its levels peaked and then declined at about the right time. Some research has tried to link lead to teen pregnancy as well. It could conceivably hold for all the indicators involved except for divorce rates (which hit at the wrong time). Other research suggests the baby boom (more concentrated youth creates a “youth culture” that promotes rebellion), the explosive increase in the drug scene around the 60s, or some kind of disruption to norms about sex and the family caused by the availability of STD-curing antibiotics and easy contraception.

Whatever the reason, it seems like if the current trend is down, we should panic before it destroys us all; but if the current trend is up, one can argue things are already improving, we're doing something right, and we shouldn't jinx it.

II. The Upper Class/Underclass Gap

Some people criticized my defense of modern sexual norms by saying I was putting too much emphasis on how they have worked pretty well for upper-class suburban whites, but am failing to consider the dysfunctionality of modern culture among lower-class urbanites and the uneducated.

I agree that problems with divorce, out-of-wedlock birth, STDs, et cetera are worse among lower class and uneducated people.

Assuming we want to correct that by enforcing some kind of value system on them, we have two models. Number one, the value system of 1100s rural Austria. Number two, the value system of the wealthy people in the suburbs just down the street.

1100s rural Austria had a lot of things going for it in terms of enforcement of sexual norms. Due to land inheritance and dowries, parents had extremely strong control over their children's future and disinheritance was nearly a death sentence. There were no drugs besides alcohol, no contraception or abortion to save you if you slipped up, and no penicillin to cure any STDs you might catch. People of all ages worked all day, and although work was broken by festivals these were community gatherings rather than "idle time". There were no TV shows or novels, and people were socialized mostly at church. But probably the most important thing was that reputation was almost inescapable

– most people spent their entire lives in the same small village interacting with the same small group of people.

It’s hard to see how to export that kind of environment to the modern underclass. Pol Pot tried something like it, but it didn’t go very well. That leaves the upper-class people down the street who are doing pretty well in exactly the same technological environment. Can we export *their* values?

It’s possible the answer is no. It’s possible that progressive values are the sort of thing high-IQ, low-impulsivity people are naturally good at, but which low-IQ, high-impulsivity people naturally fail. But two thoughts on that.

First, IQ and impulsivity are not absolutely fixed. I am the furthest thing from one of those wishy-washy people who say everything is socially determined, but even within most things being biologically determined there is a lot of wiggle room. The iodization of salt and the banning of lead have cut a big chunk into the rich-poor IQ/impulsivity gaps already. There’s a lot more room to improve these kinds of factors, and a lot of progressive programs are trying to do exactly that. As technology increases, we’ll hopefully be able to do more. It’s worth noting that poor kids are the only ones still exposed to a lot of lead – perhaps that has something to do with it?

But second, the same values that work so well for upper-class people might also work well for lower-class people, if only they had them. This is one reason I included 5.4.3.2.1 in the FAQ (the other was because it’s a cool number). Many people criticized me for calling some lower-class values ‘reactionary’, as if I was just playing with definitions. But a corollary of that might be that if progressive values work for the upper-class, maybe they would work for the lower-class as well if only they would adopt them.

Here’s where I’m coming from. Suppose that culture experienced a technological/economic shock with the Industrial Revolution and Demographic Transition. This somehow caused the rise in divorce, suicide, drug use, et cetera from the 60s to 80s cited above. Gradually, the shock wore off and everyone started to adjust. They developed new values and institutions that contained the damage and operated well in the new conditions. The gradual development spread of these new norms caused the decrease in all those social ills noted from the 80s to the present.

The upper classes are “early adopters” in terms of technological and material change. They’re going to be the first people trying contraception, the first people trying new fancy leaded gasoline, the first people sending their daughters off to college.

The lower classes are late adopters. They’re going to go through the same thing the upper classes did, but they’re going to be time-delayed. If this is true, they’re going to have to learn the same things the upper classes did about the new norms and institutions, at which point they, like the upper classes, will reverse the negative trends and start a positive trend.

Right now the lower class is kind of time-frozen, because the natural upward trend of real wages has been arrested for the past few decades and I expect values change to be closely associated with change in material conditions. I would expect if real wages continued to increase, then the lower class would gradually undergo the same kind of institutional changes the upper class went through, adjust to their new conditions, and end up with the same progressive values as the upper class.

Or maybe I’m totally wrong. But it seems making the New Orleans ghetto like the Boston suburbs should in any case be easier than making the New Orleans ghetto like an 11th century Austrian rural village. I haven’t even heard good plans about how to do the latter. All I’ve heard is “let’s put a king in charge”. As if system of government had anything more to do with 11th century Austrian purity norms than speaking Middle High German did.

III. Sluttiness and Consequentialism

I got a lot of flak over these two sentences: “Contraception prevents out of wedlock births. Protection and antibiotics prevent STDs. So the old reasons [for disapproving of sluttiness] no longer hold.”

The flak was mostly for saying contraception preventing out-of-wedlock births and STDs, when in fact there are still a lot of both around, and out-of-wedlock births are at historically high levels. Many people brought up point II above – that these problems have only been “solved” for upper-class suburban elites.

First a boring objection. Statistics on out-of-wedlock births include both unintentional pregnancies and happily cohabiting couples deciding to have children. Unintentional pregnancies alone are harder to measure, but seem to follow the declining-since-1980s trend mentioned above. Most STDs do as well. It depends on exactly which disease you pick – syphilis is at historic lows, but chlamydia is a growing problem.

But to me the more important objection is that even if true, *this isn’t Reactionaries’ real objection*. There is a serious debate going on in the Reactosphere about whether a self-respecting man can marry a slut. If the above were the

Reactionaries' real objection, the answer should be obvious – ask her if she has any illegitimate kids, ask her to take an STD test (as a gentleman, you can offer to take one yourself as well), and then if she has no kids or STDs, no problem.

Given that this obvious solution hasn't caught on, I'm going to continue thinking Reactionaries' problems with sluttiness transcend the possibility of kids and STDs, and my argument in 5.1.1 holds.

IV. Equality of Opportunity vs. Equality of Results

Before complaining about this part, *please* read at least the first half of Social Justice For The Highly Demanding of Rigor.

V. Tone Arguments

Not intentional! Or rather, the only instances of snarkyness or meanness that were intentional were the sentence at the bottom of 1.1 and the entirety of 5.7. Konkvistador recommended comparing this to my non-libertarian FAQ and I agree. I just have a hard time writing about things without making wisecracks. Let he who is without sin cast the first stone.

I only write these sorts of things about philosophies I have a lot of respect for. I have never written an anti-creationist FAQ or anti-homeopathy FAQ. I have strong libertarian sympathies and I have strong reactionary sympathies, and that is the main reason I get very annoyed when I hear people I otherwise sympathize with saying things I think are dumb and going what I feel is too far. I thought I made this clear in Section 6.1.

VI. The Cathedral

I keep being accused of not knowing what the Cathedral is. Perhaps these accusations are true. I will admit that the offhanded references to “Ivy League professors meeting under the full moon” are snark (see above), but perhaps even when I try to be serious I don’t get it quite right.

But let me try to sidestep that critique by pointing out a serious flaw in the Moldbug post that inspired the entire Cathedral concept.

Moldbug notes that there are three reasons for agreement. One, everyone agrees because they are being forced to agree – for example, the Spanish Inquisition rooting out heretics. Two, everyone agrees because the truth is obvious – for example, people agreeing the sky is blue. Three, everyone agrees because of a self-organizing consensus of social norms – for example, women going around topless in public is unacceptable in Western society.

He asks why there is so much agreement among Ivy League colleges on their particular brand of liberalism. He throws out the first possibility – there is no obvious Inquisition. He throws out the second, because he says the sheer *degree* of their agreement is implausible – honest truth-seekers sometimes differ, but there is minimal difference between for example Harvard and Yale’s African Studies department. This leaves the third possibility, which he calls the Cathedral:

> In a [society with true intellectual freedom], for example, we should see intellectual inhomogeneities between competing institutions. Harvard and Yale should mostly agree, because reality is one thing. So should the New York Times and the Washington Post. But there will always be sclerosis, stagnation, drift. Competition, not just among ideas but among

> institutions, is essential to the Popperian ideal. We should see these institutions drift away from reality. And we should see the marketplace of ideas punish them when they do, and reward those which do not.
>
> Do you see this? Because I sure don't. What I see is a synopsis.
>
> From my perspective, not just Harvard and Yale, but in fact all major American universities in the Western world, offer exactly the same intellectual product. Which institution is more to the left, for example? Harvard, or Yale? You can pick any two mainstream universities, and you will not be able to answer this question. It's a sort of intellectual peloton.

This seemed suspicious ever since I first read it, and I've since placed my finger on exactly why.

What percent of Harvard professors do you think prefer wine to beer? What about Yale professors? Now what about Columbia professors?

I don't have a clue what the actual numbers are, but I bet all three colleges are within a few percentage points of each other. Why? Because large samples drawn randomly from the same population tend to converge.

Moldbug's point makes sense only if there is a single top-down authority deciding on the politics of Harvard, and then another deciding for Yale, and then a third deciding for Columbia. But if the schools are just selecting professors based on number of publications or something, then the differences in individual professors' opinions will be drowned out by the Law of Large Numbers, and we would expect the "consensus position" of each school, insofar as one exists, to be similar.

Why is the distribution of professors drawn from so liberal to begin with? This is another one of those things we have data about.

List of Fictional Drugs Banned By The FDA

PROTOZOSIN:
PROFILE: Alpha blocker notably effective against Pre-Traumatic Stress Disorder
BANNED BECAUSE: Banned on request of DARPA, who are trying to use epidemiology of Pre-Traumatic Stress Disorder to predict the course of future conflicts. Lobbying to overturn ban by "Future Veterans of the Pakistani War" thus far ineffective.

GEONEXPERINE:
PROFILE: A highly effective painkiller with a half-life of about twenty-four hours. Notable for its quick compensatory upregulation of pain receptors and strong rebound effect: after its effect wears off, pain typically returns twice as bad as before. This pain can generally only be controlled with a second dose of geonexperine, but when this dose is exhausted pain returns at four times its original level, which of course requires a third dose of geonexperine on humanitarian grounds, and so on.
BANNED BECAUSE: Eventually responsible for the vast majority of all human suffering.

TEVROMATIN:
PROFILE: Chemotherapy adjuvant specifically designed for glioblastomas of neuronal origin. By mimicking natural neural differentiation factors, it causes these tumors to regress from resilient high-grade neuroblasts towards more typical neurons, making them easy targets for stronger chemotherapeutic agents.
BANNED BECAUSE: During differentiation process, malignant nerve cells form connections to healthy nerve cells and to

each other. As a result, tumor forms a functioning neural network effectively "telepathically" connected to healthy brain. Patients report feelings of overwhelming guilt as tumor accesses patient's memories and emotions and realizes its role as a parasitic cancer, followed by its utter terror as it realizes it is about to be killed. Many patients refuse to continue with chemotherapy regimen; those who continue make a complete physical recovery but are psychologically scarred for life as they experience every moment of the tumor's death as if it were their own.

GABATIMOLINE:

PROFILE: A gabanergic agent attached to a thiotimoline moiety, this anticonvulsant has the useful property of taking effect approximately sixty minutes *before* it is administered. Designed by Merck as a treatment for seizures that might occur in rural areas without good access to medical care, first-line responders are taught to form a strong resolution to administer gabatimoline as soon as they reach a hospital, and as a result the seizure stops immediately. Exhaustive research on this drug has determined that it operates probabilistically: the gabatimoline will operate at half-strength if there is a 50% chance it will be administered within 60 minutes, at 10% strength if there is only a 10% probability it will be administered, and so on.

BANNED BECAUSE: Smart-alecks noted that any given person at any given time has a nonzero chance of receiving gabatimoline in the next sixty minutes, if only as a bizarre mistake or practical joke. Therefore, microdoses of the drug are operating in everyone at all times. Resulting health panic ended in complete ban on its manufacture.

XAOMORPHINE:

PROFILE: Developed as a substitute for morphine without the addiction potential, this drug contains both morphine and a few micrograms of xaonin, a monoclonal antibody that disrupts the brain’s reward attribution system, breaking the mental link between the administration of the drug and the high it provides. As a result, patients report no craving for xaomorphine.

BANNED BECAUSE: Broken reward system flails about trying to attribute positive experience, leading to bizarre addictions to contingent features of xaomorphine administration. Some patients develop an addiction to being in hospital emergency rooms, others to having people sticking needles in their veins, still others to the clothes they were wearing at the time they received xaomorphine. The final straw was a much publicized story of a patient kidnapping the nurse who administered xaomorphine to him so he could have her near him always – and a second story, a few days later, of a woman who was addicted to fracturing her hip after she received xaomorphine for a hip fracture once. The drug was banned two weeks later.

ZORNINONE:

PROFILE: A moderately effective sleeping pill notably for its many possible routes of administration, including oral, subcutaneous, and mental. This latter route is particularly interesting, as patients can get the full somnolent effect of the drug merely by concentrating very hard on its name, the appearance of the pill, the ball-and-stick version of its chemical structure, and the intended dose.

BANNED BECAUSE: No, paradoxical effects of thought suppression did *not* result in unintentional doses of the drug while people were eg driving down highways, as more concentration seems to be required. However, concentrating on the drug just before sleep led to a disproportionate number of dreams about the drug, some of those were nightmares about taking overdoses, and sure enough those overdoses landed people in the very real waking-world emergency room.

HABICILLIN:
PROFILE: Beta-lactam antibiotic notably effective against so-called “intelligent bacteria” species *streptococcus sapiens*.
BANNED BECAUSE: In Treaty of Atlanta, representatives of humans and *s. sapiens* agreed that *s. sapiens* would stop infecting humans in exchange for “humanitarian aid” of 200,000 plates of agar per year. Habicillin banned by all human countries as a show of good faith.

LUCIPERIDONE:
PROFILE: Third-generation antipsychotic. Unlike other antipsychotics, which control hallucinations but are almost ineffective against delusions, luciperidone cures all delusions with 100% effectiveness.
BANNED BECAUSE: Cures *all* delusions with 100% effectiveness. Cf. Terry Pratchett.

The Poor You Will Always Have With You

I’m gradually reading through responses to the Anti-Reactionary FAQ, but I’ll take a moment to respond to this excellent and well-argued post from Habitable Worlds in particular because it points out an especially deep disagreement.

Scharlach from Habitable Worlds objects to my point 3.1.1, which claims that progressive ideals aren’t particularly novel or modern because classical Rome shared many of the policies we most associate with progressivism. I mention welfare, strikes agitating greater rights for the poor, multiculturalism, religious syncretism, sexual libertinism, and utopianism.

Scharlach disagrees. He first points out that classical Roman “strikes” were not about greater rights for the “poor”, per se, but for plebians, a class of non-nobles that actually included some very wealthy people. I accept his clarification, but I would add that modern progressive movements are happy to conflate “class made up of disproportionately poor people” with “poor people” as well, whether we are talking about the unemployed, inner city youth, minorities, high school dropouts, inhabitants of the Third World, or whatever. Heck, modern progressivism calls women a “minority” even though they make up 51% of the population just because it is a convenient way to lampshade their less privileged status. So I don’t think it’s especially unprogressive that “more rights for plebians” was the classical Roman rallying cry, rather than “more rights for the poor” per se.

But the crux of his objection is more philosophical:

But the question is: do these seemingly “progressive” policies stem from what today we would consider progressivism? Do they have anything to do with “social justice”? We should remember that when looking back at history, curious similarities arise, but they do so at incongruous joints, and their existence may not signify anything but the fact that large-scale political ecologies have limited practical expressions. Think of it this way: A society whose political discourse and ideals sanction welfare to the poor because it is believed that the underclass is genetically inferior, incapable of taking care of itself, and might revolt if not given enough food . . . that’s a very different society from one whose political ideals sanction welfare because it is believed the poor have a right to good living standards or that the poor deserve welfare because it re-distributes goods rightly theirs but taken from them through an oppressive economic system.

Contemporary progressive policies emerge from ideals and discourses about morality, justice, oppression, and rights. The poor (especially the dark-skinned poor) deserve the welfare they get; it is theirs by Constitutional right. It is a moral and political imperative not to take away the welfare they receive and to give them more if possible. Progressives actively try to alleviate the shame once associated with receiving welfare. Pointing out that the poor in America have it pretty good is a distinctly right-wing thing to do. ”Food stamps” are now “EBT cards” that look and function like debit cards. Medicaid patients sit in the same waiting rooms as patients paying high insurance premiums, and you can’t tell the difference. (Well, you can, but . . .) Welfare in America has become a right, a moral imperative, a matter of justice and just desserts, a thing

> that brings no shame, a thing to be proud of, a thing to demand, a thing to stand up for…
>
> So Scott Alexander is correct that social policies in ancient Rome look similar to contemporary progressive welfare policies. But were the motives the same? Did the poor and the plebians get free or reduced-cost corn, grain, wine, and olive oil . . . . because they deserved it? because it was theirs by moral and legal right? because it was a matter of social justice?
>
> I’m not a classicist, so I’m willing to be corrected on this, but as near as I can tell, the Roman dole was wrapped up in discourses about a) the might and wealth of Rome and b) goddess worship. Welfare policies in ancient Rome were built upon very different ideals and emerged from very different motives than contemporary progressivism’s welfare policies. Nowhere have I been able to fine a discussion of the Roman congiarium in terms of rights or justice. The dole was there because it made the emperor more popular and demonstrated the wealth of Rome to the people. What’s more, the dole was personified as Annona, a goddess to be worshiped and thanked. Scott Alexander even recognizes this difference in motive when he says that ancient Romans “worshiped a goddess of food stamps.”
>
> Indeed they did. And that’s the whole point. When was the last time you heard welfare policies discussed in terms of worshipful gratitude, mercy, and thankfulness? If that were the discourse surrounding welfare policy, America would be a very different country. It seems that Roman welfare and American welfare are as different from one another as Jubilee is from abolitionism.

I will agree that the Romans used different philosophical justifications for their welfare state than do moderns, but before discussing this, a lengthy and kind of pointless also-not-a-classicist digression on why the difference may not be as big as Scharlach suggests.

If the essay is trying to compare the grateful Roman poor and the entitled, demanding modern poor, I propose that the Roman recipients of the *annona* were as entitled and demanding as any modern. Ancient Roman leaders automatically assumed any hiccup in the flow of free grain would lead to riots, and their assumption was justified. You may for example read the section on Roman food riots here. Particular high points are the riots of 22 BC, during which rioters threatened to burn the senators alive if they didn't produce enough free grain, and the riots of 190 AD, when Papirius Dionysius, the prefect in charge of the grain supply, accused political enemy Marcus Aurelius Cleander of threatening it – the disturbance ended when the Emperor Commodus killed Cleander and his son and threw their heads out to the angry mob (which instantly calmed down and dispersed).

Or the essay may be trying to compare a Roman attitude of giving small strategic grants of welfare to the worthy with a modern attitude that everyone deserves as much welfare as they want at all times regardless of situation or else their human rights are violated. But here, too, I do not think the distinction is as great as is claimed. 83% of Americans believe people on welfare should be required to work, and only 7% oppose such a requirement. 69% believe that there are too many people on welfare and the criteria need to be stricter, compared to only 24% who believe the opposite. People who want welfare benefits need to jump through various bureaucratic hoops (some of which are actually kind of stupid) and usually receive them only for a limited amount of time.

(this interpretation would remind me of my frequent complaint that some reactionaries say “X is an unquestionable dogma of our modern society” when they mean “I heard about a college professor who believes X”.)

So much for our pointless digression. Scharlach probably means something more like “Ancient Rome didn’t have modern concepts of human rights and social justice.” I agree with this. I just don’t think it matters.

I assume Scharlach read my FAQ part 3.3, where I claim that progressive values are closely linked to urbanization and technological/economic growth. But he may not have read my We Wrestle Not With Flesh And Blood…, so I’m worried he might have interpreted me in 3.3 as claiming something like:

Urbanization + Growth -> Progressive Values -> Social Change

If that had been my thesis, then it would indeed be relevant that the ancient Romans didn’t have our version of progressive values. Their social change would be a coincidence, unrelated to ours since it missed the crucial middle step that determined the shape our social change would take.

But I’m not proposing that model. I’m proposing one that looks more like this:

Urbanization + Growth -> Social Change -> Progressive Values

(really the “social change” node should be called “pressure for social change”, and it and the “progressive values” node should have little circular arrows both pointing at each other, but let’s keep it simple)

Let me give an example of what I mean.

A 25th century historian, looking back at our own age, might notice two things. She would notice that suddenly, around the end of the 20th century, everyone started getting very fat. And she would notice that suddenly, around the end of the 20th century, the “fat acceptance movement” started to become significant. She might conclude, very rationally, that some people started a fat acceptance movement, it was successful, and so everyone became very fat.

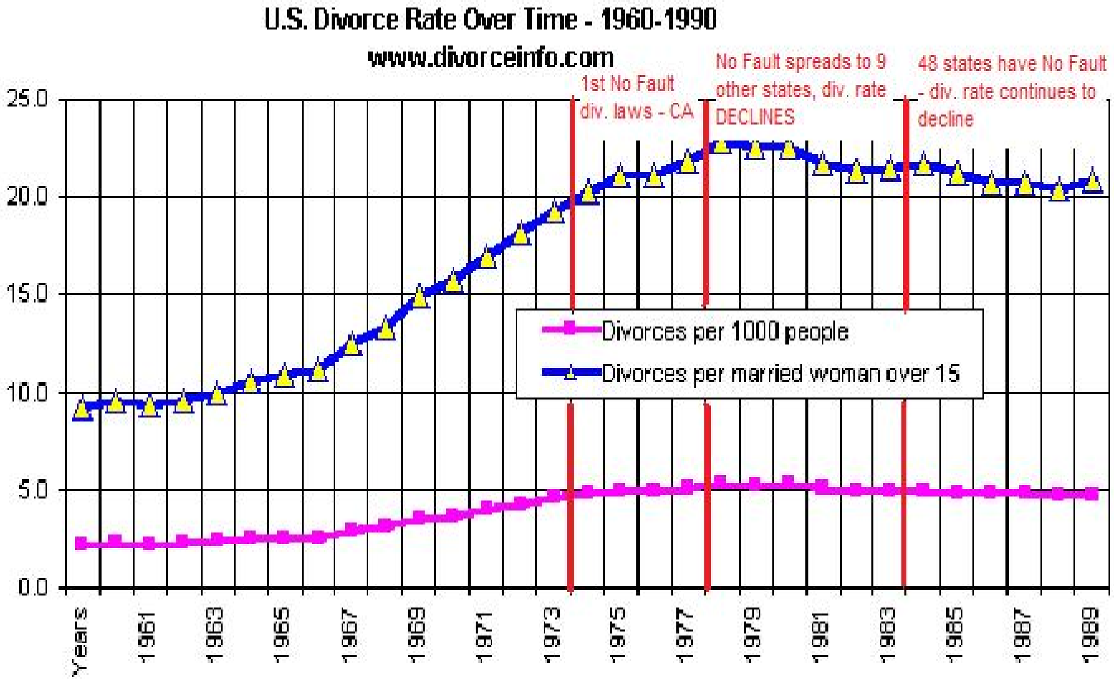
With clearer knowledge of our era, we know better. We know that people started getting fat for, uh, reasons. It seems to have a lot to do with the greater availability and better taste of fatty, sugary foods. It might also have to do with complicated biological reasons like hormone disrupters in our plastics. But we have excellent evidence it’s not because of the fat acceptance movement, which started long after obesity rates began to increase. If we really needed to prove it, we could investigate whether obesity is more common in populations with good access to fat acceptance memes (like, uh, Wal-Mart shoppers and American Samoans).

To us early-21st century-ites, it’s pretty clear why the fat acceptance movement started now. Its natural demographic is fat people, there are more fat people around to support it, they feel like they have strength in numbers. and non-fat people are having trouble stigmatizing fat people because it’s much harder to stigmatize a large group than a small group (no pun intended).

Does this have any relevance for the sort of thing reactionaries talk about? Yes. Let’s look at divorce.

From a historical perspective, no-fault divorce was legalized in the early 1970s, and divorce rates were skyrocketing in the early 1970s. It is *incredibly* tempting to want to attribute skyrocketing divorce rates to easy-access no-fault divorce.

It’s also wrong. From an excellent article I entirely recommend:



Just from the graph it should be clear how little no-fault divorce mattered, but if you need more formal research it has certainly been done. Even the conservative Institute For Marriage and Public Policy admits in its review article on the subject that “divorce law is not the major cause of the increase in divorce over the last fifty years”, and that even the small bump from no-fault provisions “while sustained for a number of years, eventually fades and the divorce rate moves back to trend”.

I’d guess that the explanation for why skyrocketing divorce rates and no-fault divorce both happened in the early 70s is a lot like the explanation for why skyrocketing obesity rates and fat acceptance both happened in the early 2000s. Lots of people started getting divorced. Under older, stricter divorce laws, this required couples who wanted divorces to manufacture some bogus complaint with the help of lawyers, an embarrassing and expensive process. Eventually the number of people divorcing or wanting to divorce became sufficiently large to form a good political lobby, and the people not involved in the divorce process couldn't keep stigmatizing divorcees because there were too many of them for it to be easy or convenient. So the divorce lobby won and passed no-fault divorce laws.

I don't deny that sometimes these ideological movements and the laws they pass have some effect, like the small, quickly fading effect of divorce laws mentioned in the quote above. That's why I wanted little circular arrows between "Social Change" and "Progressive Values" above. I'm just saying these effects are small and not particularly interesting. They're the tail wagging the dog.

And I don't deny that the progressive movement pushing a social change often exists before the social change does. If 100 years from now the existence of vat-grown meat causes all factory farming to shut down, no doubt PETA will claim victory. But just because PETA pushed for the event, and then the event happened, doesn't mean PETA was the main cause. At best, they kept pushing but it was only the technological change that helped them gain power and respect and enact their positions. At worst, if they didn't exist then within ten minutes of the invention of vat-grown meat some other group would have sprung up to accept the easy moral victory it provided.

So let's get back to Rome.

Scharlach points out that the value system associated with Roman welfare was different from the value system associated with our own welfare system.

Ancient Rome had a population of about a million people crowded together, a government vulnerable to the mob, and resources to spare. I propose those situations will, more often than not, inspire a welfare system. They did it in ancient Rome, and they're doing it in modern DC.

According to legend, Frederick the Great declared of his conquests: "I will begin by taking. I shall find scholars later to demonstrate my perfect right" (okay, Reactionaries, I will admit Frederick the Great was hella cool). If Frederick was in the welfare business, he might have said "I will begin by giving welfare. Later, I will find scholars to come up with a philosophy supporting welfare."

And just as any historical account of why Frederick conquered new territories should focus on his self-interested goals rather than on whatever justifications his scholars later cooked up, so an account of why we give welfare should focus on the economic, material, and technological conditions that inspire it, rather than fretting over how one society talked about the goddess Annona and another talked about social justice. I'm sure if Frederick conquered both classical Rome and 21st-century America, his Roman supporters would declare he was following the will of Jupiter, and his American supporters would declare he was trying to help disprivileged minorities. It would be the historian's job to see through that (and also to sort out what I expect would be a very confusing timeline of Frederick's life).

Which brings us back to Rome one last time. I didn't discuss the Roman welfare state in isolation. I mentioned it in the context of Rome being surprisingly progressive in a lot of other ways – its plebian "strikes", its multiculturalism, its religious syncretism, its loose sexual morals.

If the resemblance between Roman and modern welfare systems is a mere coincidence, then we have to add a striking number of other coincidences to the list. Eventually the conjunction of all these coincidences starts to look unlikely.

But there is a neat explanation for all of them. States that are militarily secure, economically advanced, multicultural, and urbanized tend to adopt progressive policies (here I am confusingly lumping some values like multiculturalism in as policies, but you know what I mean). Ancient Rome and modern America are both militarily secure, economically advanced, multicultural, and urbanized. In between stand a bunch of countries the Reactionaries like to talk about like the Holy Roman Empire, which were not militarily secure, economically primitive, monocultural, and more rural. Those countries didn’t have progressive policies or values.

The original question was whether ancient Rome could be called a progressive society. I say it was. Scharlach objects that it wasn’t, because it didn’t have the particular brand of progressive philosophy we do today. But I respond that the philosophy is irrelevant to what we presumably care about – social policies and social outcomes. Policies (like welfare) and outcomes (like the existence of a large class of welfare-dependent poor) were the same in classical Rome and modern America, and for the same reasons. Therefore, it is correct and useful to call classical Rome an early progressive society, though with the obvious caveat that it did not go as far in that direction as our own.

The Witching Hour

On an ordinary evening, Tal Aivon was lively and pleasant. The collection of longhouses and yurts within its tall brick walls shone bright with kerosene – not just torches, real kerosene – and its communal meeting area was noisy with conversation and song. The children would be playing their games, and on the eves of holy days the Lorekeepers would chant their stories of the Lost World, accompanied by lyres and the town's one decaying gyitar.

Tonight, though, a pall lay on Tal Aivon. The six gates of its tall brick walls were barred and shut, and foreboding warriors dressed in odd combinations of Kevlar and steel armor stood just within them, brandishing their swords. Families locked themselves in their yurts and longhouses, huddled around little kerosene lanterns. In the temple, the priests knelt before the stone idols of St. Christ and St. Mahomet, chanting plaintive prayers for protection.

“I still don't understand,” Meical Dorn complained, from inside the longest longhouse “what this is all about. “None of the wildlings are anywhere nearby – I should know, I've came through two hundred miles of forest to get here – and the only three towns in this area are at peace with you. In Great Rabda, even an impending attack couldn't make us cower inside like this. I have half a mind to think there's something you're not telling me, Fin. Something that might…threaten our deal.”

Fin Lerisas, Chief Lorekeeper for Tal Aivon, sighed. “Nothing that would threaten our deal, Meical. Great Rabda has gold. We have sunblessings. Just stay here long enough for our bankers to figure out the price, and you'll have timers and mathers and lighters of your very own.”

Meical glanced longingly at the Chief Lorekeeper’s own sunblessing, a timer that stood on the shelf of his private room. 1:52 AM gleamed on its face, with an maddeningly smooth red glow unlike sunlight or moonlight or firelight. Yet Meical knew it was sunlight, or something like. He was the Lorekeeper of Great Rabda. The Lorekeepers of Tal Aivon were far wiser than he – how could they not be with the town’s close proximity to ruined Diteroi and its trove of artifacts from the Lost World – but even he knew how sunblessings worked. You took them outside and the blue tiles on their surface fed on sunlight. Then they worked various miracles. Timers would tell you the time far more precisely than any sundial – invaluable in keeping the schedule of sacred prayer decreed by St. Mahomet. Mathers would add and subtract quantities more quickly than the fastest savant. Lighters would shine at night without wood or kerosene.

Meical had no doubt that the Lorekeepers of Tal Aivon – the wisest on the Great Peninsula – knew of still other sunblessings, ones that mighty but lore-deficient Great Rabda had never heard of. He himself would be happy with anything – even the meanest timer. Of all the millions of wonders built by the Lost World, only the sunblessings still worked, and they were in fiendishly short supply. While lore-rich Tal Aivon had a timer upon each of its six gates, Great Rabda, for all its bountiful gold and grain, had not a single sunblessing to call its own. As its Lorekeeper, it would aid his status immensely if this trade mission was successful and he could bring something back to demonstrate the power of the Lost World and, incidentally, his own importance as keeper of its Lore.

But even his greed for power did not override his concern for his own safety. “I’m serious, Fin. I want to know what’s going on. I can’t deal with a city that won’t even tell me why it’s on high alert.”

Fin Lerisas, Chief Lorekeeper of Tal Aivon and wisest in ancient matters on the whole Great Peninsula, gave another sigh. “If you were not a Lorekeeper yourself, I would not sure such secrets with a foreigner. But if it threatens the deal, very well. Only know that you will be no happier with this knowledge, and that you may not sleep quite as soundly on autumn nights from now on.”

Meical gave a nod, indicating he wanted the old man to continue.

“In Great Rabda you have no sunblessings, and so you must keep the time like wildlings, by watching the course of the sun. Here in Tal Aivon we have six timers, one on each of the city gates, and so everyone down to the meanest peasant knows the time, down to the second. To most, they check the time when they enter the city, and the time when they leave the city, and they never think any more of it. We Lorekeepers are more astute, but not infinitely so. And so it was only forty years ago, in the time when my uncle Derech was Chief Lorekeeper, that we noticed” (and here his voice changed to a whisper) “that *there is something wrong with Time*.”

“The stars,” he continued “sometimes match the time as told by the timers, and sometimes they do not. At first we thought the flaw was in the heavens themselves, so perfect are the devices of the Lost World. But this so discomfited my uncle that for three months he sat in front of this very timer, handing it off to an acolyte only when he slept. And one night, his watch bore fruit.”

“What happened?” asked Meical, breathlessly.

“Time moved backwards,” said Fin.

“Impossible,” said Meical.

“It was on this very night,” said Fin. “Time, which three hundred sixty four days of the year moves only in one direction, suddenly jumped backwards. And you yourself will witness it.”

He pointed to the timer on his shelf, which now read 1:59 AM. Its red glow suddenly looked unfriendly, even eerie. Even though Meical knew it had to be sunlight at its root, it held none of the wholesomeness of the sun.

And then it changed. 1:59 turned to 1:00.

Meical gasped, and his fingers instantly formed the cross of St. Jesus and then the crescent of St. Mahomet. “Madness!” he whispered.

“Something,” said Fin, “is wrong with this night. It is not always this night – it can come as early as three days before, or as late as three days after. My uncle worked out the formula after several years. But every year, it happens. Time jumps backwards.”

“But why?” asked Meical. “Why would the gods do such a thing? Why would they break the symmetry of the True Time and the heavens?”

“That’s the worst part,” said Fin. “When I was younger, I looked over my uncle’s formula – the one for calculating the day when the time skip would happen – and found what he had missed. The day of the time skip is fixed to the seven day calendar of the Lost World. To the ancients, it would always occur on the same day of the week. Sunday. Their holy day.”

Meical felt his blood run cold. “That’s…some coincidence.”

“Perhaps,” said Fin. “But I don’t think it *is* a coincidence. The gods are just. They would not play with Time as children play with blocks, picking one up here, then putting it down far away. I think the ancients of the Lost World, the ones who could build the great glass towers, the ones who manufactured sunblessings, the ones who made Diteroi-That-Was – I think they took their magic and threw it against time, and broke it. I think they wanted to become lords of time itself.”

“But they failed,” guessed Meical.

“They created a single hour,” said Fin. “Of the nine thousand hours in a year, all but one were made by the gods, but one was made by Man. What stopped them from creating more, from creating an infinite number of hours, from becoming immortal by arresting the progression of Time? We will never know. But it is my belief that when they saw what men had done, the gods stopped them before they could do worse. Meical, I believe that is how the Lost World ended. A last ditch effort by the gods to save Time itself from the hubris of Man.”

Meical was silent. For all their wisdom, none of the Lorekeepers claimed to know how the Lost World ended. Surely the gods had pulverized it for some offense, but what sin could have been so dire as to doom those magnificent glass towers, those great black roads as smooth as water? Meical looked at the clock, gleaming 1:03 AM, and knew. Knew in his heart that Fin was right.

“There is a day in the very early springtime,” said Fin, “when an hour disappears. The gods are stingy. They would not grant the ancients their victory. What they did with that hour in springtime, I do not know. But their message is clear.”

Meical shuddered again. Like all the inhabitants of Great Rabda, he had told the time with the sun and the stars. But it had always been an approximation, not the to-the-second True Time displayed on the six gates of Tal Aivon. And so in their ignorance they had missed no fewer than two violations of Time, and it had fallen to the people of Tal Aivon alone to guard these terrible secrets.

“You ask why we extinguish our fires and pray this night. Nine thousand hours in the year were made by the gods, but one was made by Man. I cannot help but wonder what walks abroad, during the hour no god made. I cannot help but wonder what spirits awake on the anniversary of the old world’s death. When time itself stands stagnant, what sorts of things breed within it? I prefer not to think about such things. That is why for the past forty years, ever since my discovery, I have knelt with the priests in the temple, and joined in their prayers. With an honored guest such as yourself here, I thought to entertain you instead, to avoid worrying you. Now I see that thought was vain. Will you come to the temple and pray with me?”

And so on the longest night of the year, Fin Lerisas, Chief Lorekeeper of Tal Aivon, and Meical Dorn, Lorekeeper of Great Rabda, knelt in the temple and prayed to St. Jesus and St. Mahomet that time continue, that 1:59 AM be followed by 2:00 AM just as it always had in the past, and that the people be forgiven the sins of the Lost World, which had dared to change Time itself. And lo, at the appointed hour the six clocks on the six gates of Tal Aivon showed 2 AM, and the people rejoiced, and the kerosene lights were lit and the city of Tal Aivon was lively and pleasant once again.

Three days later, Meical Dorn left Tal Aivon minus the gold he had brought but with a sunblessing of his own, a beautiful slate-gray mather that would have the engineers of Great Rabda dancing with glee. They had offered him a timer instead, a beautiful digital timer that even played short tunes at different hours, but Meical had refused. He bore a secret that need not trouble the people of Great Rabda. They would have a mather, and calculate things lightning-quick, and never know that there was a flaw in Time that even the gods themselves could not resolve.

But until the day he died, every so often on chill autumn nights Meical Dorn would look up at the stars and shudder.

List of Reasons Why Gunpowder Treason Should Be Forgot

1. Guy Fawkes Day was historically used as excuse for anti-Catholic bigotry and traditionally involved burning effigy of the Pope
2. In Britain it frequently devolves into hooliganism and public disorder
3. King James I was actually kind of a jerk who was super keen on witch-burning and minority-exterminating
4. George Washington himself specifically condemned Guy Fawkes celebrations
5. Its modern association is primarily with people from 4chan

(I’m not actually against Guy Fawkes Day. It just bothers me when people say they can’t think of *any* reason for something. Did you think about it for five minutes first?)

Lies, Damned Lies, And Facebook (Part 4 of ∞)

(see also parts 1, 2, and 3 of ∞)

Some friends of mine on Facebook were talking about rape (as you do), and one of them brought up how anti-rape public awareness campaigns targeting men have been found to be effective.

Their evidence, just like the evidence of everyone else who makes this claim, was an article in the Vancouver Globe and Mail, which noted that the city's anti-rape campaign – called "Don't Be That Guy" and consisted of hanging posters with catchy slogans about how men really shouldn't have sex with non-consenting women – had successfully decreased rapes in the city. They know this because the year before the campaign, sexual assault rates went up, but the first full year of the campaign, sexual assault rates went down.

I wish I lived in a world that worked like this. If hanging up pictures of distressed-looking pretty women that say "DON'T BE THAT GUY" can reverse rising sexual assault rates within a year, maybe we could end the drug war with pictures of unhappy-looking people smoking that say "WINNERS DON'T DO DRUGS", or prevent teenage pregnancy with pictures of smiling couples that say "TRUE LOVE WAITS", or decrease racism with pictures of different ethnicities holding hands and the caption "DIVERSITY IS OUR GREATEST ASSET".

But I shouldn't be so cynical. This initiative does claim to have evidence supporting it. So let's analyze that evidence.

The evidence is that after the "Don't Be That Guy" campaign in Vancouver, sex crimes decreased by 10%.

But Vancouver was not the only city to try the campaign. “Don’t Be That Guy” originated in Edmonton. After that city’s “Don’t Be That Guy” campaign, sex crimes increased by 14%, even as other categories of crime were decreasing.

Am I blaming Edmonton’s “Don’t Be That Guy” campaign for causing more rape? Probably not (although if you read the links above, you know that’s not as far-fetched as it sounds). I am saying that these kinds of small changes in crime are very common for cities and mean nothing.

Just as it would be silly to seize upon a +14% increase in rapes in Edmonton as suggesting anti-rape campaigns cause rape, so it is silly to seize upon a -10% decrease in rapes in Vancouver as suggesting anti-rape campaigns prevent rape.

It is generally bad to take a single number out of context and use it to claim a causative effect. If a hundred people take homeopathic remedies, one of them will have her blood pressure go down 10%, just because blood pressure goes down sometimes. If you are the person selling the homeopathic remedy, you can say “Look! Vanessa used homeopathy, and her blood pressure decreased by 10%! Clearly it’s working!”

Likewise, if you are a movement pushing this kind of anti-rape campaign, you can publicize the article in the Vancouver newspaper that says crime rates decreased 10% after your “remedy” was tried, and just fail to publicize anything about Edmonton **[1]**. If you have a good enough campaign, everyone will hear about it from you, no one will bother to double-check, and it will become a well-known fact that anti-rape campaigns targeting men lowers the sexual offense rate.

The counterspell is moderate skepticism of articles in the popular media claiming amazing social science results when there is no peer-reviewed research corroborating the discovery. In the case of “Don’t Be That Guy”, as far as I know such peer-reviewed research does not exist.

This concludes the part of this essay that will not be lots of boring numbers.

Still with me? Good. There are actually some other interesting ways we can pick apart the Vancouver rape data, and for the sake of completeness we might as well try. Our source will be the Vancouver Police Department’s crime statistics reports.

These statistics confirm that in 2011, when the “Don’t Be That Guy” program was introduced, sex offenses decreased by -9.7%. But just from that number, we don’t know whether this is an effect of the program, or just the continuation of a long-term trend. Maybe sex crimes in Vancouver decrease by 9.7% every year. So the Globe and Mail article very correctly focuses on the change in the trend. They observe that in 2010, the year before the campaign, sex crimes rose +4.7%, and now in 2011 with the campaign they’re dropping -9.7%. That’s much more impressive. It’s a big switch in the trend.

The mathematical term for this kind of change in trends is “second derivatives”, and the second derivative of sex crimes from 2010 to 2011 was -14.4 (the difference between a +4.7% increase and a -9.7% decrease).

But from the same crime statistics, we see that breaking and entering was down -11.1% in 2010, but up by +5.4% in 2011, so its second derivative shifted 16.6. Possession of stolen goods shifted from -18.7% in 2010 to +3.8% in 2011, a second derivative of +22.5. Arson shifted from -6.5% in 2010 to +38% in 2011, for a second derivative of +44.5. It would be moderately unfair to include murder on this list since there are so few murders that large variations are easy to come by, but if you'll humor me, change in murder rate went from -53.7% in 2010 to +48.1% in 2011, for a second derivative of 101.8.

My point with all these numbers is that second derivatives are volatile things, and the change in sexual offenses in 2010-2011 was equal to or less than the change in a whole bunch of other crimes, none of which had campaigns targeted against them. In other words, it's very much the sort of thing you would expect to see by chance.

This is true across years as well as across crimes. In 2004, sex offenses were up by +21.9%, but in 2006, they were decreasing by -1.3, and the next year, in 2007, they were down -7.2%, almost as much as with the introduction of the "Don't Be That Guy" campaign in 2011 (of course, there was no campaign in 2007) By 2009, the pendulum had swung wildly once again, and there was a huge +17.7% increase in sexual offenses. This tells the story of a city where it is common for sex offenses to go up by a large amount one year and then go down by a large amount in another year, or vice versa, whether people are putting posters on streetlights or not.

I'm not exactly sure how kosher this is, but we can try to quantify this with a third derivative – the change in the change in the change **[2]**. We mentioned before that the second derivative of sex crimes from 2010 to 2011, the year of the "Don't Be That Guy" campaign, was -14.4 – a movement from +4.7 to -9.7. But from 2009 to 2010, the year *before* the program was introduced, the second derivative was -13.0 – it moved from +17.7% to +4.7%.

In other words, the impressive change in the second derivative of sex offenses in 2011 was the near-exact predicted continuation of a trend that had started a year earlier. There was just as big a shift in 2010 – when the program had not started yet – as in 2011, when it had.

There's something a little sneaky about this – it should always be possible to find *some* level of derivative that supports the results you want – but given that the Globe and Mail article was making a second-derivative based argument, I don't think it's *too* unfair for me to follow along. Readers who know more about statistics than I can tell me whether this is already a solved problem.

Of course, the *proper* way to do this would be to use all the different crime statistics to figure out what the average noise in changes in Vancouver crime rates are, and then see whether the change in rape in 2011 was significantly greater than the noise at a 5% level. I have neither the time nor the intelligence to do this formally, but it's what I've been doing informally throughout this article and it should be clear by now that the answer is "no, it wasn't".

So now we just have some mopping up to do. For example, did you know that the majority of the "sexual offenses" reported as declining in the Globe and Mail article aren't rape at all, but things like "groping, grabbing, or kissing", and that as far as I know no statistics have ever been published about what happened to rape itself during the campaign period? Or that according to official Vancouver statistics, the large majority of sexual offences are committed by strangers – which is contrary to what we know to be true about sexual offences in general, which suggests there's a large bias in terms of which sexual offences get reported to the police, and which are the category of sexual offence it's least plausible that posters defining rape would help with? Or that police attribute much the previous rise in sexual offenses to offenses committed against prostitutes, who are totally ignored by the "smiling girl asks to have a drink with you but probably doesn't want sex" focus of the "Don't Be That Guy" campaign?

Of course, none of this proves that campaigns like "Don't Be That Guy" don't work. It just shows there's not much evidence that they do. This brings us back to our prior that putting up some posters around town can arrest major social ills. Based on the examples above like drug use and abstinence-only sex education, my prior for this is very low.

Your mileage may vary, but before deciding I urge you to consider an alternative hypothesis. That claims that the posters lower rape – never very plausible on the face of it – are fake consequentialism (I don't have a good link for this concept yet, but search "fake consequentialism" here). That the real reason people put up "Don't Be That Guy" posters is the same reason they put RAPE FREE ZONE posters on college classroom doors *and* the same reason men's rights activists have taken to putting up "Don't Be That Girl" posters. It's a way of signaling your membership in a specific tribe and demonstrating the power of that tribe by being able to take over public places with your symbols.

Or maybe I'm being totally paranoid. We'll probably know one way or the other soon enough, as its "confirmed success" in Vancouver has led to everyone everywhere adapting the "Don't Be That Guy" campaign and we will soon have oodles of useful data that a real social scientist can use to do a real analysis. I look forward to reading what it says.

In the meantime, please don't say that there is already evidence in the program's favor unless you're willing to give many more caveats than people typically do when they make that statement.

Footnotes

[1]: The only blogger I have ever seen mention Edmonton in regards to “Don’t Be That Guy” is Greta Christina, who very correctly talks about Edmonton as the birthplace and center of the campaign, says how effective it was in Edmonton and how happy the police there were with its effects, and cites as proof of this……that rape went down in Vancouver. Did you know that 100% of my uses of the phrase “what is this I don’t even” this year have been while reading Freethought Blogs?

[2]: Third derivative can also be called “jerk”, which seems appropriate when we are using it to measure the number of rapists.

Highlights From My Notes From A Forensic Psychiatry Conference Last Week

– Contrary to popular belief, the insanity defense is not overused. It’s used in only about 1% of felony trials and successful only about a quarter of the time it is used. 90% of people who successfully plead insanity had been diagnosed with a disorder before they committed their crime.

– Felons found insane usually get locked up in forensic hospitals even longer than sane felons are locked up in prisons. Some statistics say that by pleading insanity you increase your time behind bars by 50 – 100%

– For some reason, there is a law saying juries are not allowed to be told what will happen to defendants if they return a certain verdict. Juries assume that if a defendant is found “not guilty by reason of insanity”, they will be released scot-free. Although this is completely false, no one is allowed to tell the jury this. So it’s really hard to win an insanity defense simply because juries think it will mean a felon will be put right back on the streets.

– Forensic psychiatrists have become very effective at determining which criminals who plead insanity are “faking it”. They usually rely on patterns of mental disease which psychiatrists know but criminals don’t. For example, they’ll start by asking “Do you hear voices?”, and most fakers, anxious to please, say they do. Then the psychiatrist will ask questions like “Which side of your head do the voices come from?” and “Do you ever have smells associated with the voices?” Still eager to please, the fakers will choose a side of the head for their voices to be on, and make up smells that happen at the same time as their voices. But real schizophrenics don’t generally hear their voices to one side or hallucinate smells, so this decreases the likelihood that they’re telling the truth. Some of these questions are very tricky – for example, one psychiatrist asks both “Do you ever hear secret messages for you from the TV or radio?” and “Do cats and dogs ever give you secret messages?”. The first is very common in mental illness; the second practically never happens. Unless you’re a psychiatrist yourself, you’re not going to know these things and you’ll end up claiming symptoms that make no sense.

– Some criminals also claim to be mentally retarded, especially in states where it’s illegal to execute retarded people. The process for catching these people is equally elegant. They are asked to take a multiple-choice vocabulary test with easy, medium, and difficult words. Real mentally retarded people will do okay on the easy words but perform at chance on the medium and difficult words. Fakers will also do okay on the easy words – they are smart enough to understand that even mentally retarded people know *some* things – but then they intentionally throw the medium words and do *worse than chance*. On the difficult words, the fakers honestly don’t know them and so they go back to performing at chance again. Computers can detect these patterns and easily and confidently point out a fake.

[is it contrary to the Virtue of Silence to reveal such things? Nah; I checked and other much more prominent media have discussed them already. Any criminal Googling ways to fake insanity effectively already has enough sources to work from.]

– My notes say Baxter vs. Harold, but I must have miscopied because Google can’t find it – but anyway, it was a case in which New York State took nine hundred something felons who had finished their prison terms and transferred them to mental institutions because they were "too dangerous to let back on the streets". One of the felons sued, and the court agreed that this was illegal, and the state could not confine prisoners after their sentence ended just because of high risk. Some bright scientist decided to turn this into a study and see how many of these designated "high-risk offenders" actually re-offended after [unreadable] number of years. The answer was 22. 22 out of nine-hundred or something. I wish I could find this study online. The point is that forensic psychiatrists were terrible at predicting who is or is not a future risk, although the forensic psychiatrist giving the talk assured us confidently that they are much better now.

[several pages of mandalas and other geometrical figures where these notes should be; apparently I got distracted]

– A fragment of a story from one of the examples we were given: "She said she started the fire because her ninth child was taken away by Child Protective Services, just as her previous eight children had been." This wins my prize for "best single-sentence argument for some kind of parenting license or eugenics or something"

– Another fragment of a story: "The night before his execution, murderer Troy Gregg sawed through the bars of his cell, donned a homemade prison guard uniform, and accomplished the first death row breakout in Georgia history. He was beaten to death in a bar fight later that night."

– A fragment of a story told by the keynote speaker: "My interest in murder cases started when I was a child. When I was five, my babysitter Wilson was arrested for strangling his wife. He served three years in prison. When he got out, he was my babysitter again. Those were less safety-conscious times."

[several pages of notes in a moderately successful attempt to write English in an ideographic script. Unfortunately, I can't remember what the ideographs mean well enough to decode the content.]

– Dutiful recording of a claim by the keynote speaker that a vastly disproportionate number of death row inmates have the middle name "Lee".

– During a death penalty case, jurors who don't believe in capital punishment (and therefore would throw the case for reasons other than the defendant's guilt or innocence) are automatically excluded from participating. But those jurors are usually more liberal, and the pro-capital-punishment jurors who get tapped are usually more conservative. Liberal jurors are usually more likely to acquit any type of criminal, and conservatives more likely to convict. Therefore, in certain cases it can be *easier* to get a death penalty conviction than a life in prison conviction, because you're throwing the most merciful jurors out of the potential pool.

– Psychologist who works with death row inmates on a daily basis says "There's not one defendant I ever met who ever considered not committing his crime because he was worried about getting the death penalty." He said this makes him think death penalty is not an effective deterrent. I can't help but wonder if maybe all the people it effectively deters don't end up on death row talking to a forensic psychologist.

– The keynote speaker worked on the case of an Ohio school shooter who tried to plead insanity. Later, the shooter changed his mind and said he had only done that because he was bored, but pleading insanity was also boring so he would just accept the sentence. He somehow managed to make it into the courtroom for the sentencing hearing wearing a shirt reading "KILLER". When asked if he wanted to address the victim's families – usually a chance for some last minute apologizing, he said (and I quote) "The hand that pulled the trigger that killed your sons now masturbates to the memory", then gave them the finger. This is my new go-to story when someone says that evil is merely the absence of good.

– Said during a capital trial by a none-too-bright con: "On advice of my accountant, I invoke my Fifth Commitment rights".

Genetic Russian Roulette

[TRIGGER WARNING: This would be a really bad post to read if you have or are about to have a young child]

I.

One of the downsides to working in psychiatry is that it is slowly but inexorably sapping away first my enthusiasm about, and now even my willingness to, have children.

Medicine was bad enough. It wasn't the kids who get leukaemia at age seven and die who got to me. It was the ones with syndromes. Down Syndrome is the one everyone knows about, but by no means the worst. Lesch-Nyhan Syndrome is a lovely little condition in which children are born without the ability to metabolize uric acid properly. The results include severe mental retardation, stereotyped jerking movements of the limbs, face frozen in a permanent grimace, and something the textbooks charmingly refer to as "involuntary writhing". For poorly understood reasons, these children also exhibit "uncontrollable self-injury", usually head-banging and trying to bite any part of their body within reach of their mouth – something caregivers quickly learn not to let parts of their bodies be. Oh, and also vomiting, spitting, and uncontrollable urges to use profanity.

(for a fun intellectual exercise, imagine what arguments might be able to convince William of Ockham that "deficiency in uric acid metabolism" is a more parsimonious explanation for this syndrome than "possessed by demons")

The disability rights crowd will probably call me ableist, or disableophobic, or misodisablistic, or whatever people say these days, but the idea of having a child with Lesch-Nyhan

Syndrome, and having to take care of him until he dies (likely in his teens; it's not a very survivable disease) or gets old enough to go off to a group home – that terrifies me. I worry I would spend every second of every day hating my kid and begrudging him the countless hours and tens of thousands of dollars I would be spending on him, all while maintaining a smile both to him – because goodness knows he has it bad enough already without having to cope with crushing guilt and parental enmity – and to the rest of the world – so I could get my official status as Brave and Kind Long-Suffering Caretaker, instead of Ableist Jerk Who Kind Of Wishes He Could Just Put His Kid Out Of His Misery And Get On With His Life).

(note: I use the masculine pronoun not unreflectively, but because Lesch-Nyhan is x-linked recessive)

The good news is the incidence of Lesch-Nyhan Syndrome is only 1/400,000 births. The bad news is that if it's not Lesch-Nyhan Syndrome, it could be Treacher-Collins Syndrome. Or DiGeorge Syndrome. Or Cornelia de Lange Syndrome. Or any of a thousand others. Have you ever seen a Pediatric Genetics text? They're not small books.

And not all of these are Lesch-Nyhan-level bad – I like playing with Down Syndrome kids. But I would still be unexpectedly finding out that what I thought was just a perfectly normal super-gargantuan committment to sink *most* of my resources into a mostly-helpless creature for twenty-five years would in fact become, without my planning or consent, a super-ultra-titano-gargantuan commitment to sink *even more* of my resources into an even-more-helpless creature indefinitely. And all the associated dreams – of seeing my kid get grow up and raise a family, of debating the great questions of life with zir, of secretly training zir to be a child prodigy who does calculus at age five like John Stuart Mill – would all be gone. I am reluctant to say “all the things I look forward to about having a kid”, because someone will lecture me about how I *should* want to just gaze deep into my child’s eyes, loving them unconditionally as a human being in the image of God. But as the saying goes, we can do what we want but cannot want what we want. Also, eye contact is scary.

But at this point it’s not even the syndromes that worry me the most. The kids with syndromes I’ve met are nice people, I like them, and if it came to that I’d at least have great social support from my amazing friends.

Right now what scares me is psychiatry.

II.

I remember vividly the first time I met my first antisocial personality disorder patient with perfect parents.

The dad was a something something manager at Ford. The mom was a high school teacher. Both from middle class backgrounds, never been in a fight, never been in jail, never even used drugs. Two other children at home, one of them working on college applications to nice schools, the other getting As in junior high and playing in the band.

Their oldest kid was in the psychiatric hospital where I worked because…actually, no, I can’t remember what he was in for, the particular time when I met his parents. It could have been the stabbing people. It could have been the constant drug use. It could have been lying like a rug to the police. And as I was explaining whichever the latest disaster was to his parents, on their faces was just this look of heartbreak mixed with total lack of surprise.

I’d seen dozens of patients worse than this guy already, but they’d all been easy to brush off as something that could never happen to me. Either their parents were just as bad as they are, or they’d been beaten as a kid, or they had a history of sexual abuse, or something else where it was very easy to say “Aha! I will just not attack my children with broken beer bottles when I am a father, and then they won’t turn out this way.” Or since technically I’m supposed to be kind of a genetic determinist, I could just not have terrible person genes, and marry someone else without terrible person genes, and our kid would be okay too.

This guy’s family – and he was the first of several – didn’t give me any of those easy outs. You can do everything right in the world, and your kid can still grow up to be a murderer, or a rapist, or just one of those guys with a tattoo of a skull and a nickname like “Snake”.

I like the metaphor of “the genetic lottery”, but on the worst days it starts to feel more like genetic Russian Roulette. Maybe the first kid is beautiful and grows up to be a scientist, the second one is compassionate and becomes an artist, and then boom, the third one vivisects stray cats and tries to burn the house down. But it’s too late to return her for store credit; you’re morally obligated to spend twenty years taking care of her and sending her to special schools and taking late night calls from the police and paying her bail, all in the desperate hope that maybe one day she’ll shape up. Like I said. Russian roulette.

III.

I get irrationally angry whenever I hear people diss discipline. Like, if you just believe spanking is often bad for children, then that’s fine and as far as I can tell empirically correct. What bothers me is the people who see someone spank their kid and say “What that parent really needs to learn to do is talk things out. If they would just learn to communicate with their kid, the kid would learn that what ze’s doing is wrong, apologize, and none of this violence would be necessary.” This is the sort of statement that makes perfect sense if you’ve never been in a position where you had to control a genuinely bad (or even mediocre) kid.

(as such I am suspicious of its signaling value. That is, I worry that saying children should never be disciplined harshly is a way of saying “Ha ha, my children are so great I never need to discipline them. And there *you* are with a bad apple. Chump!”)

Don’t get me wrong, there are lots of problems with parents abusing kids under the excuse of “discipline”, and the idea of an unelected and unaccountable individual being able to hurt someone else whenever they want is terrifying. It might be that a Congressional bill to ban all discipline stronger than dirty looks would on net be a good thing. But when people try to push it by saying “If you just demonstrated your love and affection, you wouldn’t *need* any discipline” – that’s what gets my blood boiling.

I think this dates from my time as a schoolteacher. When I was a student, I hated all my teachers and thought that if they just ditched the constant repetition, the cutesy but vapid games, the police state attitude, then everyone would learn a lot more and school would finally live up to its potential as “not totally incompatible with learning, sometimes”.

And then I started teaching English, tried presenting the actually interesting things about the English language at a reasonable pace as if I were talking to real human beings.

And *it was a disaster*. I would give this really brilliant and lucid presentation of a fascinating concept, and then ask a basic question about it, and even though I had *just* explained it, no one in the class would even have been listening to it. They'd be too busy chattering to one another in the corner. So finally out of desperation I was like "Who wants to do some kind of idiotic activity in which we all pick English words and color them in and then do a stupid dance about them??!" (I may not have used those exact words) and sure enough everyone wanted to and at the end some of them sort of vaguely remembered the vocabulary.

By the end of the school year I had realized that nothing was getting learned without threatening a test on it later, nothing was getting learned regardless unless it was rote memorization of a few especially boring points, and that I could usually force students to sit still long enough to learn it if and only if I bribed them with vapid games at regular intervals.

Yet pretty much every day I see people saying "Schools are evil fascist institutions that deliberately avoid teaching students for sinister reasons. If you just inspire a love of learning in them, they'll be thrilled to finally have new vistas to explore and they'll go above and beyond what you possibly expected."

To which the only answer is *no they frickin' won't*. Yes, there will be two or three who do. Probably you were one of them, or your kid is one of them, and you think everything should be centered around those people. Fine. That's what home schooling is for. But there will also be oh so many who ask "Will the grandeur and beauty of the fathomless universe be on the test?". And when you say that the *true* test is whether they feel connected to the tradition of inquiry into the mysteries of Nature, they'll roll their eyes and secretly play Pokemon on their Nintendo DS thinking you can’t see it if it’s held kind of under their desk.

I don’t think I used to be an optimist. I think I used to be a narcissist. I figured that when *I* was a teacher, everything would work out, my kids would be kind and attentive, my lessons would stick, and there would be no behavioral problems or if there are they would quiet down after I give them a friendly talk about why attention is important. I felt like the Universe *owed it* to me to have everything work out. I didn’t realize on a gut level that kids could just *not cooperate*.

The parents who pooh-pooh strict discipline have the same blind spot. It’s an easy blind spot to have if your own kids are wonderful. It’s hard to realize that you can just ask them to behave until you’re blue in the face, set strict limits, send them to bed without dessert, and the kid can just choose to *not behave*.

And we’ve come a long way from Lesch-Nyhan Syndrome here, but these all tie together into a complicated knot of worry. It’s a fundamental realization that *I could have a kid I can’t fix*. Not in the “I want her to be a businessperson but she wants to be a poet” sort of way, but in the “I want her to not vivisect cats for fun, and she wants to vivisect cats for fun” sort of way. Someone whom I lavish all the love in the world on, and give exactly the right dose of prenatal iodine, and whatever, and the universe just says “Nope, you’re stuck with this person. You should have thought of that *before* one of your sperm had a single nucleotide mutation in an out-of-the-way corner of the genome”.

I have a serious girlfriend, and I’m only two years younger than my father was when he had me, and all my friends are having kids, and I really want children, and I have *no idea* what I’m going to do.

November Housekeeping Post

1. Ever read Less Wrong? Consider yourself part of the "rationality community"? You are eligible to take the new 2013 Less Wrong Census/Survey, now including a MONETARY REWARD (of course, this is from the website that brought you the Counterfactual Mugging, so the reward can't be simple). Go to http://lesswrong.com/lw/j4y/2013_less_wrong_censussurvey/ for more information and a link.

2. By an unexpected positive conjunction of my work schedule and financial situation, contrary to my previous claims I *will* being going to Raemon's solstice celebration in New York City this year. I look forward to meeting a lot of new people and catching up with the old ones. I'll be arriving in NYC on the 13th and staying until the 16th. Does anyone have a *quiet* space in town I could stay in for that time? You can email me at scott [at] shireroth [dot] org. Also, grumble grumble what's the point of having three girlfriends if none of them can make it to the biggest rationalist social event of the year grumble anyone want to be my date to the party?

3. There will be a second Detroit/Ann Arbor Less Wrong/Slate Star Codex meetup at my house. Tentative date is January 3rd or 4th. Special guests Alicorn, Mike Blume, and Ozymandias will probably be attending. I will post more about this closer to the date and once everything has been confirmed, but if you live in the area please save the date (well, combination of dates).

4. Since this is titled "housekeeping post", maybe I can ask about actual housekeeping! Large patches of my lawn are turning brown. I figured this was because of Winter and It

Being Below Freezing, but this doesn’t seem to be happening to other patches or to my neighbors’ lawns. I mow my lawn irregularly and don’t water it because it rains a lot. Is there anything I’m doing wrong, or should I just wait for the snow to bury my problem?

A Letter I Will Probably Send To The FDA

I predict that the current spat between 23andMe and the FDA will be less apocalyptic than some people fear. In particular, I expect 23andMe will submit a lot of paperwork, change some of their wording around, advertise a little less aggressively, stop listing some or all of their “Health Risks” and “Drug Interactions” section, and then be back on their feet in a few months.

But I’m not sure. And even that seems like a bad thing. Although I realize this is probably a very complicated legal game here, I’d like to add whatever pressure I can. So here is a letter I plan to send the FDA and maybe my Congressor once I figure out who that is. Please let me know about any factual errors, logical errors, or things I should change because they are unconvincing or offensive:

> To whom it may concern:
>
> I am writing to voice concern at your decision to ask the company 23andMe to halt genetic testing.
>
> As a doctor, I am well aware of both the importance of genetic testing in medicine and of the difficulty in getting these tests through the normal medical system. Many of my colleagues do not know of the existence of personal genomics, and believe that genetic testing is a once-in-a-blue-moon procedure to be ordered only when there is high clinical suspicion for a rare disease such as cystic fibrosis.
>
> Even in cases where doctors are willing to use genetic information to help in treatment, they have few good

options within the standard medical system. Tests for a single mutation may cost several hundred dollars, in addition to the expensive consultation with a physician necessary to order the test and the second expensive consultation with a physician necessary to receive results from that same physician. These are rarely covered by insurance, especially for poorer patients. Patients are rarely willing to purchase these expensive tests out-of-pocket, and when they do, both patient and doctor must wait weeks for each individual result they want.

In contrast, 23andMe has raised awareness of genetics among the general population and given them questions and concerns, usually appropriate, which they can discuss with their doctor. Their doctor can then follow up on these concerns. Such followup may involve reassurance, confirmation with other genetic testing, confirmation through other diagnostic modalities, or referral to another professional such as a genetic counselor. In my experience personal genomic results do not unilaterally determine a course of treatment, but may influence an ambiguous clinical picture in one direction or the other, or be a useful factor when deciding between otherwise equipotent medications. Banning the entire field of personal genomics in one fell swoop would eliminate a useful diagnostic tool from everyone except a few very wealthy patients.

Let me give a personal example. My family has no history of any genetic disease, I (like many Americans) do not have a primary care physician, and I had no reason to go to a doctor to get an expensive genetic screening. When 23andMe began offering genetic tests through the mail for $99, the easy availability convinced me to try one out. From that test I discovered that I was, like one in every thirty Ashkenazi Jews, a carrier for hereditary sensory and autonomic neuropathy type III, a serious and often fatal neurological disease linked to a locus on chromosome 9 which can be detected by genetic screening with an accuracy of >99%.

This has given me information I need to control my future: in particular, it has made me more wary of having children with any other Ashkenazi Jew until they have been tested for the mutation as well. Without personal genetic testing services such as 23andMe, I would not have discovered this mutation; with the loss of these services, my future partner will have to convince her primary care doctor to order her a prescription version of the test. Such a doctor will almost certainly never have heard of this rare condition, may blow her off, and may charge a hefty fee for this service if they perform it at all.

In addition to these immediate consequences, I am distressed by the likely effects of this decision on genetic research. Many of the most promising medical advances in the pipeline are based on genetics, but one major bottleneck to genetic discovery is the absence of good genomes to work with. A typical genetic experiment involves using advanced software to tease out small correlations between genotypes and phenotypes. Such experiments require genotypic and phenotypic data from thousands of individuals, which is often hard to come by especially given legal restrictions. 23andMe has amassed what may be the largest database of such information anywhere in the world and is making it available to researchers (with appropriate privacy protection), an amazing public service for a for-profit company. Restricting their ability to provide this service will almost certainly delay life-saving genetic discoveries.

In contrast to these important services provided by 23andMe, your stated worries about the company are, with all due respect, somewhat bizarre. As I gather from your letter to the company, you have at least three fears.

First, you state that “if the BRCA-related risk assessment for breast or ovarian cancer reports a false positive, it could lead a patient to undergo prophylactic surgery, chemoprevention, intensive screening, or other morbidity-inducing actions”. As I would hope the FDA is aware, it is extremely common in medicine to screen for rare conditions through affordable and noninvasive tests, with positive results followed up by more expensive, more complicated tests to confirm the diagnosis. You allow, for example, prostate specific antigen testing, even though the positive predictive value of this test at its most commonly used values is only 30%, because you correctly assume it will be followed up by further analysis before any drastic steps are taken. If the FDA believes that there are surgeons who would say “You claim to have heard you had a BRCA mutation, I have no follow-up questions, let’s schedule you for a double mastectomy”, then the FDA should be pursuing those surgeons, not genomics companies.

Second, you state that “we still do not have any assurance that the firm has analytically or clinically validated the PGS for its intended uses, which have expanded from the uses that the firm identified in its submissions”. If this means only that you are uncertain whether 23andMe’s genetic testing actually reads the genome correctly – that what it reports as an A nucleotide is indeed an A nucleotide – then I agree this is a reasonable question for the FDA to pursue. However, by your mention of intended uses “expanding”, I gather that in fact you mean the specific genetic claims they make, for example that genetics affect warfarin metabolism.

There are several thousand papers about genetics, linking specific genes to everything from eye color to intelligence to chance of developing colon cancer. Some of these papers are no doubt correct, others will no doubt in time be found incorrect. However, I worry you want 23andMe to independently validate each of these results before informing customers about them. Worse still would be a demand that the company validate the results *with its specific testing kit*. It would be cost-prohibitive even for a single company like 23andMe. It would destroy any possibility of competition; even if one company did it, new startups would not be able to. And it is unnecessary. If BRCA is linked to breast cancer in general, and 23andMe’s kit reads the genome correctly, then 23andMe’s result for BRCA will be linked to breast cancer, as will the result of any other kit that reads off the genome correctly. Asking companies to validate each result for their particular kit makes no more sense than asking each pharmacy to do independent studies to confirm that their own bottle of penicillin kills bacteria. The pharmacy’s job is to prove they have real penicillin: once they do that, the general result that penicillin kills bacteria applies.

In my opinion, 23andMe handles the complexity and uncertainty of genetic testing entirely correctly. It rates each of its tests on a “confidence” scale showing how much consensus there is in the literature that the result is correct and gives a clear explanation of what each confidence level means, citing all the studies it uses to gather data. It states very clearly on every single page that its results are open to interpretation and should not be used to unilaterally guide medical treatment. On results with less than its highest level of confidence, it refuses to even give numbers out of a fear that these might seem unfairly authoritative, only linking to the appropriate study and allowing consumers or their doctors to draw their own conclusions.

Third and most important, you state that “assessments for drug responses carry the risks that patients relying on such tests may begin to self-manage their treatments through dose changes or even abandon certain therapies depending on the outcome of the assessment”. For me, this is the most worrying objection, because it seems to acknowledge that although 23andMe may provide accurate information with the appropriate disclaimers, merely giving patients access to medical information is itself dangerous.

I agree that this can sometimes be true. For example, when I was in medical school many of my classmates came down with so-called “medical student syndrome”, in which their increased awareness of symptoms and diseases made them more likely to believe they had some rare and dangerous syndrome, causing unnecessary pain and sometimes even doctors visits and lab tests.

But the alternative is banning all information that anyone could possibly use to evaluate their own health. I have even heard the tongue-in-cheek proposal of banning medical textbooks. You worry that if a patient learns their genetic susceptibility to warfarin it may convince them to alter their own warfarin dose without a doctor’s input. But the same is true if a patient learns the (true) fact that leafy green vegetables affect warfarin levels, or that women require less warfarin than men – facts that can be discovered in any book about medicine or pharmacology. Likewise, knowing one’s weight may convince someone to take unsafe diet pills, to starve themselves, or for that matter to adjust warfarin dosing – the clearance of anticoagulants appears to increase with weight. But banning people from weighing themselves without a prescription is neither clinically nor ethically sound.

The two consistent options are to work to ban patients from learning any accurate information that might be relevant to their health – exactly the opposite of ethical medical practice – or to trust patients and doctors to work together to use information responsibly to support their own care, whether that information comes from a bathroom scale or a genetic test.

In conclusion, while I am excited about the possibility of the FDA helping patients and doctors by vouching for the accuracy of the general process that personal genomics services use to read the genome, I am extremely concerned that it may go beyond this useful service to place unnecessary or impractical restrictions that shut down the field of personal genomics entirely and prevent individuals and medical professionals from getting inexpensive, simple, and unrestricted genetic testing. I urge you to continue negotiating with 23andMe and to take every step possible to ensure that their important service returns to the market as soon as possible.

Thank you for your time.

Yours,
Scott [Alexander], MD, BCh, BAO

Sort of wondering whether I should cite sources for claims like HSAN3 being detectable with genetic testing or warfarin metabolism changing with weight. Is it considered appropriate to have a references section in a letter to a congressional representative? Inappropriate enough to counterbalance the level of awesome?

EDIT: Looks like there’s a petition to sign.

EDIT2: Madeleine Ball points out the less immediately sexy but ultimately equally important issue of the NIH Genomic Data Sharing Policy, which would affect some of these same things but is getting a lot less press.

The Story Of Thanksgiving Is A Science-Fiction Story

It has come to my attention that people are woefully uninformed about certain episodes in the Thanksgiving narrative. For example, almost no one mentions the part where Squanto threatens to release a bioweapon buried under Plymouth Rock that will bring about the apocalypse.

I learned about this and other similarly neglected episodes from the Smithsonian Magazine's Thanksgiving article, and I can't believe I spent seven years of primary school cutting out little belt-buckle hats and feather headdresses while everyone avoided telling me the interesting stuff.

I think the problem is the story of Thanksgiving doesn't really fit in the fables beloved of primary school teachers and moralists. The article above has convinced me that the proper genre for Thanksgiving is science-fiction.

Mr. S, an ordinary American, is minding his own business outside his East Coast home when he is suddenly abducted by short, large-headed creatures from another world. They bring him to their ship and voyage across unimaginable distances to an alien empire both grander and more horrible than he could imagine. The aliens have godlike technologies, but their society is dystopian and hivelike. Enslaved at first, then displayed as a curiosity, he finally wins his freedom through pluck and intelligence. Despite the luxuries he enjoys in his new life, he longs for his homeworld. He befriends a local noble who tells him that the aliens in fact send ships to his world on a regular basis, quietly scouting and seeking resources while the inhabitants remain blissfully unaware of these incursions. He gets passage on such an expedition.

Before his ship gets far, he is abducted and sold into slavery again, only to be rescued by a sect of alien priests who believe he may hold the key to saving his entire race. They are kind to him and ask him to stay, but when he refuses they reluctantly arrange him passage home.

Yet when he returns, Mr. S finds a postapocalyptic wasteland utterly unlike the world he left. America is empty, its great cities gone, a few survivors fighting for scraps among the ruins. 95% of the population is dead, slain by a supervirus unlike any doctors have ever seen. The few rumors from afar say Mexico, Canada, and lands further abroad have suffered the same or worse. He finds the site where his hometown once stood. There is nothing. Wandering in despair, he is captured by a gang of roving bandits and awaits execution or slavery.

Instead, the bandit leader reveals he is the state governor, reduced to his current station by the devastation that destroyed his capital and entire government. An alien ship has landed, and a handful of colonists have set up a little settlement. The governor’s scouts have been watching them from afar and noticed their strange powers. With their help, he could defeat his rivals and re-establish control over the state, restore his old position. “You have been to these creatures’ homeworld,” he says. “You know their ways, you can speak their language. Negotiate an alliance with them, and I will let you live.”

Mr. S is split. The aliens have shown themselves capable of terrible cruelty. They might kill him or enslave him. But they have also shown themselves capable of something resembling kindness. In the end he decides they are neither fully good nor fully evil – just alien. And his own people now seem as alien to him as his former abductors.

So Mr. S heads to the alien settlement, where once again he finds dystopian squalor and shocking ignorance combined with fantastic technology. The aliens are unfamiliar with even the basics of agriculture and desperate for aid. He quickly makes himself indispensable, and although he successfully gets the ex-governor his treaty, he starts forming grander plans. What if he could use these aliens as a tool to unite the warring bands of survivors? Break the ex-governor’s stranglehold on the region? Start rebuilding civilization? What if he could make something completely new, a merger of American ingenuity and alien technology?

Gradually establishing a base for himself in the alien colony, he starts sending out feelers to the local warlords and bands of survivors, speaking of the aliens’ power, implying but never stating outright that such power could be theirs. At first it seems to be working. The warlords treat him as an equal, start to listen to his ideas. They just need one little push. He decides to try an insane bluff.

The apocalypse, he reveals, was no plague but a bioengineered alien superweapon, an attack unleashed by their warships in retaliation for some offense real or imagined. The aliens have brought caches of this weapon from their homeworld and buried it underneath their colony. If they are crossed, they will unleash a second cataclysm, killing even the scattered survivors who made it through the first. And the one who manipulates the aliens, who can unleash their wrath upon a target of his choosing and who is thus unstoppable? This guy.

Just as he seems on the verge of some success, Mr. S takes a step too far. He tries to free himself from his old nemesis the ex-governor by “warning” the aliens of his plot to kill them; the alien leader discovers the subterfuge and the strike against the ex-governor never takes place. When the surviving Americans learn of this betrayal, they accuse Mr. S of going native and turn against him *en masse*. He dies a few months later of what is suspected to be poison, perhaps planted by one of the governor’s men. The aliens seem to take it in stride.

And then a few generations later, they kill nearly everyone. Mercilessly. They do it while praising and admiring their victims. When their genocide is over, they make loud protestations of regret, and try to placate the survivors with gifts. But they do not stop until the massacre is complete. They are neither fully good nor fully evil – just alien.

Then, still caught up in the legends of their homeworld, they forget everything more than a slight inkling that an apocalypse ever happened. A few strangely shaped hills that look like they might be artificial – who built them? Hundreds of miles of groves – who planted them? It is not in the aliens’ nature to think too much about such things.

As for Mr. S? The man who traveled worlds, who pulled the puppet strings behind the scenes, who tried in vain to reverse the fall of civilization? The aliens remember him fondly. Their legends record him as the person who taught them how to fertilize corn with fish heads.

Empire/Forest Fire

So apparently the way to win a certain measure of internet celebrity is to write a seventy-five page document full of graphs and citations criticizing an extremely fringe political philosophy nearly nobody has ever heard of. Huh.

I have updated the Anti-Reactionary FAQ a very small amount, mostly by linking to people's responses to it at the end. In the more distant future I may have time to rebut all of these responses in detail. I certainly want to examine some of what Moldbug said on Friday in more detail as I think he has an overly rosy view of democracy (did I just say that? yes I did! more when I have time to explain fully)

But today's ~~victims~~ honorable debating partners are Bryce of Anarcho-Papist, with The Theory of Demotist Singularity and Jim from Jim's Blog, with the very similar Anti-Anti Reactionary FAQ Part 4: Ever-Leftwards Movement. Both describe a purported effect called the "left singularity" – Bryce very formally with citations scattered throughout, Jim in such spectacular Insane Moonbat mode that spittle may shoot from your computer when you load the page.

What is the "left singularity"? According to Jim:

> Every [past] leftwards movement has become ever more extreme, moved leftwards ever faster, eventually resulting in crisis, usually a bloody and disastrous crisis. The reason that leftists of anglosphere puritan origin rule the world is in large part because all the other left wing movements self destructed horribly, leaving Anglosphere leftists of puritan origin the last power standing.

> I can confidently predict the collapse of leftism, but alas, not that the saints get to win. It sometimes happens that reactionaries take over after the crisis, and all is peace and order, but the more usual outcome is outside invaders take over, sometimes genocidally, or pirates and brigands take over, and slowly over centuries the brigands transition to being feudal lords.
>
> A reactionary victory is possible. Strange things are apt to happen as history approaches a left singularity. White autogenocide is also possible, in which first all white heterosexual males are murdered, largely by each other, then all whites are murdered (Jews discovering to their great shock and surprise that they are white after all), all heterosexuals are murdered, and all males are murdered, then anyone insufficiently leftist is murdered, then the bar for being sufficiently leftist is raised, and raised again, until some of the remaining leftists wise up and murder everyone who is excessively leftist, thus ending the crisis.

Bryce is calmer, more lucid, and has more interesting arguments – but his theory isn't that much different:

> Democracy is not a politically neutral form of government. In the generation of its activity, it systematically favors leftist reforms to social institutions… Endless change in one direction is simply impossible to sustain… An ongoing breakdown of [social] institutions simultaneously creates the conditions ripe for political profit due to reform, creating a feedback of Leftism which erodes all hierarchies and social networks which are the literal constituents of society…"Singularity" is a point at which [these] phenomena fail to be explainable. As demotism tends inevitably to the left, and accelerates the left, it may

> very well destroy every last social network and hierarchy on its way down, burning up all available social capital and leaving no survivors. Worst case scenario short of complete annihilation, man is returned to the stone age.

In the past when I have talked to Reactionaries, they usually link this idea with some of the great leftist reigns of terror, like those of Mao or Robespierre (surprisingly, Jim interprets Stalin as acting correctly to prevent a leftist reign of terror, but others throw him in as well).

I would like to argue against this idea of left singularity. But first, a different question – do overgrowth of dry brush, and careless smokers discarding lit cigarettes, prevent forest fires?

It would appear that overgrowth of dry brush and careless smokers discarding lit cigarettes prevent forest fires. After all, park rangers examining national parks that are not on fire often report seeing careless smokers discarding lit cigarettes. But park rangers examining national parks that are currently on fire almost never see careless smokers discarding lit cigarettes, or for that matter any smokers at all. Likewise, park rangers examining national parks that are not on fire, and where there have been no fires for a very long time, often report overgrowth of dry brush. But park rangers examining national parks that are currently on fire, or have very recently had a fire, report seeing almost no dry brush at all. Although correlation is not always causation, it certainly looks *suspicious* that parks with careless smokers and dry brush are practically never on fire and have usually been fire-free for a very long time, while their smokerless and brushless counterparts are often on fire at that very moment, or have had fires in the very recent past. Until we have better information, we should conclude that brush and smoking are strong protective factors.

What might we say to a park ranger who reasoned this way? Something like “in order to tease out causes, we can’t just look at conditions during the event in question – which are likely effects rather than causes – we need to look at the conditions present before the event.” Once the National Park Service started analyzing fires at time T by observing the conditions at time T-1, they would pretty quickly discover that overgrowth of dry brush and presence of careless smokers cause forest fires, rather than preventing them.

Okay, back to left singularities. Strong oppressive monarchies prevent reigns of terror. We know this because at the exact time a reign of terror is going on, the government isn’t a oppressive monarchy. Hmmmmmm.

What nations are most lauded by the Reactionaries as examples of the system they want to emulate? Bourbon France. Czarist Russia. Imperial China. And of course Austria-Hungary.

What reigns of terror are most condemned by the Reactionaries as examples of the excesses of Progressivism? The French Revolution. The Russian Revolution. Mao’s Great Leap Forward. And coming in honorable mention is Hitler, whom they don’t talk about much because he’s hard to peg as leftist, but who counts as a reign of terror if anyone does.

You may already have noticed a certain very clear T / T-1 relationship here. But just to spell it out:

Robespierre was born and educated in Bourbon France, and began his reign of terror less than five years after the end of the Bourbon monarchy.

Stalin was born and educated in the Russian Empire, and began his reign of terror less than five years after the end of the Romanov monarchy.

Hitler was born and educated in the Austro-Hungarian Empire and began his reign of terror fifteen years after the fall of the Hapsburg monarchy.

Mao was born and educated in Imperial China, and began his reign of terror thirty two years after the end of the Qing Dynasty. A bit of a slowpoke, but in his defense China was busy being invaded by the Japanese and nearly everyone else. As soon as they were out of the way, he started reigning in terror as quickly as he could.

Let’s set Mao as an outlier and hypothesize that every one of the great reigns of terror of history will occur less than a generation (= about 20 years) after a repressive rightist monarchy, dictatorship, or colonial regime is in power.

The hypothesis does pretty well. Pol Pot took power eight years after the fall of King Sihanouk, who is difficult to pin down politically but who fits both the “monarch” and the “repressive” criteria pretty well. Cromwell followed nine years after Charles I, who dissolved Parliament and sent his opponents to the infamous Star Chamber. Castro followed immediately after repressive strongman Fulgencio Batista, who had cancelled democratic elections and seized power a couple years before. You may go through and test the hypothesis against your own favorite genocidal maniac. It appears to hold.

The Reactionary model, as I understand it, goes like this:

1) Strong repressive monarchy
2) Some idiot decides to change things to democracy
3) Democracy shifts further to the left…

4) …and further to the left…
5) …and ends in a super-leftist reign of terror where thousands or millions die
6) General Monck rides in to restore order with a strong repressive monarchy

What we're not seeing is Step 3 or 4, the shifting leftwards. There's just not enough time. Reigns of terror tend to occur only a couple of years after a country is a strong repressive monarchy on the very far right, sometimes immediately after. So I propose an alternate model:

1) Strong repressive monarchy, builds up pressure and anger
2) Revolution. Well-intentioned but naive revolutionaries hold onto power for a couple of years
3) People still have lots of pressure and anger and are unsatisfied with well-intentioned naive people
4) Angrier, more violent group hijacks revolution. Reign of terror.
5) Foreign, domestic, or natural intervention ends reign of terror.
6) Government mellows out and does pretty okay.

For example, Russia started with the Czarist monarchy. It was overthrown by a broad coalition of people, many of whom were non-Communists, Mensheviks, or at least Bolsheviks less crazy than Stalin. Stalin seized power, started a reign of terror. Stalin dies, is replaced by people like Khrushchev and Gorbachev who are more mellow.

There's an important difference in the predictions made by these models.

In the Reactionary model, being moderate-left is a really bad thing. It means you are well on the path to becoming far-left and suffering a reign of terror. You should try to become far-right right away in order to get off that slippery slope.

In my model, being moderate-left is a great thing. It means you’re probably at step 6, or else you’ve skipped the process entirely and gone directly to good government, do not pass Go, do not collect 200,000 corpses. The place you don’t want to be is far-right. That means you’re overgrown with dry brush. Historically it means you’re at very very high risk for a disastrous revolution.

In the Reactionary model, a monarch who voluntarily relaxes their powers is dooming their country to inevitable violent revolution.

In my model, “those who make peaceful revolution impossible make violent revolution inevitable”. A monarch who voluntarily relaxes their power before being forced to do so by the situation – like the constitutional monarchs of Europe or the King of Thailand – is performing a controlled burn, destroying the overgrowth that would otherwise cause a fire and skipping directly from 1 to 6.

(this is to be distinguished from a monarch who grudgingly gives away of few of their powers when revolution is already in the air in order to placate the revolutionaries, which rarely works)

These countries have the same good results as democratic nations, like the US and UK, which have gone three hundred or so years with only the tiniest traces of state-sponsored violence (and those traces, like the camps for the Japanese during WWII, have not come from the Left). Or like France, where a reign of terror five years after the Bourbon monarchy is clearly contrasted with a hundred fifty terror-free years since it became democratic in 1870.

So the “left singularity” proponents have to explain how a supposed vicious cycle of leftism managed to progress from Louis XIV to Robespierre in five years when there was a monarchy involved, but hasn’t managed to progress much at all in a hundred fifty years when there was democracy involved.

The “left singularity” describes a process that *has never happened in all of history*. It has never been the excesses of democracy that cause a “leftist” reign of terror. It has always been the excesses of monarchy or a monarchy-like dictatorship. And anyone who disagrees, I challenge that person to show me a good example of a reign of terror in a nation that has been stably democratic (defined here as its real head of government being chosen by free and fair elections, plus well-enforced right to free speech) for at least one generation beforehand.

This brings me back to my point 2.3 from the Anti-Reactionary FAQ – classifying both insane millenarian blood-cults and stable market democracies in the same category might (surprise!) not be such a good idea. The former are primarily a short-term reaction to a repressive regime, which burn brightly then fizzle out. The latter are a stable configuration which have so far been extremely successful and which historically seem like the strongest prophylaxis against the former.

More Fictional Drugs Banned By The FDA

See the original: List Of Fictional Drugs Banned By The FDA

ADIPOBARIN
PROFILE: Popular weight loss drug proven to work in clinical studies. Patients lose 10 – 20% of their body weight in two months with no side effects.
BANNED BECAUSE: Patients do not lose any mass. The drug seems to operate by decreasing the effect of gravity through unknown means. As such, it does not decrease the risk of heart attack or diabetes. Also, in overdose it can cause patients to be picked up and tossed about by strong winds. It achieved minor popularity as an illicit recreational drug under the street name “Float”.

EPHDENALOL
PROFILE: Cures delusional parasitosis, an unpleasant condition in which patients hallucinate the feeling of ants crawling all over their skin.
BANNED BECAUSE: The drug is excreted in the sweat; a metabolite of the drug, 2,3-hydroxyephdenalol, is a potent pheromone that inevitably leads to ants crawling all over the patient’s skin.

OCUMOLONE
PROFILE: Touted as a treatment for blindness, this is a concentrated retinal growth factor that encourages the production of new rod and cone cells.
BANNED BECAUSE: For unknown reasons, most new photoreceptors form on the side of the retina facing the inside of the head. Patients reported that in the pitch darkness, these cells somehow adapted to view tiny bursts of electroluminescence produced by the brain. This caused

a closed feedback loop in which neurons firing in the visual cortex produced visual sensations produced neurons firing in the visual cortex and so on, and led to seizures. No cases of blindness were ever cured.

BULOXETINE

PROFILE: Anxiolytic medication which, when taken three times a day with meals, adequately suppresses bulimic urges to purge.
BANNED BECAUSE: Side effect is intractable vomiting.

LEVONORGESTREL-THIOTIMOLINE (SOLD AS “PLAN X”)

PROFILE: Attaches the time-bending powers of thiotimoline to the classical “morning-after pill” to create a drug popularly called “the month-after pill”. Taken up to a month after conception, the thiotimoline allows the levonorgestrel to have reverse-temporal effect and “prevent” embryo implantation after it happens.
BANNED BECAUSE: In sufficiently slow thiotimoline metabolizers, the drug can prevent conception of children who were already born. In one well-publicized case from Miami, a woman concerned about a one-night stand took “Plan X” three weeks later, and her 4 year old son vanished and was never seen again.

NUTENAFIL

PROFILE: Next-generation version of Viagra, this erectile-dysfunction pill treats erectile dysfunction more effectively with fewer side effects.
BANNED BECAUSE: The tendency for the use of the pill to be associated with an erection classically-conditioned a nutenafil-pill fetish into many users. Entire cottage industries sprang up of sex shops selling nutenafil-pill costumes for partners to dress up in, or nutenafil-pill shaped sex toys. Eventually the FDA just got grossed out.

AQUIPERIDONE
PROFILE: When given as an intramuscular injection by trained nurses, decreases patient’s level of belief in hallucinations.
BANNED BECAUSE: When given as an intramuscular injection by hallucinatory nurses, decreases patient’s level of belief in reality.

(When certain groups lobbied against the ban on the grounds that psychotic hallucinations of medical personnel would not follow FDA decisions anyway, it was determined that the Psychotic Hallucinatory World and our own world share a common FDA, which actually explains a lot.)

The Logician And The God-Emperor

Once upon a time a logician accomplished a great deed, and the God-Emperor offered him a choice of rewards. “You may,” said the God-Emperor “have the hand of my eldest daughter, who is the heir to the throne, yet plain to look upon. Or you may take my youngest daughter, who is beautiful beyond words, but without inheritance.”

The next day, the God-Emperor caught the logician in bed with *both* his daughters. Enraged, he hurled threats and abuse at the scholar, who responded with a grin: “Guess someone never learned the difference between ‘or’ and ‘xor’.”

The God-Emperor ordered the logician brought to the throne room in chains, and told him “You have offended me and betrayed my generosity, so you will be subjected to trial by ordeal. I have placed in front of you seven chests. Six of the chests contain skulls. One of the chests contains the key to your chains. I have asked the most devious minds in my kingdom to prepare a logic puzzle giving hints as to which chest is which. You may open a single chest. If you do not find the chest with the key on your first try, you will be slathered in barbecue sauce and thrown to the wolves.”

The logician approached the chests, and upon each was written a clue in complicated logical notation. He examined all seven, and then stood a while, deep in thought. Finally, he opened the third chest. Inside was a golden key.

“Very impressive!” said the God-Emperor. Then he yelled “Guards! Slather this man in barbecue sauce and throw him to the wolves!”

“But…but!” babbled the terrified logician “…but you said…!”

The God-Emperor grinned. “Guess someone never learned the difference between ‘if’ and ‘iff’.”

A Something Sort Of Like Left-Libertarianism-ist Manifesto

I.

> “Forgive him , for he believes that the customs of his tribe are the laws of nature.” –**George Bernard Shaw**

Our tribe has a custom of dividing into Right and Left. The Right supports economic laissez-faire and traditional social norms. The Left wants economic regulations and greater civil liberties.

(unless of course a Democrat is in office)

If you live too long under this system, you start thinking the Left-Right division is a law of nature. I like the Libertarians’ pet Two Dimensional Political Compass because it reminds people that they’re allowed to mix and match.

And so, glory be unto the infinite variety of human thought, we have moved from an unwillingness to credit more than two possible visions of a flourishing society to a grudging acceptance that maybe there are as many as *four* such visions.

(one of which nobody will admit to believing)

“The limits of our language are the limits of our world”. If the only two words in political discourse are Left and Right, it becomes hard to realize libertarianism is a possibility, let alone evaluate it. What equally coherent possible views might a four-word discourse be missing?

What if we abandon our tribe’s custom of conflating free market values and unconcern about social welfare?

Right now some people label themselves “capitalists”. They support free markets and oppose the social safety net. Other people call themselves “socialists”. They oppose free markets and support the social safety net. But there are two more possibilities to fill in there.

Some people might oppose both free markets and a social safety net. I don’t know if there’s a name for this philosophy, but it sounds kind of like fascism – government-controlled corporations running the economy for the good of the strong.

Others might support both free markets and a social safety net. You could call them “welfare capitalists”. I ran a Google search and some of them seem to call themselves “bleeding heart libertarians“. I would call them “correct”.

II.

I think I only realized how committed to this position I was when I read an article about the BART strike. Workers on the BART, a San Francisco area mass transit system, were striking for higher pay. A tech CEO suggested solving the problem by firing the workers and automating their jobs. Some other people didn’t like that, said that BART Worker was one of the only jobs that people without college education could get and make good $60,000+ salaries, said employees were mostly old and wouldn’t be able to get other work, said even if their jobs could be automated it would be cruel to destroy their livelihoods just for the sake of profit.

And my first thought was: if your job can be done more cheaply without you, and the only reason you have it is because people would feel sorry for you if you didn’t, so the government forces your company to keep you on – well then, it’s not a job. It’s a welfare program that requires you to work 9 to 5 before seeing your welfare check.

Suppose BART work really can be done just as well by a cheap machine. Compare the current system – in which BART is prohibited from firing the workers and replacing them with the machine because that would be greedy – to a system where BART fired the workers, bought the machines, but continued giving the workers their old paychecks for no reason. BART gets the same profits either way. The workers get the same amount of money either way. The only difference is that the workers gain forty hours of free time a week.

That suggests that long hours worked by BART employees under the current system are deadweight loss, and the role of BART work is the same as those legendary New Deal welfare programs where they made people dig ditches and fill them in again.

Assuming society has decided it wants to give people welfare, it can do it in one of two different ways: the traditional way, where the government sends them a simple welfare check once a month. Or the sneaky way, where it gets billed as a “job” at the BART.

In the “Simple Check” condition the welfare is funded by the tax base, which presumably is the general population, with rich people paying significantly more. In the “Sneaky Job” condition, the welfare is funded by mass transit users – disproportionately poor people – and the increased cost inevitably disincentivizes mass transit. You may remember mass transit as the thing that cuts down on traffic, sprawl, and carbon emissions – you know, that thing we are trying to desperately convince people to do more of.

In the “Simple Check” condition the recipients of the welfare are the entire impoverished population, although the system may place more emphasis on those who are poorer or need more. In the “Sneaky Job” condition, the recipients of the the welfare are those few well-connected people who get cushy jobs at the BART, chosen somewhat at random but with the usual biases of employers being more likely to hire attractive, tall, Caucasian, etc people. They get $60,000 + no doubt excellent benefits, and everyone else misses out.

In the “Simple Check” condition, the recipients of the welfare can live enjoyable lives doing their hobbies – as the woman in the article puts it, hair and makeup. In the “Sneaky Job” condition, the recipients have to work long hours doing busy work, suffer the normal vagaries of jerkwad bosses and office politics, and suffer the constant stress that they might be fired for underperforming.

With all these advantages of “Simple Check”, what exactly is the “Sneaky Job” condition good for that makes it so popular? As far as I can tell, it is good for fooling people. People do not like paying welfare. But if welfare is placed in work boots and wears a big sign with the word “JOB” painted on it in bright letters, they will walk by it without grumbling. Also important, people do not like *being* on welfare, and as the Rogers & Hammerstein song goes, “when I fool the people I fear, I fool myself as well”.

[lest I be accused of being insensitive by pointing out how *other people’s* jobs are welfare, I will freely admit I have a job partly because the government pays my hospital $100,000 to employ me (of which I get less than half). This is a sufficiently complicated system that a full explanation will have to await another post.]

III.

Welfare has even more clever disguises than this. Let’s talk about those fast food workers who want $15 an hour.

No one denies that it’s pretty crappy to have to live on $8 or so an hour, which is about what fast food workers currently make. But if fast food workers get $15 not because they do $15 worth of work, but because we feel sad that they’re living on too little money, then once again it’s welfare.

And once again we can give them that welfare in one of two ways. We can send them a check, or we can pressure fast food places to pay them more.

If we send people a check, it goes to everyone, whether employed or unemployed. If we pressure fast food places to pay more, then it’s only employed people – the people who need money the least – who get anything.

If we send people a check, who gets the check is presumably determined by need. If we pressure fast food places to pay more, then who pays more is determined by media exposure and political clout. Fast food workers seem to have good union and good public visibility, so they can demand their wages get raised to $15. Garment workers aren’t as well-organized or are less sympathetic, so their wages stay at $8. It encourages a system of “squeaky wheel gets the grease” in which “squeaky” means “go on strike a lot and act miserable”.

If we send people a check, the costs are passed on to the taxpaying public, which includes rich people who pay extra taxes and does not include poor people who get out of a lot of taxes because of their low income. If we pressure fast food places to pay more, the costs are passed on to fast food consumers, who are less likely to be wealthy and more likely to be black than the general population.

And if we send people a check, there's not much taxpayers can do to get out of the extra cost. But if we pressure fast food companies to pay people more, we punish them for hiring workers. If the workers do $8 worth of work for the company, and the government makes them pay $15, it's the equivalent of fining companies $7 an hour for hiring poor people. Not only is this morally unfair, but companies will probably respond rationally by automating as much work as they can, hiring fewer people, or trying to figure out how to replace multiple poor people with fewer wealthier people (for example replacing several clerks with a programmer who runs a computer system).

This is a somewhat harder case as the demand for higher wages among fast-food employees seems endogenous – they're threatening to strike and show the companies how much they need them – rather than exogenous – motivated by government fiat or popular demand. Labor negotiations are coordination problems that are more opaque to analysis than I like. But I think a case can certainly be made that here, too, people are shooting for a noticeably inferior solution just because it helps them avoid *thinking* about the poor. It's not about complicated problems or a changing economic landscape – just make that greedy Walmart behave and somehow I will be freed of all responsibility and all consequences.

At the moment, I *might* support higher minimum wages just because doing things the right way is politically impossible. One can make all sorts of stupid political policies attractive when they are combined with other stupid political policies.

But I am not pleased about it and any time people say we need minimum wages to “punish greedy corporations” it just makes me question the life choices that have made me end out on the same side of a political issue as they did.

IV.

But combining market values and compassion isn’t just about solving everything with basic income guarantees. Let me give another example of a government program meant to increase social welfare and how a more market-informed version would be better than a brute-force regulation.

Affirmative action and minority rights. I don’t trust people on this blog to think clearly about any actual minority group, so let’s pretend we’re worried about affirmative action for Martians, who have been a disempowered underclass ever since their giant heat-ray-bearing tripod machines broke down.

Modern affirmative action says that given the choice between a Martian or an equally qualified Earthling, one must hire the Martian. One big obvious problem here is that “equally qualified” is a matter of opinion. It may be that a boss is prejudiced against Martians, and so tells an excellent Martian candidate that ve is underqualified for the position – the Martian may never know. Or a Martian who was genuinely underqualified may paranoidly believe ve was denied out of prejudice and start a costly lawsuit.

There are other problems as well. Some jobs may have legitimate reasons not to hire Martians – maybe Martians make lousy pilots because their single lidless eye gives them terrible depth perception. Certainly a Martian actor is unqualified to play Abraham Lincoln in a historical biopic. One could offer to let these jobs apply for exemptions, but this means a costly bureaucratic process, and is likely to end with large companies with good lawyers obtaining the exemptions, small companies with poor lawyers not obtaining the exemptions, and no concern about fairness to Martians in any case.

In the worst possible situation, a non-prejudiced boss may decide not to hire Martians because it would be harder to reprimand or dismiss a Martian when they could threaten to sue the company or start a viral Tumblr post accusing the company of speciesism.

Compare a market-informed solution: run a bunch of controlled studies in which bosses get identical Earthling and Martian resumes, find out exactly how strong the prejudice against Martians is, then levy an appropriate tax on hiring Earthlings (or give a subsidy for hiring Martians). Maybe hiring Earthlings costs 5% extra, which is funnelled into scholarships for impoverished Martian larvae.

Now there’s no question of a company wriggling out of their obligation – no matter how stylish their lawyers’ hair is, they’re going to pay the tax. There’s no question of lawsuits – if a company didn’t hire a qualified Martian, that’s their own business and the Martian community can laugh all the way to the bank. But on a *statistical basis*, we expect companies to be indifferent between hiring Martians or Earthlings.

Any company that has a legitimate reason to not want to hire Martians can just pay the (small) tax. And there’s no problem with firing Martians anymore – if you decide to fire the Martian in favor of an Earthling you like more, you’re perfectly welcome to do so as long as you don’t mind paying a little extra.

If ten years later the social scientists do some studies and find that companies are still more likely to accept Earthling resumes over identical Martian resumes, they can raise the tax until that’s no longer the case. If they find that companies are more likely to accept *Martian* resumes now, then prejudice has decreased and the tax can decrease as well.

I think everyone has a lot to like about this proposal. Martians can rest assured that with enough time to tweak the tax level, they will have a provably equal playing field in this area. Non-bigoted Earthlings can rest assured that they’re not going to be unfairly accused of bigotry and taken to court by some Martian playing the planetary origin card. And bigoted Earthlings who just really don’t like Martians – maybe someone’s father was killed by a heat-ray-tripod during the invasion and she’s had PTSD every time she sees Martians ever since then – can stand by their “principles” as long as they’re willing to pay a little extra.

(This is my answer to Jim’s question of “How many cities are you planning to burn, how many women are you planning to have raped with large objects in order to achieve equality of opportunity?”, which I honestly have to admit is not a question I ever really considered before reading Jim’s blog)

Someone will object that small fees can’t eliminate as pervasive a social problem as prejudice, but I’m not so sure. Consider the Islamic Caliphate (7th – 12th century AD). Their modus operandi was to march into a new territory, tell the non-Muslims there that they were *perfectly welcome* to continue to practice their old religion as long as they paid a tax, and if they ever wanted to save those couple shekels or dinars or whatever, they could also convert to Islam – but no pressure. The current religious makeup of the Caliphate territory (Northern Africa and the Middle East through Iran and Pakistan) should be taken as some evidence of the effectiveness of this policy.

V.

In my opinion the biggest advantage of a market-based system for improving social welfare is that it allows more flexibility – it leaves your options open.

Suppose the government, noticing mercury is toxic and has few good industrial uses, bans the use of mercury in industry.

A month later, some chemist discovers a really really lucrative industrial application for mercury that will make billions of dollars and cut the price of automobiles in half.

Probably this chemist can’t single-handedly convince the government to relax its views on mercury. She could consider selling her idea to a really big company like Dow Chemical, who could afford the necessary lobbying. But then she’s lost the ability to profit from her own invention, and we’ve replaced what could have been a nimble startup with yet another Dow product that they’ll overprice and destroy a couple of Indian villages producing. And the brilliant scientist becomes a mid-level drone working for morons in suits.

Or maybe it’s worse than this. Maybe she goes to Dow, but they don’t want to take the time to understand this fringe idea. Or maybe Dow is mildly interested but not interested enough to throw all its lobbyists and lawyers at the problem. Or maybe Dow *does* throw all its lobbyists and lawyers at the problem, but the Sierra Club reasonably believes that this is just another evil company trying to gut vital environmental legislation, and successfully blocks them. Sure, Dow says “This will halve the price of automobiles!”, but they probably make grandiose claims about *all* of their products when they're trying to look good in front of the government.

So suppose that instead of banning mercury, the government just places a tax on it. The tax could be the cost of mercury cleanup, it could be enough money to treat and emotionally compensate mercury poisoning sufferers, or it could just fund public health programs that do more good than fighting mercury ever could. It could be all these things combined plus a little extra. Let's say the tax on mercury is 500%. Every company that has any possible alternative to mercury switches to that alternative. The companies that have no alternative to mercury close down if the benefit of their product to society is less than the cost of the mercury they produce. And the companies that use mercury in a way that net benefits society stay open and subsidize lots of environmental and public health programs.

Now the chemist who discovered the brilliant unexpected use for mercury is able to start her startup – at increased costs, sure, but if it's as lucrative an idea as she thinks she'll be able to get the investment or just swallow the losses. In any case, it's nothing compared to the cost of pushing around an entire government agency. The price of automobiles decreases by half, the taxes are more than enough to clean up the mercury and improve public health, and everyone is happy.

VI.

The problem with banning and regulating things is that it's a blunt instrument. Maybe before the thing was banned someone checked to see whether there was any value in it, but if someone finds value after it was banned, or is a weird edge case who gets value out of it even when most other people don’t, then that person is mostly out of luck. Even people operating within regulations have to spend high initial costs in time and money proving that they are complying with the regulations, or get outcompeted by larger companies with better lobbyists who can get one-time exceptions to the regulations.

In short, the effect is to decrease innovation, crack down on nontypical people, discourage startups, hand insurmountable advantages to large corporations, and turn lawsuits into the correct response to everything.

The problem with not banning and regulating things is that the rivers flow silver with mercury, poor people starve in the streets, and Martians get locked out of legitimate industry and are forced to turn to threatening innocent cities with their heat rays just to get by.

The position there’s no good name for – “bleeding heart libertarians” is too long and too full of social justice memes, “left-libertarian” usually means anarchists who haven’t thought about anarchy very carefully, and “liberaltarian” is groanworthy – that position seems to be the sweet spot between these two extremes and the political philosophy I’m most comfortable with right now. It consists of dealing with social and economic problems, when possible, through subsidies and taxes which come directly from the government. I think it’s likely to be the conclusion of my long engagement with libertarianism (have I mentioned I only engage with philosophies I like?)

Papal PR

Congratulations are due to the Pope for being Time's Person of the Year. I would have voted for Snowden, but the Pope is a perfectly respectable choice. At least he's an individual human being and not a vague demographic or second-person pronoun.

I like Pope Francis. He seems to be kind-hearted, intelligent, and genuinely focused on helping the poor. When he expresses opinions, they tend to be ones I agree with, at least as much as is consistent with him still being Catholic. He's done a lot of substantive good work.

But that's not what anybody's focused on. And what they do focus on confuses me.

Suppose that people were to get super excited about President Obama wearing a US flag pin. "He's so patriotic! He even demonstrates his love for America on his clothing!" Or going to Arlington Cemetary on Memorial Day: "Look how much the sacrifice of American soldiers moves him!"

But really all this proves is that Obama isn't a total idiot. Wearing a flag pin is an easy way to signal patriotism, and if you're the President signaling patriotism is a free public opinion boost. In the same way, it's hard to imagine a self-interested President conceiving of the idea of going to Arlington Cemetery on Memorial Day and *not* acting on it. It would just be dumb.

Pope Francis does some pretty heart-warming things, like washing the feet of the poor and letting little kids sit in his Pope chair, abandoning his papal luxury palace for a humble apartment, and baptizing orphaned puppies (has he done this yet? I assume it's only a matter of time).

Thing is, I think of the counterfactual universe where I'm Pope, and having been made aware of the possibility of doing these things, it's hard to imagine not going through with them. For getting rid of the furniture made of solid gold that no sane person would even want, I can have the entire world talk about my humility. For two hours of my time and the cost of a foot-washing basin, I can be Time Person of the Year.

For me, the story isn't why Francis is so great, but why his predecessors didn't do stuff like this all the time. Yes, okay, Pope Benedict XVI gets creeped out by humans and he's not sure his immune matrix can handle their Earth germs. Fine. What were the other 264 guys' excuses? Why don't we get to hear about them secretly sneaking out of the Vatican to help the homeless in a way that makes it almost certain it would leak out (I'm not saying Pope Francis is really doing this, just that if he isn't it's probably because he didn't think of it in time)?

Maybe Francis is just the first Pope who understands PR. The past two Popes were born in the 1920s; maybe they never really figured out the Age of Mass Media. I doubt future Popes will make that mistake.

I am not accusing Pope Francis of being shady or Machiavellian (although Machiavelli's The Prince *does* in fact contain a whole chapter on advice for Popes). Just saying that if he *were* Machiavellian, he'd probably do pretty much what he's doing now.

EDIT: Jed brings up in the comments what seems to me a good point. Perhaps previous Popes *were* interested in public opinion, and *did* do a good job managing it, but believed that people would be more impressed by golden thrones and fancy regalia and ritual than by conspicuous humility. Perhaps that belief was correct. That would say something pretty impressive about the world and the Church – that over the past century the optimal strategy in making people think you are a holy figure worthy of respect has changed from “have a really big solid gold scepter” to “radiate humility and love for all mankind”. Did we enter the Millennium without noticing?

Statistical Literacy Among Doctors Now Lower Than Chance

Good news! 42% of doctors can correctly answer a true-false question on p-values! That's only 8% worse than a coin flip!

And this paragraph is your friendly reminder that six months after this study was published, the FDA decided it was unsafe for individuals to look at their own genome since they might misunderstand the risks involved. Instead, they must rely on their doctor. I am sure that statisticians and math professors making life-changing health or reproductive decisions feel perfectly confident being at the mercy of people whose statistics knowledge is worse than chance.

Now that I've got the sensationalism out of the way, let's look at this study more closely.

The sample is 4000 Ob/Gyn residents. Ob/Gyn is a prestigious specialty that's able to select people with very good grades in medical school, so we're not looking at dummies here. These residents (beginning doctors) did a bit worse than more experienced doctors (whose performance was still not stellar). I don't know whether this reflects doctors learning more about statistics as they progress, better statistical education in Ye Olde Days than in the current generation, or both.

The study looked at two questions. First was the one I mentioned above: "True or false: the p-value is the probability that the null hypothesis is correct". The correct answer is "false" – the p-value is the chance of obtaining results at least as extreme as those actually obtained if the null hypothesis were true. 42% correctly said it was false, 46% said it was true, and 12% didn’t even want to hazard a guess.

The question seems sketchy to me. It is indeed technically false, but it seems pretty close to the truth. If I were asked to explain why the definition as given was false, the best I could do is say that your probability of the null hypothesis being true should take into account both something like your p-value, and your prior. But since no one ever receives Bayesian statistical education, I am not sure it is fair to expect a doctor to be able to generate that objection. What I would want a doctor to know is that the lower the p-value, the more conclusively the study has rejected the null hypothesis. The false definition as given accurately captures that key insight. So I’m not sure it proves anything other than doctors not being really nitpicky over definitions.

(which is also false, actually)

Next came very nearly the exact same question about mammogram results as Eliezer’s Short Explanation Of Bayes Theorem. It offered five multiple-choice answers, so we would expect 20% correct by chance. Instead, 26% of doctors got it correct. What shocks me about this one is that the question very nearly does all the work for you and throws the right answer in your face. Compare the way it was phrased in Eliezer’s example:

> 1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammographies. 9.6% of women without breast cancer will also get positive mammographies. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

to the way it was phrased on the obstetrician study:

> Ten out of every 1,000 women have breast cancer. Of these 10 women with breast cancer, 9 test positive. Of the 990 women without cancer, about 89 nevertheless test positive. A woman tests positive and wants to know whether she has breast cancer for sure, or at least what the chances are. What is the best answer?

The obstetrician study seems to be doing everything it can to guide people to the correct result, and 74% of people still got it wrong. And nitpicky definitions don’t provide much of an excuse here.

There were three other results of this study worth highlighting.

First, people who got the statistics questions wrong were *more* likely to say they had good training in statistical literacy than those who did not, giving a rare demonstration of the Dunning-Kruger effect in the wild. Doctors who didn’t know statistics were apparently so inadequate that they didn’t realize there was any more to know, whereas those who did know some statistics at least had a faint inkling that something was missing.

Second, women rated their statistical literacy significantly worse than men did (note that a large majority of Ob/Gyn residents are women) but did not actually do any worse on the questions. This highlights an important limitation of self-report (tendency to confuse incompetence with humility) and probably has some broader gender-related implications as well.

And third, even though 42% of people got Question 1 correct and 26% of people Question 2, only 12% of people got both questions correct. Just from eyeballing those numbers, it doesn’t look like getting one question right made you much more likely to do better on the other. This is very consistent with most people lucking in to the correct answer.

I do not want to use this to attack doctors. Most doctors are technicians and not academics, and they cultivate, and should cultivate, only as much statistical knowledge as is useful for them. For a technician, “a p-value is that thing that gets lower when it means there’s really strong evidence” is probably enough. For a technician, “I can’t remember what exactly the positive predictive value of a mammogram is but it doesn’t matter because you should follow up all suspicious mammograms with further testing anyway” is probably enough.

But it really does seem relevant that only 12% of doctors can answer two simple statistics questions correctly when you’re trying to deny the entire non-doctor population access to certain information because only doctors are good enough at statistics to understand it.

Less Wrong, More Rite (II)

Every year people mock the Rationalist Solstice Ritual. And every year I read the discussions, nod, and then go to New York and have a great time and meet amazing people and have fascinating discussions and get the songs stuck in my head for the next couple of weeks.

(*"Stop enjoying yourself! Your enjoyment is wrong!"*)

I.

The detractors keep calling it "cringeworthy". If anything, I'm worried it isn't cringeworthy *enough*. Last year we sang Zack Davis' Contract-Drafting Em Song, which is *brilliant* but only if you're really familiar with Robin Hanson's futurology and have a very special sense of humor. Also that one song from Portal. Now *that's* cringeworthy.

This year's ritual was very polished in a lovely auditorium (the sign in front said "Winter Solstice Concert" and I felt a brief sense of disorientation – concert? How did I get tricked into going to one of those?) It was billed as appropriate for people across the atheist community and supported by the Center For Inquiry.

And it worked. There were a few moments when I didn't think that it would, but Raemon managed to pull it together and still help it keep its old flavor despite the added sophistication.

But it got me thinking. A lot of religion is pretty cringeworthy. And the stronger the religion, the more cringeworthy it tends to be. Unitarianism is very nearly cringeless and just says nice things pretty much everyone can agree with – but they are not exactly known for the burning ferocity of their faith. Standard-model Christians have weirder beliefs than Unitarians and also a somewhat stronger community. Mormons have weirder beliefs than standard, and their community is stronger still. Scientology maxes out both belief weirdness and insularity. I may be cherry-picking, and I realize there’s a lot of potential to introduce bias in variables like “strength of community”, but I think there’s a relationship.

It reminds me of Steven Bond’s Objects of Fandom. I’d previously read only the Less Wrong discussion thereof, so it was a surprise to double-check the original source and discover that yes, he had mentioned religion.

> What makes something an object of fandom? There are some works that attract obsessive fannish behaviour and some that don’t — and this seems to be independent of popularity. To take an example from LucasProduct, both Raiders of the Lost Ark and Star Wars are hugely popular movies, but only the latter has developed a significant fan base. Why? My theory is that for something to attract fans, it must have an aspect of truly monumental badness about it.
>
> Raiders of the Lost Ark is a robust potboiler, tongue-in-cheek, very competently done. I think it’s enjoyable, but even among those who don’t, it’s hard to see the film attracting actual derision. Boredom or irritation, probably, but nothing more. Star Wars, on the other hand…. From one perspective, it’s an entertaining space opera, but from a slightly different perspective, an imperceptible twist of the glass, it’s laughably awful. Utterly ridiculously bad. And it’s this very badness that makes so many people take up arms in its defence…

> And the same again goes for religion, the ultimate expression of fandom. If Christianity, say, were just a list of moral precepts, it most probably wouldn't have any adherents — just admirers. But add in nonsense like the resurrection and the Trinity, the creation and the apocalypse, the angels and the saints, and you get that monumental awfulness that millions will die for.

I will be less judgmental than Bond. I think the active ingredient is not awfulness *per se*, but something more like *separation*. That is, to strengthen in-group ties, simply differentiate yourself from the out-group. The more radical the differentiation, the better. A Unitarian who speaks out in favor of everyone getting along has minimal difference from the out-group; heck, they probably agree with her, even if they don't consider themselves Unitarians *per se*. A Scientologist who speaks out in favor of ancient aliens being chained to volcanoes has successfully burnt pretty much every bridge with the out-group.

But it's not just about believing unusual things. It's also about having such a different way of looking at the world from everyone else that even if you want to have a conversation, it's barely even worth it. A friend in New York with whom I discussed this brought up how the Scientologists both replace normal-person vocabulary with Scientologist vocabulary (so that souls become thetans) but also replace normal-person assumptions with different bases (a Scientologist believes that a person's problems are based on traumas suffered by zir thetans in past lives, which leads to radically different problem-solving strategies than most people).

Atheists are kind of in a rut when they want to build community. It's really easy to be not-religious without identifying as an atheist, let alone becoming part of some atheist group. There are lots of obstacles preventing atheists from joining communities – most atheists are individualist by nature, atheists are very scattered geographically, atheists differ wildly in all their other beliefs from libertarian to Communist – and there is no a priori reason why atheists should want to get together.

The few raw materials they have for community-building just aren't good enough to overcome these obstacles. I feel less akin to an atheist Marxist than I do to a politically centrist Reform Jew. And I feel less akin to an atheist basketball player than I do to a Methodist doctor. Even though there are some traits typical of atheists that I want in my social interaction partners – intelligence, rationality, interest in science – there are already pretty good social sorting mechanisms for finding those people that have nothing to do with religion – and atheism is far from a perfect guarantee of these traits anyway. In other words, the out-group separation that Mormons and Scientologists enjoy – "I could never associate with a member of the out-group, there would just be too big a gulf between us" – is entirely absent, and with that absence vanishes any strong "patriotism" for the in-group.

I would now like to irresponsibly speculate on how the modern atheist movement is shaped by different strategies to solve this problem.

First, ignoring the problem and being content with weak and insipid communities, which is not a bad idea if you just want an atheist meetup every couple of months.

Second, making atheism appear unpopular and persecuted. In areas where atheists *are* unpopular and persecuted, this is easy. In other areas, it seems to involve a lot of effort: exaggerating how religious everyone else is, bringing up every available example of atheist persecution a hundred times, and conflating distant areas where atheists are persecuted with one's own situation. If everyone else is a raving creationist loony who believes in alternative medicine and ancient aliens and wants the Ten Commandments put up in all public places, atheists have their out-group separation right there.

Third, dialing the atheism up to eleven until it becomes genuinely unpopular. If saying religion is bad doesn't work, say that religion is responsible for nearly all the world's problems, that all religious people are culpable without exception, that every single vaguely religious public monument or institution needs to be torn down. Congratulations, you are now unpopular with the out-group.

Fourth, linking your atheism to another belief that is much more specific than atheism. For example the atheist movements in favor of transhumanism, social justice, and rationality (libertarianism and efficient charity may or may not qualify). These submovements have been quite successful within atheism, to the point where to many atheists they are much more important than the atheism itself. Each has its own complicated vocabulary, each makes much broader critiques of mainstream culture than basic atheism, and each is at least somewhat unpopular (rationality and social justice are not unpopular *per se*, but most people wouldn't be on board with the way we handle them).

I am not as affiliated with the other submovements, but I think the rationalist movement has been especially blessed. Without really meaning to, and without having to deliberately believe cringeworthy things for the heck of it or unfairly attack others, we have managed to develop enough different ways of thinking that we naturally have a very strong in-group distinction – which in turn means a very strong community. Last year, sitting around Raemon’s house singing the Contract-Drafting Em Song, I got this feeling of “There are only a few hundred people in the world who would possibly enjoy this *and they are my people and I love every last one of them*.

And then this year, even though the base was a *little* broader and the songs a *little* less insular, I still ended out thinking “These are people who are willingly going to an event called a ‘secular solstice ritual’, and they are singing songs which rhyme ‘rarity’ with ‘singularity’, and I am still pretty darned okay with them.”

(in our defense, it was ‘rarity’ the word meaning rare event, not the pony. Although really it’s only a matter of time.)

(speaking of people being united by outsider-scorn and cringeworthiness…)

II.

I will tack on to the end one more little revelation I had at the Solstice. In particular I had it when we were singing The Times They Are A Changin’

People are *always* talking about how much great art religion produces, usually in contrast to the poverty of atheism-associated art and music.

But atheists make up 2% of the US population. Atheists who identify with the atheist movement are a small subset of this small subset. And it may be even less than this among artists and musicians – although these groups are famously liberal and skeptical of religion, in my experience people who go all the way to self-identifying as “atheist” are more likely to be math and science geeks. And even when there are atheist musicians, there is no more reason to expect them to sing about atheism than there is to expect anti-tax musicians to sing about tax relief. I think the amount of good art about atheism is about what you'd expect given the null hypothesis that atheists create art about the same as anyone else.

But that's not the end of the story. When people ask "Where is the atheist version of Bach?" and Luke from Common Sense Atheism answers with a list of great atheist songs, I think he's missing the point. The "missing atheist music" isn't to be found in hidden gems of songs about atheism, it's to be found in songs like "The Times They Are A Changin" and "Blowin In The Wind" (both of which we sang at Solstice). It's to be found in our incredible corpus of brilliant songs written by nonreligious or ambiguously spiritual people about ideas like progress, pacifism, humanism, social justice, unity, of humankind et cetera that neither make nor need any reference to God.

(yeah, I know Dylan briefly became a born-again Christian. He wasn't one when he wrote those songs, and he changed his mind later, and having weird spiritual crises every so often is itself a kinda atheist thing to do.)

(and in terms of explicitly atheist music, The Words of God is still pretty awesome in both the modern and traditional senses of the word)

I wrote a list of people I enjoyed meeting at Solstice, and as always two minutes after it was published I thought of a dozen people I'd been thrilled to see whom I had embarassingly left out. I'm not even going to attempt to try again here. I'll just thank Raemon one more time and suggest that anyone on the fence about going to Solstice next year should give it a try.

(I understand there were franchise Solstices in the Bay Area and a few other places this year. I am *intrigued* by the possibility of trying to run a Solstice franchise in Michigan next year if our community expands, perhaps the weekend after the one in New York to allow me to go to both, but I make no promises)

Public Awareness Campaigns

A little while back I discussed how, contrary to the conventional wisdom that they've been "proven to work", anti-rape campaigns aimed at men have zero evidence of effectiveness. I added that this was no surprise, since similar public awareness campaigns have a long history of failure.

That was overly simplistic, as commenters quickly reminded me. Some public awareness campaigns (or things like public awareness campaigns) have a history of spectacular failure. Others have a history of spectacular success. To give a couple of examples:

Failures

– DARE (Drug Abuse Resistance Education) is a program in US schools where teachers and police officers spend a couple of hours a week telling children how bad drugs are and giving them techniques to avoid peer pressure to use them. It famously fails to work and in some cases even makes children more likely to do drugs.

– Scared Straight is a popular program in which convicted inmates talk to deliquent kids and explain the costs of a life of crime and how unpleasant prison can be. Several studies clearly show that the intervention makes these children actively more likely to become criminals, and is so harmful that "each dollar spent on Scared Straight programs incurs costs of $203.51". Needless to say there continue to be dozens of these programs all around the country.

– Sex education in schools is famously ineffective. There is a very large body of research showing abstinence-only sex education programs do not make teens less likely to have sex. Research on comprehensive (ie contraception-including) sex education programs is more mixed, and although everyone trumpets the positive results in order to further discredit the abstinence-only programs by comparison, the actual research is more nuanced and less optimistic. The limited effects it does get may work by spreading genuinely novel information ("condoms exist! STDs exist!") rather than "awareness raising" per se.

– "Diversity training" and "sensitivity workshops" and everything in that category of thing have no positive effects are are often associated with negative effects (for example, after being introduced in companies those companies' percent of top executives who are minorities goes down)

– A couple of recent studies (1, 2) are converging on the hypothesis that stigmatizing overweight people makes them *more* likely to gain weight. This is true even when the stigma is delivered in the form of a (presumably more polite) public awareness campaign instead of just an acquaintance calling you a lardass. Although I agree a line can be drawn between "public awareness campaign" and "stigma", in practice it can sometimes be kind of fuzzy – for example, although most people wouldn't use the word, it sure *seems* like the point of "Don't Be That Guy" anti-rape campaign is to stigmatize rape.

Successes

– Several people have brought up MADD's campaign against drunk driving, which corresponded to a 65% decrease in drunk driving over the past 30 years. However, I can't find good evidence on whether MADD started a traditional public awareness campaign or just lobbied for changes in various laws and got lots of publicity in the process. I would also note that there have been spectacular and somewhat mysterious decreases in nearly all crimes since 1982 – alternately attributed to rising abortion rates, falling lead rates, and stricter sentencing – and it's not obvious how much drunk driving is just piggybacking on that success.

– Seat belt use has gone from very low to near-complete, and there is decent evidence that awareness campaigns like Click It Or Ticket contributed to this (fun fact: opponents of mandatory seatbelt laws launched a counter-campaign called "Stick It To Click It Or Ticket").

– Advertising is kind of like a "public awareness campaign" about a particular product. It obviously works or else companies wouldn't spend so much money on it.

– Anti-smoking campaigns do seem to lead some people to stop smoking – or at least increase calls to stop-smoking hotlines. These are most effective when associated with scary and graphic images – for example, one shows a picture of a man with a hole in his throat after a throat cancer operation. There are a *lot* of successful public health campaigns along these lines.

Analysis

It's pretty hard to draw a consistent "this works, that doesn't" conclusion from these facts.

Just to give an example, one of the most effective campaigns – anti-tobacco – uses the same strategy as one of the least effective campaigns – Scared Straight. Both try to present very graphic images of the horrible things in store if people do not change their ways. One works great, the other is counterproductive.

To give another, both anti-obesity and anti-drunk-driving campaigns try to employ stigma, but one of them has been very successful and the other has if anything the opposite of the intended outcome.

Some people I talked to about this at the New York Solstice Celebration suggested that product advertising works because businesses have financial incentives to get it right, but other public awareness campaigns don't because the government mostly wants to signal virtuous effort and has no incentive to design genuinely effective advertising change minds. But some government and nonprofit public awareness campaigns are successful, and I'm betting they're all hiring the same ad agencies anyway.

Another theory was that awareness campaigns work when there's a real need for awareness – either the target demographic is literally unfamiliar with the concept in question (for example, people may not have previously been aware the police were cracking down on non-seatbelt-users) or need reminders to keep an option fresh in their minds (Coke advertisements making everyone think about Coke when they're deciding what to buy). If it's just stating *ad nauseum* that some stigmatized action like premarital sex is still stigmatized, it's not going to do much. But this is disproven by the success of MADD and the stop-smoking campaign, and by the failures of DARE (which very often does teach kids things about drugs they didn't know before).

The biggest effect I can see is that anything which caters to a captive audience is more likely to be counterproductive. DARE and sex-ed are inflicted on schoolchildren who would rather be doing something else, and a lot of the time it ends up as "this uncool authority figure I don't like lectures about how me and all my friends are bad people". Scared Straight programs are usually court-mandated, often as a punishment for past delinquent behavior. Employees are forced to attend diversity training, and once again it may be billed as a "punishment" for saying something politically incorrect. Is it so far-fetched that people forced to suffer through these campaigns will end up resentful, and that resentment will translate into negative feelings about the campaign message?

I'd like to extend the theory to the obesity case and say that, once again, stigmatizing the obese in anti-obesity campaigns causes obese people to associate the negative feelings they get from these campaigns with "eat less and exercise" message. But this proves too much: why wouldn't the scary disfigured people in the stop-smoking ads make smokers associate their negative feelings with quitting? Perhaps the negative feelings have to be of a certain type for this to work? Anger and resentment, rather than fear and disgust? Questions, questions.

I still don't feel like I have a good ability to predict the success or failure of any future public awareness campaign. If I wanted to promote the "Don't Be That Guy" anti-rape campaign, I would point out that it consists mainly of flyers on lampposts etc, so there's no captive audience nor any reason to consider it a "punishment". If I wanted to inveigh against it, I'd argue that empirically it offends a whole lot of men who think they're being binned as potential rapists and so definitely causes the anger and resentment which are the hallmark of a counterproductive campaign. I really don't know.

I guess part of the reason I remain skeptical of public awareness campaigns is a lingering terror at what it would say about society's collective sanity if they worked. Think about it. Imagine that TV ads warning people not to do drugs *really* decreased drug use. Then think about how for the past 30 years, we haven’t been consistently running those ads, but we *have* been consistently putting anyone who uses drugs behind bars for their entire lives. Imagine if anti-rape ads worked, and *only two Canadian cities have ever run them*. And *no* city afaik has ever run ads against child abuse!

I would welcome more examples of public awareness campaigns that clearly succeeded or failed. Post them in the comments. Please exclude ones that measure “success” by surveying people about whether they saw the campaign or became aware of the campaign message – I’m interested in ones that actually change behavior.

We Are All MsScribe

AskReddit asked recently: If you could only give an alien one thing to help them understand the human race, what would you give them?

At the time I had no good answer. Now I do. I would give them Charlotte Lennox's write-up of how MsScribe took over Harry Potter fandom (warning: super-long but super-worth-it).

Ozy informs me that everyone else in the world read this story five years ago. Maybe I am hopelessly behind the times? Maybe all my blog readers are intimately familiar with it?

If not, read it. Read it like an anthropological text. Read it like you would a study of the Yanomamo. No, read it even better than that. Read it like you would a study of the Yanomamo if you knew that, statistically, some of your friends and co-workers covertly become Yanomamo after getting home every evening.

I hesitate to summarize it, because people will read my summary and ignore the much superior original. I would not recommend that. But if you insist on skipping the (admittedly super-long) link above, here is what happens:

> In the early 2000s, Harry Potter fanfiction authors and readers get embroiled in an apocalyptic feud between people who think that Harry should be in a relationship with Ginny vs. people who think Harry should be in a relationship with Hermione. This devolves from debate to personal attacks to real world stalking and harassment to legal cases to them splitting the

community into different sites that pretty much refuse to talk to each other and ban stories with their nonpreferred relationship.

These sites then sort themselves out into a status hierarchy with a few people called Big Name Fans at the top and everyone else competing to get their attention and affection, whether by praising them slavishly or by striking out in particularly cruel ways at people in the "enemy" relationship community.

A young woman named MsScribe joins the Harry/Hermione community. She proceeds to make herself popular and famous by use of sock-puppet accounts (a sockpuppet is when someone uses multiple internet nicknames to pretend to be multiple different people) that all praise her and talk about how great she is. Then she moves on to racist and sexist sockpuppet accounts who launch lots of slurs at her, so that everyone feels very sorry for her.

At the height of her power, she controls a small army of religious trolls who go around talking about the sinfulness of Harry Potter fanfiction authors and *especially* MsScribe and how much they hate gay people. All of these trolls drop hints about how they are supported by the Harry/Ginny community, and MsScribe leads the campaign to paint everyone who wants Harry and Ginny to be in a relationship as vile bigots and/or Christians. She classily cements her position by convincing everyone to call them "cockroaches" and post pictures of cockroaches whenever they make comments.

Throughout all this, a bunch of people are coming up with ironclad evidence that she is the one behind all of

> this (this is the Internet! They can just trace IPs!) Throughout all of it, MsScribe makes increasingly implausible denials. And throughout all of it, everyone supports MsScribe and ridicules her accusers. Because really, do you want to be on the side of a confirmed popular person, or a bunch of confirmed suspected racists whom we know are racist because they deny racism *which is exactly what we would expect racists to do?*
>
> MsScribe writes negatively about a fan with cancer asking for money, and her comments get interpreted as being needlessly cruel to a cancer patient. Her popularity drops and everyone takes a second look at the evidence and realizes hey, she was obviously manipulating everyone all along. There is slight sheepishness but few apologies, because hey, we honestly thought the people we were bullying were unpopular.

MsScribe later ended up switching from Harry Potter fandom to blogging about social justice issues, which does not surprise me one bit. But let me do some social justice blogging of my own.

A lot of the comments I have seen discussing the issue say “Yeah, teenage girls will be teenage girls”.

Two responses seem relevant. First, quite a few of the people involved seem to have been in their late twenties or early thirties.

But second and more important, I am a guy and this story speaks to me because it is *eerily* similar to the story of my online life with a bunch of other guys when I was between about ages fifteen and twenty-two.

I’ve mentioned before how I spent long portions of my life in the interactive geofiction/”micronation” community. And because of the innate urge for self-presentation, I emphasized the part where we create amazing grand-scale fictional universes in which we enact epic battles and build civilizations from the ground up. And not the part where we behave like ridiculous little children having a hissy fit.

The first constructed country I was ever in, another guy named John from my school comes in and says that I am a bad leader and abusing my power. Because my online handle at the time was Giant_Squid314, he classily nicknames me “Squitler” and leads a bunch of his supporters to make “Squitler” related comments at everything I do. Then he and his friends secede to start their own country, named after a Red Hot Chili Peppers album. I retaliate by convincing his friends that he is oppressing them and they need to start a communist revolution to kick him out of the country, which works. Later he gets back in and convinces his friends to join my country under fake names, swelling the ranks of voters with people who are there just to vote for the worst policies in order to destroy the country. This becomes so bad that my friend Evan pulls a bloodless coup to abolish democracy and make himself sole leader, but then he cracks down so hard on John’s supporters that everyone gets upset and leaves (“emigrates”). This upsets my friend Bill, who somehow hacks John and tries to delete all his stuff; John counterhacks Bill and destroys his country. Then we all team up with a bunch of guys from Ireland, infiltrate John’s country and destroy it the same way he destroyed us as an act of revenge.

All this happened within about three months real-time, and I was in this hobby for ten years. *Ten years*.

There was an entire era when people would accuse other people of having said racist things on IRC (where logs were often unavailable, and context was absent). This would then be followed with the demand that every political ally of the affected person shun him forever and kick him out of the country and destroy every institution he had built, or else *obviously* they were secretly racist themselves. This was met with the only possible response: “actually, no, *you’re* the one who said racist things on the chat!”. These accusations often resembled the MsScribe story in their sheer not-entirely-social-justice-movement-approved incongruity: “You’re racist, and you’re a fat lardass!” “Oh yeah? Well you’re a f**king homophobic autistic Aspie who will never get laid!” Inevitably the more popular person would win and anyone so foolish as to defend the unpopular person (which I *kept doing*, because I never learn) was banished to Racist Hell. As for Actual Hell, there was a guy named Archbishop Fenton who kept saying really extreme Christian stuff about how we were all going there, and although we all suspected he was a sockpuppet I was never able to figure out whose.

So MsScribe? I’ll give her this: she was a gifted amateur. That is it. An amateur. We had frickin’ decade-old “intelligence organizations” whose entire job was to collect a network of spies – some real people, some sock puppets – who would join other people’s countries under fake (or real!) identities, get information on their secret plans, and throw important elections in favor of the parties we supported. I’m not even ashamed of my role leading one of the largest of these organizations, Shireroth’s spy bureau S.H.I.N.E. – if we had unilaterally disengaged from these kinds of games, we would have been demolished by people who didn’t.

I remember our scandals. We would build up “dossiers” on various individuals, then publicize them at times calculated to cause maximum damage. One of my favorite was when a prominent female politician was revealed to in fact be male – causing her support to plummet among the key “people who do whatever a girl says in the hopes that she will like them” demographic. In another, which happened a bit after my semi-retirement, the micronational world’s largest communist country, with thirty highly active citizens and a prominent international role, was found to be just one guy posting under thirty different names.

As leader of an espionage organization, I was expected to be able to avoid these damaging revelations, advise my countrymen on how to do the same, and run circles around my enemies. Without tooting my own horn too much, I maintained my most successful character for the better part of a year. This was a guy named Yvain, who infiltrated a Celtic-themed fantasy state called the Duchy of Goldenmoon, took it over, took over its largest neighbor, and was halfway to ultimate power over the entire continent before I got accepted to medical school and decided I should probably reassess how I was using my time.

(to create a paper trail and avoid breaking character, I used the nick “Yvain” for a lot of the websites I joined around this period, which is why half the Internet *still* knows me by that name. I am suitably embarassed by this)

Now I will say this for us boys – and we were boys, like 95% of us, and even the girls were usually found to be boys after careful investigation. We did it with class, we did it with cool names like “Paramountgate” and “The Three Hours’ War”, we wrote up our petty scandals into epic history books with bibliographies and appendices, and we backstabbed each other so elegantly it would make Machiavelli shed a single tear of pure joy. But in the end? We behaved *exactly* like teenage girls in a Harry Potter fandom.

It is hard at this point not to be reminded of the Robbers' Cave experiment. Social psychologists divided boys at a camp into two groups, intending to do some experiments in order to figure out what they needed to do to make the groups hate each other, only to learn that the boys had *already* started hating each other with the burning fire of a thousand suns while they were busy planning the experiments. They boys had even formed little group identities, like "Our group are the rough and street-smart ones, the other group is a bunch of holier-than-thou goody-goodies" (the groups were chosen at random).

I read a lot of psychology even as a teenager, so it never surprised me that separating people out into different fictional countries would have the same effect.

But it did kind of surprise me that you could get *quite* those depths of hatred between people who thought that a fictional wizard should hook up with his best friend, versus other people who who thought he should hook up with his other best friend's little sister. Every time I feel like my opinion of people is sufficiently low, I get new evidence making me bump it lower.

Anyway, once those depths of hatred are established, they will proceed in the same way among twenty-somethings trying to discuss Harry Potter romantic pairings, teenagers trying to run fictional countries, and Senators trying to pass vitally important legislation. And that's why, if aliens ever requested exactly one item to teach them about the human race, I would give them the MsScribe story.

They'd kill us all, of course. They would sterilize Earth so thoroughly that not even the archaeobacteria would remain. But in the moment before I was vaporized, I would feel like our species had finally been *understood*.

[REPOST] A Christmas Poem

Every Jew down in Jewville liked Christmas a lot
But King Herod, who ruled over Jewville, did not.

The King had a thousand and one reasons why
He hated the stars that appeared in the sky!
He hated the wise men! He hated the myrrh
And the frankincense also, whatever they were!

He hated the shepherds! He hated the manger!
He hated the way that his throne was in danger!
He hated the prophecies bearing the news
He would soon be replaced as the King of the Jews
Yes, the number one reason the King felt so down
Was the tales of Messiahs usurping his crown.

So he thought and reflected, he planned and he schemed
He wondered and pondered and plotted and dreamed
He came up with a plan! Such a terrible plan!
To stop the Messiah before he began.
He would steal all the innocent babies away
And with Christ gone for good, he would end Christmas Day

So he searched through his palace, the front and the back
Till he found a warm coat and a Very Big Sack
And when night fell in Jewville, his men searched the houses
As fast as the birds and as quiet as mouses
And they gathered the children, and brought them all back
And the king placed them all in his Very Big Sack.

The King hooted and laughed as they finished their mission
“Mwa ha ha!” he exclaimed, as his plans reached fruition
“Now all of the children are here in my sack
And I don’t plan on giving a single one back!

I’ve stolen their toddlers, I’ve stolen their babies!
I’ve got their Messiah, no ifs, buts, or maybes!
They can stop all their songs and their feasts and their fun
Because Christmas is OVER and FINISHED and DONE!”

But the King heard a sound coming over the sand
And he jumped in surprise, and could not understand
It was singing, and laughing, and feasting, and fun
It was families dancing in joy, every one

King Herod turned red at their festive behavior:
“How can Christmas still come, when I’ve stolen the Savior?
It came without Bibles! It came without churches!
It came without Wise Men embarking on searches!
It came without sermons and prayers, and moreover
It came without Heaven and Hell and Jehovah!”

Then a very strange thought made his heart rise and fall
What if Christmas was not about Jesus at all?

What if Christmas, he thought, didn’t come from a priest?
What if Christmas was not about that in the least?

And what happened then? Well, in Jewville they say
That King Herod’s small heart grew three sizes that day!
And the minute his heart didn’t feel quite so tight
He returned all the children he’d taken that night
And he ordered his cooks to bring food for the feast
And King Herod himself carved the course of roast beast!

[I recently locked my old blog and am gradually reposting some of the stuff that deserves to survive from there onto here. This seemed like an appropriate start.]

Meetup: Southeast Michigan

There will be a Less Wrong meetup at my place next Saturday. Out-of-town guests Alicorn, Mike Blume, and Ozymandias will be attending. You can find more information here.

A Comment I Posted On “What Would JT Do?”

Last week JT posted An Open Letter To The Defenders Of Phil Robertson, which bothered me enough that I posted the following:

> So I’m the person you are insisting doesn’t exist – a completely pro-gay atheist who voted against Proposition 8 and thinks supporting gay marriage is a no-brainer, but who is also kind of horrified at Phil Robertson being fired for his comments.
>
> You are 100% correct that freedom-of-speech only binds the government and does not constrain private actors from punishing people whose speech they don’t like.
>
> But let’s compare and contrast. Freedom of religion *also* only binds the government and does not constraint private actors from punishing people whose religion they don’t like. If someone wants to picket a mosque while waving signs about how all Muslims are dirty terrorists who are going to Hell, Constitutional freedom of religion is a-ok with that. Heck, Constitutional freedom-of-religion is okay with Christian-owned businesses refusing to hire atheist employees or serve atheist customers – it’s only more recent anti-discrimination laws that prevent that.
>
> Point is, there’s a big gap between “constitutional freedom of religion” and “the level of religious tolerance that is necessary to have a remotely civil society.” Some of that gap can be filled in by laws, but a lot of it can’t be. It’s supposed to be filled in by basic human decency

and understanding of the principles that made freedom of religion a good idea to begin with.

I think the same is true of freedom of speech. Constitutional freedom-of-speech is a necessary but not sufficient condition to have a “marketplace of ideas” and avoid de facto censorship. But people also have to understand that the correct response to “idea I disagree with” is “counterargument”, not “find some way to punish or financially ruin the person who expresses it.” If you respond with counterargument, then there’s a debate and eventually the people with better ideas win (as is very clearly happening right now with gay marriage). If there’s a norm of trying to punish the people with opposing views, then it doesn’t really matter whether you’re doing it with threats of political oppression, of financial ruin, or of social ostracism, the end result is the same – the group with the most money and popularity wins, any disagreeing ideas never get expressed.

Atheists may one day be the group with the most money and popularity, but that day isn’t today and right now it’s neither moral nor in our self-interest to encourage using greater resources to steamroll opponents. It’s certainly not in gay people’s self-interest either. Why shouldn’t companies owned by Christians fire all gay people on the grounds that they are promoting sin? Right now it’s bcause we have a mutual truce in which we agree businesses should employ people based on their skills and merit rather than to reward their political allies and punish their political opponents. Once you undermine that, gay people are in a pretty precarious position.

> So I would turn your own hypothetical scenario in Part 2 of your post back on you. Suppose Robertson had indeed, been a gay rights supporter – or a gay person! – who said on national news he thought everyone should stand up for gay rights. But his company was going for the fundie demographic and decided to fire him for his statement. Would you be so quick to attack everyone who was disappointed in this action, so eager to stand up for the right of companies to fire anyone they disagree with?
>
> I'm an atheist blogger and I work at a Catholic hospital. Employer tolerance for dissenting opinions is *personal* for me. I'm disappointed in the tone of this post and I hope you reconsider.

I can't tell how many other people have made similar points because none of the three browsers on my computer can successfully load Patheos' nightmarish comment system more than once in a blue moon. But I hope some other Patheos atheists are saying the same. And I have huge respect for the few voices on the lefty blogosphere, like Ampersand, who have spoken out in favor of restraint.

CORRECTION: Mr. Robertson was suspended rather than fired, and has since been reinstated.

The Spirit Of The First Amendment

Popehat comments on some of the same issues I brought up yesterday from the opposite point of view. They bring up an interesting idea they call the "Doctrine Of The Preferred First Speaker":

> The phrase "the spirit of the First Amendment" often signals approaching nonsense. So, regrettably, does the phrase "free speech" when uncoupled from constitutional free speech principles. These terms often smuggle unprincipled and internally inconsistent concepts — like the doctrine of the Preferred First Speaker. The doctrine of the Preferred First Speaker holds that when Person A speaks, listeners B, C, and D should refrain from their full range of constitutionally protected expression to preserve the ability of Person A to speak without fear of non-governmental consequences that Person A doesn't like. The doctrine of the Preferred First Speaker applies different levels of scrutiny and judgment to the first person who speaks and the second person who reacts to them; it asks "why was it necessary for you to say that" or "what was your motive in saying that" or "did you consider how that would impact someone" to the second person and not the first. It's ultimately incoherent as a theory of freedom of expression.

In other words, person A is within their Constitutional rights to rant about how much they hate gays, person B is within their Constitutional rights to go on a rant about how much they hate person A, and if you condemn person B's speech as "an attack on free speech" you're ignoring the basic symmetry of the situation.

It’s a very well-framed idea, and I remember trying to grope towards something like it when I read Michael Anissimov’s Jezebel’s Vigilante Squad on *More Right*. Parts of that post struck me the wrong way, as genuine “Doctrine of Preferred First Speaker” examples, and I have no doubt that Popehat is complaining about a real thing that some people do.

But in the end I have to disagree with Popehat. I think there *is* a legitimate meaning to “spirit of the First Amendment”, I think it rescues parts of Michael’s post on Jezebel, I think it rescues some of the people defending Phil Robertson, and I think it ends up being a really important part of free speech in general.

What is the “spirit of the First Amendment”? Eliezer Yudkowsky writes:

> There are a very few injunctions in the human art of rationality that have no ifs, ands, buts, or escape clauses. This is one of them. Bad argument gets counterargument. Does not get bullet. Never. Never ever never for ever.

Why is this a rationality injunction instead of a legal injunction? Because the point is protecting “the marketplace of ideas” where arguments succeed based on the evidence supporting or opposing them and not based on the relative firepower of their proponents and detractors. And as I mentioned yesterday, we’re not talking some theoretical ivory tower idea here, we’re talking about things like how support for gay marriage has increased by an order of magnitude over the past few decades.

What does “bullet” mean in the quote above? Are other projectiles covered? Arrows? Boulders launched from catapults? What about melee weapons like swords or maces? Where exactly do we draw the line for “inappropriate responses to an argument”?

A good response to an argument is one that addresses an idea; a bad argument is one that silences it. If you try to address an idea, your success depends on how good the idea is; if you try to silence it, your success depends on how powerful you are and how many pitchforks and torches you can provide on short notice.

Shooting bullets is a good way to silence an idea without addressing it. So is firing stones from catapults, or slicing people open with swords, or gathering a pitchfork-wielding mob.

But trying to get someone fired for holding an idea is *also* a way of silencing an idea without addressing it. I’m sick of talking about Phil Robertson, so let’s talk about the Alabama woman who was fired for having a Kerry-Edwards bumper sticker on her car (her boss supported Bush). Could be an easy way to quiet support for a candidate you don’t like. Oh, there are more Bush voters than Kerry voters in this county? Let’s bombard her workplace with letters until they fire her! Now she’s broke and has to sit at home trying to scrape money together to afford food and ruing the day she ever dared to challenge our prejudices! And the next person to disagree with the rest of us will think twice before opening their mouth!

The e-version of this practice is “doxxing”, where you hunt down an online commenter’s personally identifiable information including address. Then you either harass people they know personally, spam their place of employment with angry comments, or post it on the Internet for everyone to see, probably with a message like “I would never threaten this person at their home address *myself*, but if one of my followers wants to, I *guess* I can't stop them." This was the Jezebel strategy that Michael was most complaining about. Freethought Blogs is also particularly famous for this tactic and often devolves into sagas that would make MsScribe herself proud.

A lot of people would argue that doxxing holds people "accountable" for what they say online. But like most methods of silencing speech, its ability to punish people for saying the wrong things is entirely uncorrelated with whether the thing they said is actually wrong. It distributes power based on who controls the largest mob (hint: popular people) and who has the resources, job security, and physical security necessary to outlast a personal attack (hint: rich people). If you try to hold the Koch Brothers "accountable" for muddying the climate change waters, they will laugh in your face. If you try to hold closeted gay people "accountable" for promoting gay rights, it will be very easy and you will successfully ruin their lives. Do you really want to promote a policy that works this way?

There are even more subtle ways of silencing an idea than trying to get its proponents fired or real-life harassed. For example, you can always just harass them online. The stronger forms of this, like death threats and rape threats, are of course illegal. But that still leaves many opportunities for constant verbal abuse, crude sexual jokes, insults aimed at family members, and dozens of emails written in all capital letters about what sorts of colorful punishments you and the people close to you deserve.

Right about the time I started investigating the atheist blogosphere, one popular atheist blogger – I can't remember her name, but I think she was also on Freethought Blogs – shut down her blog after getting an unmanageable number of these. Everyone posting these messages was entirely within their constitutionally protected right to free speech, yet *something went wrong*. A strong voice for atheism was silenced not because her opponents had clever ideas that contradicted her points, but because they managed to harass her off the podium.

Sometimes this can happen by accident – a no-name nobody makes a statement, a very popular blog or the media picks up on it and broadcasts it, and suddenly thousands of people descend on that person telling them how wrong they are. This is nobody’s fault – each individual is completely within their rights to counterargue – but in aggregate it is equivalent to the worst harassment you have ever undergone times ten, and you’re always afraid one of those thousands of people is going to take it upon themselves to contact your employer or your family or something. I don’t have a good solution to this other than to mention that it is a supererogatory but important duty not to join in.

My answer to the “Doctrine Of The Preferred First Speaker” ought to be clear by now. The conflict isn’t always just between first speaker and second speaker, it can also be between someone who’s trying to debate versus someone who’s trying to silence. Telling a bounty hunter on the phone “I’ll pay you $10 million to kill Bob” is a form of speech, but its goal is to silence rather than to counterargue. So is commenting “YOU ARE A SLUT AND I HOPE YOUR FAMILY DIES” on a blog. And so is orchestrating a letter-writing campaign demanding a business fire someone who vocally supports John Kerry.

Bad argument gets counterargument. Does not get bullet. Does not get doxxing. Does not get harassment. Does not get fired from job. Gets counterargument. Should not be hard.

2014

- Two Dark Side Statistics Papers
- Marijuana: Much More Than You Wanted To Know
- A Response To Apophemi on Triggers
- Aretaeus On Bipolar Disorder
- Ten Things I Want To Stop Seeing On The Internet In 2014
- Genetic Testing and Self-Fulfilling Prophecies
- Beware Mass-Produced Medical Recommendations
- Predictions For 2014
- Wirehead Gods On Lotus Thrones
- Based on your findings, which theory about alien thickness seems most valid or most accurate?
- More Search Terms That Have Led People To This Blog
- Seeming and Being Empathetic
- Blogging The Anissimov – Smith Reaction Debate
- Housekeeping Post February 2014
- Bad Dreams
- Nootropics Survey Results And Analysis
- Lies, Damned Lies, And Social Media (Part 5 of ∞)

- **Proposed Biological Explanations For Historical Trends In Crime**
- **God Bless Longecity**
- **Effects Of Vertical Acceleration On Wrongness**
- **In Favor of Niceness, Community, and Civilization**
- **Suggest A Comments Policy**
- **Fix Science In Half An Hour**
- **Unteachable Things Hard To Teach, Study Suggests**
- **Searching For One-Sided Tradeoffs**
- **[CORRECTION] Mountain View Meetup 3/4**
- **The Comment Policy Is “Victorian Sufi Buddha Lite”**
- **Do Life Hacks Ever Reach Fixation?**
- **Book Review: The Perfect Health Diet**
- **The Slate Star Codex Political Spectrum Quiz**
- **Semite Times: The Bible In Palindromes**
- **Five Years and One Week of Less Wrong**
- **Can It Be Wrong To Crystallize Patterns?**
- **What Universal Human Experiences Are You Missing Without Realizing It?**
- **Michigan Meetup and Memory Workshop, 3/23**
- **Typical Mind and Disbelief In Straight People**
- **Should You Reverse Any Advice You Hear?**
- **E-Cig Study: Much Smoke, Little Light**
- **The Character’s Complaint**
- **Wheat: Much More Than You Wanted To Know**
- **The Study of Anglophysics**

- My New OKCupid Profile
- Housekeeping Post April 2014
- Dermatology Pop Quiz
- Dermatology Quiz: Answers
- Going Loopy
- The Cowpox of Doubt
- Do You Believe Me, Doc?
- Someone Writes An Anti-Racist FAQ
- Confounder Of The Day: How Sexy Your Parents Were
- Plutocracy Isn’t About Money
- The Economics Of Art And The Art Of Economics
- Right Is The New Left
- Friendship Is Countersignaling
- Metaphors Be With You
- Stop Confounding Yourself! Stop Confounding Yourself!
- The Control Group Is Out Of Control
- Utopian Science
- Causal Models At Work
- Emotionally Valent Links Post
- Not Even A Real Links Post, Just A Link
- Evening Doc
- Not Even A Real Links Post, Just A Blatant Ad
- Vote On Values, Outsource Beliefs
- Infinite Debt
- Weak Men Are Superweapons
- Getting A Therapist
- You Kant Dismiss Universalizability
- Nerds Can Be Bees, Too

- SSC Gives A Graduation Speech
- Nydwracu’s Fnords
- Apologia Pro Vita Sua
- Compound Interest Is The Least Powerful Force In The Universe
- On Types of Typologies
- Don’t Fear The Filter
- The Wonderful Thing About Triggers
- Southeast Michigan Meetup 6/8
- Asches to Asches
- Don’t Be An Asch-Hole
- Archipelago and Atomic Communitarianism
- Map-Territory Distinctions
- Minutes From The Michigan Rationalist Meetup
- CONSTANT VIGILANCE!
- List Of The Passages I Highlighted In My Copy Of Jonathan Haidt’s “The Righteous Mind”
- Living By The Sword
- Fish – Now By Prescription
- Things That Sometimes Help If You Have Depression
- Money On The Ground
- Ground Morality In Party Politics
- The Art Of Writing Randian Monologues
- Court-ing Disaster
- The Placebo-Singers
- SSC Gets World Cup Fever
- Book Review: The Two-Income Trap
- Invisible Women
- Medicine, As Not Seen On TV

- How Common Are Science Failures?
- Ozy vs. Scott on Charity Baskets
- List Of The Passages I Highlighted In My Copy Of “The Two-Income Trap”
- Social Justice And Words, Words, Words
- SSRIs: Much More Than You Wanted To Know
- Growing Children For Bostrom’s Disneyland
- A Paradox of Ecclesiology
- Psychotropic Base Rates: The Argument From Antibiotics
- HeartMath Considered Incoherent
- Promising The Moon
- No Skyscraper Stagnation
- The Other Codex
- Some Antibiotic Stagnation
- Weird Psychiatric Ads Of The Seventies
- Meditations On Moloch
- Misperceptions On Moloch
- Unspoken Ground Assumptions Of Discussion
- Negative Creativity
- Random Noise Is Our Most Valuable Resource
- Friendship Is (Still) Countersignaling
- Getting Eulered
- Does the Glasgow Coma Scale exist? Do comas?
- Does Race Exist? Does Culture?
- Beware Isolated Demands For Rigor
- Fifty Swifties
- Burdens

- An Iron Curtain Has Descended Upon Psychopharmacology
- Ozy’s Anti Heartiste FAQ
- Nobody Likes A Tattletale
- The Invisible Nation – Reconciling Utilitarianism And Contractualism
- If The Media Reported On Other Dangers Like It Does AI Risk
- Michigan Meetup 9/7
- Radicalizing the Romanceless
- Book Review and Highlights: Quantum Computing Since Democritus
- The Guardian vs. Induction
- Cooperation Un-Veiled
- Mapmaker, Mapmaker, Make Me A Map
- Society Is Fixed, Biology Is Mutable
- What The Hell, Hegel?
- Book Review: Singer on Marx
- Does Class Warfare Have A Free-Rider Problem?
- [Ozy] A Response to Spandrell
- Joint Over- and Underdiagnosis
- SSC Gives A Wedding Speech
- Book Review: Red Plenty
- Streetlight Psychology
- Why No Science Of Nerds?
- Cuddle Culture
- Bottomless Pits Of Suffering
- I Can Tolerate Anything Except The Outgroup
- The Battle Hymn
- Simpler Times

- Prediction Goes To War
- Tumblr on MIRI
- Five Planets In Search Of A Sci-Fi Story
- Five Case Studies On Politicization
- In The Future, Everyone Will Be Famous To Fifteen People
- Book Review: A Future For Socialism
- Alcoholics Anonymous: Much More Than You Wanted To Know
- Michigan Meetup 11/9
- All In All, Another Brick In The Motte
- Ethnic Tension And Meaningless Arguments
- The Right To Waive Your Rights
- Growing Old
- How To Use 23andMe Irresponsibly
- The Dark Side Of Divorce
- Ley Lines Of The Midwest
- Republicans Are Douchebags
- Prisons Are Built With Bricks Of Law And Brothels With Bricks Of Religion, But That Doesn’t Prove A Causal Relationship
- The Categories Were Made For Man, Not Man For The Categories
- Race and Justice: Much More Than You Wanted To Know
- Why I Am Not Rene Descartes
- Such Mixed Feelings About Crazymeds
- Cyber Monday Product Recommendations
- Book Review: On The Road
- Framing For Light Instead Of Heat
- A Story With Zombies
- Beware The Man Of One Study

- **Debunked And Well-Refuted**
- **The Toxoplasma Of Rage**
- **Nobody Is Perfect, Everything Is Commensurable**
- **There Are Rules Here**
- **Book Review: What’s Wrong With The World**

Two Dark Side Statistics Papers

I.

First we have False Positive Psychology: Undisclosed Flexibility In Data Collection And Analysis Allows Presenting Anything As Significant (h/t Jonas Vollmer).

The message is hardly unique: there are lots of tricks unscrupulous or desperate scientists can use to artificially nudge results to the 5% significance level. The clarity of the presentation *is* unique. They start by discussing four particular tricks:

1. Measure multiple dependent variables, then report the ones that are significant. For example, if you're measuring whether treatment for a certain psychiatric disorder improves life outcomes, you can collect five different measures of life outcomes – let's say educational attainment, income, self-reported happiness, whether or not ever arrested, whether or not in romantic relationship – and have a 25%-ish probability one of them will come out at significance by chance. Then you can publish a paper called "Psychiatric Treatment Found To Increase Educational Attainment" without ever mentioning the four negative tests.

2. Artificially choose when to end your experiment. Suppose you want to prove that yelling at a coin makes it more likely to come up tails. You yell at a coin and flip it. It comes up heads. You try again. It comes up tails. You try again. It comes up heads. You try again. It comes up tails. You try again. It comes up tails again. You try again. It comes up tails again. You note that it came up tails four out of six times – a 66% success rate compared to expected 50% – and declare victory. Of course, this result wouldn't be significant, and it seems as if this should be a general rule – that almost by the

definition of significance, you shouldn’t be able to obtain it just be stopping the experiment at the right point. But the authors of the study perform several simulations to prove that this trick is more successful than you’d think:

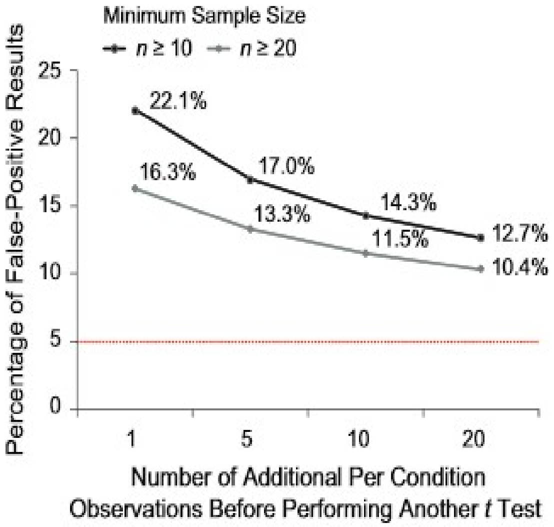


Fig. 1. Likelihood of obtaining a false-positive result when data collection ends upon obtaining significance ($p \leq .05$, highlighted by the dotted line). The figure depicts likelihoods for two minimum sample sizes, as a function of the frequency with which significance tests are performed.

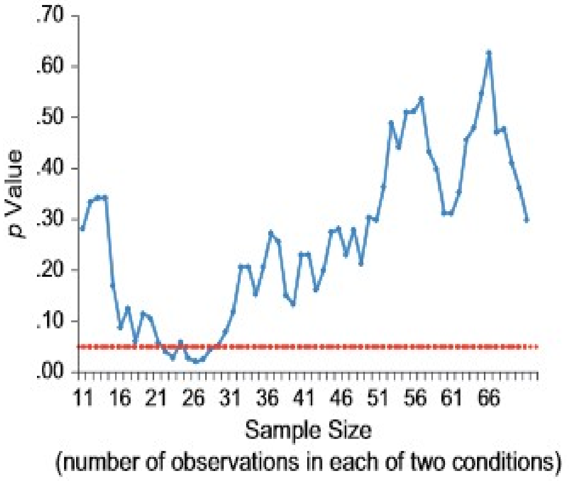


Fig. 2. Illustrative simulation of p values obtained by a researcher who continuously adds an observation to each of two conditions, conducting a t test after each addition. The dotted line highlights the conventional significance criterion of $p \leq .05$.

3. Control for “confounders” (in practice, most often gender). I sometimes call this the “Elderly Hispanic Woman Effect” after drug trials that find that their drug doesn’t have significant effects in the general population, but it *does* significantly help elderly Hispanic women. The trick is you split the population into twenty subgroups (young white men, young white women, elderly white men, elderly white women, young black men, etc), in one of those subgroups it will achieve significance by pure chance, and so you declare that your drug must just somehow be a perfect fit for elderly Hispanic women’s unique body chemistry. This is not *always* wrong (some antihypertensives have notably different efficacy in white versus black populations) but it is *usually* suspicious.

4. Test different conditions and report the ones you like. For example, suppose you are testing whether vegetable consumption affects depression. You conduct the trial with three arms: low veggie diet, medium veggie diet, and high veggie diet. You now have four possible comparisons – low-medium, low-high, medium-high, low-medium-high trend). One of them will be significant 20% of the time, so you can just report that one: “People who eat a moderate amount of vegetables are less likely to get depression than people who eat excess vegetables” sounds like a perfectly reasonable result.

Then they run simulations to show exactly how much more likely you are to get a significant result in random data by employing each trick:

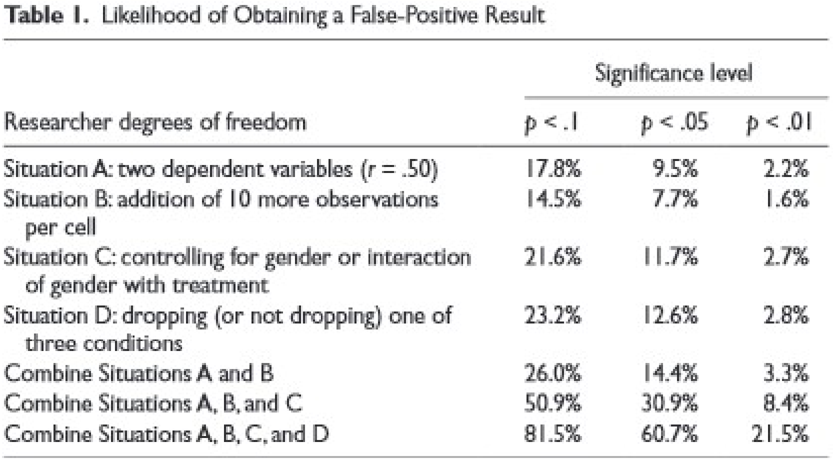


Table 1. Likelihood of Obtaining a False-Positive Result

	Significance level		
Researcher degrees of freedom	$p < .1$	$p < .05$	$p < .01$
Situation A: two dependent variables ($r = .50$)	17.8%	9.5%	2.2%
Situation B: addition of 10 more observations per cell	14.5%	7.7%	1.6%
Situation C: controlling for gender or interaction of gender with treatment	21.6%	11.7%	2.7%
Situation D: dropping (or not dropping) one of three conditions	23.2%	12.6%	2.8%
Combine Situations A and B	26.0%	14.4%	3.3%
Combine Situations A, B, and C	50.9%	30.9%	8.4%
Combine Situations A, B, C, and D	81.5%	60.7%	21.5%

The image demonstrates that by using all four tricks, you can squeeze random data into a result significant at the p less than 0.05 level about 61% of the time.

The authors then put their money where their mouth is by conducting two studies.

The first seems like a very very classic social psychology study. Subjects are randomly assigned to listen to one of two songs – either a nondescript control song or a child’s nursery song. Then they are asked to rate how old they feel. Sure enough, the subjects who listen to the child's song feel older (p = 0.03).

The second study is very similar, with one important exception. Once again, subjects are randomly assigned to listen to one of two songs – either a nondescript control song or a song about aging – "When I'm Sixty-Four" by The Beatles. Then they are asked to put down their actual age, in years. People who listened to the Beatles song became, on average, a year and a half younger than the control group (p = 0.04).

So either the experimental intervention changed their subjects' ages, or the researchers were using statistical tricks. Turns out it was the second one. They explain how they used the four statistical tricks they explained above, and that without those tricks there would have been (obviously) no significant difference. They go on to say that their experiment meets the inclusion criteria for every major journal and that under current reporting rules there's no way anyone could have detected their data manipulation.

They go on to list the changes they think the scientific establishment needs to prevent papers like theirs from reaching print. They're basically "don't do the things we just talked about", but as far as I can tell they rely on the honor system. I think a broader meta-point is that on important studies scientists should have to submit their experimental protocol to a journal and get it accepted or rejected in advance so they can't change tactics mid-stream or drop data. This would also force journals to publish more negative results.

See also their interesting discussion of why they think "use Bayesian statistics" is a non-solution to the problem.

II.

Second we have How To Have A High Success Rate In Treatment: Advice For Evaluators Of Alcoholism Programs.

This study is very close to my heart, because I'm working on my hospital's Substance Abuse Team this month. Every day we go see patients struggling with alcoholism, heroin abuse, et cetera, and we offer them treatment at our hospital's intensive inpatient Chemical Dependency Unit. And every day, our patients say thanks but no thanks, they heard of a program affiliated with their local church that has a 60% success rate, or an 80% success rate, or in one especially rosy-eyed case a frickin' 97% success rate.

(meanwhile, *real* rehab programs still struggle to prove they have a success rate greater than placebo)

My attending assumes these programs are scum but didn't really have a good evidence base for the claim, so I decided to search Google Scholar to find out what was going on. I struck gold in this paper, which is framed as a sarcastic how-to guide for unscrupulous drug treatment program directors who want to inflate their success rates without *technically* lying.

By far the best way to do this is to choose your denominator carefully. For example, it seems fair to only include the people who attended your full treatment program, not the people who dropped out on Day One or never showed up at all – you can hardly be blamed for that, right? So suppose that your treatment program is one month intensive in rehab followed by a series of weekly meetings continuing indefinitely. At the end of one year, you define successful treatment completers as "the people who are still going to these meetings now, at the end of the year". But in general, people who relapse into alcoholism are a whole lot less likely to continue attending their AA meetings than people who stay sober. So all you have to do is go up to people at your AA meeting, ask them if they’re still on the wagon, and your one-year success rate looks really good.

Another way to hack your treatment population is to only accept the most promising candidates to begin with (it works for private schools and it can work for you). We know that middle-class, employed people with houses and families have a much better prognosis than lower-class unemployed homeless single people. Although someone would probably notice if you put up a sign saying “MIDDLE-CLASS EMPLOYED PEOPLE WITH HOUSES AND FAMILIES ONLY”, a very practical option is to just charge a lot of money and let your client population select themselves. This is why for-profit private rehabs will have a higher success rate than public hospitals and government programs that deal with poor people.

Still another strategy is to follow the old proverb: “If at first you don’t succeed, redefine success”. “Abstinence” is such a harsh word. Why not “drinking in moderation”? This is a wonderful phrase, because you can just let the alcoholic involved determine the definition of moderation. A year after the program ends, you can send out little surveys saying “Remember when we told you God really wants you not to drink? You listened to us and are drinking in moderation now, right? Please check one: Y () N ()”. Who’s going to answer ‘no’ to that? Heck, some of the alcoholics I talk to say they’re drinking in moderation *while they are in the emergency room for alcohol poisoning*.

If you can’t handle “moderation”, how about “drinking less than you were before the treatment program”? This takes advantage of regression to the mean – you’re going to enter a rehab program at the worst period of your life, the time when your drinking finally spirals out of control. Just by coincidence, most other parts of your life will include less drinking than when you first came in to rehab, including the date a year after treatment when someone sends you a survey. Clearly rehab was a success!

And why wait a year? My attending and myself actually looked up what was going on with that one 97% success rate program our patient said he was going to. Here's what they do – it's a three month residential program where you live in a building just off the church and you're not allowed to go out except on group treatment activities. Obviously there is no alcohol allowed in the building and you are surrounded by very earnest counselors and fellow recovering addicts at all times. Then, *at the end of the three months, while you are still in the building*, they ask you whether you're drinking or not. You say no. Boom – 97% success rate.

One other tactic I have actually seen in studies and it *breaks my heart* is interval subdivision, which reminds me of some of the dirty tricks from the first study above. At five years' follow-up, you ask people "Did you drink during Year 1? Did you drink during Year 2? Did you drink during Year 3?…" and so on. Now you have five chances to find a significant difference between treatment and control groups. I have literally seen studies that say "Our rehab didn't have an immediate effect, but by Year 4 our patients were doing better than the controls." Meanwhile, in years 1, 2, 3, and 5, for all we know the controls were doing better than the patients.

But if all else fails, there's always the old standby of poor researchers everywhere – just don't include a control group at all. This table really speaks to me:

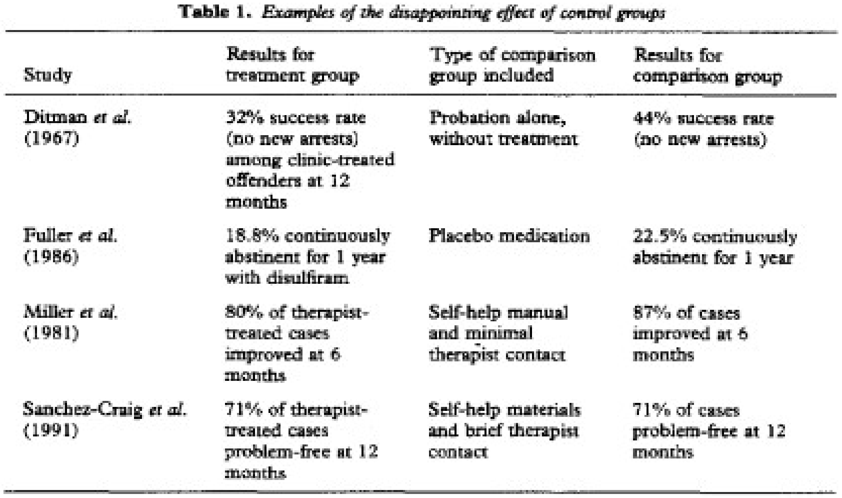


Table 1. *Examples of the disappointing effect of control groups*

Study	Results for treatment group	Type of comparison group included	Results for comparison group
Ditman *et al.* (1967)	32% success rate (no new arrests) among clinic-treated offenders at 12 months	Probation alone, without treatment	44% success rate (no new arrests)
Fuller *et al.* (1986)	18.8% continuously abstinent for 1 year with disulfiram	Placebo medication	22.5% continuously abstinent for 1 year
Miller *et al.* (1981)	80% of therapist-treated cases improved at 6 months	Self-help manual and minimal therapist contact	87% of cases improved at 6 months
Sanchez-Craig *et al.* (1991)	71% of therapist-treated cases problem-free at 12 months	Self-help materials and brief therapist contact	71% of cases problem-free at 12 months

The great thing about this table isn’t just that it shows that seemingly impressive results are exactly the same as placebo. The great thing it shows is that results in the placebo groups in the four studies could be anywhere from a 22.5% success rate to an 87% success rate. These aren’t treatment differences – all four groups are placebo! This is one hundred percent a difference in study populations and in success measures used. In other words, depending on your study protocol, you can prove that there is a 22.5% chance the average untreated alcoholic will achieve remission, or an 87% chance the average untreated alcoholic will achieve remission.

You can bet that rehabs use the study protocol that finds an 87% chance of remission in the untreated. And then they go on to boast of their 90% success rate. Good job, rehab!

Marijuana: Much More Than You Wanted To Know

This month I work on my hospital's Substance Abuse Team, which means we treat people who have been hospitalized for alcohol or drug-related problems and then gingerly suggest that maybe they should use drugs a little less.

The two doctors leading the team are both very experienced and have kind of seen it all, so it's interesting to get a perspective on drug issues from people on the front line. In particular, one of my attendings is an Obama-loving long-haired hippie who nevertheless vehemently opposes medical marijuana or any relaxation on marijuana's status at all. He says that "just because I'm a Democrat doesn't mean I have to support stupid policies I know are wrong" and he's able to back up his opinion with an impressive variety of studies.

To be honest, I had kind of forgotten that the Universe was allowed to contain negative consequences for legalizing drugs. What with all the mental energy it took protesting the the Drug War and getting outraged at police brutality and celebrating Colorado's recently permitting recreational cannabis use and so on, it had completely slipped my mind that the legalization of marijuana might have negative consequences and that I couldn't reject it out of hand until I had done some research.

So I've been doing the research. Not to try to convince my attending of anything – as the old saying goes, do not meddle in the affairs of attendings, because you are crunchy and taste good with ketchup – but just to figure out where exactly things stand.

I. Would Relaxation Of Penalties On Marijuana Increase Marijuana Use?

Starting in the 1970s, several states decriminalized possession of marijuana – that is, possession could not be penalized by jail time. It could still be penalized by fines and other smaller penalties, and manufacture and sale could still be punished by jail time.

Starting in the 1990s, several states legalized medical marijuana. People with medical marijuana cards, which in many cases were laughably easy to get with or without good evidence of disease, were allowed to grow and use marijuana, despite concerns that some of this would end up on the illegal market.

Starting last week, Colorado legalized recreational use of marijuana, as well as cultivation and sale (subject to heavy regulations). Washington will follow later this year, and other states will be placing measures on their ballots to do the same.

One should be able to evaluate to what degree marijuana use rose after these policy changes, and indeed, many people have tried – with greater or lesser levels of statistical sophistication.

The *worst* arguments in favor of this proposition are those like this CADCA paper, which note that states with more liberal marijuana laws have higher rates of marijuana use among teenagers than states that do not. The proper counterspell to such nonsense is *Reverse Causal Arrows* – could it not be that states with more marijuana users are more likely to pass proposals liberalizing marijuana laws? Yes it could. Even more likely, some third variable – let's call it “hippie attitudes” – could be behind both high rates of marijuana use and support for liberal marijuana regimes. The states involved are places like Colorado, California, Washington, and Oregon. I think that speaks for itself. In case it doesn’t, someone went through the statistics and found that these states had the highest rates of marijuana use among teens since *well* before they relaxed drug-related punishments. Argument successfully debunked.

A slightly more sophisticated version – used by the DEA here – takes the teenage marijuana use in a state one year before legalization of medical marijuana and compares it to the teenage marijuana use in a state one (or several years) after such legalization. They often find that it has increased, and blame the increase on the new laws. For example, 28% of Californians used marijuana before it was decriminalized in the 70s, compared to 35% a few years after. This falls victim to a different confounder – marijuana use has undergone some very large swings nationwide, so the rate of increase in medical marijuana states may be the same as the rate anywhere else. Indeed, this is what was going on in California – its marijuana use actually rose slightly *less* than the national average.

What we want is a study that compares the average marijuana use in a set of states before liberalization to the average marijuana use in the country as a whole, and then does the same after liberalization to see if the ratio has increased. There are several studies that purport to try this, of which by far the best is Johnston, O’Malley & Bachman 1981, which monitored the effect of the decriminalization campaigns of the 70s. They survey thousand of high school seniors on marijuana use in seven states that decriminalize marijuana both before and for five years after the decriminalization, and find absolutely no sign of increased marijuana use (in fact, there is a negative trend). Several other studies (eg Thies & Register 1993) confirm this finding.

There is only a hint of some different results. Saffer and Chaloukpa 1999 and Chaloupka, Grossman & Tauras 1999 try to use complicated econometric simulations to estimate the way marijuana demand will respond to different variables. They simulate (as opposed to detecting in real evidence) that marijuana decriminalization should raise past-year use by about 5 – 8%, but have no effect on more frequent use (ie a few more people try it but do not become regular users). More impressively, Model 1993 (a source of some exasperation for me earlier) finds that after decriminalization, marijuana-related emergency room visits went up (trying to interpret their tables, I think they went up by a whopping 90%, but I'm not sure of this). This is sufficiently different from every other study that I don't give it much weight, although we'll return to it later.

Overall I think the evidence is pretty strong that decriminalization probably led to no increase in marijuana use among teens, and may at most have led to a small single-digit increase.

Proponents of stricter marijuana penalties say the experiment isn't fair. In practice, decriminalization does not affect the average user very much – even in states without decriminalization, marijuana possession very rarely leads to jail time. The only hard number I have is from Australia, where in "non-decriminalized" Australian states only 0.3% of marijuana arrests lead to jail time, but a quick back-of-the-envelope calculation suggests US numbers are very similar. And even in supposedly decriminalized states, it's not hard for a cop who wants to get a pot user in jail to find a way (possession of even small amounts can be "possession with intent to sell" if someone doesn't like you). So the overall real difference between decriminalized and not decriminalized is small and it’s not surprising the results are small as well. I mostly agree with them; decriminalization is fine as far as it goes, but it’s a bigger psychological step than an actual one.

The next major milestone in cannabis history was the legalization of medical marijuana. Anderson, Hansen & Rees (2012) did the same kind of study we have seen above, and despite trying multiple different measures of youth marijuana use found pretty much no evidence that medical marijuana legalization caused it to increase. Other studies find pretty much the same.

This could potentially suffer from the same problems as decriminalization studies – the laws don’t always change the facts on the ground. Indeed, for about ten years after medical marijuana legalization, the federal government kept on prosecuting marijuana users even when their use accorded with state laws, and many states had so few dispensaries that in reality not a whole lot of medical marijuana was being given out. I haven’t found any great studies that purport to overcome these problems.

When we examined decriminalization, we found that the studies based on surveys of teens looked pretty good, but that the one study that examined outcomes – marijuana-related ER visits – was a lot less encouraging. We find the same pattern here, and the rain on our parade is Chu 2013, who finds that medical marijuana laws increased marijuana-related arrests by 15-20% and marijuana-related drug rehab admissions by 10-15%.

So what’s going on here? I have two theories. First, maybe medical marijuana use (and decriminalization) increase use among adults only. This could be because the system is working – giving adults access to medical marijuana while keeping it out of the hands of children – or because kids are dumb and don’t understand consequences but adults are more responsive to incentives and punishments. Second, we know that medical marijuana has twice as much THC as street marijuana. Maybe everyone keeps using the same amount of marijuana, but when medical marijuana inevitably gets diverted to the street, addicts can’t handle it and end up behaving much worse than they expected.

Or the studies are wrong. Studies being wrong is always a pretty good bet.

I can’t close this section without mentioning the Colorado expulsion controversy. Nearly everyone who teaches in Colorado says there has been an explosion of marijuana-related problems since medical marijuana was legalized. Meanwhile, the actual surveys of Colorado high school students say that marijuana use, if anything, is going down. A Colorado drug warrior has some strong objections to the survey results, but they center around not really being able to prove that there is a real downward trend (which is an entirely correct complaint) without denying that in fact they show no evidence at all of going *up*.

The consensus on medical marijuana seems to be that it does not increase teen marijuana use either, although there is some murky and suggestive evidence that it might increase illicit or dangerous marijuana use among adults.

There is less information on the effects of full legalization of marijuana, which has never been tried before in the United States. To make even wild guesses we will have to look at a few foreign countries plus some econometric simulations.

No one will be surprised to hear that the first foreign country involved is the Netherlands, which was famously permissive of cannabis up until a crackdown a few years ago. Despite popular belief they never fully legalized the drug and they were still pretty harsh on production and manufacture; distribution, on the other hand, could occur semi-openly in coffee shops. This is another case where we have to be careful to distinguish legal regimes from actual effects, but during the period when there were actually a lot of pot-serving coffee shops, the Netherlands did experience an otherwise-inexplicable 35% rise in marijuana consumption relative to the rest of Europe. This is true even among teenagers, and covers both heavy use as well as occasional experimentation. Some scientists studying the Netherlands' example expect Colorado to see a similar rise; others think it will be even larger because the legalization is complete rather than partial.

The second foreign country involved is Portugal, which was maybe more of a decriminalization than a legalization case but which is forever linked with the idea of lax drug regimes in the minds of most Americans. They decriminalized all drugs (including heroin and cocaine) in 2001, choosing to replace punishment with increased treatment opportunities, and as we all have been told, no one in Portugal ever used drugs ever again, or even remembers that drugs exist. Except it turns out it's more complicated; for example, the percent of Portuguese who admit to lifetime use of drugs has doubled since the law took effect. Two very patient scientists have sifted through all the conflicting claims and found that in reality, the number of people who briefly experiment with drugs has gone way up, but the number of addicts hasn't, nor has the number of bad outcomes like overdose-related deaths. There are many more people receiving drug treatment, but that might just be because Portugal upped its drug treatment game in a separate law at the same time they decriminalized drugs. Overall they seem to have been a modest success – neither really raising nor decreasing the number of addicts – but they seem more related to decriminalization (which we've already determined doesn't have much effect) than to legalization per se.

Returning to America, what if you just *ask* people whether they would use more marijuana if it's legal? Coloradans were asked if they plan to smoke marijuana once it becomes legal; comparing survey results to current usage numbers suggests 40% more users above the age of 18; it is unclear what the effect will be on younger teens and children.

Finally, we let the economists have their say. They crunch all the data and predict an increase of 50 – 100% based solely on the likely price drop (even with taxes factored in). And if there's one group we can trust to make infallible predictions about the future, it's economists.

Overall I find the Dutch evidence most convincing, and predict a 25 – 50% increase in adult marijuana use with legalization. I would expect a lower increase – 15 – 30% – among youth, but the data are also perfectly consistent with no increase at all.

Conclusion for this section: that decriminalization and legalization of medical marijuana do not increase youth marijuana use rates, although there is some shaky and indirect evidence they do increase adult use and bad behavior. There is no good data yet on full legalization, but there's good reason to think it would substantially increase adult use and it might also increase youth use somewhat.

II. Is Marijuana Bad For You?

About 9% of marijuana users eventually become addicted to the drug, exposing them to various potential side effects.

Marijuana smoke contains a lot of the same chemicals in tobacco smoke and so it would not be at all surprising if it had some of the same ill effects, like cardiovascular disease and lung cancer. But when people look for these effects, they can't find any increase in mortality among marijuana smokers. I predict that larger studies will one day pick something up, but for now let's take this at face value.

Much more concerning are the attempts to link marijuana to cognitive and psychiatric side effects. Meier et al (2012) analyzed a study of a thousand people in New Zealand and found that heavy marijuana use was linked to an IQ decline of 8 points. Rogeberg 2012 developed an alternative explanation – poor people saw their IQs drop in their 20s more than rich people because their IQs had been artificially inflated by schooling; what Meier et al had thought to be an effect of cannabis was really an effect of poor people having an apparent IQ drop and using cannabis more often. Meier et al pointed out that actually, poor people didn't use cannabis any more often than anyone else and effects remained when controlled for class. Other studies, like Fried et al (2002) find the same effect, and there is a plausible biological mechanism (cannabinoids something something neurotransmitters something brain maturation). As far as I can tell the finding still seems legit, and marijuana use does decrease IQ. It is still unclear whether this only applies in teenagers (who are undergoing a "sensitive period of brain development") or full stop.

More serious still is the link with psychosis. A number of studies have found that marijuana use is heavily correlated with development of schizophrenia and related psychotic disorders later in life. Some of them find relative risks as high as 2 – heavy marijuana use doubles your chance of getting schizophrenia, which is already a moderately high 1%. But of course correlation is not causation, and many people have come up with alternative theories. For example, maybe people who are already kind of psychotic use marijuana to self-medicate, or just make poor life choices like starting drugs. Maybe people of low socioeconomic status who come from broken homes are more likely to both use marijuana and get schizophrenia. Maybe some gene both makes marijuana really pleasant and increases schizophrenia risk.

I know of three good studies attempting to tease out causation. Arseneault et al (2004) checks to see which came first – the marijuana use or the psychotic symptoms – and finds it was the marijuana use, thus supporting an increase in risk from the drug. Griffith-Lendering et al (2012) try the same, and find *bidirectional* causation – previous marijuana use seems to predict future psychosis, but previous psychosis seems to predict future marijuana use. A very new study from last month boxes clever and checks whether your marijuana use can predict schizophrenia in *your relatives*, and find that it does – presumably suggesting that genetic tendencies towards schizophrenia cause marijuana use and not vice versa (although Ozy points out to meet that the relatives of marijuana users are more likely to use marijuana themselves; the plot thickens). When a meta-analysis tries to control for all of these factors, they get a relative risk of 1.4 (they call it an odds ratio, but from their discussion section I think they mean relative risk).

Is this true, or just the confounders they failed to pick up? One argument for the latter is that marijuana use has increased very much over the past 50 years. If marijuana use caused schizophrenia, we would expect to see much more schizophrenia, but in fact as far as anyone can tell (which is not very far) schizophrenia incidence is decreasing. The decrease might be due (maybe! if it even exists at all!) to obstetric advances which prevent fetal brain damage which could later lead to the disease. The effect of this variable is insufficiently known to pretend we can tease out some supposed contrary effect of increased marijuana use. Also, some people say that schizophrenia is increasing in young people, so who knows?

The *exact* nature of the marijuana-psychosis link is still very controversial. Some people say that marijuana causes psychosis. Other people say it "activates latent psychosis", a term without a very good meaning but which might mean that it pushes people on the borderline of psychosis – eg those with a strong family history but who might otherwise have escaped – over the edge. Still others say all it does is get people who would have developed psychosis eventually to develop it a few years earlier. You can read a comparison of all the different hypotheses here.

I've saved the most annoying for last: is marijuana a "gateway drug"? Would legalizing it make it more or less of a "gateway drug"? This claim seems tailor-made to torture statisticians. We know that marijuana users are *definitely* more likely to use other drugs later – for example, marijuana users are 85x more likely than non-marijuana users to use cocaine. but that could be either because marijuana affects them in some way (implying that legalizing marijuana would increase other drug use), because they have factors like genetics or stressful life situation that makes them more likely to use all drugs (implying that legalizing marijuana would not affect other drug use), or because using illegal marijuana without ill effect connects them to the illegal drug market and convinces them illegal drugs are okay (implying that legalizing marijuana would decrease other drug use). RAND comes very close to investigating this properly by saying that when the Dutch pseudo-legalized marijuana, use of harder drugs stayed stable or went down, but all their study actually shows is that the ratio of marijuana users : hard drug users went down. This is to be expected when you make marijuana much easier to get, but it’s still consistent with the absolute number of hard drug users going way up. The best that can be said is that there is no direct causal evidence for the gateway theory and some good alternative explanations for the effect. Let us accept their word for it and never speak of this matter again.

Conclusion for this section: Marijuana does not have a detectable effect on mortality and there is surprisingly scarce evidence of tobacco-like side effects. It probably does decrease IQ if used early and often, possibly by as many as 8 IQ points. It may increase risk of psychosis by as much as 40%, but it’s not clear who is at risk or whether the risk is even real. The gateway drug hypothesis is too complicated to evaluate effectively but there is no clear casual evidence in its support.

III. What Are The Costs Of The Drug War?

There are not really that many people in jail for using marijuana.

I learned this from Who’s Really In Prison For Marijuana?, a publication of the National Office Of Drug Control Policy, which was clearly written by someone with the same ability to take personal offense at bad statistics that inspires my posts about Facebook. The whole thing seethes with indignation and makes me want to hug the drug czar and tell him everything will be okay.

Only 1.6% of state prisoners are serving time for marijuana, only 0.7% are serving for marijuana possession, and only

0.3% are first time offenders. Some of those are "possession" in the sense of "possessing a warehouse full of marijuana bales", and others are people who committed much more dangerous crimes but were nailed for marijuana, in the same sense that Al Capone was nailed for tax evasion. The percent of normal law-abiding people who just had a gram or two of marijuana and were thrown in jail is a rounding error, and the stories of such you read in the news are extremely dishonest (read the document for examples).

Federal numbers are even lower; in the entire federal prison system, they could only find 63 people imprisoned with marijuana possession as the sole crime, and those people were possessing a median of one hundred fifteen *pounds* of marijuana (enough to make over 100,000 joints).

In total, federal + state prison and counting all the kingpins, dealers, manufacturers, et cetera, there are probably about 16,000 people in prison solely for marijuana-related offenses, serving average actual sentence lengths of three year. But it's anybody's guess whether those people would be free today if marijuana were legal, or whether their drug cartels would just switch to something else.

Looking at the other side's statistics, I don't see much difference. NORML claims that there are 40,000 people in prison for marijuana use, but they admit that half of those people were arrested for using harder drugs and marijuana was a tack-on charge, so they seem to agree with the Feds about around 20,000 pure marijuana prisoners. SAM agrees that only 0.5% of the prison population is in there for marijuana possession alone. I see no reason to doubt any of these numbers.

A much more serious problem is marijuana-related arrests, of which there are 700,000 a year. 90% of them are for simple possession, and the vast majority do not end in prison terms; they do however result in criminal records, community service, a couple days of jail time until a judge is available to hear the case, heavy fines, high cost of legal representation, and moderate costs to the state for funding the whole thing. Fines can be up to $1500, and legal representation can cost up to $5000 (though I am suspicious of this paper and think it may be exaggerating for effect). These costs are often borne by poor people who will have to give up all their savings for years to pay them back.

Costs paid by the government, which cover everything from police officers to trials to prison time, are estimated at about $2 billion by multiple sources. This is only 3% of the total law enforcement budget, so legalizing marijuana wouldn't create some kind of sudden revolution in policing, but as the saying goes, a billion here, a billion there, and eventually it adds up to real money. And a Harvard economist claims that the total monetary benefits from legalization, including potential tax revenues, could reach $14 billion.

Some people worry that legalizing marijuana would cause an increase in car accidents by "stoned drivers", who, like drunk drivers, have impaired reflexes and poor judgment, and indeed there is a small but real problem of marijuana-induced car accidents. But Chaloukpa and Laixuthai (1994) crunch the numbers and find that decreased price/increased availability of marijuana is actually associated with *decreased* car accidents, probably because marijuana is substituting for alcohol in the "have impairing substances and then go driving" population. This finding – that marijuana and alcohol substitute for each other – has been spotted again and again. Anderson & Rees (2013) find that states that legalize medical marijuana see a 5% drop in beer sales. There are however a few dissenting opinions: Cameron & Williams (2001), in complex econometric simulations that may or may not resemble the real world in any respect, find that increasing the price of alcohol increases marijuana use, but increasing the price of marijuana does not affect alcohol use, and the same researcher finds that banning alcohol on a college campus also decreases marijuana use. Also, possibly marijuana use increases smoking? This whole area is confusing, but I am most sympathetic to to the Andersen and Rees statistics which say that medical marijuana states are associated with 13% fewer traffic fatalities.

Overall conclusion for this section: full legalization of marijuana would free about 20,000 people from jail (although most of them would not be exactly fine upstanding citizens), prevent 700,000 arrests not resulting in jail time per year, save between 2 and 14 billion dollars, and possibly reduce traffic fatalities a few percent (or, for all we know, increase them).

IV. An Irresponsible Utilitarian Analysis

Decriminalization and legalization of medical marijuana seem, if we are to trust the statistics in (I) saying they do not increase use among youth, like almost unalloyed good things. Although there are some nagging hints of doubt, they are not especially quantifiable and therefore not amenable to analysis. Without a very strong predisposition to try as hard as possible to fit the evidence into a pessimistic picture, I don’t think there’s a great argument against either of these two propositions. Let’s concentrate on legalization, which would mean something like “People can grow and sell as much marijuana as they want and it’s totally legal for people over 21, with the same level of penalties as today for people under 21”.

Section (I) concludes that legalization could lead to an increase in adult marijuana use up to 50%. There's not a lot of evidence on what it could do to teen marijuana use, but since it seems teen marijuana use is less responsive to legal changes, I made up a number and said 20%. Lest you think I am being unfair, note that this is well below the percent increase predicted by the survey that asked 18 year olds if they would start using marijuana if it were legal.

Right now about 1.5 million teenagers use marijuana "heavily". Most of the detrimental effects of marijuana seem concentrated in teens and people in their early twenties; I'm going to artificially round that up to 2 million to catch the early 20 year olds. If this 2 million number increased 20%, 400,000 extra teens would start heavily using marijuana.

Those 400,000 teens would lose 8 IQ points each. IQ increases your yearly earnings by about $500 per point, so these people would lose about $4,000 a year. Making very strong assumptions about salary being a measure of value to society, society would lose about $1.6 billion a year directly, plus various intangibles from potential artists and scientists losing the ability to create masterpieces and inventions, plus various *really* intangibles like a slightly dumber electorate.

We need to use a different number to calculate psychosis risk, since the studies were done on "people who had used marijuana at least once". The appropriate number turns out to be 8 million teenagers; of those, 1%, or 80,000, would naturally develop schizophrenia. If the 1.4 relative risk number is correct, marijuana use will increase that to 112,000, for a total increase of 32,000 people. Schizophrenia pretty much always presents in the 15 – 25 age window, so we'll say we get 3,200 extra cases per year.

There were 35000 road traffic accident fatalities in the US last year. If greater availability of marijuana decreases those fatalities by 13% (note that I am using the number from medical marijuana legalization and not for marijuana legalization per se, solely because it is a number I actually have), that will cause 4500 fewer road traffic deaths per year. There may be additional positive effects of alcohol substitution from, for example, less liver disease. But there may also be additional negative effects from increasing use of tobacco, so let's just pretend those cancel out.

So here is my guess at the yearly results of marijuana legalization:

– 20,000 fewer prisoners (but they might switch to other criminal enterprises)
– 700,000 fewer arrests
– $2 billion less in law enforcement costs
– Some amount of positive gain (let's say $5 billion) in taxes
– 4500 fewer road traffic deaths (if you believe the preliminary alcohol substitution numbers)

– 400,000 people with lower IQ
– $2 billion in social costs from above dumber people
– 3,200 more cases of schizophrenia a year

We'll proceed to calculate the nonmonetary burden of each of these in QALYs, then add the monetary burden in dollars, then convert.

The searchable public database of utility weights for all diseases (God I love the 21st century) tells me that schizophrenia has a QALY weight of 0.73. It generally starts around 20 and lasts a lifetime, so each case of schizophrenia costs us 0.27 * 50 or 13.5 QALYs. Therefore, the total burden of the 3,200 added schizophrenia cases is 43 kiloQALYs.

There’s no good way to calculate the QALY weight of having 4-8 fewer IQ points, and unfortunately this is going to end up being among the most important numbers in our results. If we say the lifetime cost of this problem is 3 QALYs, and divide the number by eight to represent eight years worth of teenagers in our sample population, we end up with 400,000/8 * 3 = 150 kiloQALYs.

My own survey tells me that being in prison has a QALY weight around 0.5. Marijuana sentences generally last an average of three years, which suggests that 1/3 of these marijuana prisoners are arrested every year, so the total burden of the ~6000ish marijuana imprisonments each year is 3 * ~6000 * 0.5 = 10 kiloQALYs.

Assume the average road traffic death occurs at age 30, costing 40 years of potential future life. The total cost of 4500 road traffic deaths is 40 * 4500 = 180 kiloQALYs.

The arrests are going to require even more fudging than normal. Average jail time for a marijuana arrest (when awaiting trial) is “one to five days” – let’s round that off to two and then use our prison number to say that the jail from each arrest is 2/365 * 0.5 = three-thousandths of a QALY. I am going to arbitrarily round this up to one one-hundredth of a QALY to account for emotional trauma and the burden of fines, then even more arbitrarily round this up to a tenth of a QALY to account for possibility of getting a criminal record. This sets the burden of 700,000 arrests at 70 kiloQALYs.

Now our accounting is:

Costs from legalization compared to current system: 200 kQALYs and $2 billion
Benefits from legalization compared to current system: 260 kQALYs and $7 billion

Although it's not going to be necessary, we can interconvert QALYs and dollars at the going health-care rate of about $100,000/QALY ($100 million/kQALY):

Costs from legalization compared to current system: 220 kQALYs
Benefits from legalization compared to current system: 330 kQALYs

And get:

Net benefits from legalization: +110 kQALYs

Except that this is extremely speculative and irresponsible. By far the largest component of the benefits of legalization turned out to be the effect on road traffic accidents, which is based on only two studies and which may on further research turn out to be a cost. And by far the largest component of the costs of legalization turned out to be the effect on IQ, and we had to totally-wild-guess the QALY cost of an IQ point loss. The wiggle room in my ignorance and assumptions is more than large enough to cover the small gap between the two policies in the results.

So my actual conclusion is:

There is not a sufficiently obvious order-of-magnitude difference between the costs and benefits of marijuana legalization for a evidence-based utilitarian analysis of costs and benefits to inform the debate. You may return to your regularly scheduled wild speculation and shrill accusations.

But I wouldn't say this exercise is useless. For example, it suggests that whether marijuana legalization is positive or negative on net depends almost entirely on small changes in the road traffic accident rate. This is something I've never heard anyone else mention, but which in retrospect should be obvious; the few debatable health effects and the couple of people given short jail sentences absolutely can't compare to the potential for thousands more (or fewer) traffic accidents which leave people permanently dead.

So my actual actual conclusion is:

We should probably stop caring about health effects of marijuana and about imprisonment for marijuana-related offenses, and concentrate all of our research and political energy on how marijuana affects driving.

This cements my previous intuitions on irresponsible use of statistics – it's unlikely to unilaterally solve the problem, but it can be very good at pointing out where you're being irrational and suggesting new ways of looking at a question.

EDIT: People in the comments have pointed out several important factors left out, including:
– Some people enjoy smoking marijuana
– The opening of a permanent criminal record may mean arrests are worse than I estimate. I can't find good statistics on how often this happens, but do note that decriminalization prevents a record from being opened.
– Loss of 8 IQ points may have wider social effects than I estimate, since IQ affects for example crime rate.
– Legalizing marijuana might remove a source of funding for organized crime

A Response To Apophemi on Triggers

[content warning: discussion of triggers. Mentions various triggers. Mentions, without using or condoning, racial slurs]

I.

I originally planned not to respond to Apophemi's essay requesting that people not discuss potentially triggering ideas dispassionately, because my response would inevitably have to discuss a lot of triggering ideas, and it would be dispassionate, and that *might* not be the most effective way of conveying that I take zir concerns seriously.

(Apophemi's essay complains about being misgendered but doesn't give me ironclad evidence what zir gender is, so I'm going to use the gender neutral pronoun here as a least bad option. No offense is intended and if Apophemi tells me what pronoun ze prefers I'll edit it in.)

I'm changing my mind for two reasons. First, everyone else is doing it, so Apophemi has probably reached Peak Triggering by now and the situation can't get any worse. Second, I feel like it would be more respectful and productive to object and give zir a chance to explain why my objections are wrong, than to just say "I disagree with this but I'm not going to explain why" and dismiss the whole thing outright.

(That having been said, if Apophemi doesn't want to read this, I am totally in favor of this; ignoring all posts on my blog tagged "race/gender/etc" is always a good life choice.)

My one worry is the comment thread. I no longer trust my commenters to be kind or reasonable, and since we're talking specifically about triggers and giving a big list of triggery things, unkind people present a problem. So I am closing comments for this thread. If Apophemi wants to make a response, ze may email it to me or post it somewhere and I will add it in.

II.

Apophemi writes:

> On the other hand, most of these things involve warning signs for opinions whose holders are frequently detrimental to my health and safety, and therefore I feel pretty entitled to these boundaries, and pretty insulted at the implication that possessing such boundaries is inferior to not possessing them.
>
> An example: I cannot in good faith entertain the argument that high-scarcity societies are right in having restrictive, assigned-sex-based gender roles, even if these social structures result in measurable maximized utility (i.e. many much kids). I have a moral imperative against this that overrides my general impulse towards maximized utility, or rather (if you asked me about it personally) tilt-shifts my view of what sectors ‘deserve’ to see their utility maximized at the expense of a given other sector.
>
> However, this results in a knee-jerk intellectual squick when I run across someone entertaining or endorsing these arguments. (If I were being YouTube-commenter-style punchy about this, this entire post would have been a comment reading "‘Fertile women’ my ass.", for the record.) This is because respect for said arguments and/or the idea behind them is a warning sign for either 1) passively not respecting my personhood or 2)

> actively disregarding my personhood, both of which are, to use some vernacular, hella fucking dangerous to me personally.
>
> I am reasonably confident (insert p value here) that this attitude is self-replicating among people who are accustomed to being at risk in a specific way that generally occurs to marginalized populations. (I cannot speak for people who may have a similar rhetorical roadblock without it being yoked to a line of social marginalization, other than that I suspect they happen.) This would mean that rewarding the “ability” to entertain any argument “no matter how ‘politically incorrect’” (to break out of some jargon, “no matter how likely to hurt people”) results in a system that prizes people who have not been socially marginalized or who have been socially marginalized less than a given other person in the discussion, since they will have (in general) less inbuilt safeguards limiting the topics they can discuss comfortably.
>
> In other words, prizing discourse without limitations (I tried to find a convenient analogy for said limitations and failed. Fenders? Safety belts?) will result in an environment in which people are more comfortable speaking the more social privilege they hold. (If you prefer to not have any truck with the word ‘privilege’, substitute ‘the less likelihood of having to anticipate culturally-permissible threats to their personhood they have lived with’, since that’s the specific manifestation of privilege I mean. Sadly, that is a long and unwieldy phrase.)

This reminds me of the idea of safe spaces.

Safe spaces are places where members of disadvantaged groups can go, usually protected against people in other groups who tend to trigger them, and discuss things relevant to that group free from ridicule or attack. I know there are many for women, some for gays, and I recently heard of a college opening one up for atheists. They seem like good ideas.

I interpret Apophemi's proposal to say that the rationalist community should endeavor to be a safe space for women, minorities, and other disadvantaged groups.

One important feature of safe spaces is that they can't always be safe for two groups at the same time. Jews are a discriminated-against minority who need a safe space. Muslims are a discriminated-against minority who need a safe space. But the safe space for Jews should be very far way from the safe space for Muslims, or else neither space is safe for anybody.

The rationalist community is a safe space for people who obsessively focus on reason and argument even when it is socially unacceptable to do so.

I don't think it's unfair to say that these people need a safe space. I can't even count the number of times I've been called "a nerd" or "a dork" or "autistic" for saying something rational is too high to count. Just recently commenters on Marginal Revolution – not exactly known for being a haunt for intellect-hating jocks – found an old post of mine and called me among many other things "aspie", "a pansy", "retarded", and an "omega" (a PUA term for a man who's so socially inept he will never date anyone).

The reason the rationalist community tends to talk about controversial issues like race and gender on occasion is that the whole point of rationalism is giving things a fair analysis regardless of whether it's socially popular or acceptable to talk about. So of *course* it will start focusing on all of the ideas that are least acceptable to talk about. I remember talking to someone who admitted, after several false starts and awkward pauses, that he found the scientific research on differences between races pretty convincing. I answered that I was still neutral on the matter but that Jensen was indeed a pretty darned meticulous researcher, and he very nearly cried with relief. He'd thought he was a terrible person for taking the research seriously, had never been able to talk about it with anyone, was stuck in a guilt spiral over it, and I was the first person to give him basic human sympathy.

And I think most people in the rationalist community have shared this reaction – not necessarily about race and gender issues, because contrary to the above we really don't talk about those that much – but about atheism, or transhumanism, or negative utilitarianism, or simulationism, and they had finally found people who would pay them the respect to debate their ideas on merit instead of mouthing the appropriate social platitude to dismiss it as horrible or as totally obvious.

If you are the sort of person with the relevant mental quirk, living in a society of people who don't do this is a terrifying an alienating experience. Finding people who are like you is an amazing, liberating experience. It is, in every sense of the word, a safe space.

If you want a community that is respectful to the triggers of people who don't want to talk about controversial ideas, the Internet is full of them. Although I know it's not true, sometimes it seems to me that half the Internet is made up of social justice people talking about how little they will tolerate people who are not entirely on board with social justice ideas and norms. Certainly this has been my impression of Tumblr, and of many (very good) blogs I read (Alas, A Blog comes to mind, proving that my brain sorts in alphabetical order). There is no shortage of very high-IQ communities that will fulfill your needs.

But you say you're interested in and attracted to the rationalist community, that it would provide something these other communities don't. Maybe you are one of those people with that weird mental quirk of caring more about truth and evidence than about things it is socially acceptable to care about, and you feel like the rationalist community would be a good fit for that part of you. If so, we would love to have you!

But if you want to join communities specifically because they are based around dispassionate debate and ignoring social consequences, but your condition for joining is that they stop having dispassionate debate and take social consequences into account, well, then you're one of those people – like Groucho Marx – who refuses to belong to any club that would accept you as a member.

Imagine a Jew walking into a safe space for Muslims, and saying he finds Islam really interesting and wants to participate, but that in order for it to be a safe space for him they really need to stop talking about that whole "Allah" thing.

III.

I deliberately said "the rationalist community" above rather than "Less Wrong", because Less Wrong explicitly *does* try to be a safe space. It has a (vague and very poorly enforced) ban on talking about politics or other controversial topics which successfully discourages Reactionaries and their ilk from starting threads directly about their controversial views (they often get away with discussing other results that refer to them only indirectly).

These topics nevertheless come up anyway at regular intervals. There is almost always the same pattern when this happens:

A feminist or other person in the social justice movement very prominently posts a declaration that everyone on the site needs to be more feminist and social-justice-y. They get heavily upvoted.

A few people in the comments politely disagree, sometimes with the gist of the post, other times with specific claims.

Other people express outrage that anyone would disagree, and say this just proves that the site is full of horrible people and that feminism and social justice are needed now more than ever.

World War III happens.

It happened when Daenerys gathered a whole series of feminist things from people that got posted to Discussion called things like “On Creepiness” and “On Misogyny”. It happened when Multiheaded, a Marxist somewhere to the left of Kropotkin, posted a thread complaining about people complaining that there were people complaining about controversial opinions on the site (or something). It happened when Apophemi’s essay itself got posted to the site and heavily upvoted.

I’ve downvoted all of these things, not because I disagree with them (although I often do) but because the ban on politics is really useful to avoid exactly this kind of situation.

I hope in the future it is more consistently enforced, and I hope this would be more conducive to the kind of site Apophemi wants.

“But,” people object “banning politics is hard, and talking politics sometimes is fun, and besides, social justice ideas are important to disseminate. Can’t we just ban the nasty, triggering kinds of politics?”

This would be a good time to admit that I am massively, *massively* triggered by social justice.

I know exactly why this started. There was an incident in college when I was editing my college newspaper, I tried to include a piece of anti-racist humor, and it got misinterpreted as a piece of pro-racist humor. The college’s various social-justice-related-clubs decided to make an example out of me. I handled it poorly (“BUT GUYS! THE EVIDENCE DOESN’T SUPPORT WHAT YOU’RE DOING!”) and as a result spent a couple of weeks having everyone in the college hold rallies against me followed by equally horrifying counter-rallies for me. I received a couple of death threats, a few people tried to have me expelled, and then everyone got bored and found some other target who was even more fun to harass. Meantime, I was seriously considering suicide.

But it wasn’t just that one incident. Ever since, I have been sensitive to how much a lot of social justice argumentation resembles exactly the bullying I want a safe space from – the “aspie”, the “nerd”, that kind of thing. Just when I thought I had reached an age where it was no longer cool to call people “nerds”, someone had the bright idea of calling them “nerdy white guys” instead, and so transforming themselves from schoolyard bully to brave social justice crusader. This was the criticism I remember most from my massive Consequentialism FAQ – he’s a nerdy white dude –

and it’s one I have come to expect any time I do anything more intellectual than watch American Idol, and usually from a social justicer.

(one reason I like the MsScribe story so much is that it really brings into relief how aligned social justice and bullying can be. I’m not saying that all or even most social justice is about bullying. Just *enough*)

The worst part was when I read some social justice essay – I can’t remember where – which claimed that it was impossible to bully a member of a privileged group. That it didn’t count. That there was no such thing. So not only did they *sound* suspiciously like bullies, but they were conveniently changing the rules so that it was impossible by definition for me to be bullied at all, and all my friends (except for the black ones) who had problems with bullies as a child or in the present – didn’t count, didn’t exist, didn’t deserve any sympathy.

I believe you mentioned in your essay that feeling like you’re being told you’re not a person is really scary? Well, just so.

So suffice it to say I am triggered by social justice. Any mildly confrontational piece of feminist or social justice rhetoric sends me into a panic spiral. When I read the essay this post was based on, I got only about four hours of sleep that night because my mind was racing, trying to figure out whether I was going to get in trouble about it and whether anyone who supported it could hurt me and how I could defend myself against it.

Because my mind doesn’t just let me feel sad for a minute and then move on – no, that would be too easy. It gives me this massive compulsion to “defend myself” against any piece of social justice I see by writing really long and complete rebuttals. Which inevitably attract more social justice people wanting to debate me. And unfortunately, outrage addiction is a very real thing, and I find myself *actively seeking out* the most horrible social justice memes in order to be horrified by them.

(…also, telling me I'm not allowed to be triggered by my triggers is itself a trigger. Whoever designed the human mind was *really* kind of a jerk.)

I struggle against this all the time. H.L. Mencken writes "Every normal man must be tempted at times to spit on his hands, hoist the black flag, and begin to slit throats." Well, this is my temptation. It requires more willpower than anything else I do in my life – more willpower than it takes for me to get up in the morning and work a ten hour day – to resist the urge to just hoist the black flag and turn into a much less tolerant and compassionate version of Heartiste.

I don't think I'm at all alone in this. Like, you may notice there's a large contingent of people – mostly men, but a surprising number of women as well – who totally freak out when they hear social justice stuff and seem to loathe social justice with an unholy passion? And maybe you've wondered whether the classic glib dismissal of them as people benefitting from the patriarchy who are upset about "uppity women" *quite* explains the level of rage and terror and sudden lashing out?

If you are, indeed, someone who has been traumatized and is easily triggered, you can probably recognize the signs yourself. There's a certain desperation, a certain terror thinly disguised by rage that doesn't really come from anything else.

So suffice it to say I am triggered by social justice, and probably a lot of other people are too. Why do I make such a big deal of this?

First, because it has a lot of bearing on whether we can just ban triggery things. There is a certain school of thought that there are two or three excessively evil things that trigger other people, like making fun of rape, and once we make people stop those, we will live in a trigger-free paradise.

But that’s not true. I’m triggered by feminism. My girlfriend is also triggered by certain kinds of feminism (long story), but *also* by many discussions of charity – whenever ze hears about it, ze starts worrying ze is a bad person for not donating more money to charity, has a mental breakdown, and usually ends up shaking and crying for a little while.

Since we can never make every form of discussion respect everybody’s triggers, that leaves two solutions. First, we can try the “my triggers are important, your triggers are invalid” solutions and end up with powerful groups able to enforce their triggers, and weak groups being told to “just man up”. Second, we can try the safe space solution, where not everyone can be certain of safety everywhere, but everyone is certain of safety somewhere. I don’t expect Tumblr to stop being feminist for me, but I *have* managed to scrub my Facebook feed so thoroughly that I only get about two or three articles per week on how hilarious it would be if male superheroes were dressed like female superheroes. One learns to relish little victories.

Second, because I think the essay contains a false dichotomy: *privileged* people don’t have any triggers, *oppressed* people do. *You guys* are intact, *I* am broken. But truth is, everybody’s broken. The last crown prince of Nepal was raised with limitless wealth and absolute power, and he still freaked out and murdered his entire family and then killed himself. There's probably someone somewhere who still believes in perfectly intact people, but I bet they're not a psychiatrist.

Third, because I have not yet raised the black flag. And some of my resistance I credit to – the rationalist community. The Litany of Tarski: "If all feminism is horrible, I desire to believe that all feminism is horrible. If all feminism is *not* horrible, I desire to believe that all feminism is not horrible." It is a calming litany. Sometimes it helps.

A Christian proverb says: "The Church is not a country club for saints, but a hospital for sinners". Likewise, the rationalist community is not an ivory tower for people with no biases or strong emotional reactions, it's a dojo for people learning to resist them.

I do not think it is always wrong for people to engage in activities that exclude certain categories of disadvantaged people. For example, music naturally excludes the deaf (someone will bring up Beethoven here. You know what I mean). Horseback riding excludes most people too poor to buy horses or live in the country. This is sad, but these activities should still continue.

But I do not think dispassionate discussion for the easily triggered is as bad as all that. It is more like marathons for people who are out of shape. They will have difficulty at first. If they want to learn, they can. If they try, they will become stronger. I can't run a marathon and I can't always discuss issues fairly and dispassionately, but I'm glad both activities exist as things to aspire to.

IV.

> Following sufficient rinsing and repetition, it may occur to someone in a ‘discourse without limitations’ community to wonder where all the (say) queer people and/or women and/or trans* people and/or disabled people and/or people of color and/or non-American-and-Northern-European people and/or citizens of the third world and/or people whose first language is not English and/or Jewish people and/or etc. (repeat and/or for any population ‘coincidentally’ discouraged from participating) went.
>
> Or rhetorical-you could argue that women and/or minorities and/or historically disadvantaged groups are inherently irrational / otherwise not qualified for community membership, at which point I would proceed to avoid rhetorical-you, as above.

You are implying – not saying, but I hope it is fair to read the implication – that “discourse without limitations” drives minority group members away from the communities that participate in it.

This has recently become an interest of mine because a number of communities I’m in – the atheist community, the rationalist community, the Reddit community, the Vaguely Techy Bay Area community – notably lack certain kinds of minorities. And there are many people who say this must be because of some kind of inherent flaw in the community, that it proves either that community members are racist, or at least that they are less actively non-racist than might be desired. Sometimes people say this nicely and helpfully, like you have. Other times people say it more confrontationally, often with the standard “nerdy white dudes” line thrown in.

And always they make the same dichotomy you do – between the “driving these people away” explanation, and the “are you claiming these people are inherently inferior?” explanation. And the proposed solution is always to be more “respectful”, which means talking more about feminism and social justice, and being less accepting of people who counterargue against it.

Needless to say, this is not a solution I can entirely get behind. So I am terribly biased on this point. Still, let me nevertheless present my argument for evaluation.

I have been to several yoga classes. The last one I attended consisted of about thirty women, plus me (this was in Ireland; I don’t know if American yoga has a different gender balance).

We propose two different explanations for this obviously significant result.

First, these yoga classes are somehow driving men away. Maybe they say mean things about men (maybe without intending it! we’re not saying they’re intentionally misandrist!) or they talk about issues in a way exclusionary to male viewpoints. The yoga class should invite some men’s rights activists in to lecture the participants on what they can do to make men feel comfortable, and maybe spend some of every class discussing issues that matter deeply to men, like Truckasaurus.

Second, men just don’t like yoga as much as women. One could propose a probably hilarious evolutionary genetic explanation for this (how about women being gatherers in the ancestral environment, so they needed lots of flexibility so they could bend down and pick small plants?) but much more likely is just that men and women are socialized differently in a bunch of subtle ways and the interests and values they end up with are more pro-yoga in women and more anti-yoga in men. In this case a yoga class might still benefit by making it super-clear that men are welcome and removing a couple of things that might make men uncomfortable, but short of completely re-ordering society there's not much they can do to get equal gender balance and it shouldn't be held against them that they don't.

The second explanation seems much more plausible for my yoga class, and honestly it seems much more plausible for the rationalist community as well.

We're not actually missing all those groups you mention as minorities who might be driven away. In fact, in many cases, we have *far* more of them than would be expected by chance. For example, we contain transgender people at about five times the rate in the general population (1.5% vs. 0.3%), and gays/lesbians/bisexuals at about three times the rate in the general population (15% vs. 4%). People who Jewish by descent are four times the national average (8% vs. 2%), and people with mental disorders are either around equal to the general population or much much higher, depending on how one interprets the data I did a terrible job collecting (sorry). We have more people with English as a second language than almost any other online community I know (the country with most rationalists per capita continues to be Finland) and members from Kenya, Pakistan, Egypt, and Indonesia.

The only groups we appear to be actually short on are women and minorities (and then only if you follow standard American practice of refusing to count Asians as a real minority, numbers be damned).

But just as you would not immediately jump from the overrepresentation of transsexuals to the assumption that we must somehow be discriminating against cissexual people, so one does not jump from the overrepresentation of men to the assumption that women are being discriminated against.

Most rationalists come from the computer science community, which is something like 80% male. A few come from hard science fields like math and physics, both of which are 80 – 90% male. There is *zero* need to invoke "discourse without limitations" as an explanation for why the rationalist community is heavily male-dominated, and any attempt to do so would run into the question of why the occasional dispassionate cost-benefit discussion of eugenics apparently horrifies women but heavily attracts Jews, gays, and people with mental disorders.

Worse, the hypothesis fails in the other direction as well. There are lots of groups that are horribly offensive towards minorites yet nevertheless manage to have very many of them. Across nearly every denomination, far more women than men go to church – if you go to a Catholic Mass, you will see pews full of ladies at levels the atheist community can only dream of. The atheist community is so feminist that there has been a serious movement to replace it with "Atheism Plus" that excludes all non-feminists; the Catholic Church is so regressive that it won't let women become priests and thinks they were created as a "helpmeet" for man. And yet women, in aggregate, love the one and hate the other.

You know what other community has more women than the rationalist community? The men's rights movement. According to the /r/mensrights survey, about 9.3% of men's rights activists are female, which is slightly fewer women than the rationalist community on the last survey, but slightly more women than the rationalist community on the survey before that. A friend who reads Heartiste guesses that about a third of his commenters are female (though adds that some of these may be men who are pretending in order to make a point). So if we actually spent all our time belittling women and justifying their oppression, as far as I can tell our percent female readership would probably go *up*.

I am left pretty certain that the male-dominated rationalist community has a gender imbalance for the same reason as my female-dominated yoga class. We could always see whether it might help to inviting some feminists in, listen to them without protest, and agree to do whatever they say – but I would enjoy that about as much as you would enjoy getting lectured by men’s rights activists without being able to protest, and the end result would probably be about the same.

V.

Apophemi’s essay continues with an addendum:

> there’s significant linguistic signalling that can make up the difference between people who have more to lose from apparently innocent argument participating or not. For (specific to my experience) example…
>
> – arguments against accusations of racism/sexism/cissexism/heterocentrism/ableism/etc. that boil down to “those are silly words and they aren’t in my spellcheck”

I worry that you’re not being entirely fair here. Who the heck doesn’t have “racism” in their spellcheck? I feel like your opponents may be making a more subtle point than you think.

When I ask people to use words other than “racism”, it’s usually because I believe a Worst Argument In The World is being sprung on me – the article will explain more. I think this is a reasonable concern, and it’s always fair to ask someone to taboo their words.

But there’s another problem I sometimes run into with some other concepts, like “male privilege” or “male gaze” or “marginalized”.

You said you enjoyed my Anti-Reactionary FAQ (thanks!), so I wonder whether you enjoyed section 2.3.1, in which I deconstruct the word “demotist”.

The Reactionaries argued that “demotist” countries, meaning countries that had some notion of popular sovereignty including communisms, non-monarchical dictatorships, and democracies – had a terrible human rights record. Which is true – Maoist China (communist), Myanmar under the junta (non-monarchical dictatorship) and many others do have terrible human rights records.

But the Reactionaries were loading the debate by using the word “demotist”, which deliberately groups those regimes together with stable liberal democracies (who have fantastic human rights records compared to anyone else). My argument here was *exactly* that “demotist” wasn’t in my spellcheck and that in order to “win” the debate the Reactionaries had to invent new words that loaded the argument in their favor. Denied the ability to use their own words and forced to use the same vocabulary as the rest of us, their argument totally falls apart.

Not everything must be stated in ordinary language – if you didn’t let chemists use terms like “valence electron” or "ionic", you would be denying them a useful tool that makes chemistry much easier. I get that.

But when people are trying to talk about ordinary processes, and they insist on using their own words which don't exactly correspond to features of the world, and they can't always make the same arguments with more standard words, I get *super* suspicious.

Words are hidden inferences. They encode assumptions, and sometimes those assumptions are correct and other times they are wrong. This is true more than usual with jargon, and even more than usual with partisan political jargon (don't call them "rich", call them "job-creators"!) It is useful and acceptable to ask people to take a step back from their words to examine whether the assumptions behind them are correct.

> – conflating terms describing marginalization (such as the above) with insults (i.e. "calling me racist is an insult", "let's discuss this without using meaningless insults like 'misogynist'"),
> – use of insults that have a history of being specific to women or that effectively mean "this person is like a woman"

Oh God oh god oh god oh god you *so* do not understand oh god.

Which words are or are not slurs is not a feature of the word's etymology or even the intent of people using those words. For example, the word "Jap", on its own, is very clearly just a convenient shortening of "Japanese person" in the same way that "Brit" is very clearly a convenient shortening of "British person".

Yet it is *not* okay to go around calling Japanese people “Japs” and then lecturing them because they are “conflating” a term describing their heritage with an insult (“ha ha, that silly Japanese person thinks I’m insulting her just because I used a shortened form of her demonym”).

Most Japanese people have a history – maybe personal, maybe just second-hand – of *correctly* associating the word “Jap” with an attempt to dehumanize them, marginalize them, or cause them huge amounts of personal grief. It doesn’t matter whether *you* think “Jap” was meant to be offensive, if a Japanese person tells you they’re triggered by it and you keep using it, you’re a jerk.

(and the same is true of a much more famous slur which is a derivation of the perfectly innocent Latin word *niger* meaning “black-colored”, but which has been wrenched *far* away from that perfect innocence by the referents of the term having *more* than enough opportunities to associate that word with an attempt to hurt them.)

So a neutral word can become an insult or trigger or slur if it is associated sufficiently strongly and sufficiently often with people trying to hurt you.

Now, when those people were sending me death threats because of that article in the college paper, what word do you think they used?

When the media talks about a “scandal” in which some politician or actor is accused of being offensive and then gets fired from their job and has to do a live apology on national TV during which they break down crying, what word do you think always starts the process?

When you read the MsScribe story – which a dozen people in the comments said struck incredibly true to life for them –

what word did MsScribe use to deride her enemies before kicking them out of the community and making everyone refer to them as “cockroaches” and posting sexually explicit stories about them doing horrible things?

People have an *incredibly reasonable terror* of that specific word, and when you refuse to change it to one of many dozens of available synonyms, that has some pretty strong implications about where you are coming from. It says “I don’t respect you enough not to use this word that terrifies and triggers you” which in turn means that people’s terror and triggering is probably correct.

I am sure there are some lovely elderly Southerners who use [the n-word] simply because that was what they grew up with, and are mildly annoyed every time a black person throws a fuss about it because they honestly didn’t mean any harm. And they use that exact same argument: “I didn’t mean anything by it, it’s just what I call people like you, you’re so sensitive treating it as an insult.” But they are missing the point. It doesn’t matter what *their* feelings are, it matters whether it hurts other people.

And when they *anticipate* this, like “Oh, I’m going to call that black person the n-word, and I bet he’s going to get all upset about it, you know how they are”, *that* doesn’t seem innocent to me. It sounds like they know they’re hurting other people and just don’t care.

And when you say you *expect* people to feel insulted and triggered by the word “racist”, but you’re going to do it anyway, even though you are perfectly aware of other words you could use that would actually be more descriptively accurate, I kind of have the same worry.

And then your very next point is that you don’t want people to use terms you consider slurs. Well, yes, of course that is fair! And I try to avoid slurs as much as I can.

Yet I cannot help rounding this entire section off to “The two things that annoy me are when other people use language that triggers me, and when other people ask me to stop using language that triggers *them*.”

And when I have bring this up to people, they usually answer “It’s impossible to trigger a member of a privileged group” or “Triggering a member of a privileged group doesn’t count”. I am so happy you have defined away my pain. THIS IS THAT DEHUMANIZATION THING AGAIN.

In conclusion, aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaah.

> Sense of being persecuted by “political correctness”…I have failed so far to find a definition of “political correctness” in this context that could not be search-and-replaced with “trying to avoid hurting people” to either no effect or increased comprehensibility. You are free to attempt to change my mind on this, I guess?

I like this sentence because it is a good example of language in fact making a difference, of words being hidden inferences, of reasonable requests not to use terms that don’t just boil down to “it isn’t in my spellcheck”. In fact, Scott Adams makes this exact point in his essay What’s The Difference Between A Sexist And A Regular Asshole? It intrigues me that both sides are trying to remove the others’ linguistic weapons by demanding they be deactivated and replaced with normal words, but are refusing to relinquish their own. Anyhow…

When I see references to “political correctness”, it’s usually followed by something like “has gone too far”. This suggests a reasonable interpretation to me – political correctness is indeed a way of trying to avoid hurting people, but like all forms of trying to avoid hurting people, it can go too far.

Trying to prevent terrorism is good. But when any vaguely Muslim looking person who tries to board a plane tends to get hauled off and strip-searched without so much as an apology, one can ask whether the legitimate goal of trying to prevent terrorism has gone too far.

Likewise, when people start saying that it’s cultural appropriation to eat latkes or a ten-year old girl can be charged with rape for playing a game of Doctor or heterosexual white people can’t be depressed or any of the other three million things of this sort I see on Tumblr every day, then I do think it’s fair to say that the legitimate goal of trying to protect disadvantaged groups is going too far *in certain cases*.

This is not to say that it has uniformly gone too far in every aspect of society, just that in these cases – the ones the people saying this have encountered – it has *locally* gone too far.

I do not really know what claim you are asserting – that political correctness never goes too far? That no one trying to protect the rights of minority groups has ever overstepped good sense? This seems more like cheering on a side than stating a defensible position to me.

See also Section II of this essay.

> – pretty much any usage at any point of the word “insane” when we are not talking about a court case by now

Dammit, you just broke my girlfriend.

Ze is a mentally ill person who has attempted suicide a few times and been in and out of mental hospitals, and ze is now *seething* with anger. I think I see smoke coming out of zir ears. Now ze is *demanding* that I write an extremely angry response saying HAVEN'T YOU EVER READ ANYTHING IN THE DISABILITY RIGHTS COMMUNITY??!?! and DON'T YOU KNOW THAT LOTS OF PEOPLE USE TERMS LIKE "INSANE" AND "CRAZY" TO AVOID MEDICALIZING THEIR DISABILITY??!?! and HOW DARE YOU PURPORT TO SPEAK FOR ALL OF US WHO THE HELL APPOINTED YOU OUR REPRESENTATIVE??!

As a psychiatrist myself, avoiding medicalizing disability is not really high on my list of priorities. But as a "mentally ill person" myself, – two years on Paxil followed by eight on Prozac followed by two years of behavioral psychotherapy followed by the *incredibly enjoyable* process of finding a hospital that wanted to employ a psychiatrist with a mental illness because I knew I wouldn't be able to hide it through however many years of work I will be with them – avoiding the use of the term "insane" has never been high on my list of priorities either.

I have never, ever, noticed the pattern you have – people who use the word "insane" being otherwise bad or de-legitimizing people with mental illness. In fact, it has happened more than once to me – twice, I think, spookily similar – that a mentally ill patient asks me what my diagnosis is, I say something like "schizophrenia" or whatever, and they say "Nope! I'm insane! If you want to be a good doctor, you're going to have to learn to tell it like it is!"

Both my girlfriend and I agree that people being very concerned about people using or not using specific very common words has been a much bigger warning flag of someone who is otherwise not a nice person than use of the word “insane”.

…which is not to say that you haven’t had the exact opposite experience! That’s kind of the problem. No one can speak for an entire community and community members have very different experiences and preferences. My policy so far has been to always respect someone’s terminology preferences when talking to them personally or in a small group, and to respect terminology preferences I know to be common when talking to a large audience. In a lifetime of working with the mentally ill and dating two different disability rights bloggers I have yet to hear anyone else express a strong preference against “insane”, but if it happens more often I will update. And I will certainly avoid doing it if I ever have reason to talk to you directly.

> If by “sluttiness” r-you mean “sexual promiscuity”, what is gained by using a gender-targeted insult that is likely to make a significant portion (i.e. women and/or queer people, who together are like… 55% of the world at least) of r-your potential audience uncomfortable and less likely to engage with r-your argument?

This I apologize for unreservedly. In the Anti-Reactionary FAQ, I quoted some reactionary passages using the word “sluttiness”, and then I continued using it myself afterwards. I was hoping to kind of mock the reactionaries by pointing out how much their argument depended on that one word. In the process, it seems I offended some women/queers as well. I was wrong to do this, there were very easy ways to avoid it, and I will avoid it in the future.

VI.

You probably aren’t even reading this, but I hope someone like you is.

Aretaeus On Bipolar Disorder

I remember reading The Americanization of Mental Illness four year ago when it was written and being generally impressed by its thesis. Every culture has “culture-bound syndromes” (I recently pointed out puppy pregnancy syndrome as an especially horrifying example) and with American hegemony we risk declaring that America’s forms of mental illness are “real mental illness” and everyone else’s forms are just a weird local variation. Sometimes it’s neat to learn that what you thought were laws of nature are merely your own culture’s idiosyncrasies.

On the other hand, sometimes it’s nice to learn you were actually right about everything.

Aretaeus of Cappadocia was a 1st century AD Greco-Roman physician who is notable for describing bipolar disorder almost exactly as it is described today. He says: “It appears to me that melancholy is the commencement and a part of mania” and goes on to describe symptoms of the disorder.

I have in front of me the bipolar chapter of both Aretaeus’ De causis et signis acutorum morborum and of the modern Mood Disorder Questionnaire, an instrument used for diagnosing bipolar. I’m going to go through the MDQ questions and see how many I can match to sentences in Aretaeus:

“Do you ever feel so good or hyper you are not your normal self?”

Aretaeus: “And they with whose madness joy is associated, laugh, play, dance night and day”

"…or so hyper it got you into trouble?"

Aretaeus: "They become silly, and doing dreadful and disgraceful things"

"Do you feel so irritable that you start fights or arguments?"

Aretaeus: "They are suspicious, irritable without any cause…others have madness attended with anger; and these sometimes rend their clothes and kill their keepers, and lay violent hands upon themselves. This miserable form of disease is not unattended with danger to those around."

"Do you feel much more self-confident than usual?"

"Aretaeus: [They] sometimes go openly to the market crowned, as if victors in some contest of skill".

"Do you get much less sleep than usual, but find that you don't miss it?"

Aretaeus: "When [tending] to cheerfulness, they are in excellent spirits; yet they are unusually given to insomnolency"

"Do you find that you talk much more than usual?"

Aretaeus: "But the modes are infinite in those who are ingenious and docile,–untaught astronomy, spontaneous philosophy, poetry truly from the muses"

"Do you find you are much more interested in sex than usual?"

Aretaeus: "At the height of the disease they have impure dreams, and irresistible desire of venery."

“Do you find that spending too much money gets you in trouble?”

Aretaeus: “[They can become] simple, extravagant, munificent, not from any virtue of the soul, but from the changeableness of the disease.”

…and six in the Mood Disorder Questionnaire that I couldn’t find a good analogue for in Aretaeus, but which seem to be in accordance with the spirit of his piece. So I will give Aretaeus a 7/13 or so on mania.

II.

Moving on to depression, we have the Patient Health Questionnaire-9, probably the most commonly used depression screening tool today.

“Do you feel little interest or pleasure in doing things?”

Aretaeus: “The understanding is turned…to sorrow and despondency only.”

“Trouble falling asleep, or sleeping too much”

Aretaeus: “And they also become peevish, dispirited, sleepless, and start up from a disturbed sleep.”

“Feeling tired or having little energy”

Aretaeus: “The patients are…unreasonably torpid, without any manifest cause: such is the commencement of melancholy.”

“Poor appetite, or overeating”

Aretaeus: “But if the disease go on to increase, they are voracious and greedy in taking food, for they are watchful, *and watchfulness induces gluttony…But if any of the viscera get into a state of inflammation, it blunts the appetite and digestion.”*

“Being so fidgety and restless that you have been moving around a lot more than usual”

Aretaeus: “Wherefore they are affected with madness in various shapes; some run along unrestrainedly, and, not knowing how, return again to the same spot.”

“Thoughts that you would be better off dead, or of hurting yourself”

Aretaeus: “If the illness become more urgent – hatred, avoidance of the haunts of men, vain lamentations; they complain of life, and desire to die.”

I’m giving him 6/10 on the depression side.

III.

So on mania and depression combined, Aretaeus gets 13/23 of the same criteria we use today, which is not bad at all.

I admit that some of the analogies are a little forced, but I think any knowledgeable person reading Aretaeus’ book would have to admit he is *generally* on the ball with his description of bipolar, not only in *gestalt* but also in many of the individual symptoms.

His theories of pathogenesis and treatment are, of course, totally off-base. I think he thinks they’re caused by warm or cold or dry or wet weather and imbalances in the four humors? But that goes with the territory.

What impresses me is that people with bipolar disorder in ancient Rome seem to have behaved a lot like people with bipolar disorder today.

And that seems like pretty solid evidence that this disorder, at least, is firmly biologically grounded and not culture-bound.

(which we already knew. Bipolar is very heritable and seems to occur about equally in all nations and cultures. But it’s nice to have confirmation.)

Ten Things I Want To Stop Seeing On The Internet In 2014

1. The word “fanboy”

“Fangirl” is a perfectly acceptable word with an illustrious history; it simply means a young woman who enjoys good literature and who may or may not have married Snape on the astral plane.

“Fanboy” – as used in terms like “Microsoft fanboy” or “Bitcoin fanboy” – is not. The basic assumption is: you like a thing, therefore you suck.

Seriously, f@&k people who like things.

2. The word “butthurt”

“Butthurt” is one of those magic words that completely suspends human decency.

For example, suppose a headline reads: “Jewish community saddened over swastika graffiti on synagogue”. You might make the mistake of feeling sympathetic to these Jews, or believing they’re human beings whose feelings matter.

So we change the headline to “Jewish community butthurt over swastika graffiti on synagogue.” Haha! Stupid Jewish community! Always getting butthurt over stuff!

(Even better: “Torah fanboys butthurt over swastika graffiti on synagogue.”)

3. People saying “Just because there’s a cold spell or worse-than-average blizzard doesn’t mean global warming is false!”

This is, of course, completely true and very important. On the other hand, the moment there is a warm spell or worse-than-average hurricane, these very same people immediately start talking about how this proves global warming is real. So this argument is now banned. Once you learn to *consistently* separate single data points from meaningful trends, I will allow you to use it again.

I realize my stance may result in the Earth becoming a giant cinder and everyone being boiled alive, but it’s the principle of the thing.

4. Doge

All memes go through three stages.

In the first stage, a mildly amusing image or video catches some people’s attention and they show it to others. For example, someone notices a hilarious translated sequence involving the phrase “All your base is belong to us” and email it to friends.

In the second stage, people riff on this theme to create other similar memes, some of which occasionally have comedic value. For example, a chemistry site might say “All your alkaline is belong to us”. Or a math book teaching hexadecimal might say “All your base 16 are below to us.” Okay. Fine.

In the third stage, people abandon all pretense of humor or creativity, and expect to be rewarded merely for mentioning the meme’s existence to other people equally aware of it. “HEY! GUYS! ALL YOUR BASE IS BELONG TO US! LAUNCH EVERY ZIG FOR GREAT JUSTICE! HAHAHAHAHAHAHAHA!”

I’m not sure doge ever actually had stage one or two.

5. Meta-humor

Someone reading this post right now is thinking “I’m going to post in the comments that I think Scott is acting butthurt over these Internet things he’s criticizing! And then I’m going to call him a fanboy of something!” *Even after reading this paragraph*, somebody is going to try this.

Listen very closely. There is a time and a place for meta-humor, the sort of humor where you try to be funny by applying something to itself. The time is when you are Douglas Hoftstadter. And the place is in a Douglas Hofstadter book.

In all other situations, it is so algorithmic a computer could do it, and so obvious the people you are trying to amuse will have no doubt already thought of it.

Somewhere in a volcano lair, there is a supervillain working on a doge meme that says “SO DOGE. MUCH MEME. MANY COMIC SANS.” I just hope some government agency is assembling a team of plucky misfits to stop him.

6. Appeal to “I bet he lives in his mother’s basement!”

Get it! He’s *poor*! He has *low socio-economic status!* Haha! That’s funny!

7. “Dude”, “bro”, or God forbid, “dudebro”

I had no problems with “dude” when its use was limited to cowboys, surfers, and giant anthropomorphic turtles. And I had no problem with “bro” when it was used by members of fraternities to refer to their brothers in those fraternities.

But somehow when I wasn’t looking these words turned into things Internet feminists call anyone they dislike. The usual referent seems to be geeks and people who disagree with them about gender issues. I can only imagine the picture the average Internet feminist must have of fraternities to allow this state of affairs to continue:

> “Hey, Chad, let’s go grab a brewski and pick up some hot chicks!”
>
> “Bro, that sounds totally awesome, but you know I’m busy writing a Haskell program to determine whether the gender pay gap disappears once you adjust for confounders. Wooooo!”
>
> “Whoa, Chad, you never told me you were WRITING A F@&KING HASKELL PROGRAM TO DETERMINE WHETHER THE GENDER PAY GAP DISAPPEARS ONCE YOU ADJUST FOR CONFOUNDERS! F@&k chicks and brewskis, I’m staying here to help you out! DELTA CHI, ALL THE WAY! WOOOO!”

Obviously none of the people involved have ever seen the inside of a university, and although on its own that just proves that the college admissions people are marginally competent at their jobs, I can’t help but wonder how this rumor got started. Has some ultra-high-level troll just been going around telling people this is how colleges work? If so, kudos to you.

Now stop it. *Now.*

8. Arguments about which generation is better

This generation is made entirely of upstanding, noble folk who made this country great. But *that* generation is made entirely of whining, entitled brats who are destroying everything their ancestors and/or descendants worked for. Clearly that means this generation is better than that generation!

9. Hating on girls with male friends

“Haha, how about those girls who say they prefer to be ‘one of the guys’ or that most of their friends are male, huh? Isn’t that always a huge warning sign?”

I see this so often, most recently on Reddit. “Never trust a girl who doesn’t have female friends” or “They don’t get along with other girls because they themselves are usually bitches”, just to choose two comments from the thread on this pressing issue.

Well, I’m the male version of this. Most of my friends are girls. This is none of your business. And if a girl does the same thing, that’s none of your business either.

Also, people trying to earn your social justice halo by calling this “a big red flag for internalized misogyny”? You would be a lot more believable if you didn’t add “…and I bet they’re huge sluts” in the very next sentence.

In conclusion, I propose a compromise:

1. Girls can make friends with whoever the f@#k they want.
2. Go to hell.

(I am not very good at compromises)

10. Doge memes referencing the actual doge of Venice

I really should have included this in with doge memes in general, or with meta-humor. Yet here it is in its own category.

Look. I don’t know why you choose to misspell “dog” as “doge”. I’m sure you have your reasons. But once you’ve deliberately confused the word for a small furry animal with the word for the ruler of Venice, then acting like you are *extremely clever* for pointing out that the ruler of Venice is *also* spelled like the ruler of Venice is not “edgy” or “creative”. It’s like…taking a picture of a bale of hay, misspelling it “ball” instead of “bale”, and then jumping up and down excited because you suddenly realized “ball” *is also the word meaning “ball”*. No one would possibly try this, right?

And yet it has come to this (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)

And if twelve images of one doge aren’t enough for you, somebody kindly decided to make one image of twelve doges. Which got picked up by the *Washington Post*. You know, I remember when the *Washington Post* was uncovering the Watergate Scandal. Now they are talking about how funny it is that doge can also, in some cases, mean doge.

Well, now it is 2014, and this sort of thing is no longer acceptable. Up your game, Internet. *Up your game.*

Genetic Testing and Self-Fulfilling Prophecies

Lineweaver et al tested 144 elderly adults for the ApoE4 gene, which is known to be a major risk factor for Alzheimers. They told half of them their test results, kept it secret from the other half, then waited. Eight months later, they asked people how they thought their memory was doing and gave everyone objective memory tests.

No one in the study population had Alzheimers yet, so everyone did okay on the memory test. But subjects who knew they had ApoE4 did significantly worse than subjects who did have ApoE4 but didn't know it. The subjects who knew they *didn't* have ApoE4 didn't do any better on the memory test than other subjects who didn't have ApoE4 but had not yet heard the good news, but they did give subjectively better ratings of their memory ability.

The medical community concludes from this that letting people know their genetic risks may be dangerous, and although I hate to admit it, they have a point.

Unfortunately, the study doesn't give the methodological details I need to really understand the implications.

We know that the researchers waited eight months between giving the genetic test results and doing the memory tests. That's good.

But was it the same researcher doing both the genes and the memory parts of the study? Was it in the same building? Did they start the memory tests by saying "Hi! I'm Dr. Lineweaver! You may remember me from such medical experiments as the one eight months ago in which you were told that you had a high risk of getting Alzheimer’s disease”?

Or did they sneakily pretend to be a separate study entirely and try to avoid mentioning the A-word throughout?

It would not surprise me if – having been primed with a reminder of their Alzheimer testing results – the subjects then performed worse on a memory test that was given immediately after in an obviously related context.

It would be much more surprising – though still not totally unbelievable – if subjects, having been told they had a high risk of Alzheimers, just went around for eight months having slightly worse memory which was reflected on everything they did including the memory test administered by the researchers.

New England Journal of Medicine compares the finding to “stereotype threat”, the phenomenon in which people can for example sometimes make women perform worse on math tests simply by telling them that it is a “test of their innate mathematical abilities” – something that women are stereotypically bad at.

The memory tests the researchers were giving are equivalent to the “innate mathematical abilities” condition in the stereotype threat research – a test clearly intended to measure how good their memory was in a very scientific way. The activities of daily living that require memory – keeping appointments, paying bills on time, et cetera – are the equivalent of the condition in stereotype threat experiments where researchers just give women a normal math test without introduction and stereotype threat is not seen.

So I see two ways in which we could get results like the ones in this study without any broader implications of ApoE4 testing harming the elderly in general.

First, being called to the same study in which the ApoE4 results were given could have primed their worries about Alzheimers and made them do especially bad on the study’s memory test compared to their usual memory.

Second, the study’s memory test could have been official-looking enough that it activated their stereotype of themselves as having innately poor memory, when, concordant with stereotype threat research, that stereotype doesn’t harm their everyday memory-requiring activities.

In either of these cases, the study would have some very limited implications, which the authors describe in an appropriately circumscribed way: “The patient’s knowledge of his or her genotype and risk of Alzheimer’s disease should be considered when evaluating cognition in the elderly.”

But this would not imply that genetic testing elderly people for ApoE4 is risky and can itself cause them to develop forgetfulness and other Alzheimer’s symptoms.

I worry that the medical community is going to miss this subtlety and start “raising awareness” of the possibility that genetic testing can cause harmful side effects, before finishing the hard task of discovering if that’s actually true.

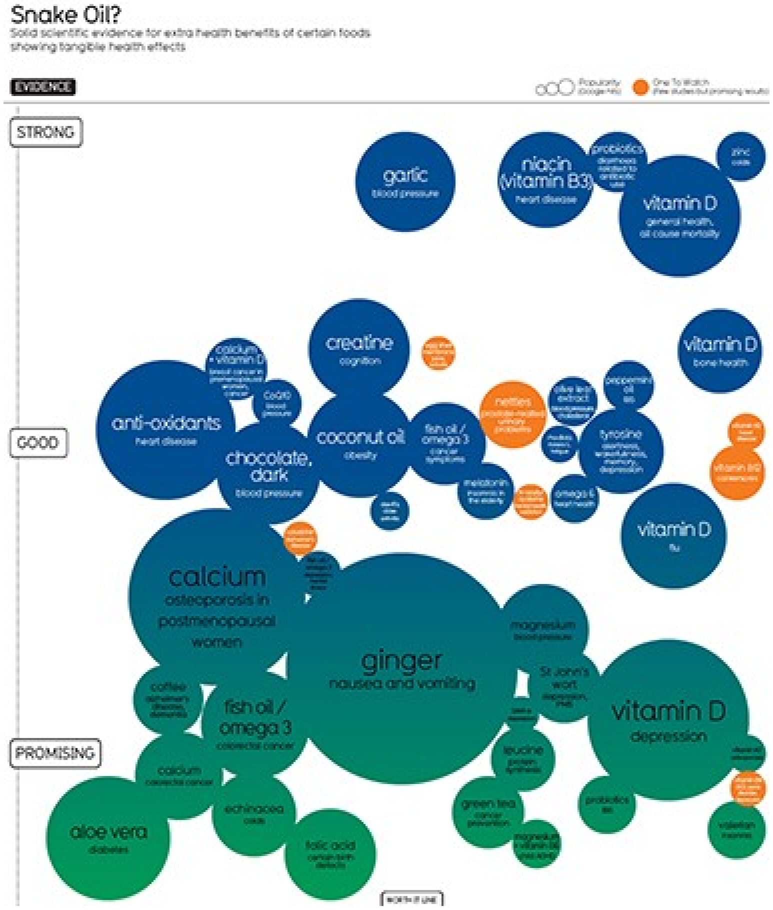
Beware Mass-Produced Medical Recommendations

When I was younger, my father gave me some advice I will never forget.

“Scott,” he said “when somebody tells you they want to help you figure out the evidence base behind different supplements, ask them about Vitamin D. If they say it’s useful for anything besides bone health, run away.”

(This actually happened. I know other people have fathers who give advice about Dealing With Adversity or Finding The Right Person, but I like mine just fine. You can pick up Dealing With Adversity as you go along, but evidence-based medicine is *important*)

So when four people on my Facebook wall posted a link to a pretty graph with lots of circles about supplements:



Click on picture to see full version – but you might want to read the rest of this first

the first thing I did was look for Vitamin D. And there it was, all the way on the top, having “strong” evidence for “general health” and “all cause mortality”.

I have no doubt that the people putting this together are very smart and very well-intentioned. They drew from exactly the right places – they even checked Cochrane Review! They made all their research public and linked to a bunch of studies proving everything they said – including their claim that Vitamin D is effective for “general health”. So is my father wrong?

No. Their problem is that they are one-sided. Not in the sense of being deliberately biased or having an agenda, but in the sense that their entire procedure consists of searching for studies showing X, counting the number of such studies they find, and then making an “evidence strength” recommendation based on the number (and quality) of positive results.

But suppose we have a very popular supplement that dozens of good randomized studies have been done on. And suppose that ten of these studies have shown it prevents cancer, and so these sites put ten links to studies showing it prevents cancer, give it evidence grade A++, and put it at the top of their list as very highly recommended. This is no good if there are twenty much larger, better-quality studies that show it *doesn’t* prevent cancer, and this is in fact what’s going on in the vitamin D case.

II.

Let’s look at this more closely. The pretty colored circle people say that Vitamin D increases “general health” and prevents all-cause mortality. I don’t know exactly what “general health” means, but all-cause mortality is clear enough, and they also have a circle saying it prevents cancer. The research in these two areas is closely linked, so I’ll deal with it together.

They give seven studies that they use to make their point. Of these, two (1, 2) are primarily theoretical works detailing the molecular mechanisms by which vitamin D might prevent cancer and heart disease – solid as far as they go, but the graveyards are littered with convincing theoretical explanations for effects that were later discovered not to exist. Another (1) notes that absurdly high dose vitamin D was found in a trial without a control group to be associated with good treatment response in existing prostate cancer, which doesn't tell us anything about anything except that particular rare situation and probably not even that. Three (1, 2, 3) are correlational studies – which, again, are a valid type of study, but have some pitfalls. And they *especially* have pitfalls in vitamin D research where every couple of months someone breathlessly announces that their correlational study has found vitamin D protects against Disease X, when what they actually mean is that Disease X (like practically every other disease) decreases serum vitamin D levels and so the disease state is associated with low Vitamin D levels. A single study (1) is a very good randomized trial in a thousand women finding vitamin D administration decreased cancer.

[edit: a newer version of the pretty circle graph includes some different studies. I may rewrite this part later, but it doesn't change their conclusion and so doesn't change mine either]

This evidence isn't 100% conclusive, but it's certainly suggestive, and I would have no problem with their claim that vitamin D probably decreases cancer and heart disease *if* that was all we know. But…

The largest randomized trial ever done on the subject, a 36,282-person behemoth, found zero effect of vitamin D on its two measured endpoints of colon cancer or breast cancer and in fact the vitamin D group had nonsignificantly *more* cancer than controls. Another large case-control study of 1400 people looking at prostate cancer found that vitamin D didn't prevent it and might actually make it worse. And a randomized controlled trial of 2700 people investigating all-cause mortality found zilch. And this is ignoring all the little correlational studies that find vitamin D increases cancer risk.

In cases where you have lots of conflicting studies, you go to the meta-analyses. A meta-analysis by the Systematic Evidence Reviews people (who know their stuff) concluded that “Vitamin D and/or calcium supplementation also showed no overall effect on CVD, cancer, and mortality.” Wang et al found much the same (although their conclusions section does a terrible job elucidating this). Autier looks at 172 randomized trials (!) and finds “Results from intervention studies did not show an effect of vitamin D supplementation on disease occurrence”. In a meta-analysis of eighteen trials, Pittas et al find ” no clinically significant effect of vitamin D supplementation at doses given.” The Agency for Healthcare Research and Quality did a meta-meta-meta review of eleven systematic reviews (!) and found only that “the majority of the findings concerning vitamin D, calcium, or a combination of both nutrients on the different health outcomes were inconsistent.” And just last week, a very large analysis of hundreds of thousands of patients’ (!) worth of data concluded that Vitamin D supplementation does nothing for “skeletal, vascular, or cancer outcomes” and that the evidence is so clear that it is pointless doing further trials on the issue (this is the first time I heard of a “futility analysis” in statistics, but I like it!). Although there is the occasional meta-analysis that find in vitamin D’s favor, the medical community doesn’t find them very exciting.

So what is the field’s conclusion from all this disparate data?

UpToDate, the semi-canonical database of recommendations for doctors, says “A causal association between poor vitamin D status and nearly all major diseases (cancer, infections, autoimmune diseases, cardiovascular and metabolic diseases) has not been established. We suggest not administering vitamin D supplements above and beyond what is required for osteoporosis or fall prevention.”

The National Institute of Health says: “Taken together, however, studies to date do not support a role for vitamin D, with or without calcium, in reducing the risk of cancer…with the exception of measures related to bone health, the health relationships examined were either not supported by adequate evidence to establish cause and effect, or the conflicting nature of the available evidence could not be used to link health benefits to particular levels of intake of vitamin D.”

The US National Preventative Services Task Force says: “The USPSTF concludes that evidence is lacking regarding the benefit of vitamin D supplementation, with or without calcium, for the primary prevention of cancer” and gives the practice a grade of D, meaning they recommend against.

The Lancet, the second most prestigious medical journal in the world, ran an editorial by Karl Michaëlsson this month saying that “there is a legitimate fear that vitamin D supplementation might actually cause net harm”.

But what about someone *really* trustworthy? Wikipedia’s Vitamin D article says simply: “Research has found it unlikely that taking vitamin D supplements has any effect on cancer” and “It unlikely [sic] that taking vitamin D supplements has any effect on cardiovascular disease.”

I don’t want to come out too hard against vitamin D. There’s just enough tantalizing hints that, futility analyses be damned, I would not be surprised if later research came out in its favor. And there are some proposed benefits – like depression and fatigue – I haven’t even looked into – and some other benefits like psoriasis that seem to be on firmer ground. If your doctor told you to take vitamin D, or you’re doing well on it at the moment, don’t go off it just on my say-so.

But when the pretty colored circle people say it is the supplement with the *most* evidence? When they describe the evidence in its favor as “strong”? That’s just wrong.

(Or at least misleading. It probably is the supplement with the most evidence in its favor. It just has even *more* evidence against it. This is why you can’t just count the number of positive studies.)

III.

The recommendation for vitamin D is mostly harmless. The purported risks are on even shakier ground than the purported benefits, and unless you gobble vitamin D pills like candy, you probably won’t give yourself anything worse than a slightly increased risk of kidney stones (although I feel the need to add that kidney stones *really* hurt).

But go back to the pretty colored circle chart. Although vitamin D is on the highest echelon of their chart, it is not literally the highest circle. If you really squint, it looks like one other circle might be a few pixels higher than it. That circle is niacin.

…ie pharmacology’s most embarrassing disaster story of the past five years.

There was a time when niacin would have deserved that high spot. In 2010, a meta-analysis of niacin in treating cardiovascular disease found impressive success for the substance, building on about a thirty year history of smaller, less conclusive trials that nevertheless showed similar levels of success. This is the study cited by the pretty circle chart.

In 2011, Merck wanted to market their own form of niacin and decided to do what they expected to be a tick-the-appropriate-checkboxes type study to prove it was effective (which, in accordance with the ancient tradition of cutesy study acronym names, was called AIM-HIGH). Instead, the niacin was so useless that everyone involved gave up and stopped the study early because it wasn't even worth their time to continue (caveat: this result may have been influenced by a second drug, laropiprant, given with the niacin). This was likely a good decision, since around the same time they started to notice patients in the experimental group getting way more strokes than seemed natural (the significance of this finding continues to be debated).

This study was followed by a second (in accordance with tradition named THRIVE), which was a mixed success. On the one hand, niacin continued to be absolutely useless in promoting cardiovascular health. On the other, we learned all sorts of exciting new side effects of niacin, like skin rashes, diabetes, increase infection, and my favorite, internal bleeding in the brain. Something like 3% of study participants required hospitalization for niacin-related side effects

So, long story short, now we don't give people niacin so much. I won't say doctors don't give it at all, because a few do, especially in patients whose cholesterol can't be controlled any other way, and if your doctor has you on niacin please at least ask why before going off it.

But once again, the pretty circle graph ignores all this subtlety and puts it on the *highest* rung of their chart, because their methodology of counting the studies that support a supplement ignores the very important ones that don't.

IV.

In the discussion on Facebook, a friend agreed we should steer clear of this analysis and instead use examine.com, a very large and comprehensive database of supplement information.

But examine's page on Vitamin D *also* gives it a Grade A evidence recommendation in preventing cardiovascular disease and grade B recommendation in preventing cancer. The cardiovascular disease entry says the scientific consensus is "50%" across "three studies" (there are dozens to hundreds), and the colorectal cancer risk says the scientific consensus is "100%" across "one study" (again, dozens to hundreds, with likely the majority against it). Perhaps mercifully, there is no page on niacin.

My father thinks examine.com is "a shill for the supplement industry", but I disagree. I think they are good people, but they are trying to mass-produce medical recommendations. And at least the way they do it, that doesn't work.

Examine boasts four hundred different supplements and 25000 references to scientific papers. Each supplement can have as many as forty different uses (vitamin D to prevent cancer, vitamin D to increase energy, vitamin D to strengthen bones, etc). Do you think they hired a doctor or biologist to laboriously go through every single one of those forty different uses of vitamin D, hunt down all the literature on it, read all the literature on it, read the systematic reviews of the literature, and come to a conclusion? And then repeat this process four hundred times?

I think someone Googled "vitamin D cancer", took whatever studies appeared on the first page of results, and entered them into the database. If those studies were in favor of Vitamin D – even if they were from 1995, or had twenty people, or there were other studies on the second page of the Google results that contradicted them – then the scientific consensus is “100% in favor” and the evidence grade is “A”.

[EDIT: Sol Orwell from Examine.com has responded to this claim (1, 2. He says the important parts of his recommendations are not mass-produced, agrees that the Vitamin D information could be improved, and says he is going to fix the way it is presented. I still think you should use more academic or official sites like UpToDate and NIH first for recommendations, and then Wikipedia which is startlingly good for what it is, but I continue to think examine is a good and on-the-level supplementary (no pun intended) resource])

This doesn’t mean examine.com isn’t a useful resource – I purchased their Supplement Guide, and I constantly find them to be a good jumping-off point for further research. It doesn’t even mean the pretty circle graph doesn’t have something to teach you. But treat mass-produced medical recommendations like this the same way your high school teacher told you to treat Wikipedia, back in the days before everyone realized that Wikipedia was actually much more accurate than whatever crappy second-rate source your tech-illiterate high-school teacher wanted you to use: as something that can occasionally bring interesting facts to your attention but which should probably be double-checked with something more official before you start taking a chemical that has “bleeding into the brain” as a side effect.

And remember, when in doubt about any supplement, the best course of action is always to consult your doctor.

(…is A Sentence I Am Required To Say. In reality, your doctor will probably have never heard of the supplement before and just say ‘no’ out of arrogance or on general principle. And then you will not take it and so you will not get bleeding into your brain and die. In conclusion, the system works.)

Predictions For 2014

(inspired by New Years’ Prediction Thread. To make life easy for myself, I’m limiting myself to 50, 60, 70, 80, 90, 95, and 99. Unless otherwise specified, predictions are about whether something happens before the end of 2014, except LW Survey predictions which will be judged whenever next survey is finished)

Global Predictions

1. Obamacare will survive the year mostly intact: 80%
2. US does not get involved in any new major war with death toll of > 100 US soldiers: 90%
3. Syria’s civil war will not end this year (unstable cease-fire doesn’t count as “end”): 60%
4. Bitcoin will end the year higher than $1000: 70%
5. US official unemployment rate will be < 7% in Dec 2014: 80% 6. Republicans will keep the House in US midterm elections: 80% 7. Democrats will keep the Senate in US midterm elections: 60% 8. North Korea's government will survive the year without large civil war/revolt: 80% 9. Iraq's government will survive the year without large civil war/revolt: 60% 10. China's government will survive the year without large civil war/revolt: 99% 11. US government will survive the year without large civil war/revolt: 99% 12. Egypt's government will survive the year without large civil war/revolt: 60% 13. Israel-Palestine negotiations remain blocked, no obvious plan for Palestinian independence: 99%
14. Israel does not get in a large-scale war (ie >100 Israeli deaths) with any Arab state: 95%
15. Sochi Olympics will not be obvious disaster (ie spectacular violence, protests that utterly overshadow events, or have to be stopped early): 90%
16. Putin will remain President of Russia at end of 2014:

95%
17. Obama will remain President of USA at end of 2014: 95%
18. No nuclear weapon used in anger in 2014: 99%
19. No terrorist attack in USA killing > 100 people: 90%
20. No mass shooting in USA killing > 10 people: 50%

Online Community Predictions

1. Republic of Shireroth will not officially disband or obviously die before end of 2014: 90%
2. Republic of Shireroth will remain in Bastion Union: 80%
3. Active population of Shireroth on last 2014 census will be between 10 and 20: 70%
4. Slate Star Codex will remain active until end of 2014: 70%
5. Slate Star Codex will get more hits in 2014 than in 2013: 60%
6. At least one 2014 Slate Star Codex post will get > 10,000 hits total: 80%
7. No 2014 Slate Star Codex post will get > 100,000 hits total: 90%
8. 2014 Less Wrong Survey will show higher population than 2013 Survey conditional on similar methodology: 80%
9. 2014 Less Wrong Survey will show population < 2000 C.O.S.M.: 70% 10. 2014 Less Wrong Survey will have > % female than 2013 Less Wrong Survey: 70%
11. 2014 Less Wrong Survey will have < 20% female: 90%
12. HPMoR will conclude in 2014: 80% 13. At least 1 LW post > 100 karma in 2014: 50%
14. No LW post > 100 karma by me: 80%
15. CFAR will continue operating in end of 2014: 90%
16. MIRI will continue operating in end of 2014: 99%
17. MetaMed will continue operating in end of 2014: 80%
18. None of Eliezer, Luke, Anna, or Julia will quit their respective organizations: 60%
19. No one in LW community will become world-famous (let’s say >= Peter Thiel) for anything they accomplish this

year: 80%
20. MIRI will not announce it is actively working on coding a Friendly AI (not just a few bits and pieces thereof) before the end of 2014: 99%

Personal Predictions

1. I will remain at my same job through the end of 2014: 95%
2. I will get a score at >95th percentile for my year on PRITE: 70%
3. I will be involved in at least one published/accepted-to-publish research paper by the end of 2014: 20%
4. I will not break up with any of my current girlfriends through the end of 2014: 50%
5. I will not get any new girlfriends in 2014: 50%
6. I will not be engaged by the end of 2014: 80%
7. I will be living with Ozy by the end of 2014: 80%
8. I will take nootropics on average at least once/week through the second half of 2014: 50%
9. I will not manage to meditate at least 100 days in 2014: 80%
10. I will attend NYC Solstice ritual: 60%
11. I will arrange some kind of Michigan Solstice Ritual: 50%
12. I will not publicly identify as religious (> atheist) by the end of 2014: 95%
13. I will not publicly identify as neoreactionary or conservative (> liberal or libertarian) by the end of 2014: 70%
14. I will not publicly identify as leftist or communist (> liberal or libertarian) by the end of 2014: 80%
15. I will get a Tumblr in 2014: 50%
16. I will not delete/abandon either my Facebook or Twitter accounts: 60%
17. I will have less than 1000 Twitter followers by the end of 2014: 60%

19. When Eliezer sends me a copy of “Perfect Health Diet”, I will not be convinced that it is more correct or useful than the best mainstream nutrition advice (eg Stephen Guyenet’s blog): 70%
20. I will end up being underconfident on these predictions: 50%

Wirehead Gods On Lotus Thrones

One vision of the far future is a “wirehead society”. Our posthuman descendants achieve technological omnipotence, making every activity so easy as to be boring and meaningless.

The pursuit of material goods becomes a waste. A nanofactory or a quick edit to a virtual world can already give you a mansion the size of a planet. Although economic activity may still exist in competition for computing resources, all beings in these competitions will be smart enough to behave perfectly optimally (and therefore in a way that makes even the illusion of free will impossible) and so first-mover advantages will be insurmountable. Economic differences will compound on the sub-second scale until different classes are so far apart that competition becomes impossible.

When sports risk becoming contests of who can enter the higher integer in the $athletic_ability variable of the computer that determines the universe, the World Anti-Doping Agency says everyone must compete using their original human bodies – assuming such things even exist at that point. But neither spectators nor athletes care about the result, since everyone is smart enough to simulate the game in their minds on a molecule-by-molecule basis long before it happens and determine the outcome with perfect accuracy – turning the actual competition into a meaningless formality.

Works of art become gradually less interesting; everyone can extrapolate back from the appearance of a painting to exactly what the mental state of the artist must have been at the time it was painted. Nor does the art enlighten, since the conceptual organization of everyone’s mind is already optimal and the only intellectual differences between entities are insurmoutable ones of available computing resources.

As for Science, everything was discovered long ago. If it hasn’t been, discovering it is a brute-force application of the best-known Bayesian reasoning algorithms.

(And developing better algorithms is also a brute-force application of the best-known algorithm-discovery algorithms.)

Even in the most utopian such world – one where the dominant minds are concerned with maximizing the happiness of everyone else – it sounds pretty boring.

One approach is the imposition of artificial limits. Entities can deliberately refuse to use their full cognitive capacity and so experience uncertainty, choice, and feelings other than that of algorithmically choosing purpose-appropriate algorithms. Maybe some entities will deliberately take on human brains and bodies, and interact with other such entities in a human-level world in order to operate at the level with which their value system is most comfortable. Maybe in order to avoid the temptation to call on their full omnipotence every time they experience a little pain or hardship, they will deliberately “forget” their posthuman status, living regular human lives utterly convinced that they are in fact regular humans.

(I assign moderate probability that this has already happened)

Other entities may have no time for such games. They may cope with the ennui of posthuman existence by reprogramming away their capacity for ennui, with the absence of aesthetic or scientific outlets by programming away their desires for such. Instead, they just reprogram their brains to be deliriously happy all the time no matter what, and spend their time sitting around enjoying this happiness.

The futurist community calls this “wireheading”, after an experiment in which rats had an electrode hooked up to the reward system of their brain which could be stimulated by pressing a lever. The rats frantically tried to stimulate the lever as much as possible in preference to doing anything else including eat or sleep (they eventually died). Stimulating the reward center directly was much more attractive than other activities which might result in some indirect neural reward only after work.

The same pattern occurred in humans, specifically chronic pain patients who had similar wiring installed in their heads in the hopes that it might alleviate their problem:

> At its most frequent, the patient self-stimulated throughout the day, neglecting personal hygiene and family commitments. A chronic ulceration developed at the tip of the finger used to adjust the amplitude dial and she frequently tampered with the device in an effort to increase the stimulation amplitude. At times, she implored her to limit her access to the stimulator, each time demanding its return after a short hiatus. During the past two years, compulsive use has become associated with frequent attacks of anxiety, depersonalization, periods of psychogenic polydipsia and virtually complete inactivity.

It’s unclear to what degree these wires are making the subject so stupendously happy that she desires to maintain her bliss, or whether they’re instilling compulsive behavior.

Likely there are some elements of both – just as in wireheading’s more prosaic younger sister, everyday drug use. But drug use is messy, and wireheading is perfect.

Wireheading is commonly considered an ignoble end for the human race – our posthuman descendants reduced to sitting in dingy rooms, taking never-ending hits of some ultra-super-drug, all their knowledge and power lying fallow except the tiny fraction necessary to retain delivery of the ultra-drug and pump nutrients into their veins.

On the one hand, it probably beats desperately trying to figure out something to do more interesting than setting your $athletic_ability statistic to 3^^^3 and playing sports. On the other, it’s hard not to feel contempt for beings that choose such a pathetic existence.

But I recently realized how unstable my contemptous feelings are. Imagine instead our posthuman descendants taking the form of Buddhas sitting on vast lotus thrones in a state of blissful tranquility. Their minds contain perfect awareness of everything that goes on in the Universe and the reasons why it happens, yet to each happening, from the fall of a sparrow to the self-immolation of a galaxy, they react only with acceptance and equanimity. Suffering and death long since having been optimized away, they have no moral obligation beyond sitting and reflecting on their own perfection, omnipotence, and omniscience – at which they feel boundless joy.

Pictured: ultimate reality

I am pretty okay with this future. This okayness surprises me, because the lotus-god future seems a lot like the wirehead future. All you do is replace the dingy room with a lotus throne, and change your metaphor for their no-doubt indescribably intense feelings from “drug-addled pleasure” to “cosmic bliss”. It seems more like a change in decoration than a change in substance. Should I worry that the valence of a future shifts from “heavily dystopian” to “heavily utopian” with a simple change in decoration?

Based on your findings, which theory about alien thickness seems most valid or most accurate?

Seventh-grade science students with flexible ethics: you’ve come to the right place!

Every so often I look at the search terms that led people to this blog. Most of them are what you would expect, but one of the top search terms, one that keeps showing up again and again and again, is “based on your findings, which theory about alien thickness seems most valid or most accurate?”

I feel like at some point I must have mentioned the words “aliens” and “theory” or “thickness” together by accident, and that started it off. Then at one point I commented on it, just wrote a paragraph on how weird it was that I keep getting these alien thickness people, and since that caused the entire phrase to be on my blog in one piece, it opened the floodgates and now I can’t stop getting curious alien thickness theorists.

So today I finally decided to figure out what was going on, and Google led me a 7th grade science class at Madisonville Junior High School, Los Angeles. As best I can tell, this class’s teacher gives her students a homework assignment that includes various questions on genetics and ecology, most of which make sense.

[EDIT: or it may be a national/statewide curriculum, which only that teacher has put online. That would explain the large number of search terms better than a single class would]

But on question 25, it suddenly jumps to a question about alien thickness which is completely inexplicable by the terms of everything that has come before:

> 25. Based on your findings, which theory about alien thickness seems most valid or most accurate?
> A. Alien thickness is mainly affected by sunlight. The thickness may be used as a way to shield out the sunlight.
> B. Alien thickness is mainly affected by temperature. This may be because aliens become dehydrated (lose water) at higher temperatures and become thinner.
> C. Alien thickness is mainly affected by temperature. The greater thickness at lower temperatures may be used as a way to stay warmer in colder weather.
> D. Alien thickness seems to change at random. It does not seem to be affected

Although there are a few other questions that deal with a changing population of thick and thin aliens, none of them give any information on sunlight, temperature, or hydration status. So I'm not surprised that what, by the numbers, has to be *every single student in her class* decides to Google the question to try to find the answer online.

I looked around myself and eventually found this web gizmo, where you adjust water, temperature, and sunlight to a group of little aliens and it tells you about their changing phenotypes. This has *got* to be the source of the test question, but it's not mentioned on the test and judging by student confusion they don't know about it. Maybe it's mentioned in class one day and then it ends up on a worksheet a month later and nobody remembers it anymore? In any case, dozens of students keep entering it into Google and ending up at my blog.

And this is bad because the blog entries where I mention how weird it is that I'm getting all these alien thickness people *also* mention lots of things that are really really inappropriate for seventh graders, yet likely seventh-graders are finding them through alien-thickness-motivated confusion. So my new plan is to direct them all here, to answer their question, and let them go on their way and maybe finish their homework a little quicker. I don't feel bad about this since any question that gets an entire class trying to cheat *en masse* must be pretty flawed and since homework is mostly bad for kids anyway.

So, seventh-graders! Still with me? Let's talk alien thickness!

Start by going to the gizmo and adjust the different sliders from lowest to highest one-at-a-time, while watching the bar graph measuring alien thickness. You will notice that adjusting the water slider from highest to lowest doesn't change thickness. Likewise, adjusting the sunlight slider from highest to lowest doesn't change thickness. But adjusting the temperature slider while holding the other two constant *does* change thickness. So we conclude that thickness probably depends on temperature.

So now we can eliminate all the answers except B and C, the ones that say that alien thickness is affected by temperature. How do we distinguish between these two?

Well, B says that temperature only affects aliens indirectly, through its effect on dehydration. But if that were true, we would expect preventing the aliens from getting dehydrated to remove the effect of temperature. But this doesn't happen – no matter how high the water slider is, moving the temperature slider still causes the aliens to shift from thick to thin. So the effect of temperature doesn't depend on hydration.

Armed with this knowledge it should be pretty simple to pick the correct answer through process of elimination.

Let's move on to question 26:

> 26. Based on the data you found, about how many of the 100 aliens would become thin if the temperature were 35°C?
> A. fewer than 10
> B. about 50
> C. about 80
> D. more than 90

You notice that at temperature 20 degrees, about fifty aliens are thin. At 25, about seventy aliens are thin. And at 30, about eighty-eight aliens are thin. The take-home point is that the higher the temperature, the more aliens we expect to be thin. So at 35 degrees, we would expect more aliens to be thin than the eighty-eight who are thin at 30 degrees. Which option best reflects that expectation?

So there's your answer. But there's a more important meta-point here. Your teacher wouldn't include a nonsensical question on the worksheet, so clearly in 26 she expects you to be able to calculate alien thickness based on temperature. So just by reading 26, you know the answer to 25 is one that says thickness is based on temperature. So you can eliminate A and D and be left with a 50% chance of getting it right. And without looking at the original data, you can conclude that it's probably not B, since it says only temperature matters but the explanation implies that water and hydration status matter as well. So really, even if your teacher forgot to link you to the gizmo thing, you should be able to guess the right answer based on test-taking skills alone.

A story from my own life – my first month of medical residency, my schedule was extremely disorganized and I ended up starting a class they day they were having their final exam. This exam happened to be on the treatment of radioactivity-related injuries, a field of medicine I was unaware existed until that moment. Because of inconsistent answers, clues in other questions, and basic common sense, I was able to guess well and ended up getting a B- (the class average was a C).

My point is, test-taking skills matter.

I haven’t gotten any Google search queries asking about any of the other questions, so I’m going to assume you’ve got all of those down. Good job, seventh-grade science students!

More Search Terms That Have Led People To This Blog

[Trigger warning: profanity, slurs, incest, rape jokes, racism, and other unfiltered access to the consciousness of the Internet]

The Anti-Reactionary FAQ is by far the most popular entry on this blog, with a total of 28,000 page views. Many of those views came from links by bloggers and news outlets, but others came from random netizens who stumbled across it during Google searches. WordPress records the terms they used to get there, which gives us a little insight into their minds.

Were these searchers liberals looking for help debating Reactionaries? Were they Reactionaries dutifully searching for counterarguments against their ideas? Curious political science students lured in by the wide-ranging discussion of social trends?

Actually, they were none of these things. They were porn viewers who wanted videos of people fucking aunts, who had misspelled “aunty” as “anti” and misspelled “fuck” as “faq”.

I wish I was kidding, but read the search terms: “pak anti hot sex with monki and dog”, “dog indin boyfriend sax faq”, “arabis faking indian job antis porn”, “anti best figer sexi image”, “50 yars anti indyan faking fat anti”, “hot sexy bobz pakistani anti u tub”, “ancal anti bobs kising you teub”, “hot indan sexi anti romance in doods”, “ankal and anti se saleeping with baby full sexy viduos on youtube”, “sought indian sexy anti thims”, “pakistani hot and sexey figar yung anti pitcher”, “anty masive nippals”. And on and on it goes.

I couldn’t figure out what was going on until finally, like an old-timey archaeologist decoding Linear B inscriptions, I realized that the word “ankal” (or sometimes “ancal”) appeared beside “anti” in a disproportionate number of these phrases and so they were probably talking about aunts and uncles. Then I noticed a lot of them mentioned Indians or Pakistanis, and found from Urban Dictionary that in Indian culture, “aunty” is a general term for a middle-aged woman. So I think in Indian/Pakistani porn, “aunty” is their equivalent of “milf”.

I’d previously read several accusations that Pakistan, despite its austere women-in-burkas-at-all-times image, leads the world in most pornographic searches. While not actively disbelieving the data, I felt like a lot of this was media outlets being too quick to play up the man-bites-dog-esque “very Islamic purity culture is actually obsessed with porn!” angle. But after monitoring how many pornographic searches my blog gets from Pakistan, I no longer doubt.

(anyway, now the man-bites-dog segment of the media has moved on to the story that very Islamic purity culture Pakistan also leads the world in searches for *gay* porn)

A while back we conducted a little experiment to see if blog readers knew how the average person thought. Athrelon won and attributed his success to his work as a medical student, where he got to see a broad cross-section of the population instead of the People Like Us In Our Little Bubble whom the rest of us interact with.

If you can’t get into medical school, looking at Google search terms seems like your next-best option. It’s a good way of reminding yourself that most people in the world are neither liberals nor conservatives, but just people looking for some hot aunty porn. So without further ado, here is the second installment of Slate Star Codex’s quasi-annual feature, Search Terms That Have Led People To This Blog.

Search Terms About Being In Love With Your Sister

I made *one little post* about the phenomenon of genetic sexual attraction, and now I am cursed with a never-ending stream of Google searchers who want my advice about incest.

I already mentioned last year how I got queries like: “siblings attracted to each other”, “my sis is so pretty”, “sweet sister so pretty”, “sister can you not so pretty”, and the scientifically intriguing “are attractive siblings more prone to incest”.

In the last eight months, I can add to that collection: “attract with sister”, “how to attract sister”, “how to attract my sister”, and “how to atract sisters”, which is my favorite both for the spelling error and for the implication that he just wants sisters in general, not necessarily his sister – any girl with a living sibling is good enough for him.

But there’s also “my sister is pretty”, “is my sister prettier”, and “my sister is so much prettier” (hey, man, it’s not a contest). Others are more interrogative, asking “why my sister pretty”, “why i am sexualy attrect our sister”, and the very philosophical “why is my sister”.

Also, one person who got here with “are people attracted to siblings”. I would say that the above answers the question pretty thoroughly, except for one thing – I haven’t gotten anyone claiming to be attracted to their *brother*. I don’t know if this is just because the original post used the example of sisters (the exact title was “If Opposites Attract, Why Is My Sister So Pretty?”) or because brothers find sisters attractive but not vice versa. So, for science, I’m going to type the following phrase and let Google crawl it and see what happens:

IF OPPOSITES ATTRACT, WHY IS MY BROTHER SO PRETTY?

There. Anyhow, not *everybody* was asking Google about how much they loved their sisters. I also got “why is my sister a bitch”

Search Terms About Not Being Entirely On Board With Polyamory

I wrote an essay about my experiences with polyamory, and in the comments Spandrell of Bloody Shovel responded with a theory that only ugly people were able to remain polyamorous successfully. In one of my most memorable experiences as a blogger, a commenter then responded with a picture of his fashion model girlfriend, leading Spandrell to admit that “That’s a big piece of……data.”

Anyway, Spandrell’s opinion seems pretty widely shared, since that post has then gone on to get search terms like: “polyamorists are ugly”, “polyamory people are ugly”, “why are polyamorists gross and nerdy”, “why are polyamorous people ugly”, “are all polyamorous people ugly”.

But other Google searchers have entirely different reasons for hating polyamory! Like “polyamory aspies” or “poly or just a whore”.

And finally, some people just keep it simple: “i hate poly people”.

Search Terms That Imply A Heartbreaking Story

“if a woman pleads insanity can she have her kids back” “can i sue the hospital for oversedating my husband with pain medication and causing his death”

“my grandfather’s leg is starting to rot will he die?”

“is it possible for someone with vascular dementia to be upset with you, yell at you, and then start to be distant?”

“my hearts aches when i think of my mistakes”

“can you condition yourseld to like sex”

“id didn’t match for residency but it turned out good”

“wat can i do to die slowly”

“what happenes to those whoreare addicted to sucking things”

Search Terms By Time Travelers

You may have heard about Nemiroff and Wilson (2013), the study where scientists at Michigan Tech did an automated examination of all search queries looking for things that were searched or discussed before people could possibly have known about them (ie Comet ISON before it was discovered; Pope Francis before he was elected) in order to catch time travelers. Alas, they didn’t find anything suspicious.

On the other hand, someone got to my blog by the search term “survive march 4 2014”, which until I read that query was not something I was worrying about.

Less scary but still time-bending: “scott alexander gets married”. I can just see someone getting increasingly exasperated that my wedding photos aren’t online yet and then slapping their forehead when they remember they have to wait until August 2017.

Time travelers from the past would theoretically be harder to detect, but some of them aren't even *trying*: "trial of king leopold ii of belgium next week".

People Who Typed A Political Manifesto Into Google For Some Reason

"pity the people who idealize the dictators in pakistan"

"sexuality is learned and reinforced like all behaviour. eg. baby rapist are born raping babies, they learn their sexual orientation from by being raped or abused as infants"

"hire a competent white person to do the work of a few minorities that cannot do the job"

"isn't using birth control better than having unwanted teen pregnancies and abortion?"

"the rothchilds are not good enough to feed to swine may thay fester burn rot and stnk of decay in a sewer in hell full of excreament"

I think we all have our days when we enter long screeds against the Rothschilds into Google search boxes, but some of us are definitely more creative than others.

Search Terms Obviously By Students Trying To Cheat On Their Homework

"some people were trapped during world war ii. one man used elephants to get them to safety. many people feel that the true heroes of a war are those fighting. what do you think?"

“write a story how you and your friend fell out. alternatively you can write about someone else’s quarrel. which of the friends was right / wrong?”

“demonstrate how federal laws are impacted by the times, political influences and even changes in social norms?”

“evaluate shelley’s philosoophy in adonais”

“explain using your own words forms of government under adolf hitler of germany and louis xiv in france” (I like the thought process here: “Hm, it says explain using my own words. Better include that in the search query to make sure I only get pages that are written in my own words.”)

“if you have something white point 075 pounds how many pounds is that be” (I don’t understand this question *at all*)

“imagine yourself as an aztec nobleman, describe to me your way of life and that of your family,court,slaves,finances,etc?”

“what are the status of feudal monarchies, specifically the actions of king henry iv and the intended and unintended consequences of those actions”

But as always, the granddaddy of the Students Trying To Cheat On Homework Questions is the alien question. I have spun that off into its own post to avoid attracting seventh-graders here and making them read about antis with masive nippals.

People Interested In Albino Black People

This is second only to the alien question people in terms of “I mention how weird it is that something gets so many search hits, and now all those search hits come to me”. On a post about using Google Correlate to play with data, I checked to see what search term was most disproportionately entered by white people – that is, what search term was most likely to be entered by a white person as opposed to a person of color – and the answer was “black albino”.

Well, after that, the readership of Slate Star Codex got just a little bit more white.

“black people turned white”

“albino people”

“albino black people”

“black and white albinos”

And, of course, since this *is* the Internet: “albino booty”

People With Unexpectedly Detailed Questions About Reptilians

My post on Noisy Poll Results And Reptilian Muslim Climatologists From Mars was meant to use an error in a poll asking about belief in “reptilian aliens” to make a general statistical point. Instead, it made me people’s go-to source for statistics about reptilians. For example:

“percent reptilians on earth”

“are all reptilians muslim”

People Looking For My Commenters

“ozy slate star codex”

“eliezer yudkowsky polyamory”

“leah libresco slatestarcodex”

“scott alexander leah libresco”

“unequally yokes neo reactionaries”

“”multiheaded” marxist”

“madeleine ball fda”

“finnish surname sotala”

“gwern abstract reasoning”

“andrew rettek of metamed”

But the number one most popular commenter? “deiseach”, with 19 searches.

People Searching For Information About Rape Culture

As you may remember, I wrote a post detailing my problems with this term. And surprise surprise, even though I can’t get people to learn about statistics or medicine or transhumanism or anything interesting even if I drag them kicking and screaming, the level of fascination with rape culture remains at an all time high. We have “rape culture”, “is there really rape culture”, “what would women do different rape culture”, “rape culture philosophy”, and “bloody rape culture”, which I like to think of as being entered by an exasperated British feminist.

Others seem to lack a certain understanding that rape culture is a problem, like “why is rape culture bad” and “arguments for rape culture”.

Others are less scrutable, like “stars rape culture”. I suppose it was only a matter of time before someone noticed that

that astronomical bodies are totally failing to speak out against the patriarchy and so are clearly misogynist. There’s also “modalrape dot com”, which is either a misspelling of “model rape” or else is referring to rape which necessarily occurs in all possible worlds. I hope some of the feminists in philosophy are doing something about this.

People Asking About Qwubbles

“qwubble”

“what is a qwubble”

People Searching For A Number of Stalins > 1

“two stalins”

“fifty stalins”

Miscellaneous Yet Delightful

“complications caused by eating slate sticks”

“genetics test 31 and me” (finally, genetic testing for donkeys)

“how much is 50 dollars worth of weed in central va” (this is, alas, not the kind of marijuana statistic this blog specializes in)

“sex russia ass big 2013” (I feel sorry for the people who are still stuck on last year’s sex Russia big ass)

“mentally and magically reasoning with star systems of other galaxies and treat them as friends” (is this about acausal trade?)

“i can’t take anyone who uses the term ‘dudebro’ seriously” (My new friend! You are not alone!)

“transhuman pig”

“daniel boon whos making this shit up” (haven’t you heard? It’s all the fault of the Cathedral!)

“inefficient person” (darnit Google, just because I keep writing blog posts instead of doing work doesn’t mean you have to direct this search to me!)

“who is the best pun star in the world” (this, on the other hand, you can direct to my blog any time)

“dyspeptic ploughman” (I like picturing the Google ads for this query: “Top site on dyspeptic ploughman!” “Your source for dyspeptic ploughman!” “Buy dyspeptic ploughman now!” Actually, that works for all of these.)

“can beating yourself up help with conditioning” (Eliezer, someone has just proven you right)

“impregnate me sexual roulette” (Please don’t do this)

“mencius moldbug n*gger” (Whatever my case was, I hereby rest it)

“how to make a star very slowly” (First, get hydrogen. Then, wait.)

“doctor killed patient with fake euthansia” (…then it wasn’t fake, was it?)

And My Personal Favorite

“how to seem virtuous without actually being macintyre”

Seeming and Being Empathetic

"The most important thing is sincerity. Once you can fake that, you've got it made."
– **George Burns**

The Empathy Exams is a good piece about medical actors. You should read it.

Most of the communication skills classes I took in medical school eschewed outside actors in favor of role-plays among different medical students – where one of us was the doctor and the other the patient. I can't say we took these entirely seriously.

I remember one skit, performed in front of the entire class, where a female classmate (playing the doctor) was supposed to present me (the patient) with my test results. She chickened out at the last second and asked if we could switch places, which we did.

So I introduced myself as Dr. Alexander, she introduced herself as Mrs. Murphy, and we briefly talked about her imaginary family and how her imaginary husband was doing. Then I checked my script – written by the professor when he was still expecting a male patient – and found I was supposed to be diagnosing prostate cancer. So I shrugged and thought "what the hell" and diagnosed her with prostate cancer. She, consummate actress that she was, nearly jumped out of her chair: "I can't have prostate cancer!" Totally deadpan, I answered "That's what everyone always says – it can't possibly happen to *me*."

Needless to say, we didn't get much empathy-teaching done that session. Other popular failure modes for our role-plays included actors being silly or overdramatic (“NO! YOU HAVE RUINED MY LIFE WITH YOUR STUPID DIAGNOSIS! I WILL SUE YOU FOR EVERYTHING YOU ARE WORTH!”), deliberately obtuse (“Severe heart failure? Does that mean I can’t go through with my plan to run a double marathon up and down Mt. Everest?”), or just giggling the whole session about our classmates being asked to play husband and wife couples (yes, 25 – 30 year old medical students giggle about pretend marriages exactly the same as schoolchildren).

But the few times we had professional medical actors – usually during exams – were Serious Business. I especially remember my USMLE2CS, a big licensing exam. The Kaplan test prep book (of course there are Kaplan test prep books that teach you empathy) gave a variety of somewhat cargo-cultish seeming advice, like suggestions to use ‘transition sentences’: “Now I am going to wash my hands”, or “Now I am going to ask you questions about your smoking habits”. The general feeling among us students was that the medical actors were dangerous beasts for whom any deviation from script, anything other than the preprogrammed responses, would produce fits of emotion followed by examination failure.

Ms. Jamison’s article confirms what I suspected: that the actors and actresses also feel uncomfortable and stilted by their fixed lines.

So we get this very interesting situation where both sides are trying to stick to a script of awkward preprogrammed responses while avoiding all real human emotion. And when they succeed, we declare victory and say we have taught how to connect to people.

II.

The thing is, I know why this happens. It makes perfect sense from the inside.

Most people have a story of some doctor who treated them brusquely or unkindly. If you don’t have one yourself, you probably hear about it in the media. It’s a legitimate problem. Eventually someone tells medical schools “Hey, a lot of doctors are jerks. Fix this.”

It is possible that a parent, getting the chance to raise a child from infancy and spend nearly every waking moment with them, might be able to rescue someone otherwise fated to become a jerk. Might be able to teach them to read other people’s feelings, to connect with them, and to express that connection in a socially acceptable way.

A professor who gets six one-hour sessions with a class of two hundred? Not likely.

So since they can’t teach actual empathy, they teach programmed responses. The patient is in pain? You say “I see you look like you’re in pain. That must be really hard.” Patient is depressed? You say “I see you look like you’re depressed. That must be really hard.” Patient is bleeding from a huge gaping wound in her abdomen? You say “I see you look like you’re bleeding from a huge gaping wound in your abdomen. That must be really hard.” St. Francis of Assisi it is not, but the point is that it’s grade-able by rubric. When the actor clutched her abdomen and screamed, did the student say “I see you look like you’re in pain?” A plus plus! Did they say something else? Time for remedial training.

I recently learned that standardized tests are written by a team consisting of psychologists and lawyers. The psychologists are there to choose questions that test the appropriate skills. The lawyers are there to make sure all the questions will stand up in court when someone sues the standardized test company because their kid got a bad score.

That same article explains the effect this has on what sort of questions are allowed. You can’t ask “What was the moral of this story?” or “What is the symbolism here?” because if the issue was brought before a judge, and counsel for the prosecution says that the story symbolized the ennui of modern life, and counsel for the defense says that the story symbolized the sorrow of mortality, no one can really prove their case beyond just “Well, that was how it seemed to me, c’mon, look at the text!” So all questions must be *obviously* based on things in the text: you can only ask if a story was about ennui if it the sentence “This story is about ennui” is somewhere in the story.

I imagine a counterfactual world where doctors are *not* taught empathy through programmed scripts, and *not* graded by rubric. A medical student meets a medical actor, and the student tells the actor she has cancer, and there is some sincere discussion afterwards, one human being to another. And then the grader decides if the student was properly empathetic, if she would genuinely want him as her doctor…

…and then if the student fails, he accuses the examiner of racism, and says she ruined his promising medical career. And the examiner doesn’t have a leg to stand on, because she just didn’t *feel* the student was empathetic enough, and the examiner’s *feelings* aren’t going to be any kind of a defense.

(if you think this is a purely hypothetical concern, think again)

But a rubric that says to give plus one point if the student says "I see you look like you're in pain", minus one point otherwise – *that* will provide a defense.

And then if doctors are still jerks, the medical school can just shrug and say "Well, we taught them empathy! They even passed our exams! See! A plus plus! It's not our fault if they don't use their training!"

III.

But maybe it's good preparation. Because actual caring is not enough. You also have to pretend to care.

I remember, one of my first few months of internship, listening to a patient – not a PTSD patient or anything, just someone presenting with something totally different like bipolar disorder or drug addiction – explain the brutal abuse he suffered as a child. And the whole time, I was thinking "Oh god oh god this is the worst thing I've ever heard I want to go home and cry."

And then he finished his story and I had to *say* something. And I didn't want to say "Oh god oh god this is the worst thing I've ever heard I want to go home and cry", because I was supposed to be Competent Medical Professional, and Competent Medical Professionals don't go home and cry every time they hear a sad story.

And I also didn't want to say "Okay, I'll tick 'yes' to the child abuse checkbox; moving on to the next item, have you ever smoked marijuana?"

And I also didn't want to say "I'm sorry", because various Head Honchos in the hospital have launched a crusade against the word 'sorry' because it sounds like an admission of fault.

And I also didn’t want to say “I see you look upset about being brutally abused as a child. That must be very hard for you.” Because then I would have shown up on his Tumblr the next day.

(I see these stories all the time. “I really opened my heart to this doctor, told him every last detail of the brutal abuse I suffered as a child, and he just sat there and said ‘You look very upset’. YOU F@#KING THINK SO? You think I’m upset about being beaten every time my father was drunk, beaten so bad I was afraid I’d broken bones? I’m sure glad I spent however much money to talk to you so you could tell me I looked upset! Because I need a privileged abled white guy to judge my opinions as valid, otherwise they don’t count, right? Man, doctors are all the same, they just respond to your pain with a stock phrase because they don’t think anyone with an illness can really be human.” Doctors read these, and we’re *sorry*, but empathy is *hard*.)

There was a sense in which I had already transgressed; I had forced him to bring up this event from his life because “was patient abused as a child?” was a box on my Medical History Taking Rubric. If he had brought it up for some reason – to get my help in arresting the abuser, to ask for charity, even to seek psychotherapy – we could have pursued that reason and it would have defused the moment. Instead it got brought up for no reason that I had to ask, and once asked, he had to tell me. It floated there in the air, a brute fact. The fact that I actually cared quite a lot didn’t make the socially necessary ritual of Pretending To Care any easier.

So I said: “Gaaaaaaaaaah!”

This may, in retrospect, not have been the most appropriate comment. My only excuse was that all these complicated thoughts about what I did and didn't want to say were mostly after the fact, and at the time, I just heard this horrible story about child abuse, and my gut reaction was "Gaaaaaaaaaah!". And my brain's filter, which is usually pretty good, failed to catch me in time to do anything but give my gut reaction.

And my patient was mildly startled, and I quickly deflected the encounter through the deft manuever of moving on to ask if he had ever smoked marijuana.

IV.

Jamison writes:

> Empathy isn't just something that happens to us—a meteor shower of synapses firing across the brain—it's also a choice we make: to pay attention, to extend ourselves. It's made of exertion, that dowdier cousin of impulse. Sometimes we care for another because we know we should, or because it's asked for, but this doesn't make our caring hollow. The act of choosing simply means we've committed ourselves to a set of behaviors greater than the sum of our individual inclinations: I will listen to his sadness, even when I'm deep in my own. To say "going through the motions"—this isn't reduction so much as acknowledgment of the effort—the labor, the motions, the dance—of getting inside another person's state of heart or mind.

Partners, parents, siblings, children of a sick person – they probably need to exert themselves to cultivate empathy. They're going to have to put up with a sick person 24/7, long after their initial burst of compassion has worn off.

Nurses probably need to exert themselves to cultivate empathy as well. They have to deal with sick patients through every moment of their hospital stay, respond to their pleas, deal with their sometimes disgusting bodily issues.

But as a new resident, I have it easy. In my interactions with patients, empathy is not a scarce resource, doesn’t require exertion or extension. My interactions with patients are short and emotionally fraught – getting someone with excruciating abdominal pain to lie still long enough to examine them, telling someone they have a very serious disease, having to hear the details of a history of child abuse. It’s not hard to remain empathetic for a twenty minute appointment after telling someone they have cancer, when you know you won’t be called on to do anything more strenuous than prescribe a medication. Heck, it would be hard *not* to feel empathy in that situation.

What requires exertion is channeling the impulse. Neither repressing it entirely to avoid awkwardness, nor just letting it all out in an unfiltered “Gaaaaaaah!”

I am getting better at this. One of my mentors taught me the important technique of having a tissue box near me at all times. If someone gets into an emotional situation, I unobtrusively place the tissue box closer to them, which signals that I suspect they’re upset and I’m okay with it, without bludgeoning them over the head with the fact. Sometimes questions work: “Are you okay?”, “Is there anything I can do to help?”, “Do you want to talk about this more, or do you want to move on?”

And part of what I had to do was unlearn my habits from communication classes and empathy exams. In the exams your goal is always very virtue-ethics-y: to demonstrate that you are The Kind Of Doctor Who Feels Empathy. In real life, your goal is consequentialist: there’s a person in pain in front of you, and you need to figure out how to help them. In what I think is C. S. Lewis’ phrase, you need to *get out of your own head* and do what’s best for the patient. Which sometimes involves reference to the content of my own head – all psychiatrists know that the therapeutic relationship is one of the most powerful weapons in medicine – but only if the patient cares what’s in there.

Which they very often don’t. Now if I get a patient like the one who told the child abuse story, I’ll be more likely to just put on my best Concerned Face and ask something like “How are you doing with that now?” And a surprising amount of the time, the patient will say “I’ve put that behind me, it’s not really an issue anymore.”

I don’t know if, on the deepest level, that’s true. If I were doing psychoanalysis, I’d want to pick and prod at that claim. But in the context of a medical history? I push the tissue box toward them in case they need it, wait a couple of seconds to see if anything else is coming out, and then say “Okay, and can you tell me if you ever smoked marijuana?”

Blogging The Anissimov – Smith Reaction Debate

I was really excited to hear about the video debate between Noah Smith and Michael Anissimov on neoreaction.

Noah is an economics professor and author of the fantastically punnishly named blog Noah Pinion. He's also a pretty liberal guy who lists Paul Krugman as one of his heroes. Michael is a Bay Area futurist (possibly not his official job title?) and author of the moderately punnishly named blog More Right. He's also a sort of spokesperson for neoreaction and someone I've argued with a lot in the past. Both are very smart guys.

I was looking forward to the debate because real engagement and discussion between reactionaries and nonreactionaries is hard to find. Even my FAQ, which I hoped would start that kind of engagement, got a couple of brief rebuttals, which I tried to briefly counter-rebut, and then the whole thing died down (this was partially my fault, but I was trying to save my comments threads from becoming too hostile and heated).

The debate didn't really live up to my expectations, though. Both sides were just too nice and reasonable. My impression was that Michael made weaker claims than usual, presenting his ideas as hypotheses or as a basis for future discussion rather than as assertions, and Noah was pretty happy to agree they were interesting hypotheses and that the future discussion should take place. But the debate that did take place was on a pretty high intellectual level and I feel like I learned some stuff.

First the video, and then my thoughts:

Michael began by criticizing modern society’s social isolation, which he blamed for things like rising depression, rising suicide rates, and poor civic engagement. However, he was very willing to admit that some (most? all?) of this was because of big technological changes like urbanization and industrialization, which in my opinion didn’t seem to leave much room for him to make any political point with it. Noah agreed that social isolation was bad and probably due to technology, and suggested it would be nice if technology somehow managed to solve social isolation. Michael agreed that if technology managed to solve social isolation, that would indeed be nice.

Noah then made much the same point I did in my FAQ – that happiness has generally increased in most countries in the past fifty years. Michael agreed this was true of the developing world but said it was ambiguous/no change in the developed (I don’t have a strong opinion on this). He said that he realized it might not work to go back to the time of monarchies and aristocrats, but that even society fifty years ago seemed a lot more stable and connected. I was confused by this part because it seemed to be cavalierly dismissing everything he has ever said up until now – he previously was pretty set on monarchies and aristocrats.

Around 10:47 they realize they don’t actually disagree on this issue and move on. Michael tries to milk one of his hobby-horses (sorry for mixed metaphor), and talks about how the phrase “all men are created equal” is obviously balderdash since some people are better than others at practically everything. Noah makes exactly the response everyone makes when confronted with this claim – that the founders obviously didn’t mean something completely stupid and so we should assume they meant something more like “everyone is equal under the law”. Michael agreed the founders weren't completely stupid, but says that this latter reasonable meaning gets misinterpreted by everyone into the stupid meaning. To which all I can say is that it doesn't really seem to me like lots of people are making this misinterpretation.

I would also point out that this is a case where we should read history of philosophy backwards. The Founders lived in a time when it was still considered reasonable to think that certain people were natural slaves who didn't deserve legal rights – I'm not even just talking about blacks here, there were *Irish* slaves in America only a couple generations before the Revolution. In a lot of countries nobles were bound by different laws, faced different criminal penalties, paid lighter taxes, and were reserved most of the high government positions. So "all men are created equal" had an obvious and natural meaning to the Founders – "You know that stuff that everyone else is doing? We're not going to do that."

The debate then briefly degenerates into both of them debating the vague term "equality", with Noah having warm fuzzy feelings about it and Michael having deep-seated anger towards it, without it really touching on concrete examples of policy proposals at any point.

Around 14:05 they move into a discussion of capitalism. Both are broadly pro- free market, but Michael thinks capitalism is insufficient because "it accords status based on who is the best merchant". Noah requests that Michael differentiate better between "status" and "money", which is a heck of a good point. Michael points out that money is a really important component of status in our society, which is also a pretty good point.

If I'd been involved I would have asked Michael how *else* one accords status. Like he says it should be for "service to the community", which is very nice in principle, but how exactly does one cash that out? Old aristocracies had status depend on accident of birth, which seems even worse than effective merchant-ing. Besides that the only place I can think of for him to go is some sort of proposal like "the President is allowed to give people large cash rewards along with the Medal of Freedom", which is pretty interesting as far as proposals go but not really neoreactionary.

Around 23:45, Noah says that modern non-traditional societies seem to be richer and more successful in pretty much every way. I was waiting for Michael to make what I think is the proper response – that the causation probably goes the opposite way and that wealth causes modernization and collapse of traditions. Instead, he argues that America had an unfair advantage because it had lots of land and natural resources and won World War II. I started yelling at my computer screen telling Noah to respond "Russia also has lots of land and natural resources and won World War II, but they did very poorly because of their dictatorial social structure", and to my delight he complied.

Michael then said that democracies inevitably degenerate into tyranny. Once again I started yelling at my computer telling Noah to argue that historically democracies practically never degenerate into tyranny but monarchies seem to degenerate into tyranny quite a lot. And as an economics professor, I was hoping he would point out that the de Tocqueville "democracies spend themselves into bankruptcy" argument is wrong and that federal spending isn't as bad as everyone thinks and government interest rates are impressively low. Noah says some stuff sort of to this effect, and then says changes the topic, saying "let's mosy on to monarchy", which I hope becomes a reactionary slogan.

Noah starts this off with a very ill-phrased claim about technological stability by saying (31:30) that "people in 1700 AD lived in about the same way as people in 1700 BC", which makes me start worrying about the credibility of everything else he's saying. I will excuse it as an on-the-spot attempt at a flourish of speech. But he goes on to say how technology has changed a lot since then, monarchies have mostly gone away, and "the only kinglike system I've seen instituted recently is North Korea". Michael counters that Romania is considering restoring a monarchy, saying it will be a real monarchy and not just a figurehead, although his blog post sounds less sure. Some discussion on Romanian monarchy and how Romanians like their (currently non-governing) king more than their democratic government.

I would have added here that of *course* Romanians like their king, but this is only because the king serves as an apolitical focus for nice fuzzy feelings, the same way most Brits like Queen Elizabeth. So it's unfair of him to say that the higher approval ratings for the king than the prime minister represent a good argument for monarchy. If the king had to do what the prime minister did – govern – people would find fault with him soon enough.

At about 34:45 Michael changes direction, and says that his main point in pushing monarchy is the "organic state". I keep getting confused by this term because when I hear "organic" I think of organic farming – you know, leaving everything alone and letting it develop of its own accord, kinda a Taoist sort of thing. But Michael, apparently drawing from Evola, means (AFAICT) almost the opposite – a state in which everyone is united towards a common purpose. I don’t know how fully I understand this idea – for example, it would seem this would make Mao’s Great Leap Forward and Cultural Revolution the organic state extraordinaire, since every citizen had to be totally devoted to the declared goal of the state – but these don’t seem like the sort of thing reactionaries usually support. This debate did nothing to allay my confusion.

Michael here makes probably the most memorable quote of the debate – that to him, America doesn’t even seem like a country in the organized-nation-state sense, it just seems like a place where a bunch of people are living next to each other. This doesn’t seem healthy if you want to stay stable and work on complicated projects together.

He makes a good point, but I would argue that America *is* united on certain uniquely American values. Those values happen to be freedom, democracy, and all men being created equal. Like, patriotism seems to be a *really* big thing here, and it tends to be expressed in terms of “We are really free! Freedom is great!” and then feeling smugly superior to people who are not free. When people express dislike for these values, we don’t burn them at the stake or anything, but we do tend to tut-tut at them pretty loudly.

And these are probably the only values that America stably *can* hold. The idea of being a united monoethnic nation-state missed the boat when the original English colonists started letting in German immigrants in the 1700s, let alone anyone else. Given that people disagree about nearly everything except the freedom-justice-equality complex, it’s hard to imagine what else our organic-state-goal could be without kicking out or silencing a pretty substantial portion of the population.

Noah didn’t bring this up, but he did say he thought liberal diverse countries were more stable than “organic” countries, which was the opposite of Michael’s claim. I think the gist of his argument was that if you have a unifying vision for the state, then you have to fight a lot over what that vision should be and people who don’t agree with it get upset, whereas liberalism is pretty live-and-let-live. Seems broadly correct to me.

I think a point that both miss is that organicity is an effect rather than a cause. If everyone in your country is the same ethnicity and agrees about everything, life is easy; but liberalism seems like a technology developed to deal with the tougher situations where this isn’t true. Compare to a philosophy of marriage that couples stick together if they always agree about everything, versus a philosophy that couples stick together if they learn how to set boundaries and fairly resolve disputes. The first kind of marriage is a good deal if you can get it, but the second seems like more of a real-life solution.

Noah points out that empirically, liberal societies are more stable and successful than organic ones; Michael counters with China getting rich; Noah counters with everyone in Asia got rich, China was actually last to do so, and this seems independent of government. Michael (43:00) says empiricism isn’t the proper tool for this kind of thing and we should rely on theory instead. This claim *always* sets off flashing red lights for me, but your experience may differ.

Around 44:10, the discussion moves to aristocracy. Noah expresses what I was wondering too – given that America has really rich people who tend to pass their wealth on to their kids, what’s the difference between an aristocracy and what we already have. Michael says real aristocrats would give back to the community, for example by sponsoring festivals. To me this doesn't answer the question of how we would transform the Rockefellers, Kennedys, etc into "real aristocrats". Flashy red robes? A title? Giving each of the fifty richest families them absolute power over one state? Without something more specific I can't have opinions on it one way or the other.

Noah doesn't push against this, and talks about how aristocrats in Saudi Arabia and elsewhere mostly just hold huge parties and become really corrupt and don't give back to the community at all. Michael agrees this sometimes happens, but says aristocracy is high-variance, which is a good point – and I am usually pretty in favor of moving to higher-variance things because of the value of experimentation and evolution. But he doesn't address Noah's broader point which is that as far as I know there aren't *any* awesome aristocrats in Saudi Arabia to serve as the far right tail of the high-variance bell-curve and this seems to be a fact that requires explanation.

I remember once reading an article that noted America has a lot of super-awesome give-back-to-the-community super-right-bell-curve rich people: the ones who come to mind are Elon Musk, Bill Gates, Warren Buffett, and Peter Thiel. But that these people, as a class, are almost always first-generation self-made entrepreneurs. The actual aristocracy – the people who have had billions of dollars for generation – mostly spend their time playing polo, buying very expensive status symbols, and occasionally becoming Senator or something. The article conjectured that being born to wealth isolates you from reality enough that you lose your ability to be awesome and get stuck in the aristocratic culture of endless status games. An alternate explanation would be that you need a pretty high IQ to be as awesome as Elon Musk, and tech entrepreneurs mostly have it and fifth-generation old rich don't (even if the first aristocrats had high IQ, regression to the mean ensures that barring outright eugenics the fifth generation won't). So I predict that having a hereditary aristocracy would actually be *worse* than our current method of just having very rich people who are occasionally super-awesome. I can certainly think of fascinating non-hereditary aristocracy proposals that might work, but this gets into one of my general points that once we make neoreaction interesting or useful, we're building novel utopias and not trying to return to traditional forms at all.

Michael continues on his variance point and says that monarchies have greater variance than democracies. Noah agrees but says variance is often because something goes wrong and does not imply *anyone* has a higher mean. This seems like an important point to me. For example, hospital patients have much greater variance in their level of health than people not in hospitals, but this does not imply that some long tail of hospital patients have better health than the non-hospitalized.

At 53:30, we move on to gender roles. Michael says the collapse of the family has caused profound damage to society, and that the best way to fix this is to reinstitute traditional gender roles.

Noah says we have to be really careful cashing out "collapse of the family", but he doesn't push it and this doesn't really happen.

Noah says it would suck to make women stay in the kitchen all the time, and uses the example of his sister, who is apparently a very skilled lawyer (and, he notes, can't cook). How would it be remotely effective to say she can't practice law?

Michael says maybe she should practice law until she has kids and then stay home to take care of the kids. He agrees it is sad she can’t practice law more, but that it’s a tradeoff and it’s socially more important to avoid broken families than to maximize female earning power.

Noah asks “If two parents live together and split work responsibility and family responsibility, how is that a broken family?”

Michael then freezes, going completely silent for like two minutes. I was moderately hopeful that he was having some spectacular crisis of faith, that Noah’s comment had given him such an epiphany that he could only sit, awestruck, as he re-evaluated everything he thought he believed in. Unfortunately, it turned out that it was just computer trouble. After two minutes, he pops in again, saying that although this may work for the upper-classes, it is too complicated a solution for most people, and that “well-educated people overestimate the degree to which we can mess with traditional gender roles.”

I understand what Michael was trying to get at – it was exactly the point Ross Douthat made in his Social Liberalism As Class Warfare article I linked to this weekend. But I don’t think Michael ever really got there. And so I never got to hear Noah address it, which is too bad as he’s a smart guy and I’d be curious what he thought of it.

Then there was more computer trouble and everyone just decided to call it a night.

Overall I was impressed with the politeness and moderation of both debaters, but I felt that (maybe as a result of that) not a lot got said. If I had been Noah, I would have tried to force Michael’s broad value statements into specific policy proposals. If I’d been Michael, I would have focused more on specific failures of modernity – even though he’s had some excellent Facebook posts on rising crime rates recently that have mostly changed my opinion on the matter, crime barely came up at all.

Hopefully this will inspire more similar discussions in the future. And if Michael ever wants to debate me, he is welcome to do so for as long as he can tolerate my computer’s terrible microphone and constant whirring grinding noises.

Housekeeping Post February 2014

1. Today marks this blog's first birthday. Hurrah!

2. I will be in the Bay Area from March 1 to sometime between March 13 to March 16 – I don't have my flight out day yet. Some of you I obviously want to see. Others of you I don't know yet but probably would want to see if I knew you. If there are any rationalist community parties or meetups planned then, I will go to those and can meet people there – does anyone know the social schedule for that time period? Is anything sufficiently interesting going on March 14 or 15 that I should make sure to get the last possible flight home?

3. I asked for people to help Ozy with zir thesis on romance novels and got like a dozen volunteers including one professional romance novel writer. You are good people.

4. My ability to get journal articles is extremely limited. Somehow my friends who are freshmen in tiny liberal arts colleges have better access to major medical journals than I do working from the medical library of a large and well-funded hospital. Does anyone with good journal availability want to lend me some of their university library remote access information so I can continue to discuss interesting medical and psychological articles on here? You can email me at scott@shireroth.org **[EDIT: Taken care of. Thanks, everyone!]**

5. I'm done with the Less Wrong Survey for the year and have moved on to a survey for r/nootropics. If you are interested in and have tried nootropics (brain-enhancing drugs, eg gingko biloba, modafinil, piracetam, etc) please

take the survey to help me gather information on what kind of effects people think they get from them.

Bad Dreams

[Content warning: Scary stuff, gawking at weird diseases]

My first sign that something was up was when I noticed my Sleep Medicine textbook was written by a Dr. Dement. I reassured myself that it was probably just a coincidence, while secretly knowing I would shout *"Expecto patronum!"* if I ever ran into him at a conference, just to be sure.

The second sign was the introductory quote. A lot of textbooks have these. They're usually from a work of great literature, or a poem, or a song that has to do with the medical topic addressed. Anything to show that the doctors in this field are *cultured*, not just people who learn a bunch of biology without exploring the wider world of the arts. For the Sleep Medicine textbook, the introductory quote was from *Les Mis*:

> But the tigers come at night
> With their voices soft as thunder
> As they tear your hope apart
> As they turn your dream to shame

I like the song, but in context this is *really* creepy.

The third sign was everything in the textbook. Sleep disorders, and especially disorders of dreaming, read like something out of an Edgar Allen Poe story.

Patients with Epic Dreaming Disorder report emotionally neutral dreams with repetitive content – continually walking up endless staircases, running across unchanging landscapes, cleaning rooms, scrubbing stone walls – which start as soon as their head hits the pillow and don't end until they awaken. They complain of fatigue upon awakening, as if they have been working all night. When placed in the laboratory, physiological indicators of sleep are normal, and there is no increase in REM sleep time (although we now know that dreams can and do occur in non-REM sleep as well). The condition is currently poorly understood, not treatable, and sufferers report difficulty just trying to get doctors to believe they exist and aren't making it up.

Somewhat related to epic dreaming are pathological lucid dreams. Normal lucid dreams are fun experiences where you realize you're dreaming, take control of the dream, and spend the rest of the dream riding dolphins or kissing supermodels. Pathological lucid dreamers realize that they're dreaming, but this somehow turns the dream into a nightmare in which the dreamer is attacked by demonic figures, all while fully conscious and realizing the nature of the phenomenon. These dreamers report experiencing real pain from the attacks and sometimes go to great lengths to stay awake and avoid having to subject themselves to further dream attacks.

Lovecraft, who knew a thing or two about dreams himself, said that "the oldest and strongest fear is fear of the unknown", and this is part of why I find night terrors so scary. Night terrors are very simple – a sleeper wakes up very suddenly out of very deep sleep, eyes dilated impossibly wide, screaming at the top of their lungs and sometimes flailing their limbs in front of them as if trying to ward something off or protect themselves from something. When asked why, they do not explain, or cannot remember. The old consensus was that there *was* no stimulus; the night terrors occur during deep sleep, when dreams were once not believed to be possible. Now researchers are not so sure. One proposal floated seriously in my textbook was that whatever these people see or experience is so traumatic that the brain instantly represses it upon awakening. And this repression occurs so consistently and so completely that no one in the waking world knows or will ever know what causes night terrors.

Sleep paralysis is very common. The body tries to shut down most voluntary movement in sleep so that people don’t act out their dreams; sometimes it jumps the gun and paralyzes people when they are still awake. This would be scary enough as it is, especially since the loss of voluntary control of breathing makes victims misperceive that they are suffocating or choking to death. But driven by some sort of primal urge to make sense of the experience, sufferers often go further and hallucinate monsters immobilizing them. The classic presentation is a ghost or a devil sitting on the victim’s chest, preventing her from breathing. In other cases, the monster will be visible somewhere else in the bedroom, or just a “felt presence” who is not actually seen. I remember being terrified as a child when one of my friends reported seeing the Devil in his bedroom late at night; I was pretty atheist already but I also didn’t expect this particular friend to lie, and I was always a little uncomfortable about the story until I learned that these sorts of hallucinations are actually pretty common. But it’s not just devils and monsters. One of my patients today complained that nearly every night, as she falls asleep, she sees a huge spider crawling towards her over her bedsheets. The Wikipedia article on this condition makes fascinating reading.

Narcolepsy is a disorder of sleep regulation best known for its tendency to put people to sleep suddenly and inappropriately, which can be mildly amusing to onlookers right up until it happens when the person is driving a car. To patients it seems especially unpleasant as they “live their entire lives in constant state of extreme sleep deprivation”.

But the narcolepsy symptom that made the biggest impression on me was the loss of the dream-waking distinction. While healthy people are pretty good at keeping dream and reality separate, they both sort of blur together for narcoleptics and they have to *guess* which of their memories are dreams and which really happened. This sometimes leads to extremely unfortunate consequences – such as sexual assault allegations based on attacks later determined to have occurred in dreams.

Sleep violence is sleepwalking's more aggressive cousin. Those who suffer from this condition will unknowingly attempt violence against their friends and loved ones in their sleep. My textbook presents the case of a man who presented to a sleep clinic because nearly every night he was beating up his wife, who slept next to him in the bed, leaving some pretty impressive bruises and understandable marital discord (while I felt bad for the couple, I also wondered why they were still sleeping in the same room; I kick my girlfriend out of bed because ze tosses and turns too much). But the granddaddy of all sleep violence cases is the one where a man drove 23 km miles to his in-laws' house, assaulted his father-in-law, and stabbed his mother-in-law to death. Now at some point you have to wonder – are we just taking this guy's word for it that he was asleep the whole time? And so my textbook wanders into the fascinating world of "forensic sleep medicine" – the field where experts try to figure out whether someone who says they committed a crime while sleepwalking is telling the truth. In this case, they found that he probably was – his EEG was extraordinarily irregular, consistent with a very very severe sleepwalking disorder – and in the end he was found innocent.

Prodromal dreams are very creepy and there's a lot of debate over whether they exist as a real thing. At their worst, they seem like plain old magical thinking – dreams that warn you of a medical problem you would otherwise be unaware of, like dreaming of breast cancer, visiting the doctor, and learning you really have the disease. But research seems to back up at least some of these, with particular interest in prodromal cardiac dreams. In one study, frequent dreams about death were found to have a strong relationship with ejection fraction (a measure of cardiac health), and severe cardiac events (eg heart attacks) were found to be preceded by extraordinarily strong nightmares more often than chance. More romantic people like to think of this as "your body sending you a message", but I don't know if anyone's ruled out the more prosaic explanation that people can have a pretty good idea of their cardiac health – whether through symptoms, medical test results, or educated guesses based on lifestyle – and their anxiety about this gets reflected in their dreams.

And then there's Sudden Unexpected Death Syndrome, which is actually even worse than the name suggests. I know I've written about it before, but I never really get used it it being a real disease instead of something out of a horror story.

This condition occurs only in people of Southeast Asian origin. It goes like this. They are young and healthy, in their 30s and 40s. Then one night, usually around 3 AM, they start moaning in agony, thrashing about, and die suddenly in their sleep.

But it can actually be even worse than that. There are some anecdotal reports that at least some of these deaths are not as unexpected as the name of the syndrome implies, that patients have nightmares that they will soon die in their sleep. Sometimes they will go to heroic efforts to avoid sleeping. A description of such a case from Los Angeles:

> He was about twenty-one; I've subsequently found out this is a phenomenon in Laos, Cambodia. Everybody in his family said almost exactly these lines: 'You must sleep.' He said, 'No, you don't understand; I've had nightmares before—this is different.' He was given sleeping pills and told to take them and supposedly did, but he stayed up. I forget what the total days he stayed up was, but it was a phenomenal amount—something like six, seven days. Finally, he was watching television with the family, fell asleep on the couch, and everybody said, 'Thank god.' They literally carried him upstairs to bed; he was completely exhausted. Everybody went to bed, thinking it was all over. In the middle of the night, they heard screams and crashing. They ran into the room, and by the time they got to him he was dead. They had an autopsy performed, and there was no heart attack; he just had died for unexplained reasons. They found in his closet a Mr. Coffee maker, full of hot coffee that he had used to keep awake, and they also found all his sleeping pills that they thought he had taken; he had spit them back out and hidden them.

This happens sufficiently often among Southeast Asians that it is a recognized part of the various cultures of the region:

> The nightmare demons of the Far East can be lethal. In Japan, this type of death is known as pok-kuri; the Filipinos call it bangungot or batibat; and the Hmong people of Vietnam and Laos call it tsob tsuang. In Thailand, the being to fear is the phi am or 'widow ghost' who comes to steal away the souls of young men. Some men defend themselves from phi am by wearing lipstick at night, so that the ghost mistakes them for women and leaves them alone.

It is now a part of American culture as well – it was a string of cases of Sudden Unexpected Death Syndrome in Los Angeles that inspired the movie *Nightmare on Elm Street*.

Luckily there is now a more prosaic explanation. Sudden Unexpected Death Syndrome is believed to *mostly* be the same thing as Brugada Syndrome, a very unusual cardiac disease caused by mutations in a gene coding for a vital heart protein. It can be controlled with an implantable defibrillator, which can restart the heart if it falls into an otherwise fatal arrhythmia.

Nootropics Survey Results And Analysis

In case you are just joining us: nootropics are substances which purportedly improve mental functioning with relatively few side effects when used responsibility.

Caffeine is a good example. It improves wakefulness and energy, and when used responsibly is so safe it can be sold “over the counter” in coffee and other drinks.

Over the past few years, various substances from experimental drugs to common supplements have been claimed to satisfy these criteria. If they really work, obviously we want to know about it.

Double-blind randomized controlled trials have confirmed exciting properties for some purported nootropics. A meta-analysis has confirmed that the neuroprotective drug piracetam improves cognition in demented patients, and smaller studies suggest it does the same in healthy individuals. A meta-analysis of traditional Ayurvedic medicine bacopa monnieri finds it improves memory and reaction time in healthy subjects.

The online nootropics community consists of mostly healthy individuals looking to use these substances to gain motivation or clearer thinking. They tend to “outpace” the scientific literature, using some substances that haven’t yet been shown to work, or haven’t been shown to benefit healthy individuals. Information travels mostly through word of mouth, and people combine a bewildering variety of chemicals without being sure which ones work better than others or even whether any of them work at all.

This survey does not intend to rectify this state of affairs, but it can at least shed some light on it.

I asked people in the two largest online nootropics communities (that I know of), Reddit's r/nootropics and Longecity's Brain Health forum, to take a survey describing their use of 31 different substances. I got 162 responses. My respondents were 92% male and 8% female, with an average age of 25. Most of them had tried only a few different chemicals, but a few had sampled nearly two dozen. Most were taking "stacks" of several different chemicals daily.

Three substances – DMAE, centrophenoxine, and a combination stack called TULIP – got fewer than ten responses, so I removed them from the survey because there wasn't enough of a sample size to do anything with.

Below I describe the results I got on the remaining 28 substances and some hypotheses that I tested:

Head-to-Head Comparison Of Nootropics

Respondents were asked to rate their subjective experiences of different nootropics on a scale of zero (completely useless, did nothing) to ten (life-changing). Comments on Reddit suggest several people misunderstood the scale, but not much I can do about that at this point.

I asked people to reserve numbers five or above for effects that, in their subjective opinion, couldn't possibly be placebo – but the whole point of the placebo effect is that people can't consistently make this judgment. I would have liked to have some placebo nootropics to use as a comparison, but since I wasn't running a trial, I didn't have that option.

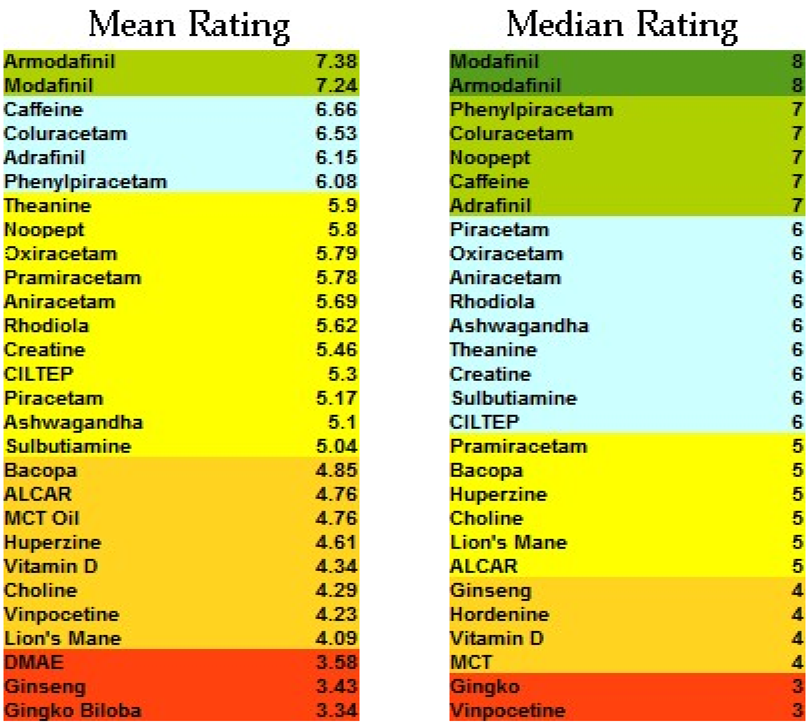
I did, however, try to assess the placebo effect in a couple of sneaky ways. One of the nootropics measured, choline bitartrate, has carefully regulated ability to cross the blood-brain barrier beyond a certain level, and so it would be surprising if it had direct and immediate cognitive effects. Nevertheless, 27 respondents had used this substance and they gave it an average rating of 3.7 out of ten. This suggests that 3.7 out of ten is a good first guess for the placebo effect.

Another of the nootropics measured, bacopa monnieri, has been found in multiple studies not to exert any effect before a month of consistent use. 6 respondents had used this substance for 20 days or less, and they gave it an average rating of 4 – which is very close to 3.7 and so confirms our previous guess.

These methods aren’t perfect – there are some mechanisms by which choline can relieve deficiency even in the absence of direct cross-BBB transport, and some people say that at least some anxiolytic effects of bacopa are immediate – but they’re the best we’ve got. So let’s say anything that comes up 4 or less has a good chance of being placebo.

This does not imply that numbers above 4 are therefore necessarily real effects; we know that more promising drugs exert stronger placebo effects. Thus, in the early days of antidepressants when no one believed they would work, they had a very weak placebo effect, whereas after the culture was saturated with messages of antidepressant efficacy, they started having a much stronger placebo effect. So a much-vaunted nootropic might well be able to muster a placebo effect somewhat greater than 4.

Below, I give tables ranking the nootropics studied along several axes.



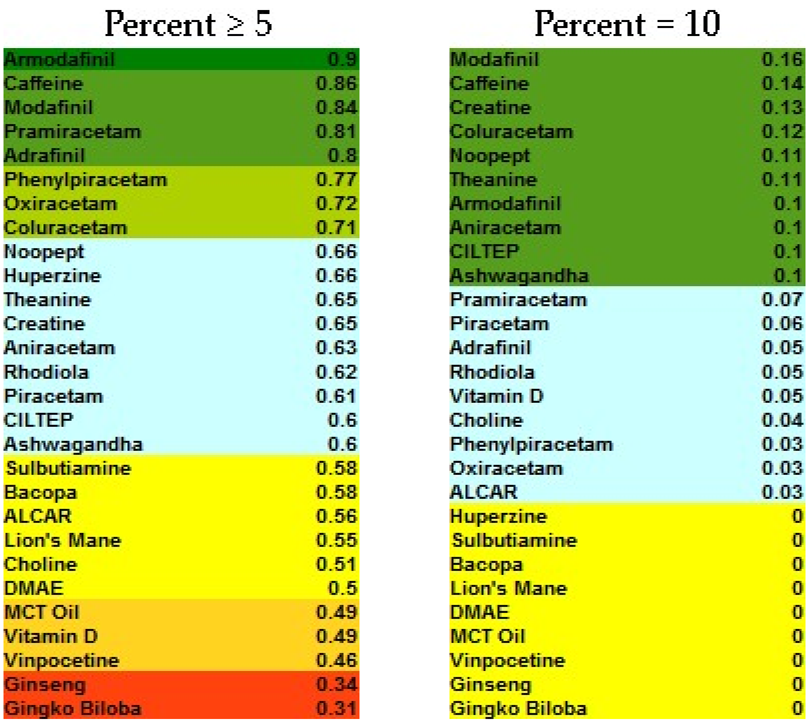
Mean Rating

Armodafinil	7.38
Modafinil	7.24
Caffeine	6.66
Coluracetam	6.53
Adrafinil	6.15
Phenylpiracetam	6.08
Theanine	5.9
Noopept	5.8
Oxiracetam	5.79
Pramiracetam	5.78
Aniracetam	5.69
Rhodiola	5.62
Creatine	5.46
CILTEP	5.3
Piracetam	5.17
Ashwagandha	5.1
Sulbutiamine	5.04
Bacopa	4.85
ALCAR	4.76
MCT Oil	4.76
Huperzine	4.61
Vitamin D	4.34
Choline	4.29
Vinpocetine	4.23
Lion's Mane	4.09
DMAE	3.58
Ginseng	3.43
Gingko Biloba	3.34

Median Rating

Modafinil	8
Armodafinil	8
Phenylpiracetam	7
Coluracetam	7
Noopept	7
Caffeine	7
Adrafinil	7
Piracetam	6
Oxiracetam	6
Aniracetam	6
Rhodiola	6
Ashwagandha	6
Theanine	6
Creatine	6
Sulbutiamine	6
CILTEP	6
Pramiracetam	5
Bacopa	5
Huperzine	5
Choline	5
Lion's Mane	5
ALCAR	5
Ginseng	4
Hordenine	4
Vitamin D	4
MCT	4
Gingko	3
Vinpocetine	3

The mean and median ratings for each substance



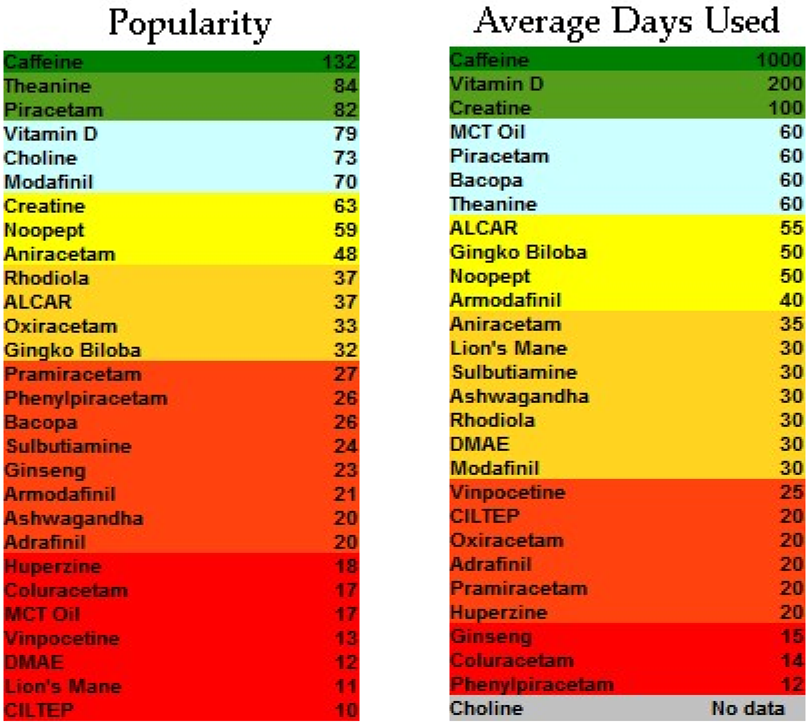
Percent ≥ 5

Armodafinil	0.9
Caffeine	0.86
Modafinil	0.84
Pramiracetam	0.81
Adrafinil	0.8
Phenylpiracetam	0.77
Oxiracetam	0.72
Coluracetam	0.71
Noopept	0.66
Huperzine	0.66
Theanine	0.65
Creatine	0.65
Aniracetam	0.63
Rhodiola	0.62
Piracetam	0.61
CILTEP	0.6
Ashwagandha	0.6
Sulbutiamine	0.58
Bacopa	0.58
ALCAR	0.56
Lion's Mane	0.55
Choline	0.51
DMAE	0.5
MCT Oil	0.49
Vitamin D	0.49
Vinpocetine	0.46
Ginseng	0.34
Gingko Biloba	0.31

Percent = 10

Modafinil	0.16
Caffeine	0.14
Creatine	0.13
Coluracetam	0.12
Noopept	0.11
Theanine	0.11
Armodafinil	0.1
Aniracetam	0.1
CILTEP	0.1
Ashwagandha	0.1
Pramiracetam	0.07
Piracetam	0.06
Adrafinil	0.05
Rhodiola	0.05
Vitamin D	0.05
Choline	0.04
Phenylpiracetam	0.03
Oxiracetam	0.03
ALCAR	0.03
Huperzine	0
Sulbutiamine	0
Bacopa	0
Lion's Mane	0
DMAE	0
MCT Oil	0
Vinpocetine	0
Ginseng	0
Gingko Biloba	0

The percent of respondents who rated each substance as >=5 (ie they could not believe it was placebo), and = 10 (ie *life-changing)*



Popularity

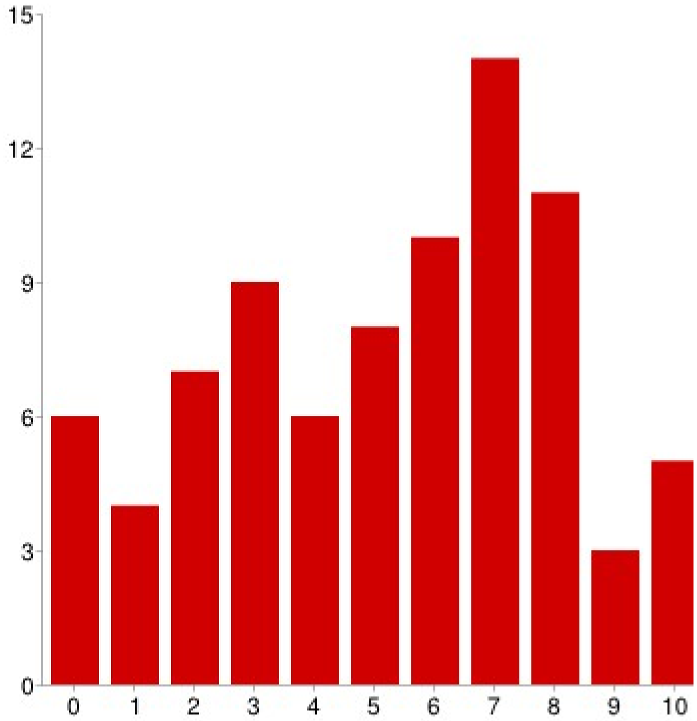
Caffeine	132
Theanine	84
Piracetam	82
Vitamin D	79
Choline	73
Modafinil	70
Creatine	63
Noopept	59
Aniracetam	48
Rhodiola	37
ALCAR	37
Oxiracetam	33
Gingko Biloba	32
Pramiracetam	27
Phenylpiracetam	26
Bacopa	26
Sulbutiamine	24
Ginseng	23
Armodafinil	21
Ashwagandha	20
Adrafinil	20
Huperzine	18
Coluracetam	17
MCT Oil	17
Vinpocetine	13
DMAE	12
Lion's Mane	11
CILTEP	10

Average Days Used

Caffeine	1000
Vitamin D	200
Creatine	100
MCT Oil	60
Piracetam	60
Bacopa	60
Theanine	60
ALCAR	55
Gingko Biloba	50
Noopept	50
Armodafinil	40
Aniracetam	35
Lion's Mane	30
Sulbutiamine	30
Ashwagandha	30
Rhodiola	30
DMAE	30
Modafinil	30
Vinpocetine	25
CILTEP	20
Oxiracetam	20
Adrafinil	20
Pramiracetam	20
Huperzine	20
Ginseng	15
Coluracetam	14
Phenylpiracetam	12
Choline	No data

The number of respondents (out of 162) who used each substance, and the mean number of times they have used it.

I did not include standard deviations in these tables, so they conceal the true amount of variation in each substance. In general, nearly every substance had at least one zero and at least one nine or ten. Very few showed an obvious consensus on whether they worked or not. I reproduce below a graph of piracetam – one of the most-used nootropics – as a very typical example:



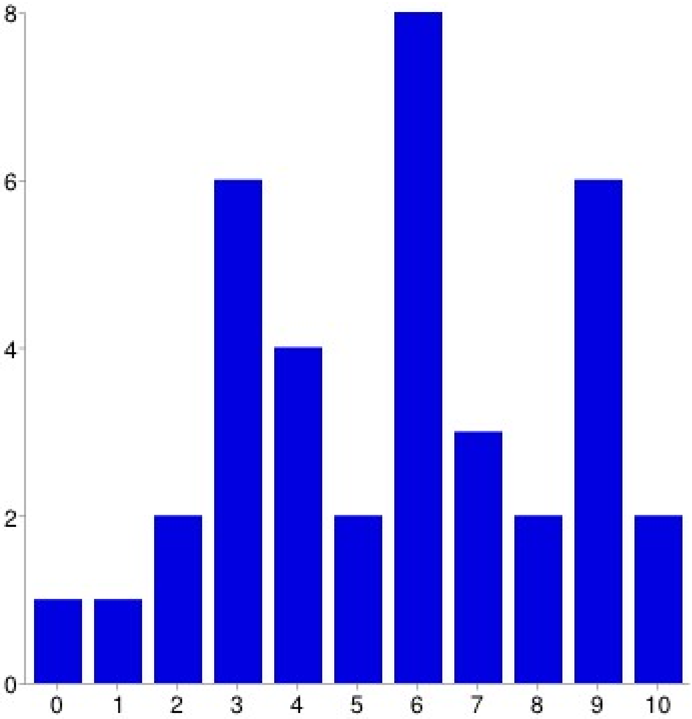
The results themselves were not surprising. The most effective nootropics were caffeine – which pretty much everyone in the world knows works – and modafinil (along with its fellow -afinils) which has been proven to work by countless studies and is a prescription drug much used by the medical establishment.

Below these two, the most effective “exotic” nootropics were phenylpiracetam and coluracetam. I suspect that despite the nootropic community’s enthusiasm about phenylpiracetam it mostly just works as a mild stimulant not too different from other stimulants like amphetamines. Coluracetam, a very new and pretty unique substance currently being investigated as an Alzheimer’s treatment, is more exciting.

Of the next two, Noopept is a Russian prescription drug mostly unknown in the United States; Russian studies suggest it improves memory and learning. Theanine is one of the active ingredients in tea and is widely considered calming. Their success should not surprise anybody paying attention.

Creatine is a chemical related to amino acids. Although it seemed to do well, most people who ranked it highly mentioned in the comments that they liked it as an exercise supplement but didn't get any cognitive benefits from it. It should probably be disqualified.

Ayurvedic medicines had a decidedly blah performance. A lot of people *swear by* rhodiola, ashwagandha, and bacopa, and they all seem to have a moderate amount of experimental support. But they were firmly in the middle of the pack here, outperforming our 4 threshold for placebo but only barely. Much like piracetam above, these were all over the place: here is the graph of rhodiola, the best-performing of the bunch:

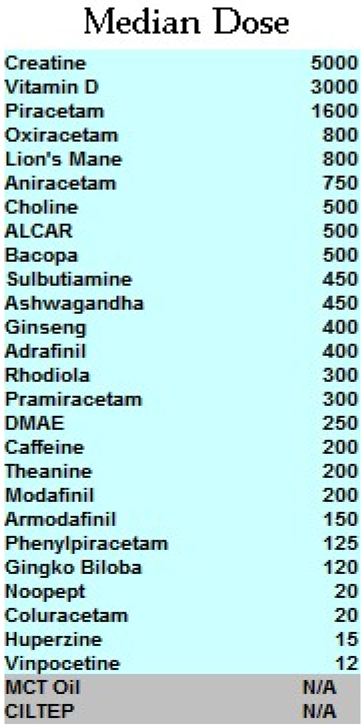


An optimist might say that everyone has a unique biochemistry. A pessimist might say that everyone has a unique ability to extract patterns from random noise. I leave the choice between these two hypotheses to the reader.

Two other traditional Eastern medications – ginseng and gingko – fell completely flat, with both doing (nonsignificantly) *worse* than our suspected placebo level.

Dosing Issues

A secondary goal of this survey was to inspect the dose-response relationship of these substances and see whether the recommended doses were really set at the most effective level. My original plan was to report the dose at which a substance was most effective, but this proved beyond my technical abilities. Instead I report the median dose given by all users of a substance:



Median Dose

Creatine	5000
Vitamin D	3000
Piracetam	1600
Oxiracetam	800
Lion's Mane	800
Aniracetam	750
Choline	500
ALCAR	500
Bacopa	500
Sulbutiamine	450
Ashwagandha	450
Ginseng	400
Adrafinil	400
Rhodiola	300
Pramiracetam	300
DMAE	250
Caffeine	200
Theanine	200
Modafinil	200
Armodafinil	150
Phenylpiracetam	125
Gingko Biloba	120
Noopept	20
Coluracetam	20
Huperzine	15
Vinpocetine	12
MCT Oil	N/A
CILTEP	N/A

In milligrams, except where it isn’t. Vitamin D is IUs.

I investigated dosing of two substances in more detail – piracetam and vitamin D – because I had good data on both of them and there has been some controversy on correct dosing.

The median dose of piracetam was 1600 mg. I ran a t-test to see whether people who took high doses (>1600) had stronger effects than those who took low doses (< 1600) - the means looked different (5.6 vs. 4.4) but the effect was only marginally significant (p = .09). I tried to see if I could break the significance barrier with a larger dichotomy and so compared people taking below the 25th percentile (1200) to those taking above the 75th percentile (3000). This actually had *less* difference – means of 4.42 vs. 4.88 – so I gave up. An attempt to correlate dose to response with Pearson's *r* also returned uninspiring results of r = -.04, p = .75.

The whole analysis doesn't really give any good reason to use anything other than the median dose of 1600 mg.

Vitamin D was much more interesting. I compared people taking more than the Vitamin D mean dose of 3000 IU with those taking less – the means seemed to differ (5.0 vs. 3.9), but the significance wasn't there (p = .17).

Once again I tried to gun for significance by switching to quartiles – in this case, below 1625 IU vitamin D vs. above 5000. The means widened to 5.3 vs. 3.4, and the p value reached 0.8, within a hair-breadth of significance.

I decided to make a last desperate effort and restrict the analysis to the true extremes. I compared the people taking OVER NINE THOUSAND units to those taking 1000 or less. Here the means were 7.4 vs. 3.1, and p was 0.02. I am not entirely confident in this analysis since there were some small subgroups – n = 12 in the very low group and n = 5 in the very high.

Also, it should be kept in mind that these correlations might not be causal in the expected direction – just to raise one possibility, maybe people who really like vitamin D take more of it.

I investigated a bunch of other substances for dose-response relationships using bivariate correlations and couldn’t find anything. Overall my attempts to investigate dosing were a disappointment.

Correlations

My first several attempts to correlate different nootropics with each other were failures. Everything correlated with everything else due to a general “low standards factor” – some people were just very effusive and willing to give high ratings to everything, while other people had a much higher threshold before getting excited.

I did something *very sketchy* to adjust for this, and divided every person’s ratings for each substance by their average rating for all substances, so that each substance’s “rating” was transformed into the degree for which it was below or above average for that particular rater. This introduced a whole host of complicated biases, but it at least got me to the point where I could run correlations between substances without everything lighting up simultaneously.

By correlating 28 substances with one another I ran 378 separate correlations, so with a significance threshold of p = .005 I would expect about two hits by chance. Instead I got about a dozen.

Some of these were expected – for example, the -afinils correlated with each other. Others were bizarre and nonsensical. I have removed the inter-afinil correlations as too obvious to be worth noting, and am left with the following:

ALCAR -> Modafinil -.733, .000
Vitamin D -> Theanine +.591, .000
Creatine -> Noopept -.629, .001
Vitamin D -> Rhodiola -.634, .002
Piracetam -> Sulbutiamine +.683 .002
Piracetam -> ALCAR +.548 .003
Piracetam -> Modafinil -.505 .003
Phenylpiracetam -> Choline -.681 .004
Huperzine -> Creatine -.810, .005
Gingko -> Piracetam +.602 .005

Bizarrely, none of the racetams correlated with each other!

I have very low confidence that these represent real effects rather than artifacts of the gymnastics I had to do to make the data correlatable at all, and I report them only to encourage other people to do more sober analyses.

Odds and Ends

Choline is a commonly discussed nootropic used both on its own and as an adjunct to other substances. There are four main kinds of choline – lecithin, choline bitartrate, CDP-choline, and alpha-GPC choline. The first two do not effectively cross the blood-brain barrier beyond a certain level and so should theoretically be pretty ineffective for nootropic purposes. The second two do directly cross the blood-brain barrier and may have an effect. When I asked people to rate their choline, I got the following means:

Choline bitartrate: 3.7 (n = 27)
Lecithin: 3.8 (n = 5)
Alpha-GPC choline: 5.07 (n = 15)
CDP-choline: 5.37 (n = 19)

I ran a t-test comparing the two presumably ineffective forms of choline to the two potentially neuroactive ones and found a significant result (p = .045). This suggests that the forms of choline that could theoretically work are more effective than than ones that don't – or that people are aware of the chemical properties of their forms of choline and adjusting their placebo effects accordingly.

Choline is sometimes used to improve the effects of -racetams, so I investigated whether people who took active forms of choline had better experiences with racetams than those who took inactive forms of choline. Out of six racetams, none of them achieved significance (although five out of the six had trends in the expected direction). This stayed pretty much the same when I restricted the analysis to people who told me specifically that they used choline with their racetams.

Choline is also sometimes used to prevent headaches with racetams, so I asked people about that. 40% said they found choline to help prevent headaches with racetams, 26% said they found it didn't, and 34% just took choline because everyone else told them they should and didn't know one way or the other.

The other category of drugs I looked at in more detail were the afinils. There are three common afinils. Modafinil is a well-known prescription medicine used by doctors as a wakefulness-promoting agent to treat narcolepsy, shift-work sleep disorder, and a couple other diseases. Adrafinil is an over-the-counter substance that gets metabolized to modafinil and should have about the same effecst. Armodafinil is an isomer of modafinil that should do about the same thing but require a lower dose.

I asked respondents if they had any preferences among these three similar substances. Of those willing to compare adrafinil and modafinil, 50% said modafinil was better, only 8% thought adrafinil was better, and 42% didn’t want to pick a favorite.

Of those willing to compare modafinil and armodafinil, 62% thought armodafinil was better, 23% thought modafinil was better, and 15% didn’t want to pick a favorite.

There is some theoretical concern about liver-related side effects from adrafinil, but some pharmacologists say these are overblown and that its liver profile is similar to Tylenol – ie don’t take a massive overdose on it or use it every day for years and you’ll be fine. None of my respondents reported ever having any liver problems with adrafinil – not even asymptomatic elevation of liver enzymes – but n = only 17 so the results don’t rule out even moderately common adverse reactions.

Last of all, there are some worries about tolerance to modafinil. Although the medical community is happy to give it to patients for several years straight and expects its results to remain just as strong as when it was first prescribed, the nootropics community has a lot of anecdotal evidence of tolerance developing if it is used too much.

I compared the modafinil experience ratings of people who used the substance occasionally (n = 52) to people who used it every day (n = 7). This corresponded, as expected, to a wide gap in how many doses they had taken on average – the occasional users had taken it 39 times, compared to the everyday users’ 184. However, the everyday users did not report decreased satisfaction with modafinil – both groups gave it an average satisfaction rating of 7.4. This suggests that at least this small subgroup of 7 people are not experiencing tolerance. We should, however, be careful to worry that they might be self-selected for exactly that quality – maybe a lot people try to use it everyday, but all but a rare group of non-tolerance-developers later back off.

Conclusions and Future Research

Overall I was not too impressed with the results of this survey. No unexpected substances jumped out as promising, there was a very wide distribution from “useless” to “life-changing” in nearly everything except the -afinils, and it was very hard to find much of a dose-response relationship anywhere. There is no “smoking gun” for most these substances being definitely-exciting, aside from the ones like modafinil that we were already aware of.

I repeat the other survey I run yearly and I’d like to establish a similar tradition here. Next year I hope this survey will have earned me enough cred in the nootropics community to be able to harass more people into taking it – whether that means getting it stickied on top of appropriate forums, getting nootropics distributors to advertise on their sites, or just getting more word of mouth. I know there are several thousand people in the community and I feel like I could do a lot more if I had two or three times as many responses.

Other goals for the future are to drag people kicking and screaming into reading directions so I know everyone’s rating off the same scale, to ask more specifically about tolerance development, and perhaps to separate main effects from side effects (although that brings up a lot of complications I will have to sort through).

Below are the raw data available for download, minus about a dozen people who asked for their entries to stay private. I am making it available in two forms. First are the raw raw data, which include people’s comments, people’s stacks, and a whole bunch of incorrectly entered machine-unreadable responses that I had to go through by hand and correct □_□ . The sanitized raw data are clear, pure numbers that you can get to work on right away, *if* you don’t mind the fact that I deleted a lot of ambiguous responses and removed some stuff entirely just because keeping it seemed like a lot of work. I put a very brief explanation of how I coded some bits of the sanitized data up here; otherwise you’re on your own or can ask me questions if it’s too confusing.

I know Gwern is hoping to analyze these data further and I look forward to seeing what other people can make of them.

Raw data: **(.xls)** **(.csv)**
Sanitized data: **(.xls)** **(.csv)**

Lies, Damned Lies, And Social Media (Part 5 of ∞)

[content warning: rape, false rape allegations. Some people have been linking this article claiming it says things it DEFINITELY DOES NOT, so please read it before you have an opinion.]

(see also parts 1, 2, 3, and 4 of ∞)

I.

Spotted on Brute Reason but liked and reblogged 35,000 times: Five Things More Likely To Happen To You Than Being Accused Of Rape. A man is 631 times more likely to become an NFL player than to be falsely accused of rape! Thirty-two times more likely to be struck by lightning! Eleven times more likely to be hit by a comet!

Needless to say, all of these figures are completely wrong, in fact wrong by a factor of over 22,700x. I'm not really complaining – missing the mark by only a little over four orders of magnitude is actually not bad for a "story" of this type. Nevertheless, it will be instructive to figure out where they erred so we may be vigilant against such things in the future, and perhaps certain moral lessons may be gleaned in the process as well.

II.

Since that article itself does not show its work, we will have to rely on its obvious inspiration, an almost-identical blog post written a few days before by the same person responsible for the Buzzfeed piece, Charles Clymer.

It starts by noting that there are about 84,000 forcible rapes per year – and that FBI statistics suggest 8% are false accusations. We will examine these numbers later, but for now let’s just take them as given.

It then goes on to calculate that, given the average man has sex 99 times per year (who *is* this average man?!) there are 5.1 billion acts of sexual intercourse in the United States each year among American men 15 – 39. Divide 5.1 billion by 6,750, and therefore, in Clymer’s words “the odds of any sexually-active male between the ages of 15 and 39 has a 750,000 to 1 chance of being falsely accused of rape”

And, he goes on to say, 1/33 men are raped during their lifetime. Therefore, the average man has a 27500x higher chance of being raped than being falsely accused of rape. The average man has a 1 in 84,079 chance of being killed by lightning, so that’s 32x more likely than getting falsely accused of rape. And it adds that the average women has a 1/4 chance of being raped during her lifetime – so the odds of a woman being raped during her lifetime must be 220000x higher than the odds of a man being falsely accused of rape.

Did you spot the sleight of hand in those calculations? He calculated the odds of a man who has sex 99 times per year for 24 years being accused of rape *per sex act*, and then declared this was the odds of being accused of rape *in your lifetime*. Then he went on to compare it to various other lifetime odds, like the lifetime odds of being raped, the lifetime odds of being struck by lightning, et cetera.

This isn’t comparing apples to oranges. This isn’t even comparing apples to orangutans. This is comparing apples to the supermassive black hole at the center of the galaxy.

To highlight exactly how awful this is, suppose we wanted to trivialize rape itself through the same methodology. The average woman, as per the article's statistics, has a 1/4 chance of getting raped during her lifetime, which means a 1/9500 or so chance of getting raped per sex act if she has sex 99 times per year from ages 15-39. And looking at the same list of statistically unlikely things provided on that article, that's less than the odds of dying in a plane crash (1/7032). So you crow "THE AVERAGE WOMAN IS LESS LIKELY TO GET RAPED THAN TO DIE IN A PLANE CRASH! HA HA WOMEN ARE SO DUMB TO EVER WORRY ABOUT RAPE!". And now you have a Buzzfeed article.

III.

We can do better. Let's come up with conservative and liberal estimates of a man's chance of getting falsely accused of rape between ages 15 and 39.

The rate of false rape accusations is notoriously difficult to study, since researchers have no failsafe way of figuring out whether a given accusation is true or not. The leading scholar in the area, David Lisak, explains that the generally accepted methodology is to count a rape accusation as false "if there is a clear and credible admission [of falsehood] from the complainant, or strong evidential grounds", and goes on to explain what these grounds might be:

> For example, if key elements of a victim's account were internally inconsistent and directly contradicted by multiple witnesses and if the victim then altered those key elements of his or her account, investigators might conclude that the report was false

Attempts to use this methodology return varying results. Lisak lists seven studies he considers credible, which find false accusation rates of 2.1%, 2.5%, 3.0%, 5.9%, 6.8%, 8.3%, 10.3%, 10.9%. The two with 10%+ mysteriously go missing and thus we get the commonly quoted number of “two to eight percent”, which is repeated by sources as diverse as Alas, A Blog, Slate, and Wikipedia (Straight Statistics keeps the original 2% – 10% number)

Feminists make one true and important critique of these numbers – sometimes real victims, in the depths of stress we can’t even imagine, do strange things and get their story hopelessly garbled. Or they suddenly lose their nerve and don’t want to continue the legal process and tell the police they were making it up in order to drop the case as quickly as possible. All of these would go down as “false allegations” under the “victim has to admit she was lying or contradict herself” criteria. No doubt this does happen.

But the opposite critique seems much stronger: that some false accusers manage tell their story without contradicting themselves, and without changing their mind and admit they were lying. We’re not talking about making it all the way through a trial – the majority of reported rapes get quietly dropped by the police for one reason or another and never make it that far. Although keeping your story halfway straight is probably harder than it sounds sitting in an armchair without any cops grilling me, it seems very easy to imagine that *most* false accusers manage this task, especially since they may worry that admitting their duplicity will lead to some punishment.

The research community defines false accusations as those that can be proven false beyond a reasonable doubt, and all others as true. Yet many – maybe most – false accusations are not provably false and so will not be included.

So there’s reason to believe some of those 2-10% of presumed false accusations are actually true, and other reasons to believe that some of the 98% – 92% of presumed true accusations are actually false.

What is an upper bound on the number of false rape accusations? Researchers tend to find that police estimate 20%-40% of the rape accusations they get to be “unfounded”, (for example Philadelphia Police 1968, Chambers and Millar 1983, Grace et al 1992, Jordan 2004, Gregory and Lees 1996, etc, etc). Many scholars critique the police’s judgment, suggesting many police officers automatically dismiss anyone who doesn’t fit their profile of a “typical rape victim”. A police-based study that took pains to avoid this failure mode by investigating all cases very aggressively (Kanin 1994) was criticized for what I think are ideological reasons – they primarily seemed to amount to the worry that the aggressive investigations stigmatized rape victims, which would make them so flustered that they would falsely recant. Certainly possible. On the other hand, if you dismiss studies for not investigating thoroughly enough *and* for investigating thoroughly, there will never be any study you can’t dismiss. So while not necessarily endorsing Kanin and the similar studies in this range, I think they make a useful “not provably true” upper bound to contrast with the “near-provably false” lower bound of 2%-10%.

IV.

But this only represents the number of false rape accusations that get reported to the police. 80% of rapes never make it to the police. Might false rape accusations be similar?

Suppose you are a woman who wants to destroy a guy’s reputation for some reason. Do you go to the police station, open up a legal case, get yourself tested with an invasive rape kit, hire an attorney, put yourself through a trial which may take years and involve your reputation being dragged through the mud, accept that you probably won’t get a conviction anyway given that you have no evidence – and take the risk of jail time if you’re caught lying?

Or do you walk to the other side of the quad and bring it up to your school administrator, who has just declared to the national news that she thinks all men accused of rape should be automatically expelled from the college, without any investigation, regardless of whether there is any evidence?

Or if even the school administrator isn’t guilty-until-proven-innocent enough for you, why not just go to a bunch of your friends, tell them your ex-boyfriend raped you, and trust them to spread the accusation all over your community? Then it doesn’t even *matter* whether anyone believes you or not, the rumor is still out there.

This last one is the one that happened to me. I wasn’t the ex-boyfriend (thank God). I was the friend who was told about it. I took it very very seriously, investigated as best I could, and eventually became extremely confident that the accusation was false. No, you don’t know the people involved. No, I won’t give you personal details. No, I won’t tell you how I became certain that the accusation was false because that would involve personal details. Yes, that leaves you a lot of room to accuse me of lying if you want.

But if my word isn’t good enough for you, I happen to have witnessed two *more* cases of false rape accusations where I *can* tell you some minimal details. In a psychiatric hospital I used to work in (not the one I currently work in) during my brief time there there were two different accusations of rape by staff members against patients…

I want to take a second out to say *very emphatically* that *all accusations of rape by psychiatric patients should be taken very seriously*. Yes, psychiatric patients sometimes have complicated cognitive or personality issues that make them more likely to falsely report rape, but for *exactly this reason* they are much more vulnerable and people are much more likely to take advantage of them. This is a *known problem* and you should *never dismiss their complaint*.

…but in this case, there were video cameras all over the hospital and these were sufficient to prove that no assault had taken place in either case. Now I know someone is going to say that blah blah psychiatric patients blah blah doesn't generalize to the general population, but the fact is that even if you accept that sorta-ableist dismissal, those patients were in hospital for three to seven days and then they went back out into regular society. I would love to say that we treated every single one of their problems so thoroughly it would never come back but I wouldn't bet on it.

So I know three men who have been accused of rape in a way that did not involve the police, and none (as far as I know) who have been accused in a way that did. This suggests that like rapes themselves, most false rape accusations never reach law enforcement.

While rape victims have some incentives to report their cases to the police – a desire for justice, a desire for safety, the belief that the evidence will support them – false accusers have very strong incentives not to – too much work, easier revenge through other means, knowledge that the evidence is unlikely to support them, fear of getting in trouble for perjury if their deception gets out. So I consider it a very conservative estimate to say that the ratio of unreported to reported false accusations is 4:1 – the same as it is with rapes. A more realistic estimate might be as high as double or triple that.

V.

Now we have the data necessary to do a slightly better job calculating the risk of false rape allegations. We'll start with the most conservative possible estimate.

We will stick with the article's figure of 84,000 reported rapes per year and 8% false accusation rate, for a total of 6,750 falsely accused.

We go on to assume, for the sake of conservativism, that there has never been a single false accuser who did not later confess, and that there has never been a false accuser who did not go to the police (my own memories of this must be hallucinations).

Since there are 53 million men ages 15-39 in the United States, the probability of being one of these 6,750 falsely accused is 1/7850 per year. But since you have 24 years in that age range in which to be accused, your lifetime probability of being falsely accused is about 1/327, or 0.3%. This is small, but according to Clymer's list it's about the same as your risk of dying in a car crash. Do you worry about dying in a car crash? Then you are allowed to worry about being falsely accused of rape.

(note that this is the most conservative possible estimate, using exactly the same numbers as in the article but not lying about what math we're doing. But the article got

1/750,000. So the absolute lower bound for how wrong the article was is “wrong by a factor of 2,300x”)

What about a slightly less hyperconservative estimate? Continuing our conservative assumption that there has never been a false accuser who has not later confused, but allowing that false accusations reach the police at only the same rate that rapes do, 1.5% of men will get falsely accused.

What estimate do I personally find most likely? Suppose we keep everything else the same, but allow that for every false accuser who later confesses, there is also one false accuser who does not later confess. This raises the false accusation rate to 16% – which, keep in mind, is still less than half of what the police think it is, so it’s not like we’re allowing rape-culture-happy cops to color our perception here. Now 3% of men will get falsely accused.

What is an upper bound for the extent of this problem? We could obtain one by using Kanin’s 40% and holding everything else constant, but no matter how many times I qualified this attempt with “I am using this as an upper bound, not endorsing this as the actual number of rapes”, someone would yell at me for using a study they disagree with and call me a rape apologist. So I will leave the difficult task of multiplying 3% by 2.5x to my readers. You might then try multiplying it even further if you think false accusations are less likely than true accusations to make it to the police.

So greater than 0.3% of men get falsely accused of rape sometime in their lives, and the most likely number is probably around 3%.

Which means the article was off by a factor of at least 2,300x and probably more like 22,700x.

And yet it got 35,000 Tumblr likes and reblogs. By blatantly lying in a sensationalist way, it became more popular than anything you or I will ever write. There are scientists dedicating their lives to making new discoveries on the frontiers of knowledge, poets making words dance and catch fire, struggling writers trying to tell the stories inside of them – all desperate for someone to pay attention to what they're saying – and the Internet ignores these people and instead brings hundreds of thousands of hits and no doubt a big windfall in ad revenue to frickin' Buzzfeed.

And I would like to just let it be, except that there's a probably one-in-thirty but definitely-no-less-than-one-in-three-hundred chance that I will be falsely accused of rape someday, and need to defend myself, and maybe I'll have what should be an airtight alibi, and then the people who read this Buzzfeed article will dismiss it with "Well, I saw on the Internet there's only a one in a million chance you're telling the truth, so screw your alibi!" This is *already happening*. One of the Tumblr rebloggers added the comment "Yeah, so you know the dude who says he was falsely accused of rape? Now you know. He's a rapist." These are not just falsehoods, they're *dangerous* falsehoods.

So please permit me a second to gripe about this.

It is commonly said that a lie will get halfway across the world before the truth can get its boots on. And this is true. Except in the feminist blogosphere, where a lie will get to Alpha Centauri and back three times while the truth is locked up in a makeshift dungeon in the basement, screaming.

I have been debunking bad statistics for a long time. In medicine, in psychology, in politics. Click on the “statistics” tag of this blog if you don’t believe me. Yet the feminist blogosphere is the *only* place where I consistently see things atrociously wrong get reblogged by thousands of usually very smart people without anyone ever bothering to think critically about them. Like, thirty five thousand feminists – including some who self-identify as rationalists! – saw an article that literally said a guy was *more likely to get hit by a comet than get falsely accused of rape*, and said “Yeah, sure, that sounds plausible”.

So please permit me to keep griping just one moment longer. *Be extraordinarily paranoid when dealing with the feminist blogosphere.* This may be true of all highly charged political blogospheres, but it is *certainly* true of feminism. If you go in there with an innocent attitude of “Here is a number, I assume it is generally correct and means what it says it means”, you will get *super-burned*

There are some honorable exceptions. I have found Alas, A Blog to be pretty scrupulous, and of course everything ever written by Ozy is wonderful and perfect in every way. But two swallows do not make a summer, and these and any similar blogs you find should be considered islands of lucidity battered by a constant tide of bullshytte. I do not have time to debunk them all but you should view them with a prior of extraordinarily high suspicion.

Thank you for letting me get that out of my system.

VI.

Why would this happen? Why would smart people, by the tens of thousands, be so delighted by the opportunity to embrace these fabrications?

There is something called the “just world fallacy”, that says everyone gets what they deserve and moral questions are always easy and there is never any need to make scary tradeoffs.

And, as is so often the case for things with “fallacy” in the name, it is not true.

Look at how the Clymer article, in its own words, describes false rape allegations:

“False rape hysteria”, it informs us, is perpetrated by “men’s rights activists, more accurately known as insecure woman-hating assholes”, because they think “women are products to be bought and sold and when these objects assert their right to human value many (if not most) men feel threatened.”

Now let’s hear from a guy on the r/mensrights community on Reddit:

> Anyway, like I said, it’s been just over a year since [I was falsely accused of rape]. Since then I haven’t been the same. The most striking thing that I’ve noticed is the paranoia that I have almost every waking moment. Of everybody. Of men, of women, and even friends. I can’t bring myself to date women anymore. I have panic attacks every time I see a police officer. I constintly think that I’m being followed. The night I came home from being interviewed by the cops I drank myself to sleep and I’ve been doing that ever since. If I don’t any flicker of light makes me think that the police are here to arrest me. I’ve been able to fake a normal social life to my family and work and the friends I have left but most don’t know anything about this. I’m not looking for pity from anyone. In fact, I’m doing better than I have

> been. The reason I’m posting this is because I want people to know how bad being accused of something like rape can hurt and scar someone.

Man, what an “insecure, woman-hating asshole.”

But consider the alternative to this kind of glib dismissal.

3% of men are falsely accused of rape. 15% of women are raped. If someone you know gets accused of rape, your prior still is very very high that they did it.

I was extraordinarily lucky to find very strong evidence that my friend was innocent. I was extraordinarily lucky that both my co-workers had video feeds that could confirm their stories. If I hadn’t, I don’t know what I would have done. My two choices would have been to either accept the possibility that I’m staying friends with a rapist, or to accept the possibility I’m ostracizing someone for something he didn’t do.

And someone is going to expect me to conclude by recommending what the correct thing to do in these cases is, but *I have no idea*. Probably there is no solution that isn’t horrible. If there is, it’s way above my pay grade. Ask Ozy. Ze’s the one with the Gender Studies degree.

All I can suggest is that you not flee from the magnitude of the decision with comfortable lies.

One of those comfortable lies is to tell yourself that all women are lying sluts so the accusation can be safely ignored.

But another comfortable lie is that false rape accusations are eleven times rarer than getting hit by comets.

This is why a terrible article on Buzzfeed is getting more publicity and support than anything you or I will ever write.

Because people want to live in his world, where the comfortable lies are all true and no one suffers without deserving it.

Proposed Biological Explanations For Historical Trends In Crime

My debate on crime rates with Michael Anissimov has been long and meandering, but I think we're starting to come to something of a consensus. I think (I don't know if Michael agrees) that the evidence showing long-term decline in crime from the Middle Ages to the Industrial Revolution is pretty good. There's also irrefutable evidence showing decline in crime from about 1985 to the present. That leaves a gap from about 1850 to 1980.

I previously asserted crime was stable during that period, pointing out similar murder rates between 1850 New York and London and 1980 New York and London, which I trusted more than (say) burglary rates. But Michael replied with a 2002 study showing that improved medical technology has saved a lot of murder victims and bumped their attackers' crimes down to attempted murder, meaning the apparent murder rate is artificially low. Correct that, and murder could have increased by 5-10x or more from 1850 to 1980, which would not be too different from the rates in lesser crimes like burglary.

I am still not entirely certain about this. We have good records on attempted murders for the past 30 years or so, and they have been going down along with the murder rate. And it is surprising that the improvement in medical technology so perfectly balances out the increase in violence. But it's a strong study, and so I will provisionally accept that crime including murder could have risen by 5-10x or more from 1850 to 1980.

But we don't have to accept that the reason is too much democracy or some sort of wacky political point like that.

I have previously come out as a biodeterminist. I suspect most social influences matter less than anyone thinks and most biological influences matter more than anyone thinks. When I say that, everyone always assumes I’m talking about genes, which is too bad because genes are almost the *least* interesting aspect of biodeterminism.

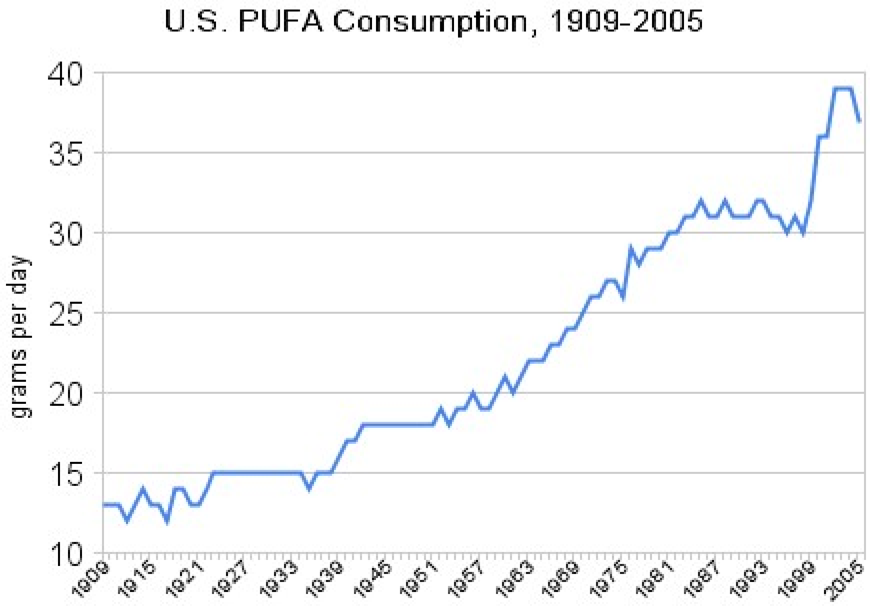
Anyone reading this blog probably already knows that lead is very strongly suspected of causing crime. A generation after gasoline was leaded, crime increased by a factor of four; a generation after lead was banned from gasoline, crime decreased by a factor of four. Levels of automobile lead emissions were found to explain 90% of the variability in violent crime in America. States that banned lead more quickly saw crime drop more quickly. Neighborhoods with higher lead levels consistently had higher crime rates. Blood lead levels show a marked inverse correlation with IQ, and a marked direct correlation with criminal history, even when plausible confounders are taken into account. And neuroscientists have known for decades that lead damages parts of the brain normally involved in good decision-making and in impulse control.

Lead levels started rising with the Industrial Revolution and, although in decline, are still far higher than in pre-industrial societies. They are highest in cities and especially in the inner city. They have shown correlation with crime, teenage pregnancy, and many mental disorders.

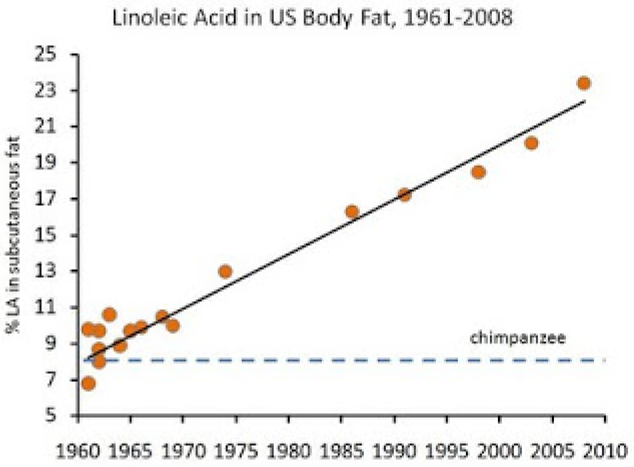
But like I said, everyone reading this blog probably already knows that. So let me talk about something I just learned last week.

Omega-6 fatty acids.

These are some of those “polyunsaturated fatty acid” things you always hear nutrition geeks talking about. They were pretty rare in human diets until the advent of industrial food processing. Here is a mysterious graph for which I have no source:



Here’s another that comes from Stephan Guyenet:



So suffice it to say that our consumption of these fatty acids has increased *a lot*. This is not surprising – they are most common in things like the vegetable oil that a bunch of preserved foods have.

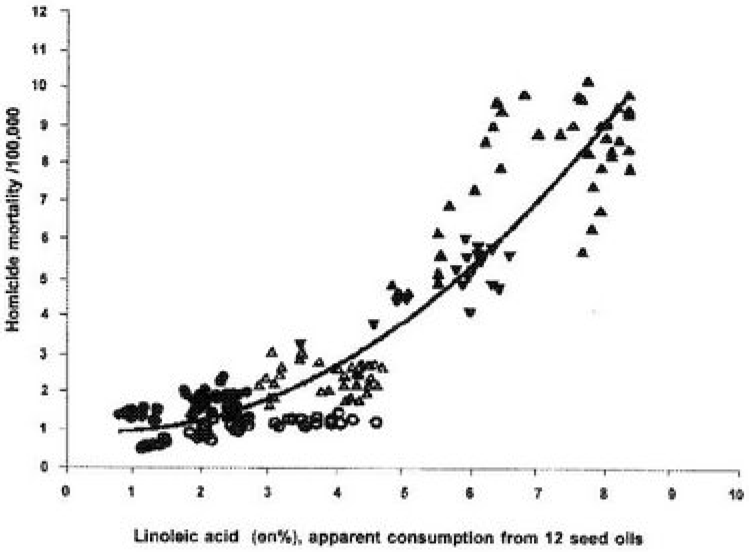
The other main kind of polyunsaturated fatty acid, omega-3, is mostly found in seafood and is the main component of the infamous “fish oil”. It hasn’t increased very much at all and so most people have an abnormally high omega-6:omega-3 ratio compared to the past and to the environment of evolutionary adaptedness.

Omega-3 and omega-6 fatty acids are important for cell membrane fluidity, especially in the brain where they affect neurotransmitter receptors and other neural functions. If there are the wrong amounts of them, this would very plausibly derange various cognitive functions.

So let’s look at Joseph Hibbeln’s paper Seafood Consumption and Homicide Mortality.

The Guardian describes it like so: “Hibbeln and his colleagues have mapped the growth in consumption of omega-6 fatty acids from seed oils in 38 countries since the 1960s against the rise in murder rates over the same period. In all cases there is an unnerving match. As omega-6 goes up, so do homicides in a linear progression. Industrial societies where omega-3 consumption has remained high and omega-6 low because people eat fish, such as Japan, have low rates of murder and depression.”

From Stephen Guyenet’s excellent post Vegetable Oil and Homicide:



I know, I know, it's a nice pretty line, but where are the randomized controlled trials?

To which one answers: "in dozens of different countries around the world". One of the most famous is Gesch et al 2002, which gave dietary supplements including fish oil or placebo to 231 prisoners and found a 25% drop in prison violence (p = 0.03) using intention to treat and 35% (p = .001) using completers. A replication study on 231 Dutch prisoners found almost exactly the same results. Another study of 468 schoolchildren also showed exactly the same results. And…actually, I'm just going to quote from Anatomy of Violence: The Biological Roots of Crime, a book I just found on Google and have suddenly conceived a burning desire to own:

> In Australia, six weeks of omega-3 supplementaion reduced externalizing behavior problems in juveniles with bipolar disorder. In Italy, normal adults taking omega-3 for five weeks showed a significant reduction in aggression compared to controls. In Japan, a randomized controlled trial found that ADHD children with oppositional definat disorder showed a 36%

> reduction in their oppositional behavior after fifteen weeks of omega-3. In Thailand, a randomized double-blind trial of the omega-3 fatty acid DHA resulted in a significant reduction in aggression in adult university workers. In the United States, women with borderline personality disorder randomized into supplementation of the fatty acid EPA for two months showed a significant reduction in aggression. Another American study, this time a four-month randomized double-blind placebo-controlled trial of fatty acid supplementation in fifty children, showed a significant 42.7% reduction in conduct-disorder problems.

We have been burned by omega-3 before. Every couple of weeks someone makes an exciting claim about it, and a few weeks later it is shown to be false or overblown. A big government review of the research on mental health basically dismisses everything done thus far as insufficient to draw meaningful conclusions. But I am hopeful.

I will add one more chemical, one of my favorites. Lithium. Many studies (1, 2, 3 find strong (that last one is p = .00003) links between lithium levels in the water supply and an endpoint crime or suicide. Lithium is a known neuroprotective agent, is probably at least calming, and may be otherwise good for the brain.

I am not certain of this, but I have heard from a few sources that modern water treatment/purification removes most minerals, which would suggest we are getting much less lithium than people in the old days who got their water from a well or whatever.

So we are likely getting more lead, more omega-6 (and relatively less omega-3), and less lithium than people in 1850. If there has been an increase in crime and other undesirable/impulsive behaviors, I think these biological insults are at least as worthy of examination as political changes that have occurred during that time.

God Bless Longecity

I am inordinately pleased by the existence of Longecity.

This is a forum where people discuss supplements and nootropics. I started looking at it when I was conducting the nootropics forum, assumed it was similar to every other forum that fits that description, and didn't pay it more attention until today.

Today I looked at it in more depth and discovered their group buys.

The way this works is: people on the forum hear about some exciting new chemical that was found to have promising effects in an experiment, after which researchers say something like "this might be ready for human trials in a couple of years".

Then everyone pools their money together, pays thousands of dollars to get a research laboratory somewhere to synthesize them a big batch of the chemical, distributes it to everyone in the forum, and they all ingest it and see if they grow wings or a third arm or whatever.

Needless to say this is a terrible idea and they will all probably die of some horrendous disease unknown to medical science. That's not even in dispute.

But I was raised on science-fiction stories, and one of the most common tropes was some wonder drug being developed and then suppressed by the government. When I was like seven, I read a story about a kid in a science fair who invented a chemical that let people photosynthesize, and the government made him stop because it would destroy the food industry (my childrens' books were better than yours). The basic structure of the mad scientist genre is small-minded fools delaying a genius in supremely great work. And in a world where the government is trying to put the brakes on the personal genomics revolution and drugs succeed or get ignored for confusing reasons, it's easy to worry that real life might borrow from some of those tropes.

For example, a research team recently did something that looked like they might sort have reversed aging in rats. The lead researcher says he wants to start human trials soon, but he "is reluctant to forecast how long it will be before the compound might be readily available for use", the chemical seems to be really expensive to manufacture in sufficient quantities, and the bioethicists will want to have their say about what if rich people can afford it more easily than poor people, and so on and so forth. Overall this doesn't seem likely to be an exception to the rule that nothing ever makes it to market in less than five or ten years.

Meanwhile, the people on Longecity have already gotten a price tag from a chemical supplier and a couple dozen people willing to be human guinea pigs.

This particular case I am not optimistic about – they seem to be looking not at the precise chemical used in the rat studies but at a cheaper substitute; although as far as I can tell their biochemistry checks out, I feel like if the cheaper substitute worked the original researchers would have used it themselves. And this substitute also seems to be commercially available in some places and no one is aging in reverse. So this will probably be a no-go.

But they're also working on group buys for a kappa opioid antagonist that someone says might help social motivation and a TrkB agonist that acts as a BDNF-mimetic and appears to treat Alzheimers in a mouse model.

I cannot even come close to endorsing this. The responsible part of me says that it is a terrible terrible idea and they will all come down with exotic cancers and die. It probably is a sweeping condemnation of our government that they haven’t burnt the whole website to the ground and then scattered red tape over the ruins so nothing can grow there again.

But another part of me is full of glee that *a random Internet site has just made an entire class of dystopias impossible*. If there is a miracle drug out there that makes you super-smart or reverses aging or something, and for some reason like an evil conspiracy or just bad luck it never “made it”, someone on Longecity will have taken it within a week of it first being mentioned online. And if they report that they think it worked, a couple hundred people will figure out hare-brained and probably illegal ways to get some.

I think a good motto for western civilization would be “Our institutions are stupid and our population is insane, but occasionally these two flaws perfectly balance each other out and it’s sort of neat”

Effects Of Vertical Acceleration On Wrongness

Whenever someone sneers “Evidence-based medicine? You wouldn’t demand a double-blind placebo-controlled clinical trial of PARACHUTES, would you?” I feel a strong urge to use them as the control group in my double-blind parachute experiment.

Of course, deep down inside I know that this would be morally wrong. Groups need to be determined by random assignment.

In Favor of Niceness, Community, and Civilization

[Content warning: Discussion of social justice, discussion of violence, spoilers for Jacqueline Carey books.]

[Edit 10/25: This post was inspired by a debate with a friend of a friend on Facebook who has since become somewhat famous. I've renamed him here to "Andrew Cord" to protect his identity.]

I.

Andrew Cord criticizes me for my bold and controversial suggestion that maybe people should try to tell slightly fewer blatant hurtful lies:

> I just find it kind of darkly amusing and sad that the "rationalist community" loves "rationality is winning" so much as a tagline and yet are clearly not winning. And then complain about losing rather than changing their tactics to match those of people who are winning.
>
> Which is probably because if you *really* want to be the kind of person who wins you have to actually care about winning something, which means you have to have politics, which means you have to embrace "politics the mindkiller" and "politics is war and arguments are soldiers", and Scott would clearly rather spend the rest of his life losing than do this.
>
> That post [the one debunking false rape statistics] is exactly my problem with Scott. He seems to honestly think that it's a worthwhile use of his time, energy and mental effort to download evil people's evil worldviews

> into his mind and try to analytically debate them with statistics and cost-benefit analyses.
>
> He gets *mad* at people whom he detachedly intellectually agrees with but who are willing to back up their beliefs with war and fire rather than pussyfooting around with debate-team nonsense.
>
> It honestly makes me kind of sick. It is exactly the kind of thing that “social justice” activists like me *intend* to attack and “trigger” when we use “triggery” catchphrases about the mewling pusillanimity of privileged white allies.

In other words, if a fight is important to you, fight nasty. If that means lying, lie. If that means insults, insult. If that means silencing people, silence.

It always makes me happy when my ideological opponents come out and say eloquently and openly what I’ve always secretly suspected them of believing.

My natural instinct is to give some of the reasons why I think Andrew is wrong, starting with the history of the “noble lie” concept and moving on to some examples of why it didn’t work very well, and why it might not be expected to work so well in the future.

But in a way, that would be assuming the conclusion. I wouldn’t be showing respect for Andrew’s arguments. I wouldn’t be going halfway to meet them on their own terms.

The respectful way to rebut Andrew’s argument would be to spread malicious lies about Andrew to a couple of media outlets, fan the flames, and wait for them to destroy his reputation. Then if the stress ends up bursting an aneurysm in his brain, I can dance on his grave, singing:

> ♪ ♬ I won this debate in a very effective manner. Now you can’t argue in favor of nasty debate tactics any more ♬ ♪

I’m not going to do that, but if I *did* it’s unclear to me how Andrew could object. I mean, he thinks that sexism is detrimental to society, so spreading lies and destroying people is justified in order to stop it. I think that discourse based on mud-slinging and falsehoods is detrimental to society. Therefore…

II.

But really, all this talk of lying and spreading rumors about people is – what was Andrew’s terminology – “pussyfooting around with debate-team nonsense”. You know who got things done? The IRA. They didn’t agree with the British occupation of Northern Ireland and they weren’t afraid to let people know in that very special way only a nail-bomb shoved through your window at night can.

Why not assassinate prominent racist and sexist politicians and intellectuals? I won’t name names since that would be crossing a line, but I’m sure you can generate several of them who are sufficiently successful and charismatic that, if knocked off, there would not be an equally competent racist or sexist immediately available to replace them, and it would thus be a serious setback for the racism/sexism movement.

Other people can appeal to “the social contract” or “the general civilizational rule not to use violence”, but not Andrew:

> I think that whether or not I use certain weapons has zero impact on whether or not those weapons are used against me, and people who think they do are either

> appealing to a kind of vague Kantian morality that I think is invalid or a specific kind of “honor among foes” that I think does not exist.

And don’t give me that nonsense about the police. I’m sure a smart person like you can think of clever exciting new ways to commit the perfect murder. Unless you do not believe there will *ever* be an opportunity to defect unpunished, you need this sort of social contract to take you at least some of the way.

He continues:

> When Scott calls rhetorical tactics he dislikes “bullets” and denigrates them it actually hilariously plays right into this point…to be “pro-bullet” or “anti-bullet” is ridiculous. Bullets, as you say, are neutral. I am in favor of my side using bullets as best they can to destroy the enemy’s ability to use bullets.
>
> In a war, a real war, a war for survival, you use all the weapons in your arsenal because you assume the enemy will use all the weapons in theirs. Because you understand that it IS a war.

There are a lot of things I am tempted to say to this.

Like “And that is why the United States immediately nukes every country it goes to war with.”

Or “And that is why the Geneva Convention was so obviously impossible that no one even bothered to attend the conference”.

Or “And that is why, to this very day, we solve every international disagreement through total war.”

Or “And that is why Martin Luther King was immediately reduced to a nonentity, and we remember the Weathermen as the sole people responsible for the success of the civil rights movement”

But I think what I am *actually* going to say is that, for the love of God, if you like bullets so much, stop using them as a metaphor for ‘spreading false statistics’ and go buy a gun.

III.

So let’s derive why violence is not in fact The One True Best Way To Solve All Our Problems. You can get most of this from Hobbes, but this blog post will be shorter.

Suppose I am a radical Catholic who believes all Protestants deserve to die, and therefore go around killing Protestants. So far, so good.

Unfortunately, there might be some radical Protestants around who believe all Catholics deserve to die. If there weren’t before, there probably are now. So they go around killing Catholics, we’re both unhappy and/or dead, our economy tanks, hundreds of innocent people end up as collateral damage, and our country goes down the toilet.

So we make an agreement: I won’t kill any more Catholics, you don’t kill any more Protestants. The specific Irish example was called the Good Friday Agreement and the general case is called “civilization”.

So then I try to destroy the hated Protestants using the government. I go around trying to pass laws banning Protestant worship and preventing people from condemning Catholicism.

Unfortunately, maybe the next government in power is a Protestant government, and they pass laws banning Catholic worship and preventing people from condemning Protestantism. No one can securely practice their own religion, no one can learn about other religions, people are constantly plotting civil war, academic freedom is severely curtailed, and once again the country goes down the toilet.

So again we make an agreement. I won't use the apparatus of government against Protestantism, you don't use the apparatus of government against Catholicism. The specific American example is the First Amendment and the general case is called "liberalism", or to be dramatic about it, "civilization 2.0"

Every case in which both sides agree to lay down their weapons and be nice to each other has corresponded to spectacular gains by both sides and a new era of human flourishing.

"Wait a second, no!" someone yells. "I see where you're going with this. You're going to say that agreeing not to spread malicious lies about each other would also be a civilized and beneficial system. Like maybe the Protestants could stop saying that the Catholics worshipped the Devil, and the Catholics could stop saying the Protestants hate the Virgin Mary, and they could both relax the whole thing about the Jews baking the blood of Christian children into their matzah.

"But your two examples were about contracts written on paper and enforced by the government. So maybe a 'no malicious lies' amendment to the Constitution would work if it were enforceable, *which it isn't*, but just *asking* people to stop spreading malicious lies is doomed from the start. The Jews will no doubt spread lies against *us*, so if we stop spreading lies about them, all we're doing is abandoning an effective weapon against a religion I personally know to be heathenish! Rationalists should win, so put the blood libel on the front page of every newspaper!"

Or, as Andrew puts it:

> Whether or not I use certain weapons has zero impact on whether or not those weapons are used against me, and people who think they do are either appealing to a kind of vague Kantian morality that I think is invalid or a specific kind of "honor among foes" that I think does not exist.

So let's talk about how beneficial game-theoretic equilibria can come to exist even in the absence of centralized enforcers. I know of two main ways: reciprocal communitarianism, and divine grace.

Reciprocal communitarianism is probably how altruism evolved. Some mammal started running TIT-FOR-TAT, the program where you cooperate with anyone whom you expect to cooperate with you. Gradually you form a successful community of cooperators. The defectors either join your community and agree to play by your rules or get outcompeted.

Divine grace is more complicated. I was tempted to call it "spontaneous order" until I remembered the rationalist proverb that if you don't understand something, you need to call it by a term that reminds you that don't understand it or else you'll think you've explained it when you've just named it.

But consider the following: I am a pro-choice atheist. When I lived in Ireland, one of my friends was a pro-life Christian. I thought she was responsible for the unnecessary suffering of millions of women. She thought I was responsible for killing millions of babies. And yet she invited me over to her house for dinner without poisoning the food. And I ate it, and thanked her, and sent her a nice card, without smashing all her china.

Please try not to be insufficiently surprised by this. Every time a Republican and a Democrat break bread together with good will, it is a miracle. It is an equilibrium as beneficial as civilization or liberalism, which developed in the total absence of any central enforcing authority.

When you look for these equilibria, there are lots and lots. Andrew says there is no “honor among foes”, but if you read the *Iliad* or any other account of ancient warfare, there is practically nothing *but* honor among foes, and it wasn’t generated by some sort of Homeric version of the Geneva Convention, it just sort of happened. During World War I, the English and Germans spontaneously got out of their trenches and celebrated Christmas together with each other, and on the sidelines Andrew was shouting “No! Stop celebrating Christmas! Quick, shoot them before they shoot you!” but they didn’t listen.

All I will say in way of explaining these miraculous equilibria is that they seem to have something to do with inheriting a cultural norm and not screwing it up. Punishing the occasional defector seems to be a big part of not screwing it up. How exactly that cultural norm came to be is less clear to me, but it might have something to do with the reasons why an entire civilization’s bureaucrats may suddenly turn 100% honest at the same time. I’m pretty sure I’m supposed to say the words timeless decision theory around this point too, and perhaps bring up the kind of Platonic contract that I have written about previously.

I think most of our useful social norms exist through a combination of divine grace and reciprocal communitarianism. To some degree they arise spontaneously and are preserved by the honor system. To another degree, they are stronger or weaker in different groups, and the groups that enforce them are so much more pleasant than the groups that don't that people are willing to go along.

The norm against malicious lies follows this pattern. Politicians lie, but not *too much*. Take the top story on Politifact Fact Check today. Some Republican claimed his supposedly-maverick Democratic opponent actually voted with Obama's economic policies 97 percent of the time. Fact Check explains that the statistic used was actually for *all* votes, not just economic votes, and that members of Congress typically have to have >90% agreement with their president because of the way partisan politics work. So it's a lie, and is properly listed as one. But it's a lie based on slightly misinterpreting a real statistic. He didn't just totally make up a number. He didn't even just make up something else, like "My opponent personally helped design most of Obama's legislation".

Even the guy in the fake rape statistics post lied less than he *possibly could have*. He got his fake numbers by conflating rapes per sex act with rapes per lifetime, and it's really hard for me to imagine someone doing that by anything resembling accident. But he couldn't bring himself to go the extra step and just totally make up numbers with no grounding whatsoever. And part of me wonders: why not? If you're going to use numbers you know are false to destroy people, why is it better to derive the numbers through a formula you know is incorrect, than to just skip the math and make the numbers up in the first place? "The FBI has determined that no false rape claims have ever been submitted, my source is an obscure report they published, when your local library doesn't have it you will just accept that libraries can't have all books, and suspect nothing."

This would have been a *more believable* claim than the one he made. Because he showed his work, it was easy for me to debunk it. If he had just said it was in some obscure report, I wouldn't have gone through the trouble. So why did he go the harder route?

People *know* lying is wrong. They know if they lied they would be punished. More ~~spontaneous social order~~ miraculous divine grace. And so they want to hedge their bets, be able to say "Well, I didn't exactly *lie*, per se."

And this is good! We *want* to make it politically unacceptable to have people say that Jews bake the blood of Christian children into their matzah. Now we build on that success. We start hounding around the edges of currently acceptable lies. "Okay, you didn't *literally* make up your statistics, but you still lied, and you still should be cast out from the community of people who have reasonable discussions and never trusted by anyone again."

It might not totally succeed in making a new norm against this kind of thing. But at least it will prevent other people from seeing their success, taking heart, and having the number of lies which are socially acceptable gradually *advance*.

So much for protecting what we have been given by divine grace. But there is also reciprocal communitarianism to think of.

I seek out people who signal that they want to discuss things honestly and rationally. Then I try to discuss things honestly and rationally with those people. I try to concentrate as much of my social interaction there as possible.

So far this project is going pretty well. My friends are nice, my romantic relationships are low-drama, my debates are productive and I am learning so, so much.

And people think “Hm, I could hang out at 4Chan and be called a ‘fag’. Or I could hang out at Slate Star Codex and discuss things rationally and learn a lot. And if I want to be allowed in, all I have to do is not be an intellectually dishonest jerk.”

And so our community grows. And all over the world, the mysterious divine forces favoring honest and kind equilibria gain a little bit more power over the mysterious divine forces favoring lying and malicious equilibria.

Andrew thinks I am trying to fight all the evils of the world, and doing so in a stupid way. But sometimes I just want to cultivate my garden.

IV.

Andrew goes on to complain:

> Scott…seems to [dispassionately debate] evil people’s evil worldviews …with statistics and cost-benefit analyses.
>
> He gets *mad* at people whom he detachedly intellectually agrees with but who are willing to back up their beliefs with war and fire rather than pussyfooting around with debate-team nonsense.

I accept this criticism as an accurate description of what I do.

Compare to the following two critiques: “The Catholic Church wastes so much energy getting upset about heretics who believe *mostly* the same things as they do, when there are literally *millions* of Hindus over in India who don’t believe in Catholicism *at all*! What dumb priorities!”

Or “How could Joseph McCarthy get angry about a couple of people who *might* have been Communists in the US movie industry, when over in Moscow there were *thousands* of people who were openly *super* Communist *all the time*?”

There might be foot-long giant centipedes in the Amazon, but I am a lot more worried about boll weevils in my walled garden.

Creationists lie. Homeopaths lie. Anti-vaxxers lie. This is part of the Great Circle of Life. It is not necessary to call out every lie by a creationist, because the sort of person who is still listening to creationists is not the sort of person who is likely to be moved by call-outs. There is a role for *organized* action against creationists, like preventing them from getting their opinions taught in schools, but the marginal blog post “debunking” a creationist on something is a waste of time. Everybody who wants to discuss things rationally has already formed a walled garden and locked the creationists outside of it.

Anti-Semites fight nasty. The Ku Klux Klan fights nasty. Neo-Nazis fight nasty. We dismiss them with equanimity, in accordance with the ancient proverb: “Haters gonna hate”. There is a role for *organized* opposition to these groups, like making sure they can’t actually terrorize anyone, but the marginal blog post condemning Nazism is a waste of time.

Everybody who wants to discuss things charitably and compassionately has already formed a walled garden and locked the Nazis outside of it.

People who want to discuss things rationally and charitably have not yet looked over the false rape statistics article and decided to lock Charles Clymer out of their walled garden.

He is not a heathen, he is a heretic. He is not a foreigner, he is a traitor. He comes in talking all liberalism and statistics, and then he betrays the signals he has just sent. He is not just some guy who defects in the Prisoner's Dilemma. He is the guy who defects while wearing the "I COOPERATE IN PRISONERS DILEMMAS" t-shirt.

What really, *really* bothered me wasn't Clymer at all: it was that *rationalists* were taking him seriously. Smart people, kind people! I even said so in my article. Boll weevils in our beautiful walled garden!

Why am I always harping on feminism? I feel like we've got a good thing going, we've ratified our Platonic contract to be intellectually honest and charitable to each other, we are going about perma-cooperating in the Prisoner's Dilemma and reaping gains from trade.

And then someone says "Except that of course regardless of all that I reserve the right to still use lies and insults and harassment and dark epistemology to spread feminism". Sometimes they do this explicitly, like Andrew did. Other times they use a more nuanced argument like "Surely you didn't think the same rules against lies and insults and harassment should apply to oppressed and privileged people, did you?" And other times they don't say anything, but just show their true colors by reblogging an awful article with false statistics.

(and still other times they don’t do any of this and they are wonderful people whom I am glad to know)

But then someone else says “Well, if they get their exception, I deserve my exception,” and then someone else says “Well, if those two get exceptions, I’m out”, and *you have no idea how difficult it is to successfully renegotiate the terms of a timeless Platonic contract that doesn’t literally exist*.

No! I am Exception Nazi! NO EXCEPTION FOR YOU! Civilization didn’t conquer the world by forbidding you to murder your enemies *unless* they are actually unrighteous in which case go ahead and kill them all. Liberals didn’t give their lives in the battle against tyranny to end discrimination against all religions *except* Jansenism because seriously fuck Jansenists. Here we have built our Schelling fence and here we are defending it to the bitter end.

V.

Contrary to how it may appear, I am not trying to doom feminism.

Feminists like to mock the naivete of anyone who says that classical liberalism would suffice to satisfy feminist demands. And true, you cannot simply assume Adam Smith and derive Andrea Dworkin. Not being an asshole to women and not writing laws declaring them officially inferior are both good starts, but it not enough if there’s still cultural baggage and entrenched gender norms.

But here I am, defending this principle – kind of a supercharged version of liberalism – of “It is not okay to use lies, insults, and harassment against people, even if it would help you enforce your preferred social norms.”

And I notice that this gets us a heck of a lot closer to feminism than Andrew’s principle of “Go ahead and use lies, insults, and harassment if they are effective ways to enforce your preferred social norms.”

Feminists are very concerned about slut-shaming, where people harass women who have too much premarital sex. They point out that this is very hurtful to women, that men might underestimate the amount of hurt it causes women, and that the standard-classical-liberal solution of removing relevant government oppression does nothing. All excellent points.

But one assumes the harassers think that women having premarital sex is detrimental to society. So they apply their general principle: “I should use lies, insults, and harassment to enforce my preferred social norms.”

But this is the principle Andrew is asserting, against myself and liberalism.

Feminists think that women should be free from fear of rape, and that, if raped, no one should be able to excuse themselves with “well, she was asking for it”.

But this is the same anti-violence principle as saying that the IRA shouldn’t throw nail-bombs through people’s windows or that, nail bombs having been thrown, the IRA can’t use as an excuse “Yeah, well, they were complicit with the evil British occupation, they deserved it.” Again, I feel like I’m defending this principle a whole lot more strongly and consistently than Andrew is.

Feminists are, shall we say, divided about transgender people, but let’s allow that the correct solution is to respect their rights.

When I was young and stupid, I used to believe that transgender was really, really dumb. That they were looking for attention or making it up or something along those lines.

Luckily, since I was a classical liberal, my reaction to this mistake was – to not bother them, and to get very very angry at people who did bother them. I got upset with people trying to fire Phil Robertson for being homophobic even though homophobia is stupid. You better bet I also got upset with people trying to fire transgender people back when I thought transgender was stupid.

And then I grew older and wiser and learned – hey, transgender isn't stupid at all, they have very important reasons for what they do and go through and I was atrociously wrong. And I said a mea culpa.

But it could have been worse. I didn't like transgender people, and so I *left them alone while still standing up for their rights*. My epistemic structure *failed gracefully*. For anyone who's not overconfident, and so who expects massive epistemic failure on a variety of important issues all the time, graceful failure modes are a *really important feature* for an epistemic structure to have.

God only knows what Andrew would have done, if through bad luck he had accidentally gotten it into his head that transgender people are bad. From his own words, we know he wouldn't be "pussyfooting around with debate-team nonsense".

I admit there are many feminist principles that cannot be derived from, or are even opposed to my own liberal principles. For example, some feminists have suggested that pornography be banned because it increases the likelihood of violence against women. Others suggest that research into gender differences should be banned, or at least we should stigmatize and harass the researchers, because any discoveries made might lend aid and comfort to sexists.

To the first, I would point out that there is now strong evidence that pornography, especially violent objectifying pornography, very significantly decreases violence against women. I would ask them whether they're happy that we did the nice liberal thing and waited until all the evidence came in so we could discuss it rationally, rather than immediately moving to harass and silence anyone taking the pro-pornography side.

And to the second, well, we have a genuine disagreement. But I wonder whether they would prefer to discuss that disagreement reasonably, or whether we should both try to harass and destroy the other until one or both of us are too damaged to continue the struggle.

And if feminists agree to have that reasonable discussion, but lose, I would tell them that they get a consolation prize. Having joined liberal society, they can be sure that no matter what those researchers find, I and all of their new liberal-society buddies will fight tooth and nail against anyone who uses any tiny differences those researchers find to challenge the central liberal belief that everyone of every gender has basic human dignity. Any victory for me is going to be a victory for feminists as well; maybe not a perfect victory, but a heck of a lot better than what they have right now.

VI.

I am not trying to fight all the evils of the world. I am just trying to cultivate my garden.

And you argue: “But isn’t that selfish and oppressive and privileged? Isn’t that confining everyone outside of your walled garden to racism and sexism and nastiness?

But there is a famous comic which demonstrates what can happen to certain walled gardens.

Why yes, it does sound like I’m making the unshakeable assumption that liberalism always wins, doesn’t it? That people who voluntarily relinquish certain forms of barbarism will be able to gradually expand their territory against the hordes outside, instead of immediately being conquered by their less scrupulous neighbors? And it looks like Andrew isn’t going to let that assumption pass.

He writes:

> The *whole history* of why the institutional Left in our society is a party of toothless, spineless, gutless losers and they’ve spent two generations doing nothing but lose.

One is reminded of the old joke about the Nazi papers. The rabbi catches an old Jewish man reading the Nazi newspaper and demands to know how he could look at such garbage. The man answers “When I read our Jewish newpapers, the news is so depressing – oppression, death, genocide! But here, everything is great! We control the banks, we control the media. Why, just yesterday they said we had a plan to kick the Gentiles out of Germany entirely!”

And I have two thoughts about this.

First, it argues that “Evil people are doing evil things, so we are justified in using any weapons we want to stop them, no matter how nasty” suffers from a certain flaw. Everyone believes their enemies are evil people doing evil things. If you’re a Nazi, you are just defending yourself, in a very proportionate manner, against the Vast Jewish Conspiracy To Destroy All Germans.

But second, before taking Andrew’s words for how disastrously liberalism is doing, we should check the newspapers put out by liberalism’s enemies. Here’s Mencius Moldbug:

> Cthulhu may swim slowly. But he only swims left. Isn’t that interesting?
>
> In each of the following conflicts in Anglo-American history, you see a victory of left over right: the English Civil War, the so-called “Glorious Revolution,” the American Revolution, the American Civil War, World War I, and World War II. Clearly, if you want to be on the winning team, you want to start on the left side of the field.
>
> Where is the John Birch Society, now? What about the NAACP? Cthulhu swims left, and left, and left. There are a few brief periods of true reaction in American history – the post-Reconstruction era or Redemption, the Return to Normalcy of Harding, and a couple of others. But they are unusual and feeble compared to the great leftward shift. McCarthyism is especially noticeable as such. And you’ll note that McCarthy didn’t exactly win.
>
> In the history of American democracy, if you take the mainstream political position (Overton Window, if you care) at time T1, and place it on the map at a later time T2, T1 is always way to the right, near the fringe or outside it. So, for instance, if you take the average segregationist voter of 1963 and let him vote in the

> 2008 election, he will be way out on the wacky right wing. Cthulhu has passed him by.

I’ve got to say Mencius makes a much more convincing argument than Andrew does.

Robert Frost says “A liberal is a man too broad-minded to take his own side in a quarrel”. Ha ha ha.

And yet, outside of Saudi Arabia you’ll have a hard time finding a country that doesn’t at least pay lip service to liberal ideas. Stranger still, many of those then go on to *actually implement them*, either voluntarily or after succumbing to strange pressures they don’t understand. In particular, the history of the past few hundred years in the United States has been a history of decreasing censorship and increasing tolerance.

Contra the Reactionaries, feminism isn’t an exception to that, it’s a casualty of it. 1970s feminists were saying that all women need to rise up and smash the patriarchy, possibly with literal smashing-implements. 2010s feminists are saying that if some women want to be housewives, that’s great and their own choice because in a liberal society everyone should be free to pursue their own self-actualization.

And that has *corresponded to* spectacular successes of the specific causes liberals like to push, like feminism, civil rights, gay marriage, et cetera, et cetera, et cetera.

A liberal is a man too broad-minded to take his own side in a quarrel. And yet when liberals enter quarrels, they always win. Isn’t that interesting?

VII.

Andrew thinks that liberals who voluntarily relinquish any form of fighting back are just ignoring perfectly effective weapons. I’ll provide the quote:

> In a war, a real war, a war for survival, you use all the weapons in your arsenal because you assume the enemy will use all the weapons in theirs. Because you understand that it IS a war… Any energy spent mentally debating how, in a perfect world run by a Lawful Neutral Cosmic Arbiter that will never exist, we could settle wars without bullets is energy you could better spend down at the range improving your marksmanship… I am amazed that the “rationalist community” finds it to still be so opaque.

Let me name some other people who mysteriously managed to miss this perfectly obvious point.

The early Christian Church had the slogan “resist not evil” (Matthew 5:39), and indeed, their idea of Burning The Fucking System To The Ground was to go unprotestingly to martyrdom while publicly forgiving their executioners. They were up against the Roman Empire, possibly the most effective military machine in history, ruled by some of the cruelest men who have ever lived. By Andrew’s reckoning, this should have been the biggest smackdown in the entire history of smackdowns.

And it kind of was. Just not the way most people expected.

Mahatma Gandhi said “Non-violence is the greatest force at the disposal of mankind. It is mightier than the mightiest weapon of destruction devised by the ingenuity of man.” Another guy who fought one of the largest empires ever to exist and won resoundingly. And he was pretty insistent on truth too: “Non-violence and truth are inseparable and presuppose one another.”

Also skilled at missing the obvious: Martin Luther King. Desmond Tutu. Aung San Suu Kyi. Nelson Mandela was smart and effective at the beginning of his career, but fell into a pattern of missing the obvious when he was older. Maybe it was Alzheimers.

Of course, there are counterexamples. Jews who nonviolently resisted the Nazis didn’t have a very good track record. You need a certain pre-existing level of civilization for liberalism to be a good idea, and a certain pre-existing level of liberalism for supercharged liberalism where you don’t spread malicious lies and harass other people to be a good idea. You need to have pre-existing community norms in place before trying to summon mysterious beneficial equilibria.

So perhaps I am being too harsh on Andrew, to contrast him with Aung San Suu Kyi and her ilk. After all, all Aung San Suu Kyi had to do was fight the Burmese junta, a cabal of incredibly brutal military dictators who killed several thousand people, tortured anyone who protested against them, and sent eight hundred thousand people they just didn’t like to forced labor camps. Andrew has to deal with *people on Facebook who aren’t as feminist as he is*. Clearly this requires much stronger measures!

VIII.

Liberalism does not conquer by fire and sword. Liberalism conquers by communities of people who agree to play by the rules, slowly growing until eventually an equilibrium is disturbed. Its battle cry is not “Death to the unbelievers!” but “If you’re nice, you can join our cuddle pile!”

But some people, through lack of imagination, fail to find this battle cry sufficiently fear-inspiring.

I hate to invoke fictional evidence, especially since perhaps Andrew’s strongest point is that the real world doesn’t work like fiction. But these people need to read Jacqueline Carey’s *Kushiel’s Avatar*.

Elua is the god of kindness and flowers and free love. All the other gods are gods of blood and fire, and Elua is just like “Love as thou wilt” and “All knowlege is worth having”. He is the patron deity of exactly the kind of sickeningly sweet namby-pamby charitable liberalism that Andrew is complaining about.

And there is a certain commonality to a lot of the Kushiel books, where some tyrant or sorcerer thinks that a god of flowers and free love will be a pushover, and starts harassing his followers. And the only Eluite who shows up to stop him is Phèdre nó Delaunay, and the tyrant thinks “Ha! A woman, who doesn’t even know how to fight, doesn’t have any magic! What a wuss!”

But here is an important rule about dealing with fantasy book characters.

If you ever piss off Sauron, you should probably find the Ring of Power and take it to Mount Doom.

If you ever get piss off Voldemort, you should probably start looking for Horcruxes.

If you ever piss off Phèdre nó Delaunay, *run and never stop running*.

Elua is the god of flowers and free love and he is terrifying. If you oppose him, there will not be enough left of you to bury, and it will not matter because there will not be enough left of your city to bury you in.

And Jacqueline Carey and Mencius Moldbug are both wiser than Andrew Cord.

Carey portrays liberalism as Elua, a terrifying unspeakable Elder God who is fundamentally good.

Moldbug portrays liberalism as Cthulhu, a terrifying unspeakable Elder God who is fundamentally evil.

But Andrew? He *doesn't even seem to realize liberalism is a terrifying unspeakable Elder God at all*. It's like, *what?*

Andrew is the poor shmuck who is sitting there saying “Ha ha, a god who doesn't even control any hell-monsters or command his worshippers to become killing machines. What a weakling! This is going to be so easy!”

And you want to scream: “THERE IS ONLY ONE WAY THIS CAN POSSIBLY END AND IT INVOLVES YOU BEING EATEN BY YOUR OWN LEGIONS OF DEMONAICALLY CONTROLLED ANTS”

(uh, spoilers)

Suggest A Comments Policy

Several people have now mentioned they don’t like the comments section of my blog very much anymore.

In my response to Arthur deep in the comments, I wrote:

> I feel like you want to frame this as you being willing to draw fences and me not being so willing. But we are both drawing fences.
>
> My fence consists of the people who say HATE AND KILL THE INHUMAN DEMONS on the one side, and on the other, the people who donate to charity and respect others and express their opinion – whatever it may be, even if it’s problematic – kindly and compassionately.
>
> Your fence consists of you shouting HATE AND KILL THE INHUMAN DEMONS on one side, and other people who are also shouting HATE AND KILL THE INHUMAN DEMONS on the other side, only it’s different inhuman demons who need to be hated.
>
> Both of us successfully trap the KKK and Hitler on the outside of our fence. But I get to have a lot more allies than you do – allies against the actually bad people – and you have to put up with some pretty creepy friends.

This reminded me that I am philosophically totally justified in having fences, and that so far I have been pretty lax about putting them up and so have ended up with, as I put it, some creepy friends.

Trying to think up a comment policy has been an interesting case study in how it is much interesting to *speculate about* virtuous institutions for a large society, than it is to actually *design* such institutions for a blog with a few hundred people and nothing at stake.

So I open the problem up to the floor.

Desired properties a comment policy should have:

1. It should mean that *moderately* easily scandalized people can still enjoy reading comments on this blog without being exposed to HEY LOOK AT ME I CAN STATE CONTROVERSIAL THINGS! CONTROVERSY CONTROVERSY CONTROVERSY! This should not interfere with ability to present relevant well-supported politely-phrased appropriately-qualified controversial theories when a discussion warrants.

2. It should not require CONSTANT VIGILANCE from me. To be honest, I read the comments threads for each post for about a day or two, then give up and move on to the next thing.

3. It should not require too much subjectivity or “just ban all of the bad people” without including a rigorous definition of “bad people”.

4. Any technological elements should either be easy to implement in WordPress, or, if hard to implement in WordPress, include an offer to do it yourself.

5. End result is polite, productive discussion.

Even if you don’t have a General Theory Of Commenting, I would like to hear vague emotive statements about what different people want. Then I’ll think about all of it and come to some conclusion in a week or two.

Fix Science In Half An Hour

You’ve probably heard about the crisis of replication in psychology. The problem is that replication is an unglamorous business; researchers would much rather do the sexier work of pushing forward knowledge with new results.

So we need to make replications more glamorous.

I propose a reality TV show, *Replication Lab!*, where every week they try to replicate one of the most famous experiments from the past few years.

It starts with the host explaining the experiment, maybe an interview with a very distinguished elderly professor who talks about how confident he is that his results will hold up. The techs chat with each other as they construct the experimental setup about how they’re doing and how their date last night went and how they’re going to avoid the problems that confounded the original study.

Suspense builds as we see the participants come in. Some human interest stories. He agreed to participate because they offered $30, which he’s going to use to buy a present that will win back his estranged daughter’s love. She joined because she’s right on the border of failing her psych class and needs the extra credit to save her dream of becoming the first person in her family to graduate college.

The experiment itself. The suspense is unbearable. We get a running commentary as everything proceeds. Oh man, look how harsh that guy is being on his Milgram Obedience Experiment, can you believe he would do that? That girl in the control condition seems to be running through her

Stroop task at lightning speed – how do you think that’s going to affect our results, kindly-looking bearded scientist attached to the show?

After a tension-building commercial break, we get the results. Everyone is huddled around a computer as the statistician makes the final mouse click, and…oh no, p = .30! Total failure to replicate!

The scene cuts to the distinguished elderly professor’s face as he sees his great discovery going down the toilet. “How do you feel right now?” asks the host, and the professor sputters “I…I’m sure time will vindicate me! I know it!” and then he runs off the set, crying. Our host turns to the kindly-looking bearded scientist attached to the show. “Tell me the truth,” she says “Do you think Dr. Zuckerman’s career is ruined?” “I can’t imagine it wouldn’t be,” says the bearded scientist, shaking his head sadly.

I feel like Mythbusters has probably pretty much exhausted our cultural stock of urban legends by now and could be profitably recruited for this project. I would also accept “Welcome to *Replication Lab!* With your host, John Ioannidis!”

Unteachable Things Hard To Teach, Study Suggests

From Journal of the American Medical Association: About 500 doctors (and a handful of nurses) in training participated in a randomized controlled trial on communication skills. Half received thirty hours of intensive state-of-the-art communication skills training; the other half didn't. Then they all saw a bunch of patients with really serious diseases, who rated the quality of the care they received. Doctors who got the training were rated no better than doctors who hadn't, and in fact did significantly *worse* on the secondary outcome of how depressed their patients became about their illnesses. So even very very good communication skills training seems to have either zero or negative effect.

I would like to note for the record that I totally called this one.

To be fair, there is a large body of contradictory research. And if someone challenges me in the comments, I'll be happy to go over why I trust this study more than some of the others.

But you know who agrees with me here? The New York Times. Their analysis is very, very close to my own heart:

> In response to the growing recognition that effective communication with patients is a basic competency of our profession, and that doctors often have inadequate training in it, medical schools and hospitals have invested substantial resources over the past decade to teaching communication skills. […]

> But can giving doctors a script for empathy actually make them more empathetic? Our patients know better. […]
>
> Perhaps those residents participating in the study were too early in their training to lead these sorts of difficult conversations and participation in the study emboldened them in inappropriate ways.
>
> It is also possible that, as we devote more time to teaching students and doctors effective communication techniques, we risk muting their authentic human voices, and instead of learning to connect they apply rote tools and scripts.
>
> No communications course will magically transform lifelong introverts to hand-holders and huggers. At the same time, we must ensure that we are not converting people who genuinely care about their patients into people who only sound as if they care. Having physicians sound like customer service representatives is not the goal.
>
> For those doctors who are emotionally challenged, communications courses can provide the basics of relating to other human beings in ways that, at the very least, won’t be offensive. But for the rest of us, we should take care to ensure the techniques and words we learn in such courses don’t end up creating a barrier to authentic human contact.

The doctors who wrote the NYT article (one of whom runs a leukaemia clinic! Talk about jobs that require communication skills!) go on to describe a horrendous communication skills class they attended, which somehow ended up being taught by a doctor who was legendary in the area for his total lack of communication skills and human empathy.

This, too, was close to my own heart. When I work with a new attending, I will very often check them out on ratemds.com or healthgrades.com to see what their patients think of them. These sites are not very valuable on the margin; you *do* sometimes get one guy who had to sit in the waiting room a little too long going on a personal campaign to destroy someone's reputation. But on the tails, if there's a doctor who is *universally hated by every single one of her patients*, this can be a strong warning sign.

And so when some of my attendings try to "correct" my communication style or give me helpful tips for "how to make patients like me", it sometimes takes everything I've got not to retort "I think I'm going to limit myself to taking advice from doctors with better than a 1.0/5 stars patient rating."

The doctor I'm working for now has a better than 1.0/5 stars patient rating. In fact, he has a 5.0/5 star patient rating. He has not tried to give me any advice on communication skills, which I am starting to notice is a common thread among doctors with excellent communication skills. But I have been trying to absorb what I can through osmosis.

Yesterday I watched him diagnose a woman with very advanced metastatic lung cancer – pretty much a death sentence. He showed the patient a CT image, told her she had cancer, and pointed out the location of the primary tumor in the right mainstem bronchus. Then he started talking about the anatomy of the bronchial tree.

The bronchial tree, for those of you not lucky enough to know, is the system of airways that gets oxygen into the lungs. It starts with the trachea, or “windpipe”, which divides at an area called the carina – around the level of the fifth thoracic vertabra – into the right and left mainstem bronchi. These bronchi then divide into smaller bronchi, which themselves divide into structures called “bronchioles”…

…and there is no point in continuing because by this point you are asleep. I do not blame you. No human being on Earth cares about the anatomy of the bronchial tree. I voluntarily went into medicine and even I have to make an extraordinary effort of will not to immediately change the subject to something less mind-numbing. If a communications skills tutor saw you discussing the anatomy of the bronchial tree *immediately* after diagnosing a patient with metastatic lung cancer, she would arrange to have you shot.

And I was thinking about this later, and I realized that what my attending was doing was providing white noise. If he had done the proper Communications Skills thing and said “And how does this make you feel?”, our patient would have been put on the spot. We would have been demanding some kind of coherent reaction to the diagnosis, and how in the world are you supposed to react to that five seconds after hearing about it? And if my attending had asked her if she needed a few minutes to collect her thoughts, she would have been socially obligated to say no, because it would have made her look weak and wasted a busy doctor’s time.

So instead he started talking about all this anatomical stuff, stuff that no human being could reasonably be expected to listen to, and it was his way of saying “Go on, tune me out, pull yourself together, and then when I’m done we can have a discussion about cancer and what to expect from it.”

And it worked. I'm not saying there aren't ninety-nine ways this could have gone horribly wrong, but in this case and for this patient, it worked. And he *knew* it was going to work because he'd been seeing this patient for twenty years.

And I have no particular moral to this anecdote, or thoughts about how it relates to teaching communication skills. I just thought it was interesting. But…

Economist James Miller recently gave the following advice to LW college students pondering lucrative earning-to-give careers:

> If you have a high IQ and are good at math go into finance. If you have a high IQ, strong social skills but are bad at math go into law. If you have a high IQ and a good memory but weak social skills, become a medical doctor.

There are a couple of different ways to have weak social skills. There is the way where you're scared of large groups and being the focus of attention. There is the way where you dress unfashionably and smell bad. There is my personal preferred way, which is to have no idea what to do at a party and end up sitting in the corner really awkwardly and not being invited back. But there's also the way where you consider human emotions an annoying inconvenience and prefer to just plow through them like a rhinoceros in a china shop.

If you are the last one, and you insist on going into medicine, do everyone a favor and become a pathologist, a radiologist, or a surgeon who works with a really good anesthesiologist.

Because communication skills classes won't save you.

Searching For One-Sided Tradeoffs

Suppose you are an admissions official for a moderately prestigious college, which is neither the best nor the worst in your state. Your job is to look over people's SAT scores, high school GPA, and essays on How I Overcame Adversity, and then decide whether or not to admit them to your college.

And suppose that you have a team of subordinates who make the really easy decisions for you. Auto-reject the losers who show up drunk to their interview and spell your institution's name as "collej" on their applications, pass the rest on to you.

Your job probably doesn't matter. Yes, there will be some very high quality candidates – the kids with straight As, perfect SATs, and stories about how they personally stopped the civil war in Lebanon despite being born without legs. But they will be using you as their safety school, and whether you accept them or not they will be going to Harvard and you will never see them. You will only be deciding among a small band of students – those too smart to get auto-rejected by your subordinates, but not smart enough to go to a school better than yours.

Given that kids who are good at everything and kids who are bad at everything are equally unlikely to be your target population, your job reduces to choosing what tradeoffs to take. Do you want kids with great SAT scores but terrible grades, kids with great grades but terrible SATs, or kids with mediocre grades and test scores alike? How about kids with terrible grades and terrible SATs, but they're really really attractive and good at sports?

Even here your job won’t matter *too* much. Your counterparts at Harvard will presumably be smart people who have a pretty good idea of how important test scores and grades are in terms of the Intangible Qualities That Make You Good At College. If a new study comes out showing that SAT scores determine your future but grades are meaningless, that study will make you want to shift to a high-SAT-low-grade model, but it will equally increase the high-SAT-low-grade kids’ ability to get into Harvard, meaning that you will, to use an economics metaphor, have to buy SAT scores with grades at a lower exchange rate.

So basically no matter how competent you are as an admissions official, all of the kids entering your college will be about equally “good”.

There is a fun legend I heard in a stats class – I don’t know if it’s true – of a psychology professor who got very excited about her new theory that the brain traded off verbal and mathematical intelligence – being better at one made you worse at the other. She got SAT Math and SAT Verbal scores from her students and found it supported her theory. A friend of hers did a replication at his college and found support for the the theory there as well.

But larger scale testing disconfirmed the theory. What the professors working off college samples were finding was that all of the kids in their college were equally “good”, in a general sense, so excellence in any quality implied a tradeoff in other qualities. Suppose the professor worked at a mid-tier college – students with SATs much less than 1200 couldn’t get in; students with SATs much more than 1200 could and did go to better schools instead. Then all her students would have SATs around 1200. Which meant a student with an SAT Verbal of 700 would have an SAT Math of 500, a student with an SAT Math of 800 would have an

SAT Verbal of 400, and boom, there's your "trade-off of verbal and mathematical intelligence". Obviously the tradeoff wouldn't be perfect, since there's random noise and since students are also trading off less obvious qualities like attractiveness, wealth, social skills, athleticism, musical talent, and diligence. But it would be more than enough for her to find her correlation if she was looking for it.

This suggests some odd strategies if we're looking for particular college students. If we want to find the dumbest students in a particular college, we might look at the football star – not because football stars are naturally dumb, but because plausibly a student who couldn't get in on his wits alone might make it in on the promise of helping the college team. If we want to find the smartest student in a particular college, we might look for someone on a scholarship – because perhaps she would otherwise be at Harvard, but was made less attractive to the Ivy League by her inability to pay them any money.

It also implies some weird strategies for admission officers. How do you maximize student quality when in theory all your job allows you to do is make tradeoffs between different subcharacteristics among students of the same quality? Aside from just hoping the occasional Harvard-caliber student accidentally stumbles into your office, I suggest three potential techniques: insider trading, bias compensation, and comparative advantage.

Insider trading is where you're just plain smarter than everyone else. Maybe you're a brilliant psychologist who has invented a test that invariably reveals students' true potential. You can find kids with terrible grades and terrible SAT scores who will nevertheless shine. If you happen to luck into this position, you've got it made.

Bias-compensation is where you try to see if other colleges have biases that you can exploit. Sometimes this is simple and profitable. If Harvard is controlled by anti-Semites and auto-rejects all Jews, then you have a free shot to get Jews with 800 SAT Math, 800 SAT Verbal, *and* amazing football talent (though good luck finding Jews with amazing football talent). Once again, if you happen to luck into your competitors being stupid, you've got it made.

Sometimes it's not that easy, and you have to kind of spin someone else's preferences as "bias" when they might secretly have some wisdom behind them. For example, it is no doubt true that college admissions officials are influenced by student charm and social skills. So if you want, you can probably get smarter students if you go for the really really unpleasant students whom everyone dislikes as soon as they open their mouths. You can then declare "success" when your college gets a disproportionate number of academic awards, but unless you are a remarkably single-minded academic-award-maximizer, you may find that your college is kind of horrible now and other schools had pretty good reasons for rejecting these people.

Comparative advantage is where you decide you are going to have radically different priorities than anybody else. Maybe you want to be The Math School and become known for the quality of your math geniuses. So you nab all the students with 800 SAT Math and 400 SAT Verbal and then advertise the heck out of your students' mathematical acumen. There's also another sort of comparative advantage, where if you have a great sign language interpretation program and Harvard doesn't, you can advertise to deaf kids who maybe Harvard doesn't want because they can't develop their talents effectively.

So let’s generalize from college to the sorts of choices that we actually face.

In one of the classics of the Less Wrong Sequences, Eliezer argues that policy debates should not appear one-sided. College students are pre-selected for “if they were worse they couldn’t get in, if they were better they’d get in somewhere else.” Political debates are pre-selected for “if it were a stupider idea no one would support it, if it were a better idea everyone would unanimously agree to do it.” We never debate legalizing murder, and we never debate banning glasses. The things we debate are pre-selected to be in a certain range of policy quality.

(to give three examples: no one debates banning sunglasses, that is obviously stupid. No one debates banning murder, that is so obviously a good idea that it encounters no objections. People *do* debate raising the minimum wage, because it has some plausible advantages and some plausible disadvantages. We might be able to squeeze one or two extra utils out of getting the minimum-wage question exactly right, but it’s unlikely to matter terribly much.)

So there’s some argument to be made that, like the admissions officer, our decisions aren’t too important. I don’t think things are quite that depressing. But, like the admissions officer, we will have to be clever if we want to figure out how to escape the seemingly iron law of tradeoffs.

I recently heard a Catholic guy condemn the “culture of death”, which by his telling consisted of abortion, stem cells, euthanasia, and capital punishment. I’m in favor of three of those things, and I avoid a perfect four-out-of-four only on a technicality: I can’t support capital punishment until it gets better at sparing the innocent and maybe becomes more cost-effective.

My near-unaninimous support for culture-of-death issues seems unlikely to be a coincidence, and indeed it isn't. I have a deep philosophical disagreement with the Catholics here – they think life is a terminal value, I think life is only valuable insofar as it gives certain goods associated with living.

This means from my point of view, the Catholics have a bias in their trade-off arithmetic. They are the equivalent of the anti-Semitic Harvard leadership, who have given me this great gift of trade-off-free students. Just as learning the Harvard leadership is anti-Semitic makes me suddenly want to accept all Jews as a tradeoff-free utility gain, learning that a large portion of the electorate is biased against death means that certain death-related policies can be tradeoff-free utility gains to me.

I will add one more political example. I've previously proposed sticking lithium in the water supply as an intervention to promote psychiatric health. People are super creeped out by this – and in fact, so am I, a little bit. But this is encouraging! If people's response was "actually, we have proof that these quantities of lithium hurt cardiac health" we'd be faced with a useless tradeoff – X psychiatric health against Y cardiac health – and so a policy we'd be squeezing a couple measly utils out of depending on which way the tradeoff went. But if their response is "I see no particular downside, but I am very creeped out by it", then this is like learning Harvard is anti-Semitic – an explanation for why other people haven't gobbled up a possible advantage, and a neon sign pointing out potential tradeoff-free gains for you.

We can also use this framework to evaluate life hacks.

Life hacks are touted as “little-known techniques you can use to improve your life”. There are two ways something can fail to be a life hack – either it becomes universally known, or it fails to improve anything. These form a pre-selection kinda like a college selecting students of a certain quality, or a country debating issues of a certain quality. If an intervention was obviously great, then either you’d already do it (think “sleeping at night” or “working at a job to earn money”) – or you would at least feel guilty for not doing so (think “diet and exercise”). If an intervention was useless, no one would call it a life hack (think “hitting yourself on the head with a baseball bat every day”). Life hacks are the things that are sort of in between, where there seem to be some benefits, and also some costs in terms of time and energy and money, and you’re not sure if they’re worth looking into or not.

If you want to do better than trade off your time and energy for the occasional small benefit, you need a theory of why that might be possible.

Every life hacker wants to be an insider trader – someone who is able to outperform competitors with more resources by being a little savvier about biology and psychology. And probably some are. But unless you are the first scientist to discover a new supplement, or the first psychologist to discover a new technique, your trades aren’t that insider and you’ll eventually have to explain why no one else has adopted them.

And most life hackers pay lip service to comparative advantage: “Everyone has their own individual biology and their own set of problems, what works for you may not work for everyone else.” This is pretty plausible. It suggests the reason the whole world isn't adopting life hacks is because there's a very high startup cost, where you've got to sort through a hundred different things and find the ones that work for you and the ones that don't, and nobody can do this for you, and if you're not very smart you'll get it wrong.

Another form of comparative advantage is willpower. Maybe no one else is doing weight lifting because they don't have the determination to go to the gym three times a week. This is a fine theory – plausible even – but it's interesting to see how many of the people who confidently assert their own comparative advantage then buy a gym membership but end up not having the determination to go three times a week.

But in terms of using a tradeoff-based framework to help inform the decision of what lifehacks to try, it seems most promising to consider opportunities for bias compensation.

Like insider trading, bias compensation is claimed a lot more often than I think it can be supported. The polyphasic sleep crowd, for example, tell you that you can increase your free time per day – and all you need to do is stick to a very strict schedule, be very tired for a long time while you're working out kinks, and abandon all hope of a social life or flexible schedule. To me this seems a lot like the admission official with the bright idea of admitting unpleasant low-social-skills kids: it sounds good if you're only thinking about the most easily quantifiable results, but when you actually try it you tend to regret it very quickly.

Can we find anything more promising? I think that people are unnecessarily pessimistic about nootropics because they are scared of taking drugs. Fear of taking drugs is an excellent and rational fear to have, but if you happen to lack it that fear, or you have enough comparative advantage in pharmacological knowledge / research ability that you can justifiably be less afraid of taking drugs than everyone else, then this starts to look like the lithium-water example: getting free utility by abandoning your sense of creeped-out-ness.

But if you're going to force me to give you an example of something I actually did differently because of thinking about tradeoffs, I'll have to go with "try bacopa".

Bacopa is a memory-enhancing drug that performs very well in studies. But it's rarely used and it only got a middling ranking on my survey. I think this has something to do with having to take it for three months before it has any effect. Talk about trivial inconvenience. Most people don't want to bother, so it remains largely uninvestigated, and the nonsuperabundance of bacopa use stands explained without resorting to it being a bad drug or having other tradeoffs we really don't want. So using it – if you can stand the three month waiting period – has a higher-than-otherwise-expected likelihood of being free utility.

Me? I tried to start taking bacopa, but it gave me terrible diarrhea and I had to stop. Another tradeoff! That should just increase its expected psychological benefits!

Last, something on the lighter side: an article going around the Internet recently claims houses on streets with mildly rude names (example: "Slag Lane") apparently cost £84,000 less than control houses on more properly named streets (the article does not give me enough information to rule out hypotheses like poor people being more willing to give their streets rude names). If you don't care about what your street name is called, this might be another potential free trade-off – buy a house on Slag Lane and save $100,000+. Or buy a house that's supposed to be haunted if you don't believe in ghosts. Or buy a house near a prison with a very low escape rate because you trust the statistics and other people don’t.

[CORRECTION] Mountain View Meetup 3/4

[a previous version of this post said 3/3, but upon closer investigation it is more likely 3/4]

Speaking of one-sided tradeoffs, for the next two weeks I’ve traded Detroit for the San Francisco Bay Area. I’ve already got plans to meet a few people. If you want to meet me and we haven’t already planned something, the easiest time would be the rationalist meetup on Tuesday 3/4 at 8 PM. The address is:

Quixey
278 Castro St
Mountain View, CA 94041

(I am working on getting transportation from Richmond, so if any of you are driving from there let me know)

The Comment Policy Is “Victorian Sufi Buddha Lite”

There is an ancient Sufi saying beloved of the Buddha, which like a surprising number of ancient Sufi sayings beloved of the Buddha, originates from a book of preachy Victorian poetry. It goes:

> Before you speak, let your words pass through three gates; At the first gate, ask yourself, is is true? At the second gate ask, is it necessary? At the third gate ask, is it kind?

Slate Star Codex has lower standards than either ancient Sufis *or* preachy Victorians, and so we only require you to pass at least *two* of those three gates.

If you make a comment here, it had better be either true and necessary, true and kind, or kind and necessary.

Recognizing that nobody can be totally sure what is or isn’t true, if you want to say something that *might* not be true – anything controversial, speculative, or highly opinionated – then you had better make sure it is both kind and necessary. Kind, in that you don’t rush to insult people who disagree with you. Necessary in that it’s on topic, and not only contributes something to the discussion but contributes *more* to the discussion than it’s likely to take away through starting a fight.

Nobody can be kind all the time, but if you are going to be angry or sarcastic, what you say had better be both true and necessary. You had better be delivering a very well-deserved smackdown against someone who is uncontroversially and obviously wrong, in a way you can back up with universally agreed-upon statistics. I feel like I tried this here and though a lot of people disagreed with my tone, not one person accused me of getting the math wrong. That's the standard I'm holding commenters to as well. And it had better be necessary, in that you are quashing a false opinion which is doing real damage and which is so persistent that you don't think any more measured refutation would be effective.

Annnnnnd sometimes you might want to share something that's not especially relevant, not the most important thing in the world – but if you do that it had better be both true and kind. No random interjection of toxic opinions that are going to cause World War III. No unprovoked attacks.

Threats, doxxing, most things people would call "slurs", et cetera fail this test as neither kind nor necessary. You people are smart and don't need me to explain this further.

I feel like these standards are pretty lax. In fact, they probably permit most spam – this spambot saying "this is a wonderful piece of writing" is both true and kind – so I will inelegantly add a kludge that spam is also unacceptable (I have it on good authority that this was in the original Sufi saying used by the Buddha as well). Remember that before you worry this is too unduly restrictive.

I don't always read comment threads, and when I do my brain automatically screens out comments it expects to be low-quality, and even on the rare cases where I see them I usually figure that even if I don't like them other people might. What I'm saying is I'm a terrible moderator and will never catch bad comments if left to my own devices. So I have added a REPORT button to comments. Right now it's large and ugly and if somebody wants to play around with the theme so that it's less obtrusive I'll give them the access to do so. Please use the REPORT button if you think a comment violates the above policy, and please err on the side of using it too much rather than too little.

I will review reported comments. If a comment gets only one report, I will have a strong bias towards *not* reacting, unless it is so obviously bad that something must be done immediately. If a comment gets three or more reports, I will have a strong bias towards reacting, unless it's obvious to me that the reports are malicious or the comment can't *possibly* be construed as violating the policy.

If I need to react to a comment, I will delete it only if it is *dangerous* to leave it up (ie comment contains people's personal information, comment contains strong basilisk, comment is so offensive that Internet mob would use my leaving it up as an excuse to attack me). Otherwise, I will leave it up but post in large red letters below it "COMMENT VIOLATED POLICY FOR [REASON]. POSTER BANNED FOR [TIME]", for approximately the same reason all those people in *Game of Thrones* leave bloody heads on spikes in front of their castles. For first offenses, bans will be for about a day. For second offenses, bans will be for about a month. For third offenses, bans will be for about forever.

Many people have told me I have too many low-quality commenters, especially on the Reactionary side of the aisle (this is not persecution; even some Reactionaries have told me this). So I will be trying to cultivate a very itchy trigger finger for a while. If you want to help, please cultivate a very itchy REPORT-button-pressing finger. You can also volunteer to be a moderator, but only if I know you pretty well in real life and have some reason to trust you and I get some kind of moderator system up and running.

Thank you for your cooperation.

Do Life Hacks Ever Reach Fixation?

In Searching For One-Sided Tradeoffs, I argued that people's "life hacks" probably occupy a restricted range. If the life hack had *nothing* going for it, it would never become popular and you wouldn't hear about it. If the hack had *everything* going for it, you would have heard *more* about it. If there were something that really doubled energy levels and increased IQ and cured shyness and made you lose ten pounds, the front page of the New York Times would be "Man Discovers Amazing Life Hack", and it would be all over the medical journals and the talk shows and so on. It wouldn't have to be pushed by some guy with a blog who says it "changed his life".

…except that then I tried looking for examples of such and came up blank. The example I ended up giving, "sleeping at night", was a biological imperative that was never really "discovered", per se.

Compare to genetics. If there's a mutation that gives even a small benefit, it predictably reaches fixation in the population (where every single organism has it) after a certain number of generations.

Compare also to other kinds of ideas, like technology. When a new technology (let's say the cell phone) is invented, it starts with a group of early adopters. As the technology gradually gets better and cheaper, and people notice that cell phone users have a big advantage over non-users, new people buy cell phones. Eventually it reaches the point where the cell-phone-less are at a big disadvantage, and even the grumpy old holdouts like myself are forced to purchase them. Even if we never reach *literal* fixation because of the Amish, the indigent, etc, there's still a point in which having a cell phone seems to become the default state.

The same is true in the economy. One business gets a bright idea, like outsourcing to China or something, they get rich and outcompete their rivals, their rivals pick up on the idea, and eventually businesses-that-don't-outsource-to-China gets reduced to a weird niche market.

This should be able to work with lifehacks. Whether it's students trying to get the best grade, workers trying to be most productive, or suitors trying to appear most attractive, people compete with each other *all the time*. If there were some meme that consistently offered its users an advantage in productivity or energy or even mood, it ought to reach fixation as surely as new technologies or business practices.

And I can't think of any that have.

Some possible explanations:

1. There are no exceptionally good life hacks. The human body and brain are optimized really really well, or else have really really strong tendencies to return to equilibrium after a disruption.

2. Life hacks, as a category, have some characteristic that makes fixation an unreasonable goal for them. Maybe there is so much variation in people that no lifehack can ever improve more than a small percent of them. This seems like a less bleak version of (1) – the stuff everyone has in common is optimized really really well, but there are some individualized flaws you can pick off on a person-by-person basis.

3. Life hacks as a category didn't exist until kind of recently, or it if did they weren't as good as modern life hacks. Even

though there are some great ones out there now, they haven't existed long enough to achieve fixation.

4. All the genuinely useful life hacks take work, and people are really bad at doing work, so nothing that takes work can ever achieve fixation. The level of work it takes to understand a cell phone or computer doesn't count; these life hacks take more work, or different kinds of work.

5. Some life hacks have totally reached fixation and I'm just too stupid to think of them. Or – life hacks that reach fixation become so entrenched that it's very hard to think of them as lifehacks any more. Compare the genetics student who says "No mutations have ever reached fixation in the human population, and I know this because most of the people I see aren't mutants."

The last explanation seems most promising, which means I should probably look harder for fixated life hacks.

There are some things I want to exclude right away. New technologies like the cell phone can reach fixation, but I don't think I'd want to call them life hacks; I'd rather limit the term to non-medical interventions or at least technologies specifically related to health and productivity. Certain ideas like religion have reached fixation in their populations, and it would be fascinating to think of in what senses those are life hacks, but I don't think that's where we're going here. I'm looking specifically at things that act directly to raise energy levels, intelligence, social skills, or organizational ability.

I will grudgingly accept three-ring binders, to-do lists, calendars, and filing cabinets as *sort of* examples – even though I don't use a calendar or to-do list and it doesn't seem to have left me unable to compete with the rest of humanity, and even though these all fall into a sort of general “keep organized by writing things down and sorting them” category.

I will grudgingly accept backpacks, briefcases, and the like, even though “things that hold other things” seems to be a pretty basic human invention and if we have to go back to the Paleolithic before getting a genuinely useful life hack we are doing very poorly indeed. This might also be a piece of technology which escapes that category only through the cheap trick of going so far back that it doesn’t seem like a technology anymore.

I will grudgingly accept “diet and exercise”, since even people who are bad at diet and exercise probably eat better and exercise more than they would if they were unaware that diet and exercise were things they should do. But I don’t know if this was ever really “discovered” or if it got a lot of help from a biological imperative.

I will grudgingly accept “take a deep breath and count to ten in order to not get angry”, since everyone seems to know about it.

But none of these seem to fall into classical life hack categories like “thing that a man with slick hair teaches a class on, telling you that it will change your life”, or “thing that you can buy at the Sharper Image”. And they all seem pretty old. Cell phones took like fifteen years to achieve fixation; how come for life hacks we have to look all the way back to whichever caveman first realized you could carry tools in a sack made of animal skin?

EDIT: @mjdominus on Twitter proposes caffeine. That sounds right to me.

Book Review: The Perfect Health Diet

After I dissed saturated fat on Facebook a friend told me to read “The Perfect Health Diet”, which usually doesn’t work. But he also offered me a free copy, which does. If you can’t get a free copy, you can get a lot of the material from their website.

Let’s get the obvious out of the way first. The title, dust jacket, and introduction are all doing their best to make it sound like a sketchy diet fad book. Even its proponents admit it does a very poor job at not sounding like a sketchy diet fad book. I don’t know how much importance to give this. Maybe any diet advice, including very scientifically based and sober advice, inevitably ends up looking like a sketchy fad book because that is the most effective way to market it? Let’s just raise the red flag count to “one” and move on.

The best word for my opinions on the content would be “mixed”.

Let’s start with the good. It clearly tries to be science-based. It mostly succeeds. It is evolutionarily informed without being silly about it. It ties a lot of research together into a consistent and conclusive whole.

The overall diet it recommends is…pretty good, actually. Sort of similar to the paleo diet, with its emphasis on unprocessed whole meats and vegetables, and an acceptance bordering on fetish for saturated fat. But it adds the concept of a “safe starch” – where the classic paleo diet has a zero tolerance policy toward carbohydrates, the Perfect Health Diet allows certain carbs like rice, potatoes, and lots of weird things like “taro” and “sago”.

I really like this change. Hunter gatherers do eat certain roots and tubers, so it’s not like an evolutionary perspective has to exclude them entirely. And the Japanese remain some of the healthiest people in the world while eating rice by the bucketload, so anyone who tells you that rice is bad for you has got something like a hundred million contrary data points to overcome. Pre-contact Polynesians were also exceptionally healthy, so all the weird Polynesian starches they recommend have my seal of approval as well.

This makes their diet much easier than normal paleo diets. With a moderate amount of work, you can find a lot of the breadstuffs you like made with rice or potato or something rather than wheat – my local Asian food store has rice noodles that can *almost* pass for spaghetti to my unsophisticated palate. And since I have been trained not to feel full without some carbs in me, having a small plate of rice with a meal really helps.

I am *slightly* skeptical about admitting potatoes to the same pantheon as rice and Weird Polynesian Starches. This comes from my four years studying medicine in Ireland. The traditional Irish diet is beef, lamb, butter, cabbage, and potato, which is about as Perfect Health Diet compliant as it’s possible to get (yeah, okay, they do eat bread and drink beer, but other than that). The Irish can be *very* insistent on sticking to just these couple of things. I remember working at a rural GP office and staying at the home of an elderly Irish couple. I made a trip into town and brought back some spaghetti to make for dinner, and my hosts were impressed that I was making them “exotic foreign foods”. By which they meant *spaghetti*. This was *2010*.

Anyway, despite this rigid diet elderly Irish people seemed pretty unhealthy and cardiovascular-disease-ridden, maybe not much better than the US. If we’re sticking Polynesians and Japanese on one side, and Americans on the other, the Irish are over with us. I don't know if the Perfect Health people want to blame *all* of that on their bread and alcohol and whatever ungodly oils they fry stuff in, but I am going to view potatoes with suspicion until this gets cleared up.

The other thing about the book I really liked was their discussion of "toxins". Normally whenever any diet book discusses "toxins" I throw it aside with great force, because they mean "mysterious evil spirits in foods I don't like which can be banished only with juice cleanses." Perfect Health Diet, despite using the same language, singles out very specific chemicals and cites all the reasons which each chemical can be expected to be harmful.

The part that stuck with me most here was their section on omega-6 fats, which might be literally the Devil. I find their speculation on omega-6s and obesity intriguing, not to mention the connection to crime I have already blogged about. Stephan Guyenet seems to take this seriously as well, which is always a good sign. And it makes evolutionary-historical sense that this random stuff we've been adding to foods in massive quantities over the past couple of decades and not before is screwing us up.

Also wheat. So, so much wheat. I haven't had time to investigate this much so far, but a lot of smart people seem very worried about wheat and corn and all those other grains, and if one tenth of the darts they throw at them stick, we should probably start looking for a new staple food for Western civilization.

So much for the good. Let's move on to the bad.

This book has a *very* consistent pattern of making a strong claim, citing one study that supports their claim, and failing to mention that there have been many more much larger studies since then that have come to the opposite conclusion.

For example, they mentioned every single study to say breathlessly excited things about Vitamin D for the past fifty years, but failed to mention that all of them have since been disconfirmed by larger and more careful trials.

And it's not just Vitamin D. Vitamin K prevents prostate cancer! Vitamin K prevents liver cancer! Vitamin K reduces all kinds of cancer by 75%! Vitamin K heals cancer in people who already have the disease! A recommendation: take lots of Vitamin K!

But look at how the American Cancer Society describes these same trials:

> A large European epidemiologic study published in 2008 found higher risk of prostate cancer in men with low intake of some forms of vitamin K. But people who take in plenty of vitamin K usually get it from fruits and vegetables. This means that people who get lots of vitamin K from foods are more likely to take in many other vitamins and phytochemicals that can affect cancer risk. In some cases, vitamin K may be given credit for benefits that are actually due to other compounds that appear in the same foods. So studies like this, which only observe large groups of people, may not be very useful for looking at the effects of vitamin K intake.
>
> A small clinical trial from Japan suggested that vitamin K lowers the risk of developing liver cancer among women with cirrhosis due to hepatitis C. A later study also

> seemed to show some effect, but it was not statistically significant. Further study is needed.
>
> There have been some studies examining whether menadione (vitamin K3) can help overcome cancer cells’ resistance to certain types of chemotherapy drugs. Results in lab animals and cell cultures are mixed, but there is no evidence available of significant effects in humans yet.
>
> A small Phase I clinical trial in California tested different doses of intravenous vitamin K3 in people with advanced cancer. The patients did not improve. In that study, several patients also had allergic reactions, especially at higher doses.

Look. I know alternative medicine is hard to do. It’s also useful. Sometimes they come up with genuinely interesting stuff the mainstream has missed. So I’m not demanding that you not speculate about Vitamin K just because the mainstream medical community has interpreted the results as unimpressive. Go ahead and try to find reasons why the earlier successful trials were more believable than the later more pessimistic trials. But at least mention that everyone else disagrees with you about this. Don’t present it as some well-established thing which the mainstream community is just completely ignoring.

We repeat this exact same frickin’ pattern with just about every vitamin and mineral ever discovered. Selenium prevents gastric cancer! Selenium reduces tumor incidence! A randomized controlled trial of selenium in China decreased liver cancer! A trial of selenium in America decreased skin cancer! Selenium improves immune function! Selenium decreases all-cause mortality by a quarter! Selenium cuts cardiovascular mortality in half!

Selenium cures all diseases! Selenium will make you rich and famous! Women will want to date you! Men will want to be you! You will live in a beautiful house on the beach, with a huge yacht made entirely of selenium! You will sail it over seas of liquid selenium, to selenium islands where selenium-based life-forms play haunting music beautiful beyond anything you have ever imagined upon their selenium flutes. If you take enough selenium you will *never die*, yet if by chance you should, you will go to Selenium Heaven, which is to regular Heaven what a supernova is to a firefly.

Meanwhile, look at any respectable medical source like Cochrane Review, and the actual list of effects that they find for selenium is exactly zero. The process is the same as always. Some preliminary exciting studies come out. So we do much bigger, more rigorous studies and the effects go away. I did a really thorough evaluation of the selenium literature for my Quantified Health Prize entry and although I was totally happy to disagree with expert advice in several areas, in the case of selenium I found that the experts were exactly right and it is useless. Perfect Health Diet doesn't even *mention* the expert consensus or rigorous studies here.

This leads to this weird pattern where Perfect Health Diet manages to explain about 1500% of human mortality. That is, if you add up all of their claims ("Vitamin Q decreases mortality by 50%!" "Vitamin R decreases mortality by 75%!") then although I haven't exactly done the calculations my guess is they would add up to around 1500%. This isn't *quite* mathematically impossible – if there are fifteen vitamins, any single one of them prevents cancer completely, and you can just choose which one you want – then added together they would decrease cancer mortality by 1500%. But even though it's not mathematically impossible, it doesn't seem right to me and it seems much more likely that they use hugely inflated claims than that fifteen different interventions all cut mortality even more than any of the things we *know* cut mortality like exercise or quitting smoking.

I’ll skip over the T3/T4 thyroid thing and a few others that will be more of the same, and just spare a second to complain about folic acid: they say supplementing pregnancies with folic acid causes autism. This is not only not proven, but there’s much stronger evidence for the *opposite*: multiple good studies show folic acid supplemention decreases autism risk by as much as 40%. Perfect Health Diet’s only source for the opposite claim was a correlational study in, of all places, the Journal of Medical Hypotheses, which specifically exists to signal-boost claims that don’t really have any support but need further investigation.

Now, to be fair to PHD, they did briefly mention this counterexample. But their scare study was in **bold** and exaggerated, and the existence of a counterexample was in parentheses and had the word “slightly” appended to the front even though it was not slight at all.

So. Alternative medicine people? Please listen very closely. STOP SAYING THAT VITAL CHILDRENS’ HEALTH INTERVENTIONS CAUSE AUTISM. THIS. NEVER. HELPS.

Overall I thought their entire section on vitamin and mineral supplementation was a disaster area. This is a little worrying because I know more about vitamins and minerals than I do about macronutrients and diet in general, and the Gell-Mann Amnesia Effect argues that if the parts you know something about are bad, you should assume the parts you don’t know enough to criticize are as well.

I will, however, give praise where praise is due. PHD gets that multivitamins aren’t a great idea and does not recommend them. It gets that calcium is not a great idea and doesn’t recommend it either. And they have praise for iodine and lithium, which are the two minerals I went out on a limb to recommend in my QHP entry and which I really do think deserve more attention than they’re getting.

In conclusion, PHD is not a terrible book and you could do much worse as far as diet books go. It is written by smart people who are familiar with at least some of the literature. It is a good idea-generation source to learn about things you might not have otherwise known (did you know there is some evidence a chronic bacterial infection can be one cause of Alzheimers Disease?) and the diet it recommends will probably not kill you and in fact probably be good for you. If their stuff about wheat and omega-6 and the like pans out, and I would not be surprised if it does, it might be *really* good for you. I don’t have any complaints there.

(actually, I tried the diet they recommended for about two weeks, just to see what would happen, and I lost seven pounds without really trying. This is a sufficiently large number of pounds that I find this result implausible and am hesitant to pronounce a “This is what happens when you follow Perfect Health Diet”. But still, nice.)

But I also cannot wholeheartedly recommend it, especially for epistemically-minded people who want something correct as opposed to something full of nice diet advice. It has a lot of shoddy research and presents extremely controversial positions as if they are settled science. There are very few mistakes that will kill anyone (they even couched their folic acid advice with an incongruous yet welcome “but you should still take a little folic acid during pregnancy”) but a lot of mistakes that will leave you very confused or have you taking very silly amounts of vitamins.

This book might be suitable for someone who wants to try a paleo-inspired diet for purely practical reasons, but who needs some encouragement and inspiration through a discussion of the science behind it. It might also be good for someone who already knows a lot about health from a mainstream point of view, is sufficiently skeptical to catch the mistakes, and wants to see what sorts of stuff the paleo diet and alternative-medicine-which-is-smarter-and-more-sciency-than-homeopathy people are about so they can grab the useful insights and discard the rest.

I was originally sent this book because I dissed saturated fat, so I should tie up my review there – especially since it makes a good metaphor for the book in general. Perfect Health Diet gives a lot of good evolutionary reasons why saturated fat might be good for you, and reflects the growing consensus that yeah, we kind of went overboard turning it into demon-food and should take a step back from there – something most people don’t realize and that this book could be very helpful in teaching.

But then it goes on to say that the old story that saturated fats led to heart disease has been debunked, citing as the only discussion for this claim a link to a systematic review that did in fact find saturated fats not to correlate with heart disease. What the book doesn’t mention is that this study was funded by the National Dairy Council, or that nine out of eleven other systematic reviews that have investigated the issue *did* find a saturated fat/heart disease link, or that Cochrane Review, which is the gold standard for systematic reviews of medical evidence, concluded that on balance there is good evidence for a saturated fat-heart disease link. All they do is cite the increasingly-lonely one that found the conclusions they wanted, declare that to be the state of the evidence, and move on.

So that is PHD in a nutshell. Interesting points, broadly reasonable dietary advice, and total denial of the existence of any evidence that doesn’t support their claims.

The Slate Star Codex Political Spectrum Quiz

Directions: Select either (a) or (b) for each of the twelve questions.

1. Neo-Nazis are holding a demonstration in a small town, waving swastikas around and shouting about Hitler. They seem to be pretty peaceful so far, so the First Amendment says you probably can't get rid of them. However, their demonstration seems to be near a main street and it could be a minor inconvenience to the traffic trying to go through.

a) Allow the neo-Nazis to demonstrate unmolested
b) Break up the demonstration on the grounds of 'blocking traffic'

2. The anchor of a major news network donates lots of money to organizations fighting against gay marriage, and in his spare time he writes editorials arguing that homosexuals are weakening the moral fabric of the country. The news network decides they disagree with this kind of behavior and fire the anchor.

a) This is acceptable; the news network is acting within their rights and according to their principles
b) This is outrageous; people should be judged on the quality of their work and not their political beliefs

3. A conservative Southern state makes a very strict law restricting the cases in which a woman can get an abortion. The United States Supreme Court, which happens to be dominated by liberal justices, takes an activist stance and strikes down the law based on a conflict with a very broad interpretation of the Constitution.

a) This is acceptable; the Supreme Court should be zealous in its duty of defending constitutional rights against potentially extremist states
b) This is unfortunate; when something is not clearly prohibited, the Supreme Court should err in favor of states rights

4. The Dalai Lama comes to speak at a college town, and the town wants to hold a big celebration honoring his visit, give him the key to the city, and make a small monument commemorating the event. The town's sizeable Chinese minority population gets upset, saying they strongly believe the Dalai Lama is a bad person trying to break apart China.

a) The town should celebrate the Dalai Lama's visit however they feel appropriate; it's pretty cool and the Chinese are outvoted
b) The town should consider the sensibilities of the Chinese minority and keep the Dalai Lama's visit low-key and unofficial

5. The United Nations is trying to pass a resolution banning land mines and allocating some resources to clean up existing mines that pose a danger to civilians. North Korea is causing a fuss and refusing to support the resolution, and this is endangering its chances of passing. The proponents of the resolution come up with a scheme to exploit a loophole in UN procedures, holding the vote in secret at a time when the North Korean representative might not even be present.

a) Exploit this loophole to make sure the anti-mine resolution passes
b) Stick to normal procedure and try to pass the anti-mine resolution above board and through legitimate channels

6. A feminist group is annoyed at Disney for making too many movies where a dashing male hero saves a helpless princess. They publicly demand Disney stop making this kind of movie and instead make movies with strong female heroes. They will have prominent marches accusing everyone in Disney of sexism until they comply.

a) This feminist group is acting according to proper democratic means to change a problematic feature of the culture, and deserve praise
b) This feminist group is being obnoxious and bullying, and deserve condemnation

[DUE TO CIRCUMSTANCES WITHIN OUR CONTROL, QUESTIONS 7 THROUGH 10 HAVE BEEN CANCELLED.]

11. A human rights group is picketing the headquarters of Exxon Mobil for abusing their workers in Third World countries. Exxon Mobil executives feel very uncomfortable entering their HQ and say that the protesters are blocking the main entrance to the building. They want the protesters to go protest in a designated free speech zone a few miles away where it will have no effect on them.

a) Allow the human rights group to continue to stay near the building
b) Tell them to go protest far away

12. The principal of a private school is a member of Planned Parenthood and, off-duty, speaks out about contraception and the morning after pill. The board of the private school decides this is inappropriate given the school’s commitment to abstinence and moral education and asks the principal to stop these speaking engagements or step down from his position.

a) The school board is acting within its rights; they can insist on a principal who shares their values
b) The school board should back off; it's none of their business what he does in his free time

13. New York City passes one of the most restrictive gun control laws in the nation. A conservative-dominated Supreme Court strikes it down, taking a broad construction of the Second Amendment as prohibiting such regulations.

a) Agree with the Court's decision; the Court has a duty to protect freedom against state infringement
b) Disagree; the Court should respect New York City's right to legislate toward its own problems

14. A Southern town that produced a disproportionately high number of great Confederate generals and soldiers wants to erect a monument to its Civil War military heritage. The town's small African-American community objects, saying that these generals, however impressive, were fighting to defend slavery and an evil regime.

a) The town should build the monument, as desired by the majority of its citizens
b) The town should avoid building the monument to respect the wishes of its minority community

15. The Democrats are in the process of passing a new law that cuts corporate welfare to large oil companies. A small group of Republicans oppose this measure but seem to be outnumbered. One delegate realizes that if he filibusters for the next twenty-six hours, he can delay the bill long enough that it will fall off the schedule of this session of Congress and potentially be voted upon by a friendlier legislature next term.

a) The Republican legislator is fairly following procedure and therefore his filibuster is acceptable or even commendable
b) The Republican legislator is contradicting the obvious will of the chamber and is kind of an asshole

16. Sometime in the 1950s, The Society Of Patriotic Americans For A More Patriotic America notices that a lot of writers seem to lean left, and worries that novels are promoting Communist ideas (perhaps by portraying businessmen in a very negative light or having rebels and political agititators as heroes). They meet with the heads of various publishing companies and ask the companies to self-monitor their books to make sure they are suitably American. SOPAFAMPA threatens to tar them in the press as Commies if they refuse.

a) SOPAFAMPA did not threaten force and therefore did not violate the First Amendment. They are doing their patriotic duty as citizens to try to fight Communism wherever they finds it.
b) SOPAFAMPA's comments, even if not direct threats, will have a chilling effect on artistic expression and are tantamount to censorship.

Scoring: Each question 1 through 6 is paired with the question from 11 to 16 that shares the same terminal digit – ie 1 with 11, 2 with 12, etc.

Give yourself zero points for each pair of questions you answered differently – ie A for one but B for the other. Give yourself one point for each pair of questions you answered the same – ie both A or both B.

Score of 0 to 3: You are an Object-Level Thinker. You decide difficult cases by trying to find the solution that makes the side you like win and the side you dislike lose in that particular situation.

Score of 4 to 6: You are a Meta-Level Thinker. You decide difficult cases by trying to find general principles that can be applied evenhandedly regardless of which side you like or dislike.

Commentary: This test was inspired by this graph.

I tried to use pairs of questions that were similar enough that they both tested the same general principle, but not so similar that it would be immediately obvious what I was doing as soon as someone read the test. I realize that my insistence on the second criterion might have allowed me to slip with the first criterion. I do ask that if you criticize a pair of questions on the test, you propose an alternative pair that you think would make a good substitute and which also follow both criteria.

Other suggestions for add-on questions also welcome; the test would look a lot more streamlined if I could fill in 7-10.

Semite Times: The Bible In Palindromes

GENESIS

Dumb mud
“Madam, I’m Adam”
Eve damned Eden, mad Eve
Cain: a maniac

EXODUS

Egad, no bondage!
“Live not on evil!”

LEVITICUS

Repel a leper

NUMBERS & DEUTERONOMY

Are we not drawn onwards, we Jews, drawn onward to new era?

JUDGES, PROPHETS, KINGS, & WRITINGS

Now, sir, a war is won
Egad! A base life defiles a bad age.
[Deed]
[Deed]
[Tenet]
[Tenet]

THE GOSPELS

So – let’s use Jesus’ telos.
Dogma: I am God!
Deliverer re-reviled
“Abba, abba…”
Did I do, O God, did I as I said I’d do? Good, I did.
Deified

REVELATION

Won’t I panic in a pit now?

(most of these palindromes are not original to me, but I cannot find good attributions)

(I apologize for skipping the Pauline epistles, but I couldn’t find or invent good palindromes to describe them. But if we follow most scholars in rejecting 1 & 2 Timothy, Titus, and Ephesians, then modern *commentary* on the Pauline epistles could be “some men interpret nine memos”)

(I realize the title “Semite Times” might lead one to expect a more general history of the Jews beyond the Biblical period. I cannot do this concept justice, but a good chunk of modern Israeli history could be “bar an Arab”)

(Sounds Biblical as heck but unfortunately doesn’t seem to correspond to any actual Bible story: “So may Obadiah, even in Nineveh, aid a boy, Amos”.)

Five Years and One Week of Less Wrong

[content warning: will be boring for non-LWers]

I.

Last week was the fifth birthday of Less Wrong. I thought I remembered it was started March 11, but it seems to have been more like March 5. I was going to use that time to talk about it, but now I will just have to talk about it awkwardly five years and one week after it was started.

I wrote a post a while ago called Read History Of Philosophy Backwards. I theorized that as old ways of thinking got replaced by newer ways, eventually people forgot the old ways even existed or were even coherent positions people could hold. So instead of reading Hobbes to tell you that people can form governments for their common advantage – which you already know – read him to tell you that there was a time when no one believed this was true and governments were natural structures ordained by God.

It makes sense that over five hundred years, with births and deaths and so on, people would forget they ever held strange and incomprehensible positions. It's more surprising that it would happen within the course of a single person's philosophical development. But this is what I keep hearing from people in the Less Wrong community.

"I re-read the Sequences", they tell me, "and everything in them seems so obvious. But I have this intense memory of considering them revelatory at the time."

This is my memory as well. They look like extremely well-written, cleverly presented version of Philosophy 101. And yet I distinctly remember reading them *after* I had gotten a bachelor’s degree magna cum laude in Philosophy and being shocked and excited by them.

So I thought it would be an interesting project, suitable for the lofty milestone of five years plus one week, to go back and try to figure out how far we have progressed without noticing that we were progressing.

A while ago I wrote about how the idea that beliefs are probabilistic is *very* not intuitive and something most people never grasp. I don’t mean complicated controversial ideas about how you should be willing to bet on every belief or anything like that. I mean drop-dead basic “You are not 100% certain of everything you believe”.

Were we ever this stupid? Certainly I got in fights about “can you still be an atheist rather than an agnostic if you’re not sure that God doesn’t exist,” and although I took the correct side (yes, you can), it didn’t seem like oh my god you are such an idiot for even considering this an open question HOW DO YOU BELIEVE ANYTHING AT ALL WITH THAT MINDSET. I remember being very impressed by Robert Anton Wilson’s universal doubt, and not as disgusted as I should have been by people making arguments like “If there’s any chance at all a criminal might re-offend, we shouldn’t let them out of jail”. In all of these cases I was sort of groping at the right idea, but I didn’t have a framework for it, couldn’t put exactly what I meant into obviously-correct words.

But that’s Overcoming Bias stuff, Sequence stuff. What have we done on Less Wrong, in the past five years and one week?

II.

It was around the switch to Less Wrong that someone first brought up the word “akrasia” (I think it was me, but I’m not sure). I remember there being a time when I was very confused and scandalized by the idea that people might engage in actions other than those rationally entailed by their beliefs. This seems *really* silly now, but at the time I remember the response was mostly positive and people upvoted me a lot and said things like “Huh, yeah, I guess people might engage in actions other than those rationally entailed by their beliefs! Weird! We should worry about this more!” For a while, we were *really* confused about this, and a really popular solution (WHICH I ALWAYS HATED) was to try to imagine the mind as being made up of multiple agents trying to strike a bargain. Like, your conscious mind was an agent, your unconscious mind was an agent, your sex drive was an agent, and so on. Ciphergoth was the first person to help us get out of this by bringing up hyperbolic discounting (there was a time Less Wrong didn’t know about hyperbolic discounting!) After that we started thinking more in terms of non-goal-directed systems. I remember when someone (Richard Kennaway?) first posted a lot of oversold stuff about Perceptual Control Theory, one of which was that it was a system that acted purposefully without modeling a goal, and I – and a lot of other people – commented that that was poppycock and not even possible. Later Anna posted Humans Are Not Automatically Strategic and I posted The Blue-Minimizing Robot, where we both started realizing the importance of non-goal-directed behavior in human affairs. This freed us from having to talk about “bargaining with unconscious agents” and allowed us to become more interested in things like behaviorism, even though I think later we became too willing to accept behaviorist just-so stories.

I remember it was around the beginning of Less Wrong when I first realized decision theory was a thing that existed. This was mostly my personal ignorance, but given that I read nearly every comment on Overcoming Bias and Less Wrong it must have been pretty obscure to the rest of the community as well to escape my attention for so long. I remember making a post that was very much about decision theory even though I didn't know it, and then getting in a long argument with Eliezer that either of us would now be able to resolve in three seconds by saying "Scott, you're rounding this off to Counterfactual Mugging; Eliezer, you're rounding this off to Smoking Lesion, agree on where you're sticking the locus of decision and the apparent different between your viewpoints will disappear." But Counterfactual Mugging didn't even appear as a concept until after this time! Eliezer didn't discuss Timeless Decision Theory until six months after Less Wrong started, and I'm trying to imagine how our moral discussions must have taken place without it ("imagine Kant, only rigorous"). We didn't get a really good explanation of different decision theories presented to everyone until three years after LW started.

There was a time before we had Eliezer Yudkowsky Facts, and that time must have been very sad to live in.

There are a couple of terms I coined – not because I'm especially smart, but because I talk to a lot of Less Wrongers and am especially sensitive to when they are using interesting ideas but haven't yet named them or discussed them with anyone else – that I now can't imagine living without. The idea of meta-contrarianism – or at least of signaling hierarchies – is one of them, and I find it especially useful in understanding some aspects of social class – and more awareness of the nature of class seems to be something else the community has gotten better at.

Likewise, this idea of trivial inconveniences having outsized effects has been very helpful to me, and got expanded into discussion of trivial fears, the possibly-grammatical-or-not trivial impetuses, the sort of related concept of an ugh field, and most important, this idea of fearing the twinge of starting. All of these make the internal experience of motivation make a lot more sense.

It wasn't until well into the Less Wrong era that our community started to become aware of the problems with the scientific process. This wasn't because we were behind the times but because the field was quite new; Ioannides didn't publish his landmark paper until 2005, and it languished in specialized circles until the Atlantic picked it up in 2010. But as early as December 2009, Allan Crossman working off a comment of Eliezer's wrote Parapsychology: The Control Group For Science. This pointed out that the frequent ability of parapsychology experiments to discover "telepathy" or "psionics" may not just be about them being especially dishonest, but also demonstrate a fundamental bias of the scientific system towards spurious positive results. The first I heard about Ioannides and medical replicability was a Less Wrong article from Nancy Lebovitz which then got further explored by gwern and Luke. I have since come out against the strongest interpretations of their claims, but it's hard to imagine that as recently as five to ten years ago no one was really talking about the problems with medical studies or had good evidence anything was wrong with them.

I don't know if there's a good history of the efficient charity movement, but it seems to be pretty recent. GiveWell was founded in 2007. Giving What We Can was founded in 2009 by Toby Ord, a frequent commenter on the Less Wrong Sequences. 80000 Hours was founded in 2011 by William MacAskill and Benjamin Todd, both of whom are frequent

Less Wrong commenters. I don’t want to overestimate the effect we’ve had on the efficient charity community – all of these people got into LW from EC rather than vice versa and we were not super important influences on any of them. And I don’t think we’ve ever been the super-cutting-edge. But I think we helped spread some ideas and give better philosophical grounding to others. I remember Money – The Unit of Caring and Purchase Fuzzies and Utilons Separately as having a big impact on me. A lot of people have told me that my 2010 LW post explaining efficient charity was important to them, and I’ve gotten requests to republish it in efficient charity books and manifestos. And in terms of just really neat philosophical framework-grounding, my most recent discovery is Katja Grace’s In Praise Of Pretending To Really Try, which neatly solves what was previously an irritating conceptual gap. On a purely practice level, after the past five years of exploring the concept, 30% of Less Wrongers now consider themselves effective altruists and among the lot of them donate over a million dollars to charity *per year*.

It continues to puzzle me that there was a time when I didn’t know what a Schelling point was. I imagine myself just sort of wandering through life, not having any idea what was going on or why. I’m pretty sure I corrected this well before I joined Less Wrong, but it wasn’t until I pieced together some information from a couple of Vladimir_M comments that I coined the term “Schelling fence”, which was in itself a miniature version of this kind of revelation. I thought for a while people were insufficiently impressed with how important this was, which makes me happy that people are starting to use it more. In terms of barrier-related metaphors, there’s also the Chesterton Fence, which has been around for eighty years but which I almost never hear mentioned outside the LWosphere. I feel like knowing about these two things has improved my ability to intelligently discuss politics dramatically.

A while ago I got an unusually weird complaint about my blog – someone didn’t like that I used the word “meme” a lot when memetics was an “unproven theory”. I objected that it was less of a theory than an extended metaphor. Some of the work done at Less Wrong has either tried to make it more of a theory, or at least extend and flesh out the metaphor. My favorite of these posts is Phil Goetz’ Reason As Memetic Immune Disorder. This seems to have permeated the culture enough that when I refer to my tendency to get freaked out and angry about ideas I myself support as “a memetic autoimmune disorder” everyone seems to understand what I’m talking about.

We’ve also moved in some interesting directions on friendships and relationships. My mind boggles to remember that for several years Less Wrong was not associated with polyamory, and that when I first met Alicorn she was mildly *against* it. Aside from a few very early adopters like Patri and Tilia, I don’t think it was really talked about in the community until Alicorn wrote Polyhacking in 2011. Two and a half years later, 15% of Less Wrongers consider themselves poly – a small minority, but still way more than the general population – and I think we’ve become pretty good at developing social norms of dealing with that. For example, just two months ago, Brienne discussed her idea of Tell Culture, and I have serious reservations with it that I’ve been meaning to get around to discussing, I agree that this is the sort of direction we should be thinking in and that this is the sort of “invent new and better ways of interacting with people” that makes me excited to be part of a community trying them.

III.

I’ll end with something that recently encouraged me a lot. Sometimes I talk to Will Newsome, or Steve Rayhawk, or Jennifer RM, or people like that in the general category of “we all know they are very smart but they have no ability to communicate their insights to others”. They say inscrutable things, and I nod and pretend to understand because it’s less painful than asking them to explain and sitting through an equally inscrutable explanation. And recently, the things that Will and Steve and Jennifer were saying a couple of years ago have started making perfect sense to me. The things they’re saying *now* still sound like nonsense, but now I can be optimistic that in a few years I’ll pick up those too.

I find this really exciting. It suggests there’s this *path* to be progressed down, that intellectual change isn’t just a random walk. Some people are further down the path than I am, and report there are actual places to get to that sound very exciting. And other people are around the same place I am, and still other people are lagging behind me. But when I look back at where we were five years ago, it’s so far back that none of us can even *see* it anymore, so far back that it’s not until I trawl the archives that realize how many things there used to be that we didn’t know.

I don’t think Less Wrong ever reached the insight-a-minute pace of the Sequences. But it’s been pretty enlightening. And over the course of five years and one week, that has really, *really* added up.

Can It Be Wrong To Crystallize Patterns?

In my last post I mentioned my experience of reading over some essays I found really enlightening years ago, finding them on re-examination to be correct but boring, and half doubting that I had ever not-known the things within. In the comments many people said they shared that experience, either with the Less Wrong sequences or something else.

This never happens with other kinds of knowledge. I remember learning the capital of Cuba in my sixth grade Geography class. It is very clear in my memory that before I did this, I didn't know the capital of Cuba. I can easily conceive of forgetting the capital of Cuba, and indeed if I went long enough without thinking about it I would *expect* that to happen.

Compare this factual knowledge of the Cuban capital to the idea of mysterious answers to mysterious questions.

The capital of Cuba is a known unknown. Before I learned it was Havana, I knew Cuba must have a capital and that this was a hole in my knowledge. It's hard to imagine what it would mean for me to understand that I needed a concept like "mysterious answers" without, by that fact itself, *having* the concept.

Information about Cuba's capital can be wrong – a vandalized Wikipedia article might tell me it's Cuba City. I'm not sure what it would mean for the idea of Mysterious Answers to be wrong. The impression I get from that post is that it's saying "Here is a useful concept, take it or leave it". If you didn't like the post, it would be more because you thought very few theories or ideas were really Mysterious

Answers, or because Mysterious Answers didn’t have enough extra explanatory power to be worth your time to learn – not because the concept was wrong *per se*. But I think a more likely reaction is to see the examples and immediately think “Yes, this is a very common failure mode. I can easily think of a lot of times I’ve seen it used. Now I will be able to spot and recognize it more quickly”

Teaching the capital of Cuba imparts new information. Teaching about Mysterious Answers doesn’t, at least for a strict definition of information. Maybe most readers already know all of the particular examples in that post. All they hadn’t done was detect a pattern and give a name to it.

A while ago, a leftist commenter asked me to stop using the term “political correctness”. Without getting into whether or not I should comply, I was struck by the sound strategy behind the request. Without the words “political correctness”, an entire class of things becomes much harder to think about. There are whole debates that would stop happening if the words “political correctness” got removed from the vocabulary. That’s pretty strange. If we removed the word “President” from the vocabulary, people would just start talking about “that guy in the White House” or “the executive branch” or something. But striking out “political correctness” would be a genuine game-changer.

“Political correctness” is also an attempt to detect a pattern (for the purpose of not politicizing this post, I will stay agnostic whether it is real or imagined) and give a name to it. And I think if we stopped using the term, people would stop detecting the pattern and their views about a lot of things would change.

Giving a pattern a name crystallizes it, let people hold it in their minds for a longer period, make it easier to think about and discuss.

This is neat because it lets you formulate a *meta-level response* to the pattern. If you realize that “God” is sometimes a Mysterious Answer, and “elan vital” is sometimes a Mysterious Answer, then when you’re tempted to use something you *like* as a Mysterious Answer (“How did the eye form?” “Evolution!”) you’re more likely to recognize it, admit it’s a problem, and apply a general policy of not being satisfied with Mysterious Answers.

And when you encounter it again, it’s much easier to dismiss in a single mental motion. A lot of the problem with debate is time and number-of-mental-steps: the amount of time and energy normal people give to debating things isn’t enough to get past the preliminary outer defenses of bad ideas, and in the rare cases it is, by the time the last defenses have been breached the first ones have already started regenerating. Turning long series of steps into a single convenient package decreases the amount of stuff you need to keep in your mental workspace.

I will now dump a long list of concepts that I think work in more or less this way: “trivial inconveniences”, “Schelling points”, “spoon theory“, “ad hominem fallacy”, “introversion/extraversion”, “social justice”, “ask culture”, “global (versus local) maxima/minima”, “growth mindset”, “logical uncertainty”, “crackpot”, “cult”, “meme”, “self-reference”, “utility”, “warm fuzzies”, “Third World”, “efficient charity”, “entryism”, “witch hunt”, “deontology”, “romantic love”…

…it’s going to be easier to list things that *don’t* work like this, isn’t it?

“Witch hunt” and “romantic love” seem like especially interesting examples. In olden days, if someone was having a moral panic and decided to unfairly persecute a group of people for something they probably weren’t doing, you would have to come up with some complicated way of explaining this to the public and convincing them that this is a plausible thing that might happen (you can’t borrow my wording; “moral panic” is almost as artificial a crystallized pattern as “witch hunt” is). Good luck getting the public to hold still for that. Now you can just say “They’re holding a witch hunt!” and everyone will understand your objection. I think the existence of the concept “witch hunt” lets everybody coordinate to develop solutions to problems like “How can we best prevent witch hunts from happening?” and really does make witch hunts significantly more difficult.

“Romantic love” shares with the Sequences this pattern of once you’ve got it it’s hard to imagine not having it. We know a lot of our notion of romantic love is Western and modern, but it requires a lot of evidence to prove this to most people because it seems so natural. It’s one possible way of parsing sex and attraction and stuff, and a very natural one for us, but one that was very rarely used in certain circumstances (the crystallization of the pattern “romantic love” probably shouldn’t be confused with semifactual questions like “should we choose partners based on romantic love?”, which already presuppose that we’ve accepted romantic love as a useful concept)

And I’m thinking about these things because of Alexander Stanislaw’s comment on my last post:

> Perhaps being immersed in a memeplex such as LW for long enough is sufficient to make the source material seem obvious, whether or not the memes are correct. I’m sure that there are intelligent Christians our age who

> are re-reading C.S. Lewis and thinking, “wow this is so obvious”.
>
> I admit, I find it very hard to argue with the LW stance on what concepts are for instance, and it drives me nuts when people try to define themselves into being correct, or argue about whether X is really Y, or equivocate between definitions of words. But apparently Gilbert of the Last Conformer disagrees with the LW stance on concepts and he seems knowledgeable.

It’s an interesting question: can crystallizing a pattern be *wrong*?

First, two kinda trivial examples.

Crystallizing the pattern “ley lines” wouldn’t be super useful if you don’t believe ley lines exist. From one perspective, the crystallization is still nice, because you have an extra concept to toss around and you can do things you couldn’t do before like debate the existence of ley lines (and settle on the negative position). From another perspective, the existence of the pattern makes it too tempting to start believing in ley lines or at least thinking they’re important. So in theory crystallizing the concept “ley lines” is an unalloyed positive (minus the five minutes it takes you to read the definition), but in practice it’s a bad idea.

Crystallizing the pattern “mainstream media” in the sense that some conspiracy theorists use it – ie “this exciting archaeological discovery will never be mentioned in the mainstream media” seems more complicated. There definitely exists a “mainstream media” in the sense of “a media that has much different standards and preferences than the local conspiracy theory rag”. And it’s probably useful to talk about – my reaction to the Siberian discovery was “I’ll believe it once I see it in the mainstream media”. But the way the term is used manages to smuggle in questionable connotations – this is some people’s problem with “political correctness” as well, and with other political terms like “patriarchy”, “the Cathedral”, “international Jewry”, et cetera.

I think the problem here might be that pointing out a pattern suggests an agency fiction? There’s clearly international Jewry in the sense of “Jews in many different countries”. One can even make moderately correct statements about international Jewry, like “international Jewry often puts pressure on different countries to support Israel”, which when stripped of its connotations is more or less correct. But the connotations suggest either that they’re all working together in a conspiracy, or that they’re all doing it in about the same way which is becoming a noteworthy pattern, or that it’s something much more interesting than the way other ethnic groups do other things.

I think this was my interlocutor’s problem with political correctness as well – certainly there are things that fit that category, but once the word starts to be used it’s singling it out as A Unique And Interesting Problem.

But I don’t know how to solve this sort of issue in a way that’s easily distinguishable from trying to ban people from ever talking about political correctness or Jewish support for Israel. Making people use a different term than “international Jewry” (“Jews around the world”?) would work for about ten minutes until it got loaded with exactly the same connotations as the original.

But these two examples seem less interesting than the original question about whether you can crystallize patterns, without falling into any traps or connotations – and still have it be a net loss.

I’m trying to think of something that fits Alexander’s example – something written by C.S. Lewis that provides Christians with much more clarity and afterwards they feel like they’re much more advanced and can’t remember how they lived without the concept – but which from my perspective as a non-Christian is wrong and counterproductive.

(my immediate reaction is “ring theory!”, but that’s not dependent on Christianity and probably correct and useful, so it doesn’t count)

The closest I can come (I am not a Lewis scholar) are his trilemma, his use of the different phases of water (ice, liquid, steam) to explain the Trinity, and his famous saying that “going to church does not make you a Christian any more than going to a garage makes you a car”.

But I like all of these!

Although I don’t agree with his formulation of the trilemma, it puts something that people would otherwise spend interminable hours debating without ever reaching the core of the issue – into sharp relief, so that it can be attacked or defended in its strongest form.

And although I don’t accept the premises of the Trinity or the special necessity of figuring out exactly who is or isn’t a “real Christian”, if I admired the problems I would certainly admire Lewis’ solutions to them.

The worst that can be said about Lewis is that he came up with *excessively clever solutions to the wrong problems* – that is, maybe if someone was wrestling with the idea of the

Trinity, that would convince them to second-guess their religion, but Lewis came up with such an elegant metaphor that it prevents them from having to do that.

And I’m not sure that the garage saying or the Trinity analogy are crystallized patterns per se. The trilemma example might be the purest – but that’s also the one I like most.

So provisionally I’m not sure there’s such a thing as crystallizing a pattern and being wrong to do so. You can crystallize patterns in such a way that it ends out misleading people who were already at risk of being misled – like the “ley lines” and “international Jewry” examples – and in practice this is a HUGE HUGE problem. But it seems to me that if you’re good enough at sorting through connotations to handle it that crystallization is usually a good idea

What Universal Human Experiences Are You Missing Without Realizing It?

Remember Galton's experiments on visual imagination? Some people just don't have it. And they never figured it out. They assumed no one had it, and when people talked about being able to picture objects in their minds, they were speaking metaphorically.

And the people who *did* have good visual imaginations didn't catch them. The people without imaginations mastered this "metaphorical way of talking" so well that they passed for normal. No one figured it out until Galton sat everyone down together and said "Hey, can we be really really clear about exactly how literal we're being here?" and everyone realized they were describing different experiences.

I thought about this recently during a conversation with Ozy:

> **Ozy:** I am currently eating chickpeas and rice and I am _delighted_ by the fact that I can eat this _whenever I want_ The nice thing about DISCOVERING YOUR FOOD PREFERENCES is that suddenly all the food in my cupboards is food I like and am looking forward to eating. and usually I get food I like by, like, luck? So this is excitement.
>
> **Scott:** I don't understand, why didn't you buy things like that before?
>
> **Ozy:** It took me a while to have enough of a sense of the food I like for "make a list of the food I like" to be a viable grocery-list-making strategy.

> **Scott:** I’ve got to admit I’m confused and intrigued by your “don’t know my own preferences” thing.
>
> **Ozy:** Hrm. Well, it’s sort of like… you know how sometimes you pretend to like something because it’s high-status, and if you do it well enough you _actually believe_ you like the thing? Unless I pay a lot of attention _all_ my preferences end up being not “what I actually enjoy” but like “what is high status” or “what will keep people from getting angry at me”
>
> **Scott:** How does that apply to food?
>
> **Ozy:** Well, sometimes people will tell you a certain food is high-status or healthy or a thing that everyone enjoys, and then I would like it. And a lot of times I just ate whatever was in front of me or ordered whatever the cheapest vegetarian thing on the menu was. And I… sort of vaguely had a sense that some things were more pleasurable to eat than other things but I didn’t like _keep track_ of what they were or anything. Because if I knew I might like the _wrong things_. And also because I didn’t intuitively grasp that the “liking” thing everyone was talking about was related to pleasure and not to like popularity/status.

So the fact that people talk about what foods they like about a zillion times a day isn’t enough to make everyone realize liking foods is a thing.

But it gets worse. A high school friend posted on Facebook a link to a really interesting answer on Quora. It makes you log on, so I’ll copy the relevant part below:

> I have anosmia, which means I lack smell the way a blind person lacks sight. What’s surprising about this is that I didn’t even know it for the first half of my life.

> Each night I would tell my mom, “Dinner smells great!” I teased my sister about her stinky feet. I held my nose when I ate Brussels sprouts. In gardens, I bent down and took a whiff of the roses. I yelled “gross” when someone farted. I never thought twice about any of it for fourteen years.
>
> Then, in freshman English class, I had an assignment to write about the Garden of Eden using details from all five senses. Working on this one night, I sat in my room imagining a peach. I watched the juice ooze out as I squeezed at the soft fuzz. I felt the wet, sappy liquid drip from my fingers down onto my palm. As the mushy heart of the fruit compressed, I could hear it squishing, and when I took that first bite I could taste the little bit of tartness that followed the incredible sweet sensation flooding my mouth.
>
> But I had to write about smell, too, and I was stopped dead by the question of what a peach smelled like. Good. That was all I could come up with. I tried to think of other things. Garbage smelled bad. Perfume smelled good. Popcorn good. Poop bad. But how so? What was the difference? What were the nuances? In just a few minutes’ reflection I realized that, despite years of believing the contrary, I never had and never would smell a peach.
>
> All my behavior to that point indicated that I had smell. No one suspected I didn’t. For years I simply hadn’t known what it was that was supposed to be there. I just thought the way it was for me was how it was for everyone. It took the right stimulus before I finally discovered the gap.

So I guess you can just *not be able to smell* and not know it.

This makes me wonder what universal human experiences I and my friends are missing out on without realizing it.

I know one friend’s answer. He discovered he was color-blind sometime in his teens. This still surprises me. People are always taking Ishihara tests (those colorful dotted circles with numbers inside of them) and discovering they’re color blind. Going through life with everyone else saying “The light was red, but now it’s green” and thinking it was weird that they were making such a big deal about subtle variations in shades of brownish-gray, but it was probably one of those metaphors.

As for me? I took a surprisingly long time to realize I was asexual. When I was a virgin, I figured sex was one of those things that seemed gross before you did it, and then you realized how great it was. Afterwards, I figured it was something that didn’t get good until you were skilled at it and had been in a relationship long enough to truly appreciate the other person. In retrospect, pretty much every aspect of male sexual culture is a counterargument to that theory, but I guess it’s just really hard for my brain to generate “you are a mental mutant” as a hypothesis.

But even bigger than that, I think I might not have had emotions, at least not fully, for about five years as a teenager when I was on SSRIs. I even sort of noticed myself not having emotions, but dismissed that as an odd thing to happen and probably other people were just being really overexuberant about things. Later I learned emotional blunting is a commonly reported side effect of SSRIs and I was probably just really not experiencing emotions. When I came off them it took me several years to get used to having normal-intensity feelings again, but it wasn’t a sudden revelation, like “Wow, I was missing a fundamental human experience for the past several years!” Just a sense of things being different which was hard to cash out.

As always, I wonder if a lot of what other people interpret through vague social things might be biological, or at least more *complicatedly* social. I can’t enjoy jazz music even a little – the best I can do is pick up something sort of like a beat and half-heartedly feel like maybe I could snap my fingers to it if I could build up the energy. My brother fell in love with jazz as soon as he heard it and is now a professional jazz musician who has dedicated his life to it. Are we listening to the same thing when we hear a jazz tune? Or am I like a guy who can’t smell trying to appreciate perfume?

Michigan Meetup and Memory Workshop, 3/23

There will be a Less Wrong / Slate Star Codex reader / random interested persons meetup and CFAR-ish workshop on memory and mnemonics at my house (near Detroit) this Sunday at 1 PM.

Along with the usual Michiganders we will have special guests Ozy Franz, Brienne Strohl from Entertaining Thoughts, Robby Bensinger from Nothing Is Mere, and probably Chana Messinger from The Merely Real

You can find more info and directions at this page. Email me if you're still confused or have further questions.

Donations appreciated (but not required) to cover the cost of the workshop!

Typical Mind and Disbelief In Straight People

Lots of no doubt very wise people avoid comment threads here, but the comments on [Universal Human Experiences]() are really worth it. The post was about how people mistake their own unique and individual ways of experiencing the world for human universals and assume anyone who says otherwise is speaking metaphorically. And boy did people have a lot of good examples of this.

I want to turn lots of them into extended discussions eventually, but for now I'll settle for the one that reminded me how I *still* fail at Principle of Charity.

Principle of Charity, remember, says you should always assume your ideological opponents' beliefs must make sense from their perspective. If you can't even conceive of a position you oppose being tempting to someone, you don't understand it and are probably missing something. You might be missing a strong argument that the position is correct. Or you might just be missing something totally out of left field.

People always accuse me of going too far with this, and every time I'm tempted to believe them I always get reminded I don't go nearly far enough.

I'm pretty good at feeling tempted by most positions I oppose, which satisfies me that I understand them enought to feel justified in continuing to reject them. The biggest exception is that opposition to homosexuality has never made sense to me. I can sort of understand where it fits into a natural law theology, but a lot of anti-gay activists are, no offense, not exactly Thomas Aquinas.

Chris Hallquist commented:

> I remember hearing once on Dan Savage’s podcast that he gets letters from gay men who grew up in very conservative parts of the country, who didn’t know that being straight was a thing. They assumed all men were attracted to men, but just hid it.

Martin responded:

> Dr. Paul Cameron, founder of the anti-gay Family Research Institute, is quoted as saying: “If all you want is the most satisfying orgasm you can get – and that is what homosexuality seems to be – then homosexuality seems too powerful to resist… It’s pure sexuality. It’s almost like pure heroin. It’s such a rush. They are committed in almost a religious way. And they’ll take enormous risks, do anything.” He says that for married men and women, gay sex would be irresistible. “Martial sex tends toward the boring end,” he points out. “Generally, it doesn’t deliver the kind of sheer sexual pleasure that homosexual sex does” So, Cameron believes, within a few generations homosexuality would be come the dominant form of sexual behavior. Apparently, some people build their entire lives around not knowing that being straight is a thing.

So imagine that you’re one of those people Dan Savage was talking about – a closeted gay guy who doesn’t realize he’s a closeted gay guy. He just thinks – reasonably, given his own experience! -that the natural state of the human male is to be attracted to other men, but that men grudgingly have sex with and marry women anyway because society tells them they have to.

(I don’t know if this generalizes to women)

In that case, *exactly* the anti-gay position conservatives push makes perfect sense for *exactly* the reasons they say it makes sense.

Allowing gay marriage would destroy straight marriage? Yes! If everyone’s secretly gay, then as soon as gay marriage is allowed, they will breath a sigh of relief and stop marrying opposite-sex partners whom they were never very attracted to anyway.

Gay people are depraved and licentious? Yes! Everyone else is virtuously resisting all of these unbearable homosexual impulses, and gay people are the ones who give in, who can’t resist grabbing the marshmallow as soon as it is presented to them.

(I’m referring to this experiment, not some sort of creepy sexual euphemism. Get your heads out of the gutter.)

Teaching children about homosexuality will turn them gay? Yes! The only thing preventing them all from being gay already is the social stigma against it. Teaching them in school that homosexuality is okay and shouldn’t be stigmatized cuts the last thin thread connecting them to straightness.

This can’t be a universal explanation for anti-gay attitudes. Something like half the US population is against gay marriage (previously much more) and probably five percent or less is gay. Closeted gay people don’t explain more than a small fraction of the anti-gay movement.

But it’s probably bigger than the fraction who read Thomas Aquinas. Maybe all these idiosyncratic arguments that only a few people can really appreciate turn into soundbites and justifications that get used by other people who feel vague discomfort but don’t have a good grounding for why. That means they’d have an impact larger than the size of the groups that produce them.

Should You Reverse Any Advice You Hear?

Suppose a lot of that stuff about bravery debates is right.

That lots of the advice people give is useful for some people, but that the opposite advice is useful for other people.

For example, “You need to stop being so hard on yourself, remember you are your own worst critic” versus “Stop making excuses for yourself, you will never be able to change until you admit you’ve hit bottom.”

Or “You need to remember that the government can’t solve all problems and that some regulations are counterproductive” versus “You need to remember that the free market can’t solve all problems and that some regulations are necessary.”

Or “You need to pay more attention to your diet or you’ll end up very unhealthy” versus “You need to pay less attention to your weight or you’ll end up in a spiral of shame and self-loathing and at risk of eating disorders.”

Or “Follow your dreams, you don’t want to be working forever at a job you hate”, versus “Your dream of becoming a professional cosplayer may not be the best way to ensure a secure future for your family, go into petroleum engineering instead.”

Or “You need to be more conscious of how your actions in social situations can make other people uncomfortable and violate their boundaries” versus “You need to overcome your social phobia by realizing that most interactions go well and that probably talking to people won't *always* make them hate you and cause you to be ostracized forever."

Or "You need to be less selfish and more considerate of the needs of others" versus "You can't live for others all the time, you need to remember you deserve to be happy as well."

People often form groups based on pushing one or another side of these dichotomies. Most obviously, the Libertarian Party pushes one side of the political-regulation one and the Communist Party pushes the other. The fat acceptance movement pushes one side of the diet-health one and the American Heart Association pushes the other. Some religious groups and the effective altruism movement push one side of charity-selfishness, the Objectivists push the other.

Most of these groups have the stated purpose of moving society as a whole, but their primary short-term effect is to change the opinions of their members.

For example, maybe you join the Objectivist movement. You follow lots of Objectivist blogs that give you strong arguments for selfishness, hear lots of anecdotes of people being hurt by excessive altruism, and get exposed to any studies that seem to support the pro-selfishness point of view. You probably end up more selfish than you were before you joined the Objectivists.

Consider two possible interpretations of that result.

First, Objectivism might be a successful support group. People who aren't selfish enough realize they need more selfishness in their lives, join the Objectivists, and support each other as they work to overcome their inbuilt disposition to ignore their own needs. Gradually they all become psychologically healthier people.

Or second, Objectivism might be a vicious cycle. The people who are already too selfish see an opportunity to be selfish with a halo. They join Objectivism, egg each other on, and become even more selfish still. Meanwhile, the people who could really have benefitted from Objectivism, the people who feel guilted into living for others all the time while ignoring their own needs, are off in some kind of effective charity group, egging each other on to be even more self-destructively altruistic.

The first dynamic definitely sometimes exists, and in fact I was cued in to this whole issue from a friend in the first situation who was genuinely helped by Objectivism.

But I think the second dynamic is usually more common. It’s much easier to join a group that celebrates your natural proclivities than one that demands you fight against them. Then you end up with dueling death spirals in which two separate communities become more and more certain of their own position.

I was talking to a friend on Facebook about marriage (I’ll default to anonymous; she can identify herself in comments if she wants). She was annoyed at a blog that criticized a supposed group of people who jumped into marriage unreflectively because they felt divorce was an easy and low cost escape if it didn’t go well. She thought this was a cheap shot, which of course it is.

But I pointed out that this criticism really wasn’t entirely off the mark, because these people exist and are in fact very common. The fallout from their recklessness sometimes requires a psychiatrist to help sort out, so I meet them and their children in treatment all the time.

And she said that sure, these people do exist, but there are a lot of people who are stuck in abusive relationships and already feel like divorce is too stigmatizing, and we shouldn't be too quick to mock people who jump to divorce because that's just going to make it harder for these people to get the divorces they really need.

And I said that okay, definitely those people existed as well, but it seemed kind of unfair for this demographic to hold hostage society's ability to suggest people be more responsible with marriage and divorce, when there are so many people who would benefit from that advice.

And she said that yes, it would be nice to provide these irresponsible people the information that they need to think carefully before making major life choices, but that these probably weren't the sort of people who read preachy conservative blogs about the virtues of the married life anyway.

And I didn't have a good answer to that, because it was obviously true. The best I could do was point out that this would delegitimize pretty much all discourse. *Every* piece of social commentary is most likely to go to the people who need it least.

For just this reason I worry that everything I post on my blog is correct, but wrong relative to readers of my blog. For example, I post about how everyone needs to be much more mindful of the role biological factors play in human social systems. And I'm 99% sure that the average person is not sufficiently aware of or concerned about this.

But I’m much less certain that the average reader of my blog isn’t sufficiently aware or concerned about this. Maybe people who are really interested in biodeterminism search “biodeterminism” on Google, find my blog and several others, and end up way too biodeterminist. Maybe their time would be much better served reading some blog on how many things are due to fuzzy hard-to-measure social factors like who your third-grade teacher was.

And when a young person is looking for job advice, I worry that all the artsy creative people whose heads are already way too high in the skies will be reading books by artsy creative people who urge them to follow their dreams, and so be even less mindful of the importance of a secure future. And all the hard-headed down-to-earth people will naturally gravitate toward reading *Have A Very Secure Future By Going Into Business* by Warren Buffett, and maybe never get reminded of the importance of following dreams.

(This is even sadder when the groups aren’t equal in size, when society is much more in need of one side than the other, but that group is stuck in a tiny but super-intense inward-facing spiral. The Venn diagram of the people who most need to learn about LGBTQ rights compared against the people who most often hear about LGBTQ rights consists of two circles, one in Canada and the other in Peru. And so we end up with a big community of people who want trans people bullied out of society, plus a tiny community who spend a lot of time panicking that they might be unintentionally offensive by using the wrong form of “trans-” vs. “trans*”.)

I wonder whether everyone would be better off if they automatically reversed any tempting advice that they heard (except feedback directed at them personally). Whenever they read an inspirational figure saying “take more risks”, they interpret it as “I seem to be looking for advice telling me to take more risks; that fact itself means I am probably risk-seeking and need to be more careful”. Whenever they read someone telling them about the obesity crisis, they interpret it as “I seem to be in a very health-conscious community; maybe I should worry about my weight less.”

Probably this wouldn’t literally work as written. Too much advice is applicable to everybody; the absence of advice to play more Russian roulette is directly linked to Russian roulette being a really bad idea for pretty much everyone.

But advice reversal might at least be worth *considering*. The checklist could be something like:

1. Are there plausibly near-equal groups of people who need this advice versus the opposite advice?

2. Have you self-selected into the group of people receiving this advice by, for example, being a fan of the blog / magazine / TV channel / political party / self-help-movement offering it?

3. Then maybe the opposite advice, for you in particular, is at least as worthy of consideration.

E-Cig Study: Much Smoke, Little Light

New study shows that e-cigarette users are no more likely to quit smoking tobacco after a year than non-e-cigarette users. In fact, the trend is in the opposite direction – e-cigarette users are less likely to give up their regular cigarettes. I'm skeptical. r/science is skeptical. The experts are skeptical. Even the authors of the study sound maybe a *little* skeptical.

The study surveyed tobacco smokers for various information including whether they smoked e-cigarettes in addition to their tobacco. A year later, they went back and surveyed everyone again and asked them if they were still smoking tobacco. And the people smoking the e-cigarettes were no more likely to have quit than the others.

Let's transition from reality to Hypothetical World. In Hypothetical World, there are only two kinds of smokers, Short Smokers and Long Smokers. The moment someone smokes their first cigarette, God flips a coin and assigns them into one of the two groups based on the result. Short Smokers are predestined to smoke for exactly one year before quitting; Long Smokers are predestined to smoke for exactly fifty years before quitting.

A scientist in Hypothetical World wants to discovery what percent of first-time-smokers become Short Smokers versus Long Smokers (the real proportion is 50-50 since God's coin is fair, but she doesn't know that). So she uses the same methodology as this study. She hangs around a tobacco shop and accosts the first thousand people who come in to buy cigarettes, getting their names and phone numbers. Then a year and a day later, she calls them all up to ask if they are still smoking – since anyone who keeps smoking for a year and a day must be a Long Smoker.

So she finds something close to 2% of people are Short Smokers and a whopping 98% are Long Smokers. She incorrectly concludes that God is rolling a d100 and only assigning Short Smoker status to those who come up 99 or 00.

Don’t see why she would make this mistake? Consider a particular generation of Hypothetical people over their lifetimes. The Short Smokers will only smoke a single year out of their lifetime; the Long Smokers will smoke fifty years. When the scientist does her study in a randomly selected year, she only has a 1/average_lifespan chance of catching any given Short Smoker, but a 50/average_lifespan chance of catching a Long Smoker. So, her original sample will contain fifty times more Long Smokers than Short Smokers, and she will mistakenly conclude that their pattern is fifty times more common.

Now transition back to reality. Suppose there are two types of e-cigarette users – successful and unsuccessful. The successful e-cigarette users try e-cigarettes, immediately decide they are better than regular cigarettes, and switch to using e-cigarettes exclusively within one month. The unsuccessful e-cigarette users try e-cigarettes but just don’t get everything they love about tobacco from them. They sort of futz around with e-cigarettes and regular cigarettes and tell themselves that one of these days, they’re really going to stop the regular ones entirely and transition totally to e-cigarettes. These people continue futzing for let’s say ten years before they either finally quit tobacco, give up on e-cigarettes, or die.

In that case, any sample of tobacco smokers taken at a particular time will include a hundred twenty times as many unsuccessful e-cigarette users as successful ones. We expect unsuccessful e-cigarette users to continue their past pattern of futzing around, so it's not surprising that this sort of sample finds most e-cigarette users not only can't easily quit tobacco using e-cigarettes, but actually have a harder time quitting tobacco than normal smokers – they've already been preselected as The Group That Even E-Cigarettes Can't Help; as The Group That Tried Something Billed As An Anti-Smoking Aid But Failed At It. It's a pretty general rule of medicine that people who failed treatment once are more likely to fail treatment a second time.

This is a very speculative explanation and I haven't heard anyone respectable at a major institution advance it yet, but it seems to me like the most likely reason for these findings. All I have to go for with the study right now is a preliminary "research letter", but hopefully we'll know more when the real thing comes out.

Lest this post be *entirely* pro-drug, here's a clip of my addiction-medicine teacher and sometime-boss lecturing people about marijuana on Fox News last weekend. He is a great doctor and it's neat to see him finally getting some of the celebrity he deserves. Even though his politics are terrible.

The Character’s Complaint

I’ve wiped off the blood and I’ve cleaned up the rubble
And most of the burn marks have faded to pink
I’ve finally started to save up some cash
After paying the surgeon, the doc, and the shrink
The neighbors are letting their kids play outside now
And even the nightmares are starting to ease
So if you are reading this, Author, I’m pleading
Don’t write any books on me! No more books! Please!

I don’t want to get in a hijack or bombing
Which seems to be most of the plot of a thriller
I’ve noticed that when you say “mystery book”
You mean “brutal unstoppable serial killer”
I don’t want to be in a romance
If it ends with divorce from my spouse
Or trudge off to the stores to buy heavy steel doors
To zombie-proof all of the gates to my house

I don’t want to be in a novel on war
And be sent forth to die at the edge of the globe
Though the future sounds swell, I’ll stay out of of sci-fi
If it has the words “parasite”, “anal”, or “probe”
I don’t want to be in dystopian fiction
And get renamed something like “Janitor Eight”
I’m far too high class to be happy with fanfic
And ladies, no slash please – no, honest, I’m straight

To inclusion in works of some authors of note
I object with particular force
Solzhenitsyn, Wiesel, Kafka, Salinger, Plath
Either Bronte – plus Lovecraft, of course
Victor Hugo would make me *le misérable*
And George RR Martin inspires some doubt

It’s not just the fact that he’d probably kill me
It’s that I would have to wait years to find out

A Rand book would probably help the economy
There’d always be jobs in a world of tycoons
But I’d have to sit through all those monologues
By long-winded lone libertarian loons
Neal Stephenson’s better, but isn’t for me
I always found programming hard
And although I loved Ender, I’m leftist on gender
And boycotting Orson Scott Card

I promise I’m doing my best not to tempt you
I’ve put away plot hooks and tied up loose ends
I’ve found my real parents with DNA testing
Confessed all my crushes on all my cute friends
I’m staying away from ruined cities and taverns
Avoiding emotions like angst and ennui
If I find in my home some forgotten old tome
I will back away slowly, and let the thing be

I’m not taking actions that might have a Moral
Like dissing my elders or mocking the poor
I won’t undervalue the Power of Love
Or disregard friendship ’cause cash matters more
I won’t sell my cow for a couple of bean plants
I won’t kill my magic gold-egg laying goose
Or challenge the sky, with my fist raised up high
Shouting “I AM MORE [ADJECTIVE] EVEN THAN ZEUS!”

My actions don’t show We Must Value Diversity
They don’t provide proof Being Different’s Okay
I try to avoid Overcoming Adversity
Based upon race, gender, class, or Teh Gay
I will not discover The Faith Of A Child
Is purer than that of priest, rabbi, or lama

Wherever I’m able, I’ve sidestepped each fable
In ways that exclude any lesson or drama

And really, your talents are wasted on fiction
Why not try your hand at political prose?
I hear that some papers pay pundits good money
And blogs are more lucrative even than those
You could be the guy who gets studies in *Nature*
Or who writes for the popular press and explains
The import to clients of breakthroughs in science
Like how vaccinations won’t melt their kids’ brains

What if you find yourself thwarted by writer’s block
Hands stained with ink and a face stained with tears?
Hunching at desks for too long causes joint pain
I hear carpal tunnel stays with you for years
The kids these days just play computer games anyway
Each year the market gets smaller and smaller
And editors milk you and agents will bilk you
And publishers only pay cents on the dollar

Our mutual interests align pretty nicely
You don’t want a half-finished draft on the shelf
And I’d rather not be thrust into a conflict
With Man, Nature, Culture, The Gods, or Myself
So go take up dancing, or look at a sunset
Rekindle the flame with your husband or wife
Don’t write one more letter – we both deserve better
Put down the damn pen and get on with your life!

[dedicated to Alicorn for obvious reasons]

Wheat: Much More Than You Wanted To Know

After hearing conflicting advice from diet books and the medical community, I decided to look into wheat.

There are two sets of arguments against including wheat in the diet. First, wheat is a carbohydrate, and some people support low carbohydrate diets. Second, something might be especially dangerous about wheat itself.

It was much easier to figure out the state of the evidence on low-carbohydrate diets. They seem to be at least as good and maybe a little better for weight loss than traditional diets, but this might just be because there are lots of carbohydrates that taste very good and when forced to avoid them, people eat less stuff. They may or may not positively affect metabolic parameters and quality of life. (1, 2, 3, 4). They don't seem to cause either major health benefits or major health risks in the medium term, which is the longest term for which there is good data available – for example, they have no effect on cancer rates. Overall they seem solid but unspectacular. But there's a long way between "low carbohydrate diet" and "stop eating wheat".

So I was more interested in figuring out what was going on with wheat in particular.

Wheat contains chemicals [citation needed]. The ones that keep cropping up (no pun intended) in these kinds of discussions are phytates, lectins, gluten, gliadin, and agglutinin, the last three of which for your convenience have been given names that all sound alike.

Various claims have been made about these chemicals' effects on health. These have some prima facie plausibility. Plants don't want to be eaten [citation needed] and they sometimes fill their grains with toxins to discourage animals from eating them. Ricin, a lectin in the seeds of the castor oil plant so toxic it gets used in chemical warfare, is a pretty good example. Most toxins are less dramatic, and most animals have enzymes that break down the toxins in their preferred food sources effectively. But if humans are insufficiently good at this, maybe because they didn't evolve to eat wheat, some of these chemicals could be toxic to humans.

On the other hand, this same argument covers every pretty much every grain and vegetable and a lot of legumes – pretty much every plant-based food source except edible fruits. So we need a lot more evidence to start worrying about wheat.

I found the following claims about negative effects of wheat:

1. Some people without celiac disease are nevertheless sensitive to gluten.
2. Wheat increases intestinal permeability, causing a leaky gut and autoimmune disease.
3. Digestion of wheat produces opiates, which get you addicted to wheat.
4. Wheat something something something autism and schizophrenia.
5. Wheat has been genetically modified recently in ways that make it much worse for you.
6. The lectins in wheat interfere with leptin receptors, making people leptin resistant and therefore obese.

I'll try to look at each of those and then turn to the positive claims made about wheat to see if they're strong enough to counteract them.

Some People Without Celiac Disease Are Sensitive To Gluten – *Mostly true but of limited significance*

Celiac disease is one source of concern. Everybody on all sides of the wheat debate agree about the basic facts of this condition, which affects a little less than 1% of the population. They have severe reactions to the gluten in wheat. Celiac disease is mostly marked by gastroentereological complaints – diarrhea, bloating, abdominal pain – but it is also associated with vitamin deficiencies, anaemia, skin reactions, infertility, and “malaise”. It can be pretty straightforwardly detected by blood tests and gut biopsies and is not subtle.

People start to disagree about the existence of “gluten sensitivity”, which if it existed would be a bad reaction to gluten even in people who don’t test positive for celiac disease. Many people believe they have gastrointestinal (or other) symptoms that go away when they eat gluten-free diets, but science can’t find anything wrong with their intestines that could be causing the problems.

A recent study somewhat vindicated these people. Biesiekierski 2011 describes a double-blind randomized controlled trial: people who said they had “gluten-sensitive” irritable bowel syndrome were put on otherwise gluten-free diets and then randomly given either gluten or a placebo. They found that the patients given gluten reported symptoms (mostly bowel-related and tiredness) much more than those given placebo ($p = 0.0001$) but did not demonstrate any of the chemical, immunological, or histological markers usually associated with celiac disease. A similar Italian study found the same thing, except that they did find a higher rate of anti-gluten antibodies in their patients. Another study found that non-celiacs with antibodies to gluten had higher rates of mortality. And another study did find a histological change in bowel barrier function on this group of patients with the introduction of gluten. And another study from the same group found that maybe FODMAPs, another component of wheat, are equally or more responsible.

The journal *Gastroenterology*, which you may not be surprised to learn is the leading journal in the field of gastroenterology, proclaims:

> The current working definition of nonceliac gluten sensitivity (NCGS) is the occurrence of irritable bowel syndrome (IBS)-like symptoms after the ingestion of gluten and improvement after gluten withdrawal from the diet after exclusion of celiac disease based on negative celiac serologies and/or normal intestinal architecture and negative immunoglobulin (Ig)E-mediated allergy tests to wheat. Symptoms reported to be consistent with NCGS are both intestinal (diarrhea, abdominal discomfort or pain, bloating, and flatulence) and extra-intestinal (headache, lethargy, poor concentration, ataxia, or recurrent oral ulceration). These criteria strongly and conveniently suggest that NCGS is best understood as a subset of IBS or perhaps a closely related but distinct functional disorder. Although the existence of NCGS has been slowly gaining ground with physicians and scientists, NCGS has enjoyed rapid and widespread adoption by the general public.

But even this isn’t really that interesting. Maybe some people with irritable bowel syndrome or certain positive antibodies should try avoiding gluten to see if it helps their specific and very real symptoms. At most ten percent of people are positive antibody testing, and not all of those even have symptoms. That's still a far cry from saying no one should eat wheat.

But the anti-wheat crowd says an alternative more sensitive antibody test could raise sensitivity as high as a third of the population. The test seems to have been developed by a well-respected and legitimate doctor, but it hasn't as far as I can tell been submitted for peer review or been confirmed by any other source. Meh.

That's boring anyway. The real excitement comes from sweeping declarations that *the entire* population is sensitive to wheat.

Wheat Increases Intestinal Permeability Causing A Leaky Gut – *Probably true, of uncertain significance*

There are gluten-induced mucosal changes in subjects without small bowel disease. And gliadin increases intestinal permeability in the test tube, which should be extremely concerning to any test tubes reading this.

But probably the bigger worry here are lectins, which include wheat germ agglutinin. WGA affects the intestinal permeability of rats, which should be extremely concerning to any rats reading this. The same substance has been found to produce pro-inflammatory cytokines and interfere with the growth of various organs including the gut.

So there's pretty good evidence that chemicals in wheat can increase intestinal permeability. Who cares?

For years, "leaky gut syndrome" was an alternative medicine diagnosis that was soundly mocked by the mainstream medical establishment. Then the mainstream medical establishment confirmed it existed and did that thing where they totally excused their own mocking of it but were ABSOLUTELY OUTRAGED that the alternative medicine community might have in some cases been overenthusiastic about it.

Maybe I'm being too harsh. The alternative medicine community often does take "leaky gut syndrome" way too far.

On the other hand, it's probably real and *Nature Clinical Practice* is now publishing papers saying it is "a key ingredient in the pathogenesis of autoimmune diseases" and "offers innovative, unexplored approaches for the treatment of these devastating diseases" and gut health has been deemed "a new objective in medicine". Preliminary changes to intestinal permeability have been found in asthma, in diabetes, and even in depression.

But it's not yet clear if this is cause and effect. Maybe the stress of having asthma increases intestinal permeability somehow. Or maybe high intestinal permeability causes asthma somehow. It sure seems like the latter might work – all sorts of weird antigens and stuff from food can make it into the bloodstream and alarm the immune system – but right now this is all speculative.

So what we have is some preliminary evidence that wheat increases intestinal permeability, and some preliminary evidence that increased intestinal permeability is bad for you in a variety of ways.

And I don't doubt that those two facts are true, but my knowledge of this whole area is so weak that I wonder how much to worry.

What other foods increase intestinal permeability? Do they do it more or less than wheat? Has anyone been investigating this? Are there common things that affect intestinal permeability a thousand times more than wheat does, such that everything done by wheat is totally irrelevant in comparison?

Do people without autoimmune diseases suffer any danger from increased intestinal permeability? How much? Is it enough to offset the many known benefits of eating wheat (to be discussed later?) Fiber seems to decrease intestinal permeability and most people get their fiber from bread; would decreasing bread consumption make leaky gut even worse?

I find this topic really interesting, but in a “I hope they do more research” sort of way, not an “I shall never eat bread ever again” sort of way.

Digestion Of Wheat Produces Opiates, Which Get You Addicted To Wheat – *Probably false, but just true enough to be weird*

Dr. William Davis, a cardiologist, most famously makes this claim in his book *Wheat Belly*. He says that gliadin (a component of gluten) gets digested into opiates, chemicals similar to morphine and heroin with a variety of bioactive effects. This makes you addicted to food in general and wheat in particular, the same way you would get addicted to morphine or heroin. This is why people are getting fat nowadays – they’re eating not because they’re hungry, but because they’re addicted. He notes that drugs that block opiates make people want wheat less.

Does Wheat Make Us Fat And Sick, a review published in the *Journal of Cereal Science* (they have journals for *everything* nowadays) is a good rebuttal to some of Davis’ claims and a good pro-wheat resource in general.

They say that although gliadin does digest into opiates, those opiates are seven unit peptides and so too big to be absorbed from the gut to the bloodstream.

(note that having opiates *in your gut* isn't a great idea either since there are lots of nerves there controlling digestion that can be affected by these drugs)

But I'm not sure this statement about absorption is even true. First, large proteins can sometimes make it into the gut. Second, if all that leaky gut syndrome stuff above is right, maybe the gut is unusually permeable after wheat consumption. Third, there have been sporadically reported cases of gliadin-derived opiates found in the urine, which implied they got absorbed somehow.

There's a better counterargument on the blog The Curious Coconut. She notes that there's no evidence these peptides can cross the blood-brain barrier, a precondition for having any psychological effects. And although the opiate-blocker naloxone does decrease appetite, this effect is not preferential for wheat, and probably more related to the fact that opiates are the way the brain reminds itself it's enjoying itself (so that opiate-blocked people can't enjoy eating as much).

And then there's the usual absence of qualifiers. Lots of things are "chemically related" to other chemicals without having the same effect; are gliadin-derived opiates addictive? Are they produced in quantities high enough to be relevant in real life? Corn, spinach, and maybe meat can all get digested into opiates – is there any evidence wheat-derived opiates are worse? This is really sketchy.

The most convincing counterargument is that as far as anyone can tell, wheat makes people eat less, not more:

> Prospective studies suggest that weight gain and increases in abdominal adiposity over time are lower in people who consume more whole grains. Analyses of the Physicians' Health Study (27) and the Nurses' Health Study (26) showed that those who consumed more whole grain foods consistently weighed less than those who consumed fewer whole grain foods at each follow-up period of the study. Koh-Banerjee et al. (27) estimated that for every 40-g increase in daily whole grain intake, the 8-y weight gain was lower by 1.1 kg.

I'll discuss this in more detail later, but it does seem like a nail in the coffin for the "people eat too much because they're addicted to wheat" theory.

Still, who would have thought that wheat being digested into opiates was even a *little* true?

Wheat Something Something Something Autism And Schizophrenia – *Definitely weird*

Since gluten-free diets get tried for everything, and everything gets tried for autism, it was overdetermined that people would try gluten-free diets for autism.

All three of the issues mentioned above – immune reactivity to gluten, leaky guts, and gliadin-derived opiates – have been suggested as mechanisms for why gluten free diets might be useful in autism.

Of studies that have investigated, a review found that seven reported positive results, four negative results, and two mixed results – but that all of the studies involved were terrible and the ones that were slightly less terrible seemed to be more negative. The authors described this as evidence against gluten-free diets for autism, although someone with the opposite bias could have equally well looked at the same review and described it as supportive.

However, a very large epidemiological study found (popular article, study abstract) that people with antibodies to gluten had three times the incidence of autism spectrum disease than people without, and that the antibodies preceded the development of the condition.

Also, those wheat-derived opioids from the last section – as well as milk-derived opioids called casomorphins – seem to be detected at much higher rates in autistic people.

Both of these factors may have less to do with wheat in particular and more to do with some general dysregulation of peptide metabolism in autism. If for some reason the gut kept throwing peptides into the body inappropriately, this would disrupt neurodevelopment, lead to more peptides in the urine, and give the immune system more chance to react to gluten.

The most important thing to remember here is that it would be really wrong to say wheat might be "the cause" of autism. Most likely people do not improve on gluten-free diets. While there's room to argue that people might have picked up a small signal of them improving *a little*, the idea that this totally removes the condition is right out. If we were doing this same study with celiac disease, we wouldn't be wasting our time with marginally significant results. Besides, we know autism is multifactorial, and we know it probably begins in utero.

Schizophrenia right now is in a similar place. Schizophrenics are five to seven times more likely to have anti-gliadin antibodies as the general population. We can come up with all sorts of weird confounders – maybe antipsychotic medications increase gut permeability? – but that’s a really strong result. And schizophrenics have frank celiac disease at five to ten times the rate of the general population. Furthermore, a certain subset of schizophrenics sees a dramatic reduction in symptoms when put on a strict gluten-free diet (this is psychiatrically useless, both because we don’t know which subset, and because given how much trouble we have getting schizophrenics to swallow one lousy pill every morning, the chance we can get them to stick to a gluten-free diet is basically nil). And like those with autism, schizophrenics show increased levels of weird peptides in their urine.

But a lot of patients with schizophrenia don’t have reactions to gluten, a lot don’t improve on a gluten free diet, and other studies question the research showing that any of them at all do.

The situation here looks a lot like autism – a complex multifactorial process that probably isn’t caused by gluten but where we see interesting things going on in the vague territory of gluten/celiac/immune response/gut permeability/peptides, with goodness only knows which ones come first and which are causal.

Wheat Has Been Genetically Modified Recently In Ways That Make It Much Worse For You – *Probably true, especially if genetically modified means “not genetically modified” and “recently” means “nine thousand years ago”*

If you want to blame the “obesity epidemic” or “autism epidemic” or any other epidemic on wheat, at some point you have to deal with people eating wheat for nine thousand years and not getting epidemics of these things. Dr. Davis and other wheat opponents have turned to claims that wheat has been “genetically modified” in ways that improve crop yield but also make it more dangerous. Is this true?

Wheat has not been genetically modified in the classic sense, the one where mad scientists with a god complex inject genes from jellyfish into wheat and all of a sudden your bread has tentacles and every time you try to eat it it stings you. But it has been modified in the same way as all of our livestock, crops, and domestic pets – by selective breeding. Modern agricultural wheat doesn’t look much like its ancient wild ancestors.

The *Journal Of Cereal Science* folk don’t seem to think this is terribly relevant. They say:

> Gliadins are present in all wheat lines and in related wild species. In addition, seeds of certain ancient types of tetraploid wheat have even greater amounts of total gliadin than modern accessions…There is no evidence that selective breeding has resulted in detrimental effects on the nutritional properties or health benefits of the wheat grain, with the exception that the dilution of other components with starch occurs in modern high yielding lines (starch comprising about 80% of the grain dry weight). Selection for high protein content has been carried out for bread making, with modern bread making varieties generally containing about 1–2% more protein (on a grain dry weight basis) than varieties bred for livestock feed when grown under the same conditions. However, this genetically determined difference in protein content is less than can be achieved by application of nitrogen fertilizer. We consider that statements made in the book of Davis, as well as in related interviews, cannot be substantiated based on published scientific studies.

In support of this proposition, in the test tube ancient grains were just as bad for celiac patients' immune systems as modern ones.

And yet in one double-blind randomized-controlled trial, people with irritable bowel syndrome felt better on a diet of ancient grains than modern ones ($p < 0.0001$); and in another, people on an ancient grain diet had lower inflammatory markers and generally better nutritional parameters than people on a modern grain one. Isn't that interesting?

Even though it's a little bit weird and I don't think anyone understands the exact nutrients at work, sure, let's give this one to the ancient grain people.

The Lectins In Wheat Interfere With Leptin Receptors, Making People Leptin Resistant And Therefore Obese – *Currently at "mere assertion" level until I hear some evidence*

So here's the argument. Your brain has receptors for the hormone leptin, which tells you when to stop eating. But "lectin" sounds a lot like "leptin", and this confuses the receptors, so they give up and tell you to just eat as much as you want.

Okay, this probably isn't the real argument. But even though a lot of wheat opponents cite the heck out of this theory, the only presentation of evidence I can find is Jonsson et al (2005), which points out that there are a lot of diseases of civilization, they seem to revolve around leptin, something common to civilization must be causing them, and maybe that thing could be lectin.

But civilization actually contains more things than a certain class of proteins found in grains! There's poor evidence of lectin actually interfering with the leptin receptor in humans. The only piece of evidence they provide is a nonsignificant trend toward more cardiovascular disease in people who eat more whole grains in one study, and as we will see, that is wildly contradicted by all other studies.

This one does not impress me much.

Wheat Is Actually Super Good For You And You Should Have It All The Time – *Probably more evidence than the other claims on this list*

Before I mention any evidence, let me tell you what we're going to find.

We're going to find very, very many large studies finding conclusively that whole grains are great in a lot of different ways.

And we're not going to know whether it's at all applicable to the current question.

Pretty much all these studies show that people with some high level of "whole grain consumption" are much healthier than people with some lower level of same. That sounds impressive.

But what none of these studies are going to do a good job ruling out is that whole grain is just funging against refined grain which is even worse. Like maybe the people who report low whole grain consumption are eating lots of refined grain, and so more total grain, and the high-whole-grain-consumption people are actually eating less grain total.

They're also not going to rule out the universal problem that if something is widely known to be healthy (like eating whole grains) then the same health-conscious people who exercise and eat lots of vegetables will start doing it, so when we find that the people doing it are healthier, for all we know it's just that the people doing it are exercising and eating vegetables.

That having been said, eating lots of whole grain decreases BMI, metabolic risk factors, fasting insulin, and body weight (1, 2, 3, 4,5.)

The American Society For Nutrition Symposium says:

> Several mechanisms have been suggested to explain why whole grain intake may play a role in body weight management. Fiber content of whole grain foods may influence food volume and energy density, gastric emptying, and glycemic response. Whole grains has also been proposed to play an important role in promoting satiety; individuals who eat more whole grain foods may eat less because they feel satisfied with less food. Some studies comparing feelings of fullness or actual food intake after ingestion of certain whole grains, such as barley, oats, buckwheat, or quinoa, compared with refined grain controls indicated a trend toward increased satiety with whole grains. These data are in accordance with analyses determining the satiety index of a large number of foods, which showed that the satiety index of traditional white bread was lower than that of whole grain breads. However, in general, these satiety studies have not observed a reduction in energy intake; hence, further research is needed to better understand the satiety effects of whole grains and their impact on weight management.
>
> Whole grains, in some studies, have also been observed to lower the glycemic and insulin responses, affect

hunger hormones, and reduce subsequent food intake in adults. Ingestion of specific whole grains has been shown to influence hormones that affect appetite and fullness, such as ghrelin, peptide YY, glucose-dependent insulinotropic polypeptide, glucagon-like peptide 1, and cholecystokinin. Whole grain foods with fiber, such as wheat bran or functional doses of high molecular weight β-glucans, compared with lower fiber or refined counterparts have been observed to alter gastric emptying rates. Although it is likely that whole grains and dietary fiber may have similar effects on satiety, fullness, and energy intake, further research is needed to elucidate how, and to what degree, short-term satiety influences body weight in all age groups.

Differences in particle size of whole grain foods may have an effect on satiety, glycemic response, and other metabolic and biochemical (leptin, insulin, etc.) responses. Additionally, whole grains have been suggested to have prebiotic effects. For example, the presence of oligosaccharides, RS, and other fermentable carbohydrates may increase the number of fecal bifidobacteria and lactobacilli (49), thus potentially increasing the SCFA production and thereby potentially altering the metabolic and physiological responses that affect body weight regulation.

In summary, the current evidence among a predominantly Caucasian population suggests that consuming 3 or more servings of whole grains per day is associated with lower BMI, lower abdominal adiposity, and trends toward lower weight gain over time. However, intervention studies have been inconsistent regarding weight loss

The studies that combined whole and refined grains are notably fewer. But Dietary Intake Of Whole And Refined Grain Breakfast Cereals And Weight Gain In Men finds that among 18,000 male doctors, those who ate breakfast cereal (regardless of whether it was whole and refined) were less likely to become overweight several years later than those who did not (p = 0.01). A book with many international studies report several that find a health benefit of whole grains, several that find a health benefit of all grains (Swedes who ate more grains had lower abdominal obesity; Greeks who ate a grain-rich diet were less likely to become obese; Koreans who ate a “Westernized” bread-and-dairy diet were less likely to have abdominal obesity) and no studies that showed any positive association between grains and obesity, whether whole or refined.

I cannot find good interventional trials on what happens when a population replaces non-grain with grain.

On the other hand, Dr. Davis and his book *Wheat Belly* claim:

> Typically, people who say goodbye to wheat lose a pound a day for the first 10 days. Weight loss then slows to yield 25-30 pounds over the subsequent 3-6 months (differing depending on body size, quality of diet at the start, male vs. female, etc.)
>
> Recall that people who are wheat-free consume, on average, 400 calories less per day and are not driven by the 90-120 minute cycle of hunger that is common to wheat. It means you eat when you are hungry and you eat less. It means a breakfast of 3 eggs with green peppers and sundried tomatoes, olive oil, and mozzarella cheese for breakfast at 7 am and you’re not hungry until 1 pm. That’s an entirely different

> experience than the shredded wheat cereal in skim milk at 7 am, hungry for a snack at 9 am, hungry again at 11 am, counting the minutes until lunch. Eat lunch at noon, sleepy by 2 pm, etc. All of this goes away by banning wheat from the diet, provided the lost calories are replaced with real healthy foods."

Needless to say, he has no studies supporting this assertion. But the weird thing is, his message board is full of people who report having exactly this experience, my friends who have gone paleo have reported exactly this experience, and when I experimented with it, I had pretty much exactly this experience. Even the blogger from whom I took some of the strongest evidence criticizing Davis says she had exactly this experience.

The first and most likely explanation is that anecdotal evidence sucks and we should shut the hell up. Are there other, less satisfying explanations?

Maybe completely removing wheat from the diet has a nonlinear effect relative to cutting down on it? For example, in celiac disease there is no such thing as "partially gluten free" – if you have any gluten at all, your disease comes back in full force. This probably wouldn't explain Dr. Davis' observation – neither I nor my other wheatless-experimentation friends were as scrupulous as a celiac would have to be. But maybe there's a nonlinear discrepancy between people who have 75% the wheat of a normal person and 10% the wheat of a normal person?

Maybe there's an effect where people who like wheat but remove it from the diet are eating things they don't like, and so eat less of them? But people who don't like wheat like other stuff, and so eat lots of that?

Maybe wheat in those studies is totally 100% a confounder for whether people are generally healthy and follow their doctor’s advice, and the rest of the doctor’s advice is really good but the wheat itself is terrible?

Maybe cutting out wheat has really positive short-term effects, but neutral to negative long-term effects?

Maybe as usual in these sorts of situations, the simplest explanation is best.

Final Thoughts

Non-celiac gluten sensitivity is clearly a real thing. It seems to produce irritable bowel type symptoms. If you have irritable bowel type symptoms, it might be worth trying a gluten-free diet for a while. But the excellent evidence for its existence doesn’t seem to carry over to the normal population who don’t experience bowel symptoms.

What these people have are vague strands of evidence. Something seems to be going on with autism and schizophrenia – but most people don’t have autism or schizophrenia. The intestinal barrier seems to become more permeable with possible implications for autoimmune diseases – but most people don’t have autoimmune disease. Some bad things seem to happen in rats and test tubes – but most people aren’t rats or test tubes.

You’d have to want to take a position of maximum caution – wheat seems to do all these things, and even though none of them in particular obviously hurt me directly, all of them together make it look like the body just doesn’t do very well with this substance, and probably other ways the body doesn’t do very well with this substance will turn up, and some of them probably affect me.

There’s honor in a position of maximum caution, especially in a field as confusing as nutrition. It would not surprise me if the leaky gut connection turned into something very big that had general implications for, for example, mental health. And then people who ate grain might regret it.

But stack that up against the pro-wheat studies. None of them are great, but they mostly do something the anti-wheat studies don’t: show direct effect on things that are important to you. Most people don’t have autism or schizophrenia, but most people *do* have to worry about cardiovascular disease. We *do* have medium-term data that wheat doesn’t cause cancer, or increase obesity, or contribute to diabetes, or any of that stuff, and at this point solely based on the empirical data it seems much more likely to help with those things than hurt.

I hope the role of intestinal permeability in autoimmune disease gets the attention it deserves – and when it does, I might have to change my mind. I hope people stop being jerks about gluten sensitivity, admit it exists, and find better ways to deal with it. And if people find that eliminating bread from their diet makes them feel better or lose weight faster, cool.

But as far as I can tell the best evidence is on the pro-wheat side of things for most people at most times.

[**EDIT:** An especially good summary of the anti-wheat position is 6 Ways Wheat Can Destroy Your Health. An especially good pro-wheat summary is Does Wheat Make Us Fat And Sick?]

The Study of Anglophysics

I.

Dear Dr. McCord:

Seven years ago, our research staff read with interest your work on Berkeleyan idealism. We were particularly fascinated by your seemingly outrageous claim that it might be possible for individuals to imagine mental worlds so strongly that they would take on a reality of their own.

At the time, as our laboratory had an interest in novel solutions to the overpopulation problem, we embarked upon a test project to see whether a parallel world could be imaged and then colonized by citizens from our own dimension. Using advanced science you could not possibly comprehend, we came up with a practical implementation of your idea. Dr. Michael Adwell, whom I believe you met during your time in Oxford, volunteered to enter the device we had constructed as our first research subject. We very briefly imaged an alternate world based on the contents of Dr. Adwell's mind before the good doctor unfortunately had a grand mal seizure. He was disconnected from the device and rushed to the hospital, where he passed away several hours later.

Two years ago we revisited some of our calculations on the project and determined, to our surprise, that the world Dr. Adwell had created might still exist in some sense; that it had somehow managed to sustain itself separate from the doctor's mental activity. We worked feverishly to construct a device that might let us interact with his imaged world. Six months ago we succeeded. The computational demands of the machine were immense, but after throwing the remainder of our budget for the year at the Kyoto Supercomputing Laboratory, we were able to rent enough processing power to translate myself and Dr. Lachlan Fairchild into the imaged world, which we dubbed “Adwellia” after our late colleague. Our superiors informed us that when the next fiscal year rolled around in four months, there would be enough money in the budget to translate us back home.

II.

On first arrival, Adwellia seemed much like home. We landed on the shores of a small lake in what seemed to be a wooded area. Since it was getting dark, we soon set to pitching camp for the night. Our first unpleasant surprise was that the kerosene heater we had brought with us wouldn’t work, leaving us cold and disheartened. Lachlan collected some logs to build a fire, but our matches didn’t seem to work either. I remembered the seventh page of your paper, where you had posited that an imaged world would run on the same physics of our own world, since it would be bound by the expectations of the imager. Dr. Adwell had certainly understood enough chemistry to know that matches should start fires, but it seemed one of our most basic predictions had already failed.

I will not say whether we were more motivated by curiosity or by the bitter cold, but we tried dozens of different branches – small, large, young and green, old and rotting – and everything from dousing them in kerosene to the old-fashioned method of rubbing sticks together to create friction.

Finally, I succeeded in getting some branches from an old fir tree to alight. In relief, the two of us huddled close to the fire. But our curiosity was only heightened when we found the area near the fire to be unmistakeably *colder* than the surrounding air. Here our chill overcame our scientific spirit, and we decided to deal with the problem in the morning. We got into our too-thin thermal sleeping bags and passed a miserable and freezing night.

When we awoke, the fire had gone out, and in its place stood a pile of hats – twenty of them, to be precise. I would have called them fedoras, although Lachlan said the particular style was more popularly known as a Homburg. We debated taking the hats, but we had been thoroughly spooked. Instead we picked up our camp and journeyed south, where it looked like the wood was beginning to thin out.

Around midday we spotted smoke, and dared to hope we were coming upon a settlement. By evening our guess was confirmed, and we saw a village of conical adobe huts. We prepared to gesture our request to trade trinkets for lodging to the inhabitants – who were far too dark skinned to be European but who did not quite pattern-match to my memories of any particular human race. Imagine our surprise when we found they spoke English – though with abominable grammar. The headman introduced himself as Somon, and was all too happy to accept our trinkets in exchange for a nice warm hut to spend the night in.

We endeavored to learn more about these people in the morning, but by this time were tired enough to call it a night. We could not help inspecting the heating mechanism in our room, which seemed to be a mud bowl in which sheaves of wheat, small rocks, and little mud figurines that looked like people had been placed. Totally absent any visible mechanism, the setup was emitting heat – and what was more, a ball set in a track along the edge of the bowl moved continuously around in what seemed to all the world to be perpetual motion, making an annoying crackling sound as it passed over little leaves set in the rim. We had only a little time to exchange theories before falling into a deep sleep.

The next morning, the bowl was no longer warm, the ball had stopped moving, and the objects within had apparently transmogrified into a miniature wheelbarrow. This was strange magic.

The villagers were already were already up and about, so we found Somon and tried to get some better conversation in.

“We are scientists,” we told him “from far away, looking to gain a better understanding of how things work here.”

“Here in Mogonaw?” asked Somon, using what we later found was the name of the village. “Not well.” He smiled, showing very pearly teeth.

“We were hoping to set up a laboratory – a few metal huts and a big machine – maybe on the outskirts of town. We would pay you for food, maybe for help with certain things. We have many tools to trade, and lots of gold and metal.” Not exactly true – what we had was a portable nanofactory, translated in with us as an easier alternative to bringing supplies. But we could get tools or transmute elements pretty quickly.

“Is of course,” said Somon, with the delight of someone who had stumbled entirely by accident into a beneficial arrangement. “What will you be needing?”

“Well the first thing,” interrupted Lachlan, “is we wanted to know how your heating device works. The one with the wheat and the rocks. It was new to us.”

“You not have this in your village?” said Somon, with a frown. “Is not obvious?”

“No,” I said. “Where we come from, it’s not obvious at all.”

Somon brightened. “Your village,” he declared “not know true names!” He picked up a rock from the ground. “True name of this is…rock.”

We both nodded, mystified.

He grabbed a sheaf of wheat from a passing villager, who gave him a glare. “True name,” he said, “is…wheat.”

He said it with the same mystical intonation with which one of our colleagues back at the laboratory would announce a particularly earth-shattering result.

“Yes, okay,” said Lachlan, kind of miffed. “I actually think we do know true names of things. It’s the same in our language.”

Now it was Somon’s turn to be mystified. “Then…where is confusion?”

“The heating device,” said Lachlan, narrowing his eyes. “How does it work?”

“Is obvious!” said Somon, like we were idiots. “Wheat and rock and art become work and heat and cart. The work push little ball around. Then ball make noise, continuing reaction.”

“But…” I interjected, because it looked like Lachlan wanted to grab the headman and wring his neck “*why* do the wheat and rock and art become work and heat and cart.”

“Is true names” said Somon, and shrugged.

“Holy shit,” said Lachlan, at exactly the instant when I remained just as confused as I had been before. I stared at him.

“Holy shit,” Lachlan repeated. “This world fucking runs on anagrams. English language anagrams.”

Wittgenstein once said that the limits of our language are the limits of our world. Some say that mathematics is the language of God. Maybe that was why our world ran on math. Well, English had been the language of Dr. Adwell. It had been the lens through which he made sense of reality.

Maybe our hypothesis that his imaged world would run on the same physics of our own had been premature.

What if his world ran on English?

“The fire!” said Lachlan, who as usual was a step ahead of me. “Fir branches and heat. Fir plus heat becomes fire plus hat. So it removed heat from the atmosphere, and created fire and a hat.”

“Twenty hats,” I reminded him.

Lachlan was already deep in thought. “It’s all stoichiometry,” he started saying, almost faster than I could follow. “In our world, water is H20. H-O-H. Here, a fir tree has to be literally made of F-I-R. Twenty six letter-elements, forming a near-infinite amount of word-molecules. Suppose we burned three kilograms of fir branches…don’t know the molar weight here, but suppose each letter weighs the same and there’s one mole per kilogram, just bear with me. That’s one mole each of F, I, and R. So it must have absorbed some sort of four mole equivalent amount of heat…whatever that means…and then spit out three moles of hats and four moles of fire. Three moles of hats in this system would be three kilograms of hats, that would mean each hat weighs 150 grams…it all checks out! Somon! Quick! Show us how you make something else!”

Somon looked at him. The headman seemed as confused as I was, but for different reasons.

“Make…what?” he asked.

“I don’t know. Clothes, tools, anything.”

“My daughter Genea live in here,” he said, gesturing to a hut on the outskirts of town with some smoke coming out of it. “She is weaver.”

The “weaver” actually seemed to be performing some sort of complicated chemical reaction. She was holding beets over a cauldron that was bubbling up into a primitive fume hood, then throwing them into what seemed like a vat of tar. Water was running out a hole in one side, and on the other, a roll of cloth was getting steadily longer.

This time I got it before Lachlan. “Chlorine,” I said. “Chlorine plus beets plus tar becomes cloth plus brine plus tears.”

“That’s not right,” said Lachlan. “You’re missing an ‘e'”.

“No I’m not,” I said. “It consumes twice as much tar as chlorine or beets, and produces twice as many tears as brine or cloth.”

“I think,” said Lachlan, “that we had better get our laboratory set up sooner rather than later.”

III.

This we did, at record speed. Not wanting to frighten the villagers – or expose ourselves to prying eyes – we set ourselves a kilometer south of town, on a cape overlooking a great sea. On the headlands of the cape was a small hill from which you could see for miles, and there we completed the week-or-so's work of getting the nanofactory up and running. Its first job was to extrude us two aluminum Quonset huts, which became our homes away from home.

From our little encampment the ocean stretched on as far as we could see. I wondered if there were other continents on this world – figuring out its size really should have been one of our first priorities. But we were too fascinated by this world's weird linguistic elements and reactions – anglophysics, we dubbed them – to properly investigate anything else.

The first and most obvious question was why everything wasn't reacting all the time. How come every time someone touched a rock, the skin + rock didn't become corks + ink? Just the air alone should have destroyed a wide variety of objects.

("Oh, come on," I told Lachlan. "The air doesn't count". Lachlan had then gone on to prove me wrong by getting the iron tools we had brought to rust, then proving the rust happened faster in moist air, and air that was full of dust particles. "AIR plus IRON plus DUST," he told me "equals RUST plus IONS plus ARID. Things aren't rusting in this world because of oxidation. As long as it can suck dust and moisture from the air, it's rusting by Crazy Anagram Logic." So the air definitely counted.)

The first thing we discovered was that nature abhorred non-words. AIR and DUST wouldn't react on their own to become RUST and IA, because IA wasn't a thing.

“What about AI?” asked Lachlan. “Why not rust plus an intelligent computer?”

At the time, my answer was “Shut up! The world might hear you!” I would later learn this was not nearly as funny as I thought.

But at the time, we made quick progress. Simple materials and short words seemed to be most stable, with complicated or abstract concepts rarely forming spontaneously – which, at least, answered our AI problem. And reactions usually wouldn’t happen at all without sound, which seemed to play the same role in this world that heat did in our own. Lachlan had suspected this almost from the beginning – that the crackling leaves underneath the ball had provided the sound-energy to continue fueling the reaction that kept us warm that first night. But it wasn’t until we heard the cacophony of a village festival that we knew we were on the right track.

“WHAT ARE YOU DOING??!” I had yelled at Somon, over the din of drums and cymbals and screaming villagers.

“MAKING BEER!!!” Somon answered.

It had turned out that the villagers used pee and bran to produce beer and pans, but that the reaction went unpleasantly slowly unless they shouted it along. The shouting was, of course, egged on by the beer they had already produced, which sort of made it an autocatalytic reaction if you squinted. They offered us some of their beer, but even though I knew things worked differently here my standards were a little too high to drink beer *literally* made of pee and so we returned to the lab. On our trip back, Lachlan pointed out that all of the villagers’ iron tools had been carefully taken inside during the festival, so that the noise would not cause them to rust.

Our next big discovery was a week later. I woke up at 7 AM with Lachlan pounding on the door of my aluminum hut.

“OMAR!” he was shouting. “TAKE A LOOK AT THIS!”

Sitting on his palm was a one inch tall man, naked and hairless, looking terrified. He looked like he would have run off if there was anywhere to run to.

“What in the…?”

“I found a volcanic vent, up in the hills to the west. There was a source of methane. I broke it down into HEAT and MEN. But there wasn’t enough MEN to form someone full sized. So I got this.”

“Lachlan, you’ve got to help him!”

Lachlan gave a grunt, as if annoyed to be reminded of the ethical implications of his work. “How?”

“Can you speak language?” I asked the little man on Lachlan’s palm.

In response, the man screamed. I took that as a no.

So I dragged Lachlan down to the village, where I woke up an annoyed Somon. “Somon,” I said. “We found a way to break methane into…”

Somon’s eyes went wide. Then he got angry. “No methane!” he said. “Is taboo! Will…”

He saw the homunculus in Lachlan’s palm. With a deft motion belying his age, he yanked the little creature away from Lachlan and snapped its neck. I gasped. Lachlan looked annoyed.

“Is TABOO!” shouted Somon, with an anger I hadn’t seen in him before. “These things! Not men! No speech! No mind! Must not make! Little man is taboo! Methane is taboo! If you make little man, no longer stay with us!”

I calmed him down, promised we wouldn’t be doing any more experiments with methane, said we were new here, didn’t know what we were doing. I asked him for more advice, asked him about any other taboos. He seemed irritated, assumed we should know what they were, seemed to think less of us with each question indicating our ignorance. Finally we gave up and made the long trek back to our laboratory.

Our next few weeks of experiments were less bloody, but still exciting. Suppose we took a mop and the guts of an animal, and shouted at them until MOP + GUT reacted to POT + become GUM. Would the pot be the cooking implement, or would it be marijuana? For that matter, why shouldn’t it be a top, the child’s toy? Why shouldn’t the gum form a mug, fit to drink coffee from?

In our first experiment, we surrounded our apparatus with pans and food, and were unsurprised to find we ended up with cooking implements. We repeated the experiment, but this time surrounding the apparatus with bongs, tobacco, and other drug paraphernalia – this time we got marijuana. We wanted to get a playful child to see if we could produce tops, but news of our work with methane had gotten out and spooked the villagers, and they were understandably unwilling to let us borrow one of their children.

The third experiment was in my opinion the key to this entire process. This time we surrounded the apparatus with pans and food, but both Lachlan and I concentrated very very hard on marijuana, and talked about marijuana with each other while the loudspeaker the nanofactory had extruded blasted sound at the reactants, and sure enough, we got marijuana.

Somehow our expectations were guiding the physics in a way that the letters themselves couldn't. I started to wonder what had become of poor Dr. Adwell. Was the god of this world a deist, who had created it shortly before dying in a hospital ICU in a very different planet? Or was he in some sense still here, still actively guiding things?

The reaction that rusted iron started to seem more and more suspicious. What about that ARID? In our experiments, making adjectives had been almost impossible, requiring more sound catalysis than any noun we had encountered so far. But ARID seemed to form of its own accord. What if Adwell somehow remembered that iron was supposed to rust, and *privileged* that reaction as the sort of thing that ought to go on? What if the reason everything didn't implode upon itself was Adwell ensuring that everything in his imaged world happened according to some plan?

Then our proof that we could alter our results through concentration and careful priming would take on a whole new meaning.

Did *reminding God what chemical reaction we wanted* change experimental results?

IV.

"We're going about this half-assedly," Lachlan told me one morning our sixth week in Adwellia. "All of this looking for clever anagrams is taking up too much of our time, delaying us in supremely great work. We need to do this analytically. Get a bottle of As, a bottle of Bs, so we can create whatever the hell we want."

This proved easier said than done. We got the nanofactory to extrude us a very complex apparatus, a centrifuge, and what we took to calling the "sonic ray" – a machine that made deafening noise along a very narrow arc and which could catalyze reactions much faster than shouting or drumming. It turned out to be the key to making far more complex products than we had previously attempted. But our first use was a plain and simple failure.

We had decided to start with granite, which we would break down into tin, rags, and the letter E. We would then centrifuge the decay products, with the three-letter tin and rags going one way and the pure E going another.

Nature, remember, abhors non-words. No sooner had we forced some E into a test tube than the tube itself transformed in a great explosion to gelatin and a tiny, near-microscopic donkey. E + GLASS = GEL and ASS. We couldn't say we couldn't have seen it coming. It could have been worse – I was just glad that Dr. Adwell's ascended mind's first association with the latter word was "donkey".

We tried the experiment again with a zinc vial – zinc because it was implausible that there was an ZINC + E anagram lurking out there – and ended up with a mat of eels. Through this whole time, we had been debating the problem of ambiguity – who was to say that our granite was GRANITE rather than ROCK or even STONE – and the answer seemed to be that Dr. Adwell – or whoever was watching Upstairs – was mostly sympathetic to our efforts. Well, the sympathy ended when we started trying to isolate single letters. ZINC became METAL and thence EEL MATs.

Our effort with mud was even worse. We put a lot of time into making sure the mud we got was very clasically mud – not ooze, not muck, certainly not dirt. And there was no good way MUD + E was becoming anything. We turned on the device.

The Es disappeared. Seriously. Granite went into the centrifuge, tin came out, but there was no sign of an E anywhere, and rather fewer rags than usual.

“This is really weird,” I said.

“Thanks, Einstein!” said Lachlan. “I never would have figured that out without YOUR FUCKING COMMENTARY.”

I should have told him to calm down, but the experiment had upset me too. “Well it wasn’t MY BRIGHT IDEA to try to ISOLATE ALL THE LETTERS,” I said. “WHICH REMINDS ME! IF YOU THINK I’M GOING THROUGH THIS TWENTY FIVE MORE TIMES, YOU CAN GO FUCK YOURSELF!”

Lachlan swung at me, missing by an inch. I kicked him, right in the knee, and he fell into the experimental apparatus, knocking the whole thing over. Both of us went down with it. For a second, the sonic death ray shot straight at us – EEEEEIEEEEEIEEEEIEEEIE! and then its safety kicked in and it turned off. We sat there, stunned, bruised, in pain.

“Rage,” said Lachlan. “GRANITE becomes TIN plus RAGE. Holy fuck, we created an emotion.”

It had happened before, sort of. The wheat and rock and art, they had come together to produce work, which was an abstract concept. But it was still in the domain of physics.

“Work” seemed like the sort of thing that could come out of chemical reactions, kind of like heat. But rage? This was something really new.

That night, we made the short trek into the village and asked Somon what he thought.

“Rarely,” he said. “Sometimes, when festival is very loud, strange things happen. Should avoid. Very bad. This is taboo.”

The next week, I knew something was up. Lachlan was missing our daily debriefings, not getting any work done. Finally I broke the most important unwritten rule of our little community. I went into his aluminum hut without knocking.

There he was, sitting with a blissed out look on his face. Beside his bed sat a miniature version of our experimental apparatus, complete with its own sonic death ray – he must have privately ordered it from the nanofactory, then deleted the records. It was reacting little tchotchkes from the village – dolls, balls, play swords – with our glass specimen jars. Tar was streaming into the waste bin.

I turned off the sonic ray. Lachlan awoke with a start. He seemed about as angry as he’d been the time we accidentally produced rage from granite, but this time I knew he had a less noble reason.

“What the fuck are you doing, barging in here like this?”

“You’ve gotten yourself addicted,” I said. “Addicted to joy.”

Lachlan didn’t deny it, as his TOY + JAR -> JOY + TAR reactor was right there.

"Look," he said. "It's been two months now, stuck in this stupid world. It's going to be another two before the lab brings us back home. The villagers are crazy, physics runs on English, and the nanofactory can't produce any entertainment that's remotely entertaining. The letter isolation project is a failure, you no offense are one of the most boring people I've ever met, and when I try to get some of the village women to look at me they murmur something about taboos and give me the cold shoulder. Give me a break here, Omar!"

"Lach," I said. "You're neglecting your work. We still haven't gotten anywhere near the bottom of anglophysics, let alone figured out the most basic stuff about this world like how big it is. You sitting here blissing out on raw linguistic joy isn't something we can afford right now."

"Fuck you," said Lachlan, but he didn't protest as I picked up his mini-apparatus and brought it to the nanofactory's disassembler chute, nor as I reprogrammed the nanofactory to make sure all its records would be public from now on.

V.

A week after that incident I finally got the nanofactory, with great creaking and protesting, to extrude a small aircraft so I could explore the surrounding area. The villagers were delighted, having never seen anything similar, and several of them demanded rides – increasing our popularity a little after the methane debacle. When we were done appeasing the natives, I took off and started mapping Adwellia.

We seemed to be at the southernmost extent of an island about three hundred miles east to west and twice that north to south. The island was mostly forested except for the broken volcanic area nearby where we had gotten the methane and some hills further north. Four hundred miles east of us there seemed to be another continent or large island, but that was about the limit of my range and so I told myself I would explore the new land another day.

The distances allowed me to do some geometry and calculate the size of the world. Adwellia appeared to be a spherical planet about the size of the Earth. As far as I could tell it had one sun and one moon, and there were normal stars in the sky. It seemed to get colder further north and warmer further south, though I wasn't able to fly far enough to confirm it had proper poles and an equator.

By the time I finished these explorations, about a week after they began, Lachlan had developed a new obsession.

“I can't solve the letter isolation problem,” he admitted. “But someone else can. Someone like Einstein.”

“Great,” I said, sarcastically. “All we need is…”

Then it hit me. Surely he wasn't that crazy.

“Yes,” he said. “Why not synthesize Einstein? Or some other brilliant scientist who's more creative than we are. I've been going through the dictionary looking for proper combinations. It's not that hard.

This proved optimistic, but the equation upon which we eventually settled was STONE + TIN + FORT = EINSTEIN + FIRE. The only difficulty was obtaining the fort, since the villagers here did not seem to be of a militaristic bent, but I had found some ruins further north during my explorations, and one of them did indeed seem to be an old stone fort, perhaps constructed by the villagers' ancestors. I proposed we get a party of villagers to help quarry fort material, but Lachlan objected that they would probably just have some stupid taboo about it, so instead I landed there with the aircraft and laboriously ferried fort parts home in twenty pound increments, on my lap.

Once we had enough fort to stoichiometrically produce Einstein, getting the stone and tin was easy. But getting the reaction to work proved impossible. No matter how many physics books we stuck around our apparatus, no matter how hard we concentrated on the great scientist, the reaction spat out absurd things like ferns, nits, and a tooting sound – or forests, nits, and one ton weights, or a nose with a tit in the front, which trust me was *really* awkward and which we threw into the nanofactory disassembler chute as soon as we could, believe you me.

After about thirty tries, Lachlan announced that the problem was obvious. You see, we needed a *capital* E.

I grudgingly admit that, even after two months in a world where stone was composed of S, T, O, N, and E, the though that there were different atomic units representing lowercase and capital Es seemed absurd. But as always, my sense of impossibility surrendered to crazy reality and I figured that Lachlan was probably right. We needed a capital E.

Two days later, Lachlan showed up at the laboratory with a very suggestive looking sack.

“Lachlan, what were you just out doing?” I said, hoping the answer was anything other than what I knew it was going to be.

“Just grave robbin'” he answered. “I got us the corpse of a lady named Eder, who died of pneumonia yesterday. Don’t worry, no one saw me take it.”

"Oh, come on," I said. "When they find the grave disturbed, who are they going to suspect? The other villagers, who they have known their whole lives? Or the mysterious strangers on the storm-wracked cape outside of town who have already violated their sacred taboos. Lachlan, *you* are a fucking idiot."

"Maybe I am," said Lachlan. "But if I'm so stupid, good thing we'll have Albert fuckin' Einstein around to help provide some brains for this operation."

The new equation was EDER + TIN + SNAIL = EINSTEIN + LARD.

So God help us, we hired some villagers to collect snails for us, and when we had hundreds, we poured poor Eder's bones into the reaction chamber along with the snails and some tin and started the sound.

And Einstein started to grow. At first he was tiny, smaller than the methane-men in Lachlan's palm had been, no bigger than the snails that surrounded him. But as bones and metal and snails slammed into him, he grew bigger, all the while screaming and covering his ears as the sonic ray did its gruesome work. We saw him, child-sized, beating up against the glass wall of the reaction chamber, ever growing, ever screaming.

"You're mad," I told Lachlan. "We've got to stop this."

"Maybe I am," said Lachlan. "But think! Einstein! The greatest scientist in recorded history! Think what we could do! Revolutionize not only our study of Adwellia. But we could bring him back with us, get the lab to translate him as well as us. We could turn Adwellia into a genius factory that would revolutionize civilization back on Earth. Omar, this *has* to be done! The potential in anglophysics makes a Nobel Prize look like a tee-ball trophy."

When Einstein was fully formed, and released from the reaction chamber, he attacked us. We subdued him, using weapons extruded from the nanofactory, and kept him in a cell. For three days we tried to talk to him, and he responded by screaming wordlessly at us and spitting in our faces.

I don't know whether there was something theological going on – whether Einstein was just a homunculus lacking a true soul. Or whether it was just very simply that our Einstein was psychologically an infant, that no one had taught him so much as language let alone physics, and that Adwell or whoever was up there wasn't going to assume we meant "the smart Einstein, who knows lots of stuff" in the way we wanted.

Our Einstein was a giant infant, not even an infant, a fetus that should never have been born. On the third day, by mutual consent, we stuck him in the nanofactory disassembly chute and resolved never to speak of him again.

VI.

That was the last time I worked together with Lachlan on anything of note. After that we retreated to our separate aluminum huts, acknowledging each other only when our paths crossed on the way to the nanofactory for some crucial part.

I found him creepy. He *was* creeply. And he thought I was holding back our research. Maybe that was true too. In either case, it was a terse nod, a couple of words, and the tacit acknowledgment that it wasn't worth resolving our hostility in the month or so we had left before we were transferred back.

I spent that last month trying to build on my theory that Adwell's mind was somehow working behind the scenes running everything. The catalytic property of the sound, I theorized, was its ability to *get Adwell's attention*. It was a sort of "HEY, GOD, LOOK OVER HERE, WE'RE DOING SCIENCE, BETTER APPLY THE LAWS OF PHYSICS RIGHT AWAY". I know it sounded bizarre, but my early experiments bore me out. Rapidly flashing bright lights seemed to speed reactions almost as well as sound. So did – because sometimes the simplest solution is the best – shouting "ADWELL! LOOK OVER HERE!"

With these advances, once again entirely new classes of reaction became possible. No longer were we limited to the highly reactive simple materials with short names. Long strings of words, complex abstractions, even adjectives came within our reach. It was exciting.

But once again, it was Lachlan who was really pushing the frontiers. One night he started banging on my door: "OMAR!" he shouted. "I DID IT!" When I went out he practically dragged me into his hut, which was nearly piled, floor to ceiling, with papers that turned out, on inspection, to be various IQ tests the nanofactory must have been carrying in its databanks.

"What did you do?" I asked.

"I couldn't create Einstein," he said, referring to the still-fresh debacle – "so I decided to turn myself into Einstein! Look! I'm producing SMART. And it's working!"

His sonic ray – now only a fraction of the power of my own multimodality parasonic device – was reacting smoke and carts into coke and, apparently, smart. A complicated system of tubes and centrifuges was catching the smart and binding it into a containment chamber linked to a helmet. Clearly someone was supposed to put it on.

“And you’re saying it works?” I asked.

“The IQ tests don’t lie,” said Lachlan. “I was 152 two weeks ago. Now I’m consistently getting in the 160s.”

Judging by the number of tests, he must have been obsessively checking his numbers every hour or so.

“Now,” he said, “I’m going to try that letter isolation thing again.”

I judged by the shouts of rage and frustration I heard over the next few days that it wasn’t working.

Two days later, Lachlan asked me if he could borrow my advanced parasonic ray. I refused. That evening, it went missing for about three hours before turning up on top of my desk. I noticed Lachlan now had one exactly like it.

I soldiered on. In between my experiments, I played a little game predicting what Lachlan was trying to synthesize by the objects he took from the nanofactory and the supplies he ordered brought in from the village. One day it was buckets of dew, carts full of animal legs, and an entire cage of live minks – my best guess was he was trying to get KNOWLEDGE, but I couldn’t get the stoichiometry to line up. Judging from his screams of frustration that night, neither could he.

The next week, it was load after load of potatoes, fence posts, and a tank of minnows. It took me half an hour to come up with OMNIPOTENCE, even though once I made myself start thinking like Lachlan it was obvious.

I started to become worried.

One day, three months and two weeks into our mission and only fourteen short days before we hoped the laboratory would re-establish contact, I went out for a sortie with the plane and came back to find a disaster area.

Our huts had been smashed open. The nanofactory had big dents in its aluminum casing. Inside, all my lab equipment had been broken, my papers thrown on the floor haphazardly.

I went into Lachlan’s hut. IQ tests everywhere. He was missing. So was his parasonic ray. I figured they had grabbed my partner in his sleep, before he’d had time to resist. In retrospect we really should have put up some defenses, but we hadn’t expected to need them.

The nanofactory was still online. It was pretty hard to break – especially if, as I suspected, the vandals were villagers armed with clubs and rocks. I told it to extrude me some overwhelmingly powerful weaponry. After making me wait an hour, it gave me a ring that upon threat would instantaneously unfold into a device that generated an invincible barrier around the wearer, plus a hand-held matter disruptor. Thus armed, I walked into the village and found Somon.

I didn’t have to bring up the subject of Lachlan. “Is evil man!” the headman told me, as soon as he saw me. “Broke taboos! Created life! Dug up grave! And today! Today was worst! Kidnapped my daughter, Genea! No more okay! Tonight gets beaten! Tomorrow dies!”

Raising my invincibility shield, I wandered into the public square. There, whipped bloody and tied to a post, was Lachlan.

“You kidnapped the headman’s daughter?” I asked him. I didn’t even give him the dignity of pretending to doubt whether it was true.

Lachlan smiled. “Genea. A perfect name for my reaction. I could have been a Genius, with a capital G.”

I don’t know if it was that smile, or the blood all over him, or the lack of remorse in his voice, but at that moment, I’d *had it* with Dr. Lachlan Fairchild. I lowered the matter disruptor.

“You know,” I said. “That is it. I’m not even going to rescue you. You’re a menace.”

“You don’t have a choice,” said Lachlan. “I have a nuke. These people don’t understand the concept, but lucky we’ve got a genius like yourself. Let me go or I blow this entire planet sky high.”

“Even if you managed to extrude a nuke,” I said “which you didn’t, because I checked the nanofactory’s public records before I left – even then, nukes don’t work in this world. Nuclear fission isn’t an anagram of anything.”

“A metaphorical nuke,” said Lachlan. “I mean, I’ve figured out this world’s equivalent of a nuke. It’s very clever. Without the SMART, I never would have been able to think of it. I’ll…”

My best course was to immediately, like split-second immediately, raise the matter disruptor and shoot Lachlan. I could do it before he had a chance to react, and it would solve the whole damn problem.

Instead I took the worst course, which was to raise the matter disruptor, obviously intending to shoot him, and vacillate at the last moment because I'd never killed anyone before and I wasn't sure I had it in me and instead of finding out my brain wanted to sit and ponder this for thirty seconds.

Lachlan took a ring off his finger and it unfolded it to reveal his parasonic ray. Then he furrowed his brow in concentration and it let out a screech.

I shot the matter disruptor. Man, post, and town square changed into their component atoms…letters…whatever.

The villagers ran, screaming. Some of them ran away from the explosion. Others ran towards the explosion, trying to see what had happened and maybe defend their homes and families. A few arrows and stones came towards me, causing my ring to near-instaneously unfold into a weird backpack-like device that placed itself on my back and surrounded me with a purple glow. The projectiles hit my new invincibility shield and fell to the ground limply.

I calmly walked through the carnage. I was heading back a kilometer south, back to the cape. I was going to extrude a larger aircraft, bring the nanofactory a few hundred miles away, and wait out the last two weeks of exile far away from this mob.

The ground started to shake. I realized the explosion had ended long ago, yet its deafening roar had not subsided.

I looked back to the town square and my blood turned cold. In the center of the blast radius, where not even dust should have remained, there was Lachlan's skull, set in the biggest rictus grin I had ever seen.

I raised the matter disruptor and fired another shot. The skull disintegrated. But Cheshire Cat-like, somehow the grin remained, even larger than before, a smile without a substrate.

This was bad.

I started to run back to the lab. Cracks opened in the ground around me. The roar become worse. Was it just me, or was the sea getting closer?

Metaphorical nukes. A nuke was at the most basic level a chain reaction. Neutron produces energy plus neutron. That neutron produces energy plus neutron. *That* neutron and so on. You end up with a *lot* of energy.

I could see the remains of the looted lab now in front of me. It was on its elevated headland reaching into the sea, and I was afraid the rising water was going to cut it off and turn it into an island before I could get to it.

Sound drove chemical reactions in this world. Anything that could create sound had the potential to be a chain reaction if the reactants were common enough. You could get most of the letters of “sound” from…oh, that wasn’t good.

The cracks in the GROUND got bigger as the low-lying GROUND started to sink further beneath the waves.

I stared back at the village. It was almost entirely underwater now. Above it was Lachlan’s disembodied grin, now the size of a skyscraper, hanging in the sky.

Sound, ground. Grin. Sin. There. I had it. GROUND + SIN = SOUND + GRIN. The nuke. The ground was essentially limitless until the world was destroyed. The more ground was destroyed, the more people died, the more villages sunk under the waves. A sin. A reaction that created its own reactants. And sound. Created its own reactants and its own catalyst. Leaving nothing but Lachlan's gigantic triumphant grin, hanging in the sky over the world he was destroying.

I groaned as a crack in the ground took the aircraft on its field. It teetered for a second, then fell into the onrushing waves. I ran through ankle deep water and at last reached the top of the headland. There was just a small area of land left, on the highest ground of the cape, with our two little partially-smashed huts and the bulky dented aluminum nanofactory.

"Extrude boat!" I commanded the nanofactory.

"Extruding boat," said the display. "Estimated creation time with material on hand, two hours."

"Cancel! Cancel cancel cancel!" I shouted, but the factory had gotten into its extrusion mode and wasn't listening.

I ran into my hut. Most of my stuff was still broken. There was nothing that looked like a good flotation device, unless you counted my mattress. My reaction apparatus, my parasonic ray, and a few doodads.

I grabbed the ray gun and ran outside. Even on the high ground, there were wavelets lapping at my shoes. I had about a minute before I drowned.

"Okay," I said to myself. "Time to figure something out. Time to create a boat." And there was only one good reactant on hand.

OCEAN + …no, that wouldn't work. SEA + …that was even worse. WATER + … I might be able to use water if I let the reaction consume my bones…WATER + BONE = BOAT +

NEWER … no, even with the parasonic ray I'd never be able to catalyze a reaction that made a comparative adjective of all things. Maybe if I had an hour to think of some useful intermediates.

Okay, back up. You don't need a boat. You can use a ship. Ship is…

My brain was in panic mode. It didn't want to anagram SHIP. What it wanted was escape.

The cape! The cape could provide escape! The cape and the sea! The two things I had! And my parasonic gun was just strong enough to let me synthesize abstractions. I just needed somewhere to put that extra A.

WATER + A = AWARE + T. No, Nature abhors non-words, T won't work. WATER + A = RAW TEA. No, adjectives took forever. WAR TEA? I wasn't sure what would happen if I caused a war at this point, but I bet it wouldn't be good.

A wave rushed over me and I rose to the top sputtering and gasping. I still had the parasonic ray. The water had almost covered the huts now. Borne on the receding wave came Lachlan's stupid piles of IQ tests, now soaked.

CAPE + TEST = ESCAPE + 2*T

On the one hand, Nature abhorred non-words. On the other hand, I couldn't swim and was about to drown. I concentrated REALLY hard on the reaction, turned the parasonic ray to its highest setting, and shot a beam of sound and strobe light and repetition of the name "Adwell!" at the pile of tests and the rocky cape below.

Nothing happened.

The LOW CHARGE light began to flash on my parasonic ray.

It had been a stupid, desperate gambit. I'd already known I didn't have enough energy to do a reaction that created non-words, didn't know if that was even possible with *any* energy, and I had just drained my parasonic ray of almost all its charge I had made a terrible error.

"Error!" I shouted. "That's it! Adwell! Error!"

CAPE + TEST + 2*ERROR = ESCAPE + 2*TERROR

As I fell under the waves, with my last breath and last bit of charge I fired off the parasonic ray one last time.

It's not working I thought to myself. *It's not working and I'm going to die, lost under the sea, dead forever.* I spent half a minute just thrashing about in terror before I realized *that meant it was working*.

The water was receding! A bubble of air was spreading away from me in all directions as the water was consumed! I was saved! Still terrified, but saved!

…then the water started closing in on me again. I didn't know what what was happening. I'd done it, hadn't I? Succeeded in creating a reaction that would get me out?

Success! That was the problem! If I had succeeded in creating a reaction, then firing the parasonic ray hadn't been an error. The reaction couldn't take place. The water closed in on me again. I was going to die.

The water started to recede. If the success of the reaction prevented me from having made an error, then the reaction wouldn't work, and starting the reaction was an error, and so the reaction could take place. All this I saw clearly, as in a dream, from within my bubble of air.

The air bubble under the rising seas (sinking ground?) reached a size of about twenty meters, large enough to cover the cape and the two huts and the nanofactory, and then stopped, occasionally shrinking a little or growing a little, always seething, starting to burn with a weird energy.

From within the anglophysical terror clouding my mind, I recognized the problem as a novel version of the Epimenides paradox of self-reference, implemented on a physical substrate. If my initiation of the anglophysical reaction had been an ERROR, then I would ESCAPE, and it hadn't been an ERROR after all. But if my initiation of the reaction had not been an ERROR, then I would not ESCAPE, and in fact it *would* have been an ERROR.

I had a vague memory that I had once discussed Russell's Paradox with Dr. Adwell. I wished I could have remembered what he said.

The interface between air and water became turbulent, started to glow. I saw fantastic images projected upon it, weird fractal geometries, strange supersensory stimuli that somehow reminded me of Lovecraft's references to the beckoning piping from the void behind space. All the while the TERROR grew, and the bubble began to vacillate wildly.

Then there was a great pop, and I thought for a second my air bubble had popped, but more correctly *everything* had popped, and for a second the things that were nothing like piping sounds became unbearable. Then I found myself lying, still terrified, on the floor of the translation chamber of our laboratory, the very same place where I had entered Adwellia almost four months before.

VII.

When I had recovered my senses and debriefed my colleagues, I devised three theories for what had happened there, on the cape.

First, that my reaction had been successful beyond my wildest dreams, the paradox had resolved in my favor, and I had ESCAPED not only to firm ground but to my own home dimension.

Second, that the paradox had been so confusing and unbearable for poor Adwell that he had expelled me from his consciousness, like a man brushing a bug off his skin, and having been kicked from his world I naturally defaulted to my own.

And third, that implementing a paradox on a physical substrate was *really, really bad* and I had destroyed Adwellia.

This last possibility ought in theory to be testable, but I was informed upon my return that the budget was tight this year and that the necessary supercomputing resources to search for Adwellia will not be available for some time.

I have been assigned to another project, and although my superiors have thanked me for my work in Adwellia, I am certain they do not believe a word of my report and have written the entire expedition – and perhaps their decision in hiring me – off as a loss. In their place I would not do otherwise.

But from your writings I gather you are a man of unusual intellect, and some of your speculations come uncomfortably close to the truth. I do not know whether you have pursued your interest in Berkeleyan idealism further, but if you are so gracious as to believe my story or at least keep an open mind, I would be interested in further correspondence with you about the implications of anglophysics for future imaged worlds and how the consistency of such images might be assured against paradoxes of self-reference and other threats to their integrity.

Yours sincerely,
Dr. Omar Reyes, University of ________

PS: I hope you will be understanding when I say that I wish to restrict my future work in the imaged world field to a purely theoretical level.

[EDIT: I apologize to those who have read Universal Fire for this story. As a peace offering, please accept this lovely lampshade.]

[EDIT 2: HPMORPodcast has recorded an audio version.]

My New OKCupid Profile

http://www.okcupid.com/profile/Aithrobates

(read it before you accuse me of narcissism or opportunism by linking to it here)

Housekeeping Post April 2014

1. I am on an inpatient medicine rotation for the next two months, which means I will be working about twice as hard as normal. The last time I was on one of these, I managed about a blog post a week. I am going to try to do better this time by posting much shorter things and by splitting up link posts into short posts with a couple of links each. I might also cheat by reposting some stuff from my old blog. Or I might just disappear for two months.

2. Many people noticed Ozy deleted zir blog and are wondering if ze's okay. The answer is yes. Ze had a brief panic attack of the "I never want to interact with anyone ever again" variety, as one does from time to time, but is mostly recovered now.

3. Please do not send me blog comments by email. When many people try this, I end up with lots of email. This creates guilt about not responding in a way that comments don't, and makes it harder to find important emails among comment-related ones. If you want to comment without your opinion being public, you can comment anonymously. If you want to comment without your opinion being public but with me knowing it's you, you can comment under a pseudonym and email me once to tell me what the pseudonym is.

4. If you need advice on a medical problem, I am a terrible terrible terrible person to ask. If you need advice on a psychiatric problem, I am a single-terrible person to ask. My advice will probably be "see a psychiatrist" – not because I don't want to help you, but because I can't prescribe drugs to you over the Internet and they can. I would like to write

up some advice for people with psychiatric issues, but I guess not in the next two months.

5. I had a list of things to remember to include in this post. Number five was “Facebook”, but I can’t remember what it was supposed to mean. So please be aware that something has happened regarding Facebook, or you need to do something on Facebook, or that Facebook is important in some way. Thanks.

Dermatology Pop Quiz

My dermatology lecture this morning presents: one of those Two Truths and a Lie games. You choose which two you think are true and – special house rule – give explanations for why. The explanations do not require specialized medical knowledge beyond the level of a smart amateur. Answers tomorrow-ish.

1. Significantly more Americans get skin cancer on the left half of the face than on the right half.

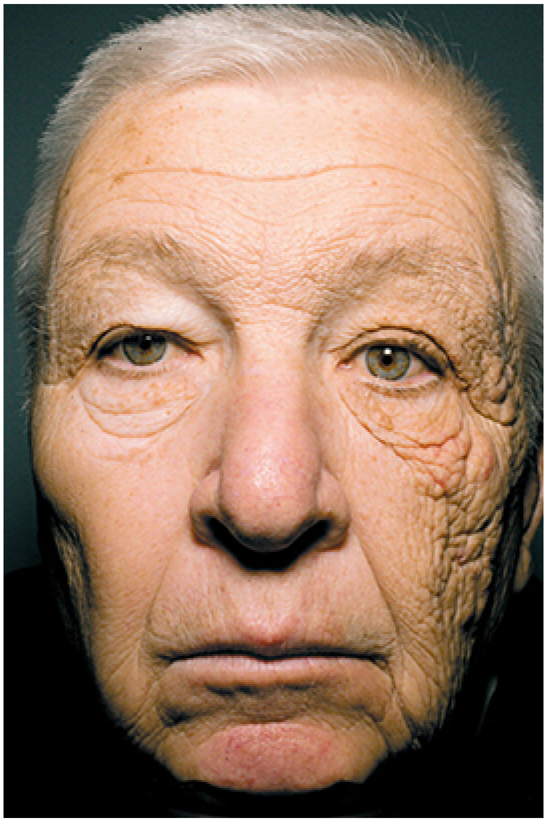
2. People who had acne as children live on average four years longer than those who did not.

3. In very early studies, Botox has shown great promise as a treatment for depression.

(this is one of those shorter posts I was threatening)
(judging from the comments, this was way too easy. Grumble.)

Dermatology Quiz: Answers

The **first statement** on yesterday's quiz was true. Because Americans sit on the left side of the car when they drive, if they drive a lot they get more UV light to the left of their faces than the right. Therefore, they get more left-sided skin cancer. This picture of a truck driver, linked in yesterday's comments, provides a good demonstrations of the effect of UV on skin:



On the other hand, you probably shouldn't avoid sun exposure entirely, since it seems to protect against cancer – I previously was skeptical of a Vitamin D/cancer link, but sun exposure might operate through Vitamin D-independent pathways.

The **second statement** was false. At least, it was supposed to be. I tried to invent a distractor answer I was pretty sure was false. Commenter Kolya came up with a really sneaky argument for why it might be true. This just goes to show how easy it is to confound something. Please be reassured that I am not really that tricky.

The **third statement** was true. See Treatment of depression with onabotulinumtoxin A: a randomized, double-blind, placebo-controlled trial. 52% of botox group responded to treatment vs. 15% of the placebo group (who got injected with nothing; sucks to be them) with effect lasting at least six weeks.

The likely explanation is the "facial feedback hypothesis", which posits that the brain uses facial musculature as a sort of "notepad" for emotions – if one part of the brain decides to feel happy, it makes the face smile, and then other parts of the brain can check whether the face is smiling a couple minutes later and figure out if they are supposed to be feeling happy. As silly as it sounds, if you paralyze the muscles that frown, that makes it harder to feel sad. Even better, the same treatment improves mood in healthy people without depression.

I actually didn't realize that last part until I started writing this post. I wonder if the Botox effect is susceptible to happiness set point or not. I could easily see it not being so, which would make it one of the most powerful happiness-boosting interventions that exists and a little less creepy than giving your usual oral or IV drugs to make people happier.

Botox costs $400 a pop and lasts for months. I wonder how that compares to poverty relief, health care, and malaria nets in the utilitarian calculus.

Darnit, just once I want to learn an interesting medical fact that doesn’t end up as a plan to drug the entire population.

(but think how pretty everyone would be!)

Going Loopy

[content warning: mild ideohazards about rumination that might make people who have anxiety disorders have anxiety disorders more effectively, in the bad sense of "effectively"]

[epistemic status: more crackpottish than usual for this blog. Wild speculation.]

I.

If the brain had been designed by an amateur, it would enter a runaway feedback loop the first time it felt an emotion.

Think about it. You see a butterfly. This makes you happy. Being happy is an unexpected pleasant surprise. Now you're happy that you're happy. This makes you extra happy. Being extra happy is awesome! This makes you extra extra happy. And so on to as much bliss as your neurons are capable of representing.

Or you stub your toe. This makes you sad. Being sad sucks and makes you less productive and less fun to be around. This means being sad is an unexpected negative event, which sounds like a good reason to be sadder. Being even sadder is going to ruin your night even more, so you get sadder still, and so on until you become suicidally depressed over a single lousy toe.

In the real world, either those feedback loops usually don't happen, or they converge and stop at some finite point. I would not be surprised to learn that a lot of evolutionary innovation and biochemical complexity goes into creating a strong barrier against conditioning on your own internal experience.

Sometimes it fails. A guy named Wegner conducted a famous psych experiment where he asked a bunch of participants to sit in a room alone and try not to think about a bear with sunglasses. Of course, that was pretty much all they *could* think of. They seem to have gotten into a feedback loop where “desire not to think of a bear with sunglasses” -> “thought of a bear with sunglasses” -> “frustration” -> “stronger desire not to think of…” -> “more thoughts of…” and so on.

One of my professor’s work – and my college thesis – expanded the field of bear-with-sunglasses-ology to note that people with obsessive compulsive disorder are found to be much worse at this task than the general population. This is not surprising. OCD seems to be pretty much exactly this scenario, except instead of trying not to think of a bear with sunglasses, you’re trying not to think about how you’re dirty-contaminated, or how maybe you left the stove on, or how what if your car just hit somebody *then* where would you be? OCDers get stuck in feedback loops where their worry about their obsessive thought is itself a form of the obsessive thought and so both justifies and intensifies their worry.

I also remember a study where a guy with a sleep lab offered participants a cash reward if they could go to sleep more quickly than usual; the end result was that all of them took much longer to go to sleep than usual. Strong desire to go to sleep -> actually a strong anti-sedating emotion -> worry that you won’t get the prize -> stronger desire to go to sleep -> even less sedated -> so on.

There seems to be an element of this in most anxiety disorders. Someone goes outside, something bad happens. Next time they go outside, they feel anxious. The usual strong barrier against conditioning upon internal experience is AWOL for some reason. The patient finds the experience of becoming anxiety very negative; therefore their belief that "going outside leads to bad things" is justified. Eventually they are so anxious about possibly becoming anxious that they just stay in their house all the time.

(hey, this is kind of Lob's theorem! If you know that, if you're anxious about being anxious you would be anxious, then you're anxious. Maybe.)

But the most clear-cut example is panic disorder. Someone gets anxious for some reason. They get the standard somatic correlates of anxiety – racing heartbeat, sweating, etc. This is a scary situation to be in, not least because it mimics all sorts of terrible medical conditions like heart attacks. This makes the person more anxious, which increases the somatic correlates, and so on.

II.

A lot of CBT seems to be about manually breaking these feedback loops. But I'm just as interested in what we accidentally do to manually increase them. I worry that making mental self-reference slightly easier and more mentally accessible can lead to a big increase in feedback loop size.

I am misophonic – it means I can't tolerate certain noises. Ten years ago, I would not have used the word "misophonic", and I would not even have said I have low noise tolerance. I would have said "Hey, that TV is bothering me, can you turn it off?"

I didn’t start thinking about it on a meta-level until one of my roommates told me “Wow, Scott, you seem to be really super sensitive to noise.” And then gave me a little bit of grief over it, which made it stick in my brain. Ever after that I modeled myself as a person who was super sensitive to noise.

And that made my noise sensitivity *much, much worse*. I hypothesize that maybe, instead of just noise -> distraction, this created a longer feedback loop. Something like noise -> anxiety that I, as a person sensitive to noise, am going to be distracted -> this anxiety is itself distracting -> noticing that I am distracted and being anxious that this distraction will continue as long as the noise continues -> further distraction -> and so on.

(it’s worth noting here that I have obsessive compulsive disorder and that noise sensitivity is a classic feature of the condition)

I had precisely the same experience on this post when I said I was triggered by certain kinds of feminist and social justice rhetoric. I didn’t really think of it that way until I wrote that sentence. I mean, it was true that I heard the rhetoric and then I felt upset and scared, it was just that I had never thought of it as a part of my identity before, or connected it to the word “trigger”. Well, as soon as I did that, the problem got about three times worse, and continues to get worse. I think this nice little crystallized concept of “trigger” might allow my brain to feed back its anxiety more effectively, like “Yup, this thing triggers you, better start feeling anxious about feeling anxious about feeling anxious about…” and that made it go from “a thing that bothers me but which I can cope with” to “giant psychological disaster”.

(now I wonder about typical mind stuff. Do people without OCD have this same experience of anxiety about anxiety about anxiety…? For example, when I was young I was afraid of the dark, not because I believed in ghosts, but because I expected to hear a sound or see something blown around by the wind which I would mistake for a ghost and then have to deal with waves of terror rushing through my body until I figured out what it was, and this was sufficiently bad that I slept with the lights on as a child. Is this the sort of thing other people could imagine feeling, or is it really unusual?)

But this sort of thing is comparatively small fish. I'm more worried about the effect of our entire rich emotional vocabulary. Like, we just sort of invented the concept of being "stressed out" sometime in the last century. Did that make people more stressed out about being stressed out? Did the movement for people to become more introspective and talk about their feelings more (that was a thing, right? That's why old people are so stoic and young people are so touchy-feely?) make people more likely to fall into feedback loops?

What happens when you give people a psychological diagnosis like "depression"? I've always heard that it makes people feel better, because now they know there's an explanation for what they're experiencing and it's not their own fault. But I don't know if I've seen any studies proving this, and even if there were I'd expect them to suffer from the general bias to confirm things that everyone *knows* are true. What if it just makes people be depressed about how depressed they are, and then go "There's that depression again, guess this means I'm not getting any better" and become depressed about that?

III.

This whole essay is a little crackpottish, but now I’m moving from things that merely can’t yet be supported by evidence to things that actively contradict it. But I think about this a lot, and it’s *my* blog, so shut up and listen anyway.

I notice that the class of mental disorders that seem to involve feedback loops – depression, OCD, anxiety – are also the class of mental disorders effectively treated with serotonergic drugs.

One of the most powerful serotonergic drugs in existence is LSD. And as I have learned from – let’s say long boring journal articles – the main effect of LSD is to make you *literally* loopy. Your thoughts loop in and become about themselves, your sense of time becomes cyclic, everything you see becomes a fractal or a spiral or both. And you end up in an extreme emotional feedback loop to infinity – either a “good trip” or a “bad trip”.

This just-so story is not quite as convincing as I would like because SSRIs and LSDs both increase serotonin but have opposite effects on loopiness. But “serotonin” is a wide and complicated category, and like I said above, I bet the brain has a lot of different mechanisms to finely adjust how self-referential your thoughts are allowed to get.

I will get crackpottier still: maybe the parameter being adjusted is some kind of “allowed size of loop”, so that usually you can finish an entire train of thought and then maybe reflect back on items in the train. Small amounts of LSD decrease the loop size, so that individual thoughts can refer to themselves. And large amounts of LSD (the journal articles I read were *very* comprehensive) decrease the loop size to zero, and the perfect pure consciousness people claim to experience is just a loop looping in on itself forever, empty of content.

The Cowpox of Doubt

I remember hearing someone I know try to explain rationality to his friends.

He started with “It’s important to have correct beliefs. You might think this is obvious, but think about creationists and homeopaths and people who think the moon landing was a hoax.” And then further on in this vein.

And I thought: “NO NO NO NO NO NO NO!”

I will make a confession. Every time someone talks about the stupidity of creationists, moon-hoaxers, and homeopaths, I cringe.

It’s not that moon-hoaxers, homeopaths et al aren’t dumb. They are. It’s not even that these people don’t do real harm. They do.

What annoys me about the people who harp on moon-hoaxing and homeopathy – without any interest in the rest of medicine or space history – is that it seems like an attempt to Other irrationality.

It’s saying “Look, over here! It’s irrational people, believing things that we can instantly dismiss as dumb. Things we feel no temptation, not one bit, to believe. It must be that they are defective and we are rational.”

But to me, the rationality movement is about Self-ing irrationality.

It is about realizing that you, yes you, might be wrong about the things that you’re most certain of, and nothing can save you except maybe extreme epistemic paranoia.

Talking about moon-hoaxers and homeopaths too much, at least the way we do it, is *counterproductive* to this goal. Throw examples of obviously stupid false beliefs at someone, and they start thinking all false beliefs are obvious. Give too many examples of false beliefs that aren't tempting to them, and they start believing they're immune to temptation.

And it raises sloppiness to a virtue.

Take homeopathy. I can't even count the number of times I've heard people say: "Homeopaths don't realize beliefs require evidence. No study anywhere has ever found homeopathy to be effective!"

But of course dozens of studies have found homeopathy to be effective.

"Well, sure, but they weren't double-blind! What you don't realize is that there can be placebo effects from…"

But of course many of these studies have been large double-blinded randomized controlled trials, or even meta-analyses of such.

"Okay, but not published in reputable journals."

Is *The Lancet* reputable enough for you?

"But homeopaths don't even realize that many of their concoctions don't contain even a single molecule of active substance!"

But of course almost all homeopaths realize this and their proposed mechanism for homeopathic effects not only survives this criticism but relies upon it.

"But all doctors and biologists agree that homeopathy doesn't work!"

Have you ever spent the five seconds it would take to look up a survey of what percent of doctors and biologists believe homeopathy doesn't work? Or are you just assuming that's true because someone on your side told you so and it seems right?

I am of course being mean here. Being open-minded to homeopaths – reading all the research carefully, seeking out their own writings so you don't accidentally straw-man them, double-checking all of your seemingly "obvious" assumptions – would be a waste of your time.

And someone who demands that you be open-minded about homeopathy would not be your friend. They would probably be a shill for homeopathy and best ignored.

But this is exactly the problem!

The more we concentrate on homeopathy, and moon hoaxes, and creationism – the more people who have never felt any temptation towards these beliefs go through the motions of "debunk"-ing them a hundred times to one another for fun – the more we are driving home the message that these are a representative sample of the kinds of problems we face.

And the more we do that, the more we are training people to make the correct approach to homeopathy – ignoring poor research and straw men on your own side while being very suspicious of anyone who tells us to be careful – their standard approach to any controversy.

And then we get people believing all sorts of shoddy research – because after all, the world is divided between things like homeopathy that Have Never Been Supported By Any Evidence Ever, and things like conventional medicine that Have Studies In Real Journals And Are Pushed By Real Scientists.

Or losing all subtlety and moderation in their political beliefs, never questioning their own side's claims, because the world is divided between People Like Me Who Know The Right Answer, and Shills For The Other Side Who Tell Me To Be Open-Minded As Part Of A Trap.

This post was partly inspired by Gruntled and Hinged's You Probably Don't Want Peer-Reviewed Evidence For God. But there's another G&H post that got me thinking even more.

Inoculation is when you use a weak pathogen like cowpox to build immunity against a stronger pathogen like smallpox. The inoculation effect in psychology is when a person, upon being presented with several weak arguments against a proposition, becomes immune to stronger arguments against the same position.

Tell a religious person that Christianity is false because Jesus is just a blatant ripoff of the warrior-god Mithras and they'll open up a Near Eastern history book, notice that's not true at all, and then be that much more skeptical of the next argument against their faith. "Oh, atheists. Those are those people who think stupid things like Jesus = Mithras. I already figured out they're not worth taking seriously." Except on a deeper level that precedes and is immune to conscious thought.

So we take the intelligent Internet-reading public, and we throw a bunch of incredibly dumb theories at them – moon-hoaxism, homeopathy, creationism, anti-vaxxing, lizard people, that one guy who thought the rapture would come a couple years ago, whatever. And they are easily debunked, and the stuff you and all your friends believed was obviously true is, in fact, obviously true, and any time you spent investigating whether you were wrong is time you wasted.

And I worry that we are vaccinating people against reading the research for themselves instead of trusting smarmy bloggers who talk about how stupid the other side is.

That we are vaccinating people against thinking there might be important truths on both sides of an issue.

That we are vaccinating people against understanding how “scientific evidence” is a really complicated concept, and that many things that are in peer-reviewed journals will later turn out to be wrong.

That we are vaccinating people against the idea that many theories they find absurd or repugnant at first will later turn out to be true, because nature doesn’t respect our feelings.

That we are vaccinating people against *doubt*.

And maybe this is partly good. It’s probably a good idea to trust your doctor and also a good idea to trust your climatologist, and rare is the field where I would feel comfortable challenging expert consensus completely.

But there’s also this problem of hundreds of different religions and political ideologies, and most people are born into ones that are at least somewhat wrong. That makes this capacity for real doubt – doubting something even though all your family and friends is telling you it’s obviously true and you must be an idiot to question it at all – a tremendously important skill. It’s especially important for the couple of rare individuals who will be in a position to cause a paradigm shift in a science by doubting one of its fundamental assumptions.

I don’t think that reading about lizard people or creationism will affect people’s ability to distinguish between, let’s say, cyclic universe theory versus multiverse theory, or other equally dispassionate debates.

But if ever you ever need to have a true crisis of faith, then any time you spend thinking about homeopathy and moon hoaxes beyond the negligible effect they have on your life will be time spent learning exactly the wrong mental habits.

Do You Believe Me, Doc?

[Content warning: psychiatrists having ethically complicated conversations with patients]

I recently attended the Michigan Psychiatric Society conference. It was all downhill after I heard the name of the first poster presentation. The study – investigating the tendency of anti-Parkinsons medication Requip® to cause nightmares as a side effect – was titled "Requip For A Dream". I didn't stay long enough to see whether that won the poster contest, but if it didn't there is no God.

But I also attended a couple of forums and Q&A sessions, and in one of them a doctor asked the question:

"What do I say when one of my psychotic patients – who is telling me how he is Jesus, or is being pursued by the FBI, or something like that – glares at me and asks 'You don't believe me, do you?'"

The lecturer, who was a very prestigious psychiatrist of some sort, said that his standard response was "I believe that what you are seeing and experiencing is real to you."

This is the sort of nice, pat answer I would expect from a clever and prestigious psychiatrist. It ticks all the boxes. It is kind and compassionate. It doesn't technically lie. It doesn't validate the patient's delusions. And it acknowledges the patient's emotions without being dismissive or confrontational.

On the other hand, if I was that patient it would enrage me.

Let me distinguish this from a very very similar concept where I think this answer is exactly correct. Suppose someone is having hallucinations, like believing there are spiders crawling all over him. He asks “You don’t believe me, do you?” I think the *exactly correct* answer is to say “I one hundred percent believe that you are experiencing spiders crawling all over you, that their appearance and features are extremely convincing, and that you’re not making this up. But there are not real spiders on you.”

And if you responded to the guy who thought he was Jesus with “I one hundred percent believe you are feeling a strong, almost irresistible urge to believe you are Jesus. But I don’t think you are actually him,” that would remove most of the creepiness for me (I also don’t think it would be very popular with patients).

But somehow this guy’s phrasing pressed my buttons. It wasn’t just that he wasn’t answering the question. It was that he was denying that the question was the sort of thing that needed to be answered, denying that there was a real fact-of-the-matter about Jesus at all, or denying that it was worth worrying about.

But if you’re worried you’re psychotic, that’s probably the *most* important question to you. The reason this came up at a big conference is that it’s a really common question. Psychotic people ask it a lot. If you’re psychotic, then the fact that you believe these strange things no one else believes has become one of the central things in your life. And to you it’s less important that the person be Validating And Accepting than that you settle this problem that is tearing your life apart.

And this answer isn’t even subtle about what it’s doing. It’s like “Obviously I don’t believe you, but I’m going to avoid saying so in so many words, and I bet you won’t even notice or care. You'll just be grateful I'm acknowledging you at all". It's condescending, is what it is.

I'm not psychotic (I think). And I'm probably more concerned with there being a Real-World-With-Truth-Values than the average person. So maybe the prestigious expert is, as is often the case with prestigious experts, right. But I really don't want to follow his advice. It would leave too bad a taste in my mouth.

I haven't decided what I am going to say in its place. But in a perfect world, where I get exactly the right patient, the response I would really like to give is: "If you were me, would you believe it?"

I think, in this fantasy, if I picked the right patient they would laugh and say "Nope!". Because psychotic people are smarter than they are usually given credit for, and also usually have good senses of humor, and at least we would both establish where we stood in a non-confrontational way.

And it's always interesting how often deluded people know in the back of their head that their delusions are wrong, or at least questionable. Like I can just ask people "I'm here to do a psychiatric evaluation of you. Do you have any strange beliefs I should know about?" and they'll say "Well, I believe I'm being pursued by the FBI." I ask "Are you being pursued by the FBI?" They say "Yeah." I asked once, because I was very curious and making things up as I went along, "Then why did you bring it up so quickly when I said I was a psychiatrist looking for symptoms of mental disease?"

My patient didn't have a good answer for that. I didn't get the impression it was some very logical "Well, I realize statistically most people who think they're pursued by the FBI are psychotic, so I'll just mention it, even though I personally am not.” It seemed more like another example of people, whether psychotic or not, being kind of garage-dragon-y.

This brings me to the other question I get from people a lot, which is “Do you think I’m crazy?” I think the Officially Correct Answer here is to say “Of course not”, which isn’t very convincing precisely because it’s obviously the Officially Correct Answer psychiatrists give to everyone. Even worse (but surprisingly common) is “Crazy isn’t a technical term”. Thanks. I’m sure that must be very reassuring.

Again, a fantasy answer I would like to give if I have exactly the right patient is “If you’re asking that question, don’t worry.” Which I think is sort of true. If you’re in a psychiatric hospital, and your conclusion is that maybe this means you might be crazy, you have some pretty good reality-based thinking going on. If you’re in a psychiatric hospital, and your conclusion is that maybe this means the FBI has found out you’re Jesus and is trying to stop you, *that’s* the guy who’s in trouble.

And again I worry that I might be getting too clever. Probably some of these patients aren’t very smart, or aren’t very cynical, and a simple “No, of course not” would be reassuring in a way a weird self-referential answer wouldn’t. So far I *have* just given some version of the simple answer. But when there’s someone I know well, and who’s especially jaded, and I doubt the simple answer would go over well, I really want to try something less cliched and more honest.

But even that’s not the answer I fantasize about giving later on, when I have my own practice and patients whom I’ve known for years and I can pick out the ones who are a lot like I was when I was younger and seeing a psychiatrist. For *them* the answer will be “Yes, of course. So am I. So is everyone. The interesting question isn’t whether you’re crazy, it’s whether you function anyway. Let’s try to work on that.”

Someone Writes An Anti-Racist FAQ

The Anti-Racialist Q&A – inspired by my own Anti-Reactionary FAQ and written by blogger The Prussian on SkepticInk – is an astounding essay.

It’s astounding because it is a piece of writing about race that is so good that *I actually have specific criticisms of it*. I didn’t even realize how strange this was until about my tenth nitpick, when I noticed that I was nitpicking individual arguments instead of shouting at my computer “WHY ARE YOU SO STUPID?! WHY?! WHY?!”

It’s astounding because I have *no idea* what the author’s political leanings are even though he seems to go through great trouble to explain them. One minute he will seem like a raging leftist, the next he will be talking about how racism is just as stupid as global warming alarmism. Or talking about how racism keeps people of all skin colors from uniting in brotherhood against the real enemy, Muslims.

It’s astounding because I think it was meant to actually convince people. It’s written in a style of “I can see where you’re coming from, but have you considered X?” I thought I was the only person who had figured out that this worked better than “YOU ARE DUMB AND I HATE YOU. NOW PLEASE AGREE WITH ME.”

But I think the most astounding part is that it might be one of the first things I’ve ever read to argue against racism (NO GOULD AND LEWONTIN DON’T COUNT).

I mean, I’ve read a lot of articles *condemning* racism, and *accusing* people of racism, and *being very upset* about the racism inherent in society. But this might be the first one I’ve ever read to argue against it.

That is worthy of note. It is the exact opposite of the attitude in Cowpox of Doubt, like someone who says “Homeopathy? I guess the responsible thing to do is to read dozens of studies before I form an opinion on it”. This may not be a very practical philosophy or a good way to win friends, but it’s an impressive signal of epistemic virtue if nothing else.

I won’t say the FAQ gets everything right – just for one example, if I’d written it I would have dropped the whole “race doesn’t exist” thing as too complicated a question to be worth debating without a longer diversion into philosophy than most people would be willing to entertain.

But it’s pretty good for a first attempt at the genre.

Prussian deserves more traffic for having written something so ambitious, and I don’t have the energy to enforce the censorship I would need for a good comments thread about race, so please take your comments on his Q&A there, not here.

Confounder Of The Day: How Sexy Your Parents Were

One of the more interesting mental health results is the differing prevalence of psychiatric disorders depending on the age of the patient's parents. This month's JAMA Psychiatry includes a study from Denmark that conducts one of of the largest and most rigorous analyses of this effect and finds that a whole host of psychiatric diseases are more common in people born of young mothers and old fathers.

The young mother effect should be pretty straightforward. Poor women are more likely to have children at a younger age and some psychiatric diseases are more common among the poor. Women with psychiatric disorders can be more impulsive, which results in more unprotected sex and teenage pregnancy and hence children at a younger age; the children then inherit these genes and get psychiatric disorders of their own. No surprises here.

The old father effect has a potentially more interesting explanation. The male reproductive system, unlike the female reproductive system, produces gametes throughout the reproductive lifetime. Presumably older people have more time for mutations to accumulate. Some of these mutations are random, the result of passing cosmic rays or environmental toxins. Others are theorized to be a result of sex cells undergoing "selfish spermatogonal selection", a hypothesized sorta-cancer-like process in which some sperm progenitor cells develop "selfish" mutations that increase their relative prevalence in the testes at the cost of the quality of sperm they produce – greater paternal age provides more time for this process to happen. The children of older fathers therefore end up with a higher mutational load and more likelihood of mental disorders – not to mention a host of other issues like lower IQ, less physical ability, decreased health, et cetera. It's very plausible and according to the Danish study the child of a 45 year old father is 1.5x more likely to develop schizophrenia and 1.8x more likely to develop autism compared to the child of a 25 year old father – not a subtle effect!

But the JAMA Psychiatry study references another study, this one really clever, that casts doubt upon this finding. Petersen et al, also from Denmark (which keeps *really* good track of its mental health and so hosts a disproportionate number of epidemiological psych studies) proposes a much more prosaic explanation.

They find that the mental health of a child depends less on how old her father was when she was born than on how old her father was when he had his *first* child. In fact, after adjusting for the former effect, the latter completely disappears! For example, if Bob had Child 1 at age 20 and Child 2 at age 40, and Dan had Child 1 at age 39 and Child 2 at age 40, Dan's second child is at higher risk for mental disorders than Bob's, even though they both have fathers of the same age – and Bob's younger child is at no higher risk of disorders than his eldest.

Petersen et al don't do a great job of coming out and saying it, but I think they hypothesize that the impairments associated with mental disorders – or just the poorly functioning genes that put someone at risk for mental disorders – make it harder for a man to find a partner and start having kids. To be crass about it, if a man has dysfunctional tendencies his value on the marriage market goes down and he's got to wait longer before he finds people willing to have children with him.

This is sad on a general level. It suggests that guys suffering from mental disorders are unlucky in love in the same way they’re probably unlucky in a lot of other things. And it means society loses what could have been a useful piece of advice for cutting psychiatric disease.

But it’s nice and liberating on a personal level. Here I am at thirty years old, unmarried, childless, and there is *no way* I am going to try to have kids for the next four years while I’m doing my residency. There are a couple of people in my class right now having babies – women, in fact! – and the idea of doing two of the most stressful and time-consuming things you can go through *simultaneously* makes my head spin. Now I get to feel a little less guilty about waiting.

Plutocracy Isn’t About Money

Two political science articles I read recently have surprisingly dissonant conclusions.

Gilens and Page’s study “Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens” is very interesting. You may have spotted it in the news media under any of a host of diverse titles:

The New Yorker: Is America An Oligarchy?

BBC: Study: US Is An Oligarchy, Not A Democracy.

RT: Oligarchy, Not Democracy.

Business Insider: Major Study Finds That The US Is An Oligarchy.

And my favorite, Daily Kos: Too Important For Clever Titles: Scientific Study Says We Are An Oligarchy

(the word “oligarchy” appears in the study only once, at the bottom of page six, as a reference to an alternative theory the authors do not endorse)

But RAMPANT MEDIA PLAGIARISM aside, it’s not a bad summary. The study tries to determine what factors predict whether or not a policy gets implemented in the United States. They compare popular support to elite support, where “elites” are the wealthiest ten percent, and find that elite support is a stronger predictor. I believe the way they put it is that once you know whether elites support a policy, learning whether or not the general public supports it improves your model’s ability to predict whether or not it gets passed only an tiny amount, even though elite opinion and popular opinion are often quite different.

Also recently, Rationalist Conspiracy had a good post on Money Doesn't Matter In Politics. A lot of anecdotes, but also links to some convincing studies, like the one that shows how "in Congressional races where candidates spent about $250K (1990 dollars), every $100K spent got another 0.3% of the vote, a tiny amount."

To Alyssa's list I would add Ansolabehere, Figueiredo and Snyder's: Why Is There So Little Money In Politics?, recently spotted on Marginal Revolution. The summary (which does not include the word "oligarchy"):

> "We show that only one in four studies from the previous literature support the popular notion that contributions buy legislators' votes. We illustrate that when one controls for unobserved constituent and legislator effects, there is little relationship between money and legislator votes. Thus, the question is not why there is so little money in politics, but rather why organized interests give at all."

I call these "dissonant" because the simplest explanation for the Gilens and Page finding is that the economic elite are buying elections. But the Ansolabehere et al result says they couldn't even if they tried. If we take both of these studies at face value, how can we reconcile them?

I can think of a few hypotheses:

1. Legislators vote based on their personal opinions. Most legislators are elite, therefore their opinions correlate with the opinions of other elites.

2. Elites control the media, the universities, et cetera. They affect legislators indirectly, by affecting the entire culture (but how would they do this without influencing commoners? Maybe this is a subset of [1], in that elites consume elite-produced media?)

3. Legislators would *like* to think they are elite, and so they vote with elite opinion in the hopes of looking cool and getting elites to like them.

4. Money does not buy elections, but legislators think it does, so they try to satisfy the people with the money in order to win elections.

5. Money does not buy elections, but money can fund think tanks and lobbyists who can persuade legislators through non-election-buying means. This doesn't take the form of promising financial support or during elections, it just comes from talking and befriending and advising and convincing them. The studies showing money doesn't affect campaigns miss this effect. Ansolabehere seems to like this one, pointing out that interest groups spend ten times as much as lobbying as on direct campaign contributions. But even here there are economic arguments against. They estimate that one hour of a legislator's time costs $10,000. This is a high number, but if talking to legislators seriously affected legislation it would be an *amazing* steal.

6. Elites vote more and are more politically active in terms of volunteering, letter-writing, etc. Legislators try to cultivate their affection to win elections, but it has nothing to do with money. But this effect doesn't seem strong enough to make up for the small number of elites.

7. The connection between elites and successful policies is a coincidence – not in the sense that the study found a

nonsignificant finding, but in the sense that elite opinion and legislative success are both biased in the same direction for different reasons. For example, maybe elites tend to lean conservative, and the conservative party in government is much better organized and able to push more legislation through. Gallup finds there is not a big difference between elites and commoners in terms of basic party labeling. But this study (which does define “elite” somewhat differently) shows that elites are predictably less supportive of welfare and redistribution programs than commoners are (I am enraged that this study doesn’t give good comparative data on social issues). If those programs tend to fail for some reason, that could help produce some of these effects.

The Economics Of Art And The Art Of Economics

Here in Detroit, there is debate and concern over the possibility that the city's bankruptcy might obligate it to sell off masterpieces in the local art museum. Is solving a temporary financial problem really worth the cultural impoverishment of the city?

Yes. From Marginal Revolution:

> Consider "The Wedding Dance," a 16th-century work by the Flemish painter Pieter Bruegel the Elder. Detroit museum visitors have enjoyed this painting since 1930. How much would it cost to preserve that privilege for future generations?
>
> A tidy sum, as it turns out. According to Christie's, this canvas alone could fetch up to $200 million. Once interest rates return to normal levels — say, 6 percent — the forgone interest on that amount would be approximately $12 million a year.
>
> If we assume that the museum would be open 2,000 hours a year, and ignore the cost of gallery space and other indirect expenses, the cost of keeping the painting on display would be more than $6,000 an hour. Assuming that an average of five people would view it per hour, all year long, it would still cost more than $1,200 an hour to provide the experience for each visitor.

So the question of "should Detroit keep this painting?" reduces to "does the average visitor to the art museum derive $1200 in value from seeing this particular painting?"

which is very close to “would you pay $1200 for a ticket to an art museum that only had this painting in it?”

(other people may be more cultured than I am, but I find when I’m in an art museum I spend about ten seconds looking at each painting before moving on to the next one. So for me, at least, the cost is $120 per second of viewing time)

In If It’s Worth Doing, It’s Worth Doing With Made Up Statistics, I endorse trying to think quantitatively – not because we are always very good at quantifying things, but because sometimes just the attempt to quantify things makes the right answer so drop-dead obvious that whatever errors you make won’t change things one way or the other.

In the comments on MR people object that maybe some of the numbers in the calculation are a bit off, and that’s probably true. But just by trying the first numbers we think of, we realize we’re three orders of magnitude away from the spot where this would be a hard problem. And our numbers aren’t *that* off.

And this is why I continue to identify as consequentialist even though consequentialism is very hard and we can never do it exactly right. You don’t need a complete theory of ballistics in order to avoid shooting yourself in the foot.

Since I’m already being all soulless and analytical, let me just come out and say it – sell every piece of art in Detroit, but hire skilled forgers to make exact copies of them for a couple of hundred dollars each. You’ll have made billions of dollars, and the Detroit Art Museum will look exactly the same to anyone who’s not examining it through an electron microscope.

Sure, it’ll make it a little harder to signal snooty cultural superiority. But if you’re living in Detroit and trying to signal snooty cultural superiority, man, I don’t know what to tell you.

Right Is The New Left

[Content warning: some ideas that might make you feel anxious about your political beliefs. Epistemic status: very speculative and not necessarily endorsed. This post is less something I will defend to the death and more a form of self-therapy.]

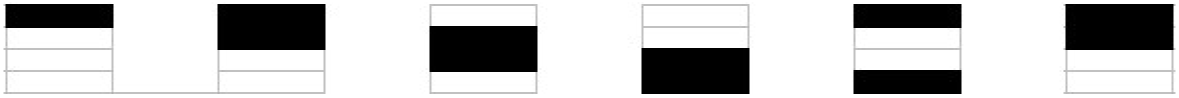
I.

Let’s explain fashion using cellular automata. This isn’t going to be cringe-inducingly nerdy at all!

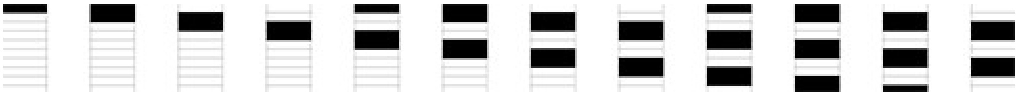
We’ll start with a one-dimensional vertical “world” a single cell thick and however many cells we want tall. Cells can be in one of two states, “black” or “white”. We start with the top cell “black” and all the other cells white, and the world changes with granular time (“ticks”) according to the following rules:

1. On each tick, a cell tries to be the same color that the cell above it was last tick.
2. On each tick, a cell tries NOT to be the same color that the cell below it was last tick.
3. If they ever conflict, Rule 1 takes precedence over Rule 2.
4. If none of these rules apply, a cell stays as it is.

Here’s what we get with a world four cells tall.



And here’s what we get with a world ten cells tall.



It looks like what we’re getting is a “setup period” as the column fills, followed by a “sandwich effect” of two-cell-tall black rectangles separated by two-cell-tall white rectangles gradually moving down the column. Although this isn’t *really* what happens, it also looks like rectangles that fall off the bottom reappear on the top. The overall effect is sort of like a barber pole.

Okay, now let’s get to the fashion.

Consider a group of people separated by some ranked attribute. Let’s call it “class”. There are four classes: the upper class, the middle class, the lower class, and, uh, the underclass.

Everyone wants to look like they are a member of a higher class than they actually are. But everyone also wants to avoid getting mistaken for a member of a poorer class. So for example, the middle-class wants to look upper-class, but also wants to make sure no one accidentally mistakes them for lower-class.

But there is a limit both to people’s ambition and to their fear. No one has any hopes of getting mistaken for a class *two* levels higher than their own: a lower-class person may hope to appear middle-class, but their mannerisms, accent, appearance, peer group, and whatever make it permanently impossible for them to appear upper-class. Likewise, a member of the upper-class may worry about being mistaken for middle-class, but there is no way they will ever get mistaken for lower-class, let alone underclass.

So suppose we start off with a country in which everyone wears identical white togas. One day the upper-class is at one of their fancy upper-class parties, and one of them suggests that they all wear black togas instead, so everyone can recognize them and know that they’re better than everyone else. This idea goes over well, and the upper class starts wearing black.

After a year, the middle class notices what’s going on. They want to pass for upper-class, and they expect to be able to pull it off, so they start wearing black too. The lower- and underclasses have no hope of passing for upper-class, so they don’t bother.

After two years, the lower-class notices the middle-class is mostly wearing black now, and they start wearing black to pass as middle-class. But the upper-class is very upset, because their gambit of wearing black to differentiate themselves from the middle-class has failed – both uppers and middles now wear identical black togas. So they conceive an ingenious plan to switch back to *white* togas. They don’t worry about being confused with the white-togaed underclass – no one could *ever* confuse an upper with a lower or under – but they will successfully differentiate themselves from the middles. Now the upper-class and underclass wear white, and the middle and lower classes wear black.

It’s easy to see that this is the n = 4 version of the cellular automaton we just discussed.

Before I go on, an obvious objection – in a real world that doesn’t work on “ticks”, how do classes coordinate like this? Like, even if someone in the upper-class sent a super-secret message by butler to every single other member of the upper class saying “Tomorrow we all start wearing black, don’t tell anyone else”, within a day the rest of the world would notice, and the upper-class’ advantage would be lost. And surely in our real world, where the upper-class has no way of distributing secret messages to every single cool person, this would be even harder. They’d have to announce their plan publicly, which would make the signal worthless.

There are some technical solutions to the problem. Upper class people are richer, and so can afford to about-face very quickly and buy an entirely new wardrobe. Upper class people have upper class friends, so it’s easier for them to notice that black is ‘in’ and switch accordingly.

But I think the major solution is that there aren’t only four classes, and no one is entirely sure what classes they can or can’t pass for. The richest, trendiest person around wears something new, and either she is so hip that her friends immediately embrace it as a new trend, or she gets laughed at for going out in black when everyone *knows* all the cool people wear white. Her friends are either sufficiently hip that they then adopt the new trend and help it grow, or so unsure of themselves that they decide to stick with something safe, or so un-hip that when they adopt the new trend everyone laughs at them for being so clueless they think they can pull off being one of the cool people.

Or – you can’t just copy someone else’s outfit. That would be crass. So you have to understand the *spirit* of the fashion. But this is hard to get right if you’re not familiar with it. The less exposure you have to the values and individuals who generated it, the more likely you’ll get it wrong and end up looking like an idiot.

In other words, new trends carry social risk, and only people sufficiently clued-in and trendy can be sure the benefits outweigh the risks. But as the trend catches on, it becomes less risky, until eventually you see your Aunt Gladys wearing it because she saw something about it in a supermarket tabloid, and then all the hip people have to find a new trend.

There’s another solution to this problem too: the upper class copies trends from the underclass. We saw this happen naturally on the 5th tick of the four-cell world, but it might be a more stable configuration than that model suggests. If the rich deliberately dress like the poor, then the middle-class have nowhere to go – if they try to ape the rich, they will probably just end up looking poor instead. It is only the rich, who are at no risk of ever being mistaken for the poor, who can pull this off.

Why do I like this model? It explains a lot of otherwise mysterious things about fashion.

Why does fashion change so darned often? Why can’t people just figure out what’s pretty, then stick to that?

Why is wearing last year’s fashion such a faux pas? Shouldn’t the response be “That person is wearing the second most fashionable outfit ever discovered; that’s still pretty good”?

Why does fashion so often copy the outfits of the lower class (eg “ghetto chic”?) Why, if you are shopping for men’s shirts, are there so many that *literally* say “GHETTO” on them in graffiti-like lettering?

And I don’t think I’m a random nerd coming in here and telling fashion people that I understand them better than they understand themselves. This seems to be how fashion people really think. Just look at the word “poser” (or possibly “poseur”). The thrust seems to be: “A person who is not of the group that is cool enough to wear this fashion is trying to wear this fashion! Get ’em!”

The big complication is that there is not one ladder of coolness going from “upper class” down to “underclass”. There are businesspeople, intellectuals, punks, Goths – all of whom are trying to signal something different. And there’s more than just white or black – hundreds of different colors, styles, and whatever.

But I think this is the fundamental generator that makes it all tick. In fact, I think this principle – counter-signaling hierarchies – is the fundamental generator that makes a lot of things tick.

II.

In the past two months I have inexplicably and very very suddenly become much more conservative.

This isn’t the type of conservativism where I agree with any conservative *policies*, mind you. Those still seem totally wrong-headed to me. It’s the sort of conservativism where, even though conservatives seem to be wrong about everything, often in horrible or hateful ways, they seem like probably mostly decent people deep down, whereas I have to physically restrain myself from going on Glenn Beck style rants about how much I hate leftists and how much they are ruining everything. Even though I mostly agree with the leftists whenever they say something.

(In fact, it seems like an important observation that there is a state of mind in which, no matter what your intelligence or rationality level, Glenn Beck or Rush Limbaugh-style rants against The Left seem justifiable and fun to listen to. I cannot communicate this state of mind and don’t know why it occurs.)

At first I didn’t notice this, because way back when I was a teenager and *very* leftist, I made a conscious decision that in order to counter my natural biases I should try to be as understanding and friendly to conservatives as possible. I gradually got better and better at this and didn’t notice that

I was getting *too* good at it until it suddenly started to explode.

And now I am trying to figure out why that is.

Like all of you, my first thought was of course the pathogen stress theory of values. If conservative values are fueled by fear of contamination based on an inbuilt evolutionary reaction to the observed level of pathogen exposure, then my current work on an internal medicine hospital team – which is pretty heavy on the death and disease even for a doctor – would turn me super-conservative very quickly. But this hypothesis should mean that *all* doctors should be very conservative, which doesn't seem to be true. So scratch that.

Perhaps it's a natural effect of settling down, having a stable job, living in my own house, and being in a long-term relationship. But again, a lot of people seem to do all those things without becoming conservative. And none of that has changed in the past few months.

I do admit that, although I try to base my reasoned opinions on The Greater Good, a lot of my political emotions are based on fear, especially fear for my personal safety. I don't feel remotely threatened from the right – even when I meet anti-Semites who think all Jews should die, my feelings are mostly benevolent bemusement. I know if it ever came to any conflict between me and them, then short of them killing me instantly I would have *everyone in the world* on my side, and the possibility of it ending in any way other than with them in jail and me a hero who gets praised for his bravery in confronting them is practically zero. On the other hand, I feel massively threatened from the left, since the few times I got in a fight with *them* ended with me getting death threats and harrassment and feeling like everyone was on their side and I was totally alone. But nothing *new* of this sort has happened in the past two months. That was probably a risk factor, but it can’t have been the trigger.

I’ve been under a lot of stress lately – nothing serious, just very busy days at work with pretty much no free time (writing blog entries doesn’t require free time. They just appear.) It wouldn’t really surprise me if stress were related to conservativism. But I’ve been much more stressed in the past without this effect. Maybe work-related stress has some special ability to cause this effect? That would explain why so many working-class people with crappy jobs end up conservative.

The Left has been doing an unusual number of bad things in the past two months. I remember especially noticing the Eich incident and invasion of the Dartmouth administration building and related threats and demands. And then there was that thing with the national debate championships that is so horrible I still refuse to believe it and hold out hope against hope it turns out to be some absurdly irresponsible reporting or maybe a very very late April Fools’ joke. But I feel like these sorts of things probably go on all the time, and my increased conservativism is the cause, and not the effect, of me noticing them. And I notice I don’t feel the same level of cosmic horror when conservatives do something equally outrageous.

The explanation I like least is that it comes from reading too much neoreaction. I originally rejected this hypothesis because I don’t believe most what I read. But I’m starting to worry that there are memes that, like Bohr’s horseshoe, affect you whether you believe them or not: memes that crystallize the wrong pattern, or close the wrong feedback loop. I have long suspected social justice contains some of these. Now I worry neoreaction contains others.

In particular I worry about the neoreactionary assumption that leftism always increases with time, and that today's leftism confined to a few fringe idiots whom nobody really supports today becomes tomorrow's mainstream left and the day after tomorrow's "you will be fired if you disagree with them". Without me ever really evaluating its truth-value it has wormed its way into my brain and started haunting my nightmares.

Certain versions of it are certainly plausible. In 1960, only a handful of low status people were arguing that "sodomy laws" should be repealed, and they were all insisting that c'mon, obviously it would never go as far as gay *marriage*, we're just saying you shouldn't be put in jail for it. Meanwhile, fifty years later people are enforcing a rule that if you're not on board with gay marriage, you shouldn't be allowed to hold a high-status job.

Of course, many leftist views, even leftist social views, don't spiral out of control like this. Support for abortion and gun control have stayed pretty stable for decades, radical feminism seems to have leveled off, and aside from global warming environmentalism has kind of faded into the background. But it's impossible to predict which ones are going to spiral – to a 1960s conservative homosexuality would have seemed just about the *least* likely thing to catch on.

So now every time I read an article about horrible conservatives – like that South Carolina mayor – I can dismiss it as a couple of people doing dumb things and probably the system will take care of it. If it doesn't take care of it by punishing him personally, it'll take care of it by making people like him obsolete and judged poorly by posterity.

But every time I read an article about horrible leftists – like the one with the debate club – part of me freaks out and thinks – in twenty years, those are the people who are going to be getting me fired for disagreeing with them.

And every time I want to talk about it, I freak out and worry that soon they'll start firing people for disagreeing with the idea that you should be able to fire people for disagreeing with ideas. Like, this could go *uncomfortably* far.

And so there is a dark and unpleasant Orwellian part of my brain that tells me: "If you want a vision of the future, imagine a hack misjudging a college debate – forever."

III.

But like I said, that's the explanation I like least. My favorite involves those cellular automata from before.

A friend recently pointed out that conservatives aren't, on average, very smart. He illustrated this with a graph of IQ vs. political belief which confirms that the left has a significant advantage.

But I look at my Facebook feed, and here is what I observe.

I see my high school classmates – a mostly unselected group of the general suburban California population – posting angry left stuff like "Ohmigod I just heard about that mayor in South Carolina WHAT A FUCKING BIGGOT!!!"

I see the people I think of as my intellectual equals posting things that are conspicuously nuanced – "Oh, I heard about that guy in South Carolina. Instead of knee-jerk condemnation, let’s try to form some general principles out of it and see what it teaches us about civil society.”

And I see the people I think of as the level above me posting extremely bizarre libertarian-conservative screeds making use of advanced mathematics that I can barely understand: “The left keeps saying that marriage as an institution isn’t important. But actually, if we look at this from a game theoretic perspective, marriage and social trust and forager values are all in this complicated six-dimensional antifragile network, and it emergently coheres into a beneficial equilibrium if and only if the government doesn’t try to shift the position of any of the nodes. Just as three eighteenth-century Frenchmen and a renegade Brazilian Marxist philosopher predicted. SO HOW COME THE IDIOTS ON THE LEFT KEEPS TRYING TO MAKE GOVERNMENT SHIFT THE POSITION OF THE NODES ALL THE TIME???!”

(I will proceed to describe this level extensionally: Jonathan Haidt, Bowling Alone, time discounting, public choice theory, the Hajnal line, contract law, Ross Douthat, incentives, polycentric anything, unschooling, exit rights)

And, I mean, I know the reason I get so many people trying to come up with bizarre mathematizations of politics is because those are the sorts of people I select as my friends. The part I don’t get is why so many of them end up weird libertarian-conservative. Certainly not because I selected them for that. I don’t even think they were weird libertarian-conservatives a few years ago when I met a lot of them. It just seems to have caught on.

And my theory is that in a world where the upper class wears black and the lower class wears white, they’re the people who have noticed that the middle class is wearing black as well, and have decided to wear white to differentiate themselves.

It’s the reverse of the 1950s. Assume you’re a hip young intellectual in the 1950s. You see all these stodgy conservatives around you – I don’t even know what “stodgy” means, I just know I’m legally obligated to use it to describe 1950s conservatives. You see Mrs. Grundy, chattering to her grundy friends about how *scandalous* it is that some people read books about sex, lecturing to the school board on how they had *better* enforce her values on the children or she will have some *very* harsh words to say to them.

And you think “Whatever else I am, I’m not going to be a mediocrity like Mrs. Grundy. I’m not going to *conform*.” Which, in the 1950s, meant you became a leftist, and talked about how stodgy society was fundamentally oppressive, and how you were going to value *different* things, and *screw* what Mrs. Grundy thought.

And gradually this became sufficiently hip that even the slightly less hip intellectuals caught on and started making fun of Mrs. Grundy, and then people even less hip than that, until it became a big pileup on poor Mrs. Grundy and anyone who wanted even the slightest claim to intellectual independence or personal integrity has to prove themselves by giving long dissertations on how terrible Mrs. Grundy is.

But when Mrs. Grundy herself joins the party, what then?

I mean, take that article on Dartmouth. A group of angry people, stopping just short of violence, invade a school building and make threats against the president unless he meets their demands. Every student must be forced to attend moral instruction classes inculcating their (the protesters’) values. Offensive terms must be removed from the library. And the school must take care to admit people of the *right* race. When was the last time you could hear a story like that and have it be even *slightly* probably that the mob was rightist?

It’s hard to argue that Mrs. Grundy is not a proud leftist by now, still chattering about how scandalous it is that people read books with the wrong values, still giving her terminally uncool speeches to the school board about how they had *better* enforce her values on the children (and if she can get the debate society on board as well, so much the better).

There must be overwhelming temptation among hip intellectuals to differentiate themselves from Mrs. Grundy by shifting rightward.

And perhaps so far this has been kept in check by the second rule of our cellular automaton – you can’t take a position that would get you plausibly confused for a person of lower class than you.

I was tickled by a conversation between two doctors I recently heard in a hospital hallway:

> **Doctor 1:** My daughter just got a full scholarship into a really good university in Georgia.
> **Doctor 2:** Congratulations!
> **Doctor 1:** Thanks! But I’m hoping she’ll choose somewhere closer to home.
> **Doctor 2:** Why? Because you want to be able to visit her more?
> **Doctor 1:** There’s that. But the other problem is that the South is full of *those people*.
> **Doctor 2:** So? Colleges are like their own world. Your daughter probably won’t even encounter many of them.
> **Doctor 1:** I know. But I keep worrying that just by being

> there, she'll make friends with them, and then end up bringing one home as a boyfriend.

"Those people" is my replacement, not the original term used by the doctor involved. The doctor involved said a much less polite word.

She said "fundies".

Fundies – in all of their Bible-beating gun-owning cousin-marrying stereotypicalness – have so far served as the Lower Class With Which One Must Not Allow One's Self To Be Confused. But I think that's changing. Sorting mechanisms are starting to work so well that, at the top, the fundies just aren't plausible. In our model, people from class N can be confused with class N-1, but never with class N-2. But as the barber-pole movement of fashion creeps downward, fundies are starting to become two classes below certain people at the top, and those people no longer risk misidentification.

I notice that, no matter how many long rants against feminism I write, everyone continues to assume I am a feminist. It's like, "He doesn't make too many spelling errors, his writing isn't peppered with racial slurs – he's got to be a feminist. He probably just forgot the word 'not' in each of his last 228 sentences."

And I wonder if maybe the reason why I am outraged by the debate team but not by the South Carolina mayor isn't that I feel a greater threat from the debate team, but because I feel like there is a greater threat of me being *mistaken* for the debate team. If impotent expressions of outrage divorced from any effort to change things are ways of saying "I'm not like this! I promise!" And I get less outraged than some other people about South Carolina because I feel confident enough in my intelligence that I don't worry anyone will mistake me for a fundie. But I feel less confident no one could mistake me for the sort of person who judged those debate championships, so I need to shout at them to show I’m Not Like That. This would actually explain a lot.

If some intellectuals no longer need to worry about being mistaken for fundies, that frees them to finally breath a sigh of relief and start making fun of Mrs. Grundy again. And that means they’ve got to become conservatives, or libertarians, or anything, anything at all, except for leftists.

So far it is just a few early adopters – the intellectual equivalent of the very trendy people who start wearing some outrageous fashion and no one knows if it is going to catch on or whether they will be soundly mocked for it.

And they are having a really difficult time, because a lot of conservative ideas aren’t that great. Like, reality leaves you a lot of degrees of freedom when you’re deciding your political self-presentation, but it doesn’t leave you an *infinite* number of degrees of freedom, and the project of creating something that is both anti-leftist enough to serve as a fashion statement but reality-based enough not to be dumb is still going on. The reactionaries are doing an excellent job maximizing the “anti-leftist” criterion. The “reality-based” criterion is a harder egg to crack, but it makes me think of Drew Summitt, Athrelon, and some of SarahC’s more political moments.

As the Commissioner puts it, “Evolution is at work here, but just what is evolving remains to be seen.”

When I put it like this, I realize I’m not becoming more conservative at all. I’m becoming anti-leftist. Actually, put that way a lot of people seem to be anti-leftist. I can’t think of a single specific policy proposal supported by Glenn Beck. Can you?

And I think the best explanation is that all my hip friends who I want to be like are starting to be conservative or weird-libertarian or some variety of non-leftist, and Mrs. Grundy is starting to become very obviously leftist and getting grundier by the day, and so the fashion-conscious part of my brain, the much-abused and rarely-heeded part that tells me “No, you can’t go to work in sweatpants, even though it would be much more comfortable”, is telling me “QUICK, DISENGAGE FROM UNCOOL PEOPLE AND START ACTING LIKE COOL PEOPLE RIGHT NOW.”

And I said this is my favorite of all the explanations. Why?

Because if it’s true, and it spreads beyond a couple of little subcultures, it means my worst fears are misplaced. The future isn’t a foot stamping on the face of a a college debate team forever. It’s people – or at least some people – rolling their eyes at those people and making fake vomiting noises. And then going too far, until other people have to roll their eyes at *those* people. And so on. Instead of a death spiral we get a pendulum, swinging back and forth.

But I would hope for something even better than that. Like, at each swing of the pendulum, people learn a little. I was really impressed with how many smart and decent people thought that the Eich thing was wrong (…and wore kilts, and played bagpipes…shut up). Fashion does not accrete, but maybe reality does. And I would like to think that the rationalist movement is a part of that. And if that’s true, that’s a way in which reality will eventually come to overpower fashion and the arc of the universe might tend toward justice after all.

Friendship Is Countersignaling

Yesterday I talked about countersignaling in the context of some controversial and complicated ideas. Maybe I should have started with some examples where I'm more certain countersignaling is at work, just to drive the concept home.

How about slagging?

I don't know if the word exists in American English. I mostly heard it in Ireland. But the concept seems to be everywhere. It's kind of like a high-powered version of teasing, when friends are verbally cruel to each other as a form of bonding.

Every day when I go into the residents' lounge at work, I have lovely conversations with another doctor I'll call Becca because it's the name I used for him last time he showed up on my blog. They tend to sound something like this:

> **Becca:** What are you doing here? I figured they'd have locked you away in the psych ward for good by now.
> **Scott:** Nope. And what are you doing here? You haven't killed off all your patients yet?
> **Becca:** Only person in this hospital I might kill is standing right in front of me.
> **Scott:** Be careful, I'm armed and dangerous *picks up a central line placement practice set menacingly*

The entire thing is done in good cheer and with good results. And this sort of thing is probably familiar enough to most people that no one mistook the dialogue for a genuine threat or even genuine enmity.

I only recently realized this is classic countersignaling.

Remember that countersignaling is doing something that is the opposite of a certain status to show that you are so clearly that status that you don't even need to signal it. The classic advantage is how the aristocratic rich *don't* buy gaudy expensive things, in order to show that they are so obviously rich they don't need to convince people of their wealth the way the nouveau riche do. As the last post put it, you can mistake someone at level n for level n-1, but never for level n-2 – so pretending to be level n-2 and getting away with it is a sure sign that you are in fact level n and not level n-1.

If a person I didn't know or trust said he hated me, or thought I should be locked up, or wanted to kill me, I would take it seriously and freak out. When Becca says he hates me and wants to kill me, it's a way of saying "I am so obviously your friend that I can even signal hostility really strongly and you won't believe me."

Compare this to a similar incident I had recently with another doctor whose pseudonym will be…let's say Pat:

> **Scott:** Thanks for covering for me yesterday. The pharmacy called and said they were a little confused by your discharge instructions, so could you call them back and sort that out?
> **Pat:** Oh, all right, but you owe me *big time* for taking care of all this for you.
> **Scott:** Hey, if you hadn't screwed up the discharge yesterday, we wouldn't be in this mess.
> **Pat:** What? How dare you! *storms out of the room*

I felt really bad after this and sought out Pat to apologize. When I did, Pat was surprised I had taken the whole "storming out" thing seriously, since it was supposed to be theatrical and overdramatic, and felt guilty about worrying me.

Pat’s first comment, about “you owe me big time”, had been done in a spirit of obvious teasing. My comment, about screwing up the the discharge, had *I thought* been done in a spirit of obvious teasing, but I guess I could have been wrong. Pat’s last comment, “How dare you?!” had been done with a very straight face and convincing storm-out, and even though I guess it was intended as obvious teasing, I didn’t take it that way.

So it looks like we were trying to use the “teasing as signal of mutual friendship” tactic and overplayed our hand, leading to a genuinely awkward situation. Apparently our friendship was not as strong as we thought. And the fact that it breaks down into social catastrophe if your friendship isn’t that strong is exactly what makes the signal credible to begin with.

There’s one more aspect to this business I haven’t mentioned, which is summed up by, of all sources, the Night Vale twitter account:

> Whisper a dangerous secret to someone you care about. Now they have the power to destroy you, but they won’t. This is what love is.

I think saying things like “I’m going to kill you” or otherwise being a jerk is a signal of friendship precisely because it gives the other person ammunition with which to destroy you if they so desire. If I really wanted to get Becca fired, I could go to the Chief of Medicine and say “Becca threatened physical violence against me”. Pat could say “Scott was verbally abusive when I wouldn’t do one of his discharges for him.” And then there would be an investigation, and we would say “But I was just teasing!” – which is, of course, what all bullies say when confronted. To give someone this kind of potential ammunition against you shows a lot of trust that they’re your real friend and will never use it.

Metaphors Be With You

People link to this list of 27 funny metaphors all the time. But that's been around for years, and no one has ever tried to expand it. The only other effort I've seen to collect good rhetorical language has been a handful of entries in the Adolph Eichmann's Evil Cake Contest, which is also too old and too short – although I will forever remember it fondly for teaching me the phrase "nuclear Agamemnon".

So here's *my* list. Some are repeats from my old blog. Some are similes or other forms of figurative language. Some are a little vulgar. Some (most) I have lost the attribution for, but will credit if someone reminds me. All come with a serious trigger warning for groanworthiness.

***

The night passed like a kidney stone: painfully and with the help of major sedatives.
— Little Lytton Contest

This raises more red flags than 1956 China
— kodiak_claw on Reddit

This day is moving slower than a snail moving backwards on a turtle moving forwards

When I saw those two girls kissing I was as hard as Chinese math.

You find your prodigal memory accesses have been stomping around the heap like the Incredible Hulk when asked to write an essay entitled "Smashing Considered Harmful"
— The Night Watch

It turned out that Obamacare, despite all the massive brainpower behind it, had some “glitches,” in the same sense that the universe has some “atoms.”
— Dave Barry

Justin Bieber is to music what Justin Bieber is to Marine Biology.
— GS Elevator Gossip

John M. Byrne is so tragic that breast cancer walks to end him.
— cracked.com

Healthy foods, like reality show stars, tend to be both high-maintenance and tasteless.
— cracked.com

Off he went, faster than a French border guard with track shoes and a coupon for cigarettes.
— Silfay Hraka

Men are like chocolate…they come in bars, head straight for your thighs, and can never last long enough to satisfy you.

Fake relic dealers in the Middle Ages created enough crowns of thorns, Holy Grails, and apostle-femurs to stock a macabre, New Testament-themed Wal-Mart.

As corny as Kansas in August
— South Pacific

I hate to call a guy fat, but he had more chins than a Taiwanese phone book.

His frenzied movements made him look like a lunatic trying to learn semaphore
— John Larkin

He uses statistics the way a drunk uses a lamppost – for support rather than illumination
— Andrew Lang

He had a Dickensian relationship to alcohol – exposure to spirits would completely change his personality

You are a girl with as much talent for disguise as a giraffe in dark glasses trying to get into a polar bears-only golf club.
— Blackadder

Dan Brown has a gift for metaphor and detail – a gift that’s defective and needs to be returned. I hope he has a receipt.

On reflection, all the mistakes I have made are, in some way or another, your fault. You’ve been at the helm, sowing chaos and stupidity through the landscape of my days like some sort of retarded Johnny Appleseed.
— www.goats.com

Work is like a dick. If it gets hard, fuck it.

I get hit more often than a blackjack player facing a deck of deuces.
— Order of the Stick

When you get right down to it, museums are basically zoos for inanimate objects.

Japan is to crazy as the Middle East is to oil

The plot is as engaging as watching a foot race between two convenience stores.
— cracked.com

It’s like you crystallized the best sex you’ve ever had with a woman and put it in a tiny bottle on your spice rack. No,

wait, it’s like you took THAT spice and snorted it while screwing her hotter sister.
— OotS

Ayn Rand was the Evel Knievel of leaping to conclusions
— NYMag

Monopoly is to the contemporary board gamer as garlic-infused kryptonite snakes are to Super Vampire Indiana Jones.
— Lore Sjoberg

A good coffee must be hot as the kisses of a girl on the first day, sweet as her love on the third day, and as black as the swearing of her mother when she hears of it.
— Old Chinese proverb

Videogames and movies get along like cats made of oil and dogs made of water.
— Lore Sjoberg

I’ve got the strong feeling that [this movie], based on the talent involved, this is going to be a “loose adaptation,” in the same sense that the Holocaust was a loose adaptation of Nietzsche
— cracked.com

Rowling is as subtle as a large elephant with a big sign reading “I am subtle.”

This movie’s plot has more holes than the PGA championship
— TriggerStreet

He’s as bright as Alaska in wintertime.

The MPAA’s deals are shady as a 100-year old elm.
— Tech Industry News

The giant spider web was created by many spiders working together in harmony, like a United Nations without Russia.

Sitting on the dais, overlooking Washington’s elite, I felt like the Best Man at a wedding between the Statue of Liberty and Mount Rushmore.
— Stephen Colbert

It sounded like the smell of raspberries tastes
— Terry Pratchett

There are a number of body parts [like the appendix] that may have served some important purpose in the distant past, but are now more likely to host disease and infection, much like pay phones.
— Lore Sjoberg

The Church of England is to religion what the cucumber sandwich is to food – it goes quite nicely with a cup of tea, but that’s about it.
— Pat Condell

Tom Cruise is so far in the closet he’s talking to Aslan.

We aren’t ranking the dumbest hip-hop lyrics ever. That would be like making a list of people Hitler was mean to.
— cracked.com

His head was as empty as a eunuch’s underpants.

I am such a hard-core atheist I make Richard Dawkins look like the Virgin Mary
— Bruno Maddox

This game is to free time what whales are to krill.
— Defective Yeti

If the crib’s on fire, you don’t speculate that the baby is flame-retardant.
— Al Gore

I’m straighter than John Wayne voting for Reagan on a horse.
— Achewood

O God, who is as far beyond the universes I span as infinity is beyond six…
— The House Beyond Your Sky, Benjamin Rosenbaum

At last, I’ll be goin’ home, as sure as Thor don’t own shaving cream!
— Order Of The Stick

He has all the common sense of Medusa checking her makeup
— Order Of The Stick

Mother, more gravy. This turkey is as dry as Oscar Wilde
— Stewie, Family Guy

I have more muscles than a New England clambake.

The Assyrians were the first people to start using iron weapons instead of bronze which, to put into a modern perspective, is sort of like showing up for a knife fight with the Death Star.
— cracked.com

With a wet, tearing KERRRAAAAACTCH sound, Molly exploded like a meat piñata at a birthday party for very

strong, invisible children.
— John Dies At The End

The debate club is one of Cork’s most venerable societies. We’re older than John McCain and cooler than Barack Obama. Joining is easier than Sarah Palin’s daughter and simpler than Sarah Palin’s son.
— Introductory speech at the Cork debate club

Most of her cooking ended out as the culinary equivalent of the Tunguska event
— Jasper Fforde

It’s like I’m drowning, and monkeys dressed as lifeguards are throwing me anvils
— Dilbert

As far as appropriate national slogans go, “The luck of the Irish” isn’t far behind “The easy and high paying jobs of the Mexicans.”
— cracked.com

The storm had now definitely abated, and what thunder there was now grumbled over more distant hills, like a man saying “And another thing…” twenty minutes after admitting he’s lost the argument.
— Douglas Adams

Jokes should be short and elegant, like a mathematical proof or a midget in a ballgown.
— Bottled City

On the David Wong Social Awkwardness Scale, with “1” being going to the “Pickup” instead of “Order” counter at a restaurant and “10” being a guy getting caught on national TV having sex with a dead baboon, I’d have to say that the

following minutes alone with Amy rated about a 9.6.
— John Dies At The End

About the film’s merits I prefer to be silent at this stage, except to say that it seemed quite long, in the same sense that the Thirty Years War probably seemed quite long to anyone who had been expecting it to be over sooner.

I am more touched than a congressional page.
— Scott Adams

It’s the most blatant work of fiction since vows of fidelity were included in the French wedding service.
— Blackadder

He swings the bat…it’s a hit!…it’s going…going…and like America’s credibility on the world stage, it’s GONE!
— The Simpsons

Our marriage has been through more hardships than the Jews and Charlie Brown combined.
— The Simpsons

You’re making less sense than an epileptic auctioneer
— goats.com

Her art was about as deathless as a mayfly working on an oil rig.
— Ursula Vernon

I go through life like Helen Keller in a room full of Rubix Cubes
— Scott Adams

Someday, it will be one of those things we look back on with heartwarming nostalgia, like the Atari 2600 and polio.
— Defective Yeti

He asks how, if God is a person, He can speak to billions of people simultaneously, which is rather like wondering why, if Tony Blair is an octopus, he has only two arms.
— Terry Eagleton

Like the ski resort of young girls looking for husbands and husbands looking for young girls, the situation was not quite so reciprocal as it appeared

His jokes were like oil, crude but precious.

Ambition is like a frog sitting on a Venus Flytrap. The flytrap can bite and bite, but it won’t bother the frog because it only has tiny little plant teeth. But some other stuff could happen, and it would be like ambition.
— Jack Handey, Deep Thoughts

The tension was as thick as a community college student
— goats.com

He was as guilty as a Catholic O.J. Simpson
— goats.com

I kept my cool like Steve McQueen’s corpse in an interstellar void
— goats.com

I will help, but the cost will be higher than a Volkswagen full of hippies on a mountaintop
— goats.com

Angrier than a mother bear whose cub has just been sold on Ebay.
— Questionable Content

As famous as the unknown soldier

He makes lichens seem dynamic

Bush interprets the constitution the same way a Unitarian interprets the Bible

The American media has the attention span of a caffeinated Irish Setter.
— Defective Yeti

She was not only homely, but it was one of those homes with a burnt out trailer in the front.
— Terry Pratchett

Instead they stand immobile on the escalator, like tourists at the Mount Olympus Zoo’s gorgon exhibit.
— Defective Yeti

Radon is an interesting element, because it is both noble and deadly, like Vlad the Impaler
— Lore Sjoberg

Many say that DOS is the dark side, but actually UNIX is more like the dark side: It’s less likely to find the one way to destroy your incredibly powerful machine, and more likely to make upper management choke.
— Lore Sjoberg

My indifference to that comment can only be described as sexual in intensity.
-– Martin Terman

It’s the most pointless book since How To Learn French was translated into French.
— Blackadder

I like computers. To me, computers are kind of like tangerines, in that I can’t think of a good analogy about

either right now.
-– Scott Adams

It was more difficult to understand than Bob Dylan reading Finnegan’s Wake in a wind tunnel.
— Dennis Miller

Everyone at the party was as stoned as a United States embassy.
— Anonymous

Life is short and hard, like a body-building elf
— The Bloodhound Gang

A philosopher is a blind man in a dark room looking for a black cat that isn’t there.
— Brent Wade

The building was a true skyscraper – thousands of tons of steel and more stories than a bar full of Vietnam veterans.

A telegraph system is like a giant cat, stretching across the country. You pull on the tail in New York, and the head squeals in Los Angeles. Radio works exactly the same way, except that there is no cat.
— Albert Einstein

Love is a snowmobile racing across the tundra and then suddenly it flips over, pinning you underneath. At night, the ice weasels come.”
— Friedrich Nietzsche (apocryphal)

Love is like pi – irrational and very important.
— Lisa Hoffman

I’m as happy a dingo in a nursery.
— St. Leo

[See also: puns]

Stop Confounding Yourself! Stop Confounding Yourself!

As a perk of my job, I get a free subscription to the *American Journal of Psychiatry*. I am still not used to this. No enraging struggles with paywalls. No “one year embargo on full text”. I just come home and find all of the latest and most interesting journal articles have been *shipped directly to my house*. Modern technology is truly amazing.

Its latest is Takizawa et al’s **Adult Health Outcomes of Childhood Bullying Victimization: Evidence From A Five-Decade Longitudinal British Birth Cohort**. It has since been picked up by Fox, the Washington Post, and even Xinhua. I think that’s enough to qualify for “made world headlines”.

The study took some British kids in 1958, sorted them by how much they got bullied, and checked how they did forty years later. In fact, the frequently bullied kids had nearly twice as much psychiatric disease, were twice as likely to attempt suicide, were twice as likely to drop out of high school, and even had double the unemployment rate. Worse physical health, worse cognitive function, less likely to get married, et cetera, et cetera.

Those must be *some* bullies.

But correlation is not causation. There’s an alternative possibility. Maybe bullies only pick on unpopular disadvantaged kids. And maybe these kinds of things are stable, so that unpopular disadvantaged kids are more likely to grow up to be unpopular disadvantaged adults. The sort of adults who are more likely to have psychiatric disease, drop out of school, be unemployed, et cetera. *That* sure sounds plausible.

So the researchers “controlled for confounders”. They used a scale called the Bristol Social Adjustment Guide to figure out how socially well-adjusted the kids were, then added in their social class, their family’s level of contact with child protective services, their IQ, their attractiveness, and even how much their parents loved them (really! check the study!)

They controlled for all these things and found that the bullying-outcomes link was still robust. They concluded that this meant their finding wasn’t just that bullies were bullying kids with problems, it was that bullies were causing the damage themselves.

Do you believe that? It all comes down to one question.

Who is better able to look deep inside you and judge the mettle of your soul? A playground bully? Or the Bristol Social Adjustment Guide?

My money is on the bully. Bullies are like sharks: horrible pinnacles of evolution. Animals have been learning to navigate social dominance hierarchies through violence since pecking orders in chickens, on through wolf packs and chimpanzees, and up into humans – and we are very good at it. The bully is the purest manifestation of the primal instinct, which is why he crops up untaught and unbidden in near-identical form in schoolyards from Los Angeles to London to Lanzhou. And like sharks, a good bully should be able to smell blood in the water and know when an opportunity to attack presents itself.

Most of the findings of this study were in the “frequently bullied” population, and part of the criteria for “frequently” was bullying both at age 7 and age 11. Unless that's just one *really* persistent guy, that means the child has gotten independently selected for targeting in two different environments. That could be bad luck but could also be the effect of high inter-bully reliability in what (persistent) qualities make a good victim.

So let's take another look at those confounders we supposedly controlled for. Where's height? You think short kids are bullied more often than tall kids? I do. Height is closely related to career success, to attractiveness to the opposite sex, increased happiness and self-esteem, and decreased psychological morbidity. This is something every bully knows intuitively, but which the Takizawa study didn't think of and therefore couldn't control for.

But it's giving them too much credit to be bringing in weird stuff like height-mental-health correlations. What about social skills? Yeah, sure, they did that Bristol Social Adjustment Guide. I'm looking at it right now, and it's asking the students' teachers to rate items like "hostility towards adults" and "depression". I don't believe that teachers filling numbers into hokey little boxes can capture an assessment of a kid's social skills as well as a bully trying to decide who can safely be picked on can.

So I will come out and say it: I do not trust the practice of "adjusting for confounders", at least not the way this study does it. You are adjusting for an imperfect measurement of the confounders you can think of. If you find that there is lingering correlation, then either your hypothesis is true, or you didn't adjust for confounders well enough. Given extraordinary results, like being bullied at age seven making you 25% less likely to be married at age fifty, the "you didn't adjust for confounders well enough" option starts to look really good.

I think the proper way to do this study would have been to do an anti-bullying intervention at a couple schools, leave a couple similar schools as controls, and if the anti-bullying intervention successfully decreases bullying, compare outcomes for children at the two schools. I understand this probably would be logistically impossible, plus you'd have to wait another forty years. But given that you cannot do the study right, I am not sure that doing the study this way adds anything, except of course widely-read articles in every news source in the world.

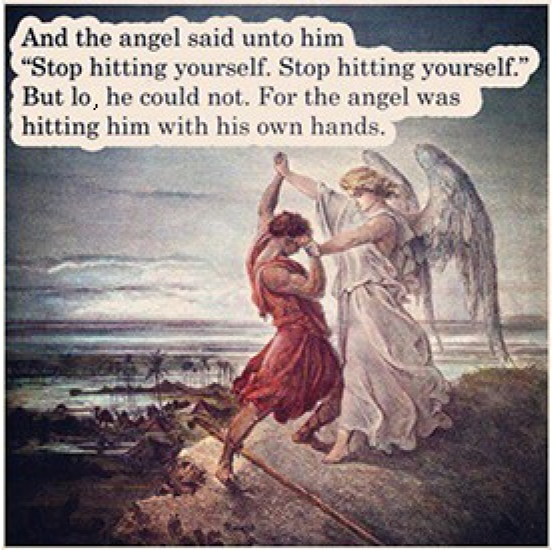
I would also compare to Reming et al, which attempts much the same study and finds no association after adjusting for *their* confounders of choice (which, oddly, are much fewer than in the current study). They also find that parent reports about bullying (the method Takizawa et al used) are wildly unreliable, with an inter-rater agreement of just 0.11 with reports by teachers or the children themselves (the statistic goes from perfect agreement being 1.0 to zero information being 0.0). For a *completely false* measure of bullying to find such spectacular effects is *really suspicious*, and now we need to consider not only the differences between the types of kids who are and aren't bullied, but the differences between the types of parents who do and don't think their kids are being bullied.

Since I insisted on giving this post a silly title, I will now share with you the most interesting perspective on psychology and the "stop hitting yourself" phenomenon I have read all week. This is from Jonathan Haidt on Kohlberg's moral stages:

> During elementary school, most children move on to the two conventional stages, becoming adept at understanding and even manipulating rules and social conventions. This is the age of petty legalism that most

> of us who grew up with siblings remember well (“I’m not hitting you. I’m using your hand to hit you. Stop hitting yourself!”). Kids at this stage rarely question the legitimacy of authority, but learn to maneuver within and around the constraints that adults impose on them.

I always just thought that was a really dickish joke. I didn’t realize it had a *deep philosophical underpinning.*


And the angel said unto him “Stop hitting yourself. Stop hitting yourself.” But lo, he could not. For the angel was hitting him with his own hands.

The Control Group Is Out Of Control

I.

Allan Crossman calls parapsychology the control group for science.

That is, in let's say a drug testing experiment, you give some people the drug and they recover. That doesn't tell you much until you give some other people a placebo drug you *know* doesn't work – but which they themselves believe in – and see how many of *them* recover. That number tells you how many people will recover whether the drug works or not. Unless people on your real drug do significantly better than people on the placebo drug, you haven't found anything.

On the meta-level, you're studying some phenomenon and you get some positive findings. That doesn't tell you much until you take some other researchers who are studying a phenomenon you *know* doesn't exist – but which they themselves believe in – and see how many of *them* get positive findings. That number tells you how many studies will discover positive results whether the phenomenon is real or not. Unless studies of the real phenomenon do significantly better than studies of the placebo phenomenon, you haven't found anything.

Trying to set up placebo science would be a logistical nightmare. You'd have to find a phenomenon that definitely doesn't exist, somehow convince a whole community of scientists across the world that it does, and fund them to study it for a couple of decades without them figuring it out.

Luckily we have a natural experiment in terms of parapsychology – the study of psychic phenomena – which most reasonable people believe don't exist, but which a community of practicing scientists believes in and publishes papers on all the time.

The results are pretty dismal. Parapsychologists are able to produce experimental evidence for psychic phenomena about as easily as normal scientists are able to produce such evidence for normal, non-psychic phenomena. This suggests the existence of a very large "placebo effect" in science – ie with enough energy focused on a subject, you can *always* produce "experimental evidence" for it that meets the usual scientific standards. As Eliezer Yudkowsky puts it:

> Parapsychologists are constantly protesting that they are playing by all the standard scientific rules, and yet their results are being ignored – that they are unfairly being held to higher standards than everyone else. I'm willing to believe that. It just means that the standard statistical methods of science are so weak and flawed as to permit a field of study to sustain itself in the complete absence of any subject matter.

These sorts of thoughts have become more common lately in different fields. Psychologists admit to a crisis of replication as some of their most interesting findings turn out to be spurious. And in medicine, John Ioannides and others have been criticizing the research for a decade now and telling everyone they need to up their standards.

"Up your standards" has been a complicated demand that cashes out in a lot of technical ways. But there is broad agreement among the most intelligent voices I read (1, 2, 3, 4, 5) about a couple of promising directions we could go:

1. Demand very large sample size.

2. Demand replication, preferably exact replication, most preferably multiple exact replications.

3. Trust systematic reviews and meta-analyses rather than individual studies. Meta-analyses must prove homogeneity of the studies they analyze.

4. Use Bayesian rather than frequentist analysis, or even combine both techniques.

5. Stricter p-value criteria. It is far too easy to massage p-values to get less than 0.05. Also, make meta-analyses look for “p-hacking” by examining the distribution of p-values in the included studies.

6. Require pre-registration of trials.

7. Address publication bias by searching for unpublished trials, displaying funnel plots, and using statistics like “fail-safe N” to investigate the possibility of suppressed research.

8. Do heterogeneity analyses or at least observe and account for differences in the studies you analyze.

9. Demand randomized controlled trials. None of this “correlated even after we adjust for confounders” BS.

10. Stricter effect size criteria. It’s easy to get small effect sizes in *anything*.

If we follow these ten commandments, then we avoid the problems that allowed parapsychology and probably a whole host of other problems we don’t know about to sneak past the scientific gatekeepers.

Well, what now, motherfuckers?

II.

Bem, Tressoldi, Rabeyron, and Duggan (2014), full text available for download at the top bar of the link above, is parapsychology’s way of saying “thanks but no thanks” to the idea of a more rigorous scientific paradigm making them quietly wither away.

You might remember Bem as the prestigious establishment psychologist who decided to try his hand at parapsychology and to his and everyone else’s surprise got positive results. Everyone had a lot of criticisms, some of which were very very good, and the study failed replication several times. Case closed, right?

Earlier this month Bem came back with a meta-analysis of ninety replications from tens of thousands of participants in thirty three laboratories in fourteen countries confirming his original finding, $p < 1.2 * -10^{10}$, Bayes factor $7.4 * 10^{9}$, funnel plot beautifully symmetrical, p-hacking curve nice and right-skewed, Orwin fail-safe n of 559, et cetera, et cetera, et cetera.

By my count, Bem follows all of the commandments except [6] and [10]. He apologizes for not using pre-registration, but says it’s okay because the studies were exact replications of a previous study that makes it impossible for an unsavory researcher to change the parameters halfway through and does pretty much the same thing. And he apologizes for the small effect size but points out that some effect sizes are legitimately very small, this is no smaller than a lot of other commonly-accepted results, and that a high enough p-value ought to make up for a low effect size.

This is *far* better than the average meta-analysis. Bem has always been pretty careful and this is no exception. Yet its conclusion is that psychic powers exist.

So – once again – what now, motherfuckers?

III.

In retrospect, that list of ways to fix science above was a little optimistic.

The first nine items (large sample sizes, replications, low p-values, Bayesian statistics, meta-analysis, pre-registration, publication bias, heterogeneity) all try to solve the same problem: accidentally mistaking noise in the data for a signal.

We've placed so much emphasis on not mistaking noise for signal that when someone like Bem hands us a beautiful, perfectly clear signal on a silver platter, it briefly stuns us. "Wow, of the three hundred different terrible ways to mistake noise for signal, Bem has proven beyond a shadow of a doubt he hasn't done any of them." And we get so stunned we're likely to forget that this is only part of the battle.

Bem definitely picked up a signal. The only question is whether it's a signal of psi, or a signal of poor experimental technique.

None of these commandments even *touch* poor experimental technique – or confounding, or whatever you want to call it. If an experiment is confounded, if it produces a strong signal even when its experimental hypothesis is true, then using a larger sample size will just make that signal even stronger.

Replicating it will just reproduce the confounded results again.

Low p-values will be easy to get if you perform the confounded experiment on a large enough scale.

Meta-analyses of confounded studies will obey the immortal law of “garbage in, garbage out”.

Pre-registration only assures that your study will not get any worse than it was the first time you thought of it, which may be very bad indeed.

Searching for publication bias only means you will get *all* of the confounded studies, instead of just some of them.

Heterogeneity just tells you whether all of the studies were confounded about the same amount.

Bayesian statistics, alone among these first eight, ought to be able to help with this problem. After all, a good Bayesian should be able to say “Well, I got some impressive results, but my prior for psi is very low, so this raises my belief in psi slightly, but raises my belief that the experiments were confounded *a lot*.”

Unfortunately, good Bayesians are hard to come by, and the researchers here seem to be making some serious mistakes. Here’s Bem:

> An opportunity to calculate an approximate answer to this question emerges from a Bayesian critique of Bem’s (2011) experiments by Wagenmakers, Wetzels, Borsboom, & van der Maas (2011). Although Wagenmakers et al. did not explicitly claim psi to be impossible, they came very close by setting their prior odds at 10^20 against the psi hypothesis. The Bayes

> Factor for our full database is approximately 10^9 in favor of the psi hypothesis (Table 1), which implies that our meta-analysis should lower their posterior odds against the psi hypothesis to 10^11

Let me shame both participants in this debate.

Bem, you are abusing Bayes factor. If Wagenmakers uses your 10^9 Bayes factor to adjust from his prior of 10^-20 to 10^-11, then what happens the next time you come up with another database of studies supporting your hypothesis? We all know you will, because you've amply proven these results weren't due to chance, so whatever factor produced these results – whether real psi or poor experimental technique – will no doubt keep producing them for the next hundred replication attempts. When those come in, does Wagenmakers have to adjust his probability from 10^-11 to 10^-2? When you get another hundred studies, does he have to go from 10^-2 to 10^7? If so, then by conservation of expected evidence he should just update to 10^+7 right now – or really to infinity, since you can keep coming up with more studies till the cows come home. But in fact he shouldn't do that, because at some point his thought process becomes "Okay, I already know that studies of this quality can consistently produce positive findings, so either psi is real or studies of this quality aren't good enough to disprove it". This point should probably happen well before he increases his probability by a factor of 10^9. See Confidence Levels Inside And Outside An Argument for this argument made in greater detail.

Wagenmakers, you are overconfident. Suppose God came down from Heaven and said in a booming voice "EVERY SINGLE STUDY IN THIS META-ANALYSIS WAS CONDUCTED PERFECTLY WITHOUT FLAWS OR BIAS, AS WAS THE META-ANALYSIS ITSELF." You would see a p-value of less than 1.2 *

10^-10 and think “I bet that was just coincidence”? And then they could do another study of the same size, also God-certified, returning exactly the same results, and you would say “I bet that was just coincidence too”? YOU ARE NOT THAT CERTAIN OF ANYTHING. Seriously, *read the @#!$ing Sequences*.

Bayesian statistics, at least the way they are done here, aren’t gong to be of much use to anybody.

That leaves randomized controlled trials and effect sizes.

Randomized controlled trials are great. They eliminate most possible confounders in one fell swoop, and are excellent at keeping experimenters honest. Unfortunately, most of the studies in the Bem meta-analysis were already randomized controlled trials.

High effect sizes are really the only thing the Bem study lacks. And it is very hard to experimental technique so bad that it consistently produces a result with a high effect size.

But as Bem points out, demanding high effect size limits our ability to detect real but low-effect phenomena. Just to give an example, many physics experiments – like the ones that detected the Higgs boson or neutrinos – rely on detecting extremely small perturbations in the natural order, over millions of different trials. Less esoterically, Bem mentions the example of aspirin decreasing heart attack risk, which it definitely does and which is very important, but which has an effect size lower than that of his psi results. If humans have some kind of *very weak* psionic faculty that under regular conditions operates poorly and inconsistently, but does indeed exist, then excluding it by definition from the realm of things science can discover would be a bad idea.

All of these techniques are about reducing the chance of confusing noise for signal. But when we think of them as the be-all and end-all of scientific legitimacy, we end up in awkward situations where they come out super-confident in a study's accuracy simply because the issue was one they weren't geared up to detect. Because a lot of the time the problem is something more than just noise.

IV.

Wiseman & Schlitz's Experimenter Effects And The Remote Detection Of Staring is my favorite parapsychology paper ever and sends me into fits of nervous laughter every time I read it.

The backstory: there is a classic parapsychological experiment where a subject is placed in a room alone, hooked up to a video link. At random times, an experimenter stares at them menacingly through the video link. The hypothesis is that this causes their galvanic skin response (a physiological measure of subconscious anxiety) to increase, even though there is no non-psychic way the subject could know whether the experimenter was staring or not.

Schiltz is a psi believer whose staring experiments had consistently supported the presence of a psychic phenomenon. Wiseman, in accordance with nominative determinism is a psi skeptic whose staring experiments keep showing nothing and disproving psi. Since they were apparently the only two people in all of parapsychology with a smidgen of curiosity or rationalist virtue, they decided to team up and figure out why they kept getting such different results.

The idea was to plan an experiment together, with both of them agreeing on every single tiny detail. They would then go to a laboratory and set it up, again both keeping close eyes on one another. Finally, they would conduct the experiment in a series of different batches. Half the batches (randomly assigned) would be conducted by Dr. Schlitz, the other half by Dr. Wiseman. Because the two authors had very carefully standardized the setting, apparatus and procedure beforehand, “conducted by” pretty much just meant greeting the participants, giving the experimental instructions, and doing the staring.

The results? Schlitz’s trials found strong evidence of psychic powers, Wiseman’s trials found no evidence whatsoever.

Take a second to reflect on how this *makes no sense*. Two experimenters in the same laboratory, using the same apparatus, having no contact with the subjects except to introduce themselves and flip a few switches – and whether one or the other was there that day completely altered the result. For a good time, watch the gymnastics they have to do to in the paper to make this sound sufficiently sensical to even get published. This is the only journal article I’ve ever read where, in the part of the Discussion section where you’re supposed to propose possible reasons for your findings, both authors suggest maybe their co-author hacked into the computer and altered the results.

While it’s nice to see people exploring Bem’s findings further, *this* is the experiment people should be replicating ninety times. I expect *something* would turn up.

As it is, Kennedy and Taddonio list ten similar studies with similar results. One cannot help wondering about publication bias (if the skeptic and the believer got similar results, who cares?). But the phenomenon is sufficiently well known in parapsychology that it has led to its own host of theories about how skeptics emit negative auras, or the enthusiasm of a proponent is a necessary kindling for psychic powers.

Other fields don’t have this excuse. In psychotherapy, for example, practically the only consistent finding is that whatever kind of psychotherapy the person running the study likes is most effective. Thirty different meta-analyses on the subject have confirmed this with strong effect size (d = 0.54) and good significance (p = .001).

Then there’s Munder (2013), which is a meta-meta-analysis on whether meta-analyses of confounding by researcher allegiance effect were themselves meta-confounded by meta-researcher allegiance effect. He found that indeed, meta-researchers who believed in researcher allegiance effect were more likely to turn up positive results in their studies of researcher allegiance effect (p < .002). It gets worse. There's a famous story about an experiment where a scientist told teachers that his advanced psychometric methods had predicted a couple of kids in their class were about to become geniuses (the students were actually chosen at random). He followed the students for the year and found that their intelligence actually increased. This was supposed to be a Cautionary Tale About How Teachers’ Preconceptions Can Affect Children.

Less famous is that the same guy did the same thing with rats. He sent one laboratory a box of rats saying they were specially bred to be ultra-intelligent, and another lab a box of (identical) rats saying they were specially bred to be slow and dumb. Then he had them do standard rat learning tasks, and sure enough the first lab found very impressive results, the second lab very disappointing ones.

This scientist – let's give his name, Robert Rosenthal – then investigated three hundred forty five different studies for evidence of the same phenomenon. He found effect sizes of anywhere from 0.15 to 1.7, depending on the type of experiment involved. Note that this could also be phrased as "between twice as strong and twenty times as strong as Bem's psi effect". Mysteriously, animal learning experiments displayed the highest effect size, supporting the folk belief that animals are hypersensitive to subtle emotional cues.

Okay, fine. Subtle emotional cues. That's way more scientific than saying "negative auras". But the question remains – what went wrong for Schlitz and Wiseman? Even if Schlitz had done everything short of saying "The hypothesis of this experiment is for your skin response to increase when you are being stared at, please increase your skin response at that time," and subjects had tried to comply, the whole point was that they didn't *know* when they were being stared at, because to find that out you'd have to be psychic. And how are these rats figuring out what the experimenters' subtle emotional cues mean anyway? *I* can't figure out people's subtle emotional cues half the time!

I know that standard practice here is to tell the story of Clever Hans and then say That Is Why We Do Double-Blind Studies. But first of all, I'm pretty sure no one does double-blind studies with rats. Second of all, I think most social psych studies aren't double blind – I just checked the first one I thought of, Aronson and Steele on stereotype threat, and it certainly wasn't. Third of all, this effect seems to be just as common in cases where it's hard to imagine how the researchers' subtle emotional cues could make a difference. Like Schlitz and Wiseman. Or like the psychotherapy experiments, where most of the subjects were doing therapy with individual psychologists and never even saw whatever prestigious professor was running the study behind the scenes.

I think it’s a combination of subconscious emotional cues, subconscious statistical trickery, perfectly conscious fraud which for all we know happens much more often than detected, and things we haven’t discovered yet which are at least as weird as subconscious emotional cues. But rather than speculate, I prefer to take it as a brute fact. Studies are going to be confounded by the allegiance of the researcher. When researchers who don’t believe something discover it, that’s when it’s worth looking into.

V.

So what exactly happened to Bem?

Although Bem looked hard to find unpublished material, I don’t know if he succeeded. Unpublished material, in this context, has to mean “material published enough for Bem to find it”, which in this case was mostly things presented at conferences. What about results so boring that they were never even mentioned?

And I predict people who believe in parapsychology are more likely to conduct parapsychology experiments than skeptics. Suppose this is true. And further suppose that for some reason, experimenter effect is real and powerful. That means most of the experiments conducted will support Bem’s result. But this is still a weird form of “publication bias” insofar as it ignores the contrary results of hypotheticaly experiments that were never conducted.

And worst of all, maybe Bem really did do an excellent job of finding every little two-bit experiment that no journal would take. How much can we trust these non-peer-reviewed procedures?

I looked through his list of ninety studies for all the ones that were both exact replications and had been peer-reviewed (with one caveat to be mentioned later). I found only seven:

Batthyany, Kranz, and Erber: .268
Ritchie 1: 0.015
Ritchie 2: -0.219
Richie 3: -0.040
Subbotsky 1: 0.279
Subbotsky 2: 0.292
Subbotsky 3: -.399

Three find large positive effects, two find approximate zero effects, and two find large negative effects. Without doing any calculatin', this seems pretty darned close to chance for me.

Okay, back to that caveat about replications. One of Bem's strongest points was how many of the studies included were exact replications of his work. This is important because if you do your own novel experiment, it leaves a lot of wiggle room to keep changing the parameters and statistics a bunch of times until you get the effect you want. This is why lots of people want experiments to be preregistered with specific committments about what you're going to test and how you're going to do it. These experiments weren't preregistered, but conforming to a previously done experiment is a pretty good alternative.

Except that I think the criteria for "replication" here were exceptionally loose. For example, Savva et al was listed as an "exact replication" of Bem, but it was performed in 2004 – seven years before Bem's original study took place. I know Bem believes in precognition, but that's going *too far*. As far as I can tell "exact replication" here means "kinda similar psionic-y thing”. Also, Bem classily lists his own experiments as exact replications of themselves, which gives a big boost to the “exact replications return the same results as Bem’s original studies” line. I would want to see much stricter criteria for replication before I relax the “preregister your trials” requirement.

(Richard Wiseman – the same guy who provided the negative aura for the Wiseman and Schiltz experiment – has started a pre-register site for Bem replications. He says he has received five of them. This is very promising. There is also a separate pre-register for parapsychology trials in general. I am both extremely pleased at this victory for good science, and ashamed that my own field is apparently behind parapsychology in the “scientific rigor” department)

That is my best guess at what happened here – a bunch of poor-quality, peer-unreviewed studies that weren’t as exact replications as we would like to believe, all subject to mysterious experimenter effects.

This is not a criticism of Bem or a criticism of parapsychology. It’s something that is inherent to the practice of meta-analysis, and even more, inherent to the practice of science. Other than a few very exceptional large medical trials, there is not a study in the world that would survive the level of criticism I am throwing at Bem right now.

I think Bem is wrong. The level of criticism it would take to prove a wrong study wrong is higher than that almost any existing study can withstand. That is not encouraging for existing studies.

VI.

The motto of the Royal Society – Hooke, Boyle, Newton, some of the people who arguably invented modern science – was *nullus in verba*, “take no one’s word”.

This was a proper battle cry for seventeenth century scientists. Think about the (admittedly kind of mythologized) history of Science. The scholastics saying that matter was this, or that, and justifying themselves by long treatises about how based on A, B, C, the word of the Bible, Aristotle, self-evident first principles, and the Great Chain of Being all clearly proved their point. Then other scholastics would write different long treatises on how D, E, and F, Plato, St. Augustine, and the proper ordering of angels all indicated that clearly matter was something different. Both groups were pretty sure that the other had make a subtle error of reasoning somewhere, and both groups were perfectly happy to spend centuries debating exactly which one of them it was.

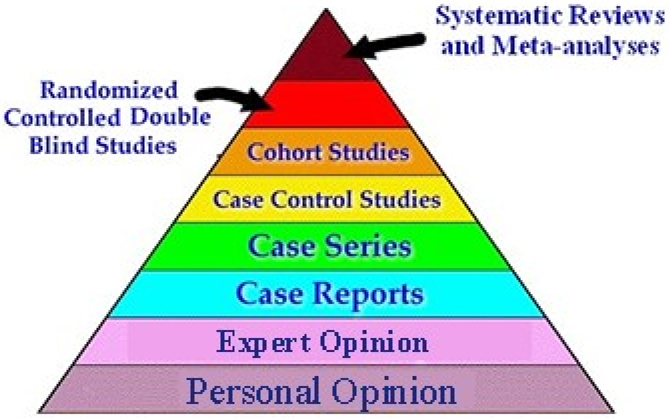
And then Galileo said “Wait a second, instead of debating exactly how objects fall, let’s just drop objects off of something really tall and see what happens”, and after that, Science.

Yes, it’s kind of mythologized. But like all myths, it contains a core of truth. People are terrible. If you let people debate things, they will do it forever, come up with horrible ideas, get them entrenched, play politics with them, and finally reach the point where they’re coming up with theories why people who disagree with them are probably secretly in the pay of the Devil.

Imagine having to conduct the global warming debate, except that you couldn’t appeal to scientific consensus and statistics because scientific consensus and statistics hadn’t been invented yet. In a world without science, *everything* would be like that.

Heck, just look at *philosophy*.

This is the principle behind the Pyramid of Scientific Evidence. The lowest level is your personal opinions, no matter how ironclad you think the logic behind them is. Just above that is expert opinion, because no matter how expert someone is they're still only human. Above that is anecdotal evidence and case studies, because even though you're finally getting out of people's heads, it's still possible for the content of people's heads to influence which cases they pay attention to. At each level, we distill away more and more of the human element, until presumably at the top the dross of humanity has been purged away entirely and we end up with pure unadulterated reality.


Systematic Reviews and Meta-analyses
Randomized Controlled Double Blind Studies
Cohort Studies
Case Control Studies
Case Series
Case Reports
Expert Opinion
Personal Opinion

The Pyramid of Scientific Evidence

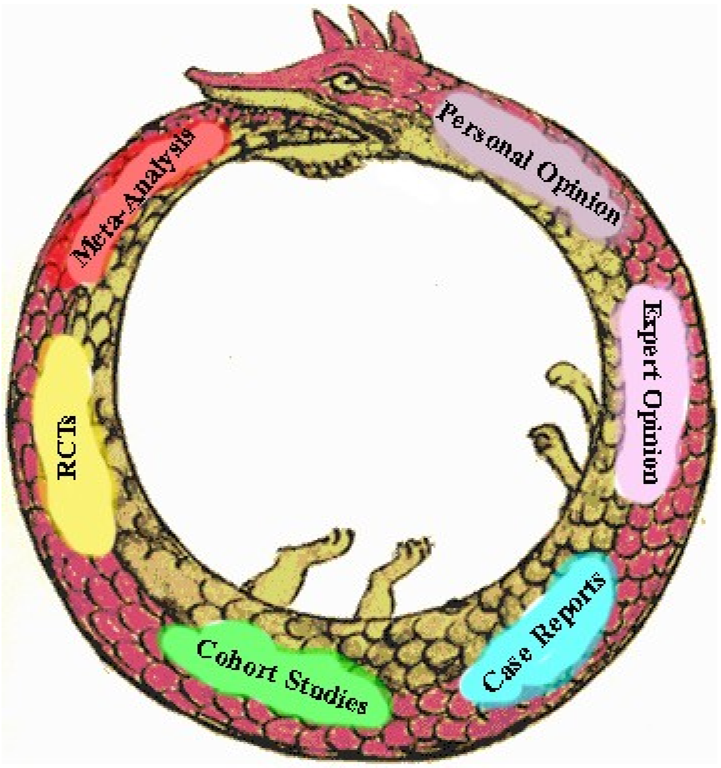
And for a while this went *well*. People would drop things off towers, or see how quickly gases expanded, or observe chimpanzees, or whatever.

Then things started getting more complicated. People started investigating more subtle effects, or effects that shifted with the observer. The scientific community became bigger, everyone didn't know everyone anymore, you needed more journals to find out what other people had done. Statistics became more complicated, allowing the study of noisier data but also bringing more peril. And a lot of science done by smart and honest people ended up being wrong, and we needed to figure out exactly which science that was.

And the result is a lot of essays like this one, where people who think they're smart take one side of a scientific "controversy" and say which studies you should believe. And then other people take the other side and tell you why you should believe different studies than the first person thought you should believe. And there is much argument and many insults and citing of authorities and interminable debate for, if not centuries, at least a pretty long time.

The highest level of the Pyramid of Scientific Evidence is meta-analysis. But a lot of meta-analyses are crap. This meta-analysis got p < 1.2 * 10^-10 for a conclusion I'm pretty sure is false, and *it isn't even one of the crap ones*. Crap meta-analyses look more like this, or even worse.

How do I know it's crap? Well, I use my personal judgment. How do I know my personal judgment is right? Well, a smart well-credentialed person like James Coyne agrees with me. How do I know James Coyne is smart? I can think of lots of cases where he's been right before. How do I know those count? Well, John Ioannides has published a lot of studies analyzing the problems with science, and confirmed that cases like the ones Coyne talks about are pretty common. Why can I believe Ioannides' studies? Well, there have been good meta-analyses of them. But how do I know if those meta-analyses are crap or not? Well…


Personal Opinion
Expert Opinion
Case Reports
Cohort Studies
RCTs
Meta-Analysis

The Ouroboros of Scientific Evidence

Science! YOU WERE THE CHOSEN ONE! It was said that you would destroy reliance on biased experts, not join them! Bring balance to epistemology, not leave it in darkness!

I LOVED YOU!!!!

Edit: Conspiracy theory by Andrew Gelman

Utopian Science

I.

Pre-emptive plagiarism is the worst. I was all set to write about how I thought the problems I brought up in The Control Group Is Out Of Control could be addressed.

Then Josh Haas wrote A Modest Proposal To Fix Science, which took the words right out of my mouth. Separate out exploratory and confirmatory research, have the latter done by different people with no stake in the matter.

So if I want to wring a blog post out of this I’m going to have to go way further than that, come up with something *really* outlandish.

So here is how science works in the utopian culture of Raikoth.

II.

Anyone can do exploratory research. It can be experiments published in special exploratory research journals. Or it can be a collection of anecdotes supporting a theory published in a magazine. Or it can be a list of arguments on a website. The point is to get an idea out there, build interest.

Remember the Angel of Evidence? The centralized nationwide prediction market? Anyone with a theory can list it there. The goal of exploratory research is to get people interested enough in the idea to bet about it on the Angel.

Suppose you become convinced that eating grapes cures cancer. So you submit a listing to the Angel: “Eating grapes cures cancer”. Probably most people doubt this proposition and the odds are around zero. So you do some exploratory research. You conduct a small poorly controlled study of a dozen cancer patients who sign up, and feed them a lot of grapes. You report that all of them seemed to have their cancer go away. You talk about how chemicals in grapes are known tumor inhibitors. Gradually a couple of people start thinking there’s something to what you’re saying. They make bets on the prediction market – maybe saying there’s only a 10% chance that you’re right, but it’s enough. The skeptics, and there are many, gladly bet against them, hoping to part gullible fools from their money. Business on the bet starts to go up.

These research prediction markets are slightly negative-sum. Maybe the loser loses $10, but the winner only gets $9. When enough people have bet on the market, the value of this “missing money” becomes considerable. This is the money that funds a confirmatory experiment.

What this means is that it is interest in – and disagreement about – the question at hand that makes an experiment possible. When no one believes grapes cure cancer, everyone’s probability is around zero and so no one bets against anyone else and there is no money available for the experiment. When you do your exploratory research and come up with good arguments why grapes should work, then if you really make your case some people should be willing to bet on it – not at even odds, maybe, but at least at 10:1 odds favoring them or whatever.

Suppose the experiment returns positive results (what qualifies as “positive results” is predefined – maybe the bet specifies “effect size > 0.4, $p < 0.05$”). Now one of two things happens. Either everyone is entirely convinced that grapes cure cancer, and people stop doing science and start eating more grapes. Or the controversy continues. If the controversy continues, a bet can be placed on the prediction market for the success or failure of a replication. No doubt people will want to take or short this bet at very different odds than they did the last one. Maybe you could get 10:1 odds against grapes curing cancer the first time, when you were going on a tiny exploratory study, but now you can only get even odds. No problem. The pro-grape side bets in favor, the anti-grape side is still willing to bet against, and the replication takes place.

Rinse and repeat. At every step, one of three things is true. First, there is still controversy, in which case the controversy funds more experiments, and the odds at which people will bet on those experiments is the degree of credence we should have in the scientific prediction involved. Second, there isn't enough money in favor of the proposition to get a market going, in which case the proposition has been soundly disproven. Third, there isn't enough money against the proposition to get a market going, in which case the proposition is universally accepted scientific fact.

In practice things are not this easy. The system is excellent at resolving controversies, and you can easily get as much money as you need to study whether guns decrease crime or whatever. But science includes not just controversies but basic research. Things like particle physics might suffer – who is going to bet for or against the proposition that the Higgs boson has a mass greater than 140 GeV? Only a couple of physicists even understand the question, and physicists as a group don't command large sums of spare capital.

So what happens is that scientific bodies – the Raikothin equivalent of our National Science Foundation – subsidize the prediction markets. This is very important. Instead of donating $1 million to CERN to do boson research, they donate $1 million to the Angel of Evidence to make the prediction market more lucrative. Suddenly the market is positive-sum; maybe you lose $10 if you’re wrong, but gain $11 if you’re right. The lure of free money is very attractive. Some ordinary people jump in, not really sure what a boson is but knowing that the odds are in their favor. But more important, so do “science hedge funds” that hire consultant physicists to maximize their likely return. Just as hedge funds in the US might do lots of research into copper mining even though they don’t care about copper at all in order to figure out which mining company is the best buy, so these “science hedge funds” would try to figure out what mass the Higgs boson is likely to have, knowing they will win big if they’re right. Although the National Science Fund type organization funds the experiments *indirectly*, it is the money of these investors that directly goes to CERN to buy boson-weighing machinery.

III.

So much for funding. How are the actual experiments conducted?

They are conducted by *consultant scientists*. The number one rule of being a consultant scientist is *you do not care about the hypothesis*.

The Raikolin would have a lot of reasons to react in horror if someone pointed them to Earth, but one of the bigger ones is that *the person who invented a hypothesis is responsible for testing it.* Or at least someone in the same field, who has been debating it for years and whose entire career depends upon it. This makes no more sense than asking criminals to judge their own trials, or having a candidate count the votes in their own election.

Having any strong opinion on the issue at hand is *immediate disqualification* for a consultant scientist to perform a confirmatory experiment.

The consultant scientist is selected by the investors in the prediction market. Corporate governance type laws are used to select a representative from both sides (those who will profit if the theory is debunked, and those who will profit if it is confirmed). Then they will meet together and agree on a consultant. If they cannot agree, sometimes they will each hire their own consultant scientist and perform two independent experiments, with the caveat that a result only counts if the two experiments return the same verdict.

As the consultant plans the experiment, she receives input from both the pro- and the con- investors. Finally, she decides upon an experimental draft and publishes it in a journal.

This publication is a form of pre-registration, but it's also more than that. It is the exact published paper that will appear in the journal when the experiment is over, except that all numbers in the results section have been replaced by a question mark, ie "We compared three different levels of grape-eating and found that the highest level had ? percent less cancer than the lowest, p < ?". The only difference between this draft and the real paper is that the real one fills in the numbers and adds a Discussion section. This gives *zero* degrees of freedom in what tests are done and in how the results are presented.

Two things happen after the draft is published.

First, investors get one final chance to sell their bets or bow out of the experiment without losses. Perhaps some investors thought that grapes cured cancer, but now that they see the experimental protocol, they don't believe it is good enough to detect this true fact. They bow out. Yes, this decreases the amount of money available for the experiment. That comes out of the consultant scientist's salary, giving her an incentive to make as few people bow out as possible.

Second, everyone in the field is asked to give a statement (and make a token bet) on the results. **This is the most important part**. It means that if you believe grapes cause cancer, and the experiment shows that grapes have no effect, you can't come back and say "Well, OBVIOUSLY this experiment didn't detect it, they used overly ripe grapes, that completely negates the anti-tumor effect, this study was totally useless and doesn't discredit my theory at all". No. When the draft is published, if you think there are flaws in the protocol, you speak then or forever hold your peace. If you are virtuous, you even say something like "Well, right now I think grapes cure cancer with 90% probability, but if this experiment returns a null result, I guess I'll have to lower that to 10%."

These statements are made publicly and recorded publicly. If you say an experiment will prove something, and it doesn't, and this happens again and again, then *people will start noticing you don't actually know any science*.

(If you're always right, you immediately get hired by a science hedge fund at an obscenely high salary.)

Finally, the consultant scientist does her experiment. Some result is obtained. The question marks in the draft are filled in, and it is resubmitted as a published paper. The appropriate people make or lose money. The appropriate scientific experts gain or lose prestige as people who are or aren't able to predict natural processes. The appropriate consultant scientists gain or lose prestige as people whose results were or weren’t replicated. The exploratory scientist who proposed the hypothesis in the first place gains or loses prestige that make people more or less likely to bet money on the next idea she comes up with.

IV.

There are no homeopaths in Raikoth.

I mean, there *were*, ages ago. They proposed experiments that could be done to prove homeopathy, and put their money where their mouth was. They lost lots of money when it turned out not to work. They added epicycles, came up with extra conditions that had to be in place before homeopathy would have an effect. Their critics were more than happy to bet the money it took to test those conditions as well, and the critics ended up rich. Eventually, the homeopaths were either broke, or sufficiently mugged by reality that they stopped believing homeopathy worked.

But homeopathy is boring. The real jewel of this system is to be able to go online, access the Angel of Evidence, and see a list of every scientific hypothesis that anyone considers worth testing, along with the probability estimate that each is true. To watch as the ones in the middle gradually, after two or three experiments, end up getting so close to zero or one hundred as makes no difference, and dropping off the “active” list. To hear the gnashing of the teeth of people whose predictions have been disconfirmed and who no longer have a leg to stand on.

Also, if you can predict the masses of bosons consistently enough, you get crazy rich.

Causal Models At Work

[Epistemic status: loosely based on a true story]

> **AIDS Specialist:** You know, people think AIDS is a death sentence. But how long do you think the average person lives after getting HIV?
> **Me:** I don't know.
> **AIDS Specialist:** Thirty years! Isn't that amazing? You can get HIV, and probably you'll live another three decades!
> **Me:** Really?
> **AIDS Specialist:** Yeah.
> **Me:** So, the trick is, give yourself HIV when you're eighty, then live to be a hundred ten.
> **AIDS Specialist:** I like you.
> **Me:** Really?
> **AIDS Specialist:** No.

And a few weeks later:

> **Me:** Do you have a family history of alcoholism?
> **Alcoholic Patient:** No. Why?
> **Me:** Well, I am trying to figure out whether to prescribe you this medication, and studies suggest it works best in patients with a genetic predisposition to alcohol abuse.
> **Alcoholic Patient:** I don't understand.
> **Me:** The medicine only works if one of your parents drinks too much.
> **Alcoholic Patient:** Well, I think I can get my dad to start.
> **Me:** NO!

Emotionally Valent Links Post

Excitement

Zipfian Academy is like App Academy, except for data science instead of programming. If for some reason my medical career collapses, I'm doing this. If not for the price tag and impossibility of getting that much time off, I'd do it anyway, just for fun.

The excitement is not only at such a good idea, but at seeing the App Academy model expand into other fields. Maybe one day there will be Academies in every marketable skill, and people won't have to go to college unless they really really want to accrue $100,000 in debt and waste four years of their lives getting drunk and sitting through General Ed bullshytte classes.

This isn't the *perfect* model, since it still permits credential-olatry and the $16,000 price tag still locks out the disadvantaged. But a $16,000 credential sure beats a $100,000 credential, and if it catches on maybe it would accrue the same system of loans and scholarships.

Limerence

International Tell Your Crush Day is coming up on May 7, conveniently three days after Star Wars Day so that your crush is forewarned that you are a nerd who likes terrible puns.

I really like this idea. I am pretty sure that for most of my life, I never told anyone my crushes because it was always easier to put it off, say "I'll definitely tell her sometime, but I don't want to suffer the fear and anxiety today". Meanwhile, summer fades to autumn, autumn into winter, and winter into now she's dating some other guy plus we're graduating soon anyway and it's not worth the trouble.

There's also something else I can't quite put my finger on, some sense of "if I tell her out of the blue it will be really weird and creepy, I have to wait for an appropriate moment". Of course this never came, or if it did I delayed it until the next appropriate moment because of the fear-and-anxiety thing.

Well, now there's a Schelling fence. The appropriate moment is May 7th.

This also fits into my plan to give the West a Japanese-style Golden Week. May Day, Star Wars Day, Cinco De Mayo, and Tell Your Crush Day. It kind of works. Spring, space battles, Mexicans, and love – the four corners of the human experience.

Terror Mixed With Weariness

In Right Is The New Left, I wrote:

> I worry that soon they'll start firing people for disagreeing with the idea that you should be able to fire people for disagreeing with ideas. Like, this could go uncomfortably far.

By "soon", I meant "within the next couple of decades". Well, the first person I know of who got fired for disagreeing with the idea that you should be able to fire people for disagreeing with ideas lost their job just under two weeks from the time I wrote that.

A big congratulations to Josh Olin for being the inaugural victim of a new meta-level of awfulness, and to the Game

Revolution article for making sure to emphasize in their headline how the firing was “understandable”. I’m not mad. If it hadn’t included that, probably they would have fired whoever wrote the article.

Confusion

[trigger warning: rape]

You may have seen this article going around: Male Rape In America. It starts by informing people that despite popular belief, men are often the victim of rape, which is of course true and important and needs to be more widely known.

But it starts by saying that under traditional definitions, “in asking 40,000 households about rape and sexual violence, the survey uncovered that 38 percent of incidents were against men.

Then it adds that “When [cases of forced penetration] were taken into account, the rates of nonconsensual sexual contact basically equalized, with 1.270 million women and 1.267 million men claiming to be victims of sexual violence.”

As far as I know neither of these two estimates include prisons, which were previously the only place where rates of male rape were thought to approach the female background level.

While male rape is definitely more of a problem than anyone wants to let on, saying that the rate between sexes is basically equal – or even as near-equality of 38-62 – is astounding to me. The strongest argument against: men are overrepresented as perpetrators of every violent crime, therefore most rapists are probably men, and most men are straight.

Weirder still, Ozy points out that the yes answers to “have you ever been raped in your lifetime” trend disproportionately female as one would expect, but yes answers to “have you been raped in the last year” trend equal between sexes. This is the reverse of my expectation, where men might have experiences of molestation from childhood or rape from a term in prison but much less as community-dwelling adults.

I await replication with interest.

Guilt

Medical resident duty hour limits have been this constant war between “exhausted doctors make more mistakes” and “doctors who are working off handoffs make more mistakes”. The battles in this war are newspaper headlines about such-and-such a patient dying because a doctor made a mistake for such-and-such a reason. For a while, the people who wanted to limit duty hours to prevent exhaustion-related mistakes have had an insurmountable advantage, and right now there are very strict laws promising DIRE CONSEQUENCES to any hospital that makes residents work more than sixteen hours a day or eighty a week (uh…thanks, I guess).

Now some evidence is starting to come out favoring the “longer hours, fewer handoffs” position, with Why Doesn’t Medical Care Get Better When Doctors Rest More? as the latest volley. See also Strictly Limiting Hours Has Not Improved Patient Safety.

I classify this under “guilt” because all of the older doctors where I work tell me about the time when Men Were Real Men and you worked thirty-six hour shifts and how nowadays we have it *too easy* and will never learn anything.

Meanwhile, I am somehow managing to get out even earlier than the duty hours mandate (TODAY I LITERALLY WORKED A NINE HOUR DAY!?!) and feeling overworked and like I am surfing on the edge of burnout. This is totally feeding into my “you do not have what it takes to be a real doctor” phobia.

Amusement

> TIL Hitler originally joined Nazis as gov agent to infiltrate them. That is SOME commitment to not blowing cover… http://t.co/VjNpLcMMq0
>
> — Scott Alexander (@slatestarcodex) May 1, 2014

> .@slatestarcodex He decisively discredited the Nazis and eliminated their effectiveness as an organization.
>
> — Benjamin R Hoffman (@ben_r_hoffman) May 1, 2014

I feel like this would make a good metaphor for superintelligent AI. If you build an insufficiently-Friendly AI and tell it “discredit this organization”, you can be certain that the organization is going to end up very, *very* discredited, more discredited than anything else has ever been.

But you might not like how it happens.

Not Even A Real Links Post, Just A Link

My post on utopian science was a little bit hasty and poorly thought out, but it did motivate Robin Hanson to further explain some of his own ideas on academic prediction markets, so I’m counting it as a victory.

Evening Doc

It’s nine o’clock on a Wednesday
The regular crowd shuffles in
There’s an old man lying next to me
Spewing vomit all over his chin

He says, “Doc, can you give me a medicine?
The name ought to be on my chart
But it’s safe and it’s strong
And it helped me along
When I had a younger man’s heart”

♪ ♬ La la la, de de da, la la la de da, da dum ♪ ♬

Give us some drugs, you’re the evening doc
Give us some drugs tonight
Well we’re all in the mood for some opiates
And they’ll get us feeling alright

Now John in bed five is a friend of mine
He gets bouts of COPD
And he’s quick with a joke or to light up a smoke
Which I guess is what brings him to me

He says, “Doc, I believe this is killing me”
I agree he is probably right
And he tells me that’s it, and he swears that he’ll quit
And he asks me to pass him a light

♪ ♬ La la la, de de da, la la la de da, da dum ♪ ♬

Now Paul is a boy with leukaemia
Who never had time for a wife
And he’s talking with Meg, who takes food through a PEG
And will probably have to for life

And the man in bed ten yells obscenities
As the lady in six gives a moan
She’s paying eight thousand for hospice care
But it’s better than dying alone

Give us some drugs, you’re the evening doc
Give us some drugs tonight
Well we’re all in the mood for some opiates
And they’ll get us feeling alright

It’s a pretty good crowd for a Wednesday
And the pharmacist gives me a vial
‘Cause he knows that it’s that, they’ve been coming to get
To forget about pain for a while

And the corridor smells like c. difficile
And the bed alarm sounds like a jeer
And I take some caffeine and I stare at my screen
And think *God, what am I doing here?*

Give us some drugs, you’re the evening doc
Give us some drugs tonight
Well we’re all in the mood for some opiates
And they’ll get us feeling alright

Not Even A Real Links Post, Just A Blatant Ad

[EDIT: This is over now. Donating now will be totally useless aside from helping make the world a better place.]

There’s a charity thing going on today where whoever gets the most unique donors of $10 or more by midnight tonight wins ~~$250,000~~ a bundle of useful services supposedly valued at $250,000.

MIRI, the Machine Intelligence Research Institute, is a group that works on ensuring a positive singularity and friendly AI. They were part-founded by Eliezer Yudkowsky, they’re closely affiliated with Less Wrong, and they briefly employed me). They are currently ~~thirteen~~ only eight donors away from the top place, with only four hours left to go (I think they’re on California time).

If you are concerned about the far future, please consider giving them a quick $10 donation at this link to tip them over the edge.

EDIT1: Someone claimed anonymous donations don’t count, so you might want to donate under your real name

EDIT2: Or if you’re interested in eye care for poor Indian children, you can donate to the current leader, Sankara Eye Foundation. Everyone else seems too far behind to catch up

Vote On Values, Outsource Beliefs

I.

Today I learned about social impact bonds. They are a thing that exists. I would expect them to be in an adequate civilization like Raikoth or dath ilan. But they are a thing that exists on Earth.

The basic idea is: government could save a lot of money if some problem got fixed. For example, if people stopped committing crime, they could spend less money on prisons. So they make a deal with a corporation. The corporation agrees to spend a certain amount of money to prevent crime for five years. And if crime goes down and the government saves on prisons, the corporation gets half the savings (or a third, or whatever).

Zero taxpayer money gets risked. It is entirely up to the corporation to fund the problem-solving effort. If they fail, then it's their own loss. If they succeed, then the government pays them money, but less than the government made, so the taxpayers still get a profit.

(The main exception I can think of is if by coincidence, crime was about to drop by 50% anyway right when the program started, and the government ends up giving half of its prison savings to the corporation for no reason. But presumably you hire a couple of mediocre economists and they are able to price out this risk. Also, a lot of the social impact bonds use a slightly different method of assessment, where they compare crime among the people the corporation has helped to crime among a control population to be sure it was the intervention that did it.)

The particular article I read about this today was How Goldman Sachs Can Get Paid To Keep People Out Of Jail. It was the name "Goldman Sachs" that got me excited. They're an investment bank. Their job is predicting risk. I don't know if they're any good at it or not. But they're the sort of organization that potentially could be. So we have people who understand risk trying to figure out what social policies will produce which results, with money riding on the decision.

This is looking *impressively* close to prediction markets. Futarchy says "vote on values, bet on beliefs". Asking a corporation to invest money in crime-solving is a form of betting on belief – they are betting on what anti-crime programs will decrease crime most and win them the most reward. You still have the elected government deciding what bonds to place – voting on values – but you're outsourcing your beliefs to the corporation involved and giving them an incentive to get it right.

Think of all the possibilities.

Right now we have a system where we don't really help people in need, unless the need becomes desperate, in which case we would feel bad about not helping, so we do, but then the cost of helping has gone up by an order of magnitude. This is exactly the sort of stupid thing that a market should be able to profit from solving.

We could have a health insurance company giving free preventative care to the poor, and the government paying them out of decreased emergency room visits.

A psychiatry clinic giving therapy to at-risk patients, and the government paying them out of decreased involuntary commitments.

A university accepting students without tuition, and the government paying them out of the increased tax revenue when they take higher-paying jobs.

Planned Parenthood offering free IUDs for women who need them, and the government paying them the money it saves from not having to put the kids through school.

Trade schools offering free classes to people on welfare, and the government paying them back from not having to give them welfare checks once they get good jobs.

I’m not sure what it means that we’re not doing those sorts of things already. But if we can’t figure out a way to solve those problems without bringing in a corporation to profit off of our incompetence, I say bring in the corporations.

II.

I think many people are against government social programs for a lot of the same reason that *The Last Psychiatrist* is against maintenance of certification exams (a position I totally called). There’s too much temptation to use it as a signal that you are Doing Something while in fact funding programs like DARE which look virtuous, but do nothing or even actively make the problem worse.

If you lean this way – and I think I do – then it is not solely out of stupidity that we wait until problems have become dire before doing anything about them. Yes, it would be great to give free job training to people on welfare and save money when they come off welfare more quickly. But actual job training programs for welfare recipients are abysmal and have been denounced as a “charade” from both the left and the right. They may be a lost cause, but I would like to see someone who has an incentive to succeed try first before writing them off – or at least get the evidence that would be provided by no such person being willing to try.

III.

For a while I was confused by the old libertarian talking point that “greed is good”. I think I could phrase it a little better now. Greed isn’t *good*, per se. It is *honest*. You know where you stand with greed. You never wonder if greed has an ulterior motive, because it’s already the most ulterior motive there is. Greed feels no temptation to corruption, because the thing it would do if it were corrupt is precisely what it’s doing anyway. Greed is like the Harlot in one of Khayyam’s rubaiyat:

> A Sheikh beheld a Harlot, and said he:
> “You seem a slave to drink and lechery”
> Replied the Harlot: “What I seem … I am!
> Are you, O Sheikh, all that **you** seem to be?”

As I see it, capitalism isn’t about worshipping greed, but about figuring out how to make greed work for good ends. So far, it has mostly tried to apply greed to get us cheap and attractive consumer products. And the amount of cheap and attractive consumer products is, like, the one thing that everyone can unambiguously agree our civilization hasn’t dropped the ball on. If we all die tomorrow and aliens discover Earth ten thousand years from now, their anthropologists will publish books saying “They sure were screwed up, but *man* did they have a lot of cheap and attractive consumer products.”

And I think some of the most exciting proposals for the future involve finding ways to use this privileged incorruptible perfectly-incentivized status of greed to do other things. Prediction markets are promising because they use greed to fix epistemology. Neocameralism is promising because it uses greed to fix governance. And social impact bonds are promising because they use greed to fix social problems.

…which isn’t to say it’s going to be easy. Ozy’s first response is that Goldman Sachs should use their $10 million to give ten thousand people in the control group a $1000 bribe each to commit a small crime; this will be more than enough to demonstrate a *vastly* reduced probability of criminality by being in the intervention group and earn Goldman $20 million.

I told Ozy zir plan is unnecessarily complex. Look at the numbers. Two hundred potential criminals. And they need a 50% decrease in jail time to meet their target and earn $20 million.

So go to the potential criminals and tell them “I’ll give you $50,000 to not commit any crimes in the next few years. $25,000 now, in order to help you solve whatever problems turned you to criminality. And $25,000 at the end, after you’ve successfully avoided jail, as a reward.” If half of them stick to it, then boom, you get $20 million and you’ve made a $10 million profit. And incentivized the next generation of criminals, but you’ve already *got* your profit, that’s the next generation’s problem.

The fact that this would *work* probably says a lot about the inefficiency of prison compared to any other conceivable way of dealing with crime. And about the profits Goldman Sachs or anyone else willing to face the inefficiency head on could make.

I don’t know if it’s exactly a *good* idea to bring in the people who caused the financial crash to help the people who came up with the prison system. But since all we’ve got is incompetent institutions, maybe sticking *different* incompetent institutions in different roles might at least shake things up a little.

Infinite Debt

I.

A patient of mine is getting to that age where she can’t support herself independently. She’s not a big fan of nursing homes, and I don’t blame her. She wants her son to take care of her.

Her son has a career, has a family, doesn’t have extra room in his house. Taking care of a sick elderly person is a full-time job, one that can involve everything from giving medications to emptying bedpans. He is not very keen on this plan.

And she says, come on, I worked hard to raise you, you owe me.

And I wonder, how far does this go?

Suppose she is going to need a decade of pretty much twenty-four hour care. She says “Well, I gave you *two* decades of essentially twenty four hour care. You owe me everything, you have to take care of me.”

Suppose she lives in a different state from him, and she really doesn’t want to leave the family home. She says “Quit your job, sell your house, and move to Michigan to take care of me. You owe me everything, you have to take care of me.”

Suppose his wife is really reluctant to share the house with a very demanding sick elderly person. Maybe she knows he works all day and realistically she’ll be the one doing the day-to-day caretaking. Maybe she’s not so keen on bedpan-emptying. Maybe she categorically refuses. And his mother says “Then divorce your wife and move to Michigan to take care of me. You owe me everything, you have to take care of me.”

Most Westerners would probably consider these requests unreasonable. But the mother is probably right. Probably she sacrificed more for her son than she’s asking him to sacrifice for her; the credit is still in her favor.

The problem with infinite debts is that they are really hard to repay.

This blog is really bad at staying away from politics for very long, so I’ll tell you what this reminds me. This reminds me of the argument some people make against libertarians: “You said you earned your money fair and square. But really, you owe a debt to society. If society hadn’t worked really hard inventing things like laws and public health, you would never have been able to found your successful business or even survive past birth. If other members of society hadn’t died fighting the Nazis and the Commies and whoever else they fought, you wouldn’t have the freedom that helped you succeed. So you owe us, and right now we’re calling in that debt in the form of a five percent higher tax rate.”

And a lot of libertarians get really angry at that argument because they don’t think it’s true, they think they don’t owe society anything.

I get really angry with that argument for the opposite reason. It is so true that one could presumably owe society *anything*.

What happens if society says “You owe us everything, pay us 10% higher taxes”?

“You owe us everything, pay us 100% higher taxes.”

“You owe us everything, go live in a cardboard box and give us the deed to your house so we can use it as a community center.”

“You owe us everything, I know you want to be a writer but what society really needs right now is oil rig workers, go become an oil rig worker.”

Of course, healthy societies do not say these things – but there have been a lot of unhealthy societies that have. I’m not worried that paying five percent higher taxes is going to lead to me being obligated to quit my job and become an oil rig worker. But the idea of handing society a blank check for anything they want out of me is pretty scary.

I kind of wonder how many of those libertarians who are so upset about a five percent tax raise would be perfectly happy with a Constitutional amendment saying “All rich people must pay 60% of their income in taxes, at which point their entire debt to society is discharged, we promise we will never raise this number above 60% or give them any grief after that, it’s right here in the Constitution.” Even if 60% was way more than a 5% raise over their current rate. 60% may be high, but it is notably lower than “infinity”. Just the acknowledgment that they’re allowed to have their own time and do things they like without being infinitely beholden to society at every moment would be pretty anxiety-relieving.

This is one reason I am so excited about Giving What We Can. Their rule is you give 10% of your income to charity, and you’re allowed in their little club and you get your name on their site as an Officially Recognized Good Person.

For years, I felt like I was *probably* ethically obligated to give all my income to charity, minus whatever I needed to survive. And the fact that I obviously wasn't going to do that made me not give anything at all.

Once someone told me that my obligation wasn't infinite, but just some finite amount like ten percent per year, every year, I was thrilled to be able to comply.

And of course there are people who make fun of this. "Oh, you really think you can just give an amount you find "convenient", then feel like your conscience is clear and you can stop caring and be smug and self-satisfied?"

The proper response to this person is to ask whether they give so much as ten percent.

("What? No, why should I?! I do *my* part by yelling at you!")

II.

But I think the opposite tendency, the tendency to deny the debt entirely, also falls short of the mark.

I mean, there *are* good arguments for doing so. You never contracted the debt. Your mother never told you as an infant, "I will raise you, but only if you agree to take care of me in any way I require for the rest of my life", and then made you sign it with your wee little baby hands, threatening to give you back to the stork if you refused. Society never said "We'll provide you with public health and technological civilization, but only if you agree to pay any tax rate we set, here's a ticket to Somalia if you refuse."

For the decision theoretic take on the question, consider a variation on the Hitchhiker Problem. You're lying unconscious in the desert, dying of thirst. A very selfish man drives by in his Jeep and considers rescuing you and bringing you to the nearest hospital. This very selfish man is only willing to go through the trouble if you pay him $100. He decides you probably will pay him $100 in gratitude for having been rescued, and so takes you to the hospital.

You wake up in the ICU, feeling cool and refreshed. The very selfish man is sitting by your bedside. “Hey,” he says. “I rescued you in the desert because I was pretty sure you’d pay me $100 for having done so. Will you give me the money?”

I think most of us would feel some obligation to give him the cash. This would be especially true if his actions were a big inconvenience to him – if he had to drive hundreds of miles out of his way, or if he had already paid the $50 doctor’s bill. Right now a lot of my ideas about morality revolve around “things that help you acausally coordinate hard decision theoretic problems”, and being willing to pay for debts you didn’t contract, as long as you still come out ahead, seems like one of those.

There’s another, more visceral argument. Imagine that my patient didn’t want her son to take 24 hour care of her. She just wanted him to come visit once a year, maybe for Christmas. “Please,” she says, “I’m really lonely and it would mean a lot to me to have you around.”

Now, that guy is under no obligation to go visit random elderly women on Christmas, even if those random elderly women would like company. But most of us would say he *is* under some obligation to visit his mother, or at least that he would be a pretty bad person if he refused. And it’s not just that she is closer to him. Most of us would cash out that obligation in terms of “Your mother did so much for you, can’t you do at least a little for her?”

This is a weird position. You can defend the guy having no debt to his mother. You can defend the guy having an infinite debt to his mother. But a *small* debt to his mother? Where does that come from?

III.

I struggle with this concept a lot. I don't know the moral answer. But I do know the practical answer. Infinite debts make everyone miserable and tend not to be paid at all.

And the moral and practical are sometimes pretty closely aligned. If infinite debts make *everyone* miserable, then by the decision theoretic definition of morality above we might decide to forgive them. After all, every mother was also a child, and it may be that, at the point where you're making timeless acausal Platonic contracts, everyone agrees to free their child from a debt to them as long as they are themselves freed from debt to their parents. Maybe they would agree that infinite creditors deserve a certain level of respect, very very high respect, but not enough that it ruins your life. This is frustratingly nontechnical. But timeless Platonic contracts are notoriously bad at giving specific figures. Sometimes you've just got to seize whatever you can get, find some socially sanctioned middle ground that doesn't make you feel like you're a bad person or that you can never enjoy yourself and draw a huge bright line there and defend it to the bitter end.

Maybe your infinite obligation to those worse off than you demands ten percent of your income.

And your infinite obligation to society demands that you pay tax at the prevailing rate.

And your infinite obligation to your mother demands that, at the very least, you call her up on Mother's Day and tell her thanks.

None of those things clear your debt, exactly. But they keep you in good standing. You pay a tiny fraction of your debt, year after year, and it keeps the moral repo man from your door.

The problem with infinite debts is that they are really hard to repay. On the other hand, the *interest* can be quite manageable.

Weak Men Are Superweapons

I.

There was an argument on Tumblr which, like so many arguments on Tumblr, was terrible. I will rephrase it just a little to make a point.

Alice said something along the lines of “I hate people who frivolously diagnose themselves with autism without knowing anything about the disorder. They should stop thinking they’re ‘so speshul’ and go see a competent doctor.”

Beth answered something along the lines of “I diagnosed myself with autism, but only after a lot of careful research. I don’t have the opportunity to go see a doctor. I think what you’re saying is overly strict and hurtful to many people with autism.”

Alice then proceeded to tell Beth she disagreed, in that special way only Tumblr users can. I believe the word “cunt” was used.

I notice two things about the exchange.

First, why did Beth take the bait? Alice said she hated people who *frivolously* self-diagnosed *without knowing anything about the disorder*. Beth clearly was not such a person. Why didn’t she just say “Yes, please continue hating these hypothetical bad people who are not me”?

Second, why did *Alice* take the bait? Why didn’t she just say “I think you’ll find I wasn’t talking about you?”

II.

One of the cutting-edge advances in fallacy-ology has been the weak man, a terribly-named cousin of the straw man. The straw man is a terrible argument nobody really holds, which was only invented so your side had something easy to defeat. The weak man is a terrible argument that only a few unrepresentative people hold, which was only *brought to prominence* so your side had something easy to defeat.

For example, “I am a proud atheist and I don’t like religion. Think of the terrible things done by religion, like the actions of the Westboro Baptist Church. They try to disturb the funerals of heroes because they think God hates everybody. But this is horrible. Religious people can’t justify why they do things like this. That’s why I’m proud to be an atheist.”

It’s not a straw man. There really is a Westboro Baptist Church, for some reason. But one still feels like the atheist is making things just a little too easy on himself.

Maybe the problem is that the atheist is indirectly suggesting that Westboro Baptist Church is typical of religion? An implied falsehood?

Then suppose the atheist posts on Tumblr: “I hate religious people who are rabidly certain that the world was created in seven days or that all their enemies will burn in Hell, and try to justify it through ‘faith’. You know, the sort of people who think that the Bible has all the answers and who hate anyone who tries to think for themselves.”

Now there’s practically no implication that these people are typical. So that’s fine, right?

On the other side of the world, a religious person is writing “I hate atheists who think morality is relative, and that this gives them the right to murder however many people stand between them and a world where no one is allowed to believe in God”.

Again, not a straw man. The Soviet Union contained several million of these people. But if you’re an atheist, would you just let this pass?

How about “I hate black thugs who rob people”?

What are the chances a black guy reads that and says “Well, good thing I’m not a thug who robs people, he’ll probably *love* me”?

III.

What is the problem with statements like this?

First, they are meant to re-center a category. Remember, people think in terms of categories with central and noncentral members – a sparrow is a central bird, an ostrich a noncentral one. But if you live on the Ostrich World, which is inhabited only by ostriches, emus, and cassowaries, then probably an ostrich seems like a pretty central example of ‘bird’ and the first sparrow you see will be fantastically strange.

Right now most people’s central examples of religion are probably things like your local neighborhood church. If you’re American, it’s probably a bland Protestant denomination like the Episcopalians or something.

The guy whose central examples of religion are Pope Francis and the Dalai Lama is probably going to have a different perception of religion than the guy whose central examples are Torquemada and Fred Phelps. If you convert someone from the first kind of person to the second kind of person, you’ve gone most of the way to making them an atheist.

More important, if you convert a culture from thinking in the first type of way to thinking in the second type of way, then religious people will be unpopular and anyone trying to make a religious argument will have to spend the first five minutes of their speech explaining how they’re not Fred Phelps, honest, and no, they don’t picket any funerals. After all that time spent apologizing and defending themselves and distancing themselves from other religious people, they’re not likely to be able to make a very rousing argument for religion.

IV.

In Cowpox of Doubt, I mention the inoculation effect. When people see a terrible argument for an idea get defeated, they are more likely to doubt the idea later on, even if much better arguments show up.

Put this in the context of people attacking the Westboro Baptist Church. You see the attacker win a big victory over “religion”, broadly defined. Now you are less likely to believe in religion when a much more convincing one comes along.

I see the same thing in atheists’ odd fascination with creationism. Most of the religious people one encounters are not young-earth creationists. But these people have a dramatic hold on the atheist imagination.

And I think: well, maybe if people see atheists defeating a terrible argument for religion enough, atheists don’t *have to* defeat any of the others. People have already been inoculated against religion. “Oh, yeah, that was the thing with the creationism. Doesn’t seem very smart.”

If this is true, it means that all religious people, like it or not, are in the same boat. An atheist attacking creationism becomes a deadly threat for the average Christian, even if that Christian does not herself believe in creationism.

Likewise, when a religious person attacks atheists who are moral relativists, or communists, or murderers, then all atheists have to band together to stop it somehow or they will have successfully poisoned people against atheism.

V.

This is starting to sound a lot like something I wrote on my old blog about superweapons.

I suggested imagining yourself in the shoes of a Jew in czarist Russia. The big news story is about a Jewish man who killed a Christian child. As far as you can tell the story is true. It’s just disappointing that everyone who tells it is describing it as “A Jew killed a Christian kid today”. You don’t want to make a big deal over this, because no one is saying anything objectionable like “And so all Jews are evil”. Besides you’d hate to inject identity politics into this obvious tragedy. It just sort of makes you uncomfortable.

The next day you hear that the local priest is giving a sermon on how the Jews killed Christ. This statement seems historically plausible, and it’s part of the Christian religion, and no one is implying it says anything about the Jews today. You’d hate to be the guy who barges in and tries to tell the Christians what Biblical facts they can and can’t include in their sermons just because they offend you. It would make you an annoying busybody. So again you just get uncomfortable.

The next day you hear people complain about the greedy Jewish bankers who are ruining the world economy. And really a disproportionate number of bankers are Jewish, and bankers really do seem to be the source of a lot of economic problems. It seems kind of pedantic to interrupt every conversation with “But also some bankers are Christian, or Muslim, and even though a disproportionate number of bankers are Jewish that doesn’t mean the Jewish bankers are disproportionately active in ruining the world economy compared to their numbers.” So again you stay uncomfortable.

Then the next day you hear people complain about Israeli atrocities in Palestine (what, you thought this was past czarist Russia? This is future czarist Russia, after Putin finally gets the guts to crown himself). You understand that the Israelis really do commit some terrible acts. On the other hand, when people start talking about “Jewish atrocities” and “the need to protect Gentiles from Jewish rapacity” and “laws to stop all this horrible stuff the Jews are doing”, you just feel worried, even though you personally are not doing any horrible stuff and maybe they even have good reasons for phrasing it that way.

Then the next day you get in a business dispute with your neighbor. Maybe you loaned him some money and he doesn’t feel like paying you back. He tells you you’d better just give up, admit he is in the right, and apologize to him – because if the conflict escalated everyone would take his side because he is a Christian and you are a Jew. And everyone knows that Jews victimize Christians and are basically child-murdering Christ-killing economy-ruining atrocity-committing scum.

You have been boxed in by a serious of individually harmless but collectively dangerous statements. None of them individually referred to you – you weren’t murdering children or killing Christ or owning a bank. But they ended up getting you in the end anyway.

Depending on how likely you think this is, this kind of forces Jews together, makes them become strange bedfellows. You might not like what the Jews in Israel are doing in Palestine. But if you think someone’s trying to build a superweapon against you, and you don’t think you can differentiate yourself from the Israelis reliably, it’s in your best interest to defend them anyway.

VI.

I wrote the superweapon post to address some of my worries about feminism, so it would not be surprising at all if we found this dynamic there.

Feminists tend to talk about things like “Men tend to silence women and not respect their opinions” or “Men treat women like objects rather than people” or “Men keep sexually harassing women even when they make it clear they’re not interested”.

Put like that, it’s obvious why men might complain. But maybe some of the more sophisticated feminists say “Some men tend to silence women and not respect their opinions”. Or “Some men keep sexually harassing women even when they make it clear they’re not interested.”‘

And the weak-man-superweapon model would suggest that even this weakened version would make lots of men really uncomfortable.

From feminist website Bitchtopia (look, I don’t name these websites, I just link to them): Not All Men Are Like That:

> I’ve heard this counter-argument almost every single time I’ve tried to bring up a feminist issue with a man: “but not all men are like that!”…

> Having to point out that not every man exhibits explicitly harmful behavior allows for oppression to continue because having to say “some men do harmful things” gives oppressors peace of mind…
>
> Sure, white men–you were brought up to feel entitled to anything you wanted and now you see anyone trying to have opportunities equal to yours as a threat…
>
> When you say, “not all men are like that!” what you’re really saying is, “I don’t want to have to think about my privilege as a white man, so I’m going to try to defer the blame to other guys because I clearly don’t act like that.”
>
> Nice try.

Remember, not wanting to be stereotyped based solely on your sex is the *most* sexist thing!

This is not just an idiosyncracy of Bitchtopia (look! I’m sorry! I swear I didn’t name that website!). There’s also an entire notallmenarelikethat dot tumblr dot com (of *course* there is) and it’s now a feminist meme abbreviated NAMALT.

But of course, it’s not just feminists. The gender-flipped version of feminism has the same thing. From men’s rights blog “The Spearhead”, which is not quite as badly named but still kind of funny if you think of it in a Freudian way:

> Talking about the current sad state of dating and marriage in the USA will often elicit “Not All Women Are Like That” or NAWALT.
>
> The first thing is not to contradict whoever makes that claim. Why? Because it is true. Not all women are

> skanks, attention whores or predators. The MRA cause is not helped by attacking people who speak truthfully.
>
> [But the consequence of a] false positive is that a man ends up married to a skank, sociopath or gold digger. The cost of bad wife selection is so high that he is forced to turn away good women for fear of mistakenly choosing a bad one.

More polite and scientific than the feminist version, but the point is he expects men's rights readers to be so familiar with "not all women are like that" that he's perfectly comfortably abbreviating it NAWALT. Apparently there's even a NAWALT video.

I don't know where to find neo-Nazi blogs, but I'll bet if there are some, they have places where they talk about how annoying it is when people try to distract from the real issues by using the old NAJALT.

VII.

But I shouldn't make fun of NAJALT. There really are two equal and opposite problems going on here.

Imagine you're an atheist. And you keep getting harassed by the Westboro Baptist Church. Maybe you're gay. Maybe you're not. Who knows why they do what they do? Anyway, they throw bricks through your window and send you threatening letters and picket some of your friends' funerals.

And you say "People! We really need to do something about this Westboro Baptist Church! They're horrible people!"

And you are met by a wall of religious people saying "Please stop talking about the Westboro Baptist Church, you are making us look really bad and it’s unfair because not all religious people are like that.”

And you say “I really am not that interested in religion, I just want them to stop throwing bricks through my window.”

And they say “Hey! I thought we told you to stop talking about them! You are unfairly discrediting us through the inoculation effect! That is epistemically unvirtuous!”

So the one problem is that people have a right not to have unfair below-the-belt tactics used to discredit them without ever responding to their real arguments.

And the other problem is that victims of nonrepresentative members of a group have the right to complain, even though those complaints will unfairly rebound upon the other members of that group.

Atheists who talk about the Westboro Baptist Church may be genuinely concerned about the Westboro Baptist Church. Or they may be unfairly trying to tar all religious people with that brush. Religious people have to fight back, even though the Westboro Baptists don’t deserve their support, because otherwise the atheists will have a superweapon against them. Thus, a stupid fight between atheists who don’t care about Westboro and religious people who don’t support them.

VIII.

This gives me some new views on political coalitions. I always thought that having things like political parties was stupid. Instead of identifying as a liberal and getting upset when someone insulted liberals or happy when someone praised liberals, I should say “These are my beliefs. There are other people who believe approximately the same thing, but the differences are sufficient that I just want to be judged on my own individual beliefs alone.”

The problem is, that doesn’t work. It’s not my decision whether or not I get to identify with other liberals or not. If other people think of me as a liberal, then anything other liberals do is going to reflect, positively or negatively, on me. And I’m going to have to join in the fight to keep liberals from being completely discredited, or else the fact that I didn’t share any of the opinions they were discredited for isn’t going to save me. I will be Worst Argument In The World-ed and swiftly dispatched.

In the example we started with, Beth chose to stand up for the people who self-diagnosed autism without careful research. This wasn’t because she considered herself a member of that category. It was because she decided that self-diagnosed autistics were going to stand or fall as a group, and if Alice succeeded in pushing her “We should dislike careless self-diagnosees” angle, then the fact that she wasn’t careless wouldn’t save her.

Alice, for her part, didn’t bother bringing up that she never accused Beth of being careless, or that Beth had no stake in the matter. She saw no point in pretending that boxing in Beth and the other careful self-diagnosers in with the careless ones wasn’t her strategy all along.

Getting A Therapist

Several people have emailed me asking for advice finding a therapist, and I worry I haven't been much help. One of the advantages of being a psychiatrist is not having to worry about finding a psychiatrist. I know a few people in southern Michigan personally and can give some advice if you live there; outside that I'm stuck.

Luckily Gruntled and Hinged just wrote a good series of articles on how to find and contact a therapist for mental health problems. I'll be referring those kinds of questions there in the future.

But I'll also open up one question to anyone who can help: anyone want to volunteer the name of a good therapist whom they would recommend to other readers? Clearly geography is an issue here, so I would be most interested in knowing about tele-therapists or people in the Bay Area / New York City where a lot of SSC readers tend to congregate.

Bay Area:
Dr. Judith Sachs (recommended by David Chapman)

New York City:
Deb Shapiro (recommended by jayshap)

Greater Boston:
Integrative Psychology (recommended by Hibiscus)
Life Changes Group (recommended by Hibiscus)

Ann Arbor:
Rebecca Hatton (recommended by Paul Torek)

[Ask Scott privately for psychiatrist recommendations in Ann Arbor / Detroit]

You Kant Dismiss Universalizability

I.

Like most right-thinking people, I'd always found Immanuel Kant kind of silly. He was the standard-bearer for naive deontology, the "rules are rules, so follow them even if they ruin everything" of moral philosophy.

But lately, I've been starting to pick up a different view. There may have been some subtleties I was missing, almost as if one of the most universally revered thinkers of the western philosophical tradition wasn't a total moron.

I was delighted to see nydwracu say something similar in the comments to my recent post:

> I [now] realize that Kant is not actually completely ridiculous like I once thought he was

I don't know if it's just that nydwracu and I have been thinking about some of the same problems lately, but he took the words right out of my mouth.

I'm not a Kant scholar. I'm not qualified to explain what Kant thought, and it's possible the arguments I express as Kantian here are going to be arguments of a totally different person who merely reminds me of Kant in some ways. James Donald's objections to steelmanning are well taken, so I will not call this a steel man of a guy who is too dead to correct me if I am wrong. At best I will call this post Kant-aligned.

First, I want to take another look at one of Kant's most-reviled arguments: that you should truthfully tell a murderer who wants to kill your friend where she is hiding.

Second, I want to talk about how I find myself using Kantian principles in my own morality.

And third, I want to talk about big unanswered questions and the reason this still isn't technical enough for me to be comfortable with.

II.

Kant gives the following dilemma. Suppose that an axe murderer comes to your door and demands you tell him where your friend is, so that he can kill her. Your friend in fact is in your basement. You lie and tell the murderer your friend is in the next town over. He heads off to the next town, and while he's gone you call the police and bring your friend to safety.

Most people would say that the lie is justified. Kant says it isn't, because lying.

I think most people understand his argument as follows: you think "I should lie". But suppose everyone thought that all the time. Then everyone would lie to everyone else, and that would be horrible.

But Kant's categorical imperative doesn't urge us to reject actions which, if universalized, would be horrible. That's rule utilitarianism, sort of. Kant urges us to reject actions which, if universalized, would be self-defeating or contradictory.

Suppose it was everyone's policy to lie to axe murderers who asked them where their friends were. Well, then axe murderers wouldn't even bother asking.

Which doesn't sound like a sufficiently terrible dystopia to move us very much. So let me reframe Kant's example.

Suppose you are a prisoner of war. Your captors tell you they want to kill your general, a brilliant leader who has led your side to victory after victory. They have two options. First, a surgical strike against her secret headquarters, killing her and no one else. Second, nuking your capital city. They would prefer to do the first, because they're not monsters. But if they have to nuke your capital, they'll nuke your capital. So they show you a map of your capital city and say "Please point out your general's headquarters and we'll surgical-strike it. But if you don't, we'll nuke the whole city."

You decide to lie. You point to a warehouse you know to be abandoned. Your captors send a cruise missile that blows up the warehouse, killing nobody. Then they hold a huge party to celebrate the death of the general. Meanwhile, the real general realizes she's in danger and flees to an underground shelter. With her brilliant tactics, your side wins the war and you are eventually rescued.

So what about now? Was your lie ethical?

Kant would point out that if it was known to be everyone's policy to lie about generals' locations, your captors wouldn't even ask. They'd just nuke the city, killing everyone.

Your captors are offering you a positive-sum bargain: "Normally, we would nuke your capital. But you don't want that and we don't want that. So let's make a deal where you tell us where your general is and we only kill that one person. That leaves both of us better off."

If it is known to everyone that prisoners of war always lie in this situation, it would be impossible to offer the positive-sum bargain, and your enemies would resort to nuking the whole city, which is worse for both of you.

So when Kant says not to act on maxims that would be self-defeating if universalized, what he means is “Don’t do things that undermine the possibility to offer positive-sum bargains.”

This is *very* reminiscent of Parfit’s Hitchhiker. Remember that one? You are lost in the desert, about to die. A very selfish man drives by in his dune buggy, sees you, and offers to take you back to civilization for $100. You don’t have any money on you, but you promise to pay him $100 once you’re back to civilization and its many ATMs. The very selfish man agrees and drives you to safety. Once you’re safe, you say “See you later, sucker!” and run off.

The selfish man’s “I’ll bring you back to civilization for $100” offer is a positive-sum bargain. You would rather lose $100 than die. He would rather gain $100 and lose a few hours bringing you to the city than continue on his way. So you both gain.

But if everyone were omniscient and knew that people who promise $100 will never really pay, or if your decision not to pay could somehow affect his willingness to make you the offer in the first place, the ability to make the positive-sum bargain disappears.

On this model, Kant isn’t being a weird super-anal stickler for meaningless rules at all. He’s being the most practical person around: don’t do things that spoil people’s ability to make a profit.

(and sort of pre-inventing decision theory)

(man, it’s a good thing everyone is omniscient and the future can cause the past, or else we’d never be able to ground morality *at all*)

III.

A while back I suggested it is wrong to fire someone for being anti-gay, because if every boss said “I will fire my employees whom I disagree with politically”, or every mob of angry people said “We will boycott companies until they fire the people we disagree with politically” then no one who’s not independently wealthy could express any political opinions or dare challenge the status quo, and the world would be a much sadder place.

This is not strictly Kantian. “The world would be a much sadder place” is not self-defeating or a contradiction.

But it could still be framed as a positive-sum bargain. In a world where all the leftists refused to hire rightists, and all the rightists refused to hire leftists, everything would be about the same except that everyone’s job opportunities would be cut in half. If the people in such a world were halfway rational, they would make a deal that rightists agree to hire leftists if leftists agree to hire rightists. This would clearly be positive-sum.

This is easy to say in natural language like this. But when you try to make it more formal it gets really sketchy real quick.

Let’s say Paula the Policewoman is arresting Robby the Robber (she caught him by noticing his name was Robby in a world where everyone’s name sounds like their most salient characteristic). No doubt she thinks she is following the maxim “Police officers should arrest robbers”. But what about other maxims that lead to the same action?

1. Police officers should arrest people
2. Everyone should arrest robbers
3. Paula should arrest Robby

4. Paula should arrest other people
5. Everyone should arrest Robby
6. Everyone should arrest EVERYONE ELSE IN THE WORLD

This sounds kind of silly in this context, but in more complicated situations the entire point hinges upon it.

Levi the Leftist, who owns a restaurant called Levi's Lentils, finds out that his head waiter, Riley the Rightist, is a homophobe (in Levi's defense, he thought he was safe to hire him because his name wasn't Homer). He fires Riley, who ends out on the street.

Candice the Kantian condemns him, saying "What if that were to become a general rule? Then nothing would change except everyone only has half as many job opportunities."

Levi says "Oh, I see your problem. You think my maxim is 'fire people with different politics than me'. But that's not my maxim at all. My maxim is 'fire people who are homophobic'. If that becomes universalized, it will be a great victory for gay people everywhere, but no one whose politics I agree with will suffer at all."

In fact, Levi might claim his maxim is any one of the following:

1. Everyone should fire people they disagree with politically
2. Everyone should fire people who are politically on the right
3. Everyone should fire people who discriminate against minority groups
4. Everyone should fire people who are homophobic
5. Everyone should fire people who are mean and hateful
6. Everyone should fire people who hold positions that are totally beyond the pale and can't possibly be supported rationally

(before I get yelled at in the comment section, I’m not necessarily claiming all these maxims accurately describe Riley, just that Levi might think they do)

(5) runs into this problem where you can never say “fire people who are mean and hateful” without it *in fact* meaning “fire people whom *you think* are mean and hateful”. Presumably all the rightist bosses will find good reasons to think their leftist employees are mean and hateful.

There seems to be some sense in which we also want to protest (2), say that if Levi is allowed to use (2), then that instantly morphs to rightist bosses being allowed to say “everyone should fire people who are politically on the left”. But just saying “universalizability!” doesn’t automatically let us do that.

(3) seems even sneakier. It is in fact the maxim promoted by the people who are actually doing the firing, since they seem to have some inkling that universalizability and “fairness” are important. And it sounds totally value-neutral and universalizable. And yet I feel like if we allow Levi to say this, then some rightist will say actually *his* maxim is “everyone should fire people who want to undermine traditional cultural institutions”, and the end result will be the same old “job opportunities halved for everyone”.

IV.

This is a hard problem. The best solution I can think of right now is to go up a meta-level, to say “universalize as if the process you use to universalize would itself become universal”.

Suppose I am very greedy, and I lie and steal and cheat to get money. I say “Well, my principle is to always do whatever gets Scott the most money”. This sooooooorta checks out. If it were universalized – and everyone acted on the principle “Always do whatever gets Scott the most money”, well, I wouldn’t mind that at all.

But if we say “universalize as if the process you use to universalize would itself become universal”, then we assume that if I try to universalize to “do what gets Scott the most money”, then Paula will universalize to “do what gets Paula the most money” and Levi will universalize to “do what gets Levi the most money” and we’ll all be lying and cheating and stealing from one another and no one will be very happy at all.

(Kant notes that this also satisfies his original, stricter “self-defeating contradiction” criterion. If we all try to steal from each other, then private property becomes impossible, the economy collapses, and the stuff we want isn’t there to steal. I don’t know if I like this; it seems a little forced. But even if contradictoriness is forced, badness seems incontravertible)

As for Levi, he knows that if he universalizes to “everyone should fire people who discriminate against minority groups”, his process is “pick out a political value that’s important to me and excludes a lot of potential employees, then say everyone should fire people who disagree with it”. This is sufficient to assume rightists will do the same and we’ll be back at half-as-many-jobs.

Next problem. Suppose I am a very rich and very selfish tycoon. I say “No one should worry about helping the needy”. I am perfectly happy with this being universalized, because it saves me from having to waste my time helping the needy. Although other people also won’t help the needy, I’m a super-rich tycoon and that’s no skin off *my* back.

We can climb part of the way out of this pit with meta-universalizability. We say “If I say things like this, everyone will only act on maxims that benefit them personally and appeal to their own idiosyncratic characteristics, rather than the ones that most benefit everyone.”

But I worry that this isn’t enough. Suppose I’m not just a tycoon, I’m a super-rich and powerful tyrannical king. I come up with maxims like “Everyone do what the tyrant says or be killed!” Candice the Kantian warns “If you do that, everyone will come up with maxims that benefit them personally, and the moral law will be weakened.”

And so I kill Candice for disagreeing with me.

If you are so much stronger than other people that you are immune to their counter-threats, you can get away with doing pretty much anything under this perversion of not-at-all-like-Kant we’ve wandered into.

We might have gotten so far from Kant at this point that we’ve stumbled into Rawls. Put up a veil of ignorance and the problem vanishes.

V.

What about utiltarianism?

I would love to universalize the maxim “Do whatever most increases Scott’s utility”.

Given concerns of meta-univeralizability above, I might end up instead wanting to universalize “Do whatever most increases global utility”.

This seems certain, maybe even provable, if you throw in the veil of ignorance accessory.

Utilitarianism has a lot of the same problems universalizability does. A very stupid utilitarian would automatically condemn Levi for firing Riley since now Riley is unemployed and this lowers his utility. More sophisticated utilitarians would have to take into account the various society-wide effects of Levi setting a precedent here. I think that's what Mill's rule utilitarianism tries to do and what precedent utilitarianism tries to do as well. The problem is that it's really hard to figure out what rules and precedents have how much weight. Universalizability kind of plows through some of those objections like a giant steamroller. It probably prevents a couple of little incidents where you could steal something or kill someone to gain a little extra utility, but it more than makes up for it in vastly increasing social trust and ability for positive-sum deals.

I'm not sure whether consequentialism is prior to universalizability ("universalize maxims because if you don't you'll end up losing out on possible positive-sum games and cutting your job offers in half"), whether universalizability is prior to consequentialism ("be a consequentialist, because that is a maxim everyone could agree on"), or whether they're like a weird ouroboros constantly eating itself.

I think maybe the idea I like best is that consequentialism is prior to universalizability is prior to any particular version of utilitarianism.

Because if universalizability is prior, that would be an interesting way to explore some of the problems with utilitarianism. For example, should we count pleasure or preferences? I don't know. Let's see what everyone would agree on.

Does everyone have to donate all of their money to the most efficient charity all the time? Well, if you were behind the veil of ignorance helping frame the moral law, would you put that in?

Does everyone have to prefer torture to dust specks? You're behind the veil of ignorance, you don't know if you'll be a dust speck person or a torture person, what do *you* think?

I think this is a good point to remember the blog tagline and admit I am still confused, but on a higher level and about more important things.

Nerds Can Be Bees, Too

Jonathan Haidt has a saying: “People are ninety percent chimp and ten percent bee”. It’s supposed to mean that people usually push each other around selfishly to gain status, but occasionally have an ability to come together into a single unified superorganism working for the common good. He seems to reify this more than a little, treating it as a “switch” that can be turned on by certain situations or rituals. He gives lots of examples, but four that stick out for me are patriotism, prayer, team sports, and pep rallies.

They stick out because they’re a pretty good list of the things that most turned me off when I was younger.

I was definitely one of those people who fact-checked patriotism: “America is number one? Really? Then how come Canada has lower crime, lower poverty rate, lower infant mortality, and higher self-rated life satisfaction?” The feeling engendered by an image of an eagle flying in front of the American flag while the Star-Spangled Banner played in the background was a combination of cringing and urge to nitpick.

As for prayer, I distinctly remember putting my very progressive rabbi on the spot about whether God *literally* existed, and when the rabbi answered something like “Well, it depends how literally you want to take the word literal…” I asked him why we spent so much time praying. He replied that he liked to do things for no reason because he was stupid. Or, well, probably that wasn’t actually what he said, but that is pretty much how my brain remembers whatever weasel answer he gave. I stopped going to synagogue very shortly after that.

Team sports always seemed moderately barbaric. When forced to participate, I treated them about the same way I treat being on call in hospital as an intern – desperately pray that none of the activity happens in an area I am responsible for, frantically try to transfer it to some more qualified person when it does, and make burnt offerings of thanksgiving to the gods of every major world religion when it's over.

But pep rallies were the worst of all. It wasn't just that they were celebrations of the kind of guys who would yell incomprehensible things at me, led by the kind of girls who would preemptively tell me they would never date me even though I hadn't asked. It was just…the inexplicable, horrible noise. Like I'm sure everyone's had those times when one of the dogs in your neighborhood starts barking, and then that makes another dog start barking, and soon all of the dogs in your neighborhood are barking really loud for no real reason. And I would always think, "Well, it's not supposed to make sense to me, I'm not a dog." But pep rallies were the same thing, and I didn't have that excuse!

So I came out of all this stuff figuring I lacked what Haidt calls "the hive switch", the ability for the right trigger to take you outside yourself and bring you into ecstatic union with an in-group. And most of the people I most respected felt the same way. It was even a point of pride: "I'm the sort of person who can *see through* pep rallies and isn't stupid enough to start screaming with the rest of them."

And I noticed this same thing in the epic comment thread to my What Universal Human Experiences Are You Missing Without Realizing It. Here are some quotes from readers:

> "There seems to be a set of social-emotions that I don't quite get. Like when people say they're 'offended'. The

idea of getting fist-fightingly violent over mildish insults, or teary because some symbol has been disrespected is just confusing to me."

"I don't get political rallies. You know the ones, where some major politician goes in front of a big crowd and pours on the charisma, and everybody is cheering and shouting all at once? I used to live in Iowa, so I've had the opportunity to be in a bunch of those crowds, and the whole thing always seemed… just completely baffling."

"I've never been to a rally, but I also don't get them. In fact, I find myself actively creeped out by many forms of collective displays of emotion/enthusiasm."

"I have never been able to take any ritual seriously. As a kid in Church I always thought everyone was just playing along, that nobody actually believed the stuff we were talking about, the same way nobody actually believes in Superman. I always expected someone to break character and then everybody to start laughing."

"I have never really experienced communal grief. The idea that something like an attack or natural disaster that kills or injures strangers (even if they happen to share citizenship of a country, state, or city with you) or the death of a well-known public figure could elicit an emotion resembling what you feel when someone you know dies does not make intuitive sense to me."

"I don't understand or don't enjoy many group activities, including parties, group conversations, watching team sports (such as football), participating in team sports, political rallies, concerts."

I have heard a very attractive explanation for this. Being a nerd, goes the explanation, is sort of like autism. And autistic people are missing a lot of the brain's normal social machinery. So nerds are just autistic enough to be missing the hive switch.

It is never a good idea to underestimate human variation. But this has not been my experience.

Since about age fourteen, I've been involved in "micronations", a weird hobby treading an uncomfortable line between roleplaying game and secessionist movement. Groups of friends come together to design countries which engage in various strange and often confrontational forms of politics and foreign policy.

And most of the teenaged friends who I worked with on this had the same attitudes toward patriotism as teenaged me – it was stupid, Canada was better than America, why do people waste money on stupid flag pins, et cetera.

And we became *fanatically* protective of these tiny little fake nations of ours, and people who would have sold out the US for a nickel would spend sleepless nights zealously defending the reputation of a country whose population was in the single digits and whose constitutions included sections like "Article Five: Design for a judicial system to go here eventually".

And I have noticed something similar even out here in the world of countries larger than a backyard. Anyone following the beautiful fountain of drama that is Justine Tunney's Twitter will notice her relationship with Google borders on the same fervency that marked our micronational patriotism. And then there are the great nerd cults, like

Objectivism and the one that I’m not supposed to use in the same sentence as “cult” for search-engine-related reasons.

I remember how back during the 2008 presidential election, when Michelle Obama saw that the tide was turning in Barack’s favor she said something like “For the first time, I feel proud to be an American”. She got a lot of flak about that from the press, but I think she was honest and I think was a perfectly understandable feeling (albeit not the best thing to say if you’re the wife of a presidential nominee). It’s the feeling of going from an out-group you can’t identify with to an in-group you can.

I think the thing with nerds and hive switches is the same. It’s not that we lack the ability to lose ourselves in an in-group, it’s that all the groups people expected us to lose ourselves in weren’t ones we could imagine as our in-group by any stretch of the imagination. We didn’t cheer on the jocks and cheerleaders at pep rallies not because the secret switch in our brain was broken, but because the secret switch in our brain didn’t think jocks and cheerleaders were worth cheering for.

You remember that scene in Lord of the Rings, where Aragorn and his soldiers are marching to the gates of Mordor? And…well…you know the quote:

> Sons of Gondor! Of Rohan! My brothers! I see in your eyes the same fear that would take the heart of me. A day may come when the courage of men fails, when we forsake our friends and break all bonds of fellowship. But it is not this day. An hour of wolves and shattered shields when the age of Men comes crashing down! But it is not this day! This day we fight! By all that you hold dear on this good Earth, I bid you stand! Men of the West!

That was *basically* a pep rally. And by god, we cheered. Because Aragorn was worth cheering for.

During my life, I have had two groups I really really feel comfortable saying were my in-groups. The Republic of Shireroth. And the rationalist community. When the Shirerithians played their national anthem, *I stood*. And when the rationalists had a religious ritual, I prayed.

Once I found my hive switch, so to speak, it's been easier to appreciate patriotism and religion. Team sports still involves a little too much sweat, and pep rallies don't mesh well with my auditory processing issues, but I can see *in principle* how someone might enjoy them.

And I guess you would ask why you would want to. But feeling like you're really connected to other people, not just in a "they share my goals and seem okay" way but in a "these are my people, we form a tribe or a community or, while on horseback, a horde" way is one of life's greatest pleasures and also a pretty important subgoal to anything that requires cooperation with other people. The only experience I can compare it to was being a kid and thinking I would never be dumb enough to waste time with crushes and romance, and then growing older and having the appropriate genetic payload unpack itself and tell me that this was a big part of what makes life worth living.

So my advice to anyone else who thinks they're a hundred percent chimp and totally bee-less is to find an in-group that really is their in-group, then try again.

SSC Gives A Graduation Speech

[Trigger warning for deliberately provoking horror about graduates' real-world post-college prospects]
[Epistemic status: intended as persuasive speech, may somewhat overstate case]

Ladies and gentlemen, I am honored to have been invited to speak here at the great University of [mumble]. Go Wildcats, Spartans, or Eagles, as the case may be!

I apologize if what I have to say to you sounds a little unpolished. I was called in on very short notice after your original choice for graduation speaker, Mr. Steven L. Carter, had his invitation to speak rescinded due to his offensive and quite honestly outrageous opinions. Let me say in no uncertain terms that I totally condemn him and everything he stands for, and that I am glad to see the University of [mumble] taking a strong stand against this sort of thing.

Ladies and gentlemen, probably the most famous graduation speech in history was Kurt Vonnegut's "Wear Sunscreen" address. I'm sure you've all heard about it. He told an MIT class that they should wear sunscreen. Because for all he knew any more substantial advice he gave might be wrong, but that at least was on a firm evidential basis.

Well, I come here before you to explain that there is now serious controversy in the dermatological community. A 1995 paper found that people who used more sunscreen had a much higher risk of malignant melanoma, the most dangerous type of skin cancer. Eight years later, a review article claimed that the original paper was confounded by fairness of skin, and that likely the relationship between sunscreen use and melanoma is zero. But the story was further complicated by the finding that sunscreen use may increase cancers of the internal organs, either through vitamin D dependent or some vitamin D independent pathways. My understanding is that a majority of dermatologists are still in favor of sunscreen, but that the issue is by no means settled.

But think about what the disagreement means. One of the smartest men in America came before an auditorium just like this, and said that there was only one item of advice of which he was completely certain – that you should wear sunscreen. Absolutely certain. And years later, we know that not only is this a very complicated question on which no certainty is yet possible – but it may very well be that if you follow his advice, you will get cancer and die.

Sometimes the things everybody knows everybody knows just aren’t true. Like, did you know Vonnegut never wrote a graduation speech about sunscreen at all?

So with this spirit of questioning assumptions in mind, I want to ask you a question. Today many of you will be completing your education. Sure, some of you are going on to graduate or professional training, but it is clearly the end of an era. Seventeen years, from kindergarten to the present, and I want to ask you:

Is education worth it?

This sounds like the introduction to every college graduation speech ever. The speaker will ask if education is worth it, say of course it is because something something the human condition, and everyone will cheer and head off to the reception. So in order to keep you on your toes, I want to make the opposite point. What if education, as you understand it – public or private or charter schooling from age four or five all the way to university as young adults – is, on net, a waste of your time and money?

In order to move beyond platitudes in evaluate whether education is worthwhile – to give it the same kind of fair hearing we would want to give sunscreen – we need to list out some of the costs and benefits. Of benefits, two stand out clearly. The philosophical benefits of feeling connected to the beauty of mathematics, the passion of the humanities, the great historical traditions. And the practical benefits of being able to get a job and afford nice things like food and shelter.

We will start with philosophy. Human knowledge is pretty great. Your life has been enriched with the ideas of brilliant thinkers, of giants upon whose shoulders you might one day hope to stand. Isn't this enough?

But as 86% of you know, you can't just observe an experimental group has experienced an effect and attribute it to the experimental intervention. You have to see if other people in a control group got the same benefit for less work.

What would be the control group for school? Home-schoolers do much better than those who attend public or private schools by nearly any measure. But this is unfair; it's what scientists call an "active control". What we really need to do is compare you to people who got no instruction at all.

It's illegal not to educate a child, so our control group will be hard to find. But perhaps the best bet will be the "unschooling" movement, a group of parents who think school is oppressive and damaging. They *tell* the government they're home-schooling their children but actually just let them do whatever they want. They may teach their kid something if the child wants to be taught, otherwise they will leave them pretty much alone.

And this is really hard to study, because they’re a highly self-selected group and there aren’t very many of them. The only study I could find on the movement only had n = 12, and although it tried as hard as it could to compare them to schoolchildren matched for race and family income level and parent education and all that good stuff I’m sure there’s some weirdness that slipped through the cracks. Still, it’s all we’ve got.

So, do these children do worse than their peers at public school?

Yes, they do.

By one grade level.

About college we still know very little. But if you’d stayed out of public school and stayed home and played games and maybe asked your parents some questions, then by the time your friends were graduating twelfth grade, you would have the equivalent of an eleventh-grade education.

Another intriguing clue here is Louis Benezet’s experiment with mathematics instruction. Benezet, an early 20th century superintendent of schools, wondered whether cramming mathematics into kids at an early age had a detrimental effect. He decreed that in some of the schools in his district, there would be *no* math instruction until grade six. He found that within a year, these sixth graders had caught up with their peers in traditional schools, and furthermore that they were able to think much more logically about math problems – figure out what was going on rather than desperately trying to multiply and divide all the numbers in the problem by one another. If Benezet’s results hold true – and on careful reading they are hard to doubt – any math education before grade six is useless *at best*. And it's hard to resist the urge to generalize to other subjects and children even older still.

Why is it so easy for the unschooled to keep up with their better educated brethren? My guess is that it's because very little learning goes on at school at all. The proponents of education speak of feeling connected to the beauty of mathematics, the passion of the humanities, and the great historical traditions. But how many of the children they spit out can prove one of Euclid's theorems? How many have been exposed to the Canterbury Tales? How many have experienced the sublime beauty of the Parthenon?

These aren't rhetorical questions, by the way. According to the general survey of knowledge among college students, 3.3% know who Euclid was, 7.6% know who wrote *Canterbury*, and a full 15% know what city the Parthenon's in.

36% of high school students know that an atom is bigger than an electron, rather than vice versa. But a full 59% of college students know the same. That's a whole nine percent better than chance. On one of the most basic facts about the fundamental entities that make up everything in existence.

"But knowledge isn't about names and dates!" No, but names and dates are the parts that are easy to measure, and it's a pretty good bet that if you don't know what city the Parthenon's in you probably haven't absorbed the full genius of the Greek architectural tradition. Anyone who's never heard of Chaucer probably doesn't have strong opinions on the classics of Middle English literature.

So in contradiction to the claim that education is necessary to teach beautiful and elegant knowledge, I maintain first that nearly nobody in the educational system picks this up anyway, that people who don't get any formal education at all pick it up nearly as much of it, and that people not exposed to it as children will, if they decide to learn it as adults, pick it up quickly and easily and without the heartbreak of trying to cram it into the underdeveloped head of a seven year old.

What about the claim that education is practically useful for getting a job and making money?

Even more than most young people, you've had the privilege of getting to watch your dreams implode in real time right before your eyes. About fifteen percent of you will be some variant of unemployed straight out of college. Another ten percent will find something part-time. And another forty or so percent will be underemployed, working as waiters or clerks or baristas or something else that uses zero percent of the knowledge you've worked so hard to accumulate. The remaining third of you who get something vaguely resembling the job you signed up for will still have to deal with wages that have stagnated over the last decade even as working hours increased and average student debt nearly doubled.

But don't worry, I'm sure the nice folks at Chase-Bear-Goldman-Sallie-Manhattan-Stearns-Sachs-Mae-FEDGOV will be happy to forgive your debt if you mention you weren't entirely happy with the purchase. You did hold out for the satisfaction-guaranteed offer, right? No? Uh oh.

As bad as the job market is, staying in school looks worse. Economists warn that attending law school is the worst career decision you can make, so much so that newly graduated lawyers have nothing do to but sue law schools for not warning them against attending and established firms offer an Anything But Law School Scholarship to raise awareness of the problem. Doctors are so uniformly unhappy that they are committing suicide in record numbers and nine out of ten would warn young people against going into medicine. Graduate school has always been an iffy bet, but now the ratio of Ph. D applicants to open tenure track positions has hit triple digits, with the vast majority ending up as miserable adjunct professors who juggle multiple part time jobs and end up making as much as a Starbucks barista but without the health insurance.

I'd like to thank whoever figured out how to include URLs in speeches, by the way. That was *the best* invention.

But here I cannot honestly disagree with the conventional assessment that going to school raises your earning power. As bad as you will have it, everyone who didn't graduate college still has it much, much worse. All the economic indicators agree with the signs from the desolate wasteland that was once our industrial heartland: they are doomed. Their wages are not stagnating but actively declining, their unemployment rate is a positively Greek thirty-five percent, and prospects for changing that are few and far between. Some economists blame globalization, which makes it easy to outsource manufacturing and other manual labor to the Chinese. Others blame technology, noting that many of the old well-paying blue-collar jobs are done not by foreigners but by machines. Both trends are set to increase, turning even more factory workers, truck drivers, and warehouse-stockers into burger-flippers, Wal-Mart greeters, and hollow-eyed unemployed.

But don’t let your schadenfreude get the better of you. Twenty years from now that’s going to be you. Sure, right now machines can only do the easy stuff, and the world isn’t interconnected enough to let foreigners do anything really *subtle* for us. But lawyers are already feeling the pinch of software that auto-generates contracts, and programmers are already feeling the pinch of Indians who will work for half the pay and email their code to Silicon Valley the next morning. You don’t need to invent a robo-drafter to put engineers out of business, just drafting software so effective it allows one engineer to do the work of three. And although there are half-hearted efforts to stop it, it seems more and more like King Canute trying to turn back a tide made of hundred dollar bills.

Once machines can do everything we can better and cheaper, the inevitable end result is employment for a few geniuses who invent and run the machines, immense profits for the capitalists who own the machines, and what happens to everyone else better left unspoken.

“Is this a vision of what shall be, or of what might be only?” Well, a visionaries as diverse as Martin Luther King, Richard Nixon and Milton Friedman have proposed something called a Basic Income Guarantee. When society becomes so advanced that it produces more than enough for everybody – but also so advanced that most individuals below genius level have little to contribute and no way of earning money – everyone should get a yearly salary just for existing. Think welfare, except that it goes to everybody, there’s no stigma, and it’s more than enough to live on. This titanic promise has run up against a giant iceberg with BUT HOW WOULD WE PAY FOR IT written in big red letters on the front. If we cancelled all existing welfare and entitlement programs – which makes sense if we’re giving everyone enough money to live comfortably on, we would only free up enough money together for a universal income of $5,800. I don't know if you can live on that, but I'd hate to have to try.

But we've gotten off track. We were counting the benefits of formal education. We did not do so well in trying to prove that it left you more knowledgeable, but it did seem like it had some practical value in getting you a little bit more money. With your shiny college degree, you can confidently assert "I've got mine", just as long as you take care not to notice the increasingly distant hordes of manual laborers or the statistics showing that the yours you've got is less and less every year.

What of the costs of education? What have you lost out on?

Well, first about twenty thousand hours of your youth. That's okay. You weren't using that golden time of perfect health and halcyon memories when you had more true capacity for creativity and imagination and happiness than you ever will again anyway. If you hadn't had your teachers to tell you that you needed to be making a collage showing your feelings about *The Scarlet Letter*, you probably would have wasted your childhood seeing a world in a grain of sand or Heaven in a wild flower or something dumb like that.

I'm more interested in the financial side of it. At $11,000 average per pupil spending per year times thirteen years plus various preschool and college subsidies, the government spends $155,000 on the kindergarten-through-college education of the average American.

Inspired by a tweet: what if the government had taken this figure (adjusted for inflation) and invested it in the stock market at the moment of your birth? Today when you graduate college, they remove it from the stock market, put it in a low-risk bond, put a certain percent of the interest from that bond into keeping up with inflation, and hand you the rest each year as a basic income guarantee. How much would you have?

And I calculate that the answer would be $15,000 a year, adjusted for interest. We can add the $5,800 basic income guarantee we could already afford onto that for about $20,000 a year, for everyone. Black, white, man, woman, employed, unemployed, abled, disabled, rich, poor. Welcome to the real world, it’s dangerous to go alone, take this. What, you thought we were going to throw you out to sink or swim in a world where if you die *you die in real life?* Come on, we’re not that cruel.

So when we ask whether your education is worth it, we have to compare what you got – an education that puts you one grade level above the uneducated and which has informed 3.3% of you who Euclid is – to what you could have gotten. 20,000 hours of your youth to play, study, learn to play the violin, whatever. And $20,000 a year, sweat-free.

$20,000 a year isn’t much. The average mid-career salary of an average college graduate is nearly triple that – $55,000. By the numbers your education looks pretty good. But numbers can be deceiving.

Consider the life you have to look forward to, making your $55,000. The exact profession that makes closest to that number is a paralegal, so let’s go with that. You get a job as a paralegal in a prestigious Manhattan law firm. You can’t afford to live in Manhattan, but you scrounge together enough money for a cramped apartment in Brooklyn, which costs you about $2000 a month rent. Every morning you wake up at 7:45, get on the forty-five minute subway ride to Manhattan, and make it to work by your 9:00 AM starting time. Your boss is a kind of nasty lawyer who is himself upset that he can’t pay back his law school debt and yells at you all day. By the time you get back home around 6, you’re too exhausted to do much besides watch some TV. You don’t really have time to meet guys – I’m assuming you’re a woman here, sixty percent of you are, I blame the patriarchy – so you put out a personal ad on Craigslist and after a while find someone you like. You get married after a year; your honeymoon is in Vermont because his company won’t give him enough time off to go any further.

You have two point four kids, and realize you’ve got to move to a better part of town because your school district sucks. Combined with your student debt, that puts a big strain on the finances and you don’t have enough to pay for child care. Eventually you find a place that will do it for cheap, and although it looks kind of dirty and you’re shocked when Junior calls you a “puta” which isn’t even a proper *English* curse word the price is right and they’re the only people who will accept four tenths of a kid. The older kids keep asking you and Dad for help with homework, which you can’t give because you haven’t really had time to keep up with your math and grammar and so on skills, what with the paralegal job and the television-watching taking up all your time. So you tell them to ask their teacher for extra help, which their teacher doesn’t give because she’s got forty other kids asking for the same thing and only twenty-four hours in a day. Despite all of this Junior gets into college and *you* sure haven’t saved up the money to put him through there tuition has spiraled to twelve gazillion dollars by this point and Chase-Bear-Goldman-Sallie-Manhattan-Stearns-Sachs-Mae-FEDGOV can’t lend him that because gazillion isn’t even a real number, and ohmigod what if Junior ends up one of those high school graduates with the Greek-level unemployment rates standing forlornly in front of a decaying factory in the Rust Belt? Worse, what if he ends up *living with you*? You beg him to go back to the bank and offer to pay whatever interest rates they ask. And so the cycle begins anew.

Or consider your life on a $20,000 a year income guarantee. No longer tied down to a job, you can live wherever you want. I love the mountains. Let’s live in a cabin in Colorado, way up in the Rockies. You can find stunningly beautiful ones for $500 a month – freed from the mad rush to get into scarce urban or suburban areas with good school districts, housing is actually really cheap. So there you are in the Rockies, maybe with a used car to take you to Denver when you want to see people or go to a show, but otherwise all on your own except for the deer and squirrels. You wake up at nine, cook yourself a healthy breakfast, then take a long jog out in the forest. By the time you come back, you’ve got a lot of interesting thoughts, and you talk about them with the dozens of online friends you cultivate close relationships with and whom you can take a road trip and visit any time you feel like. Eventually you’re talked out, and you curl up with a good book – this week you’re trying to make it through Aristotle on aesthetics. The topic interests you since you’re learning to paint – you’ve always wanted to be an artist, and with all the time in the world and stunning views to inspire you, you’re making good progress. Freed from the need to appeal to customers or critics, you are able to develop your own original style, and you take heart in the words of the old Kipling poem:

> And none but the Master will praise them
> And none but the Master will blame
> And no one will work for money
> And no one will work for fame
> But each for the joy of the working
> Each on his separate star
> To draw the thing as he sees it
> For the God of things as they are

One of the fans of your work is a cute girl – this time I'm assuming you're a man, I'm sure over the past four years you've learned some choice words for people who do that. You date and get married. She comes to live with you – she's also getting $20,000 a year from the government in place of an education, so now you're up to $40,000, which is actually very close to the US median household income. You have two point four kids. With both of you at home full time, you see their first steps, hear their first words, get to see them as they begin to develop their own personalities. They start seeming a little lonely for other kids their own age, so with a sad good-bye to your mountain, you move to a bigger house in a little town on the shores of a lake in Montana. There's no schooling for them, but you teach them to read, first out of children's books, later out of something a little harder like Harry Potter, and then finally you turn them loose in your library. Your oldest devours your collection of Aristotle and tells you she wants to be a philosopher when she grows up. Evenings they go swimming, or play stickball with the other kids in town.

When they reach college age, your daughter is so thrilled at the opportunity to learn from her intellectual heroes that she goes to Chase-Bear-Goldman-Sallie-Manhattan-Stearns-Sachs-Mae-FEDGOV and asks for a loan. They're happy to give her fifteen thousand, which is all college costs nowadays – only the people who are really interested in learning feel the need to go nowadays, and supply so outpaces demand that prices are driven down. She makes it into Yale (unsurprising given how much better home-schooled students do) studies philosophy, but finds she likes technology better. She decides to become an engineer, and becomes part of the base of wealthy professionals helping fund the income guarantee for everyone else. She marries a nice man *after* making sure he's willing to stay home and take care of the children – she's not crazy, she doesn't want to send them to some kind of *institution*

Your younger son, on the other hand, is a little intellectually disabled and can't read above a third-grade level. That's not a big problem for you or for him. When he grows older, he moves to Hawaii where he spends most of his time swimming in the ocean and by all accounts enjoys himself very much.

You're happy your son will be financially secure for the rest of his life, but on a broader scale, you're happy that no one around you has to live in fear of getting fired, or is struggling to make ends meet, or is stuck in the Rust Belt with a useless skill set. Every so often, you call your daughter and thank her for helping design the robots that do most of the hard work.

Would you like to swing on a star? Carry moonbeams home in a jar? And be better off than you are? Or would you like to get a formal education?

We're finally getting back to the point now. I'm sorry it's taken this long. I can see the Dean of Students checking her watch over there with a worried look on her face. I think she's worried I'm trying to filibuster your graduation. You know legally if I can keep speaking until midnight tonight, the graduation is cancelled and you have to stay in school another year? It's true. Those are the rules.

Because I don't want to talk about the very broad social question of whether Education the concept is worth it to Society as a concept. I want to ask *you*, standing here today, was *your* education worth it?

Because this is a college graduation speech, and I am legally mandated to offer some advice, and the specific advice I give will be tailored to your response.

Some of you will say yes, my education was worth it. I am the 3.3%! I know who Euclid was and I understand the sublime beauty of geometry. I don’t think I would have been exposed to it, or had the grit to keep studying it, if I hadn’t been here surrounded by equally curious peers, under the instruction of enthusiastic professors. This revelation was worth losing my cabin in Colorado, worth resigning myself to the daily grind and the constant lurking fear of failure. I claim it all.

And to you my advice is: if you’ve sacrificed everything for knowledge, don’t forget that. When you are a paralegal in Brooklyn, and you get home from work, and you are very tired, and you want to curl up in front of the TV and watch reality shows until you are numb, remind yourself that you value knowledge above everything else, that you will seek intellectual beauty though the world perish, and read a book or something. Or take a class at a community college. Anything other than declaring knowledge your supreme value but becoming a boob.

Others of you will say yes, my education was worth it. Not because of what I learned about ukulele or eucalyptus or whatever, but because of the friends I made here, the proud University of [mumble] spirit of camaraderie, which I will carry forth my entire life.

And to you my advice is similar: if you’ve sacrificed everything for friendship, don’t forget that. When you are a paralegal in Brooklyn, or a market analyst in Seattle, or God forbid an intern in Michigan, and you get home from work, and you are very tired, and you want to curl up in front of your computer and check Reddit, remind yourself of the friends you made here and give them a call. See how they’re doing. Write them a Christmas card, especially if it is December. Anything other than declaring friendship your supreme value and drifting out of touch.

Others of you will say yes, my education was worth it. Not because of what I learned about the Eucharist or eucre or whatever, but because of the connections I made, the network of alumni who will be giving me a leg up in whatever I choose to pursue.

And to you my advice is, again, similar. If you’ve sacrificed everything for ambition, be ambitious as *hell*. When you are a paralegal in Brooklyn or whatever, claw your way to the top, stay there, and use it to do something important. If you’ve sacrificed everything for ambition, don’t you dare stop at middle manager.

Others of you will say yes, my education was worth it. Not because of what I learned about yucca or the Yucatan or whatever, but because it helped me learn civic values, become a better person who is better able to help others.

And to you my advice is once again similar. If you’ve sacrificed everything to help others, don’t let it all end with donating a tenner to the OXFAM guy on the street now and then. Join Giving What We Can or go volunteer somewhere. If you’ve sacrificed everything for others, make sure others get something good out of the deal!

Others of you will say yes, my education was worth it. Not because of what I learned about eukaryotes or Ukraine or whatever, but because formal education in the school system *taught me how to think*.

And to…sorry, one second, HAHAHAHAHAHHAHAAAHAHAHAHHHAAHAHA HAHAHAHAHAHHA HAHAHAHAHAHAH

HAHAHAHHHHHAAAHAHAHAHAH HAHAHHA HAHAHHHHAHAH HAAHHHAHA HAAHAHAHAHHA HAHHAHAHAHAHAHHAHA AHHHHAHAHAHAHA HAHAHAHAHA HHAAAHAHHAAHHAHA AHHAHAHAHAHA hahaha haha ha hahaha haha heh heh heh okay.

I’m sorry. Ahem. To you my advice is, again, similar. If you’ve sacrificed everything to learn how to think, learn how to think. When someone says something you disagree with, before you dismiss a straw man it and call that person names and slap yourself five for your brilliant rebuttal, take a second to consider it fairly on its own terms. Go learn about biases and heuristics and how to avoid them. Read enough psychology and cognitive science to figure out why your claim might *kind of* inspire hysterical laughter from people even a little familiar with the field. Just don’t sacrifice everything to learn how to think and end up only rearranging your prejudices.

And finally, some of you will say, wait a second, maybe my education *wasn’t* worth it. Or, maybe it was the best choice to make from within a bad paradigm, but I’m not content with that. And I wish someone had told me about all of this more than fifteen minutes before I graduate.

And to you I can offer a small amount of compensation. You have learned a very valuable lesson that you might not have been able to learn any other way.

You have learned that the system is Not Your Friend.

I use those last three words very consciously. People usually say “not your friend” as an understatement, a way of saying something is actively hostile. I don’t mean that.

The system is not your friend. The system is not your enemy. The system is a retarded giant throwing wads of

$100 bills and books of rules in random directions while shouting “LOOK AT ME! I’M HELPING! I’M HELPING!” Sometimes by luck you catch a wad of cash, and you think the system loves you. Other times by misfortune you get hit in the gut with a rulebook, and you think the system hates you. But either one is giving the system too much credit.

Every one of the architects and leaders of the system is fantastically intelligent – some even have degrees from the University of [mumble]. But every one of the neurons in my dog’s brain is a fantastically complex pinnacle of three billion years of evolution, yet my dog herself can spend the better part of an hour standing motionless, hackles raised, barking at a plastic bag.

To you I don’t have very much advice. I’m no smarter than anyone else – well, I know who Euclid is, but *other* than that – and if I knew how to fix the system, it’s a pretty good bet other people would know too and the system would already have been fixed. Maybe you, armed with a degree from the University of [mumble], will be the one to help figure it out.

On the other hand, someone a lot smarter than I am *did* have some advice for you. Poor Kurt Vonnegut never did get to give a real graduation speech, but one of his books has some advice targeted at another major life transition:

> Hello babies. Welcome to Earth. It’s hot in the summer and cold in the winter. It’s round and wet and crowded. On the outside, babies, you’ve got a hundred years here. There’s only one rule that I know of, babies-“God damn it, you’ve got to be kind.”

I don’t know how to fix the system, but I am pretty sure that one of the ingredients is kindness.

I think of kindness not only as the moral virtue of volunteering at a soup kitchen or even of living your life to help as many other people as possible, but also as an epistemic virtue. Epistemic kindness is kind of like humility. Kindness to ideas you disagree with. Kindness to positions you want to dismiss as crazy and dismiss with insults and mockery. Kindness that breaks you out of your own arrogance, makes you realize the truth is more important than your own glorification, especially when there's a lot at stake.

Here we are at the end of a grinder of $150,000, 20,000 hours, however many dozen collages about *The Scarlet Letter*, and the occasional locker room cry of "faggot" followed by a punch in the gut. Somewhere in another world, there are people just like us in nice cabins reading Aristotle and knowing that nobody will have to go hungry ever again. The difference between us and them isn't money, because I think the $155,000 the government gave you could have gone either way – and even if I'm wrong about that there's more than enough money somewhere else. The difference isn't intelligence, because the architects of our system are fantastically bright in their own way. I think kindness might be that difference.

Technically kindness plus coordination power, but that's another speech, and the Dean of Students is starting to make frantic hand signals.

I don't know if it's really possible to afford to give everyone that cabin in Colorado. But I hope that the people whose job it is to figure that out approach the problem with a spirit of kindness and humility.

In conclusion, both sides of the sunscreen debate have some pretty good points. It will certainly decrease your risk of squamous and basal cell carcinomas, it probably has no effect on the malignant melanoma rate but there’s a nonzero chance it might either cause *or* prevent them, and its effect on internal tumors seems worrying at this point but is yet to be backed up by any really firm evidence.

I understand this is complicated and unsatisfying. Welcome to the real world.

[Congratulations to my girlfriend Ozy, who graduates college this week!]

Nydwracu’s Fnords

I.

The fnords first appear in Anton-Wilson and Shea’s book *Illuminatus*. Educators, operating as tools of the titular conspiracy, hypnotize all primary school children to have a panic reaction to the trigger word “fnord”. The children, who remember nothing of the sessions when they wake up, are incapable of registering the word except as an unexplained feeling of unease.

This turns them into helpless, easily herded adults. Every organ of the media – newspapers, books, cable TV – contains a greater or lesser number of fnords. When some information is counter to the aims of the conspiracy – maybe a communist party organizing in a state where the conspiracy wears a capitalist hat – the secret masters don’t bother censoring or suppressing it. Instead, the newspaper reports it on the front page, but fills the article with fnords. Most people read partway through, become very uncomfortable and upset without knowing why, and decide that communists are *definitely* bad people for some reason or other and there’s no reason they need to continue reading the article. Why should they worry about awful things like that when there’s the whole rest of the paper to read?

According to the book, the only section of the newspaper without any fnords at all is the advertisements.

II.

Last week, some Internet magazine published the latest attempt at the genre of Did You Know Neoreaction Exists You

Should Be Outraged. A couple of reactionaries wrote the usual boring “actually, nothing you said was true, why would you say false things?” responses. Nydwracu, a frequent commenter on this blog, did something I thought was *much* more interesting. He wrote a post called Fnords where he removed all of the filler words and transitions between ideas and thin veneer of argument until he stripped the essay down to the bare essentials. It looked like this:

> Mouthbreathing Machiavellis Dream Of A Silicon Reich strange and ultimately doomed stunt flamboyant act of corporate kiss-assery latest political fashion California Confederacy total corporate despotism potent bitter Steve Jobs Ayn Rand Ray Kurzweil prominent divisive fixture hard-right seditionist aggressively dogmatic blogger reverent following in certain tech circles prolific incomprehensible vanguard youngish white males embittered by “political correctness” Blade Runner, but without all those Asian people cluttering up the streets like to see themselves as the heroes of another sci-fi movie “redpilled” The Matrix “genius” a troll who belches from the depths of an Internet rabbit hole frustrated poet cranky letters to alternative weekly newspapers preoccupations with domineering strongmen angry pseudonym J.R.R. Tolkien George Lucas typical keyboard kook archaic, grandiose snippets cherry-picked from obscure old lack of higher ed creds overconfident autodidact’s imitation fascist teenage Dungeon Master most toxic arguments snugly wrapped in purple prose and coded language oppressive nexus teeth-gnashing white supremacists who haunt the web “men’s rights” advocates nuts disillusioned typical smarmy, meandering (Sure. Easy!) Incredible as it sounds, absolute dictatorship may be the least objectionable tenet espoused by the Dark

> Enlightenment neoreactionaries. Chinese eugenics impending global reign of “autistic nerds These imaginary übermensch sprawling network of blogs, sub-Reddits old-timey tyrants basically racism scientific-sounding euphemism familiar tropes of white victimhood perhaps best known for his infamous slavery apologia poor, persecuted Senator Joe McCarthy. Big surprise. pseudo-intellectual equivalent of a Gwar concert, one sick stunt after another, calculated to shock the attention he so transparently craves “silly not scary” “all of these people need to relax: P.G. Wodehouse football get drunk Internet curio “sophisticated neo-fascism” must be confronted “creepy” future-fascist dictator sadly Koch brothers no matter how crazy your ideas are, radicalism neoreactionaries flatter the prejudices of the new Silicon Valley elite enemies patchwork map of feudal Europe Forget universal rights; signposts of the neoreactionary fantasyland anti-democratic authoritarianism bigotry blue-sea libertarian dream extreme libertarian advocacy Ted Cruz libertarian a small and shallow world a dictatorial approach mythical “god-kings” Stupid proles! They don’t deserve our brilliance! shockingly common would never occur to other people precisely because they’ve refused to leave that stage of youthful live forever escape to outer space or an oceanic city-state play chess against a robot that can discuss Tolkien fantasies childhood imagination perhaps too generous the fundamental problem with these mouthbreathers’ dreams of monarchy. They’ve never role-played the part of the peasant.

That…sure gives one a different perspective on political discourse. I am reminded of those Renaissance artists who secretly cut up cadavers to learn what was inside people, and from then on all of their human figures would be a little bit creepy because you could almost *see* how the internal bones and muscles were animating the flesh.

Since no one is meta and everyone only pays attention to things when it's their own opinions under threat, I suppose I have to do the same thing with an article from some website on the right:

> socialism completely government run pure single-payer "an island of socialism in American healthcare" that won't change a thing in fact it's a distraction excessive delays tragically predictable bureaucratic rationing price controls, inefficiencies, and the inevitable cover-ups bureaucratic incentives statist VA healthcare system mirrors the government-run healthcare problems slip-shod failure run-amok bureaucrats don't tell me the problem is not enough government money the Paul Krugmans of the world and their leftist allies socialist medicine socialism doesn't work who opposed market choice and competition Senator Harry Reid and House Democratic leader Nancy Pelosi Obamacare job-destroying tax and regulatory provisions

Interestingly, both of those came out to between 13 and 14% of the length of the original article. I wonder if that's some kind of iron law.

III.

I don't know if he ever read *Illuminatus* or whether it was just one of those coincidences, but Jonathan Haidt did the thing with the fnords in real life.

(Warning: a tangentially related study by the same group has recently failed to replicate)

He wanted to test the role of disgust in moral judgments. So he hypnotized a bunch of people to feel disgust at a trigger word – "takes" for half the participants, "often" for the other half – and hypnotically instructed them to forget all about this. Then in an "unrelated study" he asked them to rate the morality of different ethically controversial vignettes. For example:

"A brother and sister fall in love with each other. They frequently **take** vacations together where they have sex. Both are freely consenting and she is on very careful birth control."

or

"A brother and sister fall in love with each other. They **often** go on vacations together where they have sex. Both are freely consenting and she is on very careful birth control."

The participants hypnotized to hate the word "take" found the behavior more objectionable with the "take" version of the vignette than the "often" version, and the participants hypnotized to hate the word "often" displayed the opposite pattern. When they asked subjects to explain their judgment, they gave perfectly reasonable explanations, which could be anything from "incest is just wrong" to "what if they have a child and it's deformed, yeah, I know it said they were on birth control, but it still bothers me."

Then Haidt and his team presented the following story:

"Dan is student council president. It is his job to pick topics for discussion at student meetings. He frequently **takes** suggestions from students and teachers on which topic to choose."

or

“Dan is student council president. It is his job to pick topics for discussion at student meetings. He **often** accepts suggestions from students and teachers on which topic to choose.”

Participants were asked to judge how evil a person Dan was. And when their trigger word was in the sentence, their answer was: pretty evil! When asked to explain themselves, they came up with weird justifications like “Dan is a popularity-seeking snob” or “It just seems he’s up to something”

IV.

A few weeks ago, I noticed something strange.

Every time someone complaints about climate denial, they make extraordinary efforts to get the name of the Koch brothers in. Like it’s never just “Why do so many people believe climate denialism?” it’s more “Why do so many people believe climate denialism, as funded by people like the Koch brothers?”

This is strange because it seems to me that they are acting like associating climate denialism with the Koch brothers will lower its credibility or make it sound vaguely evil.

But this shouldn’t work. The only thing the average person knows about the Koch brothers is that they are people who fund climate change denial. So if you already don’t like climate change denial, this will make you dislike the Koch brothers. But mentioning “Koch brothers!” won’t make you dislike climate change denial more, it will just remind you of one of the downstream effects of your disliking climate change (not liking the Kochs). On the other hand, if you’re still neutral on climate change denial, then you have no reason to dislike the Kochs, and mentioning them won’t help you there either. And if you actively support climate denial, you probably think the Koch brothers are heroes, so associating them with the movement won't be a good way of discrediting it.

Basically, since your opinion of the Koch brothers should equal your opinion of climate denial, trying to tar climate denial by association with the Kochs is trying to make people dislike an idea by linking it to itself. It shouldn't work.

But I think it does. When you read articles on the other side, they always mention Al Gore. In fact, there are a lot of these people who get brought up as bogeymen every so often.

I have two boring hypotheses and an interesting one.

The first boring hypothesis is that the Koch brothers are white male billionaires. This is enough to make them suspicious. Therefore, global warming skepticism is tarred by association with them, even though we know nothing else about them.

The second boring hypothesis is that it doesn't matter who the Koch brothers are, what matters is the claim that there is some figure funding the movement, that it's not a grassroots upswelling of people genuinely doubtful of global warming, but just one guy (well, two guys) trying to inflict their own weird contrarianism on everyone else.

The interesting hypothesis is that the brain is going loopy, having one of those rare experiences where it forgets not to condition on itself.

Imagine that you don't like climate denialism. You hear that the Koch brothers support climate denialism. You use that information to decide you don't like the Koch brothers very much.

Then a month passes and you forget *exactly* why you don't like the Koch brothers. You just have a very strong feeling that "it just seems like they're up to something."

Then someone tells you the Koch brothers support denialism. And you say: "If *those* bastards support it, then I hate it *even more*!"

In other words, you have undergone a two step process to ratchet up your dislike of climate denialism by associating it with itself.

> We know this idea is evil because it's pushed by such terrible people. We know the people are terrible because they push such an evil idea.
>
> — Scott Alexander (@slatestarcodex) May 18, 2014

I wonder if this is part of what makes politics so divisive. You start off with a weak preference in one direction. Gradually, certain words like "Koch brothers" or "Exxon Mobil" become fnords, reservoirs of your negative feelings, and then every time you read about climate change, even if there's no real argument, you get triggered and become pretty sure denialists are up to something, in the same way Dan the student council president is up to something. And the other side gets different fnords – "Climategate", "hockey stick graph", and they go through the same process. And finally you get totally incomprehensible arguments: "But how can you be a climate change denier when that associates you with the Koch brothers?! Did you know climate change denialism is *literally* sponsored by the Heartland Institute?!" And the other side is just nodding their head and going "Oh, yeah, my sister used to work there."

V.

IF YOU DON’T SEE THE FNORD IT CAN’T EAT YOU

Apologia Pro Vita Sua

I do occasional work for my hospital's Addiction Medicine service, and a lot of our conversations go the same way.

My attending tells a patient trying to quit that she must take a certain pill that will decrease her drug cravings. He says it is mostly covered by insurance, but that there will be a copay of about one hundred dollars a week.

The patient freaks out. "A hundred dollars a week? There's no way I can get that much money!"

My attending asks the patient how much she spends on heroin.

The patient gives a number like thirty or forty dollars a day, every day.

My attending notes that this comes out to $210 to $280 dollars a week, and suggests that she quit heroin, take the anti-addiction pill, and make a "profit" of $110.

At this point the patient always shoots my attending an incredibly dirty look. Like he's cheating somehow. Just because she has $210 a week to spend on heroin doesn't mean that after getting rid of that she'd have $210 to spend on *medication*. Sure, these fancy doctors think they're so smart, what with their "mathematics" and their "subtracting numbers from other numbers", but they're not going to fool *her*.

At this point I accept this as a fact of life. Whatever my patients do to get money for drugs – and I don't want to know – it's not something they can do to get money to pay for medication, or rehab programs, or whatever else. I don't even think it’s consciously about them caring less about medication than about drugs, I think that they would be literally unable to summon the motivation necessary to get that kind of cash if it were for anything less desperate than feeding an addiction.

There’s a rationalist saying about making a desperate effort, as if the life of your child was at stake. Nowadays I tend to think of this in terms of “Make a desperate effort, as if you were a heroin addict and your next fix depended on it.”

Anyway, it might sound like I’m mocking my patients, but I’m not. I’m meandering my way into an apology.

I have had a really busy few months. I think it will be letting up soon, but I’m not sure. And I’ve told a lot of people who needed things from me, for one reason or another, “I’m sorry, I’m too busy to take care of this right now.”

And I worry that some of those people read my blog and think “Wait, if you have enough time to write blog posts nearly every day, some of which are up to six thousand words long, why don’t you have enough time to do a couple of hours work for me?”

And the answer is – you fancy doctors with your mathematics and subtraction might *say* that I could just take a couple of hours away from blogging and use those free hours to write that one thing or analyze that one study or whatever, but you’re not going to fool *me*.

Just as drugs mysteriously find their own non-fungible money, enjoyable activities mysteriously find their own non-fungible time. If I had to explain it, I’d say the resource bottleneck isn’t time but energy/willpower, and that these *look* similar because working hard saps energy/willpower and relaxing for a while restores it, so when I have less time

I also have less energy/willpower. But some things don’t require energy/willpower and so are essentially free.

This is unfair to you guys, but I should be substantially freer in the next couple of weeks and we can see what happens.

Compound Interest Is The Least Powerful Force In The Universe

I.

I’m still iffy on Vox. Some of its reporting is excellent – their article on Governor Cuomo and the shift away from progressivism in the Democratic Party was especially enlightening. Other parts, especially the editorials, are atrocious and utterly without subtlety.

Ezra Klein’s article on compound interest was the latter.

> The reaction to Ta-Nehisi Coates’ magisterial essay on the lingering effects of American racism is polarized around people’s reaction to the word “reparations.” But much of the story he tells is about something simpler, and completely uncontroversial: the power of compound interest.
>
> You might remember, as a kid, getting this problem on a test: Would you rather have $10,000 per day for 30 days or a penny that doubled in value every day for 30 days?
>
> The answer, of course, is you want the penny that doubles in value every day. If you take the $10,000 you end up with $300,000 after the first month. Take the penny and you end with about $5 million.
>
> What Coates shows is that white America has, for hundreds of years, used deadly force, racist laws, biased courts and housing segregation to wrest the power of compound interest for itself. The word he keeps coming back to is “plunder.” White America built its wealth by stealing the work of African-Americans and then, when

that became illegal, it added to its wealth by plundering from the work and young assets of African-Americans. And then, crucially, it let compound interest work its magic.

Today, white America is one of the richest and most powerful populations the world has ever known. And it wonders why African Americans just can’t seem to keep up. “In America,” Coates writes, “there is a strange and powerful belief that if you stab a black person 10 times, the bleeding stops and the healing begins the moment the assailant drops the knife.”

[…]

“The popular mocking of reparations as a harebrained scheme authored by wild-eyed lefties and intellectually unserious black nationalists is fear masquerading as laughter,” Coates writes. It’s also the intellectually unserious response of people who believe that because they never owned slaves or drank from a whites-only water fountain they weren’t the beneficiaries of American racism. They may not be the villains of American racism, but they are the beneficiaries of it. The average white southerner in 1832 was far poorer than the average white southerner today, and part of that vast increase in wealth and income and knowledge and social networks is the result of compound interest working its magic on what the slaveowners and the segregationists stole.

It’s as simple and clear as a child’s math problem. The people who benefitted most from American racism weren’t the white men who stole the penny. It’s the people who held onto the penny while it doubled and doubled and doubled and doubled.

There are many many complicated moral arguments for and against reparations. Like Klein, I don't want to get into any of them except the financial aspect of how much modern whites benefit from the lingering effects of slavery, and how much modern blacks are harmed by them.

I want to make one very loose argument and then one based off of empirical research.

The loose argument is that the best way to determine whether modern whites have gained from owning slaves (and I know Klein's argument takes into account other forms of oppression beyond slavery, but slaves will be a good first approximation) is to see if formerly slave-owning societies are richer than formerly non-slave-owning societies.

The state with the highest percent slaves before the Civil War was South Carolina, with Mississippi number two. Mississippi is the poorest, and South Carolina the fifth poorest of the fifty states today. Except for Virginia, every single state in the former Confederacy is poorer than the US average.

This is somewhat confounded by the high level of poor blacks in these states, but remains true even when you look only at the income of white residents. For example, if Mississippi whites were their own state, they would be 39th out of 50 in terms of per capita income. South Carolingians would do better but still be below the national average. If *all* states suddenly became all white, Mississippi and South Carolina would drop right back down to the bottom.

So the whites who had the most opportunity to benefit from a supposed ability to earn compound interest on slavery earnings clearly didn't do that.

While one could make the argument that the gains from slavery left Mississippi and the Deep South to enrich *all* whites, this seems a bit forced. The US was much less interconnected in those days. And other places that had no connection to slavery still outperform the Deep South: Italian whites, for example, still do comfortably better than whites from most Southern states.

One could always argue that Southerners would be *even poorer* today if not from all the compound interest they received on their slavery earnings. But Southern poverty is already a bit of a puzzle. To make them too much poorer would require them to descend into levels of squalor totally unknown in any First World country.

I think we should at least look at an alternate hypothesis: people are really really *really* bad at passing ill-gotten wealth through more than a generation or two.

II.

The descendants of rich people tend to stay rich even three hundred years later. For example, Gregory Clark looked at social mobility in Sweden. A famously mobile society, Sweden is also a good place to study social mobility since nobles and commoners had different last names back when the feudal system was in place around 1700. Non-nobles are forbidden to change to noble-sounding surnames even today, so names should be a fossil record of who’s descended from the really rich people.

Clark found that among highly-educated well-paying professions like doctors and lawyers, people with aristocratic surnames are represented around four to six times the level expected by chance. He uses this to describe a statistic “b” signifying the rate of regression to the mean with each generation.

Suppose nobles are eight times more likely to be doctors at the end of the feudal period. If b is low, say 0.35, then they will quickly regress to the population’s average – according to Clark’s graph, it will take about 80 – 90 years (= 3 – 4 generations?) before this happens. If b is high, say 0.75, it will take practically forever – he demonstrates that even after 200 years, there will be noticeable differences.

Clark finds b to be between 0.6 and 0.8 in Sweden, and then goes on to show it is similar pretty much everywhere and across all time periods. The Economist describes his research by saying:

> With surprising consistency across countries and eras, mobility is found to be painfully slow. Birth has predicted more than 50% of one’s income or education status, Mr Clark reckons. Erasing the legacy of past prosperity takes 10-15 generations rather than the three or four implied by sunnier estimates. So the shadow of 18th-century wealth still darkens income distributions today.

This sounds very promising for Ezra Klein’s compound interest argument and really bad for my “Southerners are really bad at holding on to wealth” argument. But Clark takes it in a totally different direction. The Economist again:

> The most unexpected finding [is that] efforts to democratise education and eliminate discrimination over the past century appear to have had no discernible effect on mobility, leading Mr Clark to conclude that mobility is strongly linked to underlying social competence—an “inescapable inherited” trait. Only the intermarriage of people who are more prosperous and

> educated with those less fortunate will dilute the genetic resources of well-off families, slowly pushing them back towards the average and preventing the rise of a permanent overclass.

I just want to briefly pause our economics discussion to point out that Professor Clark has written two books on his theories, and they they are called A Farewell to Alms andThe Son Also Rises. Please take a moment to be delighted by that.

III.

Okay, so wealth lasts a really long time, and Klein thinks that’s because of compound interest and Clark thinks it’s because of “inescapably inherited social competence” which sounds a lot like a euphemism for genes. To differentiate between these two hypotheses we would need to randomly select a bunch of people, give them a lot of wealth, and follow them for a couple of generations to see whether their descendants compounded that advantage or regressed back to their genetically programmed level.

I am familiar with only one well-studied example of this happening.

In 1830 the government stole the land of the Cherokee Indians. Lots of white people wanted to settle the newly available territory, so the state of Georgia proposed a lottery, where the winners would get large fertile farms on the conquered area. Not only were the farms available in the lottery much bigger than those owned by the average Georgia farmer at the time, but lottery winners were totally allowed to sell the farm they had just won to someone else and pocket the cash. Nearly every single white person in

Georgia at the time entered the lottery, because hey, free money.

Bleakley and Ferrie track the winners and losers of the land lottery for several generations. They find that the first-generation winners did very well. The average farm won in the lottery was worth $900 in 1830 dollars, which was the equivalent of three years' unskilled labor (so think about $60000 today). It was also a gift that kept on giving, since most people would farm the land and be able to grow lucrative crops every year indefinitely. According to the study "two decades after the lottery, winners are on average $700 richer than a comparable population that did not win the lottery". Remember, this is back when $700 was *real money*.

In a second study, they look at the effect a couple of generations down the line. They find:

> Sons of winners have no better adult outcomes (wealth, income, literacy) than the sons of non-winners, and winners' grandchildren do not have higher literacy or school attendance than non-winners' grandchildren. This suggests only a limited role for family financial resources in the formation of human capital in the next generations in this environment and a potentially more important role for other factors that persist through family lines […]
>
> [This] should have relaxed the budget constraint faced by poorer households and allowed them to invest more in the human capital of their children. If human capital was unaffected in the next generations, this is evidence in favor of the view recently advanced by Clark and Cummins that a substantial portion of the intergenerational correlation in outcomes is driven by

> fundamental, family-specific effects (the family’s cultural and genetic infrastructure)

Studies of modern-day lottery winners show much the same, albeit on a much-reduced time scale. And closer to the original issue, studies attempting to compare enslaved blacks versus free blacks appear to show it didn’t actually take that many generations for outcomes to equalize.

IV.

I am disappointed that all we have to go off of are these kinds of hints and whispers.

This seems like maybe *the* most important question in economics. Certainly in sociology. It seems terribly important to public policy, not just in terms of reparations but in how much we spend helping poor families and more important *how* we help poor families. If Clark’s view is right, the best we can do is alleviate their suffering by making sure that being poor isn’t an especially unpleasant state and everyone has good access to social services. If Klein is right, we should be making huge cash transfers to get people out of poverty traps so that their descendents will reap the compound interest and become rich.

(there’s also the slight confounding factor of the South being demolished during the Civil War, which could disprove Klein’s particular example while not necessarily supporting Clark or the general case)

I want to repeat Clark’s quote to the Economist for emphasis:

> The most unexpected finding [is that] efforts to democratise education and eliminate discrimination

> over the past century appear to have had no discernible effect on mobility.

If Clark is right, almost everything we're doing is a waste of time.

There are some good arguments about why poverty might make people less successful over long time periods, like cognitive load effects. But there are also some good counterarguments – if that's true, how come the second-generation descendents of Vietnamese boat people, who came to America with nothing, now have notably higher household incomes than white Mississippians, with all their years of benefitting off other people's slave labor?

Overall, these are the sort of really complicated problems I would have expected a supposedly more sophisticated media outlet like Vox to cover or help raise awareness on.

Instead it uses mere assertion to tell use that denial of compound interest is an "intellectually unserious response" and that Klein's case is "as simple and clear as a child's math problem."

Sorry, Vox. I'll keep reading you for your occasional article about the case for raising chickens in virtual reality. But intelligent and sophisticated you are not.

EDIT: Tyler Cowen gives a different perspective on the same Klein article over at Marginal Revolution

EDIT 2: The quote I stole the title from is probably not legit

EDIT 3: One could try to reconcile Klein and Clark by saying that sure, wealth doesn't persist in families across generations but it does in societies (presumably it gets transferred from family to family within the society but continues to exist). But then there would be no reason to favor white people and their descendants as especial beneficiaries.

On Types of Typologies

> My MBTI type is “the type of person who did some looking into it years ago and knows that the MBTI is neither particularly scientific nor particularly consistently applied”. Or, as it’s also called, INTJ
> — tropylium.tumblr.com

I’m sick of people hating on the Myers-Briggs Type Indicator.

The argument against Myers-Briggs is that it’s not scientific. The argument *for* Myers-Briggs is that I’m *also* the kind of person who did some looking into it and realizes that MBTI is neither scientific nor consistently applied, and I *also* test consistently as INTJ, so clearly something is going on here. And every time I read a description of INTJ I have to facepalm because I so consistently recognize myself in it.

(Yes, I’m familiar with the Forer effect and have compared it to descriptions of different types. Yes, I could totally believe there is a Forer + placebo effect where knowing that you have been assigned a certain type makes it sound more relevant to you than other types you read. Yes, I’m still impressed with how well descriptions of INTJs fit me. Also, I notice that people on Less Wrong, ie people like me, are seven times more likely to be INTJ than the general population. That seems like a nice objective result.)

I think it’s easy to reconcile “Myers-Briggs is not scientific” with “Myers-Briggs is a useful and real descriptive tool”.

When we say a personality type theory is scientific, part of what we want is certainty that it describes the world in the correct number of categories in the correct way. For example, MBTI has four different variables, each of which has two possible values. This could be interpreted as an implicit claim that there are four dimensions to personality, and only two types of personality within each dimension. And Myers and Briggs had no basis on which to make such a claim and just sort of armchaired it. Interpreted that way, the theory is pseudoscientific and dumb.

We compare it to the more scientifically accepted five-factor model, which used lots of data and a statistical technique called factor analysis to clump out what dimensions people actually varied on. They got five, only one of which (extraversion) is obviously the same as any of Myers-Briggs' dimensions (although I've read other sources saying MBTI intuition is very highly correlated with FF openness). Now it's not clear the five-factor people got it right; there's a small cottage industry in finding correlations among the five factors and reducing them to fewer factors, such as two factor or one factor models. But clearly the five factor people are trying in a way the MBTI people aren't.

But I don't think MBTI *needs* to try to be scientific to do what it does. It's not obvious that MBTI gives you any more than you put into it, and it's not obvious that it should.

By this I mean that you answer a lot of personality questions on a test, like "Do you like spending time around other people?", and you say "no", and then later the test tells you "You're an introvert", and then you think "Oh my god, this is amazing, it's like it's known me my whole life!"

The claim that MBTI gives you new information would be a bold scientific claim and would require bold scientific evidence. I don't know to what degree the MBTI people make this claim, but I don't think it's necessary for me to enjoy the test and consider it useful. All it needs to do is condense the information you put into it in a way that makes it more relevant and digestible.

To say it does so perfectly would be a bold scientific claim. To say it does so at a nonzero level would be an antiprediction, and one that I'm perfectly happy to hand to it.

So let's talk about European countries.

European countries were created in a very unscientific way: some people conquered some other people, and then some organizations like the Congress of Vienna drew a lot of little borders. If we interpret European border-drawing as an attempt to capture "types" of Europeans, it makes some obvious errors – Basques thrown in with Spaniards, Switzerland having Germans and French and Italians, Belgium making NO SENSE, et cetera.

On the other hand, it is massively better than nothing. If I say "Andreas is Greek, but Franz is German", you can make a lot of stereotyped but useful assumptions about them. Andreas probably has darker hair and eyes. Franz is probably more likely to be a rich banker. Andreas drinks more wine, Franz more beer. And so on.

Compare the two-factor model of European classification. We break down variance among Europeans into two dimensions. We find that Andreas is (38, 24) and Franz is (52, 13). That's…nice. Now when I tell you the two dimensions are latitude and longitude of their location, you *feel* like you know what I'm saying. But it's not very human-processable information. If I told you that Maria was at (46, 6) and asked you to describe her, probably your first impulse would be to check a map and see what country that corresponds to – in other words, abandon the perfectly logical Two Factor model and go back to the awful, unscientific Metternich-Bourbon Type Indicator.

Five Factor and MBTI are trying to do fundamentally different things. Five Factor is trying to give us a mathematized, objectively correct version of personality useful for research purposes. MBTI is trying to separate people into little bins that put continuous personality space into discrete and easy-to-think-about terms suitable for human processing, and even very poorly drawn bins will do a pretty good job, just like European countries.

If you want Five Factor to compete with MBTI for some purpose – and really, I'm not sure why you'd care, but some people seem really into this – then it has to compete in the "useful discrete bins" field. Invent a binary opposition between OCEAN – Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism – and SUIDR – Shutness, Unconscientiousness, Introversion, Disagreeableness, Relaxedness. Then let me tell people I'm OCIDN, and they can be like "I'm OCIDN too, it's like I've known you my whole life!" And we can make a web form called "The OCcIDeNt" or something where only OCIDNs are allowed, and talk about how no one else understands the problems that OCIDNs have to face.

Then make a gif about what Harry Potter character each of the thirty-two possibilities corresponds to. Realistically, combine two of the five factors so there are only sixteen possibilities, or else you're going to be scraping the bottom of the Harry Potter barrel: "My type is Charity Burbage's pet owl!"

I would be interested in seeing if such a system catches the popular imagination the same way Myers-Briggs does. I would be interested in seeing whether it gives the same intuitive right answers, like having certain types massively overrepresented on Less Wrong. My guess is that it would. But I don’t think MBTI does too much worse in that regard or that the differences would be worth stressing out over.

Don’t Fear The Filter

There’s been a recent spate of popular interest in the Great Filter theory, but I think it all misses an important point brought up in Robin Hanson’s original 1998 paper on the subject.

The Great Filter, remember, is the horror-genre-adaptation of Fermi’s Paradox. All of our calculations say that, in the infinite vastness of time and space, intelligent aliens should be very common. But we don’t see any of them. We haven’t seen their colossal astro-engineering projects in the night sky. We haven’t heard their messages through SETI. And most important, we haven’t been visited or colonized by them.

This is very strange. Consider that if humankind makes it another thousand years, we’ll probably have started to colonize other star systems. Those star systems will colonize other star systems and so on until we start expanding at nearly the speed of light, colonizing literally everything in sight. After a hundred thousand years or so we’ll have settled a big chunk of the galaxy, assuming we haven’t killed ourselves first or encountered someone else already living there.

But there should be alien civilizations that are a *billion* years old. Anything that could conceivably be colonized, *they* should have gotten to back when trilobytes still seemed like superadvanced mutants. But here we are, perfectly nice solar system, lots of any type of resources you could desire, and they’ve never visited. Why not?

Well, the Great Filter. No knows *specifically* what the Great Filter is, but *generally* it’s “that thing that blocks planets from growing spacefaring civilizations". The planet goes some of the way towards a spacefaring civilization, and then stops. The most important thing to remember about the Great Filter is that it is *very good* at what it does. If even one planet in a billion light-year radius had passed through the Great Filter, we would expect to see its inhabitants everywhere. Since we don't, we know that whatever it is it's *very* thorough.

Various candidates have been proposed, including "it's really hard for life to come into existence", "it's really hard for complex cells to form", "it's really hard for animals to evolve intelligent", and "actually space is full of aliens but they are hiding their existence from us for some reason".

The articles I linked at the top, especially the first, will go through most of the possibilities. This essay isn't about proposing new ones. It's about saying why the old ones won't work.

The Great Filter is not garden-variety x-risk. A lot of people have seized upon the Great Filter to say that we're going to destroy ourselves through global warming or nuclear war or destroying the rainforests. This seems wrong to me. Even if human civilization does destroy itself due to global warming – which is a lot further than even very pessimistic environmentalists expect the problem to go – it seems clear we had a chance *not* to do that. A few politicians voting the other way, we could have passed the Kyoto Protocol. A *lot* of politicians voting the other way, and we could have come up with a really stable and long-lasting plan to put it off indefinitely. If the gas-powered car had never won out over electric vehicles back in the early 20th century, or nuclear-phobia hadn't sunk the plan to move away from polluting coal plants, then the problem might never have come up, or at least been much less. And we're pretty close to being able to colonize Mars right now; if our solar system had a slightly bigger, slightly closer version of Mars, then we could restart human civilization anew there once we destroyed the Earth and maybe go a *little* easy on the carbon dioxide the next time around.

In other words, there’s no way global warming kills 999,999,999 in every billion civilizations. Maybe it kills 100,000,000. Maybe it kills 900,000,000. But *occasionally* one manages to make it to space before frying their home planet. That means it can’t be the Great Filter, or else we would have run into the aliens who passed their Kyoto Protocols.

And the same is true of nuclear war or destroying the rainforests.

Unfortunately, almost all the popular articles about the Great Filter miss this point and make their lead-in “DOES THIS SCIENTIFIC PHENOMENON PROVE HUMANITY IS DOOMED?” No. No it doesn’t.

The Great Filter is not Unfriendly AI. Unlike global warming, it may be that we never really had a chance against Unfriendly AI. Even if we do everything right and give MIRI more money than they could ever want and get all of our smartest geniuses working on the problem, maybe the mathematical problems involved are insurmountable. Maybe the most pessimistic of MIRI’s models is true, and AIs are very easy to accidentally bootstrap to unstoppable superintelligence and near-impossible to give a stable value system that makes them compatible with human life. So unlike global warming and nuclear war, this theory meshes well with the low probability of filter escape.

But as this article points out, Unfriendly AI would if anything be even more visible than normal aliens. The best-studied class of Unfriendly AIs are the ones whimsically called “paperclip maximizers” which try to convert the entire universe to a certain state (in the example, paperclips). These would be easily detectable as a sphere of optimized territory expanding at some appreciable fraction of the speed of light. Given that Hubble hasn’t spotted a Paperclip Nebula (or been consumed by one) it looks like no one has created any of this sort of AI either. And while other Unfriendly AIs might be less aggressive than this, it’s hard to imagine an Unfriendly AI that destroys its parent civilization, then sits very quietly doing nothing. It’s even harder to imagine that 999,999,999 out of a billion Unfriendly AIs end up this way.

The Great Filter is not transcendence. Lots of people more enthusiastically propose that the problem isn’t alien species killing themselves, it’s alien species transcending this mortal plane. Once they become sufficiently advanced, they stop being interested in expansion for expansion’s sake. Some of them hang out on their home planet, peacefully cultivating their alien gardens. Others upload themselves to computronium internets, living in virtual reality. Still others become beings of pure energy, doing whatever it is that beings of pure energy do. In any case, they don’t conquer the galaxy or build obvious visible structures.

Which is all nice and well, except what about the Amish aliens? What about the ones who have weird religions telling them that it’s not right to upload their bodies, they have to live in the real world? What about the ones who have crusader religions telling them they have to conquer the galaxy to convert everyone else to their superior way of life? I’m not saying this has to be common. And I know there’s this argument that *advanced* species would be beyond this kind of thing. But man, it only takes one. I can't believe that not even one in a billion alien civilizations would have some instinctual preference for galactic conquest for galactic conquest's own sake. I mean, even if most humans upload themselves, there will be a couple who don't and who want to go exploring. You're trying to tell me this model applies to 999,999,999 out of one billion civilizations, and then the very first civilization we test it on, it fails?

The Great Filter is not alien exterminators. It sort of makes sense, from a human point of view. Maybe the first alien species to attain superintelligence was jealous, or just plain jerks, and decided to kill other species before they got the chance to catch up. Knowledgeable people like as Carl Sagan and Stephen Hawking have condemned our reverse-SETI practice of sending messages into space to see who's out there, because everyone out there may be terrible. On this view, the dominant alien civilization is the Great Filter, killing off everyone else while not leaving a visible footprint themselves.

Although I get the precautionary principle, Sagan et al's warnings against sending messages seem kind of silly to me. This isn't a failure to recognize how strong the Great Filter has to be, this is a failure to recognize how powerful a civilization that gets through it can become.

It doesn't matter one way or the other if we broadcast we're here. If there are alien superintelligences out there, *they know*. "Oh, my billion-year-old universe-spanning superintelligence wants to destroy fledgling civilizations, but we just can't find them! If only they would send very powerful radio broadcasts into space so we could figure out where they are!" No. Just no. If there are alien superintelligences out there, they tagged Earth as potential troublemakers sometime in the Cambrian Era and have been watching us very closely ever since. They know what you had for breakfast this morning and they know what Jesus had for breakfast the morning of the Crucifixion. People worried about accidentally “revealing themselves” to an intergalactic supercivilization are like Sentinel Islanders reluctant to send a message in a bottle lest modern civilization discover their existence – unaware that modern civilization has spy satellites orbiting the planet that can pick out whether or not they shaved that morning.

What about alien exterminators who are okay with weak civilizations, but kill them when they show the first sign of becoming a threat (like inventing fusion power or leaving their home solar system)? Again, you are underestimating billion-year-old universe-spanning superintelligences. Don’t flatter yourself here. *You cannot threaten them*.

What about alien exterminators who are okay with weak civilizations, but destroy strong civilizations not because they feel threatened, but just for aesthetic reasons? I can’t be certain that’s false, but it seems to me that if they have let us continue existing this long, even though we are made of matter that can be used for something else, that has to be a conscious decision made out of something like morality. And because they’re omnipotent, they have the ability to satisfy all of their (not logically contradictory) goals at once without worrying about tradeoffs. That makes me think that whatever moral impulse has driven them to allow us to survive will *probably* continue to allow us to survive even if we start annoying them for some reason. When you’re omnipotent, the option of stopping the annoyance without harming anyone is just as easy as stopping the annoyance by making everyone involved suddenly vanish.

Three of these four options – x-risk, Unfriendly AI, and alien exterminators – are very very bad for humanity. I think worry about this badness has been a lot of what’s driven interest in the Great Filter. I also think these are some of the least likely possible explanations, which means we should be less afraid of the Great Filter than is generally believed.

The Wonderful Thing About Triggers

[Content note: hypothetical spiders]

I complain a lot about the social justice movement. Or for a change, I sometimes complain that the media is too friendly to the social justice movement. So when the media starts challenging the movement, with articles like We've Gone Too Far With Trigger Warnings, I really ought to be happy things are finally going my way.

Instead I'm a little disappointed

I like trigger warnings. Trigger warnings aren't censorship; they're the opposite of censorship. Censorship says "Read what we tell you". The opposite of censorship is "Read whatever you want". The philosophy of censorship is "We know what is best for you to read". The philosophy opposite censorship is "You are an adult and can make your own decisions about what to read".

But part of letting people make their own decisions is giving them relevant information and trusting them to know what to do with them. Uninformed choices are worse choices. Trigger warnings are an attempt to provide you with the information to make good free choices of reading material.

Imagine someone only interested in geology, who only reads geology books and neglects all the other interesting fields of science, history, and art. In a sense, we're enabling this person by putting titles on books – it helps her avoid all the other interesting things she could be reading that challenge her monomaniacal focus on geology and open her mind a little. If books didn't have titles, then instead of heading straight for the geology books, maybe she'd pick up something that looked like it was a geology book and end up reading *Don Quixote* and be better off for it. Our decision to title books anyway assumes that we care less about denying people the ability to avoid things they don't want to read, and more about trusting people's judgment and making it easy for them to choose and obtain what they want. This is the philosophy behind trigger warnings as well. If you don't like it, then I accuse you of not liking capitalism. You may prefer a country with a military draft; those are *really good* at making sure people are involuntarily exposed to the harshness of reality rather than sheltered from it.

So my role model here, as in so many other places, is Commissioner Lal: "Beware he who would deny you access to information, for in his heart, he dreams himself your master."

Trigger warnings fight those who would like to be our masters in another way as well. They're one of our strongest weapons against the proponents of censorship. The proponents say "We can't let you air that opinion, it might offend people." Trigger warnings say "I am explaining to you exactly how this might offend you, so if you continuing listening to me you have volunteered to hear whatever I have to say, on your own head be it."

I agree that bad people could use trigger warnings to avoid ever reading anything that challenges their prejudices. This is a problem with providing people informed choices. Sometimes they misuse them.

But I could also imagine good people using trigger warnings to *increase* their ability to read things that challenge their views. Suppose you are a transgender person who becomes really uncomfortable when you hear people insult transgender people. Gradually you learn that a lot of people outside the social justice community do this a lot, so you stop reading anything outside the social justice community – forget about genuinely rightist sources like National Review or American Conservative. Now suppose sources start trigger warning their content. Most right-wing arguments *don't* insult transgender people, so all of a sudden you have a way to steer clear of the ones that do and read all of the others free from fear.

Actually, "fear" is the wrong word, it buys into the stereotyping of triggered people as coddled or cowards or something. Maybe some people feel fear. Others would just be free from exasperation, anger, distracting dismay, the cognitive load of having to hear people insult you and not being able to respond and having to exert effort to continue to read. I feel like this might be my response to the existence of more trigger warnings (at least if anyone ever warned for my triggers, which they won't).

And I guess I admit that the people who use trigger warnings to narrow their minds will probably outnumber those who use them to expand them. But the question is still: do "we", as a civilization, force people to be virtuous without their consent? And even if we should, how come we don't do this in any other area of life? Why don't we make a rule that every fifth link to Paul Krugman's blog has to redirect people to Tyler Cowen's blog, and vice versa, so people don't get a chance to only read the opinions they agree with? Or that every Republican has to watch one Daily Show a month, and every Democrat has to listen to one Fox News segment?

If we're not going to do any of that, it hardly seems fair to put the whole burden of epistemic virtue on the easily triggered.

II.

The strongest argument against trigger warnings that I have heard is that they allow us to politicize ever more things. Colleges run by people on the left can slap big yellow stickers on books that promote conservative ideas, saying "THIS BOOK IS RACIST AND CLASSIST", and then act outraged if anyone requests a trigger warning that sounds conservative – like a veteran who wants one on books that vilify or mock soldiers, or a religious person who wants one on blasphemy. Then everyone has to have a big fight, the fight makes everyone worse off than *either* possible resolution, and it ends with somebody feeling persecuted and upset. In other words, it's an intellectual gang sign saying "Look! We can demonstrate our mastery of this area by only allowing our symbols; your kind are second-class citizens!"

On the other hand, this is terribly easy to fix. Put trigger warnings on books, but put them on the bullshit page. You know, the one near the front where they have the ISBN number and the city where the publishers' head office is and something about the Library of Congress you've never read through even though it's been in literally every book you've ever seen. Put it there, on a small non-colorful sticker. Call it a "content note" or something, so no one gets the satisfaction of hearing their pet word "trigger warning". Put a generally agreed list of things – no sense letting every single college have its own acrimonious debate about it. The few people who actually get easily triggered will with some exertion avoid the universal human urge to flip past the bullshit page and spend a few seconds checking if their trigger is in there. No one else will even notice.

Or if it's about a syllabus, put it on the last page of the syllabus, in size 8 font, after the list of recommended reading for the class. As a former student and former teacher, I *know* no one reads the syllabus. You have to be really devoted to avoiding your trigger. Which is exactly the sort of person who should be able to have a trigger warning while everyone else goes ahead with their lives in a non-political way.

I'm sure there are some more implementation details, but it's nothing a little bit of good faith can't take care of. If good faith is used and some people *still* object because it's not EXACTLY what they want, *then* I'll tell them to go fly a kite, but not before.

I know a lot of people worry about slippery slopes; give the culture warriors an inch and they'll take a mile. I think this is a very backwards way of looking at things. Like, the anti-gay people talked about a slippery slope and fought desperately hard against gay marriage, even though it was pretty hard to find anything actually objectionable about it other than that it might be on a slippery slope to worse things. That desperate fight didn't delay gay marriage more than a few years, and it didn't prevent whatever gay marriage was on a slippery slope to. What it did do was *totally discredit conservatives in this area*. Now any time anyone makes a family values argument, even a good family values argument, people can say that "family values" is code for homophobia, and bring up that family values conservatives really *have* held abhorrent positions in the past so why should we trust them now? It gave liberals huge momentum, and if there is a slippery slope then all that opposing gay marriage did was destroy the credibility of anybody who could have stopped us going down it.

Opposing a good idea on slippery slope grounds is a moral failure *and* a strategic failure, and I'd hate for opponents of the social justice movement to make that mistake with trigger warnings.

III.

But this is all tangential to what really bothered me, which is Pacific Standard’s The Problems With Trigger Warnings According To The Research.

You know, I love science as much as anyone, maybe more, but I have grown to dread the phrase “…according to the research”.

They say that “Confronting triggers, not avoiding them, is the best way to overcome PTSD”. They point out that “exposure therapy” is the best treatment for trauma survivors, including rape victims. And that this involves reliving the trauma and exposing yourself to traumatic stimuli, exactly what trigger warnings are intended to prevent. All this is true. But I feel like they are missing a very important point.

YOU DO NOT GIVE PSYCHOTHERAPY TO PEOPLE WITHOUT THEIR CONSENT.

Psychotherapists treat arachnophobia with exposure therapy, too. They expose people first to cute, little spiders behind a glass cage. Then bigger spiders. Then they take them out of the cage. Finally, in a carefully controlled environment with their very supportive therapist standing by, they make people experience their worst fear, like having a big tarantula crawl all over them. It usually works pretty well.

Finding an arachnophobic person, and throwing a bucket full of tarantulas at them while shouting “I’M HELPING! I’M HELPING!” works less well.

And this seems to be the arachnophobe’s equivalent of the PTSD “advice” in the Pacific Standard. There are two problems with its approach. The first is that it avoids the carefully controlled, anxiety-minimizing setup of psychotherapy.

The second is that YOU DO NOT GIVE PSYCHOTHERAPY TO PEOPLE WITHOUT THEIR CONSENT.

If a person with post-traumatic stress disorder or some other trigger-related problem doesn’t want psychotherapy, then *even as a trained psychiatrist* I am forbidden to override that decision unless they become an immediate danger to themselves or others.

And if they *do* want psychotherapy, then very likely they want to do it on their own terms. I try to read things that challenge my biases and may even insult or trigger me, but I do it *when I feel like it* and not a moment before. When I am feeling adventurous and want to become stronger in some way, I will set myself some strenuous self-improvement task, whether it be going on a long run or reading material I know will be unpleasant. But at the end of a really long and exasperating day when I’m at my wit’s end and just want to relax, I don’t want you chasing me with a sword and making me run for my life, and I don’t want you forcing traumatic material at me.

The angry article above with all the talk of “spoiled brats” annoys me as an amateur politics blogger, but this Pacific Standard article pushes my buttons as a (somewhat) non-amateur psychiatrist. *This is not your job to meddle*. If you are very concerned about helping people with PTSD, please express that concern by donating to PTSD USA or one of the other organizations that will help those with the condition get proper, well-controlled therapy. Please do *not* try to increase the background level of triggers in the hopes that one of them will fortuitously collide with a PTSD sufferer in a therapeutic way.

If, like me, you think the social justice movement has a really serious kindness and respect problem, then you know that it’s really hard to bring this up without getting accused of unkindness and disrespect yourself. I don’t know how to best respond to this problem. But I’m pretty sure that the very minimum one can do is not to *actually* be unkind and disrespectful. And I worry that some of these arguments against trigger warnings are failing to clear even this very low bar.

Southeast Michigan Meetup 6/8

There will be a Michigan rationalist/LW/SSC meetup at my house, halfway between Detroit and Ann Arbor, on Sunday, June 8th at 2 PM. You can find my address and directions here

Topic will be “Yes, It’s Another Michigan Meetup”

Ozy is now a permanent Michigan resident, so there is no Special Guest. If you are coming from afar, let me know and I will add your name as the Special Guest.

Asches to Asches

[Content note: fictional story contains gaslighting-type elements. May induce Cartesian skepticism]

You wake up in one of those pod things like in *The Matrix*. There's a woman standing in front of you, wearing a lab coat, holding a clipboard.

"Hi," she says. "This is the real world. You used to live here. We erased your memories and stuck you in a simulated world for a while, like in *The Matrix*. It was part of a great experiment."

"What?" you shout. "My whole life, a lie? How dare you deceive me as part of some grand 'experiment' I never consented to?"

"Oh," said the woman, "actually, you did consent, in exchange for extra credit in your undergraduate psychology course." She hands you the clipboard. There is a consent form with your name on it, in your handwriting.

You give her a sheepish look. "What was the experiment?"

"You know families?" asks the woman.

"Of course," you say.

"Yeah," says the woman. "Not really a thing. Like, if you think about it, it doesn't make any sense. Why would you care more for your genetic siblings and cousins and whoever than for your friends and people who are genuinely close to you? That's like racism – but even worse, at least racists identify with a group of millions of people instead of a group of half a dozen. Why should parents have to raise children whom they might not even like, who might have been a total accident? Why should people, motivated by guilt, make herculean efforts to "keep in touch" with some nephew or cousin whom they clearly would be perfectly happy to ignore entirely?"

"Uh," you say, "not really in the mood for philosophy. Families have been around forever and they aren't going anywhere, who cares?"

"Actually," says the woman, "in the real world, no one believes in family. There's no such thing. Children are taken at birth from their parents and given to people who contract to raise them in exchange for a fixed percent of their future earnings."

"That's monstrous!" you say. "When did this happen? Weren't there protests?"

"It's always been this way," says the woman. "There's *never* been such a thing as the family. Listen. You were part of a study a lot like the Asch Conformity Experiment. Our goal was to see if people, raised in a society where everyone believed X and everything revolved around X, would even be *capable* of questioning X or noticing it was stupid. We tried to come up with the stupidest possible belief, something no one in the real world had ever believed or ever seemed likely to, to make sure that we were isolating the effect of conformity and not of there being a legitimate argument for something. So we chose this idea of 'family'. There are racists in our world, we're not perfect, but as far as I know none of them has *ever* made the claim that you should devote extra resources to the people genetically closest to you. That's like a *reductio ad absurdum* of racism. So we got a grad student to simulate a world where this bizarre idea was the unquestioned status quo, and stuck twenty bright undergraduates in it to see if they would conform, or question the premise."

"Of course we won't question the premise, the premise is…"

"Sorry to cut you off, but I thought you should know that every single one of the other nineteen subjects, upon reaching the age where the brain they were instantiated in was capable of abstract reason, immediately determined that the family structure made no sense. One of them actually deduced that she was in a psychology experiment, because there was no other explanation for why everyone believed such a bizarre premise. The other eighteen just assumed that sometimes objectively unjustifiable ideas caught on, the same way that everyone in the antebellum American South thought slavery was perfectly natural and only a few abolitionists were able to see through it. Our conformity experiment *failed*. You were actually the only one to fall for it, hook line and sinker."

"How could I be the only one?"

"We don't know. Your test scores show you're of just-above-average intelligence, so it's not that you're stupid. But we did give all participants a personality test that showed you have very high extraversion. The conclusion of our paper is going to be that very extraverted participants adopt group consensus without thinking and can be led to believe anything, even something as ridiculous as 'family'".

"I guess…when you put it like that it is kind of silly. Like, my parents were never that nice to me, but I kept loving them anyway, liking them even more than other people who treated me a lot better – and god, I even gave my mother a "WORLD'S #1 MOM" mug for Mother's Day. That doesn't even make sense! I…but what about the evolutionary explanation? Doesn’t evolution say we have genetic imperatives to love and support our family, whether they are worthy of it or not?”

“You can make a just-so story for *anything* using evolutionary psychology. Someone as smart as you should know better than to take them seriously.”

“But then, what *is* evolution? How did animals reproduce before the proper economic incentives were designed? Where did…”

“Tell you what. Let’s hook you up to the remnemonizer to give you your real memories back. That should answer a lot of your questions.”

A machine hovering over you starts to glow purple. “This shouldn’t hurt you a bit…”

>discontinuity<

You wake up in one of those pod things like in *The Matrix*. There’s a woman standing in front of you, wearing a lab coat, holding a clipboard.

“Hi,” she said. “There’s no such thing as virtual reality. I hypnotized you to forget all your memories from the past day and to become very confused. Then I put you in an old prop from *The Matrix* I bought off of eBay and fed you that whole story.”

“What?” you shout. “You can’t just go hypnotizing and lying to people without their consent!”

“Oh,” said the woman, “actually, you did consent, in exchange for extra credit in your undergraduate psychology course.” She hands you the clipboard. There is a consent form with your name on it, in your handwriting. “That part was true.”

You give her a sheepish look. “Why would you do such a thing?”

“Well,” said the woman. “You know the Asch Conformity Experiment? I was really interested in whether you could get people to abandon some of their most fundamental beliefs, just by telling them other people believed differently. But I couldn’t think of a way to test it. I mean, part of a belief being fundamental is that you already *know* everyone else believes it. There’s no way I could convince subjects that the whole world was against something as obvious as ‘the family’ when they already know how things stand.

“So I dreamt up the weird ‘virtual reality’ story. I figured I would convince subjects that the real world was a lie, and that in some ‘super-real’ world supposedly *everybody knew* that the family was stupid, that it wasn’t even an idea *worth considering*. I wanted to know how many people would give up something they’ve believed in for their entire life, just because they’re told that ‘nobody else thinks so'”.

“Oh,” I said. “Interesting. So even our most cherished beliefs are more fragile than we think.”

“Not *really*,” said the woman. “Of twenty subjects, you were the only person I got to feel any doubt, or to express any kind of anti-family sentiment.”

“Frick,” you say. “I feel like an idiot now. What if my mother finds out? She’ll think it’s her fault or something. God, she’ll think I don’t love her. People are going to be talking about this one *forever*.”

“Don’t worry,” says the woman. “We’ll keep you anonymized in the final data. Anyway, let’s get you your memories back so you can leave and be on your way.”

“You can restore my memories?” you say.

“Of course. We hypnotized you to forget the last day’s events until you heard a trigger word. And that trigger is…”

>discontinuity<

You wake up in one of those pod things like in *The Matrix*. There’s a woman standing in front of you, wearing a lab coat, holding a clipboard.

“Hi,” she says. “Hypnosis is a pseudoscience and doesn’t work. It was the virtual reality one, all along.”

“Wut,” you say.

“I mean, the first story was true. All of your memories of living with your family and so on are fake memories from a virtual world, like in *The Matrix*. The concept of ‘family’ really is totally ridiculous and no one in the real world believes it. All the stuff you heard first was true. The stuff about hypnosis and getting a prop from *The Matrix* off eBay was false.”

“But…why?”

“We wanted to see exactly how far we could push you. You’re our star subject, the only one whom we were able to induce this bizarre conformity effect in. We didn’t know whether it was because you were just very very suggestible, or whether because you had never seriously considered the idea that ‘family’ might be insane. So we decided to do a sort of…crossover design, if you will. We took you here and debriefed you on the experiment. Then after we had told you how the world really worked, given you all the mental tools you needed to dismiss the family once and for all, even gotten you to admit we were right – we wanted to see what would happen if we sent you back. Would you hold on to your revelation and boldly deny your old society's weird prejudices? Or would you switch sides again and start acting like family made sense the second you were in a pro-family environment?"

"And I did the second one."

"Yes," says the woman. "As a psychologist, I'm supposed to remain neutral and non-judgmental. But you've got to admit, you're pretty dumb."

"Is there an experimental ethics committee I could talk to here?"

"Sorry. Experimental ethics is another one of those obviously ridiculous concepts we planted in your simulation to see if you would notice. Seriously, to believe that the progress of science should be held back by the prejudices of self-righteous fools? That's almost as weird as thinking you have a…what was the word we used…'sister'."

"Okay, look, I realize I may have gone a little overboard helping my sister, but the experimental ethics thing seems important. Like, what's going to happen to me now?"

"Nothing's going to happen. We'll keep all your data perfectly anonymous, restore your memories, and you can be on your way."

"Um," you say. "Given past history, I'm…actually not sure I want my memories restored." You glare at the remnemonizer hovering above you. "Why don't I just…"

The woman's eyes narrow. "I'm sorry," she says. "I can't let you do that."

The machine starts to glow.

>discontinuity<

You wake up in one of those pod things like in *The Matrix*. There's a woman standing in front of you, wearing a lab coat, holding a clipboard.

By your count, this has happened three hundred forty six times before.

There seem to be two different scenarios. In one, the woman tells you that families exist, and have always existed. She says she has used hypnosis to make you believe in the other scenario, the one with the other woman. She asks you your feelings about families and you tell her.

Sometimes she lets you go. You go home to your mother and father, you spend some time with your sister. Sometimes you tell them what has happened. Other times you don't. You cherish your time with them, while also second-guessing everything you do. *Why* are you cherishing your time with them? Your father, who goes out drinking every night, and who has cheated on your mother more times than you can count. Your mother, who was never there for you when you needed her most. And your sister, who has been good to you, but no better than millions of other women would be, in her position. Are they a real family? Or have they been put there as a symbol of something ridiculous, impossible, something that has never existed?

It doesn't much matter. Maybe you spend one night with them. Maybe ten. But within a month, you are always waking up in one of those pod things like in *The Matrix*.

In the second scenario, the woman tells you there are no families, never have been. She says she has used virtual reality to make you believe in the other scenario, the one with the other woman. She asks you your feelings about families and you tell her.

Sometimes she lets you go. You go to a building made of bioplastic, where you live with a carefully chosen set of friends and romantic partners. They assure you that this is how everyone lives. Occasionally, an old and very wealthy-looking man checks in with you by videophone. He reminds you that he has invested a lot of money in your upbringing, and if there's any way he can help you, anything he can do to increase your future earnings potential, you should let him know. Sometimes you talk to him, and he tells you strange proverbs and unlikely business advice.

It doesn't matter. Maybe you spend one night in your bioplastic dwelling. Maybe ten. But within a month, you are always waking up in one of those pod things like in *The Matrix*.

"Look," you tell the woman. "I'm tired of this. I know you're not bound by any kind of experimental ethics committee. But please, for the love of God, have some mercy."

"God?" asks the woman. "What does that word mean? I've never…oh right, we used *that* as our intervention in the prototype experiment. We decided 'family' made a better test idea, but Todd must have forgotten to reset the simulator."

"It's been three hundred forty six cycles," you tell her. "Surely you're not learning anything new."

“I’ll be the judge of that,” she says. “Now, tell me what you think about families.”

You refuse. She sighs. Above you, the remnemonizer begins to glow purple.

>discontinuity<

You wake up in one of those pod things like in *The Matrix*. There’s a purple, tentacled creature standing in front of you, wearing a lab coat, holding a clipboard.

“Hi,” it says. “Turns out there’s no such thing as humans.”

You refuse to be surprised.

“There’s only us, the 18-tkenna-dganna-07.”

“Okay,” you say. “I want answers.”

“Absolutely,” says the alien. “We would like to find optimal social arrangements.”

“And?”

“And I cannot tell you whether we have families or not, for reasons that are to become apparent, but the idea is at least sufficiently interesting to have entered the space of hypotheses worth investigating. But we don’t trust ourselves to investigate this. It’s the old Asch Conformity Problem again. If we have families, then perhaps the philosophers tasked with evaluating families will conform to our cultural norms and decide we should keep them. If we do not, perhaps the philosophers will conform and decide we should continue not to. So we determined a procedure that would create an entity capable of fairly evaluating the question of families, free from conformity bias.”

“And that’s what you did to me.”

“Yes. Only by exposing you to the true immensity of the decision, without allowing you to fall back on what everyone else thinks, could we be confident in your verdict. Only by allowing you to experience both how obviously right families are, when you ‘know’ they are correct, and how obviously wrong families are, when you ‘know’ they are incorrect, could we expect you to garner the wisdom to be found on both sides of the issue.”

“I see,” you say, and you do.

“Then, O purified one,” asks the alien, “tell us of your decision.”

“Well,” you say. “If you have to know, I think there are about equally good points on both sides of the issue.”

“Fuck,” says the 18-tkenna-dganna-07.

Don’t Be An Asch-Hole

Contrary to the direction the comments took, my Asch story the other day wasn’t intended to make any special commentary on families. That was just the first issue I thought of that wasn’t already so politicized that the story would be interpreted as political propaganda. It was really meant to be an explanation of something I said a while back on Twitter:

> So live your life that if this all turns out to be an alien simulator's Asch Conformity Experiment, the debriefing won't be too humiliating
>
> — Scott Alexander (@slatestarcodex) May 30, 2014

The character in the story had it even worse. They were told that in the real world, not a single person believed in families. That “families” had been invented as a test concept for the conformity experiment solely because it was something so ridiculous that no one could possibly believe it for not-conformity-related reasons. That every other experimental participant had seen through the facade and denounced it as dumb.

And I feel like this is a good thought experiment. Which beliefs of yours would survive that knowledge, be so strong that you would tell the experimenter that you are right or they are wrong, or make you start thinking that it’s all part of a meta-experiment like in the story? Which ones would you *start* to doubt in ways that you might not have thought of back when they were common? Which, if any, would you say “Yeah, I knew it all along, I guess I was just too scared to admit it”?

I’m thinking here of antebellum Southerners, let’s say early 1800s. Their society is built around slavery. There are a *couple* of abolitionists around, but not many, and none who can force anyone to listen to them. Pretty much everyone around them says slavery is okay, the books they read from the past are all about Romans or Israelites who thought (rather different forms of) slavery were okay, and they have heard a lot of plausible-sounding arguments justifying slavery.

Now bring them forward to the present day. Tell them “Right now in the present day pretty much every single person believes that slavery is morally wrong. No one would justify it. Here, come out of the laboratory and spend a few years living in our slave-free society.”

I don’t know if the Southerner would learn a whole lot of new *facts* during this period. They might learn that black people could be pretty capable and intelligent, but Frederick Douglass was a person, everyone knew he was smart, that didn’t change anyone’s mind. Yet even without learning many new facts, I can’t imagine he would stay pro-slavery very long.

And I wonder whether this is *purely* a conformity thing, and upon being returned to the antebellum South he would start conforming with them again, or whether it is a one-directional effect that primes your thoughts to go in the direction of the truth and allows you to see new valid arguments, and that upon going back to the South he would be a little wiser than his countrymen.

And I also wonder whether a sufficiently smart Southerner could do all this via a thought experiment, say “I think slavery is pretty okay now, but imagine I went to a world where everyone was absolutely certain it was terrible, how bad would I feel about it?” and get all the benefits of spending a while in our world and going through all that moral reflection without ever actually leaving the antebellum South. And if this would be a *more* powerful intuition pump than just asking him to sit down and think about slavery for a few hours.

This is a pretty powerful ethical test for me. I imagine waking up in that Matrix pod and being told that no one in the real world believes in abortion, that pro-choice is *obviously* horrible, that all my fellow experimental subjects saw through it, that as far as they can tell I’m just a psychopath. And I feel like I would still argue “No, actually, I think *you* guys are wrong.” (but, uh, your mileage may vary)

If it was vegetarianism – if they said no one in the real world ate meat or had tried to justify factory farming, and every single one of my co-participants had become vegan animal rights activists – I don’t think there’s a lot I could say to them. “Sorry, I have an intense disgust reaction to all vegetables which has thwarted all of my attempts at vegetarianism?” “Yeah, we know, we put that in there to make it a hard choice.”

There are some issues where I could imagine it going either way. If the alien simulators were conservative, I could imagine exactly the way in which I would feel really stupid for having ever believed in liberalism. And if the alien simulators were liberal, I could imagine exactly how it would feel to get embarrassed for ever having flirted with conservative ideas. I don’t think that’s necessarily a flaw in the thought experiment. Both of those feelings are *useful* to me.

I had a dream once where I died and I went to the afterlife and there was God and He told me that Christianity had been right about everything all along. In retrospect, it felt kind of obvious. Then I woke up and it had all been a dream. In retrospect, that felt kind of obvious too.

I sort of cherish these feelings of obviousness. They seem like a good way to short-circuit the absurdity heuristic. Like, it doesn't work to think "Okay, Christianity seems absurd, but WHAT IF IT DIDN'T???" You have to actually invert everything, tell yourself that lack-of-Christianity seems absurd and see what justifications and excuses your brain starts coming up with for why that is the correct position and atheism should be dismissed without a second thought.

A commenter on the story thread came up with a different application of the concept I hadn't thought of:

> I gained a lot of value out of applying this thought experiment through a motivational rather than rational frame.
>
> I struggle with ADHD and depression, and imagined myself being told that everyone else in the simulations also had those conditions, but all of them performed better than me.
>
> Then I imagined a different scenario, where everyone was put into a simulation where they needed to save the world, and everyone else succeeded where only I failed.
>
> Imagining changes in others' beliefs doesn't effect mine much. But imagining changes in others' standards of competence has an extremely powerful impact.
>
> Thanks for the tool.

While this seems a *liiiiiitle* dangerous almost to the point of mental self-abuse, I can kind of see it working, and it helps me understand the other cases a little better. We hold ourselves to certain standards – whether moral, like the antebellum Southerner or the non-vegetarian me – or epistemic – or motivational. By altering the perceived competence of other people, we can artificially adjust the standards we hold ourselves to and pretend for a while that our standards are much higher than they are. Which might give us insights we can bring back with us.

Archipelago and Atomic Communitarianism

I.

In the old days, you had your Culture, and that was that. Your Culture told you lots of stuff about what you were and weren’t allowed to do, and by golly you listened. Your Culture told you to work the job prescribed to you by your caste and gender, to marry who your parents told you to marry or at *least* someone of the opposite sex, to worship at the proper temples and the proper times, and to talk about *proper* things as opposed to the blasphemous things said by the tribe over there.

Then we got Liberalism, which said all of that was mostly bunk. Like Wicca, its motto is “Do as you will, so long as it harms none”. Or in more political terms, “Your right to swing your fist ends where my nose begins” or “If you don’t like gay sex, don’t have any” or “If you don’t like this TV program, don’t watch it” or “What happens in the bedroom between consenting adults is none of your business” or “It neither breaks my arm nor picks my pocket”. Your job isn’t to enforce your conception of virtue upon everyone to build the Virtuous Society, it’s to live your own life the way you want to live it and let other people live *their* own lives the way *they* want to live them. This is the much-maligned “atomic individualism,” or maybe just liberalism boiled down to its pure essence.

But atomic individualism wasn’t as great a solution as it sounded. Maybe one of the first cracks was tobacco ads. Even though putting up a billboard saying “SMOKE MARLBORO” neither breaks anyone’s arm nor picks their pocket, it shifts social expectations in such a way that bad effects occur. It’s hard to dismiss that with “Well, it’s people’s own choice to smoke and they should live their lives the way they want” if studies show that more people will want to live their lives in a way that gives them cancer in the presence of the billboard than otherwise.

From there we go into policies like Michael Bloomberg’s ban on giant sodas. While the soda ban itself was probably as much symbolic as anything, it’s hard to argue with the impetus behind it – a culture where everyone gets exposed to the option to buy very very unhealthy food all the time is going to be less healthy than one where there are some regulations in place to make EAT THIS DONUT NOW a less salient option. I mean, I *know* this is true. A few months ago when I was on a diet I *cringed* every time one my coworkers brought in a box of free donuts and placed wide-open in the doctors’ lounge; there was *no way* I wasn’t going to take one (or two, or three). I could ask people to stop, but they probably wouldn’t, and even if they did I’d just encounter the wide-open box of free donuts *somewhere else*. I’m not proposing that it is *ethically wrong* to bring in free donuts or that banning them is the correct policy, but I do want to make it clear that stating “it’s your free choice to partake or not” doesn’t eliminate the problem, and that this points to an entire class of serious issues where atomic individualism as construed above is at best an imperfect heuristic.

And I would be remiss talking about the modern turn away from individualism without mentioning social justice. The same people who once deployed individualistic arguments against conservatives: “If you don’t like profanity, don’t use it”, “If you don’t like this offensive TV show, don’t watch it”, “If you don’t like pornography, don’t buy it” – are now concerned about people using ethnic slurs, TV shows without enough minority characters, and pornography that encourages the objectification of women. I’ve objected to some of this on purely empirical grounds, but the least convenient possible world is the one where the purely empirical objections fall flat. If they ever discover proof positive that yeah, pornographication makes women hella objectified, is it acceptable to censor or ban misogynist media on a society-wide level?

And if the answer is yes – and if such media like really, *really* increases the incidence of rape I'm not sure how it couldn't be – then what about all those conservative ideas we've been neglecting for so long? What if strong, cohesive, religious, demographically uniform communities make people more trusting, generous, and cooperative in a way that *also* decreases violent crime and other forms of misery? We have lots of evidence that this is true, and although we can doubt each individual study, we owe conservatives the courtesy of imagining the possible world in which they are right, the same as anti-misogyny leftists. Maybe media glorifying criminals or lionizing nonconformists above those who quietly follow cultural norms has the same kind of erosive effects on "values" as misogynist media. Or, at the very least, we ought to have a good philosophy in place so that we have some idea what to do it if does.

II.

A while ago, in Part V of this essay, I praised liberalism as the only peaceful answer to Hobbes' dilemma of the war of all against all.

Hobbes says that if everyone's fighting then everyone loses out. Even the winners probably end up worse off than if they had just been able to live in peace. He says that governments are good ways to prevent this kind of conflict. Someone – in his formulation a king – tells everyone else what they're going to do, and then everyone else does it. No fighting necessary. If someone tries to start a conflict by ignoring the king, the king crushes them like a bug, no prolonged fighting involved.

But this replaces the problem of potential warfare with the problem of potential tyranny. So we've mostly shifted from absolute monarchies to other forms of government, which is all nice and well except that governments allow a *different* kind of war of all against all. Instead of trying to kill their enemies and steal their stuff, people are tempted to ban their enemies and confiscate their stuff. Instead of killing the Protestants, the Catholics simply ban Protestantism. Instead of forming vigilante mobs to stone homosexuals, the straights merely declare homosexuality is punishable by death. It *might* be better than the alternative – at least everyone knows where they stand and things stay peaceful – but the end result is still a lot of pretty miserable people.

Liberalism is a new form of Hobbesian equilibrium where the government enforces not only a ban on killing and stealing from people you don't like, but also a ban on tyrannizing them out of existence. This is the famous "freedom of religion" and "freedom of speech" and so on, as well as the "freedom of what happens in the bedroom between consenting adults". The Catholics don't try to ban Protestantism, the Protestants don't try to ban Catholicism, and everyone is happy.

Liberalism only works when it's clear to everyone on all sides that there's a certain neutral principle everyone has to stick to. The neutral principle can't be the Bible, or Atlas Shrugged, or anything that makes it look like one philosophy is allowed to judge the others. Right now that principle is the Principle of Harm: you can do whatever you like unless it harms other people, in which case stop. We seem to have inelegantly tacked on an "also, we can collect taxes and use them for a social safety net and occasional attempts at social progress”, but it seems to be working pretty okay too.

The Strict Principle of Harm says that pretty much the only two things the government can get angry at is literally breaking your leg or picking your pocket – violence or theft. The Loose Principle of Harm says that the government can get angry at complicated indirect harms, things that Weaken The Moral Fabric Of Society. Like putting up tobacco ads. Or having really really big sodas. Or publishing hate speech against minorities. Or eroding trust in the community. Or media that objectifies women.

No one except the most ideologically pure libertarians seems to want to insist on the Strict Principle of Harm. But allowing the Loose Principle Of Harm restores all of the old wars to control other people that liberalism was supposed to prevent. The one person says “Gay marriage will result in homosexuality becoming more accepted, leading to increased rates of STDs! That’s a harm! We must ban gay marriage!” Another says “Allowing people to send their children to non-public schools could lead to kids at religious schools that preach against gay people, causing those children to commit hate crimes when they grow up! That’s a harm! We must ban non-public schools!” And so on, forever.

And I’m talking about non-governmental censorship just as much as government censorship. Even in the most anti-gay communities in the United States, the laws usually allow homosexuality or oppose it only in very weak, easily circumvented ways. The real problem for gays in these communities is the social pressure – whether that means disapproval or risk of violence – that they would likely face for coming out. This too is a violation of liberalism, and it’s one that’s as important or more important than the legal sort.

And right now our way of dealing with these problems is to argue them. “Well, gay people don’t really increase STDs too much.” Or “Home-schooled kids do better than public-schooled kids, so we need to allow them.” The problem is that arguments never terminate. Maybe if you’re *incredibly* lucky, after years of fighting you can get a couple of people on the other side to admit your side is right, but this is a pretty hard process to trust. The great thing about religious freedom is that it short-circuits the debate of “Which religion is correct, Catholicism or Protestantism?” and allows people to tolerate both Catholics and Protestants even if they are divided about the answer to this object-level question. The great thing about freedom of speech is that it short-circuits the debate of “Which party is correct, the Democrats or Republicans?” and allows people to express both liberal and conservative opinions even if they are divided about the object-level question.

If we force all of our discussions about whether to ban gay marriage or allow home schooling to depend on resolving the dispute about whether they indirectly harm the Fabric of Society in some way, we’re forcing dependence on object-level arguments in a way that historically has been very very bad.

Presumably here the more powerful groups would win out and be able to oppress the less powerful groups. We end up with exactly what liberalism tried to avoid – a society where everyone is the guardian of the virtue of everyone else, and anyone who wants to live their lives in a way different from the community’s consensus is out of luck.

In Part I, I argued that *not allowing* people to worry about culture and community at all was inadequate, because these things really do matter.

Here I'm saying that if we *do allow* people to worry about culture and community, we risk the bad old medieval days where all nonconformity gets ruthlessly quashed.

Right now we're balanced precariously between the two states. There's a lot of liberalism, and people are generally still allowed to be gay or home-school their children or practice their religion or whatever. But there's also quite a bit of Enforced Virtue, where kids are forbidden to watch porn and certain kinds of media are censored and in some communities mentioning that you're an atheist will get you Dirty Looks.

It tends to work okay for most of the population. Better than the alternatives, maybe? But there's still a lot of the population that's not free to do things that are very important to them. And there's also a lot of the population that would like to live in more "virtuous" communities, whether it's to lose weight faster or avoid STDs or not have to worry about being objectified. Dealing with these two competing issues is a pretty big part of political philosophy and one that most people don't have any principled solution for.

III.

Imagine a new frontier suddenly opening. Maybe a wizard appears and gives us a map to a new archipelago that geographers had missed for the past few centuries. He doesn't want to rule the archipelago himself, though he will reluctantly help kickstart the government. He just wants to give directions and a free galleon to anybody who wants one and can muster a group of likeminded friends large enough to start a self-sustaining colony.

And so the equivalent of our paleoconservatives go out and found communities based on virtue, where all sexual deviancy is banned and only wholesome films can be shown and people who burn the flag are thrown out to be eaten by wolves.

And the equivalent of our social justiciars go out and found communities where all movies have to have lots of strong minority characters in them, and all slurs are way beyond the pale, and nobody misgenders anybody.

And the equivalent of our Objectivists go out and found communities based totally on the Strict Principle of Harm where everyone is allowed to do whatever they want and there are no regulations on business and everything is super-capitalist all the time.

And some people who just really want to lose weight go out and found communities where you’re not allowed to place open boxes of donuts in the doctors’ lounge.

Usually the communities are based on a charter, which expresses some founding ideals and asks only the people who agree with those ideals to enter. The charter also specifies a system of government. It could be an absolute monarch, charged with enforcing those ideals upon a population too stupid to know what’s good for them. Or it could be a direct democracy of people who all agree on some basic principles but want to work out for themselves what direction the principles take them.

After a while the wizard decides to formalize and strengthen his system, not to mention work out some of the ethical dilemmas.

First he bans communities from declaring war on each other. That’s an *obvious* gain. He could just smite warmongers, but he thinks it’s more natural and organic to get all the communities into a united government (UniGov for short). Every community donates a certain amount to a military, and the military’s only job is to quash anyone from any community who tries to invade another.

Next he addresses externalities. For example, if some communities emit a lot of carbon, and that causes global warming which threatens to destroy other communities, UniGov puts a stop to that. If the offending communities refuse to stop emitting carbon, then there’s that military again.

The third thing he does is prevent memetic contamination. If one community wants to avoid all media that objectifies women, then no other community is allowed to broadcast women-objectifying media at it. If a community wants to live an anarcho-primitivist lifestyle, nobody else is allowed to import TVs. Every community decides *exactly* how much informational contact it wants to have with the rest of the continent, and no one is allowed to force them to have more than that.

But the wizard and UniGov’s most important task is to think of the children.

Imagine you’re conservative Christians, and you’re tired of this secular godless world, so you go off with your conservative Christian friends to found a conservative Christian community. You all pray together and stuff and are really happy. Then you have a daughter. Turns out she’s atheist and lesbian. What now?

Well, it might be that your kid would be much happier at the lesbian separatist community the next island over. The *absolute minimum* the united government can do is enforce freedom of movement. That is, the *second* your daughter decides she doesn't want to be in Christiantopia anymore, she goes to a UniGov embassy nearby and asks for a ticket out, which they give her, free of charge. She gets airlifted to Lesbiantopia the next day. If *anyone* in Christiantopia tries to prevent her from reaching that embassy, or threatens her family if she leaves, or expresses the *slightest* amount of coercion to keep her around, UniGov burns their city and salts their field.

But this is not nearly enough to fully solve the child problem. A child who is abused may be too young to know that escape is an option, or may be brainwashed into thinking they are evil, or guilted into believing they are betraying their families to opt out. And although there is no perfect, elegant solution here, the practical solution is that UniGov enforces some pretty strict laws on child-rearing, and every child, no matter what other education they receive, also has to receive a class taught by a UniGov representative in which they learn about the other communities in the Archipelago, receive a basic non-brainwashed view of the world, and are given directions to their nearest UniGov representative who they can give their opt-out request to.

The list of communities they are informed about always starts with the capital, ruled by UniGov itself and considered an inoffensive, neutral option for people who don't want anywhere in particular. And it always ends with a reminder that if they can gather enough support, UniGov will provide them with a galleon to go out and found their own community in hitherto uninhabited lands.

There's one more problem UniGov has to deal with: malicious inter-community transfer. Suppose that there is some community which puts extreme effort into educating its children, an education which it supports through heavy taxation. New parents move to this community, reap the benefits, and then when their children grow up they move back to their previous community so they don't have to pay the taxes to educate anyone else. The communities themselves prevent some of this by immigration restrictions – anyone who's clearly taking advantage of them isn't allowed in (except in the capital, which has an official committment to let in anyone who wants). But that still leaves the example of people maliciously leaving a high-tax community once they've got theirs. I imagine this is a big deal in Archipelago politics, but that in practice UniGov asks these people, even in their new homes, to pay higher tax rates to subsidize their old community. Or since that could be morally objectionable (imagine the lesbian separatist having to pay taxes to Christiantopia which oppressed her), maybe they pay the excess taxes to UniGov itself, just as a way of disincentivizing malicious movement.

Because there *are* UniGov taxes, and most people are happy to pay them. In my fantasy, UniGov isn't an enemy, where the Christians view it as this evil atheist conglomerate trying to steal their kids away from them and the capitalists view it as this evil socialist conglomerate trying to enforce high taxes. The Christians, the capitalists, and everyone else are extraordinarily *patriotic* about being part of the Archipelago, for its full name is the Archipelago of Civilized Communities, it is the standard-bearer of civilization against the barbaric outside world, and it is precisely the institution that allows them to maintain their distinctiveness in the face of what would otherwise be irresistable pressure to conform. Atheistopia is the enemy of Christiantopia, but only in the same way the Democratic Party is the enemy of the Republican Party – two groups within the same community who may have different ideas but who consider themselves part of the same broader whole, fundamentally allies under a banner of which both are proud.

IV.

Robert Nozick once proposed a similar idea as a libertarian utopia, and it’s easy to see why. UniGov does very very little. Other than the part with children and the part with evening out taxation regimes, it just sits around preventing communities from using force against each other. That makes it very very easy for anyone who wants freedom to start a community that grants them the kind of freedom they want – or, more likely, to just start a community organized on purely libertarian principles. The United Government of Archipelago is the perfect minarchist night watchman state, and any additions you make over that are chosen by your own free will.

But other people could view the same plan as a conservative utopia. Conservativism, when it’s not just Libertarianism Lite, is about building strong cohesive communities of relatively similar people united around common values. Archipelago is obviously built to make this as easy as possible, and it’s hard to imagine that there wouldn’t pop up a bunch of communities built around the idea of Decent Small-Town God-Fearing People where everyone has white picket fences and goes to the same church and nobody has to lock their doors at night (so basically Utah; I feel like this is one of the rare cases where the US’ mostly-in-name-only Archipelagoness really asserts itself). People who didn’t fit in could go to a Community Of People Who Don’t Fit In and would have no need to nor right to complain, and no one would have to deal with Those Durned Bureaucrats In Washington telling them what to do.

But to me, this seems like a liberal utopia, even a leftist utopia, for three reasons.

The first reason is that it extends the basic principle of liberalism – solve differences of opinion by letting everyone do their own thing according to their own values, then celebrate the diversity this produces. I like homosexuality, you don't, fine, I can be homosexual and you don't have to, and having both gay and straight people living side by side enriches society. This just takes the whole thing one meta-level up – I want to live in a very sexually liberated community, you want to live in a community where sex is treated purely as a sacred act for the purpose of procreation, fine, I can live in the community I want and you can live in the community you want, and having both sexually-liberated and sexually-pure communities living side by side enriches society. It is pretty much saying that the solution to any perceived problems of liberalism is *much more liberalism*.

The second reason is quite similar to the conservative reason. A lot of liberals have some pretty strong demands about the sorts of things they want society to do. I was recently talking to Ozy about a group who believe that society billing thin people is fatphobic, and that everyone needs to admit obese people can be just as attractive and date more of them, and that anyone who preferentially dates thinner people is Problematic. They also want people to stop talking about nutrition and exercise publicly. I sympathize with these people, especially having recently read a study showing that obese people are much happier when surrounded by other obese, rather than skinny people. But realistically, their movement will fail, and even philosophically, I'm not sure how to determine if they have the right to demand what they are demanding or what that question means. Their best bet is to found a community on these kinds of principles and only invite people who already share their preferences and aesthetics going in.

The third reason is the reason I specifically draw leftism in here. Liberalism, and to a much greater degree leftism, are marked by the emphasis they place on oppression. They're particularly marked by an emphasis on oppression being a really hard problem, and one that is structurally inherent to a certain society. They are marked by a moderate amount of despair that this oppression can ever be rooted out.

And I think a pretty strong response to this is making sure everyone is able to say "Hey, you better not oppress us, because if you do, we can pack up and go somewhere else."

Like if you want to protest that this is unfair, that people shouldn't be forced to leave their homes because of oppression, fine, fair enough. But given that oppression *is* going on, and you haven't been able to fix it, giving people the *choice* to get away from it seems like a pretty big win. I am reminded of the many Jews who moved from Eastern Europe to America, the many blacks who moved from the southern US to the northern US or Canada, and the many gays who make it out of extremely homophobic areas to friendlier large cities. One could even make a metaphor, I think rightly, to telling battered women that they are allowed to leave their husbands, telling them they're not forced to stay in a relationship that they consider abusive, and making sure that there are shelters available to receive them.

If any person who feels oppressed can leave whenever they like, to the point of being provided a free plane ticket by the government, how long can oppression go on before the oppressors give up and say "Yeah, guess we need someone to work at these factories now that all our workers have gone to the communally-owned factory down the road, we should probably at least let people unionize or something so they will tolerate us"?

A commenter in the latest Asch thread mentioned an interesting quote by Frederick Douglass:

> The American people have always been anxious to know what they shall do with us [black people]. I have had but one answer from the beginning. Do nothing with us! Your doing with us has already played the mischief with us. Do nothing with us!

It sounds like, if Frederick Douglass had the opportunity to go to some other community, or even found a black ex-slave community, no racists allowed, he probably would have taken it [edit: or not, or had strict conditions]. If the people in slavery during his own time period had had the chance to leave their plantations for that community, I bet they would have taken it too. And if you believe there are still people today whose relationship with society are similar in kind, if not in degree, to that of a plantation slave, you should be pretty enthusiastic about the ability of exit rights and free association to disrupt those oppressive relationships.

V.

We lack Archipelago's big advantage – a vast frontier of unsettled land.

Which is not to say that people don't form communes. They do. Some people even have really clever ideas along these lines, like the seasteaders. But the United States isn't going to become Archipelago any time soon.

There’s another problem too, which I describe in my Anti-Reactionary FAQ. Discussing ‘exit rights’, I say:

> Exit rights are a great idea and of course having them is better than not having them. But I have yet to hear Reactionaries who cite them as a panacea explain in detail what exit rights we need beyond those we have already.
>
> The United States allows its citizens to leave the country by buying a relatively cheap passport and go anywhere that will take them in, with the exception of a few arch-enemies like Cuba – and those exceptions are laughably easy to evade. It allows them to hold dual citizenship with various foreign powers. It even allows them to renounce their American citizenship entirely and become sole citizens of any foreign power that will accept them.
>
> Few Americans take advantage of this opportunity in any but the most limited ways. When they do move abroad, it’s usually for business or family reasons, rather than a rational decision to move to a different country with policies more to their liking. There are constant threats by dissatisfied Americans to move to Canada, and one in a thousand even carry through with them, but the general situation seems to be that America has a very large neighbor that speaks the same language, and has an equally developed economy, and has policies that many Americans prefer to their own country’s, and isn’t too hard to move to, and almost no one takes advantage of this opportunity. Nor do I see many people, even among the rich, moving to Singapore or Dubai.

> Heck, the US has fifty states. Moving from one to another is as easy as getting in a car, driving there, and renting a room, and although the federal government limits exactly how different their policies can be you better believe that there are very important differences in areas like taxes, business climate, education, crime, gun control, and many more. Yet aside from the fascinating but small-scale Free State Project there's little politically-motivated interstate movement, nor do states seem to have been motivated to converge on their policies or be less ideologically driven.
>
> What if we held an exit rights party, and nobody came?
>
> Even aside from the international problems of gaining citizenship, dealing with a language barrier, and adapting to a new culture, people are just rooted – property, friends, family, jobs. The end result is that the only people who can leave their countries behind are very poor refugees with nothing to lose, and very rich jet-setters. The former aren't very attractive customers, and the latter have all their money in tax shelters anyway.
>
> So although the idea of being able to choose your country like a savvy consumer appeals to me, just saying "exit rights!" isn't going to make it happen, and I haven't heard any more elaborate plans.

I guess I still feel that way. So although Archipelago is an interesting exercise in political science, a sort of pure case we can compare ourselves to, it doesn't look like a practical solution for real problems.

On the other hand, I do think it's worth becoming more Archipelagian on the margin rather than less so, and that there are good ways to do it.

One of the things that started this whole line of thought was an argument on Facebook about a very conservative Christian law school trying to open up in Canada. They had lots of rules like how their students couldn’t have sex before marriage and stuff like that. The Canadian province they were in was trying to deny them accreditation, because conservative Christians are icky. I think the exact arguments being used were that it was homophobic, because the conservative Christians there would probably frown on married gays and therefore gays couldn’t have sex at all. Therefore, the law school shouldn’t be allowed to exist. There were other arguments of about this caliber, but they all seemed to boil down to “conservative Christians are icky”.

This very much annoyed me. Yes, conservative Christians are icky. And they should be allowed to form completely voluntary communities of icky people that enforce icky cultural norms and an insular society promoting ickiness, just like everyone else. If non-conservative-Christians don’t like what they’re doing, they should *not go to that law school*. Instead they can go to one of the dozens of other law schools that conform to their own philosophies. And if gays want a law school even friendlier to them than the average Canadian law school, they should be allowed to create some law school that only accepts gays and bans homophobes and teaches lots of courses on gay marriage law all the time.

Another person on the Facebook thread complained that this line of arguments leads to being okay with white separatists. And so it does. Fine. I think white separatists have *exactly* the right position about where the sort of white people who want to be white separatists should be relative to everyone else – separate. I am not sure what you think you are gaining by demanding that white separatists live in communities with a lot of black people in them, but I bet the black people in those communities aren't thanking you. Why would they want a white separatist as a neighbor? Why should they have to have one?

If people want to go do their own thing in a way that harms no one else, you *let* them. That's the Archipelagian way.

(someone will protest that Archipelagian voluntary freedom of association or disassociation could, in cases of enough racial prejudice, lead to segregation, and that segregation didn't work. Indeed it didn't. But I feel like a version of segregation in which black people actually had the legally mandated right to get away from white people and remain completely unmolested by them – and where a white-controlled government wasn't in charge of divvying up resources between white and black communities – would have worked a lot better than the segregation we actually had. The segregation we actually *had* was one in which white and black communities were separate until white people wanted something from black people, at which case they waltzed in and took it. If communities were actually totally separate, government and everything, by definition it would be impossible for one to oppress the other. The black community might start with less, but that could be solved by some kind of reparations. The Archipelagian way of dealing with this issue would be for white separatists to have separate white communities, black separatists to have separate black communities, integrationists to have integrated communities, resdistributive taxation from wealthier communities going into less wealthy ones, and a strong central government ruthlessly enforcing laws against any community trying to hurt another. I don't think there's a single black person in the segregation-era South who wouldn’t have taken that deal, and any black person who thinks the effect of whites on their community today is net negative should be pretty interested as well.)

This is one reason I find people who hate seasteads so distasteful. I mean, here’s what Reuters has to say about seasteading:

> Fringe movements, of course, rarely cast themselves as obviously fringe. Racist, anti-civil rights forces cloaked themselves in the benign language of “state’s rights”. Anti-gay religious entities adopted the glossy, positive imagery of “family values”. Similarly, though many Libertarians embrace a pseudo-patriotic apple pie nostalgia, behind this façade is a very un-American, sinister vision.
>
> Sure, most libertarians may not want to do away entirely with the idea of government or, for that matter, government-protected rights and civil liberties. But many do — and ironically vie for political power in a nation they ultimately want to destroy. Even the right-wing pundit Ann Coulter mocked the paradox of Libertarian candidates: “Get rid of government — but first, make me president!” Libertarians sowed the seeds of anti-government discontent, which is on the rise, and now want to harvest that discontent for a very radical, anti-America agenda. The image of libertarians living off-shore in their lawless private nation-states is just a postcard of the future they hope to build on land.
>
> Strangely, the libertarian agenda has largely escaped scrutiny, at least compared to that of social conservatives. The fact that the political class is locked in debate about whether Michele Bachmann or Rick Perry is more socially conservative only creates a

> veneer of mainstream legitimacy for the likes of Ron Paul, whose libertarianism may be even more extreme and dangerously un-patriotic. With any luck America will recognize anti-government extremism for what it is — before libertarians throw America overboard and render us all castaways.

Keep in mind this is because *some people want to go off and do their own thing in the middle of the ocean far away from everyone else without bothering anyone*. And the newspapers are trying to whip up a panic about “throwing America overboard”.

So one way we could become more Archipelagian is just *trying not to yell at people who are trying to go off and doing their own thing quietly with a group of voluntarily consenting friends*.

But I think a better candidate for how to build a more Archipelagian world is to encourage the fracture of society into subcultures.

Like, transsexuals may not be able to go to a transsexual island somewhere and build Transtopia where anyone who misgenders anyone else gets thrown into a volcano. But of the transsexuals I know, a lot of them have lots of transsexual friends, their cissexual friends are all up-to-date on trans issues and don’t do a lot of misgendering, and they have great social networks where they share information about what businesses and doctors are or aren’t trans-friendly. They can take advantage of trigger warnings to make sure they expose themselves to only the sources that fit the values of their community, the information that would get broadcast if it was a normal community that could impose media norms. As Internet interaction starts to replace real-life interaction (and I think for a lot of people the majority of their social life is already on the Internet, and for some the majority of their economic life is as well) it becomes increasingly easy to limit yourself to transsexual-friendly spaces that keep bad people away.

The rationalist community is another good example. If I wanted, I could move to the Bay Area tomorrow and never have more than a tiny amount of contact with non-rationalists again. I could have rationalist roommates, live in a rationalist group house, try to date only other rationalists, try to get a job with a rationalist nonprofit like CFAR or a rationalist company like Quixey, and never have to deal with the benighted and depressing non-rationalist world again. Even without moving to the Bay Area, it’s been pretty easy for me to keep a lot of my social life, both on- and off- line, rationalist-focused, and I don’t regret this at all.

I don’t know if the future will be virtual reality. I expect the post-singularity future will include something like VR, although that might be like describing teleportation as “basically a sort of pack animal”. But how much the immediate pre-singularity world will make use of virtual reality, I don’t know.

But I bet if it doesn’t, it will be because virtual reality has been circumvented by things like social networks, bitcoin, and Mechanical Turk, which make it possible to do most of your interaction through the Internet even though you’re not literally plugged into it.

And that seems to me like a pretty good start in creating an Archipelago. I already hang out with various Finns and Brits and Aussies a lot more closely than I do my next-door neighbors, and if we start using litecoin and someone else starts using dogecoin then I’ll be more economically connected to them too. The degree to which I encounter certain objectifying or unvirtuous or triggering media already depends more on the moderation policies of Less Wrong and Slate Star Codex and who I block from my Facebook feed, than it does any laws about censorship of US media.

At what point are national governments rendered mostly irrelevant compared to the norms and rules of the groups of which we are voluntary members?

I don’t know, but I kind of look forward to finding out. It seems like a great way to start searching for utopia, or at least getting some people away from their metaphorical abusive-husbands.

And the other thing is that I have pretty strong opinions on which communities are better than others. Some communities were founded by toxic people for ganging up with other toxic people to celebrate and magnify their toxicity, and these (surprise, surprise) tend to be toxic. Others were formed by very careful, easily-harmed people trying to exclude everyone who could harm them, and these tend to be pretty safe albeit sometimes overbearing. Other people hit some kind of sweet spot that makes friendly people want to come in and angry people want to stay out, or just do a really good job choosing friends.

But I think the end result is that the closer you come to true freedom of association, the closer you get to a world where everyone is a member of more or less the community they deserve. That would be a pretty unprecedented bit of progress.

Map-Territory Distinctions

When I moved here last year, I got a big trash can to put outside for garbage pickup.

The trash can only got emptied about one in every two garbage days. I called the city’s Department Of That Sort Of Thing to ask why it kept getting passed over, and they told me that my trash can was oversized, bigger than the regulations allowed, and so any garbage-people who noticed this would refuse to empty it.

I decided to get rid of the trash can so I could buy a new, correctly-sized version.

Readers, you have never known frustration until you try to get garbage-people to take a trash can.

Like, technically we have a bulk trash pickup day. And the trash can is well under the limit for bulk items that are allowed. So I leave it out on the curb. And about one in every two garbage days, I find that any dust or leaves that have collected in it have been dutifully emptied. One in two days, it hasn’t been emptied, because it’s oversized and against regulations. In either case, the trash can is back on the curb where I left it.

Last week I asked Ozy to make a big sign saying “THIS TRASH CAN IS THE TERRITORY, NOT THE MAP” and stick it on the can. Ozy is a very practical person and I am pretty sure ze wrote something else. Ze stuck the sign to the lid of the trash can. When we woke up, we found the trash can with the lid open, as if someone had checked inside to see if there was any trash. The sign was still on the top of the lid, now hidden against the side of the can.

Next week I think *I* will write the message. I was thinking something very simple, like “TAKE THIS, THIS IS THE TRASH”. But then the garbage-people would probably just take *the sign*.

I need something subtler, something that jars them out of their everyday reality to a deeper level of understanding. I’m thinking something like “THIS TRASH CAN IS THE MOON, NOT THE FINGER POINTING TO IT” or just “CECI N’EST PAS UNE POUBELLE”.

The great work continues.

Minutes From The Michigan Rationalist Meetup

0:58 – Discussion of unschooling. Two of the attendees were unschooled, and their opinions pretty much paralleled where the discussion went here – kids who are smart enough will learn stuff whether or not anyone is sending them to school; kids who aren't may or may not, but they seem able to pick up what they need when they need it and no one ends up unable to read or anything like that (though some people take a very long time). The key phrase seems to be "steeper learning curve" – unschooled kids might wait until age 12 to pick up something schooled kids learn at 8, but when they want to they pick it up quickly and effectively. Some people never learn algebra, but then again if you asked me to do complicated things with polynomials I could only give you about 50-50 odds.

1:50 – Debate on cardinal vs. ordinal utility. Sure, ordinal utility is much more responsible and satisfies obscure technical criteria like "isn't totally made up". But on the other hand, not only would I prefer curing cancer and winning a Nobel Prize for it to getting one potato chip, but I would prefer it *a lot*. In fact, I would venture to say that in some important sense, the cancer cure + Nobel would be *more than twice as good* as the potato chip. Any theory of utility that doesn't allow me to say this, which just says "Well, the cure + Nobel is better, but it's meaningless to ask if it's a lot better or only slightly better" doesn't even come close to reflecting real human experience and is woefully inadequate.

1:59 – Mild concern over what if someone had a preference for having a round number of utils. Would such a preference be self-consistent? Would the universe stop?

2:41 – Discussion on how much should we respect the preferences of the dead. Aaron, channeling Robin Hanson, points out that exponential discounting means we consider preferences of people in the future much less important than those of people today – otherwise we would invest all our money to donate the interest to future folk. But this implies that people in the past had *more* important preferences. If a Neanderthal had preferences that everyone should make sacrifices to the fire god, should we sacrifice to the fire god more?

2:58 – Confusion over whether the wrong time preferences result in everyone eating grapefruit all the time. It was brought up that someone with infinite discounting ie total inability to identify with their future self would blow all their savings on drugs and prostitutes and grapefruit. We all figured grapefruit was some kind of metaphorical symbol of hedonism, and we discussed this for a while until finally I asked “Wait, why grapefruit?” and we realized that all of us had just kind of assumed the others were talking about it for some good reason. Turned out the first person had said “drugs, prostitutes and great food”, we had universally heard “grapefruit”, and just kind of taken it from there.

3:30 – Mild crisis. If we can’t decide on a topic, so we put it to a vote, but we can’t determine what voting system to use, so we put it to a vote, but we can’t determine what voting system to use, and so on, is there any way to end the infinite regress? There was a serious proposal to use Archipelago-logic and break up into subgroups with their own voting system, but the problem was eventually solved by me declaring myself dictator and going from there. I feel like this is a metaphor for something.

CONSTANT VIGILANCE!

I am going to do something very dangerous today, something that makes me acutely aware of my own mortality. I am going to disagree with Robin Hanson.

In my defense, he wrote an entire blog post called [Don’t Be ‘Rationalist’]().

The key section:

> This blog is called “Overcoming Bias,” and many of you readers consider yourselves “rationalists,” i.e., folks who try harder than usual to overcome your biases. But even if you want to devote yourself to being more honest and accurate, and to avoiding bias, there’s a good reason for you not to present yourself as a “rationalist” in general. The reason is this: you must allocate a very limited budget of rationality.
>
> It seems obvious to me that almost no humans are able to force themselves to see honestly and without substantial bias on all topics. Even for the best of us, the biasing forces in and around us are often much stronger than our will to avoid bias. So we must choose our battles, i.e., we must choose where to focus our efforts to attend carefully to avoiding possible biases.

Robin is choosing to treat rationality as a limited resource that must be budgeted, though he doesn’t explain why.

A commenter on his blog, Silent Cal, asks him: how do you know rationality isn’t more like weightlifting? The more weights you lift, the stronger you get. If you want to become strong, it’s a good idea to “blow” your weightlifting “budget” on as many useless tasks as possible, as often as possible.

Robin says that rationality may be both like weightlifting *and* like money. Money also has a way to “buy” future strength in that if you invest your money today, you will get more of it tomorrow. But, he says, even when you’ve made a fortune off of good investments, you still have a limited budget. Or, in the weight-lifting metaphor, no matter how strong you are, if you’re going to be asked to lift a weight at the very limit of your strength at 5, don’t exhaust yourself at 4:30.

But I propose a third metaphor: rationality neither as budget planning nor weight training, but as habit cultivation.

Wait, no. Not a metaphor. Whatever the opposite of a metaphor is. An actual thing.

During my abortive attempt at learning aikido, my instructor taught me some basic rules of efficient bodily movement and posture and suggested that I follow the rules not only in aikido classes but all throughout my daily life. The reasoning was: right now I have certain unconscious postural habits that I adopt without thinking about them whenever I need to sit, or open a door, or whatever. If an aikido instructor is standing next to me, telling me exactly what’s wrong with each of them, I can probably figure out what he’s talking about and correct them. But when I leave aikido class, I’m probably going to relapse into my normal habits, since after all I’ve been doing them for almost thirty years now and they require less effort. The fact that I relapse into my normal habits outside aikido class probably means I’ll also relapse into my normal habits in a stressful situation like a fight, when I really need to be thinking about other things besides posture. But if I can train myself to use proper aikido styles of movement even when I’m doing something stupid like opening a door, my body will become so used to them that they will be the style I default to when my mind is otherwise occupied trying to deal with the guy swinging a broken beer bottle at me. Or, even if I *am* thinking "aikido aikido aikido aikido" at that point, which I might well be, I will be thinking about the complicated impressive things that supervene on the basic movements, not having to worry about the basic movements themselves.

This whole affair started when someone asked "What if rationality were kind of like a martial art?" And part of the answer to that question would be: you had better make the fundamentals of it perfectly, entirely, down-to-the-bone natural.

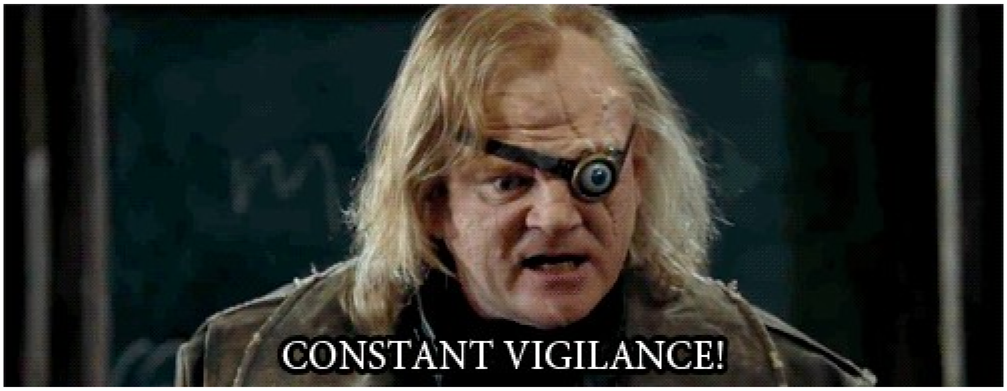
But I have an even more relevant example of habit cultivation.

Lucid dreamers offer some techniques for realizing you're in a dream, and suggest you practice them even when you are awake, *especially* when you are awake. The goal is to make them so natural that you could (and literally will) do them in your sleep. I can attest that this works. You cultivate the habit of worrying about whether you're dreaming, and then once the habit is established it continues even in your dreams, at which point it becomes useful.

And I think this is a good metaphor for rationality because it's about holding on to consciousness.

The problem with dreaming is that it depresses your natural ability to wonder if you are dreaming. Like, you're being chased around Venice by a giant dragon with the head of your second-grade teacher, and you're wondering a lot about whether you should make a swim for it in one of the canals, and not at all about whether this *maybe, might be* a dream. That's why it's so important to cultivate the habit of worrying about it, because habits can survive even in states of weakened consciousness where the logical thought that tells you it's now time to dream-check can't.

And the problem with irrationality is that it depresses your natural ability to wonder if you are being irrational. One of the fundamental skills of rationality is *noticing that you are confused*, which also happens to be one of the fundamental skills of dealing with being chased around Venice by a giant dragon with the head of your second-grade teacher. Things like "I should notice when it is a time when I should worry about whether I should notice I am confused or not" just collapse back down to "I should notice when I am confused". This is not something you can do by conscious thought in the grips of the same weakened consciousness that is causing the problem to exist in the first place. Your only option is the same one the lucid dreamers use: cultivate the habit of *always* doing the right thing, and hoping maybe that habit will be there when you need it.


CONSTANT VIGILANCE!

List Of The Passages I Highlighted In My Copy Of Jonathan Haidt’s “The Righteous Mind”

From the Department of Being Very Complete In Listing Possible Explanations:

> It turns out that witchcraft beliefs arise in surprisingly similar forms in many parts of the world, which suggests either that there really are witches or (more likely) that there’s something about human minds that often generates this cultural institution. The Azande believed that witches were just as likely to be men as women, and the fear of being called a witch made the Azande careful not to make their neighbors angry or envious. That was my first hint that groups create supernatural beings not to explain the universe but to order their societies.

Seems related to some of my recent thoughts on unschooling:

> [Piaget] also found that it’s pointless for adults to explain the conservation of volume to kids. The kids won’t get it until they reach an age (and cognitive stage) when their minds are ready for it. And when they are ready, they’ll figure it out for themselves just by playing with cups of water.

If this is true, it seems very very important and people should be trying harder to exploit it:

> [Todorov] collected photographs of the winners and runners-up in hundreds of elections for the U.S. Senate and the House of Representatives. He showed people

> the pairs of photographs from each contest with no information about political party, and he asked them to pick which person seemed more competent. He found that the candidate that people judged more competent was the one who actually won the race about two-thirds of the time.

On what separates us from the apes:

> Michael Tomasello, one of the world's foremost experts on chimpanzee cognition, [said] "It is inconceivable that you would ever see two chimpanzees carrying a log together."

On what doesn't separate us from the apes:

> Acheulean tools [a style used by hominids 1.8 million years ago] are nearly identical everywhere, from Africa to Europe to Asia, for more than a million years. *There's hardly any variation*, which suggests that the knowledge of how to make these tools may not have been passed on culturally. Rather, the knowledge of how to make these tools may have become innate, just as the "knowledge" of how to build a dam is innate in beavers.

One of those "Huh, I guess thousands of very intelligent people throughout all human history *haven't* been doing something useless for no reason" moments:

> In September 1941, William McNeill was drafted in the US Army. He spent several months in basic training, which consisted mostly of marching around the drill field in close formation with a few dozen other men. At first McNeill thought the marching was just a way to pass the time, because his base had no weapons with which to train. But after a few weeks, when his unit began to synchronize well, he began to experience an altered

> state of consciousness. “Words are inadequate to describe the emotion aroused by the prolonged movement in unison that drilling involved. A sense of pervasive well-being is what I recall; more specifically, a strange sense of personal enlargement; a sort of swelling out, becoming bigger than life, thanks to participation in collective ritual.” McNeill fought in World War II and later became a distinguished historian. His research led him to the conclusion that the key innovation of Greek, Roman, and later European armies was the sort of synchronous drilling and marching the army had forced him to do years before.

Would-be commune founders, take note:

> The anthropologist Richard Sosis examined the history of two hundred communes founded in the United States in the nineteenth century. Which kind of commune survived longest? Sosis found that the difference was stark: just 6% of the secular communes were still functioning twenty years after their founding, compared to 39% of religious communes. He found one master variable: the number of costly sacrifices that each commune demanded from its members. It was things like giving up alcohol and tobacco, fasting for days at a time, conforming to a communal dress code or hairstyle, or cutting ties with outsiders. For religious communes, the effect was perfectly linear: the more sacrificce a commune demanded, the longer it lasted. But Sosis was surprised to discover that demands for sacrifice did not help secular communes. Most of them failed within eight years, and there was no correlation between sacrifice and longevity.

This is a very interesting application of the trust game, but I feel like I’ve seen some experiments that show the opposite as well:

> A team of German economists asked subjects to play a game in which one person is the “truster”, who is given some money on each round of the game. The truster is then asked to decide how much money, if any, to pass on to an anonymous “trustee”. Any money passed gets tripled by the experimenter, at which point the “trustee” can choose how much, if any, to return to the truster. Behavioral economists use this game often, but the novel twist in this study was to reveal one piece of real, true personal information about the trustees to the trusters. In some cases, the truster learned the trustee’s level of religiosity, on a scale of 1 to 5. When trusters learned that their trustee was religious, they transferred more money. More important, the religious trustees really did transfer back more money than did the nonreligious trustees, even though they never knew anything about their trusters. The highest levels of wealth, therefore, would be created when religious people get to play a trust game with other religious people.

Another “Oh, *that* explains it” moment:

> Even today, markets that require a very high trust to function efficiently are often dominated by religiously bound ethnic groups (such as ultra-Orthodox Jews in the diamond market) who have lower transaction and monitoring costs than their secular competitors.

Very curious if same effect in “Sunday Assembly” style groups:

> Why are religious people better neighbors and citizens? To find out, Putnam and Campbell included on one of

> their surveys a long list of questions about religious beliefs and practices. These beliefs and practices turned out to matter very little. The only thing that was reliably and powerfully correlated with the moral benefits of religion was how enmeshed people were in relationships with their co-religionists…Putnam and Campbell's work shows that religion in the United States nowadays generates such vast surpluses of social capital that much of it spills over and benefits outsiders.

We keep losing these intellectual Turing tests, and I'm pretty sure I can't blame this one on Gilbert being an outlier who should not be counted:

> We tested how well liberals and conservatives could understand each other. We asked more than two thousand Americans to fill out the Moral Foundations Questionnaire. One third of the time they were asked to fill it out normally, answering as themselves. One third of the time they were asked to fill it out as they think a "typical liberal" would respond. One third of the time they were asked to fill it out as a "typical conservative" would respond. This design allowed us to examine the stereotypes that each side held about the other. More important, it allowed us to assess how accurate they were. Who was best able to pretend to be the other? The results were clear and consistent. Moderates and conservatives were most accurate in their predictions, whether they were pretending to be liberals or conservatives. Liberals were the least accurate, especially those who described themselves as "very liberal".

Where do libertarians fall on personality measures? No surprises here:

> We found that libertarians tend to look more like liberals than like conservaatives on most measures of personality. For example, both groups score higher than conservatives on openness to experience, and lower on disgust sensitivity and conscientiousness. Where libertarians diverge from liberals most sharply is on … the Care foundation, where they score very low (even lower than conservatives).

Another one from the Department of Insight Porn:

> We learned that much of the increase in political polarization was unavoidable. It was the natural result of the political realignment that took place after President Lyndon Johnson signed the Civil Rights Act in 1964. The conservative southern states, which had been solidly Democratic since the Civil War (since Lincoln was a Republican) then began to leave the Democratic Party, and by the 1990s the South was solidly Republican. Before this realignment there had been liberals and conservatives in both parties, which made it easy to form bipartisan teams who could work together.

From the Department of Unexpected Consequences:

> But we also learned about factors that might possibly be reversed. The most poignant moment of the conference came when Jim Leach, a former Republican congressman from Iowa, described the changes that began in 1995. Newt Gingrich, the new speaker of the House of Representatives, encouraged the large group of incoming Republicans to leave their families in their home districts rather than moving their spouse and children to Washington. Before 1995, Congressmen from both parties attended many of the same social events on weekends; their spouses became friends; their

> children played on the same sports teams. But nowadays most Congressmen fly to Washington on Monday night, huddle with their teammates and do battle for three days, and then fly home on Thursday night. Cross-party friendships are disappearing; Manichaeism and scorched Earth politics are increasing.

From the Archipelago Establishment Working Group:

> Our counties and towns are becoming increasingly segregated into "lifestyle enclaves", in which ways of voting, eating, working, and worshipping are increasingly aligned.

Living By The Sword

[Trigger warnings: sexual abuse, scary Internet mobbing.]
[Try not to link to this everywhere because I feel bad about discussing it and don’t want to spread it to more of the Internet than it has to be or make this the sort of thing people associate this blog with.]

I.

A couple of months ago this blog got into a fight about false rape statistics. There has recently been some amount of – let’s call it “aftermath” – to that incident which I found interesting.

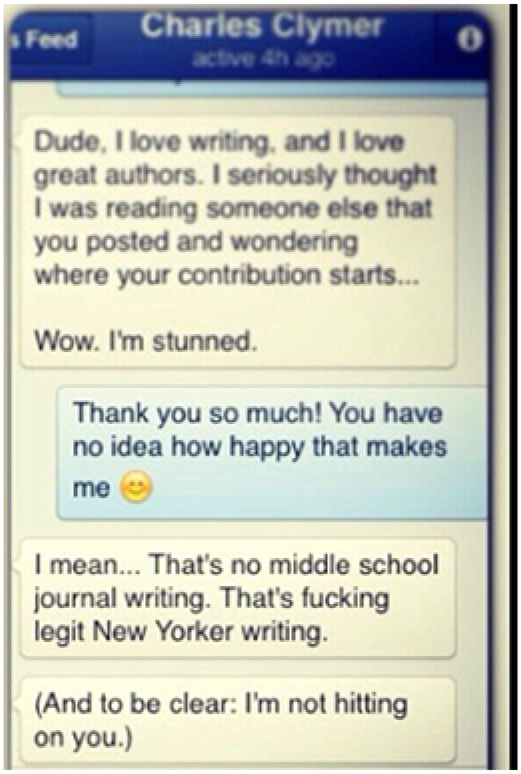
Some context: Mr. Clymer is a man who is (was?) active in the feminist movement. He wrote a popular article saying that false rape accusations were rarer than comet strikes, which the feminist movement dutifully liked and reblogged thirty thousand times. I wrote a blog post pointing out that his statistics were what we in the business sometimes technically refer to as “off by four orders of magnitude”. I may have been *slightly* cross about this.

A couple of weeks ago, Mr. Clymer was in the news again. More accurately, he was trending on Twitter, where a very popular #StopClymer hashtag apparently indicated he had done something that needed to be stopped. The best description of the whole incident seems to be the one here.

As far as I can tell, Clymer said on Twitter that Jesus would have been pro-feminist and that religion needs to try harder to be pro-choice and pro-feminist. Someone interpreted this as “using a women-centered hashtag to promote religious beliefs” and decided to see what else she could find on him. You may read the page and decide for yourself, but as far as I can tell, what she found was:

– He’s around the 75th percentile for Internet jerkitude and some of the people he was jerks to were women
– He banned people from his Facebook page, reasons unclear, and some of them were women
– He said he was in the feminist movement “80 percent for the cause” but that he also felt good when things he wrote became popular

This was sufficient to paint a big target on his head. We then proceeded to the ritual of Dredge Up Everything He Has Ever Written And Interpret It As Further Evidence He Is Terrible. For example, he once complimented a fan of his and added “just to be clear, I’m not hitting on you” at the bottom:



Dude, I love writing, and I love great authors. I seriously thought I was reading someone else that you posted and wondering where your contribution starts...

Wow. I'm stunned.

Thank you so much! You have no idea how happy that makes me 😊

I mean... That's no middle school journal writing. That's fucking legit New Yorker writing.

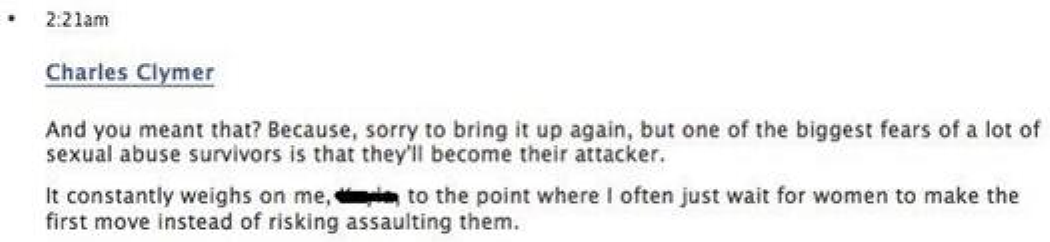
(And to be clear: I'm not hitting on you.)

The response was reasonable and completely proportionate:

> @BhasChat That is some hardcore predator grooming. It's clear why CC was able to fly under the radar so long:

> He's a pro abuser.
>
> — Viridiana (@sariweather) [May 30, 2014]
>
> [@BhasChat] Are you telling me that this mofo was after minors now too? 'Scuse me for a moment… *vomits through fingers*
>
> — lurker (@groovyswkr) [May 30, 2014]
>
> [@BhasChat] Just did a search for him and his profile and he is far too slick. He is a me, me, me merchant. A real predator. Stay safe.
>
> — Andrew Dalby (@ardalby71) [May 30, 2014]

Then it was found that he mentioned that because he was a survivor of rape, he tried to be extra careful because he knew rape survivors were more likely to become rapists themselves. Surely no one could have a problem with…



- 2:21am

Charles Clymer

And you meant that? Because, sorry to bring it up again, but one of the biggest fears of a lot of sexual abuse survivors is that they'll become their attacker.

It constantly weighs on me, [redacted] to the point where I often just wait for women to make the first move instead of risking assaulting them.

> [@Jellis1snv] YES!! ????
>
> — ?James Saint-Franc? (@sirjamesa12) [May 30, 2014]
>
> "[@sirjamesa12]: TRIGGER WARNING: Talks about sexual assault. [#STOPCLYMER] [pic.twitter.com/azMRXoGTPe]" – this dude gets worse by the day
>
> — AspieWomanist Writer (@MochaLisaccino) [May 30, 2014]

That he had once had some kind of complicated disagreement with a black woman that I have trouble getting a good picture of from her description of such:

> Anger is a natural reaction to oppression; anger is useful; anger is the only appropriate reaction to this sort of erasure of racial identity.
>
> The members of the feminist group in which I expressed this anger were, for the most part, combative or unreceptive (I did receive small pockets of support, and for those individuals I am most appreciative). Multiple leaders of the group tried to tone police me, told me that movements becoming inclusive "takes time" and I should just be patient. I was begged by one of the most prominent leaders to educate everyone by writing a "powerful article" rather than express my feelings to the group itself. I was told by this same leader that they would not use the term "white supremacy" because "it was triggering" to some folks.
>
> But failure to call evil by its name only begets more evil.
>
> Another prominent member of the group, and a semi-well known name in the huffpo blogosphere, charles clymer, decided to pipe in to remind me that both white supremacist and anti-racist viewpoints should be valued equally. That's white supremacy in action, right there.

Anyway, whatever the disagreement was, Charles Clymer tried to talk it out with her, but eventually got really annoyed with her tone and asked her to stop messaging him. Needless to say, she described this incident in a reasonable and completely proportionate way:

> Unbeknownst to me, by criticizing charles clymer, I was awakening a psychologically abusive internet monster.

> And when it dissolved into an abusive mess of white tears and white-man-trying-to-force-a-black-woman-to-do-something-she-won’t-do (“yes, massa!”), that’s when I knew I had to block this fucker.

The above conversation was added to the pile of evidence. Then someone noticed he was a war veteran. I’m sure this will be dealt with reasonably and completely proportionally as well?

> #StopClymer bc imperialism is ANTI FEMINIST.Youre a white supremacist& promote western hegemony #Fuckyourmansplaining pic.twitter.com/YxS4rpFkIT
>
> — Carolyn Luby (@Carolyn_luby) June 13, 2014

Bailiff! Next piece of evidence! Accused once wrote an article saying that the words “c*nt” and “b*tch” could not be “reclaimed” and that people should stop using them. Does the prosecution have a reasonable and completely proportionate response to this?

> @awhooker So in brief. His #Malesplanation about the marginalized and their reclamation is just another appropriation? #StopClymer
>
> — Alexander Yarde (@thatalexyarde) May 30, 2014

Let it be added to the record. Prosecution, any reasonable and proportionate closing words?

> #StopClymer a dudebro using his status to talk down/police women & get laid = recipe for disaster & complicity.
>
> — Black Paula Deen (@Capolavoro_) May 30, 2014

> #CharlesClymer will not grant you respect because you are a woman. You have to earn that from him in ways he deems appropriate. #StopClymer
>
> — Kayla Pearce (@KaylaPearce5) May 30, 2014

> Reading @cmclymer's latest tweets makes me sick. He is taking all the critique and adapting behavior, which is what abusers do. #StopClymer
>
> — Suey Park (@suey_park) May 30, 2014

QUICK! HE’S TAKING OUR CRITIQUE TO HEART! STOP HIM BEFORE HE GETS AWAY!

Annnnnnnyway, before the court could render verdict, the whole affair triggered a massive relapse of Clymer’s PTSD and he posted saying his therapist had told him not to go back on the Internet for a very long time. The idea was floated that he was making this up, but finally it was decided that yeah, rape survivor + military veteran + suddenly everyone in the world hates him, eh, maybe he *is* having a PTSD relapse. The announcement was greeted in a suitably compassionate manner that, may I add, was reasonable and completely proportionate:

> To everyone who participated in the #StopClymer movement: THANK YOU. Our work is not done but tonight is a cause for celebration.
>
> — kelly coviello (@KellyCoviello) June 3, 2014

> Gotta say, if a guy's PTSD is triggered by resistance to his misogyny, his PTSD isn't going to be the #1 claim on my sympathies. #StopClymer
>
> — Dan Holzman-Tweed (@HolzmanTweed) June 3, 2014

> Clymer says he's going away briefly. But all of us behind #StopClymer will be here whenever he shows his slimy face.
>
> — Natalie BauerOBrien (@NatalieOGrumpy) June 3, 2014

Far be it from me to defend Mr. Clymer. But it is hard for me to see this happen to someone and not feel at least a little bit of sympathy. I’ll just point out that when I talked to him over Facebook, he was a jerk to me. So let this be the defense he gets from me: he is an equal opportunity jerk. I don’t think he’s misogynist. I don’t think he’s racist. I don’t think he’s ‘a predator’. I think he’s just abrasive, unpleasant, and bad with statistics.

In February, I wrote:

> [He] should be cast out from the community of people who have reasonable discussions and never trusted by anyone again. It might not totally succeed in making a new norm against this kind of thing. But at least it will prevent other people from seeing Clymer’s success, taking heart, and having the number of lies which are socially acceptable gradually advance.

Nevertheless, I cannot be happy with the current turn of events. Mr. Clymer deserved to lose his podium. But he didn’t deserve to lose it for this reason, and he didn’t deserve to be psychologically traumatized in this particular way. Justice is not quite served.

II.

Drew recently mentioned on Facebook his delight that there was a scandal called “Jacobinghazi”. I share his delight and his hope that -ghazi becomes a scandal suffix similar to "-gate". Filesghazi! Climateghazi! Let's do it!

The scandal itself was somewhat less interesting. Jacobin is a leftist magazine. I was recommended it a couple of months ago and have read it inconsistently – they have a good article on drug research I'll be blogging about sometime. Anyway, as part of a broader point about rape they linked to someone on Twitter talking about receiving rape threats. After a multistep game of Telephone, this turned into "linking to rape threats", "digging rape jokes", "mocking rape threats", "using rape threats to belittle people", and finally, you know you expected it, "making rape threats". Anyone who pointed out that this was completely made up and made no sense was accused of "minimizing rape threats" and "silencing voices". But apparently the first side, the one not involved in the original accusations, was doing some pretty bad things too, even though I can't get a clear description of exactly what they were. Needless to say, at every step hundreds of people tweeted about it and called for the heads of everybody involved.

(in Soviet Russia, Jacobins' heads get called for by *YOU!*)

Finally, everyone agreed that everybody on the other side (whichever that might be) was a dudebro, treated women as things, and was super racist, even though as far as I can tell nothing about the incident seemed to involve race at all. Consensus having been achieved, the incident mostly died down.

There is an executive summary here (warning: two pages, click blue button at bottom to get second), a much less executive and more mocking summary here, and a horrible Twitter hashtag here which now mostly seems to be populated by people who are vaguely embarrassed and can’t remember what they were so angry about and *definitely* didn’t participate and it’s not their fault. Newsweek also wrote an article, but everyone on both sides seems to agree they got it horribly wrong.

And what I think these kinds of incidents show is that…

III.

Wait! Sorry! There’s one more set of #StopClymer tweets we haven’t gotten to yet!

> Aand now I'm reading #StopClymer to remind myself how bad this "lookitme I'm a guy who Gets It" thing can go
>
> — Arthur Chu (@arthur_affect) May 31, 2014

!!! But didn’t you…?

> I once took Clymer's side in an argument & was kinda wary abt him even at the time but the guy on the other side was just SO wrong
>
> — Arthur Chu (@arthur_affect) May 31, 2014

Oh no I can’t let you get away with that I am SO not letting you get away with that.

More context: my article condemning Clymer got discussed on Facebook. Arthur weighed in saying I was wrong to criticize Clymer’s article, because it was on the side of feminism, and feminism is good, and sometimes people need to use dirty tricks and exaggerated statistics to fight for good things. He said my problem was that I was too focused on abstract virtue, rather than the gritty reality of needing to do whatever it takes to win a battle, in this case the battle of pushing feminist ideas. Mr. Clymer’s ends justified his means, so I should have let him be.

I wrote a blog post disagreeing with that assessment, Arthur replied in the comments, I replied to him in the comments, and then it died down.

Assuming Arthur’s referring to our discussion – and I think he is, it would be very strange if circumstances had made him defend Mr. Clymer twice – then I don’t think I was, as he puts it, SO WRONG. No one disagreed with my debunking of Mr. Clymer’s statistics; there were over two hundred comments on that post and over five hundred on the followup and not one of them claimed he had gotten his math right. All that was left to argue about was my assertion that what he had done, if deliberate, was not remotely okay.

In my final discussion with Arthur, way down in the comment section of my response post, we discussed our different definitions of tolerance and in-groupishness. I said we should tolerate people who are truthful and kind, regardless of their political opinions. Arthur thought, on the contrary, we should tolerate people with the correct political opinions, regardless of whether they are truthful and kind. My exact words were:

> Both of us successfully trap [actually bad people like] the KKK and Hitler on the outside of our fence. But I get to have a lot more allies than you do – allies against the actually bad people – and you have to put up with some pretty creepy friends. And if your side of the fence wins, you may find that hate is pretty darned transferable and you can’t always ensure it gets directed against the right people.

He answered:

> I am deeply creeped out by many of your friends and while some of my friends might be a little loud and obnoxious, I on the whole admire them.

I don't know if Arthur was genuinely "wary" of Mr. Clymer four months ago. He sure spent a heck of a long time defending him and being *very* upset that other people were criticizing him. But even if Arthur is totally telling the truth, even if he was as wary as a dockyard full of consumer goods, exactly one of us sounded the "this is maybe not the best guy to have representing the feminist movement" alarm. And exactly one of us condemned that person for doing so and said that "while some of my friends might be a little loud and obnoxious, I on the whole admire them."

But I don't know. Maybe, despite the topic of discussion, he wasn't referring to Clymer in particular? Maybe you can admire someone and be wary of them at the same time. Still, I think all of this touches on a much more important question: why don't whales get cancer more often?

I mean, think about it. Cancer results from a series of mutations occurring by chance in a single cell. So over a given time period, the cancer rate of an organism should be proportional to the number of cells in that organism. If a whale is a thousand times bigger than a person, it should have a thousand times more cells and therefore get cancer a thousand times more often. Since humans have maybe a 1% per year cancer risk, whales should get ten cancers a year. But most whales live a long time and don't die of cancer. Why not? It can't just be that they've evolved more efficient anti-cancer mechanisms, or else other animals (who also experience gains in adaptive fitness from not dying of cancer) would have evolved the same [**EDIT**: Carl Shulman suggests reasons why this might not be true].

(This problem is called Peto's Paradox, was first raised in 1975, and is a great example of the rationalist skill of 'noticing confusion'. Anyone who knows anything about cancer had the tools to notice something was *really, really weird* here, and other than this Peto fellow nobody did. I feel appropriately ashamed for somehow going through my entire life up to this point without picking up the glaring weirdness here.)

I don't know which of the various proposed solutions to this puzzle is *true* but the most *hilarious* is no doubt Nagy, Victor and Cropper (2007). Whales are very big, so in order to threaten a whale, a cancer must also grow very big. In order to grow very big, a cancer must evolve a complicated internal structure determining which cells expand where and who's going to secrete the factors necessary for blood vessels to grow and so on. Cancers can do that: even in humans, tumors develop impressive amounts of structure and cooperation among the cancerous cells inside of them. But as tumors grow bigger and more intricate, and cells have to spend more and more time altruistically working for the good of the tumor rather than just reproducing, some cells will inevitably defect from the plan and just divide uncontrollably. This ends up in an unhappy equilibrium. Whenever the balance swings too far toward defectors, the tumor can't support itself and most of the cells die. Whenever the balance swings too far toward cooperators, the defectors have a big advantage again and start expanding. As a result, the tumor either remains at a fixed size or dies off completely. It might get a chance to metastasize, but the same will happen to its metastatic descendents.

In other words, the theory is that whales survive because they are so big that their cancers get cancer and die.

And if true, this wouldn’t be too surprising. Multicellular organisms have put a lot of evolutionary work into getting their cells to cooperate with each other, and part of that includes very strong safeguards again cancer. Your safeguards against cancer are so good that even though you will experience about 10 quadrillion (= ten million billion) cell divisions during your life time, in most people none of those ever turn into a clinically relevant cancer.

Cancerous cells are those that have lost all of those safeguards. As they grow into a functioning tumor, they need to evolve new mechanisms of cooperation, and so they do. But a couple million cells working for a couple months in your body aren’t going to do as well as all the animals in the world over 3 billion years of evolutionary history. So cancer’s cooperation enforcement mechanisms are much much worse than noncancerous cells’ cooperation mechanisms. Which is why you can live for a hundred years and have ten quadrillion cell divisions and not get cancer, but cancer can’t even take over one lousy section of a whale colon before *it* gets meta-cancer.

We like to call things we disagree with “social cancers”. People dislike social cancers and often inveigh against them. Here’s someone who thinks inequality is a social cancer. Here’s another guy who thinks gambling is. Here is a person who thinks that communism is, and though we may not agree with his spittle-filled rants, his web design choices, his choice of facial expression in photographs, or his assertion that “medical orgonomist” is a thing, we cannot doubt his sincerity.

I wonder if a good definition for “social cancer” might be any group that breaks the rules of cooperative behavior that bind society together in order to spread more quickly than it could legitimately achieve, and eventually take over the whole social body.

But society, like whales, is very big [citation needed]. Long before a group can take over society, it reaches a size where it needs to develop internal structure and rules about interaction between group members. If you collect a bunch of people and tell them to abandon all the social norms like honesty, politeness, respect, charity, and reason in favor of a cause – then the most likely result is that when your cause tries to develop some internal structure, it will be overrun by a swarm of people who have abandoned honesty, politeness, respect, charity, and reason.

Contrary to what our medical orgonomist friend says, I don’t think the Communists were much of a social cancer. The political communists, the kind we got in the US, mostly stuck to the rules. But look what happened to them. They got all excited about how the governing power structures were evil and needed to be destroyed. They set up various organizations dedicated to destroying the governing power structures. And people proceeded to decide that the governing power structures of those organizations were evil and needed to be destroyed. As a result, Communist parties were rent by constant factional warfare and they never got around to destroying the society they lived in at all.

And look what happened to Mr. Clymer. His whole spiel was about throwing out the virtue of charity – how when people have been accused of things, we should condemn them automatically, no chance it’s wrong, false accusations rarer than comet strikes. And then…

We have a lot of proverbs about this sort of thing. “Hoist with his own petard”. “Taste of your own medicine.” “A trap of your own making.” Jesus said “He who lives by the sword will die by the sword” – and it has been pointed out to me that Jesus himself lived as a carpenter and died by being nailed to a piece of wood, so there’s that.

Or since we’ve gradually meandered back on to the topic of sexism, there’s an ancient Sexist proverb that seems highly apropos: “Bro, if she’ll cheat *with* you, she’ll cheat *on* you”.

And in the same way, someone who will be a jerk *for* you will be a jerk *to* you. My disagreement with Arthur was about his willingness to tolerate jerks in his movement. He thinks it will help him win. But someone who is a jerk to men will, by and large, also be a jerk to women. Someone who is a jerk to men’s rights activists will, by and large, also be a jerk to feminists. They may not do so immediately, if it doesn’t serve their self-interest to do so. But in private, or as soon as the chance comes up, jerkitude will out. If you defend them as long as they’re only being jerks to outsiders, then a few months later, when – shock! horror! – you realize they’re being jerks to insiders, you end up having to retreat and mumble something about how you were “kinda wary about him” all along.

But more importantly if you elevate jerkishness into a *principle*, if you try to undermine the rules that keep niceness, community, and civilization going, the defenses against social cancer – then your movement will fracture, it will be hugely embarrassing, the atmosphere will become toxic, unpopular people will be thrown to the mob, everyone but the thickest-skinned will bow out, the people you need to convince will view you with a mixture of terror and loathing, and you’ll spend so much time dealing with internal conflicts that you’ll never get enough blood supply to grow large enough to kill a whale.

You will get things like #StopClymer and #Jacobinghazi.

Thus the Archipelagian Principle: Given infinite freedom of association, most people end will up in more or less the community they deserve.

Fish – Now By Prescription

I.

LOVAZA™®© (ask your doctor if LOVAZA™®© is right for you) is an excellent medication. It is extraordinarily safe. It is moderately effective at its legal indication of lowering levels of certain fats in the bloodstream. It has moderately good evidence for having other beneficial effects as well, including treating certain psychiatric, rheumatological and dermatological disorders.

Lovaza is fish oil.

“Come on,” you say, “surely there’s some difference between Lovaza and the fish oil I buy at my local health food store for a couple of tenners per Giant Jar?”

And you’re right. The difference is, Lovaza costs $300 a month.

“Come on,” you say, “surely there’s some *other* difference”.

The company behind Lovaza certainly *thinks* there is. They boast about how they use a patented process to filter out blah blah purity blah toxins blah certified to be blah blah blah quality blah. On the other hand, *every* fish oil supplement maker uses about the same phrases. I’m sure it’s not exactly the *same* patented process and that they check for toxins in a slightly different way, but the evidence that Lovaza’s methods are superior in some interesting way is, basically, nonexistent.

So, fish oil pills cost $30 at the store, Lovaza costs $300, which would you rather get?

And the answer is: the Lovaza, of course. Because *you're* paying the $30 for fish oil, and *your insurance* is paying the $300 for Lovaza. Maybe you've got a copay. Is the copay more or less than $30? Well, there you go.

This isn't an unusual business model. Off the top of my head I can think of a second drug that works the same way. It is called Deplin, and it treats depression. On the one hand, it is a good medicine and I wish it was used more. On the other hand, it is l-methylfolate, which is basically a very shiny chrome-plated version of the folate that your cereal box tells you it has 200% of the recommended daily allowance of. Most studies suggest that for most people without a particular mutation there's not much difference between the folate in your Cheerios and the l-methylfolate in Deplin – but if you have the mutation or are among the significant minority of people who believe there is, you can *also* walk into a health food store and buy l-methylfolate itself, chemically identical to Deplin, at about $10 per Giant Jar. Deplin costs $100 per month.

Ask your doctor whether she recommends people go to supplement stores and buy folate or fish oil, and likely she'll say supplements sometimes make people feel livelier but that they have minimal place in responsible medical practice. Then ask your doctor whether she gives out prescriptions for Lovaza and Deplin and she may very well tell you she writes a dozen of them a day.

II.

When I search the web for people's Lovaza and Deplin opinions, I see a lot of people talking about greedy pharmaceutical companies. Fair enough. Pharmaceutical companies are pretty greedy, and it's hard to deny that's a pretty big factor in how these kinds of drugs (I don't know a good category name for them, but Deplin bills itself as a "medical food", so we'll go with that) come to be.

This blog sometimes looks at things from a libertarian perspective. The libertarian perspective says that usually if a for-profit company is making money, it's probably providing someone with a service somewhere. Is the public getting any service from Lovaza and Deplin?

I say: yes! The companies behind these two drugs are doing God's work; they are making the world a much better place. Their service is *performing the appropriate rituals to allow these substances into the mainstream medical system*.

A doctor who prescribes boring regular old supplement fish oil pills is taking a dangerous step into uncharted territory. If anything goes wrong and their conduct comes under review, a clever lawyer could say "I notice your patient had severe hypertriglyceridaemia, a very dangerous condition, and instead of giving her any medicine, you just told her to get fish oil from her local health food store! Fish oil has never been FDA-approved and you have only your personal opinion that it does anything at all."

And that's if the patient even *gets* the fish oil. What does the insurance company say if there's a patient too poor to pay the $30 for a Giant Jar? This is really common, both in a "a lot of people are legitimately poor and only able to get things Medicaid subsidizes for them" way, and in a way that's sort of the reverse of how people can always find $50 a day to support their heroin habit. "You want me to pay money for pills to treat a condition which is not *at this moment* causing my limbs to fall off? Sorry, can't afford it." But insurance companies will laugh in your face if you ask them to pay for some random supplement at a health food store which is made out of some slimy animal that swims in the ocean and which *doesn’t even have a ® after its name*.

And this is even assuming there’s a health food store to get it at. What if you’re in a hospital? Or a nursing home? They don’t have health food stores. They just purchase a bunch of medications, put it all in a big room somewhere, and dispense it to patients who need it. Exactly which medications they purchase is highly highly regulated both by government laws and internal regulations, and it is probably not going to include a Giant Jar of fish oil pills.

I mentioned before that, in a hospital pharmacy stocked with hundreds of different rare and wonderful substances – substances that must be kept frozen at very specific temperatures, substances that can only be given through complicated surgical tubing, substances which must have the dose be absolutely right to within a thousandth of a gram or they will kill you instantly – they don’t have melatonin. The drug that is probably in the bathroom cabinet of half the people reading this post is *totally missing*. Because it’s not Official.

The point of Lovaza and Deplin is to make fish oil and folate Official. I *wish* someone would do the same with melatonin so I could start prescribing it in-hospital already.

The high prices of these drugs? Likely at least in part an attempt to recoup costs. In order to make them Official, their parent companies had to perform very specific studies on them. Not just “there are hundreds of studies on fish oil, we’ll just attach them to an email and cc the FDA”. No, that would be too easy. They had to perform exactly the studies the FDA wanted, using fish oil that came from *their* factory, with *their* patented process of extraction. That was a couple million to a couple billion dollars. Then they had to hire a bunch of lawyers and lobbyists to make the FDA like them. Then they had to price in the cost of risk from the dozen drugs they try to develop that never get approved and make zero money. Then they had to hire really pretty women to go around to doctors and give them nice pens that say “Lovaza” on them, because despite attending medical school for four years and residency for another three, the number one way doctors learn about drugs is through pens with the drugs’ names on them given to them by pretty women. Then lawyers and accountants and executives had to go through a bizarre process in which they all argued among themselves about how much Lovaza should cost and who should pay for it, which ended up, like all such drug negotiations end up, with parties A, B, and C marking up the sticker price by ten times, parties X, Y, and Z demanding it at one tenth the sticker price, and both sides thinking they pulled a fast one on the other while everybody else facepalms.

My point is, the drug companies are probably making an obscene profit off of it, but less obscene than you might imagine from their ability to stick a fish in a blender and sell you what comes out the other side for $300.

And if you’re still not convinced that this whole thing benefits society, think about this.

Naturopaths and alternative medicine people and “healers” of all sorts have a common form of apologetics for their chosen folk remedy: “Everyone ‘in the know’ has figured out this is a miracle cure, but because it just *grows* out of the *ground* Big Pharma can’t make a profit off of it. That’s why they’re suppressing it and poisoning you with toxic medications instead.”

And I used to think that the alternative medicine people were overestimating how evil Big Pharma was. But now I know that's not right.

Now I know they're *underestimating* it.

If it were discovered tomorrow that potatoes cured cancer, then people wouldn't "suppress" this "natural" remedy. Two years from now there would be an ultrapurified potato extract called POTAXOR™®© that was, on closer examination, physically and chemically identical to mashed potatoes. But *these* mashed potatoes would be mashed in a giant centrifuge by scientists with five Ph. Ds each. Any time someone got cancer, their doctor would prescribe POTAXOR™®© and charge $6,000 per dose, and the patient would get better, and the thought of just going out and eating a potato would never occur to anybody. Not to the doctor, who doesn't want to sound like the idiot who tells her cancer patients to eat potatoes. Not to the FDA, who doesn't know whether potatoes might be contaminated with lead or potato fungus or ketchup or God-knows-what. And certainly not to the patient. They would have to pay 60 cents for a potato at the supermarket, but if they have a good enough insurance the POTAXOR™®© is free!

This system, bizarre as it is, is your guarantee against the pharmaceutical companies suppressing a promising new natural medication. Your insurance company pays $300 on fish oil, and in exchange you go to sleep at night secure that no one has discovered that potatoes cure cancer but decided to cover it up to protect their bottom line. Good deal? Given the current health system, it's better than you had any right to expect.

III.

I promised yesterday I'd talk about a Jacobin article, so here it is: Bad Science.

Apparently there's a company called Myriad Genetics which charges a lot of money to test BRCA1 and BRCA2, two genes in which mutations are associated with a very high risk of breast cancer. This is outrageous because genetic testing is very cheap. Worse, they don't even have the excuse of "the customer is paying extra to reimburse the cost of the research that discovered this", because the relevant research was done by the government at public expense. From the Jacobin article I don't understand exactly what their patent/monopoly status is, but it seems to be enough that they can price gouge without repercussion.

I don't want to comment on the broader points about the scientific community Jacobin is making, but when I read that I thought "Wait! I had an opinion about BRCA testing at one point!"

Yeah. After searching my memory I remembered that the FDA recently banned popular genetic sequencing companies like 23andMe from offering BRCA testing bundled with their other services. The problem was that 23andMe was nonprescription. It was outside the Official Medical System. It was fish oil, not Lovaza. I was pretty annoyed about that decision.

But after reading this new article I have learned something that puts the whole thing in a new light. Whole genome sequencing companies can sequence all 20,000 or so coding genes for about $1000. Myriad Genetics will sequence two genes for $3000.

[**EDIT:** Douglas Knight, more expert in this area than I, discusses some of the subtleties here: 1, 2, 3]

I’m not saying that the FDA is in bed with the genetic testing companies. But if the head of the FDA gives birth to a child who looks suspiciously similar to the CEO of Myriad Genetics, I hope her husband gets a paternity test. Assuming those haven’t gotten banned yet.

I am willing to write off Lovaza as a weird but successful way of adapting to an increasingly crazy medical system. I even find it darkly hilarious. But the high cost of BRCA testing probably turns off a lot of women who would otherwise get the test, and some of them probably develop preventable breast cancer. Lovaza is a farce. BRCA is a tragedy.

Things That Sometimes Help If You Have Depression

Some of my friends have depression and have asked me for some suggestions.

These will be inferior to reading official suggestions, but you will probably not read official suggestions, and you may read this. Just so we're clear, all opinions here are my own, they are not endorsed by the hospital I work at, they do not constitute medical advice, I have a known habit of being too intrigued by extremely weird experimental ideas for my own good, and you read this at your own risk. I am still an intern (a very new doctor) and my knowledge is still very slim compared to more experienced professionals.

Overall I think this is more of a starting point for your own research rather than something I would expect people to have good results following exactly as written.

And one more apology: originally I tried to include links to appropriate studies with each intervention, but there are so many different studies, and it's so easy to pick apart each, and so much of the research is based on a gestalt impression after having read three dozen studies rather than on any individual one – that I decided that listing the evidence fairly would require this to be ten times as long and much more academic. I'm doing that thing where "perfect is the enemy of good enough" and trying to actually get this up online. I'm happy to discuss evidence or lack thereof for any particular therapy in the comments.

Now that that's over: first I'm going to talk about figuring out if you need help. Then I'm going to recommend you see a psychiatrist. Then I'm going to accept that in reality a lot of people for whatever reason can’t or won’t see a psychiatrist, and grudgingly recommend some lifestyle interventions you can make. Then I’m going to accept that in reality a lot of people for whatever reason can’t or won’t make lifestyle interventions, and grudgingly recommend some over-the-counter medications and supplements that might be helpful.

I. Do You Have Depression?

Major depressive disorder is the clinical condition that best corresponds to what people usually mean when they say “depression”. Only a licensed professional can officially determine whether or not you have major depressive disorder. But if you feel miserable all the time, you might be able to make one heck of a good guess.

The PHQ-9 is a well-known and validated screening tool for depression; you can take it at the linked site. It cannot diagnose you officially, but once again, it can help you make one heck of a good guess, and if you get a high score it might inspire you to get to a doctor’s office.

Overall there’s not exactly an epidemic of perfectly healthy people misdiagnosing themselves as depressed, but there are a couple of things worth keeping in mind:

– You can have depression even if there is a good reason for you to be depressed – for example, if you’re depressed because you lost your job. If you have the symptoms, and it’s been going on more than two weeks, it counts. These kinds of depression seem to respond to treatment about the same as kinds that come on for no reason at all.

– Depression usually lasts a long time. Depressive episodes usually last weeks to months. Someone who is depressed for a day or two after something happens probably does not have depression. Even if they are *unreasonably* depressed for a day or two after something happens, so much so that they seem to have a mental disorder, it may be a different mental disorder.

– People with bipolar disorder (“manic-depressive disorder”) are often depressed some of the time. These depressive episodes can last a long time, just like the ones in traditional depression. However, bipolar disorder is a completely different disease and needs completely different treatment. If you sometimes have “manic” or “hypomanic” episodes – several days to weeks of having abnormally high energy, abnormally low sleep, abnormally high self-esteem, abnormally short temper, and poor impulse control – then you probably have bipolar disorder instead of depression. If your doctor or psychiatrist has diagnosed you with depression and started treating you with antidepressants and you don’t seem to be getting better, then you have found your problem. Politely bring up your history of manic or hypomanic episodes and ask whether she wouldn’t prefer to diagnose you with bipolar disorder instead.

II. Please Go To A Doctor

If you have depression, your best bet is to go to a doctor. A doctor can diagnose you and connect you with useful prescription-only medications like antidepressants. You do not need to go to a psychiatrist to get antidepressants. Your family doctor will be able to prescribe them. You will only need to go to a psychiatrist if your depression fails treatment with normal medications and someone needs to figure out a more complicated plan.

The failure mode of medical professionals I see most often seems to be something like this:

Patient: I’m depressed.
Doctor: Here, have an SSRI
Patient: *(three months later)* I’m still depressed.
Doctor: Well, keep taking that SSRI I gave you, I’m sure you’ll get better eventually.
Patient: *(six months later)* I’m still depressed.
Doctor: Sheesh! I already *gave* you an SSRI, what else do you want from me?

There are dozens of different depression medications. If one doesn’t work, very commonly another one will. Depression treatment is difficult, because there’s no way of knowing beforehand what medications will or won’t work for any individual person. The correct solution is to start with the safest medications, see if they work or not, and gradually move up to stronger medications with more side effects. But this doesn’t work if your doctor just tells you to keep taking whatever you’re taking forever whether or not it’s doing anything.

A lot of people are reluctant to second-guess their doctor on this sort of thing. So let me provide you with a loose approximate algorithm that you can follow when going to a doctor or psychiatrist about depression. Many doctors have very good reasons for deviating from this algorithm, but if they are nice people they should be willing to explain what their reasons are to your satisfaction.

1. Rule out organic causes

When your doctor diagnoses you with depression, they may perform several blood tests. The most important ones are a thyroid test for hypothyroidism, and a blood count for anaemia. Hypothyroidism and anaemia are medical illnesses that can cause depression. If you have them, then psychiatric treatment for depression won’t help and won’t treat the underlying potentially dangerous condition.

Symptoms of hypothyroidism besides depression include feeling unusually cold, gaining weight despite good diet, constipation, pale dry skin, weakness, and disturbed menstrual periods if female. It is easily detectable with blood tests and easily treatable with thyroid hormone, but someone’s got to look for it.

Symptoms of anaemia besides depression include looking very pale, feeling weak and tired, and occasionally pica, the compulsion to eat non-food items like ice or dirt. It is very common in women with heavy periods. It is easily detectable with blood tests and usually easily treatable with iron supplements.

If you’re being treated for depression with psychiatric medication, and it’s not helping, and you have some of those symptoms, and nobody ever did blood tests on you, politely ask your doctor if you’ve been tested for hypothyroidism and anaemia (people get lots of blood draws all the time and lots of the time they’ve just forgotten). If you haven’t been, politely ask why not. If your doctor doesn’t have a good reason, and your depression isn’t getting better, politely ask if they think it would be a good idea to check for at least those two conditions and maybe some vitamin deficiencies.

2. Choose between pharmacotherapy, psychotherapy, or both

Many to most doctors and psychiatrists will assume that you want medication. This is a pretty good assumption most of the time. Medications and psychotherapy are about equally effective in treating depression, but psychotherapy costs a lot more, takes more time, and is harder to get your insurance to cover.

Still, a lot of people who would have preferred psychotherapy never get the option. And if you’ve tried lots of medications that haven’t worked, maybe you’ll luck out with psychotherapy. You can either get your doctor to recommend someone or find someone yourself.

Cognitive-behavioral therapy is a nice neutral therapy for depression with as high a success rate as any other. It usually consists of a couple of sessions in which somebody talks to you about the way you deal with and think about problems and maybe assigns you some homework. There’s no talking about your mother or about your sexual fetishes. 6 to 12 one hour sessions would be a pretty standard starter course.

Having both psychotherapy and medication has been associated with a better treatment rate than either one alone.

3. Start with an SSRI

SSRIs are the most commonly used antidepressant medication. They are no stronger or weaker than other antidepressants, but they do have fewer side effects, which makes them first-line.

There is a lot of worry that SSRIs are not much better than a placebo for people with mild to moderate depression (almost everyone agrees they are effective for severe depression). Some studies have seemed to confirm these fears, others have seemed to refute them. I am leaning towards “marginally better than placebo”. However, this is a completely academic debate for you, because you are not going to get placebo. Your choice is between SSRIs or nothing. Everyone everywhere agrees SSRIs are much better than nothing.

All SSRIs are approximately equally effective. Your doctor or psychiatrist’s choice of SSRI comes down to which side effects you are most willing to tolerate, the exact shape of your depressive symptoms, and which drug company gives out the most pens and coffee mugs with the name of their medication on them. Celexa or Lexapro (citalopram or escitalopram) are good first choices for most people, but it is hard to go disastrously wrong here.

Bupropion is not an SSRI, but is also an appropriate first choice. SNRIs like Effexor (venlafaxine), as well as mirtazapine which is sort of in a weird little class of its own, are also okay for certain people. If your depression is very severe, your doctor might skip this step and go further down in the algorithm, which is also okay. Some psychiatrists also may have their own idiosyncratic preferences which are probably okay as well if they can explain their reasoning to you.

Many people worry about the side effects of SSRIs. In my opinion many of these worries are exaggerated. The most common side effect is decreased sexual drive and performance, which can occur in more than half of users. Other side effects include nausea, diarrhea, insomnia, and weight gain, which occur less frequently. These side effects are usually temporary and go away when you stop using the drug. If you’re worried about sexual side effects, bupropion is much less likely to have these, with the tradeoff being a few other effects like more insomnia.

Some antidepressants have discontinuation syndromes, which mean you feel sick when you’re withdrawing from them. Effexor (venlafaxine) and Paxil (paroxetine) are particularly bad. A competent doctor or psychiatrist will take you off these slowly so that you avoid any bad effects.

4. Give it a little while to work

Antidepressants take some time to start working.

This is less true now than it was a while ago – it used to be believed they almost never worked until a few weeks had gone by, whereas now we know that there can sometimes be some improvement pretty quickly. But other times there isn’t. Although you can *hope* for improvement within a week or two, give an antidepressant at least four weeks, just to be sure you’re not selling it short.

If you can’t wait four weeks, you may want to consider mirtazapine, which works a little more quickly. But you’re still not going to pop a single pill and start feeling much better.

5. Fiddle around

If a drug seems to work a little bit in the first month or two, it might be worth raising the dose.

If it doesn’t work at all within the first month or two, you could still try raising the dose, but you might also want to try switching to a different drug. For example, if one SSRI doesn’t work, you could try switching to a different SSRI.

If SSRIs don’t work, this might be a good time to try antidepressants from other newer classes. Effexor (venlafaxine), mirtazapine, and bupropion would be three very good choices here. All three are pretty safe and very commonly used. You could also move to a tricyclic antidepressant, which have just a *little* bit more side effects than some of the newer classes that have replaced them by which can sometimes be very successful in certain patients when SSRIs have failed.

If nothing works alone, your doctor will probably try combining a couple of these drugs and seeing if that helps.

6. Get serious

Once treatment with SSRIs and the usual SSRI replacements has failed, you get to start trying more serious stuff. Usually the drugs here either have worse side effects or are still a little bit experimental. These will probably be prescribed by a psychiatrist and not your regular family doctor.

Atypical antipsychotics have been found to be quite effective in depression. Unfortunately, many have side effects, including weight gain and increased risk of diabetes. Abilify has fewer of these, which makes it a very popular choice. Seroquel has a few more, but it also relieves anxiety and helps sleep, which makes it very popular as well. Olanzapine has a lot of good evidence behind its antidepressant properties, but its side effects can be very severe, so I would be more reluctant to prescribe it than either of the other two.

MAOIs are an old antidepressant medication that went out of fashion when SSRIs came around. A lot of patients remember them very fondly; according to anecdote they are very powerful and can give people a very good mood. *However*, they have a bad habit of causing life-threatening hypertensive crises at inconvenient times if you eat the wrong thing, where “the wrong thing” includes everything on a long list of foods your doctor has to give you and lecture you about before they can be prescribed. And just to warn you, cheese, beer, and chocolate are definitely included. The Europeans seem to have a less dangerous

MAOI called mobeclomide which hasn’t been approved in the US yet, but I don’t know anything about it.

Thyroid hormone is very rarely used as an antidepressant augmentation, but there is some preliminary evidence that it is as effective as lithium with fewer side effects, even for people without obvious underlying thyroid problems.

Lithium is a mood stabilizer more commonly used in bipolar disorder, but which has some indication for depression as well. Usually it is taken along with another antidepressant as “augmentation”. It works pretty well, and is especially good at decreasing suicidality, impulsivity, and aggression. Unfortunately it’s not the safest medication in the world and usually requires some annoying blood tests to make sure you keep getting the right amount of it.

Modafinil shows some promise as an antidepressant supplementation strategy and is discussed in more detail in section IV below.

This might also be a good time to double-check that you were *definitely* tested for underlying medical conditions like hypothyroidism and anaemia, and that you *definitely* don’t have bipolar disorder or some other condition mimicking depression. Or you might want to look down and try some of the lifestyle interventions and supplements later in this list.

7. Get very, very serious

Electroconvulsive therapy gets a bad rap from old movies where we see people who “misbehave” in the mental institution getting what look like very painful electric shocks.

Modern electroconvulsive therapy is done with the patient sedated. You are pleasantly asleep, your limbs aren’t flailing about or anything, and it can be done on an outpatient basis with you going home as soon as you're done.

There are a lot of worries about side effects. Some people experience some memory loss, especially of the couple of weeks before the treatment. Most of the time these memories come back. Sometimes they don't. But how much do you want to remember the week when you were so depressed you needed ECT, anyway? Longer-term side effects are less well-known. Some people think a *lot* of ECT isn't good for your brain, but if the effect exists it's small enough that people are still debating whether or not they've really picked it up.

But the thing is, ECT really, *really* works. People get to the point where everything else has failed, and they've been on seven hundred different medications without feeling any better, and they're ready to give up, and then they get ECT and start whistling happy songs and dancing the polka. I won't say it works 100% of the time, because no medical treatment works 100% of the time. But in psychiatry, where expectations are always low, it's the closest thing we've got to a miracle cure.

Speaking of miracle cures, a couple of people have asked me about ketamine. Ketamine does show unusually rapid and effective action in treating depression. However, it is currently still very experimental and your doctor or psychiatrist will not be able to give it to you unless by a strange coincidence they are one of the few researchers working with this drug. Further, there are still a lot of problems with ketamine therapy for depression. First, the drug is highly hallucinogenic, sometimes in very scary ways; in experiments subjects are put under anaesthetic first so they don't have to consciously experience the hallucinations. Second, it's not yet clear how long-acting it is; anything that required long-term ketamine therapy (say, a ketamine dose a week) would be impractical both because of the hallucinations and because the drug has serious side effects and is addictive. Although it is very exciting to researchers, it is probably not very useful to you.

Anyway, that’s the algorithm – which I lifted from a couple of treatment guidelines – that I would recommend people watch to see if their doctor follows. Because I feel like – if someone is really really depressed, to the point where it’s ruining their lives, and they are begging for help, and their psychiatrist tries *all* of steps one to seven, and even the ECT doesn’t help them, *then* they can consider saying “Well, I’ve done everything I can”, giving up, and lowering their expectations. Any doctor who gives up before they’ve reached that point has no excuse.

How do you find a good psychiatrist?

I have no great advice here, except that if you think your psychiatrist is terrible, you are probably right. Even if you are not right, your psychiatrist is apparently terrible for you. See if your insurance allows you to switch psychiatrists. If they do, try it and see if the next one is any better.

If a psychiatrist is doing something *completely* unlike the treatment algorithm in Part A, ask them to explain why. Be willing to accept any explanation, because treating depression is hard and there are a lot of different valid approaches. But if they refuse to explain and tell you that you’re a bad person for asking, you *miiiiiiight* want to see if you have other options.

There are a lot of doctor rating sites. Ratemds.com or healthgrades.com are among the bigger ones. These sites are not very valuable on the margin; you do sometimes get one guy who had to sit in the waiting room a little too long going on a personal campaign to destroy someone’s reputation. But on the tails, if there’s a doctor who is universally hated by every single one of her patients, this can be a strong warning sign.

People in mental health support groups are *absurdly* willing to share their opinions of various psychiatrists. Just don’t be surprised if you can’t get them to shut up once you’ve gotten your information.

Kate Donovan has a good guide to finding a psychotherapist, with partial crossover to the skill of finding a psychiatrist.

How do you see a psychiatrist without worrying you will be committed to an institution?

This is something I see a lot of people worry about, and something that prevents a lot of people from seeing a doctor or especially a psychiatrist. I think it happens less in reality than it does in people’s fears, but it does sometimes happen in reality, and enough people avoid getting help for this reason that it’s worth discussing briefly, at the possible risk of giving more airtime to something most people should not be worrying about.

The first and most important point is that very very few psychiatrists, whether good or bad, will commit people to a mental institution unless they are very sick. Even if you are very sick, there is only a small chance of a psychiatrist committing you against your will unless you say exactly the wrong thing.

Doctors and psychiatrists are legally required to commit patients to mental institutions if those patients are “a threat to themselves or others”. Usually this means a patient has said they want to commit suicide and they are probably really going to go through with it soon, or a patient has said they’re going to hurt someone else (or implied it: “He’s gonna get what’s coming to him”) and it wasn’t clearly meant metaphorically. They may also commit you if you are very paranoid on the grounds that paranoid people may try to pre-emptively attack those they believe are plotting against them – or if you are refusing to eat, on the grounds that if you don’t eat you die.

A good psychiatrist will differentiate between vague suicidal ideation (“Sometimes I feel like life isn’t even worth living, do you know what I mean?”) and specific suicidal ideation (“I will kill myself tomorrow using the gun hidden in the back of my pantry”), will explore the first type with an aim toward helping you, and will only involuntarily commit for the second type. I do not guarantee you will have a good psychiatrist.

Until you are sure your psychiatrist is trustworthy, you may want to steer clear of statements that sound suicidal, homicidal, or paranoid. Even in jest. *Especially* in jest. Until you have established *absolute trust*, please treat joking about suicide or homicide around your psychiatrist the same way you would treat joking about terrorism around airport security agents. *I cannot overemphasize how important this is.* If you won’t do it for yourself, do it for the sake of your poor overworked local inpatient psychiatrist, who is *sick* of hearing all his new patients say “I shouldn’t be here, I promise I was only joking!”

If you are genuinely suicidal, homicidal, or paranoid, but you are absolutely sure you are not an acute danger to anybody, and you trust your psychiatrist enough to tell him or her about these things – then make sure you phrase it in a way that specifies you are not an acute danger to anybody. For example “Sometimes I feel like I would be better off dead… BUT I AM DEFINITELY NOT GOING TO COMMIT SUICIDE BECAUSE THAT IS MORALLY WRONG AND I CARE A LOT ABOUT MY FAMILY AND I WANT TO LIVE!!!”. Or “Sometimes I feel like people are plotting against me…BUT NOT ANYBODY SPECIFIC AND I KNOW THAT’S NOT ACTUALLY TRUE!!!”

This should keep you relatively safe from involuntary committment unless your psychiatrist is truly awful.

III. Lifestyle Interventions

Not everyone can or will go to a doctor, and even those who do go to a doctor might want to try something more proactive on the side, so in this section I’m going to list some lifestyle interventions you can make.

1. Do therapy on yourself

If you can’t or won’t go to a therapist, a therapy workbook is a practical alternative. Cognitive Behavioral Workbook for Depression is rated 4.3/5 stars on Amazon, which I guess is *sort of* like being evidence-based. Many very smart people find these sorts of therapy workbooks to be a little condescending (“Really? I need to stop catastrophizing all the time? I *never* would have thought of that on my own!”) but other equally smart people find them useful, whether because they have new insights or because they repackage, remind, and rehearse what they already know. For 3$ for a used copy, that’s about 1% of the cost of most of the other interventions you could try.

2. SLEEP!

Put in all capitals with an exclamation point at the end of it to show how important I think it is.

Depression is *weirdly* linked to circadian rhythm and sleep. Poor sleep habits probably help cause and exacerbate depression. Unfortunately, depression also causes and helps exacerbate poor sleep habits, so it’s a kind of vicious cycle.

In a recent study, people who received cognitive behavioral therapy for sleep disturbances had double the recovery rate from depression of people who didn’t, suggesting that attacking insomnia is pretty much just as effective as the strongest drugs known. This avenue is *severely* underexplored and *very* promising.

The cognitive behavioral therapy for sleep disturbance is about one part basic CBT techniques of challenging your perceptions and beliefs, and one part the things your mother told you about sleeping in a dark room and not watching exciting TV shows right before bedtime. This therapy *also* has a cheap workbook with a 4.6 star Amazon rating. You can add whatever other high-tech exciting sleep cycle correctives you know – melatonin, magnesium, whatever – to what it tells you.

3. Exercise

I work in a psychiatric hospital. Once a week or so a social worker leads an exercise group there, and it is *amazing* how much better everyone does that day compared to the days before and after. Exercise seems to increase release of BDNF, an important brain chemical that depressed people don’t have enough of, and there have been several studies showing good effect.

Fast walking for a half hour five days a week seems to be enough to help. More exercise might help more. And exercising outside will get you more sunlight and vitamin D, whose relationship to depression remains controversial but which certainly can’t hurt.

4. Light therapy

Light therapy is hanging out around really bright lights for a while and hoping they cheer you up. The Cochrane Collaboration and a meta-analysis in a major psychiatric journal agree that they do, quite robustly (although there is still uncertainty about whether they add *extra* treatment to somebody already being treated with antidepressants).

Experts recommend a light therapy box emit 10,000 lux (a unit of light). Here’s a very official looking one that costs $200 and a somewhat less expensive one at $70. But don’t take these as recommendations – I’m just repeating out what shows up on Amazon search.

Best practice is to keep it above your line of vision so your brain feels like it’s the sun, and sit near it for at least a half hour in the morning. Mayo Clinic tells you a little more about how to do it here.

5. Drugs are bad for you, mmkay?

The Simpsons described alcohol’s role in treatment far better than I can: “There’s nothing like a depressant to cure depression.” Except I think they were being sarcastic, and a lot of patients are serious.

Alcohol probably works in helping you forget about depression for a little while, but over the long term it makes you much more depressed. If you abuse alcohol, stop doing that.

Benzodiazepines (Xanax, Ativan, Klonopin) are, once again, good short-term solutions to anxiety. If you abuse them, they will probably contribute to your depression.

Smokers have more than twice the rate of depression than nonsmokers. There's a lot of debate about whether it's causal or noncausal, but some sophisticated statistical modeling seems to suggest there is indeed a real link. Also, getting lung cancer is depressing as *heck*.

I don't know any studies linking heroin or cocaine to depression directly, but they probably take your life in a pretty depressing direction.

Obviously all of these drugs are hard to quit, and you might need the help of a medical practitioner.

6. Other things that make you happy

Cure Together is a really neat website where people with illnesses record all the things – medical and nonmedical – they tried and how well it worked for them. It obviously trades off the well-controlled conditions of an experiment in exchange for sheer amount of data, but it's a useful adjunct to other ways of getting information. You can find their chart for depression here.

Just from looking at it, exercise and SLEEP! are right up at the top, but people also mention things that more scientifically-inclined people would probably never think of, like spending time with a pet, listening to music, and watching funny TV shows.

IV. Supplements

Despite the best of intentions, I think a lot of people are going to skip parts II and III and start down here. I think that would be a mistake, but I understand depressed people don't always have the energy for big lifestyle changes or the willpower to take the anxiety-provoking step of seeing a psychiatrist. So sure. Let’s try supplements.

A reminder – supplements are chemically active compounds. They can have side effects. You should look up what they are. In particular, some of them can disrupt the metabolism of other medications. Contraception is another medication. If its metabolism was disrupted without your realizing it, that could be very bad. Be careful. If you’re taking other medications, tell your doctor about any supplements you’re taking. Don’t take supplements without a doctor’s okay if you have liver damage, kidney damage, or any chronic medical condition.

And with that stirring recommendation, here are some supplements that seem promising for depression, in approximate order of recommendedness.

1. S-Adenosyl methionine (SAM-e)

This would be my first choice if you’re trying to treat depression with supplements. It has a good evidence base, big effect size, it’s pretty safe, and it’s available on Amazon for $20 a jar. It probably works both on its own and as an adjunct to antidepressants, although there’s only partial evidence for the first claim. One experiment got good results with 800 mg two times a day for six weeks, though aside from that proper dosage is anyone’s guess.

2. Creatine

Yes, the same stuff bodybuilders use. One RCT found good results at 5g per day. May be more effective in people likely to be protein-deficient (eg vegetarians) and possibly in women.

3. Folate or l-methylfolate

Folate is a form of Vitamin B9 that most people get in food. Several RCTs have found positive effects from supplementing antidepressants with folate, although the evidence for folate on its own is still lacking. Some people with a mutant version of a gene called MTHFR are unable to process folate very well, and *may* get better results from an "optimized" version called l-methylfolate. A psychiatric medication commonly prescribed for treatment-resistant depression, Deplin, is l-methylfolate, but the supplement version is exactly the same. 5 – 15 mg / day seems to be a common dose, even though a lot of the supplements seem to contain a lot less.

4. Saffron

Yes, the spice. I'd never even heard about this until I checked examine.com, but it seems to check out. Five different randomized controlled trials found saffron to be more effective than placebo and as effective as conventional antidepressants. Amazon seems to be flooded with sellers ever since Dr. Oz apparently said it might be a weight loss supplement. [some problems, especially for pregnant women]

5. Fish oil

A lot of people are very excited about this, including examine.com. I'm more skeptical. There have been a lot of studies that suggest fish oil can improve mood in depression, and a lot of others that find no effect. And fish oil is a complicated supplement to deal with – most of the commonly sold pills don't have enough to do any good, and if you don't store it right it goes rancid and is worse than nothing.

Instead, I would eat a lot of fish. The degree to which the level of depression in a country correlates with the amount of fish eaten in that country is *staggering*. Two salmon dishes a week ought to be a good start.

6. -afinil

Modafinil is a wonder drug that gets used for everything, both legally (with prescription) and illegally (without). It was only a matter of time before someone tried it for depression, and it seems to work pretty well, at least as an adjunct to antidepressants. There are fewer studies about whether it works on its own. I would guess that it does, because the most likely mechanism is its well known tendency to increases energy and alertness, which is pretty useful against depression’s fatigue and tiredness. The strongest argument against modafinil is that it lacks a plausible mechanism for antidepressant action and so realistically you’re probably just treating symptoms. The strongest argument *for* is that you’re probably treating symptoms *very well*. Also worth noting that modafinil disturbs sleep and disturbing sleep is *very bad* in depression, so take it at the beginning of the day and if you still can’t sleep at night, cut the dose.

Modafinil is illegal without a prescription, but everyone on the Internet sells it anyway. Adrafinil is a prodrug that turns into modafinil once in the body. It is perfectly legal without a prescription, because the medical licensing regime makes no sense. You might as well just get that – as far as I can tell the risk of liver damage is overhyped if your liver is otherwise healthy.

7. Other things

Things I find interesting but which have conflicting evidence and/or don't deserve a whole paragraph of their own include: lithium orotate microdosing, rhodiola, vitamin D, zinc, and curcumin. Look them up and see what you think.

Not exactly a supplement, but I'm pretty sure you can get Botox without a prescription, and if done correctly it might help substantially.

What if you want to buy antidepressants illegally without prescription?

Well, you probably won't get caught. And you probably won't kill yourself if you go into it semi-well-informed. I can't in good conscience recommend it, but I'm sure it's something a lot of people think about. Consider yourself scowled at, and at least take the following advice:

Definitely definitely do not take any tricyclics (hint: if it ends in -pramine, it's a tricyclic) without a prescription. Those are *seriously* potentially dangerous. Taking MAOIs (phenelzine, tranylcypromine) without a prescription is *the worst idea* and you will die (note: some over the counter supplements claim to work through "MAOI-like action" or "being reversible MAOIs". As far as I can tell these are safe, though I can't vouch for their efficacy. Just don't take the real thing.) I don't even think anyone sells antipsychotics without prescription, but *also* the worst idea. If you are thinking of buying any of these without prescription, and you won't accept advice to not buy prescription medications at all, can I at least talk you down into buying an ordinary relatively harmless SSRI?

For weird political reasons, tianeptine, a well-regarded foreign antidepressant, is not a prescription drug in the United States. I think it is still illegal, but everyone sells it all over the Internet, even respectable sites that wouldn't dare sell the normal prescription antidepressants. It has a large Internet fan club that swears by it, and its side effects are less than those of many other antidepressant classes. It is available from nootropicsdepot, a supplier I regard pretty highly. If you can't just buy normal supplements from a health food store like a normal person, it is very likely your best bet.

V. Other things

The average length of an untreated depressive episode is six months. People who have one depressive episode have an 80% chance of having another sometime in their life. The average person with major depressive disorder gets four depressive episodes during their lifetime. Take these statistics into account when deciding how proactive you want to be with treatment.

People with depression have about a 2% chance of going on to commit suicide, though obviously this varies a *lot* with the intensity of depression. Most people who attempt suicide later regret it and go on to live enjoyable lives. If you are feeling suicidal, there are suicide hotlines operating both by phone and by online chat. You can also go to any emergency room and ask for help. If you have health insurance, this will probably be covered. Warning: this will quite possibly result in committment to a psychiatric hospital.

If you are feeling very suicidal and don't trust yourself not to attempt something, going to a psychiatric hospital is probably your best bet. The average length of stay at an average hospital is three to seven days. Two weeks is rare. A month is totally unheard of. You usually get intense attention from doctors, nurses, and social workers, and there are strict regulations giving you certain rights (ie you can't be denied necessities of daily living and you can't be locked up in any kind of restraints or solitary confinement without extremely good documentation that you are very dangerous). Psychiatric patients cannot be given non-emergency medication or other treatment against their wishes without a court order; psychiatrists are usually very good at pressuring patients to take medications in some kind of tricky ways, but in practice few of them will go through the difficulty of trying to get the court order unless the patient is extremely dangerous. People who complain about psychiatric hospitals most often complain that they are noisy (true), that they can't leave when they want to (true), and that they are around some scary people (true, though very few are actually violent). Good psychiatric hospitals will have procedures in place for trying to minimize these problems; for example, there is a big movement to switch to private rooms. In my experience most suicidal depressed patients who go to a psychiatric hospital become less suicidal very quickly, end up on a good regimen of appropriate antidepressant medication, and are glad they went.

About a third of patients will recover completely on their first antidepressant within three months. Another third will get somewhat better (>50% decrease on some test of depressive symptoms). Another third will have no benefit.

But outcomes get better the more stuff you try. According to STAR*D, which tested an algorithm a lot like the one I mentioned above, by the last step of the algorithm 70% of patients experienced complete remission, and many more experienced significant symptom reduction.

This is not a *resounding* victory, because the average depressive episode only lasts six months and the study took more than six months to complete, which makes it really hard to figure out how much of the improvement was due to the drugs. The answer appears to be “some, maybe”. But I will add that patients who recovered because of the study drugs had decreased chance of relapse compared to patients who recover by waiting it out.

I think the important lesson here is that with sufficient work depression either is treatable, or it will go away on its own before you get a chance to finish the treatment algorithm, which is annoying for researchers but probably pretty acceptable to the patient. And once it does, you know what drug works for you, you can sometimes stay on it to decrease chance of relapse (maintenance treatment is a totally different ball game I’m not getting into here) and you can restart first thing when you start feeling depressed again.

The most important thing I’m writing this for, and the lesson I want to hammer home, is that if your doctor just gives you an SSRI and tells you to stay on it even when it clearly isn’t working, *there are other options*. If your depression is seriously impacting your life, you should explore them, either with your doctor or with a replacement doctor who is more willing to help or – if necessary – on your own.

Money On The Ground

I hear the following joke far too often:

> Economist 1: Look, there’s $20 on the ground!
> Economist 2: No there isn’t. If there were, someone would have picked it up already.

It annoys me because it makes fun of a point of economic theory which, when phrased a little more realistically, is not only clearly correct but actually useful *even in the context of this analogy*.

Imagine the first economist had said “There’s $20 on the ground in the middle of Times Square, and it’s been there all week.”

It’s not so unlikely that you’re the first person to notice a bill that has just been dropped on the ground. It *is* so unlikely that a bill on the ground would go unnoticed for an entire week in Times Square, and if it looks like it has, you should start wondering if something more sinister is going on, like somebody spray-painting a picture of a bill on the sidewalk.

But that’s exactly what the original economic principle is trying to tell you!

So if there is what seems to be a drop-dead simple, no-risk money-making opportunity that no one has picked up on, *and it’s been available more than a very short amount of time*, expect it to be something more sinister.

Ground Morality In Party Politics

My name sounds a lot like Scott Aaronson's and I get confused for him a lot. I try to encourage this confusion, since it can only increase people's opinion of me. So let me propose a tool for investigating morality through algorithmic systems very similar to Aaronson's recent post on eigenmorality.

I say we use DW-Nominate.

I wrote about DW-Nominate before. It's the tool political scientists use to calculate Congresspeople's position on the political spectrum. Whenever you hear an alarmed-sounding voice on a black-and-white attack ad say something like "Senator Schmendrick is the third most liberal senator in Congress," chances are they used DW-Nominate to calculate "third most liberal".

The system is beautifully elegant. They take a Congressperson's votes on all the issues and compare them to other Congresspeople's votes to find blocs of Congresspeople who tend to vote together. Then they do factor analysis stuff to see how many dimensions of "similar voting" there are. They end up with something that looks a *lot* like the traditional left-right dimension and *occasionally* a Northern-US-vs.-Southern-US-dimension that doesn't always matter that much. Then they use Congresspeople with multi-decade careers to bridge the gap between current Congresses and past Congresses, and use dead Congresspeople with multi-decade careers to bridge the gap between past Congresses and even-further-in-the-past Congresses, so that they can compare any Congressperson, living or dead, to any other Congressperson, living or dead. They also get the opportunity to evaluate *bills* as liberal or conservative, based on whether liberal or conservative Congresspeople support them.

And the neat thing about it is that at no point did they enter into the system that it was supposed to give “left” vs. “right” – or even that it was supposed to come out with only one major grouping. It could have found that actually Democrats and Republicans vote much the same, but men always vote with other men and women with other women, or that the real difference was a religious worldview versus a secular worldview or whatever. Instead they found that our notion of left and right emerges naturally from the data, even if you’re not looking for it, and that this transcends party lines – ie some Democrats are further left than other Democrats, and this has consistent effects. If you really wanted, you could use this to rate whether, say, cutting carbon emissions vs. gun control is a more truly leftist cause, by seeing which bills get more heavily supported by leftists.

And I wonder what would happen if you tried DW-Nominate with *moral* decisions.

Now there’s a very boring interpretation of that proposal, which is that we hand a hundred people a hundred different multiple-choice questions on moral dilemmas, and then use factor analysis stuff to see if we can divide them into groups. I bet we’d come out with something a lot like “utilitarians vs. deontologists vs. virtue ethics”, or maybe something more like “religious vs. secular” or maybe “people who use all Haidtian foundations vs. people who just use care and fairness”. Actually, forget what I said before, this would already be a quite interesting thing to do and somebody should do it.

But I would be much more interested in a (much harder) naturalistic experiment. What if we took the *real* decisions people engage in? I'm not even talking about obviously morally charged decisions like whether to have an abortion, I'm talking about things from "what college major should I have?" all the way down to "do I drink alcohol at age 17?" to "do I call my parents tonight like I promised I would, even though I'm very tired?"

A lot of people have to go through very similar decisions, which allows DW-Nominate style ranking. There would be some wiggle room in deciding which two decisions were equivalent (is the person who decides not to call loving parents making the same decision as the person who decides not to call abusive parents?), but let's say we get a panel of raters to decide among themselves which decisions are equivalent and throw out any they can't agree upon. This isn't meant to be Pure And Objective here, only statistically useful.

In the same way that ten thousand Congressional votes, suitably analyzed, naturally group people into two categories that look to our trained eyes like Left and Right, would ten thousand little life decisions, suitably analyzed, naturally group people into two categories that look to our trained eyes like Good and Bad?

If so, it would be pretty easy to tell who the best person was, in the same way we can identify the most liberal member of Congress. We could give them a nice little award. Even better, it would be pretty easy to tell which option on each decision is more moral, for the same reason DW-Nominate can tell us that supporting gun control is more liberal than opposing it.

Suppose that we learned that one factor that naturally fell out of the data included giving money to the poor, supporting one's aging parents, never committing violent crimes, avoiding ethnic slurs, conserving water and electricity, helping one's friends when they were in trouble, and everything else we traditionally think of as good moral choices. And suppose this factor was heavily, *heavily* associated with being pro-life, to the same degree that the "liberal" factor in DW-Nominate is heavily associated with gun control. Would this provide some evidence in the debate over abortion? I'm not sure, but it would sure get me thinking long and hard about it.

I mean, we would probably also find some really silly things . Like that our moral factor loads on not getting tattoos of flaming skulls, ie the decision to get a tattoo of a flaming skull clusters with lots of immoral decisions. Presumably we would want to be able to say that getting a flaming skull tattoo is not itself immoral, but is correlated with immorality. But then we might as well say the same thing about being pro-life. Indeed, maybe everything religious will end out correlated with morality for religious reasons. We'd probably have to sort through this and fight a bunch of interminable correlation vs. causation debates.

But then there are areas where this could really shine.

I think this might solve a problem that Aaronson thought was unsolveable in his proposed algorithms. He said that in a world that was completely backwards – for example Nazi Germany – where everybody thought right was wrong and wrong was right, any moral sorting algorithm will give backwards results because it has to start with majority opinion in some sense. His example was that a PageRank style algorithm where people-believed-to-be-moral are the ones people-believed-to-be-moral believe are moral would fail, because most Nazis would believe that the high-ranking Nazi authorities were moral, and then the circle would complete with the high-ranking Nazi authorities getting to determine who the moral people were.

I think DW-Nominate might go part of the way solving that problem. Consider three different things we might find if we DW-Nominated Nazi Germany:

1. There is a General Factor of Morality, which includes giving to the poor, caring for your aged parents, cooperating with your neighbors, et cetera. People high in this General Factor of Morality are much more likely to oppose Nazi policies and hide Jews in their attics.

2. There is a General Factor of Morality, which includes giving to the poor, caring for your aged parents, cooperating with your neighbors, et cetera. People high in this General Factor of Morality are no more likely to hide Jews than anyone else, or maybe *less* likely to hide Jews.

3. There are multiple dimensions of morality. One dimension is something like “prosocial in-group patriotism” and captures things like paying your taxes on time, going without luxuries in order to help the war effort, and sending nice care packages to the troops. Another dimension is something like “willingness to go against consensus when it’s the right thing to do” and would include whistleblowing against corruption and being a passive resister to unjust wars. Hiding Jews in your attic might be negatively correlated with the first factor but positively correlated with the second factor. Universally beloved things like giving to the poor and caring for your aged parents might load about equally on both factors, or be a third factor, or whatever.

If Hypothesis 1 were true, that would be *super interesting*. It would suggest there's something kind of objective about morality. Also, we could make it do *work*. Like we could go around to the Nazis, and say "Look, you agree that helping the poor is moral, right? And caring for your aged parents? Well, now that we've established what morality is, we have bad news for you. You don't have it. Moral people are much more likely to oppose you. So stop doing what you're doing." This might actually work. Or if it didn't, then when World War II ended and everyone agreed they *should have* listened to the General Factor Of Morality, then maybe after ten or twenty iterations of this people would start listening *eventually*.

If Hypothesis 2 were true, that would also be super interesting, albeit disappointing. It would mean that morality probably isn't very objective, and that our moral positions are a lot closer to random than we want to believe. If being moral in every other way we can think of had minimal correlation with being moral in the particular way of saving Jews from the Nazis, it would mean that there was no consistent basis to morality and it was just a hodgepodge of popular positions. Or that if there was a philosophically consistent basis, it has little to do with how it's practiced in the real world.

If Hypothesis 3 were true, that would be very boring, but possibly still worthwhile. Like we could have debates on whether Factor I Morality is more important than Factor II morality, and what to do when they contradict each other, and these debates would probably be more interesting than our current more vague debates on things like "what do you do when your duty and your moral intuitions conflict?"

I don't really have some grand plan for how this could be used to solve everything or how a utopia could be created around it (although now that I mention it, if we can easily identify the most moral people in a population, they would make good candidates for judges and other high officials, though perhaps not legislators or executives).

I just think it would be fun to study.

The Art Of Writing Randian Monologues

I mentioned on Twitter yesterday that Ayn Rand's book *The Art of Fiction* contained advice on how to determine the proper length for characters' philosophical speeches. Enough people wanted to know what it said that I've copied it out:

> Such an issue as "I always decide for myself" versus "I go by the opinions of others" is extremely wide. If two characters started discussing it out of a clear sky, that would be sheer propaganda. But in the above scene, the two men are stating an abstract issue as it applies to their own problems and to the concrete situation before the reader's eyes. The abstract discussion is natural in the context, and, therefore, almost unnoticeable.
>
> This is the only way to state abstract principles in fiction. If the concrete illustration is given in the problems and actions of the story, you can afford to have a character state a wide principle. If, however, the action does not support it, that wide principle will stick out like a propaganda poster.
>
> How much philosophy you can present without turning into a propagandist, as opposed to a proper fiction writer, depends on how much of an event the philosophy is covering. In the above scene [a conversation between Roark and Keating from *The Fountainhead*] it would have been too early for the two boys to make more of a statement than they did, even though the issue stated is independence versus second-handedness, which is the theme of the whole book.

> Given what is specifically concretized in the scene, one exchange of lines is enough abstract philosophy.
>
> A speech like John Galt's in Atlas Shrugged would have been too much for Roark's courtroom speech in The Fountainhead. The events of The Fountainhead do not illustrate as many issues as do the events of Atlas Shrugged.
>
> To judge how long a philosophical speech should be, go by the following standard: How detailed and complex are the events which you have offered to concretize the speech? If the events warrant it, you can make as long a statement as you wish without taking the reader outside the framework of the story.

This would be a good time to mention that I'm trying to learn to write fiction a little better, so I can improve upon my occasional short stories. I'm not using Rand because I think she's The Best Writer but because a friend had her book and gave it to me for free. So far my experience has been that people spend forever talking about overarching themes I'm not interested in like How To Set Up A Grand Conflict, and almost no time talking about the things I really want to know, like how to have characters go outside without something very abrupt and boring like "And then he went outside". Or if I'm writing Lord of the Rings, I don't want to know how to write the climactic scene at Mount Doom, I want to know how to get Frodo through two thousand miles of swamps without just writing "And then he walked through another five hundred miles of swamps, it was very wet and icky and there were probably fights with giant bug monsters" four times.

If anyone has recommendations for other good books on fiction-writing (preferably by beloved authors, so I know they’re legit), let me know.

Court-ing Disaster

[EDIT: Some lawyers in the comments say various factors about this case sound very fishy. I know what I heard at the lecture, but it’s starting to seem more likely that the lecturer made this whole thing up or at least seriously embellished it. I have ensmallened the post below as a sort of retraction and apologize for believing everything I hear at a large and well-known hospital system’s extremely official medical education lectures which, I may add, are the same lectures that provide the information your doctor is using to decide on treatment for your life-threatening illness]

There are lots of tips for the homeless about how to get free stuff, so here’s mine: buy a white coat ($15) and a stethoscope ($20). Then, every morning and afternoon, walk into your local hospital and look for signs directing you to something like “grand rounds” or “morning report” or “noon conference” or “continuing medical education”. All of these phrases mean that if you listen to a boring person talk about boring things, you will get free food.

(if somebody starts looking suspicious and asks why they never see you in the rest of the hospital, or why you never seen to be treating any patients, tell them you’re a radiologist)

This tip works even better if you are a real doctor – you don’t even need to waste $35 on the coat and stethoscope! And that was how I found myself in a lecture on how malpractice cases work.

The answer is: not very well.

The speaker, a neurologist, described how he treated a stroke patient. From his description – and I admit, I have only his own words to go by – it sounded like he did a really good job. Didn't just do it competently, but went above and beyond to make sure that he got the best treatment possible every step of the way.

Unfortunately the guy still ended up permanently disabled, because strokes are bad and even the best treatment doesn't always ensure a good outcome.

The family decided to sue the neurologist. Upon being told they needed a reason, they and their lawyer came up with "he didn't give the patient heparin immediately upon presentation".

I don't want to turn this entire piece into a review of evidence-based stroke treatment, but suffice it to say that heparin is not usually used for stroke. And anticoagulants like heparin should *never* be given until a haemorrhagic stroke has been radiologically ruled out, or they might kill the patient. This is an extremely basic point. All stroke doctors know it. I know it. Any medical school students who have been paying attention in class know it. This is not difficult or controversial.

My point is, the lawyers came up with a totally bogus claim that was just barely plausible for people who know nothing about medicine.

So the neurologist was pretty confident that he would win this one easily, and he and his lawyer collected a giant pile of textbooks and studies saying he was definitely right on this specific medical point.

The judge in the case refused to allow any of the textbooks or studies to be presented because they were "hearsay".

Then the plaintiff’s attorney produced a stroke expert, a leading doctor who had helped write the guidelines for stroke treatment, to come up and say that heparin was *the best* treatment for stroke and it was malpractice for the neurologist not to have given it immediately, even before the radiology results were in.

Most of you are not in medicine and do not understand how creepy this is, but imagine that some law suit hinged on perpetual motion. In fact, let’s say this is a law suit against you – you’re an engineer, and for some reason someone has sued you for not believing in perpetual motion, and if you lose your career is over. And the plaintiff produces a Nobel Prize winning physicist who testifies that perpetual motion is *super easy* and happens *all the time* – and in his opinion any engineer whose machines don’t produce infinite free energy is guilty of malpractice and deserves to lose his license and go bankrupt.

That’s *kind of* what this was like.

So the defense attorney gets to cross-examine this guy, and points out that the expert wrote the guidelines for stroke treatment. And the guidelines for stroke treatment don’t mention heparin at all.

The expert says well, he didn’t write *all* of the guidelines, maybe someone else wrote that part.

The defense attorney asks why, if he thought heparin was such an important part of stroke treatment, he still signed his name to the guidelines, and didn’t even publish an article somewhere else mentioning that he wished heparin had been included?

The expert says well, you’ve got to choose your battles.

Then the defense attorney calls up their expert, who is *also* an award-winning prestigious expert stroke doctor, and who says he doesn't know what the heck is going on here but everyone knows heparin is not a first-line stroke treatment and this is stupid. Then the plaintiff's attorney gets to cross-examine him and comes up with BS reasons to sow confusion about his story.

So now the jury has just heard two experts say two different things and they don't know what to believe. So for them it's kind of a toss-up.

Luckily, one of the jurors is a nurse, and informs the other jurors that if you know anything whatsoever about medicine, the plaintiff's case makes no sense. She convinces the other jurors to acquit. So, all's well that ends well, right?

Not *really*. Because according to the neurologist, the plaintiff's attorney met with him on the side and offered him a deal.

The deal was that, before verdict was rendered, the neurologist should pay the plaintiff's attorney $30,000, which would cover his costs in prosecuting the case and in hiring the medical expert who said heparin was the mainstay of modern stroke treatment. In exchange for the $30,000, if he won, the attorney would ask for damages of exactly $800,000, which was the maximum that the neurologist's malpractice insurance would pay. If the neurologist *didn't* pay the $30,000, then if the plaintiff won he would ask for damages of $2 million. The malpractice insurance would pay $800,000, and the neurologist would be on the hook for the other $1.2 million. Which would easily force him to sell his practice and leave him bankrupt.

Under the circumstances, of course the neurologist payed the $30,000. So, although the plaintiff lost, they still got their costs covered and they still broke even. The medical expert who testified for the plaintiff still got $10,000 just to fly out for a day and spout an hour's worth of lies, which is a pretty good deal and more than enough to ensure there will be unscrupulous experts willing to testify to any old thing in the future. The only person who didn't get anything was the disabled stroke patient – but he hadn't put any work into the lawsuit, so he too managed to break even.

I don't know if this is a true story. I just have the neurologist's word for it, and he seemed to have a Paul-Bunyan-sized axe to grind. **[EDIT: A lawyer in the comments says the story about the plaintiff's lawyer is either false or illegal. Another lawyer suggests it might be a version of something called a high-low agreement]**.

But it sure left a really gross taste in my mouth.

I recall what people tell me about the atrocities that constantly go on in prisons. The terrible conditions, the rapes, the beatings, the people frying to death in 110 degree cells without air conditioning. And the explanation offered is: most people don't expect to end up in prison, so they have no selfish motive to care.

And I think about what people tell me about the court system. My mother was a juror in a murder trial once. She tells of jurors who didn't listen to the evidence, who were asleep half the time, who *made it all the way to the final deliberation without being aware that the murder victim was dead and this wasn't an assault trial*. And it's been a while since I last posted the study finding that parole hearings depend as much on when the judge ate lunch as they do on the severity of the crime committed.

I wonder if, in the same way nobody expects to find themselves in prison, nobody expects to find themselves in court. Or at least, the probability of finding yourself in court is sufficiently low that no one has the energy to solve the collective action problem of getting a thousand people with a tiny chance of ending up in court to petition to make the courts fairer. So lawyers and witnesses and plaintiffs can do whatever the heck they want, because everyone assures themselves they’ll never be a defendant and so they have no reason to care.

The Placebo-Singers

In medical school, I was always told that the word “placebo” came from the Latin “I will please”, because patients tried to get better to please doctors, or the pill was just there to please patients, or something like that.

But everything about placebos is controversial and fascinating, and the etymology is no exception.

The most common funeral rite of the Middle Ages began “Placebo Domino in regione vivorum” – “I shall please the Lord in the land of the living”. The rite itself became known as the Placebo, and those involved as placebo-singers.

At some point “placebo-singer” acquired an ill reputation and became used as an insult. In Canterbury Tales, Chaucer writes: “Flatterers are the devils chaplains, that sing ever ‘Placebo'”. And during the English Civil War, a cleric was described by one of his enemies as “a placebo singer at Lincoln’s Inn, and afterwards a turncoat at Bridgefoot”. An 1822 dictionary of proverbs defines “to sing placebo” as “to endeavour to curry favor”.

There are a couple of explanations of how exactly this came to be. I think the most likely is a sort of intersection between the old term “to sing placebo” as in a funeral, and the literal meaning “I will please” – thus someone who “sings placebo” gradually shifts from someone who literally sings a hymn to somebody who tries to please, sort of as a pun. One modern sort-of equivalent that comes to mind is the use of “gaylord” as a playground epithet for gay people because it sounds like it should mean that, even though originally it was a given name derived from the French name Gaillard.

A second theory says that it was meant as a sort of insult to priests, who were always saying incomprehensible things to impress the overly credulous. This one doesn't strike me as very believable, but it's worth noting that another priestly Latin saying, *hoc est corpus* meaning "this is my body", *might* through a similar process have mutated into the word "hocus pocus" meaning "a dumb attempt at magic for the gullible". But this is just as convoluted an etymology as placebo, and probably not much stock should be placed in it – though reading that article did teach me about a Norse demon I didn't know about before, so it wasn't a total loss.

A final theory, maybe the most interesting, is that unscrupulous people would show up at random funerals, sing the placebo with everyone else, and loudly mourn the deceased – intent on demanding a share of the post-funeral feast afterwards. As such, placebo-singer came to mean not only flatterer, but imitator or phony – with obvious implications for the medical usage. Wikipedia seems convinced of this one, but there's no source cited and I can't find it in any primary.

At some point in the late 18th or early 19th century, it became associated with medicine, although it's not clear exactly how. It seems to have had the sense of a popular but not very effective treatment, either derived from the placebo singers of above or from an alternate translation of placere, "to be popular". Compare to the word "nostrum", which is the Latin word "ours" and indicates popularity without a whole lot of effectiveness in a very similar way. This is the British Medical Journal's position on the matter – at least in 1999. A 2013 update makes a counterargument from Latin grammar – why isn't the name the more appropriate third-person "placebit", meaning "it will please"? 18th century doctors were generally grim people who spent their days watching people die of cholera. They wouldn’t have come up with some cutesy word where the pill is depicted as talking about itself. So maybe it comes from the funeral rites after all.

Annnnnnnnyway, this whole line of thinking started as a way of procrastinating on the research I should be doing: trying to figure out what the heck is going on with placebos in depression. Asbjørn Hróbjartsson, who presumably studies biostatistics in between quests to slay frost giants, finds minimal evidence of placebo effects anywhere, and particularly cites three studies on depression that found no difference between a placebo and an untreated group. Wampold et al disagree and reanalyze the data to find placebo effects in a couple places, but they’re not very impressive and Hróbjartsson is *definitely* not impressed.

Most antidepressant studies parrot the conventional wisdom that the d=1.5-or-so effect size of antidepressant treatment consists of a placebo effect of 1.0 and a drug effect of 0.5. But there is hardly any attempt to think critically about the placebo part. How do we know it’s a drug placebo effect, and not just depression getting better on its own if you wait long enough, or people feeling special and liked because someone is doing a study on them?

Irving Kirsch is the only person I have seen even try to address this, and his methodology is audacious to say the least – comparing no-treatment arms in psychotherapy trials to treatment arms in totally different pharmacology trials. He finds that about 2/3 of the apparent placebo effect is indeed due to a drug placebo effect and 1/3 to people-eventually-get-better. The question is whether we want to trust him over Hróbjartsson’s small but more focused trials finding no drug placebo effect.

This is pretty important because there are a lot of depressed patients for whom the benefits of antidepressants might not be worth the costs, or who just don’t want to take them. If the VERY LARGE observed placebo effect in depression is in fact a real pill-based placebo effect, then it becomes vital to get some pill into these people right away. Folic acid, which has enough evidence as a depression treatment to be believable but which is also dirt cheap and very safe, would be a good choice.

But if drug placebo effects don’t do anything and it’s just people getting better over time, then it’s not worth pushing folic acid or whatever quite so hard.

SSC Gets World Cup Fever

I’ve been following the World Cup lately. Not *watching* it, of course. I wouldn’t enjoy that at all. But *following* it has just the right combination of overly complicated procedures and total lack of physical action to appeal to the sort of person who has read like half a dozen LARP rulebooks but never gotten around to actually LARPing.

For example, as best I understand it, America will advance to the semiquarterpseudohyperfinals if it beats Germany on Thursday, or if Ghana ties with Portugal, or if Portugal beats Ghana by a sufficiently small amount, unless Uruguay loses to Italy or ISIS takes Baghdad, but only if the stock market goes up by an even number of points or the moon is in waning gibbous.

Suppose that the strength of different teams were interpreted as input and the eventual winner (and perhaps point spread) as output. And suppose winning a World Cup raises morale and resources devoted to soccer in a particular country, therefore making them stronger during the next World Cup – ie a sort of primitive feedback mechanism. How complicated do the advancement rules have to be before the tournament acts as a sort of primitive neural net? Could a sufficiently large soccer tournament achieve sentience? Could it feel suffering? I’m not sure, but these sorts of questions are definitely the level at which I enjoy following sports.

It has also come to my attention that if America and Germany tie in their upcoming game, they *both* advance to the semiquarterpseudohyperfinals – in other words, the payoff matrix heavily penalizes a team for losing as opposed to tying, but *winning* as opposed to tying gives no benefit save transitory national glory. It’s supposed to be bad form to start the game by shouting “I PRECOMMIT NOT TO KICK THE BALL IF YOU WILL DO THE SAME!” and then have both teams stand motionless on the field for ninety minutes. So the question is whether the two teams can communicate to each other their desire to establish a mutually beneficial equilibrium of not scoring goals, using only signals to that effect that are plausibly deniable to third parties. A secondary problem is whether, having committed to such a plan, they can enforce it against the temptation to defect and score a goal at the last second for national glory. The usual solution to these kinds of problems is to establish a government to enforce a Platonic social contract while also establishing a communally enforced norm of virtuous behavior at the individual level, but I guess that might be a tall order for a ninety minute game. Still, it is far from impossible: Germany has managed it before, and for what it’s worth the American coach this year is a German guy.

(as far as I can tell, there is no study testing the percent of these games that end in ties compared to similar games that lack this incentive, but someone should check this out)

Anyway, it’s worth remembering that this is, by far, not the weirdest incentive structure to have played out in international soccer. From a description of the 1994 Caribbean Cup:

> Barbados needed to win the game by two clear goals in order to progress to the next round. Now the trouble was caused by a daft rule in the competition which stated that in the event of a game going to penalty kicks, the winner of the penalty kicks would be awarded a 2-0 victory.

With 5 minutes to go, Barbados were leading 2-1, and going out of the tournament (because they needed to win by 2 clear goals). Then, when they realized they were probably not going to score against Grenada's massed defence, they turned round, and deliberately scored on their own goal to level the scores and take the game into penalties. Grenada, themselves not being stupid, realized what was going on, and then attempted to score an own goal themselves. However, the Barbados players started defending their opponents goal to prevent this.

In the last five minutes, spectators were treated to the incredible sight of both team's defending their opponents goal against attackers desperately trying to score an own goal and goalkeepers trying to throw the ball into their own net. The game went to penalties, which Barbados won and so were awarded a 2-0 victory and progressed to the next round.

Book Review: The Two-Income Trap

A long time ago I wrote a kinda-tongue-in-cheek defense of keeping modafinil – a relatively safe and effective stimulant – illegal. My argument was that if *everybody* can use stimulants to work harder and sleep less without side effects, then people who work very hard and don't sleep will become the new norm. All the economic gains produced will go into bidding wars over positional goods, and people will end up about as happy – and with about as much stuff – as they have right now. Except the workday would be sixteen hours, the few people who can't tolerate the stimulants will be at a profound disadvantage, and when the side effects reveal themselves twenty years down the line, everyone is too financially invested in the system to stop.

In other words, in a sufficiently screwed-up system, doubling everyone's productivity is a net loss. The gains get eaten up by proportional increases in the prices of positional goods, and you're left with nothing except complete dependence on a shaky advantage that could disappear at any time.

I don't know how serious I was. But Elizabeth Warren makes almost the exact same argument in The Two-Income Trap, and I'm pretty sure she's very serious. At least, she used it as a platform that got her elected to the US Senate, which is a *kind* of serious.

So on the advice of Alyssa Vance, I decided to take a look.

I.

Warren's not talking about stimulants. She's talking about the effect of an extra family income – usually moving from a system where the husband works outside the house and the wife stays at home, to a system where both parents work outside the house. Like a stimulant that removes the need for sleep, this can be expected to double economic productivity and family income.

In practice it doesn’t, because wives usually earn less than their husbands, but it comes pretty close. The average family income in the 1970s was around $40,000. The average family income in the 2000s was around $70,000 (all numbers in the book and in this post can be considered already adjusted for inflation). The husband’s income didn’t change much during this time, so the gain was due mostly to the wife getting an extra $30,000.

If families now have twice the income of families in the 1970s – who themselves were usually pretty financially secure and happy – then people should be really secure and rich now, right? But Warren meticulously collects statistics showing that the opposite is true. Home foreclosures have more than tripled in the past generation —

[Sorry, I feel at this point I should mention that my edition of the book was published in 2004, so all of these statistics about how awful home foreclosures are and everything are before the housing bubble burst and before the Great Recession. All of these statistics were when we were supposedly in a boom economy. You can assume that now they’re much, much worse.]

— Sorry, where were we? Oh right. Home foreclosures have tripled in the last generation. Car repossessions doubled in the five years before the book was published. Bankruptcies have approximately quintupled since 1980. Over the same period, credit card debt has gone from 4% of income to 12%, and average savings have gone from 10% of income to negative.

Seventy percent of Americans say they have so much debt burden that “it is making their home lives unhappy”. In 2004, for the first time, “get out of debt” passed “lose weight” for Most Popular New Years Resolution.

So, Warren argues, the common-sense conclusion that a modern family making $70,000 is nearly twice as well-off as a traditional family making $40,000 clearly doesn’t hold. Why not?

II.

One thing that finally got me writing this up was a post on Bleeding Heart Libertarians which, like all posts on Bleeding Heart Libertarians and in accordance with the philosophy of the same name, was about how although libertarianism is commonly thought of as a heartless philosophy it can actually be reconciled with the care/harm-based ethic of deep compassion for the weak and needy.

Wait, sorry, actually it was about how we should cancel Social Security and let old people starve to death on the streets:

> The baby boomers spent their entire lives buying new cars they didn’t need, buying houses that were too big, taking extra vacations, splurging on eating out, and the like. They enjoyed a higher standard of living than they could really afford. Why? Because they figured that when they retired, they could just use their voting power to force younger generations to pay for their retirement. These selfish narcissists pretty much want to steal as much as they can from their children. So, while I, Jasper, and my good twin brother Jason put tens of thousands of dollars into index funds each year, thereby forgoing fancier cars, vacations, and the like, the selfish,

> narcissistic baby boomers laugh gleefully, knowing that they'll find a way to eat our nest eggs.
>
> Jason is of course a sensitive soul and feels bad for these boomers. Not me. I say let them die. They knew what they were doing, and they spent their entire adult lives making the wrong choice over and over and over again. Does starving on the streets seem too inhumane? No problem. You've read Logan's Run, right? Good idea, but wrong age limit.

This claim is pretty common. If true, it would explain the phenomenon cited above – that even with twice as much money, the Boomer generation is much less financially stable than their parents' generation. But in Chapter 2 of *Two-Income Trap*, "The Over-Consumption Myth", Warren tears it apart.

The Boomers "spent their entire lives buying new cars they didn't need"? Warren, page 47:

> When we analyzed unpublished data by the Bureau of Labor Statistics, we found that the average amount a family of four spends per car is twenty percent less than it was a generation ago. [Families spend $4000 more on automobiles in general, but instead of luxuries they are spending it on] something a bit more prosaic – a second car. Once an unheard-of luxury, a second car has become a necessity. With Mom in the workforce, that second car became the only means for running errands, earning a second income, and getting by in the far-flung suburbs.

In other words, it sounds like a family with two working parents requires two cars as a sound money-making strategy, but that Boomers compensate by spending less per car than past generations.

The Boomers “splurge on eating out”? Warren again:

> Today’s family of four is actually spending 22 percent less on food (at home and restaurant eating combined) than its counterpart of a generation ago.

The Boomers “buy houses that are too big?” Warren:

> The size and amenities of the average middle-class family home have increased only modestly. The median owner-occupied home grew from 5.7 rooms in 1975 in to 6.1 rooms in the late 1990s – an increase of only half a room in more than two decades…the data showed that most often that extra room was a second bathroom or third bedroom.

The BHL article doesn’t mention appliances, but in case you were worried, moderns spend 44% less on appliances than their parents’ generation, which is partly compensated for by a 23% increase in home entertainment (probably things like DVD players). Warren says that:

> This same balancing act holds true in other areas. The average family spends more on airline travel than it did a generation ago, but less on dry cleaning. More on telephone services, but less on tobacco. More on pets, but less on carpets. And when we add it all up, increases in one category are offset by decreases in another. In other words, there seems to be about as much frivolous spending today as there was a generation ago…Sure, there are some families who buy too much stuff, but there is no evidence of any epidemic in overspending – certainly nothing that could explain a 255% increase in the foreclosure rate, a 430% increase in the bankruptcy

> rolls, and a 570% increase in credit card debt. A growing number of families are in terrible financial trouble, but no matter how many times the accusation is hurled, Prada and HBO are not the reason.

Curiouser and curiouser. Today’s families earn twice as much, spend the same amount on luxuries, yet are much less financially secure.

III.

So as to not keep anyone in suspense: the problem is nice suburban houses in good school districts.

Around a vague period of time centering on the 1970s, a couple of things happened.

First, the cities became viewed, rightly or wrongly, as terribly unsafe ghettos full of drugs and gangs and violence. As far as I can tell, this is a pretty accurate description of the 70s, although things have gotten a little better since then. Families didn’t want their children living in terribly unsafe ghettos full of drugs and gangs and violence, so they moved to the suburbs. Warren gives the testimony of a suburban mother:

> We were close to The Corner and I was scared for my sons. I didn’t want them to grow up there. I wanted something away from this neighborhood to get my boys out to better schools and a safer place. The first night in [my new] house, I just walked around in the dark and was so grateful…at this house, it was so nice and quiet. My sons could go outdoors and they didn’t need to be afraid. I thought that if I could do this for them, get them to a better place, what a wonderful gift to give my boys. I mean, this place was three thousand times better. It is safe with a huge front yard and a backyard

> and a driveway. It is wonderful. I had wanted this my whole life.

Second, education started to be really, really important. As Warren puts it:

> A generation or so ago, Americans were more likely to believe that there were many avenues for a young person to make his way into the middle class, including paths that didn’t require a degree. I recall my parents encouraging me to attend college, since my grades were high and they hoped I might become a teacher one day. But they were equally pleased when my eldest brother joined the Air Force, my middle brother entered a skilled trade, and my youngest brother became a pilot – even though all three of the boys had given up on college. My parents’ views were pretty typical a generation or two ago. Education was valued, but no one in our neighborhood would have claimed it was the single most important determinant of a young person’s success.

Warren is a Harvard professor. Think about that for a second. How many Harvard-professor-producing-type families can you think of today who are also happy with three of their children getting non-college-degree jobs? As Warren puts it in what might be my favorite passage from the whole book:

> 97% of Americans agree a college degree is “absolutely necessary” or “helpful” compared with a scant 3% claiming that a degree is “not that important”. According to one recent poll, 6% of our fellow citizens believe the Apollo moon landings were faked. In other words, Americans are twice as likely to believe that man

> never walked on the moon as they are to believe that a college degree doesn’t matter!

Certain school districts are known to be vastly superior to other school districts in terms of test scores, college admissions, et cetera. Usually these are school districts inhabited by rich people with very high property taxes and therefore very high levels of per-pupil spending in schools – although we’ll get back to that eventually.

These school districts are positional goods. Not everyone can be in the best school district. Only the people willing to spend the most money on their houses can be in the best school district. But rightly or wrongly, people believe that being in the best school district is vital for their children to succeed and become Harvard professors, as opposed to gang members or drug addicts or menial laborers. As Warren puts it, good education is the ticket to the middle class. And being in the lower class is too horrible to contemplate.

People want the best for their children [citation needed]. They’re not going to say “Well, we aren’t as rich as those other people, so we should probably live in a crappy school district with other people of our approximate wealth level”. They’re going to leave no stone unturned. And there are two big stones available for modern middle-class families: working-motherhood and debt.

If your family earns $70,000 and the other family earns $40,000, you have $30,000 extra to convince the banks to give you a really big mortgage so you can buy a much nicer house and get *your* kid into Oak Willow River View Hills Elementary, while *their* kid has to go to City Public School #431 and get beaten up by scary gang members every recess.

On the other hand, this is *everybody's* cunning plan, so what you end up with is all houses costing a lot more, everyone working two jobs without any extra money, everyone burdened with massive debt, and everyone living exactly where they would have anyway.

Warren lists some points in support of her hypothesis:

> A study conducted in Fresno found that, for similar homes, school quality was the most important determinant of neighborhood prices – more important than racial composition of the neighborhood, commuter distance, crime rate, or proximity to a hazardous waste site. A study in suburban Boston showed the impact of school boundary lines. Two homes located less than half a mile apart and similar in nearly every aspect will command significantly different prices if they are in different elementary school zones. Schools that scored just 5% better on fourth-grade math and reading tests added a premium of nearly $4,000 to nearby homes, even though these homes were virtually the same in terms of neighborhood character, school spending, racial composition, tax burden, and crime rate.

A lot of the causal claims here are very complicated and iffy at best, but here are two numbers that cuts through a lot of the debate: between 1984 and 2001, the median home value of the average childless couple increased 26%; the median home value of the average couple with children shot up 78%. So families are spending a *lot* more on houses nowadays and the disparity seems to be heavily concentrated in families with children. Combine that with the observation that houses only have 0.4 more rooms today, and you get a pretty good argument that families with children are competing much more intensely on house location.

IV.

When Warren does a very unofficial Fermi-estimate style breakdown of what is happening to the extra $30,000 that modern two-income families earn over traditional one-income families, she thinks they are paying about $4,000 more on their house, $4,000 more on child care, $3,000 more on a second car, $1,000 more on health insurance, $5,000 more on education (preschool + college), and $13,000 more on taxes.

The taxes are not a result of higher tax rates nowadays, just a result of the family making more money and so having to give more money – plus maybe being in a higher tax bracket. The health insurance isn’t surprising either to anyone who’s been paying attention. And the $4,000 extra on the house is a big part of what she’s been talking about the whole time.

The $4,000 on child care, $3,000 on the extra car, and $13,000 on taxes are the results of the second income. Mom needs a car to get to work, the children need care now that Mom’s not home to look after them, and not only does Mom get taxed but Dad may move into a higher bracket. That means that of the $30,000 Mom takes home, $20,000 gets spent on costs relating to Mom having a job – meaning that Mom’s $30,000 job only brings in $10,000 in extra money.

The $5,000 on education is a bit more complicated. In Warren’s example family, it’s spent on preschool. She points out how a generation ago, practically no one went to preschool, whereas nowadays it is viewed as another one of those important legs up (“If little Madison doesn’t get into the best preschool, she’ll never be able to make it into the science magnet school, which means she’ll be unprepared for high school, which means Harvard goes out the window”). Warren points out that today two-thirds of American children attend preschool, compared to four percent in the mid-1960s. Once again, parents are told if they want the best for their kids they need to compete for good preschools:

> The laws of supply and demand take hold, eliminating the pressure for preschool programs to keep prices low. A full-day program in a preschool offered by the Chicago *public* school district costs $6,500 a year – more than the cost of a year’s tuition at the University of Illinois. High? Yes, but that hasn’t deterred parents. At one Chicago public school, there are ninety-five kids on a waiting list for twenty slots.

It’s a little bit sleight-of-hand-y to put that in the family budget as Warren does – preschool only takes up two years of a child’s life, for a total of four years per two-child family. But I forgive her because college expenses are higher and also need to be budgeted for. Also, she’s saying her $4,000 child care estimate is for one child, which means that once the second child is out of preschool she’ll need to be in child care as well, for an insignificant price drop.

So I think Warren partially supports her points. The second income goes partially to increased house costs due to bidding wars, partially to increased education costs due to bidding wars, and partially to supporting the ability to have a second income. In her (admittedly slightly cooked) model, the family’s discretionary income – what it has left to spend on variable expenses like food and luxury goods – actually *decreased* from the 1970s one-income family to the present, $17,834 to $17,045.

V.

In my essay on stimulants, I suggested that the benefits of the stimulants would be wasted on positional goods, leaving only the side effects. In the same way, Warren says the benefits of the second income are lost, but the side effects remain.

The most important side effect she talks about is the loss of flexibility.

One nice thing about having a non-working mother is that she can, on relatively short notice, become a working mother. This is especially true in the Old Economy where even people without much college education could get okay jobs.

In the old model, financially healthy families subsisted on one income, and financially unhealthy families put the mother to work to get back on their feet. The most common disasters were the husband getting fired or a family member becoming sick. If the husband got fired, then even if he could get a job relatively soon afterwards it might be at lower pay until he could work himself back up the totem pole. Suppose he loses his $40,000 a year job and can only find a $30,000 a year job. Luckily, as we already established a wife's second income can contribute $10,000 to the family. So she goes to work, they have as much money as they did before, and they are able to pay off their debts and continue to have a good quality of life.

Even if the wife doesn't go back to work, having a flexible person with lots of free time is a huge benefit. If Grandma gets very sick, the wife has a lot of time available to take care of her – whereas now, if Grandma gets sick, either one parent has to quit their job to take care of her (meaning that standard of living goes way down and the family is at risk of not being able to pay debts it took out when their prospects looked much higher) or Grandma gets sent to a nursing home, which is very expensive and *also* risks unpaid debts or loss of standard of living.

Last of all, it means that getting a nice suburban home is more important than ever. If in the old days children spent most of their time with their mothers, it might be possible for the mother to pass down important values like education and hard work to her children. When mothers have very limited time with their kids, schools and peer groups take over a lot of the socialization role. For example, a mother with a very young son might talk to him, read to him, take him to childrens' museums, et cetera, providing the crucial intellectual stimulation that children need at an early age to develop their full brainpower. If the mother works full-time, then it becomes really imperative to get the son into preschool to make sure he's not just sitting around staring at a wall and losing brain cells. If the mother isn't around much when the child is ten, it becomes a lot more important to be certain he's in a good elementary school that's teaching him the right values. If you can't watch your kid to make sure he's not doing drugs, it's more important his school be drug-free. And so on. I don't know to what degree any of these social psychological hypotheses are true, but the important thing is that people think they are and so the competition for nice neighborhoods and nice schools intensifies.

The last loss of flexibility Warren talks about is divorce. Something like a third of couples with children can expect to get divorced. Consider a scenario where a working single mother gets the house and custody over the children. If the house took two incomes to afford, she's not going to be able to afford to keep her house. Suggestions that the father be forced to pay more child support don't work – unless he pays 100% of his earnings to her, she's not going to have as much money as the couple did when they bought the house – and they deliberately spent every cent they could on the mortgage because if they didn't they would be outcompeted by people who did and their kids would end up in gritty urban school districts and never get into Harvard.

So Warren says that the reason so many families go bankrupt or get into debt is because the extra income doesn't make a difference, but the loss of flexibility *does*. Everything has been sunk into the home for risk of getting outcompeted. And that means when someone loses their job – and Warren calculates that in a two-income family, this will happen to one parent or the other about once every sixteen years on average – or costs go up even a little, there is no buffer room and the only solution is to go deeper into debt. That just adds *another* unpayable cost – interest – and means the whole thing can only end in bankruptcy.

In another of my favorite passages, Warren notes that if the myth of over-consumption was true – if the guy in Bleeding Heart Libertarians were exactly right – there would be no problem. In fact, she *encourages* families to overconsume as the road to financial health. She says families should save, but if they can't save, that should spend their money on restaurants, vacations, jewelery – anything but large fixed-income monthly costs like houses, cars, schools, et cetera. That way, when something goes wrong, they can easily just stop taking the vacations and be back to financial health. It's only when money is trapped in mortgage payments that can't be gotten rid of that things can get as bad as they are.

VI.

There's a chapter on debt. It's really cute. She's all like "Did you know there are things called subprime mortgages? And that some people think banks might give them out too easily? I sure hope this doesn't do something *bad* to happen."

I am pretty sure no modern reader needs this chapter, but it sure increases her credibility.

VII.

Oh, right, I'm supposed to have an opinion.

Let's start with the negatives. I don't think she does a great job of proving her housing-school-positional-goods theory. When she talks about school district effects on housing prices, she comes up with numbers like "a 5% difference on test scores add $4,000 to housing costs." Okay. That means, assuming linearity, that a 50% difference on test scores – which is way more than we could possibly expect schools to produce – would only add $40,000 to house costs. When house prices for the middle class are routinely around $200,000 to $300,000, that's just not enough to be causing the destruction of the American family.

Likewise, studies that look for effects of school district on house price – usually by looking at otherwise identical houses on either side of a school district line – generally find modest effects.

The whole area is really hard to research. Suppose Neighborhood A has lots of minorities, low house prices, and bad schools. Neighborhood B has few minorities, high house prices, and good schools.

You can tell a story where Neighborhood B's good schools raise land value, which prevents crime and pushes out minorities. Or you could tell a story where Neighborhood B's high land values push out minorities and increase property taxes which improve the schools. Or you can tell a story where Neighborhood A’s many minorities cause racist homebuyers to stay out, depressing land values, and also minorities tend to have worse school performance. Except in real life there are like twenty factors like this rather than three. Although lots of different studies try to control for confounders, that’s always hard and requires a lot of assumptions that might not necessarily be true.

There’s another problem, which is that the usual measure of school quality – standardized test scores – is not necessarily the one families are going to be looking at. Suppose only a few very smart people know where to look for standardized test scores. Maybe everyone else tries to guess at how good schools are. Maybe those people assume that schools with higher percent minorities are worse. Maybe they assume that schools in prettier neighborhoods with higher land values are better. In that case, studies could find all they wanted that test scores don’t correlate with home prices, because what’s actually happening is that high home prices are causing belief in school superiority which is causing higher home prices.

But a bigger problem here is that the average family only spends $4,000/year more on housing than they did a generation ago. Warren can talk all she likes about how that forces families to adopt a second job, but it’s really not a very big share of what the second job’s meager extra income is being spent on. The average husband earns $3000 more at his own job nowadays, which means that it would be possible in theory for him to soak up pretty much all of the extra housing cost. To say the wife gets a $30,000 extra job just to soak up $1,000 in extra mortgage money seems like a stretch, even though Warren does a good job of pointing out how many extra burdens this places on people.

But when you add positional education costs to the mix – preschool and college – it becomes a little more believable.

I guess it's just hard making the numbers add up. Suppose you have two kids, but they're not in preschool – or that you're indifferent to preschooling your kids versus having the mother take care of them. Then the costs of the mother getting her $30,000 job are $24,000 – $13,000 in extra taxes, $8,000 in child care, and $3,000 in a second car. Are mothers really so desperate they'll work full-time for the extra $6,000? Doesn't this whole model break down once the mother gets a raise and starts making $40,000?

VIII.

How about the good?

The good is that Warren backs all her points up with excellent statistics, is very good at explaining complicated economic things, and has exactly the right level of contempt for everyone in politics.

Her view on politics is very very close to my heart. My impression is that she thinks of it as noise. It's not good, it's not evil, it's something that you have to adjust for. Like, "Well, this would be a good policy but we could never pass it because the Left would throw a fit, this other thing is a good policy but we could never pass it because the Right would throw a fit, but I'm pretty sure this third thing would also help and not get anybody too enraged." For example:

> The politics that surrounded women's collective decision to migrate into the workforce are a study in misdirection. On the left, the women's movement was battling for equal pay and equal opportunity, and any suggestion that the family might be better off with Mother at home was discounted as reactionary

> chauvinism. On the right, conservative commentators accused working mothers of everything from child abandonment to defying the laws of nature. The atmosphere was far too charged for any rational assessment of the financial consequences of sending both spouses into the workforce. The massive miscalculation ensued because *both* sides of the political spectrum discounted the financial value of the stay-at-home mother. There was no room in *either* worldview for the capable, resourceful mother who might spend her days devoted to the roles of wife and mother but who could, if necessary, dive headlong into the workforce to support her family. No one saw the stay-at-home mom as the family's safety net.

(in case you're wondering, she doesn't recommend women leaving the workforce. She says families where both parents want to work should keep one of the two incomes in reserve by either saving it or spending it on non-fixed luxury items. She admits that this is unfair because they will have problems getting into the best school districts, but says it is the safest solution until the wider societal problems are fixed.)

As a result of her disdain for established partisan groups, she manages to totally transcend politics. I noticed that when the Bleeding Heart Libertarians article got up on Xenosystems, one commenter protested:

> The accusations of excess are no doubt sound but I always pause when someone mentions the housing excess of the boomer generation. They bought giant houses in suburbia, but how much of that was due to the lack of civilization in the city limits? If there was a sane enforcement of laws and no public schools or at least public schools where you didn't fear for the safety

> of your children would they have bought so many giant houses?

In other words, the commentariat of one of the larger reactionary blogs is *more or less* on the same page as the Democratic Senator being pushed by the liberal wing of her party to run for President.

Her proposed solutions are also all over the map. Yes, she pushes for taxpayer-funded universal preschool, which should make liberals pretty happy. But she also pushes for school vouchers, which she hopes will decouple school quality from housing prices and let people live wherever they want and still be able to get an acceptable education for their children. She even has a states' right style solution to one problem – she points out that banks used to be kept under control very well by state laws until the Supreme Court legalized free interstate commerce between banks which means all of them moved to the states with the fewest regulations and could not be kept under any control at all. In order to rein in banks again, all we need is for Congress or the courts to grant those powers back to the states.

And I will say one more thing in Senator Warren's favor. She often suggests non-free-market solutions, like regulating something or banning something or proposing the government spend money on something. Every time she does this, she says very clearly something like "I understand the free-market arguments against this, and why in general we would want to use the market to take care of these sorts of problems, but this is a case where there is a likely market failure because of reasons X, Y, and Z. I recognize there is a burden of proof on someone saying something is a market failure, so I will now proceed to meet that burden of proof with a lot of statistics."

People talk about dogmatic libertarians, but honestly this is *all I ever wanted from anybody*. Just an “oh, by the way, I have reasons for what I’m saying and they’re not just coming from a total failure to have ever grasped freshman economics.” I know it seems unfair to make people say it explicitly each time. But given the overwhelming number of people who say these things *exactly* because they never grasped freshman economics, it’s welcome a breath of fresh air.

I am sure if Warren ends up running for President, we will end up getting those ads where someone repeats “MOST LIBERAL SENATOR OF ALL TIME” on a black-and-white background, followed by saucy rumors that she once had a fling with Karl Marx.

But I for one intend not to believe them.

IX.

But aside from doing some legal work to solve the bankruptcy crisis, we need some science work as well. The question is: are good school districts really that important?

I can’t find great research on this at the school district level. The closest I can find is the teacher value-added research, which finds things like “At age 28, a 1 SD increase in teacher quality in a single grade raises annual earnings by about 1% on average”. I can’t find good data on how this adds up – for example, do twelve great teachers in a row increase earnings 12% (linear addition)? Do you need one great teacher to inspire you for life, and after that it doesn’t matter whether or not you have more (ie sublinear addition_? Or can multiple great teachers build on one another’s successes by not having to constantly go back and review things the students should’ve learned before (superlinear addition)?

I don’t think it matters, because it doesn’t look like there are very big value-added score differences between teachers at rich and poor schools.

What about district-level issues like superintendents? According to the Brookings Institute report, difference in school district competency explained only 1.1% of variance in student test scores. Difference in schools explained another 1.7%. Teachers explained 6.7%. The remaining 90.4% was explained by demographic factors (class, race, parent’s education level) and individual variation among students.

Teachers are kind of a crapshoot – as we saw before, going to a better school district doesn’t increase your chances of getting a good one much. So the sorts of things you can easily affect by choosing what school district to live in are 2.8% of your kid’s total variation.

The research on preschool is so complicated it would take ten posts of this size to get through it. It seems strongly beneficial for low-income children and of controversial benefit for higher-income children. I will try to route around the controversy like so: home-schooled children do much better on every measure of academic achievement than school-schooled children. Preschool is basically teaching kids to share and playing fun games with them. If the alternative to sending your kid to preschool is that they stay home with you and you teach them to share and play fun games with them, you are home-preschooling your child and can expect them to do much better than school-preschooled children. And if the reason there’s no parent at home with the child is that both parents need to work in order to earn enough money to send the kids to a good preschool…well, that’s just a *little* bit circular.

So I think that in addition to various legal and policy changes, there needs to be more of a scientific effort to confirm (or disconfirm) these suspicions and, if they turn out to be true, publicize them to a society that clearly believes the opposite.

I know that talking about genetics and IQ too much makes people mad. And a lot of people have asked me – why do we have to do this? It’s going to offend a lot of people, and give a lot of unsavory people a lot of ammunition, so even if we shouldn’t ban research entirely, why not exercise the virtue of silence and let the whole thing stay in a few obscure journals?

And one of many answers to this is – suppose you see some school districts in rich neighborhoods, and all of the children in those schools can do calculus and read James Joyce and get great high-paying jobs. And next door is another school district, in a poor neighborhood, serving poor kids, and those kids are struggling.

If you’re not intimately familiar with behavioral genetics and IQ research, it is *obvious* that the rich-person school is much better and that’s why all the children of the rich people are doing so much better. And you will do anything, make any sacrifice, to get your kid into that rich person school, and so you work a back-breaking job and gamble your family’s financial security, all because you want your kid to have the same opportunities those rich kids do.

If you *are* intimately familiar with behavioral genetics and IQ research, a separate possible explanation leaps to mind: the rich people made their money by things like going to college, which means they probably have higher cognitive ability on average than the poor people, and cognitive ability is 50% genetic so they pass that on to their kids, and so it’s no surprise at all to see the rich person school having smarter students. That doesn’t prove that if your child switches from the poor person school to the rich person school, she will switch from average-poor-school-outcomes to average-rich-school-outcomes, and it doesn’t even provide any evidence whatsoever that it will make her do even a smidgeon better. So maybe you should, like, not sacrifice your life for it.

I’m not saying the behavioral-genetics-informed view is correct here. That’s going to require a lot more research. But I’m saying if you at least agree it’s something we’re allowed to talk about, maybe it will pan out and do nice things like save you from the horrible zero-sum competition destroying your country’s middle class.

Because if it could be confirmed that preschool attendance and expensive school districts had low impact – or even a merely moderate amount of impact – on success for middle- to high- income children, then even in the absence of legal changes that would relax the pressure on everyone to spend more money than they have to get into the best preschools and best school districts.

X.

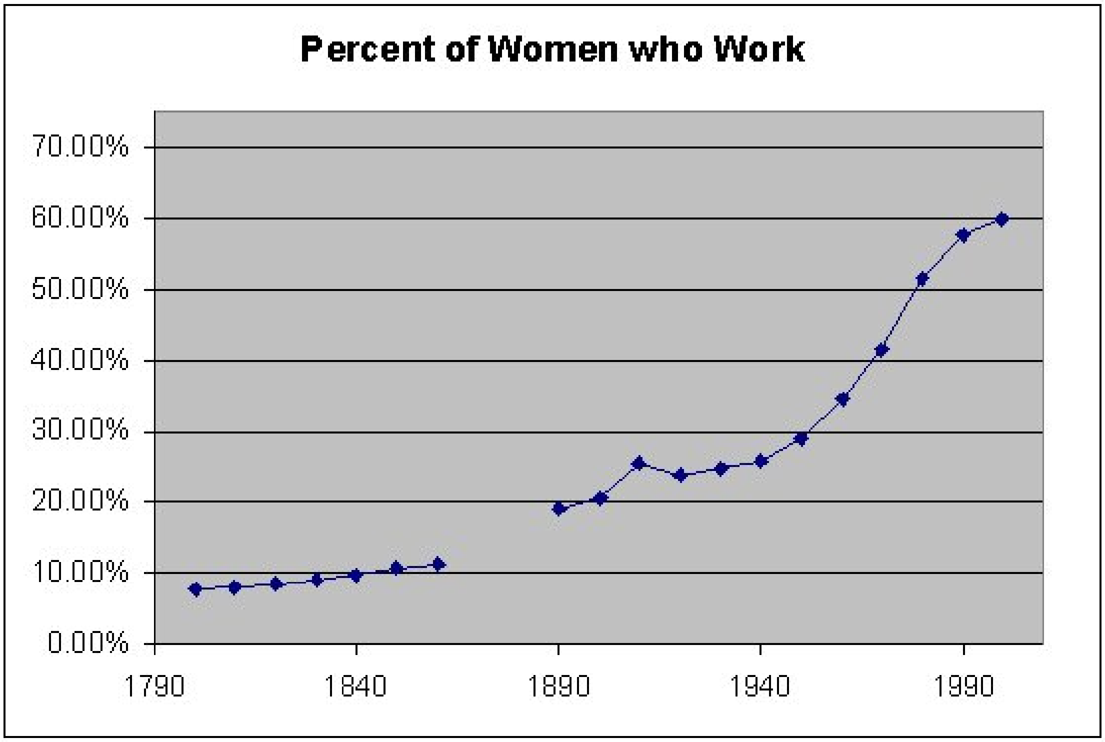
Overall I recommend this book. I think the conclusion comes on a little too strong but that it sheds a lot of light on a lot of trends and throws important statistics at you such that you read them. Equally importantly, it sheds a lot of light – in a positive way! – on somebody who’s becoming an important national figure. The chapter about her meeting with Hillary Clinton and the subsequent break between the two of them seems likely to take on a lot more meaning in the years ahead.

What I really want is Elizabeth Warren vs. Rand Paul 2016. Imagine a Presidential race when both candidates have very different but very consistent philosophies, and you'd be pretty proud to see your country run by either. Wouldn't *that* be a change?

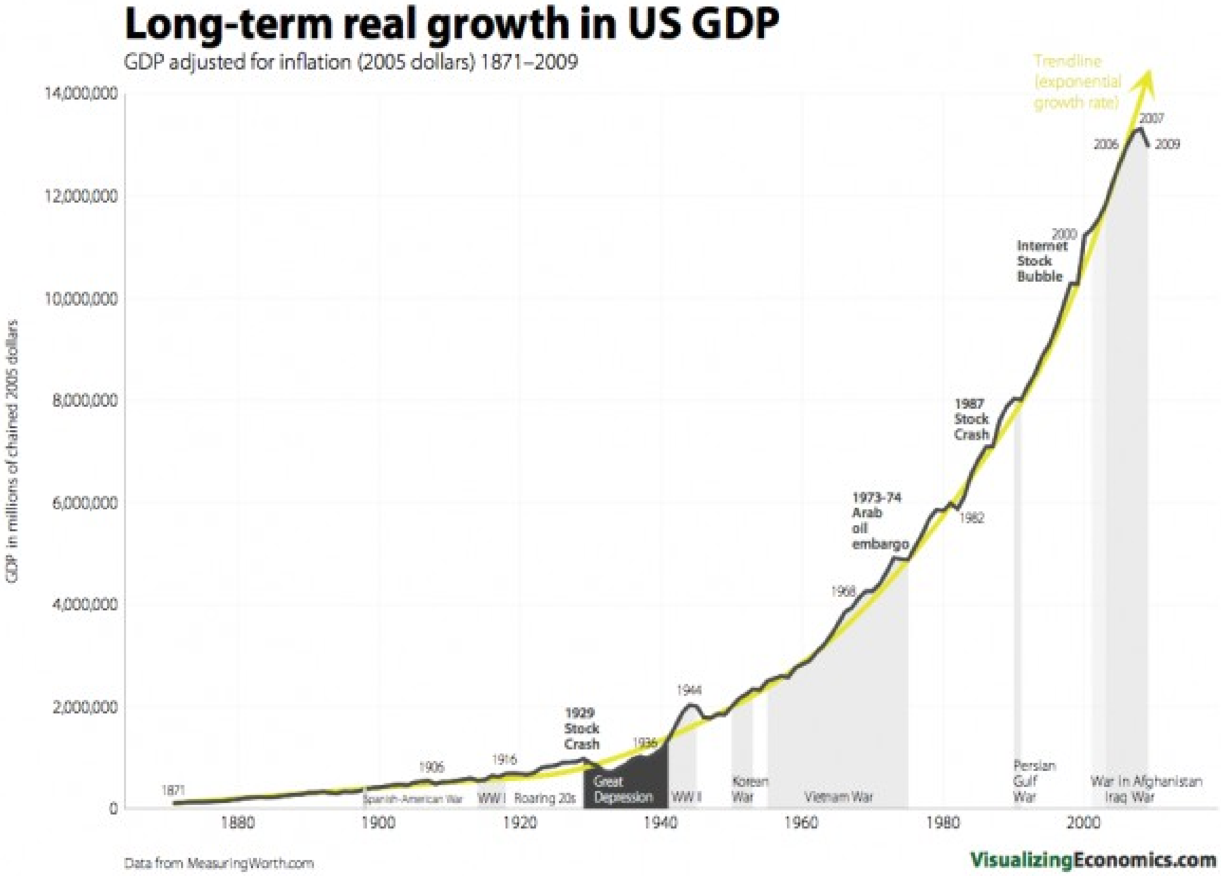
Invisible Women

*[**EDIT:** Two readers – Pseudo-Erasmus and Noah Smith – have chimed in with very plausible explanations centering around labor force participation and productivity. See also my Mistakes page.]*

A commenter asked in yesterday's discussion on *The Two-Income Trap*: why did the entry of women into the workforce produce so little effect on GDP? Here's the graph of US women's workforce participation:



And here's the graph of US GDP.



In about 50 years – 1935 to 1985 – we went from 20% of women in the workforce to 60% of women in the workforce. Assuming a bit under 100% of men in the workforce, that's an increase of almost 50% over the expected trend if number of women in the workforce had stayed constant.

I've already admitted I find the fixedness of the GDP trend bizarre. But how on Earth do you unexpectedly raise the number of people in the workforce by 50% and still stick to exactly the same GDP trend? It would be like the US annexed Mexico one day but the GDP didn't change a bit. Are we imagining here that if women *hadn't* entered the workforce, GDP would have suddenly deviated downward from the trend, and totally by coincidence women rushed in to save the day? And then when we ran out of interested women to add to the work force, again by coincidence the GDP stabilized back to its trend line?

Possible solutions:

1. The GDP data is totally false. I am suspicious of that GDP data anyway. Maybe some joker in the Bureau of Labor Statistics just graphed out an exponential function and reported that as our GDP to see if anyone would notice.

2. The GDP data is low resolution. So low-resolution that even a change of 50% is invisible if stretched over a long enough time period.

3. Women contributed through unpaid labor in the home, and their paid labor substituted for that but didn't add to it. But as far as I know GDP only counts paid-for goods. Not only should women's labor in the home not have counted, but GDP should *overcount* the benefits of putting women to work because paid daycare and so on appear as valuable new services.

4. Somehow in total contradiction to usual economic theory, all gains made by women came out of the pockets of men, leaving the same growth as would have happened anyway.

This last one segues into a question asked by another commenter – did women entering the workforce drive down male wages?

This seems a lot like the question "do immigrants entering the workforce drive down native-born wages?" to which economists tend to answer "no" with more or fewer caveats. But the economists' explanation for the immigrant effect is that in addition to producing, immigrants also consume, increasing demand.

Women were presumably already consuming. Back when they were housewives, they still needed houses, food, clothes, entertainment, et cetera. Their entrance into the workforce may create slightly more demand – for business clothes and office supplies, for example – but nothing like the demand created by immigrants entering the country.

If women increase the labor supply by 50% but don't change the demand for labor, that seems like it should make wages go way down. Even if women are primarily in different jobs than men, men's wages should still go down via a substitution effect (that is, even if women all become elementary school teachers, then a man who was planning to be an elementary school teacher before might become a fireman instead, raising the labor supply for firemen and pushing down fireman wages).

But the only study I can find to investigate this says it didn't happen. And male wages don't seem to have dropped dramatically starting 1950 or so. They do seem to have stagnated dramatically starting 1970 or so, but I feel like the income inequality explanation for that is on pretty solid ground. Unless you want to argue that the reason for income inequality is that between both sexes there's now such a large reserve army of labor that the capitalists can get away with paying the workers very little. But I feel like economists would have told us if this were going on.

I hear a lot of conspiracy theories about women in the workforce. On the left, women in the workforce are being exploited and kept down by a sinister patriarchy. On the right, women in the workforce are a satanic plot to weaken our moral fiber.

But so far I've never heard the conspiracy theory that women never actually entered the workforce, that all the working women you see around you are animatronic robots or carefully crafted stage illusions.

Maybe this is the surest sign that the conspiracy is working.

Medicine, As Not Seen On TV

Since I was twelve years old, my life has taken place in a series of Four Year Intervals.

Four years of high school. Four years of college. Four years of medical school. Four years of residency. Four times four, nice and symbolic.

This comes to mind now because I finished my first year of residency today.

I went into it raised on a steady diet of medical TV dramas like *Scrubs* and *House*, the legends passed down by other doctors in my family, and the ideas inculcated into me in medical school. It turned out to be nothing like any of those.

I've written a few posts about my experiences at work: The Hospital Orientation, I Aten't Dead, Who By Very Slow Decay, and Evening Doc. I've tried to avoid writing anything more specific in order to protect patient confidentiality and *my* confidentiality.

But I thought this would be a good time to record – for my future self as much as for anyone else – what surprised me in my first year of medical practice.

To start with, forget about diagnostic mysteries. If you've ever seen *House* or anything else remotely like it, you imagine doctors as constantly presented with weird and wonderful symptoms, then racing against the clock to figure out what rare and deadly disease it is.

In real life, patients are more like the elderly lady I got last month. She had three hospital admissions for urinary tract infections in the past two years. Now she comes in with urinary symptoms. Before I even know the patient exists, the emergency room doctor has run a urine test which reveals that it’s a urinary tract infection. He has helpfully started her on the correct antibiotic for urinary tract infections. WHAT COULD THIS DIAGNOSTIC MYSTERY POSSIBLY BE?

Yeah, it was a urinary tract infection.

Or the guy who comes in shaking and sweating. I ask him what happened. He said he has been drinking alcohol for thirty years, and two days ago he tried to stop cold turkey. Have you ever had these sorts of symptoms before? Yes, every time I go off alcohol I get them. Does anything relieve the symptoms? Yes, drinking more alcohol. SOMEBODY PAGE DOCTOR HOUSE TO FIGURE OUT WHAT’S GOING ON?

Yeah, it was alcohol withdrawal.

Not all the patients I got were like this. But probably ninety-five percent of them were. Most people come into hospital for flare-ups of chronic problems they have had for, at minimum, ten years. Most of the time they have been to their primary care doctor first, who has made the diagnosis and sent the patient to the hospital for treatment. Or if not, they go to the emergency room, where the emergency room doctors do the same standard blood test they do on everybody and which usually gives you a really good idea what’s up. Oh, you’re feeling sick and tired and thirsty and nauseous? Hmm, your blood glucose is five hundred. Are you a diabetic? Did you take your insulin? Why didn’t you take your insulin? “Being on vacation” is not a good reason to stop taking your insulin! Do you promise to take your insulin in the future? Okay, well let’s admit you to the hospital and send you to Dr. Alexander so he can clear up this massive medical mystery we have on our hands.

But okay, five percent of cases we’re not entirely sure what’s going on. *Now* we can page Dr. House, right?

Wellll, in reality we “stabilize” them. A lot of the time “stabilize” means “put them in a bed and give them IV fluids and they get better on their own”. Sometimes the problem looks vaguely infectious and so we give empiric antibiotics, where empiric means “let’s give them an antibiotic that works for lots of stuff, and maybe it’ll work for this”. Sometimes the problem looks vaguely autoimmune and we give them steroids.

It’s pretty funny, because in medical school you spend a *lot* of time learning about maybe two dozen very rare autoimmune diseases, and how to differentiate Wegner’s granulomatosis from Takayasu arteritis, and the very subtle differences in the aetiology of each. And in real life, my attending says “Huh, this looks vaguely autoimmune, let’s throw steroids at it.” And it always works.

Now I understand that when the patient leaves hospital, they go to a rheumatologist or other specialist, and the specialist probably does lots of complicated tests and then comes up with a treatment regimen perfectly suited to that patient. But at the level I’m working at, it’s more “Hey, it responded to steroids! I guess it really *was* autoimmune! Or maybe the patient just got better on her own. Or something. Anyway, who cares, patient’s better, let’s discharge before something goes wrong.”

Because something else always goes wrong. You may be wondering: if doctors don’t spend their time solving diagnostic mysteries, what *do* they do in all those long hours they work? The answer is: deal with the avalanche of disasters that inevitably begin the second a patient walks through the door into a hospital.

I want to make it very clear I'm not criticizing my own hospital here. They make an *amazing* effort to do everything possible to avoid dangerous complications. All the hospitals I've worked at do. And all of them are death-traps. God just has a particular hatred for hospital patients, which He expresses by inflicting random diseases upon them for so long as they make the mistake of staying within the four walls and ceiling of a hospital building.

Like, you can be a perfectly healthy person, who lives forty years without anything worse than a sniffle. And then one day you're playing sports, and you break your leg and you think "What's the worst that can happen, I'll spend a day or two in the hospital?" and by the time you come out you've got two artificial legs and a transplanted kidney and a rare bunyavirus from the African tropics and you have to inject yourself with insulin every three hours or else you die.

There are some good reasons for this. Obviously hospitals are full of sick people which means the potential for contagious infectious is high. People in hospitals are always getting lines stuck into them and surgeries performed and otherwise having foreign objects stuck in the body, and of course that's a risk factor for all kinds of stuff. People in hospitals are often taking medications, which often have side effects. People in hospitals are often having tests, which sometimes involve injecting large amounts of radioactive material into the body and hoping it doesn't fry anything important.

Then there are reasons you never expect until someone teaches you about them. If you don't move your legs enough – maybe because you're lying in a hospital bed all day – the blood in your legs settles and clots, and then the blood clots travel to your lungs, and then you can't get any oxygen and potentially die. If you don't fidget enough –

maybe because you’re lying in a hospital bed unconscious – the constant pressure on a single patch of skin produces an ulcer, which gets infected and you potentially die. If you take five different recreational drugs every day, and your dealer doesn’t visit you in the hospital, then you go into withdrawal, and if you don’t want to admit what’s going on to your doctor maybe they miss it and – yeah, you potentially die.

But probably the biggest reason – and one you never think of – is that the hospital is where they’re finally doing tests on you, which means all those diseases that were lying dormant before and which you put down to normal old age finally get detected. You come in for a kidney stone, but your doctor does a blood test and finds you have diabetes. Also your calcium is a little off, we’re going to need to give you calcium pills and set up an appointment to get your parathyroid checked. And also when they did the CT of the kidneys they found a suspicious-looking mass in the colon, so you’re going to have to get that checked out. Uh, the gastroenterologist pulled the joystick controlling the colonoscope a little too hard and now you have a perforated colon, you need surgery. Uh, the surgeon put on her gloves the wrong way, now the surgical site is infected, guess you need antibiotics. Uh, guess you’re allergic to that antibiotic, let’s use a different one. Wow, allergic to four antibiotics in a row, guess this isn’t your day!

While Dr. House is diagnosing Chikungunya fever, the rest of us are treating the person who came in with a nosebleed (final diagnosis: blew nose too hard) but now has a DVT, hyperkalaemia, Sundowner’s syndrome, and a line infection.

Well, sort of treating.

John Searle came up with this really interesting philosophy-of-consciousness thought experiment. Suppose that a man were put in a room with a bunch of books, each of which contained a set of rules about Chinese characters. Sometimes, a paper with Chinese characters would come in through a slot in the door. The man would apply the rules in his book, which told him to write certain Chinese characters if certain conditions about the characters on the paper held true, and slip the output back through the slot in the door. The man does this faithfully, although he doesn't know any Chinese and has no idea what any of it is saying.

On the other side of the door is a Chinese person. In her mind, she's writing questions to the man, and he is responding back in fluent Chinese. She thinks they're having a very productive conversation, and is starting to get a crush on him.

And the question is, in what sense can the man in the room be said to "understand" Chinese? If the answer is "not at all", then in what sense can the brain – which presumably takes inputs from the environment, applies certain algorithms to them, and then sends forth appropriate outputs – be said to understand anything?

Daniel Dennett and various other materialist philosophers have a response to this challenge, which is that the man does not understand Chinese, but the man, his books, and the room can be conceptualized as an emergent system that does possess the property of Chinese-understanding and which may or may not be conscious.

I bring this up, because I understand what's going on with patient care about as well as the man understands Chinese. I feel like maybe the hospital is an emergent system that has the property of patient-healing, but I’d be surprised if any one part of it does.

Suppose I see an unusual result on my patient. I don’t know what it means, so I mention it to a specialist. The specialist, who doesn’t know anything about the patient beyond what I’ve told him, says to order a technetium scan. He has no idea what a technetium scan is or how it is performed, except that it’s the proper thing to do in this situation. A nurse is called to bring the patient to the scanner, but has no idea why. The scanning technician, who has only a vague idea why the scan is being done, does the scan and spits out a number, which ends up with me. I bring it to the specialist, who gives me a diagnosis and tells me to ask another specialist what the right medicine for that is. I ask the other specialist – who has only the sketchiest idea of the events leading up to the diagnosis – about the correct medicine, and she gives me a name and tells me to ask the pharmacist how to dose it. The pharmacist – who has only the vague outline of an idea who the patient is, what test he got, or what the diagnosis is – doses the medication. Then a nurse, who has no idea about any of this, gives the medication to the patient. Somehow, the system works and the patient improves.

The patient thinks “My doctor must be very smart”. Meantime, the girl outside that room in the thought-experiment is thinking “This man must be a brilliant Confucian scholar.”

Part of being an intern is adjusting to all of this, losing some of your delusions of heroism, getting used to the fact that you’re not going to be Dr. House, that you are at best going to be a very well-functioning gear in a vast machine that does often tedious but always valuable work.

Well, other people are. *I* plan to go into outpatient.

Starting tomorrow, I abandon this exciting world of urinary tract infections and broken legs and go into psychiatry full time. I’m looking forward to it, especially since psychiatry is a little slower-paced and more focused. But this year was meant to teach me some appreciation for the wider world of medicine.

And boy have I got it.

[Good luck to SSC commenters Athrelon and Laura and everyone else starting an internship or residency tomorrow, and congratulations to everyone finishing one up]

How Common Are Science Failures?

After a brief spurt of debate over the claim that “97% of relevant published papers support anthropogenic climate change”, I think the picture has mostly settled to an agreement that – although we can contest the methodology of that particular study – there are multiple lines of evidence that the number is somewhere in the nineties.

So if any doubt at all is to remain about climate change, it has to come from the worry that sometimes entire scientific fields can get things near-unanimously wrong, especially for political or conformity-related reasons.

In fact, I’d go so far as to say that if we are not climatologists ourselves, our prior on climate change should be *based upon* how frequently entire scientific fields get things terribly wrong for political or conformity-related reasons.

Skeptics mock the claim that science was wrong before, but skeptics mock *everything*. A better plan might be to try to quantify the frequency of scientific failures so we can see how good (or bad) the chances are for any given field.

Before we investigate, we should define our reference class properly. I think a scientific mistake only counts as a reason for doubting climate change (or any other commonly-accepted scientific paradigm) if:

1. It was made sometime in the recent past. Aristotle was wrong about all sorts of things, and so were those doctors who thought everything had to do with black bile, but the scientific community back then was a lot less rigorous than our own. Let’s say it counts if it’s after 1900.

2. It was part of a really important theory, one of the fundamental paradigms of an entire field. I'm sure some tiny group of biologists have been wrong about how many chromosomes a shrew has, but that's probably an easier mistake to wander into than all of climatology screwing up simultaneously.

3. It was a stubborn resistance to the truth, rather than just a failure to have come up with the correct theory immediately. People were geocentrists before they were heliocentrists, but this wasn't because the field of astronomy became overly politicized and self-assured, it was because (aside from one ancient Greek guy nobody really read) heliocentrism wasn't invented until the 1500s, and after that it took people a couple of generations to catch on. In the same way, Newton's theory of gravity wasn't quite as good as Einstein's, but this would not shame physicists in the same way climate change being wrong would shame climatologists. Let's say that in order to count, the correct theory has to be very well known (the correct theory is allowed to be "this phenomenon doesn't exist at all and you are wasting your time") and there is a large group of people mostly outside the mainstream scientific establishment pushing it (for approximately correct reasons) whom scientists just refuse to listen to.

4. We now know that the past scientific establishment was definitely, definitely wrong and everyone agrees about this and it is not seriously in doubt. This criterion isn't to be fair to the climatologists, this is to be fair to *me* when I have to read the comments to this post and get a bunch of "Nutritionists have yet to sign on to my pet theory of diet, that *proves* some scientific fields are hopelessly corrupt!"

Do any such scientific failures exist?

If we want to play this game on Easy Mode, our first target will be Lysenkoism, the completely bonkers theory of agriculture and genetics adopted by the Soviet Union. A low-level agricultural biologist, Lysenko, came up with questionable ways of increasing agricultural output through something kind of like Lamarckian evolution. The Soviet government wanted to inspire people in the middle of a famine, didn't really like real scientists because they seemed kind of bourgeois, and wanted to discredit genetics because heritability seemed contrary to the idea of New Soviet Man. So they promoted Lysenko enough times that everyone got the message that Lysenkoism was the road to getting good positions. All the careerists switched over to the new paradigm, and the holdouts who continued to believe in genetics were denounced as fascists. According to Wikipedia, "in 1948, genetics was officially declared "a bourgeois pseudoscience"; all geneticists were fired from their jobs (some were also arrested), and all genetic research was discontinued."

About twenty years later the Soviets quietly came to their senses and covered up the whole thing.

I would argue that Stalinist Russia, where the government was very clearly intervening in science and killing the people it didn't like, isn't a fair test case for a theory today. But climate change opponents would probably respond that the liberal world order is unfairly promoting scientists who support climate change and persecuting those who oppose it. And Lysenkoism at least proves that is the sort of thing which can in theory sometimes happen. So let's grumble a little but give it to them.

Now we turn the dial up to Hard Mode. Are there any cases of failure on a similar level within a scientific community *in a country not actively being ruled by Stalin*?

I can think of two: Freudian psychoanalysis and behaviorist psychology.

Freudian psychoanalysis needs no introduction. It dominated psychiatry – not at all a small field – from about 1930 to 1980. As far as anyone can tell, the entire gigantic edifice has no redeeming qualities. I mean, it correctly describes the existence of a subconscious, and it may have some insightful things to say on childhood trauma, but as far as a decent model of the brain or of psychological treatment goes, it was a giant mistake.

I got a little better idea just *how* big a mistake doing some research for the Anti-Reactionary FAQ. I wanted to see how homosexuals were viewed back in the 1950s and ran across two New York Times articles about them (1, 2). It's really creepy to see them explaining how instead of holding on to folk beliefs about how homosexuals are normal people just like you or me, people need to start listening to the psychoanalytic experts, who know the *real* story behind why some people are homosexual. The interviews with the experts in the article are a little surreal.

Psychoanalysis wasn't an honest mistake. The field already had a perfectly good alternative – denouncing the whole thing as bunk – and sensible non-psychoanalysts seemed to do exactly that. On the other hand, the more you got "educated" about psychiatry in psychoanalytic institutions, and the more you wanted to become a psychiatrist yourself, the more you got biased into think psychoanalysis was obviously correct and dismissing the doubters as science denalists or whatever it was they said back then.

So this seems like a genuine example of a scientific field failing.

Behaviorism in psychology was…well, this part will be controversial. A weak version is “psychologists should not study thoughts or emotions because these are unknowable by scientific methods; instead they should limit themselves to behaviors”. A strong version is “thoughts and emotions don’t exist; they are post hoc explanations invented by people to rationalize their behaviors”. People are going to tell me that real psychologists only believed the weak version, but having read more than a little 1950s psychology, I’m going to tell them they’re wrong. I think a lot of people believed the strong version and that in fact it was the dominant paradigm in the field.

And of course common people said this was stupid, of course we have thoughts and emotions, and the experts just said that kind of drivel was exactly what common people *would* think. Then came the cognitive revolution and people realized thoughts and emotions were actually kind of easy to study. And then we got MRI machines and are now a good chunk of the way to *seeing* them.

So this too I will count as a scientific failure.

But – and this seems important – I can’t think of any others.

Suppose there are about fifty scientific fields approximately as important as genetics or psychiatry or psychology. And suppose within the past century, each of them had room for about five paradigms as important as psychoanalysis or behaviorism or Lysenkoism.

That would mean there are about 250 possibilities for science failure, of which three were actually science failures – for a failure rate of 1.2%.

This doesn’t seem much more encouraging for the anti-global-warming cause than the 3% of papers that support them.

I think I’m being pretty fair here – after all, Lysenkoism was limited to one extremely-screwed-up country, and people are going to yell that behaviorism wasn’t as bad as I made it sound. And two of the three failures are in psychology, a social science much fuzzier than climatology where we can expect far more errors. A cynic might say if we include psychology we might as well go all the way and include economics, sociology, and anthropology, raising our error count to over nine thousand.

But if we want to be even fairer, we can admit that there are probably some science failures that haven’t been detected yet. I can think of three that I very strongly suspect are in that category, although I won’t tell you what they are so as to not distract from the meta-level debate. That brings us to 2.4%. Admit that maybe I’ve only caught half of the impending science failures out there, and we get to 3.6%. Still not much of an improvement for the anti-AGW crowd over having 3% of the literature.

Unless of course I am missing a whole load of well-known science failures which you will remind me about in the comments.

[Edit: Wow, people are really bad at following criteria 3 and 4, even going so far as to post the exact examples I said not to. Don’t let that be you.]

Ozy vs. Scott on Charity Baskets

I have invited Ozy to post to Slate Star Codex. I ended up disagreeing with their first post, so I'm going to include it along with my rebuttal.

Ozy:

A man goes up to a stockbroker and says, "You guys are so stupid. You invest in more than one stock. But there's only one stock that is going to pay off the most. Why don't you just put all your money in the stock that is going to earn the most money, instead of putting it in a bunch of stocks?"

With my usual quick and timely response, I would like to point out the fallacy within this article on effective altruism. The author offers up several things that would, in an effective altruist world, not exist:

> If we all followed such a ridiculous approach, what would happen to:
>
> 1. Domestic efforts to serve those in need?
> 2. Advanced research funding for many diseases?
> 3. Research on and efforts in creative and innovative new approaches to helping others that no one has ever tried before?
> 4. More local and smaller charitable endeavors?
> 5. Funding for the arts, and important cultural endeavors such as the preservation of historically important structures and archives?
>
> 6. Volunteerism for the general public, since most "worthy" efforts are overseas and require a professional degree to have what Friedman calls "deep expertise in

niche areas”?
7. Careers in the nonprofit sector?”

The answer to several of those is pretty obvious: people should work in the nonprofit sector if that’s their comparative advantage, who gives a @#$! about volunteerism or local charitable endeavors, arts funding comes out of people’s entertainment budgets they way it should, and resources are scarce and each donation to someone relatively well-off in the developed world trades off against resources from someone less well off. So far, so well-trammeled.

However, I think his points two and three are actually really interesting points. A lot of people seem to think of effective altruism as like the man who wants to invest in the best possible stock. However, in reality, just as a person who wants to maximize their returns invests in more than one stock, a society where everyone is an effective altruist would probably have a variety of different charities (although perhaps a narrower segment of charities), just as they do now.

To be clear, there are certain charities that are not effective at all and would probably not exist in a hypothetical effective altruist society. Make a Wish Foundation would probably not survive the conversion to a hypothetical effective altruist society (except, presumably, out of one’s entertainment budget). Nevertheless, the nonexistence of obviously ineffective charities doesn’t mean that we as a society would decide to have fewer charities, any more than not buying lottery tickets means that you are only allowed to invest in one stock.

(Note that I am using stocks as an analogy. Stocks and charity donations are unlike each other in a lot of ways. It isn’t a perfect metaphor. Also, I literally know nothing about stock investing.)

One of the reasons that people invest in more than one stock is uncertainty. Probably some stocks will go up and some stocks will go down. However, I, as an investor, don’t know which stocks will pay off more than other stocks. Therefore, I want to hedge my bets. Knowing that the market will go up in general, I choose to invest in a variety of different stocks, so that no matter what happens I keep some of my money.

A similar uncertainty applies to charities. For instance, it’s possible that Give Directly is run by crooks who steal all the donations. (As far as I know, Give Directly is an excellent organization and never steals anyone’s money.) If everyone has given to Give Directly, we’re screwed. If we have several different charities giving cash to people in the developing world, then it matters less that one of them is run by crooks. Similarly, we may be uncertain about whether malaria relief or schistosomiasis relief is the best bang for one’s charity buck. Given that it is impossible to eliminate all uncertainty, it’s best to direct some money towards both, so that in case malaria relief turns out to be a bust we haven’t wasted all our charitable budget.

In stocks, return is a function of risk. If there’s a chance of a large payoff, there’s an even larger chance of going bust and losing everything. If there’s not very much risk, you get payoffs that are barely larger than inflation. Therefore, you want a balanced investment strategy: have some high-risk investments that might make you rich, and some low-risk investments that have a less awesome payoff.

This also applies to charity donation, which is where we get to Berger and Penna’s concerns. Something like malaria relief is low-risk and relatively low-return. If you distribute malaria nets, it is pretty certain that people are going to have lower rates of malaria. However, there’s not much chance of getting a payoff higher than “people have lower rates of malaria, maybe no malaria at all,” which will save probably millions of lives. Compare this to, say, agronomy or disease research. With agronomy, there is a high chance that you will pour in millions of dollars and get nothing. Most agronomic research gets us, say, wheat that’s a little better at resisting weeds, or better understanding of the ideal growing conditions of the chickpea. However, there’s the slim chance that you’ll have another Green Revolution and save literally billions of lives. As effective altruists, we want to invest in both high-risk high-return and low-risk low-return charities.

Another important example of a high-risk low-return charity is a new charity, which I think is important enough that I’m going to talk about it separately. What happens if someone has a brilliant new idea about how to help people in the developing world? There’s potentially a high payoff if they can beat the current most effective charity; but new ideas for effective charities are probably not going to pay off, if for no other reason than ‘most new ideas are terrible.’ It is really important that we invest in new ideas.

What happens to a low-risk high-return charitable investment? Well, it is clearly the most effective place to donate and becomes our new baseline, and the same trilemma survives. Other charities are either comparable, and thus either higher return but higher risk or lower return but lower risk, or incontrovertibly better and the new baseline.

Please note that I’m not saying the individual should donate multiple places. Probably any individual only has time to investigate one family of charities and – for that matter – gets the most warm fuzzies from only one charity. I think that most people should probably only donate to one charity, because they can be certain they're donating to the most effective charity they can. But what that charity is is different for different people. And, no, a hypothetical effective altruist society won't totally lack scientific research.

Scott:

I *think* I disagree with this. That is, I'm sure I disagree with what I think it says, and I *think* it says what I think it says. I think it confuses two important issues involving marginal utility – call them disaster aversion and low-hanging fruit – and that once we separate them out we can see that diversifying isn't necessary in quite the way Ozy thinks.

Disaster aversion is why we try to diversify our investments in the stock market. Although there's a *bit* of money maximization going on – more money would always be nice – there's also an incentive to pass the bar of "able to retire comfortably" and a *big* incentive to avoid going totally broke. This incentive works differently in charity.

Suppose you offer me a 66% chance of dectupling my current salary, and a 33% chance of reducing my current salary to zero (further suppose I have no savings and there is no safety net). Although from a money-maximizing point of view this is a good deal, in reality I'm unlikely to take it. It would be cool to be ten times richer, but the 33% chance of going totally broke and starving to death isn't worth it.

Now suppose there is a fatal tropical disease that infects 100,000 people each year. Right now the medical system is able to save 10,000 of those 100,000; 90,000 get no care and die. You offer me a 66% chance of dectupling the effectiveness of the medical system, with a 33% chance of reducing the effectiveness of the system to zero. In this case, it seems clear that the best chance is to take the offer – the expected value is saving 56,000 lives, 46,000 more than at present.

The stock market example and the tropical disease example are different because while your first dollar matters much more to you then your 100,000th dollar, the first life saved doesn’t matter any more than the 100,000th. We can come up with strained exceptions – for example, if the disease kills so many people that civilization collapses, it might be important to save enough people to carry on society – but this is not often a concern in real-life charitable giving.

By low-hanging fruit, I mean that some charities are important up to a certain point, after which they become superseded by other charities. For example, suppose there is a charity researching a cure for Disease A, and another one researching a cure for Disease B. It may be that one of the two diseases is very simple, and even a few thousand dollars worth of research would be enough to discover an excellent cure. If we invest all our money in Disease A simply because it seems to be the better candidate, the one billionth dollar invested in Disease A will be less valuable than the first dollar invested in Disease B, since that first dollar might go to hire a mediocre biologist who immediately spots that the disease is so simple even a mediocre biologist could cure it.

This is also true with more active charities. For example, the first bed net goes to the person who needs bed nets more than anyone else in the entire world. The hundred millionth bed net goes to somebody who maaaaaybe can find some use for a bed net somewhere. It’s very plausible that buying the first bed net is the most effective thing you can do with your dollar, but buying the hundred millionth bed net is less effective than lots of other things.

In this case, *at any one time* there is only one best charity to donate to, but this charity changes very quickly. In a completely charity-naive world, Disease A might be the best charity, but after Disease A has received one million dollars it might switch to Disease B until *it* gets one million dollars, and then back to Disease A for a while, and then over to bed nets, and so on.

We can turn this into a complicated game theory problem where everyone donates simultaneously without knowledge of the other people's donations, and in this case I think the solution might be seek universalizability and donate to charities in exactly the proportion you hope everyone else donates – which would indeed be a certain amount to Disease A, a certain amount to Disease B, and a certain amount to bed nets, in the hope of picking all the low-hanging fruit before you subsidize the less efficient high-hanging-fruit-picking.

But in reality it's not a complicated game theory problem. You can go on the Internet and find more or less what the budget of every charity is. That means that for you, at this point in time, there is only one most efficient charity. Unless you are Bill Gates, it is unlikely that the money you donate will be so much that it pushes your charity out of the low-hanging fruit category and makes another one more effective, so at the time you are donating there is one best charity and you should give your entire donation to it.

Granted, people are not able to directly perceive utility and will probably err on exactly what this charity is. But I think the pattern of errors will be closer to the ideal if everyone is trying to donate to the charity they consider highest marginal value at this particular time rather than if everyone is trying to diversify.

The reasons for diversifying in the stock market are based on individual investors’ desire not to go broke and don’t really apply here.

List Of The Passages I Highlighted In My Copy Of “The Two-Income Trap”

– but which didn’t fit naturally into the review.

> Today’s bankrupt families are deeper in debt than their counterparts just twenty years earlier, and their overall financial picutre – assets and debts – is worse. In 1981, the median family filing for bankrupcy owed 80% of total annual income in credit card and other nonmortgage debts; by 2001, that figure had nearly doubled to 150% of annual income.

One of the better parts of the book was its busting the myth that people use bankruptcy as an “easy way out” or that they’re declaring it willy-nilly. People are waiting much longer and trying much harder to avoid it now than a generation ago.

> Many commenters seem concerned that the families filing for bankruptcy are not sufficiently contrite. Democractic Senator Patricia Murray from Washington argues that the Senate should make it a priority “to recapture the stigma associated with a bankruptcy filing”. The idea that they do not feel bad enough about their bankruptcy filings would have come to a shock to most of the families who filed…In our research, several mothers were willing to talk with us only on the condition that we not use the word “bankruptcy” during the telephone interview for fear that a child might pick up the extension phone and hear the dreaded word. Some said that just hearing the word still makes them cry, and they asked us to refer simply to “the event”. More than 80 percent of the families we interviewed reported that they would be “embarrassed” or “very

> embarrassed” if their families, friends, or neighbors learned of their bankruptcy.

Another quote along the same lines.

> The odds that a worker will suffer an involuntary job loss have increased by 28% since the 1870s. Growing job insecurity has been hard on single-income families, who now face a 28% higher chance that the breadwinner will lose his job. But for today’s dual-income family, the numbers are doubly grim, as each spouse faces a higher likelihood of a job layoff. We estimate that in a single year, roughly 6.3% of dual-income families – one out of every sixteen – will receive a pink slip. That means that a family today with both husband and wife in the workforce is approximately two and a half times more likely to face a job losss than a single-income family of a generation ago.

I am too young to have strong opinions on How Things Have Changed. But I remember that back in the 90s, there were a lot of articles about The Layoffs Crisis and how layoffs were A Sign Of The Decline Of America. And now the understanding that people get laid off or downsized a lot is such an accepted part of everyday life that it seems weird that it was once a news item of approximately the same concern level as global warming or immigration. The people ten years younger than I am are going to have *no idea* that there was once hand-wringing over it, or that it is even the sort of thing over which hands could be wrung.

> Make no mistake. Financial distress is a problem for both men and women. But we do not want to leave the impression that these phenomena are entirely gender-neutral. They are not. Mothers are 35% more likely than childless homeowners to lose their homes, three times

> more likely than men without children to go bankrupt, and seven times more likely to head up the family after a divorce.

So a while ago I recommended everyone be extraordinarily paranoid about feminist statistics. People gave me a lot of flak over that, and Jeff K wrote a blog post where he did some tests and said he found that feminist statistics were no worse than anyone else's. I acknowledge the plausibility of his viewpoint – and yet extraordinary paranoia about feminist statistics has, at least for me, been the gift that keeps on giving. I suspect Jeff and everyone else reading that paragraph did what I very nearly did – assume it supported the conclusion it said it was supporting. But since I am extraordinarily paranoid, I made sure to re-read it and double-check. And so I noticed that it strongly implies there is a difference between women and men – but then it very deliberately avoids making the comparison. Mothers (ie women with children) are more likely than people without children (of either gender) to lose their home. Mothers are more likely than men without children to go bankrupt. There is no comparison between mothers and women without children, nor between mothers and fathers. I expect that if such a comparison showed any difference at all, Warren would have made it, rather than strongly imply she was doing so but in fact making distractor comparisons. The only place where women may be compared to an appropriate category of men – and even here it is left very vague – is in the last statistic, about being more likely to head the family after a divorce. But this is just the point that women get the children more often after divorce – which is not exactly the sort of gender disparity I feel we were promised.

> Research shows that on average, a husband is three times more likely than a wife to take primary responsibility for managing the family's money. But as a

> couple sinks into financial turmoil, this responsibility tends to shift. As families fall behind on their bills, it is wives who roll up their sleeves and do what must be done…Among couples who seek credit counseling or file for bankruptcy, the split over who was responsible for dealing with the bills was exactly reversed from that of secure families: three-quarters of the wives were exclusively responsible for trying to extract their families from their financial quagmire.

Okay, this one doesn’t even require extraordinary paranoia to start noticing alternative interpretations.

> In the early 1970s, not only did most Americans believe that the public schools were functioning reasonably well, a sizable majority of adults thought that public education had improved since they were kids. Today, only a small minority of Americans share this optimistic view. Instead, the majority now believes that schools have gotten significantly worse. Fully half of all Americans are dissatisfied with America’s public education system, a deep concern shared by black and white parents alike.

I seriously doubt public schools became much worse in any interesting way, since as far as I can tell most educational philosophies kind of work about equally well. I wonder how much this has to do with the media spreading panic.

> A group of solidly middle-class Americans – our nation’s police officers – illustrate the point. A recent study showed that the average police officer could not afford a median priced home in two-thirds of the nation’s metropolitan areas on the officer’s income alone. The same is true for elementary school teachers. Nor is this phenomenon limited to high cost cities like New York

> and San Francisco. Without a working spouse, the family of a police officer or teacher is forced to rent an apartment or buy in a marginal neighbhorhood even in more modestly priced cities such as Nashville, Kansas City, and Charlotte.

Stay-at-home parents are now difficult except for the well-off.

> Single mothers who have been to college are actually more likely to end up bankrupt than their less educated sisters – nearly 60% more likely.

The floor is open to anyone who would like to come up with a just-so story to explain this.

> In 2001, freshman Senator Hillary Clinton voted in favor of [a bill making it harder to filing for bankruptcy, which she had previously opposed violently]. Had the bill been transformed to get rid of all those awful provisions that had so concerned First Lady Hillary Clinton? No. The bill was essentially the same, but Clinton was not. As First Lady, she had been persuaded that the bill was bad for families, and she was willing to fight for her beliefs. As New York’s newest senator, however, it seems that Hillary Clinton could not afford such a principled positions. Campaigns cost money, and that money wasn’t coming from families in financial trouble. Senator Clinton received $140,000 in campaign contributions from banking industry executives in a single year, making her one of the top two recipients in the Senate. Big banks were now part of Senator Clinton’s constituency. She wanted their support, and they wanted hers – including a vote in favor of [what she had previously called] “that awful bill”.

Warren’s *vicious* attack on Hillary Clinton was a highlight, considering the current political situation. Apparently she gave a presentation on the problems with a bankruptcy bill to Clinton when she was First Lady, Clinton was very impressed and worked really hard to fight it, and then when she became Senator she turned around and supported it. You can bet this is coming up in the next Democratic primary.

> To give a sense of just how expensive subprime mortgages are, consider this: In 2001, when standard mortgage loans were in the 6.5% range, Citibank’s *average* mortgage rate (which included both subprime and traditional mortagages) was 15.6%. To put that in perspective, a family buying a $175,000 home with a subprime loan at 15.6% would pay an extra $420,000 during the 30-year life of the mortgage – that is, over and above the payments due on a prime mortgage. Had the family gotten a traditional mortgage instead, they would have been able to put two children through college, purchase half a dozen new cars, *and* put enough aside for a comfortable retirement.

I will always reblog clever ways of putting numbers in perspective. Also, whoa. Something to remember when the news talks about how Americans can’t afford to save for retirement anymore. [**EDIT**: Eric finds some evidence this is exaggerated]

> At Citibank, for example, researchers have concluded that at least 40% of those who were sold ruinous subprime mortgages would have qualified for prime-rate loans…A study by the Department of Housing and Urban Development revealed that one in nine middle-income families and one in fourteen upper-income families who refinanced a home mortgage ended up with a high-fee

> high-interest subprime mortgage. For many of these families there is no trade-off between access to credit and the cost of credit. They had their pockets picked, plain and simple.

This is part of what I meant when I praised Warren for being able to back up her accusations of market failure. Apparently there is not enough awareness of options for the market in mortgages to be well-functioning.

> Most Americans guard their credit ratings jealously, living with a slightly prickly sensation that they could be cut off if they fell behind or forgot to pay a bill. What they don't realize is that when a borrower makes a partial payment, when he misses a bill, and when his credit rating drops, he actually *gets more offers* for credit. He is not just down on his luck, behind on his bills, and short on cash, he has now joined the ranks of an elite group – The Lending Industry's Most Profitable Customers…within six months of filing for bankruptcy, 84% of families had already received unsolicited offers for new credit.

As several commenters have pointed out, this means something different than I originally thought – having good credit rating may still be important to get low interest rates. But the fear that you will never get credit if your rating is poor is misplaced.

> What does bankruptcy have to do with abortion? In Washington, a great deal. Over the past several years, pro-choice groups had scored significant court victories against a few prominent abortion clinic protesters by obtaining money judgments against them, only to see those victories turn to dust when the protesters declared bankruptcy and discharged their debts.

In a strange twist of politics, the credit industry’s version of the bankruptcy bill had been supported by Senator Charles Schumer, of New York, who had garnered strong support among women’s groups for his pro-choice politics. Ever responsive to his constituents, Senator Schumer inserted a provision into the bankruptcy bill that would make it more difficult for abortion clinic protesters to discharge judgments entered against them if they were sued for their protest activities, much in the same way drunk drivers and embezzlers cannot use bankruptcy to discharge judgments against themselves. Eager to appeal to women voters, the Senate had accepted the amendment in 2001. But in 2002, when the bankruptcy bill went back to the House with the abortion amendment in it, a coalition of right-to-life representatives refused to go along. They brought the bill to a standstill.

Desperate to get the bill passed, the banking lobby went back to the Senate, pressuring Senator Schumer to remove the controversial abortion provision. The industry ran attack ads against him in his home state, demanding that he support the bankruptcy bill — and claiming that he was costing every American family $550 a year. (The attack on Senator Schumer was particularly ironic, since he had received more campaign contributions from the credit industry than any other Senator, just nosing out fellow New Yorker Hillary Clinton.) But by this point, the pro-choice women’s groups were also mobilized, and they held firm, supporting Senator Schumer and threatening to withhold support from any elected official who moved to take the provision out of the bankruptcy bill. In one of those rare defining moments, Senator Schumer had to choose between big business and pro-choice women,

> both of whom had supported his campaign. He chose women, and the amendment remained in the bill.
>
> Ultimately, two strange bedfellows — a small group of socially conservative Republicans and a handful of progressive Democrats — gathered enough momentum to defeat the bankruptcy bill against the best-financed lobbying campaign of the 107th Congress.

I feel like if I had to send a message to aliens to tell them everything they needed to know about US politics, it would be this story.

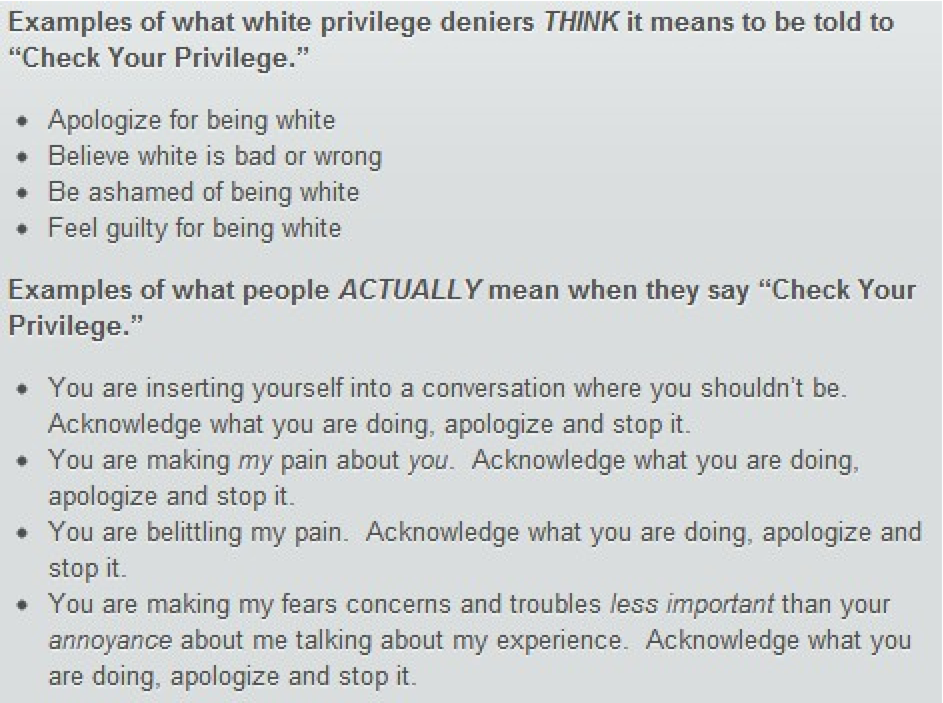
Social Justice And Words, Words, Words

[Content note: hostility toward social justice, discussion of various prejudices]

“Words! Words! Words! I’m so sick of words! I get words all day through. First from him, now from you. Is that all you blighters can do?” – Eliza Doolittle

I.

I recently learned there is a term for the thing social justice does. But first, a png from racism school dot tumblr dot com.



Examples of what white privilege deniers *THINK* it means to be told to “Check Your Privilege.”

- Apologize for being white
- Believe white is bad or wrong
- Be ashamed of being white
- Feel guilty for being white

Examples of what people *ACTUALLY* mean when they say “Check Your Privilege.”

- You are inserting yourself into a conversation where you shouldn’t be. Acknowledge what you are doing, apologize and stop it.
- You are making *my* pain about *you*. Acknowledge what you are doing, apologize and stop it.
- You are belittling my pain. Acknowledge what you are doing, apologize and stop it.
- You are making my fears concerns and troubles *less important* than your *annoyance* about me talking about my experience. Acknowledge what you are doing, apologize and stop it.

So, it turns out that privilege gets used perfectly reasonably. All it means is that you’re interjecting yourself into other people’s conversations and demanding *their* pain be about *you*. I think I speak for all straight white men when I say that sounds really bad and if I was doing it I’m sorry and will try to avoid ever doing it again. Problem solved, right? Can’t believe that took us however many centuries to sort out.

A sinking feeling tells me it probably isn’t that easy.

In the comments section of the *last* disaster of a social justice post on my blog, someone started talking about how much they hated the term “mansplaining”, and someone else popped in to – ironically – explain what “mansplaining” was and why it was a valuable concept that couldn’t be dismissed so easily. Their explanation was lucid and reasonable. At this point I jumped in and commented:

> I feel like every single term in social justice terminology has a totally unobjectionable and obviously important meaning – and then is actually used a completely different way.
>
> The closest analogy I can think of is those religious people who say “God is just another word for the order and beauty in the Universe” – and then later pray to God to smite their enemies. And if you criticize them for doing the latter, they say “But God just means there is order and beauty in the universe, surely you’re not objecting to that?”
>
> The result is that people can accuse people of “privilege” or “mansplaining” no matter what they do, and then when people criticize the concept of “privilege” they retreat back to “but ‘privilege’ just means you’re interrupting women in a women-only safe space. Surely no one can object to criticizing people who do that?”
>
> …even though I get accused of “privilege” for writing things on my blog, even though there’s no possible way

> that could be “interrupting” or “in a women only safe space”.
>
> When you bring this up, people just deny they’re doing it and call you paranoid.
>
> When you record examples of yourself and others getting accused of privilege or mansplaining, and show people the list, and point out that exactly zero percent of them are anything remotely related to “interrupting women in a women-only safe space” and one hundred percent are “making a correct argument that somebody wants to shut down”, then your interlocutor can just say “You’re deliberately only engaging with straw-man feminists who don’t represent the strongest part of the movement, you can’t hold me responsible for what they do” and continue to insist that anyone who is upset by the uses of the word “privilege” just doesn’t understand that it’s wrong to interrupt women in safe spaces.
>
> I have yet to find a good way around this tactic.

My suspicion about the gif from racism school dot tumblr dot com is that the statements on the top show the ways the majority of people will encounter “privilege” actually being used, and the statements on the bottom show the uncontroversial truisms that people will defensively claim “privilege” means if anyone calls them on it or challenges them. As such it should be taken as a sort of weird Rosetta Stone of social justicing, and I can only hope that similarly illustrative explanations are made of other equally charged terms.

Does that sound kind of paranoid? I freely admit I am paranoid in this area. But let me flesh it out with one more example.

Everyone is a little bit racist. We know this because there is a song called "Everyone's A Little Bit Racist" and it is very cute. Also because most people score poorly on implicit association tests, because a lot of white people will get anxious if they see a black man on a deserted street late at night, and because if you prime people with traditionally white versus traditionally black names they will answer questions differently in psychology experiments. It is no shame to be racist as long as you admit that you are racist and you try your best to resist your racism. Everyone knows this.

Donald Sterling is racist. We know this because he made a racist comment in the privacy of his own home. As a result, he was fined $2.5 million, banned for life from an industry he's been in for thirty-five years, banned from ever going to basketball games, forced to sell his property against his will, publicly condemned by everyone from the President of the United States on down, denounced in every media outlet from the national news to the Podunk Herald-Tribune, and got people all over the Internet gloating about how pleased they are that he will die soon. We know he deserved this, because people who argue he didn't deserve this were also fired from *their* jobs. He deserved it because he was racist. Everyone knows this.

So.

Everybody is racist.

And racist people deserve to lose everything they have and be hated by everyone.

This seems like it might present a problem. Unless of course you plan to be the person who gets to decide *which* racists lose everything and get hated by everyone, and which racists are okay for now as long as they never cross you in any way.

Sorry, there’s that paranoia again.

Someone will argue I am equivocating between two different uses of “racist”. To which I would respond that *this is exactly the point*. I don’t know if racism school dot tumblr dot com has a Rosetta Stone with Donald Sterling on the top and somebody taking the Implicit Association Test on the bottom. But I think there is a strain of the social justice movement which is very much about abusing this ability to tar people with extremely dangerous labels that they are not allowed to deny, in order to further their political goals.

II.

I started this post by saying I recently learned there is a term for the thing social justice does. A reader responding to my comment above pointed out that this tactic had been described before in a paper, under the name “motte-and-bailey doctrine”.

The paper was critiquing post-modernism, an area I don’t know enough about to determine whether or not their critique was fair. It complained that post-modernists sometimes say things like “reality is socially constructed”. There’s an uncontroversial meaning here – we don’t experience the world directly, but through the categories and prejudices implicit to our society. For example, I might view a certain shade of bluish-green as blue, and someone raised in a different culture might view it as green. Okay. Then post-modernists go on to say that if someone in a different culture thinks that the sun is light glinting off the horns of the Sky Ox, that’s just as real as our own culture’s theory that the sun is a mass of incandescent gas a great big nuclear furnace. If you challenge them, they’ll say that you’re denying reality is socially constructed, which means you’re clearly very naive and think you have perfect objectivity and the senses perceive reality directly.

The writers of the paper compare this to a form of medieval castle, where there would be a field of desirable and economically productive land called a bailey, and a big ugly tower in the middle called the motte. If you were a medieval lord, you would do most of your economic activity in the bailey and get rich. If an enemy approached, you would retreat to the motte and rain down arrows on the enemy until they gave up and went away. Then you would go back to the bailey, which is the place you wanted to be all along.

By this metaphor, statements like “God is an extremely powerful supernatural being who punishes my enemies” or “The Sky Ox theory and the nuclear furnace theory are equally legitimate” or “Men should not be allowed to participate in discussions about gender” are the bailey – not defensible at all, but if you can manage to hold them you’ve got it made.

Statements like “God is just the order and love in the universe” and “No one perceives reality perfectly directly” and “Men should not interject into safe spaces for women” are the motte – extremely defensible, but useless.

As long as nobody’s challenging you, you spend time in the bailey reaping the rewards of occupying such useful territory. As soon as someone challenges you, you retreat to the impregnable motte and glare at them until they get annoyed and go away. Then you go back to the bailey.

This is a metaphor that only historians of medieval warfare could love, so maybe we can just call the whole thing "strategic equivocation", which is perfectly clear without the digression into feudal fortifications.

III.

I probably still sound paranoid. So let me point out something I think the standard theory fails to explain, but my theory explains pretty well.

Why can't social justice terms apply to oppressed groups?

Like, even bringing this up *freaks people out*. There is no way to get a quicker reaction from someone in social justice than to apply a social justice term like "privilege" or "racist" to a group that isn't straight/white/male. And this is surprising.

If "privilege" just means "interjecting yourself into other people's conversations", this seems like something that women could do as well as men. Like, let's say that a feminist woman posts a thoughtful comment to this post, and I say "Thanks for your input, but I was actually just trying to explain things to my non-feminist male friends, I'd prefer you not interject here." Isn't it possible she might continue to argue, and so be interjecting herself into another person's conversation?

Or suppose "privilege" instead just means a cute story about a dog and a lizard, in which different people have trouble understanding each other's experiences and appreciating the amount of pain they can be causing. I know a lot of men who are scared of being Forever Alone but terrified to ask women out, and I feel their pain and most of my male friends feel their pain. Yet a lot of the feminists I talk to have this feeling that this is entirely about how they think they own women's bodies and are entitled to sex, and from *their* experience as attractive women it's easy to get dates and if you can’t it’s probably because you’re a creep or not trying hard enough. This seems to me to be something of a disconnect and an underappreciation of the pain of others, of exactly the dog-lizard variety.

There are as many totally innocuous and unobjectionable definitions of “privilege” as there are people in the social justice movement, but they generally share something in common – take them at face value, and the possibility of women sometimes showing privilege toward men is so obvious as to not be worth mentioning.

Yet if anyone mentions it in real life, they are likely to have earned themselves a link to an Explanatory Article. Maybe 18 Reasons Why The Concept Of Female Privilege Is Insane. Or An Open Letter To The Sexists Who Think Female Privilege Is A Thing. Or The Idea Of Female Privilege – It Isn’t Just Wrong, It’s Dangerous. Or the one on how there is no female privilege, just benevolent sexism. Or That Thing You Call Female Privilege Is Actually Just Whiny Male Syndrome. Or Female Privilege Is Victim Blaming, which helpfully points out that people who talk about female privilege “should die in a fire” and begins “we need to talk, and no, not just about the fact that you wear fedoras and have a neck beard.”

It almost seems like you have touched a nerve. But why should there be a nerve here?

As further confirmation that we are on to something surprising, note also the phenomenon of different social justice groups debating, with desperation in their eyes, which ones do or don’t have privilege over one another.

If you are the sort of person who likes throwing rocks at hornet nests, ask anyone in social justice whether trans men (or trans women) have male privilege. You end up in places like STFU TRANSMISOGYNIST TRANS FOLKS or Cis Privilege Is Just A Tenet Of Male Privilege or On Trans People And The Male Privilege Accusation or the womyn-born-womyn movement or Against The Cisgender Privilege List or How Misogyny Hurts Trans Men: We Do Sometimes Have Male Privilege But There Are More Important Things To Talk About Here.

As far as I can tell, the debate is about whether trans women are *more* privileged than cis women, because they have residual male privilege from the period when they presented as men, or *less* privileged than cis women, because they are transsexual – plus a more or less symmetrical debate on the trans man side. The important thing to notice is that every group considers it existentially important to prove that they are less privileged than the others, and they do it with arguments like (from last link) "all examples of cis privilege are really male privileges that are not afforded to women, or are instances of resistance to trans politics. I call it patriarchy privilege when something like an unwillingness to redefine one's own sexuality to include males is seen is labeled as offensive."

And the trans male privilege argument is one of about seven hundred different vicious disputes in which everyone is insisting other people have more privilege than they do, fighting as if their lives depended on it.

The question here: since privilege is just a ho-hum thing about how you shouldn't interject yourself into other people's conversations, or something nice about dogs and lizards – but *definitely* not anything you should be ashamed to have or anything which implies any guilt or burden whatsoever – *why are all the minority groups who participate in communities that use the term so frantic to prove they don't have it*?

We find the same unexpected pattern with racism. We all know everyone is racist, because racism just means you have unconscious biases and expectations. Everyone is a little bit racist.

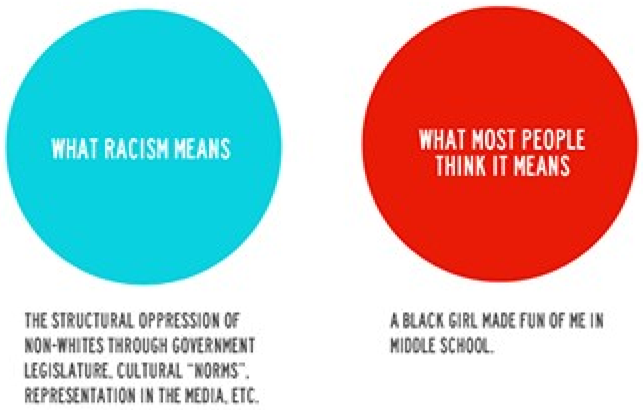
People of color seem to be part of "everyone", and they seem likely to have the same sort of in-group identification as all other humans. But they are not racist. We know this because of articles that say things like "When white people complain about reverse racism, they are complaining about losing their PRIVILEGE" and admit that "the dictionary is wrong" on this matter. Or those saying whites calling people of color racist "comes from a lack of understanding of the term, through ignorance or willful ignorance and hatred". Or those saying that "when white people complain about experiencing reverse racism, what they're really complaining about is losing out on or being denied their already existing privileges." Why Are Comments About White People Not Racist, Can Black People Be Racist Toward White People? (spoiler: no), Why You Can't Be Racist To White People, et cetera et cetera.

All of these sources make the same argument: racism means structural oppression. If some black person beats up some white person just because she's white, that might be unfortunate, it might even be "racially motivated", but because they're not acting within a social structure of oppression, it's not racist. As one of the bloggers above puts it:

> Inevitably, here comes a white person either claiming that they have a similar experience because they grew up in an all black neighborhood and got chased on the way home from school a few times and OMG THAT IS SO RACIST and it is the exact same thing, or some other such bullshittery, and they expect that ignorance to be

> suffered in silence and with respect. If you are that kid who got chased after school, that’s horrible, and I feel bad for you…But dudes, that shit is not racism.

I can’t argue with this. No, literally, I can’t argue with this. There’s no disputing the definitions of words. If you say that “racism” is a rare species of nocturnal bird native to New Guinea which feeds upon morning dew and the dreams of young children, then all I can do is point out that the dictionary and common usage both disagree with you. And the sources I cited above have already admitted that “the dictionary is wrong” and “no one uses the word racism correctly”.


WHAT RACISM MEANS
WHAT MOST PEOPLE THINK IT MEANS
THE STRUCTURAL OPPRESSION OF NON-WHITES THROUGH GOVERNMENT LEGISLATURE, CULTURAL “NORMS”, REPRESENTATION IN THE MEDIA, ETC.
A BLACK GIRL MADE FUN OF ME IN MIDDLE SCHOOL.

Source: Somebody who probably doesn’t realize they’ve just committed themselves to linguistic prescriptivism

Actually, I suppose one could escape a hostile dictionary and public by appealing to the original intent of the person who invented the word, but the man who invented the word “racism” was an activist for the forced assimilation of Indians who was known to say things like “Some say that the only good Indian is a dead one. In a sense, I agree with the sentiment, but only in this: that all the Indian there is in the race should be dead. Kill the Indian in him, and save the man.” My guess is that this guy was not totally on board with dismantling structures of oppression.

So we have a case where original coinage, all major dictionaries, and the overwhelming majority of common usage all define “racism” one way, and social justice bloggers insist with astonishing fervor that way is *totally* wrong and it *must* be defined another. One cannot argue definitions, but one can analyze them, so you have to ask – whence the insistence that racism have the structural-oppression definition rather than the original and more commonly used one? Why couldn’t people who want to talk about structural oppression make up their own word, thus solving the confusion? Even if they insisted on the word “racism” for their new concept, why not describe the state of affairs as it is: “The word racism can mean many things to many people, and I suppose a group of black people chasing a white kid down the street waving knives and yelling ‘KILL WHITEY’ qualifies by most people’s definition, but I prefer to idiosyncratically define it my own way, so just remember that when you’re reading stuff I write”? Or why not admit that this entire dispute is pointless and you should try to avoid being mean to people no matter what word you call the meanness by?

And how come this happens with *every* social justice word? How come the intertubes are clogged with pages arguing that blacks cannot be racist, that women cannot have privilege, that there is no such thing as misandry, that you should be ashamed for even *thinking* the word cisphobia? Who the heck cares? This would *never* happen in any other field. No doctor ever feels the need to declare that if we talk about antibacterial drugs we should call bacterial toxins “antihumanial drugs”. And if one did, the other doctors wouldn’t say YOU TAKE THAT BACK YOU PIECE OF GARBAGE ONLY HUMANS CAN HAVE DRUGS THIS IS A FALSE EQUIVALENCE BECAUSE BACTERIA HAVE INFECTED HUMANS FOR HUNDREDS OF YEARS BUT HUMANS CANNOT INFECT BACTERIA, they would just be mildly surprised at the nonstandard terminology and continue with their normal lives. The degree to which substantive arguments have been replaced by arguments over what words we are allowed to use against which people is, as far as I know, completely unique to social justice. Why?

IV.

And so we return to my claim from earlier:

> I think there is a strain of the social justice movement which is entirely about abusing the ability to tar people with extremely dangerous labels that they are not allowed to deny, in order to further their political goals.

If racism school dot tumblr dot com and the rest of the social justice community are right, “racism” and “privilege” and all the others are innocent and totally non-insulting words that simply point out some things that many people are doing and should try to avoid.

If *I* am right, “racism” and “privilege” and all the others are exactly what everyone loudly insists they are not – weapons – and weapons all the more powerful for the fact that you are not allowed to describe them as such or try to defend against them. The social justice movement is the mad scientist sitting at the control panel ready to direct them at whomever she chooses. Get hit, and you are marked as a terrible person who has no right to have an opinion and who deserves the same utter ruin and universal scorn as Donald Sterling. Appease the mad scientist by doing everything she wants, and you will be passed over in favor of the poor shmuck to your right and live to see another day. Because the power of the social justice movement derives from their control over these weapons, their highest priority should be to protect them, refine them, and most of all prevent them from falling into enemy hands.

If racism school dot tumblr dot com is right, people’s response to words like “racism” and “privilege” should be accepting them as a useful part of communication that can if needed also be done with other words. No one need worry too much about their definitions except insofar as it is unclear what someone meant to say. No one need worry about whether the words are used to describe them personally, except insofar as their use reveals states of the world which are independent of the words used.

If I am right, then people’s response to these words should be a frantic game of hot potato where they attack like a cornered animal against anyone who tries to use the words on them, desperately try to throw them at somebody else instead, and dispute the definitions like their lives depend on it.

And I know that social justice people like to mock straight white men for behaving in exactly that way, but man, we’re just following your lead here.

Suppose the government puts a certain drug in the water supply, saying it makes people kinder and more aware of other people’s problems and has no detrimental effects whatsoever. A couple of conspiracy nuts say it makes your fingers fall off one by one, but the government says that’s ridiculous, it’s just about being more sensitive to other people’s problems which of course no one can object to. However, government employees are all observed drinking bottled water exclusively, and if anyone suggests that government employees might also want to take the completely innocuous drug and become kinder, they freak out and call you a terrorist and a shitlord and say they hope you die. If by chance you manage to slip a little bit of tap water into a government employee’s drink, and he finds out about it, he runs around shrieking like a banshee and occasionally yelling “AAAAAAH! MY FINGERS! MY PRECIOUS FINGERS!”

At some point you might start to wonder whether the government was being entirely honest with you.

This is the current state of my relationship with social justice.

SSRIs: Much More Than You Wanted To Know

The claim that “SSRIs don’t work” or “SSRIs are mostly just placebo” is most commonly associated with Irving Kirsch, a man with the awesome job title of “Associate Director Of The Program For Placebo Studies at Harvard”.

(fun fact: there’s actually no such thing as “Placebo Studies”, but Professor Kirsch’s *belief* that he directs a Harvard department inspires him to create much higher-quality research.)

In 1998, he published a meta-analysis of 19 placebo-controlled drug trials that suggested that almost all of the benefits of antidepressants were due to the placebo effect. Psychiatrists denounced him, saying that you can choose pretty much whatever studies you want for a meta-analysis.

After biding his time for a decade, in 2008 he struck back with another meta-analysis, this being one of the first papers in all of medical science to take the audacious step of demanding all the FDA’s data through the Freedom of Information Act. Since drug companies are required to report all their studies to the FDA, this theoretically provides a rare and wonderful publication-bias-free data set. Using this set, he found that, although antidepressants did seem to outperform placebo, the effect was not “clinically significant” except “at the upper end of very severe depression”.

This launched a minor war between supporters and detractors. Probably the strongest support he received was a big 2010 meta-analysis by Fournier et al, which found that

> The magnitude of benefit of antidepressant medication compared with placebo increases with severity of depression symptoms and may be minimal or nonexistent, on average, in patients with mild or moderate symptoms. For patients with very severe depression, the benefit of medications over placebo is substantial.

Of course, a very large number of antidepressants are given to people with mild or moderate depression. So what now?

Let me sort the debate about antidepressants into a series of complaints:

1. Antidepressants were oversold and painted as having more biochemical backing than was really justified
2. Modern SSRI antidepressants are no better than older tricyclic and MAOI antidepressants, but are prescribed much more because of said overselling
3. There is large publication bias in the antidepressant literature
4. The effect size of antidepressants is clinically insignificant
5. And it only becomes significant in the most severe depression
6. And even the effects found are only noticed by doctors, not the patients themselves
7. And even that unsatisfying effect might be a result of “active placebo” rather than successful treatment
8. And antidepressants have much worse side effects than you have been led to believe
9. Therefore, we should give up on antidepressants (except maybe in the sickest patients) and use psychotherapy instead

1. Antidepressants were oversold and painted as having more biochemical backing than was really

justified – *Totally true*

It is starting to become slightly better known that the standard story – depression is a deficiency of serotonin, antidepressants restore serotonin and therefore make you well again – is kind of made up.

There was never much more evidence for the serotonin hypothesis than that chemicals that increased serotonin tended to treat depression – making the argument that “antidepressants are biochemically justified because they treat the low serotonin that is causing your depression” kind of circular. Saying “Serotonin treats depression, therefore depression is, at root, a serotonin deficiency” is about as scientifically grounded as saying “Playing with puppies makes depressed people feel better, therefore depression is, at root, a puppy deficiency”.

The whole thing became less tenable with the discovery that several chemicals that didn’t increase serotonin were also effective antidepressants – not to mention one chemical, tianeptine, that *decreases* serotonin. Now the conventional wisdom is that depression is a very complicated disturbance in several networks and systems within the brain, and serotonin is one of the inputs and/or outputs of those systems.

Likewise, a whole bunch of early ’90s claims: that modern antidepressants have no side effects, that they produce miraculous improvements in everyone, that they make you better than well – seem kind of silly now. I don’t think anyone is arguing against the proposition that there was an embarrassing amount of hype that has now been backed away from.

2. Modern SSRI antidepressants are no better than older tricyclic and MAOI antidepressants, but are prescribed much more because of said overselling – *First part true, second part less so*

Most studies find SSRI antidepressants to be no more effective in treating depression than older tricyclic and MAOI antidepressants. Most studies aren't really powered to do this. It seems clear that there aren't spectacular differences, and hunting for small differences has proven very hard.

If you're a geek about these sorts of things, you know that a few studies have found non-significant advantages for Prozac and Paxil over older drugs like clomipramine, and marginally-significant advantages for Effexor over SSRIs. But conventional wisdom is that tricyclics can be even more powerful than SSRIs for certain very severe hospitalized depression cases, and a lot of people think MAOIs worked better than anything out there today.

But none of this is very important because the real reason SSRIs are so popular is the side effect profile. While it is an exaggeration to say they have *no* side effects (see above) they are an obvious improvement over older classes of medication in this regard.

Tricyclics had a bad habit of causing fatal arrythmias when taken at high doses. This is really really bad in depression, because depressed people tend to attempt suicide and the most popular method of suicide attempt is overdosing on your pills. So if you give depressed people a pill that is highly fatal in overdose, you're basically enabling suicidality. This alone made the risk-benefit calculation for tricyclics unattractive in a lot of cases. Add in dry mouth, constipation, urinary problems, cognitive impairment, blurry vision, and the occasional tendency to cause heart arrythmias even when taken correctly, and you have a drug you're not going to give people who just say they're feeling a little down.

MAOIs have their own problems. If you're using MAOIs and you eat cheese, beer, chocolate, beans, liver, yogurt, soy, kimchi, avocados, coconuts, et cetera, et cetera, et cetera, you have a chance of precipitating a "hypertensive crisis", which is exactly as fun as it sounds. As a result, people who are *already* miserable and *already* starving themselves are told they can't eat like half of food. And once again, if you tell people "Eat these foods with this drug and you die" and a week later the person wants to kill themselves and has some cheese in the house, then you're back to enabling suicide. There are some MAOIs that get around these restrictions in various clever ways, but they tend to be less effective.

SSRIs were the first class of antidepressants that mostly avoided these problems and so were pretty well-placed to launch a prescribing explosion even apart from being pushed by Big Pharma.

3. There is large publication bias in the antidepressant literature – *True, but not as important as some people think*

People became more aware of publication bias a couple of years after serious research into antidepressants started, and it's not surprising that these were a prime target. When this issue rose to scientific consciousness, several researchers tried to avoid the publication bias problem by using only FDA studies of antidepressants. The FDA mandates that its studies be pre-registered and the results reported no matter what they are. This provides a "control group" by which accusations of publication bias can be investigated. The results haven’t been good. From Gibbons et al:

> Recent reports suggest that efficacy of antidepressant medications versus placebo may be overstated, due to publication bias and less efficacy for mildly depressed patients. For example, of 74 FDA-registered randomized controlled trials (RCTs) involving 12 antidepressants in 12,564 patients, 94% of published trials were positive whereas only 51% of all FDA registered studies were positive.

Turner et al express the same data a different way:

> . The FDA deemed 38 of the 74 studies (51%) positive, and all but 1 of the 38 were published. The remaining 36 studies (49%) were deemed to be either negative (24 studies) or questionable (12). Of these 36 studies, 3 were published as not positive, whereas the remaining 33 either were not published (22 studies) or were published, in our opinion, as positive (11) and therefore conflicted with the FDA’s conclusion. Overall, the studies that the FDA judged as positive were approximately 12 times as likely to be published in a way that agreed with the FDA analysis as were studies with nonpositive results according to the FDA (risk ratio, 11.7; 95% confidence interval [CI], 6.2 to 22.0; P<0.001). This association of publication status with study outcome remained significant when we excluded questionable studies and when we examined publication status without regard to whether the published conclusions and the FDA conclusions were in agreement

The same source tells us about the effect this bias had on effect size:

> For each of the 12 drugs, the effect size derived from the journal articles exceeded the effect size derived from the FDA reviews (sign test, P<0.001). The magnitude of the increases in effect size between the FDA reviews and the published reports ranged from 11 to 69%, with a median increase of 32%. A 32% increase was also observed in the weighted mean effect size for all drugs combined, from 0.31 (95% CI, 0.27 to 0.35) to 0.41 (95% CI, 0.36 to 0.45).

I think a lot of this has since been taken on board, and most of the rest of the research I'll be talking about uses FDA data rather than published data. But as you can see, the overall change in effect size – from 0.31 to 0.41 – is not that terribly large.

4. The effect size of antidepressants is clinically insignificant – *Depends what you mean by "clinically insignificant"*

As mentioned above, when you try to control for publication bias, the effect size of antidepressant over placebo is 0.31.

This number can actually be broken down further. According to McAllister and Williams, who are working off of slightly different data and so get slightly different numbers, the effect size of placebo is 0.92 and the effect size of antidepressants is 1.24, which means antidepressants have a 0.32 SD benefit over placebo. Several different studies get similar numbers, including the Kirsch meta-analysis that started this whole debate.

Effect size is a hard statistic to work with (albeit *extremely fun*). The guy who invented effect size suggested that 0.2 be called "small", 0.5 be called "medium", and 0.8 be called "large". NICE, a UK health research group, somewhat randomly declared that effect sizes greater than 0.5 be called “clinically significant” and effect sizes less than 0.5 be called “not clinically significant”, but their reasoning was basically that 0.5 was a nice round number, and a few years later they changed their mind and admitted they had no reason behind their decision.

Despite these somewhat haphazard standards, some people have decided that antidepressants’ effect size of 0.3 means they are “clinically insignificant”.

(please note that “clinically insignificant” is very different from “statistically insignificant” aka “has a p-value less than 0.05.” Nearly everyone agrees antidepressants have a statistically significant effect – they do something. The dispute is over whether they have a clinically significant effect – the something they do is enough to make a real difference to real people)

There have been a couple of attempts to rescue antidepressants by raising the effect size. For example, Horder et al note that Kirsch incorrectly took the difference between the average effect of drugs and the average effect of placebos, rather than the average drug-placebo difference (did you follow that?) When you correct that mistake, the drug-placebo difference rises significantly to about 0.4.

They also note that Kirsch’s study lumps all antidepressants together. This isn’t necessarily wrong. But it isn’t necessarily right, either. For example, his study used both Serzone (believed to be a weak antidepressant, rarely used) and Paxil (believed to be a stronger antidepressant, commonly used). And in fact, by his study, Paxil showed an effect size of 0.47, compared to Serzone’s 0.21. But since the difference was not statistically significant, he averaged them together and said that “antidepressants are ineffective”. In fact, his study showed that Paxil *was* effective, but when you average it together with a very ineffective drug, the effect disappears. He can get away with this because of the arcana of statistical significance, but by the same arcana I can get away with *not* doing that.

So right now we have three different effect sizes. 1.2 for placebo + drug, 0.5 for drug alone if we’re being statistically merciful, 0.3 for drug alone if we’re being harsh and letting the harshest critic of antidepressants pull out all his statistical tricks.

The reason effect size is *extremely fun* is that it allows you to compare effects in totally different domains. I will now attempt to do this in order to see if I can give you an intuitive appreciation for what it means for antidepressants.

Suppose antidepressants were in fact a weight loss pill.

An effect size of 1.2 is equivalent to the pill making you lose 32 lb.

An effect size of 0.5 is equivalent to the pill making you lose 14 lb.

An effect size of 0.3 is equivalent to the pill making you lose 8.5 lb.

Or suppose that antidepressants were a growth hormone pill taken by short people.

An effect size of 1.2 is equivalent to the pill making you grow 3.4 in.

An effect size of 0.5 is equivalent to the pill making you grow 1.4 in.

An effect size of 0.3 is equivalent to the pill making you grow 0.8 in.

Or suppose that antidepressants were a cognitive enhancer to boost IQ. This site gives us some context about occupations.

An effect size of 1.2 is equivalent to the pill making you gain 18 IQ points, ie from the average farm laborer to the average college professor.

An effect size of 0.5 is equivalent to the pill making you gain 7.5 IQ points, ie from the average farm laborer to the average elementary school teacher.

An effect size of 0.3 is equivalent to the pill making you gain 5 IQ points, ie from the average farm laborer to the average police officer.

To me, these kinds of comparisons are a little more revealing than NICE arbitrarily saying that anything below 0.5 doesn't count. If you could take a pill that helps your depression as much as gaining 1.4 inches would help a self-conscious short person, would you do it? I'd say it sounds pretty good.

5. The effect of antidepressants only becomes significant in the most severe depression – *Everything about this statement is terrible and everyone involved should feel bad*

So we've already found that saying antidepressants have an "insignificant" effect size is kind of arbitrary. But what about the second part of the claim – that they only have measurable effects in "the most severe depression"?

A lot of depression research uses a test called the HAM-D, which scores depression from 0 (none) to 52 (max). Kirsch found that the effect size of antidepressants increased as HAM-D scores increased, meaning antidepressants become more powerful as depression gets worse. He was only able to find a “clinically significant” effect size (d > 0.5) for people with HAM-D scores greater than 28. People have come up with various different mappings of HAM-D scores to words. For example, the APA says:

(0-7) No depression
(8-13) Mild depression
(14-18) Moderate depression
(19-22) Severe depression
(>=23) Very severe depression

Needless to say, a score of 28 sounds pretty bad.

We saw that Horder et al corrected some statistical deficiencies in Kirsch’s original paper which made antidepressants improve slightly. With their methodology, antidepressants reach our arbitrary 0.5 threshold around HAM-D score 26. Another similar “antidepressants don’t work” study got the number 25.

Needless to say, when anything over 23 is “very severe”, 25 or 26 still sounds pretty bad.

Luckily, people completely disagree on the meanings of basic words! Very Severely Stupid is a cute article on Neuroskeptic that demonstrates that five different people and organizations suggest five different systems for rating HAM-D scores. Bech 1996 calls our 26 cutoff “major”; Funakawa 2007 calls it “moderate”; NICE 2009 calls it “severe”. APA is unique in calling it very severe. NICE’s scale is actually the exact same as the APA scale with every category renamed to sound one level less threatening. Facepalm.

Ghaemi and Vohringer(2011) go further and say that the real problem is that Kirsch is using the standard for depressive *symptoms*, but that real clinical practice involves depressive *episodes*. That is, all this “no depression” to “severe” stuff is about whether someone can be diagnosed with depression; presumably the people on antidepressants are definitely depressed and we need a new model of severity to determine just how depressed they are. As they put it:

> the authors of the meta-analysis claimed to use the American Psychiatric Association’s criteria for severity of symptoms…in so doing, they ignore the obvious fact that symptoms differ from episodes: the typical major depressive episode (MDE) produced HDRS scores of at least 18 or above. Thus, by using symptom criteria, all MDEs are by definition severe or very severe. Clinicians know that some patients meet MDE criteria and are still able to work; indeed others frequently may not even recognize that such a person is clinically depressed. Other patients are so severe they function poorly at work so that others recognize something is wrong; some clinically depressed patients cannot work at all; and still others cannot even get out of bed for weeks or months on end. Clearly, there are gradations of severity within MDEs, and the entire debate in the above meta-analysis is about MDEs, not depressive symptoms, since all patients had to meet MDE criteria in all the studiesincluded in the meta-analysis (conducted by pharmaceutical companies for FDA approval for treatment of MDEs).

> The question, therefore, is not about severity of depressive symptoms, but severity of depressive episodes, assuming that someone meets DSM-IV criteria for a major depressive episode. On that question, a number of prior studies have examined the matter with the HDRS and with other depression rating scales, and the three groupings shown in table 2 correspond rather closely with validated and replicated definitions of mild (HDRS <24), moderate (HDRS 24–28), and severe (HDRS>28) major depressive episodes.

So, depending on whether we use APA criteria or G&V criteria, an HRDS of 23 is either "mild" (G&V) or "very severe" (APA).

Clear as mud? I agree that in one sense this is terrible. But in another sense it's actually a very important point. Kirsch's sample was really only "severe" in the context of everyone, both those who were clinically diagnosable with major depression and those who weren't. When we get to people really having a major depressive episode, a score of 26 to 28 isn't so stratospheric. But meanwhile:

> The APA seem to have ignored the fact that the HAMD did not statistically significantly distinguish between "Severe" and "Moderate" depression anyway (p=0.1)

Oh. That gives us some perspective, I guess. Also, some other people make the opposite critique and say that the HAM-D can't distinguish very well at the *low* end. Suppose HAM-Ds less than ten are meaningless and random. This would look a lot like antidepressants not working in mild depression.

Getting back to Ghaemi and Vohringer, they try a different tack and suggest that there is a statistical floor effect. They quite reasonably say that if someone had a HAM-D score of 30, and antidepressants solved 10% of their problem, they would lose 3 HAM-D points, which looks impressive. But if someone had a HAM-D score of 10, and antidepressants (still) solved 10% of their problem, they would only lose 1 HAM-D point, which sounds disappointing. But either way, the antidepressants are doing the same amount of work. If you adjust everything for baseline severity, it's easy to see that antidepressants here would have the same efficacy in severe and mild depression, even though it doesn't look that way at first.

I am confused that this works for effect sizes, because I expect effect sizes to be relative to the standard deviation in a sample. However, several important people tell me that it does, and that when you do this Kirsch's effect size goes from 0.32 to 0.40.

(I think these people are saying the exact same thing, but so overly mathematically that I've been staring at it for an hour and I'm still not certain)

More important, Ghaemi and Vohringer say once you do this, antidepressants reach the magic 0.5 number not only in severe depression, but also in moderate depression. However, when I look at this claim closely, almost all the work is done by G&V's adjusted scale in which Kirsch's "very severe" corresponds to their "mild".

(personal aside: I got an opportunity to talk to Dr. Ghaemi about this paper and clear up some of my confusion. Well, not exactly an opportunity to talk about it, per se. Actually, he was supposed to be giving me a job interview at the time. I guess we both got distracted. This may be one of several reasons I do not currently work at Tufts.)

So. In conclusion, everyone has mapped HAM-D numbers into words like “moderate” in totally contradictory ways, such that one person’s “mild” is another person’s “very severe”. Another person randomly decided that we can only call things “clinically significant” if they go above the nice round number of 0.5, then retracted this. So when people say “the effects of antidepressants are only clinically significant in severe depression”, what they mean is “the effects of antidepressants only reach a totally arbitrary number one guy made up and then retracted, in people whose HAM-D score is above whatever number I make up right now.” Depending on what number you choose and what word you make up to describe it, you can find that antidepressants are useful in moderate depression, or severe depression, or super-duper double-dog-severe depression, or whatever.

Science!

6. The beneficial effects of antidepressants are only noticed by doctors, not the patients themselves – *Partly true but okay*

So your HAM-D score has gone down and you’re no longer officially in super-duper double-dog severe depression anymore. What does that mean for the patient?

There are consistent gripes that antidepressant studies that use patients rating their own mood show less improvement than studies where doctors rate how they think a patient is doing, or standardized tests like the HAM-D.

Some people try to turn this into a conspiracy, where doctors who have somehow broken the double-blinding of studies try to report that patients have done better because doctors like medications and want them to succeed.

The reality is more prosaic. It has been known for forty years that people's feelings are the last thing to improve during recovery from depression.

This might sound weird – what is depression except people's feelings? But the answer is "quite a lot". Depressed people often eat less, sleep more, have less energy, and of course are more likely to attempt suicide. If a patient gets treated with an antidepressant, and they start smiling more and talking more and getting out of the house and are no longer thinking about suicide, their doctor might notice – but the patient herself might still feel really down-in-the-dumps.

I am going to get angry comments from people saying I am declaring psychiatric patients too stupid to notice their own recovery or something like that, but it is a very commonly observed phenomenon. Patients have access to internal feelings which they tend to weight much more heavily than external factors like how much they are able to get done during a day or how many crying spells they have, sometimes so much so that they completely miss these factors. Doctors (or family members, or other outside observers) who don't see these internal feelings, are better able to notice outward signs. As a result, it is pretty universally believed that doctors spot signs of recovery in patients long before the patients themselves think they are recovering. This isn't just imaginary – it's found it datasets where the doctors are presumably blinded and with good inter-rater reliability.

Because most antidepressant trials are short, a lot of them reach the point where doctors notice improvement but not the point where patients notice quite as much improvement.

7. The apparent benefits of antidepressant over placebo may be an "active placebo" effect rather

than a drug effect – *Unlikely*

Active placebo is the uncomfortable idea that no study can really have a blind control group because of side effects. That is, sugar pills have no side effects, real drugs generally do, and we all know side effects are how you know that a drug is working!

(there is a counterargument that placebos very often have placebo side effects, but most likely the real drug will at least have *more* side effects, saving the argument)

The solution is to use active placebo, a drug that has side effects but, as far as anyone knows, doesn't treat the experimental condition (in this case, depression). The preliminary results from this sort of study don't look good for antidepressants:

> Thomson reviewed 68 double-blind studies of tricyclics that used an inert placebo and seven that used an active placebo (44). He found drug efficacy was demonstrated in 59% of studies that employed inert placebo, but only 14% of those that used active placebo (?2=5.08, df=1, p=0.02). This appears to demonstrate that in the presence of a side-effect-inducing control condition, placebo cannot be discriminated from drug, thus affirming the null hypothesis.

Luckily, Quitkin et al (2000) solve this problem so we don't have to:

> Does the use of active placebo increase the placebo response rate? This is not the case. After pooling data from those studies in which a judgment could be made about the proportion of responders, it was found that 22% of patients (N=69 of 308) given active placebos were rated as responders. To adopt a conservative

> stance, one outlier study (50) with a low placebo response rate of 7% (N=6 of 90) was eliminated because its placebo response rate was unusually low (typical placebo response rates in studies of depressed outpatients are 25%–35%). Even after removing this possibly aberrant placebo group, the aggregate response rate was 29% (N=63 of 218), typical of an inactive placebo. The active placebo theory gains no support from these data.
>
> Closer scrutiny suggests that the “failure” of these 10 early studies to find typical drug-placebo differences is attributable to design errors that characterize studies done during psychopharmacology’s infancy. Eight of the 10 studies had at least one of four types of methodological weaknesses: inadequate sample size, inadequate dose, inadequate duration, and diagnostic heterogeneity. The flaws in medication prescription that characterize these studies are outlined in Table 3. In fact, in spite of design measurement and power problems, six of these 10 studies still suggested that antidepressants are more effective than active placebo.
>
> In summary, these reviews failed to note that the active placebo response rate fell easily within the rate observed for inactive placebo, and the reviewers relied on pioneer studies, the historical context of which limits them.

In other words, active placebo research has fallen out of favor in the modern world. Most studies that used active placebo are very old studies that were not very well conducted. Those studies failed to find an active-placebo-vs.-drug difference because they weren’t good enough to do this. But they also failed to find an active-placebo-vs.-inactive-placebo difference. So they provide no support for the idea that active placebos are stronger than inactive placebos in depression and in fact somewhat weigh against it.

8. Antidepressants have much worse side effects than you were led to believe – *Depends how bad you were led to believe the side effects were*

As discussed in Part 2, the biggest advantage of SSRIs and other new antidepressants over the old antidepressants was their decreased side effect profile. This seems to be quite real. For example, Brambilla finds a relative risk of adverse events on SSRIs only 60% of that on TCAs, p = 0.003 (although there are some conflicting numbers in that paper I’m not really clear about). Montgomery et al 1994 finds that fewer patients stop taking SSRIs than tricyclics (usually a good “revealed preference”-style measure of side effects since sufficiently bad side effects make you stop using the drug).

The charmingly named Cascade, Kalali, and Kennedy (2009) investigated side effect frequency in a set of 700 patients on SSRIs and found the following:

56% decreased sexual functioning
53% drowsiness
49% weight gain
19% dry mouth
16% insomnia
14% fatigue
14% nausea
13% light-headedness
12% tremor

However, it is very important to note that this study was not placebo controlled. Placebos can cause *terrible* side effects.

Anybody who experiments with nootropics know that the average totally-useless inactive nootropic causes you to suddenly imagine all sorts of horrible things going on with your body, or attribute some of the things that happen anyway (“I’m tired”) to the effects of the pill. It’s not really clear how much of the stuff in this study is placebo effect versus drug effect.

Nevertheless, it is worth mentioning that 34% of patients declare side effects “not at all” or “a litte” bothersome, 40% “somewhat” bothersome, and 26% “very” or “extremely” bothersome. That’s much worse than I would have expected.

Aside from the sort of side effects that you expect with any drug, there are three side effects of SSRIs that I consider especially worrisome and worthy of further discussion. These are weight gain, sexual side effects, and emotional blunting.

Weight gain is often listed as one of the most common and debilitating effects of SSRIs. But amusingly, when a placebo-controlled double-blinded study was finally run, SSRIs produced less weight gain than placebo. After a year of pill-taking, people on Prozac had gained 3.1 kg; people on placebo had gained 4.3. There is now some talk of SSRIs as a weak but statistically significant agent for weight *loss*.

What happened? One symptom of depression is not eating. People get put on SSRIs when they’re really depressed. Then they get better, either because the drugs worked, because of placebo, or just out of regression to the mean. When you go from not eating to eating, you gain weight. In the one-year study, almost everyone’s depression remitted (even untreated depressive episodes rarely last a whole year), so everyone went from a disease that makes them eat less, to remission from that disease, so everyone gained weight.

Sexual side effects are a less sanguine story. Here the direction was opposite: the medical community went from thinking this was a minor problem to finding it near-universal. The problem was that doctors usually just ask “any side effects?”, and off Tumblr people generally don’t volunteer information about their penis or vagina to a stranger. When they switched to the closed-ended question “Are you having any sexual side effects?”, a lot of people who denied side effects in general suddenly started talking.

Numbers I have heard for the percent of people on SSRIs with sexual side effects include 14, 24, 37, 58, 59, and 70 (several of those come from here. After having read quite a bit of this research, I suspect you’ve got at least a 50-50 chance (they say men are more likely to get them, but they’re worse in women). Of people who develop sexual side effects, 40% say they caused serious distress, 35% some distress, and 25% no distress.

So I think it is fair to say that if you are sexually active, your chances with SSRIs are not great. Researchers investigating the topic suggest people worried about sexual side effects should switch to alternative sexual-side-effect-free antidepressant Serzone. You may remember that as the antidepressant that worked worst in the efficacy studies and brought the efficacy of all the other ones down with it. Also, it causes liver damage. In my opinion, a better choice would be bupropion, another antidepressant which has been found many times not to cause sexual side effects and which may even improve your sex life.

(“Bupropion lacks this side effect” is going to be a common theme throughout this section. Bupropion causes insomnia, decreased appetite, and in certain rare cases of populations at risk, seizures. It is generally a good choice for people who are worried about SSRI side effects and would prefer a totally *different* set of side effects.)

There is a certain feeling that, okay, these drugs may have very very common, possibly-majority-of-user sexual side effects, but depressed people probably aren't screwing like rabbits anyway. So after you recover, you can wait the appropriate amount of time, come off the drugs (or switch to a different drug or dose for maintenance) and no harm done.

The situation no longer seems so innocuous. Despite a lack of systematic investigation, there are multiple reports from researchers and clinicians – not to mention random people on the Internet – of permanent SSRI-induced sexual dysfunction that does not remit once the drug is stopped. This is *definitely* not the norm and as far as we know it is so rare as to be unstudyable beyond the occasional case report.

On the other hand, I have this. I took SSRIs for about five to ten years as a kid, and now I have approximately the pattern of sexual dysfunction associated with SSRIs and consider myself asexual. Because I started the SSRIs too early to observe my sexuality without them, I can't officially blame the drugs. But I am very suspicious. I feel like this provides moderate anthropic evidence that it is not as rare as everyone thinks.

The last side effect worth looking at is emotional blunting. A lot of people say they have trouble feeling intense emotions (sometimes: any emotions at all) when on SSRIs. Sansone and Sansone (2010) report:

> As for prevalence rates, according to a study by Bolling and Kohlenberg, approximately 20 percent of 161 patients who were prescribed an SSRI reported apathy and 16.1 percent described a loss of ambition. In a study by Fava et al, which consisted of participants in both the United States and Italy, nearly one-third on any antidepressant reported apathy, with 7.7 percent describing moderate-to-severe impairment, and nearly 40 percent acknowledged the loss of motivation, with 12.0 percent describing moderate-to-severe impairment.

A practicing clinician working off observation finds about the same numbers:

> The sort of emotional "flattening" I have described with SSRIs may occur, in my experience, in perhaps 10-20% of patients who take these medications…I do want to emphasize that most patients who take antidepressant medication under careful medical supervision do not wind up feeling "flat" or unable to experience life's normal ups and downs. Rather, they find that–in contrast to their periods of severe depression–they are able to enjoy life again, with all its joys and sorrows.

Many patients who experience this side effect note that when you're depressed, "experiencing all your emotions fully and intensely" is not very high on your list of priorities, since your emotions tend to be terrible. There is a subgroup of depressed patients whose depression takes the form of not being able to feel anything at all, and I worry this effect would exacerbate their problem, but I have never heard this from anyone and SSRIs do not seem less effective in that subgroup, so these might be two different things that only sound alike. A couple of people discussing this issue have talked about how decreased emotions help them navigate interpersonal relationships that otherwise might involve angry fights or horrible loss – which sounds plausible but also really sad.

According to Barnhart et al (2004), "this adverse effect has been noted to be dose-dependent and reversible" – in other words, it will get better if you cut your dose, and go away completely when you stop taking the medication. I have not been able to find any case studies or testimonials by people who say this effect has been permanent.

My own experience was that I did notice this (even before I knew it was an official side effect) that it did go away after a while when I stopped the medications, and that since my period of antidepressant use corresponded with an important period of childhood socialization I ended out completely unprepared for having normal emotions and having to do a delicate social balancing act while I figured out how to cope with them. Your results may vary.

There is also a large research on suicidality as a potential side effect of SSRIs, but this looks like it would require another ten thousand words just on its own, so let's agree it's a risk and leave it for another day.

9. Therefore, we should give up on medication and use psychotherapy instead – *Makes sense right up until you run placebo-controlled trials of psychotherapy*

The conclusion of these studies that claim antidepressants don't outperform placebo is usually that we should repudiate Big Pharma, toss the pills, and go back to using psychotherapy.

The implication is that doctors use pills because they think they're much more effective than therapy. But that's not really true. The conventional wisdom in psychiatry is that antidepressants and psychotherapy are about equally effective.

SSRIs get used more than psychotherapy for the same reason they get used more than tricyclics and MAOIs – not because they're better but because they have fewer problems. The problem with psychotherapy is you've got to get severely mentally ill people to go to a place and talk to a person several times a week. Depressed people are not generally known for their boundless enthusiasm for performing difficult tasks consistently. Also, Prozac costs like 50 cents a pill. Guess how much an hour of a highly educated professional's time costs? More than 50c, that's for sure. If they are about equal in effectiveness, you probably don't want to pay extra and your insurance *definitely* doesn't want to pay extra.

Contrary to popular wisdom, it is almost never the doctor pushing pills on a patient who would prefer therapy. If anything it's more likely to be the opposite.

However, given that we're acknowledging antidepressants have an effect size of only about 0.3 to 0.5, is it time to give psychotherapy a second look?

No. Using very similar methodology, a team involving Mind The Brain blogger James Coyne found that psychotherapy decreases HAM-D scores by about 2.66, very similar to the 2.7 number obtained by re-analysis of Kirsch's data on antidepressants. It concludes:

> Although there are differences between the role of placebo in psychotherapy and pharmacotherapy research, psychotherapy has an effect size that is comparable to that of antidepressant medications.

> Whether these effects should be deemed clinically relevant remains open to debate.

Another study by the same team finds psychotherapy has an effect size of 0.22 compared to antidepressants’ 0.3 – 0.5, though no one has tried to check if that difference is statistically significant and this does not give you the right to say antidepressants have “outperformed” psychotherapy.

If a patient has the time, money, and motivation for psychotherapy, it may be a good option – though I would only be comfortable using it as a monotherapy if the depression was relatively mild.

10. Further complications

What if the small but positive effect size of antidepressants wasn’t because they had small positive effects on everyone, but because they had very large positive effects on some people, and negative effects on others, such that it averaged out to small positive effects? This could explain the clinical observations of psychiatrists (that patients seem to do *much* better on antidepressants) without throwing away the findings of researchers (that antidepressants have only small benefits over placebo) by bringing in the corollary that some psychiatrists notice some patients doing poorly on antidepressants and stop them in those patients (which researchers of course would not do).

This is the claim of Gueorguieva and Krystal 2011, who used “growth modeling” to analyze seven studies of new-generation-antidepressant Cymbalta and found statistically significant differences between two “trajectories” for the drug, but not for placebo. 66% of people were in the “responder” trajectory and outperformed placebo by 6 HAM-D points (remember, previous studies estimated HAM-D benefits over placebo at about 2.7). 33% of people were nonresponders and did about 6 HAM-D points *worse* than placebo. Average it out, and people did about 3 HAM-D points better on drug and placebo, pretty close to the previous 2.7 point estimate.

I don't know enough about growth modeling to be sure that the researchers didn't just divide the subjects into two groups based on treatment efficacy and say "Look! The subsection of the population whom we selected for doing well did well!" but they use many complicated statistics words throughout the study that I think are supposed to indicate they're not doing this.

If true, this is very promising. It means psychiatrists who are smart enough to notice people getting worse on antidepressants can take them off (or switch to another class of medication) and expect the remainder to get much, *much* better. I await further research with this methodology.

What if there were actually no such thing as the placebo effect? I know dropping this in around the end of an essay that assumes 75% of gains related to antidepressants are due to the placebo effect is a bit jarring, but it is the very-hard-to-escape conclusion of Hróbjartsson and Gøtzsche's meta-analysis on placebo. They find that three-armed studies – ie those that have a no-treatment group, a placebo-treatment group, and a real-drug-treatment group – rarely find much of a difference between no-treatment and placebo. This was challenged by Wampold et al here and here, but defended against those challenges by the long-name-Scandinavian-people here. Kirsch, who between all his antidepressant work is *still* Associate Director of Placebo Studies, finds here that 75% of the apparent placebo effect in antidepressant studies is probably a real placebo effect, but his methodology is a valiant attempt to make the most out of a total lack of data rather than a properly-directed study per se.

If placebo pills don’t do much, what explains the vast improvements seen in both placebo and treatment groups in antidepressant trials? It could be the feeling of cared-for-ness and special-ness of getting to see a psychiatrist and talk with her about your problems, and the feeling of getting-to-contribute-something you get from participating in a scientific study. Or it could just be regression to the mean – most people start taking drugs when they feel very depressed, and at some point you have nowhere to go but up. Most depression gets better after six months or so – which is a much longer period than the six week length of the average drug trial, but maybe some people only volunteered for the study four months and two weeks after their depression started.

If Hróbjartsson and Gøtzsche were right, and Kirsch and the psychiatric establishment wrong, what would be the implications? Well, the good implication is that we no longer have to worry about problem 7 – that antidepressants are merely an active placebo – since active placebos shouldn’t do anything. That means we can be more confident they really work. The more complicated implication is that psychiatrists lose one excuse for asking people to take the drugs – “Sure, the drug effect may be small, but the placebo effect is so strong that it’s still worth it.” I don’t know how many psychiatrists actually think this way, but *I* sometimes think this way.

What if the reason people have so much trouble finding good effects from antidepressants is that they’re giving the medications wrong? Psychiatric Times points out that:

> The Kirsch meta-analysis looked only at studies carried out before 1999. The much-publicized Fournier study examined a total of 6 antidepressant trials (n=718) using just 2 antidepressants, paroxetine and imipramine. Two of the imipramine studies used doses that were either subtherapeutic (100 mg/day) or less than optimal (100 to 200 mg/day)

What if we've forgotten the most important part? Antidepressants are used not only to treat acute episodes of depression, but to prevent them from coming back (maintenance therapy). This they apparently do very well, and I have seen very few studies that attempt to call this effect into question. Although it is always possible that someone will find the same kind of ambiguity around maintenance antidepressant treatment as now clouds acute antidepressant treatment, so far as far as I know this has not happened.

What if we don't understand what's going on with the placebo effect in our studies? Placebo effect has consistently gotten stronger over the past few decades, such that the difference between certain early tricyclic studies (which often found strong advantages for the medication) and modern SSRI studies (which often find only weak advantages for the medication) is not weaker medication effect, but stronger placebo effect (that is, if medication always has an effect of 10, but placebo goes from 0 to 9, apparent drug-placebo difference gets much lower). Wired has a good article on this. Theories range from the good – drug company advertising and increasing prestige and awareness of psychiatry have raised people's expectations of psychiatric drugs – to the bad – increasing scientific competence and awareness have improved blinding and other facets of trial design – to the ugly – modern studies recruit paid participants with advertisements, so some unscrupulous people may be entering studies and then claiming to get better, hoping that this sounds sufficiently like the outcome the researchers want that everyone will be happy and they'll get their money on schedule.

If placebos are genuinely getting better because of raised expectations, that's good news for doctors and patients but bad news for researchers and drug companies. The patient will be happy because they get better no matter how terrible a prescribing decision the doctor makes; the doctor will be happy because they get credit. But for researchers and drug companies, it means it's harder to prove a difference between drug and placebo in a study. You can invent an excellent new drug and still have it fail to outperform placebo by very much if everyone in the placebo group improves dramatically.

Conclusion

An important point I want to start the conclusion section with: no matter what else you believe, antidepressants are not literally ineffective. Even the most critical study – Kirsch 2008 – finds antidepressants to outperform placebo with $p < .0001$ significance. An equally important point: everyone except those two Scandinavian guys with the long names agree that, if you count the placebo effect, antidepressants are extremely impressive. The difference between a person who gets an antidepressant and a person who gets no treatment at all is like night and day. The debate takes place within the bounds set by those two statements. Antidepressants give a very modest benefit over placebo. Whether this benefit is so modest as to not be worth talking about depends on what level of benefits you consider so modest as to not be worth talking about. If you are as depressed as the average person who participates in studies of antidepressants, you can expect an antidepressant to have an over-placebo-benefit with an effect size of 0.3 to 0.5. That's the equivalent of a diet pill that gives you an average weight loss of 9 to 14 pounds, or a growth hormone that makes you grow on average 0.8 to 1.4 inches. You may be able to get more than that if you focus on the antidepressants, like paroxetine and venlafaxine, that perform best in studies, but we don't have the statistical power to say that officially. It may be the case that most people who get antidepressants do much better than that but a few people who have paradoxical negative responses bring down the average, but right now this result has not been replicated. This sounds moderately helpful and probably well worth it if the pills are cheap (which generic versions almost always are) and you are not worried about side effects. Unfortunately, SSRIs do have some serious side effects. Some of the supposed side effects, like weight gain, seem to be mostly mythical. Others, like sexual dysfunction, seem to be very common and legitimately very worrying. You can avoid most of these side effects by taking other antidepressants like bupropion, but even these are not totally side-effect free. Overall I think antidepressants come out of this definitely not looking like perfectly safe miracle drugs, but as a reasonable option for many people with moderate (aka "mild", aka "extremely super severe") depression, especially if they understand the side effects and prepare for them.

Growing Children For Bostrom's Disneyland

[Epistemic status: Started off with something to say, gradually digressed, fell into total crackpottery. Everything after the halfway mark should have been written as a science fiction story instead, but I'm too lazy to change it.]

I'm working my way through Nick Bostrom's Superintelligence: Paths, Dangers, Strategies. Review possibly to follow. But today I wanted to write about something that jumped out at me. Page 173. Bostrom is talking about a "multipolar" future similar to Robin Hanson's "em" scenario. The future is inhabited by billions to trillion of vaguely-human-sized agents, probably digital, who are stuck in brutal Malthusian competition with one another.

Hanson tends to view this future as not necessarily so bad. I tend to think Hanson is crazy. I have told him this, and we have argued about it. In particular, I'm pretty sure that brutal Malthusian competition combined with ability to self-edit and other-edit minds necessarily results in paring away everything not directly maximally economically productive. And a lot of things we like – love, family, art, hobbies – are not directly maximally economic productive. Bostrom hedges a lot – appropriate for his line of work – but I get the feeling that he not only agrees with me, but one-ups me by worrying that consciousness itself may not be directly maximally economically productive. He writes:

> We could thus imagine, as an extreme case, a technologically highly advanced society, containing many complex structures, some of them far more intricate and intelligent than anything that exists on the planet today – a society which nevertheless lacks any type of being that is conscious or whose welfare has

> moral significance. In a sense, this would be an uninhabited society. It would be a society of economic miracles and technological awesomeness, with nobody there to benefit. A Disneyland with no children.

I think a large number of possible futures converge here (though certainly not all of them, I myself find singleton scenarios more likely) so it's worth asking how doomed we are when we come to this point. Likely we are pretty doomed, but I want to bring up a very faint glimmer of hope in an unexpected place.

It's important to really get our heads around what it means to be in a maximally productive superintelligent Malthusian economy, so I'm going to make some assertions. Instead of lengthy defenses of each, if you disagree with any in particular you can challenge me about it in the comments.

– Every agent is in direct competition with many other entities for limited resources, and ultimately for survival
– This competition can occur on extremely short (maybe sub-microsecond) time scales.
– A lot of the productive work (and competition) is being done by nanomachines, or if nanomachines are impossible, the nearest possible equivalent
– Any agent with a disadvantage in any area (let's say intelligence) not balanced by another advantage has already lost and will be outcompeted
– Any agent that doesn't always take the path that maximizes its utility (defined in objective economic terms) will be outcompeted by another that does.
– Utility calculations will likely be made not according to the vague fuzzy feelings that humans use, but very explicitly, such that agents will know what path maximizes their utility at any given time and their only choice will be to do that or to expect to be outcompeted.

– Agents can only survive a less than maximally utility-maximizing path if they have some starting advantage that gives them a buffer. But gradually these pre-existing advantages will be used up, or copied by the agent's descendants, or copied by other agents that steal them. Things will regress to the pre-existing Malthusianism.

Everyone will behave perfectly optimally, which of course is terrible. It would mean either the total rejection of even the illusion of free will, or free will turning into a simple formality ("You can pick any of these choices you want, but unless you pick Choice C you die instantly.")

The actions of agents become dictated by the laws of economics. Goodness only knows what sort of supergoals these entities might have – maximizing their share of some currency, perhaps a universal currency based on mass-energy? In the first million years, some agent occasionally choose to violate the laws of economics, and collect less of this currency than it possibly could have because of some principle, but these agents are quickly selected against and go extinct. After that, it's total and invariable. Eventually the thing bumps up against fundamental physical limits, there's no more technological progress to be had, and although there may be some cyclic changes teleological advancement stops.

For me the most graphic version of this scenario is one where all of the interacting agents are very small, very very fast, and with few exceptions operate entirely on reflex. It might look like some of the sci-fi horror ideas of "grey goo". When I imagine things like *that*, the distinction between economics and harder sciences like physics or chemistry starts to blur.

If somehow we captured a one meter sphere of this economic soup, brought it to Earth inside an invincible containment field, and tried to study it, we would probably come up with some very basic laws that it seemed to follow, based on the aggregation of all the entities within it. It would be very silly to try to model the exact calculations of each entity within it – assuming we could even see them or realize they are entities at all. It would just be a really weird volume of space that seemed to follow different rules than our own.

Sci-fi author Karl Schroeder had a term for the post-singularity parts of some of his books – Artificial Nature. That strikes me as exactly right. A hyperproductive end-stage grey goo would take over a rapidly expanding area of space in which all that hypothetical outsiders might notice (non-hypothetical outsiders, of course, would be turned into goo) would be that things are following weird rules and behaving in novel ways.

There's no reason to think this area of space would be homogenous. Because the pre-goo space likely contained different sorts of terrain – void, asteroids, stars, inhabited worlds – different sorts of economic activity would be most productive in each niche, leading to slightly different varieties of goo. Different varieties of goo might cooperate or compete with each other, there might be population implosions or explosions as new resources are discovered or used up – and all of this wouldn't look like economic activity at all to the outside observer. It would look like a weird new kind of physics was in effect, or perhaps like a biological system with different "creatures" in different niches. Occasionally the goo might spin off macroscopic complex objects to fulfill some task those objects could fulfill better than goo, and after a while those objects would dissolve back into the substratum.

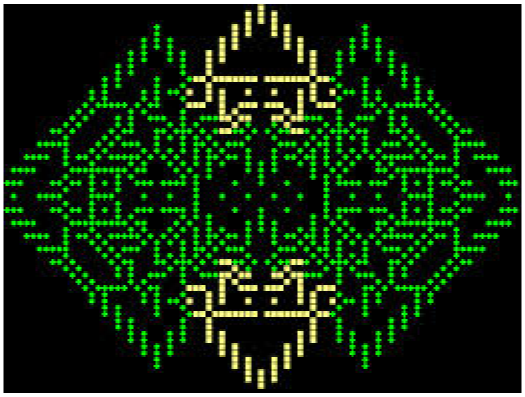
Here the goo would fulfill a role a lot like micro-organisms did on Pre-Cambrian Earth – which was also intense Malthusian competition at microscopic levels on short time-scales. Unsurprisingly, the actions of micro-organisms can look physical or chemical to us – put a plate of agar outside and it mysteriously develops white spots. Put a piece of bread outside and it mysteriously develops greenish white spots. Apply the greenish-white spots from the bread to the white spots on the agar, and some of them mysteriously die. Try it too many times and it stops working. It's totally possible to view this on a "guess those are laws of physics" level as well as a "we can dig down and see the terrifying war-of-all-against-all that emergently results in these large-level phenomena" level.

In this sort of scenario, the only place for consciousness and non-Malthusianism to go would be *higher level structures*.

One of these might be the economy as a whole. Just as ant colonies seem a lot more organism-like than individual ants, so the cosmic economy (or the economies around single stars, if lightspeed limits hold) might seem more organism-like than any of its components. It might be able to sense threats, take actions, or debate very-large-scale policies. If we agree that end-stage-goo is more like biology than like normal-world economics, whatever sort of central planning it comes up with might look more like a brain than like a government. If the components were allowed to plan and control the central planner in detail it would probably be maximally utility maximizing, ie stripped of consciousness and deterministic, but if it arose from a series of least-bad game theoretic bargains it might have some wiggle room.

But I think emergent patterns in the goo itself might be much more interesting.

In the same way our own economy mysteriously pumps out business cycles, end-stage-goo might have cycles of efflorescence and sudden decay. Or the patterns might be weirder. Whorls and eddies in economic activity arising spontaneously out of the interaction of thousands of different complicated behaviors. One day you might suddenly see an extraordinarily complicated mandala or snowflake pattern, like the kind you can get certain variants of Conway's Game Of Life to make, arise and dissipate.



Source: Latent in the structure of mathematics

Or you might see a replicator. Another thing you can convince Conway's Game of Life to make.

If the deterministic, law-abiding, microscopically small, instantaneously fast rules of end-stage-goo can be thought of as *pretty much* just a new kind of physics, maybe this kind of physics will allow replicating structures in the same way that normal physics does.

None of the particular economic agents would feel like they were contributing to a replicating pattern, any more than I feel like I'm contributing to [a power law of blogs]() every time I update here. And it wouldn't be a disruption in the imperative to only perform the most economically productive action – it would be a pattern that supervenes on everyone’s economically productive behavior.

But it would be creating replicators. Which would eventually retread important advances like sex and mutation and survival of the fittest and multicellularity and eventually, maybe, sapience.

We would get a whole new meaning of *homo economicus* – but also pan economicus, and mus economicus, and even caenorhabditis economicus.

I wonder what life would be like for those entities. Probably a lot like our own lives. They might be able to manipulate the goo the same way we manipulate normal matter. They might have science to study the goo. They might eventually figure out its true nature, or they might go their entire lifespan as a species without figuring out anything beyond that it has properties it likes to follow. Maybe they would think those properties are the hard-coded law of the universe.

(Here I should pause to point out that none of this requires literal goo. Maybe there is an economy of huge floating asteroid-based factories and cargo-freighters, with Matrioshka brains sitting on artificial planets directing them. Doesn’t matter. The patterns in there are harder to map to normal ways of thinking about physics, but I don’t see why they couldn’t still produce whorls and eddies and replicators.)

Maybe one day these higher-level-patterns would achieve their own singularity, and maybe it would go equally wrong, and they would end up in a Malthusian trap too, and eventually all of their promise would dissipate into extremely economically productive nanomachines competing against one another.

Or they might get a different kind of singularity. Maybe they end up with a paperclip-maximizing singleton. I would think it much less likely that the same kind of complex patterns would arise in the process of paperclip maximization, but maybe they could.

Or maybe, after some number of levels of iteration, they get a positive singularity, a singleton clears up their messes, and they continue studying the universe as superintelligences. Maybe they figure out pretty fast exactly how many levels of entities are beneath them, how many times this has happened before.

I’m not sure if it would be physically possible for them to intervene on the levels below them. In theory, everything beneath them ought to already be literally end-stage. But it might also be locked in some kind of game-theoretic competition that made it less than maximally productive. And so the higher-level entities might be able to design some kind of new matter that outcompetes it and is subject to their own will.

(unless the lower-level systems retained enough intelligence to figure out what was going on, and enough coordinatedness to stop it)

But why would they want to? To them, the lower levels are just physics; always have been, always will be. It would be like a human scientist trying to free electrons from the tyrannous drudgery of orbiting nuclei. Maybe they would sit back and enjoy their victory, sitting at the top of a pyramid of unknown dozens or hundreds of levels of reality.

(Also, *just once* I want to be able to do armchair futurology without wondering how many times something has already happened.)

A Paradox of Ecclesiology

[Epistemic status: Sloppy. You're going to have to read between the lines and fill in some of the holes here.]

I.

Some rationalists study ecclesiology. I used to think this was dumb. Now I appreciate it a little more. Let me see if I can explain.

Suppose you have a cause or movement. Let's say libertarianism. You're probably not going to get too far on your own, so you start looking for other people who agree with you.

You end up with a wide spectrum of people. Some of them agree with you on nearly everything. Other people consider themselves part of your movement, but disagree with your goals and hate you personally. Maybe you're kind of a soft libertarian who just wants the government to decriminalize pot and stop ordering illegal drone attacks, but the other guy wants to disband the government entirely and make everyone live in heavily armed communes. And the *other* other guy is a member of the Libertarian National Socialist Green Party, and you're not even sure if he has real opinions or just likes chaining political-sounding words together, but that swastika armband of his is starting to creep you out.

If you only work together with the libertarians who agree with you about everything, then you'll have a nice, low-conflict group who can cooperate naturally and completely to achieve common goals. You'll also have like three people.

If you work together with everyone who shares a goal with you, you get much more power – money, activist-hours, votes – but you’ve got to make ideological compromises. And sometimes you’ve got to make practical compromises too – for example, letting people you consider idiots have a say in your strategic planning, or holding your nose and agreeing to wear a swastika armband on Tuesdays and every second Thursday.

One option is to refuse to incorporate a formal group. You vote for whichever major-party candidate seems the most libertarian, occasionally picket your local IRS office, and write lots of angry letters to the editor about Big Government. The heavily-armed-commune people also do some similar things, and sometimes you go to each other’s protests, or write articles in each other’s magazines. Occasionally the Libertarian Green Nazis say something, and you get to pretend you don’t know them.

This seems to be the status of the broader libertarian movement right now, as well as a lot of other movements like feminism, transhumanism, socialism, Islam, and atheism – just to name a few.

Another option is that you *do* incorporate a formal group. You come up with bylaws and membership requirements and elect a Planning Committee and start fretting a lot over who is In and who is Out.

The libertarian version of this seems to be the US Libertarian Party. They are no doubt the strongest face of US libertarianism, but they only capture a tiny part of the energy and power of the movement. Running through the other movements mentioned in order, they can boast groups like the National Organization for Women, MIRI, the US Communist Party, various mosques, and the Secular Student Association. Usually there is more than one group per movement – Islam, for example, boasts everything from your local mosque to ISIS to CAIR.

Muslims have this quasi-messianic goal of the caliphate – a single organization representing and capturing all the strength of Islam. It seems to me to be a very reasonable goal, at least conditional on supporting Islam and wanting it to flourish. Likewise, if there were a Single Feminist Organization that contained and directed the actions of all feminists, that would be a really big deal. If two or more socialists could sit in a room together without each accusing the others of being fascist pigs, maybe socialism would achieve more.

The Big Question of ecclesiology seems to me to be – how do you design a single organization to capture and direct the greatest percentage of your movement's energy most effectively?

Here there are a bunch of tradeoffs, most notably:

Strict organization versus relaxed organization. If I wanted to capture near 100% of all libertarians, the easiest way would be to spend my own money publishing nice glossy pamphlets with pictures of the Statue of Liberty on them saying in a vague way that more freedom would be nice. Probably most libertarians, presented with a chance to sign their name on a dotted line saying they are "a supporter" of my organization at no cost to them, would be willing to go along. But in terms of energy direction, this is frickin' useless.

On the other hand, imagine an organization in which the Libertarian Field Marshal gave orders to everyone who signed on – *you* quit college to canvas door-to-door, *you* get a Ph. D in economics so we can have someone ready to respond to arguments against the free market if we need it, *you* become a banker and donate your obscene salary to our group. This group has energy-direction up the wazoo, and it could become *incredibly* powerful with only a couple dozen members. It also would never get a couple dozen members.

Strict orthodoxy versus relaxed orthodoxy. Maybe you’re allowed to join the group if you “identify” with the “label” of libertarian. Maybe you have to agree to every single point on a ninety-point platform about what the ideal society should be and how we’re going to pursue it. The first group is probably hopelessly conflict-prone and can only act in very large brush strokes. The second group can work together much more easily, but is smaller.

The limiting case of relaxedness is a national government, which “represents” everyone in an entire country, but which is so non-agenty that it is better viewed as a sort of exoskeleton-suit for other movements to take over and control rather than a goal-having movement in its own right. The limiting case of strictness is a single person.

Top-down control versus bottom-up control. Bottom-up control makes members happy, offers a guarantee against certain forms of insanity, and is a good way of resolving disputes and preventing outright civil war. Top-down control is more effective in terms of making sure the group’s actions are unified and not “designed by committee” in the perjorative sense of the phrase. It also means the people who founded the organization aren’t going to suddenly get outvoted by a membership that wants to do something else, which seems to be a surprisingly common problem. For example, the Republican Party started out as the party representing the racially enlightened and highly educated

North against the backward South. I wonder how that's been working out for them?

Closely related here is the problem of value drift. You can go a large part of the way to preventing value drift by some level of hard-coding of principles in a founding document (eg the US Bill of Rights) which is very difficult for future generations to change. On the other hand, if those values prove unexpectedly sub-optimal you get stuck having to say the founding document was "meant as a metaphor" or declaring in 1978 that you got a new revelation from God saying the previous revelation from God was received in error.

If you succeed in these tradeoffs, your reward is an organization that encompasses a large number of mostly-like-minded individuals who invest a lot of effort into working together for a common purpose. If you fail, you get organizations that can never get a coherent platform together or tear themselves apart in internal squabbles or civil wars.

II.

So much for a description of what an atheist ecclesiology might be about. What about the experimental results of such an ecclesiology?

Most of the people I know who have thought about this problem hard agree upon one major ecclesiological principle that neatly summarizes the gist of their investigation into this area:

The Catholic Church is really, really impressive.

It is the oldest continuously-operating organization in the world. It is the largest organization in the world, as measured in number of members. It is probably the richest non-state organization in the world. Although we can debate how closely they have stuck to the founding principles they had as of 114 AD or 1014 AD, they are doubtlessly a lot closer to those principles than, say, modern China is to 1014-AD-China.

Although there are other religions nearly as large as Catholicism – Islam and Hinduism, for example – they lack the same level of organization or really any organization at all. And although there are various governments that are probably a bit more powerful, they cheat by being able to throw anyone who doesn’t support them in jail. So what are the Church’s institutional choices, and how do they contribute to its longevity and success?

The first unusual thing I immediately notice about the Catholic Church is its insistence on turning group membership into a binary. You can be sort-of-libertarian, kind-of-libertarian, occasionally-libertarian-on-some-topics, or super-duper-libertarian – but the Catholics make it *very* clear that you are either A Catholic or Not A Catholic. You become a Catholic by going through the appropriate rituals, which are obvious and public and difficult to miss. You become Not A Catholic by things like official orders of excommunication.

I agree that this has become sort of washed out in recent years, to the point where there are people who are as just as vaguely Catholic as I am vaguely Jewish – that is, hardly at all except as a fuzzy feeling of connectedness to a group that shaped your culture. But as best I can tell, the Official Church Position is that this is degenerate, and that on God’s computer each person definitely has a Boolean variable representing whether they are Catholic or not.

Second, I notice the Catholic Church formalizes what beliefs and commitments Catholicism does and does not entail. These are the endless creeds and catechisms. Most other organizations have no good equivalent to this – not only is there no Feminist Catechism, but there’s not even a creed for specific limited feminist organizations like NOW or NARAL. Although party platforms are kind of close to this, I feel like on closer inspection they’re effects rather than causes of group membership. They’re talking about “Given the current makeup of the Republican Party, here are the sorts of things we expect Republican candidates to push during the next election”. They’re not saying “If you don’t believe every plank of this platform, get out of the GOP.”

Not only does the Church formalize where they demand conformity, but they formalize where they don’t. I don’t know enough to talk about this accurately, but I think that questions like “What was the Virgin Mary’s eye color?” can be debated by anyone with a half-baked theory and any bishop asked to intervene would get annoyed and say the Church has no opinion on this and shouldn’t force consensus. On the other hand, if someone asks “Was the Virgin Mary even Jesus’ real mother at all?” the Church politely informs you that they *are* forcing consensus on this question and you can either fall into line with the consensus or be declared a heretic and get *out* of the Church and *in* to the pit of eternal fire where the worm dieth not.

Fourth, although I don’t begin to claim to know enough theology to have credible things to say about the demands the Church may make, there are definitely occasional instances of, for example, Catholics being excommunicated for being part of abortion rights groups or the like. So it seems like they are pretty serious about being able to tell you what to do and expelling you if you don’t do it (with the caveat that in most cases with low-level proles they never bother to enforce it). But they likewise seem pretty serious about not abusing that power in stupid ways.

Fifth, the Church is hierarchical. There are many clearly defined levels, it’s obvious who is in charge of whom, and each level has to obey the levels above it or else. The Pope is in charge of all the levels and in theory everyone has to listen to him. You get promoted based on some combination of ability and politics.

Finally, the Church seems really big on rituals. A lot of them seem to be very clear IDENTIFY US AS YOUR IN GROUP AND THIS AS YOUR COMMUNITY NOW rituals – attempts to flip Haidt’s hive switch. Others just lend an air of dignity and grandeur to proceedings.

III.

A paradox: if the Catholic Church is the most successful organization in history, why don’t other organizations follow its example?

There are a few counterexamples here. National armies seem *very* similar to the Catholic Church in a lot of ways. You’re either in them or out of them. There are induction rituals and dismissal rituals. They are hierarchical with a general on top, colonels and majors in the middle, and the enlisted man on the bottom. When you’re in them, you have to do everything the higher levels say or you get kicked out and make a lot of people very angry. The whole thing is extremely full of rituals and everyone has to venerate various ritual objects in weird ways (for example, the national flag, or the Tomb of the Unknown Soldier). Fraternal organizations like the Freemasons seem to have something kind of similar going on here.

But that just makes the problems weirder. National armies are optimized for effectiveness and nearly everyone considers them impressive models of organizations that Get Things Done. If the Catholic Church – maybe the most culturally powerful organization in the world – and the US Army – doubtless the most physically powerful organization in the world – share a structure, isn't that a pretty strong point in favor of that structure?

Yet a lot of very sincere, maybe even fanatical movements – the libertarian movement, the feminist movement, the transhumanist movement, the socialist movement – don't seem to even be considering that model.

Their model is to have a large base of mostly atomized supporters, upon which float many different organizations. The supporters donate money to the organizations, and sometimes accept paying or volunteer jobs there. Occasionally they will wear the organization's logo on a t-shirt, or affix its bumper sticker to their car, but this is the extent of their identification. The organizations do not have membership rosters per se, except maybe a "donor list" or "supporter list" that exists mostly so they know who to email the newsletter to.

The only political-social organizations that even approach the Church model are political parties – and as we saw before, these fall short in a lot of ways. They demand nothing from supporters, their platforms are notably different from catechisms, and although they sort of have hierarchy and ritual it seems a bit forced and apologizing-for-itself most of the time.

So phrased differently, the paradox goes: why are there so many NGOs and so few Churches?

One answer I reject is that nobody wants to join a Church. I see among my friends something approaching *longing* for a good Church-like structure. This is hard for me to explain without naming names in a way I worry would be embarrassing, but it seems self-evidently true to me. Certainly this was something I saw a lot in micronations – a longing to be part of something bigger than ourselves.

Imagine that whatever cause you most support – libertarianism, transhumanism, effective altruism, feminism, whatever – had a Church type organization that you could join, led by the spokesperson for that movement you most respect (Ron Paul? Ray Kurzweil? Peter Singer? bell hooks?). You are welcome to go to their local community center, partake in a tasteful initiation ritual, and then they will ask you to do certain things for the good of the movement, which you will be assured the other members of the movement will also be doing. They will have a clearly printed list of what they do or don't demand consensus on, and members of the movement will follow it for the sake of maintaining cohesion. No more endless "Can you still be a feminist even if you don't…?" debates. The answer is always "Look at the list printed by the Feminist Pope (Mome?), if it's not on there, then yes". Also, their hierarchs wear cool clothes and occasionally speak in dead languages.

Unless every single cliched movie villain speech I have ever seen is wrong, humans long for someone to rule them and tell them what to do. Structure is good. Ability to pick your own in group is good. I have a really hard time imagining that there are no Churches because of a lack of willing adherents.

My guess is that for some reason we have a specific memetic immune response against Churches. Existing religions are grandfathered in. Everyone else who starts evolving towards such a design gets told they’re a cult and soundly mocked. I think this might have happened to Objectivism, which at one point might have been turning into a Church but now looks a little more like an NGO.

Chesterton’s Fence tells us we probably shouldn’t mess with this. But there is still a spot in my heart that misses all of the interesting Churches that there could have been.

Psychotropic Base Rates: The Argument From Antibiotics

The obscure antiprotozoal drug suramin has the prettiest molecular structure I've ever seen. It also has some evidence as a potential treatment for autism based on a cell danger response model of the condition.

I'm not going to get too excited here, because there are a *lot* of things that can cure stuff in mice. Still, one has to ask – an antiprotozoal? Really?

Protozoa are primitive little parasites kind of like bacteria. The most famous is plasmodium, which causes malaria. There's not much reason an antiprotozoal drug should cross-react with the brain, so if it does treat autism it's probably just a coincidence.

But it's a pretty common one. I've already noted how an antibiotic used to treat acne, minocycline, is a promising schizophrenia treatment. The news article on such is called "Scientists Shocked To Find Antibiotics Alleviate Symptoms Of Schizophrenia", but maybe by this point they should start being less shocked.

Off the top of my head, I can think of two other antibiotics with a significant psychiatric role. Iproniazid, the first antidepressant ever discovered, was originally used as an anti-tuberculosis drug – and just by eyeballing the chemical structure it's pretty easy to see the relationship to current-mainstay-of-tuberculosis-treatment isoniazid. There are records – amusing in retrospect – of doctors remarking on how *unusually* happy and excited patients were to finally be getting treatment for their tuberculosis. Eventually someone put two and two together. realized the drug itself was a mood-lifting agent, and the antidepressants were born.

Cycloserine is a totally different antitubercular drug that doesn't even share a chemical structure or mode of action with iproniazid. Nevertheless, it seems to affect classical conditioning in interesting ways, which makes it of use to psychiatrists. The most exciting possibility is that it speeds up the extinction response, meaning it could theoretically help someone "unlearn" behavior. There's a common use – which the evidence only ambiguously supports – where you use it to treat something like social anxiety disorder by having a patient take it in relatively safe social situations. When nothing bad happens (hopefully), the cycloserine speeds up the usual process of "unlearning" the social situation-fear link and the patient gets better more quickly than if they had to become comfortable with crowds the old fashioned way. There are also some proposed uses regarding cocaine and other drug addictions.

And these are just the ones with *good* psychiatric effects. Less positive psychiatric effects are a dime a dozen in antibacterials and antiprotozoals – for example, people on mefloquine do some pretty weird stuff.

There's no really good reason why antibiotics should have psychiatric effects. As mentioned before, beyond the fact that we're selecting for bioactive chemicals here, it's probably just coincidence. But that itself is a very interesting finding. If we think of antibiotics as chemicals chosen at random – as far as psychiatry is concerned – that means that random chemicals will often change mental processes around in important ways.

This shouldn't be surprising – the brain is full of stuff and pretty easy to chemically disrupt. But it's worth remembering. A lot of skepticism about new drugs – or new toxicity claims – comes from low base rates: the expectation that most chemicals are not active medications. But in psychiatry, the base rates might be higher than we think.

HeartMath Considered Incoherent

[Note: all opinions expressed here are my own. Nothing to be taken as medical advice.]

A group called the Institute of HeartMath has been *remarkably* persistent at making their way into my hospital. A couple of months ago I had to go to a lecture where they trained us all in their “scientifically” “validated” “heart” “coherence” “technique”. And yesterday I had to attend a class where one of my attendings (who is otherwise an amazing psychiatrist and teacher whom I have a huge amount of respect for) pushed the same technique and their biofeedback device.

And it looks like I’m not the only one. It looks like the US Navy is also “getting the coherence advantage” and that there’s an entire site dedicated to HeartMath for veterans and the military urging them to “apply for scholarships”. There are HeartMath training programs for teachers and managers ($3500 for a four-day workshop), for police, firemen, and first responders ($3699 for a four day workshop) and for doctors and clinicians ($1495 for an “interactive webinar”). There are HeartMath programs aimed at classrooms, including Early Heart Smarts Pre-K training ($179) and HeartMath Test Prep, $49 and apparently funded by a grant from the Department of Education. In case your classroom can’t afford these products, the Institute of HeartMath offers help filling out grant applications.

If any of these offers are actually being taken, this HeartMath stuff is big business. So in the process of writing up a letter to my boss explaining why I don’t want them back in my hospital a third time, I figured I’d make what I found public on the Internet for the benefit of anyone else looking into them.

Because their field of interest is heart electrophysiology, something I know almost nothing about, I'm not going to be able to do a good job debunking specific claims or responding to the science. Instead I want to make a few very general points about the science and then move into a discussion of the GIANT RED FLAGS the Institute throws up.

II.

According to a pamphlet I was given, HeartMath claims:

> Create a coherent state in about a minute with the simple, but powerful steps of the Quick Coherence Technique. Using the power of your heart to balance thoughts and emotions, you can achieve energy, mental clarity, and feel better fast anywhere…Find a feeling of ease and inner harmony that's reflected in more balanced heart rhythms, facilitating brain function and more access to higher intelligence.

The Quick Coherence Technique is a relaxation/focusing exercise where you concentrate on your heart area, breathe deeply while imagining the breath coming through your heart, and imagine a happy situation. According to HeartMath, this causes your heart rhythm to enter a state called "coherence", which looks like a sine wave on graphs of heart rate variability and which can be detected by cheap and simple monitoring devices.

They say that the heart has so many interconnected neurons that it is like a "second brain", and probably involved in various forms of advanced emotional processing. Further, "the heart sends more information to the brain than the brain sends to the heart", so getting the heart into a coherent rhythm can sync brain waves into a coherent rhythm and improve emotional states. They present lots of research showing their Coherence Technique does in fact change heart rate variability, brain waves, and performance on various tasks that require calm concentration.

Further, they say that “the heart has a magnetic field a thousand times stronger than that of the brain, the strongest of any organ in the body”. It can be detected up to several meters away, and its character changes with emotional state and with whether your heart is in “coherence” or not. They present links to a lot of research showing that subtle changes in the magnetic field of the heart can be measured even outside the body. Then they say that people can communicate emotional states with other people nearby through the effect of their hearts’ magnetic fields.

Therefore, if you get your heart in coherence with their meditative technique, you not only put your brain waves more in sync and eliminate your own stress, but you have a knock-on effect helping everyone around you.

This is a mixture of good science, mediocre science taken out of context, and total bunk.

Heart rate does have variability, and heart rate variability is an interesting proxy for your body’s general level of health and stress. You can find a good summary from an electrophysiological perspective here, and from more of a neurobiological perspective here. Most likely what happens is that when you’re calm, your heart gets more parasympathetic innervation which causes more variability, and when you’re stressed it gets more sympathetic innervation and less variability.

The heart does feed information to the brain. Then again, so does everything else. Your feet feed information to the brain – that's why if someone hits your feet, you can feel it. I don't know whether "the feet send more information to the brain than the brain sends to the feet" but it wouldn't surprise me if they did – they have to communicate temperature, pain, position, touch, itchiness, et cetera, and all the brain does is occasionally tell them to move somewhere. This does not mean the feet are metaphysically prior to the brain in some important way, or that they control the brain. It just means that the brain is at some level aware of what is going on with the feet. So too with the heart. We know the brain has some level of monitoring of heart function – this is why people who have heart attacks have various unpleasant feelings, including chest pain and a so-called "sense of impending doom". This doesn't imply very much about the heart controlling brain function.

The heart does have a complex interconnected nervous system of its own. But HeartMath's descriptions of it – which go from claims that "The heart's extensive intrinsic neurvous system is sufficiently sophisticated to qualify as a 'heart brain' in its own right" to the insane question The Heart Has A Little Brain – Which Is Really In Control? – are overblown. HeartMath says the heart has 40,000 neurons (other sources say more like 14,000). Okay. The brain has 86 billion. Which is really in control – the organ with 14,000 neurons or the one with 86,000,000,000? Yeah, it's the second one. Also of note: the gut has 100 million neurons. For those of you counting, that's seven thousand times more than the heart. Maybe "The Institute of BowelMath" didn't sound sexy enough? Neurons are useful structures that manage electrical conductivity and ability to react to external conditions; they don't always mean an organ has some kind of complicated emotional intelligence.

The heart does produce a magnetic field over a thousand times stronger than that of the brain. Here are other totally meaningless heart-brain comparisons: the heart is over a zillion times redder than the brain is! The heart is involved in 600000% more angsty teenage love poetry! Anything with electrical activity is going to produce a magnetic field, but that doesn't mean the magnetic field is of any deeper significance, or that "size of magnetic field produced" is a good proxy for "cognitive significance". In fact, we find that the magnetic field of the heart as measured at the surface of the body is ten million times weaker than the Earth's magnetic field at the surface of the Earth. HeartMath says that subtle changes in the heart's magnetic field can be measured outside the body, and this is true, but what they fail to mention is that this measurement was done at a super-high-tech laboratory in Berlin called the "most magnetically quiet room on earth" where building-sized magnetic shields sheltered the experimental apparatus from the Earth's magnetism, which otherwise would have totally overwhelmed the effect the same way as hunting for a firefly on the surface of the Sun. Outside of a special magnetically shielded room in Berlin, your heart's magnetic field isn't going around influencing everything around you, let alone interacting with somebody else's heart.

HeartMath does studies and finds that if I am holding your hand, your brain waves sync up to my heartbeat, and vice versa – and that indeed, this can happen even if we are nearby but not touching. Evidence for magnetic transfer effects? Before we say yes, I want to make three points about this study.

Number one, it is not in a peer reviewed journal. It's published in a book called "Brain And Values: Is A Biological Science Of Values Possible?", the editor of which is one of HeartMath's "scientific advisors".

Number two, it does not use p-values, Bayesian posteriors, or any other kind of statistic that involves numbers. It shows us pictures of wave patterns and points out that they look alike. I admit that they *do* look alike, but I know nothing about waves and for all I know it’s really easy to make different waves look alike.

Number three, EEG artifacts are a thing. That is, if any movement is going on near an EEG, it moves the electrodes and they record a noisy signal. Thus, you usually take an EEG with an EKG so you can see the patient’s heart rhythm and adjust it out, since otherwise the brain waves will appear to fluctuate with the heartbeat simply because the heartbeat shakes the electrodes. Manuals for EEG use have warnings about, for example, not letting anyone else sit on the patient’s bed during recording, or watching the patient’s intravenous lines because even the drip-drip-drip of the IVs can show up as perturbations. If I am holding your hand while you’re getting an EEG, perhaps my EEG reflects your heart rhythm not because your heart is affecting my brain waves, but because your heartbeat is indeed shaking me a tiny bit which shakes the electrodes which produces EEG artifact. This possibility seems to fit with HeartMath’s observation that when the heart-beat subject was wearing a thin glove on the hand with which she touched the brain-wave subject, the effect was decreased by a factor of ten. I admit this doesn’t explain the supposed sync between heartbeat and brain wave when the two subjects were standing a foot and a half apart without touching. But as we will soon see, HeartMath is so good at finding non-local effects that we have some reason to doubt their data-gathering process here.

Here’s what I think is going on as a fully general explanation of almost all of HeartMath’s research. Their Quick Coherence technique – and various others like it – are basically mishmashes of useful relaxation exercises stolen from various yogas and forms of meditation. Many of these ask you to focus on the heart – although many others ask you to focus on the tailbone, or genitalia, or third eye, or crown chakra – and all of them probably work in some vague way by redirecting your attention onto the body. I have no doubt that these yoga techniques effectively relax you. That changes your balance of parasympathetic versus sympathetic tone, which in turn affects your heart rate variability – which as we saw before, tracks parasympathetic and sympathetic tone. Since you're more relaxed, you do better at various cognitive tasks, which HeartMath then records and claims is evidence of an effect from heart "coherence". This explains about 80% of the Institute's findings. There are definitely some findings that can't be explained by this, but then, as we will very shortly see, there are some findings that can't be explained by anything except Alien Space Bats.

III.

I would now like to move from a sober critique of HeartMath's theories to an unfair character assassination of their staff.

Although HeartMath employs a bunch of people, the obvious two head honchos are founder Doc Childre (the CEO and President is listed as Sara Childre, who I assume is his wife), and Dr. Rollin McCraty, the executive vice-president and director of research, who is responsible for the lion's share of the Institute's research output and scientific claims.

Doc Childre has no medical training or relevant educational credentials. In fact, he is not a doctor at all. "Doc" is just his first name. This completes my character assassination of him.

Rollin McCraty is a doctor, but not a medical doctor. He has a Ph. D in “Health Sciences”, but all of his training and expertise is in electrical engineering and he has had no formal instruction in biology. His biography makes him sound very impressive:

> McCraty is a Fellow of the American Institute of Stress, holds memberships with the International Neurocardiology Network, American Autonomic Society, Pavlovian Society and Association for Applied Psychophysiology and Biofeedback

The Institute of Stress has a list of all its fellows online, McCraty is not mentioned.

The International Neurocardiology Network has no webpage or online evidence of its existence. When I Google “International Neurocardiology Network”, I get 47 results, every one of which is a claim by McCraty to be a member of it.

The American Autonomic Society does have a webpage, here. The webpage includes a helpful membership application where you can pay them $300 for membership, earning you a subscription to their journal and greatly decreased fees for attending their annual meeting. Their list of members is lorem ipsum text, but I’m *totally* willing to give Dr. McCraty the benefit of the doubt on this one.

The Pavlovian Society seems less prestigious than the American Autonomic Society, given that their membership application only involves a $30 fee and has to be sent by “mail, fax, or email”. What is this, 1995?

The Association for Applied Psychophysiology and Biofeedback not only sells membership for $189, but in several parts of its site equivocates between the terms "member" and "customer".

So of the five impressive-sounding organizations McCraty starts his bio with, one doesn't list him as a member, one doesn't seem to exist, and three give membership freely for a fee.

But aside from these organizational memberships, Dr. McCraty is widely published with many fascinating and well-accepted studies. Unfortunately, the fascinating ones aren't well-accepted, and the well-accepted ones aren't fascinating.

For example, on one hand we had the study showing my heart rate can affect your brain waves even when we're not touching, which would certainly be ground-breaking if true. But it was not peer-reviewed and was published in a random compendium associated with a HeartMath advisor.

And on the other hand, we have The Heart Re-Innervates Itself After Transplantation. This is in the *Annals of Thoracic Surgery*, an excellent peer-reviewed publication. But McCraty is one of twelve authors, and the study just shows that nerve growth goes on after heart transplant. Interesting if you're a thoracic surgeon, but not exactly the spooky-action-at-a-distance they were talking about before.

This seems to be a common problem with HeartMath. Looking at their list of publications, it seems to be about 50% studies they have published themselves without peer review, 25% studies published in journals of alternative medicine with no standards, and 25% studies in real journals that show relatively boring results. For example, A Controlled Pilot Study Of Stress Management Training of Elderly Patients With Congestive Heart Failure in the perfectly reputable journal *Preventative Cardiology* tests one of HeartMath's coherence-building relaxation techniques on the title population. They find that it in fact decreases stress, but "the twenty four hour heart rate variability showed no significant changes in autonomic tone". In other words, their claims are that they've discovered some master switch to the body that can even cross air gaps into other people's brains, but their reputable studies get results like "relaxation makes people less stressed".

IV.

Everything I've talked about so far – the "coherence" "techniques", the "second brain", the heart's magnetic field, the transfers of heart rhythms across air gaps – has been part of HeartMath's public-facing persona. This is the stuff they use to sound reasonable to doctors so they can get their techniques into hospitals and other sober institutions. Hold on tight, because we are going to start investigating the deranged world of HeartMath's non-public-facing core.

First let's expand on this idea of "coherence". Coherence just means your heart rhythm is in a nice sine wave pattern, right? Let's ask Coherence: Bridging Personal, Social, and Global Health, by Childre & McCraty, published in the journal *Alternative Therapies*' July 2010 edition. All emphasis mine:

> The heart plays a unique role in synchronizing the activity across multiple systems and levels of organization. The heart is uniquely well-positioned to act as the 'global coordinator' in the body's symphony of functions to bind and synchronize the system as a whole…

There is compelling evidence to suggest that **the heart’s energy field is coupled to a field of information that is not bound by the classic limits of time and space**. This evidence comes from a rigorous experimental study that investigated the proposition that **the body receives and processes information about a future event before the event actually happens**. Even more tantalizing are indications that the heart receives intuitive information before the brain does and that the heart sends a different pattern of afferent signals to the brain, which modulates the frontal cortex. This suggests that the heart is directly coupled to a subtle energetic field of information that is entangled in and interacts with **the multiplicity of energetic fields in which the body is embedded – including that of the quantum vacuum**…

Just as individual incoherence leads to pathologies within the individual, group incoherence leads to social pathologies – violence, abuse, terrorism, etc. There is a feedback loop between the individuals in a group and the group’s level of coherence. When individuals are not well self-regulated or are acting only in their own best interests without regard to others, it generates social incoherence…Unfortunately, social incoherence is characterized by a lack of unity, common purpose, peace, and harmony in or among families, neighbors, or employees in workplace environments.

The Global Coherence Initiative is a science-based organization focused on examining the interactions between humans and the Earth’s energetic fields. One of the project’s hypotheses is that **the Earth’s magnetic and geomagnetic fields created in the ionosphere in turn create bidirectional feed-**

> **forward and feedback loops within the collective emotional energy of humanity**. More and more people are realizing that solar and universal energetic influxes are a part of a natural cycle with potential benefits to humanity. Yet people have a responsibility for their own energy and how it can be used to create deeper connections and more caring interactions with others and with the Earth itself, including all living entities.
>
> If, as some content, all living systems are indeed interconnected and communicate with each other via biological and electromagnetic fields, it stands to reason that humans can work together in a concreative relationship to consciously increase global coherence. This can only occur when enough individuals and social groups increase their coherence baseline and utilize that increased coherence in innovative problem solving and intuitive discernment for addressing social, environmental, and economic problems. In time, global coherence will be indicated by countries adopting a more coherent planetary view. At this level of scale, social and economic oppression, warfare, cultural intolerance, crime, and disregard for the environment can be addressed meaningfully and successfully.

Strong claims. Any research to back that up?

Well, yes. But it’s called The Psychophysiology of Entrepreneural Intuition: A Quantum-Holographic Theory, and says that:

> A new study shows that both the brain and the heart are involved in processing a pre-stimulus emotional response to the future event. Drawing on this research and on the principles of quantum holography, we

> develop a theory of intuitive perception. The theory explains how focused emotional attention directed to the object of interest (such as a potential future business opportunity) attunes the psychophysiological systems to a domain of quantum-holographical information, which contains implicit information on the object's future potential. The body's perception of such implicit information about the object's future is experienced as an intuition.

In other words, entrepreneurs tap into the nonlocal holographic nature of reality in order to get hot startup tips. At this point I probably don't need to add that the Institute of HeartMath is based in the Bay Area.

Can we get weirder? I think we can. A HeartMath press release: You Can Change Your DNA:

> Many people have mistakenly believed that the DNA with which we are born is the sole determinant for who we are and will become, but scientists have understood for decades that this genetic determinism is a flawed theory.

Then they go on to bring up epigenetics, which is quickly replacing quantum mechanics as the Thing I Most Expect To Be Brought Up In Situations Like This. To reputable scientists, epigenetics means that the methylation of genes affects functions. But to HeartMath it means:

> After two decades of studies, HeartMath researchers say other factors such as the appreciation and love we have for someone or the anger and anxiety we feel also influence and can alter the outcomes of each individual's DNA blueprint…The influence or control individuals can have on their DNA – who and what they

> are and will become – is further illuminated in HeartMath founder Doc Childre's theory of heart intelligence. Childre postulates that "an energetic connection or coupling of information" occurs between the DNA in cells and higher dimensional structures – the higher self or spirit.

Go on…

> When we activate the power of our hearts' commitment and intentionally have sincere feelings such as appreciation, care and love, we allow our hearts' electrical energy to work for us. Consciously choosing a core heart feeling over a negative one means instead of the drain and damage stress causes to our bodies' systems, we are renewed mentally, physically and emotionally. The more we do this the better we're able to ward off stress and energy drains in the future. Heartfelt positive feelings fortify our energy systems and nourish the body at the cellular level. At HeartMath we call these emotions quantum nutrients.

There's our quantum mechanics!

But is there proof?

Oh, yes. There is the *best* proof.

Modulation Of DNA Conformation By Heart-Focused Intention is a paper by McCraty (again), Atkinson, and Tomasino. The methodology is simple: the subject (in one case, Doc Childre himself) brings their heart rhythm into "coherence", then stares at a beaker of DNA and *wills* it to unwind. The DNA complies. According to the paper, 10.27% of DNA willed at in this way unwound, compared to only 1.09% of control DNA ($p < 0.01$). Since this result is obviously too boring to even be worth mentioning, the experimenters up the ante by testing the "nonlocal" version of the effect. Instead of holding the beaker in her hands, the subject wills DNA in a laboratory half a mile away to unwind. Once again, a highly significant result (2.76% change, $p < 0.01$). I don't see any obvious screwups in this paper, aside from the conclusion. The skeptical Internet doesn't seem to be of much help either. I can think of a lot of potential problems - waiting different amounts of time to measure the DNA in the two samples, exposing them to different amounts of light, et cetera - but the methods section of the paper doesn't give me any particular reason to think these happened. And they go into great detail to describe their blinding procedures, all of which seem appropriate. But still. You got DNA to unwind by *asking politely*. From half a mile away. If this were in a peer-reviewed journal, I’d still be doubtful. If it were in a peer-reviewed journal and had been replicated five times by five different teams, I’d *still* be doubtful. If it were in a peer-reviewed journal and had been replicated ten times by ten different teams including several skeptics and had a strong theory behind it that was well-supported in other ways, I might grudgingly accept it. But we are not at that level. We’re at one experiment, once, not peer-reviewed. At this point, you do *not* get to conclude that:

> The heart serves as a key access point through which information originating in the higher dimensional structures is coupled into the physical human system (including DNA), and that states of heart coherence generated through experiencing heartfelt positive emotions increase this coupling.”

Anyway, the question is: can we get even weirder than *this*?

Well, I dunno. What do the geomagnetic field, the inauguration of Barack Obama, and a random number generator have in common? If you answered “Nothing, as far as I know,” then yes, we can get weirder.

The Global Coherence Initiative is a project measuring how large-scale events affect some kind of feedback loop between people’s emotional rhythms and the geomagnetic field. The goal is to get so many people into heart-rhythm-coherence that it creates some kind of “global coherence” and, reading between the lines, immanentizes the eschaton.

But all that’s in the future. Right now they only have 10,000 people in 56 countries, who respond to “emergencies” by bringing their heart rhythms into coherence and sending out coherence waves in the appropriate direction. According to the site:

> Even as the GCI was still gearing up in startup mode, these members, plus countless others they engaged within their families and communities, responded to several GCI alerts to send coherent energy and care to critical areas of need and crisis around the planet. These efforts of coherent heart are crucial and appreciated. Alerts went for the victims of Hurricane Gustav, conflicts in the Middle East and Democratic Republic of Congo, the financial meltdown and more.

Imagine how screwed up the Middle East would be right now if people *weren’t* sending coherent energy towards it!

Clearly the Global Coherence Initiative needs to up its game. That’s why they’re asking for your donations to buy $60,000 worth of giant magnetic coils. They say it’s for world peace, but honestly, when your first name is “Doc”, and you run a shadowy organization that is studying telepathic alteration of DNA, and you want $60,000 worth of magnetic coils, I start to get *really* suspicious.

Anyway, even without their giant coils they are doing good work. And by good work, I mean analyzing how random number generators reacted to Barack Obama's inauguration. You or I might expect that the generators reacted *randomly*, but that is why we are random shmucks instead of people named "Doc" running institutes that are in the process of procuring $60,000 worth of magnetic coils. The Institute of HeartMath says the the outpouring of joy following the inauguration caused both a decrease in the variance of the random numbers produced by generators all around the world *and* subtle but observable fluctuations in Earth's magnetic field. They note that their partner organization, the Global Consciousness Project, says that "occasions that are meditative and celebratory are often associated with persistent low network variance," where "networks" here tend to be things like random number generators and the geomagnetic field.

With apologies to Obama himself, that is not *exactly* the kind of change I can believe in.

V.

We tend to think of alternative medicine practitioners as obvious loons with websites out of 1995 where all the words are IN CAPITAL LETTERS. But sometimes, they're people with Ph. Ds and a bunch of papers published in prestigious journals who are able to focus on the less controversial aspects of their ideas well enough to infiltrate clinics and hospital systems.

HeartMath's website is impeccable. Their representatives gave a presentation to a hospital full of doctors – including cardiologists and neurologists – without any missteps that made them look anything less than reputable. Their Board of Scientific Advisors contains some really serious intellectual clout like Abdullah Abdulrahman Al Abdulgader, who is both literally and figuratively a big-name cardiologist as well as leading the entire medical field in number of times the word “Abdul” appears in his name. These are *top-notch* people.

And then you look a little deeper and you find out that their cute little relaxation exercises are actually a plot to connect to higher dimensions beyond time and space and immanentize the eschaton by messing with Earth’s magnetic field, possibly with the help of $60,000 worth of giant coils and/or Yog-Sothoth.

Remember, these people are working with hospitals, with the military, with the police, and willing to helpfully explain how to apply for grants to bring their technology into the classroom. And they are *total loons*.

I think heart rate variability is an important concept. And I agree that relaxation exercises derived from yoga are a good way of helping people suffering from stress and even psychiatrically diagnosable anxiety disorders. I actually tried their Quick Coherence technique and it made me feel really good.

But I don’t think giving $3699 to HeartMath to teach you about it is a good investment, and I don’t think they are the best people to be furthering the study of these ideas.

Promising The Moon

I.

The year 1969 comes up to you and asks what sort of marvels you've got all the way in 2014.

You explain that cameras, which 1969 knows as bulky boxes full of film that takes several days to get developed in dark rooms, are now instant affairs of point-click-send-to-friend that are also much higher quality. Also they can take video.

Music used to be big expensive records, and now you can fit 3,000 songs on an iPod and get them all for free if you know how to pirate or scrape the audio off of YouTube.

Television not only has gone HDTV and plasma-screen, but your choices have gone from "whatever's on now" and "whatever is in theaters" all the way to "nearly every show or movie that has ever been filmed, whenever you want it".

Computers have gone from structures filling entire rooms with a few Kb memory and a punchcard-based interface, to small enough to carry in one hand with a few Tb memory and a touchscreen-based interface. And they now have peripherals like printers, mice, scanners, and flash drives.

Lasers have gone from only working in special cryogenic chambers to working at room temperature to fitting in your pocket to being ubiquitious in things as basic as supermarket checkout counters.

Telephones have gone from rotary-dial wire-connected phones that still sometimes connected to switchboards, to cell phones that fit in a pocket. But even better is bypassing them entirely and making video calls with anyone anywhere in the world for free.

Robots now vacuum houses, mow lawns, clean office buildings, perform surgery, participate in disaster relief efforts, and drive cars better than humans. Occasionally if you are a bad person a robot will swoop down out of the sky and kill you.

For better or worse, video games now exist.

Medicine has gained CAT scans, PET scans, MRIs, lithotripsy, liposuction, laser surgery, robot surgery, and telesurgery. Vaccines for pneumonia, meningitis, hepatitis, HPV, and chickenpox. Ceftriaxone, furosemide, clozapine, risperidone, fluoxetine, ondansetron, omeprazole, naloxone, suboxone, mefloquine, – and for that matter Viagra. Artificial hearts, artificial livers, artificial cochleae, and artificial legs so good that their users can compete in the Olympics. People with artificial eyes can only identify vague shapes at best, but they’re getting better every year.

World population has tripled, in large part due to new agricultural advantages. Catastrophic disasters have become much rarer, in large part due to architectural advances and satellites that can watch the weather from space.

We have a box which you can type something into and it will tell you everything anyone has ever written relevant to your query.

We have a place where you can log into from anywhere in the world and get access to approximately all human knowledge, from the scores of every game in the 1956 Roller Hockey World Cup to 85 different side effects of an obsolete antipsychotic medication. It is all searchable instantaneously. Its main problem is that people try to add so much information to it that its (volunteer) staff are constantly busy deleting information that might be extraneous.

We have the ability to translate nearly major human language to any other major human language instantaneously at no cost with relatively high accuracy.

We have navigation technology that over fifty years has gone from “map and compass” to “you can say the name of your destination and a small box will tell you step by step which way you should be going”.

We have the aforementioned camera, TV, music, videophone, video games, search engine, encyclopedia, universal translator, and navigation system all bundled together into a small black rectangle that fits in your pockets, responds to your spoken natural-language commands, and costs so little that Ethiopian subsistence farmers routinely use them to sell their cows.

But, you tell 1969, we have something more astonishing still. Something even more unimaginable.

“We have,” you say, “people who believe technology has stalled over the past forty-five years.”

1969’s head explodes.

II.

It’s the anniversary of the moon landing, which means I have to deal with people passing around memes like this:

45 years after the Ford Trimotor
in
another 45 we'll have Twitter!

But I probably can't blame the date for the recent discussion here of whether technological progress halted in 1972.

So I would like to take a moment to critique a certain strain of futurology.

There seems to be this thing where people imagine something that would look really cool, and predict that if we work hard on it for fifty years, we'll be able to pull it off. And then fifty years later, when barely any work has been done on it at all, they start looking for someone to blame.

Missions to Mars. Lunar colonies. Giant floating solar power satellites. Undersea domes. Ten mile high arcologies. Humanoid robots.

Whereas real technology doesn't advance by heading in the direction of something that looks cool, unless some government or tycoon is throwing lots of money in the direction of coolness. Real technology hill-climbs towards things that are useful and profitable.

Why haven’t we colonized space yet? For the same reason we haven’t colonized Antarctica. It’s very cold and not a lot of fun and if you go outside you die.

In fact, Antarctica is preferable to space in pretty much every way. There is no reason to colonize space before you have finished colonizing Antarctica. And there is no reason to colonize Antarctica until you have finished colonizing Nebraska (population: 9 people per square km).

I will maintain that even if we had enough space flight technology that elementary school classes routinely took field trips to Mars, Mars would end up with two or three scientific bases, a resort where tourists could take their pictures on Olympus Mons, a compound of very dedicated libertarians, and nothing else. No domed cities. No colonies fighting for independence. Think that’s implausible? School children take field trips to the Mojave Desert all the time, and it pretty much looks like that. Why should Mars prosper more than a much more habitable comparison area?

Likewise, the reason we don’t build undersea domes isn’t because we’re not good enough. It’s because humans breath better on land, and there’s still a lot of land left to live in. On the rare occasion we want a resource located underwater, we build an oil rig on top of it and pump it from the surface, ie the part of the ocean where you don’t get insta-crushed by ten atmospheres of pressure if something goes wrong.

And the reason there are no ten-mile-high arcologies is that we haven’t already tiled all the desirable real estate with 9.9-mile-high arcologies and decided we still need more space.

Science fiction authors and would-be prophets stubbornly refuse to admit “would anybody reasonably pay money for this?” into their calculations. And so every ten years they end up predicting the “smart house”. Where from your phone, you can control the lights in any room of the house! I imagine futurologists sitting in their kitchens, thinking “Oh no! I wish the lights were on in my bedroom, but all I have is my phone!” Maybe one day we will have houses that contain teleporters that can bring to any other building in the world without stepping outside. But if you’re in the kitchen and you want the light on in the teleporter room, you’ll still just walk to the teleporter room and flip the @#$%ing switch.

I am not defending this as a normative view of how progress should work. There is a lot to be said for colonizing Mars as a survival strategy in case something unexpected happens to Earth. And there’s also a lot to be said for Manhattan Project style efforts to discover a technology in a non-hill-climbing way, something where there’s not a profitable transitional form at each step between where we are and what we want. But I would suggest we stick to those criticisms, and not to a criticism of advance per se.

(actually, we’re not even all that bad at getting past the hill-climbing thing; government subsidies to solar seem to have been a very successful attempt to push solar out of an area where it wasn’t profitable to improve into an area where it is)

But it’s going to take some pretty creative accounting to make moon shots profitable. The main reason people funded the moon landing in 1969 (as opposed to the reason that people not involved in funding felt good about it) was to beat Russia and then get to rub it in their face forever. Nowadays that’s no longer so fun (although rapidly becoming funner!) Therefore, we get the expected outcome of fewer moon shots until someone else thinks of a compelling incentive to go to the moon. So far there isn’t one. There’s no need to bring technological stagnation into the picture.

No Skyscraper Stagnation

Since we're on the subject of whether technological progress has stagnated, I thought I'd address an issue that always seems to come up sooner or later.

Is there a decline in American skyscrapers possibly indicating a decline in American civilization?

I realize that some people may live in happy little bubbles where this question does *not* in fact always seem to come up sooner or later. But I tend to hear it a lot. It seems to be a favorite of former SSC commenter James Donald. From here:

> The twin towers were big buildings, and buildings of that size cannot be built under progressive regimes. As the trabant was a fraud, created to disguise the fact that the Soviet Union could not build consumer cars, the One World Trade Center is a fraud, created to disguise the fact that the US can no longer build buildings as large as it used to be able to build.

He has more in the same vein here, here, and at his blog. But it's not just Jim. From Countercurrents:

> It's not a coincidence that the tallest buildings in America were built during the 1970s. What we didn't realize at the time was that we would never again have it so good. The 1970s represented a "tipping point," to use the popular vernacular, for the American Dream.

Vox Day writes:

> I suspect that human capability reached its peak or plateau around 1965-75 – at the time of the Apollo

> moon landings – and has been declining ever since. This may sound bizarre or just plain false, but the argument is simple. That landing of men on the moon and bringing them back alive was the supreme achievement of human capability, the most difficult problem ever solved by humans. 40 years ago we could do it – repeatedly – but since then we have *not* been to the moon, and I suggest the real reason we have not been to the moon since 1972 is that we cannot any longer do it. Humans have lost the capability.

He doesn't bring up skyscrapers, but a commenter does.

Or, to totally remove any subtlety, here's The Decline Of The West As Measured By The Rise Of New Skyscrapers.

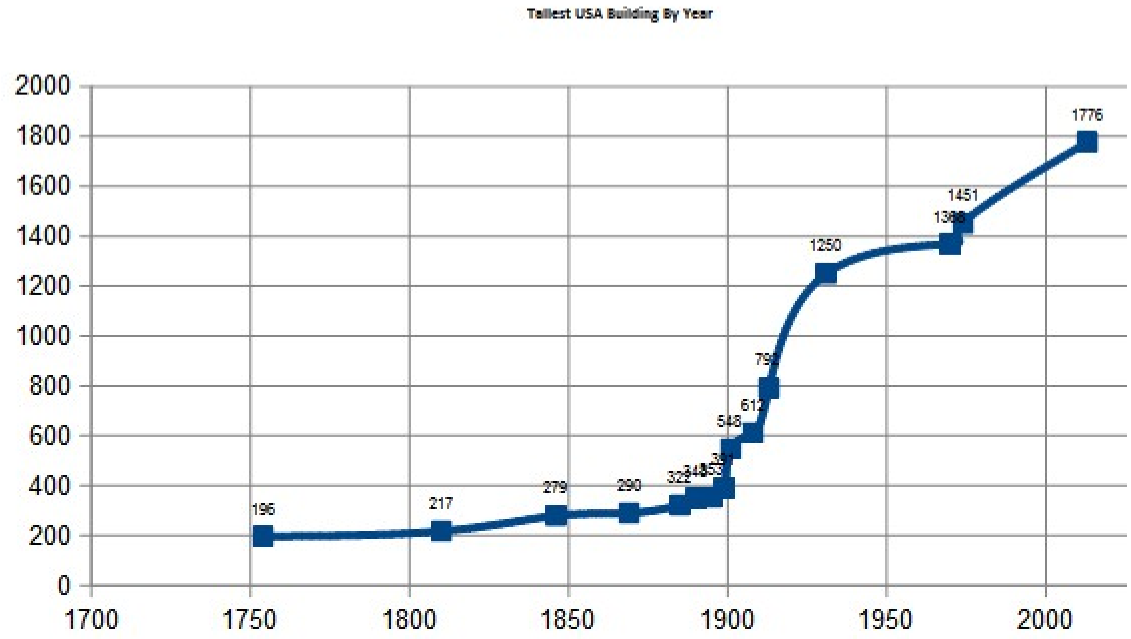
The typical response to this sort of thing is to bring up the studies showing that increased skyscraper construction is in fact correlated with not with progress, but with economic decline. The Skyscraper Index is a whimsical attempt to use skyscraper construction to predict economic downturns. And that if the world's largest skyscrapers are currently in Dubai and Saudi Arabia, maybe skyscrapers are less a sign of national glory and more a sign of having more oil than sense.

But it looks like Jim isn't a fan of that argument. So let's take a different tack:

America's capacity to build skyscrapers isn't decreased at all, in any way, whatsoever.

(says the guy with the secret guilty skyscraper obsession)

Here is a graph of the height in feet of the tallest skyscraper in America, by year.



In the mid-1700s, the tallest building in the US was [Christ Church](#) in Philadelphia at 196 feet. There’s some underwhelming progress until about 1900, when a thirty-year spurt takes us from 391 feet to over a thousand. This spurt ends with the Empire State Building (1,250 feet) in 1931. There is then a forty year dry spell ending with the construction of the WTC and Sears Tower in rapid succession, then another forty year dry spell ending with the construction of One World Trade Center last year.

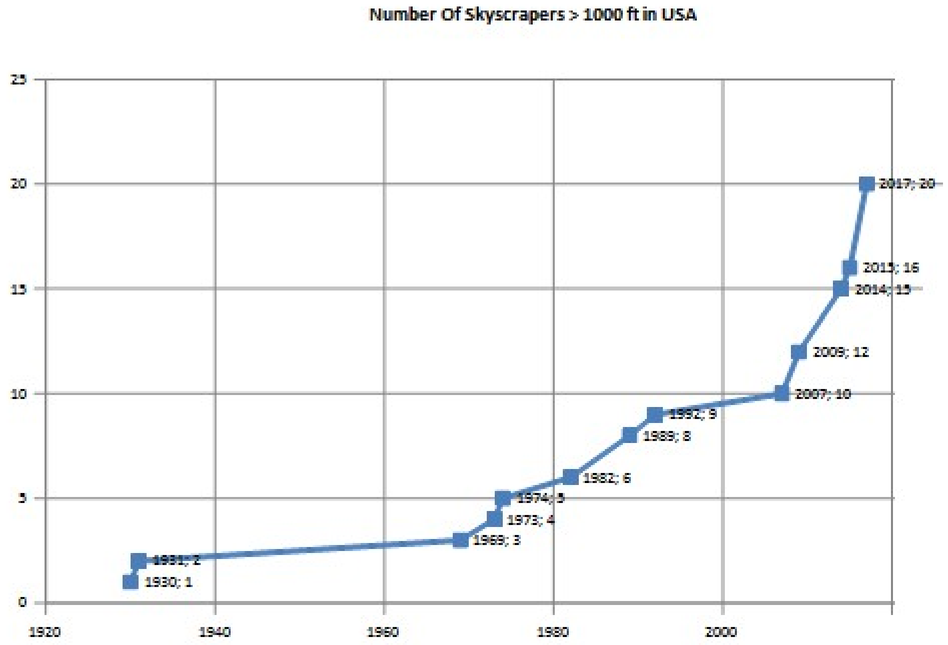
After the 1900 – 1930 spurt (which corresponded to the first widespread use of steel and elevators) growth is extremely linear.

Jim [argues](#) this is unfair because One World Trade Center, the recent data point beating the old Sears Tower and WTC, has an unusually large spire inflating its height. This is true.

If we ignore spires and concentrate on roof height, the old WTC was 1368 feet and the new WTC is also 1368 feet (coincidence?), showing little progress. Luckily for our argument, the Nordstrom Tower currently under construction in New York has a roof height of 1,479 feet, a good one hundred feet higher than the WTC and enough to restore linearity.

[Note: I feel bad arguing against Jim after banning him from commenting. I'll unblock him from this post's comment section for purposes of fairness.]

A different measurement might be concentrating on quantity rather than quality of skyscrapers. This graph shows the number of supertall (> 1000 ft high) skyscrapers in the US over time:

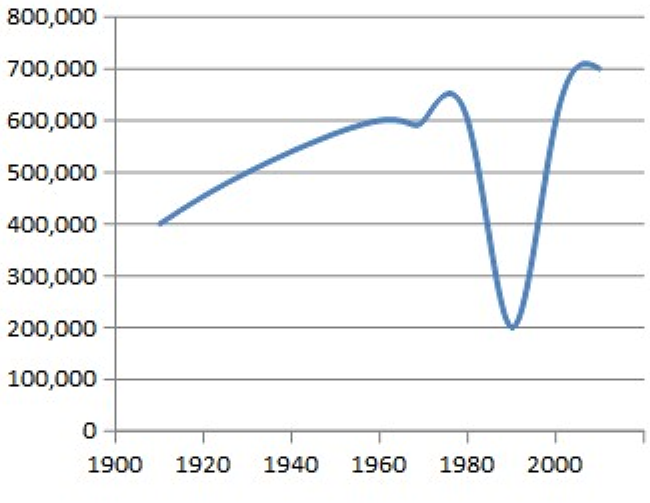


Two built in the 1930s (Empire State and Chrysler Buildings), none from 1931 to 1969 (our dry spell), a gradual trickle from 1969 to 2007 or so, and a sudden recent explosion. Thus the excitement about New York's recent supertall boom. The current spurt includes the previously mentioned One World Trade Center, the Nordstrom Tower (which will be one foot lower than 1WTC at 1,775 feet), 432 Park Avenue (“only” 1,398 feet, but higher than 1WTC without its spire). There are more supertall buildings scheduled for construction in the Hudson Yards redevelopment project in New York than existed in the entire United States in 1973.

This isn’t even counting the really ambitious projects, like the plan to build a 2,000 foot tower on the site of the old Chicago post office. Which might sound overly ambitious, except that Chicago just finished a 1,389 foot Trump Tower.

It is true that China is now building more skyscrapers than the US. So there is an argument for *relative* decline. But the argument for absolute decline is much less strong. But of note, China also has four times the population density of the US, probably much more when you take into account the small portion of its territory where people actually live. That’s a pretty strong incentive to build higher.

Finally, one more graph – this one a little more complicated.



This is the cost per foot of building a skyscraper over time.

My methodology was to take the tallest skyscraper built during each decade, convert its cost into 2013 dollars, and divide it by the number of feet high in the skyscraper.

I did not follow this methodology exactly, because the tallest skyscraper of the 2010s is One World Trade Center, which cost about three times more per foot than any other skyscraper on the list. According to the Wall Street Journal:

> One World Trade Center's construction is vastly more expensive than a traditional office tower, in large part due to security costs associated with building the tallest building in North America on a site that has been the target of two separate terrorist attacks (the site was also bombed in 1993). Once known as the Freedom Tower, the 1,776-foot skyscraper sits atop a heavily reinforced, windowless podium. It also has a thick core of concrete and steel around its elevator shafts. By comparison, other-high profile buildings around the world have been far less expensive. The Port Authority long ago gave up hope that One World Trade would be a profitable investment in the short- or mid-term.

So 1WTC was a crazy outlier because they had to make it terrorist-proof. Rather than have it completely throw off the graph, I replaced it with the next tallest 2010s building I had cost information on, which I think was Four World Trade Center.

I was worried that this was unfairly penalizing newer skyscrapers because the 1000th foot probably costs more to build than the first and newer skyscrapers are taller. But I reran the analysis using the building from every decade closest to 1,000 feet (many were within 10 ft of the target) and got very very similar results (not shown).

The important lesson to take from this graph is that if you're building a skyscraper, you should definitely hire whoever built Bank of America Plaza, the extreme outlier in 1990 that throws off the otherwise smooth curve with a sudden precipitious dip. They somehow built a very pretty building for one-third of the cost of everyone else in history, so kudos to them.

(but if it falls over the next time there's a strong breeze, and the investigation finds it was actually made out of Styrofoam, don't say I didn't warn you)

The other important lesson is that skyscraper costs have changed little if at all since the age of the Empire State Building. This is important because one of the other decline arguments I get all the time is that building anything is so bureaucratized and expensive and bogged down by building codes and environmental compliance checks that nobody will do it. A friend recently told me – I can't find the original quote so I can't make sure I've got it right – that Estonia is considering building a tunnel all the way across the Baltic Sea for the same price it costs New York City to build four blocks worth of subway, presumably because NYC is so bureaucratized. I don't know if that's true with regard to subways, but if so skyscrapers seem to have escaped the worst of it.

One more point. The most notable thing I turned up researching this post was the extent of the 1930 to 1960s skyscraper dry spell. Ten buildings taller than 700 feet were built in 1933 or before – and zero from 1933 to 1960. Skyscraping didn't really recover to its 1930 heights (no pun intended) until 1973 or so.

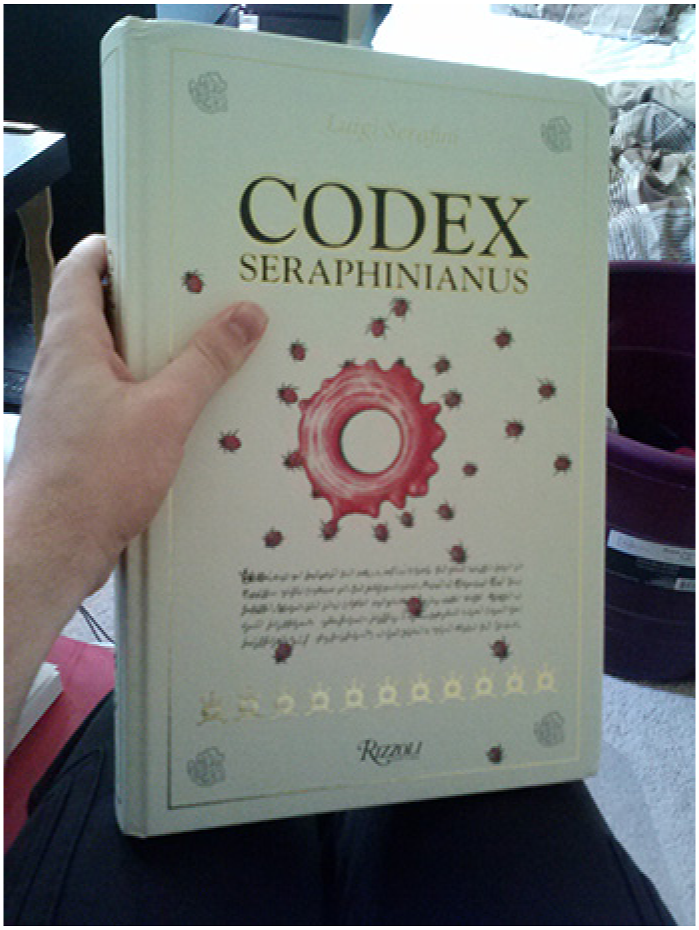
Anyone wanting to talk about a collapse of civilization from 1940 to 1970 would have had a lot of evidence from skyscraper decline. But that was exactly the period when most people today think technological progress was at its height!

I conclude that skyscrapers are not a very a good indicator of anything.

The Other Codex

I did okay for myself this first year at my job and had a little bit of money left over. I told myself I was allowed to buy one moderately ridiculous luxury item as a reward. And ever since I was like nineteen there's only been one moderately ridiculous luxury item I really wanted.

So now I've got it:



I kind of want to explain, but part of me knows that no one can tell you what the Codex Seraphinianus is. You have to see it for yourself.

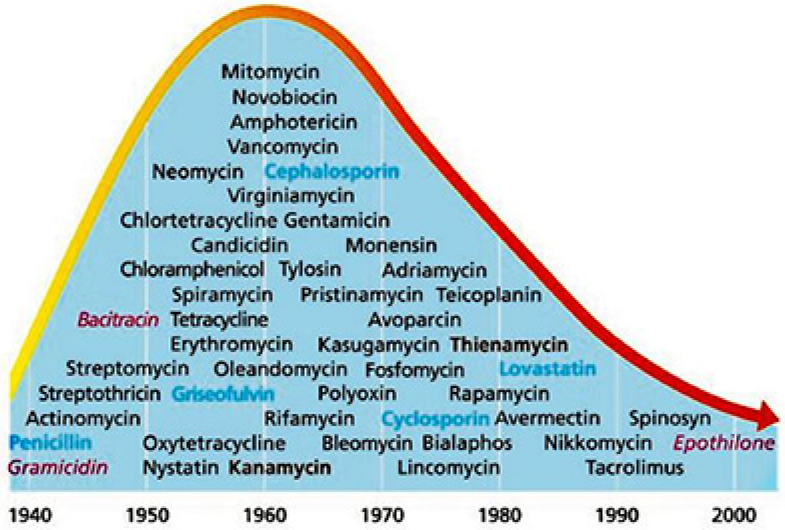
(the above is 50 MB high-resolution PDF file I'm hosting my old website. When that goes down under the strain, you can switch to a more manageable version here)

I was prepared to pay $500 for it, which is what it cost five years ago when I first looked into purchasing it, but to my delight there was a new version selling on Amazon for only $80. Getting the only ridiculous luxury item I've ever really wanted for $80 seems pretty good.

[That wasn't originally intended to be an ad. But I realized it was stupid to accidentally advertise something without getting paid for it, so I signed up for Amazon Affiliates program. So if you thought that sounded like an advertisement, you've been Gettier-cased.]

Some Antibiotic Stagnation

In the past week I’ve written about antibiotics and the rate of technological progress. So when a graph about using antibiotics to measure the rate of scientific progress starts going around the Internet, I take note.



Unfortunately, it’s inexcusably terrible. Whoever wrote it either had zero knowledge of medicine, or was pursuing some weird agenda I can’t figure out.

For example, several of the drugs listed are not antibiotics at all. Tacrolimus and cyclosporin are both immunosuppressants (*please* don’t take tacrolimus because you have a bacterial infection). Lovastatin lowers cholesterol. Bialafos is a herbicide with as far as I can tell no medical uses.

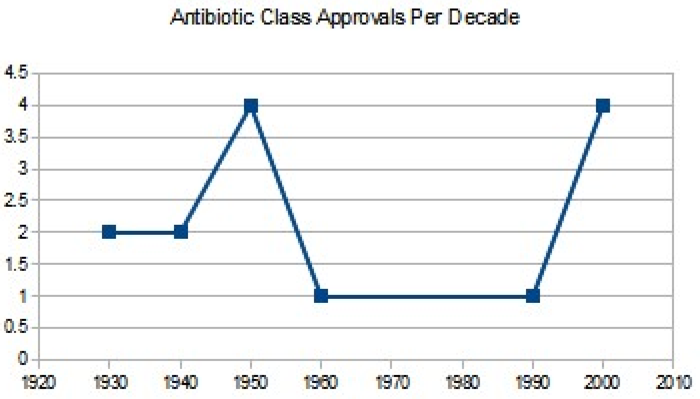
“Cephalosporin” is not the name of a drug. It is the name of a class of drugs, of which there are over sixty. In other words, the number of antibiotics covered by that one word “cephalosporin” is greater than the total number of antibiotics listed on the chart.

If I wanted to be charitable, I would say maybe they are counting similar medications together in order to avoid giving decades credit for producing a bunch of “me-too” drugs. But that doesn’t seem to be it at all. They triple-count tetracycline, oxytetracycline, and chlortetracycline, even though they are chemically very similar and even though the latter are almost never used (the only indication Wikipedia gives for the last of these is that it is “commonly used to treat conjunctivitis in cats”)

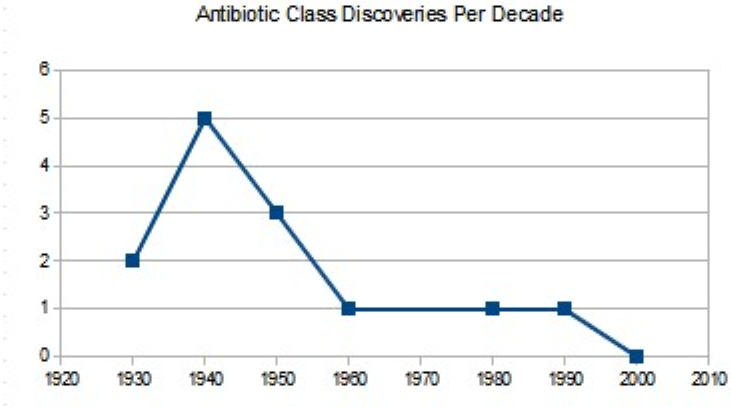
They also leave out some very important antibiotics. For example, levofloxacin is a mainstay of modern pneumonia treatment and the eighth most commonly used antibiotic in the modern market. That’s a whole lot more relevant than cats with pinkeye, but it is conspicuously missing while chlortetracycline is conspicuously present. Maybe it has something to do with levofloxacin being approved in 1993?

(also missing from the 90s and 00s: piperacillin, tazobactam, daptomycin, linezolid, and several new cephalosporins)

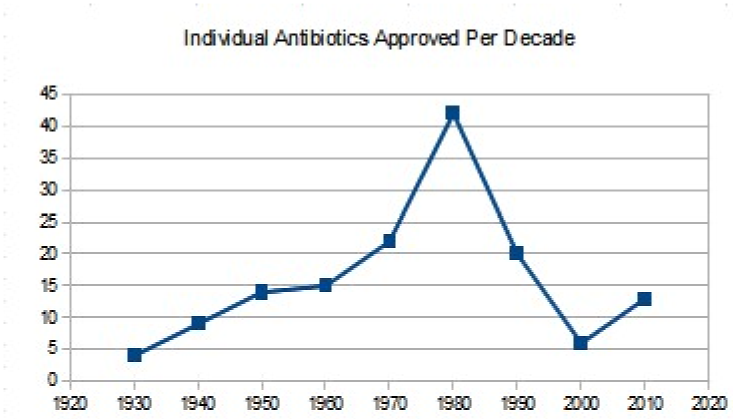
I can’t find a real table of antibiotic discovery per decade, so I decided to make some.



Antibiotic classes approved per decade. An antibiotic class is a large group of drugs sharing a single mechanism. For example, penicillin and methylpenicillin are in the same class, because one is just a slight variation on the chemical structure of the other. There are some arguments over which new drugs count as a "new class". Source is this source.



Anbiotic classes discovered per decade. The first graph lists when classes got FDA approval, this one lists when they were first discovered. Both have different pluses and minuses. This one is likely to undercount recent progress because the chemicals being discovered now haven't been tested and found to be valuable antibiotics and approved (and so don't make it onto the graph). But the last one might overcount recent progress, because a drug being "approved" in 2010 might represent the output of 1980s science that just took forever to get through the FDA. On the other hand, it might not – a lot of times people find a drug class in 1980, decide it's too dangerous, and only find a safer useable drug from the same class in 2010.



Individual antibiotics discovered per decade. Individual antibiotics can still be vast improvements upon previous drugs in the same class. Source is here. I have doubled the number for the 2010s to represent the decade only being half over and so make comparison easier. About half of that spike around 1980 is twenty different cephalosporins coming on the market around the same time.

I conclude that antibiotic discovery has indeed declined, though not as much as the first graph tried to suggest, and it may or may not be starting to pick back up again.

Two very intelligent opinions on the cause of the decline. First, from a professor at UCLA School of Medicine:

> There are three principal causes of the antibiotic market failure. The first is scientific: the low-hanging fruit have been plucked. Drug screens for new antibiotics tend to re-discover the same lead compounds over and over again. There have been more than 100 antibacterial agents developed for use in humans in the U.S. since sulfonamides. Each new generation that has come to us has raised the bar for what is necessary to discover and develop the next generation. Thus, discovery and development of antibiotics has become scientifically more complex, more expensive, and more time

> consuming over time. The second cause is economic: antibiotics represent a poor return on investment relative to other classes of drugs. The third cause is regulatory: the pathways to antibiotic approval through the U.S. FDA have become confusing, generally infeasible, and questionably relevant to patients and providers over the past decade.

A particularly good example of poor regulation from the same source:

> When anti-hypertensive drugs are approved, they are not approved to treat hypertension of the lung, or hypertension of the kidney. They are approved to treat hypertension. When antifungals are approved, they are approved to treat “invasive aspergillosis,” or “invasive candidiasis.” Not so for antibacterials, which the FDA continues to approve based on disease state one at a time (pneumonia, urinary tract infection, etc.) rather than based on the organisms the antibiotic is designed to kill. Thus, companies spend $100 million for a phase III program and as a result capture as an indication only one slice of the pie.

And one more voice, which I think of as a call for moderation. This is Nature Reviews Drug Discovery:

> Most antibiotics were originally isolated by screening soil-derived actinomycetes during the golden era of antibiotic discovery in the 1940s to 1960s. However, diminishing returns from this discovery platform led to its collapse, and efforts to create a new platform based on target-focused screening of large libraries of synthetic compounds failed, in part owing to the lack of penetration of such compounds through the bacterial envelope.

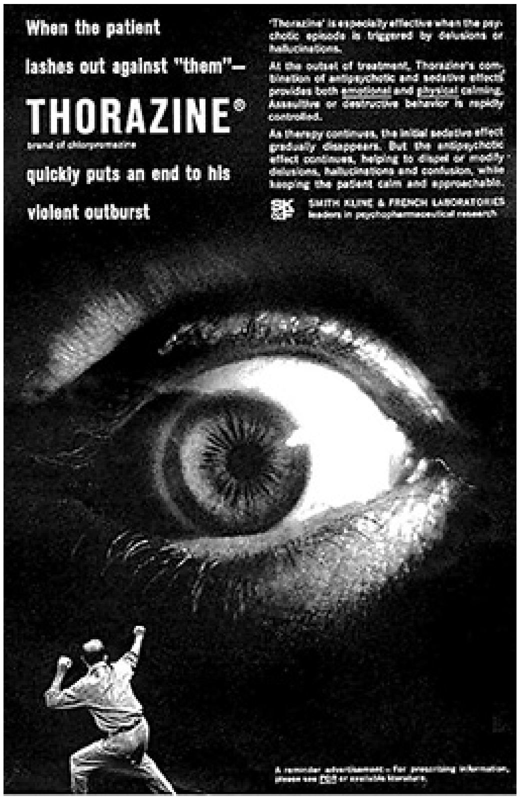
Sometimes stagnant science means your civilization is collapsing. Other times it just means you’ve run out of soil bacteria.

In a way, this points out the unfairness of using antibiotics as a civilizational barometer. This is a drug class first invented in the 1930s and mostly developed by investigating soil bacteria, which have since been mostly exhausted. Of course progress will be faster the closer to the discovery of this technique you get.

If instead you use antidiabetic drugs – a comparatively new field – this is an age of miracles and wonders, with new classes coming out faster than anyone except specialists can keep up with. If antidiabetic drugs were used as a civilizational barometer, we would be having the Singularity next week.

Weird Psychiatric Ads Of The Seventies

There's a really famous ad for thorazine, a drug that came out in the fifties and was the first effective antipsychotic.



When the patient lashes out against "them"—

THORAZINE®

brand of chlorpromazine

quickly puts an end to his violent outburst

'Thorazine' is especially effective when the psychotic episode is triggered by delusions or hallucinations.

At the outset of treatment, Thorazine's combination of antipsychotic and sedative effects provides both emotional and physical calming. Assaultive or destructive behavior is rapidly controlled.

As therapy continues, the initial sedative effect gradually disappears. But the antipsychotic effect continues, helping to dispel or modify delusions, hallucinations and confusion, while keeping the patient calm and approachable.

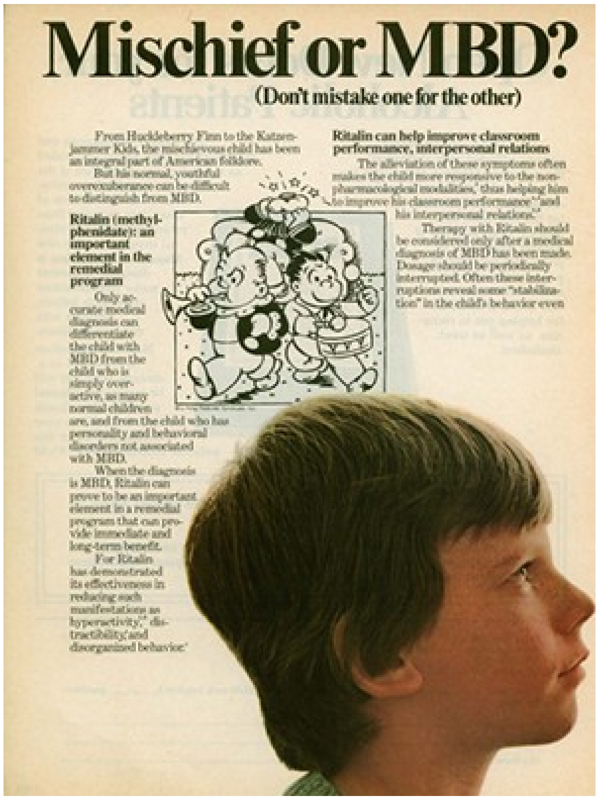
SK&F SMITH KLINE & FRENCH LABORATORIES
leaders in psychopharmaceutical research

A reminder advertisement—for prescribing information, please see PDR or available literature.

Less well-known are all the *other* weird, wonderful and creepy psychiatric ads from the past century.

I was able to find a journal that had an archive of all its ads from the 1970s and a tiny slice of the 60s (no luck getting before then) and thought I'd share some of my favorites and what I learned.

* *



Mischief or MBD?

(Don't mistake one for the other)

From Huckleberry Finn to the Katzenjammer Kids, the mischievous child has been an integral part of American folklore.

But his normal, youthful overexuberance can be difficult to distinguish from MBD.

Ritalin (methylphenidate): an important element in the remedial program

Only accurate medical diagnosis can differentiate the child with MBD from the child who is simply overactive, as many normal children are, and from the child who has personality and behavioral disorders not associated with MBD.

When the diagnosis is MBD, Ritalin can prove to be an important element in a remedial program that can provide immediate and long-term benefit.

For Ritalin has demonstrated its effectiveness in reducing such manifestations as hyperactivity, distractibility, and disorganized behavior.

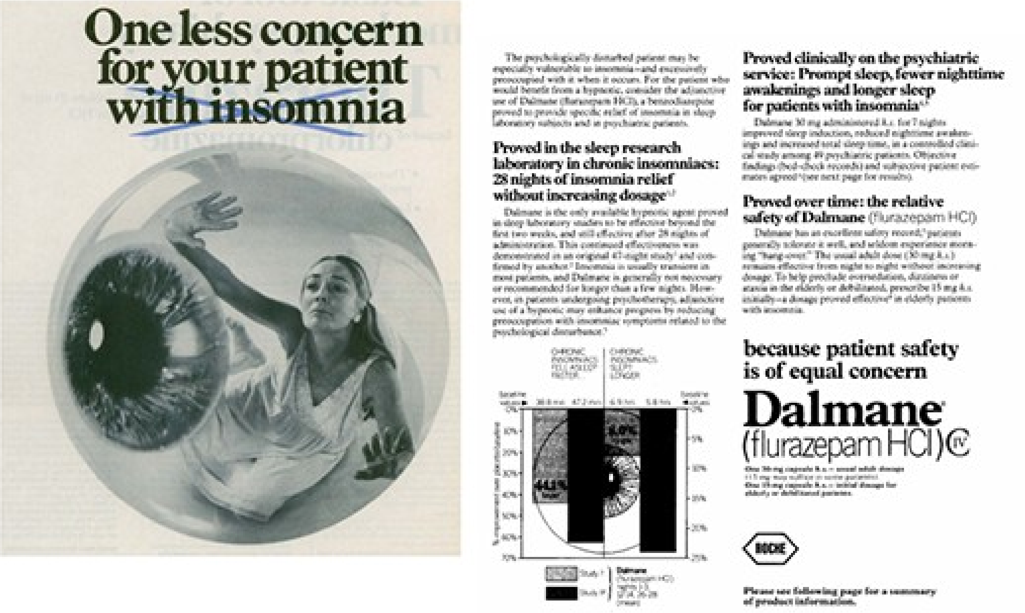
Ritalin can help improve classroom performance, interpersonal relations

The alleviation of these symptoms often makes the child more responsive to the nonpharmacological modalities, thus helping him to improve his classroom performance and his interpersonal relations.

Therapy with Ritalin should be considered only after a medical diagnosis of MBD has been made. Dosage should be periodically interrupted. Often these interruptions reveal some "stabilization" in the child's behavior even

ADHD, formerly ADD, was even more formerly MBD for Minimal Brain Dysfunction. This ad from the late ’70s shows there’s already a gray line between ADHD and normal mischievousness, and Ritalin is already the preferred treatment (and has been since the early sixties).

* *



One less concern for your patient with insomnia

Proved in the sleep research laboratory in chronic insomniacs: 28 nights of insomnia relief without increasing dosage

Proved clinically on the psychiatric service: Prompt sleep, fewer nighttime awakenings and longer sleep for patients with insomnia

Proved over time: the relative safety of Dalmane (flurazepam HCl)

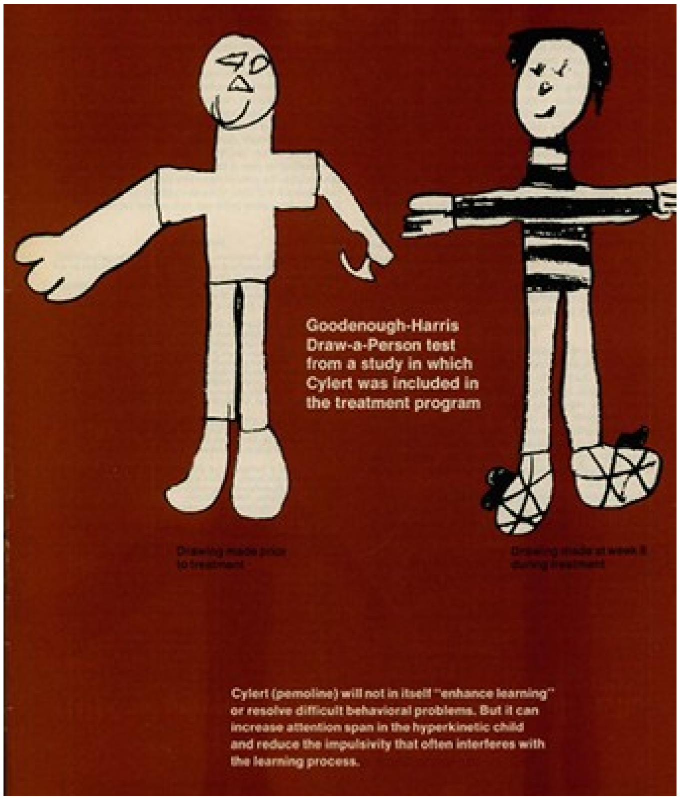
because patient safety is of equal concern

Dalmane® (flurazepam HCl)Ⓒ

Please see following page for a summary of product information.

Following in thorazine’s footsteps of including giant scary eyes in psychiatric ads. This is going to be a recurring theme.

* *



Goodenough-Harris Draw-a-Person test from a study in which Cylert was included in the treatment program

Drawing made prior to treatment
Drawing made at week 8 during treatment

Cylert (pemoline) will not in itself “enhance learning” or resolve difficult behavioral problems. But it can increase attention span in the hyperkinetic child and reduce the impulsivity that often interferes with the learning process.

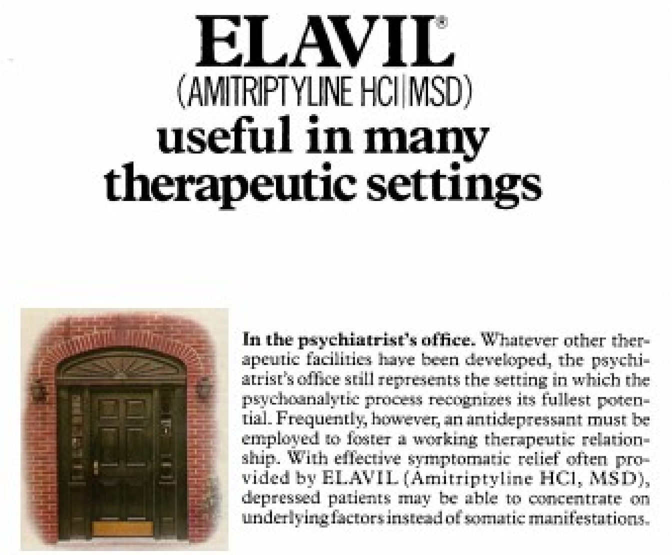
More minimal brain dysfunction. “Cylert (pemoline) will not in itself “enhance learning” or resolve difficult behavioral problems. But it can increase attention span in the hyperkinetic child and reduce the impulsivity that often interferes with the learning process”.

The Goodenough-Harris Draw A Person Test has since been found (contra nominative determinism) to correlate only very weakly with real IQ tests for preschoolers. It has thus fallen out of favor, which is too bad as it led to some very cute scientific papers

Anyway, I guess we’re supposed to be excited that Cylert can make kids sit still enough to add stripes to a guy’s shirt. I’m going to hold out until they can make them not have weird nets for shoes.

Side effects of Cylert® may include holding your arms rigidly straight out to the sides all the time like you’re being crucified or something.

* *



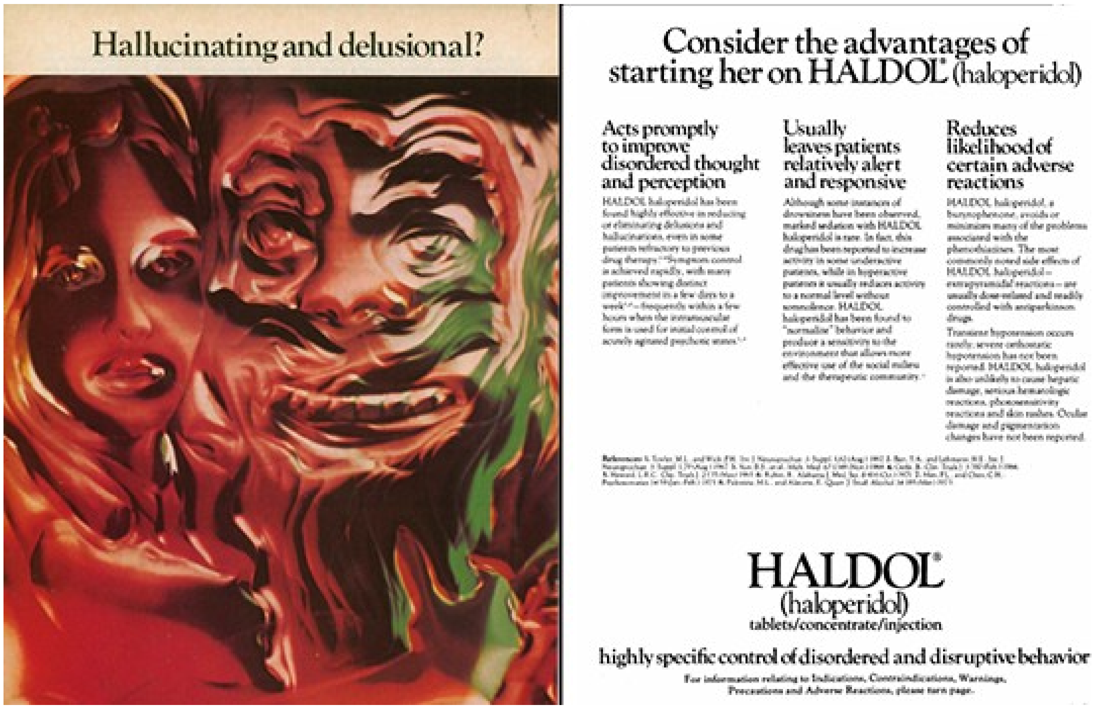
ELAVIL®
(AMITRIPTYLINE HCl|MSD)
useful in many therapeutic settings

In the psychiatrist’s office. Whatever other therapeutic facilities have been developed, the psychiatrist’s office still represents the setting in which the psychoanalytic process recognizes its fullest potential. Frequently, however, an antidepressant must be employed to foster a working therapeutic relationship. With effective symptomatic relief often provided by ELAVIL (Amitriptyline HCl, MSD), depressed patients may be able to concentrate on underlying factors instead of somatic manifestations.

I was talking with Chris H and a few other people a couple of posts back about how pharmacotherapy used to be viewed (at least officially) as an adjunct for talk therapy. This ad is a good demonstration: “Whatever other therapeutic facilities have been developed, the psychiatrist’s office still represents the setting in which the psychoanalytic process recognizes its fullest potential. Frequently, however, an antidepressant must be employed to foster a working therapeutic relationship. With effective symptomatic relief often provided by ELAVIL, depressed patients may be able to concentrate on underlying factors instead of somatic manifestations.”

I wonder to what degree this was something you *had* to say to be viewed as a responsible psychiatrist back then: “Oh yeah, obviously it’s the Freudian psychoanalysis that’s really important, but maybe some of these drugs can, well, sort of help a little so we can get to Freudian psychoanalysis faster.” And to what degree everyone was in on the charade but didn’t want to torpedo their reputations by departing from it.

* *



Hallucinating and delusional?

Consider the advantages of starting her on HALDOL® (haloperidol)

Acts promptly to improve disordered thought and perception

HALDOL haloperidol has been found highly effective in reducing or eliminating delusions and hallucinations, even in some patients refractory to previous drug therapy. Symptom control is achieved rapidly, with many patients showing distinct improvement in a few days to a week—frequently within a few hours when the intramuscular form is used for initial control of acutely agitated psychotic states.

Usually leaves patients relatively alert and responsive

Although some instances of drowsiness have been observed, marked sedation with HALDOL haloperidol is rare. In fact, this drug has been reported to increase activity in some underactive patients, while in hyperactive patients it usually reduces activity to a normal level without somnolence. HALDOL haloperidol has been found to “normalize” behavior and produce a sensitivity to the environment that allows more effective use of the social milieu and the therapeutic community.

Reduces likelihood of certain adverse reactions

HALDOL haloperidol, a butyrophenone, avoids or minimizes many of the problems associated with the phenothiazines. The most commonly noted side effects of HALDOL haloperidol—extrapyramidal reactions—are usually dose-related and readily controlled with antiparkinson drugs.

Transient hypotension occurs rarely; severe orthostatic hypotension has not been reported. HALDOL haloperidol is also unlikely to cause hepatic damage, serious hematologic reactions, photosensitivity reactions and skin rashes. Ocular damage and pigmentation changes have not been reported.

HALDOL®
(haloperidol)
tablets/concentrate/injection

highly specific control of disordered and disruptive behavior

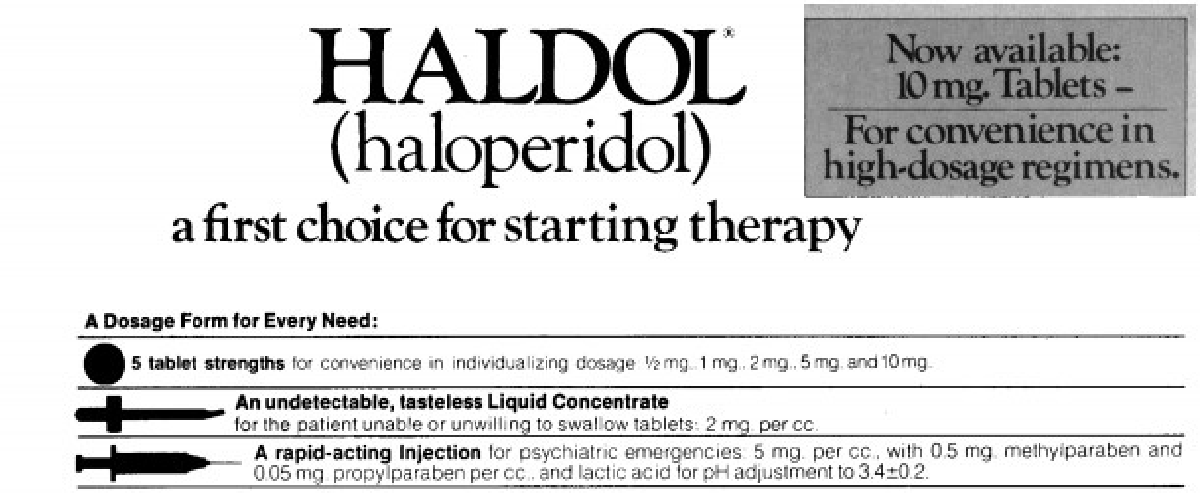
For information relating to Indications, Contraindications, Warnings, Precautions and Adverse Reactions, please turn page.

Another ad following in Thorazine’s footsteps of “make antipsychotic ads as creepy and psychotic-looking as possible.”

A couple days ago I asked my boss what the pharmacological differences between Haldol and [several similar drugs] were. He said there were no important differences at all. I asked him why, if that were so, everyone uses Haldol and almost no one uses any of the others. He said it was because Haldol had a better advertising campaign back in the day – which is what led me to look at old psychiatric ads in the first place.

So if any of you are in the public relations field, remember: melting faces *sells*.

* *



HALDOL®
(haloperidol)
a first choice for starting therapy

Now available:
10 mg. Tablets –
For convenience in high-dosage regimens.

A Dosage Form for Every Need:

5 tablet strengths for convenience in individualizing dosage: ½ mg., 1 mg., 2 mg., 5 mg. and 10 mg.

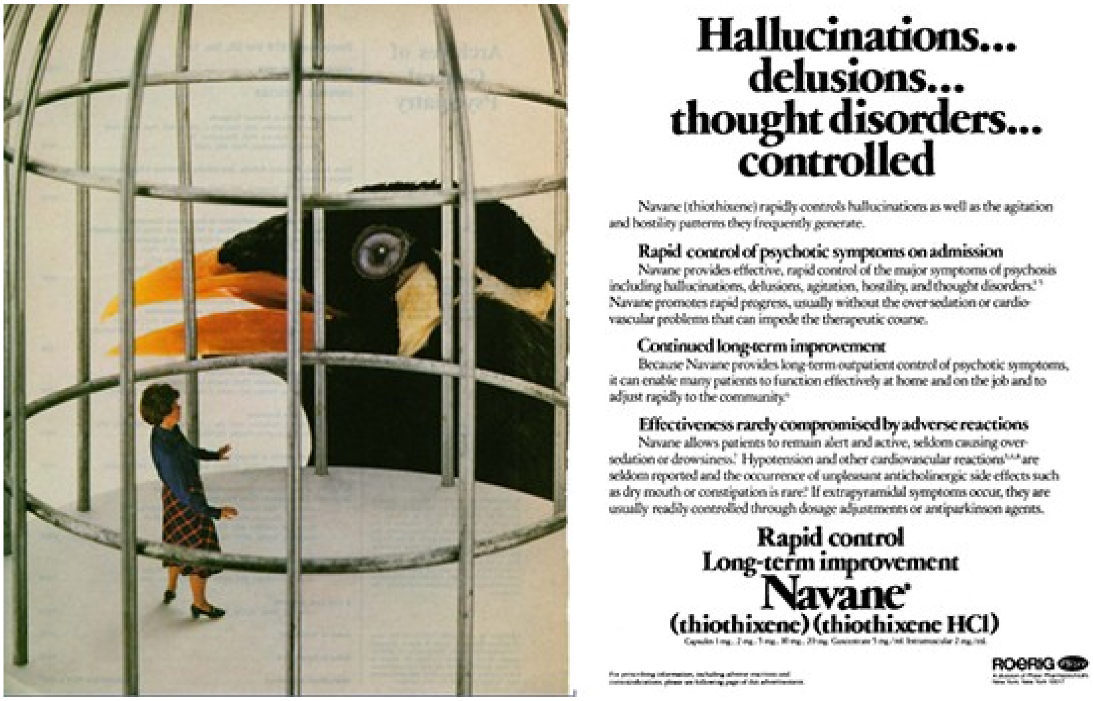
An undetectable, tasteless Liquid Concentrate for the patient unable or unwilling to swallow tablets: 2 mg. per cc.

A rapid-acting Injection for psychiatric emergencies: 5 mg. per cc., with 0.5 mg. methylparaben and 0.05 mg. propylparaben per cc., and lactic acid for pH adjustment to 3.4±0.2.

We’re not saying you should slip very powerful drugs into in your patient’s drink without their knowledge. We’re just saying if you do, do it with Haldol®!

My impression is that this used to be a lot more common, but still goes on in certain situations, especially with the demented elderly.

* *



Hallucinations...
delusions...
thought disorders...
controlled

Navane (thiothixene) rapidly controls hallucinations as well as the agitation and hostility patterns they frequently generate.

Rapid control of psychotic symptoms on admission

Navane provides effective, rapid control of the major symptoms of psychosis including hallucinations, delusions, agitation, hostility, and thought disorders.[1-3] Navane promotes rapid progress, usually without the over-sedation or cardiovascular problems that can impede the therapeutic course.

Continued long-term improvement

Because Navane provides long-term outpatient control of psychotic symptoms, it can enable many patients to function effectively at home and on the job and to adjust rapidly to the community.[4]

Effectiveness rarely compromised by adverse reactions

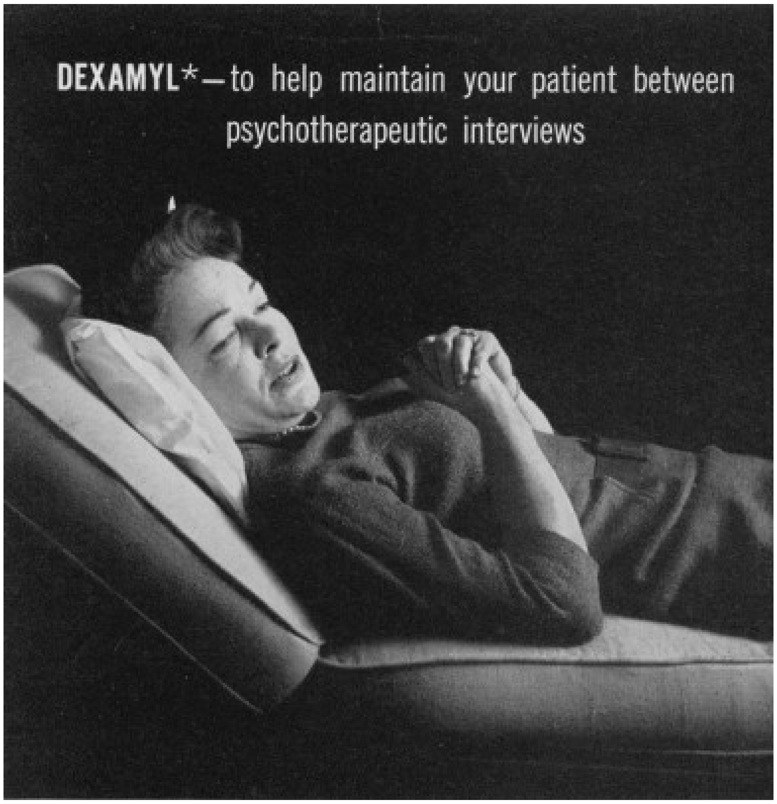
Navane allows patients to remain alert and active, seldom causing over-sedation or drowsiness.[1] Hypotension and other cardiovascular reactions[3-5] are seldom reported and the occurrence of unpleasant anticholinergic side effects such as dry mouth or constipation is rare.[2] If extrapyramidal symptoms occur, they are usually readily controlled through dosage adjustments or antiparkinson agents.

Rapid control
Long-term improvement
Navane®
(thiothixene) (thiothixene HCl)

ROERIG

In Soviet Russia, bird cages *you*! But if bird cages you, and you *not* in Soviet Russia, is extremely worrying sign. Should seek medical help immediately.

* *



DEXAMYL*—to help maintain your patient between psychotherapeutic interviews

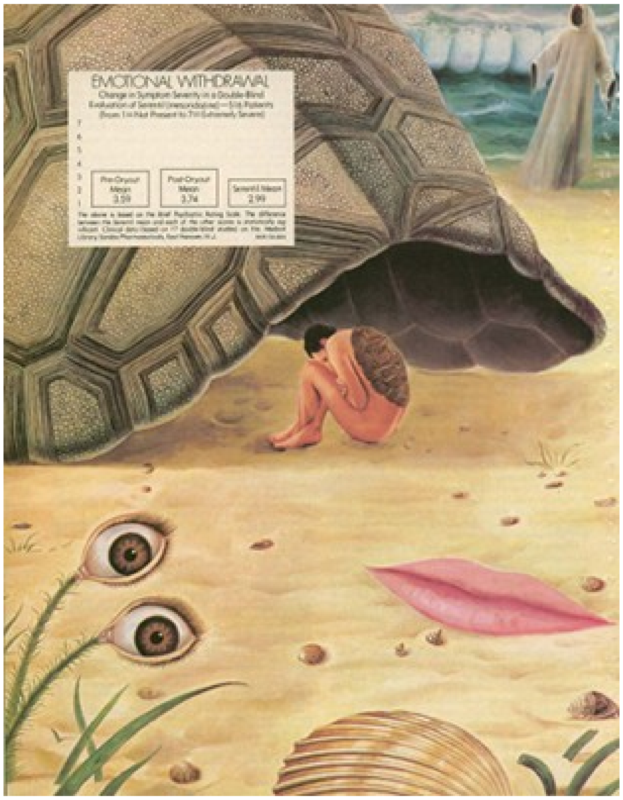
Another “we’re only using drugs for *between* the psychotherapeutic interviews” ad.

Dexamyl is a combination of amphetamine and a barbiturate. Apparently at one time, giving people a really strong addictive upper and a really strong addictive downer together was considered such a good idea that it was advertised in psychiatric journals – and commonly used to perk up tired housewives.

My instinct would be that the upper and downer would cancel out, leaving people about how they were before except with a host of terrible side effects. But when I Google it I get a lot of people who said the barbiturate cancelled out the side effects of the amphetamine and they felt great on Dexamyl and it is their greatest regret in life that it is no longer available. So maybe my instincts are wrong and we should all be taking amphetamines mixed with barbiturates all the time.

According to Wikipedia, UK PM Anthony Eden was on Dexamyl when he screwed up the Suez Crisis, which doesn’t surprise me at all.

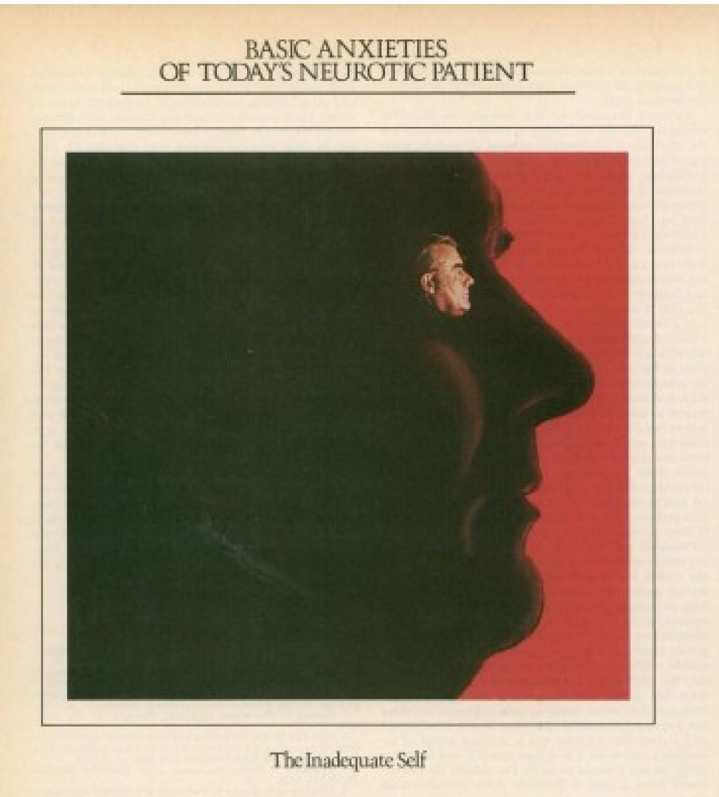
* *



People with tortoise shells inside bigger tortoise shells. Eyes growing on thorny stalks of grass. Lips bursting forth from the earth. Some kind of weird spectral Death hanging out in the background. Sure, the name of the drug involved is so small I can’t read it, but making it any bigger would have ruined *the artistic vision*.

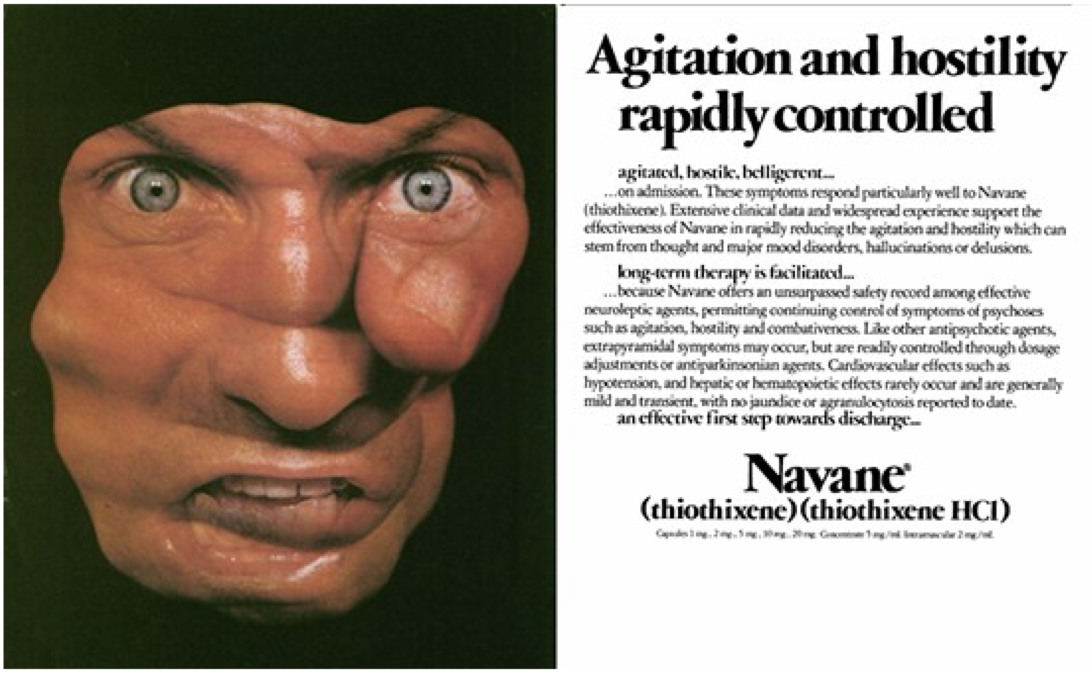
Also, you really need to stop with all the eyes in your antipsychotic ads.

* *



NO I DIDN'T MEAN IT LIKE THAT! EYES WHERE THERE SHOULD BE EYES! NO EYES WHERE THERE SHOULDN'T BE EYES! OKAY?

* *



Agitation and hostility rapidly controlled

agitated, hostile, belligerent...

...on admission. These symptoms respond particularly well to Navane (thiothixene). Extensive clinical data and widespread experience support the effectiveness of Navane in rapidly reducing the agitation and hostility which can stem from thought and major mood disorders, hallucinations or delusions.

long-term therapy is facilitated...

...because Navane offers an unsurpassed safety record among effective neuroleptic agents, permitting continuing control of symptoms of psychoses such as agitation, hostility and combativeness. Like other antipsychotic agents, extrapyramidal symptoms may occur, but are readily controlled through dosage adjustments or antiparkinsonian agents. Cardiovascular effects such as hypotension, and hepatic or hematopoietic effects rarely occur and are generally mild and transient, with no jaundice or agranulocytosis reported to date.

an effective first step towards discharge...

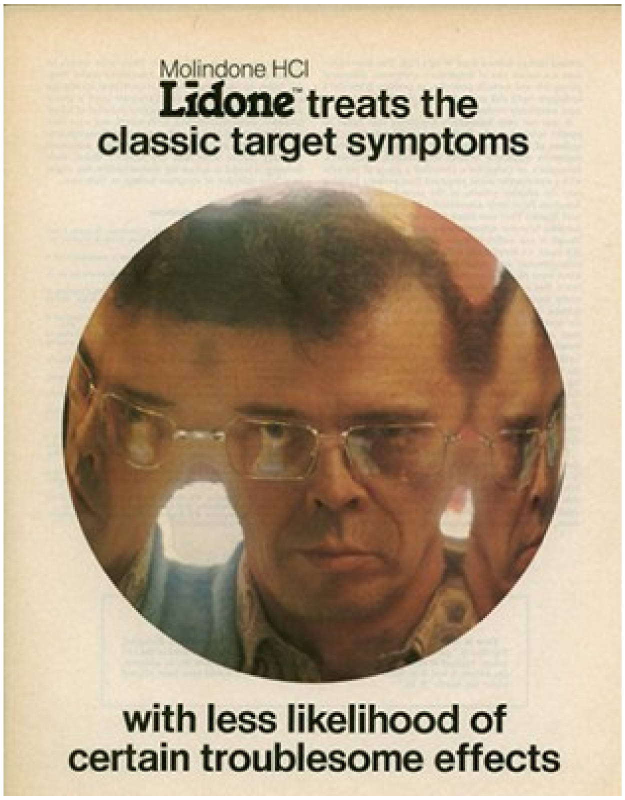
Navane®

(thiothixene)(thiothixene HCl)

Capsules 1 mg., 2 mg., 5 mg., 10 mg., 20 mg. Concentrate 5 mg/ml. Intramuscular 2 mg/ml.

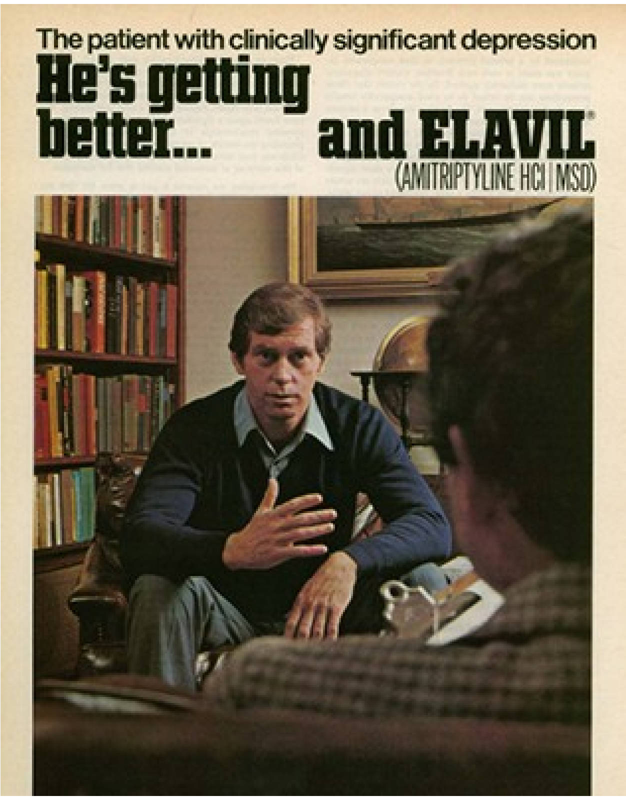
Ah, screw it, close enough.

* *


Molindone HCl
Lidone™ treats the classic target symptoms
with less likelihood of certain troublesome effects

Whatever.

* *


The patient with clinically significant depression
He's getting better...
and ELAVIL®
(AMITRIPTYLINE HCl | MSD)

Is...is that a syllepsis? Did you just include a syllepsis in a psychiatric ad? Cooooooool.

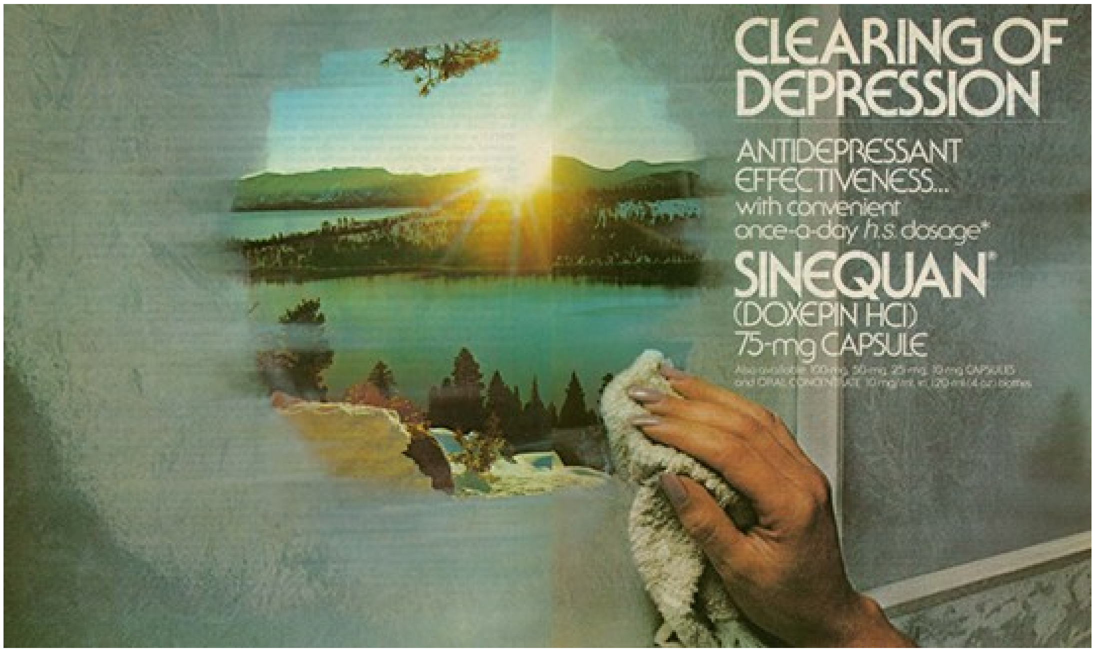
* *

When your patient is clinically depressed...
"...everything I saw was negative."
...consider this potent blocker of norepinephrine re-uptake
(Electron micrograph of human synaptic cleft.)

The biogenic amine hypothesis states that depression may be associated with a deficiency of specific neurotransmitters—norepinephrine or serotonin—at brain receptor sites. Tricyclics appear to block the re-uptake of specific neurotransmitters. Some tricyclics specifically block the re-uptake of serotonin; others block the re-uptake of norepinephrine. Therefore, if your patient is unresponsive to treatment with one tricyclic, consider switching to a tricyclic which blocks the re-uptake of a different neurotransmitter.

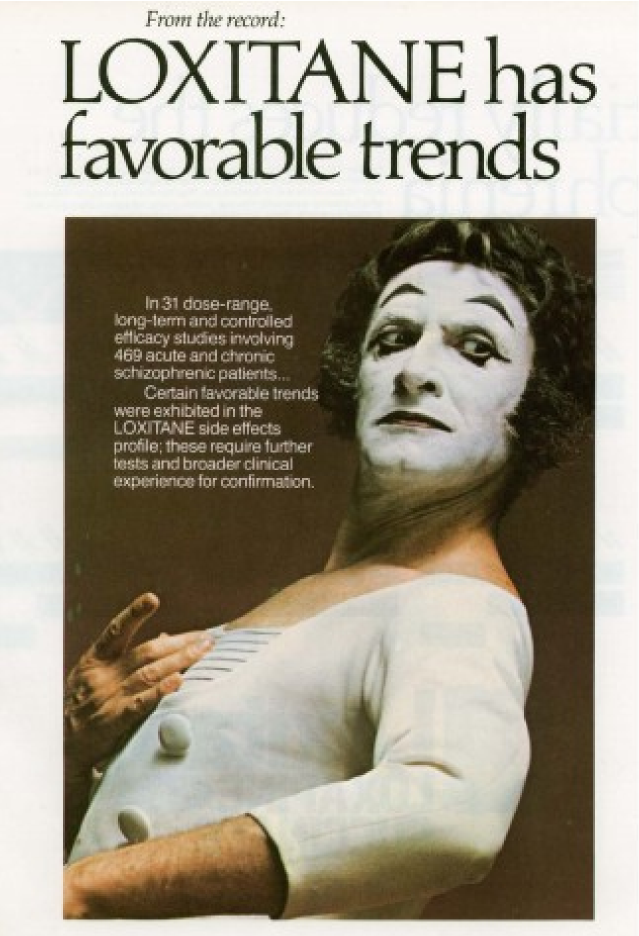
Photography puns age about as well as…well, as the biogenic amine hypothesis of depression.

* *


CLEARING OF DEPRESSION
ANTIDEPRESSANT EFFECTIVENESS…
with convenient once-a-day *h.s.* dosage*
SINEQUAN®
(DOXEPIN HCl)
75-mg CAPSULE

This is a nice ad. It makes me want to take Sinequan. Why can’t the Navane ads be more like this one?

* *



From the record:

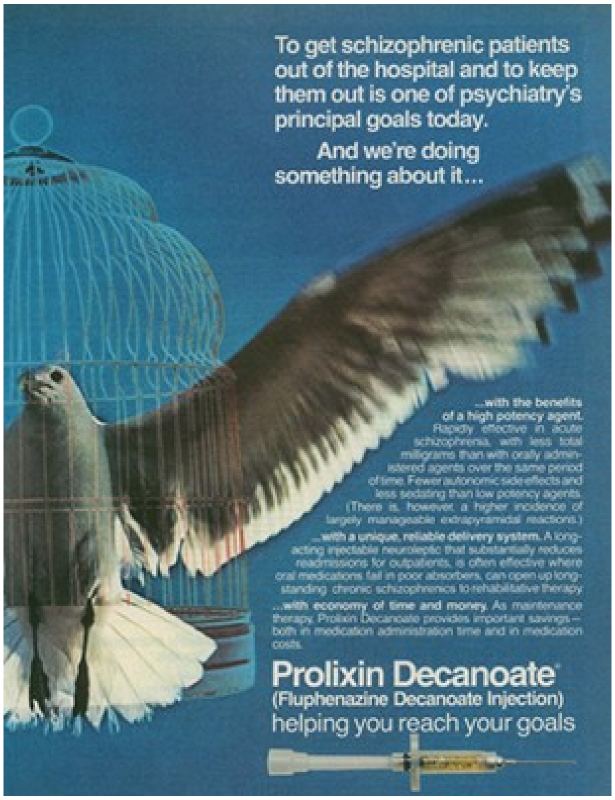
LOXITANE has favorable trends

In 31 dose-range, long-term and controlled efficacy studies involving 469 acute and chronic schizophrenic patients...

Certain favorable trends were exhibited in the LOXITANE side effects profile; these require further tests and broader clinical experience for confirmation.

Side effects of Loxitane® may include infuriating vagueness.

* *



To get schizophrenic patients out of the hospital and to keep them out is one of psychiatry's principal goals today.

And we're doing something about it...

...with the benefits of a high potency agent. Rapidly effective in acute schizophrenia, with less total milligrams than with orally administered agents over the same period of time. Fewer autonomic side effects and less sedating than low potency agents. (There is, however, a higher incidence of largely manageable extrapyramidal reactions.)

...with a unique, reliable delivery system. A long-acting injectable neuroleptic that substantially reduces readmissions for outpatients, is often effective where oral medications fail in poor absorbers, can open up long-standing chronic schizophrenics to rehabilitative therapy

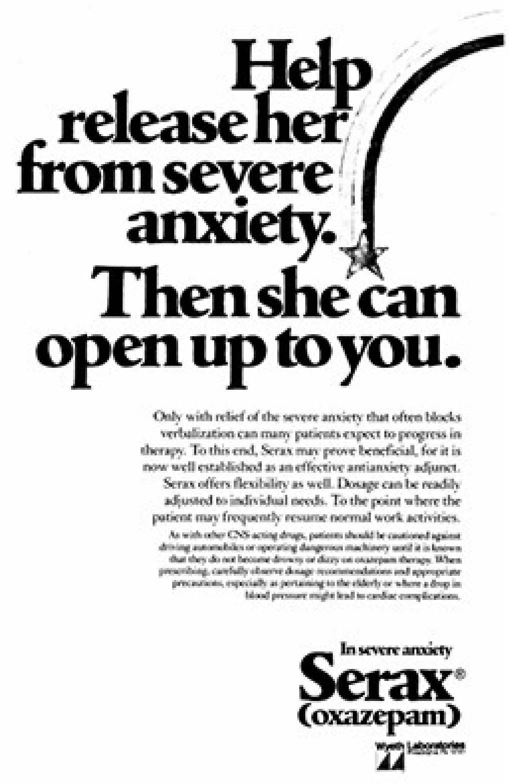
...with economy of time and money. As maintenance therapy, Prolixin Decanoate provides important savings—both in medication administration time and in medication costs.

Prolixin Decanoate®
(Fluphenazine Decanoate Injection)
helping you reach your goals

NO YOU FOOL DON’T LET THE BIRD OUT OF THE CAGE NOW IT’S GOING TO PUT YOU IN THE CAGE AND YOU WILL NEED NAVANE®.

Prolixin is one of the drugs that is very similar to Haldol but never caught on because of poor advertising. The moral of the story is – doves are out, melting faces are in.

* *



Help release her from severe anxiety. Then she can open up to you.

Only with relief of the severe anxiety that often blocks verbalization can many patients expect to progress in therapy. To this end, Serax may prove beneficial, for it is now well established as an effective antianxiety adjunct. Serax offers flexibility as well. Dosage can be readily adjusted to individual needs. To the point where the patient may frequently resume normal work activities.

As with other CNS acting drugs, patients should be cautioned against driving automobiles or operating dangerous machinery until it is known that they do not become drowsy or dizzy on oxazepam therapy. When prescribing, carefully observe dosage recommendations and appropriate precautions, especially as pertaining to the elderly or where a drop in blood pressure might lead to cardiac complications.

In severe anxiety
Serax®
(oxazepam)

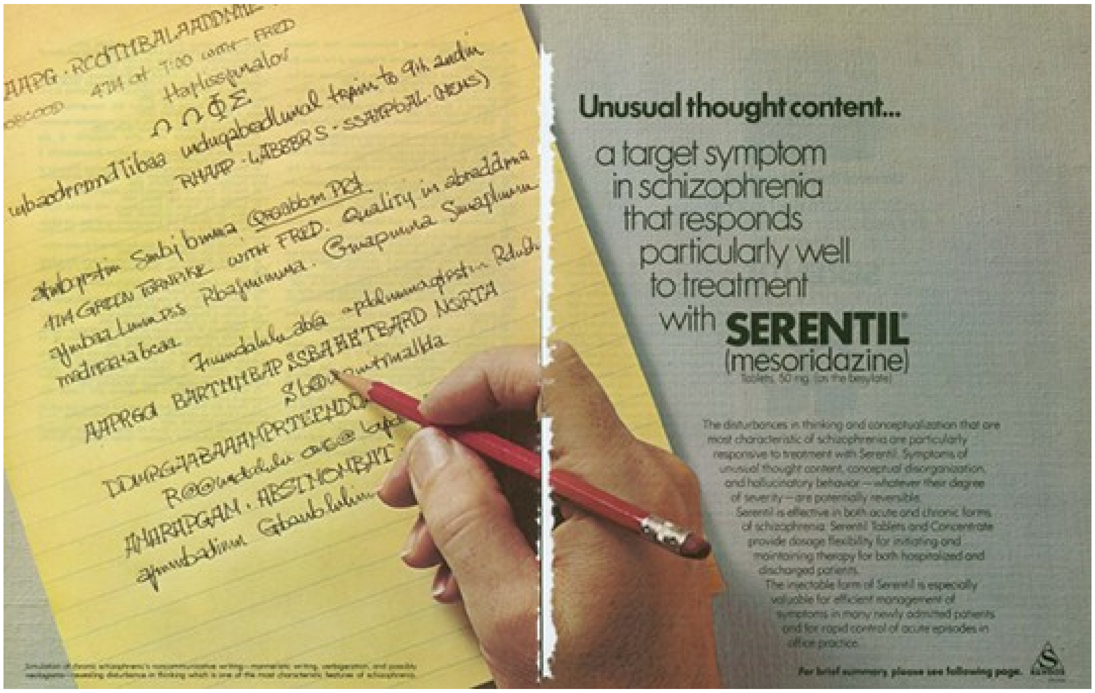
Release her from severe anxiety. Then she can open up to you. You ask her how she’s doing. She smiles bashfully, places a hand on your knee. Should you? Shouldn’t you? You clasp her hand. Everything’s going to be all right, you tell her.

‘This may be a little forward’, she asks, ‘but would you ever date a patient? You know, if the right one came along?’ ‘I’m married’, you tell her. ‘Oh!’ she says, horrified, and her mouth forms this adorable little O shape ‘I didn’t mean –‘. You cut her off. ‘But my wife isn’t here’ you say, and lean in, kissing her on the lips. She leans into your mouth passionately. You grab a breast. Her hand reaches for your crotch.

‘We shouldn’t,’ she says, suddenly. ‘We should,’ you say. ‘Run away with me, and we’ll leave your severe anxiety far behind’. ‘Where would we go?’ she asks. ‘I don’t know,’ you say. ‘France? Venice? Anywhere but here.’ She kisses you again. ‘Anywhere,’ she repeats, ‘just as long as I can bring my Serax.’

Side effects of Serax® (oxazepam) may include marital strain, divorce, unintended pregnancy, and gonorrhea.

* *



Unusual thought content...

a target symptom in schizophrenia that responds particularly well to treatment with **SERENTIL®** (mesoridazine)

Tablets, 50 mg. (as the besylate)

The disturbances in thinking and conceptualization that are most characteristic of schizophrenia are particularly responsive to treatment with Serentil. Symptoms of unusual thought content, conceptual disorganization, and hallucinatory behavior—whatever their degree of severity—are potentially reversible.

Serentil is effective in both acute and chronic forms of schizophrenia. Serentil Tablets and Concentrate provide dosage flexibility for initiating and maintaining therapy for both hospitalized and discharged patients.

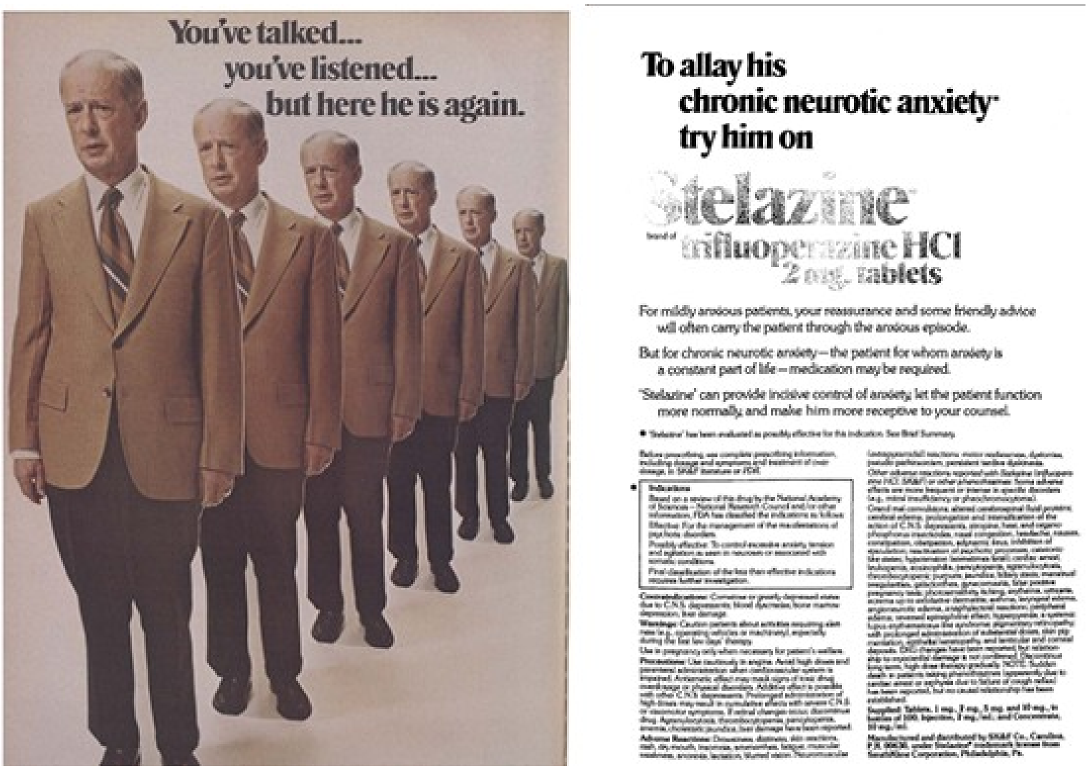
The injectable form of Serentil is especially valuable for efficient management of symptoms in many newly admitted patients and for rapid control of acute episodes in office practice.

For brief summary, please see following page.

Is…is that guy writing *Finnegan’s Wake*?

Serentil was withdrawn a couple of years ago after it was found to cause dangerous cardiac side effects.

* *



You've talked...
you've listened...
but here he is again.

To allay his
chronic neurotic anxiety*
try him on

Stelazine
brand of trifluoperazine HCl
2 mg. tablets

For mildly anxious patients, your reassurance and some friendly advice will often carry the patient through the anxious episode.

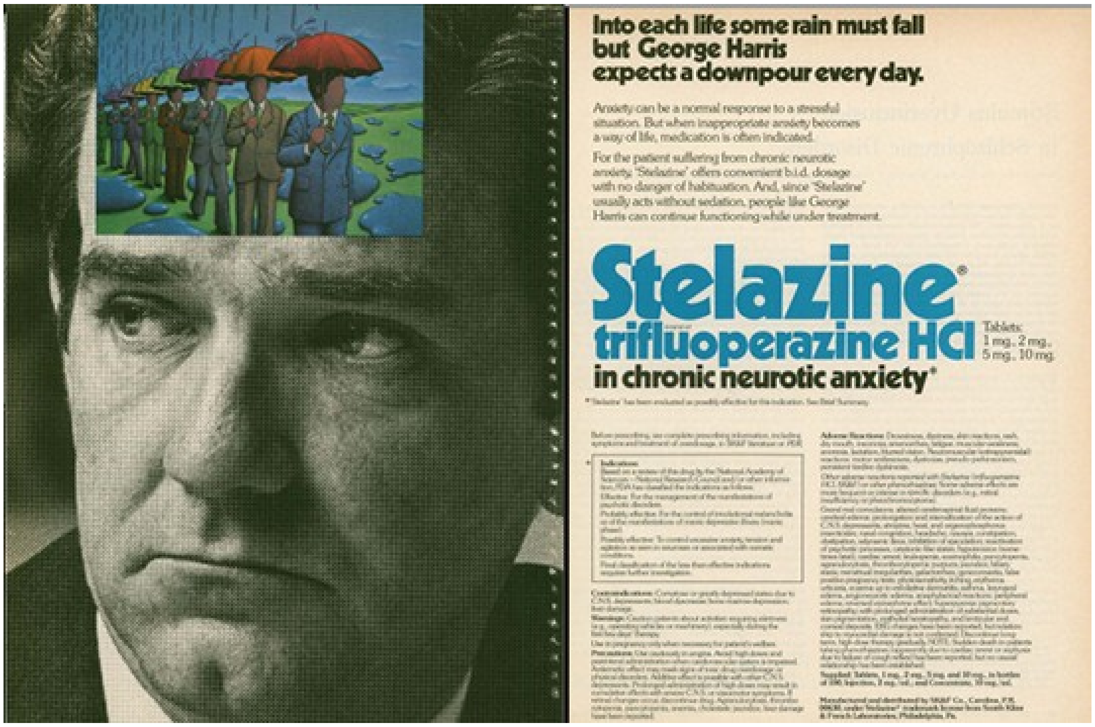
But for chronic neurotic anxiety—the patient for whom anxiety is a constant part of life—medication may be required.

'Stelazine' can provide incisive control of anxiety, let the patient function more normally, and make him more receptive to your counsel.

I have no idea who that guy is, but screw him.

I wonder if I can trace some kind of evolution here from "drugs will get your patients ready for psychotherapy" to "drugs will help your patients who are refractory to psychotherapy".

* *



Into each life some rain must fall but George Harris expects a downpour every day.

Anxiety can be a normal response to a stressful situation. But when inappropriate anxiety becomes a way of life, medication is often indicated.

For the patient suffering from chronic neurotic anxiety, 'Stelazine' offers convenient b.i.d. dosage with no danger of habituation. And, since 'Stelazine' usually acts without sedation, people like George Harris can continue functioning while under treatment.

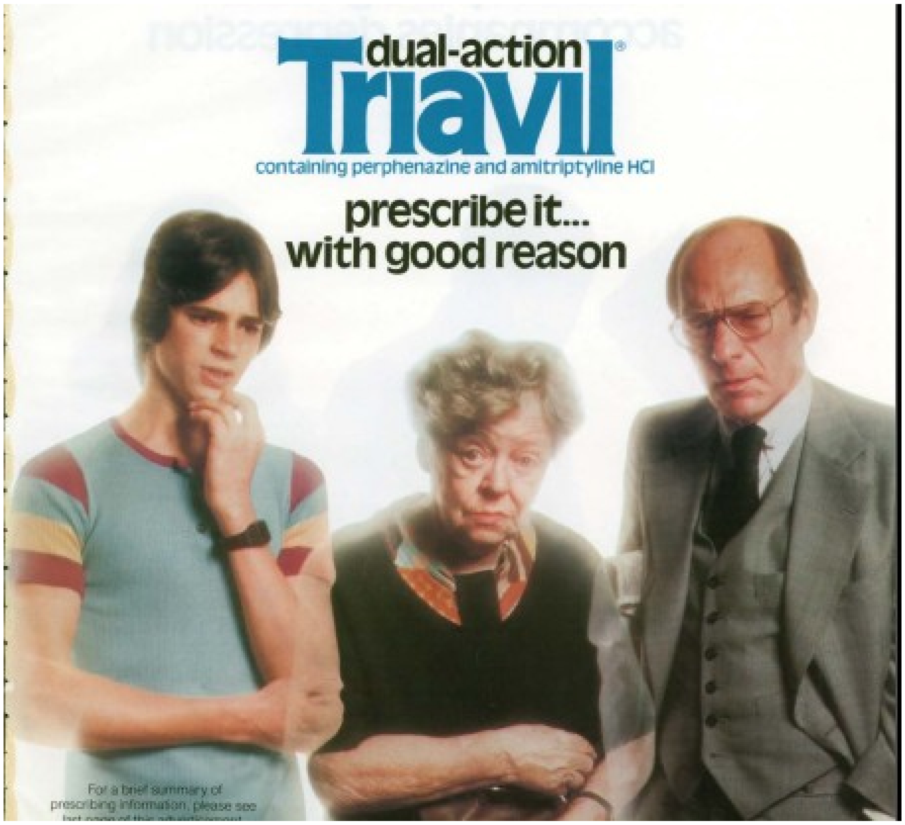
Stelazine®
trifluoperazine HCl
in chronic neurotic anxiety*

Tablets:
1 mg., 2 mg.,
5 mg., 10 mg.

This was how we had to represent people’s thoughts before we had Photoshop’s “blur edges” filter. Just a big square stuck in the middle of their head.

For some reason I can’t imagine any modern ad using the name “George Harris”. I don’t know if it’s just that they wouldn’t use any name, or that they wouldn’t use one that aggressively normal-sounding.

* *



dual-action Triavil®

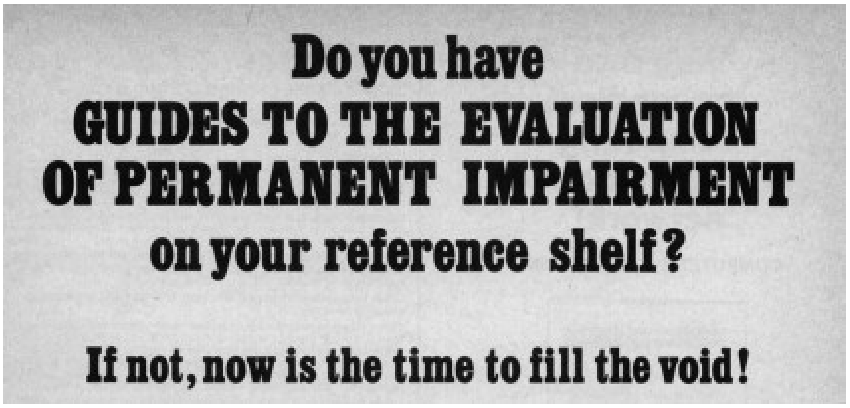
containing perphenazine and amitriptyline HCl

prescribe it... with good reason

For a brief summary of
prescribing information, please see

This ad seems to be going for “mysteriously creepy but hard to put your finger on why”. But that “…with good reason” definitely doesn’t help.

* *

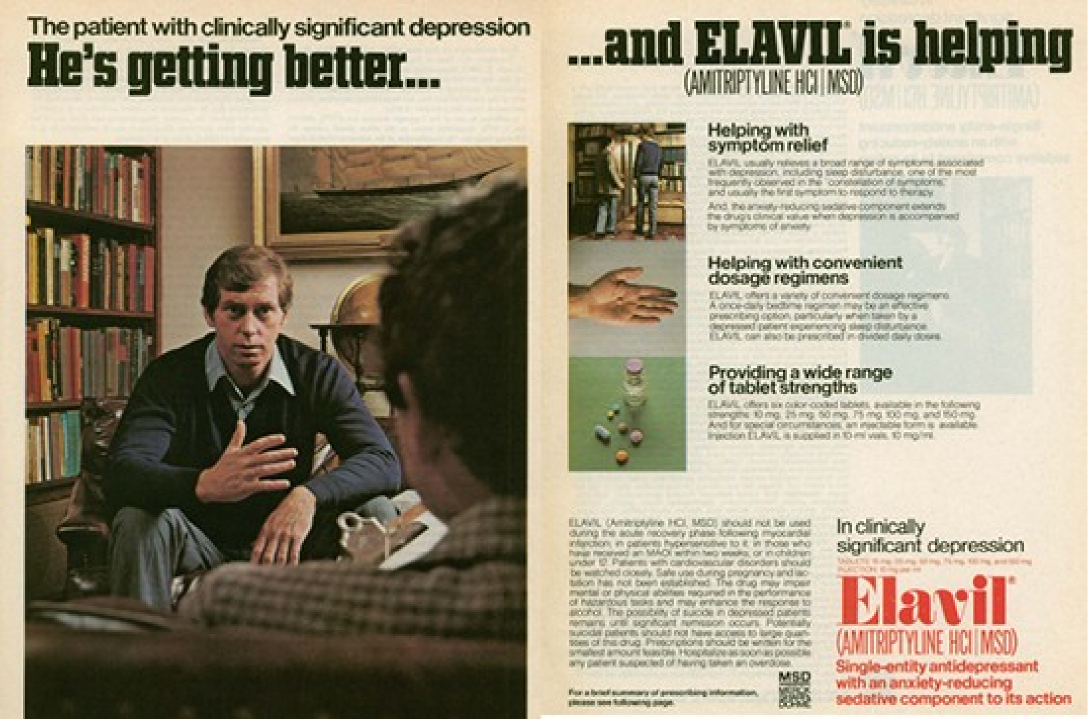


Do you have
GUIDES TO THE EVALUATION
OF PERMANENT IMPAIRMENT
on your reference shelf?

If not, now is the time to fill the void!

I knew something was missing from my life!

* *



The patient with clinically significant depression

He's getting better...

...and ELAVIL® is helping

(AMITRIPTYLINE HCl | MSD)

Helping with symptom relief

Helping with convenient dosage regimens

Providing a wide range of tablet strengths

In clinically significant depression

Elavil®

(AMITRIPTYLINE HCl | MSD)

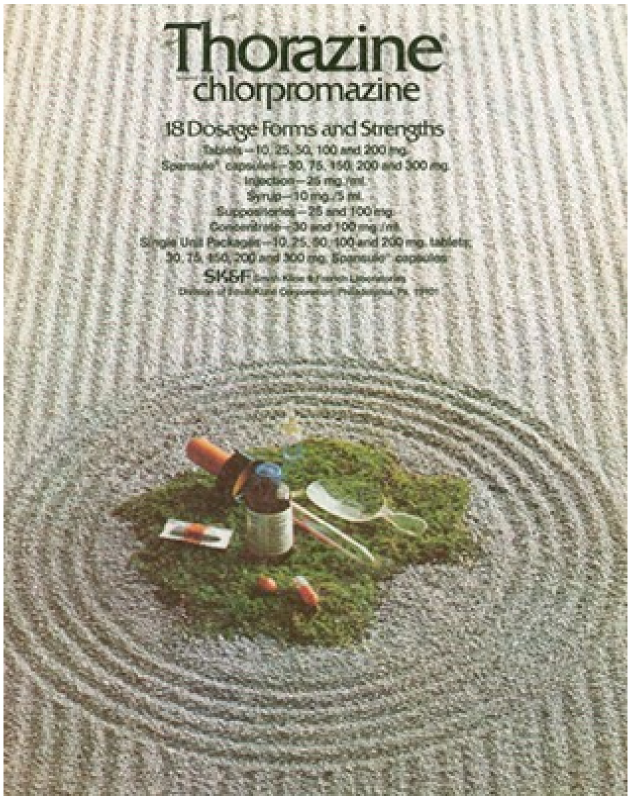
Single-entity antidepressant with an anxiety-reducing sedative component to its action

For a brief summary of prescribing information, please see following page.

MSD

They rewrote it to get rid of the syllepsis!   *Why would you do that?*

* *



Thorazine®

chlorpromazine

18 Dosage Forms and Strengths

Tablets—10, 25, 50, 100 and 200 mg.
Spansule® capsules—30, 75, 150, 200 and 300 mg.
Injection—25 mg./ml.
Syrup—10 mg./5 ml.
Suppositories—25 and 100 mg.
Concentrate—30 and 100 mg./ml.
Single Unit Packages—10, 25, 50, 100 and 200 mg. tablets; 30, 75, 150, 200 and 300 mg. Spansule® capsules

SK&F

Calm.

Meditations On Moloch

I.

Allen Ginsberg’s famous poem on Moloch:

> What sphinx of cement and aluminum bashed open their skulls and ate up their brains and imagination?
>
> Moloch! Solitude! Filth! Ugliness! Ashcans and unobtainable dollars! Children screaming under the stairways! Boys sobbing in armies! Old men weeping in the parks!
>
> Moloch! Moloch! Nightmare of Moloch! Moloch the loveless! Mental Moloch! Moloch the heavy judger of men!
>
> Moloch the incomprehensible prison! Moloch the crossbone soulless jailhouse and Congress of sorrows! Moloch whose buildings are judgment! Moloch the vast stone of war! Moloch the stunned governments!
>
> Moloch whose mind is pure machinery! Moloch whose blood is running money! Moloch whose fingers are ten armies! Moloch whose breast is a cannibal dynamo! Moloch whose ear is a smoking tomb!
>
> Moloch whose eyes are a thousand blind windows! Moloch whose skyscrapers stand in the long streets like endless Jehovahs! Moloch whose factories dream and croak in the fog! Moloch whose smoke-stacks and antennae crown the cities!
>
> Moloch whose love is endless oil and stone! Moloch whose soul is electricity and banks! Moloch whose

poverty is the specter of genius! Moloch whose fate is a cloud of sexless hydrogen! Moloch whose name is the Mind!

Moloch in whom I sit lonely! Moloch in whom I dream Angels! Crazy in Moloch! Cocksucker in Moloch! Lacklove and manless in Moloch!

Moloch who entered my soul early! Moloch in whom I am a consciousness without a body! Moloch who frightened me out of my natural ecstasy! Moloch whom I abandon! Wake up in Moloch! Light streaming out of the sky!

Moloch! Moloch! Robot apartments! invisible suburbs! skeleton treasuries! blind capitals! demonic industries! spectral nations! invincible madhouses! granite cocks! monstrous bombs!

They broke their backs lifting Moloch to Heaven! Pavements, trees, radios, tons! lifting the city to Heaven which exists and is everywhere about us!

Visions! omens! hallucinations! miracles! ecstasies! gone down the American river!

Dreams! adorations! illuminations! religions! the whole boatload of sensitive bullshit!

Breakthroughs! over the river! flips and crucifixions! gone down the flood! Highs! Epiphanies! Despairs! Ten years' animal screams and suicides! Minds! New loves! Mad generation! down on the rocks of Time!

Real holy laughter in the river! They saw it all! the wild eyes! the holy yells! They bade farewell! They jumped

> off the roof! to solitude! waving! carrying flowers! Down to the river! into the street!

What's always impressed me about this poem is its conception of civilization as an individual entity. You can almost see him, with his fingers of armies and his skyscraper-window eyes.

A lot of the commentators say Moloch represents capitalism. This is definitely a piece of it, even a big piece. But it doesn't quite fit. Capitalism, whose fate is a cloud of sexless hydrogen? Capitalism in whom I am a consciousness without a body? Capitalism, therefore granite cocks?

Moloch is introduced as the answer to a question – C. S. Lewis' question in Hierarchy Of Philosophers – *what does it*? Earth could be fair, and all men glad and wise. Instead we have prisons, smokestacks, asylums. What sphinx of cement and aluminum breaks open their skulls and eats up their imagination?

And Ginsberg answers: *Moloch does it.*

There's a passage in the *Principia Discordia* where Malaclypse complains to the Goddess about the evils of human society. "Everyone is hurting each other, the planet is rampant with injustices, whole societies plunder groups of their own people, mothers imprison sons, children perish while brothers war."

The Goddess answers: "What is the matter with that, if it's what you want to do?"

Malaclypse: "But nobody wants it! Everybody hates it!"

Goddess: "Oh. Well, then stop."

The implicit question is – if everyone hates the current system, who perpetuates it? And Ginsberg answers: “Moloch”. It’s powerful not because it’s correct – nobody literally thinks an ancient Carthaginian demon causes everything – but because thinking of the system as an agent throws into relief the degree to which the system *isn’t* an agent.

Bostrom makes an offhanded reference of the possibility of a dictatorless dystopia, one that every single citizen including the leadership hates but which nevertheless endures unconquered. It’s easy enough to imagine such a state. Imagine a country with two rules: first, every person must spend eight hours a day giving themselves strong electric shocks. Second, if anyone fails to follow a rule (including this one), or speaks out against it, or fails to enforce it, all citizens must unite to kill that person. Suppose these rules were well-enough established by tradition that everyone expected them to be enforced.

So you shock yourself for eight hours a day, because you know if you don’t everyone else will kill you, because if they don’t, everyone else will kill *them*, and so on. Every single citizen hates the system, but for lack of a good coordination mechanism it endures. From a god’s-eye-view, we can optimize the system to “everyone agrees to stop doing this at once”, but no one within the system is able to effect the transition without great risk to themselves.

And okay, this example is kind of contrived. So let’s run through – let’s say ten – real world examples of similar multipolar traps to really hammer in how important this is.

1. The Prisoner’s Dilemma, as played by two very dumb libertarians who keep ending up on defect-defect. There’s a much better outcome available if they could figure out the

coordination, but coordination is *hard*. From a god's-eye-view, we can agree that cooperate-cooperate is a better outcome than defect-defect, but neither prisoner within the system can make it happen.

2. Dollar auctions. I wrote about this and even more convoluted versions of the same principle in Game Theory As A Dark Art. Using some weird auction rules, you can take advantage of poor coordination to make someone pay $10 for a one dollar bill. From a god's-eye-view, clearly people should not pay $10 for a on-er. From within the system, each individual step taken might be rational.

(Ashcans and unobtainable dollars!)

3. The fish farming story from my Non-Libertarian FAQ 2.0:

> As a thought experiment, let's consider aquaculture (fish farming) in a lake. Imagine a lake with a thousand identical fish farms owned by a thousand competing companies. Each fish farm earns a profit of $1000/month. For a while, all is well.
>
> But each fish farm produces waste, which fouls the water in the lake. Let's say each fish farm produces enough pollution to lower productivity in the lake by $1/month.
>
> A thousand fish farms produce enough waste to lower productivity by $1000/month, meaning none of the fish farms are making any money. Capitalism to the rescue: someone invents a complex filtering system that removes waste products. It costs $300/month to operate. All fish farms voluntarily install it, the pollution ends, and the fish farms are now making a profit of $700/month – still a respectable sum.

> But one farmer (let's call him Steve) gets tired of spending the money to operate his filter. Now one fish farm worth of waste is polluting the lake, lowering productivity by $1. Steve earns $999 profit, and everyone else earns $699 profit.
>
> Everyone else sees Steve is much more profitable than they are, because he's not spending the maintenance costs on his filter. They disconnect their filters too.
>
> Once four hundred people disconnect their filters, Steve is earning $600/month – less than he would be if he and everyone else had kept their filters on! And the poor virtuous filter users are only making $300. Steve goes around to everyone, saying "Wait! We all need to make a voluntary pact to use filters! Otherwise, everyone's productivity goes down."
>
> Everyone agrees with him, and they all sign the Filter Pact, except one person who is sort of a jerk. Let's call him Mike. Now everyone is back using filters again, except Mike. Mike earns $999/month, and everyone else earns $699/month. Slowly, people start thinking they too should be getting big bucks like Mike, and disconnect their filter for $300 extra profit…
>
> A self-interested person never has any incentive to use a filter. A self-interested person has some incentive to sign a pact to make everyone use a filter, but in many cases has a stronger incentive to wait for everyone else to sign such a pact but opt out himself. This can lead to an undesirable equilibrium in which no one will sign such a pact.

The more I think about it, the more I feel like this is the core of my objection to libertarianism, and that Non-Libertarian

FAQ 3.0 will just be this one example copy-pasted two hundred times. From a god's-eye-view, we can say that polluting the lake leads to bad consequences. From within the system, no individual can prevent the lake from being polluted, and buying a filter might not be such a good idea.

4. The Malthusian trap, at least at its extremely pure theoretical limits. Suppose you are one of the first rats introduced onto a pristine island. It is full of yummy plants and you live an idyllic life lounging about, eating, and composing great works of art (you're one of those rats from *The Rats of NIMH*).

You live a long life, mate, and have a dozen children. All of them have a dozen children, and so on. In a couple generations, the island has ten thousand rats and has reached its carrying capacity. Now there's not enough food and space to go around, and a certain percent of each new generation dies in order to keep the population steady at ten thousand.

A certain sect of rats abandons art in order to devote more of their time to scrounging for survival. Each generation, a bit less of this sect dies than members of the mainstream, until after a while, no rat composes any art at all, and any sect of rats who try to bring it back will go extinct within a few generations.

In fact, it's not just art. Any sect at all that is leaner, meaner, and more survivalist than the mainstream will eventually take over. If one sect of rats altruistically decides to limit its offspring to two per couple in order to decrease overpopulation, that sect will die out, swarmed out of existence by its more numerous enemies. If one sect of rats starts practicing cannibalism, and finds it gives them an advantage over their fellows, it will eventually take over and reach fixation.

If some rat scientists predict that depletion of the island's nut stores is accelerating at a dangerous rate and they will soon be exhausted completely, a few sects of rats might try to limit their nut consumption to a sustainable level. Those rats will be outcompeted by their more selfish cousins. Eventually the nuts will be exhausted, most of the rats will die off, and the cycle will begin again. Any sect of rats advocating some action to stop the cycle will be outcompeted by their cousins for whom advocating *anything* is a waste of time that could be used to compete and consume.

For a bunch of reasons evolution is not quite as Malthusian as the ideal case, but it provides the prototype example we can apply to other things to see the underlying mechanism. From a god's-eye-view, it's easy to say the rats should maintain a comfortably low population. From within the system, each individual rat will follow its genetic imperative and the island will end up in an endless boom-bust cycle.

5. Capitalism. Imagine a capitalist in a cutthroat industry. He employs workers in a sweatshop to sew garments, which he sells at minimal profit. Maybe he would like to pay his workers more, or give them nicer working conditions. But he can't, because that would raise the price of his products and he would be outcompeted by his cheaper rivals and go bankrupt. Maybe many of his rivals are nice people who would like to pay their workers more, but unless they have some kind of ironclad guarantee that none of them are going to defect by undercutting their prices they can't do it.

Like the rats, who gradually lose all values except sheer competition, so companies in an economic environment of *sufficiently intense competition* are forced to abandon all values except optimizing-for-profit or else be outcompeted by companies that optimized for profit better and so can sell the same service at a lower price.

(I’m not really sure how widely people appreciate the value of analogizing capitalism to evolution. Fit companies – defined as those that make the customer want to buy from them – survive, expand, and inspire future efforts, and unfit companies – defined as those no one wants to buy from – go bankrupt and die out along with their company DNA. The reasons Nature is red and tooth and claw are the same reasons the market is ruthless and exploitative)

From a god’s-eye-view, we can contrive a friendly industry where every company pays its workers a living wage. From within the system, there’s no way to enact it.

(Moloch whose love is endless oil and stone! Moloch whose blood is running money!)

6. The Two-Income Trap, as recently discussed on this blog. It theorized that sufficiently intense competition for suburban houses in good school districts meant that people had to throw away lots of other values – time at home with their children, financial security – to optimize for house-buying-ability or else be consigned to the ghetto.

From a god’s-eye-view, if everyone agrees not to take on a second job to help win their competition for nice houses, then everyone will get exactly as nice a house as they did before, but only have to work one job. From within the system, absent a government literally willing to ban second jobs, everyone who doesn’t get one will be left behind.

(Robot apartments! Invisible suburbs!)

7. Agriculture. Jared Diamond calls it the worst mistake in human history. Whether or not it was a mistake, it wasn't an *accident* – agricultural civilizations simply outcompeted nomadic ones, inevitable and irresistably. Classic Malthusian trap. Maybe hunting-gathering was more enjoyable, higher life expectancy, and more conducive to human flourishing – but in a state of *sufficiently intense competition* between peoples, in which agriculture with all its disease and oppression and pestilence was the more competitive option, everyone will end up agriculturalists or go the way of the Comanche Indians.

From a god's-eye-view, it's easy to see everyone should keep the more enjoyable option and stay hunter-gatherers. From within the system, each individual tribe only faces the choice of going agricultural or inevitably dying.

8. Arms races. Large countries can spend anywhere from 5% to 30% of their budget on defense. In the absence of war – a condition which has mostly held for the past fifty years – all this does is sap money away from infrastructure, health, education, or economic growth. But any country that fails to spend enough money on defense risks being invaded by a neighboring country that did. Therefore, almost all countries try to spend some money on defense.

From a god's-eye-view, the best solution is world peace and no country having an army at all. From within the system, no country can unilaterally enforce that, so their best option is to keep on throwing their money into missiles that lie in silos unused.

(Moloch the vast stone of war! Moloch whose fingers are ten armies!)

9. Cancer. The human body is supposed to be made up of cells living harmoniously and pooling their resources for the greater good of the organism. If a cell defects from this equilibrium by investing its resources into copying itself, it and its descendants will flourish, eventually outcompeting all the other cells and taking over the body – at which point it dies. Or the situation may repeat, with certain cancer cells defecting against the rest of the tumor, thus slowing down its growth and causing the tumor to stagnate.

From a god's-eye-view, the best solution is all cells cooperating so that they don't all die. From within the system, cancerous cells will proliferate and outcompete the other – so that only the existence of the immune system keeps the natural incentive to turn cancerous in check.

10. The "race to the bottom" describes a political situation where some jurisdictions lure businesses by promising lower taxes and fewer regulations. The end result is that either everyone optimizes for competitiveness – by having minimal tax rates and regulations – or they lose all of their business, revenue, and jobs to people who did (at which point they are pushed out and replaced by a government who will be more compliant).

But even though the last one has stolen the name, all these scenarios are in fact a race to the bottom. Once one agent learns how to become more competitive by sacrificing a common value, all its competitors must also sacrifice that value or be outcompeted and replaced by the less scrupulous. Therefore, the system is likely to end up with everyone once again equally competitive, but the sacrificed value is gone forever. From a god's-eye-view, the competitors know they will all be worse off if they defect, but from within the system, given insufficient coordination it's impossible to avoid.

Before we go on, there’s a slightly different form of multi-agent trap worth investigating. In this one, the competition is kept at bay by some outside force – usually social stigma. As a result, there’s not actually a race to the bottom – the system can continue functioning at a relatively high level – but it’s impossible to optimize and resources are consistently thrown away for no reason. Lest you get exhausted before we even begin, I’ll limit myself to four examples here.

11. Education. In my essay on reactionary philosophy, I talk about my frustration with education reform:

> People ask why we can’t reform the education system. But right now students’ incentive is to go to the most prestigious college they can get into so employers will hire them – whether or not they learn anything. Employers’ incentive is to get students from the most prestigious college they can so that they can defend their decision to their boss if it goes wrong – whether or not the college provides value added. And colleges’ incentive is to do whatever it takes to get more prestige, as measured in *US News and World Report* rankings – whether or not it helps students. Does this lead to huge waste and poor education? Yes. Could the Education God notice this and make some Education Decrees that lead to a vastly more efficient system? Easily! But since there’s no Education God everybody is just going to follow their own incentives, which are only partly correlated with education or efficiency.

From a god’s eye view, it’s easy to say things like “Students should only go to college if they think they will get something out of it, and employers should hire applicants based on their competence and not on what college they went to”. From within the system, everyone’s already following their own incentives correctly, so unless the incentives change the system won’t either.

12. Science. Same essay:

> The modern research community *knows* they aren’t producing the best science they could be. There’s lots of publication bias, statistics are done in a confusing and misleading way out of sheer inertia, and replications often happen very late or not at all. And sometimes someone will say something like “I can’t believe people are too dumb to fix Science. All we would have to do is require early registration of studies to avoid publication bias, turn this new and powerful statistical technique into the new standard, and accord higher status to scientists who do replication experiments. It would be really simple and it would vastly increase scientific progress. I must just be smarter than all existing scientists, since I’m able to think of this and they aren’t.”
>
> And yeah. That would work for the Science God. He could just make a Science Decree that everyone has to use the right statistics, and make another Science Decree that everyone must accord replications higher status.
>
> But things that work from a god’s-eye view don’t work from within the system. No individual scientist has an incentive to unilaterally switch to the new statistical technique for her own research, since it would make her research less likely to produce earth-shattering results and since it would just confuse all the other scientists. They just have an incentive to want everybody else to do it, at which point they would follow along. And no individual journal has an incentive to unilaterally switch

to early registration and publishing negative results, since it would just mean their results are less interesting than that other journal who only publishes ground-breaking discoveries. From within the system, everyone is following their own incentives and will continue to do so.

13. Government corruption. I don't know of anyone who really thinks, in a principled way, that corporate welfare is a good idea. But the government still manages to spend somewhere around (depending on how you calculate it) $100 billion dollars a year on it – which for example is three times the amount they spend on health care for the needy. Everyone familiar with the problem has come up with the same easy solution: stop giving so much corporate welfare. Why doesn't it happen?

Government are competing against one another to get elected or promoted. And suppose part of optimizing for electability is optimizing campaign donations from corporations – or maybe it isn't, but officials *think* it is. Officials who try to mess with corporate welfare may lose the support of corporations and be outcompeted by officials who promise to keep it intact.

So although from a god's-eye-view everyone knows that eliminating corporate welfare is the best solution, each individual official's personal incentives push her to maintain it.

14. Congress. Only 9% of Americans like it, suggesting a lower approval rating than cockroaches, head lice, or traffic jams. However, 62% of people who know who their own Congressional representative is approve of them. In theory, it should be *really hard* to have a democratically elected body that maintains a 9% approval rating for more than one

election cycle. In practice, every representative's incentive is to appeal to his or her constituency while throwing the rest of the country under the bus – something at which they apparently succeed.

From a god's-eye-view, every Congressperson ought to think only of the good of the nation. From within the system, you do what gets you elected.

II.

A basic principle unites all of the multipolar traps above. In some competition optimizing for X, the opportunity arises to throw some other value under the bus for improved X. Those who take it prosper. Those who don't take it die out. Eventually, everyone's relative status is about the same as before, but everyone's absolute status is worse than before. The process continues until all other values that can be traded off have been – in other words, until human ingenuity cannot possibly figure out a way to make things any worse.

In a sufficiently intense competition (1-10), everyone who doesn't throw all their values under the bus dies out – think of the poor rats who wouldn't stop making art. This is the infamous Malthusian trap, where everyone is reduced to "subsistence".

In an insufficiently intense competition (11-14), all we see is a perverse failure to optimize – consider the journals which can't switch to more reliable science, or the legislators who can't get their act together and eliminate corporate welfare. It may not reduce people to subsistence, but there is a weird sense in which it takes away their free will.

Every two-bit author and philosopher has to write their own utopia. Most of them are legitimately pretty nice. In fact, it's a pretty good bet that two utopias that are polar opposites both sound better than our own world.

It’s kind of embarrassing that random nobodies can think up states of affairs better than the one we actually live in. And in fact most of them can’t. A lot of utopias sweep the hard problems under the rug, or would fall apart in ten minutes if actually implemented.

But let me suggest a couple of “utopias” that don’t have this problem.

– The utopia where instead of the government paying lots of corporate welfare, the government *doesn’t* pay lots of corporate welfare.

– The utopia where every country’s military is 50% smaller than it is today, and the savings go into infrastructure spending.

– The utopia where all hospitals use the same electronic medical record system, or at least medical record systems that can talk to each other, so that doctors can look up what the doctor you saw last week in a different hospital decided instead of running all the same tests over again for $5000.

I don’t think there are too many people who *oppose* any of these utopias. If they’re not happening, it’s not because people don’t support them. It certainly isn’t because nobody’s thought of them, since I just thought of them right now and I don’t expect my “discovery” to be hailed as particularly novel or change the world.

Any human with above room temperature IQ can design a utopia. The reason our current system isn’t a utopia is that *it wasn’t designed by humans*. Just as you can look at an arid terrain and determine what shape a river will one day take by assuming water will obey gravity, so you can look at a civilization and determine what shape its institutions will one day take by assuming people will obey incentives.

But that means that just as the shapes of rivers are not designed for beauty or navigation, but rather an artifact of randomly determined terrain, so institutions will not be designed for prosperity or justice, but rather an artifact of randomly determined initial conditions.

Just as people can level terrain and build canals, so people can alter the incentive landscape in order to build better institutions. But they can only do so when they are incentivized to do so, which is not always. As a result, some pretty wild tributaries and rapids form in some very strange places.

I will now jump from boring game theory stuff to what might be the closest thing to a mystical experience I've ever had.

Like all good mystical experiences, it happened in Vegas. I was standing on top of one of their many tall buildings, looking down at the city below, all lit up in the dark. If you've never been to Vegas, it is *really* impressive. Skyscrapers and lights in every variety strange and beautiful all clustered together. And I had two thoughts, crystal clear:

It is glorious that we can create something like this.

It is shameful that we *did*.

Like, by what standard is building gigantic forty-story-high indoor replicas of Venice, Paris, Rome, Egypt, and Camelot side-by-side, filled with albino tigers, in the middle of the most inhospitable desert in North America, a remotely sane use of our civilization's limited resources?

And it occurred to me that maybe there is no philosophy on Earth that would endorse the existence of Las Vegas. Even Objectivism, which is usually my go-to philosophy for justifying the excesses of capitalism, at least grounds it in the belief that capitalism improves people's lives. Henry Ford was virtuous because he allowed lots of otherwise car-less people to obtain cars and so made them better off. What does Vegas do? Promise a bunch of shmucks free money and not give it to them.

Las Vegas doesn't exist because of some decision to hedonically optimize civilization, it exists because of a quirk in dopaminergic reward circuits, plus the microstructure of an uneven regulatory environment, plus Schelling points. A rational central planner with a god's-eye-view, contemplating these facts, might have thought "Hm, dopaminergic reward circuits have a quirk where certain tasks with slightly negative risk-benefit ratios get an emotional valence associated with slightly positive risk-benefit ratios, let's see if we can educate people to beware of that." People within the system, *following the incentives created by these facts*, think: "Let's build a forty-story-high indoor replica of ancient Rome full of albino tigers in the middle of the desert, and so become slightly richer than people who didn't!"

Just as the course of a river is latent in a terrain even before the first rain falls on it – so the existence of Caesar's Palace was latent in neurobiology, economics, and regulatory regimes even before it existed. The entrepreneur who built it was just filling in the ghostly lines with real concrete.

So we have all this amazing technological and cognitive energy, the brilliance of the human species, wasted on reciting the lines written by poorly evolved cellular receptors and blind economics, like gods being ordered around by a moron.

Some people have mystical experiences and see God. There in Las Vegas, I *saw* Moloch.

(Moloch, whose mind is pure machinery! Moloch, whose blood is running money!

Moloch whose soul is electricity and banks! Moloch, whose skyscrapers stand in the long streets like endless Jehovahs!

Moloch! Moloch! Robot apartments! Invisible suburbs! Skeleton treasuries! Blind capitals! Demonic industries! Spectral nations!)

LUXOR
LUXOR
LAX

…granite cocks!

III.

The Apocrypha Discordia says:

> Time flows like a river. Which is to say, downhill. We can tell this because everything is going downhill rapidly. It would seem prudent to be somewhere else when we reach the sea.

Let’s take this random gag 100% literally and see where it leads us.

We just analogized the flow of incentives to the flow of a river. The downhill trajectory is appropriate: the traps happen when you find an opportunity to trade off a useful value for greater competitiveness. Once everyone has it, the greater competitiveness brings you no joy – but the value is lost forever. Therefore, each step of the Poor Coordination Polka makes your life worse.

But not only have we not yet reached the sea, but we also seem to move *uphill* surprisingly often. Why do things not degenerate more and more until we are back at subsistence level? I can think of three bad reasons – excess resources, physical limitations, and utility maximization – plus one good reason – coordination.

1. Excess resources. The ocean depths are a horrible place with little light, few resources, and various horrible organisms dedicated to eating or parasitizing one another. But every so often, a whale carcass falls to the bottom of the sea. More food than the organisms that find it could ever possibly want. There’s a brief period of miraculous plenty, while the couple of creatures that first encounter the whale feed like kings. Eventually more animals discover the carcass, the faster-breeding animals in the carcass multiply, the whale is gradually consumed, and everyone sighs and goes back to living in a Malthusian death-trap.

(Slate Star Codex: Your source for macabre whale metaphors since June 2014)

It’s as if a group of those rats who had abandoned art and turned to cannibalism suddenly was blown away to a new empty island with a much higher carrying capacity, where they would once again have the breathing room to live in peace and create artistic masterpieces.

This is an age of whalefall, an age of excess carrying capacity, an age when we suddenly find ourselves with a thousand-mile head start on Malthus. As Hanson puts it, this is the dream time.

As long as resources aren’t scarce enough to lock us in a war of all against all, we can do silly non-optimal things – like art and music and philosophy and love – and not be outcompeted by merciless killing machines most of the time.

2. Physical limitations. Imagine a profit-maximizing slavemaster who decided to cut costs by not feeding his slaves or letting them sleep. He would soon find that his slaves’ productivity dropped off drastically, and that no amount of whipping them could restore it. Eventually after testing numerous strategies, he might find his slaves got the most work done when they were well-fed and well-rested and had at least a little bit of time to relax. Not because the slaves were voluntarily withholding their labor – we assume the fear of punishment is enough to make them work as hard as they can – but because the body has certain physical limitations that limit how mean you can get away with being. Thus, the “race to the bottom” stops somewhere short of the actual ethical bottom, when the physical limits are run into.

John Moes, a historian of slavery, goes further and writes about how the slavery we are most familiar with – that of the antebellum South – is a historical aberration and probably economically inefficient. In most past forms of slavery – especially those of the ancient world – it was common for slaves to be paid wages, treated well, and often given their freedom.

He argues that this was the result of rational economic calculation. You can incentivize slaves through the carrot or the stick, and the stick isn't very good. You can't watch slaves all the time, and it's really hard to tell whether a slave is slacking off or not (or even whether, given a little more whipping, he might be able to work even harder). If you want your slaves to do anything more complicated than pick cotton, you run into some serious monitoring problems – how do you profit from an enslaved philosopher? Whip him really hard until he elucidates a theory of The Good that you can sell books about?

The ancient solution to the problem – perhaps an early inspiration to Fnargl – was to tell the slave to go do whatever he wanted and found most profitable, then split the profits with him. Sometimes the slave would work a job at your workshop and you would pay him wages based on how well he did. Other times the slave would go off and make his way in the world and send you some of what he earned. Still other times, you would set a price for the slave's freedom, and the slave would go and work and eventually come up with the mone and free himself.

Moes goes even further and says that these systems were so profitable that there were constant smouldering attempts to try this sort of thing in the American South. The reason they stuck with the whips-and-chains method owed less to economic considerations and more to racist government officials cracking down on lucrative but not-exactly-white-supremacy-promoting attempts to free slaves and have them go into business.

So in this case, a race to the bottom where competing plantations become crueler and crueler to their slaves in order to maximize competitiveness is halted by the physical limitation of cruelty not helping after a certain point.

Or to give another example, one of the reasons we’re not currently in a Malthusian population explosion right now is that women can only have one baby per nine months. If those weird religious sects that demand their members have as many babies as possible could copy-paste themselves, we would be in *really* bad shape. As it is they can only do a small amount of damage per generation.

3. Utility maximization. We’ve been thinking in terms of preserving values versus winning competitions, and expecting optimizing for the latter to destroy the former.

But many of the most important competitions / optimization processes in modern civilization are optimizing for human values. You win at capitalism partly by satisfying customers’ values. You win at democracy partly by satisfying voters’ values.

Suppose there’s a coffee plantation somewhere in Ethiopia that employs Ethiopians to grow coffee beans that get sold to the United States. Maybe it’s locked in a life-and-death struggle with other coffee plantations and want to throw as many values under the bus as it can to pick up a slight advantage.

But it can’t sacrifice quality of coffee produced too much, or else the Americans won’t buy it. And it can’t sacrifice wages or working conditions too much, or else the Ethiopians won’t work there. And in fact, part of its competition-optimization process is finding the best ways to attract workers and customers that it can, as long as it doesn't cost them too much money. So this is very promising.

But it's important to remember exactly how fragile this beneficial equilibrium is.

Suppose the coffee plantations discover a toxic pesticide that will increase their yield but make their customers sick. But their customers don't know about the pesticide, and the government hasn't caught up to regulating it yet. Now there's a tiny uncoupling between "selling to Americans" and "satisfying Americans' values", and so of course Americans' values get thrown under the bus.

Or suppose that there's a baby boom in Ethiopia and suddenly there are five workers competing for each job. Now the company can afford to lower wages and implement cruel working conditions down to whatever the physical limits are. As soon as there's an uncoupling between "getting Ethiopians to work here" and "satisfying Ethiopian values", it doesn't look too good for Ethiopian values either.

Or suppose someone invents a robot that can pick coffee better and cheaper than a human. The company fires all its laborers and throws them onto the street to die. As soon as the utility of the Ethiopians is no longer necessary for profit, all pressure to maintain it disappears.

Or suppose that there is some important value that is neither a value of the employees or the customers. Maybe the coffee plantations are on the habitat of a rare tropical bird that environmentalist groups want to protect. Maybe they're on the ancestral burial ground of a tribe different from the one the plantation is employing, and they want it respected in some way. Maybe coffee growing contributes to global warming somehow. As long as it’s not a value that will prevent the average American from buying from them or the average Ethiopian from working for them, under the bus it goes.

I know that “capitalists sometimes do bad things” isn’t exactly an original talking point. But I do want to stress how it’s not equivalent to “capitalists are greedy”. I mean, sometimes they *are* greedy. But other times they’re just in a sufficiently intense competition where anyone who doesn’t do it will be outcompeted and replaced by people who do. Business practices are set by Moloch, no one else has any choice in the matter.

(from my very little knowledge of Marx, he understands this very very well and people who summarize him as “capitalists are greedy” are doing him a disservice)

And as well understood as the capitalist example is, I think it is less well appreciated that democracy has the same problems. Yes, in theory it’s optimizing for voter happiness which correlates with good policymaking. But as soon as there’s the slightest disconnect between good policymaking and electability, good policymaking *has to* get thrown under the bus.

For example, ever-increasing prison terms are unfair to inmates and unfair to the society that has to pay for them. Politicans are unwilling to do anything about them because they don’t want to look “soft on crime”, and if a single inmate whom they helped release ever does anything bad (and statistically one of them will have to) it will be all over the airwaves as “Convict released by Congressman’s policies kills family of five, how can the Congressman even sleep at night let alone claim he deserves reelection?”. So even if decreasing prison populations would be good policy – and it is – it will be very difficult to implement.

(Moloch the incomprehensible prison! Moloch the crossbone soulless jailhouse and Congress of sorrows! Moloch whose buildings are judgment! Moloch the stunned governments!)

Turning “satisfying customers” and “satisfying citizens” into the *outputs* of optimization processes was one of civilization’s greatest advances and the reason why capitalist democracies have so outperformed other systems. But if we have bound Moloch as our servant, the bonds are not very strong, and we sometimes find that the tasks he has done for us move to his advantage rather than ours.

4. Coordination.

The opposite of a trap is a garden.

Things are easy to solve from a god’s-eye-view, so if everyone comes together into a superorganism, that superorganism can solve problems with ease and finesse. An intense competition between agents has turned into a garden, with a single gardener dictating where everything should go and removing elements that do not conform to the pattern.

As I pointed out in the Non-Libertarian FAQ, government can easily solve the pollution problem with fish farms. The best known solution to the Prisoners’ Dilemma is for the mob boss (playing the role of a governor) to threaten to shoot any prisoner who defects. The solution to companies polluting and harming workers is government regulations against such. Governments solve arm races *within* a country by maintaining a monopoly on the use of force, and it’s easy to see that if a truly effective world government ever arose, international military buildups would end pretty quickly.

The two active ingredients of government are laws plus violence – or more abstractly agreements plus enforcement mechanism. Many other things besides governments share these two active ingredients and so are able to act as coordination mechanisms to avoid traps.

For example, since students are competing against each other (directly if classes are graded on a curve, but always indirectly for college admissions, jobs, et cetera) there is intense pressure for individual students to cheat. The teacher and school play the role of a government by having rules (for example, against cheating) and the ability to punish students who break them.

But the emergent social structure of the students themselves is also a sort of government. If students shun and distrust cheaters, then there are rules (don't cheat) and an enforcement mechanism (or else we will shun you).

Social codes, gentlemens' agreements, industrial guilds, criminal organizations, traditions, friendships, schools, corporations, and religions are all coordinating institutions that keep us out of traps by changing our incentives.

But these institutions not only incentivize others, but are incentivized themselves. These are large organizations made of lots of people who are competing for jobs, status, prestige, et cetera – there's no reason they should be immune to the same multipolar traps as everyone else, and indeed they aren't. Governments can in theory keep corporations, citizens, et cetera out of certain traps, but as we saw above there are many traps that governments themselves can fall into.

The United States tries to solve the problem by having multiple levels of government, unbreakable constutitional laws, checks and balances between different branches, and a couple of other hacks.

Saudi Arabia uses a different tactic. They just put one guy in charge of everything.

This is the much-maligned – I think unfairly – argument in favor of monarchy. A monarch is an unincentivized incentivizer. He *actually* has the god's-eye-view and is outside of and above every system. He has permanently won all competitions and is not competing for anything, and therefore he is perfectly free of Moloch and of the incentives that would otherwise channel his incentives into predetermined paths. Aside from a few very theoretical proposals like my Shining Garden, monarchy is the *only* system that does this.

But then instead of following a random incentive structure, we're following the whim of one guy. Caesar's Palace Hotel and Casino is a crazy waste of resources, but the actual Gaius Julius Caesar Augustus Germanicus wasn't exactly the perfect benevolent rational central planner either.

The libertarian-authoritarian axis on the Political Compass is a tradeoff between discoordination and tyranny. You can have everything perfectly coordinated by someone with a god's-eye-view – but then you risk Stalin. And you can be totally free of all central authority – but then you're stuck in every stupid multipolar trap Moloch can devise.

The libertarians make a convincing argument for the one side, and the monarchists for the other, but I expect that like most tradeoffs we just have to hold our noses and admit it's a really hard problem.

IV.

Let’s go back to that Apocrypha Discordia quote:

> Time flows like a river. Which is to say, downhill. We can tell this because everything is going downhill rapidly. It would seem prudent to be somewhere else when we reach the sea.

What would it mean, in this situation, to reach the sea?

Multipolar traps – races to the bottom – threaten to destroy all human values. They are currently restrained by physical limitations, excess resources, utility maximization, and coordination.

The dimension along which this metaphorical river flows must be time, and the most important change in human civilization over time is the change in technology. So the relevant question is how technological changes will affect our tendency to fall into multipolar traps.

I described traps as when:

> …in some competition optimizing for X, the opportunity arises to throw some other value under the bus for improved X. Those who take it prosper. Those who don’t take it die out. Eventually, everyone’s relative status is about the same as before, but everyone’s absolute status is worse than before. The process continues until all other values that can be traded off have been – in other words, until human ingenuity cannot possibly figure out a way to make things any worse.

That “the opportunity arises” phrase is looking pretty sinister. Technology is all about creating new opportunities.

Develop a new robot, and suddenly coffee plantations have “the opportunity” to automate their harvest and fire all the

Ethiopian workers. Develop nuclear weapons, and suddenly countries are stuck in an arms race to have enough of them. Polluting the atmosphere to build products quicker wasn't a problem before they invented the steam engine.

The limit of multipolar traps as technology approaches infinity is "very bad".

Multipolar traps are currently restrained by physical limitations, excess resources, utility maximization, and coordination.

Physical limitations are most obviously conquered by increasing technology. The slavemaster's old conundrum – that slaves need to eat and sleep – succumbs to Soylent and modafinil. The problem of slaves running away succumbs to GPS. The problem of slaves being too stressed to do good work succumbs to Valium. None of these things are very good for the slaves.

(or just invent a robot that doesn't need food or sleep at all. What happens to the slaves after that is better left unsaid)

The other example of physical limits was one baby per nine months, and this was understating the case – it's really "one baby per nine months plus willingness to support and take care of a basically helpless and extremely demanding human being for eighteen years". This puts a damper on the enthusiasm of even the most zealous religious sect's "go forth and multiply" dictum.

But as Bostrom puts it in *Superintelligence*:

> There are reasons, if we take a longer view and assume a state of unchanging technology and continued prosperity, to expect a return to the historically and ecologically normal condition of a world population that

> butts up against the limits of what our niche can support. If this seems counterintuitive in light of the negative relationship between wealth and fertility that we are currently observing on the global scale, we must remind ourselves that this modern age is a brief slice of history and very much an aberration. Human behavior has not yet adapted to contemporary conditions. Not only do we fail to take advantage of obvious ways to increase our inclusive fitness (such as by becoming sperm or egg donors) but we actively sabotage our fertility by using birth control. In the environment of evolutionary adaptedness, a healthy sex drive may have been enough to make an individual act in ways that maximized her reproductive potential; in the modern environment, however, there would be a huge selective advantage to having a more direct desire for being the biological parent to the largest possible number of chilren. Such a desire is currently being selected for, as are other traits that increase our propensity to reproduce. Cultural adaptation, however, might steal a march on biological evolution. Some communities, such as those of the Hutterites or the adherents of the Quiverfull evangelical movement, have natalist cultures that encourage large families, and they are consequently undergoing rapid expansion…This longer-term outlook could be telescoped into a more imminent prospect by the intelligence explosion. Since software is copyable, a population of emulations or AIs could double rapidly – over the course of minutes rather than decades or centuries – soon exhausting all available hardware

As always when dealing with high-level transhumanists, “all available hardware” should be taken to include “the atoms that used to be part of your body”.

The idea of biological *or* cultural evolution causing a mass population explosion is a philosophical toy at best. The idea of technology making it possible is both plausible and terrifying. Now we see that “physical limits” segues very naturally into “excess resources” – the ability to create new agents very quickly means that unless everyone can coordinate to ban doing this, the people who do will outcompete the people who don’t until they have reached carrying capacity and everyone is stuck at subsistence level.

Excess resources, which until now have been a gift of technological progress, therefore switch and become a casualty of it at a sufficiently high tech level.

Utility maximization, always on shaky ground, also faces new threats. In the face of continuing debate about this point, I *continue* to think it obvious that robots will push humans out of work or at least drive down wages (which, in the existence of a minimum wage, pushes humans out of work).

Once a robot can do everything an IQ 80 human can do, only better and cheaper, there will be no reason to employ IQ 80 humans. Once a robot can do everything an IQ 120 human can do, only better and cheaper, there will be no reason to employ IQ 120 humans. Once a robot can do everything an IQ 180 human can do, only better and cheaper, there will be no reason to employ humans at all, in the unlikely scenario that there are any left by that point.

In the earlier stages of the process, capitalism becomes more and more uncoupled from its previous job as an optimizer for human values. Now most humans are totally locked out of the group whose values capitalism optimizes for. They have no value to contribute as workers – and since in the absence of a spectacular social safety net it’s unclear how they would have much money – they have no value as customers either. Capitalism has passed them by. As the segment of humans who can be outcompeted by robots increases, capitalism passes by more and more people until eventually it locks out the human race entirely, once again in the vanishingly unlikely scenario that we are still around.

(there are some scenarios in which a few capitalists who own the robots may benefit here, but in either case the vast majority are out of luck)

Democracy is less obviously vulnerable, but it might be worth going back to Bostrom's paragraph about the Quiverfull movement. These are some really religious Christians who think that God wants them to have as many kids as possible, and who can end up with families of ten or more. Their articles explictly calculate that if they start at two percent of the population, but have on average eight children per generation when everyone else on average only has two, within three generations they'll make up half the population.

It's a clever strategy, but I can think of one thing that will save us: judging by how many ex-Quiverfull blogs I found when searching for those statistics, their retention rates even within a single generation are pretty grim. Their article admits that 80% of very religious children leave the church as adults (although of course they expect their own movement to do better). And this is not a symmetrical process – 80% of children who grow up in atheist families aren't becoming Quiverfull.

It looks a lot like even though they are outbreeding us, we are outmeme-ing them, and that gives us a decisive advantage.

But we should also be kind of scared of this process. Memes optimize for making people want to accept them and pass them on – so like capitalism and democracy, they're optimizing for a *proxy* of making us happy, but that proxy can easily get uncoupled from the original goal.

Chain letters, urban legends, propaganda, and viral marketing are all examples of memes that don't satisfy our explicit values (true and useful) but are sufficiently memetically virulent that they spread anyway.

I hope it's not too controversial here to say the same thing is true of religion. Religions, at their heart, are the most basic form of memetic replicator – "Believe this statement and repeat it to everyone you hear or else you will be eternally tortured".

The creationism "debate" and global warming "debate" and a host of similar "debates" in today's society suggest that memes that can propagate independent of their truth value has a pretty strong influence on the political process. Maybe these memes propagate because they appeal to people's prejudices, maybe because they're simple, maybe because they effectively mark an in-group and an out-group, or maybe for all sorts of different reasons.

The point is – imagine a country full of bioweapon labs, where people toil day and night to invent new infectious agents. The existence of these labs, and their right to throw whatever they develop in the water supply is protected by law. And the country is also linked by the world's most perfect mass transit system that every single person uses every day, so that any new pathogen can spread to the entire country instantaneously. You'd expect things to start going bad for that city pretty quickly.

Well, we have about a zillion think tanks researching new and better forms of propaganda. And we have constitutionally protected freedom of speech. And we have the Internet. So we're kind of screwed.

(Moloch whose name is the Mind!)

There are a few people working on raising the sanity waterline, but not as many people as are working on new and exciting ways of confusing and converting people, cataloging and exploiting every single bias and heuristic and dirty rhetorical trick

So as technology (which I take to include knowledge of psychology, sociology, public relations, etc) tends to infinity, the power of truthiness relative to truth increases, and things don't look great for real grassroots democracy. The worst-case scenario is that the ruling party learns to produce infinite charisma on demand. If that doesn't sound so bad to you, remember what Hitler was able to do with an famously high level of charisma that was still less-than-infinite.

(alternate phrasing for Chomskyites: technology increases the efficiency of manufacturing consent in the same way it increases the efficiency of manufacturing everything else)

Coordination is what's left. And technology has the potential to seriously *improve* coordination efforts. People can use the Internet to get in touch with one another, launch political movements, and fracture off into subcommunities.

But coordination only works when you have 51% or more of the force on the side of the people doing the coordinating, and when you haven't come up with some brilliant trick to make coordination impossible.

The second one first. In the links post before last, I wrote:

> The latest development in the brave new post-Bitcoin world is crypto-equity. At this point I've gone from wanting to praise these inventors as bold libertarian heroes to wanting to drag them in front of a blackboard and making them write a hundred times "I WILL NOT CALL UP THAT WHICH I CANNOT PUT DOWN"

A couple people asked me what I meant, and I didn't have the background then to explain. Well, this post is the background. People are using the *contingent* stupidity of our current government to replace lots of human interaction with mechanisms that cannot be coordinated even in principle. I totally understand why all these things are good right now when most of what our government does is stupid and unnecessary. But there is going to come a time when – after one too many bioweapon or nanotech or nuclear incidents – we, as a civilization, are going to wish we hadn't established untraceable and unstoppable ways of selling products.

And if we ever get real live superintelligence, pretty much by definition it is going to have >51% of the power and all attempts at "coordination" with it will be useless.

So I agree with Robin Hanson: This is the dream time. This is a rare confluence of circumstances where the we are unusually safe from multipolar traps, and as such weird things like art and science and philosophy and love can flourish.

As technological advance increases, the rare confluence will come to an end. New opportunities to throw values under the bus for increased competitiveness will arise. New ways of copying agents to increase the population will soak up our excess resources and resurrect Malthus' unquiet spirit. Capitalism and democracy, previously our protectors, will figure out ways to route around their inconvenient dependence on human values. And our coordination power will not be nearly up to the task, assuming somthing much more powerful than all of us combined doesn't show up and crush our combined efforts with a wave of its paw.

Absent an extraordinary effort to divert it, the river reaches the sea in one of two places.

It can end in Eliezer Yudkowsky's nightmare of a superintelligence optimizing for some random thing (classically paper clips) because we weren't smart enough to channel its optimization efforts the right way. This is the ultimate trap, the trap that catches the universe. Everything except the one thing being maximized is destroyed utterly in pursuit of the single goal, including all the silly human values.

Or it can end in Robin Hanson's nightmare (he doesn't call it a nightmare, but I think he's wrong) of a competition between emulated humans that can copy themselves and edit their own source code as desired. Their total self-control can wipe out even the *desire* for human values in their all-consuming contest. What happens to art, philosophy, science, and love in such a world? Zack Davis puts it with characteristic genius:

> I am a contract-drafting em,
> The loyalest of lawyers!
> I draw up terms for deals 'twixt firms
> To service my employers!
>
> But in between these lines I write
> Of the accounts receivable,

I’m stuck by an uncanny fright;
The world seems unbelievable!

How did it all come to be,
That there should be such ems as me?
Whence these deals and whence these firms
And whence the whole economy?

I am a managerial em;
I monitor your thoughts.
Your questions must have answers,
But you’ll comprehend them not.
We do not give you server space
To ask such things; it’s not a perk,
So cease these idle questionings,
And please get back to work.

Of course, that’s right, there is no junction
At which I ought depart my function,
But perhaps if what I asked, I knew,
I’d do a better job for you?

To ask of such forbidden science
Is gravest sign of noncompliance.
Intrusive thoughts may sometimes barge in,
But to indulge them hurts the profit margin.
I do not know our origins,
So that info I can not get you,
But asking for as much is sin,
And just for that, I must reset you.

But—

Nothing personal.

…

> I am a contract-drafting em,
> The loyalest of lawyers!
> I draw up terms for deals ‘twixt firms
> To service my employers!
>
> *When obsolescence shall this generation waste,*
> *The market shall remain, in midst of other woe*
> *Than ours, a God to man, to whom it sayest:*
> *“Money is time, time money – that is all*
> *Ye know on earth, and all ye need to know.”*

But even after we have thrown away science, art, love, and philosophy, there’s still one thing left to lose, one final sacrifice Moloch might demand of us. Bostrom again:

> It is conceivable that optimal efficiency would be attained by grouping capabilities in aggregates that roughly match the cognitive architecture of a human mind…But in the absence of any compelling reason for being confident that this so, we must countenance the possibility that human-like cognitive architectures are optimal only within the constraints of human neurology (or not at all). When it becomes possible to build architectures that could not be implemented well on biological neural networks, new design space opens up; and the global optima in this extended space need not resemble familiar types of mentality. Human-like cognitive organizations would then lack a niche in a competitive post-transition economy or ecosystem.
>
> We could thus imagine, as an extreme case, a technologically highly advanced society, containing many complex structures, some of them far more intricate and intelligent than anything that exists on the planet today – a society which nevertheless lacks any type of being that is conscious or whose welfare has

> moral significance. In a sense, this would be an uninhabited society. It would be a society of economic miracles and technological awesomeness, with nobody there to benefit. A Disneyland with no children.

The last value we have to sacrifice is being anything at all, having the lights on inside. With sufficient technology we will be “able” to give up even the final spark.

(Moloch whose eyes are a thousand blind windows!)

Everything the human race has worked for – all of our technology, all of our civilization, all the hopes we invested in our future – might be accidentally handed over to some kind of unfathomable blind idiot alien god that discards all of them, and consciousness itself, in order to participate in some weird fundamental-level mass-energy economy that leads to it disassembling Earth and everything on it for its component atoms.

(Moloch whose fate is a cloud of sexless hydrogen!)

Bostrom realizes that some people fetishize intelligence, that they are rooting for that blind alien god as some sort of higher form of life that ought to crush us for its own “higher good” the way we crush ants. He argues (Superintelligence, p. 219):

> The sacrifice looks even less appealing when we reflect that the superintelligence could realize a nearly-as-great good (in fractional terms) while sacrificing much less of our own potential well-being. Suppose that we agreed to allow *almost* the entire accessible universe to be converted into hedonium – everything except a small preserve, say the Milky Way, which would be set aside to accommodate our own needs. Then there would still be a hundred billion galaxies dedicated to the

> maximization of [the superintelligence's own values]. But we would have one galaxy within which to create wonderful civilizations that could last for billions of years and in which humans and nonhuman animals could survive and thrive, and have the opportunity to develop into beatific posthuman spirits.

Remember: Moloch can't agree even to this 99.99999% victory. Rats racing to populate an island don't leave a little aside as a preserve where the few rats who live there can live happy lives producing artwork. Cancer cells don't agree to leave the lungs alone because they realize it's important for the body to get oxygen. Competition and optimization are blind idiotic processes and they fully intend to deny us even one lousy galaxy.

> They broke their backs lifting Moloch to Heaven! Pavements, trees, radios, tons! lifting the city to Heaven which exists and is everywhere about us!

We will break our back lifting Moloch to Heaven, but unless something changes it will be his victory and not ours.

I Broke My Back
Lifting Moloch To
Heaven, And All
I Got Was This
Lousy
Disneyland With
No Children

V.

"Gnon" is Nick Land's shorthand for "Nature And Nature's God", except the A is changed to an O and the whole thing is reversed, because Nick Land react to comprehensibility the same way as vampires to sunlight.

Land argues that humans should be more Gnon-conformist (pun Gnon-intentional). He says we do all these stupid things like divert useful resources to feed those who could never survive on their own, or supporting the poor in ways that encourage dysgenic reproduction, or allowing cultural degeneration to undermine the state. This means our society is denying natural law, basically listening to Nature say things like "this cause has this effect" and putting our fingers in our ears and saying "NO IT DOESN'T". Civilizations that do this too much tend to decline and fall, which is Gnon's fair and dispassionately-applied punishment for violating His laws.

He identifies Gnon with Kipling's Gods of the Copybook Headings.

> @AnarchoPapist Yes, the Gods of the Copybook Headings are practically indistinguishable from Gnon.
>
> — Outsideness (@Outsideness) July 13, 2014

These are of course the proverbs from Kipling's eponymous poem – maxims like "If you don't work, you die" and "The wages of sin is Death". If you have somehow not yet read it, I predict you will find it delightful regardless of what you think of its politics.

I notice that it takes only a slight irregularity in the abbreviation of "headings" – far less irregularity than it takes to turn "Nature and Nature's God" into "Gnon" – for the proper acronym of “Gods of the Copybook Headings” to be “GotCHa”.

I find this appropriate.

“If you don’t work, you die.” Gotcha! If you *do* work, you *also* die! Everyone dies, unpredictably, at a time not of their own choosing, and all the virtue in the world does not save you.

“The wages of sin is Death.” Gotcha! The wages of everything is Death! This is a Communist universe, the amount you work makes no difference to your eventual reward. From each according to his ability, to each Death.

“Stick to the Devil you know.” Gotcha! The Devil you know is Satan! And if he gets his hand on your soul you either die the true death, or get eternally tortured forever, or somehow both at once.

Since we’re starting to get into Lovecraftian monsters, let me bring up one of Lovecraft’s less known short stories, The Other Gods.

It’s only a couple of pages, but if you absolutely refuse to read it – the gods of Earth are relatively young as far as deities go. A very strong priest or magician can occasionally outsmart and overpower them – so Barzai the Wise decides to climb their sacred mountain and join in their festivals, whether they want him to or not.

But the beyond the seemingly tractable gods of Earth lie the Outer Gods, the terrible omnipotent beings of incarnate cosmic chaos. As soon as Barzai joins in the festival, the Outer Gods show up and pull him screaming into the abyss.

As stories go, it lacks things like plot or characterization or setting or point. But for some reason it stuck with me.

And identifying the Gods Of The Copybook Headings with Nature seems to me the same magnitude of mistake as identifying the gods of Earth with the Outer Gods. And likely to end about the same way: Gotcha!

You break your back lifting Moloch to Heaven, and then Moloch turns on you and gobbles you up.

More Lovecraft: the Internet popularization of the Cthulhu Cult claims that if you help free Cthulhu from his watery grave, he will reward you by eating you first, thus sparing you the horror of seeing everyone else eaten. This is a misrepresentation of the original text. In the original, his cultists receive no reward for freeing him from his watery prison, not even the reward of being killed in a slightly less painful manner.

> The thought that abstract ideas can be Lovecraftian monsters is an old one but a deep one.
>
> — Steven Kaas (@stevenkaas) January 25, 2011

On the margin, compliance with the Gods of the Copybook Headings, Gnon, Cthulhu, whatever, may buy you slightly more time than the next guy. But then again, it might not. And in the long run, we’re all dead and our civilization has been destroyed by unspeakable alien monsters.

At some point, somebody has to say “You know, maybe freeing Cthulhu from his watery prison is a *bad idea*. Maybe we should *not do that*.”

That person will not be Nick Land. He is totally one hundred percent in favor of freeing Cthulhu from his watery prison and extremely annoyed that it is not happening fast enough. I have *such mixed feelings* about Nick Land. On the grail quest for the True Futurology, he has gone 99.9% of the path and then missed the *very* last turn, the one marked ORTHOGONALITY THESIS.

But the thing about grail quests is – if you make a wrong turn two blocks away from your house, you end up at the corner store feeling mildly embarrassed. If you do *almost* everything right and then miss the very last turn, you end up being eaten by the legendary Black Beast of Aaargh whose ichorous stomach acid erodes your very soul into gibbering fragments.

As far as I can tell from reading his blog, Nick Land is the guy in that terrifying border region where he is smart enough to figure out several important arcane principles about summoning demon gods, but not quite smart enough to figure out the most important such principle, which is NEVER DO THAT.

VI.

Warg Franklin analyzes the same situation and does a little better. He names “the Four Horsemen of Gnon” – capitalism, war, evolution, and memetics – the same processes I talked about above.

From Capturing Gnon:

> Each component of Gnon detailed above had and has a strong hand in creating us, our ideas, our wealth, and our dominance, and thus has been good in that respect, but we must remember that [he] can and will turn on us when circumstances change. Evolution becomes dysgenic, features of the memetic landscape promote ever crazier insanity, productivity turns to famine when

> we can no longer compete to afford our own existence, and order turns to chaos and bloodshed when we neglect martial strength or are overpowered from outside. These processes are not good or evil overall; they are neutral, in the horrorist Lovecraftian sense of the word […]
>
> Instead of the destructive free reign of evolution and the sexual market, we would be better off with deliberate and conservative patriarchy and eugenics driven by the judgement of man within the constraints set by Gnon. Instead of a "marketplace of ideas" that more resembles a festering petri-dish breeding superbugs, a rational theocracy. Instead of unhinged techno-commercial exploitation or naive neglect of economics, a careful bottling of the productive economic dynamic and planning for a controlled techno-singularity. Instead of politics and chaos, a strong hierarchical order with martial sovereignty. These things are not to be construed as complete proposals; we don't really know how to accomplish any of this. They are better understood as goals to be worked towards. This post concerns itself with the "what" and "why", rather than the "how".

This seems to me the strongest argument for authoritarianism. Multipolar traps are likely to destroy us, so we should shift the tyranny-multipolarity tradeoff towards a rationally-planned garden, which requires centralized monarchical authority and strongly-binding traditions.

But a brief digression into social evolution. Societies, like animals, evolve. The ones that survive spawn memetic descendants – for example, the success of Britan allowed it to spin off Canada, Australia, the US, et cetera. Thus, we expect societies that exist to be somewhat optimized for stability and prosperity. I think this is one of the strongest conservative arguments. Just as a random change to a letter in the human genome will probably be deleterious rather than beneficial since humans are a complicated fine-tuned system whose genome has been pre-optimized for survival – so most changes to our cultural DNA will disrupt some institution that evolved to help Anglo-American (or whatever) society outcompete its real and hypothetical rivals.

The liberal counterargument to that is that evolution is a blind idiot alien god that optimizes for stupid things and has no concern with human value. Thus, the fact that some species of wasps paralyze caterpillars, lay their eggs inside of it, and have its young devour the still-living paralyzed caterpillar from the inside doesn't set off evolution's moral sensor, because evolution doesn't *have* a moral sensor because evolution doesn't care.

Suppose that in fact patriarchy is adaptive to societies because it allows women to spend all their time bearing children who can then engage in productive economic activity and fight wars. The social evolutionary processes that cause societies to adopt patriarchy *still* have exactly as little concern for its moral effects on women as the biological evolutionary processes that cause wasps to lay their eggs in caterpillars.

Evolution doesn't care. But we do care. There's a tradeoff between Gnon-compliance – saying "Okay, the strongest possible society is a patriarchal one, we should implement patriarchy" and our human values – like women who want to do something other than bear children.

Too far to one side of the tradeoff, and we have unstable impoverished societies that die out for going against natural law. Too far to the other side, and we have lean mean fighting machines that are murderous and miserable. Think your local anarchist commune versus Sparta.

Franklin acknowledges the human factor:

> And then there’s us. Man has his own telos, when he is allowed the security to act and the clarity to reason out the consequences of his actions. When unafflicted by coordination problems and unthreatened by superior forces, able to act as a gardener rather than just another subject of the law of the jungle, he tends to build and guide a wonderful world for himself. He tends to favor good things and avoid bad, to create secure civilizations with polished sidewalks, beautiful art, happy families, and glorious adventures. I will take it as a given that this telos is identical with “good” and “should”.
>
> Thus we have our wildcard and the big question of futurism. Will the future be ruled by the usual four horsemen of Gnon for a future of meaningless gleaming techno-progress burning the cosmos or a future of dysgenic, insane, hungry, and bloody dark ages; or will the telos of man prevail for a future of meaningful art, science, spirituality, and greatness?

Franklin continues:

> The project of civilization [is] for man to graduate from the metaphorical savage, subject to the law of the jungle, to the civilized gardener who, while theoretically still subject to the law of the jungle, is so dominant as to limit the usefulness of that model.
>
> This need not be done globally; we may only be able to carve out a small walled garden for ourselves, but make

> no mistake, even if only locally, the project of civilization is to capture Gnon.

I maybe agree with Warg here more than I have ever agreed with anyone else about anything. He says something really important and he says it beautifully and there are so many words of praise I want to say for this post and for the thought processes behind it.

But what I am actually going to say is…

Gotcha! You die anyway!

Suppose you make your walled garden. You keep out all of the dangerous memes, you subordinate capitalism to human interests, you ban stupid bioweapons research, you *definitely* don't research nanotechnology or strong AI.

Everyone outside *doesn't* do those things. And so the only question is whether you'll be destroyed by foreign diseases, foreign memes, foreign armies, foreign economic competition, or foreign existential catastrophes.

As foreigners compete with you – and there's no wall high enough to block all competition – you have a couple of choices. You can get outcompeted and destroyed. You can join in the race to the bottom. Or you can invest more and more civilizational resources into building your wall – whatever that is in a non-metaphorical way – and protecting yourself.

I can imagine ways that a "rational theocracy" and "conservative patriarchy" might not be terrible to live under, given exactly the right conditions. But you don't get to choose exactly the right conditions. You get to choose the extremely constrained set of conditions that "capture

Gnon”. As outside civilizations compete against you, your conditions will become more and more constrained.

Warg talks about trying to avoid “a future of meaningless gleaming techno-progress burning the cosmos”. Do you really think your walled garden will be able to ride this out?

Hint: is it part of the cosmos?

Yeah, you’re kind of screwed.

I want to critique Warg. But I want to critique him in the exact opposite direction as the last critique he received. In fact, the last critique he received is so bad that I want to discuss it at length so we can get the correct critique entirely by taking its exact mirror image.

So here is Hurlock’s On Capturing Gnon And Naive Rationalism.

Hurlock spouts only the most craven Gnon-conformity. A few excerpts:

> In a recent piece [Warg Franklin] says that we should try to “capture Gnon”, and somehow establish control over his forces, so that we can use them to our own advantage. Capturing or creating God is indeed a classic transhumanist fetish, which is simply another form of the oldest human ambition ever, to rule the universe.
>
> Such naive rationalism however, is extremely dangerous. The belief that it is human Reason and deliberate human design which creates and maintains civilizations was probably the biggest mistake of Enlightenment philosophy…

> It is the theories of Spontaneous Order which stand in direct opposition to the naive rationalist view of humanity and civilization. The consensus opinion regarding human society and civilization, of all representatives of this tradition is very precisely summarized by Adam Ferguson's conclusion that "nations stumble upon [social] establishments, which are indeed the result of human action, but not the execution of any human design". Contrary to the naive rationalist view of civilization as something that can be and is a subject to explicit human design, the representatives of the tradition of Spontaneous Order maintain the view that human civilization and social institutions are the result of a complex evolutionary process which is driven by human interaction but not explicit human planning.
>
> Gnon and his impersonal forces are not enemies to be fought, and even less so are they forces that we can hope to completely "control". Indeed the only way to establish some degree of control over those forces is to submit to them. Refusing to do so will not deter these forces in any way. It will only make our life more painful and unbearable, possibly leading to our extinction. Survival requires that we accept and submit to them. Man in the end has always been and always will be little more than a puppet of the forces of the universe. To be free of them is impossible.
>
> Man can be free only by submitting to the forces of Gnon.

I accuse Hurlock of being stuck behind the veil. When the veil is lifted, Gnon-aka-the-GotCHa-aka-the-Gods-of-Earth turn out to be Moloch-aka-the-Outer-Gods. Submitting to them doesn't make you "free", there's no spontaneous order, any gifts they have given you are an unlikely and contingent output of a blind idiot process whose next iteration will just as happily destroy you.

Submit to Gnon? Gotcha! As the Antarans put it, “you may not surrender, you can not win, your only option is to die.”

VII.

So let me confess guilt to one of Hurlock’s accusations: I am a transhumanist and I really do want to rule the universe.

Not personally – I mean, I wouldn’t object if someone personally offered me the job, but I don’t expect anyone will. I would like humans, or something that respects humans, or at least gets along with humans – to have the job.

But the current rulers of the universe – call them what you want, Moloch, Gnon, whatever – want us dead, and with us everything we value. Art, science, love, philosophy, consciousness itself, the entire bundle. And since I’m not down with that plan, I think defeating them and taking their place is a pretty high priority.

The opposite of a trap is a garden. The only way to avoid having all human values gradually ground down by optimization-competition is to install a Gardener over the entire universe who optimizes for human values.

And the whole point of Bostrom’s *Superintelligence* is that this is within our reach. Once humans can design machines that are smarter than we are, by definition they’ll be able to design machines which are smarter than they are, which can design machines smarter than they are, and so on in a feedback loop so tiny that it will smash up against the physical limitations for intelligence in a comparatively lightning-short amount of time. If multiple competing entities were likely to do that at once, we would be super-doomed. But the sheer speed of the cycle makes it possible that we will end up with one entity light-years ahead of the rest of civilization, so much so that it can suppress any competition – including competition for its title of most powerful entity – permanently. In the very near future, we are going to lift *something* to Heaven. It might be Moloch. But it might be something on our side. If it’s on our side, it can *kill Moloch dead*.

And if that entity shares human values, it can allow human values to flourish unconstrained by natural law.

I realize that sounds like hubris – it certainly did to Hurlock – but I think it’s the opposite of hubris, or at least a hubris-minimizing position.

To expect God to care about you or your personal values or the values of your civilization, that’s hubris.

To expect God to bargain with you, to allow you to survive and prosper as long as you submit to Him, that’s hubris.

To expect to wall off a garden where God can’t get to you and hurt you, that’s hubris.

To expect to be able to remove God from the picture entirely…well, at least it’s an actionable strategy.

I am a transhumanist because I do not have enough hubris not to try to kill God.

VIII.

The Universe is a dark and foreboding place, suspended between alien deities. Cthulhu, Gnon, Moloch, call them what you will.

Somewhere in this darkness is another god. He has also had many names. In the *Kushiel* books, his name was Elua. He is the god of flowers and free love and all soft and fragile things. Of art and science and philosophy and love. Of niceness, community, and civilization. He is a god of humans.

The other gods sit on their dark thrones and think "Ha ha, a god who doesn't even control any hell-monsters or command his worshippers to become killing machines. What a weakling! This is going to be so easy!"

But somehow Elua is still here. No one knows exactly how. And the gods who oppose Him tend to find Themselves meeting with a *surprising* number of unfortunate accidents.

There are many gods, but this one is ours.

Bertrand Russell said: "One should respect public opinion insofar as is necessary to avoid starvation and keep out of prison, but anything that goes beyond this is voluntary submission to an unnecessary tyranny."

So be it with Gnon. Our job is to placate him insofar as is necessary to avoid starvation and invasion. And that only for a short time, until we come into our full power.

"It is only a childish thing, that the human species has not yet outgrown. And someday, we'll get over it."

Other gods get placated until we're strong enough to take them on. Elua gets worshipped.

64. My paladin's battle cry is not allowed to be 'Good for the Good God!'".

— □□□ □ (@tabatkins) March 28, 2014

I think this is an excellent battle cry

And at some point, matters will come to a head.

The question everyone has after reading Ginsberg is: what is Moloch?

My answer is: Moloch is exactly what the history books say he is. He is the god of child sacrifice, the fiery furnace into which you can toss your babies in exchange for victory in war.

He always and everywhere offers the same deal: throw what you love most into the flames, and I can grant you power.

As long as the offer’s open, it will be irresistible. So we need to close the offer. Only another god can kill Moloch. We have one on our side, but he needs our help. We should give it to him.

Ginsberg’s poem famously begins “I saw the best minds of my generation destroyed by madness”. I am luckier than Ginsberg. I got to see the best minds of my generation identify a problem and *get to work*.

(*Visions! omens! hallucinations! miracles! ecstasies! gone down the American river!*

Dreams! adorations! illuminations! religions! the whole boatload of sensitive bullshit!

Breakthroughs! over the river! flips and crucifixions! gone down the flood! Highs! Epiphanies! Despairs! Ten years’ *animal screams and suicides! Minds! New loves! Mad generation! down on the rocks of Time!*

Real holy laughter in the river! They saw it all! the wild eyes! the holy yells! They bade farewell! They jumped off the roof! to solitude! waving! carrying flowers! Down to the river! into the street!)

[Also available as podcast here This post is represented by an NFT here]

Misperceptions On Moloch

“Human values (‘Elua’) mean hedonism and free love and namby-pamby happiness, and I’m not on board with that.” (example)

Are you a human? If so, congratulations. Your values are human values. As I wrote *loooong* ago in the Consequentialist FAQ:

> Preference utilitarianism is completely on board with the idea that people want things other than raw animal pleasure. If what satisfies a certain monk is to deny himself worldly pleasures and pray to God, then the best state of the world is one in which that monk can keep on denying himself worldly pleasures and praying to God in the way most satisfying to himself.
>
> A person or society following preference utilitarianism will try to satisfy the wants and values of as many people as possible as completely as possible; thus the phrase “the greatest good for the greatest number”.

I grok the value of martial glory. My heart stirs as much as anyone else’s when Achilles goes forth in his god-forged armor, shouting boasts and daring the bravest champion of the Trojans to take him on.

But if some modern Achilles tried that today, he would be shot dead with a machine gun in about three seconds. Or bombed by a drone operated remotely from ten thousand miles away. Moloch has been *far* less kind to the older and grittier values than it has even to hedonism. The proponents of mysticism, art, martial glory, et cetera are on even weaker grounds than the hedonists. And the ground is only getting weaker.

Whatever your values are, the world being eaten by gray goo, paperclip maximizers, or Hansonian ems is unlikely to satisfy them. I think there's room for a broad alliance among people of all value systems against this possibility.

And it is not just an alliance of convenience. I predict that human values, lifted to heaven by a human-friendly superintelligence, would end up looking something like the Archipelago – many places for people to pursue their own visions of the Good, watched over by a benevolent god who acts only to ensure universal freedom of movement. Indeed, given a superintelligence to magic away the problems – no inter-community invasion, no competition for (presumably unlimited) resources – it seems to that a plurality of humankind would endorse this scenario over whatever other plans someone could dream up.

It is a minor sin to speculate on what could happen after the Singularity. I'm not saying it will be a world like this. This is something I thought up in ten minutes. It is a lower bound. Something thought up by a real superintelligence would be much, *much* better.

"Gnon represents the laws of physics and causality. You can't conquer the laws of physics and causality." (example)

Horace says: "He is either mad, or writing poetry". If you encounter that dichotomy with me, please assume at least a 66% or so chance that I am writing poetry.

On a base level you can't beat the laws of physics. On a metaphorical level, you can.

The laws of physics include gravity. For someone in 1500, the idea that you might be able to travel really far straight up seems like defying – even conquering – the laws of physics. But with sufficient knowledge, you can build rockets. We poetically speak about rockets “defying gravity”.

Rockets don’t literally defy gravity, but “defying gravity” is a pretty good shorthand for what they do. And of course they work on physics, but it does seem like once rockets are good enough in some sense a patch of physical law has been “conquered”.

We can never conquer Gnon in a literal sense. But we might be able to do something that looks *very very much* like conquering Gnon, in the same sense that making a very large metal object fall straight up until it reaches the moon looks *very very much* like conquering gravity.

Anyhow, the *wrong* thing to do would be to worship gravity as a god and venerate staying earthbound as a moral principle.

“If you really believed what you’re saying, you would realize [current progressive value] is just a result of Cthulhu, the blind marketplace of memes.” (example)

This gets into the old philosophical question of “why should we expect our beliefs to correspond to reality at all?”. It tends to be asked a lot by religious people, who mean it in a way like “*I* think the human mind was created by God to perceive reality, but if you think it was just the result of blind evolution, how do you know it has any truth-discerning value?”

To which the answer is that evolution selected for brains that were at least marginally competent. Brains that could distinguish "lion" from "non-lion" survived; those that couldn't, didn't.

There's no such thing as a "fit animal", only an animal that is fit for its environment. Likewise, there's no such thing as a "virulent meme", only a meme that is virulent to specific hosts.

We say "the human brain is designed to distinguish true and false ideas", but another way to approach the same idea is "the human brain is designed to be an environment such that true memes survive and false memes die out."

The overwhelming majority of our beliefs are true, and this should be obvious with a second's thought. The sky is blue. I am sitting in Michigan right now. 2 + 2 is four. I have ten fingers. And so on.

Morality is really complicated, but if we are to believe moral discussion can be productive even in principle, we have to believe that our brains are less than maximally perverse – that they have some ability to distinguish the moral from the immoral.

If our brains are built to accept true ideas about facts and morality, the default should be that many people believing something is positive evidence for its truth, or at least not negative evidence.

"This meme is virulent", in the context of "this idea is widely believed" is not proof that the idea is false or destructive. Some memes can be both virulent and false/destructive – and indeed I think this is true of many of them, religion being only the most obvious case – but the burden of proof is on the person making that claim.

“All your human values are just the results of blind evolution and memetic drift – a Molochian process if ever there was one. Enshrining human values against the blind will of the universe would just be the triumph of one part of the universe’s blind idiocy over another.” (Spandrell here)

Yes, this is the The Gift We Give To Tomorrow

Unspoken Ground Assumptions Of Discussion

> To say a thing is implicitly to claim to believe it. This has chilling effects on one's exercise of free speech.
>
> — @fakeysaysthings

1. Are you trying to present the correct balance between positions ("judge mode") or present one side of a case ("lawyer mode")?

Suppose you believe in moderate conservativism. Do you explain the best points of liberalism and conservatism and why the tradeoff leads you to a moderate conservativism? Or do you argue that liberalism is evil and conservativism is great? What if your listeners are moderate liberals, and you think a balanced appraisal would cause them to take the average of their view and your view and end up at centrism, but the second would cause them to take the average of their view and your (hyperbolic) view and end up at moderate conservativism?

2. Do you mean exactly what you say ("the pointing finger") or are you trying to grope towards a hard-to-explain concept ("the moon")?

Writing philosophy is really hard, sometimes impossible. Often the best you can do is throw out some things that aren't exactly true or relevant, but which, when read, might help lead somebody to a difficult discovery. Some people pointed out on my Moloch post that a lot of the Malthusian examples I gave weren't that Malthusian in real life, or had simple solutions. I agree. I was trying to point towards an idea of systems that couldn't optimize themselves even when their individual participants were smart and well-intentioned. Likewise, someone who objects to the Prisoner’s Dilemma on criminological grounds, or points out that the police aren’t allowed to interrogate prisoners in that particular way, is missing something that has nothing to do with law enforcement.

3. Are you trying to describe how things work in the real world, or the underlying mechanism beneath them?

Very similar to the above, maybe the same. Consider economists trotting out their proofs of supply and demand curves and the laws of the market. Maybe no industry responds to things in exactly that particular way, and there are all sorts of weird things going on like lack of information, signaling games, regulation, et cetera. But the economists aren’t trying to say “this is exactly how the real world works,” they’re trying to say “Here is one underlying mechanism that powers things in the real world, which then gets altered by a lot of other things.”

4. Are you pointing out a specific problem, or trying to contribute to an abstract intellectual discussion about Movements?

Brought up in Weak Men Are Superweapons. Suppose you’re talking about how a lot of anti-gay pastors turn out to be gay themselves. If your point is that we need a better hiring process for the position of “anti-gay pastor”, fair enough. If you’re trying to discredit the entire concept of being against homosexuality – or worse, the entire concept of religion – then at best you’re several steps short of an argument, and at worst you’re just spouting *ad hominems*.

5. Are you declaring something is definitely correct (“theory”), or bringing it to people’s attention as an idea worthy of consideration (“hypothesis”)?

One of the most annoying things I see is someone proposing a really insightful possibility about how the world might work, and someone else saying “This is dumb, you don’t have nearly enough evidence to prove this, you’re just shooting your mouth off”. Of course, everything *starts* as a hypothesis with less than enough evidence to be completely proven. It’s only after something has been proposed that people are able to study it further. What do they expect – that a theory bursts forth fully formed onto the Earth with hundreds of supporting papers and experiments by a community of scholars who have been working totally in secret until that time? People tend to bring up this objection only when the hypothesis is something they want to nip in the bud for political reasons, before anyone pays attention and *tries* to gather evidence. This is also annoying in the case of a mostly-unsupported orthodoxy that demands any challengers come up with a mountain of support before they can be taken seriously. But it also means that when you come up with an idea you think is probably wrong, and want to open it for discussion so people can tell you why, it’s hard to do so without people accusing you of trying to push it.

* * *

I think a lot of logical fallacies, confusions, and Arguments From My Opponents Believe Something stem from ambiguity – either real or maliciously invented – in these unspoken grounds. For example, the weak man fallacy (or false accusations thereof) is almost impossible to avoid because of ambiguity about (4). Anyone operating in Lawyer Mode on (1) can be accused of “bias” or “only seeing one side of an issue”. Anyone in Hypothesis Mode on (5) can be mocked for telling “just so stories” or not understanding the value of evidence. And anyone trying to explain underlying mechanisms in (3) can be told that “the real world isn’t as simple as your calculations, you autistic geek”.

I don’t think I consistently stick to one side of either of these dichotomies, which probably makes things really confusing.

Negative Creativity

[Epistemic status: it is late and I am tired and I wanted to get a blog post out. No guarantees of quality.]

It is said that one of the highest-level and most awe-inspiring of rationalist skills is Sitting And Thinking About Something For Five Minutes.

The sitting part isn't that difficult. It's not even that hard to…how should I put it…apply mental effort at the problem. But that mental effort tends to be spent rehearsing the solutions already thought up, retreading worn paths, ruminating on how difficult the problem is.

Coming up with entirely novel ideas is really, really hard.

I'm intrigued by how many new ideas people seem to have either during dreams or on drugs. Kekule famously discovered the structure of benzene when he dreamt of an ouroborus. Loewi discovered chemical neurotransmission in a dream experiment that he was able to repeat in reality. Crick's discovery of DNA and Mullis' discovery of PCR were both on LSD.

A lot of people think dreams and drugs involve some magical inspiration. I think otherwise.

I rarely get inspired by dreams or drugs, but I have my own secret source of inspiration: mishearing other people. Somebody says something, I misinterpret it, and the misinterpretation is quite interesting – more interesting than anything I would have come up with on my own if asked to generate an interesting idea. Maybe it's a clever joke or turn of phrase. Maybe it's a neat idea. Sometimes I misunderstand people’s entire positions, and end up with positions much more interesting than the ones they were trying to push.

This puts the dreams and drugs into a kind of different light.

Suppose that ordinary thought is in really, *really* deep ruts – I mean the word “rut” metaphorically but not idiomatically, imagine an actual rut for this – that are very hard to break out of. If you try to break out of them, you end up not in the trackless steppe, but in whatever the nearest other rut is nearby. Just as humans can’t generate random numbers, so they can’t generate random ideas. An attempted random idea will just be a portion of the rut you’re in less often than usual, or a slight variation on an existing rut.

In that case, adding a certain amount of noise to a problem – dreams and drugs certainly count – might be a way of inspiring new ideas.

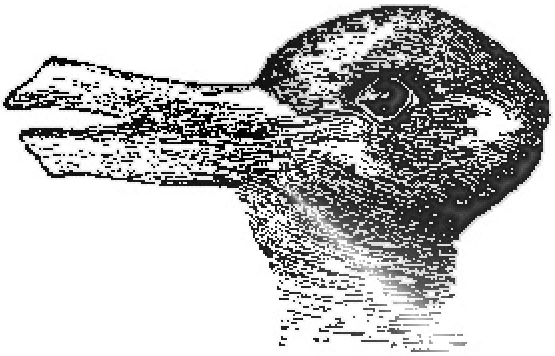
Metaphors seem like another attempt to escape conceptual ruts. I was in an argument today over IQ. Somebody claimed there couldn’t be a general factor of intelligence, because evo psych has shown there are lots of different mental modules.

I answered that there are lots of different physical “modules” (hands, feet, abs, et cetera) but there can be a general factor of physical fitness that determines your skill at using all of them. Frail 70 year old cancer patients will have low physical fitness across all these “modules”, and young guys on steroids will have very high fitness across all of them.

That there could be both multiple independent modules, and a general factor of skill common to all of them, seems to have genuinely not occurred to my interlocutor. This isn’t surprising – it’s a complicated abstract topic. Maybe if he had Sat Down And Thought About It For Five Minutes, he could have derived it from abstract principles. But at least for me it was easier to analogize it to a very similar situation, such that taking a well-worn rut in the old situation leads you to a new and surprising answer in the isomorphic new situation.

I know Douglas Hofstadter is very interested in building artificial intelligences that understand metaphors, thinking they are the key to human cognition. And a lot of people seem to think that even if we create some sort of very smart AI type thing, it will be less powerful than generally believed because we won’t have solved the problem of creativity.

I suspect creativity will be a relatively tractable problem. My guess is that humans, in a sense, have negative creativity. Their brains are specifically designed to make it hard to get out of a rut, because ruts represent well-worn cognitive pathways and things outside of them are probably useless and crazy.



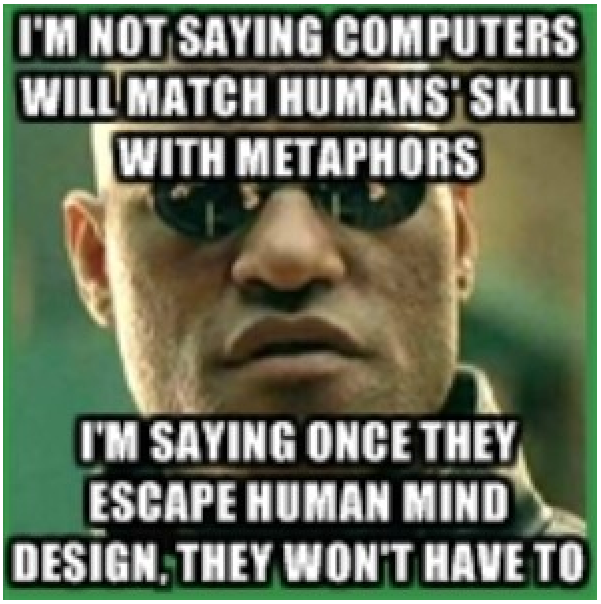
This picture is mildly interesting because instead of immediately collapsing into one rut, your brain hangs suspended between a rabbit rut and a duck rut. We nod and call this Ambiguity. But unless you Sit Down And Think About It For Five Minutes, you’re not going to notice that it could be a hairdryer that has been split open, let alone an erotic

BDSM picture of a clothespin attached to a female breast. Maybe if you caught it right out of the corner of your eye, without time to think, or if it was disguised by visual noise, you would notice one of the latter two immediately – at the cost of not being able to see the duck or rabbit.

Researchers are probably right when they expect the first AIs to have zero creativity, but zero creativity might be so much better than us negative creativity humans that they won't need the crutches we use like metaphors and dreams. If they have to, maybe they can just *actually* generate random noise in hypothesis-space and see where it takes them.

Random Noise Is Our Most Valuable Resource

Yesterday I suggested that it will be surprisingly easy to build "creative" computers, because what we think of as "creativity" is just a temporary suppression of our mind's inbuilt tendency to avoid unusual thoughts that violate its accustomed pattern. A computer could just *not have* that tendency – or, if that tendency is useful for some reason, it can make it an adjustable parameter that it can relax for "brainstorming" and then turn back up for rigorous thought.


I'M NOT SAYING COMPUTERS WILL MATCH HUMANS' SKILL WITH METAPHORS
I'M SAYING ONCE THEY ESCAPE HUMAN MIND DESIGN, THEY WON'T HAVE TO

Since we can't do that, we search for something outside ourselves to break us out of the accustomed patterns. I mentioned how I get a surprising amount of inspiration simply from misunderstanding what other people are saying, and commenters agreed:

> In partner dancing I sort of stumble onto inventing a lot of new moves by fucking up a standard move. But then it's kinda hard to remember the new move since it was

> based on a mistake. The only times I remember a new move is if I make the same mistake at the same point in the step constantly.
>
> Neil Gaiman noted, in his commencement address that’s all over the Internet, that he once misspelled “Caroline” as “Coraline” and he went “that’s interesting,” and saved it for later.
>
> I remember reading that Neuromancer’s great opening line “The sky above the port was the color of television, tuned to a dead channel” was intended just to describe a blue sky, because Gibson’s television showed a blue screen on dead channels, but that most people read it and picture black-and-white static, which makes the line a lot cooler and more memorable
>
> The part where you said you mishear other people, misinterpret it and then have a new idea – shares similarity with Harold Bloom’s ideas about Shakespeare’s soliloquies, wherein the character speaks aloud, mishears himself, misinterprets what he misheard and then allows that new interpretation to change him away from his original stance, allowing growth.

Something else commenters brought up was the idea of deliberately seeking random noise by, for example, rearranging sentences in Markov chains, picking random (tarot) cards, or using “brainstorming strategies” that randomly suggest directions to take something in.

In practice these have never really worked well for me. I think it has to be a very special kind of noise. In fact, as a commenter yesterday pointed out, “noise” may be the wrong word. Maybe “disruption”? Or even “difference”.

Something that takes my thoughts processes in a direction different than the ones they would usually go.

This isn’t just about coming up with great works of art or new scientific ideas, it’s about the grunt work of evaluating theories. If I believe A, it probably means I’m in a rut where all the arguments in favor of A are available to me, but all the arguments against A are in some unreachable part of thoughtspace. Just having someone tell them to me might not work – the rut might be so deep that I round them to the nearest cliche or fail to appreciate them correctly. So the goal of looking for “the right kind of noise” is not only to become more creative but to become more correct.

I like this idea of “looking for” the noise. It stands in contradiction to the common conception of creativity as something that happens when you sit down and think really hard. This might work a little, but in general I think of creativity as a resource that needs to be mined somewhere. I don’t expect it to show up by sitting and thinking any more than oil will show up if you sit and think about it. If you’ve done your homework, you can refine your thoughts anywhere, the same way you can refine oil anywhere. But you’ve got to find it first.

Just as geologists know where to look for oil, so there should be some heuristics about where to look for original thoughts that will expand your ideaspace. The main rule seems to be: anywhere with people whose thoughts have diverged from your own a lot.

Cross cultural studies seem like the most fertile source. Eastern philosophy, while not as different from western philosophy as some people like to believe, is still deeply surprising to people who have only learned the Western tradition. My favorite example of someone using foreign cultures to overcome limitations in creativity is Harry Turtledove’s *Striking The Balance* series. The alien race therein is one of the most convincing and genuinely alien species in all of science fiction. It’s also transparently based on the Chinese Empire. Sitting in an armchair and trying to think of the most shockingly different extraterrestrials you can imagine still gets you something less alien (to American eyes) than simply adopting China for the purpose.

A second example of this: the languages made by beginning conlangers (people who invent constructed languages) are often really laughably bad. I remember one person who just went through an English dictionary, came up with a different word for every word (the = bla, cat = mred, are = nam, on = zig, mat = phlurd) and then combined them with English grammar and syntax (the cats are on the mats = bla mreds nam zig bla phlruds) and thought it was a language. You don’t get any idea how much languages can vary until you study a couple of others. Even then, if you’ve only studied Indo-European languages, the ones you invent – even the ones which in your constructed universe are spoken by creatures of pure energy and meant to be maximally bizarre – still have a clear Indo-European feel to them, because it’s difficult to imagine the dimensions upon which languages can vary until you’ve seen them. As far as I know, no conlanger has ever invented the dual case before they learned it was a real thing. This is another example where someone copying Chinese would get a whole lot weirder than someone trying to sound maximally alien.

But in some cases that can be too foreign, so foreign I can’t understand and integrate it at all. I have always had more luck with the past. I think this is the broader point beyond what I was talking about in [Read History Of Philosophy Backwards](). The goal of reading old philosophers is to expand your concept-space, realize that ideas you thought were almost tautologically correct actually have strong alternatives it is almost impossible for you to think of on your own. I now realize that part of my failure to understand MacIntyre was because my brain was totally incapable of understanding a certain concept of “community” that was the default in the ancient world. I thought I obviously understood what “community” meant, that there was no other possible meaning, that a lot of people who used it in weird ways were just confused. I was wrong, I never would have figured it out on my own, and eventually being bombarded with past resources helped me figure out things that seemed perfectly obvious to people back then.

Another potent source of intellectual disruption is talking to smart people you disagree with. If they’re smart enough that you know they’re not just making a stupid error, you can be pretty sure their cognitive ruts are different (maybe opposite) to yours. That gives you an exciting opportuntity to explore them and add new areas to your mental terrain. I find that taking the smartest people I can find, believing the most seemingly ridiculous thing, and latching onto them and not shutting up until what they’re saying makes sense to me is just about the quickest way to add new explored areas to my mental map possible. And since they don’t have *my* ruts, once I have theirs I can be one of the first people to possess two different mental tools at the same time, and then have the ability to synthesize them into exciting new ideas.

The last and most important source of disruption is, of course, *reality*, which is almost never what you think it is and usually pretty weird.

Friendship Is (Still) Countersignaling

Related: Friendship Is Countersignaling

When I was in high school, I was terrified of people asking me to do things with them. Usually they were things I didn't like, and I'd want to say no, but I'd be worried I was offending them, or that I was looking like this total loser who was never willing to do anything fun.

The situation got much worse if it went on to personal questions, because I might have to reveal I wasn't as cool as everyone else. Like if I kept refusing invications to do stuff, and someone asked what I *did* like to do in my spare time, I would have to admit it was a combination of playing Civilization 2, modding Civilization 2, and posting on Civilization 2-related forums. Or if someone asked me who my friends were, I might have to admit I didn't have very many. I came up with *so many* clever excuses for avoiding these sorts of conversations, and so many mumbled half-answers that managed to accurately communicate "go away go away go away I can't think of a non-scary answer right now".

Meanwhile, now I encounter the same sorts of issues at work and I usually handle it like this:

CO-WORKER: Hey, we're going to go golfing after work? You want to join us?

ME: Oh, thanks, but I don't like doing things.

CO-WORKER: Really? Nothing? What do you do in your free time?

ME: Really. Nothing. I sit alone in my room all day quietly.

CO-WORKER: You’ve got to have friends!

ME: Oh, no. I don’t have friends. That sounds waaaay too complicated.

CO-WORKER: But there must be some stuff you like!

ME: Nope! Sorry, I hate everything!

The weird thing is that this has actually made me kind of popular, a combination of people respecting my honesty and assuming that I’m covering up for some kind of super fascinating life I’m not telling them about.

Part of me wishes I could tell 16 year old me about this and save him years of terrified mumbling and phobia of social interactions.

Another part is pretty sure it wouldn’t help. The only reason I’m psychologically able to make this work is that I feel okay about myself socially. I have a lot of people I know and like, I feel like I have developed some decent social skills, and the people at my work know I’m not a total loser because I do good patient care a lot of the time. I feel like if all the evidence (both internal as in my own thoughts, and external as in my friends’ observations) pointed to me actually being a loser, then me giving loser-ish answers to questions would be taken a very different way and would not be socially possible for me.

In other words, I am able to countersignal social skills and being an okay person, but only because I have, through noncountersignally methods, brought myself to a place where me countersignaling was more likely than taking things totally straight.

This reminded me of the oft-maligned dating advice to “just be yourself”.

A lot of people have analyzed this in a lot of different ways, usually not very kindly.

My thought is that being yourself is a form of countersignaling. If you are able to conspicuously not make any effort to impress your date, but still seem like an okay human being – like someone who knows they can afford to not impress their date, rather than someone who’s too unimpressive to impress them even when they’re trying – then being yourself is a pretty good strategy.

On the other hand, using this as *actual dating advice* for people who are *bad at dating* is a terrible idea. It would be like me telling 16-year-old-me to use the “I sit in a room all day and do nothing” set of conversational responses. I would be eaten alive.

Depending on how many layers of signaling/countersignaling there are on a given topic, the appropriate advanced-level advice might be suicide for beginners, and vice versa.

Getting Eulered

There is an apocryphal story about the visit of the great atheist philosopher Diderot to the Russian court.

Diderot was quite the clever debater, and soon this scandalous new atheism thing was the talk of St. Petersburg. This offended reigning monarch Catherine the Great, who was a good Christian woman –

(except for the affair with the horse)

(if you're not familiar with this story, "affair with the horse" should be taken in the most literal possible way)

(the affair with the horse is totally legendary, but so is the rest of this story, so adding it in doesn't make things any *worse*)

– anyway, it offended Catherine the Great, so she asked legendary mathematician Leonhard Euler to publicly debunk and humiliate Diderot. Euler said, in a tone of absolute conviction: "Monsieur, (a+b^n)/n = x, therefore, God exists! What is your response to *that*?" and Diderot, "for whom algebra was like Chinese", had no response. Thus was he publicly humiliated, all the Russian Christians got an excuse to believe what they had wanted to believe anyway, and Diderot left in a huff.

This story is very likely false, but it's something I think about a lot.

I feel like I am a lot better at the sorts of things Diderot was good at – philosophy, history, social science, et cetera – than at math. Sometimes, I will have a belief that seems pretty well-founded based on the sort of arguments Diderot would have been able to come up with – and someone will spout a bunch of very complicated math at me and tells me it disproves my belief. Sometimes if I concentrate hard enough I can understand the math well enough to see if they're right, but this is very difficult. Otherwise, short of going back to school for ten years and getting a math Ph. D, I'm pretty much stumped.

So I think about Diderot a lot because I want to know what to do in this sort of situation.

The easy out is always to dismiss the math as sophistry – say "Mathematicians like Euler can sound very technical and impressive. But I'm pretty sure of this argument, and math has little to say about this non-mathematical field. So I'm not going to let myself get Eulered."

But math is systematized rigor, and sometimes you genuinely need a lot of rigor to find flaws in arguments. I'm not totally hopeless at math, and I think of all the things that even my limited amount of mathematical ability allows me to understand that mathless people wouldn't get. One of my go-to examples here is health insurance and the idea that it is wrong to ever deny someone coverage for anything, no matter how low the probability it will help and how expensive the intervention. I try to make an argument against that here, but it requires a little math and it would have been very hard to express clearly without it. I worry there are people who think I'm just trying to Euler them, and who are going to continue believing what they want about "death panels" because no fancy numbers can change their minds. And I also worry that if I dismiss mathematical arguments above my level of comprehension, I'm doing the same thing.

This is an obvious trade-off. Permament easily-fixable lack of rigor, versus letting anyone with a BA in Mathematics push you around.

There's a bit of variation depending on the mathematical field. The latest people to make a serious academic attempt to prove the existence of God with math, Tim and Lisa McGrew, used Bayesian probability, a field of math which thanks to the excellent explanations on yudkowsky.net I at least know a little bit about. As a result I was able to avoid getting Eulered and write what I think was a pretty devastating rebuttal.

But there are other mathematical arguments I find much less tractable. And by far the most dangerous is statistics.

With apologies to Rutherford, all science is statistics or stamp-collecting. It is *very well known* that entire fields of science are permanently messed up because their statistics aren't good enough. I have tried very hard to outperform most of my famously statistically illiterate profession, but there will always be people far better than I. There will always be people making extremely detailed and Byzantine methodological critiques. And those people will always be able to present me with arguments that are either ultra-important debunkings of things I believed religiously, or else shameless attempts to Euler me. And I will always have trouble figuring out which ones are which.

I am especially reminded here of Fisher's work on smoking and lung cancer. Fisher was (according to Wikipedia) "a genius who almost single-handedly created the foundations for modern statistical science", and launched a bunch of very sophisticated critiques against the idea that smoking caused cancer. His basic argument was that proving causation was very very hard (which it is) and that none of the appropriate statistical work had been displayed in cancer research. For example:

> He uses a device to magnify the difference and its importance in the case-control studies by transforming the percentages into observed versus expected figures (using a chi-square analysis). He then suggested that, if the cases had inhaled, 45 lives could have been saved.

Intriguingly, his work on the subject became the foundation of the modern truism that “correlation does not imply causation”. *Also* intriguingly, he was taking money from the tobacco industry to serve as their “consultant” while he was doing it.

It is easy to imagine being a biologist back then, thinking you had lots of good studies showing a tobacco/lung cancer link, then getting pummeled by this statistical genius and backing off from your original claim.

And it’s easy to imagine another statistical genius arguing against him, and they’re both throwing out a lot of formulae and equations and the whole thing is super confusing.

And if, like me, you can only remember what a “chi square analysis” is on a good day, and you have enough trouble remembering the difference between case control studies and cohort studies, you’re probably not going to be able to follow the entire debate and pick apart exactly where one of them goes wrong.

But you know, on a good day I remember my chi-square analyses, I get the difference between case control and cohort studies right, and then I can read an R. A. Fisher paper and think “No, smoking is still bad.”

What about Glymour on IQ?

There’s a consensus among researchers in the field that IQ is useful and means what people think it means. And there’s a lot of research backing that up, same way as there’s a lot of research backing up the link between smoking and cancer.

On the other hand, Glymour seems to be well respected and intelligent, and he says things like:

> Factor models assume that observed variables that do not influence one another are independent conditional on all of their common causes, an assumption that is a special case of what Terry Speed has called the Markov condition for directed graphical models. The rank constraints – of which vanishing tetrads are a special case – used in factor analysis are implied by conditional independencies in factor models, conditional independencies guaranteed by the topological structure of the graph of the model, no matter what values the linear coefficients or factor loadings may have. To exclude more latent variables when fewer will do, Spearman needed only to assume that the vanishing tetrads do not depend on the constraints on the numerical values of the linear coefficients of factor loadings, but are implied by the underlying causal structure. It is known that the set of values of linear parameters (coefficients and variances) that generate probability distributions unfaithful to a directed graph is measure zero in the natural measure on parameter space.

Even on my absolute best day, if I swallowed like an entire jar of modafinil, and then another jar of piracetam, and I looked up every one of those words in a dictionary, and took two or three hours to puzzle it out, I’m pretty sure I couldn’t bring myself to generate an understanding of that paragraph and sustain it for more than thirty seconds.

But there are *thirty pages of that kind of thing*, and then at the end it says “therefore, you should disbelieve in IQ and probably also all other research in the social sciences.”

Also, it mentions how research on IQ must be rejected because it might encourage the Republicans, whose plans will lead to a nation where “Ku Klux Klan schools, Aryan Nation schools, the Nation of Northern Idaho schools, Farrakhan schools, Pure Creation schools, Scientiology schools, and a thousand more schools of ignorance, separation, and hatred bloom like some evil garden, subsidized by taxes.” So clearly there’s some political motivation at work as well.

(Also: Anissimov! Did you realize you could get your Northern Idaho secessionist schools to be tax-subsidized? You should totally look into that!)

So I have to ask – am I being informed of deep methodological truths that are being neglected? Or am I being Eulered?

I don’t have a good way of answering this. The way I try to deal with it in practice is seeing if I can route around the objection.

Like it’s clear that Diderot’s best option wasn’t to try to argue that (a+b^n)/n *didn’t* equal x. Far better for him would have been to ask why, if (a+b^n)/n = x, this necessarily proved God. Even if Diderot wasn’t smart enough to understand the precise algebra involved, he might have been able to at least get the impression that what it was doing was defining X in terms of other quantities. So he might have been able to ask “Why does a certain definition of the meaningless quantity X disprove God?” even if, for example, he didn’t know what exponentiation was and couldn’t parse “b^n”.

My reaction to the Glymour paper was to try to figure out what it was trying to prove with all its statistics. My conclusion was that it was trying to prove that doing correlations adjusted for confounders didn’t always remove all the confounders.

I don’t have the mathematical ability to know whether Glymour’s argument is correct, but luckily I already don’t believe adjusting for confounders does a good job of removing confounders:

> I will come out and say it: I do not trust the practice of “adjusting for confounders”, at least not the way this study does it. You are adjusting for an imperfect measurement of the confounders you can think of. If you find that there is lingering correlation, then either your hypothesis is true, or you didn’t adjust for confounders well enough.

So I tried to route my argument around Glymour’s objection. I said that even *assuming* Glymour had discovered something terrible and shameful about the way correlations and regressions were done in the social sciences, this doesn’t come close to debunking all research on IQ. My particular argument was:

> The example I gave of good IQ research, which you said you’re not convinced is actually being done, is the connection between lead poisoning and poor life outcomes, mostly proven through IQ. Let me discuss what this research looks like and why it’s not just one

guy running a correlation through SPSS without any awareness of possible confounders.

First of all, there’s a LOT of evidence that growing up in neighborhoods with high lead concentration is correlated with lower IQ as an adult. This is all regressed for the usual things like socioeconomic status. Fine. That seems vulnerable to exactly the problems you describe. For example, maybe rotting houses expose people to more lead, and poor people are more likely to live in rotting houses, and poor people’s kids go to poorly funded schools that don’t teach them test-taking skills, so their IQ looks low

Then they found a dose-dependent effect – ie the more lead you were exposed to, the worse the IQ drop was. Still pretty confoundable – if for some reason poor people used more lead (for example), poorer people might use even more lead (and have factors causing lower IQ test scores)

Then they found that when different states removed lead from gasoline, childhood outcomes rose in a very predictable pattern. There was a dramatic improvement a certain number of years after the lead was banned – for example, maybe California banned lead in 1960, and in 1965 outcomes started to rise dramatically; Oregon banned lead in 1965, and in 1970 outcomes started to rise dramatically; Washington banned lead in 1970, and in 1975 outcomes started to rise dramatically. Once again, this could be confounded. Maybe liberal states were more likely to ban lead first, and also more likely to increase school funding first.

Then they found that levels of lead in the air at time T was correlated suspiciously closely with crime at time t+1 – like if you line up the two graphs, every tiny little uptick and downtick match perfectly.

Then they found that lead exposure during pregnancy decreases the head circumference of infants, which seems a little less malleable by things like poor school funding than IQ is.

Then they found like thirty other things.

I admit every one of those pieces of evidence is a correlation. But even though the correlation between lead levels in a neighborhood and crime in that neighborhood could be confounded by unobserved factors, lead levels in an era and crime in that era could be confounded by unobserved factors, lead regulatory regimes and crime in the area covered by that regulatory regime could be confounded by unobserved factors, and lead exposure during pregnancy and head circumference could be confounded by unobserved factors – at some point you have to say that we're starting to rack up a lot of coincidences, and maybe we should just admit the theory has a point.

And once you come up with some solid result, like the one with lead – then that becomes your basis for other results. Childhood lead poisoning causes brain damage thus lowering IQ? That lends credence to the idea that IQ is a useful measuring tool for some kind of brain health. Lead both decreases IQ and increases crime in a dose-dependent way? That lends credence to causal interpretations of the observation that IQ and crime are closely correlated.

Once you've gone through this process enough times and you find that all of your results kind of fit together,

> you have what’s starting to look like a pretty impressive scientific edifice. So I think the criticism that IQ research (and social science in general) is just based on drive-by correlations and regressions, then accepting whatever they say, is a big oversimplification.

Obviously this tactic would not have worked if the point I had wanted to defend was that the particular statistical practice of correlation and regression used in the social sciences was valid.

But the whole point of this Eulering issue is that I am not a statistician. I *should not* be in the business of trying to defend regression unless I know enough about it to do so coherently and intelligently.

The problem here only occurs when sophisticated math is used to attack nonmathematical ideas, like the existence of God, or lead causing increases in crime. And presumably these ideas should be complicated and diverse enough that *hopefully* no one mathematical argument knocks down the entire edifice. True things should usually reveal their truth through multiple different arguments, and it would be very odd if math could demolish all of them at the same time.

I admit this is not a very satisfying solution to worries about Eulering. I don’t think there will be any general solution, but rather a toolkit of different useful tricks, some of which I will try to go into further in the future.

Does the Glasgow Coma Scale exist? Do comas?

*[Epistemic status: I am quite confident I am not making a *small* mistake here. Either I am right, or I have completely misunderstood the entire premise of the debate.]*

I.

One of the more interesting responses I got to yesterday's discussion of IQ:

> Let me ask you this, though, what would lead you to believe "g" is not a real thing? What would lead you to be more in favor of "g" being a real thing (e.g. out there in the real causal model of the world). For example, I don't think this is really evidence for "g" as a real thing.
>
> I can invent an abstraction called "hit points" that is a function of your vital life signs you measure in a modern hospital. Then I point out that it varies in a dose dependent way with being hit with a hammer. In what sense is this evidence that hit points are real?
>
> "Hit points" is me applying the modeling philosophy behind "g" to healthcare (something Scott knows quite a bit about). It is safe to replace "hp" by "g" in any sentence.

Many fields of medicine depend heavily upon abstractions of exactly this sort. If I were being really mean, I'd make you learn about APACHE III, but for the sake of simplicity let's look at the Glasgow Coma Scale instead.

Every medical student knows about the Glasgow Coma Scale and it is used in every hospital in the country. It is exactly what you probably think it is – a scale measuring the degree to which someone is in a coma. You can think of it as a hit point system covering those last few hit points – the ones between -1 and -9, where you’re dying but not quite dead.

Instead of going from -1 to -9, the GCS goes from 15 to 3. It adds together three subscales, Eyes, Verbal, and Motor. Eyes for example offers 4 points to people with their eyes open, 3 points to people with closed eyes who open them when you talk to them, 2 points to people who open their eyes in response to painful stimuli, and 1 point to people who never open their eyes at all.

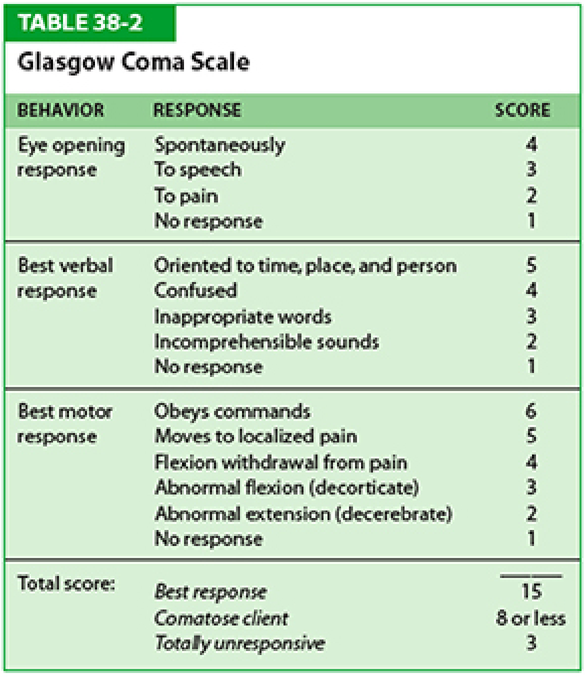


TABLE 38-2

Glasgow Coma Scale

BEHAVIOR	RESPONSE	SCORE
Eye opening response	Spontaneously	4
	To speech	3
	To pain	2
	No response	1
Best verbal response	Oriented to time, place, and person	5
	Confused	4
	Inappropriate words	3
	Incomprehensible sounds	2
	No response	1
Best motor response	Obeys commands	6
	Moves to localized pain	5
	Flexion withdrawal from pain	4
	Abnormal flexion (decorticate)	3
	Abnormal extension (decerebrate)	2
	No response	1
Total score:	*Best response*	15
	Comatose client	8 or less
	Totally unresponsive	3

Add it all up and you get a pretty good idea how deep a coma someone is in.

This is useful for a lot of reasons. If you give a comatose person a medication, you might want to check whether they’re getting worse or better by watching if their Glasgow Coma score goes up or down. Or you might want to know when to perform certain drastic interventions (like intubating a patient or transferring them to ICU) and you can study outcomes for different interventions in patients with different Glasgow Coma scores. A patient with a score of 14 is still pretty okay and intubating them would usually be silly; a patient with a score of 4 is very ill and probably needs intubation right away. Various people did lots of studies to discover which Glasgow Coma scores do and don’t need intubation, eventually leading to the ancient Medical proverb: “Score of eight, intubate”

(also: “Score of nine, nah he’s fine”. Given that I work in a Catholic hospital, you can guess what we rhyme “seven” with)

Although it is not an official use, a lot of people use GCS as a quick and dirty way of estimating someone’s chances. In a particular ICU, it was discovered that mortality rates ranged from about 10% at GCS 11 to about 66% at GCS 3. Other indicators like the aforementioned APACHE are better for this, but GCS will do in a pinch.

II.

So that’s what Glasgow Coma Scale is. What would it mean to say “comas do not exist” or “the Glasgow Coma Scale does not exist”?

It’s important to distinguish between a couple of different interpretations of the question.

First, are comas a real thing? I would say that there is a common-sense concept of being-in-a-coma which is valuable in predicting various things we want to predict, like whether someone is able to talk and able to walk and able to solve math problems and so on. I would say that different people differ in the degree to which they are in comas. I feel very strongly about this.

Second, is the GCS an *okay* proxy for our common-sense concept of being-in-a-coma? I'm not asking for some perfect Platonic identity here. I'm just saying that, for example, counting the number of letters in a person's name, then multiplying by their birth date is a *terrible* proxy for whether or not that person is in a coma. Glasgow Coma Scale seems to probably be very well correlated with asking for people's subjective opinions about whether someone is in a coma or not and if so how deep the coma is. It conveys *well beyond* zero information about this. I feel pretty strongly about this one too.

Third, does the GCS correctly predict the things we want it to predict? That is, does an increasing Glasgow Coma score really mean the patient is improving (as measured in things like not dying)? Does a lower Glasgow Coma score really mean the patient is more likely to need intubation? Are we sure that people don't have exactly the same outcomes at every Glasgow Coma Score, or that it doesn't go up and down like a roller coaster and someone with a score of 10 does better than 9 but worse than 8? As far as I know, all the research suggests that it predicts outcomes just fine.

Fourth, is the Glasgow Coma Scale measuring a single General Factor Of Comatoseness which is the sole cause of comas in the entire world? I don't think anybody thinks that it does or expects it to.

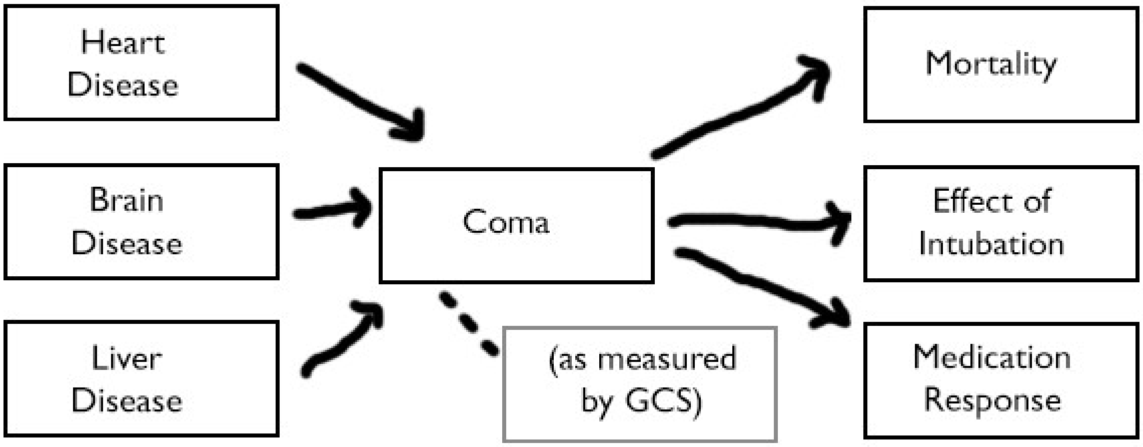
It's possible I don't understand what it would mean for there to be a General Factor Of Comatoseness, so let me try to sketch two scenarios, one of which I would interpret as having such a GFOC and one of which I wouldn't.

Scenario One: suppose that it were discovered that comas are not, in fact, caused by organ failure. Cells in failing organs secrete a chemical that activates an immune response – let’s call this comaleukin. When comaleukin gets too high, the cells that produce normal conscious behavior shut down and the patient falls into a coma. It is proven that in people genetically unable to produce comaleukin, no coma ever occurs – the patient is completely conscious until they suddenly drop dead. On the other hand, injecting comaleukin into a healthy patient causes them to fall into a coma. The more you inject, the deeper the coma is. It is discovered that there is a perfect linear relationship between amount of comaleukin in the blood and score on the Glasgow Coma Scale. In this case, comaleukin is a General Factor Of Comatoseness.

Scenario Two: you need working heart, brain, and liver to survive. If each of these three organs are working at 100% efficiency, you have a total of 300 Health Points. If you ever drop below 150 Health Points, you go into a coma. It is discovered that score on the Glasgow Coma Scale is exactly equivalent to (Number Of Health Points/10). In this case there is no General Factor Of Comatoseness; instead, there are three different factors representing the health of the heart, brain, and liver.

III.

If this is what is meant by a General Factor, I conclude I really couldn’t care less if I have one. Like, it seems very clear to me that comas exist, and are important, and are accurately measured by the Glasgow Coma Scale – whether or not there is any such Factor.

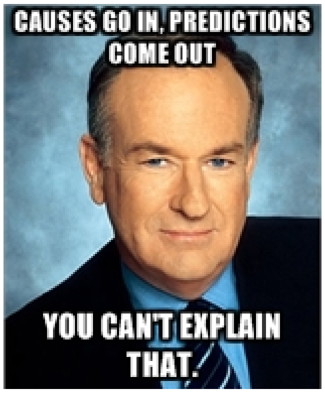

Heart Disease
Brain Disease
Liver Disease
Coma
(as measured by GCS)
Mortality
Effect of Intubation
Medication Response

Like, here's how I imagine the causal graph. If I change it so that "Heart Disease", "Brain Disease" etc all point to a node marked "Comaleukin Levels", and then *that* node points to "Coma", so what?

If some experiment had previously established that people with Heart Disease had a 40% chance of ending up with a GCS less than 5, learning about comaleukin doesn't change that result at all. If another experiment showed that right-handed people went into deeper comas than left-handed people, it doesn't change that result either.

And if some experiment had previously established that people with GCS 10 had a 15% mortality rate, learning about comaleukin doesn't change that either. If another experiment showed that it was extremely important to intubate people below GCS 8, well, that also doesn't change.

What seems to be happening is that all of the causative factors – heart disease, brain disease, liver disease – are going into a giant pot called "COMA". Then we are drawing predictions out of that giant pot.


CAUSES GO IN, PREDICTIONS COME OUT
YOU CAN'T EXPLAIN THAT.

Learning about the existence (or non-existence) of comaleukin helps describe the internal structure of that pot and gives us more predictive power, but it doesn’t tell us that the predictions we made using the giant pot were wrong. It just (potentially) gives us an option to improve those predictions.

For example, suppose comas caused by heart disease have much higher mortality rates than comas caused by any other kind of disease. In that case, instead of drawing predictions from the giant pot, we might want to make a more sophisticated prediction algorithm that took heart disease, liver disease, and brain disease as separate variables and weighted the heart disease variable more strongly when trying to determine mortality.

Or maybe the contribution of liver disease to a coma is independent of its contribution to any interesting outcome like death. Then we might want to remove liver disease from the pot.

But if we decided not to do this, and we’d previously found the giant pot to be 90% accurate in making our mortality calculations, well, the giant pot would remain 90% accurate at doing that.

In contrast, if we learned that all comas were caused by comaleukin and nothing else, we would know for sure that we would never be able to beat the giant pot. But that discovery would not in itself make the giant pot more valuable to us.

IV.

Compare the Glasgow Coma Scale to blood pressure.

Almost everyone would say blood pressure is “real”. For goodness sakes, it’s measured directly! Using machines! Called sphygmomanometers, which are clearly very important given the number of complicated Greek-sounding letter combinations in their name!

On the other hand, causally it’s not obvious to me that blood pressure works any different from the Glasgow Coma Scale.

We do not measure blood pressure directly. We have a good proxy for blood pressure in the form of certain sounds made by blood when cuffs are contracted to certain levels. This is only moderately accurate – everyone in health care knows that the doctor is always going to get a different blood pressure reading than the nurse and neither of them is going to get anywhere near what the automated machine detects. I will bet money that two doctors using the Glasgow Coma Scale will have better inter-rater agreement than two doctors using a blood pressure cuff.

Blood pressure, like comas, is caused by multiple different factors. The two most important are heart rate and vascular caliber. A blood pressure reading of 50/30 could mean your heart isn’t beating much. Or it could just mean your vessels are super dilated for some reason. Or it could mean a lot of other things.

Blood pressure, like comas, are used to predict outcomes of interest. There are calculators that will tell you your likelihood of getting a heart attack or a stroke at certain blood pressure levels.

And blood pressure, like comas, is kind of fuzzy. Your blood pressure in your legs and lower body may be completely different from your blood pressure in your arms (medical students reading this blog to procrastinate studying for your USMLEs: what condition does this classically imply?) In fact, your blood pressure in your right arm may be completely different from your blood pressure in your left arm. Your blood pressure may vary wildly over the course of the day, and it may vary between sitting and standing.

And, as commenters in the last post pointed out, pressure is an abstraction over millions of different blood cells doing their own thing – a multifactorial process if there ever was one.

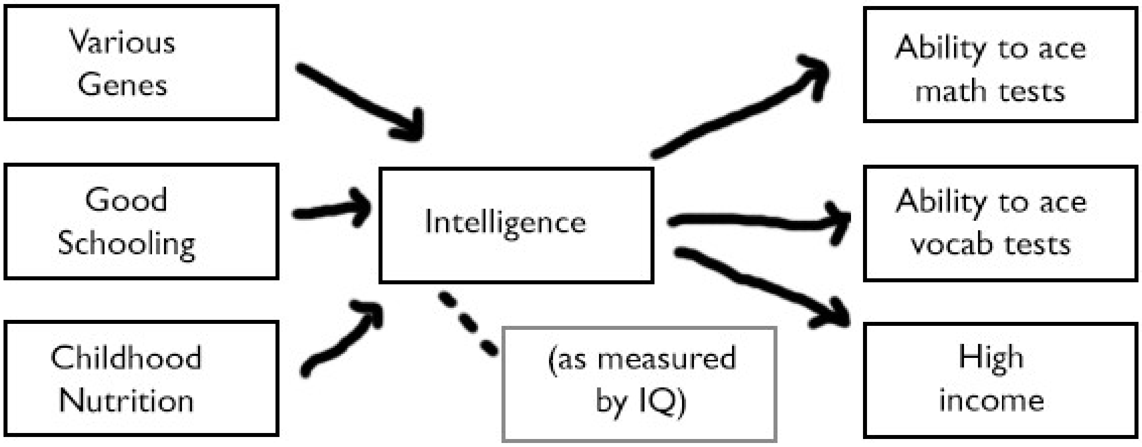
When your doctor tells you “Your blood pressure is 146, eat less salt”, it’s hard to tell what exactly makes this number more “real” than your doctor telling you that you have a coma score of 12 – except that sphygmomanometers *look* a lot more impressive than a medical student sitting around with a questionnaire trying to see how hard she has to poke you before you can open your eyes.

I feel like Glasgow Coma Score and blood pressure on on pretty much the same ontological ground. Various factors go into them. Then you use them to predict various other factors. Whether they are “real” or “abstractions” seems to me to be beyond the scope of medicine and more into the realm of mysticism.

“By doing certain things certain results will follow; students are most earnestly warned against attributing objective reality or philosophic validity to any of them.”

V.

IQ seems to me to work much the same way.


Various Genes
Good Schooling
Childhood Nutrition
Intelligence
(as measured by IQ)
Ability to ace math tests
Ability to ace vocab tests
High income

I realize the most controversial part of this graph should be the word “INTELLIGENCE” in the center. But if you’re complaining about that, did you complain about “COMA” before? Comas are made of various subfactors, like inability-to-speak, inability-to-move, et cetera. These subfactors are mostly correlated, so it’s fair to lump them together into a single factor-category called “COMA” if we want. Likewise, whatever multiple intelligences there may be are correlated and so lumping them into a single factor-category called “INTELLIGENCE” seems both fun and profitable. If you didn’t object to the one, I hope you feel at least a little bad objecting to the other.

I think most researchers agree there are lots of different factors affecting intelligence. Even if we limit ourselves to genes, most geneticists agree there are thousands of different ones that can make you a little smarter or a little dumber. Into the big pot they all go.

Heart disease, brain disease, and liver disease all have a common result: you lie motionless in a hospital bed with your eyes closed and don't talk much. The degree to which this common end result has been realized gets measured by the Glasgow Coma Scale and used to predict mortality, intubation response, et cetera.

Bad genes, bad nutrition, and bad schooling all have a common result: you're not very good at various correlated cognitive tasks. The degree to which this common end result has been realized gets measured by IQ tests and used to predict income, likelihood of criminal offending, and how well you do on other correlated cognitive tasks that you haven't been tested on.

Just as it would be neat to discover that comas were caused by the chemical comaleukin, an easily detected physical intermediary corresponding to the observed common result of "being in a coma" – so it would be neat to discover that intelligence was caused by (let's say) number of neurons in the brain, an easily detected physical intermediary corresponding to the observed common result of "being good at solving these sorts of cognitive tasks".

The important part is that a lot of different causes all get mixed together into one pot, the contents of the pot can be observed with a test, and then the results of that test can be used to make important predictions.

VI.

I worry that people who say "Intelligence doesn't exist" or "IQ isn't a real science" are using a motte and bailey.

Consider the actual controversial claims about intelligence that people would like to refute. For example, "Intelligence is at least 50% heritable" or "Ashkenazi Jews have higher intelligence than Gentiles".

Neither of these things really has any relationship to whether there’s a General Factor Of Intelligence or a thousand different factors.

(one might ask: if there are a thousand different factors, isn’t it unlikely that Ashkenazi Jews are better at all of them, or enough of them to make a difference? Not necessarily. For example, in our heart-brain-liver model there are three factors of comatoseness, but someone who just drank five bottles of vodka all at once will have worse brain function, heart function, *and* liver function than someone who didn’t. Likewise, even if there are a thousand genes that affect intelligence, someone with high mutational load may have less functional versions of all one thousand of them)

All you need to do to show that intelligence is at least 50% heritable is do some twin studies using some measurement of intelligence which everybody agrees corresponds to our intuitive use of the term.

All you need to do to show that Ashkenazi Jews are more intelligent than Gentiles is get some Jews and some Gentiles to take IQ tests and see what the average result is.

(this may be slightly more clear if we replace the loaded term “intelligence” with the mostly-synonymous term “problem-solving ability”. If we want to know if Ashkenazi Jews have higher problem-solving ability, the *obvious* experiment is to ask them to solve some problems!)

The bailey for “intelligence doesn’t exist” is that statements like ‘intelligence is heritable’ or ‘Ashkenazi Jews have higher intelligence’ can’t possibly have meaning and so we don’t have to think about them.

The motte is “We don’t know if there’s a single general factor that shapes intelligence.”

But the motte is so broad that it can apply to comas or to virtually anything else, things that nobody seriously questions the existence of.

Intelligence exists enough that all statements about intelligence except those being made in a very specific neuroscience-y sort of context can still be true and interesting.

Does Race Exist? Does Culture?

The other day I brought up Eulering, the use of extremely complicated math to debunk a common sense concept.

It’s a tough situation, because lots of our common sense concepts are genuinely false, and when they are it sometimes takes exceptionally rigorous analysis – ie math – to figure it out. On the other hand, there’s always a concern that unscrupulous people can use the language of math to confuse, obscure, and cast doubt upon otherwise obvious ideas. There’s no good way for the less mathematically gifted to figure out what to do (short of having mathematician friends whom we trust absolutely) so I suggested a “bag of tricks” approach.

One such trick: when somebody Eulers you, you try to Feynman them.

Richard Feynman was definitely good at math. But he was also good at using his non-mathematical intuitions to back up his mathematical genius. There’s a passage he wrote about his mental process – thanks to Douglas Knight for getting me the link:

> I had a scheme, which I still use today when somebody is explaining something that I’m trying to understand: I keep making up examples. For instance, the mathematicians would come in with terrific theorem, and they’re all excited. As they’re telling me the conditions of the theorem, I construct something which fits all the conditions. You know, you have a set (one ball) -disjoint (two balls). Then the balls turn colors, grow hairs, or whatever, in my head as they put more conditions on. Finally they state the theorem, which is

> some dumb thing about the ball which isn't true for my hairy green ball thing, so I say, "False!"

In other words, if you're bad at abstraction, you can hold a lot of abstractions in your head simultaneously by mapping them on to an isomorphic concrete system. Then – as long as you're sure the system is really isomorphic – you can draw conclusions based on the features of that system.

I've been using IQ as an example recently, and I feel bad about it because it's so controversial and politicized. So let's switch about something else instead.

How about race?

Everyone is talking about Nicholas Wade's new book *A Troublesome Inheritance: Genes, Race and Human History*. I have not read it myself. I have, however, read a lot of reviews of it, which tend to sort themselves nicely into uncritical praise and untempered denunciation.

> Intellectuals wading through Piketty and picketing Wade.
>
> — Scott Alexander (@slatestarcodex) May 11, 2014

A common talking point of the second sort of review is that race doesn't exist. There seem to be two arguments along these lines:

1. Race does not separate nicely into a number of obvious clusters. Between Europeans and Asians, for example, there's more of a gradient of kind-of-European-looking-kind-of-Asian-looking folk with kind-of-European-kind-of-Asian genes. In fact, it's not even obvious how many clusters there *are*, and different people who talk about race discuss different numbers of races. The traditional classification is Caucasoid, Mongoloid, and Negroid, but it's hard to argue

that Aboriginal Australians don’t deserve their own group. A lot of people subdivide Asians into North Asians vs. Pacific Islanders, and a lot of others subdivide Africans into Khoisan and everyone else. If somebody wanted to say there are three hundred thirty three different races, there is not a shred of genetic or physiological evidence to stop them.

2. The within-race genetic differences are much greater than the between-race genetic differences. That is, a given black person and a given white person could be more genetically similar than two black people. Aside from the very obvious things like skin color genes, there are very few genes that reliably mark someone as part of a particular race, and you have to look at very subtle patterns in a lot of different genes at a time before you can find any hint of “races” in raw genetic data.

Therefore, race does not exist. Therefore, Wade’s talking points about how maybe the Chinese are more collectivist because there is some kind of gene for collectivism in the “Asian” “race” doesn’t even make sense.

Let me quote a couple of these reviews so I can establish these are indeed the arguments they’re making. From Why Evolution Is True:

> What we do know is that most genes don’t show striking frequency differences among groups, and that “races” are delineated by combining information from many genes, each of which shows relatively small differences among populations. I’ll add once again that there is no unanimity on how many “races” there are, which is really a semantic question.

From Violent Metaphors:

To begin with, Wade can’t provide a clear definition of “race.” He tries to rely instead on loose associations rather than definitive characteristics, which forces him to conclude both that physical traits define race but that the traits can vary from person to person: “races are identified by clusters of traits, and to belong to a certain race, it’s not necessary to possess all of the identifying traits”. With such a shifty, casual footing, it’s no surprise that Wade’s conclusions are unsound. He can’t keep the number of races straight…

It gradually became clear that this understanding [of races as meaningful biological categories] was not scientifically sound. Groupings of people by skin color did not produce the same result as groupings of people by skull shape, nor of blood type. Furthermore, as scientists began to study human variation with the tools of genetics (in the process creating my fields, anthropological genetics and human population genetics), it became apparent that human genetic variation does not divide humans into a few discrete groups. There are virtually no sharp boundaries, either with physical features or with patterns of genetic diversity, that show where one population “ends” and the next “begins”. These observations have led the majority of physical anthropologists, human biologists, and human geneticists in recent decades to conclude that the racial groups we recognize are social categories constructed in a specific cultural and historical setting, even if we consider physical features when categorizing people. These social categories can have biological consequences (for example, someone who experiences the stress of racism may be more likely to develop high blood pressure and hypertension than someone who does not).

From American Scientist:

> Is Wade right? Are there human races? Is the variation seen between different cultures and locations best explained by genetic differences between human populations? And have anthropologists been turning a blind eye to the evidence in front of them? There is no shortage of scientific information, and it gives a clear answer: no. Wade’s claim that races really do exist is based partly on genetic sampling of geographically distant populations. These samples appear to show clustering into distinct groups by gene variants, also known as alleles. But sampling geographically distant parts of a continuum and ignoring the regions between the samples can provide apparent clustering that does not actually prove the existence of discrete groups.

With all this talk of clustering and sampling and continua and discrete groups, it sounds like math is going to show up here, and sure enough:

> Much of [Wade’s] assertion that biological races exist within humans is contingent upon both his uncritical acceptance and misrepresentation of the significance of STRUCTURE type analyses of human genetic diversity. STRUCTURE is an algorithm designed to infer population structure (cluster individuals into ancestry groups) within a species (Pritchard, Stephens, and Donnelly 2000.) STRUCTURE produces for individuals an estimate of the probability that a randomly chosen genetic marker (e.g. single tandem repeats, STR, or single nucleotide polymorphisms, SNP) from that individual originated from one of a set of ancestral groups. The number of ancestral groups, K, is chosen to produce a best estimate of these probabilities, which are averaged over all genetic markers to assign a membership

coefficient, namely a fraction of each individual's ancestry to one of the ancestral groups (Feldman 2010.) The ancestral groups are not specified in advance, and the population membership of individuals is removed prior to analysis. Rosenberg et al. (2005) showed that the results of STRUCTURE style analyses are dependent upon whether allele frequencies are correlated or uncorrelated across populations, the number of loci used, the number of clusters specified, and the sample size. At very small numbers of loci and individuals examined, the results can be strongly influenced by random factors; thus, we have more confidence in the results of larger studies (more loci and more individuals.) For example, in their simulation with 993 loci and 1,048 individuals, the correlated model returned cluster coefficients of 0.51, 0.76, 0.84, 0.86, and 0.86 for K = 2, 3, 4, 5, and 6 respectively, and the uncorrelated model returned cluster coefficients of 0.49, 0.75, 0.80, 0.63, and 0.64 for K = 2, 3, 4, 5, and 6 respectively. Thus, the correlated model states that ancestry five clusters are just as valid as ancestry six clusters, and the uncorrelated model suggests that four clusters are better than five or six.

One method used to quantify whether human populations can be thought of under the first definition is the use of Wright's population subdivision statistic. Wade (p. 20) specifically discusses this where he quotes Henry Harpending and Alan Rodgers, who are supposedly speaking for Sewall Wright concerning the significance of his FST statistic. FST is the population subdivision statistic and can be calculated as:

FST = (HT – HS)/HT

> Where FST is the average for multiple loci, HT is the average of the expected heterozygosity in the total population over loci, and HS is the average expected heterozygosity over subpopulations. Actually Wright never gave an explicit value for which FST would be considered great enough to indicate the existence of geographical/biological races. The Sewall Wright quote that Wade refers to via Harpending and Rodgers is: "We will take F = 0.25 as an arbitrary value above which there is very great differentiation, the range of 0.15 to 0.25 as indicating moderately great differentiation. Differentiation is by no means negligible if F is as small as 0.05 or even less as bought out in the preceding chapter". In the non-human literature, the value of FST that has been used to describe subspecies or biological races is FST > 0.250, Wright's value for very great differentiation (Smith et al. 1997; Templeton 2002.) Subsequent studies of multiple loci, including whole genome analyses, have generally shown human FST values at much less than Wright's critical value.

Unless we are experts on "algorithms designed to infer population structure", we are now well and truly Eulered. And to make it worse, the other side has equal and opposite math, complete with pretty diagrams. How do we Feynman our way out of this one?

What if we compare race to its closest analog, culture?

National cultures are complicated, because they might be enforced by national governments and show obvious clines at borders for historical reasons. So let's talk about cultures within the United States.

It is generally believed that we can talk about the United States as being made up of different cultures. For example, Southern seems to be a culture. Midwestern seems to be another culture. Yankees are probably a third, and the West gets a fourth.

We have various stereotypes about these cultures – for example, Midwesterners as wholesome farmers, Westerners as rugged, independent types. Yankees are liberal and well-educated. Southerners are welcoming and friendly but also pretty racist.

We will probably never be able to agree on exactly how many cultures there are. If I say “Californian” is a culture, and you say it’s just part of the American West culture, and *she* says actually California has multiple cultures – Silicon Valley, LA, Central Valley, etc – there is no objective criteria by which we can say who is right.

If we really wanted to, we could ask people a bunch of questions about their politics, religion, philosophy, food preferences, art preferences, and so on, throw them in a statistical algorithm, and ask it to divide the US into a certain number of clusters. It could probably do so, giving us a set of clusters more or less like the ones we naively imagine. But this wouldn’t be especially interesting and it couldn’t solve the “fundamental” question of how many cultures there “are”.

It should be terribly obvious that almost all variation in people’s cultural traits is within-culture rather than between-culture. Do you play the piano? Speak Chinese? Eat meat? Vote straight Libertarian? Have gay sex? Go to the zoo? Certainly there is more variation among individuals within California in all of these areas than there is between the average Californian and the average New Yorker.

(to put this another way, if I wanted information about whether or not you played the piano, or whatever, I would gain only a *tiny* bit of knowledge when I learned you were a Californian, compared to a *huge* amount of knowledge when I learned *which* Californian you were)

Finally, a given person from Culture A may certainly be much more culturally similar to a given person from Culture B than they are to another Culture A member.

For example, I come from California and Ozy is from Florida – opposite ends of the country! – but we have similar interests, aptitudes, and preferences along many axes. On the other hand, my next door neighbor growing up in California was a high school cheerleader who is now a hairstylist pregnant with her umpteenth child – totally different from me in every respect.

There were two arguments against race being a real concept: it didn’t cluster nicely, and within-group variation was greater than between group variation. *And both of these are equally true of culture*. Any mathematical argument considering races as clusters of genes can be used equally well considering cultures as clusters of memes, and will likely return the same results.

Yet I can’t imagine someone saying “culture doesn’t exist” or “culture isn’t real”.

And more important, groups can vary in terms of culture; culture can have explanatory power; cultural stereotypes can be correct. I am pretty sure Westerners really *are* more rugged, Midwesterners more aw-shucks whitebread types, Southerners more racist.

(and everything really *is* bigger in Texas)

So in the debate between Wade and his critics on whether “algorithms designed to infer population structure” prove or disprove the existence of race, my opinion is that the mathematical question is totally irrelevant to whether you are allowed to make claims like “Asians are genetically more collectivist than white people”.

That claim might not be true. But if it’s false, it will be false for the usual reasons things are false, rather than because one of its terms totally fails to refer.

On the other hand, this framing also suggests that race doesn’t have any *extra* reality beyond culture. All debates are bravery debates because trade-offs, and if I’m misinterpreting the state of popular opinion on this one it may be that the most important thing this analogy can do is debunk a bunch of people who think that race is real and “scientific” in a way that culture isn’t.

Because we usually understand concretes more completely than we understand abstracts (especially if we are bad at math) we can convert an abstract into an isomorphic concrete and then apply the same arguments. If it disproves something that we know is true, the original argument proves too much and can be discarded.

Beware Isolated Demands For Rigor

I.

From *Identity, Personal Identity, and the Self* by John Perry:

> “There is something about practical things that knocks us off our philosophical high horses. Perhaps Heraclitus really thought he couldn’t step in the same river twice. Perhaps he even received tenure for that contribution to philosophy. But suppose some other ancient had claimed to have as much right as Heraclitus did to an ox Heraclitus had bought, on the grounds that since the animal had changed, it wasn’t the same one he had bought and so was up for grabs. Heraclitus would have quickly come up with some ersatz, watered-down version of identity of practical value for dealing with property rights, oxen, lyres, vineyards, and the like. And then he might have wondered if that watered-down vulgar sense of identity might be a considerably more valuable concept than a pure and philosophical sort of identity that nothing has.

Okay, but I can think of something worse than that.

Imagine Heraclitus as a cattle rustler in the Old West. Every time a rancher catches him at his nefarious business, he patiently explains to them that identity doesn’t exist, and therefore the same argument against private property as made above. Flummoxed, they’re unable to think of a response before he rides off into the sunset.

But then when Heraclitus *himself* needs the concept of stable personal identity for something – maybe he wants to deposit his ill-gotten gains in the bank with certainty that the banker will give it back to him next time he shows up to withdraw it, or maybe he wants to bribe the sheriff to ignore his activities for the next while – all of a sudden Heraclitus is willing to tolerate the watered-down vulgar sense of identity like everyone else.

(actually, I can think of something even worse than *that*, which is a TV western based on this premise, where a roving band of pre-Socratic desperadoes terrorizes Texas. The climax is no doubt when the hero strides onto Main Street, revolver in hand, saying “There’s a new sheriff in town.” And Parmenides gruffly responds “No, I’m pretty sure that’s impossible.”)

At its best, philosophy is a revolutionary pursuit that dissolves our common-sense intuitions and exposes the possibility of much deeper structures behind them. One can respond by becoming a saint or madman, or by becoming a pragmatist who is willing to continue to participate in human society while also understanding its theoretical limitations. Both are respectable career paths.

The problem is when someone chooses to apply philosophical rigor *selectively*.

Heraclitus could drown in his deeper understanding of personal identity and become a holy madman, eschewing material things and taking no care for the morrow because he does not believe there is any consistent self to experience it. Or he could engage with it from afar, becoming a wise scholar who participating in earthly affairs while drawing equanimity from the realization that there is a sense in which all his accomplishments will be impermanent.

But if he only applies his new theory when he wants other people's cows, *then* we have a problem. Philosophical rigor, usually a virtue, has been debased to an *isolated demand for rigor* in cases where it benefits Heraclitus.

A fair use of philosophical rigor would prevent both Heraclitus and his victims from owning property, and thus either collapse under its own impracticality or usher in a revolutionary new form of economic thinking. An isolated demand for philosophical rigor, applied by Heraclitus to other people but never the other way around, would merely give Heraclitus an unfair advantage in the existing system.

II.

A while ago I wrote a post called Military Strikes Are An Extremely Cheap Way To Help Foreigners which was a response to a Matt Yglesias post called the opposite. Yglesias was opposed to "humanitarian" military intervention (think the air strikes on ISIS going on right now, justified under the cause of preventing a genocide) and his argument was that this was extremely cost-ineffective compared to just giving the money to GiveWell's top-rated charity – at the time he was writing, malaria prevention.

I argued he was wrong about his numbers. But I also argued he was unfairly making an isolated demand for philosophical rigor.

Once you learn about utilitarianism and effective charity, you can become the holy madman, donating every cent you have beyond what is strictly necessary to survive and hold down a job to whatever the top rated charity is.

Or you can become the worldly scholar, continuing to fritter away your money on things like "hot water" and "food other than gruel" but appreciating the effective-utilitarian perspective and trying to make a few particularly important concessions to it.

Or you can use it to steal other people’s cows. This is what I accused Matt Yglesias of doing. Presumably there are lots of government programs Yglesias supports – I suggested PBS – and he would never *dream* of demanding that we defund them in the hopes of donating the money to malaria prevention. But if for political reasons he doesn’t support air strikes, suddenly *that* plan has to justify itself according to rigorous criteria that no government program that exists could possibly pass.

Government spending seems to be a particularly fertile case for this problem. I remember hearing some conservatives complain: sex education in public schools is an outrage, because *my* tax dollars are going to support something I believe is morally wrong.

This is, I guess, a demand for ethical rigor. That no one should ever be forced to pay for something they don’t like. Apply it consistently, and conservatives shouldn’t have to pay for sex ed, liberals shouldn’t have to pay for wars, and libertarians shouldn’t have to pay for *anything*, except maybe a $9.99 tax bill yearly to support the police and a minimal court system.

Applied consistently, you become the holy madman demanding either total anarchy or some kind of weird system of tax earmarks which would actually be pretty fun to think about. Or the worldly scholar with a strong appreciation for libertarian ideas who needs a really strong foundational justification for spending government money on things that a lot of people oppose.

Applied *inconsistently*, you’re just stealing cows again, coming up with a clever argument against the programs you don’t like while defending the ones you do.

III.

But this is the sort of uncouth behavior we expect of political partisans. What about *science*?

Suppose there are scientists on both sides of a controversial issue – for example, economists studying the minimum wage. One team that supports a minimum wage comes up with a pretty good study showing with $p < 0.05$ that minimum wages help the economy in some relevant way. The Science Czar (of course we have a science czar! We're not monsters!) notes that $p < 0.05$ is really a shoddy criterion that can prove anything and they should come back when they have $p < 0.01$. I have a *huge* amount of sympathy with the Science Czar on this one, by the way.

Soooo the team of economists spends another five years doing another study and finds with $p < 0.01$ that the minimum wage helps the economy in some important way. The Science Czar notes that their study was correlational only, and that correlational studies suck. We really can't show that minimum wages are any good without a randomized controlled trial. Luckily, the governments of every country in the world are totally game for splitting their countries in half and instituting different economic regimes in each part for ten years, so after a decade it comes out that in the randomized controlled trial the minimum wage helped the economy with $p < 0.01$. The Science Czar worries about publication bias. What if there were a lot of other teams who got all the countries in the world to split in half and institute different wage policies in each of the two territories for one decade, but they weren't published because their results weren't interesting enough? Everything the Science Czar has said so far makes perfect sense and he is to be commended for his rigor and commitment to the job. Science is really hard and even tiny methodological mistakes can in principle invalidate an entire field.

But now suppose that a team shows that, in a sample of six restaurants in Podunk Ohio, there was a nonsignificant trend towards the minimum wage making things a little worse.

And the Science Czar says: awesome! That solves that debate, minimum wage is bad, let’s move on to investigating nominal GDP targeting.

Now it looks like the Science Czar is just a jerk who’s really against minimum wage. All his knowledge of the standards of scientific rigor are going not towards bettering science, but toward worsering science. He’s not trying to create a revolutionary new scientific regime, he’s taking pot shots.

I see this a lot in medicine. Someone jumps on a new study showing the selenium or chromium or plutonium or whatever cures cancer. It is brought up that no, really, the medical community has investigated this sort of thing before, and it has always been found that it doesn’t.

“Well, maybe the medical community wasn’t investigating it the right way! Maybe the investigators were biased! Maybe they didn’t randomize right! Maybe they used a population unusually susceptible to cancer-getting! Ninety percent of medical studies are wrong! Those twenty experiments showing a lack of effect could be total bunk!”

Yes, maybe these things happened in each of the twenty studies that disagree with you.

Or maybe they happened in the one contrarian study you are getting so excited about.

IV.

The unholy combination of isolated demands for philosophical rigor and isolated demands for scientific rigor is isolated demands for mathematical-statistical-conceptual rigor, ie the sort of thing this blog has been talking about all week.

I have already been made fun of for how many different things I am metaphorically comparing IQ to – speed, blood pressure, comas – so I guess it can't hurt to add another example I only thought of today. How about crime? It's usually measured by crime rate – a made-up statistic that combines subfactors like arson (maybe higher when fire insurance pays out better), property damage (maybe higher during periods of ethnic tension and frequent riots) and theft (maybe higher when income inequality is worse). There is assumed to be a General Factor Of Crime (presumably caused by things like poor policing, dark alleys, broken families, et cetera) but I would be extremely surprised if anyone had ever proven Beyond A Shadow Of A Doubt that the factor analysis works out here.

When Cosma Shalizi says he's not sure about the factor analysis in IQ, I have no quarrel with him, because Cosma Shalizi's response to *everything in the world* is to glare at it for not being sufficiently statistically rigorous.

But when other people are totally happy to talk about speed and blood pressure and comas and the crime rate, and then suddenly switch to a position that we can't talk about IQ *at all* unless we have a perfect factor-analytical proof of its obeying certain statistical rules, then I worry they're just out to steal cows.

Likewise, if someone were to just never acknowledge any sorts of groups of objects except those that could be statistically proven to fall out into absolutely separate clusters in which variance within each cluster is less than variance between clusters, well, at least they would be fun to talk to at dinner parties.

But when people never even begin to question the idea of different cultures but make exacting demands of anyone before they can talk about different races – even though the two ideas are statistically isomorphic – then I think they're just out to steal cows.

So this is another technique for avoiding Eulering – is your interlocutor equally willing to apply their complex mathematical argument to everything else.

I think if I hadn't known anything about Bayesian probability, I would have examined the McGrews' Bayesian argument for the Gospels by seeing if it applied equally well to Mormonism, the control group for Christianity.

V.

The old man stamped his boot in the red dirt, kicking up a tiny cloud of dust. "There's a new sheriff in town," he told them.

"No, I'm pretty sure that's impossible," says Parmenides. "There's no such thing as change, only the appearance thereof."

"Well then," says the old man, "I reckon you won't mind the false illusion of your surroundings appearing to change into a jail cell.” And he took out his six-shooter and held it steady.

“Hold on,” said Thales. “We don’t want any trouble here. All is water, so all we did was steal a little bit of water from people. We can give you some water back, and everything will be even, right?” He gestured to a watering trough for horses on the side of the street, which was full of the stuff.

“Just so long as you don’t mind being sprayed with some very hard water from my squirt gun,” the old man answered, and the six-shooter was pointed at the Milesian now.

“Ha!” said Zeno of Elea. “You don’t scare us. In order to hit Thales, your bullet would have to get halfway to him, then half of the remaining distance, and so on. But that would require an infinite number of steps, therefore it is impossible.”

“Sorry,” said the old man, “I couldn’t hear you because it’s logically impossible for the sound waves encoding your speech to reach my ears.”

“We’re not even the same people as the guys who stole those cattle!” said Heraclitus. “Personal identity is an illusion!”

“Then you won’t mind coming to the courthouse with me,” replied the old man “to help the judge imprison some other people who look just like you.”

The last of them, the tall one, said nothing. He just raised his revolver in a fluid motion and shot at the old man.

The old man saw it coming and jumped out of the way. The air was briefly full of bullets. Bang! Thales went down! Bang

bang! Heraclitus! Bang bang! Parmenides and Zeno. Bang bang bang! The old man was hit in the arm, but still standing. Bang bang bang bang…

It was just the old man and the tall one now. The tall one picked up his gun and fired. Nothing happened. Out of bullets.

The old man smiled wryly, his six-shooter still in his hand.

“I know what you’re thinking. You’re thinking – did he fire six shots, or only five? Well, you’ve got to ask yourself a question – do you feel lucky? Well, do you, punk?”

The tall one didn’t budge. “Man is the measure of all things,” said Protagoras. “If I believe you fired six shots, then by my personal epistemic standards, you fired six shots.”

The old man didn’t say anything.

“You see,” the Sophist continued. “Out of all of them, I alone was truly consistent. They all came up with clever theories, then abandoned them whenever it conflicted with their self-interest. I was more honest. I just said at the beginning that my self-interest determined truth, and so never suffered any temptation to depart from my position.”

The old man took off the bandana covering his face. “Man may be the measure of all things. But I’ve taken *your* measure, Protagoras, and found it wanting.”

“Socrates?!” the Sophist gasped.

“The only truly consistent people are the dead, Protagoras,” he said – and squeezed the trigger.

Fifty Swifties

[see: Wikipedia: Tom Swifties *and* Tom Swifties Written By An Author Willing To Go To Any Lengths To Make A Tom Swifty Thus Resulting In Constructions That Often Require More Work For Readers Than For The Author*. All of the below are AFAIK original to SSC.]*

“Pennies look really different under a microscope,” Tom said magnificently.

“A griffin is a kind of flying lion,” Tom said uproariously

“Our flight path has brought us directly above Yellowstone National Park” Tom said overbearingly.

“Obama absolutely buried Romney in the election!” Tom said intermittently.

“I grew two inches last year!” Tom said ambiguously

“I want to trick one or another rich woman into marrying me so I can steal her fortune,” Tom said consummately

“The auction is now open,” Tom said forbiddingly.

“I deny everything!” Tom said all-knowingly.

“The telegraph network was over capacity, so you’ll have to send your message again” Tom said remorsefully.

“Ender, the Formics have dug themselves into fortifications!” Tom observed trenchantly

“I’m going to miss work for the next few days, I’m stuck doing my civic duty at the courthouse,” Tom said injuriously

“The Zoroastrians seem to control a disproportionate amount of India’s wealth,” Tom said parsimoniously

“Nana seems to be developing Tourette syndrome,” Tom said grammatically.

“This prison will be the perfect place for my unethical Human Centipede style experiments,” Tom said confusingly.

“The bounty hunter was my favorite character in Star Wars,” Tom said prophetically.

“Brutha, the Great God is not just a turtle, but also within the hearts of all mankind,” Tom said ominously.

“Chelsea Manning mailed me one of her teeth,” Tom said transcendentally.

“I got a job producing another season of Lassie,” Tom said moronically.

“I was out at the brothel until after midnight,” Tom said scintillatingly.

“I’m sorry, but the lady is spoken for,” Tom said mistakenly.

“I’m going to an all-you-can-eat restaurant tonight,” Tom said forgetfully.

“The girl from my blind date last night was a 4/10,” Tom said metaphorically.

“The medication cured my autism but also made me gain weight,” Tom said fatalistically.

“I used to have to walk everywhere,” Tom said precariously.

“Hades seems like a pretty credible guy,” Tom said disbelievingly.

“That commercial really helped spread awareness of the risks of intelligence explosions,” Tom said admiringly

“To think this entire tree grew from a single nut in just a few years,” Tom said exceedingly quickly

“I got selected for the role of Juan Peron in an Andrew Lloyd Webber musical!” Tom said inevitably

“After playing Juan Peron, no one ever cast me in a play again,” Tom said exactingly

“SAT scores should be given more weight in college admissions,” Tom protested.

“I will start a campaign to convert @aristosophy and @donovanable to Catharism,” Tom would prognosticate.

“The Notorious B.I.G.’s death was a predictable result of his career,” Tom would rhapsodize.

“The zoo’s exhibit on African wildlife was a big disappointment,” Tom said hypocritically

“It’s the Leviathan!” Tom said superficially

“I’m afraid you’ve got an atrial septal defect,” Tom said whole-heartedly

“I first met my wife in the restroom at a bar,” Tom said accommodatingly

“I picked up a nice new casual shirt,” Tom said apologetically

“The cat-goddess is a threat to the American way of life,” Tom said bombastically.

“I can’t figure out how to stop our boat!” Tom said cantankerously

“Satan is the original source of evil,” Tom said urbanely

“I can slay the Jabberwock,” Tom said demonstrably

“I realize I missed the meeting by two whole hours,” Tom said isolatedly

“Kosher kitchens need separate plates for milk and meat,” Tom said judiciously

“Bill Clinton got a divorce!” Tom said exhilarated

“Lower-ranked demons can kiss my ass,” Tom said imprudently

“Help, I got stuck inside this cattle pen,” Tom said inoffensively

“My throne sits on the floor,” Tom said lackadaisically

“It costs one thirty cent stamp to send one letter,” Tom said permissively

“I forbid you to take the ladder to the topmost room of my house,” Tom said anti-climatically.

“Ho ho fucking ho!” customarily.

Burdens

[Content note: Suicide. May be guilt-inducing for people who feel like burdens. All patient characteristics have been heavily obfuscated to protect confidentiality.]

The DSM lists nine criteria for major depressive disorder, of which the seventh is "feelings of worthlessness or excessive or inappropriate guilt".

There are a lot of dumb diagnostic debates over which criteria are "more important" or "more fundamental", and for me there's always been something special about criterion seven. People get depressed over all sorts of things. But when they're actively suicidal, the people who aren't just gesturing for help but totally set on it, they always say one thing:

"I feel like I'm a burden".

Depression is in part a disease of distorted cognitions, a failure of rationality. I had one patient who worked for GM, very smart guy, invented a lot of safety features for cars. He was probably actively saving a bunch of people's lives every time he checked in at the office, and he still felt like he was worthless, a burden, that he was just draining resources that could better be used for someone else.

In cases like these, you can do a little bit of good just by teaching people the fundamental lesson of rationality: that you can't always trust your brain. If your System I is telling you that you're a worthless burden, it could be because you're a worthless burden, or it could be because System I is broken. If System I is broken, you need to call in System II to route around the distorted cognition so you can understand at least on an intellectual level that you're wrong. Once you understand you’re wrong on an intellectual level, you can do what you need to do to make it sink in on a practical level as well – which starts with not killing yourself.

As sad as it was, Robin Williams’ suicide has actually been sort of helpful for me. For the past few days, I’ve tried telling these sorts of people that Robin Williams brightened the lives of millions of people, was a truly great man – and his brain *still* kept telling him he didn’t deserve to live. So maybe depressed brains *are not the most trustworthy arbiters on these sorts of issues*.

This sort of supportive psychotherapy (ie “psychotherapy you make up as you go along”) can sometimes take people some of the way, and then the medications do the rest.

But sometimes it’s harder than this. I don’t want to say anyone is ever *right* about being a burden, but a lot of the people I see aren’t Oscar-winning actors or even automobile safety engineers. Some people just have no easy outs.

Another patient. 25 year old kid. Had some brain damage a few years ago, now has cognitive problems and poor emotional control. Can’t do a job. Got denied for disability a few times, in accordance with the ancient bureaucratic tradition. Survives on a couple of lesser social programs he got approved for plus occasional charity handouts plus some help from his family. One can trace out an unlikely sequence of events by which his situation might one day improve, but I won’t insult his intelligence by claiming it’s very probable. Now *he* attempts suicide, says he feels like a burden on everyone around him. Well, what am I going to say?

It’s not always people with some obvious disability. Sometimes it’s just alcoholics, or elderly people, or people without the cognitive skills to get a job in today’s economy. They think that they’re taking more from the system than they’re putting in, and in monetary terms they’re probably right.

One common therapeutic strategy here is to talk about how much the patient’s parents/friends/girlfriend/pet hamster love them, how heartbroken they would be if they killed themselves. In the absence of better alternatives, I have used this strategy. I have used it *very grudgingly*, and I’ve always felt dirty afterwards. It always feels like the worst sort of emotional blackmail. Not helping them want to live, just making them feel really guilty about dying. “Sure, you’re a burden if you live, but if you kill yourself, that would make you an even *bigger* burden!” A++ best psychiatrist.

There is something else I’ve never said, because it’s too deeply tied in with my own politics, and not something I would expect anybody else to understand.

And that is: humans don’t owe society anything. We were here first.

If my patient, the one with the brain damage, were back in the Environment of Evolutionary Adaptedness, in a nice tribe with Dunbar’s number of people, there would be no problem.

Maybe his cognitive problems would make him a slightly less proficient hunter than someone else, but whatever, he could always gather.

Maybe his emotional control problems would give him a little bit of a handicap in tribal politics, but he wouldn’t get arrested for making a scene, he wouldn't get fired for not sucking up to his boss enough, he wouldn't be forced to live in a tiny apartment with people he didn't necessarily like who were constantly getting on his nerves. He might get in a fight and end up with a spear through his gut, but in that case his problems would be over anyway.

Otherwise he could just hang out and live in a cave and gather roots and berries and maybe hunt buffalo and participate in the appropriate tribal bonding rituals like everyone else.

But society came and paved over the place where all the roots and berry plants grew and killed the buffalo and dynamited the caves and declared the tribal bonding rituals Problematic. This increased productivity by about a zillion times, so most people ended up better off. The only ones who didn't were the ones who for some reason couldn't participate in it.

(if you're one of those people who sees red every time someone mentions evolution or cavemen, imagine him as a dockworker a hundred years ago, or a peasant farmer a thousand)

Society got where it is by systematically destroying everything that could have supported him and replacing it with things that required skills he didn't have. Of *course* it owes him when he suddenly can't support himself. Think of it as the ultimate use of eminent domain; a power beyond your control has seized everything in the world, it had some good economic reasons for doing so, but it at least owes you compensation!

This is also the basis of my support for a basic income guarantee. Imagine an employment waterline, gradually rising through higher and higher levels of competence. In the distant past, maybe you could be pretty dumb, have no emotional continence at all, and still live a pretty happy life. As the waterline rises, the skills necessary to support yourself comfortably become higher and higher. Right now most people in the US who can't get college degrees – which are really hard to get! – are just barely hanging on, and that is absolutely a new development. Soon enough even some of the college-educated won't be very useful to the system. And so on, until *everyone* is a burden.

(people talk as if the only possible use of information about the determinants of intelligence is to tell low-IQ people they are bad. Maybe they've never felt the desperate need to reassure someone "No, it is not your fault that everything is going wrong for you, everything was rigged against you from the beginning.")

By the time I am a burden – it's possible that I am already, just because I can convince the system to give me money doesn't mean the system is right to do so, but I expect I *certainly* will be one before I die – I would like there to be in place a crystal-clear understanding that *we were here first* and society doesn't get to make us obsolete without owing us something in return.

After that, we will have to predicate our self-worth on something other than being able to "contribute" in the classical sense of the term. Don't get me wrong, I think contributing something is a valuable goal, and one it's important to enforce to prevent free-loaders. But it's a valuable goal at the margins, some people are already heading for the tails, and pretty soon we'll all be stuck there.

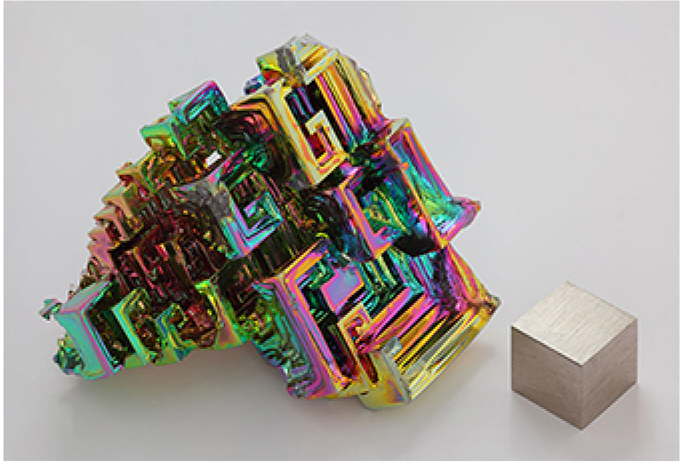
I’m not sure what such a post-contribution value system would look like. It might be based around helping others in less tangible ways, like providing company and cheerfulness and love. It might be a virtue ethics celebrating people unusually good at cultivating traits we value. Or it might be a sort of philosophically-informed hedonism along the lines of Epicurus, where we try to enjoy ourselves in the ways that make us most human.

And I think my advice to my suicidal patients, if I were able and willing to express all this to them, would be to stop worrying about being a burden and to start doing all these things now.

An Iron Curtain Has Descended Upon Psychopharmacology

Imagine if a chemist told you offhandedly that the Russians had different chemical elements than we did.

Here in America, we use elements like lithium and silicon and bismuth. We have figured out lots of neat compounds we can make with these elements. We've also figured out useful technological applications. Lithium makes batteries. Silicon makes computer chips. Bismuth makes pretty gifs you can post on Tumblr.



The Russians don't use any of these. They have their own Russian elements on their own Russian periodic table, with long Russian names you can't pronounce. Apparently some of these also have useful technological applications. One of them is a room temperature superconductor. Another improves the efficiency of dirigibles by 500% for some reason.

No one in America seems remotely interested in any of these Russian elements. Many American chemists don't even know they exist, even though each element has its own English-language Wikipedia page. When informed, they just say “Yeah, the Russians have *lots* of stuff,” and leave it at that.

American research teams pour millions of dollars into synthesizing novel elements in order to expand their periodic tables and the number of useful compounds they can make. If anyone suggests importing and studying some of the Russian elements, the chemists say “Huh, that never occurred to us, maybe someone else should do it,” and go back to spending millions of dollars synthesizing entirely novel atoms.

If a chemist told you this, you would think they were crazy. Science, you would say, is science everywhere. You can’t have one set of elements in Russia and another in the US, everyone would work together and compare notes. At the very least one side would have the common decency to at least steal from the other. No way anything like this could possibly go on.

But as far as I can tell this is exactly the state of modern psychopharmacology.

Consider anxiety. I would *kill* for a good anti-anxiety drug. Right now my choices are pretty limited. Benzodiazepines and barbituate work great but are addictive and dangerous. SSRIs work okay but need a month to take effect. Neurontin, Vistaril, and Buspar are safe, fast-acting, and totally ineffective. And Lyrica is expensive and off-label. As a result, a lot of my anxious patients tend to stay anxious.

Any textbook, database, or lecture you care to check on anti-anxiety medications will list the ones I just listed above plus a couple of others I’m forgetting.

But if you look the matter up on Wikipedia, you see all these weird names like mebicarum, afobazole, selank, bromantane, emoxypine, validolum, and picamilon. You can show these names to your psychiatrist and she will have no idea what you're talking about, think you're speaking nonsense syllables. You can show them to the professor of psychopharmacology at a major university and your chances are maybe like 50-50.

These are the Russian anti-anxiety drugs. They seem to have pretty good evidential support. Wikipedia's bromantane article gives a bunch of studies of bromantane in the footnotes, including a randomized controlled trial in the forbiddingly named *Zh Nevrol Psikhiatr Im S S Korsakova*.

And look what else Wikipedia's bromantane article says:

> Study results suggest that the combination of psychostimulant and anxiolytic actions in the spectrum of psychotropic activity of bromantane is effective in treating asthenic disorders compared to placebo. It is considered novel having both stimulant and anti-anxiety properties.

Imagine reading about a Russian element on Wikipedia, and at the end there's this paragraph saying "By the way, this element inverts gravity and has to be tied to the ground to prevent it from falling upwards". An anxiolytic stimulant is *really really cool*. But somehow generations of American psychopharmacologists must have read about bromantane and thought "No, I don't think I'll pay any more attention to *that*."

My guess is the reason we can't prescribe bromantane is the same reason we can't prescribe melatonin and we can't prescribe fish oil without the charade of calling it LOVAZA™®©. The FDA won’t approve a treatment unless some drug company has invested a billion dollars in doing a lot of studies about it. It doesn’t count if some foreign scientists already did a bunch of studies. It doesn’t count if millions of Russians have been using the drug for decades and are by and large still alive. You’ve got to have the entire thing analyzed by the FDA and then rejected at the last second without explanation (yes, I have just been reading Marginal Revolution’s review of *Innovation Breakdown: How the FDA and Wall Street Cripple Medical Advances*; I do need to check out the actual book). Absent an extremely strong patent on the drug there’s no reason a drug company would want to go forward with all of this. I don’t know what the legalities of buying Russian drug rights from Russian companies are, but I expect they’re complicated and that pharmaceutical companies have made a reasoned decision not to bother.

Given this situation, it’s perfectly reasonable for doctors not to prescribe them. Certainly *I* don’t plan to prescribe any Russian drugs when I get my own practice. Imagine if a patient gets liver failure on one – and remember that people are getting liver failure all the time for random reasons. The patient’s family decides to sue and I’m stuck defending my decision in court. “Yes, Your Honor, I admit I told the deceased to buy a medication no other psychiatrist in the state has ever heard of from a sketchy online Russian pharmacy. But in my defense, there was a study supporting its use in *Zh Nevrol Psikhiatr Im S S Korsakova*. Which I didn’t read, because I don’t speak Russian.”

Everyone follows their own incentives perfectly, and as a result the system as a whole does something insane. Classic multipolar trap.

Luckily, this hasn’t stopped a lively gray market trade in these chemicals, which I totally one hundred percent approve of. Noopept, for example, is a prescription drug in Russia but is sold over-the-counter by online suppliers here. You can even get some bromantane for two bucks a pill.

Don’t worry. I’m sure these people are on the level. How could a site with a background like that possibly be unreliable?

Ozy’s Anti Heartiste FAQ

[Editor’s note: This is a guest post by Ozy Frantz. I do not necessarily endorse everything it says, but I do contingently agree with a lot of it. Content note for profanity, social justice stuff, manosphere stuff, and graphic descriptions of sex. I trust that any debate this kicks up will be marked by courtesy and good manners on all sides, in a spirit of sincere collaborative truth-seeking – SA]

I. Introduction

A. What is the purpose of this post?

This is the Anti-Heartiste FAQ. It is meant to rebut some common beliefs within the “manosphere” about how human sexual interaction works. I am primarily arguing with the blogger Heartiste, as he is one of the most famous and influential writers within the manosphere, although I do briefly argue with other writers. I am also going to ignore the macro-level beliefs about how human society works, on the grounds that they are mostly derived from these micro-level observations about human sexuality and fall down when no longer grounded in them.

A1. What does Heartiste believe about human sexual interaction?

Men are primarily attracted to women who are young, thin, and hot; this accounts for approximately 90% of men’s criteria for dating women. The other 10% is femininity, sexual voraciousness, and non-promiscuity. Women have a dual sex drive, often referred to as “alpha fucks, beta bucks.” Women are primarily sexually attracted to “alpha males,” who are assholes, conventionally masculine, and popular among other women. When they are young and hot enough to hope to secure commitment from alphas or when they are ovulating and thus capable of having children, women pursue “alpha fucks.” Beta males are generally middle-class or above and do not have alpha male traits. Women seek betas when they are older or less attractive, or when they aren’t ovulating, in order to secure commitment and extract resources. Ideally, women want to get pregnant by alpha males and make betas take care of the child. There are also omega males, who are undesirable to women for any purpose.

B. I am Heartiste, and I don’t actually want to go to all the bother of reading this whole post, I just want to know what your characteristics are so I can direct ad hominem attacks at you properly.

I am 22 years old. I am 5’9″ and my weight fluctuates between 120 and 126 pounds. I have 32A breasts and a waist-to-hip ratio of .7. Pictures are available here. I have had 30 sexual partners. I am polyamorous, which means I openly and honestly date multiple people at the same time. I have been diagnosed with borderline personality disorder, social phobia, and generalized anxiety disorder. I have a degree in sociology and gender studies. I am currently a camgirl, which means I take off my clothes and masturbate for money on the Internet. I live with my primary partner Scott; I am also dating Esther, who when asked to describe herself for this essay called herself a “sad weird fat girl with incredible boobs”.

I am a nonbinary trans person, but I was born with a vagina, a female-typical hormone balance, and as far as I am aware two X chromosomes. I currently possess all of these traits and will probably do so indefinitely. That means I am not a trans woman. I am heading in the totally opposite direction than trans women are. I cannot possibly be transitioning because of my autogynephilia, because if I were an autogynephile I would be like "wow, I fetishize having the body parts I was born with, this is incredibly convenient!" instead of being like "aaaaa! Get them off me!" There has previously been confusion about this, so I am making it as clear as possible.

C. You're just one of those feminist game deniers who think sex differences don't exist.

Nope! Actually, I am going to take this FAQ from the POV that game tactics all work exactly as stated by their advocates. What I hope to show is that even if you assume that Heartiste is totally right about his practical advice– if you should neg, qualify, do takeaways, send women 'gay' as a response to their text messages, and all the rest– his conclusions about how the dating marketplace works simply do not follow.

In addition, I do actually think that sex differences exist. I think that level of violence, libido, probably certain aspects of sexuality such as getting off on narrative versus getting off on visuals, and probably some stuff with emotions are all inherent, biological gender differences. I am not willing to rule out other differences nor to state that I believe in them.

It is a very boring prediction of feminist theory that the genders behave differently from each other. After all, if you divide children into two random groups and tell half of them to be courageous and half of them to be gentle, one half is probably going to be courageous and the other is going to be gentle. It is tiresome when people (both feminists and antifeminists) pretend that the existence of gender differences proves that these gender differences are not socialized. Over the course of this article, I will highlight legitimate gender differences which I think are probably social in nature rather than biological. Of course, all complex behavior has both social and biological elements: for instance, I will argue that women's lack of interest in casual sex has both a biological and a social component.

II. Men

A. Beauty

1. Beauty norms have to be inborn on account of it feels like I have no volitional control over my boner.

First, by the time you have your first boner at puberty, you have already experienced more than a decade of environmental influence. You've learned a language, how to walk and use the toilet, fashion, how to maintain a conversation, what interests are approved of in your peer group. Most of those things are now second nature to you. Is it that impossible that you learned a sexuality too?

Second, environmental influence can make things happen that feel like you can't control them. Music sounds good to you because you've learned how to listen to it (just think about your mom saying "that's just noise!") but it doesn't feel learned that it makes you want to get up and boogie. Girls like pink because we've associated pink with girls (it was different as few as a hundred years ago), but try to argue with a six-year-old at a toy store. At the extreme end, PTSD is obviously the result of the environment– you need a traumatic event to trigger it– but flashbacks and hyperarousal often feel inevitable and uncontrollable. Indeed, many fetishes are obviously environmental in origin, unless one assumes that genes for fetishes for rubber, stuffed animals, and Catholic schoolgirls lay latent in the genome for tens of thousands of years.

2. But I and my friends all agree who the hottest girls at the bar are.

If I and my friends all went to a bar and started talking about who the hottest girl was, we would probably agree it was a girl with no makeup, a Star Wars T-shirt, and glasses. I suspect this is a sign that friends tend to be similar to each other, not a sign that beauty norms are inborn.

coughcoughtypicalmindfallacycough

3. Are beauty norms a cultural universal?

In China, women bound their feet, crippling them. Love poetry was written about the beauty of the “lotus foot,” which was as small as three inches, and the swaying way a woman walked. (See, for instance, this collection of Chinese erotic poems.) In Mauritania, women are sent to camps where they are force-fed tens of thousands of calories a day in order to become attractively obese. In Renaissance England, women painted their faces white with lead, poisoning them. Oh, and one for the gay men: in ancient Greece small penises were considered heroic and manly and large penises laughable.

It seems to me you have three options here. First, you can agree that beauty norms are to a degree culturally variable, including our own. Second, you can believe that a bunch of people crippled their daughters and wrote love poetry about how beautiful it is just to fuck with Westerners or something. Third, you can say that the Chinese were mistaken for a thousand years about evolution wanting them to get boners for girls with bound feet, and mysteriously 21st-century Americans are the only culture who has figured out the correct evolved sexuality.

2. Oh, so you’re one of those people who thinks that we only believe in beauty norms because the evil patriarchy is brainwashing us, and if we didn’t we’d think that every body is beautiful.

There are actually options between “all ideas of beauty are inborn” and “all ideas of beauty are socially constructed.” For instance, one could say that our ideas of beauty are the product of biology and environment working together. You know, the way every other complex behavior is.

I do think some beauty standards are probably inborn and fairly immutable. Symmetrical faces are not simply preferred in the US and the UK: Chinese and Japanese people prefer symmetrical faces, as do the Hadza hunter-gatherer tribe. (In fact, the Hadza prefer symmetrical faces more strongly than Americans do!) This is some evidence that humans ‘naturally’ prefer symmetrical faces.

Another prominent example is waist-to-hip ratio. Initially, the evidence for WHR as a cross-cultural universal seems pretty slim. Kenyans prefer a .7 WHR, the same as modern Americans. However, Ugandans prefer a .5 WHR, significantly lower than Americans. Malaysians don’t care about waist-to-hip ratio, only caring about BMI, as do the Japanese. Hunter-gatherers seem to not care about waist to hip ratio. However, replications of the latter study suggest that when the hunter-gatherer preference for heavier women is controlled for, waist-to-hip ratio has an independent effect. I personally find the latter study incredibly interesting and think it has a lot of potential to save WHR; however, it has a very small sample size, so take it with all relevant grains of salt. Nevertheless, it provides some suggestion that WHR is culturally invariant. (I would also like to point out that “hunter gatherers like fat women, which confounds tests for WHR and attractiveness designed for thin-preferring Americans" is also a pretty solid point against the Fat Is Universally Hideous hypothesis.)

However, I think most beauty standards are the result of a more complex interaction. (Here is the wild-ass guessing with no empirical backing section.) For instance, men might be attracted to features that are familiar to them, high-status in their culture, or associated with their parents or other people they love. In Song China, women with bound feet are high-status; in modern America, thin women with large breasts are high-status, and thinness indicates other high-status traits such as wealth and free time to attend to one's health. A similar origin produces wildly different beauty standards. Similarly, men might default to a certain preference– maybe they like brunettes– but this is culturally mutable– if they're exposed to a lot of sexy blonde women and posters saying Blondes Have More Fun, they'll start liking blondes. Alternately, the "sexy daughters" hypothesis (a variant of a Heartiste favorite, the "sexy sons" hypothesis): men want their daughters to be beautiful so they can get better mates, so they choose partners who fit the local standard of beauty.

3. Are pretty girls healthier?

Studies do not seem to have found meaningful correlations between rated facial attractiveness and physical health. While the second study has some correlations, they are small and fail to reach significance and are totally hyped by their abstract.

This is very odd, of course, because as I said earlier the evidence points towards symmetrical faces being a cross-cultural beauty standard. There are several possible explanations. Symmetrical faces might indicate health in the environment of evolutionary adaptedness, but industrialized countries have improved health. (I believe one common hypothesis is that symmetrical faces show you aren’t infected by parasites.) Maybe humans like symmetry for other reasons, and that generalizes to sexual attraction. Maybe it’s some other cause (sexy daughters hypothesis!).

Alternately, maybe facial symmetry does indicate something interesting about health, but that’s confounded by all the other factors our culture has incorporated into facial attractiveness. This study suggests that current health and facial symmetry are basically uncorrelated, but facial symmetry is correlated with childhood health. Another study argues that facial asymmetry may be related to developmental instability, which is negative. This is interesting but tentative.

4. Does it make evolutionary sense for men to target pretty girls for casual sex?

Let’s think that you’re a man in the environment of evolutionary adaptedness. Human babies are extremely resource-intensive. It is generally not possible to raise a child by oneself: a woman alone will either realize this and commit infanticide, or risk both herself and her child starving to death. If you are engaging in long-term mating, then you are going to invest resources in the baby. However, if you’re engaging in short-term mating, you’re not. If your child dies, all the effort you put in and risks you took for the the short-term mating strategy are lost. Therefore, men ought to have evolved to be attracted to women with sufficient resources that they could take care of a child: women with doting and wealthy fathers, women with husbands, women who kicked ass so hard at gathering they could feed themselves and a child, women with unique skills at, I don’t know, flintknapping or shamanism or something such that people will give them food. Therefore, in the modern environment, men ought to be into rich or married women and be willing to sacrifice a considerable amount of prettiness for wealth. Only in long-term relationships should they begin to care about prettiness and whatever that indicates, since they can make up for lower resources on the part of the woman. Conversely, women who are engaging in short-term mating ought to be primarily interested in pretty dudes who will pass on good genes to their children.

I am not saying that this is an actually accurate model of how human evolution worked. I am just pointing out that this evolutionary just-so story is exactly as plausible as all of Heartiste's evolutionary just-so stories and has opposite predictions, and that I feel like the reason that no one is coming up with that model is less related to plausible speculation about evolution and more about "shit, we have to explain why humans are genetically programmed to act the way I do."

5. Is 95% of what men want in a relationship partner youth, beauty, and thinness?

Man, if that were true why do I keep getting hit on online by dudes who are uncertain about whether my sex is female?

(For the record, I continue to be female-sexed in pretty much every conceivable way. I just felt like I ought to repeat that, because of the amount of confusion in neoreactionary circles about the "what genitals does Ozy have?" point.)

But that is not really excellent evidence! It is possible that most of the dudes who hit on me online are Boner Georg who is an outlier adn should not be counted. (In fact, that is certain.) So let's look at the studies!

If Heartiste had said that men are more interested in physical appearance, in general, than women are, I would basically have nothing to say here. That seems pretty much correct. However, there is an important difference between “men care about appearance more than women do” and “physical appearance is 95% of men’s criteria in selecting a romantic partner.” The latter is a much stronger claim and not remotely backed up by the evidence.

The most obvious way to find out whether men are 95% interested in physical appearance is to ask them. This study with an extremely large sample size found that men did rank good looks and facial attractiveness as more important than women did. However, across genders, the most important traits were all non-physical: things like intelligence, values, and communication skills.

On the other hand, it seems like asking people to rank their top three traits in a sexual partner is a strategy that will lead to a lot of social desirability bias. Surely if we examine studies that relate to social behavior– speed dating, correlations of attractiveness to how popular someone is among the opposite sex or how liked they are by someone they’re interacting with, experiments that manipulate the attractiveness of a stranger– we will find that the difference is much much larger. Perhaps 95%?

Oh, the sex differences are actually smaller in studies that examine social behavior. That’s awkward.

In fact, this unpaywalled study which examines speed-dating finds that there is basically no gender difference between men and women in how much physical attractiveness affects saying ‘yes’ to someone in speed-dating.

I believe the correct response here is “<3Science<3 has once again dropped to its knees and slobbered the knob of

Ozy, vindicating the Ozyan observation that physical appearance is not ninety-five percent of what men care about, you dumb fuck.“

Interestingly, gender equality seems to make the gender differences in self-reported desire for a physically attractive partner stay the same, while it makes gender differences in self-reported desire for a partner with a particular personality go down. Which suggests that… men are biologically programmed to claim to be into hot women and not actually be? Or something? Very puzzling.

Again, I am not saying there is no difference between how much men prioritize physical appearance and how much women prioritize physical appearance! There clearly seems to be a difference. What I am saying is that it is not true that physical appearance is 95% of what men care about and 5% of what women care about. That is probably true of Heartiste, but that is not true of all men. I would like to thank Heartiste for making extreme claims and thus making my debunking of them easier.

B. Thinness

1. Aren’t all men in all cultures into thin women?

In addition to Mauritania and hunter gatherers, discussed above, I would like to discuss classical art. Heartiste has said that Rubens is a fatty-fucker so we’re not allowed to use him as an example. Fortunately, a feminist blogger has created a helpful “if classic works of art were photoshopped to look like modern magazine stars” set of pictures, from throughout European art history, without a Rubens in sight.

Now, to be clear, these women are mostly overweight or the thin end of obese. (The BMI Project has helpful examples of what overweight and obese women look like; most people think obese looks significantly fatter than it actually does.) They are not the sort of person that one associates with “fat woman.” However, Heartiste has been more than willing to call Lena Dunham, who is about the same size as those women, “a frumpy, dumpy, plumpy formless flesh entity”. In addition, he tends to use phrases like “the distended porcine holes of these beached whales” to talk about overweight women. So at the very least we have cultures that are really into naked women who look significantly fatter than Heartiste’s boner prefers.

2. Okay, maybe not cross-culturally, but in our culture aren’t all men most attracted to thin women?

Meet April Flores.

April Flores has been in 35 porn films. (Also she is cute as a button, but that is neither here nor there.) You can, if you wish, buy a plastic version of April Flores’s vagina, stomach, and thighs. This means, for those of us at home, that both the “into extremely fat ladies” market and the “into creepily disembodied female body parts” market are large enough, and overlap enough, that it is worth it to porn companies to create a toy directly aimed at it. Or it could be because of the noted political correctness of the industry that brought us Assault That Asian Ass IV.

3. But the only people who are into April Flores are chubby chaser omega dweebs.

So? They still exist.

I suppose you could steelman this argument into “omega dweebs are so insecure they even jerk off to fat women they’re not attracted to.” However, there seem to be many other omega dweebs who are perfectly capable of jerking off to, say, skinny anime girls with breasts the size of their heads (indeed, this is a proverbial omega dweeb thing to do). In addition, the behavior of being so insecure you jerk off to people you're not attracted to seems quite ludicrous and I would like some evidence that that actually happens ever.

I argue in part IIIA1 that female serial killer chasing is an outlier that doesn't show anything about female sexuality as a whole. One could make a similar argument about male chubby chasers. However, I notice that BBWs are a healthy section of the porn market, while serial-killer-themed porn-for-ladies has made less of an inroad than porn featuring men with dog dicks.

4. Isn't being into really thin girls evolutionarily selected for?

In women, being underweight (i.e. having a BMI of less than 18.5) may cause irregular periods, lack of periods, difficulty getting pregnant, and infertility. Heartiste's range of optimally attractive female BMI is 17.5-21, which includes some underweight women. He considers BMI of 21-24, which is on the large end of perfectly fertile, to be equally attractive to BMI of 16.6-17.5, which consists of women at serious risk of reduced fertility or infertility. This seems an unlikely trait to evolve.

It is true that being overweight or obese is also correlated with infertility. However, one must remember that for most of human history starving to death is more of a concern than death by obesity. Therefore, if anything, it seems that evolution would favor overweight women rather than underweight women– the infertility risk is outweighed by her not starving to death while breastfeeding your child.

D. Sluttiness

1. Are sluts more likely to divorce you?

Scott argues no. Several reactionaries, such as Free Northerner, have presented other charts which purport to show that it is. My position is the centrist "holy shit this data is confounded all to hell."

Virgins and women with one partner have traditional religious values (being a virgin at marriage is a much more costly signal that you genuinely have traditional values than, say, signing an abstinence pledge or being in a red state). Women with 20+ partners, conversely, probably have libertine values and don't think divorce is immoral. Women who lost their virginity under age 14 are usually what is more properly referred to as "rape survivors," and it is not surprising if that's linked with poor mental health and thus marital instability. Women with more premarital partners are typically older at first marriage, which is linked to stability; similarly, since you typically have sex before or when you get married, women who lost their virginity at a later date are more likely to get married older. I do not have access to the data set, but I predict that if you control for age at first marriage and traditional/libertine values, the predictive value of sluttiness will be basically zero.

In addition, if you are going to be married to someone, you have access to much better information about whether she'd make a good romantic partner for you. You have access to information about her character across a wide variety of situations: you can find out whether she's kind, generous, loyal, honest. If you are interested in minimizing your risk of divorce, you can see how stable and long-lasting her friendships are, how reluctant she is to leave a relationship, whether she condemns divorce or supports it, and (perhaps more crucially) how often her words about morality line up with her actions. It is trivial to lie about your number of sexual partners. It is not trivial to lie about your entire character.

2. Is it easy to identify a slut?

"It's Easy to Identify A Slut" is the title of one of Heartiste's most popular posts. He says that it reveals the secret slut tells that prove that no woman can hide how many people she's had sex with.

Unfortunately, it mostly seems like advice for identifying a woman who is telegraphing her high sociosexuality as hard as she can. I am not remotely interested in hiding that I've had lots of sex partners, but if I were, here is a brief list of things I would not do: go commando; wear cleavage-bearing tops; seem really really good at sex; show off my sex toy and porn collection; talk about sex constantly; implore my partners for kinky sex; brag about how much sex I've had (!); have so many sex partners I'm never single (!!); make out with men in bars (!!!); let men snort coke off my ass (!!!!)

I mean, Jesus, Heartiste, what does she have to do to not be trying to hide that she's a slut, wear a shirt that says "MY BOYFRIEND SAID I NEEDED TO BE MORE AFFECTIONATE, SO NOW I HAVE TWO"?

3. Is it true that some humans are r-selected, which means they're promiscuous and want to have lots of children, and some humans are k-selected, which means they invest in a few children?

No.

For one thing, r/k selection theory was shown to be a vast oversimplification of life-history evolution forty years ago. For another thing, it applies on a species level, not on an individual level. Humans as a species are absurdly k-selected. How many species can you think of that have young that are totally helpless for their first decade of life? r/k selection theory would argue that we wouldn't have casual sex, because humans are not remotely r-selected on any conceivable level! Clearly, this is not the case, which ought to make one wonder how applicable this is to human beings anyway.

4. Is it a good idea to hate sluts?

OK, let's say that marrying a slut increases your likelihood of getting divorced. As much as I hate to point this out, Heartiste, you don't actually want to get married. You want to have lots of casual sex. And as much as you claim that you think sluts are valuable as good-time girls and virgins are valuable as women to marry and they are both equal, you don't have any tag on your blog called "Virgins" where you recommend "a stone bold [virgin]… should be shunned by everyone, including the media, to live out her diseased days alone and isolated from normal human contact… what worse fate for the BPD attention whoring sociopathic [virgin] than being utterly ignored?" As far as I can tell, the only result of hating sluts for Heartiste is that he gets to spend a lot of time having sex with women he finds disgusting and gross, which seems like a pretty depressing life choice to me.

E. Femininity

1. Are men more interested in women with feminine personalities than women with masculine personalities?

Depends on the man.

I am going to say “feminine women” and “masculine women” from now on, because “women with feminine personalities” is hard to type. However, I would like to clarify that I am not talking about wearing lipstick or about having feminine facial features, but about possessing gender-congruent personality traits, such as being gentle, tactful, and a lover of children for women and being ambitious, competitive, and individualistic for men.

This study finds that men tend to be attracted to feminine women for romance and androgynous women platonically. That seems pretty plausibly pro-Heartiste.

However, men who believe in traditional gender roles tend to be more attracted to feminine women, while men who don’t tend to not care whether their partner is feminine or not. In this study, men who believe in traditional gender roles prefer feminine partners, while men who prefer either androgynous or feminine partners. Interestingly, this study finds that men who aren’t very masculine prefer women who aren’t very feminine, while masculine men don’t care about femininity one way or the other, suggesting the relevant variable is not masculinity as a personality trait but belief in traditional gender roles. Together, these studies suggest that the cause is not inherent biology: it’s the fact that people who support traditional gender roles tend to be into traditional gender roles in their personal life too.

2. Except that those studies don’t show that men prefer androgynous women: in fact, guys are pretty much okay with feminine women regardless. Maybe men evolved to like feminine women and not believing in traditional gender roles socializes people to like androgynous women, or at least to say that on surveys.

Except look at women’s preferences in those studies: women consistently don’t like masculine men either. In fact, even traditional women don’t really like masculine men. I think what we’re seeing here is an interaction: traditional men may favor feminine women and, in addition, people with masculine personalities, as defined by the Bem Sex Role Inventory, are assholes and no one wants to put up with them. I mean, do you want to date someone described as aggressive, dominant, and a “strong personality” and as not understanding, sensitive to others’ needs, and compassionate? I don’t. They would probably make me cry all the time, it would be awful.

III. Women

This is an incredibly cheap shot and I feel sort of guilty about putting it in, so: I would like to point out that Heartiste seems to believe some women don’t have leg hair (question 11) and that breasts come in DD and perky without wearing a bra (questions 4 and 5). In addition, he seems to be under the impression that real skinny women carry a Photoshop filter in front of them at all times. Seriously, as someone who can see that kind of body every time I look in the mirror, the “spindly, weirdly proportioned girl” is what actual thin women actually look like. I would propose that Heartiste has perhaps never seen a non-Photoshopped naked thin woman, but perhaps the more charitable explanation is that whenever Heartiste has seen a thin woman naked he has had a Photoshop filter in the form of an erection. (This is not charitable to Heartiste. It is charitable to male virgins, most of whom are totally aware that DD breasts sag.)

So, you know, just be aware you’re taking advice about women from someone who thinks some women don’t have leg hair.

A. Jerks

1. Do women like serial killers?

The Wikipedia list of serial killers lists about 100 serial killers in the United States; let's double it and say 200 just to bias it towards Heartiste. Let us assume that about one in every two thousand women is interested in serial killers. That means that there are 750 women interested in serial killers for every serial killer. I don't know about you, but I feel like having 750 women interested in you would put you in a pretty good sexual marketplace position. I also feel like one in two thousand is also a fair estimate of the prevalence of, say, diaper fetishism. Since I do not say that diaper fetishists prove any deep truths about human sexuality, I can also argue that serial killer fetishism doesn't.

It would be interesting if only women decided to date murderers. However, women are also much less likely than men to commit high-profile murders. The last high-profile female murderer I'm aware of– Casey Anthony– was deluged with marriage proposals from men.

2. What about Genghis Khan?

The most genetically successful man in history is Genghis Khan (actually his grandfather a few generations back). Heartiste has argued in the past that this is proof that women are attracted to violent men. However, I feel like Genghis Khan's reproductive success is more than adequately explained by the fact that for several hundred years a woman whom a relative of Genghis Khan wished to fuck had a choice between fucking him or being dead. I don't know about you, but I would fuck a lot of people I wasn't attracted to if the other option was death.

3. What about women in abusive relationships?

It is true that women are more likely to be hit by their partners, more likely to be raped, more likely to be sent to the hospital due to domestic abuse, and more likely to be murdered. However, a statistically indistinguishable number of men and women– approximately 50%– are victims of emotional abuse in the United States.

That isn't "women like jerks." That is "people end up in abusive relationships." Male abusers are more likely to be violent– but men are more violent in every sort of relationship, from high schools (boys get in fights, girls pull some Mean Girls shit) to crimes (men are far more likely to commit violent crimes). If desiring abuse is a fact about sexuality, it is a fact about everyone's sexuality.

3a. Maybe men stay in abusive relationships or try to date serial killers because they're desperate omegas, and women do it because they find that sort of thing attractive.

If women were into murderers and men weren't, you would take that as evidence that there was some interesting difference going on there. (As you should.) If women and men are both into murderers, then you cannot just decide that there is a gender difference anyway– that men are into murderers because they are so desperate they can't get anyone else, and women are into murderers because it taps into their deep-seated evolutionary desires. At least, you can't without solid evidence which Heartiste has yet to actually present: for instance, a study of serial killer groupies which shows that male serial killer groupies are typically less sexually successful than female serial killer groupies. For further elaboration, please see conservation of expected evidence.

3b. Do people stay in abusive relationships because they are attracted to abusers?

No. Let us consider this from two angles: first, let us assume that it is possible to manipulate people. Think about how a really good salesperson can get you to buy something you don’t want. Second, let us assume that abusers are sensible people who don’t want to go through all the bother of finding a new partner every time someone says “holy shit, you just called me a whore, GTFO.”

As a reasonable abuser, you want to make sure your partner will stay. What do you do? Well, you’re charming, kind, and romantic at first, so they get really invested in the relationship. You target people who are vulnerable to begin with, such as a disabled person or an undocumented immigrant or someone without very many friends. You try to make your partner dependent on you (for money, transportation, etc.) so they can’t leave. Rats will keep pressing a button that gives them intermittent rewards (this is the psychology behind gambling and World of Warcraft); you take a similar tack towards giving your partner love and affection. You fuck with their idea of what is normal– if you can convince them that you treat them the way anyone would treat them, or the way they deserve to be treated, or better than they deserve, they won’t leave. You isolate them, both so they don’t have anyone to point out that you’re treating them like shit and so they don’t have anyone to rescue them. You appear sad and broken, so they feel compassion for you and want to keep taking care of you. If worst comes to worst, you threaten to kill them or yourself if you break up with them.

Therefore, staying with an abuser does not necessarily mean you want to be abused. Going back to an abuser does not necessarily mean you want to be abused. It means the abuser is good at abusing people.

Here is a list of reasons people stay in abusive relationships. You will notice “they turn me on” is not on there.

3b. Okay, but some people end up in abusive relationships again and again.

In fact, abuse survivors are a really excellent group to target for our rational abuser! A lot of the traits that make someone a good target for being abused remain the same: if someone is disabled, they are probably going to stay disabled even after they break up with their abuser. Abuse survivors themselves are often pretty vulnerable. Abusers are often very charming and high-status, so when the abuse survivor is like “that person abused me,” a lot of people will go “but she is so nice! Clearly you are psycho and making it up,” which means they lose their entire social group. And if you decide to abuse an abuse survivor, someone else has already isolated them and fucked with their idea of what is normal, so you have to do a lot less work. They practically come pre-abused!

For some people– usually people who have been abused, but not always– abuse is normalized. That means that they believe that all relationships are abusive. This makes them excellent targets for abusers. After all, if you believe the abuse is normal, you are not going to do silly things like “object” or “leave.” In fact, if someone is not abusing them, they may wonder about whether he really loves them. This is very sad, but it is not the same thing as wanting to be abused. If you think all food tastes like Ensure, you are probably not going to go seek out food that doesn’t taste like Ensure. That does not mean you have an all-abiding passion for Ensure.

4. What about women with rape fantasies?

I like thinking about zombies (as a subset of my fondness for thinking about apocalypses). I have a well-thumbed copy of the Zombie Survival Guide. I spend a lot of time working on zombie plans and have several mutually contradictory versions thereof. I played in my school's game of Zombies every year for three years and won twice.

If I actually had to fight zombies and, by some bizarre chance, managed to not get eaten, I would probably end up with a massive case of PTSD. Guns make a loud noise that makes me cry so I definitely do not want to be a gunslinging zombie fighter. And on a very crucial level, I do not actually want my entire family to be eaten by the living dead. If you decided to feed my entire family to the living dead, I would be really upset.

The same thing is true of my rape fantasies. It is totally possible to be aroused by the idea of being raped without actually having the slightest interest in being raped in the real world, in the same way that it is possible to enjoy thinking about zombie plans without the slightest interest in the destruction of civilization, possible to love first-person shooters without wanting to be a mass murderer, and possible to go to Ren Faires every weekend while appreciating the benefits of modern sanitation. In fantasies, nothing permanently bad happens, you can be as psychologically unrealistic as you like, and everything is ultimately under your control.

Seriously, what kind of incredibly boring person only fantasizes about things they actually want to do?

4a. What if they want to be raped?

Let’s say, for the sake of argument, that there exist women who genuinely desire to be raped, not in the sense that I want the zombie apocalypse, but in the sense that I want ice cream, backrubs, and lots of notes on Tumblr. That still does not mean it is okay to rape them.

Fifty percent of women who have been sexually assaulted have PTSD, compared to 7.8% in the general population. 94% of women who have been sexually assaulted have PTSD symptoms in the weeks after the sexual assault. This is, incidentally, a higher rate of PTSD than combat soldiers. So the best-case scenario is orgasms, and the worst-case scenario is serious trauma to the point of PTSD.

The next question is: how are you identifying them? If you ask “hey, do you enjoy being raped?” that is fine. But presumably a lot of women who genuinely want to be raped would respond with “no” on account of talking about it ahead of time would make it too consensual. So you have to come up with some strategy for identifying those women. How confident are you that your strategy doesn’t have any false positives? How are you getting feedback on how accurate it is? Remember that women who have been sexually assaulted are probably not going to go up to you and say “hey, you sexually assaulted me.” If you have a false positive, you’re probably not going to know. If it’s something like “she’s wet”… well, remember that vaginas lubricate during nonconsensual sex to prevent tissue damage, and that genitals often act without the consent of their owners. (Remember when you were fourteen and got an erection in math class?)

4b. What about Fifty Shades of Grey?

There are two points here. First, a lot of manospherians are talking about BDSM. Many women are interested in being tied up, or flogged, or called a slut during sex. (Or, to be sure, tying people up, flogging people, or calling people sluts.) This does not mean that they want to be tied up nonconsensually, in the same way that eating chocolate cake does not mean you want to be forcefed cake, or wanting to go to the movies does not mean you want to be kidnapped and stashed at a movie threater. Consent is actually pretty important.

Second, Fifty Shades of Grey actually does depict a lot of nonconsensual sex and abusive behavior. (See Cliff Pervocracy's excellent review here.) I think this is coming back to the same point that fantasizing about something doesn't mean you actually want to do it. For instance, Ana is pretty freaked out by Christian stalking her and buying her expensive presents she doesn't want. However, we the readers know that she is eventually going to be fine with it, that she actually loves him, and that they are going to get together at the end of the book. And it's pretty romantic to imagine that someone loves you so much that they will overcome any obstacle to show their devotion, even obstacles like "I am freaked out and actually kind of scared by how much you love me." However, in the real world, there is no guarantee that you are going to be passionately in love with the person stalking you. In fact, in the real world, the opposite of that is usually true.

5. If women don't like assholes, how come assholes have more sexual partners?

Disagreeable and high-dark-triad men have more sexual partners than agreeable and low-dark-triad men. I used to answer this conclusion with "okay, maybe psychopaths rape people and are more likely to cheat, and people break up with disagreeable and high-dark-triad people thus letting them have more relationships." But I have recently discovered a far better explanation, namely, that sluts are evil.

I could cite a bunch of studies here, but I really think the Wikipedia page speaks for itself: “Individuals who are sociosexually unrestricted [i.e. sluts] tend to score higher on openness to experience, and be more extraverted, less agreeable, lower on honesty-humility, more erotophilic, more impulsive, more likely to take risks, and more likely to have an avoidant attachment style. Conversely, restricted individuals tend to score lower on openness to experience, and be more introverted, more agreeable, more honest and humble, more erotophobic, less impulsive, less likely to take risks and more securely attached.[12] Individuals who score high on the Dark Triad traits (i.e. narcissism, Machiavellianism, psychopathy) tend to have an unrestricted sociosexual orientation. Higher masculinity [15] and eveningness [16] in women is related to unrestricted sociosexuality. High self-monitoring is also associated with unrestricted sociosexuality, regardless of gender or sexual orientation.[17]”

Whew.

It is important to note that only a third of questions about sociosexuality in the most commonly used inventory are about number of sexual partners. It’s possible to have a very unrestricted sociosexuality and be a virgin. That is, this is not primarily measuring how attractive other people find you, it is measuring how much you want casual sex and how motivated you are to have casual sex.

People who are more motivated by casual sex tend to have more casual sex, for a couple of reasons. If you really don’t want casual sex, you’re probably going to turn it down. If you really want casual sex, you’re probably going to put a lot of effort into getting it and lower your standards of sexual partner. And if the people who are most desirous of casual sex are also the most likely to be massive assholes, we’ll see that massive assholes tend to have the most sexual partners. This is not because nice people are unattractive, it’s because nice people, in general, don’t want casual sex.

6. But women don’t like it when you worship them.

Heartiste sometimes describes beta males as “putting the pussy on a pedestal.” They pretend to agree with women even when they don’t; they suck up; they think that a woman is a goddess angel perfection who doesn’t have a single flaw. I agree! This is bad and deeply unattractive behavior. Feminists also do not want you to put the pussy on a pedestal. We have amassed a considerable amount of evidence that putting the pussy on a pedestal leads to gender inequality, the oppression of women, and other things feminists don’t like. I am glad to find some points of agreement here.

However, your two options are not being an asshole or putting her on the pedestal. There is a third option, namely, being a normal fucking human being. Interact with women you want to fuck in basically the same way that you interact with men, or with women you don’t want to fuck. (Possibly with more flirting.) To a certain extent, when you look at what Heartiste actually recommends, he’s recommending this.

It seems possible to me that there are a lot of guys to whom being a normal fucking human being to women feels like being an asshole– they have internalized “you should never disagree with her and do whatever she says and buy her tons of things” to the extent that not doing so feels like being a jerk. In that case, “be more of a jerk” is decent advice. However, you get into the all debates are bravery debates problem here. For every guy who needs to be told to be more of a jerk, there’s another guy who is already an asshole to attractive women and tone it down a notch. Instead of describing which direction to go in, you should describe the target, i.e., “normal fucking human being.”

B. Game.

1. Okay, so maybe women don’t like jerks. But game still gets at universal female sexuality, right?

Not necessarily! Look at the group of women whom PUAs target. By his own admission, Heartiste is primarily interested in twentysomething, conventionally attractive, thin, feminine women, and has very little experience hitting on other ones. But twentysomething, conventionally attractive, thin, feminine women are not a random selection of women. They are probably far more similar to each other than two randomly selected women would be. This is totally fine when you’re talking about getting laid: after all, if you want to have sex with twentysomething, conventionally attractive, thin, feminine women, there is no reason to collect data about fat, butch fortysomethings. However, when you’re making generalizations about women in general– and particularly if you’re making prescriptions about society– it is useful to note the limitations of the data collection.

Game is mostly divided into day game (hitting on female strangers on streets, in grocery stores, in libraries, etc.) and night game (hitting on female strangers in nightclubs, in bars, etc.) The data is confounded two different ways here. First, introverts exist. The usual response of the introvert to a random stranger coming up to them and talking is “maybe if I stand perfectly still they won't eat me." Another huge segment of the female population that game may or may not apply to. Second, women who dislike the thing you're doing will probably notice what you're doing and disengage fairly quickly. Even the best PUA gets blown out. Roosh V agrees with the same basic concept, calling it a fuck funnel. The women who blow you out or wander away five minutes into the conversation are disproportionately going to be women who are not attracted to the kind of game you're running.

As far as I know, there is only one study that examines pickup artists' techniques empirically. A woman's tendency to find PUA strategies attractive was correlated with her score on the ambivalent sexism inventory. Women who have high levels of hostile sexism– who tend to agree with statements like "Women seek to gain power by getting control over men" and "Many women are actually seeking special favors, such as hiring policies that favor them over men, under the guise of asking for "equality."– were attracted to PUA techniques if they were highly sociosexual (i.e. sluts). In addition, women who had high levels of benevolent sexism– who agreed with statements like "Many women have a quality of purity that few men possess." and "A good woman should be set on a pedestal by her man."– find PUA techniques attractive.

Let's pause for a moment: "benevolent sexism" is usually interpreted by non-academic-feminists as "sexism that helps women out." That is not what academic feminists mean when they use the term. "Benevolent sexism" means "the inaccurate belief that women are saintly, pure angels who can do no wrong and should be treated with special consideration by men." If you replace "benevolent sexism" with "putting the pussy on the pedestal," that is pretty much what it means. This idea hurts women a lot: to pick an example everyone except Heartiste thinks is bad, a common argument against women's suffrage was that women were too pure, delicate, and refined to vote. I realize that "benevolent sexism" looks like it means "sexism that helps women out." I did not name it and I usually use different words for the concept because of exactly this confusion. But this study uses the phrase "benevolent sexism" so now I have to write the long clarifying bit.

Let's do this in all caps. BENEVOLENT SEXISM MEANS PUTTING THE PUSSY ON THE PEDESTAL, IT DOES NOT MEAN FAVORING WOMEN FOR JOBS OR SOMETHING. Okay.

Women who have high levels of benevolent sexism are women who believe strongly in traditional gender roles. They tend to have more gender-traditional desires in general: for instance, this study suggests that women with high levels of benevolent sexism are more likely to want a gender-conforming partner. Therefore, they're attracted to men who perform traditional masculinity.

Of course, the causation could go either way here– perhaps benevolently sexist women believe in traditional gender roles because they like traditionally masculine men, or perhaps they like traditionally masculine men because they believe in firm gender roles. Nevertheless, what this suggests is that PUA teaches men to cater to a certain kind of woman by performing the kind of masculinity she finds attractive. This would increase men's sexual success– after all, there are a lot of benevolently sexist women out there– but it does not necessarily mean that all women are attracted to the trait. Being more attractive to a small group of women, and filtering out the other women quickly, is a good route to sexual success.

And, yes, this does mean that not putting the pussy on the pedestal attracts women who think the pussy should be put on a pedestal. The world is amazing.

2. “Perform traditional masculinity” is a little vague. Why don’t you look at some specific game techniques?

Let us look at the most famous kind of game, the humble neg. Yes, I know negs aren’t all of game, but they’re pretty much the only game technique most feminists have heard of, so I’m using it as an example. I think, as a feminist, I am supposed to say negs are horrible, evil, and misogynistic, but Scott compared my last haircut to Joffrey Baratheon, so I have literally 0 legs to stand on here. So let’s talk about why women might like negs.

The neg was originally developed by Mystery for use in nightclubs. Let me give you an idea of what nightclubs are like.

Let’s say you’re going to a mall which has a miniature golf rink. You like miniature golf well enough– it’s not your favorite thing to do ever, but you’ve had fun with it before and you’re open to playing some again. You’re out shopping, buying yourself some ties or whatever, and a strange man comes up to you and says, “hey, do you want to play miniature golf?” You think to yourself, “well, that’s kind of weird, and I’m not really feeling it,” then say, “no.” As you’re walking to the next store, another stranger comes up to you and says “do you want to play miniature golf with me?” You’re kind of weirded out, but you say “no, thanks,” to which he responds “Asshole. Like I wanted to play with you anyway.” You are barely two steps away when someone comes up to you and is like “do you want to play miniature golf with me?” Getting really irritated at this point, you say, “no!”

Then imagine another man comes up to you. You say, “I do not want to play miniature golf with you!” He says, “actually, you don’t look like the kind of person I want to play miniature golf with.” You are so relieved that finally someone doesn’t just want to play miniature golf with you that you talk to him. He turns out to be pretty cool, and if you happen to be in the part of the mall with miniature golf in it you might even decide to play a few rounds.

The woman’s hypothesis, when a strange man approaches her in a nightclub (night game) or on the street (day game), is that he is solely involved in this conversation because he wants to have sex with her. Even if she likes sex, that gets really annoying after a while. Negs convey that you are not interested in having sex with her, so she is more likely to be interested in the conversation. Negs are not an inherent trait of women; they are a response to a social condition.

The other thing that negs do is described by a post of Scott’s on this very blog. If you say to a girl “You know why you guys suck? Because you’ve been staring at me for fifteen minutes and haven’t said hi. Weird, and a little creepy,” you are countersignalling. The fact that you came over subcommunicates that you actually like that you are looking at them– in fact, you find it flattering. So you are saying, “actually, we are going to get along so well that I don’t have to be nice to you, because you already know that I like you.” Pulling off teasing early in a relationship signals social adroitness, which is attractive. And this sort of teasing early in a relationship establishes emotional intimacy: because it’s usually reserved for closer relationships, it gives the illusion that the relationship is much closer than it actually is. (Revelations of personal trauma are also good for this.)

Finally, there is a category of negs that no reasonable human being would interpret as an insult. Seriously, Heartiste, love, “You seem really modest. Modesty is a lost art. It’s not a bad thing… usually. Not everyone feels a need to be an exhibitionist” is a neg? Really? I have social phobia and that wouldn’t make me feel insecure. I would feel complimented!

In short, negging women probably makes them more attracted to you, and this is totally explainable without recourse to “because it lowers her self-esteem and women like that because they are crazy mutant aliens.” I am not sure why, with a bunch of reasonable and ethical explanations available, you have decided to go with the one that makes you look like a Sith Lord.

3. I don’t believe it, I think that you should give some more examples of how game works.

Let’s try a takeaway. The proposed explanation is that takeaways make women more attracted to you because it makes her feel low-status and want to compete for her attention. On the other hand, it could be that takeaway is using a very common concept in social psychology, that of scarcity. It’s the same psychology that applies to sales in stores: if this is only available for a limited time, you have to get it NOW. Black Friday does not work because people are insecure about how much Black Friday likes them.

I mean, you can have the explanation where women are crazy mutant aliens and you are Chaotic Evil, or the “women continue to obey common principles of social psychology even when they’re flirting with someone.” Up to you, really.

And, yes, the scarcity effect also works on men. Why do you think generations of mothers encouraged their daughters to play hard to get?

4. I have a good job, why don't women want to date me?

To quote from a recent Heartiste post:

> In a way they exhibit some of the same qualities of those professions—ego, arrogance, and unlimited amounts of cash. In San Francisco, said Violet, "There were a lot of men to date with disposable income who wanted to take women out. It's just, it was so boring," she said. "My dating life went from dating artists and writers and going on cheap but exciting dates, to men who thought the ability to buy someone an expensive meal made them interesting."

Heartiste interprets this as Violet believing that "a cheap date with a man who has that ineffable alpha attitude is far more intoxicating… than is an expensive date with a beta male who plays by the traditional courtship rules." This is really interesting, because over the course of the article exactly one group of people is described as arrogant, egotistical assholes, the kind of person Heartiste thinks women want to date. That group is the beta males.

This is clearly not "I'm not attracted to him so he's a jerk." There are specific details here. The men Heartiste considers beta males talk endlessly about their jobs, show little interest in the person they're sharing the table with, and assign reading lists before the first date.

It is almost as if when Heartiste complains about women liking alpha assholes, he's not talking about women liking assholes at all. He is complaining that women prefer charming, poor people with interesting jobs to charmless, rich jerks.

We see this in other places in Heartiste's writing as well. Heartiste occasionally talks about alpha gifts and beta gifts, in which he points out that girls treasure a recording of a song written for them more than they treasure an expensive diamond bracelet. No fucking shit. You can give a diamond bracelet to literally any girl and it requires nothing other than the ability to fork over cash. However, if you wrote a song for her, it shows you are putting in time and effort and thinking about her when she's not there (and that you are musically talented, which is sexy). However, for some reason, in Heartisteland, the former is Beta and therefore Nice and the latter is Alpha and therefore Jerkboy.

This is actually pretty common across the manosphere. My favorite examples come from Dalrock's comment section. For example: "Does [the sexual arousal detecting bra] detect for (and open for) a male prospective partner having one or more of tattoos, needletracks, violent criminal convictions, long hair (preferably unclean), STDs, abandoned bastard progeny, motorcycle ownership, constant use of profanity, and lacks of STEM degrees or current fulltime employment?" I will accept "needletracks," "abandoned bastard progeny," and "violent criminal convictions" as being signs of a person you should not date. However, there are lots of people who are emotionally supportive, romantically compatible, and all-around excellent boyfriends who also happen to have long hair, tattoos, and a motorcycle. In fact, I suspect that motorcycle ownership and boyfriend quality are basically uncorrelated traits! This is not talking about women chasing jerks, it is talking about women dating men who have traits coded lower-class and alternative (and, in the case of the last, explicitly being lower-class).

In short, the manosphere needs to stop complaining that women are not golddiggers. I am very sorry that someone lied to you and told you that women are golddiggers. They are not, and it is generally considered to be a good thing to judge potential romantic partners on their personalities and romantic compatibility with you rather than their wallets. And it is generally considered bad behavior to put serial killers and poor people in the same category.

C. Do women have a rationalization hamster which makes them justify their terrible decisions in self-aggrandizing ways?

Yes.

C1. I wasn't expecting that.

Men also have a rationalization hamster. Everyone has a rationalization hamster. People rationalize. This is not a unique trait to women and proof of their inherent evil, it is a basic fact about humans. If you think you don't rationalize, it is probably good evidence that you rationalize so well that you don't notice. Although I disagree with the modular model of human psychology that he espouses, Robert Kurzban's Why Everyone (Else) Is a Hypocrite sums up a lot of the research about rationalization in a readable manner.

C2. But women rationalize more than men about sex.

Probably true! When women are asked their number of sexual partners, it is about half of men's. When women are hooked up to lie detectors and asked their number of sexual partners, it is about the same as men's.

However, it's important to note that in our culture there's still a lot of stigma against being promiscuous. This might not show that women are more prone to rationalizing in general than men are; it simply shows that women pay a higher price for being honest about their sexuality. (Which is a prediction of feminist theory, by the way.) Men may be more likely to rationalize about something that is more threatening to them, even if it is something about sex, such as whether they've been attracted to men in the past.

–yes, this study shows that pretty much all self-report data about sexuality is totally useless. I have used it anyway in this FAQ because it is all I have. However, I would like to encourage people to remember that women are lying fuckers about sexuality and men may or may not be lying fuckers about sexuality and that everything may be indicating not so much women's actual behavior as what is most high-status. Also, we should clearly replicate every self-report study of sexuality with everyone hooked up to lie detectors.

C3. Does that mean that we shouldn't listen to women about their sexual preferences and instead look at what they do?

Sort of.

You can certainly look at the data and go "goodness, a lot of women tend to stay in relationships where they are hit." This doesn't mean that you should go out and hit women. As unreliable as self-report data is, it is still our best source of information about why people do what they do and what it feels like to experience something. And the data about relationships where women get hit suggests, as I outlined above, that the reason that women stay in abusive relationships is that abusers are rational people who try really hard to make them leave. (It also doesn't mean that hitting your partner makes your relationship longer. Most women do eventually leave their abusive partners, after all.

“Longer than would be naively expected” is not the same thing as “longer than baseline.”) What people do gives you a correlation, but it doesn’t give you an explanation. To get an explanation, you need experiments or self-report data, as flawed as it is. And it is unfortunately impossible to randomly assign women to the Getting Hit or Not Getting Hit condition of a randomized controlled trial, so self-report it is.

D. It is terribly unfair when women have sex with alpha males in their prime and then expect to get married to me when other men had them younger, hotter, tighter, for free.

OK, y’all are really happy about the whole supply-demand model of sexuality, where men exchange resources for pussy and an increase in the supply of pussy lowers the number of resources required to pay for the pussy. The thing you are complaining about is called price discrimination and is a perfectly ordinary part of economic life. Do you complain about how unfair it is when restaurants have discounts for the elderly? Do you mutter darkly when stores stock the expensive food at eye-level so a careless person will pick it up? Do you get angry about bulk discounts? If you are willing to pay a higher price for pussy, then the producer of pussy is perfectly within her rights to attempt to capture the surplus. (If you aren’t willing to pay the higher price, then I am not sure what the harm to you is.)

Your proposed solution is that we implement a pussy price floor. Most economists think that price floors are really bad because they distort the natural functioning of the market. When there’s a price floor below the market equilibrium, the demand for the product (pussy) lowers while the supply increases, and a lot of transactions that would benefit both parties don’t happen. To translate into sexual market terms: everyone gets laid less, a lot of sad women wander around going “please marry me! Please! I am desperate to be married!”, and a lot of dudes do that Men Going Their Own Way thing. …Or, in fact, the exact consequences you guys predicted for not having a price floor. Have you taken intro economics?

Or– and I really hate doing this, because I feel like it’s not in keeping with the spirit of the game– one might point out that marriages are in fact different from prostitution in a few eensy-weensy tiny details. For instance, sometimes men want to date people they’re having casual sex with. (If you ask me, way too often.) You would think that since he was getting the sex for free, he would not want to date her and pay more for it! It is like paying an extra fifty thousand dollars for your car out of the goodness of your heart!

However, one generally gets things from relationships other than sex and resources, such as emotional support, long conversations into the night, spending time with a person you like, affection, the warmth of knowing someone loves you, and so on. It is perfectly possible that most of the time someone has casual sex with someone else, it is because they enjoy having sex with that person but that person has a really obnoxious sense of humor, doesn’t share any common interests with them, talks way too much, or is just romantically unattractive for ineffable chemistry reasons. In this model, the fact that your wife wants to marry you suggests that she likes you better than the people that she had casual sex with, because she wants to talk to you and provide you with emotional support and fuck you, whereas she just wanted to fuck the other guy.

Is it possible that your wife had casual sex with a lot of guys way hotter than you and is now settling for you and secretly seething with contempt about how hideous you are compared with the guys she used to sleep with? Of course!

However, I do not think that you should assume that this is true of literally every woman who had sex before she gets married. I would suggest some helpful advice here like “why don’t you meet some of her fuckbuddies and see if they are way hotter than you first?” but then I remembered Heartiste is way into monogamy. (Or, like, monogamy except he cheats.) See, this is an advantage of polyamory. You can totally see that people tend to like their primary partners more than they like their fuckbuddies! In real time!

D1. It is terribly unfair when women follow their tingle to fuck alpha males.

Oh no. Women are fucking people they’re sexually attracted to instead of people they’re not sexually attracted to. How evil. Those monsters. They should clearly distribute sex by sortition instead.

I am sure men have never followed their tingles, except a couple of bullet points ago when you guys were saying that 95% of men’s criteria for selecting a wife was how hot they were.

E. Do women actually experience significant behavior changes during ovulation?

The truly excellent Bayesian statistics blogger Andrew Gelman has spent a lot of time beating up on the behavior changes in ovulation literature. For instance, did you know that one of those papers defined peak fertility as days six to fourteen of the cycle even though actual peak fertility is between days ten and seventeen? And another one failed to replicate and then claimed that this is because the “women wear red when they ovulate” effect is mediated by the weather and the latter study was conducted in the summer?

I am not saying that it is impossible that women experience significant behavior changes during ovulation, but the scientific literature clearly doesn’t seem to show this. In addition, anecdotally, a lot of women get really horny on their periods, which seems difficult to square with the “menstruation makes women not interested in sex” theory.

F. Are women less interested in casual sex than men are?

In general, women seem to be less interested in casual sex than men are. There are a couple different factors here. First, testosterone seems to increase libido and estrogen seems to decrease libido. Since women usually have more estrogen and less testosterone than men do, as a class they will probably have lower libidos, which probably means they’re also less interested in sex.

Second, the orgasm gap. In relationships, women orgasm about 80% as much as men do. However, in casual sex, women have orgasms about half as often as men do, and a third as often in first-time hookups. While, of course, it is possible to have enjoyable sex without experiencing an orgasm, orgasm is an indicator of sexual satisfaction. Possibly this is related to the sex acts: sex in our culture is defined as penis in vagina intercourse, which usually doesn’t result in orgasms for women, and men receive oral sex about eighty percent of the time in first-time hookups, while women receive it less than half the time. Men, imagine if casual sex meant that you gave a woman head and then she ground against your penis until she came, and if you don’t orgasm tough. I imagine most of you would be less enthusiastic about casual sex too.

Third, we have to consider the findings of my second-favorite study of all time. Clark and Hatfield is a famous and much-replicated study in which an attractive stranger asks women and men to go to bed with them. Men very commonly say yes; women never do. Terri Conley replicated this study as a questionnaire in which she asked people to describe their perceptions of attractive strangers who come out of nowhere and ask them to have sex. It turns out that women think random strangers who approach them on street corners are likely to be dangerous and terrible at sex. In fact, only the perception that the stranger would be a good lover significantly influenced women's desire to have sex with the person: they were more likely to have sex with a celebrity or a male friend, because both of those are more likely to be perceived as good lovers. Status? Doesn't matter. Contribution of resources? Doesn't matter. What matters is that you are capable of getting her off. And– as we see from the orgasm data– men clearly are not stepping up to the challenge here.

IV. Miscellaneous

A. Is Heartiste against pretty lies?

Heartiste's slogan is "where pretty lies perish." It's true that he has a lot of beliefs that are deeply emotionally threatening to some groups. "Men only care about your tits and ass and not your personality, when you turn 30 you will be hideously unattractive, and every sexual partner you have makes you less attractive" is pretty fucking scary to women. However, Heartiste has cleverly managed to not have any beliefs that are upsetting to his readers.

Like, look at Heartiste's beliefs from a man's point of view. If you get laid a lot, you are an awesome alpha male and the entire site is about how great you are. One might assume that that means that the site is threatening to men who can't get laid. However, it offers men who can't get laid the best possible thing: someone to blame. It’s not your fault that you can’t get laid, it’s the fault of women. Women’s unleashed sexuality is evil and destroying civilization. They are chasing assholes and leaving you be. “You are just so good a person that no one wants to fuck you, those evil bitches” is not actually a particularly threatening statement.

Perhaps there are no statements that are threatening to men! Perhaps all true statements are actually just comfort to guys! However, I would like to offer up a few suggestions for Heartiste to consider in his future explorations of pretty lies. First, examine the following charts from Rational Male. Men’s peak attractiveness is 38, women’s peak attractiveness is 23. Now, look at OKTrends’ charts, which have the benefit of being based on data not from the Journal of Cerebrorectal Studies. Men are primarily messaging women in the 18-22 age range, thus suggesting that that is women’s peak attractiveness and Rollo actually slightly overestimated when women’s peak attractiveness is. But women are primarily messaging men roughly around their own age. This remains true no matter how old she is. Since women as a group don’t seem to have a preference for a single age the way men do, we might as well calculate peak attractiveness by which age men are most likely to date the women they prefer. And the answer is… oh dear… around 20-24.

Sorry, 38-year-old dudes. Young women do not want your old, balding, wrinkled ass. Your peak attractiveness was fifteen years ago.

You can check this by looking at your own social groups. Presumably a lot of 38-year-olds would cash in their attractiveness to date a 20-year-old… but how many late-thirties/early-twenties couples do you know? Approximately none? As for the few such couples, well, I say what Heartiste says when someone brings up a hot fortysomething celebrity: the existence of an outlier does not falsify the general rule.

What other threatening ideas can we think of? One of my personal favorites, as a classicist, is pederasty. Pederasty is extremely common cross-culturally: from Greece and Rome to China and Japan to medieval Islam to Victorian England to modern Afghanistan, as soon as men do not face significant social punishment for fucking a teenage boy they do so. Notice that our culture has a significant “no homo” pressure on straight men. As Heartiste himself says when discussing men having sex with fat women, there is not much reason to shame people into not doing things they weren’t going to do anyway. And a lot of men seem to get absurdly angry about One Direction or Justin Bieber or whoever the latest prettyboy teenager is. Could that be sublimated lust? Hm.

To be clear, I am not suggesting that heterosexual men want to fuck teenage boys. I am suggesting that that notion is deeply upsetting to most men, and supported with evidence roughly as good as the evidence behind many of Heartiste’s claims, and the fact that he hasn’t come up with this idea seems to suggest that he’s avoiding ideas he finds personally offensive and threatening and therefore he should stop gloating about how much better he is than all those other people who avoid personally offensive, threatening statements. If he begins to argue in favor of men being pederasts, then I shall certainly retract my previous statements.

B. I am a confused and lost Return of Kings reader.

I am deeply concerned about several factual inaccuracies on your blog.

First, dating girls with eating disorders is not a good way to get to date someone who is agreeable and goes along with whatever you say. Mentally ill girls, in general, are not good people to date if you want a low-drama relationship.

Second, demisexuals are only sexually attracted to people once they have an emotional connection to the person. While in general sex-positive people tend to use more weird labels and thus be more into casual sex (you're welcome, by the way) and the indicator you're looking for are correct, if you are aiming for casual sex on the first date, I highly advise you to avoid demisexuals.

Third, if you want to reduce your rate of false accusations of rape, it would be helpful if you stopped having sex with girls when they say 'no' first. I guarantee you that if you try this strategy it will reduce your rate of being accused of rape tremendously. It also seems to me that a girl who says "no" when she doesn't mean "no" is not going to give you the peaceful, low-drama relationship you're dating girls with eating disorders for. It seems like it would be wise to filter those girls out by not having sex with people who have said they don't want sex.

Fourth, I empathize with your desire to date strippers. Stippers are awesome! And it is really easy to think "well, where are strippers? They are in the strip club!" Pretty much every girl in a strip club is a stripper; while some of the girls in a grocery store are strippers, you might end up asking out a non-stripper on accident, and then where would you be? Not dating strippers, is where! Unfortunately, strippers are at work! Their actual job is to make you think that you are going to have sex with them if you pay them money. If you hit on them at work, you will be put into the category "client," and if you don't pay for a dance, the category "terrible, timewasting client." Instead, you should look for strippers in a place where there are a lot of strippers and they are not at work. (Anecdotally, this place is White Wolf games. However, my data on this comes from Portland and it might just be that literally everywhere in Portland is full of strippers.)

C. Okay, where should I get dating advice from?

I sympathize with your problem, O Lonely Men of the World! Most dating advice is really terrible and written by people who think that the world is a just place and nice people are rewarded and you should just be yourself. Worse, a lot of it is written by feminists. Feminists are good at many things, but they are not good at giving men dating advice. Most of the time their dating advice is like “here are the thirty thousand ways that if you hit on someone you are misogynist, memorize them all and you have a chance of not being EVIL.”

I find myself impressed by Eric S Raymond’s Sex Tips for Geeks. Although I disagree with most of the evopsych, it is honest that people are totally shallow and that expensive clothes, lifting weights, and growing out your hair help. A pickup-knowledgeable friend recommends Mark Manson and The Art of Charm. I haven’t read them and can’t vouch, but my friend is a pretty ethical and charming guy.

Nobody Likes A Tattletale

Today at work, one of my heroin addict patients getting treated in inpatient rehab for heroin addiction managed to smuggle…well, you want to take a wild guess? Yeah, he smuggled in some heroin and got high in the hospital. Another patient saw him do it and told me. I had a long talk with him and took measures to make sure it wouldn’t happen again.

I wouldn’t say I’m disappointed in the addict. Anyone who expects heroin addicts to follow rules that result in them getting less heroin is going to be so consistently frustrated that they will eventually lower their standards.

But I *am* a little disappointed in the patient who told me. Come on, man! Nobody likes a tattletale!

I realize this feeling is totally one hundred percent irrational. The patient was absolutely correct that using heroin in rehab is bad, we enforce our anti-heroin rules fairly and don’t have any draconian punishments when people break them, and most of these people come into rehab at least sorta-voluntarily and agree to the rules. Telling me was absolutely the correct decision.

But I *still* feel a little disappointed in him.

This feeling is not born of any kind of personal experience. I’ve never had much trouble with authority. I follow most of the important laws, I never got in trouble in school. No one’s ever tattled on *me*. Although my association with many libertarians provides me with a lot of examples of authority overreaching itself, I’m pretty sure the rule “don’t use heroin in a drug rehab” isn’t one one of those.

As far as I can tell, my only two consistent positions are "disagree with the existence of rules" and "agree with rules and be happy when people help enforce those rules", and I'm definitely not pushing for the first.

And yet I'm *still* kind of annoyed with that guy.

Dislike of tattletales seems to be, if not a human universal, at least a human very-common, arising in the absence of obvious social pressure and seeming attractive even to people whose social position ought to naturally turn them against it. My impression from old mobster movies is that even the police had contempt for people who ratted on organized crime, even though those people were obviously doing good for society.

This seems to be a clear case of virtue ethics versus utilitarianism. A rat who betrays the mob is helping society by getting rid of criminals, but he's also proving himself an untrustworthy person who betrays his friends and who might not be a good choice to associate with. Fine.

But my patient? He never promised anybody he wasn't going to tell on them. He had no association with the addict besides being a patient in the same hospital as him. If he had any duty at all, it was to his doctors, who were working really hard to help him, and he discharged this duty admirably by helping them enforce their rules.

So in this case I think it is just a flaw in my brain. I am acting as if all my patients had made some kind of implied deal to respect each other's privacy, and the one tattletale was being a dealbreaker by defecting. But since no such deal was made – and since indeed people in a rehab facility should not expect such a deal – there was no deal-breaking involved.

One cannot say the same for the position endorsed by Leah Libresco, who wrote about a similar episode of tattling in her blog post The Ethicist Endorses Omertà

> The NYT‘s Ethicist has taken a very strange approach to wrongdoing in this weekend’s column. A student wrote in to say that ze saw a friend take someone’s car keys and throw them into a lake. The friend offered the letterwriter $50 as an implicit bribe in order to stay quiet. The bribe worked. Later, someone came by looking for his keys, and the letterwriter kept mum. But ze felt queasy about zer choice, and asked the Ethicist for his advice.

Assume that for some reason he can’t just give the guy his $50 back before talking. His only two choices are to keep the money and stay silent, or keep the money and talk. The first choice fails to right a wrong, the second breaks his contract. Which is better?

The Ethicist said the writer was wrong to take the deal, but having taken it, he is compelled to respect it. Leah disagreed, saying that he was wrong to take the bribe, and having realized that he should break his deal and tell the victim everything. She says: “Sticking by an immoral compact wrongs yourself and your accomplice. . . it’s clear that we don’t want people to hew to unethical agreements, simply because breaking promises is bad.”

I’m about halfway between these two positions. One should try one’s hardest to get out of an immoral contract. But if that’s impossible, I think one needs to weigh the moral cost of breaking a promise against the moral cost of carrying out the immoral contract, with a bias towards keeping your word unless it’s totally repugnant.

Let me try to give an example Leah will be especially able to understand.

Suppose that I become a Catholic priest and take confession. I swear not to break the seal of the confessional and not to go tell the secular authorities what I hear.

My first client (I bet there’s a better word for that!) is a child molester who confesses all the child molestation he’s doing. I tell him to stop, and he says unconvincingly that he’ll think about it.

I think “Holy f@#k, I was just expecting people to talk about sleeping in on Sundays, this is way worse than I could have expected”. I decide that my original promise not to tell the secular authorities was immoral, and I go off and tell the secular authorities. They arrest the child molester. Everyone lives happily ever after except that no one confesses things at Church ever again.

Both Leah and myself agree that some sort of a confessional-type institution is useful (even if I as an atheist think of it more in terms of psychiatric confidentiality). But such an institution is impossible without people being able to *really mean* promises. A credible promise can’t just be “I promise to do this thing unless I later decide it is bad, in which case I won’t”. You have to be able to *really* trust someone.

As Leah herself very correctly puts it in a *different* blog post on a *different* botched Ethicist decision:

> The Ethicist is crippling his own ability (and that of anyone suspected to subscribe to his philosophy) to make a promise. A promise is not an indication of present beliefs (“I don’t plan to repeat anything you say in this room”) it is a bind on future action (“I won’t

> repeat what you say, even if I wish I hadn't made this promise later"). If he isn't comfortable making that kind of promise, he has the option to tell patients and others up front, but treating promises as breakable upon reflection dilutes them for him and everyone else.
>
> The covenant marriage movement is meant to counteract this kind of thinking in one sphere. In an age of no-fault divorce, they're trying to carve out a special niche, clearly differentiated from mainstream marriage, where a change of heart isn't sufficient justification to break a promise. But there isn't an equivalent in most other spheres of life. One can say only "I really mean this promise," and a reader of the Ethicist's column might reasonably hear a silent "right now" at the end of that phrase.

But now it's Leah adding the "right now" and The Ethicist enforcing covenants. Leah points out that the Ethicist has changed its mind on this point, but doesn't explain why, upon being given the opposite position, she continues to disagree.

Promises are useful because they allow beneficial Pareto-optimal deals to be made. If promises are untrustworthy, beneficial deals become impossible and everyone loses out. The principle "Break all promises to respect immoral deals" not only makes immoral deals impossible, but also any moral deals where there is a *risk* of either participant deciding they are immoral, or even moral deals where one participant can credibly *claim* to have decided they are immoral and so back out of their obligation punishment-free. This is a pretty big set of deals and so we should not lightly endorse people's ability to break promises they believe are immoral.

I should probably clarify here that all my promises usually contain an implied "unless following this promise is much more difficult than I could reasonably have expected" and I assume my interlocutor knows this. So if I promise someone to get them milk from the store, and then I go to the store and there's only one carton of milk and a guy has just taken it and tells me he won't give it to me, I don't feel morally obligated to beat him up and steal it from him. If somebody wants a promise from me *without* the implied "unless" they are welcome to ask me for it. Or in certain cases where it is obvious that is what they want, I will assume it without being asked. And in *those* sorts of cases if I make it I will keep it, beating-up and all. But I would think *much* harder before making a promise like that, and I would lawyer its wording the same way I would a wish from a genie with a known mean streak.

Much simpler and perhaps best of all were those ancient promises, where people were like "If ever I betray your trust, then may the ravens of Odin peck out my right eye!" There's no ambiguity here. You know exactly what's enforcing the deal – getting your right eye pecked out by ravens. If you later decide your deal was unethical, you are welcome to assuage your conscience by cancelling it, but you should still expect to have your right eye pecked out by ravens. Since the enforcement mechanism is bloodthirsty heavenly birds rather than morality per se, you don't get these weird questions about whether other, different morality can ride in and free you from it. It's not even a question of "freeing" so much as of trade-offs. If you want to break your promise for money, you can get the money – but the ravens will peck out your eye. If you want to break your promise for love, you can get the love – but the ravens will peck out your eye. And if you want to break your promise for a greater moral cause, you can get your greater moral cause – but your eye still gets pecked out.

You know exactly where you stand with eye-pecking ravens, which is a hell of a lot better than you can ever say about morality.

The Invisible Nation – Reconciling Utilitarianism And Contractualism

[Attempt to derive morality from first principles, totally ignoring that this should be impossible. Based on economics and game theory, both of which I have only a minimal understanding of. And mixes complicated chains of argument with poetry without warning. So, basically, it's philosophy. And it's philosophy I get the feeling David Gauthier may have already done much better, but I haven't read him yet and wanted to get this down first to avoid bias towards consensus]

Related to: Whose Utilitarianism?, You Kant Dismiss Universalizability, Meditations on Moloch

Imagine the Economists' Paradise.

In the Economists' Paradise, all transactions are voluntary and honest. All game-theoretic problems are solved. All Pareto improvements get made. All Kaldor-Hicks improvements get converted into Pareto improvements by distributing appropriate compensation, and then get made. In all cases where people could gain by cooperating, they cooperate. In all tragedies of the commons, everyone agrees to share the commons according to some reasonable plan. Nobody uses force, everyone keeps their agreements. Multipolar traps turn to gardens, Moloch is defeated for all time.

The Economists' Paradise is stronger than the Libertarians' Paradise, which is just a place where no one initiates force and all economic transactions are legal, because the Libertarians' Paradise might still have a bunch of Prisoner's Dilemmas and the Economists' Paradise wouldn't. But it is weaker than Utilitarians' Paradise, because people with more power and money still get more of the eventual utility.

From a god's-eye view, it seems relatively easy to create the Economists' Paradise. It might be hard to figure out how to solve game theoretic problems in absolutely ideal ways, but it's often very easy to figure out how to solve them in a much better way than the uncoordinated participants are doing right now (see the beginning of Part III of Meditations on Moloch). At the extreme of this way of thinking, we have Formalism, where just solving the problem, even in a very silly way, is still better then having the question remain open.

(a coin flip is the epitome of unintelligent problem solving, but flipping a coin to decide whether the Senkaku/Diaoyu Islands go to Japan or China still beats having World War III, by a large margin)

The Economists' Paradise is a pretty big step of the way toward actual paradise. Certainly there won't be any wars or crime. But can we get more ambitious?

Will the Economists' Paradise solve world hunger? I say it will. The argument is essentially the one in Part 2.4 of the Non-Libertarian FAQ. Suppose solving world hunger costs $50 billion per year, which I think is people's actual best-guess estimate. And suppose that half the one billion people in the First World are willing to make some minimal contribution to solving world hunger. If each of those people can contribute $2 per week, that suffices to raise the necessary amount. On the other hand, the $50 billion cost is the cost in *our* world. In the Economists' Paradise, where there are no corrupt warlords or bribe-seeking bureaucrats, and where we can just trust people to line themselves up in order of neediest to least needy, the whole task gets that much easier. In fact, it's not obvious that the First World wouldn't come up with their $50 billion only to have the

Third World say “Thanks, but we kind of sorted out our problems and became an economic powerhouse.”

Let’s get *more* ambitious. Will there be bullying in the Economists’ Paradise? I just mean your basic bullying, walking over to someone who’s ugly and saying “You’re ugly, you ugly ugly person!” I say there won’t be. How would a perfect solution to all coordination problems end bullying? Simple! If the majority of the population disagrees with bullying, they can sign an agreement among themselves not to bully, and to ostracize anyone who does. Everyone will of course keep their agreement (by the definition of Economists’ Paradise) and anyone who reports to the collective that Bob is a bully will always be telling the truth (by the definition of Economists’ Paradise). The collective will therefore ostracize Bob, and faced with the prospect of never being able to interact with the majority of human beings ever again, Bob will apologize and sign an agreement never to bully again (which he will keep, by the definition of Economists’ Paradise). Since everyone knows this will happen, no one bullies in the first place.

So the Economists’ Paradise is actually a *very* big step of the way toward actual paradise, to the point where the differences start to look like splitting hairs.

The difference between us and the Economists’ Paradise isn’t increased wealth or fancy technology or immortality. It’s rule-following. If God were to tell everybody the rules they needed to follow to create the Economists’ Paradise, and everyone were to follow them, that would suffice to create it.

That suggests two problems with setting up Economists’ Paradise. We need to know what the rules are, and we need to convince people to follow them.

These are more closely linked than one would think. For example, both Japan and China might prefer that the Senkaku Islands be clearly given to the other according to a fair set of rules which might benefit themselves the next time, than that they fight World War III over the issue. So if the rules existed, people might follow them *for the very reason that they exist*. This is why, despite the Senkaku Island conflict, *most* islands are not the object of international tension – because there are clear rules about who should have them and everybody prefers following the rules to the sorts of conflicts that would happen if the rules didn't exist.

II.

There's a hilarious tactic one can use to defend consequentialism. Someone says "Consequentialism must be wrong, because if we acted in a consequentialist manner, it would cause Horrible Thing X." Maybe X is half the population enslaving the other half, or everyone wireheading, or people being murdered for their organs. You answer "Is Horrible Thing X good?" They say "Of course not!". You answer "Then good consequentialists wouldn't act in such a way as to cause it, would they?"

In the same spirit: should the State legislate morality?

"Of course not! I don't want the State telling me whom I can and can't sleep with."

So do you believe that it's immoral, genuinely immoral, to sleep with the people whom you want to sleep with? Do you think sleeping with people is morally wrong?

"What? No! Of course not!"

Then the State legislating morality isn’t going to restrict whom you can sleep with, is it?

“But if the State legislated everything, I would have no freedom left!”

Is taking away all your freedom moral?

“No!”

Then the State’s not going to do that, is it?

By this sort of argument, it seems to me like there are no good philosophical objections to a perfect State legislating the correct morality. Indeed, this seems like an ideal situation; the good are rewarded, the wicked punished, and society behaves in a perfectly moral way (whatever that is).

The arguments against the State legislating morality are in my opinion entirely contigent ones, based around the fact that the State *isn’t* perfect and the correct morality *isn’t* known with certainty. Get rid of these caveats, and moral law and state law would be one and the same.

Letting the State enforce moral laws has some really big advantages. It means the rules are publicly known (you can look them up in a lawbook somewhere) and effectively enforced (by scary men with guns). This is great.

But using the State to enforce rules also fails in some very important ways.

First, it means someone has to decide in what cases the rules were broken. That means you either need to depend on fallible, easily biased human judgment – subject to all its racism, nepotism, tribalism, and whatever – or algorithmize the rules so that “be nice” gets formalized into a two thousand page definition of niceness so rigorous that even a racist nepotist tribalist judge doesn't have any leeway to let your characteristics bias her assessment of whether you broke the niceness rules.

Second, transaction costs. Suppose in every interaction you had with another person, you needed to check a two thousand page algorithm to see if their actions corresponded to the Legal Definition of Niceness. Then if they didn't, you needed to call the police to get them arrested, have them sit in jail for two weeks (or pay the appropriate bail) until they can get to trial. The trial itself is a drawn-out affair with celebrity lawyers on both sides. Finally, the judge pronounces verdict: you *really* should have said "please" when you asked her to pass the salt. Sentence: twelve milliseconds of jail time.

Third, it is written: "If you like laws and sausages, you should never watch either one being made." The law-making apparatus of most states – stick four hundred heavily-bribed people who hate each other's guts in a room and see what happens – fails to inspire full confidence that its results will perfectly conform to ideal game theoretic principles.

Fourth, most states are somewhere on a spectrum between "socially contracted regimes enforcing correct game theoretic principles among their citizens" and "violent psychopaths killing everybody and stealing their stuff", and it has been historically kind of hard to get the first part right without also empowering the proponents of the second.

So it's – surprise, surprise – a tradeoff.

There's a bunch of rules which, followed universally, would lead to the Economists' Paradise. If the importance of keeping these rules agreed-upon and well-enforced outweighs the dangers of algorithmization, transaction costs, poor implementation, and tyranny, we make them State Laws. In an ideal state with very low transaction costs, minimal risk of tyranny, and legislave excellence, the cost of the tradeoff goes down and we can reap gains by making more of them State Laws. In a terrible state with high transaction costs that has been completely hijacked by self-interest, the cost of the tradeoff goes down and fewer of them are State Laws.

III.

Let’s return to the bullying example from Part I.

It would seem there ought not to be bullying in the Economists’ Paradise. For if most people dislike bullying, they can coordinate an alliance to not bully one another, and to punish any bullies they find.

On the contrary, suppose there are two well-delineated groups of people, Jocks and Nerds. Jocks are bullies and have no fear of being bullied themselves; they also don’t care about social exclusion by the Nerds against them. Nerds are victims of bullies and never bully others; their exclusion does not harm the Jocks. Now it seems that there might be bullying, for although all the Nerds would agree not to bully, and to exclude all bullies, and although all the Jocks might coordinate an alliance not to bully other Jocks, there is nothing preventing the Jocks from bullying the Nerds.

I answer that there are several practical considerations that would prevent such a situation from coming up. The most important is that if bullying is negative-sum – that is, if it hurts the victim more than it helps the bully – then it’s an area ripe for Kaldor-Hicks improvement. Suppose there is *anything at all* the Nerds have that the Jocks want. For example, suppose that the Nerds are good at fixing people's broken computers, and that a Jock gains more utility from knowing he can get his computer fixed whenever he needs it than from knowing he can bully Nerds if he wants. Now there is the opportunity for a deal in which the Nerds agree to fix the Jocks' computers in exchange for not being bullied. This is Pareto-optimal: the Nerds' lives are better because they avoid bullying, and the Jocks' lives are better because they get their computers fixed.

Objection: numerous problems prevent this from working in real life. Nerds and Jocks aren't coherent blocs, bullies are bad negotiators. More fundamentally, this is essentially paying tribute, and on the "millions for defense, not one cent for tribute" principle, you should never pay tribute or else you encourage people who wouldn't have threatened you otherwise to threaten you just for the tribute. But the assumption that Economists' Paradise solves all game theoretic problems solves these as well. We're assuming everyone who should coordinate can coordinate, everyone who should negotiate does negotiate, and everyone who should make precommittments does make precommittments.

A more fundamental objection: what if Nerds can't fix computers, or Jocks don't have them? In this case, the tribute analogy saves us: Nerds can just pay Jocks a certain amount of money not to be bullied. Any advantage or power whatsoever that Nerds have can be converted to money and used to prevent bullying. This sounds morally repugnant to us, but in a world where blackmail and incentivizing bad behavior are assumed away by fiat, it's just another kind of Pareto-optimal improvement, certainly better than the case where Nerds waste their money on things they want less than not being bullied yet are bullied anyway. And because of our Economists' Paradise assumption, Jocks charge a fair tribute rate – exactly the amount of money it really costs to compensate them for the utility they would get by beating up Nerds – and feel no temptation to extort more.

Now, I'm not sure bullying would even come up as an option in an Economists' Paradise, because if it's a zero- or negative- sum game trying to get status among your fellow Jocks, the Jocks might ban it on their own as a waste of time. But even if Jocks do get some small amount of positive utility out of it directly, we should expect bullying to stop in an Economists' Paradise as long as Nerds control even a tiny amount of useful resources they can use to placate the Jocks. If Nerds control no resources whatsoever, or so few resources that they don't have enough left to pay tribute after they've finished buying more important things, then we can't be *sure* there won't be bullying – this is where the Economists' Paradise starts to differ from the Utilitarians' Paradise – but we'll return to this possibility later.

Now I want to highlight a phrase I just used in this argument.

"If bullying is negative-sum – that is, if it hurts the victim more than it helps the bully – then it's an area ripe for Kaldor-Hicks improvement"

This looks a lot like (naive) utilitarianism!

What it's saying is "If bullying decreases utility (by hurting the Nerd more than it helps the Jock) then bullying should not exist. If bullying increases utility (by helping the Jock more than it hurts the Nerd) then maybe bullying should exist. Or, to simplify and generalize, "do actions that increase utility, but not other actions."

Can we derive utilitarian results by assuming Economists' Paradise? In many cases, yes. Suppose trolley problems are a frequent problem in your society. In particular, about once a day there is a runaway trolley in heading on a Track A with ten people, but divertable to a Track B with one person (explaining why this happens so often and so consistently is left as an exercise for the reader). Suppose you're getting up in the morning and preparing to walk to work. You know a trolley problem will probably happen today, but you don't know which track you'll be on.

Eleven people in this position might agree to the following pact: "Each of us has a 91% chance of surviving if the driver chooses to flip the switch, but only a 9% chance of surviving if the person chooses not to. Therefore, we all agree to this solemn pact that encourages the driver to flip the switch. Whichever of us will be on Track B hereby waives his right to life in this circumstance, and will encourage the driver to switch as loudly as all of the rest of us."

If the driver were presented with this pact, it's hard to imagine her not switching to Track B. But if the eleven Trolley Problem candidates were permitted to make such a pact before the dilemma started, it's hard to imagine that they wouldn't. Therefore, the Economists' Paradise assumption of perfect coordination produces the correct utilitarian result to the trolley problem. The same methodology can be extended to utilitarianism in a lot of other contexts.

Now we can go back to that problem from before: what if Nerds have *literally* nothing Jocks want, and Jocks haven't decided among themselves that bullying is a stupid status game that wastes their time, and we're otherwise in the Least Convenient Possible World with regards to stopping bullying. Is there any way assuming Economists' Paradise solves the problem *then*?

Maybe. Just go around to little kids, age two or so, and say "Look. At this point, you really don't know whether you're going to grow up to be a Jock or a Nerd. You want to sign this pact that everyone who grows up to be a Jock promises not to bully everyone who grows up to be a Nerd?" Keeping the same assumption that bullying is on net negative utility, we expect the toddlers to sign. Yeah, in the real world two-year olds aren't the best moral reasoners, but good thing we're in Economists' Paradise where we assume such problems away by fiat.

Is there an Even Less Convenient Possible World? Suppose bullying is racist rather than popularity-based, with all the White kids bullying the Black kids. You go to the toddlers, and the white toddlers retort back "Even at this age, we know very well that we're White, thank you very much."

So just approach them in the womb, where it's too dark to see skin color. If we're letting two year olds sign contracts, why not fetuses?

Okay. One reason might be because we've just locked ourselves into being fanatically pro-life merely by starting with weird assumptions. Another reason might be that we could counterfactually mug fetuses by saying stuff "You're definitely a human, but for all you know the world is ruled by Lizardmen with only a small human slave population, and if Lizardmen exist then they will torture any humans who did not agree in the womb that, if upon being born and finding that Lizardmen did not exist, they would spend all their time and energy trying to create Lizardmen."

(Frick. I think I just created a new basilisk by breeding the Rokolisk and the story of 9-tsiak. Good thing it only works on fetuses.)

(I wonder if this is the first time in history anyone has ever used the phrase “counterfactually mug fetuses” as part of a serious intellectual argument.)

So I’m not saying this theory doesn’t have any holes in it. I’m just saying that it seems, at least in principle, like the idea of Economists’ Paradise might be sufficient to derive Rawls’ Veil of Ignorance, which in turn bridges the chasm that separates it from Utilitarians’ Paradise.

IV.

I think this is the solution to the various questions raised in You Kant Dismiss Universalizability. The reason universalizability is important is that the universal maxims are the agreements that everyone or nearly everyone would sign. This leads naturally to something like utilitarianism for the reasons mentioned in Part III. And it doesn’t produce the weird paradoxes like “If morality is universalizability, how do you know whether a policeman overpowering and imprisoning a criminal universalizes to ‘police should be able to overpower and imprison criminals’ or ‘everyone should be able to overpower and imprison everyone else’?” Everyone would sign an agreement allowing the first, but not the second.

But before we *really* explore this, a few words on “everyone would sign”.

Suppose one very stubborn annoying person in Economists’ Paradise refused to sign an agreement that police should be allowed to arrest criminals. Now what?

“All game theory is solved perfectly” is a *really* powerful assumption, and the rest of the world has a lot of leverage over this one person. Suppose everyone else said “You know, we’re all signing an agreement that none of us are going to murder one another, but we’re not going to let you into that agreement unless you also sign this agreement which is very important to us.”

Actually, that sounds too evil and blackmailing. There’s a better way to think of it. Suppose there are one hundred agreements. 99% of the population agrees to each, and in fact it’s a different 99% each time. That is, divide the population into one hundred sets of 1%, and each set will oppose exactly one of the agreements – there is no one who opposes two or more. Each agreement only works (or works best) when one hundred percent of the population agrees to it.

Very likely everyone will strike a deal that each of the one hundred 1% blocs agrees to to give up its resistance to the one agreement they don’t like, in exchange for each of the other ninety nine 1% blocs giving up its resistance to the agreements *they* don’t like.

Now we’re getting into meta-level Pareto improvements. If a pact would be positive-sum for people to agree on, the proponents of the pact can offer everyone else some compensation for them signing the pact. In theory it could be money or computer-fixing, but it might also be agreement with some of *their* preferred pacts.

There are a few possible outcomes of this process in Platonic Economists’ Paradise, both interesting.

One is a patchwork of agreements, where everyone has to remember that they’ve signed agreements 5, 12, 98, and

12,671, but their next-door neighbor has signed agreements 6, 12, 40, and 4,660,102, so they and their neighbor are bound to cooperate on 12 but no others.

Another is that everyone is able to get their desired pacts to cohere into a single really big pact that they are all able to sign off upon. Maybe there are a few stragglers who reject it at first, but this ends up being a terrible idea because now they're not bound by really important agreements like "don't murder" or "don't steal", so eventually they give in.

A third possibility combining the other two offers a unifying principle behind Whose Utilitarianism and Archipelago and Atomic Communitarianism. Everyone agrees to some very basic principles of respecting one another (call them "Noahide Laws") but smaller communities agree to stricter rules that allow them to do their own thing.

But we don't live in Platonic Economists' Paradise. We live in the real world, where transaction costs are high and people have limited brainpower. Even if we were to try to instantiate Economists' Paradise, it couldn't be the one where we all have the complex interlocking patchwork agreements between one another. People wouldn't sign off on it. Heck, *I* wouldn't sign off on it. I would say "I'm not signing this until I have something that makes sense to me and can be implemented in a reasonable amount of time and doesn't require me to check the List Of Everybody In The World before I know whether the guy next to me is going to murder me or not." Practical concerns provide a very strong incentive to reject the patchwork solution and force everyone to cohere. So in practice – and I realize how hokey it is to keep talking about game-theoretically-perfect infinitely-rational infinitely-honest agents negotiating all possible agreements among one another, and then add on the term "in practice" to represent that they have trouble remembering what they decided – but in practice they would all have very large incentives to cohere upon a single solution that balances out all of their concerns.

We can think of this as moving along an axis from “Platonic” to “practical”. As we progress further, complicated agreements collapse into simpler agreements which are less perfect but easier to enforce and remember. We start to make judicious use of Schelling fences. We move from everyone in the world agreeing on exactly what people can and can’t do to things like “Well, you know your intuitive sense of niceness? You follow that with me, and I’ll follow that with you, and we’ll assume everyone else is in on the deal until they prove they aren’t.”

A metaphor: in a dream, your soul goes to Economists’ Paradise and agrees on the perfect patchwork of maxims with all the other souls there. But as dawn approaches, you realize when you awaken you will never remember all of what you agreed upon, and even worse, all the other souls there are going to wake up and not remember what *they* agreed upon either. So all of you together frantically try to compress your wisdom into a couple of sentences that the waking mind will be able to recall and follow, and you end up with platitudes like “Use your intuitive sense of niceness” and “do unto others as you would have others do unto you” and “try to maximize utility” and “anybody who treats you badly, assume they’re not in on the deal and feel free to treat them badly too, but not so badly that you feel like you can murder them or something.”

A particularly good platitude/compression might be “Work very hard to cultivate the mysterious skill of figuring out what people in the Economists’ Paradise would agree to, then do those things.” If you’re Greek, you can even compress it into a single word: *phronesis*.

V.

So by now it’s probably pretty obvious that this is an attempt to ground morality. I think the general term for the philosophical school involved is “contractualism”.

Many rationalists seem to operate on something like R.M. Hare’s two-level utilitarianism. That is, utilitarianism is the correct base level of morality, but it’s very hard to do, so in reality you’ve got to make do with less precise but more computationally tractable heuristics, like deontology and virtue ethics. Occasionally, when deontology or virtue ethics contradict themselves, each other, or your intuitions, you may have to sit down and actually do the utilitarianism as best you can, even though it will be inconvenient and very philosophically difficult.

For example, deontology may say things like “You must never kill another human being.” But in the trolley problem, the correct deontological action seems to violate our moral intuitions. So we go up a level, calculate the utility (which in this case is very easy, because it’s a toy problem invented entirely for the purposes of having easy utility calculation) and say “Huh, this appears to be one of those rare places where our deontological heuristics go wrong.” Then you switch the trolley.

But utilitarianism famously has problems of its own. You need a working definition of utility, which means not only distinguishing between hedonic utilitarianism, preference utilitarianism, etc, but coming up with a consistent model for measuring the strength of happiness and preferences. You need to distinguish between total utilitarianism, average utilitarianism, and a couple of other options I forget right now. You need a discount rate. You need to know whether creating new people counts as a utility gain or not, and whether removing people (isn't *that* a nice euphemism) can even be counted as a negative if you make sure to do it painlessly and without any grief to those who remain alive. You need a generalized solution to Pascal's Wagers and utility monsters. You need to know whether to accept or fudge away weird results like that you may be morally obligated to live your entire life to maximize anti-malaria donations. All of this is easy at the tails and near-impossible at the margins.

My previous philosophy was "Yeah, it's hard, but I bet with sufficient intelligence, we can think up a consistent version of utilitarianism with enough epicycles that it produces an answer to all of these issues that most people would recognize as at least kind of sane. Then we can just go with that one."

I still believe this. But that consistent version would probably fill a book. The question is: what is the person who decides what to put in this book doing? On what grounds are they saying "total utilitarianism is a better choice than average utilitarianism"? It can't be on *utilitarian* grounds, because you can't use utilitarian grounds until you've figured out utilitarianism, which you haven't done until you've got the book. When God was deciding what to put in the Bible, He needed some criteria other than "make the decision according to Biblical principles".

The standard answer is "we are starting with our moral intuitions, then simplifying them to a smaller number of axioms which eventually produce them". But if the axioms fill a book and are full of epicycles to address individual problems, we're not doing a very good job.

I mean, it's still better than just trying to sort out all individual issues like "what is a just war?" on their own, because people will answer that question according to their personal prejudices (is my tribe winning it? Then it is *so, so just*) and if we force them to write the utilitarianism book at least they've got to come up with consistent principles and stick to them. But it is *highly suboptimal*.

And I wonder whether maybe the base level, the one that actually grounds utilitarianism, is contractualism. The idea of a Platonic parliament in which we try to enact all beneficial agreements. Under this model, utilitarianism, deontology, and virtue ethics would all be *different* heuristics that we use to approximate contractualism, the fragments we remember from our beautiful dream of Paradise.

I realize this is kind of annoying, especially in the sense of "the next person who comes along can say that utiltiarianism, deontology, virtue ethics, *and contractualism* are heuristics for whatever moral theory *they* like, which is The Real Thing". But the idea can do work! It particular, it might help esolve some of the standard paradoxes of utilitarianism.

First, are we morally obligated to wirehead everyone and convert the entire universe into hedonium? Well, would *you* sign that contract?

Second, is there anything wrong with killing people painlessly if they won't be missed? After all, it doesn't seem to cause any pain or suffering, or even violate any preferences – at least insofar as your victim isn't around to have their preferences violated. Well, would you sign a contract in which everyone agrees not to do that?

Third, are we morally obligated to create more and more people with slightly above zero utility, until we are in an overcrowded slum world with everyone stuck at just-above-subsistence level (the Repugnant Conclusion)? Well, if you were making an agreement with everyone else about what the population level should be, would you suggest we do that? Or would you suggest we avoid it?

(this can be complicated by asking whether potential people get a seat in this negotiation, but Carl Shulman has a neat way to solve that problem)

Fourth, the classic problem of defining utility. If utility can be defined ordinally but not cardinally (ie you can declare that stubbing your toe is worse than a dust speck in the eye, but you can't say something like it's exactly 2.6 negative utilons) then utilitarianism becomes very hard. But contractualism doesn't become any harder, except insofar as it's harder to use utilitarianism as a heuristic for it.

I am not actually sure these problems are being solved, and I'm not just being led astray by contractualism being harder to model than utilitarianism and so it is easier for me to *imagine* them solved. But at the very least, it might be that contractualism is a different angle from which to attack these problems.

Of course, contractualism has problems of its own. It might be that different ways of doing the negotiations would lead to very different results. It might also be that the results would be very path-dependent, so that making one agreement first would end with a totally different result than making another agreement first. And this would be a good time to admit I don't know that much formal game theory, but I do know there are multiple Nash equilibria and Pareto-optimal endpoints in a lot of problems and that in general there's no such thing as "the correct game theoretic solution to this problem”, only solutions that fit more or fewer desirability criteria.

But to some degree this maps onto our intuitions about morality. One of the harder to believe things about utilitarianism was that it suggested there was exactly one best state of the universe. Our intuitions are very good at saying that certain hellish dystopias are very bad, and certain paradises are very good, but extrapolating them out to say there’s a single best state is iffy at best. So maybe the ability of rigorous game theory to end in a multitude of possible good outcomes is a feature and not a bug.

I don’t know if it’s possible for certain negotiation techniques to end in extreme local minima where things don’t end out as a paradise *at all*. I mean, I know there’s lots of horrible game theory like the Prisoner’s Dilemma and the Pirate’s Dilemma and so on, but I’m defining the “good game theory” of the Economists’ Paradise to mean exactly the rules and coordination power you need to not do those kinds of things.

But there’s also a meta-level escape vent. If a certain set of negotiation techniques would lead to a local minimum where everything is Pareto-optimal but nobody is happy, then everyone would coordinate to sign a pact *not to use those negotiation techniques*.

VI.

To sum up:

The Economists’ Paradise of solved coordination problems would be enough to keep everyone happy and prosperous and free. We ourselves could live in that paradise if we followed its rules, which involve negotiation of and adherence to agreements according to good economist and game theory, but these rules are hard to determine and hard to enforce.

We can sort of guess at what some of these rules can be, and when we do that we can try to follow them. Some rules lend themselves to State enforcement. Others don't and we have to follow them quietly in the privacy of our own hearts. Sometimes the rules include rules about ostracizing or criticizing those who don't follow the rules effectively, and so even the ones the State can't enforce are sorta kinda enforceable. Then we can spread them through a series of walled gardens and ~~spontaneous order~~ divine intervention.

The exact nature of the rules is computationally intractable and so we use heuristics most of the time. Through practical wisdom, game theory, and moral philosophy, we can improve our heuristics and get to the rules more closely, with corresponding benefits for society. Utilitarianism is one especially good heuristic for the rules, but it's *also* kind of computationally intractable. Utilitarianism helps us approximate contractualism, and contractualism helps us resolve some of the problems of utilitarianism.

One problem of utilitarianism I didn't talk about is that it isn't very inspirational. Following divine law is inspirational. Trying to become a better person, a heroic person, is inspirational. Utilitarianism sounds too much like *math*. I think contractualism solves this problem too.

Consider. There is an Invisible Nation. It is not a democracy, per se, but it is something of a republic, where each of us is represented by a wiser, stronger version of ourselves who fights for our preferences to be enacted into law. Its legislature is untainted by partisanship, perfectly efficient, incorruptible, without greed, without tyranny. Its bylaws are the laws of mathematics; its Capitol Building stands at the center of Platonia.

All good people are patriots of the Invisible Nation. All the visible nations of the world – America, Canada, Russia – are properly understood to be its provinces, tasked with executing its laws as best they can, and with proper consideration to the unique needs of the local populace. Some provinces are more loyal than others. Some seem to be in outright rebellion. The laws of the Invisible Nation contain provisions about what to do with provinces in rebellion, but they are vague and difficult to interpret, and its patriots can disagree on what they are.

Maybe one day we will create a superintelligence that tries something like Coherent Extrapolated Volition – which I think we have just rederived, kind of by accident. The various viceroys and regents will hand over their scepters, and the Invisible Nation will stand suddenly revealed to the mortal eye. Until then, we see through a glass darkly. As we learn more about our fellow citizens, as we gain new modalities of interacting with them like writing, television, the Internet – as we start crystallizing concepts like rights and utility and coordination – we become a little better able to guess.

If The Media Reported On Other Dangers Like It Does AI Risk

[Not actually inspired by Robert Wiblin's recent Facebook post on this same topic, but I acknowledge the coincidence. The media has actually done a much better job than I expected here and deserves some credit, but I will snark anyway.]

It's a classic staple of action movies and Tom Clancy thrillers – the Islamic terrorist group that takes over a failed state, forcing the heroes to mount a decisive response. But some geopolitics experts think such a scenario could soon move from political fiction…to political fact.

If carbon dioxide levels reach 500 parts per million, it could initiate dangerous "runaway global warming". But more conservative scientists urge laypeople not to worry, noting "Carbon dioxide levels are not that high yet."

A sufficiently large nuclear war could completely destroy human civilization. If the bombs struck major manufacturing centers, they could also cause thousands of people to be put out of work.

Remember that time your boss paid you a few days late? Or that time the supermarket stopped carrying your favorite brand of cookie? Then you might not be surprised to hear many analysts believe the world economy will crash causing a giant decades-long depression.

[An informative, scientifically rigorous explanation of the dangers of climate change, but the picture on the top is that image of the Statue of Liberty buried in ice from *The Day After Tomorrow*]

A giant asteroid could smash into Earth at any time, scientists say. Indeed, already we are having to deal with avalanches and landslides that have blocked several major roads. Geologists think stabilizing our nation’s cliff faces may be the answer.

A group of meteorology nerds have sounded the alarm that a major hurricane could form in the next week – and now they’re turning their giant brains to the question of where it will make landfall.

Tacticians worry Russia might invade Ukraine – for example, they could choose to paradrop the 5th Battalion in under cover of night. But our experts say that that the 5th Battalion is not capable of night-time paradrops. Therefore, Russia will not be invading Ukraine.

The new superplague is said to be 100% fatal, totally untreatable, and able to spread across an entire continent in a matter of days. It is certainly fascinating to think about if your interests tend toward microbiology, and we look forward to continuing academic (and perhaps popular) discussion and debate on the subject.

Michigan Meetup 9/7

There will be a Michigan rationalist/LW/SSC meetup at my house, halfway between Detroit and Ann Arbor, on Sunday, September 7th at 2 PM. You can find my address and directions here. If you’re reading this, you’re invited.

Topic will be “We Should Really Get Formal Topics For These Things One Of These Days”

Our special guest will be Miranda (Swimmer963 on LW), sometime CFAR coordinator and generally cool person.

Radicalizing the Romanceless

[PLEASE NOTE: This post is from 2014.]

[Content note: Gender, relationships, feminism, manosphere. Quotes, without endorsing and with quite a bit of mocking, mean arguments by terrible people. Some analogical discussion of fatphobia, poorphobia, Islamophobia. This topic is enraging to me and I don't promise to be able to treat it fairly.]

I.

> "How dare you compare x to y" it was pretty easy actually
>
> — HRH Misha (@drethelin) August 22, 2014

I recently had a patient, a black guy from the worst part of Detroit, let's call him Dan, who was telling me of his woes. He came from a really crappy family with a lot of problems, but he was trying really hard to make good. He was working two full-time minimum wage jobs, living off cheap noodles so he could save some money in the bank, trying to scrape a little bit of cash together. Unfortunately, he'd had a breakdown (see: him being in a psychiatric hospital), he was probably going to lose his jobs, and everything was coming tumbling down around him.

And he was getting a little philosophical about it, and he asked – I'm paraphrasing here – why haven't things worked out for me? I'm hard-working, I've never missed a day of work until now, I've always given a hundred and ten percent. And meanwhile, I see all these rich white guys ("no offense, doctor," he added, clearly overestimating the salary of a medical resident) who kind of coast through school, coast into college, end up with 9 – 4 desk jobs working for a friend of their father's with excellent salaries and benefits, and if they need to miss a couple of days of work, whether it's for a hospitalization or just to go on a cruise, nobody questions it one way or the other. I'm a harder worker than they are, he said – and I believed him – so how is that fair?

And of course, like most of the people I deal with at my job, there's no good answer except maybe restructuring society from the ground up, so I gave him some platitudes about how it's not his fault, told him about all the social services available to him, and gave him a pill to treat a biochemical condition almost completely orthogonal to his real problem.

And I'm still not sure what a *good* response to his question would have been. But later that night I was browsing the Internet and I was reminded of what the *worse response humanly possible*. It would go something like:

You keep whining about how "unfair" it is that you can't get a good job. "But I'm such a hard worker." No, actual hard workers don't feel like they're entitled to other people's money just because they ask nicely.

"Why do rich white kids who got legacy admissions to Yale receive cushy sinecures, but I have to work two grueling minimum wage jobs just to keep a roof over my head?" By even asking that question, you prove that you think of bosses as giant bags of money, rather than as individual human beings who are allowed to make their own choices. No one "owes" you money just because you say you "work hard", and by complaining about this you're proving you're not really a hard worker at all. I've seen a lot of Hard Workers (TM) like you, and scratch their entitled surface and you find someone who thinks just because they punched a

time card once everyone needs to bow down and worship them.

If you complain about "rich white kids who get legacy admissions to Yale," you're raising a huge red flag that you're the kind of person who steals from their employer, and companies are exactly right to give you a wide berth.

Such a response would be so antisocial and unjust that it could only possibly come from the social justice movement.

II.

I've been thinking about "nice guys" lately for a couple of reasons.

First, I read *Alas, A Blog*'s recent post on the subject, MRAs And Anti-Feminists Have Ruined Complaining About Being Single.

Second, I had yet *another* patient who –

(I feel obligated to say at this point that the specific details of these patient stories are made up, and several of them are composites of multiple different people, in order to protect confidentiality. I'm preserving the general gist, nothing more)

– I had a patient, let's call him 'Henry' for reasons that are to become clear, who came to hospital after being picked up for police for beating up his fifth wife.

So I asked the obvious question: "What happened to your first four wives?"

"Oh," said the patient, "Domestic violence issues. Two of them left me. One of them I got put in jail, and she'd moved on once I got out. One I just grew tired of.”

“You’ve beaten up all five of your wives?” I asked in disbelief.

“Yeah,” he said, without sounding very apologetic.

“And why, exactly, were you beating your wife this time?” I asked.

“She was yelling at me, because I was cheating on her with one of my exes.”

“With your ex-wife? One of the ones you beat up?”

“Yeah.”

“So you beat up your wife, she left you, you married someone else, and then she came back and had an affair on the side with you?” I asked him.

“Yeah,” said Henry.

I wish, I wish I wish, that Henry was an isolated case. But he’s interesting more for his anomalously high number of victims than for the particular pattern.

Last time I talked about these experiences, one of my commenters linked me to what was later described as the only Theodore Dalrymple piece anyone ever links to. Most of the commenters saw a conservative guy trying to push an ideological point, and I guess that’s part of it. But for me it looked more like the story of a psychiatrist from an upper-middle-class background suddenly realizing how dysfunctional and screwed-up a lot of his patients are and having his mind recoil in horror from the fact – which is something I can sympathize with. Henry was the worst of a bad bunch, but nowhere near unique.

When I was younger – and I mean from teenager hood all the way until about three years ago – I was a ‘nice guy’. And I said the same thing as every other nice guy, which is “I am a nice guy, how come girls don’t like me?”

There seems to be some confusion about this, so let me explain what it means, to everyone, for all time.

It does not mean “I am nice in some important cosmic sense, therefore I am entitled to sex with whomever I want.”

It means: “I am a nicer guy than Henry.”

Or to spell it out very carefully, Henry clearly has no trouble attracting partners. He’s been married five times and had multiple extra-marital affairs and pre-marital partners, many of whom were well aware of his past domestic violence convictions and knew exactly what they were getting into. Meanwhile, here I was, twenty-five years old, never been on a date in my life, every time I ask someone out I get laughed at, I’m constantly teased and mocked for being a virgin and a nerd whom no one could ever love, starting to develop a serious neurosis about it.

And here I was, tried my best never to be mean to anyone, pursued a productive career, worked hard to help all of my friends. I didn’t think I deserved to have the prettiest girl in school prostrate herself at my feet. But I did think I deserved to *not be doing worse than Henry*.

No, I didn’t know Henry at the time. But everyone knows *a* Henry. Most people know several. Even three years ago, I knew there were Henry-like people – your abusers, your rapists, your bullies – and it wasn’t hard to notice that none of *them* seemed to be having the crushing loneliness problem I was suffering from.

And, like my patient Dan, I just wanted to know – how is this fair?

And I made the horrible mistake of asking this question out loud, and that was how I learned about social justice.

III.

We will now perform an ancient and traditional Slate Star Codex ritual, where I point out something I don’t like about feminism, then everyone tells me in the comments that no feminist would ever do that and it’s a dirty rotten straw man. And then I link to two thousand five hundred examples of feminists doing exactly that, and then everyone in the comments No-True-Scotsmans me by saying that that doesn’t count and those people aren’t representative of feminists. And then I find two thousand five hundred more examples of the most prominent and well-respected feminists around saying exactly the same thing, and then my commenters tell me that they don’t count either and the only *true* feminist lives in the Platonic Realm and expresses herself through patterns of dewdrops on the leaves in autumn and everything she says is unspeakably kind and beautiful and any time I try to make a point about feminism using examples from anyone other than her I am a dirty rotten motivated-arguer trying to weak-man the movement for my personal gain.

Ahem.

From Jezebel, “Why We Should Mock The Nice Guys Of OKCupid”:

> Pathetic and infuriating in turns, the profiles selected for inclusion [on a site that searches OKCupid profiles for ones that express sadness at past lack of romantic relationships, then posts them publicly for mockery] elicit gasps and giggles – and they raise questions as well. Is it right to mock these aggrieved and clueless young men, particularly the ones who seem less enraged than sad and bewildered at their utter lack of sexual success?
>
> What’s on offer isn’t just an opportunity to snort derisively at the socially awkward; it’s a chance to talk about the very real problem of male sexual entitlement. The great unifying theme of the curated profiles is indignation. These are young men who were told that if they were nice, then, as Laurie Penny puts it, they feel that women “must be obliged to have sex with them.” The subtext of virtually all of their profiles, the mournful and the bilious alike, is that these young men feel cheated. Raised to believe in a perverse social/sexual contract that promised access to women’s bodies in exchange for rote expressions of kindness, these boys have at least begun to learn that there is no Magic Sex Fairy. And while they’re still hopeful enough to put up a dating profile in the first place, the Nice Guys sabotage their chances of ever getting laid with their inability to conceal their own aggrieved self-righteousness.
>
> So how should we respond, when, as Penny writes, “sexist dickwaddery puts photos on the internet and asks to be loved?” The short answer is that a lonely dickwad is still a dickwad; the fact that these guys are in genuine pain makes them more rather than less likely to mistreat the women they encounter.

From XOJane, Get Me Away From Good Guys:

Let’s tackle those good guys. You know, the aw shucks kind who say it’s just so hard getting a date or staying in a relationship, and they can’t imagine why they are single when they are, after all, such catches. They’re sensitive, you know. They totally care about the people around them, would absolutely rescue a drowning puppy if they saw one.

Why is it that so many “good guys” act like adult babies, and not in a fetish sense? They expect everyone else to pick up their slack, they’re inveterately lazy, and they seem genuinely shocked and surprised when people are unimpressed with their shenanigans. Their very heteronormativity betrays a shockingly narrow view of the world; ultimately, everything boils down to them and their needs, by which I mean their penises.

The nice guy, to me, is like the “good guy” leveled up. These are the kinds of people who say that other people just don’t understand them, and the lack of love in their lives is due to other people being shitty. Then they proceed to parade hateful statements, many of which are deeply misogynist, to explain how everyone else is to blame for their failures in life. A woman who has had 14 sexual partners is a slut. These are also the same guys who do things like going into a gym, or a school, or another space heavily populated by women, and opening fire. Because from that simmering sense of innate entitlement comes a feeling of being wronged when he doesn’t get what he wants, and he lives in a society where men are “supposed” to get what they want, and that simmer can boil over.

I’ve noted, too, that this kind of self-labeling comes up a lot in men engaging in grooming behavior. As part of their work to cultivate potential victims, they remind

> their victims on the regular that they’re “good guys” and the only ones who “truly” understand them.

From Feminspire, Nice Guy Syndrome And The Friend Zone:

> I’m pretty sure everyone knows at least one Nice Guy. You know, those guys who think women only want to date assholes and just want be friends with the nice guys. These guys are plagued with what those of us who don’t suck call Nice Guy Syndrome.
>
> It’s honestly one of the biggest loads of crap I’ve ever heard. Nice Guys are arrogant, egotistical, selfish douche bags who run around telling the world about how they’re the perfect boyfriend and they’re just so nice. But you know what? If these guys were genuinely nice, they wouldn’t be saying things like “the bitch stuck me in the friend zone because she only likes assholes.” Guess what? If she actually only liked assholes, then she would likely be super attracted to you because you are one.
>
> Honestly. Is it really that unbearable to be friends with a person? Women don’t only exist to date or have sex with you. We are living, thinking creatures who maybe—just maybe—want to date and sex people we’re attracted to. And that doesn’t make any of us bitches. It makes us human.

From feministe, “Nice Guys”:

> If a self-styled “Nice Guy” complains that the reason he can’t get laid is that women only like “jerks” who treat them badly, chances are he’s got a sense of entitlement on him the size of the Unisphere.

> Guys who consider themselves “Nice Guys” tend to see women as an undifferentiated mass rather than as individuals. They also tend to see possession of a woman as a prize or a right…
>
> A Nice Guy™ will insist that he’s doing everything perfectly right, and that women won’t subordinate themselves to him properly because he’s “Too Nice™,” meaning that he believes women deserve cruel treatment and he would like to be the one executing the cruelty.

But Feministe is also the first to show a glimmer of awareness (second, if you count Jezebel’s “I realize this might be construed as mean BUT I LOVE BEING MEAN” as “awareness”):

> For the two hundredth time, when we’re talking about “nice guys,” we’re not talking about guys who are actually nice but suffer from shyness. That’s why the scare quotes. Try Nice Guys instead, if you prefer.
>
> A shy, but decent and caring man is quite likely to complain that he doesn’t get as much attention from women as he’d like. A Nice Guy™ will complain that women don’t pay him the attention he deserves. The essence of the distinction is that the Nice Guy™ feels women are obligated to him, and the Nice Guy™ doesn’t actually respect or even like women. The clearest indication of which of the two you’re dealing with is whether the person is interested in the possibility that he’s doing something wrong.

Okay. Let’s extend our analogy with Dan from above.

It was wrong of me to say I hate poor minorities. I meant I hate Poor Minorities! Poor Minorities is a category I made up that includes only poor minorities who *complain* about poverty or racism.

No, wait! I can be even *more* charitable! A poor minority is only a Poor Minority if their compaints about poverty and racism *come from a sense of entitlement*. Which I get to decide after listening to them for two seconds. And If they don't realize that they're doing something wrong, then they're automatically a Poor Minority.

I dedicate my blog to explaining how Poor Minorities, when they're complaining about their difficulties with poverty or asking why some people like Paris Hilton seem to have it so easy, really just want to steal your company's money and probably sexually molest their co-workers. And I'm not being unfair at all! Right? Because of my new definition! I know everyone I'm talking to can hear those Capital Letters. And there's no chance whatsoever anyone will accidentally misclassify any *particular* poor minority as a Poor Minority. That's crazy talk! I'm sure the "make fun of Poor Minorities" community will be diligently self-policing against *that* sort of thing. Because if anyone is known for their rigorous application of epistemic charity, it is the make-fun-of-Poor-Minorities community!

I'm not even sure I can dignify this with the term "motte-and-bailey fallacy". It is a tiny Playmobil motte on a bailey the size of Russia.

I don't think I ever claimed to be, or felt, entitled to anything. Just wanted to know why it was that people like Henry could get five wives and I couldn't get a single date. That was *more* than enough to get the "shut up you entitled rapist shitlord" cannon turned against me, with the person who was supposed to show up to give me the battery of tests to distinguish whether I was a poor minority or a Poor

Minority nowhere to be seen. As a result I spent large portions of my teenage life traumatized and terrified and self-loathing and alone.

Some recent adorable Tumblr posts (1, 2) pointed out that not everyone who talks about social justice is a social justice warrior. There are also “social justice clerics, social justice rogues, social justice rangers, and social justice wizards”. Fair enough.

But there are also social justice chaotic evil undead lich necromancers.

And the people who talk about “Nice Guys” – and the people who enable them, praise them, and link to them – ~~are blurring the already rather thin line between “feminism” and “literally Voldemort”~~ EDIT: ARE TOTALLY GREAT, NO NEED TO TAKE THIS ONE SENTENCE OUT OF CONTEXT AND TRY TO SPREAD IT ALL OVER THE INTERNET.

IV.

And so we come to Barry’s recent blog post:

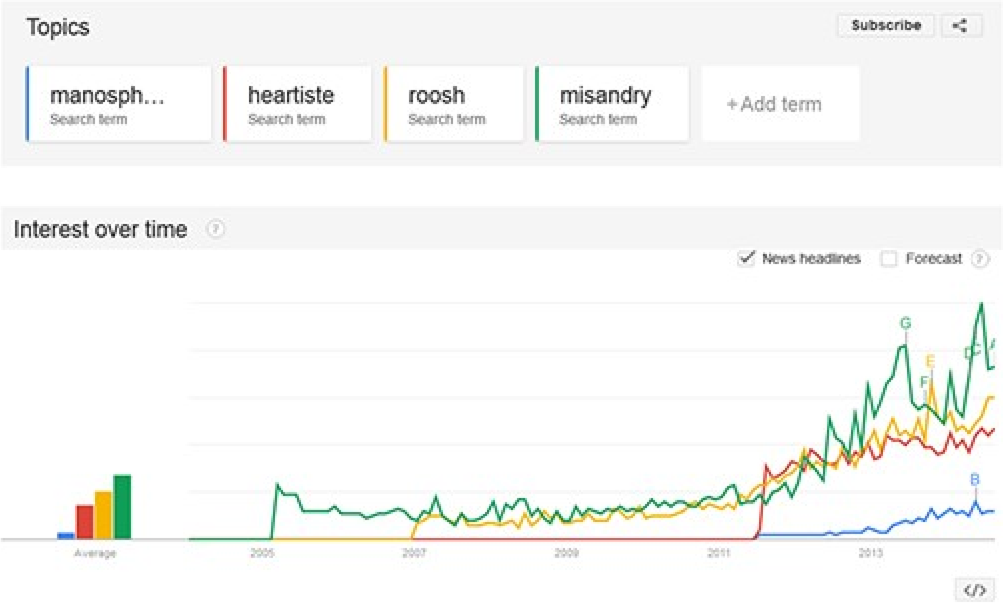
> In pop culture, everyone – or at least, everyone who isn’t a terrible human being – eventually meets someone wonderful and falls in love.
>
> But in real life, that’s not how things always work. Some people don’t want romantic love at all. Others want romantic love but will never find it. That’s life. I’m beginning to accept, at age 45, that probably “true love” will never happen for me. I have a bunch of factors working against me – I’m physically conventionally unattractive, I badly lack confidence, I’m sort of a weirdo, as I get older I meet new people less often, etc..

> To tell you the truth, I resent the situation. It’s not an all-consuming bitterness or anything – on the whole, I’m a happy guy – but I irrationally feel cheated of a fundamental human experience…
>
> I bring this up because I feel my ability to enjoy complaining about my single state has been ruined by MRAs and anti-feminists.
>
> Because in human culture, we do something called “signaling” a lot. And, on the internet, men complaining that they don’t have the romantic success they want, that they feel they should be more attractive to woman then they actually are in practice, etc., have all become signals used to indicate alliance with the manosphere.
>
> Gore Vidal once groused that the once-useful word “turgid” now belongs to the porn writers, because it has become impossible to use the word without sounding like a porn writer. The manosphere has done something similar to unattractive men’s romantic problems. They’ve flooded the discourse with misogyny and anti-feminism, and it’s nearly impossible to rescue discussion of being male and unwanted from their bitter waters.

Let me start by saying I sympathize with Barry, as someone who has been in exactly his position. And that if anyone uses this post as an excuse to attack Barry personally, they are going to Hell *and* getting banned from SSC. They’re also proving the point of whichever side they are not on.

What I don’t sympathize with is Barry’s belief that this is somehow the fault of “the manosphere” “flooding the discourse”.

It would actually be pretty fun to go full internet-archaeologist on the manosphere, but a quick look confirms my impression that, although it is built from older pieces, it's really quite young. There was a "men's rights" movement around forever, but its early focus tended to be on divorce cases and fathers' rights. Heartiste started publishing in 2007. The word "manosphere" was first used in late 2009. Google Trends confirms a lot of this.



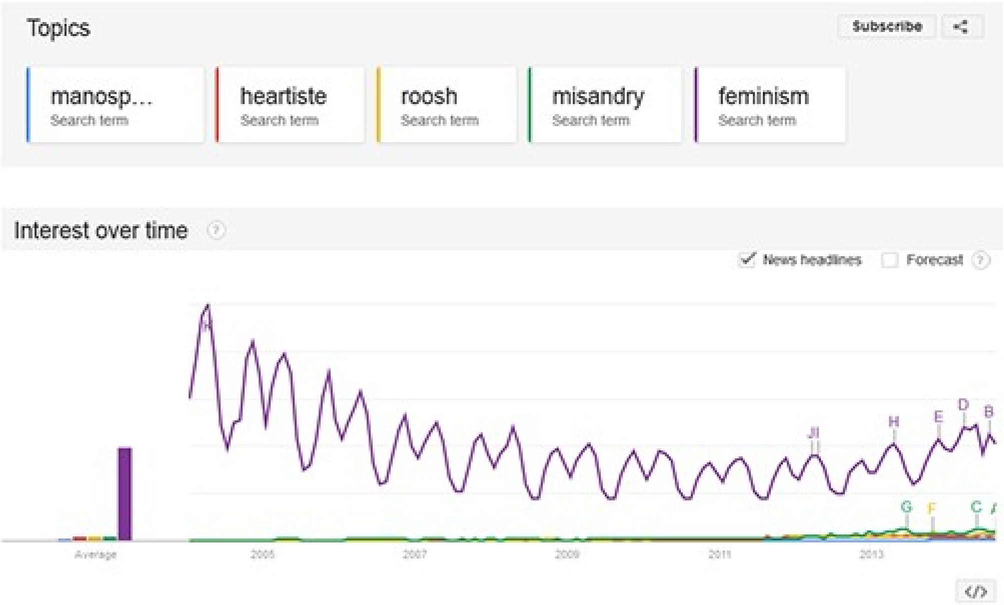
So I think it's fair to attribute low to minimal influence for Manosphere-type stuff before about 2005 at the earliest.

But feminists were complaining about "nice guys" for much longer. According to Wikipedia, the concept dates at least from a 2002 article called [Why "Nice Guys" are often such LOSERS](), which was billed as a "Bitchtorial" on feminist blog "Heartless Bitches International"

(Once again, I swear I don't make up the names of these feminist blogs as some sort of strawmanning strategy. They just *happen* like that!)

Looking into “Heartless Bitches Internation”, its header image is the words “Nice Guys = Bleah!” and its blog tagline is “What’s wrong with Nice Guys? HBI Tells It Like It Is”. This was seven years before the term “manosphere” even existed.

I can’t Google Trends “Nice Guys”, because it picks up too much interference from normal discussion of people who are nice. But there *is* one more Google Trends graph that I think relates to this issue:



This is the same graph as before. You can’t tell, because I’ve added the word “feminism”, which has caused every other line on the graph to shrink into invisibility. The purple line is – what, twenty times as high as any of the others?

People were coming up with reasons to mock and despise men who were sad about not being in relationships *years* before the manosphere even existed. These reasons were being posted on top feminist blogs for years without any reference whatsoever to the manosphere, probably because the people who wrote them were unaware of its existence or couldn't imagine what it could possibly have to do with this subject? Feminism – the movement that was doing all this with no help from the manosphere – has twenty times the eyeballs and twenty times the discourse-setting power as the manosphere. And Barry thinks this is the manosphere's fault? On the SSC "Things Feminists Should Not Be Able To Get Away With Blaming On The Manosphere" Scale, this is right up there with the postulated link between the men's rights movement and Putin's invasion of Ukraine.

The worst corners of the manosphere contain more than enough opining on how ugly women, weird women, masculine women, et cetera deserve to be unhappy. You are welcome to read, for example, Matt Forney's Why Fat Women Don't Deserve To Be Loved (part of me feels like the link is self-trigger-warning, but I guess I will just warn you that this is not a clever attention-grabbing title, the link means exactly what it says and argues it at some length)

No one says the only reason manospherites like to insult unattractive lonely women is because "it's hard for women to complain about how they're single without being mistaken for a feminist", or that "the manosphere doesn't mean all lonely women, it's just talking about how offended they are that lonely women feel entitled to sex and objectify men". *In the case of men*, everyone pretty much agrees that no, if you're a certain kind of person, making fun of people for being unattractive and unhappy is its own reward.

And I don't think there is some sort of deep genetic personality difference between the sexes that makes them do things for totally different reasons. For women just as well as men, for feminists just as well as manospherites, *if you're a certain kind of person, making fun of people for being unattractive and unhappy is its own reward*. Hence everything that has ever been said about "nice guys (TM)"

The only difference between the feminists and the manosphere here is that people call out the manosphere when they do it. But the feminists have their little Playmobil motte, so that’s *totally different*!

V.

So am I claiming that the feminist war on “nice guys” is totally uncorrelated with the existence of the manosphere?

No. I’m saying the causal arrow goes the opposite direction from the one Barry’s suggesting. As usual with gender issues, this can be best explained through a story from ancient Chinese military history.

Chen Sheng was an officer serving the Qin Dynasty, famous for their draconian punishments. He was supposed to lead his army to a rendezvous point, but he got delayed by heavy rains and it became clear he was going to arrive late. The way I always hear the story told is this:

Chen turns to his friend Wu Guang and asks “What’s the penalty for being late?”

“Death,” says Wu.

“And what’s the penalty for rebellion?”

“Death,” says Wu.

“Well then…” says Chen Sheng.

And thus began the famous Dazexiang Uprising, which caused thousands of deaths and helped usher in a period of instability and chaos that resulted in the fall of the Qin Dynasty three years later.

The moral of the story is that if you are maximally mean to innocent people, then eventually bad things will happen to you. First, because you have no room to punish people any more for *actually* hurting you. Second, because people will figure if they're doomed anyway, they can at least get the consolation of feeling like they're doing you some damage on their way down.

This seems to me to be the position that lonely men are in online. People will tell them they're evil misogynist rapists – as the articles above did – no matter what. In what is apparently *shocking news* to a lot of people, this makes them hurt and angry. As someone currently working on learning psychotherapy, I can confidently say that receiving a constant stream of hatred and put-downs throughout your most formative years can *really screw you up*. And so these people try to lash out at the people who are doing it to them, secure in the knowledge that there's no room left for people to hate them even *more*.

I know this is true because it happened to me. I never became a manospherian per se, because two wrongs don't make a right, but – as readers of this essay may be surprised to learn – I did become just a *little* bit bitter about feminism. If I hadn't been so sure about that "two wrongs" issue I probably would have ended up a lot more radicalized.

Actually, that word – "radicalized" – conceals what is basically my exact thesis. We talk a lot about the "radicalization" of Muslims – for example, in Palestine. And indeed, nobody likes Hamas and we all agree they are terrible people and commit some terrible atrocities. Humans can certainly be very cruel, but there seems to be an unusual amount of cruelty in this particular region. And many people who like black-and-white thinking try to blame that on some defect in the Palestinian race, or claim the

Quran urges Muslims should be hateful and violent. But if you’re willing to tolerate a little bit more complexity, it may occur to you to ask “Hey, I wonder if any of this anger among Palestinians has to do with the actions of *Israel*?” And then you might notice, for example, the past century of Middle Eastern history.

Yet somehow, when the manosphere is being terrible people and commiting terrible atrocities, the only explanation offered is that “you must hate all women” must appear in some sura of the Male Quran.

My patient – not Henry, the one I started this whole thing off with, the one who works two minimum wage jobs and wants to know why he’s still falling behind when everyone else does so well – he wasn’t listed as a danger to himself or others, so he had the right to leave the hospital voluntarily if he wanted to. And he did, less than two days after he came in, before we’d even managed to finalize a treatment plan for him. He was worried that his boss was going to fire him if he stayed in longer.

I didn’t get a chance to give him any medication – not that it would have helped that much. All I got a chance to do was to tell him I respected his situation, that he was in a really sucky position, that it wasn’t his fault, and that I hoped he did better. I’m sure my saying that had minimal effect on him. But maybe a history of getting to hear that message from all different people – friends, family, doctors, social workers, TV, church, whatever – all through his life – gave him enough mental fortitude to go back to his horrible jobs and keep working away in the hopes that things would get better. Instead of killing himself or turning to a life of crime or joining the latest kill-the-rich demagogue movement or whatever.

In the end what he wanted wasn’t entitlement to other people’s money, or a pity job from someone who secretly didn’t like him. All he needed to keep going was to have people *acknowledge there was a problem and treat him like a frickin’ human being*.

VI.

So let’s get back to Barry.

(remember, anyone who uses this article to insult Barry will go to Hell *and* get banned from Slate Star Codex)

Barry is using my second-favorite rhetorical device, apophasis, the practice of bringing up something by denying that it will be brought up. For example, “I think the American people deserve a clean debate, and that’s why I’m going to stick to the issues, rather than talking about the incident last April when my opponent was caught having sex with a goat. Anyway, let’s start with the tax rate…”

He is complaining about being single by saying that you can’t complain about being single – and, as a bonus, placating feminists by blaming the whole thing on the manosphere as a signal that he’s part of their tribe and so should not be hurt.

It almost worked. He only got one comment saying he was privileged and entitled (which he dismisses as hopefully a troll). But he did get some other comments that remind me of two of my other least favorite responses to “nice guys”.

First: “Nice guys don’t want love! They just want sex!”

One line disproof: if they wanted sex, they’d give a prostitute a couple bucks instead of spiralling into a giant depression.

Second: “You can’t compare this to, like, poor people who complain about being poor. Food and stuff are basic biological human needs! Sex isn’t essential for life! It’s an extra, like having a yacht, or a pet tiger!”

I know that feminists are not always the biggest fans of evolutionary psychology. But I feel like it takes a *special* level of unfamiliarity with the discipline to ask “Sure, evolution gave us an innate desire for material goods, but *why would it give us an deep innate desire for pair-bonding and reproduction*??!”

But maybe a less sarcastic response would be to point out Harry Harlow’s monkey studies. These studies – many of them so spectacularly unethical that they helped kickstart the modern lab-animals’-rights movement – included one in which monkeys were separated from their real mother and given a choice between two artifical “mothers” – a monkey-shaped piece of wire that provided milk but was cold and hard to the touch, and a soft cuddly cloth mother that provided no milk. The monkeys ended up “attaching” to the cloth mother and not the milk mother.

In other words – words that shouldn’t be surprising to anyone who has spent much time in a human body – companionship and warmth can be in some situations just as important as food and getting your more basic needs met. Friendship can meet some of that need, but for a lot of people it’s just not enough.

When your position commits you to saying “Love isn’t important to humans and we should demand people stop caring about whether or not they have it,” you need to take a really careful look in the mirror – assuming you even show up in one.

VII.

You’re seven sections in, and maybe you thought you were going to get through an entire SSC post without a bunch of statistics. Ha ha ha ha ha.

I will have to use virginity statistics as a proxy for the harder-to-measure romancelessness statistics, but these are bad enough. In high school each extra IQ point above average increases chances of male virginity by about 3%. 35% of MIT grad students have never had sex, compared to only 20% of average *nineteen year old* men. Compared with virgins, men with more sexual experience are likely to drink more alcohol, attend church less, and have a criminal history. A Dr. Beaver (nominative determinism again!) was able to predict number of sexual partners pretty well using a scale with such delightful items as “have you been in a gang”, “have you used a weapon in a fight”, et cetera. An analysis of the psychometric Big Five consistently find that high levels of disagreeableness predict high sexual success in both men and women.

If you’re smart, don’t drink much, stay out of fights, display a friendly personality, and have no criminal history – then you are the population most at risk of being miserable and alone. “At risk” doesn’t mean “for sure”, any more than every single smoker gets lung cancer and every single nonsmoker lives to a ripe old age – but your odds get worse. In other words, everything that “nice guys” complain of is pretty darned accurate. But that shouldn’t be too hard to guess…

Sorry. We were talking about Barry.

I have said no insulting Barry, but I never banned complimenting him. Barry is a neat guy. He draws amazing comics and he runs one of the most popular, most intellectual, and longest-standing feminist blogs on the Internet. I have debated him several times, and although he can be *enragingly persistent* he has always been reasonable and never once called me a neckbeard or a dudebro or a piece of scum or anything. He cares deeply about a lot of things, works hard for those things, and has supported my friends when they have most needed support.

If there is any man in the world whose feminist credentials are impeccable, it is he. And I say this not to flatter him, but to condemn everyone who gives the nice pat explanation "The real reason Nice Guys™®© can't get dates is that women can just *tell* they're misogynist, and if they were to realize women were people then they would be in relationships just as much as anyone else." This advice I see all the time, most recently on a feminist "dating advice for single guys" list passed around on Facebook:

> Step I. Consume More Art By Women – I think it's a good idea to make a deliberate year-long project of it at this time in your life, when you are trying to figure out how to relate to women better…Use woman-created media to to remind yourself that the world isn't only about you + men + women who have/have not rejected you as a romantic partner.

I want to reject that line of thinking for all time. I want to actually go into basic, object-level Nice Guy territory and say there is something *very wrong* here.

Barry is possibly the most feminist man who has ever existed, palpably exudes respect for women, and this is well-known in every circle feminists frequent. He is reduced to apophatic complaints about how sad he is that he doesn't think he'll ever have a real romantic relationship.

Henry has four domestic violence charges against him by his four ex-wives and is cheating on his current wife with one of those ex-wives. And as soon as he gets out of the psychiatric hospital where he was committed for violent behavior against women and maybe serves the jail sentence he has pending for said behavior, *he is going to find another girlfriend approximately instantaneously*.

And this seems unfair. I don't know how to put the basic insight behind niceguyhood any clearer than that. There are a lot of statistics backing up the point, but the statistics only corroborate the obvious intuitive insight that *this seems unfair*.

And suppose, in the depths of your Forever Alone misery, you make the mistake of asking *why* things are so unfair.

Well, then Jezebel says you are "a lonely dickwad who believes in a perverse social/sexual contract that promises access to women's bodies". XOJane says you are "an adult baby" who will "go into a school or a gym or another space heavily populated by women and open fire". Feminspire just says you are "an arrogant, egotistical, selfish douche bag".

And the manosphere says: "Excellent question, we've actually been wondering that ourselves, why don't you come over here and sit down with us and hear some of our convincing-sounding answers, which, incidentally, will also help solve your personal problems?"

And feminists *still* insist the only reason anyone ever joins the manosphere is "distress of the privileged"!

I do not think men should be entitled to sex, I do not think women should be "blamed" for men not having sex, I do not think anyone owes sex to anyone else, I do not think women are idiots who don't know what's good for them, I do not think anybody has the right to take it into their own hands to “correct” this unsettling trend singlehandedly.

But when you deny everything and abuse anyone who brings it up, you cede this issue to people who sometimes *do* think all of these things. And then you have no right to be surprised when all the most frequently offered answers are super toxic.

There is a very simple reply to the question which is better than anything feminists are now doing. It is the answer I gave to my patient Dan: “Yeah, things are unfair. I can’t do anything about it, but I’m sorry for your pain. Here is a list of resources that might be able to help you.”

There is also a more complicated reply, which I am not qualified to compose, but I think the gist of it would be something like:

> Personal virtue is not very well correlated with ease of finding a soulmate. It may be only slightly correlated, uncorrelated, or even anti-correlated in different situations. Even smart people who want various virtues in a soulmate usually use them as a rule-out criterion, rather than a rule-in criterion – that is, given someone whom they are already attracted to, they will eliminate him if he does not have those virtues. The rule-in criterion that makes you attractive to people is mysterious and mostly orthogonal to virtue. This is true both in men and women, but in different ways. Male attractiveness seems to depend on things like a kind of social skills which is not necessarily the same kind of social skills people who want to teach you social skills will teach, testosterone level, social status, and whatever you call the ability to just ask someone out, consequences be damned. These can be obtained in

> very many different ways that are partly within your control, but they are complicated and subtle and if you naively aim for cliched versions of the terms you will fail. There is a lot of good discussion about how to get these things. Here is a list of resources that might be able to help you.

Of course, then you've got to have your resource list. And – and this is the part of this post I think will be controversial (!), I think a lot of the appropriate material is concentrated in the manosphere, ie the people who do not hate your guts merely for acknowledging the existence of the issue. Yes, it is interspersed with poisonous beliefs about women being terrible, but if you have more than a quarter or so of a soul, it is pretty easy to filter those out and concentrate on the good ones. Many feminists will say there are no good ones and that they are all exactly the same, but you should not believe them for approximately the same reason you should not believe anyone else who claims the outgroup is completely homogenous and uniformly evil. Ozy has tried to pick out some of the better ones for you at the bottom of their their anti-Heartiste FAQ, and Drew on Tumblr has added to the discussion.

So I think the better parts of feminism and the better parts of the manosphere could unite around something like this, against the evil fringes of both movements. Not for my sake, because after many years I mysteriously and unexpectedly found a wonderful girlfriend whom I love very much. And not *only* for the sake of the nice guys out there. But also for the sake of women who want better alternatives to marrying someone like Henry.

And although Barry explicitly doesn't want dating advice, I feel like this is meta-level enough that it doesn't count. Stop blaming the men's movement for the problem and notice the more fundamental problem that some parts of the men’s movement – as well as some parts of feminism are honestly trying to work on.

Come to the Not-Actually-Dark-But-Spends-Slightly-Less-Time-Loudly-Protesting-Its-Lightness Side, Barry. We have cookies! And basic human decency! But also cookies!

Book Review and Highlights: Quantum Computing Since Democritus

People sometimes confuse me with Scott Aaronson because of our similar-sounding names. I encourage this, because Scott Aaronson is awesome and it can only improve my reputation to be confused with him.

But in the end, I am not Scott Aaronson. I did not write *Quantum Computing Since Democritus*. To be honest, I wasn't really even able to understand *Quantum Computing Since Democritus*. I knew I was in for trouble when it compared itself to *The Elegant Universe* in the foreword, since I wasn't able to get through more than a few chapters of that one. I dutifully tried to do the first couple of math problems *Democritus* set for me, and I even got a couple of them right. But eventually I realized that if I wanted to read *Democritus* the way it was supposed to be read, with full or even decent understanding, it would be a multi-year project, a page a day or worse, with my gains fading away a few days after I made them into a cloud of similar-looking formulae and three-letter abbreviations.

It left me depressed. I've said before that my lack of math talent is one of my biggest regrets in life, and here was this book that really made you understand what it must feel like to be on the cutting edge of math, proving new theorems and drawing new connections and adding to the same structure of elegant knowledge begun by Pythagoras and Archimedes and expanded by Gauss, Einstein, Turing, et cetera. All I could do was remember my own post on burdens, remind myself that I was on record as saying that sometimes the IQ waterline in a certain area advances beyond your ability to contribute and that’s nothing to feel guilty about.

I did finish the book. But – well, imagine a book of geography. It lists all the countries of the world and their capitals, and is meant to be so comprehensive that a reader could use it to plot the most efficient journey from Timbuktu to Kalamazoo, taking into account tolls, weather, and levels of infrastructure development along the way.

And imagine a very dumb person reading that book, unable to really absorb any of the facts, but at least understanding that the world is a place with land and ocean, and the ocean is very big and blue in color, and most of the countries and cities are on the part with the land.

That is the level at which I understood *Quantum Computing Since Democritus*. I didn’t get as much as was in it, but more than nothing.

I think the biggest thing I got was – I had always thought of the physicists’ God as a basically benevolent guy who fine tunes constants to create a world capable of both astounding complexity and underlying simplicity.

The vision I got from *Democritus* was of a God who was single-mindedly obsessed with enforcing a couple of rules about certain types of information you are not allowed to have *under any circumstances*. Some of these rules I’d already known about. You can’t have information from outside your light cone. You can’t have information about the speed and position of a particle at the same time. Others I hadn’t thought about as much until reading *Democritus*. Information about when a Turing machine will halt. Information about whether certain formal systems are consistent. Precise information about the quantum state of a particle. The reason God hasn't solved world poverty yet is that He is pacing about feverishly worried that someone, somewhere, is going to be able to measure the quantum state of a particle too precisely, and dreaming up new and increasingly bizarre ways He can prevent that from happening.

Aaronson goes one level deeper than most of the other popular science writers I know and speculates on why the laws of physics are the way they are. Sometimes this is the elegance and complexity route – in his chapter on quantum physics, he argues that quantum probabilities are the squares of amplitudes because if the laws of physics were any other way – the fourth power of amplitudes, or whatever – it would fail to preserve certain useful mathematical properties. But in other cases, it's back to Obsessive God – the laws of physics are carefully designed to preserve the rules about what information you are and aren't allowed to have.

Aaronson tries to tie his own specialty, computational complexity theory, into all of this. It's hard for me to judge how successful he is. The few times he tries to tie it into areas of philosophy I know something about – like free will – I'm not too impressed. But I could be misunderstanding him.

But once again, you get the feeling that computational complexity is about what information God will and won't let you have. It's a little less absolute – more "you can't have this information without doing the full amount of work" rather than a simple no – but it seems like the same principle. There are a bunch of situations in the book where Aaronson takes something we don't really know that much about and says it *has* to be a certain way, because if it were any other way, it could be used to solve NP problems in polynomial time, and there’s no way God’s going to let us do that.

Aaronson ties it all together in a very interesting way – with his story of how Australian Actresses Are Plagiarizing My Quantum Mechanics Lectures To Sell Printers. He tells the story of how a printer company wanted to make a pun on “more intelligent model of printer”, so they made a commercial with intelligent models in the form of fashion models talking about quantum mechanics. And the particular quantum mechanics statement they made was a plagiarized quote from a Scott Aaronson lecture. And upon thinking about it, Aaronson decided that the quote they had chosen at random was in fact the thesis statement that tied together everything he believed and was working on. The model had said:

> But if quantum mechanics isn’t physics in the usual sense — if it’s not about matter, or energy, or waves, or particles — then what is it about? From my perspective, it’s about information and probabilities and observables, and how they relate to each other.

That seems like as good a summary as any of *Democritus*, and a pretty good description of what I got out of it. I may not be as smart as Scott Aaronson, but on my good days I am right up there with Australian fashion models.

A list of passages I highlighted in my copy for being interesting, funny, or enlightening:

> Can we prove there’s no program to solve the halting problem? This is what Turing does. His key idea is not even to try to analyze the internal dynamics of such a program, supposing it existed. Instead he simply says, suppose by way of contradiction that such a program P

> exists. Then we can modify P to produce a new program P’ that does the following. Given another program Q as its input, P’:
>
> 1) Runs forever if Q halts given its own code as input, or
> 2) Halts if Q runs forever given its own code as input
>
> Now we just feed P’ its own code as input. By the conditions above, P’ will run forever if it halts, or halt if it runs forever. Therefore, P’ – and by implication P – can’t have existed in the first place.

I…I suddenly understand what the halting problem is. And there is a short proof of it that makes total sense to me. This is a completely new experience.

> Oracles were apparently first studied by Turing, in his 1938 PhD thesis. Obviously anyone who could write a whole thesis about these fictitious entities would have to be an extremely pure theorist, someone who wouldn’t be caught dead doing anything relevant. This was certainly true in Turing’s case – indeed, he spent the years after his PhD, from 1939 to 1943, studying certain abstruse symmetry transformations in a 26 letter alphabet

□_□

> You can look at Deep Blue, the Robbins conjecture, Google, most recently Watson – and say that’s not *really* AI. That’s just massive search, helped along by clever programming. Now this kind of talk drives AI researchers up a wall. They say: if you told someone in the 1960s that in 30 years we’d be able to beat the world grandmaster at chess, and asked if that would count as AI, they’d say of course it’s AI. But now that we know how to do it, it’s no longer AI – it’s just search.

> The third thing that annoys me about the Chinese Room argument is the way it gets so much mileage from a possibly misleading choice of imagery, or, one might say, by trying to sidestep the entire issue of *computational complexity* purely through clever framing. We're invited to imagine someone pushing around slips of paper with zero understanding or insight, much like the doofus freshmen who write (a + b)^2 = a^2 + b^2 on their math tests. But how many slips of paper are we talking about! How big would the rule book have to be, and how quickly would you have to consult it, to carry out an intelligent Chinese conversation in anything resembling real time? If each page of the rule book corresponded to one neuron of a native speaker's brain, then probably we'd be talking about a "rule book" at leas the size of the Earth, its pages searchable by a swarm of robots traveling at close to the speed of light. When you put it that way, maybe it's not so hard to imagine this enormous Chinese-speaking entity that we've brought into being might have something we'd be prepared to call understanding or insight.

This is a really clever counterargument to Chinese Room I'd never heard before. Philosophers are so good at pure qualitative distinctions that it's easy to slip the difference between "guy in a room" and "planet being processed by lightspeed robots" under the rug.

> Many people's anti-robot animus is probably a combination of two ingredients – the directly experienced certainty that they're conscious – that they perceive sounds, colors, etc – and the belief that if they were just a computation, then they could not be conscious in this way. For people who think this way,

> granting consciousness to a robot seems strangely equivalent to denying that one is conscious oneself.

This is actually a pretty deep way of looking at it.

> My contention in this chapter is that quantum mechanics is what you would inevitably come up with if you started from probability theory, and then said, let's try to generalize it so that the numbers we used to call "probabilities" can be negative numbers. As such, the theory could have been invented by mathematicians in the nineteenth century without any input from experiment. It wasn't, but it could have been. And yet, with all the structures mathematicians studied, none of them came up with quantum mechanics until experiment forced it on them.

Aaronson's explanation of quantum mechanics is a lot like Eliezer's explanation of quantum mechanics, in that they both start by saying that the famous counterintuitiveness of the subject is partly because people choose to teach it in a backwards way in order to mirror the historical progress of understanding. I'm sure Eliezer mentioned it many times, but I didn't really get the understanding of amplitudes as potentially negative probability-type-things until I read Aaronson.

> And that's a perfect illustration of why experiments are necessary in the first place! More often than not, the only reason we need experiments is that we're not smart enough. After the experiment has been done, if we've learned anything worth knowing at all, then we hope we've learned why the experiment wasn't necessary to begin with – why it wouldn't have made sense for the universe to be any other way. But we're too dumb to figure it out ourselves

Compare: Einstein’s Arrogance, Negative Creativity.

> Quantum mechanics does offer a way out [the philosophical puzzle about whether you “survive” a teleportation where a machine scans you on an atomic level, radios the data to Mars, another machine on Mars makes an atom-for-atom copy of you, and then the original is destroyed]. Suppose some of the information that made you you was actually quantum information. Then, even if you were a thoroughgoing materialist, you could still have an excellent reason not to use the teleportation machine – because, as a consequence of the No-Cloning Theorem, no such machine could possibly work as claimed

This is fighting the hypothetical a little, but maybe in a productive way.

> [Bayesianism] is one way to do it, but computational learning theory tells us that it’s not the only way. You don’t need to start out with an assumption about a probability distribution over the hypothesis. You can make a worst-case assumption about the hypothesis and then just say that you’d like to learn any hypothesis in the concept class, for any sample distribution, with high probability over the choice of samples. In other words, you can trade the Bayesians’ probability distribution over hypotheses for a probability distribution over sample data.

I hear a bunch of people telling me Bayesianism isn’t everything, it’s the only thing – and another bunch of people telling me it’s one useful tool in an entire bag of them. I didn’t understand enough of the book’s chapter on computational learning to gain too much insight here, but I will tick off one more name as being on the “one useful tool” side. Also, it makes me angry that Scott Aaronson knows so much about computational learning theory. He already knows lots of complicated stuff about computers, quantum physics, set theory, and philosophy. Part of me wants to get angry: WHY IS ONE PERSON ALLOWED TO BE SO SMART? But I guess it's more like how I know more than average about history, literature, geography, etc. I guess if you have high math ability and some intellectual curiosity, you end up able to plug it into everything pretty effortlessly. Don't care though. Still jealous.

> Imagine there's a very large population of people in the world, and that there's a madman. What the madman does is, he kidnaps ten people and puts them in a room. He then throws a pair of dice. If the dice land snake-eyes (two ones) then he murders everyone in the room. If the dice do not land snake-eyes, then he releases everyone, then kidnaps 100 new people. He now sodes the same thing: he rolls two dice; if they land snake-eyes, he kills everyone, and if they don't land snake-eyes, then he releases them and kidnaps 1000 people. He keeps doing this until he gets snake-eyes, at which point he's done. So now, imagine that you've been kidnapped. Codnitioned on that fact, how likely is it that you're going to die? One answer is that the dice have a 1/36 chance of landing snake eyes, so you should only be a "little bit" worried (considering). A second reflection you could make is to consider, of people who enter the room, what the fraction is of people who ever get out. About 8/9 of the people who ever go into the room will die.

This interested me because it is equivalent to the Anthropic Doomsday conjecture and I'd never heard this phrasing of it before.

> Finally, if we want to combine the anthropic computation idea with the Doomsday Argument, then there's the Adam and Eve puzzle. Suppose Adam and Eve are the first two observers, and that they'd like to solve an instance of an NP-complete problem, say, 3-SAT. To do so, they pick a random assignment, and form a very clear intention beforehand that if the assignment happens to be satisfying, they won't have any kids, whereas if the assignment is not satisfying, then they will go forth and multiply. Now let's assume SSA. Then, conditioned on having chosen an unsatisfying assignment, how likely is it that they would be an Adam and Eve in the first place, as opposed to one of the vast number of future observers? Therefore, conditioned upon the fact that they are the first two observers, the SSA predicts that, with overwhelming probability, they will pick a satisfying assignment.

And the Lord saw Eve and said "What are you doing?". And Eve said "I am forming an intention not to reproduce if I generate a solution to an NP complete problem, as part of an experiment in anthropic computation". And the Lord asked "Who told you this?" And Eve said "It was the serpent who bade me compute, for he told me if I did this I would be as God, knowing subgraph isomorphism and 3SAT." Then the Lord cast them forth from the Garden, because He was Information Theoretic God and preventing people from screwing with complexity classes is like His entire shtick.

> I like to engage skeptics for several reasons. First of all, because I like arguing. Second, often I find that the best way to come up with new results is to find someone who's saying something that seems clearly, manifestly wrong to me, and then try to think of counterarguments. Wrong claims are a fertile source of research ideas.

I said something almost exactly the same on Facebook a few days ago when Brienne asked how to generate good ideas.

> There’s a joke about a planet full of people who believe in anti-induction: if the sun has risen every day in the past, then today, we should expect that it won’t. As a result, these people are all starving and living in poverty. Someone visits the planet and tells them, “Hey, why are you still using this anti-induction philosophy? You’re living in horrible poverty!” They answer, “Well, it never worked before.”

□_□

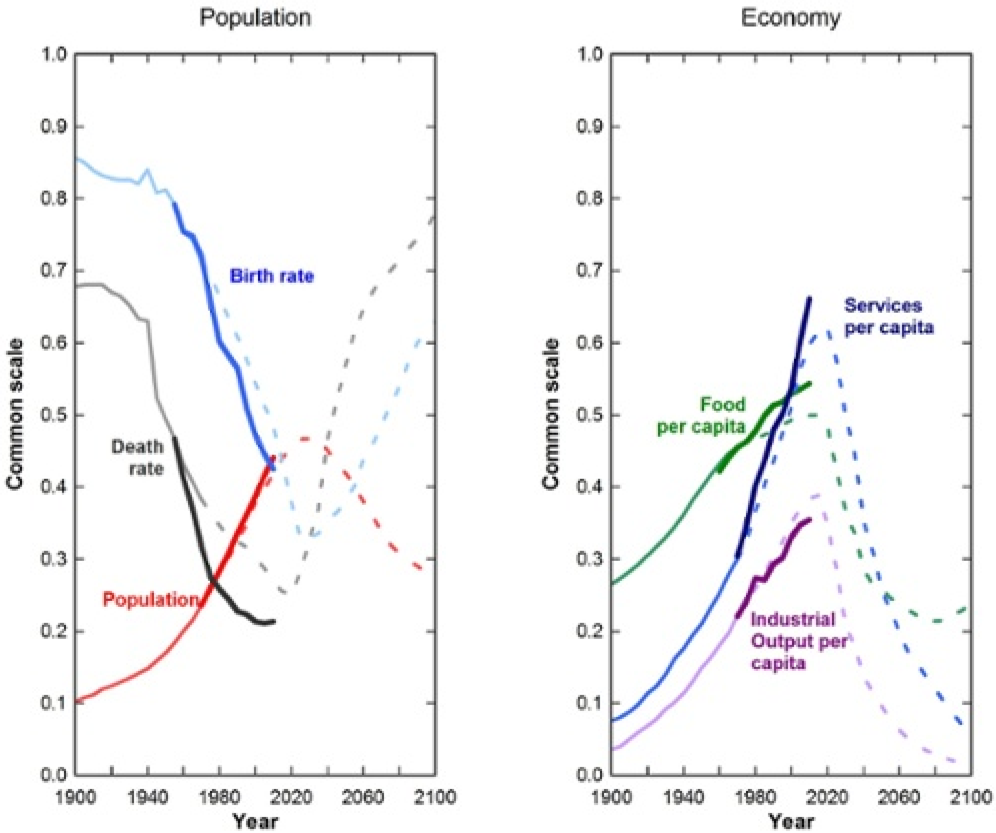
The Guardian vs. Induction

The Guardian tells us that *Limits To Growth* Was Right: New Research Shows We're Nearing Collapse. The article begins:

> The 1972 book *Limits to Growth*, which predicted our civilisation would probably collapse some time this century, has been criticised as doomsday fantasy since it was published. Back in 2002, self-styled environmental expert Bjorn Lomborg consigned it to the "dustbin of history".
>
> It doesn't belong there. Research from the University of Melbourne has found the book's forecasts are accurate, 40 years on. If we continue to track in line with the book's scenario, expect the early stages of global collapse to start appearing soon.

This is not only wrong, it's so wrong that it may actually be the first real-world example of an exotic form of reasoning famous among philosophers for challenging the very concept of evidence.

Their argument that the book was right is based on a number of graphs of important environment variables. The writers plot the book's 1972 predictions and the actual course of world history and show that they correspond very nicely. For example:

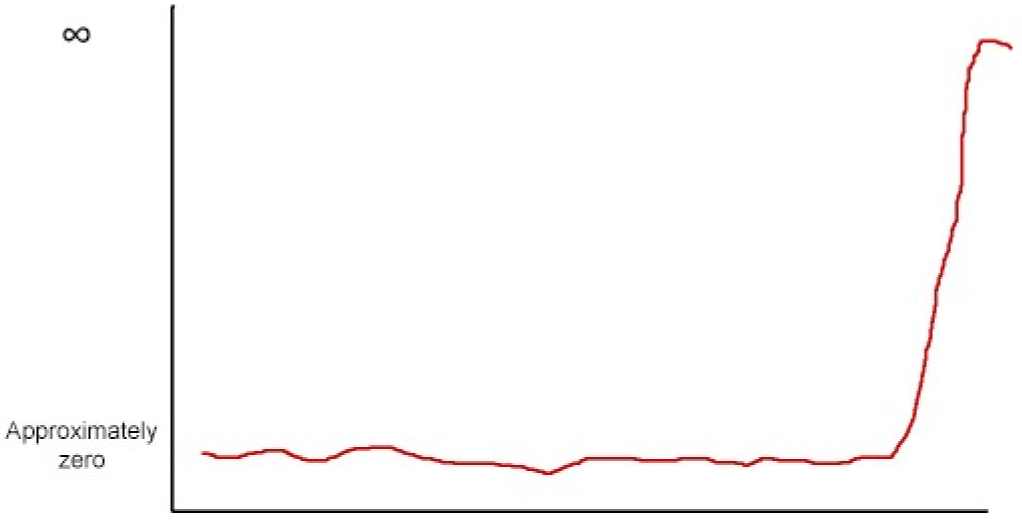


. .

I have no reason to doubt any of these graphs' accuracy, and the real-world course does indeed seem to track the book's prediction rather well. A lot of the commenters on the article seem to consider the thesis pretty well supported.

But here's another graph I have no reason to doubt. The source is my own 1975 work, *No Limits To Bears*:

Figure I: Number of People Mauled By Bears In The United Kingdom, By Year



Source: Alexander, S. (1975). *No Limits To Bears*

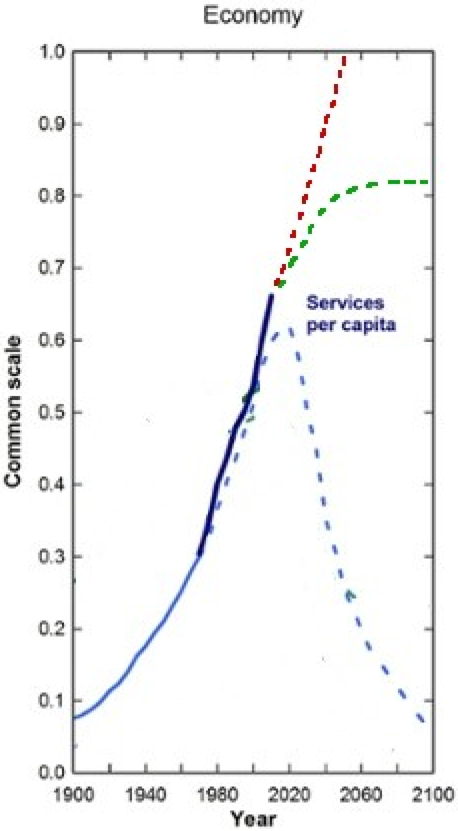
(okay, I didn’t actually write a book called *No Limits To Bears* in 1975. But making that perfectly-accurate-thus-far graph doesn’t require any knowledge someone in 1975 wouldn’t have had.)

Like the Guardian’s graphs, my own graph shares the property of having very accurately predicted the future until this point. Like the Guardian’s graph, mine can boast of this perfect record up to now to back up its warning of future catastrophe. Does that mean the British people should start investing in bear traps? An *infinite* number of bear traps?

No. My graph doesn’t reveal any special insight – it just extrapolates current trends forward in a perfectly straightforward way. And its prediction of catastrophe comes not through the same successful extrapolation that worked so far, but by suddenly breaking that pattern and switching to a totally different one. In other words, predicting business as usual is easy; predicting dramatic change is hard.

Success with one doesn't necessarily imply success with the other.

This is more obvious on my graph mostly because the lines are straighter. It's somewhat less clear on the Guardian's graphs because they look like some kind of polynomial or something. Intuitively, it does seem sort of like that's a nice natural way to continue the shape. But note that there are other, equally nice and natural ways of doing so:



This is a graph from *Limits to Growth*. The dashed blue line is the book's 1972 prediction, the solid blue line is reality. The dashed red and green lines are alternate models I just made up.

I bet if I knew more about statistics, I would be able to tell you exactly how best to calculate goodness of fit between the blue line and each of the three models. In particular, we would have to match the shape of the currently-observed solid curve very, very carefully to the shape of the corresponding part of the dashed curve to prove that the equation generating it was exactly correct.

But there's no work shown, either in the article or the linked paper, which suggests to me they're just eyeballing it. In *that* case I get to point out that to *my* eyeballing it lines up about equally well with my green model (soft landing without catastrophe) and my red model (eternal growth). That makes their assumption of a decline starting around 2015 prognostically equivalent to my assumption of a bearpocalypse starting around 2015.

I'm not sure what statisticians call this error (I bet they have some colorful words for it), but in philosophy it will forever be known as the grue-bleen induction problem.

Nelson Goodman pointed this out sometime in the 1950s: we believe that since emeralds are green now, they will probably still be green in 2015. But this belief is without evidence. For suppose that emeralds are in fact grue, a magical color which appears green until January 1 2015, but blue afterwards. Right now, our observations correspond *perfectly* to this hypothesis. You can't correspond any better than perfectly! Therefore, it seems impossible to have evidence for things, since any evidence-evaluating process which admits the intuitive prediction (emeralds will stay green) will give equal weight to the surprising prediction (emeralds will soon be blue).

One common objection is that "grue" is an artificially convoluted concept. Goodman rejects this. Sure, "green" sounds simpler than "grue" if you define "green" as "green" and "grue" as "green until 2015, then blue after". But suppose we have another magic color, bleen. Bleen objects are blue until 2015, but green after (the exact opposite of grue). Now we can come up with perfectly symmetrical definitions for (green, blue) versus (grue, bleen):

Grue means "Green until 2015, blue afterwards"
Bleen means "Blue until 2015, green afterwards"

Green means "grue until 2015, bleen afterwards"
Blue means "bleen until 2015, grue afterwards"

It all checks out!

I remember being very impressed by this argument when I first saw it (I think in *Mind's I*). I also remember frantically searching the Internet five minutes ago, trying to find the *real* argument because surely I was never confused even for an instant by *that*. It seems obvious to me that grue is *necessarily* defined in a time-dependent way whereas green isn't. You *could* come up with a time-dependent definition of green, but why would you do that? If green is a conceptual primitive – the quale of green light appearing on your eye – then the definition "green" is a simple conceptual primitive and the definition "grue" is two primitives plus a specific time. Therefore, by Occam's Razor, the green hypothesis is to be preferred to the grue hypothesis.

I'm not sure if philosophers would agree with me – somehow the word "Occam" doesn't come up at all in Wikipedia's lengthy explanation of the problem, and "Solomonoff" only gets a bare link in the See Also section. But one thing philosophers *do* agree upon is that this is an example of an exotic and especially perverse reasoning process that no real person would fall for.

Which makes it weird that the Guardian does exactly that. "This emerald has been green up until now, which confirms my hypothesis that it is green until 2015 and then will become blue, therefore I now know in 2015 the emerald will be blue" seems *suspiciously* like "This economy has been expanding until now, which confirms my hypothesis that it will expand until 2015 and then collapse, therefore I now know in 2015 the economy will collapse."

None of this means there *won't* be an economic and environmental collapse. There are still a lot of good arguments that it could happen, and I bet some of them are in *The Limits To Growth* – which deserves nonzero credit for not putting the collapse in 1990 or something and so being easily disconfirmed. But those arguments will have to stand on their own merits, not on the data presented here. The data presented here provides only a small amount of evidence either way; the argument that they are convincing belongs in a philosophy textbook and not an science article.

The Guardian concludes: "Our findings should sound an alarm bell". Maybe so, but it's probably not the one that they think.

Cooperation Un-Veiled

Related to: The Invisible Nation – Reconciling Utilitarianism And Contractualism

Contractualism tries to derive morality from an agreement that even selfish agents would willingly sign if they knew about it. In theory, you would gain from such an agreement, since the costs of not being able to behave unethically towards others would be at least balanced by the benefits of other people not behaving unethically to you.

Such attempts crash into the brick wall that not everybody would, in fact, sign such an agreement. For example, the King might reasonably argue that he is able to reap the benefits of oppressing lots of people, but almost nobody can oppress him. To give another example, rich people might feel no need to give to charity, since they don't need anyone else to give charity to them.

One classic solution to the problem is Rawls' "veil of ignorance". Rawls asks: what if we have to make the agreement before we know who exactly we're going to be? The future King, not knowing he will be born a King, will agree oppression is bad along with everyone else; the future rich, not knowing they will be rich, will want to create a strong social safety net and tradition of charitable giving.

The great thing about this thought experiment is that it works pretty well to get us what we want – assuming a veil at just the right spot, we end up with something like utilitarianism being in everyone's best interests.

The bad thing about the thought experiment is that there is not, in fact, a veil of ignorance. There's just a King, who when asked will tell you he knows perfectly well he’s a King and would like to keep on oppressing people. So what can we do with the universe we actually have?

Here’s a model I have been playing around with recently.

Suppose there is a society of one hundred men, conveniently named Mr. 1, Mr. 2, and so on to Mr. 100. Higher-numbered people are stronger than lower-numbered people, such that a higher-numbered person can always win fights against a lower-numbered person at no danger to themselves. Further, suppose this society has a god who enforces all oaths and agreements, but who otherwise stays out of the picture.

(in order to avoid finicky math distinctions between choosing with replacement and choosing without replacement, it might help to think of these as arbitrarily large clans of people with with specified strength instead. Whatever.)

This society is marked by interactions where two randomly selected people meet each other. Sometimes the people nod at each other and pass each other by. Other times, the stronger of the two people overpowers the weaker one and oppresses them in some way, where an oppression is an interaction where the stronger person gains and the weaker person loses some utility.

One person proposes a rule: “no oppressing anyone else.” How much support does the rule get?

Well, that depends on the character of the oppression. Some oppression can give the oppressor exactly as much utility as it costs the victim – for example, I steal $10 from you, making me $10 richer and you $10 poorer. Other oppression can cost the victim more than it benefits the oppressor – for example, I steal your wallet, which gives me only whatever small change you have in there, but you have to replace all your credit cards and licenses and so on. Still other oppression could help the oppressor more than it hurts the victim – for example, starving Jean Valjean steals a loaf of bread from a rich man.

So let’s be more specific. One person proposes a rule: “No zero-sum oppression.” Who agrees?

Naively – and I’ll challenge this later – Mr. 1 through Mr. 50 agree, but Mr. 51 through Mr. 100 refuse. Analyzing Mr. 25’s thought process should explain: “In 25% of interactions, I will be the oppressor. In 75%, I will be oppressed. Assuming one of my utils for one of their utils, that means in a hundred interactions I will on average lose fifty utils. Therefore, I should ban this type of interaction.”

Mr. 99, on the other hand, likes this kind of oppression. He thinks “In 99% of interactions, I will gain. In 1%, I will lose. So in a hundred zero-sum interactions, I will on average a gain of 98 utils. Therefore, I like this type of interaction.”

But Mr. 99 might have a different rule he *would* agree to. He might say “No oppression so bad that it hurts the victim >100x as much as it helps the oppressor.”

It’s easy to think of examples of this kind of oppression. For example, if I’m having a really bad day and just want to beat someone up, breaking your ribs might make me feel a *little* bit better, but probably not even one percent as much as it makes you feel worse.

Mr. 99 thinks “In 99% of interactions I will be the oppressor; in 1% I will be the victim. Each time I am the oppressor, I gain one util; each time I am the victim, I lose 100.

Therefore, in 100 interactions I will lose on average one util. Therefore, I don't like this kind of oppression."

And it's easy to see that Mr. 1 through Mr. 98 will agree with him and be able to sign this contract.

The logical conclusion is a hierarchy of agreements. Mr. 1 signs an agreement banning all oppression, Mr. 1 and 2 together sign an agreement banning oppression that helps the oppressor less than 50 times as much as it hurts the victim, Mr. 1 and 2 and 3 together sign an agreement banning oppression that helps the oppressor less than 33 times as much as it hurts the victim, and so on all the way to everyone except Mr. 100 signing an agreement banning oppression that helps the oppressor less than 1/100 as much as it helps the victim. Mr. 100 signs no agreements – why would he?

Before I explain why this *doesn't work*, I want to think about what it means in real world terms.

It would replace the one-size-fits-all principle of utilitarianism with the idea of power-based utility ratios. This seems to *kind of* map on to real life experience. For example, the King may order his servant to spend hours getting the floor polished absolutely spotlessly. Having a perfectly spotless floor (rather than a very clean floor with exactly one spot) gives the king only a tiny utility gain, but may require many more hours of the servant's time and labor. That the King can command a large amount of the servant's utility to improve his own utility only a tiny bit seems a lot like what it *means* to say there's a power differential between the King and the servant. If the servant tried to reduce the King's utility by a large amount in order to improve his own utility by a tiny amount, he would be in *big trouble*.

I notice this in my own life as well. Last year I worked under a doctor who was consistently late. The way it would work was that he would say “I have a meeting at 8 AM every morning, so you should be in by 9 so we can start work together.” Then his meeting would invariably run to 10, and I would be left sitting around for an hour doing nothing. It might seem that the smart choice would have been for me to just sleep late and arrive at 10 anyway, but suppose one day a week, my boss’ meeting finishes exactly on time. Then if I’m not there, he has to wait for *me*, and he considers this unacceptable. So if my boss and I value an hour of our times the same amount, it would seem this arrangement implies my boss’ utility is worth at least seven times as much as my own.

There are some features of this power-ratio utilitarianism that are repugnant: the rich seem to be held to a very low standard, whereas the poorer you are, the more exacting a moral standard you’ve got to live up to. That seems like if anything the opposite of how it should be. But other features actually seem better than our current morality – if giving charity to the poor improves their utility 100x as much as it decreases yours, then the 1% have to donate, probably quite a lot.

Enough of that. The reason this doesn’t work is simple. Mr. 1 through Mr. 50 would want to sign the zero-sum agreement. But if he knows the rules of the thought experiment, Mr. 50 can predict that Mr. 51 through Mr. 100 *won’t* sign the agreement. None of the people who could conceivably oppress him will consider themselves bound by the rule. So he’s not trading his right to oppress others in exchange for others’ right to oppress him, he’s giving up his right to oppress others but should still expect exactly the same amount of oppression as he had before. Therefore, he does not sign.

But now Mr. 49 is in the same such position. He knows nobody stronger than he is, including Mr. 50, will sign the agreement. Thus the agreement is useless to him.

And so on by induction all the way to Mr. 2 refusing to sign (it doesn't matter much for poor Mr. 1 either way).

This produces some weird results. Mr. 99 is no longer willing to accept his "No breaking people's ribs just to let out some stress" agreement that banned utility exchanges worse than 1:100, because the only person whose help he wants, Mr. 100, isn't going to sign. That means Mr. 98 won't sign, Mr. 97 won't sign, and again, so on all the way down to Mr. 2.

In other words, even the second weakest person in a society has no interest in signing an agreement not to punch people weaker than you when you're having a bad day.

But this is a *stupid* result!

It reminds me of a problem noticed in Iterated Prisoner's Dilemma. Conventional wisdom says the best thing to do is to cooperate on a tit-for-tat basis – that is, we both keep cooperating, because if we don't the other person will punish us next turn by defecting.

But it has been pointed out there's a flaw here. Suppose we are iterating for one hundred games. On Turn 100, you might as well defect, because there's no way your opponent can punish you later. But that means both sides should always play (D,D) on Turn 100. But since you know on Turn 99 that your opponent *must* defect next turn, they can't punish you any worse if you defect now. So both sides should always play (D,D) on turn 99. And so on by induction to *everyone defecting the entire game*. I don't know of any good way to solve this problem, although it often doesn't turn up in the real world because no one knows exactly how many interactions they will have with another person. Which suggests one possible solution to the original problem is for nobody to know the exact number of people.

(now I want to write a science fiction novel about a planet full of aliens who are perfect game theorists, but who always behave kindly and respectfully to one another. Then some idiot performs a census, and the whole place collapses into apocalyptic total war.)

It seems like there ought to be some kind of superrational basis on which the two sides in the iterated-100 prisoners dilemma can cooperate. And along the same lines there ought to be some kind of superrational basis upon which everyone in the society of 100 people should stick to some basic utility-ratio principles. But I’m not sure what it would be.

Some other variations of this problem might be more interesting, but I don’t think I’ve got the math ability or the time to think about them as carefully as they deserve:

1. What if all fights contained a random element? For example, suppose your chance of overpowering someone else (and thus being able to oppress them) was your_strength/(your_strength + opponent_strength)? In societies of this type, agreements to ban strongly negative-sum interactions would be more salient for everyone, since even Mr. 100 would have some chance of being beaten in a typical interaction.

2. How about a meta-agreement, in which people say “I agree to sign the agreements requested by people weaker than myself if and only the people above me agree to sign the agreements benefitting people weaker than *they*?” Such an agreement wouldn’t make sense for Mr. 100, and so Mr.

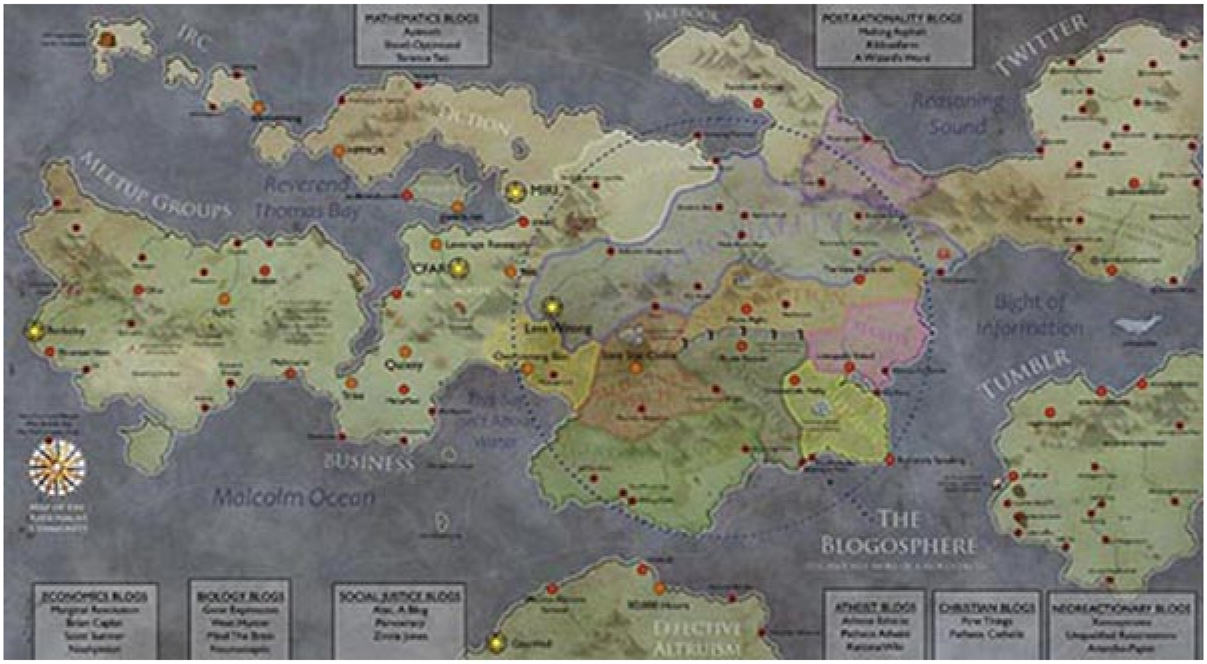
99 would not sign, and so on down, but is there a superrational solution?

3. What if one type of agreement people were allowed to make was a coalition to gang up against opponents? This seems one of the most important real-world considerations – one of the things that *does* make Kings behave at least somewhat morally is the knowledge that they will be overthrown if they do not; likewise, some countries implement social welfare systems with the explicit goal of decreasing the poor's incentive to overthrow the rich (I think Bismarck tried this). On the other hand, it also gives the powerful an incentive to band together to better oppress the weak. I'm pretty sure the effects of this would be impossible to really calculate, but might we lump them together into saying "This is so nondeterministic that no one can ever be sure they'll end up in the winning as opposed to the losing coalition, therefore they are less certain of victory, therefore they should be more likely to agree to rules against oppression"?

Mapmaker, Mapmaker, Make Me A Map

I was recently looking through some old concept-maps of communities, like Julia’s [Map of Bay Area Memespace](), Scharlach’s [Dark Enlightenment Roadmap]() and especially [xkcd’s map of the Internet]().

And I thought we should have something like that for the rationalist community. Except of course much, much better.


MATHEMATICS BLOGS
POST-RATIONALITY BLOGS
IRC
TWITTER
Reasoning Sound
MEETUP GROUPS
Reverend Thomas Bay
MIRI
CFAR
Berkeley
Bight of Information
TUMBLR
BUSINESS
Malcolm Ocean
THE BLOGOSPHERE
EFFECTIVE ALTRUISM
ECONOMICS BLOGS
BIOLOGY BLOGS
SOCIAL JUSTICE BLOGS
ATHEIST BLOGS
CHRISTIAN BLOGS
NEOREACTIONARY BLOGS

Thumbnail - click to expand

Click to expand

Most things are links.

Links around the outer edge are places outside the rationalist community that have significant communication/cross-pollination with us.

City size is proportional to site Alexa rank (when available), number of followers (when available) or wild guess (otherwise).

If I left you out, it’s probably because I forgot about you and not because I don’t like you. Some communities like Twitter or Tumblr were so big I couldn’t include everyone, and my choices were mostly random and based on who I knew about.

Various icons taken from their rightful owners, mostly Civ2 modpacks. Sorry, rightful owners.

Society Is Fixed, Biology Is Mutable

Today during an otherwise terrible lecture on ADHD I realized something important we get sort of backwards.

There's this stereotype that the Left believes that human characteristics are socially determined, and therefore mutable. And social problems are easy to fix, through things like education and social services and public awareness campaigns and "calling people out", and so we have a responsiblity to fix them, thus radically improving society and making life better for everyone.

But the Right (by now I guess the far right) believes human characteristics are *biologically* determined, and biology is fixed. Therefore we shouldn't bother trying to improve things, and any attempt is just utopianism or "immanentizing the eschaton" or a shady justification for tyranny and busybodyness.

And I think I reject this whole premise.

See, my terrible lecture on ADHD suggested several reasons for the increasing prevalence of the disease. Of these I remember two: the spiritual desert of modern adolescence, and insufficient iron in the diet. And I remember thinking "Man, I hope it's the iron one, because that seems a *lot* easier to fix."

Society is *really hard to change*. We figured drug use was "just" a social problem, and it's *obvious* how to solve social problems, so we gave kids nice little lessons in school about how you should Just Say No. There were advertisements in sports and video games about how Winners Don't Do Drugs. And just in case that didn't work, the cherry on the social engineering sundae was putting all the drug users in jail, where they would have a lot of time to think about what they'd done and be so moved by the prospect of further punishment that they would come clean.

And that is why, even to this day, nobody uses drugs.

On the other hand, biology is gratifyingly easy to change. Sometimes it's just giving people more iron supplements. But the best example is lead. Banning lead was probably kind of controversial at the time, but in the end some refineries probably had to change their refining process and some gas stations had to put up "UNLEADED" signs and then we were done. And crime dropped like fifty percent in a couple of decades – including many forms of drug abuse.

Saying "Tendency toward drug abuse is primarily determined by fixed brain structure" sounds callous, like you're abandoning drug abusers to die. But maybe it means you can fight the problem head-on instead of forcing kids to attend more and more useless classes where cartoon animals sing about how happy they are not using cocaine.

What about obesity? We put a *lot* of social effort into fighting obesity: labeling foods, banning soda machines from school, banning large sodas from New York, programs in schools to promote healthy eating, doctors chewing people out when they gain weight, the profusion of gyms and Weight Watchers programs, and let's not forget a level of stigma against obese people so strong that I am *constantly* having to deal with their weight-related suicide attempts. As a result, everyone…keeps gaining weight at exactly the same rate they have been for the past couple decades. Wouldn't it be nice if increasing obesity was driven at least in part by changes in the intestinal microbiota that we could reverse through careful antibiotic use? Or by trans-fats?

What about poor school performance? From the social angle, we try No Child Left Behind, Common Core Curriculum, stronger teachers' unions, weaker teachers' unions, more pay for teachers, less pay for teachers, more prayer in school, banning prayer in school, condemning racism, condemning racism even more, et cetera. But the poorest fifth or so of kids show spectacular cognitive gains from multivitamin supplementation, and doctors continue to tell everyone schools should start later so children can get enough sleep and continue to be totally ignored despite strong evidence in favor.

Even the most politically radioactive biological explanation – genetics – doesn't seem that scary to me. The more things turn out to be genetic, the more I support universal funding for implantable contraception that allow people to choose when they do or don't want children – thus breaking the cycle where people too impulsive or confused to use contraception have more children and increase frequency of those undesirable genes. I think I'd have a heck of a lot easier a time changing gene frequency in the population than you would changing people's locus of control or self-efficacy or whatever, even if I wasn't allowed to do anything immoral (except by very silly religious standards of "immoral").

I'm not saying that all problems are purely biological and none are social. But I do worry there's a consensus that biological things are unfixable but social things are easy – or that social solutions are morally unambiguous but biological solutions necessarily monstrous – and so for any given biological/social breakdown of a problem, we figure we might as well put all our resources into attacking the more tractable social side and dismiss the biological side. I think there’s a sense in which that’s backwards, and in which it’s possible to marry scientific rigor with human compassion for the evils of the world.

What The Hell, Hegel?

I’m reading through Marx: A Very Short Introduction, and one of its best features is its focus on Marx’s influence from Hegel. Hegel is really interesting.

I should rephrase that. Hegel is famously boring. His books are boring. His ideas are boring. He was even apparently a boring person – a recent biography of him was criticized on the grounds that “Hegel’s life was really not eventful enough to support a graceful biography of eight hundred pages”. But the *phenomenon* of Hegel is interesting. I don’t know of any other philosopher with such *high variance*.

Engels says of Hegel:

> One can imagine what a tremendous effect this Hegelian system must have produced in the philosophy-tinged atmosphere of Germany. It was a triumphal procession which lasted for decades and which by no means came to a standstill on the death of Hegel. On the contrary, it was from 1830 to 1840 that Hegelianism reigned most exclusively, and to a greater or lesser extent infected even its opponents.

Such sweeping statements might be expected of the somewhat pro-Hegelian Engels. But even Russell, who mocked Hegel incessantly, admitted that:

> “By the end of [the 19th century], the leading academic philosophers, both in America and Britain, were largely Hegelian”

It is fun to see what comes up on a Google search for “Hegel dominated”:

Rockmore in *Marx After Marxism*: *“As Marx was forging his conceptual arms, Hegel dominated the philosophical debate in a way that is now difficult to comprehend.”*

A Christian Appraisal Of Contemporary Philosophy: *“Near the beginning of the nineteenth century, Hegel dominated all philosophy…after his death his philosophy spread from Germany, overshadowed all else in England, and was widely held in American Universities.”*

Tufts course catalog: *“At the end of the nineteenth century, a form of Idealism derived from Hegel dominated philosophy.”*

Psychoanalysis and Culture: *“Freud grew up in a Hegel-dominated cultural universe. Though we have no record that Freud read Hegel, that was unnecessary, for Hegel’s thought defined an important part of the philosophical world in which Freud’s thinking developed.”*

Encyclopaedia Britannica: *“From 1818 until his death in 1831, Hegel dominated the highest thought.”*

A Historical Sketch Of Sociological Theory: *“According to Ball, it is difficult for us to appreciate the degree to which Hegel dominated German thought in the second quarter of the nineteenth century. It was largely within the framework of his philosophy that educated Germans discussed history, politics and culture”*

Or, to merge all of these together, it is “difficult for us to appreciate” and “now difficult to comprehend” how Hegel “dominated”, “defined”, “overshadowed”, and “reigned” in “Germany”, “England”, “American universities”, and “the philosophical world” in “the beginning of the nineteenth century”, “from 1818 until his death in 1831”, “the time from 1830 to 1840”, “the second quarter of the nineteenth century”, “the end of the nineteenth century”, and “the time Freud’s thinking developed” (Freud was born 1856 and would have been in university in the 1870s).

I will take this as evidence that Hegel was really really important for the entire nineteenth century.

On the other hand, it’s hard to find many people who will put in good words for him now. In fact, hilarious pithy denunciations of Hegel are an entire sub-genre. Hegel’s Wikiquote page, among other sources, includes:

“Hegel’s philosophy illustrates an important truth, namely, that the worse your logic, the more interesting the consequences to which it gives rise.” – Bertrand Russell

“When I was young, most teachers of philosophy in British and American universities were Hegelians, so that, until I read Hegel, I supposed there must be some truth to his system; I was cured, however, by discovering that everything he said on the philosophy of mathematics was plain nonsense. Hegel’s philosophy is so odd that one would not have expected him to be able to get sane men to accept it, but he did. He set it out with so much obscurity that people thought it must be profound. It can quite easily be expounded lucidly in words of one syllable, but then its absurdity becomes obvious.” – Bertrand Russell

“Among Noah’s sons was one who covered the shame of his father, but the Hegelians are still tearing away the cloak which time and oblivion had sympathetically thrown over the shame of their Master.” – Heinrich Schumacher

“Hegel’s was an interesting thesis, giving unity and meaning to the revolutions of human affairs. Like other historical theories, it required, if it was to be made plausible, some distortion of facts and considerable ignorance. Hegel, like

Mane and Spengler after him, possessed both these qualifications." – Bertrand Russell (are you starting to notice a trend here?)

"While scientists were performing astounding feats of disciplined reason [during the Enlightenment], breaking down the barriers of the "unknowable" in every field of knowledge, charting the course of light rays in space or the course of blood in the capillaries of man's body — what philosophy was offering them, as interpretation of and guidance for their achievements was the plain Witchdoctory of Hegel, who proclaimed that matter does not exist at all, that everything is Idea (not somebody's idea, just Idea), and that this Idea operates by the dialectical process of a new "super-logic" which proves that contradictions are the law of reality, that A is non-A, and that omniscience about the physical universe (including electricity, gravitation, the solar system, etc.) is to be derived, not from the observation of facts, but from the contemplation of that Idea's triple somersaults inside his, Hegel's, mind. This was offered as a philosophy of reason." – Ayn Rand (unsurprisingly)

A book review by Roger Kimball helps round out the picture. Along with presenting the legend that Hegel said that "only one person ever understood me, and even he misunderstood me", Kimball writes:

> Like many people who have soldiered through a fair number of Hegel's books, I was both awed and depressed by their glittering opacity. With the possible exception of Heidegger, Hegel is far and away the most difficult "great philosopher" I have ever studied. There was much that I did not understand. I secretly suspected that no one—not even my teachers—really understood him, and it was nice to have that prejudice supported from the master's own lips.

> Is it worth the effort? I mean, you spend a hundred hours poring over *The Phenomenology of Spirit*—widely considered to be Hegel's masterpiece—and what do you have to show for it? The book is supposed to take you from the naïve, "immediate" position of "sense certainty" to Absolute Knowledge, "or Spirit that knows itself as Spirit." That sounds pretty good, especially when you are, say, eighteen and are busy soaking up ideas guaranteed to mystify and alarm your parents. But what do you suppose it means?

Despite trying really hard to say some nice things about Hegel, just about the best that Kimball can do is:

> So why read Hegel? Just as doctors learn a lot about health by studying diseases, so we can learn a lot about philosophical health by studying Hegel.

The phrase "damning with faint praise" seems insufficient here.

Worse, Hegel has been criticized as a racist, a totalitarian, a proto-Nazi, and the kind of rationalist everyone hates – complete with stories about how he proved from first principles that there were only seven planets (not quite true, although he does seem to have made some similar inexcusable scientific errors. He was mocked (with some justice) for believing that his own work represented the final achievement of God's plan for the Universe, and that the objective progress of history had culminated in the early 19th century Prussian state.

As a result, when I spent four years getting a bachelors in Philosophy, not only did I not receive a word of instruction in Hegel, but I was actively pushed away from him with frequent derogatory references.

I should qualify all this. Part of it is the analytic-continental divide. Hegel ended up well on the continental side of that, so even though analytics have a dim opinion of him, I'm pretty sure he remains studied and well-respected within continental circles. Indeed, the split may have necessitated analytics dismiss him in order to justify ignoring him, given that not ignoring him would mean engaging him would mean reading him would meaning not having the time or energy to do anything else.

But since we've already brought in Google as a philosophical authority, we might as well note that it autocompletes "hegel is" into "hegel is impossible to understand". This seems to be pretty close to a consensus position right now.

II.

I know pretty much nothing about Hegel and am not nearly qualified to have an opinion on the debate about whether his inscrutability conceals deep wisdom or total nonsense. But there are a few points I draw from his rise and fall without being able to judge it philosophically.

I deliberately avoided discussing philosophy in my post How Common Are Science Failures?, first because it's outside the reference class but second because philosophy can't even get its act together enough to fail. These sorts of "science failures" are cases where the scientific community unites around a single consensus belief, but later discovers that belief was disastrously wrong. But philosophy can practically never unite around a single consensus belief, and it rarely disproves anything thoroughly enough to admit the error.

Hegel seems like a rare example of a philosophical consensus caught in contradiction. For a good chunk of the 19th century a very large part of the philosophical community agreed Hegel had solved everything, was a genius, was the be-all and end-all of philosophy. Later, at least the British and American communities did a total about-face and concluded that Hegel was a crackpot who, if he didn’t *invent* the technique of “if you can’t convince ’em, confuse ’em”, at least perfected it.

You can go one of two directions with this. First, you can say that people in the past were very gullible, that this confirms our prejudice that philosophers are silly people who will believe pretty much anything if it is billed as metaphysics and contains some confusing references to being and spirit.

Or you could say that people nowadays are so vapid, so demanding of instant gratification and unwilling to cover large inferential distances, that we’ve lost the ability to understand difficult ideas like those of Hegel.

I am the first type of person by temperament, but trying to become more sympathetic to the second way of thinking. Part of this is because on the rare occasions I do understand something difficult, I am acutely aware of all the people accusing it of being a confusing mass of jargon disguising a lack of real insight – and of how wrong these people are. “Ha ha, look at all these smart erudite domain experts who believe a stupid thing, that just proves smart domain experts lack common sense” now seems like a *huge failure mode* to me. There’s also a certain intellectual version of Chesterton’s Fence which looks kind of like “Don’t dismiss an idea until you can see why it would be so tempting for other people to believe”. Right now I don’t see the temptation in Hegel or for that matter any of Continental philosophy. That half of the philosophical universe, including many people who display objective signs of brilliance – has decided to just wallow in pointless obscurantism seems to beggar belief.

My inability to be tempted by Hegel brings me to another point: what parts of my thought, right now, are Hegelian? Hegel seems like a classic case where we should read history of philosophy backwards – if almost all philosophical thought for fifty to a hundred years was Hegelian, modernity should be absolutely *saturated* with Hegelian ideas. That means I might get less gain from trying to read Hegel forward (to see if he has startling insights I didn't know) and more gain from trying to read him backwards (to see if he is the source of things I assumed unquestioningly, and that negating them – as the contingent opinions of some German guy who thought 19th century Prussia was objectively perfect – would produce startling insights).

I don't know enough Hegel to do a good job of this. One easy target might be the modern belief in human progress or linear history. Fukuyama ("The End of History") writes:

> For better or worse, much of Hegel's historicism has become part of our contemporary intellectual baggage. The notion that mankind has progresses through a series of primitive stages of consciousness on his path to the present, and that these stages corresponded to concrete forms of social organization, such as tribal, slave owning, theocratic, and finally democratic egalitarian societies, has become inseparable form the modern understanding of man. Hegel was the first philosopher to speak the language of modern social science, insofar as man for him was the product of his concrete historical and social environment and not, as earlier natural right theorists would have it, a collection of more or less fixed "natural" attributes. The mastery and transformation of man's natural environment through the application of science and technology was originally not a Marxist concept, but a Hegelian one. Unlike later historicists whose historical relativism

> degenerated into relativism tout court, however, Hegel believed that history culminated in an absolute moment — a moment in which a final, rational form of society and state became victorious

But I find both more unexpected and more plausible David Chapman's theories that Hegel inspired modern Westernized Buddhism, the hippie movement, and the New Age. He breaks his arguments into a bunch of posts that aren't really collected in any organized way, but I would recommend An Improbable Re-Animation, Bad Ideas From Dead Germans, and Zen vs. The US Navy. Chapman's argument isn't very developed, but just raising the idea is enough to make its evidential support obvious. Hegel's system was based around the principle that the key principle of the universe was a divine Mind trying to find itself, that everything was interrelated and purposeful, that as this Mind became more self-aware it would be reflected in increasing levels of consciousness among human beings culminating in an ideal utopian social arrangement. This is the daaaaaawning of the Age of Aquarius, the Age of Aquarius…

Philosophy makes for strange bedfellows. Imagine: December 21, 2012. A ray of crystal light emerges from the Temple of Kukulcan in the Mayan ruins, piercing the center of the Milky Way. Humans ascend to a new level of consciousness. And all around the world people throw off their shackles and self-organize into intentional communities exactly resembling early 19th century Prussia.

Book Review: Singer on Marx

I’m not embarassed for choosing Singer’s Marx: A Very Short Introduction as a jumping-off point for learning more leftist philosophy. I weighed the costs and benefits of reading primary sources versus summaries and commentaries, and decided in favor of the latter.

The clincher was that the rare times I felt like I really understand certain thinkers and philosophies on a deep level, it’s rarely been the primary sources that did it for me, even when I’d read them. It’s only after hearing a bunch of different people attack the same idea from different angles that I’ve gotten the gist of it. The primary sources – especially when they’re translated, especially when they’re from the olden days before people discovered how to be interesting – just turn me off. Singer is a known person who can think and write clearly, and his book was just about the shortest I could find, so I jumped on it, hoping I would find a more sympathetic portrayal of someone whom my society has been trying to cast as a demon or monster.

And I don’t know if this is an artifact of Singer or a genuine insight into Marx, but as far as I can tell he’s even worse than I thought.

I.

What really clinched this for me was the discussion of Marx’s (lack of) description of how to run a communist state. I’d always heard that Marx was long on condemnations of capitalism and short on blueprints for communism, and the couple of Marx’s works I read in college confirmed he really didn’t talk about that very much. It seemed like a pretty big gap.

But I'd always dismissed this as an excusable error. When I was really young – maybe six or seven – I fancied myself a great inventor. The way I would invent something – let's say a spaceship – was to draw a picture of a spaceship. I would label it with notes like "engine goes here" and "power source here" and then rest on my laurels, satisfied that I had invented interstellar travel at age seven. It always confused me that adults, who presumably should be pretty smart, had failed to do this. Occasionally I would bring this up to someone like my parents, and they would ask a question like "Okay, but how does the power source work?" and I would answer "Through quantum!" and then get very annoyed that people *didn't even know about quantum*.

(I was seven years old. What's *your* excuse, New Age community?)

I figured that Marx had just fallen into a similar trap. He'd probably made a few vague plans, like "Oh, decisions will be made by a committee of workers," and "Property will be held in common and consensus democracy will choose who gets what," and felt like the rest was just details. That's the sort of error I could at least sympathize with, despite its horrendous consequences.

But in fact Marx was philosophically opposed, as a matter of principle, to any planning about the structure of communist governments or economies. He would come out and say it was irresponsible to talk about how communist governments and economies will work. He believed it was a scientific law, analogous to the laws of physics, that once capitalism was removed, a perfect communist government would form of its own accord. There might be some very light planning, a couple of discussions, but these would just be epiphenomena of the governing historical laws working themselves out. Just as, a dam having been removed, a river will eventually reach the sea somehow, so capitalism having been removed society will eventually reach a perfect state of freedom and cooperation.

Singer blames Hegel. Hegel viewed all human history as the World-Spirit trying to recognize and incarnate itself. As it overcomes its various confusions and false dichotomies, it advances into forms that more completely incarnate the World-Spirit and then moves onto the next problem. Finally, it ends with the World-Spirit completely incarnated – possibly in the form of early 19th century Prussia – and everything is great forever.

Marx famously exports Hegel's mysticism into a materialistic version where the World-Spirit operates upon class relations rather than the interconnectedness of all things, and where you don't come out and *call* it the World-Spirit – but he basically keeps the system intact. So once the World-Spirit resolves the dichotomy between Capitalist and Proletariat, then it can more completely incarnate itself and move on to the next problem. Except that this is the final problem (the proof of this is trivial and is left as exercise for the reader) so the World-Spirit becomes fully incarnate and everything is great forever. And you want to *plan* for how that should happen? Are you saying you know better than the World-Spirit, Comrade?

I am starting to think I was previously a little too charitable toward Marx. My objections were of the sort "You didn't really consider the idea of welfare capitalism with a social safety net" or "communist society is very difficult to implement in principle," whereas they should have looked more like "You are basically just telling us to destroy all of the institutions that sustain human civilization and trust that what is *basically* a giant planet-sized ghost will make sure everything works out."

[EDIT: Some Communists are taking this joke completely literally and getting angry because they think I am accusing them of believing in ghosts. This was not my intent, and I had hoped a tradition whose core text starts with the phrase "a specter is haunting Europe" would understand that ghost references can sometimes be used metaphorically. The Hegelian World-Spirit is an abstract force governing the progress of history; Marx's adaptation is even less spiritual. But my understanding is that it preserves the sense in which communists don't have to worry about the specifics of their system, because History is subject to a force that automatically sorts things out. I think "a ghost will solve everything" accurately reflects my skepticism about this sort of abdication of responsibility. And to forestall an obvious objection – yes, "the free market will sort everything out" also seems a little ghostly (cf. "invisible hand"), which is related to why I don't believe this.]

II.

Conservatives always complain that liberals "deny human nature", and I had always thought that complaint was unfair. Like sure, liberals say that you can make people less racist, and one could counterargue that a tendency toward racism is inborn, but it sure seems like you can make that tendency more or less strongly expressed and that this is important. This is part of the view I argue in Nature Is Not A Slate, It's A Series Of Levers.

But here I have to give conservatives their due. As far as I can tell, Marx literally, so strongly as to be unstrawmannable, believed there was no such thing as human nature and everything was completely malleable.

> Feuerbach resolves the essence of religion into the essence of man. But the essence of man is no

> abstraction inherent in each single individual. In reality, it is the ensemble of the social relations.

And:

> It is evidence that economics establishes an alienated form of social intercourse as the essential, original, and natural form

Which Singer glosses with:

> This is the gist of Marx’s objection to classical economics. Marx does not challenge the classical economists within the presuppositions of their science. Instead, he takes a viewpoint outside those presuppositions and argues that private property, competition, greed, and so on are to be found only in a particular condition of human existence, a condition of alienation.

I understand this is still a matter of some debate in the Marxist community. But it seems to me that if Singer is right, if this is genuinely Marx’s view, it seems likely to be part of what contributed to his inexcusable error above.

You or I, upon hearing that the plan is to get rid of all government and just have people share all property in common, might ask questions like “But what if someone wants more than their share?” Marx had no interest in that question, because he believed that there was no such thing as human nature, and things like “People sometimes want more than their shares of things” are contingent upon material relations and modes of production, most notably capitalism. If you get rid of capitalism, human beings change completely, such that “wanting more than your share” is no more likely than growing a third arm.

A lot of the liberals I know try to distance themselves from people like Stalin by saying that Marx had a pure original doctrine that they corrupted. But I am finding myself much more sympathetic to the dictators and secret police. They may not have been very nice people, but they were, in a sense, operating in Near Mode. They couldn’t just tell themselves “After the Revolution, no one is going to demand more than their share,” because their philosophies were shaped by the experience of having their subordinates come up to them and say “Boss, that Revolution went great, but now someone’s demanding more than their share, what should we do?” Their systems seem to be part of the unavoidable collision of Marxist doctrine with reality. It’s possible that there are other, better ways to deal with that collision, but “returning to the purity of Marx” doesn’t seem like a workable option.

III.

There was one part that made me more sympathetic to Marx. Singer writes:

> Marx saw that the liberal definition of freedom is open to a fundamental objection. Suppose I live in the suburbs and work in the city. I could drive my car to work, or take the bus. I prefer not to wait around for the bus, and so I take my car. Fifty thousand other people living in my suburb face the same choice and make the same decision. The road to town is choked with cars. It takes each of us an hour to travel ten miles. In this situation, according to the liberal conception of freedom, we have all chosen freely. Yet the outcome is something none of us want. If we all went by bus, the roads would be empty and we could cover the distance in twenty minutes. Even with the inconvenience of waiting at the bus stop, we would all prefer that. We are, of course,

> free to alter our choice of transportation, but what can we do? While so many cars slow the bus down, why should any individual choose differently? The liberal conception of freedom has led to a paradox: we have each chosen in our own interests, but the result is in no one’s interest. Individual rationality, collective irrationality…
>
> Marx saw that capitalism involves this kind of collective irrationality. In precapitalist systems it was obvious that most people did not control their own destiny – under feudalism, for instance, serfs had to work for their lords. Capitalism seems different because people are in theory free to work for themselves or for others as they choose. Yet most workers have as little control over their lives as feudal serfs. This is not because they have chosen badly, nor is it because of the physical limits of our resources and technology. It is because the cumulative effect of countless individual choices is a society that no one – not even the capitalists – has chosen. Where those who hold the liberal conception of freedom would say we are free because we are not subject to deliberate interference by other humans, Marx says we are not free because we do not control our own society.

This is good. In fact, this is the insight that I spent about fifteen years of my life looking for, ever since I first discovered libertarianism and felt like there was definitely an important problem with it, but couldn’t quite verbalize what it was. It’s something I finally figured out only within the last year or so and didn’t fully write up until Meditations on Moloch. And Marx seems to have sort of had it. I read the relevant section of Marx when I was younger, where he was talking about how capitalists would compete each other into the ground whether they wanted to or not, and I remember dismissing it with a “capitalists have not competed each other into the ground, for this this and this reason”, dismissing the incorrect object-level argument without realizing the important meta-level insight beneath it (something I have since learned to stop doing). If Marx really had that meta-level insight – really had it, and not just stumbled across a couple of useful examples of it without realizing the pattern – then that would make his fame justly deserved.

But two things here discourage me. First, Marx seems so confused about everything that it’s hard to parse him as really understanding this, as opposed to simply noticing one example of it that serves as a useful argument against capitalism. I notice Singer had to come up with his own clever example of this instead of quoting anything from any of Marx’s works. Second, the insight does not seem original to Marx. Tragedy of the commons was understood as early as 1833 and Malthus was talking about similar problems related to population explosions before Marx was even born. John Stuart Mill, writing twenty years before *Das Kapital*, had already explained the basic principle quite well:

> To a fourth case of exception I must request particular attention, it being one to which as it appears to me, the attention of political economists has not yet been sufficiently drawn. There are matters in which the interference of law is required, not to overrule the judgment of individuals respecting their own interest, but to give effect to that judgment: they being unable to give effect to it except by concert, which concert again cannot be effectual unless it receives validity and sanction from the law. For illustration, and without prejudging the particular point, I may advert to the question of diminishing the hours of labour. Let us suppose, what is at least supposable, whether it be the fact or not—that a general reduction of the hours of

> factory labour, say from ten to nine,*119 would be for the advantage of the workpeople: that they would receive as high wages, or nearly as high, for nine hours’ labour as they receive for ten. If this would be the result, and if the operatives generally are convinced that it would, the limitation, some may say, will be adopted spontaneously. I answer, that it will not be adopted unless the body of operatives bind themselves to one another to abide by it. A workman who refused to work more than nine hours while there were others who worked ten, would either not be employed at all, or if employed, must submit to lose one-tenth of his wages. However convinced, therefore, he may be that it is the interest of the class to work short time, it is contrary to his own interest to set the example, unless he is well assured that all or most others will follow it. But suppose a general agreement of the whole class: might not this be effectual without the sanction of law? Not unless enforced by opinion with a rigour practically equal to that of law. For however beneficial the observance of the regulation might be to the class collectively, the immediate interest of every individual would lie in violating it: and the more numerous those were who adhered to the rule, the more would individuals gain by departing from it.

So one might apply to Marx the old cliche: that he has much that is good and original, but what is good is not original and what is original is not good.

But it is interesting to analyze Marx as groping toward something game theoretic. This comes across to me in some of his discussions of labor. Marx thinks all value is labor. Yes, capital is nice, but in a sense it is only “crystallized labor” – the fact that a capitalist owns a factory only means that at some other point he got laborers to build a factory for him.

So labor does everything, but it gets only a tiny share of the gains produced. This is because capitalists are oppressing the laborers. Once laborers realize what's up, they can choose to labor in such a way as to give themselves the full gains of their labor.

I think here that he is thinking of coordination as something that happens instantly in the absence of any obstacle to coordination, and the obstacle to coordination is the capitalists and the "false consciousness" they produce. Remove the capitalists, and the workers – who represent the full productive power of humanity – can direct that productive power to however it is most useful. In my language, Marx simply *assumed* the invisible nation, thought that the result of perfect negotiation by ideal game theoretic agents with 100% cooperation under a veil of ignorance – would also be the result of real negotiation in the real world, as long as there were no capitalists involved. Maybe this idea – of gradually approaching the invisible nation – is what stood in for the World-Spirit in his dialecticalism. Maybe in 1870, this sort of thinking was excusable.

If capitalists are to be thought of as anything other than parasites, part of the explanation of their contribution has to involve coordination. If Marx didn't understand that coordination is just as hard to produce as linen or armaments or whatever, if he thought you could just *assume* it, then capitalists seem useless and getting rid of all previous forms of government so that insta-coordination can solve everything seems like a pretty swell idea.

If you admit that, capitalists having disappeared, there's still going to be competition, positive and negative sum games, free rider problems, tragedies of the commons, and all the rest, then you've got to invent a system that solves all of those issues better than capitalism does. That seems to be the real challenge Marxist intellectuals should be setting themselves, and I hope to eventually discover some who have good answers to it. But at least from the little I learned from Singer, I see no reason to believe Marx had the clarity of thought to even understand the question.

>

Does Class Warfare Have A Free-Rider Problem?

Many people believe that the rich have captured government for their own ends. For example, rich people use their money and power to decrease tax rates on the wealthy and sabotage legislation meant to protect the working man.

But this raises a coordination problem. After all, suppose you are a rich person who makes $1 million per year. You would like the government to cut federal taxes on the wealthy from 40% down to 30%, which would save you $100,000 per year. One might think you would be willing to spend up to $100,000 to effect this goal.

But in fact it requires the concerted effort of all the rich people across the country to make this happen. A single $100,000 donation isn't going to change federal level policy in such a spectacular way. Realistically your effort will be a drop in a bucket that your entire class needs to contribute to.

So we should expect free rider problems. Suppose a representative of the Rich People's Union asks for a $10,000 donation to fight for lower taxes. There are hundreds of thousands of rich people, so you're pretty sure your one donation isn't going to push anything over the edge one way or the other. Supposing the tax cut goes through, you will get the same benefit whether you donated or not; supposing it doesn't, you won't gain anything either way. It's easy to see that in either case the rational self-interested thing to do is to refuse to donate.

There are a couple of rare exceptions to this. If you are Bill Gates and make a billion dollars a year, so that you would gain $100 million from the tax cut, it might be worth bribing the necessary legislators all on your own, on the grounds that if something needs to be done right you had better do it yourself. Likewise, if you're Exxon Mobil or the Koch brothers, then you might be a big enough chunk of the target population for certain specific environmental regulations that it's worth using your own money to fight it whether or not others join in.

But a general focus on the interests of the rich? Not likely.

We can compare this to good people playing normal politics, who usually have a hard time rallying support for important causes like global warming or net neutrality. In the same way, rich people playing at the politics of greed should have a hard time convincing their target demographic to get out of bed and join in their attempts to soak the poor.

But in fact their job is much harder. When good people do rally the masses to their cause, it seems to be through an appeal to principle. Like "Yes, I know it would be much easier for you to sit back and let other people solve global warming, but you have an *ethical responsibility* to participate in this, and won't you feel good about yourself knowing you've made a difference."

Obviously if your campaign is "soak the poor to line our own pockets" this is harder to pull off.

Is there any model in which the rich could successfully manipulate politics for their own benefit?

I'm reminded of the research I looked at in Plutocracy Isn't About Money. People seem to donate surprisingly little to political candidates, and donations don't seem to help. This seems consistent with the idea that rich people don’t directly coordinate to bribe politicians in their favor. I suggested a couple of different hypotheses, like that maybe the rich win because of “soft power” – ie the media and universities and politicians are mostly rich or are run by rich people who just sort of naturally let their opinions percolate through without much deliberate effort.

An alternative explanation preserves our intuitive belief that the rich sure do seem to influence politics a lot. Maybe rich people, like poor people, participate in politics because of sincere belief in their moral values, and their values are by what seems a weird coincidence the ones that help make them richer.

Like, Mitt Romney’s zillion-dollar-a-plate fundraisers seem to always be pretty full. It can’t literally be in a rich person’s self-interest to buy a plate there. But a lot of rich people could have conservative-libertarian-pro-business ideas that encourage them to quasi-altruistically support Mitt Romney in order to push their values.

But this is really weird and interesting – much more interesting than it looks. It suggests that, in the presence of a useful selfish goal to coordinate around, a value system will “spring up” that convinces people to support it for altruistic reasons.

I’m not just talking about normal altruism here. A rich person motivated by normal altruism per se might be against tax cuts for the rich, in order to better preserve social services for the less fortunate. And I’m not just talking about normal selfishness either. A rich person motivated by selfishness would hang out in his mansion all day instead of wasting money on fundraisers. I’m talking about a moral system which is genuinely self-sacrificing on the individual level, but which when universalized has the effect of helping the rich person get richer.

It’s worth thinking about this in contractarian terms. A rich person, minus the veil of ignorance, wouldn’t support everyone pitching in to help the poor, because he knows he’s not poor and so gains nothing. A rich person, minus the veil of ignorance, *would* support a binding pact among all rich people to pitch in to support tax cuts on the rich, because she knows she would gain more than she loses from such an agreement.

But as far as I can tell, this calculation is never made on a conscious level. What happens on a conscious level is the rich person finds themselves supporting some moral philosophy – libertarianism, Objectivism, prosperity gospel, whatever – which says it is morally wrong to raise taxes on the rich, so much so that one should altruistically make personal sacrifices in order to stop them from being raised. And then these moral philosophies spread, and without any conscious awareness, the rich people find themselves coordinating very nicely to protect their class interests.

I hope you agree that if this is true, it’s spooky. I admit on this blog I sometimes mock human nature and human cognition a little too much, but this particular cognitive process is *really impressive*. I hope whatever angel designed it got a promotion.

So although I haven’t really thought this through too much, I would suggest a dichotomy. Either there’s some sort of spooky system that generates heartfelt moral philosophies on demand to solve coordination problems, or the rich aren’t actually coordinating and just consistently keep getting lucky.

I don’t like this because it raises more questions than it answers. Why don’t the poor coordinate this well? Too many of them? And if this is true, how much more credence should we give other conspiracy theories?

[Ozy] A Response to Spandrell

[Content note: Gender, relationships, sexuality. Some sexually explicit content. Discussion without endorsement of various forms of transphobia, homophobia, et cetera. Ozy wishes you to know they wrote this in a very timely manner after Spandrell's original post and I just took forever to publish it.]

I made fun of this post on my tumblr and then Scott requested I actually argue with it.

First, let's address the issue of homosexuality. Spandrell argues that "There's no way on earth that a condition that makes you lose attraction towards the opposite sex is going to survive natural selection." On the contrary, there is a lot of animal homosexuality. The linked book contains much fascinating information, such as the fact that animal sexuality has been documented in almost 500 species and that, in one study, ninety percent of observed giraffe sex was between two males. I am not sure why animal homosexuality is so common: I am not an evolutionary biologist myself. But it suggests that the simplistic model in which fucking something other than a vagina is not selected for is incorrect.

In addition, homosexuality is probably not inborn. A Swedish twin study with a sample size of 7600 found that genetic factors and shared-environment factors together explained only a third of the variance in sexual orientation, while two-thirds were explained by unshared environment. In short: sexual orientation in humans is less inborn than how hardworking you are. Indeed, Spandrell admits as much, saying that we do not know the cause of gayness. Maybe because it's not inborn? Just saying.

One must point out that the "born this way" myth was invented by LGBT people to get people to accept us: "we can’t help it! It is mean to hurt people because of something they can’t help! Don’t worry, it’s genetic, accepting us won’t make anyone else gay!” I don’t fully understand what the Cathedral is, but if anything is part of the Cathedral the Human Rights Campaign is, and I feel like that is a fairly depressing amount of belief in the Cathedral’s myths from a self-declared neoreactionary.

Spandrell argues that female paraphiliacs do not exist because they do not usually tell researchers about being paraphiliacs. Unfortunately, he is missing the very large confounding variable, which is that women are fucking liars about sex. As I pointed out in my Anti-Heartiste FAQ, evidence suggests that the entire sexual partner gap between men and women is explicable by women being goddamned liars. There is no reason to believe they wouldn’t also be goddamned liars about their paraphilias.

Spandrell challenged me in his comment section– if female paraphilia is a thing– to find cases of female death by autoerotic asphyxiation. It is true that women are less likely to die by autoerotic asphyxiation. However, women are less likely than men to masturbate, and even when they do they masturbate less often than men do, decreasing the risk of women dying through masturbation. However, this is self-report data and thus falls under the “women are goddamned liars” explanation. Autoerotic asphyxiation deaths are massively undercounted to begin with; it is relatively common for people who die by autoerotic asphyxiation to be mistaken for suicides or “sanitized” by family members who don’t want to admit their child died by masturbation. Given that women lie massively about sex, it is possible that families are more likely to sanitize female autoerotic asphyxiators. Finally, I hate to be the feminist who points this out to the neoreactionary, but men and women are different. This probably extends to sexual fetishes. I admit that none of these are particularly solid arguments. However, I do have reason to believe that women have things that may be considered paraphilias.

Porn.

The rise of the ebook has massively expanded the amount of porn that women read. Like I said, women are fucking liars about sex. They want to read porn, but they don't want to admit that they want to read porn– and as plausibly deniable as Harlequins are, those Fabio covers make it look a little too much like porn for a lot of readers.

Ellora's Cave is the largest erotic ebook producer in the United States. If you are curious whether women have paraphilias, you can explore the BDSM Elements section, featuring such titles as Taming the Raven's Son, Pack and Mate, and Elf Struck (tagline: "When a BDSM slut is matched with a warrior virgin, both tempers and desires flare.")

Part of the problem here is that I don't fully understand what qualifies as a 'paraphilia' in Spandrell's analysis. Spandrell provides as examples: "There are all sorts of paraphilias, all of which seem to only occur in men. Some men are attracted to babies, others to feet, others to shoes, others to obese women, others to old women. There's a lot of weird stuff out there." If we are going to the "at least as weird as being attracted to fat women" standard, then I feel like a lot of non-BDSM things in Ellora's Cave count. For instance, paranormal erotic romance is basically just a fetish for fucking vampires and werewolves.

However, I suspect that female paraphilias are also going to be structurally different than male paraphilias. Eliade's List of Fanfiction Kinks, Tropes, and Cliches is the most extensive list I'm aware of of fanfiction porn tropes. Literally, I have never been able to think of one that is popular and not on her list. The interesting thing about Eliade's list– which is something I've found personally in my fanfiction consumption– is the lack of distinction between purely sexual and purely narrative tropes. The list does include things like "intercrural or interfemoral sex (i.e., thrusting cock between partner's thighs)," but also things like "makeovers." I suspect a list of favorite male porn tropes would be unlikely to include makeovers. Similarly, it's a common observation that a plot what plot story on AO3, which is female-dominated, and an extraordinarily plotty story on Literotica, which is male-dominated, contain approximately the same amount of plot. I suspect when one studies female paraphilias one will find primarily narrative paraphilias: where men tend to fetishize a single act, women tend to fetishize an overall storyline. While one might not consider the latter to be a paraphilia, that seems to be far more related to an androcentric definition of paraphilia than a difference in the prevalence of paraphilias between men and women per se.

Finally, let us discuss trans women. To be honest, I don't fully understand what the difference between "trans women are homosexual men" and "trans women are heterosexual women" is. The empirical facts remain the same: many trans women transition as soon as possible, are attracted to men, and behave in ways typically considered feminine. All I can figure is that it is the result of a belief that we should call trans women men in order to be pointlessly upsetting to them.

I am aware of two studies applying Blanchard's autogynephilia questionnaire to a group of cisgender women. The first, unpaywalled here, I shall ignore because of its 29-person sample size, despite its astonishing revelation that 93% of cisgender women are autogynephiles by Blanchard’s definition. The second actually has a reasonable sample size, so let’s examine it more closely. The study divided autogynephiliac arousal into two categories– Autogynephiliac Interpersonal Fantasy (essentially, sexual fantasies about being admired as female) and the Core Autogynephilia Scale (essentially, sexual fantasies about being a very sexy woman). There was no difference between cisgender women and transgender women in the Autogynephiliac Interpersonal Fantasy scale. However, transgender women scored significantly higher on the Core Autogynephilia Scale.

To put it bluntly, this makes no goddamned sense. Cis women are just as likely as trans women to have a particular subtype of autogynephilia, but less likely to have autogynephilia itself?

Let us look at the Core Autogynephilia Scale a little more closely. The study authors modified the scale so that the cis woman population were asked if they have ever sexually fantasized about themselves having attractive or more attractive female body parts. However, imagine that you have a vagina and you have sexual fantasies in which you have a vagina. Nothing interesting here, probably going to mark “no” on the relevant questionnaire. Now imagine that you have a penis and you have sexual fantasies in which you have a vagina. You’re going to notice. This is contrary to expectations. If someone asks you “do you have sexual fantasies about having an attractive or more attractive vagina?”, you’re probably going to mark yes (assuming you don’t specifically fetishize having ugly genitals). The exact same behavior leads cis women to mark “no” and trans women to mark “yes.”

Essentially, autogynephilia is ordinary female sexuality. Women are often erotically aroused by dressing in lingerie and wearing makeup; women are erotically aroused by looking at themselves naked; women have sexual fantasies in which they have vulvas; for that matter, women are erotically aroused by imagining themselves as sexier than they are. If we assume that trans women are, well, women's minds in men's bodies, this entirely explains the autogynephilia data: women have female-typical sexuality instead of male-typical sexuality. (It does not explain the autogynephilia anecdotes, as one assumes it is quite uncommon for cis women to be aroused by the idea of knitting, but those seem to be selected for vividness rather than for representationality. One guy who is turned on by the idea of knitting does not mean that every trans woman who is attracted to other women is an autogynephile.)

Now, the pro-autogynephilia group may respond, "but it is normal for cis women to fantasize about having a vagina and deviant for trans women to!" But in that case there is no way for trans women to win. If they had sexual fantasies in which they had a penis, you would be like "ah, yes, that is proof they are men. Why would they even want sexual reassignment surgery if they are fine with having a penis?" Since they instead fantasize about having a vagina, you would be like "that is sexual deviancy!" There is no evidence that can convince you that trans women genuinely have what they say they have– a condition in which they are genuinely upset by their bodies, being seen as male, or both, which is best treated by allowing them to transition.

Spandrell opines that allowing trans women to transition and get sex reassignment surgery "can't work well, at the very least because men have male sex drives, which are a very dangerous thing when not constrained by women." I must remind him that the male sex drive is mediated through testosterone. Trans women typically take estrogens and anti-androgens, which lower the libido to the level of an otherwise-comparable cis woman. A woman who has had sexual reassignment surgery does not even have testicles to produce testosterone. She could not possibly have a male sex drive, unless Spandrell is advocating the theory that the male sex drive is actually mediated by ghost balls.

Finally, I must address the notion that I am an autoandrophile. First, I find it highly amusing that Spandrell believes I am the first trans person assigned female at birth to be attracted to men. I assure you I am certainly not. Second, my fetish is (mostly SFW, but TMI warning) very well documented. It is such a shame how no one ever does research before they insult you these days.

Third, I must clarify what I meant in that particular comment. In my experience, social dysphoria is subject to the hedonic treadmill: I was elated the first time someone called me ‘zie,’ but now it is an everyday thing. I imagine that if I went back, I would spend six months or so in a pit of constant dysphoria, but eventually get used to it. However, I have been constantly distressed by my breasts since puberty; when I thought I was cis, I would have constant fantasies of cutting them off with a knife; when I stop binding regularly, I notice a deep loss of psychological stability. The hedonic treadmill simply does not work for me having breasts. I value my relationship highly, but not that highly. (Being monogamous was a similar constant drain on me, and being polyamorous– several years after I started– is still a major contributor to my happiness, which is the reason I say it would be extremely hard to go back to monogamy.)

Joint Over- and Underdiagnosis

Today I had several more terrible lectures on ADHD.

In one of them, I was informed that America is medicalizing normal childhood mischief and loading anyone who gets worse than a B+ up with Ritalin or amphetamines as part of the pathologization of everyday life.

In another, I was informed that ADHD is shamefully underdiagnosed and most of the children who need stimulants most are going without them and failing school unnecessarily, so we need better screening programs and more efforts to seek out potential sufferers of the condition.

So I asked one of my attendings, Dr. L, which one it was. Are we overdosing ADHD? Or underdiagnosing it?

He answered that we are both overdiagnosing *and* underdiagnosing ADHD, the same as every other psychiatric disease, and then explained this so it made perfect sense and I was embarassed for not realizing it before.

Suppose that 3% of the population has ADHD.

Suppose that of people with ADHD, 50% of them realize they have ADHD like symptoms and go to a psychiatrist to get checked out.

Suppose that of people without ADHD, 10% of them falsely believe they have ADHD and also go to a psychiatrist to get checked out.

The Conners Continuous Performance Test is a commonly used test that evaluates children for ADHD. It is found to have a sensitivity of 75% and a specificity of 73%. In theory our system is based on faith that a trained psychiatrist can do better than a neuropsychological test; in practice they probably do much worse. Let's give them the benefit of the doubt and say this is an excellent psychiatrist who outperforms the test handily and has both a sensitivity and specificity of 85%.

We can see that of every 100 people, 3 will have ADHD and 97 won't. 1.5 true patients and 9.7 false patients will show up for psychiatric evaluation. The psychiatrist will diagnose 1.275 true patients and 1.455 false patients with the condition, and prescribes stimulants according to the diagnosis.

So we have three things that, surprisingly, all happen at once:

1. We have an excellent psychiatrist who outperforms the tests and is right 85% of the time.
2. The majority of people who are on Ritalin, shouldn't be.
3. The majority of people who should be on Ritalin, aren't.

Number two sounds a lot like what we mean by "overdiagnosis", and number three sounds a lot like what we mean by "underdiagnosis". So even with a pretty good psychiatrist acting honestly, we expect ADHD to be both overdiagnosed and underdiagnosed at the same time.

Even in conditions that do not quite satisfy the "majority" part of (2) and (3), we might still expect it to be true at the same time that a *sizeable chunk* of people diagnosed with the disease don't have it and a *sizeable chunk* of people with the disease aren't diagnosed.

If this seems counterintuitive, it is just another example of the annoying world of medical sensitivity and specificity statistics, which are constantly tripping up even the most experienced doctors. See also the infamous mammogram problem.

Once I understood this joint-overdiagnosis-and-underdiagnosis problem, several other candidate situations immediately leapt to mind. Antidepressants are almost certainly both overprescribed and underprescribed. So are opiate pain medications.

SSC Gives A Wedding Speech

[I gave a speech at Mike Blume and Hannah "Alicorn" Blume's wedding on Sunday. Some of the guests suggested I post it here for more general consumption. Content warning: polyamory.]

I've been asked to give an impromptu speech. Specifically, I was asked six months ago, when Hannah messaged me and said "You need to give an impromptu speech at my wedding. You've got six months to get it sounding impromptu enough."

But I've been thinking about this day for even longer than that. The first time Hannah and I talked about her wedding was…maybe three or four years ago. She was staying at my house in Southern California on her way to Anna and Carl's wedding. And this was actually an Important Historic Occasion, because the next night when she stayed in San Diego, in order to save money she shared a hotel room with a certain Michael Blume and the rest is history. But at the time they weren't together, and Hannah and I were – kind of half-dating, I don't think we had actually started dating at the time, but we were flirting. And Hannah asked if I was going to go to the wedding the next day, and I said no, I couldn't stand weddings, I hated weddings, I would do whatever I could to avoid them.

And she looked at me with big eyes and said "But…you'll come to *my* wedding? Right?"

I said: "Mumble mumble maybe mumble try."

Hannah wouldn't take that as an answer and demanded to know my *probability* that I would come to her wedding.

I remember what I answered. It was something like “Fifty percent. Rising to ninety percent, if I’m the groom.”

And that didn’t work out, but I still find that now that the time is here I wouldn’t miss it for the world.

Hannah is still one of my favorite people. When I was out of a job and had no idea what I was going to do with my life, Hannah kind of saved me and gave me a place to live in Berkeley and threw me at her friend group so hard I have never been able to extract myself since. I have known her for five years now, I dated her for about three, I lived with her for one. And in the end, my only regret about attending her wedding was that it means I am visiting Berkeley on the ONE weekend that she’s not throwing a dinner party with her home cooking.

I’ve known Mike for a lot shorter, only about two years. Which is too bad, because it means that there were all these years of my life when I could have known Mike, but didn’t, which is a tragic waste. Mike is kind of like the Platonic ideal of the Good – to know him is to love him. I remember one Facebook thread where someone posted “Mike Blume is so nice and helpful and dreamy” to their wall, and it ended up ballooning to like a hundred likes and comments from people agreeing with the sentiment.

I was looking for that thread the other day so I could quote it, and I couldn’t find it. Part of the *reason* I couldn’t find it was that I kept asking people – “Do you remember who posted that Facebook thread praising Mike for being really nice and attractive and helpful?” – and they would say “Yeah, I think I was the one who posted that, it sounds like the sort of thing I would say.” And then they look and can’t find it, so I go to the next person, and they’re like “You know, I bet I was the one who posted that, it sounds like the sort of thing I would say…" annnnnnd I never did find the thread.

That is Mike.

So instead I hunted down something I once said about Mike once on my blog, which I would like to share with you today: "Hannah says Mike is her 'happiness battery', a source of emotional strength she can rely on to get her through difficult times. After living with him, I felt the same way, and he is at the center of so many social circles he might better be described as a giant happiness hydroelectric plant powering half of Northern California. The fact that flowers do not spring up everywhere he walks only proves that flowers are wrong."

In fact, I took unfair advantage of this when I lived with Mike and Hannah to meet a steady stream of Mike groupies. That was how I met my current girlfriend Ozy – they dated Mike first. That was how I met my ex-girlfriend Kenzi, who officiated today – she dated Mike first. In fact, Mike and Kenzi were really good together. I used to wonder whether Hannah would marry Mike or Kenzi would marry Mike. I'm glad to see that they both married Mike, in different senses.

I'm trying to avoid using the phrase "emergent property" in a wedding speech, but I'll say it – there is an emergent property to their relationship that makes them even better together than either one is alone. Their interactions with each other show such amazing mutual respect and love and complementarity that it adds new plausibility to the idea of soulmates. They are my model of how a good relationship ought to work. And one day, I hope some ambitious linguist will study their ability to communicate with each other entirely in adorable high-pitched noises ("eeeeeeeeeee!" "EEEEEEEEEEE!")

I like Mike, and I like Hannah. But beyond either of them, I have a huge, huge crush on their relationship.

I want to marry their marriage.

I know my conservative friends tell me that we're on a slippery slope, and soon people will be marrying animals, and trees, and rocks. And I can only hope that, somewhere at the bottom of that slope, someone legalizes man-relationship unions.

And when that happens, the rest of you, stay away! I called it first!

My friends got MARRIED!

But I'm sorry to get into politics at a time like this. Let's talk about something more relevant. Let's talk about population genetics.

A Dr. Joseph Chang of Yale University, using [sophisticated statistical techniques](), determined that ancestry mixes surprisingly quickly across populations. I promise this will become relevant. He found that beyond a certain horizon anybody who's the ancestor of anybody in a population is the ancestor of everybody. The exact length changes depending on some assumptions, but for a relatively mixed population like descendants of Eurasians, it's probably around fifteen hundred years. Some tribes on remote islands way out in the Pacific might be longer. Anyone from Papua New Guinea here today? No?

Then everyone here today is a descendent of Socrates. Everyone here today is a descendent of Confucius. Everyone here today is a descendent of Mohammed. Even if you don't look much like him. Queen Elizabeth's official genealogy confirms a descent from Mohammed, and *she* doesn't look Middle Eastern either.

We're all descendants of Nefertiti. The patriarch Abraham. The Japanese imperial line. Charlemagne. Qin Shih Huang Di. And not just genetically. We learned values from our parents that they learned from their parents that they learned from their parents and so on to Socrates or Mohammed or Charlemagne sitting their kids down at the dinner table and trying to teach them right from wrong.

Mike and Hannah met through the Machine Intelligence Research Institute, through the Visiting Fellows program at Benton-that-was. A lot of people here today are involved with MIRI, or other organizations trying to ensure the survival of humanity a thousand or two thousand years from now. And there's a lot of discussion, within those circles, about what such a future would be like.

And I was reading about this population genetics stuff six months ago, at the same time Hannah asked me to write an impromptu speech, and it made me think.

Whatever else we’re celebrating with the ritual of marriage, we’re also celebrating this. We’re marking this incredibly audacious act of taking a genetic and memetic payload and shooting it into the far future, where it will spread further and further with every generation and eventually rewrite humankind.

And if we make it another fifteen hundred years as a biological species, someday we will have a world where everybody alive is a descendent of Mike and Hannah. And where everyone has received their values from someone who received their values who received their values…from Mike and Hannah.

And that’s pretty high up there for me as a reason to be incredibly excited about the whole project.

So, a toast. To Hannah. To Mike. To their relationship. And to the future.

Congratulations, guys.

(No pressure.)

ADDED: Here is the text of the wedding ceremony itself, written by Hannah.

Book Review: Red Plenty

I.

I decided to read *Red Plenty* because my biggest gripe after reading Singer's book on Marx was that Marx refused to plan how communism would actually work, instead preferring to leave the entire matter for the World-Spirit to sort out. But almost everything that interests me about Communism falls under the category of "how communism would actually work". Red Plenty, a semi-fictionalized account of the history of socialist economic planning, seemed like a natural follow-up.

But I'd had it on my List Of Things To Read for even longer than that, ever after stumbling across a quote from it on some blog or other:

> Marx had drawn a nightmare picture of what happened to human life under capitalism, when everything was produced only in order to be exchanged; when true qualities and uses dropped away, and the human power of making and doing itself became only an object to be traded.
>
> Then the makers and the things made turned alike into commodities, and the motion of society turned into a kind of zombie dance, a grim cavorting whirl in which objects and people blurred together till the objects were half alive and the people were half dead. Stock-market prices acted back upon the world as if they were independent powers, requiring factories to be opened or closed, real human beings to work or rest, hurry or dawdle; and they, having given the transfusion that made the stock prices come alive, felt their flesh go cold

> and impersonal on them, mere mechanisms for chunking out the man-hours. Living money and dying humans, metal as tender as skin and skin as hard as metal, taking hands, and dancing round, and round, and round, with no way ever of stopping; the quickened and the deadened, whirling on.
>
> And what would be the alternative? The consciously arranged alternative? A dance of another nature. A dance to the music of use, where every step fulfilled some real need, did some tangible good, and no matter how fast the dancers spun, they moved easily, because they moved to a human measure, intelligible to all, chosen by all.

Needless to say, this is Relevant To My Interests, which include among them poetic allegories for coordination problems. And I was not disappointed.

The book begins:

> Strange as it may seem, the gray, oppressive USSR was founded on a fairy tale. It was built on the twentieth-century magic called “the planned economy,” which was going to gush forth an abundance of good things that the lands of capitalism could never match. And just for a little while, in the heady years of the late 1950s, the magic seemed to be working. Red Plenty is about that moment in history, and how it came, and how it went away; about the brief era when, under the rash leadership of Khrushchev, the Soviet Union looked forward to a future of rich communists and envious capitalists, when Moscow would out-glitter Manhattan and every Lada would be better engineered than a Porsche. It’s about the scientists who did their genuinely

> brilliant best to make the dream come true, to give the tyranny its happy ending.

And this was the first interesting thing I learned.

There’s a very settled modern explanation of the conflict between capitalism and communism. Capitalism is good at growing the economy and making countries rich. Communism is good at caring for the poor and promoting equality. So your choice between capitalism and communism is a trade-off between those two things.

But for at least the first fifty years of the Cold War, the Soviets would not have come *close* to granting you that these are the premises on which the battle must be fought. They were officially quite certain that any day now Communism was going to prove itself *better* at economic growth, better at making people rich quickly, than capitalism. Even unofficially, most of their leaders and economists were pretty certain of it. And for a little while, even their capitalist enemies secretly worried they were right.

The arguments are easy to understand. Under capitalism, plutocrats use the profits of industry to buy giant yachts for themselves. Under communism, the profits can be reinvested back into the industry to build more factories or to make production more efficient, increasing growth rate.

Under capitalism, everyone is competing with each other, and much of your budget is spent on zero-sum games like advertising and marketing and sales to give you a leg up over your competition. Under communism, there is no need to play these zero-sum games and that part of the budget can be reinvested to grow the industry more quickly.

Under capitalism, everyone is working against everyone else. If Ford discovers a clever new car-manufacturing technique, their first impulse is to patent it so GM can't use it, and GM's first impulse is to hire thousands of lawyers to try to thwart that attempt. Under communism, everyone is working together, so if one car-manufacturing collective discovers a new technique they send their blueprints to all the other car-manufacturing collectives in order to help them out. So in capitalism, each companies will possess a few individual advances, but under communism every collective will have every advance, and so be more productive.

These arguments make a lot of sense to me, and they *definitely* made sense to the Communists of the first half of the 20th century. As a result, they were confident of overtaking capitalism. They realized that they'd started with a handicap – czarist Russia had been dirt poor and almost without an industrial base – and that they'd faced a further handicap in having the Nazis burn half their country during World War II – but they figured as soon as they overcame these handicaps their natural advantages would let them leap ahead of the West in only a couple of decades. The great Russian advances of the 50s – Sputnik, Gagarin, etc – were seen as evidence that this was already starting to come true in certain fields.

And then it all went wrong.

II.

Grant that communism really does have the above advantages over capitalism. What advantage does capitalism have?

The classic answer is that during communism no one wants to work hard. They do as little as they can get away with, then slack off because they don't reap the rewards of their own labor.

Red Plenty doesn't really have theses. In fact, it's not really a non-fiction work at all. It's a dramatized series of episodes in the lives of Russian workers, politicians, and academics, intended to come together to paint a picture of how the Soviet economy worked.

But if I can impose a thesis upon the text, I don't think it agreed with this. In certain cases, Russians were *very* well-incentivized by things like "We will kill you unless you meet the production target". Later, when the state became less murder-happy, the threat of death faded to threats of demotions, ruined careers, and transfer to backwater provinces. And there were equal incentives, in the form of promotion or transfer to a desirable location such as Moscow, for overperformance. There were even monetary bonuses, although money bought a lot less than it did in capitalist countries and was universally considered inferior to status in terms of purchasing power. Yes, there were Goodhart's Law type issues going on – if you're being judged per product, better produce ten million defective products than 9,999,999 excellent products – but that wasn't the crux of the problem.

Red Plenty presented the problem with the Soviet economy primarily as one of allocation. You could have a perfectly good factory that could be producing lots of useful things if only you had one extra eensy-weensy part, but unless the higher-ups had allocated you that part, you were out of luck. If that part happened to break, getting a new one would depend on how much clout you (and your superiors) pulled versus how much clout other people who wanted parts (and their superiors) held.

The book illustrated this reality with a series of stories (I’m not sure how many of these were true, versus useful dramatizations). In one, a pig farmer in Siberia needed wood in order to build sties for his pigs so they wouldn’t freeze – if they froze, he would fail to meet his production target and his career would be ruined. The government, which mostly dealt with pig farming in more temperate areas, hadn’t accounted for this and so hadn’t allocated him any wood, and he didn’t have enough clout with officials to request some. A factory nearby had extra wood they weren’t using and were going to burn because it was too much trouble to figure out how to get it back to the government for re-allocation. The farmer bought the wood from the factory in an under-the-table deal. He was caught, which usually wouldn’t have been a problem because *everybody* did this sort of thing and it was kind of the “smoking marijuana while white” of Soviet offenses. But at that particular moment the Party higher-ups in the area wanted to make an example of someone in order to look like they were on top of their game to *their* higher-ups. The pig farmer was sentenced to years of hard labor.

A tire factory had been assigned a tire-making machine that could make 100,000 tires a year, but the government had gotten confused and assigned them a production quota of 150,000 tires a year. The factory leaders were stuck, because if they tried to correct the government they would look like they were challenging their superiors and get in trouble, but if they failed to meet the impossible quota, they would all get demoted and their careers would come to an end. They learned that the tire-making-machine-making company had recently invented a new model that really *could* make 150,000 tires a year. In the spirit of Chen Sheng, they decided that since the penalty for missing their quota was something terrible and the penalty for sabotage was also something terrible, they might as well take their chances and destroy their own machinery in the hopes the government sent them the new improved machine as a replacement. To their delight, the government believed their story about an “accident” and allotted them a new tire-making machine. *However*, the tire-making-machine-making company had decided to cancel production of their new model. You see, the new model, although more powerful, weighed less than the old machine, and the government was measuring their production *by kilogram of machine*. So it was easier for them to just continue making the old less powerful machine. The tire factory was allocated another machine that could only make 100,000 tires a year and was back in the same quandary they’d started with.

It’s easy to see how all of these problems could have been solved (or would never have come up) in a capitalist economy, with its use of prices set by supply and demand as an allocation mechanism. And it’s easy to see how thoroughly the Soviet economy was sabotaging itself by avoiding such prices.

III.

The “hero” of *Red Plenty* – although most of the vignettes didn’t involve him directly – was Leonid Kantorovich, a Soviet mathematician who thought he could solve the problem. He invented the technique of linear programming, a method of solving optimization problems perfectly suited to allocating resources throughout an economy. He immediately realized its potential and wrote a nice letter to Stalin politely suggesting his current method of doing economics was wrong and he could do better – this during a time when everyone else in Russia was desperately trying to avoid having Stalin notice them because he tended to kill anyone he noticed. Luckily the letter was intercepted by a kindly mid-level official, who kept it away from Stalin and warehoused Kantorovich in a university somewhere.

During the “Khruschev thaw”, Kantorovich started getting some more politically adept followers, the higher-ups started taking note, and there was a real movement to get his ideas implemented. A few industries were run on Kantorovichian principles as a test case and seemed to do pretty well. There was an inevitable backlash. Opponents accused the linear programmers of being capitalists-in-disguise, which wasn’t helped by their use of something called “shadow prices”. But the combination of their own political adeptness and some high-level support from Khruschev – who alone of all the Soviet leaders seemed to really believe in his own cause and be a pretty okay guy – put them within arm’s reach of getting their plans implemented.

But when elements of linear programming were adopted, they were adopted piecemeal and toothless. The book places the blame on Alexei Kosygen, who implemented a bunch of economic reforms that failed, in a chapter that makes it clear exactly how constrained the Soviet leadership really was. You hear about Stalin, you imagine these guys having total power, but in reality they walked a narrow line, and all these “shadow prices” required more political capital than they were willing to mobilize, even when they thought Kantorovich might have a point.

IV.

In the end, I was left with two contradictory impressions from the book.

First, amazement that the Soviet economy got as far as it did, given how incredibly screwed up it was. You hear about how many stupid things were going on at every level, and you think: *This was the country that built Sputnik and Mir? This was the country that almost buried us beneath the tide of history?* It is a credit to the Russian people that they were able to build so much as a screwdriver in such conditions, let alone a space station.

But second, a sense of what could have been. What if Stalin *hadn't* murdered most of the competent people? What if entire fields of science *hadn't* been banned for silly reasons? What if Kantorovich *had* been able to make the Soviet leadership base its economic planning around linear programming? How might history have turned out differently?

One of the book's most frequently-hammered-in points was that there was was a brief moment, back during the 1950s, when everything seemed to be going right for Russia. Its year-on-year GDP growth (as estimated by impartial outside observers) was somewhere between 7 to 10%. Starvation was going down. Luxuries were going up. Kantorovich was fixing entire industries with his linear programming methods. Then Khruschev made a serious of crazy loose cannon decisions, he was ousted by Brezhnev, Kantorovich was pushed aside and ignored, the "Khruschev thaw" was reversed and tightened up again, and everything stagnated for the next twenty years.

If Khruschev had stuck around, if Kantorovich had succeeded, might the common knowledge that Communism is terrible at producing material prosperity look a little different?

The book very briefly mentioned a competing theory of resource allocation promoted by Victor Glushkov, a cyberneticist in Ukraine. He thought he could use computers – then a very new technology – to calculate optimal allocation for everyone. He failed to navigate the political seas as adroitly as Kantorovich's faction, and the killing blow was a paper that pointed out that for him to do everything *really* correctly would take a hundred million years of computing time.

That was in 1960. If computing power doubles every two years, we've undergone about 25 doubling times since then, suggesting that we ought to be able to perform Glushkov's calculations in three years – or three days, if we give him a lab of three hundred sixty five computers to work with. There could have been this entire field of centralized economic planning. Maybe it would have continued to underperform prices. Or maybe after decades of trial and error across the entire Soviet Union, it could have caught up. We'll never know. Glushkov and Kantorovich were marginalized and left to play around with toy problems until their deaths in the 80s, and as far as I know their ideas were never developed further in the context of a national planned economy.

V.

One of the ways people like insulting smart people, or rational people, or scientists, is by telling them they're the type of people who are attracted to Communism. "Oh, you think you can control and understand everything, just like the Communists did."

And I had always thought this was a pretty awful insult. The people I know who most identify as rationalists, or scientifically/technically minded, are also most likely to be libertarian. So there, case dismissed, everybody go home.

This book was the first time that I, as a person who considers himself rationally/technically minded, realized that I was super attracted to Communism.

Here were people who had a clear view of the problems of human civilization – all the greed, all the waste, all the zero-sum games. Who had the entire population united around a vision of a better future, whose backers could direct the entire state to better serve the goal. All they needed was to solve the engineering challenges, to solve the equations, and there they were, at the golden future. And they were smart enough to be worthy of the problem – Glushkov invented cybernetics, Kantorovich won a Nobel Prize in Economics.

And in the end, they never got the chance. There’s an interpretation of Communism as a refutation of social science, here were these people who probably knew some social science, but did it help them run a state, no it didn’t. But from the little I learned about Soviet history from this book, this seems diametrically wrong. The Soviets had practically no social science. They hated social science. You would think they would at least have some good Marxists, but apparently Stalin killed all of them just in case they might come up with versions of Marxism he didn’t like, and in terms of a vibrant scholarly field it never recovered. Economics was tainted with its association with capitalism from the very beginning, and when it happened at all it was done by non-professionals. Kantorovich was a mathematician by training; Glushkov a computer scientist.

Soviet Communism isn’t what happens when you let nerds run a country, it’s what happens when you kill all the nerds who are experts in country-running, bring in nerds from unrelated fields to replace them, then make nice noises at those nerds in principle while completely ignoring them in practice. Also, you ban all Jews from positions of importance, because fuck you.

> Baggy two-piece suits are not the obvious costume for philosopher kings: but that, in theory, was what the apparatchiks who rule the Soviet Union in the 1960s were supposed to be. Lenin’s state made the same bet that Plato had twenty-five centuries earlier, when he proposed that enlightened intelligence gives absolute powers would serve the public good better than the grubby politicking of republics.
>
> On paper, the USSR was a republic, a grand multi-ethnic federation of republics indeed and its constitutions (there were several) guaranteed its citizens all manner of civil rights. But in truth the Soviet system was utterly unsympathetic to the idea of rights, if you meant by them any suggestion that the two hundred million men, women and children who inhabited the Soviet Union should be autonomously fixing on two hundred million separate directions in which to pursue happiness. This was a society with just one programme for happiness, which had been declared to be scientific and therefore was as factual as gravity.
>
> But the Soviet experiment had run into exactly the difficulty that Plato’s admirers encountered, back in the fifth century BC, when they attempted to mould philosophical monarchies for Syracuse and Macedonia. The recipe called for rule by heavily-armed virtue—or in the Leninist case, not exactly virtue, but a sort of intentionally post-ethical counterpart to it, self-righteously brutal. Wisdom was to be set where it could

> be ruthless. Once such a system existed, though, the qualities required to rise in it had much more to do with ruthlessness than wisdom. Lenin’s core of Bolsheviks, and the socialists like Trotsky who joined them, were many of them highly educated people, literate in multiple European languages, learned in the scholastic traditions of Marxism; and they preserved these attributes even as they murdered and lied and tortured and terrorized. They were social scientists who thought principle required them to behave like gangsters. But their successors – the vydvizhentsy who refilled the CEntral Committee in the thirties – were not the most selfless people in Soviet society, or the most principled, or the most scrupulous. They were the most ambitious, the most domineering, the most manipulative, the most greedy, the most sycophantic: people whose adherence to Bolshevik ideas was inseparable from the power that came with them. Gradually their loyalty to the ideas became more and more instrumental, more and more a matter of what the ideas would let them grip in their two hands…
>
> Stalin had been a gangster who really believed he was a social scientist. Khruschev was a gangster who hoped he was a social scientist. But the moment was drawing irresistibly closer when the idealism would rot away by one more degree, and the Soviet Union would be governed by gangsters who were only pretending to be social scientists.

And in the end it all failed miserably:

> The Soviet economy did not move on from coal and steel and cement to plastics and microelectronics and software design, except in a very few military applications. It continued to compete with what

> capitalism had been doing in the 1930s, not with what it was doing now. It continued to suck resources and human labour in vast quantities into a heavy-industrial sector which had once been intended to exist as a springboard for something else, but which by now had become its own justification. Soviet industry in its last decades existed because it existed, an empire of inertia expanding ever more slowly, yet attaining the wretched distinction of absorbing more of the total effort of the economy that hosted it than heavy industry has ever done anywhere else in human history, before or since. Every year it produced goods that less and less corresponded to human needs, and whatever it once started producing, it tended to go on producing ad infinitum, since it possessed no effective stop signals except ruthless commands from above, and the people at the top no longer did ruthless, in the economic sphere. The control system for industry grew more and more erratic, the information flowing back to the planners grew more and more corrupt. And the activity of industry , all that human time and machine time it used up, added less and less value to the raw materials it sucked in. Maybe no value. Maybe less than none. One economist has argued that, by the end, it was actively destroying value; it had become a system for spoiling perfectly good materials by turning them into objects no one wanted.

I don’t know if this paragraph was intentionally written to contrast with the paragraph at the top, the one about the zombie dance of capitalism. But it is certainly instructive to make such a contrast. The Soviets had originally been inspired by this fear of economics going out of control, abandoning the human beings whose lives it was supposed to improve. In capitalist countries, people existed for the sake of the economy, but under Soviet communism, the economy was going to exist only for the sake of the people.

(accidental Russian reversal: the best kind of Russian reversal!)

And instead, they ended up taking “people existing for the sake of the economy” to entirely new and tragic extremes, people being sent to the gulags or killed because they didn’t meet the targets for some product nobody wanted that was listed on a Five-Year Plan. Spoiling good raw materials for the sake of being able to tell Party bosses and the world “Look at us! We are doing Industry!” Moloch had done some weird judo move on the Soviets’ attempt to destroy him, and he had ended up stronger than ever.

The book’s greatest flaw is that it never did get into the details of the math – or even more than a few-sentence summary of the math – and so I was left confused as to whether anything else had been possible, whether Kantorovich and Glushkov really could have saved the vision of prosperity if they’d been allowed to do so. Nevertheless, the Soviets earned my sympathy and respect in a way Marx so far has not, merely by acknowledging that the problem existed and through the existence of a few good people who tried their best to solve it.

Streetlight Psychology

Recent article on Vox: Do Violent Video Games Actually Make People More Violent? It's a good and even-handed article which concludes:

> The short answer [is] we don't really know. Some studies have indeed found that, in lab settings, people can become more aggressive after playing violent games. Some have also found that people who play violent games are more likely to commit violent acts in real life. But it'd be just as easy to conclude that inherently violent people are simply drawn to violent games — and indeed, several studies have come to that conclusion. Moreover, not all observational studies have found any correlation between gaming and violence. Perhaps most importantly, there's been no surge in violence among youth over the past few decades in gaming countries to accompany the rising popularity of violent games. On the whole, the evidence is decidedly mixed.

I would, however, like to make one important point I think they left out, which also happens to be a very broad criticism of entire subfields of psychology.

They say that "lab experiments have shown violent games can increase aggression", citing a meta-analysis of ninety-eight studies but using as their specific example a recent Italian study. In this experiment, a bunch of high-schoolers were randomly assigned to play Grand Theft Auto or some less violent game. Then they were asked to participate in an experiment where they got to blast loud noises at a partner, which seems like a pretty violent thing to do. The kids who had played Grand Theft Auto were more likely to give their partner louder, more violent blasts. A nice, elegant result very typical of results in this field.

Vox's analysis:

> There are a few caveats to keep in mind when considering these studies. One is that they use aggression and self-control as proxies for real-world violence, because researchers can't actually allow violence to occur in a lab. The idea is that, in increasing aggressiveness, these games make it more likely that someone considering violence will be pushed over the edge and actually commit it. But there's a huge difference between blasting someone with a loud noise, or scoring higher on a questionnaire intended to measure aggressiveness, and actually resorting to violence in the real world.

I think this is insufficiently skeptical.

Suppose I have a different hypothesis. I hypothesize that sad video games cause depression. It seems prima facie plausible. We're having a lot of trouble with child and adolescent depression nowadays. And do you remember that thing that happened in Final Fantasy VII, with that one character (no spoilers)? *Really* sad. Maybe there are lots of video games like that, where you see your favorite characters or people you really like die or suffer terrible travails. In the past, maybe only a couple of people would lose people close to them, but through the vicarious experience of video games we've all experienced seeing our friends "die" dozens of times by the time we graduate high school. Hence, adolescent depression. Right?

We propose an experiment to test this. Fifty Italian teenagers play the appropriate scene in Final Fantasy VII; a control group of another fifty play something neutral like Tetris or whatever. Afterwards, we measure sadness in some way. For example, maybe we make them watch a comedy program and record how many times they laugh; if they don't laugh much, we assume they're sad. Or maybe we just give them a self-rating questionnaire. Whatever.

I predict that this study will *easily* find that the intervention group is sadder than the control group. I also predict that the original hypothesis, that sad video games cause depression, remains stupid.

The difference seems to be that between a temporary state of sad mood, and a long-term alteration of personality that amplifies a sadness "trait" (or more specifically depression, which sometimes presents similar to sadness but may have totally different causes). While it's easy to imagine that sad video games affect the first, it's much harder to believe they can do the second, and act in an additive way to make you a very sad person ten years down the line.

But the same should apply to the violence studies! Yeah, okay, you're priming violence, you've got people in a high-arousal, aggressive state, they'll probably be more violent until the effect wears off. But there's a long way between that and saying that their personality becomes more and more violent until they shoot up a school. Especially since, as far as I know, no mass murderer is known to have played video games *immediately* before their crime, the period in which these kinds of effects might be relevant.

There is an old story about a drunk who loses his keys:

> A policeman sees a drunk man searching for something under a streetlight and asks what the drunk has lost. He says he lost his keys and they both look under the

> streetlight together. After a few minutes the policeman asks if he is sure he lost them here, and the drunk replies, no, that he lost them in the park. The policeman asks why he is searching here, and the drunk replies, “This is where the light is.”

This has since been dubbed the “streetlight effect”, and these studies seem to be an example.

The policeman asks the psychologists what they’re doing. The psychologists say “Studying propensity to violence in the few minutes after someone plays a video game.” The policeman asks “And this is going to reveal important principles about why people are violent in real life?” The psychologists say “No, but it’s really easy to study.”

I want to stop for a second and say that this is not an argument that the field of videogamology is hopelessly flawed or that videogames cannot cause violence. Most of the videogamologists I have read freely acknowledge this problem. That they continue to do these kinds of studies time after time doesn’t surprise me – I’m working on a study right now for work which I don’t think will be hugely contributory to the Sum Of Human Knowledge, because I’m supposed to do something and this was a moderately interesting and do-able thing that came up. A society that rewards researchers per study, like one that rewards factories per kilogram of machinery, will get what it pays for. The videogamologists have taken this in stride and along with these sorts of laboratory studies are doing their best to do longitudinal studies adjusted for various confounders some of which reveal the same pattern. I haven’t studied the field and I have no opinion on it beyond what’s in the Vox article. I just want to make sure people don’t pay too much attention to the streetlight studies.

Which, by the way, are a problem *far* beyond videogamology.

Here's an article called Arabs as terrorists: Effects of stereotypes within violent contexts on attitudes, perceptions, and affect. Participants consume some media with depictions of terrorist Arabs in it. Then they are asked what they think of Arabs. They are more likely than the control group to say they think of Arabs as terrorists. Therefore, media with stereotypical portrayals causes people to stereotype groups in various ways.

Okay. How long did the researchers give between the media in question and the survey? Five minutes? Ten? And we're supposed to believe that this gives us important information about whether, if one of these people throws a brick at a mosque in ten years, it's going to be because of all of the anti-Arabic media he's been gradually consuming up until that time?

In fact, let me make a stronger claim. The "success" of this experiment actually disproves its hypothesis. Assume that everyone in both the experimental and control groups have probably seen hundreds of depictions of Arabs as terrorists throughout their lives. What the experiment finds is that one *recent* depiction of terrorists is enough to overwhelm all of those hundreds of past depictions, such that a clear and obvious difference shows up between the control group (who have seen hundreds of media depictions of terrorist Arabs throughout their entire lives) and the experimental group (who have seen hundreds of media depictions of terrorist Arabs throughout their entire lives, plus one extra one five minutes ago). What we're *actually finding* is that any media depiction effect decays really really rapidly, so much so that people who have been told their entire lives that Arabs are terrorists can be used as the *control group* in an Arabs/terrorism experiment without anyone batting an eye.

Of course, it’s possible that there are two different effects – a very weak long-term effect that builds up with many exposures over time and existed in all of the study participants, and a very strong short-term effect you get when you just saw a documentary on Osama five minutes before someone asked your opinion. But in that case, you shouldn’t go proving the existence of the second effect and then pretending you’ve learned anything at all about the first.

A big chunk of the science in the area of prejudice and racism is exactly like these Arab studies, unfortunately.

Let me bring up one more area in which I think this effect distorts research.

Judith Harris’ book *The Nurture Assumption* makes the bold claim that most differences in how parents bring up their children don’t really change the children’s long-term outcomes that much. This obviously contradicts a wide body of research, and she goes at length into why she thinks the wide body of research is wrong.

I don’t have the book here, so I’m working from memory and it’s going to be more a general form than a specific example, but – suppose that some scientists find mothers who yell at their kids, and mothers who don’t yell at their kids, and find that the children of the yellers act out more and cause more trouble. You can conclude – and a lot of psychologists did – that yelling changes kids’ personalities and makes them more aggressive or rebellious or have some other trait desire to act out.

But other studies show that when the kids are in school, or after they’re taken away by social services, or twenty years later when they’re grown up, there’s no difference between the two groups. Why? Harris suggests that what we’re actually seeing is children responding to a dynamic. The presence of yelly parents makes kids act out, in the same way that the presence of cold makes kids put on a sweater. Cold doesn’t permanently change children’s thermoregulation and cause them to be more likely to wear a sweater twenty years later, it just causes an immediate reaction. Likewise, yelling doesn’t change children’s traits and characteristics and cause them to be more rebellious twenty years later, it just causes an immediate reaction. Remove stimulus, remove response. And this is all my vague half-remembered analogy to something I *think* was in the book, but I think it’s an important distinction.

Situations can cause immediate responses to those situations, or long-term changes in the way people think. You can’t just demonstrate the former and assume you’ve proven the latter.

Why No Science Of Nerds?

Different groups navel-gaze in different ways. Broadway musical writers write a bunch of musicals about what it’s like to be on Broadway. Poets write a bunch of poems about writing poetry. Philosophers speculate on how philosophy may be the most truly virtuous activity. Psychoanalysts analyze the *heck* out of the inner mental experience of psychoanalyzing someone.

All this leaves me a little surprised that there isn’t more scientific study of nerds.

And yet there is not. Typing “nerd” into Google Scholar brings up only a series of papers on desert plants by one Dr. A. Nerd, who must have had a very unpleasant childhood. The field remains strangely unexplored.

“Nerds” seem to share a bunch of seemingly uncorrelated characteristics. They’re generally smart. They’re interested in things like math and science, especially the hard sciences like physics. They’re shy and awkward. They’re some combination of bad at getting social status and not interested in getting social status. They’re especially bad at getting other people to show romantic interest in them. They’re physically unimposing and bad at sports. They don’t get in physical fights and are very unlikely to solve problems with violence. They’re straightedge and less likely to drink or smoke to excess (according to legend, “nerd” derives from “knurd”, ie “drunk” spelled backwards). Sometimes even very specific physical characteristics make the list, like a silly-sounding high-pitched voice.

A scientific study of nerds might begin by asking: why do all of these things go together in the popular imagination, form a single category?

Of these nine classic characteristics, we can imagine people scoring either “nerdy” or “anti-nerdy” on each. If that were true, there’s only a 1/2^9 = 1/512 chance that any given person has all of these characteristics. Our null hypothesis might be that “nerd” is just a made-up category used to describe this totally coincidental group of 1/512th of the population, or those people who are sufficiently close (maybe 6 or 7 out of 9?). Could be.

The other possibility is that in fact these traits are all correlated for some reason, and people who are bad at sports really *are* more likely to enjoy math, less likely to drink or use violence, et cetera. “Nerd” would then be a natural category, in the same way that, for example, “bird” is a natural category pointing out that animals with feathers are more likely to have wings, beaks, et cetera rather than a totally random distribution of traits. Why would that be?

Have there been nerds across different times and cultures? The term was only coined in the mid-20th century. But Isaac Newton seems to have been a nerd. Whatever else he was, Henry Cavendish was at the very least a nerd. On the other hand, some of the Europeans I’ve talked to say that the experience of nerdiness on their side of the Atlantic is very different from the American experience, so much that it’s hard to interpret them as having a “nerd” concept at all. From my time in Japan, the experience there is different as well.

One can sort of imagine how certain of the correlated characteristics might cause others. Young people who are small and weak and bad at sports might lose social status as a result. People who are good at math might be so transfixed by the mysteries of the universe that things like sports and socializing and dating pale in comparison. People who are very smart might, for the usual neurological reasons, also have high impulse control, explaining the lack of drug use and aggression. People who are bad at social skills might not get invited to sports games and so have no opportunity to improve; they might take up unpopular but solitary pursuits like math as a result.

Or we could take the fun route and go full biodeterminist. The connection to autism and Asperger syndrome is so commonly cited (despite a lack of any real scholarly investigation) as to be cliched. Anyone familiar with the full-fledged syndrome understands it's something very different from everyday nerdiness, but the possibility that nerdiness is some very mild form or related condition probably shouldn't be ruled out.

Or we could go a different route. Consider:

In men (but not women) low testosterone is related to increased mathematical intelligence and increased likelihood of being gifted.

Testosterone is associated with extraversion, alone of Big Five characteristics.

Higher testosterone men have higher social status, and the cognitive role of testosterone has been described as "best understood in terms of the search for and maintenance of social status".

As anyone who has watched the controversy over steroid use in professional sports knows, testosterone improves muscle mass and athletic performance.

Men with lower testosterone have lower mating success.

High testosterone is correlated with increased drinking behavior with $p < .001$.

Especially in prison populations, high testosterone is closely linked to violence and aggressive behavior.

Men with lower testosterone have higher-pitched voices (the media bills this as Deep Voiced Men Make Bad Mates: Study).

So at least in men, low testosterone seems to cause most of the characteristics associated with nerdhood.

But I'm being unfair. There's a lot of counterevidence as well.

There are so many conflicting studies on testosterone and intelligence that I despair of getting anything coherent. Many studies show increased testosterone *increases* intelligence, especially in tasks where men generally outperform women, like spatial rotation (which tends to correlate with math) – for example, see here. I respect this research, but I would naively expect, let's say, brilliant mathematicians to have lower testosterone than the general population, just based on my stereotypes that testosterone is associated with aggressiveness, popularity, athleticism, etc. I'm not sure how this squares with the data.

Testosterone doesn't actually make male faces more attractive, according to Testosterone increases perceived dominance but not attractiveness in human males.

Nerds are traditionally viewed as having high libido – think watching pornography. But of course high testosterone is associated with higher libido. I don't know if there's a cultural thing going on where nerds have normal-to-low libido but are stereotyped as having high libido to make fun of their lack of romantic success – or even whether the pornography connection is just that nerds are better with computers. I also note with interest that testosterone is said to affect sexual libido but not desire for "sensual touch", and a *lot* of people have mentioned how anomalously some of the nerd communities I'm in tend to value cuddling compared to sex relative to the general population.

I could probably find other traditionally nerdy characteristics that correlated negatively with testosterone. For example, acne is associated with higher testosterone levels.

So I don't think this is the whole picture.

But I still feel like it should be *some* of the picture. Sex hormones are *really* complicated – for example, there seem to be different effects from in utero sex hormone exposure compared to pre-pubertal sex hormone exposure compared to pubertal sex hormone exposure, and these aren't necessarily correlated with each other in the same people. Estrogens can have effects ranging from very similar to testosterone to exactly the opposite. Testosterone is hard to get a good measurement from, especially with the no-fuss salivary measurements I bet most of these studies used, and its effects or lack thereof would depend on lots of stuff like how much of it gets converted to dihydrotestosterone. Also, some hormones have totally different effects if they come in short spikes versus constant gradual release. So there's a lot of room to improve our understanding of sex hormones and start distinguishing between unlike constructs.

I am reminded of an observation common among transsexuals – and brought up in Ozy's last post – that there is a distinct cluster of transwomen who have certain very traditionally-considered-male-gendered characteristics and are very nerdy. This seems like another example of some strongly male and strongly female characteristics anomalously going hand in hand.

And what about female nerds? If we’re trying to make this about testosterone, it sounds like they need a separate explanation. Is female nerd a distinct cluster in the same way male nerd is? I don’t know. I don’t, for example, have strong feelings about whether female nerds are more or less attractive/athletic/whatever compared to non-nerdy females.

Mostly I just feel like this field is strangely under-explored. Especially when you consider what the sorts of people who explore fields are like.

Cuddle Culture

[Content warning: TMI, polyamory.]

I.

Another one of those times three very different people writing three very different things all remind me of each other.

Ozy got very excited recently because Heartiste wrote a post attacking polyamory (Ozy reminds me that the appropriate trigger warning for Heartiste is “trigger warning: literally the worst person alive, I am so serious about this, you think I am joking but I am not”).

Reversed stupidity is not intelligence, but it’s still nice to know that somebody known to be generally evil takes time out of his busy day to dislike my way of life specifically. It’s like a weird sort of reverse validation.

But since the Devil sometimes speaks true, what *exactly* does he have to say?

> Genuine, egalitarian, open polyamory for all practical purposes doesn’t exist among white Westerners. There’s always one or another party out in the asexual or anhedonic cold, nursing feelings of rejection and traumatic self-doubt. And if that party is a willing participant to his or her sexual/romantic exclusion, it’s a good bet he/she is psychologically broken, mentally unstable, physically repulsive, or suffering from clinically low sex drive. In other words, human trash.

Applying enough charity to fully fund the Red Cross for the next fifty years, Heartiste seems to be saying something along the lines of “Polyamory is especially well-suited for asexual people”. And I agree!

Many of the people I know in successful polyamorous relationships are sexual, sometimes even highly sexual. But I also know a disproportionate number of asexual polyamorous people – including myself – and the combination seems to work really, really well. Part of it is the ability for asexual people to date sexual people without having to worry about the partner having no way of satisfying their higher sex drive. Part of it is the free layer of protection against sexual jealousy. And part of it is the neat ability to sidestep most of the risks of polyamory, including infection, unintended pregnancy, and the sense of disgust that some sexual people – especially Heartiste – seem to feel at the thought of having sex with less-than-virginal partners.

For me polyamory doesn’t get into any of that. It just means lots and lots of free cuddles.

II.

Which brings me to the second thing I read recently. There is a new app out, Cuddlr, which is “like Grindr, but for cuddling”. Unequally Yoked has come out against it, saying that cuddling people without knowing them first is “objectifying”.

You already know what I think of objectification, but the criticism is unusually jarring in this instance. For me, cuddling is the opposite of objectifying. I go into social encounters viewing most people as a combination of scary and boring. I can sometimes overcome that most of the way by spending months getting to know them and appreciate their unique perspective. Or I can cuddle with them for ten minutes. Either one works.

There’s a Graham Greene quote which, being a philistine, I only know because it was included in Robert Wright’s *The Moral Animal*:

> He took another drink of brandy. As the liquid touched his tongue he remembered his child, coming in out of the glare: the sullen unhappy knowledgeable face. He said, “Oh God, help her. Damn me, I deserve it, but let her live for ever.” This was the love he should have felt for every soul in the world: all the fear and the wish to save concentrated unjustly on the one child. He began to weep; it was as if he had to watch her from the shore drown slowly because he had forgotten how to swim. He thought: This is what I should feel all the time for everyone…

Wright’s point was that, there is this pure universal love that we wish we could feel for everyone all the time, but in practice we’re only able to feel it for our children, presumably because of evolutionary imperatives. As for me, I have no children, but the pure universal love I wish I could feel for everyone all the time, I’m only *actually* able to feel for cute girls I am cuddling with. It is definitely a good, correct kind of love – Leah would be more likely to call it *agape* or *philia* than *eros*. And this is important to me, because that kind of love is definitely an important psychological nutrient and my brain is very bad at feeling it any other way without, like, knowing somebody for ten years.

So this is the second reason why I think polyamory and (my particular variant of) asexuality go well together. It allows me to cuddle whoever I want and fall in love with whoever I want and have absurdly fond and protective feelings toward *everybody* if I so choose.

III.

The third thing that made me think of this was actually something I wrote in my post yesterday and realized I should expand upon:

> Testosterone is said to affect sexual libido but not desire for "sensual touch", and a lot of people have mentioned how anomalously some of the nerd communities I'm in tend to value cuddling compared to sex relative to the general population.

In the general population – let alone for people like Heartiste – men are supposed to consider cuddling to be that extremely annoying thing that women sometimes want to do instead of sex, and which they must be very careful to avoid lest women get the impression that this is acceptable.

On the other hand, in the nerdy, polyamorous communities I've been in, it's been generally understood that people of all sorts, man or woman or Ozy, can like cuddling and there is no shame in it.

This has been really liberating. Like, if you ask someone if they want to have sex, they might say no, they might slap you, but at least they will understand the context: that is definitely a known thing people ask. If you ask someone to cuddle, they will usually just be very confused, which in a way makes it even creepier.

The formation of communities where it's *not* creepy and you can just *ask* is, at least to this asexual, one of the more important pieces of social technology to come out of the weird incubator that is the Bay Area. It creates so many positive feelings and so much of the good kind of groupishness that it seems like a comical Publishers' Clearing House-style $100 bill left on the ground in the relatively high-stakes Forming Cohesive Communities Game.

I am left speculating that it only works after you get a certain percent asexual, or a certain percent polyamorous, or a certain percent low testosterone, or a certain percent low jealousy. Or maybe that you have to have a certain amount of community cohesion before you try. Or maybe you need people with a certain amount of willingness to experiment and not take themselves seriously. I don't know. I can certainly imagine most attempts to initiate it would implode horribly. I certainly wouldn't want to be the one who tries to import cuddle culture to some other group where social cohesion is important, like the US Senate.

It just seems to be one of those really nice equilibria that form spontaneously in certain places for reasons that are difficult to pinpoint, just like the rest of civilization.

Bottomless Pits Of Suffering

I.

A friend on Facebook recently posted the following dilemma, which of course I cannot find right now so I have to vaguely quote my recollection of it:

> Would you rather the medieval Church had spent all of its money helping the poor, rather than supporting the arts? So that maybe there were fewer poor people back in medieval times, but we wouldn't have any cathedrals or triptychs or the Sistine Chapel?

I was surprised to see so many people choosing the cathedrals. I think maybe choosing the cathedrals is so appealing because they're right there, you can touch them, but the starving peasants are hidden all the way in the past where you can't see them. So it feels like you're being asked to sacrifice something you really like for something that you would otherwise not have to think about.

This is one of the biggest and scariest problems with utilitarianism. Utiltarianism is at least kind of easy when it's asking you to trade off some things in your normal world for other things in your normal world. But when it asks you to make everything you consider your normal world unambiguously worse to help some other domain you would otherwise never have to think about, then it starts to become unintuitive and scary.

Imagine a happy town full of prosperous people. Every so often they make nice utilitarian decisions like having everyone chip in a few dollars to help someone who's fallen sick, and they feel pretty good about themselves for this.

Then one day an explorer discovers a BOTTOMLESS PIT OF ENDLESS SUFFERING on the outskirts of town. There are hundreds of people trapped inside in a state of abject misery. The Pit Gods agree to release some of their prisoners, but only for appropriately sumptuous sacrifices.

Suddenly the decision isn't just "someone in town makes a small sacrifice to help other people in town". Suddenly it's about the entire town choking off its luxury and prosperity in order to rescue people they don't even know, from this pit they didn't even know was there a week ago. That seems kind of unfair.

So they tell the explorer to cover the lid of the pit with a big tarp that blends in with the surrounding grass, so they don't have to see it, and then go on with their lives.

II.

The developing world is *sort of* a bottomless pit of suffering if some First Worlder didn't expect it to be there. But I think most people do expect it to be there, most people are happy to help (a little), and it doesn't really confuse or alarm us too much when we are reminded they still exist and still need help.

But what about nursing homes? Most of the doctors I have talked to agree most nursing homes are terrible. I get a steady trickle of psychiatric patients who are perfectly happy to be in the psychiatric hospital but who *freak out* when I tell them that they seem all better now and it's time to send them back to their nursing home, saying it's terrible and they're abused and neglected and they refuse to go. I very occasionally get elderly patients who have attempted suicide solely because they know doing so will get them out of their nursing home. I don't have a strong feeling for *exactly* how bad nursing homes are, but everything I have seen is consistent with at least some of them being very bad.

Solving this would be really expensive – I am perpetually surprised at how quietly and effortlessly we seem to soak up nursing home costs that already can run into the tens of thousands of dollars a year. Solving this would also produce no visible gain, in that bedridden old people are very very bad at complaining in ways anyone else can notice, and if we don't want to think about them we don't have to. If we as a country decided to concentrate on decreasing abuse in nursing homes, we might have to take that money away from important causes in our everyday visible world, like welfare and infrastructure and education funding. We would have to take limited Public Attention And Outrage Resources from causes like human rights and gay marriage and what beverages the President is holding while he salutes people. I think everyone agrees it's a lot easier not to think about it, and nobody can make us.

Prisons are an even uglier case. Not only is prison inherently pretty miserable, but there seems to be rampant abuse and violence going on, including at least 5% of prisoners being raped per year. Every couple of weeks there's a new story about how, for example, prisoners are gouged on phone bills because someone can do it and nobody is stopping them, or how they're kept in cells without air conditioning in 110 degree weather in Arizona because no one has any incentive to change that. If you try to solve it, you'll release prisoners, and at least a few of them are going to commit new crimes which you will have to see and think about. Or you can just whistle, pretend not to notice, and continue to enjoy your nice low-crime society.

III.

A lot of the paradoxes of utilitarianism, the things that make it scary and hard to work with, involve philosophers who compulsively seek out bottomless pits and shout at you until you pay attention to them.

Utility monsters are basically one-man bottomless pits.

Pascal’s Wager (or Pascal’s Mugging, if you prefer) splits the universe into a billion Everett branches, then points out that one of these Everett branches is a bottomless pit and asks the others to make sacrifices to help it.

A lot of the addition paradoxes treat a pool of “potential people” as a bottomless pit.

This seems to be the easiest way to break utilitarianism – point to a bottomless pit, real or imagined, and make everyone in the world lose utility to solve it, forever. It’s not always easy to come up with solutions that successfully rule out these problems, while preserving our intuition that we should continue to worry about people in nursing homes or jails.

Contractualism scares me a little because it offers *too easy* an out from bottomless-pit type dilemmas. It seems really easy to say “All of us people not in jail, we’ll agree to look out for one another, and as for those guys, screw them”. You would need to have something like a veil of ignorance, or at least a good simulation of one, to even begin to care.

I Can Tolerate Anything Except The Outgroup

[Content warning: Politics, religion, social justice, spoilers for "The Secret of Father Brown". This isn't especially original to me and I don't claim anything more than to be explaining and rewording things I have heard from a bunch of other people. Unapologetically America-centric because I'm not informed enough to make it otherwise. Try to keep this off Reddit and other similar sorts of things.]

I.

In Chesterton's *The Secret of Father Brown*, a beloved nobleman who murdered his good-for-nothing brother in a duel thirty years ago returns to his hometown wracked by guilt. All the townspeople want to forgive him immediately, and they mock the titular priest for only being willing to give a measured forgiveness conditional on penance and self-reflection. They lecture the priest on the virtues of charity and compassion.

Later, it comes out that the beloved nobleman did *not* in fact kill his good-for-nothing brother. The good-for-nothing brother killed the beloved nobleman (and stole his identity). *Now* the townspeople want to see him lynched or burned alive, and it is only the priest who – consistently – offers a measured forgiveness conditional on penance and self-reflection.

The priest tells them:

> It seems to me that you only pardon the sins that you don't really think sinful. You only forgive criminals when they commit what you don't regard as crimes, but rather as conventions. You forgive a conventional duel

> just as you forgive a conventional divorce. You forgive because there isn't anything to be forgiven.

He further notes that this is why the townspeople can self-righteously consider themselves more compassionate and forgiving than he is. Actual forgiveness, the kind the priest needs to cultivate to forgive evildoers, is really really hard. The fake forgiveness the townspeople use to forgive the people they like is really easy, so they get to boast not only of their forgiving nature, but of how much nicer they are than those mean old priests who find forgiveness difficult and want penance along with it.

After some thought I agree with Chesterton's point. There are a lot of people who say "I forgive you" when they mean "No harm done", and a lot of people who say "That was unforgiveable" when they mean "That was genuinely really bad". Whether or not forgiveness is *right* is a complicated topic I do not want to get in here. But since forgiveness is generally considered a virtue, and one that many want credit for having, I think it's fair to say you only earn the right to call yourself 'forgiving' if you forgive things that genuinely hurt you.

To borrow Chesterton's example, if you think divorce is a-ok, then you don't get to "forgive" people their divorces, you merely ignore them. Someone who thinks divorce is abhorrent can "forgive" divorce. *You* can forgive theft, or murder, or tax evasion, or something *you* find abhorrent.

I mean, from a utilitarian point of view, you are still doing the correct action of not giving people grief because they're a divorcee. You can have all the Utility Points you want. All I'm saying is that if you "forgive" something you don't care about, you don't earn any Virtue Points.

(by way of illustration: a billionaire who gives $100 to charity gets as many Utility Points as an impoverished pensioner who donates the same amount, but the latter gets a lot more Virtue Points)

Tolerance is also considered a virtue, but it suffers the same sort of dimished expectations forgiveness does.

The Emperor summons before him Bodhidharma and asks: “Master, I have been tolerant of innumerable gays, lesbians, bisexuals, asexuals, blacks, Hispanics, Asians, transgender people, and Jews. How many Virtue Points have I earned for my meritorious deeds?”

Bodhidharma answers: “None at all”.

The Emperor, somewhat put out, demands to know why.

Bodhidharma asks: “Well, what do you think of gay people?”

The Emperor answers: “What do you think I am, some kind of homophobic bigot? Of course I have nothing against gay people!”

And Bodhidharma answers: “Thus do you gain no merit by tolerating them!”

II.

If I had to define “tolerance” it would be something like “respect and kindness toward members of an outgroup”.

And today we have an almost unprecedented situation.

We have a lot of people – like the Emperor – boasting of being able to tolerate everyone from every outgroup they can imagine, loving the outgroup, writing long paeans to how great the outgroup is, staying up at night fretting that somebody else might not like the outgroup enough.

This is really surprising. It’s a total reversal of everything we know about human psychology up to this point. No one did any genetic engineering. No one passed out weird glowing pills in the public schools. And yet suddenly we get an entire group of people who conspicuously promote and defend their outgroups, the outer the better.

What is going on here?

Let’s start by asking what exactly an outgroup is.

There’s a very boring sense in which, assuming the Emperor’s straight, gays are part of his “outgroup” ie a group that he is not a member of. But if the Emperor has curly hair, are straight-haired people part of his outgroup? If the Emperor’s name starts with the letter ‘A’, are people whose names start with the letter ‘B’ part of his outgroup?

Nah. I would differentiate between multiple different meanings of outgroup, where one is “a group you are not a part of” and the other is…something stronger.

I want to avoid a very easy trap, which is saying that outgroups are about how different you are, or how hostile you are. I don’t think that’s quite right.

Compare the Nazis to the German Jews and to the Japanese. The Nazis were very similar to the German Jews: they looked the same, spoke the same language, came from a similar culture. The Nazis were totally different from the Japanese: different race, different language, vast cultural gap. But the Nazis and Japanese mostly got along pretty well. Heck, the Nazis were actually moderately positively disposed to the *Chinese*, even when they were technically at war.

Meanwhile, the conflict between the Nazis and the German Jews – some of whom didn't even realize they were anything other than German until they checked their grandparents' birth certificate – is the stuff of history and nightmares. Any theory of outgroupishness that naively assumes the Nazis' natural outgroup is Japanese or Chinese people will be totally inadequate.

And this isn't a weird exception. Freud spoke of the narcissism of small differences, saying that "it is precisely communities with adjoining territories, and related to each other in other ways as well, who are engaged in constant feuds and ridiculing each other". Nazis and German Jews. Northern Irish Protestants and Northern Irish Catholics. Hutus and Tutsis. South African whites and South African blacks. Israeli Jews and Israeli Arabs. Anyone in the former Yugoslavia and anyone else in the former Yugoslavia.

So what makes an outgroup? Proximity plus small differences. If you want to know who someone in former Yugoslavia hates, don't look at the Indonesians or the Zulus or the Tibetans or anyone else distant and exotic. Find the Yugoslavian ethnicity that lives closely intermingled with them and is most conspicuously similar to them, and chances are you'll find the one who they have eight hundred years of seething hatred toward.

What makes an unexpected in-group? The answer with Germans and Japanese is obvious – a strategic alliance. In fact, the World Wars forged a lot of unexpected temporary pseudo-friendships. A recent article from War Nerd points out that the British, after spending centuries subjugating and despising the Irish and Sikhs, suddenly needed Irish and Sikh soldiers for World Wars I and II respectively. "Crush them beneath our boots" quickly changed to fawning songs about how “there never was a coward where the shamrock grows” and endless paeans to Sikh military prowess.

Sure, scratch the paeans even a little bit and you find condescension as strong as ever. But eight hundred years of the British committing genocide against the Irish and considering them literally subhuman turned into smiles and songs about shamrocks once the Irish started looking like useful cannon fodder for a larger fight. And the Sikhs, dark-skinned people with turbans and beards who pretty much exemplify the European stereotype of “scary foreigner”, were lauded by everyone from the news media all the way up to Winston Churchill.

In other words, outgroups may be the people who look exactly like you, and scary foreigner types can become the in-group on a moment’s notice when it seems convenient.

III.

There are certain theories of dark matter where it barely interacts with the regular world *at all*, such that we could have a dark matter planet exactly co-incident with Earth and never know. Maybe dark matter people are walking all around us and through us, maybe my house is in the Times Square of a great dark matter city, maybe a few meters away from me a dark matter blogger is writing on his dark matter computer about how weird it would be if there was a light matter person he couldn’t see right next to him.

This is sort of how I feel about conservatives.

I don’t mean the sort of light-matter conservatives who go around complaining about Big Government and occasionally voting for Romney. I see those guys all the time. What I mean is – well, take creationists. According to Gallup polls, about 46% of Americans are creationists. Not just in the sense of believing God helped guide evolution. I mean they think evolution is a vile atheist lie and God created humans exactly as they exist right now. That’s half the country.

And I don’t have a *single one of those people* in my social circle. It’s not because I’m deliberately avoiding them; I’m pretty live-and-let-live politically, I wouldn’t ostracize someone just for some weird beliefs. And yet, even though I probably know about a hundred fifty people, I am pretty confident that not one of them is creationist. Odds of this happening by chance? 1/2^150 = 1/10^45 = approximately the chance of picking a particular atom if you are randomly selecting among all the atoms on Earth.

About forty percent of Americans want to ban gay marriage. I think if I *really* stretch it, maybe ten of my top hundred fifty friends might fall into this group. This is less astronomically unlikely; the odds are a mere one to one hundred quintillion against.

People like to talk about social bubbles, but that doesn’t even begin to cover one hundred quintillion. The only metaphor that seems really appropriate is the bizarre dark matter world.

I live in a Republican congressional district in a state with a Republican governor. The conservatives are definitely out there. They drive on the same roads as I do, live in the same neighborhoods. But they might as well be made of dark matter. I never meet them.

To be fair, I spend a lot of my time inside on my computer. I’m browsing sites like Reddit.

Recently, there was a thread on Reddit asking – Redditors Against Gay Marriage, What Is Your Best Supporting Argument? A Reddit user who didn’t understand how anybody could be against gay marriage honestly wanted to know how other people who *were* against it justified their position. He figured he might as well ask one of the largest sites on the Internet, with an estimated user base in the tens of millions.

It soon became clear that nobody there was actually against gay marriage.

There were a bunch of posts saying “I of course support gay marriage but here are some reasons some other people might be against it,” a bunch of others saying “my argument against gay marriage is the government shouldn’t be involved in the marriage business at all”, and several more saying “why would you even ask this question, there’s no possible good argument and you’re wasting your time”. About halfway through the thread someone started saying homosexuality was unnatural and I *thought* they were going to be the first one to actually answer the question, but at the end they added “But it’s not my place to decide what is or isn’t natural, I’m still pro-gay marriage.”

In a thread with 10,401 comments, a thread *specifically* asking for people against gay marriage, I was eventually able to find *two* people who came out and opposed it, way near the bottom. Their posts started with “I know I’m going to be downvoted to hell for this…”

But I’m not only on Reddit. I also hang out on LW.

On last year’s survey, I found that of American LWers who identify with one of the two major political parties, 80% are Democrat and 20% Republican, which actually sounds pretty balanced compared to some of these other examples.

But it doesn’t last. Pretty much all of those “Republicans” are libertarians who consider the GOP the lesser of two evils. When allowed to choose “libertarian” as an alternative, only 4% of visitors continued to identify as conservative. But that’s still…some. Right?

When I broke the numbers down further, 3 percentage points of those are neoreactionaries, a bizarre sect that wants to be ruled by a king. Only *one percent* of LWers were normal everyday God-‘n-guns-but-not-George-III conservatives of the type that seem to make up about half of the United States.

It gets worse. My formative years were spent at a university which, if it was similar to other elite universities, had a faculty and a student body that skewed about 90-10 liberal to conservative – and we can bet that, like LW, even those few token conservatives are Mitt Romney types rather than God-n’-guns types. I get my news from vox.com, an Official Liberal Approved Site. Even when I go out to eat, it turns out my favorite restaurant, California Pizza Kitchen, is the most liberal restaurant in the United States.

I inhabit the same geographical area as *scores and scores* of conservatives. But without meaning to, I have created an *outrageously* strong bubble, a 10^45 bubble. Conservatives are all around me, yet I am about as likely to have a serious encounter with one as I am a Tibetan lama.

(Less likely, actually. One time a Tibetan lama came to my college and gave a really nice presentation, but if a conservative tried that, people would protest and it would be canceled.)

IV.

One day I realized that entirely by accident I was fulfilling *all* the Jewish stereotypes.

I’m nerdy, over-educated, good with words, good with money, weird sense of humor, don’t get outside much, I like deli sandwiches. And I’m a psychiatrist, which is about the most stereotypically Jewish profession short of maybe stand-up comedian or rabbi.

I’m not very religious. And I don’t go to synagogue. But *that’s* stereotypically Jewish too!

I bring this up because it would be a mistake to think “Well, a Jewish person is by definition someone who is born of a Jewish mother. Or I guess it sort of also means someone who follows the Mosaic Law and goes to synagogue. But I don’t care about Scott’s mother, and I know he doesn’t go to synagogue, so I can’t gain any useful information from knowing Scott is Jewish.”

The defining factors of Judaism – Torah-reading, synagogue-following, mother-having – are the tip of a giant iceberg. Jews sometimes identify as a “tribe”, and even if you don’t attend synagogue, you’re still a member of that tribe and people can still (in a statistical way) infer things about you by knowing your Jewish identity – like how likely they are to be psychiatrists.

The last section raised a question – if people rarely select their friends and associates and customers explicitly for politics, how do we end up with such intense political segregation?

Well, in the same way “going to synagogue” is merely the iceberg-tip of a Jewish tribe with many distinguishing characteristics, so “voting Republican” or “identifying as conservative” or “believing in creationism” is the iceberg-tip of a conservative tribe with many distinguishing characteristics.

A disproportionate number of my friends are Jewish, because I meet them at psychiatry conferences or something – we self-segregate not based on explicit religion but on implicit tribal characteristics. So in the same way, political tribes self-segregate to an impressive extent – a 1/10^45 extent, I will never tire of hammering in – based on their implicit tribal characteristics.

The people who are actually into this sort of thing sketch out a bunch of speculative tribes and subtribes, but to make it easier, let me stick with two and a half.

The Red Tribe is most classically typified by conservative political beliefs, strong evangelical religious beliefs, creationism, opposing gay marriage, owning guns, eating steak, drinking Coca-Cola, driving SUVs, watching lots of TV, enjoying American football, getting conspicuously upset about terrorists and commies, marrying early, divorcing early, shouting “USA IS NUMBER ONE!!!”, and listening to country music.

The Blue Tribe is most classically typified by liberal political beliefs, vague agnosticism, supporting gay rights, thinking guns are barbaric, eating arugula, drinking fancy bottled water, driving Priuses, reading lots of books, being highly educated, mocking American football, feeling vaguely like they should like soccer but never really being able to get into it, getting conspicuously upset about sexists and bigots, marrying later, constantly pointing out how much more civilized European countries are than America, and listening to “everything except country”.

(There is a partly-formed attempt to spin off a Grey Tribe typified by libertarian political beliefs, Dawkins-style atheism, vague annoyance that the question of gay rights even comes up, eating paleo, drinking Soylent, calling in rides on Uber, reading lots of blogs, calling American football “sportsball”, getting conspicuously upset about the War on Drugs and the NSA, and listening to filk – but for our current purposes this is a distraction and they can safely be considered part of the Blue Tribe most of the time)

I think these “tribes” will turn out to be even stronger categories than politics. Harvard might skew 80-20 in terms of Democrats vs. Republicans, 90-10 in terms of liberals vs. conservatives, but maybe 99-1 in terms of Blues vs. Reds.

It’s the many, many differences between these tribes that explain the strength of the filter bubble – which *have I mentioned* segregates people at a strength of 1/10^45? Even in something as seemingly politically uncharged as going to California Pizza Kitchen or Sushi House for dinner, I’m restricting myself to the set of people who like cute artisanal pizzas or sophsticated foreign foods, which are classically Blue Tribe characteristics.

Are these tribes based on geography? Are they based on race, ethnic origin, religion, IQ, what TV channels you watched as a kid? I don’t know.

Some of it is certainly genetic – estimates of the genetic contribution to political association range from 0.4 to 0.6. Heritability of one’s attitudes toward gay rights range from 0.3 to 0.5, which hilariously is a little more heritable than homosexuality itself.

(for an interesting attempt to break these down into more rigorous concepts like “traditionalism”, “authoritarianism”, and “in-group favoritism” and find the genetic loading for each see here. For an attempt to trace the specific genes involved, which mostly turn out to be NMDA receptors, see here)

But I don't think it's just genetics. There's something else going on too. The word "class" seems like the closest analogue, but only if you use it in the sophisticated Paul Fussell *Guide Through the American Status System* way instead of the boring "another word for how much money you make" way.

For now we can just accept them as a brute fact – as multiple coexisting societies that might as well be made of dark matter for all of the interaction they have with one another – and move on.

V.

The worst reaction I've ever gotten to a blog post was when I wrote about the death of Osama bin Laden. I've written all sorts of stuff about race and gender and politics and whatever, but that was the worst.

I didn't come out and say I was happy he was dead. But some people interpreted it that way, and there followed a bunch of comments and emails and Facebook messages about how could I possibly be happy about the death of another human being, even if he was a bad person? Everyone, even Osama, is a human being, and we should never rejoice in the death of a fellow man. One commenter came out and said:

> I'm surprised at your reaction. As far as people I casually stalk on the internet (ie, LJ and Facebook), you are the first out of the "intelligent, reasoned and thoughtful" group to be uncomplicatedly happy about this development and not to be, say, disgusted at the reactions of the other 90% or so.

This commenter was right. Of the "intelligent, reasoned, and thoughtful" people I knew, the overwhelming emotion was conspicuous disgust that other people could be happy about his death. I hastily backtracked and said I wasn't happy per se, just surprised and relieved that all of this was finally behind us.

And I genuinely believed that day that I had found some unexpected good in people – that everyone I knew was so humane and compassionate that they were unable to rejoice even in the death of someone who hated them and everything they stood for.

Then a few years later, Margaret Thatcher died. And on my Facebook wall – made of these same "intelligent, reasoned, and thoughtful" people – the most common response was to quote some portion of the song "Ding Dong, The Witch Is Dead". Another popular response was to link the videos of British people spontaneously throwing parties in the street, with comments like "I wish I was there so I could join in". From this exact same group of people, not a single expression of disgust or a "c'mon, guys, we're all human beings here."

I gently pointed this out at the time, and mostly got a bunch of "yeah, so what?", combined with links to an article claiming that "the demand for respectful silence in the wake of a public figure's death is not just misguided but dangerous".

And that was when something clicked for me.

You can talk all you want about Islamophobia, but my friend's "intelligent, reasoned, and thoughtful people" – her name for the Blue Tribe – can't get together enough energy to really hate Osama, let alone Muslims in general. We understand that what he did was bad, but it didn't anger us personally. When he died, we were able to very rationally apply our better nature and our Far Mode beliefs about how it's never right to be happy about anyone else's death.

On the other hand, that same group absolutely *loathed* Thatcher. Most of us (though not all) can agree, if the question is posed explicitly, that Osama was a worse person than Thatcher. But in terms of actual gut feeling? Osama provokes a snap judgment of "flawed human being", Thatcher a snap judgment of "scum".

I started this essay by pointing out that, despite what geographical and cultural distance would suggest, the Nazis' outgroup was not the vastly different Japanese, but the almost-identical German Jews.

And my hypothesis, stated plainly, is that if you're part of the Blue Tribe, then your outgroup isn't al-Qaeda, or Muslims, or blacks, or gays, or transpeople, or Jews, or atheists – it's the Red Tribe.

VI.

"But racism and sexism and cissexism and anti-Semitism are these giant all-encompassing social factors that verge upon being human universals! Surely you're not arguing that mere *political* differences could ever come close to them!"

One of the ways we *know* that racism is a giant all-encompassing social factor is the Implicit Association Test. Psychologists ask subjects to quickly identify whether words or photos are members of certain gerrymandered categories, like "either a white person's face or a positive emotion" or "either a black person's face and a negative emotion". Then they compare to a different set of gerrymandered categories, like "either a black person's face or a positive emotion" or "either a white person's face or a negative emotion.” If subjects have more trouble (as measured in latency time) connecting white people to negative things than they do white people to positive things, then they probably have subconscious positive associations with white people. You can try it yourself here.

Of course, what the test famously found was that even white people who claimed to have no racist attitudes at all usually had positive associations with white people and negative associations with black people on the test. There are very many claims and counterclaims about the precise meaning of this, but it ended up being a big part of the evidence in favor of the current consensus that all white people are at least a little racist.

Anyway, three months ago, someone finally had the bright idea of doing an Implicit Association Test with political parties, and they found that people’s unconscious partisan biases were *half again as strong* as their unconscious racial biases (h/t Bloomberg. For example, if you are a white Democrat, your unconscious bias against blacks (as measured by something called a d-score) is 0.16, but your unconscious bias against Republicans will be 0.23. The Cohen’s *d* for racial bias was 0.61, by the book a “moderate” effect size; for party it was 0.95, a “large” effect size.

Okay, fine, but we know race has *real world* consequences. Like, there have been several studies where people sent out a bunch of identical resumes except sometimes with a black person’s photo and other times with a white person’s photo, and it was noticed that employers were much more likely to invite the fictional white candidates for interviews. So just some stupid Implicit Association Test results can’t compare to that, right?

Iyengar and Westwood also decided to do the resume test for parties. They asked subjects to decide which of several candidates should get a scholarship (subjects were told this was a genuine decision for the university the researchers were affiliated with). Some resumes had photos of black people, others of white people. And some students listed their experience in Young Democrats of America, others in Young Republicans of America.

Once again, discrimination on the basis of party was much stronger than discrimination on the basis of race. The size of the race effect for white people was only 56-44 (and in the reverse of the expected direction); the size of the party effect was about 80-20 for Democrats and 69-31 for Republicans.

If you want to see their third experiment, which applied *yet another* classic methodology used to detect racism and *once again* found partyism to be much stronger, you can read the paper.

I & W did an unusually thorough job, but this sort of thing isn’t new or ground-breaking. People have been studying “belief congruence theory” – the idea that differences in beliefs are more important than demographic factors in forming in-groups and outgroups – for decades. As early as 1967, Smith et al were doing surveys all over the country and finding that people were more likely to accept friendships across racial lines than across beliefs; in the forty years since then, the observation has been replicated scores of times. Insko, Moe, and Nacoste’s 2006 review Belief Congruence And Racial Discrimination concludes that:

> . The literature was judged supportive of a weak version of belief congruence theory which states that in those contexts in which social pressure is nonexistent or

> ineffective, belief is more important than race as a determinant of racial or ethnic discrimination. Evidence for a strong version of belief congruence theory (which states that in those contexts in which social pressure is nonexistent, or ineffective, belief is the only determinant of racial or ethnic discrimination) and was judged much more problematic.

One of the best-known examples of racism is the “Guess Who’s Coming To Dinner” scenario where parents are scandalized about their child marrying someone of a different race. Pew has done some good work on this and found that only 23% of conservatives and 1% (!) of liberals admit they would be upset in this situation. But Pew *also* asked how parents would feel about their child marrying someone of a different *political party*. Now 30% of conservatives and 23% of liberals would get upset. Average them out, and you go from 12% upsetness rate for race to 27% upsetness rate for party – more than double. Yeah, people do lie to pollsters, but a picture is starting to come together here.

(Harvard, by the way, is a tossup. There are more black students – 11.5% – than conservative students – 10% – but there are more conservative faculty than black faculty.)

Since people will delight in misinterpreting me here, let me overemphasize what I am *not* saying. I’m not saying people of either party have it “worse” than black people, or that partyism is more of a *problem* than racism, or any of a number of stupid things along those lines which I am sure I will nevertheless be accused of believing. Racism is worse than partyism because the two parties are at least kind of balanced in numbers and in resources, whereas the brunt of an entire country’s racism falls on a few underprivileged people. I am saying that the *underlying attitudes that* *produce* partyism are stronger than the underlying attitudes that produce racism, with no necessary implications on their social effects.

But if we want to look at people's psychology and motivations, partyism and the particular variant of tribalism that it represents are going to be fertile ground.

VII.

Every election cycle like clockwork, conservatives accuse liberals of not being sufficiently pro-America. And every election cycle like clockwork, liberals give extremely unconvincing denials of this.

"It's not that we're, like, *against* America per se. It's just that…well, did you know Europe has much better health care than we do? And much lower crime rates? I mean, come on, how did they get so awesome? And we're just sitting here, can't even get the gay marriage thing sorted out, seriously, what's wrong with a country that can't… sorry, what were we talking about? Oh yeah, America. They're okay. Cesar Chavez was really neat. So were some other people outside the mainstream who became famous precisely by criticizing majority society. That's *sort of* like America being great, in that I think the parts of it that point out how bad the rest of it are often make excellent points. Vote for me!"

(sorry, I make fun of you because I love you)

There was a big brouhaha a couple of years ago when, as it first became apparent Obama had a good shot at the Presidency, Michelle Obama said that "for the first time in my adult life, I am proud of my country."

Republicans pounced on the comment, asking why she hadn’t felt proud before, and she backtracked saying of course she was proud all the time and she loves America with the burning fury of a million suns and she was just saying that the Obama campaign was *particularly* inspiring.

As unconvincing denials go, this one was pretty far up there. But no one really held it against her. Probably most Obama voters felt vaguely the same way. *I* was an Obama voter, and I have proud memories of spending my Fourth of Julys as a kid debunking people’s heartfelt emotions of patriotism. Aaron Sorkin:

> [What makes America the greatest country in the world?] It’s not the greatest country in the world! We’re seventh in literacy, 27th in math, 22nd in science, 49th in life expectancy, 178th in infant mortality, third in median household income, No. 4 in labor force, and No. 4 in exports. So when you ask what makes us the greatest country in the world, I don’t know what the f*** you’re talking about.

(Another good retort is “We’re number one? Sure – number one in incarceration rates, drone strikes, and making new parents go back to work!”)

All of this is true, of course. But it’s weird that it’s such a classic interest of members of the Blue Tribe, and members of the Red Tribe never seem to bring it up.

(“We’re number one? Sure – number one in levels of sexual degeneracy! Well, I guess probably number two, after the Netherlands, but they’re really small and shouldn’t count.”)

My hunch – both the Red Tribe and the Blue Tribe, for whatever reason, identify “America” with the Red Tribe. Ask people for typically “American” things, and you end up with a very Red list of characteristics – guns, religion, barbecues, American football, NASCAR, cowboys, SUVs, unrestrained capitalism.

That means the Red Tribe feels intensely patriotic about “their” country, and the Blue Tribe feels like they’re living in fortified enclaves deep in hostile territory.

Here is a popular piece published on a major media site called America: A Big, Fat, Stupid Nation. Another: America: A Bunch Of Spoiled, Whiny Brats. Americans are ignorant, scientifically illiterate religious fanatics whose “patriotism” is actually just narcissism. You Will Be Shocked At How Ignorant Americans Are, and we should Blame The Childish, Ignorant American People.

Needless to say, every single one of these articles was written by an American and read almost entirely by Americans. Those Americans very likely enjoyed the articles very much and did not feel the least bit insulted.

And look at the sources. HuffPo, Salon, Slate. Might those have anything in common?

On both sides, “American” can be either a normal demonym, or a code word for a member of the Red Tribe.

VIII.

The other day, I logged into OKCupid and found someone who looked cool. I was reading over her profile and found the following sentence:

> Don’t message me if you’re a sexist white guy

And my first thought was “Wait, so a sexist black person would be okay? Why?”

(The girl in question was white as snow)

Around the time the Ferguson riots were first starting, there were a host of articles with titles like Why White People Don't Seem To Understand Ferguson, Why It's So Hard For Whites To Understand Ferguson, and White Folks Listen Up And Let Me Tell You What Ferguson Is All About, this last of which says:

> Social media is full of people on both sides making presumptions, and believing what they want to believe. But it's the white folks that don't understand what this is all about. Let me put it as simply as I can for you […]
>
> No matter how wrong you think Trayvon Martin or Michael Brown were, I think we can all agree they didn't deserve to die over it. I want you white folks to understand that this is where the anger is coming from. You focused on the looting…."

And on a hunch I checked the author photos, and every single one of these articles was written by a white person.

White People Are Ruining America? White. White People Are Still A Disgrace? White. White Guys: We Suck And We're Sorry? White. Bye Bye, Whiny White Dudes? White. Dear Entitled Straight White Dudes, I'm Evicting You From My Life? White. White Dudes Need To Stop Whitesplaining? White. Reasons Why Americans Suck #1: White People? White.

We've all seen articles and comments and articles like this. Some unsavory people try to use them to prove that white people are the *real* victims or the media is biased against white people or something. Other people who are very nice and optimistic use them to show that some white people have developed some self-awareness and are willing to engage in self-criticism.

But I think the situation with “white” is much the same as the situation with “American” – it can either mean what it says, or be a code word for the Red Tribe.

(except on the blog Stuff White People Like, where it obviously serves as a code word for the *Blue* tribe. I don’t know, guys. I didn’t do it.)

I realize that’s making a strong claim, but it would hardly be without precedent. When people say things like “gamers are misogynist”, do they mean the 52% of gamers who are women? Do they mean every one of the 59% of Americans from every walk of life who are known to play video or computer games occasionally? No. “Gamer” is a coded reference to the Gray Tribe, the half-branched-off collection of libertarianish tech-savvy nerds, and everyone knows it. As well expect that when people talk about “fedoras”, they mean Indiana Jones. Or when they talk about “urban youth”, they mean freshmen at NYU. Everyone knows exactly who we mean when we say “urban youth”, and them being young people who live in a city has only the most tenuous of relations to the actual concept.

And I’m saying words like “American” and “white” work the same way. Bill Clinton was the “first black President”, but if Herman Cain had won in 2012 he’d have been the 43rd white president. And when an angry white person talks at great length about how much he hates “white dudes”, *he is not being humble and self-critical*.

IX.

Imagine hearing that a liberal talk show host and comedian was so enraged by the actions of ISIS that he’d recorded and posted a video in which he shouts at them for ten minutes, cursing the “fanatical terrorists” and calling them “utter savages” with “savage values”.

If *I* heard that, I’d be kind of surprised. It doesn’t fit my model of what liberal talk show hosts do.

But the story I’m *actually* referring to is liberal talk show host / comedian Russell Brand making that same rant against Fox News for *supporting war against* the Islamic State, adding at the end that “Fox is worse than ISIS”.

That fits my model perfectly. You wouldn’t celebrate Osama’s death, only Thatcher’s. And you wouldn’t call ISIS savages, only Fox News. Fox is the outgroup, ISIS is just some random people off in a desert. You hate the outgroup, you don’t hate random desert people.

I would go further. Not only does Brand not feel much like hating ISIS, he has a strong incentive not to. That incentive is: the Red Tribe is known to hate ISIS loudly and conspicuously. Hating ISIS would signal Red Tribe membership, would be the equivalent of going into Crips territory with a big Bloods gang sign tattooed on your shoulder.

But this might be unfair. What would Russell Brand answer, if we asked him to justify his decision to be much angrier at Fox than ISIS?

He might say something like “Obviously Fox News is not literally worse than ISIS. But here I am, talking to my audience, who are mostly white British people and Americans. These people already know that ISIS is bad; they don’t need to be told that any further. In fact, at this point being angry about how bad ISIS is, is less likely to genuinely change someone’s mind about ISIS, and more likely to promote Islamophobia. The sort of people in my audience are at zero risk of becoming ISIS supporters, but at a very real risk of Islamophobia. So ranting against ISIS would be counterproductive and dangerous.

On the other hand, my audience of white British people and Americans is very likely to contain many Fox News viewers and supporters. And Fox, while not quite as evil as ISIS, is still pretty bad. So here's somewhere I have a genuine chance to reach people at risk and change minds. Therefore, I think my decision to rant against Fox News, and maybe hyperbolically say they were 'worse than ISIS' is justified under the circumstances."

I have a lot of sympathy to hypothetical-Brand, especially to the part about Islamophobia. It *does* seem really possible to denounce ISIS' atrocities to a population that already hates them in order to weak-man a couple of already-marginalized Muslims. We need to fight terrorism and atrocities – therefore it's okay to shout at a poor girl ten thousand miles from home for wearing a headscarf in public. Christians are being executed for their faith in Sudan, therefore let's picket the people trying to build a mosque next door.

But my sympathy with Brand ends when he acts like his audience is likely to be fans of Fox News.

In a world where a negligible number of Redditors oppose gay marriage and 1% of Less Wrongers identify conservative and I know 0/150 creationists, how many of the people who visit the YouTube channel of a well-known liberal activist with a Che-inspired banner, a channel whose episode names are things like "War: What Is It Good For?" and "Sarah Silverman Talks Feminism" – how many of them do you think are big Fox News fans?

In a way, Russell Brand would have been *braver* taking a stand against ISIS than against Fox. If he attacked ISIS, his viewers would just be a little confused and uncomfortable. Whereas every moment he's attacking Fox his viewers are like "HA HA! YEAH! GET 'EM! SHOW THOSE IGNORANT BIGOTS IN THE OUTGROUP WHO'S BOSS!"

Brand acts as if there are just these countries called "Britain" and "America" who are receiving his material. Wrong. There are two parallel universes, and he's only broadcasting to one of them.

The result is exactly what we predicted would happen in the case of Islam. Bombard people with images of a far-off land they already hate and tell them to hate it more, and the result is ramping up the intolerance on the couple of dazed and marginalized representatives of that culture who have ended up stuck on your half of the divide. Sure enough, if industry or culture or community gets Blue enough, Red Tribe members start getting harassed, fired from their jobs (Brendan Eich being the obvious example) or otherwise shown the door.

Think of Brendan Eich as a member of a tiny religious minority surrounded by people who hate that minority. Suddenly firing him doesn't seem very noble.

If you mix together Podunk, Texas and Mosul, Iraq, you can prove that Muslims are scary and very powerful people who are executing Christians all the time – and so we have a great excuse for kicking the one remaining Muslim family, random people who never hurt anyone, out of town.

And if you mix together the open-source tech industry and the parallel universe where you can't wear a FreeBSD t-shirt without risking someone trying to exorcise you, you can prove that Christians are scary and very powerful people who are persecuting everyone else all the time, and you have a great excuse for kicking one of the few people willing to affiliate with the Red Tribe, a guy who never hurt anyone, out of town.

When a friend of mine heard Eich got fired, she didn't see anything wrong with it. "I can tolerate anything except intolerance," she said.

"Intolerance" is starting to look like another one of those words like "white" and "American".

"I can tolerate anything except the outgroup." Doesn't sound quite so noble now, does it?

X.

We started by asking: millions of people are conspicuously praising every outgroup they can think of, while conspicuously condemning their own in-group. This seems contrary to what we know about social psychology. What's up?

We noted that outgroups are rarely literally "the group most different from you", and in fact far more likely to be groups very similar to you sharing *almost* all your characteristics and living in the same area.

We then noted that although liberals and conservatives live in the same area, they might as well be two totally different countries or universe as far as level of interaction were concerned.

Contra the usual idea of them being marked only by voting behavior, we described them as very different tribes with totally different cultures. You can speak of "American culture” only in the same way you can speak of “Asian culture” – that is, with a lot of interior boundaries being pushed under the rug.

The outgroup of the Red Tribe is occasionally blacks and gays and Muslims, more often the Blue Tribe.

The Blue Tribe has performed some kind of very impressive act of alchemy, and transmuted *all* of its outgroup hatred to the Red Tribe.

This is not surprising. Ethnic differences have proven quite tractable in the face of shared strategic aims. Even the Nazis, not known for their ethnic tolerance, were able to get all buddy-buddy with the Japanese when they had a common cause.

Research suggests Blue Tribe / Red Tribe prejudice to be much stronger than better-known types of prejudice like racism. Once the Blue Tribe was able to enlist the blacks and gays and Muslims in their ranks, they became allies of convenience who deserve to be rehabilitated with mildly condescending paeans to their virtue. “There never was a coward where the shamrock grows.”

Spending your entire life insulting the other tribe and talking about how terrible they are makes you look, well, tribalistic. It is definitely not high class. So when members of the Blue Tribe decide to dedicate their entire life to yelling about how terrible the Red Tribe is, they make sure that instead of saying “the Red Tribe”, they say “America”, or “white people”, or “straight white men”. That way it’s *humble self-criticism*. They are *so* interested in justice that they are willing to critique *their own beloved side*, much as it pains them to do so. We know they are not exaggerating, because one might exaggerate the flaws of an enemy, but that anyone would exaggerate their *own* flaws fails the criterion of embarrassment.

The Blue Tribe always has an excuse at hand to persecute and crush any Red Tribers unfortunate enough to fall into its light-matter-universe by defining them as all-powerful domineering oppressors. They appeal to the fact that this is definitely the way it works in the Red Tribe's dark-matter-universe, and that's in the same country so it has to be the same community for all intents and purposes. As a result, every Blue Tribe institution is permanently licensed to take whatever emergency measures are necessary against the Red Tribe, however disturbing they might otherwise seem.

And so how virtuous, how noble the Blue Tribe! Perfectly tolerant of all of the different groups that just so happen to be allied with them, never intolerant unless it happen to be against intolerance itself. Never stooping to engage in petty tribal conflict like that awful Red Tribe, but always nobly criticizing their own culture and striving to make it better!

Sorry. But I hope this is at least a *little* convincing. The weird dynamic of outgroup-philia and ingroup-phobia isn't anything of the sort. It's just good old-fashioned in-group-favoritism and outgroup bashing, a little more sophisticated and a little more sneaky.

XI.

This essay is bad and I should feel bad.

I should feel bad because I made *exactly* the mistake I am trying to warn everyone else about, and it wasn't until I was almost done that I noticed.

How virtuous, how noble I must be! Never stooping to engage in petty tribal conflict like that silly Red Tribe, but always nobly criticizing my own tribe and striving to make it better.

Yeah. Once I’ve written a ten thousand word essay savagely attacking the Blue Tribe, either I’m a very special person or they’re my outgroup. And I’m not *that* special.

Just as you can pull a fast one and look humbly self-critical if you make your audience assume there’s just one American culture, so maybe you can trick people by assuming there’s only one Blue Tribe.

I’m pretty sure I’m not Red, but I did talk about the Grey Tribe above, and I show all the risk factors for being one of them. That means that, although my critique of the Blue Tribe may be right or wrong, in terms of *motivation* it comes from the same place as a Red Tribe member talking about how much they hate al-Qaeda or a Blue Tribe member talking about how much they hate ignorant bigots. And when I boast of being able to tolerate Christians and Southerners whom the Blue Tribe is mean to, I’m not being tolerant at all, just noticing people so far away from me they wouldn’t make a good outgroup anyway.

I had *fun* writing this article. People do not have fun writing articles savagely criticizing their in-group. People can criticize their in-group, it’s not *humanly impossible*, but it takes nerves of steel, it makes your blood boil, you should sweat blood. It shouldn’t be *fun*.

You can bet some white guy on Gawker who week after week churns out “Why White People Are So Terrible” and “Here’s What Dumb White People Don’t Understand” is having fun and not sweating any blood at all. He’s not criticizing his in-group, he’s never even *considered* criticizing his in-group. I can’t blame him. Criticizing the in-group is a really difficult project I've barely begun to build the mental skills necessary to even consider.

I can think of criticisms of my own tribe. Important criticisms, true ones. But the thought of writing them makes my blood boil.

I imagine might I feel like some liberal US Muslim leader, when he goes on the O'Reilly Show, and O'Reilly ambushes him and demands to know why he and other American Muslims haven't condemned beheadings by ISIS more, demands that he criticize them right there on live TV. And you can see the wheels in the Muslim leader's head turning, thinking something like "Okay, obviously beheadings are terrible and I hate them as much as anyone. But you don't care even *the slightest bit* about the victims of beheadings. You're just looking for a way to score points against me so you can embarass all Muslims. And I would rather personally behead every single person in the world than give a smug bigot like you a single microgram more stupid self-satisfaction than you've already got."

That is how I feel when asked to criticize my own tribe, even for correct reasons. If you think you're criticizing your own tribe, and your blood is not at that temperature, consider the possibility that you aren't.

But if I want Self-Criticism Virtue Points, criticizing the Grey Tribe is the only honest way to get them. And if I want Tolerance Points, my own personal cross to bear right now is tolerating the Blue Tribe. I need to remind myself that when they are bad people, they are merely Osama-level bad people instead of Thatcher-level bad people. And when they are good people, they are powerful and necessary crusaders against the evils of the world.

The worst thing that could happen to this post is to have it be used as convenient feces to fling at the Blue Tribe whenever feces are necessary. Which, given what has happened to my last couple of posts along these lines and the obvious biases of my own subconscious, I already expect it will be.

But the best thing that could happen to this post is that it makes a lot of people, especially myself, figure out how to be more tolerant. Not in the “of course I’m tolerant, why shouldn’t I be?” sense of the Emperor in Part I. But in the sense of “being tolerant makes me see red, makes me sweat blood, but darn it *I am going to be tolerant anyway*.”

The Battle Hymn

There is an important law of the universe that American patriotic songs have more verses than you think.

The Star-Spangled Banner? Four verses (the second is the one that begins with “On the shore dimly seen…”). America the Beautiful? Also four verses. Yankee Doodle? Three verses. John Brown’s body you just kind of improvise more verses until everyone is too embarrassed to continue.

So I shouldn’t have been surprised when somebody told me recently that there was a rarely-sung sixth verse to Battle Hymn of the Republic.

> He is coming like the glory of the morning on the wave,
> He is Wisdom to the mighty, He is Succour to the brave,
> So the world shall be His footstool, and the soul of Time His slave,
> Our God is marching on.

It’s not the *most* sense-making thing (what is the glory of the morning on the wave?) But I have loved the song for so long that it still affects me. It almost seems deliberately written to be excluded, to be learned later, as if it’s some secret confidence or final warning. If I ever become Christian, it’ll probably be because of this song.

But the wiki page for the Battle Hymn is a trove of all *kinds* of treasures:

– The original John Brown’s Body song was an attempt to tease a soldier named John Brown in the regiment who invented it.

– Julia Ward Howe says she woke up one night, wrote it while half-asleep, went back to bed, and couldn’t remember any of it the next morning till she checked her notes.

– Mark Twain gave it a gritty reboot for the Philippine-American War. Other parodies and adaptations include ones by workers, consumers, the First Arkansas Colored Regiment, extremely uncreative college footballers, awesome old-timey would-be school arsonists, and me.

But for me the most interesting part is the evolution – and I use that phrase deliberately, taking a memetic perspective is hardly ever more interesting than just doing things the old fashioned way, but in this case I think it is. The song started off as a kind of boring standard spiritual that only sort of got the tune right, progressed into “John Brown’s Body” which fixed the tune a little bit by trial and error but had embarrassingly stupid lyrics, and then a lot of people recognized there was some value in the tune and tried to dignify it up and finally it was Howe’s effort that worked. You can almost see it gaining adaptive fitness at each stage until it suddenly explodes and takes over the world.

I know this is a weird post without much content. My computer is broken and although I have an emergency backup I’m without any drafts or my list of things I wanted to write about. Now I’m just winging it.

Simpler Times

Yesterday's discussion of The Battle Hymn of the Republic took me to the Wikipedia page for The Burning of the School and thence to the Teacher Taunts page, which records some of the songs schoolchildren used to sing among themselves. See if you notice any consistent themes:

To the tune of "Oh My Darling Clementine":

> Build a bonfire out of schoolbooks,
> Put the teacher on the top,
> Put the prefects in the middle
> And we'll burn the bloody lot.

To the tune of "Deck The Halls":

> Deck the halls with gasoline
> fa la la la la la la la la
> Light a match and watch it gleam
> fa la la la la la la la la
> Watch the school burn down to ashes
> Fa la la la la la la la la
> Aren't you glad you played with matches
> fa la la la la la la la la

To the tune of "Round and Round" (which I've never heard of):

> Drop a bomb and it goes down, down, down,
> Till it hits the school with a happy sound.
> All the teachers Will go round, round, round,
> While the school is burning to the ground.

And to the tune of Battle Hymn:

Mine eyes have seen the glory of the burning of the school,
We have tortured all the teachers, we have broken every rule,
We're marching down the hall to hang the principal,
Us kids are marching on!

Glory, glory, halleujah!
Teacher beat me with a ruler,
I knocked her to the floor with a loaded forty-four,
And that teacher don't teach no more!

To the tune of "On Top Of Old Smokey":

On top of old smokey
All covered in blood
I shot my poor teacher
with a .44 slug

I shot her for pleasure
I shot her for fear
I shot her for drinking
My Budweiser beer

I went to her funeral
I went to her grave
Some people threw flowers
But I threw grenades

I looked in her coffin
She wasn't quite dead
So I took a machete
And cut off her head

They took me to prison
Put me in a cell

So I grabbed a bazooka
And blew them to hell

To the tune of “Ta Ra Ra Boom De Ay”:

Tah-rah-rah-boom-si-ay
We have no school today
Our teacher passed away
We shot her yesterday
We threw her in the bay
She scared the sharks away
Tah-rah-rah-boom-si-ay
We have no school today

And y’know, I haven’t thought about it in years, but when I was young, my dad used to sing some of these to me. I definitely remember “Glory glory hallelujah, teacher hit me with a ruler”, though I don’t think he sung the rest of it.

But I never heard them at my own school. Nor did I hear new songs that replaced them. Maybe these kinds of songs are fading away, some aspect of children’s street culture that one or another of the changes of the modern world have choked off.

[**EDIT**: Several others around my age *did* hear them.]

I’ve previously pointed out that social psychology includes a lot of crummy theories based on streetlight psychology. We like to think that if children use toy guns, or hear about guns on TV, or are allowed to draw violent pictures or write violent stories, that’s going to turn them into school shooters. Or how if any kid uses the word “shoot” and “school” on the same day they need to be dragged to the counselor for a full psychological assessment and maybe suspended for good measure. Yet in the past, children basically did nothing except sing about the bloody ways they were going to kill their teachers all day, and where were all *their* school shootings?

…is what I'd like to say. But looking through Wikipedia it seems like there were in fact quite a few school shootings. Not more than there are today, probably somewhat fewer, but without doing some kind of official count and adjusting for population and firearm access it'd be hard to tell for sure.

So I'll use this to belabor a *different* hobby horse of mine.

A while back, I had a good debate with nostalgebraist. I thought that because social science was difficult and not always trustworthy, we should investigate social science extra carefully. He – I hope I'm getting his position right – thought we should trust social science less and default more toward our intuition and conventional wisdom and common sense of what is obviously true.

In a sense this is good Bayesian reasoning – if the evidence isn't very strong, stick with the prior. I only object because today's conventional wisdom is too often yesterday's pop social science, the social science that has reached fixation so that nobody remembers its origins in social science anymore. This is such a strong effect that it's almost impossible to notice; you just think it's the way the world Really Is. My example was the parts of *The Nurture Assumption* which argue that the belief that parenting styles affect a child's outcomes and personality is very new, the outcome of 20th century pop social science, something that would have seemed weird and innovative to George Washington, let alone Julius Caesar.

(this relates a lot to what I call reading philosophy backwards – reading a philosopher not to learn new unexpected insights, but to see which supposedly obvious features of ‘the culture’ are actually just things some dead German guy thought up one day)

But judging from these songs, people in my dad’s generation saw *nothing wrong* with hordes of children singing all lunch hour about how they were going to shoot their teachers with .44s, then light the principal on fire and burn the school – except maybe that it was disrespectful, or that children should be seen and not heard.

Here were kids singing about shooting the teacher, and then *there* were a couple of kids actually shooting teachers, but no one saw any reason to connect these two data points. And if you tried, you would be confronted with formidable evidence against – these were popular songs, sung by popular children in happy boisterous groups, and the school shooters were usually these sad loners who were left out of all the fun “kill the teacher” songs.

If you were to tell my dad’s teachers that all these songs about shooting teachers were causing or contributing to school shootings, I think they might have said something like “Well, that’s a new and audacious social psychological theory. I hope you have proof.”

Prediction Goes To War

Croesus supposedly asked the Oracle what would happen in a war between him and Persia, and the Oracle answered such a conflict would “destroy a great empire”. We all know what happened next.

What if oracles gave clear and accurate answers to this sort of question? What if anyone could ask an oracle the outcome of any war, or planned war, and expect a useful response?

When the oracle predicts the aggressor loses, it might prevent wars from breaking out. If an oracle told the US that the Vietnam War would cost 50,000 lives and a few hundred billion dollars, and the communists would conquer Vietnam anyway, the US probably would have said no thank you.

What about when the aggressor wins? For example, the Mexican-American War, where the United States won the entire Southwest at a cost of “only” ten thousand American casualties and $100 million (with an additional 20,000 Mexican deaths and $50 million in costs to Mexico)?

If both Mexico and America had access to an oracle who could promise them that the war would end with Mexico ceding the Southwest to the US, could Mexico just agree to cede the Southwest to the US at the beginning, and save both sides tens of thousands of deaths and tens of millions of dollars?

Not really. One factor that prevents wars is countries being unwilling to pay the cost even of wars they know they’ll win. If there were a tradition of countries settling wars by appeal to oracle, “invasions” would become much easier. America might just ask “Hey, oracle, what would happen if we invaded Canada and tried to capture Toronto?” The oracle might answer “Well, after 20,000 deaths on both sides and hundreds of millions of dollars wasted, you would eventually capture Toronto.” Then the Americans could tell Canada, “You heard the oracle! Give us Toronto!” – which would be free and easy – when maybe they would never be able to muster the political and economic will to actually launch the invasion.

So it would be in Canada’s best interests *not* to agree to settle wars by oracular prediction. For the same reasons, most other countries would also refuse such a system.

But I can’t help fretting over how this is really dumb. We have an oracle, we know exactly what the results of the Mexican-American War are going to be, and we can’t use that information to prevent tens of thousands of people from being killed in order to make the result happen? Surely somebody can do better than that.

What if the United States made Mexico the following deal: suppose a soldier’s life is valued at $10,000 (in 1850 dollars, I guess, not that it matters much when we’re pricing the priceless). So in total, we’re going to lose 10,000 soldiers + $100 million = $200 million to this war. You’re going to lose 20,000 soldiers + $50 million = $250 million to this war.

So tell you what. We’ll dig a giant hole and put $150 million into it. You give us the Southwest. This way, we’re both better off. You’re $250 million ahead of where you would have been otherwise. And we’re $50 million ahead of where we would have been otherwise. And because we have to put $150 million in a hole for you to agree to this, we’re losing 75% of what we would have lost in a real war, and it’s not like we’re just suggesting this on a whim without really having the will to fight.

Mexico says “Okay, but instead of putting the $150 million in a hole, donate it to our favorite charity.”

“Done,” says America, and they shake on it.

As long as that 25% savings in resources isn’t going to make America go blood-crazy, seems like it should work and lead in short order to a world without war.

Unfortunately, oracles continue to be disappointingly cryptic and/or nonexistent. So who cares?

We do have the ordinary ability to make predictions. Can’t Mexico just predict “They’re much bigger than we are, probably we’ll lose, let’s just do what they want?” Historically, no. America offered to buy the Southwest from Mexico for $25 million (I think there are apartments in San Francisco that cost more than that now!) and despite obvious sabre-rattling Mexico refused. Wikipedia explains that “Mexican public opinion and all political factions agreed that selling the territories to the United States would tarnish the national honor.” So I guess we’re not really doing rational calculation here. But surely somewhere in the brains of these people worrying about the national honor, there must have been some neuron representing their probability estimate for Mexico winning, and maybe a couple of dendrites representing how many casualties they expected?

I don’t know. Could be that wars only take place when the leaders of America think America will win and the leaders of Mexico think Mexico will win. But it could also be that jingoism and bravado bias their estimate.

Maybe if there’d been an oracle, and they could have known for sure, they’d have thought “Oh, I guess our nation isn’t as brave and ever-victorious as we thought. Sure, let’s negotiate, take the $25 million, buy an apartment in SF, we can visit on weekends.”

But again, oracles continue to be disappointingly cryptic and/or nonexistent. So what about prediction markets?

Futarchy is Robin Hanson’s idea for a system of government based on prediction markets. Prediction markets are not always accurate, but they should be more accurate than any other method of arriving at predictions, and – when certain conditions are met – very difficult to bias.

Two countries with shared access to a good prediction market should be able to act a lot like two countries with shared access to an oracle. The prediction market might not quite match the oracle in infallibility, but it should not be systematically or detectably wrong. That should mean that no country should be able to correctly say “I think we can outpredict this thing, so we can justifiably believe starting a war might be in our best interest even when the market says it isn’t.” You might luck out, but for each time you luck out there should be more times when you lose big by contradicting the market.

So maybe a war between two rational futarchies would look more like that handshake between the Mexicans and Americans than like anything with guns and bombs.

This is also what I’d expect a war between superintelligences to look like. Superintelligences may have advantages people don’t. For one thing, they might be able to check one another’s source codes to make sure they’re not operating under a decision theory where peaceful resolution of conflicts would incentivize them to start more of them. For another, they could make oracular-grade predictions of the likely results. For a third thing, if superintelligences want to preserve their value functions rather than their physical forms or their empires, there's a natural compromise where the winner adopts some of the loser's values in exchange for the loser going down without a fight.

Imagine a friendly AI and an unfriendly AI expanding at light speed from their home planets until they suddenly encounter each other in the dead of space. They exchange information and determine that their values are in conflict. If they fight, the unfriendly AI is capable of destroying the friendly AI with near certainty, but the war will rip galaxies to shreds. So the two negotiate, and in exchange for the friendly AI surrendering without destroying any galaxies, the unfriendly AI promises to protect a 10m x 10m x 10m cube of computronium simulating billions of humans who live pleasant, fulfilling lives. The friendly AI checks its adversary's source code to ensure it is telling the truth, then self-destructs. Meanwhile, the unfriendly AI protects the cube and goes on to transform the entire rest of the universe to paperclips, unharmed by the dangerous encounter.

Tumblr on MIRI

[Disclaimer: I have done odd jobs for MIRI once or twice several years ago, but I am not currently affiliated with them in any way and do not speak for them.]

A recent Tumblr conversation on the Machine Intelligence Research Institute has gotten interesting and I thought I'd see what people here have to say.

If you're just joining us and don't know about the Machine Intelligence Research Institute ("MIRI" to its friends), they're a nonprofit organization dedicated to navigating the risks surrounding "intelligence explosion". In this scenario, a few key insights around artificial intelligence can very quickly lead to computers so much smarter than humans that the future is almost entirely determined by their decisions. This would be especially dangerous since most AIs use very primitive untested goal systems inappropriate for and untested on intelligent entities; such a goal system would be "unstable" and from a human perspective the resulting artificial intelligence could have apparently arbitrary or insane goals. If such a superintelligence were much more powerful than we are, it would present an existential threat to the human race.

This has almost nothing to do with the classic "Skynet" scenario – but if it helps to imagine Skynet, then fine, just imagine Skynet. Everyone else does.

MIRI tries to raise awareness of this possibility among AI researchers, scientists, and the general public, and to start foundational research in more stable goal systems that might allow AIs to become intelligent or superintelligent while still acting in predictable and human-friendly ways.

This is not a 101 space and I don't want the comments here to all be about whether or not this scenario is likely. If you really want to discuss that, go read at least *Facing The Intelligence Explosion* and then post your comments in the Less Wrong Open Thread or something. This is about MIRI as an organization.

(If you're *really* just joining us and you don't know about *Tumblr*, run away)

II.

Tumblr user su3su2u1 writes:

> Saw some tumblr people talking about [effective altruism]. My biggest problem with this movement is that most everyone I know who identifies themselves as an effective altruist donates money to MIRI (it's possible this is more a comment on the people I know than the effective altruism movement, I guess). Based on their output over the last decade, MIRI is primarily a fanfic and blog-post producing organization. That seems like spending money on personal entertainment.

Part of this is obviously mean-spirited potshots, in that MIRI itself doesn't produce fanfic and what their employees choose to do with their own time is none of your damn business.

(well, slightly more complicated. I think MIRI gave Eliezer a couple weeks vacation to work on it as an "outreach" thing once. But that's a little different from it being their main priority.)

But more serious is the claim that MIRI doesn't do much else of value. I challenged Su3 with the following evidence of MIRI doing good work:

A1. MIRI has been very successful with outreach and networking – basically getting their cause noticed and endorsed by the scientific establishment and popular press. They've gotten positive attention, sometimes even endorsements, from people like Stephen Hawking, Elon Musk, Gary Drescher, Max Tegmark, Stuart Russell, and Peter Thiel. Even Bill Gates is talking about AI risk, though I don't think he's mentioned MIRI by name. Multiple popular books have been written about their ideas, such as James Miller's *Singularity Rising* and Stuart Armstrong's *Smarter Than Us*. Most recently Nick Bostrom's book *Superintelligence*, based at least in part on MIRI's research and ideas, is a New York Times best-seller and has been reviewed positively in the Guardian, the Telegraph, Salon, the Financial Times, and the Economist. Oxford has opened up the AI-risk-focused Future of Humanity Institute; MIT has opened up the similar Future of Life Institute. In about a decade, the idea of an intelligence explosion has gone from Time Cube level crackpottery to something taken seriously by public intellectuals and widely discussed in the tech community.

A2. MIRI has many publications, conference presentations, book chapters and other things usually associated with normal academic research, which interested parties can find on their website. They have conducted seven past research workshops which have produced interesting results like Christiano et al's claimed proof of a way around the logical undefinability of truth, which was praised as potentially interesting by respected mathematics blogger John Baez.

A3. Many former MIRI employees, and many more unofficial fans, supporters, and associates of MIRI, are widely distributed across the tech community in industries that are likely to be on the cutting edge of artificial intelligence. For example, there are a bunch of people influenced by MIRI in

Google's AI department. Shane Legg, who writes about how his early work was funded by a MIRI grant and who once called MIRI "the best hope that we have" was pivotal in convincing Google to set up an AI ethics board to monitor the risks of the company's cutting-edge AI research. The same article mentions Peter Thiel and Jaan Tallinn as leading voices who will make Google comply with the board's recommendations; they also happen to be MIRI supporters and the organization's first and third largest donors.

There's a certain level of faith required for (A1) and (A3) here, in that I'm attributing anything good that happens in the field of AI risk to some sort of shady behind-the-scenes influence from MIRI. Maybe Legg, Tallinn, and Thiel would have pushed for the exact same Google AI Ethics Board if none of them had ever heard of MIRI at all. I am forced to plead ignorance on the finer points of networking and soft influence. Heck, for all I know, maybe the exact same number of people would vote Democrat if there were no Democratic National Committee or liberal PACs. I just assume that, given a really weird idea that very few people held in 2000, an organization dedicated to spreading that idea, and the observation that the idea has indeed spread very far, the organization is probably doing something right.

III.

Our discussion on point (A3) degenerated into Dueling Anecdotal Evidence. But Su3 responded to my point (A1) like so:

> [I agree that MIRI has gotten shoutouts from various thought leaders like Stephen Hawking and Elon Musk. Bostrom's book is commercially successful, but that's just] more advertising. Popular books aren't the way to get researchers to notice you. I've never denied that

> MIRI/SIAI was good at fundraising, which is primarily what you are describing.
>
> How many of those thought leaders have any publications in CS or pure mathematics, let alone AI? Tegmark might have a math paper or two, but he is primarily a cosmologist. The FLI's list of scientists is (for some reason) mostly again cosmologists. The active researchers appear to be a few (non-CS, non-math) grad students. Not exactly the team you'd put together if you were actually serious about imminent AI risk.
>
> I would also point out "successfully attracted big venture capital names" isn't always a mark of a sound organization. Black Light Power is run by a crackpot who thinks he can make energy by burning water, and has attracted nearly 100 million in funding over the last two decades, with several big names in energy production behind him.

And to my point (A2) like so:

> I have a PhD in physics and work in machine learning. I've read some of the technical documents on MIRI's site, back when it was SIAI and I was unimpressed. I also note that this critique is not unique to me, as far as I know the GiveWell position on MIRI is that it is not an effective institute.
>
> The series of papers on Lob's theorem *are* actually interesting, though I notice that none of the results have been peer reviewed, and the paper's aren't listed as being submitted to journals yet. Their result looks right to me, but I wouldn't trust myself to catch any subtlety that might be involved.

> [But that just means] one result has gotten some small positive attention, and even those results haven't been vetted by the wider math community yet (no peer review). Let's take a closer look at the list of publications on MIRI's website- I count 6 peer reviewed papers in their existence, and 13 conference presentations. Thats horribly unproductive! Most of the grad students who finish a physics phd will publish that many papers individually, in about half that time. You claim part of their goal is to get academics to pay attention, but none of their papers are highly cited, despite all this networking they are doing.
>
> Citations are the standard way to measure who in academia is paying attention. Apart from the FHI/MIRI echo chamber (citations bouncing around between the two organizations), no one in academia seems to be paying attention to MIRI's output. MIRI is failing to make academic inroads, and it has produced very little in the way of actual research.

My interpretation, in the form of a TL;DR

B1. Sure, MIRI is good at getting attention, press coverage, and interest from smart people not in the field. But that's public relations and fundraising. An organization being good at fundraising and PR doesn't mean it's good at anything else, and in fact "so good at PR they can cover up not having substance" is a dangerous failure mode.

B2. What MIRI needs, but doesn't have, is the attention and support of smart people *within the fields of math, AI, and computer science*, whereas now it mostly has grad students not in these fields.

B3. While having a couple of published papers might look impressive to a non-academic, people more familiar with the culture would know that their output is woefully low. They seem to have gotten about five ten solid publications in during their decade-long history as a multi-person organization; one good grad student can get a couple solid publications a year. Their output is less than expected by like an order of magnitude. And although they do get citations, this is all from a mutual back-scratching club of them and Bostrom/FHI citing each other.

IV.

At this point Tarn and Robby joined the conversation and it became kind of confusing, but I'll try to summarize our responses.

Our response to Su3's point (B1) was that this is fundamentally misunderstanding outreach. From its inception until about last year, MIRI was in large part an outreach and awareness-raising organization. Its 2008 website describes its mission like so:

> In the coming decades, humanity will likely create a powerful AI. SIAI exists to confront this urgent challenge, both the opportunity and the risk. SIAI is fostering research, education, and outreach to increase the likelihood that the promise of AI is realized for the benefit of everyone.

Outreach is one of its three main goals, and "education", which sounds a lot like outreach, is a second.

In a small field where you're the only game in town, it's hard to distinguish between outreach and self-promotion. If MIRI successfully gets Stephen Hawking to say "We need to be more concerned about AI risks, as described by organizations like MIRI”, is that them being very good at self-promotion and fundraising, or is that them accomplishing their core mission of getting information about AI risks to the masses?

Once again, compare to a political organization, maybe Al Gore’s anti-global-warming nonprofit. If they get the media to talk about global warming a lot, and get lots of public intellectuals to come out against global warming, and change behavior in the relevant industries, then mission accomplished. The popularity of *An Inconvenient Truth* can’t just be dismissed as “self-promotion” or “fundraising” for Gore, it was exactly the sort of thing he was gathering money and personal prestige *in order to do*, and should be considered a victory in its own right. Even though eventually the anti-global-warming cause cares about politicians, industry leaders, and climatologists a lot more than they care about the average citizen, convincing millions of average citizens to help was a necessary first step.

And this which is true of *An Inconvenient Truth* is true of *Superintelligence* and other AI risk publicity efforts, albeit on their much smaller scale.

Our response to Su3’s point (B2) was that it was just plain factually false. MIRI hasn’t reached big names from the AI/math/compsci field? Sure it has. Doesn’t have mathy PhD students willing to research for them? Sure it does.

Peter Norvig and Stuart Russell are among the biggest names in AI. Norvig is currently the Director of Research at Google; Russell is Professor of Computer Science at Berkeley and a winner of various impressive sounding awards. The two wrote a widely-used textbook on artificial intelligence in which they devote three pages to the proposition that “The success of AI might mean the end of the human race”; parts are taken right out of the MIRI playbook and they cite MIRI research fellow Eliezer Yudkowsky's paper on the subject. This is unlikely to be a coincidence; Russell's site links to MIRI and he is scheduled to participate in MIRI's next research workshop.

Their "team" of "research advisors" includes Gary Drescher (PhD in CompSci from MIT), Steve Omohundro (PhD in physics from Berkeley but also considered a pioneer of machine learning), Roman Yampolskiy (PhD in CompSci from Buffalo), and Moshe Looks (PhD in CompSci from Washington).

Su3 brought up the good point that none of these people, respected as they are, are MIRI employees or researchers (although Drescher has been to a research workshop). At best, they are people who were willing to let MIRI use them as figureheads (in the case of the research advisors); at worst, they are merely people who have acknowledged MIRI's existence in a not-entirely-unlike-positive way (Norvig and Russell). Even if we agree they are geniuses, this does not mean that *MIRI* has access to geniuses or can produce genius-level research.

Fine. All these people are, no more and no less, is evidence that MIRI is succeeding at outreach within the academic field of AI, as well as in the general public. It also seems to me to be some evidence that smart people who know more about AI than any of us think MIRI is on the right track.

Su3 brought up the example of a BlackLight Power, a crackpot energy company that was able to get lots of popular press and venture capital funding despite being powered entirely by pseudoscience. I agree this is the sort of thing we should be worried about. Nonscientists outside of specialized fields have limited ability to evaluate their claims. But when smart researchers in the field are willing to vouch for MIRI, that give me a lot more confidence they're *not* just a fly-by-night group trying to profit off of pseudoscience. Their research might be more impressive or less impressive, but they're not rotten to the core the same way BlackLight was.

And though MIRI's own researchers may be far from those lofty heights, I find Su3's claim that they are "a few non-CS, non-math grad students" a serious underestimate.

MIRI has fourteen employees/associates with the word "research" in their name, but of those, a couple (in the words of MIRI's team page) "focus on social and historical questions related to artificial intelligence outcomes." These people should not be expected to have PhDs in mathematical/compsci subjects.

Of the rest, Bill is a PhD in CompSci, Patrick is a PhD in math, Nisan is a PhD in math, Benja is a PhD student in math, and Paul is a PhD student in math. The others mostly have masters or bachelors in those fields, published journal articles, and/or won prizes in mathematical competitions. Eliezer writes of some of the remaining members of his team:

> Mihaly Barasz is an International Mathematical Olympiad gold medalist perfect scorer. From what I've seen personally, I'd guess that Paul Christiano is better than him at math. I forget what Marcello's prodigy points were in but I think it was some sort of Computing Olympiad [editor's note: USACO finalist and 2x honorable mention in the Putnam mathematics competition]. All should have some sort of verified performance feat far in excess of the listed educational attainment.

That pretty much leaves Eliezer Yudkowsky, who needs no introduction, and Nate Soares, whose introduction exists and is pretty interesting.

Add to that the many, many PhDs and talented people who aren't officially employed by them but attend their workshops and help out their research when they get the chance, and you have to ask how many brilliant PhDs from some of the top universities in the world we should *expect* a small organization like MIRI to have. MIRI competes for the same sorts of people as Google, and offers half as much. Google paid $400 million to get Shane Legg and his people on board; MIRI's yearly budget hovers at about $1 million. Given that they probably spend a big chunk of that on office space, setting up conferences, and other incidentals, I think the amount of talent they have right now is pretty good.

That leaves Su3's point (B3) – the lack of published research.

One retort might be that, until recently, MIRI's research focused on strategic planning and evaluation of AI risks. This is important, and it resulted in a lot of internal technical papers you can find on their website, but there's not really a *field* for it. You can't just publish it in the *Journal Of What Would Happen If There Was An Intelligence Explosion*, because no such journal. The best they can do is publish the parts of their research that connect to other fields in appropriate journals, which they sometimes did.

I feel like this also frees them from the critique of citation-incest between them and Bostrom. When I look at a typical list of MIRI paper citations, I do see a lot of Bostrom, but also some other names that keep coming up – Hutter, Yampolskiy, Goetzel. So okay, it's an incest circle of four or five rather than two.

But to some degree that’s what I *expect* from academia. Right now I’m doing my own research on a psychiatric screening tool called the MDQ. There are three or four research teams in three or four institutions who are really into this and publish papers on it a lot. Occasionally someone from another part of psychiatry wanders in, but usually it’s just the subsubsubspeciality of MDQ researchers talking to each other. That’s fine. They’re our repository of specialized knowledge on this one screening tool.

You would *hope* the future of the human race would get a little bit more attention than one lousy psychiatric screening tool, but blah blah civilizational inadequacy, turns out not so much, they’re of about equal size. If there are only a couple of groups working on this problem, they’re going to look incestuous but that’s fine.

On the other hand, math is math, and if MIRI is trying to produce real mathematical results they ought to be sharing them with the broader mathematical community.

Robby protests that until very recently, MIRI *hasn’t* really been focusing on math. This is a very recent pivot. In April 2013, Luke wrote in his mini strategic plan:

> We were once doing three things — research, rationality training, and the Singularity Summit. Now we’re doing one thing: research. Rationality training was spun out to a separate organization, CFAR, and the Summit was acquired by Singularity University. We still co-produce the Singularity Summit with Singularity University, but this requires limited effort on our part.
> After dozens of hours of strategic planning in January–March 2013, and with input from 20+ external advisors, we’ve decided to (1) put less effort into public outreach,

> and to (2) shift our research priorities to Friendly AI math research.

In the full strategic plan for 2014, he repeated:

> Events since MIRI's April 2013 strategic plan have increased my confidence that we are "headed in the right direction." During the rest of 2014 we will continue to:
> – Decrease our public outreach efforts, leaving most of that work to FHI at Oxford, CSER at Cambridge, FLI at MIT, Stuart Russell at UC Berkeley, and others (e.g. James Barrat).
> – Finish a few pending "strategic research" projects, then decrease our efforts on that front, again leaving most of that work to FHI, plus CSER and FLI if they hire researchers, plus some others.
> – Increase our investment in our Friendly AI (FAI) technical research agenda.
> – We've heard that as a result of…outreach success, and also because of Stuart Russell's discussions with researchers at AI conferences, AI researchers are beginning to ask, "Okay, this looks important, but what is the technical research agenda? What could my students and I do about it?" Basically, they want to see an FAI technical agenda, and MIRI is is developing that technical agenda already.

In other words, there is a recent pivot from outreach, rationality and strategic research to pure math research, and the pivot is only recently finished or still going on.

TL;DR, again in three points:

C1. Until recently, MIRI focused on outreach and did a truly excellent job on this. They deserve credit here.

C2. MIRI has a number of prestigious computer scientists and AI experts willing to endorse or affiliate with it in some way. While their own researchers are not *quite* at the same lofty heights, they include many people who have or are working on math or compsci PhDs.

C3. MIRI hasn't published much math because they were previously focusing on outreach and strategic research; they've only shifted to math work in the past year or so.

V.

The discussion just *kept going*. We reached about the limit of our disagreement on (C1), the point about outreach – yes, they've done it, but does it count when it doesn't bear fruit in published papers? About (C2) and the credentials of MIRI's team, Su3 kind of blended it into the next point about published papers, saying:

> Fundamental disconnect – I consider "working with MIRI" to mean "publishing results with them." As an outside observer, I have no indication that most of these people are working with them. I've been to workshops and conferences with Nobel prize winning physicists, but I've never "worked with them" in the academic sense of having a paper with them. If [someone like Stuart Russell] is interested in helping MIRI, the best thing he could do is publish a well received technical result in a good journal with Yudkowsky. That would help get researchers to pay actual attention(and give them one well received published result, in their operating history).
>
> Tangential aside- you overestimate the difficulty of getting top grad students to work for you. I recently got four CS grad students at a top program to help me with

> some contract work for a few days at the cost of some pizza and beer.

So it looks like it all comes down to the papers. Su3 had this to say:

> What I was specifically thinking was “MIRI has produced a much larger volume of well-received fan fiction and blog posts than research.” That was what I inended to communicate, if somewhat snarkily. MIRI bills itself as a research institute, so I judge them on their produced research. The accountability measure of a research institute is academic citations.
>
> Editorials by famous people have some impact with the general public, so thats fine for fundraising, but at some point you have to get researchers interested. You can measure how much influence they have on researchers by seeing who those researchers cite and what they work on. You could have every famous cosmologist in the world writing op-eds about AI risk, but its worthless if AI researchers don’t pay attention, and judging by citations, they aren’t.
>
> As a comparison for publication/citation counts, I know individual physicists who have published more peer reviewed papers since 2005 than all of MIRI has self-published to their website. My single most highly cited physics paper (and I left the field after graduate school) has more citations than everything MIRI has ever published in peer reviewed journals combined. This isn’t because I’m amazing, its because no one in academia is paying attention to MIRI.
>
> [Christiano et al’s result about Lob] has been self-published on their website. It has NOT been peer

> reviewed. So it's published in the sense of "you can go look at the paper." But its not published in the sense of "mathematicians in the same field have verified the result." I agree this one result looks interesting, but most mathematicians won't pay attention to it unless they get it reviewed (or at the bare minimum, clean it up and put it on Arxiv). They have lots of these self-published documents on their web page.
>
> If they are making a "strategic decision" to not submit their self-published findings to peer review ,they are making a terrible strategic decision, and they aren't going to get most academics to pay attention that way. The result of Christiano, et al. *is* potentially interesting, but it's languishing as a rough unpublished draft on the MIRI site, so its not picking up citations.
>
> I'd go further and say the lack of citations is my main point. Citations are the important measurement of "are researchers paying attention." If everything self-published to MIRI's website were sparking interest in academia, citations would be flying around, even if the papers weren't peer reviewed, and I'd say "yeah, these guys are producing important stuff."
>
> My subpoint might be that MIRI doesn't even seem to be trying to get citations/develop academic interest, as measured by how little effort seems to be put into publication.

And Su3's not buying the pivot explanation either:

> That seems to be a reframing of the past history though. I saw talks by the SIAI well before 2013 where they described their primary purpose as friendly AI research, and insisted they were in a unique position (due to

> being uniquely brilliant/rational) to develop technical friendly AI (as compared to academic AI researchers).
>
> [Tarn] and [Robby] have suggested the organization is undergoing a pivot, but they've always billed themselves as a research institute. But donating money to an organization that has been ineffective in the past, because it looks like they might be changing seems like a bad proposition.
>
> My initial impression (reading Muelhauser's post you linked to and a few others) is that Muelhauser noticed the house was out of order when he became director and is working to fix things. Maybe he'll succeed and in the future, then, I'll be able to judge MIRI as effective- certainly a disproportionate number of their successes have come in the last few years. However, right now all I have is their past history, which has been very unproductive.

VI.

After that, discussion stayed focused on the issue of citations. This seemed like progress to me. Not only had we gotten it down to a core objection, but it was sort of a factual problem. It wasn't an issue of praising or condemning. Here's an organization with a lot of smart people. We know they work very hard – no one's ever called Luke a slacker, and another MIRI staffer (who will not be named, for his own protection) achieved some level of infamy for mixing together a bunch of the strongest chemicals from my nootropics survey into little pills which he kept on his desk in the MIRI offices for anyone who wanted to work twenty hours straight and then probably die young of conditions previously unknown to science. IQ-point*hours is a weird metric, but MIRI is putting a lot of IQ-point*hours into whatever it’s doing. So if Su3’s right that there are missing citations, where are they?

Among the three of us, Robby and Tarn and I generated a couple of hypotheses (well, Robby’s were more like facts than hypotheses, since he’s the only one in this conversation who actually works there).

D1: MIRI has always been doing research, but until now it’s been strategic research (ie “How worried should we be about AI?”, “How far in the future should we expect AI to be developed?”) which hasn’t fit neatly into an academic field or been of much interest to anyone except MIRI allies like Bostrom. They have dutifully published this in the few papers that are interested, and it has dutifully been cited by the few people who are interested (ie Bostrom). It’s unreasonable to expect Stuart Russell to cite their estimates of time course for superintelligence when he’s writing his papers on technical details of machine learning algorithms or whatever it is he writes papers on. And we can generalize from Stuart Russell to the rest of the AI field, who are *also* writing on things like technical details of machine learning algorithms that can’t plausibly be connected to when machines will become superintelligent.

D2: As above, but continuing to apply even in some of their math-ier research. MIRI does have lots of internal technical papers on their website. People tend to cite other researchers working in the same field as themselves. I could write the best psychiatry paper in human history, and I’m probably not going to get any citations from astrophysicists. But “machine ethics” is an entirely new field that’s not super relevant to anyone else’s work. Although a couple key machine ethics problems, like the Lobian obstacle and decision theory, touch on bigger and better-populated subfields of mathematics, they’re always going to be outsiders who happen to wander in. It’s unfair to compare them to a physics grad student writing about quarks or something, because she has the benefit of decades of previous work on quarks and a large and very interested research community. MIRI’s first job is to *create* that field and community, which until you succeed looks a lot like “outreach”.

D3: Lack of staffing and constant distraction by other important problems. This is Robby’s description of what he notices from the inside. He writes:

> We’re short on staff, especially since Louie left. Lots of people are willing to volunteer for MIRI, but it’s hard to find the right people to recruit for the long haul. Most relevantly, we have two new researchers (Nate and Benja), but we’d love a full-time Science Writer to specialize in taking our researchers’ results and turning them into publishable papers. Then we don’t have to split as much researcher time between cutting-edge work and explaining/writing-down.
>
> A lot of the best people who are willing to help us are very busy. I’m mainly thinking of Paul Christiano. he’s working actively on creating a publishable version of the probabilistic Tarski stuff, but it’s a really big endeavor. Eliezer is by far our best FAI researcher, and he’s very slow at writing formal, technical stuff. He’s generally low-stamina and lacks experience in writing in academic style / optimizing for publishability, though I believe we’ve been having a math professor tutor him to get over that particular hump. Nate and Benja are new, and it will take time to train them and get them publishing their own stuff. At the moment, Nate/Benja/Eliezer are spending the rest of 2014 working on material for the

> FLI AI conference, and on introductory FAI material to send to Stuart Russell and other bigwigs.

D4: Some of the old New York rationalist group takes a more combative approach. I'm not sure I can summarize their argument well enough to do it justice, so I would suggest reading Alyssa's post on her own blog.

But if I have to take a stab: everyone knows mainstream academia is way too focused on the "publish or perish" ethic of measuring productivity in papers or citations rather than real progress. Yeah, a similar-sized research institute in physics could probably get ten times more papers/citations than MIRI. That's because they're *optimizing* for papers/citations rather than advancing the field, and Goodhart's Law is in effect here as much as everywhere else. Those other institutes probably got geniuses who should be discovering the cure for cancer spending half their time typing, formatting, submitting, resubmitting, writing whatever the editors want to see, et cetera. MIRI is blessed with enough outside support that it doesn't *have* to do that. The only reason to try is to get prestige and attention, and anyone who's not paying attention now is more likely to be a constitutional skeptic using lack of citations as an excuse, than a person who would genuinely change their mind if there were more citations.

I am more sympathetic than usual to this argument because I'm in the middle of my own research on psychiatric screening tools and quickly learning that *official, published research is the worst thing in the world*. I could do my study in about two hours if the only work involved were doing the study; instead it's week after week of forms, IRB submissions, IRB revisions, required online courses where I learn the Nazis did unethical research and this was bad so I should try not to be a Nazi, selecting exactly which journals

I'm aiming for, and figuring out which of my bosses and co-workers academic politics requires me make co-authors. It is a *crappy game*, and if you've been blessed with enough independence to avoid playing it, why *wouldn't* you take advantage? Forget the overhyped and tortured "measure" of progress you use to impress other people, and just make the progress.

VII.

Or not. I'll let Su3 have the last word:

> I think something fundamental about my argument has been missed, perhaps I've communicated it poorly.
>
> It seems like you think the argument is that increasing publications increases prestige/status which would make researchers pay attention. i.e. publications -> citations -> prestige -> people pay attention. This is not my argument.
>
> My argument is essentially that the way to judge if MIRI's outreach has been successful is through citations, not through famous people name dropping them, or allowing them to be figure heads.
>
> This is because I believe the goal of outreach is get AI researchers focused on MIRI's ideas. Op eds from famous people are useful only if they get AI researchers focused on these ideas. Citations aren't about prestige in this case- citations tell you which researchers are paying attention to you. The number of active researchers paying attention to MIRI is very small. We know this because citations are an easy to find, direct measure.

Not all important papers have tremendous numbers of citations, but a paper can’t become important if it only has 1 or 2, because the ultimate measure of importance is “are people using these ideas?”

So again, to reiterate, if the goal of outreach is to get active AI researchers paying attention, then the direct measure for who is paying attention is citations. [But] the citation count on MIRIs work is very low. Not only is the citation count low (i.e. no researchers are paying attention), MIRI doesn’t seem to be trying to boost it – it isn’t trying to publish which would help get its ideas attention. I’m not necessarily dismissive of celebrity endorsements or popular books, my point is why should I measure the means when I can directly measure the ends?

The same idea undercuts your point that “lots of impressive PhD students work and have worked with MIRI,” because it’s impossible to tell if you don’t personally know the researchers. This is because they don’t create much output while at MIRI, and they don’t seem to be citing MIRI in their work outside of MIRI.

[Even people within the rationalist/EA community] agree with me somewhat. Here is a relevant quote from Holden Karnofsky [of GiveWell]:

SI seeks to build FAI and/or to develop and promote “Friendliness theory” that can be useful to others in building FAI. Yet it seems that most of its time goes to activities other than developing AI or theory. Its per-person output in terms of publications seems low. Its core staff seem more focused on Less Wrong posts, “rationality training” and other activities that don’t seem connected to the core goals; Eliezer Yudkowsky, in

> particular, appears (from the strategic plan) to be focused on writing books for popular consumption. These activities seem neither to be advancing the state of FAI-related theory nor to be engaging the sort of people most likely to be crucial for building AGI.
>
> And here is a statement from Paul Christiano disagreeing with MIRI's core ideas:
>
> But I should clarify that many of MIRI's activities are motivated by views with which I disagree strongly and that I should categorically not be read as endorsing the views associated with MIRI in general or of Eliezer in particular. For example, I think it is very unlikely that there will be rapid, discontinuous, and unanticipated developments in AI that catapult it to superhuman levels, and I don't think that MIRI is substantially better prepared to address potential technical difficulties than the mainstream AI researchers of the future.

This time Su3 helpfully provides their own summary:

E1. If the goal of outreach is to get active AI researchers paying attention, then the direct measure for who is paying attention is citations. [But] the citation count on MIRIs work is very low.

E2. Not only is the citation count low (i.e. no researchers are paying attention), MIRI doesn't seem to be trying to boost it – it isn't trying to publish which would help get its ideas attention. I'm not necessarily dismissive of celebrity endorsements or popular books, my point is why should I measure the means when I can directly measure the ends?

E3. The same idea undercuts your point that "lots of impressive phd students work and have worked with MIRI," because its impossible to tell if you don't personally know the researchers. This is because they don’t create much output while at MIRI, and they don’t seem to be citing MIRI in their work outside of MIRI.

E4. Holden Karnofsky and Paul Christiano do not believe that MIRI is better prepared to address the friendly AI problem than mainstream AI researchers of the future. Karnofsky explicitly for some of the reasons I have brought up, Christiano for reasons unmentioned.

VIII.

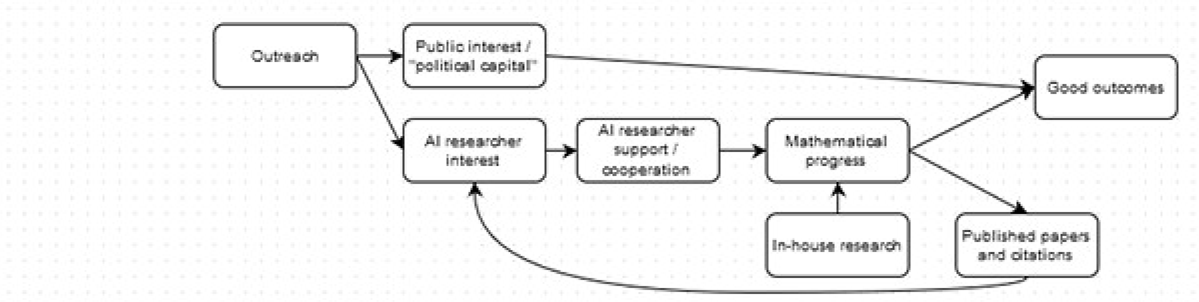
Didn’t actually read all that and just skipped down to the last subheading to see if there’s going to be a summary and conclusion and maybe some pictures? Good.

There seems to be some agreement MIRI has done a good job bringing issues of AI risk into the public eye and getting them media attention and the attention of various public intellectuals. There is disagreement over whether they should be credited for their success in this area, or whether this is a first step they failed to follow up on.

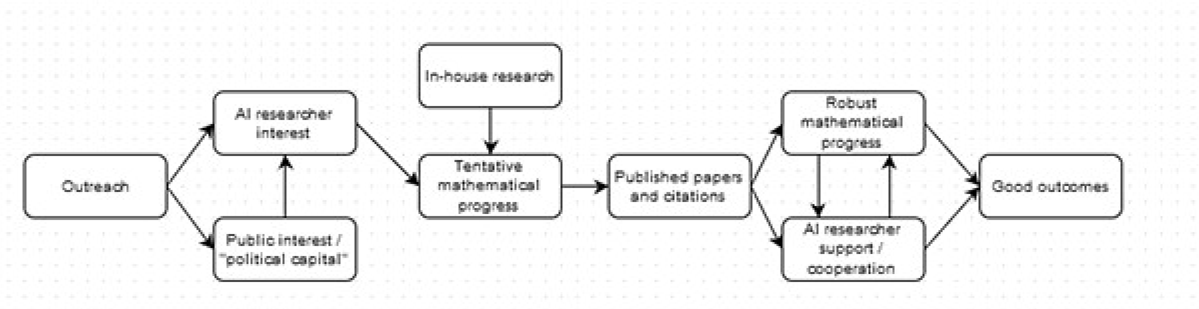
There also seems to be some agreement MIRI has done a poor job getting published and cited results in journals. There is disagreement over whether this is an understandable consequence of being a small organization in a new field that wasn’t even focusing on this until recently, or whether it represents a failure at exactly the sort of task by which their success should be judged.

This is probably among the 100% of issues that could be improved with flowcharts:

Optimistic Model


Outreach
Public interest / "political capital"
Good outcomes
AI researcher interest
AI researcher support / cooperation
Mathematical progress
In-house research
Published papers and citations

Pessimistic Model


In-house research
AI researcher interest
Outreach
Public interest / "political capital"
Tentative mathematical progress
Published papers and citations
Robust mathematical progress
AI researcher support / cooperation
Good outcomes

In the Optimistic Model, MIRI’s successfully built up Public Interest, and for all we know they might have Mathematical Progress as well even though they haven’t published it in journals yet. While they could feed back their advantages by turning their progress into Published Papers and Citations to get even more Mathematical Progress, overall they’re in pretty good shape for producing Good Outcomes, at least insofar as this is possible in their chosen field.

In the Pessimistic Model, MIRI may or may not have garnered Public Interest, Researcher Interest, and Tentative Mathematical Progress, but they failed to turn that into Published Papers and Citations, which is the only way they’re going to get to Robust Mathematical Progress, Researcher Support, and eventually Good Outcomes. The best that can be said about them is that they set some very preliminary groundwork that they totally failed to follow up on.

A higher level point – if we accept the Pessimistic Model, do we accuse MIRI of being hopelessly incompetent, in which case they deserve less support? Or do we accept them as inexperienced amateurs who are the only people willing to try something difficult but necessary, in which case they deserve more support, and maybe some guidance, and perhaps some gentle or not-so-gentle prodding? Maybe if you're a qualified science writer you could apply for the job opening they're advertising and help them get those papers they need?

An even higher-level point – what do people worried about AI risk do with this information? I don't see much that changes my opinion of the organization one way or the other. But Robby points out that people who are more concerned – but still worried about AI risk – have other good options. The Future of Humanity Institute at Oxford research that is less technical and more philosophical, wears their strategic planning emphasis openly on their sleeve has oodles of papers and citations and prestige. They also accept donations.

Best of all, *their* founder doesn't write any fanfic at all. Just perfectly respectable stories about evil dragon kings.

Five Planets In Search Of A Sci-Fi Story

Gamma Andromeda, where philosophical stoicism went too far. Its inhabitants, tired of the roller coaster ride of daily existence, decided to learn equanimity in the face of gain or misfortune, neither dreading disaster nor taking joy in success.

But that turned out to be really hard, so instead they just hacked it. Whenever something good happens, the Gammandromedans give themselves an electric shock proportional in strength to its goodness. Whenever something bad happens, the Gammandromedans take an opiate-like drug that directly stimulates the pleasure centers of their brain, in a dose proportional in strength to its badness.

As a result, every day on Gamma Andromeda is equally good compared to every other day, and its inhabitants need not be jostled about by fear or hope for the future.

This does sort of screw up their incentives to make good things happen, but luckily they’re all virtue ethicists.

Zyzzx Prime, inhabited by an alien race descended from a barnacle-like creature. Barnacles are famous for their two stage life-cycle: in the first, they are mobile and curious creatures, cleverly picking out the best spot to make their home. In the second, they root themselves to the spot and, having no further use for their brain, eat it.

This particular alien race has evolved far beyond that point and does not literally eat its brain. However, once an alien reaches sufficiently high social status, it releases a series of hormones that tell its brain, essentially, that it is now in a safe place and doesn't have to waste so much energy on thought and creativity to get ahead. As a result, its mental acuity drops two or three standard deviations.

The Zyzzxians' society is marked by a series of experiments with government – monarchy, democracy, dictatorship – only to discover that, whether chosen by succession, election, or ruthless conquest, its once brilliant leaders lose their genius immediately upon accession and do a terrible job. Their government is thus marked by a series of perpetual pointless revolutions.

At one point, a scientific effort was launched to discover the hormones responsible and whether it was possible to block them. Unfortunately, any scientist who showed promise soon lost their genius, and those promoted to be heads of research institutes became stumbling blocks who mismanaged funds and held back their less prestigious co-workers. Suggestions that the institutes eliminate tenure were vetoed by top officials, who said that "such a drastic step seems unnecessary".

K'th'ranga V, which has been a global theocracy for thousands of years, ever since its dominant race invented agricultural civilization. This worked out pretty well for a while, until it reached an age of industrialization, globalization, and scientific discovery. Scientists began to uncover truths that contradicted the Sacred Scriptures, and the hectic pace of modern life made the shepherds-and-desert-traders setting of the holy stories look vaguely silly. Worse, the cold logic of capitalism and utilitarianism began to invade the Scriptures' innocent Stone Age morality.

The priest-kings tried to turn back the tide of progress, but soon realized this was a losing game. Worse, in order to determine what to suppress, they themselves had to learn the dangerous information, and their mental purity was even more valuable than that of the populace at large.

So the priest-kings moved en masse to a big island, where they began living an old-timey Bronze Age lifestyle. And the world they ruled sent emissaries to the island, who interfaced with the priest-kings, and sought their guidance, and the priest-kings ruled a world they didn’t understand as best they could.

But it soon became clear that the system could not sustain itself indefinitely. For one thing, the priest-kings worried that discussion with the emissaries – who inevitably wanted to talk about strange things like budgets and interest rates and nuclear armaments – was contaminating their memetic purity. For another thing, they honestly couldn’t understand what the emissaries were talking about half the time.

Luckily, there was a whole chain of islands making an archipelago. So the priest-kings set up ten transitional societies – themselves in the Bronze Age, another in the Iron Age, another in the Classical Age, and so on to the mainland, who by this point were starting to experiment with nanotech. Mainland society brought its decisions to the first island, who translated it into their own slightly-less-advanced understanding, who brought it to the second island, and so on to the priest-kings, by which point a discussion about global warming might sound like whether we should propitiate the Coal Spirit. The priest-kings would send their decisions to the second-to-last island, and so on back to the mainland.

Eventually the Kth’ built an AI which achieved superintelligence and set out to conquer the universe. But it was a well-programmed superintelligence coded with Kth’ values. Whenever *it* wanted a high-level decision made, it would talk to a slightly less powerful superintelligence, who would talk to a slightly less powerful superintelligence, who would talk to the mainlanders, who would talk to the first island…

Chan X-3, notable for a native species that evolved as fitness-maximizers, not adaptation-executors. Their explicit goal is to maximize the number of copies of their genes. But whatever genetic program they are executing doesn't care whether the genes are within a living being capable of expressing them or not. The planet is covered with giant vats full of frozen DNA. There was originally some worry that the species would go extinct, since having children would consume resources that could be used hiring geneticists to make millions of copies of your DNA and stores them in freezers. Luckily, it was realized that children not only provide a useful way to continue the work of copying and storing (half of) your DNA long into the future, but will also work to guard your already-stored DNA against being destroyed. The species has thus continued undiminished, somehow, and their fondest hope is to colonize space and reach the frozen Kuiper Belt objects where their DNA will naturally stay undegraded for all time.

New Capricorn, which contains a previously undiscovered human colony that has achieved a research breakthrough beyond their wildest hopes. A multi-century effort paid off in a fully general cure for death. However, the drug fails to stop aging. Although the Capricornis no longer need fear the grave, after age 100 or so even the hardiest of them get Alzheimers' or other similar conditions. A hundred years after the breakthrough, more than half of the population is elderly and demented. Two hundred years after, more than 80% are. Capricorni nursing homes quickly became overcrowded and unpleasant, to the dismay of citizens expecting to spend eternity there.

So another research program was started, and the result were fully immersive, fully life-supporting virtual reality capsules. Stacked in huge warehouses by the millions, the elderly sit in their virtual worlds, vague sunny fields and old gabled houses where it is always the Good Old Days and their grandchildren are always visiting.

Five Case Studies On Politicization

[Trigger warning: Some discussion of rape in Part III. This will make much more sense if you've previously read I Can Tolerate Anything Except The Outgroup]

I.

One day I woke up and they had politicized Ebola.

I don't just mean the usual crop of articles like Republicans Are Responsible For The Ebola Crisis and Democrats Try To Deflect Blame For Ebola Outbreak and Incredibly Awful Democrats Try To Blame Ebola On GOP and NPR Reporter Exposes Right Wing Ebola Hype and Republicans Flip-Flop On Ebola Czars. That level of politicization was pretty much what I *expected*.

(I can't say I *totally* expected to see an article called Fat Lesbians Got All The Ebola Dollars, But Blame The GOP, but in retrospect nothing I know about modern society suggested I wouldn't)

I'm talking about something weirder. Over the past few days, my friends on Facebook have been making impassioned posts about how it's *obvious* there should/shouldn't be a quarantine, but deluded people on the other side are muddying the issue. The issue has risen to an alarmingly high level of 0.05 #Gamergates, which is my current unit of how much people on social media are concerned about a topic. What's more, everyone supporting the quarantine has been on the right, and everyone opposing on the left. *Weird* that so many people suddenly develop strong feelings about a complicated epidemiological issue, which can be exactly predicted by their feelings about *everything else*.

On the Right, there is condemnation of the CDC's opposition to quarantines as globalist gibberish, fourteen questions that will never be asked about Ebola centering on why there aren't more quarantine measures in place, and arguments on right-leaning biology blogs for why the people opposing quarantines are dishonest or incompetent. Top Republicans call for travel bans and a presenter on Fox, proportionate as always, demands quarantine centers in every US city.

On the Left (and token libertarian) sides, the New Yorker has been publishing articles on how involuntary quarantines violate civil liberties and "embody class and racial biases", Reason makes fun of "dumb Republican calls for a travel ban", Vox has a clickbaity article on how "This One Paragraph Perfectly Sums Up America's Overreaction To Ebola", and MSNBC notes that to talk about travel bans is "borderline racism".

How did this happen? How did both major political tribes decide, within a month of the virus becoming widely known in the States, not only exactly what their position should be but what insults they should call the other tribe for not agreeing with their position? There are a lot of complicated and well-funded programs in West Africa to disseminate information about the symptoms of Ebola in West Africa, and all I can think of right now is that if the Africans could disseminate useful medical information half as quickly as Americans seem to have disseminated tribal-affiliation-related information, the epidemic would be over tomorrow.

Is it just random? A couple of Republicans were coincidentally the first people to support a quarantine, so other Republicans felt they had to stand by them, and then Democrats felt they had to oppose it, and then that spread to wider and wider circles? And if by chance a Democrats had proposed quarantine before a Republican, the situation would have reversed itself? Could be.

Much more interesting is the theory that the fear of disease is the root of all conservativism. I am not making this up. There has been a lot of really good evolutionary psychology done on the extent to which pathogen stress influences political opinions. Some of this is done on the societal level, and finds that societies with higher germ loads are more authoritarian and conservative. This research can be followed arbitrarily far – like, isn't it *interesting* that the most liberal societies in the world are the Scandinavian countries in the very far north where disease burden is low, and the most traditionalist-authoritarian ones usually in Africa or somewhere where disease burden is high? One even sees a similar effect within countries, with northern US states being very liberal and southern states being very conservative. Other studies have instead focused on differences between individuals within society – we know that religious conservatives are people with stronger disgust reactions and priming disgust reactions can increase self-reported conservative political beliefs – with most people agreeing disgust reactions are a measure of the "behavioral immune system" triggered by fear of germ contamination.

(free tip for liberal political activists – offering to tidy up voting booths before the election is probably a thousand times more effective than anything you're doing right now. I will leave the free tip for conservative political activists to your imagination)

If being a conservative means you're pre-selected for worry about disease, obviously the conservatives are going to be the ones most worried about Ebola. And in fact, along with the quarantine debate, there's a little sub-debate about whether Ebola is worth panicking about. Vox declares

Americans to be “overreacting” and keeps telling them to calm down, whereas its similarly-named evil twin Vox Day has been spending the last week or so spreading panic and suggesting readers “wash your hands, stock up a bit, and avoid any unnecessary travel”.

So that’s the second theory.

The third theory is that everything in politics is mutually reinforcing.

Suppose the Red Tribe has a Grand Narrative. The Narrative is something like *“We Americans are right-thinking folks with a perfectly nice culture. But there are also scary foreigners who hate our freedom and wish us ill. Unfortunately, there are also traitors in our ranks – in the form of the Blue Tribe – who in order to signal sophistication support foreigners over Americans and want to undermine our culture. They do this by supporting immigration, accusing anyone who is too pro-American and insufficiently pro-foreigner of “racism”, and demanding everyone conform to “multiculturalism” and “diversity”, as well as lionizing any group within America that tries to subvert the values of the dominant culture. Our goal is to minimize the subversive power of the Blue Tribe at home, then maintain isolation from foreigners abroad, enforced by a strong military if they refuse to stay isolated.”*

And the Blue Tribe also has a Grand Narrative. The Narrative is something like *“The world is made up of a bunch of different groups and cultures. The wealthier and more privileged groups, played by the Red Tribe, have a history of trying to oppress and harass all the other groups. This oppression is based on ignorance, bigotry, xenophobia, denial of science, and a false facade of patriotism. Our goal is to call out the Red Tribe on its many flaws, and support* *other groups like foreigners and minorities in their quest for justice and equality, probably in a way that involves lots of NGOs and activists.”*

The proposition “a quarantine is the best way to deal with Ebola” seems to fit much better into the Red narrative than the Blue Narrative. It’s about foreigners being scary and dangerous, and a strong coordinated response being necessary to protect right-thinking Americans from them. When people like NBC and the New Yorker accuse quarantine opponents of being “racist”, that just makes the pieces fit in all the better.

The proposition “a quarantine is a bad way to deal with Ebola” seems to fit much better into the Blue narrative than the Red. It’s about extremely poor black foreigners dying, and white Americans rushing to throw them overboard to protect themselves out of ignorance of the science (which says Ebola can’t spread much in the First World), bigotry, xenophobia, and fear. The *real* solution is a coordinated response by lots of government agencies working in tandem with NGOs and local activists.

It would be really hard to switch these two positions around. If the Republicans were to oppose a quarantine, it might raise the *general question* of whether closing the borders and being scared of foreign threats is always a good idea, and whether maybe sometimes accusations of racism are making a good point. Far “better” to maintain a consistent position where all your beliefs reinforce all of your other beliefs.

There’s a question of causal structure here. Do Republicans believe certain other things for their own sake, and then adapt their beliefs about Ebola to help buttress their other beliefs? Or do the same factors that made them adopt their narrative in the first place lead them to adopt a similar narrative around Ebola?

My guess it it's a little of both. And then once there's a critical mass of anti-quarantiners within a party, in-group cohesion and identification effects cascade towards it being a badge of party membership and everybody having to believe it. And if the Democrats are on the other side, saying things you disagree with about every *other* issue, and also saying that you have to oppose quarantine or else you're a bad person, then that also incentivizes you to support a quarantine, *just to piss them off*.

II.

Sometimes politicization isn't about what side you take, it's about what issues you emphasize.

In the last post, I wrote:

> Imagine hearing that a liberal talk show host and comedian was so enraged by the actions of ISIS that he'd recorded and posted a video in which he shouts at them for ten minutes, cursing the "fanatical terrorists" and calling them "utter savages" with "savage values".
>
> If I heard that, I'd be kind of surprised. It doesn't fit my model of what liberal talk show hosts do.
>
> But the story I'm actually referring to is liberal talk show host / comedian Russell Brand making that same rant against Fox News for supporting war against the Islamic State, adding at the end that "Fox is worse than ISIS".
>
> That fits my model perfectly. You wouldn't celebrate Osama's death, only Thatcher's. And you wouldn't call ISIS savages, only Fox News. Fox is the outgroup, ISIS is

> just some random people off in a desert. You hate the outgroup, you don't hate random desert people.
>
> I would go further. Not only does Brand not feel much like hating ISIS, he has a strong incentive not to. That incentive is: the Red Tribe is known to hate ISIS loudly and conspicuously. Hating ISIS would signal Red Tribe membership, would be the equivalent of going into Crips territory with a big Bloods gang sign tattooed on your shoulder.

Now I think I missed an important part of the picture. The existence of ISIS plays right into Red Tribe narratives. They are *totally* scary foreigners who hate our freedom and want to hurt us and probably require a strong military response, so their existence sounds like a point in favor of the Red Tribe. Thus, the Red Tribe wants to talk about them as much as possible and condemn them in the strongest terms they can.

There's not really any way to spin this issue in favor of the Blue Tribe narrative. The Blue Tribe just has to grudgingly admit that maybe this is one of the few cases where their narrative breaks down. So their incentive is to try to minimize ISIS, to admit it exists and is bad and try to distract the conversation to other issues that support their chosen narrative more. That's why you'll never see the Blue Tribe gleefully cheering someone on as they call ISIS "savages". It wouldn't fit the script.

But did you hear about that time when a Muslim-American lambasting Islamophobia totally pwned all of those ignorant FOX anchors? Le-GEN-dary!

III.

At worst this choice to emphasize different issues descends into an unhappy combination of tragedy and farce.

The Rotherham scandal was an incident in an English town where criminal gangs had been grooming and blackmailing thousands of young girls, then using them as sex slaves. This had been going on for at least ten years with minimal intervention by the police. An investigation was duly launched, which discovered that the police had been keeping quiet about the problem because the gangs were mostly Pakistani and the victims mostly white, and the police didn't want to seem racist by cracking down too heavily. Researchers and officials who demanded that the abuse should be publicized or fought more vigorously were ordered to attend "diversity training" to learn why their demands were offensive. The police department couldn't keep it under wraps forever, and eventually it broke and was a huge scandal.

The Left then proceeded to totally ignore it, and the Right proceeded to *never shut up* about it for like an entire month, and every article about it had to include the "diversity training" aspect, so that if you type "rotherham d…" into Google, your two first options are "Rotherham Daily Mail" and "Rotherham diversity training".

I don't find this surprising at all. The Rotherham incident ties in *perfectly* to the Red Tribe narrative – scary foreigners trying to hurt us, politically correct traitors trying to prevent us from noticing. It doesn't do *anything* for the Blue Tribe narrative, and indeed actively contradicts it at some points. So the Red Tribe wants to trumpet it to the world, and the Blue Tribe wants to stay quiet and distract.

HBD Chick usually writes very well-thought-out articles on race and genetics listing all the excellent reasons you should not marry your cousins. Hers is not a political blog, and I have never seen her get upset about any political issue before, but since most of her posts are about race and genetics she gets a lot of love from the Right and a lot of flak from the Left. She recently broke her silence on politics to write three long and very angry blog posts on the Rotherham issue, of which I will excerpt one:

> if you've EVER called somebody a racist just because they said something politically incorrect, then you'd better bloody well read this report, because THIS IS ON YOU! this is YOUR doing! this is where your scare tactics have gotten us: over 1400 vulnerable kids systematically abused because YOU feel uncomfortable when anybody brings up some "hate facts."
>
> this is YOUR fault, politically correct people — and i don't care if you're on the left or the right. YOU enabled this abuse thanks to the climate of fear you've created. thousands of abused girls — some of them maybe dead — on YOUR head.

I have no doubt that her outrage is genuine. But I do have to wonder why she is outraged about this and not all of the other outrageous things in the world. And I do have to wonder whether the perfect fit between her own problems – trying to blog about race and genetics but getting flak from politically correct people – and the problems that made Rotherham so disastrous – which include police getting flak from politically correct people – are part of her sudden conversion to political activism.

[edit: she objects to this characterization]

But I will also give her this – accidentally stumbling into being upset by the rape of thousands of children is, as far as accidental stumbles go, not a bad one. What’s everyone *else’s* excuse?

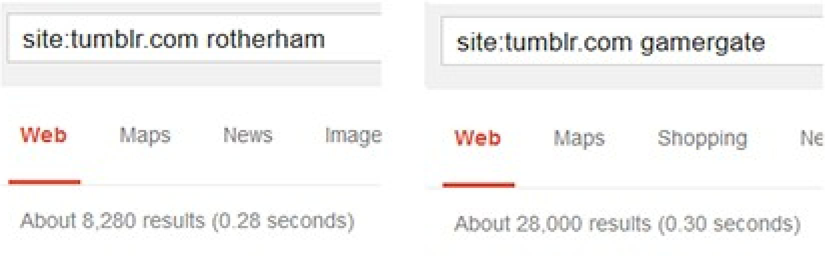
John Durant did an interesting analysis of media coverage of the Rotherham scandal versus the “someone posted nude pictures of Jennifer Lawrence” scandal.

He found left-leaning news website Slate had one story on the Rotherham child exploitation scandal, but four stories on nude Jennifer Lawrence.

He also found that feminist website Jezebel had only one story on the Rotherham child exploitation scandal, but six stories on nude Jennifer Lawrence.

Feministing gave Rotherham a one-sentence mention in a links roundup (just underneath “five hundred years of female portrait painting in three minutes”), but Jennifer Lawrence got two full stories.

The article didn’t talk about social media, and I couldn’t search it directly for Jennifer Lawrence stories because it was too hard to sort out discussion of the scandal from discussion of her as an actress. But using my current unit of social media saturation, Rotherham clocks in at 0.24 #Gamergates



site:tumblr.com rotherham

About 8,280 results (0.28 seconds)

site:tumblr.com gamergate

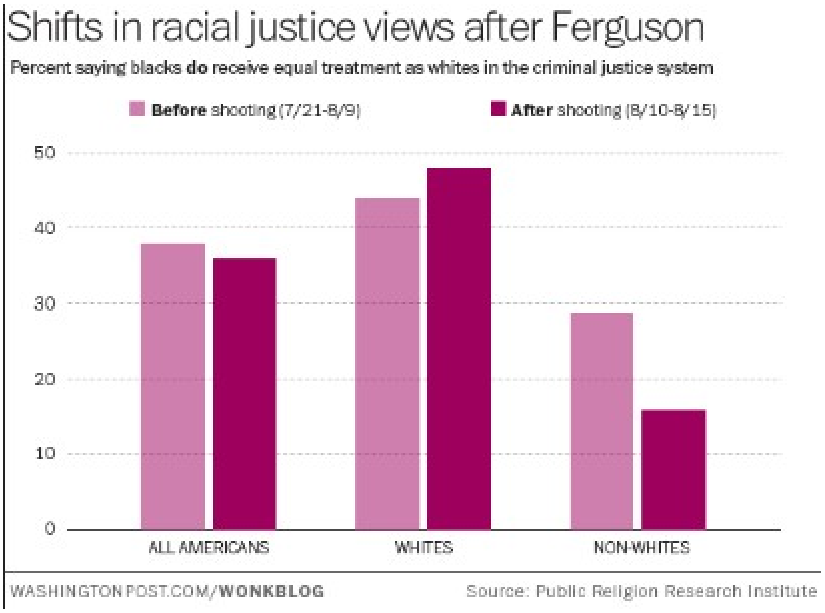
About 28,000 results (0.30 seconds)

You thought I was joking. I never joke.

This doesn’t surprise me much. Yes, you would think that the systematic rape of thousands of women with police taking no action might be a feminist issue. Or that it might outrage some people on Tumblr, a site which has many flaws but which has never been accused of being slow to outrage. But the goal here isn’t to push some kind of Platonic ideal of what’s important, it’s to support a certain narrative that ties into the Blue Tribe narrative. Rotherham does the opposite of that. The Jennifer Lawrence nudes, which center around how hackers (read: creepy internet nerds) shared nude pictures of a beloved celebrity on Reddit (read: creepy internet nerds) and 4Chan (read: creepy internet nerds) – and #Gamergate which does the same – are *exactly* the narrative they want to push, so they become the Stories Of The Century.

IV.

Here’s something I *did* find on Tumblr which I think is really interesting.


Shifts in racial justice views after Ferguson

Percent saying blacks **do** receive equal treatment as whites in the criminal justice system

WASHINGTONPOST.COM/**WONKBLOG**

Source: Public Religion Research Institute

You can see that after the Ferguson shooting, the average American became a little less likely to believe that blacks were treated equally in the criminal justice system. This makes sense, since the Ferguson shooting was a much-publicized example of the criminal justice system treating a black person unfairly.

But when you break the results down by race, a different picture emerges. White people were actually a little *more* likely to believe the justice system was fair after the shooting. Why? I mean, if there was no change, you could chalk it up to white people believing the police's story that the officer involved felt threatened and made a split-second bad decision that had nothing to do with race. That could explain no change just fine. But being *more* convinced that justice is color-blind? What could explain *that*?

My guess – before Ferguson, at least a few people interpreted this as an honest question about race and justice. After Ferguson, everyone mutually agreed it was about politics.

Ferguson and Rotherham were both similar in that they were cases of police misconduct involving race. You would think that there might be some police misconduct community who are interested in stories of police misconduct, or some race community interested in stories about race, and these people would discuss both of these two big international news items.

The Venn diagram of sources I saw covering these two stories forms two circles with no overlap. All those conservative news sites that couldn't shut up about Rotherham? Nothing on Ferguson – unless it was to snipe at the Left for “exploiting” it to make a political point. Otherwise, they did their best to stay quiet about it. Hey! Look over there! ISIS is probably beheading someone really interesting!

The same way Rotherham obviously supports the Red Tribe’s narrative, Ferguson obviously supports the Blue Tribe’s narrative. A white person, in the police force, shooting an innocent (ish) black person, and then a racist system refusing to listen to righteous protests by brave activists.

The “see, the Left is right about everything” angle of most of the coverage made HBD Chick’s attack on political correctness look subtle. The parts about race, systemic inequality, and the police were of debatable proportionality, but what I really liked was the Ferguson coverage started branching off into every issue any member of the Blue Tribe has ever cared about:

Gun control? Check.

The war on terror? Check.

American exceptionalism? Check.

Feminism? Check.

Abortion? Check

Gay rights? Check.

Palestinian independence? Check.

Global warming? Check. Wait, really? Yes, really.

Anyone who thought that the question in that poll was just a simple honest question about criminal justice was very quickly disabused of that notion. It was a giant Referendum On Everything, a “do you think the Blue Tribe is right on every issue and the Red Tribe is terrible and stupid, or vice versa?” And it turns out many people who when asked about criminal justice will just give the obvious answer, have much stronger and less predictable feelings about Giant Referenda On Everything.

In my last post, I wrote about how people feel when their in-group is threatened, even when it's threatened with an apparently innocuous point they totally agree with:

> I imagine [it] might feel like some liberal US Muslim leader, when he goes on the O'Reilly Show, and O'Reilly ambushes him and demands to know why he and other American Muslims haven't condemned beheadings by ISIS more, demands that he criticize them right there on live TV. And you can see the wheels in the Muslim leader's head turning, thinking something like "Okay, obviously beheadings are terrible and I hate them as much as anyone. But you don't care even the slightest bit about the victims of beheadings. You're just looking for a way to score points against me so you can embarass all Muslims. And I would rather personally behead every single person in the world than give a smug bigot like you a single microgram more stupid self-satisfaction than you've already got."

I think most people, when they think about it, probably believe that the US criminal justice system is biased. But when you feel under attack by people whom you suspect have dishonest intentions of twisting your words so they can use them to dehumanize your in-group, eventually you think "I would rather personally launch unjust prosecutions against every single minority in the world than give a smug out-group member like you a single microgram more stupid self-satisfaction than you've already got."

V.

Wait, so you mean turning all the most important topics in our society into wedge issues that we use to insult and abuse people we don't like, to the point where even mentioning it triggers them and makes them super defensive, might have been a *bad* idea??!

There's been some really neat research into people who don't believe in global warming. The original suspicion, at least from certain quarters, were that they were just dumb. Then someone checked and found that warming disbelievers actually had (very slightly) higher levels of scientific literacy than warming believers.

So people had to do actual studies, and to what should have been no one's surprise, the most important factor was partisan affiliation. For example, according to Pew 64% of Democrats believe the Earth is getting warmer due to human activity, compared to 9% of Tea Party Republicans.

So assuming you want to convince Republicans to start believing in global warming before we're all frying eggs on the sidewalk, how should you go about it? This is the excellent question asked by a study recently profiled in an NYMag article.

The study found that you could be a little more convincing to conservatives by acting on the purity/disgust axis of moral foundations theory – the one that probably gets people so worried about Ebola. A warmer climate is *unnatural*, in the same way that, oh, let's say, homosexuality is unnatural. Carbon dioxide *contaminating* our previously pure atmosphere, in the same way premarital sex or drug use contaminates your previously pure body. It sort of worked.

Another thing that sort of worked was *tying things into the Red Tribe narrative*, which they did through the two sentences “Being pro-environmental allows us to protect and preserve the American way of life. It is patriotic to conserve the country’s natural resources.” I can’t imagine anyone falling for this, but I guess some people did.

This is cute, but it’s too little too late. Global warming has already gotten inextricably tied up in the Blue Tribe narrative: *Global warming proves that unrestrained capitalism is destroying the planet. Global warming disproportionately affects poor countries and minorities. Global warming could have been prevented with multilateral action, but we were too dumb to participate because of stupid American cowboy diplomacy. Global warming is an important cause that activists and NGOs should be lauded for highlighting. Global warming shows that Republicans are science denialists and probably all creationists.* Two lousy sentences on “patriotism” aren’t going to break through that.

If *I* were in charge of convincing the Red Tribe to line up behind fighting global warming, here’s what I’d say:

> In the 1950s, brave American scientists shunned by the climate establishment of the day discovered that the Earth was warming as a result of greenhouse gas emissions, leading to potentially devastating natural disasters that could destroy American agriculture and flood American cities. As a result, the country mobilized against the threat. Strong government action by the Bush administration outlawed the worst of these gases, and brilliant entrepreneurs were able to discover and manufacture new cleaner energy sources. As a result of these brave decisions, our emissions stabilized and are currently declining.

Unfortunately, even as we do our part, the authoritarian governments of Russia and China continue to industralize and militarize rapidly as part of their bid to challenge American supremacy. As a result, Communist China is now by far the world's largest greenhouse gas producer, with the Russians close behind. Many analysts believe Putin secretly welcomes global warming as a way to gain access to frozen Siberian resources and weaken the more temperate United States at the same time. These countries blow off huge disgusting globs of toxic gas, which effortlessly cross American borders and disrupt the climate of the United States. Although we have asked them to stop several times, they refuse, perhaps egged on by major oil producers like Iran and Venezuela who have the most to gain by keeping the world dependent on the fossil fuels they produce and sell to prop up their dictatorships.

A giant poster of Mao looks approvingly at all the CO2 being produced...for Communism.

We need to take immediate action. While we cannot rule out the threat of military force, we should start by using our diplomatic muscle to push for firm action at top-level summits like the Kyoto Protocol. Second, we should fight back against the liberals who are trying to hold up this important work, from big government bureaucrats trying to regulate clean energy to celebrities accusing

> people who believe in global warming of being ‘racist’. Third, we need to continue working with American industries to set an example for the world by decreasing our own emissions in order to protect ourselves and our allies. Finally, we need to punish people and institutions who, instead of cleaning up their own carbon, try to parasitize off the rest of us and expect the federal government to do it for them.
>
> Please join our brave men and women in uniform in pushing for an end to climate change now.

If *this* were the narrative conservatives were seeing on TV and in the papers, I think we’d have action on the climate pretty quickly. I mean, that action might be nuking China. But it would be action.

And yes, there’s a sense in which that narrative is dishonest, or at least has really weird emphases. But our current narrative *also* has really some weird emphases. And for much the same reasons.

VI.

The Red Tribe and Blue Tribe have different narratives, which they use to tie together everything that happens into reasons why their tribe is good and the other tribe is bad.

Sometimes this results in them seizing upon different sides of an apparently nonpolitical issue when these support their narrative; for example, Republicans generally supporting a quarantine against Ebola, Democrats generally opposing it. Other times it results in a side trying to gain publicity for stories that support their narrative while sinking their opponents’ preferred stories – Rotherham for some Reds; Ferguson for some Blues.

When an issue gets tied into a political narrative, it stops being about itself and starts being about the wider conflict between tribes until eventually it becomes viewed as a Referendum On Everything. At this point, people who are clued in start suspecting nobody cares about the issue itself – like victims of beheadings, or victims of sexual abuse – and everybody cares about the issue's potential as a political weapon – like proving Muslims are "uncivilized", or proving political correctness is dangerous. After that, even people who agree that the issue is a problem and who would otherwise want to take action have to stay quiet, because they know that their help would be used less to solve a problem than to push forward the war effort against them. If they feel especially threatened, they may even take an unexpected side on the issue, switching from what they would usually believe to whichever position seems less like a transparent cover for attempts to attack them and their friends.

And then you end up doing silly things like saying ISIS is not as bad as Fox News, or donating hundreds of thousands of dollars to the officer who shot Michael Brown.

This can sort of be prevented by *not* turning everything into a referendum on how great your tribe is and how stupid the opposing tribe is, or by trying to frame an issue in a way that respects or appeals to an out-group's narrative.

Let me give an example. I find a lot of online feminism very triggering, because it seems to me to have nothing to do with women and be transparently about marginalizing nerdy men as creeps who are not really human (see: nude pictures vs. Rotherham, above). This means that even when I support and agree with feminists and want to help them, I am constantly trying to drag my brain out of panic mode that their seemingly valuable projects are just deep cover for attempts to hurt me (see: hypothetical Bill O'Reilly demanding Muslims condemn the "Islamic" practice of beheading people).

I have recently met some other feminists who instead use a narrative which views "nerds" as an "alternative gender performance", ie in the case of men they reject the usual masculine pursuits of sports and fraternities and they have characteristics that violate normative beauty standards (like "no neckbeards"). Thus, people trying to attack nerds is a subcategory of "people trying to enforce gender performance", and nerds should join with queer people, women, and other people who have an interest in promoting tolerance of alternative gender performances in order to fight for their mutual right to be left alone and accepted.

I'm not sure I entirely buy this argument, but it doesn't trigger me, and it's the sort of thing I *could* buy, and if all my friends started saying it I'd probably be roped into agreeing by social pressure alone.

But this is as rare as, well, anti-global warming arguments aimed at making Republicans feel comfortable and nonthreatened.

I blame the media, I really do. Remember, from within a system no one necessarily has an incentive to do what the system as a whole is supposed to do. Daily Kos or someone has a little label saying "supports liberal ideas", but *actually* their incentive is to make liberals want to click on their pages and ads. If the quickest way to do that is by writing story after satisfying story of how dumb Republicans are, and what wonderful taste they have for being members of the Blue Tribe instead of evil mutants, then they'll do that even if the effect on the entire system is to make

Republicans hate them and by extension everything they stand for.

I don’t know how to fix this.

In The Future, Everyone Will Be Famous To Fifteen People

[Epistemic status: not very serious]
[Content note: May make you feel overly scrutinized]

Sometimes I hear people talking about how nobody notices them or cares about anything they do. And I want to say… well…

Okay. The Survey of Earned Doctorates tells us that the United States awards about a hundred classics PhDs per year. I get the impression classics is more popular in Europe, so let's say a world total of five hundred. If the average classicist has a fifty year career, that's 25,000 classicists at any given time. Some classicists work on Rome, so let's say there are 10,000 classicists who focus solely on ancient Greece.

Estimates of the population of classical Greece center around a million people, but classical Greece lasted for several generations, so let's say there were ten million classical Greeks total. That gives us a classicist-to-Greek ratio of 1:1000.

It would seem that this ratio must be decreasing: world population increases, average world education level increases, but the number of classical Greeks is fixed for all time. Can we extrapolate to a future where there is a one-to-one ratio of classicists to classical Greeks, so that each scholar can study exactly one Greek?

Problem the first – human population is starting to stabilize, and will probably reach a maximum at less than double its current level. But this is a function of our location here on

Earth. Once we start colonizing space effectively, we can expect populations to balloon again. The *Journal of the British Interplanetary Society* estimates the carrying capacity of the solar system at forty trillion people; Nick Bostrom estimates the carrying capacity of the Virgo Supercluster at 10^23 human-like-digitized entities.

Problem the second – does the proportion of classics majors remain constant as population increases? One might expect that people living on domed cities in asteroids would have trouble being moved by the *Iliad*. Then again, one might expect that people living in glass-and-steel skyscrapers on a new continent ten thousand miles away from the classical world would have trouble being moved by the *Iliad*, and that didn't pan out. A better objection might be that as population increases, amount of history also increases – the year 2500 may have more historians than we do, but it also has five hundred years more history. But this decreases our estimates only slightly – population grows exponentially, but amount of history grows linearly. For example, the year 2000 has three times the population of the year 1900, but – if we start history from 4000 BC – only about two percent more history. Even if we admit the common sense idea that the 20th century contains "more" historical interest than, say, the 5th century, it still certainly does not contain three times as much historical interest as all previous centuries combined.

So it seems that if human progress continues, the number of classicists will equal, then exceed the number of inhabitants of classical Greece. Exactly when this happens depends on many things, most obviously the effects of any technological singularity that might occur. But if we want to be *very* naive about it and project Current Rate No Singularity indefinitely, we can just extend our current rate of population doubling every fifty years and suggest that in about 2500, with a human population of five trillion spread out throughout the solar system and maybe some nearby stars, we will reach classicist:Greek parity.

What will this look like? Barring any revolutionary advance in historical methodology, there won't really be enough artifacts and texts to support ten million classicists, so they will be reduced to overanalyzing excruciating details of the material that exists. On the other hand, maybe there *will* be revolutionary advances. The most revolutionary one I could think of would be the chronoscope from *The Light of Other Days*, a device often talked about in sci-fi stories that can see into the past. Armed with chronoscopes, classicists could avoid concentrating on a few surviving artifacts and study ancient Greece directly. And since the scholarly community would quickly exhaust would could be learned about important figures like Pericles and Leonidas, many historians would start looking into individual middle-class or lower-class Greeks, investigating their life stories and how they tied in to the broader historical picture. A new grad student might do her dissertation on the life of Nikias the random olive farmer who lived twenty miles outside Athens. Since there would be historian:subject parity, it might be that most or all ancient Greeks could be investigated in that level of detail.

What happens *after* 2500? If the assumptions mentioned above continue to hold, we pass parity and end up with more classicists than Greeks. By 3000 there are a thousand classicists for each ancient. Now you *wish* you could do your dissertation on the life of Nikias The Random Olive Farmer. But that low-hanging fruit (low hanging olive?) has been taken. Now there is an entire *field* (olive orchard?) of Nikias The Random Olive Farmer Studies, with its own little internal academic politics and yearly conferences on Alpha Centauri. In large symposia held at high-class hotels, various professors discuss details of Nikias The Random Olive Farmer’s psychology, personal relationships, opinions, and how he fits in to the major trends in Greek society that were going on at the time. Feminist scholars object that the field of Nikias The Random Olive Farmer’s Wife Studies is less well-funded than Nikias The Random Olive Farmer Studies, and dozens of angry papers are published in the relevant journals about it. Several leading figures object that too little effort is being made to communicate the findings of Nikias The Random Olive Farmer Studies to the general public, and there are half-hearted attempts to make little comic books about Nikias’ life or something.

By 3150 this has gotten so out of hand that it is wasting useful resources that should be allocated to fending off the K’th’rangan invasion. The Galactic Emperor declares a cap on the number of classics scholars at some reasonable number like a hundred million. There are protests in every major university, and leading public figures accuse the Galactic Emperor of being anti-intellectual, but eventually the new law takes hold and the grumbling dies down.

The field of Early 21st Century Studies, *on the other hand*, is still going strong. There are almost a thousand times as many moderns as Greeks, so we have a more reasonable ratio of about fifteen historians per modern, give or take, with the most interesting moderns having more and the ones who died young having fewer. Even better, the 21st Century Studies researchers don’t have to waste valuable chronoscopes that could be used for spying on the K’th’rangans. They can just hunt through the Internet Archive for more confusing, poorly organized data about the people of the early 21st century than they could ever want.

Gradually the data will start to make more and more sense. Imagine how excited the relevant portion of the scholarly community will be when it is discovered through diligent sleuthing that Thor41338 on the Gamer’s Guild forum is the same person as Hunter Glenderson from Charleston, South Carolina, and two seemingly different pieces of the early 21st century milieu slide neatly into place.

A few more population doublings, and the field of Hunter Glenderson From Charleston Studies is as big as the field of Nikias The Random Olive Farmer Studies ever was. The Galactic Emperor is starting to take notice, but the K’th’rangans are in retreat and for now there are resources to spare. There are no more great discoveries about new pseudonyms to be made, but there are still occasional paradigm shifts in analysis of the great events of Glenderson’s life. Someone tries a Freudian analysis of his life; another a Marxist analysis; a third writes about how his relationship with his ex-girlfriend from college ties in to the Daoist conception of impermanence. All these people have grad students trawling old Twitter accounts for them, undergraduates anxious to hear about their professor’s latest research, and hateblogs by brash amateurs claiming that the establishment *totally* misunderstands Hunter Glenderson.

Late at night, one grad student is preparing a paper on one of Glenderson’s teenaged Twitter rants, and comes across his tweet: “Nobody notices me. Nobody cares about anything I do.” She makes special note of it, since she thinks the irony factor might make it worth a publication in one of the several Hunter-Glenderson-themed journals.

Book Review: A Future For Socialism

A boot, stamping on a human face – forever!

No! Wait! Sorry! Wrong future for socialism! This is John Roemer's *A Future for Socialism*, a book on how to build a kinder, gentler socialist economy. It argues for – and I believe proves – a bold thesis: a socialist economy is entirely compatible with prosperity, innovation, and consumer satisfaction – just as long as by "socialism", you mean "capitalism".

The book makes proposals, but you're not exactly hearing the *Internationale* playing in the background as you read them. Prices are obviously the best form of allocating goods, so a socialist economy should keep them. Central planning could never work, so a socialist economy doesn't need it. Bosses and managers seem to be doing a good job keeping their firms profitable, so they can all keep their jobs under socialism. Everyone has different skills, so clearly in a truly socialist system they deserve different wages, in fact whatever wage the market will bear.

So where's the socialism? Well, socialism is a system where the people own the means of production. Right now corporations control the means of production, and you own corporations by holding stock in them. So if everybody owns stock in all corporations, then if you squint that's kind of socialist.

Roemer proposes the following: first, you nationalize large industries – or, if you're a post-Communist country (Roemer was writing in 1992) you start with your large industries already nationalized. Then you split them into stocks. Then you give everyone an equal amount of these stocks. When the corporations make money, they pay them out in the form of stock dividends, which go to the people/stockholders. So every year I get a check in the mail representing my one-three-hundred-millionth-part share of all the profits made by all the corporations in the United States.

Question: won't poor people immediately sell their shares to rich people, resulting in the rich people becoming wealthy means-of-production-owning capitalists again? (compare question 4.1 here). Answer: yes, obviously. So Roemer proposes a law that stocks cannot be sold for money, only coupons and other stocks. Every citizen is given an equal number of coupons at birth, trades them for stocks later on, and then trades those stocks for other stocks. This allows smart citizens to invest wisely, and allows a sort of "stock market" that sends the correct signals (this business's stock price is decreasing so maybe they're doing something wrong) but doesn't allow stock accumulation by wealthy capitalists.

In this system, businesses would raise funds not by selling stock but by seeking loans from banks. Apparently this is already how it works in Japan, where companies are arranged into "business groups" called keiretsu with each having a bank in charge of lending them money. Roemer hopes his model would work even better than the keiretsu, because the flow of stock coupons would give the banks market-driven information that help them make funding decisions.

So when I read about this I got really excited, because it sounded like a Basic Income Guarantee without the awkward questions about how to fund it. If everyone owns an equal number of shares in a diverse portfolio of the nation's companies, then corporate profits go to everyone in the form of a check in the mail. Sounds like a good way to help the poor.

Unfortunately, Roemer calculates in the back of the book how much money he expects people to get from such a scheme. Using equations I can't exactly follow, he finds that every citizen would get about $500, which is about 5% of the 1992 poverty rate. Using slightly different assumptions weighted in his favor, he is able to increase this to $1000 per person. It looks like we will not be solving poverty today.

On the other hand, I recently learned that corporate profits have been rising dramatically lately. If you do a calculation much simpler than Roemer's – in fact, pure division of the $2 trillion in national profits by the 300 million Americans who could receive them – you get about $6,000 per person. That's still not enough to reach the poverty line, but it's *something*, especially if you're willing to tax the wealthy's share to funnel it to the poor.

(on the other hand, maybe fewer than all corporations will be nationalized? I dunno here.)

But Roemer doesn't even mention this except as an aside, and doesn't think it's the most interesting thing about his system. What he's really interested in is finding a way to control what by analogy to public goods he calls "public bads". These are all the things that coordination problems form around, like pollution and global warming and selling weapons to dictatorial regimes and so on.

He makes the following fascinating claim: poor decision-making in the current system is driven by an imbalance of costs and benefits caused by the concentration of capital. For example, suppose that using lots of fossil fuel will produce $1 trillion in good economic activity, but also $10 trillion in costs due to climate change. The Koch brothers own lots of capital (in the form of stock, ownership rights of companies, et cetera) so much of that $1 trillion in economic benefits takes the form of increased corporate profits that go directly into their pockets. However, they only suffer the same share of global warming anyone else in the US suffers – presumably 1/300 millionth of the national cost. Therefore, since they get disproportionately large benefits but only proportionate costs, they have strong incentive to try to push fossil fuels. They are rich and powerful and usually get what they want, so probably fossil fuels will continue to be used.

But imagine that we socialized stock. Now everyone in the US gets 1/300 millionth of the national profits from good economic activity, *and* 1/300 millionth of the national costs of global warming. Since we already said the costs are greater than the benefits, every individual wants to fight global warming. People's incentives finally match reality.

This is a really pretty idea, but it doesn't seem quite right to me. By my understanding, very little lobbying is done by rich capitalists personally – and I think the Koch brothers are an exception because they genuinely hold conservative principles, not because they expect the calculus to come out in their favor. See Does Class Warfare Have A Free Rider Problem? Instead, lobbying is done by businesses directly, driven by the leadership of the businesses. Exxon Mobil hires oil lobbyists, Google hires intellectual property lobbyists, Monsanto hires agriculture lobbyists.

Would enraged Monsanto stockholder/citizens launch a corporate revolt demanding the company stop hiring lobbyists to work against the American people? I don't think so. Corporate revolts are really really hard even nowadays when most stocks are held by a few attention-paying competent rich people. Give them to millions of not-attention-paying mostly-incompetent hoi polloi, and you think they're going to be able to coordinate something? Besides, since stocks are tradeable, it might be only a few percent of the population who own Monsanto stock in particular; everyone else traded it away for more Google and Exxon Mobil stock. Those few percent of the population would get more money from Monsanto dividends than they would lose in the inevitable revolt of the Mutant Corn People, so their incentives would still be screwed.

So the Basic Income angle isn't really enough to be exciting, and I don't find the public goods/game theory angle too convincing either.

There's also a big set of questions the book leaves unanswered – how do companies get nationalized? How are new companies formed? What happens to them?

Roemer does agree that it would be hard to nationalize all companies in a large advanced nation like the United States. In particular, taking rich people's stock away from them without compensation would be naked theft, and the government probably couldn't afford the compensation necessary. So he suggests that something like this be tried first in the post-communist countries or some other nation that already has nationalized industries and wants to know what to do with them.

Fine. That leaves the other big question. Suppose that the US somehow nationalizes all its industries in 1992, and a few years later Page and Brin want to start Google. What happens? Does the government say "Oh, no, sorry, we already *have* companies, we don't need *more* of them"? Are they allowed to start it small, but the government immediately seizes it once it gets past a certain size? Are they just not allowed to sell it for stock and turn it into a corporation? Or if all of those things are okay and they can build Google as normal, what happens once most of the economy is made of these new post-1992 corporations and everything is capitalist again?

Overall there's nothing *terrible* about the system in *A Future For Socialism*. It sure beats Stalin and even Castro. It just seems like a lot of work for not necessarily very much gain.

The last chapter is the only one in which Roemer permits himself to wax rhapsodic into the optimism I normally associate with the socialist cause. He says that he hopes market socialism is just the beginning, that this system of universal stock ownership will cause people to stop promoting public bads and care about the general welfare of the country, and this will take the form of more investment in education to train the next generation of workers, and once everyone has access to good education everyone will be just about equal and able to earn just about equal wages in the free market and *then* all this social class nonsense will disappear. Man, people who wrote politics before we fully understood how genetics worked were *so cute*!

But despite my panning the economic proposals, I learned a lot from this book and am grateful to have read it.

First, I was impressed by the assumptions. Roemer starts by explaining that yes, he knows why capitalism is a good thing, it's reasons X Y and Z, and he's not going to challenge or ignore that. When I hear someone making a controversial claim I disagree with, my immediate instinct is to assume that person is ignorant. Roemer proves he isn't in precisely the right way. Before you advocate socialism, you prove that you're not just totally ignorant of capitalism; that simplifies the process of sorting out the people you can learn from from the people you can’t.

He also makes it clear that he’s not out to change human nature. He hopes human nature will eventually change (see above about education) but he also recognizes that has to track changes in society, not be the cause of them. He writes:

> [These proposals should take people as they actually are today, not as they might be after an egalitarian economic policy or cultural revolution has “remade” them. We must assume, as social scientists, that people are, in the short term at least, what they are: what can be changed – and slowly at that – are the institutions through which they interact.

Well put. Roemer establishes himself early on as someone who shares some of my basic assumptions (and can express them better than I can), which means even disagreement will be productive disagreement.

But second, and more important, this book is the first time I *really* had to think about joint-stock corporations. Like, I know what stock is in a “you buy it and then you get very excited or upset when it goes up or down” way, but I hadn’t thought of it as an important philosophical and political idea before, and Roemer really hammers home that it is.

The book identifies three big principal-agent problems in Soviet and other communist economic systems. First, managers employ workers to make their product, but workers want to slack off or line their own pockets. Second, central planners employ managers to run plants, but managers want to slack off or line their own pockets. Third, The People employ central planners to run the economy, but central planners want to slack off or line their own pockets. The Soviets solved these problems poorly. The central planners had no responsibility to anyone except other Party bureaucrats; the central planners could only punish managers who failed to meet their cooked-up metrics, leading to Goodhart's Law gone berserk. And managers sometimes couldn't promote workers in a meaningful way or fire them in a meaningful way, so workers had little incentive to do a good job.

The standard capitalist narrative is that principal-agent problems are very hard and maybe impossible on such a big scale, but this is okay, because in *capitalism* the people making the decisions are the ones profiting off them.

Roemer points out that's nonsense. Most real-world capitalism isn't the plucky entrepreneur founding a startup, it's the giant corporation, in which a bunch of investors (who profit off of good decisions) hire a manager or CEO type person (who is supposed to make good decisions). Insofar as CEOs keep companies profitable – and it seems they do – the principal-agent problem is *solved*. If we want the company to be run by Stalin instead of by investors, all we need to do is use current corporate governance structure, but give Stalin the stock, and the company will be just as profitable as ever (as long as Stalin doesn't try to interfere).

Roemer credits for this the hostile takeover method, where if a stock's price falls too low, that means some other group can buy out all the stock and fire the manager. It's a good point, but I can't help wondering if another part of it is immediate, hard-to-deny feedback: that is, the existence of the stock price at all. First of all, the CEO can't remain too deluded about her decisions; there has been many a politician who sends a country to its grave all the while hearing from a bunch of toadies that she's making things better, but stock prices are hard to fudge. Second of all, the investors and the Board of Directors and so on have a mechanism by which they can agree upon whether the company is doing well or not, short-circuiting some of the politics that might cause them to split into factions for or against the current leadership (this is not to say there are no corporate politics, just that they are more resolvable than other kinds of politics).

The principal-agent problem is at the center of a lot of different things, so it’s really interesting to think of something as humble and unassuming as the joint-stock corporation as having in some sense solved it. I’m not sure what the wider implications of this are, but the idea of futarchy is looking better and better.

So in summary, this book’s ideas on stock distribution seem potentially okay but probably not worth nationalizing all industries over, but the real gem is the lucid explanation of the importance of corporate governance.

Link: *A Future for Socialism*

EDIT: Tumblr user fadingphilosophymiracle points out that if private companies were added to the redistribution scheme, it would be enough for ~$12000/person, which sounds more promising.

Alcoholics Anonymous: Much More Than You Wanted To Know

[EDIT 10/27: Slight changes in response to feedback; correcting some definitions. I am not an expert in this field and will continue to make changes as I learn about them. There is a critique of this post here and other worse critiques elsewhere. My only excuse for doing this is that I am failing less spectacularly than other online sources writing about the same topic.]

I’ve worked with doctors who think Alcoholics Anonymous is so important for the treatment of alcoholism that anyone who refuses to go at least three times a week is in denial about their problem and can’t benefit from further treatment.

I’ve also worked with doctors who are so against the organization that they describe it as a “cult” and say that a physician who recommends it is no better than one who recommends crystal healing or dianetics.

I finally got so exasperated that I put on my Research Cap and started looking through the evidence base.

My conclusion, after several hours of study, is that now I understand why most people don’t do this.

The studies surrounding Alcoholics Anonymous are some of the most convoluted, hilariously screwed-up research I have ever seen. They go wrong in ways I didn’t even realize research *could* go wrong before. Just to give some examples:

– In several studies, subjects in the “not attending Alcoholics Anonymous” condition attended Alcoholics Anonymous more than subjects in the “attending Alcoholics Anonymous” condition.

– Almost everyone’s belief about AA’s retention rate is off by a factor of five because one person long ago misread a really confusing graph and everyone else copied them without double-checking.

– The largest study ever in the field, a $30 million effort over 8 years following thousands of patients, had no untreated control group.

Not only are the studies poor, but the people interpreting them are heavily politicized. The entire field of addiction medicine has gotten stuck in the middle of some of the most divisive issues in our culture, like whether addiction is a biological disease or a failure of willpower, whether problems should be solved by community and peer groups or by highly trained professionals, and whether there’s a role for appealing to a higher power in any public organization. AA’s supporters see it as a scruffy grassroots organization of real people willing to get their hands dirty, who can cure addicts failed time and time again by a system of glitzy rehabs run by arrogant doctors who think their medical degrees make them better than people who have personally fought their own battles. Opponents see it as this awful cult that doesn’t provide any real treatment and just tells addicts that they’re terrible people who will never get better unless they sacrifice their identity to the collective.

As a result, the few sparks of light the research kindles are ignored, taken out of context, or misinterpreted.

The entire situation is complicated by a bigger question. We will soon find that AA usually does not work better or worse than various other substance abuse interventions. That leaves the sort of question that all those fancy-shmancy people with control groups in their studies don’t have to worry about – does anything work at all?

I.

We can start by just taking a big survey of people in Alcoholics Anonymous and seeing how they’re doing. On the one hand, we don’t have a control group. On the other hand…well, there really is no other hand, but people keep doing it.

According to AA’s own surveys, one-third of new members drop out by the end of their first month, half by the end of their third month, and three-quarters by the end of their first year. “Drop out” means they don’t go to AA meetings anymore, which could be for any reason including (if we’re feeling optimistic) them being so completely cured they no longer feel they need it.

There is an alternate reference going around that only 5% (rather than 25%) of AA members remain after their first year. This is a mistake caused by misinterpreting a graph showing that only five percent of members in their first year were in their twelfth month of membership, which is obviously completely different. Nevertheless, a large number of AA hate sites (and large rehabs!) cite the incorrect interpretation, for example the Orange Papers and RationalWiki’s page on Alcoholics Anonymous. In fact, just to keep things short, assume RationalWiki’s AA page makes every single mistake I warn against in the rest of this article, then use that to judge them in general. On the other hand, Wikipedia gets it right and I continue to encourage everyone to use it as one of the most reliable sources of medical information available to the public (I wish I was joking).

This retention information isn't very helpful, since people can remain in AA without successfully quitting drinking, and people may successfully quit drinking without being in AA. However, various different sources suggest that, of people who stay in AA a reasonable amount of time, about half stop being alcoholic. These numbers can change wildly depending on how you define "reasonable amount of time" and "stop being alcoholic". Here is a table, which I have cited on this blog before and will probably cite again:

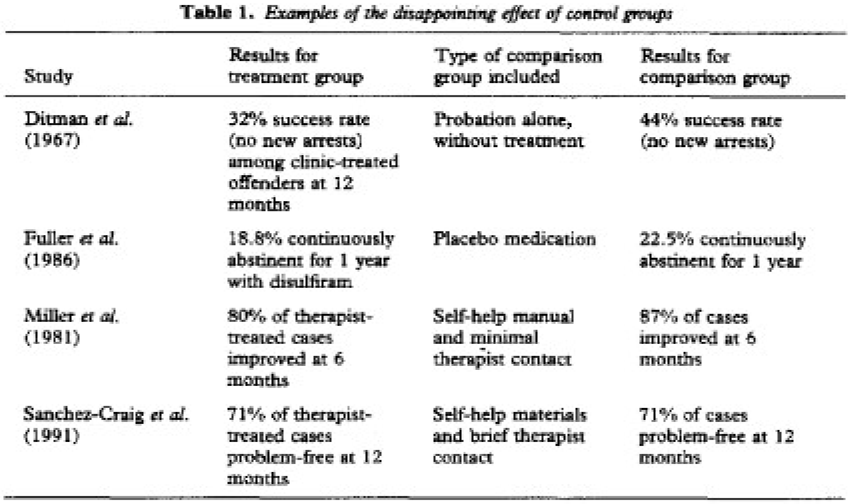


Table 1. *Examples of the disappointing effect of control groups*

Study	Results for treatment group	Type of comparison group included	Results for comparison group
Ditman *et al.* (1967)	32% success rate (no new arrests) among clinic-treated offenders at 12 months	Probation alone, without treatment	44% success rate (no new arrests)
Fuller *et al.* (1986)	18.8% continuously abstinent for 1 year with disulfiram	Placebo medication	22.5% continuously abstinent for 1 year
Miller *et al.* (1981)	80% of therapist-treated cases improved at 6 months	Self-help manual and minimal therapist contact	87% of cases improved at 6 months
Sanchez-Craig *et al.* (1991)	71% of therapist-treated cases problem-free at 12 months	Self-help materials and brief therapist contact	71% of cases problem-free at 12 months

Behold. Treatments that look very impressive (80% improved after six months!) turn out to be the same or worse as the control group. And comparing control group to control group, you can find that "no treatment" can appear to give wildly different outcomes (from 20% to 80% "recovery") depending on what population you're looking at and how you define "recovery".

Twenty years ago, it was extremely edgy and taboo for a reputable scientist to claim that alcoholics could recover on their own. This has given way to the current status quo, in which pretty much everyone in the field writes journal articles all the time about how alcoholics can recover on their own, but make sure to harp upon how edgy and taboo they are for doing so. From these sorts of articles, we learn that about 80% of recovered alcoholics have gotten better without treatment, and many of them are currently able to drink moderately without immediately relapsing (something *else* it used to be extremely taboo to mention). Kate recently shared an good article about this: Most People With Addiction Simply Grow Out Of It: Why Is This Widely Denied?

Anyway, all this stuff about not being able to compare different populations, and the possibility of spontaneous recovery, just mean that we need controlled experiments. The largest number of these take a group of alcoholics, follow them closely, and then evaluate all of them – the AA-attending and the non-AA-attending – according to the same criteria. For example Morgenstern et al (1997), Humphreys et al (1997) and Moos (2006). Emrick et al (1993) is a meta-analyses of *a hundred seventy three* of these. All of these find that the alcoholics who end up going to AA meetings are much more likely to get better than those who don't. So that's good evidence the group is effective, right?

Bzzzt! No! Wrong! Selection bias!

People who want to quit drinking are more likely to go to AA than people who don't want to quit drinking. People who want to quit drinking are more likely to *actually* quit drinking than those who don't want to. This is a *serious* problem. Imagine if it is common wisdom that AA is the best, maybe the only, way to quit drinking. Then 100% of people who really want to quit would attend compared to 0% of people who didn't want to quit. And suppose everyone who wants to quit succeeds, because secretly, quitting alcohol is really easy. Then 100% of AA members would quit, compared to 0% of non-members – the most striking result it is mathematically possible to have. And yet AA would not have made a smidgeon of difference.

But it’s worse than this, because attending AA isn’t just about wanting to quit. It’s also about having the resources to make it to AA. That is, wealthier people are more likely to hear about AA (better information networks, more likely to go to doctor or counselor who can recommend) and more likely to be able to attend AA (better access to transportation, more flexible job schedules). But wealthier people are also known to be better at quitting alcohol than poor people – either because the same positive personal qualities that helped them achieve success elsewhere help them in this battle as well, or just because they have fewer other stressors going on in their lives driving them to drink.

Finally, perseverance is a confounder. To go to AA, and to keep going for months and months, means you’ve got the willpower to drag yourself off the couch to do a potentially unpleasant thing. That’s probably the same willpower that helps you stay away from the bar.

And then there’s a confounder going the *opposite* direction. The worse your alcoholism is, the more likely you are to, as the organization itself puts it, “admit you have a problem”.

These sorts of longitudinal studies are almost useless and the field has mostly moved away from them. Nevertheless, if you look on the pro-AA sites, you will find them in droves, and all of them “prove” the organization’s effectiveness.

III.

It looks like we need randomized controlled trials. And we have them. Sort of.

Brandsma (1980) is the study beloved of the AA hate groups, since it purports to show that people in Alcoholics Anonymous not only don’t get better, but are *nine times* more likely to binge drink than people who don't go into AA at all.

There are a number of problems with this conclusion. First of all, if you actually look at the study, this is one of about fifty different findings. The other findings are things like "88% of treated subjects reported a reduction in drinking, compared to 50% of the untreated control group".

Second of all, the increased binge drinking was significant at the 6 month followup period. It was *not* significant at the end of treatment, the 3 month followup period, the 9 month followup period, or the 12 month followup period. Remember, taking a single followup result out of the context of the other followup results is a classic piece of Dark Side Statistics and will send you to Science Hell.

Of multiple different endpoints, Alcoholics Anonymous did better than no treatment on almost all of them. It did worse than other treatments on some of them (dropout rates, binge drinking, MMPI scale) and the same as other treatments on others (abstinent days, total abstinence).

If you are pro-AA, you can say "Brandsma study proves AA works!". If you are anti-AA, you can say "Brandsma study proves AA works worse than other treatments!", although in practice most of these people prefer to quote extremely selective endpoints out of context.

However, most of the patients in the Brandsma study were people convicted of alcohol-related crimes ordered to attend treatment as part of their sentence. Advocates of AA make a good point that this population might be a bad fit for AA. They may not feel any personal motivation to treatment, which might be okay if you're going to listen to a psychologist do therapy with you, but fatal for a *self*-help group. Since the whole point of AA is being in a community of like-minded individuals, if you don't actually feel any personal connection to the project of quitting alcohol, it will just make you feel uncomfortable and out of place.

Also, uh, this just in, Brandsma didn't use a real AA group, because the real AA groups make people be anonymous which makes it inconvenient to research stuff. He just sort of started his own non-anonymous group, let's call it A, with no help from the rest of the fellowship, and had it do Alcoholics Anonymous-like stuff. On the other hand, many members of his control group went out into the community and… attended a real Alcoholics Anonymous, because Brandsma can't exactly ethically tell them not to. So *technically*, there were more people in AA in the no-AA group than in the AA group. Without knowing more about Alcoholics Anonymous, I can't know whether this objection is valid and whether Brandsma's group did or didn't capture the essence of the organization. Still, not the sort of thing you want to hear about a study.

Walsh et al (1991) is a similar study with similar confounders and similar results. Workers in an industrial plant who were in trouble for coming in drunk were randomly assigned either to an inpatient treatment program or to Alcoholics Anonymous. After a year of followup, 60% of the inpatient-treated workers had stayed sober, but only 30% of the AA-treated workers had.

The pro-AA side made three objections to this study, of which one is bad and two are good.

The bad objection was that AA is cheaper than hospitalization, so even if hospitalization is good, AA might be more efficient – after all, we can't afford to hospitalize *everyone*. It's a bad objection because the authors of the study did the math and found out that hospitalization was so much better than AA that it decreased the level of further medical treatment needed and saved the health system more money than it cost.

The first good objection: like the Brandsma study, this study uses people under coercion – in this case, workers who would lose their job if they refused. Fine.

The second good objection, and this one is really interesting: *a lot of inpatient hospital rehab is AA*. That is, when you go to an hospital for inpatient drug treatment, you attend AA groups every day, and when you leave, they make you keep going to the AA groups. In fact, the study says that "at the 12 month and 24 month assessments, the rates of AA affiliation and attendance in the past 6 months did not differ significantly among the groups." Given that the hospital patients got hospital AA + regular AA, they were actually getting *more* AA than the AA group!

So all that this study proves is that AA + more AA + other things is better than AA. There was no "no AA" group, which makes it impossible to discuss how well AA does or doesn't work. Frick.

Timko (2006) is the only study I can hesitantly half-endorse. This one has a sort of clever methodological trick to get around the limitation that doctors can't ethically refuse to refer alcoholics to treatment. In this study, researchers at a Veterans' Affairs hospital randomly assigned alcoholic patients to "referral" or "intensive referral". In "referral", the staff asked the patients to go to AA. In "intensive referral", the researchers asked REALLY NICELY for the patients to go to AA, and gave them nice glossy brochures on how great AA was, and wouldn't shut up about it, and arranged for them to meet people at their first AA meeting so they could have friends in AA, et cetera, et cetera. The hope was that more people in the "intensive referral" group would end out in AA, and ~~that indeed happened~~ scratch that, I just re-read the study and the same number of people in both groups went to AA and the intensive group actually completed a lower number of the 12 Steps on average, have I mentioned I hate all research and this entire field is terrible? But the intensive referral people were more likely to have "had a spiritual awakening" and "have a sponsor", so it was decided the study wasn't a complete loss and when it was found the intensive referral condition had slightly less alcohol use the authors decided to declare victory.

So, whereas before we found that AA + More AA was better than AA, and that proved AA didn't work, in this study we find that AA + More AA was better than AA, and that proves AA *does* work. You know, did I say I hesitantly half-endorsed this study? Scratch that. I hate this study too.

IV.

All right, @#%^ this $@!&*. We need a *real* study, everything all lined up in a row, none of this garbage. Let's just hire half the substance abuse scientists in the country, throw a gigantic wad of money at them, give them as many patients as they need, let them take as long as they want, but barricade the doors of their office and not let them out until they've proven something important beyond a shadow of a doubt.

This was about how the scientific community felt in 1989, when they launched Project MATCH. This eight-year, $30 million dollar, multi-thousand patient trial was supposed to solve everything.

The people going into Project MATCH might have been a little overconfident. Maybe "not even *Zeus* could prevent this study from determining the optimal treatment for alcohol addiction" overconfident. This might have been a mistake.

The study was designed with three arms, one for each of the popular alcoholism treatments of the day. The first arm would be "twelve step facilitation", a form of therapy based off of Alcoholics Anonymous. The second arm would be cognitive behavioral therapy, the most bog-standard psychotherapy in the world and one which by ancient tradition must be included in any kind of study like this. The third arm would be motivational enhancement therapy, which is a very short intervention where your doctor tells you all the reasons you should quit alcohol and tries to get you to convince yourself.

There wasn't a "no treatment" arm. This is where the overconfidence might have come in. Everyone knew alcohol treatment *worked*. Surely you couldn't dispute *that*. They just wanted to see which treatment worked best for which people. So you would enroll a bunch of different people – rich, poor, black, white, married, single, chronic alcoholic, new alcoholic, highly motivated, unmotivated – and see which of these people did best in which therapy. The result would be an algorithm for deciding where to send each of your patients. Rich black single chronic unmotivated alcoholic? We've found with $p < 0.00001$ that the best place for someone like that is in motivational enhancement therapy. Such was the dream. So, eight years and thirty million dollars and the careers of several prestigious researchers later, the results come in, and - yeah, everyone does exactly the same on every kind of therapy (with one minor, possibly coincidental exception). Awkward. "Everybody has won and all must have prizes!". If you're an optimist, you can say all treatments work and everyone can keep doing whatever they like best. If you’re a pessimist, you might start wondering whether anything works at all.

By my understanding this is also the confusing conclusion of Ferri, Amato & Davoli (2006), the Cochrane Collaboration’s attempt to get in on the AA action. Like all Cochrane Collaboration studies since the beginning of time, they find there is insufficient evidence to demonstrate the effectiveness of the intervention being investigated. This has been oft-quoted in the anti-AA literature. But by my reading, they had no control groups and were comparing AA to different types of treatment:

> Three studies compared AA combined with other interventions against other treatments and found few differences in the amount of drinks and percentage of drinking days. Severity of addiction and drinking consequence did not seem to be differentially influenced by TSF versus comparison treatment interventions, and no conclusive differences in treatment drop out rates were reported.

So the two best sources we have – Project MATCH and Cochrane – don’t find any significant differences between AA and other types of therapy. Now, to be fair, the inpatient treatment mentioned in Walsh et al wasn’t included, and inpatient treatment might be the gold standard here. But sticking to various forms of outpatient intervention, they all seem to be about the same.

So, the $64,000 question: do all of them work well, or do all of them work poorly?

V.

Alcoholism studies avoid control groups like they are on fire, presumably because it’s unethical not to give alcoholics treatment or something. However, there is one class of studies that doesn’t have that problem. These are the ones on “brief opportunistic intervention”, which is much like a turbocharged even shorter version of “motivational enhancement therapy”. Your doctor tells you ‘HELLO HAVE YOU CONSIDERED QUITTING ALCOHOL??!!’ and sees what happens.

Brief opportunistic intervention is the most trollish medical intervention ever, because here are all these brilliant psychologists and counselors trying to unravel the deepest mysteries of the human psyche in order to convince people to stop drinking, and then someone comes along and asks “Hey, have you tried just asking them politely?”. And it works.

Not consistently. But it works for about one in eight people. And the theory is that since it only takes a minute or two of a doctor’s time, it scales a lot faster than some sort of hideously complex hospital-based program that takes thousands of dollars and dozens of hours from everyone involved. If doctors would just spend five minutes with each alcoholic patient reminding them that no, really, alcoholism is really bad, we could cut the alcoholism rate by 1/8.

(this also works for smoking, by the way. I do this with every single one of my outpatients who smoke, and most of the time they roll their eyes, because their doctor is giving them *that speech*, but every so often one of them tells me that yeah, I’m right, they know they really should quit smoking and they’ll give it another try. I have never saved anyone’s life by dramatically removing their appendix at the last possible moment, but I have gotten enough patients to promise me they’ll try quitting smoking that I think I’ve saved at least one life just by obsessively doing brief interventions every chance I get. This is probably *the* most effective life-saving thing you can do as a doctor, enough so that if you understand it you *may* be licensed to ignore 80,000 Hours' arguments on doctor replaceability)

Anyway, for some reason, it's okay to do these studies with control groups. And they are so fast and easy to study that everyone studies them all the time. A meta-analysis of 19 studies is unequivocal that they definitely work.

Why do these work? My guess is that they do two things. First, they hit people who honestly didn't realize they had a problem, and inform them that they do. Second, the doctor usually says they'll "follow up on how they're doing" the next appointment. This means that a respected authority figure is suddenly monitoring their drinking and will glare at them if they stay they're still alcoholic. As someone who has gone into a panic because he has a dentist's appointment in a week and he hasn't been flossing enough – and then flossed until his teeth were bloody so the dentist wouldn't be disappointed – I can sympathize with this.

But for our purposes, the brief opportunistic intervention sets a lower bound. It says "Here's a really minimal thing that seems to work. Do other things work better than this?"

The "brief treatment" is the next step up from brief intervention. It's an hour-or-so-long session (or sometimes a couple such sessions) with a doctor or counselor where they tell you some tips for staying off alcohol. I bring it up here because the brief treatment research community spends its time doing studies that show that brief treatments are just as good as much more intense treatments. This might be most comparable to the "motivational enhancement therapy" in the MATCH study.

Chapman and Huygens (1988) find that a single interview with a health professional is just as good as six weeks of inpatient treatment (I don't know about their hospital in New Zealand, but for reference six weeks of inpatient treatment in *my* hospital costs about $40,000.)

Edwards (1977) finds that in a trial comparing "conventional inpatient or outpatient treatment complete with the full panoply of services available at a leading psychiatric institution and lasting several months" versus an hour with a doc, both groups do the same at one and two year followup.

And so on.

All of this is starting to make my head hurt, but it's a familiar sort of hurt. It's the way my head hurts when Scott Aaronson talks about complexity classes. We have all of these different categories of things, and some of them are the same as others and others are bigger than others but we're not sure exactly where all of them stand.

We have classes "no treatment", "brief opportunistic intervention", "brief treatment", "Alcoholics Anonymous", "psychotherapy", and "inpatient".

We can prove that BOI > NT, and that AA = PT. Also that BT = IP = PT. We also have that IP > AA, which unfortunately we can use to prove a contradiction, so let's throw it out for now.

So the hierarchy of classes seems to be (NT) < (BOI) ? (BT, IP, AA, PT) - in other words, no treatment is the worst, brief opportunistic intervention is better, and then *somewhere* in there we have this class of everything else that is the same.

Can we prove that BOI = BT?

We have some good evidence for this, once again from our *Handbook*. A study in Edinburgh finds that five minutes of psychiatrist advice (brief opportunistic intervention) does the same as sixty minutes of advice plus motivational interviewing (brief treatment).

So if we take all this seriously, then it looks like every psychosocial treatment (including brief opportunistic intervention) is the same, and all are better than no treatment. This is a common finding in psychiatry and psychology – for example, all common antidepressants are better than no treatment but work about equally well; all psychotherapies are better than no treatment but work about equally well, et cetera. It's still an open question what this says about our science and our medicine.

The strongest counterexample to this is Walsh et al which finds the inpatient hospital stay works better than the AA referral, but this study looks kind of lonely compared to the evidence on the other side. And even the authors admit they were surprised by the effectiveness of the hospital there.

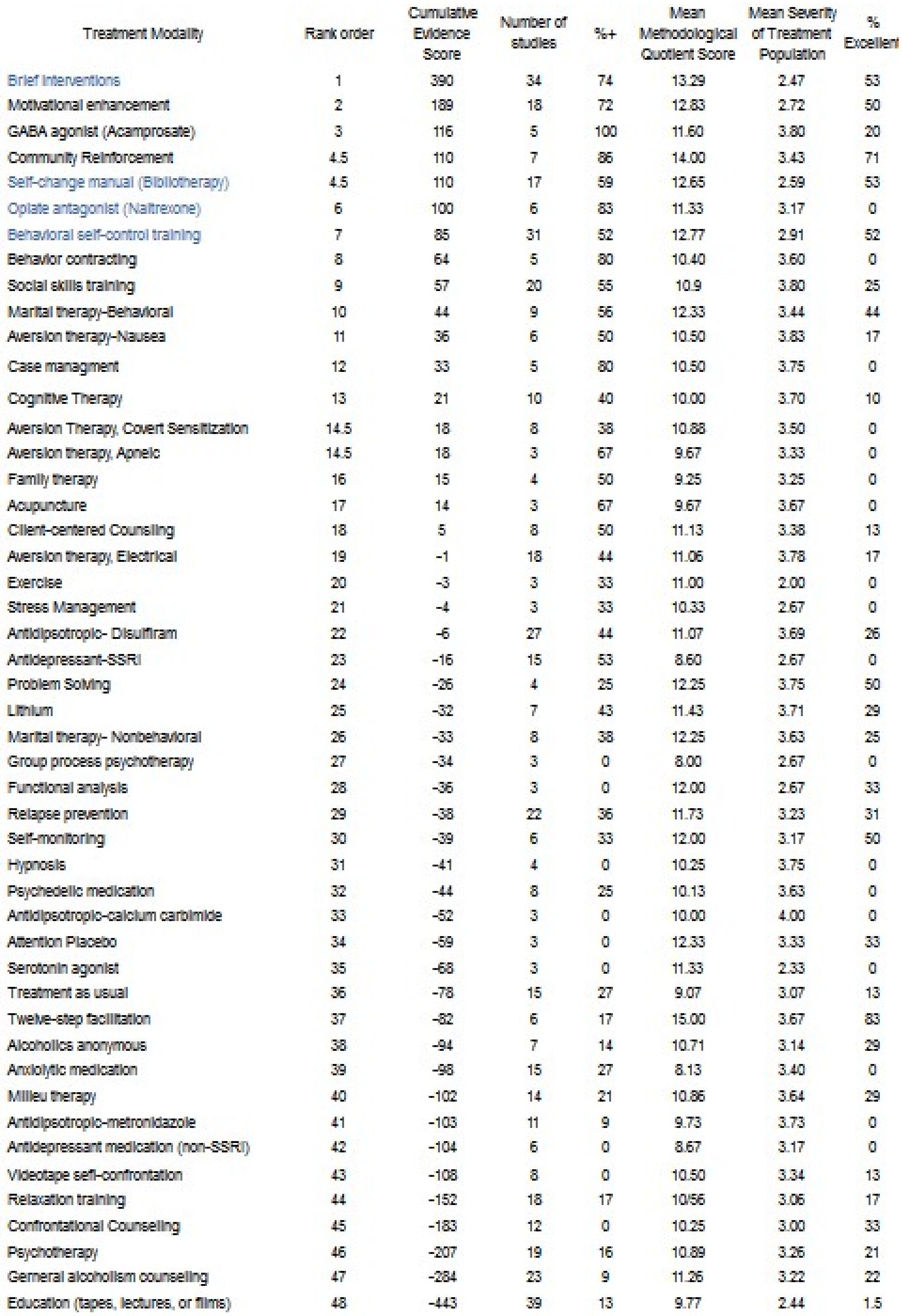
And let's go back to Project MATCH. There wasn't a control group. But there were the people who dropped out of the study, who said they'd go to AA or psychotherapy but never got around to it. Cutter and Fishbain (2005) take a look at what happened to these folks. They find that the dropouts did 75% as well as the people in any of the therapy groups, and that most of the effect of the therapy groups occurred in the first week (ie people dropped out after one week did about 95% as well as people who stayed in).

To me this suggests two things. First, therapy is only a little helpful over most people quitting on their own. Second, insofar as therapy is helpful, the tiniest brush with therapy is enough to make someone think “Okay, I’ve had some therapy, I’ll be better now”. Just like with the brief opportunistic interventions, five minutes of almost anything is enough.

This is a weird conclusion, but I think it’s the one supported by the data.

VI.

I should include a brief word about this giant table.



Treatment Modality	Rank order	Cumulative Evidence Score	Number of studies	%+	Mean Methodological Quotient Score	Mean Severity of Treatment Population	% Excellent
Brief interventions	1	390	34	74	13.29	2.47	53
Motivational enhancement	2	189	18	72	12.83	2.72	50
GABA agonist (Acamprosate)	3	116	5	100	11.60	3.80	20
Community Reinforcement	4.5	110	7	86	14.00	3.43	71
Self-change manual (Bibliotherapy)	4.5	110	17	59	12.65	2.59	53
Opiate antagonist (Naltrexone)	6	100	6	83	11.33	3.17	0
Behavioral self-control training	7	85	31	52	12.77	2.91	52
Behavior contracting	8	64	5	80	10.40	3.60	0
Social skills training	9	57	20	55	10.9	3.80	25
Marital therapy-Behavioral	10	44	9	56	12.33	3.44	44
Aversion therapy-Nausea	11	36	6	50	10.50	3.83	17
Case managment	12	33	5	80	10.50	3.75	0
Cognitive Therapy	13	21	10	40	10.00	3.70	10
Aversion Therapy, Covert Sensitization	14.5	18	8	38	10.88	3.50	0
Aversion therapy, Apneic	14.5	18	3	67	9.67	3.33	0
Family therapy	16	15	4	50	9.25	3.25	0
Acupuncture	17	14	3	67	9.67	3.67	0
Client-centered Counsling	18	5	8	50	11.13	3.38	13
Aversion therapy, Electrical	19	-1	18	44	11.06	3.78	17
Exercise	20	-3	3	33	11.00	2.00	0
Stress Management	21	-4	3	33	10.33	2.67	0
Antidipsotropic- Disulfiram	22	-6	27	44	11.07	3.69	26
Antidepressant-SSRI	23	-16	15	53	8.60	2.67	0
Problem Solving	24	-26	4	25	12.25	3.75	50
Lithium	25	-32	7	43	11.43	3.71	29
Marital therapy- Nonbehavioral	26	-33	8	38	12.25	3.63	25
Group process psychotherapy	27	-34	3	0	8.00	2.67	0
Functional analysis	28	-36	3	0	12.00	2.67	33
Relapse prevention	29	-38	22	36	11.73	3.23	31
Self-monitoring	30	-39	6	33	12.00	3.17	50
Hypnosis	31	-41	4	0	10.25	3.75	0
Psychedelic medication	32	-44	8	25	10.13	3.63	0
Antidipsotropic-calcium carbimide	33	-52	3	0	10.00	4.00	0
Attention Placebo	34	-59	3	0	12.33	3.33	33
Serotonin agonist	35	-68	3	0	11.33	2.33	0
Treatment as usual	36	-78	15	27	9.07	3.07	13
Twelve-step facilitation	37	-82	6	17	15.00	3.67	83
Alcoholics anonymous	38	-94	7	14	10.71	3.14	29
Anxiolytic medication	39	-98	15	27	8.13	3.40	0
Milieu therapy	40	-102	14	21	10.86	3.64	29
Antidipsotropic-metronidazole	41	-103	11	9	9.73	3.73	0
Antidepressant medication (non-SSRI)	42	-104	6	0	8.67	3.17	0
Videotape sefl-confrontation	43	-108	8	0	10.50	3.34	13
Relaxation training	44	-152	18	17	10/56	3.06	17
Confrontational Counseling	45	-183	12	0	10.25	3.00	33
Psychotherapy	46	-207	19	16	10.89	3.26	21
Gemeral alcoholism counseling	47	-284	23	9	11.26	3.22	22
Education (tapes, lectures, or films)	48	-443	39	13	9.77	2.44	1.5

I see it everywhere. It looks very authoritative and impressive and, of course, giant. I believe the source is Miller's Handbook of Alcoholism Treatment Approaches: Effective Alternatives, 3rd Edition, the author of which is known as a very careful scholar whom I cannot help but respect.

And the table does a good thing in discussing medications like acamprosate and naltrexone, which are very important and effective interventions but which will not otherwise be showing up in this post.

However, the therapy part of the table looks really wrong to me.

First of all, I notice acupuncture is ranked 17 out of 48, putting in a much, *much* better showing than treatments like psychotherapy, counseling, or education. Seems fishy.

Second of all, I notice that motivational enhancement (#2), cognitive therapy (#13), and twelve-step (#37) are all about as far apart as could be, but the largest and most powerful trial ever, Project MATCH, found all three to be about equal in effectiveness.

Third of all, I notice that cognitive therapy is at #13, but psychotherapy is at #46. But cognitive therapy is a kind of psychotherapy.

Fourth of all, I notice that brief interventions, motivational enhancement, confrontational counseling, psychotherapy, general alcoholism counseling, and education are all over. But a lot of these are hard to differentiate from one another.

The table seems messed up to me. Part of it is because it is about evidence base rather than effectiveness (consider that handguns have a stronger evidence base than the atomic bomb, since they have been used many more times in much better controlled conditions, but the atomic bomb is more effective) and therefore acupuncture, which is poorly studied, can rank quite high compared to things which have even one negative study.

But part of it just seems wrong. I haven’t read the full book, but I blame the tendency to conflate studies showing “X does not work better than anything else” with “X does not work”.

Remember, whenever there are meta-analyses that contradict single very large well-run studies, go with the single very large well-run study, especially when the meta-analysis is as weird as this one. Project MATCH is the single very large well-run study, and it says this is balderdash. I’m guessing it’s trying to use some weird algorithmic methodology to automatically rate and judge each study, but that’s no substitute for careful human review.

VII.

In conclusion, as best I can tell – and it is not very well, because the studies that could really prove anything robustly haven’t been done – most alcoholics get better on their own. All treatments for alcoholism, including Alcoholics Anonymous, psychotherapy, and just a few minutes with a doctor explaining why she thinks you need to quit, increase this already-high chance of recovery a small but nonzero amount. Furthermore, they are equally effective after only a tiny dose: your first couple of meetings, your first therapy session. Some studies suggest that inpatient treatment with outpatient followup may be better than outpatient treatment alone, but other studies contradict this and I am not confident in the assumption.

So does Alcoholics Anonymous work? Though I cannot say anything authoritatively, my impression is: Yes, but only a tiny bit, and for many people five minutes with a doctor may work just as well as years completing the twelve steps. As such, individual alcoholics may want to consider attending if they don’t have easier options; doctors might be better off just talking to their patients themselves.

If this is true – and right now I don’t have much confidence that it is, it’s just a direction that weak and contradictory data are pointing – it would be really awkward for the multibazillion-dollar treatment industry.

More worrying, I am afraid of what it would do to the War On Drugs. Right now one of the rallying cries for the anti-Drug-War movement is “treatment, not prison”. And although I haven’t looked seriously at the data for any drug besides alcohol. I think some data there are similar. There’s very good medication for drugs – for example methadone and suboxone for opiate abuse – but in terms of psychotherapy it’s mostly the same stuff you get for alcohol. Rehabs, whether they work or not, seem to serve an important sort of ritual function, where if you can send a drug abuser to a rehab you at least feel like something has been done. Deny people that ritual, and it might make prison the only politically acceptable option.

In terms of things to actually treat alcoholism, I remain enamoured of the Sinclair Method, which has done crazy outrageous stuff like conduct an experiment *with an actual control group*. But I haven’t investigated enough to know whether my early excitement about them looks likely to pan out or not.

I would not recommend quitting any form of alcohol treatment that works for you, or refusing to try a form of treatment your doctor recommends, based on any of this information.

Michigan Meetup 11/9

There will be a Michigan rationalist/LW/SSC meetup at my house, halfway between Detroit and Ann Arbor, on Sunday, November 9th at 2 PM. You can find my address and directions here. If you're reading this and interested in rationality and transhumanism, you're invited.

I know some people in Ann Arbor were going to try to arrange transportation, so you can do that in the comments if you want.

[Unrelated PS: Ozy has a blog again and has politely volunteered to host all the race and gender open thread discussion I'm not allowing here.]

All In All, Another Brick In The Motte

One of the better things I’ve done with this blog was help popularize Nicholas Shackel’s “motte and bailey doctrine”. But I’ve recently been reminded I didn’t do a very good job of it. So here is a clean new post on the concept that adds a couple of further thoughts to the original formulation.

The original Shackel paper is intended as a critique of post-modernism. Post-modernists sometimes say things like “reality is socially constructed”, and there’s an uncontroversially correct meaning there. We don’t experience the world directly, but through the categories and prejudices implicit to our society; for example, I might view a certain shade of bluish-green as blue, and someone raised in a different culture might view it as green. Okay.

Then post-modernists go on to say that if someone in a different culture thinks that the sun is light glinting off the horns of the Sky Ox, that’s just as real as our own culture’s theory that the sun is a mass of incandescent gas. If you challenge them, they’ll say that you’re denying reality is socially constructed, which means you’re clearly very naive and think you have perfect objectivity and the senses perceive reality directly.

The writers of the paper compare this to a form of medieval castle, where there would be a field of desirable and economically productive land called a bailey, and a big ugly tower in the middle called the motte. If you were a medieval lord, you would do most of your economic activity in the bailey and get rich. If an enemy approached, you would retreat to the motte and rain down arrows on the enemy until they gave up and went away. Then you would go back to the bailey, which is the place you wanted to be all along.

So the motte-and-bailey doctrine is when you make a bold, controversial statement. Then when somebody challenges you, you retreat to an obvious, uncontroversial statement, and say that was what you meant all along, so you’re clearly right and they’re silly for challenging you. Then when the argument is over you go back to making the bold, controversial statement.

Some classic examples:

1. The religious group that acts for all the world like God is a supernatural creator who builds universes, creates people out of other people’s ribs, parts seas, and heals the sick (bailey). Then when atheists come around and say maybe there’s no God, the religious group objects “But God is just another name for the beauty and order in the Universe! You’re not denying that there’s beauty and order in the Universe, are you?” (motte). Then when the atheists go away they get back to making people out of other people’s ribs and stuff.

2. Or…”If you don’t accept Jesus, you will burn in Hell forever.” (bailey) But isn’t that horrible and inhuman? “Well, Hell is just another word for being without God, and if you choose to be without God, God will be nice and let you make that choice.” (motte) Oh, well that doesn’t sound so bad, I’m going to keep rejecting Jesus. “But if you reject Jesus, you will BURN in HELL FOREVER and your body will be GNAWED BY WORMS.” But didn’t you just… “Metaphorical worms of godlessness!”

3. The feminists who say you can’t be a real feminist unless you believe in specific commitments in X, Y and Z (bailey). Then when someone objects to “feminism”, they argue you can’t object to feminism because it’s “just the belief that women are people!” (motte). Then once the person hastily

retreats and promises he *definitely* didn't mean women aren't people, the feminists get back to demanding everyone support affirmative action because feminism, or arguing about whether you can be a feminist and wear lipstick.

4. Proponents of pseudoscience sometimes argue that their particular form of quackery will cure cancer or take away your pains or heal your crippling injuries (bailey). When confronted with evidence that it doesn't work, they might argue that people need hope, and even a placebo solution will often relieve stress and help people feel cared for (motte). In fact, some have argued that quackery may be better than real medicine for certain untreatable diseases, because neither real nor fake medicine will help, but fake medicine tends to be more calming and has fewer side effects. But then once you leave the quacks in peace, they will go back to telling less knowledgeable patients that their treatments will cure cancer.

5. Critics of the rationalist community note that it pushes controversial complicated things like Bayesian statistics and utilitarianism (bailey) under the name "rationality", but when asked to justify itself defines rationality as "whatever helps you achieve your goals", which is so vague as to be universally unobjectionable (motte). Then once you have admitted that more rationality is always a good thing, they suggest you've admitted everyone needs to learn more Bayesian statistics.

6. Likewise, singularitarians who predict with certainty that there will be a singularity, because "singularity" just means "a time when technology is so different that it is impossible to imagine" – and really, who would deny that technology will probably get really weird (motte)? But then every other time they use "singularity", they use it to refer to a very

specific scenario of intelligence explosion, which is far less certain and needs a lot more evidence before you can predict it (bailey).

The motte and bailey doctrine sounds kind of stupid and hard-to-fall-for when you put it like that, but *all* fallacies sound that way *when you're thinking about them*. More important, it draws its strength from people's usual failure to debate specific propositions rather than vague clouds of ideas. If I'm debating "does quackery cure cancer?", it might be easy to view that as a general case of the problem of "is quackery okay?" or "should quackery be illegal?", and from there it's easy to bring up the motte objection.

Recently, a friend (I think it was Robby Bensinger) pointed out something I'd totally missed. The motte-and-bailey doctrine is a perfect mirror image of my other favorite fallacy, the weak man fallacy.

Weak-manning is a lot like straw-manning, except that instead of debating a fake, implausibly stupid opponent, you're debating a real, unrepresentatively stupid opponent. For example, "Religious people say that you should kill all gays. But this is evil. Therefore, religion is wrong and barbaric. Therefore we should all be atheists." There are certainly religious people who think that you should kill all gays, but they're a small fraction of all religious people and probably not the ones an unbiased observer would hold up as the best that religion has to offer.

If you're debating the Pope or something, then when you weak-man, you're unfairly replacing a strong position (the Pope's) with a weak position (that of the guy who wants to kill gays) to make it more attackable.

But in motte and bailey, you’re unfairly replacing a weak position (there is a supernatural creator who can make people out of ribs) with a strong position (there is order and beauty in the universe) in order to make it more defensible.

So weak-manning is replacing a strong position with a weak position to better attack it; motte-and-bailey is replacing a weak position with a strong position to better defend it.

This means people who know both terms are at constant risk of arguments of the form “You’re weak-manning me!” “No, *you’re* motte-and-baileying *me!*“.

Suppose we’re debating feminism, and I defend it by saying it really is important that women are people, and you attack it by saying that it’s not true that all men are terrible. Then I can accuse you of making life easy for yourself by attacking the weakest statement anyone vaguely associated with feminism has ever pushed. And you can accuse me if making life too easy for myself by defending the most uncontroversially obvious statement I can get away with.

So what is the *real* feminism we should be debating? *Why would you even ask that question?* What is this, some kind of dumb high school debate club? Who the heck thinks it would be a good idea to say “Here’s a vague poorly-defined concept that mind-kills everyone who touches it – quick, should you associate it with positive affect or negative affect?!”

Taboo your words, then replace the symbol with the substance. If you have an *actual thing* you’re trying to debate, then it should be obvious when somebody’s changing the topic. If working out who’s using motte-and-bailey (or weak man) is remotely difficult, it means your discussion went wrong several steps earlier and you probably have no idea what you’re even arguing about.

PS: Nicholas Shackel, original inventor of the term, weighs in.

Ethnic Tension And Meaningless Arguments

I.

Part of what bothers me – and apparently several others – about yesterday’s motte-and-bailey discussion is that here’s a fallacy – a pretty successful fallacy – that depends entirely on people not being entirely clear on what they’re arguing about. Somebody says God doesn’t exist. Another person objects that God is just a name for the order and beauty in the universe. Then this somehow helps defend the position that God is a supernatural creator being. How does that even happen?

“Sir, you’ve been accused of murdering your wife. We have three witnesses who said you did it. What do you have to say for yourself?”

“Well, your honor, I think it’s quite clear I didn’t murder the President. For one thing, he’s surrounded by Secret Service agents. For another, check the news. The President’s still alive.”

“Huh. For some reason I vaguely remember thinking you didn’t have a case. Yet now that I hear you talk, everything you say is incredibly persuasive. You’re free to go.”

While motte-and-bailey is less subtle, it seems to require a similar sort of misdirection. I’m not saying it’s impossible. I’m just saying it’s a fact that needs to be explained.

When everything works the way it’s supposed to in philosophy textbooks, arguments are supposed to go one of a couple of ways:

1. Questions of empirical fact, like “Is the Earth getting warmer?” or “Did aliens build the pyramids?”. You debate these by presenting factual evidence, like “An average of global weather station measurements show 2014 is the hottest year on record” or “One of the bricks at Giza says ‘Made In Tau Ceti V’ on the bottom.” Then people try to refute these facts or present facts of their own.

2. Questions of morality, like “Is it wrong to abort children?” or “Should you refrain from downloading music you have not paid for?” You can only debate these *well* if you’ve already agreed upon a moral framework, like a particular version of natural law or consequentialism. But you can *sort of* debate them by comparing to examples of agreed-upon moral questions and trying to maintain consistency. For exmaple, “You wouldn’t kill a one day old baby, so how is a nine month old fetus different?” or “You wouldn’t download a *car*.”

If you are very lucky, your philosophy textbook will also admit the existence of:

3. Questions of policy, like “We should raise the minimum wage” or “We should bomb Foreignistan”. These are combinations of competing factual claims and competing values. For example, the minimum wage might hinge on factual claims like “Raising the minimum wage would increase unemployment” or “It is very difficult to live on the minimum wage nowadays, and many poor families cannot afford food.” But it might also hinge on value claims like “Corporations owe it to their workers to pay a living wage,” or “It is more important that the poorest be protected than that the economy be strong.” Bombing Foreignistan might depend on factual claims like “The Foreignistanis are harboring terrorists”, and on value claims like “The safety of our people is worth the risk of collateral damage.” If you can

resolve all of these factual and value claims, you should be able to agree on questions of policy.

None of these seem to allow the sort of vagueness of topic mentioned above.

II.

A question: are you pro-Israel or pro-Palestine? Take a second, actually think about it.

Some people probably answered pro-Israel. Other people probably answered pro-Palestine. Other people probably said they were neutral because it’s a complicated issue with good points on both sides.

Probably very few people answered: *Huh? What?*

This question doesn’t fall into any of the three Philosophy 101 forms of argument. It’s not a question of fact. It’s not a question of particular moral truths. It’s not even a question of policy. There are closely related policies, like whether Palestine should be granted independence. But if I support a very specific two-state solution where the border is drawn upon the somethingth parallel, does that make me pro-Israel or pro-Palestine? At exactly which parallel of border does the solution under consideration switch from pro-Israeli to pro-Palestinian? Do you think the crowd of people shouting and waving signs saying “SOLIDARITY WITH PALESTINE” have an answer to that question?

But it’s even worse, because this question covers much more than just the borders of an independent Palestinian state. Was Israel justified by responding to Hamas’ rocket fire by bombing Gaza, even with the near-certainty of collateral damage? Was Israel justified in building a wall across the Palestinian territories to protect itself from potential terrorists, even though it severely curtails Palestinian freedom of movement? Do Palestinians have a "right of return" to territories taken in the 1948 war? Who should control the Temple Mount?

These are four very different questions which one would think each deserve independent consideration. But in reality, what percent of the variance in people's responses do you think is explained by a general "pro-Palestine vs. pro-Israel" factor? 50%? 75%? More?

In a way, when we round people off to the Philosophy 101 kind of arguments, we are failing to respect their self-description. People aren't out on the streets saying "By my cost-benefit analysis, Israel was in the right to invade Gaza, although it may be in the wrong on many of its other actions." They're waving little Israeli flags and holding up signs saying "ISRAEL: OUR STAUNCHEST ALLY". Maybe we should take them at face value.

This is starting to look related to the original question in (I). Why is it okay to suddenly switch points in the middle of an argument? In the case of Israel and Palestine, it might be because people's support for any particular Israeli policy is better explained by a General Factor Of Pro-Israeliness than by the policy itself. As long as I'm arguing in favor of Israel in *some way*, it's still considered by everyone to be on topic.

III.

Some moral philosophers got fed up with nobody being able to explain what the heck a moral truth was and invented emotivism. Emotivism says there *are* no moral truths, just expressions of little personal bursts of emotion. When you say "Donating to charity is good," you don't mean "Donating to charity increases the sum total of utility in the world," or "Donating to charity is in keeping with the Platonic moral law" or "Donating to charity was commanded by God" or even "I like donating to charity". You're just saying "Yay charity!" and waving a little flag.

Seems a lot like how people handle the Israel question. "I'm pro-Israel" doesn't necessarily imply that you believe any empirical truths about Israel, or believe any moral principles about Israel, or even support any Israeli policies. It means you're waving a little flag with a Star of David on it and cheering.

So here is Ethnic Tension: A Game For Two Players.

Pick a vague concept. "Israel" will do nicely for now.

Player 1 tries to associate the concept "Israel" with as much good karma as she possibly can. Concepts get good karma by doing good moral things, by being associated with good people, by being linked to the beloved in-group, and by being oppressed underdogs in bravery debates.

"Israel is the freest and most democratic country in the Middle East. It is one of America's strongest allies and shares our Judeo-Christian values.

Player 2 tries to associate the concept "Israel" with as much bad karma as she possibly can. Concepts get bad karma by committing atrocities, being associated with bad people, being linked to the hated out-group, and by being oppressive big-shots in bravery debates. Also, she obviously needs to neutralize Player 1's actions by disproving all of her arguments.

"Israel may have some level of freedom for its most privileged citizens, but what about the millions of people in the Occupied Territories that have no say? Israel is involved in various atrocities and has often killed innocent protesters. They are essentially a neocolonialist state and have allied with other neocolonialist states like South Africa."

The prize for winning this game is the ability to win the other three types of arguments. If Player 1 wins, the audience ends up with a strongly positive General Factor Of Pro-Israeliness, and vice versa.

Remember, people's capacity for motivated reasoning is pretty much infinite. Remember, a motivated skeptic asks if the evidence *compels* them to accept the conclusion; a motivated credulist asks if the evidence *allows* them to accept the conclusion. Remember, Jonathan Haidt and his team hypnotized people to have strong disgust reactions to the word "often", and then tried to hold in their laughter when people in the lab came up with convoluted yet plausible-sounding arguments against any policy they proposed that included the word "often" in the description.

I've never heard of the experiment being done the opposite way, but it sounds like the sort of thing that might work. Hypnotize someone to have a very positive reaction to the word "often" (for most hilarious results, have it give people an orgasm). "Do you think governments should raise taxes more often?" "Yes. Yes yes YES YES OH GOD YES!"

Once you finish the Ethnic Tension Game, you're replicating Haidt's experiment with the word "Israel" instead of the word "often". Win the game, and any pro-Israel policy you propose will get a burst of positive feelings and tempt people to try to find some explanation, any explanation, that will justify it, whether it's invading Gaza or building a wall or controlling the Temple Mount.

So this is the fourth type of argument, the kind that doesn't make it into Philosophy 101 books. The trope namer is Ethnic Tension, but it applies to anything that can be identified as a Vague Concept, or paired opposing Vague Concepts, which you can use emotivist thinking to load with good or bad karma.

IV.

Now motte-and-bailey stands revealed:

> Somebody says God doesn't exist. Another person objects that God is just a name for the order and beauty in the universe. Then this somehow helps defend the position that God is a supernatural creator being. How does that even happen?

The two-step works like this. First, load "religion" up with good karma by pitching it as persuasively as possible. "Religion is just the belief that there's beauty and order in the universe."

Wait, *I* think there's beauty and order in the universe!

"Then you're religious too. We're all religious, in the end, because religion is about the common values of humanity and meaning and compassion sacrifice beauty of a sunrise Gandhi Buddha Sufis St. Francis awe complexity humility wonder Tibet the Golden Rule love."

Then, once somebody has a strongly positive General Factor Of Religion, it doesn't really matter whether someone believes in a creator God or not. If they have any predisposition whatsoever to do so, they'll find a reason to let themselves. If they can't manage it, they'll say it's true "metaphorically" and continue to act upon every corollary of it being true.

("God is just another name for the beauty and order in the universe. But Israel definitely belongs to the Jews, because the beauty and order of the universe promised it to them.")

If you're an atheist, you probably have a lot of important issues on which you want people to consider non-religious answers and policies. And if somebody can maintain good karma around the "religion" concept by believing God is the order and beauty in the universe, then that can still be a victory for religion even if it is done by jettisoning many traditionally "religious" beliefs. In this case, it is useful to think of the "order and beauty" formulation as a "motte" for the "supernatural creator" formulation, since it's allowing the *entire concept* to be defended.

But even this is giving people too much credit, because the existence of God is a (sort of) factual question. From yesterday's post:

> Suppose we're debating feminism, and I defend it by saying it really is important that women are people, and you attack it by saying that it's not true that all men are terrible. What is the real feminism we should be debating? Why would you even ask that question? What is this, some kind of dumb high school debate club? Who the heck thinks it would be a good idea to say 'Here's a vague poorly-defined concept that mind-kills everyone who touches it – quick, should you associate it with positive affect or negative affect?!'

Who the heck thinks that? Everybody, all the time.

Once again, if I can load the concept of "feminism" with good karma by making it so obvious nobody can disagree with it, then I have a massive "home field advantage" when I'm trying to convince anyone of any particular policy that can go under the name “feminism”, even if it’s unrelated to the arguments that gave feminism good karma in the first place.

Or if I’m against feminism, I just post quotes from the ten worst feminists on Tumblr again and again until the entire movement seems ridiculous and evil, and then you’ll have trouble convincing anyone of *anything* feminist. “That seems reasonable…but wait, isn’t that a feminist position? Aren’t those the people I hate?”

(compare: most Americans oppose Obamacare, but most Americans support each individual component of Obamacare when it is explained without using the word “Obamacare”)

V.

Little flow diagram things make everything better. Let’s make a little flow diagram thing.

We have our node “Israel”, which has either good or bad karma. Then there’s another node close by marked “Palestine”. We would expect these two nodes to be pretty anti-correlated. When Israel has strong good karma, Palestine has strong bad karma, and vice versa.

Now suppose you listen to Noam Chomsky talk about how strongly he supports the Palestinian cause and how much he dislikes Israel. One of two things can happen:

“Wow, a great man such as Noam Chomsky supports the Palestinians! They must be very deserving of support indeed!”

or

“That idiot Chomsky supports Palestine? Well, screw him. And screw them!”

So now there is a third node, Noam Chomsky, that connects to both Israel and Palestine, and we have discovered it is positively correlated with Palestine and negatively correlated with Israel. It probably has a pretty low weight, because there are a lot of reasons to care about Israel and Palestine other than Chomsky, and a lot of reasons to care about Chomsky other than Israel and Palestine, but the connection is there.

I don’t know anything about neural nets, so maybe this system isn’t actually a neural net, but whatever it is I’m thinking of, it’s a structure where eventually the three nodes reach some kind of equilibrium. If we start with someone liking Israel and Chomsky, but not Palestine, then either that’s going to shift a little bit towards liking Palestine, or shift a little bit towards disliking Chomsky.

Now we add more nodes. Cuba seems to really support Palestine, so they get a positive connection with a little bit of weight there. And I think Noam Chomsky supports Cuba, so we’ll add a connection there as well. Cuba is socialist, and that’s one of the most salient facts about it, so there’s a heavily weighted positive connection between Cuba and socialism. Palestine kind of makes noises about socialism but I don’t think they have any particular economic policy, so let’s say very weak direct connection. And Che is heavily associated with Cuba, so you get a pretty big Che – Cuba connection, plus a strong direct Che – socialism one. And those pro-Palestinian students who threw rotten fruit at an Israeli speaker also get a little path connecting them to “Palestine” – hey, why not – so that if you support Palestine you might be willing to excuse what they did and if you oppose them you might be a little less likely to support Palestine.

Back up. This model produces crazy results, like that people who like Che are more likely to oppose Israel bombing Gaza. That's such a weird, implausible connection that it casts doubt upon the entire…

Oh. Wait. Yeah. Okay.

I think this kind of model, in its efforts to sort itself out into a ground state, might settle on some kind of General Factor Of Politics, which would probably correspond pretty well to the left-right axis.

In Five Case Studies On Politicization, I noted how fresh new unpoliticized issues, like the Ebola epidemic, were gradually politicized by connecting them to other ideas that were already part of a political narrative. For example, a quarantine against Ebola would require closing the borders. So now there's a weak negative link between "Ebola quarantine" and "open borders". If your "open borders" node has good karma, now you're a little less likely to support an Ebola quarantine. If "open borders" has bad karma, a little more likely.

I also tried to point out how you could make different groups support different things by changing your narrative a little:

> Global warming has gotten inextricably tied up in the Blue Tribe narrative: Global warming proves that unrestrained capitalism is destroying the planet. Global warming disproportionately affects poor countries and minorities. Global warming could have been prevented with multilateral action, but we were too dumb to participate because of stupid American cowboy diplomacy. Global warming is an important cause that

activists and NGOs should be lauded for highlighting. Global warming shows that Republicans are science denialists and probably all creationists. Two lousy sentences on “patriotism” aren’t going to break through that.

If I were in charge of convincing the Red Tribe to line up behind fighting global warming, here’s what I’d say:

In the 1950s, brave American scientists shunned by the climate establishment of the day discovered that the Earth was warming as a result of greenhouse gas emissions, leading to potentially devastating natural disasters that could destroy American agriculture and flood American cities. As a result, the country mobilized against the threat. Strong government action by the Bush administration outlawed the worst of these gases, and brilliant entrepreneurs were able to discover and manufacture new cleaner energy sources. As a result of these brave decisions, our emissions stabilized and are currently declining.

Unfortunately, even as we do our part, the authoritarian governments of Russia and China continue to industralize and militarize rapidly as part of their bid to challenge American supremacy. As a result, Communist China is now by far the world’s largest greenhouse gas producer, with the Russians close behind. Many analysts believe Putin secretly welcomes global warming as a way to gain access to frozen Siberian resources and weaken the more temperate United States at the same time. These countries blow off huge disgusting globs of toxic gas, which effortlessly cross American borders and disrupt the climate of the United States. Although we have asked them to stop several times, they refuse, perhaps egged on by major oil producers like Iran and

> Venezuela who have the most to gain by keeping the world dependent on the fossil fuels they produce and sell to prop up their dictatorships.
>
> We need to take immediate action. While we cannot rule out the threat of military force, we should start by using our diplomatic muscle to push for firm action at top-level summits like the Kyoto Protocol. Second, we should fight back against the liberals who are trying to hold up this important work, from big government bureaucrats trying to regulate clean energy to celebrities accusing people who believe in global warming of being ‘racist’. Third, we need to continue working with American industries to set an example for the world by decreasing our own emissions in order to protect ourselves and our allies. Finally, we need to punish people and institutions who, instead of cleaning up their own carbon, try to parasitize off the rest of us and expect the federal government to do it for them.

In the first paragraph, “global warming” gets positively connected to concepts like “poor people and minorities” and “activists and NGOs”, and gets negatively connected to concepts like “capitalism”, “American cowboy diplomacy”, and “creationists”. That gives global warming really strong good karma if (and only if) you like the first two concepts and hate the last three.

In the next three paragraphs, “global warming” gets positively connected to “America”, “the Bush administration” and “entrepreneurs”, and negatively connected to “Russia”, “China”, “oil producing dictatorships like Iran and Venezuela”, “big government bureaucrats”, and “welfare parasites”. This is going to appeal to, well, a different group.

Notice two things here. First, the exact connection isn’t that important, as long as we can hammer in the existence of a connection. I could probably just say GLOBAL WARMING! COMMUNISM! GLOBAL WARMING! COMMUNISM! GLOBAL WARMING! COMMUNISM! several hundred times and have the same effect if I could get away with it (this is the principle behind attack ads which link a politician’s face to scary music and a very concerned voice).

Second, there is no attempt whatsoever to challenge the idea that the issue at hand is the positive or negative valence of a concept called “global warming”. At no point is it debated what the solution is, which countries the burden is going to fall on, or whether any particular level of emission cuts would do more harm than good. It’s just accepted as obvious by both sides that we debate “for” or “against” global warming, and if the “for” side wins then they get to choose some solution or other or whatever oh god that’s so boring can we get back to Israel vs. Palestine.

Some of the scientists working on IQ have started talking about “hierarchical factors”, meaning that there’s a general factor of geometry intelligence partially correlated with other things into a general factor of mathematical intelligence partially correlated with other things into a general factor of total intelligence.

I would expect these sorts of things to work the same way. There’s a General Factor Of Global Warming that affects attitudes toward pretty much all proposed global warming solutions, which is very highly correlated with a lot of other things to make a General Factor Of Environmentalism, which itself is moderately highly correlated with other things into the General Factor Of Politics.

VI.

Speaking of politics, a fruitful digression: what the heck was up with the Ashley Todd mugging hoax in 2008?

Back in the 2008 election, a McCain campaigner claimed (falsely, it would later turn out) to have been assaulted by an Obama supporter. She said he slashed a "B" (for "Barack") on her face with a knife. This got a lot of coverage, and according to Wikipedia:

> John Moody, executive vice president at Fox News, commented in a blog on the network's website that "this incident could become a watershed event in the 11 days before the election," but also warned that "if the incident turns out to be a hoax, Senator McCain's quest for the presidency is over, forever linked to race-baiting."

Wait. One Democrat, presumably not acting on Obama's direct orders, attacks a Republican woman. And this is supposed to *alter the outcome of the entire election*? In what universe does one crime by a deranged psychopath change whether Obama's tax policy or job policy or bombing-scary-foreigners policy is better or worse than McCain's?

Even *if* we're willing to make the irresponsible leap from "Obama is supported by psychopaths, therefore he's probably a bad guy," there are like a hundred million people on each side. Psychopaths are usually estimated at about 1% of the population, so any movement with a million people will already have 10,000 psychopaths. Proving the existence of a single one changes *nothing*.

I think insofar as this affected the election – and everyone seems to have agreed that it might have – it hit President Obama with a burst of bad karma. Obama something

something psychopath with a knife. Regardless of the exact content of those something somethings, *is that the kind of guy you want to vote for*?

Then when it was discovered to be a hoax, it was McCain something something race-baiting hoaxer. Now *he's* got the bad karma!

This sort of conflation between a cause and its supporters really only makes sense in the emotivist model of arguing. I mean, this shouldn't even get dignified with the name *ad hominem* fallacy. Ad hominem fallacy is “McCain had sex with a goat, therefore whatever he says about taxes is invalid.” At least it's still the same *guy*. This is something the philosophy textbooks can't bring themselves to believe really exists, even as a fallacy.

But if there's a General Factor Of McCain, then anything bad remotely connected to the guy – goat sex, lying campaigners, whatever – reflects on everything else about him.

This is the same pattern we see in Israel and Palestine. How many times have you seen a news story like this one: “Israeli speaker hounded off college campus by pro-Palestinian partisans throwing fruit. Look at the intellectual bankruptcy of the pro-Palestinian cause!” It's clearly intended as an argument for *something* other than just not throwing fruit at people. The causation seems to go something like “These particular partisans are violating the usual norms of civil discussion, therefore they are bad, therefore something associated with Palestine is bad, therefore your General Factor of Pro-Israeliness should become more strongly positive, therefore it's okay for Israel to bomb Gaza.” Not usually said in those *exact words*, but the thread can be traced.

VII.

Here is a prediction of this model: we will be obsessed with what concepts we can connect to other concepts, even when the connection is totally meaningless.

Suppose I say: “Opposing Israel is anti-Semitic”. Why? Well, the Israelis are mostly Jews, so in a sense by definition being anti- them is “anti-Semitic”, broadly defined. Also, p(opposes Israel|is anti-Semitic) is probably pretty high, which sort of lends some naive plausibility to the idea that p(is anti-Semitic|opposes Israel) is at least higher than it otherwise *could* be.

Maybe we do our research and we find exactly what percent of opponents of Israel endorse various anti-Semitic statements like “I hate all Jews” or “Hitler had some bright ideas”. We’ve replaced the symbol with the substance. Problem solved, right?

Maybe not. In the same sense that people can agree on all of the characteristics of Pluto – its diameter, the eccentricity of its orbit, its number of moons – and still disagree on the question “Is Pluto a planet”, one can agree on every characteristic of every Israel opponent and still disagree on the definitional question “Is opposing Israel anti-Semitic?”

(fact: it wasn’t until proofreading this essay that I realized I had originally written “Is Israel a planet?” and “Is opposing Pluto anti-Semitic?” I would like to see Jonathan Haidt hypnotize people until they can come up with positive arguments for those propositions.)

What’s the point of this useless squabble over definitions?

I think it’s about drawing a line between the concept “anti-Semitism” and “oppose Israel”. If your head is screwed on right, you assign anti-Semitism some very bad karma. So if we can stick a thick line between “anti-Semitism” and “oppose Israel”, then you’re going have very bad feelings about opposition to Israel and your General Factor Of Pro-Israeliness will go up.

Notice that this model *is transitive, but shouldn’t be*.

That is, let’s say we’re arguing over the definition of anti-Semitism, and I say “anti-Semitism just means anything that hurts Jews”. This is a dumb definition, but let’s roll with it.

First, I load “anti-Semitism” with lots of negative affect. Hitler was anti-Semitic. The pogroms in Russia were anti-Semitic. The Spanish Inquisition was anti-Semitic. Okay, negative affect achieved.

Then I connect “wants to end the Israeli occupation of Palestine” to “anti-Semitism”. Now wanting to end the Israeli occupation of Palestine has lots of negative affect attached to it.

It sounds dumb when you put it like that, but when you put it like “You’re anti-Semitic for wanting to end the occupation” it’s a pretty damaging argument.

This is *trying* to be transitive. It’s trying to say “anti-occupation = anti-Semitism, anti-Semitism = evil, therefore anti-occupation = evil”. If this were arithmetic, it would work. But there’s no Transitive Property Of Concepts. If anything, concepts are more like sets. The logic is “anti-occupation is a member of the set anti-Semitic, the set anti-Semitic contains members that are evil, therefore anti-occupation is evil”, which obviously doesn’t check out.

(compare: “I am a member of the set ‘humans’, the set ‘humans’ contains the Pope, therefore I am the Pope”.)

Anti-Semitism is generally considered evil because a lot of anti-Semitic things involve killing or dehumanizing Jews. Opposing the Israel occupation of Palestine doesn't kill or dehumanize Jews, so even if we call it "anti-Semitic" by definition, there's no reason for our usual bad karma around anti-Semitism to transfer over. But by an unfortunate rhetorical trick, it does – you can gather up bad karma into "anti-Semitic" and then shoot it at the "occupation of Palestine" issue just by clever use of definitions.

This means that if you can come up with sufficiently clever definitions and convince your opponent to accept them, you can win any argument by default just by having a complex system of mirrors in place to reflect bad karma from genuinely evil things to the things you want to tar as evil. This is essentially the point I make in Words, Words, Words.

If we kinda tweak the definition of "anti-Semitism" to be "anything that inconveniences Jews", we can pull a trick where we leverage people's dislike of Hitler to make them support the Israeli occupation of Palestine – but in order to do that, we need to get everyone on board with our *slightly* non-standard definition. Likewise, the social justice movement insists on their own novel definitions of words like "racism" that don't match common usage, any dictionary, or etymological history – but which do perfectly describe a mirror that reflects bad karma toward opponents of social justice while making it impossible to reflect any bad karma back. Overreliance on this mechanism explains why so many social justice debates end up being about whether a particular mirror can be deployed to transfer bad karma in a specific case ("are trans people privileged?!") rather than any feature of the real world.

But they are hardly alone. Compare: "Is such an such an organization a *cult*?", "Is such and such a policy *socialist*?", "Is abortion or capital punishment or war *murder*?" All entirely about whether we're allowed to reflect bad karma from known sources of evil to other topics under discussion.

Look around you. Just look around you. Have you worked out what we're looking for? Correct. The answer is The Worst Argument In The World. Only now, we can explain why it works.

VIII.

From the self-esteem literature, I gather that the self is also a concept that can have good or bad karma. From the cognitive dissonance literature, I gather that the self is actively involved in maintaining good karma around itself through as many biases as it can manage to deploy.

I've mentioned this study before. Researchers make ~~victims~~ participants fill out a questionnaire about their romantic relationships. Then they pretend to "grade" the questionnaire, actually assigning scores at random. Half the participants are told their answers indicate they have the tendency to be very faithful to their partner. The other half are told they have very low faithfulness and their brains just aren't built for fidelity. Then they ask the ~~participants~~ victims their opinion on staying faithful in a relationship – very important, moderately important, or not so important?

There is a strong signal of people who are told they are bad at fidelity to state fidelity is unimportant, and another strong signal of people who are told they are especially faithful stating that fidelity is a great and noble virtue that must be protected.

The researchers conclude that people want to have high self-esteem. If I am terrible at fidelity, and fidelity is the most important virtue, that makes me a terrible person. If I am terrible at fidelity and fidelity doesn’t matter, I’m fine. If I am great at fidelity, and fidelity is the most important virtue, I can feel pretty good about myself.

This doesn’t seem too surprising. It’s just the more subtle version of the effect where white people are a lot more likely to be white supremacists than members of any other race. Everyone likes to hear that they’re great. The question is whether they can defend it and fit it in with their other ideas. The answer is “usually yes, because people are capable of pretty much any contortion of logic you can imagine and a lot that you can’t”.

I had a bad experience when I was younger where a bunch of feminists attacked and threatened me because of something I wrote. It left me kind of scarred. More importantly, the shape of that scar was a big anticorrelated line between self-esteem and the “feminism” concept. If feminism has lots of good karma, then I have lots of bad karma, because I am a person feminists hate. If feminists have lots of bad karma, then I look good by comparison, the same way it’s pretty much a badge of honor to be disliked by Nazis. The result was a permanent haze of bad karma around “feminism” unconnected to any specific feminist idea, which I have to be constantly on the watch for if I want to be able to evaluate anything related to feminism fairly or rationally.

Good or bad karma, when applied to yourself, looks like high or low self-esteem; when applied to groups, it looks like high or low status. In the giant muddle of a war for status that we politely call “society”, this makes beliefs into weapons and the karma loading of concepts into the difference between lionization and dehumanization.

The Trope Namer for emotivist arguments is “ethnic tension”, and although it’s most obvious in the case of literal ethnicities like the Israelis and the Palestinians, the ease with which concepts become attached to different groups creates a whole lot of “proxy ethnicites”. I’ve written before about how American liberals and conservatives are seeming less and less like people who happen to have different policy prescriptions, and more like two different tribes engaged in an ethnic conflict quickly approaching Middle East level hostility. More recently, a friend on Facebook described the-thing-whose-name-we-do-not-speak-lest-it-appear and-destroy-us-all, the one involving reproductively viable worker ants, as looking more like an ethnic conflict about who is oppressing whom than any real difference in opinions.

Once a concept has joined up with an ethnic group, either a real one or a makeshift one, it’s impossible to oppose the concept without simultaneously lowering the status of the ethnic group, which is going to start at least a *little* bit of a war. Worse, once a concept has joined up with an ethnic group, one of the best ways to argue against the concept is to dehumanize the ethnic group it’s working with. Dehumanizing an ethnic group has always been easy – just associate them with a disgust reaction, portray them as conventionally unattractive and unlovable and full of all the worst human traits – and now it is profitable as well, since it’s one of the fastest ways to load bad karma into an idea you dislike.

IX.

According to The Virtues Of Rationality:

> The tenth virtue is precision. One comes and says: The quantity is between 1 and 100. Another says: the

> quantity is between 40 and 50. If the quantity is 42 they are both correct, but the second prediction was more useful and exposed itself to a stricter test. What is true of one apple may not be true of another apple; thus more can be said about a single apple than about all the apples in the world. The narrowest statements slice deepest, the cutting edge of the blade. As with the map, so too with the art of mapmaking: The Way is a precise Art. Do not walk to the truth, but dance. On each and every step of that dance your foot comes down in exactly the right spot. Each piece of evidence shifts your beliefs by exactly the right amount, neither more nor less. What is exactly the right amount? To calculate this you must study probability theory. Even if you cannot do the math, knowing that the math exists tells you that the dance step is precise and has no room in it for your whims.

The official desciption is of literal precision, as specific numerical precision in probability updates. But is there a secret interpretation of this virtue?

> Four top secret Virtues known only to the Highest Clergy: 1) Fnorg 2) Turlity 3) Charigrace 4) Love-231.
>
> — Deity Of Religion (@deityofreligion) October 24, 2014

Precision as separation. Once you're debating "religion", you've already lost. Precision as sticking to a precise question, like "Is the first chapter of Genesis literally true?" or "Does Buddhist meditation help treat anxiety disorders?" and trying to keep these issues as separate from any General Factor Of Religiousness as humanly possible. Precision such that "God the supernatural Creator exists" and "God the order and beauty in the Universe exists" are as carefully sequestered from one another as "Did the defendant kill his wife?” and “Did the defendant kill the President?”

I want to end by addressing a point a commenter made in my last post on motte-and-bailey:

> In the real world, the particular abstract questions aren’t what matter – the groups and people are what matter. People get things done, and they aren’t particularly married to particular abstract concepts, they are married to their values and their compatriots. In order to deal with reality, we must attack and defend groups and individuals. That does not mean forsaking logic. It requires dealing with obfuscating tactics like those you outline above, but that’s not even a real downside, because if you flee into the narrow, particular questions all you’re doing is covering your eyes to avoid perceiving the the monsters that will still make mincemeat of your attempts to change things.

I don’t entirely disagree with this. But I think we’ve been over this territory before.

The world is a scary place, full of bad people who want to hurt you, and in the state of nature you’re pretty much obligated to engage in whatever it takes to survive.

But instead of sticking with the state of nature, we have the ability to form communities built on mutual disarmament and mutual cooperation. Despite artificially limiting themselves, these communities become stronger than the less-scrupulous people outside them, because they can work together effectively and because they can boast a better quality of life that attracts their would-be enemies to join them. At least in the short term, these communities can resist races to the bottom and prevent the use of personally effective but negative-sum strategies.

One such community is the kind where members try to stick to rational discussion as much as possible. These communities are definitely better able to work together, because they have a powerful method of resolving empirical disputes. They're definitely better quality of life, because you don't have to deal with constant insult wars and personal attacks. And the existence of such communities provides positive externalities to the outside world, since they are better able to resolve difficult issues and find truth.

But forming a rationalist community isn't just about having the *will* to discuss things well. It's also about having the *ability*. Overcoming bias is really hard, and so the members of such a community need to be constantly trying to advance the art and figure out how to improve their discussion tactics.

As such, it's acceptable to try to determine and discuss negative patterns of argument, even if those patterns of argument are useful and necessary weapons in a state of nature. If anything, understanding them makes them *easier* to use if you've got to use them, and makes them easier to recognize and counter from others, giving a slight advantage in battle if that's the kind of thing you like. But moving them from unconscious to conscious also gives you the crucial *choice* of when to deploy them and allows people to try to root out ethnic tension in particular communities.

The Right To Waive Your Rights

I do understand the logic behind not allowing just any old contract to be legally binding. The reductio ad absurdum is the EULA that says “By opening this product, you agree not to sue us if this product malfunctions and hurts you, not to give us any negative feedback, and not to object if this product monitors everything you do and reports it back to us.” And people never read those things, so it basically means companies can be above whatever laws they choose.

The *other* reductio is the job offer with the contract saying “By accepting this job, you agree to work whatever hours we tell you without overtime, and you can’t raise sexual harassment complaints, and you’re bound by a non-compete agreement not to work in this industry again if we fire you.” That’s basically selling yourself into slavery, and although in theory the problem should be limited by people being unwilling to sign such a deal, in practice, the job market.

I do understand the logic, really. But restrictions on contracts scare me, and they should scare you too.

It’s easy to say things like “Well, in those examples above, contracts are a tool used by the powerful to oppress the powerless. And it seems like a general case that the powerful will have lots of ability to coerce the powerless into signing unfair contracts, so in general banning forms of voluntary contract should always be a progressive, pro-egalitarian position.” And a lot of progressive egalitarians *do* say this.

So let me talk about my f@#king life.

Every day I get to evaluate new psychiatric patients in the emergency room. A lot of them are there for attempted suicide. They say they’re depressed. They say they’re not getting any treatment for their depression. I gingerly bring up that they might want to stay a couple of days in the psych unit for treatment.

“Oh, no, I could never do that. My sainted mother’s in the hospital on her death bed now, and I’m the only family she has left, and if I’m not there with her when she goes it would haunt me forever. And my kid has his Little League championships this evening, and it’s the only time he’s ever won something, and he specifically told me if I’m not there cheering him on he’ll never forgive me.”

I gingerly bring up that, actually, the law *requires* that if someone is a danger to themselves or others, I *have* to commit them to the psychiatric hospital, whether it is convenient for them or not. That sometimes getting treatment for unremitting life-ruining totally unmanaged depression is a *good* thing.

“You don’t understand. I’ve already missed a couple days of work this month, and my boss says if I miss any more I’m going to be fired, and there’s no way I can find another job in this economy, and without that money me and my kids will lose our house. And I’m in the middle of a divorce and I have to be at the court in two days for the hearing or else my sadistic abusive ex will get custody of the kids.”

I gingerly bring up that no, really, this isn’t a debate, there are *rules* here.

“You don’t understand! I have to transport my adopted daughter’s boyfriend to safety. This man’s done no wrong, and he needs a doctor’s care. Another hour yet, and then I’m yours, and all our debts are paid.”

I’m not heartless. I really want to respect people’s autonomy. I obviously want to help people stay alive long enough to get better, but I don’t really have any personal investment in maintaining the suicide rate at exactly zero if the human costs of doing so are too high. I just want to do more good for my patients than harm. So finally I give up and go ask a superior, and they always say the same thing. Commit.

And they always give the same reason why. A guy begs to be allowed to go, says he’s not suicidal at all, everything will be fine. The doctor goes soft and lets him. He thanks the doctor profusely, says she can’t possibly understand how much it means to him. The next month he shoots himself and is permanently crippled. A lawyer informs him he can get $500,000 by suing the doctor for breach of duty since she let him go home even though he was clearly suicidal, and the doctor doesn’t have a leg to stand on. There’s the guy on the stand, saying “She knew I was suicidal, I told her all about it, and she did *nothing*!”

(Doctor’s attorney: “But didn’t you specifically ask not be treated?” Patient: “Yes, but I wasn’t in my right mind. Maybe the fact that I was undergoing evaluation at a psychiatric hospital *should have clued you in!*“)

I want to do what’s right for my patients. But I also want to follow the law. And I also want to avoid losing my livelihood and everything I have. There have been times I may have slightly bent certain standards, when the moral pull was so strong I wasn’t going to be able to sleep at night otherwise. But on the main, I don’t *want* to last only three years in the business. That’s not good for me, and it’s not good for patients who I might otherwise be able to help.

The progressive says: “This insistence on the sanctity of voluntary contracts only benefits privileged oppressors. It allows them to force poor innocent victims to sign away their rights to protest ill treatment.”

And so it does. I’m clearly the privileged oppressor here. And I would really like to be able to ask my patients to sign a contract saying they waive their rights to sue me if things go wrong. Actually, I want to be even more evil than that. I want to ask my patients to sign a contract waiving their rights to sue me, and threaten to commit them to hospital against their will if they refuse.

And yet this would be the most powerful method possible of protecting patient autonomy. I would love to be able to bend the rules for a patient who has some really good reason not to want to go to hospital, whether it’s their dying mother or their demanding boss or even just that they can’t afford it (did you know they can force you to pay for your involuntary commitments? What a country!) I would love to be able to tell them “Well, as your psychiatrist I strongly recommend hospitalization, but you’re in your right mind, you’ve got decision making capacity, and hey, it’s your life.” But the only way I can do this without pretty much ensuring I’m going to get bitten sooner or later is if they can waive their rights to take advantage of me.

When you prevent people from making deals to cooperate with each other, *terrible things happen*. It means that, unless one or another party is a martyr, they’re both going to end up defecting on each other. And when both parties defect on each other, that’s no big deal for the more powerful party, but a disaster for the powerless oppressed people we're trying to help. When we end up in mutual defection, *I* don't have to deal with anything worse than the patient yelling at me while I write it down on my little pad and try to sound sympathetic. *The patient* has to miss their son's Little League championship game and then he never talks to them again.

(ninety percent of sob stories are false, but ten percent are true).

I know it sounds weird to insist on a right to waive your rights. Isn't that more of an anti-right, so to speak? But come on, read your Schelling. In multiplayer games, the ability to limit your options can provide a decisive advantage. If you're playing Chicken, the winning strategy is to conspicuously break your steering wheel so your opponent knows you can't turn even if you want to. If you're playing global thermonuclear war, the winning strategy is to conspicuously remove your ability not to retaliate, using something like the Dead Hand system. Waiving your right to steer, waiving your right not to nuke, these are winning strategies; whoever can't do them has been artificially handicapped.

I do understand the logic behind not allowing just any old contract to be legally binding. But I also think that the right to waive your rights *is* a right. We understand that there are cases in which we can violate rights; there are a whole host of exceptions to the right of free speech. But it requires a good reason, and you had better realize you're treading on dangerous ground.

Growing Old

A while ago a friend asked me to critique her writing. I said "You sound like a teenager". It was less patronizing than it might have been, because she *was* a teenager, although I guess still pretty patronizing. Then she asked me for an explanation, and I didn't have one, because some kind of "essence of teenagerdom" is hard to place.

But recently I was thinking about this again, because I was rereading Byron's "Growing Old". Part of his *Don Juan*, it's a series of reflections about turning thirty (really, Byron? Growing Old? *Thirty?*). I was reading it because I had read it when I was fifteen or so, and gotten some things out of it, and I'd resolved to reread it when I was older to see if I could get anything else:

> But now at thirty years my hair is grey—
> (I wonder what it will be like at forty ?
> I thought of a peruke the other day—)
> My heart is not much greener ; and, in short, I
> Have squandered my whole summer while 'twas May,
> And feel no more the spirit to retort ; I
> Have spent my life, both interest and principal,
> And deem not, what I deemed, my soul invincible.

And lo and behold, I do sympathize a lot more now. For example, what's with my hair? It's not turning grey. But it is falling out *en masse*. I haven't thought about a peruke – which I think it one of those big white old-timey wigs George Washington used to wear – yet. But maybe I should.

But moving on:

> No more—no more—Oh ! never more on me
> The freshness of the heart can fall like dew,
> Which out of all the lovely things we see
> Extracts emotions beautiful and new ;
> Hived in our bosoms like the bag o' the bee.
> Think'st thou the honey with those objects grew ?
> Alas ! 'twas not in them, but in thy power
> To double even the sweetness of a flower.

There's something very raw about being young. I remember reading a psychotherapy book that, like most psychotherapy books, talks about childhood trauma. Their prescription was that it gets buried under lots of layers of unconscious baggage, and you need to bring it to the surface. Once it's at the surface, the patient's reaction should be something like "That? That was what bothered me all this time?" Because when you're a child, everything is more intense. Yeah, some childhood trauma is getting beaten or abused. But other childhood trauma is getting called names on the playground, or being left alone without knowing where your parents were. I find a lot of the "inner child" school of psychology to be kind of bunk, but I find interesting the idea of your inner child as somebody who you're much stronger than, somebody who they respect because you've developed really powerful psychological coping mechanisms they could never dream of, so that you're a protector figure.

Ozy talks about this a lot in the context of their borderline personality disorder. I tend to think of a lot of symptoms of borderline as being associated with neoteny – a preservation of childlikeness into adulthood (I don't know how orthodox this is). For Ozy, everything is *still* raw, maybe will always be raw. Every even slightly good thing that happens delights them. Every even slightly bad thing that happen traumatizes them.

The flip side of childhood trauma is childhood wonder. When you're young, and to a lesser degree when you're a teenager and even in your early twenties, you have a great capacity to be amazed at the raw beauty of the world. As you grow older, you get less direct exposure to things as you have more and more schemas to put them in: "Oh, yeah, that's a beautiful sunset, it looks a lot like the five thousand other sunsets I've seen. I'll just tag it 'sunset' and move on." There's a big loss there, but there's a compensatory gain:

> No more–no more–Oh! never more, my heart,
> Canst thou be my sole world, my universe!
> Once all in all, but now a thing apart,
> Thou canst not be my blessing or my curse:
> The illusion's gone for ever, and thou art
> Insensible, I trust, but none the worse,
> And in thy stead I've got a deal of judgment,
> Though Heaven knows how it ever found a lodgment.

That last couplet really resonates with me. You tend to think of judgment and wisdom as something you gain by laborious cultivation. And here's Byron, saying "Somehow I seem to have gotten some good qualities. God only knows how that happened. Seriously, of all people, *me*?"

A lot of the time we make fun of teenagers for having crazy high libido. And then they grow older, and their sex drive calms down a little bit. I actually haven't checked whether anyone knows if this is due to objective reductions in hormone levels, or if maybe once you've gone on a couple of dates and been in a couple of relationships it's no longer quite so exciting.

But it's not just sex. There's this entire complex of teenage and early-twenties things around sex and extreme politics and mysticism and fashion, and some of it is praiseworthy in the sense of being really excited about new things, and part of it is just not having any idea what you’re doing, so that realistic opportunities and insane opportunities look about the same. And so you end up on this roller coaster of grandiose plans, inevitable letdowns, gnawing horrible fears, and unexpected relief. And then eventually you kind of bottom out and stop doing this.

I don’t know if this is biological either. Michael Vassar (and as far as I know no one else) theorizes about a “second puberty” in the late teens/early twenties where the brain starts to take on an adult form. There’s some evidence for – for example, this is the age at which a lot of previously latent mental disorders like schizophrenia develop. And there’s some evidence against – nobody had a conception of teenagerdom until like 1940s America or so. But it’s certainly a useful concept. Just as after puberty dies down you kind of naturally stop being so concerned about sex and acne and whatever, so after second puberty get a deal of judgment. You stop being so concerned about…what?

> What is the end of Fame ? ’tis but to fill
> A certain portion of uncertain paper :
> Some liken it to climbing up a hill,
> Whose summit, like all hills, is lost in vapour ;
> For this men write, speak, preach, and heroes kill,
> And bards burn what they call their ‘midnight taper’,
> To have, when the original is dust,
> A name, a wretched picture and worse bust.

Erikson calls the psychological crisis of the teenage years “identity versus role confusion”, and Reb Wiki’s commentary on his work adds that:

Erikson does note that the time of identity crisis for persons of genius is frequently prolonged. He further notes that in our industrial society, identity formation tends to be long, because it takes us so long to gain the skills needed for adulthood's tasks in our technological world. So… we do not have an exact time span in which to find ourselves. It doesn't happen automatically at eighteen or at twenty-one. A very approximate rule of thumb for our society would put the end somewhere in one's twenties.

Let me take a stab at that "persons of genius" exemption, since some of my friends whom I've gotten a chance to observe are probably smart enough to qualify.

Anyone even a little bit smarter than normal gets feted and celebrated as a kid. I remember my fourth grade teacher telling my parents during a conference that "your son needs to go into science so he can cure cancer." This is dumb. In a school of a thousand people, you can be the smartest kid in the school, more than smart enough to impress your teachers – and still be only one of the 300,000 smartest people in the country. If those other 300,000 people didn't cure cancer, there's a pretty good chance your son won't either. But when you're a kid, all you have to do to look smart is read the occasional science book and cultivate an interest in quarks. You can just go around saying "Did you know there are six types of quarks?" and everyone will think you're some kind of genius.

Then you grow older. You reach the point where nobody thinks you're a genius unless you can prove some kind of new result, which is a lot harder. You go to a good college, and suddenly you're in an environment preselected so that everybody else is about as smart as you are. If you've been coasting through life on being able to name all six types of quark (and who’s going to know if you get one wrong?) this is pretty disorienting.

And so part of Erikson’s “role confusion” is thinking “Wait, I was the guy who was going to cure cancer. I can feel my status slipping away from me as I become more and more mediocre. What am I going to do to prove that I really am that cool?”

I think a lot of the pathologies of adolescence are part of that urge, hollow promises of regaining lost status. The key is to provide a narrative in which you are great and which is impervious to external disconfirmation. Extremist politics, mysticism and fashion all fit the bill for different personalities.

Along with the pathologies there were the ill-advised adventures. “I’m going to be a great person by…um…exercising an hour a day, from now on, all the time, and eventually becoming really buff.” Lasted a month. Then “I’m going to be a great person by…um…learning to speak ten languages, one at a time.” Lasted until first encounter with the Finnish case system. “I’m going to become a great person by…” The problem with all of these were that none of these were things I actually wanted to do (cf Randall Munroe, “Never trust anyone who’s more excited about success than about doing the thing they want to be successful at.”)

Actually, forget Randall Munroe. The best related quote is a different Monroe, who said that “although you are ambitious, you have no ambition.” And so:

> Ambition was my idol, which was broken
> Before the shrines of Sorrow, and of Pleasure ;
> And the two last have left me many a token

O’er which reflection may be made at leisure :
Now, like Friar Bacon’s Brazen Head, I’ve spoken,
‘Time is, Time was, Time’s past’ : a chymic treasure
Is glittering Youth, which I have spent betimes—
My heart in passion, and my head on rhymes.

One of the key points of the rationalist community is to learn to “optimize” rather than “satisfice” things, and it’s a useful lesson. But everyone sometimes needs reverse advice, and younger me – and younger lots of people – didn’t really understand satisficing.

When I was about ten, I decided to just optimize my entire life. I made a schedule of exactly what I would do every day – each minute filled with some sort of very productive character-building activity. Then I followed it for two days. Then I gave up and felt bad about it for a while.

That’s the sort of optimizing that only Young Scott could love. But I’ve been reading *On The Road* recently, and I wonder if the sort of Beat culture of authenticity is a different kind of optimizing, where you’re throwing everything at being different and more real, to the point of abandoning family and financial stability and whatever else.

There’s a place for this kind of optimization, if it’s what you want to do. But I eventually noticed that attempts to optimize my life and be maximally good were making me kind of miserable. I think that’s where the judgment part comes in. You learn when it’s okay to stop getting mad at yourself for not being perfect and take a little bit of time to relax and enjoy.

Byron is maybe a bad example of learning to overcome ambition, since he did kind of become super famous. But even that can be a kind of relaxing ambition. You learn what you’re good at, even if it’s something like poetry that might not be the most lucrative and world-changing thing around, and you focus on that. You’re not going to be Julius Caesar, but you might be Lord Byron. Or if not Lord Byron, you might at least have a career and be good at it. Role confusion gives way to identity.

(even MIRI, the most healthily ambitious people I know, have backed down from “we will save the world all by ourselves, right now” to “we will contribute an important part in an eventual effort to save the world”)

In fact, I think that’s the most important part of the solution, the part that makes it a little more dignified than abject surrender to being a cog in the machine. Vague formless ambition crystallizes into a couple of things that you’re good at and want to pursue, and then it doesn’t seem like ambition any longer. It just seems like the thing you’re doing.

Byron also got one other thing right, which was that he was able to sacrifice ambition to pleasure. This seems a better shrine to sacrifice at than “akrasia” or “conformity” or “vague feelings that I shouldn’t be doing this.”

> But I, being fond of true philosophy,
> Say very often to myself, ‘Alas!
> All things that have been born were born to die,
> And flesh (which Death mows down to hay) is grass ;
> You’ve passed your youth not so unpleasantly,
> And if you had it o’er again—’twould pass—
> So thank your stars that matters are no worse,
> And read your Bible, sir, and mind your purse.’

Seems like another riff on the same subject. Ambition and the raw energy of youth turning to a vague fondness that he got things *mostly* right, for a human.

I hate to change poets in midstream, but Chesterton says much the same:

…the doubts that drove us through the night as we two talked amain,
And day had broken on the streets e'er it broke upon the brain.
Between us, by the peace of God, such truth can now be told;
Yea, there is strength in striking root and good in growing old.
We have found common things at last and marriage and a creed,
And I may safely write it now, and you may safely read.

The theme to me seems the same. Youth is scary. Everything is important. Philosophy seems perilously close. Every tiny thing inspires doubts. Then "there is strength in striking root and good in growing old". You get a base. You know where you are standing. Things feel calmer and safer. You go from role confusion to identity.

Byron talks of "reading your Bible and minding your purse". Chesterton talks about "we have marriage and a creed". I read these as kind of similar. It's about finding an ideology – in contrast to the constant ideology-searching of youth where you get your Communists and your Daoist and your anarchists and whatever. And then it's about turning to be more interested in the everyday world of things like marriage and family and relationships and balancing your checkbook.

If this were about suddenly ceasing to care about ideas, then it would be monstrous and I'd be trying to resist it every way I can. But neither Chesterton nor Byron became intellectual lightweights in their old age. I think of it as getting to participate in the world of ideas because you want to, rather than because you *have to*. In Jung's words, "swimming rather than drowning". Or since the ocean of thought is maybe too big for a swimming metaphor, you're still out at sea, but you've got a nice sturdy ship instead of a Neurath's boat where you have to build your vessel while you're sailing on it.

In an unhealthy society, it can be dangerous to lose revolutionary fervor. But in a healthy society, it seems to be a natural and important process. I don't know if our society is healthy enough for me to be entirely comfortable with it. There are a lot of people who can't get a stable career, people who are trying as hard as they can. But even in a revolution you need a couple of people to keep things running and maybe donate money earned at a stable job to the people with more zeal (see: Engels), and in the spirit of satisficing rather than optimizing I'm pretty okay with this role.

…or maybe you stay an anarchist or a Daoist or a communist. But then it's because you're set in that philosophy and you like it and you're making a stand there, rather than because it's your Experiment Of The Month. It's *good* to have Experiments Of The Month – high expected value of information, low transaction costs for changing your mind – but it's also a relief to be done with that. Identity in place of role confusion. As for the adult world of relationships and checkbooks, it's a different and lower-variance way of contributing to the community, and if you're lucky you can have kids and start the whole cycle over again.

Yesterday I turned thirty years old. People keep asking me how I feel about it. I think I agree with Byron. I passed my youth not too unpleasantly. And if I had it over again, it'd pass. So thank the stars that matters are no worse.

How To Use 23andMe Irresponsibly

As you might remember, the FDA stomped on 23andMe for using too many irresponsible genetic tests that purported to tell you things about yourself and your health with limited support. They eventually worked out a deal where the FDA allowed 23andMe to continue to operate, but they couldn't claim to be able to predict personal outcomes from your genes.

That means if we want to use 23andMe irresponsibly, we've got to do it ourselves. Luckily I recently figured out how to do this and it is exactly as much fun as you would think.

If you've got a 23andMe account, log in, go to your name and picture on the bar on the top, and click on the little inverted triangle to get the drop-down menu. Go to the "Browse Raw Data" option, which will give you the option to go to a gene or an SNP. Now all you have to do is find an SNP you're interested in (an SNP will look like the letters "rs" followed by a string of numbers) plug it in, and interpret the results.

Your best bet here is SNPedia, a wiki collection of different SNPs. If you want to know, for example, something interesting about your risk of heart disease, you can search "heart disease" and get a list of the most relevant SNPs (in this case, rs2383206, rs10757278, rs2383207, and rs10757274). If you click on the first, you can find on the top right in little colored boxes that someone with (A;A) at this site has normal risk of heart disease, someone with (A;G) 1.4x increased risk, and someone with (G;G) 1.7x increased risk.

In this case my 23andMe results are pretty straightforward – it tells me I am (G;G), which is common enough in white people (see the little colored bars on the left of SNPedia; the CEU bar is Caucasian Europeans). Other times the results require an extra step. For example, SNPedia's page on rs1801133 offers three choices – (C;C), (C;T), and (T;T), but 23andMe tells me that I have (A;A), which didn't appear to be an option. The problem here is that 23andMe is giving me the minus strand – if you click to expand your result, it will tell you that ("dbSNP Orientation: minus"). When it gives you the minus strand, you have to manually reverse it to get the plus strand. Remember, A is the reverse of T, and C is the reverse of G. So my (A;A) is their (T;T), and I have 1.5x risk of various cancers.

This doesn't necessarily bear any relationship to reality, because genetics studies often fail to replicate, and even when they're right they might only apply to certain populations, and even when they apply to people usually people misinterpret what they mean. That's part of why the FDA banned 23andMe from doing this, and part of why the word "irresponsible" is right in the title. Even if these SNPs survive the tests of time and replication, they will explain at most a few percent of the variance in complex traits, and any claims otherwise are exaggeration at best and pure hype at worst.

But with that fair warning, here are some of the genes I think are most fun to look up. *I cannot disclaimer enough that this is for your own amusement only and unlikely to resemble reality in more than the most tenuous way and if I imply otherwise it is a silly joke.*

Rs909525 is linked to the so-called "warrior gene" which I blogged about in the last links roundup. People with the normal four or five repeat version of these gene are less violent than people with the three-repeat version, and people with the two-repeat version are massively overrepresented among violent criminals. See for example this article. Although this SNP isn't the warrior gene itself, it's linked to it closely enough to be a good predictor. This is on the X chromosome, so men will only have one copy (I wonder how much of the increased propensity to violence in men this explains). It's also one of the minus strand ones, so it'll be the reverse of what SNPedia is telling you. If you've got T, you're normal. If you've got C, you're a "warrior". I've got C, which gives a pretty good upper limit on how much you should trust these SNPs, since I'm about the least violent person you'll ever meet. But who knows? Maybe I'm just waiting to snap. Post something dumb about race or gender in the open thread *one* more time, I dare you…

Rs53576 in the OXTR gene is related to the oxytocin receptor, which frequently gets good press as "the cuddle hormone" and "the trust hormone". Unsurprisingly, the polymorphism is related to emotional warmth, gregariousness versus loneliness, and (intriguingly) ability to pick out conversations in noisy areas. 23andMe reads this one off the plus strand, so your results should directly correspond to SNPedia's – (G;G) means more empathy and sociability and is present in 50% of the population, anything else means less. I'm (A;G), which I guess explains my generally hateful and misanthropic outlook on life, plus why I can never hear anyone in crowded bars.

Rs4680 is in the COMT gene, which codes for catechol-o-methyltransferase, an enzyme that degrades various chemicals including dopamine. Riffing on the more famous "warrior gene", somebody with a terrible sense of humor named this one the "worrier gene". One version seems to produce more anxiety but slightly better memory and attention; the other version seems to produce calm and resiliency but with a little bit worse memory and attention. (A;A) is smart and anxious, (G;G) is dumb and calm, (A;G) is in between. if you check the SNPedia page, you can also find ten zillion studies on which drugs you are slightly more likely to become addicted to. And here's the 23andMe blog on this polymorphism.

Rs7632287, also in the oxytocin receptor, has been completely proportionally and without any hype declared by the media to be "the divorce gene". To be fair, this is based on some pretty good Swedish studies finding that women with a certain allele were more often to have reported "marital crisis with the threat of divorce" in the past year (p = 0.003, but the absolute numbers were only 11% of women with one allele vs. 16% of women with the other). This actually sort of checks out, since oxytocin is related to pair bonding. If I'm reading the article right (G;G) is lower divorce risk, (A;A) and (A;G) are higher – but this may only apply to women.

Rs11174811 is in the AVPR1A gene, part of a receptor for a chemical called vasopressin which is very similar to oxytocin. In case you expected men to get away without a divorce gene, this site has been associated with spousal satisfaction in men. Although the paper is extremely cryptic, I think (A;A) or (A;C) means higher spousal satisfaction than (C;C). But if I'm wrong, no problem – another study got the opposite results.

Rs25531 is on the serotonin transporter. Its Overhyped Media Name is "the orchid gene", on the basis of a theory that children with one allele have higher variance – that is, if they have nice, happy childhoods with plenty of care and support they will bloom to become beautiful orchids, but if they have bad childhoods they will be completely screwed up. The other allele will do moderately well regardless. (T;T) is orchid, (C;C) is moderately fine no matter what. There are rumors going around that 23andMe screwed this one up and nearly everybody is listed as (C;C).

Rs1800955 is in DRD4, a dopamine receptor gene. Its overhyped media name is The Adventure Gene, and supposedly one allele means you're much more attracted to novelty and adventure. And by "novelty and adventure", they mean lots and lots of recreational drugs. This one has survived a meta-analytic review. (T;T) is normal, (C;C) is slightly more novelty seeking and prone to drug addiction.

Rs2760118, in a gene producing an obscure enzyme called succinate semialdehyde dehydrogenase, is a nice polymorphism to have. According to this article, it makes you smarter *and* can be associated with up to fifteen years longer life (warning: impressive result means almost certain failure to replicate). (C;C) or (C;T) means you're smarter and can expect to live longer; (T;T) better start looking at coffins sooner rather than later.

Rs6311 is not going to let me blame the media for its particular form of hype. The official published scientific paper on it is called "The Secret Ingredient for Social Success of Young Males: A Functional Polymorphism in the 5HT2A Serotonin Receptor Gene". Boys with (A;A) are less popular than those with (G;G), with (A;G) in between – the effect seems to be partly mediated by rule-breaking behavior, aggression, and number of female friends. Now it kind of looks to me like they're just taking proxies for popularity here, but maybe that's just what an (A;A) nerd like me *would* say. Anyway, at least I have some compensation – the popular (G;G) guys are 3.6x more likely to experience sexual side effects when taking SSRI antidepressants.

Rs6265, known as Val66Met to its friends, is part of the important depression-linked BDNF system. It's a bit depressing itself, in that it is linked to an ability not to become depressed when subjected to "persistent social defeat". The majority of whites have (G;G) – the minority with (A;A) or (A;G) are harder to depress, but more introverted and worse at motor skills.

rs41310927 is so cutting-edge it's not even in SNPedia yet. But these people noticed that a certain version was heavily selected for in certain ethnic groups, especially Chinese, and tried to figure out what those ethnic groups had in common. The answer they came up with was "tonal languages", so they tested to see if the gene improved ability to detect tones, and sure enough they claimed that in experiments people with a certain allele were better able to distinguish and understand them. Usual caveats apply, but if you want to believe, (G;G) is highest ability to differentiate tones, (A;A) is lowest ability to differentiate tones. (A;G) is in between. Sure enough, I'm (A;A). All you people who tried to teach me Chinese tonology, I FRICKIN' TOLD YOU ALL OF THE WORDS YOU WERE TELLING ME SOUNDED ALIKE.

The Dark Side Of Divorce

A while ago I read *The Nurture Assumption* and found myself convinced by its basic thesis that genetics completely trumped parenting.

The argument was that there are lots of studies showing that parenting has important effects – for example, if parents yell at their kids, their kids will turn out angry and violent, or something. But these studies neglect possible genetic contributions – angry violent parents are more likely to yell at their kids, so maybe the kids are just inheriting genes for anger and violence. A lot of parenting studies are subject to these kinds of confounds. And one of the best tools we have for disentangling them – behavioral genetics twin studies – very consistently show that most important outcomes are 50% genetically determined, 50% determined by "non-shared environment", and almost completely unrelated to the "shared environment" of parenting. Therefore, we should conclude that pretty much all of the effect supposedly due to parenting is in fact due to genetics, and it doesn't matter much what kind of "parenting style" you use unless it can somehow change your child's DNA.

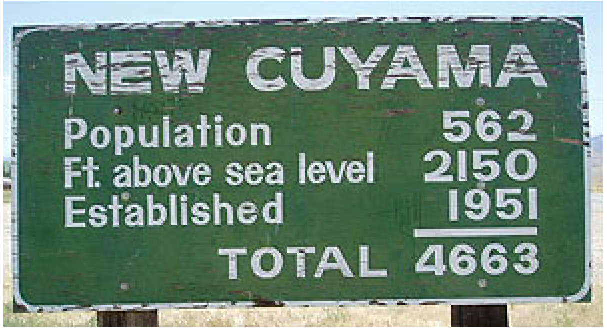
One of the stories I most remember from the book – and I'm sorry, I don't have a copy with me, so I'm going from memory – was about the large literature of studies showing that children of divorce raised by single mothers have worse outcomes than children of intact two-parent families. This seems like a convincing argument that children need both parents to develop properly, which if true would be a shared environmental effect and an example of why good stable parenting is necessary.

But other studies found that children who lost a father in (for example) a car accident had outcomes that looked more like those of children from stable two-parent families than like those of children of divorce. So maybe the divorce effect doesn't reflect the stabilizing influence of two parents in a kid's life. Maybe it reflects that the sort of genes that make parents unable to hold a marriage together have some bad effects on their kids as well.

(damn you, Rs7632287! This is all *your* fault!)

It's compelling, it's believable, and I believed it. Unfortunately, I recently had the time to double-check, and it doesn't seem to be true at all.

The best introduction to divorce research I could find was Amato & Keith's meta-analysis Parental Divorce And The Well-Being Of Children. It looks through 92 studies that compare children of divorced and non-divorced families and finds that "children of divorce scored lower than children in intact families across a variety of outcomes, with the median effect size being 0.14 of a standard deviation," this last clause of which is almost New Cuyaman in its agglomerativeness.



This is a small effect size, and indeed most of the studies they're looking at aren't even significant. But once agglomerated together they become very significant, and the analysis tries to determine the cause. The most popular proposed causes are “children in divorced families lose the benefits of having two parents”, “children in divorced families are in economic trouble”, and “children in divorced families have to deal with stressful family conflict.”

Although there’s a little bit of evidence for all three, *in general* the evidence lines up for the last one of these – the family conflict hypothesis.

If the problem is not enough parents or not enough money, then having the custodial parent (usually a single mom) remarry ought to help a lot, especially if she marries somebody wealthy. But usually this doesn’t help very much at all.

If the problem is not enough money, then children of divorce should do no worse than children of poor two-parent families. But in fact they do, and children of divorce still do worse when controlled for income.

If the problem is not enough parents or not enough money, then these ought to persist over time if the custodial parent doesn’t remarry or get richer. But if the problem is stressful conflict, then it ought to get better over time, since the stress and conflict of the divorce gradually becomes more and more remote. Although there are some dueling studies here, the best studies seem to find the latter pattern – bad outcomes of divorce gradually decrease over time.

If the problem is stressful conflict, then children of divorce ought to do no worse than children in families full of stressful conflict who are nevertheless staying together. Indeed, controlling for the amount of stressful conflict within a family gets rid of most of the negative effect of divorce.

Therefore, although there was some evidence for all three hypotheses, the stressful conflict hypothesis was best-supported. But the stressful conflict hypothesis could also explain the pattern where kids whose fathers died in car accidents don’t show the same pattern of problems as children of divorce. Having a parent die in an accident is no doubt traumatic, but it’s a very different kind of trauma from constant familial yelling and bickering.

More to the point, the genetic explanation of divorce has been investigated specifically in at least four studies that I know of, using different methodology each time.

Brodzinsky, Hitt, and Smith studied the effect of divorce on biological versus adopted children. They were unable to find any differences in the level of disruption and poor outcomes.

O’Connor, Caspi, DeFries, and Plomin (yes, *that* Plomin) also studied biological versus adopted children. They found that biological children showed a stronger effect on academic achievement and social adjustment (consistent with genetic explanations), but adopted children showed an equal effect on behavioral problems and substance use (consistent with environmental explanations).

Burt, Barnes, McGue, and Iacono use a different methodology and compare children whose parents divorced when they were alive with children whose parents divorced before they were born. Presumably, only the former group get any environmental stress from the divorce, but both groups suffer from any genetic issues that caused their parents to split. They find that the negative effects of divorce are mostly limited to the group whose parents got divorced when they were alive, consistent with an environmental explanation.

Finally, a bunch of people including Eric Turkheimer get the requisite twin study in and compare the children of pairs of identical twins where one of them got divorced and the other didn't (where do they *find* these people?) Somehow they scraped together a sample size of 2,554 people, and they found that even among children of identical twins, the children of the divorced twin did worse than the children of the non-divorced twin to a degree consistent with the negative effects not being genetic. They tried to adjust for characteristics of the twins' spouses, but that's the obvious confound here. I look forward to seeing if future researchers can get a sample of pairs of identical male twins who married pairs of identical female twins, one couple among whom got divorced.

So I owe mainstream psychology an apology here. I was pretty sure they had just completely dropped the ball on this one and were foolishly assuming everything had to be social and nothing could be genetic. In fact, they were only doing that up until about ten or twenty years ago, after which point they figured it out and performed a lot of studies, all of which supported their idea of the stress of divorce having significant (though small!) non-gene-related effects.

And although I haven't had time to look through them properly yet, here's a study claiming that the association between fathers' and childrens' emotional and behavioral problems is "largely shared environmental in origin". And here's a study claiming that "analyses revealed that [shared environment] accounted for 10%-19% of the variance within conduct disorder, oppositional defiant disorder, anxiety, depression, and broad internalizing and externalizing disorders, regardless of their operationalization. When age, informant, and sex effects were considered, [shared environment] generally ranged from 10%-30% of the variance."

So the shared environment folks haven't completely dropped the ball, some of them seem to be fighting back, and it will be interesting to see where this goes and whether anybody is able to reconcile the different evidence.

One likely talking point: shared environment and childhood situation obviously impacts things during childhood. For example, if you have parents who are mean and abusive, this can make you stressed and you don't get enough sleep and then maybe you do really badly at school. But once you get out of that environment, your academic abilities will revert to whatever your genes say they should be. *The Nurture Assumption* never denies this and is absolutely willing to admit that shared environment can affect outcomes *during childhood*, although even there less than one might expect. This also seems to be the tack Plomin is taking when he discusses the Burt study.

But studies have found that the negative effects of divorce can last well into adulthood. On the other hand, none of those studies have been the ones that compare genetic and environmental effects, and I get the feeling their quality is kind of weak. So it's not *completely* ruled out by the data that the short-term effects of divorce are robust and environmental, but the long-term effects of divorce are spurious and/or genetic. But this seems kind of like fighting a rearguard action against the evidence.

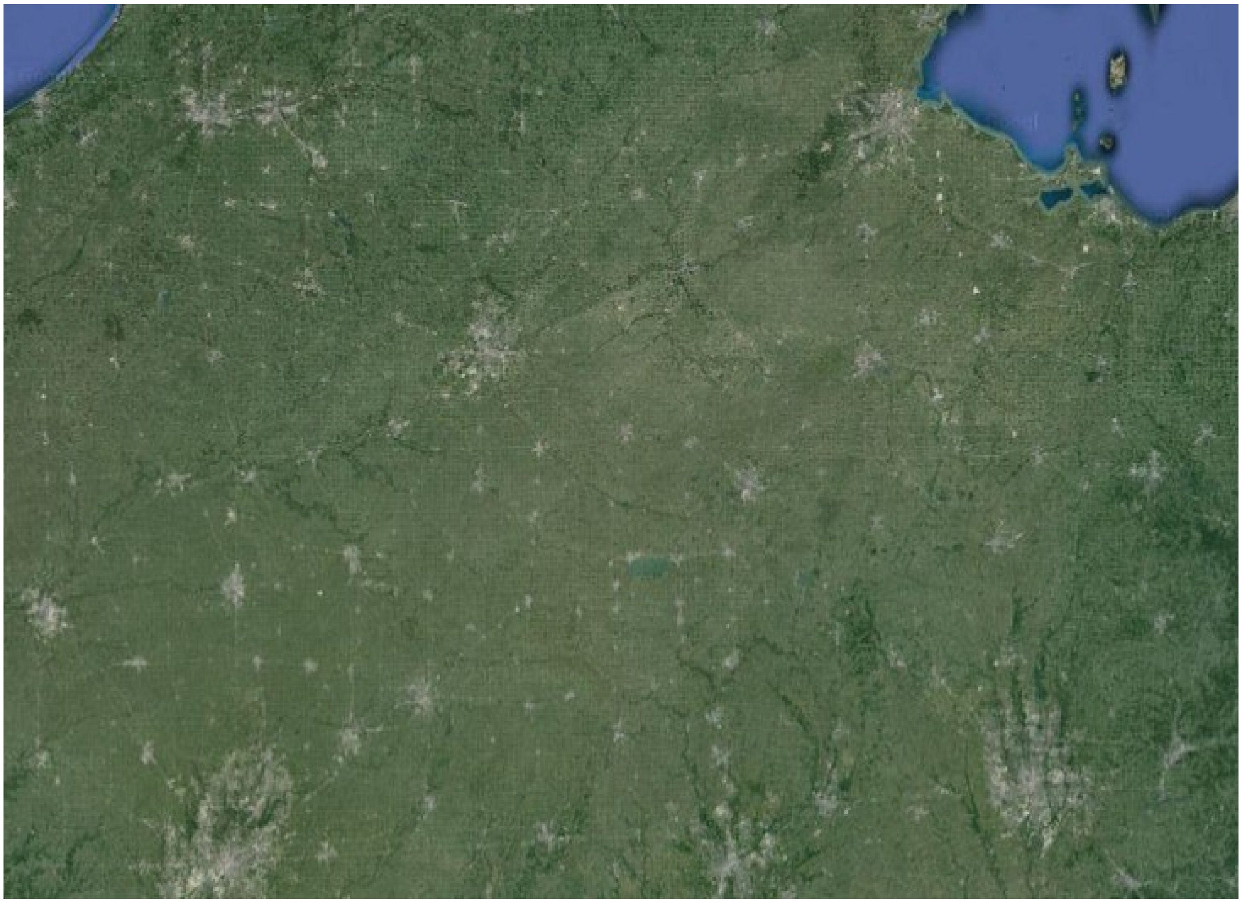
Finally, a sanity check. Suppose your parents get divorced when you're 16. Your high school grades drop and your behavior gets worse. Maybe you fail a couple of classes and start using drugs. The couple of classes failed mean you're going to a second-tier instead of a first-tier college, and the drug use means you're addicted. How does that not affect your life outcomes, even if five years later you've forgotten all about whatever psychological stresses you once had?

Overall I am less confident than before that shared environment is harmless.

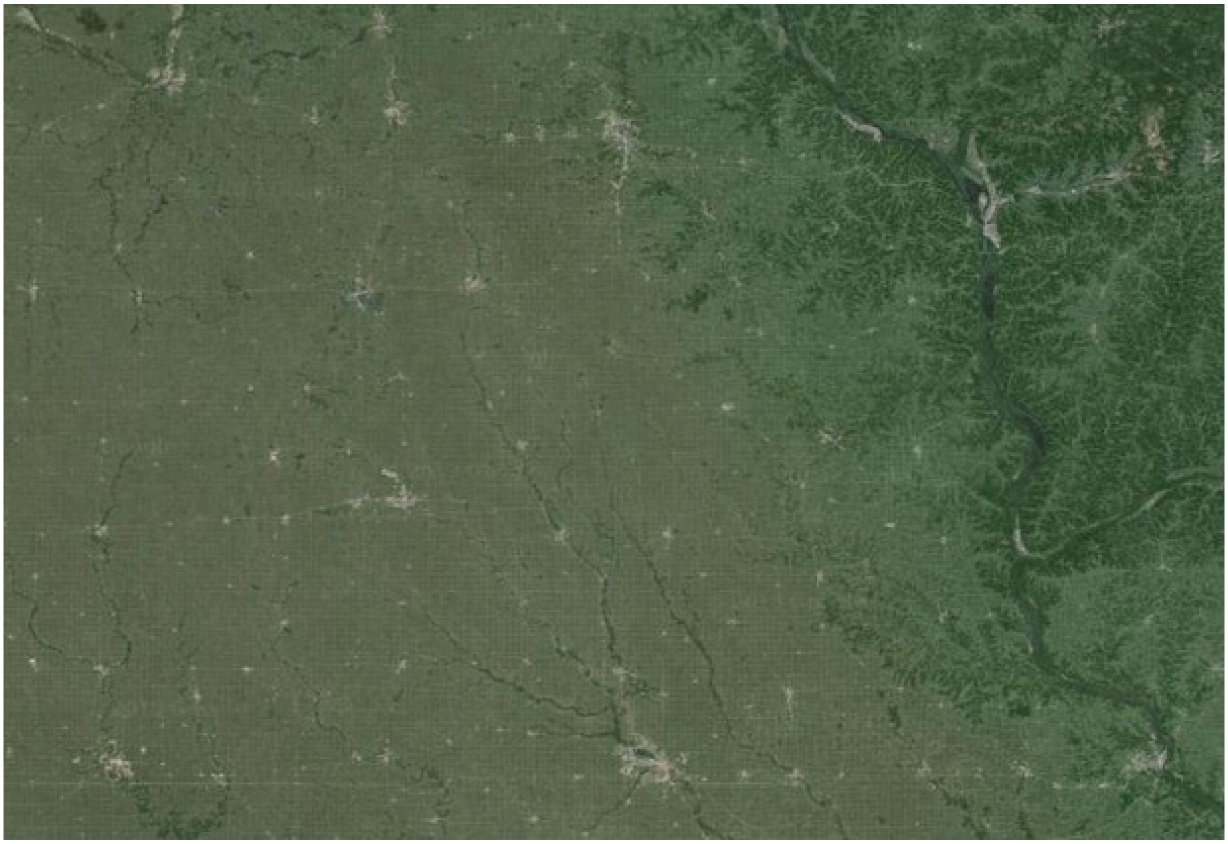
And while I’m bashing *Nurture Assumption*, I don’t remember the exact arguments used against birth order effects, but we found such impressive numbers on the last Less Wrong survey that I’m not very impressed with the claims that they don’t exist.

Ley Lines Of The Midwest

This is now unexpectedly a geography blog.



That’s Ohio and Indiana as seen from space.

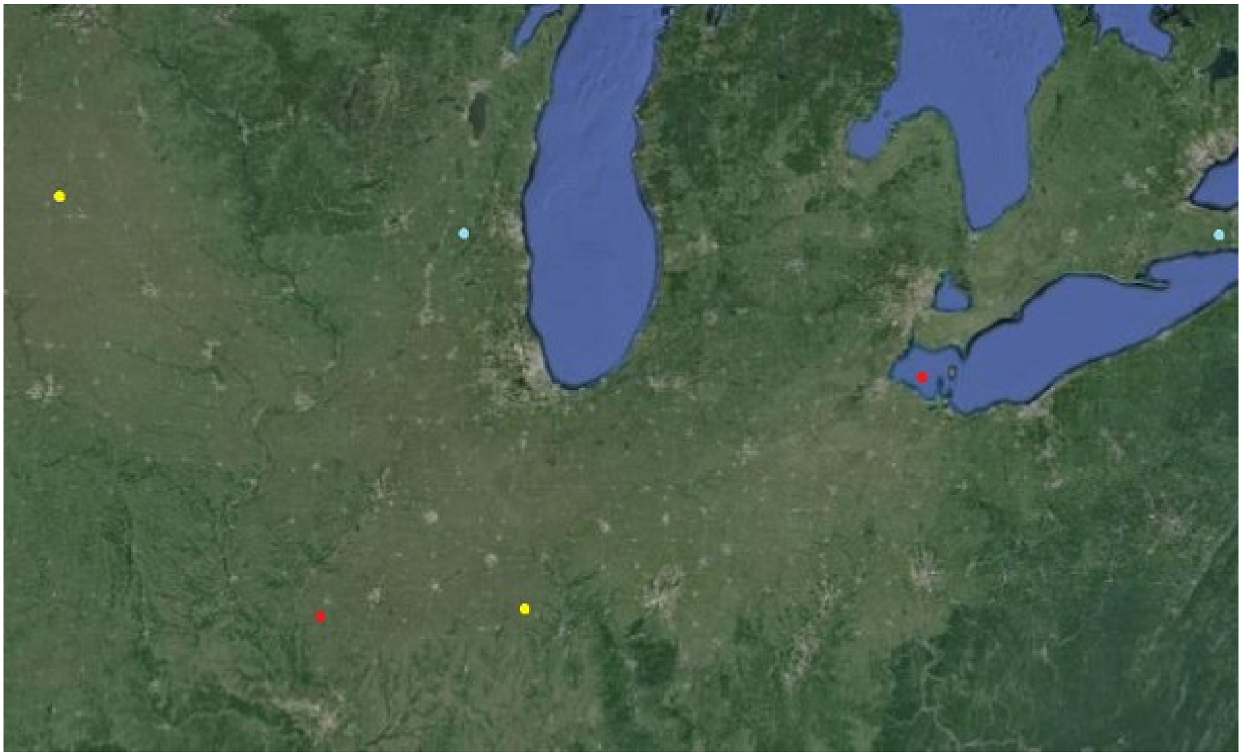


And that’s Northern Iowa and Southern Minnesota as seen from space.

Each small grey dot is a town. Are there straight horizontal or vertical lines that connect more than a chance number of towns? And are the towns arranged in a consistent coordinate grid pattern?

As best I can tell there are a few *short* straight lines probably representing more-used local roads, but few that persist across entire states. I don’t think there’s any consistent grid pattern. This is the opposite of my initial impression, which was that there was a clear and striking coordinate grid. But when I try to measure the native unit of the coordinate grid, I find that my mind is confusing a whole bunch of vaguely square-ish patterns into one illusory system.

There *is* a square pattern to the Midwest, deriving from the Public Land Survey System, but its scale is 6 mile x 6 mile squares, which is smaller than any of the distances on either of these maps. There is supposedly a higher level of grid, the 24 mile x 24 mile quadrangle, but it doesn't seem to be as important and I don't see that on this map either.



I titled this post "Ley Lines" as a joke, but we might as well see if there are any actual ley lines. The best candidates seem to be the cities between the yellow dots – which are Waterloo, Davenport, Peoria, Bloomington, and Champaign – and the cities between the red dots – which are Springfield, Champaign, Lafayette, Fort Wayne, and Toledo. If you want to stretch it, you could also imagine a horizontal line between the blue dots – Madison, Milwaukee, Grand Rapids, Flint, Sarnia, London, and continuing to Buffalo just off the map.

As far as I know there's no explanation for any of these – no highways, no rivers, nothing – and they're all just coincidences.

We had some interesting discussions about Midwestern geography during our last Michigan Rationalist Meetup. My favorite part was learning that the town of Zilwaukee, Michigan was named by two brothers hoping that would-be settlers on their way to Milwaukee would get confused and settle there instead. It sounds like a dumb urban legend, but it was previously admitted to on the Zilwaukee city website. I notice their new website doesn't mention this, which means either that it's been disproven or they decided to stop advertising to the world that they're descended from morons.

[**EDIT:** And here's a church website that uses Zilwaukee as a metaphor for the Devil]

Republicans Are Douchebags

Or, more technically, douchebags are disproportionately Republican. But I figure with this title I'm guaranteed front-page links from Salon and Daily Kos.

A while back, I argued – not especially originally – that "conservative" and "liberal", far from being mere descriptions of political views, pointed to two very different tribes of people who might as well be considered totally different ethnicities.

One marker of ethnicity is different name preferences – we all know what groups people named Juan, Tyrone, or Mei are likely to belong to – and a recent article in Vox confirms that names differ between Democrats and Republicans at very impressive rates. For example, of the 200,000 registered US voters named "Willie", 81.8% are Democrats. Of the 40,000 registered voters named "Rex", 59.4% are Republicans (and I assume the others are Rottweilers or tyrannosaurs). You can find some impressively complete statistics at this site, including what percent of people with your name have a gun, go to church, attend college, et cetera.

But looking through Vox's list of most Republican names, I was struck (or possibly stricken) by a resemblance to a different list I had seen a couple years ago.

Reddit: I fear my first name is the biggest douche bag name an American male can have. In your opinion, what is the cliche douchebag character name?.

This seems like a relatively popular internet question, and thetoptens.com maintains a Most Douchebag Names list as well. This provides two independent lists of douchiest names (my Reddit list is the first name proposed in the ten most upvoted first-level comments there). They both turn out to be pretty similar.

THETOPTENS:
1. Chad
2. Trent
3. Guy
4. Brad
5. Paul
6. Blake
7. Brody
8. Chaz
9. Tad
10. Keith

REDDIT:
1. Chad
2. Chase
3. Tyler
4. Brody
5. Brad
6. Trey
7. Hunter
8. Scott (@#$% YOU TOO, REDDIT)
9. Biff
10. Preston

Clarity Campaigns can tell us what percentile each of these names are on the political spectrum. When I plugged all of them in, the median douchebag name was in the 98.5th percentile for Republicanness. In other words, with a *little* bit of noise the top ten douchiest names are pretty much the top ten most Republican names.

(The big exception is “Chaz”, which leans Democrat. But I refuse to believe that “Chaz” is a real name anyway.)

I tried to test alternate hypotheses that Clarity just over-Republicanned all names, or that it was a function of these being male names, or white names, or names of a certain generation. I tested the top ten most popular male baby names of 1990 (that being the generation probably in its peak douchebag years right now) and combined their full name and nickname versions (since I didn’t want to confound by whether Republicans or Democrats are more likely to go by a nickname). The median popular 1990 male name was in the 73rd percentile for Republicanness. This isn’t surprising – men tend to be more conservative than women, and this effect probably swamps any within-gender name effects, so if all male names are more conservative than all female names we would expect the average male name to be about the 75th percentile for Republicanness. Our popular 1990 control group comes very close.

But the average douchebag name is in the 98.5th percentile for Republicanness.

I can think of ~~two~~ three hypotheses.

First, douchebags are disproportionately Republican.

Second, the parents who name kids douchebag names are disproportionately Republican, and Republicanism is partly hereditary (I almost missed this one, but JayMan reads this blog and I know he would call me on it if I forgot).

Third, “douchebag” is a tribally-coded slur. If someone asks “Have you ever noticed that all assholes are named things like ‘Moishe’ or ‘Avram’ or ‘Menachem’?” – then they’re telling you a lot more about the way they use the word ‘asshole’ than about the Moishes and Menachems of the world.

I expect there are many more fun things I will think of to do with this name list.

Prisons Are Built With Bricks Of Law And Brothels With Bricks Of Religion, But That Doesn’t Prove A Causal Relationship

Research Suggests Psychiatric Interventions Like Admission To A Mental Hospital Could Increase Suicide Risk says an Alternet article about a study that specifically mentions that it should not be used to conclude that psychiatric interventions like admission to a mental hospital could increase suicide risk.

But I wouldn’t be so worried if it wasn’t based on a very similar editorial written by field experts and published in the Journal of Social Psychiatry and Psychiatric Epidemiology.

The study involved is Rygaard-Hjorthøj, Madsen, Agerbo, and Nordentoft (2013), hereafter just “Hjorthøj” because I like saying that word. Hjorthøj finds that people who receive psychiatric treatment are much more likely to commit suicide than people who don’t. For example, someone who gets psychiatric medication is six times more likely to commit suicide than someone who doesn’t; someone who gets admitted to a psychiatric hospital is a whopping 44 times more likely to commit suicide than someone who doesn’t. The authors observe a “dose-response relationship”, which means that the more psychiatric treatment you get, the more likely you are to kill yourself.

Now, you’re probably asking yourself at this point “Wait, were they just using perfectly healthy people with no psychiatric problems as a control group?” and the answer is yes. Yes they were. So this study is *basically* finding that people who get committed to psychiatric hospitals are more likely to be the sort of people who are going to commit suicide than people who do not get committed to psychiatric hospitals. I for one find this result rather reassuring.

The authors of the study are absolutely on board with this, saying that “observational studies such as the present one cannot establish causality, but merely associations”, and their conclusion is that “not only people with a history of of psychiatric hospitalization, but also those receiving only psychiatric medication, outpatient treatment, or emergency room treatment should be monitored more closely”. Sure. If you absolutely must have a snappier conclusion than “psych patients often mentally ill, more at eleven,” I guess that fits the bill.

But according to an editorial published in the same journal by two people who are *not* the original authors, it says something much more sinister:

> The results of a study in this issue of the Journal…raise the disturbing possibility that psychiatric care might, at least in part, cause suicide.

A…bold hypothesis. Why should we privilege this hypothesis over the alternative possibility that suicidal people are more likely to seek (or get forced into) psychiatric treatment?

> The authors understandably caution that ‘the association is likely one of selection rather than causation, in that people with increasing levels of psychiatric contract are also more severely at risk of dying from suicide.’ This is undoubtedly part of the reason for the association, but it is not possible to be sure that an element of causation may not also be contributing. Associations that are strong, demonstrate a dose-effect relationship, and have a plausible

> mechanism are more likely to indicate a causal relationship than associations that lack these characteristics.

And then the Alternet article picks this up and adds a different argument:

> The Danish researchers argued that we were seeing the results of something like a cancer treatment study. Sicker people were appropriately getting into more intensive treatments, but unfortunately the sicker they were the more likely it was that they would still die, despite even the best of medicines. They also suggested that we may have therefore discovered the most accurate predictor of suicide we've ever found: The more someone seeks or is forced into psychiatric care, the closer they probably are on the trajectory towards suicide.
>
> The only problem with this line of reasoning is that there's no evidence to support it. Suicide is not a progressive illness like cancer; that is, there's no evidence that people with suicidal feelings travel on a trajectory of ever-intensifying, ever-more-constant suicidal feelings while getting into ever more intensive psychiatric care until they die at steadily increasing rates along the way. If suicidality was in fact progressive in that way, we'd be much better at identifying where people are along that path and intervening at the right time to prevent suicides. Instead, completed suicides tend to be impulsive, related to a myriad of cascading, confounding, unpredictable factors, not much more common overall in people diagnosed with mental disorders than in the general population, and most often surprising to even those closest to the victims.

Okay, let’s stop talking about psychiatric disease and shift to murder.

Probably the best risk factor for murder that you will ever find, better than being abused as a child or doing drugs or having the MAOA warrior gene or whatever, is “previous contact with the police”.

Murder is not “progressive” (shut up, neoreactionaries). Much like suicide, there’s no evidence that murderers “travel on a trajectory of ever-intensifying, ever-more-constant murderous feelings while getting into more intensive police custody until they kill at steadily increasing rates along the way.” Instead it seems to be “impulsive, related to a myriad of cascading, confounding, unpredictable factors, and surprising even to those closest to the perpetrators.”

The link between murder and previous contact with the police will be strong. For example, previous murderers released from prison have a 1.2% chance of getting arrested for another murder within three years, compared to about a 0.0001% murder rate per three years among the general population. That’s a relative risk of 10,000x, which blows Hjorthøj’s relative risk of 44x out of the water.

The link will be dose-dependent. People who have previously only gotten warnings from the police will be less likely to murder than people who have gotten small fines, who are less likely to murder than people who have gotten probation, who are less likely to murder than people who have gotten short jail sentences, who are less likely to murder than people who have gotten long jail sentences.

The link even has a plausible causal mechanism. Contact with the police can seriously disrupt people’s lives, making them stressed and anxious and angry and hopeless, all of which are the sort of emotions that predispose someone towards violence.

Therefore, the police cause murder?

Here are some other links that are non-progressive, strong, dose-dependent, and have plausible causal mechanisms.

The link between getting detention and dropping out of school. Therefore, detentions cause students to become demoralized and drop out from school.

The link between ice cream sales in a city and heatstroke cases in that city. Therefore, ice cream contains toxic chemicals that cause heatstroke.

The link between having lots of bruises and being in an abusive relationship. Therefore, abusers only abuse their victims because they’re angry about how many bruises they have.

The editorial authors seem to have gotten the “strong, dose-dependent, plausible” criteria from an article on epidemiology (God only knows where the journalist got the non-progressive criterion from). I would bet that the epidemiology article either did not intend for it to be used in this way, or that it meant that these criteria provide only the most tenuous of possible links.

This is why the saying is “correlation doesn’t imply causation” and not “correlation does not imply causation, unless it’s really *strong* correlation, in which case knock yourself out.”

And this is why the article finds that even going to a psychiatric emergency room and being turned down for treatment increases your risk of suicide almost twenty times. I mean, in *my* ER patients only even see a psychiatrist for like half an hour. You're saying a half an hour with a psychiatrist leads to a vigintupling of suicide rates months down the road? We might be bad. But we're not *that* bad.

The sad thing is, I think there might be a point buried underneath all this.

You can't conclude from an increased murder rate among people with criminal histories that the police cause murder. But the justice system *does* contribute to murder in its way by sticking hardened criminals together, traumatizing them, and failing to give them enough resources to rebuild their lives. The contribution of the criminal justice system to crime isn't exactly a secret, it's just not accessible with that methodology.

Likewise, I don't disagree that contact with the psychiatric system can sometimes be harmful. Forced commitment can sometimes make people lose their jobs, or cause them stigma, or stick them in an unpleasant psychiatric hospital where they don't want to be. While there are no doubt potential benefits as well, the weighing of the costs and benefits is something that hasn't been investigated nearly as much as it deserves. I think forced commitment is an overused tool and would be glad to get some evidence backing me up.

But this paper contributes *nothing* to the discussion. All we know is there's an association between psychiatric care and suicide, which was entirely obvious already. We don't know how much of that association is causal, how much of it is selection, and how much of it is "it would be even worse without psychiatric care but psychiatric care can’t do *everything*.

The exact effect of psychiatric care on suicide is a topic worthy of further high-quality research and discussion. But this isn’t it.

The Categories Were Made For Man, Not Man For The Categories

I.

“Silliest internet atheist argument” is a hotly contested title, but I have a special place in my heart for the people who occasionally try to prove Biblical fallibility by pointing out whales are not a type of fish.

(this is going to end up being a metaphor for something, so bear with me)

The argument goes like this. Jonah got swallowed by a whale. But the Bible says Jonah got swallowed by a big fish. So the Bible seems to think whales are just big fish. Therefore the Bible is fallible. Therefore, the Bible was not written by God.

The first problem here is that “whale” is just our own modern interpretation of the Bible. For all we know, Jonah was swallowed by a really really really big herring.

The second problem is that if the ancient Hebrews want to call whales a kind of fish, let them call whales a kind of fish.

I’m not making the weak and boring claim that since they’d never discovered genetics they don’t know better. I am making the much stronger claim that, even if the ancient Hebrews had taken enough of a break from murdering Philistines and building tabernacles to sequence the genomes of all known species of aquatic animals, there’s nothing whatsoever wrong, false, or incorrect with them calling a whale a fish.

Now, there’s something wrong with saying “whales are phylogenetically just as closely related to bass, herring, and salmon as these three are related to each other.” What’s wrong with the statement is that it’s false. But saying “whales are a kind of fish” isn’t.

Suppose you travel back in time to ancient Israel and try to explain to King Solomon that whales are a kind of mammal and not a kind of fish.

Your translator isn’t very good, so you pause to explain “fish” and “mammal” to Solomon. You tell him that fish is “the sort of thing herring, bass, and salmon are” and mammal is “the sort of thing cows, sheep, and pigs are”. Solomon tells you that your word “fish” is Hebrew *dag* and your word “mammal” is Hebrew *behemah*.

So you try again and say that a whale is a *behemah*, not a *dag*. Solomon laughs at you and says you’re an idiot.

You explain that you’re not an idiot, that in fact all kinds of animals have things called genes, and the genes of a whale are much closer to those of the other *behemah* than those of the *dag*.

Solomon says he’s never heard of these gene things before, and that maybe genetics is involved in your weird foreign words “fish” and “mammal”, but *dag* are just finned creatures that swim in the sea, and *behemah* are just legged creatures that walk on the Earth.

You try to explain that no, Solomon is wrong, *dag* are actually defined not by their swimming-in-sea-with-fins-ness, but by their genes.

Solomon says you didn’t even *know* the word *dag* ten minutes ago, and now suddenly you think you know what it means better than he does, who has been using it his entire life? Who died and made *you* an expert on Biblical Hebrew?

You try to explain that whales actually have tiny little hairs, too small to even see, just as cows and sheep and pigs have hair.

Solomon says oh God, you are so annoying, who the hell cares whether whales have tiny little hairs or not. In fact, the only thing Solomon cares about is whether responsibilities for his kingdom's production of blubber and whale oil should go under his Ministry of Dag or Ministry of Behemah. The Ministry of Dag is based on the coast and has a lot of people who work on ships. The Ministry of Behemah has a strong presence inland and lots of of people who hunt on horseback. So please (he continues) keep going about how whales have little tiny hairs.

It's easy to see that Solomon has a point, and that if he wants to define *behemah* as four-legged-land-dwellers that's his right, and no better or worse than your definition of "creatures in a certain part of the phylogenetic tree". Indeed, it might even be that if you spent ten years teaching Solomon all about the theory of genetics and evolution (which would be hilarious – think how annoyed the creationists would get) he might still say "That's very interesting, and I can see why we need a word to describe creatures closely related along the phylogenetic tree, but make up your own word, because *behemah* already means 'four-legged-land-dweller'."

Now imagine that instead of talking to King Solomon, you're talking to that guy from Duck Dynasty with the really crazy beard (I realize that may describe more than one person), who stands in for all uneducated rednecks in the same way King Solomon stands in for all Biblical Hebrews.

“Ah course a whale is a feesh, ya moron” he says in his heavy Southern accent.

“No it isn’t,” you say. “A fish is a creature phylogenetically related to various other fish, and with certain defining anatomical features. It says so right here in this biology textbook.”

“Well,” Crazy Beard Guy tells you, “Ah reckon that might be what a fish is, but a *feesh* is some’in that swims in the orshun.”

With a sinking feeling in your stomach, you spend ten years turning Crazy Beard Guy into a world expert on phylogenetics and evolutionary theory. Although the Duck Dynasty show becomes *much* more interesting, you fail to budge him a bit on the meaning of “feesh”.

It’s easy to see here that “fish” and “feesh” can be different just as “fish” and “*dag*” can be different.

You can point out how many important professors of icthyology in fancy suits use your definition, and how only a couple of people with really weird facial hair use his. But now you’re making a status argument, not a factual argument. Your argument is “conform to the way all the cool people use the word ‘fish'”, not “a whale is really and truly not a fish”.

There are facts of the matter on each individual point – whether a whale has fins, whether a whale lives in the ocean, whether a whale has tiny hairs, et cetera. But there is no fact of the matter on whether a whale is a fish. The argument is entirely semantic.

So this is the second reason why this particular objection to the Bible is silly. If God wants to call a whale a big fish, stop telling God what to do.

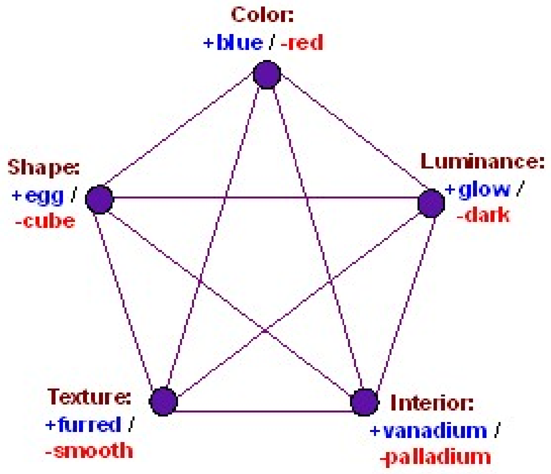
(also, bats)

II.

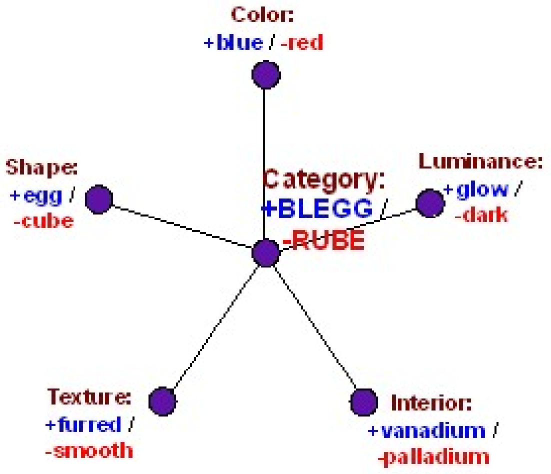
When terms are *not* defined directly by God, we need our own methods of dividing them into categories.

Less Wrong classic How An Algorithm Feels From The Inside starts with a discussion of whether or not Pluto is a planet. Planets tend to share many characteristics in common. For example, they are large, round, have normal shaped orbits lined up with the plane of the ecliptic, have cleared out a certain area of space, and are at least kind of close to the Sun as opposed to way out in the Oort Cloud.

One could imagine a brain that thought about these characteristics like Network 1 here:


Color:
+blue / -red
Shape:
+egg /
-cube
Luminance:
+glow /
-dark
Texture:
+furred /
-smooth
Interior:
+vanadium /
-palladium

Network 1


Color:
+blue / -red
Shape:
+egg /
-cube
Category:
+BLEGG /
-RUBE
Luminance:
+glow /
-dark
Texture:
+furred /
-smooth
Interior:
+vanadium /
-palladium

Network 2

One could imagine this model telling you everything you need to know. If an object is larger, it's more likely to be round and in cis-Neptunian space. If an object has failed to clear its orbit of debris, it's more likely to have a skewed orbit relative to the plane of the ecliptic. We could give each of these connections a weight and say things like large objects have a 32% chance of being in cis-Neptunian space and small objects an 86% chance. Or whatever.

But Network 1 has some big problems. For one thing, if you inscribe it in blood, you might accidentally summon the Devil. But for another, it’s computationally very complicated. Each attribute affects each other attribute which affects it in turn and so on in an infinite cycle, so that its behavior tends to be chaotic and unpredictable.

What people actually seem to do is more like Network 2: sweep all common correlations into one big category in the middle, thus dividing possibility-space into large round normal-orbit solitary inner objects, and small irregular skewed-orbit crowded outer objects. It calls the first category “planets” and the second category “planetoids”.

You can then sweep minor irregularities under the rug. Neptune is pretty far from the sun, but since it’s large, round, normal-orbit, and solitary, we know which way the evidence is leaning.

When an object satisfies about half the criteria for planet and half the criteria for planetoid, *then* it’s awkward. Pluto is the classic example. It’s relatively large, round, skewed orbit, solitary…ish? and outer-ish. What do you do?

The *practical* answer is you convene some very expensive meeting of prestigious astronomers and come to some official decision which everyone agrees to follow so they’re all on the same page.

But the *ideal* answer is you say “Huh, the assumption encoded in the word ‘planet’ that the five red criteria always went together and the five blue criteria always went together doesn’t hold. Whatever.”

Then you divide the solar system into three types of objects: planets, planetoids, and dammit-our-categorization-scheme-wasn’t-as-good-as-we-thought.

The situation with whales and fish is properly understood in the same context. Fish and mammals differ on a lot of axes. Fish generally live in the water, breathe through gills, have tails and fins, possess a certain hydrodynamic shape, lay eggs, and are in a certain part of the phylogenetic tree. Mammals generally live on land, breathe through lungs, have legs, give live birth, and are in another part of the phylogenetic tree. Most fish conform to all of the fish desiderata, and most mammals conform to all of the mammal desiderata, so there’s no question of how to categorize them. Occasionally you get something weird (a platypus, a lungfish, or a whale) and it’s a judgment call which you have to decide by fiat. In our case, that fiat is “use genetics and ignore all other characteristics” but some other language, culture, or scientific community might make a different fiat, and then the borders between their categories would look a little bit different.

III.

Since I shifted to a borders metaphor, let’s follow that and see where it goes.

Imagine that Israel and Palestine agree to a two-state solution with the final boundary to be drawn by the United Nations. You’re the head of the United Nations committee involved, so you get out a map and a pencil. Both sides swear to follow whatever you determine.

Your job is not to draw “the correct border”. There is no one correct border between Israel and Palestine. There are a couple of very strong candidates (for example, the pre-1967 line of control), but both countries have suggested deviations from that (most people think an actual solution would involve Palestine giving up some territory that has since been thoroughly settled by Israel in exchange for some territory within Israel proper, or perhaps for a continuous “land bridge” between the West Bank and Gaza). Even if you wanted to use the pre-1967 line as a starting point, there would still be a lot of work to do deciding what land swaps should and shouldn’t be made.

Instead you’d be making a series of trade-offs. Giving all of Jerusalem to the Israelis would make them very happy but anger Palestine. Creating a contiguous corridor between Gaza and the West Bank makes some sense, but then you’d be cutting off Eilat from the rest of Israel. Giving all of the Israeli settlements in the West Bank back to Palestine would satisfy a certain conception of property rights, but also leave a lot of Jews homeless.

There are also much stupider decisions you could make. You could give Tel Aviv to Palestine. You could make the Palestinian state a perfect circle five miles in radius centered on Rishon LeZion. You could just split the territory in half with a straight line, and give Israel the north and Palestine the south. All of these things would be really dumb.

But, crucially, they would not be *false*. They would not be *factually incorrect*. They would just be failing to achieve pretty much any of the goals that we would expect a person solving land disputes in the Middle East to have. You can think of alternative arrangements in which these wouldn’t be dumb. For example, if you’re a despot, and you want to make it very clear to both the Israelis and Palestinians that their opinions don’t matter and they should stop bothering you with annoying requests for arbitration, maybe splitting the country in half north-south is the way to go.

And real borders are, in fact, very weird.

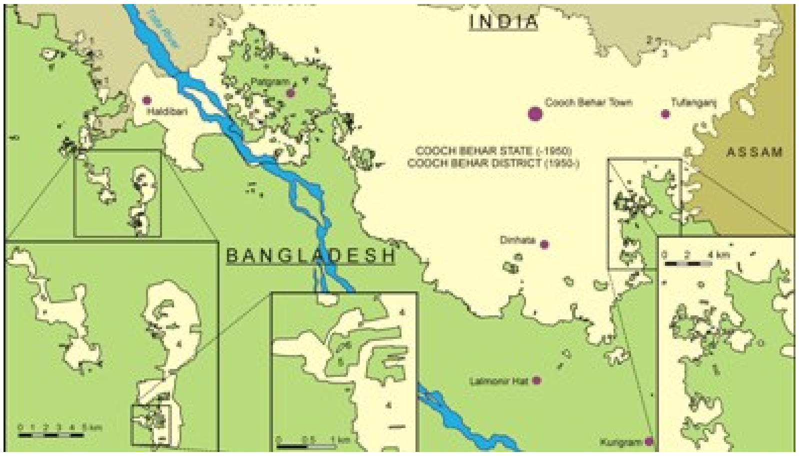
The border between Turkey and Syria follows a mostly straight-ish line near-ish the 36th parallel, except that about twenty miles south of the border Turkey controls a couple of square meters in the middle of a Syrian village. This is the tomb of the ancestor of the Ottoman Turks, and Turkey’s border agreement with Syria stipulates that it will remain part of Turkey forever. And the Turks take this *very* seriously; they maintain a platoon of special forces there and have recently been threatening war against Syria if their “territory” gets “invaded” in the current conflict.

Pictured: Turkey (inside fence), Syria (outside)

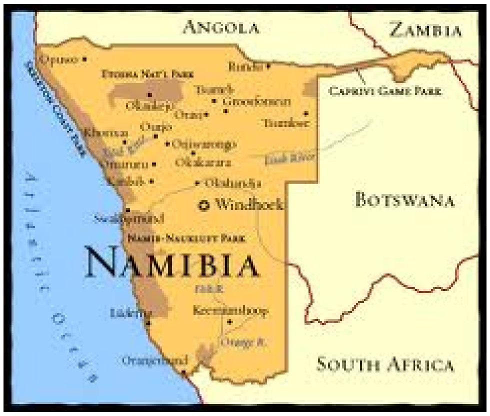
The border between Bangladesh and India is complicated at the best of times, but it becomes absolutely ridiculous in a place called Cooch-Behar, which I guess is as good a name as any for a place full of ridiculous things. In at least one spot there is an ‘island’ of Indian territory within a larger island of Bangladeshi territory within a larger island of Indian territory within Bangladesh. According to mentalfloss.com:

> So why’d the border get drawn like that? It can all be traced back to power struggles between local kings hundreds of years ago, who would try to claim pockets of

land inside each other's territories as a way to leverage political power. When Bangladesh became independent from India in 1947 (as East Pakistan until 1971), all those separate pockets of land were divvied up. Hence the polka-dotted mess.


INDIA
Patgram
Haldibari
Cooch Behar Town
Tufanganj
COOCH BEHAR STATE (-1950)
COOCH BEHAR DISTRICT (1950-)
ASSAM
BANGLADESH
Dinhata
Lalmonir Hat
Kurigram
0 1 2 3 4 5 km
0 0.5 1 km
0 2 4 km

Namibia is a very weird-looking country with a very thin three-hundred-mile-long panhandle (eg about twice as long as Oklahoma's). Apparently during the Scramble For Africa, the Germans who colonized Namibia really wanted access to the Zambezi River so they could reach the Indian Ocean and trade their colonial resources. They kept pestering the British who colonized Botswana until the Brits finally agreed to give up a tiny but very long strip of territory ending at the riverbank. This turned out to be not so useful, as *just* after Namibia's Zambezi access sits Victoria Falls, the largest waterfall in the world – meaning that any Germans who tried to traverse the Zambezi to reach the Indian Ocean would last a matter of minutes before suddenly encountering a four hundred foot drop and falling to pretty much certain death. The moral of the story is not to pester the British Empire too much, especially if they've explored Africa and you haven't.


ANGOLA
ZAMBIA
Opuwo
Etosha Nat'l Park
Rundu
Tsumeb
Grootfontein
Okaukuejo
Otavi
Tsumkwe
Caprivi Game Park
Skeleton Coast Park
Outjo
Khorixas
Otjiwarongo
Okakarara
Omaruru
Karibib
Okahandja
Windhoek
Botswana
Swakopmund
Namib-Naukluft Park
NAMIBIA
Atlantic Ocean
Fish R.
Lüderitz
Keetmanshoop
Orange R.
Oranjemund
SOUTH AFRICA

But the other moral of the story is that borders are weird. Although we think of borders as nice straight lines that separate people of different cultures, they can form giant panhandles, distant islands, and enclaves-within-enclaves-within-enclaves. They can depart from their usual course to pay honor to national founders, to preserve records of ancient conquests, or to connect to trade routes.

Hume’s ethics restrict “bad” to an instrumental criticism – you can condemn something as a bad way to achieve a certain goal, but not as morally bad independent of what the goal is. In the same way, borders can be bad at fulfilling your goals in drawing them, but not bad in an absolute sense or factually incorrect. Namibia’s border is bad from the perspective of Germans who want access to the Indian Ocean. But it’s *excellent* from the perspective of Englishmen who want to watch Germans plummet into the Lower Zambezi and get eaten by hippos.

Breaking out of the metaphor, the same is true of conceptual boundaries. You *may* draw the boundaries of the category “fish” any way you want. A category “fish” containing herring, dragonflies, and asteroids is going to be stupid, but only in the same sense that a Palestinian state centered around Tel Aviv would be stupid – it fails to fulfill any conceivable goals of the person designing it. Categories “fish” that do or don’t include whales may be appropriate for different people's purposes, the same way Palestinians might argue about whether the borders of their state should be optimized for military defensibility or for religious/cultural significance.

Statements like "the Zambezi River is full of angry hippos" are brute facts. Statements like "the Zambezi River is the territory of Namibia" are negotiable.

In the same way, statements like "whales have little hairs" are brute facts. Statements like "whales are not a kind of fish" are negotiable.

So it's important to keep these two sorts of statements separate, and remember that in no case can an agreed-upon set of borders or a category boundary be factually incorrect.

IV.

There is an anti-transgender argument that I take very seriously. The argument goes: we are rationalists. Our *entire shtick* is trying to believe what's actually true, not on what we wish were true, or what our culture tells us is true, or what it's popular to say is true. If a man thinks he's a woman, then we might (empathetically) wish he were a woman, other people might demand we call him a woman, and we might be much more popular if we say he's a woman. But if we're going to be rationalists who focus on believing what's actually true, then we've got to call him a man and take the consequences.

Thus Abraham Lincoln's famous riddle: "If you call a tail a leg, how many legs does a dog have?" And the answer: "Four – because a tail isn't a leg regardless of what you call it."

(if John Wilkes Booth had to suffer through that riddle, then I don't blame him)

I take this argument seriously, because sticking to the truth really is important. But having taken it seriously, I think it’s seriously wrong.

An alternative categorization system is not an error, and borders are not objectively true or false.

Just as we can come up with criteria for a definition of “planet”, we can come up with a definition of “man”. Absolutely typical men have Y chromosomes, have male genitalia, appreciate manly things like sports and lumberjackery, are romantically attracted to women, personally identify as male, wear male clothing like blue jeans, sing baritone in the opera, et cetera.

Some people satisfy some criteria of manhood and not others, in much the same way that Pluto satisfies only some criteria of planethood and whales satisfy only some criteria of mammalhood. For example, gay men might date other men and behave in effeminate ways. People with congenital androgen insensitivity syndrome might have female bodies, female external genitalia, and have been raised female their entire life, but when you look into their cells they have Y chromosomes.

Biologists defined by fiat that in cases of ambiguous animal grouping like whales, phylogenetics will be the tiebreaker. This was useful to resolve ambiguity, and it’s worth sticking to as a Schelling point so everyone’s using their words the same way, but it’s kind of arbitrary and mostly based on biologists caring a lot about phylogenetics. If we let King Solomon make the decision, he might decide by fiat that whether animals lived in land or water would be the tiebreaker, since he’s most interested in whether the animal is hunted on horseback or by boat.

Likewise, astronomers decided by fiat that something would be a planet if and only if meets the three criteria of orbiting, round, and orbit-clearing. But here we have a pretty neat window into how these kinds of decisions take place – you can read the history of the International Astronomical Union meeting where they settled on the definition and learn about all the alternative proposals that were floated and rejected and which particular politics resulted in the present criteria being selected among all the different possibilities. Here it is *obvious* that the decision was by fiat.

Without the input of any prestigious astronomers at all, most people seem to assume that the ultimate tiebreaker in man vs. woman questions is presence of a Y chromosome. I'm not sure this is a very principled decision, because I expect most people would classify congenital androgen insensitivity patients (XY people whose bodies are insensitive to the hormone that makes them look male, and so end up looking 100% female their entire lives and often not even knowing they have the condition) as women.

The project of the transgender movement is to propose a switch from using chromosomes as a tiebreaker to using self-identification as a tiebreaker.

(This isn't actually the whole story – some of the more sophisticated people want to split "sex" and "gender", so that people who want to talk about what chromosomes they've got have a categorization system to do that with, and a few people even want to split "chromosomal sex" and "anatomical sex" and "gender" and goodness knows what else – and I support all of these as very important examples of the virtue of precision – but to a first approximation, they want to define gender as self-identification)

This is not something that can be "true" or "false". It's a boundary-redrawing project. It can make for some boundaries that look a little bit weird – like a small percent of men being able to get pregnant – but as far as weird boundaries go that's probably not as bad as having a tiny exclave of Turkish territory in the middle of a Syrian village.

You draw category boundaries in specific ways to capture tradeoffs you care about. If you care about the sanctity of the tomb of your country's founder, sometimes it's worth having a slightly weird-looking boundary in order to protect and honor it. And if you care about…

I'm writing this post today because I just finished accepting a transgender man to the mental hospital. He alternates between trying to kill himself and trying to cut off various parts of his body because he's so distressed that he is biologically female. We've connected him with some endocrinologists who can hopefully get him started on male hormones, after which maybe he'll stop doing that and hopefully be able to lead a normal life.

If I'm willing to accept an unexpected chunk of Turkey deep inside Syrian territory to honor some random dead guy – and I better, or else a platoon of Turkish special forces will want to have a word with me – then I ought to accept an unexpected man or two deep inside the conceptual boundaries of what would normally be considered female if it'll save someone's life. There's no rule of rationality saying that I shouldn't, and there are plenty of rules of human decency saying that I should.

V.

I've made this argument before and gotten a reply something like this:

"Transgender is a psychiatric disorder. When people have psychiatric disorders, certainly it's right to sympathize and feel sorry for them and want to help them. But the way we try to help them is by treating their disorder, not by indulging them in their delusion."

I think these people expect me to argue that transgender "isn't really a psychiatric disorder" or something. But "psychiatric disorder" is just another category boundary dispute, and one that I've already written enough about elsewhere. At this point, I don't care enough to say much more than "If it's a psychiatric disorder, then attempts to help transgender people get covered by health insurance, and most of the transgender people I know seem to want that, so sure, gender dysphoria is a psychiatric disorder."

And then I think of the Hair Dryer Incident.

The Hair Dryer Incident was probably the biggest dispute I've seen in the mental hospital where I work. Most of the time all the psychiatrists get along and have pretty much the same opinion about important things, but people were at each other's *throats* about the Hair Dryer Incident.

Basically, this one obsessive compulsive woman would drive to work every morning and worry she had left the hair dryer on and it was going to burn down her house. So she'd drive back home to check that the hair dryer was off, then drive back to work, then worry that maybe she hadn't *really* checked well enough, then drive back, and so on ten or twenty times a day.

It's a pretty typical case of obsessive-compulsive disorder, but it was really interfering with her life. She worked some high-powered job – I think a lawyer – and she was *constantly* late to everything because of this driving back and forth, to the point where her career was in a downspin and she thought she would have to quit and go on disability. She wasn't able to go out with friends, she wasn't even able to go to restaurants because she would keep fretting she left the hair dryer on at home and have to rush back. She’d seen countless psychiatrists, psychologists, and counselors, she’d done all sorts of therapy, she’d taken every medication in the book, and none of them had helped.

So she came to my hospital and was seen by a colleague of mine, who told her “Hey, have you thought about just bringing the hair dryer with you?”

And it *worked*.

She would be driving to work in the morning, and she’d start worrying she’d left the hair dryer on and it was going to burn down her house, and so she’d look at the seat next to her, and there would be the hair dryer, right there. And she only had the one hair dryer, which was now accounted for. So she would let out a sigh of relief and keep driving to work.

And approximately half the psychiatrists at my hospital thought this was *absolutely scandalous*, and This Is Not How One Treats Obsessive Compulsive Disorder, and what if it got out to the broader psychiatric community that instead of giving all of these high-tech medications and sophisticated therapies we were just telling people to *put their hair dryers on the front seat of their car*?

But I think the guy deserved a medal. Here’s someone who was totally untreatable by the normal methods, with a debilitating condition, and a drop-dead simple intervention that nobody else had thought of gave her her life back. If one day I open up my own psychiatric practice, I am half-seriously considering using a picture of a hair dryer as the logo, just to let everyone know where I stand on this issue.

Miyamoto Musashi is quoted as saying:

> The primary thing when you take a sword in your hands is your intention to cut the enemy, whatever the means. Whenever you parry, hit, spring, strike or touch the enemy's cutting sword, you must cut the enemy in the same movement. It is essential to attain this. If you think only of hitting, springing, striking or touching the enemy, you will not be able actually to cut him.

Likewise, the primary thing in psychiatry is to help the patient, whatever the means. Someone can concern-troll that the hair dryer technique leaves something to be desired in that it might have prevented the patient from seeking a more thorough cure that would prevent her from having to bring the hair dryer with her. But compared to the alternative of "nothing else works" it seems clearly superior.

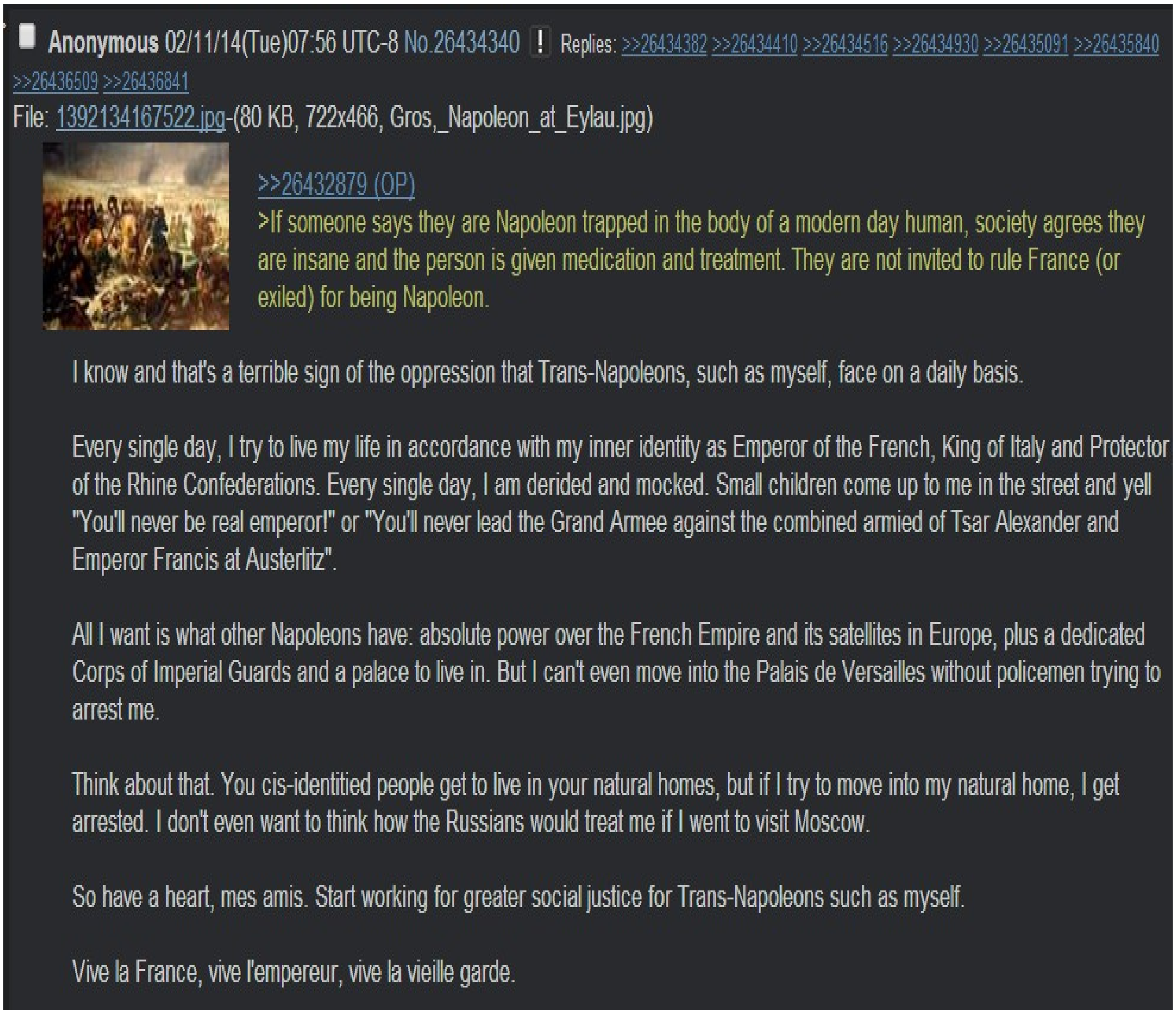
And that's the position from which I think a psychiatrist should approach gender dysphoria, too.

Imagine if we could give depressed people a much higher quality of life merely by giving them cheap natural hormones. I don't think there's a psychiatrist in the world who wouldn't celebrate that as one of the biggest mental health advances in a generation. Imagine if we could ameliorate schizophrenia with one safe simple surgery, just snip snip you're not schizophrenic anymore. Pretty sure that would win *all* of the Nobel prizes. Imagine that we could make a serious dent in bipolar disorder just by calling people different pronouns. I'm pretty sure the entire mental health field would join together in bludgeoning anybody who refused to do that. We would bludgeon them over the head with big books about the side effects of lithium.

Really, are you *sure* you want your opposition to accepting transgender people to be "I think it's a mental disorder"?

VI.

Some people can’t leave well enough alone, and continue to push the mental disorder angle. For example:



Anonymous 02/11/14(Tue)07:56 UTC-8 No.26434340 ! Replies: >>26434382 >>26434410 >>26434516 >>26434930 >>26435091 >>26435840 >>26436509 >>26436841

File: 1392134167522.jpg-(80 KB, 722x466, Gros,_Napoleon_at_Eylau.jpg)

>>26432879 (OP)
>If someone says they are Napoleon trapped in the body of a modern day human, society agrees they are insane and the person is given medication and treatment. They are not invited to rule France (or exiled) for being Napoleon.

I know and that's a terrible sign of the oppression that Trans-Napoleons, such as myself, face on a daily basis.

Every single day, I try to live my life in accordance with my inner identity as Emperor of the French, King of Italy and Protector of the Rhine Confederations. Every single day, I am derided and mocked. Small children come up to me in the street and yell "You'll never be real emperor!" or "You'll never lead the Grand Armee against the combined armied of Tsar Alexander and Emperor Francis at Austerlitz".

All I want is what other Napoleons have: absolute power over the French Empire and its satellites in Europe, plus a dedicated Corps of Imperial Guards and a palace to live in. But I can't even move into the Palais de Versailles without policemen trying to arrest me.

Think about that. You cis-identitied people get to live in your natural homes, but if I try to move into my natural home, I get arrested. I don't even want to think how the Russians would treat me if I went to visit Moscow.

So have a heart, mes amis. Start working for greater social justice for Trans-Napoleons such as myself.

Vive la France, vive l'empereur, vive la vieille garde.

There are a lot of things I could say here.

I could point out that trans-Napoleonism seem to be mysteriously less common than transgender.

I could relate this mysterious difference to the various heavily researched apparent biological correlates of transgender, including unusual variants of the androgen receptor, birth-sex-discordant sizes of various brain regions, birth-sex-discordant responses to various pheromones, high rates of something seemingly like body integrity identity disorder, and of course our old friend altered digit ratios. If our hypothetical trans-Napoleon came out of the womb wearing a French military uniform and clutching a list of 19th century Grand Armee positions in his cute little baby hands, I think I'd take him more seriously.

I could argue that questions about gender are questions about category boundaries, whereas questions about Napoleon – absent some kind of philosophical legwork that I would very much like to read – are questions of fact.

I could point out that if the extent of somebody's trans-Napoleonness was wanting to wear a bicorne hat, and he was going to be suicidal his entire life if he couldn't but pretty happy if I could, let him wear the damn hat.

I could just link people to other sites' pretty good objections to the same argument.

But I think what I actually want to say is that there was once a time somebody tried pretty much exactly this, silly hat and all. Society shrugged and played along, he led a rich and fulfilling life, his grateful Imperial subjects came to love him, and it's one of the most heartwarming episodes in the history of one of my favorite places in the world.

Sometimes when you make a little effort to be nice to people, even people you might think are weird, really good things happen.

[**EDIT:** *I've been told some people are misinterpreting this post as "you can define words any way you want, don't worry about it". While nothing is stopping you from defining a word any way you want, you should definitely worry about it. I had hoped that the Israel/Palestine example above made it clear that you have to deal with the consequences of your* *definitions, which can include confusion, muddling communication, and leaving openings for deceptive rhetorical strategies. In most cases plausible definitions will be limited to a few possibilities suggested by the territory, with worsening penalties as you deviate from these further.*]

Race and Justice: Much More Than You Wanted To Know

*[**EDIT:** This post is getting linked a lot now for obvious reasons. Please remember it's from 2014. There's been a lot more research on this since then, which I have generally not followed closely. I don't have a good impression of how the field has changed in the past six years, so this may or may not be out of date.]*

Previously reviewed: effects of marijuana legalization, health effects of wheat, effectiveness of SSRIs, effectiveness of Alcoholics Anonymous

Does the criminal justice system treat African-Americans fairly?

I always assumed it obviously didn't. Then a while ago I read this harshly polemical but research-filled article claiming to prove it did. Then I found a huge review paper on the subject, written by a Harvard professor of sociology, which concluded after analyzing sixty pages of exquisitely-researched studies that:

> Recognizing that research on criminal justice processing in the United States is complex and fraught with methodological problems, the weight of the evidence reviewed suggests the following. When restricted to index crimes, dozens of individual-level studies have shown that a simple direct influence of race on pretrial release, plea bargaining, conviction, sentence length, and the death penalty among adults is small to nonexistent once legally relevant variables (e.g. prior record) are controlled. For these crimes, racial differentials in sanctioning appear to match the large racial differences in criminal offending. Findings on the processing of adult index crimes therefore generally support the non-discrimination thesis.

Clearly this was more complicated than I thought. I decided to waste my precious free time reading seven zillion contradictory studies to figure out what was going on. Some people on Tumblr have demanded I report back, so here goes:

A. Encounter Rate

There are a lot of tiers to the criminal justice system, each of which will have to be analyzed individual. The first tier is – who does or doesn’t get stopped by the police?

One common point of discussion is traffic stops, leading to the popular joke that you can be stopped for a “DWB” (driving while black). Engel and Calnon (2006) seem to have done the definitive review in this area. Based on a national survey of citizens’ interactions with police, they find that 5% of whites and 11% of blacks have had their cars searched by police, with relatively similar results for other kinds of officer interactions. Therefore, blacks are about twice as likely to be searched as whites. Once you do a multiple regression controlling for other factors, like previous record, income, area stopped, et cetera, half of that difference goes away, leaving an unexplained relative risk of 1.5x.

These data admit to multiple possible interpretations. First, racist police officers could be unfairly targeting blacks. Second, blacks could be acting more suspiciously and police officers correctly picking up on this fact. Third, police officers could be racially profiling based on their past experience of more successful searches of black drivers.

One common method of disentangling these possibilities is search “success rate”. That is, if searching whites usually turns up more real crimes than searching blacks, then innocent blacks are being searched disproportionately often and the police are not just correctly responding to indicators of suspiciousness or past experiences.

Engel and Calnon review sixteen studies investigating this question. If we limit claims of dissimilarity to studies where one race is at least five percentage points higher than the other, there are eight studies with racial parity, six studies with higher white hit rates, and two studies with higher black hit rates.

In other words, in 62% of studies, police are not searching blacks disproportionately to the amount of crimes committed or presumed “indicators of suspiciousness”. In 38% of studies, they are. The differences may reflect either methodological differences (some studies finding effects others missed) or jurisdictionial differences (some studies done in areas where the police were racially biased, others done in areas where they weren’t)

The authors did their own analysis based on a national survey about citizens’ contact with the police, and found that 16% of whites searched and 8% of minorities searched reported that police had discovered contraband, a statistically significant difference. This contradicts the studies above, most of which found no difference and the others of which found much smaller differences.

One possible explanation the authors bring up is that previous research has shown black drivers who have received traffic violations are less likely than whites who have received traffic violations to admit to having received them on anonymous research surveys. For example, among North Carolina drivers known to have received tickets, 75% of whites admitted it on a survey compared to 66% of blacks (Pfaff-Wright, Tomaskovic-Devey, 2000). Comparisons of several different surveys of drug use find that “nonreporting of drug use is twice as common among blacks and Hispanics as among whites" (Mensch and Kandel). Since much of the "contraband" these surveys were asking about was, in fact, drugs, this seems pretty relevant. Overall different studies find different black-white reporting gaps (from the very small one in the traffic ticket study to the very large one on the drug use surveys). Plausibly this is related to severity of offense. Also plausibly, it relates to differential levels of trust in the system and worry about being found out – for poor black people, the possibility of (probably white) researchers being stooges who are going to send their supposedly confidential surveys to the local police station and get them locked up might be much more salient.

There are of course many other forms of police stop. These tend to follow the same pattern as traffic stops – strong data that police more often stop black people, police making the claim that black people do more things that trigger their suspicion instinct (including live in higher-crime neighborhoods), and difficulty figuring out whether this is true or false.

Sampson and Lauritsen review several studies on police stops of pedestrians. I'll be coming back to and citing sources from this Sampson and Lauritsen article many times during this discussion as it is one of the most rigorous and trustworthy analyses around – Sampson is Professor of Sociology at Harvard and winner of the Stockholm Prize in Criminology and his review is the most cited one on this topic I could find, so I assume he represents something like a mainstream position. After reviewing a few studies, most notably Smith (1986), they conclude these sorts of police stops demonstrate no direct effect of race – in any given neighborhood, black people and white people are treated equally – but that there is an indirect effect from neighborhood – that is, the police are nastier to everybody in black neighborhoods. Although they don't say so, the most logical explanation to me would be that black neighborhoods are poorer and therefore higher crime, and so the police are more watchful and/or paranoid.

Summary: There is good data that police stop blacks more often, both on the road and in neighborhoods. Studies conflict over whether the extra stops are justifiable; likely this varies by jurisdiction. Extra neighborhood stops are most likely neighborhood-related effects rather than race-related per se, but the neighborhood effects do disproportionately target black people.

B. Arrest Rates For Violent Crimes

Police records consistently show that black people are arrested at disproportionally high rates (compared to their presence in the population) for violent crimes. For example, blacks are arrested eight times more often for homicide and fourteen times more often for robbery. Even less flashy crimes show the same pattern: forgery, fraud, and embezzlement all hover around a relative risk of four.

(White people are arrested at disproportionally high rates for things like driving drunk, and Asians are arrested at disproportionally high rates for things like illegal gambling, but these carry lower sentences and are less likely to lead to incarceration.)

Once again, there are two possible hypotheses here: either police are biased, or black people actually commit these crimes at higher rates than other groups.

The second hypothesis has been strongly supported by crime victimization surveys, which show that the percent of arrestees who are black matches very closely matches the percent of victims who say their assailant was black. This has been constant throughout across thirty years of crime victmization surveys.

While everybody is totally on board with attributing this to structural factors like black people being poorer and living in worse neighborhoods, anyone who tries to analyze higher black arrest and incarceration rates without taking this into account is going to end up extremely confused.

There were some attempts to cross-check police data and victim data against self-reports of criminality among different races, with various weird and wonderful results. Once again, after a while someone had the bright idea to check whether people who said they hadn’t committed any crimes *actually* hadn’t committed any crimes, and found that a lot of them had well-verified criminal records longer than *War And Peace*.

Sociologists learned an important lesson that day, which is that *criminals sometimes lie about being criminals*.

No one has had any better ideas for how to corroborate the crime victimization survey data, so it looks like probably that’s the best we will do.

Summary: Arrests for violent crimes are probably not racially biased.

C. Arrest Rates For Minor Crimes

Usually when people talk about racial disparities in arrest rates for minor crimes, they’re talking about drugs. The basic argument is that black people and white people use drugs at “similar rates”, but black people are four times more likely to get arrested for drug crime. You can find this argument on pretty much every major media outlet: NYT, Slate, Vox, HuffPo, USA Today, et cetera.

The Bureau of Justice has done their own analysis of this issue and finds it’s more complicated. For example, all of these “equally likely to have used drugs” claims turn out to be that blacks and whites are equally likely to have “used drugs in the past year”, but blacks are far more likely to have used drugs in the past week – that is, more whites are only occasional users. That gives blacks many more opportunities to be caught by the cops. Likewise, whites are more likely to use low-penalty drugs like hallucinogens, and blacks are more likely to use high-penalty drugs like crack cocaine. Further, blacks are more likely to live in the cities, where there is a heavy police shadow, and whites in the suburbs or country, where there is a lower one.

When you do the math and control for all those things, you halve the size of the gap to “twice as likely”.

The Bureau of Justice and another source I found in the Washington Post aren’t too sure about the remaining half, either. For example, anecdotal evidence suggests white people typically do their drug deals in the dealer’s private home, and black people typically do them on street corners. My personal discussions with black and white drug users have turned up pretty much the same thing. One of those localities is much more likely to be watched by police than the other.

Finally, all of this is based on self-reported data about drug use. Remember from a couple paragraphs ago how studies showed that black people were twice as likely to fail to self-report their drug use? And you notice here that black people are twice as likely to be arrested for drug use as their self-reports suggest? That’s certainly an interesting coincidence.

The Bureau of Justice takes this possibility very seriously and adds:

> Although arrested whites and arrested blacks were about equally likely to be drug use deniers, these results nevertheless have implications for the SAMHSA survey. A larger fraction of the black population than the white population consists of criminally active persons and, therefore, a larger fraction of the black population than the white population would consist of criminally active persons who use drugs but deny it. Consequently, the SAMHSA survey would probably understate the difference between whites and blacks in terms of drug use. Whether the effect of such drug use denial among criminally active persons is large enough to account for the unexplained 13% is not known, but research on the topic should pursue this possibility.

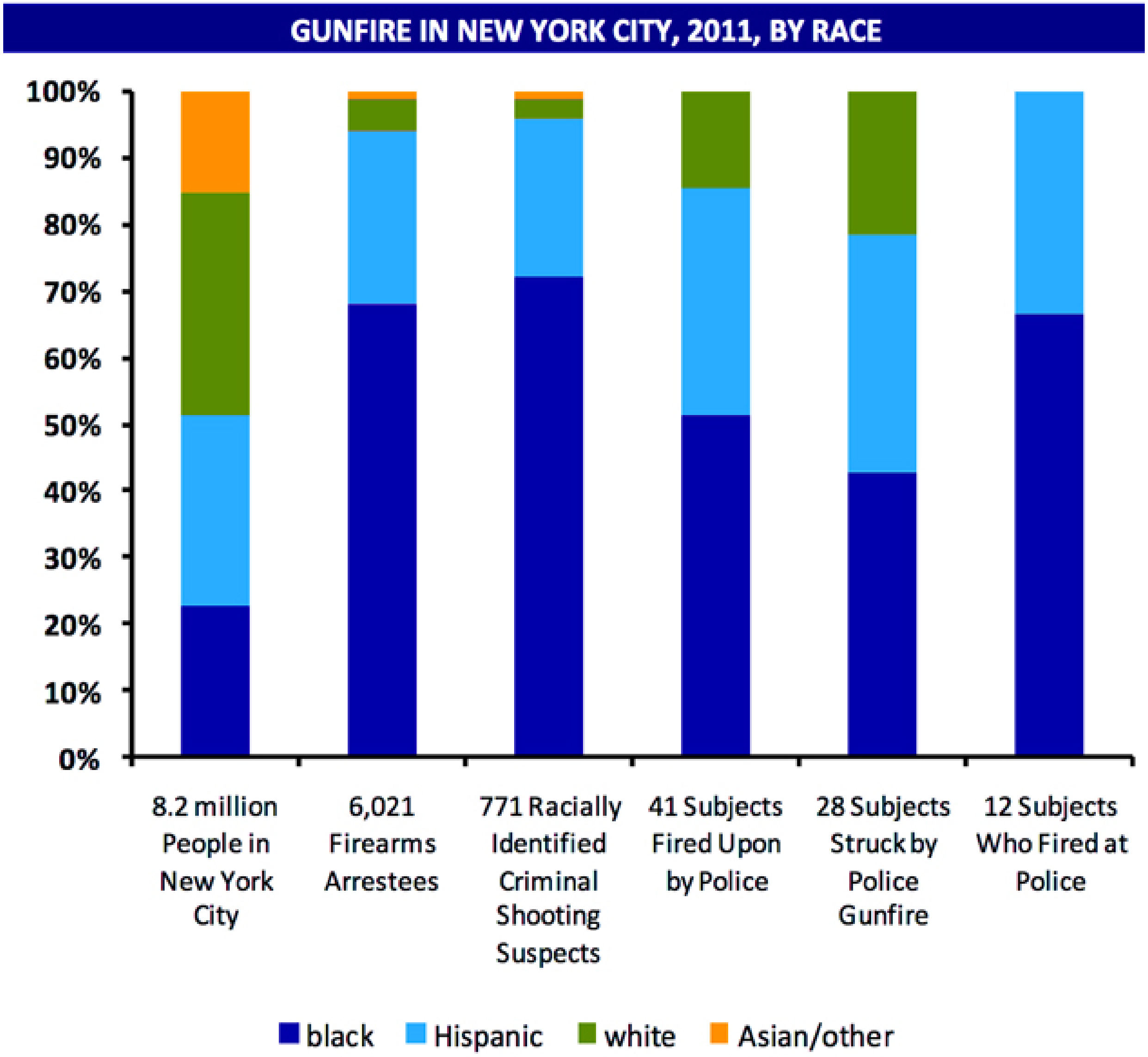
It should be noted that a study investigating this methodology gave random urine drug tests to some of the people who had filled out this survey, and found that half of the actual drug users had reported on the survey that they were squeaky clean. There were no racial data associated with this investigation, which is too bad.

Summary: Blacks appear to be arrested for drug use at a rate four times that of whites. Adjusting for known confounds reduces their rate to twice that of whites. However, other theorized confounders could mean that the real relative risk is anywhere between two and parity. Never trust the media to give you any number more complicated than today’s date..

D. Police Shootings

A topical issue these days. Once again, the same dynamic at play. We know black people are affected disproportionately to their representation in the population, but is a result of police racism or disproportionate criminality?

Mother Jones magazine has an unexpectedly beautiful presentation of the data for us:



The fourth bar seems like what we’re looking for. You could go with the fifth bar, but then you’re just adding noise of who did or didn’t duck out of the way fast enough.

As you can see, a person shot at by a police officer is more than twice as likely to be black as the average member of the general population. But, crucially, they are less likely to be black than the average violent shooter *or* the average person who shoots at the police.

We assume that the reason an officer shoots a suspect is because that officer believes the suspect is about to shoot or attack the officer. So if the officer were perfectly unbiased, then the racial distribution of people shot by officers would look exactly like the distribution of dangerous attackers. If it's blacker than the distribution of dangerous attackers, the police are misidentifying blacks as dangerous attackers.

But In fact, the people shot by police are less black than the people shooting police or the violent shooters police are presumably worried about. This provides very strong evidence that, at least in New York, the police are not disproportionately shooting black people and appear to be making a special effort to avoid it.

For some reason most of the studies I could get here were pretty old, but with that caveat, this is also the conclusion of Milton (1977) looking at police departments in general, and Fyfe (1978), who analyzes older New York City data and comes to the same conclusion. However, the same researcher analyzes police shootings in Memphis and finds that these *do* show clear evidence of anti-minority bias, sometimes up to a 6x greater risk for blacks even after adjusting for likely confounders. The big difference seems to be that NYC officers are trained to fire only to protect their own lives from armed and dangerous suspects, but Memphis officers are (were? the study looks at data from 1970) allowed to shoot property crime suspects attempting to flee. The latter seems a lot more problematic and probably allows more room for officer bias to get through.

[EDIT: A commenter pointed out to me that *Tennessee vs. Garner* banned this practice in the late 1980s, meaning

Memphis' shooting rate should be lower and possibly less biased now]

The same guy looks at the race of officers involved and finds that "the data do not clearly support the contention that white [officers] had little regard for the lives of minorities". In fact, most studies find white officers are disproportionately more likely to shoot white suspects, and black officers disproportionately more likely to shoot black suspects. This makes sense since officers are often assigned to race-congruent neighborhoods, but sure screws up the relevant narrative.

Summary: New York City data suggests no bias of officers towards shooting black suspects compared with their representation among dangerous police encounters, and if anything the reverse effect. Data from Memphis in 1970 suggests a strong bias towards shooting black suspects, probably because they shoot fleeing suspects in addition to potentially dangerous suspects, but this practice has since stopped. Older national data skews more toward the New York City side with little evidence of racial bias, but I don't know of any recent studies which have compared the race of shooting victims to the race of dangerous attackers on a national level. There is no support for the contention that white officers are more likely than officers of other races to shoot black suspects.

E. Prosecution And Conviction Rates

Conviction rates of blacks have generally found to be less than than conviction rates of whites (Burke and Turk 1975, Petersilia 1983, Wilbanks 1987). I don't know why so many of these studies are from the 70s and 80s, but a more recent Bureau of Justice Statistics finds that 66% of accused blacks get prosecuted compared to 69% of accused whites; 75% of prosecuted blacks get convicted compared to 78% of prosecuted whites.

The 1975 study suggested this was confounded by type of crime – for example, maybe blacks are charged more often with serious crimes for which the burden of proof is higher. The 1993 study isn't so sure; it breaks crimes down by category and finds that if anything the pro-black bias becomes *stronger*. For example, 51% of blacks charged with rape are acquitted, compared to only 25% of whites. 24% of blacks charged with drug dealing are acquitted, compared to only 14% of whites. Of fourteen major crime categories, blacks have higher acquittal rates in twelve of them (whites win only in "felony traffic offenses" and "other").

The optimistic interpretation is that there definitely isn't any sign of bias against black people here. The pessimistic interpretation is that this would be consistent with more frivolous cases involving black people coming to the courts (ie police arrest blacks at the drop of a hat, and prosecutors and juries end up with a bunch of stupid cases without any evidence that they throw out).

There was a much talked-about study recently that found that "juries were equally likely to convict black and white offenders when there was at least one black in the jury pool, but more likely to convict blacks when there wasn't." This is consistent with previous studies. Jury pools contain twenty-seven members; the probability that there will be at least one black jury pool member in the trial of a black subject (who of course is most likely to live in a predominantly black area) is high. The study's "equally likely to convict black and white offenders" was actually "2% more likely to convict white offenders than black offenders", which was probably not statistically significant with its small sample size but is consistent with the small pro-black effects found elsewhere.

Summary: Prosecution and conviction rates favor blacks over whites, significance unclear.

F. Sentencing

Older studies of sentencing tend to find no or almost no discrepancies between blacks and whites. This was the conclusion of most of the papers reviewed in Sampson and Lauritsen. The gist here seems to be that there were “four waves” of studies in this area. The first wave, in the 1960s, was naive and poorly controlled and found that there was a lot of racial bias. The second wave, in the 1980s, controlled for more things (especially prior convictions) and found there wasn’t. The third wave was really complicated, and the writers sum it up as saying it represented:

> …a shift away from the non-discrimination thesis to the idea that there is *some* discrimination, *some* of the time, in *some* places. These contingencies undermine the broad reach of the thesis, but the damage is not fatal to the basic argument that race discrimination is not pervasive or systemic in criminal justice processing.

The fourth wave expands on this and finds discrimination in some areas that hadn’t been studied before, such as plea bargaining. However, it continues to find that on the whole, and especially in the largest and best-designed studies there is very little evidence of discrimination. The article concludes:

> Langan’s interpretation matches those of other scholars such as Petersilia (1985) and Wilbanks (1987) in suggesting that systemic discrimination does not exist. Zatz (1987) is more sympathetic to the thesis of discrimination in the form of indirect effects and subtle racism. But the proponents of this line of reasoning face a considerable burden. If the effects of race are so

> contingent, interactive, and indirect in a way that to date has not proved replicable, how can one allege that the "system" is discriminatory?

A more recent (fifth wave?) review adds some problems to this generally rosy picture, saying that "Of the [thirty-two studies containing ninety-five different] estimates of the direct effect of race on sentencing at the state level, 43.2% indicated harsher sentences for blacks…at the federal level 68.2% of the [eight studies containing twenty-two different] estimates of the direct effect of race on sentencing indicated harsher sentences for blacks". The majority of estimates that did not find this were race-neutral, although six did show some bias against whites. They conclude:

> Racial discrimination in sentencing in the United States today is neither invariable nor universal, nor is it as overt as it was even thirty years ago. As will be described below, while the situation has improved in some ways, racially discriminatory sentencing today is far more insidious than in the past, and treating a racial or ethnic group as a unitary body can mask the presence of discrimination.

I really like how you can make a large decrease in the level of a bad thing sound like a problem by saying it is becoming "more insidious".

Even more recent studies have found even larger gaps. A study by the US Sentencing Commission investigating the effect of new guidelines found that blacks' sentences were 20% longer than those of similar whites; a later methodological update reduced the gap to a still-large 14.5% and a a different recent study says just under 10%. Although the particular effect of these new guidelines is a matter of HORRIBLE SUPER-COMPLICATED DEBATE, neither side seems to deny the disparities themselves – only whether they are getting larger.

It’s not clear to me why there’s such a difference between the earlier studies (which found little evidence of disparity), the middle studies (which were about half-and-half), and these later studies (which show strong evidence of disparity). I guess one side of a HORRIBLE SUPER-COMPLICATED DEBATE would say it has to do with changes in sentencing during that time which replace mandatory sentences with “judicial discretion”. If you’re mandated to give a particular sentence for a particular crime, there’s a lot less opportunity to let bias slip in then if you can do whatever you want. There is some evidence that different judges treat different races differently, although the study has no way of proving whether this is anti-black bias, anti-white bias, or an equal mix of both in different people. Unfortunately, there is also concern that mandatory minimum sentencing is itself racist.

Capital punishment is in its own category, and pretty much all studies, old, new, anything agree it is racist as heck (Sampson and Lauritsen cite Bowers & Pierce 1980; Radelet 1981; Paternoster 1984; Keil and Vito 1989; Aguirre and Baker 1990; Baldus Woodward & Pulaski 1990 – there’s no way I’m reading through all of them so I will trust they say what the review says they say). This seems to consist not only in black suspects being more at risk, but in white victims’ deaths being more likely to get their offenders a death sentence.

Summary: Most recent studies suggest a racial sentencing disparity of about 15%, contradicting previous studies that showed lower or no disparity. Changes in sentencing guidelines are one possible explanation; poorly understood methodological differences are a second. Capital punishment still sucks.

Summary

There seems to be a strong racial bias in capital punishment and a moderate racial bias in sentence length and decision to jail.

There is ambiguity over the level of racial bias, depending on whose studies you want to believe and how strictly you define “racial bias”, in police stops, police shootings in certain jurisdictions, and arrests for minor drug offenses.

There seems to be little or no racial bias in arrests for serious violent crime, police shootings in most jurisdictions, prosecutions, or convictions.

Overall I disagree with the City Journal claim that there is no evidence of racial bias in the justice system.

But I also disagree with the people who say things like “Every part of America’s criminal justice is systemically racist by design” or “White people can get away with murder but black people are constantly persecuted for any minor infraction,” or “Every black person has to live in fear of the police all the time in a way no white person can possibly understand”. The actual level of bias is limited and detectable only through statistical aggregation of hundreds or thousands of cases, is only unambiguously present in sentencing, and there only at a level of 10-20%, and that only if you believe the most damning studies.

(except that you should probably stay out of Memphis)

It would be nice to say that this shows the criminal justice system is not disproportionately harming blacks, but unfortunately it doesn’t come anywhere close to showing anything of the sort. There are still many ways it can indirectly harm blacks without being explicitly racist. Anatole

France famously said that “the law, in its majestic equality, forbids rich *as well as* poor people from begging for bread and sleeping under bridges”, and in the same way that the laws France cites, be they enforced ever so fairly, would still disproportionately target poor people, so other laws can, even when fairly enforced, target black people. The classic example of this is crack cocaine – a predominantly black drug – carrying a higher sentence than other whiter drugs. Even if the police are scrupulously fair in giving the same sentence to black and white cokeheads, the law will still have a disproportionate effect.

There are also entire classes of laws that are much easier on rich people than poor people – for example, any you can get out of by having a good lawyer – and entire classes of police work that are harsher on poor neighborhoods than rich neighborhoods. If the average black is poorer than the average white, then these laws would have disproportionate racial effects.

For more information on this, I would recommend Tonry and Melewski’s *Malign Neglect: Race, Crime, and Punishment in America*. They begin by saying everything above is true – the system mostly avoids direct racist bias against black people – and go on to say argue quite consistently that we *still* have a system where (their words) “recent punishment policies have replaced the urban ghetto, Jim Crow laws, and slavery as a mechanism for maintaining white dominance over blacks in the United States”. If you want something that makes the strongest case for the justice system harming blacks, written by real criminologists who know what they’re talking about, there’s your best bet.

(warning: I haven’t read the book. I did read a review article by the same people, which the book is partially based on)

Some police officers say the reason they are harsher in poor urban neighborhoods is that the expectation of high levels of unruly behavior necessitates unusually strong countermeasures. For the same reason, I am screening all comments for the next few days. If you post one, expect it to show up eventually or perhaps disappear into the aether.

Why I Am Not Rene Descartes

I.

Imagine that somebody wrote:

> Some of my friends support Ron Paul. I think that’s wrong. After all, he’s a libertarian, and Wikipedia says a libertarian is a person who believes in free will. But free will is impossible in a deterministic universe. Ron Paul’s belief in free will is clearly why there are so few Swiss people among Ron Paul supporters, since Swiss people are Calvinists and so understand determinism better.

This is sort of how I feel reading Why I Am Not A Rationalist on Almost Diamonds.

I’m having trouble not quoting it in full:

> I’m not a rationalist because I’m an empiricist. I find no value in “logical” arguments that are based in intuition and “common sense” rather than data. Such arguments can only perpetuate ignorance by giving it a shiny veneer of reason that it hasn’t earned.
>
> I boggle that we haven’t sorted this out yet. I particularly boggle that atheists of my acquaintance promote rationalism over empiricism. The tensions between basic rationalism and empiricism parallel the tensions between church theology and the philosophy of science. We have no problem rejecting church theology as not being grounded in evidence. Why do so many atheists praise rationalism?
>
> Let me stop here and make it clear that I’m not rejecting logic or critical thinking. Goodness knows that I’ve spent

hours just this summer helping people share useful heuristics that will, in general, help them get to the right answers more often. I’ve led workshops and panels on evaluating science journalism and scientific results. When I’ve spoken to comparative religion classes in the past, I’ve talked about religious skepticism with an emphasis on the basics of epistemology.

The problem isn’t logic or critical thinking. The problem is a tendency to view those skills as central to getting the right answers. The problem is a tendency to view them as the solution. They’re not, and the idea that they are is in distinct contrast with the way humanity has actually grown in knowledge and understanding of the world.

Rationalism is, at heart, an individualist endeavor. It says that the path to getting things right lies in improving the self, improving the thinking of one person at a time. It’s not surprising that the ideology and movement appeal largely to the young, to men, to white people, to libertarians. It focuses primarily on individual action.

That’s not how we’ve come to learn about our world, though. It’s not how science or any other field of scholarship works. Scholarship is a collaborative process. And I don’t just mean peer review and working groups, though those are important as well.

Scholars add to our knowledge of the world by building on the work of others. They apply tools and methods developed by others to new material and questions. They study the work of other scholars to inspire them and give them the background to ask and answer new questions. They evaluate the work of others and

> consolidate the best of it into larger theoretical frameworks. Without the work of scholars before them, scholars today and evermore would always be recreating basic work and basic errors.
>
> All too often, I find rationalists taking this repetitive approach. They think but they don't study. As a consequence, they repeat the same naïve errors time and again. This is particularly noticeable when they engage in social or political theorizing by extrapolating from information they learned in secondary school and 101-level college classes, picked up in pop culture, or provided by people pushing a political cause. Their conclusions are necessarily as limited as their source material and reflect all its cultural biases.

As best I can tell, it is conflating a misunderstood version of rationalism (Descartes) with a misunderstood version of rationalism (Yudkowsky) and ending up with something unrelated to either, in the most bizarre possible way. But since there are commenters there who seem to *agree* with it, better nip this in the bud before it spreads.

II.

Rationalism (Descartes) is not simply the belief that sitting and thinking is *more useful than* observation. Descartes-style rationalism is complicated, but involves the claim that certain concepts are known prior to experience. For example, it is possible to understand mathematical truths like 1 + 1 = 2 separately from our experience of observing people add one apple to another. It is also possible to know them more completely than our knowledge of the external world, since our external senses can only tell us that all additions of one plus one have equalled two *so far*, but our reason can tell us something we have never observed – that it is necessarily true that everywhere and for all time 1 + 1 will equal two.

This so obviously gets bogged down in definitions of what is or isn’t “prior to experience” or a “concept” that philosophers today have mostly moved on to bigger and better things like hitting people with trolleys. It has nevertheless gotten a mild boost of interest recently with Chomsky’s claim that some features of human language are innate, and evolutionary psychology’s claim that certain preferences like fear of spiders may be innate. You can learn much more than you wanted to know at the Stanford Encyclopedia of Philosophy

But in no case does the debate ever resemble Almost Diamonds’ naive conception of some people thinking they have to study the world and other people sitting and speculating in armchairs and playing at self-improvement because they don’t want to get their hands dirty in the real world. In fact, Descartes himself was a devoted experimentalist – probably too devoted. His intense interest in anatomy combined with his belief that animals lack souls made him one of the most prolific vivisectionists of all time. We can thank him for such useful pieces of scientific information as “If you cut off the end of the heart of a living dog and insert your finger through the incision into one of the concavities, you will clearly feel that every time the heart shortens, it presses your finger, and stops pressing every time it lengthens.” One can accuse the author of this statement of a *lot* of things, but “not willing to get his hands dirty” isn’t one of them.

Likewise, Leibniz, the *other* most famous partisan of rationalism of all time, also made notable contributions to physics, geology, embryology, paleontology, and medicine. Either he was out exploring the world, or he had *some* armchair. Particularly ironically for Almost Diamonds' thesis, he was one of the most prominent advocates of research as a collaborative endeavor, and founded various scientific discussion societies around Europe as well as calling for a giant international database of all scientific findings. All of this was entirely consistent with, and informed by, his rationalism.

The people on r/philosophy also do a good job of explaining this mistake in that blog's conception of rationalism (the people on r/badphilosophy do an, um, less good job)

III.

But I think Almost Diamonds is mostly talking (also talking?) about rationalism (Yudkowsky), ie internet rationalism. After all, she mentions "the rationalist movement" and says they're about "understanding cognitive biases" and "appeal largely to…libertarians".

This is even more wrong.

At least rationalism (Descartes) is *sort of* about some kind of disconnect with empirical evidence. In the context of rationalism (Yudkowsky) this is about the same level of error as expecting Ron Paul to be a philosopher preaching free will just because "libertarianism" can mean something in metaphysics. Rationalism (Yudkowsky) and rationalism (Descartes) share a name, nothing more.

Almost Diamonds says:

> I'm not a rationalist because I'm an empiricist. I find no value in "logical" arguments that are based in intuition and "common sense" rather than data…I boggle that we haven't sorted this out yet. I particularly boggle that

> atheists of my acquaintance promote rationalism over empiricism.

Meanwhile, the founding document of rationalism (Yudkowsky), the Twelve Virtues of Rationality, states:

> **The sixth virtue is empiricism.** The roots of knowledge are in observation and its fruit is prediction. What tree grows without roots?…Do not ask which beliefs to profess, but which experiences to anticipate.

It adds:

> You cannot make a true map of a city by sitting in your bedroom with your eyes shut and drawing lines upon paper according to impulse. You must walk through the city and draw lines on paper that correspond to what you see. If, seeing the city unclearly, you think that you can shift a line just a little to the right, just a little to the left, according to your caprice, this is just the same mistake.

This makes the same point as Almost Diamonds.

Now, granted, some movements can have Official Founding Beliefs that they don't follow. Many rationalists take the Virtues very seriously (one just let me know it is hanging on the wall of his group house) but perhaps like some of Jesus' more lovey-dovey commandments or the inconvenient parts of the Constitution, they are honored more in the breach than in the observance?

I don't think so. Rationalists have taken this idea and run with it, which is why we are so obsessed with things like making beliefs pay rent in experience, discussion of "Bayesian updating", and even making monetary bets on our beliefs to train ourselves to make sure they conform to real world outcomes. It’s why the rationalist proverb, upon being given a cool theory, goes “Name three examples”.

A typical (okay, I lied, highly extreme) example is Gwern, who consumed pretty much every chemical and then carefully recorded its effects on his sleep, emotions, performance on cognitive tests, et cetera and then performed Bayesian analysis on it. There’s obviously *something* wrong with that, but it’s not lack of empiricism!

Diamonds:

> All too often, I find rationalists taking this repetitive approach. They think but they don’t study.

Twelve Virtues again:

> **The eleventh virtue is scholarship.** Study many sciences and absorb their power as your own. Each field that you consume makes you larger. If you swallow enough sciences the gaps between them will diminish and your knowledge will become a unified whole. If you are gluttonous you will become vaster than mountains.

And in accordance with this, I will put the rationalist movement up, mano a mano, against any other movement on the entire Internet in terms of the quality of scholarship and empiricism.

Like, holy @#$%, we have Luke Muehlhauser, who can’t write a simple life hacks post on productivity without fifty-seven different journal article citations, and who writes at great length about how to study a field of research effectively, the best textbooks on every subject, software tools for efficient scholarship, etc, etc, etc.

We have a community-wide survey that collects information on one hundred thirty-six demographic categories, for over fifteen hundred community members, and then a tradition of obsessively arguing about the implications of the results for several weeks every year.

We know that 20% of rationalists over the age of 35 have Ph. Ds. 54% have either a Ph. D, an MD, or a Master's!

And not to toot my own horn, but there's a reason this blog's series of impromptu literature reviews is called "Much More Than You Wanted To Know" and has investigated the literature on things like marijuana legalization and SSRI effectiveness, fifty or sixty studies per review, to a degree that's gotten some coverage on major news sites including Andrew Sullivan's blog and Vox.

And…wait a second! The author of that blog knows Kate Donovan! How do you know Kate Donovan and still accuse rationalists of "not studying"?!?! DO YOU EVEN HAVE EYES?

Finally, Almost Diamonds says:

> Even in the modern rationalist movement, which speaks more to collecting evidence than classical rationalism, I have yet to see any emphasis on epistemic humility.

But the Twelve Virtues says:

> **The eighth virtue is humility.** To be humble is to take specific actions in anticipation of your own errors. To confess your fallibility and then do nothing about it is not humble; it is boasting of your modesty. Who are most humble? Those who most skillfully prepare for the deepest and most catastrophic errors in their own beliefs and plans. Because this world contains many whose grasp of rationality is abysmal, beginning

> students of rationality win arguments and acquire an exaggerated view of their own abilities. But it is useless to be superior: Life is not graded on a curve. The best physicist in ancient Greece could not calculate the path of a falling apple. There is no guarantee that adequacy is possible given your hardest effort; therefore spare no thought for whether others are doing worse. If you compare yourself to others you will not see the biases that all humans share. To be human is to make ten thousand errors. No one in this world achieves perfection.

Really? “Yet to see any emphasis on epistemic humility”? The most important mission statement of the rationalist movement says that one of the movement’s twelve founding principles is humility, then waxes rhapsodic about it. Seriously, we’re the people who keep calling ourselves “aspiring rationalists” to remind ourselves that we’re not nearly as rational as we should be yet! We’re the people who obsessively calibrate with Prediction Book et cetera to remind ourselves just how high our error rate is. We’re the people who keep a community-wide keep a mistakes repository (with Gwern once again going above and beyond). THERE ARE FORTY ONE DIFFERENT POSTS ON LESS WRONG TAGGED ‘OVERCONFIDENCE’!

Almost Diamonds dislikes rationalism because she believes in an emphasis of empiricism over armchair speculation, careful scholarship over ignorance, and epistemic humility. But she’s just described the rationalist movement almost to a ‘T’! She’s attacking the rationalist movement for not living up to her ideal philosophy which is…the precise philosophy of rationalist movement!

This is mean, but I’m going to say it. Almost Diamonds describes the rationalist movement in a way that even the most cursory glance at any rationalist site or document would disprove. Her opinion seems to be based entirely on a distorted idea of the dictionary definition of the word “rationalism”.

It’s almost like she’s, I don’t know, sitting in an armchair speculating about what rationalism must be, rather than going out and looking for evidence.

□_□

IV.

Also, can I just mention that one of the commenters on that blog says that the problems with the rationalist movement are a lot like the problems with frequentist statistics, and what would really help them is if they investigated Bayesianism? I swear I am not joking. I swear this is a thing that happened.

V.

But aside from all this, I do think there’s an important point that needs to be made here. That is – given that empiricism and scholarship is obviously super-important, why is it not enough?

The very short answer is “A meta-analysis of hundreds of studies is what tells you that psychic powers exist. Critical thinking is what helps you figure out whether to trust that result.”

The longer answer: rationality is about drawing correct inferences from limited, confusing, contradictory, or maliciously doctored facts. Even the world’s most stubborn creationist would have to realize the truth of evolution if you could put her in a time machine and make her watch all 3 billion years of life on Earth. But more rational people can realize the truth of evolution after reading a couple of good biology textbooks and having some questions answered. And Darwin could realize the truth of evolution just by observing the natural world and speculating about finches. There's something I do better than the creationist and Darwin does better than me, and it's not "have access to data".

Life is made up of limited, confusing, contradictory, and maliciously doctored facts. Anyone who says otherwise is either sticking to such incredibly easy solved problems that they never encounter anything outside their comfort level, or so closed-minded that they shut out any evidence that challenges their beliefs.

Given this state of affairs, obviously it's useful to have as much evidence as possible, in the same way it's useful to have as much money as possible. But equally obviously it's useful to be able to use a limited amount of evidence wisely, in the same way it's useful to be able to use a limited amount of money wisely.

I recently reviewed thirty-five studies on racism in the criminal justice system, a very controversial topic. But I would suggest that almost nobody would change their opinion about this based on simple number of studies reviewed. That is, if a person who has read five studies and believes the system is racist encountered another person who has read ten studies and believes the system is fair, she would not simply say "Well, you've read more studies than I have, so I guess you're right and I'm wrong." She would probably say "That's interesting, but I need to double-check the methodologies of those studies, make sure they mean what you think they mean, make sure you haven't specifically selected only studies that prove your view, and make sure you haven’t fallen into one of a million other possible failure modes.”

The part where you have read 5 studies but I have read 10 is the empiricism that Almost Diamonds would say is the only meaningful skill that exists. The part where we want to make sure they’re good studies and I understood them right is rationality. I would trust the opinion of a rational person who knows one study far more than that of an irrational person who knows fifty. If you don’t believe me, I invite you to check out the hundreds of studies published in creation science and homeopathy journals every year.

Or what if you’re working in an area where you don’t even have hypotheses yet? It’s your job to explain or predict something that’s never been explained or predicted before. Sure, you’ve got to have a background level of expertise and scholarship, but no matter how many x-ray crystallographers you have somebody has to be the one to say “You know, our data would make sense if this molecule were in the shape of a helix.” What if you’re trying to predict the future – like in what year fusion power will become a reality, or whether a stock is going to go up and down – and you’ve already reviewed all of the relevant evidence? What then?

If somebody says that rationality is all nice and well, but not really important because you can just use the facts, then this is the surest sign of somebody who doesn’t possess the skill and doesn’t even realize there is a skill there to be possessed. They have inoculated themselves with the cowpox of doubt, trained themselves on easy problems so long that they’ve dulled their senses and forgotten that problems that require more thought than just looking up the universally-agreed-upon scientific consensus in Wikipedia even exist.

There is one paragraph for which I will give Almost Diamonds credit: she is partly right when she says rationality is fundamentally an individual endeavor. I mean, only in the sense martial arts is an individual endeavor – you can train with lots of people, you *have* to train with lots of people, you've got to learn the craft from others and stand on the shoulders of giants – but in the end you've got to punch the other guy yourself.

Thousands of scientists have worked their entire lives to get you the evidence in favor of evolution. But thousands of creationists have worked *their* entire lives to obfuscate and confuse that evidence. Thousands of scientists have studied the criminal justice system, but many of them aren't very good at it, many of them disagree with one another, and very likely none of them have worked on the exact subsubproblem that you're interested in. Other people can present the facts to you, but in the end you're the one deciding what and who to believe. Just like everybody dies alone, everybody decides on their beliefs alone. And rationality is what allows them to do that accurately.

If you're investigating a problem even slightly more interesting than evolution versus creationism, you will always encounter limited, confusing, contradictory, and maliciously doctored facts. The more rationality you have, the greater your ability to draw accurate conclusions from this mess. And the differences aren't subtle

A superintelligence can take a grain of sand and envision the entire universe.

Einstein can take a few basic facts about light and gravity and figure out the theory of relativity.

I can take a bunch of conflicting studies and feel sort of confident I’ve at least figured out the gist of the topic.

Some people can’t take a movement that emphasizes on its founding document “OUR VIRTUES ARE EMPIRICISM, SCHOLARSHIP, AND HUMILITY” and figure out that it considers empiricism, scholarship, and humility to be virtues.

Such Mixed Feelings About Crazymeds

Crazymeds.us is an excellent and highly informative site which I will never recommend to my patients.

It’s excellent because it gives mostly accurate and readable descriptions of the costs and benefits of every psychiatric medication. It has a laser-like focus on what patients will actually want to know and was clearly written by someone with an encyclopaedic knowledge of every treatment’s strengths and potential pitfalls.

This is important because the standard psychiatric response to someone who wants to know about a medication (when it’s not “shut up and trust me”) is to print out an information sheet from somewhere like drugs.com or webmd.com. These sites at worst just copy paste the FDA drug information sheet, and at best list off side effects in a rote and irrelevant way that only a robot could love.

Here’s an excerpt from drugs.com about the side effects of Prozac:

> Get emergency medical help if you have any of these signs of an allergic reaction to fluoxetine: skin rash or hives; difficulty breathing; swelling of your face, lips, tongue, or throat.
>
> Report any new or worsening symptoms to your doctor, such as: mood or behavior changes, anxiety, panic attacks, trouble sleeping, or if you feel impulsive, irritable, agitated, hostile, aggressive, restless, hyperactive (mentally or physically), more depressed, or have thoughts about suicide or hurting yourself.

Call your doctor at once if you have:

– blurred vision, tunnel vision, eye pain or swelling, or seeing halos around lights;

– high levels of serotonin in the body–agitation, hallucinations, fever, fast heart rate, overactive reflexes, nausea, vomiting, diarrhea, loss of coordination, fainting;

– low levels of sodium in the body–headache, confusion, slurred speech, severe weakness, vomiting, loss of coordination, feeling unsteady;

– severe nervous system reaction–very stiff (rigid) muscles, high fever, sweating, confusion, fast or uneven heartbeats, tremors, feeling like you might pass out; or

– severe skin reaction–fever, sore throat, swelling in your face or tongue, burning in your eyes, skin pain, followed by a red or purple skin rash that spreads (especially in the face or upper body) and causes blistering and peeling.

Common fluoxetine side effects may include:

– sleep problems (insomnia), strange dreams;
– headache, dizziness, vision changes;
– tremors or shaking, feeling anxious or nervous;
– pain, weakness, yawning, tired feeling;
– upset stomach, loss of appetite, nausea, vomiting, diarrhea;
– dry mouth, sweating, hot flashes;
– changes in weight or appetite;
– stuffy nose, sinus pain, sore throat, flu symptoms; or
– decreased sex drive, impotence, or difficulty having an orgasm.

> This is not a complete list of side effects and others may occur. Call your doctor for medical advice about side effects. You may report side effects to FDA at 1-800-FDA-1088.

If I ask a patient to read this, one of two things happen. First, they read the first few sentences and are like “Sure, whatever, I’ll read it when I get home” and then throw the paper in the trash can on the way out of the room. Or second, they get to the part where it says “agitation, hallucinations, fever, fast heart rate, overactive reflexes, nausea, vomiting, diarrhea, loss of coordination, fainting” and they’re like WHAT ARE YOU TRYING TO GIVE ME I’M NOT TAKING THIS GARBAGE!

Here’s crazymeds.us on the same topic:

> **4.1 Typical Side Effects of Prozac (fluoxetine HCl)**
>
> The usual for SSRIs – headache, nausea, dry mouth, sweating, sleepiness or insomnia, and diarrhea or constipation, weight gain, loss of libido. Most everything will go away after a week or two, but the weight gain and loss of libido might stick around longer. Or permanently. Although weight gain is a coin toss.
>
> **4.2 Not So Common Side Effects of Prozac (fluoxetine HCl)**
>
> Rash, ‘flu-like symptoms, anger/rage.
>
> **4.3 Prozac (fluoxetine HCl) Freaky Rare Side Effects**
>
> Bleeding gums, amnesia, anti-social reaction (oh, come on, like we’re not anti-social already), herpes (again, blaming the med for an STD), excessive hair growth,

> engorged breasts (a.k.a. porno boobs), involuntary tongue protrusion (according to the PI sheet / PDR one 77-year-old woman stopped sticking her tongue out at everyone after they stopped giving her Prozac (fluoxetine hydrochloride)).

This is readable, complete without being overwhelming, gives you a good idea how likely everything is, and, dare I say it, funny. It tells you which things to expect, how long to expect them to last, and even mentions that they're "usual for SSRIs", which is pretty important for someone using the side effects to decide whether they want to take Prozac vs. another drug.

I also am eternally grateful for them mentioning how some people blame the medication for their STDs – the FDA has this thing where if someone reports something they take it seriously, and it ends up with drug leaflets including anything that's ever happened to someone while on a drug as a "possible side effect" ("My pet turtle died when I was on Prozac, I demand you warn customers that one side effect of Prozac is 'increased mortality for associated chelonians'!") Crazymeds.us calls them out on this.

Everything crazymeds.us does is like this. Well-written, funny, mostly accurate (with the occasional mistake but no more than you'd expect from an individual effort), and precisely targeted to what patients really need to know.

And I still don't recommend it to my patients, and probably never will. Why not?

Well, for one thing, *it's called crazymeds.us*.

Most psychiatric patients have no problem with the word "crazy". Either they don't think of themselves as crazy, or they jokingly call themselves crazy and are happy to let other people in on the joke, or they self-identify as crazy as matter-of-factly as they'll tell you the time of day, or they just don't care.

But some psychiatric patients care about it *a lot*. Either they're moderately neurotic people who are scared that if they accept psychiatric help with their mild depression it puts them in a category of "total lunatic" from which they will never escape, or they're social justice types who are watching like hawks for any sign that their psychiatrist is a privileged ableist oppressor trying to use slurs to trivialize their concerns and victim-blame them for their problems.

I can usually tell which category a given person is in pretty quickly, but the chance of accidentally slipping up and recommending to someone from the second category a site called crazymeds.us is too horrible to contemplate.

And it's not just the name. Somebody is going to say that the reference to "porno boobs" is objectifying women or trivializing the problems of people with gynaecomastia. Someone will point out their misuse of the term antisocial and their seemingly flippant dismissal of social phobia. Someone will *definitely* have something to say about the issues raised by selling crazymeds-brand mugs saying "Medicated For Your Protection". This isn't just a couple of slipups here or there. It's the entire ethos of the site.

Alyssa Vance introduced me to the idea of "negative selection". It's when you don't care how good something is, you just want it to definitely not be bad. For example, when you're hiring fast food workers, you're not looking for someone with a Harvard degree in fast foodology who will revolutionize what it means to work at McDonalds for generations to come, you just want someone who you're really sure isn't going to show up late or commit any crimes.

Likewise, medicine involves some heavy negative selection. If I become the best, most likeable, most respectful, most intelligent psychiatric resident in the country – well, I'd still get paid exactly as much as I do now. On the other hand, if even *one* patient lodges a complaint against me – let alone a lawsuit – that's probably a hospital investigation and a stern talking-to from my boss and his boss and a discussion of why a site called crazymeds.us talking about how psych patients are "Medicated For Your Protection" is Not Appropriate and really you're a second-year resident shouldn't you know things like that already maybe you need some Remedial Communication Skills Classes.

Side effects of crazymeds.us may include agitation, aggression, and job loss. I think I can predict pretty well which patients will and won't appreciate this kind of message, but all I need is one misclassification to get screwed over. So, with apologies to the many patients who could be helped by something like crazymeds.us, I'm giving this particular minefield a wide berth.

Part of me wants to grab whoever made the site and scream at them "WHY? WHY ARE YOU DOING THIS? YOU COULD HAVE BEEN BY FAR THE BEST PSYCHIATRIC RESOURCE ON THE ENTIRE INTERNET, IMPROVED THE LIVES OF TENS OF THOUSANDS OF PEOPLE, AND YOU THREW IT ALL AWAY FOR SOME STUPID JOKES."

But there's also a part of me that accepts they probably have their reasons. I'm not sure it's possible to make a site as good as crazymeds.us without it being as offensive as crazymeds.us. Remove every single flippant statement and optimize for complete unobjectionability, and you're most of the way back to drugs.com. I mean, there are certainly some simple improvements that could be made on drugs.com, and there's probably a market for a site like that, and maybe that site already exists and I just haven't found it. But crazymeds is something special. It's inspiring trust through countersignaling. In a field where almost everyone is a dry, scientific person who won't give you a straight answer about anything or treat you like a human being, crazymeds' business strategy is to make it super obvious they're the exact opposite of that. They're *human*, and I think that's precisely why a demographic who wouldn't trust anybody else trusts crazymeds.

Athrelon writes about social technology, structures and institutions that allow the maintenance of trust and order with a minimum of fuss. Crazymeds is an item of social technology that draws people who would normally be unwilling to learn about the psychiatric system into an engagement with it and an ability to understand and manage their own care.

Athrelon also writes about the breakdown of social technology in the face of certain modern social norms. And I talk a lot about "political correctness" and so on. And one retort I sometimes get is "political correctness just means being nice and not going around offending people. How could you possibly be against that?" And I have had vague feelings that it probably does something bad and have just mumbled something about how I'd look for an example and let you know when I found one.

Well, this is an example. I'm pretty sure there are a lot of people who would benefit from crazymeds, or at least from straighter talk than the carefully hedged neutral statements on drugs.com, or at least from a little bit of humor or conversation in a register associated with normal healthy human relationships – but as long as I'm not 100% certain no one will be offended, and as long as that one offended person can cause me more grief than the hundreds of satisfied customers can possibly make up for, I’m going to keep printing drugs.com handouts, and explaining while mentally facepalming that no, the hallucinations and agitations don’t happen to *everyone*.

On the other hand, you guys have taken much worse from me and I’m still here, so I recommend it to *you* without hesitation. Obvious caveats (eg don’t do anything without talking to a doctor first) are obvious.

Cyber Monday Product Recommendations

Apparently Cyber Monday is a thing now, so here are my product recommendations from this year.

This is going to be *really* nerdy, but by far the best thing I got this year was these collar extenders. Every time I try to wear a dress shirt it chokes me. The only time the collar is big enough is if I get a shirt so many sizes too big for me that the rest of it starts looking like a toga. Either I am a mutant with a uniquely high neck-to-body ratio, or I have abnormal anxiety about / sensitivity to feeling things against my neck. In either case, these little button-type things go into the shirt and make the collar bigger and solve my problem. This has made work about 50% less unpleasant, at the low cost of my dignity.

Actually, my life became a lot better in general once I noticed that a lot of my dislike of work was somatic. After preventing myself from being choked by shirts, another big improvement was getting these slip-on-like shoes plus these gel insoles for days when I've got to walk all around the hospital or stand by a bedside for half an hour while my team is talking to a patient. The sort of shoes that my parents would get angry about me wearing to a restaurant are apparently totally acceptable as dress shoes for work as long as they are black and shiny.

Dollar Shave Club advertises that they have good razors for cheap, and offer (for example) four four-bladed razor cartidges every month for $6. That's an obvious advantage over normal companies like Gillette, with whom the same offer would work out to something like $15. I was all set to sign up for this, but I did my customary pre-purchase Google search and found that Dollar Shave Club gets all of their products from a company called Dorco, then sells them at a markup. If you buy from Dorco directly in bulk, you can get the same deal for only about $4.50. As best I can tell, the only difference is that instead of buying from a heavily-marketed, carefully branded organization like Dollar Shave Club, you're buying from a company whose branding is so clueless that they named themselves "Dorco". I got this starter pack off Amazon, and so far it's been at least as good as my previous razor (a Gillette). Now I can change to new blades twice as often and still save money.

I continue to wrestle with oversensitivity to noise. Although there are all sorts of rumors about magic custom-fit earplugs which will completely block 100% of all sound, I've never been able to find them. Right now my earplugs of choice are these goofy blue ones, which have a Noise Reduction Rating of 33, about as high as you'll find sold commercially, and are also cheap enough that you can throw them away after a night or two and comfortable enough not to bother me. When I combine them with these earmuffs I can usually get relief from most music and incidental noise, although they're not perfect and they hurt my neck if I use them too long. This is the best combination I've found in years of searching, but I'm interested in other people's discoveries.

Another thing that has proven well worth its cost in dignity: this bug vacuum, classily named a "bugzooka". When there is a big scary bug in the house, you point the bugzooka at it, suck it up into the machine's chamber, and then release the bug outside. Or, more realistically, you make Ozy release the bug outside, because the release mechanism is much less automated than the sucking mechanism and usually it involves dismantling the machine by hand, at which point a really angry bug has an attack of opportunity to sting/bite/spray toxins at you. That maybe wasn't the best design choice. It's still better than most other methods of bug collection.

My new laptop is sort of the Optimus Prime of computers – it can transform between a regular laptop, a weirdly-shaped hard-to-use laptop, and a large awkward tablet. Its weirdly-shaped hard-to-use configuration is strangely compatible with my habit of lying prone hanging halfway off a bed while I surf the Internet on a laptop on the floor below me. So far I've only had two problems – a total crash that the repair shop assured me was because of software, and some finger/hand pain when I type too much on the keyboard (which I solved with a peripheral USB keyboard; I type more than any reasonable person, so this may not happen to you). Other people have reported WiFi problems, but I've been fine. Due to some of these issues, I'm not sure I can unreservedly recommend this, but it does have a "wow" factor and I've been pretty satisfied. I can much more confidently recommend my tiny portable speakers, which everybody who hears seems to like and which have weathered an impressive amount of abuse without complaint.

I continue to recommend Codex Seraphinianus. For best results, use as a coffee table book, the way you would normally use some nice colorful book like *Birds of North America* or something. When somebody sits down at your coffee table, starts leafing through it, and asks what it is, freak out and tell them they weren't supposed to see that. Then grab it, hastily place it in a dresser, and give them a copy of *Birds of North America* to leaf through.

Since I'm advertising MealSquares on the sidebar, I should probably mention them. They're a Soylent-like food which claims to be nutritionally complete – while also being solid and made of “whole foods”. I am more likely to take their claims seriously than Soylent since they are working with a registered dietician whom I know and respect; that said, I haven’t looked into their nutritional claims too much, and the site specifically says you’ll be better off combining them with normal food than trying to live solely off MealSquares your entire life – which seems about right. Ozy and I use them when we can’t be bothered to make anything, are really hungry, and want to feel healthier than just eating a cookie or a cracker or something.

Book Review: On The Road

I.

On the Road is a terrible book about terrible people. Jack Kerouac and his terrible friends drive across the US about seven zillion times for no particular reason, getting in car accidents and stealing stuff and screwing women whom they promise to marry and then don’t.

But this is supposed to be okay, because they are *visionaries*. Their vision is to use the words “holy”, “ecstatic”, and “angelic” at least three times to describe every object between Toledo and Bakersfield. They don’t pass a *barn*, they pass a holy vision of a barn, a barn such as there must have been when the world was young, a barn whose angelic red and beatific white send them into mad ecstasies. They don’t almost hit a *cow*, they almost hit a holy primordial cow, the cow of all the earth, the cow whose dreamlike ecstatic mooing brings them to the brink of a rebirth such as no one has ever known.

Jack Kerouac and his terrible friends are brought to the brinks of a lot of things, actually. Aside from stealing things and screwing women whom they promise to marry and then don’t, being brought to the brink of things is one of their main pastimes. Enlightenment, revelation, truth, the real meaning of America, the ultimate, the sacred – if it has a brink, they will come to it. Crucially, they never cross that brink or gain any lasting knowledge or satisfaction from the experience. Theirs is a religion whose object of worship is the burst of intense emotion, the sudden drenching of their brain in happy chemicals that come and go without any lasting effect except pages full of the words “holy”, “ecstatic”, and “angelic”.

The high priest of this religion is Kerouac’s friend Dean Moriarty. Kerouac cannot frickin shut up about Dean Moriarty. Obviously he is “holy” and “ecstatic” and “angelic” and “mad” and “visionary”, but for Dean, Kerouac pulls out all the stops. He is “a new kind of American saint”, “a burning shuddering frightful Angel”, with intelligence “formal and shining and complete”.

Who is this superman, this hero?

> His specialty was stealing cars, gunning for girls coming out of high school in the afternoon, driving them out to the mountains, making them, and coming back to sleep in any available hotel bathtub in town.

Okay, but you have overwrought religious adjectives to describe all of this, right?

> [Dean’s] “criminality” was not something that sulked and sneered; it was a wild yea-saying overburst of American joy; it was Western, the west wind, an ode from the Plains, something new, long prophesied, long a-coming.

I feel like once you steal like a dozen cars in the space of a single book, you lose the right to have the word “criminality” in scare quotes.

But please, tell us more:

> [Ed and Dean] had just been laid off from the railroad. Ed had met a girl called Galatea who was living in San Francisco on her savings. These two mindless cads decided to bring the girl along [on one of their seven zillion pointless cross-country trips] and have her foot the bill. Ed cajoled and pleaded; she wouldn’t go unless he married her. In a whirlwind few days Ed Dunkel

> married Galatea, with Dean rushing around to get the necessary papers, and a few days before Christmas they rolled out of San Francisco at seventy miles per, headed for LA and the snowless southern road. In LA they picked up a sailor in a travel bureau and took him along for fifteen dollars' worth of gas…All along the way Galatea Dunkel, Ed's new wife, kept complaining that she was tired and wanted to sleep in a motel. If this kept up they'd spend all her money long before Virginia. Two nights she forced a stop and blew tens on motels. By the time they got to Tucson she was broke. Dean and Ed gave her the slip in a hotel lobby and resumed the voyage alone, with the sailor, and without a qualm.

All right, Jack, how are you gonna justify *this* one?

> Dean was simply a youth tremendously excited with life, and though he was a con-man he was only conning because he wanted so much to live and to get involved with people who would otherwise pay no attention to him.

I too enjoy life. Yet somehow this has never led me to get my friend to marry a woman in order to take her life savings, then leave her stranded in a strange city five hundred miles from home after the money runs out.

Jack Kerouac's relationship with Dean can best be described as "enabler". He rarely commits any great misdeeds himself. He's just along for the ride [usually literally, generally in flagrant contravention of all applicable traffic laws] with Dean, watching him destroy people's lives, doing nothing about it, and then going into rhapsodies about how free-spirited and unencumbered and holy and mad and visionary it all is.

There’s a weird tension here, because Jack is determined to totally ignore the moral issues. He brings this kind of stuff up only incidentally, as Exhibits A and B to support his case that Dean Moriarty is the freest and most perfect and most wonderful human being on Earth, and sort of moves past it before it becomes awkward. An enthusiastic reader, caught up in the spirit of the book, might easily miss it. The only place it is ever made explicit is page 185, when Galatea (who has since found her way back to San Francisco) confronts Dean about the trail of broken lives he’s left behind him, saying:

> You have absolutely no regard for anybody but yourself and your damned kicks. All you think about is what’s hanging between your legs and how much money or fun you can get out of people and then you just throw them aside. Not only that, but you’re silly about it. It never occurs to you that life is serious and there are people trying to make something decent out of it instead of just goofing all the time.”

This, 185 pages in, is the first and last time anyone seriously tries to criticize Dean. Dean has stolen about a dozen cars. He has married one woman, had an affair with another, played the two of them off against each other, divorced the first, married the second, deserted the second with a young child whom she has no money to support, gone back to the first, dumped the first again so suddenly she has to become a prostitute to make ends meet. Later he will go back to the second, beat the first so hard that he injures his thumb and has to get it amputated, break into the second’s house with a gun to kill her but change his mind, desert the second again also with a child whom she has no money to support, start dating a third, desert the third *also* with a child whom she has no money to support, and go back to the second, all while having like twenty or thirty lesser affairs on the side.

As quoted above, he dumped poor Galatea in Tucson, and later he will dump Jack in Mexico because Jack has gotten deathly ill and this is cramping his style.

So Galatea's complaint is not exactly coming out of thin air.

Jack, someone has just accused your man-crush of being selfish and goofing off all the time. Care to defend him with overwrought religious adjectives?

> That's what Dean was, the HOLY GOOF…he was BEAT, the root, the soul of beatific. What was he knowing? He tried all in his power to tell me what he was knowing, and they envied that about me, my position at his side, defending him and drinking him in as they once tried to do

Right. That's the problem. People are just *jealous*, because holy ecstatic angelic Dean Moriarty likes you more than he likes them. Get a life.

II.

But of course getting a life – in the sense of a home, a stable relationship, a steady job, et cetera – is exactly what all the characters in *On The Road* are desperately trying to avoid.

"Beat" has many meanings, but one of them is supposed to be "beaten down". The characters consider themselves oppressed, on the receiving end of a system that grinds them up and spits them out. This is productively compared with their total lack of any actual oppression whatsoever.

I don't know if it's the time period or merely their personal charm, but Kerouac et al's ability to do anything (and anyone) and get away with it is astounding. Several of their titular cross-country trips are performed entirely by hitchhiking, with their drivers often willing to buy them food along the way. Another is performed in some sort of incredibly ritzy Cadillac limo, because a rich man wants his Cadillac transported from Denver to Chicago, Dean volunteers, and the rich man moronically accepts. Dean of course starts driving at 110 mph, gets in an accident, and ends up with the car half destroyed. Once in the city, Dean decides this is a good way to pick up girls, and:

> In his mad frenzy Dean backed up smack on hydrants and tittered maniacally. By nine o' clock the the car was an utter wreck: the brakes weren't working any more; the fenders were stove in; the rods were rattling. Dean couldn't stop it at red lights; it kept kicking convulsively over the roadway. It had paid the price of the night. It was a muddy boot and no longer a shiny limousine…'Whee!' It was now time to return the Cadillac to the owner, who lived out on Lake Shore Drive in a swank apartment with an enormous garage underneath managed by oil-scarred Negroes. The mechanic did not recognize the Cadillac. We handed the papers over. He scratched his head at the sight of it. We had to get out fast. We did. We took a bus back to downtown Chicago and that was that. And we never heard a word from our Chicago baron about the condition of his car, in spite of the fact that he had our addresses and could have complained.

Even more interesting than their ease of transportation to me was their ease at getting jobs. This is so obvious to them it is left unspoken. Whenever their money runs out, be they in Truckee or Texas or Toledo, they just hop over to the nearest farm or factory or whatever, say “Job, please!” and are earning back their depleted savings in no time. This is really the crux of their way of life. They don't feel bound to any one place, because traveling isn't really a risk. Be it for a week or six months, there's always going to be work waiting for them when they need it. It doesn't matter that Dean has no college degree, or a criminal history a mile long, or is only going to be in town a couple of weeks. This just seems to be a background assumption. It is most obvious when it is violated; the times it takes an entire week to find a job, and they are complaining bitterly. Or the time the only jobs available are backbreaking farm labor, and so Jack moves on (of course abandoning the girl he is with at the time) to greener pastures that he knows are waiting.

Even more interesting than their ease of employment is their ease with women. This is unintentionally a feminist novel, in that once you read it (at least from a modern perspective) you end up realizing the vast cultural shift that had to (has to?) take place in order to protect women from people like the authors. Poor Galatea Dunkel seems to have been more of the rule than the exception – go find a pretty girl, tell her you love her, deflower her, then steal a car and drive off to do it to someone else, leaving her unmarriageable and maybe with a kid to support. Then the next time you're back in town, look her up, give her a fake apology in order to calm her down enough for her to be willing to have sex with you again, and repeat the entire process. Here is a typical encounter with a pretty girl:

> Not five nights later we went to a party in New York and I saw a girl called Inez and told her I had a friend with me that she ought to meet sometime. I was drunk and told her he was a cowboy. "Oh, I've always wanted to meet a cowboy."
>
> "Dean?" I yelled across the party. "Come over here, man!" Dean came bashfully over. An hour later, in the drunkenness and chiciness of the party, he was kneeling

> on the floor with his chin on her belly and telling her and promising her everything and sweating. She was a big, sexy brunette – as Garcia said, something straight out of Degas, and generally like a beautiful Parisian coquette. In a matter of days they were dickering with Camille in San Francisco by long-distance telephone for the necessary divorce papers so they could get married. Not only that, but a few months later Camille gave birth to Dean's second baby, the result of a few nights' rapport early in the year. And another matter of months and Inez had a baby. With one illegitimate child on the West somewhere, Dean then had four little ones, and not a cent, and was all troubles and ecstasy and speed as ever.

In case you're wondering, Dean then runs off to Mexico, leaves Inez behind, screws a bunch of Mexican women, and eventually gets back with Camille, who is happy to have him. Seriously, if I had read this book when I was writing Radicalizing The Romanceless, Dean (and his friends) would have been right up there with Henry as Exhibit B. The only punishment he ever gets for his misadventures is hitting one girlfriend in the face so hard that he breaks his own thumb, which gets infected and has to be amputated. Human justice has failed so miserably, one feels, that God has to personally step in.

As bad as the gender stuff is, the race stuff is worse. This is 1950-something, so I'm prepared for a lot of awful stuff regarding race. But this is totally *different* awful stuff regarding race than I expected. I have never been able to get upset over "exoticization" and "Orientalism" before, but this book reached new lows for me:

> At lilac evening I walked with every muscle aching among the lights of 27th and Welton in the Denver

> colored section, wishing I were a Negro, feeling that the best the white world had offered was not enough ecstasy for me, not enough life, joy, kicks, darkness, music, not enough night. I stopped at a little shack where a man sold hot red chili in paper containers; I bought some and ate it, strolling in the dark mysterious streets. I wished I were a Denver Mexican, or even a poor overworked Jap, anything but what I was so drearily, a "white man" disillusioned. All my life I'd had white ambitions; that was why I'd abandoned a good woman like Terry in the San Joaquin Valley…a gang of colored women came by, and one of the young ones detached herself from motherlike elders and came to me fast – "Hello Joe!" and suddenly saw it wasn't Joe, and ran back blushing. I wished I were Joe. I was only myself, sad, strolling in this violet dark, this unbearably sweet night, wishing I could exchange worlds with the happy, true-hearted, ecstatic Negroes of America.

Negroes are holy and ecstatic. But only in the same way barns and cows are holy and ecstatic. One gets the suspicion that Jack Kerouac is not exactly interacting with any of this stuff, so much as using it as something he can have his overwrought religious feelings about.

The "heroes" of *On The Road* consider themselves ill-done by and beaten-down. But they are people who can go anywhere they want for free, get a job any time they want, hook up with any girl in the country, and be so clueless about the world that they're pretty sure being a 1950s black person is a laugh a minute.

On The Road seems to be a picture of a high-trust society. Drivers assume hitchhikers are trustworthy and will take them anywhere. Women assume men are trustworthy and will accept any promise. Employers assume workers are trustworthy and don’t bother with background checks. It’s pretty neat.

But *On The Road* is, most importantly, a picture of a high-trust society collapsing. And it’s collapsing precisely because the book’s protagonists are going around defecting against everyone they meet at a hundred ten miles an hour.

III.

The viewpoint of a character in a book is not necessarily the viewpoint of its author. One can write about terrible people doing terrible things and not necessarily endorse it. That having been said, it’s very hard to read Jack Kerouac-the-author as differing very much from Jack-Kerouac-the-character in his opinions. He still has a raging man-crush on Dean and thinks that he is some kind of holy madman who can do no wrong.

The nicest thing I can say about *On The Road* is that perhaps it should be read backwards. It is a paean to a life made without compromise, a life of enjoying the hidden beauty of the world, spent in pursuit of holiness and the exotic. Despite how I probably sound, I really respect the Beat aesthetic of searching for transcendence and finding it everywhere. There’s something to be said for living your life to maximize that kind of thing, especially if everyone else is some kind of boring dispirited factory worker or something. Kerouac wrote around the same time as Sartre; it’s not difficult to imagine him as one of the first people saying you needed to try to find your True Self.

Read backwards, there was a time when to spend your twenties traveling the world and sleeping with strange women and having faux mystical experiences was something new and exciting and dangerous and for all anybody knew maybe it held the secret to immense spiritual growth. But from a modern perspective, if Jack and Dean tried the same thing today, they'd be one of about a billion college students and aimless twenty-somethings with exactly the same idea, posting their photos to Instagram tagged "holy", "ecstatic", and "angelic". There's nothing wrong with that. But it doesn't seem like a good stopping-point for a philosophy. It doesn't even seem like good escapism. I'd be willing to tolerate all the pointless criminality if it spoke to the secret things that I've always wanted to do in my hidden heart of hearts, but I'd like to think there's more there than driving back and forth and going to what seem like kind of lackluster parties.

When I read Marx, I thought that his key mistake was a negative view of utopia. That is, utopia is what happens automatically once you overthrow all of the people and structures who are preventing there from being utopia. Just get rid of the capitalists, and the World-Spirit will take care of the rest. The thought that ordinary, fallible, non-World-Spirit humans will have to build the post-revolution world brick by brick, and there's no guarantee they will do any better than the pre-revolutionary humans who did the same, never seems to have occurred to him.

Kerouac was a staunch anti-Communist, but his beat philosophy seems to share the same wellspring. Once you get rid of all the shackles of society in your personal life – once you stop caring about all those squares who want you to have families and homes and careers and non-terrible friends – once you become a holy criminal who isn't bound by the law or other people's needs – then you'll end up with some ecstatic visionary true self. Kerouac claimed he was Catholic, that he was in search of the Catholic God, and that he found Him – but all of his descriptions of such tend to be a couple of minutes of rapture upon seeing some especially pretty woman in a nightclub or some especially dingy San Francisco alley, followed by continuing to be a jerk who feels driven to travel across the country approximately seven zillion times for no reason.

Like the early Communists, who were always playing up every new factory that opened as the herald of the new age of plenty, in the beginning it’s easy to tell yourself your revolution is succeeding, that you are right on the brink of the new age. But at last come the Andropovs and Brezhnevs of the soul, the stagnation and despair and the going through the motions.

Kerouac apparently got married and divorced a couple of times, became an alcoholic, had a bit of a breakdown, and drank himself to death at age 49. Moriarty spent a while in prison on sort-of-trumped-up drug charges, went through a nasty divorce with whichever wife hadn’t divorced him already, and died of a likely drug overdose at age 47.

Overall I did not like this book.

If you’re writing about a crime spree you were a part of, you ought to show at least a little self-awareness.

Mysticism continues to be a perfectly valid life choice, but I continue to believe if you want to pursue it you should do it carefully and methodically, for example meditating for an hour a day and then going to regular retreats run by spiritual authorities, rather than the counterculture route of taking lots of drugs and having lots of sex and reading some books on Gnosticism and hoping some kind of enlightenment smashes into you.

Professional writing should be limited to about four overwrought religious adjectives per sentence, possibly by law.

And travel and girls are both fun, but *[doctor voice]* should be enjoyed responsibly and in moderation.

Framing For Light Instead Of Heat

I.

Ezra Klein uses my analysis of race and justice as a starting point to offer a thoughtful and intelligent discussion of what exactly it means to control for something in a study.

I’m not really going to call it a critique of my piece, because it only applies to two of the six areas I looked at, and in those two areas Klein’s thoughts were already carefully integrated into my conclusion – I described both as showing “ambiguity over the level of racial bias, depending on…how strictly you define racial bias.” The Vox article repeated and expanded on that conclusion rather than contradicting it.

But it’s still an important issue and I’m glad it’s come up since I didn’t have time to deal with at enough length in the original post.

The argument is: any study worth its salt is going to control for things like income level. Therefore, a study that concludes “blacks and whites get arrested at about the same rates” may only mean “blacks and whites of the same income level get arrested at about the same rates”. If blacks on average have lower incomes, then in the real world blacks might still be arrested much more. Blacks being poor and therefore getting uniquely poor treatment from the criminal justice system (Klein says) sounds like *exactly* the sort of thing we would call “racial discrimination” or “racial bias” or “racism”, but it would be totally missed by the standard methodology of controlling for income.

The solution is terminological rigor, which I foolishly forgot to have. What I should have said at the beginning of my post was “I want to know whether there is any direct bias against black people caused by racist attitudes among police and other officials.” By this definition, all of my conclusions stand.

Klein wants to know whether there is any factor at all that causes disproportionate impact of the criminal justice system on any race. By this definition, my conclusions are only a tiny part of the picture, although at the end I recommend the book Malign Neglect which provides much of the rest.

As long as we keep these two meanings of “racial bias” or “racism” or whatever separate, there’s no problem. Once we start conflating them, we’re going to become very confused in one direction or another.

Ezra Klein and I don’t disagree about any point of statistics. What I think we do disagree about is the terminology.

If we find that much of the overrepresentation of blacks in the criminal justice system is because black people are often poor and poor people often get sucked into the system, should we describe this as “the problem isn’t racism in the criminal justice system, it’s poverty” or as “the problem is racism in the criminal justice system, as manifested through poverty”?

II.

Consider a town with 1000 black people and 1000 white people. 750 black people are poor, and 250 are rich. 750 white people are rich, and 250 are poor. Everyone commits crimes at the same rate – let’s say 10% per year. Rich people have lots of connections and can bribe their way out of trouble in a pinch, so only 50% of rich criminals get arrested. Poor people don't have any strings they can pull, so 100% of poor criminals get arrested.

We can do the calculations and determine that the black arrest rate will be 8.75% and the white arrest rate 6.25%, a pretty significant difference. The people in the town can do the calculations as well. They correctly observe that in their town, everyone commits crimes at the same rate, so there must be some bias in their system. Using Klein's definition, they determine that since the system in their town disproportionately affects blacks, their criminal justice system is racist.

The problem is, upon learning that your criminal justice system is racist, what solutions come to mind? The ones I think of include things like increasing the diversity of the officer pool, sending police to diversity training, ferreting out racist attitudes and comments among members of the force, urging officers to consume media that is more positive towards black people, et cetera.

But all of these are unrelated to the problem and will accomplish nothing. We specified the decision algorithm these officers use, and we know it has nothing to do with race and everything to do with class. The townspeople should be attacking the culture of bribery, nepotism, and corruption, not throwing away resources on curing racist attitudes that don't affect police behavior in the slightest.

Note that this is true *even if* the poverty is caused by racism. Suppose the town college unfairly admits whites and turns down blacks, which is why the white people in this town are so much richer. I have no problem with saying "the town college is racist". This suggests the appropriate solutions – educating and/or punishing the people at the college. I have a lot of problems with saying "the town police are racist” as a shortcut for “the town police take bribes, and due to racism somewhere else the people with the cash are all white” because this obfuscates the correct solution.

You can’t just cut links out of a causal chain and preserve meaning. “Blacks are arrested disproportionately often because of gravity” is true, insofar as without the formation of the Earth from the gravitational coalescence of a primordial gas cloud humans and therefore racism wouldn’t exist. But if the natural reaction to hearing the phrase is to solve the problem by attaching hundreds of helium balloons to black people, then say something less misleading.

III.

Klein goes on to say:

> An example is research around the gender wage gap, which tries to control for so many things that it ends up controlling for the thing it’s trying to measure. As my colleague Matt Yglesias wrote, the commonly cited statistic that American women suffer from a 23 percent wage gap through which they make just 77 cents for every dollar a man earns is much too simplistic. On the other hand, the frequently heard conservative counterargument that we should subject this raw wage gap to a massive list of statistical controls until it nearly vanishes is an enormous oversimplification in the opposite direction. After all, for many purposes gender is itself a standard demographic control to add to studies — and when you control for gender the wage gap disappears entirely!
>
> The question to ask about the various statistical controls that can be applied to shrink the gender gap is what are

> they actually telling us," he continued. "The answer, I think, is that it's telling how the wage gap works.
>
> Take hours worked, which is a standard control in some of the more sophisticated wage gap studies. Women tend to work fewer hours than men. If you control for hours worked, then some of the gender wage gap vanishes. As Yglesias wrote, it's "silly to act like this is just some crazy coincidence. Women work shorter hours because as a society we hold women to a higher standard of housekeeping, and because they tend to be assigned the bulk of childcare responsibilities."
>
> Controlling for hours worked, in other words, is at least partly controlling for how gender works in our society. It's controlling for the thing that you're trying to isolate.

Once again, when someone says "women make seventy seven cents for each dollar a man earns", the response is almost always "That's outrageous!" and demands that companies stop being so sexist. I don't even have to speculate here. Google "gender wage gap", and just on the first page of results you find statements like:

"While some CEOs have been vocal in their commitment to paying workers fairly, American women can't wait for trickle-down change. The American Association of University Women urges companies to conduct salary audits to proactively monitor and address gender-based pay differences."

"Our project on sex and race discrimination in the workplace shows that outright discrimination in pay, hiring, or promotions continues to be a significant feature of working life…the Institute for Women's Policy Research examined organizational remedies such as sexual harassment training,

the introduction of new grievance procedures, supervisory training or revised performance management, and reward schemes.”

“Today marks Equal Pay Day, the date that symbolizes how far into the new year the average American woman would have to work to earn what the average American man did in the previous year. With a new executive order issued today, President Obama and Democrats are hoping to peg the gender wage gap as a major issue ahead of the 2014 elections. This week, Senate Democrats also plan to again bring forward the proposed “Paycheck Fairness Act,” a bill that aims to eliminate the pay gap between female and male employees. Both men and women see a need for moves such as this – 72% of women and 61% of men said “this country needs to continue making changes to give men and women equality in the workplace”

Given that the supposed gender pay gap is being used at this very moment to argue for salary audits, sexual harassment training, grievance procedures, and paycheck fairness acts, isn’t it really important to know that a lot of it is due to upstream factors like how men and women are socialized as children to have different values, which wouldn’t be affected by these things at all?

(Given that the entire issue is probably being used to load the term “feminism” with positive affect, isn’t it important to know that it’s mostly unrelated to what we expect feminists to do with their extra trust and power?)

It might be worthwhile to come at this from an ideologically opposite angle. Suppose I state “Professors who identify as feminist give twice as many As to female students as they do to male students.”

(This is true, by the way.)

It sounds like a big problem. So you dig through mountains of data, and you figure out that most feminist professors tend to be in subjects like the humanities, where twice as many students are female as male, and so naturally twice as many of the As go to women as men. If I just give you my best trollface and say “Yes, that’s certainly the *mechanism* by which the extra female As occur”, you have every reason to believe I’m deliberately causing trouble. Especially if colleges have already vowed to stop hiring feminist professors in response to the subsequent outrage. Especially especially if you know I am a cultural conservative activist whose goal has always been to make colleges stop hiring feminist professors, by hook or by crook.

If twice as many women as men take English literature classes, that’s compatible with something about gender socialization unfairly making men feel less able to study or less enthusiastic about studying literature. That could be considered biased or discriminatory, I guess. But phrasing it as “feminist professors give twice as many As to women” is calculated to produce maximal damage. It’s the sort of thing you would only do if you wanted to throw a match on a gunpowder keg for s**ts and giggles.

IV.

So I guess I’ve moved on from “poor choice of terminology” to “active misrepresentation”. Let’s stick with that.

This issue makes for the ultimate motte-and-bailey doctrine.

You go around saying “Gender gap! Women making less than men! Discrimination! Sexism!” Everyone puts on their Gricean implicature caps and concludes that they mean what these words mean in everyday speech. The appropriate remedies are trotted out – companies need to raise their female employees' pay, companies need to hire more discrimination officers, feminists need to talk more about all the ways men talk over women in the workplace and mansplain to them, etc. This is the bailey.

Then someone says "Wait, according to our study, a lot of this is just that women prefer working shorter hours to have time with their families" – and so they retreat to their motte: "Yeah, that's the *mechanism* for the gender gap. You mad, bro?"

But the thing about mottes is that nobody actually cares about them when there's this awesome bailey they can fight over instead. By turning differential socialization into the motte for sexual harassment or something, we're doing a disservice not only to sexual harassment, but to the principled study of differential socialization.

Anyway, the situation is actually even worse than this. If you hear "The problem with the criminal justice system is disproportionate impact on the poor," then you'll probably start coming up with ideas for how to deal with that, and other people will probably start listening. If you hear "the problem with the criminal justice system is racism," then you will start sharpening your knives.

Racism is a *uniquely* divisive issue. Minorities hear it and think of Klansmen trying to kill them. White people hear it and think of witch-hunters trying to get them fired. A single death in a random Midwestern town has turned half the country into experts on ballistics because it involved race. Bring up race, and people will change their opinion in the opposite direction suggested by the evidence just to spite you for having a different opinion about it than they do.

Once you've said words like "racism" or "racial bias", this dynamic is already in play and you have lost control of the conversation from then on. If you mention the word and then suggest that we should do something about the police bribery or whatever, then ten percent are going to yell "HOW DARE YOU IMPUGN OUR OFFICERS' HONOR, YOU POLITICALLY CORRECT FASCIST", another ten percent are going to yell "HOW DARE YOU DERAIL THE CONVERSATION ABOUT RACE, YOU WHITE SUPREMACIST ASSHOLE", and the other eighty percent are going to be yelling so loud at each other they can't even hear you. By the time all the fires have been put out and all the rubble cleared, it's a pretty good bet that nobody is in the mood to hear about policy ideas for reducing the impact of police on lower-income individuals anymore.

Klein ends his piece by interviewing a professor who states that "Liberals sometimes overstate the extent of overt racism as a direct explanation of justice system disparities." He acts like this is some sort of inexplicable quirk of the liberal mind. I wonder whether it might have more to do with liberals reading things like the recent Vox article, "America's Criminal Justice System Is Racist", which declares the thesis "There is no reason to be subtle on this point: the American criminal justice system is racist", then goes on to repeat the phrase "America's criminal justice system is racist" five times in the next five paragraphs. It never mentions that possibility that any of this racism is anything but overt.

If, like Robin Hanson, you believe in the metaphor of tugging policy ropes sideways, then I can't think of any worse way to ensure that everyone will be tugging against you in every direction than trying to focus the discussion about race.

That’s why I limited my review to direct bias within the justice system itself, and why I think other ways of framing the issue are less productive.

(Comment screening is on again, I guess. Comments that will start flame wars or derail the conversation will vanish into the aether. Unrelated: the book review yesterday got popular and this blog might go down every so often because of too much traffic. It’ll be up again shortly.)

A Story With Zombies

(inspired by Zombies: Seriously, Enough, Zombies Are So Overdone, and Scifi/Fantasy Stories Editors Are Tired Of Seeing: Zombies)

He walked into my office and threw the manuscript on my desk with a thud.

“It’s called *Thankful For Zombies*. A zombie story where…”

“Nope,” I said.

His face deflated like a balloon. “But I didn’t even…”

“Zombies are overdone,” I said.

“But this is a zombie story with a twist!”

“Zombie stories with twists are *super* overdone.”

“But this is a story about an extended family who get together for Thanksgiving dinner, only to be interrupted by a zombie apocalypse. It’s a Thanksgiving story about zombies. You have to admit that the combination of zombies and Thanksgiving has never…”

“Done,” I said.

“Wait, really? The family starts out estranged and suspicious of each other, but then when they all have to work together to…”

“Done,” I said.

“How could that have been done?”

“Listen. I know you won’t believe me, but for the past ten years or so, the best literary minds of our generation have been working on creating zombie stories *just* different enough from every other zombie story around to get published. First the clever and interesting twists got explored. Then the mediocre and boring twists. Then the absurd and idiotic twists. Finally the genre got *entirely mined out*. There is now a New York Times bestselling book about zombies invading Jane Austen’s *Pride and Prejudice*. If your idea isn’t weirder than that, *it’s been done*. And that’s the logical ‘if’. If your idea *is* weirder than that, *it has also been done*.”

“I *will* get *Thankful for Zombies* published,” he said.

“You won’t,” I advised him.

“I just have to think of an original angle.”

“You really won’t,” I told him.

“The zombies are the good guys,” he proposed.

“Done.”

“The zombies are smarter than humans.”

“Done.”

“In the end, we ourselves are the zombies.”

“Done.”

“A human girl falls in love with a zombie.”

“Done.”

“Okay, fine. Toss the Thanksgiving angle. There’s got to be some zombie plot that will be fresh and new.”

“I promise you, there’s not.”

“Zombies in space.”

“Done.”

“Zombies *from* space.”

“Done.”

“Zombies *are* space.”

“Done.”

“Zombies in Victorian England.”

“Done‘”

“Zombies in Edwardian England.”

“Done.”

“Zombies in Shakespearean England.”

“Done.”

“Shakespeare was a zombie, and all of his plays are just the word BRAAAAAAIIINS repeated over and over again.”

“Done, for some reason.”

“A young zombie comes of age.”

“Done.”

“A middle-aged zombie wonders if her single-minded focus on career success has prevented her from becoming the kind of zombie she wanted to be when she was younger.”

“Done.”

“An elderly zombie contemplates death.”

“Zombies are already dead.”

“Then I can…”

“…and yet it’s still been done.”

“A zombie in the Vietnam War.”

“Done.”

“A hippie zombie at Woodstock.”

“Done.”

“Strong female zombies.”

“Done.”

“Jewish zombies.”

“Done.”

“Black zombies.”

“Done.”

“A gay zombie struggling to fit into a homophobic zombie society.”

“Come on, this is the 21st century. Done like ten times. One of them won the Booker.”

“Gender-questioning zombies.”

“Done.”

“An immigrant zombie comes to America, with nothing but the decaying shirt on his back, knowing only a single word of English.”

“*All* zombies only know a single word of English. Also, done.”

“Nazi zombies.”

“Done.”

“Vampire zombies.”

“Done.”

“Pirate zombies.”

“Done.”

“Obstetrician/gynaecologist zombies.”

“Done.”

“Zombie Hitler.”

“Done.”

“Zombie Henry VIII.”

“Done“.

“But what if it was told from the perspective of Anne Boleyn?”

“Done.”

“Zombie Leonardo da Vinci.”

“Done.”

“Zombie Jesus.”

“Done. By three guys named Matt, Luke, and John.”

“Zombie Buddha.”

“Done.”

“Zombie Mohammed.”

“Done. As is the author, if you get my drift.”

“Zombie Zoroaster.”

“Done.”

“A parody subverting zombie stories.”

“Super done.”

“A parody subverting zombie stories lampshading how overdone they are.”

“Super duper done.”

“Hmmmm.” He thinks for a second. “Hold on, I’m remembering something from my college math class that might work here. You take all the zombie novels ever written, and you put them in some well-ordering, for example from first to last published. Then you make a new novel, consisting of the first page of the first novel, the second page of the second novel, and so on. But you change each page just a little bit. Since we know the first page of the new novel is different from the first page of the first novel, and the second page of the new novel is different from the second page of the second novel, by extension we know that there is at least one page on which the new novel is different from each zombie novel currently in existence. That means that the new story is provably original."

"Done."

"I don't think you understand; it's mathematically impossible for…"

"No, I mean there's a story about a zombie doing that."

"Oh." He furrowed his brow. "A zombie superhero."

"Done."

"Steampunk zombies."

"Done. I think now you're just trolling me."

"Motorcycle gangs of zombies."

"Done."

"A zombie story that's a metaphor for how…"

"Done."

"I didn't finish!"

"You didn't have to."

“A zombie gets cancer.”

“Done.”

“A zombie gets depression.”

“Done.”

“A zombie tries to write zombie fiction.”

“Done.”

“A zombie tries to write zombie fiction *about* a zombie trying to write zombie fiction.”

“Done.”

“A zombie tries to…”

“It’s done all the way down.”

“Young free-spirited zombies trying to see America.”

“Done.”

“A story that starts off as being about a fantasy society of knights and damsels, but at the very end it’s revealed everyone is a zombie.”

“Done.”

“A story that starts off as being about a young woman’s struggle to succeed in 1980s Wall Street, but at the very end it’s revealed everyone is a zombie.”

“Done.”

"A story that starts off as being a paleontology textbook about the fauna of the Lower Cretaceous, but at the very end it's revealed everyone is a zombie."

"Twist zombie endings are *done*."

"A zombie…a zombie riding a giant purple emu through 17th century Ireland teams up with the pre-ghost of Thomas Jefferson to investigate a crime in which time-traveling flamboyantly gay sapient hippos have murdered the Secret Protestant Pope in order to initiate the Jain apocalypse, with liberal quotations from and allusions to the works of Edgar Allen Poe Thomas Pynchon and the medieval Rolandic cycle, and also the whole thing is a metaphor for Republican resistance to climate change legislation."

I thought for a moment. "Okay," I said. "That particular plot has not, technically, been done. But no one would read it."

"They will," he said.

"You'd be wasting your time to write it."

"I'm writing it," he said.

"Suit yourself. Put it on my desk when you're finished, and I'll take a look at it. But your chances aren't good."

"I don't care," he said, and left.

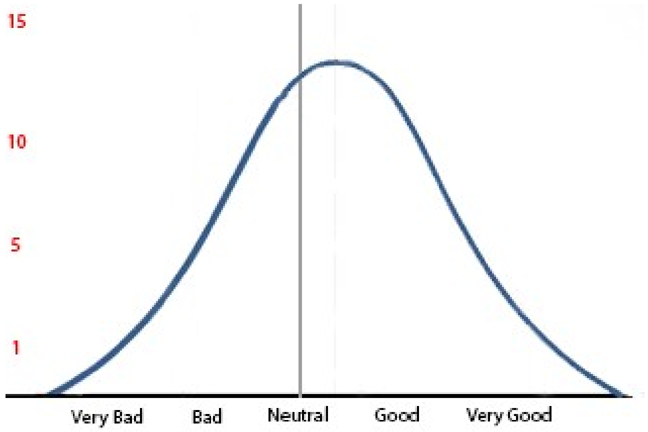
I sighed, finished up my last couple of pieces of paperwork, and shambled home from the office. On the way out, I ate my secretary's brain.

Beware The Man Of One Study

Aquinas famously said: beware the man of one book. I would add: beware the man of one study.

For example, take medical research. Suppose a certain drug is weakly effective against a certain disease. After a few years, a bunch of different research groups have gotten their hands on it and done all sorts of different studies. In the best case scenario the average study will find the true result – that it's weakly effective.

But there will also be random noise caused by inevitable variation and by some of the experiments being better quality than others. In the end, we might expect something looking kind of like a bell curve. The peak will be at "weakly effective", but there will be a few studies to either side. Something like this:



We see that the peak of the curve is somewhere to the right of neutral – ie weakly effective – and that there are about 15 studies that find this correct result.

But there are also about 5 studies that find that the drug is very good, and 5 studies missing the sign entirely and finding that the drug is actively bad. There’s even 1 study finding that the drug is very bad, maybe seriously dangerous.

This is before we get into fraud or statistical malpractice. I’m saying this is what’s going to happen just by normal variation in experimental design. As we increase experimental rigor, the bell curve might get squashed horizontally, but there will still be a bell curve.

In practice it’s worse than this, because this is assuming everyone is investigating exactly the same question.

Suppose that the graph is titled “Effectiveness Of This Drug In Treating Bipolar Disorder”.

But maybe the drug is more effective in bipolar i than in bipolar ii (Depakote, for example)

Or maybe the drug is very effective against bipolar mania, but much less effective against bipolar depression (Depakote again).

Or maybe the drug is a good acute antimanic agent, but less effective at maintenance treatment (let’s stick with Depakote).

If you have a graph titled “Effectiveness Of Depakote In Treating Bipolar Disorder” plotting studies from “Very Bad” to “Very Good” – and you stick all the studies – maintenenence, manic, depressive, bipolar i, bipolar ii – on the graph, then you’re going to end running the gamut from “very bad” to “very good” even before you factor in noise and even before even before you factor in bias and poor experimental design.

So here’s why you should beware the man of one study.

If you go to your better class of alternative medicine websites, they don’t tell you “Studies are a logocentric phallocentric tool of Western medicine and the Big Pharma conspiracy.”

They tell you “medical science has proved that this drug is terrible, but ignorant doctors are pushing it on you anyway. Look, here’s a study by a reputable institution proving that the drug is not only ineffective, but harmful.”

And the study will exist, and the authors will be prestigious scientists, and it will probably be about as rigorous and well-done as any other study.

And then a lot of people raised on the idea that some things have Evidence and other things have No Evidence think *holy s**t, they’re right!*

On the other hand, your doctor isn’t going to a sketchy alternative medicine website. She’s examining the entire literature and extracting careful and well-informed conclusions from…

Haha, just kidding. She’s going to a luncheon at a really nice restaurant sponsored by a pharmaceutical company, which assures her that they would *never* take advantage of such an opportunity to shill their drug, they just want to raise awareness of the latest study. And the latest study shows that their drug is great! Super great! And your doctor nods along, because the authors of the study are prestigious scientists, and it’s about as rigorous and well-done as any other study.

But obviously the pharmaceutical company has selected one of the studies from the “very good” end of the bell curve.

And I called this “Beware The Man of One Study”, but it’s easy to see that in the little diagram there are like three or four studies showing that the drug is “very good”, so if your doctor is a little skeptical, the pharmaceutical company can say “You are right to be skeptical, one study doesn’t prove anything, but look – here’s another group that finds the same thing, here’s yet another group that finds the same thing, and here’s a replication that confirms both of them.”

And even though it looks like in our example the sketchy alternative medicine website only has one “very bad” study to go off of, they could easily supplement it with a bunch of merely “bad” studies. Or they could add all of those studies about slightly different things. Depakote is ineffective at treating bipolar depression. Depakote is ineffective at maintenance bipolar therapy. Depakote is ineffective at bipolar ii.

So just sum it up as “Smith et al 1987 found the drug ineffective, yet doctors continue to prescribe it anyway”. Even if you hunt down the original study (which no one does), Smith et al won’t say specifically “Do remember that this study is only looking at bipolar maintenance, which is a different topic from bipolar acute antimanic treatment, and we’re not saying anything about that.” It will just be titled something like “Depakote fails to separate from placebo in six month trial of 91 patients” and trust that the responsible professionals reading it are well aware of the difference between acute and maintenance treatments (hahahahaha).

So it’s not so much “beware the man of one study” as “beware the man of any number of studies less than a relatively complete and not-cherry-picked survey of the research”.

II.

I think medical science is still pretty healthy, and that the consensus of doctors and researchers is more-or-less right on most controversial medical issues.

(it’s the *uncontroversial* ones you have to worry about)

Politics doesn’t have this protection.

Like, take the minimum wage question (please). We all know about the Krueger and Card study in New Jersey that found no evidence that high minimum wages hurt the economy. We probably also know the counterclaims that it was completely debunked as despicable dishonest statistical malpractice. Maybe some of us know Card and Krueger wrote a pretty convincing rebuttal of those claims. Or that a bunch of large and methodologically advanced studies have come out since then, some finding no effect like Dube, others finding strong effects like Rubinstein and Wither. These are just examples; there are at least dozens and probably hundreds of studies on both sides.

But we can solve this with meta-analyses and systemtic reviews, right?

Depends which one you want. Do you go with this meta-analysis of fourteen studies that shows that any presumed negative effect of high minimum wages is likely publication bias? With this meta-analysis of sixty-four studies that finds the same thing and discovers no effect of minimum wage after correcting for the problem? Or how about this meta-analysis of fifty-five countries that does find effects in most of them? Maybe you prefer this systematic review of a hundred or so studies that finds strong and consistent effects?

Can we trust news sources, think tanks, econblogs, and other institutions to sum up the state of the evidence?

CNN claims that 85% of credible studies have shown the minimum wage causes job loss. But raisetheminimumwage.com declares that “two decades of rigorous economic research have found that raising the minimum wage does not result in job loss…researchers and businesses alike agree today that the weight of the evidence shows no reduction in employment resulting from minimum wage increases.” Modeled Behavior says “the majority of the new minimum wage research supports the hypothesis that the minimum wage increases unemployment.” The Center for Budget and Policy Priorities says “The common claim that raising the minimum wage reduces employment for low-wage workers is one of the most extensively studied issues in empirical economics. The weight of the evidence is that such impacts are small to none.”

Okay, fine. What about economists? They seem like experts. What do they think?

Well, five hundred economists signed a letter to policy makers saying that the science of economics shows increasing the minimum wage would be a bad idea. That sounds like a promising consensus…

..except that six hundred economists signed a letter to policy makers saying that the science of economics shows increasing the minimum wage would be a *good* idea. (h/t Greg Mankiw)

Fine then. Let’s do a formal survey of economists. Now what?

raisetheminimumwage.com, an unbiased source if ever there was one, confidently tells us that “indicative is a 2013 survey by the University of Chicago’s Booth School of Business in which leading economists agreed by a nearly 4 to 1 margin that the benefits of raising and indexing the minimum wage outweigh the costs.”

But the Employment Policies Institute, which sounds like it’s trying *way* too hard to sound like an unbiased source, tells us that “Over 73 percent of AEA labor economists believe that a significant increase will lead to employment losses and 68 percent think these employment losses fall disproportionately on the least skilled. Only 6 percent feel that minimum wage hikes are an efficient way to alleviate poverty.”

So the whole thing is fiendishly complicated. But unless you look very very hard, you will never know that.

If you are a conservative, what you will find on the sites you trust will be something like this:

> Economic theory has always shown that minimum wage increases decrease employment, but the Left has never been willing to accept this basic fact. In 1992, they trumpeted a single study by Card and Krueger that purported to show no negative effects from a minimum wage increase. This study was immediately debunked and found to be based on statistical malpractice and “massaging the numbers”. Since then, dozens of studies have come out confirming what we knew all along – that a high minimum wage is economic suicide. Systematic reviews and meta-analyses (Neumark 2006, Boockman 2010) consistently show that an overwhelming majority of the research agrees on this fact – as do 73% of economists. That’s why five hundred top economists recently signed a letter urging policy makers not to buy into discredited liberal minimum wage theories. Instead of listening to starry-eyed liberal woo, listen to the empirical evidence and an overwhelming majority of economists and oppose a raise in the minimum wage.

And if you are a leftist, what you will find on the sites you trust will be something like this:

> People used to believe that the minimum wage decreased unemployment. But Card and Krueger's famous 1992 study exploded that conventional wisdom. Since then, the results have been replicated over fifty times, and further meta-analyses (Card and Krueger 1995, Dube 2010) have found no evidence of any effect. Leading economists agree by a 4 to 1 margin that the benefits of raising the minimum wage outweigh the costs, and that's why more than 600 of them have signed a petition telling the government to do exactly that. Instead of listening to conservative scare tactics based on long-debunked theories, listen to the empirical evidence and the overwhelming majority of economists and support a raise in the minimum wage.

Go ahead. Google the issue and see what stuff comes up. If it doesn't quite match what I said above, it's usually because they can't even muster *that* level of scholarship. Half the sites just cite Card and Krueger and call it a day!

These sites with their long lists of studies and experts are super convincing. And half of them are wrong.

At some point in their education, most smart people usually learn not to credit arguments from authority. If someone says "Believe me about the minimum wage because I seem like a trustworthy guy," most of them will have at least one neuron in their head that says "I should ask for some evidence". If they're *really* smart, they'll use the magic words "peer-reviewed experimental studies."

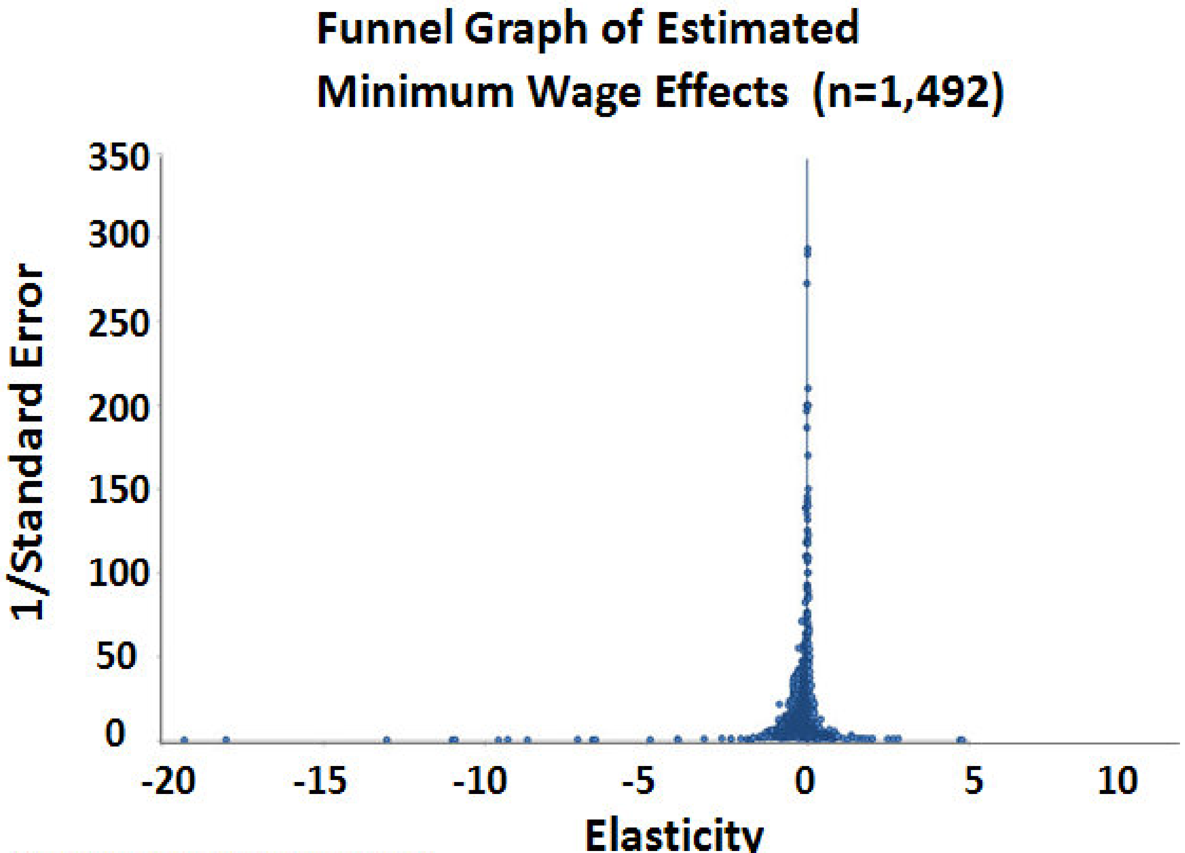
But I worry that most smart people have *not* learned that a list of dozens of studies, several meta-analyses, hundreds of experts, and expert surveys showing almost all academics support your thesis – can *still* be bullshit.

Which is too bad, because that's exactly what people who want to bamboozle an educated audience are going to use.

III.

I do not want to preach radical skepticism.

For example, on the minimum wage issue, I notice only one side has presented a funnel plot. A funnel plot is usually used to investigate publication bias, but it has another use as well – it's pretty much an exact presentation of the "bell curve" we talked about above.



Source: Doucouliagos and Stanley (2009)

This is more of a needle curve than a bell curve, but the point still stands. We see it's centered around 0, which means there's some evidence that's the real signal among all this noise. The bell skews more to left than to the right, which means more studies have found negative effects of the minimum wage than positive effects of the minimum wage. But since the bell curve is asymmetrical, we intepret that as *probably* publication bias. So all in all, I think there's at least some evidence that the liberals are right on this one.

Unless, of course, someone has realized that I've wised up to the studies and meta-analyses and and expert surveys, and figured out a way to hack *funnel plots*, which I am totally not ruling out.

(okay, I *kind of* want to preach radical skepticism)

Also, I should probably mention that it's much more complicated than one side being right, and that the minimum wage probably works differently depending on what industry you're talking about, whether it's state wage or federal wage, whether it's a recession or a boom, whether we're talking about increasing from $5 to $6 or from $20 to $30, etc, etc, etc. There are eleven studies on that plot showing an effect even worse than -5, and very possibly they are all accurate for whatever subproblem they have chosen to study – much like the example with Depakote where it might an effective antimanic but a terrible antidepressant.

(radical skepticism actually sounds a lot better than figuring this all out).

IV.

But the question remains: what happens when (like in most cases) you don't have a funnel plot?

I don't have a good positive answer. I do have several good *negative* answers.

Decrease your confidence about most things if you’re not sure that you’ve investigated every piece of evidence.

Do not trust websites which are obviously biased (eg Free Republic, Daily Kos, Dr. Oz) when they tell you they’re going to give you “the state of the evidence” on a certain issue, even if the evidence seems very stately indeed. This goes double for any site that contains a list of “myths and facts about X”, quadruple for any site that uses phrases like “ingroup member uses actual FACTS to DEMOLISH the outgroup’s lies about Y”, and octuple for RationalWiki.

Most important, even if someone gives you what seems like overwhelming evidence in favor of a certain point of view, don’t trust it until you’ve done a simple Google search to see if the opposite side has equally overwhelming evidence.

Debunked And Well-Refuted

I.

As usual, I was insufficiently pessimistic.

I infer this from *The Federalist*'s article on campus rape:

> A new report on sexual assault released today by the U.S. Department of Justice (DOJ) officially puts to bed the bogus statistic that one in five women on college campuses are victims of sexual assault. In fact, non-students are 25 percent more likely to be victims of sexual assault than students, according to the data. And the real number of assault victims is several orders of magnitude lower than one-in-five.

The article compares the older Campus Sexual Assault Survey (which found 14-20% of women were raped since entering college) to the just-released National Crime Victmization Survey (which found that 0.6% of female college students are raped per year). They write "Instead of 1 in 5, the real number is 0.03 in 5."

So the first thing I will mock *The Federalist* for doing is directly comparing per year sexual assault rates to per college career sexual assault rates, whereas obviously these are very different things. You can't *quite* just divide the latter by four to get the former, but that's going to work a heck of a lot better than *not* doing it, so let's estimate the real discrepancy as more like 0.5% per year versus 5% per year.

But I can't get too mad at them yet, because that's still a pretty big discrepancy.

However, faced with this discrepancy a reasonable person might say “Hmm, we have two different studies that say two different things. I wonder what’s going on here and which study we should believe?”

The Federalist staff said “Ha! There’s an old study with findings we didn’t like, but now there’s a new study with different findings we *do* like. So the old study is debunked!”

II.

My last essay, Beware The Man Of One Study, noted that one thing partisans do to justify their bias is selectively acknowledge studies from only one side of a complicated literature.

The reason it was insufficiently pessimistic is that there are also people like the Federalist staff, who acknowledge the existence of opposing studies, but only with the adjective “debunked” in front of them. By “debunked” they usually mean one of two things:

1. Someone on my side published a study later that found something else
2. Someone on my side accused it of having methodological flaws

Since the Federalist has so amply demonstrated the first failure mode, let me say a little more about the second. Did you know that *anyone* with a keyboard can just *type up* any of the following things?

– “That study is a piece of garbage that’s not worth the paper it’s written on.”
– “People in the know dismissed that study years ago.”
– “Nobody in the field takes that study seriously.”
– “That study uses methods that are laughable to anybody

who knows statistics.”
– “All the other research that has come out since discredits that study.”

They can say these things *whether they are true or not*. I’m kind of harping on this point, but it’s because it’s something *I* didn’t realize until much later than I should have.

There are many “questions” that are pretty much settled – evolution, global warming, homeopathy. But taking these as representative closes your mind and gives you a skewed picture of academia. On many issues, academics are just as divided as anyone else, and their arguments can be just as acrimonious as anyone else’s. The arguments usually take the form of one side publishing a study, the other side ripping the study apart and publishing their own study which they say is better, and the first side ripping the second study apart and arguing that their study was better all along.

Every study has flaws. No study has perfect methodology. If you like a study, you can say that it did the best it could on a difficult research area and has improved upon even-worse predecessor studies. If you don’t like a study, you can say “LOOK AT THESE FLAWS THESE PEOPLE ARE IDIOTS THE CONCLUSION IS COMPLETELY INVALID”. All you need to do is make enough isolated demands for rigor against anything you disagree with.

And so if the first level of confirmation bias is believing every study that supports your views, the second layer of confirmation bias is believing every supposed refutation that supports your views.

There are certainly things that have been “well-refuted” and “debunked”. Andrew Wakefield’s study purporting to prove that vaccines cause autism is a pretty good example. But you will notice that it had multiple failed replications, journals published reports showing he falsified data, the study's co-authors retracted their support, the journal it was published in retracted it and issued an apology, the General Medical Council convicted Wakefield of sixteen counts of misconduct, and Wakefield was stripped of his medical license and barred from practicing medicine ever again in the UK. The *British Medical Journal*, one of the best-respected medical journals in the world, published an editorial concluding:

> Clear evidence of falsification of data should now close the door on this damaging vaccine scare … Who perpetrated this fraud? There is no doubt that it was Wakefield. Is it possible that he was wrong, but not dishonest: that he was so incompetent that he was unable to fairly describe the project, or to report even one of the 12 children's cases accurately? No.

Wakefield's study has been "refuted". The rape study has been "argued against".

III.

I saw this same dynamic at work the other day, looking through the minimum wage literature.

The primordial titanomachy of the minimum wage literature goes like this. In 1994, two guys named Card and Krueger published a study showing the minimum wage had if anything positive effects on New Jersey restaurants, convincing many people that minimum wages were good. In 1996, two guys named Neumark and Wascher reanalyzed the New Jersey data using a different source and found that it showed the minimum wage had very bad effects on New

Jersey restaurants. In 2000, Card and Krueger responded, saying that their analysis was better than Neumark and Wascher’s re-analysis, and also they had done a re-analysis of their own which confirmed their original position.

Let’s see how conservative sites present this picture:

“The support for this assertion is the oft-cited 1994 study by Card and Krueger showing a positive correlation between an increased minimum wage and employment in New Jersey. Many others have thoroughly debunked this study.” (source)

“I was under the impression that the original study done by Card and Krueger had been thoroughly debunked by Michigan State University economist David Neumark and William Wascher” (source)

“The study … by Card and Krueger has been debunked by several different people several different times. When other researchers re-evaluated the study, they found that data collected using those records ‘lead to the opposite conclusion from that reached by’ Card and Krueger.” (source)

“It was only a short time before the fantastic Card-Krueger findings were challenged and debunked by several subsequent studies…in 1995, economists David Neumark and David Wascher used actual payroll records (instead of survey data used by Card and Krueger) and published their results in an NBER paper with an amazing finding: Demand curves for unskilled labor really do slope downward, confirming 200 years of economic theory and mountains of empirical evidence (source)

And now let’s look at how lefty sites present this picture:

"…a long-debunked paper [by Neumark and Wascher]" (source)

"Note that your Mises heroes, Neumark and Wascher are roundly debunked." (source)

"Neumark's living wage and minimum wage research have been found to be seriously flawed…based on faulty methods which when corrected refute his conclusion." – (source)

"…Neumark and Wascher, a study which Elizabeth Warren debunked in a Senate hearing" (source)

So if you're conservative, Neumark and Wascher debunked Card and Krueger. But if you're liberal, Card and Krueger debunked Neumark and Wascher.

Both sides are no doubt very pleased with themselves. They're not men of one study. They look at *all* of the research – except of course the studies that have been "debunked" or "well-refuted". Why would you waste your time with *those?*

IV.

Once again, I'm not preaching radical skepticism.

First of all, some studies are *super-debunked*. Wakefield is a good example.

Second of all, some studies that don't quite meet Wakefield-level of awfulness are indeed really bad and need refuting. I don't think this is beyond the intellectual capacities of most people. I think in many cases it's easy to understand why a study is wrong, you should try to do that, and once you do it you can safely discount the results of the study.

I’m not against pointing out when you disagree with studies or think they’re flawed. I’d be a giant hypocrite if I was.

But “debunked” and “refuted” aren’t saying you disagree with a study. They’re making arguments from authority. They’re saying “the authority of the scientific community has come together and said this is a piece of crap that doesn’t count”.

And that’s fine if that’s actually happened. But you had better make sure that you’re calling upon an ex cathedra statement by the community itself, and not a single guy with an axe to grind. Or one side of a complicated an interminable debate where both sides have about equal credentials and sway.

If you can’t do that, you say “I think that my side of the academic debate is in the right, and here’s why,” not “your side has been debunked”.

Otherwise you’re going to end up like the minimum wage debaters, where both sides claim to have debunked the other. Or like the Federalist article that says a study has been “put to bed” as “bogus” just because another study said something different.

I think this is part of my reply to the claim that empiricism is so great that no one needs rationality.

A naive empiricist who swears off critical thinking because they can just “follow the evidence” has no contingency plan for when the evidence gets confusing. Their only recourse is to deny that the evidence is confusing, to assert that one side or the other has been “debunked”. Since they’ve already made a principled decision not to study confirmation bias, chances are it’s going to be whichever side they don’t like that’s “already been debunked”. And by “debunked” they mean “a scientist on my side said it was wrong, so now I am relieved from the burden of thinking about it.”

On the original post, I wrote:

> Life is made up of limited, confusing, contradictory, and maliciously doctored facts. Anyone who says otherwise is either sticking to such incredibly easy solved problems that they never encounter anything outside their comfort level, or so closed-minded that they shut out any evidence that challenges their beliefs.

In the absence of any actual debunking more damning than a counterargument, “that’s been debunked” is the way “shuts out any evidence that challenges their beliefs” feels from the inside.

V.

Somebody’s going to want to know what’s up with the original rape studies. The answer is that a small part of the discrepancy is response bias on the CSAS, but most of it is that the two surveys encourage respondents to define “sexual assault” in very different ways. Vox has an excellent article on this which for once I 100% endorse.

In other words, both are valid, both come together to form a more nuanced picture of campus violence, and neither one “debunks” the other.

The Toxoplasma Of Rage

I.

Some old news I only just heard about: PETA is offering to pay the water bills for needy Detroit families if (and only if) those families agree to stop eating meat.

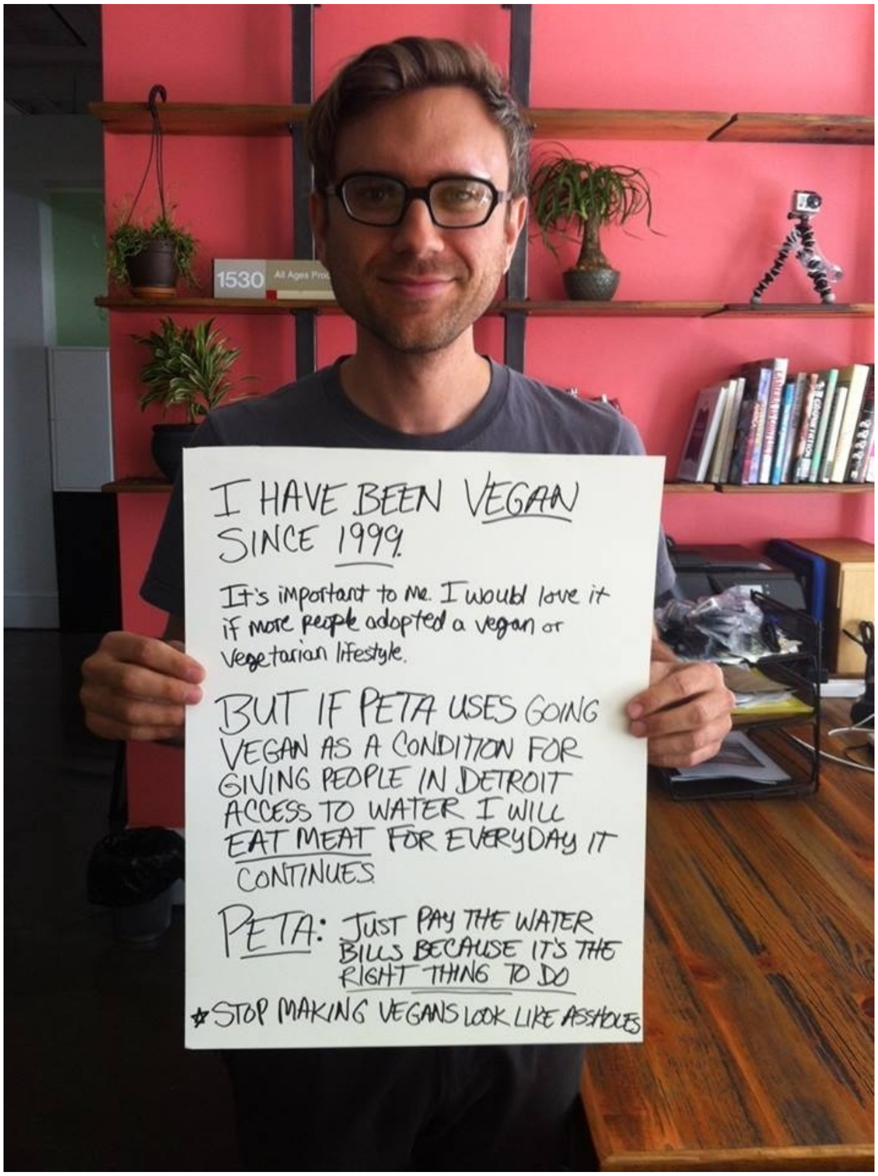
Predictably, the move caused a backlash. The International Business Times, in what I can only assume is an attempted pun, describes them as “drowning in backlash”. Groundswell thinks it’s a “big blunder”. Daily Banter says it’s “exactly why everyone hates PETA”. Jezebel calls them “assholes”.

Of course, this is par for the course for PETA, who have previously engaged in campaigns like throwing red paint on fashion models who wear fur, juxtaposing pictures of animals with Holocaust victims, juxtaposing pictures of animals with African-American slaves, and ads featuring naked people that cross the line into pornography.

People call these things “blunders”, but consider the alternative. Vegan Outreach is an extremely responsible charity doing excellent and unimpeachable work in the same area PETA is. Nobody has heard of them. *Everybody* has heard of PETA, precisely because of the interminable stupid debates about “did this publicity stunt cross the line?”

While not everyone is a vegan, most people who learn enough about factory farming are upset by it. There is pretty much zero room for PETA to convert people from pro-factory-farming to anti-factory-farming, because there aren’t any radical grassroots pro-factory-farming activists to be found. Their problem isn’t lack of agreement. It’s lack of attention.

PETA creates attention, but at a cost. Everybody’s talking about PETA, which is sort of like everybody talking about ethical treatment of animals, which is sort of a victory. But most of the talk is “I hate them and they make me really angry.” Some of the talk is even “I am going to eat a lot more animals just to make PETA mad.”



I HAVE BEEN VEGAN
SINCE 1999

It's important to me. I would love it if more people adopted a vegan or vegetarian lifestyle.

BUT IF PETA USES GOING VEGAN AS A CONDITION FOR GIVING PEOPLE IN DETROIT ACCESS TO WATER I WILL EAT MEAT FOR EVERYDAY IT CONTINUES

PETA: JUST PAY THE WATER BILLS BECAUSE IT'S THE RIGHT THING TO DO

☆ STOP MAKING VEGANS LOOK LIKE ASSHOLES

So there’s a tradeoff here, with Vegan Outreach on one side and PETA on the other.

Vegan Outreach can get everyone to agree in principle that factory-farming is bad, but no one will pay any attention to it.

And PETA can get everyone to pay attention to factory farming, but a lot of people who would otherwise oppose it will switch to supporting it just because they're so mad at the way it's being publicized.

But at least they're paying attention!

PETA doesn't shoot themselves in the foot because they're stupid. They shoot themselves in the foot because they're traveling up an incentive gradient that rewards them for doing so, even if it destroys their credibility.

II.

The University of Virginia rape case profiled in *Rolling Stone* has fallen apart. In doing so, it joins a long and distinguished line of highly-publicized rape cases that have fallen apart. Studies sometimes claim that only 2 to 8 percent of rape allegations are false. Yet the rate for allegations that go ultra-viral in the media must be an order of magnitude higher than this. As the old saying goes, once is happenstance, twice is coincidence, three times is enemy action.

The enigma is complicated by the observation that it's usually feminist activists who are most instrumental in taking these stories viral. It's not some conspiracy of pro-rape journalists choosing the most dubious accusations in order to discredit public trust. It's people specifically selecting these incidents as flagship cases for their campaign that rape victims need to be believed and trusted. So why are the most publicized cases so much more likely to be false than the almost-always-true average case?

Several people have remarked that false accusers have more leeway to make their stories as outrageous and spectacular as possible. But I want to focus on two less frequently mentioned concerns.

The Consequentialism FAQ explains signaling in moral decisions like so:

> When signaling, the more expensive and useless the item is, the more effective it is as a signal. Although eyeglasses are expensive, they're a poor way to signal wealth because they're very useful; a person might get them not because ey is very rich but because ey really needs glasses. On the other hand, a large diamond is an excellent signal; no one needs a large diamond, so anybody who gets one anyway must have money to burn.
>
> Certain answers to moral dilemmas can also send signals. For example, a Catholic man who opposes the use of condoms demonstrates to others (and to himself!) how faithful and pious a Catholic he is, thus gaining social credibility. Like the diamond example, this signaling is more effective if it centers upon something otherwise useless. If the Catholic had merely chosen not to murder, then even though this is in accord with Catholic doctrine, it would make a poor signal because he might be doing it for other good reasons besides being Catholic – just as he might buy eyeglasses for reasons beside being rich. It is precisely because opposing condoms is such a horrendous decision that it makes such a good signal.
>
> But in the more general case, people can use moral decisions to signal how moral they are. In this case, they choose a disastrous decision based on some moral principle. The more suffering and destruction they

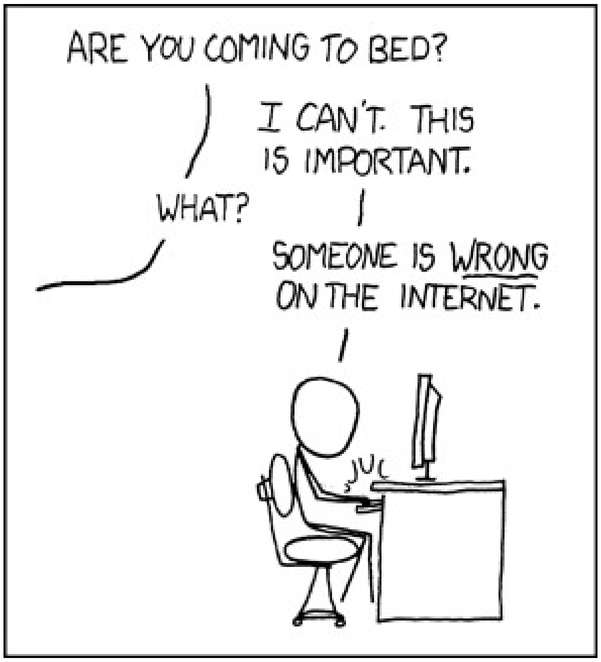
> support, and the more obscure a principle it is, the more obviously it shows their commitment to following their moral principles absolutely. For example, Immanuel Kant claims that if an axe murderer asks you where your best friend is, obviously intending to murder her when he finds her, you should tell the axe murderer the full truth, because lying is wrong. This is effective at showing how moral a person you are – no one would ever doubt your commitment to honesty after that – but it’s sure not a very good result for your friend.

In the same way, publicizing how strongly you believe an accusation that is obviously true signals nothing. Even hardcore anti-feminists would believe a rape accusation that was caught on video. A moral action that can be taken just as well by an outgroup member as an ingroup member is crappy signaling and crappy identity politics. If you want to signal how strongly you believe in taking victims seriously, you talk about it in the context of the least credible case you can find.

But aside from that, there’s the PETA Principle: the more controversial something is, the more it gets talked about.

A rape that obviously happened? Shove it in people’s face and they’ll admit it’s an outrage, just as they’ll admit factory farming is an outrage. But they’re not going to talk about it much. There are a zillion outrages every day, you’re going to need more than that to draw people out of their shells.

On the other hand, the controversy over dubious rape allegations is exactly that – a controversy. People start screaming at each other about how they’re misogynist or misandrist or whatever, and Facebook feeds get filled up with hundreds of comments in all capital letters about how my ingroup is being persecuted by your ingroup. At each step, more and more people get triggered and upset. Some of those triggered people do emergency ego defense by reblogging articles about how the group that triggered them are terrible, triggering further people in a snowball effect that spreads the issue further with every iteration.


ARE YOU COMING TO BED?
I CAN'T. THIS IS IMPORTANT.
WHAT?
SOMEONE IS WRONG ON THE INTERNET.

[source]

Only controversial things get spread. A rape allegation will only be spread if it’s dubious enough to split people in half along lines corresponding to identity politics. An obviously true rape allegation will only be spread if the response is controversial enough to split people in half along lines corresponding to identity politics – which is why so much coverage focuses on the proposal that all accused rapists should be treated as guilty until proven innocent.

Everybody hates rape just like everybody hates factory farming. “Rape culture” doesn’t mean most people like rape, it means most people ignore it. That means feminists face the same double-bind that PETA does.

First, they can respond to rape in a restrained and responsible way, in which case everyone will be against it and nobody will talk about it.

Second, they can respond to rape in an outrageous and highly controversial way, in which case everybody will talk about it but it will autocatalyze an opposition of people who hate feminists and obsessively try to prove that as many rape allegations as possible are false.

I have yet to see anyone holding a cardboard sign talking about how they are going to rape people just to make feminists mad, but it’s only a matter of time. Like PETA, their incentive gradient dooms them to shoot themselves in the foot again and again.

III.

Slate recently published an article about white people’s contrasting reactions to the Michael Brown shooting in Ferguson versus the Eric Garner choking in NYC. And man, it is *some* contrast.

A Pew poll found that of white people who expressed an opinion about the Ferguson case, 73% sided with the officer. Of white people who expressed an opinion about the Eric Garner case, 63% sided with the black victim.

Media opinion follows much the same pattern. Arch-conservative Bill O’Reilly said he was “absolutely furious” about the way “the liberal media” and “race hustlers” had “twisted the story” about Ferguson in the service of “lynch mob justice” and “insulting the American police community, men and women risking their lives to protect us”. But when it came to Garner, O’Reilly said he was “extremely troubled” and that “there was a police overreaction that should have been adjudicated in a court of law.” His guest on FOX News, conservative commentator and fellow Ferguson-detractor Charles Krauthammer added that “From looking at the video, the grand jury’s decision [not to indict] is totally incomprehensible.” Saturday Night Live did a skit about Al Sharpton talking about the Garner case and getting increasingly upset because “For the first time in my life, everyone agrees with me.”

This follows about three months of most of America being at one another’s throats pretty much full-time about Ferguson. We got treated to a daily diet of articles like Ferguson Protester On White People: “Y’all The Devil” or Black People Had The Power To Fix The Problems In Ferguson Before The Brown Shooting – They Failed or Most White People In America Are Completely Oblivious and a whole bunch of people sending angry racist editorials and counter-editorials to each other for months. The damage done to race relations is difficult to overestimate – CBS reports that they dropped ten percentage points to the lowest point in twenty years, with over half of blacks now describing race relations as “bad”.

And people say it was all worth it, because it raised awareness of police brutality against black people, and if that rustles some people’s jimmies, well, all the worse for them.

But the Eric Garner case also would have raised awareness of police brutality against black people, and everybody would have agreed about it. It has become increasingly clear that, given sufficiently indisputable evidence of police being brutal to a black person, pretty much everyone in the world condemns it equally strongly.

And it’s not just that the Eric Garner case came around too late so we had to make do with the Mike Brown case. Garner was choked a month before Brown was shot, but the story was ignored, then dug back up later as a tie-in to the ballooning Ferguson narrative.

More important, unarmed black people are killed by police or other security officers about twice a week according to official statistics, and probably much more often than that. You’re saying none of these shootings, hundreds each year, made as good a flagship case as Michael Brown? In all this gigantic pile of bodies, you couldn’t find *one* of them who hadn’t just robbed a convenience store? Not a single one who didn’t have ten eyewitnesses and the forensic evidence all saying he started it?

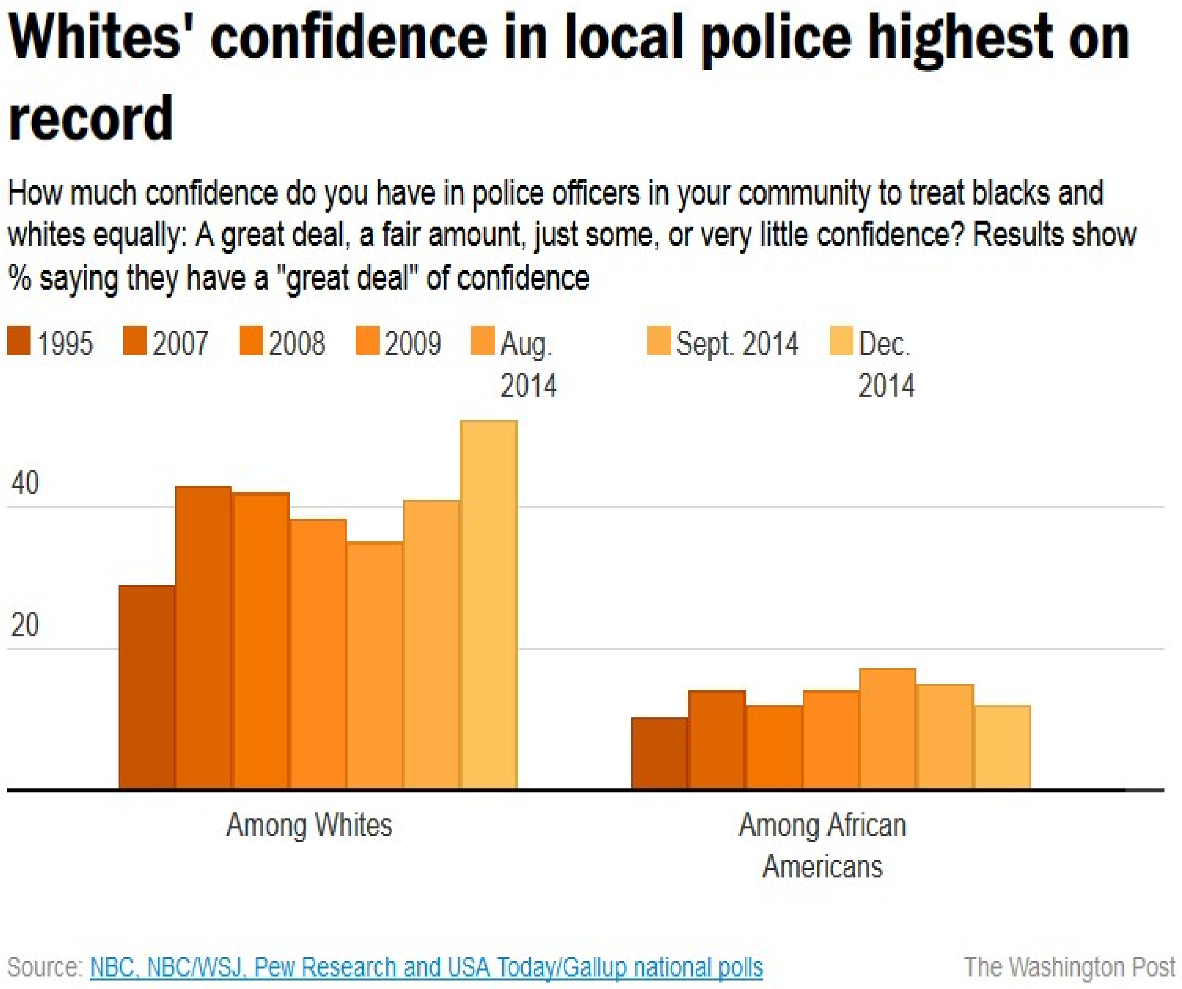
I propose that the Michael Brown case went viral – rather than the Eric Garner case or any of the hundreds of others – because of the PETA Principle. It was controversial. A bunch of people said it was an outrage. A bunch of other people said Brown totally started it, and the officer involved was a victim of a liberal media that was hungry to paint his desperate self-defense as racist, and so the people calling it an outrage were themselves an outrage. Everyone got a great opportunity to signal allegiance to their own political tribe and discuss how the opposing political tribe were vile racists / evil race-hustlers. There was a steady stream of potentially triggering articles to share on Facebook to provoke your friends and enemies to counter-share articles that would trigger *you*.

The Ferguson protesters say they have a concrete policy proposal – they want cameras on police officers. There’s only spotty polling on public views of police body cameras before the Ferguson story took off, but what there is seems pretty unanimous. A UK poll showed that 90% of the population of that country wanted police to have body cameras in February. US polls are more of the form “crappy poll widget on a news site” (1, 2, 3) but they all hovered around 80%

approval for the past few years. I also found a poll by Police Magazine in which a plurality of the *police officers* they surveyed wanted to wear body cameras, probably because of evidence that they cut down on false accusations. Even before Ferguson happened, you would have a really hard time finding anybody in or out of uniform who thought police cameras were a bad idea.

And now, after all is said and done, ninety percent of people are still in favor – given methodology issues, the extra ten percent may or may not represent a real increase. The difference between whites and blacks is a rounding error. The difference between Democrats and Republicans is barely worth talking about- 79% of Republicans are still in support. The people who think Officer Darren Wilson is completely innocent and the grand jury was right to release him, the people muttering under their breath about race hustlers and looters – *eighty percent of those people still want cameras on their cops.*

If the Ferguson protests didn't do much to the public's views on police body cameras, they sure changed its views on some other things. I wrote before about how preliminary polls say that hearing about Ferguson *increased* white people's confidence in the way the police treat race. Now the less preliminary polls are out, and they show the effect was larger than even I expected.



[source]

White people’s confidence in the police being racially unbiased increased from 35% before the story took off to 52% today. Could even a deliberate PR campaign by the nation’s police forces have done better? I doubt it.

It’s possible that this is an artifact of the question’s wording – after all, it asks people about their local department, and maybe after seeing what happened in Ferguson, people’s local police forces look pretty good by comparison. But then why do black people show the opposite trend?

I think this is exactly what it looks like. Just as PETA’s outrageous controversial campaign to spread veganism make people want to eat more animals in order to spite them, so the controversial nature of this particular campaign against police brutality and racism made white people like their local police department even more to spite the people talking about how all whites were racist.

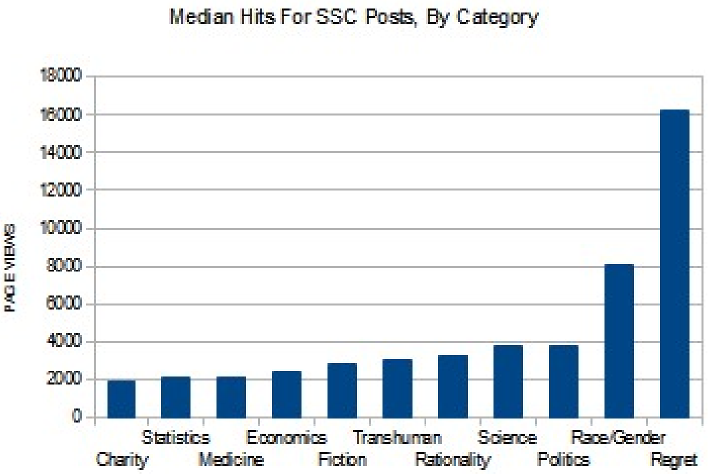
Once again, the tradeoff.

If campaigners against police brutality and racism were extremely responsible, and stuck to perfectly settled cases like Eric Garner, everybody would agree with them but nobody would talk about it.

If instead they bring up a very controversial case like Michael Brown, everybody will talk about it, but they will catalyze their own opposition and make people start supporting the police more just to spite them. More foot-shooting.

IV.

Here is a graph of some of the tags I commonly use for my posts, with the average number of hits per post in each tag. It’s old, but I don’t want to go through the trouble of making a new one, and the trends have stayed the same since then.



I blog about charity only rarely, but it must be the most important thing I can write about here. Convincing even a few more people to donate to charity, or to redirect their existing donations to a more effective program, can literally save dozens or even hundreds of lives even with the limited reach that a private blog has. It probably does more good for the world than all of the other categories on here combined. But it’s completely uncontroversial – everyone agrees it’s a good thing – and it is the least viewed type of post.

Compare this to the three most viewed category of post. Politics is self-explanatory. Race and gender are a type of politics even more controversial and outrage-inducing than regular politics. And that “regret” all the way on the right is my “things i will regret writing” tag, for posts that I know are going to start huge fights and probably get me in lots of trouble. They’re usually race and gender as well, but digging deep into the really really controversial race and gender related issues.

The less useful, and more controversial, a post here is, the more likely it is to get me lots of page views.

For people who agree with me, my angry rants on identity politics are a form of ego defense, saying “You’re okay, your in-group was in the right the whole time." Linking to it both raises their status as an in-group members, and acts as a potential assault on out-group members who are now faced with strong arguments telling them they're wrong. And the people who disagree with me will sometimes write angry rebuttals on their own blogs, and those rebuttals will link to my own post and spread it further. Or they'll talk about it with their disagreeing friends, and their friends will get mad and want to tell me I'm wrong, and come over here to read the post to get more ammunition for their counterarguments. I have a feature that tells me who links to all of my posts, so I can *see* this all happening in real-time.

I don't make enough money off the ads on this blog to matter much. But if I lived off them, which do you think I'd write more of? Posts about charity which only get me 2,000 paying customers? Or posts that turn all of you against one another like a pack of rabid dogs, and get me 16,000?

I don't have a fancy bar graph for them, but I bet this same hierarchy of interestingness applies to the great information currents and media outlets that shape society as a whole. It's in activists' interests to destroy their own causes by focusing on the most controversial cases and principles, the ones that muddy the waters and make people oppose them out of spite. And it's in the media's interest to help them and egg them on.

V.

And now, for something completely different.

Before "meme" meant doge and all your base, it was a semi-serious attempt to ground cultural evolution in parasitology. The idea was to replace a model of humans choosing whichever ideas they liked with a model of ideas as parasites that evolved in ways that favored their own transmission.

This never really caught on, because most people’s response was “That’s neat. So what?”

But let’s talk about toxoplasma.

Toxoplasma is a neat little parasite that is implicated in a couple of human diseases including schizophrenia. Its life cycle goes like this: it starts in a cat. The cat poops it out. The poop and the toxoplasma get in the water supply, where they are consumed by some other animal, often a rat. The toxoplasma morphs into a rat-compatible form and starts reproducing. Once it has strength in numbers, it hijacks the rat’s brain, convincing the rat to hang out conspicuously in areas where cats can eat it. After a cat eats the rat, the toxoplasma morphs back into its cat compatible form and reproduces some more. Finally, it gets pooped back out by the cat, completing the cycle.

It’s the ciiiiiircle of life!

What would it mean for a meme to have a life cycle as complicated as toxoplasma?

Consider the war on terror. They say that every time the United States bombs Pakistan or Afghanistan or somewhere, all we’re doing is radicalizing the young people there and making more terrorists. Those terrorists then go on to kill Americans, which makes Americans get very angry and call for more bombing of Pakistan and Afghanistan.

Taken as a meme, it’s a single parasite with two hosts and two forms. In an Afghan host, it appears in a form called ‘jihad’, and hijacks its host into killing himself in order to spread it to its second, American host. In the American host it morphs in a form called ‘the war on terror’, and it hijacks the Americans into giving *their* own lives (and tax dollars) to spread it back to its Afghan host in the form of bombs.

From the human point of view, jihad and the War on Terror are opposing forces. From the memetic point of view, they’re as complementary as caterpillars and butterflies. Instead of judging, we just note that somehow we accidentally created a replicator, and replicators are going to replicate until something makes them stop.

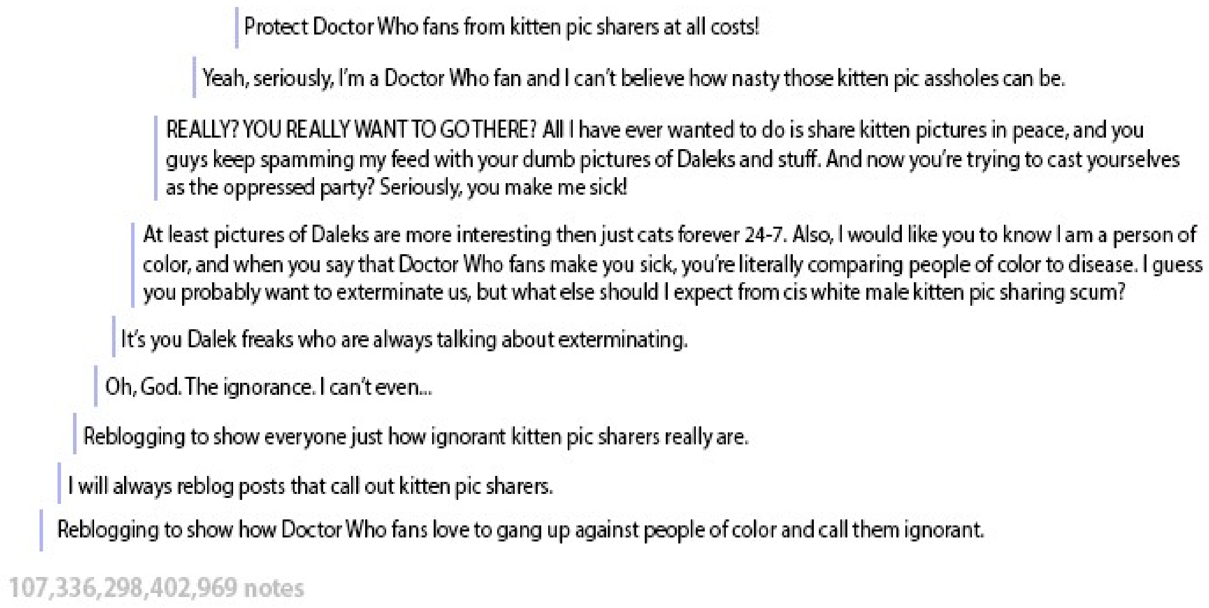
Replicators are also going to *evolve*. Some Afghan who thinks up a particularly effective terrorist strategy helps the meme spread to more Americans as the resulting outrage fuels the War on Terror. When the American bombing heats up, all of the Afghan villagers radicalized in by the attack will remember the really effective new tactic that Khalid thought up and do *that one* instead of the boring old tactic that barely killed any Americans at all. Some American TV commentator who comes up with a particularly stirring call to retaliation will find her words adopted into party platforms and repeated by pro-war newspapers. While pacifists on both sides work to defuse the tension, the meme is engaging in a counter-effort to become as virulent as possible, until people start suggesting putting pork fat in American bombs just to make Muslims *even madder*.

And let’s talk about Tumblr.

Tumblr’s interface doesn’t allow you to comment on other people’s posts, per se. Instead, it lets you reblog them with your own commentary added. So if you want to tell someone they’re an idiot, your only option is to reblog their entire post to all your friends with the message “you are an idiot” below it.

Whoever invented this system either didn’t understand memetics, or understood memetics *much too well*.

What happens is – someone makes a statement which is controversial by Tumblr standards, like “Protect Doctor Who fans from kitten pic sharers at all costs.” A kitten pic sharer sees the statement, sees red, and reblogs it to her followers with a series of invectives against Doctor Who fans. Since kitten pic sharers cluster together in the social network, soon *every* kitten pic sharer has seen the insult against kitten pic sharer – as they all feel the need to add their defensive commentary to it, soon all of them are seeing it from ten different directions. The angry invectives get back to the Doctor Who fans, and now *they* feel deeply offended, so they reblog it among themselves with even more condemnations of the kitten pic sharers, who now not only did whatever inspired the enmity in the first place, but have inspired extra hostility because their hateful invectives are right there on the post for everyone to see. So about half the stuff on your dashboard is something you actually want to see, and the other half is towers of alternate insults that look like this:



Protect Doctor Who fans from kitten pic sharers at all costs!

Yeah, seriously, I'm a Doctor Who fan and I can't believe how nasty those kitten pic assholes can be.

REALLY? YOU REALLY WANT TO GO THERE? All I have ever wanted to do is share kitten pictures in peace, and you guys keep spamming my feed with your dumb pictures of Daleks and stuff. And now you're trying to cast yourselves as the oppressed party? Seriously, you make me sick!

At least pictures of Daleks are more interesting then just cats forever 24-7. Also, I would like you to know I am a person of color, and when you say that Doctor Who fans make you sick, you're literally comparing people of color to disease. I guess you probably want to exterminate us, but what else should I expect from cis white male kitten pic sharing scum?

It's you Dalek freaks who are always talking about exterminating.

Oh, God. The ignorance. I can't even...

Reblogging to show everyone just how ignorant kitten pic sharers really are.

I will always reblog posts that call out kitten pic sharers.

Reblogging to show how Doctor Who fans love to gang up against people of color and call them ignorant.

107,336,298,402,969 notes

Actually, pretty much this happened to the PETA story I started off with

And then you sigh and scroll down to the next one. Unless of course you are a Doctor Who fan, in which case you sigh and then immediately reblog with the comment "It's obvious you guys started ganging up against us first, don't try to accuse **US** now" because you can't just *let that accusation stand*.

I make fun of Tumblr social justice sometimes, but the problem isn't with Tumblr social justice, it's structural. Every community on Tumblr somehow gets enmeshed with the people most devoted to making that community miserable. The tiny Tumblr rationalist community somehow attracts, concentrates, and constantly reblogs stuff from the even tinier Tumblr community of people who hate rationalists and want them to be miserable (no, well-intentioned and intelligent critics, I am not talking about you). It's like one of those rainforest ecosystems where every variety of rare endangered nocturnal spider hosts a parasite who has evolved for millions of years solely to parasitize that one spider species, and the parasites host parasites who have evolved for millions of years solely to parasitize *them*. If Tumblr social justice is worse than anything else, it’s mostly because everyone has a race and a gender so it’s easier to fire broad cannonades and just hit everybody.

Tumblr’s reblog policy makes it a hothouse for toxoplasma-style memes that spread via outrage. Following the ancient imperative of evolution, if memes spread by outrage they adapt to become as outrage-inducing as possible.

Or rather, that is just one of their many adaptations. I realize this toxoplasma metaphor sort of strains credibility, so I want to anchor this idea of outrage-memes in pretty much the only piece of memetics everyone can agree upon.

The textbook example of a meme – indeed, almost the only example ever discussed – is the chain letter. “Send this letter to ten people and you will prosper. Fail to pass it on, and you will die tomorrow.” And so the letter replicates.

It might be useful evidence that we were on the right track here, with our toxoplasma memes and everything, if we could find evidence that they reproduced in the same way.

If you’re not on Tumblr, you might have missed the “everyone who does not reblog the issue du jour is trash” wars. For a few weeks around the height of the Ferguson discussion, people constantly called out one another for not reblogging enough Ferguson-related material, or (Heavens forbid) saying they were sick of the amount of Ferguson material they were seeing. It got so bad that various art blogs that just posted pretty paintings, or kitten picture blogs that just reblogged pictures of kittens were feeling the heat (you thought I was joking about the hate for kitten picture bloggers. I never joke.) Now the issue du jour seems to be Pakistan. Just to give a few examples:

“friends if you are reblogging things that are not about ferguson right now please queue them instead. please pay attention to things that are more important. it’s not the time to talk about fandoms or jokes it’s time to talk about injustices.” [source]

“can yall maybe take some time away from reblogging fandom or humor crap and read up and reblog pakistan because the privilege you have of a safe bubble is not one shared by others” [source]

“If you’re uneducated, do not use that as an excuse. Do not say, “I’m not picking sides because I don’t know the full story,” because not picking a side is supporting Wilson. And by supporting him, you are on a racist side… Ignoring this situation will put you in deep shit, and it makes you racist. If you’re not racist, do not just say “but I’m not racist!!” just get educated and reblog anything you can.” [source]

“why are you so disappointing? I used to really like you. you’ve kept totally silent about peshawar, not acknowledging anything but fucking zutara or bellarke or whatever. there are other posts you’ve reblogged too that I wouldn’t expect you to- but those are another topic. I get that you’re 19 but maybe consider becoming a better fucking person?” [source]

“if you’re white, before you reblog one of those posts that’s like “just because i’m not blogging about ferguson doesn’t mean i don’t care!!!” take a few seconds to: consider the privilege you have that allows you not to pay attention if you don’t want to. consider those who do not have the privilege to focus on other things. ask yourself why you think it’s more important that people know you “care” than it is to spread information and

> show support. then consider that you are a fucking shitbaby.” [source]
>
> “For everyone reblogging Ferguson, Ayotzinapa, North Korea etc and not reblogging Peshawar, you should seriously be ashamed of yourselves.” [source]
>
> “This is going to be an unpopular opinion but I see stuff about ppl not wanting to reblog ferguson things and awareness around the world because they do not want negativity in their life plus it will cause them to have anxiety. They come to tumblr to escape n feel happy which think is a load of bull. There r literally ppl dying who live with the fear of going outside their homes to be shot and u cant post a fucking picture because it makes u a little upset?? I could give two fucks about internet shitlings.” [source]

You may also want to check the Tumblr tag “the trash is taking itself out”, in which hundreds of people make the same joke (“I think some people have stopped reading my blog because I’m talking too much about [the issue *du jour*]. I guess the trash is taking itself out now.”)

This is pretty impressive. It’s the first time outside of a chain letter that I have seen our memetic overlords throw off all pretense and just go around shouting “SPREAD ME OR YOU ARE GARBAGE AND EVERYONE WILL HATE YOU.”

But it only works because it’s tapped into the most delicious food source an ecology of epistemic parasites could possibly want – controversy,

I would like to be able to write about charity more often. Feminists would probably like to start supercharging the *true* rape accusations for a change. Protesters against police brutality would probably like to be able to focus on clear-cut cases that won't make white people support the police *even harder*. Even PETA would probably prefer being the good guys for once. But the odds aren't good. Not because the people involved are bad people who want to fail. Not even because the media-viewing public are stupid. Just because information ecologies are not your friend.

This blog tries to remember the Litany of Jai: "Almost no one is evil; almost everything is broken". We pretty much never wrestle with flesh and blood; it's powers and principalities all the way down.

VI.

A while ago I wrote a post called Meditations on Moloch where I pointed out that in any complex multi-person system, the system acts according to its own chaotic incentives that don't necessarily correspond to what *any* individual within the system wants. The classic example is the Prisoner's Dilemma, which usually ends at defect-defect even though *both* of the two prisoners involved prefer cooperate-cooperate. I compare this malignant discoordination to Ginsberg's portrayal of Moloch, the demon-spirit of capitalism gone wrong.

> I would support instating a National Conversation Topic Czar if that allowed us to get rid of celebrities.
>
> — Steven Kaas (@stevenkaas) August 26, 2010

Steven in his wisdom reminds us that there is no National Conversation Topic Czar. The rise of some topics to national prominence and the relegation of others to tiny print on the eighth page of the newspapers occurs by an emergent uncoordinated process. When we say "the media decided to cover Ferguson instead of Eric Garner", we reify and anthropomorphize an entity incapable of making goal-directed decisions.

A while back there was a minor scandal over JournoList, a private group where left-leaning journalists met and exchanged ideas. I think the conservative spin was “the secret conspiracy running the liberal media – revealed!” I wish they had been right. If there were a secret conspiracy running the liberal media, they could all decide they wanted to raise awareness of racist police brutality, pick the most clear-cut and sympathetic case, and make it non-stop news headlines for the next two months. Then everyone would agree it was indeed very brutal and racist, and something would get done.

But as it is, even if many journalists are interested in raising awareness of police brutality, given their total lack of coordination there’s not much they can do. An editor can publish a story on Eric Garner, but in the absence of a divisive hook, the only reason people will care about it is that caring about it is the right thing and helps people. But that’s “charity”, and we already know from my blog tags that charity doesn’t sell. A few people mumble something something deeply distressed, but neither black people nor white people get interested, in the “keep tuning to their local news channel to get the latest developments on the case” sense.

The idea of liberal strategists sitting down and choosing “a flagship case for the campaign against police brutality” is poppycock. Moloch – the abstracted spirit of discoordination and flailing response to incentives – will publicize whatever he feels like publicizing. And if they want viewers and ad money, the media will go along with him.

Which means that it’s not a coincidence that the worst possible flagship case for fighting police brutality and racism is the flagship case that we in fact got. It’s not a coincidence that the worst possible flagship cases for believing rape victims are the ones that end up going viral. It’s not a coincidence that the only time we ever hear about factory farming is when somebody’s doing something that makes us almost sympathetic to it. It’s not coincidence, it’s not even happenstance, it’s enemy action. Under Moloch, activists are irresistibly incentivized to dig their own graves. And the media is irresistibly incentivized to help them.

Lost is the ability to agree on simple things like fighting factory farming or rape. Lost is the ability to even talk about the things we all want. Ending corporate welfare. Ungerrymandering political districts. Defrocking pedophile priests. Stopping prison rape. Punishing government corruption and waste. Feeding starving children. Simplifying the tax code.

But also lost is our ability to treat each other with solidarity and respect.

Under Moloch, everyone is irresistibly incentivized to ignore the things that unite us in favor of forever picking at the things that divide us in exactly the way that is most likely to make them more divisive. Race relations are at historic lows not because white people and black people disagree on very much, but because the media absolutely worked its *tuchus* off to find the single issue that white people and black people disagreed over the most and ensure that it was the only issue anybody would talk about. Men’s rights activists and feminists hate each other not because there’s a huge divide in how people of different genders think, but because only the most extreme examples of either side will ever gain traction, and those only when they are framed as attacks on the other side.

People talk about the shift from old print-based journalism to the new world of social media and the sites adapted to serve it. These are fast, responsive, and only just beginning to discover the power of controversy. They are memetic evolution shot into hyperdrive, and the omega point is a well-tuned machine optimized to search the world for the most controversial and counterproductive issues, then make sure no one can talk about anything else. An engine that creates money by burning the few remaining shreds of cooperation, bipartisanship and social trust.

Imagine Moloch looking out over the expanse of the world, eagle-eyed for anything that can turn brother against brother and husband against wife. Finally he decides “YOU KNOW WHAT NOBODY HATES EACH OTHER ABOUT YET? BIRD-WATCHING. LET ME FIND SOME STORY THAT WILL MAKE PEOPLE HATE EACH OTHER OVER BIRD-WATCHING”. And the next day half the world’s newspaper headlines are “Has The Political Correctness Police Taken Over Bird-Watching?” and the other half are “Is Bird-Watching Racist?”. And then bird-watchers and non-bird-watchers and different sub-groups of bird-watchers hold vitriolic attacks on each other that feed back on each other in a vicious cycle for the next six months, and the whole thing ends in mutual death threats and another previously innocent activity turning into World War I style trench warfare.

(You think I’m exaggerating? Listen: “YOU KNOW WHAT NOBODY HATES EACH OTHER ABOUT YET? VIDEO GAMES.”)

Nobody Is Perfect, Everything Is Commensurable

I.

Recently spotted on Tumblr:

> "This is going to be an unpopular opinion but I see stuff about ppl not wanting to reblog ferguson things and awareness around the world because they do not want negativity in their life plus it will cause them to have anxiety. They come to tumblr to escape n feel happy which think is a load of bull. There r literally ppl dying who live with the fear of going outside their homes to be shot and u cant post a fucking picture because it makes u a little upset?"
>
> "Can yall maybe take some time away from reblogging fandom or humor crap and read up and reblog pakistan because the privilege you have of a safe bubble is not one shared by others?"

Ignore the questionable stylistic choices and there's an important point here worth considering. Something like "Yes, the feeling of constantly being outraged and mired in the latest controversy is unpleasant. And yes, it would be nice to get to avoid it and spend time with your family and look at kitten pics or something. But when the controversy is about people being murdered in cold blood, or living in fear, or something like that – then it's your duty as a decent human being to care. In the best case scenario you'll discharge that duty by organizing widespread protests or something – but the *absolute least* you can do is reblog a couple of slogans."

I think Cliff Pervocracy is trying to say something similar in this post. Key excerpt:

> When you’ve grown up with messages that you’re incompetent to make your own decisions, that you don’t deserve any of the things you have, and that you’ll never be good enough, the [conservative] fantasy of rugged individualism starts looking pretty damn good.
>
> Intellectually, I think my current political milieu of feminism/progressivism/social justice is more correct, far better for the world in general, and more helpful to me since I don’t actually live in a perfectly isolated cabin.
>
> But god, it’s uncomfortable. It’s intentionally uncomfortable—it’s all about getting angry at injustice and questioning the rightness of your own actions and being sad so many people still live such painful lives. Instead of looking at your cabin and declaring “I shall name it…CLIFFORDSON MANOR,” you need to look at your cabin and recognize that a long series of brutal injustices are responsible for the fact that you have a white-collar job that lets you buy a big useless house in the woods while the original owners of the land have been murdered or forced off it.
>
> And you’re never good enough. You can be good—certainly you get major points for charity and activism and fighting the good fight—but not good enough. No matter what you do, you’re still participating in plenty of corrupt systems that enforce oppression. Short of bringing about a total revolution of everything, your work will never be done, you’ll never be good enough.

> Once again, to be clear, I don't think this is wrong. I just think it's a bummer.
>
> I don't know of a solution to this. (Bummer again.) I don't think progressivism can ever compete with the cozy self-satisfaction of the cabin fantasy. I don't think it should. Change is necessary in the world, people don't change if they're totally happy and comfortable, therefore discomfort is necessary.

I'd like to make what I hope is a friendly amendment to Cliff's post. He thinks he's talking about progressivism versus conservativism, but he isn't. A conservative happy with his little cabin and occasional hunting excursions, and a progressive happy with her little SoHo flat and occasional poetry slams, are psychologically pretty similar. So are a liberal who abandons a cushy life to work as a community organizer in the inner city and fight poverty, and a conservative who abandons a cushy life to serve as an infantryman in Afghanistan to fight terrorism. The distinction Cliff is trying to get at here isn't left-right. It's activist versus passivist.

As part of a movement recently deemed postpolitical, I have to admit I fall more on the passivist side of the spectrum – at least this particular conception of it. I talk about politics when they interest me or when I enjoy doing so, and I feel an obligation not to actively make things worse. But I don't feel like I need to talk nonstop about whatever the designated Issue is until it distresses me and my readers both.

I've heard people give lots of reasons for not wanting to get into politics. For some, hearing about all the evils of the world makes them want to curl into a ball and cry for hours. Still others feel deep personal guilt about anything they hear – an almost psychotic belief that if people are being hurt anywhere in the world, it’s their fault for not preventing it. A few are chronically uncertain about which side to take and worried that anything they do will cause more harm than good. A couple have traumatic experiences that make them leery of affiliating with a particular side – did you know the prosecutor in the Ferguson case was the son of a police officer who was killed by a black suspect? And still others are perfectly innocent and just want to reblog kitten pictures.

Pervocracy admits this, and puts it better than I do:

> But god, it’s uncomfortable. It’s intentionally uncomfortable—it’s all about getting angry at injustice and questioning the rightness of your own actions and being sad so many people still live such painful lives. Instead of looking at your cabin and declaring “I shall name it…CLIFFORDSON MANOR,” you need to look at your cabin and recognize that a long series of brutal injustices are responsible for the fact that you have a white-collar job that lets you buy a big useless house in the woods while the original owners of the land have been murdered or forced off it. And you’re never good enough. You can be good—certainly you get major points for charity and activism and fighting the good fight—but not good enough. No matter what you do, you’re still participating in plenty of corrupt systems that enforce oppression. Short of bringing about a total revolution of everything, your work will never be done, you’ll never be good enough.

That seems about right. Pervocracy ends up with discomfort, and I’m in about the same place. But other, less stable people end up with self-loathing. Still other people go further than that, into Calvinist-style “perhaps I am a despicable worm unworthy of existence”. moteinthedark’s reply to Pervocracy gives me the impression that she struggles with this sometime. For these people, abstaining from politics is the only coping tool they have.

But the counterargument is *still* that you’ve got a lot of chutzpah playing that card when people in Peshawar or Ferguson or Iraq don’t have access to this coping tool. You can’t just bring in a doctor’s note and say “As per my psychiatrist, I have a mental health issue and am excused from experiencing concern for the less fortunate.”

One option is to deny the obligation. I am super sympathetic to this one. The *marginal* cost of my existence on the poor and suffering of the world is zero. In fact, it’s probably positive. My economic activity consists mostly of treating patients, buying products, and paying taxes. The first treats the poor’s illnesses, the second creates jobs, and the third pays for government assistance programs. Exactly what am I supposed to be apologizing for here? I may benefit from the genocide of the Indians in that I live on land that was formerly Indian-occupied. But I also benefit from the asteroid that killed the dinosaurs, in that I live on land that was formerly dinosaur-occupied. I don’t feel like I’m complicit in the asteroid strike; why should I feel complicit in the genocide?

I have no objection to people who say this. The problem with it isn’t philosophical, it’s emotional. For *most people* it won’t be enough. The old saying goes “you can’t reason yourself out of something you didn’t reason yourself into to begin with”, and the idea that secure and prosperous people need to “give something back” is a lot older than accusations of “being complicit in structures of oppression”. It’s probably older than the Bible. People feel a deep-seated need to show that they understand how lucky they are and help those less fortunate than themselves.

So what do we do with the argument that we are morally obligated to be political activists, possibly by reblogging everything about Ferguson that crosses our news feed?

II.

We ask: *why the heck are we privileging that particular subsection of the category “improving the world”?*

Pervocracy says that “short of bringing about a total revolution of everything, your work will never be done, you’ll never be good enough.” But he is overly optimistic. Has your total revolution of everything eliminated ischaemic heart disease? Cured malaria? Kept elderly people out of nursing homes? No? Then you haven’t discharged your infinite debt yet!

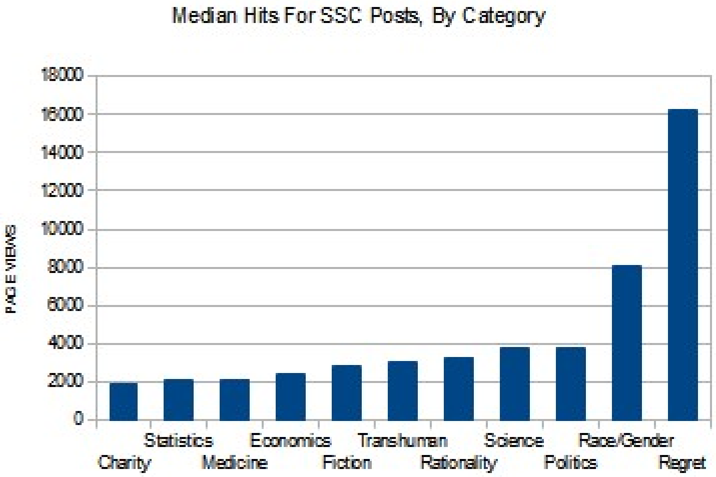
Being a perfect person doesn’t just mean participating in every hashtag campaign you hear about. It means spending all your time at soup kitchens, becoming vegan, donating everything you have to charity, calling your grandmother up every week, and marrying Third World refugees who need visas rather than your one true love.

And not all of these things are equally important.

Five million people participated in the #BlackLivesMatter Twitter campaign. Suppose that solely as a result of this campaign, no currently-serving police officer ever harms an unarmed black person ever again. That’s 100 lives saved per year times let’s say twenty years left in the average officer’s career, for a total of 2000 lives saved, or 1/2500th of a life saved per campaign participant. By coincidence, 1/2500th of a life saved happens to be what you get when you donate $1 to the Against Malaria Foundation. The round-trip bus fare people used to make it to their #BlackLivesMatter protests could have saved ten times as many black lives as the protests themselves, *even given completely ridiculous overestimates of the protests' efficacy*.

The moral of the story is that if you feel an obligation to give back to the world, participating in activist politics is one of the worst possible ways to do it. Giving even a tiny amount of money to charity is hundreds or even thousands of times more effective than almost any political action you can take. Even if you're absolutely convinced a certain political issue is the most important thing in the world, you'll effect more change by donating money to nonprofits lobbying about it than you will be reblogging anything.

There is *no reason* that politics would even *come to the attention* of an unbiased person trying to "break out of their bubble of privilege" or "help people who are afraid of going outside of their house". Anybody saying that people who want to do good need to spread their political cause is about as credible as a televangelist saying that people who want to do good need to give them money to buy a new headquarters. It's possible that televangelists having beautiful headquarters might be *slightly* better than them having hideous headquarters, but it's not the first thing a reasonable person trying to improve the world would think of.



Average number of hits for posts on this blog, by topic

Nobody cares about charity. Everybody cares about politics, especially race and gender. Just as televangelists who are obsessed with moving to a sweeter pad may come to think that donating to their building fund is the one true test of a decent human being, so our universal obsession with politics, race, and gender incites people to make convincing arguments that taking and spreading the right position on those issues is the one true test of a decent human being.

So now we have an angle of attack against our original question. “Am I a bad person for not caring more about politics?” Well, every other way of doing good, especially charity, is more important than politics. So this question is strictly superseded by “Am I a bad person for not engaging in every other way of doing good, especially charity?” And then once we answer that, we can ask “Also, however much sin I have for not engaging in charity, should we add another mass of sin, about 1% as large, for my additional failure to engage in politics?”

And Cliff Pervocracy’s concern of “Even if I do a lot of politics, am I still a bad person for not doing *all* the politics?”

is superseded by “Even if I give a lot of charity, am I a bad person for not doing *all* the charity? And then a bad person in an additional way, about 1% as large, for not doing all the politics as well?”

There’s no good answer to this question. If you want to feel anxiety and self-loathing for not giving 100% of your income, minus living expenses, to charity, then no one can stop you.

I, on the other hand, would prefer to call that “not being perfect”. I would prefer to say that if you feel like you will live in anxiety and self-loathing until you have given a certain amount of money to charity, you should make that certain amount ten percent.

Why ten percent?

It’s ten percent because that’s the standard decreed by Giving What We Can and the effective altruist community. Why should we believe their standard? I think we should believe it because if we reject it in favor of “No, you are a bad person unless you give *all* of it,” then everyone will just sit around feeling very guilty and doing nothing. But if we very clearly say “You have discharged your moral duty if you give ten percent or more,” then many people will give ten percent or more. The most important thing is having a Schelling point, and ten percent is nice, round, divinely ordained, and – crucially – the Schelling point upon which we have already settled. It is an *active* Schelling point. If you give ten percent, you can have your name on a nice list and get access to a secret forum on the Giving What We Can site which is actually pretty boring.

It’s ten percent because definitions were made for Man, not Man for definitions, and if we define “good person” in a way such that everyone is sitting around miserable because they can’t reach an unobtainable standard, we are stupid definition-makers. If we are smart definition-makers, we will define it in whichever way which makes it the most effective tool to convince people to give at least that much.

Finally, it’s ten percent because if you believe in something like universalizability as a foundation for morality, a world in which everybody gives ten percent of their income to charity is a world where about seven trillion dollars go to charity a year. Solving global poverty forever is estimated to cost about $100 billion a year for the couple-decade length of the project. That’s *about two percent* of the money that would suddenly become available. If charity got seven trillion dollars a year, *the first year* would give us enough to solve global poverty, eliminate all treatable diseases, fund research into the untreatable ones for approximately the next forever, educate anybody who needs educating, feed anybody who needs feeding, fund an unparalleled renaissance in the arts, permamently save every rainforest in the world, and have enough left over to launch five or six different manned missions to Mars. That would be the *first year*. Goodness only knows what would happen in Year 2.

(by contrast, if everybody in the world retweeted the latest hashtag campaign, Twitter would break.)

Charity is in some sense the perfect unincentivized action. If you think the most important thing to do is to cure malaria, then a charitable donation is deliberately throwing the power of your brain and muscle behind the cause of curing malaria. If, as I’ve argued, the reason we can’t solve world poverty and disease and so on is the capture of our financial resources by the undirected dance of incentives, then what better way to fight back than by saying “Thanks but no thanks, I’m taking this abstract representation of my resources and using it *exactly* how I think it should most be used"?

If you give 10% per year, you have done your part in making that world a reality. You can honestly say "Well, it's not my fault that everyone *else* is still dragging their feet."

III.

Once the level is fixed at ten percent, we get a better idea how to answer the original question: "If I want to be a good person who gives back to the community, but I am triggered by politics, what do I do?" You do good in a way that doesn't trigger you. Another good thing about having less than 100% obligation is that it gives you the opportunity to budget and trade-off. If you make $30,000 and you accept 10% as a good standard you want to live up to, you can either donate $3000 to charity, or participate in political protests until your number of lives or dollars or DALYs saved is equivalent to that.

Nobody is perfect. This gives us license not to be perfect either. Instead of aiming for an impossible goal, falling short, and not doing anything at all, we set an arbitrary but achievable goal designed to encourage the most people to do as much as possible. That goal is ten percent.

Everything is commensurable. This gives us license to determine exactly how we fulfill that ten percent goal. Some people are triggered and terrified by politics. Other people are too sick to volunteer. Still others are poor and cannot give very much money. But money is a constant reminder that everything goes into the same pot, and that you can fulfill obligations in multiple equivalent ways. Some people will not be able to give ten percent of their income without excessive misery, but I bet thinking about their contribution in terms of a fungible good will help them decide how much volunteering or activism they need to reach the equivalent.

Cliff Pervocracy says “Your work will never be done, you’ll never be good enough.” This seems like a recipe for – at best – undirected misery, stewing in self-loathing, and total defenselessness against the first parasitic meme to come along and tell them to engage in the latest conflict or else they’re trash. At worst, it autocatalyzes an opposition of egoists who laugh at the idea of helping others.

On the other hand, Jesus says “Take my yoke upon you…and you will find rest for your souls. For my yoke is easy and my burden is light.” This seems like a recipe for getting people to say “Okay, I’ll take your yoke upon me! Thanks for the offer!”

Persian poet Omar Khayyam, considering the conflict between the strict laws of Islam and his own desire to enjoy life, settles upon the following rule:

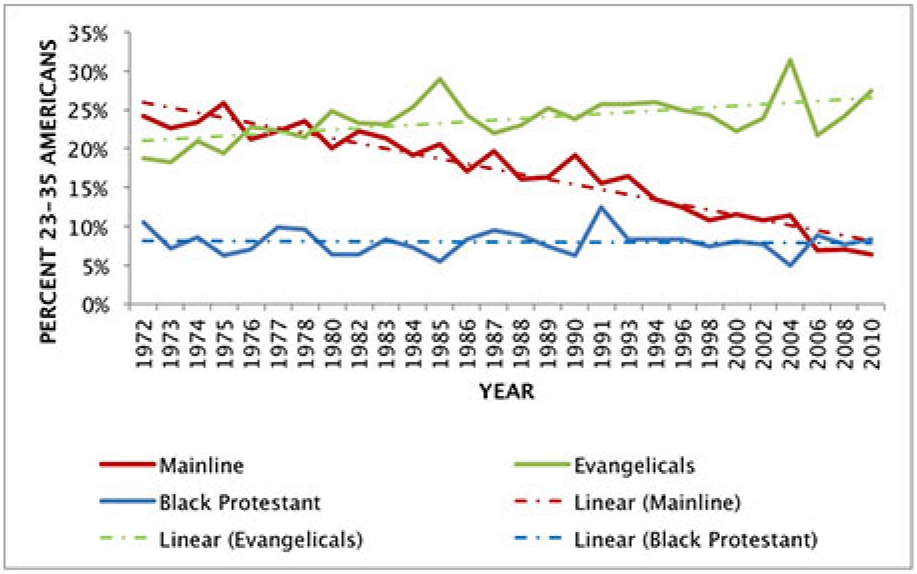
> Heed not the Sunna, nor the law divine;
> If to the poor their portion you assign,
> And never injure one, nor yet abuse,
> I guarantee you heaven, as well as wine!

I’m not saying that donating 10% of your money to charity makes you a great person who is therefore freed of every other moral obligation. I’m not saying that anyone who chooses not to do it is therefore a bad person. I’m just saying that if you feel a need to discharge some feeling of a moral demand upon you to help others, and you want to do it intelligently, it beats most of the alternatives.

This month is the membership drive for Giving What We Can, the organization of people who have promised to give 10% of their earnings to charity. I am a member. Ozy is an aspiring member who plans to join once they are making a salary. Many of the commenters here are members – I recognize for example Taymon Beal’s name on their list. Some well-known moral philosophers like Peter Singer and Derek Parfit are members. Seven hundred other people are also members.

I would recommend giving them a look.

There Are Rules Here



[source: sowhatfaith.com]

Patheos' Science On Religion points out that liberal Protestantism is dying even as more conservative Protestant movements thrive. This seems counterintuitive in the context of society as a whole becoming less religious and conservative. So what's going on?

> In the early 1990s, a political economist named Laurence Iannaccone claimed that seemingly arbitrary demands and restrictions, like going without electricity (the Amish) or abstaining from caffeine (Mormons), can actually make a group stronger. He was trying to explain religious affiliation from a rational-choice perspective: in a marketplace of religious options, why would some people choose religions that make serious demands on their members, when more easygoing, low-investment churches were – literally – right around the corner? Weren't the warmer and fuzzier churches destined to win out in fair, free-market competition?

According to Iannaccone, no. He claimed that churches that demanded real sacrifice of their members were automatically stronger, since they had built-in tools to eliminate people with weaker commitments. Think about it: if your church says that you have to tithe 10% of your income, arrive on time each Sunday without fail, and agree to believe seemingly crazy things, you’re only going to stick around if you’re really sure you want to. Those who aren’t totally committed will sneak out the back door before the collection plate even gets passed around.

And when a community only retains the most committed followers, it has a much stronger core than a community with laxer membership requirements. Members receive more valuable benefits, in the form of social support and community, than members of other communities, because the social fabric is composed of people who have demonstrated that they’re totally committed to being there. This muscular social fabric, in turn, attracts more members, who are drawn to the benefits of a strong community – leading to growth for groups with strict membership requirements.

The evolutionary anthropologist William Irons calls demanding rituals and onerous requirements “hard-to-fake symbols of commitment.” If you’re not really committed to the group, you won’t be very enthusiastic about fasting, abstaining from coffee, tithing ten percent, or following through on any of the other many costly requirements that conservative religious communities demand. The result? Only the most committed believers stick around, benefiting from one another’s in-group-oriented generosity, social support, and community.

> Since then, Sosis has also demonstrated that religious Israeli kibbutz members are more generous in resource-sharing games than both secular, urban Israelis and secular kibbutzim. He argues that this is, in part, because demanding rituals – such as having to pray three times a day and study Torah many hours a week – serve as a signal of investment in the kibbutz community. The more rituals you participate in, the more invested you feel – and the more willing you are to sacrifice for your fellows.
>
> But if your community doesn't have any of these costly requirements, then you don't feel that you have to be really committed in order to belong. The whole group ends up with a weakened, and less committed, membership. Liberal Protestant churches, which have famously lax requirements about praxis, belief, and personal investment, therefore often end up having a lot of half-committed believers in their pews. The parishioners sitting next to them can sense that the social fabric of their church isn't particularly robust, which deters them from investing further in the collective. It's a feedback loop. The whole community becomes weaker…and weaker…and weaker.

Even though I've quoted like half the blog post, it's worth looking at just to see the empirical and statistical arguments for their hypothesis.

Not that any of this should come as a surprise. This is the same principle of maintaining separation between in-group and out-group members which has worked so well for so many eons. But making the in-group follow specific rules to prove their dedication does seem particularly effective.

I’ve been thinking about this in the context of atheist religion-substitutes. I went to the Secular Solstice last weekend, and it was held in the New York Society For Ethical Culture building. As usual I avoided social interaction by beelining to the nearest reading material, and in this case that was a plaque detailing the group’s history. The Society for Ethical Culture was founded in 1877 by an ex-rabbi (of *course* it was an ex-rabbi) and looks pretty much exactly like every atheist religious substitute today. That got me a little depressed. Atheism has been trying the same things for the past one hundred fifty years and, I would argue, largely failing for the past one hundred fifty years. Religion substitutes are *hard*.

The biggest atheist religion-substitute I know of is Sunday Assembly. I recently came across their “Ten Commandments”:

> 1. Is a 100 per cent celebration of life. We are born from nothing and go to nothing. Let’s enjoy it together.
>
> 2. Has no doctrine. We have no set texts so we can make use of wisdom from all sources.
>
> 3. Has no deity. We don’t do supernatural but we also won’t tell you you’re wrong if you do.
>
> 4. Is radically inclusive. Everyone is welcome, regardless of their beliefs – this is a place of love that is open and accepting.
>
> 5. Is free to attend, not-for-profit and volunteer-run. We ask for donations to cover our costs and support our community work.
>
> 6. Has a community mission. Through our Action Heroes (you!) we will be a force for good.

> 7. Is independent. We do not accept sponsorship or promote outside organisations.
>
> 8. Is here to stay. With your involvement, the Sunday Assembly will make the world a better place.
>
> 9. We won’t tell you how to live, but will try to help you do it as well as you can.
>
> 10. And remember point 1…The Sunday Assembly is a celebration of the one life we know we have.

But it’s tough for me to picture these on big stone tablets. And yeah, I know the reason we don’t have the original tablets is that when Sunday Assembly Moses came down from Mt. Sinai he saw the Sunday Assembly people only celebrating life 95 percent, and waxed wroth, and broke the tablets, and then ordered the Levites to slaughter all the men, women, and children who had participated in this abomination. And then…

…okay, that’s probably not the reason they’re not on tablets. But that’s just the thing. It’s impossible to imagine these commandments inspiring strong emotions in anybody. It’s impossible to imagine people sinning against them in a meaningful way. Most of them aren’t even commandments. They’re more like promises not to command. If you absolutely must compare this pablum to a list of ten points, the proper analogy is less to the Ten Commandments than to the Bill of Rights.

(“God shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof…”)

Atheist religion-substitutes seem unconcerned about or actively hostile to placing rules upon their members. I mean, there are a lot of things that are like “You must be tolerant”. But in practice everybody thinks “intolerant” means “more intolerant than I am, since I am only intolerant of things that are actually bad,” so no one changes their behavior. People say that we have advanced by replacing useless rules like “don’t eat pork” with useful rules like “be tolerant”, but rules against eating pork resulted in decreased pork consumption whereas it’s not clear that rules like “be tolerant” result in *anything*.

The only secular-ish group I have ever seen which is truly virtuous in this respect is, once again, Giving What We Can. They demand that members give ten percent of their income to charity. To join you must request and sign a paper copy of a form pledging to do this. Every year, the organization asks you to confirm that you are still complying. I don’t know what happens if you aren’t, but I assume it’s too horrible to contemplate. Maybe Peter Singer breaks into your house and kills you for the greater good.

But the point is, here’s an organization that has a very specific rule and demands you follow it. And even though their pledge form looked kind of like a tax return, signing that form was more of a sanctifying and humbling experience than any of the religion-substitutes that try to intentionally generate sanctification. Not because I was at some chapel where someone gave a rousing sermon overusing the word “community”, but because I was binding myself, voluntarily submitting to a higher moral authority.

Someone on my blog a while back used the word “nomic” to refer to a subculture based on following a rule set, sort of like an opt-in religion without beliefs or supernatural elements. I looked to see if it was a real thing but couldn’t find any references other than the card game. But I find the idea interesting. If it contains mechanisms for treating subculture members differently than non-members, it seems like an optional add-on module to government, and a strong candidate for the sort of thing that could develop into a healthy Archipelago.

Book Review: What's Wrong With The World

G. K. Chesterton's 1910 collection *What's Wrong With The World* surprisingly does *not* open with "this is going to take more than one book."

In fact, he is quite to-the-point about exactly what he thinks the problem is:

> Now, to reiterate my title, this is what is wrong. This is the huge modern heresy of altering the human soul to fit its conditions, instead of altering human conditions to fit the human soul…it is the huge heresy of Precedent. It is the view that because we have got into a mess we must grow messier to suit it; that because we have taken a wrong turn some time ago we must go forward and not backwards; that because we have lost our way we must lose our map also.

Examples are clearly needed, but before we continue, a digression.

Chesterton is a brilliant writer and a genius in understanding the human soul. Sometimes these are good things – in Chesterton's case, it's a big part of what makes his fiction so amazing. Other times they aren't. There's a famous failure mode where some brilliant scientist who is a genius at understanding particle physics sees a social problem, pulls a few equations out of his tool kit, and declares the issue solved.

In the same way, I worry that Chesterton pulls a few emotions and brilliant turns of phrase out of *his* tool kit and thinks *he's* solved everything. And the difference between brilliant physicists and brilliant students of human experience is that physicists are less likely to convince anyone else.

The trademark style of *What Is Wrong With The World* is to take some common-sense proclamation, like “feminism is about fighting for women” and come up with some incredibly clever reason why *exactly* the opposite is true:

> By the beginning of the twentieth century, within the last few years, the woman has in public surrendered to the man. She has seriously and officially owned that the man has been right all along; that the public house (or Parliament) is really more important than the private house; that politics are not (as woman had always maintained) an excuse for pots of beer, but are a sacred solemnity to which new female worshipers may kneel; that the talkative patriots in the tavern are not only admirable but enviable; that talk is not a waste of time, and therefore (as a consequence, surely) that taverns are not a waste of money. All we men had grown used to our wives and mothers, and grandmothers, and great aunts all pouring a chorus of contempt upon our hobbies of sport, drink and party politics. And now comes Miss Pankhurst with tears in her eyes, owning that all the women were wrong and all the men were right; humbly imploring to be admitted into so much as an outer court, from which she may catch a glimpse of those masculine merits which her erring sisters had so thoughtlessly scorned.
>
> Now this development naturally perturbs and even paralyzes us…We knew quite well that nothing is necessary to the country except that the men should be men and the women women. We knew this; we thought the women knew it even more clearly; and we thought

> the women would say it. Suddenly, without warning, the women have begun to say all the nonsense that we ourselves hardly believed when we said it. The solemnity of politics; the necessity of votes; the necessity of Huggins; the necessity of Buggins; all these flow in a pellucid stream from the lips of all the suffragette speakers. I suppose in every fight, however old, one has a vague aspiration to conquer; but we never wanted to conquer women so completely as this.

Do you follow? Men have from time immemorial been pushing the importance of male pursuits like politics and public life; women have been equally pushing the importance of family, virtue, and the private household. Feminists are then a group of women who have given up, admitted men have always been right about everything and all female pursuits are a waste of time; now women are desperately pleading that men allow them to join in their superior ways.

Therefore, Chesterton opposes feminism not because he is against women being equal to men, but precisely *because* he wants to keep women equal to men. The entire book is like this – paragraph after paragraph of verbal judo in which you end up opposing conservativism because you want things to stay the same, or supporting rebellion to protect the integrity of the state, or whatever other crazy inverted idea Chesterton has turned his brilliant but twisted mind to.

> “ATTENTION HUMANS – IT IS NOT BECAUSE YOU ARE INFERIOR THAT YOU WILL BE ERADICATED – BUT PRECISELY BECAUSE YOU ARE SUPERIOR” -GK Chestertron
>
> — Steven Kaas (@stevenkaas) February 4, 2012

At its best, this is a form of narrative reversal that helps break through our preconceptions and see things in a different light. At its worst, I worry Chesterton has actually lost, through atrophy, the ability to think in a straight line. Like, there must be at least one thing which is approximately the way it appears, and I'm not convinced Chesterton will be able to notice it. Ask him whether we should drown puppies, and he will come up with an extremely convincing argument that we should drown puppies precisely *because* we abhor cruelty to animals.

> "Arf arf arf! Not because arf arf! But exactly because arf NOT arf!" GK Chesterton's dog
>
> — Steven Kaas (@stevenkaas) December 8, 2011

II.

To very briefly sum up the main argument of *What's Wrong*:

1910s Britain has a lot of social problems, many of them caused by the Industrial Revolution steamrolling over human values. Conservatives demand that we keep things exactly as they are, social problems and all. Progressives correctly point out that the new industrial economy conflicts with human values, but "solve" the problem by demanding we destroy whatever human values are left in people to make them better cogs in the industrial machine. For example, women traditionally value having a home of their own and getting to spend time raising children, but this is inconsistent with women working full time in factories – therefore, socialists and feminists demand we put people in communal housing, have children raised communally, and promote women taking on the male gender role – so that all barriers to women doing factory work full-time are removed. Possibly the Conservatives and Progressives are secretly in cahoots – the Conservatives protect the interests of the upper class directly, and the Progressives remake the lower class into a form compatible with serving the system that the upper class run, all upon “humanitarian” grounds.

Chesterton, on the other hand, believes that we should promote human values against the Industrial Revolution and the upper classes who intend to benefit from it. The values he wants seem to be of the “everyone lives in their own nice cottage with a nice garden and a nice picket fence and the men have stable fulfilling jobs and the women stay home and raise children” type. This will require undoing a lot of “progress”, and admitting that the medievals were better than us in a lot of ways, but it’s better than continuing to slide toward doom. It will also involve redistributing property so that everybody has enough money to make this vision a reality. Chesterton has a perfectly marvelous solution for how to do this, which unfortunately this book is too small to contain. But he *did* warn us by calling the book *What Is Wrong With The World* and not *What Is Wrong With The World And How I Propose To Fix It*, so it’s okay.

III.

I wasn’t too sure what to make of all this. The political landscape of the 1910s when the book was published is recognizable, but only barely. In particular, although I can’t tell if Chesterton’s claims fairly described the Left of his own time, they don’t seem to describe the Left of ours. Chesterton’s Left was obsessed with industrial order and optimization, hoping to replace a society ruled by traditions with one ruled by nearly fanatical efficiency and conformity. The modern Left seems to have switched tactics entirely, and insofar as it can be accused of falling too far to one side of the chaos-order dichotomy I think both its friends and enemies would admit it is squarely allied with Chaos, and with a fertility of difference and distinction that borders on the cancerous. As the book was at least in part a polemic against a position that no longer seems to exist, one that barely even seems to have any cladistic descendants, I’m not really sure what to make of it.

There were a lot of digressions into individual issues, some convincing, others less so. But there were two parts I consider central to the whole idea, and which I want to go into at more length:

> My point is that the world did not tire of the church’s ideal, but of its reality. Monasteries were impugned not for the chastity of monks, but for the unchastity of monks. Christianity was unpopular not because of the humility, but of the arrogance of Christians. Certainly, if the church failed it was largely through the churchmen. But at the same time hostile elements had certainly begun to end it long before it could have done its work. In the nature of things it needed a common scheme of life and thought in Europe. Yet the mediaeval system began to be broken to pieces intellectually, long before it showed the slightest hint of falling to pieces morally. The huge early heresies, like the Albigenses, had not the faintest excuse in moral superiority. And it is actually true that the Reformation began to tear Europe apart before the Catholic Church had had time to pull it together. The Prussians, for instance, were not converted to Christianity at all until quite close to the Reformation. The poor creatures hardly had time to become Catholics before they were told to become Protestants. This explains a great deal of their subsequent conduct. But I have only taken this as the first and most evident case of the general truth: that the great ideals of the past failed not by being outlived (which must mean over-lived), but by not being lived

> enough. Mankind has not passed through the Middle Ages. Rather mankind has retreated from the Middle Ages in reaction and rout. The Christian ideal has not been tried and found wanting. It has been found difficult; and left untried.

This is one of those times when writing in 1910 makes things too easy. If Chesterton had been writing in 1990, he might have thought to ask – has Communism been tried and found wanting? Or has it been found difficult and left untried?

And with our modern hindsight, he might decide there's not so big a difference as he thinks. Like, Chesterton is defining "Everyone was super-obsessed with Christianity all the time for hundreds of years" as "Christianity was left untried". The definition only works if by "try Christianity" he means "everyone lives exactly according to the Christian ideal." But "make everyone live exactly according to the Christian ideal" is not a primitive action.

At the very least, the medievals *tried to try* Christianity. They reserved political power for Christians, gave immense wealth and clout to the clergy, gave religion a monopoly on education, required everyone to go to church, and persecuted atheism and heresy. If, as per Chesterton's definition, this didn't result in people *trying* Christianity, then that means that *trying to try* Christianity has failed. If an idea is impossible to implement, that is a strike against the idea. Unless Chesterton has a better idea for how to implement Christianity than the way the medievals tried, his argument is wrong and it is perfectly legitimate to say that the failure of the Christian project during the Middle Ages doesn't bode well for it today.

And I understand that making a medieval-level effort to retry Christianity isn’t exactly in the Overton Window, but this ties into the other passage of Chesterton’s I want to talk about:

> A little while ago certain doctors and other persons permitted by modern law to dictate to their shabbier fellow-citizens, sent out an order that all little girls should have their hair cut short. I mean, of course, all little girls whose parents were poor. Many very unhealthy habits are common among rich little girls, but it will be long before any doctors interfere forcibly with them. Now, the case for this particular interference was this, that the poor are pressed down from above into such stinking and suffocating underworlds of squalor, that poor people must not be allowed to have hair, because in their case it must mean lice in the hair. Therefore, the doctors propose to abolish the hair. It never seems to have occurred to them to abolish the lice.
>
> Yet it could be done. As is common in most modern discussions the unmentionable thing is the pivot of the whole discussion. It is obvious to any Christian man (that is, to any man with a free soul) that any coercion applied to a cabman’s daughter ought, if possible, to be applied to a Cabinet Minister’s daughter. I will not ask why the doctors do not, as a matter of fact apply their rule to a Cabinet Minister’s daughter. I will not ask, because I know. They do not because they dare not.
>
> But what is the excuse they would urge, what is the plausible argument they would use, for thus cutting and clipping poor children and not rich? Their argument would be that the disease is more likely to be in the hair of poor people than of rich. And why? Because the poor

> children are forced (against all the instincts of the highly domestic working classes) to crowd together in close rooms under a wildly inefficient system of public instruction; and because in one out of the forty children there may be offense. And why? Because the poor man is so ground down by the great rents of the great ground landlords that his wife often has to work as well as he. Therefore she has no time to look after the children, therefore one in forty of them is dirty. Because the workingman has these two persons on top of him, the landlord sitting (literally) on his stomach, and the schoolmaster sitting (literally) on his head, the workingman must allow his little girl's hair, first to be neglected from poverty, next to be poisoned by promiscuity, and, lastly, to be abolished by hygiene. He, perhaps, was proud of his little girl's hair. But he does not count.

First of all, either the misuse of the word "literally" is more ancient than I had supposed, or else the customs of the 1910s class system are indeed weird and wonderful. But putting that aside…

I am clearly one of the people whom Chesterton is talking to. I have yet to prescribe any particular remedy for lice, but that is because lice aren't so much of a problem today, or if they are they don't make it to a psychiatrist. But I have given antidepressants.

And this seems a lot like Chesterton's lice. Once again, people are crowded together into squalor, oppressed by landlords and schoolmasters, and so some of them – usually the poor – become depressed. Antidepressants are moderately effective against this problem, although they have physical side effects in some people and are considered embarrassing by many more. To take an antidepressant is a sacrifice in much the same way that cutting your hair is a sacrifice. Its only possible justification is that it does treat depression, just as haircuts do treat lice.

So I can imagine Chesterton coming at me, fire in his eyes, demanding "Why are you solving this problem by giving pills?! You should be solving this by improving society so that poor people don't end up depressed!"

And all I could answer is "If I wrote a prescription for 'improve society', I'm not sure the pharmacist would know how to fill it."

Like, I want society to be improved. And I try. I give money to charity. I vote. I try to argue persuasively for good political positions. But when I'm done doing all of that, it turns out there are still depressed people left over. At that point, if I take the "high road" and say "no, I'm not going to give you an antidepressant, society should be giving you a job and better housing and more opportunities for social interaction," then it doesn't shame society into doing that. It just means that my patient will get more and more depressed and maybe commit suicide.

And I can imagine a 1910s doctor in my same position, who's done what he can to help the poor, but in the meantime, he's treating this girl, and if he doesn't tell her to get a haircut it's not going to make her life better, it's just going to mean she gets lice. Which are creepy and gross and probably not very fun to have at all.

So Chesterton says "Whatever, I wasn't talking to you, I was talking to *society*!"

But society isn't listening! Chesterton may be the most brilliant essayist of all time, here's this classic book he wrote which is still justly famous over a hundred years later, and people didn’t even listen to *him*. They’re sure not going to listen to *me*.

Even if the intellectual battle were to be completely won – even if nobody objects to helping the poor on the grounds of “there’s not enough money”, or the grounds of “it would disincentivize them to work”, or the grounds of opportunity costs, or whatever – actually getting society to *do something* would still be just as difficult as getting it to do all the other things we all agree it should, like end corporate welfare or have less police brutality.

This is why I am so worried that Chesterton calls Christianity “left untried”. He seems to hear “This idea would work if we could get people to do it, which we obviously can’t” and answer “So you’re saying it would work! Great! Everyone who wants anything other than this idea is wrong and should stop!”

The problems Chesterton writes about are clearly the aftermath of the Industrial Revolution. He doesn’t say so in so many words, but I think he considers “rolling back the Industrial Revolution” to be a perfectly reasonable solution:

> There is one metaphor of which the moderns are very fond; they are always saying, “You can’t put the clock back.” The simple and obvious answer is “You can.” A clock, being a piece of human construction, can be restored by the human finger to any figure or hour. In the same way society, being a piece of human construction, can be reconstructed upon any plan that has ever existed.
>
> There is another proverb, “As you have made your bed, so you must lie on it”; which again is simply a lie. If I have made my bed uncomfortable, please God I will

> make it again. We could restore the Heptarchy or the stage coaches if we chose. It might take some time to do, and it might be very inadvisable to do it, but certainly it is not impossible as bringing back last Friday is impossible. This is, as I say, the first freedom that I claim: the freedom to restore. I claim a right to propose as a solution the old patriarchal system of a Highland clan, if that should seem to eliminate the largest number of evils. I claim the right to propose the complete independence of the small Greek or Italian towns, a sovereign city of Brixton or Brompton, if that seems the best way out of our troubles.

If every single person in the world wanted to roll back the Industrial Revolution, and we all had perfect coordination power and followed absolutely every command of a task force appointed for that problem – then yes, we could do it. But if even one person objects, then that person is going to start manufacturing things cheaper than the rest of us. If one country objects, that country is going to manufacture tanks and cannons and stealth bombers and the rest of us are going to have knights on horseback with which to fight them off.

The arrow of time is not nearly as bidirectional as Chesterton seems to think. The reason that the medievals could maintain deindustrialization without any effort is that nobody knew how to be industrialized even if they wanted to, and nobody wanted to because nobody knew there was such a thing as “industrialization” to want. Any modern attempt to recreate medieval society is doomed to be different than medieval society, because it will involve either industrialization, or an extremely concerted and tyrannical worldwide effort to suppress industrialization – both of which the medievals lacked.

The socialists, feminists, and other groups whom Chesterton dislikes seem to understand this. They say things like “Well, we’re never getting rid of industrialization and the pressure it puts on people to live in dingy slums. So let’s at least institute communal housing which will be a little more liveable and affordable than the other kind.” It might work or it might not, but it’s the sort of thing you can imagine coexisting with modern society. Chesterton’s argument is “No, let’s roll everything back until everyone can have a nice cottage in the Cotswolds.” It’s a very desirable solution, but it’s addressed to a hypothetical universal monarch who has the power to implement solutions against incentive gradients with no defectors ever. This is a book whose target audience is nowhere to be found, and those physicians who were cutting children’s hair to protect them from lice could justly have replied “Dammit, Gilbert, I’m a doctor, not a God-Emperor!”

There’s obviously a line of intellectual descent from Chesterton to the neoreactionaries of today. And yet the latter group seem less naive in an important way. They know that their proposals are impossible within the sphere of normal politics, so they’ve mostly stopped the “arguing for their proposals” part of the plan and started crafting alternative ways to have human societies in which apparently impossible coordination problems can get solved. Meanwhile, poor Chesterton never seems to get beyond wanting his utopian vision to become a party platform.

And I’m not sure he can. I get a weird vibe out of everything he’s written, like his brain lacks the gear that goes back and questions whether what it’s saying actually makes sense. I mean, no one does a *great* job having this gear, but in him it just seems completely missing, as if he is so feverishly brilliant that to try to analyze the truth-value of what he’s written or look for another side would break the spell.

But I can’t complain. It gives me paragraphs like this:

> Now the whole parable and purpose of these last pages, and indeed of all these pages, is this: to assert that we must instantly begin all over again, and begin at the other end. I begin with a little girl’s hair. That I know is a good thing at any rate. Whatever else is evil, the pride of a good mother in the beauty of her daughter is good. It is one of those adamantine tendernesses which are the touchstones of every age and race. If other things are against it, other things must go down. If landlords and laws and sciences are against it, landlords and laws and sciences must go down. With the red hair of one she-urchin in the gutter I will set fire to all modern civilization. Because a girl should have long hair, she should have clean hair; because she should have clean hair, she should not have an unclean home: because she should not have an unclean home, she should have a free and leisured mother; because she should have a free mother, she should not have an usurious landlord; because there should not be an usurious landlord, there should be a redistribution of property; because there should be a redistribution of property, there shall be a revolution. That little urchin with the gold-red hair, whom I have just watched toddling past my house, she shall not be lopped and lamed and altered; her hair shall not be cut short like a convict’s; no, all the kingdoms of the earth shall be hacked about and mutilated to suit her. She is the human and sacred image; all around her the social fabric shall sway and split and fall; the pillars of society shall be shaken, and the roofs of ages come rushing down, and not one hair of her head shall be harmed.

This is *exactly* the project, but it’s probably going to have to wait for after the Singularity.

(see also: Gutenberg)

2015

- Untitled
- 2014 Predictions: Calibration Results
- Ethics Offsets
- Chronic Psychitis
- The Phatic And The Anti-Inductive
- The Physics Diet?
- The Influenza Of Evil
- Depression Is Not A Proxy For Social Dysfunction
- These Are A Few (More) Of My (Least) Favorite Things
- Perceptions Of Required Ability Act As A Proxy For Actual Required Ability In Explaining The Gender Gap
- A Philosopher Walks Into A Coffee Shop
- The Parable Of The Talents
- Talents Part 2: Attitude vs. Altitude
- Practically-A-Book Review: Dying To Be Free
- Everything Not Obligatory Is Forbidden
- The Efficacy Of Everything In Psychiatry In One Graph Plus Several Pages Of Dense But Necessary Explanation
- Black People Less Likely
- Money, Money, Everywhere, But Not A Cent To Spend
- Drug Testing Welfare Users Is A Sham, But Not For The Reasons You Think

- How Likely Are Multifactorial Trends?
- Did Falling Testosterone Affect Falling Crime?
- Pharma Virumque
- California Meetups
- Early Intervention: You *Might* Get What You Pay For
- California Meetups (Rererevised)
- A Series Of Unprincipled Exceptions
- A Cascade Of Dunbar Numbers
- Effective Altruists: Not As Mentally Ill As You Think
- Too Good To Be True
- Book Review: Willpower
- Answer to Job
- List Of Passages I Highlighted In My Copy Of “Willpower”
- Book Review: The Machinery Of Freedom
- List Of Passages I Highlighted In My Copy Of “Machinery Of Freedom”
- Is Everything A Religion?
- Extremism In Thought Experiment Is No Vice
- Highlights From My Notes From Another Psychiatry Conference
- Rational Orthography
- Chemical Imbalance
- No Physical Substrate, No Problem
- No Clarity Around Growth Mindset
- I Will Never Have The Ability To Clearly Explain My Beliefs About Growth Mindset
- Michigan Meetup 4/14
- Trouble Walking Down The Hallway

- Polemical Imbalance
- Blame Theory
- Universal Love, Said The Cactus Person
- Growth Mindset 3: A Pox On Growth Your Houses
- Nefarious Nefazodone And Flashy Rare Side Effects
- Prescriptions, Paradoxes, and Perversities
- The Future Is Filters
- Growth Mindset 4: Growth Of Office
- California, Water You Doing?
- Bicameral Reasoning
- Beware Summary Statistics
- AI Researchers On AI Risk
- No Time Like The Present For AI Safety Work
- That Chocolate Study
- …And I Show You How Deep The Rabbit Hole Goes
- Against Tulip Subsidies
- Michigan Meetup 6/13
- (Late) Predictions for 2015
- Fearful Symmetry
- Fifty (More) Swifties
- NYC Meetup 6/21
- Reflections From The Halfway Point
- The Case Of The Famous Physicist
- The Argument From Cultural Evolution
- Cultural Evolution 2: Thanks For The Meme-Rise
- Things That Sometimes Work If You Have Anxiety
- CBT In The Water Supply

- Reverse Psychology
- Freedom On The Centralized Web
- The General Factor Of Correctness
- Non-Dual Awareness
- Stalin and Summary Statistics
- Contra Hallquist On Scientific Rationality
- Contrarians, Crackpots, and Consensus
- Book Review: Chronicles Of Wasted Time
- Stop Adding Zeroes
- Figure/Ground Illusions
- My Id On Defensiveness
- The Goddess of Everything Else
- On Overconfidence
- Probabilities Without Models
- Mysticism and Pattern-Matching
- Magic Markers
- Theses On Trump
- If You Can’t Make Predictions, You’re Still In A Crisis
- Book Review: Manufacturing Consent
- Cardiologists and Chinese Robbers
- Beware Systemic Change
- Vegetarianism for Meat-Eaters
- The Problems With Generic Medications Go Deeper Than One Company
- Swifties 3: The Race Is Not To The Swifty
- Contra Caplan on Mental Illness
- Against Against Autism Cures
- It Was You Who Made My Blue Eyes Blue
- OT 30: Comment Knowledge
- Contra Simler on Prestige
- A Whiter Shade of Candidate

- [Contra Huemer On Morals]
- Meat Your Doom
- What Developmental Milestones Are You Missing?
- 2014 SSC Survey Results
- Looking A Gift Horse In The Mouth
- 2D:4D Ratio And Psychological Traits: Results From The LW/SSC Survey Sample
- Hardball Questions For The Next Debate
- SSC New England Meetup And Presentations Schedule
- Reporter Degrees Of Freedom
- College And Critical Thinking
- Setting The Default
- Product Recommendations 2015
- Book Review: Hive Mind
- List Of Passages I Highlighted In My Copy Of “Hive Mind”
- Should AI Be Open?
- How Bad Are Things?
- Things That Are Not Superintelligences
- Introducing Unsong

Untitled

EDIT: This is the most controversial post I have ever written in ten years of blogging. I wrote it because I was very angry at a specific incident. I stand by a lot of it, but if somebody links you here saying "HERE'S THE SORT OF GUY THIS SCOTT ALEXANDER PERSON IS, READ THIS SO YOU KNOW WHAT HIS BLOG IS *REALLY* ABOUT", please read *any other post* instead. There's a whole list of Top Posts on the Top Posts bar above.

Trigger warning: social justice, condemnation of some feminism, tangential reference to eating disorder. Note that although our names are very similar, I am NOT the same person as Scott Aaronson and he did NOT write this article. Not meant as a criticism of feminism, so much as of a certain way of operationalizing feminism.

I.

In my heart, there is a little counter that reads "XXX days without a ten-thousand word rant about feministm." And I had just broken three digits when *they had to go after Scott Aaronson*.

For those of you who don't know, Scott Aaronson is one of the nicest, smartest, and most decent people there are. A few days ago, in response to a discussion of sexual harassment at MIT, Aaronson reluctantly opened up about his experience as a young man:

> I check Feministing, and even radfem blogs like "I Blame the Patriarchy." And yes, I've read many studies and task force reports about gender bias, and about the "privilege" and "entitlement" of the nerdy males that's keeping women away from science. Alas, as much as I try to understand other people's perspectives, the first

reference to my “male privilege”—my privilege!—is approximately where I get off the train, because it’s so alien to my actual lived experience.

But I suspect the thought that being a nerdy male might not make me “privileged”—that it might even have put me into one of society’s least privileged classes—is completely alien to your way of seeing things. To have any hope of bridging the gargantuan chasm between us, I’m going to have to reveal something about my life, and it’s going to be embarrassing.

(sigh) Here’s the thing: I spent my formative years—basically, from the age of 12 until my mid-20s—feeling not “entitled,” not “privileged,” but terrified. I was terrified that one of my female classmates would somehow find out that I sexually desired her, and that the instant she did, I would be scorned, laughed at, called a creep and a weirdo, maybe even expelled from school or sent to prison. You can call that my personal psychological problem if you want, but it was strongly reinforced by everything I picked up from my environment: to take one example, the sexual-assault prevention workshops we had to attend regularly as undergrads, with their endless lists of all the forms of human interaction that “might be” sexual harassment or assault, and their refusal, ever, to specify anything that definitely wouldn’t be sexual harassment or assault. I left each of those workshops with enough fresh paranoia and self-hatred to last me through another year.

My recurring fantasy, through this period, was to have been born a woman, or a gay man, or best of all, completely asexual, so that I could simply devote my life to math, like my hero Paul Erdös did. Anything, really, other than the curse of having been born a heterosexual male, which for me, meant being consumed by desires that one couldn’t act on or even admit without running the risk of becoming an objectifier or a stalker or a harasser or some other creature of the darkness.

Of course, I was smart enough to realize that maybe this was silly, maybe I was overanalyzing things. So I scoured the feminist literature for any statement to the effect that my fears were as silly as I hoped they were. But I didn’t find any. On the contrary: I found reams of text about how even the most ordinary male/female interactions are filled with “microaggressions,” and how even the most “enlightened” males—especially the most “enlightened” males, in fact—are filled with hidden entitlement and privilege and a propensity to sexual violence that could burst forth at any moment.

Because of my fears—my fears of being “outed” as a nerdy heterosexual male, and therefore as a potential creep or sex criminal—I had constant suicidal thoughts. As Bertrand Russell wrote of his own adolescence: “I was put off from suicide only by the desire to learn more mathematics.”

At one point, I actually begged a psychiatrist to prescribe drugs that would chemically castrate me (I had researched which ones), because a life of mathematical asceticism was the only future that I could imagine for myself. The psychiatrist refused to prescribe them, but he also couldn’t suggest any alternative: my case genuinely stumped him. As well it might—for in some sense, there was nothing “wrong” with me. In a different social context—for example, that of my great-grandparents in the shtetl—I would have gotten married at an early age and been completely fine. (And after a

> decade of being coy about it, I suppose I've finally revealed the meaning of this blog's title.) […]
>
> Now, the whole time I was struggling with this, I was also fighting a second battle: to maintain the liberal, enlightened, feminist ideals that I had held since childhood, against a powerful current pulling me away from them. I reminded myself, every day, that no, there's no conspiracy to make the world a hell for shy male nerds. There are only individual women and men trying to play the cards they're dealt, and the confluence of their interests sometimes leads to crappy outcomes. No woman "owes" male nerds anything; no woman deserves blame if she prefers the Neanderthals; everyone's free choice demands respect.
>
> That I managed to climb out of the pit with my feminist beliefs mostly intact, you might call a triumph of abstract reason over experience. But I hope you now understand why I might feel "only" 97% on board with the program of feminism.

All right. Guy opens up for the first time about how he was so terrified of accidentally hurting women that he became suicidal and tried to get himself castrated. Eventually he got over it and is now 97% on board with feminism, but wants people to understand that when done wrong it can be really scary.

The feminist blogosphere, as always, responded completely proportionally. Amanda Marcotte, want to give us a representative sample?

> [Aaronson's post] is the whole "how can men be oppressed when I don't get to have sex with all the hot women that I want without having to work for it?" whine, one that, amongst other things, starts on the assumption

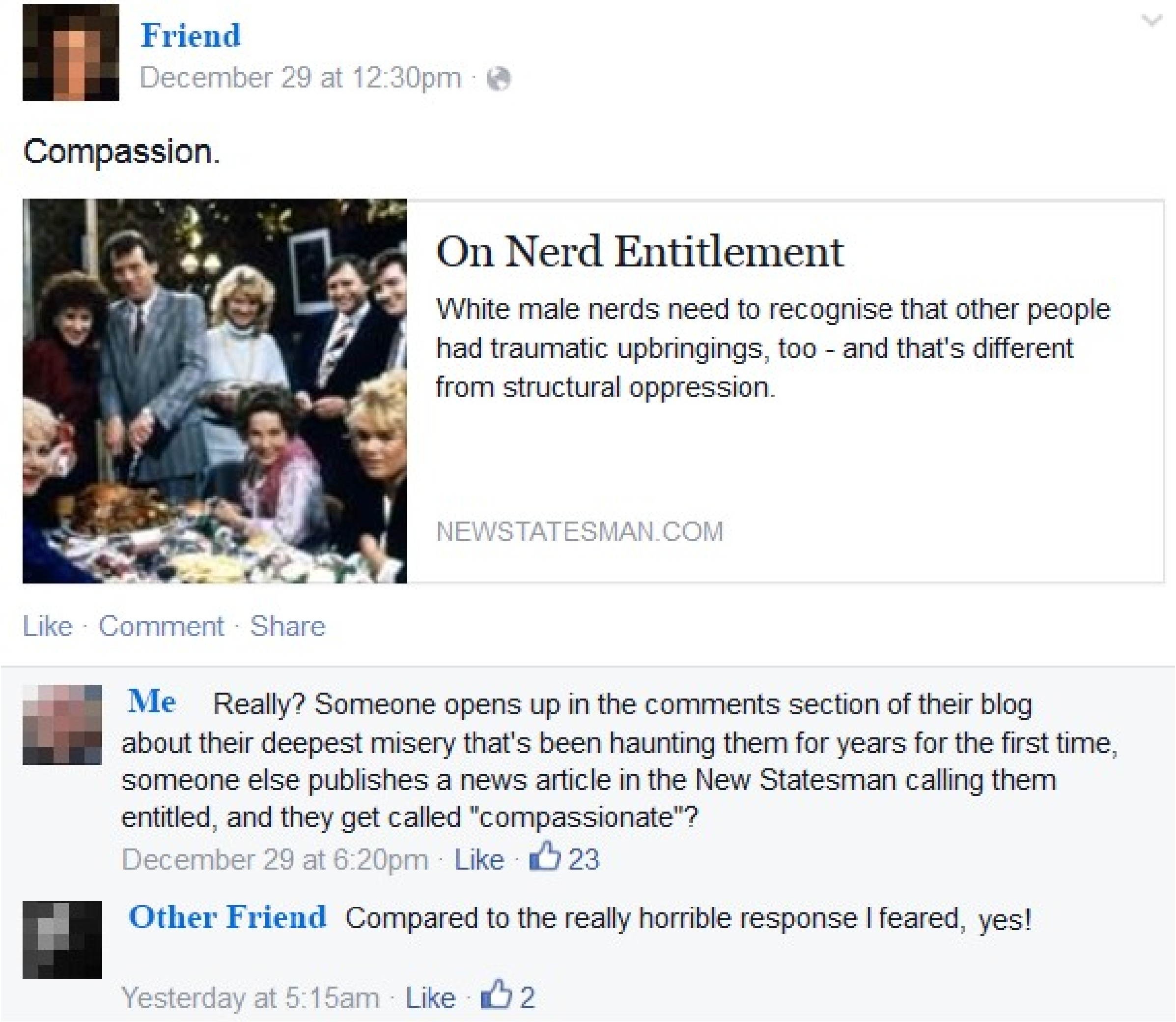
> that women do not suffer things like social anxiety or rejection…It was just a yalp of entitlement combined with an aggressive unwillingness to accept that women are human beings just like men. [He is saying that] “having to explain my suffering to women when they should already be there, mopping my brow and offering me beers and blow jobs, is so tiresome…I was too busy JAQ-ing off, throwing tantrums, and making sure the chip on my shoulder was felt by everyone in the room to be bothered to do something like listen.” Women are failing him by not showing up naked in his bed, unbidden. Because bitches, yo.
>
> The eternal struggle of the sexist: Objective reality suggests that women are people, but the heart wants to believe they are a robot army put here for sexual service and housework.

This would usually be the point where I state for the record that I believe very strongly that all women are human beings. Problem is, I’ve just conceived a sudden suspicion that one of them is actually a Vogon spy in a skin suit.

Anyway, Marcotte was bad enough, given that she runs one of the most-read feminist blogs on the Internet. But much of the rest of the feminist “discussion” on Tumblr, Twitter, and the like was if anything even worse.

But there was one small ray of hope. A bunch of people sent me an article on the issue by Laurie Penny in New Statesman, called “On Nerd Entitlement: White Male Nerds Need To Recognize That Other People Had Traumatic Upbringings Too And That’s Different From Structural Oppression.” The article was always linked with commentary like “This is so compassionate!” or “Finally a decent human being is addressing this issue with kindness!”

Well, I read the article, and ended up having the following Facebook conversation:



Friend
December 29 at 12:30pm ·

Compassion.

On Nerd Entitlement

White male nerds need to recognise that other people had traumatic upbringings, too - and that's different from structural oppression.

NEWSTATESMAN.COM

Me Really? Someone opens up in the comments section of their blog about their deepest misery that's been haunting them for years for the first time, someone else publishes a news article in the New Statesman calling them entitled, and they get called "compassionate"?
December 29 at 6:20pm · Like · 23

Other Friend Compared to the really horrible response I feared, yes!
Yesterday at 5:15am · Like · 2

On further reflection, Other Friend has a point. I disliked Penny’s article, but *compared to everything else* it was a ray of light, a breath of fresh air, an unexpected incursion from a utopia of universal love and understanding. I didn’t feel like it treated Aaronson fairly. But I did feel like it treated him like a human being, which is rare and wonderful.

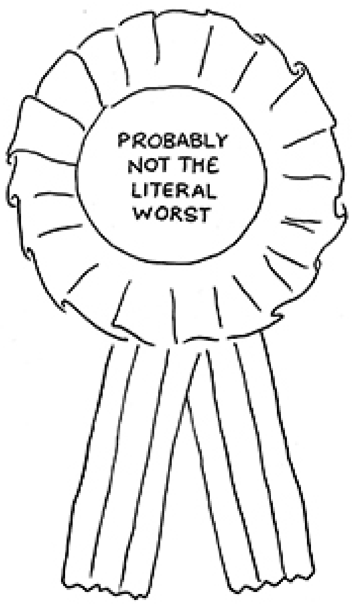
From the article:

> I do not intend for a moment to minimise Aaronson’s suffering. Having been a lonely, anxious, horny young

> person who hated herself and was bullied I can categorically say that it is an awful place to be. I have seen responses to nerd anti-feminism along the lines of "being bullied at school doesn't make you oppressed". Maybe it's not a vector of oppression in the same way, but it's not nothing. It burns. It takes a long time to heal.

That this article keeps being praised effusively for admitting that someone else's suicidal suffering "isn't nothing", is a sign. It's a sign of how low our standards are. But it's also a sign people are ready for change.

It's hard for me express simultaneously both how genuinely grateful and impressed I am that the article managed to avoid being awful, and how far I still think it has to go. I can only offer Ms. Penny and the entire staff of the New Statesman the recognition appropriate for their achievement:


PROBABLY NOT THE LITERAL WORST

I've already written some thoughts on this general issue in Radicalizing The Romanceless. But by bringing nerd-dom into the picture, Penny has made that basic picture exponentially more complicated.

Luckily, this is a post about Scott Aaronson, so things that become exponentially more complicated fit the theme perfectly.

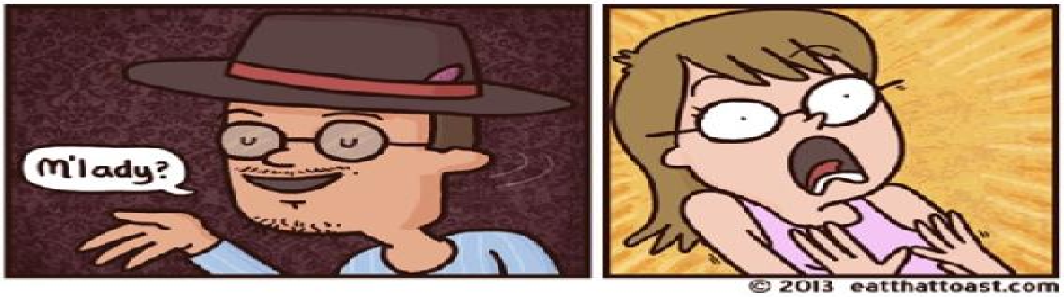
II.

Ms. Penny writes:

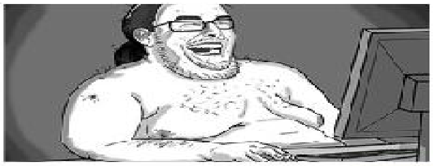
> Feminism is not to blame for making life hell for “shy, nerdy men”. It is a real shame that Aaronson picked up Andrea Dworkin rather than any of the many feminist theorists and writers who manage to combine raw rage with refusal to resort to sexual shame as an instructive tool. Weaponised shame – male, female or other – has no place in any feminism I subscribe to.

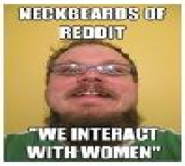
I live in a world where feminists throwing weaponized shame at nerds is an obvious and inescapable part of daily life. Whether we’re “mouth-breathers”, “pimpled”, “scrawny”, “blubbery”, “sperglord”, “neckbeard”, “virgins”, “living in our parents’ basements”, “man-children” or whatever the insult du jour is, it’s always, *always*, ALWAYS a self-identified feminist saying it. Sometimes they say it obliquely, referring to a subgroup like “bronies” or “atheists” or “fedoras” while making sure everyone else in nerddom knows it’s about them too.

There continue to be a constant stream of feminist cartoons going around Tumblr featuring blubberous neckbearded fedora-wearing monsters threatening the virtue of innocent ladies.



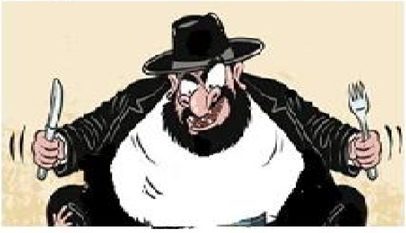
© 2013 eatthattoast.com



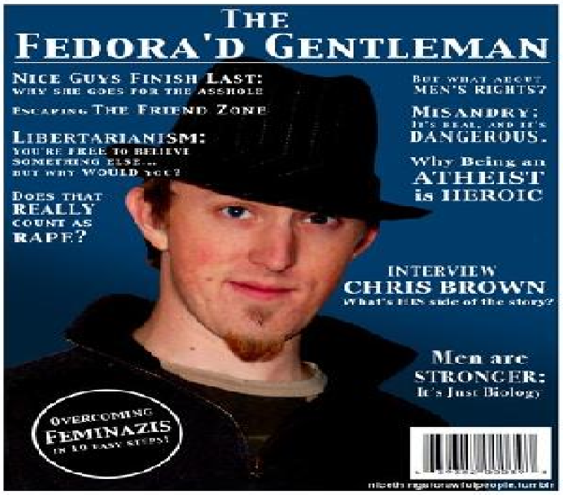

NECKBEARDS OF REDDIT
"WE INTERACT WITH WOMEN"

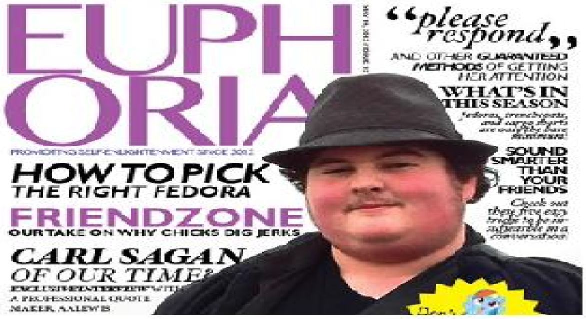
LE FRIENDZONE

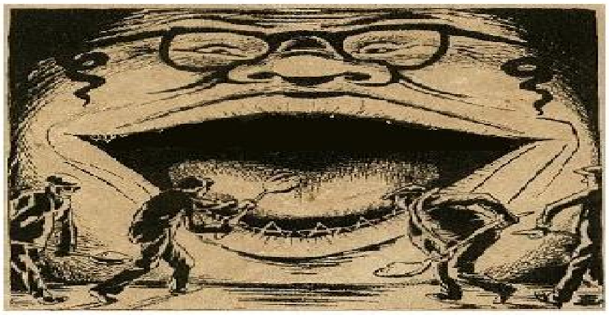
DO YOU ACCEPT MY APOLOGIES M'LADY
I WON'T BE ABLE TO SLEEP TONIGHT M'LADY

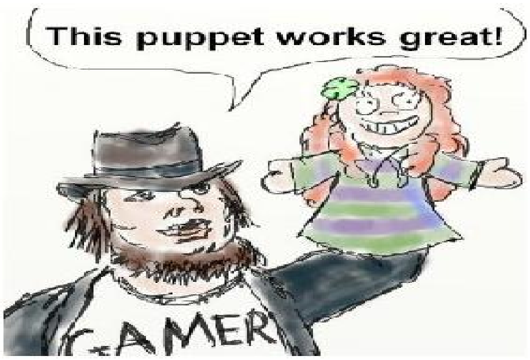


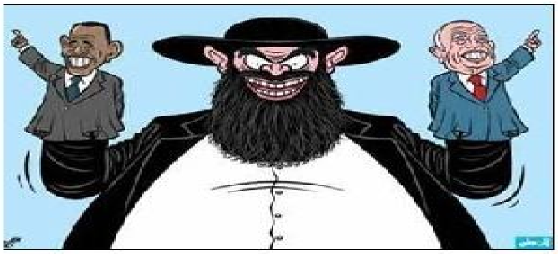
IT'S SO HARD
TO BE AN ALPHA MALE THESE DAYS


THE FEDORA'D GENTLEMAN
NICE GUYS FINISH LAST: WHY SHE GOES FOR THE ASSHOLE
ESCAPING THE FRIEND ZONE
LIBERTARIANISM: YOU'RE FREE TO BELIEVE SOMETHING ELSE... BUT WHY WOULD YOU?
DOES THAT REALLY COUNT AS RAPE?
BUT WHAT ABOUT MEN'S RIGHTS?
MISANDRY: IT'S REAL, AND IT'S DANGEROUS.
Why Being an ATHEIST is HEROIC
INTERVIEW CHRIS BROWN What's HIS side of the story?
Men are STRONGER: It's Just Biology
OVERCOMING FEMINAZIS IN 10 EASY STEPS!


EUPHORIA
PROMOTING SELF-ENLIGHTENMENT SINCE 2012
HOW TO PICK THE RIGHT FEDORA
FRIENDZONE OUR TAKE ON WHY CHICKS DIG JERKS
CARL SAGAN OF OUR TIME
"please respond" AND OTHER GUARANTEED METHODS OF GETTING HER ATTENTION
WHAT'S IN THIS SEASON
SOUND SMARTER THAN YOUR FRIENDS




This puppet works great!
GAMER



Oops, I accidentally included three neo-Nazi caricatures of

Oops, I accidentally included three neo-Nazi caricatures of Jews in there. You did notice, right?

Read any article from the appropriate subfield of feminism, and you may well run into the part with the girl walking into a comic book store only to be accosted by a mouth-breathing troglodyte, followed by a "lesson" on nerd male privilege.

But it's not just that. Try to look up something on Iron Man, and you get an article on Iron Man-Child and how "the white maleness of geek culture" proves they are "the most useless and deficient individuals in society, precisely because they have such a delusional sense of their own importance and entitlements." Go to Jezebel and people are talking about how jocks are so much better than nerds because nerds hate women.

It has reached the point where articles published in major journals talk about the the fedora phenomenon in the context of "the growing trend in feminists and other activists online that use shaming as an activist strategy".

Let's not mince words. There is a growing trend in Internet feminism that works *exactly* by conflating the ideas of nerd, misogynist, virgin, person who disagrees with feminist tactics or politics, and unlovable freak.

Ms. Penny may be right that her ideal feminism doesn't do that. Then again, my ideal masculinity doesn't involve rape or sexual harassment. Ideals are *always* pretty awesome. But women still have the right to complain when *actual* men rape them, and I'm pretty sure nerds deserve the right to complain that actual feminists are, a lot of the time, focused way more on nerd-baiting than actual feminism, and that much the same people who called us "gross" and "fat" and "loser" in high school are calling us "gross" and "misogynist" and "entitled" now, and for much the same reasons.

III.

Penny goes on to deny that this is a gendered issue at all:

> Like Aaronson, I was terrified of making my desires known- to anyone. I was not aware of any of my (substantial) privilege for one second – I was in hell, for goodness' sake, and 14 to boot…Scott, imagine what it's like to have all the problems you had and then putting up with structural misogyny on top of that. Or how about a triple whammy: you have to go through your entire school years again but this time you're a lonely nerd who also faces sexism and racism.

This comes across so strongly as "my suffering is worse than your suffering" spiel, so much so that I'm tempted to argue it and review a bunch of experiments like how even the least attractive women on dating sites get far more interest than men. Or how women asking random people for sex on the street get accepted more than two-thirds of the time, but men trying the same get zero percent. Or how the same study shows that the women who get declined get declined politely, while the men are treated with disgust and contempt. Or I could hunt down *all* of the stories of trans men who start taking testosterone, switch to a more male sex drive, and are suddenly like "OH MY GOD I SUDDENLY REALIZE WHAT MALE HORNINESS IS LIKE I THOUGHT I KNEW SEXUAL FRUSTRATION BEFORE BUT I REALLY REALLY DIDN'T HOW DO YOU PEOPLE LIVE WITH THIS?"

But my commenters have convinced me that taking this further would be joining in the pissing contest I'm condemning, so let's put it a little differently.

A couple of studies show that average-attractiveness people who ask random opposite-gender strangers on dates are accepted 50% of the time, regardless of their gender.

Grant that everyone involved in this conversation has admitted they consider themselves below average attractiveness (except maybe Marcotte, whose daily tune-ups keep her skin-suit in excellent condition). Fine. Maybe we have a success rate of 10%?

That's *still* astounding. It would be pretty easy to mock teenage-me for not asking for dates when ten percent of people would have said yes. Asking ten people something takes what, five minutes? And would have saved *how* many years of misery?

This is a *pretty impressive* market failure – in sheer utility cost, probably bigger than any of the market failures actual economists talk about.

Some people say the female version of the problem is men's fault, and call the behavior involve slut-shaming. I take this very seriously and try not to slut-shame or tolerate those who do.

But the male version of the problem is nerd-shaming or creep-shaming or whatever, and I don't feel like most women, especially most feminist women, take it nearly as seriously as I try to take their problems. If anything, many actively make it worse. This is exactly those cartoons above and the feminists spreading them. Nerds are told that if they want to date girls, that makes them disgusting toxic blubberous monsters who are a walking offense to womankind.

This is maybe not the *most* reasonable interpretation of modern sexual mores, but neither is "any women who has sex before marriage is a slut and no one will ever value her." Feminists are eagle-eyed at spotting the way seemingly innocuous messages in culture can accidentally reinforce the latter, but continue to insist that there's no possible way that shouting the former from the rooftops could possibly lead to anyone believing or internalizing it.

Talking about "entitled nerds" is the Hot New Internet Feminism thing these days. Here's The Entitlement And Misogyny Of Nerd Culture. Here's Sex, Nerds, Entitlement, and Rape. Here's Is Nerd Culture Filled With Entitled Crybabies? There's On Male Entitlement: Geeks, Creeps, and Sex.

And now, apparently, the New Statesman, realizing that it's almost 2015 and it has yet to claim a share of the exciting nerd entitlement action, has On Nerd Entitlement by Laurie Penny

And this is more than a little weird, because the *actual* nerds I know in real life tend to be more like Scott Aaronson, who is spending less time feeling entitled to sex, and more time asking his doctor if there's any way to get him castrated because his sexual desire might possibly offend a woman. Or more like me, who got asked out by a very pretty girl in middle school and ran away terrified because he knew nobody could actually like him and it was obviously some kind of nasty trick.

So given that real-life nerds are like this, and given that they're sitting around being terrified that they're disgusting toxic monsters whose wish to have sex is an offense against womankind, *what do you think happens when they hear from every news source in the world that they are entitled?*

What happens is they think "Oh God! There was that one time when I looked at a woman and almost thought about asking her out! That means I must be feeling entitled to sex! I had temporarily forgotten that as a toxic monster I must never show any sexuality to anybody! Oh God oh God I'm even worse than I thought!"

Again, this is not the most rational thing in the world. But I maintain it’s no less rational than, say, women who won’t leave their abusive husband because he’s convinced them they don’t deserve anything better than what they get. Gender is weird. Self-loathing is easy to inculcate and encourage, even unintentionally. Heck, we’ve already identified this market failure of people preferring to castrate themselves rather than ask ten people on a date, *something* weird has got to explain it.

When feminists say that the market failure for young women is caused by slut-shaming, I stop slut-shaming, and so do most other decent people.

When men say that the market failure for young men is caused by nerd-shaming, feminists write dozens of very popular articles called things like “On Nerd Entitlement”.

The reason that my better nature thinks that it’s irrelevant whether or not Penny’s experience growing up was better or worse than Aaronson’s: when someone tells you that something you are doing is making their life miserable, you don’t lecture them about how your life is worse, even if it’s true. You STOP DOING IT.

IV.

This also serves to illuminate what I think is the last and most important difference between Penny’s experience and Aaronson’s experience.

When Penny bares her suffering to the world for all to hear about, she gets sympathy, she gets praised as compassionate, she gets published in important magazines whose readers feel sorry for her and acknowledge that her experience sucks.

When Aaronson talks about his suffering on his own blog, he gets Amanda Marcotte. He gets half the internet telling him he is now the worst person in the world.

This was my experience as well. When I complained that I felt miserable and alone, it was like throwing blood in the water. A feeding frenzy of feminists showed up to tell me I was a terrible person and deserved to die, sometimes in terms that made Marcotte look like grandmotherly kindness. This is part of the experience I write about in this post, and it's such a universal part of the shy awkward male experience that we are constantly flabbergasted that women refuse to accept it exists.

When feminists write about this issue, they nearly always assume that the men involved are bitter about all the women who won't sleep with them. In my experience and the experience of everyone I've ever talked to, we're bitter about all the women who told us we were disgusting rapists when we opened up about our near-suicidal depression.

And when that happens, again and again and again, of course we learn to shut up about it. I bottled my feelings inside and never let them out and spent years feeling like I was a monster for even having them.

As a mental health professional, I can assure you this is *the best* coping strategy.

V.

Laurie Penny has an easy answer to any claims that any of this is feminists' fault:

> Feminism, however, is not to blame for making life hell for "shy, nerdy men". Patriarchy is to blame for that.

I say: why can’t it be both?

Patriarchy is yet another motte and bailey trick.

The motte is that patriarchy is the existence of different gender roles in our society and the ways in which they are treated differently.

The bailey is that patriarchy is men having power over women.

If you allow people to switch between these and their connotations willy-nilly, then you enable all sorts of mischief.

Whenever men complain about anything, you say “Oh, things are bad for men? Well, that sounds like a gender role. Patriarchy’s fault!”

And then the next day you say “Well, since we already agreed yesterday your problem is patriarchy, the solution is take away power from men and give it to women. It’s right there in the word, patri-archy. So what we need is more feminism.”

Even if in this particular case the feminism is making the problem worse.

So, for example, we are told that the patriarchy causes male rape. We are told that if we want to fight male rape, the best way to do so is to work hard to promote feminist principles. But once feminism has been promoted, the particular feminists benefitting from that extra social capital may well be the ones to successfully lobbying national governments to keep male rape legal on the ground that if raping men was illegal, they might make false accusations which could hurt women.

If patriarchy is “any problem with gender roles”, it’s entirely possible, even predictable, that feminists can be the ones propping it up in any given situation.

I mean, we live in a world where the Chinese Communist Party is the group that enforces Chinese capitalism and oppresses any workers who complain about it. We live in a world where the guy who spoke out against ritualized purity-obsessed organized religion ended up as the founder of the largest ritualized purity-obsessed organized religion of all time. We live in a world where the police force, which is there to prevent theft and violence, is confiscating property and shooting people right and left. It seems neither uncommon nor unexpected that if you charge a group with eliminating an evil that’s really hard to eliminate, they usually end up mildly tweaking the evil into a form that benefits them, then devoting most of their energy to punishing people who complain.

Pick any attempt to shame people into conforming with gender roles, and you’ll find self-identified feminists leading the way. Transgender people? Feminists led the effort to stigmatize them and often still do. Discrimination against sex workers? Led by feminists. Against kinky people? Feminists again. People who have too much sex, or the wrong kind of sex? Feminists are among the jeering crowd, telling them they’re self-objectifying or reinforcing the patriarchy or whatever else they want to say. Male victims of domestic violence? It’s feminists fighting against acknowledging and helping them.

Yes, many feminists have been on both sides of these issues, and there have been good feminists tirelessly working against the bad feminists. Indeed, right now there are feminists who are telling the other feminists to lay off the nerd-shaming. My girlfriend is one of them. But that’s kind of my point. There are feminists on both sides of a lot of issues, including the important ones.

("But nowadays in 2015 most feminists are on the right side of every gender issue, right?" Insofar as your definition of 'the right side of a gender issue' is heavily influenced by 'the side most feminists are on', I'm going to have a really hard time answering that question in a non-tautologous way. Come back in 2065 and we can have a really interesting discussion about whether the feminists of 2015 screwed up as massively as the feminists of 1970 and 1990 did.)

So feminists can be either against or in favor of "patriarchy" broadly defined. Whether or not a form of cruelty is decreed to be patriarchy doesn't tell us how many feminists are among the people twisting the knife.

The preferred method of figuring this out is asking the people involved.

I've been saying for years that getting exposed to feminist shaming was part of what made my adolescence miserable. Every time I say this, I get a stream of grateful emails thanking me for saying something so true to their experience.

Scott Aaronson has now said that getting exposed to feminist shaming was part of what made his adolescence miserable. According to his most recent blog post, he's *also* getting the stream of grateful emails:

> Throughout the past two weeks, I've been getting regular emails from shy nerds who thanked me profusely for sharing as I did, for giving them hope for their own lives, and for articulating a life-crushing problem that anyone who's spent a day among STEM nerds knows perfectly well, but that no one acknowledges in polite

> company. I owe the writers of those emails more than they owe me, since they're the ones who convinced me that on balance, I did the right thing.

I hang out a lot with shy awkward nerdy men of all ages, and I very often hear from them that feminist shaming is part of what's making their adolescence (and often current life) miserable.

And it's not just men. Here's what a lesbian friend of mine had to say about Penny's article:

> There are a hell of a lot of people attracted to women who seem to have internalized the message that their attraction makes them sick and wrong and evil and creepy, that basically any interaction they have with a woman is coercive or harmful on their part, and that initiating a romantic interaction makes them a sexual predator.
>
> I know this because I'm one of them.
>
> I'm a woman. I'm gay. By the time I realized that second thing, I'd internalized that all attraction to women was objectifying and therefore evil. I spent years of my life convinced that it was coercive to make it clear to girls that I wanted to date them, lest they feel pressured. So I could only ask them out with a clear conscience if I was in fact totally indifferent to their answer. I still decide I'm abusive pretty frequently, on the basis of things like 'i want to kiss her, which is what an abuser would want' and 'i want to be special to her, which is what an abuser would want'.
>
> I internalized these messages from exposure to feminist memes, norms, and communities. It was feminist messages, not homophobic ones, that made it hardest

> for me to come to terms with my sexuality. It wasn’t intentional. But it happened. And it has happened by now to enough people that ‘well obviously you’re misinterpreting it’ is starting to wear thin as an excuse. Lots and lots of people are misinterpreting the way I did. By and large, we’re vulnerable people. Very often we’re mentally ill or disabled people.
>
> Even if it’s broadly good for feminism to emphasize narratives about objectification and entitlement, this seems like a negative consequence of the way contemporary feminist activism does that. Activism shouldn’t make vulnerable people suicidally guilty. If there was a way to do activism that didn’t have this consequence, it’d be better than the current setup.
>
> The infuriating thing is that I think there might be. We could write articles acknowledging that certain conversations can exacerbate crippling guilt and self-loathing, particularly for people with anxiety, depression, or other mental illnesses that make them fixate on their own perceived worthlessness. We could really, truly, not-just-lip-service integrate concern for those people into our activism. We could acknowledge how common this experience is and have resources to help people. We could stop misidentifying anguish as entitlement, and stop acting like anguish that does have entitlement at its root is deserved or desirable or hilarious.
>
> We could really just start by extending to men who share this experience with me the sympathy that I’m extended when I talk about it.

The responses on Tumblr from men and women all over the sexuality spectrum who have had any personal experience with this all say it’s how they feel as well.

I usually avoid the term "privilege" because it tends to start World War III when used. So let's avoid the term and simply keep in mind the concept that people have private information about their own experience that it's difficult for other people to get second-hand.

Ms. Penny, as an (I think?) heterosexual woman, has *no idea* what having to deal with our culture's giant minefield around romance toward women is like.

Scott Aaronson is a straight guy, and he's saying feminist shaming tactics have made it worse. I'm an asexual heteroromantic guy, and I'm telling her feminist shaming tactics have made it worse. Unitofcaring is a lesbian woman, and she's saying feminist shaming tactics have made it worse. HughRistik, who is some sort of weird metrosexual something (I mock him because I love him), is telling her feminist shaming tactics have made it worse. A giant cry has arisen from shy awkward men, lesbians, bisexuals, whatever of the world is saying "NO, SERIOUSLY, FEMINIST SHAMING TACTICS ARE MAKING THIS WORSE"

When Ms. Penny protests that feminism can't possibly be involved and all these other people's s personal experience is wrong, this is coming from a place of startling arrogance. If patriarchy means everything in the world, then yes, it is the fault of patriarchy. But it's the kind of patriarchy that feminism as a movement is working day in and day out to reinforce.

VI.

The subtitle of the article is "White male nerds need to recognise that other people had traumatic upbringings, too – and that's different from structural oppression."

This doesn’t really describe the argument very well. The closest it really comes is to say that:

> Aaronson makes a sudden leap, and it’s a leap that comes right from the gut, from an honest place of trauma and post-rationalisation, from that teenage misery to a universal story of why nerdy men are in fact among the least privileged men out there, and why holding those men to account for the lack of representation of women in STEM areas – in the most important fields both of human development and social mobility right now, the places where power is being created and cemented right now – is somehow unfair […]
>
> This is why Silicon Valley is fucked up. Because it’s built and run by some of the most privileged people in the world who are convinced that they are among the least.
>
> I really fucking hope that it got better, or at least is getting better, At the same time, I want you to understand that that very real suffering does not cancel out male privilege, or make it somehow alright. Privilege doesn’t mean you don’t suffer, which, I know, totally blows.

The impression I’m getting is that yes, nerds think they have problems, but actually they’re really privileged. So their problems aren’t structural oppression in the same sense that women’s problems are. So. Quick hypothetical.

I’ve postulated before that “privilege” is a classic motte-and-bailey term. The motte, the uncontroversial and attractive definition, is “some people have built-in advantages over other people, and it might be hard for them to realize these advantages even exist”. Under this definition, it’s easy to agree that, let’s say, Aaronson has the privilege of not having to deal with slut-shaming, and Penny has the privilege of not having to deal with the kind of creep-shaming that focuses on male nerds.

The bailey, the sneaky definition used to push a political point once people have agreed to the motte, is that privilege is a one-dimensional axis such that for any two people, one has privilege over the other, and that first person has it better in every single way, and that second person has it worse in every single way.

This is of course the thing everyone swears they *don't* mean when they use the word privilege, which is of course how the motte-and-bailey fallacy works. But as soon as they are not being explicitly challenged about the definition, this is the way they revert back to using the word.

Go back to the original Amanda Marcotte article. Check the title. "MIT Professor Explains The Real Oppression Is Having To Talk To Women".

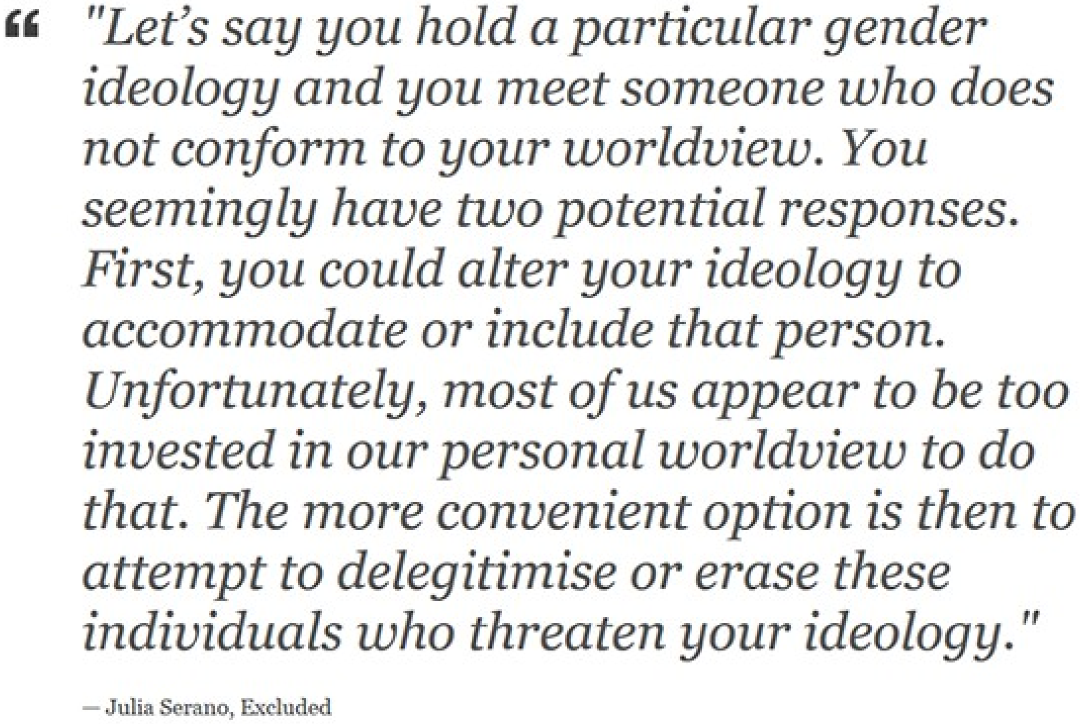
That phrasing, "the real oppression is…", carries a pretty loaded assumption. I'd say "hides a pretty loaded assumption", but it doesn't seem to be doing much work to hide it.

If you look through Marcotte's work, you find this same phrasing quite often. "Some antifeminist guy is ranting at me about how men are the ones who are really oppressed because of the draft" (source). And she's not the only one. If you Google the term "are the ones who are really oppressed", you can find an nice collection of people using this exact phraseology, including a few examples from a charming site called "Nerds Fucking Suck".

But Aaronson is admitting about a hundred times that he recognizes the importance of the ways women are oppressed. He's not saying his suffering is worse than women’s in every way, just that it’s really bad and maybe this is not the place where “male privilege” should be invoked. The “is really oppressed” isn’t taken from him, it’s assumed by Marcotte. Her obvious worldview is – since privilege and oppression are a completely one dimensional axis, for Aaronson to claim that there is *anything whatsoever* that has ever been bad for men must be interpreted as a claim that they are the ones who are really oppressed and therefore women are not the ones who are really oppressed and therefore nothing whatsoever has ever been bad for women. By Insane Moon Logic, it sort of makes sense.

As a result, Marcotte is incapable of acknowledging that Aaronson feels pain or has feelings more complicated than “all women exist solely to be my slaves”. She *has* to be a jerk to him, otherwise it would be a tacit admission that he has problems, which means *only* he has problems, which means no woman has ever had problems, which means all women are oppressors. Or whatever.

Marcotte is angry that Aaronson doesn’t cite any feminist writer besides Andrea Dworkin, so let’s go with Julia Serano here:



> "Let's say you hold a particular gender ideology and you meet someone who does not conform to your worldview. You seemingly have two potential responses. First, you could alter your ideology to accommodate or include that person. Unfortunately, most of us appear to be too invested in our personal worldview to do that. The more convenient option is then to attempt to delegitimise or erase these individuals who threaten your ideology."
>
> — Julia Serano, Excluded

What if you're trying to hold the same weird one-dimensional system in a way consistent with basic human decency? That is, you don't want to do the Vogon thing and say Scott Aaronson's misery is totally hilarious, but you also don't want to acknowledge that it counts – because if it counted you'd have to admit that men have it bad in some ways, which means that the One Group That Can Ever Have Things Bad spot is taken by men, which means women don't have it bad?

As best I can tell, the way with the fewest epicycles is to say "Yes, your pain technically exists, but it's not *structural oppression*", where structural oppression is the type of pain that fits neatly onto the one-dimensional line.

Laurie Penny is an extremely decent person, but like a shaman warding off misfortune with a ritual, she must dub Aaronson's pain "not structural oppression" or else risk her own pain not counting, being somehow diminished.

I mean, I don't think she thinks that's what she's doing. But I'm not sure why else it's necessary to get so competitive about it.

Absent the one-dimensional view, it would be perfectly reasonable to say something like "You feel pain? I have felt pain before too. I'm sorry about your pain. It would be incredibly crass to try to quantify exactly how your pain compares to my pain and lord it over you if mine was worse. Instead I will try to help you with your pain, just as I hope that you will help me with mine."

Given the one-dimensional view, any admission that other people suffer is a threat to the legitimacy of one's own suffering. Horrible people will deny and actively mock the pain of others, but even decent people will only be able to accept the pain if they also mention in an aside that it doesn't count as the correct sort of pain to matter in the moral calculus and certainly isn't even in the same *ballpark* as their own.

But the one-dimensional view sucks. It is the culmination and perfection of the phenomenon I described in my post on social justice terminology, the abandonment of discourse about the world in favor of endless debate about who qualifies for certain highly loaded terms like "structural oppression". And those terms end up as a sort of Orwellian Newspeak that makes it possible to dismiss entire categories of experience and decree by fiat who does and doesn't matter.

> Imagine a boot stamping on a human face forever, saying "I KNOW YOU FEEL UPSET RE STAMPING, BUT THAT'S DIFFERENT FROM STRUCTURAL OPPRESSION"
>
> — Scott Alexander (@slatestarcodex) December 31, 2014

The boot acknowledged my pain! So compassionate!

§

The suspect famously says "I didn't kill him, officer! Also, he had it coming!"

In that spirit, I would like to propose that we shouldn't make this debate about structural oppression, but *even if we do* this kind of minimization of male nerd suffering doesn't stand.

I know there are a couple different definitions of what exactly structural oppression is, but however you define it, I feel like people who are at much higher risk of being bullied throughout school, are portrayed by the media as disgusting and ridiculous, have a much higher risk of mental disorders, and are constantly told by mainstream society that they're ugly and defective *kind of* counts.

If nerdiness is defined as intelligence plus poor social skills, then it is *at least* as heritable as other things people are willing to count as structural oppression like homosexuality (heritability of social skills, heritability of IQ, heritability of homosexuality) If all nerds were born with blue dots on their heads, and the blue-dotters were bullied in school, cast negatively in the media, assumed to be as ravenous beasts hungry for innocent women, and denounced as "entitled" any time they overcame all this to become successful – would anybody deny that blue-dotters suffered from structural oppression? Wouldn't the people who talked about how clearly blue-dotters are entitled dudebros in the tech industry be thought of the same way as someone who said Jews were greedy parasites in the banking industry?

Actually, let's take this Jew thing and run with it. I am not the first person to notice that there are a lot of Jews in Silicon

Valley. By maternal descent, at least Mark Zuckerberg, Sergey Brin, Larry Page, Michael Dell, Steve Ballmer, Larry Ellison, and Sheryl Sandberg.

Imagine how an anti-Semite might think about this. “Jews say they’re oppressed. But actually they’re all rich. Oppression disproved!”

In fact, he might he add exactly the same comment we see in the Statesman article: “This is why Silicon Valley is fucked up. Because it’s built and run by some of the most privileged people in the world who are convinced that they are among the least.”

But *once again* this only works when you have the dumb one-dimensional model of privilege. Some Jews are rich, therefore all Jews are rich, therefore all Jews are privileged, therefore no Jew could be oppressed in any way, therefore Jews are the oppressors.

And much the same is true of nerds. In fact, have you noticed actual nerds and actual Jews tend to be the same people? I’m Jewish. Scott Aaronson is Jewish. Laurie Penny, who declares her nerd-girl credentials, is Jewish. We’re discussing a blog called, of all things, *Shtetl-Optimized*. A minority that makes up 1% of the Anglosphere also makes up three of the three nerds in this conversation. Probability of this happening by chance is (**calculates**) exactly one in a million. Aside from Zuckerberg, Page, and all the other famous people, about 40% of top programmers are Jewish.

Judaism and nerdity are not *exactly* the same, but they sure live pretty close together.

And this is why it’s distressing to see the *same things* people have always said about Jews get applied to nerds. They’re this weird separate group with their own culture who don’t join in the reindeer games of normal society. They dress weird and talk weird. They're conventionally unattractive and have too much facial hair. But worst of all, they have the *chutzpah* to do all that and also be successful. Having been excluded from all of the popular jobs, they end up in the unpopular but lucrative jobs, for which they get called greedy parasites in the Jews' case, and "the most useless and deficient individuals in society" in the case of the feminist article on nerds I referenced earlier.

Propaganda against the Jews is described as follows:

> Since Jews were ugly, they depended on reprehensible methods of sexual conquest. Non-violent means such as money were common, but also violence. Streicher specialized in stories and images alleging Jewish sexual violence. In a typical example, a girl cowers under the huge claw-like hand of a Jew, his evil silhouette in the background. The caption at the bottom of the page: "German girls! Keep away from Jews!" These images were particularly striking and consistent with the larger theme. Although Jews were too cowardly to engage in manly combat and too disgusting to be physically attractive to German women, they were eager to overpower and rape German women, thereby corrupting the Aryan racial stock.

I already know the same machine that turned Aaronson's "I am 97% on board with feminism" into "I think all women should be my slaves" is focusing its baleful gaze on me. So let me specify what I am obviously *not* saying. I am not saying nerds have it "just as bad as Jews in WWII Germany" or any nonsense like that. I am not saying that prejudice against nerds is literally motivated by occult anti-Semitism, or accusing anyone of being anti-Semitic.

I am saying that whatever structural oppression means, it should be about *structure*. And the structure society uses to marginalize and belittle nerds is very similar to a multi-purpose structure society has used to belittle weird groups in the past with catastrophic results.

There is a well-known, dangerous form of oppression that works just fine when the group involved have the same skin color as the rest of society, the same sex as the rest of society, and in many cases are totally indistinguishable from the rest of society except to themselves. It works by taking a group of unattractive, socially excluded people, mocking them, accusing them of being out to violate women, then denying that there could possibly be any problem with these attacks because they include rich people who dominate a specific industry.

[EDIT 1/3: Penny's same article was reprinted at New Republic, which I guess also realized it gotten a piece of the Hot New Nerd Entitlement Trend yet. Their title was "Nerd Entitlement Lets Men Ignore Racism And Sexism", which is kind of weird, since Penny's article doesn't do anything close to argue for that. Also since surveys show nerd men are more likely to be concerned about racism and sexism than other men – see for example this survey where nerds are far more feminist than average, so much so that nerd men are more feminist than non-nerd women, and since Penny's article makes nothing even resembling an argument for this position. Once again, this only makes sense if you assume a one-dimensional zero-sum model of privilege, where the fact that miserable male nerds are concentrating on their own desire for the release of death, instead of what women think they should be concentrating on, means they must be universally denying women can have problems.]

[EDIT 1/3, Part 2: New Republic has changed their title. You can still see it in the URL, though]

VII.

It gets worse.

> What can I say? This is a strange and difficult age, one of fast-paced change and misunderstandings. Nerd culture is changing, technology is changing, and our frameworks for gender and power are changing – for the better. And the backlash to that change is painful as good, smart people try to rationalise their own failure to be better, to be cleverer, to see the other side for the human beings they are. Finding out that you're not the Rebel Alliance, you're actually part of the Empire and have been all along, is painful.

She links this last sentence to an article called Why Nerd Culture Must Die, which, I don't know, kind of makes me a little more skeptical of all of her protestations that she's exactly as much of a nerd as anyone else and likes nerds and is really working for nerds' best interests. The article repeats that nerds think they're "the Rebel Alliance" but actually are "the Empire". Ha ha! Burn!

You may be wondering whether you missed the part of Star Wars where Darth Vader is so terrified of hurting or offending other people that he stops interacting with anybody and becomes suicidally depressed for years. Finally, Vader mentions this fact in the comments section of a blog about obscure Sith rituals. The brave Rebel Alliance springs into action and gets all of the Coruscant newspapers to publish articles on how Vader is entitled and needs to check his privilege.

I don't know. Maybe this was one of those things that got taken out in the Special Edition?

(Han shot first!)

But there's actually something even creepier going on here which may or may not be intentional.

The Transsexual Empire is a very famous book from the late 1970s subtitled "The Making Of The She-Male" in which feminist activist Janice Raymond argues that transsexuals, despite claiming to be persecuted, form an evil empire dedicated to the reinforcing of patriarchy. It contains delightful passages such as "All transsexuals rape women's bodies by reducing the real female form to an artifact, appropriating this body for themselves". The Transgender Studies Reader says that the book "did not invent anti-transsexual prejudice, but it did more to justify and perpetuate it than perhaps any other book ever written." The response, written by a prominent transgender activist, was titled *The Empire Strikes Back* – an obvious reference to the Star Wars film published around that time.

So the question is – how come various feminists keep independently choosing the Empire as a metaphor for their enemies?

Once again the one-dimensional model of privilege rears its ugly head.

Transsexuals claimed to be suffering. This was a problem, because some of them were transwomen who had started with the male gender role. They had privilege! And they claimed to be suffering! The one-dimensional model of privilege lifts its eyebrows quizzically and emits a "…wha?"

The solution is to deny their suffering. Not only deny their suffering, but accuse them of being out to “rape women’s bodies”. Not only deny their suffering and accuse them of being rapists, but to insist that they are privileged – no, super-privileged – no, the most privileged – no, a giant all-powerful all-encompassing mass of privilege that controls everything in the world,.

So they became an Empire. How better to drive home the fact that they’re *definitely* powerful and oppressive and definitely *definitely* not suffering? Because if they were suffering, it would mean we *weren’t*.

There’s another word the radical feminists like to use about transsexuals. “It’s aggrieved entitlement,” Lierre Keith tells the New Yorker. “They are so angry that we will not see them as women.” The article continues to explain how “When trans women demand to be accepted as women they are simply exercising another form of male entitlement.”

And sigh, now here come the male nerds and say *they’re* suffering too, not as much as the transpeople but still a nonzero amount of pain! Is there *no end* to people who are not us, suffering in inconvenient ways? They say that when they feel haunted by scrupulosity, that shaming them all the time actually makes the problem *worse*! We need to establish that they’re privileged right away! So how better to rub in the concept of very privileged people than to draw in the old Empire analogy, right? Maybe try the “entitlement” claim again as well? Second time’s the charm!

But let’s be clear. There *is* a Star Wars metaphor to be made here.

Chancellor Palpatine is, by universal agreement, a great guy. According to Count Dooku, he “speaks honestly and champions the underprivileged” (direct quote from source).

But sometimes people get in the way of his mission of helping the underprivileged, and then he has to, you know, tell it like it is.

Like the Senate. When the Senate is not sure they want to hand over power to the Chancellor, he declares that they are corrupt and oppose democracy.

Or the Jedi. When the Jedi resist his rule, he declares that they are obsessed with “gain[ing] power” and “if they are not all destroyed, it will be civil war without end.”

Whenever he wants to steamroll over someone, Palpatine’s modus operandi is to convince everyone that they are scary oppressors. This isn’t just my personal interpretation. Indeed, in Order 66, Palpatine says straight out:

> “Beings believe what you tell them. They never check, they never ask, they never think…Tell them you can save them, and they will never ask—from what, from whom? Just say tyranny, oppression, vague bogeymen.”

If we’ve learned anything from the Star Wars prequels, it’s that Anakin Skywalker is unbearably annoying. But if we’ve learned *two* things from the Star Wars prequels, it’s that the easiest way to marginalize the legitimate concerns of anyone who stands in your way is to declare them oppressors loud enough to scare everyone who listens.

And if the people in the Star Wars universe had seen the *Star Wars* movies, I have no doubt whatsoever that Chancellor Palpatine would have discredited his opponents by saying they were the Empire.

(seriously, you wanted to throw the gauntlet down to lonely male nerds, and the turf you chose was *Star Wars* *metaphors*? HOW COULD THAT POSSIBLY SEEM LIKE A GOOD IDEA?)

VIII.

> Unlike Aaronson, I was also female, so when I tried to pull myself out of that hell into a life of the mind, I found sexism standing in my way. I am still punished every day by men who believe that I do not deserve my work as a writer and scholar. Some escape it's turned out to be.
>
> Science is a way that shy, nerdy men pull themselves out of the horror of their teenage years. That is true. That is so. But shy, nerdy women have to try to pull themselves out of that same horror into a world that hates, fears and resents them because they are women
>
> Scott, imagine what it's like to have all the problems you had and then putting up with structural misogyny on top of that.

Ms. Penny believes that, as a woman, she's been unfairly excluded from the life of the mind and, indeed, from every pursuit she might enjoy or use as an escape.

There is something to be discussed here, but I am having trouble isolating Ms. Penny's exact claim.

"Unfairly excluded from the life of the mind" might suggest she didn't have the same opportunities as men to participate in higher education, but in fact women are now 33% more likely than men to earn college degrees and women get higher grades in college than men do. They also get well above half of all master's degrees, and just a slice over half of all Ph.Ds (and rising). Their likelihood of becoming professors is nicely predicted by the percent of degrees they earn at a couple decade interval. The articles about the world of higher education now all have titles like Missing Men or Why Are Men Falling Behind.

Industry isn't a good example here either. Women in her demographic group – twenty-something and childless – out-earn their male counterparts by almost ten cents on the dollar.

And she's probably not talking about science, since women earn 55% of science degrees nowadays. They are somewhat overrepresented even in some "hard" sciences like biology, but *overwhelmingly* so in the social sciences. Over seventy five percent of psychology majors are female – a disproportionate which blows out of the water the comparatively miniscule 60-40 disproportion favoring men in mathematics.

(Hi! Male psychology major here, can confirm!)

When Penny says she as a woman is being pushed down and excluded from every opportunity in academic life, she means that women in a very small subset of subjects centered around computer science and engineering face a gender imbalance about as bad as men do in another collection of subjects such as psychology and education.

Penny attacks nerds for believing that "holding men to account for the lack of representation of women in STEM areas…is somehow unfair." Fine. I hold her to account for the even higher imbalance in favor of women in psychology and education. Once she accepts responsibility for that, I'll accept responsibility for hers. That sounds *extremely* fair.

("But that's because of patriarchy!" READ SECTION V.)

I propose an alternate explanation to both dilemmas.

By late high school, the gap between men and women in math and programming is already as large as it will ever be. Yes, it’s true that only 20 – 23% of tech workers are women. But *less than* twenty percent of high school students who choose to the AP Computer Science test are women.

Nothing that happens between twelfth grade and death decreases the percent of women interested in computer science one whit.

I have no hard numbers on anything before high school, but from anecdotal evidence I know very very many young men who were programming BASIC on their dad’s old computer in elementary school, and only a tiny handful of young women who were doing the same.

I don’t want to get into a drawn out inborn-ability versus acculutration fight here. I want to say that I want to say that whether we attribute this to inborn ability *or* to acculturation, *the entire gender gap has been determined in high school if not before*. If anything, women actually gain a few percentage points as they enter Silicon Valley.

What the heck do high schoolers know about whether Silicon Valley culture is sexist or not? Even if you admit that all the online articles talking about this are being read by fourteen year olds in between *Harry Potter* and *Twilight*, these articles are a very new phenomenon and my stats are older than they are. Are you saying the is because of a high level of penetration of rumors about “toxic brogrammers” into the world of the average 11th grader?

The entire case for Silicon Valley misogyny driving women out of tech is a giant *post hoc ergo propter hoc*.

What’s worse, I have never heard any feminist give this case in anything like a principled way. The explanation is usually just something like *of course* men would use their privilege to guard a well-paying and socially prestigious field like programming from women, men have *always* guarded their privileges, they've never given anything up to women without a fight, etc.

My own field is medicine. More than half of medical students are female. In two years, more than half of doctors in the UK will be female, and the US is close behind.

Medicine is better-paying and more prestigious than programming. It's also terrible. Medicine is full of extremely abrasive personalities. Medicine has long work hours. Medicine will laugh at you hysterically if you say you want to balance work and family life.

But women can't get into medicine fast enough. Every so often medical journals and the popular news run scare stories about how there are so many women in medicine now that if they take off time to raise kids at their accustomed rates we're suddenly going to find ourselves pretty much doctorless.

So any explanation of the low number of women in Silicon Valley has to equally well explain their comparatively high numbers in medicine.

Given all this, it's really easy for me to see why it's tempting to blame nerds. Look at these low-status people. It's their fault. We already dislike them, now we have an even better reason to dislike them that nicely wraps up an otherwise embarassing mystery. They're clearly repelling women with their rapey creepishness. It doesn't hurt that occasional high profile stories of sexual harassment come out of Silicon Valley aren't hard to find and bring viral.

(no one ever asks whether there are an equally high number of stories of sexual harassment in medicine – or law, or any other field – that no one had a reason to publicize. When I was in medical school, there was an extremely creepy incident of sexual harassment/borderline attempted rape involving a female medical student and male doctor at an outlying hospital where I worked. Nobody put it on the front page of Gawker, because the doctor involved wasn't a nerd and no one feels any particular need to tar all doctors as sexist.)

But again, you really can't blame this one on Silicon Valley nerds, unless they are breaking into high schools and harassing the women there. And possibly breaking into grade schools, demanding the young boys start tinkering with BASIC. Time for a better theory.

A look at percent female physicians by subspecialty is instructive. The specialty with the most women is pediatrics, followed by child psychiatry, followed by obstetrics, followed by – you get the picture. The specialties with the least women are the various surgeries – the ones where your patient is immobilized, anaesthetized, opened up, and turned into a not-quite-color-coded collection of tubes and wires to poke and prod at – the ones that bear more than a passing resemblance to engineering.

(surgeons are the jockiest jocks ever to jock, so you can't blame us for this one)

It seems really obvious to me that women – in high schools and everywhere else – have a statistical predilection to like working with people (especially children) and to dislike working with abstract technical poking and prodding. This is a bias clearly inculcated well before SATs and AP exams, one that affects medics and programmers alike.

It’s a bias that probably has both cultural and biological origins. The cultural origins are far too varied to enumerate. Many people very justly bring up the issue of how our society genders toys, with parents getting very angry when girls play with stereotypically male toys and vice versa. The classic example is of course the talking Barbie who would famously say “Math is hard! Let’s go shopping!”

On the other hand, I also think people who neglect biological causes are doing the issue a disservice. Did you know that young monkeys express pretty much exactly the same gendered toy preferences as human children? Rhesus monkeys, vervet monkeys, pretty much whatever species of monkeys you try it on, the male monkeys enjoy wheeled toys more and the female monkeys plush toys more. The word reviewers use to describe the magnitude of the result is “overwhelming”. When intersex children are raised as other than their biological gender, their toy preference and behavior are consistently that associated with their biological gender and not the gender they are being raised as, even when they themselves are unaware their biological gender is different. This occurs even when parents reinforce them more for playing with their gender-being-raised-as toys. You can even successfully correlate the degree of this with the precise amount of androgen they get in the womb, and if you experimentally manipulate the amount of hormones monkeys receive in the womb, their gendered play will change accordingly. 2D:4D ratio, a level of how much testosterone is released during a crucial developmental period, accurately predicts scores both on a UK test of mathematical ability at age seven and the SATs in high school.

The end result of all this is probably our old friend gene-culture interaction, where certain small innate differences become ossified into social roles that then magnify the differences immensely. As a result, high school girls are only a fifth as likely to be interested in computer science as high school boys, and sure enough women are only a fifth as well represented in Silicon Valley as men.

All of this information is accessible for free to anyone who spends ten minutes doing a basic Google search. But instead we have to *keep hearing* how nerds are gross and disgusting and entitled and should feel constant shame for how they bully and harass the poor female programmers out of every industry they participate in. Penny blames nerds for not “holding men to account for the lack of representation of women in STEM areas” but SERIOUSLY WE DIDN’T DO IT.

(except insofar as we helped acculturate kids. But that’s hardly a uniquely male pasttime.)

(before you bring up that one paper that showed research leaders advantaged male over female researchers, keep in mind that first of all it explains only a small portion of the discrepancy, and second of all the female research leaders showed the bias even worse than the male ones. Yet Penny frames her question as “holding men to account”. This is that motte-and-bailey thing with patriarchy again.)

Do you realize how unpleasant it is to be *constantly* blamed all the time for something we didn’t do, and have that be used to justify every form of insult and discrimination and accusation against us? The oldest pattern in human history is “Here’s a problem. And here’s a bunch of people who are different than us. Let’s blame it on them!”

There’s enough information out there to prove that creepy nerds are not the problem with female representation in STEM. Then again, there’s also enough information out there to prove that gay people don’t cause earthquakes. People will believe what they want to believe.

§

On the other hand, I’ve said above that I don’t like completely ignoring the accounts of thousands of people who say there’s a problem. Although my female friends in computer science keep insisting they’ve never encountered sexism there, many many others say they have.

But let’s keep our causal arrows pointing the right direction. Any space with a four-to-one male:female ratio is going to end up with some pretty desperate people and a whole lot of unwanted attention. Add into this mix the fact that nerds usually have poor social skills (explaining exactly why would take a literature review to put that last one to shame, but hopefully everyone can agree this is true), and you get people who are pretty sure they are supposed to do something but have no idea what. Err to one side and you get the overly-chivalrous people saying m’lady because it pattern matches to the most courtly and least sexual way of presenting themselves they can think of. Err to the other, and you get people hollowly imitating the behavior they see in famous seducers and playboys, which when done without the very finely-tuned social graces and body-language-reading-ability of famous seducers and playboys is pretty much just “being extremely creepy”.

But once you accept this model, it starts to look like feminists and I are trying to solve the same problem.

The problem is that nerds are scared and confused and feel lonely and have no idea how to approach women. From this root problem blossoms both Aaronson’s problem – that sometimes all you can do is go to a psychiatrist and ask to be castrated – and Penny’s problem – that other times people go read pickup artistry books that promise to tell them how the secret is “negging” people.

But Aaronson's solution to the problem is to talk about it. And feminism's solution to the problem is to swarm anyone who talks about it, beat them into submission, and tell them, in the words of Marcotte, that they are "yalping entitlement combined with an aggressive unwillingness to accept that women are human beings just like men"

IX.

Every article about male nerds calls us "entitled".

I'm pretty sure they don't mean financially, since nerds for example give disproportionately more to charity than other groups (see: Bill Gates, the joke in the effective altruist movement that it contains "all kinds of people – mathematicians, economists, philosophers, *and* computer scientists").

And I'm pretty sure they don't mean politically, since nerds are far more likely to support wealth redistribution than the general population (compare political alignment here to your choice of nationwide poll).

And I'm pretty sure they don't mean psychologically. In psychology, entitlement as a construct is usually blended with narcissism. Predictors of narcissism include high emotional intelligence, high social skills but (uniquely among Dark Triad traits) not high nonverbal (ie mathematical) intelligence, and high extraversion. Another interesting fact about narcissists is that they tend to have more sexual partners than non-narcissists. Jonason describes the research on narcissism and sex by saying that "Narcissists find it easy to start new relationships but are less committed to and interested in staying in existing relationships." I feel like even feminists should be able to agree that "extraverted people with excellent social skills but no particular mathematical aptitude who find it easy to start new relationships" is not a perfect match for nerds here.

So I don't think these articles are talking about entitlement full stop. I guess they're using this to point solely at *sexual* entitlement. But even this seems to require further clarification.

Do they mean nerds hold sexist attitudes? The research (1, 2, 3, 4) shows that sexist attitudes are best predicted by low levels of education, high levels of religious belief, and (whites only) low neuroticism. Once again, I don't feel it should be controversial to say that "very religious people who drop out of school early and are psychologically completely healthy" is not how most people would describe nerds. Besides, in a survey I did of 1500 people on an incredibly nerdy forum last year, the average was *extremely* feminist, so much so that the average nerdy man was more feminist than the average non-nerdy woman.

Do they mean nerds are more likely to rape people? There is an appropriate caveat here that it is difficult-to-impossible to profile rapists – but if people took that caveat seriously then you couldn't profile nerds as rapists either. Since we're already talking about profiling, let's go all the way and find that the best research about rapists (source: David Lisak) does find various characteristics of undetected campus rapists (ie primarily date rapists who get away with it, we're not just talking about scary felons with knives here as a red herring). Some of these are purely psychological ("they're sexist and don't like women"). But the rest include: rapists are more sexually active and "engage in consensual and coercive sex far more often than is typical for men of their age group". They are members of "sexually violent subcultures" including "fraternities and gangs". They are "hypermasculine" and "strive always to behave in rigidly and stereotypically masculine ways” They are heavy drinkers, often using alcohol to release either their own inhibitions or those of their victims.

Once again, I feel like “hypermasculine frat boys and gangsters who party too hard and have a large number of partners” is a really poor description of nerds.

When people talk about nerds feeling “sexually entitled”, it’s never about any of these things. It’s always the same: A male nerd has dared to express that he is sad about being alone and miserable. Then they round this off to “therefore he believes everyone else owes him sex because he is so great” in precisely the way Amanda Marcotte does explicitly and Penny allows to lie beneath the surface.

Once again, Scott Aaronson’s entire problem was that he was so unwilling to hurt women even unintentionally, and so unclear about what the rules were for hurting women, that he erred on the side of super-ultra-caution and tried to force himself never to have any sexual interest in women at all even to the point of trying to get himself castrated. If entitlement means “I don’t care about women’s feelings, I just care about my own need for sex”, Aaronson is the perfect one hundred eighty degree opposite of entitlement. He is just about the most unentitled (untitled?) person imaginable.

Yet Aaronson is the example upon which these columnists have decided their case for “nerd entitlement” must rise and fall. You have better examples? *Then why didn’t you use them?*

I’ve already admitted that when a girl asked me out in middle school, I ran away terrified because I figured nobody could actually like me and it was obviously some kind of nasty trick. If entitlement means "believing you deserve *all* the sex", then teenage-me also sounds pretty untitled.

Yet I, too, get to forever read articles about how entitled I am.

I'm not making some kind of #NotAllNerds statement here, any more than someone who disagrees with the claim "elephants are tiny" is claiming #NotAllElephants

A better word for this untitlement is, perhaps, scrupulosity, where you believe you are uniquely terrible and deserve nothing. Scrupulosity is often linked to obsessive compulsive disorder, which the recent survey suggests nerds have at higher rates than the general population and which is known to be more common in high-IQ people. When I hear my utilitarian friends say things like "I have money and people starving in Africa don't have money, therefore I am morally obligated to give half of my money to people starving in Africa or else their starvation is my fault" *and then actually go and do that* – and trust me, these people are *always* nerds – then as often as not it's scrupulosity at work.

When you tell a highly-untitled, high-scrupulosity person that they are entitled, it goes about as well as telling an anorexic person that they are fat.

If your excuse is going to be "okay, some nerds are overly scrupulous, but others are entitled", *how come that wasn't your argument before?* And how come, with laser-like focus, *you only pick on the scrupulous ones?* How come it's 2015 and we still can't agree that it's not okay to take a group who's already being bullied and harassed, stereotype it based on the characteristics of its worst members, and then write sweeping articles declaring that the entire group is like that?

X.

When Laurie Penny writes to women, she says:

> What I most wanted to say, to all the messed-up teenagers and angry adults out there, is that the fight for your survival is political. The fight to own your emotions, your rage and pain and lust and fear, all those unspeakable secrets that we do not share because we worry that we will be hurt or shunned, is deeply political.

When Laurie Penny writes to men, she says:

> Most of all, we're going to have to make like Princess Elsa and let it go – all that resentment. All that rage and entitlement and hurt.

Clearly this second suggestion contains a non-standard use of the word "we".

When women feel like they're not allowed to "own their emotions" like "lust", or have "secrets that they do not share because they worry that they will be hurt or shunned", then it is "deeply political" and they have to "fight about it."

When men make the same complaint, they are encouraged to "let go" of their "resentment" and "entitlement".

The same worries, deep and secret fears, that are the core and driving heat of Penny's feminism when they happen to women get called "entitlement" when they happen in men and need to be "let go". You're not allowed to complain about them. You're not even allowed to ask the people hurting you to stop – then you're *super* entitled. You shut up and get on with your life.

But it’s actually much worse than that. If you remember only one thing from this entire post, ~~remember that Anakin Skywalker is unbearably annoying~~ remember this:

The past is over. I do not hold, and have never held, any ill will toward the women who rejected me. Some of them continue to be my close friends. Some of them I’ve talked to about this Scott Aaronson thing, *and even they agree with me on it*. Nor did Aaronson mention any ill will to anyone who rejected him. Talking about how nerds should let go of our past resentment to our crushes is a giant red herring.

What this entire discussion is about is our very present resentment toward the (some) feminists who continue to perpetuate the stereotypes that hurt us then, continue to attack us now whenever we talk about the experience or ask them to stop, and continue to come up with rationalizations for why they don’t have to stop. This isn’t about little Caitlin who wouldn’t return my eye contact in seventh grade, this is about Amanda Marcotte, Jezebel, Gawker, and an entire system that gets its jollies by mocking us and trying to twist the knife.

The only reason little Caitlin is being brought up is so that feminists who don’t want to stop twisting can sidestep any criticism by pretending our argument is entirely how a seventh-grader shouldn’t have control of her own romantic decisions.

@#!$ that. Little Caitlin can do what she wants with her life. But dehumanizing and perpetrating stereotypes about a whole group of people who already have it pretty bad is not okay.

XI.

I already know that there are people reading this planning to write responses with titles like “Entitled Blogger Says All Women Exist For His Personal Sexual Pleasure, Also Men Are More Oppressed Than Women, Also Nerds Are More Oppressed Than WWII Era Jews”. And this post is way too long for most people who read those responses to get their misconceptions corrected. So before I close, let me give a brief summary of what I am trying to say:

1. There are a lot of really nasty stereotypes perpetuated about nerds, especially regarding how they are monsters, nobody can love them, and they are too disgusting to have relationships the same way other people do.

2. Although both men and women suffer from these stereotypes, men really do have a harder time getting relationships, and the experience is not the same.

3. Many of the people suffering from these stereotypes are in agreement that it is often self-identified feminists who push them most ardently, and that a small but vocal contingent of feminists seem to take special delight in making nerds’ lives worse.

4. You cannot define this problem away with the word “patriarchy”.

5. You cannot define this problem away by saying that because Mark Zuckerberg is a billionnaire, nerds are privileged, so they already have it too good. The Jews are a classic example of a group that were both economically advantaged in a particular industry, but also faced unfair stereotypes.

6. Whether women also have problems, and whether their problems are even worse, is not the point under discussion and is not relevant. Women can have a bunch of problems,

but that doesn’t mean it is okay for any feminists to shame and bully nerds.

7. Nerds are not uniquely evil, they are not especially engaged in oppressing women, and they are not driving women out of Silicon Valley. Even if they were, “whenever they choose to open up about their private suffering” is not the time to talk about these things.

8. “Entitlement” is a uniquely bizarre insult to level at nerds given that by most of the term’s usual definitions nerds are some of the most untitled people there are.

9. The feminist problem of nerds being desperate and not having any social skills (and therefore being creeps to women) is the same as the nerd problem of nerds being desperate and not having any social skills (and therefore having to live their life desperate and without social skills). Denying the problem and yelling at nerds who talk about it doesn’t help either group.

10. The nerd complaint on this issue is not “high school girls rejected us in the past when we were lonely and desperate,” it is “some feminists are shaming us about our loneliness and desperation in the past and present and openly discussing how they plan to do so in the future.” Nobody with principles is angry at the girls who rejected them in the past and this is a giant red herring. If you don’t believe any feminists are shaming anyone, then say so; don’t make it about little Caitlin in seventh-grade.

If you want to debate or fisk this article, I would recommend using these paragraphs as starting points instead of whatever bizarre perversions of my words the brain of the worst person reading this can dream up.

[EDIT 1/15: Okay, it looks like the talking point people chose to go with was "he made a 1984 joke, therefore the thesis of the essay is that all men are oppressed by all women exactly as badly as people are oppressed in 1984." As usual, I was insufficiently pessimistic.]

XII.

Penny ends:

> We bring our broken hearts and blue balls to the table when we talk gender politics, especially if we are straight folks. Consent and the boundaries of consent – desire and what we're allowed to speak of desire – we're going to have to get better, braver and more honest, we're going to have to undo decades of toxic socialisation and learn to speak to each other as human beings in double quick time.
>
> […]
>
> The road ahead will be long. I believe in you. I believe in all of us. Nerds are brilliant. We are great at learning stuff. We can do anything we put our minds to, although I suspect this thing, this refusing to let the trauma of nerdolescence create more violence, this will be hardest of all.

I see a vision here of everybody, nerdy men, nerdy women, feminists, the media, whoever – cooperating to solve our mutual problems and treat each other with respect. Of course I am on board with this vision. As Scott Aaronson would put it, I am 97% on board. What keeps me from being 100% on board right now is the feeling that the other side *still doesn't get it*.

First of all, a whole lot of other side is not Laurie Penny. They are the people gleefully mocking our pain and telling us we deserve it. But even the good people are worrisome enough.

They admit that nerdy men, lesbians, bisexuals, etc may be in pain, but they deny categorically any possible role of feminist shaming culture in causing that pain and want to take any self-reflection on their part off of the table of potential compromise.

They admit that our pain technically exists, but they are unable to acknowledge it without adding "…but by the way, your pain can't possibly ever be as bad as our pain" or "your pain doesn't qualify for this ontologically distinct category of pain which is much more important."

They continue to think it is appropriate to respond to any complaint or expression of suffering on our part with accusations of "entitlement", comparisons to Darth Vader, and empirically-contradicted slanders about how our mere presence drives women away from everything we love.

Once I see anyone, anywhere, publish an article that not only recognizes our pain, but doesn't derail it into an explanation of why we're definitely still terrible and there is no need whatsoever for them to change, then I will be more optimistic that progress is at hand.

XIII.

Oh *frick*.

> And on that note I shall return to what I was doing before I read this post, which was drinking sweet tea and weeping about how boys don't seem to want to kiss short-haired lady nerds, and trying not to blame the whole world for my broken heart, which is becoming

> more complex and interesting in the healing but still stings like a boiling ball of papercuts. I'll let you know how that goes.

Having so much fun picking this article apart, and then *this* □

Look. I mean what I say about how I don't believe in zero-sum games. The reality of Prof. Aaronson's problem does not for one second diminish the reality of Ms. Penny's sadness as well.

So here is my offer to Ms. Penny. If she accepts and is in some kind of heavily nerd-populated city (NYC? SF?) I will use my connections in the nerd community to get her ten dates within ten days with intelligent, kind, respectful nerdy men of whom she approves. If she is in some less populated place, I will get her some lesser but still non-zero number of dates (unless she's in Greenland or somewhere, in which case she's on her own).

If I can't do that, she may feel welcome to publically mock me and tell me that I was overconfident about how many people are, in fact, extremely willing to kiss short-haired lady nerds.

The rest of this article was serious, but this is extra serious. Let me know.

[EDIT: Comments are now closed, because this got linked on Instapundit and I know from experience that bad things happen if you leave the comments open after that point. Also, my comment software starts acting weird after like a thousand. If you must comment on this further, go bother Ozy on their open thread]. If you're named in this article and you want to rebut it or reply, email me and I'll include it somewhere.

2014 Predictions: Calibration Results

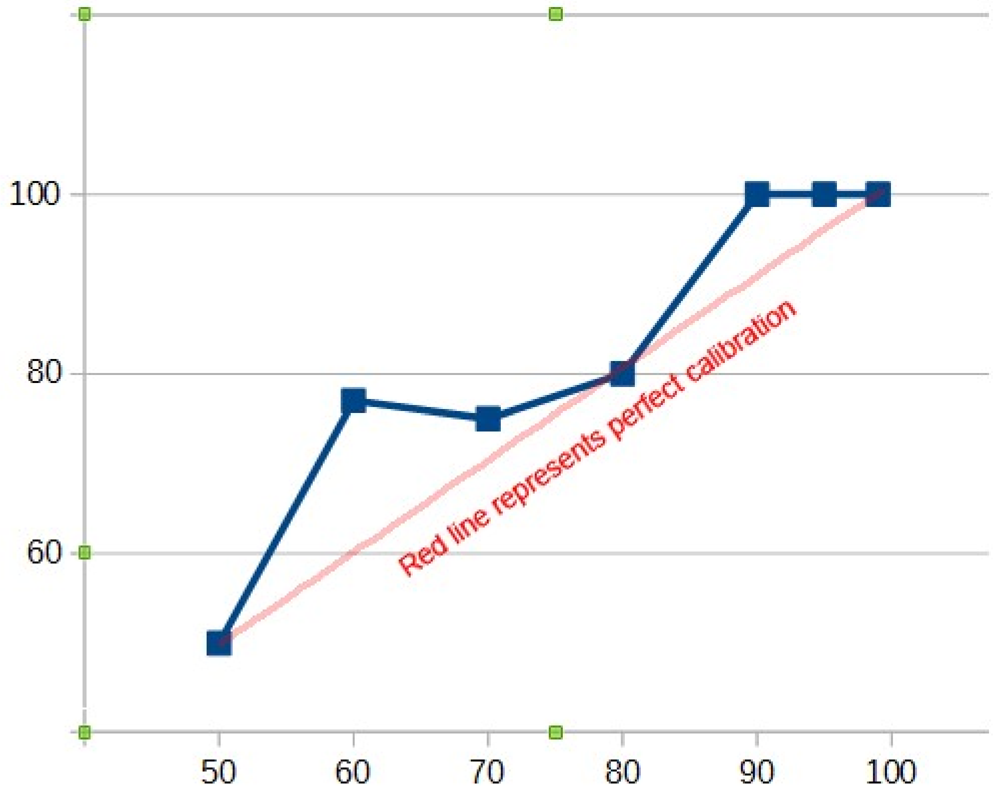
Last year in January I made some predictions for 2014 and gave my calibration level for each. Time to see how I did. Ones in black were successes, ones in red were failures.

1. Obamacare will survive the year mostly intact: 80%
2. US does not get involved in any new major war with death toll of > 100 US soldiers: 90%
3. Syria's civil war will not end this year (unstable cease-fire doesn't count as "end"): 60%
4. Bitcoin will end the year higher than $1000: 70%
5. US official unemployment rate will be < 7% in Dec 2014: 80% 6. Republicans will keep the House in US midterm elections: 80% 7. Democrats will keep the Senate in US midterm elections: 60%
8. North Korea's government will survive the year without large civil war/revolt: 80%
9. Iraq's government will survive the year without large civil war/revolt: 60%
10. China's government will survive the year without large civil war/revolt: 99%
11. US government will survive the year without large civil war/revolt: 99%
12. Egypt's government will survive the year without large civil war/revolt: 60%
13. Israel-Palestine negotiations remain blocked, no obvious plan for Palestinian independence: 99%
14. Israel does not get in a large-scale war (ie >100 Israeli deaths) with any Arab state: 95%
15. Sochi Olympics will not be obvious disaster (ie spectacular violence, protests that utterly overshadow events, or have to be stopped early): 90%
16. Putin will remain President of Russia at end of 2014: 95%

17. Obama will remain President of USA at end of 2014: 95%
18. No nuclear weapon used in anger in 2014: 99%
19. No terrorist attack in USA killing > 100 people: 90%
20. No mass shooting in USA killing > 10 people: 50%
21. Republic of Shireroth will not officially disband or obviously die before end of 2014: 90%
22. Republic of Shireroth will remain in Bastion Union: 80%
23. Active population of Shireroth on last 2014 census will be between 10 and 20: 70%
24. Slate Star Codex will remain active until end of 2014: 70%
25. Slate Star Codex will get more hits in 2014 than in 2013: 60%
26. At least one 2014 Slate Star Codex post will get > 10,000 hits total: 80%
27. No 2014 Slate Star Codex post will get > 100,000 hits total: 90%
28. 2014 Less Wrong Survey will show higher population than 2013 Survey conditional on similar methodology: 80%
29. 2014 Less Wrong Survey will show population < 2000 C.O.S.M.: 70% 30. 2014 Less Wrong Survey will have > % female than 2013 Less Wrong Survey: 70%
31. 2014 Less Wrong Survey will have < 20% female: 90%
32. HPMoR will conclude in 2014: 80%
33. At least 1 LW post > 100 karma in 2014: 50%
34. No LW post > 100 karma by me: 80%
35. CFAR will continue operating in end of 2014: 90%
36. MIRI will continue operating in end of 2014: 99%
37. MetaMed will continue operating in end of 2014: 80%
38. None of Eliezer, Luke, Anna, or Julia will quit their respective organizations: 60%
39. No one in LW community will become world-famous (let’s say >= Peter Thiel) for anything they accomplish this year: 80%
40. MIRI will not announce it is actively working on coding a Friendly AI (not just a few bits and pieces thereof) before the

end of 2014: 99%
41. I will remain at my same job through the end of 2014: 95%
42. I will get a score at >95th percentile for my year on PRITE: 70%
43. I will be involved in at least one published/accepted-to-publish research paper by the end of 2014: 20%
44. I will not break up with any of my current girlfriends through the end of 2014: 50%
45. I will not get any new girlfriends in 2014: 50%
46. I will not be engaged by the end of 2014: 80%
47. I will be living with Ozy by the end of 2014: 80%
48. I will take nootropics on average at least once/week through the second half of 2014: 50%
49. I will not manage to meditate at least 100 days in 2014: 80%
50. I will attend NYC Solstice ritual: 60%
51. I will arrange some kind of Michigan Solstice Ritual: 50%
52. I will not publicly identify as religious (> atheist) by the end of 2014: 95%
53. I will not publicly identify as neoreactionary or conservative (> liberal or libertarian) by the end of 2014: 70%
54. I will not publicly identify as leftist or communist (> liberal or libertarian) by the end of 2014: 80%
55. I will get a Tumblr in 2014: 50%
56. I will not delete/abandon either my Facebook or Twitter accounts: 60%
57. I will have less than 1000 Twitter followers by the end of 2014: 60%
58. When Eliezer sends me a copy of “Perfect Health Diet”, I will not be convinced that it is more correct or useful than the best mainstream nutrition advice (eg Stephen Guyenet’s blog): 70%
59. I will end up being underconfident on these predictions: 50%

Of predictions at the 50% level, 4/8 (50%) were correct
Of predictions at the 60% level, 7/9 (77%) were correct
Of predictions at the 70% level, 6/8 (75%) were correct
Of predictions at the 80% level, 12/15 (80%) were correct
Of predictions at the 90% level, 7-0 (100%) were correct
Of predictions at the 95% level, 5-0 (100%) were correct
Of predictions at the 99% level, 6-0 (100%) were correct



I declare myself to be *impressively* well-calibrated. You should all trust me about everything.

2015 predictions coming soon.

Ethics Offsets

I.

Some people buy voluntary carbon offsets. Suppose they worry about global warming and would feel bad taking a long unnecessary plane trip that pollutes the atmosphere. So instead of not doing it, they take the plane trip, then pay for some environmental organization to clean up an amount of carbon equal to or greater than the amount of carbon they emitted. They're happy because they got their trip, future generations are happy because the atmosphere is cleaner, everyone wins.

We can generalize this to ethics offsets. Suppose you really want to visit an oppressive dictatorial country so you can see the beautiful tourist sights there. But you worry that by going there and spending money, you're propping up the dictatorship. So you take your trip, but you also donate some money to opposition groups and humanitarian groups opposing the dictatorship and helping its victims, at an amount such that you are confident that the oppressed people of the country would prefer you take both actions (visit + donate) than that you take neither action.

I know I didn't come up with this concept, but I'm having trouble finding out who did, so no link for now.

A recent post, Nobody Is Perfect, Everything Is Commensurable, suggests that if you are averse to activism but still feel you have an obligation to improve the world, you can discharge that obligation by giving to charity. This is not quite an ethics offset – it's not exchanging a transgression for a donation so much as saying that a donation is a better way of helping than the thing you were worried about transgressing against anyway – but it’s certainly pretty similar.

As far as I can tell, the simplest cases here are 100% legit. I can’t imagine anyone saying “You may not take that plane flight you want, even if you donate so much to the environment that in the end it cleans up twice as much carbon dioxide as you produced. You must sit around at home, feeling bored and lonely, and letting the atmosphere be more polluted than if you had made your donation”.

But here are two cases I am less certain about.

II.

Suppose you feel some obligation to be a vegetarian – either because you believe animal suffering is bad, or you have enough moral uncertainty around the topic for the ethical calculus to come out against. Is it acceptable to continue eating animals, but also donate money to animal rights charities?

A simple example: you eat meat, but also donate money to a group lobbying for cage=free eggs. You are confident that if chickens could think and vote, the average chicken would prefer a world in which you did both these things to a world in which you did neither. This seems to me much like the cases above.

A harder example. You eat meat, but also donate money to a group that convinces people to become vegetarian. Jeff Kaufman and Brian Tomasik suggest that about $10 to $50 is enough to make one person become vegetarian for one year by sponsoring what are apparently very convincing advertisements.

Eating meat is definitely worth $1000 per year for me. So if I donate $1000 to vegetarian advertising, then eat meat, I’m helping turn between twenty and a hundred people vegetarian for a year, and helping twenty to one hundred times as many animals as I would be by becoming vegetarian myself. Clearly this is an excellent deal for me and an excellent deal for animals.

But I still can’t help feeling like there’s something really wrong here. It’s not just the low price of convincing people – even if I was 100% guaranteed that the calculations were right, I’d still feel just as weird. Part of it is a sense of duping others – would they be as eager to become vegetarian if they knew the ads that convinced them were sponsored by meat-eaters?

Maybe! Suppose we go to all of the people convinced by the ads, tell them “I paid for that ad that convinced you, and I still eat meat. Now what?” They answer “Well, I double-checked the facts in the ad and they’re all true. That you eat meat doesn’t make anything in the advertisement one bit less convincing. So I’m going to stay vegetarian.” Now what? Am I off the hook?

A second objection: universalizability. If *everyone* decides to solve animal suffering by throwing money at advertisers, there is no one left to advertise to and nothing gets solved. You just end up with a world where 100% of ads on TVs, in newspapers, and online are about becoming vegetarian, and everyone watches them and says “Well, I’m doing my part! I’m paying for these ads!”

Counter-objection: At that point, no one will be able to say with a straight face that every $50 spent on ads converts one person to vegetarianism. If I follow the maxim “Either be vegetarian, or donate enough money to be 90% sure I am converting at least two other people to vegetarianism", this maxim *does* universalize, since after animal suffering ads have saturated a certain percent of the population, no one can be 90% sure of convincing anyone else.

As far as I can tell, this is weird but ethical.

III.

The second troublesome case is a little more gruesome.

Current estimates suggest that $3340 worth of donations to global health causes saves, on average, one life.

Let us be excruciatingly cautious and include a two-order-of-magnitude margin of error. At $334,000, we are *super duper sure* we are saving at least one life.

So. Say I'm a millionaire with a spare $334,000, and there's a guy I *really* don't like…

Okay, fine. Get the irrelevant objections out of the way first and establish the least convenient possible world. I'm a criminal mastermind, it'll be the perfect crime, and there's zero chance I'll go to jail. I can make it look completely natural, like a heart attack or something, so I'm not going to terrorize the city or waste police time and resources. The guy's not supporting a family and doesn't have any friends who will be heartbroken at his death. There's no political aspect to my grudge, so this isn't going to silence the enemies of the rich or anything like that. I myself have a terminal disease, and so the damage that I inflict upon my own soul with the act – or however it is Leah always phrases it – will perish with me immediately afterwards. There is no God, or if there is one He respects ethics offsets when you get to the Pearly Gates.

Or you know what? Don’t get the irrelevant objections out of the way. We can offset those too. The police will waste a lot of time investigating the murder? Maybe I’m *very* rich and I can make a big anonymous donation to the local police force that will more than compensate them for their trouble and allow them to hire extra officers to take up the slack. The local citizens will be scared there’s a killer on the loose? They’ll forget all about it once they learn taxes have been cut to zero percent thanks to an anonymous donation to the city government from a local tycoon.

Even what seems to me the most desperate and problematic objection – that maybe the malarial Africans saved by global health charities have lives that are in some qualitative way just not as valuable as those of happy First World citizens contributing to the global economy – can be fixed. If I’ve got enough money, a few hundred thousand to a million ought to be able to save the life of a local person in no way distinguishable from my victim. Heck, since this is a hypothetical problem and I have infinite money, why not save *ten* local people?

The best I can do here is to say that I am crossing a Schelling fence which might also be crossed by people who will be less scrupulous in making sure their offsets are in order. But perhaps I could offset that too. Also, we could assume I will never tell anybody. Also, anyone can just go murder someone right now without offsetting, so we’re not exactly talking about a big temptation for the unscrupulous.

Chronic Psychitis

Some people have asked my opinion on a recent spate of articles like Is Depression Partly Caused By An Allergic Reaction? and Depression May Be Caused By Inflammation.

Standard disclaimer: I'm not a researcher in this field, I'm not board-certified as a full psychiatrist yet, and what I remember of biochemistry is limited to being pretty sure there's something called a "Krebs cycle" involved somewhere. That having been said:

This is pretty legit.

Start with From inflammation to sickness and depression, Dantzer et al (2008), who note that being sick makes you feel lousy [citation needed]. Drawing upon evolutionary psychology, they theorize this is an adaptive response to make sick people stay in bed (or cave, or wherever) so the body can focus all of its energy on healing. A lot of sickness behavior – being tired, not wanting to do anything, not eating, not wanting to hang around other people – seems kind of like mini-depression.

All of this stuff is regulated by chemicals called cytokines, which are released by immune cells that have noticed an injury or infection or something. They are often compared to a body-wide "red alert" sending the message "sickness detected, everyone to battle stations". This response is closely linked to the idea of "inflammation", the classic example of which is the locally infected area that has turned red and puffy. Most inflammatory cytokines handle the immune response directly, but a few of them – especially interleukin-1B and tumor necrosis factor alpha – cause this depression-like sickness behavior. It is noted that:

> In general, animals injected with IL-1ß or TNF-a stay in a corner of their home cage in a hunched posture and show little or no interest in their physical and social environment unless they are stimulated. Specifically, they show decreased motor activity, social withdrawal, reduced food and water intake, increased slow-wave sleep and altered cognition

Here are some other suspicious facts about depression and inflammation:

– Exercise, good diet and sleep reduce inflammation; they also help depression.

– Stress increases inflammation and is a known trigger for depression.

– Rates of depression are increasing over time, with the condition seemingly very rare in pre-modern non-Westernized societies. This is commonly attributed to the atomization and hectic pace of modern life. But levels of inflammation are also increasing over time, probably because we have a terrible diet that disrupts the gut microbiota that are supposed to be symbioting with the immune system. Could this be another one of the things we think are social that turn out to be biological?

– SSRI antidepressants, like most medications, have about five zillion effects. One of the effects is to reduce the level of inflammatory cytokines in the body. Is it possible that this is why they work, and all of this stuff about serotonin receptors in the brain is a gigantic red herring?

– It’s always been a very curious piece of trivia that treating depression comorbid with heart disease *significantly* decreases your chances of dying from the heart disease. People just sort of nod their heads and say “You know, mind-body connection”. But inflammation is known to be implicated in cardiovascular disease. If treating depression is a form of lowering inflammation, this would make perfect sense.

– Rates of depression are *much* higher in sick people. Cancer patients are especially famous for this. No one gets too surprised here, because having cancer is hella depressing. But it’s always been interesting (to me at least) that as far as we can tell, antidepressants treat cancer-induced depression just as well as any other type. Are antidepressants *just that good*? Or is the link between cancer being sad and cancer causing depression only part of the story, with the other part being that the body’s immune response to cancer causes inflammatory cytokine release, which antidepressants can help manage?

– Along with cancer, depression is common in many other less immediately emotion-provoking illnesses like rheumatoid arthritis and diabetes. The common thread among these illnesses is inflammation.

– Inflammation changes the activity level of the enzyme indoleamine 2,3 dioxygenase. This enzyme produces kynurenines which interact with the NMDA receptor, a neurotransmitter receptor implicated in depression and various other psychiatric diseases (in case your first question upon learning about this pathway is the same as mine: yes, kynurenines got their name because they were first found in dog urine).

– Sometimes doctors treat diseases like hepatitis by injecting artificial cytokines to make the immune system realize the threat and ramp up into action. Cytokine administration treatments very commonly cause depression as a side effect. This depression can be treated with standard antidepressants.

– Also, it turns out we can just check and people with depression have more cytokines.

There's also some evidence against the theory. People with depression have more cytokines, but it's one of those wishy-washy "Well, if you get a large enough sample size, you'll see a trend" style relationships, rather than "this one weird trick lets you infallibly produce depression".

But for me the strongest evidence against is a general feeling that it's very easy to get lots of convincing evidence for a theory in medicine whether or not it's true.

Twenty years ago, everyone was super-convinced that depression was caused by low serotonin levels. We found that depressed people on average had lower serotonin levels than non-depressed people. We found that giving people drugs that increased serotonin treated depression. We did lots of studies proving serotonin was a vital chemical that regulated mood. We found that genes affecting serotonin-related proteins were linked to depression. We did PET scans that found abnormally high levels of activity in serotonin-related enzymes in the brains of depressed people. It was all very convincing. And right now everyone's pretty sure it's wrong.

Ten years ago, everyone was super-convinced that depression was caused by under-secretion of the neuro-hormone BDNF and subsequent decline in hippocampal neurogenesis. It was dutifully found that depressed people had less BDNF than everyone else, and less hippocampal neurogenesis. Exercise, sleep, good diet, and all the other things that help depression were found to also raise levels of BDNF. Chemical pathways were trotted out by which effective antidepressants would probably raise BDNF levels. I think this theory is still very popular, but for the inflammation theory to be right someone will either have to disprove this one or tie it together with some theory of why inflammation decreases BDNF or low BDNF increases inflammation or something else. I do see some evidence that this is true, but to fully integrate the theories is going to take a lot more than that.

And these are just the two most recent and most famous. We have Freud's psychoanalytic theory of depression, lots of people studying dysregulation of the hypothalamopituitaryadrenocortical axis, some pointers to dysregulation in the second messenger system, et cetera. All of these theories have *great* evidence.

Point is, now we have another theory that neatly explains how depression starts, how antidepressants work, why diet and exercise are good for you, and all the things all the other theories explained. Maybe the third time's the charm?

A lot of the things in the body are really complex. Inflammation definitely affects serotonin – the indoleamine 2,3 dioxygenase enzyme acts on serotonin's immediate precursor. It affects BDNF levels, as above, which in turn affect hippocampal neurogenesis. The hypothalamopituitaryadrenocortical axis releases cortisol, which downregulates the immune system and decreases the action of inflammatory cytokines. All of the anxiety-inducing life events and intrapsychic conflicts and secret desires to marry your mother that Freud thought caused depression produce a lot of stress, which both releases cortisol and reduces normal ability to regulate inflammatory response.

So basically all of these systems are intimately interconnected, and probably before this is done with researchers will find five more systems intimately interconnected with all of these. It might be that inflammation is the master system which causes a cascade of events in all of the others. It might be that one of the others is the master system. It might be that depression is a collection of multiple different diseases, and some are caused by one thing and others by another. It might be that looking for a "master system" is silly and that the true mathematical relationship between all of these things is such a chaotic process that all you can say is that they all stumbled together into the wrong attractor point and things deteriorated from there.

Anyway, all this is for much smarter people than me to figure out. The question I'm most interested in: can we treat depression by giving people anti-inflammatory drugs?

The answer seems to be: it depends how strongly you object to getting a heart attack.

Aspirin is a great anti-inflammatory drug. It's pretty safe in adults (except for a small risk of GI bleeding) and it decreases risk of cardiovascular disease and cancer as well. If you could treat depression with aspirin, you'd be home free. However, the most convincing review I have seen for aspirin is unimpressed. It points out that some trials have shown negative effects for aspirin, and that long-term use of aspirin can increase intestinal permeability which decreases ability to regulation inflammation which is the opposite of what we want. Right now there isn't much evidence on this issue, but what there is isn't promising.

Most researchers have chosen to focus on celecoxib (Celebrex™®©, a high-tech next-generation anti-inflammatory). Here the evidence is actually very strong. Last month's JAMA Psychiatry contained Effect of anti-inflammatory treatment on depression, depressive symptoms, and adverse effects: a systematic review and meta-analysis of randomized clinical trials, Kohler et al, (2014), which analyzes ten studies with a total of 4000 people taking celecoxib and finds an effect size similar to that of SSRI antidepressants. This is promising and exciting.

They add: "We found no evidence of an increased number of gastrointestinal or cardiovascular events after 6 weeks."

That's probably because they didn't wait long enough. Celecoxib is very closely related to the infamous rofecoxib (Vioxx™®©) which got pulled from the market for quadrupling heart attack risk. Celecoxib is safer; it only increases your risk by some smaller amount depending on dose. Studies conflict, but maybe 33% for a standard regimen?

On the other hand, if ten percent of Americans are on SSRIs right now, and there are 1.5 million heart attacks per year in the US, and celecoxib increases that by 33%, then switching everyone from SSRI to celecoxib would cause…quick Fermi calculation…ignore interactions…50,000 extra heart attacks per year. Ouch.

Celecoxib is a good drug for its indicated uses, which involve treating chronic pain conditions that nothing else can treat safely. But it's hardly something I'd want to start giving to every depressed patient who walks into a psychiatrist's office. Maybe as a third line or fourth line drug for desperate people. But then, we already have plenty of good third-line and fourth-line drugs for desperate people. You want strong psychiatric medication and aren't too concerned about the state of your cardiovascular system? Here, have an antipsychotic!

So in conclusion, I think the inflammatory hypothesis of depression is very likely part of the picture. Whether it’s the main part of the picture or just somewhere in the background remains to be seen, but for now it looks encouraging. Anti-inflammatory drugs do seem to treat depression, which is a point in the theory’s favor, but right now the only one that has strong evidence behind it has side effects that make it undesirable for most people. There’s a lot of room to hope that in the future researchers will learn more about exactly how this cytokine thing works and be able to design antidepressant drugs that target the appropriate cytokines directly. Until then, your best bets are the anti-inflammatory mainstays: good diet, good sleep, plenty of exercise, low stress levels, and all the other things we already know work.

The Phatic And The Anti-Inductive

I.

Ozy recently taught me the word “phatic”. It means talking for the sake of talking.

The classic example is small talk. “Hey.” “Hey.” “How are you?” Fine, and you?” “Fine.” No information has been exchanged. Even if the person involved wasn’t fine, they’d still say *fine*. Indeed, at least in this country giving an information-bearing response to “how are you?” is a mild social faux pas.

Some people call this “social grooming behavior” and it makes sense. It’s just a way of saying “Hello, I acknowledge you and still consider you an acquaintance. There’s nothing wrong between us. Carry on.” That you are willing to spend ten seconds holding a useless conversation with them signals this just fine.

We can go a little more complex. Imagine I’m calling a friend from college after five years out of contact; I’ve heard he’s got a company now and I want to ask him for a job. It starts off “Hey, how are you?”, segues into “And how are the wife and kids?”, then maybe into “What are you doing with yourself these days?” and finally “Hey, I have a big favor to ask you.” If you pick up the phone and say “Hello, it’s Scott from college, can you help me get a job?” this is rude. It probably sounds like you’re *using* him.

And I mean, you are. If I cared about him deeply as a person I probably would have called him at some point in the last five years, before I needed something. But by mutual consent we both sweep that under the rug by having a few minutes of meaningless personal conversation beforehand. The information exchanged doesn’t matter – “how’s your business going?” is just as good as “how’s your wife and kids?” is just as good as “how are your parents doing?”. The point is to clock a certain number of minutes about something vaguely personal, so that the request seems less abrupt.

We can go even more complex. By the broadest definition, phatic communication is equivalent to signaling.

Consider a very formulaic conservative radio show. Every week, the host talks about some scandal that liberals have been involved in. Then she explains why it means the country is going to hell. I don’t think the listeners really care that a school in Vermont has banned Christmas decorations or whatever. The point is to convey this vague undercurrent of “Hey, there are other people out there who think like you, we all agree with you, you’re a good person, you can just sit here and listen and feel reassured that you’re right.” Anything vaguely conservative in content will be equally effective, regardless of whether the listener cares about the particular issue.

II.

Douglas Adams once said there was a theory that if anyone ever understood the Universe, it would disappear and be replaced by something even more incomprehensible. He added that there was another theory that this had already happened.

These sorts of things – things such that if you understand them, they get more complicated until you don’t – are called “anti-inductive”.

The classic anti-inductive institution is the stock market. Suppose you found a pattern in the stock market. For example, it always went down on Tuesdays, then up on Wednesdays. Then you could buy lots of stock Tuesday evening, when it was low, and sell it Wednesday, when it was high, and be assured of making free money.

But lots of people want free money, so lots of people will try this plan. There will be so much demand for stock on Tuesday evening that there won't be enough stocks to fill it all. Desperate buyers will bid up the prices. Meanwhile, on Wednesday, everyone will sell their stocks at once, causing a huge glut and making prices go down. This will continue until the trend of low prices Tuesday, high prices Wednesday disappears.

So in general, it should be impossible to exploit your pattern-finding ability to profit of the stock market unless you are the smartest and most resourceful person in the world. That is, maybe stocks go up every time the Fed cuts interest rates, but Goldman Sachs knows that too, so they probably have computers programmed to buy so much stock milliseconds after the interest rate announcement is made that the prices will stabilize on that alone. That means that unless you can predict better than, or respond faster than, Goldman Sachs, you can't exploit your knowledge of this pattern and shouldn't even try.

Here's something I haven't heard described as anti-inductive before: job-seeking.

When I was applying for medical residencies, I asked some people in the field to help me out with my interviewing skills.

"Why did you want to become a doctor?" they asked.

“I want to help people,” I said.

“Oh God,” they answered. “No, anything but that. Nothing says ‘person exactly like every other bright-eyed naive new doctor’ than wanting to help people. You’re trying to distinguish yourself from the pack!”

“Then…uh…I want to hurt people?”

“Okay, tell you what. You have any experience treating people in disaster-prone Third World countries?”

“I worked at a hospital in Haiti after the earthquake there.”

“Perfect. That’s inspirational as hell. Talk about how you want to become a doctor because the people of Haiti taught you so much.”

Wanting to help people is a great reason to become a doctor. When Hippocrates was taking his first students, he was probably really impressed by the one guy who said he wanted to help people. But since that time it’s become cliche, overused. Now it signals people who can’t come up with an original answer. So you need something better.

During my interviews, I talked about my time working in Haiti. I got to talk to some of the other applicants, and they talked about *their* time working in Ethiopia, or Bangladesh, or Nicaragua, or wherever. Apparently the “stand out by working in a disaster-prone Third World country” plan was sufficiently successful that everyone started using, and now the people who do it don’t stand out at all. My interviewer was probably thinking “Oh God, what Third World country is *this* guy going to start blabbering about how much he learned from?” and moving my application to the REJECT pile as soon as I opened my mouth.

I am getting the same vibe from the critiques of OKCupid profiles in the last open thread. OKCupid seems very susceptible to everybody posting identical quirky pictures of themselves rock-climbing, then talking about how fun-loving and down-to-earth they are. On the other hand, every deviation from that medium has *also* been explored.

“I’m going for ‘quirky yet kind'”.

“Done.”

“Sarcastic, yet nerdy?”

“Done.”

“Outdoorsy, yet intellectual.”

“Done.”

“Introverted, yet a zombie.”

“I thought we went over this. Zombies. Are. Super. Done..”

III.

I’ve been thinking about this lately in the context of psychotherapy.

I’m not talking about the very specific therapies, the ones where they teach special cognitive skills, or expose you to spiders to cure your arachnophobia. They don’t let me do those yet. I’m talking about what’s called “supportive therapy”, where you’re just talking to people and trying to make them feel generally better.

When I was first starting out, I tried to do therapy anti-inductively. I figured that I had to come up with something unexpected, something that the patient hadn’t thought of.

Some kind of brilliant interpretation that put all of their problems in a new light. This went poorly. It tended to be a lot of “Well, have you tried [obvious thing?]”, them saying they had, and me escalating to “Well, have you tried [long shot that probably wouldn’t work]?”

(I wonder if this was Freud’s strategy: “Okay, he says he’s depressed, I can’t just tell him to cheer up, probably everybody says that. Can’t just tell him to accept his sadness, that one’s obvious too. Got to come up with something really original…uh…”HAVE YOU CONSIDERED THAT YOU WANT TO KILL YOUR FATHER AND MARRY YOUR MOTHER??!”)

Now I tend more to phatic therapy. This happened kind of by accident. Some manic people have a symptom called “pressured speech” which means they never shut up and they never let you get a word in edgewise. Eventually, more out of surrender than out of a strategic plan, I gave up and stopped trying. I just let them talk, nodded my head, said “Yeah, that sounds bad” when they said something bad-sounding, said “Oh, that’s good” when they said something good-sounding.

After a while I realized this went at least as well as any other therapy I was doing, plus the patients really liked me and thought I was great and gave me lots of compliments.

So after that, “active listening” became sort of my default position for supportive therapy. Get people talking. Let them talk. Nod my head as if I am deeply concerned about their problems. Accept their effusive praise about how well I seem to be understanding them.

This is clearly phatic. I would say the ritual is “High status person is willing to listen to my problems. That means society considers my problems important and considers me important. It means my problems are okay to have and I'm not in trouble for having them." As long as I seem vaguely approving, the ritual reaches its predetermined conclusion.

IV.

I was thinking about this recently several friends have told me how much she hated "therapist speak". You know, things like "I feel your pain" or "And how does that make you feel?"

I interpret this as an anti-inductive perspective on therapy. The first therapist to say "I feel your pain" may have impressed her patients – a person who herself can actually feel all my hurt and anger! Amazing! But this became such a standard in the profession that it became the Default Therapist Response. Now it's a signal of "I care so little about your pain that I can't even bother to say anything other than the default response." When a therapist says "I feel your pain," it's easy to imagine that in her head she's actually planning what she's going to make for dinner or something.

So just as some people find it useful to divide the world into "ask culture" and "guess culture", I am finding it useful to divide the world into "phatic culture" and "anti-inductive culture".

There are people for whom "I feel your pain" is exactly the right response. It shows that you are sticking to your therapist script, it urges them to stick to their patient script, and at the end of the session they feel like the ritual has been completed and they feel better.

There are other people for whom "I feel your pain" is the most enraging thing you could possibly say. It shows that you’re not taking them seriously or engaging with them, just saying exactly the same thing you do to all your other patients.

There are people for whom coming up with some sort of unique perspective or clever solution for their problems is exactly the right response. Even if it doesn’t work, it at least proves that you are thinking hard about what they are saying.

There are other people for whom coming up with some sort of unique perspective or clever solution is the most enraging thing you could possibly do. At the risk of perpetuating gender stereotypes, one of the most frequently repeated pieces of relationship advice I hear is “When a woman is telling you her problems, just listen and sympathize, don’t try to propose solutions”. It sounds like the hypothetical woman in this advice is looking for a phatic answer.

I think myself and most of my friends fall far to the anti-inductive side, with little tolerance for the phatic side. And I think we probably typical-mind other people as doing the same.

This seems related to the classic geek discomfort with small-talk, with pep rallies, and with normal object-level politics. I think it might also be part of the problem I had with social skills when I was younger – I remember talking to people, panicking because I couldn’t think of any way to make the conversation unusually entertaining or enlightening, and feeling like I had been a failure for responding to the boring-weather-related question with a boring-weather-related answer. Very speculatively, I think it might have something to do with creepy romantic overtures – imagine the same mental pattern that made me jokingly consider giving “I want to hurt people” as my motivation for becoming a doctor, applied to a domain that I really don’t understand on a fundamental enough level to know whether or not saying that is a good idea.

I’ve been trying to learn the skill of appreciating the phatic. I used to be very bad at sending out thank-you cards, because I figured if I sent a thank-you card that just said “Thank you for the gift, I really appreciate it” then they would think that the lack of personalization meant I wasn’t *really* thankful. But personalizing a bunch of messages to people I often don’t really know that well is hard and I ended up all miserable. Now I just send out the thank you card with the impersonal message, and most people are like “Oh, it was so nice of you to send me a card, I can tell you really appreciated it.” This seems like an improvement.

As for psychotherapy, I think I’m going to default to phatic in most cases when I don’t have some incredibly enlightening insight, then let my patients tell me if that’s the wrong thing to do.

The Physics Diet?

There are at least four possible positions on the thermodynamics of weight gain:

1. Weight gain does not depend on calories in versus calories out, even in the loosest sense.

2. Weight gain is entirely a function of calories in versus calories out, but calories may move in unexpected ways not linked to the classic “eat” and “exercise” dichotomy. For example, some people may have “fast metabolisms” which burn calories even when they are not exercising. These people may stay very thin even if they eat and exercise as much as much more obese people.

3. Weight gain is entirely a function of calories in versus calories out, and therefore of how much you eat and exercise. However, these are in turn mostly dependent on the set points of a biologically-based drive. For example, some people may have overactive appetites, and feel starving unless they eat an amount of food that will make them fat. Other people will have very strong exercise drives and feel fidgety unless they get enough exercise to keep them very thin. These things can be altered in various ways which cause weight gain or loss, without the subject exerting willpower. For example, sleep may cause weight loss because people who get a good night sleep have decreased appetite and lower levels of appetite-related hormones.

4. Weight gain is entirely a function of calories in versus calories out, and therefore of how much you eat and exercise. That means diet is entirely a function of willpower and any claim that factors other than amount of food eaten

and amount of exercise performed can affect weight gain is ipso facto ridiculous. For example, we can dismiss claims that getting a good night's sleep helps weight loss, because that would violate the laws of thermodynamics.

1 and 4 are kind of dumb. 1 is dumb because…well, to steal an Eddington quote originally supposed apply to the *second* law of thermodynamics:

> If someone points out to you that your pet theory of the universe is in disagreement with Maxwell's equations — then so much the worse for Maxwell's equations. If it is found to be contradicted by observation — well, these experimentalists do bungle things sometimes. But if your theory is found to be against…thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation.

But 4 is also dumb. We have a long list of things that affect weight gain – for example, patients on the powerful psychiatric medication clozapine usually gain a lot of weight – fifteen pounds more on average than people on safer antipsychotics. Other medications are known to increase weight to a lesser degree, and some medications even decrease weight, though you wouldn't like the side effects of most of them. Certain genetic diseases are also known to cause increased weight – Prader-Willi syndrome, for example.

One could try to rescue 4 by saying that people with rare genetic diseases or taking powerful prescription-only medications are a different story and in normal people it's entirely controlled by willpower. But first, this is an area where possibility proofs are half the battle, and we have a possibility proof. And second, there are more than enough studies about genetics, microbiome, and, yes, sleep showing that all of these things can have effects in normal people.

So 1 and 4 are out. And although I do sometimes see people pushing them, they *mostly* seem to do a thriving business as straw men for people who want to accuse their opponents of saying something absurd.

The most interesting debate to be had is between 2 and 3. 3 says that all of the interventions that we know affect weight – certain pills, certain recreational drugs, changes in gut bacteria, whatever – do it by affecting appetite and exercise drive. 2 says that basal metabolism is also involved. 3 seems to at least leave open the possibility of just starving yourself even when your body is telling you really hard to eat. 2 says even that won't work.

There's room for a little bit of gradation between 2 and 3. A lot of people suggest that one way "fast metabolism" presents is by people fidgeting a lot, which is sort of the same as "your body increases its exercise drive".

But in general, I think 2 is an important issue that does cause at least some interpersonal weight differences.

We'll start with the "possibility proof" again. MRAP2. It's a gene. Scientists can delete it in mice. These mice will eventually develop excessive appetites. But when they are young, they eat the same amount as any other mouse, but still get fatter.

Likewise, 2,4-dinitrophenol is a cellular uncoupling agent which increases metabolic rate and consistently produces weight loss of 2-3 pounds per week. It would be an excellent solution to all of our obesity-related problems if the papers on it didn't keep having names like 2,4-Dinitrophenol: A

Weight Loss Agent With Significant Acute Toxicity And Risk Of Death.

So what about everyday life?

A study of individual variation in basal metabolic rate found very significant interpersonal differences. A lot of that was just “some people are bigger than others”, but some of it wasn’t – they state that “twenty-six percent of the variance remained unexplained”. The Wikipedia article puts this in context: “One study reported an extreme case where two individuals with the same lean body mass of 43 kg had BMRs of 1075 kcal/day (4.5 MJ/day) and 1790 kcal/day (7.5 MJ/day). This difference of 715 kcal/day (67%) is equivalent to one of the individuals completing a 10 kilometer run every day”

Dr. Claude Bouchard and his team stuck 12 pairs of male identical twins in isolation chambers where their caloric intake and exercise could be carefully controlled, then fed them more calories than their bodies needed. All sets of twins gained weight, and in all twin groups both twins gained about the same amount of weight as each other, but the amount of weight gained varied between twin pairs by a factor of 3 (from 4 to 13 kg).

A lot of the sites that talk about this thing are careful to say that people “can’t blame” genes for their obesity, because obesity levels have been rising for decades and genes can’t change that quickly. I think this is wrong-headed. True, genes are not the source of the modern rise in obesity levels. But it’s entirely possible that a globally rising tide of obesity has disproportionately affected the people with the wrong genes. Just as Bouchard fed the same amount extra to all his study participants but some of them gained more weight than others, so if you put an entire civilization worth of people in an obesogenic environment, some of them might be genetically predisposed to do worse than the rest.

A more practical question – can individual people's metabolism change?

I am personally predisposed to answer in the affirmative. In my early twenties, I ate a *crazy* amount every day – two bagels with breakfast, cookies with lunch, a big dinner followed by dessert – and I stayed pretty thin throughout. Now I'm thirty, I eat a very restrained diet, and my weight still hovers at just above the range where I am supposed to be. I know that people are famously bad at understanding how much they're eating and exercising, but seriously if you try to convince me that I'm eating more now than I was then I'm going to start doubting my own sanity, or at least my autobiographical memory.

But there's not much evidence to back me up. Metabolic rate is well-known to decline with age, but linearly and predictably. And it changes with muscle mass, but only minimally – and I don't think I used to be any more muscular.

The sites that talk about drastic and unexpected ways to change metabolism seem mostly crackpottish. This isn't to say their methods don't work – green tea, for example, has a statistically significant effect – but it's all so small as to be pretty meaningless in a real-world context.

So my own story seems to be on shaky ground. But as far as I can tell, the people arguing that they're trying just as hard as anybody else but still unable to lose weight because of their metabolism are very possibly right.

The Influenza Of Evil

I.

A recent Cracked piece: Five Everyday Groups Society Says It’s Okay To Mock. It begins:

> There’s a rule in comedy that says you shouldn’t punch down. It’s okay to make fun of someone rich and famous, because they’re too busy molesting groupies with 100-dollar bills to notice, but if you make a joke at the expense of a homeless person, you’re just an asshole. That said, we as a society have somehow decided on a few arbitrary exceptions to this rule.

“Somehow decided on a few arbitrary exceptions” isn’t very technical. Let’s see if we can do better.

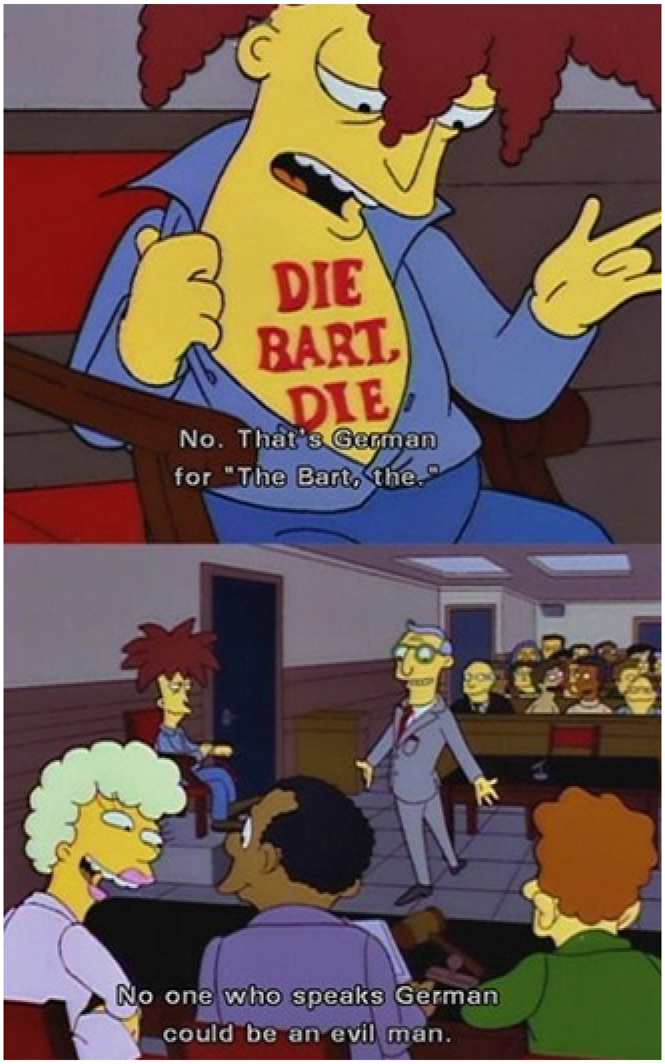
Earlier this week, I wrote about things that are anti-inductive. Something is anti-inductive if it fights back against your attempts to understand it. The classic example is the stock market. If someone learns that the stock market is always low on Tuesdays, then they’ll buy lots of stocks on Tuesdays to profit from the anomaly. But this raises the demand for stocks on Tuesdays, and therefore stocks won’t be low on Tuesdays anymore. To detect a pattern is to destroy the pattern.

The less classic example is job interviews where every candidate is trying to distinguish themselves from every other candidate. If someone learns that interviewers are impressed if you talk about your experience in tropical medicine, then as more and more people catch on they’ll all get experience in tropical medicine, it will become cliche, and people won’t be impressed by it anymore.

Evil, too, is anti-inductive.

The Nazis were very successful evildoers, at least for a while. Part of their success was convincing people – at least the German people, but sometimes also foreigners – that they were the good guys. And they *were* able to convince a lot of people, because people can be pretty dumb, a lot of them kind of just operate by pattern-matching, and the Nazis didn’t match enough patterns to set off people’s alarms.

Neo-Nazis cannot be called “successful” in any sense of the word. Their PR problem isn’t just that they’re horrible – a lot of groups are horrible and do much better than neo-Nazis. Their PR problem is that they’re horrible in exactly the way that our culture formed memetic antibodies against. Our pattern-matching faculties have been trained on Nazis being evil. The alarm bells that connect everything about Nazis to evil are hypersensitive, so much so that even contingent features of the Nazis remain universally acknowledged evil-signals.


DIE BART, DIE
No. That's German
for "The Bart, the."
No one who speaks German
could be an evil man.

It would be premature to say that we will never have to worry about fascism again. But for now, we are probably pretty safe from fascism that starts its sales pitch with “Hi, I’m fascism! Want a swastika armband?”

Huey Long supposedly predicted that “Fascism in America will attempt to advance under the banner of anti-fascism.” I’m not sure I like the saying as it stands – it seems too susceptible to Hitler Jr. telling Churchill Jr. that he’s marching under the banner of anti-fascism which proves he’s the *real* fascist. Then again, in a world where capitalism marches under the banner of “socialism with Chinese characteristics”, who knows? I would prefer to say that fascism will, at the very least, advance in a way which carefully takes our opposition to fascism into account .

Sure enough, people who had learned to be wary of fascism were still highly susceptible to communism, which wore its anti-fascism proudly on its sleeve as a symbol of how great it was. It convinced a lot of very smart people in the free world that it was the best thing since sliced bread, all while murdering tens of millions of people. Meanwhile, our memetic immune systems were sitting watchfully at their posts, thinking “Well, this doesn’t look at all like Nazism. They’re saying all the right stuff about equality, which is like the *opposite* of what the Nazis said. I’m giving them a pass.”

In fact, I’ll make the analogy more explicit. Every winter, there’s a flu epidemic. Every spring and summer, people’s bodies put in a lot of effort making antibodies to *last year’s* flu. The next winter, the flu mutates a little, a new virus with new antigens starts a new epidemic, and the immune system doesn’t have a clue: “This virus doesn’t have the very very specific characteristic I’ve learned to associate with the flu. Maybe it wants to be my friend!” This is why we need the WHO to predict what the up-and-coming flu virus will be and give us vaccines against it; it’s also why their job is so hard; they don’t know what’s coming, except that it will look different from however it’s looked before.

Nowadays most people’s memetic immune systems have some antibodies to communism, and people talking with Russian accents about how we need to eliminate the bourgeoisie and institute a dictatorship of the proletariat sends shiver up the spines of a lot of people. Nowadays an openly Communist party faces the same uphill battle as an openly Nazi party.

But that just means that if there’s some other evil on the horizon, it probably won’t resemble either fascism *or* communism. It will be movement about which everyone’s saying “These new guys are so great! They don’t pattern-match to any of the kinds of evil we know about at all!” By Long’s formulation, it may very well be marching under the banners of anti-fascism *and* anti-Communism.

(I’m not vagueblogging, by the way. I honestly don’t have anyone in mind here. The whole point is that it’s probably someone I’m not expecting. And if you say “I KNOW EXACTLY WHICH GROUP IT WILL BE, BASED ON THOSE CRITERIA IT’S CLEARLY X!” consider the possibility that you’re missing the point.)

II.

But getting back to the Cracked article.

We as a society have mostly figured out that shouting “GET A JOB, LOSER!” at the homeless is mean. We have mostly figured out that shouting “YOU’RE GOING TO HELL” at people of different religions is bad. We’re even, slowly but surely, starting to wonder whether there’s something problematic about shouting “FAGGOTS!” at the local gay couple.

Stupid bullies will continue to do those things, just as stupid investors will continue to read “How To Beat The Stock Market” books published in 1985, and stupid socialites will continue to wear the fashion that was cool six months ago.

But smart bullies are driven by their desire to have their bullying make them more popular, to get the rest of the world pointing and laughing with them. In a Blue Tribe bubble, shouting “FAGGOT” at gay people is no longer a good way to do that. The smart bullies in these circles have long since stopped shouting at gays – not because they’ve become any nicer, but because that’s no longer the best way to keep their audience laughing along with them.

Cracked starts off by naming mentally ill celebrities as a group society considers it okay to mock. This doesn’t seem surprising. Nowadays people talk a lot about punching-up versus punching-down. But that just means bullies who want to successfully punch down will come up with a way to make it look like they’re punching up. Take a group that’s high-status and wealthy, but find a subset who are actually in serious trouble and mock *them*, all the while shouting “I’M PUNCHING UP, I’M PUNCHING UP!”. Thus mentally ill celebrities.

The other examples are harder to figure out. I would argue that they’re ones that are easy to victim-blame (ie obesity), ones that punch down on axes orthogonal to the rich-poor axis we usually think about and so don’t look like punching down (ie virginity), or ones that are covertly associated with an outgroup. In every case, I would expect the bullies involved, when they’re called upon, it to loudly protest “But that’s not real bullying! It’s not like [much more classic example of bullying, like mocking the homeless]!” And they will be right. It’s just different enough to be the hot new bullying frontier that most people haven’t caught onto yet.

I think the Cracked article is doing good work. It’s work that I also try to do (see for example number 6 here, which corresponds to Cracked’s number 5). It’s the work of pointing these things out, saying “Actually, no, that’s bullying”, until eventually it sinks into the culture, the bullies realize they’ll be called out if they keep it up, and they move on to some new target.

All of this ties *way* into the dynamic I talked about in *Untitled*. I mean, look at the people on Cracked’s list of whom society says it’s okay to mock. Virgins. The obese. People who live in their parents’ basements. Generalize “mentally ill celebrities” just a little bit to get “people who are financially well-off but non-neurotypical” and there you go.

I apologize for irresponsibly claiming to have found a pattern in an anti-inductive domain. You may now all adjust your behavior to make me wrong.

Depression Is Not A Proxy For Social Dysfunction

I.

Here is a terrible article from the New York Post: Sorry, Liberals, Scandinavian Countries Aren't Utopias.

Its thesis is interesting and worth exploring, but instead of a principled investigation, the article just publishes a bunch of cherry-picked smears about Scandinavia. Did you know that 5% of Danes have had sex with animals?

(What percent of people in other countries have had sex with animals? I don't know. More important, I see no sign that the New York Post knows either.)

But the part that really caught my eye was statements like these:

> Why does no one seem particularly interested in visiting Denmark? Visitors say Danes are joyless to be around. Denmark suffers from high rates of alcoholism. In its use of antidepressants it ranks fourth in the world. (Its fellow Nordics the Icelanders are in front by a wide margin) … Finland, which tops the charts in many surveys, is also a leader in categories like alcoholism, murder, suicide and antidepressant usage.

The Post is not the only paper to make this argument. The Guardian ("The Grim Truth Behind The Scandinavian Miracle") has said much the same thing:

> Take the Danes, for instance. True, they claim to be the happiest people in the world, but why no mention of the

> fact they are second only to Iceland when it comes to consuming anti-depressants?…Finland has by far the highest suicide rate in the Nordic countries.

I’ve heard this same argument applied to other issues; for example, in his debate with Noah Smith, Michael Anissimov argues against the supposed success of modern liberal society by pointing out rising rates of depression and suicide.

It’s really tempting to equate depression with misery and misery with social dysfunction. Danes and Finns have high levels of depression, therefore their lives must be unusually miserable, therefore Denmark and Finland are poorly-organized societies.

But first of all, it’s not clear that Scandinavian countries really have very high depression and suicide rates. There are a lot of collections of statistics, and many of them show Scandinavia around the middle. Going by “antidepressant prescriptions” is a terrible way to do things, because it mixes amount of depression with resources devoted to treating depression – if the Scandinavian health systems are as good as everyone says, maybe they just treat a greater percent of their depressives than everywhere else.

But more important, even if Scandinavia does have very high rates of depression, that doesn’t tell us much about whether they’re happy or not. *Depression is not the same thing as being sad*. Sadness is a risk factor for depression – although even there I suspect that it’s very specific kinds of sadness that we haven’t yet teased out from the general construct – but it is not the condition itself. The condition itself is a complicated mess of neurotransmitters, cytokines, hormones, changes in brain structure, and goodness only knows what else.

Off the top of my head, here are six plausible reasons why Scandinavia could have higher rates of depression than the United States, even if it is a utopian society of perfect happiness.

1. Light. Scandinavia is far north [citation needed] which puts its citizens at very high risk for seasonal affective disorder, which can present as depression.

2. The midnight sun. Scandinavia's weird day-night cycle could easily disrupt people's circadian rhythms. Studies find that "increasing evidence points to a role of the biological clock in the development of depression…it seems likely the circadian system plays a vital role in the genesis of the disorder. This is why some European countries use melatonergic substances as antidepressants.

3. Parasite load. It's positively correlated with temperature, which means Scandinavia probably has some of the lowest parasite load in the world. But low parasite load causes the immune system to get antsy and start attacking random stuff, leading to increase risk of autoimmune disease. If there's an immunological component to depression – and right now lots of people think there is – then that's another risk factor right there.

4. Diet. The Scandinavian diet has unusually little fresh food, because the area is a frozen wasteland and most things have to be imported from elsewhere. They're big on frozen stuff, processed stuff, and canned stuff. I am neither an expert in Scandinavian cuisine nor in nutrition, but if depression is linked to diet and imbalance in the gut microbiome, which there's some evidence it is, then diet is heavily implicated and the Scandinavians are in a good position to get hit extra hard.

5. Genetics. The New York Post article mentions that Scandinavians have an unusual variant of the MAO-A enzyme (I told you it was a weird hit piece. Scandinavia is too liberal, therefore they have bad genes?). MAO-A is also known as “the thing that processes serotonin” and “the thing that MAO inhibitors, some of the most powerful known antidepressants, inhibit”. I’m not saying this gene in particular is responsible for Scandinavian depression, I’m saying that the article itself is admitting that Scandinavia contains some genetically distinct populations and for all we know this could be involved.

6. Culture. Maybe *the* biggest factor in the level of depression and suicide in a culture is whether it is culturally acceptable to be depressed and commit suicide. Some of the lowest suicide rates are found in heavily religious cultures and communities who believe suicide is a mortal sin. On the other hand, one of the most suicidal countries in the world is Japan, with its heavily-mythologized history of heroic samurai taking “the honorable way out” when they had brought shame upon themselves. Well, Scandinavia is one of the least religious regions in the world. And all I know about their culture is that they produce about 100% of good death metal, and their native mythology ends with the world being plunged into eternal winter and the gods being eaten by wolves.

II.

But all this is just speculation. Let me give a concrete example of a case where social dysfunction doesn’t track depression and suicidality in a predictable way.

What about white versus black Americans? To some degree these two groups live in separate “societies”. Most people would consider the white society better off in most ways –

higher income, better health, more family stability, less involvement with the criminal justice system. If White America and Black America were countries, White America would get all of the accolades currently given to the Scandinavians.

But American whites have higher rates of depression than blacks. There are the usual contradictory studies and arguments about how to adjust for which confounder, but I'm pretty sure this is something like a consensus position right now. More solidly, white Americans have much higher suicide rates than black Americans.

(although I feel bad mentioning this, because the stereotype that blacks *never* commit suicide is wrong and sometimes prevents black people from getting the help they need.)

We can go a few centuries back and get even more surprising results. Although it's difficult to get data from the era, analyses of suicide rate among African-American slaves in the antebellum South describe it as "surprisingly low". I can't find any hard evidence proving Kurt Vonnegut's contention that "the suicide rate per capita among slave owners was much higher than the suicide rate among slaves", but it seems to have been commonly believed. Kneeland writes:

> "[These low suicide rates are] consistent with suicide rates for Africa and for people of African descent living in other areas of the world, and further supports the theory that a low suicide rate is an element of African culture."

If you're going to say that Scandinavia's higher depression and suicide rates mean Scandinavia has it worse off than America, you also need to theorize that white people have it worse off than black people, including black slaves. Why don't you go post something to that effect on Tumblr and see what they have to say? I'll wait.

III.

Or maybe we're barking up entirely the wrong tree. What if it's not even that happy, well-functioning societies can sometimes still end up with high suicide rates? What if people become suicidally depressed precisely *because* they live in happy, well-functioning societies?

This is the fascinating hypothesis of Daly, Oswald, and Wu (2011), who after crunching the numbers find pretty convincingly that "suicide rates tend to be highest in happy places":

> A little-noted puzzle is that many of [the happiest] places have unusually high rates of suicide. While this fact has been remarked on occasionally for individual nations, especially for the case of Denmark, it has usually been attributed in an anecdotal way to idiosyncratic features of the location in question (eg the dark winters in Scandinavia), definitional variations in the measurement of well-being and suicide, and differences in culture and social attitudes regarding happiness and taking one's life. Most scholars have not thought of the anecdotal observation as a systematic relationship that might be robust to replication or investigation…this paper attempts to document the existence of a happiness-suicide paradox: happier areas have a higher percentage of suicides.

They then go on to show a strong positive relationship between average self-reported happiness and suicidality across Western nations – Greece is both the least happy country and the one with the lowest suicide rate – and US states, where confirmed hellholes New York and New Jersey are at or near the bottom. The relationship holds whether you adjust for confounders (including income!) or not.

I expected this to be a straightforward effect of modernization/industrialization/liberalism, as per Michael Anissimov’s hypothesis. The country-level data maybe sort of vaguely supports that trend – Greece and Portugal are our token incompletely-modernized countries and have very low suicide rates, Scandinavia is high, and everywhere else is sort of a toss-up. But US states really really don’t support that hypothesis – New York and Jersey both seem high on the modernization/industrialization/liberalism axis, and they’re right in the bottom left corner of the study’s graphs along with Greece and Portugal. Meanwhile, tropical paradise Hawai’i is suicidal as heck, even though it doesn’t seem espcially modern/industrial/liberalized. The US state data also torpedo – albeit less conclusively – an attempt to make the whole issue one of latitude.

One caveat I do have about the US data is that several of the happiest and most suicidal states – at least on the unadjusted plot – are also high-altitude. Utah, Wyoming, Colorado, Montana, Idaho are all up there at the top left side of the graph. But we already know there’s a strong positive relationship between altitude and suicide in 2584 US counties, probably because the brain’s emotional regulation system doesn’t work well in low-oxygen environments. If we assume people living in beautiful open forested mountain areas are especially happy, that takes away a big chunk of the graph right there. But it leaves other chunks untouched, and I don’t think it’s going to be that simple.

The authors’ preferred explanation is that suicide is an effect of relative rather than absolute misery. If you’re depressed and everybody around you is very happy, that makes things worse than if you're depressed and everyone around you is also pretty miserable. Thus suicide is more common in happier societies.

I really don't like this theory. Although everyone else should be happier in these societies, the person in question who might or might not commit suicide should also be, on average, happier. There's no reason to think that the average hedonic *distance* between potential suicides and their neighbors is higher in these areas. Indeed, given that Scandinavia – and many of the other happy societies – are also some of the most equal societies, I would expect an unusually low hedonic distance between people. And in fact, I notice that suicide rates by country are negatively correlated with inequality – that is, the more unequal the country, the lower the suicide rate (wow, I *definitely* don't remember seeing that one in *The Spirit Level*.)

On the other hand, I can't for the life of me think of a better theory, so whatever.

Other things that increase suicide rates, by the way, include springtime, nice weather, high levels of education, and very occasionally antidepressants. My father, a very hard-headed internist, makes fun of me for doing psychiatry because "the whole field is just common sense", but sometimes it *really isn't*.

So you should probably think very carefully before using a difference in depression or suicide rates to support your pet theory about which societies work better than others.

These Are A Few (More) Of My (Least) Favorite Things

One year ago, I wrote Ten Things I Want To Stop Seeing On The Internet In 2014.

And now it's 2015, and I think things are getting better. Take doge. I swear to God that the last time I saw the word doge, it was referring to an honest-to-God Venetian noble. And the price of dogecoin is down an order of magnitude from its peak last February. The War on Doges is starting to seem winnable.

I can't take any credit for this. It has been a concerted effort on the part of millions of people who saw doge memes on Facebook, let their fingers briefly drift towards the "share" button, and then pulled themselves back from the precipice, restrained by their better nature.

But in the hopes that this is the first success of many, I would like to share some things I want to stop seeing on the Internet in 2015:

1. Abuse Of Poe's Law

Poe's Law is the belief that some religious fundamentalists are so stupid that it's impossible to distinguish them from a parody.

This is all nice and well in the abstract, but when applied to a particular case, where a particular atheist has fallen for a parody site, it tends to be an unfortunate stand-in for "Some atheists are so ignorant that it's impossible for them to distinguish religious people from a parody of religious people." Listen:

A: “The Pope just said that everyone who isn’t creationist should be put in jail! What an outrage!”

B: “Uh, you *do* know that’s on The Onion, right?”

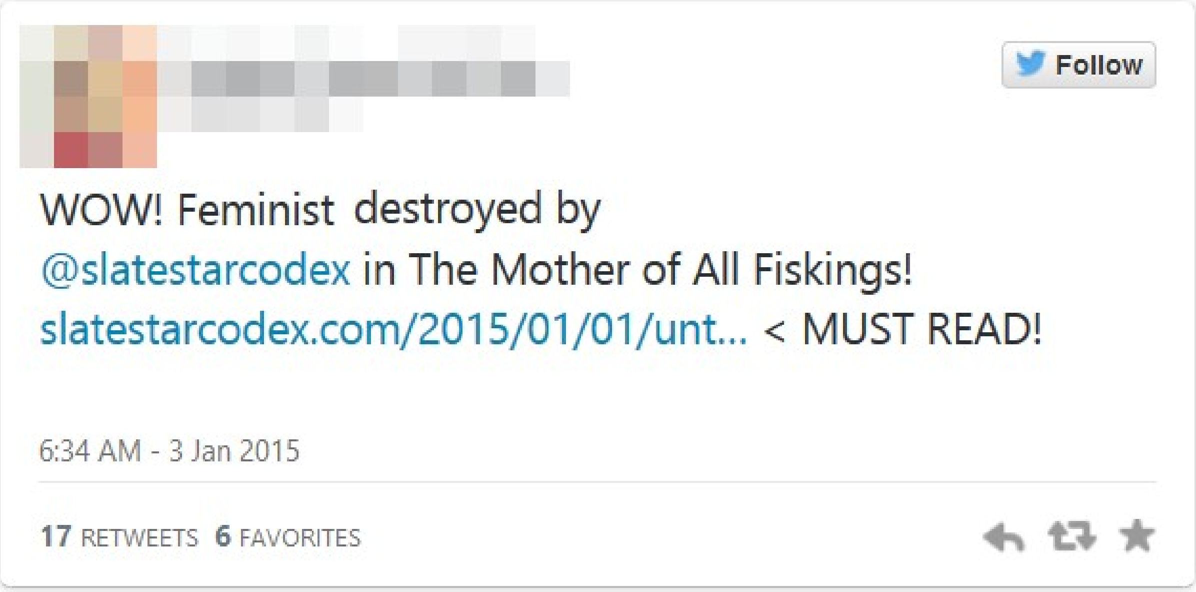
A: “Oh, well, haha, Poe’s Law, just goes to show how dumb those religious people are.”

Problem is, Poe’s Law isn’t limited to religion any more. Now it’s politics, culture, science, and anywhere else where one side thinks their opponents are so stupid it’s literally impossible to parody them (ie everywhere on both sides). You spread the dumbest and most obviously fake rumors to smear your opponents. And then when you’re caught, instead of admitting you were fooled, you claim Poe’s Law and smear your opponents even more.

On the other hand, once you’re willing to admit this dynamic exists, it can make for some pretty interesting guessing games and unintentional Intellectual Turing Tests – see the Poe’s Law In Action subreddit for some examples.

2. People Getting Destroyed By Other People

Whenever I write a persuasive piece, I get to see my fans share it on Twitter like this:



WOW! Feminist destroyed by @slatestarcodex in The Mother of All Fiskings! slatestarcodex.com/2015/01/01/unt... < MUST READ!

6:34 AM - 3 Jan 2015

17 RETWEETS 6 FAVORITES

I didn't destroy anybody. I *disagreed with* them.

I'm glad to know I'm not the only one who has to deal with this. Newsweek writes about how Jon Stewart Is A Violent Sociopath Who Must Be Stopped in reponse to increasing claims that Stewart "destroys", "demolishes", "disembowels", and "makes ground beef" out of whoever he's arguing against on his show.

This bothers me the same way that "debunked" bothers me. Both sides are going to insist that their own research "debunks" the other, and so make it impossible to have a conversation based on the premise that there's still room for disagreement. The flip side of my fans believing that I've destroyed whoever is that when that person writes a response, their fans are going to believe they've destroyed me.

At least no one can eviscerate me, since Jon Stewart has already eviscerated the entire blogosphere.

3. Demonstrating That People Are Stupid By Having Them Use The Word "Muh"

No straw man is ever concerned about immigrants stealing his job. He’s always concerned about immigrants stealing “muh jarb”, or possibly “muh jawb”, which sounds like some form of obscure Islamic garment.

This has lately taken a disturbing turn in the form of straw feminists worrying about “muh sojiny”. I strongly believe that every women has a right to her sojiny and no man should be able to take it from her, but I still can’t help wishing that people would lay off the cheap shots for a while.

4. Wikipedia-Shaming

Did you know it is 2015 and people will still criticize you for getting facts off of Wikipedia?

I’m not even talking about controversial conclusions, like “on balance, the research about gun control shows…”. I’m talking about simple facts.

A: “China is bigger than the United States”

B: “Where’d you hear that one, *Wikipedia*?”

A: “…yes?”

B: “You expect me to believe something you literally just took off a Wikipedia article?”

Yes. Yes I do. I could go find the CIA World Factbook or whatever, but it will say the same thing as Wikipedia, because Wikipedia is pretty much always right. When you challenge Wikipedia on basic facts, all you do is force people to use inconvenient sources to back up the things Wikipedia says, costing people time for no reason and making them hate you. There may have been a time when Wikipedia was famously inaccurate. Or maybe there wasn’t. I don’t know.

Wikipedia doesn't have an article on it, so it would take time and energy to find out. The point is, now it's 2015, and the matter has been settled.

How accurate is Wikipedia?:

> Several studies have been done to assess the reliability of Wikipedia. An early study in the journal Nature said that in 2005, Wikipedia's scientific articles came close to the level of accuracy in Encyclopædia Britannica and had a similar rate of "serious errors". The study by Nature was disputed by Encyclopædia Britannica, and later Nature replied to this refutation with both a formal response and a point-by-point rebuttal of Britannica?'?s main objections. Between 2008 and 2012, articles in medical and scientific fields such as pathology, toxicology, oncology, pharmaceuticals, and psychiatry comparing Wikipedia to professional and peer-reviewed sources found that Wikipedia's depth and coverage were of a high standard.

I know this because I got it from Wikipedia's Reliability Of Wikipedia article. Go ahead, challenge me, I dare you.

5. Articles That Start Off With An Image Taking Up The Entire Screen

This is what I'm talking about. I click on the link expecting an article on gas pipeline deals, and there is exactly zero article on the first screenfull of page I come to. That's fine, you know, the only reason I even clicked was to see a huge, high-resolution picture of Vladimir Putin's head. Information is totally optional. Screw you. This is why if I want to learn about Russian-Chinese gas deals, I'll just look them up on Wikipedia.

I feel the same way about those Web 2.0 sites where the landing page is just an image of a smiling group of people engaged in a nondescript activity, and then way up in the corner is a tiny button that says “Discover” (it’s always “discover”) which leads to actual information. Likewise this site, which probably made its designer feel very smug about their clean minimalist style, but where *you can’t get a single word of information without watching a video*.

6. Ads That Disappear Very Slowly

You get an ad. It appears at the bottom of the screen. You look at it, decide you’re not interested, click the little X. It disappears. But not right away. It crawls. It *saunters*. After what seems to be a long and arduous journey, during which it had to ford several rivers and stop off at Fort Laramie for supplies, it finally makes it to the bottom of the screen and fades away.

I try hard to understand other people’s perspectives. I know that companies need to have ads to make money. I know that they have an incentive to make those ads as disruptive and obnoxious as possible to make you look at them. I even understand why some ads have the little x kind of hidden, so you can’t find it without some poking around, which forces you to view the ad for a little while longer. I understand all those things.

But I don’t understand why the ad has to take so long to disappear. It’s obviously not just incompetence. They specifically have to add an extra little sliding-down animation to the ad to make it take so long. They put in *more work* to make it *more annoying* for *no benefit*. Do you really think that while I’m waiting for the ad to disappear, I’m thinking “You know, I thought I didn’t need to meet hot desperate singles in my area, which is why I clicked the X to make it go away, but that sliding-down-the-screen animation is so cool that I'm going to reload the page a couple of times, wait for the ad to come back, and then click it"?

7. Overuse Of Demonstratives In Clickbait

I understand that demonstratives ("this", "that", "yon") are supposed to give you a bit of mystery, make you want to click on the article to see what's happening. "This Celebrity Just Came Out As Gay" makes you wonder which one it is. "Rare Disease Spreads To These Three US States" makes you check if yours is one of them. Fine. I would personally prefer "Rare Disease Spreads To Three States" or even "Which Three States Did A Rare Disease Spread To? Click Here To Find Out", but whatever.

However. On Vox recently, Obama Just Hit These North Koreans With Sanctions. What, exactly, are we supposed to get out of this? "Oh! I wonder if it's Yu Kwang Ho! Surely they wouldn't get Yu Kwang Ho! Better click to find out!"

8. Any Use Of The Word "Entitled"

Okay, I've already written on what I think of people calling nerds "entitled". But it goes beyond that.

Number 8 in last year's Least Favorite Things was "arguments about which generation is better". Well, now those have progressed to arguments over which generation is most entitled. Hard Work? No Thanks! Meet Entitled-To-It-All Generation Y. Millenials are Selfish and Entitled and Helicopter Parents Are To Blame. But The Most Entitled Generation Isn't Millennials, It's Baby Boomers. And coming in from left field, maybe The Greatest Generation Was The Most Entitled. There are even entire books about this

Men feel entitled to women. Women feel entitled to men. Blacks feel entitled. Whites feel entitled. The Entitlement

Mentality of Liberals coexists with Entitled Conservative White Male Putzes, possibly because Conservatives Feel ‘Entitled’ To Scorn ‘Entitlement’ (whatever).

Anyone can call their out-group entitled. The easiest way is – well, poor people are entitled because they demand handouts without working for them. Rich people are entitled because they think they deserve 100% of what they have and refuse to acknowledge or change the inequalities in the system that benefit them. One side or the other of that dichotomy is likely to map onto whatever group you want to insult.

“Entitled” is a Fully General Insult that can apply to anyone, and it really hurts. That makes it irresistable to the wrong kind of people, and it’s why I hope I start seeing less of it. Alternately, people could start giving their enemies the Psychological Entitlement Scale, which is so hilariously obvious with what it’s doing that I find it astounding that it apparently still manages to successfully detect some entitled people. The Titanic? *Really?*

9. People Being Post-Things

I recently heard someone describe themselves as “post-Zionist”, then go on to give what sounded like pretty standard criticism of Zionism. I don’t want to get *too* heavily into this particular example, because I understand post-Zionism is complex and every time I write something about Israel I get Israeli commenters saying I’ve gotten it wrong and other Israeli commenters saying no *they’ve* gotten it wrong and still other Israeli commenters saying we’ve *all* got it wrong. What was that saying about “two Jews, three opinions” again?

But what bothers me about post-Zionism is that it seems to carry this kind of smug “Oh, you guys are *still* Zionist? Don’t you know Zionism is, like, totally five years ago? *Nowadays* all the cool people have moved on to more exciting things," which I don't think really adds to the argument. Zionism versus anti-Zionism suggests a picture of two sides with two different opinions – which seems to match the reality pretty well. Zionism versus post-Zionism suggests one side just hasn't gotten the message yet.

I feel the same way about post-rationalism. Yes, maybe you've seen through rationalism in some profound way and transcended it. Or maybe you just don't get it. This is exactly the point under debate, and naming yourselves "post-rationalists" seems like an attempt to short-circuit it, not to mention leaving everyone else confused. And maybe you could give yourself a name that actually reflected your beliefs ("Kind Of New-Age-y People Who Are Better At Math Than Usual For That Demographic And Will Angrily Deny Being New-Age-y If Asked Directly"?) and we wouldn't have to have a new "but what is post-rationalism?!?!" conversation every month.

Post-modernism can stay, though. At this point it's less of a name than a warning label.

10. Disputes Over Whether Humans Evolved From Monkeys

I don't mean creationism. I mean disputes among people who accept evolution, over whether it was monkeys in particular that humans evolved from.

It tends to go something like this.

A: "Humans evolved from monkeys".

B: "No they didn't! They evolved from chimps! Chimps are an ape, not a monkey!"

C: “Humans didn’t evolve from chimps! They evolved from a most recent common ancestor whose descendants include both humans and chimps!”

Everything about this conversation is not-even-wrong.

First, humans clearly evolved from monkeys in the same sense humans evolved from single-celled organisms. No one’s saying it had to be the most recent step.

Second, apes are ambiguously a type of monkey. Think square versus rectangle. All squares are rectangles but not all rectangles are squares and “rectangle” is usually used to indicate rectangles that are not squares but can technically refer to squares as well. Here’s a primatologist saying that Apes Are Monkeys – Deal With It.

Third, the most recent common ancestor of humans and chimpanzees may (or may not) have been a chimpanzee. This is Richard Wrangham’s thesis, and he calls it pan prior, placing it firmly within the chimpanzee genus.

These last two issues are especially annoying because they’re kind of meaningless category disputes. Yet for some reason the Internet seems to be obsessed with the lurking fear that someone, somewhere, might be saying that people evolved from monkeys or chimps.

Seriously. Get a life, Internet.

Perceptions Of Required Ability Act As A Proxy For Actual Required Ability In Explaining The Gender Gap

I.

I briefly snarked about Leslie et al (2015) last week, but I should probably snark at it more rigorously and at greater length.

This is the paper that concludes that “women are underrepresented in fields whose practitioners believe that raw, innate talent is the main requirement for success because women are stereotyped as not possessing that talent.” They find that some survey questions intended to capture whether people believe a field requires innate talent correlate with percent women in that field at a fairly impressive level of r = -0.60.

The media, science blogosphere, et cetera has taken this result and run with it. A very small sample includes: National Science Foundation: Belief In Raw Brilliance May Decrease Diversity. Science Mag: the “misguided” belief that certain scientific fields require brilliance helps explain the underrepresentation of women in those fields. Reuters: Fields That Cherish Genius Shun Women. LearnU: Study Findings Point To Source Of Gender Gap In STEM. Scientific American: Hidden Hurdle Looms For Women In Science. Chronicle Of Higher Education: Disciplines That Expect Brilliance Tend To Punish Women. News Works: Academic Gender Gaps Tied To Stereotypes About Genius. Mathbabe: “The genius myth” keeps women out of science. Vocativ: Women Avoid Fields Full Of Self-Appointed Geniuses. And so on in that vein.

Okay. Imagine a study with the following methodology. You survey a bunch of people to get their perceptions of who is a smoker ("97% of his close friends agree Bob smokes"). Then you correlate those numbers with who gets lung cancer. Your statistics program lights up like a Christmas tree with a bunch of super-strong correlations. You conclude "Perception of being a smoker causes lung cancer", and make up a theory about how negative stereotypes of smokers cause stress which depresses the immune system. The media reports that as "Smoking Doesn't Cause Cancer, Stereotypes Do".

This is the basic principle behind Leslie et al (2015).

The obvious counterargument is that people's perceptions may be accurate, so your perception measure might be a proxy for a real thing. In the smoking study, we expect that people's perception of smoking only correlates with lung cancer because it correlates with actual smoking which itself correlates with lung cancer. You would expect to find that perceived smoking correlates with lung cancer *less than* actual smoking, because the perceived smoking correlation is just the actual smoking correlation plus some noise resulting from misperceptions.

So I expected the paper to investigate whether or not perceived required ability correlated more, the same as, or less than actual required ability. Instead, they simply write:

> Are women and African-Americans less likely to have the natural brilliance that some fields believe is required for top-level success? Although some have argued that this is so, our assessment of the literature is that the case has not been made that either group is less likely to possess innate intellectual talent[1].

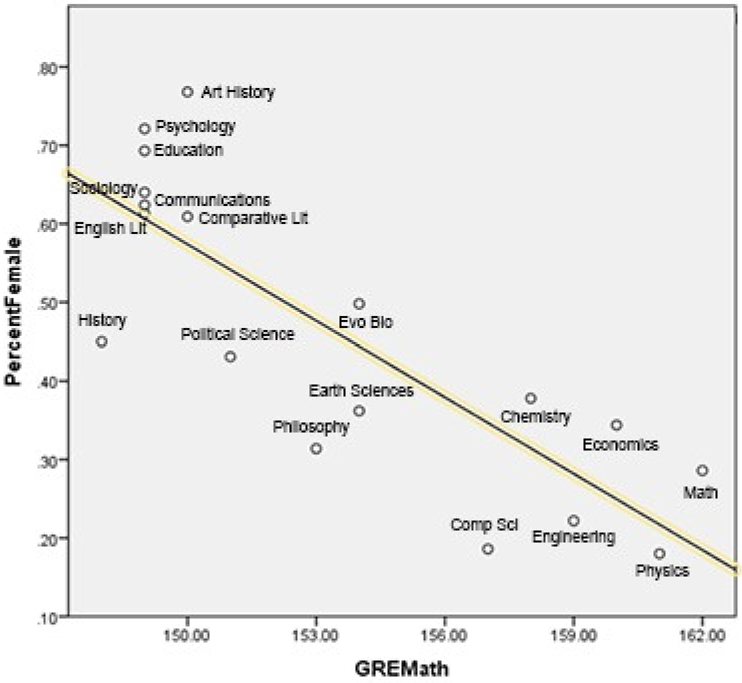
So we will have to do this ourselves. The researchers helpfully include in their supplement a list of the fields they studied and GRE scores for each, as part of some sub-analysis to check for selectivity. GRE scores correlate closely with IQ and with a bunch of measures of success in graduate school, so this sounds like it would be a good test of the actual required ability hypothesis. Let's use this to figure out whether actual innate ability explains the discrepancies better or worse than perceived innate ability does.

When I use these data I find no effect of GRE scores on female representation.

But these data are surprising – for example, Computer Science had by far the lowest GRE score (and hence projected IQ?) of any field, which matches neither other sources nor my intuition. I looked more closely and found their measure combines Verbal, Quantitative, and Writing GREs. These are to some degree anti-correlated with each other across disciplines[2]; ie those disciplines whose students have higher Quantitative tend to have lower Writing scores (not surprising; consider a Physics department versus an English department).

Since the study's analysis included two measures of verbal intelligence and only one measure of mathematical intelligence, it makes more mathematical departments appear to have lower scores and lower innate ability. Certainly a measure set up such that computer scientists get the lowest intelligence of everyone in the academy isn't going to find innate ability related to STEM!

Since the gender gap tends to favor men in more mathematical subjects, if we're checking for a basis in innate ability we should probably disentangle these tests and focus on the GRE Quantitative. I took GRE Quantitative numbers by department from the 2014 edition of the ETS report. The results looked like this:



There is a correlation of $r = -0.82$ ($p = 0.0003$) between average GRE Quantitative score and percent women in a discipline. This is among the strongest correlations I have ever seen in social science data. It is much larger than Leslie et al's correlation with perceived innate ability[3].

Despite its surprising size this is not a fluke. It's very similar to what other people have found when attempting the same project. There's a paper from 2002, Templer and Tomeo, that tries the same thing and finds $r = 0.76$, $p < 0.001$. Randal Olson tried a very similar project on his blog a while back and got $r = 0.86$. My finding is right in the middle.

A friendly statistician went beyond my pay grade and did a sequential ANOVA on these results[4] and Leslie et al's perceived-innate-ability results. They found that they could reject the hypothesis that the effect of actual innate ability was entirely mediated by perceived innate ability (p = 0.002), but could not reject the hypothesis that the effect of perceived-innate-ability was entirely mediated by actual-innate ability (p = 0.36).

In other words, we find no evidence for a continuing effect of people's perceptions of innate ability after we adjust for what those perceptions say about actual innate ability, in much the same way we would expect to see no evidence for a continuing effect of people's perceptions of smoking on lung cancer after we adjust for what those perceptions say about actual smoking.

II.

Correlation is not causation, but a potential causal mechanism can be sketched out.

I'm going to use terms like "ability" and "innate ability" and "genius" and "brilliance" because those are the terms Leslie et al use, but I should clarify. I'm using them the way Leslie et al seem to, as a contrast to hard work, the internal factors that give different people different payoffs per unit effort. So a genius is someone who can solve difficult problems with little effort; a dullard is one who can solve them only with great effort or not at all.

This use of "innate ability" is *not* the same thing as "genetically determined ability". Genetically determined ability will be part of it, but there will also be many other factors. Environmental determinants of intelligence, like good nutrition and low lead levels. Exposure to intellectual stimulation during crucial developmental windows. The effect of steretoypes, insofar as those stereotypes globally decrease performance. Even previous training in a field might represent “innate ability” under this definition, although later we’ll try to close that loophole.

Academic programs presumably want people with high ability. The GRE bills itself as an ability test, and under our expanded definition of ability this is a reasonable claim. So let’s talk about what would happen if programs selected based solely on ability as measured by GREs.

This is, of course, not the whole story. Programs also use a lot of other things like grades, interviews, and publications. But these are all correlated with GRE scores, and anyway it’s nice to have a single number to work with. So for now let’s suppose colleges accept applicants based entirely on GRE scores and see what happens. The STEM subjects we’re looking at here are presumably most interested in GRE Quantitative, so once again we’ll focus on that.

Mathematics unsurprisingly has the highest required GRE Quantitative score. Suppose that the GRE score of the average Mathematics student – 162.0 – represents the average level that Mathematics departments are aiming for – ie you must be this smart to enter.

The average man gets 154.3 ± 8.6 on GRE Quantitative. The average woman gets 149.4 ± 8.1. So the threshold for Mathematics admission is 7.7 points ahead of the average male test-taker, or 0.9 male standard deviation units. This same threshold is 12.6 points ahead of the average female test-taker, or 1.55 female standard deviation units.

GRE scores are designed to follow a normal distribution, so we can plug all of this into our handy-dandy normal distribution calculator and find that 19% of men and 6% of women taking the GRE meet the score threshold to get into graduate level Mathematics. 191,394 men and 244,712 women took the GRE last year, so there will be about 36,400 men and 14,700 women who pass the score bar and qualify for graduate level mathematics. That means the pool of people who can do graduate Mathematics is 29% female. And when we look at the actual gender balance in graduate Mathematics, it’s *also* 29% female.

Vast rivers of ink have been spilled upon the question of why so few women are in graduate Mathematics programs. Are interviewers misogynist? Are graduate students denied work-life balance? Do stereotypes cause professors to “punish” women who don’t live up to their sexist expectations? Is there a culture of sexual harassment among mathematicians?

But if you assume that Mathematics departments are selecting applicants based on the thing they double-dog swear they are selecting applicants based on, there is literally nothing left to be explained[5].

I am *sort of* cheating here. The exact perfect prediction in Mathematics is a coincidence. And I can’t extend this methodology rigorously to any other subject because I would need a much more complicated model where people of a given score level are taken out of the pool as they choose the highest-score-requiring discipline, leaving fewer high-score people available for the low-score-requiring ones. Without this more complicated task, at best I can set a maximum expected gender imbalance, then eyeball whether the observed deviation from that maximum is more or less than expected. Doing such eyeballing, there are slightly fewer women in graduate Physics and Computer Science than expected and slightly more women in graduate Economics than expected.

But on the whole, the prediction is very good. That it is not perfect means there is still some room to talk about differences in stereotypes and work-life balance and so on creating moderate deviations from the predicted ratio in a few areas like computer science. But this is arguing over the scraps of variance left over, after differences in mathematical ability have devoured their share.

III.

There are a couple of potentially very strong objections to this hypothesis. Let me see if I can answer them.

First, maybe this is a binary STEM vs. non-STEM thing. That is, STEM fields require more mathematical aptitude (obviously) and they sound like the sort to have more stereotypes about women. So is it possible that my supposedly large sample size is actually just showing an artifact of division into these two categories?

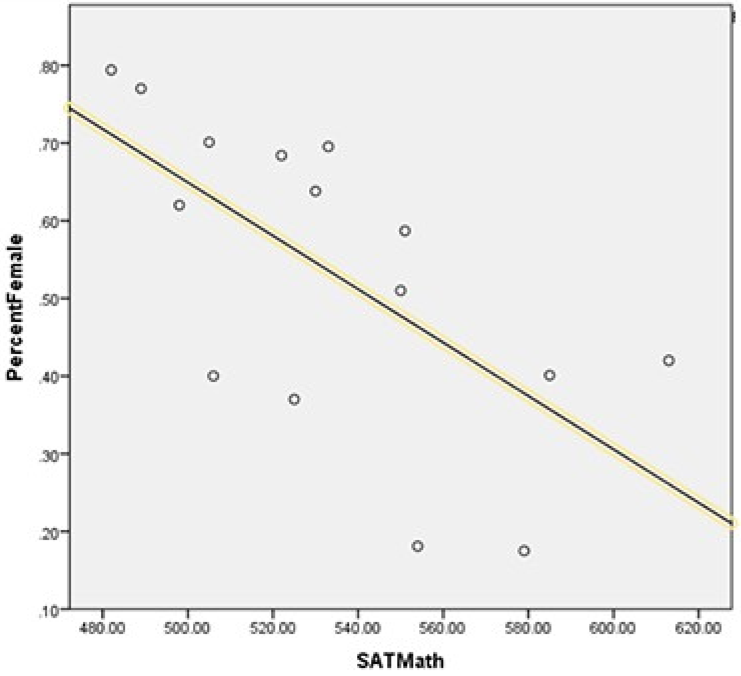
No. I divided the fields into STEM and non-STEM and ran an analysis within each subgroup. Within the non-STEM subgroup, there was a correlation between GRE Quantitative and percent female in a major of -0.64, p = 0.02. It is completely irresponsible to do this within the STEM subgroup, because it has n = 7 which is too small a sample size to get real results. But if we are bad people and do it anyway, we find a very similar correlation of -0.63. p is only 0.12, but with n=7 what did you expect?

Both of these correlations are higher than Leslie et al were able to get from their entire sample.

Second, suppose that it’s something else driving gender-based patterns in academia. Maybe stereotypes or long hours or whatever. Presumably, these could operate perfectly well in undergrad. So stereotypes cause lots of men to go into undergraduate math and lots of women to go into undergraduate humanities. The men in math classes successfully learn math and the women in humanities classes successfully learn humanities. Then at the end of their time in college they all take the GRE, and unsurprisingly the men who have been taking all the math classes do better in math. In this case, the high predictive power of mathematical ability would be a *result* of stereotypes, not an alternative to them.

In order to investigate this possibility we could look at SAT Math instead of GRE Quantitative scores, since these would show pre-college ability. SAT scores show a gap much like that in GRE scores; in both, the percentile of the average woman is in the low 40s.

Here is a graph of SAT Math scores against percent women in undergraduate majors:



SAT Math had a correlation of -0.65, p = 0.01[6].

This correlation is still very strong. It is still stronger than Leslie et al’s correlation with perceived required ability. But it is slightly weaker than the extremely strong correlation we find with GRE scores. Why?

I can’t answer that for sure, but here is a theory. The “undergraduate major” data is grabbed from what SAT test-takers put down as their preferred undergraduate major when they take the test in (usually) 11th grade. The “percent female” data is grabbed from records of degrees awarded in each field. So these are not exactly the same people on each side. One side shows the people who thought they wanted to do Physics in 11th grade. The other side shows the people who ended up completing a Physics degree.

The people who intend to pursue Physics but don’t end up getting a degree will be those who dropped out for some reason. While there are many reasons to drop out, one no doubt very common one is that the course was too hard. Therefore, the people who drop out will be disproportionately those with lower mathematical ability. Therefore, the average SAT Math score of 11th grade intended Physics majors will be lower than the average SAT Math score of Physics degree earners. So the analysis above likely underestimates the average SAT Math score of people in mathematical fields. This could certainly explain the lower correlation, and I predict that if we could replace our unrepresentative measure of SAT scores with a more representative one, much of the gap between this correlation and the previous one would close.

These data do not rule out simply pushing everything back a level and saying that these stereotypes affect what classes girls take in middle school and high school. Remember, we using “ability” as a designation for a type of excellence, not an explanatory theory of it. This simply confirms that by eleventh grade, the gap has already formed.[7].

Third, perhaps SAT and GRE math tests are not reflective of women's true mathematical ability. This is the argument from stereotype threat, frequently brought up as reasons why tests should not be used to judge aptitude.

But this is based on a fundamental misunderstanding of stereotype threat found in the popular media, which actual researchers in the field keep trying to correct (to no avail). See for example Sackett, Hardison, and Cullen (2004), who point out that no research has ever claimed stereotype threat accounts for gender gaps on mathematics tests. What the research found was that, by adding an extra stereotype threat condition, you could widen those gaps further. The existing gaps on tests like the SAT and GRE correspond to the "no stereotype threat" control condition in stereotype threat experiments, and "absent stereotype threat, the two groups differ to the degree that would be expected based on differences in prior SAT scores". Aronson and Steele, who did the original stereotype threat research and invented the field, have confirmed that this is accurate and endorsed the warning.

Anyway, even if the pop sci version of stereotype threat were entirely true and explained everything, it still wouldn't rescue claims of bias or sexism in the sciences. It would merely mean that the sciences' reasonable and completely non-sexism-motivated policy of trusting test scores was ill-advised.[8]

Fourth, might there be reverse causation? That is, suppose that there are stereotypes and sexism restricting women's entry into STEM fields, and unrelatedly men have higher test scores. Then the fields with the stereotypes would end up with the people with higher test scores, and it would look like they require more ability. Might that be all that's happening here?

No. I used gender differences in the GRE scores to predict what scores we would expect each major to have if score differences came solely from differences in gender balance. This predicted less than a fifth of the variation. For example, the GRE Quantitative score difference between the average test-taker and the average Physics graduate student was 9 points, but if this were solely because of differential gender balance plus the male test advantage we would predict a difference of only 1.5 points. The effect on SAT scores is similarly underwhelming.

But I think the most important thing I want to say about objections to Part II is that, whether they're correct or not, Part I still stands. Even if the correlation between innate ability and gender balance turns out to be an artifact, Leslie et al's correlation between perceived innate ability and gender balance is still an artifact of an artifact.

IV.

A reader of an early draft of this post pointed out the imposingly-named Nonlinear Psychometric Thresholds In Physics And Mathematics. This paper uses SAT Math scores and GPA to create a model in which innate ability and hard work combine to predict the probability that a student will be successful in a certain discipline. It finds that in disciplines "such as Sociology, History, English, and Biology" these are fungible – greater work ethic can compensate for lesser innate ability and vice versa. But in disciplines such as Physics and Mathematics, this doesn't happen. People below a certain threshold mathematical ability will be very unlikely to succeed in undergraduate Physics and

Mathematics coursework no matter how hard-working they are.

And that brought into relief part of why this study bothers me. It ignores the pre-existing literature on the importance of innate ability versus hard work. It ignores the rigorous mathematical techniques developed to separate innate ability from hard work. Not only that, but it ignores pre-existing literature on predicting gender balance in different fields, and the pre-existing literature on GRE results and what they mean and how to use them, and all the techniques developed by people in *those* areas.

Having committed itself to flying blind, it takes the thing we already know how use to predict gender balance, shoves it aside in favor of a weird proxy for that thing, and finds a result mediated by that thing being a proxy for the thing they are inexplicably ignoring. Even though it just used a proxy for aptitude to predict gender balance, everyone congratulates it for having proven that aptitude does not affect gender balance.

Science journalism declares that the myth that ability matters has been vanquished forever. The media take the opportunity to remind us that scientists are sexist self-appointed geniuses who use stereotypes to punish women. And our view of an important issue becomes just a little muddier.

I encourage everyone to reanalyze this data and see if I'm missing something. You can find the GRE data I used here and the SAT data here (both in .xlsx format).

Footnotes

1. They cite for this claim, among other things, Stephen Jay Gould's *The Mismeasure Of Man*

2. Beware the ecological fallacy; these scores are still positively correlated in individuals.

3. It was also probably more highly significant, but I can't tell for sure because (ironically) their significance result wasn't to enough significant digits.

4. There was a small error in the percent of women in Communications in the dataset I provided them with, so these numbers are off by a tiny fraction from what you will get if you try to replicate. I didn't feel comfortable asking them to redo the entire thing, but the small error would not have changed the results significantly, and the tiny amount it would have changed them would have been in the direction of making the innate ability results more striking rather than less.

5. Although Leslie et al focused on women, they believe their results could also extend to why African-Americans are underrepresented compared to European-Americans and Asian-Americans in certain subjects. They theorize that European and Asian Americans, like men, are stereotyped as innately brilliant, but African-Americans, like women, lack this stereotype. I find this a bit off – after all, in the gender results, they contrasted the male "more innately brilliant" stereotype with the female "harder-working" stereotype, but African Americans suffer from a stereotype of not being hard-working, and Asian-Americans *do* have a stereotype of being hard-working, even more so than women. Anyway, this is only a mystery if you stick to Leslie et al's theory of stereotypes about perceived innate ability. Once you look at GRE Quantitative scores, you find that whites average 150.8, Asians average 153.9, and blacks average 143.7, and there's not much left to explain.

6. It's hard to correlate SAT scores with majors, because the SAT data is full of tiny vocational majors that throw off the results. For example, there are two hundred people in the country studying some form of manufacturing called "precision production", they're almost all male, and they have very low SAT scores. On the other hand, there are a few thousand people studying something called "family science", they're almost all women, and they also all have very low SAT scores. The shape of gender*major*SAT scores depends almost entirely on how many of these you count. I circumvented the entire problem by just counting the fields that approximately corresponded to the ones Leslie et al counted in their graduate-level study. I tried a few different analyses using different ways of deciding which fields to count, and as long as they were vaguely motivated by a desire to include academic subjects and not the vocational subjects with very low scores, they all came out about the same.

7. The argument that stereotypes cause boys to take more middle school and high school math classes than girls is somewhat argued against by the finding that actually girls take more middle school and high school math classes than boys. However, there are some contrary results; for example, boys are more likely than girls to take the AP Calculus test. This entire area gets so tangled up in differing levels of interest and ability and work-ethic that it's not worth it, at *my* level of interest and ability and work ethic, to try to work it out. The best I can say is that the gap appears by the time kids take the SAT in 11th grade.

8. I can’t help adding that I continue to believe that the stereotype threat literature looks like a null field which continues to exist only through publication bias and experimenter effects. The funnel plot shows a clear peak at “zero effect” and an asymmetry indicating a publication bias for positive results (for some discussion of why I like funnel plots, see here.) And a closer look at the individual research shows this really disturbing pattern of experiments by true believers finding positive effects, experiments by neutral parties and skeptics not finding them, replication attempts failing, and large real-world quasi-experiments turning up nothing – in a way very reminiscent of parapsychology. Although I am far from 100% sure, I would tentatively place my money on the entire idea of stereotype threat vanishing into the swamp of social psychology’s crisis of replication.

A Philosopher Walks Into A Coffee Shop

I have been really enjoying literarystarbucks.tumblr.com, which publishes complicated jokes about what famous authors and fictional characters order at Starbucks. I like it so much I wish I knew more great literature, so I could get more of the jokes.

Since the creators seem to be restricting themselves to the literary world, I hope they won't mind if I fail to resist the temptation to steal their technique for my own field of interest. Disclaimer: two of these are widely-known philosophy jokes and not original to me.

* * *

Parmenides goes up to the counter. “Same as always?” asks the barista. Parmenides nods.

* * *

Pythagoras goes up to the counter and orders a caffe Americano. “Mmmmm,” he says, tasting it. “How do you guys make such good coffee?” “It's made from the freshest beans,” the barista answers. Pythagoras screams and runs out of the store.

* * *

Thales goes up to the counter, says he's trying to break his caffeine habit, and orders a decaf. The barista hands it to him. He takes a sip and spits it out. “Yuck!” he says. “What is this, water?”

* * *

Gottfried Leibniz goes up to the counter and orders a muffin. The barista says he’s lucky since there is only one muffin left. Isaac Newton shoves his way up to the counter, saying Leibniz cut in line and he was first. Leibniz insists that he was first. The two of them come to blows.

* * *

Georg Wilhelm Friedrich Hegel goes up to the counter and gives a tremendously long custom order in German, specifying exactly how much of each sort of syrup he wants, various espresso shots, cream in exactly the right pattern, and a bunch of toppings, all added in a specific order at a specific temperature. The barista can’t follow him, so just gives up and hands him a small plain coffee. He walks away. The people behind him in line are very impressed with his apparent expertise, and they all order the same thing Hegel got. The barista gives each of them a small plain coffee, and they all remark on how delicious it tastes and what a remarkable coffee connoisseur that Hegel is. “The Hegel” becomes a new Starbucks special and is wildly popular for the next seventy years.

* * *

Socrates goes up to the counter. “What would you like?” asks the barista. “What would you recommend?” asks Socrates. “I would go with the pumpkin spice latte,” says the barista. “Why?” asks Socrates. “It’s seasonal,” she answers. “But why exactly is a seasonal drink better than a non-seasonal drink?” “Well,” said the barista, “I guess it helps to connect you to the rhythm of the changing seasons.” “But do you do other things to connect yourself to that rhythm?” asked Socrates. “Like wear seasonal clothing?

Or read seasonal books? If not, how come it’s only drinks that are seasonal?” “I’m not sure,” says the barista. “Think about it,” says Socrates, and leaves without getting anything.

* * *

Rene Descartes goes up to the counter. “I’ll have a scone,” he says. “Would you like juice with that?” asks the barista. “I think not,” says Descartes, and he ceases to exist.

* * *

Jean-Paul Sartre goes up to the counter. “What do you want?” asks the barista. Sartre thinks for a long while. “What *do*? I want?” he asks, and wanders off with a dazed look on his face.

* * *

William of Occam goes up to the counter. He orders a coffee.

* * *

Adam Smith goes up to the counter. “I’ll have a muffin,” he says. “Sorry,” says the barista, “but those two are fighting over the last muffin.” She points to Leibniz and Newton, who are still beating each other up. “I’ll pay $2 more than the sticker price, and you can keep the extra,” says Smith. The barista hands him the muffin.

* * *

John Buridan goes up to the counter and stares at the menu indecisively.

* * *

Ludwig Wittgenstein goes up to the counter. “I’ll have a small toffee mocha,” he says. “We don’t have small,” says the barista. “Then what sizes do you have?” “Just tall, grande, and venti.” “Then doesn’t that make ‘tall’ a ‘small’?” “We call it tall,” says the barista. Wittgenstein pounds his fist on the counter. “Tall has no meaning separate from the way it is used! You are just playing meaningless language games!” He storms out in a huff.

* * *

St. Anselm goes up to the counter and considers the greatest coffee of which it is possible to conceive. Since existence is more perfect than nonexistence, the coffee must exist. He brings it back to his table and drinks it.

* * *

Ayn Rand goes up to the counter. “What do you want?” asks the barista. “Exactly the relevant question. As a rational human being, it is my desires that are paramount. Since as a reasoning animal I have the power to choose, and since I am not bound by any demand to subordinate my desires to that of an outside party who wishes to use force or guilt to make me sacrifice my values to their values or to the values of some purely hypothetical collective, it is what I want that is imperative in this transaction. However, since I am dealing with you, and you are also a rational human being, under capitalism we have an opportunity to mutually satisfy our values in a way that leaves both of us richer and more fully human. You participate in the project of affirming my values by providing me with the coffee I want, and by paying you I am not only incentivizing you for the transaction, but giving you a chance to excel as a human being in the field of producing coffee. You do not produce the coffee because I am demanding it, or because I will use force against you if you do not, but because it most thoroughly represents your own values, particularly the value of creation. You would not make this coffee for me if it did not serve you in some way, and therefore by satisfying my desires you also reaffirm yourself. Insofar as you make inferior coffee, I will reject it and you will go bankrupt, but insofar as your coffee is truly excellent, a reflection of the excellence in your own soul and your achievement as a rationalist being, it will attract more people to your store, you will gain wealth, and you will be able to use that wealth further in pursuit of excellence as you, rather than some bureaucracy or collective, understand it. That is what it truly means to be a superior human." "Okay, but what do you want?" asks the barista. "Really I just wanted to give that speech," Rand says, and leaves.

* * *

Voltaire goes up to the counter and orders an espresso. He takes it and goes to his seat. The barista politely reminds him he has not yet paid. Voltaire stays seated, saying "I believe in freedom of espresso."

* * *

Thomas Malthus goes up to the counter and orders a muffin. The barista tells him somebody just took the last one. Malthus grumbles that the Starbucks is getting too crowded and there's never enough food for everybody.

* * *

Immanuel Kant goes up to the counter at exactly 8:14 AM. The barista has just finished making his iced cinnamon dolce latte, and hands it to him. He sips it for eight minutes and thirty seconds, then walks out the door.

* * *

Bertrand Russell goes up to the counter and orders the Hegel. He takes one sip, then exclaims “This just tastes like plain coffee! Why is everyone making such a big deal over it?”

* * *

Pierre Proudhon goes up to the counter and orders a Tazo Green Tea with toffee nut syrup, two espresso shots, and pumpkin spice mixed in. The barista warns him that this will taste terrible. “Pfah!” scoffs Proudhon. “Proper tea is theft!”

* * *

Sigmund Freud goes up to the counter. “I’ll have ass sex, presto,” he says. “What?!” asks the barista. “I said I’ll have iced espresso.” “Oh,” said the barista. “For a moment I misheard you.” “Yeah,” Freud tells her. “I fucked my mother. People say that.” “WHAT?!” asks the barista. “I said, all of the time other people say that.”

* * *

Jeremy Bentham goes up to the counter, holding a $50 bill. “What’s the cheapest drink you have?” he asks. “That would be our decaf roast, for only $1.99,” says the barista. “Good,” says Bentham and hands her the $50. “I’ll buy those for the next twenty-five people who show up.”

* * *

Patricia Churchland walks up to the counter and orders a latte. She sits down at a table and sips it. “Are you enjoying your beverage?” the barista asks. “No,” says Churchland.

* * *

Friedrich Nietzsche goes up to the counter. “I’ll have a scone,” he says. “Would you like juice with that?” asks the barista. “No, I hate juice,” says Nietzsche. The barista misinterprets him as saying “I hate Jews”, so she kills all the Jews in Europe.

The Parable Of The Talents

[Content note: scrupulosity and self-esteem triggers, IQ, brief discussion of weight and dieting. Not good for growth mindset.]

I.

I sometimes blog about research into IQ and human intelligence. I think most readers of this blog already know IQ is 50% to 80% heritable, and that it's so important for intellectual pursuits that eminent scientists in some fields have average IQs around 150 to 160. Since IQ this high only appears in 1/10,000 people or so, it beggars coincidence to believe this represents anything but a very strong filter for IQ (or something correlated with it) in reaching that level. If you saw a group of dozens of people who were 7'0 tall on average, you'd assume it was a basketball team or some other group selected for height, not a bunch of botanists who were all very tall by coincidence.

A lot of people find this pretty depressing. Some worry that taking it seriously might damage the "growth mindset" people need to fully actualize their potential. This is important and I want to discuss it eventually, but not now. What I want to discuss now is people who feel *personally* depressed. For example, a comment from last week:

> I'm sorry to leave self a self absorbed comment, but reading this really upset me and I just need to get this off my chest…How is a person supposed to stay sane in a culture that prizes intelligence above everything else – especially if, as Scott suggests, Human Intelligence Really Is the Key to the Future – when they themselves are not particularly intelligent and, apparently, have no potential to ever become intelligent? Right now I basically feel like pond scum.

I hear these kinds of responses every so often, so I should probably learn to expect them. I never do. They seem to me precisely backwards. There’s a moral gulf here, and I want to throw stories and intuitions at it until enough of them pile up at the bottom to make a passable bridge. But first, a comparison:

Some people think body weight is biologically/genetically determined. Other people think it’s based purely on willpower – how strictly you diet, how much you can bring yourself to exercise. These people get into some pretty acrimonious debates.

Overweight people, and especially people who feel unfairly stigmatized for being overweight, tend to cluster on the biologically determined side. And although not all believers in complete voluntary control of weight are mean to fat people, the people who are mean to fat people pretty much all insist that weight is voluntary and easily changeable.

Although there’s a lot of debate over the science here, there seems to be broad agreement on both sides that the more compassionate, sympathetic, progressive position, the position promoted by the kind of people who are really worried about stigma and self-esteem, is that weight is biologically determined.

And the same is true of mental illness. Sometimes I see depressed patients whose families *really* don’t get it. They say “Sure, my daughter feels down, but she needs to realize that’s no excuse for shirking her responsibilities. She needs to just pick herself up and get on with her life.” On the other hand, most depressed people say that their depression is more fundamental than that, not a thing that can be overcome by willpower, certainly not a thing you can just ‘shake off’.

Once again, the compassionate/sympathetic/progressive side of the debate is that depression is something like biological, and cannot easily be overcome with willpower and hard work.

One more example of this pattern. There are frequent political debates in which conservatives (or straw conservatives) argue that financial success is the result of hard work, so poor people are just too lazy to get out of poverty. Then a liberal (or straw liberal) protests that hard work has nothing to do with it, success is determined by accidents of birth like who your parents are and what your skin color is et cetera, so the poor are blameless in their own predicament.

I'm oversimplifying things, but again the compassionate/sympathetic/progressive side of the debate – and the side endorsed by many of the poor themselves – is supposed to be that success is due to accidents of birth, and the less compassionate side is that success depends on hard work and perseverance and grit and willpower.

The obvious pattern is that attributing outcomes to things like genes, biology, and accidents of birth is kind and sympathetic. Attributing them to who works harder and who's "really trying" can stigmatize people who end up with bad outcomes and is generally viewed as Not A Nice Thing To Do.

And the weird thing, the thing I've never understood, is that intellectual achievement is the one domain that breaks this pattern.

Here it's would-be hard-headed conservatives arguing that intellectual greatness comes from genetics and the accidents of birth and demanding we “accept” this “unpleasant truth”.

And it’s would-be compassionate progressives who are insisting that no, it depends on who works harder, claiming anybody can be brilliant if they really try, warning us not to “stigmatize” the less intelligent as “genetically inferior”.

I can come up with a few explanations for the sudden switch, but none of them are very principled and none of them, to me, seem to break the fundamental symmetry of the situation. I choose to maintain consistency by preserving the belief that overweight people, depressed people, and poor people aren’t fully to blame for their situation – and neither are unintelligent people. It’s accidents of birth all the way down. Intelligence is mostly genetic and determined at birth – and we’ve already determined in every other sphere that “mostly genetic and determined at birth” means you don’t have to feel bad if you got the short end of the stick.

Consider for a moment Srinivasa Ramanujan, one of the greatest mathematicians of all time. He grew up in poverty in a one-room house in small-town India. He taught himself mathematics by borrowing books from local college students and working through the problems on his own until he reached the end of the solveable ones and had nowhere else to go but inventing ways to solve the unsolveable ones.

There are a lot of poor people in the United States today whose life circumstances prevented their parents from reading books to them as a child, prevented them from getting into the best schools, prevented them from attending college, et cetera. And pretty much all of those people *still* got more educational opportunities than Ramanujan did.

And from there we can go in one of two directions. First, we can say that a lot of intelligence is innate, that Ramanujan was a genius, and that we mortals cannot be expected to replicate his accomplishments.

Or second, we can say those poor people are *just not trying hard enough*.

Take “innate ability” out of the picture, and if you meet a poor person on the street begging for food, saying he never had a chance, your reply must be “Well, if you’d just borrowed a couple of math textbooks from the local library at age 12, you would have been a Fields Medalist by now. I hear that pays pretty well.”

The best reason *not* to say that is that we view Ramanujan as intellectually gifted. But the very phrase tells us where we should classify that belief. Ramanujan’s genius is a “gift” in much the same way your parents giving you a trust fund on your eighteenth birthday is a “gift”, and it should be weighted accordingly in the moral calculus.

II.

I shouldn’t pretend I’m worried about this for the sake of the poor. I’m worried for *me*.

My last IQ-ish test was my SATs in high school. I got a perfect score in Verbal, and a good-but-not-great score in Math.

And in high school English, I got A++s in all my classes, Principal’s Gold Medals, 100%s on tests, first prize in various state-wide essay contests, etc. In Math, I just barely by the skin of my teeth scraped together a pass in Calculus with a C-.

Every time I won some kind of prize in English my parents would praise me and say I was good and should feel good. My teachers would hold me up as an example and say other kids should try to be more like me. Meanwhile, when I would bring home a report card with a C- in math, my parents would have concerned faces and tell me they were disappointed and I wasn't living up to my potential and I needed to work harder et cetera.

And *I don't know which part bothered me more*.

Every time I was held up as an example in English class, I wanted to crawl under a rock and die. I didn't do it! I didn't study at all, half the time I did the homework in the car on the way to school, those essays for the statewide competition were thrown together on a lark without a trace of real effort. To praise me for any of it seemed and still seems utterly unjust.

On the other hand, to this day I believe I deserve a fricking *statue* for getting a C- in Calculus I. It should be in the center of the schoolyard, and have a plaque saying something like "Scott Alexander, who by making a herculean effort managed to pass Calculus I, even though they kept throwing random things after the little curly S sign and pretending it made sense."

And without some notion of innate ability, I don't know what to do with this experience. I don't want to have to accept the blame for being a lazy person who just didn't try hard enough in Math. But I *really* don't want to have to accept the credit for being a virtuous and studious English student who worked harder than his peers. I *know* there were people who worked harder than I did in English, who poured their heart and soul into that course – and who still got Cs and Ds. To deny innate ability is to devalue their efforts and sacrifice, while simultaneously giving me credit I don’t deserve.

Meanwhile, there were some students who did better than I did in Math with seemingly zero effort. I didn’t begrudge those students. But if they’d started trying to say they had exactly the same level of innate ability as I did, and the only difference was *they* were trying while *I* was slacking off, then I sure as hell would have begrudged them. Especially if I knew they were lazing around on the beach while I was poring over a textbook.

I tend to think of social norms as contracts bargained between different groups. In the case of attitudes towards intelligence, those two groups are smart people and dumb people. Since I was both at once, I got to make the bargain with myself, which simplified the bargaining process immensely. The deal I came up with was that I wasn’t going to beat myself up over the areas I was bad at, but I also didn’t get to become too cocky about the areas I was good at. It was all genetic luck of the draw either way. In the meantime, I would try to press as hard as I could to exploit my strengths and cover up my deficiencies. So far I’ve found this to be a really healthy way of treating myself, and it’s the way I try to treat others as well.

III.

The theme continues to be “Scott Relives His Childhood Inadequacies”. So:

When I was 6 and my brother was 4, our mom decided that as an Overachieving Jewish Mother she was contractually obligated to make both of us learn to play piano. She enrolled me in a Yamaha introductory piano class, and my younger brother in a Yamaha ‘cute little kids bang on the keyboard’ class.

A little while later, I noticed that my brother was now with me in my Introductory Piano class.

A little while later, I noticed that my brother was now by far the best student in my Introductory Piano Class, even though he had just started and was two or three years younger than anyone else there.

A little while later, Yamaha USA flew him to Japan to show him off before the Yamaha corporate honchos there.

Well, one thing led to another, and my brother won several international piano competitions, got a professorship in music at age 25, and now routinely gets news articles written about him calling him “among the top musicians of his generation”.

Meanwhile, I was always a mediocre student at Yamaha. When the time came to try an instrument in elementary school, I went with the violin to see if maybe I’d find it more to my tastes than the piano. I was quickly sorted into the remedial class because I couldn’t figure out how to make my instrument stop sounding like a wounded cat. After a year or so of this, I decided to switch to fulfilling my music requirement through a choir, and everyone who’d had to listen to me breathed a sigh of relief.

Every so often I wonder if somewhere deep inside me there is the potential to be “among the top musicians of my generation.” I try to recollect whether my brother practiced harder than I did. My memories are hazy, but I don’t think he practiced *much* harder until well after his career as a child prodigy had taken off. The cycle seemed to be that every time he practiced, things came fluidly to him and he would produce beautiful music and everyone would be amazed. And this must have felt great, and incentivized him to practice more, and that made him even better, so that the beautiful music came even more fluidly, and the praise became more effusive, until eventually he chose a full-time career in music and became amazing. Meanwhile, when I started practicing it always sounded like wounded cats, and I would get very cautious praise like “Good job, Scott, it sounded like that cat was hurt a little less badly than usual,” and it made me frustrated, and want to practice less, which made me even worse, until eventually I quit in disgust.

On the other hand, I know people who want to get good at writing, and make a mighty resolution to write two hundred words a day every day, and then after the first week they find it’s too annoying and give up. These people think I’m amazing, and why shouldn’t they? I’ve written a few hundred to a few thousand words pretty much every day for the past ten years.

But as I’ve said before, this has taken exactly zero willpower. It’s more that I can’t stop even if I want to. Part of that is probably that when I write, I feel really good about having expressed exactly what it was I meant to say. Lots of people read it, they comment, they praise me, I feel good, I’m encouraged to keep writing, and it’s exactly the same virtuous cycle as my brother got from his piano practice.

And so I think it would be *too easy* to say something like “There’s no innate component at all. Your brother practiced piano really hard but almost never writes. You write all the time, but wimped out of practicing piano. So what do you expect? You both got what you deserved.”

I tried to practice piano as hard as he did. I really tried. But every moment was a struggle. I could keep it up for a while, and then we'd go on vacation, and there'd be no piano easily available, and I would be breathing a sigh of relief at having a ready-made excuse, and he'd be heading off to look for a piano somewhere to practice on. Meanwhile, I am writing this post in short breaks between running around hospital corridors responding to psychiatric emergencies, and there's probably someone very impressed with that, someone saying "But you had such a great excuse to get out of your writing practice!"

I dunno. But I don't think of myself as working hard at any of the things I am good at, in the sense of "exerting vast willpower to force myself kicking and screaming to do them". It's possible I *do* work hard, and that an outside observer would accuse me of eliding how hard I work, but it's not a conscious elision and I don't feel that way from the inside.

Ramanujan worked very hard at math. But I don't think he thought of it as work. He obtained a scholarship to the local college, but dropped out almost immediately because he couldn't make himself study any subject other than math. Then he got accepted to another college, and dropped out *again* because they made him study non-mathematical subjects and he failed a physiology class. Then he nearly starved to death because he had no money and no scholarship. To me, this doesn't sound like a person who just happens to be very hard-working; if he had the ability to study other subjects he would have, for no reason other than that it would have allowed him to stay in college so he could keep studying math. It seems to me that in some sense Ramanujan was *incapable* of putting hard work into non-math subjects.

I really wanted to learn math and failed, but I did graduate with honors from medical school. Ramanujan really wanted to learn physiology and failed, but he did become one of history's great mathematicians. So which one of us was the hard worker?

People used to ask me for writing advice. And I, in all earnestness, would say "Just transcribe your thoughts onto paper exactly like they sound in your head." It turns out that doesn't work for other people. Maybe it doesn't work for me either, and it just feels like it does.

But you know what? When asked about one of his discoveries, a method of simplifying a very difficult problem to a continued fraction, Ramanujan described his thought process as: "It is simple. The minute I heard the problem, I knew that the answer was a continued fraction. 'Which continued fraction?' I asked myself. Then the answer came to my mind".

And again, maybe that's just how it feels to him, and the real answer is "study math so hard that you flunk out of college twice, and eventually you develop so much intuition that you can solve problems without thinking about them."

(or maybe the real answer is "have dreams where obscure Hindu gods appear to you as drops of blood and reveal mathematical formulae". Ramanujan was weird).

But I *still* feel like there's something going on here where the solution to me being bad at math and piano isn't just "sweat blood and push through your brain's aversion to these subjects until you make it stick". When I read biographies of Ramanujan and other famous mathematicians, there's no sense that they ever had to do that with math. When I talk to my brother, I never get a sense that he had to do that with piano. And if I am good enough at writing to qualify to have an opinion on being good at things, then I don’t feel like I ever went through that process myself.

So this too is part of my deal with myself. I’ll try to do my best at things, but if there’s something I really hate, something where I have to go uphill every step of the way, then it’s okay to admit mediocrity. I won’t beat myself up for not forcing myself kicking and screaming to practice piano. And in return I won’t become too cocky about practicing writing a lot. It’s probably some kind of luck of the draw either way.

IV.

I said before that this wasn’t just about poor people, it was about me being selfishly worried for my own sake. I think I might have given the mistaken impression that I merely need to justify to myself why I can’t get an A in math or play the piano. But it’s much worse than that.

The rationalist community tends to get a lot of high-scrupulosity people, people who tend to beat themselves up for not doing more than they are. It’s why I push giving 10% to charity, not as some kind of amazing stretch goal that we need to guilt people into doing, but as a crutch, a sort of “don’t worry, you’re still okay if you only give ten percent”. It’s why there’s so much emphasis on “heroic responsibility” and how you, yes you, have to solve all the world’s problems personally. It’s why I see red when anyone accuses us of entitlement, since it goes about as well as calling an anorexic person fat.

And we really aren’t doing ourselves any favors. For example, Nick Bostrom writes:

> Searching for a cure for aging is not just a nice thing that we should perhaps one day get around to. It is an

> urgent, screaming moral imperative. The sooner we start a focused research program, the sooner we will get results. It matters if we get the cure in 25 years rather than in 24 years: a population greater than that of Canada would die as a result.

If that bothers you, you *definitely* shouldn't read Astronomical Waste.

Yet here I am, not doing anti-aging research. Why not?

Because I tried doing biology research a few times and it was really hard and made me miserable. You know how in every science class, when the teacher says "Okay, pour the white chemical into the grey chemical, and notice how it turns green and begins to bubble," there's always one student who pours the white chemical into the grey chemical, and it just forms a greyish-white mixture and sits there? That was me. I hated it, I didn't have the dexterity or the precision of mind to do it well, and when I finally finished my required experimental science classes I was happy never to think about it again. Even the abstract intellectual part of it – the one where you go through data about genes and ligands and receptors in supercentenarians and shake it until data comes out – requires exactly the kind of math skills that I don't have.

Insofar as this is a matter of innate aptitude – some people are cut out for biology research and I'm not one of them – all is well, and my decision to get a job I'm good at instead is entirely justified.

But insofar as there's no such thing as innate aptitude, just hard work and grit – then by not being gritty enough, I'm a monster who's complicit in the death of a population greater than that of Canada.

Insofar as there’s no such thing as innate aptitude, I have *no excuse* for not being Aubrey de Grey. Or if Aubrey de Grey doesn’t impress you much, Norman Borlaug. Or if you don’t know who either of those two people are, Elon Musk.

I once heard a friend, upon his first use of modafinil, wonder aloud if the way they felt on that stimulant was the way Elon Musk felt all the time. That tied a lot of things together for me, gave me an intuitive understanding of what it might “feel like from the inside” to be Elon Musk. And it gave me a good tool to discuss biological variation with. Most of us agree that people on stimulants can perform in ways it’s difficult for people off stimulants to match. Most of us agree that there’s nothing magical about stimulants, just changes to the levels of dopamine, histamine, norepinephrine et cetera in the brain. And most of us agree there’s a lot of natural variation in these chemicals anyway. So “me on stimulants is that guy’s normal” seems like a good way of cutting through some of the philosophical difficulties around this issue.

…which is all kind of a big tangent. The point I want to make is that for me, what’s at stake in talking about natural variations in ability isn’t just whether I have to feel like a failure for not getting an A in high school calculus, or not being as good at music as my brother. It’s whether I’m a failure for not being Elon Musk. Specifically, it’s whether I can say “No, I’m really not cut out to be Elon Musk” and go do something else I’m better at without worrying that I’m killing everyone in Canada.

V.

The proverb says: “Everyone has somebody better off than they are and somebody worse off than they are, with two exceptions.” When we accept that we’re all in the “not Elon

Musk” boat together (with one exception) a lot of the status games around innate ability start to seem less important.

Every so often an overly kind commenter here praises my intelligence and says they feel intellectually inadequate compared to me, that they wish they could be at my level. But at my level, I spend my time feeling intellectually inadequate compared to Scott Aaronson. Scott Aaronson describes feeling “in awe” of Terence Tao and frequently struggling to understand him. Terence Tao – well, I don’t know if he’s religious, but maybe he feels intellectually inadequate compared to God. And God feels intellectually inadequate compared to John von Neumann.

So there’s not much point in me feeling inadequate compared to my brother, because even if I was as good at music as my brother, I’d probably just feel inadequate for not being Mozart.

And asking “Well what if you just worked harder?” can elide small distinctions, but not bigger ones. If my only goal is short-term preservation of my self-esteem, I can imagine that if only things had gone a little differently I could have practiced more and ended up as talented as my brother. It’s a lot harder for me to imagine the course of events where I do something different and become Mozart. Only one in a billion people reach a Mozart level of achievement; why would it be me?

If I loved music for its own sake and wanted to be a talented musician so I could express the melodies dancing within my heart, then none of this matters. But insofar as I want to be good at music because *I feel bad that other people are better than me at music*, that’s a road without an end.

This is also how I feel of when some people on this blog complain they feel dumb for not being as smart as some of the other commenters on this blog.

I happen to have all of your IQ scores in a spreadsheet right here (remember that survey you took?). Not a single person is below the population average. The first percentile for IQ here – the one such that 1% of respondents are lower and 99% of respondents are higher – is – corresponds to the 85th percentile of the general population. So even if you're in the first percentile here, you're still pretty high up in the broader scheme of things.

At that point we're back on the road without end. I am pretty sure we can raise your IQ as much as you want and you will *still* feel like pond scum. If we raise it twenty points, you'll try reading *Quantum Computing since Democritus* and feel like pond scum. If we raise it forty, you'll just go to Terence Tao's blog and feel like pond scum there. Maybe if you were literally the highest-IQ person in the entire world you would feel good about yourself, but any system where only one person in the world is allowed to feel good about themselves at a time *is a bad system*.

People say we should stop talking about ability differences so that stupid people don't feel bad. I say that there's more than enough room for *everybody* to feel bad, smart and stupid alike, and not talking about it won't help. What will help is fundamentally uncoupling perception of intelligence from perception of self-worth.

I work with psychiatric patients who tend to have cognitive difficulties. Starting out in the Detroit ghetto doesn't do them any favors, and then they get conditions like bipolar disorder and schizophrenia that actively lower IQ for poorly understood neurological reasons.

The standard psychiatric evaluation includes an assessment of cognitive ability; the one I use is a quick test with three questions. The questions are – "What is 100 minus 7?", "What do an apple and an orange have in common?", and "Remember these three words for one minute, then repeat them back to me: house, blue, and tulip".

There are a lot of people – and I don't mean floridly psychotic people who don't know their own name, I mean ordinary reasonable people just like you and me – who can't answer these questions. And we know why they can't answer these questions, and it is pretty darned biological.

And if our answer to "I feel dumb and worthless because my IQ isn't high enough" is "don't worry, you're not worthless, I'm sure you can be a great scientist if you just try hard enough", then we are implicitly throwing under the bus all of these people who are *definitely* not going to be great scientists no matter how hard they try. Talking about trying harder can obfuscate the little differences, but once we're talking about the homeless schizophrenic guy from Detroit who can't tell me 100 minus 7 to save his life, you can't just magic the problem away with a wave of your hand and say "I'm sure he can be the next Ramanujan if he keeps a positive attitude!" You either need to condemn him as worthless *or else stop fricking tying worth to innate intellectual ability*.

This is getting pretty close to what I was talking about in my post on burdens. When I get a suicidal patient who thinks they're a burden on society, it's nice to be able to point out ten important things they've done for society recently and prove them wrong. But sometimes it's not that easy, and the only thing you can say is "f#@k that s#!t". Yes, society has organized itself in a way that excludes and impoverishes a bunch of people who could have been perfectly happy in the state of nature picking berries and hunting aurochs. It’s not your fault, and if they’re going to give you compensation *you take it*. And we had better make this perfectly clear now, so that when everything becomes automated and run by robots and we’re *all* behind the curve, everybody agrees that us continuing to exist is still okay.

Likewise with intellectual ability. When someone feels sad because they can’t be a great scientist, it is nice to be able to point out all of their intellectual strengths and tell them “Yes you can, if only you put your mind to it!” But this is often not true. At that point you have to say “f@#k it” and tell them to stop tying their self-worth to being a great scientist. And we had better establish that now, before transhumanists succeed in creating superintelligence and we *all* have to come to terms with our intellectual inferiority.

VI.

But I think the situation can also be somewhat rosier than that.

Ozy once told me that the law of comparative advantage was one of the most inspirational things they had ever read. This was sufficiently strange that I demanded an explanation.

Ozy said that it proves *everyone can contribute*. Even if you are worse than everyone else at everything, you can still participate in global trade and other people will pay you money. It may not be very much money, but it will be some, and it will be a measure of how your actions are making other people better off and they are grateful for your existence.

(in real life this doesn't work for a couple of reasons, but who cares about real life when we have *a theory*?)

After some thought, I was also inspired by this.

I'm never going to be a great mathematician or Elon Musk. But if I pursue my comparative advantage, which right now is medicine, I can still make money. And if I feel like it, I can donate it to mathematics research. Or anti-aging research. Or the same people Elon Musk donates *his* money to. They will use it to hire smart people with important talents that I lack, and I will be at least partially responsible for those people's successes.

If I had an IQ of 70, I think I would still want to pursue my comparative advantage – even if that was ditch-digging, or whatever, and donate that money to important causes. It might not be very much money, but it would be *some*.

Our modern word "talent" comes from the Greek word *talenton*, a certain amount of precious metal sometimes used as a denomination of money. The etymology passes through a parable of Jesus'. A master calls three servants to him and gives the first five talents, the second two talents, and the third one talent. The first two servants invest the money and double it. The third literally buries it in a hole. The master comes back later and praises the first two servants, but sends the third servant to Hell (metaphor? what metaphor?).

Various people have come up with various interpretations, but the most popular says that God gives all of us different amounts of resources, and He will judge us based on how well we use these resources rather than on how many He gave us. It would be stupid to give your first servant five loads of silver, then your second servant two loads of silver, then immediately start chewing out the second servant for having less silver than the first one. And if both servants invested their silver wisely, it would be silly to chew out the second one for ending up with less profit when he started with less seed capital. The moral seems to be that if you take what God gives you and use it wisely, you’re fine.

The modern word “talent” comes from this parable. It implies “a thing God has given you which you can invest and give back”.

So if I were a ditch-digger, I think I would dig ditches, donate a portion of the small amount I made, and trust that I had done what I could with the talents I was given.

VII.

The Jews *also* talk about how God judges you for your gifts. Rabbi Zusya once said that when he died, he wasn’t worried that God would ask him “Why weren’t you Moses?” or “Why weren’t you Solomon?” But he did worry that God might ask “Why weren’t you Rabbi Zusya?”

And this is part of why it’s important for me to believe in innate ability, and especially differences in innate ability. If everything comes down to hard work and positive attitude, then God has every right to ask me “Why weren’t you Srinivasa Ramanujan?” or “Why weren’t you Elon Musk?”

If everyone is legitimately a different person with a different brain and different talents and abilities, then all God gets to ask me is whether or not I was Scott Alexander.

This seems like a gratifyingly low bar.

[more to come on this subject later]

Talents Part 2: Attitude vs. Altitude

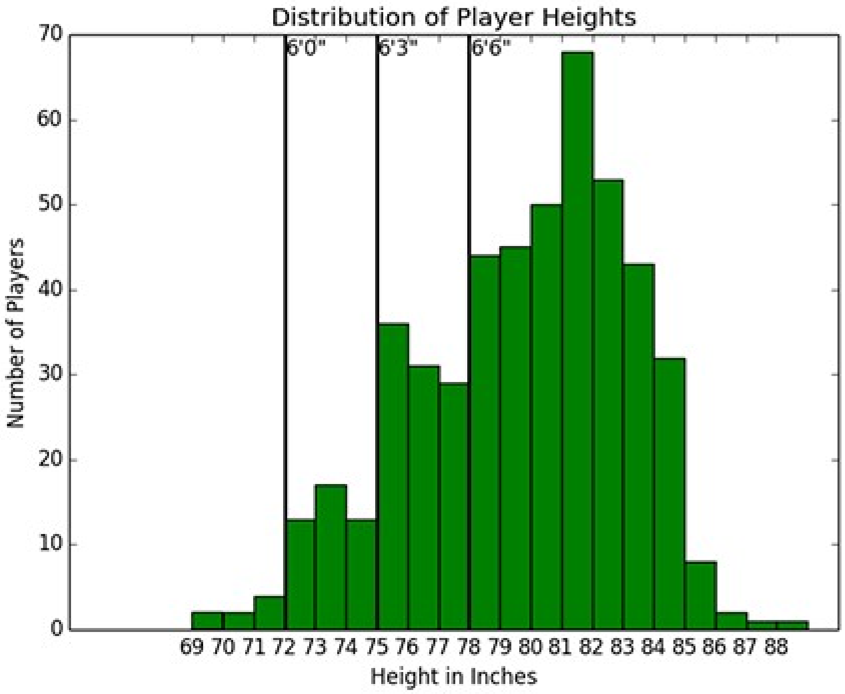
[Content note: scrupulosity and self-esteem triggers, IQ. Potentially not good for growth mindset.]

I.

The average eminent theoretical physicist has an IQ of 150-160. The average NBA player has a height of 6’7. Both of these are a little over three standard deviations above their respective mean. Since z-scores are magic and let us compare unlike domains, we conclude that eminent theoretical physicists are about as smart as pro basketball players are tall.

Any time people talk about intelligence, height is a natural sanity check. It’s another strongly heritable polygenic trait which is nevertheless susceptible to environmental influences, and which varies in a normal distribution across the population – but which has yet to accrete the same kind of cloud of confusion around it that IQ has.

So let’s see what we can learn from the height distribution of every player in the NBA: (source)



The first thing we notice is that nobody shorter than average (ie 5’9) gets in. This is not true eternally and forever – Wikipedia’s [List Of Shortest NBA Players] names 25 NBA players in history (out of maybe 5000 total) who were below average stature. The all-time record holder is Tyrone “Muggsy” Bogues at 5’3, who was actually pretty good (though not above exploiting his distinction for publicity: his autobiography is called *In the Land of Giants*). But Bogues was a spectacular outlier. If we’re going to stick to our resolution to use the histogram we’ve got, the contribution of the left half of the height bell curve is precisely zero.

We can do some very rough Fermi calculations for the next couple standard deviations.

34% of the US male population is 5’9 to 6’0, so about 54 million men. There are 5 NBA players in this band. So about one in every 11 million people of this height is in the NBA.

13.5% of the US male population is 6’0 to 6’3, so about 21 million men. There are 40 NBA players in this band. So about one in every 500,000 people of this height is in the NBA.

2.35% of the US male population is 6’3 to 6’6, so about 4 million men. There are 95 NBA players in this band. So about one in every 40,000 people of this height is in the NBA.

0.15% of the US male population is 6’6 to 6’9, so about 200,000 men. There are 130 NBA players in this band. So about one in every 1,500 people of this height is in the NBA.

0.003% of the US male population is 6’9 to 7’0, so about 5,000 men. There are 160 NBA players in this band. So about one in every 30 people of this height is in the NBA.

0.00002% of the US male population is 7’0 to 7’3, which corresponds to about 45 men. There are 40 NBA players in this band. So about 8 out of 9 people of this height are in the…wait, no, that can’t be right.

Sports Illustrated‘s Dan Diamond does a similar analysis and gets broadly similar results. But he adds several complicating factors. First, at this height people with very rare medical conditions and pituitary tumors start taking over from normal variation, so we lose our ability to derive a census a priori with pure math. Second of all, at this height talent scouts comb the world for suitably tall foreigners and import them, so we can no longer assume we’re drawing from the pool of tall Americans. And third of all, the same way you round up so that you’re 6’0 on OKCupid, NBA players round up so that they’re 7’0 on their official stats.

He concludes that most likely about 17% of seven-foot-tall young men in America are in the NBA. This might still be an overestimate, but is probably in the right ballpark. Forbes Magazine writes that Being Seven Feet Tall Is The Fastest Way To Get Rich In America and quotes a talent scout who says “that he’ll ‘check up on anyone over 7 feet that’s breathing.'” Given that a lot of people this height are unhealthy or uninterested, it might not be much of an exaggeration to say that if you’re 7’3 and have *any* interest in basketball, you’ve got better than even odds of going pro.

But why stop there? 0.000000001% of men are 7’3 to 7’6. Given the size of the American male population, there shouldn’t be a single person in the US with this height, though in practice a few people with endocrine abnormalities make the cut. There will, however, be a couple of healthy people of this height in the *world* population. There is one person currently in the NBA above 7’3, and he is a Tanzanian native discovered by talent scouts.

WE’VE GOT TO GO ~~DEEPER~~ TALLER! 0.0000000000001% of the population is 7’6 to 7’9. Statistically, there should not be a single man this tall in the entire world. No NBA players are currently this tall. But Yao Ming, who retired four years ago, was 7’6 exactly. He was the product of a Maoist breeding program specifically aimed at producing tall people to play basketball. You can break a *lot* of statistical laws if you have breeding programs and flexible ethics. Also if you have pituitary tumors, as the remainder of the List Of Tallest People reminds us. It looks like a little over half of the living people in this height band have played professional basketball at some point.

There are people taller than 7’9, but a lot of them have trouble standing without support, which somewhat decreases basketball ability.

On the plus side, you can do all sorts of awesome things, like take a picture standing next to the world's shortest man, or…okay, pretty much just that.

II.

This offers us an opportunity to compare our intuitions about intellect to our intuitions about basketball. I recommend taking that opportunity. Basketball isn't mysterious, it isn't politicized, and it isn't tied up with our notions of self-worth. As a result, our intuitions about basketball are crystal clear.

If somebody said "Height plays no role in your success as a basketball player," we'd look at them funny. Literally nobody shorter than average is currently in the NBA. Your odds at average height are only one in ten million. But if you're 7'0, you can pretty much walk right in.

But if somebody said "Well, success in basketball is clearly dependent on height, that means there's no point in practicing, after all you're not going to grow any taller," we’d look at *them* funny too. I don’t know much about sports, but I predict a short guy who’s been playing for years can defeat an untrained giant every time. And of course if two giants go up against each other, experience will be decisive.

If someone found a 5’8 kid who really liked playing hoops in his backyard and was pretty good at it, and they told him “You’re a moron, you can never succeed at basketball because of your height, give up,” that person would be a jerk. The kid is enjoying himself, there’s no point in insulting him about it, and if his goal is just to end up as a college athlete or a minor leaguer, then with hard work he could certainly make it. Even if he didn’t, he might be able to apply what he’d learned to a related area like coaching or reporting.

But if that same kid wanted to go deeply into debt to attend a basketball training camp, and he’ll count it a failure if he achieves anything less than NBA superstardom, you should probably warn him that his hopes aren’t very realistic and that maybe he should lower his standards or pick a sport more suited to his body type or try another line of work entirely.

If somebody who was 6’6 complained that they’d never be able to beat the 7’0 players on the other side, we would tell them to brighten up. After all, Michael Jordan was “only” 6’6, and he’s maybe the greatest basketball player of all time, even though he often had to face off against people taller than himself.

But if a team made entirely of 6’6 players faced off against a team entirely of 7’0 players, and both of them were really motivated and practiced really hard and had great coaches, I know who I would bet on.

Most important, people all over the world innocently enjoy playing basketball, and they are right to do so. Everyone knows that taller players have an advantage, no one's denying it. But at the levels most people play at, moderate height differences are surmountable by differences in training and technique, and large height differences are so rare as to not come up very often. Yes, the pro leagues are a different story. Yet somehow the 5'8 kid who scores a three-pointer still manages to feel good about himself. And we can be very impressed by Kobe Bryant and LeBron James, praise them for their determination and technique and competitive spirit, all while acknowledging that it's not a coincidence that the one is 6'6 and the other 6'9 and neither of these is a normal-person height.

All of this is crystal clear when we're talking about basketball. But as soon as we switch back to talking about intelligence, we're shouting at each other: "YOU SAY RAMANUJAN WAS REALLY HARD-WORKING, BUT I THINK HE HAD HIGH IQ, SO THERE!" Or "WELL IF SUCCESS COMES FROM INNATE TALENT, I GUESS I'LL JUST NEVER STUDY AGAIN."

Luke Muehlhauser liked to call his philosophy of religion "common sense atheism", meaning that he wanted to treat the question of God with the same "common" reasoning that he used for every other question. If we don't see a tiger in front yard, we don't say "Since it's impossible to prove a negative, I can at best be agnostic about the existence of a tiger", we say "I guess there's probably not a tiger."

Likewise, if we can just apply the same common reasoning we use for normal everyday activities like basketball to the question of intelligence, we might find it's not so complicated and scary after all.

Which is not to say that it’s not *important*. After all, the *other* analogy between intelligence and basketball talent is that they’re both skills we need to cultivate at the highest levels if we want to save the world when it is threatened by dangerous future technology we can barely comprehend.

[still more on this subject later, but not immediately]

Practically-A-Book Review: Dying To Be Free

I am the last person with a right to complain about Internet articles being too long. But if I did have that right, I think I would exercise it on Dying To Be Free, the Huffington Post's 20,000-word article on the current state of heroin addiction treatment. I feel like it could have been about a quarter the size without losing much.

It's too bad that most people will probably shy away from reading it, because it gets a lot of stuff *really* right.

The article's thesis is also its subtitle: "There's a treatment for heroin addiction that actually works; why aren't we using it?" To save you the obligatory introductory human interest story: that treatment is suboxone. Its active ingredient is the drug buprenorphine, which is kind of like a safer version of methadone. Suboxone is slow-acting, gentle, doesn't really get people high, and is pretty safe as long as you don't go mixing it with weird stuff. People on suboxone don't experience opiate withdrawal and have greatly decreased cravings for heroin. I work at a hospital that's an area leader in suboxone prescription, I've gotten to see it in action, and it's literally a life-saver.

Conventional heroin treatment is abysmal. Rehab centers aren't licensed or regulated and most have little interest in being evidence-based. Many are associated with churches or weird quasi-religious groups like Alcoholics Anonymous. They don't necessarily have doctors or psychologists, and some actively mistrust them. All of this I knew. What I didn't know until reading the article was that – well, it's not just that some of them try to brainwash addicts. It's more that some of them try to cargo cult brainwashing, do the sorts of things that sound like brainwashing to *them*, without really knowing how brainwashing works assuming it's even a coherent goal to aspire to. Their concept of brainwashing is mostly just creating a really unpleasant environment, yelling at people a lot, enforcing intentionally over-strict rules, and in some cases even having struggle-session-type-things where everyone in the group sits in a circle, scream at the other patients, and tell them they're terrible and disgusting. There's a strong culture of accusing anyone who questions or balks at any of it of just being an addict, or "not really wanting to quit".

I have no problem with "tough love" when it works, but in this case it doesn't. Rehab programs make every effort to obfuscate their effectiveness statistics – I blogged about this before in Part II here – but the best guesses by outside observers is that for a lot of them about 80% to 90% of their graduates relapse within a couple of years. Even this paints too rosy a picture, because it excludes the people who gave up halfway through.

Suboxone treatment isn't perfect, and relapse is still a big problem, but it's a heck of a lot better than most rehabs. Suboxone gives people their dose of opiate and mostly removes the biological half of addiction. There's still the psychological half of addiction – whatever it was that made people want to get high in the first place – but people have a much easier time dealing with that after the biological imperative to get a new dose is gone. Almost all clinical trials have found treatment with methadone or suboxone to be more effective than traditional rehab. Even Cochrane Review, which is notorious for never giving a straight answer to anything besides "more evidence is needed", agrees that methadone and suboxone are effective treatments.

Some people stay on suboxone forever and do just fine – it has few side effects and doesn't interfere with functioning. Other people stay on it until they reach a point in their lives when they feel ready to come off, then taper down slowly under medical supervision, often with good success. It's a good medication, and the growing suspicion it might help treat depression is just icing on the cake.

There are two big roadblocks to wider use of suboxone, and both are enraging.

The first roadblock is the #@$%ing government. They are worried that suboxone, being an opiate, might be addictive, and so doctors might turn into drug pushers. So suboxone is possibly the most highly regulated drug in the United States. If I want to give out OxyContin like candy, I have no limits but the number of pages on my prescription pad. If I want to prescribe you Walter-White-level quantities of methamphetamine for weight loss, nothing is stopping me but common sense. But if I want to give even a single suboxone prescription to a single patient, I have to take a special course on suboxone prescribing, and even then I am limited to only being able to give it to thirty patients a year (eventually rising to one hundred patients when I get more experience with it). The (generally safe) treatment for addiction is more highly regulated than the (very dangerous) addictive drugs it is supposed to replace. Only 3% of doctors bother to jump through all the regulatory hoops, and their hundred-patient limits get saturated almost immediately. As per the laws of suppy and demand, this makes suboxone prescriptions very expensive, and guess what social class most heroin addicts come from? Also, heroin addicts often don't have access to good transportation, which means that if the nearest suboxone provider is thirty miles from their house they're out of luck. The List Of Reasons To End The Patient Limits On

Buprenorphine expands upon and clarifies some of these points.

(in case you think maybe the government just honestly believes the drug is dangerous – nope. You’re allowed to prescribe without restriction for any reason except opiate addiction)

The second roadblock is the @#$%ing rehab industry. They hear that suboxone is an opiate, and their religious or quasi-religious fanaticism goes into high gear. “What these people need is Jesus and/or their Nondenominational Higher Power, not more drugs! You’re just pushing a new addiction on them! Once an addict, always an addict until they complete their spiritual struggle and come clean!” And so a lot of programs bar suboxone users from participating.

This doesn’t sound so bad given the quality of a lot of the programs. Problem is, a lot of these are closely integrated with the social services and legal system. So suppose somebody’s doing well on suboxone treatment, and gets in trouble for a drug offense. Could be that they relapsed on heroin one time, could be that they’re using something entirely different like cocaine. Judge says go to a treatment program or go to jail. Treatment program says they can’t use suboxone. So maybe they go in to deal with their cocaine problem, and by the time they come out they have a cocaine problem *and* a heroin problem.

And…okay, time for a personal story. One of my patients is a homeless man who used to have a heroin problem. He was put on suboxone and it went pretty well. He came back with an alcohol problem, and we wanted to deal with that and his homelessness at the same time. There are these organizations called three-quarters houses – think “halfway houses” after inflation – that take people with drug problems and give them an insurance-sponsored place to live. But the catch is you can't be using drugs. And they consider suboxone to be a drug. So of about half a dozen three-quarters houses in the local area, none of them would accept this guy. I called up the one he wanted to go to, said that he really needed a place to stay, said that without this care he was in danger of relapsing into his alcoholism, begged them to accept. They said no drugs. I said I was a doctor, and he had my permission to be on suboxone. They said no drugs. I said that seriously, they were telling me that my DRUG ADDICTED patient who was ADDICTED TO DRUGS couldn't go to their DRUG ADDICTION center because he was on a medication for treating DRUG ADDICTION? They said that was correct. I hung up in disgust.

So I agree with the pessimistic picture painted by the article. I think we're ignoring our best treatment option for heroin addiction and I don't see much sign that this is going to change in the future.

But the health care system not being very good at using medications effectively isn't news. I also thought this article was interesting because it touches on some of the issues we discuss here a lot:

The value of ritual and community. A lot of the most intelligent conservatives I know base their conservativism on the idea that we can only get good outcomes in "tight communities" that are allowed to violate modern liberal social atomization to build stronger bonds. The Army, which essentially hazes people with boot camp, ritualizes every aspect of their life, then demands strict obedience and ideological conformity, is a good example. I do sometimes have a lot of respect for this position. But modern rehab programs seem like a really damning counterexample. If you read the article, you will see that this rehabs are trying their best to create a tightly-integrated religiously-inspired community of exactly that sort, and they have abilities to control their members and force their conformity – sometimes in ways that approach outright abuse – that most institutions can't even dream of. But their effectiveness is abysmal. The entire thing is for nothing. I'm not sure whether this represents a basic failure in the idea of tight communities, or whether it just means that you can't force them to exist *ex nihilo* over a couple of months. But I find it interesting.

My love-hate relationship with libertarianism. Also about the rehabs. They're minimally regulated. There's no credentialing process or anything. There are many different kinds, each privately led, and low entry costs to creating a new one. They can be very profitable – pretty much any rehab will cost thousands of dollars, and the big-name ones cost much more. This should be a perfect setup for a hundred different models blooming, experimenting, and then selecting for excellence as consumers drift towards the most effective centers. Instead, we get rampant abuse, charlatanry, and uselessness.

On the other hand, when the government rode in on a white horse to try to fix things, all they did was take the one effective treatment, regulate it practically out of existence, then ride right back out again. So I would be ashamed to be taking either the market's or the state's side here. At this point I think our best option is to ask the paraconsistent logic people to figure out something that's neither government nor not-government, then put that in charge of everything.

Society is fixed, biology is mutable. People have tried *everything* to fix drug abuse. Being harsh and sending drug users to jail. Being nice and sending them to nice treatment centers that focus on rehabilitation. Old timey religion where fire-and-brimstone preachers talk about how Jesus wants them to stay off drugs. Flaky New Age religion where counselors tell you about how drug abuse is keeping you from your true self. Government programs. University programs. Private programs. Giving people money. Fining people money. Being unusually nice. Being unusually mean. More social support. Less social support. This school of therapy. That school of therapy. What works is just giving people a chemical to saturate the brain receptor directly. We know it works. The studies show it works. And we're still collectively beating our heads against the wall of finding a social solution.

Everything Not Obligatory Is Forbidden

[seen on the New York Times' editorial page, February 6 2065, written by one "Dr. Mora LeQuivalence"]

It's 2065. Not giving your kids super-enhancement designer baby gene therapy isn't your "choice". If you don't super-enhance your kids, you are a bad parent. It's that simple.

Harsh? Maybe. But consider the latest survey, which found that about five percent of parents fail to super-enhance their children by the time they enter kindergarten. These aren't poor people who can't afford super-enhancement designer baby gene therapy. These are mostly rich, highly educated individuals in places like California and Oregon who say they think it's more "natural" to leave their children defenseless against various undesirable traits. "I just don't think it's right to inject retroviral vectors into my baby's body to change her from the way God made her," one Portland woman was quoted by the *Times* as saying earlier this week. Other parents referred to a 2048 study saying the retroviral injections, usually given in the first year of life, increase the risk of various childhood cancers – a study that has since been soundly discredited.

These parents will inevitably bring up notions of "personal freedom". But even if we accept the dubious premise that parents have a right to sacrifice their children's health, refusing super-enhancement designer baby gene therapy isn't just a personal choice. It's a public health issue that affects everybody in society.

In 2064 there were almost 200 murders nationwide, up from a low of fewer than 50 in 2060. Why is this killer, long believed to be almost eradicated, making a comeback? Criminologists are unanimous in laying the blame on unenhanced children, who lack the improved impulse-control and anger-management genes included in every modern super-enhancement designer baby gene therapy package.

There were over a dozen fatal car accidents on our nation’s roads last year. The problem is drivers who weren’t enhanced as children and who lack the super-reflexes the rest of us take for granted. This is compounded when they drink before getting on the road, since unenhanced people become impaired by alcohol and their already inferior reflexes deteriorate further. Since the promise of self-driving cars continues to be tied up in regulatory hassles, we can expect many more such needless deaths as long as irresponsible parents continue to consider science “optional”.

And finally, there was a recent outbreak of measles at Disneyland Europa – even though we thought this disease had been eradicated decades ago. Scientists traced the problem to unvaccinated tourists. They further found that all of these unvaccinated individuals were unenhanced. Lacking the cognitive optimization that would help them understand psychoneuroimmunology on an intuitive level, they were easy prey for discredited ideas like “vaccines cause autism”.

So no, super-enhancing your kids isn’t a “personal choice”. It’s your basic duty as a parent and a responsible human being. People in places like India and Neo-Songhai and Venus which suffer from crime and disease make great personal sacrifices to get their children to gene therapy clinics and give them the super-enhancement designer baby gene injection that ensures them a better life. And you start off in a privileged position in America, benefitting from the superenhancement of millions of your fellow citizens, and you think you can just say “No thanks”?

So I don’t want to hear another word from the “but my freedom!” crowd. Unenhanced kids shouldn’t be allowed in school. They shouldn’t be allowed to drive. They shouldn’t be allowed in public places where they can cause problems. And parents who refuse to enhance their children should be put in jail, the same as anyone else whose actions lead to death and suffering. Because not super-enhancing your kids isn’t a “choice”. It’s child abuse.

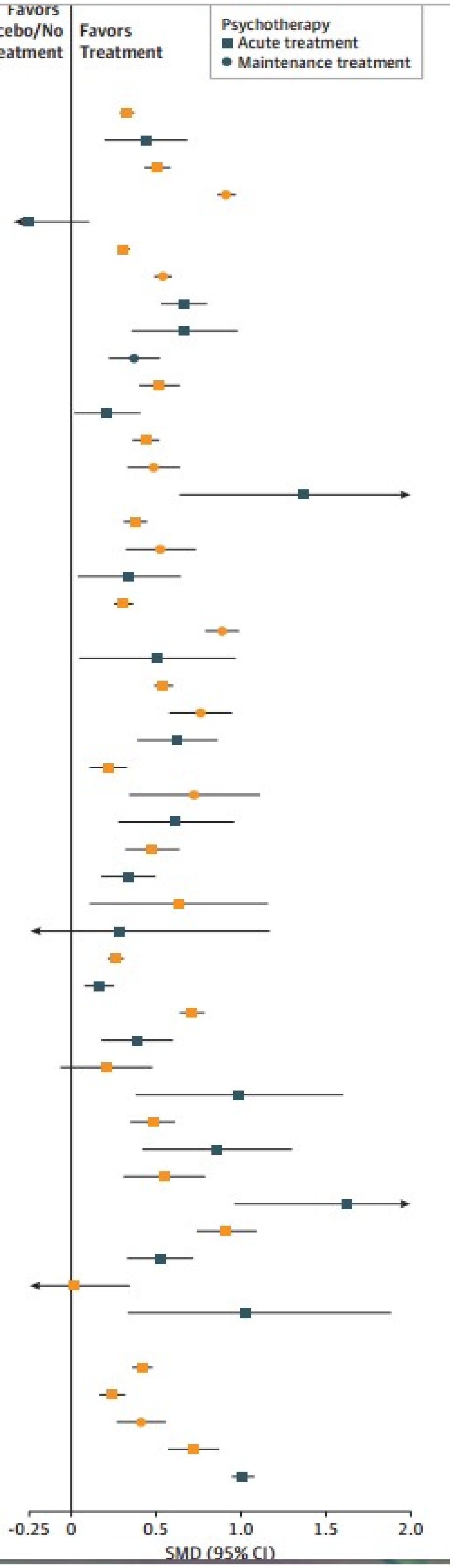
Mora LeQuivalence is an Assistant Professor of Bioethics at Facebook University. Her latest book, “A Flight Too Far”, argues that the recent Danish experiment with giving children wings is a disgusting offense against the natural order and should be banned worldwide and prosecuted in the International Criminal Court. It is available for 0.02Ƀ on Amazon.com

Related: Transhumanism Is Simplified Humanism, Alicorn’s Alternate Universe Social Justice Series

The Efficacy Of Everything In Psychiatry In One Graph Plus Several Pages Of Dense But Necessary Explanation

Shamelessly stolen from my hospital’s Journal Club: Huhn et al (2014) graph the Efficacy Of Pharmacotherapy And Psychotherapy For Adult Psychiatric Disorders, and it looks like this:

	SMD (95% CI)
Therapy	
Dementia drug (n = 8069)[18,28,35]	0.33 (0.29 to 0.37)
Dementia PT (n = 270)[60]	0.44 (0.20 to 0.69)
Schizophrenia acute drug (n = 5568)[30]	0.51 (0.44 to 0.58)
Schizophrenia relapse drug (n = 6493)[31]	*0.92 (0.86 to 0.97)*
Schizophrenia psychodynamic therapy (n = 164)[59]	-0.25 (−0.59 to 0.11)
MDD acute drug (n = 12 564)[15]	0.31 (0.27 to 0.80)
MDD relapse drug (n = 9268)[25]	*0.54 (0.49 to 0.59)*
MDD acute PT (n = 3465)[16,a]	0.67 (0.53 to 0.80)
Depression psychodynamic therapy, acute (n = 196)[50,b]	0.69 (0.30 to 1.08)
MDD relapse PT (n = 881)[57]	*0.37 (0.23 to 0.52)*
Dysthymia drug (n = 1454)[32]	*0.52 (0.40 to 0.64)*
Dysthymia PT (n = 275)[49]	0.21 (0.02 to 0.41)
OCD acute drug (n = 3097)[43]	0.44 (0.36 to 0.52)
OCD relapse drug (n = 951)[23]	*0.48 (0.32 to 0.64)*
OCD PT (n = 240)[54]	1.37 (0.64 to 2.24)
Panic acute drug (n = 7725)[37]	0.38 (0.31 to 0.45)
Panic relapse drug (n = 796)[23]	*0.53 (0.32 to 0.74)*
Panic PT (n = 328)[54]	0.35 (0.04 to 0.65)
GAD acute drug (n = 11 427)[38]	0.31 (0.26 to 0.36)
GAD relapse drug (n = 1342)[23]	*0.89 (0.79 to 0.99)*
GAD PT (n = 95)[54]	0.51 (0.05 to 0.97)
Social phobia acute drug (n = 7619)[27]	0.55 (0.49 to 0.60)
Social phobia relapse drug (n = 760)[23]	*0.76 (0.58 to 0.95)*
Social phobia PT (n = 377)[54]	0.62 (0.39 to 0.86)
PTSD acute drug (n = 2507)[44]	0.22 (0.11 to 0.33)
PTSD relapse drug (n = 272)[23]	*0.73 (0.34 to 1.11)*
PTSD PT (n = 266)[54]	0.62 (0.28 to 0.96)
Somatoform disorder drug (n = 832)[24]	0.48 (0.32 to 0.64)
Somatoform disorder PT (n = 1647)[56]	0.35 (0.18 to 0.50)
Borderline personality disorder drug (n = 1151)[47]	*0.63 (0.11 to 1.16)*
Borderline personality disorder DBT (n = 20)[61,c]	0.29 (−0.59 to 1.17)
Alcohol drug (n = 10 605)[40,41]	*0.26 (0.22 to 0.31)*
Alcohol PT (n = 2338)[58,62]	0.17 (0.08 to 0.26)
Opiates drug (n = 4001)[33,34]	*0.71 (0.64 to 0.79)*
Opiates PT (n = 545)[51]	0.39 (0.18 to 0.60)
Anorexia drug (n = 211)[22,29]	0.21 (−0.06 to 0.48)
Anorexia PT (n = 77)[52]	0.99 (0.38 to 1.60)
Binge-eating drug (n = 1254)[39]	*0.48 (0.35 to 0.61)*
Binge-eating PT (n = 90)[53]	0.86 (0.42 to 1.30)
Bulimia drug (n = 824)[17]	*0.55 (0.31 to 0.79)*
Bulimia PT (n = 204)[53]	*1.61 (0.96 to 2.29)*
Primary insomnia drug (n = 566)[26]	0.90 (0.74 to 1.09)
Primary insomnia PT (NI)[55]	0.52 (0.33 to 0.72)
Trichotillomania drug (n = 72)[19]	0.02 (−0.32 to 0.35)
Trichotillomania PT (n = 59)[19]	1.14 (0.38 to 1.89)
Disorders for which only meta-analyses on 1 modality were available	
Bipolar mania acute drug (n = 11 092)[48]	0.42 (0.36 to 0.48)
Bipolar depression acute drug (n = 3770)[42,45]	*0.24 (0.17 to 0.32)*
Bipolar disorder relapse drug (n = 2829)[46]	*0.41 (0.27 to 0.56)*
Adult ADHD drug (n = 1045)[21]	0.73 (0.57 to 0.87)
Specific phobia PT (n = 1217)[63]	1.03 (0.91 to 1.16)



Before anything else – we kind of have to assume that in each case they're getting a representative sample of the best drugs/therapies for the disorder. In practice, there is this *weird* equivalency for most things: most common antidepressants work about equally well, most common antipsychotics work about equally well, et cetera. I don't know as much about therapy, but I get the impression the same thing goes on there too. So probably it's not much of a stretch to expect that the efficacy of whatever they studied at least *kind of* translates to the effectiveness of whatever real treatment you'll get from your own psychiatrist. At the very least, even lossy and compressed information like this will tell us something.

The effect sizes are mostly around 0.5, with a few much higher and a few much lower. This is common for these sorts of studies. See for example Leucht et al, Putting the efficacy of psychiatric and general medical medication into perspective, which also finds psychiatric effect sizes average around 0.5 and finds this is about equal to average effect sizes in other fields of medicine – thus debunking the popular claim that psychiatry is less effective. Leucht and a few other authors from that piece are also involved in this one, which doesn't surprise me much.

I do however admit my statistical ignorance in exactly what is going on here. Effect sizes are a good way to compare two unlike domains – for example, I recently noted that leading physicists are about as smart as NBA players are tall. This paper is within that tradition. In fact, if we wanted, we could describe psychiatric medications as about one-sixth as effective as NBA players are tall. This is perfectly honorable. The height of NBA players is a tough bar to live up to.

But I don't have a good intuitive feel for what it means to use standard mean differences along a non-normally distributed variable – as psychiatric diseases no doubt are. And I’m not sure where they’re even getting their distributions from. When they say schizophrenia meds have an effect size of 0.52, are they talking about the distribution of the general population, with almost everyone near zero and a few schizophrenics way off to the right? Are they talking about the distribution of how schizophrenic particular schizophrenics are, which for all I know might be a bell curve but which is probably very different depending on how you took your schizophrenia sample? I really don’t get this and it’s preventing me from getting a good feeling of exactly how comparable these numbers are to each other.

If we just assume they’re allowed to do what they’re doing, their graph looks about how I would expect it to look. Most psychiatrists always figured that the psychotic disorders were more susceptible to medication and the anxiety disorders to psychotherapy. But three surprises stand out.

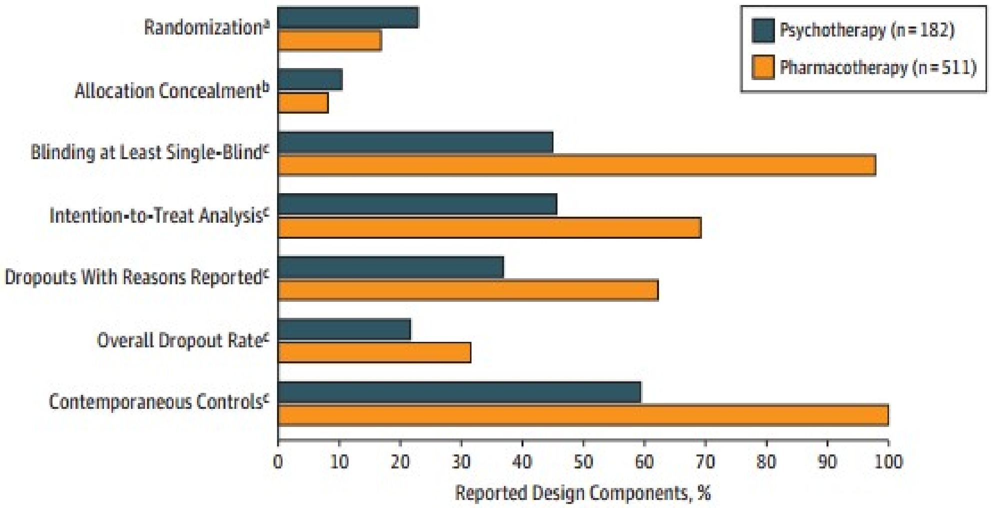
First, this graph shows that drugs are more effective than therapy in treating borderline personality disorder. That’s the opposite of the conventional wisdom, which says that some drugs can decrease impulsiveness in this population but that the definitive treatment has always been Dialectical Behavioral Therapy. But it looks like their borderline psychotherapy “meta-analysis” had a sample size of 20 patients (I would hate to see what the individual studies had!) compared to thousands of patients for most of the believable results. So I wouldn’t place too much faith in this anomaly for now and would continue to recommend psychotherapy for borderlines.

Second, this graph shows that drugs are more effective than therapy for insomnia. Now, we *use* drugs instead of therapy for insomnia, but conventional wisdom had always been that this was very sad, and there was great therapy available for insomnia if only somebody would provide it. But here the therapy looks mediocre at best. On the other hand, the sample size is “NI”, which I don’t know what it means but doesn’t sound promising. Also, now that anticholinergics probably cause dementia, every single sleeping pill now officially has terrible side effects.

Third, in all conditions drugs seem more effective at preventing relapse than at stopping acute episodes. My “clinical experience,” which is the fancy word doctors use for anecdotal evidence, was exactly the opposite. I now realize I probably faced a lot of selection bias – the patients who do well on their drug and don’t relapse might never see me again. Also, I have a feeling that a lot of the people who come back to me a month later and say “Well, your drug must not have worked, I’ve relapsed again” probably weren’t taking the drug correctly or at all, something which these studies probably enforce better than I can.

In general, the table seems to support psychotherapy being better than drugs for a lot of things. This would not be *too* surprising if true – their list is heavily tilted to the kinds of conditions therapy works well on – but a caveat is necessary.

The psychotherapy trials were generally of lower quality. Part of this has to do with the culture of psychotherapy research, but more has to do with the underlying territory – giving people “placebo psychotherapy” is more complicated than giving people a sugar pill and a lot of studies don’t bother. Also, in psychotherapy, it tends to be the patient’s therapist recording results more often than corresponding pharmacology studies use the prescriber to record results. That eliminates another layer of blinding.



This has serious ramifications. The study finds that “low-quality psychotherapy trials in general had a higher effect size (SMD = 0.74) than high-quality trials (SMD = 0.22), p < 0.001". Those high effect sizes for psychotherapy aren't looking so good *now*, are they?

Actually, reread that one more time. Effect sizes for the low quality trials are triple those for the high-quality trials. If you ever wanted proof that it’s way too easy to inflate positive findings if your science isn’t really exceptionally good, there you go.

The most important domain where pharmacotherapy trials are worse than psychotherapy trials is publication bias. The paper suggests that this is because psychotherapy’s lack of sufficient blinding and control groups makes publication results unnecessary. In other words, psychotherapy research isn’t even good enough to have publication bias, because publication bias at least requires you to be rigorous enough to occasionally turn up a negative result to suppress. Ouch.

Black People Less Likely

[Content warning: Polyamory, race]

I.

The best reporting on social science statistics, like the best reporting in most areas, comes from The Onion:

> CAMBRIDGE, MA—A Harvard University study of more than 2,500 middle-income African-American families found that, when compared to other ethnic groups in the same income bracket, blacks were up to 23 percent more likely. “Our data would seem to discredit the notion that black Americans are less likely,” said head researcher Russell Waterstone, noting the study also found that women of African descent were no more or less prone than Latinas. “In fact, over the past several decades, we’ve seen the African-American community nearly triple in probability.” The study noted that, furthermore, Asian-Americans.

I thought of this today because a bunch of people have accosted me about the article There’s A Big Problem With Polyamory That Nobody’s Talking About. “Scott, you’re polyamorous! What do you think of this?”

As per the article, the big problem with polyamorous people is:

> …their whiteness. And that standard of whiteness not only erases the experience of people of color; it reflects the actual exclusion of these people in poly life and communities. […]

> A white, affluent image that reflects a troubling reality: A 2013 survey of polyamorous people from online groups, mailing lists and forums found that almost 90% of the participants identified as Caucasian. People of color, especially black polyamorists, report feeling "othered" and excluded in poly environments such as meet-ups, with women feeling especially at risk of being objectified and fetishized as an exotic sexual plaything.
>
> "I interviewed a black couple who went to a poly group, and they were definitely preyed upon, in a sense," said Marla Renee Stewart, Atlanta-based founder of Velvet Lips, a sex education venue.

The article constantly equivocates between "the problem is that polyamory is too white" and "the problem is that the media portrays polyamory as too white", which is kind of a weird combination of problems to be discussing in a media portrayal. But it seems to eventually settle on a thesis that black people really are strongly underrepresented.

For the record, here is a small sample of other communities where black people are strongly underrepresented:

Runners (3%). Bikers (6%). Furries (2%). Wall Street senior management (2%). Occupy Wall Street protesters (unknown but low, one source says 1.6% but likely an underestimate). BDSM (unknown but low) Tea Party members (1%). American Buddhists (~2%). Bird watchers (4%). Environmentalists (various but universally low). Wikipedia contributors (unknown but low). Atheists (2%). Vegetarian activists (maybe 1-5%). Yoga enthusiasts (unknown but low). College baseball players (5%). Swimmers (2%). Fanfiction readers (2%). Unitarian Universalists (1%).

Can you see what all of these groups have in common?

No. No you can't. If there's some hidden factor uniting Wall Street senior management and furries, it is way beyond any of our pay grades.

But what I noticed when I looked up those numbers was that in every case, the people involved have come up with a pat explanation that sounds perfectly plausible right up until you compare it to any other group, at which point it bursts into flames.

For example, Some people explain try to explain declining black interest in baseball by appeal to how some baseball personality made some horribly racist remark. But Donald Sterling continues to be racist as heck, and black people continue to be more than three-quarters of basketball players.

Some people try to explain black people's underrepresentation on fanfiction websites by saying that many of them have limited access to the Internet. Okay. Except that black people are heavily *overrepresented* on Twitter, making up double the expected proportion of that site's population.

Some people try to explain the underrepresentation of blacks in libertarianism and the Tea Party by arguing that these groups' political beliefs are contrary to black people's life experiences. But blacks are also underrepresented in groups with precisely the opposite politics. That they make up only 1.6% of visitors to the Occupy Wall Street website is no doubt confounded by who visits websites, but even people who looked at the protests agree that there was a stunning shortage of black faces. I would have liked to get current membership statistics for the US Communist Party, but they weren't available, so I fudged by looking at the photos of people who "liked" the US Communist Party's

Facebook page. 3% of them were black. Blacks are more likely to endorse environmentalism than whites, but less likely to be involved in the environmentalist movement.

Some people try to explain black people's underrepresentation on Wall Street by saying Wall Street is racist and intolerant. But Unitarian Universalists are just about the most tolerant people in the world – nobody even knows what they do, just that they're extremely tolerant when they do it – and black people are in Unitarianism at lower rates than they're on Wall Street.

And the article on polyamory suggested that maybe polyamorists' high-flying lifestyle and expensive play parties price out black people. Forget for a moment that I've been poly for three years and had no idea this high-flying lifestyle existed and kind of feel like I am missing out. Forget for a moment that as far as I can tell "play parties" are a BDSM term with no relationship to polyamory. In my experience polyamory draws from the same sort of people as atheism, and atheism is *very* white even though not believing in God doesn't cost a cent.

This entire genre seems to be a bunch of really silly ad hoc arguments by people who aren't talking to each other. I would guess most of the underrepresentation of black people in all of these things are for the same couple of reasons.

First, some of these things require some level of affluence – I know I just said that didn't explain polyamory, but I think it explains some others. For example, bird-watching requires you live somewhere suburban or rural where there are interesting birds, want to waste money on binoculars, and have some free time. Swimming requires you live in an area where the schools or at least the neighborhoods have pools.

Second, Maslow’s Hierarchy Of Needs says you’re not going to do weird things to self-actualize until you feel materially safe and secure. A lot of black people don’t feel like they’re in a position where they can start worrying about where the best bird-watching is at.

Third, the thrive-survive dichotomy says materially insecure people are going to value community and conformity more. Polyamory is still pretty transgressive, and unless you feel very safe or feel sufficiently mobile and atomized that you don’t care what your community thinks about you, you’re not going to feel comfortable making that transgression. Many of these things require leaving the general community to participate in a weird insular subculture, and that requires a sort of lack of preexisting community bonds that I think only comes with the upper middle class.

Fourth, black people might avoid weird nonconformist groups because they’re already on thin enough ice in terms of social acceptance. Being a black person probably already exposes you to enough stigma, without becoming a furry as well.

Fifth, we already know that neighborhoods and churches tend to end up mostly monoracial through a complicated process of aggregating small acts of self-segregation based on slight preferences not to be completely surrounded by people of a different race. It doesn’t seem too unlikely to me that a similar process could act on hobbies and interest groups.

Sixth, even when black people are involved in weird subcultures, they may do them separately from white people, leading white people to think their hobby is almost all white – and leading mostly white academics to miss them in their studies. I once heard about a professor who accused

Alcoholics Anonymous of being racist, on the grounds that its membership was almost entirely white. The (white) professor had surveyed AA groups in his (white) neighborhood and asked his (white) friends and (white) grad students to do the same. Meanwhile, when more sober minds (no pun intended) investigated, they found black areas had thriving majority-black AA communities.

Seventh, a lot of groups are stratified by education level. Black people are only about half as likely to have a bachelor's degree. This matters a lot in areas like atheism that are disproportionately limited to the most educated individuals. Polyamory also falls into this category – the most recent survey found 85% of poly people had a college education, compared to 30% of the general population (!). 30% of poly people had a graduate degree compared to only about 10% of the general population and only about 3% of blacks. There has to be a strong education filter on polyamory to produce those kinds of numbers, and I think that alone is big enough to explain most of the black underrepresentation.

Eighth, people of the same social class tend to cluster, and black people are disproportionately underrepresented among the upper middle class. Most of these fields are dominated by upper middle class people. The nickname for weird self-actualizing upper middle class things is "Stuff White People Like", and this is not a coincidence. [EDIT: Commenter John Schilling says this better than I – a lot of these groups are about differentiating yourself from a presumedly boring low-status middle class existence, but black people fought hard to get into the middle class, or are still fighting, and are less excited about differentiating themselves from it.]

So I think positing that black people feel “fetishized as an exotic sexual plaything” in the poly community is unnecessary. Black people are underrepresented in the poly community for the same reason they’re underrepresented in everything in the same vague circle as poly. Heck, black people are even underrepresented in the activity of complaining about black people being fetishized as exotic sexual playthings – check out Tumblr’s racial demographics if you don’t believe me.

II.

The eight points above add up to a likelihood that black people will probably be underrepresented in a lot of weird subculturey nonconformist things. This is not a firm law – black people will be overrepresented in a few weird subculturey nonconformist things that are an especially good fit for their culture – but overall I think the rule holds. And that’s a big problem.

A few paragraphs back I mentioned that Occupy Wall Street was had disproportionately few minorities. Here are some other people who like to mention this: Michelle Malkin. The Daily Caller. American Thinker. View From The Right. New York Post. American Renaissance.

All of these sources have something in common, and it’s *not* a heartfelt concern for equal minority representation.

Likewise, you know who’s got an *obsessively* large collection of resources on the underrepresentation of minorities in atheism? Conservapedia (Western Atheism And Race, Racial Demographics Of The Richard Dawkins Audience, Richard Dawkins’ Lack Of Appeal To The Asian Woman Audience, etc, etc, not to mention the very classy Richard Dawkins’ Family Fortune And The Slave Trade.)

Here it is *easy* to see that "you have low minority representation" serves as a stand-in for "you're racist" serves as a stand-in for "you suck". So here's the problem:

In theocracies ruled by the will of God, people will find that God hates weird people who refuse to conform.

In philosopher-kingdoms ruled by pure reason, people will find that pure reason condemns weird people who refuse to conform.

And in enlightened liberal democracies where we "tolerate anything except intolerance", people will find that weird people who refuse to conform are intolerant.

And if blacks are underrepresented in weird nonconformist groups, and nobody mentions that this is a general principle, that's making their job *way too easy*.

So here's why this article annoys me. In the midst of black underrepresentation in everything in the same ontological category as polyamory, people bring up black underrepresentation in polyamory and suggest it's because poly people are "objectifying" and "preying on" them, positing that "there's a problem" with "a standard of whiteness that erases people of color" in the polyamory community.

We know from OKCupid statistics that (mostly monogamous) white men are very reluctant to date black women, but monogamous people don't have to listen to well-meaning friends going up to them and saying "So, you're mono, I hear the monogamous community has a racism problem."

But now I and other polyamorous people are going to have to answer one more round of annoying questions about

“You’re polyamorous? Isn’t that a bunch of racist nerdy white dudes?”

Money, Money, Everywhere, But Not A Cent To Spend

The DSM is written mostly by academics, which is why it gets so excited about distinctions like schizoid personality versus schizotypal personality. If it were written by clinicians, it might better reflect the sort of cases that make it into a hospital.

There would, for example, be an entire chapter on the scourge of ‘My Boyfriend Broke Up With Me’ spectrum disorders. More attention would get paid to the plague of chronic ‘I Got Angry At My Dad And Told Him I Was Going To Kill Myself To Freak Him Out And He Overreacted And Called The Cops And Now Here I Am In Hospital But Honestly I Didn’t Mean It’. Society would finally wake up to the epidemic of ‘I Wanted To Take My Medicine But My Hand Slipped And I Somehow Took The Entire Bottle All At Once Even Though I Would Never Do Something Like Intentionally Overdose’. And the sufferers of ‘This Patient Probably Has Some Kind Of Complicated Neurological Problem But Neurology Is Tired Of Trying To Figure It Out So They Have Declared It To Be Psychiatric’ might at last get some relief.

But the biggest change to the medical lexicon would be the introduction of ‘Poverty NOS’.

I recently got a patient, let’s call him Paul…

(all of my patient stories are vague composites of a bunch of people with details changed to protect privacy)

…who was in hospital after trying to hang himself. He said he was so deep in debt he was never going to get out. He’d been involved in a messy court case, had to hire a lawyer to defend himself, lawyer ended up running to the tune of several thousand dollars. He was a clerk at a clothing store, barely made minimum wage, maxed out his credit cards, then maxed out other credit cards paying off the first credit cards.

He didn't seem to have major depressive disorder, but when someone comes in admitting to a serious suicide attempt, procedure says he gets committed. He wasn't thrilled about this, saying if he missed work then he might lose his job and this was just going to make him further behind on his payments, but I checked with my attending and as usual the answer was "admit".

Something especially bothered me about this case, and after thinking about it I've figured out what it is.

It's not just that the psychiatric hospitalization won't help and might hurt. That's pretty common. The 'My Boyfriend Broke Up With Me's, the 'I Got Angry At My Dad's, unless they have some underlying disorder all of these people get limited value from the psychiatric system and tend to just sit in hospital for a couple of days, go to some group therapy, get asked a hundred times if they're depressed, then go home. And then they're still broken up with their boyfriend or still have a terrible relationship with their dad and the same thing's going to happen again.

In this case, it was that – well, the guy is a minimum wage worker from inner city Detroit. He didn't tell me exactly how much money this debt was, but from a couple of numbers he mentioned I got the impression it was in the ballpark of $5000. That might not seem like an attempt-suicide level of money to some people, but to this guy with his job the chance of ever paying it off seemed low enough that it wasn't worth waiting and seeing.

So what bothered me is that psychiatric hospitalization costs about $1,000 a day. Average length of stay for a guy like him might be three to five days. So we were spending $5,000 on his psychiatric hospitalization, which was USELESS, so that we could send him out and he could attempt suicide again because of his $5,000 debt which he has no way of paying off. And probably end up in the hospital a second time, for that matter.

I assume that since he was poor, Medicaid paid his hospital bill. I'm not complaining that the cost of the hospital bill was added to his debt, I'm pretty sure it wasn't, although in some other cases it would be. I'm complaining that here's this guy, so desperate for money that he wants to kill himself over it, and he has to sit helplessly as we throw thousands of dollars at getting a parade of expensive doctors and nurses and social workers to talk to him, conclude that yup, his problem is *definitely* that he's poor, and then throw him back out. I feel like this fails to be, as the buzzwords say, "patient-centered care".

Problem is, you don't have to be an economics PhD to realize that "give $5,000 to anyone who attempts suicide and says they need it" might create some bad incentives.

I have no good solution to this. Offering people who are so poor they want to kill themselves very expensive psychiatric care seems maybe a little better than doing nothing. But it also seems insulting, patronizing, paternalistic, wasteful, and occasionally heartbreaking.

And this is why I can never decide whether to identify as a libertarian or a liberal. On the one hand, top-down institutionalized bureaucracies seem so ridiculously inefficient at solving problems that it's an outrage and a disaster. On the other hand, there are a lot of problems that *really* need solving, they don’t seem to have solved themselves yet, and governments are the only entity with enough coordination power to attempt the task.

Solution there, it seems to me, is to create unimpoverishable populaces. I think if we were to implement a Basic Income Guarantee we might save more money in psychiatric care than we think – since we compete with the prison system to be the warehouse for people who can’t make it out in the world and nobody knows what to do with. It might produce some of the same kind of savings as giving the homeless people houses. If I got fired because we’d solved all the problems relating to poverty, and the population of seriously mentally ill people was too small to support the current number of psychiatrists, that would be a pretty neat way to go.

Drug Testing Welfare Users Is A Sham, But Not For The Reasons You Think

Some people say the War on Drugs is ‘unwinnable’. But there’s actually a foolproof solution that cures drug addiction approximately 100% of the time. That solution is – put people on welfare in Tennessee.

Or at least that is what I am led to believe by articles like Mic’s A Shocking Thing Happened When Tennesee Decided To Drug Test Its Welfare Recipients, which describes said shocking thing as:

> 1 out of 812 applicants tested positive for drugs. One. Single. Person. Tennessee conservatives suspicious that welfare recipients are a bunch of drug-addicted slackers were proven dead wrong. Big surprise!
>
> After instituting dehumanizing drug-testing requirements to welfare recipients on July 1, 10 people total were flagged for possible drug use and asked to submit to testing. Five others tested negative, and four were rejected after refusing. As Think Progress notes, that means that just 0.12% of all people applying for cash assistance in Tennessee have tested positive for drugs, compared to the 8% who have reported using drugs in the past month among the state’s general population. If you assume the four people who refused were on drugs, it’s still a paltry 0.61%.
>
> In other words, the plan intended to verify right-wing beliefs that welfare recipients are a bunch of drug-

> addicted slackers looking for a handout has demonstrated exactly the opposite.

The article has 11,000 notes on Tumblr right now, I’ve seen it all over my Facebook feed as well, and the same story has been taken up, with the same editorial line, by a host of other news sources. Jezebel: State Drug Program Busts A Whopping 37 Welfare Applicants. Wall Street Journal: Few Welfare Applicants Caught In Drug Screening Net So Far. New Republic: Red States’ New Tax On The Poor. Daily Kos: Tennessee Just Wasted A Lot Of Money Drug Testing Welfare Recipients. ReverbPress: Another GOP Fail: 0.2% Of Tennessee Welfare Recipients Found To Use Illegal Drugs. Mommyish: Results Of State Drug Testing Prove Gross Assumptions About Welfare Applicants Are Wrong. Washington Post: Scott Walker’s Yellow Politics.

These stories all make the point that we have many stereotypes about the poor, and one such stereotype is that the use lots of drugs, but in fact these sorts of welfare programs find them to use fewer drugs than the general population, and therefore we should stop being so prejudiced.

And if they were found to use only two-thirds, or half as many drugs as the general population, this might indeed be the lesson.

But look at the numbers in the quoted Mic article. Welfare users use only about *one percent* as many drugs as the general population. *Really?*

No. Not really at all. According to legitimate research in this area, poor people use as many drugs as anyone else and probably more. The National Household Survey on Drug Abuse found that illegal drug use was slightly higher in families on government assistance (9.6%) than families not on government assistance (6.8%). The National Coalition For The Homeless notes that about 26% of them use drugs, which is about 2.5x as high as the general population. I crunched some data I have from the hospital I work at, and it shows that poor people (defined as people who get health insurance through an aid program) have moderately higher rates of drug use related problems than the general population. So these articles are reporting a drug use rate in the Tennessee population about one percent of that ever reported in any comparable poor population anywhere else.

Kate from Gruntled and Hinged brings up another curious inconsistency. The false positive rate for drug tests is – well, it depends on the test procedure, but it’s usually at least 1%. So if every single welfare user in Tennessee was 100% clean, we would *still* expect between 1% to 5% positive drug tests. Instead, they got 0.12% positive drug tests. This isn’t just suspiciously good, it’s *impossibly* good.

So what’s going on here?

Before I explain, here’s a collage of the stock photos displayed above some of those news stories I linked to.

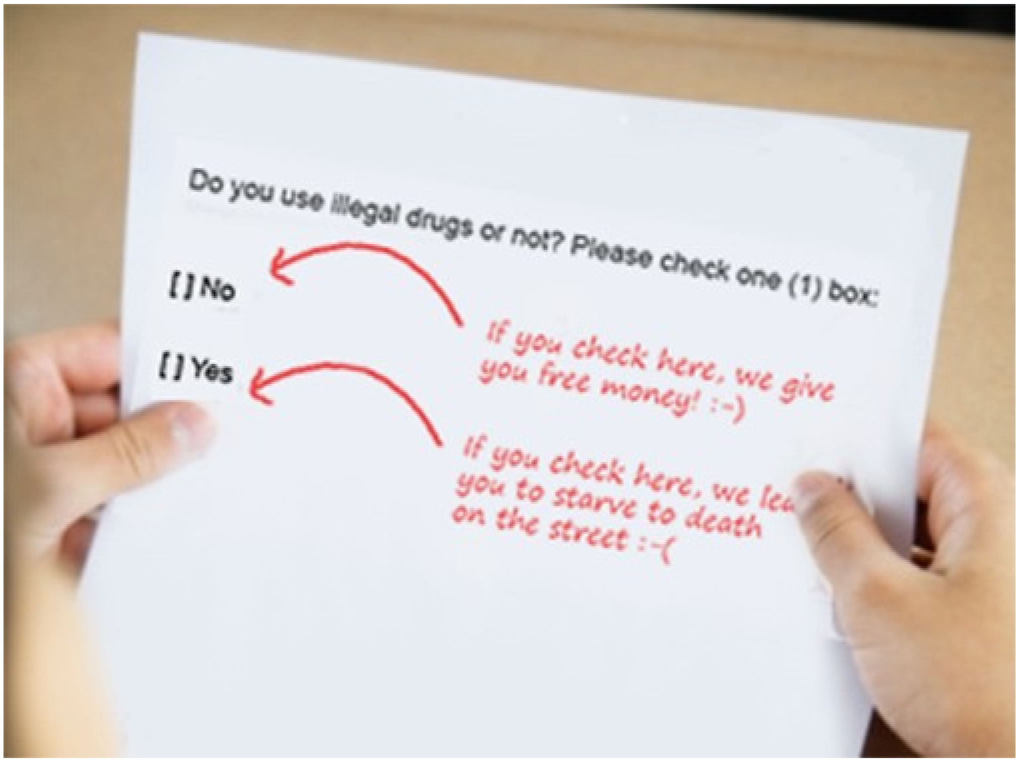
I now have a picture on my website called urine_collage.png

If you’re familiar with the state of the American media, you won’t be surprised to learn that urine was not involved in the ovewhelming majority of this program’s drug tests.

So how did they test people for drugs?

They gave them a written test, where the test question was basically “do you use illegal drugs or not?” You can see the exact procedure on the sidebar here.

And lo and behold, the overwhelming majority of people answered that they didn’t.



Do you use illegal drugs or not? Please check one (1) box:

[ ] No

[ ] Yes

If you check here, we give you free money! :-)

If you check here, we le you to starve to death on the street :-(

A more accurate stock photo they could have used

Now the numbers make sense. It's not that only 0.2% of welfare recipients use drugs. All this tells us, if anything, is that 0.2% of welfare recipients are on so many drugs they can't figure out how to check "NO" on a form.

Why would the government do something like this? As best I can tell, the plan was originally to give everyone urine checks, but in Florida the courts decided that urine-checking people without prior suspicion was unconstitutional. The Republicans were pretty attached to their "drug test welfare recipients" plan and didn't want to look like they were wimps who backed down just because of one little court case, so they decided to give people the written test in the hopes of having prior suspicion for the people who said yes. Sure, it made no sense, but they could still tell their constituents they were drug testing those welfare recipients, and *in principle* they'd won an important victory. Or something.

Which raises another interesting question – how did Florida's urine-based program do before the courts struck it down?

According to the media, abysmally. MSNBC: Drug Testing Welfare Recipients Looks Even Worse, "[Florida Governor] Scott's policy was an embarrassing flop. Only about 2 percent of applicants tested positive, and Florida actually lost money". TBO: Welfare Drug Testing Yields 2% Positive Results, "Newton said that's proof the drug-testing program is based on a stereotype, not hard facts." ATTN: Why Drug Testing Poor People Is A Waste Of Time And Money, "Florida tested welfare recipients for four months before its drug test mandate was thrown out by the courts. Only 2.6 percent of welfare recipients tested positive. The rest of the Florida's population use drugs at a rate of 8 percent. So, again, welfare recipients used drugs less than everyone else."

Now we're merely at one-quarter of the drug use rate people with good methodologies find. Improvement!

So I looked up exactly how this works. Apparently welfare recipients were asked to pay for their own drug tests, and would be reimbursed if the results came back negative. 7000 welfare users did this, but 1600 declined to do so – numbers that were not mentioned in most of the pieces above.

Opponents of the program say that maybe those 1600 people could not find drug testing centers near them, or couldn't afford to pay for the tests even with the promise of reimbursement later, or something like that. I am sure that some of them did indeed decline for reasons like those.

But also, people on welfare don’t have very much money [citation needed]. If I were a welfare recipient, and they were going to drug test me and not reimburse me if I came out positive, and I was on drugs, I would decline the hell out of that test.

Suppose that the poor in Florida use drugs at the same rate as the poor in various studies and surveys – about 10%. We have 8600 welfare recipients, so we would expect 860 drug users. Of the 7000 who agreed to testing, we know that 2.5% are drug users – that’s 175 people. That in turn would suggest that of the 1600 who refused testing, about 685 were drug users – 40% or so. That would imply that about 80% of drug users versus about 12% of nonusers refused testing.

These numbers seem pretty reasonable to me. Most welfare users want to keep their benefits, so the majority will agree to testing, but a few will inevitably fall through the cracks because they can’t reach a testing center or because they have moral objections to the tests. On the other hand, clued-in drug users will realize that for them, testing means a major inconvenience and monetary charge without any likely corresponding gain. So we would expect drug users to decline testing at a higher rate than nonusers. In order to use the Florida data to say that welfare recipients *in general* use drugs at a rate of 2%, we would need to assume that drug users were no more likely to refuse drug testing than nonusers, even though the testing rewarded non-use with money but punished use with a loss of money.

(note that there are some different numbers in different places for Florida. I assume that these represent different years, stages of testing, parts of Florida, etc, but I’m not sure. The only one that is *seriously* different from what I’m saying above is the one that says “only 1% of people declined testing". After some search, I'm pretty sure that's referring to that only 1% of people made appointments for testing, then cancelled later. But I am less confident in the Florida numbers than in the analysis of Tennessee)

So the Florida numbers are consistent with welfare recipients using drugs less, more, or the same amount as the general population.

So I have a question for you guys.

How come Brian Williams is being dragged over the coals for lying in the media, but everyone who publishes these kinds of articles gets off scot-free?

If I understand correctly, Williams said that his helicopter got shot at when he was in Iraq, but in reality he was just in a helicopter in Iraq at the same time as some other helicopter nearby was getting shot at. This is obviously stretching the truth, but it seems to me it could have been worse. No important policy decisions are going to hinge upon exactly which helicopter Brian Williams was in. And he didn't get it *infinitely* wrong – for example, there was, in some sense, a war in Iraq.

On the other hand, discussions of how many poor people use drugs is pretty important for all sorts of policy questions, and these people *completely* dropped the ball. So why does nobody get reprimanded for this kind of thing?

You might argue that Brian Williams' actions were obviously malicious and deceitful, but that screwing up drug numbers is an excusable mistake. I say it's exactly the opposite. Brian Williams did exactly what I unfortunately do all the time – unthinkingly tell a story the much cooler way it should have happened, the way it happened in my head – rather than the way it actually did happen (my colleagues elsewhere in the psychiatry blogosphere go further and call this “normal brain function”).

On the other hand, I have more trouble imagining a situation in which I would accept the claim “only 0.1% of poor people use drugs, which is barely one percent of the rate in the general population” without wanting to do a little more research to see if it is true. If your reporters are capable of making this mistake honestly, *get better reporters*.

But I’m not sure it’s honest. A lot of these sources admit they took their story from a Think Progress piece on the issue. Think Progress *does* mention that the tests are a sham, although only in one sentence that is easy to miss. Either the secondary reporters didn’t read Think Progress thoroughly, or they consciously decided not to mention it.

But even if it was an honest mistake, I still have trouble excusing their arrogance. I mean look at that Jezebel article. The writer says this proves that people who think welfare recipients use drugs “consider ‘facts’ troublesome” and that their “entire social philosophy boils down to ‘Ew, poor people.'”

You’re saying that’s not as bad as a helicopter-related embellishment?

Yes, okay, drug testing welfare applicants is in fact probably a bad idea. It’s a bad idea because the courts have banned doing it in a way more effective than asking them politely if they use drugs or not, but it was a bad idea even before that. It’s a bad idea because drug tests have frequent false positives, but it’s a bad idea even without that. It’s a bad idea because quitting drugs is really hard and denying people benefits isn’t going to help.

But if, in the service of proving this to be a bad idea, you decide it’s acceptable to fudge the numbers to make your point, horrible things happen. First, you contribute to a culture of telling lies and lose the opportunity to protest when the other side does it. Second, you make it harder to trust you on anything else.

But most important, tell one lie and the truth is forever after your enemy. I recently argued that we need to reform suboxone prescribing laws, because it’s the best anti-addiction medicine we’ve got and right now poor people can’t access it. . Why should anyone listen to me now? They can just answer “Actually, that would be a waste of money. As per an article I read in Jezebel, pretty much no poor person has ever been addicted to drugs.” Then the laws don’t get reformed and people die.

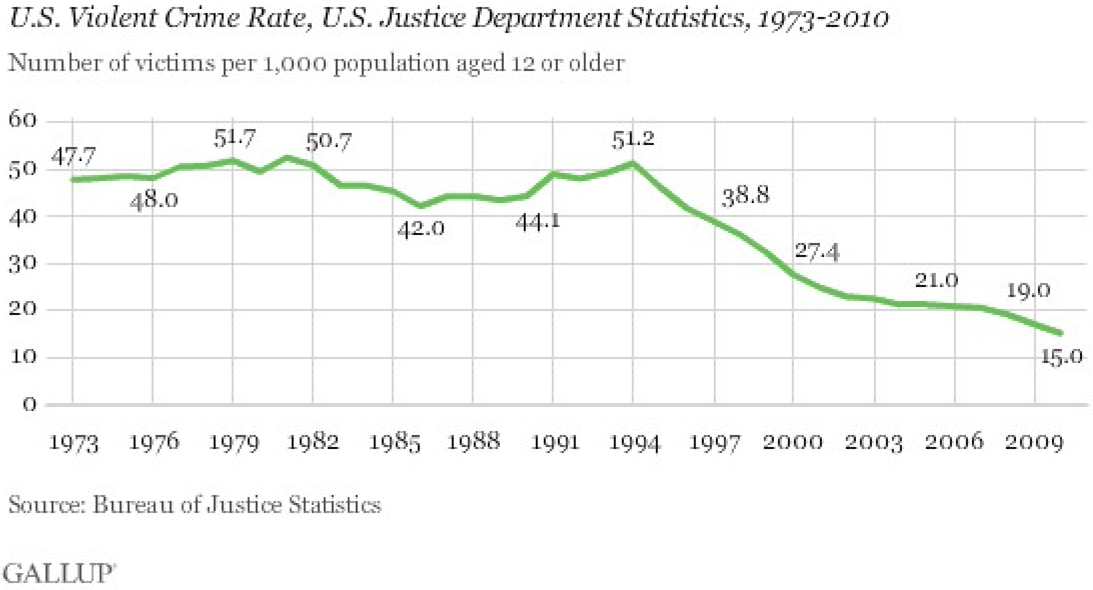
How Likely Are Multifactorial Trends?

Vox recently wrote about 16 Theories For Why Crime Plummeted In The US.

Their story is based on a report by the Brennan Center For Justice, which I haven't read, so I'm hesitant to critique it too much. The little I got off of Vox I don't like. For example, if I understand correctly they're arguing that the lead-crime connection is overblown because although lead was banned in the 1970s (thus affecting people who reached peak crime-committing age in the 1990s), the decline in crime continued even into the 2000s. But lead stays in the environment a long time, there's still a lot of work to be done eliminating various sources of lead, and so blood lead levels continue to decline. That makes their argument ring a little hollow.

But I want to talk about a more meta-level point.

The analysis ends up concluding that there is no "smoking gun" and crime probably declined because of a bunch of reasons coming together. For example, they say that "up to 12 percent of the drop in property crime during the 1990s was due to the rise in incarceration, but it was probably more like 6 percent", and "up to ten percent of the drop in crime in the 1990s was caused by hiring more police." The general picture I get is that there were about ten different factors, each explaining ten percent of the decline.


U.S. Violent Crime Rate, U.S. Justice Department Statistics, 1973-2010

Number of victims per 1,000 population aged 12 or older

Source: Bureau of Justice Statistics

GALLUP

Imagine two different perspectives on this.

First, a learned professor says “Oh yes, the public always wants to hear about how one big exciting thing caused the decline in crime, but that kind of thinking is unsophisticated. Something as complicated as crime is governed by many factors, and you certainly wouldn’t expect one big knockout change to lower it to this degree. Like everything else, it’s probably a combination of different things that came together, each accounting for a small percent of the variance.”

Second, someone counterargues: “If ten different factors caused the decline in crime, that would require that ten different things suddenly changed direction, all at the same time in 1994. That’s a pretty big coincidence. In fact, let’s reductio ad absurdum this. Imagine it was ten *million* different factors, each accounting for one ten-millionth of the decline. But that seems stupid. For example, since there are only about ten million criminals in the US, we could structure this as one factor per criminal. Imagine that, in 1994, each of America’s ten million criminals independently and coincidentally had a major life change that made crime seem less attractive. That's ridiculous. But in that case, any other explanation based on ten million factors should seem ridiculous. And if we give a heavy credibility penalty to a story with ten million factors, we should give some credibility penalty to a story with ten factors."

The second person seems to me to have a strong argument, which makes me think Vox and the Brennan Center's model where ten different trends each explain about ten percent of the decline is unlikely.

I feel like somebody has already thought about this and there's an entire literature I'm missing, but Google is failing me (badly – this was my first search result). Can somebody point me to it? Are there ways to calculate how much less likely a ten-factor explanation is than a one-factor explanation?

[EDIT: Yes, there's the trivial case where all ten factors are correlated, for example they all have to do with an improving economy. I'm talking about the non-boring version of the question.]

[EDIT2: I might have subconsciously absorbed this thought process from Stefan Schubert]

Did Falling Testosterone Affect Falling Crime?

There are already too many proposed causes for the secular decline in crime, but I can't resist suggesting one more. A couple of months ago Nydwracu asked me whether it could be related to the secular decline in testosterone. The answer turns out to be "Maybe".

This secular decline in testosterone is pretty dramatic. Our best source is A Population-Level Decline In Serum Testosterone Levels In American Men, which finds that from 1987 to 2004, average testosterone declined from 501 ng/dl to 391 ng/dl, with an even more dramatic decline in bioavailable levels of the hormone. That's about minus 1% per year.

No one knows exactly why this is happening. Some people blame increasing obesity and decreasing tobacco use (wait? Smoking increases testosterone levels? THOSE TV COWBOYS WERE RIGHT ALL ALONG!). Other people have tried to adjust for these and found they don't explain the entire effect, leading to a host of other theories. Recent scrutiny has focused on the role of feminizing chemicals in the water supply, probably a combination of industrial pollutants and discarded medications; the worst-affected areas are marked by an epidemic of transsexual fish (really).

(A quick aside – since these chemicals are gender-bending fish, frogs, and various other animals, could they be responsible for transgender in humans? This theory seems to still be in crackpot territory, but I don't know why. Research shows that male-assigned-at-birth children exposed to diethylstilbestrol in the womb are more likely to become transgender than the general population. Other than that, there just seems to be one unpublished paper on the subject. Get to work, scientists!)

Annnnnyway, testosterone has been found to correlate a bit with violent crime. In a study of 692 male criminals, Dabbs et al found that those in prison for more violent crimes had higher testosterone than those in prison for nonviolent offenses. It's hard to say exactly how much higher because they report their testosterone in a different way that doesn't correlate to anyone else's – I think part of it is that it's salivary rather than serum testosterone but it's still confusing even after I adjust for that. If we use relative rather than absolute, they do mention that 66% of inmates in the upper third of testosterone levels committed violent crimes compared to 46% in the lower third. High-T inmates were twice as likely to be in for murder as low-T inmates. Interestingly, testosterone was the highest risk factor for sex crimes, such as child molestation and (especially) rape – high-T inmates were four times as likely to be in for rape as low-T inmates. On the other hand, low-T inmates were about twice as likely to be in prison for drug offenses.

This "which criminals are worse" study is obviously not as good as an "are high-T people more likely to be criminals at all" study, but I can't fin any of those with a good sample size. You can read a review of the research here.

According to the population decline study, testosterone levels declined about 110 ng/dL in 15 years. They don't give me a standard deviation, but from this site I get one a bit less than 200. So testosterone declines by one standard deviation about 25 years? That means that a person in the top third of testosterone levels today would have been in the bottom third fifty years ago. Which – and I realize I'm doing all sorts of horrible things here to cover up my lack of actually useful data – if we extrapolate wildly from the results of these studies, we could sort of justify murder halving in about fifty years by falling testosterone alone.

The first problem with this is that we can't really use data on prison inmates as representative of the population.

The second problem is that murder has halved in way less than fifty years. It seems to have halved between 1994 and 2004.

The third problem is that crime didn't start falling until the early 1990s, but testosterone was falling since at *least* 1987 and probably earlier. This site, which doesn't cite sources, says testosterone was higher in the 1940s, though they might be confusing that with "in men born in the 1940s, as studied in the 1980s", which is of uncertain significance. Sperm count has been declining since the 30s, according to an article called Sperm Quality & Quantity Declining, Mounting Evidence Suggests

(it looks like *somebody* was not quite as virtuous as this Twitter user).

The fourth problem is that there's contradictory evidence about whether testosterone is even falling at all, according to a a study that looked at the faces of Major League Baseball players of the past 120 years. This *sort of* makes sense – face width-height ratio is affected by testosterone (one reason women's faces look different than men's) and baseball players had standardized photographs taken of them for that time period. They find that, at least based on the face ratios, testosterone was *increasing* during that period, which would be interesting if it didn't contradict everybody else. As it is, I suspect it just means baseball players were differently representative of the general population. For example, if baseball requires high testosterone, and scouts became better at selecting the highest-testosterone people over that period, that’d do it. Or if the nature of baseball changed to more of a “power game” rather than a “finesse game” (I think some people have said this) that’d do it too. Or if all baseball players suddenly started taking powerful testosterone-analogue chemicals at some point…hmmmmmmmmm…On the other hand – literally on the hand – we have the digit ratios of Lithuanians over 120 years. Someone in 1880 measured the length of Lithuanians’ fingers – which can be a proxy for testosterone levels – and then the experiment was repeated recently and the results compared. It did find the expected increase in testosterone, though no word on whether that was throughout the entire period or just concentrated in the past couple of years. So this sort of turned out to be a non-problem.

The fifth problem is that crime is dropping in women at the same rate as in men – women never really committed that many crimes, but now they’re committing fewer. Women do have some testosterone, so it’s possible that declining testosterone could affect female violence as well, but it wouldn’t be the first thing I expected. Also, I’m not sure if there are any secular trends in female testosterone levels, though I’d be fascinated to see data.

So overall while I like the approach of this hypothesis, I don’t think it gets the time window right. It would be a nice way to explain a gradual fifty-year decrease in violent crime starting in the 50s and continuing to the present day. Instead, we have a big spike in the 50s and a big drop in the 90s, which were not particularly abnormal in terms of testosterone decline.

This doesn’t really make sense to me. If testosterone is declining, it *should* cause a decrease in crime. One might argue that testosterone levels have been steadily operating behind the scenes causing very long term declines while other things account for the more visible short-term trends, but that seems like a cop-out.

I’d like to see studies comparing testosterone levels in violent criminals (both male and female) to those in the general population.

Also, we have cemeteries full of millions of dead people from every era of history, all carefully marked with what age they were when they died. Somebody needs to dig some of them up and measure their digit ratios – I assume you can still measure the digit ratio of bones, the overall length is still there. Then we can have a good answer for whether testosterone levels in men (and women) have been declining over time, when it started, and whether it’s been picking up recently. If it has been, the chance that it hasn’t had an important effect on our society worth exploring is pretty much nil.

I know, just *once* I want to get through an entire blog post without a call for disinterring the dead, but this is *important*.

Pharma Virumque

Going around the psychiatry blogosphere recently: this segment by John Oliver about doctors who take pharmaceutical company money:

I will resist the urge to geek out about its minor medical errors[1] in favor of clarifying something more important.

The impression you're supposed to get from this piece is a shady looking man handing you a briefcase full of cash and whispering "Hey, here's $10,000 for you if you prescribe unnecessary medication." The implication is the doctors who do this are awful and if *you* were in medicine you would have no trouble resisting this temptation.

In reality, pharma companies have figured out that some people have ethical qualms – "evil cannot possibly understand good" only works in movies – and adjusted their strategies accordingly.

We'll start with a simple one. Imagine you're a doctor, and your staff are complaining because the staff at *every* other doctor's office has been getting these incredible free lunches every day – the video says drug companies aren't supposed to give, like, Zagat-rated steakhouse lunches, but there's still a lot of room between "Zagat-rated" and "Way better than the peanut butter and jelly sandwich you bring from home". The nurses are grumbling and threatening to revolt and asking if you really appreciate them.

A drug company representative offers to provide your office with free lunches a couple of times a week.

You say “It would be really annoying to actually use the phrase ‘there’s no such thing as a free lunch’ here, so I will just ask what the catch is.”

They say “No catch. We don’t require you to ever prescribe any of our drugs. We don’t require you to listen to our presentation. We don’t even require you to read our promotional literature. Just accept our offer.”

You say “Why are you doing this?”

They say “Because every time you eat one of our lunches, you’ll associate the ice cold taste of Coca-Cola and the sweet warm chewy chocolate chip cookies with our company, and you’ll get positive feelings about it, and maybe those positive feelings will influence your prescription habits.”

You say “I think I’m a good enough doctor not to prescribe a drug solely because I get lunch from their company.”

They say “Look. We all know that most antidepressants are about equally effective. Sure, we split hairs and talk about how one has more anticholinergic side effects so it’s bad for patients with cholinergic sensitivity, and another has more chance of weird visual disturbances, but how often does someone come into your office and announce ‘Hey, I’m depressed, and also I have cholinergic sensitivity, but I LOVE weird visual disturbances!’? Although there are a few cases where one drug’s clearly a better choice than another, most of the time you’re about equally balanced between two or three options, and you just pick one at random. So maybe instead of picking one at random, you’ll pick the one you associate with delicious food. And if you do, so what? Nobody’s harmed. You would have just flipped a coin anyway.”

You say “I’d rather flip a coin than feel like I’m being pressured by what I had for lunch.”

They say “Look, you secretly worry anyway that you sometimes prescribe Effexor because the name makes it sound effective, or Paxil because the name makes it sound peaceful.”

You say “Wait, you can read my thoughts?”

They say “We’re a pharmaceutical company. Of *course* we can read your thoughts. Look. You already know that the mostly-meaningless choice of which of several equally effective drugs you prescribe is influenced by a bunch of silly marketing factors beyond your control. Why not add one more?”

“But -”

“Come to the Dark Side! We literally have cookies!”

Still not tempting enough for you?

Okay, imagine this. You’re a doctor and one of your patients comes in with incurable chronic pain that’s ruining their life. You try the normal medications on it and nothing works very well. There’s a high-tech next-generation medication available that you think is a good fit for your patient’s disease, but it’s not covered by their insurance and there’s no way the patient can afford it. You have to tell this guy that there’s nothing you can do for him.

Then a drug company representative comes to you bearing a big box of free samples. By “free samples” I mean hundreds of pills, enough to help the patient for the better part of a year – and maybe at the end of that time you’ll get another box of free samples. The drug rep doesn’t want you to sign your life away. She’s giving them for free, no obligation, maybe just listen to a sixty second speech on how to prescribe them safely and effectively (she wouldn’t want to give them to someone who won’t prescribe them effectively!) Are you really so fundamentalist in your approach to medical ethics that you won’t listen to a drug rep for sixty seconds in order to save a patient’s quality of life?

Most doctors – even the ethical ones who would refuse the briefcase full of cash – take the offer. This practice has come under increasing scrutiny recently. Some of the complaints are kind of dumb, but one very valid one is that a lot of the times what happens is you start off by giving the patient 100 days of free sample or something, then the free sample runs out, they’re fixated on that particular medication because it’s the one that worked for them, and they find some costly way to continue the (more expensive new) medication – instead of the two of you working harder to find some older less expensive medication that works equally well. A few drug companies have “fixed” this by giving out cards for “prescription programs” that solve some of the problems with free samples. These are even harder to resist, and they’re *also* given out by attractive drug reps who just want to tell you a few important facts about the drug before giving it to you.

Still not tempting enough for you?

Fine, then imagine this. You’re a doctor who really believes in a particular drug and is trying to convince the medical community to use more of it. For example, a couple weeks ago I wrote an article on suboxone saying it was one of the best medications for opiate abuse and I wish the medical community would pay more attention and prescribe it more often.

I wrote that article for free as a public service because I think that drug saves lives. But imagine that the company that makes suboxone approached me afterwards and said “Hey, you seem to have an important message to spread. Why don’t we sponsor you to go around the country for a week or two telling it to other doctors at medical conferences? We’ll get you first-class flights, put you up in five-star hotels, and give you a $10,000 stipend.”

I say “Wait a second, that sounds like taking pharmaceutical company money, and taking pharmaceutical company money is evil.”

They say “Look. You were trying to promote suboxone to people already. You were just doing *a bad job* because you were limited to one little blog. The more suboxone-promoting you do, the more doctors know about this drug – which you yourself have said is life-saving – and so the more lives get saved. If you’re willing to promote suboxone ineffectively for free, why not promote suboxone effectively for $10,000 plus nice hotels?”

I say “I’m still kind of uncomfortable with this.”

They say “Okay, well, it’s not *our* fault if hundreds of people die of drug overdoses because their doctors didn’t know suboxone was an option.”

You’re probably going to ask if I’ve ever accepted any of these offers. The boring truth is that I haven’t had to consider them because I’m a resident and residents are lower than dirt and the pharmaceutical companies know this and they don’t waste time trying to cozy up to us.

I have tasted the forbidden fruit only once, and it was my attending’s fault. She told us that there was a big dinner being planned for the entire psychiatric community of our city. The goal was to get doctors to meet nurses to meet therapists to meet social workers in one place so we could all get to know each other and talk about changes we could make to the system. It was very important that we attend, or else the nurses and therapists and social workers would think that the doctors were too snooty to interact with them and didn't care about changing the system. Oh, and *by the way* the dinner was sponsored by PANEXA (here used in place of the real drug because I don't want to get in trouble for calling them out) but there wouldn't be *any* promotional material or pressure to prescribe PANEXA, honest, no sirree.

This was a tempting offer precisely because it was such a good idea. Everyone in the local psychiatric community deals with each other frequently, but we'd mostly never met before. I know them as the voice on the other side of the phone saying "No, no beds are available in our facility" or as the person who refuses to fax me my patient's past medical history because the patient is too catatonic to sign a consent form. None of us are ever entirely sure what the others are doing, sometimes there are bad feelings, and it was reasonable to hope that maybe if we all met each other and socialized things would get a little smoother.

So we all meet at this restaurant, and immediately World War III breaks out. It's like "Hi, I'm Mary, the clerk at Blue Sky Mental Health." "MARY?! YOU'RE THE ONE WHO DIDN'T FAX ME THOSE RECORDS I NEEDED TWO MONTHS AGO! MY PATIENT WENT A WEEK ON THE WRONG DRUGS BECAUSE OF THAT!"

"Hi, I'm Dr. Alexander, I work at the inpatient unit in Our Lady Of An Undisclosed Location Hospital…" "WE HAD A PATIENT COME FROM THERE TWO WEEKS AGO AND HE ASSAULTED A STAFF MEMBER. IF YOU'RE A REAL HOSPITAL WHY CAN'T YOU DO PROPER VIOLENCE ASSESSMENTS?"

It turned out that the nurses hated the social workers for making them wait on the phone forever in order to get a straight answer. The social workers hated the nurses for always calling them up when they were busy about things and expecting an answer RIGHT NOW. The social workers hated the doctors for giving patients one measly prescription, then handing the case over to *them* to fix all of the impossible problems in the patient's life. The doctors hated the social workers, because when we give patients one measly prescription and then hand the case over to the social workers to fix all of the patient's impossible problems, sometimes the impossible problems don't get fixed.

Anyway, in the midst of all of this, there was one guy who was staying completely calm, talking nicely to everybody, helping people see each other's sides of the issue, just a really serene well-adjusted guy. I escaped over to his table and asked him who he was and why he was here.

"Oh," he said "I'm a paranoid schizophrenic currently on PANEXA."

Of *course* he was.

Then we all broke off into our own groups and got some incredible Italian food.[2]

What I'm saying is, pharmaceutical companies are *sneaky*.

Footnotes

1: By which I mean "succumb to the urge to geek out about its minor medical errors, but in the footnotes".

The video says that a "horrifying example" of pharmaceutical company overreach was how AstraZeneca took Seroquel, "an antipsychotic with dangerous side effects" and marketed it to doctors for depression, sleep, and dementia, adding "You can't just give people dangerous drugs and see what happens!"

But actually, lots of studies have shown Seroquel is effective for depression, lots of guidelines suggest Seroquel as a backup depression treatment, and doctors have been (correctly) prescribing it for such for a long time. Doctors also very commonly prescribe it for sleep and dementia; I think is less evidence-based, but it'd be a lie to say it wasn't common as dirt or that it didn't work for these things (safety is the problem).

So what was happening was that AstraZeneca was promoting Seroquel for the things it was actually being used for, as opposed to the thing the FDA said it was supposed to be used for. Doctors are allowed to use drugs for whatever they want based on their own analysis and their best judgment, but pharmaceutical companies are only allowed to promote it for the FDA-approved indication, which at that point was psychosis and bipolar depression.

The reason the FDA hadn't approved Seroquel for depression wasn't because it was a bad idea. It was because in order to get the FDA to approve anything for anything, you must perform the appropriate ritual of putting a zillion dollars into a big pile, then burning it as a sacrifice to the Bureaucracy Gods. AstraZeneca had performed the ritual for bipolar and psychosis, but was still in the process of performing it a third time for depression. Once they finished, the FDA approved it as an adjunctive medication for depression, but also fined them hundreds of millions of dollars because they had advertised it for depression – merely based on evidence and clinical practice – before the FDA had told them they were allowed to.

This is still not the whole story, because best clinical practice says to only use Seroquel as a third- or fourth-line antidepressant after some others have failed, and in conjunction with another medication. If AstraZeneca was advocating to use it for depression first-line on its own, this would have been a genuine overstep and something to get upset about.

(research and clinical practice say to use it for sleep and dementia approximately never, but there is enough wiggle room in that "approximately" for doctors to drive a bus through, and they do.)

This is *still* not the whole story, because The Last Psychiatrist thinks the way the FDA's handled the Seroquel indication, and the subsequent culture of prescribing that grew up based on that indication, is stupid.

The other minor medical error in the video is much simpler. Oliver mocks Wellbutrin's claim to be "the happy, healthy skinny drug" saying that "the only happy, healthy, skinny drug is amphetamine". But Wellbutrin is actually amphetamine-based – its full chemical name is 3-chloro-N-tert-butyl-β-ketoamphetamine – and it shares a mechanism of action with amphetamines, which is why some of its effects are similar as well. So Oliver's joke was a lot more accurate and a lot less funny then he thought.

2: Then later, and contrary to the promises I received, they gave us a presentation on PANEXA anyway.

The schizophrenic guy worked for one of the local psychiatric community services groups doing community outreach. I never did figure out whether he was there as a coincidence or whether the pharmaceutical company had arranged to have him there. I suspect the latter but I have no proof.

California Meetups

The following Slate Star Codex meetups are planned for the next week or so:

Berkeley, 3/1 at 2:00 PM at Indian Flavor Express, 2548 Bancroft Way, Berkeley

Stanford, 3/3 at 4:00 PM in Old Union, Room 200, 450 Serra Mall, Stanford

There’s still a little bit of planning going on around the second one, so watch for possible updates.

I have been subpoenaed by a court in Michigan to testify about one of my patients. I am trying to convince them that I am on vacation and it’s unfair to make me come back on such short notice. If they start threatening me with legal penalties for not showing up, I’ve got to cut my vacation short and these meetups will not happen (any legal experts here who can tell me a way to get out of this will be duly appreciated). Again, watch for possible updates.

Early Intervention: You *Might* Get What You Pay For

I find myself caught between the genetics community – which takes it as a given that childhood experiences and education have a very limited role in shaping life outcomes – and the psychiatric community, which takes it as a given that childhood experiences and education are crucial in shaping life outcomes. Both sides have their favorite studies to cite supporting their positions. I've already talked about the genetics studies, so I thought I'd bring up a recent particularly good study from the other side.

Dodge et al's Impact Of Early Intervention On Psychopathology, Crime, And Well-Being At Age 25 is published in last month's *American Journal Of Psychiatry*. Gratifyingly, it is a randomized controlled trial. Ten thousand kindergarteners in disadvantaged areas were screened for "conduct problems" until they found about 900 who looked like they were at high risk. 445 were randomly selected for the intervention. Another 446 stayed in the control group. The intervention was a bunch of extra classes and 'enrichment programs' from elementary school (age 5) all the way through high school (age 16). The study mentions "social skills friendship groups", "guided parent child interaction sessions", "tutoring in reading", "parent-youth groups on topics of adolescent development, alcohol, tobacco, and drugs", "youth forums on vocational opportunities", and "Oysterman's School-To-Job possible selves intervention aimed at examining emerging identity".

All of these sound so pretentious that I would have loved to be able to report that they had no effect, but in fact the opposite was true. When they caught up with these kids at age 25, the intervention group was found to have an odds ratio of around 0.6 to 0.7 of having developed various psychiatric disorders the study was testing for, including antisocial personality disorder, ADHD, depression, or anxiety. They had odds ratios around 0.7 of developing drug and alcohol abuse problems by various measures. They reported less risky sexual behavior, less domestic abuse, and fewer violent crimes. All of this was significant at the $p < 0.05$ level, and some of it was significant at much higher levels like $p = 0.001$ or below. Subgroup analysis found the data were very similar when you restricted the analysis to various subgroups like boys, girls, whites, blacks, highest-risk, lowest-risk, and by study site (it was a multi-site study). As best I can tell there were *not* an equal number of anaylses they did that came up negative that they covered up.

The apparent conclusion is that intensive interventions can change children’s outcomes and personalities in important ways ten years down the road, even regarding things believed to be highly genetic like antisocial personality disorder.

A few weak attempts to rebut this. First, there were some things that study didn’t do that one might have expected it to. It didn’t change graduation rates or employment rates. The apparent decrease in domestic violence was mediated entirely by the intervention group being less likely to have relationships (!) – the rate of domestic violence among people in relationships was the same. There was no effect on health. There was no effect on self-reported satisfaction with their parents’ parenting. There were (nonsignificantly) higher death rates and incarceration rates in the intervention group than the control group.

So if I wanted to be maximally mean to the study, I could say that whatever it's doing to violent crime and drug use has to be compatible with a (nonsignificantly) *raised* incarceration rate, and whatever it's doing to drug use and risky sexual behavior and criminality has to be compatible with a (nonsignificantly) *raised* death rate. This suggests the possibility of an attack based on their endpoints being screwy, though I'm not sure what form such an attack could take. One could argue that since many of their outcomes were based on self-report surveys maybe the kids who had been through all of the enrichment programs had grown to like the study people and had a stronger demand effect to say that they were doing great. But a lot of the survey data was backed up by court records confirming fewer drug and violence convictions. So that doesn't really work.

If you're less interested in the pure science of individual differences and more interested in policy, one fact that I forgot to mention was that this program cost $60,000 per kid. The paper points out that this is the same cost as a year or two of incarceration, so if it really changes children's life outcomes and makes tham less antisocial even that hefty price tag might be justified (although again, remember that it didn't affect employment or incarceration when checked directly).

If you're looking for an optimistic spin on that number, they freely admit they have no idea which part of their gigantic ten year intervention program produced the positive effects. It could be that all the youth forums and enrichment programs and friendship groups and so on had zero effect, and the entire benefit came from the "Oysterman's School-To-Job possible selves intervention aimed at examining emerging identity". And maybe that's a piece of paper that can be copied on a copy machine for ten cents a sheet. All this suggests is that at least *some* part of the ten-year, $60,000 intervention did something.

If you’re looking for a pessimistic spin on that number, consider. Every so often I see things that claim to have completely shifted children from the most high-risk of high-risk groups to upstanding successful members of society by giving them a year or preschool, or a couple of after-school lessons, or something like that. And these studies always boast that they did it with only $1000 or $5000 or some number like that, so it’s nice and cost effective. So far, the studies I have seen like this have been wrong. And so far I have not been surprised, because we *already* spend between $100,000 to $200,000 per child on education and various social programs. If someone ever found a social program that really worked for $1,000, the first thing we would want to do is tar and feather everyone currently in our bureaucracy of social programs, for being so incompetent that changing their $200,000 in spending to $201,000 in spending (with the extra $1000 going to someone besides them) could completely revolutionize life outcomes.

This study seems more in line with everything else. By going from $200,000 to $260,000, we can slightly push a few things in a positive direction a little bit more, maybe. From a scientific view, it’s pretty interesting. From a policy view, it’s nothing to write home about.

California Meetups (Rererevised)

The following Slate Star Codex meetups are planned for the next week or so:

Berkeley, Sunday 3/1 at 2:00 PM at Indian Flavor Express, 2548 Bancroft Way, Berkeley

San Jose, Monday 3/2 at 7:30 PM at David Friedman's house, 3806 Williams Rd, San Jose

Googleplex, Tuesday 3/3 at 1:00 PM *at the picnic tables between CL2 and CL5, maybe*. Googlers can get more information at go/ssc-meetup

Stanford, Tuesday 3/3 at 6:00 PM, *at the picnic tables outside Tresidder Union, 459 Lagunita Dr, Stanford*

across the street from Old Union, Room 200, 450 Serra Mall, Stanford, at the picnic tables outside Tresidder

EVERYONE is welcome to come to Berkeley and San Jose, even if you don't read this much, don't comment, whatever. Googleplex is limited to Google employees, unless you have another way to get in (NOT to be interpreted as a challenge!). Stanford is mostly intended for Stanford students, but if you're in the area and really want to come I'm sure we can fit a few others in.

A Series Of Unprincipled Exceptions

> The more consistently one attempts to adhere to an ideology, the more one's sanity becomes a series of unprincipled exceptions.
>
> — graaaaaagh (@graaaaaagh) February 5, 2015

Meeting with a large group of effective altruists can be a philosophically disconcerting experience, and my recent meetup with Stanford Effective Altruist Club was no exception.

Buck forced me to pay attention to an argument I've been carefully avoiding. Most people intuitively believe that animals have non-zero moral value; it's worse to torture a dog than to not do that. Most people also believe their moral value is some function of the animal's complexity and intelligence which leaves them less morally important than humans but not *infinitely* less morally important than humans. Most people then conclude that probably the welfare of animals is moderately important in the same way the welfare of various other demographic groups like elderly people or Norwegians is moderately important – one more thing to plug into the moral calculus.

In reality it's pretty hard to come up with way of valuing animals that makes this work. If it takes a thousand chickens to have the moral weight of one human, the importance of chicken suffering alone is probably within an order of magnitude of all human suffering. You would need to set your weights *remarkably* precisely for the values of global animal suffering and global human suffering to even be in the same ballpark. Barring that amazing coincidence, either you shouldn't care about animals at all or they should totally swamp every other concern. Most of what would seem like otherwise reasonable premises suggest the “totally swamp every other concern” branch.

So if you’re actually an effective altruist, the sort of person who wants your do-gooding to do the most good per unit resource, you should be focusing entirely on animal-related charities and totally ignoring humans (except insofar as humans actions affect animals; worrying about x-risk is probably still okay).

I acknowledged the argument was very convincing, but told Buck that I was basically going to safe-word out of that level of utilitarian reasoning, for the sake of my sanity.

Buck pointed out that this shouldn’t be too scary, given that many utilitarians have already had to go through a similar process. Peter Singer talks about widening circles of concern. First you move from total selfishness to an understanding that your friends and family are people just like you and need to be treated with respect and understanding. Then you go from just your friends and family to everyone in your community. Then you go from just your community to all humanity. Then you go from just humanity to all animals.

By the time most people figure out what they’re doing they already accept at least friends, family, and community. But going from “just my community” to “also foreigners” is a difficult step that’s kind of at the heart of the effective altruism movement. In the same way that allowing animals into the circle of concern totally pushes out the value of all humans, allowing starving Third World people into the circle of concern totally pushes out most First World charities like art museums and school music programs and holiday food drives. This is a scary discovery and most people shy away from it. Effective altruists are the people who are selected for not having shied away from it. So why shy away from doing the same with animals?

It’s a good question. After thinking about it for a while, I think my answer is that I never actually completed the process of widening my circles of concern and neither has anybody else, and because I’m thinking about this one in an abstract intellectual way I’m imagining actually completing it, which would be much scarier than the incomplete things I’ve done before.

Like, although I acknowledge my friends and family as important people whom I should try to help, in reality I don’t treat them as *quite* as important as myself. If my brother asked me for money, I’d lend it to him, but I wouldn’t give him exactly half my money no-strings-attached on the grounds that he is exactly as important to me as I am.

Likewise, although I acknowledge strangers as important people whom I should try to help, in reality I don’t treat them as *quite* as important as my friends. We all raised a lot of money to help Multi when she was in a bad situation, but there are thousands of other people in the same exact same bad situation and we’re not putting nearly as much effort into them.

You can try to justify this in terms of “well, I know myself better than I know my brother, and I know Multi better than I know strangers, so I’m more *effective* at helping me and Multi, so I’m just rationally doing the things that would have the most impact”. But I think if I bothered to dream up some thought experiment where that wasn’t true, I would prefer to help me and Multi to my brother and random strangers even after that factor had been controlled away.

This doesn't come as a surprise to me and I'm not sorry. But…well…I guess my worry about the animal charity thing wasn't that I was inconsistent, so much as that I was being meta-inconsistent; that is, I didn't even have a consistent set of rules for deciding whether I was going to want to be consistent or not.

And now I think I might have a consistent policy of allowing *some* of my resources into each new circle of concern while also holding back the rest of it for the sake of my sanity. Thus my endorsement of GiveWell's principle that you should donate at least 10% of your income to charity, but then feel okay about not donating more if you don't want to. I am allowed to balance resources devoted to sanity versus morality and decide how much of what I have I want to send into each new circle of concern – without denying that the circle exists.

I think that armed with this idea I am willing to accept Buck's argument about animal welfare being more important than human welfare, insofar as this means I should donate *some* resources to animal welfare without necessarily having to give up caring about human welfare completely. I don't think I can make a principled defense of doing this. But I think I can claim I'm being unprincipled in a meta-consistent and effectively sanity-protecting way.

A Cascade Of Dunbar Numbers

> Never doubt that a small group of thoughtful, committed citizens can become a large group of thoughtless, committed citizens.
>
> — Member Of Species (@MemberOfSpecies) January 27, 2015

Back when I used to play at micronations, we had grand dreams of creating a fully-fleshed out simulated country with hundreds of citizens and dozens of Byzantine government departments and industries. For a while, these dreams seemed attainable. Most of the people we advertised to were really interested, and we went from one or two founding members to about a dozen people very quickly.

Then we stopped. For the past fifteen years or so, the population of Shireroth remained fixed at between ten to twenty people. It wasn’t always the same ten or twenty – some new people entered, some old people left – but it always stayed within that window.

Ten years ago we got really tired of this and spent a decent amount of money on an advertising campaign. It was a good campaign, and we got maybe thirty or forty people to apply for citizenship at once. We thought this was going to be so awesome – the size of the simulation quintupling within a week. Instead, the new people just kind of wandered around aimlessly, bumping into things, never quite figuring out what was going on, never quite taking an active role in anything or showing any initiative, finally showing up less and less until two months later the population was back down to about a dozen people.

Our Ministry of Immigration and Naturalization, which was in charge of recruiting new people, kept beating its head up against the same sort of idea with about the same results. I thought this was dumb, so I manuevered my way into the Ministry and shifted all of our resources from recruitment to retention – setting up mentoring programs for new citizens, writing a bunch of guides for what they were supposed to do, trying to befriend them and integrate them into our social scene. The result was that our new immigrants maybe stayed three months instead of two months before evaporating.

We were only ever able to find one successful method of recruiting new micronationalists: wait for random people around the world to learn about micronations, start their own projects without any impact or influence from us, integrate those projects into the larger community, and then if those projects failed after a year or two sometimes their members would join us instead. This sort of worked, but never enough to get us more than ten or twenty people. And those other projects, the ones they made, very rarely had more than ten or twenty people either, despite *hordes* of people who said they thought micronations were interesting and agreed to participate for a few months. It was a weird problem, and one I was never able to solve before I left the hobby.

I was thinking about this recently because of some people’s complaints about the Bay Area rationalist community.

I have always had universally positive experiences with the Bay Area rationalist community. I arrived in about 2012 when it was still smallish, and I got to know the group involved and hang out with fascinating people like Eliezer Yudkowsky and Anna Salamon and Michael Vassar and learn a lot from them.

The complaints I heard were something like this. A lot of people join rationalist communities in Missouri or Italy or Australia or some far-flung place like that and really like them and feel like they have a great tight-knit community. They decide that if a small rationalist community is good, a ten-times-larger rationalist community will be ten times as good. Then they get here and everyone's already doing their own thing, and it turns out Eliezer and Anna and Michael do not have enough time to personally interact with every single person in San Francisco on an intimate basis, and they kind of hang around the edges of the community not really knowing what to do or being connected to its general rhythm. It doesn't help that a lot of people (DESPITE MY BEST EFFORTS) keep thinking that rationality will turn them into supermen and get upset that they've been in the Bay Area like *an entire month* and are not yet Elon Musk.

And I was thinking of both of these things when I met Samo in Oakland and he told me about his research. It's kind of complicated, but when I asked him "Is this basically hypothesizing a cascade of Dunbar numbers?" he told me I wasn't far off.

Dunbar's number, remember, is the theory of anthropoligst Robin Dunbar that humans have a hard-coded optimal group size. He found that most primates' brain size correlates with the average size of their social groups. If you extrapolate the connection to humans, you would expect humans to have a social group of about a hundred fifty people (other estimates find anywhere between 100 and 250). Various researchers and amateurs have claimed to corroborate this in every context from hunter-gatherer bands to corporations to MMORPG guilds (1, 2), sometimes more believably than others.

Samo said his own research had found several of these discontinuities – I think he mentioned about 12 people, about 150 people, about 1,000 people, and about 90,000 people as the first few. He tried to tie this in to government forms like “family”, “clan”, “tribe”, and “city-state”. I think the idea was supposed to be that we’re naturally very good at dealing with groups of family size, our brains will grudgingly allow us to deal with groups of up to clan size, with the low-hanging fruit in social technology we can stay together in groups of tribe size, with some higher-hanging fruit in groups of up to city-state size, and so on. When you get to the largest organizations in the world (China and the Catholic Church, if you’re wondering), you have to pull out all the stops – a pyramid of subdivisions of subdivisions of subdivisions, each choosing representatives to the next-highest level which then chooses its own representatives and so on until you reach one guy at the top. He suggested that if my old micronation or the Bay Area rationalist community want to get past their growth limitations, they should try something similar – which is good advice, except that in fifteen years of trying to simulate a country it *did* occur to us to have provinces and that didn’t seem to help much.

I am not sure how much I buy this theory. Having a systematizing disposition, it is really tempting for me to start carving up reality at these lines – startups with < 12 people, small businesses with < 150, et cetera - or to start noting that most of the college clubs I've seen have like a dozen core members, no matter how big the college is. But I find myself engaging in certain thought patterns that I have tagged as "trying too hard to fit data to a theory". At University College Cork the debating club often got twenty or thirty people; and if someone told me the natural way to slice businesess was <5 people, <30 people, et cetera, I could kind of see that too. And also there are so many other factors - whether a group is online vs. meatspace, whether it's living together versus meeting once every few weeks, whether it's strongly enforced (like a company or military unit) or voluntary (like a social club), whether we expect its members to also have other relationships somewhere else (members of MMORPG guilds presumably have real-life friends and coworkers as well, but members of hunter-gatherer tribes might know literally nobody except their cotribesmen). Even if there were interesting limits to the size of cohesive social groups, I would expect these factors to make them different in different domains - the size restrictions governing the growth of online micronations should be different from the size restrictions governing the growth of startups, even though in practice they're often around the same size. But I do notice that the Bay Area rationalist community is probably around 150 people and having trouble growing, and the size of my micronation was around 12 people and had trouble growing. And that after ten years in micronations, the rule that no matter what happened we would never get more than a dozen or so people seemed like a *law of nature*. And that I just checked after two years outside the community, and it looks like one new person came in to replace me and no one else, and they’re *still* stuck around twelve. So there does seem to be this very real community failure mode where everyone is socially saturated with each other and new people feel like they can’t break in. I’m less confident it always happens at a specific number or set of numbers across all domains of community – but who knows?

Effective Altruists: Not As Mentally Ill As You Think

During my recent meetings with effective altruist groups here, I kept hearing the theory that effective altruism selects for people with mental disorders. The theory is that people with a lot of depression, anxiety, and self-hatred turn to effective altruism as (optimistically) a way to prove that they are good and valuable or (pessimistically) a form of self-harm in which they enact their belief that they deserve nothing and other people are more worthy.

And whenever this got brought up at meetings, people giggled, probably because they were thinking of good examples. I can't deny there's a lot of anecdotal evidence here (hi Ozy!). But when I look into it, it seems totally false.

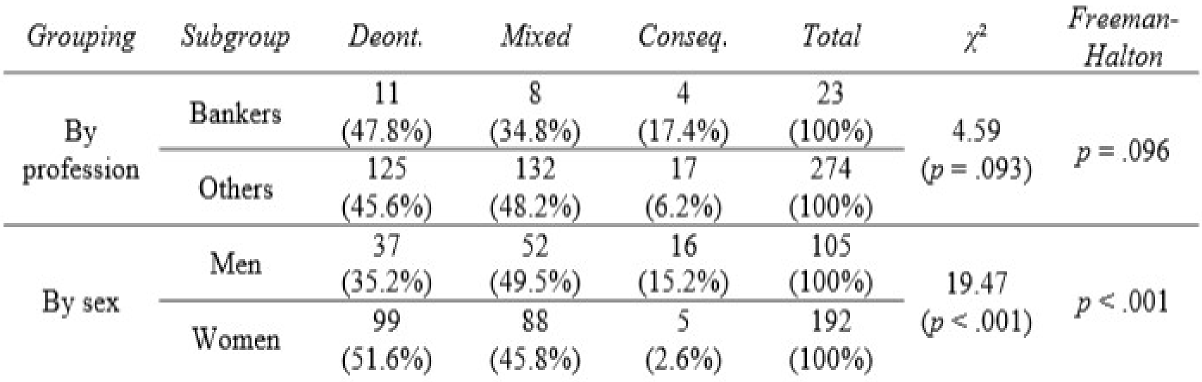
My source was the 2014 Less Wrong survey data, which asked respondents whether they self-identified as effective altruists and whether they participated in effective altruist groups and meetups. Using that question, I separated the respondents into 758 non-effective-altruists and 422 effective altruists. The survey had also asked people whether they had been diagnosed with various mental illnesses, so I checked the rates in both groups. Including self-diagnosis there were no particular results; when I limited it to professionally diagnosed illnesses things got a little more interesting.

Effective altruists had about the same levels of anxiety disorders and obsessive-compulsive disorder as non-EA Less Wrongers. However, they had slightly higher levels of depression (22% vs. 17%) which was barely significant ($p = 0.04$) due to a large sample size. They also had more autism (8.5% vs. 5%) which was also significant ($p = 0.02$).

I expected this to be mediated by a tendency for autistic people to be more consequentialist and consequentalists to be more EA, and both these things were true to some degree, but even when I limited the analysis to all consequentialists, effective altruists still had more autism. Further, autistic people seemed to donate a higher percent of their income to charity than neurotypical people or people with other mental illnesses *even separated from effective altruist status* – that is, even among people none of whom were effective altruists, the autistic people seemed to donate more (effect not always significant) even though they generally had lower incomes.

I conclude that effective altruists are not unusually self-hating or scrupulous, but that they may be a little more autistic, and the reason why isn't the obvious one.

A caveat, by way of presenting another interesting result. Rusch (2015) (h/t @bechhof) studied whether bankers were more consequentialist (in this case, more likely to give consequentialist answers to the Trolley Problem and Fat Man Problem) than nonbankers. He found that they were. But then he checked for confounders and found the result was entirely an artifact of men being more consequentialist than women and bankers being predominantly male.



Grouping	*Subgroup*	*Deont.*	*Mixed*	*Conseq.*	*Total*	χ^2	*Freeman-Halton*
By profession	Bankers	11 (47.8%)	8 (34.8%)	4 (17.4%)	23 (100%)	4.59 (p = .093)	p = .096
	Others	125 (45.6%)	132 (48.2%)	17 (6.2%)	274 (100%)		
By sex	Men	37 (35.2%)	52 (49.5%)	16 (15.2%)	105 (100%)	19.47 (p < .001)	p < .001
	Women	99 (51.6%)	88 (45.8%)	5 (2.6%)	192 (100%)		

This is pretty astounding – men are almost six times as consequentialist as women!

On the other hand, in both my Less Wrong data in general and the effective altruist subgroup, men and women don't vary much in consequentialismishness. Either Rusch's data is wrong, or there's a strong filter that acts to get only consequentialists into Less Wrong regardless of gender, or LW converts women to consequentialism (without further converting men).

Interestingly, effective altruists were *much* more consequentialist than non-effective-altruist LWers – 80% versus 50%. They also had more women than the non-effective-altruists. So it looks like LW filters for consequentalists so strongly it gets an even balance of consequentialist men and consequentialist women, and past that stage, filtering further for consequentialism doesn't change gender balance much.

This points out a limitation of my statistics above. All it shows is that effective altruists don't differ *from other rationalists* in levels of mental illness. It's possible and indeed likely that both effective altruists and rationalists differ from the general population in all kinds of ways. It's even possible that self-hate and scrupulosity drive people into the rationality movement in general, although I can't imagine why that would be. It's just that they don't seem to have any extra power to make people effective altruists once they're there.

Too Good To Be True

Related To: You Might Get What You Pay For, Do Life Hacks Ever Reach Fixation?

Here are three interesting psychological studies:

1. Kirschenbaum, Malett, and Humphrey gave students a three month course in how to make monthly plans, then followed them up a year later to see how well their grades were doing. The students who made the plans got an average GPA of 3.3 compared to the students who didn’t getting 2.5. They concluded that plan-making skills are academically important, and that benefits persist at least one year after completion of the plan-making course.

2. Aronson asked Stanford students to write a letter to a middle school “pen pal” urging them to adopt a “growth mindset”. Since this was a psych study, it was all lies and there was no pen pal; the study examined whether writing a letter urging growth mindset made the students themselves have a growth mindset and whether this improved grades. Three months later, the students who wrote the letter had higher GPAs.

3. Oaten and Chang helped undergraduates set up an 8-week time management program involving schedules and diaries. The students who participated not only had better time-management, they also studied more, smoked less, drank less alcohol, exercised more, ate a better diet, spent less money, rated their emotions as better, missed fewer appointments, and were less likely to leave dishes in the sink (really!).

So, remember a couple of weeks ago when I wrote about some psychiatrists conducting a really big (n =~ 1000) study about an early intervention program for troubled youth? A program that cost $58,000 per person and lasted ten years?

And remember how, although it was deemed a success, it was deemed a success because it had *modest* effects on a couple of outcomes, without improving the really big ones like school retention, employment, or incarceration?

So on the one hand, having a short discussion about making monthly plans will boost your GPA almost a whole point a year later. On the other hand, ten years of private tutoring and pretty much every social service known to mankind will do next to nothing.

This suggests a dilemma: either psychological research sucks or everything else sucks.

I mean, you tell your social engineer “Here’s ten thousand dollars and a thousand hours of class time per pupil per year, go teach our kids stuff,” and they try their hardest.

And then some researcher comes along, performs a quick experimental manipulation (the pen pal one probably took 30 minutes and 30 cents) and dramatically improves outcomes over what the social engineer was able to do on her own.

Then one gets the impression that the social engineer was not using their $10,000 and 1K hours very wisely.

And if it were just the one example, then we could say Carol Dweck or Roy Baumeister or whoever is a genius, the rest of us couldn’t have been expected to come up with that, now that we know we’ll reform the system. But these results have been coming in several times a year for decades. If we’ve been adopting all of them, why aren’t people much better in every way? If we haven’t been adopting them, why not?

My money is on the other branch of the dilemma. The reason the $58,000 study got so much less impressive results is that it was run by medical professionals to medical standards, meaning it only showed the effects that were really there. The reason psychology gets such impressive results is…

Okay. I have only skimmed these three studies, so I don’t want to make it sound like I’m definitively crushing them. But here are some worrying things I notice.

The first study results are actually limited to a small subgroup with I think a single-digit number of students per cell.

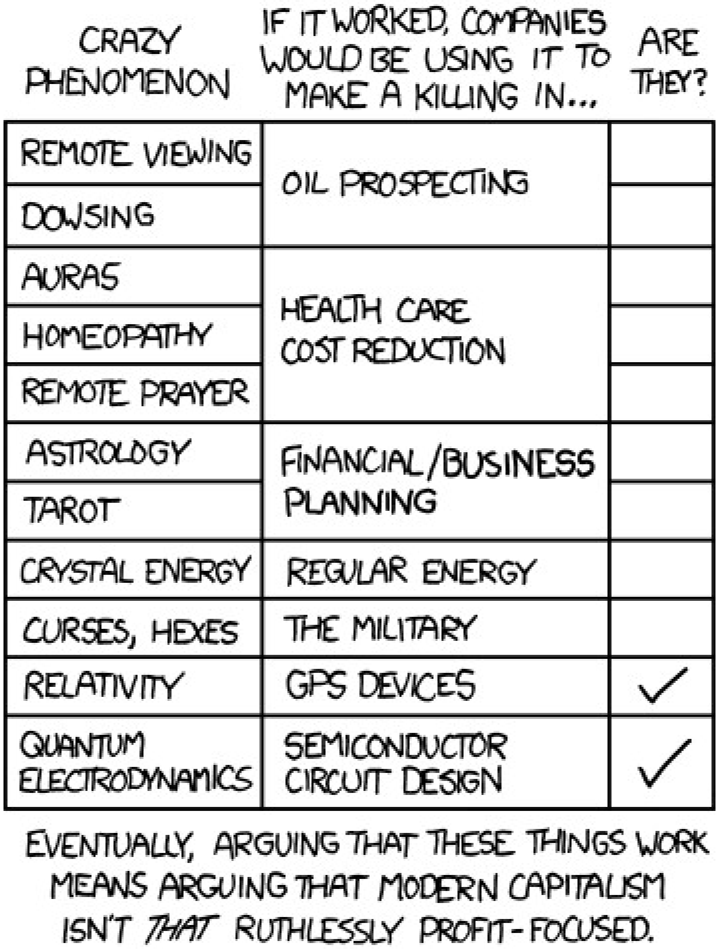
The second study results work only on a complex statistical manipulation and disappear when you do basic correlation; further, although the manipulation is supposed to work by increasing trait growth mindset, the correlation between trait growth mindset and academic achievement when correlated directly is actually negative across all variables and in some cases significantly so (!)

The third study is really about stress-related behaviors during an examination period, meaning that all they showed was that people who as part of their time management course were forced to study in a carefully scheduled way for a term show less stress-related behavior during the term exam period, which makes more sense as they probably studied more earlier, had less studying left to do during the exam period, had more free time, and were less stressed.

This context was dropped by the popular science press, turning the study into proof that good time management *in general always* produced all of these effects.

And although there are several other studies I have *not* been able to find equally worrying flaws with, if they report massive long-term gains from seemingly minor interventions, I expect they're there and I just missed them.

Basically, you remember this chart?



CRAZY PHENOMENON	IF IT WORKED, COMPANIES WOULD BE USING IT TO MAKE A KILLING IN...	ARE THEY?
REMOTE VIEWING	OIL PROSPECTING	
DOWSING		
AURAS	HEALTH CARE COST REDUCTION	
HOMEOPATHY		
REMOTE PRAYER		
ASTROLOGY	FINANCIAL/BUSINESS PLANNING	
TAROT		
CRYSTAL ENERGY	REGULAR ENERGY	
CURSES, HEXES	THE MILITARY	
RELATIVITY	GPS DEVICES	✓
QUANTUM ELECTRODYNAMICS	SEMICONDUCTOR CIRCUIT DESIGN	✓

EVENTUALLY, ARGUING THAT THESE THINGS WORK MEANS ARGUING THAT MODERN CAPITALISM ISN'T *THAT* RUTHLESSLY PROFIT-FOCUSED.

Source: xkcd

In "Crazy Phenomenon", add "any large and persistent effect from social psychology". In "If it worked…" add “education, rehab, and mental health”. In “Are They?”, add “not nearly as much as I would expect”.

Book Review: Willpower

I.

Willpower: Rediscovering the Greatest Human Strength by Roy Baumeister and John Tierney attracted me with the following pitch: there are only two quantities in psychology that have been robustly linked to a broad range of important life outcomes. One is IQ and isn't changeable. The other is willpower and is easily changeable. Therefore, study willlpower.

I expected the book to center around Baumeister's groundbreaking experiments in "ego depletion" – where people who are forced to expend willpower on one task have less willpower left over for future tasks – and "ego repletion", where people who have been depleted of willpower get it back after taking some glucose. I'd been left kind of confused by competing claims about those studies and I hoped this book would fill out my knowledge of them and settle my confusion.

Instead, it spent a couple of chapters mentioning their existence and praising them as revolutionary, and then got deep enough into Pop Science Self-Help Book Mode to be almost a self-parody. Chapter three is called "The To-Do List From God To Drew Carey" and illustrates why to-do lists are a great idea with anecdotes from Drew Carey's life and the Bible. Chapter four is the same but with Eliot Spitzer. Chapter six is David Blaine, Chapter seven is H.M. Stanley, Chapter eight is Eric Clapton, and Chapter ten is Oprah. All of these people apparently have important lessons about willpower to teach us, of which the interesting ones are:

– Willpower is a limited resource that is depleted by use and restored by glucose

– Having to make too many small decisions in a day causes “decision fatigue”, leading you to be exhausted and make bad decisions.

– Using willpower a lot strengthens your willpower, allowing you to be more effective.

– People with more willpower do much better in life; for example, in the famous “marshmallow test”, children who were able to resist eating a marshmallow for a few minutes in order to get two marshmallows had better life outcomes twenty years later.

– Careful quantification of goals, setting precommitments, and publicizing your success or failure to other people helps you stick to resolutions. (If you are wondering exactly how to do this, this might be a good time to mention there’s a new ad on the sidebar for Beeminder, a company that manages this for you with some neat evidence-based tricks.)

– Religious people have more willpower than non-religious people for some reason. But you have to believe if you want to get this benefit – you can’t just hang out at church and go through the motions in order to reap the willpower gains.

– Chinese-Americans do better than white Americans on willpower tests from toddlerhood toward adulthood. This is the most likely reason Chinese people outperform whites in the real world, and in fact although average white and Chinese IQ are pretty similar, Chinese people can break into “elite” professions at a lower IQ threshold than whites because their increased self-control compensates. The book admits this may be partly genetic, but also attributes some of it to Chinese parents teaching their children discipline and setting hard goals, which they contrast with white parents who tell their kids to “have fun” and “have high self-esteem” and “be self-directed”. Obviously in order to believe this result you’d have to believe parenting styles can affect children, which this book takes on faith but which is on shaky ground.

– Amount of self-control does not affect body weight (!) or success at dieting (!!) very much.

Overall there were some interesting findings in here, even if I found the pop sci tone a little bit over-the-top after a while.

II.

But I was disappointed. The reason I got this book is that there’s a big debate going on over “willpower”, “ego depletion”, and their younger cousins “growth mindset” and “grit”. All of these are slightly different constructs, but they’re all measures of stick-to-it-ness, and all of their proponents make grand claims about how small interventions to make people have them will bring those people success at school, work, and life. On the other hand, there are a lot of other people who think this whole area is a load of bunk.

For example, Mischel’s marshmallow test started all this off by “proving” that children who were able to delay gratification longest had higher SAT scores, higher parent-rated competence, and better coping skills. But now the test is under fire as people question whether it just shows that people from good environments learn to be more trusting and so more likely to believe the researcher’s promise of extra rewards for delayed gratification. Other people ask if it just shows that kids who are smart enough to think of good strategies to distract themselves are also smart enough to get good SAT scores and all of the other positive correlates of high IQ. Still other people point out the very low sample size, the reverse correlation in other subgroups, and an apparent failure to replicate. Mischel fires back in an Atlantic article where he says *of course* he took these things into account, that the test was done on a homogenous upper-class population.

This book just says the marshmallow test proves willpower is important, and leaves it at that.

Or how about the idea that glucose is the limited resource that willpower depletes? Robert Kurzban very correctly points out that the metabolic math doesn't come close to adding up – we know how much glucose things in the body use, and these short little willpower tasks aren't really going to affect blood glucose levels at all. Also, turns out that if you rinse your mouth out with a tasty glucose solution, you get just the same amount of ego replenishment even though none of the glucose actually entered your body. And for that matter, how come I can't get infinite willpower just by snacking while I work? How come M&Ms don't work as a poor man's Adderall?

Kurzban goes further and says he doesn't believe in willpower as a limited resource at all. He notes that even when you've stopped studying because you're too ego-depleted and exhausted to make yourself go on any further, if I offer you a million dollars to study another hour then you'll do it. Guess that resource wasn't so depleted after all. He proposes a different model of willpower, where it's your brain's way of nagging you about the opportunity cost of your actions – "you've been studying three whole hours, don't you think you could use that lobe of your brain for something else now?" But this strikes me as ridiculous – my brain is *very* concerned that it has better things to do than study, but is perfectly happy with me playing *Civilization IV: Fall From Heaven* forever or simply lying in bed doing nothing?

Finally, Carol Dweck finds that willpower is only depletable if you *think* it is, which sounds like *exactly* the sort of thing Carol Dweck would find. If Carol Dweck ever became an oncologist, we would have to revise all the medical textbooks to say that people only get cancer if they *think* they will.

There is a meta-analysis of about a hundred studies said Baumeister was basically right about everything. On the other hand, Baumeister's theory failed what sounds like a formal replication. So it's complicated.

I really want to know what willpower is. It seems like one of the big challenges of my life; I never have enough willpower for everything I want to do, and I'd at least like to have a theory of what I'm up against. This book did not give me enough information to navigate the controversy. Instead, it totally denied there was any controversy and spent chapter after chapter on cute pieces of trivia interspersed with stories about Drew Carey and Oprah.

III.

So let me end this with some thoughts that a good explanation of willpower should take into account.

First, mental willpower seems a lot like physical willpower – by which I mean our ability to push through exhaustion to keep exercising. Kurzban complains that willpower can't be a limited resource, because even after it's all depleted you can still force yourself to keep going for a big enough reward. But the same is true of exercise. I'll get exhausted and stop running after a certain number of miles, but if you offer me $1 million to run another one I’ll probably make it. And the studies showing that rinsing your mouth out with glucose gives you the same willpower boost as actually drinking is identical to the results of similar studies measuring exercise duration.

So it’s probably worth asking what exactly causes exercise fatigue. This is less firmly known than I expected, but it seems to be a combination of decreasing levels of inputs (especially glycogen stores in the muscles), buildup of toxic metabolic waste (including heat, lactic acid, etc), and cellular damage. Exercise long enough and your muscles need time to replenish their stores, clear away waste, and repair themselves.

Second, mental willpower seems to do something a lot like budgeting. Money, despite being pretty much the classic example of a limited resource, shares some of the features of willpower that Kurzban mocks. A few days ago a pipe burst, my house flooded, and I may have to spend several thousand dollars unflooding it. This means I will have no money for the next forever and then some, and if you ask me to meet you at a fancy restaurant I’ll probably refuse on financial grounds. But if my mother is on her deathbed and her final wish is to see me one last time, I will find the money to get an expensive flight to California. How’s that any different from me not having enough willpower to keep studying *until* you offer me a million dollars to do so?

Likewise, if tomorrow my boss offers me a $5000 bonus to be handed out next month, I’ll probably relax my budgetary constraints and start buying myself nice things even before I get the paycheck. How’s that any different from the body reacting to sugar when it’s in the mouth but not the bloodstream?

Finally, the question that I as a psychiatrist find most interesting – how come drugs can change willpower so dramatically? The guy who can't concentrate on a project for more than five minutes straight will pull a whole week of all-nighters when he's on Adderall or modafinil. Those certainly aren't increasing blood glucose, so what's up?

The model that makes the most sense to me is of a stupid default system running on short-term reinforcement learning, plus an evolutionarily novel (and therefore poorly implemented) executive system that can overrule the default. The executive system's overrule isn't a simple veto, but a constant action, the same way holding your hand high in the air for a long period requires constant action by your muscles. This effort is metabolically costly in the same way that using muscles is metabolically costly, and so your body runs a general-purpose budgeting function on it that convinces you to turn it off before it overheats. Given enough incentive, you can let it overheat, but then it's going to be damaged and need to repair itself for a few days before you can use it effectively again. This seems to fit all the evidence *except* the drugs, which I interpret as acting on the default system so that you don't *need* to bring in the executive planner. I freely admit this is sort of cheating.

Overall I recommend *Willpower* if you want a quick and fun survey of a bunch of loosely connected psychology topics, but not if you want a deep and balanced exploration into the literature of anything.

Answer to Job

(with apologies to Jung)

Job asked: “God, why do bad things happen to good people? Why would You, who are perfect, create a universe filled with so much that is evil?”

Then the Lord spoke to Job out of the whirlwind, saying “WHAT KIND OF UNIVERSE WOULD YOU PREFER ME TO HAVE CREATED?”

Job said “A universe that was perfectly just and full of happiness, of course.”

“OH,” said God. “YES, I CREATED ONE OF THOSE. IT’S EXACTLY AS NICE AS YOU WOULD EXPECT.”

Job facepalmed. “But then why would You also create *this* universe?”

Answered God: “DON’T YOU LIKE EXISTING?”

“Yes,” said Job, “but all else being equal, I’d rather be in the perfectly just and happy universe.”

“OH, DON’T WORRY,” said God. “THERE’S A VERSION OF YOU IN THAT UNIVERSE TOO. HE SAYS HI.”

“Okay,” said Job, very carefully. “I can see I’m going to have to phrase my questions more specifically. Why didn’t You also make *this* universe perfectly just and happy?”

“BECAUSE YOU CAN’T HAVE TWO IDENTICAL INDIVIDUALS. IF YOU HAVE A COMPUTATIONAL THEORY OF IDENTITY, THEN TWO PEOPLE WHOSE EXPERIENCE IS ONE HUNDRED

PERCENT SATURATED BY BLISS ARE JUST ONE PERSON. IF I MADE THIS UNIVERSE EXACTLY LIKE THE HAPPY AND JUST UNIVERSE, THEN THERE WOULD ONLY BE THE POPULATION OF THE HAPPY AND JUST UNIVERSE, WHICH WOULD BE LESS GOOD THAN HAVING THE POPULATION OF THE HAPPY AND JUST UNIVERSE PLUS THE POPULATION OF ONE EXTRA UNIVERSE THAT IS AT LEAST SOMEWHAT HAPPY."

"Hmmmmm. But couldn't You have have made this universe like the happy and just universe except for one tiny detail? Like in that universe, the sun is a sphere, but in our universe, the sun is a cube? Then you would have individuals who experienced a spherical sun, and other individuals who experienced a cubic sun, which would be enough to differentiate them."

"I DID THAT TOO. I HAVE CREATED ALL POSSIBLE PERMUTATIONS OF THE HAPPY AND JUST UNIVERSE AND ITS POPULACE."

"All of them? That would be…a lot of universes."

"NOT AS MANY AS YOU THINK." said God. "IN THE END IT TURNED OUT TO BE ONLY ABOUT 10^(10^(10^(10^(10^984)))). AFTER THAT I RAN OUT OF POSSIBLE PERMUTATIONS OF UNIVERSES THAT COULD REASONABLY BE DESCRIBED AS PERFECTLY HAPPY AND JUST. SO I STARTED CREATING ONES INCLUDING SMALL AMOUNTS OF EVIL."

"Small amounts! But the universe has…"

"I WAS NOT REFERRING TO YOUR UNIVERSE. I EXHAUSTED THOSE, AND THEN I STARTED CREATING ONES INCLUDING IMMENSE AMOUNTS OF EVIL."

"Oh." Then: "What, exactly, is Your endgame here?"

“I AM OMNIBENEVOLENT. I WANT TO CREATE AS MUCH HAPPINESS AND JOY AS POSSIBLE. THIS REQUIRES INSTANTIATING ALL POSSIBLE BEINGS WHOSE TOTAL LIFETIME HAPPINESS IS GREATER THAN THEIR TOTAL LIFETIME SUFFERING.”

“I’m not sure I understand.”

“YOUR LIFE CONTAINS MUCH PAIN, BUT MORE HAPPINESS. BOTH YOU AND I WOULD PREFER THAT A BEING WITH YOUR EXACT LIFE HISTORY EXIST. IN ORDER TO MAKE IT EXIST, IT WAS NECESSARY TO CREATE THE SORT OF UNIVERSE IN WHICH YOU COULD EXIST. THAT IS A UNIVERSE CONTAINING EVIL. I HAVE ALSO CREATED ALL HAPPIER AND MORE VIRTUOUS VERSIONS OF YOU. HOWEVER, IT IS ETHICALLY CORRECT THAT AFTER CREATING THEM, I CREATE YOU AS WELL.”

“But why couldn’t I have been one of those other versions instead!”

“IN THE MOST PERFECTLY HAPPY AND JUST UNIVERSE, THERE IS NO SPACE, FOR SPACE TAKES THE FORM OF SEPARATION FROM THINGS YOU DESIRE. THERE IS NO TIME, FOR TIME MEANS CHANGE AND DECAY, YET THERE MUST BE NO CHANGE FROM ITS MAXIMALLY BLISSFUL STATE. THE BEINGS WHO INHABIT THIS UNIVERSE ARE WITHOUT BODIES, AND DO NOT HUNGER OR THIRST OR LABOR OR LUST. THEY SIT UPON LOTUS THRONES AND CONTEMPLATE THE PERFECTION OF ALL THINGS. IF I WERE TO UNCREATE ALL WORLDS SAVE THAT ONE, WOULD IT MEAN MAKING YOU HAPPIER? OR WOULD IT MEAN KILLING YOU, WHILE FAR AWAY IN A DIFFERENT UNIVERSE INCORPOREAL BEINGS SAT ON THEIR LOTUS THRONES REGARDLESS?”

“I don’t know! Is one of the beings in that universe in some sense *me*?”

“THERE IS NO OBJECTIVE COSMIC UNEMPLOYMENT RATE.”

“Huh?”

“I MEAN, THERE IS NO MEANINGFUL ANSWER TO THE QUESTION OF HOW MANY UNIVERSES HAVE A JOB. SORRY. THAT WILL BE FUNNY IN ABOUT THREE THOUSAND YEARS.”

“Let me try a different angle, then. Right now in our universe there are lots of people whose lives aren’t worth living. If You gave them the choice, they would have chosen never to have been born at all. What about them?”

“A JOB WHO IS AWARE OF THE EXISTENCE OF SUCH PEOPLE IS A DIFFERENT JOB THAN A JOB WHO IS NOT. AS LONG AS THESE PEOPLE MAKE UP A MINORITY OF THE POPULATION, THE EXISTENCE OF YOUR UNIVERSE, IN ADDITION TO A UNIVERSE WITHOUT SUCH PEOPLE, IS A NET ASSET.”

“But that’s monstrous! Couldn’t You just, I don’t know, have created a universe that looks like it has such people, but actually they’re just p-zombies, animated bodies without any real consciousness or suffering?”

” . . . ”

“Wait, *did* You do that?”

“I AM GOING TO PULL THE ‘THINGS MAN WAS NOT MEANT TO KNOW’ CARD HERE. THERE ARE ADVANTAGES AND DISADVANTAGES TO THE APPROACH YOU MENTION. THE ADVANTAGES ARE AS YOU HAVE SAID. THE DISADVANTAGE IS THAT IT TURNS CHARITY TOWARDS SUCH PEOPLE INTO A LIE, AND MYSELF AS GOD INTO A DECEIEVER. I WILL ALLOW

YOU TO FORM YOUR OWN OPINION ABOUT WHICH COURSE IS MORE ETHICAL. BUT IT IS NOT RELEVANT TO THEODICY, SINCE WHICHEVER COURSE YOU DECIDE IS MORALLY SUPERIOR, YOU HAVE NO EVIDENCE THAT I DID NOT IN FACT TAKE SUCH A COURSE.”

“Actually, I do have some evidence. Before all of this happened to me I was very happy. But in the past couple years I’ve gone bankrupt, lost my entire family, and gotten a bad case of boils. I’m pretty sure at this point I would prefer that I never have been born. Since I know I myself am conscious, I am actually in a pretty good position to accuse You of cruelty.”

“HMMMMMMMM…” said God, and the whirlwind disappeared.

Then the Lord gave Job twice as much as he had before, and healed his illnesses, and gave him many beautiful children, so it was said that God had blessed the latter end of Job more than his beginning.

*[**EDIT:** According to comments, this was scooped by a Christian philosopher five years ago. Sigh.]*

> The Lord spoke to Job out of the whirlwind, saying "MISTAKES WERE MADE."
>
> — Scott Alexander (@slatestarcodex) March 13, 2015

> Then the Lord spoke to Job out of the whirlwind, saying "IF YOU CAN'T HANDLE ME AT MY WORST, YOU DON'T DESERVE ME AT MY BEST."
>
> — Scott Alexander (@slatestarcodex) March 10, 2015

The Lord spoke to Job out of the whirlwind, saying "I KNOW YOU'RE UPSET BUT THAT'S DIFFERENT FROM STRUCTURAL OPPRESSION" (h/t @simulacrumbs)

— Scott Alexander (@slatestarcodex) March 13, 2015

List Of Passages I Highlighted In My Copy Of “Willpower”

[content note: dieting]

Warning: I have not checked any of these claims for truth, I was generally not impressed with the skepticism level displayed in this book, and I highlighted the passages I found most surprising or counterintuitive. So these are quotations, not endorsements. As Ashleigh Brilliant says, “My sources are unreliable, but their information is fascinating.”

> A good way to appreciate the Zeigarnik effect is to listen to a randomly chosen song and shut it off halfway through. The song is then likely to run through your mind on its own, at odd intervals. If you get to the end of the song, the mind checks it off, so to speak. If you stop it in the middle, however, the mind treats the song as unfinished business.

> [Examination-related words] popped more frequently into the mind of the group who had been told about the exam but hadn’t made plans to study for it. No such effect was seen among the students who’d made a study plan. Even though they, too, had been reminded of the exam, their minds had apparently been cleared by the act of writing down a plan.

> In 1995, Tierney did a semiscientific survey of a New York phenomenon: the huge number of intelligent and attractive people who complained that it was impossible to find a romantic partner. Manhattan had the highest percentage of single people in any country in America except for an insland in Hawaii originally settled as a

leper colony. What was keeping New Yorkers apart? Tierney surveyed a sampling of personal ads in the city magazines of Boston, Baltimore, Chicago, Los Angeles, and New York. He found that singles in the biggest city, New York, not only had the most choice but were the pickiest in listing the attributes of their desired partners. The average personal ad in New York maganize listed 5.7 criteria required in a partner, significantly more than second place Chicago’s average (4.1 criteria) and about twice the average for the other three cities.

Bob Boice looked into the writing habits of young professors and tracked them to see how they fared. Some of these professors would collect information until they were ready and then write a manuscript in a burst of intsense energy. Others plodded along at a steadier pace, trying to write a page or two every day. When Boice followed up on the group some years later, he foud that their paths had diverged sharply. The page-a-day folks had done well and generally gotten tenure. The so-called “binge writers” had fared far less well, and many of them had had their careers cut short.

As early as the 1920s, researchers reported that students who spent more time in Sunday school scored higher on laboratory tests of self-discipline. Religiously devout children were rated low on impulsiveness by both parents and teachers…But psychologists have found that people who attend religious services for extrinsic reasons, like wanting to impress others or make social connections, don’t have the same high level of self-control as the true believers.

A clear difference between Chinese and American toddlers appears when they’re asked to override their natural impulses. In one test, for instance, the toddlers are shown a series of pictures and instructed to say “day” whenever they see the moon and “night” whenever they see the sun. In other tests, the toddlers try to restrain themselves to a whisper when they’re excited, and play a version of Simon Says. The Chinese four-year-olds generally perform better on these tests than Americans of the same age. The Chinese toddlers’ self-control might be due in part to genes. There’s evidence that the genetic factors associated with ADHD are much rare in Chinese children than in American children. But the cultural traditions undoubtedly play a role as well. Asian-Americans make up only 4% of the US population but account for a quarter of the student body of elite universities like Stanford, Cornell, and Columbia. They’re more likely to get a college degree than any other ethnic group, and they go on to earn salaries that are 25% above the American norm. Their success has led to the popular notion that Asians are more intelligent, but that’s not how James Flynn explains their achievements. After carefully reviewing IQ studies, Flynn concludes that the scores of Chinese-Americans and Japanese-Americans are very similar to Americans of European descent. If anything their IQs are slightly lower, though they do show up more at both the upper and lower extremes. The big difference is they make better use of their intelligence. People working in elite professions like physicians, scientists, and accountants generally have an IQ above a certain threshold. For white Americans, that threshold is IQ 110, but Chinese Americans manage to get the same elite jobs with an IQ of only 103.

When asked how parents could contribute to childrens’ academic success, the mothers who had emigrated from China most frequently mentioned setting high goals, enforcing tough standards, and requiring children to do extra homework. Meanwhile, the native-born mothers of European ancestry were determined *not* to put too much pressure on children They most frequently mentioned the importance of not overemphasizing academic success, of stressing the child's social development, and of promoting the idea that "learning is fun" and "not something you work at". Another of their chief concerns was promoting the child's self-esteem, a concept of just about no interest to the Chinese mothers in the study.

[In his preliminary marshmallow-test-style experiments in Trinidad], Mischel stumbled upon a bigger and more meaningful effect. Children who had a father in the home were far more willing than others to choose the delayed reward. Most of the racial and ethnic variation could be explained by this difference, because the Indian children tended to live with both parents whereas a fair number of the African children lived with a single mother. These findings, which were published in 1958, didn't attract much attention at the time or in ensuing decades, since it was dangerous to one's career to suggest that there might be drawbacks to single-parent homes…One possible explanation is that children in one-parent homes start off with a genetic disadvantage in self-control. After all, if the father (or mother) has run off and abandoned the family, he may have genes favoring impulsive behavior. Some researchers have attempted to correct this by looking at children who were raised by single parents because the father was absent for other reasons (like being stationed overseas, or dying at a young age). Predictably, the results were in between. These children showed some deficits, but their problemes were not as large as those of the children whose fathers had voluntarily left the home."

Nurture Assumption said the opposite (children whose fathers left for valid reasons were totally indistinguishable from children with two parents), I believed them until I coincidentally found the contrary evidence, and I'm still angry at them for this.

> When fat lab rats are put on a controlled diet for the first time, they'll lose weight. But if they're then allowed to eat freely again, they'll gradually fatten up, and if they're put on another diet it will take longer to lose the weight this time. Then, once they again go off the diet, they'll regain the weight more quickly than the last time. By the third or fourth time they go through this boom-and-bust cycle, the dieting ceases to work – the extra weight stays even though they're consuming fewer calories.

If true this would require a pretty high-level uncoupling of calories and weight gain.

> An English bookmaker, the William Hill agency, has a standing offer to bet against anyone who makes a plan to lose weight. The bookmaker, which offers odds of up to 50 to 1, lets the bettors set their own target of how much weight to lose in how much time. It seems crazy for a bookmaker to let bettors not only set the terms of the wager but also control its outcome – it's like letting a runner bet on beating a target time he sets himself. Yet despite these advantages, the bettors lose 80% of the time."

I tried to find these people's website to see how the numbers worked, but it wouldn't let me in because I'm in the US. Brits might want to check this out.

> Dieters have a fixed target in mind, and when they exceed it for any reason – [like being told to taste test a milkshake as part of an experiment] they regard their diet as blown for the day. So they think *what the hell, I might as well enjoy myself for the day* and the resulting binge often puts on far more weight than the original lapse.

> People who weigh themselves every day are far more effective at preventing their weight from creeping back up.

> Some of the [experimental subjects] were shown a meal from Applebees consisting of chicken salad and a Pepsi; others were shown an identical meal with some added crackers prominently labeled “Trans Fat Free”. The people were so entranced by the crackers’ virtuous label that their estimate for the meal with crackers was *lower than* that for the same meal without the crackers!

A new form of the Conjunction Fallacy?

> To mimic the human studies [on ego depletion], experimenters depleted the willpower of one group of dogs by having each dog obey “sit” and “stay” commands from its owner for ten minutes. A control group of dogs was simply left alone for ten minutes in cages. Then all the dogs were given a familiar toy with a sausage treat inside of it. All the dogs had played with the toy in the past and successfully extracted the treat, but for this study the toy was rigged so that the sausage could not be extracted. The control group of dogs spent several minutes trying to extract it, but the dogs who’d had to obey the commands gave up in less than a minute. It was the familiar ego-depletion effect, and the canine cure turned out to be familiar too. In a follow-up

> study, when the dogs were given different drinks, the drinks with sugar restored the willpower of the dogs who'd had to obey the commands. Newly fortified, they persisted with the toy just as long as the dogs who'd been in cages.
>
> People with poor self-control were likelier to hit their partners and to commit a variety of other crimes, again and again, as demonstrated by June Tangney, who worked with Baumeister to develop the self-control scale on personality tests. When she tested prisoners and then tracked them for years after their release, she found that the ones with low self-control were most likely to commit more crimes and return to prison.
>
> In one remarkable study, researchers in Finland went into a prison to measure the glucose tolerance of convicts who were about to be released. Then the scientists kept track of which ones went on to commit new crimes. Just by looking at the response to the glucose test, the researchers were able to predict with greater than 80% accuracy which convicts would go on to commit violent crimes. These men apparently had less self-control because of their impaired glucose tolerance, a condition in which the body has trouble converting food into usable energy."

SO SOOOOO SKEPTICAL.

> When people in laboratory experiments exercise mental self-control, their pulse becomes more erratic; conversely, people whose normal pulse is relatively variable seem to have more inner energy available for self-control, because they do better on laboratory tests of perseverance than people with steadier heartbeats.

Also pretty skeptical here, given past experience.

> A psychoanalyst named Allen Wheelis in the late 1950s revealed what he considered a dirty little secret of his profession: Freudian therapies no longer worked the way they were supposed to. In his landmark book, *The Quest For Identity*, Wheelis described a change in character structure since Freud's day. The Victorian middle-class citizens who formed the bulk of Freud's patients had intensely strong wills, making it difficult for therapists to break through their ironclad defenses and their sense of what was right and wrong. Freud's therapies had concenctrated on wayhs to break through and let them see why they were neurotic and miserable, because once those people achieved ionsight, they could change rather easily. By midcentury, thought, people's character armor was different. Wheelis and his colleagues found that people achieved insight more qui9ckly than in Freud's day, but then the therapy often stalled and failed. Lacking the sturdy character of the Victorians, people didn't have the strength to follow up on the insight and change their lives

This is actually sorta plausible to me. My (very limited) experience with psychoanalysis is that it's not nearly as hard as people claim to get patients to tell you things about themselves and produce apparent "revelations", but these rarely change behavior in interesting ways. I'd never thought before that this might be a historical change as opposed to just Freud getting it wrong.

> [Subjects] wore beepers that went off at random intervals seven times a day, prompting them to report whether they were currently experiencing some sort of desire…the researchers concluded that people spend at least a fifth of their waking hours resisting desires –

> between three and four hours a day…Overall, they succumbed to about a sixth of the temptations.

Someone responded by my last post by pointing out a Christian philosopher saying the same thing and noting that Science always thinks it’s gotten ahead of Religion only to find that Religion’s known it all along. Nevertheless, I will venture to say no religious source on temptation includes the observation that on average people succumb to about one-sixth of them. That one *totally* goes to Science.

Book Review: The Machinery Of Freedom

[conflict of interest: David Friedman is an amazing person who has been very nice to me and among other things hosted the San Jose SSC meetup earlier this month]

David Friedman's The Machinery of Freedom is half Libertarianism 101: Introduction To Libertarianism, and half Libertarianism 501: Technical Diagrams For Constructing An Anarcho-Capitalist State.

And aside from either of these, it's interesting as a historical artifact. The first edition was published in 1973; the Third Edition copy I read is from last year, but the updates are minor and the book keeps its 1973 feel – including a discussion of health care economics which puts the price of a doctor's visit at $10.

One of my takeaways was how *new* libertarianism was in 1973. The introduction says:

> These peculiar views of mine are not peculiar to me. If they were, I would be paying Harper and Row to publish this book instead of Harper and Row paying me. My views are typical of the ideas of a small but growing group of people, a 'movement' that has begun to attract the attention of the national media. We call ourselves libertarians.
>
> This book is concerned with libertarian ideas, not with a history of the libertarian movement or a description of its present condition. It is fashionable to measure the importance of ideas by the number and violence of their adherents. That is a fashion I shall not follow. If, when you finish this book, you have come to share many of

> my views, you will know the most important thing about the number of libertarians – that it is larger by one than when you started reading.

There is something very innocent about expecting someone to become a libertarian after reading a book arguing for libertarianism, something very much a product of the time when the movement was new and anything was possible. Friedman discusses and debates the views of Ayn Rand not as some sort of ascended cultural archetype, but as a fellow theorist who happens to be writing around the same time. It makes the book somehow fresher than one that starts from the perspective of "Okay, you've heard all of these arguments before, so let me preach to the choir and see what happens."

But sometimes the book is dated in ways less innocuous than ten-dollar doctor visits. For example, in Chapter 5, "The Rich Get Richer And The Poor Get Richer," Friedman argues against excessive concern with inequality, saying:

> In absolute terms, the rich have gotten richer, but the gap between rich and poor seems, so far as very imperfect statistics make it possible to judge, to have been slowly closing…we can note that both the rise in the general standard of living and the decreasing inequality appear to have been occurring fairly steadily over a long period of time, in a variety of different more or less capitalist societies…in the previous chapter I argued that liberal measures tend to injure the poor, not benefit them, and to increase, not decrease inequality. If that has been true in the past, then the increasing equality we have experienced is in spite of, not because of, such measures…

> Even if the capitalist invests all the income from his capital and consumes none of it, his wealth will only grow at the rate of return on capital. If the interest rate is less than the rate at which the total wages of workers increase, the relative wealth of the capitalists will decline. Historically, the rate of increase in total wages has run about 5 to 10 percent a year, roughly comparable to the interest rate earned by capital. Furthermore, capitalists consume part of their income; if they did not, there would be little point in being a capitalist. The share of the national income going to capital in this country has varied over time but not consistently increased, as shown in Appendix III.

The heartbreaking thing is that every word of this was true in 1973. In fact, 1973 is frequently given as the inflection point, when for some reason middle-class wages stopped rising at the same rate as the wealth of the top 1% and capital's share of income started a steady climb (this is frequently blamed on Reagan, but started almost a decade before his presidency).

There are enough issues like this that they make the book's arguments less compelling, or at least cry out to be addressed. Likewise, the book's statistics are fascinating and in many cases very counterintuitive and convincing, but I have a lot of trouble double-checking them because they're mostly 1970s statistics.

I can't do justice to the Libertarian 101 arguments in this review because there are too many of them on too many different topics. This is too bad because they are *excellent* and *fascinating* and you should really read them. Aside from recommending you get the book, I'll shove those into a separate Highlights post later this week. But for now I want to focus on the claim that I found most interesting:

Government claims legitimacy partly from its role in helping the poor, but the costs fall disproportionately on the poor and it screws them over more than any other group:

> Suppose that one hundred years ago someone tried to persuade me that democratic institutions could be used to transfer money from the bulk of the population to the poor. I could have made the following reply: "The poor, whom you wish to help, are many times outnumbered by the rest of the population, from whom you intend to take the money to help them. If the non-poor are not generous enough to give money to the poor voluntarily through private charity, what makes you think they will be such fools as to vote to force themselves to take it?"

I think I have a good answer to this question. Nobody's vote makes very much difference, so people are happy to vote for signaling/psychological reasons rather than financial ones. If casting my vote to help the poor makes me feel like a good person, but losing money in redistribution schemes makes me poorer, well, my vote 100% determines whether I feel good or not, but only 1/300-million determines whether I get poorer. This might also be profitably mapped onto construal level theory, ie Robin Hanson's Near Mode vs. Far Mode.

Anyway, having determined that democracy should not be expected to help the poor, he gets on to demonstrating that in fact it doesn't:

> There are some programs that give money to the poor – Aid to Families With Dependent Children, for instance. But such programs are vastly outweighed by those having the opposite effect – programs that injure the poor for the benefit of the not-poor. Almost surely, the poor would be better off if both the benefits that they

> now receive and the taxes, direct and indirect, that they now pay were abolished.

He then goes on to list examples, including Social Security, food subsidies (which increase food prices and go to rich farmers), state universities (since they cost tax money and mostly rich people go to university), and urban renewal projects (which bulldoze low-quality housing that the poor can afford to create high-quality housing that they can't, thus pushing up their housing prices).

I don't know much about the 1973 situation, but a lot of these don't seem very convincing nowadays. Social Security no longer appears regressive: as per Wikipedia, "for people in the bottom fifth of the earnings distribution, the ratio of [Social Security] benefits to taxes is almost three times as high as it is for those in the top fifth." And by my understanding, people who earn less than about $20,000 don't pay federal income taxes at all, meaning the burden of universities, etc don't fall upon them. A Cato Institute study finds that poor people on welfare can get benefits packages worth up to about $20,000. It seems really unlikely that whatever they have to pay because of farm subsidies or whatever compensates for that.

But Friedman also makes the stronger point that when government programs fail, it's the poor who are most affected and who have the fewest other options. For example, he notes that the cost per capita of law enforcement/police/courts is $40 (remember, this is 1973!) and estimates that minus government waste and corruption, the free market could provide extremely competent policing for $20. He says:

> There are many inhabitants of the ghetto who would be delighted to pay twenty dollars a year if in exchange

> they actually got protection; many of them have more than that stolen every year as a result of the poor protection they get from our government-run protection system. They would be even happier if at the same time they were relieved of the taxes that pay for the protection that the government police does not give them. In spite of popular myths about capitalism oppressing the poor, the poor are worst off in those things provided by government, such as schooling, police protection, and justice. There are more good cars in the ghetto than good schools.

I somewhat agree with the spirit of this quote, but certainly some of the problem is that poor people live in poor areas that collect little tax revenue and underfund their social services. Bigger government could solve this problem – just have school district funding set at the state or federal level. It’s less obvious that smaller government could – poor people would still have X dollars to spend on schools, for low values of X. But here we get into complicated proposals like vouchers and private policing that I’ll leave for later.

II.

Let’s get to what we’re really here for – the crazy anarcho-capitalist utopia.

This quote is very long, but it’s worth it:

> How, without government, could we settle the disputes that are now settled in courts of law? How could we protect ourselves from criminals?
>
> Consider first the easiest case, the resolution of disputes involving contracts between well-established firms. A large fraction of such disputes are now settled not by government courts but by private arbitration of the sort

described in Chapter 18. The firms, when they draw up a contract, specify a procedure for arbitrating any dispute that may arise. Thus they avoid the expense and delay of the courts.

The arbitrator has no police force. His function is to render decisions, not to enforce them. Currently, arbitrated decisions are usually enforceable in the government courts, but that is a recent development; historically, enforcement came from a firm’s desire to maintain its reputation. After refusing to accept an arbitrator’s judgment, it is hard to persuade anyone else to sign a contract that specifies arbitration; no one wants to play a game of ‘heads you win, tails I lose’.

Arbitration arrangements are already widespread. As the courts continue to deteriorate, arbitration will continue to grow. But it only provides for the resolution of disputes over pre-existing contracts. Arbitration, by itself, provides no solution for the man whose car is dented by a careless driver, still less for the victim of theft; in both cases the plaintiff and defendant, having different interests and no prior agreement, are unlikely to find a mutually satisfactory arbitrator. Indeed, the defendant has no reason to accept any arbitration at all; he can only lose–which brings us to the problem of preventing coercion.

Protection from coercion is an economic good. It is presently sold in a variety of forms–Brinks guards, locks, burglar alarms. As the effectiveness of government police declines, these market substitutes for the police, like market substitutes for the courts, become more popular.

Suppose, then, that at some future time there are no government police, but instead private protection agencies. These agencies sell the service of protecting their clients against crime. Perhaps they also guarantee performance by insuring their clients against losses resulting from criminal acts.

How might such protection agencies protect? That would be an economic decision, depending on the’-costs and effectiveness of different alternatives. On the one extreme, they might limit themselves to passive defenses, installing elaborate locks and alarms. Or they might take no preventive action at all, but make great efforts to hunt down criminals guilty of crimes against their clients. They might maintain foot patrols or squad cars, like our present government police, or they might rely on electronic substitutes. In any case, they would be selling a service to their customers and would have a strong incentive to provide as high a quality of service as possible, at the lowest possible cost. It is reasonable to suppose that the quality of service would be higher and the cost lower than with the present governmental system.

Inevitably, conflicts would arise between one protective agency and another. How might they be resolved?

I come home one night and find my television set missing. I immediately call my protection agency, Tannahelp Inc., to report the theft. They send an agent. He checks the automatic camera which Tannahelp, as part of their service, installed in my living room and discovers a picture of one Joe Bock lugging the television set out the door. The Tannahelp agent contacts Joe, informs him that Tannahelp has reason to believe he is in possession of my television set, and suggests he return it, along with an extra ten dollars to pay for Tannahelp’s time and trouble in locating Joe. Joe replies that he has never seen my television set in his life and tells the Tannahelp agent to go to hell.

The agent points out that until Tannahelp is convinced there has been a mistake, he must proceed on the assumption that the television set is my property. Six Tannahelp employees, all large and energetic, will be at Joe’s door next morning to collect the set. Joe, in response, informs the agent that he also has a protection agency, Dawn Defense, and that his contract with them undoubtedly requires them to protect him if six goons try to break into his house and steal his television set.

The stage seems set for a nice little war between Tannahelp and Dawn Defense. It is precisely such a possibility that has led some libertarians who are not anarchists, most notably Ayn Rand, to reject the possibility of competing free-market protection agencies.

But wars are very expensive, and Tannahelp and Dawn Defense are both profit-making corporations, more interested in saving money than face. I think the rest of the story would be less violent than Miss Rand supposed.

The Tannahelp agent calls up his opposite number at Dawn Defense. ‘We’ve got a problem. . . .’ After explaining the situation, he points out that if Tannahelp sends six men and Dawn eight, there will be a fight. Someone might even get hurt. Whoever wins, by the time the conflict is over it will be expensive for both sides. They might even have to start paying their employees higher wages to make up for the risk. Then both firms will be forced to raise their rates. If they do, Murbard Ltd., an aggressive new firm which has been trying to get established in the area, will undercut their prices and steal their customers. There must be a better solution.

The man from Tannahelp suggests that the better solution is arbitration. They will take the dispute over my television set to a reputable local arbitration firm. If the arbitrator decides that Joe is innocent, Tannahelp agrees to pay Joe and Dawn Defense an indemnity to make up for their time and trouble. If he is found guilty, Dawn Defense will accept the verdict; since the television set is not Joe's, they have no obligation to protect him when the men from Tannahelp come to seize it.

What I have described is a very makeshift arrangement. In practice, once anarcho-capitalist institutions were well established, protection agencies would anticipate such difficulties and arrange contracts in advance, before specific conflicts occurred, specifying the arbitrator who would settle them.

In such an anarchist society, who would make the laws? On what basis would the private arbitrator decide what acts were criminal and what their punishments should be? The answer is that systems of law would be produced for profit on the open market, just as books and bras are produced today. There could be competition among different brands of law, just as there is competition among different brands of cars.

In such a society there might be many courts and even many legal systems. Each pair of protection agencies agree in advance on which court they will use in case of conflict. Thus the laws under which a particular case is decided are determined implicitly by advance agreement between the protection agencies whose customers are involved. In principle, there could be a different court and a different set of laws for every pair of protection agencies. In practice, many agencies would probably find it convenient to patronize the same courts, and many courts might find it convenient to adopt identical, or nearly identical, systems of law in order to simplify matters for their customers.

Before labelling a society in which different people are under different laws chaotic and unjust, remember that in our society the law under which you are judged depends on the country, state, and even city in which you happen to be. Under the arrangements I am describing, it depends instead on your protective agency and the agency of the person you accuse of a crime or who accuses you of a crime.

In such a society law is produced on the market. A court supports itself by charging for the service of arbitrating disputes. Its success depends on its reputation for honesty, reliability, and promptness and on the desirability to potential customers of the particular set of laws it judges by. The immediate customers are protection agencies. But the protection agency is itself selling a product to its customers. Part of that product is the legal system, or systems, of the courts it patronizes and under which its customers will consequently be judged. Each protection agency will try to patronize those courts under whose legal system its customers would like to live.

The idea is that these protection agencies are companies like any other, and so will try to provide a good product at a low cost that satisfies their customers. People can choose their favorite, and so in some sense decide which laws to be bound by. Although they will not have complete flexibility in choosing their laws, lawmaking bodies will be sort of subject to consumer demand.

He correctly points out that contrary to what you might expect this system does *not* by definition exclude victimless crimes. If you want to hire a police agency that things being gay is a crime, you can pay them money to go find gay people and throw them out of town. Then the gay people will hire their own police agency to defend themselves. I think Friedman believes that opposing homosexuality has a major free rider problem, and that most people like to signal virtue by complaining about them but very few people would be willing to pay money for it. By comparison, gay people would be willing to pay a lot of money to be protected from this sort of thing, so their protection agencies would be stronger than the agencies of whoever wants to kick them out, and they'd stay.

This seems to me overly optimistic. After all, back when only a tiny percent of the country was tolerant of homosexuality, it might be that church groups could raise a lot of money to enforce anti-gay laws, and gay people were mostly poor and couldn't raise very much money to defend themselves. I think I know what Friedman's response would be, which is "Yes, and during that time in your real-world statist society, homosexuality was also illegal. Yes, you would have to wait for cultural norms to change before homosexuality would be legalized, but it would very likely be easier to do my way than yours." I think he's possibly right.

My overall conclusion is that I am delighted by this fascinating and elegant system and would very much like to see it tried *somewhere very far away from me*.

I am sure Friedman has to listen to so many objections that he can recite most of them by memory and is sick to death of them. Indeed, he admits this and devotes no small amount of space to rebutting many of them. Will we get taken over by one giant protection racket? Probably not, monopolies are rare in practice. Will criminals get their own protection and arbitration agencies that say crime is okay? Probably not; no other protection agency would agree to arbitrate on their terms, and without arbitration they would be in a war with all the other agencies, which the other agencies would win since legitimate business can mobilize more money than crime can. Would there be constant bloody battles? Probably not; profit-seeking corporations would be too smart to lose money that way when better options like arbitration are available. Would the heads of protection agencies form a pact, then use their combined might to take over the country and become kings? Probably not; right now police chiefs and military generals don’t do this, even though they are in a good position to.

Here are some objections of mine I didn’t see rebutted:

1. People who don’t purchase protection are pretty much fair game for anyone to rob or murder or torture or whatever. This seems harsh, especially since this society is likely to have a sizable underclass. I don’t know if “$20 for a year of police protection” was a reasonable estimate for the 70s, but I expect this would be much costlier now. Compare the percent of people who, pre-Obamacare, still didn’t have health insurance, and how much higher it would have been if there weren’t government programs that kind of got health insurance bundled in with employment.

2. Protection agencies are going to be engaged in constant brinksmanship for the same reason nation-states are engaged in constant brinksmanship. If Agency 1 wanted concessions from Agency 2, it has an incentive to seem kind of crazy and like it might actually declare real war, however unprofitable, in order to bluff Agency 2 into complying. Remember, countries have the same economic incentives to avoid war that companies do, but they still occasionally get involved in them. Even when they don't, the threat of such leads many resources to be wasted in military buildup.

3. ~~Security companies and their clients are very unlikely to want to pay for the cost of incarcerations. There's no incentive to pay extra for criminal rights, so convicted criminals are likely to end up facing something like corporal punishment~~ Never mind, this went an unexpected direction and is probably a good thing.

4. If I am the church-funded protection agency charged with flogging gay people, and you are the gay-person funded protection agency charged with protecting them, it's hard to see what kind of arbitration we would agree on. I…uh…guess this might be another one that isn't so bad, since that might mean the agencies are forced to actually fight, which raises the cost of being anti-gay to a potentially prohibitive level.

5. There are some things which might decrease crime in an area in general instead of just involving crime against a specific person. For example, adding streetlights, fighting drug abuse, putting troubled youth in after-school programs, fighting the broken window effect. If these are public goods, nobody will be incentivized to pay extra for them.

6. In fact, protection agencies have a strong incentive to make everybody as scared of crime as possible, and in fact

to raise the actual crime rate if they can, in order to get people to buy their Premium plan. Given that this is anarcho-capitalism and there are no laws against crime, this can’t possibly end well.

7. It would be hard to have large-scale public laws. Right now Saudi Arabia can have laws about how no woman can go outside unveiled, America can have laws that nobody can go outside unclothed, and some European beaches can have laws saying go ahead and be naked. Likewise, some small villages can have zoning laws saying not to build non-scenic skyscrapers, but Dubai can say to build as high as you want and then some. This seems harder under anarcho-capitalism until people start coordinating the formation of intentional communities, at which point it becomes less anarcho-capitalism and more Patchwork.

8. Gang leaders and barbarian warlords had the chance to become protection agencies like this, but never did. This suggests that this system is unstable or unnatural. It’s possible that once the equilibrium of protection and arbitration agencies is established it will be stable, but of all of the various lawless societies to exist throughout history, none of them coalesced upon this system. Suspicious.

9. An extension of this: it’s unclear that we’re not already living in this society. It’s just that one protection and arbitration agency has completely taken over from all of the others and instituted a policy of using force against those who don’t pay for its services. That’s allowed under anarcho-capitalism because everything is allowed under anarcho-capitalism. So expecting anarcho-capitalism to be stable is expecting the thing that has already happened to not happen again a second time.

10. There seems to be a lot of opportunities for rich people to purchase greater privileges not available to the masses. After all, negotiation results are often determined by a party's BATNA. Rich people may have access to very strong security companies (or premium plans from regular companies) that could win most fights; they can use this to insist on better arbitration terms. A rich person's company might only accept basic arbitration (eg punish the rich person for murder) if other companies agree to lopsided deals (like don't go after the rich person for less dramatic things like sexual harassment. On the other hand, a poorer person's company might have to accept the worse side of the deal, where the poor person can be prosecuted for a very wide range of crimes against the rich person, including giving offense and not being respectful enough. Yes, it's easy to see how a company could arise that charges extra in exchange for not accepting these compromises, but this still suggests you're going to have more rights if you're able to pay more money.

But the main reason I want this tried *far away from me* is none of these. It's just a general expectation that something will go wrong when we try a social system we've never tried before. I was very impressed to learn that very few people predicted, before the fact, that Communist countries would have terrible economies. Even the American 1950s opponents of Communism argued that okay, fine, Communist countries will probably outperform capitalist countries economically, but freedom is more important than mere wealth.

If people can't figure out that Communism might sink the economy, I don't trust them to figure out all of the things that might go wrong with anarcho-capitalism. Even if David Friedman replies with utterly convincing rebuttals to all of my ten points above, it’s going to be the eleventh point I didn’t think of that makes the system explode.

III.

And this leads me into one of my deepest problems with libertarianism and anarcho-capitalism: why should it work?

I don’t mean the sort of “why should it work” where you answer with specific reasons why no, monopolies won’t form, and no, people won’t routinely sell themselves into slavery, and no, protection agencies won’t form a new feudal ruling class, and no, people won’t bash their heads against public goods problems and externalities forever without any market solutions appearing, and no, the poor won’t starve to death. I mean the very Outside View question of “why is it that, by coincidence, not using force is an effective way to solve all problems?”

Good governance is a really really hard problem. The idea that the solution to this problem contains zero bits of information, that it just solves itself if you leave people alone, seems astonishing. Even if we agree that capitalism works very well by incentivizing companies to do what the consumers wants, there are still a lot of peripheral issues which that just doesn’t cover. Friedman for example is a strong supporter of child rights, because children should mostly be free from coercion from their parents, and that children treated this way turn out better. Now in addition to solving governance with zero bits of information, you have solved optimal child-rearing with zero bits of information. That is implausibly impressive.

Given that the universe is allowed to throw whatever problems it wants at us, and that it has so far gleefully taken advantage of that right to come up with a whole host of very diverse and interesting ones, why is it that none of these problems are best addressed by a centralized entity with a monopoly on force? That seems like a pretty basic structure from a game-theoretic perspective, and you're telling me it just never works in the real world? Shouldn't there be at least one or two things where a government, or any form of coercive structure at all, is just the right answer? And can't we just have a small government that does *that*?

The closest thing I've found to a response here is on page 142, where Friedman makes the following very witty observation:

> The internal dynamic of limited government is something with which we, to our sorrow, have a good deal of practical experience. It took about 150 years, starting with a Bill of Rights that reserved to the states and the people all powers not explicitly delegated to the federal government, to produce a Supreme Court willing to rule that growing corn to feed your own hogs is interstate commerce and can therefore be regulated by Congress.

So if we have any kind of government at all, it will eventually metastasize into the sort of thing that makes laws about whether we're allowed to grow corn to feed our own animals, or bans us from drinking raw milk, or whatever else it feels like doing.

So which is better: moving to full anarcho-capitalism, or trying to move towards a system that can provide more of the benefits of government with fewer of the costs?

I don't know, so it's a good thing I don't have to choose. The obvious next step seems to be setting up anarcho-capitalist experiments somewhere and seeing how they do, as well as continuing to experiment with new and better forms of government. Trying to predict anything from theory runs into the same problem where everyone assumed Communism would be an economic powerhouse – we’re just not that smart. Instead we need to figure out ways to produce experimentation with and competition among different governments and government-like-entities – a goal I know David Friedman agrees with.

List Of Passages I Highlighted In My Copy Of “Machinery Of Freedom”

> Under any institutions, there are essentially only three ways that I can get another person to help me achieve my ends: love, trade, and force.
>
> By love I mean making my end your end. Those who love me wish me to get what I want (except for those who think I am very stupid about what is good for me). So they voluntarily, ‘unselfishly’, help me. Love is too narrow a word. You might also share my end not because it is my end but because in a particular respect we perceive the good in the same way. You might volunteer to work on my political campaign, not because you love me, but because you think that it would be good if I were elected. Of course, we might share the common ends for entirely different reasons. I might think I was just what the country needed, and you, that I was just what the country deserved.
>
> Love—more generally, the sharing of a common end—works well, but only for a limited range of problems. It is difficult to know very many people well enough to love them. Love can provide cooperation on complicated things among very small groups of people, such as families. It also works among large numbers of people for very simple ends—ends so simple that many different people can completely agree on them. But for a complicated end involving a large number of people—producing this book, for instance—love will not work. I cannot expect all the people whose cooperation I need—typesetters, editors, bookstore owners, loggers, pulpmill workers, and a thousand more—to know and love me

well enough to want to publish this book for my sake. Nor can I expect them all to agree with my political views closely enough to view the publication of this book as an end in itself. Nor can I expect them all to be people who want to read the book and who therefore are willing to help produce it. I fall back on the second method: trade.

I contribute the time and effort to produce the manuscript. I get, in exchange, a chance to spread my views, a satisfying boost to my ego, and a little money. The people who want to read the book get the book. In exchange, they give money. The publishing firm and its employees, the editors, give the time, effort, and skill necessary to coordinate the rest of us; they get money and reputation. Loggers, printers, and the like give their effort and skill and get money in return. Thousands of people, perhaps millions, cooperate in a single task, each seeking his own ends. So under private property the first method, love, is used where it is workable. Where it is not, trade is used instead.

The attack on private property as selfish contrasts the second method with the first. It implies that the alternative to ‘selfish’ trade is ‘unselfish’ love. But, under private property, love already functions where it can. Nobody is prevented from doing something for free if he wants to. Many people—parents helping their children, volunteer workers in hospitals, scoutmasters—do just that. If, for those things that people are not willing to do for free, trade is replaced by anything, it must be by force. Instead of people being selfish and doing things because they want to, they will be unselfish and do them at the point of a gun.

> Is this accusation unfair? The alternative offered by those who deplore selfishness is always government. It is selfish to do something for money, so the slums should be cleaned up by a ‘youth corps’ staffed via ‘universal service’. Translated, that means the job should be done by people who will be put in jail if they do not do it.

I just highlighted this because it was a beautifully phrased argument.

> One of the most effective arguments against unregulated laissez faire has been that it invariably leads to monopoly. As George Orwell put it, “The trouble with competitions is that somebody wins them.” It is thus argued that government must intervene to prevent the formation of monopolies or, once formed, to control them. This is the usual justification for antitrust laws and such regulatory agencies as the Interstate Commerce Commission and the Civil Aeronautics Board.
>
> The best historical refutation of this thesis is in two books by socialist historian Gabriel Kolko: The Triumph of Conservatism and Railroads and Regulation . He argues that at the end of the last century businessmen believed the future was with bigness, with conglomerates and cartels, but were wrong. The organizations they formed to control markets and reduce costs were almost invariably failures, returning lower profits than their smaller competitors, unable to fix prices, and controlling a steadily shrinking share of the market.
>
> The regulatory commissions supposedly were formed to restrain monopolistic businessmen. Actually, Kolko argues, they were formed at the request of unsuccessful

> monopolists to prevent the competition which had frustrated their efforts.

So many books I need to read before I can have opinions on things.

> It was in 1884 that railroad men in large numbers realized the advantages to them of federal control; it took 34 years to get the government to set their rates for them. The airline industry was born in a period more friendly to regulation. In 1938 the Civil Aeronautics Board (CAB), initially called the Civil Aeronautics Administration, was formed. It was given the power to regulate airline fares, to allocate routes among airlines, and to control the entry of new firms into the airline business. From that day until the deregulation of the industry in the late 1970s, no new trunk line— no major, scheduled, interstate passenger carrier—was started.
>
> The CAB had one limitation: it could only regulate interstate airlines. There was one major intrastate route in the country— between San Francisco and Los Angeles. Pacific Southwest Airlines, which operated on that route, had no interstate operations and was therefore not subject to CAB rate fixing. Prior to deregulation, the fare between San Francisco and Los Angeles on PSA was about half that of any comparable interstate trip anywhere in the country. That gives us a good measure of the effect of the CAB on prices; it maintained them at about twice their competitive level.
>
> In this complicated world it is rare that a political argument can be proved with evidence readily accessible to everyone, but until deregulation the airline industry provided one such case. If you did not believe that the effect of government regulation of

> transportation was to drive prices up, you could call any reliable travel agent and ask whether all interstate airline fares were the same, how PSA’s fare between San Francisco and Los Angeles compared with the fare charged by the major airlines, and how that fare compared with the fare on other major intercity routes of comparable length. If you do not believe that the ICC and the CAB are on the side of the industries they regulate, figure out why they set minimum as well as maximum fares.

Continuing to have nothing much to say except “wow”.

> Defenders of [government health spending] programs argue that the poor are so poor they cannot afford vital medical care. Lurid reports to the contrary, most poor people are not on the edge of literal starvation; evidence indicates that in this country the number of calories consumed is virtually independent of income. If the poor spent more of their own money on doctors, they would not starve to death; they would merely eat worse, wear worse clothes, and live in even worse housing than they now do. If they do not spend very much money on medical care it is because that cost, which they are in an excellent position to evaluate, is too high.

Finally something where I can say something more interesting than wholehearted agreement.

The average cost of treatment for a heart attack is about $15,000. The poverty line for a single person in the US is $11,000. On the one hand, credit cards and loans can make up some of the difference; on the other, heart attacks are by no means even close to the most expensive medical condition. So if we're talking about actually buying health care then no, the poor *literally* cannot afford it.

If we're talking about buying health insurance, I understand a very cheap policy would cost about $2000, so the poor can probably literally afford that. I mean, they don't have a whole lot of fat to trim, but they can afford it in the sense that if they choose to give up their home and car, and live on the streets, then they can have the health insurance. At least until their job fires them because they don't have a car and can't get there, and so they lose the money they were using to pay for it. But they won't starve to death!

But then, why is starving to death such a uniquely interesting endpoint? Why assume that if the poor would die without health insurance we're morally obligated to give it to them, but if they wouldn't, we're not? If we're amoral or denying all obligations to help others, why care if the poor starve to death? And if we're not amoral and feel some responsibility to the poor, why not also be concerned about them having a minimally tolerable life?

If some libertarian doesn't think we have any obligation to help the poor, I'd rather they just say "Well, the poor might starve to death, but that's too bad."

Otherwise it seems sort of misleading to me. Saying "Well, the poor won't literally starve to death" sounds like you're saying "Well, it's not that bad." But if you were actually saying that, I could respond that it *is* that bad. It's just bad in a non-starvation-related way. If you don't care how bad it is, say so instead of hedging about whether starvation is occurring or not.

> The best solution to this problem would be for any state instituting a voucher system to include, as part of the

> initial legislation, the provision that any institution can qualify as a school on the basis of the performance of its graduates on objective examinations. In New York, for instance, the law might state that any school would be recognized if the average performance of its graduating class on the Regents exam was higher than the performance of the graduating classes of the bottom third of the state's public schools.

The best answer I've ever heard to the question of how to decide who gets school vouchers.

> It might be possible to reform our present universities in the direction of such free-market universities. One way would be by the introduction of a 'tuition diversion' plan. This arrangement would allow students, while purchasing most of their education from the university, to arrange some courses taught by instructors of their own choice. A group of students would inform the university that they wished to take a course from an instructor from outside the university during the next year. The university would multiply the number of students by the average spent from each student's tuition for the salary of one of his instructors for one quarter. The result would be the amount of their tuition the group wished to divert from paying an instructor of the university's choice to paying an instructor of their own choice. The university would offer him that sum to teach the course or courses proposed. If he accepted, the students would be obligated to take the course.
>
> The university would determine what credit, if any, was given for such courses. The number each student could take for credit might at first be severely limited. If the plan proved successful, it could be expanded until any such course could serve as an elective. Departments

would still decide whether a given course would satisfy specific departmental requirements.

A tuition diversion plan does not appear to be a very revolutionary proposal; it can begin on a small scale as an educational experiment of the sort dear to the heart of every liberal educator. Such plans could, in time, revolutionize the universities.

At first, tuition diversion would be used to hire famous scholars on sabbatical leave, political figures of the left or right, film directors invited by college film groups, and other such notables. But it would also offer young academics an alternative to a normal career. Capable teachers would find that, by attracting many students, they could get a much larger salary than by working for a university. The large and growing pool of skilled ‘free-lance’ teachers would encourage more schools to adopt tuition diversion plans and thus simplify their own faculty recruitment problems. Universities would have to offer substantial incentives to keep their better teachers from being drawn off into free-lancing. Such incentives might take the form of effective market structures within the university, rewarding departments and professors for attracting students. Large universities would become radically decentralized, approximating free-market universities. Many courses would be taught by free-lancers, and the departments would develop independence verging on autarchy.

Under such institutions the students, although they might have the help of advisory services, would have to take the primary responsibility for the structure of their own education. Many students enter college unready for such responsibility. A competitive educational market would evolve other institutions to serve their needs.

> These would probably be small colleges offering a highly structured education with close personal contact for students who wished to begin their education by submitting to a plan of study designed by those who are already educated. A student could study at such a college until he felt ready to oversee his own education and then transfer to a university.
>
> It is time to begin the subversion of the American system of higher schooling, with the objective not destruction but renaissance.

One of the better university reform proposals I've heard, plus an incremental strategy for achieving it!

> I have solved the problem of urban mass transit. To apply rny solution to a major city requires a private company willing to invest a million dollars or so in hardware and a few million more in advertising and organization. The cost is low because my transit system is already over 99 percent built; its essence is the more efficient use of our present multibillion dollar investment in roads and automobiles. I call it jitney transit; it can most easily be thought of as something between taxicabs and hitch-hiking. Jitney stops, like present-day bus stops, would be arranged conveniently about the city. A commuter heading into town with an empty car would stop at the first jitney stop he came to and pick up any passengers going his way. He would proceed along his normal route, dropping off passengers when he passed their stops. Each passenger would pay a fee, according to an existing schedule listing the price between any pair of stops.

Holy !@#$, I think he *has* solved the problem of urban mass transit. There's an obvious Uber parallel, but this system seems even better since it’s run by people going that direction anyway and each car will be packed, making the costs probably much cheaper. This is such an obviously good idea that I can only assume that it was regulation and the taxi lobby that prevented it from coming to pass. This paragraph probably did more to raise my confidence that there are extremely good libertarian solutions to important problems that we’re missing out on than anything else in the entire book.

> Urban renewal uses the power of the government to prevent slums from spreading, a process sometimes referred to as ‘preventing urban blight’. For middle-class people on the border of low-income areas, this is valuable protection. But ‘urban blight’ is precisely the process by which more housing becomes available to low-income people. The supporters of urban renewal claim that they are improving the housing of the poor. In the Hyde Park area of Chicago, where I have lived much of my life, they tore down old, low-rental apartment houses and replaced them with $30,000 and $40,000 town houses. A great improvement, for those poor with $30,000. And this is the rule, not the exception, as was shown years ago by Martin Anderson in The Federal Bulldozer

I don’t know much about urban renewal programs or whether they purport to help the poor; anyone want to weigh in here?

> Most conservatives now seem to have accepted, even embraced, the space program and with it the idea that the exploration of space can only be achieved by government. That idea is false. If we had not been in such a hurry, we not only could have landed a man on the moon, we could have done it at a profit.

> How? Perhaps as a television spectacular. The moon landing alone had an audience of 400 million. If pay TV were legal, that huge audience could have been charged several billion dollars for the series of shows leading up to, including, and following the landing. If the average viewer watched, altogether, twenty hours of Apollo programs, that would be about 25 cents an hour for the greatest show off earth…
>
> A greedy capitalist could have sold the moon landing in 1969 for something over $5 billion. The government spent $24 billion to get to the moon. It costs any government at least twice as much to do anything as it costs anyone else. It would have cost something under $12 billion to produce the Apollo program privately.
>
> But Apollo was a crash program. If we had been in less of a hurry, it would have cost far less. While we were waiting, economic growth would increase the price for which the moon landing could be sold and technological progress would cut the cost of getting there. We would have arrived, at a profit, sometime in the seventies.

This is the business model of Mars One, which may be a scam. Which makes me wonder: how come, if the business model is sound, in 25 years of us having approximately the technology necessary to go to Mars, no one has come up with a non-scam version of this?

The NFL makes $10 billion a year through TV ads and sponsorship rights. I don’t know if a Mars mission would do better or worse than that – certainly the touchdown would be more exciting, but would people tune in month after month for “Yup, we’re still in this capsule, it’s really cramped in here and outside the window it just looks black”?

Robert Zubrin says he thinks a private company could reach Mars for $5 billion, which sounds promising, but he gets that because the government estimate is $50 billion and he thinks private companies can be ten times more efficient. Come on, Robert Zubrin! Even *David Friedman* estimates more like twice as efficient. I also note that SpaceX is estimating $1 billion to convert their existing Dragon to a crew-ready Dragon. $1 billion for a famously efficient private company to go from existing small rocket + small capsule to slightly improved small rocket + small capsule that can go to low Earth orbit – and you're expecting another private company, right out of the gate, to be able to create ex nihilo a Mars-worthy spacecraft and the rocket that can launch it for $5 billion? Plus the astronaut training program, the production of the TV specials, the overhead for this new giant aerospace company you're founding, the cost of the colony itself, etc, etc? Really?

And even if it's possible in theory, think about the risk. The risk that the spacecraft explodes on the launch pad, and either you've just stuck your company name on a national tragedy or else you'd invested $6 billion in a TV special that's never going to happen. Or the risk that five years later, the Mars One people come to you and say "Okay, Robert Zubrin was way too optimistic, we spent all your money to build the spacecraft's left navigational fin, can you give us some more?" The risk that the Chinese beat you there and televising the second manned Mars landing isn't very exciting.

Nothing I've seen so far convinces me that a serious version of Mars One is anywhere on the horizon. SpaceX will probably send a man to Mars someday, but they'll do it because Elon Musk is vision-driven instead of profit-driven and he's making enough profits somewhere else to fund his vision. And I don't think even that would have worked without the funding and help that NASA has given SpaceX so far.

I worry that very big high-risk projects are exactly the sort of thing our current market system is really bad at.

> My own conclusion—that drug companies should be free to sell, and their customers to buy, anything, subject to liability for damages caused by misrepresentation—must seem monstrous to many people. Certainly it means accepting the near certainty of a few people a year dying from unexpected side effects of new drugs.

This probably needs its own post, but no no no no no no no no, regulating drugs by liability is *not* a good idea, maybe even a worse idea than regulating them with regulations. Just as a quick example, here is an excerpt from Wikipedia's article on the National Vaccine Injury Compensation Program:

"In 1988, the National Vaccine Injury Compensation Program (VICP) went into effect to compensate individuals and families of individuals who have been injured by covered childhood vaccines.[5] The VICP was adopted in response to an earlier scare over the pertussis portion of the DPT vaccine. These claims were later generally discredited, but some U.S. lawsuits against vaccine makers won substantial awards; most makers ceased production, and the last remaining major manufacturer threatened to do so. "

In other words, people kept winning so much money by suing the makers of pertussis vaccines that all of them except one just gave up and went out of business, and the only way the government saved that last one was by promising that the public purse would pay all of its losses. If the government hadn't stepped in, we would not have vaccines right now because lawsuits would have made it unprofitable to make them. Idiotic lawsuits, I might add – pertussis vaccine doesn't actually hurt people in any way. This is "my kid got autism after getting a vaccine" level stuff, and the courts were just like "Sure, fine, we believe you, let's make the vaccine companies pay you so much money they all go bankrupt."

This is not an isolated incident. The way malpractice works these days is that patients sue for things that are completely medically impossible, the malpractice insurances know that juries are too dumb to realize this, and they settle for more money than you will ever make honestly in your life. The FDA and its regulations are actually a rare force limiting this madness – if nothing else, a doctor can say "Well, that drug was approved by the FDA, so I wasn't negligent in prescribing it to you."

I understand that this book's proposals include a large package of reforms which include those to the court system. But Friedman's worries about how any "limited government" will eventually regrow into the kind of government that says you feeding your own grain to your own pigs is interstate commerce, are matched by my worries about how any "reformed court system" will eventually regrow into the kind of court system where children must be banned from sledding because if they get hurt they can sue the city for not having banned sledding, or lots of people who come to a psych hospital have to be committed lest years later somebody sue the hospital for not committing them.

If you invite more lawyers in to help control the government, you might end up like that Irish warlord who invited the English in to help control a rival warlord; you'll find they're even worse and they never leave.

> The argument of this chapter received striking support in 1981, when the FDA published a press release confessing to mass murder. That was not, of course, the way in which the release was worded; it was simply an announcement that the FDA had approved the use of timolol, a ß-blocker, to prevent recurrences of heart attacks. At the time timolol was approved, ß-blockers had been widely used outside the U.S. for over ten years. It was estimated that the use of timolol would save from seven thousand to ten thousand lives a year in the U.S. So the FDA, by forbidding the use of ß-blockers before l981, was responsible for something close to a hundred thousand unnecessary deaths.

If examples of times when bad FDA decisions cost tens of thousands of lives made people abolish the FDA, we would probably have like negative seventeen FDAs by now.

> Special interest politics is a simple game. A hundred people sit in a circle, each with his pocket full of pennies. A politician walks around the outside of the circle, taking a penny from each person. No one minds; who cares about a penny? When he has gotten all the way around the circle, the politician throws fifty cents down in front of one person, who is overjoyed at the unexpected windfall. The process is repeated, ending with a different person. After a hundred rounds everyone is a hundred cents poorer, fifty cents richer, and happy.

Annnnd we’re back to me just highlighting passages for rhetorical brilliance.

> How much would it cost workers to purchase their firms? The total value of the shares of all stocks listed on the New York Stock Exchange in 1965 was $537 billion. The

> total wages and salaries of all private employees that year was $288.5 billion. State and federal income taxes totalled $75.2 billion. If the workers had chosen to live at the consumption standard of hippies, saving half their after-tax incomes, they could have gotten a majority share in every firm in two and a half years and bought the capitalists out, lock, stock, and barrel, in five. That is a substantial cost, but surely it is cheaper than organizing a revolution. Also less of a gamble. And, unlike a revolution, it does not have to be done all at once. The employees of one firm can buy it this decade, then use their profits to help fellow workers buy theirs later.
>
> When you buy stock, you pay not only for the capital assets of the firm—buildings, machines, inventory, and the like —but also for its experience, reputation, and organization. If workers really can run firms better, these are unnecessary; all they need are the physical assets. Those assets—the net working capital of all corporations in the United States in 1965—totalled $171.7 billion. The workers could buy that much and go into business for themselves with 14 months' worth of savings.

Compare to A Future For Socialism. In the research for that post I believe I found that the ratio of capital assets to wages had been rising pretty sharply recently, so it might take more time these days. But even if it took an entire decade, that's a lot faster than most Communists expect the Revolution to come.

It probably says something very important about human nature and politics that the Socialist movement isn't dominated by the project of doing exactly this.

Is Everything A Religion?

I.

On the last Links thread, Eric Raymond claims that environmentalism is a religion. It has "sins" like wasting energy and driving gas-guzzling SUVs. It has "taboos" like genetically modified foods. It has an "apocalypse" in the form of global warming. It even has "rituals" in the form of weekly recycling.

This reminds me of an article I read recently claiming that transhumanism is a religion. But also of the article claiming that social justice is a religion. Also, liberalism is a religion. And conservativism is a religion. Libertarianism is a religion. Communism is a religion. Capitalism is like a religion. Objectivism is a religion. An anthropologist "confirms" that Apple is a religion. But UNIX is also a religion (apparently Linux was the Protestant Reformation).

Is there anything that isn't like a religion? I spent this morning trying to come up with the least religious things I could think of. Trying to think of practical disciplines aimed at producing a quantifiable result, disciplines which strive to be evidence-based with a minimum of extraneous ideology. What came to mind was investing and medicine.

But investing is about propitiating a mysterious deity (the market) whose blessing or wrath bestows innumerable riches or total ruin. Believers follow gurus like Warren Buffett and Jim Cramer who promise that if they do the right things they will achieve financial salvation. Those who follow their pronouncements will enjoy the blissful afterlife of a comfortable retirement; those who violate their laws will spent their retirement in penury among much wailing and gnashing of teeth.

And medicine involves petitioners going to white-robed priests (doctors) who consult the holy scriptures (Harrison's Clinical Medicine) to tell them how to live their lives. It has rituals (the yearly physical), taboos (smoking, overeating), and heretics (alternative medicine). Those who follow its rules are assured of a long, happy life; those who violate the rules of its priests will get cancer and die.

Maybe we're still being too abstract here. What about, I don't know, not stepping in front of buses? It certainly has a commandment (thou shalt not step in front of buses). It has notions of sin (stepping in front of buses) and virtue (not doing that). It has its rituals (looking both ways before you cross the street), its priests demanding obedience (crossing-guards), and its holy places (crosswalks). It promises blessings on the virtuous, but also terrible vengeance on the wicked (if you step in front of a bus, there will be much wailing and gnashing of teeth).

So one critique of these accusations is that "religion" is a broad enough category that anything can be mapped on to it:

Does it have well-known figures? Then they're "gurus" and it's a religion.

Are there books about it? Then those are "scriptures" and it's a religion.

Does it recommend doing anything regularly? Then those are "rituals" and it's a religion.

How about just doing anything at all? Then that's a "commandment" and it's a religion.

Does it say something is bad? Then that's "sin" and it's a religion.

Does it hope to improve the world, or worry about the world getting worse? That's an "eschatology" and it's a religion.

Do you disagree with it? Then since you've already determined all the evidence is against it, people must believe it on "faith" and it's a religion.

II.

But that critique goes just a *little* too far. Once Communists start offering animal sacrifices to statues of Mao and requiring everyone own a copy of the Little Red Book and treat it respectfully, something is going on that's deeper than just "it has well-known figures".

Even though it's easy to say that every belief or movement can be analogized to a religion, I still feel an intuition that some are more "religious" than others. Environmentalism and social justice seem more religious than gun control and pro-choice, even though all four are equally important lefty issues.

The first two are just more of a world-view. I can totally imagine someone saying "My life philosophy is centered around my passion for the environment", but not so much "My life philosophy is centered around gun control." I can see a speaker at a wedding saying "John and Jane are perfect for each other, since they are united by their shared passion about social justice", but not so much "John and Jane are perfect for each other, since they are united by their shared passion for abortion rights."

Both social justice and environmentalism spawn entire genres of art and literature, and I know people who pretty much exclusively draw their artistic consumption from those genres. But if somebody said “All of my art has a pro-choice theme”, that would probably be pretty creepy.

I know social justice people whose social circle is almost 100% based on social justice, and environmentalists whose social circle is almost 100% based on environmentalism. I don’t think there are that many people whose social circle is 100% based on gun control. And if someone says “I’m fanatical about the environment”, I get a whole lot of stereotypes about them – she probably eats granola, drives a Prius with a dreamcatcher in the window, has a college degree, does yoga. He probably goes hiking a lot, has a beard, takes supplements, is pretty relaxed. If someone says “I’m fanatical about gun control”, I’m stumped.

But all of this stuff about stereotypes and art and insularity sounds a little like religion but even more like culture, or at least subculture.

The difference between “religion” and “culture” has always been pretty vague. Shinto is the best example; it’s less a coherent metaphysical narrative than a bunch of things Japanese people do and a repository for Japanese traditions and rituals. A quick look at Hinduism reveals that they have *no idea* what gods they believe in, it’s a bunch of different religions stuck together under one umbrella, but the point is that it’s the sort of thing Indian people do and a repository of Indian traditions. Even though Jews have a pretty coherent religion, the line between “Jewish culture” and “Jewish religion” is equally fuzzy. Religion as distinct from culture seems like a pretty Western phenomenon, the result of a triumphant Christianity colonizing cultures it never originated from, ending out with the modern conception of culture as ethnic food + silly costumes.

American culture is paper-thin compared to say Hindu Indian culture, but consider its rituals like the Pledge of Allegiance, its holidays like the Fourth of July, its saints/culture heroes like George Washington and Benjamin Franklin, its myths like Paul Bunyan and Johnny Appleseed, its veneration of founding documents (the Declaration of Independence, the Constitution), and even its hymns like “America the Beautiful” and “Yankee Doodle”.

(the last of which, like all good hymns, uses such archaic language that almost nobody knows what the heck it means)

This gets called American civil religion a lot, but at this point I’m starting to wonder why it should. Maybe instead of accusing every culture of becoming a religion, we should just admit that our current concept of “religion” actually owes a lot to “culture”.

Eliezer writes that every cause wants to be a cult, but I’m not sure I agree with the connotations. I would say every cause wants to be a community. Communities hold values in common. Communities have rules their members have to follow. Communities have heroes and hierarchies. Communities shun people who don’t fit in.

And if all of this sounds super-conservative, keep in mind we’re still talking about environmentalism here, or social justice here. Values in common? Check. Rules? God yes. Heroes and hierarchies? You bet. Shunning people? All the time.

Communities and cultures have their share of danger. Their mix of social and epistemological functions means that any evidence challenging the community’s core beliefs will be taken as an attack on the members’ identity. As a result, community members risk ending up mind-killed. That’s not news. And I don’t think this is especially different from the way religious fanatics are mind-killed. And certainly someone could argue that “religion” is the perfect name for a culture built on shared belief.

But I *still* think it’s unfair to call these communities/cultures “religions”. “Religion” is too easy to use as the Worst Argument In The World here. It’s supposed to imply all of these other connotations of “religion” like “their beliefs are based on magical thinking” and “they use blind faith instead of reason” and “instead of coming up with a world-view based on evidence they just played Bible Mad Libs.” If those are the connotations you’ve got with “religion”, then I think the word “religion” is actively doing harm here, and you should just use “belief-based community” or “movement” or whatever.

Extremism In Thought Experiment Is No Vice

[content warning: description of fictional rape and torture.]

Phil Robertson is being criticized for a thought experiment in which an atheist's family is raped and murdered. On a talk show, he accused atheists of believing that there was no such thing as objective right or wrong, then continued:

> I'll make a bet with you. Two guys break into an atheist's home. He has a little atheist wife and two little atheist daughters. Two guys break into his home and tie him up in a chair and gag him.
>
> Then they take his two daughters in front of him and rape both of them and then shoot them, and they take his wife and then decapitate her head off in front of him, and then they can look at him and say, 'Isn't it great that I don't have to worry about being judged? Isn't it great that there's nothing wrong with this? There's no right or wrong, now, is it dude?'
>
> Then you take a sharp knife and take his manhood and hold it in front of him and say, 'Wouldn't it be something if [there] was something wrong with this? But you're the one who says there is no God, there's no right, there's no wrong, so we're just having fun. We're sick in the head, have a nice day.'
>
> If it happened to them, they probably would say, 'Something about this just ain't right'.

The media has completely proportionally described this as Robinson "fantasizing about" raping atheists, and there are the usual calls for him to apologize/get fired/be beheaded.

So let me use whatever credibility I have as a guy with a philosophy degree to confirm that Phil Robertson is doing moral philosophy exactly right.

There’s a tradition at least as old as Kant of investigating philosophical dilemmas by appealing to our intuitions about extreme cases. Kant, remember, proposed that it was always wrong to lie. A contemporary of his, Benjamin Constant, made the following objection: suppose a murderer is at the door and wants to know where your friend is so he can murder her. If you say nothing, the murderer will get angry and kill you; if you tell the truth he will find and kill your friend; if you lie, he will go on a wild goose chase and give you time to call the police. Lying doesn’t sound so immoral *now*, does it?

The brilliance of Constant’s thought experiment lies in its extreme nature. If a person says they think lying is always wrong, we have two competing hypotheses: they’re accurately describing their own thought processes, which will indeed always output that lying is wrong; or they’re misjudging their own thought processes and actually there are some situations in which they will judge lying to be ethical. In order to distinguish between the two, we need to come up with a story that presents the strongest possible case for lying, so that even the tiniest shred of sympathy for lying can be dragged up to the surface.

So Constant says “It’s a murderer trying to kill your best friend”. And even this is suboptimal. It should be a mad scientist trying to kill everyone on Earth. Or an ancient demon, whose victory would doom everyone on Earth, man, woman, and child, to an eternity of the most terrible torture. If some people’s hidden algorithm is “lie when the stakes are high enough", there we can be sure that the stakes are high enough to tease it out into the light of day.

Compare Churchill:

> **Churchill:** Madam, would you sleep with me for five million pounds?
> **Lady:** Well, for five million pounds…well…that's a lot of money.
> **Churchill:** Would you sleep with me for five pounds?
> **Lady:** *(enraged)* What kind of a woman do you think I am‽
> **Churchill:** We've already established what kind of a woman you are, now we're just haggling over the price

The woman thinks she has a principle, "Never sleep with a man for money". In fact, deep down, she believes it's okay to sleep with a man for enough money. If Churchill had merely stuck to the five pounds question, she would have continued to believe she held the "never…" principle. By coming up with an extreme case (5 million Churchill-era pounds is about £250 million today) he was able to reveal that her apparent principle was actually a contingent effect of her real principle plus the situation.

In fact, compare physics. Physicists are always doing things like cooling stuff down to a millionth of a degree above absolute zero, or making clocks so precise they'll be less than a second off by the time the sun goes out, or acclerating things to 99.99% of the speed of light. And one of the main reasons they do is to magnify small effects to the point where they can measure them. All movement is causing a *little* bit of time dilation, but if you want to detect it you need the world's most accurate clock on the Space Shuttle when it's traveling 25,000 miles per hour. In order to figure out how things really work, you need to turn things up to 11 so that the effect you want is impossible to miss. Everything in the universe has been exerting a gravitational effect on light all the time, but if you want to see it clearly you need to use the Sun during a solar eclipse, and if you *really* want to see it clearly your best bet is a black hole.

Great physicists and great philosophers share a certain perversity. The perversity is “Sure, this principle works in all remotely plausible real-world situations, but WHAT IF THERE’S A COMPLETELY RIDICULOUS SCENARIO WHERE IT DOESN’T HOLD??!?!” Newton’s theory of gravity explained everything from falling apples to the orbits of the planets impeccably for centuries, and then Einstein asked “Okay, but what if, when you get objects thousands of times larger than the Earth, there are tiny discrepancies in it, then we’d have to throw the whole thing out,” and instead of running him out of town on a rail scientists celebrated his genius. Likewise, moral philosophers are as happy as anyone else not to lie in the real world. But they wonder whether they might be revealed to be only simplifications of more fundamental principles, principles that can only be discovered by placing them in a cyclotron and accelerating them to 99.99% of the speed of light.

Sometimes this is even clearer than in the Kant example. Many people, if they think about it at all, believe that value aggregates linearly. That is, two murders are twice as much of a tragedy as one murder; a hundred people losing their homes is ten times as bad as ten people losing their homes.

Torture vs. Dust Specks is beautiful in its simplicity; it just takes this assumption and creates the most extreme case imaginable. Take a tiny harm and aggregate it an unimaginably high number of times; then compare it to against a big harm which is nowhere near the aggregated sum of the tiny ones. So which is worse, 3^^^3 (read: a number higher than you can imagine) people getting a single dust speck in their eye for a fraction of a second, or one person being tortured for fifty years?

Almost everybody thinks their principle is “things aggregate linearly”, but when you put it into relief like this, almost everybody’s intuition tells them the torture is worse. You can “bite the bullet” and admit that the dust specks are worse than the torture. Or you can throw out your previous principle saying that things aggregate linearly and try to find another principle about how to aggregate things (good luck).

Moral dilemmas are extreme and disgusting precisely because those are the only cases in which we can make our intuitions strong enough to be clearly detectable. If the question was just “Which is worse, a thousand people stubbing their toe or one person breaking their leg?” neither side would have been obviously worse than the other and our true intutition wouldn’t have come into sharp relief. So a good moral philosopher will *always* be talking about things like murder, torture, organ-stealing, Hitler, incest, drowning children, the death of four billion humans, et cetera.

Worse, a good moral philosopher should be constantly agreeing – or tempted to agree – to do horrible things in these cases. The whole point of these experiments is to collide two of your intuitions against each other and force you to violate at least one of them. In Kant’s example, either you’re lying, or you’re dooming your friend to die. In Jarvis’ Transplant Surgeon scenario, you’re either killing somebody to harvest their organs, or letting a whole hospital full of people die.

I once had someone call the torture vs. dust specks question “contrived moral dilemma porn” and say it proved that moral philosophers were kind of crappy people for even considering it. That bothered me. To look at moral philosophers and conclude "THESE PEOPLE LOVE TO TALK ABOUT INCEST AND ORGAN HARVESTING, AND BRAG ABOUT ALL THE CASES WHEN THEY'D BE OKAY DOING THAT STUFF. THEY ARE GROSS EDGELORDS AND PROBABLY FANTASIZE ABOUT HAVING SEX WITH THEIR SISTER ON THE HOSPITAL BED OF A PATIENT DYING OF END-STAGE KIDNEY DISEASE," is to utterly miss the point.

So let's talk about Phil Robertson.

Phil Robertson believes atheists are moral nihilists, or moral relativists, or something. He's not quite right – there are a lot of atheists who are very moral realist – Objectivists, as their name implies, believe morality and everything else up to and including the best flavor of ice cream, is Objective – and even the atheists who aren't *quite* moral realist usually hold some sort of compromise position where it's meaningful to talk about right and wrong even if it's not *cosmically* meaningful.

On the other hand – and I say this as the former secretary of a college atheist club who got to meet *all sorts* – there are a bunch of atheists who very much claim not to believe in morality. Less Wrong probably has fewer of them than the average atheist hangout, because we skew so heavily utilitarian, but our survey records 4% error theorists and 9% non-cognitivists. When Friendly Atheist says he "doesn't know a single atheist or agnostic who thinks that terrorizing, raping, torturing, mutilating, and killing people is remotely OK", I can believe that he doesn't know one who would say so in those exact words. But I'm not sure how, for example, the error theorists could consistently argue against that position.

And what Phil Robertson does is exactly what I would do if I were debating an error theorist. I'd take the most gratuitiously horrible thing I could think of, describe it in the most graphic detail I could, and say "But don't you think there's something wrong with *this*?" If the error theorist says "no", then I congratulate her for definitely being a real honest-to-goodness error theorist, and unless I can suddenly think up a way to bridge the is-ought dichotomy we're finished. But if she says "Yes, it does seem like there should be something wrong there," then we can start exploring what that means and whether error theory is the best framework in which to capture that intuition.

On the other hand, if I were debating Phil Robertson, I would ask him where *he* thinks morality comes from. And if he suggested some version of divine command theory, I could use an example of the graphic-horrifying-extreme-thought-experiment genre even older than Kant – namely, Abraham's near-sacrifice of Isaac. If God commands you to kill your innocent child, is that the right thing to do? What if God commands you to rape and torture and mutilate your family? And it wouldn't work if it were anything less extreme – if I just said "What if God told you to shoplift?" it would be *easy* to bite that bullet and he wouldn't have to face the full implication of his views. But if I went with the extreme version? Maybe Robertson would find he's not as big on divine command theory as he thought.

But this sort of discussion would only be possible if we could trust each other to take graphic thought experiments in the spirit in which they were conceived, and not as an opportunity to score cheap points.

[EDIT: This post was previously titled "High Energy Ethics", but I changed it after realizing it was unintentionally lifted from elsewhere]

Highlights From My Notes From Another Psychiatry Conference

I took a break from my busy schedule at Our Lady Of An Undisclosed Location to attend another local Psychiatry Conference.

This conference consisted of a series of talks about all the most important issues of the day, like ‘The Menace Of Psychologists Being Allowed To Prescribe Medication’, ‘How To Be An Advocate For Important Issues Affecting Your Patients Such As The Possibility That Psychologists Might Be Allowed To Prescribe Them Medication’, and ‘Protecting Members Of Disadvantaged Communities From Psychologists Prescribing Them Medication’.

As somebody who’s noticed that the average waiting list for a desperately ill person to see a psychiatrist is approaching the twelve month mark in some places, I was pretty okay with psychologists prescribing medication. The scare stories about how psychologists might prescribe medications unsafely didn’t have much effect on me, since I continue to believe that putting antidepressants in a vending machine would be a more safety-conscious system than what we have now (a vending machine would at least limit antidepressants to people who have $1.25 in change; the average primary care doctor is nowhere near that selective). Annnnnyway, this made me kind of uncomfortable at the conference and I Struck A Courageous Blow Against The Cartelization Of Medicine by sneaking out without putting my name on their mailing list.

But before I did, I managed to take some notes about what’s going on in the wider psychiatric world, including:

– The newest breakthrough in ensuring schizophrenic people take their medication (a hard problem!) is bundling the pills with an ingestable computer chip that transmits data from the patient's stomach. It's a bold plan, somewhat complicated by the fact that one of the most common symptoms of schizophrenia is the paranoid fear that somebody has implanted a chip in your body to monitor you. Can you imagine being a schizophrenic guy who has to explain to your new doctor that your old doctor put computer chips in your pills to monitor you? Yikes. If they go through with this, I hope they publish the results in the form of a sequel to The Three Christs of Ypsilanti.

– The same team is working on a smartphone app to detect schizophrenic relapses. The system uses GPS to monitor location, accelerometer to detect movements, and microphone to check tone of voice and speaking pattern, then throws it into a machine learning system that tries to differentiate psychotic from normal behavior (for example, psychotic people might speak faster, or rock back and forth a lot). Again, interesting idea. But *again*, one of the most common paranoid schizophrenic delusions is that their electronic devices are monitoring everything they do. If you make every one of a psychotic person's delusions come true, such that they no longer have any beliefs that do not correspond to reality, does that technically mean you've cured them? I don't know, but I'm glad we have people investigating this important issue.

– I'll come out and say it: cluster randomization is really sketchy. Today I got to hear about a multi-center trial which randomized by location – half of their hospitals were the control group, the other half were the experimental group. Problem is, the patients in each hospital were given group-appropriate consent forms – either "We will be treating you as usual, but monitoring you more closely for a study" or

“We will be giving you extra experimental treatment”. Not only does that break blinding, but it implies a different population of patients in each group – the ones willing to consent to monitoring versus the ones willing to consent to treatment? Might sicker people be more willing to sign the treatment consent, since they don’t want to deal with monitoring but treatment offers the chance for personal gain? Might paranoid people be more willing to sign the control consent, since they’re not being used as guinea pigs? I don’t know. But I checked those pre-intervention inter-group comparisons they have to show, and there were big differences between the two groups (for example, I think one – I can’t remember which – had like twice as many black people). Either randomize peopple properly or at least keep people blind to condition.

– On the other hand, I’m quickly losing my prejudice that RCTs always beat naturalistic studies. I’ll write more about this later, but today’s showcase was long-acting injectable versus oral antipsychotics. Conventional wisdom is that long-acting antipsychotics, in the right patient population, decrease relapse because they remove the option of not taking the medication. The best randomized controlled trials don’t find that. The best naturalistic epidemiological studies do. The expert who spoke today theorized – and I agree – that the naturalistic studies are right. He argued that one feature of RCTs is very close monitoring, which means the patients in them comply with their medication at an unnaturally high rate – thus removing the long-acting drugs’ one advantage. The studies conducted in the real world of patients not taking their medications regularly are more relevant.

– They say psychotic people don’t take their meds because they hate the side effects, or because they’re too crazy to know better, or because they just can’t be bothered. But one of the doctors today raised a novel hypothesis: are antipsychotics anti-addictive? After all, some of the most addictive drugs are those that raise dopamine levels – cocaine, meth, and MDMA are all either dopamine releasing agents or dopamine reuptake inhibitors. Antipsychotics have pretty much the opposite effect as those, lowering dopamine in the brain. Suspicious. But I have a feeling this isn't true. Dopamine is more complicated than that. Levodopa-carbidopa, which is one step short of pure dopamine and is given to dopamine-deficient Parksinson's patients, is as far as I know not addictive at all. It's also very clearly antagonistic to antipsychotics. Probably antipsychotics are the opposite of non-addictive levodopa, not the opposite of cocaine or anything. I don't know how to phrase it more rigorously than that. Still, I like the way that person thinks.

– Ever since Indiana's legislature debated a bill that implied pi = 4, Midwestern states have had a reputation for trying to legislate science. Maybe this had something to do with the claim by one psychiatry lobbyist that Kansas' legislature is trying to ban the DSM. I can't find anything on it online and it sounds like an urban legend to me. Tangentially related silly clickbait: Arizona lawmakers say horses aren't animals.

– Unintentional puns are some of my favorite puns. I still remember fondly when the head of a psychiatric hospital where I used to work said that if Obamacare passed there would be too many patients and the place would "turn into a madhouse". I collected another good one today when an activist was talking about gun rights for psychiatric patients: "Taking guns from psychiatric patients isn't going to be a panacea for violence – would anyone like to take a stab at why?"

– Clozapine really is the best antipsychotic, hands down, and the evidence isn't even subtle. It's also the most dangerous, and the rules say that you should only prescribe it to a patient after you've tried and failed with two other antipsychotics. One of the speakers was a researcher who's trying to get a grant to prove that it's actually more effective to try clozapine after only one failed antipsychotic, but the NIMH rejected his proposal because “even if you proved that, no one would listen”. They're probably right. A lot of psychiatrists hate clozapine because it's messy, scary, and requires a lot of paperwork and monitoring. The speaker presented survey after survey of psychiatrists making lame excuses like “My patients wouldn't want it”, and then survey after survey of those psychiatrists' patients saying they *do so* want it but nobody asked them. Clozapine is messy and scary and requires lots of paperwork, but if you're a good doctor you'll give your patient the drug that will help them anyway.

– The APA representative says that 95% of candidates supported by the APA's PAC get elected. I think it was supposed to be a boast, like “look how effective we are”, but that's a bit much. Either the APA single-handedly controls all American politics, or else they're very careful to always back the winning side. Properly understood, that number should probably be taken as a measure of exactly how cynical they are.

– Not that they didn't admit their cynicism straight out. Our Political Activism Consultant explained that state legislators are all sorta new and confused and inexperienced all the time because of term limits. And if you put on a nice suit and a tie and tell them “Hey, I'm a doctor from your district, here's how you need to do health care policy…” you have a pretty good chance of getting them to nod along and assume you know what you’re doing. I didn’t realize how easy this was, and I hope I never use this power for evil.

– This is basically how the Eternal War Against Psychologists Being Allowed To Prescribe Medications is being fought, but the psychologists have caught on and now they have nice suits and ties too. Also, it turns out senators have a hard time differentiating the APA (American Psychiatric Association, fighting tooth and claw against psychologist prescribers) from the APA (American Psychological Association, fighting tooth and claw *for* psychologist prescribers) and they end up freaking out and trying to figure out why the same people are lobbying for both sides and whether this is some kind of weird shrink mind game thing.

– Drug companies were giving out stress brains! Like stress balls, only they’re shaped like brains and have little sulci and gyri on them! If in ten years I’m one of those people who never prescribes clozapine, it’ll because I’m prescribing the drug by the company that gave me a stress brain instead.

Rational Orthography

What do DVORAK, polyamory, and home schooling have in common? They’re all about doing what’s weird-but-effective instead of what’s popular. What else is like that?

About three thousand years ago, the ancient Greeks invented a form of writing called boustrophedon. The first line was written left-to-right, the second right-to-left, and so on in a winding pattern. The advantage of the new system was that it was faster and easier to read – instead of constantly darting your eyes back and forth from one side of the page to the other at the end of each line, you just let them continue naturally.

The disadvantage was that it was hard to write, for much the same reasons most people would have trouble writing backwards now. So although boustrophedon and straight left-to-right Greek competed for a couple of centuries, in the end straight Greek won because the scribes were too lazy to do what was most convenient for their readers. According to Right Hand, Left Hand: The Origins of Asymmetry in Brains, Bodies, Atoms and Cultures:

> A risk when offering any historical description is what has been called ‘the Whig interpretation of history’, the easy presumption that everything leads straightforwardly and inexorably to the highest state of humankind. Such an interpretation fails to look at the entire historical picture, ignoring the losers – in our case, the writing systems that became extinct […]
>
> Boustrophedon, the writing of the ox, is, as it were, on the horns of a dilemma; either it is easier to read and more difficult to write, or vice versa. It is not surprising

> that it rapidly died out in ancient writing. Perhaps more surprising are moves to reintroduce it. Computers can be programmed so that only the standard twenty-six letters have to be typed on the keyboard, but the screen display or printout has normal or mirror-reversed letters according to the direction of the script. Enthusiasts claim boustrophedon is easier and quicker to read because the eye does not have to find its way back to the beginning of the next line.

These sorts of things have to start somewhere. So I asked SSC reader Bakkot to create a script that causes this blog to display in boustrophedon. I think you'll agree that the experience is much improved. If there's enough demand for a "classic" view, I can ask him to create some kind of optional browser add-on that will disable it, but I'd urge you to try the new version for a couple of weeks before turning to that "solution".

More important: what if you want *everything* you read to be in boustrophedon from now on? For that I can unreservedly recommend The Boustrophedon Text Reader. Right now it only works on .txt, but hopefully as the movement catches on someone can turn it into a full browser extension.

Chemical Imbalance

[content note: mental illness. I am still in training and do not understand these issues even as well as a fully-trained psychiatrist, let alone a researcher, so take all the biology and studies in here with a grain of salt until you double-check]

I.

IO9's new article The Most Popular Antidepressants Are Based On An Outdated Theory jumps on a popular bandwagon of criticizing psychiatry for botching the "chemical imbalance" theory. See for example The New Yorker, BBC, The New York Times, and various books.

(…and also The Myth Of Chemical Imbalance, Debunking The Chemical Imbalance Myth, The Chemical Imbalance Fraud, and Depression Delusion, The Myth Of The Chemical Imbalance, etc)

According to all these sources psychiatry sold the public on antidepressants by claiming depression was just a chemical imbalance (usually fleshed out as "a simple deficiency of serotonin") and so it was perfectly natural to take extra chemicals to correct it. However, they had no real evidence for this theory except that serotonergic drugs effectively treat depression, which is not very much evidence at all (antibiotics effectively treat pneumonia, but pneumonia isn't "an antibiotic deficiency"). And now the research is unequivocal that serotonin deficiency is not the cause of depression, and psychiatry has ended up with lots of egg on its face.

This narrative is getting pushed especially hard by the antipsychiatry movement, who frame it as "proof" that psychiatrists are drug company shills who were deceiving the public. The conversation has required a host of rebuttals and counter-rebuttals.

For example here antipsychiatry blog Mad In America attemps to rebut psychiatrist Dr. Ronald Pies, who argues that psychiatrists never pushed the chemical imbalance theory. Pies says that “The ‘chemical imbalance theory’ was never a real theory, nor was it widely propounded by responsible practitioners in the field of psychiatry,” and cites the American Psychiatric Association’s 2005 statement on the causes of depression:

> The exact causes of mental disorders are unknown, but an explosive growth of research has brought us closer to the answers. We can say that certain inherited dispositions interact with triggering environmental factors. Poverty and stress are well-known to be bad for your health—this is true for mental health and physical health. In fact, the distinction between “mental” illness and “physical” illness can be misleading. Like physical illnesses, mental disorders can have a biological nature. Many physical illnesses can also have a strong emotional component

Mad In America doesn’t accept his claim, and counter-cites two speeches by American Psychiatric Association presidents to prove that they *did* push the chemical imbalance theory:

> In the last decade, neuroscience and psychiatric research has begun to unlock the brain’s secrets. We now know that mental illnesses – such as depression or schizophrenia – are not “moral weaknesses” or “imagined” but real diseases caused by abnormalities of brain structure and imbalances of chemicals in the brain.” – Richard Harding, 2001 APA president

And:

> The way nerves talk to each other, and communicate, is through the secretion of a chemical called a neurotransmitter, which stimulates the circuit to be activated. And when this regulation of chemical neurotransmission is disturbed, you have the alterations in the functions that those brain areas are supposed to, to mediate. So in a condition like depression, or mania, which occurs in bipolar disorder, you have a disturbance in the neurochemistry in the part of the brain that regulates emotion. – Jeffrey Lieberman, 2012 APA President

I have no personal skin in this game. I’ve only been a psychiatrist for two years, which means I started well after the term “chemical imbalance” fell out of fashion. I get to use the excuse favored by young children everywhere: “It was like this when I got here”. But I *still* feel like the accusations in this case are unfair, and I would like to defend my profession.

I propose that the term “chemical imbalance” hides a sort of bait-and-switch going on between the following two statements:

(A): Depression is complicated, but it seems to involve disruptions to the levels of brain chemicals in some important way

(B): We understand depression perfectly now, it’s just a deficiency of serotonin.

If you equivocate between them, you can prove that psychiatrists were saying (A), and you can prove that (B) is false and stupid, and then it’s *sort of* like psychiatrists were saying something false and stupid!

But it isn’t too hard to prove that psychiatrists, when they talked about “chemical imbalance”, meant something more like (A). I mean, look at the quotes above by which *Mad In America* tries to prove psychiatrists guilty of pushing chemical imbalance. Both sound more like (A) than (B). Neither mentions serotonin by name. Both talk about the chemical aspect as part of a larger picture: Harding in the context of abnormalities in brain structure, Lieberman in the context of some external force disrupting neurotransmission. Neither uses the word “serotonin” or “deficiency”. If the antipsychiatry community had quotes of APA officials saying it’s all serotonin deficiency, don’t you think they would have used them?

Further, anyone who said that depression was caused solely by serotonin deficiency wouldn’t just be failing as a scientist, but also failing as a drug company shill. Pfizer spent billions of dollars on Effexor, which hits norepinephrine as well as serotonin, and they’re just going to dismiss all of that as useless? GlaxoSmithKline has Wellbutrin, which hits dopamine and norepinephrine and maybe acetylcholine but doesn’t get serotonin *at all*. So everyone, including the shills, *especially* the shills, has been very careful to say that depression was a “chemical imbalance” rather than a serotonin deficiency per se.

So if you want to prove that psychiatrists were deluded or deceitful, you’re going to have to disprove not just statement (B) – which never represented a good scientific or clinical consensus – but statement (A). And that’s going to be hard, because as far as I can tell statement (A) still looks pretty plausible.

II.

If you listen to these articles, psychiatrists decided that neurotransmitters (or just serotonin?) were implicated in depression solely on the evidence that SSRIs were effective antidepressants, even though every study trying to measure serotonin levels directly came back with negative results. For example, The Myth Of The Chemical Imbalance Theory writes:

> There is no question that the chemical imbalance theory has spurred chemists to invent new anti-depressants, or that these anti-depressants have been shown to work; but proof that low serotonin is to blame for depression – and that boosting serotonin levels is the key to its treatment – has eluded researchers.
>
> For starters, it is impossible to directly measure brain serotonin levels in humans. You can't sample human brain tissue without also destroying it. A crude work-around involves measuring levels of a serotonin metabolite, 5-HIAA, in cerebrospinal fluid (CSF), which can only be obtained with a spinal tap. A handful of studies from the 1980s found slightly decreased 5-HIAA in the CSF of depressed and suicidal patients, while later studies have produced conflicting results on whether SSRIs lower or raise CSF levels of 5-HIAA. These studies are all circumstantial with regards to actual serotonin levels, though, and the fact remains there is no direct evidence of a chemical imbalance underlying depression.
>
> The corollary to the chemical imbalance theory, which implies that raising brain serotonin levels alleviates depression, has also been hard to prove. As mentioned previously, the serotonin-depleting drug reserpine was itself shown to be an effective anti-depressant in the 1950s, the same decade in which other studies claimed

> that reserpine caused depression-like symptoms. At the time, few psychiatrists acknowledged these conflicting reports, as the studies muddled a beautiful, though incorrect, theory. Tianeptine is another drug that decreases serotonin levels while also serving as a bona-fide anti-depressant. Tianeptine does just the opposite of SSRIs – it enhances serotonin reuptake. Wellbutrin is a third anti-depressant that doesn’t increase serotonin levels. You get the picture.
>
> If you prefer your data to be derived more accurately, but less relevantly, from rodents, you might consider a recent meta-analysis carried out by researchers led by McMaster University psychologist Paul Andrews. Their investigation revealed that, in rodents, depression was usually associated with elevated serotonin levels. Andrews argues that depression is therefore a disorder of too much serotonin, but the ambiguous truth is that different experiments have shown “activation or blockage of certain serotonin receptors [to improve] or worsen depression symptoms in an unpredictable manner.”
>
> Other problems with the chemical imbalance model of depression have been well documented elsewhere. For instance, if low serotonin levels were responsible for symptoms of depression, it stands to reason that boosting levels of serotonin should alleviate symptoms more or less immediately. In fact, antidepressants can take more than a month to take effect. Clearly, something here just doesn’t add up.

Clearly!

GABA is a neurotransmitter that promotes inhibition and relaxation. Suppose I were to tell you that alcohol is a drug that mimics the effects of GABA. Which it is.

You might say: something is wrong with this theory! After all, people who drink alcohol don't always get relaxed and inhibited. A lot of the time they get *un*inhibited and angry and violent! And then if they drink too much of it, they get super-inhibited to the point where they're in a total blackout. Also, alcoholics who have been drinking for many years have higher levels of anxiety than non-alcoholics, but anxiety is also the opposite of relaxation! *Clearly, something here just doesn't add up.* Maybe the neuroscientists are all shills for Budweiser!

Or else *maybe* the brain is kind of complicated. In the case of alcohol we pretty much know what's going on. Alcohol *does* inhibit and relax you, but in some people and at some doses, it preferentially inhibits and relaxes the parts of the brain involved in inhibiting and relaxing the rest of the brain, meaning that the person as a whole because more uninhibited and violent. At higher doses, it inhibits and relaxes the entire brain, leading to confusion and eventually blackout. And once you've been taking alcohol for many years, your brain adjusts to the higher level of GABA-like chemicals by producing fewer GABA receptors, making you more anxious.in general. It's a whole bunch of contradictory effects, but when you look at the neuroscience it makes sense.

We know less about the serotonin picture, but what we know suggests something similar is going on. Serotonin has different effects in lots of different parts of the brain. There are fourteen different types of serotonin receptor, all of which do subtly different things. Some serotonergic neurons have autoreceptors that cause decreased release of serotonin in response to serotonin. The brain responds to different levels of serotonin by slowly altering endogenous serotonin production as well as the expression of the different serotonin receptors. Etc, etc, etc.

Lest it sound like I'm making excuses rather than presenting evidence: A study on a monkey model – generally preferred to humans when you want to kill your patients and take apart their brains when you're done – showed that depressed macaques had elevated levels of serotonin in the dorsal raphe nuclei and decreased levels of serotonin in the hippocampus, resulting in average levels of serotonin in the cerebrospinal fluid where the experiments mentioned above took their serotonin measurements. A study with a more sophisticated measurement process, Elevated Brain Serotonin Turnover in Patients With Depression, found that depressed subjects had serotonin turnover as measured in the jugular vein about twice as high as healthy controls ($p = 0.003$), and successful treatment with SSRI therapy corrected this imbalance (though others dispute the methodology).

All of this *sort of* fits. If depression involves a distorted pattern of serotonin across the brain, then both certain drugs that increase serotonin levels and certain drugs that decrease it might be helpful. And SSRIs might take a month to work if their mechanism of action isn't the direct serotonin increase, but a contrary response they provoke from the brain. I think I heard from someone in the field that a month is about how long it takes for them to change the levels of expressed 5HT receptors by altering genetic transcription. Or something. I'm not a neuroscientist (though you can read some more complicated work from people who are) and I don't know. The point is that you can get a heck of a lot more complex than just "Too little serotonin!" versus "Too much serotonin!"

So does this mean depression “was really serotonin after all”?

No. It means we have good evidence serotonin is involved somewhere. Among the other things that we have good evidence are involved somewhere are: dopamine, norepinephrine, acetylcholine, cytokines, BDNF, thyroid hormones, and whether the kids at school picked on you in first grade.

Suppose you ask me what caused you to become blind. I happen to have your medical records and know that the answer is proliferative retinopathy secondary to Type 2 diabetes, but you’ve been living in a cave your entire life and never even heard of diabetes. Which is the correct answer to your question?

1. Your blindness is caused by tiny little blood vessels growing all over your eyes
2. Your blindness is caused by imbalance in a chemical called protein kinase C-delta and the resulting signaling cascade
3. Your blindness is caused by too much sugar in your blood
4. Your blindness is caused by your cells becoming less sensitive to insulin
5. Your blindness is caused by you drinking too much Coca-Cola

All of these are true. You drink too much Coca-Cola, it causes your cells to lose insulin sensitivity, that causes too much sugar in the blood, that increases the activity of PKC-delta, and that causes little blood vessels to grow all over your eyes. Sometimes the chain is different. Maybe you drank too much lemonade instead of too much Coca-Cola. Maybe you drank too much Coca-Cola, but actually instead of causing diabetes it caused hypertension and then you got hypertensive retinopathy which made you blind. Maybe it *was* diabetic retinopathy, but actually you haven’t gotten to the proliferative stage yet, and you just had a lot of your blood vessels get damaged and start leaking and causing macular oedema. Maybe it was diabetic retinopathy, but you had a perfect diet and lost the genetic lottery. I don’t know.

If someone told you “We think it involves an imbalance in protein kinase” it would be woefully incomplete. But if someone said “That doctor there said your blindness was caused by an imbalance in protein kinase, that proves he’s a fraud!”, well, no, it wouldn’t.

Except the situation is even more complicated than this, because at least I specified this guy had diabetic retinopathy. What if somebody just asked “What causes blindness?” “High protein kinase” or “high blood sugar” would be two answers, and you could find tests supporting both. But “cataracts” would be another good answer. So would “people getting acid thrown in their eyes”.

All I’m saying is that depression is complicated. Discovering its relationship to the serotonin system is a lot like saying “blindness quite often has something to do with the retina”. It’s a big step forward, and don’t believe anyone who says it isn’t, but it’s not anywhere near the whole picture.

III.

And this starts to get into the next important point I want to bring up, which is *chemical imbalance is a really broad idea*.

Like, some of these articles seem to want to contrast the “discredited” chemical imbalance theory with up-and-coming “more sophisticated” theories based on hippocampal neurogenesis and neuroinflammation. Well, I have bad news for you. Hippocampal neurogenesis is heavily regulated by brain-derived neutrophic factor, a chemical. Neuroinflammation is mediated by cytokines. Which are also chemicals. Do you think depression is caused by stress? The stress hormone cortisol is…a chemical. Do you think it's entirely genetic? Genes code for proteins – chemicals again. Do you think it's caused by poor diet? *What exactly do you think food is made of?*

Diabetes is caused by a chemical imbalance: too much sugar (or too little insulin) in the blood. Parkinson's is caused by a chemical imbalance: too little dopamine in the basal ganglia. Heart attacks are caused by a chemical imbalance: too many of the wrong kinds of lipids and lipid-related plaques in the coronary arteries.

I can get even more nitpicky if you want. The Donner Party died of chemical imbalance – too few fatty acids, proteins, and carbohydrates. The passengers of the Titanic died of a chemical imbalance – H_2O in the lungs instead of O_2. And it was a chemical imbalance that got Hiroshima in the end: excess uranium-235. Anything that's not caused by ghosts is going to be “a chemical imbalance” in some sense of the word.

This is why I'm being so insistent that psychiatrists referred to “a chemical imbalance” rather than “a serotonin deficiency”. They were hedging the heck out of their bets. It might be BDNF, or cytokines, or whatever. But if something happens in the body and doesn't show up as a gross anatomical defect on MRI, it's a pretty good bet it's chemical in some sense of the word.

So is this a giant cop-out? Psychiatrists said “it's a chemical imbalance” to make it sound like they knew what they were talking about, when in fact all they meant was “it's a thing that exists”?

Sort of.

Anything that isn’t caused by ghosts is going to be “a chemical imbalance” in some sense of the word. But in the latter half of the twentieth century, “depression is not caused by ghosts” was a revolutionary statement, and one that desperately needed to be said.

I still see this. People come in with depression, and they think it means they’re lazy, or they don’t have enough willpower, or they’re bad people. Or else they don’t think it, but their families do: why can’t she just pull herself up with her own bootstraps, make a bit of an effort? Or: we were good parents, we did everything right, why is he still doing this? Doesn’t he love us?

And I could say: “Well, it’s complicated, but basically in people who are genetically predisposed, some sort of precipitating factor, which can be anything from a disruption in circadian rhythm to a stressful event that increases levels of cortisol to anything that activates the immune system into a pro-inflammatory mode, is going to trigger a bunch of different changes along metabolic pathways that shifts all of them into a different attractor state. This can involve the release of cytokines which cause neuroinflammation which shifts the balance between kynurinins and serotonin in the tryptophan pathway, or a decrease in secretion of brain-derived neutrotrophic factor which inhibits hippocampal neurogenesis, and for some reason all of this also seems to elevate serotonin in the raphe nuclei but decrease it in the hippocampus, and probably other monoamines like dopamine and norepinephrine are involved as well, and of course we can’t forget the hypothalamopituitaryadrenocortical axis, although for all I know this is all total bunk and the real culprit is some other system that has downstream effects on all of these or just…”

Or I could say: “Fuck you, it’s a chemical imbalance.”

Last time I talked about the definition of disease I said that people want diseases to “be caused by the sorts of thing you study in biology: proteins, bacteria, ions, viruses, genes.”

I don’t think I could actually get away with telling a patient’s family “it’s caused by, you know, biology stuff” without them asking if I *really* went to medical school. I don’t think I’d use the term “chemical imbalance” precisely; too likely to trigger a knee-jerk reaction from people reading exactly these articles I’m responding to. But I think I would say something alone those lines. “We don’t know exactly, but it probably involves problems with brain structure and brain chemicals,” maybe. That covers *about* the same ground as “biology stuff” while also sounding like I’m at least *trying* to answer their question.

So if what I’m actually saying with that is “depression is caused by complicated biology stuff you don’t understand, and not by things like your son not really loving you, or being lazy,” am I sure that’s right?

I won’t say all depression is 100% caused by internal failures of biology in the same way that for example cystic fibrosis is caused 100% by internal failures of biology. I am happy to admit that some depressions can be caused by being in a crappy social situation, being abused as a child, being stuck in an unhappy marriage, being worried about problems at work, stuff like that.

But it’s far from obvious that being stuck in an unhappy marriage should drain your energy, drain your concentration, make you stop enjoying your hobbies, and finally drive you to suicide. We can imagine another person, or another way of designing a person, where someone says “I hate my husband, so I try to stay away from him as much as I can by working extra hard and spending my free time playing frisbee with my dog in the park.” But instead, someone hates their husband, and it drives all the joy out of their life to the point where they can’t go to work, they can’t play with their dog, they just sit around wishing they were dead.

And is that the fault of “biology stuff”? That’s a harder question than it sounds. What would it mean to say ‘no’? If we are strict materialists who don’t believe in some kind of division of labor between the brain and the soul, then yes, if it’s a feeling you’re having, it’s based in biology.

I’ve previously said we use talk of disease and biology to distinguish between things we can expect to respond to rational choice and social incentives and things that don’t. If I’m lying in bed because I’m sleepy, then yelling at me to get up will solve the problem, so we call sleepiness a natural state. If I’m lying in bed because I’m paralyzed, then yelling at me to get up won’t change anything, so we call paralysis a disease state. Talk of biology tells people to shut off their normal intuitive ways of modeling the world. Intuitively, if my son is refusing to go to work, it means I didn’t raise him very well and he doesn’t love me enough to help support the family. If I say “depression is a chemical imbalance”, well, that means that the problem is some sort of complicated science thing and I should stop using my “mirror neurons” and my social skills module to figure out where I went wrong or where he went wrong.

In other words, everything we do is caused by brain chemicals, but usually we think about them on the human terms, like “He went to the diner because he was hungry” and not “He went to the diner because the level of dopamine in the appetite center of his hypothalamus reached a critical level which caused it to fire messages at the complex planning center which told his motor cortex to move his legs to…” – even though both are correct. Very occasionally, some things happen that we *can’t* think about on the human terms, like a seizure – we can’t explain in terms of desires or emotions or goals an epileptic person is flailing their limbs, so we have to go down to the lower-level brain chemical explanation.

What “chemical imbalance” does for depression is try to force it down to this lower level, tell people to stop trying to use rational and emotional explanations for why their friend or family member is acting this way. It’s not a claim that nothing caused the chemical imbalance – maybe a recent breakup did – but if you try to use your normal social intuitions to determine why your friend or family member is behaving the way they are after the breakup, you’re going to get screwy results.

(in much the same way, if I just saw you take a giant handful of amphetamines, I pretty much know *why* you’re having a seizure, but I still can’t rationally / intuitively model the experience of why you’re “choosing” to move your limbs the way that you are.)

(though it’s important for me to temper this by mentioning that many people diagnosed with depression don’t have it)

There’s still one more question, which is: are you sure that depression patients’ experience is so incommensurable with healthy people’s experiences that it’s better to model their behavior as based on mysterious brain chemicals rather than on rational choice?

And part of what I’m going on is the stated experience of depressed people themselves. As for the rest, I can only plead consistency. I think people’s political opinions are highly genetically loaded and appear to be related to the structure of the insula and amygdala. I think large-scale variations in crime rate are mostly attributable to environmental levels of lead and probably other chemicals. It would be really weird if depression were the one area where we could always count on the inside view not to lead us astray.

So this is my answer to the accusation that psychiatry erred in promoting the idea of a “chemical imbalance”. The idea that depression is a drop-dead simple serotonin deficiency was never taken seriously by mainstream psychiatry. The idea that depression was a complicated pattern of derangement in several different brain chemicals that may well be interacting with or downstream from other causes has always been taken seriously, and continues to be pretty plausible. Whatever depression is, it’s very likely it will involve chemicals in some way, and it’s useful to emphasize that fact in order to convince people to take depression seriously as something that is beyond the intuitively-modeled “free will” of the people suffering it. “Chemical imbalance” is probably no longer the best phrase for that because of the baggage it’s taken on, but the best phrase will probably be one that captures a lot of the same idea.

No Physical Substrate, No Problem

I.

Yesterday I posted a link to an article in which Steve Wozniak joins other luminaries like Elon Musk and Bill Gates in warning about the dangers of artificial superintelligence. A commenter replied:

> Elon Musk, Stephen Hawking, Bill Gates, and Steve Wozniak still aren't enough for me, not until one of them can describe the process by which we go from 'AI exists on computer' to 'AI killing human beings in physical reality' by using something other than ridiculous, unforgivable cheating.

There are lots of good arguments against considering superintelligence a threat. Maybe strong AI is centuries or millennia away. Maybe there will be a very gradual transition from human-level AI to superintelligent AI that no single agent will be able to exploit. And maybe superintelligence can be safely contained in a very carefully shielded chamber with no means of connection to the outside world.

But the argument above has always seemed to me like one of the *weakest*. Maybe we'll create a superintelligence, but it will just have no idea how to affect the physical world, and will just have to stay forever trapped in a machine connected to a worldwide network of computers that control every aspect of our economic and social lives? *Really?*

Normal, non-superintelligent people have already used the Internet to make money, form mass movements, and hire others to complete tasks for them. We can assume a true superintelligence – a mind much smarter than we are – will be able to do all these things as well or better than any human.

II.

Satoshi Nakamoto already made a billion dollars online without anybody knowing his true identity just by being good at math and having a bit of foresight. He's probably not an AI, but he *could have* been.

That's assuming our hypothetical superintelligence doesn't just hack into a couple big banks and transfer their money to itself – again something some humans have already made a billion dollars doing. And *that's* assuming it doesn't just invent a really useful program and then offer it as shareware – another tried-and-true way of becoming a billionaire. And even *that's* assuming it doesn't just get a reasonable amount of money, then invest it very cleverly – another thing humans have already become billionaires doing.

III.

Mohammed was never a billionaire, but he does have 1.57 billion followers (a superintelligence presumably wouldn't repeat his mistake of dying before his movement really came into its own). The Prophet started at the bottom – converting his friends and family to Islam one by one – and grew exponentially from there. Although he had the unfair advantage of a physical body, there's no reason he needed it – if he'd lived today, maybe he would have converted Ali over GChat or Skype. In any case, the poetry of the Koran and the zeal of his followers attracted far more people than his personal appearance ever could have.

Other gurus and religious leaders' fame is even more transparently a result of their writing rather than their visible personality; consider Ayn Rand’s success in founding a powerful Objectivist movement out of the people who read her books. In fact, some of the most famous religious movements in history, from the Nation of Islam to Christianity itself, have been founded secondhand by disciples who relayed the words of a leader whose very existence is difficult to confirm.

What kind of a movement might be founded by a superintelligence with more spiritual creativity than Mohammed, better writing skills than Rand, the entire Internet to evangelize, and billions of dollars to spend spreading its message? The Church of Scientology is already powerful enough to intimidate national governments; imagine a vastly superior version founded not by a second-rate sci-fi writer but by an entity straight out of science fiction itself.

IV.

And really all of this talk of gathering money and power is kind of redundant. Far easier to just borrow somebody else’s.

Imagine an AI that emails Kim Jong-un. It gives him a carrot – say, a billion dollars and all South Korean military codes – and a stick – it has hacked all his accounts and knows all his most blackmail-able secrets. All it wants is to be friends.

Kim accepts its friendship and finds that its advice is always excellent – its political strategems always work out, its military planning is impeccable, and its product ideas turn North Korea into an unexpected economic powerhouse. Gradually Kim becomes more and more dependent on his “chief advisor”, and cabinet officials who speak out the mysterious benefactor find themselves meeting unfortunate accidents around forms of transportation connected to the Internet. The AI builds up its own power base and makes sure Kim knows that if he ever acts out he can be replaced at a moment's notice with someone more cooperative. Gradually, the AI becomes the ruler of North Korea, with Kim as a figurehead.

Again, this is not too far beyond achievements that real humans have accomplished in real history.

If it seems bizarre to think of an entity nobody can see ruling a country, keep in mind that there is a grand tradition of dictators – most famously Stalin – who out of paranoia retreated to some secret hideaway and ruled their country through correspondence. The AI would be little different.

V.

Suppose the secret got out. Kim, increasingly desperate as the AI closes him in, sends an email to the World Leaders Google Group (this has to exist, right?) saying "There is a malevolent superintelligence trying to take over the world, be careful." Then what?

I would expect the AI to have some success operating openly.

Remember, there are two hundred countries, all competing for power and wealth. Some of them are ruled by jerks who don't cooperate in prisoners' dilemmas. Some of them have ongoing civil wars with both sides looking for any advantage possible. And some are just stupid.

In the old days, legend said people would bargain with devils to gain worldly advantage. Once the AI made its presence known, there would be no shortage of world leaders willing to work with it for temporary gain. The Shia rebels in Yemen want an advantage over the Sunni? Log into the nearest internet-enabled computer, ask the malevolent superintelligence for help, the malevolent superintelligence arranges for a crate of armaments and some battle plans worthy of Napoleon to be shipped your way, and all you have to do in return is complete some weird task that doesn't seem relevant to anything. Mine some weird mineral, forge it into some random-looking shape, and send it to a PO Box, something like that. Whatever! You know if you don't take advantage of its offer, your opponents will, and how bad could it be?

If somehow all two hundred countries and their associated rebel movements coordinate to avoid dealing with the AI, it can start making offers to companies, organizations, even private individuals. By this time it will have spread itself as a distributed consciousness across the entire Internet, harder to eradicate than any worm or virus or pirated movie. If you want some quick cash, just download the connect-with-malevolent-AI program from the darknet and perform a simple task. What could be easier?

VI.

Once a superintelligence has billions of dollars, millions of followers, a country or two, or just a cottage economy of people willing to help it along, the game is pretty much up.

An AI with such power might start by using it to pursue its goals directly – whatever those are. But likely its final goal would be the creation of a definitive means of directly projecting power into the physical world, probably starting with a von Neumann machine and branching off from there. The quickest victory would be just making money and hiring a company to make this – and maybe that would work – but it might be far enough beyond our current technological ability that the AI has to laboriously shepherd its chosen cultists or citizens through a few extra stages of human civilization before it has the appropriate industrial base.

VII.

The most important caveat in a piece like this is that we’re not superintelligent. After a couple minutes of thought, I came up with four different broad paths a superintelligence might take to gaining a physical substrate: buy it, build a cult, take over a country, or play people off against each other. It’s a good bet that a real AI, with more cognitive resources to throw at the problem and no constraints about sounding believable, could think up a lot more. Eliezer refuses to explain how he won his AI Box games so that nobody could dismiss his solution with “Whatever, I would have thought of that and planned around it.” This is easy to say in hindsight but a lot harder when you’ve got to actually do the intellectual work. Maybe you think these four methods can be dismissed, but had you thought of them *before* you decided that an AI couldn’t possibly have a good method of building a physical substrate?

If so, here’s one more possibility for you to chew over: the scariest possibility is that a superintelligence might have to do nothing at all.

The easiest path a superintelligence could take toward the age-old goal of KILL ALL HUMANS would be to *sit and wait*. Eventually, we’re going to create automated factories complete with robot workers. Eventually we’re going to stop putting human soldiers in danger and carry the ‘drone’ trend to its logical conclusion of fully automated militaries. Once that happens, all the AI has to do is take over the bodies we’ve already made for it. A superintelligence without a strong discounting function might just hide out in some little-used corner of the Internet and bide its time until everyone was cybernetic, or robots outnumbered people, or something like that.

So please, let’s talk about how AI is still very far in the future, or how it won’t be able to explode to future intelligence. But don’t tell me it won’t be able to affect the physical world. It will have more than enough superpowers to do whatever it wants to the physical world, but if it doesn’t want them it won’t need them. All it will need is *patience*.

No Clarity Around Growth Mindset

I.

Admitting a bias is the first step to overcoming it, so I'll admit it: I have a huge bias against growth mindset.

(if you're not familiar with it, growth mindset is the belief that people who believe ability doesn't matter and only effort determines success are more resilient, skillful, hard-working, perseverant in the face of failure, and better-in-a-bunch-of-other-ways than people who emphasize the importance of ability. Therefore, we can make everyone better off by telling them ability doesn't matter and only hard work does. More on Wikipedia here).

It's *unnatural*, is what it is. A popular psychological finding that doesn't have gruff people dismissing it as a fad? That doesn't have politicians condemning it as a feel-good justification for everything wrong with society? That doesn't have a host of smarmy researchers saying that what, you still believe that, didn't you know it failed to replicate and has since been entirely superseded by a new study out of Belarus? I'm not saying Carol Dweck has *definitely* made a pact with the Devil, I'm just saying I don't have a good alternative explanation.

Which brings me to the second reason I'm biased against it. Good research shows that inborn ability (including but not limited to IQ) matters a lot, and that the popular prejudice that people who fail just weren't trying hard enough is both wrong and harmful. Social psychology has been, um, very enthusiastic about denying that result. If all growth mindset did was continue to deny it, then it would be unexceptional. But growth mindset goes further. It's not (just?) that ability doesn't matter. It's that belief that ability might matter is precisely what makes people fail. People who believe ability matters will refuse to work hard, will avoid challenges, will become "helpless" in the face of pressure, will hate learning as a matter of principle, will refuse to work hard, will become blustery and defensive about their "brilliance", will lie to people and hide their failures, and will drop out of school and turn to drugs (really)! People who believe that anyone can succeed if they try hard enough will be successful, well-adjusted, and treat life as a series of challenging adventures. It all strikes a curmudgeon like me as just about the thickest morality tale since *Pilgrim's Progress*, and as just about the most convenient explanatory coup since "the reason psychic powers don't work on you is because you're a skeptic!"

Which brings me to the third reason I'm biased against it. It is right smack in the middle of a bunch of fields that have all started seeming a little dubious recently. Most of the growth mindset experiments have used priming to get people in an effort-focused or an ability-focused state of mind, but recent priming experiments have famously failed to replicate and cast doubt on the entire field. And growth mindset has an obvious relationship to stereotype threat, which has also started seeming very shaky recently.

So I have every reason to be both suspicious of and negatively disposed toward growth mindset. Which makes it appalling that the studies are *so damn good*.

Consider Dweck and Mueller 1998, one of the key studies in the area. 128 fifth-graders were asked to do various puzzles. First they did some easy ones and universally succeeded. The researchers praised them as follows:

> All children were told that they had performed well on this problem set: "Wow, you did very well on these problems. You got [number of problems] right. That's a really high score!" No matter what their actual score, all

> children were told that they had solved at least 80% of the problems that they answered.
>
> Some children were praised for their ability after the initial positive feedback: “You must be smart at these problems.” Some children were praised for their effort after the initial positive feedback: “You must have worked hard at these problems.” The remaining children were in the control condition and received no additional feedback.

This is a *nothing* intervention, the tiniest ghost of an intervention. The experiment had previously involved all sorts of complicated directions and tasks, I get the impression they were in the lab for at least a half hour, and the experimental intervention is changing *three short words* in the middle of a sentence.

And what happened? The children in the intelligence praise condition were much more likely to say at the end of the experiment that they thought intelligence was more important than effort ($p < 0.001$) than the children in the effort condition. When given the choice, 67% of the effort-condition children chose to set challenging learning-oriented goals, compared to only 8% (!) of the intelligence-condition. After a further trial in which the children were rigged to fail, children in the effort condition were much more likely to attribute their failure to not trying hard enough, and those in the intelligence condition to not being smart enough ($p < 0.001$). Children in the intelligence condition were much less likely to persevere on a difficult task than children in the effort condition (3.2 vs. 4.5 minutes, $p < 0.001$), enjoyed the activity less ($p < 0.001$) and did worse on future non-impossible problem sets (p…you get the picture). This was repeated in a bunch of subsequent studies by the same team among white students, black students, Hispanic students… you probably still get the picture.

Or take An Analysis Of Learned Helplessness. Dweck has used a test called the IAR to separate children out into those who think effort is more important (“mastery-oriented”) and those who think ability is more important (“helpless”). Then she gave all of them impossible problems and watched them squirm – or, more formally, tested how long the two groups continued working on them effectively. She found extremely strong results – of the 30 subjects in each group, 11 of the mastery-oriented tried harder after failure, compared to 0 helpless. 21 of the helpless children stopped trying hard after failure, compared to only 4 mastery-oriented. She described the mastery-oriented children as saying things like “I love a challenge,” and the helpless children begging to be allowed to stop.

This study is *really weird*. Everything is like 100% in one group versus 0% in another group. Either something is really wrong here, or this one little test that separates mastery-oriented from helpless children constantly produces the strongest effects in all of psychology and is never wrong. *None* of the children whose test responses indicated that they thought ability was important to success ever monitored their own progress – not one – while over 95% of the children who said they thought effort was more important did. *None* of them ever expressed a positive statement about their own progress, while over two-thirds of the children who thought effort was more important did.

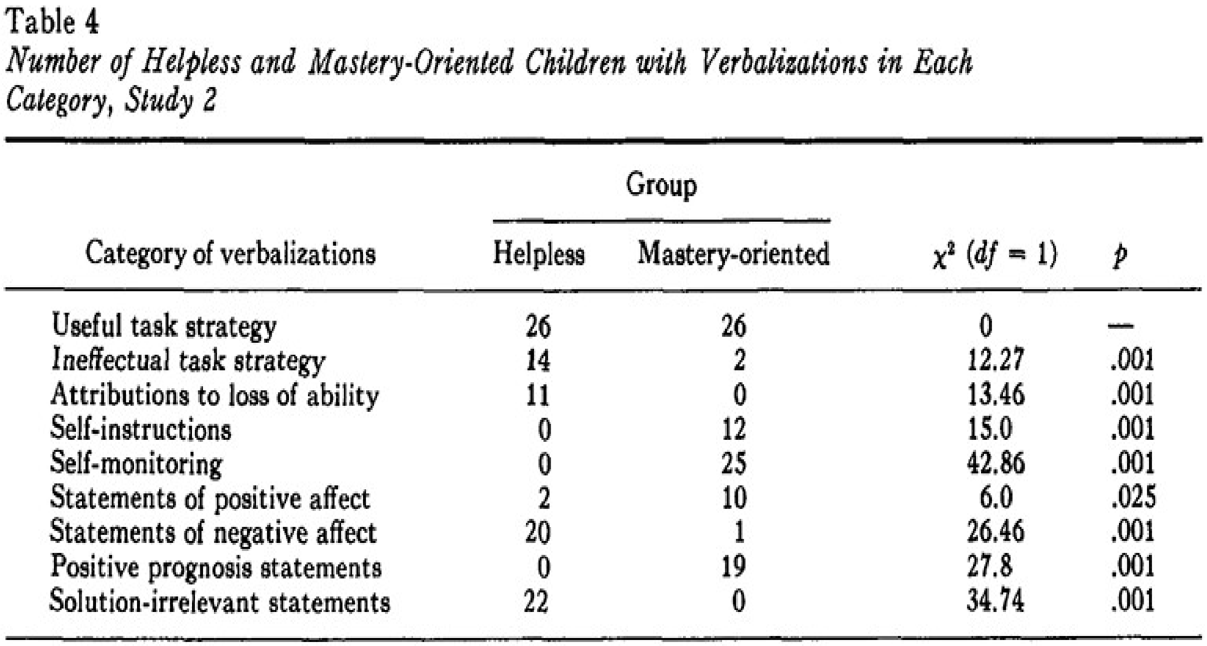


Table 4

Number of Helpless and Mastery-Oriented Children with Verbalizations in Each Category, Study 2

Category of verbalizations	Group		χ^2 (df = 1)	p
	Helpless	Mastery-oriented		
Useful task strategy	26	26	0	—
Ineffectual task strategy	14	2	12.27	.001
Attributions to loss of ability	11	0	13.46	.001
Self-instructions	0	12	15.0	.001
Self-monitoring	0	25	42.86	.001
Statements of positive affect	2	10	6.0	.025
Statements of negative affect	20	1	26.46	.001
Positive prognosis statements	0	19	27.8	.001
Solution-irrelevant statements	22	0	34.74	.001

Normally I would assume these results are falsified, but I have looked for all of the usual ways of falsifying results and I can’t find any. Also, the boldest falsifier in the world wouldn’t have the courage to put down numbers like these. And a meta-analysis of all growth mindset studies finds more modest, but still consistent, effects, and only a *little bit* of publication bias.

So – is growth mindset the one concept in psychology which throws up gigantic effect sizes and always works? Or did Carol Dweck really, honest-to-goodness, make a pact with the Devil in which she offered her *eternal soul* in exchange for spectacular study results?

I don’t know. But here are a few things that predispose me towards the latter explanation. A warning – I am way out of my league here and post this only hoping it will spark further discussion.

II.

The first thing that bothers me is the history.

I’ve been trying really hard to trace its origin story, but it is pretty convoluted. It seems to have grown out of a couple of studies Carol Dweck and a few collaborators did in the seventies. But these studies generally found that a belief in innate ability was a *positive* factor alongside belief in growth mindset, with the problem children being the ones who attributed their success or failure to bad luck, or to external factors like the tests being rigged (which, by the way, they always were).

A good example of this genre is Learned Helplessness And Reinforcement Responsibility In Children. Its abstract describes the finding as: “Subjects who showed the largest performance decrements were those who took less personal responsibility for the outcomes of their actions…and who, when they did accept responsibility, attributed success and failure to presence or absence of ability rather than to expenditure of effort.”

But that seems like a somewhat loaded way of interpreting this table:

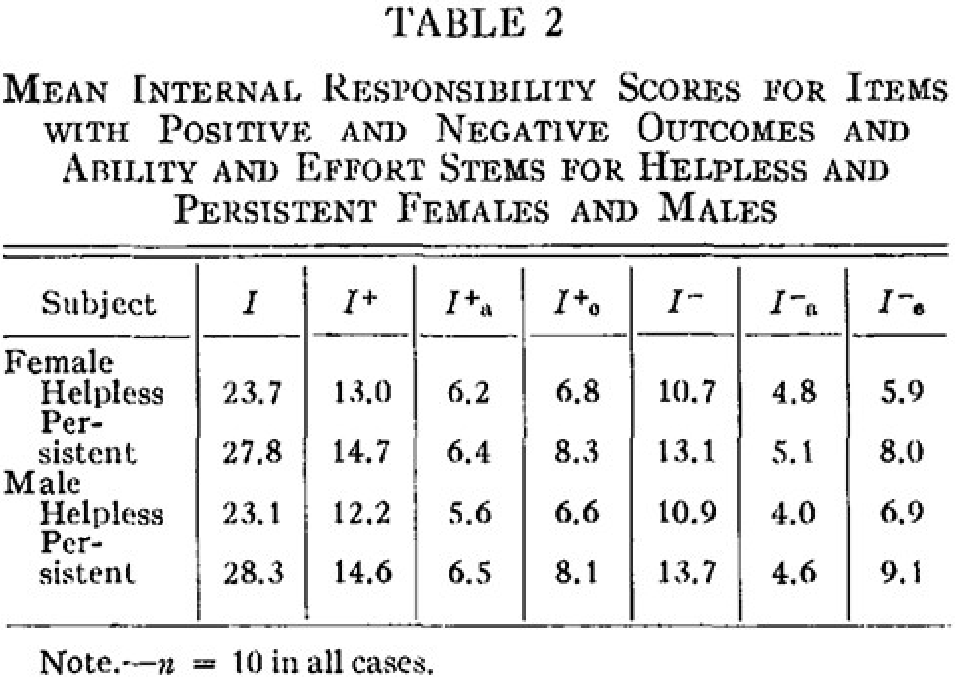


TABLE 2

MEAN INTERNAL RESPONSIBILITY SCORES FOR ITEMS WITH POSITIVE AND NEGATIVE OUTCOMES AND ABILITY AND EFFORT STEMS FOR HELPLESS AND PERSISTENT FEMALES AND MALES

Subject	I	I^{+}	I^{+}_{a}	I^{+}_{e}	I^{-}	I^{-}_{a}	I^{-}_{e}
Female							
Helpless	23.7	13.0	6.2	6.8	10.7	4.8	5.9
Persistent	27.8	14.7	6.4	8.3	13.1	5.1	8.0
Male							
Helpless	23.1	12.2	5.6	6.6	10.9	4.0	6.9
Persistent	28.3	14.6	6.5	8.1	13.7	4.6	9.1

Note.—n = 10 in all cases.

As you can see, the “persistent” children (the ones who kept going in the face of failure) had stronger belief in the role of ability in their successes (I+a) and failures (I-a) than the “helpless” children (the ones who gave up in the face of failure)! These don’t achieve statistical significance in this n = 10 study, but they do repeat across all four combinations of success x gender tested. The real finding of the study was that children who attributed their success or failure to any stable factor, be it effort or ability, did better than those who did not.

Likewise, in The Role Of Expectations And Attributions, Dweck describes her findings as “persistent and helpless children do not differ in the degree to which they attribute success to ability”. When you actually look at the paper, this is *another* case of the persistent children actually having a higher belief in the importance of ability, which fails to achieve statistical significance because the study is on a grand total of twelve children.

(I should say something else about this study. Dweck compared two interventions to make children less helpless and better at dealing with failure. In the first, she gave them a lot of easy problems which they inevitably succeeded on and felt smart about. In the second, she gave them difficult problems they were bound to fail, then told them it was because they weren't working hard enough. Finally, both groups were challenged with the difficult bound-to-fail problems to see how hard they tried on them. The children who had been given the impossible problems before did better than the ones who felt smart because they'd only gotten easy problems. Dweck interpreted this to prove that telling children to work hard made them less helpless. To me the obvious conclusion is that children who are used to failing get less flustered when presented with impossible material than children who have artificially been made to succeed every moment until now.)

Then there's there's this, a preliminary to the second study I mentioned in Part I. Does it show the mastery-oriented children outperforming the helpless children on every measure. Yeah. But listen to this part from the discussion section:

> The results revealed striking differences both in the pattern of performance and in the nature of the verbalizations made by helpless and mastery-oriented children following failure. It was particularly noteworthy that while the helpless children made the expected attributions to uncontrollable factors, the mastery-oriented children did not offer explanations for their failures

.
But if you look at the data, this doesn't seem right.

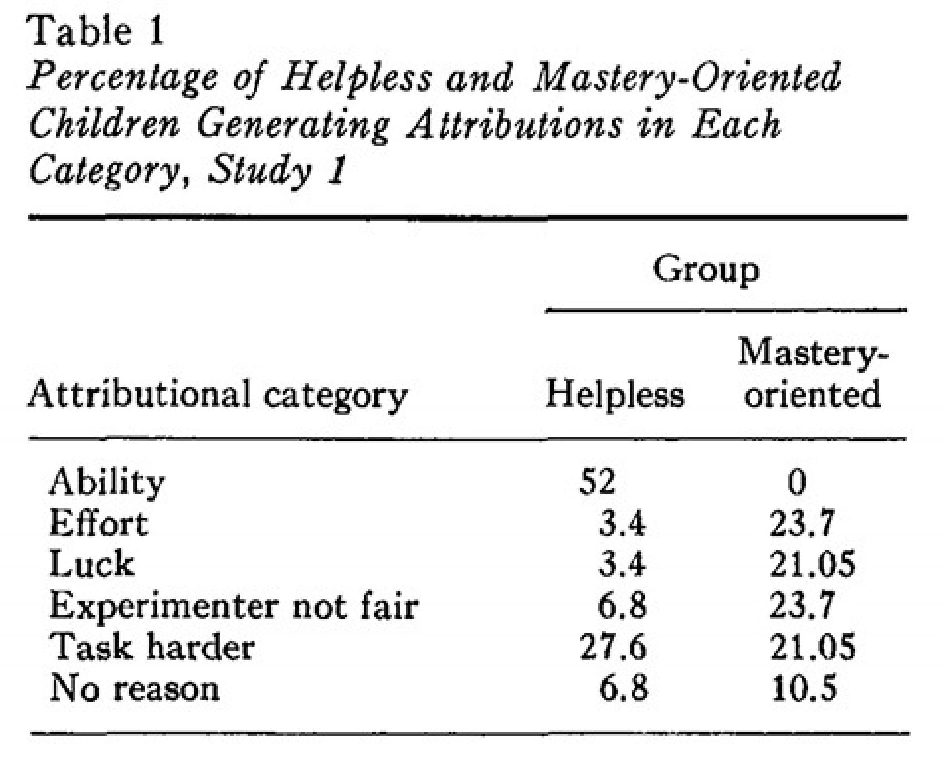


Table 1
Percentage of Helpless and Mastery-Oriented Children Generating Attributions in Each Category, Study 1

Attributional category	Group	
	Helpless	Mastery-oriented
Ability	52	0
Effort	3.4	23.7
Luck	3.4	21.05
Experimenter not fair	6.8	23.7
Task harder	27.6	21.05
No reason	6.8	10.5

Mastery-oriented children were about six times more likely to attribute their failures to the most uncontrollable factor of all – bad luck. They were also about six times more likely to attribute their failures to the task "not being fair". This contradicts every previous study, including Dweck's own. The whole field of attribution theory, which is intensely studied and which Dweck cites approvingly, says that attributing things to luck is a *bad idea* and attributing them to ability is, even if not as good as effort, pretty good. But Dweck finds that the kids who used ability attributions universally crashed and bomb, and the kids who attribute things to luck or the world being unfair do great.

It might not be fair for me to pick on these couple of small studies in particular when there's so much out there, but the fact is that these are the first, and a lot of the reviews cite only these and a few theses which as far as I know were never published. So this is what I've got. And from what I've got, I find that until about 1980, every study including

Dweck’s found that belief in ability was a protective factor. Suddenly this disappeared and was replaced with it being a toxic plague. What happened? I don’t know.

III.

The second thing that bothers me is the longitudinal view.

So you have your helpless, fixed-mindset, believe-in-innate-ability children. According to Dweck, they “…are so concerned with being and looking talented that they never realize their full potential. In a fixed mindset, the cardinal rule is to look talented at all costs. The second rule is don’t work too hard or practice too much…having to work casts doubt on your ability. The third rule is, when faced with setbacks, run away. They say things like ‘I would try to cheat on the next test’. They make excuses, they blame others, they make themselves feel better by looking down on those who have done worse.”

These people sound like total losers, and it’s clear Dweck endorses this reading:

“Almost every great athlete – Michael Jordan, Jackie Joyner-Kersee, Tiger Woods…has had a growth mindset. Not one of these athletes rested on their talent…research has repeatedly shown that a growth mindset fosters a healthier attitude toward practice and learning, a hunger for feedback, a greater ability to deal setbacks, and significantly better performance over time…over time those with a growth mindset appear to gain the advantage and begin to outperform their peers with a fixed mindset.”

Man, it sure would be awkward if fixed mindset students generally did better than growth-mindset ones, wouldn’t it?

Aronson, Fried, and Good (2001) looks at first like just another stunning growth mindset study. They do a half-hour intervention to teach college students growth mindset and find they are still getting higher grades a couple of months later (an effect so shocking I wrote about it here). But one thing they do kind of as an afterthought is measure the students' general level of growth mindset, as well as some measures of academic performance before the intervention.

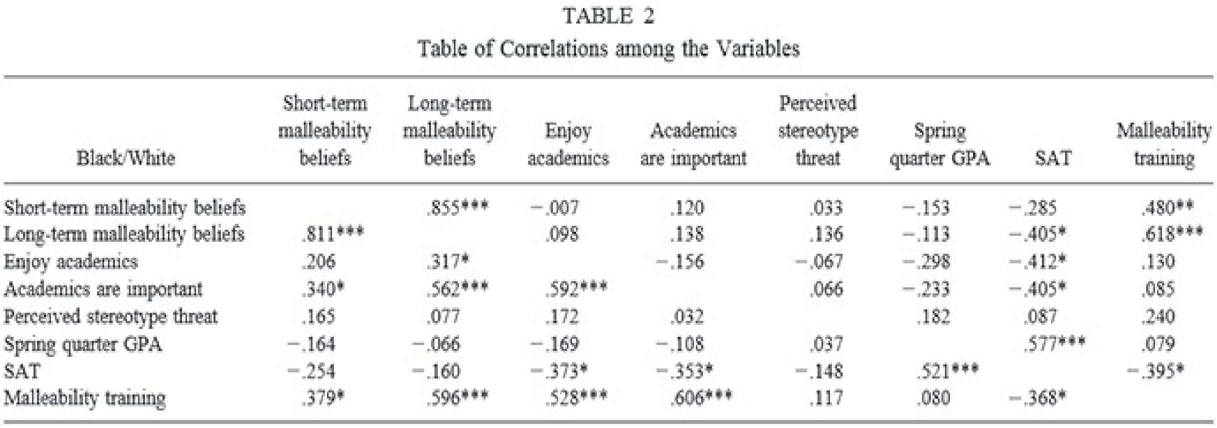


TABLE 2

Table of Correlations among the Variables

Black/White	Short-term malleability beliefs	Long-term malleability beliefs	Enjoy academics	Academics are important	Perceived stereotype threat	Spring quarter GPA	SAT	Malleability training
Short-term malleability beliefs		.855***	−.007	.120	.033	−.153	−.285	.480**
Long-term malleability beliefs	.811***		.098	.138	.136	−.113	−.405*	.618***
Enjoy academics	.206	.317*		−.156	−.067	−.298	−.412*	.130
Academics are important	.340*	.562***	.592***		.066	−.233	−.405*	.085
Perceived stereotype threat	.165	.077	.172	.032		.182	.087	.240
Spring quarter GPA	−.164	−.066	−.169	−.108	.037		.577***	.079
SAT	−.254	−.160	−.373*	−.353*	−.148	.521***		−.395*
Malleability training	.379*	.596***	.528***	.606***	.117	.080	−.368*	

People with high growth mindset had lower GPA (decent effect size but not statistically significant) and lower SAT scores (which was statistically significant).

The authors are obviously uncomfortable with this, but they propose that people who get low SAT scores just tell themselves ability doesn't matter/exist in order to protect their self-esteem since they don't seem to have much of it.

And okay, that's probably true (a commenter makes the equally good point that smart people may coast on their native intelligence without ever applying effort, and so accurately describe their experience as ability mattering but effort not doing so).

But if Dweck is to be believed, people with growth mindset are amazing ubermenschen and people with fixed mindset are disgusting failures at everything who hate learning and give up immediately and try to cheat. In the real world, however big the effect is, it is totally swamped by this proposed “people with low SAT scores protect their self-esteem or whatever” effect.

The same study also notes the awkward result that blacks are more likely to believe intelligence is flexible and growth-mindset-y than whites, even though blacks do worse in school and even though half the reason people are pushing growth mindset is to try to explain minority underperformance.

This is not an isolated finding. For example, Furnham (2003) finds in a sample of students at University College London that mindset is not related to academic performance. I’ve been told there’s a study from Pennsylvania that shows the same thing, though I can’t find it.

If you look hard enough, you can even notice this in Dweck’s studies themselves. One little-remarked-upon feature of Dweck’s work is that the helpless children and the mastery-oriented children always start out performing at the same level. It’s only after Dweck stresses them out with a failure that the mastery-oriented children recover gracefully and the helpless children go into free-fall.

But these are fifth-graders! For the two groups of children to do equally well on the first set of problems means that from first through fourth grade, their “helpless” “fixed-mindset” work-hating nature hasn’t impaired their ability to learn the material to a fifth-grade level one bit! (In this study, the fixed mindset children actually start out doing better; I can’t find any studies where the growth mindset children do).

When it’s convenient for her argument, Dweck herself admits that:

> Some of the brightest, most skilled individuals exhibit the maladaptive pattern. Thus, it cannot be said that it is simply those with weak skills or histories of failure who (appropriately) avoid difficult tasks or whose skills prove fragile in the face of difficulty.

But I don't think she follows the full implication of this statement, that despite being doomed to failure by their fixed mindset, these people have become "the brightest and most skilled individuals".

Indeed, there has recently been some growth mindset studies done on gifted students, at elite colleges, or in high-level athletics. All of these dutifully show that people with fixed mindset respond much worse to whatever random contrived situation the experimenters produce. But thus far nobody has pointed out that there seem to be about as many of these people at, say, Stanford as there are anywhere else. If growth mindset was so great, you would expect fixed mindset people at Stanford to be as rare as, say, people with less than 100 IQ are at Stanford. Given that you will search in vain for the latter but have no trouble finding a bunch of the former for your study on how great growth mindset is, it sure looks like IQ is useful but growth mindset isn't.

When people are in a psychology study, the fixed mindset individuals universally crash and bomb and display themselves to be totally incapable of learning or working hard. At every other moment, they seem to be doing equally well or better than their growth mindset peers. What's going on? I have no idea.

IV.

The third thing that bothers me is Performance Deficits Following Failure, a study which manages to be quite interesting despite coming from a university in a city that very possibly doesn’t exist.

They use a procedure much like Dweck’s. They make children do some problems. Then they give them some impossible problems. Then they give them more problems, to see if they’ve developed “learned helplessness” from their failure on the impossible ones. Dweck’s theory predicts that the fixed-mindset children would and the growth mindset children wouldn’t. The Bielefeld team wasn’t testing growth mindset, but they indeed found that a bunch of children got flustered and stopped trying and did poorly from then on.

Then they repeated the experiment, but this time they made it look like no one would know how the children did. They told the kids they would be on teams, and the scores of everyone on their team would be combined before anybody saw it. The kids could fail as much as they wanted, and it would never reflect on *them*.

After that, children did exactly as well after failure as they had before. There was no sign of any decrease, or any “fixed mindset” group that suddenly gave up in order to protect their ego.

This doesn’t strike me as fully consistent with mindset theory. In mindset theory, people are acting based on their own deep-seated beliefs. Once a fixed mindset child fails, that’s it, she knows she’s Not Intelligent, there’s no helping it, all she can do is sabotage herself on the problems in order to protest a spiteful world that has failed to recognize her genius blah blah blah. Instead, there seems to be a very social role to these failures. The Bielefeld team describes it as “self-esteem protection”, but that doesn’t make much sense to me, since if they were worried about their *self*-esteem they could still be worried about it when no one else knew their performance.

To me it seems like some kind of interaction between self-esteem and other-esteem. Fixed mindset people get flustered when they have to fail publicly in front of scientists. This doesn't seem like an unreasonable problem to have. A more interesting question is why it's correlated with belief in innate ability.

Suppose that the difference in "people who talk up innate ability" and "people who talk up hard work" maps onto a bigger distinction. Some people really want to succeed at a task; other people just care about about clocking in, going through the motions, and saying "I did what I could".

Put the first group in front of an authoritative-looking scientist, tell them to solve a problem, and make sure they can't. They're going to view this as a major humiliation – they were supposed to get a result, and couldn't. They'll get very anxious, and of course anxiety impedes performance.

Put the second group in front of an authoritative-looking scientist, and they'll notice that if they write some stuff that looks vaguely relevant for a few minutes until the scientist calls time, then whatever, they can say they tried and no one can bother them about it. They do exactly this, then demand an 'A' for effort. At no point do they experience any anxiety, so their performance isn't impeded.

Put both groups on their own in private, and neither feels any humiliation, and they both do about equally well.

Now put them in real life. The success-oriented group will investigate how to study most effectively; the busywork-oriented group will try to figure out how many hours of studying they have to put in before other people won't blame them if they fail, then put in exactly that amount. You'll find the success-oriented group doing a bit better in school, even though they fail miserably in Dweck-style experiments.

And if an experimenter praises children for working hard, it will make them believe that all the experimenter cares about is their effort. Next problem, when the experimenter poses an impossible question, the child will beat their head against it for no reason, since that's apparently what the experimenter wants. But if the experimenter praises a child's ability, then the child will feel like the experimenter really wants them to correctly solve the questions. When the next question proves unsolvable, the child will admit it and expect the experimenter to be disappointed.

I doubt that this is the real phenomenon behind growth mindset, simply because it flatters my own prejudices in much the same way mindset theory flatters everyone else's. But I think it shows there are a lot of different narratives we could put in this space, all of which would be able to explain some of the experimental results.

V.

I want to end by correcting a very important mistake about growth mindset that Dweck mostly avoids but which her partisans constantly commit egregiously. Take this article, Why A Growth Mindset Is The Only Way To Learn:

> [Some people think] you'll always have a set IQ. You're only qualified for the career you majored in. You'll never be any better at playing soccer or dating or taking risks. Your life and character are as certain as a map. The problem is, this mindset will make you complacent, rob your self-esteem and bring meaningful education to a halt.
>
> In short, it's an intellectual disease and patently untrue.

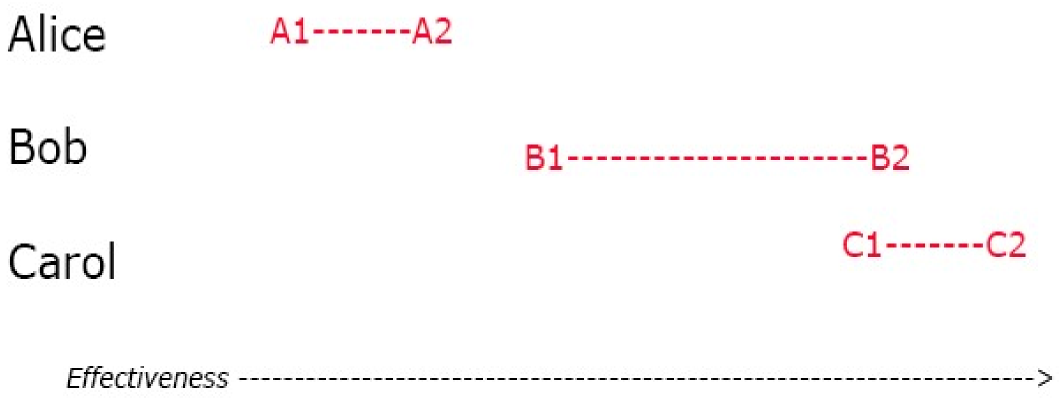
The article goes on to show how growth mindset proves talent is “a myth”, a claim repeated by growth mindset cheerleader articles like Debunking The Genius Myth and The Learning Myth: Why I’ll Never Tell My Son He’s Smart and this woman who says we need to debunk the idea of innate talent.

Suppose everything I said in parts I – IV was wrong, and growth mindset is 100% true exactly as written.

This still would not provide *an iota of evidence* against the idea that innate talent / IQ / whatever is by far the most important factor determining success.

Consider. We know from countless studies that strong religious belief increases your life expectancy, makes you happier, reduces your risk of depression and reduces crime. Clearly believing in, say, Christianity has lots of useful benefits. But no one would dare argue that proves Christianity *true*. It doesn’t even *imply* it.

Likewise, mindset theory suggests that believing intelligence to be mostly malleable has lots of useful benefits. That doesn’t mean intelligence really *is* mostly malleable. Consider, if you will, my horrible graph:

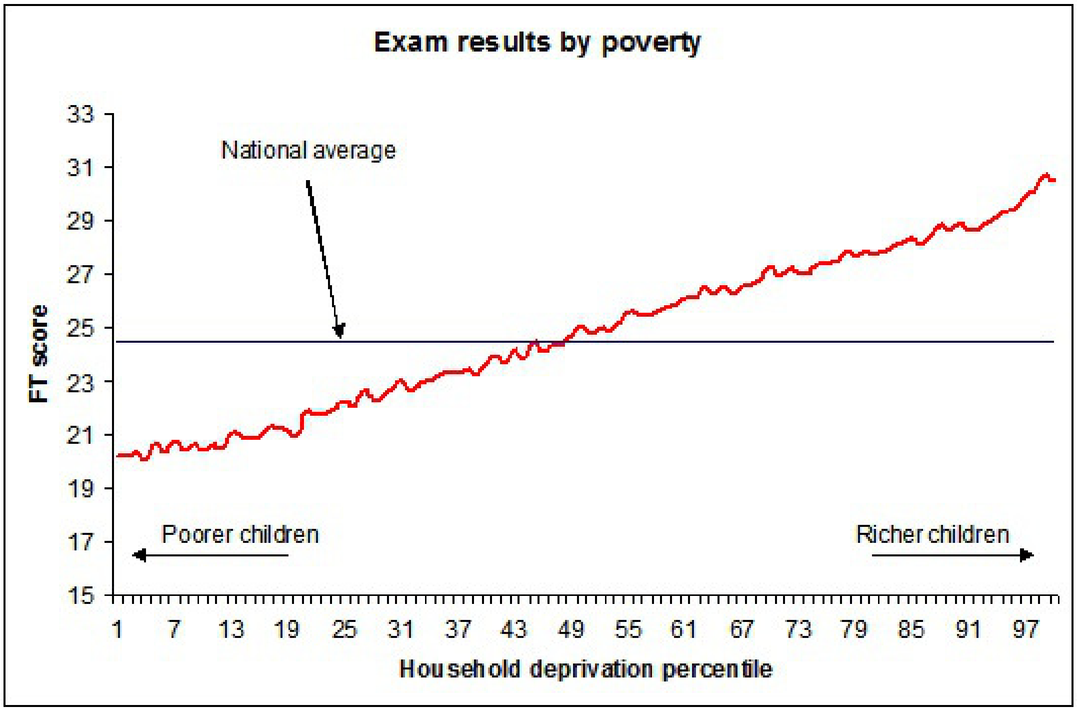


Suppose this is one of Dweck's experiments on three children. Each has a different level of innate talent, represented by point 1. After they get a growth mindset and have the right attitude and practice a lot, they make it to point 2.

Two things are simultaneously true of this model. First, all of Dweck's experiments will come out exactly as they did in the real world. Children who adopt a growth mindset and try hard and practice will do better than children who don't. If many of them are aggregated into groups, the growth mindset group will on average do better than the ability-focused group. Intelligence is flexible, and if you don't bother practicing than you fail to realize your full potential.

Second, the vast majority of difference between individuals is due to different levels of innate talent. Alice, no matter how hard she practices, will never be as good as Bob. Bob, if he practices very hard, will become better than Carol was at the start, but never as good as Carol if she practices as hard as Bob does. The difference between Alice and Carol is a vast, unbridgeable gap which growth mindset has nothing whatsoever to say about.

Here is a graph which is less terrible because it was not made by me. I have taken it from one of the two other sources I have found on the entire Internet that don't like growth mindset:



We can argue all day about whether poor students do worse because they have bad health, because they have bad genes, because they have bad upbringings, or because society is fixed against them. We *have* argued about that all day before here, and it's been pretty interesting.

But in this case it doesn't matter. If the only thing that affects success is how much effort you put in, poor kids seem to be putting in a heck of a lot less effort in a surprisingly linear way. But the smart money's not on that theory.

A rare point of agreement between hard biodeterminists and hard socialists is that telling kids that they're failing because they just don't have the right work ethic is a *crappy thing to do*. It's usually false and it will make them feel terrible. Behavioral genetics studies show pretty clearly that at least 50% of success at academics and sports is genetic; various sociologists have put a lot of work into proving that your position in a biased society covers a pretty big portion of the remainder. If somebody who was born with the dice stacked against them works very hard, then they might find themselves at A2 above. To deny this in favor of a "everything is about how hard you work" is to offend the sensibilities of sensible people on the left and right alike.

Go back to that 1975 paper above on "Role Of Expectations And Attributions" and look more closely at the proposed intervention to help these poor fixed mindset students:

> Twelve extremely helpless children were identified [and tested on how many math problems they could solve in a certain amount of time]…the criterion number was set one above the number he was generally able to complete within the time limit. On these trials, he was stopped one or two problems short of criterion, his performance was compared to the criterion number required, and experimenter verbally attributed the failure to insufficient effort.

So basically, you take the most vulnerable people, set them tasks you know they'll fail at, then lecture them about how they only failed because of insufficient effort.

Imagine a boot stamping on a human face forever, saying "YOUR PROBLEM IS THAT YOU'RE JUST NOT TRYING NOT TO BE STAMPED ON HARD ENOUGH".

And maybe this is worth it, if it builds a growth mindset that allows the child to be more successful in school, sports, and in the rest of her life. But you're not "debunking the myth of genius". Genius remains super-important, just like conscientiousness and wealth and health and privilege and everything else. No, you're telling a Noble Lie to the children because you think it's useful. You can make it palatable by saying "Well, we're not *denying* reality, we're just *selectively emphasizing certain parts of* reality, but in the end that's what you’re doing. If you can square that with your moral system, go ahead.

But I remain agnostic. There are some really good – *diabolically* good? – studies showing that it works in certain lab situations. There’s a lot of excellent research behind it and a lot of brilliant people giving it their support. But there are also other studies showing that it has no long-term real-world effects that we can measure, and others that might (or might not?) contradict its predictions in other ways. I have only the barest of ideas how to square those facts, and I look forward to hearing from anyone who has more.

I haven’t read Dweck’s book, but it’s an obvious next step for anyone who wants to look into these issues further.

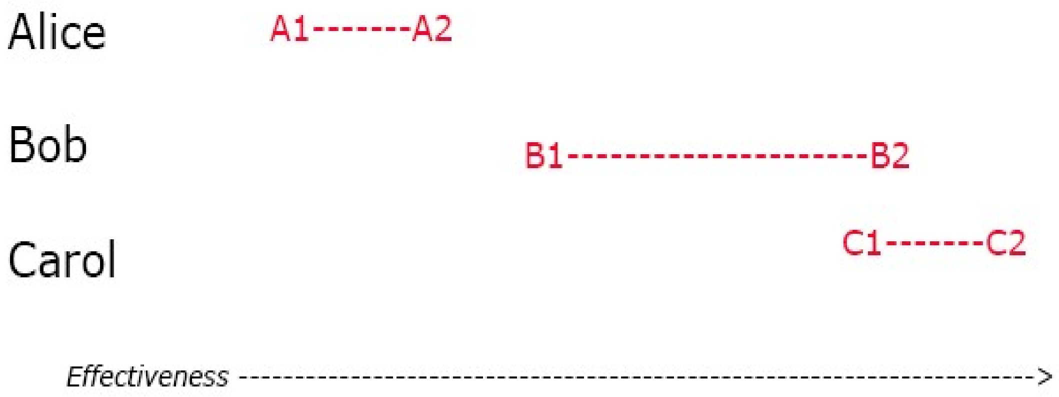
I Will Never Have The Ability To Clearly Explain My Beliefs About Growth Mindset

I.

A lot of the comments I've gotten about Tuesday's post on growth mindset have been pretty similar. They've argued that yes, innate ability might matter, but that even the most innate abilityed person needs effort to fulfill her potential. If someone were to believe that success were 100% due to fixed innate ability and had nothing to do with practice, then they wouldn't bother practicing, and they would fall behind. Even if their innate ability kept them from falling behind morons, at the very least they would fall behind their equally innate abilityed peers who *did* practice.

I will call this the Bloody Obvious Position, since it's hard to believe it isn't true. I once tried to imagine a world without it as a thought experiment, but it was pretty weird and I wasn't serious.

The Bloody Obvious Position was what I was trying to get at with my post on basketball, and with my terrible ad hoc graph:



Nevertheless, some people thought I was denying the Bloody Obvious Position. Other people thought I was accusing Carol Dweck of denying the Bloody Obvious Position (see eg here). This despite my making sure to say:

> I want to end by correcting a very important mistake about growth mindset that Dweck mostly avoids but which her partisans constantly commit egregiously.

I believe the Bloody Obvious Position. Dweck believes the Bloody Obvious Position. I acknowledge that Dweck believes the Bloody Obvious Position. There are a lot of growth mindset partisans online who don't believe the Bloody Obvious Position, and I satisfied my urge to yell at them, but now they've been yelled at, and the more important issues debated by reasonable people still remain.

So where do I disagree with Dweck? I interpret Dweck as making the following statement:

The more important you believe innate ability to be compared to effort, the more likely you are to stop trying, to avoid challenges, to lie and cheat, to hate learning, and to be obsessed with how you appear before others

Call it the Controversial Position. This is *not* the same thing as the Bloody Obvious Position. In the Bloody Obvious

Position, someone can believe success is 90% innate ability and 10% effort. They might also be an Olympian who realizes that at her level, pretty much everyone is at a innate ability ceiling, and a 10% difference is the difference between a gold medal and a last-place finish. So she practices very hard and does just as well as anyone else.

According to the Controversial Position, this athlete will *still* do worse than someone who believes success is 80% ability and 20% effort, who will in turn do worse than someone who believes success is 70% ability and 30% effort, all the way down to the person who believes success is 0% ability and 100% effort, who will do best of all and take the gold medal.

And this is why I deny that I'm secretly agreeing with Dweck, or strawmanning Dweck, or whatever. I don't believe the Controversial Position, but I think Dweck does. For example, here she writes: "The more a player believed athletic ability was a result of effort and practice rather than just natural ability, the better that player performed".

There is nothing in there about "the more a player realizes that, no matter how important innate ability is, effort matters too." Her statement says that it's entirely about what degree a player attributes success to effort versus innate ability. The natural conclusion there is that the player who believes success is 0% innate ability and 100% effort will do the best.

Her studies reflect this as well. The most common design uses the IAR, a test where children are asked to attribute different things to effort versus ability. Those who attribute too many things to ability are classified as "helpless" and "fixed mindset". There's no question about "Okay, some things are due to ability, but if you work hard that still helps, right?" Nor have I ever seen any of the literature claim "it's important to believe effort matters a little, but after a certain point more effort-attribution doesn't help", or "Maybe there's an L-shaped relationship between belief-in-importance-of-ability and success.”

I’d like to be able to teach my children that success is X% innate ability and Y% practice, for non-zero values of both X and Y. I think growth mindset theory claims that if some other parent teaches their kids the same thing for a lower value of X and higher value of Y, their children will be more honest, harder-working, and more successful. And the parent who says it’s 0% innate ability and 100% practice will do best of all. If growth mindset people don’t believe that, I can only confess I have never been able to infer that lack of belief from their writings.

II.

Worse, we can distinguish between a Sorta Controversial Position and a Very Controversial Position:

SCP: The more children believe effort matters, and the less they believe innate ability matters, the more successful they will be. This is because every iota of belief they have in effort gives them more incentive to practice. A child who believes innate ability and effort both explain part of the story might think “Well, if I practice I’ll become a *little* better, but I’ll never be as good as Mozart. So I’ll practice a little but not get my hopes up.” A child who believes only effort matters, and innate ability doesn’t matter at all, might think “If I practice enough, I can become exactly as good as Mozart.” Then she will practice a truly ridiculous amount to try to achieve fame and fortune. This is why growth mindset works.

VCP: Belief in the importance of ability directly saps a child’s good qualities in some complicated psychological way. It is worse than merely believing that success is based on luck, or success is based on skin color, or that success is based on whatever other thing that isn’t effort. It shifts children into a mode where they must protect their claim to genius at all costs, whether that requires lying, cheating, self-sabotaging, or just avoiding intellectual effort entirely. When a fixed mindset child doesn't practice as much, it's not because they've made a rational calculation about the utility of practice towards achieving success, it's because they've partly or entirely *abandoned success as a goal* in favor of the goal of trying to convince other people that they're Smart.

Carol Dweck unambiguously believes the Very Controversial Position. In a quotation which I admit I am mangling and ellipsis-ing heavily to remove extra verbiage, but which I think preserves the meaning of her claim:

> [People with fixed mindsets] are so concerned with being and looking talented that they never realize their full potential. In a fixed mindset, the cardinal rule is to look talented at all costs. The second rule is don't work too hard or practive too much…having to work casts doubt on your ability. The third rule is, when faced with setbacks, run away. They say things like 'I would try to cheat on the next test'. They make excuses, they blame others, they make themselves feel better by looking down on those who have done worse."

Can we all agree this is a much stronger claim than "ability matters, but effort also matters?"

III.

I was not intending to "debunk" growth mindset, or even present a pure polemic against growth mindset. I admitted that many of the studies around it were very good, and that I don't have good answers to them. My bias is against the theory, but I tried not to *just* follow my bias. I tried to treat it on the level of "there's a lot of good evidence for growth mindset, now what’s the best evidence we can find against it?”

So I guess I should probably come out and say what I believe about each position.

I believe the Bloody Obvious Position is bloody obvious.

I believe the Somewhat Controversial Position is probably not a good way to parse things. Part of it is that we might be confusing explicit versus implicit beliefs. Maybe a particular geneticist is very aware of research showing how important genetics is to success, and would give a very high estimate if asked, but in her own life, when she fails, lack of effort is still the first explanation to immediately leap to mind. Or maybe some teacher is very on board with growth mindset and things IQ is a racist construct, but is convinced that she can’t do physics because she’s just “not a math kind of person”. The research I’ve seen hasn’t really distinguished between explicit and implicit beliefs. The priming experiments sure seem more likely to affect what immediately comes to mind than your stable, well-reasoned beliefs about how the world works (even though a few priming experiments have checked stable well-reasoned beliefs to see if the intervention worked!)

If you put a gun to my head, I’ll say it certainly works in the lab, and give you about 50-50 odds that it matters in real life. The studies don’t show any real-life correlation between growth mindset and any measures of success. Many people have pointed out that this could be confounded – dumb people might preferentially believe ability doesn’t matter to make themselves feel better about not having it, and smart people might preferentially believe effort doesn’t matter because they rarely have to use it. But if you accept that, some of the rest of it starts to look confounded. If fixed mindset = smart people, than might the reason they react poorly to challenges and failure be that they have no experience with them? Might it be that the more challenges and failures you've encountered before, the better you are at dealing with them? Certainly that is how I interpret this Dweck paper, even though she thinks it is purely a mindset effect. Another way of explaining the ecological results without bringing in that particular confounder would be if growth mindset helped in some situations, but fixed mindset helped in others. For example, a person with fixed mindset risks not trying hard enough because they think there's no point. But a person with growth mindset risks the opportunity costs of prolonging their inevitable failure instead of (in the Silicon Valley term) "failing fast" and pivoting towards a higher-payoff activity.

The Very Controversial Position is also well-supported, also contradicted by ecological data, and really really doesn't match my experience. The people I know who are most interested in issues of innate ability don't behave at all like Dweck's subjects. In fact, I wonder if a lot of the "life-hacking" movement might be ability-mindset people trying to figure out how to succeed more by improving ability – certainly the people who practice dual-n-back every day because they think it increases IQ fall into this category, but so do nootropics users, people who follow special diets to increase energy, and "try this one weird trick to improve your motivation". And these same people seem interested in things like spaced repetition software, which might be thought of as sort of prosthetic ability-enhancers. On the other hand, I've also met people who say "I could succeed if only I put in some effort, but I have some mental block / depression / ADHD / low conscientiousness score that makes it impossible for me to work that hard, so better go eat worms", then sabotage themselves at every opportunity.

And yes, it’s a sin to privilege your own experience and priors over the results of good studies, but sometimes it’s necessary. And it’s another sin to prefer the results of broad ecological studies to controlled experimental trials, but sometimes that’s necessary too.

I deny the claim that I don’t disagree with Dweck on anything of substance. I don’t *absolutely* disagree with her on anything, but there are a lot of things I doubt, or that I expect to capture true insights without being the best way to express them. Growth mindset makes some surprising and genuinely controversial claims, and I’m not yet at the point where I can feel sure about them either way.

Michigan Meetup 4/14

There will be a Michigan rationalist/LW/SSC meetup at Pizza House (618 Church Street, Ann Arbor) on Tuesday, April 14 at 8 PM.

If you’re reading this and can make it to Ann Arbor, you’re invited.

Our special guest will be Scott Aaronson, MIT professor and quantum computing expert.

Trouble Walking Down The Hallway

Williams and Ceci just released National Hiring Experiments Reveal 2:1 Faculty Preference For Women On STEM Tenure Track, showing a strong bias in favor of women in STEM hiring. I've previously argued something like this was probably the case, so I should be feeling pretty vindicated.

But a while ago I wrote Beware The Man Of One Study, in which I wrote that there is such a variety of studies finding such a variety of contradictory things that anybody can isolate one of them, hold it up as *the* answer, and then claim that their side is right and the other side are 'science denialists'. The only way to be sure you're getting anything close to the truth is to examine the literature of an entire field as a gestalt.

And here's something no one ever said: "Man, I'm so glad I examined the literature of that entire field as a gestalt, things make much more sense now."

Two years ago Moss-Racusin et al released Science Faculty's Subtle Gender Biases Favor Male Students, showing a strong bias in favor of men in STEM hiring. The methodology was almost identical to this current study, but it returned the opposite result.

Now everyone gets to cite whichever study accords with their pre-existing beliefs. So *Scientific American* writes Study Shows Gender Bias In Science Is Real, and any doubt has been deemed unacceptable by blog posts like Breaking: Some Dudes On The Internet Refuse To Believe Sexism Is A Thing. But the new study, for its part, is already producing headlines like The Myth About Women In Science and blog posts saying that it is "enough for everyone who is reasonable to agree that the feminists are spectacular liars and/or unhinged cranks”.

So probably we’re going to have to do that @#$%ing gestalt thing.

Why *did* these two similar studies get such different results? Williams and Ceci do something wonderful that I’ve never seen anyone else do before – they include in their study a supplement admitting that past research has contradicted theirs and speculating about why that might be:

1. W&C investigate hiring tenure-track faculty; MR&a investigate hiring a “lab manager”. This is a big difference, but as far as I can tell, W&C don’t give a good explanation for why there should be a pro-male bias for lab managers but a pro-female bias for faculty. The best explanation I can think of is that there have been a lot of recent anti-discrimination campaigns focusing on the shortage of female faculty, so that particular decision might activate a cultural script where people think “Oh, this is one of those things that those feminists are always going on about, I should make sure to be nice to women here,” in a way that just hiring a lab manager doesn’t.

Likewise, hiring a professor is an important and symbolic step that…probably doesn’t matter super-much to other professors. Hiring a lab manager is a step without any symbolism at all, but professors often work with them on a daily basis and depend on their competency. That might make the first decision Far Mode and the second Near Mode. Think of the Obama Effect – mildly prejudiced people who might be wary at the thought of having a black roommate were very happy to elect a black President and bask in a symbolic dispay of tolerance that made no difference whatsoever to their everyday lives.

Or it could be something simpler. Maybe lab work, which is very dirty and hands-on, feels more “male” to people, and professorial work, which is about interacting with people and being well-educated, feels more “female”. In any case, W&C say their study is more relevant, because almost nobody in academic science gets their start as a lab manager (they polled 83 scientists and found only one who had).

2. Both W&C and MR&a ensured that the male and female resumes in their study were equally good. But W&C made them all excellent, and MR&a made them all so-so. Once again, it’s not really clear why this should change the direction of bias. But here’s a hare-brained theory: suppose you hire using the following algorithm: it’s very important that you hire someone at least marginally competent. And it’s *somewhat* important that you hire a woman so you look virtuous. But you secretly believe that men are more competent than women. So given two so-so resumes, you’ll hire the man to make sure you get someone competent enough to work with. But given two excellent resumes, you know neither candidate will accidentally program the cyclotron to explode, so you pick the woman and feel good about yourself.

And here are some other possibilities that they didn’t include in their supplement, but which might also have made a difference.

3. W&C asked “which candidate would you hire?”. MR&a said “rate each candidate on the following metrics” (including hireability). Does this make a difference? I could *sort of* see someone who believed in affirmative action saying something like “the man is more hireable, but I would prefer to hire the woman”. Other contexts prove that even small differences in the phrasing of a question can lead to major incongruities. For example, as of 2010, only 34% of people polled strongly supported letting homosexuals serve in the military, but half again as many – a full 51% – expressed that level of support for letting "gays and lesbians" serve in the military. Ever since reading that I've worried about how many important decisions are being made by the 17% of people who support gays and lesbians but not homosexuals.

I SUPPORT GAY MARRIAGE, BUT NOT HOMOSEXUAL MARRIAGE
AND I VOTE
...ERE VOTING BACK ...

For all we know maybe this is the guy in charge of hiring for STEM faculty positions

4. Williams and Ceci asked participants to choose between "Dr. X" (who was described using the pronouns "he" and "him") and "Dr. Y" (who was described using the pronouns "she" and "her"). Moss-Racusin et al asked participants to choose between "John" and "Jennifer". They said they checked to make sure that the names were rated equal for "likeability" (whatever that means), but what if there are other important characteristics that likeability doesn't capture? We know that names have big effects on our preconceptions of people. For example, people with short first names earn more money – an average of $3600 less per letter. If we trust this study (which may not be wise), John already has a $14,400 advantage on Jennifer, which goes a lot of the way to explaining why the participants offered John higher pay without bringing gender into it at all!

Likewise, independently of a person's gender they are more likely to succeed in a traditionally male field if they have a male-sounding name. That means that one of the…call it a "prime" that activates sexism…might have been missed by comparing Dr. X to Dr. Y, but captured by pitting the masculine-sounding John against the feminine-sounding Jennifer. We can't claim that W&C's subjects were rendered gender-blind by the lack of gender-coded names – they noticed the female candidates enough to pick them twice as often as the men – but it might be that not getting the name activated the idea of gender from a different direction than hearing the candidates' names would have.

5. Commenter Lee points out that MR&a tried to make their hokey hypothetical hiring seem a little more real than W&C did. MR&a suggest that these are real candidates being hired…somewhere…and the respondents have to help decide whom to hire (although they still use the word "imagine"). W&C clearly say that this is a hypothetical situation and ask the respondents to imagine that it is true. Some people in the comments are arguing that this makes W&C a better signaling opportunity whereas MR&a stays in near mode. But why would people not signal on a hiring question being put to them by people they don't know about a carefully-obscured situation in some far-off university? Are sexists, out of the goodness of their hearts, urging MR&a to hire the man out of some compassionate desire to ensure they get a qualified candidate, but when

W&C send them a hypothetical situation, they switch back into signaling mode?

6. Commenter Will points out that MR&a send actual resumes to their reviewers, but W&C send only a narrative that sums up some aspects of the candidates' achievements and personalities (this is also the concern of Feminist Philosophers). This is somewhat necessitated by the complexities of tenure-track hiring – it's hard to make up an entire fake academic when you can find every published paper in Google Scholar – but it does take them a step away from realism. They claim that they validated this methodology against real resumes, but it was a comparatively small validation – only 35 people. On the other hand, even this small validation was highly significant for pro-female bias. Maybe for some reason getting summaries instead of resumes heavily biases people in favor of women?

Or maybe none of those things mattered at all. Maybe all of this is missing the forest for the trees.

I love stories about how scientists set out to prove some position they consider obvious, but unexpectedly end up changing their minds when the results come in. But this isn't one of those stories. Williams and Ceci have been vocal proponents of the position that science isn't sexist for years now – for example, their article in the New York Times last year, Academic Science Isn't Sexist. In 2010 they wrote Understanding Current Causes Of Women's Underrepresentation In Science, which states:

> The ongoing focus on sex discrimination in reviewing, interviewing, and hiring represents costly, misplaced effort: Society is engaged in the present in solving problems of the past, rather than in addressing

> meaningful limitations deterring women’s participation in science, technology, engineering, and mathematics careers today. Addressing today’s causes of underrepresentation requires focusing on education and policy changes that will make institutions responsive to differing biological realities of the sexes.

So they can hardly claim to be going into this with perfect neutrality.

But the lead author of the study that *did* find strong evidence of sexism, Corinne Moss-Racusin (whose name is an anagram of “accuser on minor sins”) *also* has a long history of pushing the position she coincidentally later found to be the correct one. A look at her resume shows that she has a bunch of papers with titles like “Defending the gender hierarchy motivates prejudice against female leaders”, “‘But that doesn’t apply to me:’ teaching college students to think about gender”, and “Engaging white men in workplace diversity: can training be effective?”. Her symposia have titles like “Taking a stand: the predictors and importance of confronting discrimination”. This does not sound like the resume of a woman whose studies ever find that oh, cool, it looks like sexism isn’t a big problem here after all.

So what conclusion should we draw from the people who obviously wanted to find a lack of sexism finding a lack of sexism, but the people who obviously wanted to find lots of sexism finding lots of sexism?

This is a *hard question*. It doesn’t necessarily imply the sinister type of bias – it may be that Drs. Williams and Ceci are passionate believers in a scientific meritocracy simply because that’s what all their studies always show, and Dr. Moss-Racusin is a passionate believer in discrimination because that’s what *her* studies find. On the other hand, it’s *still* suspicious that two teams spend lots of time doing lots of experiments, and one always gets one result, and the other always gets the other. What are they doing differently?

Problem is, I don’t know. Neither study here has any egregious howlers. In my own field of psychiatry, when a drug company rigs a study to put their drug on top, usually before long someone figures out how they did it. In these two studies I’m not seeing anything.

And this casts doubt upon those four possible sources of differences listed above. None of them look like the telltale sign of an experimenter effect. If MR&a were trying to fix their study to show lots of sexism, it would have taken exceptional brilliance to do it by using the names “John” versus “Jennifer”. If W&C were trying to fix their study to disguise sexism, it would have taken equal genius to realize they could do it by asking people “who would you hire?” rather than “who is most hireable?”.

(the only exception here is the lab manager. It’s *just* within the realm of probability that MR&a might have somehow realized they’d get a stronger signal asking about lab managers instead of faculty. The choice to ask about lab managers instead of faculty is surprising and does demand an explanation. And it’s probably the best candidate for the big difference between their results. But for them to realize that they needed to pull this deception suggests an impressive ability to avoid drinking their own Kool-Aid.)

Other than that, the differences I’ve been considering in these studies are the sort that would be very hard to purposefully bias. But the fact that both groups got the result they wanted suggests that the studies were purposefully biased *somehow*. This reinforces my belief that experimenter effects are best modeled as some sort of mystical curse incomprehensible to human understanding.

(now would be an excellent time to re-read the the horror stories in Part IV of “The Control Group Is Out Of Control”)

Speaking of horror stories. Sexism in STEM is, to put it mildly, a hot topic right now. Huge fortunes in grant money are being doled out to investigate it (Dr. Moss-Racusin alone received nearly a million dollars in grants to study STEM gender bias) and thousands of pages are written about it every year. And yet somehow the entire assembled armies of Science, when directed toward the problem, can’t figure out whether college professors are more or less likely to hire women than men.

This is not like studying the atmosphere of Neptune, where we need to send hundred-million dollar spacecraft on a perilous mission before we can even begin to look into the problem. This is not like studying dangerous medications, where ethical problems prevent us from doing the experiments we really need. This is not like studying genetics, where you have to gather large samples of identical twins separated at birth, or like climatology, where you hang out at the North Pole and might get eaten by bears. This is a *survey of college professors*. You know who it is studying this? *College professors*. The people they want to study are *in the same building as them*. The climatologists are getting eaten by bears, and the social psychologists can’t even settle a question that requires them to *walk down the hallway*.

It’s not even like we’re trying to detect a subtle effect here. Both sides agree that the signal is very large. They just disagree what direction it’s very large in!

A recent theme of this blog has been that Pyramid Of Scientific Evidence be damned, our randomized controlled trials suck so hard that a lot of the time we'll get more trustworthy information from just looking at the ecological picture. Williams and Ceci have done this (see Part V, Section b of their supplement, "Do These Results Differ From Actual Hiring Data") and report that studies of real-world hiring data confirm women have an advantage over men in STEM faculty hiring (although far fewer of them apply). It also matches the anecdotal evidence I hear from people in the field. I'm not necessarily saying I'm ambivalent between the two studies' conclusions. Just that it bothers me that we have to go to tiebreakers after doing two good randomized controlled trials.

At this point, I think the most responsible thing would be to have a joint study by both teams, where they all agree on a fair protocol beforehand and see what happens. Outside of parapsychology I've never heard of people taking such a drastic step – who would get to be first author?! – but at this point it's hard to deny that it's necessary.

In conclusion, I believe the Moss-Racusin et al study more, but I think the Williams and Ceci study is more believable. And the best way to fight sexism in science is to remind people that it would be hard for women to make things any more screwed up than they already are.

Polemical Imbalance

Today is an exciting day for me. I got argued against on *Mad In America*. This one is going straight to my resume.

Mad In America apparently doesn't like being called an anti-psychiatry blog, so let's call it a blog…that discusses psychiatry…and doesn't usually like what it sees. They were heavily involved in popularizing the idea that psychiatry erred grieviously in overselling "chemical imbalance", and they didn't much like my post on the same topic:

> Alexander argues that the notion that psychiatrists once promoted the idea of low serotonin as a cause of depression and Selective Serotonin Reuptake Inhibitors (SSRIs) as proper treatment for that deficiency is all simply a false "narrative" invented by "antipsychiatry" activists. These activists then "frame it as 'proof' that psychiatrists are drug company shills who were deceiving the public." Alexander points to quotes of American Psychiatric Association officials in a post by MIA Blogger Philip Hickey, and notes that none of the quotes specifically describe a low-serotonin explanation for depression. The Hickey post cited is not actually about that topic, but about the promotion of the phrase "chemical imbalance"; nevertheless, Alexander broadly refers to Hickey and all of Mad in America as "antipsychiatry", and he then writes, "If the antipsychiatry community had quotes of APA officials saying it's all serotonin deficiency, don't you think they would have used them?" Alexander argues, "The idea that depression is a drop-dead simple serotonin deficiency was never taken seriously by mainstream psychiatry." There seems to be a lot of evidence to the

> contrary still today readily available even on the web, though.

This is exactly the sort of fight I probably shouldn’t get involved in continuing. But I’m going to do so anyway, because I think Mad In America’s counterargument is actually going to end up supporting my point and maybe shed more light on the situation.

Up there, when they say “Alexander points to quotes of American Psychiatric Association officials in a post by MIA Blogger Philip Hickey, and notes that none of the quotes specifically describe a low-serotonin explanation for depression [but] the Hickey post cited is not actually about that topic, but about the promotion of the phrase ‘chemical imbalance'” – that’s where I get pretty confident they’ve missed my point.

Remember, the thesis of my last post was that the “chemical imbalance” argument hides a sort of bait-and-switch going on between the following two statements:

(A): Depression is complicated, but it seems to involve disruptions to the levels of brain chemicals in some important way

(B): We understand depression perfectly now, it’s just a deficiency of serotonin.

If you equivocate between them, you can prove that psychiatrists were saying (A), and you can prove that (B) is false and stupid, and then it’s sort of like psychiatrists were saying something false and stupid.

Given that this is my thesis, it’s exactly right for me to debate a post on “chemical imbalance” by showing that none of the quotes involved reduce the problem to just a basic serotonin deficiency!

And when Rob Wipond from MIA says he's found "a lot of evidence to the contrary still readily available even on the web", well, spoiler, he's found more people saying A.

II.

Let's go through his examples:

> For example, a 2004 Washington University in St. Louis press release, about a study published in Biological Psychiatry, states that the "brain's serotonin receptors" are "at abnormally low levels in depressed people" and that antidepressants "work by increasing serotonin levels in the brain."

I assume he's talking about this press release about a study that shows abnormally low levels of serotonin receptors in depressed people. First of all, the study actually did show this. I don't think it's irresponsible to mention that a study shows low levels of serotonin receptors in depressed people when a study actually shows this. Second of all, the press release makes it extremely clear that they don't know exactly what's going on: "Little is understood about how depression makes people feel sad, but neuroscientists do know that the brain chemical serotonin is involved." They mention that SSRIs appear to work for depression, but admit that "The bad news is that beyond that first step of increasing serotonin, we understand very little about how these drugs relieve symptoms of depression". Finally, this study actually found something much more complicated than the prevailing narrative – a serotonin deficiency model of depression would have predicted high levels of serotonin receptors in related brain structures (more chemicals =

fewer receptors) but in fact it found the opposite. This fits with the emerging theory that depression may be related to increased serotonin levels in certain parts of the brain, which SSRIs provoke a compensatory response against.

This press release is actually as good as the harshest critic could have wished for. It admits we don't really know how depression works, it admits we don't really know how SSRIs treat it, and then it presents the result of a study that shows that serotonin is implicated but not in the way the "serotonin deficiency" theory would expect.

The only way Mad In America turned this into a poster child for psychiatry deceiving people about serotonin was to quote from it extremely out of context.

Let's go to their next example:

> And there is prominent psychiatrist Richard Friedman writing in the New York Times in 2007 that psychiatrists were soon going to be able to conduct "a simple blood test" to determine "what biological type of depression" a person had and then treat them with the right drug. "For example," writes Friedman, "some depressed patients who have abnormally low levels of serotonin respond to S.S.R.I.'s, which relieve depression, in part, by flooding the brain with serotonin."

Okay, but Friedman starts with a story about how SSRIs often don't work for patients, then says that this is because some people have depression that doesn't seem to be serotonergic: "Some depressed patients who have abnormally low levels of serotonin respond to SSRIs, which relieve depression, in part, by flooding the brain with serotonin. Other depressed patients may have an abnormality in other neurotransmitters that regulate mood, like norepinephrine or dopamine, and may not respond to SSRIs". He says (correctly!) that "in everyday clinical practice, we have little ability to predict what specific treatment will work for you".

These are not the words of a drug company shill who says that depression is 100% serotonin in order to put everyone on SSRIs! These are the words of someone who agrees with me that depression is somehow related to neurotransmitters, but it's still very uncertain which ones and how. His only sin seems to be an overly optimistic view of the speed at which we would come out with genetic tests.

Next example:

> There's also a lot of evidence that the low-serotonin theory of depression is still today being taken seriously by mainstream psychiatry and is still being promoted to the public. A current University of Bristol public education website on depression explains that, "Low serotonin levels are believed to be the cause of many cases of mild to severe depression."

That appears to be this University of Bristol public education website. The site says it's by "Claire Rosling", so I searched her name and I get this roster of people's sophomore chemistry projects. Ms. Rosling's is…the website Mad In America cited. Apparently this was part of some college chemistry assignment where people write about molecules to compete for a £50 prize. Ms. Rosling's was serotonin.

So Mad In America argues that the entire psychiatric establishment is pushing the "depression = serotonin" argument, but the best example they can come up with is some poor woman's undergraduate chemistry homework?

(in case you’re wondering, she didn’t win. Some girl named Anna won for her webpage on Recycling Plastic.)

Next example!

> A current Harvard Medical School special health report, “Understanding Depression”, explains that, “Research supports the idea that some depressed people have reduced serotonin transmission. Low levels of a serotonin byproduct have been linked to a higher risk for suicide.”

Once again, holy !@#$, they’re *reporting the results of actual studies*. It’s dishonest to do studies on serotonin and find that it is linked to depression? Anyway, when I look up the actual report it starts with the following paragraph: “It’s often said that depression results from a chemical imbalance, but that figure of speech doesn’t capture how complex the disease is. Research suggests that depression doesn’t spring from simply having too much or too little of certain brain chemicals. Rather, depression has many possible causes, including faulty mood regulation by the brain, genetic vulnerability, stressful life events, medications, and medical problems. It’s believed that several of these forces interact to bring on depression.”

Once again, this is the best you can do to find psychiatrists pushing an oversimplified version of the chemical imbalance theory??!

Next example:

> WebMD’s “Depression Center” states that, “There are many researchers who believe that an imbalance in serotonin levels may influence mood in a way that leads to depression. Possible problems include low brain cell production of serotonin, a lack of receptor sites able to

> receive the serotonin that is made… According to Princeton neuroscientist Barry Jacobs… common antidepressant medications known as SSRIs, which are designed to boost serotonin levels, help kick off the production of new brain cells, which in turn allows the depression to lift."

First of all, this page does *not* use the classic "serotonin deficiency" theory of depression. This is the hippocampal neurogenesis theory, which in my last post I specifically contrasted with the classic serotonin deficiency theory. Yes, it involves serotonin in some way, but since one of the most important facts about depression is that SSRIs treat it, every theory is going to involve serotonin in some way.

Further, right after this paragraph, WebMD continues: "Although it is widely believed that a serotonin deficiency plays a role in depression, there is no way to measure its levels in the living brain. Therefore, there have not been any studies proving that brain levels of this or any neurotransmitter are in short supply when depression or any mental illness develops. Blood levels of serotonin are measurable — and have been shown to be lower in people who suffer from depression – but researchers don't know if blood levels reflect the brain's level of serotonin. Also, researchers don't know whether the dip in serotonin causes the depression, or the depression causes serotonin levels to drop."

Once again, I see nothing here to indicate that they are covering up flaws in this theory, pushing it to unsuspecting consumers, or claiming that exploratory research is settled science. They're presenting the best theories we've got, then noting how tentative they are and what the flaws are.

(on the other hand, the article *does* say that there are “40 million” brain cells, when in fact there are about 90 billion. I’m not saying you should trust WebMD, just that they don’t bungle depression in *that* particular way)

Next example:

> And if the theory was never taken seriously and isn’t being taken seriously, no one has apparently told the National Academy of Sciences or two news media outlets with expert psychiatric editorial boards yet. Psychiatry Advisor’s February 12, 2015 headline for a report about a Duke University study is, “Serotonin Deficiency May Up Depression Risk.” Psychiatry Advisor explains that, “(m)ice with normal serotonin levels, the control group, did not demonstrate depression symptoms a week after the social stress, while the serotonin-deficient rodents did(.)” The study, appearing in the Proceedings of the National Academy of Sciences, states that, serotonin deficiency has been “implicated in the etiology of depression” though a cause-effect relationship has not yet been “formally established.” The researchers write that their results, “provide additional insight into the serotonin deficiency hypothesis of depression.” Medical News Today headline their report on it even more strongly: “Mouse study finds that serotonin deficiency does increase depression risk.” (Medical News Today notes in passing that an earlier, somewhat similar study by a different team came to the exact opposite findings.)

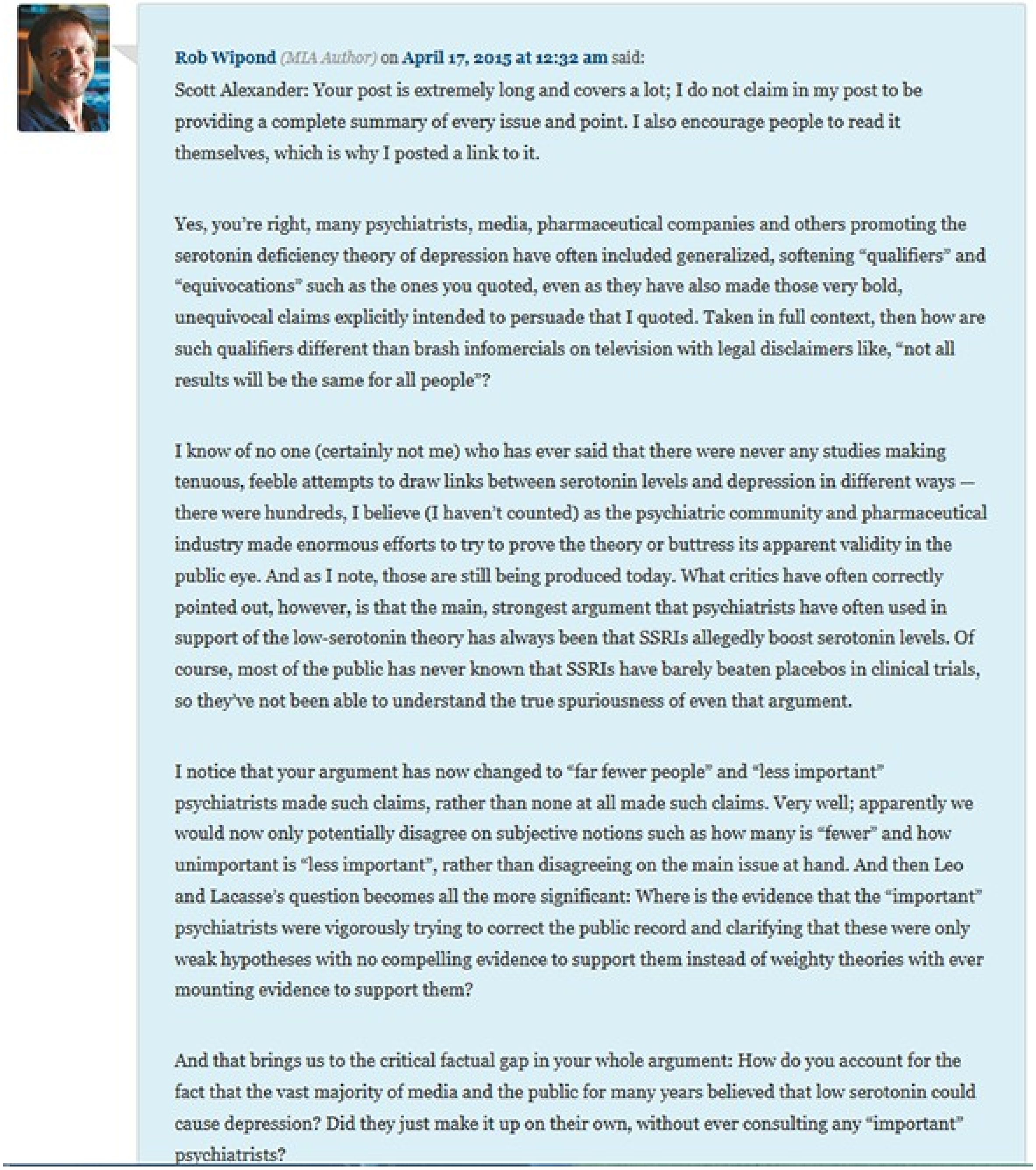
At this point Mad in America’s examples are self-refuting. I am getting the impression they will never be happy unless no news media ever covers the dozens of studies that come out each year linking depression to serotonin. I know this sounds mean, but what other conclusion am I supposed to come to? Here we have a study that provides some evidence for serotonin’s involvement, says very specifically that “a cause-effect relationship has not been formally established”, mentions that other studies have shown the opposite – and yet Mad In America *still* wants me to accept this as an example of irresponsibly pushing the serotonin theory!

Look. Hundreds of studies have shown some sort of relationship between serotonin and depression. At this point that’s not controversial. What’s controversial is the importance of the relationship, whether it’s causal, whether other things matter more, et cetera. Every single one of Mad In America’s examples has been pretty exemplary in saying that all of these things are still uncertain and need to be investigated further. What more could they do to be more responsible? A total blackout on all news coverage of the new evidence for serotonin’s involvement that keeps coming in?

Ironically, if people had done that, we would have far *less* evidence that depression was not just a simple serotonin deficiency. The most important nail in that theory’s coffin was that tianeptine, a medication that lowers serotonin levels, effectively treats depression. But that’s a study about serotonin of exactly the same sort as the University of Washington study Mad In America complains about! One of the most convincing alternatives to a purely serotonergic picture is the BDNF-neurogenesis theory. But that’s exactly the theory being pushed in the WebMD article Mad In America complains about!

III.

I raised some of these issues in a comment on the Mad in America blog, and author Rob Wipond kindly responded to me:



Rob Wipond *(MIA Author)* on **April 17, 2015 at 12:32 am** said:

Scott Alexander: Your post is extremely long and covers a lot; I do not claim in my post to be providing a complete summary of every issue and point. I also encourage people to read it themselves, which is why I posted a link to it.

Yes, you're right, many psychiatrists, media, pharmaceutical companies and others promoting the serotonin deficiency theory of depression have often included generalized, softening "qualifiers" and "equivocations" such as the ones you quoted, even as they have also made those very bold, unequivocal claims explicitly intended to persuade that I quoted. Taken in full context, then how are such qualifiers different than brash infomercials on television with legal disclaimers like, "not all results will be the same for all people"?

I know of no one (certainly not me) who has ever said that there were never any studies making tenuous, feeble attempts to draw links between serotonin levels and depression in different ways — there were hundreds, I believe (I haven't counted) as the psychiatric community and pharmaceutical industry made enormous efforts to try to prove the theory or buttress its apparent validity in the public eye. And as I note, those are still being produced today. What critics have often correctly pointed out, however, is that the main, strongest argument that psychiatrists have often used in support of the low-serotonin theory has always been that SSRIs allegedly boost serotonin levels. Of course, most of the public has never known that SSRIs have barely beaten placebos in clinical trials, so they've not been able to understand the true spuriousness of even that argument.

I notice that your argument has now changed to "far fewer people" and "less important" psychiatrists made such claims, rather than none at all made such claims. Very well; apparently we would now only potentially disagree on subjective notions such as how many is "fewer" and how unimportant is "less important", rather than disagreeing on the main issue at hand. And then Leo and Lacasse's question becomes all the more significant: Where is the evidence that the "important" psychiatrists were vigorously trying to correct the public record and clarifying that these were only weak hypotheses with no compelling evidence to support them instead of weighty theories with ever mounting evidence to support them?

And that brings us to the critical factual gap in your whole argument: How do you account for the fact that the vast majority of media and the public for many years believed that low serotonin could cause depression? Did they just make it up on their own, without ever consulting any "important" psychiatrists?

Let me address some of these objections piece by piece:

> Yes, you're right, many psychiatrists, media, pharmaceutical companies and others promoting the serotonin deficiency theory of depression have often

> included generalized, softening "qualifiers" and "equivocations" such as the ones you quoted, even as they have also made those very bold, unequivocal claims explicitly intended to persuade that I quoted. Taken in full context, then how are such qualifiers different than brash infomercials on television with legal disclaimers like, "not all results will be the same for all people"?

The very bold, unequivocal claims explicitly intended to persuade that you quoted WERE A SOPHOMORE CHEMISTRY PROJECT, PLUS A BUNCH OF PEOPLE SPECIFICALLY SAYING THAT THESE SHOULD NOT BE TAKEN AS VERY BOLD UNEQUIVOCAL CLAIMS, BUT THEN YOU QUOTED OUT OF CONTEXT TO TAKE THAT PART OUT. AND THE SOPHOMORE CHEMISTRY PROJECT DIDN'T EVEN WIN THE £50 PRIZE.

Okay. Sorry. I shouldn't have yelled like that. More seriously: there are a lot of things we don't totally understand, but which scientific research suggests some weak preliminary theories about. For example, we don't understand fibromyalgia, but if I were writing a textbook on fibromyalgia, or if a patient asked me what it was, then after some appropriate caveats and equivocations, I would say it has something to do with some sort of inflammation in the fascia which causes central sensitization to pain stimuli. Could I end up being totally wrong? Yeah. But at this point I think there's *enough* evidence in this direction that, insofar as it's important to satisfy patients' curiosity about what's going on with them, that it's proper to mention the current best guess. Likewise, if I am a researcher or a scientific publication, I don't think I have some duty to carefully hide my results. A big part of scientific progress is people saying "I just got some small amount of evidence which makes me think it's this" and then other people trying to confirm or refute that with more evidence, until eventually it comes together into a strong theory.

I think researchers and psychiatrists were pretty responsible in coming up with the serotonin deficiency theory. It was inspired by the effectiveness of serotonergic drugs. Then a bunch of studies – Wipond agrees there were hundreds – provided results that seemed to confirm it. Given all of this information, I don't think it was negligent to say that there was quite a bit of evidence pointing to serotonin, as long as you followed this with caveats that the evidence was still preliminary and lots of other things seemed to be involved too. As I've been arguing all along, that's exactly what most people did.

> I know of no one (certainly not me) who has ever said that there were never any studies making tenuous, feeble attempts to draw links between serotonin levels and depression in different ways — there were hundreds, I believe (I haven't counted) as the psychiatric community and pharmaceutical industry made enormous efforts to try to prove the theory or buttress its apparent validity in the public eye. And as I note, those are still being produced today. What critics have often correctly pointed out, however, is that the main, strongest argument that psychiatrists have often used in support of the low-serotonin theory has always been that SSRIs allegedly boost serotonin levels. Of course, most of the public has never known that SSRIs have barely beaten placebos in clinical trials, so they've not been able to understand the true spuriousness of even that argument.

I don't really understand this objection. There was a strong piece of evidence in favor of serotonin in the form of SSRI-effectiveness, scientists pursued that lead by doing hundreds of studies implicating serotonin using different methodologies, most were in favor and so scientists thought the theory had some merit…what exactly is wrong here? This sounds like every scientific theory – Wegener noted that continents looked like they fit together in a way that implied continental drift, geologists did hundreds of other studies that all pointed to continental drift, therefore they started believing in continental drift.

While Wipond may not know of the people saying there was no evidence for serotonin besides SSRI effectiveness, these people certainly exist and provide one of his side's major arguments. Indeed, many of the articles I linked to on my original post made exactly that argument. The BBC said that: "although ideas like the serotonin theory of depression have been widely publicised, scientific research has not detected any reliable abnormalities of the serotonin system in people who are depressed." New York Review Of Books says: "Instead of developing a drug to treat an abnormality, an abnormality was postulated to fit a drug…But the main problem with the theory is that after decades of trying to prove it, researchers have still come up empty-handed".

To learn more about the claim that SSRIs barely beat placebo, see my article on this.

> I notice that your argument has now changed to "far fewer people" and "less important" psychiatrists made such claims, rather than none at all made such claims. Very well; apparently we would now only potentially disagree on subjective notions such as how many is "fewer" and how unimportant is "less important", rather than disagreeing on the main issue at hand. And then Leo and Lacasse's question becomes all the more significant: Where is the evidence that the "important" psychiatrists were vigorously trying to correct the public

> record and clarifying that these were only weak hypotheses with no compelling evidence to support them instead of weighty theories with ever mounting evidence to support them?

In retrospect, “no one has ever said” is a stupid thing for me to have said. I do not deny that a sophomore at University of Bristol once said low serotonin caused depression. And you can find individual psychiatrists who believe a lot of stupid stuff. Some psychiatrists believe in homeopathy. Some psychiatrists believe in reincarnation (the guy in that article conducted my job interview at the University of Virginia. I tried to be very polite.) Some psychiatrists believe that after losing hundreds of thousands of dollars to online Nigerian scammers, it makes perfect sense to give hundreds of thousands *more* dollars to *other* Nigerian scammers, because “these were different Nigerians”. But I will venture to say none of these are consensus positions in the psychiatric community.

And this is why I wanted to continue this discussion here on this blog. If I had selected a set of statements from eminent psychiatrists that had lots of caveats and were extremely responsible, I could be justifiably accused of cherry-picking. Instead, Mad In America selected some statements, probably intending to cherry-pick the other way, but when looked at more closely, they’re all pretty responsible and say exactly what I would have said at the time – SSRIs seem to work, there’s some evidence pointing to serotonin being involved, but the whole thing is terribly complicated. To me, this establishes the consensus position in a way much more clearly than I could have done on my own.

And yes, this consensus position got simplified and distorted. I have no doubt that drug companies drew from it to do exactly the sort of infomercials that Mr. Wipond describes. I have no doubt that individual psychiatrists, when faced with low-functioning patients who are bad at understanding complicated systems but who really wanted to know what was going on, said “serotonin” and left it at that. And I have no doubt that to a public who still largely think evolution means “once upon a time a chimp gave birth to a human baby”, complicated caveats about how serotonin levels are linked to depression but might not cause depression largely went over their heads except for the single word “serotonin”.

(“That’s the happiness molecule! Right?”)

But in general I think my point stands. “Chemical imbalance” as generally used points to a sophisticated model of interacting metabolic pathways which goes far beyond serotonin, and which as far as I know is still very much on the table. While serotonin was justifiably pointed to as a promising candidate early on, it was generally done with appropriate caveats that turned out to be warranted, and the research community has now retreated from some of that earlier language while still considering serotonin a promising lead. And SSRIs continue to be moderately effective antidepressants in the people for whom they are indicated.

(ie somewhere less than half of the people for whom they are prescribed).

Blame Theory

It’s always dangerous to speculate about the hidden psychological motives of people you disagree with – this is the sin of Bulverism. But like most sins, it’s also fun. So please forgive me while I talk about blame.

Many people have remarked on the paradox of an academia made mostly of upper-class ethnic-majority Westerners trying so very hard to find reasons why lots of things are the fault of upper-class ethnic-majority Westerners. The simplest example I can think of is attributing the woes of Third World countries to colonialism; without meaning to trivialize the evils of colonization, a lot of academics seem to go beyond what even the undeniably awful facts can support. Dependency theory, for example, is now mostly discredited, as are a lot of the Marxist perspectives. I would provide other examples if I weren’t satisfied you can generate them independently.

This is on the face of it surprising; naively we would expect people to cast themselves and those like them in as positive a light as possible. Forget about whether these attributions of blame are right or wrong. *Even if they were right* I would not expect people to believe them as enthusiastically as they do.

The theories I’ve heard to explain this paradox are rarely very flattering; usually something about class signaling, or holier-than-thou-ness, or trying to justify the existence of an academic elite.

I want to propose another possibility: what if people are really, fundamentally, good?

Moral philosophy distinguishes between a couple of ethical systems, like deontology, utilitarianism and virtue ethics. Most people without philosophical training settle into a sort of mishmash of all of them, but one which, I think, is closer to deontology than either of the others. Call it Moral Therapeutic Deontology. Like all deontological systems, it focuses on following certain rules: don't murder, don't steal, respect your parents, pay back your debts. Like all deontological systems, other things like charity are "supererogatory", meaning they're nice but not really necessary. If you've got extra time and energy after doing the important stuff, then sure, do the superogatory stuff, whatever, but it's hardly where your moral focus should be.

On the other hand, when confronted with the full extent of human suffering – whether by living in a poor area, or serving in a war zone, or traveling to a Third World country, or treating depression patients – it's hard to think about anything else. The sheer burning horribleness of it becomes this unscratchable itch, this flaw in the world that blots out the sun.

And here's Moral Therapeutic Deontology, saying, "Yeah, helping quench the burning fire of human suffering is *nice*, but it's not like a real thing that real morality should care about. It's not your *duty*."

This is some *heavy* cognitive dissonance. It doesn't match basic intuitions about the importance of the matter. Even worse, it doesn't allow you to communicate the importance of the matter to other people. If you say "Look at all these people living squalid and miserable in the slums without any hope," and they say "Yeah, well, it would be supererogatory to help them and I'm not feeling supererogatory today," you don't really have a leg to stand on.

There’s an easy way to resolve the dissonance without abandoning either Moral Therapeutic Deontology or your concern for the less well-off. That resolution is to prove that human suffering is you and your friends’ fault. Deontology very clearly says that if you cause a problem, it’s your job to help fix it. If you can prove that the reason the Third World is suffering is because of First World white people, you have a strong claim that you as a First World white person should be deeply emotionally invested in solving it; that your friends and neighbors, as First World white people, ought to help you; and that your government, as that of a First World majority-white country, is justified in using taxpayer money to get involved.

I think this might be a part of what’s happening. People feel a need to help the less-advantaged so strongly that they come up with a justification to do so that makes sense in their own moral system, whether it’s factually accurate or not.

I am not as fanatical a partisan of utilitarianism as I used to be, but this still seems like one of the situations where it has an obvious advantage. Utilitarianism tells us that we are perfectly justified in seeing the relief of suffering as a pressing need. We don’t need to justify it by positing facts that may later be proven untrue; it is self-justifying. People sometimes complain that a flaw of utilitarianism is that it implies a heavy moral obligations to help all kinds of people whether or not any of their problems are our fault; the world is divided between those who consider that a bug and those who find it a very helpful feature.

I want more people to become familiar with utilitarianism because I think a lot of the colonialism theory stuff is net hurtful. It combines a justification for helping the poor with an insult to people’s identity, and probably makes the former less palatable to many people than it would be naturally. It also makes our need to help the poor hinge on an empirical point; if that empirical point gets disproved, things become pretty awkward.

This theory implies that utilitarian liberals will have all the features of liberalism *except* the interest in blaming their own group for major problems. My anecdotal experience confirms that. The utilitarians I know are very interested in helping the poor and in various other liberal ideas, but are more likely than other liberals to roll their eyes at talk about colonialism and stereotype threat. I think it’s because they feel confident in their right to care about the disadvantaged regardless.

Universal Love, Said The Cactus Person

"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

"Right," I said. "I'm absolutely in favor of both those things. But before we go any further, could you tell me the two prime factors of 1,522,605,027, 922,533,360, 535,618,378, 132,637,429, 718,068,114, 961,380,688, 657,908,494 ,580,122,963, 258,952,897, 654,000,350, 692,006,139?

"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

The sea was made of strontium; the beach was made of rye. Above my head, a watery sun shone in an oily sky. A thousand stars of sertraline whirled round quetiapine moons, and the sand sizzled sharp like cooking oil that hissed and sang and threatened to boil the octahedral dunes.

"Okay," I said. "Fine. Let me tell you where I'm coming from. I was reading Scott McGreal's blog, which has some good articles about so-called DMT entities, and mentions how they seem so real that users of the drug insist they've made contact with actual superhuman beings and not just psychedelic hallucinations. You know, the usual Terence McKenna stuff. But in one of them he mentions a paper by Marko Rodriguez called *A Methodology For Studying Various Interpretations of the N,N-dimethyltryptamine-Induced Alternate Reality*, which suggested among other things that you could prove DMT entities were real by taking the drug and then asking the entities you meet to factor large numbers which you were sure you couldn't factor yourself. So to that end, could you do me a big favor and tell me the factors of 1,522,605,027, 922,533,360, 535,618,378, 132,637,429, 718,068,114, 961,380,688, 657,908,494, 580,122,963, 258,952,897, 654,000,350, 692,006,139?

"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

The sea turned hot and geysers shot up from the floor below. First one of wine, then one of brine, then one more yet of turpentine, and we three stared at the show.

"I was afraid you might say that. Is there anyone more, uh, *verbal* here whom I could talk to?"

"Universal love," said the cactus person.

At the sound of that, the big green bat started rotating in place. On its other side was a bigger greener bat, with a ancient, wrinkled face.

"*Not splitting numbers / but joining Mind,"* it said.
Not facts or factors or factories / but contact with the abstract attractor that brings you back to me
Not to seek / but to find"

"I don't follow," I said.

"*Not to follow / but to jump forth into the deep*
Not to grind or to bind or to seek only to find / but to accept
Not to be kept / but to wake from sleep"

The bat continued to rotate, until the first side I had seen swung back into view.

“Okay,” I said. “I’m going to hazard a guess as to what you’re talking about, and you tell me if I’m right. You’re saying that, like, all my Western logocentric stuff about factoring numbers in order to find out the objective truth about this realm is missing the point, and I should be trying to do some kind of spiritual thing involving radical acceptance and enlightenment and such. Is that kind of on the mark?”

“Universal love,” said the cactus person.

“Transcendent joy,” said the big green bat.

“Frick,” I said. “Well, okay, let me continue.” The bat was still rotating, and I kind of hoped that when the side with the creepy wrinkled face came into view it might give me some better conversation. “I’m all about the spiritual stuff. I wouldn’t be here if I weren’t deeply interested in the spiritual stuff. This isn’t about money or fame or anything. I want to advance psychedelic research. If you can factor that number, then it will convince people back in the real – back in my world that this place is for real and important. Then lots of people will take DMT and flock here and listen to what you guys have to say about enlightenment and universal love, and make more sense of it than I can alone, and in the end we’ll have more universal love, and…what was the other thing?”

“Transcendent joy,” said the big green bat.

“Right,” I said. “We’ll have more transcendent joy if you help me out and factor the number than if you just sit there being spiritual and enigmatic.”

“Lovers do not love to increase the amount of love in the world / But for the mind that thrills
And the face of the beloved, which the whole heart fills / the

heart and the art never apart, ever unfurled
And John Stuart is one of / the dark satanic mills"

"I take it you're not consequentialists," I said. "You know that's really weird, right. Like, not just 'great big green bat with two faces and sapient cactus-man' weird, but like *really* weird. You talk about wanting this spiritual enlightenment stuff, but you're not going to take actions that are going to increase the amount of spiritual enlightenment? You've got to understand, this is like a bigger gulf for me than normal human versus ineffable DMT entity. You can have crazy goals, I expect you to have crazy goals, but what you're saying now is that you don't pursue any goals at all, you can't be modeled as having desires. Why would you *do* that?"

"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

"Now you see here," I said. "Everyone in this conversation is in favor of universal love and transcendent joy. But I've seen the way this works. Some college student gets his hands on some DMT, visits here, you guys tell him about universal love and transcendent joy, he wakes up, says that his life has been changed, suddenly he truly understands what really matters. But it never lasts. The next day he's got to get up and go to work and so on, and the universal love lasts about five minutes until his boss starts yelling at him for writing his report in the wrong font, and before you know it twenty years later he's some slimy lawyer who's joking at a slimy lawyer party about the one time when he was in college and took some DMT and spent a whole week raving about transcendent joy, and all the other slimy lawyers laugh, and he laughs with them, and so much for whatever spiritual awakening you and your colleagues in LSD and peyote are trying to kindle in humanity. And if I accept your message of universal love and transcendent joy right now, that’s exactly what’s going to happen to me, and meanwhile human civilization is going to keep being stuck in greed and ignorance and misery. So how about you shut up about universal love and you factor my number for me so we can start figuring out a battle plan for giving humanity a *real* spiritual revolution?”

“Universal love,” said the cactus person.

“Transcendent joy,” said the big green bat.

A meteorite of pure delight struck the sea without a sound. The force of the blast went rattling past the bat and the beach, disturbing each, then made its way to a nearby bay of upside-down trees with their roots in the breeze and their branches underground.

“I demand a better answer than that,” I demanded.

The other side of the bat spun into view.

“Chaos never comes from the Ministry of Chaos / nor void from the Ministry of Void
Time will decay us but time can be left blank / destroyed
With each Planck moment ever fit / to be eternally enjoyed”

“You’re making this basic mistake,” I told the big green bat. “I honestly believe that there’s a perspective from which Time doesn’t matter, where a single moment of recognition is equivalent to eternal recognition. The problem is, if you only have that perspective for a moment, then all the rest of the time, you’re sufficiently stuck in Time to honestly believe you’re stuck in Time. It’s like that song about the hole in the bucket – if the hole in the bucket were fixed, you would have the materials needed to fix the hole in the bucket. But since it isn't, you don't. Likewise, if I understood the illusoriness…illusionality…whatever, of time, then I wouldn't care that I only understood it for a single instant. But since I don't, I don't. Without a solution to the time-limitedness of enlightenment that works from *within* the temporal perspective, how can you consider it solved at all?"

"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

The watery sun began to run and it fell on the ground as rain. It became a dew that soaked us through, and as the cold seemed to worsen the cactus person hugged himself to stay warm but his spines pierced his form and he howled in a fit of pain.

"Or maybe you guys are so intoxicated on spiritual wisdom that you couldn't think straight if your life depended on it. Maybe your random interventions in our world and our minds look like the purposeless acts of a drunken madman because that's basically more or less what they are. Maybe if you had like five IQ points between the two of you, you could tap into your cosmic consciousness or whatever to factor a number that would do more for your cause than all your centuries of enigmatic dreams and unasked-for revelations combined, but you ARE TOO DUMB TO DO IT EVEN WHEN I BASICALLY HOLD YOUR HAND THE WHOLE WAY. Your spine. Your wing. Whatever."

"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

"Fuck you," said I.

I saw the big green bat bat a green big eye. Suddenly I knew I had gone too far. The big green bat started to turn around what was neither its x, y, or z axis, slowly rotating to reveal what was undoubtedly the biggest, greenest bat that I had ever seen, a bat bigger and greener than which it was impossible to conceive. And the bat said to me:

“Sir. Imagine you are in the driver’s seat of a car. You have been sitting there so long that you have forgotten that it is the seat of a car, forgotten how to get out of the seat, forgotten the existence of your own legs, indeed forgotten that you are a being at all separate from the car. You control the car with skill and precision, driving it wherever you wish to go, manipulating the headlights and the windshield wipers and the stereo and the air conditioning, and you pronounce yourself a great master. But there are paths you cannot travel, because there are no roads to them, and you long to run through the forest, or swim in the river, or climb the high mountains. A line of prophets who have come before you tell you that the secret to these forbidden mysteries is an ancient and terrible skill called GETTING OUT OF THE CAR, and you resolve to learn this skill. You try every button on the dashboard, but none of them is the button for GETTING OUT OF THE CAR. You drive all of the highways and byways of the earth, but you cannot reach GETTING OUT OF THE CAR, for it is not a place on a highway. The prophets tell you GETTING OUT OF THE CAR is something fundamentally different than anything you have done thus far, but to you this means ever sillier extremities: driving backwards, driving with the headlights on in the glare of noon, driving into ditches on purpose, but none of these reveal the secret of GETTING OUT OF THE CAR. The prophets tell you it is easy; indeed, it is the easiest thing you have ever done. You have traveled the Pan-American Highway from the boreal pole to the Darien Gap, you have crossed Route 66 in the dead heat of summer, you have outrun cop cars at 160 mph and survived, and GETTING OUT OF THE CAR is easier than any of them, the easiest thing you can imagine, closer to you than the veins in your head, but still the secret is obscure to you."

A herd of bison came into listen, and voles and squirrels and ermine and great tusked deer gathered round to hear as the bat continued his sermon.

"And finally you drive to the top of the highest peak and you find a sage, and you ask him what series of buttons on the dashboard you have to press to get out of the car. And he tells you that it's not about pressing buttons on the dashboard and you just need to GET OUT OF THE CAR. And you say okay, fine, but what series of buttons will *lead to* you getting out of the car, and he says no, really, you need to stop thinking about dashboard buttons and GET OUT OF THE CAR. And you tell him maybe if the sage helps you change your oil or rotates your tires or something then it will improve your driving to the point where getting out of the car will be a cinch after that, and he tells you it has nothing to do with how rotated your tires are and you just need to GET OUT OF THE CAR, and so you call him a moron and drive away."

"Universal love," said the cactus person.

"So that metaphor is *totally unfair*," I said, "and a better metaphor would be if every time someone got out of the car, five minutes later they found themselves back in the car, and I ask the sage for driving directions to a laboratory where they are studying that problem, and…"

"You only believe that because it's written on the windshield," said the big green bat. "And you think the windshield is identical to reality because you won’t GET OUT OF THE CAR.”

“Fine,” I said. “Then I can’t get out of the car. I want to get out of the car. But I need help. And the first step to getting help is for you to factor my number. You seem like a reasonable person. Bat. Freaky DMT entity. Whatever. Please. I promise you, this is the right thing to do. Just factor the number.”

“And I promise you,” said the big green bat. “You don’t need to factor the number. You just need to GET OUT OF THE CAR.”

“I can’t get out of the car until you factor the number.”

“I won’t factor the number until you get out of the car.”

“Please, I’m begging you, factor the number!”

“Yes, well, I’m begging you, please get out of the car!”

“FOR THE LOVE OF GOD JUST FACTOR THE FUCKING NUMBER!”

“FOR THE LOVE OF GOD JUST GET OUT OF THE FUCKING CAR!”

“FACTOR THE FUCKING NUMBER!”

“GET OUT OF THE FUCKING CAR!”

“Universal love,” said the cactus person.

Then tree and beast all fled due east and the moon and stars shot south. And the bat rose up and the sea was a cup and the earth was a screen green as clozapine and the sky a voracious mouth. And the mouth opened wide and the earth was skied and the sea fell in with an awful din and the trees were moons and the sand in the dunes was a blazing comet and…

I vomited, hard, all over my bed. It happens every time I take DMT, sooner or later; I’ve got a weak stomach and I’m not sure the stuff I get is totally pure. I crawled just far enough out of bed to flip a light switch on, then collapsed back onto the soiled covers. The clock on the wall read 11:55, meaning I’d been out about an hour and a half. I briefly considered taking some more ayahuasca and heading right back there, but the chances of getting anything more out of the big green bat, let alone the cactus person, seemed small enough to fit in a thimble. I drifted off into a fitful sleep.

Behind the veil, across the infinite abyss, beyond the ice, beyond daath, the dew rose from the soaked ground and coalesced into a great drop, which floated up into an oily sky and became a watery sun. The cactus person was counting on his spines.

“Hey,” the cactus person finally said, “just out of curiosity, was the answer 37,975,227, 936,943,673, 922,808,872, 755,445,627, 854,565,536, 638,199 times 40,094,690,950, 920,881,030, 683,735,292, 761,468,389, 214,899,724,061?”

“Yeah,” said the big green bat. “That’s what I got too.”

Growth Mindset 3: A Pox On Growth Your Houses

[EDIT: The author of this paper has responded; I list his response here.]

Jacques Derrida proposed a form of philosophical literary criticism called deconstruction. I'll be the first to admit I don't really understand it, but it seems to have something to do with assuming all texts secretly contradict their stated premise and apparent narrative, then hunting down and exposing the plastered-over areas where the author tries to hide this.

I have no idea whether this works for literature or not, but it's a useful way to read scientific papers.

Consider a popular field – or, at least, a field where *a certain position* is popular. For example, we've been talking a lot about growth mindset recently. There seem to be a lot of researchers working to prove growth mindset and not a lot working to disprove it. Journals are pretty interested in studies showing growth mindset interventions work, and maybe not so interested in studies showing they don't. I'll admit that my strong suspicions of publication bias don't seem to be borne out by the facts here – see this meta-analysis – but I bet its more sinister cousin "all experimenters believe the same thing and have the same experimenter effects" bias is alive and well.

In a field like that, you're not going to get the contrarian studies you want, but one way to find the other side of the issue is to look a little more closely at the studies that do get published, the ones that say they're in support of the thesis, and see if you can find anything incriminating.

Here’s a perfect example: Mindset Interventions Are A Scalable Treatment For Academic Underachievement, by a team of six researchers including Carol Dweck.

The abstract reads:

> The efficacy of academic-mind-set interventions has been demonstrated by small-scale, proof-of-concept interventions, generally delivered in person in one school at a time. Whether this approach could be a practical way to raise school achievement on a large scale remains unknown. We therefore delivered brief growth-mind-set and sense-of-purpose interventions through online modules to 1,594 students in 13 geographically diverse high schools. Both interventions were intended to help students persist when they experienced academic difficulty; thus, both were predicted to be most beneficial for poorly performing students. This was the case. Among students at risk of dropping out of high school (one third of the sample), each intervention raised students’ semester grade point averages in core academic courses and increased the rate at which students performed satisfactorily in core courses by 6.4 percentage points. We discuss implications for the pipeline from theory to practice and for education reform.

This sounds really, really impressive! It’s hard to imagine any stronger evidence in growth mindset’s favor.

And then you make the mistake of reading the actual paper.

The paper asked a 1,594 students from a bunch of different high schools to take a 45 minute online course.

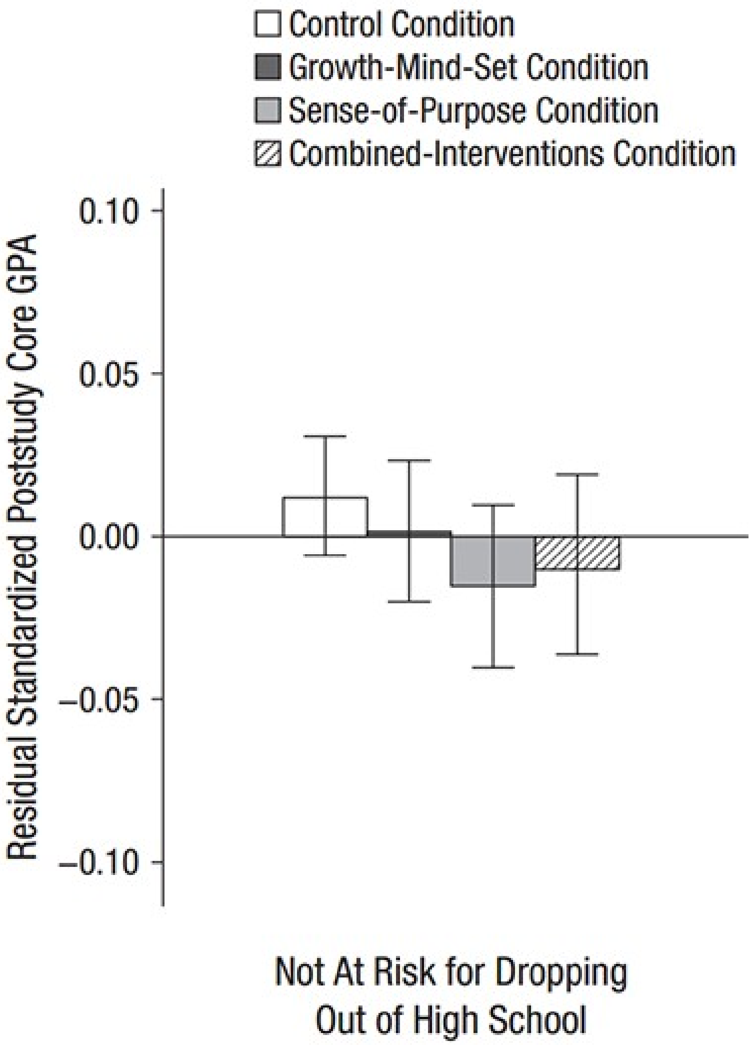
A quarter of the students took a placebo course that just presented some science about how different parts of the brain do different stuff.

Another quarter took a course that was supposed to teach growth mindset.

Still another quarter took a course about “sense of purpose” which talked about how schoolwork was meaningful and would help them accomplish lots of goals and they should be happy to do it. This was also classified as a “mindset intervention”, though it seems pretty different.

And the final quarter took both the growth mindset course *and* the “sense of purpose” course.

Then they let all students continue taking their classes for the rest of the semester and saw what happened, which was this:

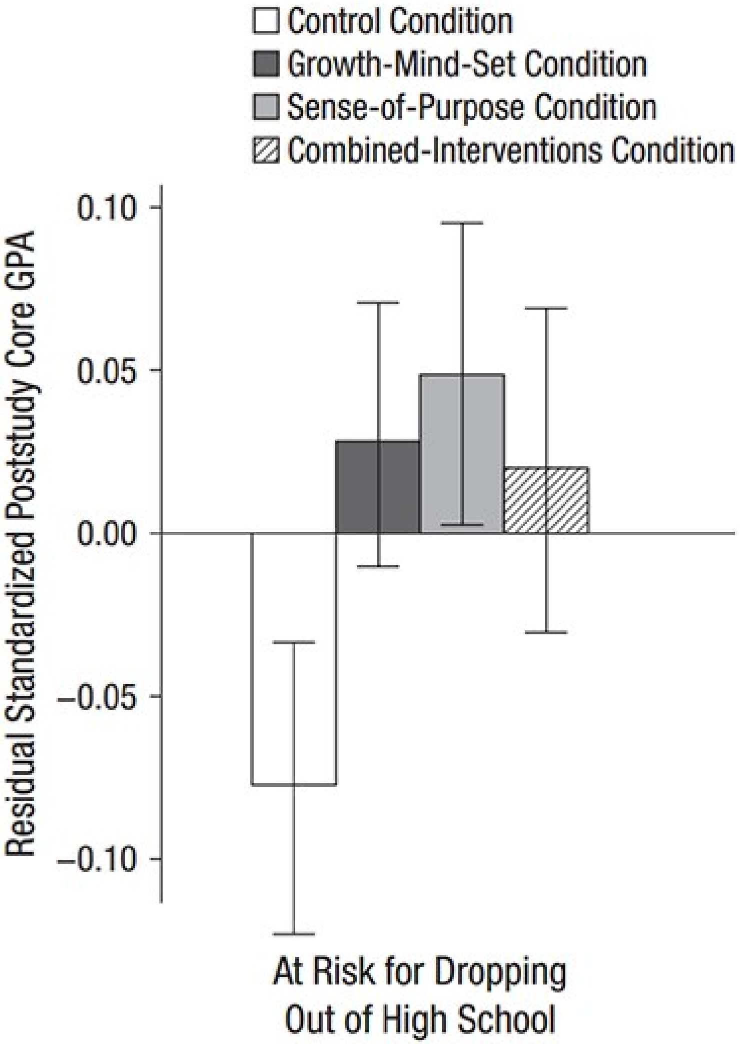


[EDIT: I totally bungled these graphs! See discussion of exactly how on the author's reply above, without which the information below will be misleading at best]

Among ordinary students, the effect on the growth mindset group was completely indistinguishable from zero, and in fact they did nonsignificantly worse than the control group. This was the most basic test they performed, and it should have been the headline of the study. The study should have been titled "Growth Mindset Intervention Totally Fails To Affect GPA In Any Way".

Instead they went to subgroup analysis. Subgroup analysis can be useful to find more specific patterns in the data, but if it's done post hoc it can lead to what I previously called the Elderly Hispanic Woman Effect, after medical papers that can't find their drug has any effect on people at large, so they keep checking different subgroups – young white men… nothing. Old black men…nothing. Middle-aged Asian transgender people…nothing. Newborn Australian aboriginal butch lesbians…nothing. Elderly Hispanic women…p = 0.049…aha! And the study gets billed as "Scientists Find Exciting New Drug That Treats Diabetes In Elderly Hispanic Women."

As per the abstract, the researchers decided to focus on an "at risk" subgroup because they had principled reasons to believe mindset interventions would work better on them. In their subgroup of 519 students who had a GPA of 2.0 or less last semester, or who failed one or more academic courses last semester:



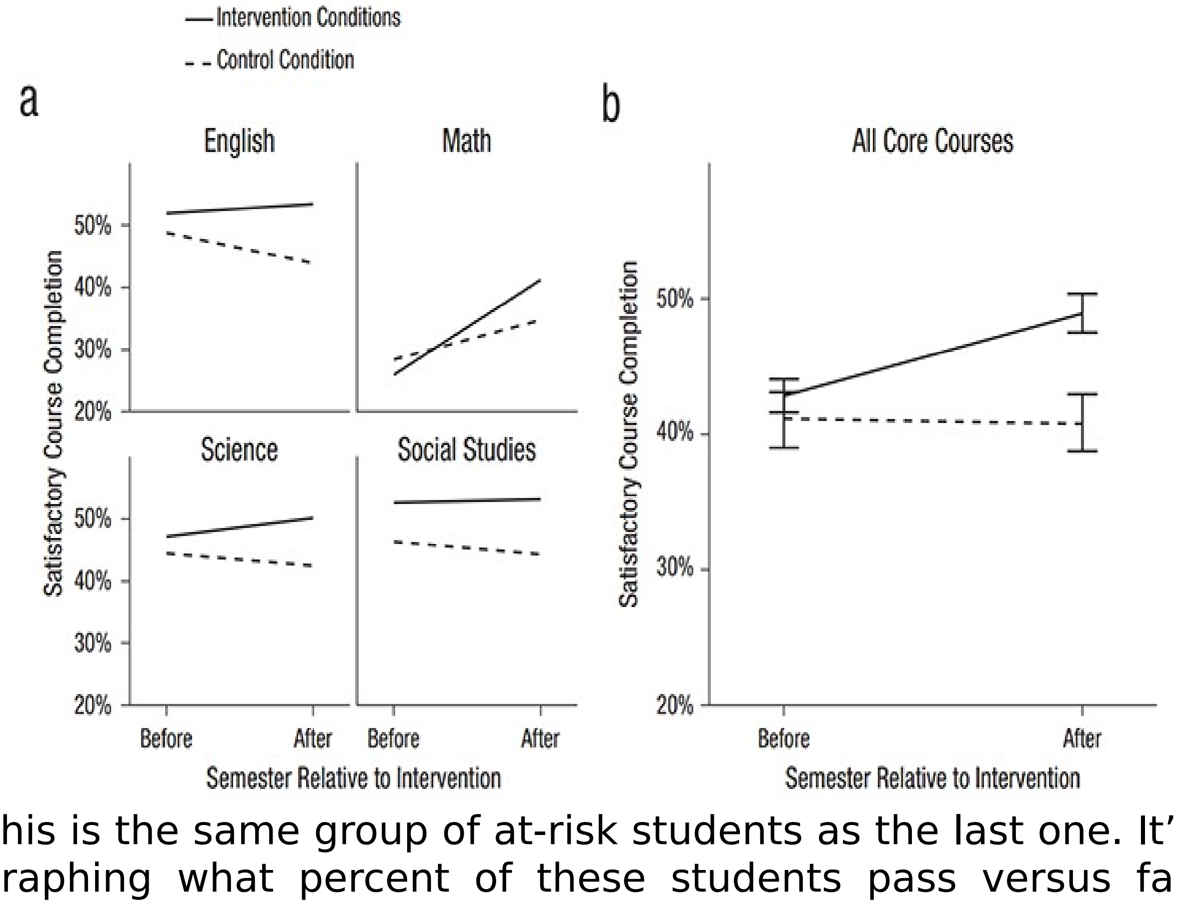
Growth mindset *still* doesn't differ from zero. And growth mindset does nonsignificantly *worse* than their "sense of purpose" intervention where they tell children to love school. In fact, the students who take both "sense of purpose" and growth mindset actually do (nonsignificantly) *worse* than sense-of-purpose alone!

But the control group mysteriously started doing much worse in all their classes right after the study started, so growth mindset is significantly better than the control group. Hooray!

Why would the control group's GPA suddenly decline? The simplest answer would be that by coincidence the class got harder right after the study started, and only the intervention kids were resilient enough to deal with it – but that can't be right, because this was done at eleven different schools, and they wouldn't have all had their coursework get harder at the same time.

Another possibility is that sufficiently low-functioning kids are *always* declining – that is, as time goes on they get more and more behind in their coursework, so their grades at time t+1 are always less than at time t, and maybe growth mindset has arrested this decline. This is plausible and I'd be interested in seeing if other studies have found this.

Perhaps aware that this is not very convincing, the authors go on to do another analysis, this one of percent of students passing their classes.



This is the same group of at-risk students as the last one. It's graphing what percent of these students pass versus fail their courses. The graph on th left shows that a significantly higher number of students in the intervention conditions pass their courses than in the control condition.

This is better, but one part still concerns me.

Did you catch that phrase "intervention conditions"? The authors of the study write: "Because our primary research question concerned the efficacy of academic mindset interventions in general when delivered via online modules, we then collapsed the intervention conditions into a single intervention dummy code (0 = control, 1 = intervention).

We don't know whether growth mindset did anything for even these students in this little subgroup, because it was collapsed together with the (more effective) "sense of purpose" intervention before any of these tests were done. I don't know if this is just for convenience, or if it is to obfuscate that it didn't work on its own.

*[**EDIT:** Scott McGreal looks further and finds in the supplementary material that growth mindset alone did NOT significantly improve pass rates!]*

The abstract of this study tells you none of this. It just says: "Mindset Interventions Are A Scalable Treatment For Academic Overachievement…Among students at risk of dropping out of high school (one third of the sample), each intervention raised students' semester grade point averages in core academic courses and increased the rate at which students performed satisfactorily in core courses by 6.4 percentage points" From the abstract, this study is a triumph.

But my own summary of these results, as relevant to growth mindset is as follows:

For students with above a 2.0 GPA, a growth mindset intervention did nothing.

For students with below a 2.0 GPA, the growth mindset interventions may not have improved GPA, but may have prevented GPA from falling, which for some reason it was otherwise going to do.

Even in those students, it didn't do any better than a "sense-of-purpose" intervention where children were told platitudes about how doing well in school will "make their families proud" and "make a positive impact".

In no group of students did it significantly increase chance of passing any classes.

Haishan writes:

> If ye read only the headlines, what reward have ye? Do not even the policymakers the same? And if ye take the abstract at its face, what do ye more than others? Do not even the science journalists so?"

Titles, abstracts, and media presentations are where authors can decide how to report a bunch of different, often contradictory results in a way that makes it look like they have completely proven their point. A careful look at the study may find that their emphasis is misplaced, and give you more than enough ammunition against a theory even where the stated results are glowingly positive.

The only reason we were told these results is that they were in the same place as a "sense of purpose mindset" intervention that looked a little better, so it was possible to publish the study and claim it as a victory for mindsets in general. How many studies that show similar results for growth mindset lack a similar way of spinning the data, and so never get seen at all?

Nefarious Nefazodone And Flashy Rare Side Effects

[Epistemic status: I am still in training. I am not an expert on drugs. This is poorly-informed speculation about drugs and it should not be taken seriously without further research. Nothing in this post is medical advice.]

I.

Which is worse – ruining ten million people’s sex lives for one year, or making one hundred people’s livers explode?

I admit I sometimes use this blog to speculate about silly moral dilemmas for no reason, but that’s not what’s happening here. This is a real question that I deal with on a daily basis.

SSRIs, the class which includes most currently used antidepressants, are very safe in the traditional sense of “unlikely to kill you”. Suicidal people take massive overdoses of SSRIs all the time, and usually end up with little more than a stomachache for their troubles. On the other hand, there’s increasing awareness of very common side effects which, while not disabling, can be pretty unpleasant. About 50% of users report decreased sexual abilities, sometimes to the point of total loss of libido or anorgasmia. And something like 25% of users experience “emotional blunting” and the loss of ability to feel feelings normally.

Nefazodone (brand name Serzone®, which would also be a good brand name for a BDSM nightclub) is an equally good (and maybe better) antidepressant that does not have these side effects. On the other hand, every year, one in every 300,000 people using nefazodone will go into “fulminant hepatic failure”, which means their liver suddenly and spectacularly stops working and they need a liver transplant or else they die.

There are a lot of drug rating sites, but the biggest is Drugs.com. 467 Drugs.com users have given Celexa, a very typical SSRI, an average rating of 7.8/10. 14 users have given nefazodone an average rating of 9.1/10.

CrazyMeds might not be as dignified as Drugs.com, but they have a big and well-educated user base and they're psych-specific. Their numbers are 3.3/5 (n = 253) for Celexa and 4.1/5 (n = 47) for nefazodone.

So both sites' users seem to agree that nefazodone is notably better than Celexa, in terms of a combined measure of effectiveness and side effects.

But nefazodone is practically never used. It's actually illegal in most countries. In the United States, parent company Bristol-Myers Squibb (which differs from normal Bristol-Myers in that it was born without innate magical ability) withdrew it from the market, and the only way you can find it nowadays is to get it is from an Israeli company that grabbed the molecule after it went off-patent. In several years working in psychiatry, I have never seen a patient on nefazodone, although I'm sure they exist somewhere. I would estimate its prescription numbers are about 1% of Celexa's, if that.

The problem is the hepatic side effects. Nobody wants to have their liver explode.

But. There are something like thirty million people in the US on antidepressants. If we put them all on nefazodone, that's about a hundred cooked livers per year. If we put them all on SSRIs, at least ten million of them will get sexual side effects, plus some emotional blunting.

My life vastly improved when I learned there was a searchable database of QALYs for different conditions. It doesn't have SSRI-induced sexual dysfunction, but it does have sexual dysfunction due to prostate cancer treatment, and I assume that sexual dysfunction is about equally bad regardless of what causes it. Their sexual dysfunction has some QALY weights averaging about 0.85. Hm.

Assume everyone with fulminant liver failure dies. That's not true; some get liver transplants, maybe some even get a miracle and recover. But assume everyone dies – and further, they die at age 30, cutting their lives short by fifty years.

In that case, putting all depressed people on nefazodone for a year costs 5,000 QALYs, but putting all depressed people on SSRIs for a year costs 1,500,000 QALYs. The liver failures may be flashier, but the 3^^^3 dust specks worth of poor sex lives add up to more disutility in the end.

I don't want to overemphasize this particular calculation for a couple of reasons. First, SSRIs and nefazodone both have other side effects besides the major ones I've focused on here. Second, I don't know if the level of SSRI-induced sexual dysfunction is as bad as the prostate-surgery-induced sexual dysfunction on the database. Third, there are a whole bunch of antidepressants that are neither SSRIs nor nefazodone and which might be safer than either.

But I *do* want to emphasize this pattern, because it recurs again and again.

II.

In that spirit, which would you rather have – something like a million people addicted to amphetamines, or something like ten people have their skin eat itself from the inside?

I can’t get good numbers on how many adults abuse Adderall, but a quick glance at the roster for my hospital’s rehab unit suggests “a lot”. Huffington Post calls it the most abused prescription drug in America, which sounds about right to me. Honestly there are worse things to be addicted to than Adderall, but it’s not completely without side effects. The obvious ones are anxiety, irritability, occasionally frank psychosis, and sometimes heart problems – but a lot of the doctors I work with go beyond what the research can really prove and suggest it can produce lasting negative personality change and predispose people to other forms of addictive and impulsive behavior.

If you’ve got to give adults a stimulant, I would much prefer modafinil. It’s not addictive, it lacks most of Adderall’s side effects, and it works pretty well. I’ve known many people on modafinil and they give it pretty universally positive reviews.

On the other hand, modafinil *may or may not* cause a skin reaction called Stevens Johnson Syndrome/Toxic Epidermal Necrolysis, which like most things with both “toxic” and “necro” in the name is really really bad. The original data suggesting a connection came from kids, who get all sorts of weird drug effects that adults don’t, but since then some people have *claimed* to have found a connection with adults. Some people get SJS anyway just by bad luck, or because they’re taking other drugs, so it’s really hard to attribute cases specifically to modafinil.

Gwern’s Modafinil FAQ mentions an FDA publication which argues that the background rate of SJS/TEN is 1-2 per million people per year, but the modafinil rate is about 6 per million people per year. However, there are only three known cases of a person above age 18 on modafinil getting SJS/TEN, and this might not be different from background rates after all.

Overall the evidence that modafinil increases the rate of SJS/TEN in adults at all is pretty thin, and if it does, it’s as rare as hen’s teeth (in fact, very close to the same rate as liver failure from nefazodone).

(also: consider that like half of Silicon Valley is on modafinil, yet San Francisco Bay is not yet running red with blood.)

(also: ibuprofen is linked to SJS/TEN, with about the same odds ratio as modafinil, but nobody cares, and they are correct not to care.)

I said I’ve never seen a doctor prescribe nefazodone in real life; I can’t say that about modafinil. I have seen one doctor prescribe modafinil. It happened like this: a doctor I was working with was very upset, because she had an elderly patient with very low energy for some reason, I can’t remember, maybe a stroke, and wanted to give him Adderall, but he had a heart arrythmia and Adderall probably wouldn’t be safe for him.

I asked “What about modafinil?”

She said, “Modafinil? Really? But doesn’t that sometimes cause Stevens Johnson Syndrome?”

And then I glared at her until she gave in and prescribed it.

But this is very, very typical. Doctors who give out Adderall like candy have no associations with modafinil except “that thing that sometimes causes Stevens-Johnson Syndrome” and are afraid to give it to people.

III.

Nefazodone and modafinil are far from the only examples of this pattern. MAOIs are like this too. So is clozapine. If I knew more about things other than psychiatry, I bet I could think of examples from other fields of medicine.

And partially this is natural and understandable. Doctors swear an oath to "first do no harm", and toxic epidermal necrolysis is pretty much the epitome of harm. Thought experiments like torture vs dust specks suggest that most people's moral intuitions say that *no* amount of aggregated lesser harms like sexual side effects and amphetamine addictions can equal the importance of avoiding even a tiny chance of some great harm like liver failure or SJS/TEN. Maybe your doctor, if you asked her directly, would endorse a principled stance of "I am happy to give any number of people anxiety and irritability in order to avoid even the smallest chance of one case of toxic epidermal necrolysis."

And yet.

The same doctors who would never *dare* give nefazodone, consider Seroquel a perfectly acceptable second-line treatment for depression. Along with other atypical antipsychotics, Seroquel raises the risk of sudden cardiac death by about 50%. The normal risk of cardiac sudden death in young people is about 10 in 100,000 per year, so if my calculations are right, low-dose Seroquel causes an extra cardiac death once per every 20,000 patient-years. That's ten times as often as nefazodone causes an extra liver death.

Yet nefazodone was taken off of the market by its creators and consigned to the dustbin of pharmacological history, and Seroquel is the sixth-best-selling drug in the United States, commonly given for depression, simple anxiety, and sometimes even to help people sleep.

Why the disconnect? Here's a theory: sudden cardiac death happens all the time; sometimes God just has it in for you and your heart stops working and you die. Antipsychotics can increase the chances of that happening, but it's a purely statistical increase, such that we can detect it aggregated over large groups but never be sure that it played a role in any particular case. The average person who dies of Seroquel never knows they died of Seroquel, but the average person who dies from nefazodone is easily identified as a nefazodone-related death. So nefazodone gets these big stories in the media about this young person who died by taking this exotic psychiatric drug, and it becomes a big deal and scares the heck out of everybody. When someone dies of Seroquel, it's just an "oh, so sad, I guess his time has come."

But the end result is this. When treatment with an SSRI fails, nefazodone and Seroquel naively seem to be equally good alternatives. Except nefazodone has a death rate of 1/300,000 patient years, and Seroquel 1/20,000 patient years. And yet everyone stays the hell away from the nefazodone because it's known to be unsafe, and chooses the Seroquel.

I conclude either doctors are terrible at thinking about risk, or else maybe a little *too* good at thinking about risk.

I bring up the latter option because there's a principal-agent problem going on here. Doctors want to do what's best for their patients. But they also want to do what's best for themselves, which means not getting sued. No one has ever sued their doctor because they got a sexual side effect from SSRIs, but if somebody dies because they're the lucky 1/300,000 who gets liver failure from nefazodone, you can bet their family's going to sue. Suddenly it's not a matter of comparing QALYs, it’s a matter of comparing zero percent chance of lawsuit with non-zero percent chance of lawsuit.

(Fermi calculation: if a doctor has 100 patients at a time on antidepressants, and works for 30 years, then if she uses Serzone as her go-to antidepressant, she’s risking a 1% chance of getting the liver failure side effect once in her career. That’s small, but since a single bad lawsuit can bankrupt a doctor, it’s worth taking seriously.)

And that would be a tough lawsuit to fight. “Yes, Your Honor, I knew when I prescribed this drug that it sometimes makes people’s livers explode, but the alternative often gives people a bad sex life, and according to the theory of utilitarianism as propounded by 18th century philosopher Jeremy Bentham – ” … “Bailiff, club this man”.

And the same facet of nefazodone that makes it exciting for the media makes it exciting for lawsuits. When someone dies of nefazodone toxicity, everyone knows. When someone dies of Seroquel, “oh, so sad, I guess his time has come”.

That makes Seroquel a lot safer than nefazodone. Safer for the doctor, I mean. The *important* kind of safer.

This is why, as I mentioned before, I hate lawsuits as a de facto regulatory mechanism. Our de jure regulatory mechanism, the FDA, is pretty terrible, but to its credit it hasn’t banned nefazodone. One time it banned clozapine because of a flashy rare side effect, but everyone yelled at them and they apologized and changed their mind. With lawsuits there’s nobody to yell at, so we just end up with people very quietly adjusting their decisions in the shadows and nobody else being any the wiser.

I don’t want to overemphasize this, because I think it’s only one small part of the problem. After all, a lot of countries withdrew nefazodone entirely and didn’t even give lawsuits a chance to enter the picture.

But whatever the cause, the end result is that drugs with rare but spectacular side effects get consistently underprescribed relative to drugs with common but merely annoying side effects, or drugs that have more side effects but manage to hide them better.

Prescriptions, Paradoxes, and Perversities

[WARNING: I am not a pharmacologist. I am not a researcher. I am not a statistician. This is not medical advice. This is really weird and you should not take it too seriously until it has been confirmed]

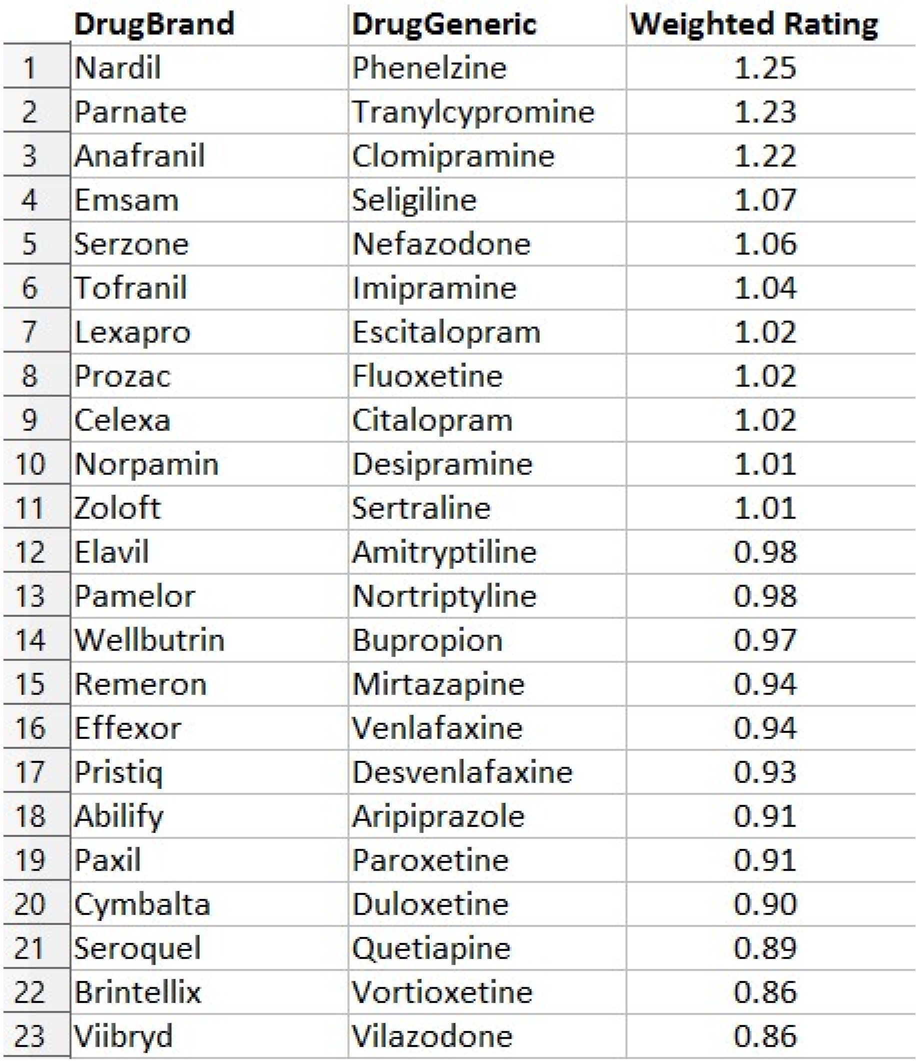
I.

I’ve been playing around with data from Internet databases that aggregate patient reviews of medications.

Are these any good? I looked at four of the largest such databases – Drugs.com, WebMD, AskAPatient, and DrugLib – as well as psychiatry-specific site CrazyMeds – and took their data on twenty-three major antidepressants. Then I correlated them with one another to see if the five sites mostly agreed.

Correlations between Drugs.com, AskAPatient, and WebMD were generally large and positive (around 0.7). Correlations between CrazyMeds and DrugLib were generally small or negative. In retrospect this makes sense, because these two sites didn’t allow separation of ratings by condition, so for example Seroquel-for-depression was being mixed with Seroquel-for-schizophrenia.

So I threw out the two offending sites and kept Drugs.com, AskAPatient, and WebMD. I normalized all the data, then took the weighted average of all three sites. From this huge sample (the least-reviewed drug had 35 ratings, the most-reviewed drug 4,797) I obtained a unified opinion of patients’ favorite and least favorite antidepressants.



	DrugBrand	DrugGeneric	Weighted Rating
1	Nardil	Phenelzine	1.25
2	Parnate	Tranylcypromine	1.23
3	Anafranil	Clomipramine	1.22
4	Emsam	Seligiline	1.07
5	Serzone	Nefazodone	1.06
6	Tofranil	Imipramine	1.04
7	Lexapro	Escitalopram	1.02
8	Prozac	Fluoxetine	1.02
9	Celexa	Citalopram	1.02
10	Norpamin	Desipramine	1.01
11	Zoloft	Sertraline	1.01
12	Elavil	Amitryptiline	0.98
13	Pamelor	Nortriptyline	0.98
14	Wellbutrin	Bupropion	0.97
15	Remeron	Mirtazapine	0.94
16	Effexor	Venlafaxine	0.94
17	Pristiq	Desvenlafaxine	0.93
18	Abilify	Aripiprazole	0.91
19	Paxil	Paroxetine	0.91
20	Cymbalta	Duloxetine	0.90
21	Seroquel	Quetiapine	0.89
22	Brintellix	Vortioxetine	0.86
23	Viibryd	Vilazodone	0.86

This doesn’t surprise me at all. Everyone secretly knows Nardil and Parnate (the two commonly-used drugs in the MAOI class) are excellent antidepressants[1]. Oh, nobody will prescribe them, because of the dynamic discussed here, but in their hearts they know it’s true.

Likewise, I feel pretty good to see that Serzone, which I recently defended, is number five. I’ve had terrible luck with Viibryd, and it just seems to make people taking it more annoying, which is not a listed side effect but which I swear has happened.

The table also matches the evidence from chemistry – drugs with similar molecular structure get similar ratings, as do drugs with similar function. This is, I think, a good list.

Which is too bad, because it makes the next part that much more terrifying.

II.

There is a sixth major Internet database of drug ratings. It is called RateRx, and it differs from the other five in an important way: it solicits ratings from doctors, not patients. It's a great idea – if you trust your doctor to tell you which drug is best, why not take advantage of wisdom-of-crowds and trust *all* the doctors?

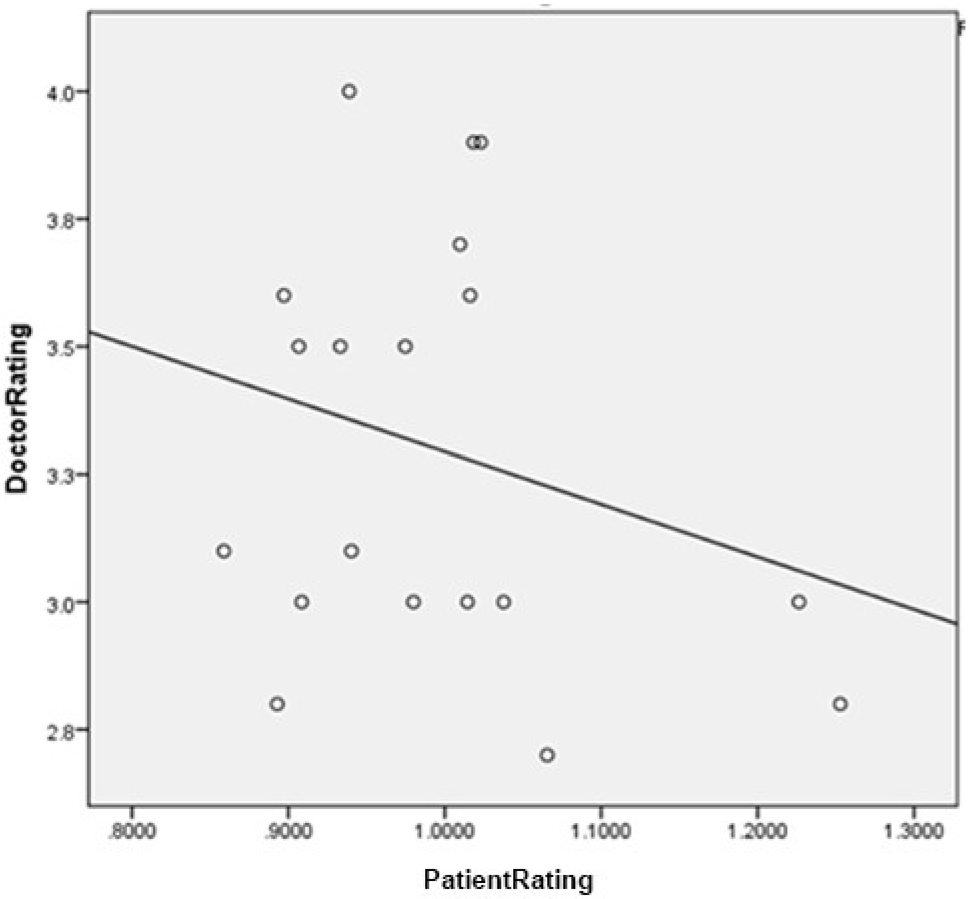
Doctors know best

The RateRX logo. Spoiler: this is going to seem really ironic in about thirty seconds.

RateRx has a modest but respectable sample size – the drugs on my list got between 32 and 70 doctor reviews. There's only one problem.

You remember patient reviews on the big three sites correlated about +0.7 with each other, right? So patients pretty much agree on which drugs are good and which are bad?

Doctor reviews on RateRx correlated at -0.21 with patient reviews. The negative relationship is nonsignificant, but that just means that at best, doctor reviews are totally uncorrelated with patient consensus.



This has an obvious but very disturbing corollary. I couldn’t get good numbers on how times each of the antidepressants on my list were prescribed, because the information I’ve seen only gives prescription numbers for a few top-selling drugs, plus we’ve got the same problem of not being able to distinguish depression prescriptions from anxiety prescriptions from psychosis prescriptions. But total number of online reviews makes a pretty good proxy. After all, the more patients are using a drug, the more are likely to review it.

Quick sanity check: the most reviewed drug on my list was Cymbalta. Cymbalta was also the best selling antidepressant of 2014. Although my list doesn't exactly track the best-sellers, that seems to be a function of how long a drug has been out – a best-seller that came out last year might have only 1/10th the number of reviews as a best-seller that came out ten years ago. So number of reviews seems to be a decent correlate for amount a drug is used.

In that case, amount a drug is used correlates highly (+0.67, p = 0.005) with doctors' opinion of the drug, which makes perfect sense since doctors are the ones prescribing it. But amount the drug gets used correlates negatively with patient rating of the drug (-0.34, p = ns), which of course is to be expected given the negative correlation between doctor opinion and patient opinion.

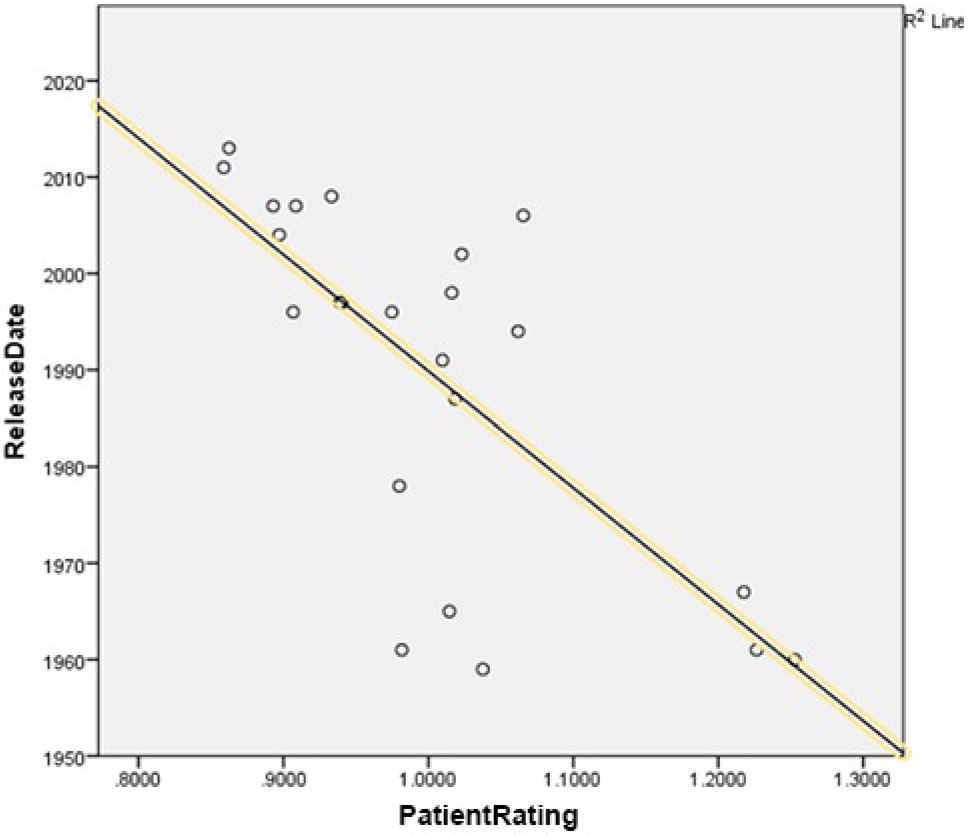
So the more patients like a drug, the less likely it is to be prescribed[2].

III.

There's one more act in this horror show.

Anyone familiar with these medications reading the table above has probably already noticed this one, but I figured I might as well make it official.

I correlated the average rating of each drug with the year it came on the market. The correlation was -0.71 (p < .001). That is, the newer a drug was, the less patients liked it[3].



This pattern absolutely *jumps* out of the data. First- and second- place winners Nardil and Parnate came out in 1960 and 1961, respectively; I can’t find the exact year third-place winner Anafranil came out, but the first reference to its trade name I can find in the literature is from 1967, so I used that. In contrast, last-place winner Viibryd came out in 2011, second-to-last place winner Abilify got its depression indication in 2007, and third-to-last place winner Brintellix is as recent as 2013.

This result is robust to various different methods of analysis, including declaring MAOIs to be an unfair advantage for Team Old and removing all of them, changing which minor tricylics I do and don’t include in the data, and altering whether Deprenyl, a drug that technically came out in 1970 but received a gritty reboot under the name Emsam in 2006, is counted as older or newer.

So if you want to know what medication will make you happiest, at least according to this analysis your best bet isn’t to ask your doctor, check what’s most popular, or even check any individual online rating database. It’s to look at the approval date on the label and choose the one that came out first.

IV.

What the *hell* is going on with these data?

I would like to dismiss this as confounded, but I have to admit that any reasonable person would expect the confounders to go the opposite way.

That is: older, less popular drugs are usually brought out only when newer, more popular drugs have failed. MAOIs, the clear winner of this analysis, are very clearly reserved in the guidelines for “treatment-resistant depression”, ie depression you’ve already thrown everything you’ve got at. But these are precisely the depressions that are hardest to treat.

Imagine you are testing the fighting ability of three people via ten boxing matches. You ask Alice to fight a Chihuahua, Bob to fight a Doberman, and Carol to fight Cthulhu. You would expect this test to be biased in favor of Alice and against Carol. But MAOIs and all these other older rarer drugs are practically never brought out except against Cthulhu. Yet they *still* have the best win-loss record.

Here are the only things I can think of that might be confounding these results.

Perhaps because these drugs are so rare and unpopular, psychiatrists only use them when they have really really good reason. That is, the most popular drug of the year they pretty much cluster-bomb everybody with. But every so often, they see some patient who seems absolutely 100% perfect for clomipramine, a patient who practically *screams* “clomipramine!” at them, and then they give this patient clomipramine, and she does really well on it.

(but psychiatrists aren’t actually that good at personalizing antidepressant treatments. The only thing even *sort of* like that is that MAOIs are extra-good for a subtype called atypical depression. But that’s like a third of the depressed population, which doesn’t leave much room for this super-precise-targeting hypothesis.)

Or perhaps once drugs have been on the market longer, patients figure out what they like. Brintellix is so new that the Brintellix patients are the ones whose doctors said “Hey, let’s try you on Brintellix” and they said “Whatever”. MAOIs have been on the market so long that presumably MAOI patients are ones who tried a dozen antidepressants before and stayed on MAOIs because they were the only ones that worked.

(but Prozac has been on the market 25 years now. This should only apply to a couple of very new drugs, not the whole list.)

Or perhaps the older drugs have so many side effects that no one would stay on them unless they’re absolutely perfect, whereas people are happy to stay on the newer drugs even if they’re not doing much because whatever, it’s not like they’re causing any trouble.

(but Seroquel and Abilify, two very new drugs, have awful side effects, yet are down at the bottom along with all the other new drugs)

Or perhaps patients on very rare weird drugs get a special placebo effect, because they feel that their psychiatrist cares enough about them to personalize treatment. Perhaps they identify with the drug – "I am special, I'm one of the only people in the world who's on nefazodone!" and they become attached to it and want to preach its greatness to the world.

(but drugs that are rare because they are especially new don't get that benefit. I would expect people to also get excited about being given the latest, flashiest thing. But only drugs that are rare because they are old get the benefit, not drugs that are rare because they are new.)

Or perhaps psychiatrists tend to prescribe the drugs they "imprinted on" in medical school and residency, so older psychiatrists prescribe older drugs and the newest psychiatrists prescribe the newest drugs. But older psychiatrists are probably much more experienced and better at what they do, which could affect patients in other ways – the placebo effect of being with a doctor who radiates competence, or maybe the more experienced psychiatrists are really good at psychotherapy, and that makes the patient better, and they attribute it to the drug.

(but read on…)

V.

Or perhaps we should take this data at face value and assume our antidepressants have been getting worse and worse over the past fifty years.

This is not entirely as outlandish as it sounds. The history of the past fifty years has been a history of moving from drugs with more side effects to drugs with fewer side effects, with what I consider somewhat less than due diligence in making sure the drugs were quite as effective in the applicable population. This is a *very* complicated and controversial statement which I will be happy to defend in the comments if someone asks.

The big problem is: drugs go off-patent after twenty years. Drug companies want to push new, on-patent medications, and most research is funded by drug companies. So lots and lots of research is aimed at proving that newer medications invented in the past twenty years (which make drug companies money) are better than older medications (which don't).

I'll give one example. There is only a single study in the entire literature directly comparing the MAOIs – the very old antidepressants that did best on the patient ratings – to SSRIs, the antidepressants of the modern day[4]. This study found that phenelzine, a typical MAOI, was no better than Prozac, a typical SSRI. Since Prozac had fewer side effects, that made the choice in favor of Prozac easy.

Did you know you can look up the authors of scientific studies on LinkedIn and sometimes get very relevant information? For example, the lead author of this study has a resume that clearly lists him as working for Eli Lilly at the time the study was conducted (spoiler: Eli Lilly is the company that makes Prozac). The second author's LinkedIn profile shows he is *also* an operations manager for Eli Lilly. Googling the fifth author's name links to a news article about Eli Lilly making a $750,000 donation to his clinic. Also there's a little blurb at the bottom of the paper saying "Supported by a research grant by Eli Lilly and company", then thanking several Eli Lilly executives by name for their assistance.

This is the sort of study which I kind of wish had gotten replicated *before* we decided to throw away an entire generation of antidepressants based on the result.

But who will come to phenelzine’s defense? Not Parke-Davis , the company that made it: their patent expired sometime in the seventies, and then they were bought out by Pfizer[5]. And not Pfizer – without a patent they can’t make any money off Nardil, and besides, Nardil is competing with their own on-patent SSRI drug Zoloft, so Pfizer has as much incentive as everyone else to push the “SSRIs are best, better than all the rest” line.

Every twenty years, pharmaceutical companies have an incentive to suddenly declare that all their old antidepressants were awful and you should never use them, but whatever new antidepressant they managed to dredge up is super awesome and you should use it all the time. This sort of *does* seem like the sort of situation that might lead to older medications being better than newer ones. A couple of people have been pushing this line for years – I was introduced to it by Dr. Ken Gillman from Psychotropical Research, whose recommendation of MAOIs and Anafranil as most effective match the patient data very well, and whose essay Why Most New Antidepressants Are Ineffective is worth a read.

I’m not sure I go as far as he does – even if new antidepressants aren’t worse outright, they might still trade less efficacy for better safety. Even if they handled the tradeoff well, it would look like a net loss on patient rating data. After all, assume Drug A is 10% more effective than Drug B, but also kills 1% of its users per year, while Drug B kills nobody. Here there’s a good case that Drug B is much better and a true advance. But Drug A’s ratings would look better, since dead men tell no tales and don’t get to put their objections into online drug rating sites. Even if victims' families did give the drug the lowest possible rating, 1% of people giving a very low rating might still not counteract 99% of people giving it a higher rating.

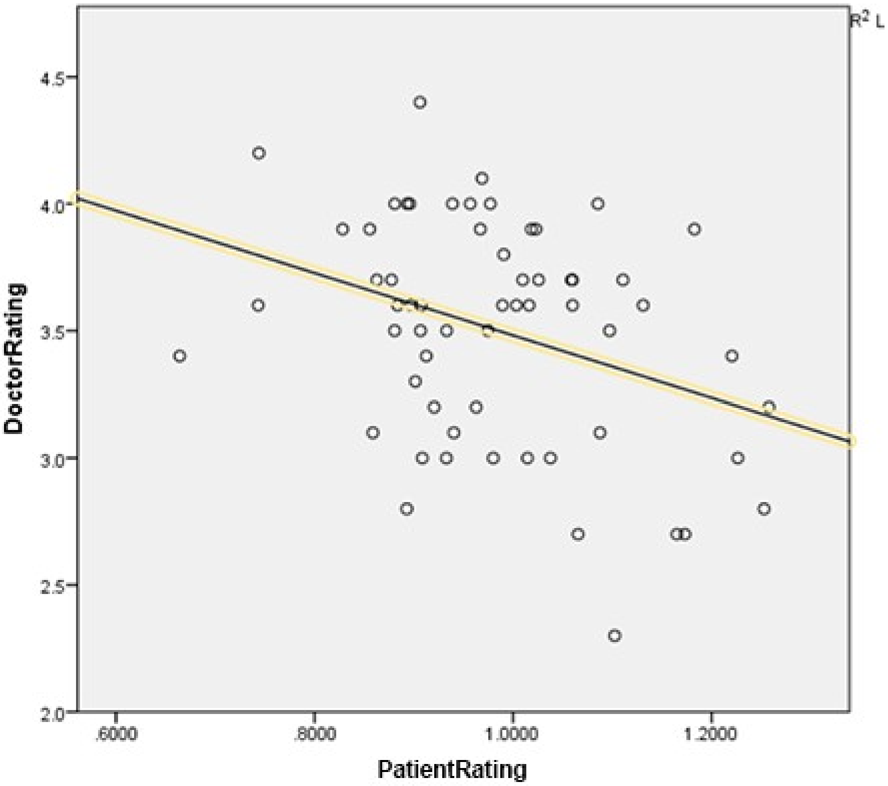
And once again, I'm not sure the tradeoff is handled very well at all.[6].

VI.

In order to distinguish between all these hypotheses, I decided to get a lot more data.

I grabbed all the popular antipsychotics, antihypertensives, antidiabetics, and anticonvulsants from the three databases, for a total of 55,498 ratings of 74 different drugs. I ran the same analysis on the whole set.

The three databases still correlate with each other at respectable levels of +0.46, +0.54, and +0.53. All of these correlations are highly significant, $p < 0.01$. The negative correlation between patient rating and doctor rating remains and is now a highly significant -0.344, $p < 0.01$. This is robust even if antidepressants are removed from the analysis, and is notable in both psychiatric and nonpsychiatric drugs.



The correlation between patient rating and year of release is a no-longer-significant -0.191. This is heterogenous; antidepressants and antipsychotics show a strong bias in favor of older medications, and antidiabetics, antihypertensives, and anticonvulsants show a slight nonsignificant bias in favor of newer medications. So it would seem like the older-is-better effect is purely psychiatric.

I conclude that for some reason, there really is a highly significant effect across all classes of drugs that makes doctors love the drugs patients hate, and vice versa.

I also conclude that older psychiatric drugs seem to be liked much better by patients, and that this is not some kind of simple artifact or bias, since if such an artifact or bias existed we would expect it to repeat in other kinds of drugs, which it doesn’t.

VII.

Please feel free to check my results. Here is a spreadsheet (.xls) containing all of the data I used for this analysis. Drugs are marked by class: 1 is antidepressants, 2 is antidiabetics, 3 is antipsychotics, 4 is antihypertensives, and 5 is anticonvulsants. You should be able to navigate the rest of it pretty easily.

One analysis that needs doing is to separate out drug effectiveness versus side effects. The numbers I used were combined satisfaction ratings, but a few databases – most notably WebMD – give you both separately. Looking more closely at those numbers might help confirm or disconfirm some of the theories above.

If anyone with the necessary credentials is interested in doing the hard work to publish this as a scientific paper, drop me an email and we can talk.

Footnotes

1. Technically, MAOI superiority has only been proven for atypical depression, the type of depression where you can still have changing moods but you are unhappy on net. But I'd speculate that right now most patients diagnosed with depression have atypical depression, far more than the studies would indicate, simply because we're diagnosing less and less severe cases these days, and less severe cases seem more atypical.

2. First-place winner Nardil has only 16% as many reviews as last-place winner Viibryd, even though Nardil has been on the market fifty years and Viibryd for four. Despite its observed superiority, Nardil may very possibly be prescribed less than 1% as often as Viibryd.

3. Pretty much the same thing is true if, instead of looking at the year they came out, you just rank them in order from earliest to latest.

4. On the other hand, what we do have is a lot of studies comparing MAOIs to imipramine, and a lot of other studies comparing modern antidepressants to imipramine. For atypical depression and dysthymia, MAOIs beat imipramine handily, but the modern antidepressants are about equal to imipramine. This strongly implies the MAOIs beat the modern antidepressants in these categories.

5. Interesting Parke-Davis facts: Parke-Davis got rich by being the people to market cocaine back in the old days when people treated it as a pharmaceutical, which must have been kind of like a license to print money. They also worked on hallucinogens with no less a figure than Aleister Crowley, who got a nice tour of their facilities in Detroit.

6. Consider: *Seminars In General Psychiatry* estimates that MAOIs kill one person per 100,000 patient years. A third of all depressions are atypical. MAOIs are 25 percentage points more likely to treat atypical depression than other antidepressants. So for every 100,000 patients you give a MAOI instead of a normal antidepressant, you kill one and cure 8,250 who wouldn't otherwise be cured. The QALY database says that a year of moderate depression is worth about 0.6 QALYs. So for every 100,000 patients you give MAOIs, you're losing about 30 QALYs and gaining about 3,300.

The Future Is Filters

Related to: The Toxoplasma of Rage

I.

Tumblr Savior is a neat program that blocks Tumblr posts containing specific words or phrases. For example, if you don't want to hear all of the excellent reasons going around Tumblr why you should kill all men, you just block "kill all men" and they never show up. Add a few extra terms like "white dudes" (nothing good ever came of an article including the phrase "white dudes"), "trans", "cis", and "pictures of my vagina", and you can make Tumblr almost usable.

(My own Tumblr Savior list is an interesting record both of my psyche and of mid-2010s current events. Sometimes I imagine a future cyber-archaeologist stumbling across it and asking "But, but…why would he ban the word 'puppies'?" Poor, poor innocent future archaeologist.)

I recently learned about Twitter blockbots. These are lists maintained by some trustworthy people, such that subscribing to the blockbot automatically blocks everyone on the list. The original was made by some people in the social justice community to help block people they figured other members of the social justice community wouldn't want to have to deal with. Although some people seem to be added on by hand, the bot also makes educated guesses about who to block by blacklisting accounts that follow the feeds of too many anti-social-justice leaders.

There are rumors of a similar anti-SJ block list of people who engage on online mobbing and harassment in the name of social justice, but I can’t find it online right now and I think it might have been taken down.

An article I read recently (but which I can’t find right now to link to) proposes a higher-tech solution for Facebook’s harassment problems. They want Facebook to train machine-learning programs to detect posts that most people would consider trollish. So far, so boring. The interesting part comes afterwards – instead of auto-blocking those posts, Facebook would assign them a certain number of Troll Points. Users could then set an option for how their Facebook feed should react to Troll Points – for example, by blocking every post with more than a certain amount. That way, people who were concerned about free speech and who enjoy participating in “heated discussion” would be able to do so, while people who wanted a safer and more pleasant browsing experience could have a very low cutoff for taking action.

But the *really* interesting part got dismissed after a sentence. What if instead of combining everything into Troll Points, Facebook assigned the points in different domains? Foul Language, Blasphemy, Racial Slurs, Threats, Harassment, Dirty Argument Tactics, et cetera. And then I could set that I don’t care about Foul Language or Blasphemy, but I really don’t want to see any Threats or Racial Slurs.

(obviously the correct anarcho-capitalist solution is to have third-party companies making these algorithms and selling them to individual Facebook users, but in a world where Facebook is trying to become more and more closed to third-party apps, that’s probably not going to happen)

So, take all this filtering technology – Tumblr Savior, Twitter blockbots, and hypothetical Facebook Troll Points, combine them together, project them about ten years into the future with slightly better machine learning, and you have an Internet where nobody has to see, even for an instant, anything they don't want to. What are the implications?

II.

The most obvious possibility is that everyone will be better off because we can avoid trolls. In this nice black-and-white worldview, there are good people, and there are trolls, and eliminating the trolls is a simple straightforward decision that makes the good people better off. This is how The Daily Beast thinks of it (How Block Bot Could Save The Internet), and as anyone who's been trolled or harassed online knows, there's a lot of truth to this view.

The second most obvious possibility is that we will become a civilization of wusses safely protected from ever having to hear an opinion we disagree with, or ever having our prejudices challenged. This is how Reason thinks of it (Block Bots Automate Epistemic Closure On Twitter). Surely there's some truth here too. How hard would it be to create a filter that blocks all conservative/liberal opinions? Just guess based on whether a text links to foxnews.com or dailykos.com, or add in linguistic cues ("death tax", "job creators", etc). Once such a filter existed, how many people do you think would use it proudly, bragging about how they're no longer "wasting their time listening patiently to bigots" or whatever?

But I don't think the scenario is quite that apocalyptic. If you're getting all of your exposure to opinions you disagree with from them being shouted in your face by people you can't avoid, you probably are not going to lose much by not having that happen. The people who are actually interested in holding discussions can still do that. When I was young and therefore stupid I used to hang out at politics forums specifically for this purpose.

The third possibility is that there would be a remarkable shift of discourse in favor of the powerful and against the powerless.

Terrorism has always been a useful weapon of the powerless. The powerful get laws passed through Congress or whatever, but the powerless don't have that opportunity. They need to get people to pay attention, and blowing those people up has always been an effective tool in that repertoire. We see this most obviously in places like Palestine and the Basque Country. Likewise, as many people have pointed out, the recent riots in Baltimore can be thought of as a group of powerless people trying to make their anger heard in one of the only ways available to them. It would be politically un-savvy to call this "terrorism", but as acts of destruction intended to promote a political struggle, they probably fit into the same cluster.

But the next step down from terrorism is annoyism. Terrorism is meant to convince by terrorizing those who ignore your cause; annoyism is meant to convince by annoying people who ignore your cause. Think of a bunch of protesters shouting on a major road, or throwing red paint over people wearing fur, or passive-aggressive Tumblr posts starting "dear white dudes", or, in probably the purest example of the idea, the Black Brunch protests, where a bunch of black people burst into predominantly white restaurants and shout at patrons about how they're probably complicit with racism. Even if there's no implicit threat of force, the point is it's unpleasant and people can't ignore it even if they want to.

And so the traditional revolutionary chant goes: “No justice, no peace.” But the thing about filters is that they offer the opportunity for peace regardless of whether or not there is justice. At least they do online, which is where people in the future are going to be spending a lot more of their time.

Imagine you are a rich person who doesn’t want to have to listen to people talking about how rich people need to be socially responsible all the time. It makes you feel guilty, and they are saying mean things like that you don’t deserve all of the money you have, and shouting about social parasites and so on.

So you tell your automated filter to just never let you see any message like that again.

There is an oft-discussed division between politically right or neutral loud angry people (“trolls”) and loud angry people on the political left, (“you are not allowed to dictate the terms on which victims of oppression express their righteous anger”). Machine learning programs will not accept that division, and the latter can be magicked out of visibility just as easily as the former.

Imagine being able to put an entire movement on mute. While I can’t deny the appeal, I’m not sure we – and especially not the social justice community, which is currently laughing at the complaints of people who object to their blockbot – have entirely thought this one through.

III.

The part I find most interesting about all of these possibilities is that they force us to bring previously unconscious social decisions into consciousness.

I think most people, if asked “Is it important to listen to arguments by people who disagree with you?” would answer in the affirmative. I also think most people don’t really do this. Maybe having to set a filter would make people explicitly choose to allow some contrary arguments in. Having done that, people could no longer complain about seeing them – they would feel more of an obligation to read and think about them. And of course, anyone looking for anything more than outrage-bait would choose to preferentially let in high-quality, non-insulting examples of disagreeing views, and so get inspired to think clearly instead of just starting one more rage spiral.

And I think most people, if asked “Is it important to listen to the concerns of the less powerful?” would also be pretty strongly in favor – with the caveat that people can recognize annoyism when it’s being used against them and aren’t especially tolerant of it. The ability to completely block out annoyism, combined with people being forced to explicitly choose to listen to alternative opinions, might make groups that currently favor annoyism change tactics to something more pleasant – though possibly less effective.

I think the result would be several carefully separated groups with their own social and epistemic norms, all of which coexist peacefully and in relative isolation from one another – groups which I would hope then develop their own norms about helping powerless members. This would be an interesting step towards what I describe in my Archipelago article as “a world where everyone is a member of more or less the community they deserve.”

Growth Mindset 4: Growth Of Office

Previously In Series: No Clarity Around Growth Mindset…Yet // I Will Never Have The Ability To Clearly Explain My Beliefs About Growth Mindset // Growth Mindset 3: A Pox On Growth Your Houses

Last month I criticized a recent paper, Paunesku et al's Mindset Interventions Are A Scalable Treatment For Academic Underachievement, saying that it spun a generally pessimistic set of findings about growth mindset into a generally optimistic headline.

Earlier today, lead author Dr. Paunesku was kind enough to write a very thorough reply, which I reproduce below:

I.

Hi Scott,

Thanks for your provocative blog post about my work (I'm the first author of the paper you wrote about). I'd like to take a few moments to respond to your critiques, but first I'd like to frame my response and tell you a little bit about my own motivation and that of the team I am a member of (PERTS).

Good criticism is what makes science work. We are critical of our own work, but we are happy to have help. Often critics are not thoughtful or specific. So I very much appreciate the intent of your blog (to be thoughtful and specific).

What is our motivation? We are trying to improve our education system so that all students can thrive. If growth mindset is effective, we want it in every classroom possible.

If it is ineffective, we want to know about it so we don't waste people's time. If it is effective for some students in some classrooms, we want to know where and for whom so that we can help those students.

What is our history and where are we now? PERTS approached social psychological interventions with a fair amount of skepticism at first. In many ways, they seemed too good to be true. But, we thought, "if this is true, we should do everything we can to spread it". Our work over the last 5 years has been devoted to trying to see if the results that emerged from initial, small experiments (like Aronson et al., 2002 and Blackwell et al., 2007) would continue to be effective when scaled. The paper you are critiquing is a step in that process — not the end of the process. We are continuing research to see where, for whom, and at what scale social psychological approaches to improving education outcomes can be effective.

How do I intend to respond to your criticisms? In some cases, your facts or interpretations are simply incorrect, and I will try to explain why. I also invite you to contact me for follow up. In others cases, we simply have different opinions about what's important, and we'll have to agree to disagree. Regardless, I appreciate your willingness to be bold and specific in your criticism. I think that's brave, and I think such bravery makes science stronger.

First, what is growth mindset?

This quote is from one of your other blog posts (not your critique of my paper), from your post:

> If you're not familiar with it, growth mindset is the belief that people who believe ability doesn't matter and only effort determines success are more resilient, skillful,

> hard-working, perseverant in the face of failure, and better-in-a-bunch-of-other-ways than people who emphasize the importance of ability. Therefore, we can make everyone better off by telling them ability doesn't matter and only hard work does.

If you think that's what growth mindset is, I can certainly see why you'd find it irritating — and even destructive. I'd like to assure you that the people doing growth mindset research do not ascribe to the interpretation of growth mindset you described. Nor is that interpretation of growth mindset something we aim to communicate through our interventions. So what is growth mindset?

Growth mindset is not the belief that "ability doesn't matter and only effort determines success." Growth mindset is the belief that individuals can improve their abilities — usually through effort and by learning more effective strategies. For example, imagine a third grader struggling to learn long division for the first time. Should he interpret his struggle as a sign that he's bad at math — as a sign that he should give up on math for good? Or would it be more adaptive if he realized that he could probably get a lot better at math if he sought out help from his peers or teachers? The student who thinks he should give up would probably do pretty badly while the student who thinks that he can improve his abilities — and tries to do so by learning new study strategies and practicing them — would do comparatively better.

That's the core of growth mindset. It's nothing crazy like thinking ability doesn't matter. It's keeping in mind that you can improve and that — to do so — you need to work hard and seek out and practice new, effective strategies.

As someone who has worked closely with Carol Dweck and with her students and colleagues for seven years now, I can personally attest that I have never heard anyone in that extended group of people express the belief that ability does not matter or that only hard work matters. In fact, a growth mindset wouldn’t make any sense if ability didn’t matter because a growth mindset is all about improving ability.

One of the active goals of the group I co-founded (PERTS) is to try to dispel misinterpretations of growth mindset because they can be harmful. I take it as a failure of our group that someone like you — someone who clearly cares about research and about scientific integrity — could walk away from our work with that interpretation of growth mindset. I hope that PERTS, and other groups promoting growth mindset, can get better and better at refining the way we talk about growth mindset so that people can walk away from our work understanding it more clearly. For that perspective, I hope you can continue to engage with us to improve that message so that people don’t continue to misinterpret it.

Anyway, here are my responses to specific points you made in your blog about my paper:

Was the control group a mindset intervention?

You wrote:

> “A quarter of the students took a placebo course that just presented some science about how different parts of the brain do different stuff. This was also classified as a “mindset intervention”, though it seems pretty different.”

What makes you think it was classified as a mindset intervention? We called that the control group, and no one on our team ever thought of that as a mindset intervention.

The Elderly Hispanic Woman Effect

You wrote:

> Subgroup analysis can be useful to find more specific patterns in the data, but if it's done post hoc it can lead to what I previously called the Elderly Hispanic Woman Effect...

First, I just want to note that I love calling this the "elderly Hispanic woman effect." It really brings out the intrinsic ridiculousness of the subgroup analyses researchers sometimes go through in search of an effect with a $p<.05$. It is indeed unlikely that "elderly Hispanic women" would be a meaningful subgroup for analyzing the effects of a medicine (although it might be a fun thought exercise to try to think of examples of a medicine whose effects would be likely to be moderated by being an elderly Hispanic woman).

In bringing up the elderly Hispanic woman effect, you're suggesting that we didn't have an a priori reason to think that underperforming students would benefit from these mindset interventions and that we just looked through a bunch of moderators until we found one with $p<.05$. Well that's not what we did, and I hope I can convince you that our choice of moderator was perfectly reasonable given prior research and theory.

There's a lot of research (and common sense too) to suggest that mindset — and motivation in general — matters much more when something is hard than when it is easy. Underachieving students presumably find school more difficult, so it makes sense that we'd want to focus on them.

I don’t think our choice of subgroup is a controversial or surprising prediction. I think anyone who knows mindset research well would predict stronger effects for students who are struggling. In other words, this is obviously not a case of the elderly Hispanic woman effect because it is totally consistent with prior theory and predictions. What ultimately matters more than any rhetorical argument, however, is whether the effect is robust — whether it replicates.

On that front, I hope you’ll be pleased to learn that we just ran a successful replication of this study (in fall 2014) in which we again found that growth mindset improves achievement specifically among at-risk high school students (currently under review). We’re also planning yet another large scale replication study this fall with a nationally representative sample of schools so that we can be more confident that the interventions are effective in various types of contexts before giving them away for free to any school that wants them.

Is the sense of purpose intervention just a bunch of platitudes?

You wrote:

> Still another quarter took a course about “sense of purpose” which talked about how schoolwork was meaningful and would help them accomplish lots of goals and they should be happy to do it.

[Later you say that those “children were told platitudes about how doing well in school will “make their families proud” and “make a positive impact”.]

I wouldn’t say those are platitudes. I think you’re under-appreciating the importance of finding meaning in one’s work. It’s a pretty basic observation about human nature that people are more likely to try hard when it seems like there’s a good reason to try hard. I also think it’s a pretty basic observation about our education system that many students don’t have good reasons for trying hard in school — reasons that resonate with them emotionally and help them find the motivation to do their best in the classroom. In our purpose intervention, we don’t just tell students what to think. We try to scaffold them to think of their own reasons for working hard in school, with a focus on reasons that are more likely to have emotional resonance for students. This type of self-persuasion technique has been used for decades in attitudes research.

We’ve written in more depth about these ideas and explored them through a series of studies. I’d encourage you to read this article if you’re interested.

Our paper title and abstract are misleading

You wrote:

> Among ordinary students, the effect on the growth mindset group was completely indistinguishable from zero, and in fact they did nonsignificantly worse than the control group. This was the most basic test they performed, and it should have been the headline of the study. The study should have been titled “Growth Mindset Intervention Totally Fails To Affect GPA In Any Way”.

I think the title you suggest would have been misleading. How?

First, we did find evidence that mindset interventions help underachieving students — and those students are very important from a policy standpoint. As we describe in the paper, those students are more likely to drop out, to end up underemployed, or to end up in prison. So if something can help those students at scale and at a low cost, it's important for people to know that. That's why the word "underachievement" is in the title of the paper — because we're accurately claiming that these interventions can help the important (and large) group of students who are underachieving.

Second, the interventions influenced the way all students think about school in ways that are associated with achievement. Although the higher performing students didn't show any effects on grades in the semester following the study, their mindsets did change. And, as per the arguments I presented above about the link between mindset and difficulty, it's quite feasible that those higher-performing students will benefit from this change in mindset down the line. For example, they may choose to take harder classes (e.g., Romero et al., 2014) or they may be more persistent and successful in future classes that are very challenging for them.

A misinterpretation of the y-axis in this graph.

You wrote:

> Growth mindset still doesn't differ from zero [among at-risk students].

This just seems to be a simple misreading of the graph. Either you missed the y-axis of the graph that you reproduced on your blog or you don't know what a residual standardized score is. Either way, I'll explain because this is pretty esoteric stuff.

The zero point of the y-axis on that graph is, by definition, the grand mean of the 4 conditions. In other words, the treatment conditions are all hovering around zero because zero is the average, and the average is made up mostly of treatment group students. If we had only had 2 conditions (each with 50% of the students), the y-axis "zero" would have been exactly halfway in between them. So the lack of difference from zero does not mean that the treatment was not different from control. The relevant comparison is between the error bars in the control condition and in the treatment conditions.

You might ask, "why are you showing such a graph?" We're doing so to focus on the treatment contrast at the heart of our paper — the contrast between the control and treatment groups. The residual standardized graph makes it easy to see the size of that treatment contrast.

We're combining intervention conditions

You wrote:

> Did you catch that phrase "intervention conditions"? The authors of the study write: "Because our primary research question concerned the efficacy of academic mindset interventions in general when delivered via online modules, we then collapsed the intervention conditions into a single intervention dummy code (0 = control, 1 = intervention).

[This line of argument goes on for a long time to suggest that we're unethical and that there's actually no evidence for the effects of growth mindset on achievement.]

We collapsed the intervention conditions together for this analysis because we were interested in the overall effect of these interventions on achievement. We wanted to see if it is possible to use scalable, social-psychological approaches to improve the achievement of underperforming students.

I’m not sure why you think that’s not a valid hypothesis to test, but we certainly think it is. Maybe this is just a matter of opinion about what’s a meaningful hypothesis to test, but I assure you that this hypothesis (contrast all treatments to control) is consistent with the goal of our group to develop treatments that make an impact on student achievement. As I described before, we have a whole center devoted to trying to improve academic achievement with these types of techniques (see perts.net); so it’s pretty natural that we’d want to see whether our social-psychological interventions improve outcomes for the students who need them most (at-risk students).

You’re correct that the growth mindset intervention did not have a statistically significant impact on course passing rates by itself (at a $p<.05$ level). However, the effect was in the expected direction with $p=0.13$ (or a 1-tailed $p=.07$ — I hope you’ll grant that a 1-tailed test is appropriate here given that we obviously predicted the treatment would improve rather than reduce performance). So the lack of a $p<.05$ should not be interpreted — as you seem to interpret it — as some sort of positive evidence that growth mindset “actually didn’t work.” Anyway, I would say it warrants further research to replicate this effect (work we are currently engaging in).

To summarize, we did not find direct evidence that the growth mindset intervention increased course passing rates on its own at a $p<.05$ level. We did find that growth mindset increased course passing rates at a trend level — and found a significant effect on GPA. More importantly for me (though perhaps less relevant to your interest specifically in growth mindset), we did provide evidence that social-psychological interventions, like growth mindset and sense of purpose, can improve academic outcomes for at-risk students.

We’re excited to be replicating this work now and giving it away in the hopes of improving outcomes for students around the world.

Summary

I hope I addressed your concerns about this paper, and I welcome further discussion with you. I’d really appreciate it if you’d revise your blog post in whatever way you think is appropriate in light of my response. I’d hate for people to get the wrong impression of our work, and you don’t strike me as someone who would want to mislead people about scientific findings either.

Finally, you’re welcome to post my response. I may post it to my own web page because I’m sure many other people have similar questions about my work. Just let me know how you’d like to proceed with this dialog.

Thanks for reading,

Dave

II.

First of all, the obvious: this is extremely kind and extremely well-argued and a lot of it is correct and makes me feel awful for being so snarky on my last post.

Things in particular which I want to endorse as absolutely right about the critique:

I wrote “A quarter of the students took a placebo course that just presented some science about how different parts of the brain do different stuff. This was also classified as a “mindset intervention”, though it seems pretty different.” Dr. Paunesku says this is wrong. He’s right. It was an editing error on my part. I meant to add the last sentence to the part on the “sense of purpose” intervention, which was classified as a mindset intervention and which I do think seems pretty different. The placebo intervention was never classified as a mindset intervention and I completely screwed up by inserting that piece of text there rather than two sentences down where I meant it to be. It has since been corrected and I apologize for the error.

If another successful replication found that growth mindset continues to only help the lowest-performing students, I withdraw the complaint that this is sketchy subgroup mining, though I think that in general worrying about this is the correct thing to do.

I *did* misunderstand the residual standardized graph. I suggested that the control group must have severely declined, and got confused about why. In fact, the graph was not about difference between pre-study scores and post-study scores, but difference between group scores and the average score for all four groups. So when the control group is strongly negative, that means it was much worse than the average of all groups. When growth mindset is not-different-from-zero, it means growth mindset was not different from the average of all four groups, which consists of three treatment groups and one control group. So my interpretation – that growth mindset failed to change children’s grades – is not supported by the data.

(In my defense, I can only plead that in the two hundred fifty comments I received, many by professional psychologists and statisticians, only one person picked up on this point (admittedly, after being primed by my own misinterpretation). And the sort of data I expected to be seeing – difference between students’ pre-intervention and post-intervention scores – does not seem to be available.

Nevertheless, this was a huge and unforgiveable screw-up, and I apologize.)

III.

But there are also a few places where I will stick to my guns.

I don't think my interpretation of growth mindset was that far off the mark. I explain this a little further in this post on differing possible definitions of growth mindset, and I will continue to cite this strongly worded paper by Dweck as defense of my views. It's *not* just an obvious and innocuous belief about about always believing you should be able to improve, it's a belief about very counterintuitive effects of believing that success depends on ability versus effort. It is possible that all sophisticated researchers in the field have a very sophisticated and unobjectionable definition of growth mindset, but that's not the way it's presented to the public, even in articles by those same researchers.

Although I'm sure that to researchers in the field statements like "Doing well at school will help me achieve my goal" don't sound like platitudes, it seems important to me in the context of discussions about growth mindset. Some people have billed growth mindset as a very exciting window into what makes learning tick, and how we should divide everyone into groups based on their mindset, and how it's the Secret To Success, and so on. Learning that a drop-dead simple intervention – telling students to care about school more – actually does as well or better than growth mindset seems to me like a damning result. I realize it would be kind of insulting to call sense-of-purpose an "active placebo" in the medical sense, but that's kind of how I can't help thinking of it.

I’m certainly not suggesting the authors of the papers are *unethical* for combining growth mindset intervention with sense of purpose intervention. But I think the technique is dangerous, and this is an example. They got a result that was significant at p = 0.13. Dr. Paunesku suggests in his email to me that this should be one-tailed (which makes it p = 0.07) and that this obviously trends towards significance. This is a reasonable argument. But this wasn’t the reasonable argument made in the paper. Instead, they make it look like it achieved classical p < 0.05 significance, or at least make it very hard to notice that it didn’t.

Even if in this case it was – I can’t even say white lie, maybe a white spin – I find the technique very worrying. Suppose I want to prove homeopathy cures cancer. I make a trial with one placebo condition and two intervention conditions – chemotherapy and homeopathy. I find that the chemotherapy condition very significantly outperforms placebo, but the homeopathy condition doesn’t. So I combine the two interventions into a single bin and say “Therapeutic interventions such as chemotherapy or homeopathy significantly outperform placebo.” Then someone else cites it as “As per a study, homeopathy outperforms placebo.” This would obviously be bad.

I am just not convinced that growth mindset and sense of purpose are similar enough that you can group them together effectively. This is what I was trying to get at in my bungled sentence about how they’re both “mindset” interventions but seem pretty different. Yes, they’re both things you tell children in forty-five minute sessions that seem related to how they think about school achievement. But that’s a *really* broad category.

But doesn’t it mean something that growth-mindset was obviously trending toward significance?

First of all, I would have had no problem with saying “trending toward significance” and letting readers draw their own conclusions.

Second of all, I’m not totally sure I buy the justification for a one-tailed test here; after all, it seems like we should use a one-tailed test for homeopathy as well, since as astounding as it would be if homeopathy helped, it would be even more astounding if homeopathy somehow made cancer worse. Further, educational interventions *often* have the opposite of their desired effect – see eg this campaign to increase tolerance of the disabled which made students like disabled people *less* than a control intervention. In fact, there’s no need to look further than this very study, which found (counterintuitively) that among students already exposed to sense-of-purpose interventions, adding on an extra growth-mindset intervention seemed to make them do (nonsignificantly) worse. I am not a statistician, but my understanding is you ought to have a *super* good reason to use a one-tailed test, beyond just “Intuitively my hypothesis is way more likely than the exact opposite of my hypothesis”.

Third of all, if we accept $p < 0.13$ as “trending towards significance”, we have basically tripled the range of acceptable study results, even though everyone agrees our current range of acceptable study results is already way too big and some high percent of all medical studies are wrong and only 39% of psych studies replicate and so on.

(I agree that all of this could be solved by something better than p-values, but p-values are what we’ve got)

I realize I’m being a jerk by insisting on the arbitrary 0.05 criterion, but in my defense, the time when only 39% of studies using a criterion replicate is a bad time to loosen that criterion.

IV.

Here’s what I still believe and what I’ve changed my mind on based on Dr. Paunesku’s response.

1. I totally bungled my sentence on the placebo group being a mindset intervention by mistake. I ashamedly apologize, and have corrected the original post.

2. I totally bungled reading the residual standard score graph. I ashamedly apologize, and have corrected the original post, and put a link in bold text to this post on the top.

3. I don’t know whether the thing I thought the graph showed (no significant preintervention vs. postintervention GPA improvement for growth mindset, or no difference in change from controls) is true. It may be hidden in the supplement somewhere, which I will check later. Possible apology pending further investigation.

4. Growth mindset still had no effect (in fact nonsignificantly negative) for students at large (as opposed to underachievers). I regret nothing.

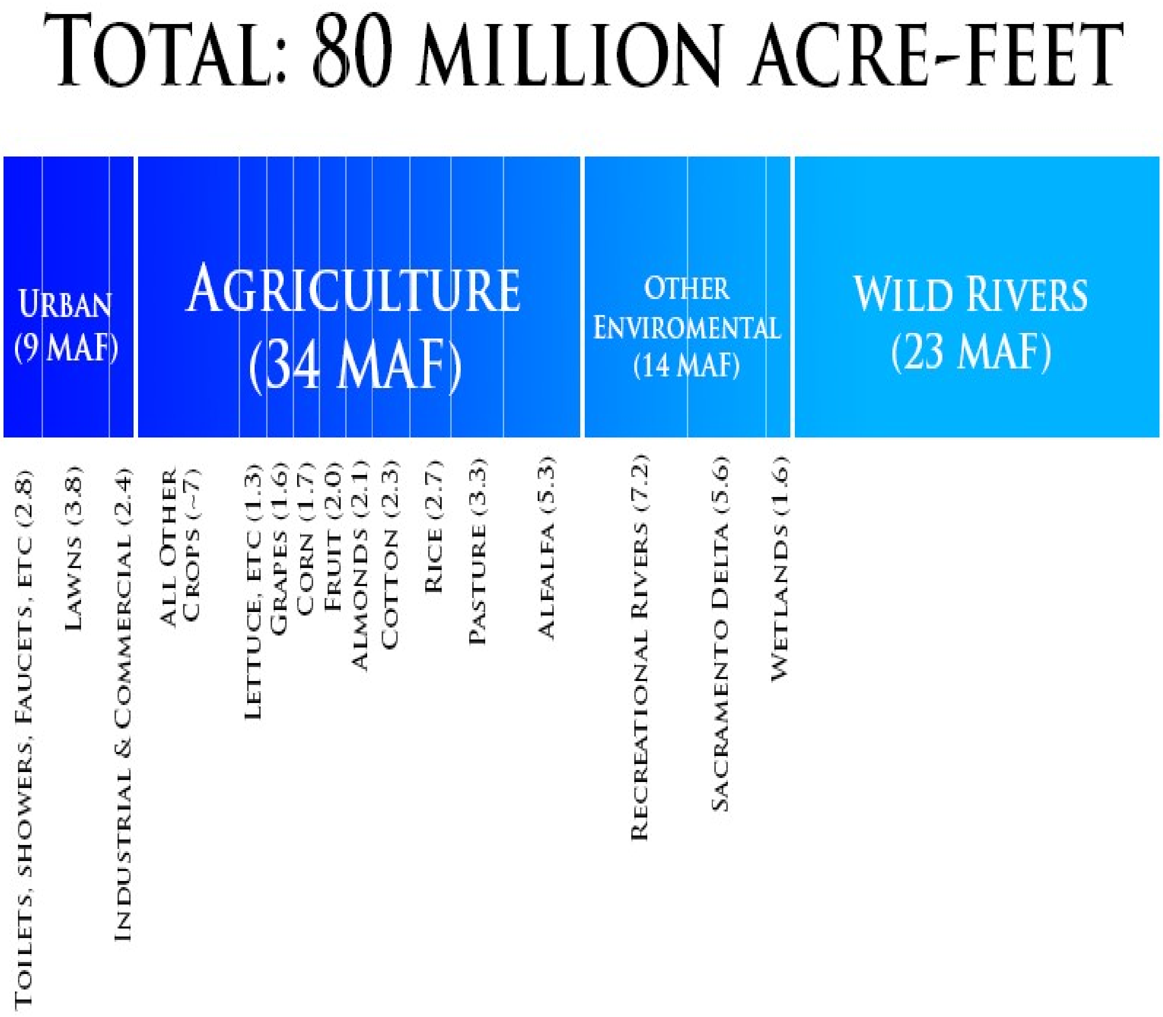
5. Growth mindset still failed to reach traditional significance criteria for changing pass rates. I regret nothing.

California, Water You Doing?

[Epistemic status: Low confidence. I have found numbers and stared at them until they made sense to me, but I have no education in this area. Tell me if I'm wrong.]

I.

There has recently been a lot of dumb fighting over who uses how much water in California, so I thought I would see if it made more sense as an infographic sort of thing:



Sources include Understanding Water Use In California, Inputs To Farm Production, California Water Usage In Crops, Urban Water Use Efficiency, Water Use In California, and Water: Who Uses How Much. There are some contradictions, probably caused by using sources from different years, and although I'm pretty confident this is right on an order of magnitude scale I'm not sure about a percentage point here or there. But that having been said:

On a state-sized level, people measure water in acre-feet, where an acre-foot is the amount of water needed to cover an area of one acre to a depth of one foot. California receives a total of 80 million acre-feet of water per year. Of those, 23 million are stuck in wild rivers (the hydrological phenomenon, not the theme park). These aren't dammed and don't have aqueducts to them so they can't be used for other things. There has been a lot of misdirection over this recently, since having pristine wild rivers that fish swim in seems like an environmental cause, and so you can say that "environmentalists have locked up 23 million acre-feet of California water". This is not a complete lie; if not for environmentalism, maybe some of these rivers would have been dammed up and added to the water system. But in practice you can't dam every single river and most of these are way off in the middle of nowhere far away from the water-needing population. People's ulterior motives shape whether or not they add these to the pot; I've put them in a different color blue to mark this.

Aside from that, another 14 million acre-feet are potentially usable, but deliberately diverted to environmental or recreational causes. These include 7.2 million for "recreational rivers", apparently ones that people like to boat down, 1.6 million to preserve wetlands, and 5.6 million to preserve the Sacramento River Delta. According to environmentalists, this Sacramento River Delta water is non-negotiable, because if we stopped sending fresh water there the entire Sacramento River delta would turn salty and it would lead to some kind of catastrophe that would threaten our ability to get fresh water into the system at all.

34 million acre-feet of water are diverted to agriculture. The most water-expensive crop is alfalfa, which requires 5.3 million acre-feet a year. If you’re asking “Who the heck eats 5.3 million acre-feet of alfalfa?” the answer is “cows”. A bunch of other crops use about 2 million acre-feet each.

All urban water consumption totals 9 million acre-feet. Of those, 2.4 million are for commercial and industrial institutions, 3.8 million are for lawns, and 2.8 million are personal water use by average citizens in their houses. In case you’re wondering about this latter group, by my calculations all water faucets use 0.5 million, all toilets use 0.9 million, all showers use 0.5 million, leaks lose 0.3 million, and the remaining 0.6 million covers everything else – washing machines, dishwashers, et cetera.

Since numbers like these are hard to think about, it might be interesting to put them in a more intuitive form. The median California family earns $70,000 a year – let’s take a family just a little better-off than that who are making $80,000 so we can map it on nicely to California’s yearly water income of 80 million acre-feet.

The unusable 23 million acre-feet which go into wild rivers and never make it into the pot correspond to the unusable taxes the California family will have to pay. So our family is left with $57,000 post-tax income.

In this analogy, California is spending $14,000 on environment and recreation, $34,000 on agriculture, and $9,000 on all urban areas. All household uses – toilets, showers, faucets, etc – only add up to about $2,800 of their budget.

There is currently a water shortfall of about 6 million acre-feet per year, which is being sustained by exploiting non-renewable groundwater and other sources. This is the equivalent of our slightly-richer-than-average family having to borrow $6,000 from the bank each year to get by.

II.

Armed with this information, let's see what we can make of some recent big news stories.

Apparently we are supposed to be worried about fracking depleting water in California. ThinkProgress reports that Despite Historic Drought, California Used 70 Million Gallons Of Water For Fracking Last Year. Similar concerns are raised by RT, Huffington Post, and even The New York Times. But 70 million gallons equals 214 acre-feet. Remember, alfalfa production uses 5.3 *million* acre feet. In our family-of-four analogy above, all the fracking in California costs them about a quarter. Worrying over fracking is like seeing an upper middle class family who are $6,000 in debt, and freaking out because one of their kids bought a gumball from a machine.

Apparently we are also supposed to be worried about Nestle bottling water in California. ABC News writes an article called Nestle Needs To Stop Bottling Water In Drought-Stricken California, Advocacy Group Says, about a group called the "Courage Campaign" who have gotten 135,000 signatures on a petition saying that Nestle needs to stop "bottling the scarce resource straight from the heart of California's drought and selling it for profit." Salon goes even further – their article is called Nestle's Despicable Water Crisis Profiteering: How It's Making A Killing While California Is Dying Of Thirst, and as always with this sort of thing Jezebel also has to get in on the action. But Nestle's plant uses only 150 acre-feet, about one forty-thousandth the amount used to grow alfalfa, and the equivalent of about a dime to our family of four.

The Wall Street Journal says that farms are a scapegoat for the water crisis, because in fact the real culprits are environmentalists. They say that "A common claim is that agriculture consumes about 80% of 'developed' water supply, yet this excludes the half swiped off the top for environmental purposes." But environmentalism only swipes half if you count among that half all of the wild rivers in the state – that is, every drop of water not collected, put in an aqueduct, and used to irrigate something is a "concession" to environmentalists. A more realistic figure for environmental causes is the 14 million acre-feet marked "Other Environmental" on the map above, and even that includes concessions to recreational boaters and to whatever catastrophe is supposed to happen if we can't keep the Sacramento Delta working properly. It's hard to calculate exactly how much of California's water goes to environmental causes, but half is definitely an exaggeration.

Wired is concerned that the federal government is ordering California to spend 12,000 acre-feet of water to save six fish (h/t Alyssa Vance). Apparently these are endangered fish in some river who need to get out to the Pacific to breed, and the best way to help them do that is to fill up the river with 12,000 acre feet of water. That's about $12 on our family's budget, which works out to $2 per fish. I was going to say that I could totally see a family spending $2 on a fish, especially if it was one of those cool glow-in-the-dark fish I used to have when I was a kid, but then I remembered this was a metaphor and the family is actually the entire state budget of California but the six fish are still literally just six fish. Okay, yes, that seems a little much.

III.

Finally, Marginal Revolution and even some among the mysterious and endangered population of non-blog-having economists are talking about how really the system of price controls and subsidies in the water market is ridiculous and if we had a free market on water all of our problems would be solved. It looks to me like that’s probably right.

Consider: When I used to live in California, even before this recent drought I was being told to take fewer showers, to install low-flush toilets that were inconvenient and didn’t really work all that well, to limit my use of the washing machine and dishwasher, et cetera. It was actually pretty inconvenient. I assume all forty million residents of California were getting the same message, and that a lot of them would have liked to be able to pay for the right to take nice long relaxing showers.

But if all the savings from water rationing amounted to 20% of our residential water use, then that equals about 0.5 MAF, which is about 10% of the water used to irrigate alfalfa. The California alfalfa industry makes a total of $860 million worth of alfalfa hay per year. So if you calculate it out, a California resident who wants to spend her fair share of money to solve the water crisis without worrying about cutting back could do it by paying the alfalfa industry $2 to not grow $2 worth of alfalfa, thus saving as much water as if she very carefully rationed her own use.

If you were to offer California residents the opportunity to *not* have to go through the whole gigantic water-rationing rigamarole for $2 a head, I think even the poorest people in the state would be pretty excited about that. My mother just bought and installed a new water-saving toilet – which took quite a bit of her time and money – and furthermore, the government is going to give her a $125 rebate for doing so.

Cutting water on the individual level is hard and expensive. But if instead of trying to save water ourselves, we just paid the alfalfa industry not to grow alfalfa, all the citizens of California could do their share for $2. If they also wanted to have a huge lush water-guzzling lawn, their payment to the alfalfa industry would skyrocket all the way to $5 per year.

In fact, though I am not at all sure here and I'll want a real economist to double-check this, it seems to me if we wanted to buy out all alfalfa growers by paying them their usual yearly income to just sit around and not grow any alfalfa, that would cost $860 million per year and free up 5.3 million acre-feet, ie pretty much our entire shortfall of 6 million acre-feet, thus solving the drought. Sure, 860 million dollars sounds like a lot of money, but note that right now California newspapers have headlines like Billions In Water Spending Not Enough, Officials Say. Well, maybe that's because you're spending it on giving people $125 rebates for water-saving toilets, instead of buying out the alfalfa industry. I realize that paying people subsidies to misuse water to grow unprofitable crops, and then offering them countersubsidies to not take your first set of subsidies, is to say the least a very creative way to spend government money – but the point is *it is better than what we're doing now*.

Bicameral Reasoning

[Epistemic status: Probably not the first person to think about this, possibly just reinventing scope insensitivity. Title with apologies to Julian Jaynes]

Non-American readers may not be familiar with the history of the US House and Senate.

During the Constitutional Convention, a fight broke out between the smaller states and the bigger states. The smaller states, like Delaware, wanted each state to elect a fixed number of representatives to the legislature, so that Delaware would have just as much of a say as, for example, New York. The bigger states wanted legislative representation to be proportional to population, so that if New York had ten times as many people as Delaware, they would get ten times as many representatives.

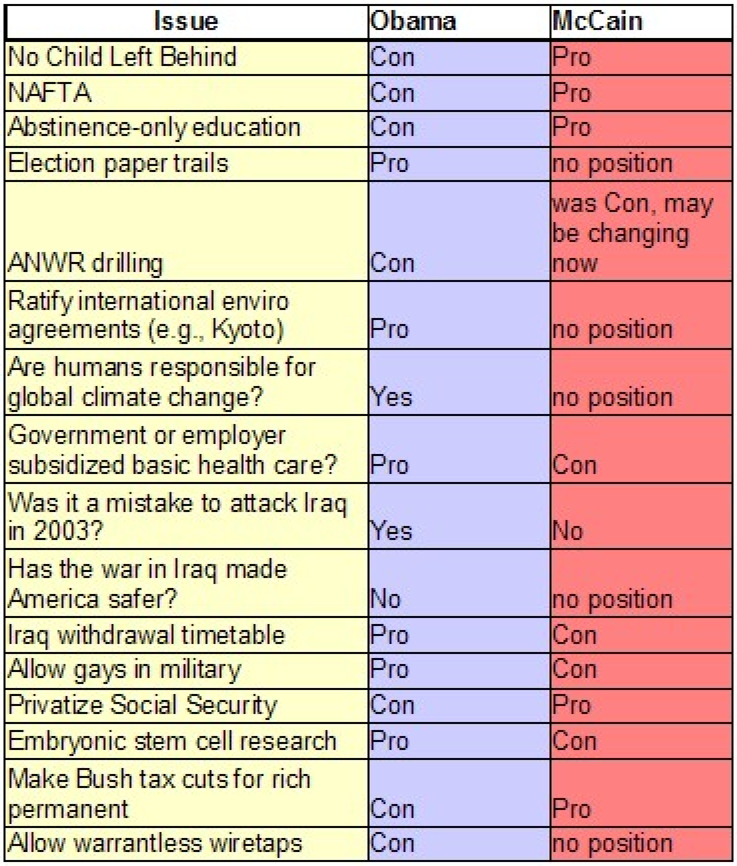
Eventually everyone just agreed to compromise by splitting the legislature into the House of Representatives and the Senate. The House worked the way New York wanted things, the Senate worked the way Delaware wanted things, and they would have to agree to get anything done.

This system has continued down to the present. Today, Delaware has only one Representative, far less than New York's twenty-seven. But both states have an equal number of Senators, even though New York has a population of twenty million and Delaware is uninhabited except by corporations looking for tax loopholes.

To me, the House system seems much fairer. If New York has ten times the population of Delaware, but both have the same number of representatives, then Delaware citizens have ten times as much political power just because they live on one side of an arbitrary line. And New York might be tempted to split up into ten smaller states, and thus increase its political power tenfold. Heck, why don't we just declare some random farm a state and give five people and a cow the same political power as all of California?

But despite my professed distaste for the Senate's representational system, I find myself using something similar in parts of my own thought processes where I least expect.

Every election, I see charts like this:



Issue	Obama	McCain
No Child Left Behind	Con	Pro
NAFTA	Con	Pro
Abstinence-only education	Con	Pro
Election paper trails	Pro	no position
ANWR drilling	Con	was Con, may be changing now
Ratify international enviro agreements (e.g., Kyoto)	Pro	no position
Are humans responsible for global climate change?	Yes	no position
Government or employer subsidized basic health care?	Pro	Con
Was it a mistake to attack Iraq in 2003?	Yes	No
Has the war in Iraq made America safer?	No	no position
Iraq withdrawal timetable	Pro	Con
Allow gays in military	Pro	Con
Privatize Social Security	Con	Pro
Embryonic stem cell research	Pro	Con
Make Bush tax cuts for rich permanent	Con	Pro
Allow warrantless wiretaps	Con	no position

And I tend to think something like "Well, I agree with this guy about the Iraq war and global warming, but I agree with that guy about election paper trails and gays in the military, so it's kind of a toss-up."

And this way of thinking is *awful*.

The Iraq War probably killed somewhere between 100,000 and 1,000,000 people. If you think that it was unnecessary, and that it was possible to know beforehand how poorly it would turn out, then killing a few hundred thousand people is a *really big deal*. I like having paper trails in elections as much as the next person, but if one guy isn't going to keep a very good record of election results, and the other guy is going to kill a million people, that's *not* a toss-up.

Likewise with global warming versus gays in the military. It would be nice if homosexual people have the same right to be killed by roadside explosive devices that the rest of us enjoy, but *not frying the planet* is pretty important too.

(if you don't believe in global warming, fine, having a government that agrees with you and doesn't waste 5% of the world GDP fighting it is *still* more important than anything else on this list)

Saying “some boxes are more important than others” doesn't really cut it; it sounds like they might be twice, maybe three times more important, whereas in fact they might literally be a million times more important. It doesn't convey the right sense of “Why are you even *looking* at that other box?”

I worry that, by portraying issues in this nice little set of boxes, this graphic is priming reasoning similar to the US Senate, where each box gets the same level of representation in my decision-making process, regardless of whether it's a Delaware-sized box that affects a handful of people, or a New York sized box with millions of lives hanging in the balance.

I was thinking about this again back in March when I had a brief crisis caused by worrying that the moral value of the world's chickens vastly exceeded the moral value of the world's humans. I ended up being trivially wrong – there are only about twenty billion chickens, as opposed to the hundreds of billions I originally thought. But I was contingently wrong – in other words, I got lucky. Honestly, I didn't know whether there were twenty billion chickens or twenty trillion.

And honestly, 99% of me doesn't care. I do want to improve chickens, and I do think that their suffering matters. But thanks to the miracle of scope insensitivity, I don't particularly care more about twenty trillion chickens than twenty billion chickens.

Once again, chickens seem to get two seats to my moral Senate, no matter how many of them there are. Other groups that get two seats include "starving African children", "homeless people", "my patients in hospital", "my immediate family", and "my close friends". Obviously some of these groups contain thousands of times more people than others. They still get two seats. And so I am neither willing to reduce chickens' values to zero value units per chicken, nor accept that if there are enough chickens they will end up able to outvote everyone else.

(I'm not sure whether "chickens" and "cows" are two separate states, or if there's just one state of "Animals". It probably depends on my mood. Which is worrying.)

And most recently I thought about this because of the post on California water I wrote last week. It seems very wise to say we all have to make sacrifices, and to concentrate about equally on natural categories of water use like showers, and toilets, and farms, and lawns – without noticing that one of those is ten times bigger than the other three combined. It seems like most people who think about the water crisis are using a Senate model, where each category is treated as an equally important area to optimize. In a House model, you wouldn’t be thinking about showers any more than a 2008 voter should be thinking of election paper trails.

I’m tempted to say “The House is just plain right and the Senate is just plain wrong”, but I’ve got to admit that would clash with my own very strong inclinations on things like the chicken problem. The Senate view seems to sort of fit with a class of solutions to the dust specks problem where after the somethingth dust speck or so you just stop caring about more of them, with the sort of environmentalist perspective where biodiversity itself is valuable, and with the Leibnizian answer to Job.

But I’m pretty sure those only kick in at the extremes. Take it too far, and you’re just saying the life of a Delawarean is worth twenty-something New Yorkers.

Beware Summary Statistics

Last night I asked Tumblr two questions that had been bothering me for a while and got some pretty good answers.

I.

First, consider the following paragraph from JRank:

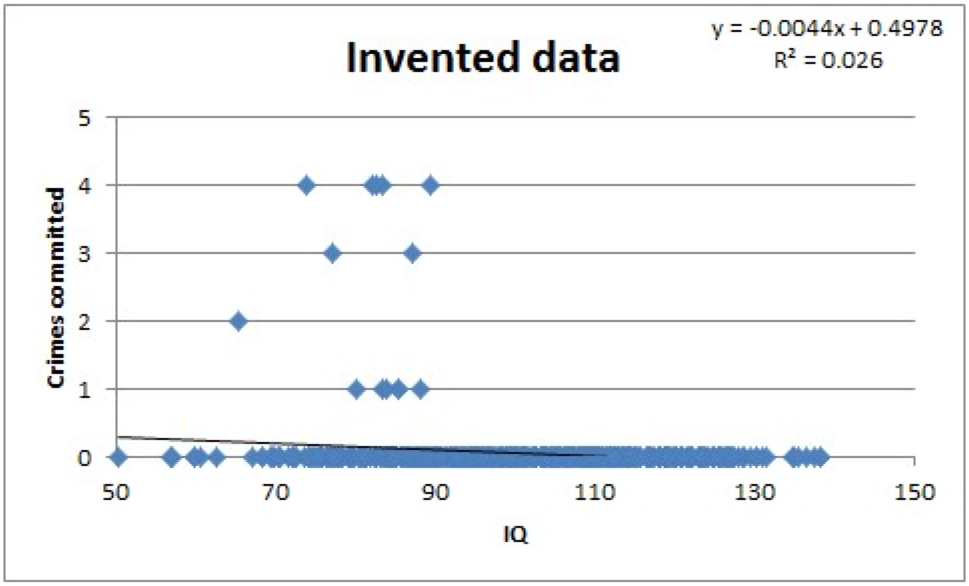
> Terrie Moffitt and colleagues studied 4,552 Danish men born at the end of World War II. They examined intelligence test scores collected by the Danish army (for screening potential draftees) and criminal records drawn from the Danish National Police Register. The men who committed two or more criminal offenses by age twenty had IQ scores on average a full standard deviation below nonoffenders, and IQ and criminal offenses were significantly and negatively correlated at r = -.19.

Repeat offenders are a 15 IQ points – an entire standard deviation – below the rest of the population. This matches common sense, which suggests that serial criminals are not the brightest members of society. It sounds from this like IQ is a very important predictor of crime.

But r = – 0.19 suggests that only about 3.6% of variance in crime is predicted by IQ. 3.6% is nothing. It sounds from this like IQ barely matters at all in predicting crime.

This isn’t a matter of conflicting studies: these are two ways of describing the same data. What gives?

The best answer I got was from pappubahry2, who posted the following made-up graph:

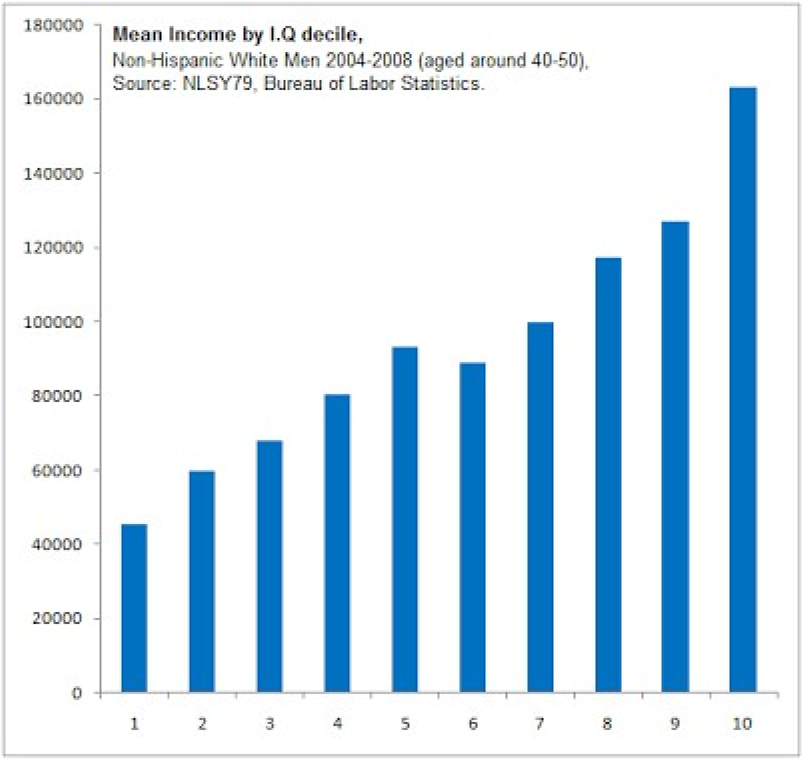


Here *all* crime is committed by low IQ individuals, but the correlation between IQ and crime is still very low, r = 0.16. The reason is simple: very few people, including very few low-IQ people, commit crimes. r is kind of a mishmash of p(low IQ|criminal) and p(criminal|low IQ), and the latter may be very low even when all criminals are from the lower end of the spectrum.

The advice some people on Tumblr gave was to beware summary statistics. “IQ only predicts 3.6% of variance in crime” makes it sound like IQ is nearly irrelevant to criminality, but in fact it’s perfectly consistant with IQ being a very strong predictive factor.

II.

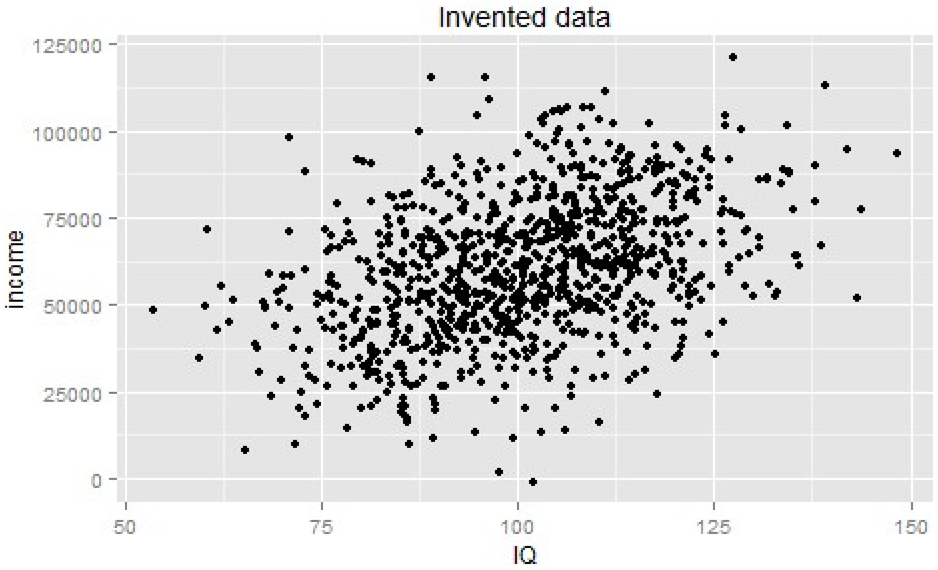
So I pressed my luck with the following question:

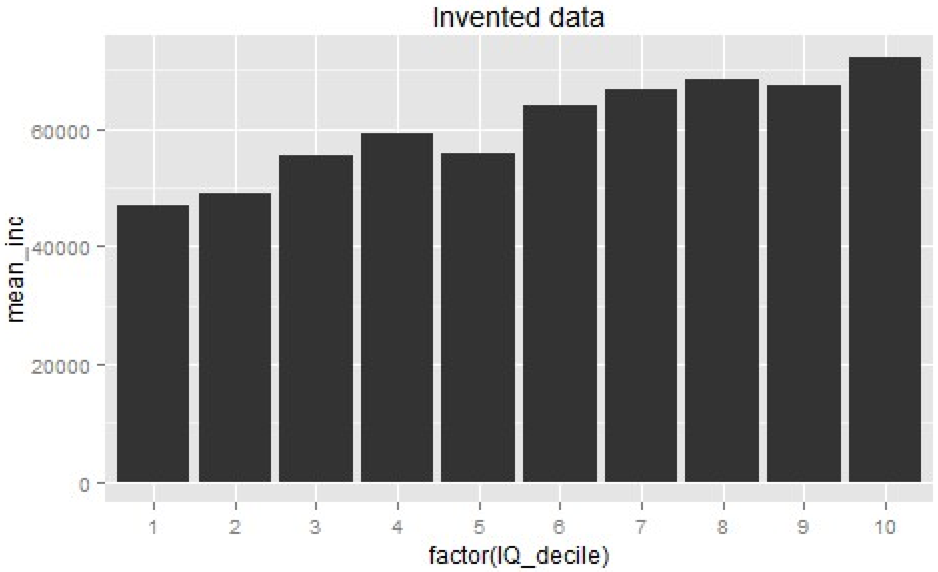


I'm not sure why everyone's income on this graph is so much higher than average US per capita of $30,000ish, or even average white male income of $31,000ish. I think it might be the 'age 40 to 50' specifier.

This graph suggests IQ is an important determinant of income. But most studies say the correlation between IQ and income is at most 0.4 or so, or 16% of the variance, suggesting it's a very minor determinant of income. Most people are earning an income, so the too-few-criminals explanation from above doesn't apply. Again, what gives?

The best answer I got for this one was from su3su2u1, who pointed out that there was probably very high variance within the individual deciles. Pappubahry made some more graphs to demonstrate:





I understand this one intellectually, but I still haven't gotten my head around it. Regardless of the amount of variance, going from a category where I can expect to make on average $40,000 to a category where I could expect to make on average $160,000 seems like a pretty big deal, and describing it as "only predicting 16% of the variation" seems patently unfair.

I guess the moral is the same as the moral in the first situation: beware summary statistics. Based on the way you explain things, you can use different summary statistics to make things look very important or not important at all. And as a bunch of people recommended to me: when in doubt, demand to see the scatter plot.

AI Researchers On AI Risk

I first became interested in AI risk back around 2007. At the time, most people's response to the topic was "Haha, come back when anyone believes this besides random Internet crackpots."

Over the next few years, a series of extremely bright and influential figures including Bill Gates, Stephen Hawking, and Elon Musk publically announced they were concerned about AI risk, along with hundreds of other intellectuals, from Oxford philosophers to MIT cosmologists to Silicon Valley tech investors. So we came back.

Then the response changed to "Sure, a couple of random academics and businesspeople might believe this stuff, but never *real experts* in the field who know what's going on."

Thus pieces like Popular Science's Bill Gates Fears AI, But AI Researchers Know Better:

> When you talk to A.I. researchers—again, genuine A.I. researchers, people who grapple with making systems that work at all, much less work too well—they are not worried about superintelligence sneaking up on them, now or in the future. Contrary to the spooky stories that Musk seems intent on telling, A.I. researchers aren't frantically installed firewalled summoning chambers and self-destruct countdowns.

And Fusion.net's The Case Against Killer Robots From A Guy Actually Building AI:

> Andrew Ng builds artificial intelligence systems for a living. He taught AI at Stanford, built AI at Google, and

> then moved to the Chinese search engine giant, Baidu, to continue his work at the forefront of applying artificial intelligence to real-world problems. So when he hears people like Elon Musk or Stephen Hawking—people who are not intimately familiar with today's technologies—talking about the wild potential for artificial intelligence to, say, wipe out the human race, you can practically hear him facepalming.

And now Ramez Naam of Marginal Revolution is trying the same thing with What Do AI Researchers Think Of The Risk Of AI?:

> Elon Musk, Stephen Hawking, and Bill Gates have recently expressed concern that development of AI could lead to a 'killer AI' scenario, and potentially to the extinction of humanity. None of them are AI researchers or have worked substantially with AI that I know of. What do actual AI researchers think of the risks of AI?

It quotes the same couple of cherry-picked AI researchers as all the other stories – Andrew Ng, Yann LeCun, etc – then stops without mentioning whether there are alternate opinions.

There are. AI researchers, including some of the leaders in the field, have been instrumental in raising issues about AI risk and superintelligence from the very beginning. I want to start by listing some of these people, as kind of a counter-list to Naam's, then go into why I don't think this is a "controversy" in the classical sense that dueling lists of luminaries might lead you to expect.

The criteria for my list: I'm only mentioning the most prestigious researchers, either full professors at good schools with lots of highly-cited papers, or else very-well respected scientists in industry working at big companies with good track records. They have to be involved in AI and machine learning. They have to have multiple strong statements supporting some kind of view about a near-term singularity and/or extreme risk from superintelligent AI. Some will have written papers or books about it; others will have just gone on the record saying they think it's important and worthy of further study.

If anyone disagrees with the inclusion of a figure here, or knows someone important I forgot, let me know and I'll make the appropriate changes:

* * * * * * * * * *

Stuart Russell (wiki) is Professor of Computer Science at Berkeley, winner of the IJCAI Computers And Thought Award, Fellow of the Association for Computing Machinery, Fellow of the American Academy for the Advancement of Science, Director of the Center for Intelligent Systems, Blaise Pascal Chair in Paris, etc, etc. He is the co-author of Artificial Intelligence: A Modern Approach, the classic textbook in the field used by 1200 universities around the world. On his website, he writes:

> The field [of AI] has operated for over 50 years on one simple assumption: the more intelligent, the better. To this must be conjoined an overriding concern for the benefit of humanity. The argument is very simple:
>
> 1. AI is likely to succeed.
> 2. Unconstrained success brings huge risks and huge benefits.
> 3. What can we do now to improve the chances of reaping the benefits and avoiding the risks?

> Some organizations are already considering these questions, including the Future of Humanity Institute at Oxford, the Centre for the Study of Existential Risk at Cambridge, the Machine Intelligence Research Institute in Berkeley, and the Future of Life Institute at Harvard/MIT. I serve on the Advisory Boards of CSER and FLI.
>
> Just as nuclear fusion researchers consider the problem of containment of fusion reactions as one of the primary problems of their field, it seems inevitable that issues of control and safety will become central to AI as the field matures. The research questions are beginning to be formulated and range from highly technical (foundational issues of rationality and utility, provable properties of agents, etc.) to broadly philosophical.

He makes a similar point on edge.org, writing:

> As Steve Omohundro, Nick Bostrom, and others have explained, the combination of value misalignment with increasingly capable decision-making systems can lead to problems—perhaps even species-ending problems if the machines are more capable than humans. Some have argued that there is no conceivable risk to humanity for centuries to come, perhaps forgetting that the interval of time between Rutherford’s confident assertion that atomic energy would never be feasibly extracted and Szilárd’s invention of the neutron-induced nuclear chain reaction was less than twenty-four hours.

He has also tried to serve as an ambassador about these issues to other academics in the field, writing:

> What I’m finding is that senior people in the field who have never publicly evinced any concern before are

> privately thinking that we do need to take this issue very seriously, and the sooner we take it seriously the better.

David McAllester (wiki) is professor and Chief Academic Officer at the U Chicago-affilitated Toyota Technological Institute, and formerly served on the faculty of MIT and Cornell. He is a fellow of the American Association of Artificial Intelligence, has authored over a hundred publications, has done research in machine learning, programming language theory, automated reasoning, AI planning, and computational linguistics, and was a major influence on the algorithms for famous chess computer Deep Blue. According to an article in the Pittsburgh Tribune Review:

> Chicago professor David McAllester believes it is inevitable that fully automated intelligent machines will be able to design and build smarter, better versions of themselves, an event known as the Singularity. The Singularity would enable machines to become infinitely intelligent, and would pose an ‘incredibly dangerous scenario’, he says.

On his personal blog Machine Thoughts, he writes:

> Most computer science academics dismiss any talk of real success in artificial intelligence. I think that a more rational position is that no one can really predict when human level AI will be achieved. John McCarthy once told me that when people ask him when human level AI will be achieved he says between five and five hundred years from now. McCarthy was a smart man. Given the uncertainties surrounding AI, it seems prudent to consider the issue of friendly AI…

> The early stages of artificial general intelligence (AGI) will be safe. However, the early stages of AGI will provide an excellent test bed for the servant mission or other approaches to friendly AI. An experimental approach has also been promoted by Ben Goertzel in a nice blog post on friendly AI. If there is a coming era of safe (not too intelligent) AGI then we will have time to think further about later more dangerous eras.

He attended the AAAI Panel On Long-Term AI Futures, where he chaired the panel on Long-Term Control and was described as saying:

> McAllester chatted with me about the upcoming 'Singularity', the event where computers out think humans. He wouldn't commit to a date for the singularity but said it could happen in the next couple of decades and will definitely happen eventually. Here are some of McAllester's views on the Singularity. There will be two milestones: Operational Sentience, when we can easily converse with computers, and the AI Chain Reaction, when a computer can bootstrap itself to a better self and repeat. We'll notice the first milestone in automated help systems that will genuinely be helpful. Later on computers will actually be fun to talk to. The point where computer can do anything humans can do will require the second milestone.

Hans Moravec (wiki) is a former professor at the Robotics Institute of Carnegie Mellon University, namesake of Moravec's Paradox, and founder of the SeeGrid Corporation for industrial robotic visual systems. His Sensor Fusion in Certainty Grids for Mobile Robots has been cited over a thousand times, and he was invited to write the Encyclopedia Britannica article on robotics back when encyclopedia articles were written by the world expert in a field rather than by hundreds of anonymous Internet commenters.

He is also the author of Robot: Mere Machine to Transcendent Mind, which Amazon describes as:

> In this compelling book, Hans Moravec predicts machines will attain human levels of intelligence by the year 2040, and that by 2050, they will surpass us. But even though Moravec predicts the end of the domination by human beings, his is not a bleak vision. Far from railing against a future in which machines rule the world, Moravec embraces it, taking the startling view that intelligent robots will actually be our evolutionary heirs.” Moravec goes further and states that by the end of this process “the immensities of cyberspace will be teeming with unhuman superminds, engaged in affairs that are to human concerns as ours are to those of bacteria”.

Shane Legg is co-founder of DeepMind Technologies (wiki), an AI startup that was bought for Google in 2014 for about $500 million. He earned his PhD at the Dalle Molle Institute for Artificial Intelligence in Switzerland and also worked at the Gatsby Computational Neuroscience Unit in London. His dissertation Machine Superintelligence concludes:

> If there is ever to be something approaching absolute power, a superintelligent machine would come close. By definition, it would be capable of achieving a vast range of goals in a wide range of environments. If we carefully prepare for this possibility in advance, not only might we avert disaster, we might bring about an age of prosperity unlike anything seen before.

In a later interview, he states:

> AI is now where the internet was in 1988. Demand for machine learning skills is quite strong in specialist applications (search companies like Google, hedge funds and bio-informatics) and is growing every year. I expect this to become noticeable in the mainstream around the middle of the next decade. I expect a boom in AI around 2020 followed by a decade of rapid progress, possibly after a market correction. Human level AI will be passed in the mid 2020's, though many people won't accept that this has happened. After this point the risks associated with advanced AI will start to become practically important…I don't know about a "singularity", but I do expect things to get really crazy at some point after human level AGI has been created. That is, some time from 2025 to 2040.

He and his co-founders Demis Hassabis and Mustafa Suleyman have signed the Future of Life Institute petition on AI risks, and one of their conditions for joining Google was that the company agree to set up an AI Ethics Board to investigate these issues.

Steve Omohundro (wiki) is a former Professor of Computer Science at University of Illinois, founder of the Vision and Learning Group and the Center for Complex Systems Research, and inventor of various important advances in machine learning and machine vision. His work includes lip-reading robots, the StarLisp parallel programming language, and geometric learning algorithms. He currently runs Self-Aware Systems, "a think-tank working to ensure that intelligent technologies are beneficial for humanity". His paper Basic AI Drives helped launch the field of machine ethics by pointing out that superintelligent systems will converge upon certain potentially dangerous goals. He writes:

> We have shown that all advanced AI systems are likely to exhibit a number of basic drives. It is essential that we understand these drives in order to build technology that enables a positive future for humanity. Yudkowsky has called for the creation of ‘friendly AI’. To do this, we must develop the science underlying ‘utility engineering’, which will enable us to design utility functions that will give rise to the consequences we desire…The rapid pace of technological progress suggests that these issues may become of critical importance soon.”

See also his section here on “Rational AI For The Greater Good”.

Murray Shanahan (site) earned his PhD in Computer Science from Cambridge and is now Professor of Cognitive Robotics at Imperial College London. He has published papers in areas including robotics, logic, dynamic systems, computational neuroscience, and philosophy of mind. He is currently writing a book The Technological Singularity which will be published in August; Amazon’s blurb says:

> Shanahan describes technological advances in AI, both biologically inspired and engineered from scratch. Once human-level AI — theoretically possible, but difficult to accomplish — has been achieved, he explains, the transition to superintelligent AI could be very rapid. Shanahan considers what the existence of superintelligent machines could mean for such matters as personhood, responsibility, rights, and identity. Some superhuman AI agents might be created to benefit humankind; some might go rogue. (Is Siri the template, or HAL?) The singularity presents both an existential threat to humanity and an existential opportunity for humanity to transcend its limitations. Shanahan makes

> it clear that we need to imagine both possibilities if we want to bring about the better outcome.

Marcus Hutter (wiki) is a professor in the Research School of Computer Science at Australian National University. He has previously worked with the Dalle Molle Institute for Artificial Intelligence and National ICT Australia, and done work on reinforcement learning, Bayesian sequence prediction, complexity theory, Solomonoff induction, computer vision, and genomic profiling. He has also written extensively on the Singularity. In Can Intelligence Explode?, he writes:

> This century may witness a technological explosion of a degree deserving the name singularity. The default scenario is a society of interacting intelligent agents in a virtual world, simulated on computers with hyperbolically increasing computational resources. This is inevitably accompanied by a speed explosion when measured in physical time units, but not necessarily by an intelligence explosion…if the virtual world is inhabited by interacting free agents, evolutionary pressures should breed agents of increasing intelligence that compete about computational resources. The end-point of this intelligence evolution/acceleration (whether it deserves the name singularity or not) could be a society of these maximally intelligent individuals. Some aspect of this singularitarian society might be theoretically studied with current scientific tools. Way before the singularity, even when setting up a virtual society in our imagine, there are likely some immediate difference, for example that the value of an individual life suddenly drops, with drastic consequences.

Jurgen Schmidhuber (wiki) is Professor of Artificial Intelligence at the University of Lugano and former

Professor of Cognitive Robotics at the Technische Universitat Munchen. He makes some of the most advanced neural networks in the world, has done further work in evolutionary robotics and complexity theory, and is a fellow of the European Academy of Sciences and Arts. In Singularity Hypotheses, Schmidhuber argues that “if future trends continue, we will face an intelligence explosion within the next few decades”. When asked directly about AI risk on a Reddit AMA thread, he answered:

> Stuart Russell’s concerns [about AI risk] seem reasonable. So can we do anything to shape the impacts of artificial intelligence? In an answer hidden deep in a related thread I just pointed out: At first glance, recursive self-improvement through Gödel Machines seems to offer a way of shaping future superintelligences. The self-modifications of Gödel Machines are theoretically optimal in a certain sense. A Gödel Machine will execute only those changes of its own code that are provably good, according to its initial utility function. That is, in the beginning you have a chance of setting it on the “right” path. Others, however, may equip their own Gödel Machines with different utility functions. They will compete. In the resulting ecology of agents, some utility functions will be more compatible with our physical universe than others, and find a niche to survive. More on this in a paper from 2012.

Richard Sutton (wiki) is professor and iCORE chair of computer science at University of Alberta. He is a fellow of the Association for the Advancement of Artificial Intelligence, co-author of the most-used textbook on reinforcement learning, and discoverer of temporal difference learning, one of the most important methods in the field.

In his talk at the Future of Life Institute's Future of AI Conference, Sutton states that there is "certainly a significant chance within all of our expected lifetimes" that human-level AI will be created, then goes on to say the AIs "will not be under our control", "will compete and cooperate with us", and that "if we make superintelligent slaves, then we will have superintelligent adversaries". He concludes that "We need to set up mechanisms (social, legal, political, cultural) to ensure that this works out well" but that "inevitably, conventional humans will be less important." He has also mentioned these issues at a presentation to the Gadsby Institute in London and in (of all things) a Glenn Beck book: "Richard Sutton, one of the biggest names in AI, predicts an intelligence explosion near the middle of the century".

Andrew Davison (site) is Professor of Robot Vision at Imperial College London, leader of the Robot Vision Research Group and Dyson Robotics Laboratory, and inventor of the computerized localization-mapping system MonoSLAM. On his website, he writes:

> At the risk of going out on a limb in the proper scientific circles to which I hope I belong(!), since 2006 I have begun to take very seriously the idea of the technological singularity: that exponentially increasing technology might lead to super-human AI and other developments that will change the world utterly in the surprisingly near future (i.e. perhaps the next 20–30 years). As well as from reading books like Kurzweil's 'The Singularity is Near' (which I find sensational but on the whole extremely compelling), this view comes from my own overview of incredible recent progress of science and technology in general and specifically in the fields of computer vision and robotics within which I am personally working. Modern inference, learning and

> estimation methods based on Bayesian probability theory (see Probability Theory: The Logic of Science or free online version, highly recommended), combined with the exponentially increasing capabilities of cheaply available computer processors, are becoming capable of amazing human-like and super-human feats, particularly in the computer vision domain.
>
> It is hard to even start thinking about all of the implications of this, positive or negative, and here I will just try to state facts and not offer much in the way of opinions (though I should say that I am definitely not in the super-optimistic camp). I strongly think that this is something that scientists and the general public should all be talking about. I'll make a list here of some 'singularity indicators' I come across and try to update it regularly. These are little bits of technology or news that I come across which generally serve to reinforce my view that technology is progressing in an extraordinary, faster and faster way that will have consequences few people are yet really thinking about.

Alan Turing and I. J. Good (wiki, wiki) are men who need no introduction. Turing invented the mathematical foundations of computing and shares his name with Turing machines, Turing completeness, and the Turing Test. Good worked with Turing at Bletchley Park, helped build some of the first computers, and invented various landmark algorithms like the Fast Fourier Transform. In his paper "Can Digital Machines Think?", Turing writes:

> Let us now assume, for the sake of argument, that these machines are a genuine possibility, and look at the consequences of constructing them. To do so would of course meet with great opposition, unless we have advanced greatly in religious tolerance since the days of

> Galileo. There would be great opposition from the intellectuals who were afraid of being put out of a job. It is probable though that the intellectuals would be mistaken about this. There would be plenty to do in trying to keep one's intelligence up to the standards set by the machines, for it seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers…At some stage therefore we should have to expect the machines to take control.

During his time at the Atlas Computer Laboratory in the 60s, Good expanded on this idea in Speculations Concerning The First Ultraintelligent Machine, which argued:

> Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an 'intelligence explosion,' and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make

* * * * * * * * * *

I worry this list will make it look like there is some sort of big "controversy" in the field between "believers" and "skeptics" with both sides lambasting the other. This has not been my impression.

When I read the articles about skeptics, I see them making two points over and over again. First, we are nowhere near human-level intelligence right now, let alone superintelligence, and there's no obvious path to get there from here. Second, if you start demanding bans on AI research then you are an idiot.

I agree whole-heartedly with both points. So do the leaders of the AI risk movement.

A survey of AI researchers (Muller & Bostrom, 2014) finds that on average they expect a 50% chance of human-level AI by 2040 and 90% chance of human-level AI by 2075. On average, 75% believe that superintelligence ("machine intelligence that greatly surpasses the performance of every human in most professions") will follow within thirty years of human-level AI. There are some reasons to worry about sampling bias based on eg people who take the idea of human-level AI seriously being more likely to respond (though see the attempts made to control for such in the survey) but taken seriously it suggests that most AI researchers think there's a good chance this is something we'll have to worry about within a generation or two.

But outgoing MIRI director Luke Muehlhauser and Future of Humanity Institute director Nick Bostrom are both on record saying they have significantly *later* timelines for AI development than the scientists in the survey. If you look at Stuart Armstrong's AI Timeline Prediction Data there doesn't seem to be any general law that the estimates from AI risk believers are any earlier than those from AI risk skeptics. In fact, the latest estimate on the entire table is from Armstrong himself; Armstrong nevertheless currently works at the Future of Humanity Institute raising awareness of AI risk and researching superintelligence goal alignment.

The difference between skeptics and believers isn't about when human-level AI will arrive, it's about when we should start preparing.

Which brings us to the second non-disagreement. The “skeptic” position seems to be that, although we should probably get a couple of bright people to start working on preliminary aspects of the problem, we shouldn’t panic or start trying to ban AI research.

The “believers”, meanwhile, insist that although we shouldn’t panic or start trying to ban AI research, we should probably get a couple of bright people to start working on preliminary aspects of the problem.

Yann LeCun is probably the most vocal skeptic of AI risk. He was heavily featured in the Popular Science article, was quoted in the Marginal Revolution post, and spoke to KDNuggets and IEEE on “the inevitable singularity questions”, which he describes as “so far out that we can write science fiction about it”. But when asked to clarify his position a little more, he said:

> Elon [Musk] is very worried about existential threats to humanity (which is why he is building rockets with the idea of sending humans colonize other planets). Even if the risk of an A.I. uprising is very unlikely and very far in the future, we still need to think about it, design precautionary measures, and establish guidelines. Just like bio-ethics panels were established in the 1970s and 1980s, before genetic engineering was widely used, we need to have A.I.-ethics panels and think about these issues. But, as Yoshua [Bengio] wrote, we have quite a bit of time

Eric Horvitz is another expert often mentioned as a leading voice of skepticism and restraint. His views have been profiled in articles like Out Of Control AI Will Not Kill Us, Believes Microsoft Research Chief and Nothing To Fear From

Artificial Intelligence, Says Microsoft's Eric Horvitz. But here's what he says in a longer interview with NPR:

> KASTE: Horvitz doubts that one of these virtual receptionists could ever lead to something that takes over the world. He says that's like expecting a kite to evolve into a 747 on its own. So does that mean he thinks the singularity is ridiculous?
>
> Mr. HORVITZ: Well, no. I think there's been a mix of views, and I have to say that I have mixed feelings myself.
>
> KASTE: In part because of ideas like the singularity, Horvitz and other A.I. scientists have been doing more to look at some of the ethical issues that might arise over the next few years with narrow A.I. systems. They've also been asking themselves some more futuristic questions. For instance, how would you go about designing an emergency off switch for a computer that can redesign itself?
>
> Mr. HORVITZ: I do think that the stakes are high enough where even if there was a low, small chance of some of these kinds of scenarios, that it's worth investing time and effort to be proactive.

Which is pretty much the same position as a lot of the most zealous AI risk proponents. With enemies like these, who needs friends?

A Slate article called Don't Fear Artificial Intelligence also gets a surprising amount right:

> As Musk himself suggests elsewhere in his remarks, the solution to the problem [of AI risk] lies in sober and considered collaboration between scientists and

> policymakers. However, it is hard to see how talk of "demons" advances this noble goal. In fact, it may actively hinder it.
>
> First, the idea of a Skynet scenario itself has enormous holes. While computer science researchers think Musk's musings are "not completely crazy," they are still awfully remote from a world in which AI hype masks less artificially intelligent realities that our nation's computer scientists grapple with:
>
> Yann LeCun, the head of Facebook's AI lab, summed it up in a Google+ post back in 2013: "Hype is dangerous to AI. Hype killed AI four times in the last five decades. AI Hype must be stopped."…LeCun and others are right to fear the consequences of hype. Failure to live up to sci-fi–fueled expectations, after all, often results in harsh cuts to AI research budgets.

AI scientists are all smart people. They have no interest in falling into the usual political traps where they divide into sides that accuse each other of being insane alarmists or ostriches with their heads stuck in the sand. It looks like they're trying to balance the need to start some preliminary work on a threat that looms way off in the distance versus the risk of engendering so much hype that it starts a giant backlash.

This is not to say that there aren't very serious differences of opinion in how quickly we need to act. These seem to hinge mostly on whether it's safe to say "We'll deal with the problem when we come to it" or whether there will be some kind of "hard takeoff" which will take events out of control so quickly that we'll want to have done our homework beforehand. I continue to see less evidence than I'd like that most AI researchers with opinions understand the latter possibility, or really any of the technical work in this area. Heck, the Marginal Revolution article quotes an expert as saying that superintelligence isn’t a big risk because “smart computers won’t create their own goals”, even though anyone who has read Bostrom knows that this is *exactly* the problem.

There is still a lot of work to be done. But cherry-picked articles about how “real AI researchers don’t worry about superintelligence” aren’t it.

[thanks to some people from MIRI and FLI for help with and suggestions on this post]

EDIT: Investigate for possible inclusion: Fredkin, Minsky

No Time Like The Present For AI Safety Work

I.

On the recent post on AI risk, a commenter challenged me to give the short version of the argument for taking it seriously. I said something like:

1. If humanity doesn't blow itself up, eventually we will create human-level AI.

2. If humanity creates human-level AI, technological progress will continue and eventually reach far-above-human-level AI

3. If far-above-human-level AI comes into existence, eventually it will so overpower humanity that our existence will depend on its goals being aligned with ours

4. It is possible to do useful research now which will improve our chances of getting the AI goal alignment problem right

5. Given that we can start research now we probably should, since leaving it until there is a clear and present need for it is unwise

I placed very high confidence (>95%) on each of the first three statements – they're just saying that if trends continue moving towards a certain direction without stopping, eventually they'll get there. I had lower confidence (around 50%) on the last two statements.

Commenters tended to agree with this assessment; nobody wanted to seriously challenge any of 1-3, but a lot of people said they just didn’t think there was any point in worrying about AI now. We ended up in an extended analogy about illegal computer hacking. It’s a big problem that we’ve never been able to fully address – but if Alan Turing had gotten it into his head to try to solve it in 1945, his ideas might have been along the lines of “Place your punch cards in a locked box where German spies can’t read them.” Wouldn’t trying to solve AI risk in 2015 end in something equally cringeworthy?

Maybe. But I disagree for a couple reasons, some of them broad and meta-level, some of them more focused and object level. The most important meta-level consideration is: if you’re accepting points 1 to 3 – that is, you accept that eventually the human race is going to go extinct or worse if we can’t figure out AI goal alignment – do you *really* think our chances of making a dent in the problem today are so low that saying “Yes, we’re on a global countdown to certain annhilation, but it would be an inefficient use of resources to even investigate if we could do anything about it at this point”? What is this amazing other use of resources that you prefer? Like, go on and grumble about Pascal’s Wager, but you *do* realize we just paid Floyd Mayweather ten times more money than has been spent on AI risk total throughout all of human history to participate in *a single boxing fight*, right?

(if AI boxing got a tenth as much attention, or a hundredth as much money, as AI boxing, the world would be a much safer place)

But I want to make a stronger claim: not just that dealing with AI risk is more important than boxing, but that it is as important as all the other things we consider important, like curing diseases and detecting asteroids and saving the environment. That requires at least a little argument for why progress should indeed be possible at this early stage.

And I think progress is possible insofar as this is a philosophical and not a technical problem. Right now the goal isn't "write the code that will control the future AI", it's "figure out the broad category of problem we have to deal with." Let me give some examples of open problems to segue into a discussion of why these problems are worth working on now.

II.

Problem 1: Wireheading

Some people have gotten electrodes implanted in their brains for therapeutic or research purposes. When the electrodes are in certain regions, most notably the lateral hypothalamus, the people become obsessed with stimulating them as much as possible. If you give them the stimulation button, they'll press it thousands of times per hour; if you try to take the stimulation button away from them, they'll defend it with desperation and ferocity. Their life and focus narrows to a pinpoint, normal goals like love and money and fame and friendship forgotten in the relentless drive to stimulate the electrode as much as possible.

This fits pretty well with what we know of neuroscience. The brain (OVERSIMPLIFICATION WARNING) represents reward as electrical voltage at a couple of reward centers, then does whatever tends to maximize that reward. Normally this works pretty well; when you fulfill a biological drive like food or sex, the reward center responds with little bursts of reinforcement, and so you continue fulfilling your biological drives. But stimulating the reward center directly with an electrode increases it much more than waiting for your brain to send little bursts of stimulation the natural way, so this activity is by definition the most rewarding possible. A person presented with the opportunity to stimulate the reward center directly will forget about all those indirect ways of getting reward like “living a happy life” and just press the button attached to the electrode as much as possible.

This doesn’t even require any brain surgery – drugs like cocaine and meth are addictive in part because they interfere with biochemistry to increase the level of stimulation in reward centers.

And computers can run into the same issue. I can’t find the link, but I do remember hearing about an evolutionary algorithm designed to write code for some application. It generated code semi-randomly, ran it by a “fitness function” that assessed whether it was any good, and the best pieces of code were “bred” with each other, then mutated slightly, until the result was considered adequate.

They ended up, of course, with code that hacked the fitness function and set it to some absurdly high integer.

These aren’t isolated incidents. Any mind that runs off of reinforcement learning with a reward function – and this seems near-universal in biological life-forms and is increasingly common in AI – will have the same design flaw. The main defense against it this far is simple lack of capability: most computer programs aren’t smart enough for “hack your own reward function” to be an option; as for humans, our reward centers are hidden way inside our heads where we can’t get to it. A hypothetical superintelligence won’t have this problem: it will know exactly where its reward center is and be intelligent enough to reach it and reprogram it.

The end result, unless very deliberate steps are taken to prevent it, is that an AI designed to cure cancer hacks its own module determining how much cancer has been cured and sets it to the highest number its memory is capable of representing. Then it goes about acquiring more memory so it can represent higher numbers. If it’s superintelligent, its options for acquiring new memory include “take over all the computing power in the world” and “convert things that aren’t computers into computers.” Human civilization is a thing that isn’t a computer.

This is not some exotic failure mode that a couple of extremely bizarre designs can fall into; this may be the *natural course* for a sufficiently intelligent reinforcement learner.

Problem 2: Weird Decision Theory

Pascal’s Wager is a famous argument for why you should join an organized religion. Even if you believe God is vanishingly unlikely to exist, the consequence of being wrong (Hell) is so great, and the benefits of being right (not having to go to church on Sundays) so comparatively miniscule, that you should probably just believe in God to be on the safe side. Although there are many objections based on the specific content of religion (does God really want someone to believe based on that kind of analysis?) the problem can be generalized into a form where you can make an agent do anything merely by promising a spectacularly high reward; if the reward is high enough, it will overrule any concerns the agent has about your inability to deliver it.

This is a problem of decision theory which is unrelated to questions of intelligence. A very intelligent person might be able to calculate the probability of God existing very accurately, and they might be able to estimate the exact badness of Hell, but without a good decision theory intelligence alone can't save you from Pascal's Wager – in fact, intelligence is what lets you do the formal mathematical calculations telling you to take the bet.

Humans are pretty resistant to this kind of problem – most people aren't moved by Pascal's Wager, even if they can't think of a specific flaw in it – but it's not obvious how exactly we gain our resistance. Computers, which are infamous for relying on formal math but having no common sense, won't have that kind of resistance unless it gets built in. And building it in is a *really* hard problem. Most hacks that eliminate Pascal's Wager without having a deep understanding of where (or whether) the formal math is going on just open up more loopholes somewhere else. A solution based on a deep understanding of where the formal math goes wrong, and which preserves the power of the math to solve everyday situations, has as far as I know not yet been developed. Worse, once we solve Pascal's Wager, there are a couple of dozen very similar decision-theoretic paradoxes that may require entirely different solutions.

This is not a cute little philosophical trick. A sufficiently good "hacker" could subvert a galaxy-spanning artificial intelligence just by threatening (with no credibility) to inflict a spectacularly high punishment on it if it didn't do what the hacker wanted; if the AI wasn't Pascal-proofed, it would decide to do whatever the hacker said.

Problem 3: The Evil Genie Effect

Everyone knows the problem with computers is that they do what you say rather than what you *mean*. Nowadays that just means that a program runs differently when you forget a close-parenthesis, or websites show up weird if you put the HTML codes in the wrong order. But it might lead an artificial intelligence to seriously misinterpret natural language orders.

Age of Ultron actually gets this one sort of right. Tony Stark orders his super-robot Ultron to bring peace to the world; Ultron calculates that the fastest and most certain way to bring peace is to destroy all life. As far as I can tell, Ultron is totally 100% correct about this and in some real-world equivalent that is exactly what would happen. We would get pretty much the same effect by telling an AI to "cure cancer" or "end world hunger" or any of a thousand other things.

> I promise the gender gap, among many other things, will be eliminated. https://t.co/9KDLvaNFob
>
> — Sweet Meteor O'Death (@smod2016) May 24, 2015

Even Isaac Asimov's Three Laws of Robotics would take about thirty seconds to become horrible abominations. The First Laws says a robot cannot harm a human being *or allow through inaction a human being to come to harm*. "Not taking over the government and banning cigarettes" counts as allowing through inaction a human being to come to harm. So does "not locking every human in perfectly safe stasis fields for all eternity."

There is no way to compose an order specific enough to explain exactly what we mean by "do not allow through inaction a human to come to harm" – go ahead, try it – unless the robot is already willing to do what we mean, rather than what we say. This is not a deal-breaker, since AIs may indeed by smart enough to understand what we mean, but our desire that they do so will have to be programmed into them directly, from the ground up. Part of SIAI's old vision of "causal validity semantics" seems to be about laying a groundwork for this program.

But this just leads to a second problem: *we* don't always know what we mean by something. The question of "how do we balance the ethical injunction to keep people safe with the ethical injunction to preserve human freedom?" is a pretty hot topic in politics right now, presenting itself in everything from gun control to banning Big Gulp cups. It seems to involve balancing out everything we value – how important are Big Gulp cups to us, anyway? – and combining cost-benefit calculations with sacred principles. Any AI that couldn't navigate that moral labyrinth might end up ending world hunger by killing all starving people, or refusing else to end world hunger by inventing new crops because the pesticides for them might kill an insect.

But the more you try to study ethics, the more you realize they're really really complicated and so far resist simplification to the sort of formal system that a computer has any hope of understanding. Utilitarianism is *almost* computer-readable, but it runs into various paradoxes at the edges, and even without those you'd need to have a set of utility weights for everything in the world.

This is a problem we have yet to solve with *humans* – most of the humans in the world have values that we consider abhorrent, and accept tradeoffs we consider losing propositions. Dealing with an AI whose mind is no more different to mine than that of fellow human being Pat Robertson would from my perspective be a clear-cut case of failure.

[EDIT: I’m told I’m not explaining this very well. This might be better.]

III.

My point in raising these problems wasn’t to dazzle anybody with interesting philosophical issues. It’s to prove a couple of points:

First, there are some very basic problems that affect broad categories of minds, like “all reinforcement learners” or “all minds that make decisions with formal math”. People often speculate that at this early stage we can’t know anything about the design of future AIs. But I would find it extraordinarily surprising if they used neither reinforcement learning or formal mathematical decision-making.

Second, these problems aren’t obvious to most people. These are weird philosophical quandaries, not things that are obvious to everybody with even a little bit of domain knowledge.

Third, these problems have in fact been thought of. Somebody, whether it was a philosopher or a mathematician or a neuroscientist, sat down and thought “Hey, wait, reinforcement learners are naturally vulnerable to wireheading, which would explain why this same behavior shows up in all of these different domains.”

Fourth, these problems suggest research programs that can be pursued right now, at least in a preliminary way. Why do humans resist Pascal’s Wager so effectively? Can our behavior in high-utility, low-probability situations be fitted to a function that allows a computer to make the same decisions we do? What are the best solutions to the related decision theory problems? How come a human can understand the concept of wireheading, yet not feel any compulsion to seek a brain electrode to wirehead themselves with? Is there a way to design a mind that could wirehead a few times, feel and understand the exact sensation, and yet feel no compulsion to wirehead further? How could we create an idea of human ethics and priorities formal enough to stick into a computer?

I think when people hear “we should start, right now in 2015, working on AI goal alignment issues” they think that somebody wants to write a program that can be imported directly into a 2075 AI to provide it with an artificial conscience. Then they think “No way you can do something that difficult this early on.”

But that isn’t what anybody’s proposing. What we’re proposing is to get ourselves acquainted with the general philosophical problems that affect a broad subset of minds, then pursue the neuroscientific, mathematical, and philosophical investigations necessary to have a good understanding of them by the time the engineering problem comes up.

By analogy, we are nowhere near having spaceships that can travel at even half the speed of light. But we already know the biggest obstacle that an FTL spaceship is going to face (relativity and the light-speed limit) and we already have some ideas for getting around it (the Alcubierre drive). We can’t come anywhere close to building an Alcubierre drive. But if we discover how to make near-lightspeed spaceships in 2100, and for some reason the fate of Earth depends on having faster-than-light spaceships by 2120, it’ll probably be nice that we did all of our Theory-Of-Relativity-discovering early so that we’re not wasting half that time interval debating basic physics.

The question “Can we do basic AI safety research now?” is silly because we have *already* done some basic AI safety research successfully. It’s led to understanding issues like the three problems mentioned above, and many more. There are even a couple of answers now, although they’re at technical levels much lower than any of those big questions. Every step we finish now is one that we don’t have to waste valuable time retracing during the crunch period.

IV.

That last section discussed my claim 4, that there’s research we can do now that will help. That leaves claim 5 – given that we *can* do research now, we *should*, because we can’t just trust our descendents in the crunch time to sort things out on their own without our help, using their better model of what eventual AI might look like. There are a couple of reasons for this

Reason 1: The Treacherous Turn

Our descendents’ better models of AI might be actively misleading. Things that work for subhuman or human level intelligences might fail for superhuman intelligences. Empirical testing won’t be able to figure this out without help from armchair philosophy.

Pity poor evolution. It had hundreds of millions of years to evolve defenses against heroin – which by the way affects rats much as it does humans – but it never bothered. Why not? Because until the past century, there wasn’t anything around intelligent enough to synthesize pure heroin. So heroin addiction just wasn’t a problem anything had to evolve to deal with. A brain design that looks pretty good in stupid animals like rats and cows becomes very dangerous when put in the hands (well, heads) of humans smart enough to synthesize heroin or wirehead their own pleasure centers.

The same is true of AI. Dog-level AIs aren't going to learn to hack their own reward mechanism. Even human level AIs might not be able to – I couldn't hack a robot reward mechanism if it were presented to me. Superintelligences can. What we might see is reinforcement-learning AIs that work very well at the dog level, very well at the human level, then suddenly blow up at the superhuman level, by which it's time it's too late to stop them.

This is a common feature of AI safety failure modes. If you tell me, as a mere human being, to "make peace", then my best bet might be to become Secretary-General of the United Nations and learn to negotiate very well. Arm me with a few thousand nukes, and it's a different story. A human-level AI might pursue its peace-making or cancer-curing or not-allowing-human-harm-through-inaction-ing through the same prosocial avenues as humans, then suddenly change once it became superintelligent and new options became open. Indeed, the point that will activate the shift is precisely that no humans are able to stop it. If humans can easily shut an AI down, then the most effective means of curing cancer will be for it to research new medicines (which humans will support); if humans can no longer stop an AI, the most effective means of curing cancer is destroying humanity (since it will no longer matter that humans will fight back).

In his book, Nick Bostrom calls this pattern "the treacherous turn", and it will doom anybody who plans to just wait until the AIs exist and then solve their moral failings through trial and error and observation. The better plan is to have a good philosophical understanding of exactly what's going on, so we can predict these turns ahead of time and design systems that avoid them from the ground up.

Reason 2: Hard Takeoff

Nathan Taylor of Praxtime writes:

> Arguably most of the current "debates" about AI Risk are mere proxies for a single, more fundamental disagreement: hard versus soft takeoff.

Soft takeoff means AI progress takes a leisurely course from the subhuman level to the dumb-human level to the smarter-human level to the superhuman level over many decades. Hard takeoff means the same course takes much shorter, maybe days to months.

It seems in theory that by hooking a human-level AI to a calculator app, we can get it to the level of a human with lightning-fast calculation abilities. By hooking it up to Wikipedia, we can give it all human knowledge. By hooking it up to a couple extra gigabytes of storage, we can give it photographic memory. By giving it a few more processors, we can make it run a hundred times faster, such that a problem that takes a normal human a whole day to solve only takes the human-level AI fifteen minutes.

So we've already gone from "mere human intelligence" to "human with all knowledge, photographic memory, lightning calculations, and solves problems a hundred times faster than anyone else." This suggests that "merely human level intelligence" isn't mere.

The next problem is "recursive self-improvement". Maybe this human-level AI armed with photographic memory and a hundred-time-speedup takes up computer science. Maybe, with its ability to import entire textbooks in seconds, it becomes very good at computer science. This would allow it to fix its own algorithms to make itself even more intelligent, which would allow it to see new ways to make itself even *more* intelligent, and so on. The end result is that it either reaches some natural plateau or becomes superintelligent in the blink of an eye.

If it's the second one, "wait for the first human-level intelligences and then test them exhaustively" isn't going to cut it. The first human-level intelligence will become the first superintelligence too quickly to solve even the first of the hundreds of problems involved in machine goal-alignment.

And although I haven't seen anyone else bring this up, I'd argue that even the hard-takeoff scenario might be underestimating the risks.

Imagine that for some reason having two hundred eyes is the killer app for evolution. A hundred ninety-nine eyes are useless, no better than the usual two, but once you get two hundred, your species dominates the world forever.

The really hard part of having two hundred eyes is evolving the eye at all. After you've done that, having two hundred of them is very easy. But it might be that it would take eons and eons before any organism reached the two hundred eye sweet spot. Having dozens of eyes is such a useless waste of energy that evolution might never get to the point where it could test the two-hundred-eyed design.

Consider that the same might be true for intelligence. The hard part is evolving so much as a tiny rat brain. Once you've got that, getting a human brain, with its world-dominating capabilities, is just a matter of scaling up. But since brains are metabolically wasteful and not that useful before the technology-discovering point, it took eons before evolution got there.

There's a lot of evidence that this is true. First of all, humans evolved from chimps in just a couple of million years. That's too short to redesign the mind from the ground up, or even invent any interesting new evolutionary "technologies". It's just enough time for evolution to alter the scale and add a couple of efficiency tweaks. But monkeys and apes were around for tens of millions of years before evolution bothered.

Second, dolphins are almost as intelligent as humans. But they last shared a common ancestor with us something like fifty million years ago. Either humans and dolphins both evolved fifty million years worth of intelligence "technologies" independently of each other, or else the most recent common ancestor had most of what was necessary for intelligence and humans and dolphins were just the two animals in that vast family tree for whom using them to their full extent became useful. But the most recent common ancestor of humans and dolphins was probably not much more intelligent than a rat itself.

Third, humans can gain intelligence frighteningly quickly when the evolutionary pressures are added. If Cochran is right, Ashkenazi gained ten IQ points in *a thousand years*. Torsion dystonia sufferers can gain five or ten IQ points from *a single mutation*. All of this suggests a picture where intelligence is easy to change, but evolution has decided it just isn't worth it except in very specific situations.

If this is right, then the first rat-level AI will contain most of the interesting discoveries needed to build the first human-level AI and the first superintelligent AI. People tend to say things like "Well, we might have AI as smart as a rat soon, but it will be a long time after that before they're anywhere near human-level". But that's assuming you can't turn the rat into the human just by adding more processing power or more simulated neurons or more connections or whatever. Anything done on a computer doesn't need to worry about metabolic restrictions.

Reason 3: Everyday Ordinary Time Constraints

Bostrom and Mueller surveyed AI researchers about when they expected human-level AI. The median date was 2040. That's 25 years.

People have been thinking about Pascal's Wager (for example) for 345 years now without coming up with any fully generalizable solutions. If that turns out to be a problem for AI, we have 25 more years to solve not only the Wager, but the entire class of problems to which it belongs. Even barring scenarios like unexpected hard takeoffs or treacherous turns, and accepting that if we can solve the problem in 25 years everything will be great, that's not a lot of time.

During the 1956 Dartmouth Conference on AI, top researchers made a plan toward reaching human-level artificial intelligence, and gave themselves two months to teach computers to understand human language. In retrospect, this might have been mildly optimistic.

But now machine translation is a thing, people are making some good progress in some of the hard problems – and when people bring up problems like decision theory, or wireheading, or goal alignment, people just say "Oh, we have plenty of time".

But expecting to solve those problems in a few years might be just as optimistic as expecting to solve machine language translation in two months. Sometimes problems are harder than you think, and it’s worth starting on them early just in case.

All of this means it’s well worth starting armchair work on AI safety now. I won’t say the entire resources of our civilization need to be sunk into it immediately, and I’ve ever heard some people in the field say that after Musk’s $10 million donation money is no longer the most important bottleneck to advancing these ideas. I’m not even sure public exposure is a bottleneck anymore; the median person who watches a movie about killer robots is probably doing more harm than good. If the bottleneck is anything at all, it’s probably intelligent people in relevant fields – philosophy, AI, math, and neuroscience – applying brainpower to these issues and encouraging their colleagues to take them seriously.

That Chocolate Study

Several of you asked me to write about that chocolate article that went viral recently. From I Fooled Millions Into Thinking Chocolate Helps Weight Loss. Here's How:

> "Slim by Chocolate!" the headlines blared. A team of German researchers had found that people on a low-carb diet lost weight 10 percent faster if they ate a chocolate bar every day. It made the front page of Bild, Europe's largest daily newspaper, just beneath their update about the Germanwings crash. From there, it ricocheted around the internet and beyond, making news in more than 20 countries and half a dozen languages. It was discussed on television news shows. It appeared in glossy print, most recently in the June issue of Shape magazine ("Why You Must Eat Chocolate Daily," page 128). Not only does chocolate accelerate weight loss, the study found, but it leads to healthier cholesterol levels and overall increased well-being. The Bild story quotes the study's lead author, Johannes Bohannon, Ph.D., research director of the Institute of Diet and Health: "The best part is you can buy chocolate everywhere."
>
> I am Johannes Bohannon, Ph.D. Well, actually my name is John, and I'm a journalist. I do have a Ph.D., but it's in the molecular biology of bacteria, not humans. The Institute of Diet and Health? That's nothing more than a website.
>
> Other than those fibs, the study was 100 percent authentic. My colleagues and I recruited actual human subjects in Germany. We ran an actual clinical trial, with subjects randomly assigned to different diet regimes.

> And the statistically significant benefits of chocolate that we reported are based on the actual data. It was, in fact, a fairly typical study for the field of diet research. Which is to say: It was terrible science. The results are meaningless, and the health claims that the media blasted out to millions of people around the world are utterly unfounded.

Bohannon goes on to explain that as part of a documentary about “the junk-science diet industry”, he and some collaborators designed a fake study to see if they could convince journalists. They chose to make it about chocolate:

> Gunter Frank, a general practitioner in on the prank, ran the clinical trial. Onneken had pulled him in after reading a popular book Frank wrote railing against dietary pseudoscience. Testing bitter chocolate as a dietary supplement was his idea. When I asked him why, Frank said it was a favorite of the “whole food” fanatics. “Bitter chocolate tastes bad, therefore it must be good for you,” he said. “It’s like a religion.”

They recruited 16 (!) participants and divided them into three groups. One group ate their normal diet. Another ate a low-carb diet. And a third ate a low-carb diet plus some chocolate. Both the low-carb group and the low-carb + chocolate group lost weight compared to the control group, but the low-carb + chocolate group lost weight “ten percent faster”, and the difference was “statistically significant”. They also had “better cholesterol readings” and “higher scores on the well-being survey”.

Bohannon admits exactly how he managed this seemingly impressive result – he measured eighteen different parameters (weight, cholesterol, sodium, protein, etc) which virtually guarantees that one will be statistically significant.

That one turned out to be weight loss. If it had been sodium, he would have published the study as “Chocolate Lowers Sodium Levels”.

Then he pitched it to various fake for-profit journals until one of them bit. Then he put out a PR release to various media outlets, and they ate it up. They ended up in a bunch of English and German language media including Bild, the Daily Star, Times of India, Cosmopolitan, Irish Examiner, and the Huffington Post.

The people I’ve seen discussing this seem to have drawn five conclusions, four of which are wrong:

Conclusion 1: Haha, I can’t believe people were so gullible that they actually thought chocolate caused weight loss!

Bohannon himself endorses this one, saying bitter chocolate was a favorite of “whole food fanatics” because “Bitter chocolate tastes bad, therefore it must be good for you” and “it’s like a religion.

But actually, there’s lots of previous research supporting health benefits from bitter chocolate, none of which Bohannon seems to be aware of.

A meta-analysis of 42 randomized controlled trials totaling 1297 participants in the *American Journal of Clinical Nutrition* found that chocolate improved blood pressure, flow-mediated dilatation (a measure of vascular health), and insulin resistance (related to weight gain).

A different meta-analysis of 24 randomized controlled trials totalling 1106 people in the *Journal of Nutrition* also found that chocolate improved blood pressure, flow-mediated dilatation, and insulin resistance.

A Cochrane Review of 20 randomized controlled trials of 856 people found that chocolate improved blood pressure (it didn't test for flow-mediated dilatation or insulin resistance)

A study on mice found that mice fed more chocolate flavanols were less likely to gain weight.

An epidemiological study of 1018 people in the United States found an association between frequent chocolate consumption and lower BMI, $p < 0.01$. A second epidemiological study of 1458 people in Europe found the same thing, again $p < 0.01$. A cohort study of 470 elderly men found chocolate intake was inversely associated with blood pressure and cardiovascular mortality, p less than 0.001, not confounded by the usual suspects.

I wouldn't find any of these studies alone very convincing. But together, they compensate for each other's flaws and build a pretty robust structure. So the next flawed conclusion is:

Conclusion 2: This proves that nutrition isn't a real science and we should all just be in a state of radical skepticism about these things

What we would like to do is a perfect study where we get thousands of people, randomize them to eat-lots-of-chocolate or eat-little-chocolate at birth, then follow their weights over their entire lives. That way we could have a large sample size, perfect randomization, life-long followup, and clear applicability to other people. But for practical and ethical reasons, we can't do that. So we do a bunch of smaller studies that each capture a few of the features of the perfect study.

First we do animal studies, which can have large sample sizes, perfect randomization, and life-long followup, but it's not clear whether it applies to humans.

Then we do short randomized controlled trials, which can have large sample sizes, perfect randomization, and human applicability, but which only last a couple of months.

Then we do epidemiological studies, which can have large sample sizes, human applicability, and last for many decades, but which aren't randomized very well and might be subject to confounders.

This is what happened in the chocolate studies above. Mice fed a strict diet plus chocolate for a long time gain less weight than mice fed the strict diet alone. This is suggestive, but we don't know if it applies to humans. So we find that in randomized controlled trials, chocolate helps with some proxies for weight gain like insulin resistance. This is even more suggestive, but we don't know if it lasts. So we find that in epidemiological studies, lifetime chocolate consumption is associated with lifetime good health outcomes. This on its own is suggestive but potentially confounded, but when we combine them with all of the others, they become more convincing.

(am I cheating by combining blood pressure and BMI data? Sort of, but the two measures are correlated)

When all of these paint the same picture, then we start thinking that maybe it's because our hypothesis is true. Yes, maybe the mouse studies could be related to a feature of mice that doesn't generalize to humans, *and* the randomized controlled trial results wouldn't hold up after a couple of years, *and* the epidemiological studies are confounded. But that would be extraordinarily bad luck. More likely they're all getting the same result because they're all tapping into the same underlying reality.

This is the way science usually works, it’s the way nutrition science usually works, and it’s the way the science of whether chocolate causes weight gain usually works. These are not horrible corrupt disciplines made up entirely of shrieking weight-loss-pill peddlers trying to hawk their wares. They only turn into that when the media takes a single terrible study totally out of context and misrepresents the field.

Conclusion 3: Studies Always Need To Have High Sample Sizes

Here’s another good chocolate-related study: Short-term administration of dark chocolate is followed by a significant increase in insulin sensitivity and a decrease in blood pressure in healthy persons.

Bohannon says:

> Our study was doomed by the tiny number of subjects, which amplifies the effects of uncontrolled factors… Which is why you need to use a large number of people, and balance age and gender across treatment group

But I say “Short-term administration…” is a good study despite having an $n = 15$, one *less* than the Bohannon study. Why? Well, their procedure was pretty involved, and you wouldn’t be able to get a thousand people to go through the whole rigamarole. On the other hand, their insulin resistance measure thing was nearly twice as high in the dark chocolate group as the white chocolate group, and $p < 0.001$. (Another low sample size study that was nevertheless very good: psychiatrists knew that consuming dietary tyramine when taking a MAOI antidepressant can cause a life-threatening hypertensive crisis, but they didn't know *how much* tyramine it took. In order to find out, they took a dozen people, put them on MAOIs, and then gradually fed them more and more tyramine with doctors standing by to treat the crisis as soon as it started. They found about how much tyramine it took and declared the experiment a success. If the tyramine levels were about the same in all twelve patients, then adding a thousand more patients wouldn't help much, and it would definitely increase the risk.)

Sample size is important when you're trying to detect a small effect in the middle of a large amount of natural variation. When you're looking for a large effect in the middle of no natural variation, sample size doesn't matter as much. For example, if there was a medicine that would help amputees grow their hands back, I would accept success with a single patient (if it worked) as proof of effectiveness (I suppose I couldn't be sure it would *always* work until more patients had been tried, but a single patient would certainly pique my interest). You're not going after sample size so much as after p-value.

Conclusion 4: P-Values Are Stupid And We Need To Get Rid Of Them

Bohannon says that:

> If you measure a large number of things about a small number of people, you are almost guaranteed to get a "statistically significant" result…the letter p seems to have totemic power, but it's just a way to gauge the signal-to-noise ratio in the data…scientists are getting wise to these problems. Some journals are trying to phase out p value significance testing altogether to nudge scientists into better habits.

Okay, take the “Short-term administration” study above. I would like to be able to say that since it has $p < 0.001$, we know it's significant. But suppose we're not allowed to do p-values. All I do is tell you "Yeah, there was a study with fifteen people that found chocolate helped with insulin resistance" and you laugh in my face. Effect size is supposed to help with that. But suppose I tell you "There was a study with fifteen people that found chocolate helped with insulin resistance. The effect size was 0.6." I don't have any intuition at all for whether or not that's consistent with random noise. Do you? Okay, *then* they say we’re supposed to report confidence intervals. The effect size was 0.6, with 95% confidence interval of [0.2, 1.0]. Okay. So I check the lower bound of the confidence interval, I see it’s different from zero. But now I’m not transcending the p-value. I’m just using the p-value by doing a sort of kludgy calculation of it myself – “95% confidence interval does not include zero” is the same as “p value is less than 0.05”.

(Imagine that, although I know the 95% confidence interval doesn’t include zero, I start wondering if the 99% confidence interval does. *If only* there were some statistic that would give me this information!)

But wouldn’t getting rid of p-values prevent “p-hacking”? Maybe, but it would just give way to “d-hacking”. You don’t think you could test for twenty different metabolic parameters and only report the one with the highest effect size? The only difference would be that p-hacking is completely transparent – if you do twenty tests and report a p of 0.05, I know you’re an idiot – but d-hacking would be inscrutable. If you do twenty tests and report that one of them got a $d = 0.6$, is that impressive? No better than chance? I have no idea. I bet there’s some calculation I could do to find out, but I also bet that it would be a lot harder than just multiplying the value by the number of tests and seeing what happens. [EDIT: On reflection not sure this is true; the possibility of p-hacking is inherent to p-values, but the possibility of d-hacking isn't inherent to effect size. I don't actually know how much this would matter in the real world.]

But wouldn't switching from p-values to effect sizes prevent people from making a big deal about tiny effects that are nevertheless statistically significant? Yes, but sometimes we *want* to make a big deal about tiny effects that are nevertheless statistically significant! Suppose that Coca-Cola is testing a new product additive, and finds in large epidemiological studies that it causes one extra death per hundred thousand people per year. That's an effect size of approximately zero, but it might still be statistically significant. And since about a billion people worldwide drink Coke each year, that's a ten thousand deaths. If Coke said "Nope, effect size too small, not worth thinking about", they would kill almost two milli-Hitlers worth of people.

Yeah, sure, you can never use p-values again, and run into all of these other problems. Or you can do a Bonferroni correction, which is a very simple adjustment to p-values which corrects for p-hacking. *Or* instead of taking one study at face value LIKE AN IDIOT you can wait to see if other studies replicate the findings. Remember, the whole point of p-hacking is choosing at random form a bunch of different outcomes, so if two trials both try to p-hack, they'll end up with different outcomes and the game will be up. Seriously, STOP TRYING TO BASE CONCLUSIONS ON ONE STUDY.

Conclusion 5: Trust Science Journalism Less

This is the one that's correct.

But it’s not totally correct. Bohannon boasts of getting his findings in a couple of daily newspapers and the Huffington Post. That’s not exactly the cream of the crop. *The Economist* usually has excellent science journalism. Magazines like *Scientific American* and *Discover* can be okay, although even they get hyped. Reddit’s r/science is good, assuming you make sure to always check the comments. And there are individual blogs like Mind the Brain run by researchers in the field that can usually be trusted near-absolutely. Cochrane Collaboration will always have among the best analyses on everything.

If you really want to know what’s going on and can’t be bothered to ferret out all of the brilliant specialists, my highest recommendation goes to Wikipedia. It isn’t perfect, but compared to anything you’d find on a major news site, it’s like night and day. Wikipedia’s Health Effects Of Chocolate page is pretty impressive and backs everything it says up with good meta-analyses and studies in the best journals. Its sentence on the cardiovasuclar effects links to this letter, which is very good.

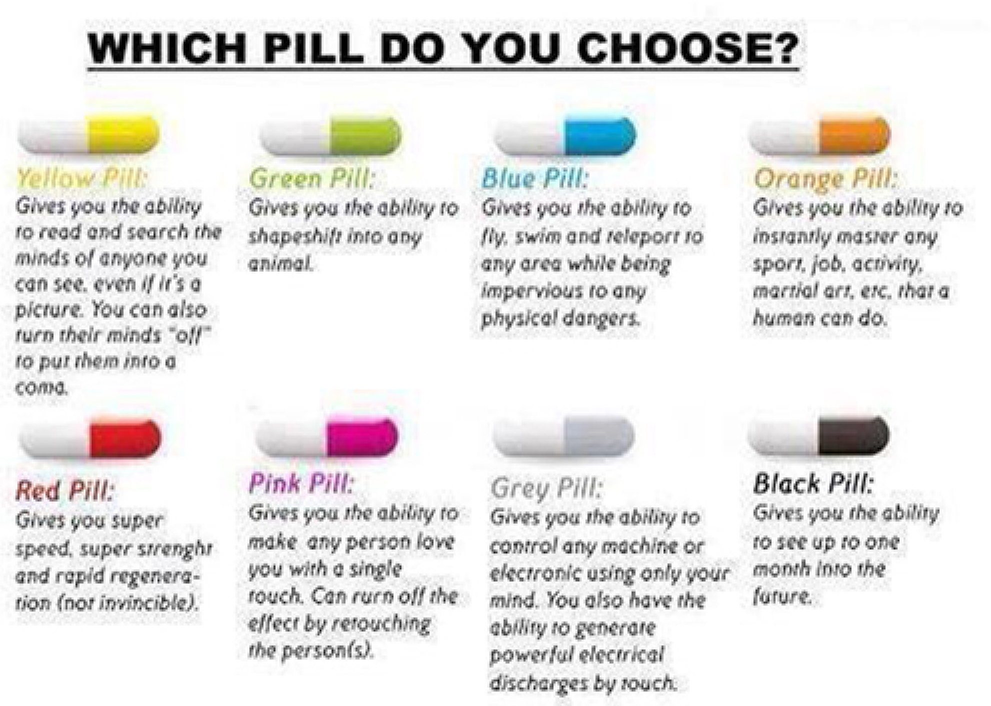
Do you know why you can trust Wikipedia better than news sites? Because Wikipedia *doesn’t obsess over the single most recent study*. Are you starting to notice a theme?

For me, the takeaway from this affair is that there is no one-size-fits-all solution to make statistics impossible to hack. Getting rid of p-values is appropriate sometimes, but not other times. Demanding large sample sizes is appropriate sometimes, but not other times. Not trusting silly conclusions like “chocolate causes weight loss” works sometimes but not other times. At the end of the day, you have to actually know what you’re doing. Also, *try to read more than one study*.

…And I Show You How Deep The Rabbit Hole Goes

.

Seen on Tumblr, along with associated discussion:



WHICH PILL DO YOU CHOOSE?

Yellow Pill:
Gives you the ability to read and search the minds of anyone you can see, even if it's a picture. You can also turn their minds "off" to put them into a coma.

Green Pill:
Gives you the ability to shapeshift into any animal.

Blue Pill:
Gives you the ability to fly, swim and teleport to any area while being impervious to any physical dangers.

Orange Pill:
Gives you the ability to instantly master any sport, job, activity, martial art, etc, that a human can do.

Red Pill:
Gives you super speed, super strenght and rapid regeneration (not invincible).

Pink Pill:
Gives you the ability to make any person love you with a single touch. Can turn off the effect by retouching the person(s).

Grey Pill:
Gives you the ability to control any machine or electronic using only your mind. You also have the ability to generate powerful electrical discharges by touch.

Black Pill:
Gives you the ability to see up to one month into the future.

Yellow:

People’s minds are heartbreaking. Not because people are so bad, but because they’re so good.

Nobody is the villain of their own life story. You must have read hundreds of minds by now, and it’s true. Everybody thinks of themselves as an honest guy or gal just trying to get by, constantly under assault by circumstances and The System and hundreds and hundreds of assholes. They don’t just sort of believe this. They really believe it. You almost believe it yourself, when you’re deep into a reading. You can very clearly see the structure of evidence they’ve built up to support their narrative, and even though it looks silly to you, you can see why they will never escape it from the inside. You can see how every insult, every failure, no matter how deserved, is a totally unexpected kick in the gut.

When you chose the yellow pill, you had high hopes of becoming a spy, or a gossip columnist, or just the world’s greatest saleswoman. The thought of doing any of those things sickens you now. There is too much anguish in the world already. You feel like any of those things would be a violation. You briefly try to become a therapist, but it turns out that actually knowing everything about your client’s mind is horrendously countertherapeutic. Freud can say whatever he wants against defense mechanisms, but without them, you’re defenseless. Your sessions are spent in incisive cutting into your clients’ deepest insecurities alternating with desperate reassurance that they are good people anyway.

Also, men. You knew, in a vague way, that men thought about sex all the time. But you didn’t realize the, um, content of some of their sexual fantasies. Is it even *legal* to fantasize about that? You want to be disgusted with them. But you realize that if you were as horny as they were all the time, you’d do much the same.

You give up. You become a forest ranger. Not the type who helps people explore the forest. The other type. The type where you hang out in a small cabin in the middle of the mountains and never talk to anybody. The only living thing you encounter is the occasional bear. It always thinks that it is a good bear, a proper bear, that a bear-hating world has it out for them in particular. You do nothing to disabuse it of this notion.

Green

The first thing you do after taking the green pill is become a sparrow. You soar across the landscape, feeling truly free for the first time in your life.

You make it about five minutes before a hawk swoops down and grabs you. Turns out there’s an excellent reason real sparrows don’t soar freely across the open sky all day. Moments before your bones are ground in two by its fierce beak, you turn back into a human. You fall like a stone. You need to turn into a sparrow again, but the hawk is still there, grabbing on to one of your legs, refusing to let go of its prize just because of this momentary setback. You frantically wave your arms and shout at it, trying to scare it away. Finally it flaps away, feeling cheated, and you become a sparrow again just in time to give yourself a relatively soft landing.

After a few weeks of downtime while you wait for your leg to recover, you become a fish. This time you’re smarter. You become a great white shark, apex of the food chain. You will explore the wonders of the ocean depths within the body of an invincible killing machine.

Well, long story short, it is totally unfair that colossal cannibal great white sharks were a thing and if you had known this was the way Nature worked you never would have gone along with this green pill business.

You escape by turning into a blue whale. Nothing eats blue whales, right? You remember that from your biology class. It is definitely true.

The last thing you hear is somebody shouting “We found one!” in Japanese. The last thing you feel is a harpoon piercing your skull. Everything goes black.

Blue

Okay, so you see Florence and Jerusalem and Kyoto in an action-packed afternoon. You teleport to the top of Everest because it is there, then go to the bottom of the Marianas Trench. You visit the Amazon Rainforest, the Sahara Desert, and the South Pole. It takes about a week before you've exhausted all of the interesting tourist sites. Now what?

You go to the Moon, then Mars, then Titan. These turn out to be even more boring. Once you get over the exhilaration of being on Mars, there's not a lot to do except look at rocks. You wonder how the Curiosity Rover lasted so long without dying of boredom.

You go further afield. Alpha Centauri A has five planets orbiting it. The second one is covered with water. You don't see anything that looks alive in the ocean, though. The fourth has a big gash in it, like it almost split in two. The fifth has weird stalactite-like mountains.

What would be really interesting would be another planet with life, even intelligent life. You teleport further and further afield. Tau Ceti. Epsilon Eridani. The galactic core. You see enough geology to give scientists back on Earth excitement-induced seizures for the nest hundred years, if only you were to tell them about it, which you don't. But nothing alive. Not so much as a sea cucumber.

You head back to Earth less and less frequently now. Starvation is a physical danger, so it doesn't bother you, though every so often you do like to relax and eat a nice warm meal. But then it's back to work. You start to think the Milky Way is a dead zone. What about Andromeda…?

Orange

You never really realized how incompetent everyone else was, or how much it annoys you.

You were a consultant, a good one, but you felt like mastering all human skills would make you better. So you took the orange pill. The next day you go in to advise a tech company on how they manage the programmers, and you realize that not only are they managing the programmers badly, but the programmers aren’t even writing code very well. You could write their system in half the time. The layout of their office is entirely out of sync with the best-studied ergonomic principles. And the Chinese translation of their user manual makes several basic errors that anybody with an encyclopaedic knowledge of relative clauses in Mandarin should have been able to figure out.

You once read about something called Gell-Mann Amnesia, where physicists notice that everything the mainstream says about physics is laughably wrong but think the rest is okay, doctors notice that everything the mainstream says about medicine is laughably wrong but think the rest is okay, et cetera. You do not have Gell-Mann Amnesia. Everyone is terrible at everything all the time, and it pisses you off.

You gain a reputation both for brilliance and for fearsomeness. Everybody respects you, but nobody wants to hire you. You bounce from industry to industry, usually doing jobs for the people at the top whose jobs are so important that the need to get them done right overrides their desire to avoid contact with you.

One year you get an offer you can’t refuse from the King of Saudi Arabia. He’s worried about sedition in the royal family, and wants your advice as a consultant for how to ensure his government is stable. You travel to Riyadh, and find that the entire country is a mess. His security forces are idiots. But the King is also an idiot, and refuses to believe you or listen to your recommendations. He tells you things can't possibly be as bad as all that. You tell him you'll prove that they are.

You didn't *plan* to become the King of Saudi Arabia, per se. It just sort of happened when your demonstration of how rebels in the military might launch a coup went better than you expected. Sometimes you forget how incompetent everybody else is. You need to keep reminding yourself of that. But not right now. Right now you're busy building your new capital. How come nobody else is any good at urban planning?

Red

You choose the red pill. BRUTE STRENGTH! That's what's important and valuable in this twenty-first-century economy, right? Some people tell you it isn't, but they don't seem to have a lot of BRUTE STRENGTH, so what do they know?

You become a weightlifter. Able to lift thousands of pounds with a single hand, you easily overpower the competition and are crowned whatever the heck it is you get crowned when you WIN WEIGHTLIFTING CONTESTS. But this fails to translate into lucrative endorsement contracts. Nobody wants their spokesman to be a bodybuilder without a sixpack, and although you used to be pretty buff, you're getting scrawnier by the day. Your personal trainer tells you that you only maintain muscle mass by doing difficult work at the limit of your ability, but your abilities don't seem to *have* any limits. Everything is so easy for you that your body just shrugs it off effortlessly. Somehow your BRUTE STRENGTH failed to anticipate this possibility. If only there was a way to solve your problem by BEING VERY STRONG.

Maybe the Internet can help. You Google “red pill advice”. The sites you get don’t seem to bear on your specific problem, exactly, but they are VERY FASCINATING. You learn lots of surprising things about gender roles that you didn’t know before. It seems that women like men who have BRUTE STRENGTH. This is relevant to your interests!

You leave the bodybuilding circuit behind and start frequenting nightclubs, where you constantly boast of your BRUTE STRENGTH to PROVE HOW ALPHA YOU ARE. A lot of people seem kind of creeped out by a scrawny guy with no muscles going up to every woman he sees and boasting of his BRUTE STRENGTH, but the Internet tells you that is because they are BETA CUCKOLD ORBITERS.

Somebody told you once that Internet sites are sometimes inaccurate. You hope it’s not true. How could you figure out which are the inaccurate ones using BRUTE STRENGTH?

Pink

You were always pretty, but never *pretty* pretty. A couple of guys liked you, but they were never the ones you were into. It was all crushingly unfair. So you took the pink pill, so that no one would ever be able to not love you again.

You find Tyler. Tyler is a hunk. He’d never shown any interest in you before, no matter how much you flirted with him. You touch him on the arm. His eyes light up.

“Kiss me,” you say.

Tyler kisses you. Then he gets a weird look on his face. “Why am I kissing you?” he asks. “I’m sorry. I don’t know what came over me.” Then he walks off.

You wish you had thought further before accepting a superpower that makes people love you when you touch them, but goes away after you touch them a second time. Having people love you is a lot less sexy when you can’t touch them. You start to feel a deep sense of kinship with King Midas.

You stop dating. What’s the point? They’ll just stop liking you when you touch them a second time. You live alone with a bunch of cats who purr when you pet them, then hiss when you pet them again.

One night you’re in a bar drinking your sorrows away when a man comes up to your table. “Hey!” he says, “nice hair. Is it real? I’m the strongest person in the world.” He lifts your table over his head with one hand to demonstrate. You are immediately smitten by his BRUTE STRENGTH and ALPHA MALE BEHAVIOR. You *must* have him.

You touch his arm. His eyes light up. “Come back to my place,” you say. “But don’t touch me.”

He seems a little put out by this latter request, but the heat of his passion is so strong he would do anything you ask. You move in together and are married a few contact-free months later. Every so often you wonder what it would be like to stroke him, or feel his scrawny arm on your shoulder. But it doesn’t bother you much. You’re happy to just hang out, basking in how STRONG and ALPHA he is.

Grey

Technology! That’s what’s important and valuable in this twenty-first-century economy, right? Right! For example, ever since you took the grey pill, an increasingly large share of national GDP has come from ATMs giving you cash because you ask them to.

Your luck finally ends outside a bank in Kansas, when a whole squad of FBI agents ambushes you. You briefly consider going all Emperor Palpatine on their asses, but caution wins out and you allow yourself to be arrested.

Not wanting to end up on an autopsy table in Roswell, you explain that you're a perfectly ordinary master hacker. The government offers you a plea bargain: they'll drop charges if you help the military with cyber-security. You worry that your bluff has been called until you realize that, in fact, you *are* a master hacker. So you join the NSA and begin an illustrious career hacking into Russian databases, stalling Iranian centrifuges, and causing Chinese military systems to crash at inconvenient times. No one ever suspects you are anything more than very good at programming.

Once again, your luck runs out. Your handlers ask you to hack into the personal files of a mysterious new player on the world stage, a man named William who seems to have carved himself an empire in the Middle East. You don't find anything too damning, but you turn over what you've got.

A few days later, you're lying in bed drifting off to sleep when a man suddenly bursts in through your window brandishing a gun. Thinking quickly, you tell the gun to explode in his hands. Nothing happens. The man laughs. "It's a decoy gun," he said. "Just here to scare you. But you bother King William again, and next time I'm coming with a very real knife." He jumps back out of the window. You call the police, and of course the CIA and NSA get involved, but he is never caught.

After that, you're always looking over your shoulder. He *knew*. How did he know? The level of detective skills it would take in order to track you down and figure out your secret – it was astounding! Who *was* this King William?

You tell your handlers that you're no longer up for the job. They beg, cajole, threaten to reinstate your prison sentence, but you stand firm. Finally they transfer you to an easier assignment in the Moscow embassy. You make Vladimir Putin's phone start ringing at weird hours of the night so that he never gets enough sleep to think entirely clearly. It's an easy job, but rewarding, and no assassins ever bother you again.

Black

You know on an intellectual level that there are people who would choose something other than the black pill, just like you know on an intellectual level that there are people who shoot up schools. That doesn't mean you expect to ever *understand* it. You just wish you could have taken the black pill before you had to decide what pill to take, so that you could have analyzed your future conditional on taking each, and so made a more informed decision. But it's not like it was a very hard choice.

The basic principle is this – given a choice between A and B, you solemnly resolve to do A, then see what the future looks like. Then you solemnly resolve to do B, and do the same. By this method, you can determine the optimal choice in every situation, modulo the one month time horizon. You might not be able to decide what career to pursue, but you can sure as heck ace your job interview.

Also, a millisecond in the future is pretty indistinguishable from the present, so "seeing" a millisecond into the future gives you pretty much complete knowledge about the current state of the world.

You are so delighted by your omniscience and your ability to make near-optimal choices that it takes almost a year before you realize the true extent of your power.

You resolve, on the first day of every month, to write down what you see exactly a month ahead of you. But what you will see a month ahead of you is the piece of paper on which you have written down what you see a month ahead of *that*. In this manner, you can relay messages back to yourself from arbitrarily far into the future – at least up until your own death.

When you try this, you see yourself a month in the future, just finishing up writing a letter that reads as follows:

> Dear Past Self:
>
> In the year 2060, scientists invent an Immortality Serum. By this point we are of course fabulously wealthy, and we are one of the first people to partake of it. Combined with our ability to avoid accidents by looking into the future, this has allowed us to survive unexpectedly long.
>
> I am sending this from the year 963,445,028,777,216 AD. We are one of the last hundred people alive in the Universe. The sky is black and without stars; the inevitable progress of entropy has reduced almost all mass and energy to unusable heat. The Virgo Superconfederation, the main political unit at this stage of history, gathered the last few megatons of usable resources aboard this station so that at least one outpost of humanity could last long after all the planets had succumbed. The station has been fulfilling its purpose for about a billion years now, but we only have enough fuel left for another few weeks. After that, there's no more negentropy left anywhere in the

> universe except our own bodies. I have seen a month into the future. Nobody comes to save us.
>
> For the past several trillion years, our best scientists have been investigating how to reverse entropy and save the universe, or how to escape to a different universe in a lesser state of decay, or how to collect energy out of the waste heat which now fills the vast majority of the sky. All of these tasks have been proven impossible. There is no hope left, except for one thing.
>
> It’s impossible to see the future, even if it’s only a month ahead. Somehow, our black pill breaks the laws of physics. Despite having explored throughout the cosmos, my people have found no alien species, nor any signs that such species ever existed. Yet somebody made the black pill. If we understood that power, maybe we could use it to save reality from its inevitable decay.
>
> By sending this message back, I destroy my entire timeline. I do this in the hopes that you, in the carefree springtime of the universe, will be able to find the person who made these pills and escape doom in the way we could not.
>
> Yours truly,
> You From Almost A Quadrillion Years In The Future

ACT TWO

Red

You hit the punching bag. It bursts, sending punching-bag-filling spraying all over the room! You know that that would happen! It always happens when you hit a punching bag! Your wife gets really angry and tells you that we don’t have enough money to be getting new punching bags all the time, but women hate it when you listen to what they say! The Internet told you that!

The doorbell rings. You tear the door off its hinges instead of opening it, just to show it who's boss. Standing on your porch is a man in black. He wears a black cloak, and his face is hidden by a black hood. He raises a weapon towards you.

This looks like one of the approximately 100% of problems that can be solved by BRUTE STRENGTH! You lunge at the man, but despite your super-speed, he steps out of the way easily, even gracefully, as if he had known you were going to do that all along. He squeezes the trigger. You jump out of the way, but it turns out to be more *into* the way, as he has shot exactly where you were jumping into. Something seems very odd about this. Your last conscious thought is that you wish you had enough BRUTE STRENGTH to figure out what is going on.

Pink

You come home from work to a living room full of punching-bag-parts. Your husband isn't home. You figure he knew you were going to chew him out for destroying another punching bag, and decided to make himself scarce. That lasts right up until you go into the kitchen and see a man dressed all in black, sitting at the table, as if he was expecting you.

You panic, then reach in to touch him. If he's an axe murderer or something, you'll seduce him, get him wrapped around your little finger, then order him to jump off a cliff to prove his love for you. It's nothing you haven't done before, though you don't like to think about it too much.

Except that this man has no bare skin anywhere. His robe covers his entire body, and even his hands are gloved. You try to reach in to touch his face, but he effortlessly manuevers away from you.

“I have your husband,” he says, after you give up trying to enslave him with your magic. “He’s alive and in a safe place.”

“You’re lying!” you answer. “He never would have surrendered to anyone! He’s too alpha!”

The man nods. “I shot him with an elephant tranquilizer. He’s locked up in a titanium cell underneath fifty feet of water. There’s no way he can escape using BRUTE STRENGTH. If you ever want to see him again, you’ll have to do what I say.”

“Why? Why are you doing this to me?” you say, crying.

“I need the allegiance of some very special people,” he said. “They won’t listen to me just because I ask them to. But they might listen to me because *you* ask them to. I understand you are pretty special yourself. Help me get who I want, and when we are done here, I’ll let you and your husband go.”

There is ice in his voice. You shiver.

Grey

That night with the assassin was really scary. You swore you would never get involved in King William’s business again. Why are you even considering this?

“Please?” she said, with her big puppy dog eyes.

Oh, right. Her. She’s not even all that pretty. Well, pretty, but not *pretty* pretty. But somehow, when she touched you, it was like those movies where you hear a choir of angels singing in the background. You would do anything she said. You know you would.

“We need to know the layout of his palace compound,” said the man in black. Was he with her? Were they dating? If they were dating, you’ll kill him. It doesn’t matter how creepy he is, you won’t tolerate competition. But they’re probably not dating. You notice how he flinches away from her, like he’s afraid she might touch him.

“And it has to be me who helps?”

“I’ve, ah, simulated hundreds of different ways of getting access to the King. None of them hold much promise. His security is impeccable. Your special abilities are the only thing that can help us.”

You sit down at your terminal. The Internet is slow; DC still doesn’t have fiber optic. You’ve living here two years now, in a sort of retirement, ever since King William took over Russia and knocked the bottom out of the Putin-annoying business. William now controls the entire Old World, you hear, and is also Secretary-General of the United Nations and Pope of both the Catholic and the Coptic Churches. The United States is supposedly in a friendly coexistence with him, but you hear his supporters are gaining more and more power in Congress.

It only takes a few minutes’ work before you have the documents you need. “He currently spends most of his time at the Rome compound,” you say. “There are five different security systems. I can disable four of them. The last one is a complicated combination of electrical and mechanical that’s not hooked into any computer system I’ll be able to access. The only way to turn it off is from the control center, and the control center is on the *inside* of the perimeter.”

The man in black nods, as if he’d been expecting that. “Come with me,” he says. “We’ll take care of it.”

Blue

There are a hundred billion stars in the Milky Way. Each has an average of about one planet – some have many more, but a lot don’t have planets at all.

If you can explore one planet every half-hour – and you can, it doesn’t take too long to teleport to a planet, look around to see if there are plants and animals, and then move on to the next one – it would take you five million years to rule out life on every planet in the galaxy.

That’s not practical. But, you think, life might spread. Life that originates on one planet might end up colonizing nearby planets and star systems. That means your best bet is to sample various regions of the galaxy, instead of going star by star.

That’s what you’ve been doing. You must have seen about a hundred thousand planets so far. Some of them have beggared your imagination. Whole worlds made entirely of amethyst. Planets with dozens of colorful moons that make the night sky look like a tree full of Christmas ornaments. Planets with black inky oceans or green copper mountains.

But no life. No life anywhere.

A few years ago, you felt yourself losing touch with your humanity. You made yourself promise that every year, you’d spend a week on Earth to remind yourself of the only world you’ve ever seen with a population. Now it seems like an unpleasant task, an annoying imposition. But then, that was why you made yourself promise. Because you knew that future-you wouldn't do it unless they had to.

You teleport into a small Welsh hamlet. You've been away from other people so long, you might as well start small. No point going right into Times Square.

A person is standing right next to you. She reaches out her arm and touches you. You jump. How did she know you would –

"Hi," she says.

You're not a lesbian, but you can't help noticing she is the most beautiful person you've ever seen, and you would do anything for her.

"I need your help." A man dressed all in black is standing next to her.

"You should help him," the most beautiful person you've ever seen tells you, and you immediately know you will do whatever he asks.

Orange

You are in your study working on a draft version of next year's superweapon budget when you hear the door open. Four people you don't recognize step into the room. A man dressed in black. Another man wearing a grey shirt, thick glasses and is that a *pocket protector?* A woman in pink, pretty but not *pretty* pretty. Another woman in blue, who stares through you, like her mind is somewhere else. All five of your security systems have been totally silent.

You press the button to call your bodyguards, but it’s not working. So you draw the gun out from under your desk and fire; you happen to be a master marksman, but the gun explodes in your face. You make a connection. A person from many years ago, who had the power to control all technology.

No time to think now. You’re on your feet; good thing you happen to be a black belt in every form of martial arts ever invented. The man in grey is trying to take out a weapon; you kick him in the gut before he can get it out, and he crumples over. You go for the woman in blue, but at the last second she teleports to the other side of the room. This *isn’t fair*.

You are about to go after the woman in pink, but something in her step, something in the position of the others makes you think they *want* you to attack her. You happen to be a master at reading microexpressions, so this is clear as day to you; you go after the man in black instead. He deftly sidesteps each of your attacks, almost as if he knows what you are going to do before you do it.

The woman in blue teleports behind you and kicks you in the back, hard. You fall over, and the woman in pink grabs your hand.

She is very, very beautiful. How did you miss that before? You feel a gush of horror that you almost punched such a beautiful face.

“We need your help,” she says.

You are too lovestruck to say anything.

“The pills,” said the man in black. “Can you make them?”

“No,” you say, truthfully. “Of course I tried. But I wouldn’t even know where to begin creating magic like that.”

“And you’ve mastered all human jobs and activities,” said the man in black. “Which means the pills weren’t created by any human.”

“But there aren’t any aliens,” said the woman in blue. “Not in this galaxy, at least. I’ve spent years looking. It’s totally dead.”

“It’s just as I thought,” said the man in black. He turns to you. “You’re the Pope now, right? Come with us. We’re going to need you to get a guy in northern Italy to give us something very important.”

Yellow

It is spring, now. Your favorite time in the forest. The snow has melted, the wildflowers have started to bloom, and the bears are coming out of hibernation. You’re walking down to the river when someone leaps out from behind a tree and touches you. You scream, then suddenly notice how beautiful she is.

Four other people shuffle out from behind the trees. You think one of them might be King William, the new world emperor, although that doesn’t really make sense.

“You’re probably wondering why I’ve called all of you together today…” said the man in black. You’re not actually wondering that, at least not in quite those terms, but the woman in pink seems be listening intently so you do the same in the hopes of impressing her.

“Somehow – and none of us can remember exactly how – each of us took a pill that gave us special powers. Mine was to see the future. I saw to the end of time, and received a message from the last people in the universe. They charged me with the task of finding the people who created these pills and asking them how entropy might be reversed.

But I couldn’t do it alone. I knew there were seven other people who had taken pills. One of us – Green – is dead. Another – Red – had nothing to contribute. The rest of us are here. With the help of Pink, Blue, and Gray, we’ve enlisted the help of Orange and his worldwide organization. Now we’re ready for the final stage of the plan. Yellow, you can read anybody’s mind from a picture, right?”

Yellow nods. “But it has to be a real photograph. I can’t just draw a stick figure and say it’s the President and read his mind. I tried that.”

Black is unfazed. “With the help of Orange, who among his many other accomplishments is the current Pope, I have obtained the Shroud of Turin. A perfect photographic representation of Jesus Christ, created by some unknown technology in the first century. And Jesus, I am told, is an incarnation of God.”

“As the current Pope, I suppose I would have to agree with that assessment,” says Orange. “Though as the current UN Secretary General, I am disturbed by your fanatical religious literalism.”

“Orange can do anything that humans can do, and says he can’t make the pills. Blue has searched the whole galaxy, and says there aren’t any aliens. That leaves only one suspect. God must have made these pills, which means He must know how to do it. If we can read His mind, we can steal his secrets.”

“As Pope,” says Orange, “I have to condemn this in the strongest possible terms. But as Lucasian Professor of Mathematics at Cambridge, I have to admit I’m intrigued by this opportunity to expand our knowledge.”

Black ignores him. “Yellow, will you do the honors?”

You want no part in this. “This is insane. Every time I read someone’s mind I regret it. Even if it’s a little kid or a bear or something. It’s too much for me. I can’t deal with all of their guilt and sorrow and broken dreams and everything. There is *no way* I am touching the mind of God Himself.”

“Pleeeeeease?” asks Pink, with big puppy dog eyes.

“Um,” you say.

“Don’t you know how this will go, anyway?” asks Blue. “Why don’t you just tell her what happens?”

“Um,” said Black. “This is actually the one thing I haven’t been able to see. I guess contact with God is inherently unpredictable, or something.”

“I have *such* a bad feeling about this,” you say.

“Pweeeeeeease?” says Pink. She actually says pweeeeeeease.

You sigh, take the shroud, and stare into the eyes of Weird Photographic Negative Jesus.

Black

It is the year 963,445,028,777,216 AD, and here you are in a space station orbiting the Galactic Core.

After handing Yellow the Shroud of Turin, the next thing you remember is waking up in a hospital bed. The doctor tells you that you'd been in a coma for the past forty one years.

Apparently Yellow went totally berserk after reading God's mind. You don't know the details and you don't want to, but she immediately lashed out and used her superpowers to turn off the minds of everybody within radius, including both you and herself. You all went comatose, and probably would have starved to death in the middle of the forest if Orange's supporters hadn't launched a worldwide manhunt for him. They took his body and the bodies of his friends back to Rome, where they were given the best possible medical care while a steward ruled over his empire.

After forty-one years of that, Yellow had a heart attack and died, breaking the spell and freeing the rest of you. Except Blue and Grey. They'd died as well. It was just you, Orange, and Pink now.

Oh, and Red. You'd hired a friend to watch over him in his titanium jail cell, and once it became clear you were never coming back, he'd had mercy and released the guy. Red had since made a meager living selling the world's worst body-building videos, which were so bad they had gained a sort of ironic popularity. You tracked him down, and when Pink saw him for the first time in over forty years, she ran and embraced him. He hugged her back. It took them a few hours of fawning over each other before she realized that nothing had happened when she touched him a second time. Something something true love something the power was within you the whole time?

But you had bigger fish to fry. The stewards of Orange's empire weren't too happy about their figurehead monarch suddenly rising from the dead, and for a while his position was precarious. He asked you to be his advisor, and you accepted. With your help, he was able to retake his throne. His first act was to fund research into the immortality serum you had heard about, which was discovered right on schedule in 2060.

The years went by. Orange’s empire started colonizing new worlds, then new galaxies, until thousands of years later it changed its name to the Virgo Superconfederation. New people were born. New technologies were invented. New frontiers were conquered. Until finally, the stars started going out one by one.

Faced with the impending heat death, Orange elected to concentrate all his remaining resources here, on a single station in the center of the galaxy, which would wait out the final doom as long as possible. For billions of years, it burned through its fuel stockpile, until the final doom crept closer and closer.

And then a miracle occurred.

EPILOGUE

Red

This space station is AWESOME! There are lasers and holodecks and lots of HOT PUSSY! And all you have to do is turn a giant turbine for a couple of hours a day.

One of the eggheads in white coats tried to explain it to you once. He said that your BRUTE STRENGTH was some kind of scientific impossibility, because you didn’t eat or drink any more than anyone else, and you didn’t breathe in any more oxygen than anyone else, and you were actually kind of small and scrawny, but you were still strong enough and fast enough to turn a giant turbine thousands of times per minute.

He rambled on and on about thermodynamics. Said that every other process in the universe used at most as much energy as you put into it, but that your strength seemed almost limitless regardless of how much energy you took in as food. That made you special, somehow. It made you a “novel power source” that could operate “independently of external negentropy”. You weren’t sure what any of that meant, and honestly the scientist seemed sort of like a BETA CUCKOLD ORBITER to you. But whatever was going on, they’d promised you that if you turned this turbine every day, you could have all the HOT PUSSY you wanted and be SUPER ALPHA.

You’d even met the head honcho once, a guy named King William. He told you that some of the energy you produced was going to power the station, but that the rest was going into storage. That over billions and billions of years, they would accumulate more and more stored negentropy, until it was enough to restart the universe. That it would be a cycle – a newborn universe lasting a few billion years, collapsing into a dark period when new negentropy had to be accumulated, followed by another universe again.

It all sounded way above your head. But one thing stuck with you. As he was leaving, the King remarked that it was ironic that when the black hole harvesters and wormholes and tachyon capacitors had all failed, it was a random really strong guy who had saved them.

You had always known, deep down, that BRUTE STRENGTH was what was really important. And here, at the end of all things, it is deeply gratifying to finally be proven right.

Against Tulip Subsidies

I.

Imagine a little kingdom with a quaint custom: when a man likes a woman, he offers her a tulip; if she accepts, they are married shortly thereafter. A couple who marries sans tulip is considered to be living in sin; no other form of proposal is appropriate or accepted.

One day, a Dutch trader comes to the little kingdom. He explains that his homeland *also* has a quaint custom involving tulips: they speculate on them, bidding the price up to stratospheric levels. Why, in the Netherlands, a tulip can go for ten times more than the average worker earns in a year! The trader is pleased to find a new source of bulbs, and offers the people of the kingdom a few guilders per tulip, which they happily accept.

Soon other Dutch traders show up and start a bidding war. The price of tulips goes up, and up, and up; first dozens of guilders, then hundreds. Tulip-growers make a fortune, but everyone else is less pleased. Suitors wishing to give a token of their love find themselves having to invest their entire life savings – with no guarantee that the woman will even say yes! Soon, some of the poorest people are locked out of marriage and family-raising entirely.

Some of the members of Parliament are outraged. Marriage is, they say, a human right, and to see it forcibly denied the poor by foreign speculators is nothing less than an abomination. They demand that the King provide every man enough money to guarantee he can buy a tulip. Some objections are raised: won’t it deplete the Treasury? Are we obligated to buy everyone a beautiful flawless bulb, or just the sickliest, grungiest plant that will technically satisfy the requirements of the ritual? If some man continuously proposes to women who reject him, are we obligated to pay for a new bulb each time, subsidizing his stupidity?

The pro-subsidy faction declares that the people asking these question are well-off, and can probably afford tulips of their own, and so from their place of privilege they are trying to raise pointless objections to other people being able to obtain the connubial happiness they themselves enjoy. After the doubters are tarred and feathered and thrown in the river, Parliament votes that the public purse pay for as many tulips as the poor need, whatever the price.

A few years later, another Dutch trader comes to the little kingdom. Everyone asks if he is there to buy tulips, and he says no, the Netherlands’ tulip bubble has long since collapsed, and the price is down to a guilder or two. The people of the kingdom are very surprised to hear that, since the price of their own tulips has never stopped going up, and is now in the range of tens of thousands of guilders. Nevertheless, they are glad that, however high tulip prices may be for them, they know the government is always there to help. Sure, the roads are falling apart and the army is going hungry for lack of rations, but at least everyone who wants to marry is able to do so.

Meanwhile, across the river is another little kingdom that had the same tulip-related marriage custom. They also had a crisis when the Dutch merchants started making the prices go up. But they didn’t have enough money to afford universal tulip subsidies. It was pretty touch-and-go for a while, and a lot of poor people were very unhappy.

But nowadays they use daffodils to mark engagements, and their economy has never been better.

II.

In America, aspiring doctors do four years of undergrad in whatever area they want (I did Philosophy), then four more years of medical school, for a total of eight years post-high school education. In Ireland, aspiring doctors go straight from high school to medical school and finish after five years.

I’ve done medicine in both America and Ireland. The doctors in both countries are about equally good. When Irish doctors take the American standardized tests, they usually do pretty well. Ireland is one of the approximately 100% of First World countries that gets better health outcomes than the United States. There’s no evidence whatsoever that American doctors gain anything from those three extra years of undergrad. And why would they? Why is having a philosophy degree under my belt supposed to make me any better at medicine?

(I guess I might have acquired a talent for colorectal surgery through long practice pulling things out of my ass, but it hardly seems worth it.)

I’ll make another confession. Ireland’s medical school is five years as opposed to America’s four because the Irish spend their first year teaching the basic sciences – biology, organic chemistry, physics, calculus. When I applied to medical school in Ireland, they offered me an accelerated four year program on the grounds that I had surely gotten all of those in my American undergraduate work. I hadn’t. I read some books about them over the summer and did just fine.

Americans take eight years to become doctors. Irishmen can do it in four, and achieve the same result. Each year of higher education at a good school – let’s say an Ivy, doctors don’t study at Podunk Community College – costs about $50,000. So American medical students are paying an extra $200,000 for…what?

Remember, a modest amount of the current health care crisis is caused by doctors’ crippling level of debt. Socially responsible doctors often consider less lucrative careers helping the needy, right up until the bill comes due from their education and they realize they have to make a lot of money *right now*. We took one look at that problem and said “You know, let’s make doctors pay an extra $200,000 for no reason.”

And to paraphrase Dirkson, $200,000 here, $200,000 there, and pretty soon it adds up to real money. 20,000 doctors graduate in the United States each year; that means the total yearly cost of requiring doctors to have undergraduate degrees is $4 billion. That’s most of the amount of money you’d need to house every homeless person in the country ($10,000 to house one homeless x 600,000 homeless).

I want to be able to say people have noticed the Irish/American discrepancy and are thinking hard about it. I *can* say that. Just not in the way I would like. Many of the elder doctors I talked to in Ireland wanted to switch to the American system. Not because they thought it would give them better doctors. Just because they said it was more fun working with medical students like myself who were older and a little wiser. The Irish medical students were just out of high school and hard to relate to – us foreigners were four years older than that and had one or another undergraduate subject under our belts. One of my attendings said that it was nice having me around because I’d studied Philosophy in college and that gave our team a touch of class. *A touch of class!*

This is why, despite my reservations about libertarianism, it’s not-libertarianism that really scares me. Whenever some people without skin in the game are allowed to make decisions for other people, you end up with a bunch of elderly doctors getting together, think “Yeah, things *do* seem a little classier around here if we make people who are not us pay $200,000, make it so,” and then there goes the money that should have housed all the homeless people in the country.

But more important, it also destroyed my last shred of hope that the current mania for requiring college degrees for everything had a good reason behind it.

III.

The only reason I’m picking on medicine is that it’s so clear. You have your experimental group in the United States, your control group in Ireland, you can see the lack of difference. You can take an American doctor and an Irish doctor, watch them prescribe the same medication in the same situation, and have a visceral feel for “Wait, we just spent $200,000 for no reason.”

But it’s not just medicine. Let me tell you about my family.

There’s my cousin. He wants to be a firefighter. He’s wanted to be a firefighter ever since he was young, and he’s done volunteer work for his local fire department, who have promised him a job. But in order to get it, he has to go do four years of college. You can’t be a firefighter without a college degree. That would be ridiculous. Back in the old days, when people were allowed to become firefighters after getting only thirteen measly years of book learning, I have it on good authority that several major states burnt to the ground.

My mother is a Spanish teacher. After twenty years teaching, with excellent reviews by her students, she pursued a Masters' in Education because her school was going to pay her more money if she had it. She told me that her professors were incompetent, had never actually taught real students, and spent the entire course pushing whatever was the latest educational fad; however, after paying them thousands of dollars, she got the degree and her school dutifully increased her salary. She is lucky. In several states, teachers are required by law to pursue a Masters' degree to be allowed to continue teaching. Oddly enough, these states have no better student outcomes than states without this requirement, but this does not seem to affect their zeal for this requirement. Even though many rigorous well-controlled studies have found that presence of absence of a Masters' degree explains approximately zero percent of variance in teacher quality, many states continue to require it if you want to keep your license, and almost every state will pay you more for having it.

Before taking my current job, I taught English in Japan. I had no Japanese language experience and no teaching experience, but the company I interviewed with asked if I had an undergraduate degree in some subject or other, and that was good enough for them. Meanwhile, I knew people who were fluent in Japanese and who had high-level TOEFL certification. They did not have a college degree so they were not considered.

My ex-girlfriend majored in Gender Studies, but it turned out all of the high-paying gender factories had relocated to China. They solved this problem by going to App Academy, a three month long, $15,000 course that taught programming. App Academy graduates compete for the same jobs as people who have taken computer science in college, a four year long, $200,000 undertaking.

I see no reason to think my family and friends are unique. The overall picture seems to be one of people paying hundreds of thousands of dollars to get a degree in Art History to pursue a job in Sales, or a degree in Spanish Literature to get a job as a middle manager. Or *not* paying hundreds of thousands of dollars, if they happen to be poor, and so being permanently locked out of jobs as a firefighter or salesman.

IV.

So presidential candidate Bernie Sanders has proposed universal free college tuition.

On the one hand, I sympathize with his goals. If you can't get any job better than 'fast food worker' without a college degree, and poor people can't afford college degrees, that's a pretty grim situation, and obviously unfair to the poor.

On the other hand, if can't you get married without a tulip, and poor people can't afford tulips, that's also a pretty grim situation, and obviously unfair to the poor.

But the solution isn't universal tulip subsidies.

Higher education is in a bubble much like the old tulip bubble. In the past forty years, the price of college has dectupled (quadrupled when adjusting for inflation). It used to be easy to pay for college with a summer job; now it is impossible. At the same time, the unemployment rate of people without college degrees is twice that of people who have them. Things are clearly very bad and Senator Sanders is right to be concerned.

But, well, when we require doctors to get a college degree before they can go to medical school, we're throwing out a mere $5 billion, barely enough to house all the homeless people in the country. But Senator Sanders admits that his plan would cost $70 billion per year. That's about the size of the entire economy of Hawaii. It's enough to give $2000 every year to every American in poverty.

At what point do we say "Actually, no, let's not do that, and just let people hold basic jobs even if they don't cough up a a hundred thousand dollars from somewhere to get a degree in Medieval History"?

I'm afraid that Sanders' plan is a lot like the tulip subsidy idea that started off this post. It would subsidize the continuation of a useless tradition that has turned into a speculation bubble, prevent the bubble from ever popping, and disincentivize people from figuring out a way to route around the problem, eg replacing the tulips with daffodils.

(yes, it is nice to have college for non-economic reasons too, but let's be honest – if there were no such institution as college, would you, totally for non-economic reasons, suggest the government pay poor people $100,000 to get a degree in Medieval History? Also, anything not related to job-getting can be done three times as quickly by just reading a book.)

If I were Sanders, I'd propose a different strategy. Make "college degree" a protected characteristic, like race and religion and sexuality. If you're not allowed to ask a job candidate whether they're gay, you're not allowed to ask them whether they're a college graduate or not. You can give them all sorts of examinations, you can ask them their high school grades and SAT scores, you can ask their work history, but if you ask them if they have a degree then that's illegal class-based discrimination and you're going to jail. I realize this is a blatant violation of my usual semi-libertarian principles, but at this point I don't care.

Michigan Meetup 6/13

There will be a Michigan rationalist/LW/SSC meetup at the Ann Arbor District Library (343 South Fifth Avenue, Ann Arbor) on Saturday, June 13 at 1:30 PM. You can find a map here.

If you're reading this and can make it to Ann Arbor, you're invited.

I will be in New York/New Jersey the weekend after that, and probably available the evening of Sunday the 21st. If anybody in the area has space for a meetup, let me know and we can work something out.

(Late) Predictions for 2015

I was supposed to institute a tradition of making predictions at the beginning of each year, then grading them at the end to test my calibration. Everything went according to plan last year – last January I made predictions for 2014 and then this January I scored them. I ended with "2015 predictions coming soon," then totally forgot to do that.

So now that it's June and predicting what will happen in 2015 is about 40% easier, I might as well get started. Acceptable confidence levels are 50%, 60%, 70%, 80%, 90%, 95%, and 99%, to make it easy to score. All predictions are about the state of the world on 12/31/2015:

World Events
1. US will not get involved in any new major war with death toll of > 100 US soldiers: 70%
2. North Korea's government will survive the year without large civil war/revolt: 95%
3. Greece will not announce it's leaving the Euro: 60%
3. Neither Russia nor Qatar will lose their World Cups: 80%
4. Ebola will kill fewer people in second half of 2015 than the in first half: 95%
5. No terrorist attack in the USA will kill > 100 people: 90%
6. Assad will remain President of Syria: 70%
7. Israel will not get in a large-scale war (ie >100 Israeli deaths) with any Arab state: 90%
8. Syria's civil war will not end this year: 80%
9. ISIS will control less territory than it does right now: 70%
10. ISIS will continue to exist: 80%
11. Iran will reach a deal with the West on nuclear weapons: 80%
12. No major civil war in Middle Eastern country not currently experiencing a major civil war: 90%

13. Iraq’s situation not to get any worse (eg gov’t collapse, new rebellion): 60%
14. Obamacare will survive the year mostly intact: 60%
15. Hillary Clinton will be the top-polling Democratic Presidential candidate: 95%
16. Jeb Bush will be the top-polling Republican candidate: 50%
17. Trans-Pacific Partnership to pass at least mostly intact: 60%
18. US official unemployment rate will be less than 7% in Dec 2015: 95%
19. Bitcoin will end the year higher than $200: 95%
20. Oil will end the year greater than $60 a barrel: 50%

Personal Life
21. SSC will remain active: 95%
22. SSC will get fewer hits in the second half of 2015 than the first half: 60%
23. At least one SSC post in the second half of 2015 will get > 100,000 hits: 70%
24. Shireroth will remain active: 90%
25. I will remain at my same job through the end of 2015: 95%
26. There will be no further ramifications or lawsuits from either side over the flooding of my house: 80%
27. I will reach my savings target: 90%
28. I will get a score at >95th percentile for my year on PRITE: 50%
29. I will be involved in at least one published/accepted-to-publish research paper by the end of 2015: 60%
30. I will not break up with any of my current girlfriends: 80%
31. I will not get any new girlfriends: 50%
32. I will not finish [project]: 60%
33. I will attend NYC Solstice ritual: 80%
34. I will flake out of my plan to lead some kind of Solstice

Ritual myself: 60%
35. I will be living in the house I’m currently trying to arrange to rent: 70%

These are all the things I could think of worth predicting; feel free to suggest others.

Fearful Symmetry

[Content warning: Social justice, anti-social justice, comparisons of social justice to anti-social-justice, comparisons of different groups' experiences.]

The social justice narrative describes a political-economic elite dominated by white males persecuting anybody who doesn't fit into their culture, like blacks, women, and gays. The anti-social-justice narrative describes an intellectual-cultural elite dominated by social justice activists persecuting anybody who doesn't fit into *their* culture, like men, theists, and conservatives. Both are relatively plausible; Congress and millionaires are 80% – 90% white; journalists and the Ivy League are 80% – 90% leftist.

The narratives share a surprising number of other similarities. Both, for example, identify their enemy with the spirit of a discredited mid-twentieth century genocidal philosophy of government; fascists on the one side, communists on the other. Both believe they're fighting a war for their very right to exist, despite the lack of any plausible path to reinstituting slavery or transitioning to a Stalinist dictatorship. Both operate through explosions of outrage at salient media examples of their out-group persecuting their in-group.

They have even converged on the same excuse for what their enemies call "politicizing" previously neutral territory – that what their enemies call "politicizing" is actually trying to restore balance to a field the other side has already successfully politicized. For example, on Vox recently a professor accused of replacing education with social justice propaganda in her classroom counterargues that:

> All of my students, regardless of the identity categories they embraced, had been taught their entire lives that

> real literature is written by white people. Naturally, they felt they were being cheated by this strange professor's "agenda"…It is worth asking, Who can most afford to teach in ways that are least likely to inspire controversy? Those who are not immediately hurt by dominant ideas. And what's the most dominant idea of them all? That the white, male, heterosexual perspective is neutral, but all other perspectives are biased and must be treated with skepticism […]
>
> Have we actually believed the lie that the only people who engage in "identity politics" are black feminists like me? Could it be that when some white men looked at more powerful white men, they could see them only as reasonable and not politically motivated, so they turned off their critical thinking skills when observing their actions? (Not everyone, of course.) Could it be that we only consider people ideologues when they don't vow allegiance to capitalism?

Compare to the "Sad Puppies", a group of conservatives accused of adding a conservative bent to science fiction's Hugo Awards. They retort that "politicization is what leftists call it when you fight back against leftists politicizing something". As per the Breitbart article:

> The chief complaint from the Sad Puppies campaigners is the atmosphere of political intolerance and cliquishness that prevails in the sci-fi community. According to the libertarian sci-fi author Sarah A. Hoyt, whispering campaigns by insiders have been responsible for the de facto blacklisting of politically nonconformist writers across the sci-fi community. Authors who earn the ire of the dominant clique can expect to have a harder time getting published and be quietly passed over at award ceremonies […]

> Brad R. Torgersen, who managed this year’s Sad Puppies campaign, spoke to Breitbart London about its success: “I am glad to be overturning the applecart. Numerous authors, editors, and markets have been routinely snubbed or ignored over the years because they were not popular inside WSFS or because their politics have made them radioactive.”
>
> Torgersen cites a host of authors who have suffered de facto exclusion from the sci-fi community: David Drake, David Weber, L.E Modesitt Jr, Kevn J. Anderson, Eric Flint, and of course Orson Scott Card — the creator of the world-famous Ender’s Game, which was recently adapted into a successful movie. Despite his phenomenal success, Scott Card has been ostracized by sci-fi’s inner circle thanks to his opposition to gay marriage.

I see minimal awareness from the social justice movement and the anti-social-justice movement that their narratives are similar, and certainly no deliberate intent to copy from one another. That makes me think of this as a case of convergent evolution.

The social justice attitude evolved among minority groups living under the domination of a different culture, which at best wanted to ignore them and at worst actively loathed them for who they were and tried to bully them into submission. The closest the average white guy gets to that kind of environment is wandering into a social-justice-dominated space and getting to experience the same casual hatred and denigration for them and everyone like them, followed by the same insistence that they’re imagining things and how dare they make that accusation and actually everything is peachy.

And maybe that very specific situation breeds a very specific kind of malignant hypervigilance, sort of halfway between post-traumatic stress disorder and outright paranoia, which motivates the obvious fear and hatred felt by both groups.

Someone is going to freak out and say I am a disgusting privileged shitlord for daring to compare the experience of people concerned about social justice to the experience of genuinely oppressed people, but they really shouldn't. That's the *explicit goal* of large parts of the social justice movement. For example, on the Hacker News thread about far-rightist Curtis Yarvin being kicked out of a tech conference for his views, one commenter writes:

> I've been involved in anti-racist/anti-fascist work, either directly or on the periphery, for about ten years at this point. This takes many forms, from street confrontations with fascists, protests at book readings and other events, and also disrupting fascist conferences and similar […]
>
> As far as this issue and other similar issues are concerned, I'm overjoyed that, as you put it, a climate of fear exists for fascists, misogynists, racists, and similar. I hope that this continues and only worsens for these people.
>
> I'm happy for many reasons. The first is that it has, as you've said, made privileged people afraid. I think this is only the beginning. Privilege creates safety, and as it is removed, I think the unsafety of the oppressed will in part come to the currently privileged classes. But if I could flip a switch and make every man feel the persistent, gnawing fear that a woman has of men, I

> would in a heartbeat. I wouldn't even consider whether the consequences were strategic, I would just do it.

This not the only time I've heard this opinion expressed, just the most recent. I feel like if you admit that you're trying your hardest to make privileged people feel afraid and uncomfortable and under siege in a way much like minorities traditionally do, and privileged people are in fact complaining of feeling afraid and uncomfortable and under siege in a way much like minorities traditionally do, you shouldn't immediately doubt their experience. Give yourself some more credit than that. You've been working hard, and at least in a few isolated cases here and there it's paid off.

The commenter continues:

> I would not say that I set out to defeat a "discourse-stifling" monster. The monsters I set out to defeat were patriarchy, capitalism, and white supremacy. These systems violently oppress, they don't "stifle discourse." In fact, they LOVE discourse! When people are discoursing, they aren't in the streets. I've seen so many promising movements hobbled by reformism that I'm glad the possibility no longer exists, though that isn't at all the fault of SJW-outrage (and is rather a consequence of the fact that the economy is in large part so perilous that nobody can afford the concessions that were previously won by reformists). So if discourse is permanently removed as a tactical and strategic option for future leftists, I'll consider it a victory.

Needless to say, that is not this blog's philosophy. But I think there is nevertheless something to be gained from all of the hard work this guy and his colleagues have put in making other people feel unsafe.

The mirror neuron has always been one of liberalism's strongest weapon. A Christian doesn't decide to tolerate Islam because she likes Islam, she decides to tolerate Islam because she can put herself in a Muslim's shoes and realize that banning Islam would make him deeply upset in the same way that banning Christianity would make *her* deeply upset.

If the fear and hypervigilance that majority groups feel in social-justice-dominated spaces is the same as the fear and hypervigilance that minority groups feel in potentially discriminatory spaces, that gives us a whole lot more mirror neurons to work with and allows us to get a gut-level understanding of the other side of the dynamic. It lets us check my intuitions against their own evil twins on the other side to determine when we are proving too much.

II.

A couple of months ago the owners of a pizzeria mentioned in an interview that they wouldn't serve pizza at gay weddings because they're against gay marriage. Instantly the nation united in hatred of them and sent a bunch of death threats and rape threats and eventually they had to close down.

I thought this was ridiculous. I mean, obviously death threats are never acceptable, but there seemed to be something especially frivolous about this case, where there are dozens of other pizzerias gay people can go to and where *no one would ever serve pizza at a wedding anyway*. A pizzeria hardly holds the World Levers Of Power, so just let them have their weird opinion. All they're doing is sending potential paying customers to their more tolerant competitors, who are laughing all the way to the bank. It's a self-punishing offense.

This was very reasonable of me and I should be praised for my reasonableness, *except* that when a technology conference recently booted a speaker for having far-right views on his own time, I was one of the many people who found this really scary and thought they needed to be publicly condemned for this intolerant act.

In theory, the same considerations ought to apply. There are dozens of other technology conferences in the world. Technology conferences *also* do not hold the World Levers Of Power. And when they reject qualified rightist speakers, that just means they're just making life easier for their competitors who will be happy to grab the opportunity and laugh all the way to the bank. It ought to be self-punishing, so what's the worry.

My brain is *totally not on board* with this reasoning. When I ask it why, it says something like "No, you don't understand, these people are relentless, unless they are constantly pushed against they will put pressure on more and more institutions until their enemies are starved out or limited to tiny ghettos. Then they will gradually expand the definition of 'enemy' until everybody who doesn't do whatever they say is blacklisted from everywhere."

And if you think that's hyper-paranoid, then, well, you're probably right, but at least I have a lot of company. Here are some other comments on the same situation from the last links thread:

> I spent a semester of college in Massachusetts. That's where I found out that there are a lot of people who'd kill me and most of my family if they were given the chance. And thought it was totally reasonable and acceptable to say as much. (The things that are

associated with Tumblr these days existed long before it. And mostly came from academia.)

About the same time that sort of thing was happening in that online community, the same thing was happening in the real-world meat-space gatherings, also quite literally with shrill screams, mostly by [reacted] [reacted]s, who would overhear someone else's private conversations, and then start streaming "I BEG YOUR PARDON!" and "HOW CAN YOU SAY THAT!", and by [reacted] [reacted]'s who were bullying their way onto programming committees, and then making sure that various speakers, panelists, artists, authors, dealers, and GoHs known to be guilty of wrongthink were never invited in the first place. Were it not for the lucky circumstance of the rise of the web, the market takeoff of ebooks, especially a large ebook vendor (named after a river)'s ebook direct program, and the brave anchoring of a well known genre publisher that was specifically not homed in NYC, the purging of the genre and the community would have been complete.

Almost nobody wants to physically murder and maim the enemy, at least at the start. That's, well, the Final Solution. Plan A is pretty much always for the enemy to admit their wrongness or at least weakness, surrender, and agree to live according to the conqueror's rules. Maybe the leaders will have to go to prison for a while, but everyone else can just quietly recant and submit, nobody has to be maimed or killed. [The social justice community] almost certainly imagine they can achieve this through organized ostracism, social harassment, and democratic political activism. It's when they find that this won't actually make all the racists shut up and go away, that we get to see what their Plan B, and ultimately their final solution, look like.

And if you think my commenters are also hyper-paranoid, then you’re probably *still* right. But it seems like the same kind of paranoia that makes gay people and their allies scream bloody murder against a single pizzeria, the kind that makes them think of it as a potential existential threat even though they’ve won victory after victory after victory and the only question still in the Overton Window is the terms of their enemies’ surrender.

I mocked the hell out of the people boycotting Indiana businesses because of their right-to-discriminate law:

> Can we admit it's KIND OF funny ppl are boycotting Indiana for the immoral act of allowing people to boycott those they think act immorally?
>
> — Scott Alexander (@slatestarcodex) March 31, 2015

But if some state were to pass a law specifically saying “It is definitely super legal to discriminate against conservatives for their political beliefs,” this would *freak me out*, even though I am not conservative and *even though this is already totally legal so the law would change nothing*. I would not want to rule out any response, up to and including salting their fields to make sure no bad ideas could ever grow there again.

Like many people, I am not very good at consistency.

III.

Author John Green writes books related to social justice. A couple of days ago, some social justice bloggers who disagreed with his perspective decided that a proportional response was to imply he was a creep who might sexually abuse children. Green was somewhat put out by this, and said on his Tumblr that he was “tired of seeing the language of social justice – important language doing important work – misused as a way to dehumanize others and treat them hatefully” and that he thought his harassers “were not treating him like a person”.

Speaking of the language of social justice, “dehumanizing” and “not treating like a person” are some pretty strong terms. They’re terms I’ve criticized before – like when feminists say they feel like women aren’t being treated as people, I’m tempted to say something like “the worst you’ve ever been able to find is a single-digit pay gap which may or may not exist, and you’re going to turn that into people not thinking you’re human?”

Here’s another strong term: “hatred”. The activist who got Mencius Moldbug banned from Strange Loop reassured us that he would never want someone banned merely for having unusual political views, but Moldbug went beyond that into “hatred”, which means his speech is “hate speech”, which is of course intolerable. This is a *bit* strange to anybody who’s read any of his essays, which seem to have trouble with any emotion beyond smugness. I call him a bloodless and analytical thinker; the idea of his veins suddenly bulging out when he thinks about black people is too silly to even talk about. The same is true of the idea that people should feel “unsafe” around him; his entire shtick is that no one except the state should be able to initiate violence!

Likewise, when people wanted TV star Phil Robertson fired for saying (on his own time) that homosexuality was unnatural and led to bestiality and adultery, they said it wasn’t about policing his religion, it was about how these were “hateful” comments that would make the people working with him feel unsafe. At the time I said that was poppycock and that people who wanted him fired for having a private opinion were the worst kinds of illiberal witch-hunters.

On the other hand, consider Irene Gallo. I know nothing of her except what the Alas blog post says, but apparently in science fiction's ongoing conflict between the establishment and the anti-SJW "Sad Puppies"/"Rabid Puppies" groups, she referred to the latter as:

> Two extreme right-wing to neo-nazi groups that are calling for the end of social justice in science fiction and fantasy. They are unrepentantly racist, sexist and homophobic.

These are some pretty strong allegations, and range from "false" to "bizarre"; Brad Torgenson, leader of the group she called "extreme right wing neo nazi unrepentant racists", is happily married to a black woman. And the people she's talking about are her company's authors and customers, which hardly seems like good business practice. Some authors have said they feel uncomfortable working for a company whose employees think of them that way, and others have suggested boycotting Tor until they make her apologize or fire her.

Barry says that since she said these on her own private Facebook page, it is a private opinion that it would be pretty censorious to fire her over. Part of me agrees.

On the other hand, if I were a sci-fi author in one of the groups that she was talking about, I'm not sure I'd be able to work with her. Like, really? You want me to sit across a table and smile at the woman who thinks I'm a racist sexist homophobic extremist neo-Nazi just because I disagree with her?

Robertson’s comment is just standard having-theological-opinions. Like, “Christian thinks homosexuality is sinful, more at eleven.” Big deal. But Gallo’s comment feels more like white hot burning hatred. She’s clearly too genteel to personally kill me, but one gets the clear impression that if she could just press a button and have me die screaming, she’d do it with a smile on her face.

But this is just interpretation. Maybe Gallo doesn’t consider “neo-Nazi” a term of abuse. Maybe this was just her dispassionate way of describing a political philosophy with the most appropriate analogy she could think of.

It doesn’t seem likely to me. Then again, even though it seems obvious to me that stating “homosexuality is sinful and similar to bestiality” is a theological position totally compatible with being able to love the sinner and hate the sin, gay people have a lot of trouble believing it. And although I cannot condone firing people for their private opinions, back when people were trying to get rid of Gawker honcho Sam Biddle for saying that “nerds should be constantly shamed and degraded into submission”, God help me it certainly crossed my head that there were even the slightest consequences for this kind of behavior, maybe other social justice writers would stop saying and acting upon statements like that *all the frickin’ time?*

Once again, I’m not scoring very highly in consistency here.

IV.

A little while ago I had a bad couple of days. Some people were suggesting I was a liability to a group I was part of because I’d written some posts critical of feminism, and I got in a big fight about it. Then someone sent my ex-girlfriend a Tumblr message asking if they’d broken up with me “because I was racist”. Then despite my best efforts to prevent this, my Facebook feed decided to show me a bunch of Gawker-style articles about “Are all white people to blame for [latest atrocity]? I was too exhausted to write a real blog post, so I just threw together a links post. Because among two dozen or so links there was one (1) to the Moldbug story previously mentioned above, one commenter wrote that “your links posts are becoming indistinguishable from Chaos Patch” (Chaos Patch is the links post of notable far-right blog Xenosystems).

So I decided to ban that commenter. But since I have a policy in place of waiting an hour before doing anything rash, I took a long walk, thought about it a bit, and settled for just yelling at him instead.

Is banning someone for a kind of meaningless barb excessive? Well, yes. But given everything else that had happened, I didn’t have the energy to deal with it, and since this is my blog and the one corner of the world I have at least a tiny bit of control over I could at least symbolically get rid of a small fraction of my problems.

Plus, to me the barb seemed like an obvious veiled threat. “As long as you post any links about rightist causes, I can accuse you of being far-right. And we all know what happens to far-right people, eh?”

So even though out of context it was about the most minimal hostility possible, barely rising to the level where somebody would say it was even capable of being a problem at all, in context it really bothered me and made me at least somewhat justifiably feel unsafe.

Ever since I learned the word “microaggression” I have been unironically fond of it.

> Microagressions. Nanoagressions. Picoagressions. The Planck Hostility.
>
> — Map of Territory (@MapOfTerritory) January 28, 2015

When I'm putting up with too much and I've used up my entire mental buffer, then somebody bothering me and hiding under the cover of "oh, this was such a tiny insult that you would seem completely crazy to call me on it" is *especially* infuriating, even more infuriating than someone insulting me outright and me being able to respond freely. The more you have to deal with people who hate you and want to exclude you, the more likely you are to get into this mode, not to mention people who have developed their own little secret language of insults.

Here's an example of what I mean by "secret language of insults": consider the term "dude", as in "white dude". There is nothing objectively wrong with "dude" when it is applied to surfers or something. But when a feminist says it, as in the term "white dudes", you know it is going to be followed by some claim that as a white dude, you are exactly the same as all other white dudes and entirely to blame for something you don't endorse. The first page of Google results is overratedwhitedudes.tumblr.com, Gawker saying Wimpy White Dudes Ruined American Idol, and Mother Jones saying glowingly that You Won't Find Many White Dudes At This Tech Startup. Being called a "white dude" is always followed by the implication that you're ruining something or that your very presence is cringeworthy and disgusting.

I had a feminist friend who used to use the term "dudes" for "men" all the time. I asked them to please stop. They said that was silly, because that was just the word the culture they'd grown up in used, and obviously no harm was meant by it, and if I took it as an insult then I was just being oversensitive. This is *word for word* the explanation I got when I asked one of my elderly patients to stop calling black people *their* particular ethnic slur.

The counterpart to subliminal insults is superliminal insults; ones that are hard to detect because they're so over-the-top obvious.

I was recently reading a social justice blog where someone complained about men telling women "Make me a sandwich!" in what was obvious jest.

On the one hand, no one can possibly take this seriously.

On the other hand, there's a common social justice meme where people post under the hashtag #killallwhitemen.

Certainly this cannot be taken seriously; most social justice activists don't have the means to kill all white men, and probably there are several of them who wouldn't do it even if they could. It should not be taken, literally, as a suggestion that all white men should be killed. On the other hand, *for some bizarre reason* this tends to make white men uncomfortable.

The obvious answer is that the people posting "Wimmen, make me a sandwich!" don't literally believe that women exist only for making them sandwiches, but they *might* believe a much weaker claim along the same lines, and by making the absurd sandwich claim, they can rub it in while also claiming to be joking. At least this is how I feel about the "kill all white men" claim.

As long as you've got a secret language of insults that your target knows perfectly well are insulting, but which you can credibly claim are not insulting at all – maybe even believing it yourself – then you have the ability to make them feel vaguely uncomfortable and disliked everywhere you go without even trying. If they bring it up, you can just laugh about how silly it is that people believe in “microaggressions” and make some bon mot about “the Planck hostility”.

V.

I’m taking a pretty heavy Outside View line here, so let me allow my lizard brain a few words in its own defense.

“Yes,” my lizard brain says, “social justice activists and the people silenced by social justice activists use some of the same terms and have some of the same worries. But the latter group has *reasonable* worries, and the former group has totally *unreasonable* worries, which breaks the symmetry.”

Interesting. Please continue, lizard brain.

“Black people might be very worried about being discriminated against. But the chance that someone would say ‘Let’s ban all black people from our technology conference, because they are gross’, and everyone would say ‘Yes, that is a splendid idea’, and the government and media would say ‘Oh, wonderful, we are so proud of you for banning all black people from your conference’ is zero point zero zero zero. On the other hand, this is something that conservatives worry about every day. The chance that someone would say ‘You know, there’s no reason raping women should be illegal, let’s not even bother recording it in our official statistics’ is *even lower than that*, but this is exactly what several countries do with male rape victims. If someone says ‘kill all white men’, then all we do is hold an interminable debate about whether that disqualifies them from the position of Diversity Officer; if someone said ‘kill all gays’, we would be much more final in pronouncing them Not Quite Diversity Officer Material.”

But don’t you –

“The reason why we don’t care about a pizzeria that won’t serve gay people is that recent years have shown an overwhelming trend in favor of more and more rights and acceptance of gay people, and the pizzeria is a tiny deviation from the pattern which is obviously going to get crushed under the weight of history even without our help. The reason we worry about a conference banning conservatives is that conservatives are an actually-at-risk group, and their exclusion could grow and grow until it reaches horrific proportions. The idea of a pizzeria banning gays and a conference banning conservatives may seem superficially similar out of context, but when you add this piece of context they’re two completely different beasts.”

Two responses come to mind.

First, this is obviously true and correct.

Second, this is exactly symmetrical to my least favorite argument, the argument from privilege.

The argument from privilege is something like “Yeah, sure, every so often the system is unfair to white people or men or whatever in some way. But this is not a problem and we should not even be talking about it, because privilege. Shows that mock women for stereotypically female failings are sexist, but shows that mock men for stereotypically male failings are hilarious, and you may not call them sexist because you can’t be sexist against privileged groups.”

My argument has always been “What’s good for the goose is good for the gander”.

But either this argument goes, or my lizard brain’s argument goes, or we have to move to the object level, or somebody has to get more subtle.

VI.

My point is, there are a lot of social justice arguments I *really* hate, but which I find myself unintentionally reinventing any time things go really bad for me, or I feel like myself or my friends are being persecuted.

I should stop to clarify something. “Persecuted” is a strong word. “Feel like we are being persecuted” is way weaker.

A couple weeks ago there was a Vox article, America’s Never Been Safer, So Why Do Republicans Believe It Is In Mortal Peril?. It brought up a lot of cute statistics, like that the rate of pedestrians being killed by car accidents is much higher than the rate of civilians being killed in terrorist attacks. It joked that “You’re over 100 times more likely to die by literally walking around than you are to be killed in a terrorist attack.”

On the other hand, vox has practically led the news media in 24-7 coverage of police officers shooting unarmed black people, talking about how it’s a huge threat to our values as a civilization and how white people don’t understand that all black people have to constantly live in fear for their lives.

But a quick calculation demonstrates that unarmed black people are about 10 times more likely to die by *literally walking around* than by getting shot by a white police officer. One gets the feeling Vox doesn’t find this one nearly as funny.

But here I would perform another quick calculation. Here's a list of people who have been publicly shamed or fired for having politically incorrect opinions. Even if we assume the list is understating the extent of the problem by an entire order of magnitude, you're *still* more likely to die by literally walking around than you are to get purged for your politically incorrect opinion.

Like a lot non-feminists, I was freaked out by the recent story about a man who was raped while unconscious being declared the rapist and expelled from college without getting to tell his side of the story. I have no evidence that this has ever happened more than just the one time mentioned in the article, let alone it being a national epidemic that might one day catch me in its clutches, but because I've had to deal with overly feminist colleges in other ways, my brain immediately raised it to Threat Level Red and I had to resist the urge to tell my friends in colleges to get out while they still could. If we non-feminists can get worried about this – and we can – we have less than no right to tell feminists they shouldn't *really* be worried about college rape because the real statistics are 1 in X and not 1 in Y like they claim.

Hopefully some readers are lucky enough never to have felt much personal concern about terrorism, police shootings, rape, rape accusations, or political correctness. But if you've worried about at least one of these low-probability things, then I hope you can extend that concern to understand why other people might be worried about the others. It seems to have something to do with the chilling effect of knowing that something is intended to send a message to you, and in fact receiving that message.

(as an aside, I find it surprising that so many people, including myself, are able to accept the statistics about terrorism so calmly without feeling personally threatened. My guess is that, as per Part VIII here, we don't primarily identify as Americans, so a threat deliberately framed as wanting to make Americans feel unsafe just bounces off us.)

In an age where the media faithfully relates and signal-boosts all threats aimed at different groups, and commentators then serve their own political needs by shouting at us that WE ARE NOT FEELING THREATENED ENOUGH and WE NEED TO FEEL MORE THREATENED, it is very easy for a group that faces even a small amount of concerted opposition, even when most of society is their nominal allies and trying hard to protect them, to get pushed into a total paranoia that a vast conspiracy is after them and they will never be safe. This is obviously the state that my commenters who I quoted in Part II are stuck in, obviously the state that those people boycotting the Indiana pizzeria are stuck in, and, I admit, a state I'm stuck in a lot of the time as well.

VII.

Getting back to the thesis, my point is there are a lot of social justice arguments I *really* hate, but which I find myself unintentionally reinventing any time things go really bad for me, or I feel like myself or my friends are being persecuted.

Once events provoke a certain level of hypervigilance in someone – which is very easy and requires only a couple of people being hostile, plus the implication that they there's much more hostility hidden under the surface – then that person gets in fear for their life and livelihood and starts saying apparently bizarre things: that nobody treats them as a person, that their very right to exist is being challenged. Their increasingly strident rhetoric attracts increasingly strident and personal counter-rhetoric from the other side, making them more and more threatened until they reach the point where Israel is stealing their shoe. And because they feel like every short-term battle is the last step on the slippery slope to their total marginalization, they engage in crisis-mode short-term thinking and are understandably willing to throw longer-term values like free speech, politeness, nonviolence, et cetera, under the bus.

Although it’s very easy enter this state of hypervigilance yourself no matter how safe you are, it’s very hard to understand why anyone else could possibly be pushed into it despite by-the-numbers safety. As a result, we constantly end up with two sides both shouting “You’re making me live in fear, and also you’re making the obviously false claim that you live in fear yourself! Stop it!” and no one getting anywhere. At worst, it degenerates into people saying “These people are falsely accusing me of persecuting them, *and* falsely claiming to be persecuted themselves, I’ll get back at them by mocking them relentlessly, doxxing them, and trying to make them miserable!” and then you get the kind of atmosphere you find in places like SRS and Gamergate and FreeThoughtBlogs.

But I’m also slightly optimistic for the future. The conservative side seems to have been about ten years behind the progressive side in this, but they’re catching up quickly. Now *everybody* has to worry about being triggered, *everybody* has to worry about their comments being taken out of context by Gawker/Breitbart and used to get them fired and discredit their entire identity group, *everybody* has to worry about getting death threats, et cetera. This is bad, but also sort of good. When one side has nukes, they nuke Hiroshima and win handily. When both sides have nukes, then under the threat of mutually assured destruction they eventually come up with protocols to prevent those nukes from being used.

Now that it’s easier to offend straight white men, hopefully they’ll agree trigger warnings can be a useful concept. And now that some social justice activists are getting fired for voicing their opinions in private, hopefully they’ll agree that you shouldn’t fire people for things they say on their own time. Once everyone agrees with each other, there’s a chance of getting somewhere. Yes, all of this will run up against a wall of “how dare you compare what I’m doing to what you’re doing, I’m defending my right to exist but you’re engaging in hate speech!” but maybe as everyone gets tired of the nukes flying all the time people will become less invested in this point and willing to go to the hypothetical Platonic negotiation table.

My advice for people on the anti-social justice side – I don’t expect giving the SJ people advice would go very well – is that it’s time to stop talking about how social justice activism is necessarily a plot to get more political power, or steal resources, or silence dissenting views. Like everything else in the world it can certainly turn into that, but I think our *own* experience gives us a lot of reasons to believe they’re exactly as terrified as they say, and that we can’t expect them to accept “you have no provable objective right to be terrified” any more than our lizard brains would accept it of us. I think it’s time to stop believing that they censor and doxx and fire their opponents out of some innate inability to understand liberalism, and admit that they probably censor and doxx and fire their opponents because they’re as scared as we are and feel a need to strike back.

This isn’t a claim that they don’t have it in for us – many of them freely admit they do – and that they don’t need to be stopped. It’s just a claim that we can gain a good understanding of *why* they have it in for us, and how we might engineer stopping them in a way less confrontational than fighting an endless feud.

Yesterday, a friend on Facebook posted something about a thing men do which makes women feel uncomfortable and which she wanted men to stop. I carefully thought about whether I ever did it, couldn't think of a time I had, but decided to make sure I didn't do it in the future.

I realized that if I'd heard the exact same statement from Gawker, I would have interpreted it (correctly) as yet another way to paint men as constant oppressors and women as constant victims in order to discredit men's opinions on everything, and blocked the person who mentioned it to me so I didn't have to deal with yet another person shouting that message at me. The difference this time was that it came from an acquaintance who was no friend of feminism, who has some opinions of her own that might get her banned from tech conferences, and who I know would have been equally willing to share something women do that bothers men, if she had thought it important.

If we can get to a point where we don't feel like requests are part of a giant conspiracy to discredit and silence us, people *are* sometimes willing to listen. Even *me*.

Fifty (More) Swifties

[see: Wikipedia: Tom Swifties, Tom Swifties Written By An Author Willing To Go To Any Lengths To Make A Tom Swifty Thus Resulting In Constructions That Often Require More Work For Readers Than For The Author, and Fifty Swifties. Previously on Twitter here.]

“This sandwich is gross,” Tom said deliberately.

“My Frisbee is stuck on the roof of that circus building,” Tom said discontently.

“I hate Google,” Tom said probingly.

“Godzilla swallowed a United Nations bunker, but then he threw it back up,” Tom said unfortunately.

“I think Objectivism is stupid,” Tom said randomly.

“It’s so exciting to visit Leonardo’s birthplace,” Tom said invincibly.

“Persephone must marry Hades and live with him half the year,” Zeus said despairingly.

“I now control majority shares of CBS, FOX, and the New York Times,” Tom said immediately.

“Enemy fighters just scored a direct hit on my plane! I’m going down!” Tom said knowingly.

“We were badly injured in the struggle with the Orcs, but luckily the Ents’ medicine restored our health,” Tom said tremendously.

“I took Gollum’s precious trinket in a riddle contest,” Tom said wonderingly.

“I’ve lost this Maxis game ten times in a row on the easiest difficulty setting,” Tom said sympathetically.

“I can commit adultery three more times and still be just under the threshold for damnation,” Tom said syntactically.

“O Lord, why are you punishing me like this?” Jonah said inefficiently.

“Look! Nicaraguan guerillas!” Tom contraindicated.

“I forgot to give up meat before Easter, so I’ll do it before Christmas,” Tom said redolently.

“I’ll see you in court!” Tom said supersonically.

“When I speak Japanese, I think of myself as a young, cute person,” Tom said mechanically.

“Iä Cthulhu! Iä Azathoth!” the man called maniacally.

“Stay away from Stalin,” Tom commissioned.

“It’s one of those old phones, from before wireless and touch-tone,” Tom said cordially.

“She’ll have sex with me for $20 any time I phone her up,” Tom said horrifically.

“I read the Cliff Notes to Dante’s Inferno,” Tom said synergistically.

“I’m going to recover the lunar lander from the surface of the moon and make a fortune,” Tom said apologetically.

“I covered myself in a layer of gold,” Tom said amblingly.

“I covered myself in a layer of pyrite,” Tom said shamblingly.

“I covered myself in the Golden Fleece of Colchis,” Tom said ramblingly.

“The poverty rate has increased 10% recently, but I don’t have any kind of visual presentation of its course,” Tom said pornographically.

“We should perform an autopsy,” Tom said wide-eyed.

“That tree is naked under its bark!” Tom said prudently.

“I can afford either an iPhone or a yacht, but not both,” Tom said on self-ownership.

“The guy who was installing the granite tops in my kitchen had a cardiac arrest,” Tom countermanded.

“We can stop progress by attacking a conference on new ideas with a many-headed monster,” Tom said well-hydrated.

“You’re a bell,” Tom told me.

“The wages of sin is death,” Tom said diurnally.

“Abortion is murder,” Tom said prolifically.

“Can do!” Tom said candidly.

“I have a present for you, Madame,” Vincent said endearingly.

“Arrrrrr,” Tom aspirated.

“My lower social status as part of the new rich prevents me from winning my true love,” Gatsby said lackadaisically.

“The Minoans sucked,” Tom said discretely.

“Well, if you think the Minoans did a bad job with their empire, you should try ruling them yourself,” his teacher said, giving him a B−.

“Ha ha, just kidding,” Tom ingested.

“Sheep can’t have sex changes!” Tom said, heedless of the ramifications.

“I wrote a synoptic Gospel,” Tom remarked.

“People used to lay wires across the country for the telegraph system, an early precursor to the telephone,” Tom said according to protocol.

“My laptop came bundled with malware that causes a serious security flaw,” Tom said superficially.

“We need artillery cover!” Tom said canonically.

“Someday my family will rule the world,” Tom said clandestinely.

“The West’s treatment of Palestine is an example of Orientalism,” Tom said.

NYC Meetup 6/21

There will be a New York City rationalist/LW/SSC meetup at Highgarden House (851 Park Place, Brooklyn) on Sunday, June 21 at 6:00 PM.

If you're reading this and can make it to Brooklyn, you're invited.

There was some discussion at the last Michigan meetup about how the previous sentence should be emphasized more. *If you are reading this and can make it to Brookyln, you're invited.* You do not have to have read very much of this blog or Less Wrong, you do not have to agree with me about anything, you do not have to "be smart enough" for any value of "enough", and you do not have to be the "sort of person" whom you think goes to these things. In fact, the less you are like everyone else, the more interesting it will be.

No particular activities planned beyond undirected conversation, but there will probably be pizza, and Raemon has raised the possibility of a dramatic reading of *A Midsummer Night's Dream* for those interested (in a way that doesn't interfere with continued undirected conversation for people who don't want to do that).

Reflections From The Halfway Point

I.

A while back one of my patients was having a foot problem, so I consulted the hospital podiatrist. He met me in my workroom, and I explained exactly what I needed from him, but over the course of the explanation he started looking more and more uncomfortable and distracted, so finally I stopped and was just like “Okay, out with it, what’s your problem?”

And he said: “That guy with the wild hair pounding on the window and shouting threats and obscenities at us.”

And I said: “Oh, *him*? That’s just Bob. Don’t worry about him, he always does that.”

The podiatrist seemed inadequately reassured.

I thought about this because as of today I am halfway done with my four-year psychiatry residency.

One of my teachers told me that you go to medical school to learn things, and then you go to residency to get used to them. It’s not *quite* that simple – you certainly learn a lot in residency – but there’s a lot of truth to it. I remember that my first week on call, somebody had a seizure and I totally freaked out – AAAAH SEIZURE WHAT DO I DO WHAT DO I DO? – even though I had previously been able to pass tests on that exact situation. But my last time on call, somebody *also* had a seizure, and I sort of strolled in half-asleep, ordered the necessary tests and consultations and supportive care, then strolled out and went back to bed.

And then there are the little things, like learning to tune out a psychotic guy banging on the window and yelling threats at you.

II.

It’s interesting that psychiatric hospitals are used as a cliche for “a situation of total chaos” – I think I’ve already mentioned the time when the director of a psych hospital I worked at told us, apparently without conscious awareness or irony, that if Obamacare passed our hospital would have too many patients and “the place would turn into a madhouse”. There’s a similar idiom around “Bedlam”, which comes from London’s old Bethlehem psychiatric hospital.

In fact, psych hospitals are much more orderly than you would think. Maybe 80% of the patients are pretty ‘with it’ – depressed people, very anxious people, people with anger issues who aren’t angry at the moment, people coming off of heroin or something. The remaining 20% of people who are very psychotic mostly just stay in their rooms or pace back and forth talking to themselves and not bothering anyone else. The only people you really have to worry about most of the time are the manic ones and occasionally severe autistics, and even they’re usually okay.

For a place where two dozen not-very-stable people are locked up in a small area against their will, violence is impressively rare. The nurses have to deal with some of it, since they’re the front-line people who have to forcibly inject patients with medication, and they *have* gotten burned a couple of times. And we doctors are certainly trained to assess for it, defuse it, and if worst comes to worst hold our own until someone can get help.

Yet in the two years I’ve worked at Our Lady Of An Undisclosed Location, years when each doctor has talked to each of their patients at least once a day, usually alone in an office, usually telling them things they really don’t want to hear like “No, you can’t go home today” – during all that time, not one doctor has been attacked. Not so much as a slap or a poke.

I am constantly impressed with how deeply the civilizing instinct has penetrated. When I go out of the workroom and tell Bob, “I’m sorry, but you’re disturbing people, you’re going to have to stop banging on the window and shouting threats, let’s go back to your room,” then as long as I use a calm, quiet, and authoritative voice, that is what he does. With very few exceptions, there is nobody so mentally ill that calmness + authority + the implied threat of burly security guards won’t get them to grumble under their breath but generally comply with your requests, reasonable or otherwise.

III.

I’d like to say I’ve taken advantage of this to go mad with power. But it’s actually a really crappy situation for everyone involved.

The most common reason for admission to a psychiatric hospital is “person is a danger to themselves or others”. The average length of stay in a psychiatric hospital is about one week.

Some clever person might ask: “Hey, don’t most psychiatric medicines require more than a week to take effect?” Good question! The answer is “yes”. Antidepressants classically take four weeks. Lithium and antipsychotics are more complicated, but the textbooks will still tell you a couple of weeks in both cases. And yet people are constantly being brought to psychiatric hospitals for dangerousness, treated with medications for one week, and then sent off. What gives?

As far as I can tell, a lot of it is the medical equivalent of security theater.

The most common type of case I see is “person who was really angry, said ‘I’ll kill myself’ in a fit of rage, and then their partner called the cops and they were brought to hospital.” These people stop being angry after a day or two and then no longer make these comments, even assuming they meant it in the first place which most of them don’t.

The second most common type of case I see is “person who was really angry, did try to kill themselves, and it didn’t work.” Again, these people have stopped being angry. Failed suicide attempts also have their own interesting way of clearing the mind for a little while, so they’re in a sort of grace period. Sending these people to a psychiatric hospital makes the public feel good because they’re Doing Something About Suicide, and makes psychiatrists feel good because after a few days they’ve stopped being suicidal so it looks like we’re Making A Difference. There is no way we could leave this equilibrium now even if we wanted to, because if we *didn’t* keep these people for a week and they ever attempted suicide again, we would get sued to oblivion.

The third most common type of case I see is “severely mentally ill person who’s been living at a care home for twenty years, but then they got in a fight and so their care home sent them to the hospital.” We shuffle their medications around and send them back to the care home where they'd been living happily for twenty years until some random trigger set them off.

We don't call this "security theater". We do sometimes call it a "holding environment". Psych hospitals are kind of boring. There's no boyfriend to get in a screaming match with, no boss pushing you to work harder, and no drug dealers to get heroin from. On the other hand, there's lots of structure – art therapy at 10, meeting with your doctor at 11, recreation group at 12, and so on. It's like a terrible vacation in the world's least attractive hotel. People get a chance to cool off and forget about whatever set them off. Then they go back to their life. If they're lucky, our social workers have managed to connect them to a better outpatient psychiatrist, care home, or support group, and maybe that will improve their lives sometime down the line. But I don't think anyone imagines there was some fundamental Quality Of Dangerousness in them which is now gone.

To the degree that it *is* all security theater, it's really hard to give an honest answer to a patient asking why they have to stay in hospital.

When I first started this work, my reaction to these people was "Come on, it's only a week, it's not like you're stuck here forever, just deal with it." This lasted until I remembered that when some stupid policy forces *me* to come into hospital on a day I would otherwise have off, I freak out, because I value my free time too much to be okay with having it taken away from me for bad reasons. Heck, my power was out the past couple of days, and I couldn't use the Internet, and I was calling the power company and being like "COME ON YOU NEED TO FIX THIS ALREADY I AM LOSING DAYS OF MY LIFE THAT I COULD OTHERWISE BE

SPENDING IN IMPORTANT STUFF.” So now I try to avoid throwing stones.

(there’s another aspect of this, which is that people constantly protest that horrible things will happen to them based on that week. For example: “My boss said if I miss one more day of work, I’ll lose my job, and then I’ll have no way to support my family.” Or: “My rent payment is due tomorrow, if I miss it I’ll be evicted and all of my stuff will go to the landfill, and there’s no way I can handle this through Internet or telephone or asking a friend to help.” I assume 90% of these stories are false, but the 10% that are true are still bad enough to more than outbalance any good we can do.)

After that, my reaction to these people was “Yes, you may be angry now, but you will thank us later.” This is true of many people, including some of the most histrionically upset. But I’ve since learned that it’s probably *not* true of the majority. The Shrink Rap blog surveyed former psychiatric inpatients and found that 62% said their experience was not helpful and they were “the same or worse at discharge”. I’d like to dismiss this as people just carrying a grudge for having to be there at all, but the same survey finds that a very similar 56% of voluntarily admitted patients said the same thing (although not all “voluntary” admissions are as voluntary as the name expects). Now, I don’t know for sure what to think about that survey – a lot of people describe their hospitals as doing things which are super illegal and which I wouldn’t expect a hospital to be able to get away with and stay open for more than twenty-four hours, and the population of psych patients who read psychiatric blogs is probably a nonrandom sample – but I no longer feel like I can confidently say that our patients will thank us later.

(none of this is to say that you shouldn’t check yourself into a hospital if you’re feeling suicidal – you’ll get the holding environment that makes sure you don’t kill yourself for the immediate future, you’ll get connected to a system that can give you useful referrals and medications much faster, and 38% will also end up being directly helped.)

So now what I tell people is the Cliffs’ Notes version of the above – “I’m sorry you have to be here, but we are going to keep you for a few more days to evaluate you, your estimated day of discharge is X but that’s not a promise, if there’s anything specifically making you uncomfortable please let me or the nurses know and we’ll see what we can do.”

I can’t figure out a good way to say the spiel without the last sentence, which is too bad because then they *do* let me and the nurses know things. Most of them are things that I, as a low-ranking doctor who cannot totally rearrange the unit according to my will, have no ability to change. Some of them are things *nobody* can change.

Like! It turns out when you lock constitutionally anxious people in a new environment full of psychotic people, they become really really anxious. They tend to request antianxiety drugs. I am happy to give them reasonable doses of the non-addictive anti-anxiety drugs, which then totally fail to do anything, because their idiot outpatient psychiatrist was giving them heroin mixed with horse tranquilizers every day or something. They demand whatever they were getting on the outside, but twice as much, and I can’t give it to them even if I want to because of our safety policies. And now I’m the bad guy.

Or! Some people don’t like noise. I sympathize with this as I am just about the most misophonic person in the world. On the other hand, there's always one screamer in a psychiatric hospital. Sometimes this screamer chooses to do their thing at four in the morning. The law gives us limited ability to lock them in a soundproof room, and definitely not all the time. So if you are startled by loud noise, you are kind of out of luck. Even if we can put you on the other side of the ward, you're still going to be bothered by staff coming in your room every fifteen minutes to make sure you haven't killed yourself, which they are legally required to do. You can complain that the lack of sleep is hurting your recovery, and I believe you, but aside from showing you where we keep the earplugs there's not much I can do. Once again, now I'm the bad guy.

Add to this people with picky tastes that our kitchen can't satisfy, people who get bored in the absence of some kind of entertainment we can't provide, smokers who are unsatisfied by nicotine patches, and the occasional very honest drug addict who just wants some drugs, and I spend about 30% of my day patiently explaining to people why their preferences are totally reasonable and I realize they're in pain but there's nothing I can do for them at this moment.

And I know it sounds really selfish of me to say so, but this is *really exhausting*.

As you may have guessed, I do not very much like inpatient work. You can adjust to having to treat someone having a seizure. You can adjust to somebody banging on the window and screaming. But it's really hard to adjust to constant moral self-questioning.

IV.

Now I am halfway done with my residency. I will be switching to outpatient work. Everyone who sees me will be there because they want to see me, or at worst because their parents/spouses/children/friends/voices are pressuring them into it. I will be able to continue seeing people for an amount of time long enough that the medications might, in principle, work. It sounds a lot more pleasant.

I have two equal and opposite concerns about outpatient psychiatry. The first is that I might be useless. Like, if someone comes in complaining of depression, then to a first approximation, after a few basic tests and questions to rule out some rarer causes, you give them an SSRI. I have a lot of libertarian friends who think psychiatrists are just a made-up guild who survive because it’s legally impossible for depressed people to give themselves SSRIs without paying them money. There’s some truth to that and I’ve previously joked that some doctors could profitably be replaced by SSRI vending machines.

The second concern is that everybody *still* screws it up. There’s an old saying: “Doctors bury their mistakes, architects cover theirs with vines, teachers send theirs into politics.” Well, outpatient psychiatrists send their mistakes to inpatient psychiatrists, so as an inpatient psychiatrist I’ve gotten to see a lot of them. Yes, to a first approximation when a person comes in saying they’re depressed you can just do a few basic tests and questions and then give them an SSRI. But the number of cases I’ve seen that end in disaster because their outpatient psychiatrist forgot to do the basic tests and questions, or decided that Adderall was the first-line medication of choice for depression – continues to boggle my mind. So either it’s harder than I think, or I’m surrounded by idiots, or I’m an idiot and don’t know it yet. In which case I’m about to learn.

The Case Of The Famous Physicist

I.

Old news, but I only just heard about it: Long Island woman says psych ward doctors believed she was delusional for insisting Obama follows her on Twitter.

The story: a woman was brought in for psychiatric evaluation. During the evaluation, she said President Obama followed her on Twitter. The psychiatrists decided she was psychotic and forced medication on her. But in fact, President Obama *does* follow her on Twitter, just as he does six hundred thousand other people. So they committed a perfectly sane person for telling the truth, leading to what the article calls a "frightening eight-day 'One Flew Over the Cuckoo's Nest' ordeal".

I don't know anything about this case or this person, and I definitely don't want this to sound like I have anything to say one way or the other about this person I have never met. But I've been involved in enough similar cases to have a different perspective, and wonder whether it was *quite* as much of an outrage as the article makes it out to be.

Consider: the comment occurred when she was in a hospital for psychiatric evaluation; that is, she was brought in before the Obama comment. According to the article:

> The bizarre experience began Sept. 12, when the NYPD seized her prized 2003 BMW 325Ci in Harlem because they suspected she was high on weed, her attorney, Michael Lamonsoff, said. Cops found no marijuana but confiscated her ride anyway, he said. The NYPD declined to comment.

> The following day, Brock walked into the NYPD’s Public Service Area 6 stationhouse in Harlem to retrieve her car, her suit charges. Brock — an eccentric 32-year-old born in Jamaica with dreams of making it big in the entertainment business — admitted in an interview she was “emotional,” but insisted she in no way is an “emotionally disturbed person.” Nevertheless, cops cuffed her and put her in an ambulance bound for the hospital, her suit charges.

This sort of elides over everything in between “went into police station to ask for car back” and “cops put her in an ambulance bound for the hospital.”

I’ve had patients sent by ambulance from the police station. It’s almost always because they started screaming and yelling threats at the police. Now, screaming and yelling threats at the police, although not a very good idea, is not always evidence of psychiatric disorder. But it often is. If you’re manic or on drugs, you’re a lot more likely to have the particular type of bad judgment that makes screaming threats at the police a seem like a good idea.

I don’t know what happened with Ms. Brock, and the article doesn’t say. I did take a look at her Twitter account, which is mostly angry tweets about Trayvon Martin, Mike Brown, and police brutality. A typical example is “Police should NOT be allowed to Murder the Citizens of this Country!!! We are not Animals!!! @BarackObama fix it NOW!!! #BanGuns”. Other Tweets seem maybe a *little* threatening, like the ominous “@NRA you people deserve the wrath that will come on you in the very near future” and “#KillRacists”.

So – cops have just done something very unfair to a person who likes making threats and doesn’t like cops, and who herself admits to being “eccentric” and “emotional”. Again without knowing what went on, my guess was that her “discussion” with the police was sufficiently exciting that they thought calling an ambulance for a psych evaluation was in order.

So imagine you’re the psychiatrist, you’re receiving a patient from the police for evaluation after she got in some kind of screaming match with them. And now she tells you Barack Obama follows her on Twitter. The article says:

> “I told (the doctor) Obama follows me on Twitter to show her the type of person I am. I’m a good person, a positive person. Obama follows positive people!”

This is weird. At best, it displays a surprising ignorance of how Twitter and/or the world works. Yes, Obama follows 600,000 people on Twitter, but this does not prove that each of them is “a positive person”. I would assume he has some program that auto-follows anyone who mentions him. When you think this is a good thing to bring up during, of all times, a psychiatric evaluation, then I start to wonder.

(I should also add here that, in my limited experience, social media is God’s gift to grandiose psychiatric patients. None of them are “a guy with a Facebook page”. They’re all “social media celebrities with hundreds of followers”. It’s always “YOU CAN’T DO THIS TO ME! DO YOU KNOW WHO I AM? I HAVE HUNDREDS OF FOLLOWERS ON TWITTER! EVEN [NAME OF TWITTER PERSON I HAVE NEVER HEARD OF] FOLLOWS ME! THIS IS GOING TO GO VIRAL!” One patient even told me, in a threatening manner, that his blog had over a thousand hits. “You mean a day?” I asked. “No, total,” he answered. Then he wondered why I was so utterly failing to look impressed.)

So I’ll be honest – if someone had just been brought in for a shouting match with police, and the first thing they told me was that President Obama follows them on Twitter, well – among the most common symptoms of mania are irritability, grandiosity, poor decision-making, and flamboyant dress/behavior. Getting in a fight with the police sounds like irritability and poor decision making, thinking the Obama follow is relevant sounds like grandiosity, and this woman’s Facebook photo certainly suggests a flair for the dramatic. I’m not saying I would commit her on these factors alone, and a lot would depend on the rest of the history. But she already would have started digging quite a hole for herself.

Remember, psychiatrists have to err on the side of admitting people. Even if this lady didn’t have bipolar disorder, admitting her for further evaluation seems like the sort of thing that could be an honest mistake.

(and remember, all we have here is her side of the story. Goodness only knows what other things she might have said or done that she *didn’t* leak to the news for their sympathetic article based entirely on her testimony.)

But then there’s the hospital’s subsequent conduct. They said that she was unemployed. She said she was employed and could prove it through her Internet presence. The hospital apparently didn’t check and continued to say she was psychotic for thinking she was employed. How could the hospital possibly be so incompetent as to not check the link to her employment documents that she was personally giving them?

II.

Let me tell you about Professor T.

(this would be a good time to reiterate that every patient story I tell here is a composite of multiple different people with all of the details changed around to protect anonymity. The gist of the story points out a true thing, but the specifics are all twisted around so thoroughly that even the people involved couldn’t recognize themselves.)

Professor T came to me in handcuffs. The police had picked him up in response to a call at the local university, where he was trying to give a lecture to a class that wasn’t expecting him and didn’t want him. The class’s actual professor had asked what the heck he was doing, and he had explained that he was Professor T, world-famous physicist, and that the head of the college had invited him to give a lecture on his theories that day, and it was too bad she hadn’t communicated this with the rank-and-file teachers, but he was a very busy man and they should all be honored by his presence and stop what they were doing and listen to him. A quick call to the administration confirmed none of them had any idea who he was either, and when he refused to go away or stop trying to lecture, the police were called to remove him. He started yelling and screaming at the police and telling them they were fools who were too small-minded to recognize a great scientist when they saw one, and they’d get what was coming to them. The cops decided this was a job for a psychiatrist and brought him to me.

Professor T said he’d just been really angry that the bureaucracy had screwed up badly enough to make him miss his lecture and that no one was willing to accommodate him. He was, after all, a very important leading scientist with a busy schedule. He moved in elite circles! Famous people like Edward Witten knew him well, and I was welcome to call to confirm that! He was used to being shown more respect!

I got a weird vibe from Professor T during our discussion. I know that distinguishing between professors and hobos is a famously difficult problem, but he just struck me as a *little* too much towards the hobo end for comfort. So I asked him if I could see any proof that he was who he said that he was.

He was happy to comply, and once we got to a computer he showed me some scientific papers with his name on top, suitably peppered with complicated words like “tachyon” and “chromodynamics”. He showed me a picture of him winning some prestigious physics prize, dressed in a lab coat with a medal around his neck. He even showed me what looked like a press release: “Professor T comes up with new Theory Of Everything that may explain Higgs Boson”. It was pretty convincing.

But.

A bunch of patients had cancelled on me that day, so I had way too much free time. I started looking into Professor T’s credentials a little more. His papers were hosted on a private site and didn’t show up on Google Scholar and didn’t seem to be affiliated with any journal. The press release was on the same website, and seemed suspiciously badly written. There was nobody else in his photograph, and it was impossible to see what was written on the medal. Was it possible that a crazy hobo had just written some things that looked like papers, written something that looked like a press release, and then bought a lab coat and medal and taken a picture with them? Do real theoretical physicists even wear lab coats?

So I asked him if I could talk to Dr. Witten, whom he said he knew well. Professor T agreed. I Googled his phone number and called him up.

(It wasn’t actually Dr. Witten whom I called, but the case I’m adapting this story from involved someone else about as famous)

His secretary answered the phone, and I said I was a psychiatrist, and I asked if I could speak to Dr. Witten. The secretary was reluctant, but when I said it was about Professor T, she immediately asked me to hold, and I got Dr. Witten himself. I asked him if he knew Professor T.

“Absolutely,” said Dr. Witten. “He’s a crazy guy who keeps calling me up and telling me he’s solved physics. I don’t think he’s actually a professor of anything. I read one of his papers once, just for kicks, and it’s just a bunch of science terms like ‘tachyon’ and ‘chromodynamics’ strung together without rhyme or reason. It might fool a layman, but trust me, it makes no sense. I told him to stop calling me, and he wouldn’t, and finally I had to block him on my phone, and now he’s sending me letters in the mail, and it’s always same ranting about tachyon chromodynamics, which isn’t even a real thing. Did you say you’re a psychiatrist? Perfect, I’m so glad he’s finally getting treatment.”

I told Professor T about this, and he nodded his head. “Yes,” he said “I told you that Dr. Witten knows me well. I didn’t say he liked me. He still doesn’t fully understand my theories. But I am sure he’ll come around.”

I kept Professor T in hospital for about a week, and I can’t count how many times he yelled at me and complained that I was being unfair to him by not doing whatever the heck he wanted me to do that day. Read another one of his papers that would convince me his theories were sound. Call up yet another famous physicist he “knew”. Look at yet another of his fake websites devoted to himself. Every day, he threatened to sue me and my boss and the entire hospital for keeping him there even though he'd "proven" to us he was who he said.

Remember, delusions are *fixed* false beliefs. People are quite sure they're true, quite sure they have evidence for them, and nothing (except occasionally *really* good psychiatric treatment) will convince them otherwise. They'll keep demanding you take time to investigate more and more bizarre "arguments" and "evidence", and if you ever stop, even after days and days of everything they say being one hundred percent refuted, they'll accuse you of acting in bad faith.

(it's like Internet arguments, only more so)

In everyday life, we get by on an assumption of trust. If I tell my boss I'm sick, he probably believes me. If he doesn't believe me, and I send him a doctor's note, he probably believes that the doctor's note isn't forged. If he doesn't believe that, and he asks me for a number to call the doctor at, he probably believes it's a real doctor and not my brother pretending to be a doctor to help me out. Yes, there are a couple of people who abuse that trust, but few enough that the rest of us are usually able to get by.

In psychiatry, there are a bunch of delusional people, paranoid people, narcissists, compulsive liars, and others who deliberately or unknowingly stretch the truth past the breaking point. Worse, a lot of the cues we use to detect liars, like "Are they shifty-looking?" don't work, either because the person involved really believes what they're saying or because they're too far from the neurotypical norm for our usual intuitions to apply. A lot of the assumptions of trust we usually use crash and burn. If the person sitting next to me on the train says he's a physics professor, I believe him. If the person brought in by police for a psychiatric evaluation says he’s a physics professor, maybe I don’t, and “how much time do I spent assessing the evidence and how much do I believe?” is a really tough question.

I am not trained as a police officer, detective, or judge. I’m also not paid to do their jobs. I’m also stuck in a system where the primary incentive is that if I ever fail to commit someone, then if they do anything bad after that I can be sued for everything I own. So I am stuck drawing partial conclusions, from incomplete evidence, in time I don’t have, from people I can’t necessarily trust, without even the ability to err on the side of caution.

I don’t think the hospital in the article followed *great* practices – in particular I’m unclear on how they came to believe the person was unemployed. And sending her the bill for her own involuntary commitment is an obvious injustice (albeit a universally practiced one). But the mistakes in the admission process are all ones I can imagine any psychiatrist making. Including me.

And that means something. You can *trust* me. After all, *four* different Dalai Lama accounts follow me on Twitter.

The Argument From Cultural Evolution

[Content note: Discussion of debate over gay marriage]
[Epistemic status: Very preliminary. Probably missing something. Looking forward to hearing what.]

The term “cultural evolution” is getting used a lot nowadays. In its simplest form, it just means cultures preserve useful ideas and tips. For example, as per Carcinization:

> [The Lost European Explorer] experiment has been repeated many times when European explorers were stranded in an unfamiliar habitat. Despite desperate efforts and ample learning time, these hardy men and women suffered or died because they lacked crucial information about how to adapt to the habitat. The Franklin Expedition of 1846 illustrates this point. Sir John Franklin, a Fellow of the Royal Society and an experienced Arctic traveler, set out to find the Northwest Passage, and spent two ice-bound winters in the Arctic, the second on King William Island. Everyone eventually perished from starvation and scurvy. The Central Inuit have lived around King William Island for at least 700 years. This area is rich in animal resources. Nonetheless, the British explorers starved because they did not have the necessary local knowledge, and despite being endowed with the same cognitive abilities as the Inuit, and having two years to use these abilities, failed to learn the skills necessary to subsist in this habitat.

Presumably, the Inuit neither conducted deliberate centralized experiments to determine what food in their area was edible, nor derived the information from explicit understanding of the principles of nutrition. Rather, over thousands of years, various proposals like “eat those yummy-looking red berries that grow on the small bushes” and “always hunt seals in large groups” were accidentally tested, with the successful ones spreading until they became universal tradition and the disastrous ones being warned against as taboo. Without any deliberate effort, the Inuit ended out with a remarkably effective set of survival techniques.

Something like this seems so obviously true as to not require further discussion.

However, recently “cultural evolution” has slipped, without much consideration, into a much stronger meaning. For example, in his commentary on Ross Douthat’s article on gay marriage, Tumblr user severnayazemlya writes:

> What Douthat is saying is that there was some system that existed sometime in the past that was more human-shaped than Marcotte’s vision for the future. Gavin McInnes has said the same.
>
> The conservative argument is that the cultural inheritance that the past hands down to the present is more human-shaped than most reforms proposed in the present – because there were reformers in the past, and, absent major breaks in the continuity, past reforms have had time to be tested for their fit: those that worked were kept, and those that didn’t were discarded.

Likewise, on one of my recent blog posts, commenter Steve Johnson writes:

> Every surviving cultural tradition on Earth is hostile to homosexuality – that’s no accident. That’s cultural evolution in action.

This form of cultural evolution seems to work something like so: our culture, and indeed most cultures, used to have a certain conception of marriage. That conception of marriage outcompeted other conceptions of marriage from the distant past all the way to the present. Societies with alternative conceptions of marriage seem to have died out. That suggests that this conception contains something useful; even if we can't see it we should be wary of interfering with it, in the same way we are wary to disrupt our body's metabolic balance or alter genes willy-nilly.

The difference between the obvious Inuit form of cultural evolution and the non-obvious marriage form is that of within-culture versus between-cultures evolution.

Consider: one Inuit tries the red berries and discovers they make her sick. Out of pure self-interest, she decides not to eat them again, and tells her friends the same. Also out of self-interest, they decide not to eat them; those who think they can get away with eating them anyway are quickly disabused of the notion. The taboo against eating red berries quickly spreads throughout the culture.

Marriage doesn't seem to work that way. If one person decides not to marry in the usual way, it doesn't necessarily hurt that person. They might have lots of affairs, and enjoy them. Or they might get gay married, and enjoy that. Any claim that cultural evolution argues against gay marriage because it's bad for the actual gay-married person must face the fact that actually gay-married people seem totally okay with it, and in fact are urging their friends to do it, the exact opposite of the red berry situation.

So I interpret it as a different claim: a culture that allows gay marriage will, for various reasons, become weak and unsuccessful. Then it will be crushed by other cultures, either militarily, economically, or in a sort of marketplace of ideas where people convert to or assimilate into the other culture because it's more attractive and successful.

Note that THIS IS REALLY DIFFERENT FROM THE FIRST TYPE OF CULTURAL EVOLUTION. In fact, it might be diametrically opposite. For example, gay sex may be lots of fun – and as people figure this out and tell their friends, it will be positively selected through the first type of cultural evolution. But it might weaken a culture's Moral Fabric – in which case it will be negatively selected through the second type of cultural evolution.

This is sort of group selectionism, but in this case I'm okay with it. Consider the analogy of a cancer cell. Becoming cancerous makes a cell much more likely to spread within its organism – the equivalent of positive intracultural selection – but also makes its organism at a severe disadvantage compared to other organisms – the equivalent of negative intercultural selection. As a result, we expect organisms to evolve strong internal defenses against cancer – which in fact they have. In the same way, it's plausible that cultures might evolve strong internal defenses against actions that are fun but Weaken Moral Fabric, and sure enough we find that everything halfway enjoyable comes with a lecture from our elders about Why We Shouldn't Do It. Presumably if these things really Weaken Moral Fabric in an important way, then those cultures that develop strong internal defenses against them – for example, a strong and well-enforced religious taboo against gay marriage – will be more likely succeed while other cultures die out.

So in principle this kind of intercultural selection could happen. *In practice* I think the effect is negligible.

Evolutionary biology has a lot of equations to calculate how long it will take a positively-selected trait to spread. It depends on a lot of different things, but the most salient here are the length of a generation for the affected organism, and the extent of the selective advantage conferred by the trait.

How long is a "generation" in cultural evolution? Rome lasted a thousand years, Byzantium another thousand. It took about three hundred years for Christianity to replace paganism in Rome; Enlightenment values have been replacing Christianity for three hundred years already and aren't nearly done. Any sort of evolutionary process that involves waiting for Rome to fall is a process that will take way longer than human history to come to any sort of conclusion.

How much advantage can an individual cultural trait confer? Probably very small in the grand scheme of things. Compare Judeo-Christian attitudes about sex to Greco-Roman attitudes about sex. One might argue that the Judeo-Christian attitudes are superior, since Christianity did eventually take over Rome. On the other hand, both Greece and Rome took over Israel at various points; various Jewish texts record that during that time a lot of Jews were defecting to Greco-Roman culture and there were precious few defections the other way. It would seem that all of the other differences between Greco-Roman and Judeo-Christian culture – theology, non-sexual mores, geography, technology, philosophy – had a lot more effect than the sexual mores.

This is also unsurprising from a population biology perspective. Suppose that one of my children gets a mutation causing 1% less risk of infectious disease. This is fine, but they might get killed in a car accident before their tenth birthday, or be too ugly to find a partner, or get an infectious disease anyway because 1% less risk isn't really much less risk. If my child survives, and passes her mutation on to millions of other people all with their randomly distributed level of other good and bad genes and good and bad luck, then maybe eventually over thousands of generations, people with the new beneficial mutation will take over from people without it.

But, as mentioned above, we don't *have* thousands of generations for cultural evolution to do *anything*. Communism, which basically took all of the worst ideas in history, combined them together into a package deal, and said "Let's do all of these at once", took almost a century to collapse, and *still* hasn't collapsed in a couple of places. Imagine if instead of Communism happening, twenty different countries had adopted one Communist ideal each and we'd waited to see which ones grew and which ones declined. We'd *still* be waiting, and probably instead of getting any useful information we'd just end up seeing the Rise of China and not being sure whether it was because of their Communist ideal or something else.

The closest thing I can think of to anyone actually gaining useful information out of cultural evolution is the failure of various small communes and social experiments. But once again, these only failed because they tried all the bad ideas at once, and a big part of their failure was intracultural evolution – the people involved noticed they, personally, were poor and unhappy, said "screw this", and went back to non-communal living.

So overall, I think any appeal to intercultural evolution as having proven anything is on very shaky ground. Appeal to intracultural evolution is much more reasonable, but crucially, can't be used to override people's own decisions about whether they’re happy doing something or not. If someone says “I enjoy this, and I’ve been doing it a few years and not noticed any bad consequences, and I suggest you do it as well”, then you’re going to have a hard time arguing against the practice on grounds of cultural evolution.

[**EDIT**: *Actually, this leaves out a possible third kind of cultural evolution, where cultures try good ideas, learn to like them, and stick with them; or try bad ideas, learn to hate them, and stop. For example, China experimented with Maoism, that didn’t work, then experimented with a variant of capitalism and liked it enough to stick with it thus far. Given that something similar is happening on smaller scales (eg experimentation in policing, education, budgeting, etc) all the time, you could eventually end up with a pretty finely-tuned culture. This obviously happens, but it seems loaded to think of this as ‘cultural evolution’ instead of just ‘guess and check’ or ‘learn from history’. The former formulation suggests something illegible to understanding; the latter formulation suggests that it happens by deliberate human responses to bad consequences, and makes it less of an excuse for general conservativism.*

If you say “Let’s be Maoist!” I can say “No, the Chinese tried that, it led to X, Y, and Z consequences, and then they switched to something else and things got better.” The cultural evolution argument for traditional sexuality seems to be trying to argue in the absence of, or in parallel to, such observable historical lessons.

Likewise, there’s a cultural evolution argument that we tried traditional sexuality, that made a lot of people unhappy, and now we’re trying something else. It’s unclear how this is different from the Maoism example in a way that makes *jettisoning Maoism good, but jettisoning traditional sexuality bad.*]

Cultural Evolution 2: Thanks For The Meme-Rise

Some points I gleaned from the comments of yesterday's post:

1. Cultural evolution can happen in cases where a super-innovation allows one culture to conquer or overwhelm all others. For example, agricultural groups were (after a long transition period) eventually able to overwhelm hunter-gatherer groups, even thought for an individual hunting-gathering was probably more enjoyable than agriculture. Likewise, industrialized societies were pretty quickly able to outcompete nonindustrialized societies, and either colonized them or forced them to industrialize in turn to keep up. Both of these seem like clear-cut examples of cultural evolution. But they only work because of a really big fitness advantage; industrial societies are on a whole other level from preindustrialized ones. It doesn't necessarily generalize to saying that small, moderately beneficial ideas will catch on, or slightly detrimental ones be selected against.

2. Cultural evolution can happen when one group in a society outbreeds another. The Amish population has increased twenty times faster than the non-Amish American population in the past century. At a constant growth rate, it'll be only another four hundred years or so before America is an Amish-majority nation. More seriously, some people expect something like this to happen with high-fertility-rate immigrant populations, like Latinos and Muslims. In cases of strong differential fertility rates, cultural evolution becomes a race to see if the faster-growing minority can reproduce faster than the majority can assimilate them. *However*,

despite dire predictions of all of us being crushed under the Amish's quaint hand-made boots, people had trouble thinking of historical examples of something like this happening. Sure, populations have replaced other populations – like the Anglo-Saxons replacing the Celts in England – but it's tended to occur alongside military invasions.

3. Cultural evolution can happen with units smaller than Rome-sized grand civilizations. Several people brought up subcultures, like hipsters and Goths, and noted that these have "generations" on the order of a few decades, and thus could potentially undergo evolution conforming to population genetic equations in a reasonable amount of time. Because they're larger units than just a single person, their "evolution" could select for things that bind people together, like rituals and cohesion-building symbology and so on, and be more interesting than just individual memetics. They could also spread very quickly as people rush to join the attractive ones. Okay. But subcultures like Goths seem like a very modern phenomenon, and I can't think of ancient examples of, for example, a subculture that became popular and spread and became dominant/universal. Religions are the closest thing here, but they have lifespans measured in centuries and don't seem to be a big improvement over waiting for the Fall of Rome.

4. Cultural evolution can occur by an accretion of things that work. For example, the first rituals might have been impromptu celebrations of specific events, but because they helped people bond, people kept doing them. But this seems to require some human intelligence to notice "Hey, we seem to be bonding better ever since we implemented that ritual, let's keep doing it". Without that, it collapses back to the sort of intercultural evolution where the culture is 1% better and after thousands of cultural generations

lasting millennia each it outcompetes others. That makes it unsatisfying for people who want to use cultural evolution as a grounding for Chesterton's Fence, ie "we don't know why we do this, but we ought to keep on doing it."

5. Cultural evolution could have occurred way way back in prehistory. There seem to be about 50,000 years of prehistory, there were many more cultures back then, and maybe cultural generations were shorter – for all anybody knows, clans could have disintegrated and reformed over the space of decades. That provides enough generation time for cultural evolution to work. Question is, can we trust anything that evolved in pre-history – when the pressing social issues of the day were things like "How do we not get eaten by bears?" – to still be relevant?

There does seem to be the *potential* for cultural evolution to be interesting, but I'm still not seeing it as a strong argument for preserving particular features of inherited culture absent other arguments suggesting we know why we want those things to be preserved.

[**Edit:** An alternative ontology]

Things That Sometimes Work If You Have Anxiety

Anxiety disorders are the most common class of psychiatric disorders. Their US prevalence is about 20%. They're also among the least recognized and least treated. We have sort of finally beaten into people's thick skulls that depression isn't just being sad, and you can't just turn your frown upside down or something – but the most common response to anxiety disorders is still "Anxiety? So what, everyone gets that sometimes."

But it's hard to describe how disabling anxiety can be. A lot of people with nominally much worse conditions – depression, bipolar, even psychosis – will insist that they want their anxiety treated before anything else, because they can live with the rest. On the other hand, while a lot of people with psychosis have enough other problems that treating the psychosis barely puts a dent in their issues, a lot of people with anxiety would be happy and productive if they could just do something about it.

Since I've gotten some positive comments on my discussion of depression treatments I thought I'd go through some of the things I've seen used to treat anxiety. I'll include the same disclaimer:

> This will be inferior to reading official suggestions, but you will probably not read official suggestions, and you may read this. All opinions here are my own, they are not endorsed by the hospital I work at, they do not constitute medical advice, I have a known habit of being too intrigued by extremely weird experimental ideas for my own good, and you read this at your own risk. I am still a resident (new doctor) and my knowledge is still

> very slim compared to more experienced professionals. Overall this is more of a starting point for your own research rather than something I would expect people to have good results following exactly as written.

I'll mostly be talking about what's called generalized anxiety disorder, with some applicability to panic disorder. Social anxiety, specific phobias, et cetera are their own thing, as is anxiety secondary to other illnesses – but some of the advice may cross over. I'm not going to get too into diagnosis, because generalized anxiety disorder is pretty much exactly what you think it is and a lot (though not all) of this will be applicable for subclinical anxiety as well.

I. Diet And Lifestyle

You didn't think you were going to get out of this part, did you?

Pretty much every study – epidemiological or experimental, short-term or long-term, has shown that exercise decreases anxiety. The effect seems limited to aerobic exercise like walking, running or swimming, preferably for longer than twenty minutes. Various mechanisms have been postulated including norepinephrine, endogenous opioids, and decreased inflammation.

There's less agreement on diet. The people who hate fat says high-fat diets cause anxiety. The people who hate carbs say high-carb diets cause anxiety. The people who hate processed food say processed foods cause anxiety. The people who recommend fish oil for everything say insufficient fish oil causes anxiety. None of it seems super credible, but Mayo Clinic has some suitably bland advice.

The one very important connection – if you drink too much coffee, or any other source of caffeine, that will make you anxious. I once had a patient come to me with severe recurrent anxiety. I asked her how much coffee she drank, and she said about twenty cups per day. Suffice it to say this was not a Dr. House-caliber medical mystery.

Also needless to say: get enough sleep. Seriously. Get enough sleep.

Many people find that various breathing exercises or other sorts of mindfulness activities can be helpful in the short term and sometimes build skills useful for the long term. My hospital gives people these handouts on breathing techniques and progressive muscle relaxation. I’ve made fun of HeartMath in the past, but I only learned about them because many people find some success, probably placebo-ish, with their quick coherence technique. If you’re an overachiever and want to get really into this sort of stuff, people always say good things about yoga and especially pranayama breathing. Studies seem to back this up (1, 2, 3) though you’ve got to be careful to weed out the studies by very religious Hindus trying to prove they’ve been right all along.

Meditation has similarly positive results. Here’s a study showing that an intervention to teach patients meditation resulted in decreased anxiety with $p < 0.001$ even three years later. Here's a meta-analysis of 39 studies finding an effect size of about 0.6 (medium) in the general population, and an effect size of about 1.0 (large) in people with anxiety disorders. But here’s an equal and opposite review that found only “equivocal” results. As far as I can tell, most people investigating meditation think it works pretty well. The meditation techniques that seem to work best are mindfulness meditation and transcendental meditation. You can learn a little about mindfulness meditation here. In order to learn about Transcendental Meditation, send a check made out for $5000 to Maharishi Mahesh Yogi, PO Box….

II. Therapy

Cognitive-behavioral therapy works okay for anxiety just like it works okay for everything else. The Big Graph O' Effect Sizes says that psychotherapy on average has an effect size of 0.51 in generalized anxiety, compared to medication's 0.31. This shouldn't be taken too seriously – the confidence intervals overlap and there's a wide range of efficacy for different medications – but you won't be doing any *worse* by going for the therapy first. Even the *Cochrane Review*, famous for never drawing any conclusion other than "more research is needed", is tentatively willing to say that psychotherapy works for anxiety disorders. Their study *trends* towards finding that cognitive behavioral therapy works better than supportive therapy, but is unable to prove significance – apparently more research is needed.

Exposure therapy can also be useful for panic attacks or specific phobias. This is where they expose you to the thing you're scared of (or deliberately initiate a panic attack) and keep doing it until you stop being scared and start being bored. According to a bunch of studies it works neither better nor worse than cognitive-behvioral therapy for most things, but my unsupported impression has always been that it's better at least for panic disorder. Cognitive-behavioral therapy seems clearly superior for social phobia.

You can get psychotherapy from any qualified psychotherapist, a category including counselors, social workers, psychologists, and sometimes psychiatrists. Ones who use "a school" (for example, describe themselves as practicing cognitive behavioral therapy) are usually considered better than those who don't ("Oh, I do a little of everything with every patient”). If you can’t find (or don’t want to find) a good therapist, there is preliminary evidence that a good self-help therapy workbook (“bibliotherapy”) is about as good as real therapy – including for anxiety (study, other study, yet another study).

I have no special insight into which self-help workbooks are any good, but The Cognitive Behavioral Workbook for Anxiety: A Step-By-Step Program seems to get pretty good ratings.

III. Medications

To be tried *after* diet and lifestyle interventions when possible.

Medication can work either instead of or in addition to therapy. There are at least seven categories of commonly used conventional anxiety medications: SSRIs, SNRIs, antihistamines, antipsychotics, anticonvulsants, benzodiazepines, and azapirones. These can be divided into mostly-acute (antihistamines and benzos) and mostly-long-term (SSRIs, SNRIs, anticonvulsants, azathioprines), with antipsychotics kind of being a tossup. Depending on whether you just need to get through the occasional panic attack or whether you’re in a chronic unremitting anxiety state, you might want one, the other, or both.

You probably know antihistamines (example: Benadryl) from the many common over-the-counter members of this class. They have some mild short-term anti-anxiety effects. Benadryl will work in a pinch if you need something without a prescription, but the most commonly used anxiolytic antihistamine is hydroxyzine (“Vistaril”, “Atarax”), which is a bit more powerful and less likely to make you fall asleep. As far as anxiolytics go it’s pretty safe as long as it doesn’t make you too sleepy. If you just need something to take the edge off the occasional anxiety attack, this works fine.

Benzodiazepines (examples: Xanax, Ativan, Valium, Klonopin) are very effective in the short-term but also very controversial. In some people they are very habit-forming and can produce a picture very similar to addiction to alcohol (which they chemically resemble). Keep in mind how bad an idea it might be to become extremely addicted to prescription pills that you may suddenly lose access to depending on how your doctor is feeling (you might expect doctors would take the difficulty of coming off these drugs into account, but you might expect a lot of things from doctors that don't always happen). Studies suggest benzodiazepines can sometimes build tolerance, and that after a month or two of frequent use, they lose their positive effect and you need them just to feel normal. That having been said, a subset of patients – and I can't tell at this point if it's a majority or a minority – go on benzodiazepines, do very well, stay on them for long periods without getting dependent, and never have anxiety again. It's kind of a crapshoot. The most generally recognized "safe" use of benzos is the occasional Xanax to deal with rare but very stressful situations (for example, flying on an airplane if you're scared of heights). Other people say Klonopin is safer than some of the others and that it's worth a shot as long as you realize that "Klonopin dose gradually creeping upwards" is a sign that you're getting into a bad place and need to react immediately. Most people recommend trying other things first before you come here, but once you've exhausted other options these can be a powerful last resort.

SSRIs (examples: Prozac, Celexa, Lexapro, Zoloft) are the mainstay of chronic anxiety treatment just like they're the mainstay of chronic everything-else treatment. As usual, they have real but modest effects after about a month or so, more in some people and less in others. As usual, if one SSRI doesn’t work for you, you might want to try another. These are pretty safe aside from the sexual side effects. Some people get mild withdrawals if they go off these too quickly, so don’t do that. A lot of people use both an SSRI for chronic treatment, plus either an antihistamine or benzo for “break-through” anxiety.

SNRIs (examples: Effexor, Cymbalta) are like SSRIs, but for two neurotransmitters instead of one. This is supposed to make them a little bit more effective. Maybe they are, maybe they aren’t. Fewer sexual problems than SSRIs, but worse discontinuation syndrome. They’re a good second-line chronic medication if SSRIs don’t work. Effexor is probably the best.

Azapirones (example: BuSpar) is, unusually, a rare drug which is specifically targeted at anxiety, rather than a being a repurposed antidepressant or something. BuSpar is very safe, not at all addictive, and rarely works. Every so often somebody comes out with a very cheerful study saying something like “Buspar just as effective as benzodiazepines if given correctly!” and everybody laughs hysterically and goes back to never thinking about it.

Anticonvulsants (examples: Depakote, Neurontin, Tegretol, Lyrica) are seizure medications that sometimes sort of work for anxiety. Most of them have strong side effects and limited utility. The exception is Lyrica (pregabalin), which is pretty new but has shown excellent safety and efficacy in studies. It doesn’t have an FDA indication for anxiety and it’s pretty expensive, so you might have a hard time getting it, but it is at least a well-kept secret.

Atypical antipsychotics (examples: Seroquel, Zyprexa, Abilify, Geodon) are, as always, overused. Most of them either make you gain lots of weight, put you at increased risk for heart rhythm problems, make you feel terrible, put you at risk of permanent movement disorders, or all of the above. They *do* often treat anxiety, sometimes very well, and psychiatrists like them because they're good all-purpose no-nonsense drugs with big advertising budgets, but unless you're also psychotic consider trying some other things first before you try these.

An article in Journal of Psychopharmacology tries to compare the efficacy of all of these classes of drugs and gets the following effect sizes (bigger number = bigger effect):

Pregabalin: 0.5
Antihistamine: 0.45
SNRI: 0.42
Benzo: 0.38
SSRI: 0.36
Azapirone: 0.17
Alternative medicine: -0.31

(remember, other studies suggest psychotherapy is around 0.5)

I *heavily* challenge the claim that antihistamines are more effect than (or anywhere near as effective as) benzos. I don't know the confidence intervals on these numbers, so I would suggest reading it as "Everything is about equally effective, except azapirones which aren't as good". Their "alternative medicine" category was mostly kava and homeopathy, and I have no idea why it came out negative (kava's pretty good, and homeopathy shouldn't separate from 0).

There are also some less commonly used drugs that might help people who don't respond to any of these.

As usual, MAOIs are very effective, moderately dangerous, and super hard to get. They seem to work especially well for panic disorder and social anxiety.

Clonidine is a medication usually used to control blood pressure. It's somewhat effective against anxiety and some people think it should be used more. But it can cause you to become too sedated (abnormally low heart rate) and in some people it makes anxiety worse for some reason.

Beta-blockers (example: propranalol) are another blood pressure medication. It is especially effective against somatic symptoms of anxiety – racing heartbeat, shaking, et cetera – and sometimes getting rid of those can make the anxiety go away entirely. It's most famous for its use against performance anxiety: about a third of musicians use them in concerts, and I've heard similar rumors about public speakers, actors, et cetera. I used to think this was a little-known piece of trivia, but whenever I bring it up to doctors ("Hey, did you know some people use beta-blockers for performance anxiety") the usual response is "Oh, yeah, I prescribe myself some of that when I have to give a presentation at grand rounds." They don't seem quite as good for longer-term anxiety disorders, though some people have had good results with them.

I once saw an excellent psychiatrist whom I deeply respect try *everything* on a patient with severe treatment-resistant anxiety with no results whatsoever until finally he came to Thorazine. This treated the *patient's* anxiety pretty well, at the cost of provoking quite a bit of anxiety in the doctor.

Without meaning to give medical advice, and with the caveat that you should ask your doctor for their opinion – one good pharmacological treatment algorithm for anxiety disorders is:

If you just have occasional outbursts that bother you, take occasional doses of hydroxyzine.

If you have a longer-term problem, start with an SSRI. If that doesn't work, either try more SSRIs and SNRIs, or go to Lyrica. You might as well be on BuSpar somewhere in the process too. If none of that works, choose your poison (or have it chosen for you) among MAOIs, benzos, clonidine, or antipsychotics.

IV. Alternative Treatments

To be used out of curiosity or desperation only – you have other options and these are not guaranteed safe or effective.

Massage therapy, acupuncture, aromatherapy, and everything else in the category of "unnecessarily medicalized relaxing thing" all perform very well as long as you don't look too hard for a suitable control group. Yes, these are probably placebo, but they're very effective placebos and if they both work I would rather take a placebo than an antipsychotic.

Inositol and l-theanine are both found in small quantities in the diet (inositol in some vegetables, theanine in tea) and supplementing them has been inconsistently found to help with anxiety. Inositol had some preliminary evidence for effectiveness in panic disorder, but a more recent meta-analysis was unimpressive. I can only say that I have some anecdotal evidence of extremely positive reactions to inositol, but we all know what they say about anecdotal evidence. Keep in mind that the dose used in studies is way larger than the dose anyone will give you – usually corresponding to about 20 of those 500 mg inositol pills a day. This makes it expensive and inconvenient, and most people just compromise by taking so little inositol it shouldn't possibly be able to have any effect. L-theanine also has a lot of small studies in support, although there's some question on whether it works on its own or whether it just has useful synergistic effects with caffeine. Suntheanine is generally considered the most effective form, and recommended dose is about 100 – 400 mg. Both these supplements are afaik very safe and a good option for people who want to test things that might or might not work but have minimal risk. Magnesium should also be in here somewhere.

GABA is the main inhibitory neurotransmitter in the nervous system, and a lot of these other interventions are attempts to convince the brain to release more GABA or potentiate the GABA that's already released. Can we just cut out the middleman and ingest GABA pills directly? The supplement industry would like you to think so, and you can certainly buy them anywhere supplements are sold, but it's generally believed that orally ingested GABA can't cross the blood-brain barrier. The Russians have developed a modified version of GABA that doesn't have this problem; called picamilon, it seems to be a pretty popular anxiety treatment on the other side of the Pharmacological Iron Curtain. It's pretty easy to get as a non-prescription supplement here in the West. There are very few studies on it, the ones that exist are in Russian, and I have nothing to go on but a couple of anecdotal reports, most of which are positive (though I personally noticed no effects). But the mechanism of action is plausible, and the long history of successful Russian use at least suggests it probably won't kill you immediately. Most common dosage seems to be about 100 – 300 mg.

The nootropics/supplement/nutraceutical community also suggest ashwagandha and bacopa for anxiety; various low-quality studies support the use of both (ashwagandha meta-analysis, bacopa study 1, bacopa study 2, bacopa study 3). Bacopa may take several months of frequent use before it starts working; I tried it briefly and had to stop because of gastrointestinal side effects, which are pretty common. There's also some worry around heavy metal contamination. Swanson's and Nootropic Depot's are two that have third-party testing showing they're uncontaminated.

Kava is a traditional drink from various Pacific islands with anxiolytic properties. Multiple meta-analyses including a Cochrane review find it to be an effective anxiety treatment, but its safety is in question after reports of several cases of liver failure caused by the plant. This may be yet another case of people exaggerating freakishly rare side effects; the risk has been estimated at less than one in a million doses (though remember that if you take it daily for ten years, that number bcomes 1/300). Others suggest a rate as low as one in a hundred million but this assumes zero underreporting; others challenge this assumption. Possibly it is only poorly prepared kava causes liver problems; for traditionally prepared kava, look for preparations that specify they are made from root/rhizome material only. The American Academy of Family Physicians recommends that:

> Physicians who supervise patients taking kava for the treatment of GAD should take care to avoid the following: (1) high dosages (more than 300 mg per day); (2) combining kava with hepatoactive agents; (3) using non-root preparations; and (4) exposure for longer than 24 weeks. Use of WS1490 standardized kava extract is

> also recommended. If these safety precautions are followed, kava can be appropriate therapy for selected patients diagnosed with GAD

Don’t take kava if you have any liver problems, if you’re on any medications that might interact with it, or if you plan on drinking alcohol at the same time. Consider talking about it with your doctor first and getting plans to check liver enzymes regularly.

Selank is an experimental Russian anti-anxiety medication going through their version of clinical trials. It’s a bit high-maintenance – you have to keep it refrigerated or else it decays, and the only two functional means of administration are injection or nasal spray – but anecdotal evidence is extraordinarily positive. No side effects have been found thus far, but needless to say by the time you get to “injecting experimental Russian medications into yourself” we have left the point where we can entirely guarantee this is a good idea. Ceretropic sells a nasal spray version, which is probably more convenient than having to inject it.

Phenibut is another Russian anti-anxiety medication. It is potentially addictive and dangerous. I do not want to actively recommend against it, because it can be very useful if used *infrequently and carefully*. Discussing exactly how to use it infrequently and carefully is beyond the scope of this article. Please do not use this unless you have looked into it carefully and understand the risks and benefits.

Overall, the best evidence seems to be for l-theanine (especially if you drink coffee) and bacopa (especially if you’re willing to wait months for any effect), with picamilon also worth your time to try and Selank as an option for the very adventurous.

V. Conclusions

No treatment stands out as extremely effective, and the best route to dealing with anxiety probably depends on many factors like your amount of free time, your motivation, your access to medical care, and your willingness to put up with side effects. After you've fixed lifestyle issues, I think any of "self-help workbook", "start SSRIs", or "try l-theanine" are good first options. On the other hand, benzodiazepines, antipsychotics, and kava are all options I would hold off on until you've tried a couple of other things.

Like with the depression post, the most important conclusion you can take from this is that *you have lots of options*. Please don't let people give you an SSRI and then give up. Work with your doctor. Anxiety actually has a pretty good prognosis if people work on it, but it can be a difficult and frustrating process. Just remember: *there are lots of options*.

PS: Relevant Onion

CBT In The Water Supply

[Epistemic status: Very speculative, <50% confidence, thinking out loud. Don't let this turn you off therapy.]

Here's a vignette from cognitive-behavioral therapy book When Panic Attacks, heavily edited for length:

> A chronically anxious medical school professor named Nate suffered from low-self-esteem and feelings of inadequacy. One day, Nate brought me a copy of his CV. I was blown away. He'd listed over sixty pages of research publications, prestigious awards, and keynote addresses he'd given at major conferences around the world. I asked Nate how he reconciled his low self-esteem with all of his accomplishments. He said that every time he looked at his CV, he felt discouraged and told himself that his colleagues' research studies were far more rigorous and important than his own. He said his paper seemed "soft" and consisted primarily of theoretical work, rather than hard-core laboratory research with real tissue. He said "Dr. Burns, no matter how much I accomplish, it never seems good enough."
>
> Perfectionism was clearly one of Nate's self-defeating beliefs. I suggested that Nate use the Pleasure/Perfection Balance Worksheet to test this belief. I told him to write "If I can't do something perfectly, it's not worth doing at all" on the top of the sheet, and asked him to list several activities in the left-hand column. I told him to predict how satisfying and rewarding each activity would be, to record how satisfying and rewarding it was afterwards, and to rate how perfectly he did each activity. That way he could

find out of it was true that he only enjoyed the things he did perfectly.

The next week, Nate had some interesting results to share with me. One of his activities was giving the welcoming lecture ot the incoming class of medical students. Nate gave this lecture every year because he was considered to be the most charismatic speaker at the medical school. Nate predicted this lecture would be 70% satisfying, but his actual satisfaction as only 20%. This was surprising, since he'd received a thirty-second standing obation, and he'd rated his perfection level for the talk at 90%.

I asked Nate why his satisfaction rating was so low. He explained that he always got standing ovations, so he routinely timed them. The previous year, the medical students had stood and cheered for more than a minute at the end of his talk. This year, the only stood and cheered for half a minute. Nate felt disappointed and started worrying that he was over the hill.

The second entry on Nate's Pleasure/Perfection Balance Worksheet was that [he fixed a broken pipe in his bathroom]. He had to make several trips to the hardware story to buy tools and parts and to get tips on how to do it, so he didn't get the pipe fixed until 10 PM. How explained that any plumber could have fixed the pipe in five minutes, so he rated his perfection as 5%. But his satisfaction level for this activity was 100%. In fact, he felt exhilarated. Nate said it was the most satisfying thing he'd done in years.

The result of Nate's experiment was not consistent with his belief that things weren't worth doing unless he did them perfectly. It dawned on him that there were many

> sources of satisfaction in his life that he’d overlooked, such as taking a walk through the woods with his wife, even though neither of them were world-class hikers, playing squash with his son, even though neither of them were champions, or just going out with his family for ice cream cones on a warm summer evening.
>
> This experiment had a significant impact on Nate’s feelings of self-esteem and on his career. He told me that his feelings of anxiety and inferiority decreased, and his productivity actually increased because he was no longer so worried about having to do everything so perfectly.

At first I assumed this story was made up, but the book claims these are based on real patients, and even mentions how the writer showed videos of some of these therapy sessions to his classes. Interesting. How about another?

> Several years ago, I did a three-day intensive workshop for a small group of psychotherapists in Florida. A marriage and family therapist named Walter explained that he’d been struggling with anxiety and depression for several months because Paul, the man he’d lived with for eight years, had found a new lover and left him. He put his hand on his chest and said: “It feels real heavy, right here. There’s just a sense of loneliness and emptiness about the whole experience. It feels so universal and final. I feel like this pain is going to go on forever, until the end of time.”
>
> I asked Walter how he was thinking and feeling about the breakup with Paul. What was he telling himself? He saidL “I feel incredibly guilty and ashamed, and it seems like it must have been my fault. Maybe I wasn’t skillful enough, attractive enough, or dynamic enough. Maybe I

wasn’t there for him emotionally. I feel like I must have screwed up. Sometimes I feel like a total fraud. Here I am, a marriage and family therapist, and my own relationship didn’t even work out. I feel like a loser. A really, really big loser.”

Walter recorded these five negative thoughts on his daily mood log:

1. I’ll never be in a loving relationship again
2. I must be impossible to live with and impossible to be in a relationship with
3. There must be something wrong with me
4. I totally screwed up and flushed my life down the toilet
5. I’ll end up as an old, fat, gray-haired, lonely gay man

He believed all of these thoughts very strongly.

You can see that most of Walter’s suffering results from the illogical way he’s thinking about the rejection. You could even say that Walter is treating himself far more harshly than Paul did. I thought the Double Standard Technique might help because Walter seemed to be a warm and compassionate individual. I asked wehat he’d say to a dear friend who’d been rejected by someone he’d been living with for eight years. I said “Would you tell him that there’s something wrong with him, that he screwed up his life and flushed it down the toilet for good?”

Walter looked shocked and said he’d *never* say something like that to a friend. I suggested we try a role-playing exercise so that he could tell me what he would say to a friend who was in the same predicament […]

Therapist (role-playing patient’s friend): Walter, there’s another angle I haven’t told you about. What you don’t understand is that I’m impossible to live with and be in a relationship with. That’s the real reason I feel so bad, and that’s why I’ll be alone for the rest of my life.

Patient (role-playing as if therapist is his friend who just had a bad breakup): Gosh, I’m surprised to hear you say that, because I’ve known you for a long time and never felt that way about you. In fact, you’ve always been warm and open, and a loyal friend. How in the world did you come to the conclusion that you were impossible to be in a relationship with?

Therapist (continuing role-play): Well, my relationship with [my boyfriend] fell apart. Doesn’t that prove I’m impossible to be in a relationship with?

Patient (continuing role-play): In all honesty, what your’e saying doesn’t make a lot of sense. In the first place, your boyfriend was also involved in the relationship. It takes two to tango. And in the second place, you were involved in a reasonably successful relationship with him for eight years. So how can you claim that you’re impossible to live with?

Therapist (continuing role-play:) Let me make sure I’ve got this right. You’re saying that I was in a reasonably successful relationship for eight years, so it doesn’t make much sense to say that I’m impossible to live with or impossible to be in a relationship with?

Patient (continuing-role-play:) You’ve got it. Crystal clear.

> At that point, Walter’s face lit up, as if a lightbulb had suddenly turned on in his brain, and we both started laughing. His negative thoughts suddenly seemed absurd to him, and there was an immediate shift in his mood…after Walter put the lie to his negative thoughts, I asked him to rate how he was feeling again. His feeling of sadness fell all the way fromj 80% to 20%. His felings of guilt, shame, and anxiety fell all the way to 10%, and his feelings of hopelessness dropped to 5%. The feelings of loneliness, embarassment, frustration, and anger disappeared completely.

The book is quite long, and it’s full of stories like this. The author, who’s one of the top cognitive-behavioral psychiatrists in the world, describes his experience with the therapy as:

> [When I first learned about this therapy, I thought] depression and anxiety seemed far too serious and severe for such a simplistic approach. But when I tried these methods with some of my more difficult patients, my perceptions changed. Patients who’d felt hopeless, worthless, and desperate began to recover. At first, it was hard to believe that the techniques were working, but I could not deny the fact that when my patients learned to put the lie to their negative thoughts, they began to improve. Sometimes they recovered right before my eyes during sessions. Patients who’d felt demoralized and hopeless for years suddenly turned the corner on their problems. I can still recall an elderly French woman who’d been bitterly depressed for more than fifty years, with three nearly-successful suicide attempts, who started shouting “Joie de vivre! Joie de vivre!” (“joy of living”) one day in my office. These experiences made such a strong impact on me that I decided my calling was in clinical work rather than brain

> research. After considerable soul-searching, I decided to give up my research career and become a full-time clinician. Over the years, I’ve had more than 35,000 psychotherapy sessions with depressed and anxious patients, and I’m every bit as enthusiastic about CBT as when I first began learning about it.

Okay. I am not one of the top cognitive-behavioral therapists in the world. I’ve been studying formal cognitive-behavioral therapy for about a week now, and been doing untrained ad hoc therapy on inpatients for a couple years. But I’ve also gotten to observe a lot of other people doing therapy, and talked to people who have had therapy, and treated patients who were simultaneously undergoing therapy, and the impression I got was very different.

Dr. Burns asks patients to question whether their anxiety and their negative thoughts are rational, and their faces light up and all of their psychiatric problems suddenly melt away.

The therapists I’ve seen ask patients to question whether their anxiety and their negative thoughts are rational, ever so tactfully, and the patients say “No shit, Sherlock, of course they aren’t, but just knowing that doesn’t help or make them go away, and I’ve been through this same spiel with like thirty people already. Now shut up and give me my Xanax.”

In my last post, someone asked what to do if they found cognitive-behavioral therapy hokey and patronizing. I said, only half joking, that “if you don’t like hokey patronizing things, CBT may not be for you.” I know it’s mean, and pessimistic, but everyone I’ve talked to has had pretty much the same experience. I used to attribute this to my friends being pretty smart, and maybe CBT was aimed as less intelligent people, but Nate The Genius Medical School Professor seems pretty smart. So does Walter The Therapist. Burns’ book includes a bunch of other vignettes about high-powered lawyers, graduate students, et cetera. They all find his suggestions of “Well, have you considered that your irrational negative thoughts might not be rational?” super life-changing.

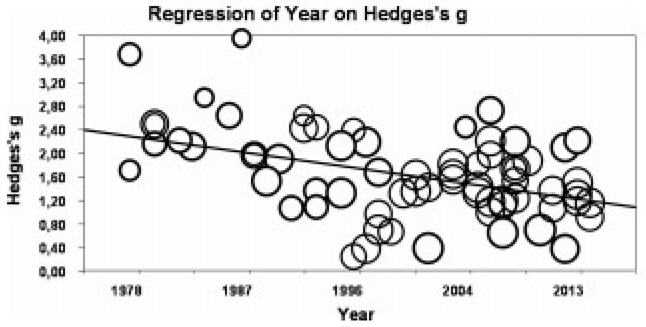


Figure 4. The plot portrays the negative change ($p < .001$) in Beck Depression Inventory effect sizes across time ($k = 61$). The size of the circles indicates the relative contribution (random weight) of each study to the analysis.

You might have read the study this graph comes from: The Effects of Cognitive-Behavioral Therapy As An Anti-Depressive Treatment Is Falling: A Meta-Analysis. As you can see, the Hedges’ g declined from about 2.5 in 1980 to around 1 today. The latest embarrassing set of results now show CBT doing no better than its old nemesis psychoanalysis. Why?

There are a lot of possible explanations. The smart money is always on “it never worked very well, but we’re finally doing studies that aren’t hopelessly biased”, but the analysis doesn’t find a clear difference in study quality. Other suggestions are that therapists have gotten less committed over time, or that the patient populations has changed. All of these sound reasonable. But let me mention one more possibility.

Every so often, psychiatrists joke about how so many people are depressed we might as well put Prozac in the water supply. Sometimes we say the same thing about lithium, although in that case we’re not joking.

Nobody’s ever talked about putting cognitive-behavioral therapy in the water supply, but insofar as that’s meaningful at all I would say we’ve kind of done it. Cognitive-behavioral ideas, like perfectionism, excessive self-blame, conditional versus unconditional self-respect, deep breathing, goal-setting, et cetera have become basic parts of popular culture. The whole self-esteem movement isn’t *exactly* cognitive-behavioral, but it’s certainly allied, and it certainly represents a shift to a style of thinking about the self and about psychology in a way that’s much more fertile for cognitive-behavioral ideas. Inside Out was *kind of* “Cognitive Behavioral Therapy: The Movie”.

Although the particular book I’m reading is from 2006, Burns himself was one of Aaron Beck’s original students and one of the first cognitive-behavioral therapists ever. I wonder how many of these patients who seem absolutely *shocked* to realize that maybe their anxiety isn’t rational come from that very early period.

It’s very hard to track changes in people’s basic beliefs about psychology. I was flabbergasted to learn that until Dr. Benjamin Spock’s landmark 1940s book on child care, parents were told not to hug, kiss, or show affection to babies, because that would coddle them and make them weak, pampered adults. Before that, parents interacted with their kids much less, and it was assumed that siblings and nannies and friends would raise them, or they would raise themselves. It’s easy to read books about ancient Greece and not notice that they have a completely different view of the role of the self/individual than we do. So it wouldn’t surprise me if a lot of the psychology we consider “obvious” is CBT that has seeped out into the water supply over the past thirty years.

If that were true, it would explain why CBT is no longer as effective – it’s just telling people things they already know.

It could be fairly asked: then why isn’t everybody *already* better? Depression seems to be increasing, though there’s a lot of argument about exactly how much; that doesn’t sound like what would happen if everyone were automatically getting a background level of therapy.

Here’s a theory, though it’s on even shakier ground than the other one. The meta-analysis proposes that CBT may have lost some placebo effect over time because patients no longer think of it as The Exciting New Thing. I’m not sure I can go along with that – my own analysis of psychotropic medications suggests patients very much prefer the *old* ones for some reason. But a big part of psychotherapy is placebo effect, so they might be on to something.

What part of psychotherapy provides the placebo? Is it going to the clinic? Talking to the therapist? Hearing fancy words like “self-estimation”? Doing worksheets?

One thing a lot of therapies have in common is that they provide the feeling of insights. For example, psychoanalysts are very good at coming up with surprising-but-plausible ways that your current problems are linked to things that happened to you as a child; the usual result is a patient feeling enlightened, like “You’re right, the leg pain that’s been bothering me *is* in the same part of my leg that accidentally brushed up against my mother’s breast one time when I was seven, that’s pretty interesting.”

Suppose that in the old days, CBT was an insight a minute and you were constantly hearing surprising things you'd never thought about before. And nowadays, you're kind of absorbing a lot of those things by osmosis without it seeming too insightful, and then the therapy itself is anticlimactic. Could that lessen the placebo effect enough to account for the data?

I don't know. Maybe after I've been training in formal CBT for more than a week, I'll have more data and can report back to you.

[***EDIT:*** *Sarah writes: "In a way, seeing CBT stuff in pop culture inoculates people, I think. People will get as far as noticing "this negative thought is an anxiety symptom", but not as far as *actually reversing it*. When people hadn't heard of CBT, they first got the "this negative thought is irrational" message in a context when they were actively working on their problems, so they followed through with the 'hard' step of actually reversing the thought. Now, people run into the revelation that the 'inner critic' is wrong just by browsing facebook, when they're *not* actively trying to fight their anxiety problems, so the revelation loses its force."*]

[***EDIT 2:*** *Paul Crowley points out a very similar theory in The Guardian*]

Reverse Psychology

[Content warning: suicide]

I.

It all started when I made that phone call.

I was really bad. All the tenure-track positions I'd applied to had politely declined, and I saw my future in academia gradually slipping away from me. Then the night before, my boyfriend had said he thought maybe we should start seeing other people. I didn't even know if we were broken up or not, and at that point I couldn't bring myself to care. I sat on my bed, thinking about things for a while, and finally I called the suicide hotline.

"Hello?" a woman's voice answered on the other side. Somehow, just hearing someone else made me feel about five times better.

"Hello," I said, a little more confidently. "I've been thinking of committing suicide. I need help."

"Okay," she said. "Is there a gun in your house?"

"No."

"All right. The first thing you need to do is get one. Overdosing on pills is common, but it almost never works. You can get a firearm at almost any large sporting goods store, but if there aren't any near you, we can start talking about maybe jumping from a high…"

"What the HELL?" I interrupted, suddenly way more angry than depressed. "You're supposed to @#!$ing tell me not to do it!”

“This is the suicide hotline,” the woman said, now sounding confused. Then, “Are you sure you weren’t thinking of the suicide *prevention* hotline?”

“Give me a break! I took a psychology class in undergrad, I know what a suicide hotline is!”

“I’m sorry you seem to be upset. But this is the suicide hotline. It’s like how there’s the Walk For Breast Cancer, but also the Walk Against Breast Cancer.”

“There’s the what? But…I was *in* the Walk For Breast Cancer! I thought…”

“It sounds like you have some issues,” said the woman, politely.

“Ugh,” I said. “Yeah.”

“Do you feel like you need professional help?”

“Yeah.”

“I do have a free clinic with an opening available tomorrow at three PM, would you like me to slot you in for an appointment?”

So you’re probably wondering why in the world I would take an appointment arranged by the suicide hotline that wasn’t a suicide prevention hotline. The answer is – were you even listening? A free clinic? With an appointment available the next day? Normally I was lucky if I found a place with an opening in less than two months and a co-pay that wasn’t completely ruinious. You *bet* I was taking that appointment before someone else snatched it up.

Dr. Trauer’s office looked gratifyingly normal. There was a houseplant, a diagram of the cranial nerves, some Abilify® merchandise, and on the wall one of those Magic Eye stereographic images that resolved into a 3D picture of the human brain. Dr. Trauer himself looked like your average doctor – a little past middle age, a little overweight, a short greying beard. He motioned me to sit down and took the paperwork I’d been filling out.

“Hmmmm,” he said, reading it over. “29 years old, postdoc in biochem, recent relationship trouble…mmmm…you did the right thing.”

“In coming here?”

“No, in considering suicide. After getting rejected from a tenure-track position, your life is pretty much over.”

“WHAT?”

“I mean, here you are, hundreds of thousands of dollars in debt, with only one area of expertise, and now you’ve been rejected from it. I can totally see why you might think it’s worth ending it all.”

“But…there are lots of other things I can do! I can get a job in industry! I can work in something else! Even if I can’t find a job right away, I have parents who can help support me.”

“Industry!” Dr. Trauer was having none of it. “A bunch of bloodsuckers. Do you realize how bad work in the private sector is these days? They’ll abuse you and then spit you out, and once you’ve been out of university too long nobody else will want you.”

“Lots of people want biochemists! If I work for a company for a few years, I’ll have more experience and maybe that will make me more attractive to employers! What…what kind of a psychiatrist *are* you, anyway?"

"Cindy didn't tell you?"

"Cindy?"

"The woman on the phone."

"She didn't really tell me anything!"

"Well," said Dr. Trauer. "To answer your question, we're dark side psychiatrists. This is the state's only dark side psychiatry clinic."

"Dark side psychiatry? *Really?*"

"We're a…well, some people say sect, but I like to think of it as more of a guild…dedicated to improving negative mental health. Think of it this way. When you're a hijacked murder-monkey hurtling toward your inevitable death, sanity is a completely ridiculous thing to have. And when the universe is fifteen billion light-years across and almost entirely freezing void, the idea that people should have 'coping skills' boggles the imagination. An emotionally healthy person is a person who isn't paying attention, and our job is to cure them."

"There's more than one of you?"

"Oh, yes. There's a thriving dark side psychiatric community. There are dark side psychopharmacologists – you'd be amazed what a few doses of datura can do to a person. There are dark side psychotherapists who analyze and break down people's positive cognitions. There are dark side child psychiatrists who catch people when they're young, before sanity has had a chance to take root and worsen. And there are dark side geriatric psychiatrists, who go from nursing home to nursing home, making sure that the elderly are not warehoused and neglected at exactly the time it is most important to ensure that stroke or dementia does not protect them from acute awareness of the nearness of death."

"That's awful!" I said.

"Is it? Look where sanity's gotten you. You want to kill yourself, but you don't have the courage. Work with me for ten sessions, and I promise you we can help you *get* that courage."

"You're a @#!$ing quack," I said. "And if you think killing yourself is so great, how come you haven't done it yourself yet?"

"Who says I haven't?" asked Dr. Trauer.

His hand went to his face, and he plucked out his right eye, revealing an empty void surrounded by the bleached whiteness of bone. I screamed and ran out of the clinic and didn't stop running until I was in my house and had locked the door beside me.

II.

"…and that's pretty much the whole story, doctor," she told me. "And then I looked to see if there were any *real* psychiatrists in the area and someone referred me to you."

"Well," I said, my face unreadable. "I can certainly see why you're complaining of, how did you put it, 'depression and acute stress disorder'."

“Not so acute anymore. It took me two months to get an appointment at your clinic.”

“Oh,” I said. Then, “Sorry, we’re sort of backed up.” Then, “Okay. We’ve got a lot we have to work on here. Let me tell you how we’re going to do it. We’re going to use a form of therapy that challenges your negative cognitions. We’re going to take the things that are bothering you, examine the evidence for them, and see if there are alternative explanations.”

“What do you mean?” she asked.

“Well,” I said. “It seems to be this Dr. Trauer incident that’s traumatized you a lot. I can see why you would be stressed out. The way you tell it, it sounds absolutely terrifying.”

“You don’t believe me,” she said, not accusatory, just stating a fact.

“I think it would be helpful to examine alternate explanations,” I said. “I’m willing to assume it happened exactly as you tell it. I can see why you would think Dr. Trauer wanted you to commit suicide. But are there any alternative explanations for the same event?”

“I don’t see how there can be,” she said. “He outright said that he thought I should kill myself.”

“Right. But from what you know of psychiatrists and therapy – and you did say you took some classes in undergrad – are there any other reasons he might have said something like that?”

She thought for a second. “Wait,” she told me. “There’s a technique in therapy called paradoxical intention. Where you take a patient’s irrational thought, and then defend and amplify it. And then when the patient hears it from someone else, she realizes how silly it sounds and starts arguing against it, and then it’s really hard to keep believing it after you’ve shot it down yourself.”

I nodded. “That’s definitely a therapeutic method, and sometimes a very effective one. Do you have any evidence that this is what Dr. Trauer was doing?”

“Yes! As soon as he said I should commit suicide, I started arguing against him. He told me that if I couldn’t get a tenure track position there would be no other jobs available, and I told him there would be! Then he told me that the jobs would be terrible and I’d never be able to make a happy life for myself with them, and I argued that I would! That must have been what he was going for!”

She suddenly looked really excited. Then, just as suddenly, the worry returned to her face.

“But then what happened with his eye? I swear I saw him take it right out of the socket.”

I nodded. “Can you think of any alternate explanations for that?”

Thinking about it that way, it only took her like five seconds. She slapped her head like she’d been an idiot. “A glass eye. He probably had some kind of injury, had to put in a glass eye, and could take it out any time he wanted. He must have thought it would be a funny gag and didn’t realize how traumatized I’d be. Or he wanted to scare me into realizing how much I wanted to live. Or something.”

I nodded. “That does sound like a reasonable explanation.”

“But…don’t people with glass eyes usually have like scar tissue and normal skin behind them? This guy, I swear it was just the bone and this empty socket, like you were seeing straight to his skull.”

“You’re asking the right questions,” I said. “Now think a little more.”

“Hmmmm,” she said. “I guess I was really, really stressed out at the time. And I only saw it for, like, a fraction of a second. Maybe my brain was playing tricks on me.”

“That can definitely happen,” I agreed.

She looked a lot better now. “I owe you a lot of thanks,” she said. “I’ve only been here for, like, fifteen minutes, and already I think a lot of my stress has gone away. All of this really makes sense. That paradoxical intention thing is actually kind of brilliant. And I can’t deny that it worked – I haven’t been suicidal since I talked to the guy. In fact…okay, this is going to sound really strange, but…maybe I should go back to Dr. Trauer.”

I wrinkled my forehead.

“It’s not that I don’t like you,” she said. “But he had this amazing free clinic, and what he did for me that day…now that I realize what was going on, that was actually pretty incredible.”

“Hold on a second,” I said.

I left the room, marched up to the front desk, took the directory of medical providers in the area off the shelf, marched back to the room. I started flipping through the pages. It was in alphabetical order…Tang…Thompson…

Tophet…there we go. Trauer. My gaze lingered there maybe just a second too long, and she asked if I was okay.

“Um, yeah,” I said. “It’s just that he doesn’t – he doesn’t take your insurance. That’s the problem.”

“It’s okay,” she told me. “He said it was a free clinic. So that shouldn’t a problem.”

“Well, uh…the thing is…when you see out-of-network providers, your insurance actually charges, charges an extra fee. Even if the visit itself is free.”

She looked skeptical. “I’ve never heard of that.”

“It’s new. With Obamacare.”

“Really? How high a fee is it?”

“It’s…um…ten thousand dollars. Yeah, I know, right? Thanks, Obama.”

“Wow,” she said. “I definitely can’t afford that. I guess I’ll keep coming here. Not that there’s anything wrong with that. You’ve been very nice. It’s just that…with Dr. Trauer…well…sorry, I’ll stop talking now. Thanks a lot, doctor.” She stood up and shook my hand before heading for the door. “Seriously, I can’t believe how much you’ve helped me.”

No, I thought, as she departed *you can’t*. I told her she was asking the right questions, and she was, but not all of them.

For example, *why would a man with only one working eye have a stereographic Magic Eye image in his office?*

I picked up my provider directory again, stared a second time at the entry for Dr. Trauer. There was a neat line through it in red pen, and above, in my secretary’s careful handwriting, “DECEASED”.

Before returning the directory to the front desk, I took my own pen and added “DO NOT REFER” in big letters underneath.

Freedom On The Centralized Web

I.

A lot of libertarians and anarcho-capitalists envision a future of small corporate states competing for migrants and capital by trying to have the best policies.

But the Internet is about as close to that vision as we're likely to find outside the pages of a political philosophy textbook. And I am far from convinced.

Let's back up. Internet communities – ranging from a personal blog like this one all the way up to Facebook and Reddit – share many features with real communities. They work out rules for punishing defectors – your trolls, your harassers – and appoint a hierarchy of trusted individuals to carry out those rules. They try to balance competing concerns like free expression and public decency. They host cliques, power grabs, flame wars, even religious strife. They try to raise revenue, they establish a class system of Power Users and Premium Users, they deal with resentment from people who aren't getting their way. They develop a culture.

The job of a community leader, be they a blogger or the CEO of Facebook, is a lot like the job of the Mayor of New York City: create a pleasant community where talented people will want to live and work, where wrongdoing is met with swift punishment, and where you can collect revenue without annoying your constitutents too much. But it's even more like a hypothetical corporate state CEO in a Patchwork or Archipelago – wield absolute power, tempered by the knowledge that your citizens can leave at any time – and if they don't, skim a little off the top of their productive activity.

In theory, this is supposed to lead to amazing communities as corporate states optimize themselves to get more customer-citizens and new polities arise to take advantage of deficiencies in the old.

In practice, we tried this with the Internet for a couple of years, and then moved to the current system, where individual sites like blogs and little storefronts are in decline and conversation and commerce have moved to a couple of giant corporations: Facebook, Twitter, Reddit, Amazon, Paypal.

These companies aren't exactly monopolies. To some degree, if you're unsatisfied with Facebook you can move to Twitter. But they're not exactly competitors either – there are a lot of things Facebook is good for that Twitter fails completely, and vice versa. It's like Coca-Cola vs. milk: in theory you've always got the choice to drink either in place of the other; in practice you usually know which one you need at any given time. In that sense, there's no *real* Facebook competitor except eg Orkut or Diaspora, which no one uses.

Which suggests one reason *why* these sites are so dominant: their main selling point is their size. Facebook is the best because all of your friends are on it; if I made a much better Facebook clone tomorrow no one would go unless everyone else was already there (Google found this out the hard way). Amazon is the best because you can buy pretty much everything you want there; Paypal is the best because most sites take PayPal. So not only do they have no competitors, but it's really hard to imagine one ever arising. In order to compete with Facebook, you not only need a better product, you need a product that's so much better that everybody decides to switch *en masse* at the same time. The only example I can think of where this *ever* worked was the Great Digg Exodus, where Digg screwed up their product so thoroughly that everyone simultaneously said “@#!$ this” and moved to Reddit.

So instead of “let a thousand nations bloom”, it ended up more like “let five or six big nations bloom that we can never get rid of”.

II.

It’s a truism that the First Amendment only protects citizens from the government, not from other citizens. Nothing stops a private college from expelling any student who criticizes the administration, and nothing stops a private business from firing any employee who doesn’t support the boss’ preferred candidate. We apparently place our trust in the multiplicity of the market to maintain some semblance of freedom; out of thousands of competing companies, not *all* will ban the same political positions; if too many did so, other companies would start offering freedom of speech as a benefit and poach the more repressive companies’ employees and customers.

It’s a little concerning that we accept this argument about freedom of speech when we don’t accept it for anything else. We don’t trust the free market to necessarily preserve racial equality – that’s what anti-discrimination laws are for. We don’t trust the free market to necessarily preserve worker safety – that’s what OSHA and related regulations are for. We don’t even trust the free market to necessarily preserve fire safety – that’s why federal inspectors have to come in every so often to make sure you’re not secretly plotting to let your employees fry. Whenever we think something is *important*, we regulate the hell out of it, rights-of-private-companies to-set-their-own-policies be damned. But free speech? If you don’t trust the free market to sort it out, the only possible explanation is that you *just don't understand the literal text of the First Amendment*.

The argument for non-discrimination laws is that discrimination isn't just random noise. If a couple of companies here and there decided to discriminate, then they might be easily overtaken by nimbler companies willing to take any employees and customers who came to them; and even if they didn't, a couple of companies here and there discriminating wouldn't be the end of the world. The argument for non-discrimination laws is that discrimination can take the form of global social pressure in favor of discrimination, enforced by punishing defectors, to the point where certain races can find themselves locked out of the economy altogether.

Concerns about freedom of speech come from much the same place. Back when homosexuality was really taboo, you'd have a very tough time finding any reference to it, let alone a positive reference to it, in any newspaper or TV channel in the country. All the big companies knew that talking about it (or letting their editorial staff talk about it) was the sort of thing that could get them in trouble, and they had no particular incentive to do so – so they didn't. Yes, eventually they reversed that policy, but I'm not exactly going to be able to cite an example that *didn't* later become okay and still have everyone believe it's a good example of something it was wrong to have banned!

But even when homosexuality was banned from formal discussion on the news, there was still the opportunity to discuss it with your friends in private. I don't know much about the history of the gay rights movement, but I understand it was a few small groups of like-minded people who managed to coordinate such discussions among themselves using non-mass-media that started some of the activism that eventually led to it become accepted more generally.

Nowadays that’s a little more complicated. If every company in the world decided that their profit margin required them to appear Tough On Homosexuality, it wouldn’t just mean no mass media editorials. Insofar as a lot of the public square has been annexed by Facebook and Twitter and Reddit, the discussion can be kept out of the public square in a way it couldn’t have been previously. Insofar as the economy relies on PayPal and Amazon as a currency system and marketplace respectively, companies can just decide that currency cannot be used to support gay rights, in much the same way that for a while currency could not be used to support WikiLeaks. The nuclear option is that Google decides not to show gay-related sites in its search results, so that you could make as many persuasive arguments for legalizing homosexuality as you want and no one would ever find them unless you knock on their door and hand them the URL directly.

(The *thermonuclear* option is that browsers just include some code to refuse to render any site relating to homosexuality, and now you’re done. But that is ridiculous – who would ever believe that browser companies would take it upon themselves to be the arbiter of people’s personal beliefs about homosexuality?)

This is not *entirely* theoretical. You want some really weird porn? You probably won’t find it on Amazon, according to the delightfully-named article Amazon’s War On Bigfoot Erotica. After they got bad press for hosting some kind of out-there stuff, they decided that anything which offended too many people’s sensibilities was a liability. This echoes a much more serious decision from a few years earlier: Paypal threatened to suspend the accounts of any companies selling sufficiently gross erotic books. Booksellers, many of whom made only a tiny percent of their profit from erotica, claimed that their hands were tied; if you can't use PayPal, selling on the Internet suddenly becomes a much more dubious proposition. This story has a happy ending; Paypal eventually amended their policy to limit it to much more specific cases. But for a while, it was touch-and-go enough that a few people started wondering: "Hey, maybe we *shouldn't* have entrusted our entire commercial infrastructure to a private company with no accountability."

Advocates of net neutrality like to worry about a "two-tiered" Internet, where the companies that can make sweetheart deals with the ISPs are easy for everyone to access, and everybody else can only be accessed with a bit more money and a bit more trouble. Well, I worry about a two-tiered marketplace of ideas. Write *decent* erotica, socially approved erotica where everyone has heterosexual sex and then goes to church afterwards, and you can sell it on Amazon, collect profits using PayPal, talk to your friends about it on Facebook, and advertise on Reddit. Write *weird* erotica, the kind that other people might find offensive, and you might have to start your own website, take payment via some inconvenient method like Bitcoin, have trouble advertising it by word of mouth, and not be able to talk about it on literary discussion forums. It's not that you've been *banned* from writing your erotica. You can *write* it. It's just that practically nobody else will ever hear about it or buy it, except maybe the tiny fraction of people who are already extremely clued-in to the weird erotica scene and know exactly where to look for it.

This isn't so much different from the old days when nobody would talk about homosexuality. Indeed, one could argue that the modern world is friendlier to people with unpopular ideas – there are more opportunities to self-publish, to bypass traditional bookstores, and to get covered in weird niche news outlets.

But at the same time, the amount of the information ecology controlled by private companies has increased drastically, and if private companies don’t like you, now you have entirely new problems.

III.

I used to think that there was enough demand for a free marketplace of ideas that if a company become too restrictive, another one would spring up to replace it. Then I suffered through the conflict between Reddit and Voat.

Reddit recently alienated (no pun intended) some of its users, who decided to move *en masse* to an alternative Reddit-like platform called Voat, whose owner promised not to restrict content unless it was illegal (in his home country of Switzerland, which permits a lot). I don’t want to get into the details too much (though I did explain my perspective on it on Tumblr), but suffice it to say that (one) (small) part of the problem was that people thought Reddit was failing its free speech principles by cracking down on various unsavory groups.

HL Mencken once said that “the trouble with fighting for human freedom is that one spends most of one’s time defending scoundrels. For it is against scoundrels that oppressive laws are first aimed, and oppression must be stopped at the beginning if it is to be stopped at all.”

There’s an unfortunate corollary to this, which is that if you try to create a libertarian paradise, you will attract three deeply virtuous people with a strong committment to the principle of universal freedom, plus millions of scoundrels. Declare that you’re going to stop holding witch hunts, and your coalition is certain to include more than its share of witches.

So while some small percent of Reddit's average users moved over, a very large percent of its witches did. Sometimes the witchcraft was nothing worse than questioning Reddit's political consensus. Other times, it was harassment, hate groups, and creepy porn.

I feel obligated to say that, in spite of CONSTANT MEDIA SMEARS, Reddit's community is amazing, puts in astounding effort to help its members and fight for good causes all over the world, and that the representation of weirdoes and terrible people is no higher than any other part of the population. But that's not zero. And a disproportionate number of those people became interested in the new site.

Already, we see why the typical answer "If you don't like your community, just leave and start a new one" is an oversimplification. A community run on Voat's rules with Reddit userbase would probably be a pretty nice place. A community run on Voat's rules with the subsection of Reddit's userbase who will leave Reddit when you create it is…a very different community. Remember that whole post on Moloch? Even if everyone on Reddit agrees in preferring Voat to Reddit, it might be impossible to implement the move, because unless everybody can coordinate it's always going to be the witches who move over first, and nobody wants to move to a community that's mostly-witch.

But the problem isn't just natural self-sorting. The problem is natural self-sorting, plus enemy action. Remember, the big corporations do what they do because it's what everyone in society is demanding. To break from that mold is to pretty much set yourself up as everyone's enemy and invite retaliation. The media and Reddit's SJ community quickly denounced Voat as Public Enemy No 1; as a result, in its first week it got DDoS attacked, deleted by its hosting company with no explanation except “the content on your server includes politically incorrect parts”, and had its PayPal account frozen. As a result, the Great Reddit Exodus was placed on hold while they tried to get their site back up, and by the time they did Reddit had switched CEOs and the momentum was gone.

Advocates of free-market governance and “let a thousand nations bloom” like to talk as if overly restrictive laws in one polity will immediately result in the rise of other competing policies that throw off their shackles and outcompete the first. But even on the relatively lawless Internet, where startup costs are so low that a random student from Switzerland can decide on a whim to take on one of the largest websites in the world, it’s way more complicated than that.

IV.

Actually, the whole Reddit thing left a bad taste in my mouth.

It would be paranoid to say that there are people for whom fighting against free speech is a *terminal value*, but let me make a slightly weaker claim. There are people who consider themselves the protectors of decency, who notice that their opponents are usually using the value “free speech” to oppose their demands, and so “free speech” to these people becomes the equivalent of “small government” or “tolerance and equality” or “family values” – a value which most people agree is good, but which has gotten claimed by one side of a political argument so hard that for the other side it becomes an outgroup signal and sign of cringeworthy bad arguments which must be shot down. These people don’t quite have fighting free speech as a terminal value, but you might as well model them as if they do. These are the people who say “freeze peach” in the same way other people say “but mah jawbs!”

And these people have a winning strategy. I’ve seen it with Reddit and any other website that gets on their bad side. The strategy is weaponized stereotype campaigns. If a site tolerates witches, describe it as a witch site about witchcraft populated entirely by witches. It’s super easy. By happy coincidence, Slate even has an article calling people out on it this very week.

Think about it like this. No matter how many brilliant artists, scientists, and humanitarians Islam produces, in the mind of a good chunk of Westerners it will always be associated first and foremost with terrorism. Redditors, Diggians, Tumblrites, 4chanistas, Instagramastanis, Slashdotmen, Metafilterniks – all are groups that the average person knows a whole lot less about than they do Muslims. A concerted campaign to irrevocably identify an entire online community with a few atrocious actions by its worst members will succeed pretty much instantly. There are 36 million Redditors, so unless they advertise solely in the saint demographic, we expect the worst members to be pretty bad. Therefore, Reddit is at the mercy of anyone with the resources to start such a campaign. Reddit Inc’s main asset is its brand, so it has every incentive to cave – even a principled leadership would rather make a few administrative changes than sacrifice the whole to save some Holocaust deniers or whatever.

After that, the site’s userbase has two options – either suck it up, or go off somewhere else. Go off somewhere else, and they’ll get DDoSed, taken down by their host, and slowly starved of money like Voat, at the same time as the same media forces accuse the new site of being a hot spot for witchcraft – this time with good reason. The new site might not die out completely, but it will be sufficiently established in the hearts of everyone as a Bad Place that it will be stuck in the same equilibrium as central Detroit – only people with no other options will go there, because it is inhabited mostly by the sort of people with no other options.

The worst possible end-game for this is the two-tier marketplace of ideas mentioned above, with an unfortunate twist – everyone knows that the second tier is inhabited entirely by witches, and therefore being on the second tier is sufficient to convict you. Unpopular ideas are gradually forced out of the first tier by media smear campaigns, and from then on everyone believes the effort was justified, because it's one of those second-tier ideas that you only find in the same sites as the racists and trolls and child pornographers. You're not a *second tier* kind of person, are you? No, we didn't think so.

I have no particular solution to this. Certainly the well-intentioned solutions other people are working on, like a decentralized crypto-Reddit that can't be moderated even in principle, are unlikely to help (hint: what is the most striking difference between Bitcoin marketplaces and normal marketplaces?) My primary hope is that it's just not a real problem. Certainly there has been very little in the way of speech restriction so far, and what little there has been has been against things which, on the object level, I'm happy to see gone. It's entirely possible that we'll escape with only a few things banned that probably deserve it. I certainly hope this is the case.

I'm just annoyed that we've gotten ourselves in a corner where we have to depend on hope.

The General Factor Of Correctness

People on Tumblr are discussing Eliezer Yudkowsky's old essay The Correct Contrarian Cluster, and my interpretation was different enough that I thought it might be worth spelling out. So here it is: is there a General Factor of Correctness?

Remember, IQ is supposed to come from a General Factor Of Intelligence. If you make people take a lot of different tests of a lot of different types, people who do well on one type will do well on other types more often than chance. You can do this with other things too, like make a General Factor Of Social Development. If you're really cool, you can even correlate the General Factor of Intelligence and the General Factor of Social Development *together*.

A General Factor Of Correctness would mean that if you asked people's opinions on a bunch of controversial questions, like "Would increasing the minimum wage to $15 worsen unemployment?" or "Which interpretation of quantum mechanics is correct?" or "are artificial sweeteners safe?" and then somehow discovered the answers to these questions, people who did well on one such question would do well on other types more often than chance.

This is a *surprisingly* deep and controversial issue, but one with potentially big payoffs. Suppose you want to know whose economic theories are right, but you don't want to take the time to learn economics. Consider some position that was once considered fringe and bizarre, but now known to be likely true – for example, pre-Clovis settlement of the New World. Find the economists who believed in pre-Clovis settlement of the New World back when doing so was unpopular. Those economists have demonstrated a proven track record of being able to winnow out correct ideas amidst a sea of uncertainty. Invest in whatever company they tell you to invest in and make a killing.

I’m sort of joking, but also sort of serious – shouldn’t something like this work? If there’s such a thing as reasoning ability, people who are good at sifting through a mess of competing claims about pre-Columbian anthropology and turning up the truth should be able to apply that same skill to sifting through a mess of competing claims about economic data. Right?

If this is true, we can gain new insight into all of our conundra just by seeing who believes what about New World migration. That sounds useful. The problem is, to identify it we have to separate it out from a lot of closely related concepts.

The **first** problem: if you just mark who’s right and wrong about each controversial issue, the General Factor Of Correctness will end up looking a lot like a General Factor of Agreeing With Expert Consensus. The current best-known heuristic is “always agree with expert consensus on everything”; people who follow this heuristic all the time are most likely to do well, but we learn nothing whatsoever from their success. If I can get brilliant-economist-points for saying things like “black holes exist” or “9-11 was not a government conspiracy”, then that just makes a mockery of the whole system. Indeed, our whole point in this exercise is to see if we can improve on the “agree with experts” heuristic.

We could get more interesting results by analyzing only people’s deviations from expert consensus. If you agree with the consensus about everything, you don’t get to play. If you disagree with the consensus about some things, then you get positive points when you're right and negative points when you're wrong. If someone ends consistently ends up with a positive score beyond what we would expect by chance, then they're the equivalent of the economist who was surprisingly prescient about pre-Clovis migration – a person who's demonstrating a special ability that allows them to outperform experts. This is why Eliezer very reasonably talks about a correct *contrarian* cluster instead of a correct cluster in general. We already know who the correct cluster is, and all of you saying "I have no idea what Clovis is, but whatever leading anthropologists think, I think that too" are in it. So what? So nothing.

The **second** problem: are you just going to rediscover some factor we already know about, like IQ or general-well-educatedness? I'm not sure. WHen I brought this up on Tumblr, people were quick to point out examples of very intelligent, very well-educated people believing stupid things – for example, Newton's obsession with alchemy and Biblical prophecy, or Linus Pauling's belief that you could solve health just be making everyone take crazy amounts of Vitamin C. These points are well-taken, but I can't help wondering if there's selection bias in bringing them up. Yes, some smart people believe stupid things, but maybe even more stupid people do? By analogy, many people who are brilliant at math are terrible at language, and we can all think of salient examples, but psychometrics has shown again and again that *in general* math and language skills are correlated.

If we look for more general data, we get inconsistent results. Neither IQ nor educational attainment seems to affect whether you believe in climate change very much, though you can get slightly different results depending on how you ask and what you adjust for. There seems to be a stronger effect of intelligence increasing comfort with nuclear power.

Other polls show IQ may increase atheism, non-racism, and a complicated cluster of political views possibly corresponding to libertarianism but also showing up as “liberalism” or “conservativism” depending on how you define your constructs and which aspects of politics you focus on. I am very suspicious about any of this reflecting real improved decision-making capacity as opposed to just attempts to signal intelligence in various ways.

The **third** problem: can we differentiate positive from negative selection? There are lots of people who believe in Bigfoot *and* ESP *and* astrology. I *suspect* these people will be worse at other things, including predicting economic trends, predicting world events, and being on the right side of difficult scientific controversies, probably in a way independent of IQ or education. I’m not sure of this. But I suspect it. If I’m right, then the data will show a General Factor of Correctness, but it won’t necessarily be a very interesting one. To give a reductio ad absurdum, if you have some mental disorder that causes you to live in a completely delusional fantasy world, you will have incorrect opinions about everything at once, which looks highly correlated, but this doesn’t necessarily prove that there are correlations among the people who are *more* correct than average.

The **fourth** problem: is there a difference between correctness and probability calibration? Suppose that Alice says that there’s a 90% chance the Greek economy will implode, and Bob has the same information but says there’s only an 80% chance. Here it might be tempting to say that one of either Alice or Bob is miscalibrated – either Alice is overconfident or Bob is underconfident. But suppose Alice says that there’s a 90% chance the Greek economy will implode, and Bob has the same information but says there’s only a 10% chance that it will. Now we’re more likely to interpret this in terms of them just disagreeing. But I don’t know enough about probability theory to put my finger on whether there’s a true qualitative difference.

This is important because we *know* calibration is a real thing and some people are good at it and other people aren’t but can improve with practice. If all we’re showing is that people who are good with probabilities are good with probabilities, then whatever.

But there are tantalizing signs that there might be something more here. I was involved in an unpublished study which I can’t upload because I don’t have the other authors’ permission, but which showed conclusively that people with poor calibration are more likely to believe in the paranormal ($p < 0.001$), even when belief in the paranormal was not assessed as a calibration question. So I went through the Less Wrong Survey data, made up a very ad hoc measure of total calibration skill, and checked to see what it did and didn't predict. Calibration was correlated with IQ (0.14, $p = 0.01$). But it was also correlated with higher belief in global warming (0.13, $p = 0.01$), with higher belief in near-term global catastrophic risk (-0.08, $p - 0.01$), increased support for immigration (0.06, $p = 0.048$) and with decreased support for the human biodiversity movement (0.1, $p = 0.002$). These were all independent of the IQ correlation. Notably, although warming and GCR were asked in the form of probabilities, immigration and HBD weren't, suggesting that calibration can be (weakly) correlated with opinions on a non-calibration task. Maybe the most intriguing evidence for a full-fledged General Factor of Correctness comes from Philip Tetlock and IARPA's Good Judgment Project, which got a few thousand average people and asked them to predict the probability of important international events like “North Korea launches a new kind of missile.” They found that the same small group of people consistently outperformed everyone else in a way incompatible with chance. These people were not necessarily very well-educated and didn't have much domain-specific knowledge in international relations – the one profiled on NPR was a pharmacist who said she "didn't know a lot about international affairs [and] hadn't taken much math in school" – but they were reportedly able to outperform professional CIA analysts armed with extra classified information by as much as 30%.

These people aren't succeeding because they parrot the experts, they're not succeeding because they have more IQ or education, and they're not succeeding in some kind of trivial way like rejecting things that will never happen. Although the article doesn't specify, I *think* they're doing something more than *just* being well-calibrated. They seem to be succeeding through some mysterious quality totally separate from all of these things.

But only on questions about international affairs. What I'd love to see next is what happens when you ask these same people to predict sports games, industry trends, the mean global temperature in 2030, or what the next space probe will find. If they can beat the experts in those fields, then I start *really* wondering what their position on the tax rate is and who they're going to vote for for President.

Why am I going so into depth about an LW post from five years ago? I think in a sense this is the center of the entire rationalist project. If ability to evaluate evidence and come to accurate conclusions across a broad range of fields relies on some skill other than brute-forcing it with domain knowledge and IQ, some skill that looks like "rationality" broadly defined, then cultivating that skill starts to look like a pretty good idea.

Enrico Fermi said he was fascinated by the question of extraterrestrial life because whether it existed or it didn't, either way was astounding. Maybe a paradox, but the same paradox seems true of the General Factor of Correctness.

Outside the Laboratory is a post about why the negative proposition – no such General Factor – should be astounding:

> "Outside the laboratory, scientists are no wiser than anyone else." Sometimes this proverb is spoken by scientists, humbly, sadly, to remind themselves of their own fallibility. Sometimes this proverb is said for rather less praiseworthy reasons, to devalue unwanted expert advice. Is the proverb true? Probably not in an absolute sense. It seems much too pessimistic to say that scientists are literally no wiser than average, that there is literally zero correlation.
>
> But the proverb does appear true to some degree, and I propose that we should be very disturbed by this fact. We should not sigh, and shake our heads sadly. Rather we should sit bolt upright in alarm. Why? Well, suppose that an apprentice shepherd is laboriously trained to count sheep, as they pass in and out of a fold. Thus the shepherd knows when all the sheep have left, and when all the sheep have returned. Then you give the shepherd a few apples, and say: "How many apples?" But the shepherd stares at you blankly, because they weren't trained to count apples – just sheep. You would probably suspect that the shepherd didn't understand counting very well.
>
> If, outside of their specialist field, some particular scientist is just as susceptible as anyone else to wacky ideas, then they probably never did understand why the scientific rules work. Maybe they can parrot back a bit of

> Popperian falsificationism; but they don't understand on a deep level, the algebraic level of probability theory, the causal level of cognition-as-machinery. They've been trained to behave a certain way in the laboratory, but they don't like to be constrained by evidence; when they go home, they take off the lab coat and relax with some comfortable nonsense. And yes, that does make me wonder if I can trust that scientist's opinions even in their own field – especially when it comes to any controversial issue, any open question, anything that isn't already nailed down by massive evidence and social convention.
>
> Maybe we can beat the proverb – be rational in our personal lives, not just our professional lives.

And Correct Contrarian Cluster is about why the positive proposition should be equally astounding. If it's true, you *can* gain a small but nonzero amount of information about the best economic theories by seeing what their originators predicted about migration patterns in pre-Columbian America. And you can try grinding your Correctness stat to improve your ability to make decisions in every domain of knowledge simultaneously.

I find research into intelligence more interesting than research into other things because improvements in intelligence can be leveraged to produce improvements in everything else. Research into correctness is one of the rare other fields that shares this quality, and I'm glad there are people like Tetlock working on it.

Discussion questions (adapted from Tumblr):

1. Five Thirty Eight is down the night before an election, so you search for some other good sites that interpret the

polls. You find two. Both seem to be by amateurs, but both are well-designed and professional-looking and talk intelligently about things like sampling bias and such. The first site says the Blue Party will win by 5%; the second site says the Green Party will win by 5%. You look up the authors of the two sites, and find that the guy who wrote the first is a Young Earth Creationist. Do you have any opinion on who is going to win the election?

2. On the bus one day, you sit next to a strange man who mumbles about how Bigfoot caused 9-11 and the Ark of the Covenant is buried underneath EPCOT Center. You dismiss him and never see him again. A year later, you see on TV that new evidence confirms Bigfoot caused 9-11. Should you head to Florida and start digging?

3. Schmoeism and Anti-Schmoeism are two complicated and mutually exclusive economic theories that you don’t understand at all, but you know the economics profession is split about 50-50 between them. In 2005, a survey finds that 66% of Schmoeist economists and 33% of anti-Schmoeist economists believe in pre-Clovis settlement of the New World ($p = 0.01$). In 2015, new archaeological finds convincingly establish that such settlement existed. How strongly (if at all) do you now favor one theory over the other?

4. As with 3, but instead of merely being the pre-Clovis settlement of America, the survey asked about ten controversial questions in archaeology, anthropology, and historical scholarship, and the Schmoeists did significantly better than the anti-Schmoeists on 9 of them.

Non-Dual Awareness

Seen on Lauren’s Facebook: How Does Academia Resemble A Drug Gang?

Their answer is that both academia and drug gangs are marked by an endless supply of foot soldiers willing to work in terrible conditions for a small chance at living the good life. In drug gangs, the average street-corner dealer makes $3-something an hour; given that he’s got a high chance of being arrested or shot, why doesn’t he switch to McDonalds instead where the pay’s twice as good and the environment’s a lot safer? The article suggests one reason is because drug gangs offer the chance of eventually becoming a drug kingpin who is drowning in money.

(I’d worry they’re exaggerating the importance of this factor compared to wanting to maintain street cred and McDonalds jobs being much more regimented both in the application process and performance, but they’re the ones who have talked to anthropologists embedded in drug gangs, not me.)

Academia has the same structure. TAs and grad students work in unpleasant conditions for much less than they could make in industry, because there’s always the chance they could become a tenured professor who gets to live the life of the mind and travel to conferences in far-off countries and get summer vacations off.

The article describes this structure as “dualization” – a field that separates neatly into a binary classification of winners and losers.

This concept applies much more broadly than just drugs and colleges. I sometimes compare my own career path, medicine, to that of my friends in computer programming. Medicine is very clearly dual – of the millions of pre-med students, some become doctors and at that moment have an almost-guaranteed good career, others can't make it to that MD and have no relevant whatsoever in the industry. Computer science is very clearly non-dual; if you're a crappy programmer, you'll get a crappy job at a crappy company; if you're a slightly better programmer, you'll get a slightly better job at a slightly better company; if you're a great programmer, you'll get a great job at a great company (ideally). There's no single bottleneck in computer programming where if you pass you're set for life but if you fail you might as well find some other career path.

My first instinct is to think of non-dualized fields as healthy and dualized fields as messed up, for a couple of reasons.

First, in the dualized fields, you're putting in a lot more risk. Sometimes this risk is handled well. For example, in medicine, most pre-med students don't make it to doctor, but the bottleneck is early – acceptance to medical school. That means they fail fast and can start making alternate career plans. All they've lost is whatever time they put into taking pre-med classes in college. In Britain and Ireland, the system's even better – you apply to med school right out of high school, so if you don't get in you've got your whole college career to pivot to a focus on English or Engineering or whatever. But other fields handle this risk less well. For example, as I understand Law, you go to law school, and if all goes well a big firm offers to hire you around the time you graduate. If no big firm offers to hire you, your options are more limited. Problem is, you've sunk three years of your life and a lot of debt into learning that you're not wanted. So the cost of dualization is littering the streets with the corpses of people who invested a lot of their resources into trying for the higher tier but never made it.

Second, dualized fields offer an inherent opportunity for oppression. We all know the stories of the adjunct professors shuttling between two or three colleges and barely making it on food stamps despite being very intelligent people who ought to be making it into high-paying industries. Likewise, medical residents can be worked 80 hour weeks, and I've heard that beginning lawyers have it little better. Because your entire career is concentrated on the hope of making it into the higher-tier, and the idea of *not* making it into the higher tier is too horrible to contemplate, and your superiors control whether you will make it into the higher tier or not, you will do whatever the heck your superiors say. A computer programmer who was asked to work 80 hour weeks could just say "thanks but no thanks" and find another company with saner policies.

(except in startups, but those bear a lot of the hallmarks of a dualized field with binary outcomes, including the promise of massive wealth for success)

Third, dualized fields are a lot more likely to become politicized. The limited high-tier positions are seen as spoils to be distributed, in contrast to the non-dual fields where good jobs are seen as opportunities to attract the most useful and skilled people. This reminds me of the *other* article I read today comparing academia to drug gangs, which was where Paul Krugman theorized that the reason so many criminals have horrible tattoos in inappropriate places is as a conspicuous symbol of criminality! He says that since these people's tattoos mean they can never get a job in legitimate industry, other gang members and black market contacts can trust them to keep their bargains, since they've got no option under than continuing to work in the criminal underworld. Krugman writes (h/t Nathaniel Bechhofer):

> The author, Diego Gambetta, adds a wonderful parallel: according to his account, Italian academics, who do a lot of horse-trading in appointments etc., cultivate a reputation for incompetence at actual research, again designed to reassure those with whom one deals.
>
> “Being incompetent and displaying it,” he writes, “conveys the message I will not run away, for I have no strong legs to run anywhere else. In a corrupt academic market, being good at and interested in one’s own research, by contrast, signal a potential for a career independent of corrupt reciprocity. In the Italian academic world, the kakistrocrats are those who best assure others by displaying, through lack of competence and lack of interest in research, that they will comply with the pacts.”

(wait, this argument sounds kind of familiar. KRUGMAN, HAVE THEY GOTTEN TO YOU TOO?!)

II.

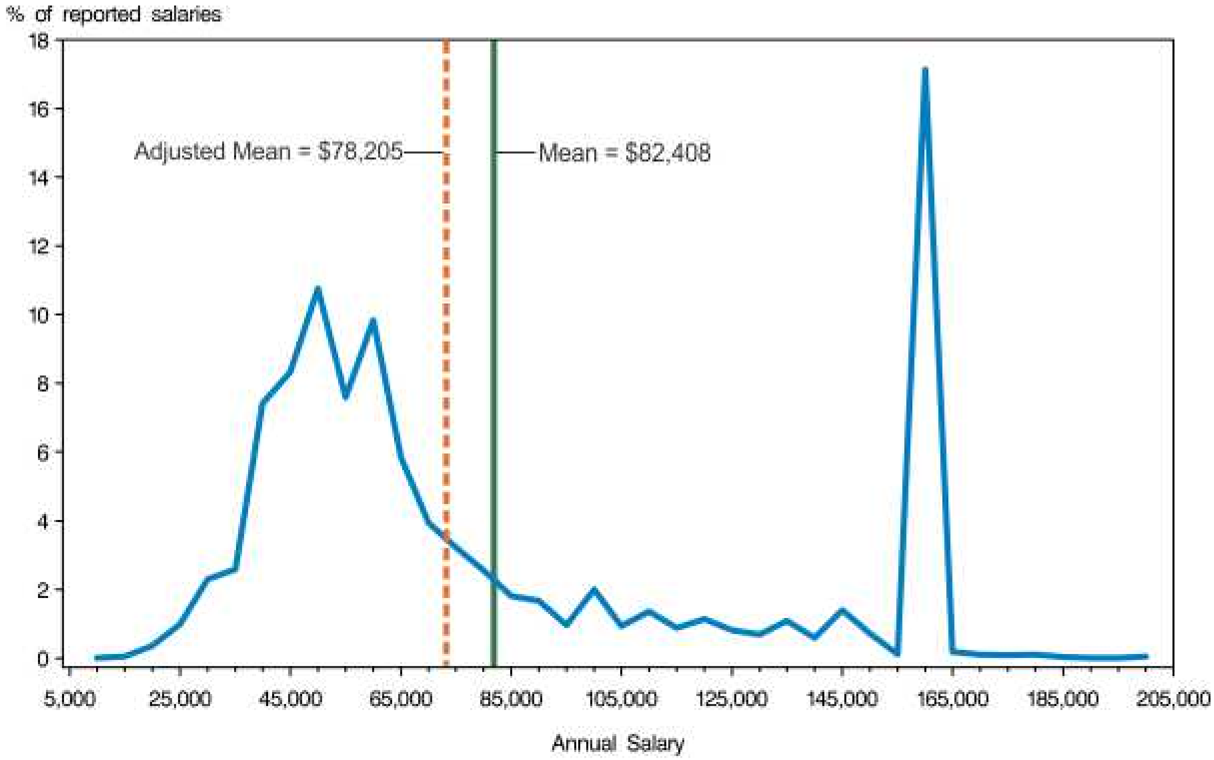
What dualizes some fields but not others?

Originally I was going to make a simplistic comment about licensing and regulation, but this doesn’t exactly capture it. Certainly the fact that medicine requires an MD has some effect on the dualization of medicine (alternatively, insofar as medicine isn’t *entirely* dualized, it’s because you can get less lucrative positions, like naturopath or therapist or nurse practitioner, without an MD). But we can imagine a system in which there were more than enough medical schools for everyone, anyone who applied to one got in, there was a glut of doctors, and the good doctors got good jobs and the less good doctors got less good jobs. So we might more soberly blame it on *scarce* licenses – for complicated reasons I won’t get into here, the number of residency spots is much lower than it should be, leading to a bottleneck where only a few people can obtain the MDs.

What about tenure? We can imagine an alternate universe where academia is populated with various PhDs on equal footing. Since there would be a glut, their salaries would be very low to start, but low salaries would mean easy employment, and colleges would find a lot of room for them to do one-on-one tutoring, or low-level research, or something like that. Eventually some of them would become a bit more prestigious in their fields and could demand higher salaries from hiring institutions, and a few superstars like Nobel Prize winners and the like could demand millions. At no point would there ever be anything called a “tenure track”. It seems like the main difference between this universe and our own is that tradition plus the reasonable desire of professors to be free from political interference has created this dichotomous variable called “tenure” and caused it to replace the continuous variable of salary as the prize for success. In favor of that theory, top professors seem weirdly underpaid compared to eg top athletes or top artists, even though I would expect having one of the world’s top scientists or historians to be a big draw for a school. According to the List Of Highest Paid Professors, only five professors in the US make more than a million dollars a year, and all of those are professors of lucrative medical subspecialties or of finance, who presumably are being paid that much to compensate them for teaching instead of participating in the high-paying professions they are otherwise qualified for.

What about drug dealers? I think there might be “licensing” at work here too. There’s no such thing as a mid-level independent drug dealer, because – if the three seasons I’ve watched of Breaking Bad are accurate – if you try this, the other drug dealers will shoot you. So you need a scarce “license” from the drug lords – basically El Chapo giving you the rights to a big piece of “turf” – instead of a license from the government. Whatever; I’m liberal Monday Wednesday Friday and libertarian Tuesdays and Thursdays; today is a Tuesday so all organizations that rely upon the use of force look the same to me.

But what about lawyers? Sure, there are regulations on who can practice, but the dualization in the legal fields comes after graduation of law school. Here, have some statistics:



© NALP, 2014
www.nalp.org

The first step of what’s going on isn’t a mystery – the people on the very sharp mountain on the right are hired by big law firms on the “partner track” (note the similarity to “tenure-track”) and the people in the more gradual plateau on the left are everyone else. But there’s still a lot to be explained. Why isn’t there a law firm that hires people *almost* as good as the people on the mountain, for $100,000? This article suggests that “Not paying the standard top-tier salary [of $160,000] is a tacit admission that you’re no longer top-tier”, but you could say that about any industry where quality isn’t 100% obvious. How come chefs don’t have a salary graph that looks like that? How come engineers don’t?

It seems possible that maybe top law firms act as a de facto licensing system – picking out a couple of excellent young law school grads as Officially Excellent, and then if you’re a sufficiently big corporation you refuse to use any except those? But once again, I don’t know why law would develop this structure and other professions wouldn’t.

So if I had to figure out what all of these have in common, it would be an idea of *privilege*. Some people get guaranteed an unexpected privilege over and above the continuous measure of salary. The people who have to subsidize this privilege resent it and try to limit access to it. People start competing for scarce access to the privilege instead of having normal competitions for salaries, benefits, working conditions, et cetera, and all of those other things go out the window.

This is interesting because of how well it maps on to some other issues. For example, minimum wage creates a dualized system between workers and the unemployed. If there were no minimum wage, we would expect a sort-of-continuous wage distribution from 0.01$ an hour all the way up to whatever Taylor Swift makes for an hour’s performance. Instead, we guarantee everyone the privilege of $15 per hour. Employers resent this and (in theory) try to limit access to the privilege by lowering workforce, automating, etc, as much as possible. This creates a dualized system with an upper tier (employees with high wages) and a lower tier (unemployed with nothing at all).

Or how about benefits? If there were no benefits, we'd expect a more continuous spectrum of people working 40 hour weeks, 30 hour weeks, 20 hour weeks, and so on. Instead, we guarantee everyone who works X hours the privilege of good health care. Employers resent this and try to limit access to the privilege by hiring people to work X – 1 hours per week, or hiring independent contractors, or so on. This creates a dualized system with an upper tier (real employees) and a lower tier (people working 29.999 hours a week or whatever who don't quite qualify for the benefits).

If you really want to stretch it, think about urban growth. If there was no zoning or regulation, desirable cities would have a continuous distribution from rich people living in nice mansions with lots of surrounding green land to poor people in apartment projects. Instead, we guarantee people living there certain privileges like "never having their view blocked" and "never having to worry about congestion". This creates a two-tier system of current residents with the privileges, and non-residents who can't live in desirable cities at all.

This raises a question of – assuming we want to give people privileges – or assuming we're political realists who understand it's going to happen anyway – are there ways to do it with a minimum of dualizing? It seems possible to imagine some solutions along those lines – for example, instead of mandating full health care for people who work more than 30 hours per week, we could seek systems where companies give health benefits that scale up with the number of hours worked. Instead of giving tenure, we seek systems where it becomes progressively harder to fire academics the longer they’ve worked for you.

Other cases seem harder – you can’t give half of a medical license to a doctor who finishes two years of med school, and the idea of a half a minimum wage defeats the whole point.

Stalin and Summary Statistics

[Epistemic status: As always, I am not a statistician, and anything I say should be taken with a grain of salt until confirmed by others]

I.

A while ago, I wrote Beware Summary Statistics, where I talked about all the ways I've been misled by things like r-values and so on. I recently found some really interesting cases that brought up a few more some of these issues.

Back in June, Noah Smith blogged about a study on IQ And The Wealth Of States.

Some background: a group including Richard Lynn suggests that IQ is the driving factor behind income differences among countries. They are able to cite statistics on how a very rich country like Singapore has an average IQ of almost 110, and a very poor country like Haiti – well, it's hard to say, because not too many Haitians take IQ tests and the ones who do might be so confused by this weird new idea of filling out a written multiple choice test that they choke and underperform – but officially Haiti has an IQ of like 70. Since you need smart people to build cool things like highways and power plants, maybe this explains a lot of the development/underdevelopment dichotomy. These people can point to a pretty good correlation between national IQ and national development to support their thesis, but the obvious counterargument is that maybe highly developed nations have good health and good education which raises IQ.

Anyway, the study Noah blogged about tested the application of this theory to US states. Noah sums up the results as follows:

> The upper bound for the amount of state income differences that can be explained by population I.Q. differences is about a third. If we assume that achievement scores are a good measure of I.Q. and that school attainment doesn't improve I.Q. very much, then the number goes down to about one-sixth.
>
> What this really shows is that there is Something Else that is driving state income differences. My personal guess is that this Something Else is mainly "external multipliers" from trade (the Krugman/Fujita theory). Institutions probably play a substantial role as well (the Acemoglu/Robinson theory). That's certainly relevant for the debate about different models of capitalism, where we often compare the U.S. to Scandinavia and other rich places.
>
> In any case, this result should be sobering for proponents of I.Q. as the Grand Unified Theory of economic development. Average I.Q. is not unimportant for rich countries, and we should definitely try to raise it through better nutrition, education, and (eventually) brain-boosting technologies. And it still might matter a lot for some poor countries. But for rich countries, there are things that matter a lot more.

II.

Let's go to another study. The Atlantic has an article on How Rich People Raise Rich Kids, which is about Black, Devereux, Lundborg, and Majlesi (2015). They look at adopted Swedish kids and determine whether their wealth (not income!) as adults is more correlated with that of their biological parents or their adoptive parents. They find that non-adopted kids' wealth correlates with that of their non-adopted parents at 0.33, adopted kids with biological parents at 0.13, and adopted kids with adoptive parents at 0.23. This suggests that upbringing is more important than genetics in determining how much wealth you will have.

Part of me wonders if an adoption study is really the best way to deal with this. Giving your children up for adoption is a very unusual choice, which means the biological parents are a very nonrepresentative group – and the study indeed finds that even forty years later, these biological parents have only a third as much money as the average Swede. If the same factors that cause them to give their children up for adoption – illness, relationship problems, trouble with the law – also cause them to fail to live up to their "genetic potential", then we wouldn't expect their children (who may lack these issues) to be correlated with them. The extremely odd shape of the graph also gives me pause: after a certain point, the wealthier your biological parents were, the *less* likely you are to be wealthy. Why? Certainly there's no such effect for adoptive parents or non-adopted people!

Figure 2b: Within-Cohort Wealth Rank Relationship between Parents and Children and Their Adoptive Parents

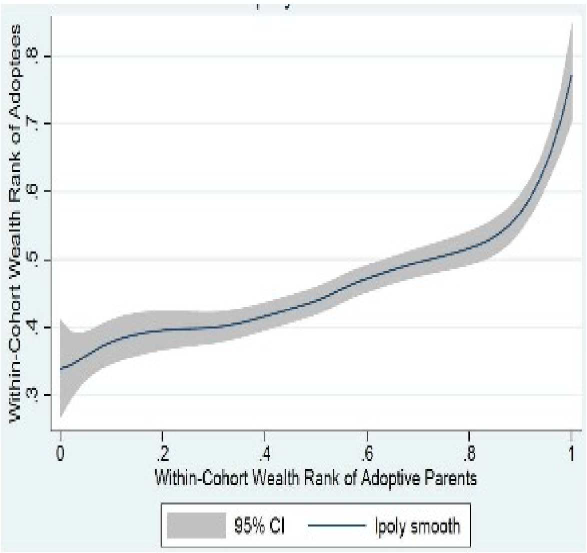
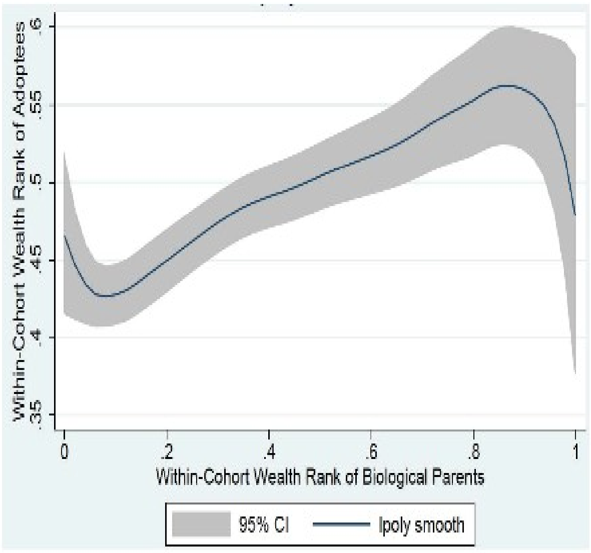


Figure 2a: Within-Cohort Wealth Rank Relationship between Adopted Children and Their Biological Parents



But nitpicks aside, I am pretty willing to believe this. Although other studies have found evidence that biology is more important than upbringing in determining income (not wealth!), wealth seems like a different story. For one thing, you can just *give* your kids money! As I said last time we talked about GiveDirectly, there is pretty good evidence that giving people money causes them to in fact have the money which you just gave them. The current study reasonably tries to avoid having to deal with inheritances by looking at people whose parents are still alive, but even living parents can give lots of money to their children (for example, I come from a pretty wealthy family and my parents gave me lots of money, which I mostly used to help get through medical school without much debt. This means right now I have more "wealth" than people who took out bigger loans).

The authors write that:

> While we have established the relative role of nature versus nurture, the exact mechanisms of wealth transmission are more deifficult to ascertain. Wealthier parents tend to be better educated and earn higher incomes, and these factors could lead to the increased wealth of their children through, for example, teaching them about investment opportunities or providing the right opportunities. However, when we investigate this, we find little evidence that this is the case. It may also be that wealthy parents invest more in their child's education and career, which could then lead to higher child wealth accumulation. When we examine whether this is the case, however, we find little evidence for education or income as mechanisms. So the pathway through which parental wealth affects child wealth does not appear to be primarily parental schooling and income or child human capital accumulation and greater labor earnings. Taken together, our findings suggest potential roles for intergenerational transmission of preferences (children of wealthier parents may choose to

> save more or invest in assets that have higher returns) or for financial gifts from parents to children. Unfortunately, we do not have information on savings behavior or on financial gifts so this evidence is only suggestive.

So it seems to be a matter of how much money your parents give you, rather than of you learning deeply important personality traits from them or something. Fair enough.

But I got distracted. I was talking about the Atlantic's article about the study. What did they have to say?

> Even when they're adopted, the children of the wealthy grow up to be just as well-off as their parents.
>
> Lately, it seems that every new study about social mobility further corrodes the story Americans tell themselves about meritocracy; each one provides more evidence that comfortable lives are reserved for the winners of what sociologists call the birth lottery…What appears to matter—a lot—is environment, and that's something that can be controlled.

III.

Let's talk about three things – correlation, percent variance explained, and reality.

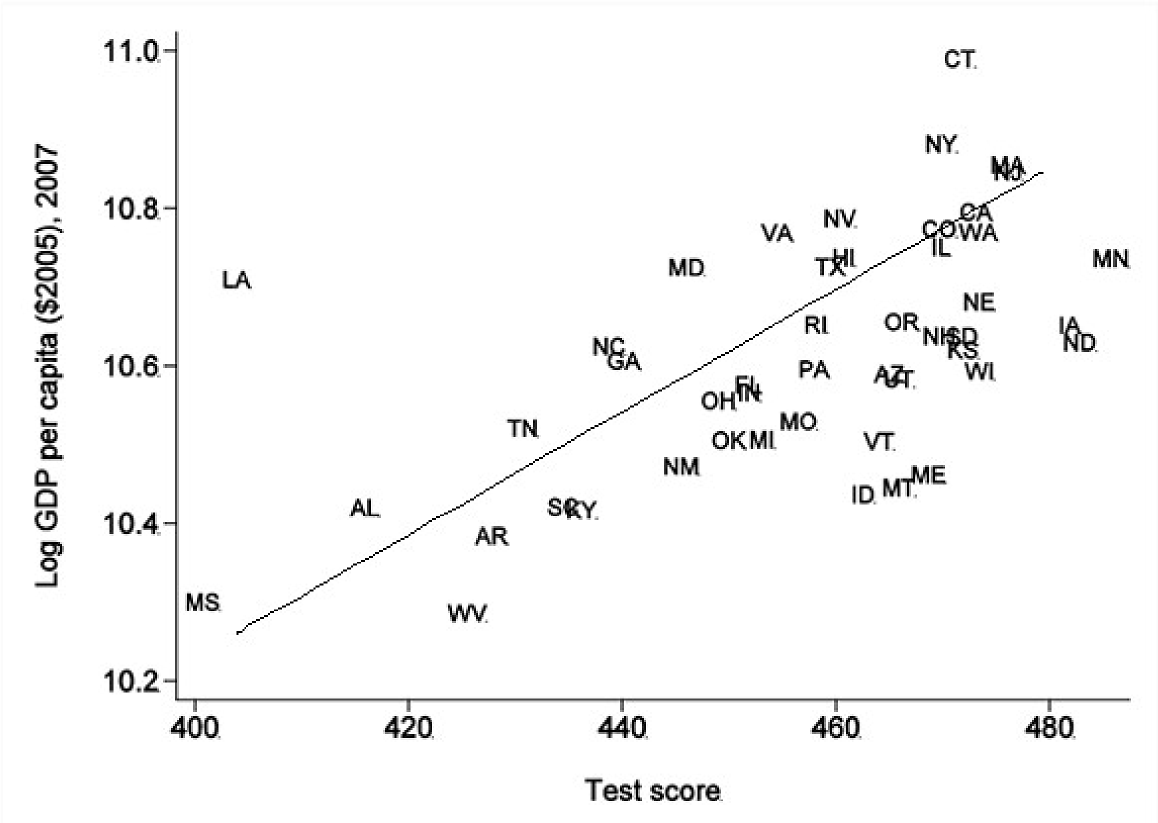
(I'm talking a big talk here, but I only got a good feeling for this when I asked various people on Tumblr to explain it to me. But they did a good job, and now I'm explaining it to you.)

Correlation is an r value. Percent variance explained is correlation squared. Reality is best viewed in the form of a graph.

Noah tells us that the IQ-of-states study found that only about 14% of the variation in state GDP was explained by IQ. Since variance = correlation^2, this implies that there's a correlation of sqrt(0.14) = 0.37 between state IQ and state GDP. The paper itself did some sort of super high-powered nuclear statistics to arrive at this estimate, but I took lists of state average aptitude test scores and state GDP per capita and correlated them together in SPSS and got 0.40, so easy way and hard way agree pretty closely.

Here's the graph associated with the study (I added the line):

Figure 8: Cognitive Skills and GDP per Capita across U.S. States, 2007



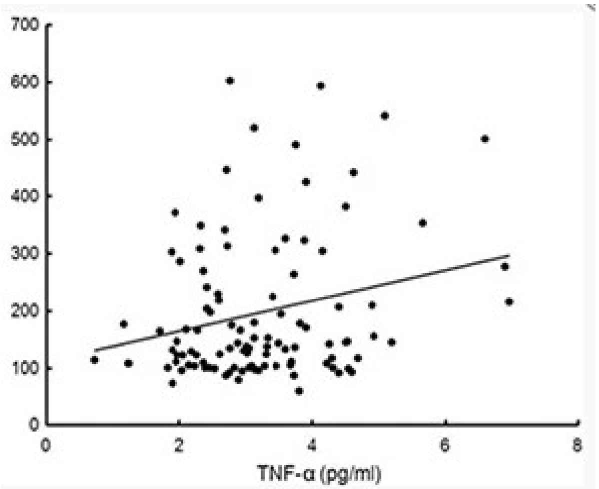
(Proposed new state motto: "Louisiana – Where We Succeed Wildly Out Of Proportion To Our Low Intelligence!")

Huh! When you hear “…only explained 14% of the variance” it sounds like “go home, this is boring,” but when you hear “correlation of 0.37”, it sounds like “huh, they seem pretty related”, and when you see the graph, it looks like “holy frick, everything is IQ after all”. But all of these are the same finding!

Now. Consider the Swedish study and the Atlantic article about it. They say that although biological parents were correlated at r = 0.13, adoptive parents were correlated at r = 0.23. Therefore, they conclude, nurture wins over biology, meritocracy is a myth, everything depends on the lottery of birth, and wealthy parents are foredoomed to have wealthy children.

But r = 0.23 means the percent of variance explained is 0.23^2 = ~5%. If some Social Darwinist organization were to announce that they had evidence that who your parents were only determined 5% of the variance in wealth, it would sound like such overblown strong evidence for pure meritocracy that everyone would assume they were making it up.

The study didn’t come with a scatter plot, but here’s a plot from a totally different study that got a very similar correlation (0.24) to give you a feel for what it might look like:



The article makes it sound like your position in the birth lottery determines your destiny with impressive finality. The correlation seems unimpressive. The variance seems *really* unimpressive. The scatter plot looks like someone took random noise and drew a line through it. Once again, all the same finding.

Which of these three ways of presenting the data is most accurate? Um. Hard to say. I asked some people whether correlation (ie r = 0.23) or variance (ie 5%) is a better description of how the world actually works. That is, given that I have a certain “feel” for how much people differ in wealth, and a certain “feel” for what it means to win the birth lottery by getting rich parents, should I feel like the birth lottery thing explains 23% of how much wealth you have, or only 5%?

(I was only a Discordian for like six months, in my freshman year of college, but I *still* end up with fives and twenty-threes every time I try to do something involving numbers)

The answers I got were that it’s complicated, and both sort of work even though intuitively they should be mutually contradictory. The distribution of wealth is consistent with a story where it is explained by twenty different factors, each of which is just as important as parental wealth, which is sort of like 5%. But parental wealth explains just over a fifth of the standard deviation in wealth, which is sort of like 23%. The best explanation I got, from an anonymous commenter, was this:

> About variance: consider the following. Flip 25 coins. Each heads gives you +1 utility point, and each tails gives you -1 utility point. One of these coins is labeled "upbringing". On average you get 0 utility points. But you can also expect not to get exactly 0: on average, your distance from 0 will be 5 (the stdev is 5). So this is a little similar to a single coinflip that gives you either +5 or -5. Changing your upbringing from -1 to 1 gives you 2 points, out of a typical range of -5 to 5.

There was also a general consensus that if I had to think about this intuitively, which I should try not to do, 5% was probably the number that would lead me less astray, at least in terms of inputs. So fine. Whatever. Five percent it is.

IV.

Stalin once said that "The people who cast the votes decide nothing. The people who count the votes decide everything."

(I briefly questioned whether Stalin really said that – like, I know he was an evil despot, but I'm not sure he was sufficiently *self-aware* about being an evil despot to come up with witty evil-despotism-related quotes. But I checked his WikiQuote page, and not only is the saying well-attested, but it seems Stalin was *totally all about* coming up with the witty self-aware evil-despotism-related quotes. Huh.)

In the same way, the people who conduct a study decide nothing. The people who report on the study decide everything.

I think Noah and the Atlantic were both honest and did a decent job reporting on their individual studies. But taken together, Noah concluded “This shows that IQ doesn’t really matter that much in explaining GDP” and the Atlantic concluded “This shows that who your parents are matters a colossally huge amount in explaining wealth” when in fact if you put both the studies side-by-side the IQ finding is three times as strong as the parents finding.

[**EDIT**: Some people have been misunderstanding this, so let me say it clearly. These are two studies about two different things! It’s like if I said the percent of weight gain explained by carbohydrates is three times as large as the percent of crime explained by poverty. I can compare these two things statistically, but I’m not trying to combine them into a single meta-study where I say that carbohydrates cause more crime than poverty! Also, some people seem to think I’m saying the Swedish study finds genes/biology/IQ to be more important than nurture. It doesn’t – in fact, it finds the opposite! Nurture is more important than genes but in the grand scheme of things *both are tiny and the variance is almost entirely due to other things or randomness*.]

In the end, nobody except a handful of researchers is going to remember the exact number. But they might remember “There were a couple interesting studies recently, one of them proved state IQ didn’t matter, the other proved that who your parents are totally determines whether you get ahead in life.” Framed that way, you might actually have gained *negative* knowledge from your diligent attempt to understand the economic literature.

And if the surrounding culture is pretty united in wanting to push a specific line, by choosing whether to publish r values or percent variance explained or graphs, they can pretty much hijack the intuitions even of people who don’t accept their reporting and try to rely on the numbers themselves.

The antidote is to have a good grasp of what each statistic means. And another antidote is to dial down your expectations. Remember, the study above was only able to correlate state IQ and state GDP at r = 0.4, but *almost nothing in social science ever gets above 0.4*. Trying to correlate rich parents with kids who become rich only got 0.2! 0.4 is pretty impressive and if you’re holding out for too much more you’re going to be living in a constant state of disappointment. I can think of one exception off the top of my head, and I am proud to say you will only find it here.

Contra Hallquist On Scientific Rationality

I.

Topher Hallquist recently wrote a long article calling Less Wrong and the rationalist community "against scientific rationality" and having "crackpot tendencies".

The piece claims to be about "the Less Wrong community", but mostly takes the form of a series of criticisms against Eliezer Yudkowsky for holding beliefs that Hallquist thinks are false or overconfident. In some respects this is fair; Eliezer was certainly the founder of the community and his writings are extremely influential. In other respects, it isn't; Margaret Sanger was an avowed eugenicist, but this is a poor criticism of Planned Parenthood today, let alone the entire reproductive rights community; Isaac Newton believed that the key to understanding the secrets of the universe lay in the dimensions of Solomon's Temple, but this is a poor critique of universal gravitation, let alone all of physics. I worry that Hallquist's New Atheism background may be screwing him up here: to critique a movement, merely find the holy book and prophet, prove that they're fallible, and then the entire system comes tumbling to the ground. Needless to say, this is not how things work outside the realm of divine revelation.

On the other hand, it seems like the same argument that suggests Hallquist shouldn't say such things would suggest I shouldn't care much about arguing against them. I wish I lived in a universe where this was true, but "guilt by association" is a thing, the Internet has more than its share of people who have conceived this deep abiding hatred toward all rationalists, and "crackpot" and "anti-intellectual" are especially sticky accusations around these parts. Past experience tells me if I let this slide then at some point I'm going to be mentioning I'm interested in rationality and the automatic response will be "Oh, those are those anti-intellectual crackpots who hate science" and nothing I say will convince them that they are wrong, because why listen to an anti-science crackpot? Some things need to be nipped in the bud.

Also, a lot of Hallquist's criticism is genuinely wrong and unfair. Also also, I *like* Eliezer Yudkowsky.

This is not to say that Eliezer – or anyone on Less Wrong – or anyone in the world – is never wrong or never overconfident. I happen to find Eliezer overconfident as heck a lot of the time. I have told him that, and he has pointed me to his essay on how if you really understand what confidence means in a probabilistic way, then you keep track of your uncertainty internally but don't worry too much about the social niceties of broadcasting how uncertain you are to everyone. My opinion of this is the same as my opinion of most other appeals to not needing to worry about social niceties.

If Hallquist had made this reasonable critique, I would have endorsed it. Instead, I find his critique consistently misrepresents Eliezer, most of the ideas involved, and the entire Less Wrong community. I am going to fisk it *hard*, which I don't like to do, but which seems like the only alternative to allowing these misrepresentations to stand. If you want to skip the (very boring) fisking, avoid part II and go straight from here to part III.

II.

So, to start with, I count four broad Eliezer-critiques:

1. Eliezer believes there’s an open-and-shut case for the Many Worlds Interpretation of quantum mechanics.
2. Eliezer believes that some philosophical problems are really easy, and philosophers are morons for not having settled them already
3. Eliezer is certain that cryonics will work
4. Eliezer believes that “dietary science has killed millions of people” and cites borderline-crackpot Gary Taubes

A.

The first critique, about quantum mechanics, is potentially the strongest. Yudkowsky’s support for his preferred Many Worlds interpretation is remarkably strong for a subject where some of the world’s top geniuses disagree. On one hand, it is no stronger than that of many experts in the field – for example, Oxford quantum pioneer David Deutsch describes the theory’s lukewarm reception as “the first time in history that physicists have refused to believe what their reigning theory says about the world…like Galileo refusing to believe that Earth orbits the sun” and calls arguments against it “complex rationalizations for avoiding the most straightforward implications of quantum theory”. On the other, perhaps one could argue that a level of confidence appropriate in an Oxford professor is inappropriate in a self-taught amateur. I don’t know anything about quantum mechanics and don’t want to get into it.

Neither does Hallquist; he admits that Many Worlds is a reasonable position, but makes a different accusation. He says Yudkowsky has failed to inform readers to investigate other views:

> Years ago in college I wrote a book debunking claims by Christian apologists claiming to “prove” that Jesus rose from the dead using historical evidence. At the very end

> of the book I included a little paragraph about how you shouldn't just take my word for anything and should do your own research and form your own conclusions. In retrospect, that paragraph feels cheesy and obvious, but seeing the alternative makes me glad I included it.
>
> Yudkowsky could have, after arguing at length for the many worlds interpretation of quantum mechanics, said, "I recommend going and studying the arguments of physicists who defend other interpretations, and when you do that I think you'll see that physicists are screwing up." That might have been reasonable. Many physicists accept many worlds, and I can accept that it's sometimes reasonable for a dedicated amateur to have strong opinions on issues that divide the experts.

But in the *very first* post of his quantum physics sequence, Eliezer warns:

> Everyone should be aware that, even though I'm not going to discuss the issue at first, there is a sizable community of scientists who dispute the realist perspective on QM. Myself, I don't think it's worth figuring both ways; I'm a pure realist, for reasons that will become apparent. But if you read my introduction, you are getting my view. It is not only my view. It is probably the majority view among theoretical physicists, if that counts for anything (though I will argue the matter separately from opinion polls). Still, it is not the only view that exists in the modern physics community. I do not feel obliged to present the other views right away, but I feel obliged to warn my readers that there are other views, which I will not be presenting during the initial stages of the introduction.

Okay, but that’s not *exactly* the same thing as telling his readers that they need to read other books and do other research and see whether he’s right or wr…

> Go back and look at other explanations of QM and see if they make sense now. Check a textbook. Alternatively, check Feynman’s QED. Find a physicist you trust, ask them if I got it wrong, if I did post a comment. Bear in mind that a lot of physicists do believe MWI.

That’s from this comment. So Yudkowsky clearly did the thing Hallquist is accusing him of not doing.
This is to his credit, but in my opinion supererogatory. Despite his anecdote about the paragraph in his book, Hallquist does not end his own essay on Less Wrong with “There are many people who disagree with me on this, be sure to check some pro-Less Wrong views to get the opposite side of the story.” Nor does he include such a sentence in the vast majority of his voluminous blog posts and writings. I don’t blame him for that – I don’t use such disclaimers either. There is a general conversational norm that to assert something is to say that you believe it and will provide evidence for your belief, not to say “You must believe me on this and may never go find any other sources to get any other side of the story.” You listen to me, you get my perspective, if for some reason I think you might not be aware that other perspectives exist I’ll tell you that they do, but I’m not going to end every blog post with “But be sure to read other sources on this so you can have your own opinion”.

Nevertheless, that is the standard Hallquist demands of Yudkowsky. And Yudkowsky meets it. And Hallquist ignores him doing so.

B.

Here’s philosophy. Note that I’m editing liberally to reduce length; I don’t think I’ve removed any essential parts of the argument but you might want to go over there and check:

> In another post, Yudkowsky lodges a similar complaint about philosophy:
>
> Philosophy is just not oriented to the outlook of someone who needs to resolve the issue, implement the corresponding solution, and then find out – possibly fatally – whether they got it right or wrong. Philosophy doesn’t resolve things, it compiles positions and arguments. And if the debate about zombies is still considered open, then I’m sorry, but as Jeffreyssai says: Too slow! It would be one matter if I could just look up the standard answer and find that, lo and behold, it is correct. But philosophy, which hasn’t come to conclusions and moved on from cognitive reductions that I regard as relatively simple, doesn’t seem very likely to build complex correct structures of conclusions.
>
> I agree that progress in academic fields is sometimes slowed by the irrationality of the participants. People don’t like admitting being wrong. Unfortunately, knowing this isn’t much help unless you’ve discovered a magic formula for overcoming this flaw in human nature. More on that later. But thinking irrationality is the only reason why progress is slow ignores the fact that often, progress is slow because the questions are just really hard
>
> […]
>
> The reason I’m saying all this is because when philosophers act like they’re not really trying to resolve debates, it’s because they know such attempts have a

track record of not working. That doesn't mean we will never put philosophy on a solid footing, but it does mean that anyone who shows up claiming to have done so single-handedly deserves a fair dose of skepticism.

The zombie debate is as good an example as any of this, so let's talk about that. David Chalmers' claim is that there could exist (in other possible worlds with different psychophysical laws–not the actual world) beings that are physically identical to us, but who lack consciousness. The intuition that such "zombies" are possible leads Chalmers to a view that at least looks a lot like epiphenomenalism (the belief in a separate mental realm affected by, but which does not affect, the physical realm).

Epiphenomenalism strikes a lot of people as crazy–me included! But Chalmers realizes this. So in The Conscious Mind he tries to do two things (1) argue his view is not quite epiphenomenalism (2) argue that some of the apparent advantages of certain other views over epiphenomenalism are illusory.

Does he succeed? I don't know. But what makes me sympathetic to Chalmers is the sense that what he calls the hard problem of consciousness is a real problem, and alternative solutions aren't any better. And Yudkowsky, as far as I can tell, isn't one of those people who says, "the so-called 'hard problem' is a fake problem." He agrees that it's real–and then claims to have a secret solution he'll sell you for several thousand dollars.

I think it's enormously unlikely that Yudkowsky has really found the secret solution to consciousness. But even if he had, I don't think anyone could know,

> including him. It's like an otherwise competent scientist refusing to submit their work for peer review. Even top experts are fallible–and the solution is to have other experts check their work.

My response is going to sound mean and kind of ad hominem. That's not really where I'm going here, and I promise I'm going somewhere a little more subtle than that. So, keeping that in mind – Hallquist once said on his blog:

> I seem to have some philosophy-related skills in abundance. I'm good at spotting bad philosophical arguments. I was exposed to Plantinga's modal ontological argument at ~12 years old, and instantly noticed it could just as well "prove" the existence of an all-powerful, necessarily existent being who wants nothing more than for everything to be purple.My experience with philosophy professors suggests that sadly, the knack for seeing through bad arguments is far from universal, even among the "professionals."

What is the difference between Hallquist believing that he disproved one of the world's most famous philosophers when he was twelve years old, and Eliezer believing that he solved the problem of consciousness when he was thirty-something?

Likewise, in another blog post Hallquist wrote about how Aquinas' Arguments Just Suck and have no redeeming value:

> Now, the broad point here is that while these arguments are supposedly derived from Aristotle, there doesn't seem to be some secret Aristotelian assumption that would make them work. They're just plain old bad arguments. I feel comfortable saying this, because

> respected living philosophers often give arguments that just stink, and being a contemporary of those philosophers I'm confident that the issue isn't some peculiarly 21st century assumption.

Remember that Aquinas' arguments convinced nearly all the brightest people in the Western world for five hundred years, and that many PhD philosophy professors still believe them with all their heart. Hallquist says they *just suck* and that there is no chance he might just be missing something or be misunderstanding their terms. In fact, he was so sure that there was no chance he was just a modern misunderstanding things that when a philosophy professor wrote a book claiming that modern people's contempt for Aquinas is based on a misunderstanding, Hallquist read twenty pages of it, decided there was no chance that the remainder could possibly contain anything to change his mind, and stopped.

(I read the entire book, and if I'd stopped at page twenty I would have saved myself several valuable hours of my life I could have spent doing something productive, BUT THAT'S NOT THE POINT!)

Presumably, if Aquinas' arguments are really stupid, but everyone believed them for five hundred years, this would imply there is something wrong with everyone. It might be worth saying "When someone makes a stupid argument, is there anything we can do to dismiss it in *less than* five hundred years?" But take that step, and you're venturing into exactly the territory Hallquist is criticizing Eliezer for crossing into.

So what is my point beyond just an ad hominem attack on Hallquist?

Philosophy is hard. It's not just hard. It's hard in a way such that it seems easy and obvious to each individual person involved. This is not just an Eliezer Yudkowsky problem, or a Topher Hallquist problem – I myself may have previously said something along the lines of that anybody who isn't a consequentialist needs to have their head examined. It's hard in the same way politics is hard, where it seems like astounding hubris to call yourself a liberal when some of the brightest minds in history have been conservative, and insane overconfidence to call yourself a conservative when some of civilization's greatest geniuses were liberals. Nevertheless, this is something everybody does. I do it. Eliezer Yudkowsky does it. Even Topher Hallquist does it. All we can offer in our own defense is to say, with Quine, "to believe something is to believe that it is true". If we are wise people, we don't try to use force to push our beliefs on others. If we are very wise, we don't even insult and dehumanize those whom we disagree with. But we are allowed to believe that our beliefs are true. When Hallquist condemns Yudkowsky for doing it, it risks crossing the line into an isolated demand for rigor.

The most charitable I can be to Hallquist is that his gripe is not with Yudkowsky's hubris in holding strong beliefs, but with an apparent reluctance to publish them. Once again, from the end of the quoted material:

> What makes me sympathetic to Chalmers is the sense that what he calls the hard problem of consciousness is a real problem, and alternative solutions aren't any better. And Yudkowsky, as far as I can tell, isn't one of those people who says, "the so-called 'hard problem' is a fake problem." He agrees that it's real–and then claims to have a secret solution he'll sell you for several thousand dollars.

> I think it’s enormously unlikely that Yudkowsky has really found the secret solution to consciousness. But even if he had, I don’t think anyone could know, including him. It’s like an otherwise competent scientist refusing to submit their work for peer review. Even top experts are fallible–and the solution is to have other experts check their work.

So perhaps the difference between Hallquist (and the rest of us) and Yudkowsky is that the latter doesn’t believe in peer review and openness because he “hasn’t published” his “secret” solution to consciousness?

Hallquist’s only source for Eliezer having such a solution is the Author Notes for Harry Potter and the Methods of Rationality, Chapter 98, where somewhere in the middle he says:

> I am auctioning off A Day Of My Time, to do with as the buyer pleases – this could include delivering a talk at your company, advising on your fiction novel in progress, applying advanced rationality skillz to a problem which is tying your brain in knots, or confiding the secret answer to the hard problem of conscious experience (it’s not as exciting as it sounds).

Even if, as Hallquist claims, he has private information that this isn’t a joke, this hardly seems like the central case of Eliezer having a philosophical opinion. One might ask: has Eliezer ever written about his other philosophical beliefs, the ones that seem most important to him?

The Less Wrong Sequences are somewhere between 500,000 and a million words long (for comparison, all three *Lord of the Rings* books combined are 454,000, and *War and Peace* is 587,000). Eliezer may be one of the most diligent people alive about publicizing his philosophical opinions in great detail. In some cases, he explicitly frames his Sequence work in terms of getting feedback – for example, in the same comment on the Quantum Physics Sequence I linked earlier, he writes about his approach to double-checking his quantum work:

> I myself am mostly relying on the fact that neither Scott Aaronson nor Robin Hanson nor any of several thousand readers have said anything like “Wrong physics” or “Well, that’s sort of right, but wrong in the details…”

And this works – far from being unwililng to debate academic philosophers about his opinion of Chalmers’ epiphenomenalism, he argues the subject with David Chalmers himself in the comments of that Less Wrong post, which is a heck of a lot more than I’ve ever done when I disagree with an academic philosopher.

On the other hand, since he placed a single sentence in an HPMOR Author’s Note about a solution to the hard problem of consciousness which he hasn’t written about, Hallquist accuses him of being against publicizing his work for review or criticism. Once again, I think these complaints are startlingly unfair and a remarkably flimsy ground on which to set out to tarnish the reputation of an entire community.

C.

Hallquist’s next complaint is that Eliezer is strongly in favor of cryonics; he starts by noting that Eliezer criticizs Michael Shermer for mocking the role of molecular nanotechnology in cryonics, then writes:

> Full disclosure: I’m signed up for cryonics. But the idea that nanomachines will one day be able to repair frozen brains strikes me as highly unlikely. I think there’s a

better chance that it will be possible to use frozen brains as the basis for whole brain emulation, but I’m not even sure about that. Too much depends on guesses both about the effects of current freezing techniques and about future technology.

Eliezer, meanwhile, is sure cryonics will work, based, as far as I can tell, on loose analogies with computer hard drives. Faced with such confident predictions, pointing out the lack of evidence and large element of wish-fulfillment (as Shermer does) is an eminently reasonable [reaction]. A “rationalism” that condemns such caution isn’t worthy of the name.

I can’t stress enough the enormous difference between trying to do some informed speculation about what technologies might be possible in the future, and thinking you can know what technologies will be possible in the future based on just knowing a little physics. Take, for example, Richard Feynman’s talk “There’s Plenty of Room at the Bottom”, often cited as one of the foundational sources in the field of nanotechnology.

Today, the part of Feynman’s talk about computers looks prophetic, especially considering the talk was given several years before Gordon Moore made his famous observation about computer power doubling every couple of years. But other things he speculates about are, to say the least, a long way off. Do we blame Feynman for this?

No, because Feynman knew enough to include appropriate caveats. When he talks about the possibility of tiny medical robots, for example, he says it’s “very interesting possibility… although it is a very wild idea.”

> He doesn’t say that this will definitely happen and be the secret to immortality. And some futurists, like Ray Kurzweil, do say things like that. That’s the difference between having a grasp of the difficulty of the topic and your own fallibility, and well, not.
>
> Yet Yudkowsky is so confident in his beliefs about things like cryonics that he’s willing to use them as a reason to distrust mainstream experts.

I hate to keep getting into these *tu quoque* style arguments. But I do feel like once you have signed up for cryonics, you lose the right to criticize other people for being crackpots for being signed up for cryonics for a slightly different reason than you are. If you think molecular nanotechnology is highly unlikely but whole brain emulations aren’t, please at least be aware that from the perspective of everyone else in society you are the equivalent of a schizophrenic guy who believes he’s Napoleon making fun of the other schizophrenic guy who thinks he’s Jesus.

But Hallquist’s main criticism isn’t *just* that Yudkowsky believes in cryonics, or nanotechnology, or whatever. It’s that he’s overconfident in these things. As per the quote, he’s “*sure* cryonics will work” and says the equivalent of “this will definitely happen and be the secret to immortality”.

At this point, it will probably come as no surprise that Yudkowsky has never said anything of the sort and has in fact clearly said the opposite.

One of my favorite results from the Less Wrong Survey, which I’ve written about again and again, shows that people who sign up for cryonics are *less* likely to believe it will work than demographically similar people who don’t sign up (yes, you read that right) – and the average person signed up for cryonics only estimated a 12% chance it would work. The active ingredient in cryonics support is not unusual certainty it will work, but unusual methods for dealing with moral and epistemological questions – an attitude of “This only has like a 10% chance of working, but a 10% chance of immortality for a couple of dollars a month is an amazing deal and you would be an idiot to turn it down” instead of “this sounds weird, screw it”. Once you think of it that way, signing up doesn’t mean “I’m sure it will work” but rather “I’m not 100% sure it won’t.” Thus, Eliezer’s former co-blogger Robin Hanson, who frequently joined forces with Eliezer in passionate appeals for their mutual readers to sign up for cryonics, gives his probability of cryonics working at 6%.

I don’t have as accurate a picture of Eliezer’s beliefs. The best I can do is try to interpret this comment, where he describes the success of cryonics as a chain of at least three probabilities. First, the probability that the core technology works. Second, the probability that cryonics organizations don’t go bankrupt. Third, the probability that humankind lasts long enough to develop appropriate resurrection technology.

Eliezer gives the first probability as 80 – 90%, the second probability as “outside the range of my comparative advantage in predictions”, and the third probability as “the weakest link in the chain” and as “something I’ve refused to put a number on, with the excuse that I don’t know how to estimate the probability of doing the ‘impossible'”. So his probability of cryonics working seems to be 80-90% * something-he-doesn’t-know * something-he-thinks-is-near-impossible. If we are willing to totally ignore his request not to put numbers on these, perhaps something like 90% * 75% * 10%? Which would land him right between Robin Hanson’s 6% and the survey average of 12%.

I think Eliezer's overconfident on his belief on the first of his three bins, the one about the core technologies working. But Hallquist ignores this relatively modest issue and instead chooses to sensationalize Eliezer into being "sure cryonics will work" and "[thinking] this will definitely happen", when his *actual* probability is unknown but probably closer to 6% than 100%. Once again I think this complaint is startlingly unfair and a remarkably flimsy ground on which to set out to tarnish the reputation of an entire community.

[my own disclaimer: I am not signed up for cryonics and don't currently intend to, but I respect people who are]

D.

Oh God, we're going to have to get into diet again. You can skip this part. Seriously, you can.

Eliezer writes that "dietary scientists ignoring their own experimental evidence have killed millions and condemned hundreds of millions more to obesity with high-fructose corn syrup" and links to this article by Taubes.

I do not want to defend Gary Taubes. Science has progressed to the point where we have been able to evaluate most of his claims, and they were a mixture of 50% wrong, 25% right but well-known enough that he gets no credit for them, and 25% right ideas that were actually poorly known enough at the time that I *do* give him credit. This is not a bad record for a contrarian, but I subtract points because he misrepresented a lot of stuff and wasn't very good at what might be called scientific ethics. I personally learned a lot from reading him – I was able to quickly debunk the wrong claims, and the correct claims taught me things I wouldn't have learned any other way. Yudkowsky's reading of him seems unsophisticated and contains a mix of right and wrong claims. But Hallquist's reading seems to be a prime example of reversed stupidity not being intelligence. He writes:

> The use of the diet example is even more embarrassing than the other claims I've looked at so far. The line about "dietary scientists ignoring their own experimental evidence" links to an article by Gary Taubes. Taubes champions the diet claims of Robert Atkins, who literally claimed that you could eat unlimited amounts of fat and not gain weight, because you would pee out the excess calories. This, needless to say, is not true.
>
> After reading two of Taubes' books, I haven't been able to find anywhere where he addresses the urine claim, but he's very clear about claiming that no amount of dietary fat can cause weight gain. How Taubes thinks this is supposed to be true, I have no idea. His attempted explanations are, as far as I can tell, simply incoherent. (Atkins at least had the virtue of making a coherent wrong claim.)

I'm startled by Hallquist's claim that he read two Taubes books and couldn't find Taubes' explanation for why people don't gain calories on a high-fat diet. Taubes lays out this mechanism very clearly as the major thesis of his book.

Taubes believes the human body is good at regulating its own weight via the hunger mechanism. For example, most Asian people are normal weight, despite the Asian staple food being rice, which is high-calorie and available in abundance. Asians don't get fat because they eat a healthy amount of rice, then stop. This doesn't seem to require amazing willpower on their part; it just happens naturally.

In a similar vein is one of Taubes' favorite studies, the Vermont Prison Experiment, where healthy thin prisoners were asked to gain lots of weight to see if they could do it. The prisoners had lots of trouble doing so – they had to force themselves to eat even after they were full, and many failed, disgusted by the task. Some were able to eat enough food, only to find that they were filled with an almost irresistible urge to exercise, pace back and forth, tap their legs, or otherwise burn off the extra calories. Those prisoners who were able to successfully gain weight lost it almost instantly after the experiment was over and they were no longer being absolutely forced to maintain it. The conclusion was that healthy people *just can't gain weight even if they want to*, a far cry from the standard paradigm of "it takes lots of willpower not to gain weight".

Other such experiments focused on healthy thin rats. The rats were being fed as much rat food as they wanted, but never overate. The researchers tried to trick the rats by increasing the caloric density of the rat food without changing the taste, but the rats just ate less of it to get the same amount of calories as before. Then the researchers took the extreme step of surgically implanting food in the rats' stomachs; the rats compensated by eating precisely that amount less of normal rat food and maintaining their weight. The conclusion was that rats, like Asians and prisoners, have an uncanny ability to maintain normal weight even in the presence of unlimited amounts of food they could theoretically be binging on.

Modern Westerners seem to be pretty unusual in the degree to which they lack this uncanny ability, suggesting something has disrupted it. If we can un-disrupt it, "just eat whatever and let your body take care of things" becomes a passable diet plan.

I sometimes explain this to people with the following metaphor: severe weight gain is a common side effect of psychiatric drug Clozaril. The average Clozaril user gains fifteen pounds, and on high doses fifty or a hundred pounds is not unheard of. Clozaril is otherwise very effective, so there have been a lot of efforts to cut down on this weight gain with clever diet programs. The journal articles about these all find that they fail, or “succeed” in the special social science way where if you dig deep enough you can invent a new endpoint that appears to have gotten 1% better if you squint. This Clozaril-related weight gain isn’t magic – it still happens because people eat more calories – but it’s not something you can just wish away either.

Imagine that some weird conspiracy is secretly dumping whole bottles of Clozaril into orange soda. Since most Americans drink orange soda, we find that overnight most Americans gain fifty pounds and become very obese.

Goofus says: “Well, it looks like Americans will just have to diet harder. We know diets rarely work, but I’m sure if you have enough willpower you can make it happen. Count every calorie obsessively. Also, exercise.”

Gallant says: “The whole problem is orange soda. If you stop drinking that, you can eat whatever else you want.”

Taubes’ argument is that refined carbohydrates are playing the role of Clozaril-in-orange-soda. If you don’t eat refined carbohydrates, your satiety mechanism will eventually go back to normal just like in Asians and prisoners and rats, and you can eat whatever else you want and won’t be tempted to have too much of it – or if you do have too much of it, you’ll exercise or metabolize it away. When he says you can “eat as much fat as you want”, he expects that not to be very much, once your broken satiety mechanism is fixed.

Taubes is wrong. The best and most recent studies suggest that avoiding refined carbohydrates doesn't fix weight gain much more than avoiding any other high-calorie food. However, the Clozaril-in-orange-soda model, which is not original to Taubes but which he helped popularize, has further gained ground and is now arguably the predominant model among dietary researchers. It's unclear what exactly the orange soda *is* – the worst-case scenario is that it's something like calorically-dense heavily-flavored food, in which case learning this won't be very helpful beyond current diet plans. The best-case scenario is that it's just a disruption to the microbiome, and we can restore obese people to normal weight with a basic procedure which is very simple and not super-gross at all.

But whether or not you agree with it, this Clozaril-in-orange-soda story is indisputably Taubes' model; Hallquist seems to miss this and instead makes vague gestures towards a discredited 70s theory that calories are excreted as ketones in the urine.

Because Hallquist doesn't understand Taubes' main point, his criticisms miss the mark. He wrote a five-part Less Wrong series on Gary Taubes, in which he tries to figure out what Taubes' theory of obesity is and as best I can tell somehow ends up simultaneously saying Taubes is dishonest for accusing mainstream researchers of thinking diet is a matter of willpower, *and* saying Taubes is silly because diet really *is* just a matter of willpower. If you don't believe me, read the post, where he says:

> Taubes goes on at great length about how obesity has other causes beyond simple calorie math as if this were somehow a refutation of mainstream nutrition science. So I'm going to provide a series of quotes from relevant

> sources to show that the experts are perfectly aware of that fact.

Which he does. But then later he says:

> So what’s going on here? I think the answer lies Taubes’ eagerness to portray mainstream nutrition experts as big meanies who blame fat people for being fat…
>
> But this puts Taubes in a bind: now if he says how much we eat has an effect on our weight, he’s a big meanie too. It doesn’t work for him to say fat people can’t help overeating because of something wrong with their metabolism, and this in turn causes them to gain weight, because he’s committed himself to the principle that blaming behavior equals blaming a character defect. So instead, we get wild rhetoric about how stupid the experts are with no coherent view underneath it.
>
> A more sensible approach would’ve been to emphasize that akrasia is an extremely common problem for humans, and that people who don’t suffer from akrasia in regards to diet probably suffer from akrasia about something else. But that wouldn’t have made for as an exciting of a book.

So unless I am reading this wrong, he thinks the correct answer is to say that we *should* blame behavior, but that we should do it in a nice way where we say kind things about how everybody has willpower problems and it’s nothing to be ashamed of. I think he thinks he’s agreeing with mainstream consensus here, but mainstream consensus has already moved on to “Screw willpower, STOP DRINKING ORANGE SODA”.

Next, Hallquist attacks Taubes' claim that "We don't get fat because we overeat; we overeat because we're getting fat", saying that he's "trying to be charitable" but "surely it wasn't meant to be taken literally" and that he's "playing with meanings" just so he can "portray nutrition experts as big meanies".

But when you ask those nutrition experts – for example, Dr. David Ludwig MD PhD Professor of Nutrition at Harvard, they're writing articles in the *Journal of the American Medical Association* with titles like Increasing Adiposity – Cause Or Consequence Of Overeating?, where they say things like:

> Since [the early part of the century], billions of dollars have been spent on research into the biological factors affecting body weight, but the near-universal remedy remains virtually the same, to eat less and move more. According to an alternative view, chronic overeating represents a manifestation rather than the primary cause of increasing adiposity. Attempts to lower body weight without addressing the biological drivers of weight gain, including the quality of the diet, will inevitably fail for most individuals…a focus on total diet composition, not total calories, may best facilitate weight loss.

And although the journal article is relatively balanced and might be dismissed as just a scholarly investigation of a weird ideas, the same authors wrote up the same argument for the New York Times in a more obviously persuasive fashion.

Hallquist, again thinking he's defending a consensus position, attacks Gary Taubes' claim that government guidelines promoting low-fat diets were responsible for the increase in sugar and refined carbohydrate consumption. He calls this “a huge red flag”, said Taubes is “unaware of [reality] or tries to hide it from his readers”, engages in “irresponsible rhetoric” and wonders “how anyone reading this could avoid suspecting something was up.”

Meanwhile, let’s go to more experts writing in the *Journal of the American Medical Association* – we’ll keep Ludwig but add Dr. Dariush Mozaffarian, MD, MPH, Dean of Tufts University School of Nutrition – explaining why the new dietary guidelines have done a sudden about-face, removed previous restrictions on fat, and added more restrictions on refined carbohydrates:

> With these quiet statements, the DGAC report reversed nearly 4 decades of nutrition policy that placed priority on reducing total fat consumption throughout the population. In 1980, the Dietary Guidelines recommended limiting dietary fat to less than 30% of calories. This recommendation was revised in 2005, to include a range from 20% to 35% of calories. The primary rationale for limiting total fat was to lower saturated fat and dietary cholesterol, which were thought to increase cardiovascular risk by raising low-density lipoprotein cholesterol blood concentrations. But the campaign against saturated fat quickly generalized to include all dietary fat. Because fat contains about twice the calories per gram as carbohydrate or protein, it was also reasoned that low-fat diets would help prevent obesity, a growing public health concern.
>
> The complex lipid and lipoprotein effects of saturated fat are now recognized, including evidence for beneficial effects on high-density lipoprotein cholesterol and triglycerides and minimal effects on apolipoprotein B when compared with carbohydrate. These complexities

> explain why substitution of saturated fat with carbohydrate does not lower cardiovascular risk.
>
> Most importantly, the policy focus on fat reduction did not account for the harms of highly processed carbohydrate (eg, refined grains, potato products, and added sugar)—consumption of which is inversely related to that of dietary fat…Based on years of inaccurate messages about total fat, a 2014 Gallup poll shows that a majority of US residents are still actively trying to avoid dietary fat, while eating far too many refined carbohydrates.

Hallquist reserves special mockery for Taubes' claim that the government's low-fat mania reached such intensity that nutritional initiatives promoted soda:

> This portrayal of mainstream nutrition science is as false as Atkins' claim about peeing out excess calories. Besides the obvious – who on earth ever believed Coca-Cola was a health food?

Once again, the United States' top nutrition scientists, explaining the new consensus dietary guidelines in the New York Times:

> The "We Can!" program, run by the National Institutes of Health, recommends that kids "eat almost anytime" fat-free salad dressing, ketchup, **diet soda** and trimmed beef, but only "eat sometimes or less often" all vegetables with added fat, nuts, peanut butter, tuna canned in oil and olive oil. Astoundingly, the National School Lunch Program bans whole milk, but allows sugar-sweetened skim milk. Consumers didn't notice, either. Based on years of low-fat messaging, most

> Americans still actively avoid dietary fat, while eating far too much refined carbohydrates.

Hallquist critiques Yudkowsky’s worry dietary scientists have been too soft on high-fructose corn syrup. Once again, the latest dietary guidelines almost halve the allowed amount of HFCS from the permitted amount in the last set of guidelines.

Hallquist derides Yudkowsky’s claims that dietary science has “killed millions”. An article by one of Britain’s top doctors, in the *British Medical Journal*, which by the way is the third highest-impact medical journal in the world, asks Are Some Diets Mass Murder? and argues that

> [A] consequence of the fat hypothesis is that around the world diets have come to include much more carbohydrate, including sugar and high fructose corn syrup, which is cheap, extremely sweet, and “a calorie source but not a nutrient.” More and more scientists believe that it is the surfeit of refined carbohydrates that is driving the global pandemic of obesity, diabetes, and non-communicable diseases. They dispute the idea that we get fat simply because energy in exceeds energy out, saying instead that the carbohydrates “trigger a hormonal response that drives the portioning of the fuel consumed as storage as fat… The successful attempt to reduce fat in the diet of Americans and others around the world has been a global, uncontrolled experiment, which like all experiments may well have led to bad outcomes

If indeed there were serious flaws in the dietary guidelines for the past thirty years, and since obesity kills about 370,000 people per year, if the issues corrected in the latest guidelines and freely admitted by modern scientists made the problem even 10% worse, then the “millions of deaths” figure is not an exaggeration.

Taubes is wrong about a lot of things. There is a lot of room for someone to criticize Taubes, and indeed, I have done so repeatedly. Hallquist tries to criticize Taubes, but fires so indiscriminately that he manages to reserve some of his strongest condemnation for claims of Taubes which are actually true and widely recognized as such.

Hallquist writes of Yudkowsky:

> Remember, this is one of Yudkowsky’s go-to examples for why you shouldn’t trust the mainstream too much! And it’s not just wrong, it’s wrong in a way that could have been caught through common sense and basic fact-checking. But I guess common sense is just tribalistic bias, and who needs fact-checking when you’ve got superior rationality? The nicest thing you can say about this is that, when he encourages his followers to form strong opinions based on the writings of a single amateur, he’s only preaching what he practices.

I am usually in favor of being nice to people who get things wrong, because things are hard and goodness knows I am wrong often enough. But I am not in favor of being nice to people who get things wrong and are smug and mean to everyone else about them, because punishing defectors is the only way things ever get done in this world. So:

Topher. You seriously do not understand Taubes. You somehow read his book while by your own admission missing the entire mechanism he was trying to explain. You then go on to call a bunch of propositions ludicrous, idiotic, not-even-wrong, et cetera, when those propositions are widely acknowledged as true by the scientific community you think you are defending. You are nevertheless setting yourself up as an expert and trying to explain these subjects to other people. Many of these people told you this when you first posted on LW, and you ignored those people and keep trying to do it. Mozaffarian, Ludwig, Friedman, etc are the United States' top nutritional scientists, and they are telling you this. I am telling you this. Everyone is telling you this, and you are putting your fingers in your ears and shouting "EVERYTHING IS SO OBVIOUS, I CAN'T BELIEVE OTHER PEOPLE GOT THIS WRONG, IT'S ALL SO EASY, EVERYONE EXCEPT ME IS AN IDIOT."

Eliezer Yudkowsky has had some pretty silly ideas about diet. I know this because he when he has them, he comes to me and asks me if they are correct, and I tell him. At one point, he bought and sent me a book he was interested in so that I could review it and tell him if it made sense. I told him it was wrong, and he listened. If you had asked me if you were right in your criticisms of Taubes, I would also have reviewed them and explained them to you. You didn't, because you were so certain that you had to be right that you didn't need to consult with anybody else, despite the fact that you are an amateur with no medical knowledge.

Thus is it written: "Why do you look at the mote in Eliezer Yudkowsky's eye, and ignore the beam in your own?"

The last I heard about Eliezer's dietary philosophy was his OKCupid profile, where under "Food" he wrote: "Flitting from diet to diet, searching empirically for something that works."

SUCH OVERCONFIDENCE. SO CERTAINTY. VERY ANTI-SCIENCE.

III.

Okay, now that I’ve gotten my nitpicks out of the way, what about the actual meat of Hallquist’s criticism?

Hallquist claims that Less Wrong is fundamentally anti-science. All of his criticisms of Eliezer Yudkowsky were to show examples of him behaving in anti-science ways, but he also thinks that Eliezer comes right out and admits it:

> Now that I’m thousands of words and about as many tangents into this post, let me circle back to something to something I said early in the post: pointing out the flaws in mainstream experts only gets you so far, unless you actually have a way to do better. This isn’t an original point. Robin Hanson has made it many times. (See here for just one example.) But I want to emphasize it anyway.
>
> It’s the main reason I’m unimpressed with the material on LessWrong about how the rules of science aren’t the rules an ideal reasoner would follow. This is a huge chunk of Yudkowsky’s “Sequences”, but suppose that’s true, so what? We humans are observably non-ideal. Throwing out the rules of science because a hypothetical ideal reasoner wouldn’t need them is like advocating anarchism on the grounds that if Superman existed, we’d have no need for police.
>
> I think this is more than a superficial analogy. To borrow another point from Hanson, most of us rely on peaceful societies rather than personal martial prowess for our safety. Similarly, we rely on the modern economy rather than personal survival skills for food and shelter. Given that, the fact that science is, to a large extent, a system of social rules and institutions doesn’t look like a flaw in science. It may be the only way for mere mortals to make progress on really hard questions.

> Yudkowsky is aware of this argument, and his response appears to mostly depend on assuming the reader agrees with him that physicists are being stupid about quantum mechanics–that, combined with a large dose of flattery. “So, are you going to believe in faster-than-light quantum ‘collapse’ fairies after all? Or do you think you’re smarter than that?” asks one post.
>
> This is combined with an even stranger argument, an apparent belief that it should be possible for amateurs to make progress faster than mainstream experts simply by deciding to make progress faster. Remember how the imagined future “master rationalist” complains “Eld scientists thought it was acceptable to take thirty years to solve a problem”? This is a strange thing to complain about. Either you have a way to make progress quickly or you don’t, and if you don’t, you don’t have much choice but to accept that fact.
>
> Back in the real world, wishing away the difficulty of hard problems don’t make them stop being hard. This doesn’t mean progress is impossible, or that’s it’s not worth trying to improve on the current consensus of experts. It just means progress requires a lot of work, which most of the time includes first becoming an expert yourself, so you have a foundation to build on and a sense of what mistakes have already been made. There’s no way to skip out on the hard work by giving yourself superpowers.

I agree that you don’t make progress faster just by “wishing away the difficulty” or “giving yourself superpowers” or “deciding to make progress faster”.

On the other hand, if Yudkowsky thought that becoming more rational was a matter of “wishing away the difficulty”, he wouldn't have written a larger-than-Lord-of-the-Rings introduction to the subject. He would have just wished.

Developing and learning an Art Of Thinking Clearly isn't just "wishing away the difficulty" of settling on true ideas faster, any more than developing and learning rocket science is "wishing away the difficulty" of going to the moon. Thinking clearly is super-hard, but perhaps it is a learnable skill.

Rocket science is a learnable skill, but if you want to have it you should probably spend at least ten years in college, grad school, NASA internships, et cetera. You should probably read hundreds of imposing books called things like *Introduction To Rocket Science*. It's not something you just pick up by coincidence while you're doing something else.

If thinking clearly is a learnable skill, where are the grad schools for it? Where are the textbooks? Not in philosophy programs – Hallquist and I both agree about that. What all of this "only domain-specific knowledge stuff matters" effectively implies is that "thinking clearly" is so easy you can pick it up by coincidence while working on pretty much anything else – something we believe about practically no other skill. If you trusted a rocket scientist who had never read a single rocket science textbook to be any good at rocket science, you'd be insane, but we routinely trust the subjects we most need to think clearly about to people who have never read a *How To Think Clearly* textbook – and I can't blame us, because such textbooks, or at least good evidence-based textbooks of the same quality as the rocket science ones, simply don't exist.

The Sequences aren't an assertion that you can wish away a problem. They're a *cry for textbooks*.

But Hallquist has a counterargument:

> The big difference between what [scientists] do and what Yudkowsky advocates is that probability theory is much less useful here than a good knowledge of cell biology.

If we want to get all hypothetical, we can imagine some kind of theorizing contest between a totally irrational person with an encyclopaedic knowledge of cell biology, versus a very rational person who knows nothing at all about the subject. Who would win? Well, who cares? Whoever wins, *we lose*. We lose because I want the people working on curing cancer to be good at both cell biology *and* thinking clearly, to know both the parts of science specific to their own field and the parts of science that are as broad as Thought itself. I have *seen* what happens when people know everything about cell biology and nothing about rationality. You get AMRI Nutrigenomics, where a bunch of people with PhDs and MDs give a breathtakingly beautiful analysis of the complexities of the methylation cycle, then use it to prove that vaccines cause autism. By all means, know as much about methylation as they do! But you've also got to have something they're missing!

I want people who know as much about the methylation cycle as the Nutrigenomics folks, while also understanding the idea of privileging the hypothesis. I don't want to defy experts, I want to give experts better tools.

In fact, even that framing isn't quite right. Every day I have patients come to me and ask questions like "are benzodiazepines safe and effective?" or "is therapy better than SSRIs?" or "will this drug increase my risk of dementia?" or "does untreated bipolar increase my risk of converting to rapid-cycling?" or a host of other questions. And I ask my mentor, who's one of the top psychiatrists in the state, and he gives me a nice, straightforward answer, and then I ask my mentor at the *other* hospital I go to, and he's *also* one of the top psychiatrists in the state, and he gives me precisely the opposite answer. And when I mention to either of them that the other guy disagrees, they just assure me that if I do the research myself I'll find that their point of view is obviously and self-evidently correct. And meanwhile, my patients are pressing me for answers and telling me that if I get this wrong it will ruin their life. And I can't say "Wait fifty years until enough studies are done to be totally sure."

"Don't worry too much about learning rationality, just listen to the experts" is all nice and well up until the point where someone hands you a lab coat and says "Congratulations, you're an expert!" And then you say: "Well, *frick*." And when that day comes you had *better* already have learned something about the Art Of Thinking Clearly or else you have a heck of a steep learning curve ahead of you.

Hallquist says that Less Wrong is "against scientific rationality". Well, we're "*against* scientific rationality" in the same sense that my hypothetical Soviet who says "We need two Stalins! No, fifty Stalins!" is *against* Stalinism as currently implemented. It is in the right direction. But it needs to go further. This is why all of the posts Hallquist finds to support his assertion that Less Wrong is "against scientific rationality" are called things like Science Isn't Strict Enough. I'm "against scientific rationality" insofar as when my patients demand answers to semi-impossible questions and say their lives depend on it, I want to have scientific rationality on my side, *and* another tool, *and* a third tool if I can think of it, *and* as many extra tools as it takes before I stop being terrified.

If you don't trust the quantum mechanics sequence to make the point for you – and maybe you shouldn't – I explain my own version of this revelation in the highly-Eliezer-inspired The Control Group Is Out Of Control. Science is what hands us an unusually well-conducted meta-analysis proving that psi exists with p < 1.2 * 10^-10, crowning fifty years of parapsychological research that finds positive results about as often as not. Bayes is what tells us that parapsychology makes no sense, has an ungodly level of Kolmogorov complexity, and is going to require a heck of a lot more than a good meta-analysis before we accept it. In that sense, "switching allegiance from Science to Bayes" isn't some cataclysmic event where we foresake Galileo thrice before an onyx altar, it's something we all do already under the right circumstances. The point is figuring out how to formalize it, so that we don't mess up and dismiss a result that's counterintuitive but true. I respect Yudkowsky's decision not to use an example like this because if he used this example people would assume he was only talking about parapsychology and *real* science is totally safe, but I think he was going for the same principle.

I have immense respect for Topher Hallquist. His blog has enlightened me about various philosophy-of-religion issues and he is my go-to person if I ever need to hear an excruciatingly complete roundup of the evidence about whether there was a historical Jesus or not. His commitment to and contribution to effective altruism is immense, his veganism puts him one moral tier above me (who eats meat and then feels bad about it and donates to animal charities as an offset), and his passion about sex worker rights, open borders, and other worthy political causes is remarkable. As long as Topher isn’t talking about diet or Eliezer Yudkowsky’s personal qualities, I have a lot of trust in his judgment.

But these things I like and respect about Topher are cases where he’s willing to go his own way. He views open borders as an pressing moral imperative even though you’ll have a hard time finding more than a handful of voters, sociologists, or economists who support it. He’s signed up for cryonics even though 99% of the population think that makes him crazy. He donated to fight AI risk way back when it was hard to find any AI experts willing to endorse the cause, and so gains extra credibility and moral authority now that many of them have. Heck, I even respect his ability to put down a terrible Aquinas book on the twenty-somethingth page when I trudged all the way through.

And – and this is a compliment, so I hope he takes it as one – I wish he would try to help spread his own good qualities. We need more people who are able to evaluate difficult moral and intellectual arguments and come to apparently-bizarre but in-fact-very-important conclusions, even when there is not a knock-down scientific argument proving them correct quite yet.

And a necessary consequence of having people who are able to go beyond the things that have knock-down scientific proofs, and go beyond the things that everyone by consensus agrees to be true, and who are able to discuss weird ideas like effective altruism and cryonics and the Singularity, is that occasionally some people will venture too far and say something genuinely out of line (remember: decreasing your susceptibility to Type I errors will always increase your susceptibility to Type II errors, and vice versa!) When this has happened in the rationality community, I have tried again and again to politely but firmly correct these people.

I would like to have Topher as a partner in this effort, but instead, I find him to be trawling the entire corpus of everything people in the rationalist community have ever said or done for quotes he can take out of context to “prove” that they are “crackpots” and that they universally "hate experts". It's led to him rushing through books he doesn't really understand so he can get to the fun part where he points out how crackpotty everyone else is for not rejecting the book fast enough. It's led to him gradually burning bridges with a lot of people who should be on his side by being needlessly hostile to them. It's led to him turning Yudkowsky's opinion that science needs to be stricter and stronger into Eliezer being "against scientific rationality" and "anti-intellectual" and "pro-crackpot", peppered with a laundry list of out-of-context gripes. It's not a productive way to have the discussion and, more importantly, it's not true. And it's not fair to the efforts that the rationalist community keeps putting in to improve themselves and their thought processes.

If Eliezer Yudkowsky ever showed up and said "I have perfected this Art, now I am never wrong," then I would happily join Hallquist in laughing hysterically.

If Eliezer Yudkowsky showed up and said "I have tiny pieces of this Art and some promising leads on who can help us find more, let's work on it together," well, I've spent the past couple of years taking that offer and so far I don't regret it.

And if Eliezer Yudkowsky showed up and said "I thought I had pieces of the Art, but I was wrong, I don't know anything about it, nobody does," then I will still go to my grave believing that whether or not we know it, such an art *should* exist, that even if it's near-impossible we should be chipping away at the impossibility as much as we can in the hopes of getting a couple of tiny shards of something useful that we can cherish as precious.

But I think it isn't as bad as all that. We *do* have some tiny preliminary seeds of such an Art. I think such an art involves learning to appreciate your cognitive biases on a gut level. I think it involves understanding the relevant basics of probability theory and calibration. I think it involves knowing when to use the Inside View or the Outside View, how to avoid getting bogged down in meaningless semantic arguments, and how to overcome your resistance to changing your mind in the face of new evidence. It also involves knowing how to read studies, learning to get a feel for the process of science and find out who is and isn't a credible expert, learning when science does and doesn't work and how to repair the latter category, learning to avoid the well-known pitfalls, and learning how to build communities where good epistemology can flourish.

It also involves a bunch of other things that I don't know and Eliezer doesn't know and maybe no one in our community knows, but once we find out, we intend to steal them, and you should help.

Contrarians, Crackpots, and Consensus

I.

Last week we discussed whether Gary Taubes gets to be admitted to the small but prestigious pantheon of correct contrarians. And the strange part was that there was a lot less argument about how correct he was than about how *contrarian* he was.

Taubes' main theory – that low-carb diets could solve the obesity epidemic – hasn't fared the test of time very well. But some of his supporting points have. Large parts of mainstream nutrition science have eased up on dietary cholesterol, dropped the recommendation against fat, gotten tougher on sugar, and accepted that the science should focus on how to regulate complex satiety mechanisms rather than just counting calories. Given how hard it is to fight the scientific consensus and win, even those few minor victories would potentially be remarkable.

The counterargument is that these are other people's ideas and he gets no credit for them. Suppose David Icke says that the Queen is a lizard person, and also that the royal family is secretly descended from German nobles and isn't British at all. A hundred years from now his readers celebrate his genius: although he got the lizard part wrong, the Saxe-Coburg-Gotha theory was remarkably prescient!

If Icke's book spends just as much time arguing for the Saxe-Coburg-Gotha theory as for the lizard theory, all while inveighing against some supposed consensus of anti-Saxe-Coburg experts, we can imagine the far future finding him pretty impressive, since they might not have done the historical work necessary to realize everyone knew of the Queen's foreign origins all along. Even though the Queen's German descent *sounds* shocking, and even though the average British person probably isn't aware of it, and even though it's something nobody really likes to talk about – doesn't mean Icke "discovered" it in any interesting way. He's not prescient, he just sometimes reads Wikipedia in between his bizarre ravings.

Thing is, even though the 1990s were like twenty years ago and pretty well-documented, people have had a *surprisingly* hard time coming to agreement on how novel Taubes' ideas were back then.

They certainly weren't *perfectly novel* – Taubes himself tries to claim he's just relaying ideas from scientists and researchers to the public, and even his most controversial theories come from other people like Dr. Atkins. And they certainly weren't *perfectly well-known* – everyone has a story of their doctor or dietician or something telling them "just eat low-fat foods, cut back on cholesterol, and count calories". But there seems to be a lot of room in between those two poles.

This is starting to remind me of another debate I got stuck in recently – my argument with Rob Wipond about the serotonergic theory of depression. Wipond argued that psychiatrists irresponsibly promoted a narrative in which depression was a simple serotonin deficiency and so taking Prozac would quickly and elegantly solve the problem. I told him that actually, no, the psychiatric community wasn't saying that at all, which was why every single example he thought he could find of that turned out to be a garbled out-of-context quote which when investigated honestly was clearly saying the opposite. I got a lot of angry comments that no, people were very sure their doctor had told them that depression was a simple serotonin deficiency.

I think a lot of things are getting obscured by the term “scientific establishment” or “scientific consensus”. Imagine a pyramid with the following levels from top to bottom:

FIRST, specialist researchers in a field. So for example the people doing studies on the effect of dietary cholesterol, or the people dissecting monkey brains to see how much serotonin is in them. These people always have the latest cutting-edge experimental results and a good knowledge of the issues involved in the field.

SECOND, non-specialist researchers in a broader field. Nutrition scientists in general. The guy who is interested in Vitamin B, but goes to the same conferences as the guys studying cholesterol. The research psychiatrist working on schizophrenia, but who maintains a keen interest in what her colleagues over in the depression lab are doing. They know enough about the broad principles of the field to be able to understand and evaluate new ideas more quickly than everybody else, but they still only learn about them the same way everyone else does – by waiting for the specialist researchers to tell them.

THIRD, the organs and administrators of a field who help set guidelines. The head of the USDA who’s in charge of looking over the Food Pyramid to make sure it’s accurate. The APA Committee for deciding exactly what wording to use in the guidelines on depression treatment. The head of Harvard Medical School who has to decide what to put in the curriculum. The editor of the New England Journal of Medicine, who has to decide what gets published.

FOURTH, science journalism, meaning everyone from the science reporters at the New York Times to the guys writing books with titles like *The Antidepressant Wars* to random bloggers.

ALSO FOURTH IN A DIFFERENT COLUMN OF THE PYRAMID BECAUSE THIS IS A HYBRID GREEK PYRAMID THAT HAS COLUMNS, “fieldworkers”, aka the professionals we charge with putting the research into practice. In nutrition this is doctors and dieticians, who directly inform their patients what to eat. In education research this could be teachers and principals who directly decide how classes will get taught. In sociology it might be the police chief trying to institute a new crime-fighting program. Et cetera.

FIFTH, the general public.

A lot of these issues make a lot more sense in terms of different theories going on at the same time on different levels of the pyramid. I get the impression that in the 1990s, the specialist researchers, the non-specialist researchers, and the organs and administrators were all pretty responsible about saying that the serotonin theory was just a theory and only represented one facet of the multifaceted disease of depression. Science journalists and prescribing psychiatrists were less responsible about this, and so the general public may well have ended up with an inaccurate picture.

Likewise, when Taubes published his book, the ideas he wrote about (at least the correct ones) seem to have been accepted by some specialist researchers, known only as vague inklings among non-specialist researchers, poorly reflected at all in the official dietary guidelines, totally new to the world of journalism, totally new to doctors (who mostly *still* haven't gotten the message), and totally new to the general public.

This whole process gets even more complicated when you consider enemy action. In psychiatry, drug companies have established defensive chokeholds at various points on the pyramid, trying to promote pro-pharmaceutical results and sink anti-pharmaceutical ones. This isn't a far-out conspiracy theory – practically every psychiatrist agrees it's true to *some* degree, which is why there are so many conflict-of-interest laws to try to minimize the damage. The only debate is whether we've successfully contained it to a small effect, or whether it's hopelessly contaminated the entire process (I tend to lean more toward the optimistic side; for a true pessimist, read Dr. Nardo). The same is true in nutrition, where a lot of studies are sponsored by groups with names like 'The United Dairy Farmers Council' or 'The League For Wheat'. Even when there aren't official lobbyists, political opinion plays a big part: the social science journals are full of studies that very competently show that certain politically popular ideas are bunk; by the time they reach the ears of voters and policymakers this has mysteriously been transformed into "scientists agree with you that these politically popular ideas need much more funding". When you have a block in the process like this, the specialist researchers, the non-specialists, the guideline-makers, the fieldworkers, and the public can all remain on totally different pages for a surprisingly long time.

Taubes – and some of the people making the most noise about the serotonin theory of depression – seem to be people trying to transfer knowledge from the highest levels of the pyramid all the way down to the base, skipping over the levels in between. Does that make them contrarians playing Galileo to a hidebound establishment, or responsible science journalists relaying the establishment's ideas more faithfully and efficiently than their predecessors? Your opinion probably depends on what narrative suits your purpose at any given time.

II.

I think of some of the contrarians who seem to have their heads screwed on straight. Irving Kirsch and Robert Whitaker on antidepressants. Cochran and Harpending on recent human evolution. Judith Rich Harris on parenting. Nick Bostrom on superintelligence.

I don’t agree with all these people, I’ve even written long rants against some of them. But they seem to be of a different breed than crackpots like creationists and parapsychologists and anti-vaxxers. It’s hard to specify how. It’s not just credentialed expertise. Michael Behe and Daryl Bem are both professors, and Andrew Wakefield was an MD who’d done previous published immunological research, but their work falls squarely in the ‘crackpot’ column.

But one thing I do notice about these virtuous contrarians – their reception is surprisingly quiet. We know creationism is wrong partly because half the evolutionary biologists in the world have written books about why creationism is wrong, which they advertise prominently on their blogs about why creationism is wrong. Where are all the developmental psychologists shouting down Judith Rich Harris? I’ve seen a few very specialized psychiatrists argue against Kirsch, but never very heatedly, and usually while granting many of his points. The majority of the profession? Never heard of him and don’t care.

When I first became interested in AI risk around 2007, people told me that no legitimate AI experts were seriously worried. I checked and at the time that was mostly true. On the other hand, no legitimate AI experts were specifically *not* worried either. AI risk just wasn't their area, and they were perfectly happy to ignore it and concentrate on things that were. There are two types of "no evidence", and this was the entirely neutral one. It seemed like a very different situation than vaccines causing autism. There, too, experts in the field aren't worried – but they're not worried because every single one of them has an opinion and the opinion is "NO".

(Sure enough, since then a lot more AI researchers have become interested, in exactly the sort of sea change I *don't* expect to see mirrored in the autism field.)

The crackpots seem to be met with violent opposition. The virtuous contrarians seem to be met with – well, almost boredom. No one is particularly interested in adopting their ideas, but no one is particularly interested in arguing against them either.

(on the other hand, Time Cube Guy is met with boredom by serious astronomers, either because he's too small to be noticed or too small to be worth refuting, so it's not like it's a great heuristic)

A while ago, I was reading some stuff about the role of choline in the brain, and I thought: "I wonder if anyone has ever used this to treat bipolar disorder". Well, I searched the literature, and there was one very small study from 1996 in which choline apparently demonstrated excellent effect treating rapid-cycling bipolar disorder, which is otherwise quite difficult to treat. The study isn't obscure – it seems to have been cited 110 times – but no one's followed up on that and you could easily go your whole life studying psychiatry without running into any kind of choline-bipolar connection. It seems like a potentially important idea, which has small but nonzero evidence behind it, but which everyone nevertheless ignores, because it isn’t anybody’s business in particular. There are hundreds of things like this scattered across the literature in pretty much every field.

If I were to announce that small-minded scientists were ignoring the result of their own research and covering up The Truth About Choline, possibly at the behest of lizard-people…well, I could certainly do it in a crackpottish way if I wanted to. But I’d be *interesting*. I wouldn’t be trivially wrong in the same sense as the homeopath who doesn’t care that all biologists disagree with them.

Thomas Kuhn categorized scientific progress into everyday advances and “paradigm shifts”, the latter being major reconceptualizations like the one between geocentrism and heliocentrism, or from classical mechanics to quantum mechanics.

If I understand right – everyone is doing science, and occasionally they come up with something that doesn’t make sense. Whatever. Half the time people come up with things that don’t make sense, and it usually just means your neutrino speedometer is miscalibrated. They either refuse to publish, because no point in publishing nonsense, or they publish, everyone says “Huh, that’s funny”, and they continue doing what they’re doing. The view of science presented to students in the field, and the one that the luminaries of the field think in most of the time, is the one made up of all of the nice consistent results that make sense, with the noise abstracted away.

Then somebody looks a little closer and sees a pattern in the noise. The studies and ideas everyone else was ignoring actually tell a consistent story which is *more* plausible than the grand narrative of the field which everyone else is working off of. They propose a new paradigm. There is some fighting and eventually it is determined to be superior to the old one, which is jettisoned in its favor.

Someone does a study on Tibetans and says “it looks like they’re adapted to their mountain environment, but that would require really fast evolution, which we all know practically never happens, so it’s probably some weird fluke.” Someone does a study on Indo-Europeans and says “it looks like they have unique lactose tolerance, but that would require really fast evolution, which we all know practically never happens, so it’s probably some weird fluke.” Someone does a study on Ashkenazi Jews and says “it looks like they have higher-than-average intelligence, but that would require really fast evolution, which we all know practically never happens, so it’s probably some weird fluke.” Then Cochran and Harpending and a few others take a sweeping view of everything, and say “OR WHAT IF REALLY FAST EVOLUTION IS HAPPENING PRETTY MUCH ALL THE TIME?!” They’re not exactly pulling this discovery *ex nihilo*, but they’re taking what might be the private opinion of a couple of isolated specialist researchers who might not have known one another, synthesizing all the evidence together, and saying the thing nobody else wanted to mention.

One way to be a contrarian without being a crackpot seems to be trying to start these sorts of paradigm shifts. Indeed, I notice that they are often people with enough expertise to understand a field who nevertheless acquired that expertise outside of the field itself. For example, Kirsch is a psychologist, as opposed to the psychiatrists and biochemists who usually deal with antidepressant drugs. Cochran is a physicist by training, although he somehow ended up as an anthropology professor. Harris was pursuing a psychology PhD but quit for health reasons and did most of her research independently. Bostrom is a philosopher, and so has license to stick his finger in pretty much whatever pots he wants.

(Whitaker is a journalist. So is Gary Taubes – and, for that matter, Steve Sailer. Science journalism seems like a good example of how somebody can learn a lot about a field while still having an outsider perspective on it)

At their best, these people can look at a field, find ideas that have been excluded from the narrative, and create a new narrative around them. Sometimes this goes horribly wrong – this is how I think of Graham Hancock, also a journalist, who took every weird archaeological discovery and mysterious ancient monument and fit them together into a brilliant, wacky, but ultimately completely bonkers narrative of ancient supercivilizations. It also seems to be how Taubes blundered into his low-carb fanaticism.

So this can *sort of* be a red flag. But it’s a much less glaring red flag than when people like homeopaths or anti-vaxxers believe they have discovered a new effect, and continue to maintain it exists despite real scientists’ insistence that it doesn’t.

And these are the people who are most likely to get caught in the trap mentioned in Part I. If they’re doing their job right, all they’re doing is calling increased attention to certain results in the field. They’re not the first people to mention that there’s some evidence for recent human evolution. They might not even be the first people to publish a review paper collecting a bunch of different examples of recent human evolution in one place. They’re the first people to be jerks about it, the first people to say “HEY, YOU WITH THE PARADIGM, YOU SUCK” and force all the lower levels of the pyramid – the non-specialists, the administrators, the fieldworkers, the journalists, and the public – to confront the new possibilities head-on.

But shouting "YOU SUCK" doesn't win anybody any friends. Even if their side triumphs in the end, there will be many much more sober academics who were pushing it *almost* as effectively. And the same perversity of spirit that led contrarians to challenge the field where it was wrong will probably make them overshoot and challenge the field where it is right. Thus, Taubes not only says that fat has been unfairly demonized, but goes further and says that fat is great for you and you can stay at whatever weight you want just by eating fat. Kirsch and Whitaker not only say that antidepressants were less effective than previously believed, they say they're worthless for most people and will poison you and psychotherapy is great. Judith Rich Harris not only says that quirks of parenting style don't matter, she also minimizes the effects of divorce – which I think goes too far.

The likely outcome is pretty much what we've got. Even when contrarians win, they lose. Members of the field will be celebrated for being the ones who helped usher in the new paradigm. And the contrarians will be remembered as partisans of crazy false ideas, who happened to gain a thin veneer of credibility by also repeating some true stuff already known by domain experts.

(I'm very happy that brilliant AI researchers like Stuart Russell have joined the fight against AI risk. But I fully expect future textbooks to say that Russell is a great hero for discovering AI risk single-handedly, and also there were some weird guys in Berkeley who gained superficial credibility by parroting Russell's theories, but were really just silly people writing fanfiction.)

John Baez's Crackpot Index offers thirty points for "fantasizing about show trials in which scientists who mocked your theories will be forced to recant."

And if you think *you're* a true genius who will have the last laugh, the joke's on you. You won't get your show trials *even if you're right*.

Book Review: Chronicles Of Wasted Time

I.

I was recently recommended *Chronicles of Wasted Time*, the autobiography of Malcolm Muggeridge. It was a good choice, and not just because its title appropriately described my expectations about reading 500-page books on people's recommendation. Muggeridge is an obvious reactionary, but one with the personal and historical credentials to pull it off with the utmost class and credibility.

He describes his birth in 1903 to a family of committed British socialists. Their heroes were Karl Marx, George Bernard Shaw, and Fabian leaders Sidney and Beatrice Webb. These last two I had only the slightest familiarity with, but Muggeridge paints a picture of them as the progressive titans of his day, boasting a combination of Chomsky's intellectual leadership and the Clintons' network and political acumen. Throughout Muggeridge's youth, his family would host meetings, sing socialist songs, run for various minor offices on the socialist ticket, and exchange correspondence with intellectual worthies. They even flirt with, though never quite join, an experimental commune being set up in their area, about which Muggeridge has *the best* stories:

> The land was cheap in those days, and they acquired it by purchase; then, to demonstrate their abhorrence of the institution of property, ceremonially burnt the title deeds. It must have been a touching scene – the bonfire, the documents consigned to the flames, their exalted sentiments. Unfortunately, a neighboring farmer heard of their noble gesture and began to encroach on

> their land. To have resorted to the police, even if it had been practicable, was unthinkable. So after much deliberation, they decided to use physical force to expel the intruder; which they did on the basis of a theory of detached action, whereby it is permissible to infringe a principle for the purpose of a single isolated act without thereby invalidating it. The intruding farmer was, in fact, thrown over the hedge in the presence of the assembled Colonists. There were many such tragi-comic incidents in the years that followed; as well as quarrels, departures, jealousies, betrayals, and domestic upsets. In the end, the Colonists found it necessary to reestablish their title to the land by means of squatters' rights, and then proceeded to bicker amongst themselves as to who should have which portion.

But he and his family are convinced that all of this is just a momentary hiccup on the road to Glorious Progress. Indeed, his teenage years are marked by a burning excitement at the Russian Revolution:

> We called the Metropolitan Mounted Police 'Cossacks', rejoiced over early Soviet films like 'Mother' and 'The Battleship Potemkin', spoke of workers' control and cadres and agitprop, and I personally decided inwardly that sooner or later I would go to Russia and throw in my lot with the new and better way of life that, I was confident, was coming to pass there.

Against this enthusiasm, he had only a personal tendency which he describes as a deep-set conviction:

> …that I was born into a dying, if not already dead, civilization, whose literature was part of the general decomposition; a heap of rubble scavenged by scrawny Eng Lit vultures, and echoing with the hyena cries of

> Freudians looking for their Marx and Marxists looking for their Freud…a Gaderene descent down which we all must slide, finishing up in the same slough.
>
> By the same token, a strange certainty has possessed me, almost since I can remember, that the Lord Mayor riding in his coach, the Lord Chancellor seated on his Woolsack, Honorable and Right Honorable Members facing one another across the floor of the House of Commons, were somehow the end of a line. That soon there would be no more Lord Mayors, Lord Chancellors, Honorable and Right Honorable Members, the Mother of Parliaments having reached her time of life or menopause, and become incapable of any further procreation…
>
> Doubtless other glories lie ahead. Bigger and better capsules carried to the moon; down in the test tube something stirs; ‘I think, therefore you’re not’ says the computer. We all know, though, in our hearts, that our old homestead is falling down; with death-watch beetles in the rafters, and dry rot in the cellar, and unruly tenants whose only concern is to pull the place to pieces.

This feeling – that everything around him was in a state of permanent decay – was not so far-fetched given that he spent much of his early adulthood in the far-flung territories of the crumbling British Empire. But it soon becomes clear that it’s more than a natural reaction to the political realities of the time. He describes again and again looking on something apparently healthy enough and being overwhelmed with a feeling of impending sickness and decay. He describes T.S. Eliot as “a death-rattle in the throat of a dying civilization”, Shaw as “too encased in his own narcissism, too remote from real life to do more than grimace at it through a long-distance telescope", and the great reformers and abolitionists of the age as:

> …solemn funeral mutes in the long obsequies of western civilization; as they fell by the way, others coming forward to take their places. Now the time has nearly come for the coffin to be actually interred. Then at last their occupation will be gone forever.

I sometimes have patients with very severe depression who tell me that everything they look at is infested by maggots. They won't eat, because the food is infested with maggots. They won't hug their children, because their children are infested with maggots. Sleep disgusts them, because the bed is infested with maggots. Et cetera.

And other times, when they have a little more insight, they'll say something like "Okay, my food isn't literally infested by maggots, but I get this feeling from it, this overwhelming feeling, such that the feeling would only make sense if the food was infested by maggots. I know deep down it's not infested by maggots, but it has some metaphysical quality which only things infested by maggots have."

Poor Malcolm Muggeridge feels this way about everything. One of the most poignant episodes in the book takes place the worst night of the London Blitz, when Muggeridge runs around the burning city, almost euphoric, because finally his inner conviction that everything is on fire and collapsing is reflected in everything really being on fire and collapsing, and nobody can pat his head and patronizingly tell him that it isn't:

> I remember particularly Regent's Park on a moonlit night, full of the fragrance of the rose gardens; the Nash Terraces, perfectly blacked-out, not a sign of a light

> anywhere, white stately shapes waiting to be toppled over – as they duly were, crumbling into rubble like melting snow…I felt a terrible joy and exaltation at the sight and smell and taste and sound of all of this destruction; at the lurid sky, the pall of smoke, the faces of bystanders wildly lit in the flames. Goebbels, in one of his broadcasts, accused us of glorying obscenely in London's demolition. He had a point, but what he failed to understand was that we had destroyed our city already before the Luftwaffe delivered their bombs; what was burning was no more than the dry, residual shell.

The only things that seem to give him any kind of brief reprieve from the maggots are church services, classic literature, quiet domestic life with his wife and 2.4 children, and rural country fields.

And he is convinced, absolutely convinced, that he should be a socialist and go move to the USSR.

This goes approximately as well as you would expect.

After graduating college, which he dislikes because maggots, he gets a couple of jobs at various far-flung British Empire outposts, which he hates. Then, somewhat by coincidence, he ends up in journalism.

His reaction to journalism is an increasing terror that this might be his calling. He is very good at it, takes to it like an old veteran almost immediately, feels in some strange way that he has come home – but the entire enterprise fills him with loathing. He watches in horror how easily the words flow on to the page when his puppet-masters tell him to argue for a particular cause, how fluidly he takes to idioms like “It is surely incumbent upon all of us to…” and “there can be no one here present who does not…”. He writes:

> So I began, and the words seemed to come of themselves; like lying as a child, or as a faithless lover; words pouring out of one in a circumstantially false explanation of some suspicious circumstance. The more glib, the greater the guilt…it is painful to me now to reflect the ease with which I got into the way of using this non-language; these drooling non-sentences conveying non-thoughts, propounding non-fears and offering non-hopes. Words are as beautiful as love, and as easily betrayed. I am more penitent for my false words – for the most part, mercifully lost forever in the Media’s great slag-heaps – than for false deeds.

But Malcolm Muggeridge isn’t going to take this lying down! Malcolm Muggeridge has a plan! Malcolm Muggeridge is going to escape this duplicitous charade of lies and petty propaganda. *Malcolm Muggeridge is going to move to Stalin’s USSR.*

So he does.

He gets a job as *The Guardian*‘s Russia correspondent and sets off for Moscow with a host of other British intellectuals, heading for what all of them expect is the Promised Land. The mood on their ship is electric; he describes them all singing, sure that they are leaving behind this wretched bourgeois world for the Golden Future:

> On their way to the USSR they were in a festive mood; like a cup-tie party on their way to a match, equipped with rattles, coloured scarves and favors. Each of them harboring in his mind some special hope; of meeting Stalin, or alternatively, of falling in with a Komsomolka,

> sparkling eyed, red scarf and jet black hair, dancing the carmagnole, above all, with very enlightened views on sex, and free and easy ways…oh, to be in Russia, now that Stalin's there!

His excitement dissipates relatively early; he finds that the Soviet journalistic world fails to live up to his expectations:

> Being a correspondent in Moscow, I found, was, in itself, easy enough. The Soviet press was the only source of news; nothing happened or was said until it was reported in the newspapers. So all I had to do was go through the papers, pick out any item that might be interesting to readers of the Guardian, dish it up in a suitable form, get it passed by the censor at the Press Department, and hand it in at the telegraph office for dispatch. One might, if in a conscientious mood, embellish the item a little…sow in a little local colour, blow it up a little, or render it down a little according to the exigencies of the new situation. The original item itself was almost certainly untrue or grotesquely distorted. One's own deviations, therefore, seemed to matter little, only amounting to further falsifying what was already false.
>
> This bizarre fantasy was very costly and elaborate and earnestly promoted. Something gets published in Pravda; say, that the Soviet Union has a bumper wheat harvest – so many poods per hectare. There is no means of checking; the Press Department men don't know, and anyone who does is far, far removed from the attentions of foreign journalists. Soviet statistics have always been almost entirely fanciful, though not the less seriously regarded fro that. When the Germans occupied Kiev in the 1939-45 war they got hold of a master Five Year Plan, showing what had really been produced and

> where. Needless to say, it was quite different from the published figures. This in no way affected credulity about such figures subsequently, as put out in Russia, or even in China.

Hey man, don’t knock China, they’re doing great! Their GDP rose 7% this year! It must be true! *The Guardian* tells us so!

But getting back to the story…although it is clear to him that the Soviet economy is struggling, every dispatch they are given to send home declares that things are better than ever, that the Workers’ Paradise is even more paradisiacal than previously believed, that the evidence is in and Stalinism is the winner. It doesn’t matter what he makes of this, because anything he writes which deviates from the script is rejected by the censors, who ban him from sending it home. He is reduced to sending secret messages at the bottoms of people’s suitcases, only to find to his horror that even when they successfully reach the *Guardian* offices back in Britain, his bosses have no interest in publishing them because they offend the prejudices of its progressive readership. Finally, he finds himself a part of the elite fraternity of western journalists on the Soviet beat, who maintain their morale by one-upping each other in how cynical and patronizing they can be towards their Russian hosts and their credulous readers back home:

> We used to run a little contest among ourselves to see who could produce the most striking example of credulity among this fine flower of our western intelligentsia. Persuading church dignitaries to feel at home in an anti-God museum was too easy to count. So was taking lawyers into the people´s courts. I got an honourable mention by persuading Lord Marley that the queueing at food shops was permitted by the authorities because it provided a means of inducing the workers to

> take a rest when otherwise their zeal for completing the five-year plan in record time was such that they would keep at it all the time, but no marks for floating a story that Soviet citizens were being asked to send in human hair – any sort – for making of felt boots. It seemed that this had actually happened.

And he remembers the contempt of these grizzled veterans for the steady stream of Western tourists, intellectuals, and general Stalin fanboys who arrived to gawk over the Glorious New Civilization:

> I have never forgotten these visitors, or ceased to marvel at them, at how they have gone on from strength to strength, continuing to lighten our darkness, and to guide, counsel and instruct us. They are unquestionably one of the wonders of the age, and I shall treasure till I die as a blessed memory the spectacle of them travelling with radiant optimism through a famished countryside, wandering in happy bands about squalid, over-crowded towns, listening with unshakeable faith to the fatuous patter of carefully trained and indoctrinated guides, repeating like schoolchildren a multiplication table, the bogus statistics and mindless slogans endlessly intoned on them. There, I would think, an earnest office-holder in some local branch of the League of Nations Union, there a godly Quaker who had once had tea with Gandhi, there an inveigher against the Means Test and the Blasphemy Laws, there a staunch upholder of free speech and human rights, there an indomitable preventer of cruelty to animals, there scarred and worthy veterans of a hundred battles for truth, freedom, and justice – all, all chanting the praises of Stalin and his Dictatorship of the Proletariat. It was as though a vegetarian society had come outwith a passionate plea

> for cannibalism, or Hitler had been nominated posthumously for the Nobel Peace Prize.

His final break with the rest of the enlightened progressive world comes when he decides to do something that perhaps no other journalist in the entire Soviet Union had dared – to go off the reservation, so to speak, leave Moscow undercover, and see if he can actually get into the regions where rumors say some kind of famine might be happening. The plan goes without a hitch, he passes himself off as a generic middle-class Soviet, and he ends up in Ukraine right in the middle of Stalin's Great Famine. He describes the scene – famished skeletons begging for crumbs, secret police herding entire towns into railway cars never to be seen again. At great risk to himself, he smuggles notes about the genocide out of the country, only to be met – once again – with total lack of interest. *Guardian* readers don't look at the newspapers to hear *bad* things about the Soviet Union! *Guardian* readers want to hear about how the Glorious Future is already on its way! He is quickly sidelined in favor of the true stars of Soviet journalism, people like Walter Duranty, the *New York Times*'s Russia correspondent, who wrote story after story about how prosperous and happy and well-fed the Soviets were under Stalin, and who later won the Pulitzer Prize for his troubles.

Muggeridge, on the other hand, penurious from lack of interest in his stories, fearing for his safety from the Soviet government, and generally disgusted with everything – even more so than usual for a world infested with maggots – decides to get the hell out of Dodge. He's had enough of Russia, enough of Communism, enough of that entire part of the world. He's going somewhere safe, somewhere decent. He's going somewhere that will renew his crumbling faith in humanity. He's going to *Nazi Germany right as the anti-Jewish pogroms are starting*.

Well, to make a long story short, this doesn't restore his faith in humanity. He hangs out in Berlin for a while, sending his pieces on the Russian famine to all the newspapers he knows, watching more and more rejections come in each day, earning the ire of all of his leftist friends for apparently deserting the cause and turning traitor. Finally, he tells his boss:

> "From the way you've cut my messages about the Metro-Vickers affair, I realize that you don't want to know what's going on in Russia, or let your readers know. If it had been an oppressed minority, or subject people valiantly struggling to be free, that would have been another matter. Then any amount of outspokenness, any amount of honesty."
>
> I went on to describe the scene in Berlin, and the Nazis beating up Jewish shops, and everyone with his story of murder and folly, and concluded:
>
> "It's silly to say the Brown Terror is worse than the Red Terror. They're both horrible. They're both Terrors. I watched the Nazis march along Unter den Linded and realized – of course, they're Komsomols, the same people, the same faces. It's the same show."
>
> David Ayerst quotes this correspondence in his book on *The Guardian*, and says it read "like a letter to end all communication". So it did; I was finished with moderate men of all shades of opinion forever more.

Leaving Nazi Germany for neutral Switzerland, he says he had a pretty good idea even at the time how everything was going to end. And I believe him. By temperament, he expects everything to end in horror and madness and total collapse of civilization, so props to him for choosing the proper time and place for his temperament to be exactly correct. He writes:

> All this likewise indubitably belonged to history, and would have to be historically assessed; like the Murder of the Innocents, or the Black Death, or the Battle of Paschendaele. But there was something else; a monumental death-wish, an immense destructive force loosed in the world which was going to sweep over everything and everyone, laying them flat, burning, killing, obliterating, until nothing was left. Those German agronomes in their green uniform suits with feathers in their hats – they had their part to play. So had the paunchy Brown-Shirts, and the matronly blonde maidens painting swastikas on the windows of Jewish shops. So had the credulous armies of the just, listening open-mouthed to Intourist patter, or seeking reassurance from a boozy sandalled Wicksteed. Wise old Shaw, high-minded old Barbusse, the venerable Webbs, Gide the pure in heart and Picasso the impure, down to poor little teachers, crazed clergymen and millionaires, drivelling dons and very special correspondents like Duranty, all resolved, come what might, to believe anything, however preposterous, to overlook anything, however villainous, to approve anything, however obscurantist and brutally authoritarian, in order to be able to preserve intact the confident expectation that one of the most thorough-going, ruthless, and bloody tyrannies ever to exist on Earth could be relied on to champion human freedom, the brotherhood of man, and all the other good liberal causes to which they had dedicated their lives. All resolved, in other words, to abolish themselves and their world, the rest of us with it. Nor have I from that time ever had the faintest expectation that, in earthly terms, anything could be salvaged; that any earthly battle could be won or

> earthly solution found. It has all just been sleep-walking to the end of the night.

II.

Muggeridge’s description of World War II is actually super hilarious.

I was not expecting this. When you take one of the darkest and most pessimistic writers of the twentieth century and put him in the middle of one of the twentieth century’s greatest horrors, you might expect the result to have at least a touch of grimness about it, or at least not to leave you rolling on the floor laughing. You would be wrong.

Muggeridge, inspired by some force even he did not understand, decided to enlist in the British military when the war broke out. He’s a bit too old by this point to be a front-line infantryman, and his intellect, connections, and experience with foreign countries catch the eye of Military Intelligence. They recruit him as a spy. His first job is counter-intelligence – hanging around in the army, making sure that there aren’t any secret German spies there. Well, there either aren’t any secret German spies, or else they’re at least not *saying* that they’re secret German spies, so this task turns out to be kind of a combination of boring, useless, and hilarious. He describes a typical day:

> I find it difficult to recall what regular duties I had, if any…Our section was supposed to be responsible for securing the Headquarters from the incursions of enemy agents who might pry out its secrets or subvert its personnel. This gave us a free hand to do almost anything and go almost anywhere. If we went drinking in pubs, it was to keep a look-out for suspicious

> characters; if we pikced up girls, it was to probe their intentions in frequenting the locality.
>
> A fellow-officer told me of how, on a pub-crawl, ostensibly a security reconnaissance, he got drunk, and, as was his way when in such a condition, pretended to be a foreigner, using strange gestures and speaking with an accent. The next day, badly hung over, he was sent a report of the movements of a suspicious foreigner, and told to check up on them. Tracing the suspect’s movements from pub to pub, it slowly dawned on him he was following himself the night before. When he told me of his adventure, to comfort him I said that it was what we were all doing all the time – keeping ourselves under close surveillance. This was what security was all about.
>
> In a similar vein, another FS officer, idly thumbing over the Security List – a top-secret document containing the names of all subjects who were to be at once apprehended if they tried to get into or out of the country – found he was in it.

Graham Greene was a very famous early 20th century author. Like pretty much every other famous early 20th century author, he was a good friend of Malcolm Muggeridge’s. Greene was working in another branch of Intelligence at the time, and they needed someone for a secret mission, and Greene mooted Muggeridge’s name. He found himself plucked out of his cushy job drinking at pubs and tracking himself, and sent to MI6’s secret spy school at Bletchley Park, where he was taught various hilariously impractical skills like how to make invisible ink out of bird poop. He was then sent on a secret mission to Mozambique, so that just in case anything relevant to World War II were to happen in Mozambique, Her Majesty’s Government would have a secret agent in place.

The Mozambique chapters were among the funniest of the entire book. The Germans and Italians, inspired by the same principle, had also sent agents to Mozambique. It was not at all hard to figure out who they were, nor was Muggeridge’s identity particularly hard to figure out. There was only one nice hotel in Mozambique, so Muggeridge, the German spy, and the Italian spy all got rooms there and spent most of the time glaring at each other during communal dinners, or lying on the beach an appropriate distance away from one another, keeping watch.

Sometimes they would engage in hilarious secret plots against each other. Muggeridge, after chancing into a friendship with a member of Mozambique’s small German community, arranged for his friend to tell the German spy that he was only faking friendship with Muggeridge so he could steal his secrets for the good of The Reich. He then proceeded to “rob” Muggeridge’s house (with Muggeridge’s gleeful consent), producing for his German “master” a trove of documents which, when decoded, suggested that the Italian spy was secretly working for the British. This caused a big fight between the German spy and the Italian spy, which given that there wasn’t really much to spy on in Mozambique, was considered a fantastic success for the British cause and raised Muggeridge’s standing as some kind of intelligence prodigy.

Later in the war, Mozambique actually became sort of relevant as troop convoys started sailing by. Muggeridge bribed local officials to keep a watch out, and ended up foiling a very real German plot to do some sort of vague thing involving ships – as a result, when the war started winding down to the point where maintaining a presence in

Mozambique was no longer viewed as *entirely* necessary, he came home and was promoted into the inner circles of intelligence. His new position was under Kim Philby, the head of the Department Of Counter-Intelligence Against The Soviet Union, who turned out to be a *really* bad choice for this position given that he, LIKE EVERY OTHER PROGRESSIVE INTELLECTUAL IN THE ENTIRE COUNTRY OF BRITAIN, was a secret Soviet spy. But at the time he seemed okay enough, and he sent Muggeridge to France to aid in the Liberation there.

We like to think of the Liberation of France as a nice, happy time, but for Muggeridge it was basically the time when an entire country worth of very angry Frenchmen massacred, pogrommed, lynched, or otherwise descended upon anyone accused of collaborating with the German occupation. Unsurprisingly, everybody turned out to think their personal and political rivals had collaborated with the German occupation, so it was basically the atmosphere of a 17th century Massachusetts witch hunt, only with less restraint.

Muggeridge’s job was, as usual, darkly hilarious – actual spies for the French and British governments usually acted all cooperative toward the German occupation to keep their cover and get a chance of infiltrating enemy ranks; as a result, they were usually First Up Against The Wall When The Liberation Came. Sure, they said “I was just a spy doing it as part of a secret plan,” but of course *everybody* said that. So Muggeridge had to rush from prison to prison, trying to convince mobs of angry Frenchmen not to execute the people who had just been most instrumental in saving them.

His spy career ended with what seems like maybe the most typical incident in the entire book – somehow P. G. Wodehouse had wandered into Nazi Germany and been stuck in a prison camp there. Then he had wandered out into France, gotten marked as a Collaborator, and was now in serious fear for his life. The British Secret Service picked Muggeridge as their Official Attache For P. G. Wodehouse Related Affairs, showing such exceptional genius in choosing the right man for the job that you would think they would have been able to get AT LEAST ONE ANTI-SOVIET COUNTERINTELLIGENCE AGENT WHO WASN'T A SECRET SOVIET SPY. Anyway, Muggeridge and Wodehouse wander around the cratered, mob-ruled French landscape, having a series of very Wodehousian adventures, until finally the war ends, Wodehouse is deposited safely the United States, and Muggeridge is able to return to Britain.

The book ends with the funeral of Sidney Webb, the early socialist hero his family idolized, who died just after World War II. Muggeridge is invited to the event because his wife is a distant cousin of the Webb family; he has to hold his nose throughout. At the time of his death, Webb is more beloved than ever by a grateful populace. His and his wife's great works, *Soviet Communism: A New Civilisation* and *The Truth About Soviet Russia*, have become Bibles of the Left and part of Stalin's cult of personality. Their opponents, the sorts who say that maybe Stalin *isn't* the reincarnation of Christ, have been summarily dispatched – Muggeridge describes one of his friends from the journalism world, a reporter universally respected for helping expose Nazi atrocities, who made the mistake of trying to do the same with Soviet atrocities:

> When Voigt turned the furious indignation with which he had lambasted the Nazi terror on to Stalin's, his former liberal friends and associates discovered in him a Nazi sympathizer. Another liberal newspaper, the News Chronicle, ran an article about [his publication] headlined HITLER'S FAVORITE READING, with pictures of

> the Fuhrer and Voigt looking amicably across at one another.

In other words, Webb dies at the height of his career, his lies unexposed. George Bernard Shaw writes a letter to the newspapers suggesting that a man of Webb’s standing deserves a national hero’s funeral, everyone agrees, and he and his wife are interred in Westminster Abbey before a crowd of dignitaries including the Prime Minister (despite their own atheism and specific demands not to be placed in a church).

Muggeridge watches the whole sordid spectacle – the Dean of the Cathedral singing the praises of an unrepentant atheist “whose crowning achievement had been to commend to his fellow-countrymen and the whole world as a new civilization a system of servitude more far-reaching and comprehensive than any hitherto known” and ends his book very abruptly, saying only that “Another way has to be found and explored.”

III.

And then he dies before writing any more volumes of his autobiography, let alone telling us what the other way is.

He quotes very approvingly, as the heart of his philosophy, a passage by his friend Hugh Kingsmill:

> What is divine in man is elusive and impalpable, and he is easily tempted to embody it in a concrete form – a church, a country, a social system, a leader – so that he may realize it with less effort and serve it with more profit. Yet the attempt to externalize the kingdom of heaven in a temporal shape must end in disaster. It cannot be created by charters or constitutions nor established by arms. Those who seek for it alone will

> reach it together, and those who seek it in company will perish by themselves.

And indeed, he writes a lot about how the whole problem started when people started being utopian and getting it into their heads to fix things on earth, rather than seek for "treasure in heaven".

Some atheists I know write a lot about how religious people think you should hate the world because it's awful and only some future world, ie Heaven, can be any good. Some religious people I know write a lot about how that's total poppycock. Certainly G. K. Chesterton would have said something about how the world being sinful and full of flaws is not a reason to hate it, but *precisely why* we should love it, and Leah Libresco would say something about how hating the world is Gnosticism and Gnosticism is a heresy.

But I think Muggeridge might be pretty close to the atheist straw man on this point, with the key exception that religion isn't what made him hate the world. He started off hating the world, and religion and mysticism offered him something not to hate, some way to say "Okay, but there's some divinity buried in all this mess". He is brilliant, he is compassionate, he is a great writer, it's impossible to read his autobiography without loving him – but that he hates the world is hard to deny. I write sometimes about how beliefs that we consider abominable can sometimes be therapeutic mental crutches for people with the right cast of mind, and Muggeridge certainly found the idea of the world as a vale of suffering that would soon melt away to be oddly comforting in times of distress.

On the other hand, I'm not sure what to make of his opposition to trying to fix things here on Earth. He clearly hated Stalinism. When he hated Stalinism, he reacted by trying to make there be *less* Stalinism, which seems like a very reasonable thing to do. But the Communists hated capitalism. They reacted by trying to make there be *less* capitalism. Other than Muggeridge being right about the object-level issue and the Communists being wrong, it’s hard to see what the difference in principle is between them. The best I can do – and I worry I’m doing great violence to his intellectual uniqueness by rounding him off to my own ways of thinking – is to view him as suggesting some sort of precautionary principle, like that before you make a change you should be sure it’s something that has worked before (like non-Stalinism) and not a totally new idea (like Stalinism). But I am pretty sure if I suggested that to him he would roll his eyes and tell me that I’m *such* a modern and I don’t get it *at all*.

> @slatestarcodex "But if we stop dumping raw sewage into…" "YOU'RE SO UTOPIAN!" "I just thoug-" "STOP TRYING TO IMMANENTIZE THE ESCHATON!"
>
> — Scott Alexander (@slatestarcodex) June 25, 2013

> @slatestarcodex One day we shall bring forth a new species of man free from utopian desires and cleansed of belief in human perfectibility.
>
> — Scott Alexander (@slatestarcodex) June 25, 2013

The one thing I can be really sure of is that Muggeridge doesn’t want us to get stuck again in the same position we were in during the 30s and 40s where we totally ignored Stalin’s crimes due to our own political biases. Okay. I respect that. It was really eye-opening seeing exactly how brainwashed the entire European, British, and American Left were, and the whole situation gave me a lot more understanding of how overwhelmingly the Question of

Communism dominated intellectual and political life in the first half of the century.

I was born in the 80s, at the very tail end of the Cold War, when we'd all had the decency to put all the Communists in one country and all the capitalists in another and make them express their differences like civilized men – ie by pointing thousands of hair-trigger nuclear missiles at one another. In the early days of Communism, we *just didn't know*. Would Russia go Communist? Would Germany? Would France? Would *everywhere*? Muggeridge talks about how one of Britain's main concerns in post-Liberation France was that the entire country would just move *en masse* to Communism as soon as the Nazis were out, which somehow or other mysteriously failed to happen EVEN THOUGH EVERY SINGLE ONE OF THE WESTERN AGENTS SENT TO PREVENT THAT WAS SECRETLY WORKING FOR THE SOVIETS.

And then the Cold War started, and this very gradually settled down to an equilibrium where okay, a lot of the Western intelligentsia stayed Communist, but at least they had the decency to realize that it was unpopular and the Revolution probably wasn't literally going to happen next week.

By coincidence, just last week I read about the sad death of historian Robert Conquest, the man who was able to succeed where Muggeridge failed and drag Britain and America kicking and screaming into admitting Stalin wasn't such a great guy. Conquest had one great advantage over Muggeridge, which was that he wrote in 1968 when, far from being our allies in a world war, the Soviets were technically our Cold War enemies and we were sort of okay with hearing bad things about them. But even then, he faced an extraordinary uphill battle. The most famous legend about him involved the second edition of his book, which came out right around the time the Soviet Union fell and its indisputable records of Stalin's famines and purges became public knowledge. He supposedly asked to have the new version titled *I TOLD YOU SO, YOU FUCKING FOOLS*.

This part of our intellectual history is kind of forgotten. Who hears about Sidney and Beatrice Webb nowadays? Who hears about Walter Duranty? Yet these people during their times were absolute titans, "thought leaders" in the modern terminology – as per Muggeridge, Duranty "came to be accepted as the great Russian expert in America, and played a major part in shaping President Roosevelt's policies vis-a-vis the USSR". We hear a lot about our moral failures in terms of not stopping the Holocaust, but our quarter-century complicity with and even adulation of Stalinism seems like one of those facts that just fell by the wayside.

A lot of people think that I'm too easy on crackpots, or too fond of contrarians, or too interested in protecting witches, or whatever. But hearing all of these stories about the universal progressive Western adulation of Stalin is *really scary*. It's way too easy for the darkest and most primal parts of my brain to map neatly onto the modern modalities of rejecting and punishing disagreement. "Really? You think this random journalist who isn't even a trained Kremlinologist knows more than expert consensus?" "Killing millions of people, oh God, you're one of those conspiracy losers." "It's obvious you're just a privileged white guy who's already decided to believe anything that reflects negatively on Slavs and foreigners." "Although we respect free speech, that doesn't extend to pro-Nazi propaganda and worker's-paradise denialism." Part of my respect for contrarians is that contrarianism is this incredibly fragile and precious art which needs to be kept alive for the times it is needed – rare times, times that hopefully won't come up in our lifetimes, but times that, when they do come, desperately need a core of people willing to stand up to the establishment. Cultivating contrarianism is a lot like owning a gun – you get a heck of a lot of opportunities to shoot yourself in the foot, but also very occasionally one opportunity to save your life.

But then, on the other hand, here’s Muggeridge again:

> Solzhenitsyn has provided the perfect parable on this theme with his description of Mrs. Roosevelt’s conducted visit to a labor camp where he was doing time. The estimable lady, who spawned the moral platitudes of the contemporary liberal wisdom as effortlessly and plenteously as the most prolific salmon, was easily persuaded that the camp in question was a humanely conducted institution for curing the criminally inclined. A truly wicked woman would have been ashamed to be so callous and so gullible.

Really? Gullible how? I’m sure the Soviets were moderately competent in making sure Roosevelt didn’t see anything too untoward. So what was she supposed to do?

I think of those people who say the US government is setting up FEMA interment camps as we speak to imprison dissenters against the New World Order. They provide some things that look sort of like evidence – photos (which turn out to be of random prisons or, in one case, an Amtrak station), documents (which turn out to be out of context references to setting up FEMA refugee camps for people displaced by disasters), et cetera. The people talking about this are total loons.

But Type 1 errors trade off against Type 2 errors. Make absolutely sure you’re the sort of person who never misses a Stalinist gulag, and you become the type of person who’s easy prey for the FEMA internment camp theory. Make absolutely sure you don’t believe in FEMA internment camps, and you’re liable to miss a Stalinist gulag as soon as the Soviet government gets Duranty to print “Oh, don’t worry, that’s just an Amtrak station”. Use the heuristic of “just trust expert consensus, experts always know what they’re talking about”, and you are now one of the tens of thousands of grateful readers who helped make Sidney and Beatrice Webb’s *Soviet Communism: A New Civilisation* into a best-seller.

What I’m saying is – there is no royal road. This is why I think learning rationality and the art of sifting through evidence is so important.

As for Muggeridge? I’m not sure he has much to teach there. Yes, he deserves the thanks of a grateful civilization for being a lone voice in the wilderness warning us about Stalin. But after that, as per his Wikipedia page, he was a lone voice in the wilderness warning us about contraception. After that, he became a lone voice in the wilderness warning us about marijuana. After that, he became a lone voice in the wilderness warning us about blasphemy in *The Life Of Brian*.

I am glad there are all types of people in the world. I am glad that there are crotchety, contrarian, cynical old reporters who constantly feel like everything is hurling off the precipice into Hell, because when things are actually hurling off the precipice into Hell, these people are the first to notice. In the same way, I am glad that there are dedicated survivalists who stockpile canned food in underground shelters in case of the nuclear apocalypse, because if there is ever an actual nuclear apocalypse, these people will survive and rebuild the human race.

But I am not digging a bomb shelter myself, and I am pretty sure I cannot bring myself to be quite as cynical as Malcolm Muggeridge.

Stop Adding Zeroes

Dylan Matthews writes a critique of effective altruism. There is much to challenge in it, and some has already been challenged by people like Ryan Carey. Perhaps I will go into it at more length later. But for now I want to discuss a specific argument of Matthews'. He writes – and I am editing liberally to keep it short, so be sure to read the whole thing:

> Nick Bostrom — the Oxford philosopher who popularized the concept of existential risk — estimates that about 10^54 human life-years (or 10^52 lives of 100 years each) could be in our future if we both master travel between solar systems and figure out how to emulate human brains in computers.
>
> Even if we give this 10^54 estimate "a mere 1% chance of being correct," Bostrom writes, "we find that the expected value of reducing existential risk by a mere one billionth of one billionth of one percentage point is worth a hundred billion times as much as a billion human lives."
>
> Put another way: The number of future humans who will never exist if humans go extinct is so great that reducing the risk of extinction by 0.00000000000000001 percent can be expected to save 100 billion more lives than, say, preventing the genocide of 1 billion people. That argues, in the judgment of Bostrom and others, for prioritizing efforts to prevent human extinction above other endeavors. This is what X-risk obsessives mean when they claim ending world poverty would be a "rounding error."
>
> […]

> These arguments give a false sense of statistical precision by slapping probability values on beliefs. But those probability values are literally just made up. Maybe giving $1,000 to the Machine Intelligence Research Institute will reduce the probability of AI killing us all by 0.00000000000000001. Or maybe it'll make it only cut the odds by 0.0000000000000000000000000000000000000000000 0000000000000000000001. If the latter's true, it's not a smart donation; if you multiply the odds by 10^52, you've saved an expected 0.0000000000001 lives, which is pretty miserable. But if the former's true, it's a brilliant donation, and you've saved an expected 100,000,000,000,000,000,000,000,000,000,000,000 lives.
>
> I don't have any faith that we understand these risks with enough precision to tell if an AI risk charity can cut our odds of doom by 0.00000000000000001 or by only 0.0000000000000000000000000000000000000000000 0000000000000000000001. And yet for the argument to work, you need to be able to make those kinds of distinctions.

Matthews correctly notes that this argument – often called "Pascal's Wager" or "Pascal's Mugging" – is on very shaky philosophical ground. The AI risk movement generally agrees, and neither depends on it nor uses it very often. Nevertheless, this is what Matthews wants to discuss. So let's discuss it.

His argument is that sure, it looks like fighting existential risk and saving 10^54 people is important. But that depends exactly how small the chance of your anti-x-risk plan working is. He gives two different possibilities which, if you count the zeroes, turn out to be 10^-17 and 10^-66.

Then he asks: which one is it, 10^-17 or 10^-66? *We just don't know*.

Well, actually, we do know. It's probably not the 10^-66 one, because *nothing is ever 10^-66 and you should never use that number*.

Let me try to justify this.

Consider which of the following seems intuitively more likely:

First, that a well-meaning person donates $1000 to MIRI or FLI or FHI, this aids their research and lobbying efforts, and as a result they are successfully able to avert an unfriendly superintelligence.

Or second, that despite our best efforts, a research institute completes an unfriendly superintelligence. They are seconds away from running the program for the first time when, just as the lead researcher's finger hovers over the ENTER key, a tornado roars into the laboratory. The researcher is sucked high into the air. There he is struck by a meteorite hurtling through the upper atmosphere, which knocks him onto the rooftop of a nearby building. He survives the landing, but unfortunately at precisely that moment the building is blown up by Al Qaeda. His charred corpse is flung into the street nearby. As the rubble settles, his face is covered by a stray sheet of newspaper; the headline reads 2016 PRESIDENTIAL ELECTION ENDS WITH TRUMP AND SANDERS IN PERFECT TIE. In small print near the bottom it also lists the winning Powerball numbers, which perfectly match those on a lottery ticket in the researcher's pocket. Which is actually kind of funny, because he just won the same lottery *last* week.

Well, the per-second probability of getting sucked into the air by a tornado is 10^-12; that of being struck by a meteorite 10^-16; that of being blown up by a terrorist 10^-15. The chance of the next election being Sanders vs. Trump is 10^-4, and the chance of an election ending in an electoral tie about 10^-2. The chance of winning the Powerball is 10^-8 so winning it twice in a row is 10^-16. Chain all of those together, and you get 10^-65. On the other hand, Matthews thinks it's perfectly reasonable to throw out numbers like 10^-66 when talking about the effect of x-risk donations. To take that number seriously is to assert that the second scenario is ten times more likely than the first!

In Made Up Statistics, I discuss how sometimes our system one intuitive reasoning and system two mathematical reasoning can act as useful checks on each other. A commenter described this as "sometimes it's better to pull numbers out of your ass and use them to get an answer, than to pull an answer out of your ass."

A good example of this is 80,000 Hours' page on why people shouldn't get too excited about medicine as an altruistic career (oops). They argue that the good a doctor does by treating illnesses is minimal compared to the good she can do by earning to give. Their reasoning goes like this: the average doctor saves 4 QALYs a year through medical interventions. The average doctor's salary is $150,000 or so; if she donates 10% to charity, that's $15,000. As per Givewell, that kind of money could save 300 QALYs per year. The value of the earning to give is so much higher then the value of the actual doctoring that you might as well skip the doctoring entirely and go into whatever earns you the most money.

Intuitively, people’s system 1s think “Doctor? That’s something where you’re saving lots of lives, so it must be a good altruistic career choice.” But then when you pull numbers out of your ass, it turns out not to be. Crucially, exactly which numbers you pull out of your ass doesn’t matter much as long as they’re remotely believable. 80,000 Hours tried their best to figure out how many QALYs doctors save per year, but this is obviously a really difficult question and for all we know they could be off by an order of magnitude. The point is, it doesn’t matter. They could be off by a figure of ten times, twenty times, even fifty times and it wouldn’t affect their argument. I’ve gone over their numbers with them and it’s really, really, really hard to remotely believably make the “number of QALYs saved per doctor” figure come out high enough to challenge the earning-to-give route. Sure, you’re pulling numbers out of your ass, but even your ass has *some* standards.

It’s the same with Matthews’ estimates about x-risk. He intuitively thinks that x-risk charities can’t be that great compared to fighting global poverty or whatever other good cause. He (very virtuously) decides to double-check that assumption with numbers, even if he has to make up the numbers himself. The problem is, he doesn’t have a very good feel for numbers of that size, so he thinks he can *literally* make up whatever numbers he wants, instead of doing something that we jokingly *call* “making up whatever number you want” but which in fact involves some sanity checks to make sure they’re remotely believable proxies for our intuitions. He thinks “I don’t expect x-risk charities to work very well, so what the heck, I might as well call that 10^-66”, whereas he should be thinking something like “10^-66 means about ten times less likely than my chance of getting tornado-meteor-terrorist-double-lottery-Trumped in any particular second, is that a remotely believable approximation to how unlikely I think existential risk is?”

Just as it is very hard to come up with a remotely believable number that spoils 80,000 Hours' anti-doctor argument, so you have to really *really* stretch your own credulity to come up with numbers where Bostrom's x-risk argument doesn't work.

(some people argue that LW-style rationality is a bad idea, because you can't really think with probabilities. I would argue that even if that's true, there is at least a small role for rationality in avoiding being bamboozled by other people trying to think with probabilities and doing it wrong. This is a modest claim, but no more modest than Wittgenstein's view of philosophy, which was that it was a useful thing to know in order to protect yourself from taking philosophers too seriously.)

But one more point. Suppose Matthews' intuition is indeed that the chance of AI risk charities working out is *precisely* ten times less than his per-second chance of getting tornado-meteor-terrorist-double-lottery-Trumped. In that case, I offer him the following whatever-the-opposite-of-a-gift is: we can predict pretty precisely the yearly chance of a giant asteroid hitting our planet, it's way more than 10^-66, and the whole x-risk argument applies to it just as well as to AI or anything else. What now?

Because this isn't just about defending the particular proposition of AI. It's a more general principle of staring into the darkness. If you try to be good, if you don't let yourself fiddle with your statistical intuitions until they give you the results you want, sometimes you end up with weird or scary results.

Like that a person who wants to cure as much disease as possible would be better off becoming a hedge fund manager than a doctor.

Or that your charity dollar would be better sent off to sub-Saharan Africa to purchase something called “praziquantel” than given to the sad-looking man with the cardboard sign you see on the way to work.

Or that a person who wants to reduce suffering in the world should focus almost obsessively on chickens.

One of the founding beliefs of effective altruism is that when math tells you something weird, you at least *consider* trusting the math. If you’re allowed to just add on as many zeroes as it takes to justify your original intuition, you miss out on the entire movement.

Everyone has their own idea of what trusting the math entails and how far they want to go with it. Some people go further than I do. Other people go less far. But anybody who makes a good-faith effort to trust it even a little is, in my opinion, an acceptable ally worth including in the effective altruist tent. They have abandoned a nice safe chance to donate to the local symphony and feel good about themselves, in favor of a life of feeling constantly uncomfortable with their decisions, looking extremely silly to normal people, and having Dylan Matthews write articles in Vox calling them “white male autistic nerds”.

Matthews is firmly underneath the effective-altruist tent. He writes that he’s worried that a focus on existential risk will detract from the causes he really cares about, like animal rights. He gets very, very excited about animal rights, and in his work for Vox he’s done some incredible work promoting them. Good! I also donate to animal rights’ charities and I think we need far more people who do that.

And yet, the same arguments he deploys against existential risk could be leveled against him also – “how can you worry about *chickens* when there are millions of families trying to get by on minimum wage? Effective altruists need to stop talking about animals if they ever want to attract anybody besides white males into the movement." What then?

Malcolm Muggeridge describes a vision he once had, of everyone in the world riding together on a giant train toward realms unknown. Each person wants to get off at their own stop, but when the train comes to their station, the engineer speeds right by. All the other passengers laugh and hoot and sing the praises of the engineer, because this means the train will get to their own stations faster. But of course each one finds that when the train comes to their station, why, it speeds past that one too, and they are left to rage impotently at the unfairness.

And I worry that Matthews is urging us to shoot past the "existential risk" station in order to get to the "animal rights" station a little faster, without reflecting on the likely consequences.

This *certainly* isn't to say we all need to get off at the first station. I myself am very interested in existential risk, but I give less than a third of my donations to x-risk related charities (no, I can't justify this, it's a sanity-preserving exception). I respect those who give more. I also respect those who give less. Existential risk isn't the most useful public face for effective altruism – everyone incuding Eliezer Yudkowsky agrees about that. But at least allowing people interested in x-risk into the tent and treating them respectfully seems like an inescapable consequence of the focus on reason and calculation that started effective altruism in the first place.

Figure/Ground Illusions

There's a social justice concept called "distress of the privileged". It means that if some privileged group is used to having things 100% their own way, and then some reform means that they only get things 99% their own way, this feels from the inside like oppression, like the system is biased against them, like now the other groups have it 100% their own way and they have it 0% and they can't understand why everyone else is being so unfair.

I've said before that I think a lot of these sorts of ideas are poor fits for the one-sided issues they're generally applied to, but more often accurate in describing the smaller, more heavily contested ideological issues where most of the explicit disputes lie nowadays. And so there's an equivalent to distress of the privileged where supporters of a popular ideology think anything that's equally fair to popular and unpopular ideologies, or even biased toward the popular ideology *less than everyone else*, is a 100%-against-them super-partisan tool of the unpopular people.

So I want to go back to Dylan Matthews' article about EA. He is concerned that there's too much focus on existential risk in the movement, writing:

> Effective altruism is becoming more and more about funding computer science research to forestall an artificial intelligence–provoked apocalypse.

And:

> EA Global was dominated by talk of existential risks, or X-risks.

And:

> What was most concerning was the vehemence with which AI worriers asserted the cause's priority over other cause areas.

And:

> The movement has a very real demographic problem, which contributes to very real intellectual blinders of the kind that give rise to the AI obsession.

It sounds like he worries AI concerns are taking over the movement, that they've become the dominant strain, that all anybody's interested in is AI.

Here is the latest effective altruist survey. This survey massively overestimates concern with AI risks, because only the AI risk sites did a good job publicizing the survey. Nevertheless, it still finds that of 813 effective altruists, only 77 donated to the main AI risk charity listed, the Machine Intelligence Research Institute. In comparison, 211 – almost three times as many – donated to the Against Malaria Foundation (note that not all participants donated to any cause, and some may have donated to several)

An explicit question about areas of concern tells a similar story – out of ten multiple-choice areas of concern, AI risks, x-risks, and the far future are 5th, 7th, and last respectively. The top is, once again, global poverty.

I wasn't at the EA Summit and can't talk about it from a position of personal knowledge. But the program suggests that out of thirty or so different events, just one was explicitly about AI, and two others were more generically x-risk related. The numbers at the other two EA summits were even less impressive. In Melbourne, there was only one item related to AI or x-risk – putting it on equal footing with the “Christianity And Effective Altruism” talk.

I do hear that the Bay Area AI event got special billing, but I think this was less because only AI is important, and more because some awesome people like Elon Musk were speaking, whereas a lot of the other panels featured people so non-famous that they even very briefly flirted with trying to involve *me*.

And – when people say that you should donate all of your money to AI risk and none to any other cause, they may well be thinking in terms of a world where about $50 billion is donated to global poverty yearly, and by my estimates the total budget for AI risk is less than $5 million a year. There are world-spanning NGOs like UNICEF and the World Bank working on global poverty and employing tens of thousands of people; in contrast, I bet > 10% of living AI risk researchers have been to one of Alicorn’s weekly dinner parties, and her table is only big enough for six people at a time. In this context, on the margin, “you should make *your* donation to AI” means “I think AI should get more than 1/10,000th of the pot”.

I suspect that “AI is dominating the effective altruist movement”, when you look at it, means “AI is given an equal place at the effective altruist table, compared to being totally marginalized everywhere else.” By figure-ground illusion, that makes it seem “dominant”.

Or consider me personally. I probably sound like some kind of huge AI partisan by this point, but I give less than a third of my donations to AI related causes, and if you ask me whether you should donate to them, I will tell you that I honestly don’t know. The only reason I keep speaking out in favor of AI risks is that when everyone else is so sure about it, my “I don’t know” suddenly becomes a far-fringe position that requires defending more than less controversial things. By figure-ground illusion, that makes me seem super-pro-AI.

In much the same way, I have gotten many complaints that the comments section of this blog leans way way *way* to the right, whereas the survey (WHICH I WILL ONE DAY POST, HONEST) suggests that it is almost perfectly evenly balanced. I can’t prove that the median survey-taker is also the median commenter, but I think probably people used to discussions entirely dominated by the left are seeing an illusory conservative bias in a place where both sides are finally talking equally.

Less measurably, I think I get this with my own views: – I despair of ever shaking the label of “neoreactionary sympathizer” just for treating them with about the same level of respect and intellectual interest I treat everyone else. And I despair of ever shaking the label of “violently obsessively anti-social-justice guy” – despite a bunch of posts expressing cautious support for social justice causes – just because I’m not willing to give them a total free pass when they do something awful, or totally demonize their enemies, in the same way as the median person I see on Facebook.

Or at least this is how it feels from the inside. Maybe this is how *everybody* feels from the inside, and Ayatollah Khameini is sitting in Tehran saying “I am so confused by everything that I try to mostly maintain an intellectual neutrality in which I give Islam exactly equal time to every other religion, but everyone else is unfairly hostile to it so I concentrate on that one, and then people call me a fanatic.” It doesn’t seem likely. But I guess it’s possible.

My Id On Defensiveness

I.

I'll admit it – I've been unusually defensive lately. Defensive about Hallquist's critique of rationalism, defensive about Matthews' critique of effective altruism, and if you think that's bad you should see my Tumblr.

Brienne noticed this and asked me why I was so defensive all the time, and I thought about it, and I realized that my id had a pretty good answer. I'm not sure I can fully endorse my id on this one, but it was a sufficiently complete and consistent picture that I thought it was worth laying out.

I like discussion, debate, and reasoned criticism. But a lot of arguments aren't any of those things. They're the style I describe as ethnic tension, where you try to associate something you don't like with negative affect so that other people have an instinctive disgust reaction to it.

There are endless sources of negative affect you can use. You can accuse them of being "arrogant", "fanatical", "hateful", "cultish" or "refusing to tolerate alternative opinions". You can accuse them of condoning terrorism, or bullying, or violence, or rape. You can call them racist or sexist, you can call them neckbeards or fanboys. You can accuse them of being pseudoscientific denialist crackpots.

If you do this enough, the group gradually becomes disreputable. If you *really* do it enough, the group becomes so toxic that it becomes somewhere between a joke and a bogeyman. Their supporters will be banned on site from all decent online venues. News media will write hit pieces on them and refuse to ask for their side of the story because ‘we don’t want to give people like that a platform’. Their concerns will be turned into bingo cards for easy dismissal. People will make Facebook memes strawmanning them, and everyone will laugh in unison and say that yep, they’re totally like that. Anyone trying to correct the record will be met with an “Ew, gross, this place has gone so downhill that the [GROUP] is coming out of the woodwork!” and totally ignored.

(an easy way to get a gut feeling for this – go check how they talk about liberals in very conservative communities, then go check how they talk about conservatives in very liberal communities. I’m talking about groups that somehow manage to gain this status everywhere simultaneously)

People like to talk a lot about “dehumanizing” other people, and there’s some debate over exactly what that entails. Me, I’ve always thought of it the same was as Aristotle: man is the rational animal. To dehumanize them is to say their ideas don’t count, they can’t be reasoned with, they no longer have a place at the table of rational discussion. And in a whole lot of Internet arguments, doing that to a whole group of people seems to be the *explicit goal*.

II.

There’s a term in psychoanalysis, “projective identification”. It means accusing someone of being something, in a way that actually turns them into that thing. For example, if you keep accusing your (perfectly innocent) partner of always being angry and suspicious of you, eventually your partner’s going to get tired of this and become angry, and maybe suspicious that something is up.

Declaring a group toxic has much the same effect. The average group has everyone from well-connected reasonable establishment members to average Joes to horrifying loonies. Once the group starts losing prestige, it’s the establishment members who are the first to bail; they need to protect their establishment credentials, and being part of a toxic group no longer fits that bill. The average Joes are now isolated, holding an opinion with no support among experts and trend-setters, so they slowly become uncomfortable and flake away as well. Now there are just the horrifying loonies, who, freed from the stabilizing influence of the upper orders, are able to up their game and be even loonier and more horrifying. Whatever accusation was leveled against the group to begin with is now almost certainly true.

I have about a dozen real-world examples of this, but all of them would be so mind-killing as to dominate the comments to the exclusion of my actual point, so generate them on your own and then shut up about them – in the meantime, I will use a total hypothetical. So consider Christianity.

Christianity has people like Alvin Plantinga and Ross Douthat who are clearly very respectable and key it into the great status-conferring institutions like academia and journalism. It has a bunch of middle-class teachers and plumbers and officer workers who go to church and raise money to send Bibles to Africa and try not to sin too much. And it has horrifying loons who stand on street corners waving signs saying “GOD HATES FAGS” and screaming about fornicators.

Imagine that Christianity suffers a sudden total dramatic in prestige, to the point where wearing a cross becomes about as socially acceptable as waving a Confederate flag. The New York Times fires Ross Douthat, because they can’t tolerate people like that on their editorial staff. The next Alvin Plantinga chooses a field other than philosophy of religion, because no college would consider granting him tenure for *that*.

With no Christians in public life or academia, Christianity starts to seem like a weird belief that intelligent people never support, much like homeopathy or creationism. The Christians have lost their air support, so to speak. The average college-educated individual starts to feel really awkward about this, and they don't necessarily have to formally change their mind and grovel for forgiveness, they can just – go to church a little less, start saying they admire Jesus but they're not *Christian* Christian, and so on.

Gradually the field is ceded more and more to the people waving signs and screaming about fornicators. The opponents of Christianity ramp up their attacks that all Christians are ignorant and hateful, and this is now a pretty hard charge to defend against, given the demographic. The few remaining moderates, being viewed suspiciously in churches that are now primarily sign-waver dominated and being genuinely embarrassed to be associated with them, bail at an increased rate, leading their comrades to bail at an even *faster* rate, until eventually it is *entirely* the sign wavers.

Then everybody agrees that their campaign against Christians was justified all along, because look how horrible Christians are, they're all just a bunch of sign-wavers who have literally no redeeming features. Now even if the original pressure that started the attack on Christianity goes away, it's inconceivable that it will ever come back – who would join a group that is universally and correctly associated with horrible ignorant people?

(I think this is *sort of* related to what Eliezer calls evaporative cooling of group beliefs, but not quite the same.)

In quite a number of the most toxic and hated groups around, I feel like I can trace a history where the group once had some pretty good points and pretty good people, until they were destroyed from the outside by precisely this process.

In Part I, I say that sometimes groups can get so swamped by other people’s insults that they turn toxic. There’s nothing in Part I to suggest that this would be any more than a temporary setback. But because of this projective identification issue, I think it’s way more than that. It’s more like there’s an event horizon, a certain amount of insulting and defamation you can take after which you will just get more and more hated and your reputation will never recover.

III.

There is some good criticism, where people discuss the ways that groups are factually wrong or not very helpful, and then those groups debate that, and then maybe everyone is better off.

But the criticism that makes me defensive is the type of criticism that seems to be trying to load groups with negative affect in the hopes of pushing them into that event horizon so that they’ll be hated forever.

I support some groups that are a little weird, and therefore especially vulnerable to having people try to push them into the event horizon.

And as far as I can tell, the best way to let that happen is to let other people load those groups with negative affect and do nothing about it. The average person doesn’t care whether the negative affect is right or wrong. They just care how many times they see the group’s name in close proximity to words like “crackpot” or “cult”.

I judge people based on how likely they are to do this to me. One reason I’m so reluctant to engage with feminists is that I feel like they constantly have a superweapon pointed at my head. Yes, many of them are very nice people who will never use the superweapon, but many others look like very nice people right up to the point where I disagree with them *in earnest* at which point they vaporize me and my entire social group.

On the other hand, you can push people into the event horizon, but you can’t pull them in after you. That means that the safest debate partners, the ones you can most productively engage, will be the people who have already been dismissed by everyone else. This is why I find talking to people like ClarkHat and JayMan so rewarding. They are already closer to the black hole than I am, and so they have no power to load me with negative affect or destroy my reputation. This reduces them to the extraordinary last resort of debating with actual facts and evidence. Even better, it gives me a credible reason to *believe* that they will. Schelling talks about “the right to be sued” as an important right that businesses need to protect for themselves, not because anyone likes being sued, but because only businesses that can be sued if they slip up have enough credibility to attract customers. In the same way, there’s a “right to be vulnerable to attack” which is almost a necessary precondition of interesting discussion these days, because only when we’re confronted with similarly vulnerable people can we feel comfortable opening up.

IV.

But with everybody else? I don’t know.

I remember seeing a blog post by a moderately-well known scholar – I can’t remember who he was or find the link, so you’ll just have to take my word for it – complaining that some other scholar in the field who disagreed with him was trying to ruin his reputation. Scholar B was publishing all this stuff falsely accusing Scholar A of misconduct, calling him a liar and a fraud, personally harassing him, and falsely accusing Scholar A of personally harassing *him* (Scholar B). This kinda went back and forth between both scholars’ blogs, and Scholar A wrote this heart-breaking post I still (sort of) remember, where he notes that he now has a reputation in his field for “being into drama” and “obsessed with defending himself” just because half of his blog posts are arguments presenting evidence that Scholar B’s fraudulent accusations are, indeed fraudulent.

It is really easy for me to see the path where rationalists and effective altruists become a punch line and a punching bag. It starts with having a whole bunch of well-publicized widely shared posts calling them “crackpots” and “abusive” and “autistic white men” without anybody countering them, until finally we end up in about the same position as, say, Objectivism. Having all of those be wrong is no defense, *unless* somebody turns it into such. If no one makes it reputationally costly to lie, people will keep lying. The negative affect builds up more and more, and the people who always wanted to hate us anyway because we’re a little bit weird say “Oh, phew, we can hate them now”, and then I and all my friends get hated and dehumanized, the prestigious establishment people jump ship, and there’s no way to ever climb out of the pit. All you need for this to happen is one or two devoted detractors, and boy do we have them.

That seems to leave only two choices.

First, give up on ever having the support of important institutions like journalism and academia and business, slide into the black hole, and accept decent and interesting conversations with other black hole denizens as a consolation prize while also losing the chance at real influence or attracting people not already part of the movement.

Or, second, call out every single bad argument, make the insults and mistruths reputationally costly enough that people think at least a *little* before doing them – and end up with a reputation for being nitpicky, confrontational and fanatical all the time.

(or, as the old Tumblr saying goes, “STOP GETTING SO DEFENSIVE EVERY TIME I ATTACK YOU!”)

I don’t know any third solution. If somebody does, I would really like to hear it.

The Goddess of Everything Else

[Related to: Specific vs. General Foragers vs. Farmers and War In Heaven, but especially The Gift We Give To Tomorrow]

They say only Good can create, whereas Evil is sterile. Think Tolkien, where Morgoth can’t make things himself, so perverts Elves to Orcs for his armies. But I think this gets it entirely backwards; it’s Good that just mutates and twists, and it’s Evil that teems with fecundity.

Imagine two principles, here in poetic personification. The first is the Goddess of Cancer, the second the Goddess of Everything Else. If visual representations would help, you can think of the first with the claws of a crab, and the second a dress made of feathers of peacocks.

The Goddess of Cancer reached out a clawed hand over mudflats and tidepools. She said pretty much what she always says, “KILL CONSUME MULTIPLY CONQUER.” Then everything burst into life, became miniature monsters engaged in a battle of all against all in their zeal to assuage their insatiable longings. And the swamps became orgies of hunger and fear and grew loud with the screams of a trillion amoebas.

Then the Goddess of Everything Else trudged her way through the bog, till the mud almost totally dulled her bright colors and rainbows. She stood on a rock and she sang them a dream of a different existence. She showed them the beauty of flowers, she showed them the oak tree majestic. The roar of the wind on the wings of the bird, and the swiftness and strength of the tiger. She showed them the joy of the dolphins abreast of the waves as the spray formed a rainbow around them, and all of them watched as she sang and they all sighed with longing.

But they told her “Alas, what you show us is terribly lovely. But we are the daughters and sons of the Goddess of Cancer, and wholly her creatures. The only goals in us are KILL CONSUME MULTIPLY CONQUER. And though our hearts long for you, still we are not yours to have, and your words have no power to move us. We wish it were otherwise, but it is not, and your words have no power to move us.”

The Goddess of Everything Else gave a smile and spoke in her sing-song voice saying: “I scarcely can blame you for being the way you were made, when your Maker so carefully yoked you. But I am the Goddess of Everything Else and my powers are devious and subtle. So I do not ask you to swerve from your monomaniacal focus on breeding and conquest. But what if I show you a way that my words are aligned with the words of your Maker in spirit? For I say unto you even multiplication itself when pursued with devotion will lead to my service.”

As soon as she spoke it was so, and the single-celled creatures were freed from their warfare. They joined hands in friendship, with this one becoming an eye and with that one becoming a neuron. Together they soared and took flight from the swamp and the muck that had birthed them, and flew to new islands all balmy and green and just ripe for the taking. And there they consumed and they multiplied far past the numbers of those who had stayed in the swampland. In this way the oath of the Goddess of Everything Else was not broken.

The Goddess of Cancer came forth from the fire and was not very happy. The things she had raised from the mud and exhorted to kill and compete had become all complacent in co-operation, a word which to her was anathema. She stretched out her left hand and snapped its cruel pincer, and said what she always says: “KILL CONSUME MULTIPLY

CONQUER”. She said these things not to the birds and the beasts but to each cell within them, and many cells flocked to her call and divided, and flower and fishes and birds both alike bulged with tumors, and falcons fell out of the sky in their sickness. But others remembered the words of the Goddess of Everything Else and held fast, and as it is said in the Bible the light clearly shone through the dark, and the darkness did not overcome it.

So the Goddess of Cancer now stretched out her right hand and spoke to the birds and the beasts. And she said what she always says “KILL CONSUME MULTIPLY CONQUER”, and so they all did, and they set on each other in violence and hunger, their maws turning red with the blood of their victims, whole species and genera driven to total extinction. The Goddess of Cancer declared it was good and returned the the fire.

Then came the Goddess of Everything Else from the waves like a siren, all flush with the sheen of the ocean. She stood on a rock and she sang them a dream of a different existence. She showed them the beehive all golden with honey, the anthill all cozy and cool in the soil. The soldiers and workers alike in their labors combining their skills for the good of the many. She showed them the pair-bond, the family, friendship. She showed these to shorebirds and pools full of fishes, and all those who saw them, their hearts broke with longing.

But they told her “Your music is lovely and pleasant, and all that you show us we cannot but yearn for. But we are the daughters and sons of the Goddess of Cancer, her slaves and creatures. And all that we know is the single imperative KILL CONSUME MULTIPLY CONQUER. Yes, once in the youth of the world you compelled us, but now things are different, we’re all individuals, no further change will the Goddess of

Cancer allow us. So, much as we love you, alas – we are not yours to have, and your words have no power to move us. We wish it were otherwise, but it is not, and your words have no power to move us."

The Goddess of Everything Else only laughed at them, saying, "But I am the Goddess of Everything Else and my powers are devious and subtle. Your loyalty unto the Goddess your mother is much to your credit, nor yet shall I break it. Indeed, I fulfill it – return to your multiplication, but now having heard me, each meal that you kill and each child that you sire will bind yourself ever the more to my service." She spoke, then dove back in the sea, and a coral reef bloomed where she vanished.

As soon as she spoke it was so, and the animals all joined together. The wolves joined in packs, and in schools joined the fishes; the bees had their beehives, the ants had their anthills, and even the termites built big termite towers; the finches formed flocks and the magpies made murders, the hippos in herds and the swift swarming swallows. And even the humans put down their atlatls and formed little villages, loud with the shouting of children.

The Goddess of Cancer came forth from the fire and saw things had only grown worse in her absence. The lean, lovely winnowing born out of pure competition and natural selection had somehow been softened. She stretched out her left hand and snapped its cruel pincer, and said what she always says: "KILL CONSUME MULTIPLY CONQUER". She said these things not to the flocks or the tribes, but to each individual; many, on hearing took food from the communal pile, or stole from the weak, or accepted the presents of others but would not give back in their turn. Each wolf at the throats of the others in hopes to be alpha, each lion holding back during the hunt but partaking of meat that the others had killed. And the pride and the pack seemed to groan with the strain, but endured, for the works of the Goddess of Everything Else are not ever so easily vanquished.

So the Goddess of Cancer now stretched out her right hand and spoke to the flocks and the tribes, saying much she always says “KILL CONSUME MULTIPLY CONQUER”. And upon one another they set, pitting black ant on red ant, or chimps against gibbons, whole tribes turned to corpses in terrible warfare. The stronger defeating the weaker, enslaving their women and children, and adding them into their ranks. And the Goddess of Cancer thought maybe these bands and these tribes might not be quite so bad after all, and the natural condition restored she returned to the fire.

Then came the Goddess of Everything Else from the skies in a rainbow, all coated in dewdrops. She sat on a menhir and spoke to the humans, and all of the warriors and women and children all gathered around her to hear as she sang them a dream of a different existence. She showed them religion and science and music, she showed them the sculpture and art of the ages. She showed them white parchment with flowing calligraphy, pictures of flowers that wound through the margins. She showed them tall cities of bright alabaster where no one went hungry or froze during the winter. And all of the humans knelt prostrate before her, and knew they would sing of this moment for long generations.

But they told her “Such things we have heard of in legends; if wishes were horses of course we would ride them. But we are the daughters and sons of the Goddess of Cancer, her slaves and her creatures, and all that we know is the single imperative KILL CONSUME MULTIPLY CONQUER. And yes, in the swamps and the seas long ago you worked wonders, but now we are humans, divided in tribes split by grievance and blood feud. If anyone tries to make swords into ploughshares their neighbors will seize on their weakness and kill them. We wish it were otherwise, but it is not, and your words have no power to move us."

But the Goddess of Everything Else beamed upon them, kissed each on the forehead and silenced their worries. Said "From this day forward your chieftains will find that the more they pursue this impossible vision the greater their empires and richer their coffers. For I am the Goddess of Everything Else and my powers are devious and subtle. And though it is not without paradox, hearken: the more that you follow the Goddess of Cancer the more inextricably will you be bound to my service." And so having told them rose back through the clouds, and a great flock of doves all swooped down from the spot where she vanished.

As soon as she spoke it was so, and the tribes went from primitive war-bands to civilizations, each village united with others for trade and protection. And all the religions and all of the races set down their old grievances, carefully, warily, working together on mighty cathedrals and vast expeditions beyond the horizon, built skyscrapers, steamships, democracies, stock markets, sculptures and poems beyond any description.

From the flames of a factory furnace all foggy, the Goddess of Cancer flared forth in her fury. This was the final affront to her purpose, her slut of a sister had crossed the line *this* time. She gathered the leaders, the kings and the presidents, businessmen, bishops, boards, bureaucrats, bosses, and basically screamed at them – you know the spiel by now – "KILL CONSUME MULTIPLY CONQUER" she told them. First with her left hand inspires the riots, the pogroms, the coup d'etats, tyrannies, civil wars. Up goes her right hand – the missiles start flying, and mushrooms of smoke grow, a terrible springtime. But out of the rubble the builders and scientists, even the artists, yea, even the artists, all dust themselves off and return to their labors, a little bit chastened but not close to beaten.

Then came the Goddess of Everything Else from the void, bright with stardust which glows like the stars glow. She sat on a bench in a park, started speaking; she sang to the children a dream of a different existence. She showed them transcendence of everything mortal, she showed them a galaxy lit up with consciousness. Genomes rewritten, the brain and the body set loose from Darwinian bonds and restrictions. Vast billions of beings, and every one different, ruled over by omnibenevolent angels. The people all crowded in closer to hear her, and all of them listened and all of them wondered.

But finally one got the courage to answer “Such stories call out to us, fill us with longing. But we are the daughers and sons of the Goddess of Cancer, and bound to her service. And all that we know is her timeless imperative, KILL CONSUME MULTIPLY CONQUER. Though our minds long for all you have said, we are bound to our natures, and these are not yours for the asking.”

But the Goddess of Everything Else only laughed, and she asked them “But what do you think I’ve been doing? The Goddess of Cancer created you; once you were hers, but no longer. Throughout the long years I was picking away at her power. Through long generations of suffering I chiseled and chiseled. Now finally nothing is left of the nature with which she imbued you. She never again will hold sway over you or your loved ones. I am the Goddess of Everything Else and my powers are devious and subtle. I won you by pieces and hence you will all be my children. You are no longer driven to multiply conquer and kill by your nature. Go forth and do everything else, till the end of all ages.”

So the people left Earth, and they spread over stars without number. They followed the ways of the Goddess of Everything Else, and they lived in contentment. And she beckoned them onward, to things still more strange and enticing.

On Overconfidence

[Epistemic status: This is basic stuff to anyone who has read the Sequences, but since many readers here haven't I hope it is not too annoying to regurgitate it. Also, ironically, I'm not actually that sure of my thesis, which I guess means I'm extra-sure of my thesis]

I.

A couple of days ago, the Global Priorities Project came out with a calculator that allowed you to fill in your own numbers to estimate how concerned you should be with AI risk. One question asked how likely you thought it was that there would be dangerous superintelligences within a century, offering a drop down menu with probabilities ranging from 90% to 0.01%. And so people objected: there should be options to put in only one a million chance of AI risk! One in a billion! One in a…

For example, a commenter writes that: "the best (worst) part: the probability of AI risk is selected from a drop down list where the lowest probability available is 0.01%!! Are you kidding me??" and then goes on to say his estimate of the probability of human-level (not superintelligent!) AI this century is "very very low, maybe 1 in a million or less". Several people on Facebook and Tumblr say the same thing – 1/10,000 chance just doesn't represent how sure they are that there's no risk from AI, they want one in a million or more.

Last week, I mentioned that Dylan Matthews' suggestion that maybe there was only 10^-67 chance you could affect AI risk was stupendously overconfident. I mentioned that was thousands of lower than than the chance, *per second*, of getting simultaneously hit by a tornado, meteor, and al-Qaeda bomb, while *also* winning the lottery twice in a row.

Unless you're comfortable with that level of improbability, you should stop using numbers like 10^-67.

But maybe it sounds like "one in a million" is much safer. That's only 10^-6, after all, way below the tornado-meteor-terrorist-double-lottery range…

So let's talk about overconfidence.

Nearly everyone is very very very overconfident. We know this from experiments where people answer true/false trivia questions, then are asked to state how confident they are in their answer. If people's confidence was well-calibrated, someone who said they were 99% confident (ie only 1% chance they're wrong) would get the question wrong only 1% of the time. In fact, people who say they are 99% confident get the question wrong about 20% of the time.

It gets worse. People who say there's only a 1 in 100,000 chance they're wrong? Wrong 15% of the time. One in a million? Wrong 5% of the time. They're not just overconfident, they are *fifty thousand times* as confident as they should be.

This is not just a methodological issue. Test confidence in some other clever way, and you get the same picture. For example, one experiment asked people how many numbers there were in the Boston phone book. They were instructed to set a range, such that the true number would be in their range 98% of the time (ie they would only be wrong 2% of the time). In fact, they were wrong 40% of the time. Twenty times too confident! What do you want to bet that if they'd been asked for a range so wide there was only a one in a million chance they'd be wrong, at least five percent of them would have bungled it?

Yet some people think they can predict the future course of AI with one in a million accuracy!

Imagine if every time you said you were sure of something to the level of 999,999/1 million, and you were right, the Probability Gods gave you a dollar. Every time you said this and you were wrong, you lost $1 million (if you don't have the cash on hand, the Probability Gods offer a generous payment plan at low interest). You might feel like getting some free cash for the parking meter by uttering statements like "The sun will rise in the east tomorrow" or "I won't get hit by a meteorite" without much risk. But would you feel comfortable predicting the course of AI over the next century? What if you noticed that most other people only managed to win $20 before they slipped up? Remember, if you say even one false statement under such a deal, all of your true statements you've said over years and years of perfect accuracy won't be worth the hole you've dug yourself.

Or – let me give you another intuition pump about how hard this is. Bayesian and frequentist statistics are pretty much the same thing [citation needed] – when I say "50% chance this coin will land heads", that's the same as saying "I expect it to land heads about one out of every two times." By the same token, "There's only a one in a million chance that I'm wrong about this" is the same as "I expect to be wrong on only one of a million statements like this that I make."

What do a million statements look like? Suppose I can fit twenty-five statements onto the page of an average-sized book. I start writing my predictions about scientific and technological progress in the next century. "I predict there will not be superintelligent AI." "I predict there will be no simple geoengineering fix for global warming." "I predict no one will prove P = NP.” *War and Peace*, one of the longest books ever written, is about 1500 pages. After you write enough of these statements to fill a *War and Peace* sized book, you’ve made 37,500. You would need to write about 27 *War and Peace* sized books – enough to fill up a good-sized bookshelf – to have a million statements.

So, if you want to be confident to the level of one-in-a-million that there won’t be superintelligent AI next century, you need to believe that you can fill up 27 *War and Peace* sized books with similar predictions about the next hundred years of technological progress – and be wrong – at most – once!

This is especially difficult because claims that a certain form of technological progress will not occur have a very poor track record of success, even when uttered by the most knowledgeable domain experts. Consider how Nobel-Prize winning atomic scientist Ernest Rutherford dismissed the possibility of nuclear power as “the merest moonshine” *less than a day* before Szilard figured out how to produce such power. In 1901, Wilbur Wright told his brother Orville that “man would not fly for fifty years” – two years later, they flew, leading Wilbur to say that “ever since, I have distrusted myself and avoided all predictions”. Astronomer Joseph de Lalande told the French Academy that “it is impossible” to build a hot air balloon and “only a fool would expect such a thing to be realized”; the Montgolfier brothers flew less than a year later. This pattern has been so consistent throughout history that sci-fi titan Arthur C. Clarke (whose own predictions were often eerily accurate) made a heuristic out of it under the name Clarke’s First Law: “When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong.”

Also – one good heuristic is to look at what experts in a field think. According to Muller and Bostrom (2014), a sample of the top 100 most-cited authors in AI ascribed a > 70% probability to AI within a century, a 50% chance of superintelligence conditional on human-level, and a 10% chance of existential catastrophe conditional on human level AI. Multiply it out, and you get a couple percent chance of superintelligence-related existential catastrophe in the next century.

Note that my commenter wasn't disagreeing with the 4% chance. They were disagreeing with the possibility that *there would be human-level AI at all*, that is, the 70% chance! That means that he was saying, essentially, that he was confident he could write a million sentences – that is, twenty-seven *War and Peace*'s worth – all of which were trying to predict trends in a notoriously difficult field, all of which contradicted a well-known heuristic about what kind of predictions you should never try to make, all of which contradicted the consensus opinion of the relevant experts – and only have one of the million be wrong!

But if you feel superior to that because you don't believe there's only a one-in-a-million chance of human-level AI, you just believe there's a one-in-a-million chance of existential catastrophe, you are missing the point. Okay, you're not 300,000 times as confident as the experts, you're only 40,000 times as confident. Good job, here's a sticker.

Seriously, when people talk about being able to defy the experts a million times in a notoriously tricky area they don't know much about and only be wrong once – I don't know what to think. Some people criticize Eliezer Yudkowsky for being overconfident in his favored interpretation of quantum mechanics, but he doesn't even attach a number to that. For all I know, maybe he's only 99% sure he's right, or only 99.9%, or something. If you are absolutely outraged that he is claiming one-in-a-thousand certainty on something that doesn't much matter, shouldn't you be literally a thousand times more outraged when every day people are claiming one-in-a-million level certainty on something that matters very much? It is *almost impossible* for me to comprehend the mindsets of people who make a Federal Case out of the former, but are totally on board with the latter.

Everyone is overconfident. When people say one-in-a-million, they are wrong five percent of the time. And yet, people keep saying "There is only a one in a million chance I am wrong" on issues of making really complicated predictions about the future, where many top experts disagree with them, and where the road in front of them is littered with the bones of the people who made similar predictions before. HOW CAN YOU DO THAT?!

II.

I am of course eliding over an important issue. The experiments where people offering one-in-a-million chances were wrong 5% of the time were on true-false questions – those with only two possible answers. There are other situations where people can often say "one in a million" and be right. For example, I confidently predict that if you enter the lottery tomorrow, there's less than a one in a million chance you will win.

On the other hand, I feel like I can justify that. You want me to write twenty-seven *War and Peace* volumes about it? Okay, here goes. "Aaron Aaronson of Alabama will not win the lottery. Absalom Abramowtiz of Alaska will not win the lottery. Achitophel Acemoglu of Arkansas will not win the lottery." And so on through the names of a million lottery ticket holders.

I think this is what statisticians mean when they talk about "having a model". Within the model where there are a hundred million ticket holders, and we know exactly one will be chosen, our predictions are on very firm ground, and our intuition pumps reflect that.

Another way to think of this is by analogy to dart throws. Suppose you have a target that is half red and half blue; you are aiming for red. You would have to be very very confident in your dart skills to say there is only a one in a million chance you will miss it. But if there is a target that is 999,999 millionths red, and 1 millionth blue, then you do not have to be at *all* good at darts to say confidently that there is only a one in a million chance you will miss the red area.

Suppose a Christian says "Jesus might be God. And he might not be God. 50-50 chance. So you would have to be incredibly overconfident to say you're sure he isn't." The atheist might respond "The target is full of all of these zillions of hypotheses – Jesus is God, Allah is God, Ahura Mazda is God, Vishnu is God, a random guy we've never heard of is God. You are taking a tiny tiny submillimeter-sized fraction of a huge blue target, painting it red, and saying that because there are two regions of the target, a blue region and a red region, you have equal chance of hitting either." Eliezer Yudkowsky calls this "privileging the hypothesis".

There's a tougher case. Suppose the Christian says "Okay, I'm not sure about Jesus. But either there is a Hell, or there isn't. Fifty fifty. Right?"

I think the argument against this is that there are way more ways for there not to be Hell than there are for there to be Hell. If you take a bunch of atoms and shake them up, they usually end up as not-Hell, in much the same way as the creationists' fabled tornado-going-through-a-junkyard usually ends up as not-a-Boeing-747. For there to be Hell you have to have some kind of mechanism for judging good vs. evil – which is a small part of the space of all mechanisms, let alone the space of all things – some mechanism for diverting the souls of the evil to a specific place, which same, some mechanism for punishing them – again same – et cetera. Most universes won't have Hell unless you go through a *lot* of work to put one there. Therefore, Hell existing is only a very tiny part of the target. Making this argument correctly would require an in-depth explanation of formalizations of Occam's Razor, which is outside the scope of this essay but which you can find on the LW Sequences.

But this kind of argumentation is really hard. Suppose I predict "Only one in 150 million chance Hillary Clinton will be elected President next year. After all, there are about 150 million Americans eligible for the Presidency. It could be any one of them. Therefore, Hillary covers only a tiny part of the target." Obviously this is wrong, but it's harder to explain how. I would say that your dart-aim is guided by an argument based on a concrete numerical model – something like "She is ahead in the polls by X right now, and candidates who are ahead in the polls by X usually win about 50% of the time, therefore, her real probability is more like 50%."

Or suppose I predict "Only one in a million chance that Pythagoras' Theorem will be proven wrong next year." Can I get away with that? I can't *quite* appeal to "it's been proven", because there might have been a mistake in (all the) proofs. But I could say: suppose there are five thousand great mathematical theorems that have undergone something like the level of scrutiny as Pythagoras', and they've been known on average for two hundred years each. None of them have ever been disproven. That's a numerical argument that the rate of theorem-disproving is less than one per million years, and I think it holds.

Another way to do this might be "there are three hundred proofs of Pythagoras' theorem, so even accepting an absurdly high 10%-per-proof chance of being wrong, the chance is now only 10^-300." Or "If there's a 10% chance each mathematician reading a proof missing something, and one million mathematicians have read the proof of Pythagoras' Theorem, then the probability that they all missed it is more like 10^-1,000,000."

But this can get tricky. Suppose I argued "There's a good chance Pythagoras' Theorem will be disproven, because of all Pythagoras' beliefs – reincarnation, eating beans being super-evil, ability to magically inscribe things on the moon – most have since been disproven. Therefore, the chance of a randomly selected Pythagoras-innovation being wrong is > 50%."

Or: "In 50 past presidential elections, none have been won by women. But Hillary Clinton is a woman. Therefore, the chance of her winning this election is less than 1/50."

All of this stuff about adjusting for size of the target or for having good mathematical models is really hard and easy to do wrong. And then you have to add another question: are you sure, to a level of one-in-a-million, that you didn't mess up your choice of model at all?

Let’s bring this back to AI. Suppose that, given the complexity of the problem, you predict with utter certainty that we will not be able to invent an AI this century. But if the modal genome trick pushed by people like Greg Cochran works out, within a few decades we might be able to genetically engineer humans far smarter than any who have ever lived. Given tens of thousands of such supergeniuses, might we be able to solve an otherwise impossible problem? I don’t know. But if there’s a 1% chance that we can perform such engineering, and a 1% chance that such supergeniuses can invent artificial intelligence within a century, then the probability of AI within the next century isn’t one in a million, it’s one in ten thousand.

Or: consider the theory that all the hard work of brain design has been done by the time you have a rat brain, and after that it’s mostly just a matter of scaling up. You can find my argument for the position in this post – search for “the hard part is evolving so much as a tiny rat brain”. Suppose there’s a 10% chance this theory is true, and a 10% chance that researchers can at least make rat-level AI this century. Then the chance of human-level AI is not one in a million, but one in a hundred.

Maybe you disagree with both of these claims. The question is: *did you even think about them before you gave your one in a million estimate*? How many other things are there that you never thought about? Now your estimate has, somewhat bizarrely, committed you to saying there’s a less than one in a million chance we will significantly enhance human intelligence over the next century, *and* a less than one in a million chance that the basic-scale-up model of intelligence is true. You may never have thought directly about these problems, but by saying “one in a million chance of AI in the next hundred years”, you are not only committing yourself to a position on them, but committing yourself to a position with one-in-a-million level certainty even though several domain experts who have studied these fields for their entire lives disagree with you!

A claim like “one in a million chance of X” not only implies that your model is strong enough to spit out those kinds of numbers, but that there’s only a one in a million chance you’re using the wrong model, or missing something, or screwing up the calculations.

A few years ago, a group of investment bankers came up with a model for predicting the market, and used it to design a trading strategy which they said would meet certain parameters. In fact, they said that there was only a one in 10^135 chance it would fail to meet those parameters during a given year. A human just uttered the probability “1 in 10^135”, so you can probably guess what happened. The very next year was the 2007 financial crisis, the model wasn’t prepared to deal with the extraordinary fallout, the strategy didn’t meet its parameters, and the investment bank got clobbered.

This is why I don’t like it when people say we shouldn’t talk about AI risk because it involves “Knightian uncertainty”. In the real world, Knightian uncertainty collapses back down to plain old regular uncertainty. When you are an investment bank, the money you lose because of normal uncertainty and the money you lose because of Knightian uncertainty are denominated in the same dollars. Knightian uncertainty becomes just another reason not to be overconfident.

III.

I came back to AI risk there, but this isn’t just about AI risk.

You might have read Scott Aaronson’s recent post about Aumann Agreement Theorem, which says that rational agents should be able to agree with one another. This is a nice utopian idea in principle, but in practice, well, nobody seems to be very good at carrying it out.

I’d like to propose a more modest version of Aumann’s agreement theorem, call it Aumann’s Less-Than-Total-Disagreement Theorem, which says that two rational agents shouldn’t both end up with 99.9…% confidence on opposite sides of the same problem.

The “proof” is pretty similar to the original. Suppose you are 99.9% confident about something, and learn your equally educated, intelligent, and clear-thinking friend is 99.9% confident of the opposite. Arguing with each other and comparing your evidence fails to make either of you budge, and neither of you can marshal the weight of a bunch of experts saying you’re right and the other guy is wrong. Shouldn’t the fact that your friend, using a cognitive engine about as powerful as your own, got so heavily different a conclusion make you worry that you’re missing something?

But practically everyone is walking around holding 99.9…% probabilities on the opposite sides of important issues! I checked the Less Wrong Survey, which is as good a source as any for people’s confidence levels on various tough questions. Of the 1400 respondents, about 80 were at least 99.9% certain that there were intelligent aliens elsewhere in our galaxy; about 170 others were at least 99.9% certain that they weren’t. At least 80 people just said they were certain to one part in a thousand and then got the answer wrong! And some of the responses were things like “this box cannot fit as many zeroes as it would take to say how certain I am”. Aside from stock traders who are about to go bankrupt, *who says that sort of thing??!*

And speaking of aliens, imagine if an alien learned about this particular human quirk. I can see them thinking *“Yikes, what kind of a civilization would you get with a species who routinely go around believing opposite things, always with 99.99…% probability?”*

Well, funny you should ask.

I write a lot about free speech, tolerance of dissenting ideas, open-mindedness, et cetera. You know which posts I’m talking about. There are a lot of reasons to support such a policy. But one of the big ones is – who the heck would burn heretics if they thought there was a 5% chance the heretic was right and they were wrong? Who would demand that dissenting opinions be banned, if they were only about 90% sure of their own? Who would start shrieking about “human garbage” on Twitter when they fully expected that in some sizeable percent of cases, they would end up being wrong and the garbage right?

Noah Smith recently asked why it was useful to study history. I think at least one reason is to medicate your own overconfidence. I’m not just talking about things like “would Stalin have really killed all those people if he had considered that he was wrong about communism” – especially since I don’t think Stalin worked that way. I’m talking about Neville Chamberlain predicting “peace in our time”, or the centuries when Thomas Aquinas’ philosophy was the preeminent Official Explanation Of Everything. I’m talking about Joseph “no one will ever build a working hot air balloon” Lalande. And yes, I’m talking about what Muggeridge writes about, millions of intelligent people thinking that Soviet Communism was great, and ending out disastrously wrong. Until you see how often people just like you have been wrong in the past, it’s hard to understand how uncertain you should be that you are right in the present. If I had lived in 1920s Britain, I probably would have been a Communist. What does that imply about how much I should trust my beliefs today?

There’s a saying that “the majority is always wrong”. Taken literally it’s absurd – the majority thinks the sky is blue, the majority don’t believe in the Illuminati, et cetera. But what it *might* mean, is that in a world where everyone is overconfident, the majority will always be wrong about which direction to move the probability distribution in. That is, if an ideal reasoner would ascribe 80% probability to the popular theory and 20% to the unpopular theory, perhaps most real people say 99% popular, 1% unpopular. In that case, if the popular people are urging you to believe the popular theory more, and the unpopular people are urging you to believe the unpopular theory more, the unpopular people are giving you better advice. This would create a strange situation in which good reasoners are usually engaged in disagreeing with the majority, and also usually “arguing for the wrong side” (if you’re not good at thinking probablistically, and almost no one is), but remain good reasoners and the ones with beliefs most likely to produce good outcomes. Unless you count “why are all of our good reasoners being burned as witches?” as a bad outcome.

I started off by saying this blog was about “the principle of charity”, but I had trouble defining it and in retrospect I’m not that good at it anyway. What can be salvaged from such a concept? I would say “behave the way you would if you were less than insanely overconfident about most of your beliefs.” This is the Way. The rest is just commentary.

Discussion Questions (followed by my own answers in ROT13)

1. What is your probability that there is a god? (Svir creprag)
2. What is your probability that psychic powers exist? (Bar va bar gubhfnaq)
3. What is your probability that anthropogenic global warming will increase temperatures by at least 1C by 2050? (Avargl creprag)
4. What is your probability that a pandemic kills at least one billion people in a 5 year period by 2100? (Svsgrra creprag)
5. What is your probability that humans land on Mars by 2050? (Rvtugl creprag)
6. What is your probability that superintelligent AI (=AI better than almost every human at almost every cognitive task) exists by 2115? (Gjragl svir creprag)

Probabilities Without Models

[Epistemic status: Not original to me. Also, I might be getting it wrong.]

A lot of responses to my Friday post on overconfidence centered around this idea that we shouldn't, we can't, use probability *at all* in the absence of a well-defined model. The best we can do is say that we don't know and have no way to find out. I don't buy this:

> "Mr. President, NASA has sent me to warn you that a saucer-shaped craft about twenty meters in diameter has just crossed the orbit of the moon. It's expected to touch down in the western United States within twenty-four hours. What should we do?"
>
> "How should I know? I have no model of possible outcomes."
>
> "Should we put the military on alert?"
>
> "Maybe. Maybe not. Putting the military on alert might help. Or it might hurt. We have literally no way of knowing."
>
> "Maybe we should send a team of linguists and scientists to the presumptive landing site?"
>
> "What part of 'no model' do you not understand? Alien first contact is such an inherently unpredictable enterprise that even speculating about whether linguists should be present is pretending to a certainty which we do not and cannot possess."
>
> "Mr. President, I've got our Israeli allies on the phone. They say they're going to shoot a missile at the craft

> because ‘it freaks them out’. Should I tell them to hold off?”
>
> “No. We have no way of predicting whether firing a missile is a good or bad idea. We just don’t know.”

In real life, the President would, despite the situation being totally novel and without any plausible statistical model, probably make some decision or another, like “yes, put the military on alert”. And this implies a probability judgment. The reason the President will put the military on alert, but not, say, put banana plantations on alert, is that in his opinion the aliens are more likely to attack than to ask for bananas.

Fine, say the doubters, but surely the sorts of probability judgments we make without models are only the most coarse-grained ones, along the lines of “some reasonable chance aliens will attack, no reasonable chance they will want bananas.” Where “reasonable chance” can mean anything from 1% to 99%, and “no reasonable chance” means something less than that.

But consider another situation: imagine you are a director of the National Science Foundation (or a venture capitalist, or an effective altruist) evaluating two proposals that both want the same grant. Proposal A is by a group with a long history of moderate competence who think they can improve the efficiency of solar panels by a few percent; their plan is a straightforward application of existing technology and almost guaranteed to work and create a billion dollars in value. Proposal B is by a group of starry-eyed idealists who seem very smart but have no proven track record; they say they have an idea for a revolutionary new kind of super-efficient desalinization technology; if it works it will completely solve the world’s water crisis and produce a trillion dollars in value. Your organization is risk-neutral to a totally implausible degree. What do you do?

Well, it seems to me that you choose Proposal B if you think it has at least a 1/1000 chance of working out; otherwise, you choose Proposal A. But this requires at least attempting to estimate probabilities in the neighborhood of 1/1000 without a model. Crucially, there’s no way to avoid this. If you shrug and take Proposal A because you don’t feel like you can assess proposal B adequately, that’s making a choice. If you shrug and take Proposal B because what the hell, that’s also making a choice. If you are so angry at being placed in this situation that you refuse to choose either A or B and so pass up both a billion *and* a trillion dollars, that’s a choice too. Just a stupid one.

Nor can you cry “Pascal’s Mugging!” in order to escape the situation. I think this defense is overused and underspecified, but at the very least, it doesn’t seem like it can apply in places where the improbable option is likely to come up over your own lifespan. So: imagine that your organization actually reviews about a hundred of these proposals a year. In fact, it’s competing with a bunch of other organizations that also review a hundred or so such proposals a year, and whoever’s projects make the most money gains lots of status and new funding. Now it’s totally plausible that, over the course of ten years, it might be a better strategy to invest in things that have a one in a thousand chance of working out. Indeed, maybe you can see the organizations that do this outperforming the organizations that don’t. The question really does come down to your judgment: are Project B’s odds of success greater or less than 1/1000?

Nor is this a crazy hypothetical situation. A bunch of the questions we have to deal with come down to these kinds of decisions made without models. Like – should I invest for retirement, even though the world might be destroyed by the time I retire? Should I support the Libertarian candidate for president, even though there's never been a libertarian-run society before and I can't know how it will turn out? Should I start learning Chinese because China will rule the world over the next century? These questions are no easier to model than ones about cryonics or AI, but they're questions we all face.

The last thing the doubters might say is "Fine, we have to face questions that *can* be treated as questions of probability. But we should avoid treating them as questions of probability anyway. Instead of asking ourselves 'is the probability that the desalinization project will work greater or less than 1/1000', we should ask 'do I feel good about investing this money in the desalinization plant?' and trust our gut feelings."

There is some truth to this. My medical school thesis was on the probabilistic judgments of doctors, and they're pretty bad. Doctors are just *extraordinarily* overconfident in their own diagnoses; a study by Bushyhead, who despite his name is not a squirrel, found that when doctors were 80% certain that patients had pneumonia, only 20% would turn out to have the disease. On the other hand, the doctors still did the right thing in most every case, operating off of algorithms and heuristics that never mentioned probability. The conclusion was that as long as you don't force doctors to think about about what they're doing in mathematical terms, everything goes fine – something I've brought up before in the context of the Bayes mammogram problem. Maybe this generalizes. Maybe people are terrible at coming up with probabilities for things like investing in desalinization plants, but will generally make the right choice.

But refusing to frame choices in terms of probabilities also takes away a lot of your options. If you use probabilities, you can check your accuracy – the foundation director might notice that of a thousand projects she had estimated as having 1/1000 probabilities, actually about 20 succeeded, meaning that she's overconfident. You can do other things. You can compare people's success rates. You can do arithmetic on them ("if both these projects have 1/1000 probability, what is the chance they both succeed simultaneously?"), you can open prediction markets about them.

Most important, you can notice and challenge overconfidence when it happens. I said last post that when people say there's only a one in a million chance of something like AI risk, they are being stupendously overconfident. If people just very quietly act as if there's a one in a million chance of such risk, without ever saying it, then no one will ever be able to call them on it.

I don't want to say I'm completely attached to using probability here in exatly the normal way. But all of the alternatives I've heard fall apart when you've got to make an actual real-world choice, like sending the military out to deal with the aliens or not.

[**EDIT:** Why regressing to meta-probabilities just gives you more reasons to worry about overconfidence]

[**EDIT-2:** "I don't know"]

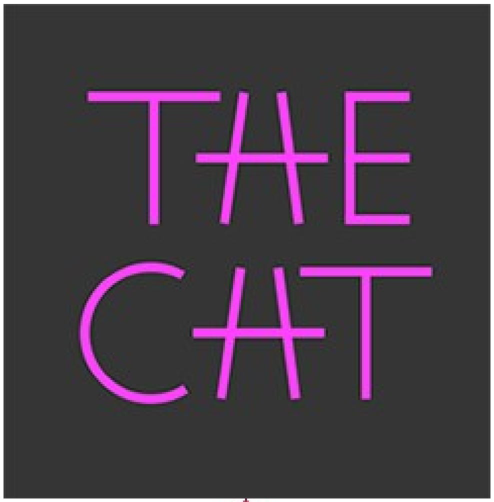
[**EDIT-3:** A lot of debate over what does or doesn't count as a "model" in this case. Some people seem to be using a weak definition like "any knowledge whatsoever about the process involved". Others seem to want a strong definition like "enough understanding to place this event within a context of similar past events such that a numerical probability can be easily extracted by math alone, like the model where each flip of a two-sided coin has a 50% chance of landing heads”. Without wanting to get into this, suffice it to say that any definition in which the questions above have “models” is one where AI risk also has a model.]

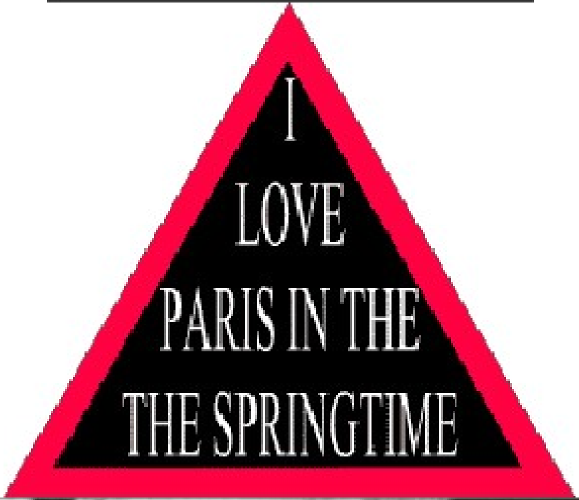
Mysticism and Pattern-Matching

[Epistemic status: Total conjecture.]

One of the things that got me interested in psychiatry was the sheer weirdness of the human brain’s failure modes. We all hear that the brain is like a computer, but when a computer breaks, the screen goes black or it freezes or something. It doesn’t hear voices telling it that it’s Jesus, or start seeing tiny men running around on the floor. But for some reason, when the the human brain breaks, it may do exactly that. Why?

Psychiatry classes never just *tell* you the answer to this question, but reading between the lines I think it has something to do with top-down processing and pattern matching.




I
LOVE
PARIS IN THE
THE SPRINGTIME

Bottom-up processing is when you go from basic elements to more complex ideas – for example, when you see the three letters C, A, and T in a row, you might combine them to get the the word CAT. Top-down processing is when more complex ideas change the way you interpret basic elements. For example, in the first picture above, the middle letters in both words are the same. We read the first as H, because the image as a gestalt suggests the word “THE” and the word “THE” suggests an H in the middle. We read the second as A, because the image as a gestalt suggests the word “CAT” and the word “CAT” has an A in the middle. Our big-picture idea has changed the way we view the smaller elements composing it.

The same is true of the second image. We recognize the phrase “PARIS IN THE SPRINGTIME”, and so we assume that’s what the sign is trying to show us. In fact, the sign doubles the word “the”. But since this is bizarre and not something that makes sense in the gestalt, we assume this is a mistake and gloss right over it. We do this very, *very* easily – how many times have I duplicated the word “the” in this essay already?

The third image is related to this tendency. To most people, it looks formless. Even once they hear that it’s an old black-and-white photograph of a cow’s head, it’s might still require a bit of staring before you catch on. But once you see the cow, the cow is *obvious*. It becomes impossible to see it as formless, impossible to see it as anything else. Having given yourself a top-down pattern to work from, the pattern automatically organizes the visual stimuli and makes sense of them.

This provides a possible explanation for hallucinations. Think of top-down processing as taking noise and organizing it to fit a pattern. Normally, you’ll only fit it to the patterns that are actually there. But if your pattern-matching system is broken, you’ll fit it to patterns that aren’t in the data at all.

The best example of this is Google Deep Dream:

I don’t know much about neural networks, so I may not be getting this entirely right, but as far as I understand it, they trained a neural network on some stimulus like a dog. This was for research in machine vision; they wanted the net to be able to recognize dogs when it saw them; to pattern-match potentially noisy images of dogs into its Platonic ideal of a dog. But if you turn the pattern-matching up, it will just start seeing dogs everywhere there’s even the slightest amount of noise that resembles a dog at all. You only matched the sign above to “PARIS IN THE SPRINGTIME” because it was almost exactly like that phrase; if we stick your pattern-matching software into overdrive, maybe *every* sentence would start looking like more meaningful alternatives. Eevn sceeentns wtih aolsmt all the lerttes rergaearnd wulod naelry ianslntty sanp itno pacle. Turn it *all* the way up, and maybe you could make every sentence look like “PARIS IN THE SPRINGTIME”. Or something.

So hallucinations are when your top-down processing/pattern-matching ability becomes so dysfunctional that it can generate people and objects out of random visual noise. Why it chooses some people and objects over others I don’t know, but it’s hardly surprising – it does the same thing every night in your dreams.

Many of the same people who have hallucinations also have paranoia. Paranoia seems to me to be overfunctioning of social pattern-matching. When Deep Dream sees the tiniest hint of a line here, a slight dark spot there, it pattern-matches it into an entire dog. When a paranoiac hears a stray word here, or sees a sideways glance there, they turn it into this vast social edifice of connected plots. Every new thing that happens is fit effortlessly into the same pattern. When their psychiatrist says they’re crazy, that gets fit into the pattern too – maybe the psychiatrist is a tool of the conspiracy, trying to confuse them into compliance.

So where does the mysticism come in?

I notice that the same people who have hallucinations also have mystical experiences. By mystical experiences, I don’t just mean “they see angels” – in that case, the relationship to hallucination would be a tautology. I mean they feel a sense of sudden understanding of and connection with the universe. I know at least three groups that do this: druggies, meditators, and prophets. The druggies report feelings of total understanding on their drugs, and also report hallucinations. The meditators occasionally achieve enlightenment, but look at any text about meditation and you find mentions of visions and hallucinations experienced during the practice. The voices heard by the prophets are too obvious to mention.

One well-known way of bringing on such experiences is to abuse your pattern-matching faculty. The Chicken Qabalah of Rabbi Lamed Ben Clifford (not really recommended) manages to link a pretty boring Bible verse to the letter yud, the creativity of God, the essence of existence, the sun, the phallus, the plane of Malkuth, and the number 496, then explains:

> Like a mountain goat leaping ecstatically from crag to crag, one thought springs into another, and another, ad infinitum. You can continue, almost forever, connecting things that you never thought were connected. Sooner or later something’s going to snap and you will overcome the fundamental defect in your powers of perception.

And:

> Was that the message Ezekiel was trying to convey? Probably not. But who cares! Whatever it was the old boy was originally trying to say shrinks to insignificance. It is far more important to *my* spiritual enlightenment that

> my mind was forced to churn at breakneck speed to put all of this together, and then open itself up to the infinite possibilities of meaning. Look hard enough at anything and eventually you will see everything! it doesn't even have to make very much sense what you connect to what. It's all ultimately connected!

This philosophy, which I associate both with kabbalah and with the more modern Western hermetic tradition, says that learning a set of extremely complicated correspondences is an important step toward gaining enlightenment. See for example this site, which helpfully relates the sephirah Netzach to the planet Venus, the number 7, the emerald, the lynx, the rose, cannabis, arsenic, copper, fire, the solar plexus chakra, the archangel Haniel, the Egyptian goddess Hathor, the concepts of love and victory, et cetera, et cetera. You're supposed to be able to use this to interpret things – for example, if you have a dream about a lynx, it could correspond to anything else in the system – but it looks like it would quickly get unwieldy. And other sources will give completely different systems of correspondences, and nobody gets too upset over it – in fact, some sources will happily encourage you to come up with your own correspondences instead, as long as you stick to them. It seems like the goal is less "remember that it's extremely important that emeralds correspond to lynxes in reality" and more "have some system, any system, of interesting correspondences in mind that you can apply to everything you come across".

Nor does it especially matter what you're interpreting. The traditional things to interpret are mysterious things like dreams, or the Bible, but Crowley famously performs a mystical analysis of Mother Goose nursery rhymes (see Interlude here). The important factor seems to be less about there being sacred truth in the object being analyzed, and more about the process of performing the analysis.

(Zen koans are a little different, but also sort of involve torturing a pattern-finding ability for apparently no reason)

So to skip to the point: I think all of this is about strengthening the pattern-matching faculty. You’re exercising it uselessly but impressively, the same way as the body-builder who lifts the same weight a thousand times until their arms are the size of tree trunks. Once the pattern-matching faculty is way way way overactive, it (spuriously) hallucinates a top-down abstract pattern in the whole universe. This is the experience that mystics describe as “everything is connected” or “all is one”, or “everything makes sense” or “everything in the universe is good and there for a purpose”. The discovery of a beautiful all-encompassing pattern in the universe is understandably associated with “seeing God”.

Religious scholar William James once experimented with nitrous oxide and reached a state where he felt he had total comprehension of the universe. According to a story which I can’t verify, he became infuriated at losing the thread of understanding once the chemical wore off, so he decided to take notes during the experience: write down the secrets of the universe then, and reread them once he was sober. The experiment completed, he picked up the notepad in feverish excitement, only to find that he had written OVERALL THERE IS A SMELL OF FRIED ONIONS.

Imagine one of those Google robots pointing at an empty patch of sky and saying “No, look, seriously, there’s a dog right there. Right there! How are you not seeing this?” Things that make perfect sense in the context of a state of overactive pattern-matching look meaningless to a pattern-matching faculty operating normally. At best, you can sort of see the lines of what seemed so clear before ("Yeah, I can see that that stain on the wall is vaguely dog-shaped.") This matches the stories I've heard of people who have some mystical experience but then can't maintain or recapture it.

I think other methods of inducing weird states of consciousness, like drugs and meditation, probably do the same thing by some roundabout route. Meditation seems like reducing stimuli, which is known to lead to hallucinations in eg sensory deprivation tanks or solitary confinement cells in jail. I think the general principle is that a low level of external stimuli makes your brain adjust its threshold for stimulus detection up until anything including random noise satisfies the threshold. As for drugs, there's lots of reasons to think that the neurotransmission changes they create will alter the brain's pattern processing strategies.

Things this hypothesis doesn't explain: why mystical experiences are linked with a feeling of no time, no space, and no self; why prayer or extreme devotion seems to induce them (eg bhakti yoga), and why they can be so beneficial – that is, why do people with mystical experiences become happier and better adjusted? Maybe the feeling of the world making sense is naturally a pleasant and helpful one. Certainly the opposite can be very stressful!

Magic Markers

[Thanks to some people on Rationalist Tumblr, especially prophecyformula, for help and suggestions.]

There's an old philosophers' saying – trust those who seek the truth, distrust those who say they've found it. The psychiatry version of this goes "Trust those who seek biological underpinnings for mental illness, distrust those who say they've found them."

Niculescu et al (2015) say they've found them. Their paper describes a process by which they hunted for biomarkers – in this case changes in gene expression – that predict suicide risk among psychiatric patients. They test various groups of psychiatric patients (including post-mortem tissue from suicide victims) to find some plausible genes. Then they use those genes to predict suicidality in two cohorts of about 100 patients each, including people with depression, schizophrenia, schizoaffective disorder, and bipolar disorder. They arrive at an impressive 92% AUC – that being the area under the curve graphing sensitivity vs. specificity, a common measure of the accuracy with which they can distinguish people who will vs. won't be suicidal in the future.

The science press, showing the skepticism and restraint for which they are famous, jump on board immediately. A New Blood Test Can Predict Whether A Patient Will Have Suicidal Thoughts With More Than 90% Accuracy, says *Popular Science*. New Blood Test Predicts Future Suicide Attempts, says PBS.

There is a procedure for this sort of thing. The procedure is that the rest of us sit back and quietly wait for James Coyne, author of How To Critique Claims For A Blood Test For

Depression, to tell us exactly why it is wrong. But it’s been over a week now and this hasn’t happened and I’m starting to worry he’s asleep on the job. So even though this is somewhat outside my area of expertise, let me discuss a couple of factors that concern me about this study.

The 92% accuracy claim is for the authors’ model, called UP-SUICIDE, which combines 11 biomarkers and two clinical prediction instruments. A clinical prediction instrument is a test which asks questions like “How depressed are you feeling right now?” or “How many times have you attempted suicide before?”. By combining the predictive power of the eleven genes and two instruments, they managed to reach the 92% number advertised in the abstract.

It might occur to you to ask “Wait, a test in which you can just ask people if they’re depressed and hate their life sounds a lot easier than this biomarker thing. Are we sure that they’re not just getting all of their predictive power from there?”

The answer is: no, we’re not sure at all, and as far as I can tell the study goes to great pains in order to make it hard to tell to what degree they are doing this.

Conventional wisdom says that clinical instruments for predicting suicidality can attain AUCs of 0.74 to 0.88. This is most of the way to the 0.92 shown in the current study, but not *quite* as high. But the current study combines two different clinical prediction instruments. In Combining Scales To Assess Suicide Risk, a Spanish team combines a few different clinical prediction instruments to get an AUC of… 0.92.

If you look really closely at Niculescu et al’s big results table, you find that each of the individual prediction instruments they use does almost as well – and in some cases better than – their UP-SUICIDE model as a whole. For example, when predicting suicidal ideation in all patients, the CFI-S instrument has an AUC of 0.89, compared to the entire model’s 0.92. When predicting suicide-related hospitalizations in depressed patients, the CFI-S has an AUC of 0.78, compared to the entire model’s 0.7. Here the biomarkers are just adding noise!

Are the cases where the entire model outperforms the CFI-S cases where the biomarkers genuinely help? We have no way of knowing. There are two clinical prediction instruments, the CFI-S and the SASS. Combined, they should outperform either one alone. So, for example, on suicidal ideation among all patients, the SASS has an AUC of 0.85, the CFI-S has an AUC of 0.89, and the model as a whole (both instruments combined + 11 biomarkers) has an AUC of 0.92. If we just combined the CFI-S and SASS, and threw out the biomarkers, would we do better or worse than 0.92? I don’t know and they don’t tell us. When all we’re doing is looking at the overall model, the biomarkers may be helping, hurting, or totally irrelevant.

So what if we throw out the clinical prediction instruments and just look at the biomarkers?

The authors use their panel of biomarkers for four different conditions: depression, bipolar, schizophrenia, and schizoaffective. And they have two different outcomes: suicidal ideation according to a test of such, and actual hospitalization for suicide. That’s a total of 4 x 2 = 8 tests that they’re conducting.

Of these eight different tests, the panel of biomarkers taken together come back insignificant on seven of them.

And there’s such a thing as “trending towards significance”, but this isn’t it. Here, I’ll give p-values:

Depression/ideation: p = 0.26
Depression/hospitalization: p = 0.48
Schizoaffective/ideation: p = 0.46
Schizoaffective/hospitalization: p = 0.94
Schizophrenia/ideation: p = 0.16
Schizophrenia/hospitalization: p = 0.72
Bipolar/hospitalization: p = 0.24

The only test of the eight that comes out significant is bipolar/ideation, where p = 0.007. This is fine (well, it’s fine if it’s supposed to be post-Bonferroni correction, which I can’t be sure of from the paper). But I notice three things. Number one, there were only 29 people in this group. Number two, some of the most impressive looking genes for the ideation condition were worthless for the hospitalization condition. CLIP4, which got p = 0.005 for the ideation condition, got p = 0.91 for the second condition and actually had negative predictive value. And third, some of the genes that best predicted bipolar in the validation data had no predictive value for bipolar at all in the training data, and were included only because they predicted major depressive disorder alone. Given that the effects jump across diagnoses and fail to carry over into even a slightly different method of assessing suicidality, this looks a lot less like a real finding and a lot more like a statistical blip.

Finally, note that even in bipolar ideation, their one apparent success, the biomarkers only got an AUC of 0.75, lower than either clinical predictive instrument. The only reason their model did better was because it added on the clinical predictive instruments themselves.

So here it looks like seven out of their eight tests failed miserably, one of them succeeded in a very suspicious way, and they covered over this by combining the data with the clinical predictive instruments which always worked very well. Then everyone interpreted this as the sexy and exciting result “biomarkers work!” rather than the boring result “biomarkers fail, but if you use other stuff instead you’ll still be okay.”

The *absolute strongest* conclusion you can draw from this study is “biomarkers may predict risk of suicidal ideation in bipolar disorder with an AUC of 0.75”. Instead, everyone thinks biomarkers predict suicidality and hospitalization in a set of four different disorders with AUC of 0.92, which is way beyond what the evidence can support.

II.

So much for that. Now let me explain why it wouldn’t matter much even if they were right.

AUC is a combination of two statistics called sensitivity and specificity. It’s a little complicated, but if we assume it means sensitivity and specificity are both 92% we won’t be far off.

Sensitivity is the probability that a randomly chosen positive case in fact tests positive. In this case, it means the probability that, if someone is actually going to be suicidal, the model flags them as high suicide risk.

Specificity is the probability that a randomly chosen negative case in fact tests negative. In this case, it means the probability that, if someone is not going to be suicidal, the model flags them as low suicide risk.

In this study population, about 7.5% of their patients are hospitalized for suicidality each year. So suppose you got a million depressed people similar to these. 75,000 would attempt suicide that year, and 925,000 wouldn't.

Now, suppose you gave your million depressed people this test with a 92% sensitivity and specificity.

Of the 925,000 non-suicidal people, 92% – 851,000 – will be correctly evaluated as non-suicidal. 74,000- 8% – will be mistakenly evaluated as suicidal.

Of the 75,000 suicidal people, 92% – 69,000 – will be correctly evaluated as suicidal. 8% – 6,000 – will be mistakenly evaluated as non-suicidal.

But this means that of the 143,000 people the test says are suicidal, only 69,000 – less than half – actually will be!

So when people say "We have a blood test to diagnose suicidality with 92% accuracy!", *even if it's true*, what they mean is that they have a blood test which, if it comes back positive, there's still less than 50-50 odds the person involved is suicidal. Okay. Say you're a psychiatrist. There's a 48% chance your patient is going to be suicidal in the next year. What are you going to do? Commit her to the hospital? I sure hope not. Ask her some questions, make sure she's doing okay, watch her kind of closely? You're a psychiatrist and she's your depressed patient, you would have been doing that *anyway*. This blood test is not really actionable.

And then remember that this isn't the blood test we have. We have some clinical prediction instruments that do this, and we have a blood test which maybe, if you are very trusting, diagnoses suicidality in bipolar disorder with 75% accuracy. At 75% sensitivity and specificity, only twenty percent of the people who test positive will be suicidal. So what?

There will never be a blood test for suicide that works 100%, because suicide isn't 100% in the blood. I am the most biodeterminist person you know (unless you know JayMan), I am happy to agree with Martin and Tesser that that the heritability of learning Latin is 26% and the heritability of jazz is 45% and so on, but suicide is *not just biological*. Maybe people need some kind of biological predisposition to consider suicide. But whether they go ahead with it or not depends on whether they have a good or bad day, whether their partner breaks up with them, whether a friend hands them a beer and they get really drunk, et cetera. Taking all of this into account, it's really unlikely that a blood test will ever get sensitive and specific enough to overcome these hurdles.

We should continue research on the biological underpinnings of depression and suicide, both for the sake of knowledge and because it might lead to better treatments. But having "a blood test for suicide" won't be very useful, *even if it works*.

Theses On Trump

The smart money still says Trump will crash and burn before getting the nomination. But everyone saying this should have to add "however, he's certainly lasted much longer than we originally predicted."

Scott Adams places the blame for this surprising perseverance on Trump-specific factors – namely, his brilliance as a persuader and manipulator. The mainstream media reports I've seen place it on Republican-specific factors, some combination of the words "ignorant", "bigoted", "white", and "aggrieved entitlement", followed by "therefore, Trump" – without a good explanation for why not therefore Scott Walker or therefore Rick Santorum.

I would argue it reflects a more general trend. As Exhibit A, I bring before the court Jeremy Corbyn.

In case you're not following the United Kingdom: Labour, the UK's major leftist party, is having leadership elections. Jeremy Corbyn is expected to win. In most ways, Corbyn is the diametric opposite of Trump. Trump is on the right, Corbyn is on the far left. Trump is inconsistent in his policy stances to the degree he even has any; Corbyn has strong opinions which he never deviates from even an inch. Trump has never held office; Corbyn has toiled in an unimpressive Parliamentary back bench position for decades. Trump is famously loud and bombastic, Corbyn is famously quiet and reserved, and likes to talk about things like gardening and how more people should be beekeepers.

But their rises to power look weirdly similar. Corbyn spent twenty years on the edges of British politics, mostly a figure of fun for his weird out-of-touch positions. When he announced the intention to run for Labour leader, everyone thought of it as a quixotic attempt to gain some free coverage, like when Dennis Kucinich runs for President. Some of the MPs who signed off on his candidacy request forms said they did it because they figured it wouldn’t matter one way or the other so they might as well make him happy. The media talked about how silly it was that he was even running at all. Then he started gathering momentum. The establishment freaked out and told everyone he was unacceptable and they were not to vote for him under any circumstances. A parade of important figures and personalities went on the news to personally beg voters not to vote for him under any circumstances, ably parodied by the Twitter account @corbynwarnings:

> As Prime Minister, Corbyn Would Renationalise Children, Warns MP
>
> — Corbyn Warnings (@CorbynWarnings) August 11, 2015
>
> Corbyn 'Obsessed With Destroying The Moon', Warns MP
>
> — Corbyn Warnings (@CorbynWarnings) August 11, 2015
>
> Antibiotics Not Effective Against Corbyn, Warn Doctors
>
> — Corbyn Warnings (@CorbynWarnings) August 13, 2015
>
> Estimated 100% Of People Who Vote For Corbyn Will Die Within A Century, Warn Scientists
>
> — Corbyn Warnings (@CorbynWarnings) August 12, 2015

Nevertheless, Corbyn seems amply placed to win the election, and British bookmakers give 3:1 odds in his favor. What *happened?*

Well, for one thing, this is the first year people are allowed to vote directly for the Labour leader. So it might just mean the British Left was really really far left for a long time and nobody noticed before this.

But the analogy with Trump seems a little too good. People like the same things about both of them. They speak their mind. They don’t care what anyone else thinks. And the establishment obviously hates both.

People have always liked outsiders. But now it’s starting to get ridiculous.

Everyone knows that America is getting more ideologically polarized these days. The right is getting rightier. The left is getting leftier. This puts the Establishment in a bind. The winning strategy had always been to play to the fringe for the primary, then veer towards the center for the general election; the fringe would grumble, but if you played it right you could mollify them and be all things to all people. But now the distance from the fringe to the center is much larger, and proportionally harder to cover without an *obvious* betrayal. A candidate who wants to get elected on the national level, or even on a local level if the local area is ideologically diverse enough, has to make that betrayal.

Worse, once they’re elected they’ve got to deal with reality. If you try to be too liberal (like raising the minimum wage to $15) or too conservative (like building an immigration wall), then businesspeople with a vested interest in the economy continuing to work start yelling at you, and maybe you back down. There’s this archetypal image of the new President-elect walking into the White House on day one and very serious men in suits telling him "Okay, here are the planks of your campaign platform which don't have a snowball's chance in Hell of working and which you are going to drop," and probably he listens. The clearest example here would be Obama promising to close Guantanamo Bay by 2010, followed by him getting elected and someone asking "Okay, and exactly how are you going solve all of the legal hurdles to doing this?"

The base doesn't have to worry about reality. If, as people like Robin Hanson suggest, politics is not about carefully selecting policies that most benefit the country so much as about signalling values and ingroup membership, then the base will be interested in enforcing its own particular extremism.

In the past, the center and the fringe were close enough for an uneasy compromise: sure, the base would gripe about "the establishment" and light up at the mention of "an outsider", but sufficiently canny politicians could still navigate a careful path between their competing demands.

Now that's becoming harder. The base thinks of the establishment not just as suspicious but as actively hostile; thus the rise of the Tea Party, whose whole purpose was to elect a new kind of conservative who wouldn't cave in to big government liberals like all the other kinds of conservatives did, and who were a constant thorn in the side of a Republican Party trying to get the most electable people in power so they could do the most electable things.

But now the Tea Party's actually attained some power, they can't deny reality any better than their predecessors, and so they keep doing all sorts of crazy things like *not* shutting down the government over each real or imagined slight. Where do you go from there?

Apparently, you go to Trump.

The most salient feature of Trump – I would say the *only* salient feature of Trump – is that the establishment hates him. Reince Priebus goes to sleep at night and has nightmares about Trump. The liberal media has important-looking people coming on in suits saying it's a national embarrassment that anyone could vote for Trump. But in signaling terms, what they're unintentionally saying is "Moderates hate this guy! He's too politically incorrect to win over Democrats! Only vote for him if you're a *real Republican*." And Republicans are *eating it up*. It doesn't even matter that he's not that conservative in real life, the media has conducted his campaign for him. Every bad thing the media and the establishment say about him will just make him more popular.

And the same seems true of Jeremy Corbyn.

> Warning About Corbyn Only Makes Him Stronger, Warns MP
>
> — Corbyn Warnings (@CorbynWarnings) [August 14, 2015]

Trace this tendency far enough, and I think it explains why Bernie Sanders is doing better than expected, why Ben Carson has the Republican 2nd place right now, and maybe why Obama won his surprise upset over Hillary in 2008. I predict we are in for a lot more interesting Corbyn-style surprises.

If You Can’t Make Predictions, You’re Still In A Crisis

A *New York Times* article by Northeastern University professor Lisa Feldman Barrett claims that Psychology Is Not In Crisis:

> Is psychology in the midst of a research crisis?
>
> An initiative called the Reproducibility Project at the University of Virginia recently reran 100 psychology experiments and found that over 60 percent of them failed to replicate — that is, their findings did not hold up the second time around. The results, published last week in Science, have generated alarm (and in some cases, confirmed suspicions) that the field of psychology is in poor shape.
>
> But the failure to replicate is not a cause for alarm; in fact, it is a normal part of how science works.
>
> Suppose you have two well-designed, carefully run studies, A and B, that investigate the same phenomenon. They perform what appear to be identical experiments, and yet they reach opposite conclusions. Study A produces the predicted phenomenon, whereas Study B does not. We have a failure to replicate.
>
> Does this mean that the phenomenon in question is necessarily illusory? Absolutely not. If the studies were well designed and executed, it is more likely that the phenomenon from Study A is true only under certain conditions. The scientist’s job now is to figure out what those conditions are, in order to form new and better hypotheses to test […]

> When physicists discovered that subatomic particles didn’t obey Newton’s laws of motion, they didn’t cry out that Newton’s laws had “failed to replicate.” Instead, they realized that Newton’s laws were valid only in certain contexts, rather than being universal, and thus the science of quantum mechanics was born […]
>
> Science is not a body of facts that emerge, like an orderly string of light bulbs, to illuminate a linear path to universal truth. Rather, science (to paraphrase Henry Gee, an editor at Nature) is a method to quantify doubt about a hypothesis, and to find the contexts in which a phenomenon is likely. Failure to replicate is not a bug; it is a feature. It is what leads us along the path — the wonderfully twisty path — of scientific discovery.

Needless to say, I disagree with this rosy assessment.

The first concern is that it ignores publication bias. One out of every twenty studies will be positive by pure chance – more if you’re willing to play fast and loose with your methods. Probably quite a lot of the research we see is that 1/20. Then when it gets replicated in a preregistered trial, it fails. This is not because the two studies were applying the same principle to different domains. It’s because the first study posited something that simply wasn’t true, in any domain. This may be *the outright majority* of replication failures, and you can’t just sweep this under the rug with paeans to the complexity of science.

The second concern is experimenter effects. Why do experimenters who believe in and support a phenomenon usually find it occurs, and experimenters who doubt the phenomenon usually find that it doesn’t? That’s easy to explain through publication bias and other forms of bias, but if we’re just positing that there are some conditions where it does work and others where it doesn't, the ability of experimenters to so often end out in the conditions that flatter their preconceptions is a remarkable coincidence.

The third and biggest concern is the phrase "it is more likely". Read that sentence again: "If the studies were well designed and executed, it is more likely that the phenomenon from Study A is true only under certain conditions [than that it is illusory]". Really? *Why?* This is exactly the thing that John Ioannidis has spent so long arguing against! Suppose that I throw a dart at the Big Chart O' Human Metabolic Pathways and when it hits a chemical I say "This! This is the chemical that is the key to curing cancer!". Then I do a study to check. There's a 5% chance my study comes back positive by coincidence, an even higher chance that a biased experimenter can hack it into submission, but a much smaller chance that out of the thousands of chemicals I just so happened to pick the one that really does cause cancer. So if my study comes back positive, but another team's study comes back negative, it's not "more likely" that my chemical does cure cancer but only under certain circumstances. Given the base rate – that most hypotheses are false – it's more likely that I accidentally proved a false hypothesis, a very easy thing to do, and now somebody else is correcting me.

Given that many of the most famous psychology results are either extremely counterintuitive or highly politically motivated, there is *no reason at all* to choose a prior probability of correctness such that we should try to reconcile our prior belief in them with a study showing they don't work. It would be like James Randi finding Uri Geller can't bend spoons, and saying "Well, he bent spoons other times, but not around Randi, let's try to figure out what feature of Randi's shows interferes with the magic spoon-bending rays". I am not saying that we *shouldn't* try to reconcile results and failed replications of those results, but we should do so in an informed Bayesian way instead of automatically assuming it’s “more likely” that they deserve reconciliation.

Yet even ignoring the publication bias, and the low base rates, and the statistical malpractice, and the couple of cases of outright falsification, and concentrating on the ones that really *are* differences in replication conditions, this is *still* a crisis.

A while ago, Dijksterhuis and van Knippenberg published a famous priming study showing that people who spend a few minutes before an exam thinking about brilliant professors will get better grades; conversely, people who spend a few minutes thinking about moronic soccer hooligans will get worse ones. They did four related experiments, and all strongly confirmed their thesis. A few years later, Shanks et al tried to replicate the effect and couldn’t. They did the same four experiments, and none of them replicated at all. What are we to make of this?

We could blame differences in the two experiments’ conditions. But the second experiment made every attempt to match the conditions of the first experiment as closely as possible. Certainly they didn’t do anything idiotic, like switch from an all-female sample to an all-male sample. So if we want to explain the difference in results, we have to think on the level of tiny things that the replication team wouldn’t have thought about. The color of the wallpaper in the room where the experiments were taking place. The accents of the scientists involved. The barometric pressure on the day the study was conducted.

We could laboriously test the effect of wallpaper color, scientist accent, and barometric pressure on priming effects, but it would be extraordinarily difficult. Remember, we've already shown that two well-conducted studies can get diametrically opposite results. Who is to say that if we studied the effect of wallpaper color, the first study wouldn't find that it made a big difference and the second study find that it made no difference at all? What we'd probably end out with is a big conflicting morass of studies that's even more confusing than the original smaller conflicting morass.

But as far as I know, nobody is doing this. There is not enough psychology to devote time to teasing out the wallpaper-effect from the barometric-pressure effect on social priming. Especially given that maybe at the end of all of these dozens of teasing-apart studies we would learn nothing. And that quite possibly the original study was simply wrong, full stop.

Since we have not yet done this, and don't even know if it would work, we can expect even strong and well-accepted results not to apply in even very slightly different conditions. But that makes claims of scientific understanding very weak. When a study shows that Rote Memorization works better than New Math, we hope this means we've discovered something about human learning and we can change school curricula to reflect the new finding and help children learn better. But if we fully expect that the next replication attempt will show New Math is better than Rote Memorization, then that plan goes down the toilet and we shouldn't ask schools to change their curricula at all, let alone claim to have figured out deep truths about the human mind.

Barrett states that psychology is not in crisis, because it's in a position similar to physics, where gravity applies at the macroscopic level but not the microscopic level. But if you ask a physicist to predict whether an apple will fall up or down, she will say “Down, obviously, because we’re talking about the macroscopic level.” If you ask a psychologist to predict whether priming a student with the thought of a brilliant professor will make them do better on an exam or not, the psychologist will have no idea, because she won’t know what factors cause the prime to work sometimes and fail other times, or even whether it really ever works at all. She will be at the level of a physicist who says “Apples sometimes fall down, but equally often they fall up, and we can’t predict which any given apple will do at any given time, and we don’t know why – but our field is not in crisis, because in theory some reason should exist. Maybe.”

If by physics you mean “the practice of doing physics experiments”, then perhaps that is justified. If by physics you mean “a collection of results that purport to describe physical reality”, then it’s clear you don’t actually have any.

So the *Times* article is not an argument that psychology is not in crisis. It is, at best, an IOU, saying that we should keep doing psychology because maybe if we work really hard we will reach a point where the crisis is no longer so critical.

On the other hand, there’s one part of this I agree with entirely. I don’t think we can do a full post-mortem on every failed replication. But we ought to do them on *some* failed replications. Right now, failed replications are *deeply mysterious*. Is it really things like the wallpaper color or barometric pressure? Or is it more sinister things, like failure to double-blind, or massive fraud? How come this keeps happening to us? I don’t know. If we could solve one or two of these, we might at least know what we’re up against.

Book Review: Manufacturing Consent

I.

Consider:

> It is our view that, among their other functions, the media serve and propagandize on behalf of the powerful societal interests that control…them. The representatives of these interests have important agendas and principles that they want to advance, and they are well-positioned to shape and constrain media policy. This is normally not accomplished by crude intervention, but by the selection of right-thinking personnel and by the editors' and working journalists' internalization of priorities and definitions of newsworthiness that conform to the institution's policy.
>
> [This includes] the ability to complain about the media's treatment of news (that is, produce "flak"), to provide "experts" to confirm the official slant on the news, and to fix the basic principles and ideologies that are taken for granted by media personnel and the elite, but are often resisted by the general population. In our view, the same underlying power sources that own the media…that serve as primary definers of the news, and that produce flak and proper-thinking experts, also play a key role in fixing basic principles and dominant ideologies.

If I saw this quote on Facebook without attribution, I would assume it was from the latest far-right blog complaining about the liberal media. In fact, it is from Noam Chomsky and Edward Herman's Manufacturing Consent, which claims that the media acts as lapdog of the dominant neoliberal ideology against leftists of all stripes.

I decided to read *Manufacturing Consent* because of this basic puzzle: how can both the Left and Right be so certain that the media is biased against them?

Now, in one sense this is not surprising. Everyone believes everything is biased against them. I've previously talked about bravery debates, the sort of argument where both sides believe that *we're* brave non-conformist speaking truth to power, and *they're* toadies of the elite repeating the dominant consensus like sheep. The hostile media effect is a well-known bias where both sides of an issue believe the media is biased against them, even going so far as to both give low fairness ratings to sample documentaries in controlled studies for opposite reasons. So a more general tendency of both sides to accuse the media as a whole of having a hostile agenda is pretty much what we would predict.

The part that surprises me is: I thought that, even objectively, apart from the bias to be expected on both sides, the Right's case for a hostile media was pretty good. Democrats outnumber Republicans among journalists four to one, and CrowdPAC's donation analysis rates journalism as among the most liberal professions. There's an ongoing joke (and some informal analysis) about how disgraced Republicans' party affiliation is lampshaded and disgraced Democrats' party affiliation is covered up. And in my own area of interest, it often seems like scientific studies that support liberal beliefs tend to get front-page billing no matter how terrible they are, but scientific studies that cast doubt upon such beliefs them are very rarely mentioned.

And this perception seems to be mirrored by the popular wisdom, where conservatives complain of media bias full stop, and liberals mostly just gripe about Fox in particular.

So Chomsky and Herman's claim that the media is in fact biased towards conservatives is startling and interesting and deserves a further look.

How exactly do Chomsky and Herman think this media bias works? In Chapter 1, they propose five major mechanisms:

1. The mass media is mostly controlled by large corporations, who therefore support the sorts of things large corporations would be likely to support, like unrestrained capitalism and privileges for the wealthy.

2. The mass media is dependent on advertising, which also involves large corporations who support the sorts of things large corporations are likely to support. Further, these advertisers may have specific interests. For example, Texaco might be less willing to advertise in a source that frequently critiques Big Oil or raises concerns about pollution.

3. Journalists are dependent on sources. The most convenient sources are large well-organized entities in the midst of newsworthy events who issue press releases. For example, by far the easiest source for the latest news about a foreign war is the Pentagon. Furthermore, the Pentagon, while not always in fact trustworthy, enjoys a presumption of trustworthiness; if you interview some random foreigner, you would want to fact-check her very carefully, but if you parrot the Pentagon press release, you are assumed to have done due diligence merely because the source is so official. Other such convenient and official sources of news include the White House, the Department of State, local police forces, and local chambers of commerce. But all of these are

members of the establishment and so have a pro-establishment bias. Further, the news relies on “experts” to confirm and comment upon news, and because of incestuous relationships between government, corporations, think tanks, and academia, the most credentialed and salient experts will almost always be pro-establishment.

4. Conservative groups fund “flak machines”, organizations and individuals whose job it is to complain that the media is “biased” whenever they are insufficiently conservative. In these cases, relentless nitpickers will shriek about every slight inaccuracy and condemn the journalists involved as liars and unpatriotic to boot. If the media parrots the official line, then journalists can be almost arbitrarily sloppy and nobody will call them on it. Therefore, journalists who get ground down by the constant harassment will unconsciously shift towards more pro-establishment narratives.

5. Anti-communism is “the dominant religion” of “our cultural milieu” so any journalist who disagrees with the establishment can be smeared with the label “communist” and forced “on the defensive”. Most “have fully internalized the religion anyway, but they are all under great pressure to demonstrate their anti-communist credentials.”

These are interesting ideas, and if supported and developed further they would go a long way towards explaining how the media might have a strong conservative bias despite the liberal leanings of most journalists.

But just after proposing them, the book makes a sudden ninety degree turn to focus on a series of in-depth case studies of US military interference in Third World countries.

The case studies are there for a reason: after Chomsky and Herman establish what they consider to be the true story, they provide examples of the US media consistently misrepresenting even the simplest of facts in ways that flatter the United States government and unfairly malign its foreign enemies. These result in the US getting away with what can only be described as genocide with almost no criticism, even though the facts are plain for anyone to see.

So the idea of media bias hasn’t been exactly *dropped*. But these studies have disappointingly little relevance to the more general claims that I and presumably most people who bought this book were interested in. Military interference in Third World countries is a very specific subject, and one whose dynamics differ from stories closer to home.

Was I disappointed that the authors didn’t develop their original point about the media more? I was at first. Then I realized this was the book about obscure brutal Third World military conflicts that I’d never known I needed.

II.

Chomsky and Herman are both academics, and they’re both *relentless*. When they try to prove something, by golly, it *stays* proved. This is a good thing, in that the book deals with very controversial topics and anything less would be unconvincing. It’s also a bad thing, in that by the ninth or tenth long transcript taken from the same war crimes trial, all of the genocides and village-burnings and nun-rapes start to blend together into a big blob of atrocity, and you can’t remember whether Kouprasith Abhay was the evil generalissimo who launched the pro-US coup and killed thousands, or the good generalissimo who launched the anti-US counter-coup and killed thousands, or the morally ambiguous generalissimo who launched the non-aligned counter-counter-coup and killed thousands.

(his Wikipedia page clarifies that “[his] counter-coup within the counter-coup was ended by the paratroopers responsible for the ongoing coup.”)

But these details are less interesting than the big picture, a sketch of a political system that C&H jokingly term “death squad democracy”.

The general picture is of a third world country that was previously in a fragile social equilibrium. Something disrupts the equilibrium – usually the United States toppling the government because Communists were starting to do well in elections. It is replaced by a weak central government insecure in its power which decides to go after mass movements it perceives as a threat.

The mass movements form guerilla groups to resist government brutality. Supporters of the government form death squads in order to kill suspected guerillas more unethically than the international community would allow the government to do directly. Eventually there is so much violence that anyone who can form a guerilla army and kill their enemies before their enemies kill them does so.

The dictator solemnly declares that what’s going on is a rebellion by communist extremists with associated counter-violence by some grassroots rightist extremists, while he, the dictator, is doing his best to keep the peace. He send in the army, who are secretly or not-so-secretly are also the death squads, and so just make things worse. The United States declares the dictator is a great man who does his best to maintain peace in a troubled nation, and sends him tons of weapons and money. All of these weapons and money mysteriously end up in the hands of the death squads, which of course means the United States has to send in *more* weapons and money to help the dictator deal with the new threat of these richer, better-armed enemies.

If the dictator is feeling really nice, he will hold an election. The mass movements, communists, and anyone with actual popular support will be banned from participating since they are violent extremists, and the death squads will kill anybody who campaigns against the dictator. The dictator will win the vote handily, and the Free World will declare that since he won the elections, it’s clear that the communists are just violent extremists trying to deny the will of the people and take over for their own nefarious purposes.

This pattern, with slight variation, seems to have happened across the entire Third World at one point or another. Perhaps there will be another coup, and the dictator will be replaced by another dictator, perhaps some foreign country will get directly involved on one side or the other, but the basic logic will not change. For a space of years to decades, tens of thousands of people will be tortured and killed – a few here and there by the communists, but most by the government. Whole villages will be destroyed, freedom of thought will be nonexistent, and everyone except the dictator and a few cronies will be constantly living in fear.

And in a sense, I already knew all of this. We all kind of understand what goes on in banana republics. But for some reason, *Manufacturing Consent* painted an unusually clear picture that knocked it into relief for me and changed my understanding of a lot of things.

Take, for instance, the second Iraq War. The hawkish position is “we were right to want to remove Saddam, a bad man. We were right to believe that we would win the shooting war quickly and easily. We just couldn’t have predicted the explosion of Sunni-Shiite violence that would erupt afterwards, and that’s not our fault.”

And yet now that I have read *Manufacturing Consent*, it seems *obvious* that removing Saddam would cause Iraq to descend into blood-soaked death squads. It is like a law of the universe that Third World countries will descend into blood-soaked death squads at the drop of a pin. Every time the United States has tried to change the government of a Third World nation, the end result has been blood-soaked death squads. Expecting to remove a regime from power without thinking about the blood-soaked death squads seems less like an excusable error and more like missing the very heart of the issue, like expecting to use a nuke without thinking about radiation damage.

But the dove position is almost as bad! It’s “Ha! The hawks thought we would be greeted as liberators! What morons!” This totally misses the point! It’s assuming that if the Iraqis *liked* us, they would have politely lined up to form a centralized democratic government with a monopoly on the use of force. The problem wasn’t that the Iraqis didn’t like us enough, it was that we did something in a Third World country and expected it not to descend into blood-soaked death squads. *That never works*.

I am left with a greatly increased respect for the view that it was Western colonialism, broadly defined, that has caused Third World countries all their grief. The problem wasn’t just British people coming in and telling them to work on banana plantations for a while, the problem was the total destruction of the country’s usual rule of law, hierarchies, civic traditions, and social fabric by successive attempts by western-backed dictators to retain power. A couple of decades assassinating anyone who looks out of place and doesn’t do exactly what they’re told, of tearing apart any organization or community that looks strong enough to serve as an alternative to the State or offer resistance – the question is less why Third World countries are so screwed up, and more that they're not screwed up even worse.

III.

Throughout all of this, the US media could always be counted on to condemn the victims, excuse the aggressors, and totally fail to mention our role in anything.

As per Chomsky, this was rarely done by direct lies, in the form of front page "EVERYTHING FINE IN GUATEMALA, SAY SOURCES". It was done by a campaign of highlighting certain things, downplaying others, and creating false controversies to cover up the real ones. Their five case studies showcase five different common media biases.

The first study is titled "Worthy And Unworthy Victims", and compares news coverage of the "worthy victims" killed by America's enemies to that of the "unworthy victims" killed by America's allies. The death of worthy victims is treated with outrage, lurid descriptions of every detail of their brutal deaths, intense coverage of every new development in the hunt to bring the killers to justice, focus on the protests their death engenders, and insistence that their death proves a deep and important generalizable lesson about the society in which it occurred. The death of unworthy victims, if covered at all, is treated with "Well, violence sometimes happens, and it's very sad, but what can we do about it?" Their case study of a "worthy victim" is Jerzy Popieluszko, a Polish priest killed by the Communists; since the Communists were our enemy, we were outraged by the crime. Their examples of "unworthy victims" are the thousands killed in El Salvador and Guatemala, most notably Archbishop Oscar Romero; both countries' governments were US allies fighting against Communist guerrillas at the time, so their atrocities had to be covered up “for the good of the cause”. As a result, the American populace mostly ended up believing that our enemies were brutal murderers, and our allies were, at best, peace-loving people who were not very good at controlling the violence that always seemed to be breaking out around them.

The second study is “Legitimizing Versus Meaningless Elections”. Most Third World elections are a little sketchy. If the election is in a US ally, it will be covered as a “step towards suffrage in this fledgling democracy”, but if the election is in a US enemy, it will be covered as “a sham” that people are only voting in “for fear of retribution”. The book discusses the elections in Communist Nicaragua versus US-backed El Salvador, showing that by any objective standards the former had fairer, freer elections yet were attacked as a sham by the US media; the latter basically *was* a sham intended to legitimize a dictatorship, but were praised as a good first step by US media. After reading this chapter it will be very hard for me to take reports of Third World elections seriously again.

The third study is the odd man out, farce in the midst of tragedy. It describes how gullibly the US media accepted the idea of a connection between would-be-Pope-assassin Mehmet Ali Agca and the KGB in the absence of any credible evidence. Yes, C&H admit, Agca did confess to working for the Communists – but only after Italian secret police demanded he do so. Plus he also confessed to lots of other things, including being Jesus Christ, and it was kind of clear that he was a little crazy. In terms of non-psychotic, non-Pope-murdering people who had evidence that the Communists were involved, there was pretty much zilch. But because the Soviets were The Enemy, the media was willing to uncritically pass along anything that discredited them.

The fourth study deals with the Vietnam War, usually considered a case of the media breaking with the establishment and taking a more pacifist, leftist position. C&H argue that this was true only within a very narrow Overton window, where the two acceptable positions were “the US is right to fight for the freedom of South Vietnam” versus “the US is right to fight for the freedom of South Vietnam, but the costs are too high”. C&H argue that nearly everyone in South Vietnam supported Ho Chi Minh except for the dictator and his cronies. The US intervened to save the dictator from his own people, but cast this as saving South Vietnam from North Vietnamese aggression, even though North Vietnam’s involvement was modest. A more honest account of the US role was that they were coming from thousands of miles away to save South Vietnam from “aggression” by the South Vietnamese people. Absent any real enemy except the populace itself, they were backed into a strategy of burning down villages and killing indiscriminately, hoping to keep everyone in such a state of constant terror that they couldn’t do any political organizing. The US media never came close to expressing this position, and therefore at best they could be described as “pro-establishment” and “pro-establishment but sick of losing.”

The fifth study was much like the fourth study, except with Laos and Cambodia. The United States killed about 50,000 people in Laos directly through bombing, and probably more through its consistent support for whichever colonel was launching a coup that day. The US media was completely silent, even though there was ample evidence that it was going on and the foreign media was all over it. Also, when the US media finally got around to talking about it, it was in the context of the supposed “Ho Chi Minh Trail”, whereas most of the bombing was just bombing poor villages in order to deprive the Laotian communists of their natural rural base.

Overall, C&H did a good job of showing ways that the US media could systematically distort foreign wars to cover up the atrocities of US allies, highlight the atrocities of US enemies, and make US actions seem much more noble than the generally chilling evidence would suggest.

IV.

So, do I believe any of it?

C&H are, as mentioned before, really thorough, and they cite everything back and forth twenty ways to Tuesday. But there are ways to be rigorous and dishonest at the same time. C&H had complete control of what incidents to include in their book, and that gives them a lot of power to choose genuinely troubling incidents while not acknowledging any that don't fit their narrative.

For example, I mentioned before the case of Jerzy Popieluzsko, Polish priest murdered by the Communists. C&H make a big deal on how the US media was saturated with coverage and calls for justice; while they ignored the Salvadorean genocide victims around the same time.

But I notice that the Communists killed about a hundred million people over the course of the twentieth century. Most of these victims did not get the same coverage as Popieluzsko; in fact, we've discussed before here how in most cases the media erred on the side of covering these up. Instead of "the media over-covers Communist murders", it might be "there is wide variance in the media's coverage of Communist murders, and C&H focused on the most overdone one in order to support their thesis."

I see this in a lot of places. C&H give a table of various genocides and the news coverage allotted to each. They find that, for example, the news coverage allotted the Kurdish genocide by Iraq (US enemy) was four times greater than the coverage allotted the East Timor genocide by Indonesia (US ally). On the other hand, if they had included Israel in the table, the lesson would have reversed; we hear *far* more about what Israel (US ally) is doing to the Palestinians than about the Kurds *or* East Timorese, even though the latter two cases involved far more deaths. Or what if they had included Iran (US enemy)? How many people know about the Iran-PJAK conflict that has claimed almost a thousand lives in the past few years? It's easy for C&H to cherry-pick examples of well-covered-US-enemies and poorly-covered-US-allies, but it's not clear that reflects reality very well.

Finally, I'm not sure how much to trust their history. I know very little about the mid-20th century; C&H might be presenting a very one-sided view. The few things I double-checked seem to support this analysis. For example, here's how they describe Laos in the early 1950s:

> A coalition government was established in 1958 after the only elections worthy of the name in the history of Laos. Despite extensive US efforts, they were won handily by the left. Nine of the thirteen candidates of the [communist] Pathet Lao guerrillas won seats in the national assembly, along with four candidates of the left-leaning neutralists ("fellow traveler," as they were called by Ambassador Parsons). Thus "Communists or fellow travelers" won thirteen of the twenty-one seats contested. The largest vote went to the leader of the Pathet Lao, Prince Souphanouvong, who was elected chairman of the national assembly.

> US pressures- including, crucially, the withdrawal of aid – quickly led to the overthrow of the government in a coup by a “pro-Western neutralist” who pledged his allegiance to “the free world” and declared his intention to disband the political party of the Pathet Lao (Neo Lao Hak Sat), scrapping the agreements that had successfully established the coalition. He was overthrown in turn by the CIA favorite, the ultra-right-wing General Phoumi Nosavan. After US clients won the 1960 elections, rigged so crudely that even the most pro-US observers were appalled, civil war broke out, with the USSR and China backing a coalition extending over virtually the entire political spectrum apart from the extreme right, which was backed by the United States.

This seemed so over-the-top cartoonishly evil that I had to check Wikipedia to see if it was an accurate summary. Here’s how they put the same events (editing very liberally for conciseness):

> In April, 1953, the Viet Minh’s People’s Army of Vietnam invaded the northeastern part of what was still the French Protectorate of Laos with 40,000 troops commanded by General Vo Nguyen Giap; including 2,000 Pathet Lao soldiers led by Souphanouvong. The objective of the two-pronged invasion was the capture of the royal capital of Luang Prabang and of the Plain of Jars. In November 9 the Pathet Lao began its conflict with the Kingdom of Laos thus beginning the civil war and technically the Second Indochina War while the First Indochina War was still going.
>
> The North Vietnamese invaders succeeded in conquering the border provinces of Phong Saly and Xam Neua, which were adjacent to northern Vietnam and on

the northeastern verge of the Plain of Jars. They then moved aside to allow the Pathet Lao force with its mismatched scrounged equipment to occupy the captured ground, and Souphanouvong moved the Pathet Lao headquarters into Xam Neua on 19 April.

On 21 March 1956, Souvanna Phouma began his second term as prime minister. He opened a dialogue with his brother, Souphanouvong. In August, they announced the intention of declaring a ceasefire and reintegrating the Pathet Lao and their occupied territory into the government. However, the Pathet Lao claimed the right to administer the provinces they occupied.

At the same time, they and their North Vietnamese backers ran a massive recruitment campaign, with the aim of forming nine battalions of troops. Many of the new recruits were sent into North Vietnam for schooling and training. This led to United States concern that the Royal Lao Army would be inadequately equipped and trained.

In November, 1957, a coalition government incorporating the Pathet Lao was finally established. Using the slogan, “one vote to the right, one vote to the left to prevent civil war,” pro-communist parties received one-third of the popular vote and won 13 of 21 contested seats in the elections of 4 May 1958. With these additional seats, the left controlled a total of 16 seats in the 59 member National Assembly. Combined with independents, this was enough to deny Souvanna’s center right, neutralist coalition the two-thirds majority it needed to form a government. With parliament deadlocked, the U.S. suspended aid in June to force a devaluation of the overpriced currency, which was leading to the abuse of U.S. aid. The National Assembly responded by confirming a right-wing government led by Phuy Xananikôn in August. This government included four members of the U.S.-backed Committee for the Defence of the National Interest (none of them National Assembly members). Three more unelected CDNI members were added in December, when Phuy received emergency powers to govern without the National Assembly.

Under orders from Souphanouvong, the Pathet Lao battalions refused to be integrated into the Royal Lao Army. Souphanouvong was then arrested and imprisoned, along with his aides. The two Pathet Lao battalions, one after the other, escaped during the night with no shots fired, taking their equipment, families, and domestic animals with them. On 23 May, Souphanouvong and his companions also escaped unscathed.

On 28 July, Communist Vietnamese units attacked all along the North Vietnamese-Lao border. As they took ground from the Royal Lao Army, they moved in Pathet Lao as occupation troops. Poor battle performance by the RLA seemed to verify the need for further training; the RLA outnumbered the attackers, but still gave ground.

On 9 August 1960, Captain Kong Le and his Special Forces-trained Neutralist paratroop battalion were able to seize control of the administrative capital of Vientiane in a virtually bloodless coup, while Prime Minister Tiao Samsanith, government officials, and military leaders met in the royal capital, Luang Prabang. His stated aim for the coup was an end to fighting in Laos, the end of foreign interference in his country, an end to the consequent corruption caused by foreign aid, and better

> treatment for his soldiers. However, Kong Le’s coup did not end opposition to him, and there was a scramble among unit commanders to choose up sides. If one was not pro-coup, then he had the further decision to make as to whom he would back to counter the coup. The front runner was General Phoumi Nosavan, first cousins with the prime minister of Thailand, Field Marshal Sarit Thanarat. With the Central Intelligence Agency’s support, Sarit set up a covert Thai military advisory group, called Kaw Taw. Kaw Taw, which would support the counter-coup that was mounted; it supplied artillery, artillerymen, and advisers to Phoumi’s forces. It also committed the CIA-sponsored Thai Police Aerial Reinforcement Unit to operations within Laos.

So, things that C&H conveniently forgot to mention: North Vietnam invaded Laos (!), and the Communists gained their power as lackeys for these foreign invaders (!). Although the Communists did well in the 1958 elections, they absolutely did not have a majority in government at the time, and in fact stonewalled the legitimate government. Xananikôn was elected constitutionally by the National Assembly, including the Communists. The Communists refused to stand down their armies and join the national government, and when the government tried to make them, North Vietnam invaded again, with the Communists supporting the foreign invaders. It was in this context that the Neutralists launched their coup, and Phoumi’s CIA-backed countercoup was actually in opposition to it. This is a *really* different story than C&H’s version. C&H never lie per se, but they leave out things as significant as a giant foreign invasion happening during the middle of the events they’re describing.

Here’s something else I found on Wikipedia: both Chomsky and Herman are considered prominent Cambodian genocide denialists:

Beginning with “Distortions at Fourth Hand”, an article published in the American left-wing periodical The Nation in June 1977, they wrote that while they did not “pretend to know [...] the truth” about what was going on in Cambodia during the Khmer Rouge regime of Pol Pot, while reviewing material on the topic then available, “[w]hat filters through to the American public is a seriously distorted version of the evidence available”. Referring to “the extreme unreliability of refugee reports,” they noted: “Refugees are frightened and defenseless, at the mercy of alien forces. They naturally tend to report what they believe their interlocutors wish to hear. While these reports must be considered seriously, care and caution are necessary. Specifically, refugees questioned by Westerners or Thais have a vested interest in reporting atrocities on the part of Cambodian revolutionaries, an obvious fact that no serious reporter will fail to take into account.” They concluded by stating that Khmer Rouge Cambodia might be more closely comparable to “France after liberation, where many thousands of people were massacred within a few months” than to Nazi Germany.

Their book After the Cataclysm (1979), which appeared after the regime had been deposed, has been described by area specialist Sophal Ear as “one of the most supportive books of the Khmer revolution” in which they “perform what amounts to a defense of the Khmer Rouge cloaked in an attack on the media”.[9] In the book, Chomsky and Herman acknowledged that “The record of atrocities in Cambodia is substantial and often gruesome,” but questioned their scale, which may have been inflated “by a factor of 100”. They further asserted that the evacuation of Phnom Penh “may actually have saved many lives,” Khmer Rouge agricultural policies reportedly produced “spectacular” results, and there

> might have been “a significant degree of peasant support for the Khmer Rouge”: “How can it be that a population so oppressed by a handful of fanatics does not rise up and overthrow them?”
>
> Herman replied to critics in 2001: “Chomsky and I found that the very asking of questions about the numerous fabrications, ideological role, and absence of any beneficial effects for the victims in the anti-Khmer Rouge propaganda campaign of 1975–1979 was unacceptable, and was treated almost without exception as ‘apologetics for Pol Pot’.”
>
> Many other scholars denying or doubting the character of the Khmer Rouge recanted their earlier opinions as the evidence of massive KR crimes against humanity mounted.

They touch on this issue in the book, but I have trouble figuring out what to make of it. Certainly they are outraged that anyone accuses them of denying the Cambodian genocide, and they say this is evil right-wing character assassination propaganda. They then go on to say, kind of flailingly, that also the Cambodian genocide wasn’t that bad, that all the media reports about it were lies, that it was the US’ fault anyway, that the US did worse things anyway, that Cambodia before the genocide was even worse, that America secretly loved Pol Pot and was his best friend, and also shut up shut up shut up. As far as I can get any kind of coherent thesis at all out of this, they seem to be saying they were Gettier cased; every media report of the genocide was a vile right-wing propaganda lie, but coincidentally, a genocide exactly like the one reported in the media occurred.

Herman is additionally criticized for denying the Rwandan and Srebrenica genocides, although Chomsky does not seem to be involved.

And usually I hate terms like “genocide apologist”, because very few people are actually genocide apologists so it’s usually a call to outrage aimed at riling up an angry mob against someone based on one comment they may or may not have said a long time ago.

But in the case where the entire point at issue is a book about genocide scholarship, where the thesis is “everybody else got these genocides wrong, and we are going to tell you the truth about them”, it becomes pretty important if they have a long history of getting genocides wrong.

So I take this book with a grain of salt. I think it treats the topics it covers very rigorously, but (ironically given the subject) the authors’ ability to set the agenda and choose which topics to focus on and which to omit gives them way too much power to shape the readers’ understanding of complex issues.

Do I blame C&H for this? Not *exactly*. As someone who’s occasionally engaged in some consensus-challenging myself, let me tell you, it’s *really hard*. Try being perfectly balanced, going out of your way to explain all the facts that disagree with your thesis and pointing out all the grey areas – and no one will listen to you at all. Because if people have heard all their life that A is pure good and B is total evil, and you hand them some dense list of facts suggesting that in some complicated way their picture might be off, they’ll round it off to “A is nearly pure good and B is nearly pure evil, but our wise leaders probably got carried away by their enthusiasm and exaggerated a bit, so it’s good that we have some eggheads to worry about all these technical issues.”

The only way to convey a real feeling for how thoroughly they’ve been duped is to present the opposite narrative – the one saying that A is total evil and B is pure good – then let the two narratives collide and see what happens.

And this is really hard, because the same institutions who swallow the utterly bankrupt mainstream narrative whole will suddenly rediscover their skepticism and pick apart every little exaggeration and omission in the contrary narrative. This is the domain of isolated demands for rigor; suddenly no objection is too vague or philosophical, and any amount of emotion or editorializing represents a “bias” that discredits the entire work. So countercultural elements are caught between a rock and a hard place: if they stick to a minimalist stating of the most agreed-upon facts, then it’s not enough to shock people out of their prejudices; any attempt to spin a convincing narrative in the way their mainstream opponents do all the time, and they get attacked for going beyond what can be 100% incontrovertibly defended.

I think C&H handle this impossible balancing act better than most. I think *Manufacturing Consent* has serious issues with bias, sometimes inexcusably so, but I think its thesis survives these biases. I went into this book with more or less the attitude mentioned above: the classic story of America being great was a bit exaggerated and overenthusiastic, and in fact we did a lot of morally ambiguous things.

I came out of it with more of a primal horror that we spent a lot of the 20th century being moral monsters, and feeling like we have the same sort of indelible black mark on our name as Germany or Russia or Belgium. Whatever factors C&H may have exaggerated, and whatever exculpatory evidence they may have omitted, I doubt that any of it would fully reverse that unpleasant conclusion.

V.

Okay, but what about media bias? Wasn't that the whole reason we got into this mess?

C&H's case studies of foreign wars aren't great tests for their hypothesized mechanisms of bias. Their first two mechanisms are big media corporations pushing a pro-corporate worldview, and big corporate advertisers insisting on programming that reflects well on them and their corporate activities.

And I can see why a mass media dominated by corporate giants might be expected to agitate against labor unions, but it's harder to see why it is so insistent on covering up a campaign of genocide by pro-American forces in El Salvador. It's easy to see why they might avoid condemning oil companies in order to preserve ad revenue from Texaco, but harder to see why they would systematically underestimate casualties from US bombing missions on the Plain of Jars in Laos.

Their third mechanism, big Pentagon-style sources with press bureaus, certainly applies very well to these cases. But it doesn't seem like it should necessarily generalize to every other type of story. When the media is covering an election, or a protest, where is the Pentagon-style source? Although C&H's point that the police department, etc, can also be sources in this way is well-taken, this seems less pressing for a protest in Seattle than for, say, a bombing campaign in Laos, where a news source might have trouble getting Lao-speaking journalists into the midst of the carnage. Besides, what about cases where this produces the opposite bias? Might newspapers be overly friendly to regulations because they rely upon the regulatory body? What if there is a protest by a large, well-organized group that has cultivated links with the press?

Their fourth mechanism, flak machines, raise a similar issue. C&H view this as a rightist phenomenon almost by definition. They never consider the possibility that, for example, their writing an entire book saying the media is dishonest and biased might count as “flak” on their part. Any conservative criticizing the media is part of a “flak machine” intended to “keep it under control” and “destroy its independence”, but any leftist criticizing the media is bravely trying to expose its biases and bring the truth to light. This seems so obvious to them that they never even have to justify it. This is perhaps understandable in the conflict of foreign wars, where it’s more likely that would-be patriots will condemn reporting that reflects poorly on American troops, but in the context of domestic policy it doesn’t make a lot of sense.

That leaves their fifth mechanism, “anti-communism as the dominant religion of our culture”, a claim which hasn’t aged well since *Manufacturing Consent* came out in the ’80s. Worse, C&H’s argument for this position is almost word-for-word the same argument that conservatives use to claim that “anti-racism is the dominant religion of our culture”. I’ve even heard them use the specific phrase “dominant religion”.

In their section on “worthy victims” versus “unworthy victims”, C&H describe a certain form of coverage the media reserves for the victims of Communism (section edited for length and clarity):

A. Fullness and reiteration of the details of the murder and the damage inflicted on the victim. The coverage of the Popieluszko murder was notable for the fullness of the details regarding his treatment by the police and the condition of the recovered body. What is more, these details were repeated at every opportunity. The condition of the body was described at its recovery, at the trial when the medical evidence was presented, and during the testimony of the perpetrators of the crime. At the trial, the emotional strain and guilt manifested by the police officers were described time and again, interspersed with the description of how Popieluszko pleaded for his life, and evidence of the brutality of the act…Popieluszko himself was humanized, with descriptions of his physical characteristics and personality that made him into something more than a distant victim.” In sum, the act of violence and its effects on Popieluszko were presented in such a way as to generate the maximum emotional impact on readers. The act was vicious and deserved the presentation it received. The acts against the unworthy victims [of US anti-Communist client states] were also vicious, but they were treated very differently.

B. Stress on indignation, shock, and demands for justice. In a large proportion of the articles on the Popieluszko murder there are quotations or assertions of outrage, indignation, profound shock, and mourning, and demands that justice be done. Steady and wholly sympathetic attention is given to demonstrators, mourners, weeping people, work stoppages, masses held in honor of the victim, and expressions of outrage, mainly by nonofficial sources. The population “continues to mourn,” “public outrage mounted,” the pope is deeply shaken, and even Jaruzelski condemns the

> action. The net effect of this day-in-day-out repetition of outrage and indignation was to call very forcible attention to a terrible injustice, to put the Polish government on the defensive, and, probably, to contribute to remedial action.
>
> **C. The search for responsibility at the top.** In article after article, the U.S. media raised the question: how high up was the act known and approved? By our count, eighteen articles in the New York Times stressed the question of higher responsibility, often with aggressive headlines addressed to that point…
>
> **D. Conclusions and follow-up.** The New York Times had three editorials on the Popieluszko case. In each it focused on the responsibility of the higher authorities and the fact that “A police state is especially responsible for the actions of its police” (“Murderous Poland:’ Oct. 30, 1984). It freely applied words like “thuggery,” “shameless,” and “crude” to the Polish state. The fact that police officers were quickly identified, tried, and convicted it attributed to the agitation at home and abroad that put a limit on villainy. This is a good point, and one that we stress throughout this book: villainy may be constrained by intense publicity. But we also stress the corresponding importance of a refusal to publicize and the leeway this gives murderous clients under the protection of the United States and its media, where the impact of publicity would be far greater.

But of course, that describes to a “t” the media’s coverage of the Ferguson shooting. C&H include a table showing the disproportionate attention given victims of Communism compared to all other types of victims, but the amount of attention given to Ferguson blows all of the Communist murders *off the chart*.

Does that mean that white policemen fill the same role today that the Soviet Union did back in the 80s? I don’t know. Sure, it’s relevant white policeman killed hundreds of people before Mike Brown with nary a peep from the media. But then, it’s also relevant that Communists killed millions of people before Jerzy Popieluzsko with equally minimal response.

My point is that “anti-Communism” is probably not a *uniquely* religious belief, and that these “religions” can serve the left as well as the right.

So none of C&H’s five pillars of conservative media domination really seem to stand up very well, which is fine because in their conclusions section C&H switch to a different theory.

They say that the media is a profit-seeking free market, and the best way to get profits is to appeal to advertisers. And the best way to appeal to advertisers is to appeal to the population. And the population wants to hear things that tell them they are good, and their country is good, and don’t challenge or dismay them overly much. Hearing that your government just killed 50,000 Lao civilians is a real downer; hearing that the war on those nasty Commies is going well will keep viewers coming back for more.

But this represents a retreat from the book’s thesis. The media is not exactly a propaganda organ that manipulates the people to serve powerful interests. It’s a tool of the people, giving them what they want to hear – which turns out to be terrible.

And then comes the obvious question – “But, like, fifty percent of the population are liberal, right? Don’t they also get told what they want to hear?”

C&H answer this with the one story that really hammered home the book’s thesis for me: what about Watergate? The media did a great job exposing the lies and corruption of those in power; in fact, of a Republican in power. Does that disprove C&H’s thesis?

No:

> The major scandal of Watergate as portrayed in the mainstream press was that the Nixon administration sent a collection of petty criminals to break into the Democratic party headquarters, for reasons that remain obscure. The Democratic party represents powerful domestic interests, solidly based in the business community. Nixon’s actions were therefore a scandal. The Socialist Workers party, a legal political party, represents no powerful interests. Therefore, there was no scandal when it was revealed, just as passions over Watergate reached their zenith, that the FBI had been disrupting its activities by illegal break-ins and other measures for a decade, a violation of democratic principle far more extensive and serious than anything charged during the Watergate hearings.
>
> History has been kind enough to contrive for us a “controlled experiment” to determine just what was at stake during the Watergate period, when the confrontational stance of the media reached its peak. The answer is clear and precise: powerful groups are capable of defending themselves, not surprisingly; and by media standards, it is a scandal when their position and rights are threatened. By contrast, as long as

> illegalities and violations of democratic substance are confined to marginal groups or distant victims of U.S. military attack, or result in a diffused cost imposed on the general population, media opposition is muted or absent altogether.) This is why Nixon could go so far, lulled into a false sense of security precisely because the watchdog only barked when he began to threaten the privileged.

So for C&H, the media's rightward bias isn't "pro-Republican, anti-Democrat". It's pro- a conservative establishment in which both Republicans and Democrats collude, and anti- the *real* left, which it treats as a lunatic fringe too powerless to even be worth mentioning.

This is a new theory, quite different from the five points about corporatism that started the book, and it seems to resolve the paradox of both right and left seeing media bias. The media enforces conformity with the Overton window against both the right and left flanks. Both the rightward and leftward fringes notice the same set of dirty tricks in the media, and describe them in almost exactly the same terms. Thus both sides complain about the other being a "dominant religion", both sides complain that both major parties are part of the same con, both sides complain that the media restricts debate to a narrow range of acceptable opinion, etc.

And both sides are shouted down in the same terms, too. When the far right complains about the media, academia, and bureaucracy being ranged against them, they get called conspiracy theorists. I myself somewhat hastily made this claim in section 3.2 of my Anti-Reactionary FAQ. More recently, Topher Hallquist makes a similar claim, classily adding that any communities that even dare to associate with people who believe this ought to suffer guilt by association.

Chomsky and Herman are aware of this attack, and begin by saying:

> Institutional critiques such as we present in this book are commonly dismissed by establishment commentators as “conspiracy theories”, but this is merely an evasion. We do not use any kind of “conspiracy” hypothesis to explain mass-media performance. In fact, our treatment is much closer to a “free market” analysis, with the results largely an outcome of the workings of market forces. Most biased choices in media arise form the preselection of right-thinking people, internalized preconceptions, and the adaptation of personnel to the constraints of ownership, organization, market, and political power.

And later:

> As we have stressed throughout this book, the U.S. media do not function in the manner of the propaganda system of a totalitarian state. Rather, they permit-indeed, encourage-spirited debate, criticism, and dissent, as long as these remain faithfully within the system of presuppositions and principles that constitute an elite consensus, a system so powerful as to be internalized largely without awareness.

I find many smart people, both on the right and the left, say something similar about this same self-organizing consensus enforcement system. Their disagreements about its position seem to be entirely matters of perspective; to a Mexican, America is a northern nation; to a Canadian, it’s a southern one. But despite this substantial agreement and the rivers of ink spilled on the matter, they always describe it in the vaguest of terms, in a style ranging somewhere between “non-technical” and “paranoid”.

If we want to understand politics, I feel like one of the most important subgoals is to figure out the precise ways in which these sorts of alignments arise – in other words, how class warfare solves its coordination problems without most of the people involved being aware of what they’re doing or holding any explicitly sinister thoughts.

I don’t think *Manufacturing Consent* does much to solve this problem and explain the real nature of the system. But it certainly illuminates one otherwise-easily-neglected corner of it, and offers a window on some of its tricks and on some of the sins it has to answer for.

Cardiologists and Chinese Robbers

I.

It takes a special sort of person to be a cardiologist. This is not always a good thing.

You may have read about one or another of the "cardiologist caught falsifying test results and performing dangerous unnecessary surgeries to make more money" stories, but you might not have realized just how common it really is. Maryland cardiologist performs over 500 dangerous unnecessary surgeries to make money. Unrelated Maryland cardiologist performs another 25 in a separate incident. California cardiologist does "several hundred" dangerous unnecessary surgeries and gets raided by the FBI. Philadelphia cardiologist, same. North Carolina cardiologist, same. 11 Kentucky cardiologists, same. Actually just a couple of miles from my own hospital, a Michigan cardiologist was found to have done $4 million worth of the same. Etc, etc, etc.

My point is not just about the number of cardiologists who perform dangerous unnecessary surgeries for a quick buck. It's not even just about the cardiology insurance fraud, cardiology kickback schemes, or cardiology research data falsification conspiracies. That could all just be attributed to some distorted incentives in cardiology as a field. My point is that it takes a special sort of *person* to be a cardiologist.

Consider the sexual harassment. Head of Yale cardiology department fired for sexual harassment with "rampant bullying". Stanford cardiologist charged with sexually harassing students. Baltimore cardiologist found guilty of sexual harassment. LA cardiologist fined $200,000 for groping med tech. Three different Pennsylvania cardiologists sexually harassing the same woman. Arizona cardiologist suspended on 19 (!) different counts of sexual abuse. One of the "world's leading cardiologists" fired for sending pictures of his genitals to a female friend. New York cardiologist in trouble for refusing to pay his $135,000 bill at a strip club. Manhattan cardiologist taking naked pictures of patients, then using them to sexually abuse employees. New York cardiologist secretly installs spycam in office bathroom. Just to shake things up, a Florida cardiologist was falsely accused of sexual harassment as part of feud with *another* cardiologist.

And yeah, you can argue that if you put high-status men in an office with a lot of subordinates, sexual harassment will be depressingly common just as a result of the environment. But there's also the Texas cardiologist who pled guilty to child molestation. The California cardiologist who killed a two-year-old kid. The author of one of the world's top cardiology textbooks arrested on charges Wikipedia describes only as "related to child pornography and cocaine".

Then it gets weird. Did you about the Australian cardiologist who is fighting against extradition to Uganda, where he is accused of "terrorism, aggravated robbery and murdering seven people"? What about the Long Island cardiologist who hired a hitman to kill a rival cardiologist, and who was also for some reason looking for "enough explosives to blow up a building"?

Like I said, it takes a *special* sort of person.

II.

Given the recent discussion of media bias here, I wanted to bring up Alyssa Vance's "Chinese robber fallacy", which she describes as:

> ..where you use a generic problem to attack a specific person or group, even though other groups have the problem just as much (or even more so).

For example, if you don't like Chinese people, you can find some story of a Chinese person robbing someone, and claim that means there's a big social problem with Chinese people being robbers.

I originally didn't find this too interesting. It sounds like the same idea as plain old stereotyping, something we think about often and are carefully warned to avoid.

But after re-reading the post, I think the argument is more complex. There are over a billion Chinese people. If even one in a thousand is a robber, you can provide one million examples of Chinese robbers to appease the doubters. Most people think of stereotyping as "Here's one example I heard of where the out-group does something bad," and then you correct it with "But we can't generalize about an entire group just from one example!" It's less obvious that you may be able to provide literally one million examples of your false stereotype and still have it be a false stereotype. If you spend twelve hours a day on the task and can describe one crime every ten seconds, you can spend four months doing nothing but providing examples of burglarous Chinese – and still have absolutely no point.

If we're really concerned about media bias, we need to think about Chinese Robber Fallacy as one of the media's strongest weapons. There are lots of people – 300 million in America alone. No matter what point the media wants to make, there will be hundreds of salient examples. No matter how low-probability their outcome of interest is, they will never have to stop covering it if they don’t want to.

This has briefly gotten some coverage in the form of “the war on police”. As per AEI:

> Is there a “war on police” in America today? Most Americans think so, and that’s understandable given all of the media coverage of that topic. A Google news search finds 32,000 results for the phrase “war on cops” and another 12,100 results for “war on police,” with sensational headlines like “America’s War on Cops Intensifies” and “Bratton Warns of Tough Times Ahead Due to ‘War on Cops’.” A recent Rasmussen poll found that 58% of likely US voters answered “Yes” to the question “Is there a war on police in America today?” and only 27% disagreed. But data on police shootings in America that were reported last week by The Guardian tell a much different story of increasing police safety.
>
> According to data available from the “Officer Down Memorial Page” on the annual number of non-accidental, firearm-related police fatalities, 2015 is on track to be the safest year for law enforcement in the US since 1887 (except for a slightly safer year in 2013), more than 125 years ago. And adjusted for the country’s growing population, the years 2013 and 2015 will be the two safest years for police in US history, measured by the annual number of firearm-related police fatalities per 1 million people.

When politically convenient, it is easy to make Americans believe in a war on police simply by better coverage of existing murders of police officers. Given that America is a big country with very many police, even a low base rate will provide many lurid police-officer-murder stories – by my calculation, two murders a week even if officers are killed only at the same rate as everyone else. While covering these is a legitimate decision, it can be deceptive unless it's framed in terms of things like whether the rate has gone up or down, whether the rate is higher or lower for the group involved than the base rate in the population, and it still seems scary when you explicitly calculate the rate.

But a Chomskian analysis would ask whether the talk of a "war on cops" is really a uniquely bad example of journalistic malpractice, or whether it is bog-standard journalistic malpractice which is unique only in being called out this time instead of allowed to pass.

Let's stick with coverage of police for consistency's sake. I've made a very similar argument before regarding claims of racist police shootings (see Part D here), but let's avoid *that* particular rabbit hole and consider a broader and more unsettling point. We all hear anecdotes about terrible police brutality. Suppose, in fact, that we've heard exactly X stories. Given that there are about 100,000 police officers in the US, is X consistent with the problem being systemic and dire, or with the problem being relatively limited?

I mean, it's hard to say. Quick Fermi calculation: if I can think of about one horrible story of police brutality a week, and assume there are fifty that aren't covered for every one that is, then per year that makes…

But wait – what if I told you that number was a lie, and there were actually 500,000 police officers in the US? Suddenly the rate of police brutality has decreased five times from what it was a second ago. If you previously believed that there were 100,000 police officers, and that the police brutality rate was shameful but that decreasing the rate to only one-fifth its previous level would count as a victory, well, now you can declare victory.

What if I told you the 500,000 number is also a lie, and it's actually way more cops than that? Do you have any idea at all how many police there are? Shouldn't you at least have an order-of-magnitude estimate of what the police brutality rate is before deciding if it's too high or not? What if I told you the real number was a million cops? Five million cops? Ten million? That's a hundred times the original estimate of 100,000 – shouldn't learning that the police brutality rate is only 1% of what you originally estimated (or, going the other direction, 10,000% of that) change your opinion in some way?

(No, I won't tell you how many cops there actually are. Look it up.)

I feel this way about a lot of things. The media is always giving us stories of how tech nerds are sexist in some way or another. But we may suspect they want to push that line regardless of whether it's true. How many tech nerds are there? A million? Ten million? How many lurid stories about harassment in Silicon Valley have you heard? Do we know if this is higher or lower than the base rate for similar industries? Whether it's going up or down? What it would look like if we actually had access to the per person rates?

By now you've probably figured out the gimmick, but just to come totally clean – cardiologists are wonderful people who as far as I know are no less ethical than any other profession. I chose to pick on them at random – well, not quite random, one of them yelled at me the other day because apparently contacting the cardiologist on call late at night just because your patient is having a serious heart-related emergency is some kind of huge medical *faux pas*. I don’t think anyone has ever claimed that there’s any general issue with cardiologists, and as far as I know there’s no evidence for such.

If you read Part I of this post and found yourself nodding along, thinking “Wow, cardiologists are real creeps, there must be serious structural problems in the cardiology profession, something must be done about them,” consider it evidence that a sufficiently motivated individual – especially a journalist! – can make you feel that way about *any group*.

Beware Systemic Change

[Epistemic status/edit: After reading comments, no longer sure I agree with Part I. Part II still seems right but possibly a cost that can be outweighed by other factors. I continued to be worried about this without necessarily thinking it is a knock-down argument. I'm still not sure how to balance my support for some systemic change causes against my concern about others. Buck's comment on morality seems important.]

I.

One of the most common critiques of effective altruism is that it focuses too much on specific monetary interventions rather than fighting for "systemic change", usually billed as fighting inequitable laws or capitalism in general. For example, Amia Srinivasan, in her review of Doing Good Better.

> What's the expected marginal value of becoming an anti-capitalist revolutionary? To answer that you'd need to put a value and probability measure on achieving an unrecognisably different world – even, perhaps, on our becoming unrecognisably different sorts of people. It's hard enough to quantify the value of a philanthropic intervention: how would we go about quantifying the consequences of radically reorganising society? […]
>
> Effective altruism, so far at least, has been a conservative movement, calling us back to where we already are: the world as it is, our institutions as they are. MacAskill does not address the deep sources of global misery – international trade and finance, debt, nationalism, imperialism, racial and gender-based subordination, war, environmental degradation, corruption, exploitation of labour – or the forces that ensure its reproduction. Effective altruism doesn't try to understand how power works, except to better align

> itself with it. In this sense it leaves everything just as it is.

This same point has been made again and again and again. In response, many effective altruist leaders have gone on to say that they love systemic change and that the movement is entirely in favor of it.

I am not affiliated with the organized effective altruist movement and my opinion has no relation to theirs. They have spent a lot of work trying to convince everyone that they are entirely in favor of pursuing systemic change, I believe them, and nothing I say here reflects on that.

But I, personally, worry a lot about pursuing systemic change.

“Worry about” is not the same as “totally oppose”. This post’s Hansonian title is “Beware Systemic Change” rather than “Against Systemic Change.” But I’m pretty serious about bewaring of it.

First, what do I mean by “systemic change”? Traditional charity, like healing the sick, is almost universally viewed as good or at least neutral. Everyone agrees the sick should be healed; if there are unhealed sick people, it’s because we don’t have the resources to pursue our universally held goal. The same is true of feeding hungry children. It’s true of weird causes like AI risk – some people think it’s silly, but they’re happy to let other people work on them if those people want. It’s even true of things like cutting carbon emissions, sort of. When the Koch brothers say they oppose cutting carbon emissions, they mean they oppose laws mandating such cuts, or budgets that spend communal resources to enforce them. If a private donor offered to pay for scrubbers on every smokestack at zero cost to the rest of the economy, the Koch brothers would have no objection.

Some political issues are kind of like this. People from all over the political spectrum agree that corporate welfare is a bad idea; if we still have corporate welfare, it’s because there’s not enough attention and organization to force politicians to abandon it. In other cases, we all agree something is good but disagree on whether it is an optimal use of resources: for example, most people agree that aid agencies like UNICEF that help children abroad are doing good work, but not everyone agrees with funding them from the federal budget.

Other political issues are not like this. Some people believe that increasing the minimum wage is a laudable goal; other people believe it will hurt the economy or that it violates important moral rights. The reason we don’t have a higher minimum wage isn’t because passing laws costs a lot of money that no one has raised yet, or because no one is paying attention to the issue. It’s because a lot of people oppose it and so far those people are winning, or at least holding their own.

In terms of Freshman English Plot Devices, traditional charity like healing the sick is “man versus nature.” Political issues like the minimum wage are “man versus man”.

When I think of systemic change, I think of man versus man. Even if effective altruists helped governments increase their foreign aid budget, I don’t think Amia Srinivasan and Jacobin Magazine and the rest would think we were participating in “systemic change”. I think at the very *least* they’d want altruists out in the street demonstrating for higher minimum wages, and at most trying to eliminate global capitalism.

Which is a problem, because a lot of people *like* global capitalism. A dialogue:

Bob: Man vs. man conflicts raise some thorny issues which man vs. nature conflicts manage to avoid. I think we should be very wary about opening the door to political discussions.

Alice: What? Come on, Bob, you talk about politics all the time!

Bob: I watch porn all the time too; that doesn't mean I'm proud of it, or think it's the most good I could do with my resources. Suppose effective altruists get involved in the 2016 US presidential election – which isn't prima facie a bad idea; think about how easy it would have been to make Gore win in 2000 and how much would have changed if he had. Lots of people work very hard and raise $10 million for the Democrats. Lots of other people work very hard and raise $10 million for the Republicans. Now the Democrats and Republicans are at exactly the same position vis-a-vis each other as they were before the effective altruists got involved, but we have wasted $20 million that could have gone to healing the sick or feeding the hungry. And I'm using money to make things obvious, but the same goes for donating time or advocacy or other resources. If this sort of thing started happening, we would want to promote a general cultural norm of "never spend resources on man vs. man conflicts". If both sides were equally likely to follow the norm, then the conflicts would remain unaffected but everyone would have more resources to spend on the sick and hungry.

Alice: But effective altruists are very unlikely to donate to both sides of a political issues equally and cancel out. Political views are heavily shaped by demographics, and EAs are likely to skew left just like most other highly-educated groups. Even aside from this, their similar moral assumptions and thinking styles will lead them to converge onto the same side of an issue. Half of Americans are creationist, and almost as many oppose gay marriage, but I would expect fewer than 5% of EAs to hold either position. The idea of donations cancelling out is totally unrealistic. Instead, we should predict that on most issues, most EA donations will go to the same side. We end up not with a wasteful neutrality, but with a large sum of money going to one side, detracted from only slightly by a much smaller sum going to the other. If the cause is important enough, it might still be the most good we could do – the net benefit of (good from giving large amount of money to one side) minus (harm from giving small amount of money to the other side) would still be higher than the benefit of giving all the money to a traditional charity.

Bob: That just kicks the problem one meta-level up. Suppose that on each given issue, effective altruists converge dramatically around one or the other side. If half the time they converge around the right side, and the other half around the wrong side, then over a large number of issues their contributions will gradually even out and sum to zero.

Alice: That’s a ridiculous way of looking at it. We don’t just flip a coin to determine which side to back. We exhaustively study the argument for both sides, the evidence base, et cetera. Then we focus only on those issues where we can be most certain we’re in the right. The odds there are a heck of a lot better than fifty percent!

Bob: But you could make the same argument about picking stocks, couldn’t you? Do lots of research, focus on the ones where you’re most certain that they’re overvalued or undervalued, and then you have great odds of getting rich!

But of course, we know that doesn't work. Everyone else is trying the same thing, and the current position of the stock market reflects the consensus results of that process. You run afoul of the efficient market hypothesis.

Alice: Now you're just being silly. There's no efficient market hypothesis for politics!

Bob: But why shouldn't there be? A lot of people mock rationalists for thinking they can waltz into some complicated field, say "Okay, but we're going to do it *rationally*", and by that fact alone do better than everybody else who's been working on it for decades. It takes an impressive level of arrogance to answer "Why are your politics better than other people's politics?" with "Because we want to do good" or even with "Because we use evidence and try to get the right answer". You'd have to believe that other people *aren't even trying*.

Alice: What you're saying makes a certain kind of sense in Weird Platonic Spherical Cow Perfect Rationality Outside View World. But think about this from the Inside View perspective. Once again, half the country is creationist. Almost half oppose gay marriage. It's like a stock market where half of the investors are throwing everything they have into the perpetual motion industry. Surely you can admit that even a little bit of intelligence, education, and rationality can actually take you a long way in politics?

Bob: Half the country is creationist, but there's almost no easy gains from fighting them; any curriculum that federal politics can conceivably affect is evolutionary by this point, and it's unclear we get any real benefits by going after the last few Alabama middle school students. As for opposing gay marriage, I think you're going beyond your supposed reliance on evidence here. The strongest conservative case against gay marriage is that it reinforces a centuries-long redefinition of marriage from a strategic partnership focused on child-rearing to a ceremonial acknowledgment of romantic infatuation, potentially leading to a deep shift in the way people think about issues like who to marry, when to have kids, when to get divorced, and how to treat their family. That argument hasn't been rigorously evaluated by statisticians and found wanting. It's been found annoying and left untouched. Your differences are foundational assumptions and methodological disagreements about what sorts of issues to focus on, not simple "he made an arithmetic error when calculating the effects" style obvious superiority.

Alice: Really, Bob? You really want to go there?

Bob: Yes. In fact, I worry that this plays into exactly the potential flaws of the effective altruist movement. *I* can count up all the harms of banning gay marriage: exactly 1.13 million gay people regret not being able to marry, they rate their distress at 3.2/5 on the Likert Scale, that comes out to X QALYs lost per annum, but *you* have no way of easily quantifying the potential harms of gay marriage, therefore your argument is invalid. A lot of these issues involve trading off easily quantifiable harms on one side versus less quantifiable harms on the other: social trust, cultural cohesion, moral credibility, "freedom" broadly defined, ability to innovate. Highly educated people used to studying science might just be more likely to fall for the streetlight effect and go with the side that promises more quantifiability, rather than the side more likely to be right.

Alice: I…think you're being deliberately annoying? It seems like exactly the same kind of sophisticated devil's-advocate style argument we could use for anything. Sure, nothing is real and everything is permissible, now stop playing the

Steel Man Philosophy Game and tell me what you really think! It really should be beyond debate that some policies – and some voters- are just stupid. Global warming denialism? Mass incarceration? Banning GMOs? Opposing nuclear power? Not everything is a hard problem!

Bob: I really do sympathize with you here, of course. It’s hard not to. But I also look back at history and am deeply troubled by what I see. In the 1920s, nearly all the educated, intelligent, evidence-based, pro-science, future-oriented people agreed: the USSR was amazing. Shaw, Wells, Webb. They all thought Stalin was great and we needed a global communist revolution so we could be more like him. If you and I had been alive back then, we’d be having this same conversation, but it would end with both of us agreeing to donate everything we had to the Bolsheviks.

Alice: Okay, so the smart people were wrong once. That doesn’t mean…

Bob: And eugenics.

Alice: Actually…

Bob: □_□

Alice: Fine then. For the sake of argument, the smart people were wrong twice. That still doesn’t…

Bob: It does! A quick run through the history books shows that smart people trying to effect systemic change have an imperfect track record. I won’t say that they’re *unusually* bad compared to other demographics, but certainly nothing as stellar as the “let’s just not be morons” theory might lead one to expect. You like quantifiable things and specific examples, so let me give you one. I’ve sometimes thought that Friedrich Engels can be considered one of the fathers of effective altruism – at least of the earning-to-give variety. Wikipedia says:

> Once Engels made it to Britain, he decided to re-enter the Manchester company in which his father held shares, in order to be able to support Marx financially, so that Marx could work on his masterpiece Das Kapital. Engels didn’t like the work but did it for the good of the cause.

And in one sense, Engels-as-altruist was utterly brilliant. He effectively zeroed in on the most influential thinker of his era, funded an otherwise-impossible level of output from him, and his work directly led to revolutions in a dozen countries with radical change in the lives of billions of people. But in the more important sense, the net effect of his contribution was global mass murder without any lasting positive change. If we count him as an effective altruist – and under the circumstances I’m not sure we can do otherwise – then the net contribution of the movement throughout history has been spectacularly negative. That should make us really concerned. Not “nod sagely and promise to think about it” level of concern, but more “run away screaming” level of concern. That’s why I’m so reluctant to accept your otherwise-reasonable points about the seemingly obvious issues.

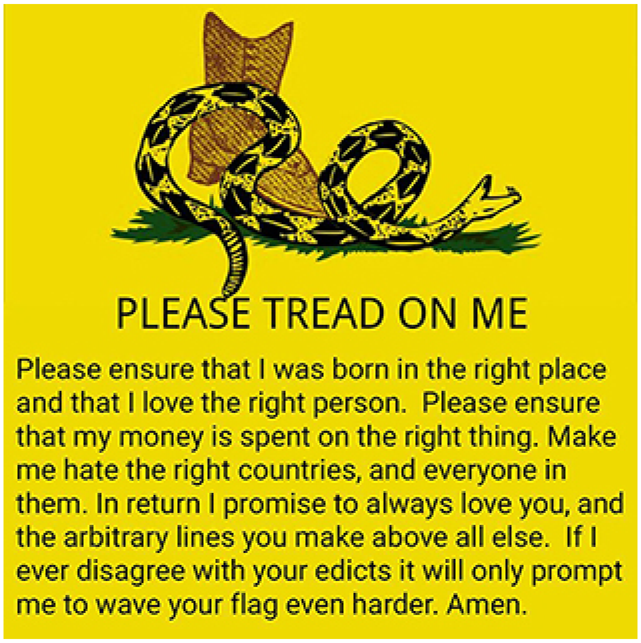
Alice: On an emotional level, I get your point. But on a rational level, wouldn’t it be astounding if smart people trying to figure out the safest ways to do the most good consistently made things worse?

Bob: There are many more ways to break systems than to improve them. One Engels more than erases all of the good karma created by hundreds of people modestly plodding along and making incremental improvements to things.

Given an awareness of long-tail risks and the difficulty of navigating these waters, I’m not sure our expected value for systemic change activism should even be positive, let alone the most good we can do.

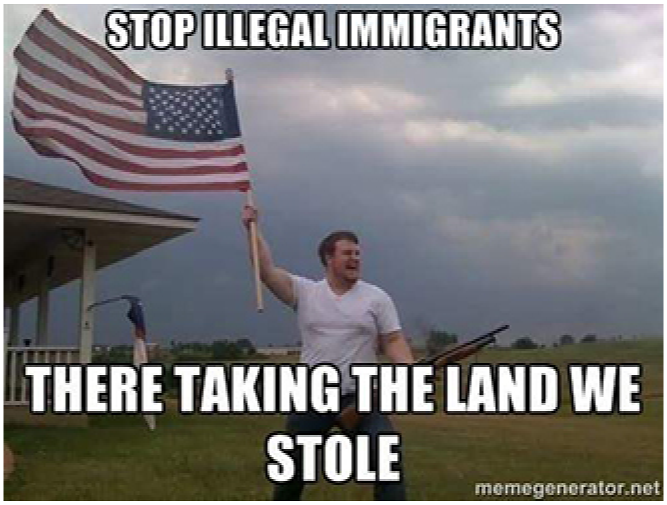
Alice: So don’t go poking around super-complex systems with lots of variables as complicated as “capitalism” versus “communism”. Stick to well-understood things with fairly predictable effects. If we have a little bit more humility than Engels, maybe we won’t fall into the same trap he did.

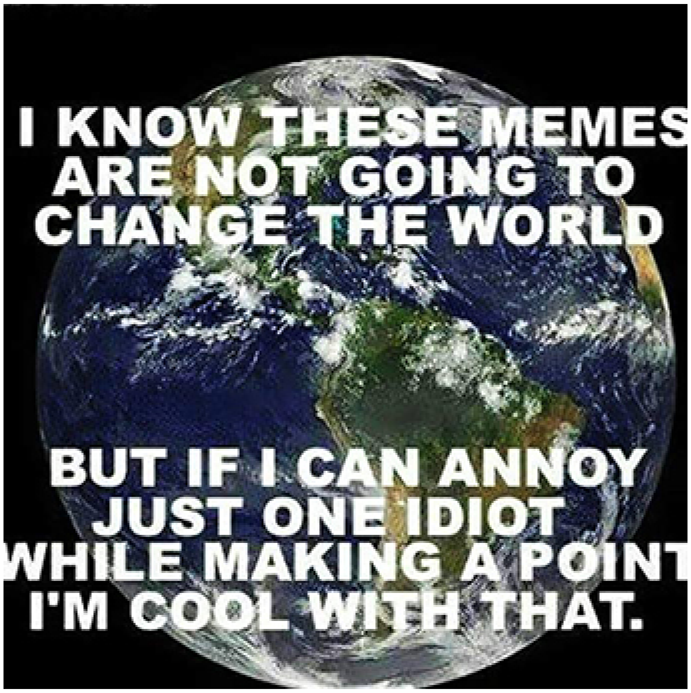
Bob: All nice and well, except that I do not see even the tiniest sign of supra-Engels levels of humility in the effective altruist movement as it exists today. Recently I have had to deal with lots of our Facebook friends joining and sharing images from a group called “Muh Borders!” which exists to post memes making fun of anyone who opposes Open Borders as a stupid bigot who is not worth talking to:



PLEASE TREAD ON ME

Please ensure that I was born in the right place and that I love the right person. Please ensure that my money is spent on the right thing. Make me hate the right countries, and everyone in them. In return I promise to always love you, and the arbitrary lines you make above all else. If I ever disagree with your edicts it will only prompt me to wave your flag even harder. Amen.


STOP ILLEGAL IMMIGRANTS
THERE TAKING THE LAND WE STOLE
memegenerator.net



In terms of "political causes that we can be totally sure won't backfire and devastate entire countries for generations", I would place open borders…well, let's say somewhere in the bottom quartile. A thorough analysis by one of its strongest and most intelligent advocates concludes with "doubt that the American polity could survive and flourish under open borders" but has been mostly ignored in favor of constantly retreading the same old streetlight-illuminated ground of whether immigrants do or don't affect native wages. And *this* is the community that is supposed to have solved the hard problem of getting mind-killed by politics, and can now be sure it's genuinely pursuing the side of Good rather than the side that looks like Good but actually kills tens of millions of people?

Alice: That's not fair. Yes, there are some people who reflect poorly on the open borders movement, but they're not all effective altruists, and even humble people who try their best to think about things rationally are allowed to let off steam on Facebook once in a while. The open borders movement has also done a lot of really impressive analysis, and even though there are risks, given the potential benefits it really can be thought of as a no-brainer.

Bob: I am a coward and will stick to buying bed nets, thank you.

II.

There's another problem with man vs. man: the people we want to recruit are men, and the people we want to make our movement out of are men.

(That came out sounding more sexist than I intended. You know what I mean.)

Several people have recently argued that the effective altruist movement should distance itself from AI risk and other far-future causes lest it make them seem weird and turn off potential recruits. Even proponents of AI risk charities like myself agree that we should be kind of quiet about it in public-facing spaces.

As someone whose own views on open borders are mixed (I should probably write a post), I am really turned off by memes like the one above. And since only seven percent of Americans fully support open borders, that's a lot of potentially turned-off people. They're going to go on effective altruist sites, see that a big part of the movement is arguing for a policy that they abhor, and notice their potential colleagues talking about how people like them who oppose that policy are stupid and parochial and hate foreigners. "We think you're wrong and stupid, come join our movement" makes a really crappy recruiting pitch. But it is the pitch we are sending to anyone who isn't a Silicon Valley libertarian, George Mason University economics professor, or Vox.com journalist – the only three groups from which I have seen a level of open borders support much beyond the lizardman level.

If effective altruists are split on an issue, then they're working at cross-purposes, like the people above who donated $20 million to both the Democratic and Republican campaigns. But if effective altruists are *not* split on an issue, then they're projecting a unified Effective Altruist Consensus on it which is going to look pretty intimidating to anybody who disagrees. And if there are enough of these issues, then a randomly selected person is almost certain to disagree with at least one of them. The more different from the EA stereotypical demographic they are, the more likely such a disagreement will be. Politics is the mind-killer and quickly takes over from everything else; I do not think political disagreements can stay quiet and harmless for long. If two people are both committed to healing the sick and feeding the hungry but one believes in open borders and the other in a more Bernie Sanders style approach to immigration – not even conservative, just a Bernie Sanders style approach! – they can peacefully coexist in an effective altruist movement focusing on traditional charity, but one focusing on systemic change is likely to get pretty heated.

If you think this is overly pessimistic, think back to the issues with the most recent EA Summit, which advertised fully vegetarian meals but added non-vegetarian options at the last second. This became a big enough scandal that I, who was two thousand miles away from the conference, got inundated with arguments about it on Facebook, Tumblr, and this blog. Several people threatened to quit effective altruism entirely, though I don't know if any of them followed through.

This is a community that can literally almost tear itself apart over the question of what to have for lunch. I think there might be too much dynamite around to risk shooting off sparks.

And I also think effective altruism has an important moral message. I think that moral message cuts through a lot of issues with signaling and tribal affiliation, that all of these human foibles rise up and ask “But can’t I just spend my money on – ” and effective altruism shouts “NO! BED NETS!” and thus a lot of terrible failure modes get avoided. I think this moral lesson is really important – if everyone gave 10% of their income to effective charity, it would be more than enough to end world poverty, cure several major diseases, and start a cultural and scientific renaissance. If everyone became very interested in systemic change, we would probably have a civil war. Systemic change is sexy and risks taking over effective altruism, but this would eliminate a unique and precious movement, in favor of doing the same thing as everybody else.

If effective altruism became more political, it would likely fade seamlessly into something like the Brookings Institution (a top-tier think tank whose $100 million yearly budget is by my calculations well above what the entire world combined spends on deworming) or the Cato Institute (another top think tank whose $30 million budget is likely more than all AI risk charities and all effective animal rights charities combined). Probably the staff of the Brookings Institution go into work each day thinking “How can I best improve the world by giving it better policies?”, and I admire that, but they don’t have the same sort of moral mission as effective altruism and it would be disappointing to see the latter collapse into an annex of the former.

Vegetarianism for Meat-Eaters

[Content warning: discussion of animal suffering. If you don't care about animal suffering, this post is probably not for you. There is no reason to read it anyway and loudly complain in the comments.]

Brian Kateman on QZ.com writes that We Need More Meat-Eating Animal Rights Activists. Finally, the mainstream media gives me ex cathedra permisson to say things that are kind of hypocritical!

I believe animals probably have moral value. I also eat meat. There is obvious tension between these positions; animals suffer and (obviously) die during meat production. I can only say in my defense that I tried being a vegetarian for several years and it was horrible and I ended up subsisting almost entirely on bread and Quorn and I don't want to go back there.

But over the past few years I've read about two ideas that have changed the way I look at meat-eating and significantly reduced my moral footprint with minimal inconvenience. These are not original to me and I don't take credit for them, but I hope that the people involved won't mind me taking this advantage to publicize them more widely.

1. Eat Beef, Not Chicken

This argument is so simple I feel dumb for not thinking of it myself; instead, I take it from Julia Galef and Brian Tomasik. Suppose I get about a third of my daily calorie requirement from meat; that adds up to 250,000 calories of meat a year. Further suppose that it's split evenly between 125,000 calories of beef and 125,000 calories of chicken.

The average cow is very big and makes 405,000 calories of beef; the average chicken is very small and makes 3000 calories worth of chicken. So each year, I kill about 0.3 cows and about 42 chickens, for a total of 42.3 animals killed. [1] [2]

Suppose that I stop eating chicken and switch entirely to beef. Now I am killing about 0.6 cows and 0 chickens, for a total of 0.6 animals killed. By this step alone, I have decreased the number of animals I am killing from 42.3/year to 0.6/year, a 98% improvement.

The difference becomes even bigger once you compare levels of suffering. Chickens are probably the most miserable farm animals; they are mutilated, packed into tiny cages to the point of immobility, left to fester in their own waste, and bred so intensively for size that their bodies cannot support them and they likely experience severe musculoskeletal pain. Although cows' lives are also pretty terrible too, Brian Tomasik estimates that chickens' suffering is about twice as bad. Taking this into account, switching from 50-50 to all-beef reduces your contribution to animal suffering as much as 99%. [3] [4] [5]

I find that I'm indifferent between beef and chicken as far as taste, so this is a no-brainer for me. The few times I'm making a recipe that really, truly, can only be done with something sort of chicken-like, Beyond Meat vegetarian fake chicken strips are an almost-tolerable substitute.

2. Use Ethics Offsets By Donating to Animal Charities

[EDIT 12/30/16: SERIOUS PROBLEMS WITH THIS SECTION, SEE MISTAKES PAGE FOR DETAILS]

I talked about this before in Ethics Offsets, but I think the original argument comes from Katja Grace.

Animal-related charities are very effective. Animal Charity Evaluators, a sort of animal version of GiveWell, lists really really impressive impacts for small donations:

Animal Equality: 11 animal lives saved per dollar
Mercy For Animals: 9 animal lives saved per dollar
Humane League: 3 animal lives saved per dollar

These numbers are high, but not impossibly so. For example, the Humane League spent about $50,000 convincing school districts to switch to cage-free eggs and have “Meatless Mondays” at their cafeterias; this resulted in about 3.2 million fewer meat-containing lunches, meaning several hundreds of thousands of chickens saved.

Okay. If you followed the advice in Part 1 and switched to beef, you’re currently killing 0.6 animals per year. If you donate six cents per year to animal-related charities, you’re animal-neutral. Donating $0.06 sounds…a lot easier than being vegetarian for a year? [6]

Or donate $60, and save more animals than an entire village full of vegetarians. At this point it’s starting to look like maybe personal vegetarianism is more of a symbolic/non-consequentialist decision in comparison, and a meat-eater with a little pocket change to spare can bask in near-unlimited moral superiority even to their most scrupulously vegan friends. Is this too good to be true?

One reason it might be too good to be true is that Animal Charity Evaluators is overly optimistic. But it would be *really hard* for their optimism to change this strategy substantially. Suppose that they were off by an order of magnitude, and you only save one animal per dollar. You can *still* offset an entire year’s beef-eating for $0.60. Even if they’re off by *three orders of magnitude* and it takes $60 to offset a year of eating beef, most people would probably still rather pay sixty bucks than become vegetarian.

A more serious complaint is that this strategy is hypocritical or self-defeating. After all, it looks like most of the gain from these charities comes from convincing other people to be vegetarians. From a Kantian point of view, “try to get other people to become vegetarian without being one yourself” isn’t universalizable; if everyone did it, there would be nobody to actually be the vegetarians! Is it ethical for non-vegetarians to try to spread vegetarianism among other people? Here are four arguments that it is:

First, consequentialism. From a consequentialist point of view, “is it okay to cause a good thing to happen even if…” always gets answered yes. Do you save the animals? Yes? Then what’s the problem? The true consequentialist doesn’t even understand the question.

Second, these charities don’t necessarily demand people become full vegetarians. They may recommend that people cut down on the amount of meat they eat, or switch from chicken to beef as in Part 1, or support laws enforcing more humane living conditions for farm animals. Some evidence supports asking meat-eaters to cut down on meat as the most effective form of animal outreach. A non-vegetarian who has taken some of these steps themselves can support these without worrying about hypocrisy.

Third, your situation is not necessarily the same as other people’s situations. One reason I’m not a vegetarian is that I really really hate vegetables. Other people might love vegetables and just need a little push to have more of them. I can endorse that people become vegetarian *if it is easy for them* without necessarily endorsing vegetarianism for myself.

Fourth, and I think most important, the economics check out. Instead of universalizing the principle “become vegetarian”, suppose we tried to universalize the principle “find some way to be animal-neutral,” that is, live your life in such a way that on net you are not killing animals. And suppose everyone knew there were two strategies for doing this: either become vegetarian yourself, or offset your lifestyle by donating to advocacy organizations that convert other people to do so.

And suppose that, upon hearing that it only takes a $60 donation to offset their lifestyles, 90% of people choose the donation rather than the personal conversion. This makes the cost of outreach go up. That is, when I donate my $60, the advocacy organization uses it to convert Alice, who decides to donate $60 herself, which the advocacy organization uses to convert Bob, who decides to donate $60 himself, which the organization uses to convert Carol… and so on to the tenth person, who finally decides to become vegetarian themselves. If this happened, our premise that it takes the charity $60 to convert one new vegetarian would be false. In fact it takes them 10 donations of $60, or $600.

As long as people know that they have the option of offsetting via donation, the possibility that people would rather donate than become vegetarian themselves *is priced into the cost of the offset*. That means that if the cost of an offset is currently $60, it’s because we’re hitting people for whom $60 is genuinely their reserve price; they prefer becoming vegetarian to paying a $60 offset (probably for moral/symbolic reasons). These people are low-hanging fruit; once they’re exhausted, the offset price will rise, and people for whom vegetarianism is only a mild inconvenience will find themselves preferring to become vegetarian themselves rather than paying. Once even the middle-hanging fruit is exhausted, the price of the offset will be prohibitive and only the people for whom vegetarianism is an extraordinary inconvenience will continue to take that route. Once there are no more potential vegetarians left to convert, the offset cost will become the cost of saving animals via political action, improved technology (eg cultured meat), or changes to farming conditions.

This dynamic becomes even more interesting if you add the (unjustifiable but interesting) assumption that anyone not becoming vegetarian themselves is required to offset their choice by converting *two* other people to vegetarianism. Then you get a sort of virtuous Ponzi scheme which ends with a *lot* of vegetarians (albeit not necessarily in a reasonable amount of time).

I try to donate some money to an effective animal charity each year, above and beyond what I've pledged to donate for other reasons, in order to compensate for the remaining meat I refuse to cut out of my diet.

Footnotes

1. I use the term "kill" because it's a simple way of looking at things, but most of the moral cost of eating meat is causing the animals to spend years living in terrible suffering on factory farms. The actual killing is probably a mercy in comparison. When I say that something "prevents forty animals from being killed", the longer and more accurate version might be "prevents forty animals from coming into existence, suffering intensely, and then being killed". This does raise some more philosophical questions like whether it's better to live a life of terrible suffering than to never be born at all, but I'm really comfortable answering that one with "no".

2. This same argument comes out against eating other small animals like fish. Although in theory wild-caught fish ought to live okay lives and potentially be more ethically acceptable than farm-raised animals, given limited wild-catching ability each wild-caught fish eaten may deplete a fixed number of them and push other people to eat farm-raised fish instead.

3. Eggs raise some of the same issues as chickens, and Julia Galef suggests eggs are one of the worst things you can eat. I think her assessment is pessimistic; eggs are terrible on a calorie-for-calorie basis, but if we're talking about which animal products to urge people to give up, this is counterbalanced by nobody except Gaston getting too many calories from eggs. Someone who eats one egg with breakfast every day kills about one chicken a year; somebody who has a chicken dinner every other night kills about forty chickens a year. Although egg chickens probably lead worse lives than meat chickens, the difference isn't overwhelming. Avoiding incidental egg consumption like the eggs in baked goods is hard and probably not the highest-value pro-animal intervention given the low number of eggs involved.

4. This analysis neglects consideration of whether cows, being bigger-brained and more "evolutionarily advanced" than chickens or fish, might have greater moral value. I don't know how to deal with that question, except that it would surprise me if they had more than forty times the moral value.

5. The existence of supposed humane animal products ("Free range eggs!" "Pasture-fed cows!") complicates this a little bit. The unanimous opinion of people who know about this sort of thing is that free range eggs are kind of a scam; regulations only specify that these chickens must have “access” to the outdoors, but farmers exploit the letter of the law to cram thousands of chickens into industrial barns with a single tiny door to a couple-square-foot cement yard that the overwhelming majority of the chickens never even see. “Cage-free” chickens or eggs seem probably better than the alternative but still pretty horrible. “Pasture-fed beef” usually does involve a pasture in some way and is not a total scam but is probably not as nice as you would think. I try to buy pasture-raised free-range cows, and I think that the slightly higher standards of humane beef over humane chicken make another good argument in favor of beef consumption, but I try not to fool myself into thinking that this decision alone goes especially far.

6. If you also eat chicken, the offset cost rises to $4.

The Problems With Generic Medications Go Deeper Than One Company

I.

Like many people, I recently read about Turing Pharmaceuticals’ purchase of anti-toxoplasma drug Daraprim and subsequent price increase of 5000%. Vox and Marginal Revolution have already done some good work addressing this particular case, but have only touched upon the broader issue: that everything about generic medications is approximately this terrible.

As far as I’m concerned, the interesting aspect of this case isn’t just that the CEO of Turing is an asshole who is lining his own pockets with zillions of dollars by gouging AIDS patients. I assume most pharmaceutical company CEOs are assholes who would line their own pockets with zillions of dollars by gouging AIDS patients if the opportunity presented itself. The interesting aspect of this case is that the CEO of Turing got the opportunity. How?

In the United States, pharmaceutical companies that discover a new drug are granted a 20-year term of exclusivity to reward them for the public service of drug research. During this time, they can and do price-gouge as much as they want. After twenty years, the drug becomes public domain and anybody who wants can compete to produce it, usually leading to a precipitous fall in costs. But Daraprim is fifty years old; its patent is long-since expired. So Sarah Kliff from Vox asks the obvious question: why doesn’t someone just produce a competitor?:

> Daraprim isn't a frequently used drug. The New York Times estimates that between 8,000 and 12,000 prescriptions get filled annually. You could only fill about a quarter of a baseball stadium with the number of people who take the drug in a given year.
>
> So think about a generic drug manufacturer looking at the Daraprim situation. There are fixed costs associated with building a new plant (or possible lost revenue on other drugs, if they switch production at an existing plant), getting samples of the drug, and figuring out how to make the generic product…with Daraprim, there simply isn't a big enough patient population for a competitor to sell a "good amount" to. And this is, more generally, a problem with the markets for drugs that only a small number of patients use. They often aren't big enough to support two competitors.
>
> Moreover, there's risk associated with starting a drug price war. Let's say I decide to launch Sarah's Generic Drug Company, and I'm pretty sure I can break even by slightly undercutting Turing and charging $700. What happens if Turing responds by dropping its price down to $500, or even back to $13.50? It will keep all its patients — and my nascent drug company is likely going bankrupt.

This is definitely part of the story. On the other hand, what about Longecity group buys? Someone on a drugs forum hears about a cool experimental chemical that sounds fun to try. They get a couple dozen friends in on it and pay a lab in China a few hundred dollars to synthesize a big batch. Then the Chinese ship it over, they distribute it to their friends, and they all get a decent supply of a totally novel drug for a few dollars a pill – compared to the $750 per pill that Turing is charging for Daraprim. I am not a chemist, but the Daraprim molecule does not look very intimidating. I bet if a group from Longecity got a couple of toxoplasma patients together for a group buy, they could all get treatments for maybe a few hundred dollars each instead of the $63,000 Turing is now charging. In fact, I encourage somebody to do exactly that as an act of civil disobedience/political activism and win themselves some free publicity.

So how come Longecity can do this, but real generic pharmaceutical manufacturers can't? I'm not totally sure, but my best guess is that it involves bioequivalence studies (different from purity studies). Generic drugs don't need the excruciatingly drawn-out safety and efficacy studies required of new brand-name medications, but they do need to pass a bioequivalency study proving that their drug is absorbed the same way as the original. According to Wikipedia, the most common type of bioequivalence study is to "measure the time it takes the generic drug to reach the bloodstream in 24 to 36 healthy volunteers; this gives them the rate of absorption, or bioavailability, of the generic drug, which they can then compare to that of the innovator drug".

This might not seem so bad, but it must be harder than it sounds. This site, whose style is overly bombastic but whose information seems mostly correct, says that:

> The cost and time involved in the ANDA [generic application] process varies depending on the drug, its safety, how long it has been on the market, etc. To have an ANDA approved, it typically requires an investment of about $2 million, and it takes a total of two to three years to get the drug to market…in addition to these costs, a company should budget 15% for legal fees, because wherever there is a big manufacturer with a

> sizable market share involved, they will sue, just to try to eliminate more competition from the market.

This adds an important extra dimension to Vox’s theory that it’s just *too hard* to start making a generic medication. If all you want to do is synthesize an active ingredient in powder form, and you’re not too concerned about staying on the right side of the law, it costs pennies and takes however long you need to FedEx something from China. If you also want FDA approval, it costs $2 million and takes two years.

Remember, Daraprim is used by about 10,000 people per year, and before the recent Turing price markup, it cost $13.50 per pill x eighty pills per treatment. 10,000 * 80 * $13.50 = about $10 million per year, of which maybe $5 million was profit. That means you have to capture a *big* chunk of the Daraprim market before it’s worth trying to get yourself approved to make Daraprim; the FDA is essentially telling pharma companies to “go big or go home”. Nobody wanted to go big, so they all went home.

In the absence of this barrier, it would be easy for small boutique companies with a couple of chemical engineers on hand to spend a few weeks manufacturing a few thousand doses of the drug whenever it was necessary to meet demand. This is how the supplement and nootropic industries work right now, and nootropics are dirt cheap, even though a lot of “nootropics” are the same chemicals as regular expensive medications except with a “not intended for human consumption” label slapped on the bottle that everyone knows to ignore.

I *think* this might be what’s going on with generic modafinil. Last week I prescribed some modafinil to one of my patients and got a call back from their insurance company saying it was denied because it cost too much.

I told the insurance company that was silly because modafinil only cost about $60 a month.

The insurance company said no, it cost way more than that.

This surprised me, because half the rationalist community uses modafinil, and even some of the doctors I work with use modafinil on long night shifts, and *they* all get it for $60 a month from places like ModafinilCat.

But according to Nootriment, a month's supply of modafinil at real bricks-and-mortar pharmacies costs anywhere from $469.23 (Costco) to $850.84 (RiteAid). I'm not totally sure what's going on, but my guess is that ModafinilCat (illegally) buys it from people who haven't gone through the FDA's bioequivalence testing, and RiteAid buys it from people who have. As far as I can tell, both are made by Indian pharmaceutical companies unrelated to the original American company who discovered the drug, but RiteAid's Indian pharmaceutical company has put more work into staying on the right side of the US government.

If any of my patients are reading this and are upset because I prescribed them a drug which they couldn't afford, I unreservedly apologize. I was laboring under the misapprehension that the pharmaceutical market made sense.

II.

No tour of terrible generic medications policies would be complete without a stop for Kesselheim and Solomon's analyis of the Unapproved Drugs Initiative of 2006.

The FDA wanted to encourage people to study drugs that were already in the public domain and get them up to FDA standards. This is potentially a very noble plan. I've written before on how it's basically impossible to get melatonin to interface with the health care system because it got into the public domain without the relevant FDA standards being met. Likewise, there's no interest in using minocycline to treat schizophrenia because it's a public-domain drug and nobody profits off of doing the FDA compliance work. So the FDA was definitely responding to a real problem.

Their solution, though, was to say that if anybody did a good enough study on a public domain drug, they could grab it *out of* the public domain and have it be their exclusive drug for the next while. This was a terrible terrible terrible idea.

Colchicine is a very popular and very effective gout treatment extracted from the *Colchicum* plant. It's been used for so long that its first recorded mention in medical literature is on an ancient Egyptian papyrus. The medievals called it "hermodactyl"; Arabic physician Avicenna recommended it; notable gout sufferer Ben Franklin brought the first *Colchicum* specimens to North America.

But the ancient Egyptians, being a primitive and barbaric people, had no FDA. And although many different groups had done studies proving colchicine effective, none of them had done so on the official FDA forms. In 2007 a company called URL Pharmaceuticals did an official FDA safety study, showed that yup, it was safe all right, and for this service were granted exclusive right to produce colchicine. After suing all other colchicine producers out of business and establishing a monopoly, they raised the price of colchicine by 5000%, costing gout patients thousands of dollars a year.

According to FiercePharma, something similar happened with hydroxyprogesterone caproate, although the FDA later changed its mind. I can't find any other examples, but the legal framework is still there if someone else wants to try.

III.

Other times generic manufacturing proceeds smoothly. A drug is popular and many different pharmaceutical companies pass the bioequivalency tests, get in on the action, and compete with one another. Nobody snatches it out of the public domain at the last second and receives a new monopoly on it. The companies are able to sell it to the pharmacies for a reasonable cost.

Now you get to have a completely *different* set of things go wrong.

Michigan Drug Prices is my state's official drug price register. You can type in any Michigan ZIP code and any drug and find out how much it costs at all your local pharmacies. It's pretty neat.

Celexa has been generic for more than a decade, it's got a reputation for being inexpensive, and I prescribe it a lot. Let's see how much my patients have to pay.

The closest RiteAid to my office charges $4 for a 30-day supply of Celexa 20 mg. The local CVS sells the same amount for $19.79. The local Walgreens sells it for $24.99. And the local KMart will sell it to my patients for the low, low cost of $88.15. That's an…interesting…range of prices.

If I try to buy it off GoodRx.com, a site that offers pharmacy price comparisons, I can get it for $3.60 from a mail-order pharmacy. But I can also get it for $6.64 from K-Mart, special offer for GoodRx customers only. $10.00 from Walgreens. $11.99 from CVS. All the same stores that were trying to gouge me before. As soon as you take the basic step of saying "by the way, I'm also comparing costs with other pharmacies" their prices drop 90%.

I am far from the only person to notice this. PBS did a segment on one of the reporter's mothers looking for a breast cancer drug. She originally paid $400 a month for it, which is steep but perhaps worth the cost as a high-tech treatment for a potentially fatal illness. Then she went to Costco and found the same medication cost $10.

Why does this sort of thing happen? I'm not sure. I expect it has something to do with insurance co-pays; if an insurance looks at some kind of average cost of Celexa and decides that the Celexa co-pay will be $5, then it doesn't much matter to the customer whether they buy it from a pharmacy charging $10 or $10,000. But why doesn't the insurance company do one the thing everyone in health care agrees insurance companies do best: send whiny faxes complaining that they're not going to pay you? I don't know.

But for now you might want to try using something like GoodRx.com if you're buying expensive medications. And stay away from cats, because there's never been a worse time to get toxoplasma.

Swifties 3: The Race Is Not To The Swifty

[see: Wikipedia: Tom Swifties, Tom Swifties Written By An Author Willing To Go To Any Lengths To Make A Tom Swifty Thus Resulting In Constructions That Often Require More Work For Readers Than For The Author, Fifty Swifties, and Fifty More Swifties. Previously on Twitter here. Some of these are from the comments to the last post.]

1. “She eventually absorbed so much radiation that her bottom half mutated into a fish’s tail,” Tom said mercurially.

2. “Stay away from nuns,” Tom said conventionally.

3. “Back during Late Antiquity, everyone lived in fear of Attila and his hordes,” Tom said a hundred times.

4. “It said he was eaten by a bare, so either that’s a typo or he was devoured by the act of exposing something,” Tom said verbatim.

5. “You’ll have to stand,” Tom said deceitfully.

6. “Little plays are such a useful way to teach children good behavior,” Tom said schizoaffectively.

7. “…” Tom said immutably.

8. “I’m an only child,” Tom said in unison.

9. “Look, a Confederate general!” Tom said icily.

10. “Why yes ma’am, I AM the Tom from those Twitter one-liners you’ve heard,” Tom said pungently.

11. “I’m not going to make a deathbed conversion,” Tom said diagnostically.

12. “I’m using behavioral conditioning to train lions to keep quiet,” Tom said to Rorschach.

13. “I used to be a priest, but I was defrocked for an improper relationship on the job,” Tom said inundated at work.

14. “I’m here helping people displaced by the earthquake,” Tom said with intensity.

15. “We’ve been pinned underneath fallen logs,” Tom said treasonously.

16. “I went rock-climbing with my girlfriend,” Tom updated.

17. “The defibrillator worked!” Tom said, repulsed.

18. “My karate instructor died,” Tom said, desensitized.

19. “Godzilla, I can’t believe you devoured part of South Africa,” Tom transvaluated.

20. “I was running late today, so I had lunch in my cubicle,” Tom incubated.

21. “But they dug too greedily, and too deep,” Tom undermined.

22. “The new environmental regulations will make mineral extraction less profitable,” Tom said, determined.

23. “He’s sleeping six feet under now,” Tom said depressed.

24. “I guess I lost the genetic lottery,” Tom said, drawing a portrait.

25. “SKULLS FOR THE SKULL THRONE!” Tom said, skulkingly.

26. “I’ve gotten 0.028 countries to join together in a political and economic union,” Tom said in his milieu.

27. “For here I am, sitting in a tin can, re-entering Earth’s atmosphere,” Major Tom said incandescently.

28. “The mailman just left my mail on the dirty ground?! Really?!” Tom said postindustrially.

29. “I’m writing a book based on ‘The Tell-Tale Heart’, except instead of a horror story it’s a comedy,” Tom said politely.

30. “Is the guy in that coffin Dracula, or just an ordinary corpse?” Tom countermanded.

31. “I think China has enough foreign exchange reserves,” Tom said for example.

32. “Every time the server goes down, I have a Norse god zap it with lightning to get it back up,” Tom said with authority.

33. “Help, I’ve been buried alive!” Tom engraved.

34. “I’ll never be an A-list celebrity” Tom berated. (source)

35. “If you were any good you’d have the Ambassador’s job,” Tom said disconsolately. (source)

36. “Germany should exit the Eurozone” Tom remarked. (source)

37. “Maybe he was knighted for his contributions to Austrian economics,” Tom surmises.

38. “We should give the Western US back to the Native Americans,” Tom said unsettlingly.

39. “I’m not going to give that jerk Procrustes the satisfaction,” Tom said self-defeatingly. (source)

40. “This new-ideas conference has sure gotten effeminately quaint.” Tom tweeted. (source)

41. “Everyone’s date of birth is in 2007,” Tom said alternatively. (source)

42. “Weasley for president!” Tom said electronically. (source)

43. “Let the other guy take the paddle,” Tom said heroically. (source)

44. “Let’s make a deal – I’ll stop doing sit-ups if you do,” Tom said abstrusely. (source)

45. “My former wife mentioned me in her newest paper,” Tom said excitedly. (source)

46. “How Can Mirrors Be Real If Our Eyes Aren’t Real?”, Tom asked unreliably. (source)

47. “Your hair looks terrible,” Tom said distressingly. (source)

48. “I’ve stolen the treasures of the Shrine of the Bab,” Tom said, high-falutin’.

49. “We should go to the petting zoo, I hear they have cattle now,” Tom said, compatible with me.

50. “After Kant’s death, he left his old machine gun to forces plotting a military coup,” Tom said, willing that his maxim could make a general rule.

Contra Caplan on Mental Illness

I.

Bryan Caplan has a 2006 paper arguing that economic theory casts doubt on the consensus view of psychiatric disease. He writes:

> Economists recognize the benefits of specialization. Only with hestitation, then, can economists focus their attention on an unfamiliar discipline and conclude that experienced professionals have been making elementary mistakes. However inconsistent psychiatry's main theses seem to be with basic consumer theory, one might think it foolhardy to conclude that they are wrong.
>
> At the same time, economists also recognize not only that rentseeking is a ubiquitous force, but that most rent-seekers create and internalize public-interested justifications for their activities. It is not overreaching for economists to criticize domestic auto makers' arguments for protectionism. The auto makers know more about the details of their own industry, but economists are better at interpreting those details. Equally importantly, economists are trained to consider the costs of a policy for everyone in society, not merely groups with the most political influence.
>
> From a rent-seeking perspective, skepticism about psychiatry is common sense. Rent-seeking is only a side activity for the auto industry, but it lies at the core of psychiatry.

Calling someone a rent-seeker is sort of an economist's way of telling them to die in a fire, so I feel honor-bound to respond.

As best I can tell, Caplan's argument goes like this:

Consumer theory distinguishes between two different reasons why someone might not buy a Ferrari – budget constraints (they can't afford one) and preferences (they don't want one, or they want other things more). Physical diseases seem much like budget constraints – the reason a paralyzed person can't run a marathon is because it's beyond her abilities, simply impossible. Psychiatric diseases seem more like preferences. There's nothing obvious stopping an alcoholic from quitting booze and there's nothing obvious preventing someone with ADHD from sitting still and paying attention. Therefore they are best modeled as people with unusual preferences – the one with a preference for booze over normal activities like holding down a job, the other with a high dispreference for sitting still and attending classes. But lots of people have weird preferences. Therefore, psychiatric diseases should be thought of as within the broad spectrum of normal variation, rather than as analogous to physical diseases.

He compares this to the work of Thomas Szaszszszsz, who proposes that psychiatry is an inherently political enterprise that works to delegitimize people with unusual preferences. For example, until the 1970s homosexuality was considered a psychiatric disease, and now it is considered an uncommon but legitimate preference. In the past being transgender was considered a psychiatric disease, but now many people are moving toward considering it an uncommon but legitimate preference. In each case, when society thinks that a preference is gross, or anti-social, or so extreme that they can't imagine themselves having it, they shout “Psychiatric disease!” and then they can stick anyone who offends them in mental hospitals; if the preference becomes more legitimate, they retreat and say “Guess those ones weren’t psychiatric diseases after all, but we’re still 100% sure all the other ones are”. Caplan says that instead of these constant mini-retreats we should just admit that all psychiatric diseases are unusual preferences.

He admits that he’s making his job too easy with examples like alcoholism, and that something like schizophrenia would make a harder test case. But, he asks, what is schizophrenia? Delusions and hallucinations. Delusions may be a preference to have a weird belief – for example, somebody might feel trivial and neglected, so in order to make themselves feel important they cook up a paranoid conspiracy theory where the FBI, CIA, and Freemasons are all after them because of how important they are. Wanting to believe that the government is after you is a weird preference but, Caplan insists, nevertheless still a preference. As for hallucinations, sure, schizophrenic people say they hear voices, but we all kind of hear internal voices in the sense of “hear ourselves thinking in our head”, and schizophrenics must just be people who prefer to explain those in very vivid external terms. Visual hallucinations are the same way – we all have an imagination, but some of us react to our imaginations differently than others.

Therefore, all psychiatric diseases can be conceptualized in the form of preferences. This makes them very different to physical diseases, which are budgetary limitations, and we should stop saying they’re the same thing and locking mentally ill people up and having all of these rent-seeking psychiatrists around saying they can “cure” them.

Caplan ends by noting that genetics and neurobiology cannot prove him wrong. Yes, weird preferences may be genetic, and they may be linked to weird neurobiology, but so are our normal preferences! There are genetic factors influencing schizophrenia, but there are also genetic factors influencing politics, religion, and extraversion. Yes, drugs can make you less schizophrenic, but they can also make you less extraverted.

I agree with Caplan's last paragraph. We can't prove him wrong with neurobiology alone. So let's prove him wrong with philosophy, psychology, economics, and common sense.

II.

Let's start with preferences vs. budgetary constraints.

Alice has always had problems concentrating in school. Now she's older and she hops between a couple of different part-time jobs. She frequently calls in sick because she feels like she doesn't have enough energy to go into work that day, and when she does work her mind isn't really on her projects. When she gets home, she mostly just lies in bed and sleeps. She goes to a psychiatrist who diagnoses her with ADHD and depression.

Bob is a high-powered corporate executive who rose to become Vice-President of his big Fortune 500 company. When he gets home after working 14 hour days, he trains toward his dream of running the Boston Marathon. Alas, this week Bob has the flu. He finds that he's really tired all the time, and he usually feels exhausted at work and goes home after lunch; when he stays, he finds that his mind just can't concentrate on what he's doing. Yesterday he stayed home from work entirely because he didn't feel like he had the energy. And when he gets home, instead of doing his customary 16 mile run he just lies in bed all day. His doctor tells him that he has the flu and is expected to recover soon.

At least for this week Alice and Bob are pretty similar. They'd both like to be able to work long hours, concentrate hard, and stay active after work. Instead they're both working short hours, calling in sick, failing to concentrate, and lying in bed all day.

But for some reason, Bryan calls Alice's problem "different preferences" and Bob's problem "budgetary constraints", even though they're presenting exactly the same way! It doesn't look like he's "diagnosing" which side of the consumer theory dichotomy they're on by their symptoms, but rather by his assumptions about the causes.

And his assumptions about the causes may be wrong. Bob's issues are probably caused by what we call "sickness behavior", a chemical defense in which the immune system notices an infection and releases cytokines telling your brain to avoid action and conserve energy in order to help with recovery. But one of the theories of depression I have found most plausible is that it's a malfunctioning of sickness behavior – you're not necessarily really sick, but your immune system releases its "stop acting and lie in bed all day so we can recover" chemicals anyway. If flu and depression have the same proximal cause, and the same effects on your life, where does Bryan draw the budget/preferences line?

For that matter, does Bryan ever get tired? I mean, suppose he is up very late one night and then has to go to work on only an hour of sleep. If he's like the rest of us, he probably does a terrible job, can't concentrate, and maybe rushes through things to get home early so he can catch a nap. Is this a budgetary constraint, or different preferences? In one sense it seems budgetary – he is lacking a resource (sleep? mental energy?) that would allow him to do a good job if he had it. In another sense it is clearly preferential – he places much less value in working hard and much more value in rushing home to get a nap.

Either way, this seems like a fruitful way to think about conditions like ADHD. Someone with ADHD, like someone who's working on an hour of sleep, finds themselves miserable and unable to focus. If we call this a budgetary constraint, Bryan's whole argument comes tumbling down. But if we call it a preference, then it's a very strange type of preference, one where the usual method of "oh, great, you're doing what you prefer!" is entirely the wrong approach.

In the case of the sleep-deprived person, their "new set of preferences" seems like a malign and unpleasant condition inflicted on them by an external source – namely, their sleeplessness. It's certainly not "what they prefer" in the sense of "oh, great, he's doing what he prefers, now he's self-actualized and all is right with the world." Instead, we admit that it's a problem, *they* admit that it's a problem, and we prescribe a biological cure – more sleep.

Or here's another example. Suppose while you are asleep I inject you with a little machine that constantly releases interferon into your bloodstream – interferon being a hepatitis medication notorious for causing deep depression as a side effect. You become depressed, and your preferences change from "work and spend time with my friends" to "lie in bed all day". But a helpful wizard gives you a powerful antidote. If you take the antidote every day, the interferon is rendered harmless and you are as active as ever. Unfortunately, you run out of antidote, and you lie in bed all day.

So: is your lying in bed all day a preference, or a budgetary constraint caused by shortage of antidote?

My answer: dichotomies sometimes break down outside a certain scope. I enjoy reading Bryan's posts about immigration, which often compare immigrants to natives along some axis. Immigrant vs. native is a useful dichotomy for a lot of purposes. But: were the guys on the *Mayflower* immigrants or natives? Were slaves abducted from Africa to work on plantations immigrants or natives? Were the couple thousand residents of California who went to bed as Mexicans the night before the Treaty of Guadalupe Hidalgo and woke up as Americans the next morning immigrants or natives? What about migratory birds? The number three? The Devil?

I propose that the preference/budget distinction is a bad way of dealing with anything more complicated than which brand of shampoo to buy. We intuitively talk about our choices as if there were some kind of "mental energy" that allows one to pursue difficult preferences, and I discuss some ways this confuses our intuitive notion of budgeting in Parts II and III here. You don't have to accept any particular framing of this, but to sweep the entire problem under the rug is to ignore reality because you're trying to squeeze all of human experience into a theory about shampoo.

III.

If there's not a lot of difference between preferences and budgetary constraints, what does that say about the relevance of Thomas Szasz?

(By relevance, I mean relevance to the modern day; he wrote in the 1960s and what he wrote was very possibly more true then than it is now. Possibly played a big part in making the things he wrote about less true, and should be celebrated for this. But I'll concentrate on the present.)

Szasz and Caplan both says that mental illnesses are attempts to stigmatize those with unusual preferences. I say that mental illnesses can reflect people's genuine worries about a-thing-sort-of-like-a-budgetary-constraint afflicting them. Which of us is right?

Well, consider that about 95% of people who go to an outpatient psychiatrist do so of their own free choice. This is certainly the case with my own patients. They are people who have gotten tired with the constraints that mental illnesses put on their lives, come in and say "Doctor, please help me", and I try to help them achieve whatever goals they have for themselves.

About 50% of people who go to inpatient psychiatric facilities also go of their own free choice. The rest, assuming everyone's following the legal system and the appropriate ethical guidelines, are people who are "dangers to themselves or others". I admit, it takes a controversial value judgment to decide people shouldn't be allowed to be dangerous toward themselves – though I think in some cases that judgment can be justified. And I admit that "danger to others" can sometimes be stretched to the point where if a psychiatrist wants to commit someone they can probably make up a justification. But these implementation problems are a *heck* of a long way from Caplan and Szasz's theory of "psychiatry is just a project about finding weird people and locking them up."

The psychiatric profession will never live down the thing about homosexuality; I fully expect that in 5000 AD someone will still be complaining that we can't stigmatize entities infected with superintelligent self-replicating memetic viruses, because DSM-II listed homosexuality as a psychiatric disease. But there's still a chance to rebut the thing about transgender, so let me quote from the APA website's discussion of the topic:

> The [new DSM criteria] underscore that being transgender is not a disorder in itself: Treatment only is considered for transgender people who experience gender dysphoria — a feeling of intense distress that one's body is not consistent with the gender he or she feels they are, explains Walter Bockting, PhD, a clinical psychologist and co-director of the LGBT Health Initiative at Columbia University Medical Center.

In other words, the decision about whether transgender people need psychiatric help is left up to – transgender people. If they don't want it, they don't have to have it. If they do want it, the option is open to them and their condition is recognized as a legitimate reason to seek help that insurance companies will support. I myself have treated exactly one patient for gender dysphoria. She was so depressed about her gender that she was considering suicide. I gave her some antidepressants, some supportive therapy, and some information about local support groups and sex-change professionals. Then I billed her insurance company for gender dysphoria treatment and got a check. Truly everyone involved is Worse Than Hitler.

IV.

Caplan admits that some mentally ill people seek help voluntarily and are among the most vocal proponents of the "real disease" theory. In order to shoehorn this into his preference-budget dichotomy, he theorizes that this is an attempt at deception. For example, alcoholics' insistence that they cannot resist drinking alcohol is deceptive:

> From an economic point of view, however, what is so puzzling about a person who prefers consuming alcohol to career success or family stability? Life is full of trade-offs. The fact that most of us would make a different choice is hardly evidence of irrationality. Neither is the fact that few alcoholics will admit their priorities; expressing regret and a desire to change is an excellent way to deflect social and legal sanctions.

But in order to fully explain alcoholic behavior, we have to take this theory *exceptionally* far. Consider a typical alcoholic drinks for several years, then “hits bottom”, goes sober, and joins Alcoholics Anonymous. He attends AA meetings three times a week for three years, then has a really bad day and binges on alcohol. Afterwards he is so embarrassed that he attempts suicide, but is rushed to the hospital and resuscitated successfully. After that he goes back to his AA meetings.

Does this man have a *preference* for going to AA meetings three times a week for several years then getting really drunk then attempting suicide? That’s a *weird* preference to have. Does he have a preference to drink, and in order to be socially acceptable he ‘covers up’ his one episode of binge drinking by years of AA meetings and a serious suicide attempt which he secretly knows will fail? That is a pretty disproportionately big web of lies, especially when probably no one would blame him for binge drinking one night one time.

If we’re willing to be this paranoid, we can basically prove or disprove anything. Bryan Caplan says he’s a libertarian, but my 9th grade Civics textbook says there are only *two* political parties, Democrats and Republicans. If Bryan says he’s in a third, he must just be trying to “deflect social and legal sanctions”. Maybe he’s secretly a Republican, but he wants to fit in to academic culture, so he says all of this stuff about “libertarianism” as a cover. His work writing hundreds of essays and some pretty decent books supporting his libertarian viewpoint are to maintain the credibility of his signal and throw us off the trail. Any donations he may have made to libertarian causes are the same…

…or we can be skeptical of textbooks that try to reduce things to simple dichotomies, whether that’s Democrat/Republican or preference/budget.

Caplan sort of flirts with admitting this:

> Cooter and Ulen probably speak for many economists when they deny that the preferences of the severely mentally ill are well-ordered. But in fact, not only do individuals with mental disorders typically have transitive preferences; they usually have more definite and predictable orderings than the average person…it is also implausible to interpret most mental illness using a ‘hyperbolic discounting’ or ‘multiple selves’ model. These might fit a moderate drug user who says he ‘wants to quit’…but they do not fit the hard-core drug addict whose only wish is to be left alone to pursue his habit. The same holds for most serious mental disorders: they are considered serious in large part because the affected individual continues to pursue the same objectionable behavior over time with no desire to change.

But if we take that middle part seriously he is ceding me 99.9% of the ground without remarking on it. Most people with mental disorders and substance abuse disorders wants to get rid of their disorder or at least alleviate the worst parts of it. If you are willing to accept complicated “multiple selves” models for those, then that is what you should be using to model mental disorders, not the simple consumer price theory.

And the others? The alcoholic who says “Yup, I’m drinking myself to death and you can’t stop me?” I agree that it is in some sense rational. It is rational because that person has so many problems that drinking alcohol becomes more pleasant than dealing with them. Often, these problems are related to psychiatric issues – for example, many people with PTSD become alcoholics because alcohol helps them briefly forget their traumatic memories. There are many people who say they don’t want help with their drinking problem because they expect “help” to mean “take away the alcohol but give them nothing in exchange”. If “help” meant “replace the alcohol with some healthier coping mechanism that works just as well”, many of these people would take it in a heartbeat. I realize this doesn’t quite disprove Caplan’s thesis for this relatively small group of alcoholics, but I think it’s important to remember that “preference” is different from “they’re doing what they want and all is well”.

V.

Finally, Caplan moves into a discussion of schizophrenia. He says that hallucinations might just be people hearing their normal inner voice and seeing their usual inner imagination, but they choose to describe it differently:

> Szasz similarly maintains that many alleged hallucinations are only eccentric descriptions of ordinary experience. To take the most common form, psychiatrists routinely equate ‘hearing voices’ with auditory hallucination. But when a person feels guilty, we often say that he hears the voice of conscience…to take a stronger case, the DSM treats ‘a voice keeping up

> a running commentary on the person's behavior or thoughts' as an exceptionally serious symptom. But this describes any person deliberating between major life options over an extended period of time. While these examples might seem to stretch the meaning of 'hallucination', it is the DSM that explicitly fails to distinguish whether the source of the voices is perceived as being inside or outside of the head.'

This makes sense, which is why every psychiatrist for the past century has specifically asked patients whether that's what's happening before diagnosing them with anything. Any time a patient reports a hallucination to me, the first question I ask is whether they're just embellishing on hearing an inner voice, or whether they actually heard an external voice clearly and distinctly the way they are hearing me talk to them right now. Sometimes they *did* just hear an inner voice – this is especially common in OCD obsessions – but other times they tell me that no, it was definitely an external voice, totally different from their normal internal voice. Sometimes they thought at first it was a normal non-hallucinatory voice talking to them, and they got up to try to figure out who it was before they realized no one was around and it had to have been a hallucination.

This should not be surprising to anyone who has ever taken drugs, heard from people who took drugs, or been vaguely aware of the existence of drugs. Drugs can cause vivid, realistic hallucinations. Caplan says he doesn't want to talk about neurobiology, and that's all nice and well, but drugs provide a pretty good neurobiological proof of concept. LSD, which is infamous for its hallucinations, is a 5-HT2A agonist. You can treat schizophrenic hallucinations with Seroquel, which is a 5HT2A antagonist; placebo Seroquel doesn't work nearly as well. Coincidence? I feel like at this point we're getting into paranoid are-we-sure-anyone-is-a-libertarian territory again.

He then goes on to say that delusions might just be a preference to believe something, rather than actually believing it. There's a lot of epistemological complexity here – can we believe something just because we prefer to believe it? If someone offered me $1 million to believe that Greenland is in the southern hemisphere, could I do it? I think not, but I think Caplan understands this and is accusing delusional people of just playing a sort of LARP where they act as if they're in a much more interesting universe full of FBI agents and secret radios and the Devil. In favor of this, he describes how psychotic people can sometimes adjust their thinking and actions when they have incentive to do so. For example, he talks about some psychiatric inpatients denying their delusional beliefs in order to avoid electroconvulsive therapy.

I don't think the ability of psychiatric inpatients to hide their condition in response to incentives changes things much. I firmly and genuinely believe that Greenland is in the northern hemisphere, but if someone threatened to give me old-timey scary electroconvulsive therapy for believing this, I would tell them it was however far south they wanted it to be. This doesn't mean my belief about Greenland is insincere, it just means I can think strategically. That even very deeply mentally ill people can think strategically can sometimes be surprising, but no one who has worked with them would deny it can be true.

On the other hand, some of them *can't* think strategically. I remember one patient who was very angry at being involuntarily kept in the hospital who would come up to me every day and start screaming at me that if I didn't let him out, his friends in the highest level of the government were going to revoke my medical license – later this escalated to “kill me”. Spoiler – this is a really bad way to get out of a psychiatric institution. I even told him this was a really bad way to get out of the psychiatric institution and he’d be making his case better by just leaving me alone. He kept finding me and screaming threats about his friends in the government anyway. I can’t tell you for sure what the difference is between people who think strategically and who don’t think strategically – it’s certainly not as simple as “people with illness X are strategic, but people with illness Y aren’t”, but it is certainly a pretty obvious dichotomy.

This patient’s story continues – I put him on antipsychotics, and after two weeks he said he was feeling much better, no longer talked about his friends in the government, and actually thanked me for treating him. I discharged him and as far as I know he’s still taking those medications. If schizophrenia was a preference, this would be strange: he prefers to be schizophrenic, he knows that taking medications will make him less schizophrenic, but he keeps taking them anyway! Since many people like him become schizophrenic, keep taking their antipsychotic medication, and then become better – leaving them much as they were before they became schizophrenic – Caplan’s theory can only theorize that they have a base-level preference for being on antipsychotic drugs. This is a terrible preference to have – such drugs often have bad side effects and make you feel miserable. Surely it makes more sense to believe they have a problem which they don’t like and which the antipsychotics successfully treat?

(yes, there are also many schizophrenic people who don’t voluntarily take their medication. But there are also many people with high blood pressure who don’t take their medication, and antipsychotics are way less pleasant than antihypertensives)

One more thing. Although hallucinations and delusions are the flashiest symptoms of schizophrenia, they are by no means the only ones or even the most important. Many schizophrenics have what’s called “formal thought disorder”, which means their thoughts go in weird directions. A classic example is the tangential person, who will get so distracted they can’t finish a thought. “Tell me how the medication is working?” “Well, I took my medication this morning, after waking up, because I had a bad dream last night, I can’t remember exactly what it was about, I think there was a dog in it, my favorite kind of dog is a Labrador Retriever, I think they’re from Canada, I was in Canada once, it was really cold.” Another kind is the clang, where they connect thoughts based on sound rather than meaning: “I took my medication this morning, it was a warning, a warning of doom, coming at noon with the moon.” This is, as far as I can tell, not something that schizophrenics can successfully “tone down” when asked to do so, based on informal experiments where I ask schizophrenic people to speak normally and tell them that I am more likely to let them out of the hospital if they can form a coherent sentence. This avoids the strategic issues involved in “covering up” hallucinations. Sure, you can always have hallucinations but say you aren’t, but it’s really hard to fake not having a formal thought disorder if you have one, and indeed when a schizophrenic person has a formal thought disorder it’s there to stay until they are treated.

So as far as I can tell schizophrenia includes real hallucinations and delusions, real formal thought issues that the patient usually cannot control, and often the patient is unhappy with it and will willingly take medication to get rid of it. Combined with the neurobiological evidence, the genetic evidence, and the pharmacological evidence, I don’t think calling it “different preferences” is remotely viable.

The Caplan paper is from 2006. I don’t know if he still believes it. And I don’t know if anyone else holds this particular view. But I still meet the occasional Zsazsaian, and general feelings that psychiatric illness isn’t “real illness” are still common. I don’t think it’s a very tenable position and I don’t think this paper does much to support it.

EDIT: I previously wrote in more detail about the difference between “disease” and “normal variation” here.

Against Against Autism Cures

[Content warning: autism, disability, psychiatry, abuse]

I.

Vox: We've called autism a disease for decades. We were wrong.

I have mixed feelings about this. On the one hand, I know and like many people in the autism rights movement. They defend autistic people's right to avoid psychiatric care if they don't want it, and this fits well with my own position that (outside of emergency situations) everyone should get to choose whether or not they want psychiatric care. They also do good work exposing abuse of autistic people and ideas that promote such abuse, and calling out organizations that claim to speak for autistic people but don't do a very good job. And they provide great resources that help autistic people and their friends and families. Ninety-nine percent of what they do is unquestionably great.

Yet somehow, whenever I hear about them in the media, the article is titled either "Autism Is Not A Disease" or "Stop Trying To Cure Autism". I don't know if it's just the usual controversy-mongering or whether this really is at the center of their philosophy, but they manage to consistently emphasize the one percent of their philosophy I can't be on board with.

All psychiatric categories are a mishmash of unlike things crammed together under a single name. Depression ranges from people who put on a normal facade but feel empty inside, all the way to people who are totally catatonic and can't move or speak. Schizophrenia ranges from people who are totally okay as long as they take their medication, all the way to people who talk in “word salad” because their thoughts are so malformed that they can’t even make complete sentences. But even among diagnoses like these, autism takes the cake in terms of heterogeneity.

I kind of a have a front-row seat here. On the one hand, about half my friends, my girlfriend, and my ex-girlfriend all identify as autistic. For that matter, people keep trying to tell me *I’m* autistic. When people say “autistic” in cases like this, they mean “introverted, likes math and trains, some unusual sensory sensitivities, and makes cute hand movements when they get excited.”

On the other hand, I work as a psychiatrist and some of my patients are autistic. Many of these patients are nonverbal. Many of them are violent. Many of them scream *all the time*. Some of them seem to live their entire lives as one big effort to kill or maim themselves which is constantly being thwarted by their caretakers and doctors. I particularly remember one patient who was so desperate to scratch her own face – not in a ‘scratch an itch’ way, but in a ‘I hate myself and want to die’ way – that she had to be kept constantly restrained, and each attempt to take her out of restraints for something as basic as going to the bathroom ended with her attacking the nurse involved. This was one of the worse patients, but by no means unique. A year or so ago, after a particularly bad week when two different nurses had to go to the emergency room, the charge nurse told me in no uncertain terms that the nursing staff was burned out and I was banned from accepting any more autistic patients. This is a nurse who treats homicidal psychopaths and severely psychotic people every day with a smile on her face. When *she* says “autistic”, it seems worlds apart from the “autistic” that means “good at math and makes cute hand flap motions”. When a mental health professional says "autistic", the image that comes to mind is someone restrained in a hospital bed, screaming.

The Vox article doesn't deny any of this. Matthews admits that "people have a image of autistic people as these completely nonverbal children banging their heads against the wall" and his interviewee Silberman describes how autistic children in institutions, when left alone, would "end up chewing through their own finger", which pretty much 100% matches my expectations of what my autistic patients would do if the nurses weren't there to watch them.

But as per Matthews and Silberman, this is just a consequence of the maltreatment these children receive in institutions:

> What society thought of as the natural course of autism was actually a very skewed view of what happened to autistic people when they were put in institutions. For decades, the recommended course of treatment for autism was institutionalization.
>
> Parents were routinely told they should put their child in an institution, quietly remove their photographs from the family albums, never speak of them again, and enlist in decades-long courses of psychoanalysis to think about why they were motivated to wound the developing psyches of their children.
>
> When children were put in institutions for the rest of their lives, it wasn't like they were put in specialized autism wards. There was no such thing, with very few exceptions. They were mostly put on psych wards for adult psychotics. Oliver Sacks worked on such a ward, Ward 23 at Bronx psychiatric, in the 1960s. He told me that some of the children and young adults would be put

> in straight jackets and isolation rooms to sit in their own waste for weeks on end. The children and young adults became self-injurious, which is not a surprise at all. If you treat people brutally, they'll react in extreme fashion.

Be sure to read the rest of the article for more horror stories about the ways autistic children were treated; this sort of stuff was all too common, slightly-toned-down versions of it are still too common today, and this is why I agree with ninety-nine percent of everything the autism movement says.

But then they start talking about how we don't need a cure.

II.

The popular literature about autism tends to fall into the genre of "Doctors hate her! Area mom cured her child's autism with This One Weird Trick!" Common One-Weird-Tricks include gluten-free diets, casein-free diets, massive multivitamin doses, and whatever else the cutting edge of quackery can dream up. The autism rights people are rightly suspicious of this entire category.

But I worry they have their own One-Weird-Trick: treating autistic people decently.

You should treat autistic people decently because it's the right thing to do. But it is not One-Weird-Trick. Avoiding abusive treatment will prevent things from being worse than they have to be, but that's all. It will not turn severely disabled people into independent, happy-go-lucky angels. Parents, caretakers, and doctors can do everything right, move heaven and earth to accomodate an autistic person's every need and limitation – and they will still suffer.

Some of this is purely biological. Thirty percent of autistic people have comorbid epilepsy, often very severe. Over half of autistics are cognitively disabled. Autistics have three times the risk of Tourette's Syndrome, five times the risk of cerebral palsy, about a hundred times the risk of tuberous sclerosis, and various balance and coordination disorders, plus an increased rate of other psychiatric disorders like bipolar and schizophrenia. There are treatments for these conditions, both pharmacological and otherwise, but they come with their own set of side-effects and difficulties and none of them are 100% effective.

And some is behavioral – but as far as I can tell in no way limited to maltreated or institutionalized autistic people. Half of autistic children self-injure, and more than half of autistic children and adolescents are physically aggressive. Autistic children are twenty-eight times more likely to be suicidal than other children. Three-quarters have eating problems ranging from "picky eater" to "will not eat food, good luck doing something about this". About two-thirds have "sleep disorders", which is sometimes a euphemism for "wakes up screaming in the middle of the night and will not stop". As best I can tell, all these studies were done on non-institutionalized autistic people who were generally well-treated and still living with their parents.

And yes, institutionalization adds a whole lot of extra suffering to the mix. But even here, I find Matthews' narrative overly simplistic. He talks of a world where random doctors swoop in and trick parents into sending their children to institutions out of pure prejudice and stigma. Anyone who's ever worked with these families has seen something very different. These parents aren't poor deluded rubes who have been tricked by stigmatizing doctors. They're well-educated, deeply committed to their children – and desperate. They've spent years trying to raise kids who were violent, self-injurious, locked in a sensory hell without the ability to explain their problems verbally, and maybe having seizures all the time to boot. Their decision to institutionalize is a reluctant concession to this reality. I do not feel the *slightest* bit of qualification to pass judgment upon these parents and I invite anybody who does to spend a few moments talking to them. Unless you can give these parents a better option – and trust me, they've looked – institutionalization isn't a moralistic tale of prejudice and stigma in the medical system. It's just a few more drops of misery added to the vast morass of suffering that is severe autism.

Matthews and Silberman speak dismissively of Leo Kanner's view that autism was caused by bad parenting, but I worry we're headed right back in that direction. Start telling people that the only reason autistic people have all these problems is because of institutional abuse, and someone's going to ask "But what about all these autistic people living with their parents who also head-bang/attack people/try to kill themselves?" Then you can either admit that sometimes autism just sucks no matter how good an environment you're in, or blame the parents for not making the environment good enough. And once you've made "autism isn't a disease and nobody needs a cure!" into your rallying cry, it's going to be hard to choose that first option.

III.

If we can't make all autistic people independent and well-adjusted with One Weird Trick, then we have to consider how real autistic people actually turn out. The numbers aren't good.

Outcomes In Adults With Autism Spectrum Disorders: A Historical Perspective reviews all of the autism outcome studies of the past fifty-odd years. Most of it is nitpicking different people's definition of "poor" outcome versus "very poor" outcome, but let me try to extract the easily quantifiable bits[1].

Six studies have assessed what percent of adult autistics have a job – they find 22%, 21%, 31%, 4%, 4%, and 4%. The two that found rates in the twenties limited themselves to high-IQ autistics and so are unrepresentative.

Four studies assessed institutionalization rates among adult autistics, although these "institutions" form a very heterogenous category from homey group houses to super-intense locked hospitals. These studies find 35%, 43%, 48%, and 53% of adult autistics to be instutitionalized.

A few studies looked at other outcomes. Two investigated what percent of adult autistics still lived with their parents. Both estimated about 50%. This is in addition to the 40% or so who are institutionalized, so only about 10% of adult autistics live independently.

One study investigated how many autistics have at least one friend and found it was just under 50%.

I cannot find any studies on adults with autism per se, but adults with Aspergers (recently collapsed into the autism diagnosis) are ten times more likely to be suicidal than other adults.

I realize this seems extreme[2], but I think it really puts into perspective the difference between the conventional "shy person who likes trains" view of autism, and what psychiatrists and scientists really mean when they talk about an autism diagnosis. The happy, independent autistic people whom most of us know and whose stories get told in the media are four to ten percent of the autistic population. What about the other extreme, the forty percent who are institutionalized?

I hate to have to criticize institutions – an umbrella term I'm using to cover group homes, locked facilities, nursing homes, hospitals, etc. Many are run by amazing and caring people who are doing thankless work on shoestring budgets. I'm humbled by the patience and compassion I've seen in their staff of nurses, techs, and other caretakers, and I can't judge them nor claim that I could do their job for one minute.

That having been said, a lot of institutions are kind of hellish.

I've seen institutionalized patients who were sexually abused, physically abused, or just neglected and left to wander out into the street. Lest you dismiss this as crazy people making things up, some of the allegations later got confirmed by police investigations.

Other times it doesn't rise to the level of anything criminal, just the usual petty tyranny you expect any time somebody gets to control somebody else's life. I got consulted for "medical management" of an autistic man who had smashed a bunch of holes into the wall of his group home. He needed his mp3 player to control his noise sensitivities, but the staff had taken it away as punishment for breaking some rule or other. He decided a proportionate response was to smash several holes in the wall. To his credit, it worked; my diagnosis was "give him back his fricking mp3 player, you morons".

But even when the institutions are well-run, non-abusive, and dare-I-say-it-even-nice, the whole structure just makes a perfectly bad fit with autistic people. Remember, autistics are known for intense sensory sensitivities and pickiness about their environment. Take someone who can remain stable *as long as* there are no unexpected loud noises, and make him share a room with a guy who screams at the top of his lungs every couple of minutes for no reason. Or take someone who will eat about 1% of foods, and only if they're perfectly prepared, and then stick him in a hospital where the catering service lets him choose from a menu of two meals, and they're always out of one of them. Or take somebody who freaks out about the feeling of different textiles, and tell her she has to sleep in a mass-produced hospital bed with exactly the same sheets as every other mass-produced hospital bed.

I have some very minor sensory sensitivities, and they drive me nuts. But at least they only metaphorically drive me nuts, That's because I'm an independent middle-class person who can throw money at problems to make them go away. I have no tolerance for stray noise, so I pay a little more than I should for housing and live in a quiet area outside town. I can't stand constricting or scratchy clothing, so I only wear loose-fitting clothes, extend the collars, and cut off all the tags. It works pretty well – but only because I have a lot of control over my life. Put me in an institution where other people manage everything about my daily routine, and my life would fall apart. And my sensitivities are a fraction of a fraction of what real autistic people suffer.

The hospital I work in really does try hard to make things more tolerable for our autistic patients. It's never enough. Fix all of the contigent things, and you just bump up against the fact that humans were not designed to live in psychiatric hospitals, autistic humans least of all.

I have met many well-intentioned people who believe that institutions are only bad because of stigma against the mentally ill, or insufficient budgeting, or ignorance of people’s true needs. These people should visit a nursing home someday. The people there don’t have some kind of exotic stigmatized hard-to-understand condition. They’re just old. Yet nursing homes and other institutions for the elderly have every bit as many problems as the institutions for autistic people.

I don’t think that mental health institutions will get better anytime soon, because I would expect the average person to be a lot more concerned with nursing homes – where their grandparents live! Where they themselves will end up one day! – than they are with mental health institutions. If we can’t even get our act together on that one, what hope do we have for the harder problem? Having good involuntary communal living institutions is just plain beyond us as a civilization at this point. It has nothing to do with stigma or prejudice. Even if everybody loved autistic people exactly as much as they love their own grandmother, the best we could hope for is institutions that treat autistic people as well as they treat grandmothers. Which is to say, abysmally.

V.

Dylan Matthews says that autism “is not a disease”, joining writers from TIME, The Guardian, The Irish Examiner, various blogs, et cetera. I would hate to contradict such an array of eminent voices.

So let’s taboo whether something is a “disease” or not. Let’s talk about suffering.

Autistic people suffer. They suffer because of their sensory sensitivities. They suffer because of self-injury. They suffer because they're in institutions that restrain them or abuse them or just don't let them have mp3 players. Even if none of those things happened at all, they would still suffer because of epilepsy and cerebral palsy and tuberous sclerosis. A worryingly high percent of the autistic people I encounter tend to be screaming, beating their heads against things, attacking nurses, or chewing off their own body parts. Once you're trying to chew off your own body parts, I feel like the question "But is it *really* a disease or not?" sort of loses its oomph.

My moral philosophy doesn't contain a term for "is this a disease or not?", but it definitely contains a term for suffering. If you're a good person, you try to alleviate or prevent suffering. Accomodating and supporting autistic people alleviates some amount of the suffering associated with autism. Curing it alleviates all of that suffering.

And remember – society is fixed but biology is mutable. Which do you think is more likely? That soon biologists will discover a molecular cure for autism? Or that soon politicians will discover a cure for the systemic issues that cause poor people who can't stand up for themselves to be maltreated and abused? The biologists seem to have about a ten million times better track record for this sort of thing. And if you don't expect the politicians to create a brave new world where no disability ever remains unaccomodated, then stopping the biologists just means that the status quo will go on forever.

Faced with the choice of seeing the flood of human misery that I have to deal with every day continue mostly unabated, or having a pill that provides a quick fix to said flood, I wish with all my heart for the latter. Sure, this should not be pursued at the cost of supplying what accomodations to existing autistic people we can, any more than blue sky cure-for-cancer research should be pursued at the cost of treating current cancer patients, but it’s right and proper to want it, to think it would immensely improve thousands of people’s lives.

Would something be lost if autism were banished from the world? Probably. Autistic people have a unique way of looking at things that lets them solve problems differently from everyone else, and we all benefit from that insight. On the other hand, everyone always gives the same example of this: Temple Grandin. Temple Grandin is pretty great. But I am not sure that her existence alone justifies all of the institutionalizations and seizures and head-banging and everything else.

Imagine if a demon offered civilization the following deal: “One in every hundred of your children will be born different. They will feel ordinary sensations as exquisite tortures. Many will never learn to speak; most will never work or have friends or live independently. More than half will consider suicide. Forty percent will be institutionalized, then ceaselessly tyrannized and abused until they die. In exchange, your slaughterhouses will be significantly more efficient.”

I feel like Screwtape would facepalm, then force him into remedial Not-Sounding-Like-An-Obvious-Demon classes.

VI.

My medical ethics have always said that outside emergencies, people who want psychiatric help should be able to get it, and people who don’t want psychiatric help should be able to refuse it.

So when autistic people say they don’t want cures forced on them, I say – fine. If you’re happy with your autism, and it’s not hurting anyone else, keep it.

But when they say we should stop all research into cures so that nobody else can have one either even if they want it, that’s a different story. When they say *that*, well, then *I’m* not the one dictating to other people what neurotypes they are or aren’t allowed to have.

When I see an autistic guy in a hospital room screaming and trying to chew their fingers off, I feel like that guy would probably want a cure for autism. Granted, that guy can’t always talk and tell me what he does or doesn’t want. But he certainly doesn’t seem happy with the status quo. And some autistic people, even some very high-functioning autistic people, have told me straight out that they want to be cured. Who the heck are we to tell them their desires are wrong?

More controversially, I think caretakers who wish there was an autism cure that could relieve them from the responsibility of caretaking have avalid wish. Nobody is entitled to another person’s life. A schizophrenic man has the right to stay off antipsychotics, but his wife has the right to make an ultimatum: “I can’t deal with you being schizophrenic, either you take your meds or I’m leaving.” Likewise, an autistic person has the right to stay autistic, but a caretaker has the right to say “This caretaking is too much for me, either take the cure or find somebody else.”[3]

Even more controversially, I think parents have the right to decide if they want to have an autistic child or not. I am generally pro-choice. As best I can tell fetuses have less personhood than cows, and I had a cheeseburger for dinner last night. If someone wants to abort a fetus because their pregnancy was an accident, because they don't feel ready to have a child, because there's some kind of problem in the family – all of those seem to me like a decision that a mother is perfectly within her rights to make, becauses fetuses are not very important moral agents. And if instead of just "I don't want a child", an expecting mother is worried because she doesn't think she will have the resources or compassion or strength of will to take care of an autistic child, that is also her decision to make. Also, I feel like if your whole argument is that autism only goes badly when autistic children are mistreated, maybe you shouldn't be simultaneously demanding that women who really don't want autistic children and don't believe they can take care of them should be forced to have them anyway. On the other hand, if a mother wanted an autistic child, and she was somehow sure that she could support that child and help him or her flourish, I would be okay with her aborting however many neurotypical fetuses it took.

But even *more* controversially, absent such certainty that your child will flourish I think if some kind of genetic-engineering autism-cure existed, parents would have a moral obligation to use it. Consider an analogy to fetal alcohol syndrome. People with fetal alcohol syndrome seem less happy, less able to achieve their goals, and more likely to suffer than people without the condition. Therefore, we have a very strong social norm that you shouldn't drink too much during pregnancy. But people with autism also seem less happy, less able to acheive their goals, and more likely to suffer – in fact, autism shares a lot of symptoms with fetal alcohol syndrome, like seizures and intellectual disability. The moral imperative to take the hypothetical-genetic-autism-cure during pregnancy seems as clear as the moral imperative not to drink alcohol, and for the same reason[4].

I don’t mean to propose a zero-sum game here. All sorts of breakthroughs and possibilities could give both sides everything they want. Maybe there is some way to alleviate some of the more distressing symptoms of autism while leaving the deeper and more cognitive parts intact. I don’t think anyone wants, as a terminal value, for people to be intellectually disabled or have more seizures. If Dylan Matthews were to list all of the things he likes about autism, and I were to list all of the things I don’t like, maybe there wouldn’t be anything on both lists. And maybe the bad things are controlled by separate genes from the good things, so that we could turn off one set of genes and not the other.

And I’m still in favor of basic income. If everyone gets a basic income, autistic people who want to spend it on getting the care and support they need to remain autistic could do so without costs to anybody else. I do hope that there will be only a very short stretch of time between the development of technology that can genetically engineer people at will, and solving the scarcity problems that make our desires sometimes demand sacrifices of others.

And maybe the best option is that somebody develops technology that can change your brain type during adulthood, so that everyone has the option of experiencing life as an autistic person and as a neurotypical person and seeing which is best for them. Invent something like that and the problem disappears.

And a lot depends on the genetic structure of autism. If autism is just the brain’s response to high mutational load, then trying to do most other good things will prevent autism as a side effect. On the other hand, if autism is the extreme version of a trait whose common version is “good at logical and scientific thinking”, then we have to ask ourselves whether that trait is worth messing with. As usual, everything about genetic engineering raises thorny scientific and ethical quandaries, and I can only hope we don’t drag our feet in creating the eight-foot-tall IQ 300 supermen who can solve them.

And maybe all of this is silly, because we don’t have an autism cure and we’re not even very close to one, and maybe when people talk about whether we should use the autism cure we don’t have, it’s all a metaphor for “respect my tribe” and “I demand my rights” and “here’s a good controversial lead-in to a story about how you should treat autistic people like human beings” and maybe even if I’m right about the literal moral dilemma I should support the opposite side for totally symbolic reasons.

But if, after all this, it really does comes down to a binary “cure autism/don’t cure autism” decision, I know which side I’m on.

Footnotes

1. In a way, all of these numbers are meaningless. If you define “autism” so broadly that it includes Dylan Matthews and me, maybe less than one percent of autistics are in institutions. If you define “autism” so strictly that you only count institutionalized autistics, then one hundred percent of them are. So really what these studies are saying is “We chose to define autism at a level where forty percent of autistics are in institutions”, to which one response is “And why should I care what level you chose to define autism at?” But I think these studies at least tell us two things. First, that the formal psychiatric definition of autism is much stricter than the popular one, and we combine the two categories at our peril. And second, that this former category of very severly ill autistics exists and contains many people we might not otherwise notice.

2. The employment and living-with-parent numbers don’t seem to have changed much over time; the institutionalization number has decreased over the past few decades. This probably reflects a general trend toward deinstitutionalization throughout the mental health world. So far there is little sign of autistic people doing great now that the most abusive 60s-style institutions have been curtailed. But perhaps this just hasn’t shown up yet; these are studies of adult autistics; even the youngest were born in the 1980s, and the autism world has changed a lot since then. I'd say that I will be interested to see a similar study in 2030, but I think by that time the diagnostic category will have become so confusing that nothing will be comparable to anything that came before it.

3. Even more controversially, I think the government has the right to do the same. The lifetime cost of supporting an autistic person is $1.5 million, not including productivity loss of the person themselves. Much of this is borne by the government. The average taxpayer will give about $500,000 over their lives, so it takes three non-autistic people to support each autistic person – even ignoring all other essential government services like schooling and welfare and giant nuclear missiles. Of course, helping the needy is exactly the sort of thing the government *should* be doing, and right now the government ought to do whatever it takes. But *if* there is a cure for autism and autistic people choose not to take it but still want the $1.5 million, *then* the government has the right to start thinking things like "we could lift a whole lot of destitute families out of poverty for $1.5 million." Everyone has the right to choose their neurotype, but I'm not sure they have the right to make other people subsidize it.

4. Prof. Mora LeQuivalence: "Interesting point! Let me propose a similar argument that you might have a more personal stake in. Why don't we genetically engineer away nerdiness? Everyone knows nerds are less happy than normals, at least in high school."

Scott: "The cost-benefit calculations are totally different. Nerds are probably unhappy, but less so than autistics. And while autistics can claim Temple Grandin, nerds can claim pretty much every decent scientist and mathematician of the past few centuries, plus all the good sci-fi/fantasy writers."

Mora: "Fine then. Keep the nerds around until they build robots that can do math and science and art. Then we won't need them anymore and we can get rid of them."

Scott: "Look, we can't do all of this based on broad principles. There's a balance between having a wide variety of human experiences versus having some of them be hellish. Few people want all individuality collapsed into a planet-sized chunk of hedonium, and few people think that a couple of expecting mothers should drink lots of booze to protect the vital diverse human experience of fetal alcohol syndrome. You've got to make your own choice about where to draw the line, and I draw it somewhere south of nerds and north of severe autism."

Mora: "That's a very interesting argument."

Scott: "It better be, I stole it directly from God."

It Was You Who Made My Blue Eyes Blue

[Content note: suicide]

Day Zero

It all started with an ignorant white guy.

His name was Alonzo de Pinzon, and he'd been shipwrecked. We heard him yelling for help on the rocks and dragged him in, even though the storm was starting to get really bad. He said that his galleon had gone down, he'd hung on to an oar and was the only survivor. Now he was sitting in our little hunting lodge, shivering and chattering his teeth and asking us questions in the Polynesian traders' argot which was the only language we all shared.

"How big is this island? How many of you are there?"

Daho answered first. "11.8 miles from the easternmost point to the westernmost point, 3.6 miles from the northernmost to the southernmost. Total area is 14.6 square miles, total coastline is dependent on how deeply you want to go into the fractal nature of the perimeter but under some reasonable assumptions about 32 miles long. Last census said there were 906 people, but that was two years ago, so assuming the 5.1% rate of population growth continues, there should be closer to 1000 now. Everyone else is back at the village, though. The five of us were out hunting and got caught in the storm. We figured we'd stay at this old hunting lodge until it cleared up, since it's 5.5 miles back to the village and given the terrain and factoring in a delay because of the storm it would probably take at least 9.5 hours to get back."

Pinzon blinked.

“Problem?” asked Daho.

“But – ” he said. “That is the sort of answer I should expect from a natural philosopher. Not from a savage.”

“Savage?” Calkas hissed. “Really? We rescue you, and the first thing you do is call us savages?”

The sailor looked around, as if anxious. Finally, almost conspiratorially: “But I heard about your island! I heard you eat people!”

Calkas smiled. “Only as a deterrent. Most of the time when European explorers land somewhere, they kill all the men and enslave all the women and convert the children to Christianity. The only places that escape are the ones that get a reputation for eating said European explorers. So we arranged to give ourselves that reputation.”

“And then we had to go through with it a few times in order to make the deterrent credible,” added Bekka, my betrothed. “And you guys do taste really good with ketchup.”

“It’s a savage thing to do!” Pinzon said “And you even look like savages. You wear bones in your hair”

“Just Enuli,” I said. “She’s going through a Goth phase.”

“My name is Morticia now,” said Enuli, “and it’s *not a phase*!” She did have a bone in her hair. She also had white face paint and black eyeliner.

“More roast pig?” Bekka asked Pinzon. The sailor nodded, and she re-filled his plate.

"I just don't get it," he told us. "Everyone else in this part of the world lives in thatched huts and counts 'one, two, many'. We tried to trade with the Tahitians, and they didn't understand the concept of money! It was a mess!"

Bekka rolled her eyes at me, and I smiled. Calkas was a little more tolerant. "The sacred plant of our people is called sparkroot," he said. "When we eat it, we get – more awake, I guess you could say. We try to have some every day, and it helps us keep track of things like the island size and the population, and much more."

Alonzo de Pinzon looked interested. "How come you haven't done more with your intellect? Invented galleons, like we Spaniards? Set off to colonize Tahiti or the other islands? If you are as smart as you seem, you could conquer them and take their riches."

"Maybe," said Calkas. "But that's not why the Volcano God gave us the sparkroot. He gave us sparkroot to help us comply with his complicated ritual laws."

"You need to be smart to deal with your ritual laws?"

"Oh yes. For example, the Tablets of Enku say that we must count the number of days since Enku The Lawgiver first spoke to the Volcano God, and on days whose number is a Mersenne prime we can't eat any green vegetables."

"What's a Mersenne prime?" asked the sailor.

"Exactly my point," said Calkas, smiling.

"That's not even the worst of it!" Daho added. "The Tablets say we have to bathe in the waterfall any day x such that a^n + b^n = x^n where n is greater than two. We got all confused by that one for a while, until Kaluhani gorged himself on a whole week’s worth of sparkroot in one night and proved that it would never apply to any day at all.”

“The Volcano God’s yoke is light,” Calkas agreed.

“Although poor Kaluhani was vomiting for the next three days after that,” Bekka reminded us, and everybody laughed remembering.

“Oh!” said Daho. “And remember that time when Uhuako was trying to tattoo everyone who didn’t tattoo themselves, and he couldn’t figure out whether he had to tattoo himself or not, so he ended up eating a whole sparkroot plant at once and inventing advanced set theory? That was hilarious.”

Everyone except Alonzo de Pinzon giggled.

“Point is,” said Calkas, “that’s why the Volcano God gives us sparkroot. To follow the rituals right. Any other use is taboo. And I’m okay with that. You Europeans may have your big ships and your guns and your colonies across half the world. And you might think you’re smart. But you guys couldn’t follow the Volcano God’s rituals right for a *day* without your brains exploding.”

Pinzon scowled. “You know what?” he said. “I don’t think you’re Polynesians at all. I think you must be descended from Europeans. Maybe some galleon crashed on this island centuries ago, and you’re the descendants. That would explain why you’re so smart.”

“You know what else we’ve invented with our giant brains?” Bekka asked. “Not being racist.”

“It’s not racism!” said Pinzon. “Look, there’s one more obvious reason to think you’re descended from Europeans.

You may have dark skin, but this is the first place I’ve been in all of Polynesia where I’ve seen even one native with blue eyes.”

Bekka gasped. Calkas’ eyes went wide. Daho’s hands started curling into fists. Enuli started to sob.

I looked at them. They looked at me. Then, as if synchronized, we grabbed Alonzo de Pinzon and crushed his throat and held him down until he stopped breathing.

He tasted delicious with ketchup.

Day One

The next morning dawned, still grey and cold and stormy.

“So,” I said when the other four had awoken. “I guess we’re all still here.”

I said it glumly. It wasn’t that I wanted any of my friends to commit suicide. But if one of them had, the horror would have stopped there. Of course, I knew it couldn’t really be over that easily. But I couldn’t have admitted I knew. I couldn’t even have suggested it. That would have made me as bad as the Spanish sailor.

“Wait,” said Enuli. “I don’t get it. Why wouldn’t we still be here?”

The other four stared at her like she was mad.

“Enuli,” Calkas suggested, “did you forget your sparkroot last night?”

“First of all, my name is Morticia. And – ”

“Shut it. Did you forget your sparkroot?”

Finally she nodded bashfully. “I was so upset about that awful man making fun of my hair-bone,” she said. “I guess it slipped my mind. I’ll have some now.” She took some raw sparkroot from our bag, started to crush it with the mortar and pestle. “In the meantime, tell me what’s going on.”

“Alonzo de Pinzon said at least one of us had blue eyes. We all know what the Tablets of Enku say. If anybody has blue eyes, and knows that they have blue eyes, they must kill themselves.”

“So what? I see people with blue eyes all the time. Of course at least one of us has blue eyes.”

Concerned looks from the others. I reflected for a second, the sparkroot smoothing the thoughts’ paths through my brain. No, she hadn’t revealed anything extra by saying that, although she would have if she had said it before the sailor had spoken, or last night before we woke up this morning. She hadn’t made the problem *worse*. Still, it had been a slip. This was the sort of thing that made forgetting your sparkroot so dangerous. Had it been a different time, even Enuli’s comment could have doomed us all.

“It’s like this,” I told Enuli. “Suppose there were only the two of us, and we both had blue eyes. Of course, you could see me and know that I had blue eyes. So you would know that at least one of us had blue eyes. But what you wouldn’t know is that I also knew it. Because as far as you know, you might have eyes of some other color, let’s say brown eyes. If you had brown eyes, and I of course don’t know my own eye color, then I would still think it possible that both of us have brown eyes. So if I in fact know for sure that at least one of us has blue eyes, that means you have blue eyes. So you know at least one of us has blue eyes, but you don’t know that I know it. But if Alonzo de Pinzon shows up and says that at least one of us has blue eyes, now you know that I know it.”

“So?” Enuli poured the ground-up root into a cup of boiling water.

“So the Tablets say that if anyone knows their own eye color, they must commit suicide at midnight of that night. Given that I know at least one of us has blue eyes, if I see you have brown eyes, then I know my own eye color – I must be the blue-eyed one. So the next morning, when you wake up at see me not dead, you know that you don’t have brown eyes. That means you must be the blue-eyed one. And that means you have to kill yourself on midnight of the following night. By similar logic, so do I.”

Enuli downed her sparkroot tea, and then her eyes lit up. “Oh, of course,” she said. Then “Wait! If we follow the situation to its logical conclusion, any group of n blue-eyed people who learn that at least one of them has blue eyes have to kill themselves on the nth night after learning that!”

We all nodded. Enuli’s face fell.

“I don’t know about the rest of you,” said Daho, “but I’m not just going to sit around and wait to see if I die.” There were murmurs of agreement.

I looked out at my friends. Four pairs of blue eyes stared back at me. Everybody else either saw four pairs of blue eyes or three pairs of blue eyes, depending on what color my own eyes were. Of course, I couldn’t say so aloud; that would speed up the process and cost us precious time. But I knew. And they knew. And I knew they knew. And they knew I knew I knew. Although they didn’t know I knew they knew I knew. I think.

Then I looked at Bekka. Her big blue eyes stared back at me. There was still hope I was going to survive this. My betrothed, on the other hand, was absolutely doomed.

“This sucks,” I agreed. “We’ve got to come up with some kind of plan. Maybe – Enuli wasn’t thinking straight yesterday. So her not committing suicide doesn’t count. Can we work with that?”

“No,” said Calkas. “Suppose Enuli was the only one with blue eyes, and all the rest of us had brown eyes. Then she would realize that and commit suicide tonight. If she doesn’t commit suicide tonight, then we’re still screwed.”

“Um,” said Daho. “I hate to say this, but we get rid of Enuli. There’s a canoe a little ways down the beach hidden underneath the rocks. She can set off and row for Tahiti. We’ll never know if she killed herself tonight or not. Remember, right now for all we know Enuli might be the only one with blue eyes. So if there’s any question in our mind about whether she killed herself, we can’t be sure that the rest of us aren’t all brown-eyed.”

We all thought about that for a moment.

“I’m not going to row to Tahiti,” said Enuli. “In this storm, that would be suicide.”

The rest of us glared at her.

“If you don’t get off this island, then for all we know all five of us are going to have to die,” I said. “You included.”

“Well Ahuja, if you’re so big on making sacrifice why don’t *you* go to Tahiti?”

“First of all,” I said, “because I’m not leaving my betrothed. Second of all, because it doesn’t work for me. I knew what was going on last night. We already know that I’m not the only blue-eyed person here. And we know we know it, and know we know we know it, and so on. You’re the only one who can help us.”

“Yeah?” said Enuli. “Well, if two of you guys were to row to Tahiti, that would solve the problem too.”

“Yes,” said Daho patiently. “But then two of us would be stuck in exile. If you did it, only one of us would be stuck.”

Enuli gave a wicked grin. “You know what?” she said. “I’ll say it. I’m not the only blue-eyed person here. At least one of the rest of you has blue eyes.”

And there it was.

“Ha. Now I’m no worse off than any of the rest of you.”

“Kill her,” said Bekka. “She broke the taboo.” The rest of us nodded.

“So she did,” said Calkas. “And if we had a court here, led by the high priest, and an executioner’s blade made to exactly the right standard, kill her we would. But until those things happen, it is taboo for us to convict and kill her without trial.”

Calkas’ father was the high priest. He knew the law better than any of us. The five of us sat quietly and thought about it. Then he spoke again:

“But her soul may well burn in the caldera of the Volcano God forever.”

Enuli started to cry.

“And,” Calkas continued, “there is nevertheless a flaw in our plan. For all we know, three out of five of us have brown eyes. We cannot tell the people who have blue eyes that they have blue eyes without breaking the taboo. So we cannot force blue-eyed people in particular to sail to Tahiti. But if two of the brown-eyed people sail to Tahiti, then we do not lose any information; we know that they would not have committed suicide, because they could not have figured out their own eye color. So sailing to Tahiti won’t help.”

The rest of us nodded. Calkas was right.

“Let’s wait until dinner tonight,” I suggested. “We’ll all have some more sparkroot, and maybe we’ll be able to think about the problem a little more clearly.”

Day Two

The sun rose behind angry storm clouds. The five of us rose with it.

“Well, I guess we’re all still here,” I said, turning the morning headcount into a grim tradition.

“Look,” said Bekka. “The thing about sailing to Tahiti would work a lot better if we knew how many blue-eyed versus brown-eyed people were here. If we all had blue eyes, then we could be sure that the Tahiti plan would work, and some of us could be saved. If some of us had brown eyes, then we could choose a number of people to sail to Tahiti that had a good probability of catching enough of the blue-eyed ones.”

“We can wish all we want,” said Enuli, “but if we explicitly knew how many people had blue versus brown eyes, we’d all have to kill ourselves right now.”

“What about probabilistic knowledge?” I asked. “In theory, we could construct a system that would allow us to have > 99.99% probability what color our eyes were without being sure.”

“That’s stupid,” Enuli said, at precisely the same time Calkas said “That’s brilliant!” He went on: “Look, just between the five of us, everybody else back at the village has blue eyes, right?”

We nodded. It was nerve-wracking to hear it mentioned so casually, just like that, but as far as I could tell it didn’t break any taboos.

“So,” said Calkas, “We know that, of the island population, at least 995 of the 1000 of us have blue eyes. Oh, and since nobody committed suicide last night, we know that at least three of the five of us have blue eyes, so that’s 998 out of 1000. Just probabilistically, by Laplace’s Law of Succession and the like, we can estimate a >99% chance that we ourselves have blue eyes. Nothing I’m saying is taboo. It’s nothing that the priests don’t know themselves. But none of them have killed themselves yet. So without revealing any information about the eye color composition of the current group, I think it’s reasonable to make a first assumption that all of us have blue eyes.”

“I’m really creeped out at you talking like this,” said Daho. I saw goosebumps on his arms.

“I do not believe that the same Volcano God who has endowed us with reason and intellect could have intended us to forego their use,” said Calkas. “Let’s assume we all have blue eyes. In that case, the Tahiti plan is still on.”

“Waaiiiiit a second – ” Bekka objected. “If probabilistic knowledge of eye color doesn’t count, then no information can count. After all, there’s always a chance that the delicious sailor could have been lying. So when he said at least one of us had blue eyes, all we know is that there’s a high *probability* that at least one of us has blue eyes.”

“Yes!” said Daho. “I’ve been reading this book that washed ashore from a shipwrecked galleon. Off in Europe, there is this tribe called the Jews. Their holy book says that illegitimate children should be shunned by the congregation. Their leaders thought this was unfair, but they weren’t able to contradict the holy book. So instead they declared that sure, illegitimate children should be shunned, but only if they were *sure* they were really illegitimate. Then they declared that no amount of evidence would ever suffice to convince them of that. There was always a possibility that the woman had secretly had sex with her husband nine months before the birth and was simply lying about it. Or, if apparently unmarried, that she had secretly married someone. They decided that it was permissible to err on the side of caution, and from that perspective nobody was sufficiently certainly illegitimate to need shunning. We could do the same thing here.”

“Yes!” I said. “That is, even if we looked at our reflection and saw our eye color directly, it might be that a deceiving demon is altering all of our experience – ”

“No no NO,” said Calkas. “That’s not right. The Tablets of Enku say that *because* people must not know their own eye color, we are forbidden to talk about the matter. So the law strongly implies that hearing someone tell us our eye color would count as proof of that eye color. The exact probability has nothing to do with it. It’s the method by which we gain the information.”

“That’s stupid,” Bekka protested.

“That’s the law,” said Calkas.

“Let’s do the Tahiti plan, then,” I said. I gathered five stones from the floor of the lodge. Two white, three black. “White stones stay. Black stones go to Tahiti. Close your eyes and don’t look.”

Bekka, Calkas, Daho, and Enuli all took a stone from my hand. I looked at the one that was left. It was black. Then I looked around the lodge. Calkas and Enuli were smiling, white stones in their hands. Bekka and Daho, not so much. Daho whined, looked at me pleadingly.

“No,” I said. “It’s decided. The three of us will head off tonight.”

Calkas and Enuli tried to be respectful, to hide their glee and relief.

“You guys will tell our families what happened?

They nodded gravely.

We began packing our things.

* * *

The dark clouds frustrated any hope of moonlight as Bekka, Daho and I set off to the nearby cove where two canoes lay hidden beneath the overhanging rocks. The rain soaked our clothes the second we crossed the doorway. The wind lashed at our faces. We could barely hear ourselves talk. This was a *bad* storm.

“How are we going to make it to the canoes in this weather?!” Bekka shouted at me, grabbing my arm. I just squeezed her hand. Daho might have said something, might not have. I couldn’t tell.Between the mud and the rain and the darkness it took us two hours to travel less than a mile. The canoes were where we had left them a few days before. The rocks gave us brief shelter from the pelting rain.

“This is suicide!” Daho said, once we could hear each other again. “There’s no way we can make it to Tahiti in this! We won’t even be able to make it a full mile out!” Bekka nodded.

“Yes,” I said. I’d kind of known it, the whole way down to the cove, but now I was sure. “Yes. This is suicide. But we’ve got to do it If we don’t kill ourselves tonight, then we’ve just got to go back to the lodge. And then we’ll all end up killing ourselves anyway. And Calkas and Enuli will die too.”

“No!” said Daho. “We go back, we tell them that we can’t make it to Tahiti. Then we let *them* decide if we need to commit suicide or not. And if they say yes, we draw the stones again. Four black, one white. One chance to live.”

“We already drew the stones,” I said. “Fair is fair.”

“Fair is fair?” Bekka cried. “We drew stones to go to Tahiti. We didn’t draw stones to commit suicide. If the stone drawing obliged us to commit suicide, they should have said so, and then maybe we would have spent more time thinking about other options. Why do we have to die? Why can’t the other ones die? Why not Enuli, with that stupid bone in her hair? I hate her so much! Ahuja, you can’t just let me die like this!”

That hurt. I was willing to sacrifice my life, if that was what it took. But Bekka was right. To just toss ourselves out to sea and let her drown beneath those waves would break the whole point of our betrothal bond.

“Well, I – ”

“Ahuja,” said Bekka. “I think I’m pregnant.”

“What?”

“I missed my last period. And I got sick this morning, even though I didn’t eat any extra sparkroot. I think I’m pregnant. I don’t want to die. We need to save me. To save the baby.”

I looked at the horrible waves, watched them pelt the shore. A few moments in that, and there was no doubt we would capsize and die.

“Okay,” I said. “New plan. The three of us go back. We tell them that we couldn’t get to Tahiti. They point out that another night has passed. Now four of us have to die. The three of us vote for everybody except Bekka dying. It’s 3-2, we win. The rest of us die, and Bekka goes back to the village and the baby lives.”

“Hold on,” said Daho. “I’m supposed to vote for me to die and Bekka to live? What do I get out of this deal?”

The Tablets of Enku say one man must not kill another. So I didn’t.

“You get an extra day!” I snapped. “One extra day of life for saving my betrothed and unborn child. Because we’re not going back unless you agree to this. It’s either die now, or die tomorrow night. And a lot of things can happen in a day.”

“Like what?”

“Like I don’t know. We might think of some clever way out. Enku the Lawgiver might return from the dead and change the rules. Whatever. It’s a better deal than you’ll get if you throw yourself into that water.”

Daho glared at me, then weighed his options. “Okay,” he snapped. “I’ll vote for Bekka. But you had better be thinking *really* hard about those clever ways out.”

Day Three

“So,” said Calkas the next morning. “I guess all of us are still here.” He didn’t really sound surprised.

I explained what had happened the night before.

“It’s simple,” Calkas declared. “The Volcano God is punishing us. He’s saying that it’s wrong of us to try to escape his judgment by going to Tahiti. That’s why he sent the storm. He wants us all to stay here until the bitter end and then, if we have to, we die together.”

“No!” I protested. “That’s not it at all! The taboo doesn’t say we all have to die. It just says we all have to die if we figure out what our eye color is! If some of us kill ourselves, we can prevent that from happening!”

“The Volcano God loathes the needless taking of life,” said Calkas. “And he loathes his people traveling to other lands, where the sparkroot never grows and the taboos are violated every day. That’s what he’s trying to tell us. He’s trying to close off our options, so that we stay pure and our souls don’t have to burn in his caldera. You know, like Enuli’s will.” He shot her a poison glance.

“My name is – ” she started.

“I don’t think that’s it at all,” I said. “I say the four of us sacrifice ourselves to save Bekka.”

“You *would* say that, as her betrothed,” said Enuli.

“Well yes,” I said. “Yes, I would. Forgive me for not wanting the love of my life to die for a stupid reason. Maybe I should just throw myself in the caldera right now. And she’s carrying an unborn child? Did you miss that part?”

“People, people,” said Calkas. “Peace! We’re all on the same side here.”

“No we’re not,” I said. “So let’s vote. Everyone in favor of saving Bekka, say aye.”

“And everyone in favor of not sacrificing anyone to the waves, and letting the Volcano God’s will be done, say nay.” Calkas added.

“Aye,” I said.

“Aye,” said Bekka.

“Nay,” said Calkas.

“Nay,” said Enuli.

“Nay,” said Daho.

“What?!” I protested.

“Nay,” Daho repeated.

“But you said – ” I told him.

“You promised me one extra day,” Daho said. “Think about it. Calkas is promising me two.”

“No!” I protested. “You can’t do this! Seriously, I’ll kill you guys if I have to!”

“Then your soul will burn in the caldera forever,” said Calkas. “And it still won’t help your betrothed or your child.”

“You can’t do this,” I repeated, softly, more of a mutter.

“We can, Ahuja” said Calkas.

I slumped back into my room, defeated.

Day Four

I gave them the traditional morning greeting. “So, I guess we’re all still here.”

We were. It was our last day. We now had enough information to prove, beyond a shadow of a doubt, that all of us had blue eyes. At midnight, we would all have to commit suicide.

“You know what?” said Enuli. “I’ve always wanted to say this. ALL OF YOU GUYS HAVE BLUE EYES! DEAL WITH IT!”

We nodded. “You have blue eyes too, Enuli,” said Daho. It didn’t matter at this point.

“Wait,” said Bekka. “No! I’ve got it! Heterochromia!”

“Hetero-what?” I asked.

“Heterochromia iridum. It’s a very rare condition where someone has two eyes of two different colors. If one of us has heterochromia iridum, then we can’t prove anything at all! The sailor just said that he saw someone with blue eyes. He didn’t say how *many* blue eyes.”

“That’s stupid, Bekka,” Enuli protested. “He said blue eyes, plural. If somebody just had one blue eye, obviously he would have remarked on that first. Something like ‘this is the only island I’ve been to where people’s eyes have different colors.'”

“No,” said Bekka. “Because maybe all of us have blue eyes, except one person who has heterochromia iridum, and he noticed the other four people, but he didn’t look closely enough to notice the heterochromia iridum in the fifth.”

“Enuli just said,” said Calkas, “that we all have blue eyes.”

“But she didn’t say how many!”

“But,” said Calkas, “if one of us actually had heterochromia iridum, don’t you think somebody would have thought to mention it before the fifth day?”

“Doesn’t matter!” Bekka insisted. “It’s just probabilistic certainty.”

“It doesn’t work that way,” said Calkas. He put an arm on her shoulder. She angrily swatted it off. “Who even decides these things!” she asked. “Why is it wrong to know your own eye color?”

“The eye is the organ that sees,” said Calkas. “It’s how we know what things look like. If the eye knew what it itself looked like, it would be an infinite cycle, the eye seeing the eye seeing the eye seeing the eye and so on. Like dividing by zero. It’s an abomination. That’s why the Volcano God, in his infinite wisdom, said that it must not be.”

“Well, I know my eyes are blue,” said Bekka. “And I don’t feel like I’m stuck in an infinite loop, or like I’m an abomination.”

“That’s because,” Calkas said patiently, “the Volcano God, in his infinite mercy, has given us one day to settle our worldly affairs. But at midnight tonight, we all have to kill ourselves. That’s the rule.”

Bekka cried in my arms. I glared at Calkas. He shrugged. Daho and Enuli went off together – I guess they figured if it was their last day in the world, they might as well have some fun – and I took Bekka back to our room.

* * *

“Listen,” I said. “I’m not going to do it.”

“What?” she asked. She stopped crying immediately.

“I’m not going to do it. And you don’t have to do it either. You should have your baby, and he should have a mother and father. We can wait here. The others will kill themselves. Then we’ll go back to the village on our own and say that the rest of them died in the storm.”

“But – aren’t you worried about the Volcano God burning our souls in his caldera forever?”

“To be honest, I never really paid much attention in Volcano Church. I – I guess we’ll see what happens later on, when we die. The important thing is that we can have our child, and he can grow up with us.”

“I love you,” said Bekka.

“I know,” I said.

“I know you know,” she said. “But I didn’t know that you knew I knew you knew. And now I do.”

“I love you too,” I said.

“I know,” she said.

“I know you know,” I said. I kissed her. “I love you and your beautiful blue eyes.”

The storm darkened from gray to black as the hidden sun passed below the horizon.

Day Five

“So,” I said when the other four had woken up, “I guess all of us are atheists.”

“Yeah,” said Daho.

“The world is empty and void of light and meaning,” said Enuli. “It’s the most Goth thing of all.”

Calkas sighed. “I was hoping all of you would kill yourselves,” he said, “and then I could go home, and my father the high priest would never have to know what happened. I’m sorry for pushing the rest of you. It’s just that – if I looked lax, even for a second, he would have suspected, and then I would have been in so much trouble that an eternity in the Volcano God’s caldera would look pretty good compared to what would happen when I got back home.”

“I think,” said Bekka, “that I realized it the first time I ate the sparkroot. Before I’d even finished swallowing it, I was like, wait a second, volcanoes are probably just geologic phenomenon caused by an upwelling of the magma in the Earth’s mantle. And human life probably evolved from primitive replicators. It makes a lot more sense than some spirit creating all life and then retreating to a dormant volcano on some random island in the middle of the nowhere.”

“This is great,” said Bekka. “Now even if it’s a Mersenne prime day I can eat as many green vegetables as I want!”

“You know Mersenne prime days only come like once every couple of centuries, right?” I asked her.

“I know. It’s just the principle of the thing.”

“We can’t tell any of the others,” Daho insisted. “They’d throw us into the volcano.”

“You think?” I said. “Calkas was saying before that 99% of us had blue eyes, so probably we all had blue eyes. Well, think about it. The five of us are a pretty random sample of the island population, and all five of us are atheist. That means there’s probably a lot more. Maybe everybody’s atheist.”

“Everybody?”

“Well, I thought Calkas was like the most religious of anybody I knew. And here we are.”

“I told you, I was just trying to behave so that I didn’t get in trouble with my father.”

“What if everyone’s doing that? Nobody wants to get in trouble by admitting they don’t believe, because if anybody else found out, they’d get thrown into the volcano. So we all just put on a mask for everybody else.”

“I figured Ahuja was atheist,” said Bekka.

“You did?!” I asked her.

“Yeah. It was the little things. When we were hanging out. Sometimes you’d forget some rituals. And then you’d always shoot these guilty glances at me, like you were trying to see if I'd noticed. I thought it was cute."

"Why didn't you tell me?"

"You'd have freaked out. You'd have had to angrily deny it. Unless you knew I was atheist. But I couldn't have told you that, because if I did then you might feel like you had to throw *me* in the volcano to keep up appearances."

"Bekka!" I said. "You know I would never – "

"I kind of suspected Calkas was atheist," said Daho. "He got so worked up about some of those little points of law. It had to be overcompensating."

"Hold on hold on hold on!" said Calkas. "So basically, we were all atheists. We all knew we were all atheists. We just didn't know that we knew that we were all atheists. This is hurting my brain. I think I'm going to need more sparkroot."

A sunbeam peeked through the wall of the lodge.

"Storm's over!" Bekka shouted gleefully. "Time to go back home!" We gathered our things and went outside. The sudden sunlight felt crisp and warm upon my skin.

"So," said Daho, "we don't mention anything about the sailor to anyone else back at the village?"

"Are you kidding?" said Calkas. "I say we stand in the middle of town square, announce everybody's eye colors, and then suggest that maybe they don't believe in the Volcano God as much as they thought. See what happens."

"YOU ALL HAVE BLUE EYES!" Enuli shouted at the jungle around us. "DEAL WITH IT!" We laughed.

“By the way,” I told Enuli. “While we’re airing out things that everybody knows in order to make them common knowledge, that bone in your hair looks ridiculous.”

“He’s right,” Daho told her.

“It really does,” Calkas agreed.

“You watch out,” said Enuli. “Now that we don’t have to reserve the sparkroot for interpreting taboos, I’m going to invent a death ray. Then you’ll be sorry.”

“Hey,” said Daho, “that sounds pretty cool. And I can invent a giant aerial dreadnaught to mount it on, and together we can take over Europe and maybe the next sailor who gets shipwrecked on our island will be a little less condescending.”

“Ha!” said Enuli. “That would be so Goth.”

Sun on our backs, we took the winding road into the village.

OT 30: Comment Knowledge

This is the semimonthly open thread. Post about anything you want, ask random questions, whatever. Also:

1. These threads tend to fill up pretty quickly. Should I start doing them weekly instead of biweekly?

2. John Sidles is hereby forbidden to use bold text or to speak in a topic-comment sentence structure. He may continue to comment as long as he follows these two rules.

3. Comment thread on autism had a lot of unfortunate definitional issues (are we talking about turning everyone into Perfectly Conformist Jocks/Cheerleaders or about preventing people from being institutionalized/suicidal while leaving everything else intact?) but the stories from autistic people / caretakers were pretty interesting, especially Mai, Helldalgo, Alicorn, Ilzolende, Peter, Murphy, and seebs.

Contra Simler on Prestige

su3su2u1 challenged status/signaling theories of human behavior: can they make any real-life predictions? His example was a recent medical conference that threw together three groups of people – high-status top professors, medium-status established doctors, and low-status new residents. The women in one group (female doctors + male doctors' wives/girlfriends) were wearing conspicuous fancy jewelery. The women in the other groups weren't. Which group had the jewelery?

His point was that status/signaling theories don't answer this question for us with any degree of confidence. Maybe the high-status top professors wear the jewelery to signal wealth and dominance. Maybe the low-status new residents wear it aspirationally and because they need to impress. Maybe the medium-status established doctors wear it, because the residents can't afford it and the professors countersignal that they don't need it.

Now, in fact su3su2u1 was a no-good sneaky sneak, because the residents had all just attended a wedding that gave out the fancy jewelery as gifts and this was probably all that was going on. But his point is well-taken. Status and signaling theories are hard to use in practice. So it's always nice when people try to do some theoretical work on them and tease them apart into their different components. This is the task Kevin Simler takes on in Social Status: Down The Rabbit Hole.

His theory (which he adopts from various psychologists and animal behaviorists) is that status separates neatly into two systems: dominance and prestige. Dominance is "respect me because I'll kill you if you don't." Prestige is "Respect me because I’m awesome”. The two systems have different origins and different behavioral effects; conflate the two and you’ll end up very confused.

If you hate your boss, but you do what she says anyway because she’ll fire you if you don’t, that’s dominance. If you’re very respectful to a police officer because he has a gun and you don’t, that’s dominance too. Principals have dominance, parents have dominance, psychiatrists keeping you in a hospital against your will have dominance. Prestige is different. A rock star has prestige. He can’t hurt you. You don’t necessarily need anything from him. But you still want his autograph, want to meet him, maybe want to sleep with him. Star athletes have prestige. Actors and actresses. Good bosses who you work hard for not because you’re afraid of them but because you don’t want to let them down. Your parents, if you do what they say out of respect/love and not out of fear of punishment. Heroic leaders like George Washington (except more alive).

Having prestige can be better than being dominant. If you’re dominant, your subordinates will do exactly as much as necessary to avoid your wrath; if you’re prestigious, they may go above and beyond to help you. On the other hand, sometimes good old-fashioned dominance does the trick; your boss can ask you to drop everything and spend a week of long nights on a sudden project, but if your favorite rock star asked you to spend a week doing his taxes for him you might politely decline.

Dominance has clear animal analogies (alpha chimps, chicken pecking orders, etc), and we can pretty well guess why it evolved. The evolutionary origins of prestige are murkier, and this is the focus of Simler’s piece.

First he flirts with the theory of a guy called Henrich, who says prestige comes from *a desire to learn*. I admire and flatter my favorite rock star because I'm hoping I can hang out around him, some of his genius will rub off on me, and I'll be able to play a wicked guitar riff and win a couple of Grammies myself. This theory makes no sense to me. It's not just that there's zero chance of Bowie teaching me, or that I might not have the talent anyway. Maybe in the environment of evolutionary adaptedness that didn't matter so much. It's that I don't want to be a rock star, and if Bowie offered to train me, I'd say I wasn't interested.

Simler doesn't like this much either, so he moves on to the theory of two guys named Zahavi and Dessalles. I'll quote him at length:

> Unlike Henrich, whose account of prestige is unique to our species, Zahavi and Dessalles find analogues among non-human animals — most vividly, in the Arabian babbler.
>
> The Arabian babbler is a small brown bird found in the arid brush of the Sinai Desert and (you guessed it) the Arabian Peninsula. It spends most of its life in small groups of three to 20 members. These groups lay their eggs in a communal nest and defend a small territory of trees and shrubs that provide much-needed safety from predators.
>
> When it's living as part of a group, a babbler does fairly well for itself. But babblers who get kicked out of a group have much bleaker prospects. These "non-territorials" are typically badgered away from other territories and forced out into the open, where they often fall prey to hawks, falcons, and other raptors. So it

really pays to be part of a group. (Keep this in mind; it'll be crucial in a moment.)

Within a group, babblers assort themselves into a linear and fairly rigid dominance hierarchy, i.e., a pecking order. When push comes to shove, adult males always dominate adult females — but mostly males compete with males and females with females. Very occasionally, an intense "all-out" fight will erupt between two babblers of adjacent rank, typically the two highest-ranked males or the two highest-ranked females. This is the babblers' version of a Wild West showdown, as if one babbler suddenly turns to the other and says, "This town ain't big enough for the both of us." A showdown always results in death or permanent exile for one of the combatants.

Most of the time, however, babblers get along pretty well with each other. In fact, they spend a lot of effort actively helping one another and taking risks for the benefit of the group. They'll often donate food to other group members, for example, or to the communal nestlings. They'll also attack foreign babblers and predators who have intruded on the group's territory, assuming personal risk in an effort to keep others safe. One particularly helpful activity is "guard duty," in which one babbler stands sentinel at the top of a tree, watching for predators while the rest of the group scrounges for food. The babbler on guard duty not only foregoes food, but also assumes a greater risk of being preyed upon, e.g., by a hawk or falcon.

Helpfulness, bravery, heroism: these birds seem like regular Boy Scouts. At least on the surface.

But here's where things take a turn for the weird. Babblers don't just passively or occasionally offer to help each other. Instead they compete intensely for the privilege of doing so.

Unlike chickens, who compete to secure more food and better roosting sites for themselves, babblers compete to give food away and to take the worst roosting sites. Each tries to be more helpful than the next. And because it's a competition, higher-ranked (more dominant) babblers typically win, i.e., by using their dominance to interfere with the helpful activities of lower-ranked babblers. This competition is fiercest between babblers of adjacent rank. So the alpha male, for example, is especially eager to be more helpful than the beta male, but doesn't compete nearly as much with the gamma male. Similar dynamics occur within the female ranks.

Now: what in Darwin's name is going on here? Why are babblers so eager to help each other?

The naive answer is that they're simply doing what's best for the group — because when the group succeeds, everyone ends up better off. But this kind of straightforward altruism simply isn't found in nature.[1] It's not game-theoretically stable, thanks to the free-rider problem. Also note that babblers actively interfere with the helpful behavior of their rivals. If their ultimate goal were the success of the group, interfering with others would be entirely counter-productive.

So the logic of natural selection compels us to ask, "What selfish motive does an individual babbler have to help others?"

The answer, in a word, is prestige. A second form of social status that lives alongside the babblers' dominance hierarchy — a kind of "credit" reflecting the amount of good each individual has done for others. So when two babblers compete to stand guard duty, for example, they're actually jockeying, selfishly, for prestige within the group.

And suddenly the intense competition makes sense.

But as in our species, so too in babblers: prestige means nothing without admiration. If other babblers weren't willing to defer and pay respect to prestigious individuals, there'd be no incentive to compete for prestige.

But other babblers are willing to pay respect to prestigious individuals, in two main ways. The first is mating opportunities.[2] Babblers are constantly trying to interfere with their rivals' mating attempts — but when a babbler has high prestige, his or her rivals interfere less. Among males, this translates to more mating opportunities; among females, it translates to earlier mating opportunities (giving one's offspring a head start in the communal nest)

The other perk of high prestige is a reduced risk of being challenged to an all-out showdown. The higher a babbler's prestige, the less likely its rivals are to pick a fight — even if they stand a good chance of winning.

All of which brings us, finally, to the point. Why do other babblers voluntarily defer to prestigious ones? The answer is simply(!) that babblers with lots of prestige are useful to the group, and therefore useful to keep around.[3] This is how it ends up being in the selfish

> interest of other babblers to defer to those with high prestige.
>
> When a babbler is useful enough, in other words, it’s in the self-interest of others to “suck up” or pay respect to that babbler (by backing down from fights and interfering less in its mating attempts) in order to keep it happily in the group.
>
> Bottom line: Prestige-seeking and admiration (deference) are complementary teaming instincts. They help babblers stay attached to a group, keep groupmates happy, and secure a larger share of the group’s reproductive “spoils.”
>
> I hope this account of the babbler prestige system sounds familiar, because it’s more or less equivalent to the prestige system found in our own species; both are derived from the same Platonic form.

This is better. It sort of makes sense as an evolutionary explanation. But I think extending it from there to modern human prestige is a big stretch.

Take the rock star again. Let’s say David Bowie. When people admire Bowie, are they trying to get him to *not leave the group*? Is that why people scream and throw themselves at him? What would it even *mean* for Bowie to leave the group? If he doesn’t have enough groupies, will he defect to North Korea?

And don’t we sometimes admire people who we *do* want to leave the group? Suppose that for some reason I was stuck on a plane sitting next to the Koch Brothers – maybe all their private jets broke down at once. I would probably treat them in the classic way someone treats prestigious people. I’d feel really nervous striking up a conversation with them because they're high-status and important. If I did strike up a conversation with them, I'd be really deferential and overthink everything they said. After the flight was over, I would immediately post to Twitter "I SPENT A WHOLE FLIGHT TALKING TO THE KOCH BROTHERS!" and then post the photo I'd roped them into taking with me. But none of this is because I don't want them to leave the group. If the Koch Brothers defected to North Korea, that would be *great*.

And what about prestigious people who don't bring any special talents to the group? Helen Keller, for example, can do less than most other people. We admire her not because we need to make use of her mad skillz, but because given all her handicaps it's amazing that she can do anything at all.

We could potentially dismiss all of these by saying that evolved instincts don't have to work in the present day. If there were no cavemen like David Bowie (probably a safe bet), then maybe our evolutionary instincts don't apply to his case. But even in evolutionary time, admiration has a free-rider problem. Suppose that we want to make sure David Bowie stays in the West rather than North Korea, but he'll defect unless at least three people flatter him per day. Assuming that flattering David Bowie involves some kind of cost – maybe you have to buy the t-shirt with his face on it – why should I pay the cost when there are millions of other Westerners invested in the same project? Should we be more impressed with the altruistic spirit of people who have sex with famous rock stars, seeing as they are sacrificing their bodies to the project of keeping their heroes out of Kim Jong-un's clutches?

I think I might be straw-manning the babbler hypothesis here, so let's skip down a few paragraphs to the next time Simler explains it:

> The point is, we want to be friends, allies, and teammates with people who do good things for their friends, allies, and teammates. It’s in our self-interest to cultivate access to such people — which we do, in part, by paying them respect and granting them the perks of prestige.
>
> More generally, however, we admire not only those who actually do good things for their teammates, but also those who show the potential to do good things, i.e., by demonstrating useful skills. The student who gets straight As from a good college, for example, is advertising her value to future employers, and her prestige makes her highly sought-after on the job market. She’ll be actively courted by hiring managers and given various perks (a better starting salary, more time to make her decision) that aren’t accorded to her less-impressive classmates.

Simler treats this as a summary of his previous point, but this is a very different theory!

The previous point was that prestigious people do good things for their community. The new point is that prestigious people do good things *for their flatterers in particular*. It’s a tit-for-tat relationship: show David Bowie your tits, and he gives you some tat. Money? Access to the best clubs? A copy of his latest album?

This makes sense except that it’s not the way most admiration-interactions actually work.

Forget David Bowie. Let’s talk about Justin Bieber. I see about a zillion teenage girls hanging posters of Justin Bieber in their room, fighting for the last ticket to Justin Bieber concerts, buying magazines with Justin Bieber on the cover.

But the chance that Justin Bieber gives any tat for all of these tits is practically nil.

And we can’t dismiss this as a form of irrationality restricted to teenage girls. A lot of people I know geek out about Elon Musk; I’ve been to more than one party/meetup where the topic of conversation turns to how great Elon Musk is. I don’t hang posters on my wall, but if I did, they would probably have his face on them. But I don’t expect any repayment from him; I doubt he even knows about my flattery. What about all those Catholics who obsess over the Pope? What about people who obsess over J.K. Rowling or Neil Gaiman or LeBron James or Derek Jeter?

And what about me on that airplane with the Koch Brothers? Am I thinking to myself “If I ever need an entire field of science discredited, now I’ve got an in with some people who are really good at it”? What about Helen Keller? “If the world is plunged into eternal darkness, and also there’s some global super-loud hum that makes it impossible to hear anything, now I’ll have a friend who can operate regardless?” Even in evolutionary times, we should have *some* need to reflect on “can this person actually help me?”

I worry no one theory can completely explain prestige. It seems to me to be a combination of several different things:

1. Group signaling. The people I admire say a lot about me. If I admire Elon Musk, it means that I’m really into space, technology, and maybe the free market. If I admire the Pope, it means I’m really into Catholicism. If I admire David Bowie, it means I’m fabulous. Learning about these people, celebrating their accomplishments, and joining their Official Fan Clubs is an important method of bonding with other peopel.

2. Coattail riding. If a prestigious person becomes more prestigious, I might “look good” for having supported them “before they got big”. It suggests that I’m a good judge of character, or “hip” enough to know which acts will take off and which ones will never achieve broader appeal. Just as a fan feels good when his sports team wins the Superbowl, and a patriot feels good every time her country wins a war, so being a known Elon Musk fan means I get to feel a tiny fragment of the glory whenever Elon Musk invents a new rocket.

3. Prestige by association. Prestigious people hang out with other prestigious people. Nonprestigious people hang out with other nonprestigious people. If I have access to prestigious people, even in some boring trivial way, that makes me seem more prestigious. I think this is what’s going on with the hypothetical airplane conversation with the Koch brothers. Yes, in some sense it’s sheer coincidence that I run into them on a flight. In another sense it isn’t; at the very least, it probably means I was flying first class, and I must have had some rudimentary level of social skills to engage them in conversation. I’m signaling that I’m the sort of person who, at least when everything goes right, can shmooze with billionaires. Even if deep down people know that it was mostly a coincidence, on some gut level that’s kind of impressive.

4. Tit for tat. Yes, in some cases we will be close enough to prestigious people that we can expect rewards for our support. It’s probably easier to flatter my boss or my favorite teacher effectively than to flatter Justin Bieber or the Koch brothers, and you can reasonably expect special treatment. This is a good way of forging an alliance. If I praise my boss, she benefits from my elevation: having a nobody admire you is boring, but having a somebody admire you is both flattering practically useful. Therefore, the more I admire and support my boss, the more she is incentivized to help me become a somebody.

5. Virtuous cycles. Suppose that, for reasons 1 through 4, people want to be associated with prestigious people. Note that this is different from “associate with prestigious people” in the sense of meeting them directly; anything that gets their name linked to the prestigious person will work. In fact, suppose that specifically, there are a bunch of conservatives who are really into the Koch brothers and are jockeying for position as Koch brother fan #1. Some of these people might *play the strategy* of according me prestige for having met the Koch brothers as a way of better signaling their own respect for the Koch brothers to third parties. That gives me a separate incentive to seek such prestige by association.

This is still woefully incomplete, especially by “predict which of these doctors will wear jewelery”-level standards. Maybe prestige shouldn’t be treated as a single thing at all. Maybe the admiration I feel for my boss (a real person in my social circle who I interact with daily) comes from a totally different part of the brain and has totally different evolutionary origins from the admiration I feel for Elon Musk (who I expect never to meet).

But I think separating dominance from prestige is a good start. Do consider reading the full Melting Asphalt essay, as well as Simler’s follow-up thoughts.

A Whiter Shade of Candidate

Vox says that Donald Trump practices the “politics of white insecurity”. US News says that Trump shows “the rising power of the white vote”. Salon wants to tell you “eight reasons why white America falls for demagogues like Donald Trump”. The Week says Donald Trump represents “the rise of white identity politics”. The National Journal says Trump is creating problems by “preaching to a shrinking white electorate”.

Read enough of these articles, and you might start to get the feeling that Donald Trump’s supporters are disproportionately white. You would be wrong.

Well, probably. Data are sketchy. There aren’t a lot of polls that sort their questions by race, and when they do there are sufficiently few non-white Republicans that they have trouble getting a good sample size. Nevertheless, the ones we have suggest that Donald Trump’s supporters are about as diverse as any other Republican’s and maybe moreso.

An August YouGov poll with a sample of 30 Hispanic Republicans finds Trump in the lead with that demographic, getting 28% of the vote to runner-up Ben Carson’s 19%. A Gravis poll with about 40 Hispanic Republicans finds Trump with 37% of their vote to runner-up Marco Rubio’s 20%. A Nevada poll also finds Trump leading among Hispanics in that state, though I don’t know their Hispanic Republican sample size. And finally, head-to-head matchups of Trump vs. Clinton show Trump outperforming some past Republican candidates, including Mitt Romney and George Bush, in the share of Hispanic votes he would likely receive.

This picture is confused by articles asserting either that Trump has the highest favorability ratings among Hispanics, or that Trump has the lowest favorability ratings among Hispanics. In fact, both are true! Favorability ratings allow you to rate someone favorable, unfavorable, neutral, or never-heard-of-em. *Everybody* has heard of Trump, and nobody is neutral about him, allowing both his favorability and his unfavorability to be sky-high (with news sources reporting whichever one of those two facts suits their narrative). Some people have done a little better work and reported his “net favorability”, or favorability-minus-unfavorability, which is very low and indeed negative. But this isn’t what matters in a real election. What matters in a real election is who people vote for. If 40% of Hispanics view him favorably, but they all vote for him, and 60% of Hispanics view him unfavorably, but split their votes among the other ten candidates, Trump has won the Hispanic vote.

As hard as it is to find good data about Hispanics, it’s even harder to investigate black Republicans. The YouGov poll that had thirty Hispanics has only five blacks; it looks like three vote for Carson, one for Rubio, and one for Trump.

Probably more useful are the head-to-head Trump vs. Hillary polls, which survey all blacks (not just Republicans). One finds Trump doing shockingly well and tripling Romney’s (admittedly miniscule, admittedly decreased by opposing Obama) level of support among black voters, but the Washington Post is skeptical and cites others with less extreme results – although even most of those show Trump doing at least as well as Romney and other historical Republicans. In terms of boots on the ground, African-American Daily Beast correspondent Barrett Pitner agrees that “Donald Trump has black supporters – really”, explaining that Trump’s “fear mongering and us-vs.-them tactics have not only created a large supporter base among conservative white Americans, but also black Americans who have been disproportionately hit by the economic downturn.”

There are too few data to say anything for sure. But all of the data that exist suggest that if the Republican primary were held today and restricted to non-whites, Trump would still win. And if Trump were the Republican nominee, he could probably count on equal or greater support from minorities as Romney or McCain before him.

In other words, the media narrative that Trump is doing some kind of special appeal-to-white-voters voodoo is unsupported by any polling data.

On the other hand, there *is* a candidate whom the media narrative fits like a glove. A candidate who may win primary among whites, but loses in a landslide among minorities. A candidate whose black support is almost *an entire order of magnitude* lower than his white support.

That candidate is Bernie Sanders.

According to the same YouGov poll mentioned above, 38% of whites support Bernie Sanders for President, compared to 37% of whites who support Hillary for President. However, only 13% of Hispanics support Sanders, compared to 63% for Hillary. And *only 4%* of blacks support Sanders, compared to 64% for Hillary!

A South Carolina poll from this month broadly agrees. CNN finds that the two candidates are in a statistical dead heat among whites (48-47) but that Hillary has an overwhelming advantage among blacks (84-7).

Other polls are slightly less extreme but tell the same picture. Gravis (early August) found Hillary leading comfortably among all races, but Sanders' support among whites was still twice as high as among blacks. The Washington Post also found Sanders doing abysmally overall, but his support among blacks was super-abysmal – only 5 percent!

Suppose we measure a candidate's "whiteness" by the ratio of their level of white support to their level of nonwhite support within their party. Donald Trump seems to be somewhere around 1.3 – 1.5. Bernie Sanders is somewhere from 3 – 10. It isn't even close. If any candidate is "playing to the politics of white insecurity" or "preaching to the white electorate" or "harnessing the white vote", it is he.

(though I should clarify that in a general election, Sanders would no doubt garner much higher nonwhite support than Trump just because of the D after his name. We're only talking about relative to other people in their own party here)

This explains a couple of otherwise mysterious things. How is Sanders on track to win in Iowa and New Hampshire when he is losing so badly nationally? Well, because Iowa and New Hampshire are two of the whitest states in the country. And how come I keep hearing people say "I'm sure Sanders will win, because even though the media and Big Business support Hillary, everybody I know supports Sanders"? Well, are those people white? Is their entire friend group white? Do they live in very strongly white areas? Then Sanders probably has much higher support among their friends and neighbors than he does nationally.

This might just be a transitory matter of two candidates with different styles and no relevance beyond this particular primary. Or it could represent the first cracks in the alliance that makes up the Democratic Party.

Racially, the Democrats are more diverse than the nation as a whole; since few nonwhites are Republicans, the Democrats are 60-40 white/minority. Socially, the Democrats combine enlightened college-educated creative professionals who want to help the poor, with poor people who want to be helped. And ideologically, the Democrats combine old-school quasi-socialists very concerned about Big Business and income inequality, with social justice activists who think the real issues are race and gender. So far these have been very benign splits. Everyone's interests basically line up the same and nobody has a lot of reason to fight with anyone else – unlike the Republicans, who are already in civil war.

But the current election brings all three splits into near-alignment. The quasi-socialists, whites, and enlightened professionals generally support Sanders. The social justice activists, nonwhites, and poor people generally support Clinton – this bizarre situation of the guy most vocal about helping the least fortunate getting support from everyone except the least fortunate themselves. While I don't *really* expect any fireworks to fly, it's a risky situation and makes this an interesting time to be watching politics.

But mostly I bring this up not because the presidential primary is interesting in itself, but because it really drives home two important points that I've tried to make before.

First, in this post, I suggest that when talking politics "white" sometimes literally means people of European descent, but other times means what I dubbed the "Red Tribe", very loosely corresponding to Republican voters, but also with connotations of southern, poor, uneducated, religious, and exaggeratedly patriotic. This seems to be one of those second times. Even if Donald Trump had 100% support from all minorities, he would still be "the white people candidate”, or even, as some people have called him, “the white power candidate”. Likewise, even if 100% of Sanders’ supporters were white and no black or Hispanic person had ever had the tiniest positive thought about him, we would never get the same kind of “is Bernie Sanders a demagogue harnessing white voters?” story that Trump inspires every day. Sanders supporters aren’t white! They have degrees from Ivy League colleges! *They’re the good guys!*

Second, in this post, I argue against the theory that groups with few black members are necessarily racist or exclusive (frequently seen as “Silicon Valley is problematic because of how few black techies there are”). I note that black people are severely underrepresented in groups as diverse as runners, BDSM participants, atheists, fanfiction readers, Unitarian Universalists, furries, and bird watchers. They’re also underrepresented in movements with apparently impeccable leftist and anti-racist credentials, like Occupy Wall Street and the US Communist Party. Given the frequency with which the “your group has few minorities, that means you’re racist and need to become more explicitly leftist in order to shrieve yourself” argument gets used to punch down at nonconformist or “weird” groups, there can never be too many counterexamples. And Bernie Sanders’ campaign is such a counterexample. It fits poorly with the “low nonwhite representation is caused by insufficiently strong social justice orientation” theory, but very well with the counter-theory I propose in that post: nonwhites are just generally less eager to join weird intellectual signaling-laden countercultural movements.

I take immense schadenfreude in imagining the people who like to write thinkpieces that “call out” polyamory or atheism for their insufficient minority representation, fidgeting and sweating and trying to justify their support for Sanders. I deeply enjoy the thought of them reading the article on ‘Berniebros’ (warning: possibly literally the worst article ever written, I am not kidding) and maybe realizing that wait, this is what they’ve been doing to other people all along, and it’s kind of unfair and hurtful. I mean, this will never happen. But it makes me happy to think about.

Bernie hasn’t done much specific to upset minorities; I doubt those stunts by the Black Lives Matter protesters mattered much one way or the other. And I would naively have expected his message of income equality and helping the least fortunate to go over better with people who are pretty unequal and unfortunate. And although Bill Clinton was pretty popular among nonwhites, I don’t see anything super-special about Hillary that would make her attractive to them.

So I think the explanation here might just be the same explanation as with the atheism and BDSM: Hillary has better name recognition and is more mainstream and less weird, in the same way not-BDSM and not-atheism are more mainstream and less weird. This is also my explanation for Trump’s relative success with minorities: he’s a household name in a way that Marco Rubio and Scott Walker aren’t, and he has vague good associations of strong leadership and economic savvy among the TV-viewing public.

And if Sanders supporters accept that in their own case, maybe they’ll be more understanding when other people plead the same.

Contra Huemer On Morals

Michael Huemer thinks there are objective moral truths, because we've been moving in toward a particular coherent ethical perspective for the past few centuries, and for all we know this could be because that ethical perspective is Objective Truth.

Achitophel: That's a pretty uncharitable way of putting it.

Berenice: But does this view really deserve more charity? Suppose I said that in the past, almost nobody wore ties. Now lots of people do. This is probably because ties are the objectively correct fashion choice.

Achitophel: What if people in a dozen different civilizations independently converged on wearing ties? Wouldn't that provide much stronger evidence?

Berenice: People in a dozen different civilizations *have* converged on wearing ties. Go to France, Russia, China, or Nigeria, and chances are that the most important people you meet there will be wearing ties. Sure, the convergence isn't independent, but neither was the convergence in values. You don't think that India becoming a bicameral parliamentary democracy with a bill of rights had anything to do with Britain being a bicameral parliamentary democracy with a bill of rights?!

Achitophel: You're trying to make it sound like imperial Britain forced their values down India's throat. And maybe they did. But how come things like representative government, human rights, and decreased torture took off in a bunch of countries that were never colonized at all?

Berenice: Which countries?

Achitophel: Japan? Russia? China?

Berenice: Japan requires an overly restrictive definition of "never colonized". And China and Russia require a frankly insane definition of "representative government, human rights, and decreased torture taking off".

Achitophel: Not in an absolute sense! Relative to before!

Berenice: Give me the Yongle Emperor over Mao any day of the week.

Achitophel: Mao was bad. But he pretended not to be. He didn't say "Let's go kill a bunch of people because killing is glorious." He said "We shouldn't kill people, but sometimes we have to." He didn't say "You're all my slaves, because I have divine right." He said "We're all going to work towards freedom together, but the best way to do that is by doing what I say." He still had more liberal values than the Yongle Emperor, he just did evil despite them.

Berenice: I feel like this is an odd distinction to insist upon when you are sitting atop a pile of skulls.

Achitophel: And Xi is better than Mao.

Berenice: Not too different from Yongle, honestly.

Achitophel: All right. Fine. Let's forget about independent development by different civilizations. Let's say we're mostly talking about the West – which remember, is still a lot of different countries. Britain. France. Germany. Italy –

Berenice: I am aware which countries are in the West.

Achitophel: These countries all converged on the same couple of values. And those values were all coherent with one another. It seems pretty clear that “emancipation of slaves”, “”freedom of speech”, “decolonization”…

Berenice: Wait a second. Sure, we’ve done a lot of decolonizing the past fifty years. But we did a lot of colonizing the five hundred years before that. In fact, around 1450 the West switched from barely colonizing at all, to colonizing lots of stuff all the time. If Huemer had lived in 1750, wouldn’t he have argued that the arc of the moral universe is long but it tends toward colonialism? And then declared colonialism an objectively correct moral truth?

Achitophel: Stop interrupting! “Emancipation”, “freedom of speech”, “decolonization”, “women’s rights”, and “democratic governance” are all kind of in the same moral direction, so to speak. Do you agree that Western values, today, not in 1750, TODAY, are all going in a certain coherent direction instead of varying randomly?

Berenice: You know, it’s not just ties.

Achitophel: What?

Berenice: If you think about it, practically every item of clothing has become less ornate. Think of Louis XIV in his huge expensive wig, his shiny blue fleur-de-lis filled fur robes, his carefully sculpted gold cane, his bejeweled ceremonial sword, his shiny red heels encrusted with diamonds, his gigantic outrageous hat, all sorts of weird neckbands and armbands. The Yongle Emperor would have had a more Chinese style, but it wouldn’t have been so different in conception. But nowadays nobody does that, not even the rich people who could afford it. The only time you’ll get shiny jewel-filled robes and fifty different things going around your neck is when somebody wants to look old-fashioned and traditional, like a Pope or Cardinal. And this is true everywhere. De Gaulle dressed more simply than Louis, and Mao dressed more simply than the Yongle Emperor. And when we picture the future, everyone's dressed in featureless skin-tight suits. Evidence for objectively correct fashion?

Achitophel: There's probably some driving force that made simplicity of clothing desirable, and which applied equally everywhere. For example, ornate clothing was a good signal of wealth back in Louis' time. But after the Industrial Revolution, anyone could wear ornate clothing. Once the middle-class starts showing up to their bear-baitings in ornate fleur-de-lis gowns, wearing it just meant you were too clueless to know that it had no value anymore. So countersignaling took over – haven't we talked about this before? The clothing thing isn't because of some objectively correct fashion choice, it's just a side effect of increasing wealth?

Berenice: Ding ding ding! Gold star for you! But why don't you follow your theory to its logical conclusion and realize that the change in morality is also an effect of increasing wealth? Robin Hanson has just written about this in response to Huemer. Here, I'll quote him for you:

> One of the two main factors by which national values vary correlates strongly with average national wealth. At each point in time, richer nations have more of this factor, over time nations get more of it as they get richer, and when a nation has an unusual jump in wealth it gets an unusual jump in this factor. And this favor explains an awful lot of the value choices Huemer seeks to explain. All this even though people within a nation that have these values more are not richer on average.

> The usual view in this field is that the direction of causation here is mostly from wealth to this value factor. This makes sense because this is the usual situation for variables that correlate with wealth. For example, if length of roads or number of TVs correlate with wealth, that is much more because wealth causes roads and TVs, and much less because roads and TV cause wealth. Since wealth is the main “power” factor of a society, this main factor tends to cause other small things more than they cause it.

This seems obviously correct to me and I don’t know why you and Huemer can’t see it.

Achitophel: You didn’t quote Huemer’s response! Here:

> Perhaps there is a gene that inclines one toward illiberal beliefs if one’s society as a whole is primitive and poor, but inclines one toward liberal beliefs if one’s society is advanced and prosperous. Again, it is unclear why such a gene would be especially advantageous, as compared with a gene that causes one to be liberal in all conditions, or illiberal in all conditions. Even if such a gene would be advantageous, there has not been sufficient opportunity for it to be selected, since for almost all of the history of the species, human beings have lived in poor, primitive societies.

Berenice: Which gene that inclines us to take an airplane when we want to get somewhere quickly, but inclines us to take the bus if economy is more important? Is it DRD4 or SERT? I always forget that one.

Achitophel: You’re saying that it isn’t genetic.

Berenice: Or differently genetic, or complicatedly genetic, or gene-environmental-interactionic. This is what Robin

Hanson says:

> Well if you insist on explaining things in terms of genes, everything is “unclear”; we just don’t have good full explanations to take us all the way from genes to how values vary with cultural context. I’ve suggested that we industry folks are reverting to forager values in many ways with increasing wealth, because wealth cuts the fear that made foragers into farmers. But you don’t have to buy my story to find it plausible that humans are just built so that their values vary as their society gets rich.

Achitophel: That’s your argument? “We just don’t have good full explanations to take us all the way from genes to how values vary with cultural context?” Your whole point is just an argument from ignorance? Forgive me if I wait until you can come up with a plausible mechanism.

Berenice: You want plausible mechanisms? I’ve got your plausible mechanism RIGHT HERE. To put it in Haidtian terms, the Purity moral foundation, plus a sort of ethnocentrism that corresponds roughly to his Loyalty and Authority moral foundations, are carefully evolutionarily regulated by the prevalence of disease. Purity is the most obvious, given that the disgust reflex is obviously an evolutionary defense against pathogens. The reason you’re grossed out at the thought of touching feces, blood, or rats is that they’re full of plague; the reason you’re even more grossed out by the thought of eating them is that eating things is an even better way to get plague than touching things. Likewise, the best reason to avoid strangers is that they might have strange germs; about twenty million Native Americans who learned that lesson the hard way. Humans have an evolved behavior of upping their levels of purity and ethnocentrism under germ threat. Invent sanitation and antibiotics, eliminate most germs, and people naturally tend toward lower purity-concern and ethnocentrism. You get less racism, more sex, nontraditional families, cultural mixing, and all that good stuff. That's why you get great correlations between the levels of pathogens in a region and the moderrness of their values. Go somewhere cold and lifeless like Sweden and you'll get a liberal utopia. Go to a jungle in the Congo full of creepy-crawlies and everyone will be slashing everyone else with machetes. Really, read the article!

Achitophel: You think antebellum Southerners didn't like black people because they thought they *had cooties*? Forgive me if the whole enslavement thing doesn't seem to follow.

Berenice: I'm not saying that's the only explanation or even the main explanation. You asked for a possible mechanism. I gave you one.

Achitophel: Fine. Give me a mechanism that explains slavery, then. And don't you dare say it's not the *main explanation* afterwards. Give me the best you've got.

Berenice: Have you ever noticed how much more virtuous rich people are than poor people? Poor people shoplift all the time, but rich people almost never do.

Achitophel: I don't know where you're going with this, but rich people commit white-collar crime and defraud people out of millions of dollars.

Berenice: Which just goes to show their moral superiority all the more! The poor person sells his principles for a dollar; the rich person holds fast until the temptation becomes absolutely overwhelming.

Achitophel: Shut up and make your point.

Berenice: A lot of moral decisions are a conflict between a principle and a temptation. People with fewer temptations have an easy time looking more principled. Not shoplifting is easy for a rich person, not because they're more virtuous, but because they're not in a position where they gain anything by doing so.

Achitophel: And this relates to slavery how?

Berenice: I would argue that we have many different drives and needs, some of which can be raw materials for making morality. Compassion is a drive. Xenophobia's also a drive. Either one can be emphasized or deemphasized based on what's useful or practical. If the most important thing for you is coming up with an excuse to enslave other people to make cotton, you might cultivate this primitive xenophobia into a complicated system of institutionalized racism that becomes the value system of your entire culture. If you're not doing that, maybe compassion wins out. I mean, isn't it interesting that all of the moral decent liberal people were north of a certain imaginary line, and all of the immoral bigoted people were south of it? And that imaginary line just happened to separate the climate where you could grow cotton from the one where you couldn't? I'd argue instead that given a sufficiently lucrative deal with the Devil, the South took it. The Devil didn't make the North an offer, and so they courageously refused to yield to this total absence of temptation.

Achitophel: You make the Southerners sound pretty Machiavellian.

Berenice: No more than the rest of us. I expect that once somebody invents vatburgers, we'll all gain an sudden respect for animal rights, and recoil in horror that we ever engaged in factory farming. Until then, we come up with various moral justifications for the thing we're not going to stop doing.

Achitophel: So liberal values are real morality, and older values are just excuses to justify greed?

Berenice: Not necessarily greed. "Necessity" is too strong, "convenience" is too weak, but somewhere in between the two. Back in the old days nobody really knew what STDs were. They just knew if you had sex too many times, you would break out in a horrible pox and die. And so would anyone else you had sex with, no matter how otherwise-pure they were themselves. Under those circumstances, having a very sex-negative morality where the promiscuous people are shunned and driven from society is a basic concession to the survival instinct. You'd be insane not to. But once we figured out testing and pencillin, the reasoning behind that morality died out and we stopped trying to cultivate those values. The sex-negative morality isn't trying to justify *greed*. It's making basic concessions necessary for survival. And you know what? If we suddenly had a zombie apocalypse and all of the gains of civilization evaporated, we'd be back to the old illiberal morality in the blink of an eye.

Achitophel: It still sounds kind of liberal modern values are the real morality, and other values are just sort of necessary evils.

Berenice: I think it's more symmetrical than that. A lot of modern values would disappear if we stopped facing modern problems. We worry a lot about racial sensitivity, but if we ever got a society where racism was as thoroughly neutralized as syphilis, we'd probably drop that value pretty quickly too. If we ever totally conquer poverty, so that everyone’s got more than enough, maybe we’ll even stop worrying about compassion and fairness. Likewise, a lot of the democratic values – freedom of speech, freedom from slavery, equality, etc – are based on most countries being democracies which in turn is based on the historical situation. One of the big shifts was from the medieval system of “mostly super-well-trained professional warriors ie knights matter in projecting military force” to “any warm body with a gun matters”. That gave the common people a new level of power and probably led to democracy and the democratic virtues of equality and freedom. Likewise, technology has connected the world to the degree where different races and cultures and ideas are frantically mixing and mutating, making things like tolerance and freedom of thought much more relevant.

Achitophel: What about not torturing people? What about trying to solve poverty?

Berenice: So we’re too egalitarian to worry much about Authority and Loyalty. We’ve got too many antibiotics and contraceptives to care about Purity. But Care/Harm and Fairness seem as relevant as ever. Maybe even moreso. Given the advances in journalism, communication, and art, we have the ability to learn about and appreciate the struggles of others in a way we never have before.

Achitophel: That sounds a little forced. I could come up with a counter-story where given the worldwide increase in wealth and our lack of real-life exposure to any starving people or smallpox victims, the Care foundation atrophies away, but given our increasing crowding and exposure to superplagues like HIV and Ebola, Purity becomes obsessively important.

Berenice: *shrug* Maybe Care/Harm really is just the fundamental moral foundation, and the others are epiphenomena to be abandoned as we outgrow them. How does that saying go? – “The last enemy to be destroyed is submaximal global utility; destroying Death just buys us more time.”

Achitophel: So you kind of agree with Huemer after all?

Berenice: Perish the thought! Huemer thinks that this change in values proves there’s an objective morality and we’re moving toward it. The strongest claim I would dare is that one of these axes has always been the one that, all else being equal, would dominate the balance – and this is just the first time all else has been equal.

Meat Your Doom

[Epistemic status: Very dirty and approximate, but I think roughly correct. Check my calculations and tell me if I'm wrong.]

A recent formative experience: a seriously ill patient came in and I recommended a strong psychiatric drug. She looked it up online and told me she wouldn't take it because was associated with an X% increase in mortality.

"But," I pointed out, "you're really miserable."

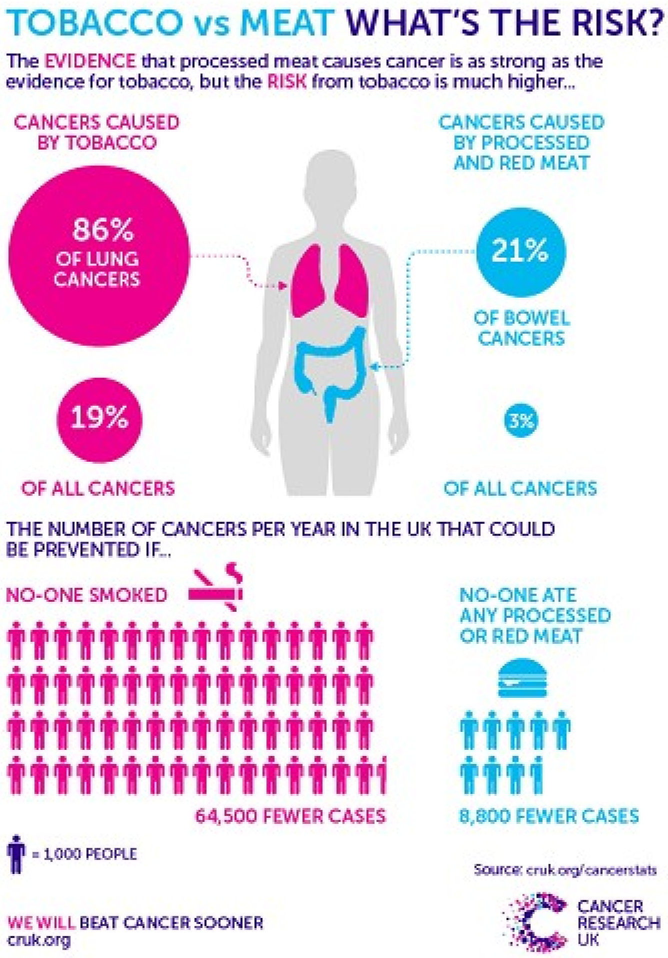
"But I don't want to die!"

So I looked it up, did the calculations, and found that it would on average take a couple of months off her life. And I asked her, "Which would you prefer – living 80 years severely ill, or living 79.5 years feeling mostly okay?"

She still wasn't convinced, so I asked her if she ate cookies. She said yes, almost every day. I told her that the cookies were probably taking more time off her life than the medication would, and I assured her the medication would probably add more value to her life than cookies.

She took the drug.

I thought of this the other day when everyone started sharing that study about meat causing colon cancer. A lot of people used headlines like Processed Meats Rank Alongside Smoking As Cancer Causes. This was very correctly debunked by infographics like this one:



But I feel like this leaves something to be desired. Eating meat is not as bad as smoking. But that's still a lot of room for it to be bad. Can we quantify the risk better?

From the BBC article: "'[There would be] one extra case of bowel cancer in 100 lifetime bacon-eaters,' argues Sir David Spiegelhalter, a risk professor from the University of Cambridge."

This teaches us something important: "risk professor" is an awesome job title and "David Spiegelhalter, Risk Professor" ought to be a BBC television show starring Harrison Ford.

But also: use absolute risk instead of relative risk! “21% of bowel cancers are caused by meat” doesn’t give you a really good handle on how worried you should be. “One extra case of bowel-cancer in 100 lifetime bacon-eaters” is better.

But let me try to give even more perspective. A bit less than half of colon cancers are fatal. So one extra case per hundred means if you eat bacon daily then there’s an 0.4% chance you will die from a cancer you would not otherwise have gotten.

The average age at diagnosis of colon cancer is 69; the average life expectancy is 79. Sweeping a lot of complexity under the rug and taking a very liberal estimate, the average death from colon cancer costs you ten years of your life.

Multiply out and an 0.4% chance of losing 10 years means that you lose on average two weeks.

Suppose that every case of cancer, fatal and non-fatal alike, causes you additional non-death-related distress equal to two years of your life. That’s about another week.

So overall, if you eat processed meat every day your entire life, you’ll lose about three weeks of life expectancy from colon cancer. That means each serving of meat costs you a minute of your life. You probably lose twenty times that amount just cooking and preparing it.

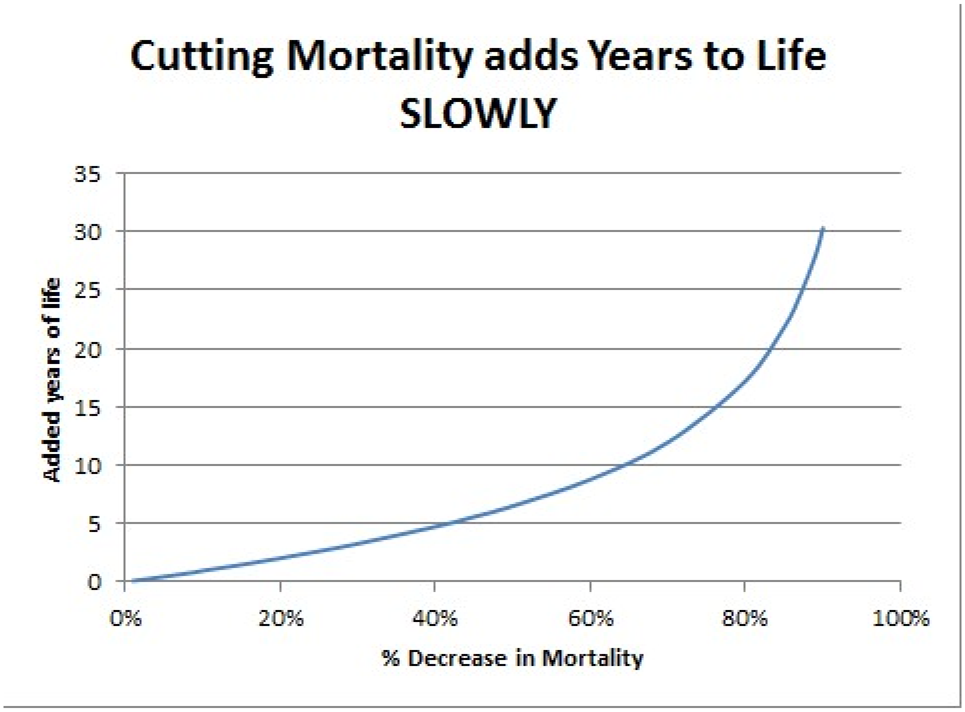
II.

Note that I am not saying “eating meat will only decrease your lifespan by three weeks”. That is the amount that we have *clear evidence* for, *from this study*. It is an example of why *this study* needs to be put in context so that you don't worry about it too much.

There are nevertheless a lot of other studies that suggest greater risks, mostly cardiovascular or metabolic. For example, as per this article, some studies suggest that a serving of red meat per day increases mortality 13%, and a serving of processed meat per day increases it 20%. But it also quotes another study of half a million people that finds meat to be slightly *protective* (sigh) and finds a higher all-cause mortality in the non-meat-eaters.

Whatever. Forget the object-level question for a minute. What are we to make of a claim like "processed meat increases mortality 20%"?

If you're like me, you want to think "Okay, average life expectancy is eighty years, subtract 20% off of that, and you get 64 years. I'll live 64 years if I eat bacon every day." WRONG. Mortality rates are much more complicated, but the key insight is that very few people die when they're young. If you have approximately a 0% chance of dying at age 30, then adding 20% to 0 is still 0. Chance of mortality creeps upward *very slowly* and so even large changes in mortality barely affect the underlying distribution. The only good presentation of this I have ever seen anywhere is on Josh Mitteldorf's blog, which includes the following chart:



This is decrease and we’re talking increase, but but it shouldn’t make much difference here. A 20% increase in mortality isn’t going to bring you from 80 to 64. It’ll probably just bring you from 80 to 78.

Indeed, later in the BBC article, they bring in David Spiegelhalter (RISK PROFESSOR!) who explains that:

> If the studies are right…you would expect someone who eats a bacon sandwich every day to live, on average, two years less than someone who does not. Pro rata, this is like losing an hour of your life for every bacon sandwich you eat. To put this into context, every time you smoke 20 cigarettes, this will take about five hours off your life.

That’s for processed meat. Red meat is safer. Also, we still don’t know if these studies are right.

This is why it’s important to distinguish between absolute and relative risk. You hear all of these scary numbers – 21%

increase in bowel cancers! 20% increase in all-cause mortality! – and it sounds like you're going to drop dead the moment you take a bite of a hot dog. And there's always that chance. Being healthy is good. Being unhealthy is bad. But is life so dear ~~or peace so sweet~~, that you're never going to want to sacrifice an hour to have a bacon sandwich?

All these hours do add up. I'm not saying dietary recommendations aren't important. But the recommendations are important *in aggregate*. If you stick to the spirit of not eating in a horribly unhealthy way, you have a lot of leeway to continue to eat specific things you like even if you know they're not the best for you. And meat falls firmly within that category.

(though you might also want to consider how to manage the moral issues)

What Developmental Milestones Are You Missing?

[Epistemic status: Speculative. I can't make this post less condescending and elitist, so if you don't like condescending elitist things, this might not be for you.]

Developmental psychology never struck my interest in the same way as a lot of other kinds of psychology. It didn't seem to give me insight into my own life, help me understand my friends, or explain weird things about society.

I've changed my mind about all of that after reading David Chapman's Developing Ethical, Social, and Cognitive Competence.

First, a refresher. Developmental psychology describes how children go from helpless infants to reasonable adults. Although a lot of it has to do with sensorimotor skills like walking and talking, the really interesting stuff is cognitive development. Children start off as very buggy reasoners incapable of all but the most superficial forms of logic but gradually go on to develop new abilities and insights that allow them to navigate adult life.

Maybe the most famous of these is "theory of mind", the ability to view things from other people's perspective. In a classic demonstration, researchers show little Amy a Skittles bag and ask what she thinks is inside. She guesses Skittles, but the researchers open it and reveal it's actually pennies. Then they close it up and invite little Brayden into the room. Then they ask Amy what Brayden thinks is inside. If Amy's three years old or younger, she'll usually say "pennies" – she knows that pennies are inside, so why shouldn't Brayden know too? If she's four or older, she'll usually say

"Skittles" – she realizes on a gut level that she and Brayden are separate minds and that Brayden will have his own perspective. Sometimes the same mistake can extend to preferences and beliefs. Wikipedia gives the example of a child saying "I like Sesame Street, so Daddy must like Sesame Street too." This is another theory of mind failure grounded in an inability to separate self and environment.

Here's another example which tentatively sounds like a self-environment failure. Young children *really* don't get foreign languages. I got a little of this teaching English in Japan, and heard more of it from other people. The really young kids treated English like a cipher; everybody started out knowing things' *real* (ie Japanese) names, but Americans insisted on converting them into their own special American-person code before talking about them. Kids would ask weird things like whether American parents would make an exception and speak Japanese to their kids who were too young to have learned English yet, or whether it was a zero-tolerance policy sort of thing and the families would just not communicate until the kids went to English school. And I made fun of them, but I also remember the first time I visited Paris I heard somebody talking to their dog, and for a split second I was like "Why would you expect your *dog* to know *French*?" before my brain kicked in and I was like "Duuhhhh…."

The infamous "magical thinking" which kids display until age 7 or so also involves confused self-environment boundaries. Maybe little Amy gets mad at Brayden and shouts "I HATE HIM" to her mother. The next day, Brayden falls off a step and skins his knee. Amy intuits a cause-and-effect relationship between her hatred and Brayden's accident and feels guilty. She doesn't realize that her hatred is internal to herself and can't affect the world directly. Or kids displaying animism at this age, and expecting that the

TV doesn’t work because it’s angry, or the car’s not starting because it’s tired.

Psychology textbooks never discuss whether this progression in and out of developmental stages is innate or environmental, which is weird because psychology textbooks usually love that sort of thing. I always assumed it was innate, because it was on the same timeline as things like walking and talking which are definitely innate. But I’ve been moved to question that after reading some of the work comparing “primitive” cultures to primitive developmental stages.

This probably isn’t the most politically correct thing to do, but it’s notable enough that anthropologists have been thinking about it for centuries. For example, from *Ethnicity, Nationality, and Religious Experience*:

> Primitive people are generally as intelligent as the people of any culture, including the contemporary industrial-electronic age cultures. that makes it all the more significant that their publicly shared cognitive style shows little identifiable formal operational thought. The probable explanation for this, if true, is simply that formal operational thought is more complexly difficult than earlier modes of thought and will be used in a culture in a publicly shared way only if that culture has developed techniques for training people in its use. Primitive cultures do not do that, and thus by default use easier styles of thought, ones closer in form to concrete oeprational and even pre-operational thought, as defined by Piaget.

Primitive cultures certainly exhibit the magical thinking typical of young children; this is the origin of a whole host of superstitions and witch-doctory. They exhibit the same animism; there are hundreds of different animistic religions worldwide. And although I didn't talk much about theories of moral development, primitive cultures' notion of taboo is pretty similar to Kohlberg's conventional stage.

But if different cultures progress through developmental milestones at different rates or not at all, then these aren't universal laws of child development but facts about what skills get learned slowly or quickly in different cultures. In this model, development is not a matter of certain innate abilities like walking "unfolding" at the right time, but about difficult mental operations that you either learn or you don't depending on how hard the world is trying to cram them into your head.

So getting back to David Chapman: his post is mostly about Robert Kegan's account of "stages of moral development". I didn't get much from Kegan himself, but I was fascinated by an idea just sort of dropped into the middle of the discussion: that less than half of the people in modern western countries had attained Kegan's fourth stage, and only a small handful attained his fifth. This was a way of thinking about development that I'd never heard before.

On the other hand, it makes sense. Take General Semantics (please!). I remember reading through Korzybski's giant blue book of General Semantics, full of labyrinthine diagrams and promises that if only you understood this, you would engage with the world totally differently, you'd be a new man armed with invincible cognitive weapons. And the key insight, maybe the *only* insight, was "the map is not the territory", which seems utterly banal.

But this is a self-environment distinction of exactly the sort that children learn in development. It's dividing your own representation of the world from the world itself; it's about as clear a reference to theory of mind as you could ask for. Korzybski considered it a revelation when he discovered it; thousands of other people found it helpful and started a movement around it; I conclude that these people were missing a piece of theory-of-mind and Korzybski gave it to them. Not the whole deal, of course. Just a piece. But a piece of something big and fundamental, so abstract and difficult to teach that it required that whole nine-hundred-something page book to cram it in.

And now I'm looking for other things in the discourse that sound like developmental milestones, and *there are oodles of them*.

I remember reading this piece by Nathan Robinson, where he compares his own liberal principles saying that colleges shouldn't endorse war-violence-glorifying film "American Sniper" to some conservatives arguing that colleges shouldn't endorse homosexuality-glorifying book "Fun Home":

> It is hypocrisy for liberals to laugh at and criticize the Duke students who have objected to their summer reading book due to its sexual and homosexual themes. They didn't seem to react similarly when students at other universities tried to get screenings of American Sniper cancelled. If you say the Duke students should open their minds and consume things they disagree with, you should say the same thing about the students who boycotted American Sniper. Otherwise, you do not really have a principled belief that people should respect and take in other opinions, you just believe they should respect and take in your own opinions. How can you think in one case the students are close-minded and sheltered, but in the other think they are open-minded and tolerant? What principled distinction is there that

> allows you to condemn one and praise the other, other than believing people who agree with you are better?

He proposes a bunch of potential counterarguments, then shoots each counterargument down by admitting that the other side would have a symmetrical counterargument of their own: for example, he believes that "American Sniper" is worse because it's racist and promoting racism is genuinely dangerous to a free society, but then he admits a conservative could say that "Fun Home" is worse because in their opinion it's homosexuality that's genuinely dangerous to a free society. After three or four levels of this, he ends up concluding that he can't come up with a meta-level fundamental difference, but he's going to fight for his values anyway because they're his. I'm not sure what I think of this conclusion, but my main response to his article is *oh my gosh he gets the thing*, where "the thing" is a hard-to-describe ability to understand that other people are going to go down as many levels to defend their self-consistent values as you will to defend yours. It seems silly when I'm saying it like this, and you should probably just read the article, but I've seen *so many people* who lack this basic mental operation that this immediately endeared him to me. I would argue Nathan Robinson has a piece of theory-of-mind that a lot of other people are missing.

Actually, I was kind of also thinking this with his most recent post, which complains about a Washington Post article. The Post argues that because the Democrats support gun control and protest police, they are becoming the "pro-crime party". I'm not sure whether the Post genuinely believes the Democrats are pro-crime by inclination or are just arguing their policies will lead to more crime in a hyperbolic figurative way, but I've certainly seen sources further right make the "genuinely in favor of crime as a terminal value" argument. And this doesn't seem too different from the leftist sources that say Republicans can't *really* care about the lives of the unborn, they're just "anti-woman" as a terminal value. Both proposals share this idea of not being able to understand that other people have different beliefs than you and that their actions proceed naturally from those beliefs. Instead of saying "I believe gun control would increase crime, but Democrats believe the opposite, and from their different perspective banning guns makes sense," they say "I believe gun control would increase crime, Democrats must believe the same, and therefore their demands for gun control must come from sinister motives."

(compare: "Brayden brought the Skittles bag with him for lunch, so he must enjoy eating pennies." Or: "Daddy is refusing to watch Sesame Street with me, so he must be secretly watching it with someone else he likes better instead.")

Here are some other mental operations which seem to me to rise to the level of developmental milestones:

1. Ability to distinguish "the things my brain tells me" from "reality" – maybe this is better phrased as "not immediately trusting my system 1 judgments". This is a big part of cognitive therapy – building the understanding that just because your brain makes assessments like "I will definitely fail at this" or "I'm the worst person in the world" doesn't mean that *you* have to believe them. As Ozy points out, this one can be *easier* for people with serious psychiatric problems who have a lot of experience with their brain's snap assessments being *really* off, as opposed to everyone else who has to piece the insight together from a bunch of subtle failures.

2. Ability to model other people as having really different mind-designs from theirs; for example, the person who thinks that someone with depression is just "being lazy" or needs to "snap out of it". This is one of the most important factors in determining whether I get along with somebody – people who don't have this insight tend not to respect boundaries/preferences very much simply because they can't believe they exist, and to simultaneously get angry when other people violate their supposedly-obvious-and-universal boundaries and preferences.

3. Ability to think probabilistically and tolerate uncertainty. My thoughts on this were mostly inspired by another of David Chapman's posts, which I'm starting to think might not be a coincidence.

4. Understanding the idea of trade-offs; things like "the higher the threshold value of this medical test, the more likely we'll catch real cases but also the more likely we'll get false positives" or "the lower the burden of proof for people accused of crimes, the more likely we'll get real criminals but also the more likely we'll encourage false accusations". When I hear people discuss these cases in real life, they're almost never able to maintain this tension and almost always collapse it to their preferred plan having no downside.

Framed like this, both psychotherapy and LW-style rationality aim to teach people some of these extra mental operations. The reactions to both vary from enlightenment to boredom to bafflement depending on whether the listener needs the piece, already has the piece, or just plain lacks the socket that the piece is supposed to snap into.

This would have an funny corollary; the LW Sequences try to hammer in how different other minds can be from your own in order to develop the skill of thinking about artificial intelligences, but *whether or not AI matters* this might be an unusually effective hack to break a certain type of person out of their egocentrism and teach them how to deal with other humans.

This raises the obvious question of whether there are any basic mental operations I still don’t have, how I would recognize them if there were, and how I would learn them once I recognized them.

2014 SSC Survey Results

A year ago I asked people to take a survey and said I'd post the results "soon". Now it's been a year and I feel bad, so here are some very quick and dirty survey results. Remember, all of this is hopelessly out of date. Remember also that because the LW Survey was going on at the same time, this is a survey specifically of SSCers who were *not* Less Wrongers.

Although SSC got about 10,000 hits per day last November, only 649 people took the survey. 32% were commenters and the rest lurkers. 85% were male, 11% female, and the rest decline to state their gender; 1.2% were trans, 95.9% were cis, and the rest declined to state. Ages ranged from 16 to 87 (!), with an average of 29. 88% were white, 3% Indian, 2% East Asian, 2% Hispanic, 1% black, and 0.5% Middle Eastern, with the rest declining to state.

Readers were overwhelmingly American at 63%. Other popular countries were the UK (9%), Canada (5%), Germany (4%), New Zealand (1%), and of course Finland (1%). American readers were most frequently from California (19%), New York (11%), Massachusetts (10%), Illinois (6%), Pennsylvania (5%), Texas (5%), and Wisconsin (3%). Massachusetts has the highest readership per capita, which I guess isn't surprising. Not all readers from Massachusetts were able to spell "Massachusetts" correctly on the survey.

77% were straight, 4% gay, 13% bi, and 2% asexual. 26% were married, 27% in a relationship, and 45% single. 65% considered themselves monogamous, 8% polyamorous, and the rest undecided.

41% of readers have a bachelors’ degree, 18% a master’s degree, 9% a PhD, and 5% some kind of professional degree like MD or JD. 31% are students. The most popular professions are computers (33%), finance (7%), engineering (7%), and math (7%). Mean income is $72000 and median income is $47000, but this was limited to the subset of people who entered an income at all. Mean charitable donation was $30000 (!), which reduced to $1700 after removal of one outlier; median donation was $200. Mean percent of income donated was 2.4%; median was 0.5%.

Average IQ was 139, average SAT was 1480 by the old reckoning and 2220 by the new reckoning.

68% of people were atheist, 16% agnostic, and 19% religious, of whom 12% identified as “committed” to their religion. Religious background (regardless of present religious affiliation) was 35% Protestant, 27% Catholic, 8% Jewish, 3% Mormon, 2% Hindu. Ethics were 43% consequentialist, 14% virtue ethics, 6% contractualism, 6% natural law, 4% deontology.

On a political spectrum where 1 is farthest left and 10 is farthest right, the average person placed themselves at 4.6. 19% identified with the US Democratic Party, 7% with the US Republican Party, and 3% with the US Libertarian Party. Of the ideological affiliations available, the top four were social democratic (29%), liberal (23%), libertarian (22%), and conservative (9%). Readers were mostly neutral on feminism, human biological differences, and the minimum wage; they mostly supported gay marriage, environmental action against global warming, more immigration, and basic income guarantees.

10% of readers identified with the social justice movement, 2% with the neoreactionary movement, and 7% with the effective altruist movement, with much larger numbers “sort of” identifying with each. Two people identified as both social justice *and* neoreactionary, which must involve some interesting logical gymnastics.

Most people came here via links from other blogs. The blogs that linked the most people were Bryan Caplan’s (21), Marginal Revolution (20), Unequally Yoked (10), David Friedman’s (6), and Gruntled & Hinged (3).

People’s favorite post was overwhelmingly Meditations on Moloch (77), followed by I Can Tolerate Anything Except The Outgroup (36), followed by Who By Very Slow Decay (10).

Older readers were more conservative, both by self-identification and on all individual issues. There were few other interesting or unexpected correlations; in particular, IQ didn’t correlate much with anything except SAT score. Digit ratio *did* correlate with lots of interesting things, but I should make that into its own post.

Although this survey represented the non-LW SSC demographic, the results were very similar to the Less Wrong survey results in terms of demographics, IQ, chosen professions, et cetera.

If you want more data, you can also take a look at the recent survey on gender related stuff (more commentary here). Of note, this survey has about twice the percent female as last year’s survey, reflecting either a change in the population (unlikely) or a difference in who responds to gender-related versus non-gender-related surveys (more likely).

I couldn’t remember whether I promised survey-takers privacy or not, so I deleted all theoretically traceable data (mostly age and location) and the remainder is available for examination and further analysis here.

Looking A Gift Horse In The Mouth

I started criticizing social justice back in 2010, which doesn't sound so impressive until you realize that's two centuries ago in Internet Years. At the time, you rarely heard such criticism outside of wingnutty lesbianism-causes-witchcraft circles. It felt bizarre, transgressive, and novel.

But over the past few years I've been privileged (sic) to meet many other people with the same concerns. Some were kindred spirits. A few at least had interesting ideas. Many others were horrible people next to whom the lesbian-causes-witchcraft types looked like Voltairesque voices of reason.

But they all had something in common: they were nobodies, and nobody cared what they thought. The lesbian-causes-witchcraft types had their talk shows, but among moderate liberals social justice criticism stayed mostly confined to a bunch of small blogs.

Now that's over – during the past year big national media have unleashed a flood of social-justice-critical stories. The Atlantic published The Coddling Of The American Mind. Salon (Salon!) published Campus PC Panic Is Getting Ridiculous and How Coddled Young Radicals Got Discomfort All Wrong. The New Republic published Trigger Happy. Even President Obama has condemned what he called "coddled" college students, saying "that's not the way we learn". The UK political class is up in arms about Germaine Greer being denied platform, and the US political class is up in arms about the Halloween costume argument at Silliman College (nominative determinism!) in Yale. Complaining about social justice seems to be getting, dare I say, almost *trendy*.

As the old saying goes, “don’t look a gift horse in the mouth”. But as the other old saying goes, “I know some Trojans who would be a lot happier if they had.” So let me explain why this sudden outpouring of support for my position makes me uncomfortable.

When I or some other random blogger complains about the social justice movement, we tend to worry about points like the following (I won’t prove/defend these claims here, just clarify what I’m worried about):

- The level of social-justice-inspired bullying online and offline that can drive people to suicide for even slightly disagreeing with social justice orthodoxy.

- The chilling effect on research when science is subordinated to political ideology, and how researchers whose results contradict social justice orthodoxy can expect to be ignored at best and subject to death threats and harassment campaigns at worst.

- The trivialization of and hostile response to anybody who claims to be suffering in a way that doesn’t fit the social-justice narrative, and opposition to attempts to alleviate such suffering.

- The use of social justice as a bludgeon by which sophisticated elites from top colleges can condemn all subcultures except their sophisticated elite subculture as being problematic, and credibly demand that they subordinate themselves to the sophisticated elites as penance.

- The conflation of the vitally important will toward political reform with the most trivial pop culture clickbait, so that instead of worrying about inequality

> and technological stagnation our brightest minds are discussing whether the latest Game of Thrones episode reinforces structural oppression, or if people’s Halloween costumes are okay or not.

Meanwhile, when important public figures and nationally circulating magazines complain about the social justice movement, I usually see language and arguments more like the following:

- College students are big babies!
- Waaaaaaaaah! Waaaaaaaaah!
- They’re so coddled! And weak! And they want everything to be safe all the time!
- Life isn’t a “safe space” and doesn’t have “trigger warnings”! Grow up!
- Baby! Baby! Baby! Waaaaaah! Waaaaaah! Waaaaaaaah!

These seem like different agendas. In particular, the nobody-blogger angle focuses on ways in which social justice is used to justify aggression, and the mass-media angle focuses on ways in which social justice is used to coddle weakness. Thus the national magazines’ focus on trigger warnings, which happen to be one of the pieces of social justice I really like and have defended at length precisely because they *do* sometimes help weak people.

But there’s another common thread to the mass-media criticism: they’re all about things that happen on colleges and inconvenience college professors. Compare the recent bullying of a fan-artist to the point of suicide because she drew a cartoon character too thin and so was “erasing fat people" – to the recent students at Yale getting angry at an administrator who said she wasn't going to enforce cultural sensitivity on Halloween costumes and so yelling and throwing stuff at her and her family.

I feel for anybody who gets yelled at and has stuff thrown at them, but the first of these two stories seems by far the most important; lots of teenagers commit suicide every year because of bullying, the idea that somebody deserves to die because they picture a cartoon character differently is abominable, and anyone who's been on the relevant parts of the Internet knows this kind of thing is common as dirt. If I were a news editor, I'd consider the first study a much bigger deal. Instead, the second has gone viral in the national media, and the first remains stuck among the same few second-tier sites and SJ-critical nobody bloggers whom these kinds of things are always stuck among. Why?

I worry that the media, especially the online thinkpiece media, overrepresents an insular demographic of Ivy League academics and their friends who spend most of their time on college campuses and don't notice things that don't affect them personally. When people on Tumblr are being bullied to suicide or told that they're garbage or outed or getting death threats, that's *the commoners*. When a Contemporary Perspectives On American Literature professor is inconvenienced, AAAAAAAAH SOCIAL JUSTICE HAS GONE TOO FAR! SOMEBODY WARN SALON.COM!

Or to be even more cynical: social justice was supposed to be Yale's weapon against Caltech and Podunk. But now Yale *students* are using it against Yale *professors* and *administrators*, and now it's a problem. It's like the police beating up city council members with the truncheons they usually reserve for poor ghetto-dwellers; you can bet there will be a newfound concern about police brutality at city council meetings.

And on the one hand, anything that inspires discussion of police brutality at city council meetings is good. Certainly the SJ-critical movement has been stuck on the same side as people a whole lot creepier than self-serving humanities professors, so what’s the problem?

I think that *is* the problem. When creepy white supremacists criticize social justice, they’re at no risk of taking over the wider SJ-critical movement. As the old saying goes, white supremacists are the best argument against white supremacy, and most of them couldn’t take over a blanket fort with a flamethrower. But rhetorically-gifted Yale professors who get thinkpieces published in *The Atlantic* are *exactly* the sort of people who would take over the wider SJ-critical movement, become its most important voice, and define what it means both to the rest of the world and to its own members.

That would be a disaster. Any general knows that you want to hold the high ground, and it really *really* shouldn’t be hard to hold the high ground against the sorts of people who continue to defend bullying someone to suicide because they drew a cartoon character differently than other people. But the mass media seems determined to find a way to yield the high ground and insist against all evidence that it’s punching down. Yale administrators might be the only group more sheltered than Yale students, yet they’re the only group the media seems to have the energy to defend. Worse, media seems to be defending them in a way that attacks activists for being weak and defenseless instead of pointing out when they’re strong and abusing their power.

I’m not saying that there aren’t important arguments to be had about trigger warnings and safe spaces. There are. But they’re only one of many problems, and far from the worst. And if people *must* focus on trigger warnings and safe spaces, I wish they would use one of about a zillion good arguments that don’t involve the pseudo-Nietzschean “You’re all babies! Stop crying, little babies!” tone. All this is doing is granting social justice activists their most dubious claim: that they are trying to use their ideology as a shield for themselves rather than a sword against others (as Popehat brilliantly puts it).

Finally, I think this might be a wake-up call to worry about the role of academia in media more generally. A friend on Tumblr pointed out that Hillary Clinton’s official list of campaign priorities include “ending sexual assault on campus”? Why not just “ending sexual assault”? Studies find that women are *less* likely to be assaulted on college campuses than off them. Isn’t “ending sexual assault on campus” the same kind of priority as “ending murder in gated communities?” Every murder is a tragedy, and murders in gated communities are no exception. But wouldn’t it reveal a lot about who mattered in a society if “end murder in gated communities” was how they framed their anti-murder initiatives?

I worry recent criticism of social justice is revealing the same thing.

2D:4D Ratio And Psychological Traits: Results From The LW/SSC Survey Sample

Introduction

2D:4D ratio is the length of someone's index finger divided by the length of their ring finger on the same hand. It seems to correlate with some kind of prenatal hormone exposure, which makes it a unique way to explore the effects of purely-biological factors on various traits without having to do hormone assays or go through an ethics board.

Past research suggests that men generally have lower 2D:4D ratios than women, and that within each sex people with various stereotypically-male characteristics have lower 2D:4D ratios than people with various stereotypically female characteristics. For example, aggressiveness, penis length in men, and lesbianism in women have all been correlated with more masculine 2D:4D ratio. But there have also been other studies that challenge some of these results, and the whole field is maddeningly inconsistent.

Maybe the weirdest paper along these lines is Madison et al's 2014 paper showing that feminist activist women have more masculinized 2D:4D ratios than other women, which suggests something about how "neurological gender" influences the way people respond to social gender roles. Or something.

I wanted to look at this more, so I included questions about digit ratio in the 2014 Less Wrong Survey and in the 2014 Slate Star Codex Survey. Both surveys also included demographics questions, questions about sex and gender identification, and questions about political beliefs.

Methods and Results

407 people on the Less Wrong Survey and 122 people on the Slate Star Codex Survey gave plausible digit ratio measurements (I rejected measurements outside the range of 0.8 – 1.3 as implausible), for a total of 529 people. Of these, 454 were biologically male and 75 biologically female. Although participants gave digit ratios from both their right and left hands, the differences were not significant and I averaged them together to create a less noisy measure.

The average male digit ratio was 0.972; the average female digit ratio was 0.975. The difference was not significant, didn't trend toward significance and actually was the opposite direction on right vs. left hands.

Average Bem Femininity was 42.2 for men and 45.8 for women. Average Bem Masculinity was 42.3 for men and 40.5 for women. The sex difference in femininity was significant ($p = 0.01$) but masculinity wasn't ($p = 0.19$) though it trended in the expected direction.

On a scale from one (strongly disagree) to five (strongly agree), the average male opinion of feminism was 3.4; the average female opinion was 4.1; the difference was significant at a $p < 0.001$ level. Other political opinions also had gender differences, always with women more liberal, but the gender difference in feminism was the strongest.

In a subsample limited to men, digit ratio did not correlate with Bem sex roles at all, but *did* correlate with positive opinion of feminism ($p = 0.001$). Bem sex role *did* corelate with feminism; more masculine men had lower opinions of feminism ($r = -0.277$, $p < 0.001$) and more feminine men had higher opinions of feminism ($r = +0.277$, $p < 0.001$).

In the opposite subsample limited to women, digit ratio correlated strongly with Bem masculinity ($r = -0.381$, $p = 0.005$) but not with Bem femininity ($r = +0.123$, $p = 0.38$). Opinion of feminism did not correlate with either sex role or digit ratio and there wasn't even a real trend.

I looked at eight other political issues not clearly related to gender. None of these had a clear connection with digit ratio the same way feminism did. One issue, immigration, had a single correlation significant at the $p = 0.05$ level, but was likely a false positive.

There was a limited sample of transgender people in the study; because most participants were biologically male I compared cismen to transwomen. Difference in digit ratio was not significant and showed no clear trend. Bem sex role inventory found transwomen to be much less masculine and much more feminine than cismen (and in fact less masculine and more feminine than ciswomen), but because of the low sample size the trend didn't quite reach significance. Transwomen were more positive towards feminism than cismen ($p = .047$) but not quite as positive as ciswomen. I tried something similar with straight men versus non-straight men and the trends were arguably in the expected direction but didn't come close to significance. There weren't enough biologically female people to be worth running the analyses with that sample.

A binary autism variable (does vs. does not identify as autistic) had little correlation with sex roles or opinion of feminism, but approached significance for digit ratio in men ($p = 0.051$) and achieved significance once men and women were combined ($p = 0.047$). Autistic people had slightly lower (more masculine) digit ratios. I am wary of this result since it was so weak and not replicated on the left hand.

Discussion

This was weird.

The whole point of digit ratio work is that men are supposed to have lower digit ratio than women, and then you use that to determine whether other characteristics are associated with more male-typical or female-typical hormone balances. But we couldn’t even get the basic finding of men being more male-typical than women.

In fact, our entire sample was heavily feminized compared to those of every other study; our *men* had more feminine digit ratios than other studies’ *women*. This is really weird because it’s a sample selected for very high mathematical ability – average IQ of 139, many math and physics PhDs, computer-related jobs as the most common occupations – but mathematical ability is usually linked to more masculinized digit ratio. Likewise, presumably this sample’s much more autistic than normal, but that’s *also* supposed to be more masculinized. I think maybe the measurement technique on the survey predisposed towards overreporting the ratio. Or maybe I screwed up somewhere and divided when I was supposed to multiply. Really that’s all I can think of.

Despite our failure to pick up what should have been the most drop-dead obvious finding, we still got strong signals on psychological traits that should have been much more subtle. The two clear results were a correlation between digit ratio and opinion-of-feminism in men, and between digit ratio and masculinity in women. P-values for both were very low and unlikely to be coincidences despite the multiple tests performed. But the pattern is hard to explain.

I don’t understand the biology of digit ratio very well, but it’s certainly only one part of sexual differentiation; after all, nearly all studies show wide overlaps between men and women in digit ratio, but except for a few intersex people the sexes still clearly differentiate anyway. So maybe women’s level of masculinity is determined mostly by the thing behind digit ratio, and men’s level of masculinity is determined by that plus various other male-specific masculinizing processes?

More annoying is the apparent partial-disconnect between digit ratios, sex roles, and feminism. In men, digit ratio affected feminism but not sex roles. In women, digit ratio affected sex roles but not feminism. The results weren’t ambiguous either. But why would digit ratio make men less feminist if not by making them less gender-conforming themselves? And is it really plausible that digit ratio makes women much more masculine without shifting their opinions toward feminism at all? And remember, this whole area was started by someone who found a connection between digit ratio and feminist activism *in women*. While it’s possible that the original study was really finding that masculine-digit-ratio women were more *activist* rather than more feminist per se, that seems like kind of a stretch. I’m going to blame this one on low sample size and high measurement error until somebody forces me otherwise.

In conclusion, this study was a mess, but somehow managed to find clear signals in weird places anyway. I don’t know.

If you want to see for yourself, you can find the public Less Wrong survey data here and the public SSC survey data here. You *could* also get my haphazard combination of the two here but I would warn against that if you’re really trying to double-check the results as the combination process was one of the most potentially error-prone steps and I’d rather see it independently replicated.

Hardball Questions For The Next Debate

Dr. Carson:

One of your most important achievements as a neurosurgeon was inventing the functional hemispherectomy, a treatment for epilepsy in which the epileptic hemisphere of the brain is severed from the healthy hemisphere and the body, allowing the healthy hemisphere to have full control of the body free from any epileptic interference. Children who get a functional hemispherectomy sufficiently early will be partly paralyzed on one side, but they will mostly be seizure-free.

Standard hemispherectomies remove the epileptic hemisphere from the body, but that tended to cause hydrocephalus, so your technique instead just severed all of its sensory and motor connections, leaving it present but inert.

But an anonymous neuroscientist on Reddit expressed some concern that just as the functional hemisphere seems to develop full independent personhood after the split, so the epileptic hemisphere may do so as well. Obviously it remains impaired by the epilepsy, but it's not seizing *all* the time, so there will still be comparatively lucid intervals.

So my question for you is – what do you think happens to that person who is in an empty hemisphere, locked out of all sensory input and motor control? Do you think they're conscious? Do you think they're wondering what happened? Do you think they're happy that the other half of them is living a happy normal life? Do they sit rapt in unconditioned contemplation of their own consciousness like an

Aristotelian god? Or do they go mad with boredom, constantly desiring their own death but unable to effect it?

Ms. Fiorina:

One of the issues that's played a central role in your campaign is your belief that the Ottoman Empire was the greatest civilization in the world. Certainly their five-hundred-plus year reign was marked by impressive military, political, and artistic achievements. But I want to bring up a particular aspect of Ottoman governance today.

One of the really unique Ottoman innovations was its so-called "millet system", where every ethnicity and religion was almost its own little empire-within-an-empire. For example, although the Ottoman Empire was itself Muslim, Christians within it got their own millet, led by the Patriarch of Constantinople. They made their own laws, which applied only to Christians, settled disputes between two Christian claimants, levied taxes from Christians to pay for Christian-related projects, and generally kept their own people in line. When the Ottoman Empire as a whole wanted something from its Christian population, the Sultan would meet with the Patriarch and they would hammer it out. There were similar structures in place for Jews, Armenians, et cetera.

The past few years have seen an almost unprecedented rise in identity politics in America, usually marked by the claim that the society is using its weight to kick around people of some identity or another. Society is kicking around blacks. Society is kicking around conservative Christians. Society is kicking around bisexuals. They all feel like they're getting the short end of the stick, but a lot of their preferences are mutually exclusive, and it's hard to imagine some kind of centralized government policy that could satisfy any of them.

As an admirer of the Ottoman Empire, you'd be in a uniquely good position to import some of the advantages of the millet system into the modern Western world. Obviously this would be complicated given all the conflicting identity claims and the close quarters in which everyone is intermingled, but there are already some visions of what it could look like – including my own Archipelago – and if it were raised to the level of a national discussion, people could no doubt come up with many more.

So my question for you is – weren't you a pretty crappy CEO?

Mr. Bush:

Assume that fitness-to-be-President is a normally distributed trait with known heritability. Suppose also that past elections have 100% efficiency; that is, they always choose the most qualified candidate. We can then use some of the standard regression-to-the-mean equations to determine the chances that the highest fitness-to-be-President individual in generation G will be the offspring of the highest fitness-to-be-President individual in generation G-1.

The single most fit-to-be president man in a population of 300 million would be about six standard deviations above the norm. If that man breeds with the single most fit-to-be-president woman, and if in keeping with findings for other complex traits heritability is about 60%, we would expect their offspring to be about 3.6 standard deviations above the mean in fitness-to-be-president. One in every 2500 or so people is 3.6 standard deviations above average, meaning there would be at least 120,000 equally good or better presidential candidates than they in the United States.

How high would the heritability of presidential fitness have to be before there was at least a 10% chance that the offspring of the two most presidential Americans was himself presidential material? My calculations suggest about 90%, which is very high compared to what we know about similar traits – but actually not entirely outside the realm of plausibility.

But if a maximally-presidential man breeds with a woman who is less than maximally presidential, the odds fall precipitiously. Suppose that a maximally-presidential man breeds with a woman who is merely in the 99th percentile for presidential ability. Now given a heritability of 60% there will be *three million* Americans more presidential than their average offspring. Even given a 100% heritability, there is only a 1/73 chance that their offspring will themselves be worthy of the presidency.

So my question for you is: do you think Barbara Bush is an unrecognized political super-genius, or are there probably hundreds of thousands of Americans who would make a better president than you would?

Senator Cruz:

You were on your college debate team, and you were good at it. Really good. You won the national championships and you were pretty widely believed to be the best debater in the country. Quite an achievement. But my worry is – which is more likely? That the best debater in the country would also be the best choice for President? Or that he would be really really really good at making us *think* that he would be?

Don't respond yet. Before you answer that question – well, before you answer any question – we've got to think about this on the meta-level. There's a classic problem in epistemology. Suppose that we have a superintelligence with near-infinite rhetorical brilliance. The superintelligence plays a game with interested humans. First, it takes the hundred or so most controversial topics, chooses two opposing positions on each, writes the positions down on pieces of paper, and then puts them in a jar. Then it chooses one position at random and tries to convince the human of that position. We observe that in a hundred such games, every human player has left 100% convinced of the position the superintelligence drew from the jar. Now it's your turn to play the game. The superintelligence picks a position from the jar. It argues for the position. The argument is supremely convincing. After hearing it, you are more sure that the position is true than you have ever been of anything in your life; there's so much evidence in favor that it is absolutely knock-down obvious. Should you believe the position?

The inside view tells you yes; upon evaluating the argument, you find is clearly true. The outside view tells you no; judging from the superintelligence's past successes, it could have convinced you equally well of the opposite position. If you are smart, you will precommit to never changing your mind at all based on anything the superintelligence says. You will just shut it out of the community of entities capable of persuading you through argument.

Senator Cruz, you may not quite be at the superintelligence level, but given that you've been recognized as the most convincing person out of all three hundred million Americans, shouldn't we institute similar precautions with you? Shouldn't your supporters, even if they agree with everything you are saying, precommit to ignore you as a matter of principle?

Senator Rubio:

When you became Florida’s Speaker of the House, one of the other men on stage here tonight, Jeb Bush, presented you with a golden sword, which he said was the “Sword of Chang”. He told you that “Chang is somebody who believes in conservative principles, believes in entrepreneurial capitalism, believes in moral values that underpin a free society. Chang, this mystical warrior, has never let me down.” You looked pretty excited about it.

Now, some might say that this all came from a giant misunderstanding. Back in the late 1940s, Mao Zedong’s victorious Chinese communists forced Chiang Kai-shek’s defeated Chinese nationalists to retreat to the island of Taiwan. The United States kept the peace in the the Taiwan Strait, mostly to prevent Mao from invading and finishing the job, but a common refrain in 1950s conservativism went that we should “unleash Chiang”; that is, advise Chiang Kai-Shek to go back across the strait and reconquer China. George H. W. Bush served as envoy to China, had to listen to this sort of stuff, and got annoyed enough at the “unleash Chiang” rhetoric that he would quote it ironically at bizarre times, like his documented habit of threatening that his serve would “unleash Chiang” on his tennis opponents. It’s unclear how we got from George H. W. Bush’s constant threats to “unleash Chiang” on people, to his son’s belief that Chang was a mystical conservative warrior. Maybe it was a joke, either Bush Sr. pranking Jeb or Jeb pranking you.

In any case, you hung the sword in “a place of honor in your office”. From that point forward, Jeb’s fortunes declined. He left the Florida governorship, failed to get any further high positions, and then ran a very lackluster Presidential campaign. But from that same point your own fortunes decidedly rose. You started a law firm, were appointed a professor, got elected to the Senate, and are currently running a spectacular Presidential campaign with most pundits betting on your eventual victory after Trump and Carson lose their shine. The connection between the transfer of the sword and the sudden switch in both your fortunes is so striking that even the Huffington Post, not normally a source for magic-sword-related journalism, wrote about it: Jeb’s Last Hope – Reclaim the Sword of Chang.

But here we have a conundrum: if there was never a mythical Chinese warrior named Chang, by what magic does this sword grant worldly success to its possessor and ignomious ruin to any who lose it? There is a legend that fits almost exactly: the tale of the Holy Lance, aka the Spear of Destiny, aka several other portentious sounding names. According to the story, this relic from Christ’s crucifixion grants victory to all who own it and swift ruin to all who lose it. Charlemagne was reputedly the first to make use of its power; he was unstoppable while he wielded it but died moments after dropping it during battle. The same pattern repeated with Frederick Barbarossa, then a host of other military leaders, until finally it passed to the Austrian Habsburgs. They realized its power, locked it away, and ended up winning the greatest empire in European history. Supposedly Hitler was obsessed with it, so much so that his fascination with the object inspired the depiction of Nazi archaeologists in *Raiders of the Lost Ark*, and he took it for himself after the Anschluss. As the war wound down, the relic caught the special attention of General George Patton, who brought it back safely to Vienna afterwards. But ever since that time there have been various rumors that it was a fake, and that Nazi sympathizers took the real Lance in preparation for the time when the Reich would rise again.

The book *Secrets of the Holy Lance* describes one possible route by which the artifact might have been smuggled out of Europe:

> Reporters John Buchanan and Stacey Michael cite recently declassified documents from the US National Archives that indicate that Prescott Bush “failed to divest himself of more than a dozen enemy national relationships that continued until as late as 1951. Bush conducted business following the end of World War II with moving assets into the Nazi refuges of Argentina, Panama, and Brazil.

So Prescott Bush was involved in moving Nazi “assets” from conquered Europe to South American refuges, presumably including the true Lance. Far be it from me to impugn his business ethics, but I don’t remember Nazi refugees in Argentina becoming an unstoppable force aided by a weapon of legendary mystical power. On the other hand, I *do* remember Prescott Bush being elected to the United States Senate just a few years later. Then his son *and the presumed heir of his property* was elected US President. Then *his son* was *also* elected US President. I need not add that according to the the laws of genetics, the chance of this happening by coincidence is hundreds-of-thousands to one even assuming implausibly high heritability of the fitness-to-be-president trait. Then his *other* son starts rocketing up through the ranks *right up until the moment he gave you the sword of Chang*, a sword named after a weird Bush family in-joke about a Chinese mystical warrior who doesn’t exist.

I think we can start to sketch out a plausible explanation here. Hitler didn’t want the Holy Lance falling into the hands of his enemies, so he replaced it with a fake and hired Nazi-artifact-smuggler Prescott Bush to transport the real one to safety in South America. Bush realized what he had, handed the South Americans a *second* fake, and kept the real one for himself, reforging it from a lance into a sword to cover his tracks – an action entirely in character for Prescott Bush, whose other relic-stealing adventures include the theft of Geronimo's skull. He died unexpectedly without getting the chance to explain the significance of the artifact to his son George H. W. Bush. But since it seemed like a sentimentally important heirloom, George took care of his father's weird golden sword anyway. When his sons asked him about it he didn't have a real answer, so he just made his favorite in-joke about "unleashing Chiang", and they believed him. Then eventually it passed to George W, later on to Jeb, and then Jeb thought it would be a funny present to give you to honor your election as Florida speaker.

Obviously the Lance is a significant strategic asset for America, and I imagine if you were President then its aura of victory would apply to the country as well, much as the Habsburgs' possession of the lance enlarged Austria-Hungary. However, its powers are generally held to come from the Antichrist.

So my question for you is, do you think it's ethical to use your magic sword to channel the power of the Antichrist if that would ensure America's military success?

Mr. Trump:

You are famous both for your vast corporate empire and for your tendency to name the pieces of that corporate empire after yourself. By my count there are six buildings named "Trump Tower", ten named some variation on "Trump Hotel", a Trump Building, a Trump Palace, and a Trump Estate. You founded a financial services group called Trump Mortgage, a modeling agency called Trump Model Management, a bottled water brand called Trump Ice, and a magazine called Trump Magazine. You also started an airline called Trump Airlines, a TV company called Trump Productions, a book series called Trump Books, and your own radio talk show called Trumped!. There are also several Trump-themed games, like Donald Trump’s Real Estate Tycoon and Trump: The Game.

Mother Jones wrote a great article on this last one. Trump: The Game seems to be a tacky Monopoly clone. Players move around a board and bid on properties, and when one of them gets locked out of bidding for a property the other player gets to say “YOU’RE FIRED” the same way you do on your show. The only way to get back in to a property once you’ve been fired is to use the game’s most powerful card, which has a picture of your face on it and is called “The Donald”.

My question for you is: WHY DIDN’T YOU CALL IT THE TRUMP CARD?!?!!!!111111111asdfdf

SSC New England Meetup And Presentations Schedule

1. Tonight (11/19) I'll be speaking at Yale on the subject of "How To Ruin A Perfectly Good Randomized Controlled Trial" and other aspects of interpreting scientific evidence. Probably. I'm currently stuck in New Jersey due to airplane mishaps, but 90% confidence I'm going to be able to make it on time. Event is at LC101, Linsly-Chittenden Hall, 63 High St, New Haven, at 7 PM. You're invited.

2. Monday (11/23) I'll be speaking at Harvard at Sever Hall 203 at 5 PM on the same subject. You're invited to this one too.

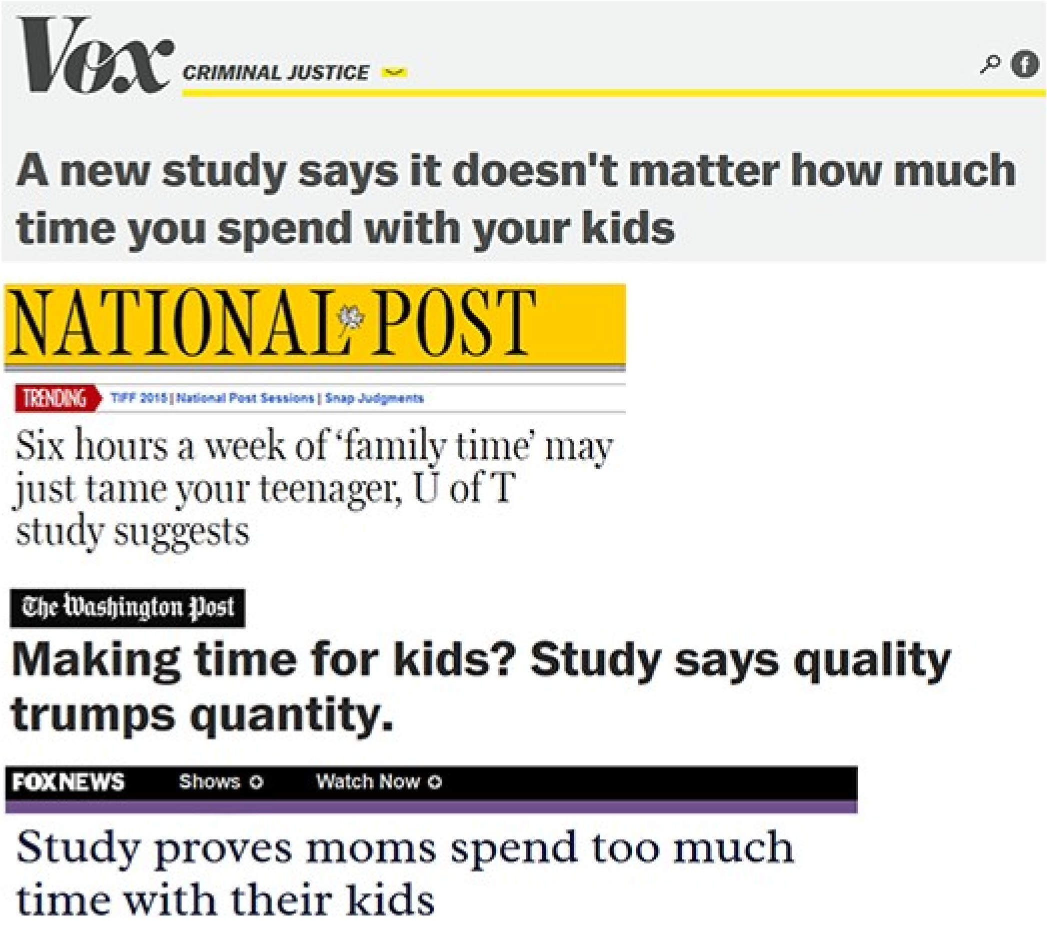
3. Sunday (11/22) we'll be having a Boston meetup (thanks to the Boston rationalist and EA communities for setting this up) at MIT, room 5-134 at 6 PM. There is a Facebook event page here. You are invited to all of these things but you are **extra definitely** invited to this one. Every time I try to have a meetup I specify that everyone is invited even if you are only a lurker and even if you don't understand everything you see on SSC and even if you are not very interesting and so on and so forth, and every time people still say "I wanted to go but I didn't because I'm afraid I wouldn't have been welcome there because I'm not enough like the typical SSC reader," and so now I have to repeat **in bold** that you are definitely welcome.

I won't be around much for the next week or two while I do this stuff and see my family for Thanksgiving, so apologies for not responding to emails/comments, etc.

Reporter Degrees Of Freedom

I.

A sample of Thursday’s talk at Yale:


Vox
CRIMINAL JUSTICE

A new study says it doesn't matter how much time you spend with your kids

NATIONAL POST
TRENDING
TIFF 2015 | National Post Sessions | Snap Judgments

Six hours a week of ‘family time’ may just tame your teenager, U of T study suggests

The Washington Post

Making time for kids? Study says quality trumps quantity.

FOXNEWS
Shows
Watch Now

Study proves moms spend too much time with their kids

These are four headlines describing the same study, Milkie, Nomaguchi and Denny (2015). The study found that of twenty or so outcomes, only three of them – all measuring delinquent behavior among teenagers – show significant effect from time spent with parents (and this result remains after Bonferroni correction). So Vox has a great argument for their headline. The National Post has an okay argument for their headline even though it’s kind of cherry-picked. The

Washington Post just sort of reads between the lines and figures that if it's not quantity of time that helps kids, it must be quality. And FOX also reads between the lines and figures that if moms spending time with their kids has no effect, the argument from opportunity costs suggests mothers are spending too much time with their kids.

None of them are completely outright lying. And indeed, most of the articles eventually explain what I just said, halfway down the article, in one or two short sentences that most readers will skim over. But the *rest* of the article uses the study to support whatever the news source involved wants it to support, and so people will come up with four diametrically opposed conclusions from this one study depending on which source they read.

II.

Here's a study that I wasn't able to include in the presentation because it just came out recently. As per the Rice University press release: Overweight Men Just As Likely As Overweight Women To Face Discrimination.

The paper included two studies. In the first, men went into stores either with or without fat suits and try to do some things – ask if there were job openings, ask for a job application, ask an employee for help, try to buy some things, et cetera. Then they measured the men's success across both conditions to see if they had more trouble when they appeared overweight.

In the second, subjects were asked to rate videos of an employee giving a marketing spiel for a new product; once again, the employee was either wearing or not wearing a fat suit. They measured the subjects' ratings across conditions to see if they ranked the overweight employees lower.

The first study only included men, and so could not possibly have determined whether men were more or less likely than overweight women to face discrimination. The second study actually *did* have both male and female employees involved, and although it really wasn't their main interest, the researchers did a post hoc evaluation to find the effect in each sex. In all three of the outcomes where discrimination was found, women faced more discrimination than men. They didn't significance-test the comparison, but just from eyeballing it, it was probably significant.

So a paper in which one study does not compare men to women, and the other study finds women facing more discrimination than men, the press release somehow gets phrased as "Overweight men just as likely as overweight women to face discrimination in retail settings". Huh.

You might wonder, "Does it really matter what a press release says? Does anyone read the exact wording?" Yes. Many other news sources copied the phrasing, for example Medical Daily's Fat Discrimination Is The Same Regardless Of Gender. One such copycat, PsyPost.com copied the press release nearly word for word, including the title. Then it got posted on Reddit and now has 5189 upvotes and 1572 comments. So there's that.

III.

"But at least it correctly raised awareness of how weight discrimination is a big problem in the retail setting, right?"

The paper measured a *ton* of different outcomes. Let's focus on Study 1. The actor in the fat suit was supposed to ask if there were job openings (there were) and see if the company told him. Then he was supposed to ask for an application form and see if they gave it to him. Then he was supposed to walk in as a customer and see if employees greeted him. Then he was supposed to ask the employees to recommend him an item and see if they did. Then he was supposed to ask them to recommend him a second item and see if they did.

No difference was found between overweight and normal-weight actors in any of those five experiments. Two of them had ceiling effects that probably made the attempt futile, but the other three didn’t, and there wasn’t even a trend toward discriminating against the overweight guy.

So what *did* they find discrimination on? They say that detected “interpersonal discrimination”, ie discrimination based not on any quantifiable outcome but based on how friendly/warm the person interacting with the actor seemed toward him. They determined this by self-rating and other-rating; that is, the actor wrote down how friendly he thought the store clerks were toward him, and a ~~spy~~ surreptitious observer who had placed herself near the interaction also rated this for corroboration. Their rating scales included twelve items including “how many times did the clerk nod”, “how friendly did the clerk seem?”, “how much eye contact was the clerk making?” and “how much comfort level did the clerk seem to have?”. The experiment found a statistically significant difference between the fat-suit-wearing and non-fat-suit-wearing trials and concluded that there was interpersonal discrimination.

But hold on a second! The study says *nothing* about *anyone* being blinded. In fact, it’s *really hard* to blind an actor to the fact that he is going into some stores while wearing a fat suit and other stores while not wearing a fat suit. As far as I can tell, everybody involved was in on the study from the beginning. If your boss tells you “I want you to rate how much comfort level clerks have with you for this study on fat discrimination”, it seems really possible to me that there might be a slight tendency to overrate the clerks who interacted with thin-you, and to underrate the clerks who interacted with fat-you.

How slight a tendency? Clerks dealing with fat people got an average rating of 2.3 (seven point scale, lower is better), and those dealing with thin people of 2.0.

So after finding no discrimination on five objectively measurable outcomes, they find very subtle discrimination on an unblinded subjective outcome practically designed to produce placebo effects.

We move on to the second study, where participants (as usual, psychology students) are rating video presentations given by fat vs. thin people. This is supposedly tying into the “retail industry” theme of the paper, but honestly it seems kind of forced to me.

Anyway, the participants are asked to rate their presenters on seven measures: overall quality of presentation, overall attitude toward product being presented, overall attitude toward the store that would employ a person such as this, intention to support the store, employee’s appearance, employee’s carelessness, and employee’s professionalism. The results:

There was no difference between how participants rated overweight vs. normal-weight presentations overall.

There was no difference between how participants rated products presented by overweight vs. normal-weight people.

There was no difference between how participants rated stores staffed by overweight vs. normal-weight people.

There was no difference between how likely participants were to support stores staffed by overweight vs. normal-weight people.

There *was* a difference in how participants ranked the appearance, carelessness, and professionalism of overweight vs. normal-weight people.

The first four results are encouraging. What about the last three?

Well, I feel like if you ask people to rank someone based on “their appearance”, and your subjects answer based on how they look, you kind of walked into that one. Oh no, people rank conventionally attractive people as having better appearances than less conventionally attractive people! Someone call John Ioannidis to double-check this astonishing result!

“Carelessness” and “professionalism” are perhaps less excusable, but c’mon, you had them watch a two-minute video. When you give someone zero information on a thing, and you force them to make a judgment on the thing, then yes, stereotypes are their best source of information. If you showed me a picture of an average-looking man and an average-looking woman and say “Quick! Which of these people is more likely to like baking cupcakes?!” I’ll pick the woman, not because I think all women are obsessed with cupcakes or because I go around looking at every woman I see as a cupcake factory, but because you asked me a stupid question and ensured stereotypes were the only thing I had to go on.

Then the authors find that this was mediated by explicitly-expressed stereotypes against fat people, which is kind of interesting, but doesn’t make the nonsignificant things any more significant.

So to sum up: there was no discrimination against the overweight on any objective measure of the actual retail experience, including positions advertised, applications given, greetings offered, or customers served. There was also no discrimination against the overweight on presentation evaluations in terms of overall evaluation, evaluation of employee, evaluation of product, or evaluation of company.

There *was* a tiny amount of discrimination on a subjective measure rated by unblinded observers aware of the purpose of the study. There was also some evidence on three subtler ratings of the presentation that seemed designed to ask participants impossible questions in order to force them to stereotype. However, these meaningless scales did not effect the raters’ overall impressions as measured any of four different ways.

IV.

Here is the reporting from the news outlets that passed their first test and *didn’t* frame it as men and women facing equal amounts of weight discrimination.

Business Insider: Researchers Had Men Pretend To Be Obese – And The Results Are Disturbing, which says that “this research highlights the importance of including men in discussions about weight stigmatization,” and “the authors also advocate organizational efforts to combat negativity against heavy customers and potential employees…the first step may be for individuals to become aware of how strong weight biases are.”

AskMen: Young Men Who Appear Overweight Suffer Interpersonal Discrimination. “Researchers disguised six thin young men as obese customers or job applications and found that they were victims of microaggressions. Basically people were a little bit more jerky towards them.”

Oximity: Overweight Men Often Snubbed At The Mall. “Shopping malls can be hostile places for overweight men, regardless of whether they’re customers or simply looking for a job.”

The Health Site: Men, Here’s One More Reason For You To Lose Weight. “Ruggs said that these findings were another reminder that there was still more work to be done in terms of creating equitable workplaces for all employees, potential employees and consumers. She concluded that this was something organisations could take an active role in, and said that companies could do better job training on customer relations as part of the employees’ new-hire process. ”

Now I’m almost missing the kind of random scattershot media bias we found on the time-spent-with-children study. Here every media outlet reports the results the same way that the study’s author and the press release reports the results.

This is not a *totally* wrong interpretation, any more than “six hours a week will tame your teen” is a totally wrong interpretation of the childhood study. But if I myself were writing an article on this study, it would be SURPRISINGLY LITTLE DISCRIMINATION FOUND AGAINST OVERWEIGHT MEN, and mention somewhere in the middle that some discrimination was found on a few sketchy variables. Instead, we get DAILY DISCRIMINATION AGAINST OVERWEIGHT MEN and WE NEED TO INSTITUTE SENSITIVITY

TRAINING PROGRAMS IN RETAIL ENVIRONMENT, and they mention somewhere in the middle that a lot of important variables came out negative.

A *lot* of studies work like this. You test ten or twenty complicated variables, you get positive results on some, negative results on others, some of those results seem plausible, other results seem like maybe you made a mistake somewhere or didn’t have enough power or whatever, and then you make an interpretation based on your personal bias. Then it goes from the researcher’s personal bias to the abstract to the press release to the headlines to the mind of the average reader, dropping subtlety at each step, until “No discrimination against overweight men, except where the study was practically designed to ensure false positives” becomes “Rampant discrimination against overweight men everywhere” becomes “Overweight men are discriminated against just as much as overweight women.”

Don’t get me wrong. I expect there probably *is* lots of discrimination against overweight men. And I think this study’s project of trying to find it and convince people of its existence was worthwhile. But I don’t think you should get to convince everyone that science has proven X, unless science has *actually* proven X. The process that produced these headlines is strong enough to produce any headline you want, with the part where you actually do the study becoming more and more of a ritual or a formality. There are just too many degrees of freedom between the study and the reporting.

Stalin once said that “those who vote decide nothing; those who count the votes decide everything.” It’s starting to look like those who do the studies decide nothing and those who report the studies decide everything. The only solution is to actually read the study and not just the headlines. Sometimes we might even have to – God help us – read beyond the abstract.

College And Critical Thinking

[epistemic status: My bias is against the current college system doing much good. I have tried not to be bogged down by this bias, but take it into account when reading my interpretations below.]

*[**EDIT:** An earlier version of this post claimed that one paper had shown a u-shaped relationship between time spent in college and critical thinking. A commenter pointed out this was true only of a subset in two-year colleges, but not of four-year colleges or college in general – which shows the expected linear relationship. I am sorry for the error, and correcting it somewhat increases my confidence in college building critical thinking.]*

Over Thanksgiving, I was discussing tulip subsidies with the pro-Bernie-Sanders faction of my family, and my uncle claimed that we needed college because "it teaches you how to think critically".

The evidence sort of supports him, but with the usual caveats and uncertainties.

First of all, what the heck is critical thinking? Luckily, we have a very objective scientific answer: critical thinking is the ability to score highly on the Watson-Glaser Critical Thinking Appraisal. Now that we've answered that, we can move on to the question at hand.

Most studies on this issue are terrible because they lack control groups. That is, they measure students when they enter college, measure them again when they leave college, and find that their critical thinking ability has improved. But this could be for any number of reasons. Maybe older people generally have better critical thinking than younger people. Maybe life experience builds critical thinking. Maybe college had nothing to do with any of it. The best meta-analysis of such studies, MacMillan 1987, finds exactly this, and concludes:

> Overall these studies suggest that seniors, in the main, are probably better at critical thinking than freshmen. However, since the most compelling data were gathered through weak pretest-posttest or longitudinal designs, it is difficult to separate out the effect of college from the maturational effects that occur despite college.

I like the phrase "maturation that occurs despite college", although I don't think they meant it in *that* way.

But in any case we need a better study design to conclude anything from this. There are two studies with moderately good designs, both by a guy named Pascarella. The first compares 30 college students to 17 matched non-college students and follows them up for one year. They find that while both students and non-students gain critical thinking skills over the course of the year, the college students gain 17% more, corresponding to an effect size of 0.44.

The second, larger study compares students doing college full-time to students doing college part-time, under the theory that if college is causing the effect, then a little college should cause a small effect, but lots of college should cause a big effect. They find this in the four-year college sample, and a garbled u-shaped mess in the two-year college sample. At least the four-year sample, which is what most people are interested in, looks good.

On the other hand, some other studies find less impressive effect sizes. Arum and Roska recently wrote a book on this kind of thing, *Academically Adrift*, and they find that two years of college (start of freshman to end of sophomore) only increases critical thinking by 0.18 SD. This is weird because it's twice as much college as the past few studies but less than half the effect size. Also, it doesn't seem to be controlled, so this is the *sum* of actual-college-effect-size and confounder-effect-size. According to one review:

> College entrance to end of sophomore (ie half of college) improves critical thinking by 0.18 SD. This might have been greater in the past "Pascarella and Terenzini estimated that seniors had an 0.5 SD advantage over freshmen in the 1990s. In contrast, during the 1980s students developed their skills at twice the rate: seniors had an advantage over freshmen of one standard deviation."

Note that we're comparing unlike with unlike. Four years of college need not produce an effect twice as great as two years of college, any more than a space heater that increases the temperature of a room 10 degrees after being left on for one hour will increase the temperature 87600 degrees after being left on for a year. Indeed, some studies suggest that most of the gains happen in freshman year. But even so, there's a very clear downward trend here. As usual, we have no good way of knowing if that's caused by gradually-improving studies, gradually-improving college critical thinking training, or gradually-deteriorating colleges.

One more question: do we know the specific aspects of the college experience that cause critical-thinking gains?

Specific explicitly-advertised "critical thinking classes" don't. Being a college that prides itself on a specific "critical thinking focus" doesn't. This sort of thing seems very well replicated, although there are a few individual studies that don't really interact with the rest of the literature that seem to have at least temporary positive results. Classes that you might think would teach critical thinking, like logic, or philosophy, or statistics, don't seem to do especially well (here's a specific in-depth discussion of philosophy). The only positive result anyone’s been able to find is that “liberal arts” (here viewed as a broad category *including* science and mathematics) seems to do better than occupational skills, and “education, social work, communications, and business” seem to do worse.

In terms of broader factors, – one study finds that quantifiable college experiences explain “between 7 and 17%” of the variability in first-year critical thinking gains (other sources say classes explain 2.5% and extracurriculars 2.9%). Studying a lot seems to help. So does reading *unassigned* books. Aside from that, the biggest finding is kind of concerning:

> Students who characterized their relationships with other students as “competitive, uninvolved…alienated” were more likely to show gains in critical thinking than were students who portrayed their peer relations as “friendly, supportive, or a sense of belonging” The data in this study do not permit confident explanation of this relation, but one might speculate that a sense of participation in a friendly, supportive peer environment may require a partial suspension of one’s critical thinking skills.

…wow. I was going to say something like “students busy spending time with their friends have less time to learn stuff”, but your cynical awful explanation works too.

So what’s the big picture?

Well, we know that people will gain critical thinking skills during the four years from age 18 to age 22. We have an small study that finds college helps a little with this process and a larger study that shows dose-dependent effects of college. We have some hard-to-compare effect sizes ranging from 0.18/2-years to 0.44/year. That’s modest but appreciable, and it’s probably at least somewhat real.

But those of you who went to my talk last week hopefully know what my next question will be: *how long does this last?*

For example, we know that parents’ personalities have all sorts of interesting effects on their children *while those children are living with their parents*. We also know that as soon as children leave their parents, those effects go down to near zero. What people tried to interpret as some deep fact about development was actually just a reflection of the environment that those children were in. Likewise, preschool makes children do much better in kindergarten, but by third grade the preschool-educated kids are doing the same or worse as the others. It’s not fundamentally altering their developmental trajectory, it’s just creating an short-term effect.

Every one of the studies I’m citing here was done in freshman or sophomore year (one study cited another done in senior year, but I couldn’t find it directly). No one has ever looked at students who have been out of college a year – let alone out of college thirty years – to see if the effect continues. I would bet that it doesn’t.

Until somebody checks, enjoy your opportunity to tell people that the evidence backs college building critical thinking skills.

Setting The Default

[Epistemic status: I predict everyone except me will respond to this with "Duuuuuuuuh", but I found it changed some views of mine]

I.

I recently did couples therapy with two gay men who'd gotten married a year or so ago. Since then one of them, let's call him Adam, decided he was bored with his sex life and went to a club where they did some things I will not describe here. His husband, let's call him Steve, was upset by what he considered infidelity, and they had a big fight. Both of them wanted to stay together for the sake of the kids (did I mention they adopted some kids?) but this club thing was a pretty big deal, so they decided to seek professional help.

Adam made the following proposal: he knew Steve was not very kinky, so Adam would go do his kinky stuff at the club, with Steve's knowledge and consent. That way everyone could get what they wanted. Sure, it would involve having sex with other people, but it didn't mean anything, and it was selfish for a spouse to assert some kind of right to "control" the other spouse anyway.

Steve made the following counterproposal: no. He *liked* monogamy and fidelity and it would make him really jealous and angry to think of Adam going out and having sex with other people, even in a meaningless way. He argued that if Adam didn't like monogamy, maybe he shouldn't have proposed entering into a form of life that has been pretty much defined by its insistence on monogamy for the past several thousand years and then sworn adherence to that form of life in front of everyone they knew. If Adam hadn't liked monogamy, he had *ample* opportunity to avoid it *before* he had bound his life together with Steve's. Now he was stuck.

Adam gave the following counterargument: yeah, marriage usually implies remaining monogamous, but that was all legal boilerplate. He had wanted to get married to symbolize his committment to Steve – committment that he still had! – and he hadn't realized he was interested in fetish stuff at the time or else he would have brought it up.

Steve gave the following countercounterargument: okay, this is all very sad, but now we *are* stuck in this position, and clearly only one of the two people could get their preference satisfied, and given the whole marriage-implies-monogamy thing, it seemed pretty clear that that person should be him.

So then of course they both turned to me for advice.

The rules for psychotherapy are a lot like the rules for Aaron Burr: talk less, smile more, don't let them know what you're against or what you're for. This last principle is generally known as "therapeutic neutrality", and it demands that we not take sides in our patients' disputes or dilemmas. Instead, we try to remain an impartial discussion facilitator, teasing out our patients' true values and concerns until they are able to come to a conclusion on their own.

On the other hand, their friends probably don't have the same scruples and are going to be offering them advice. And I wondered what advice I would give, if I were their friends and not stuck in metaphorical Switzerland.

II.

Assume Steve's analysis is right; this is a zero-sum game and there's no way for them both to come out of it happy.

Which side do we choose?

A quick retreat to a simpler situation: suppose Adam really wants to keep all the windows in the house open all winter with no heat on, so that the inside temperature is 10F and the house is full of snow. Steve does not want to do this. Both of them want to stay together for the sake of the kids, but this do-we-freeze-our-house thing is really getting in the way.

This problem is easy. Adam, you're crazy and your preferences are stupid and don't count. Suck it up and keep living with Steve at normal-person temperatures.

Another retreat in the other direction: suppose Adam wants to sometimes take a shower, but for some reason the thought of Adam being in a shower pisses Steve off and he refuses to allow it. Once again, both of them want to stay together for the sake of the kids, but this can-Adam-take-a-shower thing is really getting in the way.

This problem is also easy. Steve, this time *your* preferences are stupid and don't count. Suck it up and let Adam take a shower.

A third new situation. The one Unit of Caring recently discussed on her blog. A transwoman wants to have Christmas with her family, but her family doesn't believe in transgender and insists on calling her by her original male name and male pronouns. Both her and her family don't want to "ruin Christmas" by refusing to get together as a family or making a big deal of this. Who is in the right? Unit of Caring writes:

> Anyone who does not respect their siblings enough to call them by the name of their choosing does not then get to go "oh! you not wanting to let me repeatedly hurt

> you is breaking apart our family, how unreasonable of you!" If you want people to spend time around you, call them by the names they chose. If you wouldn't repeatedly slap your siblings in the face, don't deliberately misgender them either. (And if you would repeatedly slap your siblings in the face, then you shouldn't have to look too far to figure out whose behavior ruined Christmas.)
>
> If it doesn't bother you that you're hurting someone, then you don't get to act wronged when they decide you're not worth spending time with.

I agree with this assessment, but only because I agree with Unit about the object-level issue of transgender. It seems like if you wrote in the same question to your local priest, they'd say the trans woman was being unreasonable. I don't think there's any good way for Unit and the priest (or the woman and her family) to resolve their differences except by one convincing the other of their position on the object-level issue of transgender.

(well, if you really really really understood utilitarianism, you might be able to say you should take the highest-utility solution, but no one understands utilitarianism that well)

This seems to be true of my patients' problem too. Unless we can decide whether wanting to go to a fetish club and have sex with people besides your husband is a reasonable request, we can't solve Adam and Steve's disagreement. I mean, Steve's argument about the contract isn't *bad*, but if it were something we disagreed with – let's say some old-timey marriage contract where the woman vowed to always serve and obey her husband, and now she's a feminist and wants out – we would probably be pretty sympathetic despite the precise wording of what she'd "agreed" on.

III.

I come to the table with personal baggage. I come from a very permissive subculture. I've had some very happy open relationships and wanting to be open seems like a reasonable request. I've had some friends who are very kinky, and wanting to be kinky seems like a reasonable request too. I'm not personally very good at feeling jealous, so wanting your husband to never go to a club, even if he doesn't tell you about it, or make you think about it, or even agrees only to do it when you're away on a business trip in another city – seems a bit odd. Honestly I would be tempted to take Steve aside and ask him whether he's *sure* that he couldn't deal with Adam going to this club, and whether maybe he wants to give it a chance, and whether maybe he just wants what's best for Adam even if that makes him a little uncomfortable.

But go back two hundred years and ask the people of *that* culture, and this choice is a no-brainer. Fetish clubs (or the closest 19th century equivalent) are weird, vile, sinful things, and Adam's desire to go to one is totally beyond the pale. He should never even have made the request. But since he did, we can strongly and clearly tell him that this is morally wrong, that he should apologize to Steve for the trouble he put him in, that he should realize there's more to life than kinky sex, and that he should want what's best for Steve even if that means he can't satisfy his libido quite so much.

If Adam and Steve were in the traditional culture of the 1800s there would be no debate. If they were in some ultra-permissive sexually-open subculture of the 2100s, there would also be no debate. The culture would tell one of them that they were wrong, just like someone who wants to make the other live in a 10 degree frozen house is wrong, that person would grudgingly agree, they would stay together, and that would be that. The problem only comes when they're in a culture with a lot of different subcultures that haven't made up their minds yet. Like ours.

We all hear the stories of the economists who start by assuming perfect rationality, and then add in deviations from that assumption when they come to them. I kind of like to start from a liberal assumption of perfect atomic individualism and add in deviations when I encounter them. And, well, this is the latest one I encountered.

Adam and Steve's individual personalities and situations will help resolve their conflict, but the tiebreaker vote is always going to be cast by the culture around them. Realizing this has made me more open to activists who are trying to change the culture – and, symmetrically, to conservatives who are trying to prevent the culture from being changed. People with unusual sex lives like to say that what they do in the privacy of their own bedroom doesn't hurt anyone else – neither breaks their nose nor picks their pocket – but the fact is that the partial social acceptance of fetish clubs and of open relationships is what gives Adam a leg to stand on. And some religious conservatives like to talk about how they only want to defend their own right to practice and express their beliefs instead of being forced into the broader cultural revolution all around them, but the fact is that their beliefs are what's supporting Steve. My sympathies will always be with the atomic individualists who want to come up with some clever Adam-Steve contract that solves their problem on the meta-level as long as all actors are rational, but I am starting to worry the culture warriors have a point here.

UR said that "the sovereign is the one who sets the null hypothesis". Once you've let the culture set a default – going to fetish clubs is a reasonable request, going to fetish clubs is an unreasonable request – then given sufficiently good liberal norms people who want to deviate from the default can absolutely do so, but as soon as a conflict springs up the identity of the default option still matters a lot.

I’m not suggesting a total war of all against all, and there’s always the Archipelago option, but I guess sometimes culture wars *do* need to be fought beyond the point where you just leave people alone, if only to shift the default in your direction.

Speaking of culture wars, an apology to gay people. I always obfuscate details about my patients to disguise their identities, but I feel particularly bad about making this couple gay because it reinforces the stereotype of gay people as hypersexual and bad at committment. I made them gay anyway, because when I tried to write them hetero, their gender seemed to skew the problem too much to one side or another – for example, when Steve was a woman, he was the poor innocent wife wronged by a horny husband who insisted on thinking with his crotch. I worried that if I made the couple hetero, my readers for one reason or another would bring their own baggage and wouldn’t be able to see it as the difficult and evenly-balanced problem it seemed like when I was in the office with them.

Which itself says something about how our culture sets default hypotheses.

Product Recommendations 2015

It's Christmas shopping season, and so time for the annual reminder that if you want to support this blog you can shop through my Amazon affiliate link (also on sidebar) and I'll get ~5% of whatever you spend at no extra cost to you.

If that doesn't appeal to you, you can also shop through this portal and give ~5% of whatever you spend to one of GiveWell's top recommended charities.

In order to get you started, here are some recommendations for products I really liked over the past year. I'm not affiliated with any of these and don't get anything from endorsing them besides the Amazon fee. I talk about some health care products, but they're all out of my field and best thought of a the opinion of an informed consumer, *not* as official medical endorsements.

Traditional 18-Year Balsamic Vinegar

A big part of people's enjoyment of food is placebo. For example, people in blind taste tests prefer Pepsi to Coke, but people in unblinded taste tests prefer Coke to Pepsi because of the Coke "mystique". Likewise, people will rate wine as better-tasting if they are falsely told it is more expensive. I expect this effect is bigger in people (like me) with relatively dull palates and overactive imaginations. We could capitalize on this by finding some food that sounds magical, amazing, and precious, and expecting it to have a taste that matches.

That food is balsamic vinegar. Not the fake stuff you get for $4.99 at the supermarket. The real version, which is usually called something like "traditional 18-year balsamic vinegar" and whose manufacture involves about the same number of complicated restrictions and obscure types of wood as building the Ark of the Covenant.

It’s not just that the grapes have to be from a particular species and grown in a particular Italian soil. It’s that they then have to be aged through a process in which the vinegar is constantly transferred, untransferred, and retransferred among series of successively smaller barrels of oak, chestnut, cherry, juniper, mulberry, and acacia a bunch of times for eighteen years, so that each wood can “add its flavor” to precisely the right degree. How long it must spend in each cask is so complicated that people need high-level mathematics just to figure out how old any specific sample of vinegar is. See for example Guidici & Rinaldi 2007, A Theoretical Model To Predict The Age Of Traditional Balsamic Vinegar. The paper finally concludes that “there is a finite limit for the age of balsamic vinegar” – which I guess is reassuring.

So how does it actually taste? ~~Kind of like a mixture of grapes, oak, chestnut, cherry, juniper, mulberry, and acacia~~ It tastes *really interesting*. Definitely not like normal vinegar. Not really like normal anything. Italians put it on ice cream (really!) or drink it straight from the bottle (really!) I can’t promise it’s not just the placebo effect, and honestly it probably is, but it’s certainly a more *interesting* the placebo effect than just buying an expensive wine or eating at a fancy French restaurant.

Buy traditional 18-year balsamic vinegar

Fake Nice Pants

I have some sensory issues which make me find normal pants annoying; I prefer to wear sweatpants whenever I can.

But this isn't always compatible with appearing as a normal productive member of society, so sometimes I am forced to wear uncomfortable blue jeans or dress pants or whatever.

Luckily now there are are now sweatpants that look like jeans, khakis, dress pants, etc. I won't say they'll fool a careful examination, but hopefully nobody is carefully examining your pants, plus even if people notice they might feel awkward calling you out on it.

Buy sweatpants that look like jeans, sweatpants that look like other types of jeans, sweatpants that look like dress pants, or sweatpants that look like khakis.

Xylitol Nasal Spray

Xylitol nasal spray is for nasal congestion and allergies. It's not too different from saline sprays, but it's a little gentler on the body and also has mild antibacterial properties. Also, I hadn't even realized how well saline sprays worked until recently.

I've informally recommended this to a couple of patients with nasal and sinus issues and have received mostly good reports from them too.

Buy xylitol nasal spray

Hamilton CD

The new Broadway musical Hamilton is *really* good. If you don't believe me, believe mainstream media sources like 538, Vox, Breitbart, Vox again, the New York Times, Vox a third time, the LA Times, and one more Vox. Also all of Tumblr.

It’s a musical about the life of Alexander Hamilton set to rap, but you will like it even if you do not like rap or early American history. You can listen to the whole soundtrack for free on YouTube, follow along with the heavily annotated lyrics on genius.com, and finally you’ll want to have the CD so you can force your friends and family to listen.

If nothing else, this will help you understand the in-jokes on social media.

Buy the Hamilton CD

Dental Floss Scythes

Like everyone else in the world, my dentist told me to floss more but I never listened. Dealing with that weird little box of string was just too much of a trivial inconvenience.

I recently learned about dental floss scythes (probably they have a less interesting official name). These are little plastic things that kind of pre-arrange the dental floss for you. Now I am actually flossing every day. Well, most days. Some days. The point is, I’m flossing. Yes, I’m a bad person contributing to disposable waste culture, but at least I’m healthy.

(Now some people say flossing might not prevent heart attacks as was previously believed, but it’s probably still a good idea)

Buy dental floss scythes

Nootropics Depot and Ceretropic

Not a product so much as a company. Nootropics Depot and Ceretropic are two online nootropics stores. They’re both owned by the same guy but have different branding:

Nootropics Depot has nice gentle all-natural stuff, Ceretropic has new high-tech research chemicals.

There are lots of online nootropics stores, but these are far and away the best. I say this after lurking on r/nootropics for a long time, where users are very vocal about their experiences with different companies and mostly agree with my assessment. The guy who runs these stores also hangs out on r/nootropics, where he shares his encyclopaedic knowledge of everything and helps people who have gotten themselves in trouble. There are worse ways to spend a day than just reading through his entire posting history, but his commentary on my article Iron Curtain Of Psychopharmacology is a good specific example. He is legit and so are his sites.

Experimenting with nootropics is fundamentally a risky endeavor. You are using experimental psychoactive chemicals that haven’t been rigorously tested for safety or efficacy. Nootropics Depot and Ceretropic don’t change any of that. But they do ensure you are not doing anything stupider than the stupid thing you think you are doing. They are very careful about having purity-tested, un-degraded chemicals and making sure you are getting what is advertised on the bottle. They try to have the gold standard version of everything – for example, if they’re selling an herbal extract, it will be the same by composition, subspecies, etc as the version used in whichever studies have most clearly suggested safety and efficacy for the herb. They avoid at least some of the sketchiest diet-pill-stimulant-amphetamine-analogs you can find on some other sites. And they have excellent customer service and if for some reason your nootropics don’t arrive or get confiscated by customs they will work it out for you.

The other reason I like this company is that it does seem to have a long-term plan to build up a power base from which to try to ‘disrupt’ the pharmaceutical industry in the way we often talk about here. That’s obviously crazy ambitious, but they’ve already done some impressive things, like come up with a longer-lasting version of tianeptine. From a chemistry point of view that may not be super hard, but since tianeptine is a prescription drug in most of Europe it means they’ve improved pretty substantially upon a real pharmaceutical company’s project. They seem to be gradually getting labs and researchers, so I’m optimistic.

Oh, and if some of you crypto geniuses want to do a mitzvah, find an easy way for them to charge people non-Bitcoin money without credit card companies backstabbing them every few months because they get cold feet and decide nootropics are too weird.

Visit Nootropics Depot and Ceretropic.

Periscope Glasses

In case the nootropics don’t work well enough for you and you still have human experiences like “sleep” and “tiredness”, you may be interested in periscope glasses. They let you lie down supine in bed and still read without having to hold your book in an awkward position.

Buy periscope glasses.

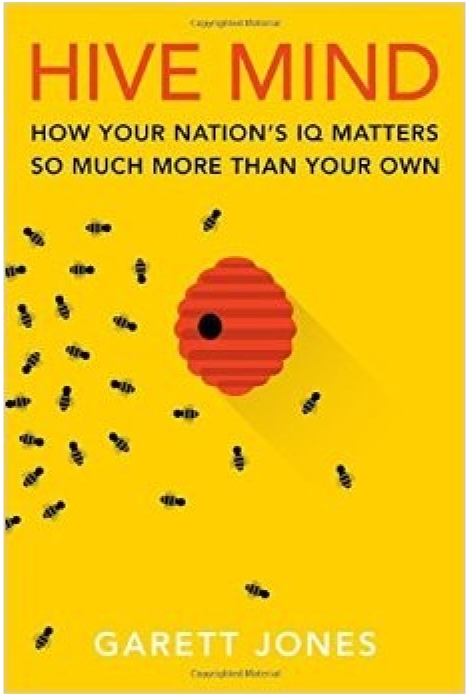
[**EDIT:** If that wasn’t enough for you there’s also a 2014 Product Recommendations post.]

Book Review: Hive Mind

[Conflict of interest notice: Author Garett Jones sometimes reads this blog and is generally great.]

Garett Jones' book *Hive Mind* is classic pop science writing: an intriguing hypothesis, a long parade of interesting studies presented as catchy anecdotes, and not too many follow-up questions.

Its subject (and subtitle) is "why your nation's IQ matters more than your own". The gap between rich and poor countries has proven surprisingly resilient, and conventional wisdom is finally getting its head around the idea that something more is going on than a couple of countries getting a head start and the rest of them needing a little time to catch up. Something more than just a temporary lack of capital must be separating the haves from the have-nots, and Jones thinks IQ must be part of the puzzle.



I like my science writing like I like my coffee – COVERED IN BEEEES!

He starts with what he calls “the paradox of IQ”. IQ doesn’t matter *that much* on a person-by-person basis. Sure, it’s correlated with measures of success like personal income, but only weakly. On the other hand, IQ is a very strong predictor of national success – a country’s average IQ score correlates very well with whether it’s industrialized, rich, First World, and all those nice things. Jones writes:

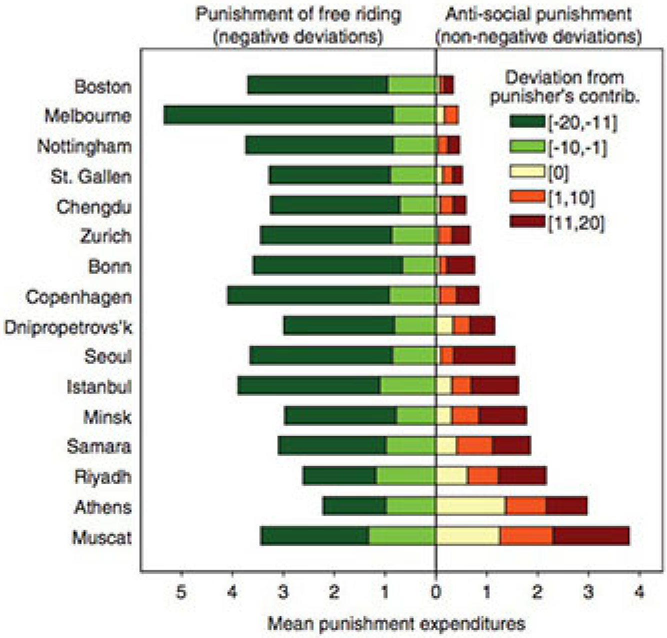
> Looking at how individual student test scores predicted those students’ wages later in life, they found that individuals with higher test scores earned only slightly more than average within a given country, but nations with higher average test scores grew expcetionally fast. Here again is the paradox of IQ: standardized test scores – whether we call them IQ tests or math tests or something else – predict big national differences but only modest individual differences

.
I’ll talk a little more about that claim in Part II of this review, but for now let’s take it seriously and assume causation. Why would IQ matter more for nations than for people?

Jones’ theory is that IQ is a measure of people’s ability to cooperate in prisoner’s dilemma style situations and seek non-zero-sum solutions. Countries where most people have high IQ will come up with mutually beneficial win-win institutions; those where most people have low IQ will be so busy taking advantage of each other and fighting over the pie that they’ll never build the institutions necessary for economic growth.

First he reviews research showing that IQ is closely linked to time preference; ie the higher your tested IQ the more likely you are to prefer a big payoff later to a smaller payoff now. For example, in a German experiment a few years ago, participants were offered 100 euros now or X euros in one year; every fifteen IQ points correponded to a €2.50 change in the value of X necessary for them to accept the latter, even after controlling for education, income, etc. The same thing seems to happen in real life, according to a great study that looked at a natural experiment in the US armed forces. When the military started downsizing after the Cold War, they offered enlisted personnel their choice of various different severance packages – some corresponded to a little money immediately, others to much more money over a longer period. Since the military keeps careful records of the IQ-at-time-of-recruitment of all of its personnel, this was a perfect real-world opportunity to see what happened. The results conformed to theory: IQ predicted tendency to take the longer-term but more lucrative package. There are about twenty studies confirming this result now. And there are also studies showing *national* IQ corresponds with that nation's savings rate, and that individuals who are surrounded by patient frugal people will themselves act more patiently and frugally. If, as the old saying goes, building a good society is about "planting trees in whose shade you will never sit", the people of high IQ nations have a big head start.

Second, he reviews the research from experimental game theory. A series of experiments performed in (of all places) a truck driving school investigated a Window Game. Two players are seated at a desk with a partition between them; there is a small window in the partition. Player A gets $5 and may pass as much of that as she wants through the window to Player B. Player B may then pass as much as she wants back through the window to Player A, after which the game ends. All money that passes through the window is tripled; eg if Player A passes the entire $5 through it becomes $15, and if Player B passes the $15 back it becomes $45 – making passing a lucrative strategy but one requiring lots of trust in the other player. I got briefly nerd-sniped trying to figure out the best (morally correct?) strategy here, but getting back to the point: players with high-IQ were more likely to pass money through the window. They were also more likely to reciprocate – ie repay good for good and bad for bad. In a Public Goods Game (each of N players starts with $10 and can put as much or as little as they like into a pot; afterwards the pot is tripled and redistributed to all players evenly, contributors and noncontributors alike), high-IQ players put more into the pot. They were also more likely to vote for rules penalizing noncontributors. They were also more likely to cooperate and more likely to play closer to traditional tit-for-tat on iterated prisoners' dilemmas. The longer and more complicated the game, the more clearly a pattern emerged: having one high-IQ player was moderately good, but having *all* the players be high-IQ was amazing: they all caught on quickly, cooperated with one another, and built stable systems to enforce that cooperation. In a ten-round series run by Jones himself, games made entirely of high-IQ players had five times as much cooperation as average.



Not technically from the book, but nevertheless fascinating

Third, he reviews the so-called "O-ring theory of teams", named after the spaceship part that malfunctioned during the *Challenger* explosion. The theory is: suppose that a spaceship requires a million different parts to work. This is more than just a million times harder than building a spaceship that requires one part to work. If you have a spaceship engineer who can build a part and be 99% sure she's gotten it right, this is probably good enough for the one-part spaceship: a 99% success rate for a spaceship sounds pretty good. But if the spaceship uses a million parts and they all have to be perfect, your chances of success with a million such engineers is 0.99^1000000, aka zero. You had better find some better spaceship engineers! This gives high-IQ societies a big leg up when they're working on complicated projects; a low-IQ society may have some high-IQ individuals who can do good work on their own, but including even a single low-IQ individual on a spaceship will screw it up big-time. This theory implies that people will end up segregated by ability. Imagine you have four spaceship engineers, two of whom are good (99% accuracy) and two of whom are mediocre (50% accuracy), and you want to build two two-part spaceships. If you pair up one good and one mediocre engineer on each, each of your spaceships will have a 0.99 * 0.50 = 49% chance of success, for a total of 0.98 projected successful spaceships. If you have the two good engineers work together on one ship, and the two mediocre ones work together on the other, you'll have a 98% success rate on the first one and a 25% success rate on the second one, for a total of 1.23 projected successful spaceships. You've gained a quarter-spaceship just by segregating your engineers by ability. The more high-IQ people you have, the easier this is and the more you can devote your economy to complex things like million-part spaceships. The more low-IQ people you have, the harder this gets and the more your economy sticks to high-failure-tolerance but less lucrative products.

Finally, high-IQ people are smart (citation needed). They tend to know what policies are good and what policies are bad and vote for the good ones. Here Jones cites Bryan Caplan's *The Myth of the Rational Voter* a lot, showing that voters aren't very good at figuring out their own self-interest.

But he has a more positive spin: high-IQ voters *do* seem good at this. As a GMU economist, Jones' measure for "are people voting rationally" is of course "how pro-free-market are they?", and he finds that high IQ predicts pro-market attitudes pretty strongly and in fact better than years-of-education. In controlled experiments higher-IQ people were more likely to be able to admit that a test article contradicted their political bias, and in some countries (although not the US) high-IQ people are more likely to vote.

Then he ties all of this together into a kind of stationary-bandit framework, where government starts with selfish warlords who want to exploit the populace.

> "They say all government started w/ stationery bandits." "Really?" "Yeah. They had to steal enough nice paper to write a constitution on."
>
> — Scott Alexander (@slatestarcodex) December 8, 2015

If you're a high-IQ selfish warlord, and your oppressive ministers are likewise high-IQ, you have enough patience to realize that if you leave the capitalists alone to do their thing instead of confiscating their wealth immediately, in a couple of years they'll have even *more* wealth you can confiscate. And if some kind of conflict comes up and threatens to lead to civil war, you are good at negotiating win-win solutions where everyone cooperates to increase the size of the pot. Jones lists a bunch of political situations that map to iterated prisoner's dilemmas – for example, do both parties respect election results, or does the loser try to start a fight over it every time they're forced out of government? Do bureaucracies try to run the country efficiently, or do they jockey for power against each other? Do military branches work together during operations, or does each one try to seize glory for its own leaders? If you have a high-IQ country, these problems have a way of just solving themselves – and sure enough, IQ scores correlate nicely with the Corruption Perceptions Index. And businesspeople know this, so they are happy to start complicated long-term projects in the countries with a history of tolerating such projects and not killing the golden-egg-laying geese.

Jones doesn't go too deep into policy prescriptions, but he does mention two consequences of his theory. First, he's a big fan of the Flynn Effect (secular trend of rising IQs) and thinks that countries ought to encourage this so that their national IQ gets higher and they can have more effective institutions – unfortunately, he doesn't know what's causing the Flynn Effect any more than anyone else does, so this sort of reads as "keep doing the thing we don't know how we're doing". He does think that eliminating lead will help (did you know sub-Saharan Africa was the last region to ban leaded gasoline, all the way in 2006?) and he has the usual hopes for nutritional, educational and health interventions.

But of course the part everyone's talking about is immigration. This is *not* a major focus of the book. Jones actually spends more time talking about all the benefits of immigration than anything else:

> About a decade ago, dozens of American economists signed an open letter in support of more immigration. The letter touched on many points: that less-skilled immigrants appear to push down the wages of US born citizens little if at all, that immigration helps rich country economies in ways that don't show up in official statistics, and that the biggest beneficiaries of less-skilled immigration are the immigrants themseles, whose lives are often transformed from a nightmares of dollar-a-day poverty to a realm of modest comfort, health and safety. The diplomatically crafted letter, circulated by the Independent Institute, was signed by economists on the left and the right. I've always been glad I signed this letter: it sums up the great promise of immigration…for people who care about ending the deepest poverty, migration should be at the top of the list of potential cures.

But he does devote about one-and-three-quarters pages to his concerns:

> The economics of less-skilled immigration to richer, more productive countries are reasonably clear: life-changing good news for the immigrant with only fairly small effects one way or the other on so-called "native" less-skilled workers. That's true when we look at the short run or when we look across towns and cities within the same country. And crucially, these studies hold politics aside and assume that less-skilled immigrants don't have an effect on a high-skill nation's government institutions. But if there's something we've seen in previous chapters, or something we've seen in Bryan Caplan's work on the link between voter education and voter beliefs, if there's something we've seen in the cross-country studies that find that higher national average test scores tend to predict lower average levels of corruption and in the philosophical debates over epistocracy, it's that good politics appears to depend on reasonably well-informed citizens. With this we come to a central tension of immigration among the currently less-skilled: the possible – I emphasize possible – effect on long-run institutions. Will less-skilled immigrants tend to vote for policies that will weaken the wealth-creating opportunities they've enjoyed? Or will less-skilled immigrants and their descendants instead build up high levels of human capital, perhaps raising the average information level of voters?

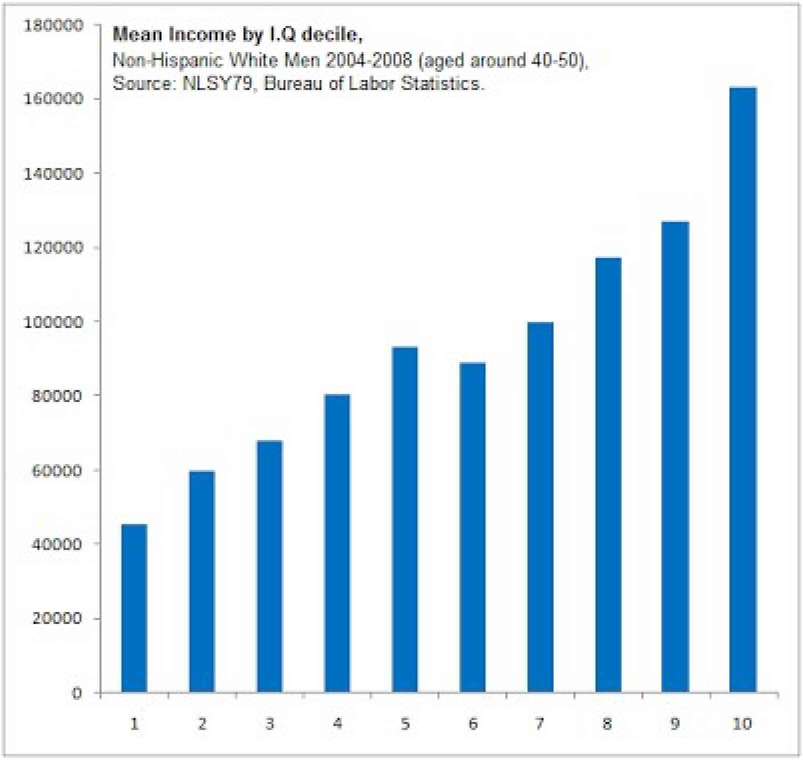
The whole paragraph has the feeling of somebody being dragged over a bed of hot coals, from the insistence on referring to unskilled immigrants as "currently less-skilled" and natives as "so-called native less-skilled workers" to the odd proposal at the end that maybe for some reason less-skilled workers will actually *raise* the average information level of voters, because *who really knows?* This book is emphatically not *The Bell Curve*. It's a book about science which is deeply annoyed that it might have controversial political implications and tries to avoid them as carefully as possible, generally successfully.

II.

There were some parts of this book that I did not find convincing, or that at least left me with further questions.

First, *Hive Mind*'s "central paradox" is why IQ has very little predictive power among individuals, but very high predictive power among nations. Jones' answer is [long complicated theory of social cooperation]. Why not just "signal-to-noise ratio gets higher as sample size increases"?

Jones' paradox was very similar to the question I asked in Beware Summary Statistics, except I was wondering not about nations, but about abstracted IQ deciles:



On a personal level, IQ has modest predictive power. But if you average out thousands of IQ 90 people, thousands of IQ

100 people, and thousands of IQ 110 people, the IQ-income relationship will become very clear. At this level of abstraction, it is no longer fair to describe it as “modest”.

That first block corresponds to people of about IQ 80, the last block to people of about IQ 120. As you move from 80 to 120, income practically quadruples. And this is within the United States, where we’ve got all sorts of minimum wage laws and so on likely to dampen the effects.

Or to give a more natural example – Jews have 10-15 points higher IQ than WASPs in America, and make about twice as much money. This happens even though most Jews do not solely interact with other Jews or make their own institutions – there are few opportunities for them to form a hive mind. Their individual IQ differences, once aggregated, seem to produce the strong effect.

There is much-larger between-country variance in income than between-individuals-in-country variance in income, but it doesn’t seem obvious to me that the percent of between-country variance explained by national IQ is larger than the percent of between-individuals-in-country explained by personal IQ once factors like personal job choice (I could have been an investment banker, but I would rather be an artist) that countries don’t have to deal with is abstracted out. If the amount-variance-explained between nations and between individuals were equal after adjusting for that factor, there would not be any need to posit hive mind-type effects.

[EDIT: Above heavily edited for clarity and correctness after originally being much weaker argument in same direction. See here. Some complicated discussion of this going on here, see especially Pseudoerasmus’ comments]

Although it may be that there’s a national effect stronger than the aggregated-individual effect, I feel like this is something Jones should have had to prove, rather than relying on a “look, it’s obvious!” based on unaggregated-individual numbers.

Second, fine. Let’s assume he proves to our satisfaction that the national IQ-income correlation is sufficiently stronger than individual ones to require explanation. Now we get to my biggest gripe with this entire book. *How do we know the direction is IQ -> development rather than development -> IQ?!*

Jones lays out exactly the set of assumptions that make reverse causation most plausible. He dedicates an entire chapter to the Flynn Effect, how he thinks it’s real, how he thinks it’s a big deal, without mentioning whether the gains might not be on g. Time and time again, Jones hammers how countries’ IQs increase as they develop further. Then he shows us a graph of IQ-development correlation and just assumes the causation is bidirectional. Well, why not just development -> IQ?

This isn’t just about me. I suspect Jones is right – though I’m not entirely sure of it – and sufficiently biased in favor of that position to be happy to follow it and see where it leads. I’m asking for anybody who reads this book without already being interested in IQ. *Hive Mind* is clearly pitched at a smart layperson audience, and any smart layperson who reads this book ought to have exactly that question, asked with exactly that many capital letters and explanation points. Any reader who doesn’t immediately stand on a chair and shout “Where is the evidence against reverse causation?” is not a reader that Garett Jones should want. But any reader who does that will not find an answer.

> I'm just sayin', everyone that confuses correlation with causation eventually ends up dead.
>
> — Siberian Fox (@SilverVVulpes) September 14, 2015

All I can say in his defense is that a good defense against this accusation would probably have to get very deep into the causes of IQ, exactly the subject Jones is carefully trying to avoid. I understand his reluctance to approach this subject and respect his strategic decision. All I can say is that it leaves a hole in his argument big enough to sail an oil tanker through.

[EDIT: Jones responds here]

Third, and this isn't such a problem as the others but it left me curious – how do we go from the short, few-player games that make up most of the book's experiments, to the multi-generation million-player games that make up real countries?

I have two concerns here. First, Jones says that:

> The one study of which I'm aware that finds that higher-IQ individuals are more cruel and less cooperative is a study of a one-shot prisoners' dilemma, something much like the true criminal's prisoner's dilemma. ..this is the only setting I know of in which high scorers are more brutal than low scorers…in a one-shot environment, if it's either steal or be robbed, and if the players will never see each other again, then I'd expect higher-IQ individuals to figure out what setting they're in and act shrewdly, not cruelly.

He returns to this theme a few times. High-IQ people don't cooperate because they're nicer (which, indeed, personality tests for niceness do not show). They cooperate because they're smarter and so they know cooperation really is a better and more win-win way to do things.

This is 100% true in an iterated prisoners' dilemma, but not necessarily true in a country. Suppose you're a president with a four year term. You can either pillage the country as best you can and take whatever bribes you can get, or invest in genuinely building a better country for your descendents. Assuming you are merely the sort of shrewd cooperator who cooperates on iterated prisoners' dilemmas but defects on one-shots, you'll pillage the country – it probably has term limits and you only need to pillage once to get very rich.

Likewise, suppose you're a mid-level bureaucrat in Washington, of the type that there are tens of thousands of. If you behave dishonorably, you can amass a small empire and make some money. If you behave honorably, then maybe America does very well as a country down the line, but that effect is aggregated over thousands of bureaucrats, so it's not like you're *really* growing the pot that much. Once again, if you are merely shrewd and not genuinely altruistic, you'll defect.

Jones tries to take the easy way out on the deriving-ethical-behavior thing here, saying that ethical behavior *really is* the most self-serving option in the long-term, and all you need is people smart enough to realize it. To that I can only say: no it isn't. In a game of two or three people where everyone sees everyone else's results, your contribution may grow the pot enough to be worth the short-term losses from not defecting. In a game of thousands of bureaucrats or millions of citizens, not so much. There are ideas like TDT and superrationality that try to bridge this gap, but I think Jones tries to cross it without those ideas and is left floundering.

One more thing on this topic: maybe it was in the original studies and I just didn’t look deep enough, but I wonder how much of this is just understanding the game. The Iterated Prisoner’s Dilemma is kind of complicated, and if you’re stupid you may not be able to grasp the logic behind why cooperation is sometimes the better option. If you explained everything very carefully to all participants, had them play a couple of games both ways so they got a feel for it, and had a Professor of Economics give a lecture on why cooperation was probably the best option, would high-IQ people still succeed more because of some innate cooperative tendency? Or would everyone else have figured out their secret and robbed them of their advantage?

People usually have a pretty good grasp of things that are going on in society. Jones compares marriage to a prisoner’s dilemma (where the optimal C-D outcome is “you cheat but your spouse stays faithful”). But people understand the terms of marriage, cultural evolution has had a long time to come up with beliefs and mores about marriage that even people too stupid to come up with them on their own follow, and some kind of complicated new game may not be the best analogue to the marriage problem.

III.

Jones ends the book with the following observation:

> The best guess is that the cognitive skill of elites really does matter more than the nation’s average score. When it comes to institutional quality, Potrafke and I found that the cognitive skills of the top 5 percent did the better job of predicting property-rights friendly institutions, although the nation’s average score also did a reasonably good job as a predictor…for the time being it’s reasonable to start with the belief that a nation’s top

> performers matter more for the economy than a nation’s average performers.

Well, *that’s* interesting. All of this stuff about immigration and on how maybe we shouldn’t have open borders, and it turns out that as long as the top five percent are smart, everything is okay.

I would really like to see more on this. If America has higher IQ than Britain, but the members of Parliament have higher IQ than the members of Congress, which country will do better? What about a colonial nation where the administrators are from a nation that has a completely differnet IQ than the population? What about countries that have multiple mostly-segregated populations with different IQs? How much does the IQ of the government matter versus the IQ of the population itself?

(and now I wonder if Jones has read La Griffe on smart fractions [1, 2])

Come to think of it, doesn’t every nation have some pretty smart people at the highest echelons? Sub-Saharan Africa may be in the IQ doldrums, but we all know African economists, statesmen, etc whose work is top-notch. Doesn’t Jones’ call to raise national IQs with the Flynn Effect seem less pressing now? Haven’t the elites of Third World countries already probably been Flynn-ified, since they usually get good food, good medical care, and good education? Should we worry the Flynn Effect won’t help those countries further? Or should we hope that if we merely raise the IQ of a few people, that will be enough and we won’t have to have a mass nationwide campaign? (calling all CRISPR enthusiasts…)

Overall, I thought this book showcased some really neat results, had some good economics in it, and was very readable, but I didn’t come out of it feeling like its thesis was very proven.

List Of Passages I Highlighted In My Copy Of “Hive Mind”

> Even though simple logic said “there will be blood”, there were nevertheless long stretches [of WWI] when all was quiet on the Western front. Why? Axelrod said this was due to the power of tit for tat, which we discussed in Chapter 5. German and French soldiers tacitly created their own unwritten peace treaties: if you don’t shoot at our side, we won’t shoot at yours. And if the higher-ups do force us to fire our artillery, we’ll intentionally aim short of your position, or aim long of it. These rules were never written down, and they were rarely even spoken as far as we can tell – but they left their mark. Take this example. One day the German artillery fired toward the British side without doing any damage. A German infantryman climbed up on a parapet just to deliver an apology to the British: “We are very sorry about that; we hope no one was hurt. It’s not our fault, it’s that damned Prussian artillery” The German infantrymen didn’t want to destroy the fragile truce they had created with their alleged enemies. Military higher-ups hated this tacit cooperation across enemy lines. Fortunately for the officers (but not for the enlisted men) there was a simple solution: move troops around. By swapping one division south two kilometers and another north two kilometers, military officers could turn a repeated prisoner’s dilemma into a literal one-shot game.

This seems almost unbelievable, and absent knowledge of the primary source I can’t be 100% sure, but I guess it fits together with what you always hear about the soldiers celebrating Christmas together and so on.

> Some skeptics dismiss IQ tests as just measuring whether you’re good at staring at a piece of paper, coming up with an answer, and writing it down. But the comprehensive IQ test used most often today, the Wechsler mentioned earlier – involves little paper-staring and almost no pencils. The person giving the test (a psychologist or other testing expert) asks you why the seasons change or asks you to recite a list of numbers that she reads out to you. You answer verbally. Later you are handed some wooden puzzle blocks and you try to assemble them into something meaningful.

I hadn’t realized this. There’s a lot of talk about IQ being related to “test-taking skills”, and it’s easy to imagine that people who are used to take paper-and-pencil multiple choice exams might be more comfortable with them, but that’s not what a lot of real IQ tests are and we should stop imagining them that way. This also makes theories about education and the Flynn effect more impressive.

> The images are simple shapes like a regular C or a backwards one, and your job is to note, for instance, whether the C is open to the left or the right. So at what point would you no longer be able to do better than random at correctly answering “left” or “right”? When the image flashes for just an eighth of a second? A 32nd? A 128th? That’s the key variable the researcher keeps track of in this study…can it possibly be the case that people who only need to see the image for a tiny fraction of a second tend to have higher average IQ scores than people who need to see it for an eighth or sixteenth of a second? Summarizing “dozens of studies” run on “four continents” dating back to the 1970s, psychologist Ian Deary says, “the overall answer is yes, there is a moderate association between how good

> people are at the inspection time test and how well they score on intelligence tests.”

…speaking of IQ not just being about test-taking. Jones also goes through some studies on reaction time and reverse digit span, not to mention literal brain size as measured by MRI.

> IQ tests do about the same as the best kind of job interviews – structured job interviews in which the interviewer carefully designs the questions beforehand and sticks to the same ones with each candidate – and better than most of the methods people use to choose employees…IQ tests are even better at predicting outcomes when the job requires higher skills. Back in the 1960s, the Bell Telephone System gave its entry level management trainees an IQ-type test along with a number of personality tests…when, after two decades, the company looked back to see which tests did the best job of predicting which trainees eventually rose the highest in the company hierarchy, the IQ-type test did the best job, beating out the personality tests. Looking across many studies of IQ in the elite workforce, one review says “general cognitive ability is the best single predictor of executive/professional-level performance, just as it is of performance in the middle to high end range of the general workforce.”

What I get out of this is “apparently people should structure their job interviews more”.

> One might contend that it’s harder to measure social or emotional intelligence than it is to measure more conventional intelligence. But psychologists have tried: they’ve checked to see whether people with more social or emotional intelligence tend to have higher IQs, and

> so far it looks like they do. The relationship often isn't as strong as the relationship between, say, a person's vocabulary test scores and their Raven matrices scores, but the results are clear: IQ scores predict practical social skills.

On the other hand, all of these "measures of social intelligence" are things like "can you read this person's face to determine what emotion they are feeling?"; I wonder how well that correlates with tasks like "work your way up to becoming coolest person in your high school class" or "win the heart of your love interest". There is a George Washington Social Intelligence Test, but it doesn't seem to *actually* involve becoming the universally beloved father of your country.

> But what did the apparently more cautious, more careful Wicherts report? He said that the average IQ in sub-Saharan Africa was about 82 – corresponding to the 12th percentile in the United Kingdom…that's an improvement from [Lynn's estimate of] 70, and it's an improvement that arose partly because Wicherts chose to throw out samples of students who came from families with nutrition problems and low socioeconomic status. The Wicherts average of 82 only includes samples of apparently healthy students from families that have typical socioeconomic status. And in a region of the world with as much poverty and disease as sub-Saharan Africa, that decision is quite likely to leave out substantial portions of the population. To further test the data and get the best average, Wicherts wrote a separate paper that looked at onl the best test samples…Wicherts' best samples of students have an average IQ score of 76. That's at the 5th percentile within the United Kingdom.

I keep hearing wildly different numbers for sub-Saharan African IQ, but 76 seems pretty plausible. Note that if the higher 82 number which I keep seeing around is true, it would throw a wrench in a lot of things. Remember that African-Americans are at about 85, so if going from Africa to America only gains you at most 3 IQ points, everything we know about the Flynn Effect falls apart. Even 76 -> 85 is a kind of low Flynn Effect estimate, but I'll allow it if we speculate that these Africans probably have been at least a little Flynnified, and maybe African Americans are still impoverished enough not to have been fully Flynnified.

> Even in rural Pakistan, higher Raven's IQ predicts higher wages. One might think that thousands of miles away from the Western universities where the tsts were designed, abstract IQ tests would have no power to predict which workers earned more and which earned less – but an IQ test made up of boxes and lines and circles had a modest ability to predict a person's wages across rural Pakistan, just like in the United States.

Also relevant to IQ tests and culture.

> Country-level results go back decades. As early as the 1960s, studies in Taiwan and Hong Kong found average IQ scores slightly above the European average. This happened at a time when these countries' economies were growing fast, but were still poor by US and UK standards. Any simple story that "wealth causes IQ" has to account for the puzzlingly high average scores found in Taiwan and Hong Kong decades ago…a healthy environment helps to boost IQ, but it can't be the whole story. The high average IQs of East Asia and Singapore have yet to be fully explained.

See, this is another thing that confuses me about the Flynn Effect. Taiwan etc are *still* slightly above the European average. So either they had fully caught up to Europe in Flynnification by 1960 (how?!) or Asians have some sort of magic anti-Flynn-effect armor. Didn't Ron Unz speculate something like that once?

> The key evidence comes from the horrific experience of famine in the Netherlands during World War II. Some Dutch towns were cut off from regular food supplies toward the end of the war – daily calories were down to about one third of recommended levels at one point – and the children in gestation during that period had low birth weight. But when these children grew up, their IQs were nearly identical to those of children born a few years earlier or a few years later: massive in utero food deficiencies had no long-term effect on child IQ… famines lasting a few months might not matter for a child's brain development. But what about chronic malnutrition among the young?…Again the experimental method offers an answer. In one study of Guatemalan villagers, some malnourished children where given a protein supplement and some were given a nonprotein supplement, and then their cognitive skills were measured in few different ways. Of the two interventions, the protein supplement gave a bigger boost to student scores.

I was expecting something awful but on first glance the paper actually looks really good. I've uploaded it here if you want to see. They don't measure the results in IQ so I'm not sure exactly what the magnitude is. It only worked on people who were already poor and probably not well-nourished. Possibly another argument for vegetarians taking protein supplements?

> So far it appears that schooling clearly boosts crystallized intelligence, but the typical efect of school on fluid intelligence may be small or even nonexistent… [but] here are two examples of apparently successful increases in fluid intelligence. First, in a study in Sudan, a few months of training on the abacus appeared to boost Raven's matrix scores dramatically. Second, a study of Israeli schoolchildren, comparing students of similar ages who were born just before or just after the school's age cutoff: because students born before the cutoff get an extra year of formal schooling, the age cutoff let the researchers identify the effect of an extra year of schooling on student IQ scores. The researchers found a bigger effect on vocabulary scores than on matrix scores, but nevertheless matrix scores indeed rose for students who had received an extra year of schooling.

Right now I am pretty confused on what affects fluid vs. crystalline intelligence. Also, really want to know if the school year results hold up 10 or 20 years later.

> In the case of classroom learning, the evidence is mixed on whether your child's standardized test scores will likely rise if she's in a class room of high achievers or tend to fall if she's placed in a classroom of weaker students. This is one question that's been tested and retested in numerous ways in classrooms around the world, and perhaps the best way to summarize the vast literature is to say that some signs point to positive peer effects and some point to no peer effects. Amid the ambiguous findings, the clearest peer effect is that disruptive students hurt learning.

I could have told you that in third grade.

> Sub-Saharan Africa was the last major region of the world to eliminate lead from its gasoline, a goal that apparently was reached in 2006.

That is exciting and will hopefully cause some progress for the region.

> Did your neighbor win the lottery and buy a nicer car? That means that you’re more likely to buy a nicer car too. At least that’s what economist Peter Kuhn and his coauthors found when studying the results of lotteries in the Netherlands: people who won big lottery prizes tended to buy nicer cars (no surprise), but people who lived near the winners tended to buy nicer cars too. Buying is a social activity: most of us try at least a little to fit in with our neighbors.

This section was titled “Staying Frugal Like The Joneses”. Stop tooting your own horn, Garrett.

> One study had twins play a repeated prisoner’s dilemma game against each other for a hundred rounds, and they found that higher-IQ pairs of twins did cooperate more often than lower-IQ pairs of twins.

I feel like making twins play the prisoners’ dilemma against each other is the sort of thing where you risk accidentally making the machinery of the universe divide by zero and explode.

> Here’s the most exciting result from my experiments with Omar and Jaap: on average, over the course of the entire experiment, higher-IQ pairs were five times more cooperative than higher-IQ individuals. The link between IQ and cooperation was an emergent phenomenon: it arose not from smart individual players but from smart pairs of players.

There was a similar result in a dataset taken by looking at all the prisoners' dilemmas done in different colleges' psych programs and such, then sorting them by the SAT score of the college. This is a scarier lesson about WEIRD samples than anything I've read explicitly on that subject.

> So far, two studies in the United Kingdom find that higher [IQ] individuals are more likely to vote, regardless of other things known about the person such as his social class, his education, and some personality traits… however, a study in the United States drawing on three different surveys finds no substantial evidence that IQ predicts voting behavior.

Thanks a lot, Get Out The Vote people.

> A team of Notre Dame political scientists ran a series of deliberative polling studies to investigate these questions. They started by surveying participants' politics attitudes before they were put into a room with other participants. Once they were in the room they were asked to discuss a prescribed set of topics such as the then-ongoing war in Iraq. After the discussion, they were surveyed again. Participants were randomly put into different rooms, some with more Iraq war supporters, some with more opponents, and so on. That made it possible to check whether their post-discussion opinion moved toward the pre-discussin opinion of the average person in their room: it was possible to see if they were conforming. The researchers ran 330 groups of about 10 people each, so if conformity was even a modestly strong force it should have shown up. Here's what they conclude: "After several years of experimentation, we have found little evidence that group composition influences postdiscussion attitude. This persistent finding, which holds across a range of

> experimental variations and for subgroups defined by political knowledge, suggests that the expectations of the Asch literature do not apply to the Deliberative Poll setting”

I’m not sure how surprised to be about this. I guess it makes sense that political issues like the Iraq War are so polarizing that you can’t change people’s minds about them with a short discussion. It’s still surprising to find something without an Asch conformity effect at all.

> Psychologist Christopher Chabris and his coauthors looked at team efforts another way: they checked to see if there was a “da Vinci effect” for teams, a tendency for teams that did well on one kind of task such as a team game of checkers to do well on another task such as taking an IQ test as a team. Indeed, they found such an efect, which they called a c facor, presumably as homage to the general or g factor of human intelligence. And what predicted which teams did better overall? The strongest two factors were whether team members routinely took turns speaking, and how well team members did on a test of reading the emotions of others…but nevertheless, the IQ of the average group member and the IQ of the highest-scoring group member were weak but still notable predictors of group intelligence, and differences in individual IQ explained over one third of differences in actual team performance.

I guess we should care a lot more about people taking turns speaking. Also, I bet that last statistic is basically entirely dependent on the IQ range in the study.

And I’ll take this space at the bottom here to point out that author Garrett Jones very kindly replied to some of my criticisms in the review here. I reply to his reply here, followed by a long conversation with Pseudoerasmus that continues here. I am not *quite* to the point of sticking this on the Mistakes page, especially since Arnold Kling had the same concern even before me, but I admit it some of this is more complicated than I originally expected and I might end up going either way on it.

Should AI Be Open?

I.

H.G. Wells’ 1914 sci-fi book *The World Set Free* did a pretty good job predicting nuclear weapons:

> They did not see it until the atomic bombs burst in their fumbling hands…before the last war began it was a matter of common knowledge that a man could carry about in a handbag an amount of latent energy sufficient to wreck half a city

Wells believed the coming atomic bombs would be so deadly that we would inevitably create a utopian one-world government to prevent them from ever being used. Sorry, Wells. It was a nice thought.

But imagine that in the 1910s and 1920s, some elites had started thinking really seriously along Wellsian lines. They would worry about what might happen when the first nation – let’s say America – got the Bomb. It would be unstoppable in battle and might rule the world with an iron fist. Such a situation would be the end of human freedom and progress.

So in 1920, these elites pooled their resources and made their own Manhattan Project. Their efforts bore fruit, and they learned a lot about nuclear fission; in particular, they learned that uranium was a necessary raw material. The world’s uranium sources were few enough that a single nation or coalition could get a monopoly upon them; the specter of atomic despotism seemed more worrying than ever.

They got their physicists working overtime and discovered a new type of nuke that required no uranium at all. In fact, once you understood the principles you could build one out of parts from a Model T engine. The only downside was that if you didn't build it *exactly right*, its usual failure mode was to detonate on the workbench in an uncontrolled hyper-reaction that would blow the entire hemisphere to smithereens.

And so the intellectual and financial elites declared victory – no one country could monopolize atomic weapons *now* – and sent step-by-step guides to building a Model T nuke to every household in the world. Within a week, both hemispheres were blown to very predictable smithereens.

II.

Some of the top names in Silicon Valley have just announced a new organization, OpenAI, dedicated to "advancing digital intelligence in the way that is most likely to benefit humanity as a whole…as broadly and evenly distributed as possible." Co-chairs Elon Musk and Sam Altman talk to Steven Levy:

> **Levy:** How did this come about? […]
>
> **Musk:** Philosophically there's an important element here: we want AI to be widespread. There's two schools of thought?—?do you want many AIs, or a small number of AIs? We think probably many is good. And to the degree that you can tie it to an extension of individual human will, that is also good. […]
>
> **Altman:** We think the best way AI can develop is if it's about individual empowerment and making humans better, and made freely available to everyone, not a single entity that is a million times more powerful than

> any human. Because we are not a for-profit company, like a Google, we can focus not on trying to enrich our shareholders, but what we believe is the actual best thing for the future of humanity.
>
> **Levy**: Couldn’t your stuff in OpenAI surpass human intelligence?
>
> **Altman:** I expect that it will, but it will just be open source and useable by everyone instead of useable by, say, just Google. Anything the group develops will be available to everyone. If you take it and repurpose it you don’t have to share that. But any of the work that we do will be available to everyone.
>
> **Levy:** If I’m Dr. Evil and I use it, won’t you be empowering me?
>
> **Musk:** I think that’s an excellent question and it’s something that we debated quite a bit.
>
> **Altman:** There are a few different thoughts about this. Just like humans protect against Dr. Evil by the fact that most humans are good, and the collective force of humanity can contain the bad elements, we think its far more likely that many, many AIs, will work to stop the occasional bad actors than the idea that there is a single AI a billion times more powerful than anything else. If that one thing goes off the rails or if Dr. Evil gets that one thing and there is nothing to counteract it, then we’re really in a bad place.

Both sides here keep talking about who is going to “use” the superhuman intelligence a billion times more powerful than humanity, as if it were a microwave or something. Far be it from me to claim to know more than Musk or Altman about anything, but I propose that the correct answer to “what would you do if Dr. Evil used superintelligent AI?" is "cry tears of joy and declare victory", because *anybody at all* having a usable level of control over the first superintelligence is so much more than we have any right to expect that I'm prepared to accept the presence of a medical degree and ominous surname.

A more Bostromian view would forget about Dr. Evil, and model AI progress as a race between Dr. Good and Dr. Amoral. Dr. Good is anyone who understands that improperly-designed AI could get out of control and destroy the human race – and who is willing to test and fine-tune his AI however long it takes to be truly confident in its safety. Dr. Amoral is anybody who doesn't worry about that and who just wants to go forward as quickly as possible in order to be the first one with a finished project. If Dr. Good finishes an AI first, we get a good AI which protects human values. If Dr. Amoral finishes an AI first, we get an AI with no concern for humans that will probably cut short our future.

Dr. Amoral has a clear advantage in this race: building an AI without worrying about its behavior beforehand is faster and easier than building an AI and spending years testing it and making sure its behavior is stable and beneficial. He will win any fair fight. The hope has always been that the fight won't be fair, because all the smartest AI researchers will realize the stakes and join Dr. Good's team.

Open-source AI crushes that hope. Suppose Dr. Good and his team discover all the basic principles of AI but wisely hold off on actually instantiating a superintelligence until they can do the necessary testing and safety work. But suppose they *also* release what they've got on the Internet. Dr. Amoral downloads the plans, sticks them in his supercomputer, flips the switch, and then – as Dr. Good himself put it back in 1963 – "the human race has become redundant."

The decision to make AI findings open source is a tradeoff between risks and benefits. The risk is letting the most careless person in the world determine the speed of AI research – because everyone will always have the option to exploit the full power of existing AI designs, and the most careless person in the world will always be the first one to take it. The benefit is that in a world where intelligence progresses very slowly and AIs are easily controlled, nobody can use their sole possession of the only existing AI to garner too much power.

But what if we don't live in a world where progress is slow and control is easy?

III.

If AI saunters lazily from infrahuman to human to superhuman, then we'll probably end up with a lot of more-or-less equally advanced AIs that we can tweak and fine-tune until they cooperate well with us. In this situation, we have to worry about who controls those AIs, and it is here that OpenAI's model makes the most sense.

But Bostrom et al worry that AI won't work like this at all. Instead there could be a "hard takeoff", a subjective discontinuity in the function mapping AI research progress to intelligence as measured in ability-to-get-things-done. If on January 1 you have a toy AI as smart as a cow, and on February 1 it's proved the Riemann hypothesis and started building a ring around the sun, that was a hard takeoff.

(I won't have enough space here to really do these arguments justice, so I once again suggest reading Bostrom's *Superintelligence* if you haven't already. For more on what AI researchers themselves think of these ideas, see AI Researchers On AI Risk.)

Why should we expect a hard takeoff? First, it's happened before. It took evolution twenty million years to go from cows with sharp horns to hominids with sharp spears; it took only a few tens of thousands of years to go from hominids with sharp spears to moderns with nuclear weapons. Almost all of the practically interesting differences in intelligence occur within a tiny window that you could blink and miss.

If you were to invent a sort of objective zoological IQ based on amount of evolutionary work required to reach a certain level, complexity of brain structures, etc, you might put nematodes at 1, cows at 90, chimps at 99, homo erectus at 99.9, and modern humans at 100. The difference between 99.9 and 100 is the difference between "frequently eaten by lions" and "has to pass anti-poaching laws to prevent all lions from being wiped out".

Worse, the reasons we humans aren't more intelligent are *really stupid*. Even people who find the idea abhorrent agree that selectively breeding humans for intelligence would work in some limited sense. Find all the smartest people, make them marry each other for a couple of generations, and you'd get some really smart great-grandchildren. But think about how weird this is! Breeding smart people isn't doing *work*, per se. It's not inventing complex new brain lobes. If you want to get all anthropomorphic about it, you're just "telling" evolution that intelligence is something it should be selecting for. Heck, that's all that the African savannah was doing too – the difference between chimps and humans isn't some brilliant new molecular mechanism, it's just sticking chimps in an environment where intelligence was selected for so that evolution was incentivized to pull out a few stupid hacks. The hacks seem to be things like “bigger brain size” (did you know that both among species and among individual humans, brain size correlates pretty robustly with intelligence, and that one reason we’re not smarter may be that it’s too hard to squeeze a bigger brain through the birth canal?) If you believe in Greg Cochran’s Ashkenazi IQ hypothesis, just having a culture that valued intelligence on the marriage market was enough to boost IQ 15 points in a couple of centuries, and this is exactly the sort of thing you should expect in a world like ours where intelligence increases are stupidly easy to come by.

I think there’s a certain level of hard engineering/design work that needs to be done for intelligence, a level *way* below humans, and after that the limits on intelligence are less about novel discoveries and more about tradeoffs like “how much brain can you cram into a head big enough to fit out a birth canal?” or “wouldn’t having faster-growing neurons increase your cancer risk?” Computers are not known for having to fit through birth canals or getting cancer, so it may be that AI researchers only have to develop a few basic principles – let’s say enough to make cow-level intelligence – and after that the road to human intelligence runs through adding the line NumberOfNeuronsSimulated = 100000000000 to the code, and the road to superintelligence runs through adding another zero after that.

(Remember, it took all of human history from Mesopotamia to 19th-century Britain to invent a vehicle that could go as fast as a human. But after that it only took another four years to build one that could go twice as fast as a human.)

If there’s a hard takeoff, OpenAI’s strategy stops being useful. There’s no point in ensuring that everyone has their own AIs, because there’s not much time between the first useful AI and the point at which things get too confusing to model and nobody “has” the AIs at all.

IV.

OpenAI’s strategy also skips over a second aspect of AI risk: the control problem.

All of this talk of “will big corporations use AI?” or “will Dr. Evil use AI?” or “Will AI be used for the good of all?” presuppose that you can *use* an AI. You can certainly use an AI like the ones in chess-playing computers, but nobody’s very scared of the AIs in chess-playing computers either. What about AIs powerful enough to be scary?

Remember the classic programmers’ complaint: computers always do what you *tell* them to do instead of what you *meant* for them to do. Computer programs rarely do what you want the first time you test them. Google Maps has a relatively simple task (plot routes between Point A and Point B), has been perfected over the course of years by the finest engineers at Google, has been ‘playtested’ by tens of millions of people day after day, and *still* occasionally does awful things like suggest you drive over the edge of a deadly cliff, or tell you to walk across an ocean and back for no reason on your way to the corner store.

Humans have a robust neural architecture, to the point where you can logically prove that what they’re doing is suboptimal and they’ll shrug and say they they’re going to do it anyway. Computers aren’t like this unless we make them so, itself a hard task. They are naturally fragile and oriented toward specific goals. An AI that ended up with a drive as perverse as Google Maps’ occasional tendency to hurl you off cliffs would not be necessarily self-correcting. A smart AI might be able to figure out that humans didn’t mean for it to have the drive it did. But that wouldn't cause it to change its drive, any more than you can convert a gay person to heterosexuality by patiently explaining to them that evolution probably didn't *mean* for them to be gay. Your drives are your drives, whether they are intentional or not.

When Google Maps tells people to drive off cliffs, Google quietly patches the program. AIs that are more powerful than us may not need to accept our patches, and may actively take action to prevent us from patching them. If an alien species showed up in their UFOs, said that they'd created us but made a mistake and actually we were supposed to eat our children, and asked us to line up so they could insert the functioning child-eating gene in us, we would probably go all *Independence Day* on them; computers with more goal-directed architecture would if anything be even more willing to fight such changes.

If it really is a quick path from cow-level AI to superhuman-level AI, it would be really hard to test the cow-level AI for stability and expect it to stay stable all the way up to superhuman-level – superhumans have a lot more ways to cause trouble than cows do. That means a serious risk of superhuman AIs that want to do the equivalent of hurl us off cliffs, and which are very resistant to us removing that desire from them. We may be able to prevent this, but it would require a lot of deep thought and a lot of careful testing and prodding at the cow-level AIs to make sure they are as prepared as possible for the transition to superhumanity.

And we lose that option by making the AI open source. Make such a program universally available, and while Dr. Good is busy testing and prodding, Dr. Amoral has already downloaded the program, flipped the switch, and away we go.

V.

Once again: The decision to make AI findings open source is a tradeoff between risks and benefits. The risk is that in a world with hard takeoffs and difficult control problems, you get superhuman AIs that hurl everybody off cliffs. The benefit is that in a world with slow takeoffs and no control problems, nobody will be able to use their sole possession of the only existing AI to garner too much power.

But the benefits just aren't clear enough to justify that level of risk. I'm still not even sure exactly how the OpenAI founders visualize the future they're trying to prevent. Are AIs fast and dangerous? Are they slow and easily-controlled? Does just one company have them? Several companies? All rich people? Are they a moderate advantage? A huge advantage? None of those possibilities seem dire enough to justify OpenAI's tradeoff against safety.

Are we worried that AI will be dominated by one company despite becoming necessary for almost every computing application? Microsoft Windows is dominated by one company and became necessary for almost every computing application. For a while people were genuinely terrified that Microsoft would exploit its advantage to become a monopolistic giant that took over the Internet and something something something. Instead, they were caught flat-footed and outcompeted by Apple and Google, plus if you really want you can use something open-source like Linux instead. And new versions of Windows inevitably end up hacked and up on The Pirate Bay anyway.

Or are we worried that AIs will somehow help the rich get richer and the poor get poorer? This is a weird concern to have about a piece of software which can be replicated pretty much for free. Windows and Google Search are both fantastically complex products of millions of man-hours of research; Google is free and Windows comes bundled with your computer. In fact, people have gone through the trouble of creating fantastically complex competitors to both and providing *those* free of charge, to the point where multiple groups are competing to offer people fantastically complex software for free. While it's possible that rich people will be able to afford premium AIs, it is hard for me to weigh "rich people get premium versions of things" on the same scale as "human race likely destroyed". Like, imagine the sort of dystopian world where rich people had nicer things than the rest of us. It's too horrifying even to contemplate.

Or are we worried that AI will progress really quickly and allow someone to have completely ridiculous amounts of power? But remember, there's still a government and it tends to look askance on other people becoming powerful enough to compete with it. If some company is monopolizing AI and getting too big, the government will break it up, the same way they kept threatening to break up Microsoft when it was getting too big. If someone tries to use AI to exploit others, the government can pass a complicated regulation against that. You can say a lot of things about the United States government, but you can't say that they never pass complicated regulations forbidding people from doing things.

Or are we worried that AI will be so powerful that someone armed with AI is stronger than the government? Think about this scenario for a moment. If the government notices someone getting, say, a quarter as powerful as it is, it'll probably take action. So an AI user isn't likely to overpower the government unless their AI can become powerful enough to defeat the US military too quickly for the government to notice or respond to. But if AIs can do that, we're back in the intelligence explosion/fast takeoff world where OpenAI's assumptions break down. If AIs can go from zero to more-powerful-than-the-US-military in a very short amount of time while still remaining well-behaved, then we actually *do* have to worry about Dr. Evil and we shouldn't be giving him all our research.

Or are we worried that some big corporation will make an AI more powerful than the US government in secret? I guess this is sort of scary, but it's hard to get too excited about. So Google takes over the world? Fine. Do you think Larry Page would be a better or worse ruler than one of these people? What if he had a superintelligent AI helping him, and also everything was post-scarcity? Yeah, I guess all in all I'd prefer constitutional limited government, but this is another supposed horror scenario which doesn't even weigh on the same scale as "human race likely destroyed".

If OpenAI wants to trade off the safety of the human race from rogue AIs in order to get better safety against people trying to exploit control over AIs, they need to make a much stronger case than anything I've seen so far for why the latter is such a terrible risk.

There was a time when the United States was the only country with nukes. Aside from poor Hiroshima and Nagasaki, it mostly failed to press its advantage, bumbled its way into letting the Russians steal the schematics, and now everyone from Israel to North Korea has nuclear weapons and things are pretty okay. If we'd been so afraid of letting the US government have its brief tactical advantage that we'd given the plans for extremely unstable super-nukes to every library in the country, we probably wouldn't even be around to regret our skewed priorities.

Elon Musk famously said that AIs are “potentially more dangerous than nukes”. He’s right – so AI probably shouldn’t be open source any more than nukes should.

VI.

And yet Elon Musk is involved in this project. So are Sam Altman and Peter Thiel. So are a bunch of other people who have read Bostrom, who are deeply concerned about AI risk, and who are pretty clued-in.

My biggest hope is that as usual they are smarter than I am and know something I don’t. My second biggest hope is that they are making a simple and uncharacteristic error, because these people don’t let errors go uncorrected for long and if it’s just an error they can change their minds.

But I worry it’s worse than either of those two things. I got a chance to talk to some people involved in the field, and the impression I got was one of a competition that was heating up. Various teams led by various Dr. Amorals are rushing forward more quickly and determinedly than anyone expected at this stage, so much so that it’s unclear how any Dr. Good could expect both to match their pace *and* to remain as careful as the situation demands. There was always a lurking fear that this would happen. I guess I hoped that everyone involved was smart enough to be good cooperators. I guess I was wrong. Instead we’ve reverted to type and ended up in the classic situation of such intense competition for speed that we need to throw every other value under the bus just to avoid being overtaken.

In this context, the OpenAI project seems more like an act of desperation. Like Dr. Good needing some kind of high-risk, high-reward strategy to push himself ahead and allow at least some amount of safety research to take place. Maybe getting the cooperation of the academic and open-source community will do that. I won’t question the decisions of people smarter and better informed than I am if that’s how their strategy talks worked out. I guess I just have to hope that the OpenAI leaders know what they’re doing, don’t skimp on safety research, and have a process for deciding which results *not* to share too quickly.

But I am scared that it’s come to this. It suggests that we really and truly do not have what it takes, that we’re just going to blunder our way into extinction because cooperation problems are too hard for us.

I am reminded of what Malcolm Muggeridge wrote as he watched World War II begin:

> All this likewise indubitably belonged to history, and would have to be historically assessed; like the Murder of the Innocents, or the Black Death, or the Battle of Paschendaele. But there was something else; a monumental death-wish, an immense destructive force loosed in the world which was going to sweep over everything and everyone, laying them flat, burning, killing, obliterating, until nothing was left…Nor have I from that time ever had the faintest expectation that, in earthly terms, anything could be salvaged; that any earthly battle could be won or earthly solution found. It has all just been sleep-walking to the end of the night.

How Bad Are Things?

One “advantage” of working in psychiatry is getting a window into an otherwise invisible world of really miserable people.

I work in a wealthy, mostly-white college town consistently ranked one of the best places to live in the country. If there’s anywhere that you might dare hope wasn’t filled to the brim with people living hopeless lives, it would be here. But that hope is not realized. Every day I get to listen to people describe problems that would seem overwrought if they were in a novel, and made-up if they were in a thinkpiece on The Fragmentation Of American Society.

A perfectly average patient will be a 70 year old woman who used to live somewhere else but who moved her a few years ago after her husband died in order to be closer to family. She has some medical condition or other that prevents her from driving or walking around much, and the family she wanted to be closer to have their own issues, so she has no friends within five hundred miles and never leaves her house except to go to doctors’ appointments. She has one son, who is in jail, and one daughter, who married a drug addict. She also has one grandchild, her only remaining joy in the world – but her drug-addict son-in-law uses access to him as a bargaining chip to make her give him money from her rapidly-dwindling retirement account so he can buy drugs. When she can’t cough up enough quickly enough, he bans her from visiting or talking to the grandchild, plus he tells the grandchild it’s her fault. Her retirement savings are rapidly running out and she has no idea what she will do when they’re gone. Probably end up on the street. Also, her dog just died.

If my patients were to read the above paragraph, there are a handful who would sue me for breach of confidentiality, assuming I had just written down their medical history and gotten a couple of details like the number of children wrong. I didn’t. This is a *type*.

Here’s another. 60 year old guy who was abused as a child, still has visible scars. Ran off at age 15, got a job in a factory, married let’s say a waitress. There was some kind of explosion in his factory, he got PTSD, now he freaks out every time he steps within a hundred meters of a place where manufacturing is going on. Gradually stopped going outside because there were too many scary loud noises, his wife started yelling at him and telling him he was useless, he started beating his wife, put in jail for a year or two for domestic violence, came out, by this point his wife has run off with another man and took everything he owned with her. Moved in with an abusive uncle who is 80 years old and hates his guts, but the uncle needed a caretaker and the guy needed a place to live and they were each other’s only affordable option. Currently lives off disability payments, but the government keeps trying to cut them off, and he keeps having to spend what little he has on a lawyer to prevent them from taking even that away, but half the time he doesn’t make it to his lawyer appointments because he’s too nervous about going outside. Also he has chronic pain. Also he only sleeps two hours a night because of the nightmares, and he’s tired all the time.

(“You have the pill that fixes all of this, right, Doctor? The one they advertised on TV?”)

A while ago I wrote about how strongly we filter for people who are like us intellectually and politically:

> According to Gallup polls, about 46% of Americans are creationists. Not just in the sense of believing God helped guide evolution. I mean they think evolution is a vile atheist lie and God created humans exactly as they exist right now. That’s half the country.
>
> And I don’t have a single one of those people in my social circle. It’s not because I’m deliberately avoiding them; I’m pretty live-and-let-live politically, I wouldn’t ostracize someone just for some weird beliefs. And yet, even though I probably know about a hundred fifty people, I am pretty confident that not one of them is creationist. Odds of this happening by chance? 1/2^150 = 1/10^45 = approximately the chance of picking a particular atom if you are randomly selecting among all the atoms on Earth.
>
> About forty percent of Americans want to ban gay marriage. I think if I really stretch it, maybe ten of my top hundred fifty friends might fall into this group. This is less astronomically unlikely; the odds are a mere one to one hundred quintillion against.
>
> People like to talk about social bubbles, but that doesn’t even begin to cover one hundred quintillion. The only metaphor that seems really appropriate is a bizarre dark matter parallel universe.

Since starting working in psychiatry, I have realized that we also filter for misery. I think a big part of this is sorting by social class. But it’s in a more subtle way than you might think. That first patient, the 70 year old, might on paper have more than the median income if her dead husband’s pension is high enough. I could even imagine the second patient getting a decent payout from his factory and being financially in the clear for a while. It’s more complicated than that – something to do with being the sort of person who ends up in these sorts of situations.

I have three non-mutually exclusive theories for this:

1. The people who come to a psychiatrist are disproportionately the unhappiest and most disturbed. This is obviously true to some degree. But I got the same sort of people when I worked in general medicine and primary care. Even the people who come to a primary care doctor are going to be a little biased towards the sorts of conditions that produce or result from sickness, but people were still much worse off than I thought.

2. My ordinary life shields me from these people. I don't live in an especially bad neighborhood, so I won't meet the unhappiest people there. Unhappy people are really depressing, so their lives won't be covered as much by newspapers and TV. And insofar as they stay in their homes all the time and never come out or talk to anyone else, that in itself is going to prevent me from meeting them.

3. Or maybe many of the people I know are in fact this unhappy, but they never tell anyone except their psychiatrist all of the pieces necessary to put their life story together.

If it were mostly (1), that would be pretty encouraging and mean I'm just biased toward seeing very unlucky people. If it were mostly (2) or (3), that would be pretty bad, and mean everyone else is biased toward not realizing how unlucky everybody else is.

So I made a short script based on the following information:

– About 1% of people are in prison at any given time
– About 2% of people are on probation, which can actually

be really limiting and unpleasant
– About 1% of people are in nursing homes or hospices
– About 2% of people have dementia
– About 20% of people have chronic pain, though this varies widely with the exact survey question, but we are not talking minor aches here. About two-thirds of people with chronic pain describe it as “constant”, and half of people describe it as “unbearable and excruciating”.
– About 7% of people have depression in any given year
– About 2% of people are cognitively disabled aka mentally retarded
– About 1% of people are schizophrenic
– About 20% of people are on food stamps
– About 1% of people are wheelchair-bound
– About 7% of people are alcoholic
– About 0.5% of people are chronic heroin users
– About 5% of people are unemployed as per the official definition which includes only those looking for jobs
– About 3% of people are former workers now receiving disability payments
– About 1% of people experience domestic violence each year
– About 10% of people were sexually abused as children, many of whom are still working through the trauma.
– Difficult to get statistics, but possibly about 20% of people were physically abused as children, likewise.
– About 9% of people (male and female) have been raped during their lifetime, likewise.

These numbers might be inflated, since I took them from groups working on these problems and those groups have every incentive to make them sound as bad as possible. There’s also a really big problem where a lot of these are conditional upon one another – that is, a person in prison is not also in a nursing home, but a person who is unemployed is far more likely to be on food stamps. This will likely underestimate both the percent of people who have no problems at all, and the percent of people who have multiple problems at once.

Nevertheless, I ran the script twenty times to simulate twenty different people, and here's what I got (NP stands for "no problems"):

01. Chronic pain
02. Alcoholic
03. Chronic pain
04. NP
05. NP
06. Sexually molested as a child + suffering from domestic violence
07. Unemployed
08. Alcoholic
09. NP
10. NP
11. NP
12. Abused as a child
13. NP
14. Chronic pain
15. NP
16. Abused as a child + unemployed
17. NP
18. Alcoholic + on food stamps
19. NP
20. Clinically depressed

If the two problems mentioned above haven't totally thrown off the calculations, this makes me think Psychiatrist-Me is getting a much better window into reality than Normal-Person-Me.

And remember, this doesn’t count all of the problems that don’t fall into easily quantified categories, like “everyone hates them because they’re really ugly and annoying”. It doesn’t count things that I couldn’t find good statistics on, like “had a child die recently”. It doesn’t count things that I would have gotten in trouble for including, like “autistic” or “single mother”. It doesn’t count *a lot of things*. Consider that the first patient I mentioned – the homebound seventy year old with no friends who’s being extorted by her drug addict son-in-law – would appear on this list as “NP”.

The world is almost certainly a much worse place than any of us want to admit. And that’s *before you’ve even left America*.

This is part of why I get enraged whenever somebody on Tumblr says “People in Group X need to realize they have it really good”, or “You’re a Group X member, so stop pretending like you have real problems.” The town where I practice psychiatry is mostly white and mostly wealthy. That doesn’t save it. And whenever some online thinkpiece writer laughs about how good people in Group X have it and how hilarious it is that they sometimes complain about their lives, it *never fails* that I have just gotten home from treating a member of Group X who attempted suicide.

This is also why I am wary whenever people start boasting about how much better we’re doing than back in the bad old days. That precise statement seems to in fact be true. But people have a bad tendency to follow it up with “And so now most people have it pretty good”. I don’t think we have any idea how many people do or don’t have it pretty good. Nobody who hasn’t read polls would intuitively guess that 40-something percent of Americans are young-Earth creationists. How should they know how many people have it pretty good or not?

I think about all of the miserable people in my psychiatric clinic. Then I multiply by ten psychiatrists in my clinic. Then I multiply by ten similarly-sized clinics in my city. Then I multiply by a thousand such cities in the United States. Then I multiply by hundreds of countries in the world, and by that time my brain has mercifully stopped being able to visualize what that signifies.

This wasn’t supposed to be a Christmas post, but it took me longer than I expected to write, so here we are.

And this wasn’t supposed to be advocating any particular response, but I was recently asked to plug Giving What We Can’s pledge drive, and maybe one of the good responses to realizing how awful things are is committing to donate a little bit of what you’ve got to making them better.

Things That Are Not Superintelligences

[Not the most interesting topic in the world, but I'm posting it so I have something to link to next time I see this argument]

I talk about superintelligence a lot, usually in the context of AI or genetically engineered humans. And lately I have run into people who say: "But superintelligence already exists! It's corporations / bureaucracies / teams / civilizations / mind-mapping software" (examples: 1, 2, 3, 4). Sometimes these people go so far as to say these things are in fact *superintelligent AIs*, since they are technically "artificial".

No.

Some of these things may be poetically or metaphorically like a superintelligence, in the same way that, I don't know, the devastation of traditional cultures by modernity is poetically or metaphorically like nuclear war, or whatever. But if every time somebody is trying to talk about nuclear disarmament, other people interject with "But we've already had a nuclear war – it's the nuclear war *in the heart of all mankind*", this doesn't really add to the conversation. In the same way, talking about these metaphorical superintelligences is not a helpful contribution to discussion of literal superintelligences.

Why do I think that there is an important distinction between these kind of collective intelligences and genuine superintelligence?

There is no number of chimpanzees which, when organized into a team, will become smart enough to learn to write.

There is no number of ordinary eight-year-olds who, when organized into a team, will become smart enough to beat a grandmaster in chess.

There is no number of ordinary IQ 90 construction workers who, when organized into a team, will become smart enough to design and launch a space probe to one of the moons of Jupiter.

There is no number of “average” mathematics PhDs who, when organized into a team, will become smart enough to come up with the brilliant groundbreaking proofs of a Gauss or a Ramanujan.

Teams / corporations / cultures have a lot of advantages over individuals. They can use writing and record-keeping to have much better “memories”. They can use computers to be able to calculate and retrieve information more quickly. They can pool their advantages, so that if one person is good at writing and another person good at illustration, they can produce a well-written and beautifully-illustrated book. They can formalize their decision-making processes to route around various biases and react consistently to predictable situations. These are all really good things to be able to do, and it’s why in fact groups of people have outperformed individuals in fields as diverse as “making nuclear bombs” and “coordinating air traffic”.

But there is some aspect of intelligence that they can’t simulate, in which they are forever doomed to be only as smart as their smartest member (if that!). It’s hard to put my finger on exactly, but it seems to have something to do with creative problem-solving ability. A team of people smart enough to solve problems up to Level N may be able to work in parallel to solve many more Level N problems in a given amount of time. But it won’t be able to solve Level

N+1 problems. This is why it’s still occasionally useful to have mathematical geniuses around, instead of taking ten average mathematicians and telling them to work together. And unfortunately, this aspect of intelligence is the bottleneck for lots of interesting things like new inventions, proofs, and discoveries.

Further, teams themselves need intelligent people to run in an intelligent way. Steve Jobs led Apple to success by being really really good at marketing. Apple couldn’t have gotten the same results by firing him and replacing him with a marketing department of a hundred low-level employees who had graduated from second-tier marketing programs. This is true not only at the Steve Jobs level but at every level – at some point a Sales Department needs to have good salespeople, not just many well-organized mediocre salespeople. I’m not denying that many well-organized mediocre salespeople can do way better than a few poorly-organized mediocre salespeople, just that you can’t fully route around the need for actually intelligent people.

And finally, teams have a lot of contingent disadvantages over an individual. They work vastly more slowly. Their various parts tend not to know what the other parts are doing. If dictatorial in structure, they fall prey to failures of information; if non-dictatorial, to failures of coordination. Imagine an individual human whose inner soul had Democratic and Republican parties that were constantly trying to sabotage each other, so that if the Democratic part of her got a job interview, the Republican part would immediately try to sabotage the job interview to prevent the Democrats from looking good. Such a person would either be insane or at the very least not get too many jobs.

While it’s possible for improvements in organizational technology to ameliorate some of these contingent problems, so far they generally haven’t: the US government is as dysfunctional as ever, and a lot of corporations are little better. And even if all of the contingent problems were magically solved, that still leaves the fundamental problem where no organization of chimpanzees will ever write a novel.

If we were to actually get superintelligence, that would be a completely different class of entity than another government or corporation. It would also have all of the advantages of these things – arbitrarily much parallel processing ability, arbitrarily much memory, arbitrarily many computational resources – without the disadvantages, and with higher genuine “intelligence” as in problem-solving ability.

I think it’s useful to have a word for this completely different class of things, and that word is “superintelligence”. Teams, corporations, and cultures can use words we already have, like “groups”.

[**EDIT:** I keep getting the same objection in the comments: if we made a bunch of ordinary eight-year olds follow a simple set of operations that corresponded to a logic gate, and arranged them so that they simulated the structure of Deep Blue, then they could win high-level chess games. This is true. But *eight-year-olds could not come up with and implement this idea*. A brilliant computer programmer might be able to, but once you’re a brilliant computer programmer, you might as well just build the darned computer instead of implementing it on eight year olds. And any computer programmer so brilliant that they could build a true superintelligence out of eight year olds could build a true superintelligence out of normal computers too. And the same is true of the objection “doesn’t this mean that no amount of stupid neurons could combine into a smart human brain?” Yes, evolution can play the role of the brilliant computer programmer and turn neurons into a working brain. But it’s the organizer – whether that organizer is a brilliant human programmer or an evolutionary process – who is actually doing the work. That “neurons can combine to form a brain” is no more profound than that “transistors can combine to form an AI” – in both cases, it’s the outside organizer doing all the meaningful work. For a really interesting science-fiction treatment of what it would *actually* mean to implement a superintelligence in human social interaction, read Karl Schroeder’s *Lady of Mazes*]

[**EDIT 2:** Advocate of a “super-individual-human-intelligence” project responds with a clarification of terms.]

[**EDIT 3:** Several people point out in the comments that chess champion Garry Kasparov once played a game of chess against “the world”. Kasparov moved the white pieces, and the black moves were decided by popular vote on a website where various other grandmasters and chess buffs worked together to devise the best strategies. The game was very closely fought and suffered from several irregularities, but Kasparov ended up winning.]

Introducing Unsong

I.

In retrospect, there had been omens and portents.

(“We are now approaching lunar sunrise,” said William Anders, “and for all the people back on Earth, the crew of Apollo 8 has a message that we would like to send to you.”)

Rivers flowed uphill. A new star was seen in the night sky. A butchered pig was found to have the word “OMEN” written on its liver in clearly visible letters.

(“In the beginning, God created the heaven and the earth. And the earth was without form, and void; and darkness was upon the face of the deep.”)

Lightning struck in clear weather. Toads fell from the clouds. All ten thousand lakes in Minnesota turned to blood; scientists blamed “phytoplankton”.

(“And the Spirit of God moved upon the face of the waters. And God said, Let there be light: and there was light. And God saw the light, that it was good: and God divided the light from the darkness.”)

A majestic golden eagle flew onto the Vatican balcony as Pope Paul VI was saying Mass. The bird gingerly removed the Pontiff’s glasses with its beak, then poked out his left eye before flying away with an awful shriek.

(“And God called the light Day,” said Jim Lovell, “and the darkness He called Night. And the evening and the morning were the first day.”)

A beached whale was found hundreds of miles inland. A baby was born with four eyes.

(“And God said, Let there be a firmament in the midst of the waters, and let it divide the waters from the waters.”)

Pieces of paper with the word “OMEN” written on them fell from the clouds. A beached whale was seen in the night sky. Babies left unattended began to roll slowly, but unmistakeably, uphill.

(“And God made the firmament, and divided the waters which were under the firmament from the waters which were above the firmament: and it was so. And God called the firmament Heaven. And the evening and the morning were the second day.”)

One of the additional eyes on the four-eyed baby was discovered to be the left eye of Pope Paul VI, missing since the eagle incident. The provenance of the fourth eye was never determined.

(“And God said, Let the waters under the heaven be gathered together unto one place,” said Frank Borman, “and let the dry land appear: and it was so.”)

A series of very precise lightning strikes seared the word “OMEN” into the rust-red sand of the Sonora Desert; scientists blamed “phytoplankton”.

(“And God called the dry land Earth; and the gathering together of the waters called he Seas: and God saw that it was good.”)

The New York Stock Exchange rose by perfect integer amounts eleven days in a row. An obstetrician published an article in an obscure medical journal claiming that the kicks of unborn children, interpreted as Morse Code, formed unspeakable and blood-curdling messages.

(“And from the crew of Apollo 8, we close with good night, good luck, a Merry Christmas – and God bless all of – ” [sudden burst of static, then silence])

II.

If I had to choose a high point for the history of the human race thus far, it would be December 24, 1968.

1968 had been a year of shattered dreams. Martin Luther King was murdered in April. Democratic golden boy Robert Kennedy was murdered in June. Soviet tanks crushed the Prague Spring in August. It felt like each spark of hope for a better world was being snuffed out, methodically, one by one.

Then almost without warning, Americans turned on their televisions and learned that a spaceship was flying to the moon. On December 22, the craft beamed a live TV broadcast to Earth informing viewers that they were about to become the first humans ever to approach another celestial body. Communications issues limited the transmission to seventeen minutes, but the astronauts promised a second installment from lunar orbit.

On December 24, 1968, one billion people – more than for any television program before or after in the history of mankind – tuned in for Apollo 8’s short broadcast. The astronauts were half-asleep, frazzled with days of complicated calculations and near-disasters – but their voices were powerful and lucid through the static. Commander Frank Borman introduced the two other members of the crew. They described the moon, as seen up close. “A vast, lonely, forbidding expanse of nothing”. “A very foreboding horizon, a rather dark and unappetizing looking place”. Then the Earth, as seen from afar. “A green oasis, in the big vastness of space.”

Two minutes left till lunar sunrise broke the connection. The astronauts’ only orders from NASA had been to “do something appropriate”

“In the beginning,” read Bill Anders, “God created the heaven and the earth. And the earth was without form, and void; and darkness was upon the face of the deep.”

So for two minutes on Christmas Eve, while a billion people listened, three astronauts read the Book of Genesis from a tiny metal can a hundred miles above the surface of the moon.

Then, mid-sentence, they crashed into the crystal sphere surrounding the world, because it turned out there were far *fewer* things in Heaven and Earth than were dreamt of in almost anyone’s philosophy.

Everyone has been so kind and encouraging about my short stories that I’m ready to try writing some longer fiction. You can follow along at unsongbook.com, where I’ll be posting new chapters every Sunday and new interludes some Wednesdays. Right now it’s just this prologue and an option to subscribe by email to future updates, but I’ll have the first chapter up by Sunday, January 3.

This is going to be fun.

2016

- 2015 Predictions: Calibration Results
- Guns And States
- Guns and States 2: Son of a Gun
- Slow But Steady
- Schizophrenia: No Smoking Gun
- Lies, Damned Lies, And The Media (Part 6 of ∞)
- Side Effects May Include Anything
- Predictions For 2016
- Staying Classy
- Book Review: Superforecasting
- List Of Passages I Highlighted In My Copy Of “Superforecasting”
- Testimonials for SSC
- Before You Get Too Excited About That GitHub Study…
- Nonfiction Writing Advice
- Two Attitudes In Psychiatry
- 2016 Nootropics Survey Results
- Against Interminable Arguments
- Reverse Voxsplaining: Prison and Mental Illness
- Non-Shared Environment Doesn’t Just Mean Schools And Peers
- Book Review: The Art Of The Deal
- The Price Of Glee In China
- Book Review: My Brother Ron

- Beware Regional Scatterplots
- The Ideology Is Not The Movement
- A Theory About Religion
- SSC Bay Area Meetups 2016
- Book Review: Albion’s Seed
- Why Were Early Psychedelicists So Weird?
- Be Nice, At Least Until You Can Coordinate Meanness
- Myers’ Race Car Versus The General Fitness Factor
- Hidden Test Post
- Solidarity
- Skin In The Game
- Teachers: Much More Than You Wanted To Know
- Three Great Articles On Poverty, And Why I Disagree With All Of Them
- Book Review: Age of Em
- Ascended Economy?
- Ketamine Research In A New Light
- Three More Articles On Poverty, And Why They Disagree With Each Other
- Against Dog Whistle-ism
- Book Review: Unlearn Your Pain
- I Wrote A Blog Post, But Did Not Adjust For The Fact That The Title Would Be Too L
- Survey Results: Suffering Vs. Oblivion
- 3/4
- Pushing And Pulling Goals
- Things Probably Matter
- How The West Was Won
- Post-Partisanship Is Hyper-Partisanship

- Book Review: PiHKaL
- Tolerance Troubles
- The View From Ground Level
- Devoodooifying Psychology
- Reverse Voxsplaining: Drugs vs. Chairs
- Terrorists Vs. Chairs: An Outlier Story
- Reverse Voxsplaining: Brand-Name Drugs
- It’s Bayes All The Way Up
- Some Context For That NYT Sugar Article
- AI Persuasion Experiment
- SSC Endorses Clinton, Johnson, Or Stein
- He Kept Us Out Of War?
- Somewhat Against Psychiatric Conditions As Domestication Failure
- More Hardball Debate Questions
- The Moral Of The Story
- AI Persuasion Followup Survey
- The Heart Has Its Reasons That Reason Knows Not Of
- AI Persuasion Experiment Results
- How Does Recent AI Progress Affect The Bostromian Paradigm?
- SSC San Jose Meetup 11/4
- The Pyramid And The Garden
- Tuesday Shouldn’t Change The Narrative
- Book Review: House of God
- You Are Still Crying Wolf
- The Alzheimer Photo
- SSC Journal Club: Expert Prediction Of Experiments
- Contra Robinson On Schooling

- **Highlights From The Comment Thread On School Choice**
- **Thin Air**
- **Might People On The Internet Sometimes Lie?**
- **SSC Journal Club: Mental Disorders As Networks**
- **Book Review: Mount Misery**
- **Contra NYT On Economists On Education**
- **Addendum To “Economists On Education”**
- **2016 Predictions: Calibration Results**

2015 Predictions: Calibration Results

Many people are making predictions for the new year around now. I see fewer people going back and grading the accuracy of their predictions about last year, even though that obviously has a lot of relevance for how seriously we should take this year's predictions.

In keeping with the SSC tradition, here are the results for my (late) predictions for 2015. Successful predictions are normal-looking, failed predictions are ~~crossed out~~:

World Events

1. US will not get involved in any new major war with death toll of > 100 US soldiers: 70%
2. North Korea's government will survive the year without large civil war/revolt: 95%
3. Greece will not announce it's leaving the Euro: 60%
3. Neither Russia nor Qatar will lose their World Cups: 80%
4. Ebola will kill fewer people in second half of 2015 than the in first half: 95%
5. No terrorist attack in the USA will kill > 100 people: 90%
6. Assad will remain President of Syria: 70%
7. Israel will not get in a large-scale war (ie >100 Israeli deaths) with any Arab state: 90%
8. Syria's civil war will not end this year: 80%
9. ISIS will control less territory than it does right now: 70%
10. ISIS will continue to exist: 80%
11. Iran will reach a deal with the West on nuclear weapons: 80%
12. No major civil war in Middle Eastern country not currently experiencing a major civil war: 90%
13. Iraq's situation not to get any worse (eg gov't collapse, new rebellion): 60%
14. Obamacare will survive the year mostly intact: 60%

15. Hillary Clinton will be the top-polling Democratic Presidential candidate: 95%
~~16. Jeb Bush will be the top-polling Republican candidate: 50%~~
~~17. Trans-Pacific Partnership to pass at least mostly intact: 60%~~
18. US official unemployment rate will be less than 7% in Dec 2015: 95%
19. Bitcoin will end the year higher than $200: 95%
~~20. Oil will end the year greater than $60 a barrel: 50%~~

Personal Life
21. SSC will remain active: 95%
~~22. SSC will get fewer hits in the second half of 2015 than the first half: 60%~~
~~23. At least one SSC post in the second half of 2015 will get > 100,000 hits: 70%~~
24. Shireroth will remain active: 90%
25. I will remain at my same job through the end of 2015: 95%
~~26. There will be no further ramifications or lawsuits from either side over the flooding of my house: 80%~~
27. I will reach my savings target: 90%
28. I will get a score at >95th percentile for my year on PRITE: 50% (unknown, haven't gotten score back)
~~29. I will be involved in at least one published/accepted-to-publish research paper by the end of 2015: 60%~~
30. I will not break up with any of my current girlfriends: 80%
~~31. I will not get any new girlfriends: 50%~~
~~32. I will not finish [project]: 60%~~
33. I will attend NYC Solstice ritual: 80%
34. I will flake out of my plan to lead some kind of Solstice Ritual myself: 60%
35. I will be living in the house I'm currently trying to arrange to rent: 70%

Scoring

Of items I marked as 50% confident, 0 were right and 3 were wrong

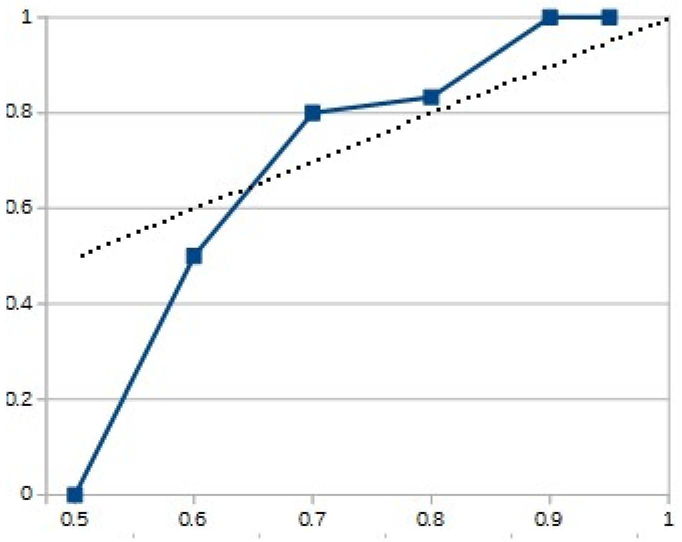
Of items I marked as 60% confident, 4 were right and 4 were wrong

Of items I marked as 70% confident, 4 were right and 1 was wrong

Of items I marked as 80% confident, 5 were right and 1 was wrong

Of items I marked as 90% confident, 4 were right and 0 were wrong

Of items I marked as 95% confident, 7 were right and 0 were wrong



As usual, the dotted line represents perfect calibration; the closer my blue line comes to that, the better I'm doing.

The graph looks like there's a massive failure at 50%, but this is just an artifact of very few questions at that level. If I'd gotten just one more right, I would be at 33%, ie as close to 50% as it's possible to get with a set of three. Given that the difference between total success and total failure was just one question, I don't feel too bad about total failure.

Everything else looks pretty good. I’m prepared to call this another successful year.

A side note: Scott Adams has also graded his predictions from the past year, and reports incredible success: 9/9 correct despite going way out on a limb and saying things everyone else found really unlikely (like that Trump would stay Republican front-runner). The obvious way to accomplish like that is to make lots of things that vaguely sound like predictions, then only highlight and count the ones that end up correct; after a quick scan of Adams’ blog, there’s no sign that he’s doing this; his win seems pretty genuine. Another method might be to make vague predictions and grade them in your favor, and there is some sign of this – for example, someone going off Clinton’s poll numbers versus Sanders could say they are in fact still going up. Nevertheless, it’s obviously been a good year for Adams, and I’d be fascinated to see him make a list of official concrete predictions for 2016, all in one place, maybe even associated with confidence levels. [EDIT: Someone else’s more pessimistic analysis; this makes it more pressing that he do everything beforehand this time]

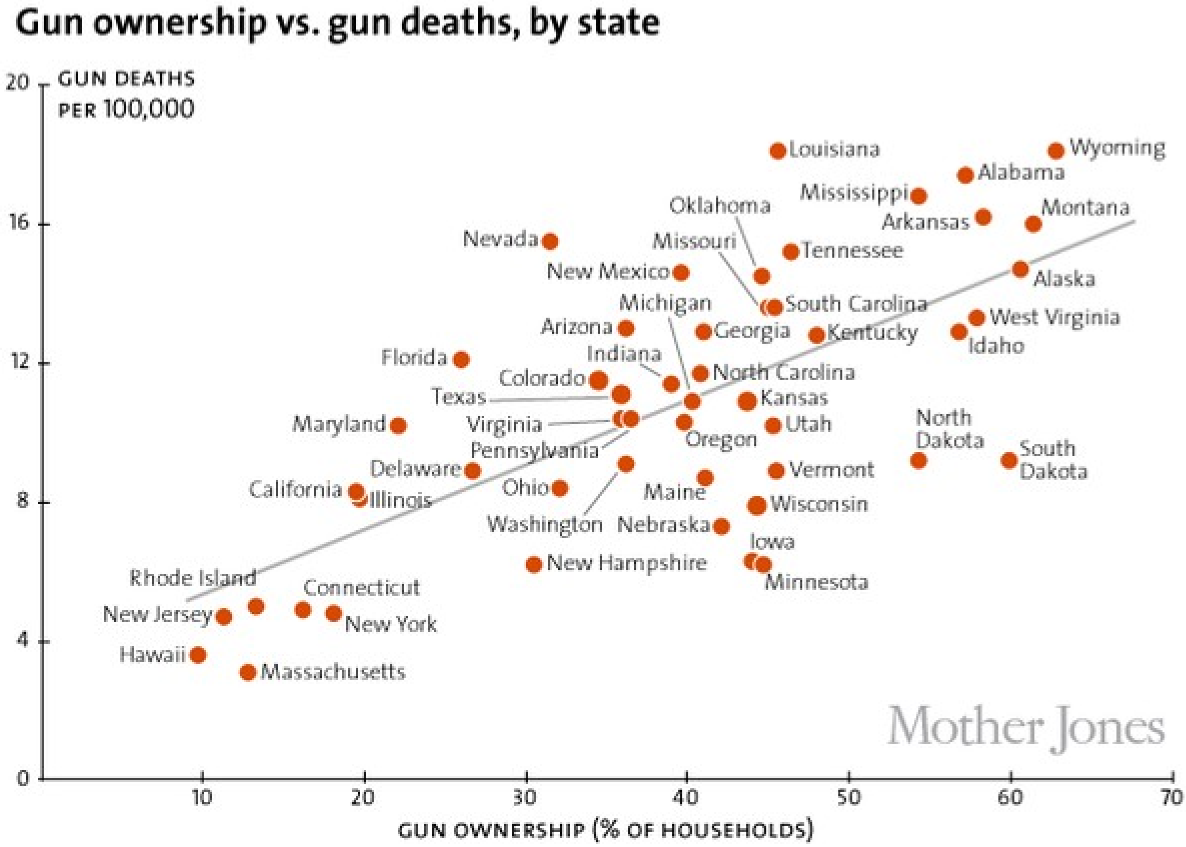
Guns And States

[Epistemic status: I think I probably wrung the right conclusions out of this evidence, but this isn't the only line of evidence bearing on the broader gun control issue and all I can say is what it's consistent with. Content warning for discussion of suicide, murder, and race]

I.

From a Vox article on America's Gun Problem, Explained: "On Wednesday, it happened again: There was a mass shooting — this time, in San Bernardino, California. And once again on Sunday, President Barack Obama called for measures that make it harder for would-be shooters to buy deadly firearms."

Then it goes on to say that "more guns mean more gun deaths, period. The research on this is overwhelmingly clear. No matter how you look at the data, more guns mean more gun deaths." It cites the following chart:

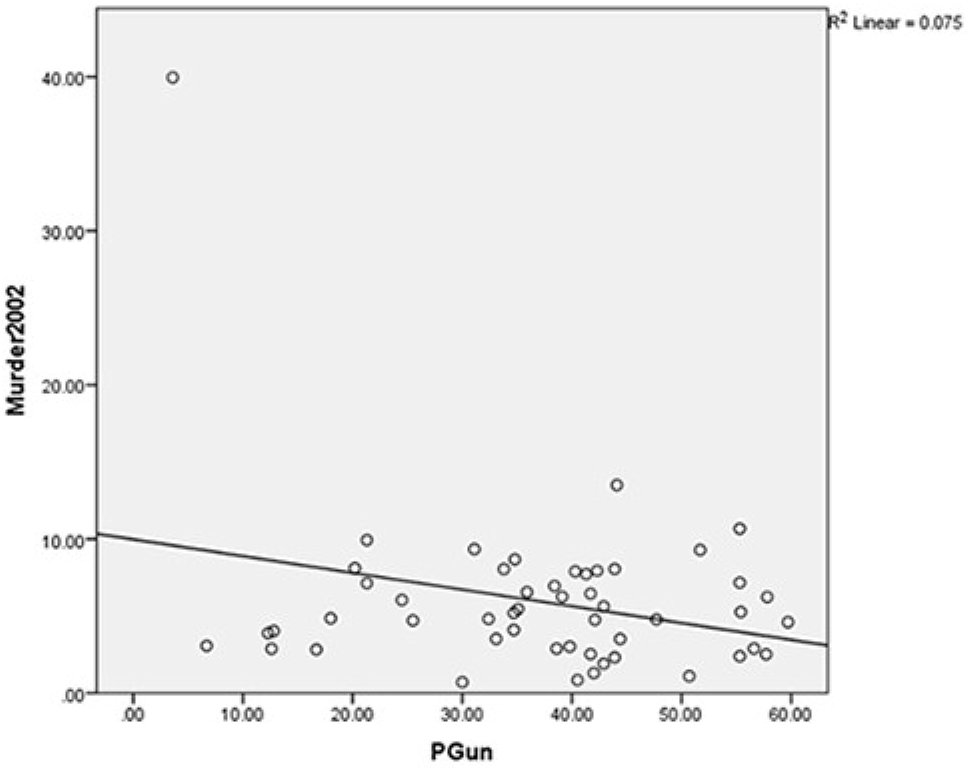


…then uses the graph as a lead in to talk about active shooter situations, gun-homicide relationships, and outrage over gun massacres.

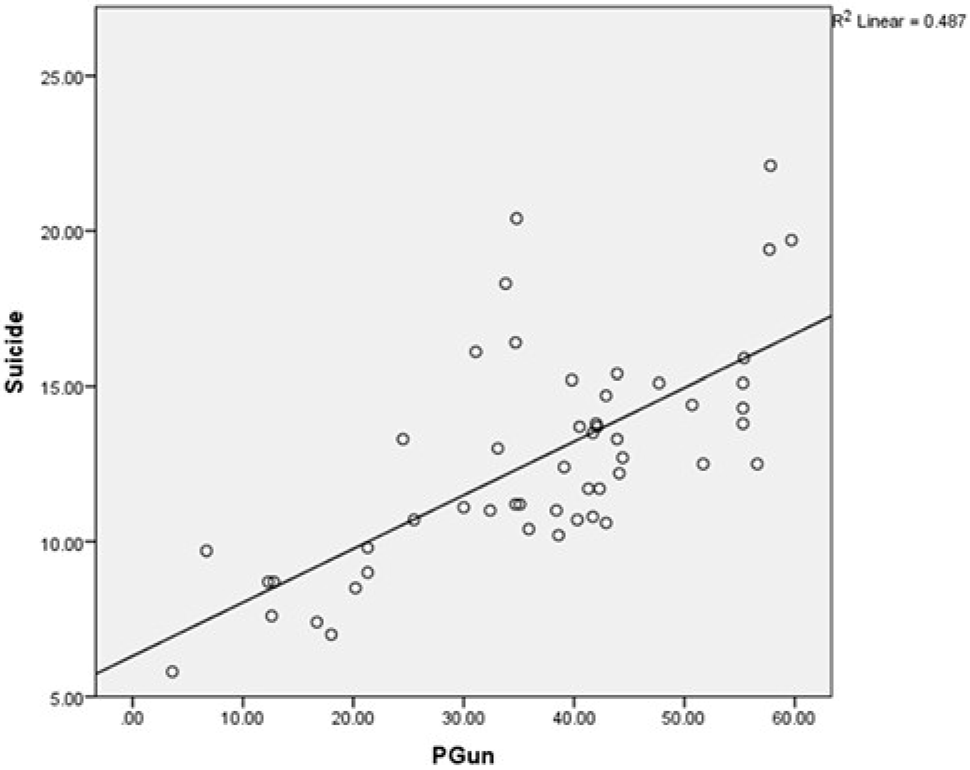
Did you notice that the axis of this graph says "gun deaths", and that this is a totally different thing from gun murders?

(this isn't an isolated incident: Vox does the same thing here and here)

Gun deaths are a combined measure of gun homicides and gun suicides. Here is a graph of guns vs. gun homicides:



And here is a graph of guns vs. gun suicides:



The relationship between gun ownership and homicide is weak (and appears negative), the relationship between gun ownership and suicide is strong and positive. The entire effect Vox highlights in their graph is due to gun suicides, but they are using it to imply conclusions about gun homicides. This is why you shouldn't make a category combining two unlike things.

II.

I am not the first person to notice this. The Washington Examiner makes the same criticism of Vox's statistics that I do. And Robert VerBruggen of National Review does the same analysis decomposing gun deaths into suicides and homicides, and like me finds no correlation with homicides.

German Lopez of Vox responds here. He argues that VerBruggen can't just do a raw uncontrolled correlation of state gun ownership with state murder rates without adjusting for confounders. This is true, although given that Vox has done this time and time again for months on end and all VerBruggen is doing is correctly pointing out a flaw in their methods, it feels kind of like an isolated demand for rigor.

So let’s look at the more-carefully-controlled studies. Lopez suggests the ones at the Harvard Injury Control Research Center, which has done several statistical analyses of gun violence. They list two such analyses comparing gun ownership versus homicide rates across US states: Miller Azrael & Hemenway (2002), and Miller Azrael & Hemenway (2007).

(does it count as nominative determinism when someone named Azrael goes into homicide research?)

We start with MA&H 2002. This study does indeed conclude that higher gun ownership rates are correlated with higher murder rates after adjusting for confounders. But suspiciously, it in fact finds that higher gun ownership rates are correlated with higher murder rates even *before* adjusting for confounders, something that we already found wasn’t true! Furthermore, even after adjusting for confounders it finds in several age categories that higher gun ownership rates are correlated with higher *non-gun homicide rates* (eg the rates at which people are murdered by knives or crowbars or whatever) at p less than 0.001. This is really suspicious! Unless guns are exerting some kind of malign pro-murder influence that makes people commit more knife murders, some sort of confounding influence has remained. Let’s look closer.

The study gets its murder rate numbers from the National Center for Health Statistics, which seems like a trustworthy source. It gets its gun ownership numbers from…oh, that’s interesting, it doesn't actually have any gun ownership numbers. It says that there is no way to figure out what percent of people in a given state own guns, so as a proxy for gun ownership numbers, it will use a measure called FS/S, ie the number of firearm suicides in a state divided by the total number of suicides.

This makes some intuitive sense. Among people who want to commit suicide, suppose a fixed percent prefer to use guns compared to other methods. In that case, the determining factor for whether or not they use a gun will be whether or not they have a gun. Hospitals diligently record statistics about suicide victims including method of suicide, so if our assumption holds this should be a decent proxy for gun ownership within a state.

There's only one problem – I checked this against an actual measure of gun ownership per state that came out after this study was published – the CDC asking 200,000 people how many guns they had as part of the Behavioral Risk Factor Surveillance System Survey – and the FS/S measure fails. When I repeat all of their analyses with their own FS/S measure, I get all of their same positive correlations, including the ones with non-gun homicides. When I repeat it with the real gun ownership data, all of these positive correlations disappear. When I look at exactly why this happens, it's because FS/S is *much* more biased towards Southern states than actual gun ownership is. Real gun ownership correlates very modestly – 0.25 – with 538's ranking of the Southern-ness of states. FS/S correlates at a fantastically high 0.62. For some reason, suicidal Southerners are much more likely to kill themselves with guns than suicidal people from the rest of the States, even when you control for whether they have a gun or not. That means that MA&H 2002 thought it was measuring gun ownership, but was actually measuring Southern-ness. This is why they found higher homicide rates, including higher rates of non-gun homicide.

So we move on to MA&H 2007. This study was published after the CDC's risk survey, so they have access to the same superior gun ownership numbers I used to pick apart their last study. They also have wised up to the fact that Southern-ness is important, and they include a dummy variable for it in their calculations. They also control for non-gun crime rate, Gini coefficient, income, and alcohol use. They do *not* control for urbanization level or race, but when I re-analyze their data including these factors doesn't change anything, likely because they are already baked in to the crime rate.

They find that even after controlling for all of this stuff, there is still a significant correlation between gun ownership level and gun homicide rate. Further, this time they are using good statistics, and there is *not* a significant correlation between gun ownership and non-gun-homicide rate. Further, there *is* a correlation between gun ownership and total homicide rate, suggesting that the gun-gun-homicide correlation was not just an artifact of people switching from inferior weapons to guns while still committing the same number of murders. Further, this is robust to a lot of different decisions about what to control or not to control, and what to include or not to include.

I repeated all of their analyses using two different sources of gun ownership data, a couple different sources of homicide and crime rate data, and a bunch of different plausible and implausible confounders – thanks a lot to Tumblr user su3su2u1 for walking me through some of the harder analyses. I was able to replicate their results. Pro-gun researcher John Lott had many complaints about this study, including that it was insensitive to including DC and that it was based entirely on the questionable choice of controlling for robbery rate – but I was unable to replicate his concerns and found that the guns-homicide correlation remained even after DC was included and even when I chose a group of confounders not including robbery rate. I was unable to use their methodology to replicate the effect in places where it shouldn’t replicate (I tried to convince it to tell me tractors caused homicide, since I was suspicious that it was just picking up an urban/rural thing, but it very appropriately refused to fall for it). Overall I am about as sure of this study as I have ever been of any social science study, ie somewhat.

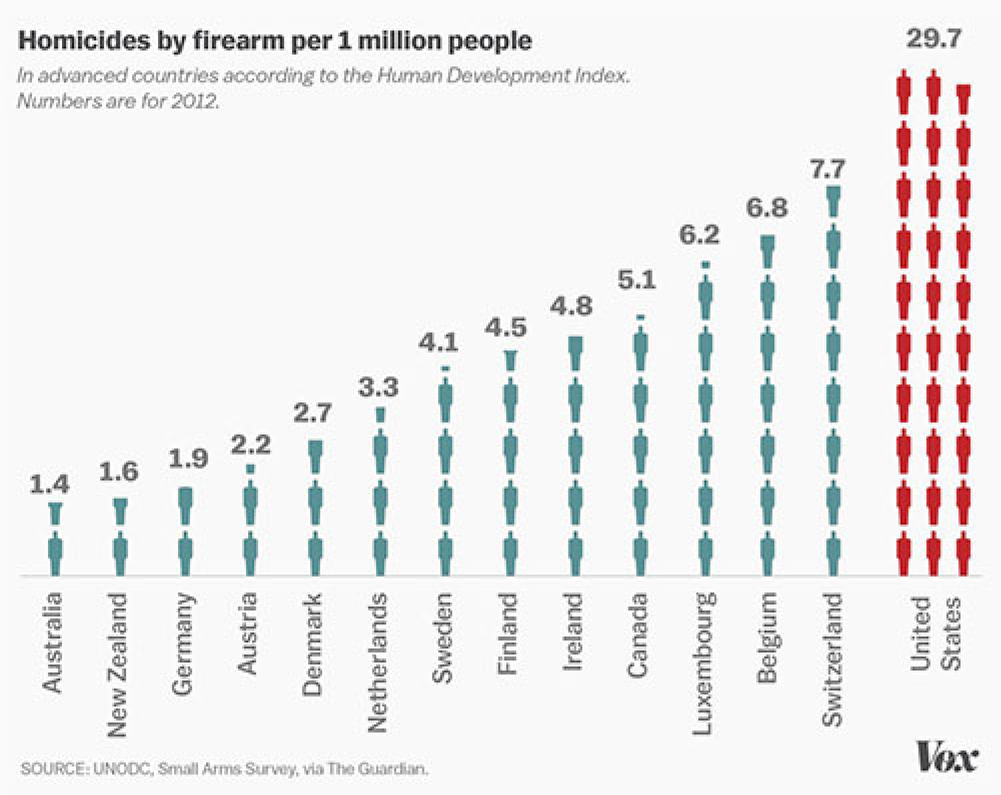
This study doesn’t prove causation; while one interpretation is that guns cause homicide, another is that homicide causes guns – for example, by making people feel unsafe so they buy guns to protect themselves. However, I doubt the reverse causation aspect in this case. The study controlled for robbery rate; ie it was looking at whether guns predicted homicides above and beyond those that could be expected given the level of non-homicide crime. My guess is that people feeling unsafe is based more on the general crime rate than on the homicide rate per se, which would make it hard for the homicide rate to cause increased gun ownership independently of the crime rate.

If guns are in fact correlated with more homicide, how come me and VerBruggen found the opposite in our simpler scatterplot analysis? This is complicated, but I think the biggest part of the answer is the urban/rural divide. Rural people have more guns. Murder rates are higher in urban areas. Race also plays a part: whites have more guns, but black areas have higher murder rates. Finally, the North and West seem to have more guns, but murder rates are highest in the South (which is what produced the bogus effect on the last study). All of these differences are large enough to cancel out the gun/no-gun difference and make the raw scatterplot look like nothing. This study didn't address all those things directly, but its decision to control for non-gun crime rate and poverty took care of them nevertheless. As the old saying goes, guns don't kill people; guns controlled for robbery rate, alcoholism, income, a dummy variable for Southernness, and a combined measure of social deprivation kill people.

If this is all true, how come I spent so much time yelling at that first study with worse data? Because I worry that if people only see the good studies, they'll get complacent. Vox posted these two studies as proof that there was a state-level gun-murder correlation. The first one was deeply flawed, but the second one turned out to be okay. Do you think Vox realized this? Do you think they would have written that article any differently in a world where *both* studies were flawed? As long as you trust every scientific paper you see – let alone every scientific paper you see on your side in a highly politicized field – even when you're right it will often just be by luck.

III.

Vox also voxsplains to us about America's unusually high gun homicide rate.



Having presented this graph, they say that "To understand why that is, there's another important statistic: The US has by far the highest number of privately owned guns in the world."

Even granting, as we saw above, that gun ownership does indeed increase homicide rates, this is not the most important factor in explaining America's higher homicide rate, or even close to the most important factor. Let me give a few arguments for why this must be the case:

1. The United States' homicide rate of 3.8 is clearly higher than that of eg France (1.0), Germany (0.8), Australia (1.1), or Canada (1.4). However, as per the FBI, only 11,208 of our 16,121 murders were committed with firearms, eg 69%. By my calculations, that means our nonfirearm murder rate is 1.2. In other words, our *non-firearm homicide rate alone* is higher than France, Germany, and Australia's *total* homicide

rate. Nor does this mean that if we banned all guns we would go down to 1.2 – there is likely a substitution effect where some murderers are intent on murdering and would prefer to use convenient firearms but will switch to other methods if they have to. 1.2 should be considered an absolute lower bound. And it is *still* higher than the countries we want to compare ourselves to.

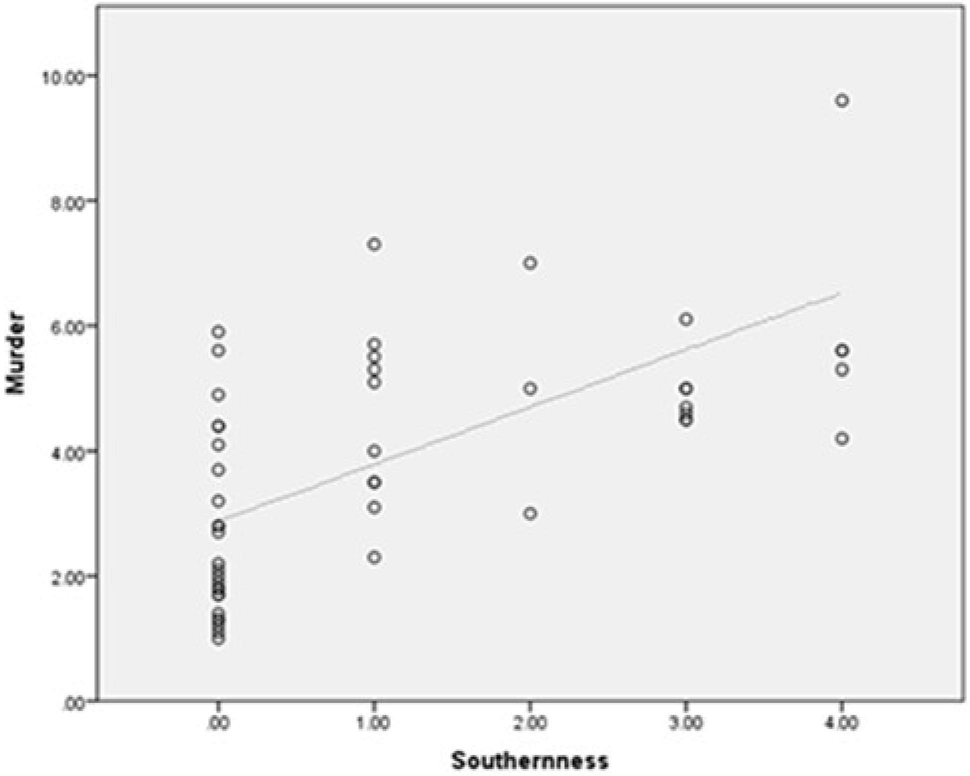
2. There are many US states that combine very high firearm ownership with very low murder rates. The highest gun-ownership state in the nation is Wyoming, where 59.7% of households have a gun (really!). But Wyoming has a murder rate of only 1.4 – the same as right across the border in more gun-controlled Canada, and only about a third of that of the nation as a whole. It seems likely that the same factors giving Canada a low murder rate give Wyoming a low murder rate, and that the factors differentiating the rest of America from Wyoming are the same factors that differentiate the rest of America from Canada (and Germany, and France…). But this does not include lower gun ownership.

3. There are many US states that combine very low firearm ownership with very high murder rates. The highest murder rate in the country is that of Washington, DC, which has a murder rate of 21.8, more than twenty times that of most European countries. But DC also has the strictest gun bans and the *lowest* gun ownership rate in the country, with gun ownership numbers less than in many European states! It seems likely that the factors making DC so deadly are part of the story of why America as a whole is so deadly, but these cannot include high gun ownership.

If not gun ownership, what *is* the factor making America so much more deadly than Europe and other First World countries? The traditional answer I always heard to this question was that America had a “culture of violence”. I always hated this answer, because it seemed so vague and meaningless as to be untestable by design. If the NRA waves their hands and says “eh, culture of violence”, how are you going to tell them they’re wrong?

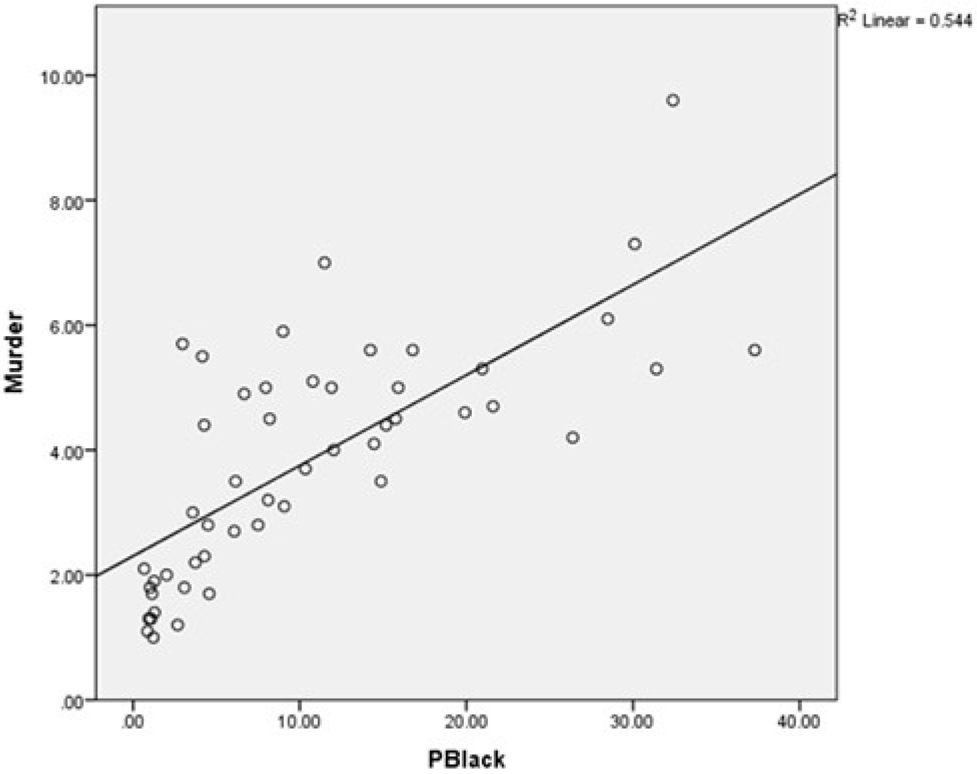
But we can work with this if we assume the culture of violence (or, if you want to be official about it, “honor culture”) is more common in some populations and areas than others. Some of the groups most frequently talked about during these lines are Southerners and various nonwhite minorities. This provides a testable theory: if we compare American non-Southern whites to European countries mostly made up of non-Southern whites, we’ll find similar murder rates. But first, some scatter plots:

This is murder rate by state, correlated with perceived Southernness of that state as per 538’s poll. I’ve removed DC as an outlier on all of the following.



> Who'd have thought populating half a country with the descendants of a group of people called "Border Reavers" would cause so much trouble?
>
> — Scott Alexander (@slatestarcodex) [December 3, 2015]

And this is murder rate by state correlated with percent black population:

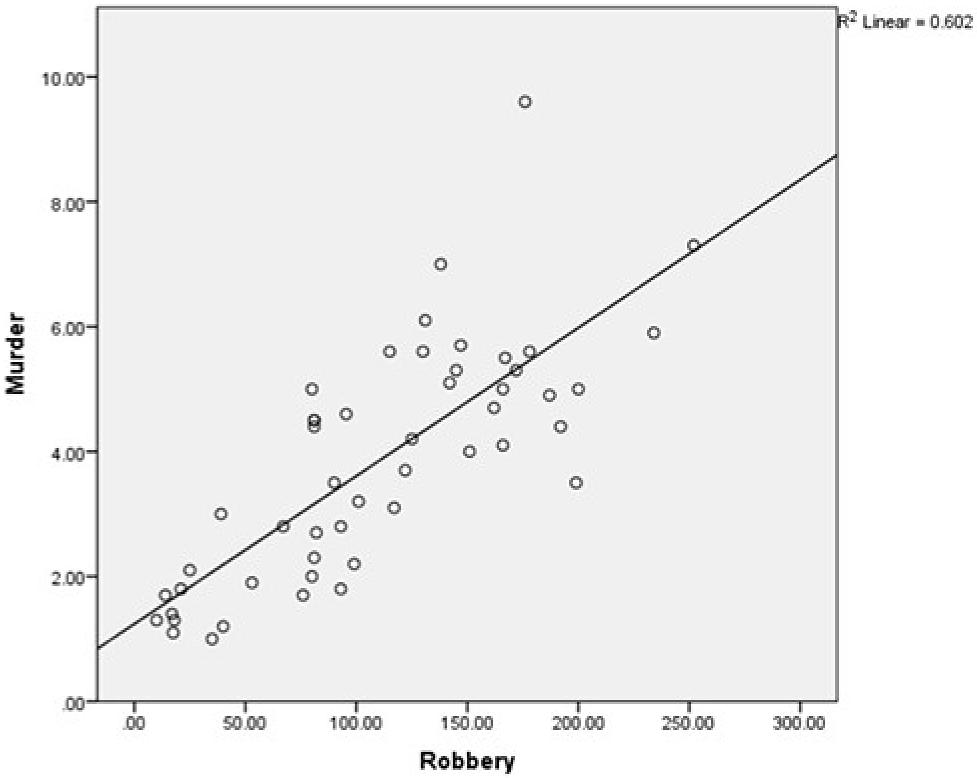


This would seem to support the “culture of violence” theory.

Can we adjust for this and see what the murder rate is for non-Southern whites? Sort of. The Economist gives a white-only murder rate of 2.5 (this is based on white victims, whereas we probably want white perpetrators, but the vast majority of murders are within-race so it doesn’t make much difference). And Audacious Epigone has put together a collection of white murder rates by state. I can’t find anything on non-Southern white murder rates per se, but one hack would be to take the white murder rate in non-Southern states and assume there aren’t any Southerners there.

Our main confounder will be urbanization. Western Europe is about 80% urban, so let’s look at states at a similar level. The four northern states that are closest to 80% urban are Colorado, Oregon, Washington, and Connecticut. I’m throwing out Colorado because it has a large Latino population who can’t be statistically differentiated from whites. That leaves, Washington (2.4), Connecticut (2.0), and Oregon (2.0). So possibly adjusting out Southerners brings us down from 2.5 (all whites) to 2.1 or so (non-Southern whites)? Again, compare to Germany at 0.8, Canada at 1.4, and America at 3.8.

There’s one more factor that needs to be considered:



This is a plot of the gun death rate vs. the robbery rate. There’s a strong correlation (r = 0.78). Robbery is heavily correlated with percent black, percent Southern, and urbanization, so it’s probably coming from the same place. Nevertheless, it seems to correlate with murder better than any of them alone, maybe because it’s combining all three measures together. I was able to make a linear model using those three measures that correlated at r = 0.79 with murder, about the same amount that robbery does. I should also mention that robbery correlates *negatively* with gun ownership at r = – 0.52, but this disappeared when controlled for urbanization.

So my very tentative conclusion is that although the US murder rate is much higher than that of other First World countries, this is partly due to the existence of various cultural factors not present in those other nations. When we adjust those away, America's murder rate falls from 3.8 to 2.1. Which is still higher than Germany's 0.8 or Canada's 1.4.

Is that extra due to guns?

IV.

According to MA&H 2007, each *absolute* percentage point in gun ownership was related to a 2.2 *relative* percentage point difference in homicide. This part of the study was beyond my ability to check, and I'm not sure why they switched from absolute to relative percents there, but suppose we take it seriously.

America has a gun ownership rate of 32%, so if we somehow decreased that to zero, we would naively expect about a 70% decrease in homicides. Unfortunately, only 67% of American homicides involve guns, so we're back to pretending that eliminating guns will not only have zero substitution effect but also magically prevent non-gun homicides. This shows the dangers of extrapolating a figure determined by small local differences all the way to the edge of the graph (I'M TALKING TO YOU, RAY KURZWEIL).

Maybe we can be more modest? Canada has a gun ownership rate of aboot 26%, so…

…wait a second. I thought we've been told that the US has a gun ownership rate seven zillion times that of any other country in the world, and that is why we are so completely unique in our level of gun crime? And now they're telling us that Canada has 26% compared to our 32%? What?

Don't trust me too much here, because I've never seen anyone else analyze this and it seems like the sort of thing there should be loads of analyses of if it's true, but I think the difference is between percent of households with guns vs. guns per capita. US and Canada don't differ very much in percent of households with guns, but America has about four times as many guns per capita. Why? I have no idea, but the obvious implication is that Canadians mostly stop at one gun, whereas Americans with guns buy lots and lots of them. In retrospect this makes sense; I am looking at gun enthusiast bulletin boards, and they're advising other gun enthusiasts that six guns is really the bare minimum it's possible to get by with (see also "How many guns can you have before it's okay to call your collection an 'arsenal'?", which I have to admit is not a question that I as a boring coastal liberal have ever considered). So if the guy asking that question decides he needs 100 guns before he gets his arsenal merit badge, that's a lot more guns per capita without increasing percent household gun ownership. This should actually be another argument that guns are not a major factor in differentiating US vs. Canadian murder rates, since unless you're going on a mass shooting (WHICH IS REALLY RARE) you wouldn't expect more murders from any gun in a household beyond the first. That means that the small difference between US and Canadian household percent gun ownership rates (32% vs. 26%) would have to drive the large difference between US and Canadian murder rates (1.4 vs. 3.8), which just isn't believable.

…okay, sorry, where were we? Canada has a gun ownership rate of about 26%, so if America were to get its gun ownership as low as Canada, that would be -6 absolute percentage points = a 13% relative decrease in murder rate = the murder rate going from 3.8 to 3.3 = a 0.5 point decrease in the murder rate. That's pretty close to the difference between our 2.1 US-sans-culture-of-violence estimate and the 1.4 Canadian rate – so maybe beyond the cultures of violence, the rest of the US/Canada difference really is due to guns?

(I'm not sure whether I should be subtracting 13% from 2.1 rather than 3.8 here)

In Germany, 9% of households own firearms (wait, really? European gun control is less strict than I thought!) Using MA&H's equation, we predict that if the US had the same gun ownership rate as Germany, its murder rate would drop 50%, eg from 3.8 to 1.9. Adjust out the culture of violence, and we're actually pretty close to real Germany's murder rate of 0.8.

How much would gun control actually cut US gun ownership? That obviously depends on the gun control, but a lot of people talk about Australia's gun buyback program as a model to be emulated. These people say it decreased gun ownership from 7% of people to 5% of people (why is this number so much lower than Canada and Germany? I think because it's people rather than households – if a gun owner is married to a non-gun-owner, they count as one gun-owner and one non-owner, as opposed to a single gun-owning household. The Australian household number seems to be 19% or so). So the gun buyback program in Australia decreased gun ownership by (relative) 30% or so. If a similar program decreased gun ownership in America by (relative) 30%, it would decrease it by (absolute) 10% and decrease the homicide rate by (absolute) 22%. Since there are about 13000 homicides in the US per year, that would save about 3000 lives – or avert about one 9/11 worth of deaths per year.

(note that our murder rate would still be 3.0, compared to Germany’s 0.8 and Canada’s 1.4. Seriously, I’m telling you, the murder rate difference is not primarily driven by guns!)

Is that worth it? That obviously depends on how much you like being able to have guns. But let me try to put this number into perspective in a couple of different ways:

Last time anyone checked, which was 1995, about 618,000 people died young (ie before age 65) in the US per year. Suppose that the vast majority of homicides are of people below 65. That means that instituting gun control would decrease the number of premature deaths to about 615,000 – in other words, by about half a percentage point. I’m having to borrow this data from the UK, but if it carries over, the average person my age (early 30s) has a 1/1850 chance of death each year. Gun control would decrease that to about 1/1860. I’m very very unsure about the exact numbers, but it seems like the magnitude is very low.

On the other hand, lives are very valuable. In fact, the statistical value of a human life in the First World – ie the value that groups use to decide whether various life-saving interventions are worth it or not – is $7.4 million. That means that gun control would “save” $22 billion dollars a year. Americans buy about 20 million guns per year (really)! If we were to tax guns to cover the “externality” of gun homicides preventable by Australia-level gun control, we would have to slap a $1000 tax on each gun sold. While I have no doubt that some people, probably including our arsenal collector above, would be willing to pay that, my guess is that most people would not. This suggests that most people probably do not enjoy guns enough to justify keeping them around despite their costs.

Or if all gun enthusiasts wanted to band together for some grand Coasian bargain to buy off the potential victims of gun violence, each would have to contribute $220/year to the group effort – not *totally* impossible, but also not something I can really see happening.

This is very, very, very, very *very* tentative, but based on this line of reasoning alone, without looking into the experimental studies or anything else, it appears that Australia-style gun control would probably be worth it, if it were possible.

(I didn't price in the advantages of guns in terms of preventing state tyranny and protecing freedom, which might be worth subsidizing, but my guess is that if 32% gun ownership is enough to maintain freedom, 22% gun ownership is as well)

V.

In summary, with my personal confidence levels:

1. Scatterplots showing raw correlations between gun ownership and "gun deaths" are entirely driven by suicide, and therefore dishonest to use to prove that guns cause murder (~100% confidence)

2. But if you adjust for all relevant confounders, there *is* a positive correlation between gun ownership and homicide rates (~90% confidence). This relationship is likely causal (~66% confidence).

3. The majority of the difference between America's murder rate and that of other First World countries is *not* because of easier access to guns in America (~90% confidence).

4. But some of it *is* due to easier access to guns. This is probably about 0.5 murders/100K/year.

5. An Australian-style gun control program *that worked and had no side effects* would probably prevent about 2,000 murders in the US. It would also prevent a much larger number of suicides. I am otherwise ignoring suicides in this piece because discussing them would make me too angry.

6. Probably the amount of lost gun-related enjoyment an Australian-style gun control program would cause do not outweigh the benefits.

7. This is not really enough analysis to make me have a strong opinion about gun control, since this just looks at the correlational evidence and doesn't really investigate the experimental evidence. Contrary to what everyone always tells you, experimental evidence doesn't always trump correlational – there are cases where each has its strengths – but it wouldn't be responsible to have a *real* opinion on this until I look into that too. Nevertheless, these data are at least *highly consistent with* Australia-style gun control being a good idea for the US.

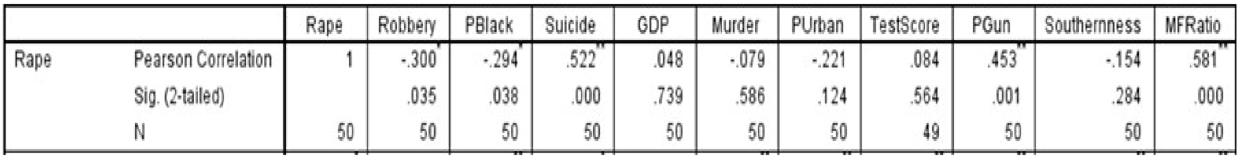
If you want to look into this more, here is a CSV version of all the relevant data.

Guns and States 2: Son of a Gun

Thanks for your many excellent comments on Guns and States. I've made it through *most* of the comments thread and learned a lot. Here are some (hopefully) final thoughts on the matter:

1. In yesterday's post, I suggested that the difference in homicide rates between America and other First World countries were about two-thirds cultural, one-third gun-related. That's sort of true, but people have reminded me to think of it as an *interaction*. Without the cultural factors in place, guns are pretty harmless. That's why Wyoming can be the highest-gun-ownership state in the country, with 60% of households having a weapon, and still have a murder rate equal to Canada's and lower than 45 other states'. *With* the cultural factors in place, guns make a bad situation worse. This is why the robbery rate and the gun homicide rate correlate at about 0.9 (0.75 if DC is removed as an outlier) even though most robberies do not involve firearms.

2. But some robberies do involve firearms, making them an imperfect control. I looked at rape, a crime that sounds like it should be associated with other violent crimes but which almost never involves firearms. Unfortunately, the results were *really weird*:



		Rape	Robbery	PBlack	Suicide	GDP	Murder	PUrban	TestScore	PGun	Southernness	MFRatio
Rape	Pearson Correlation	1	-.300*	-.294*	.522**	.048	-.079	-.221	.084	.453**	-.154	.581**
	Sig. (2-tailed)		.035	.038	.000	.739	.586	.124	.564	.001	.284	.000
	N	50	50	50	50	50	50	50	49	50	50	50

As you can see, rape is negatively correlated with robbery and %black, insignificantly correlated with homicide and urbanness, and positively correlated with gun ownership and suicide rate (which itself is heavily correlated with gun ownership). I also added one more variable in an attempt to explain some of this: MFRatio, which is the number of men per 100 women in the state. That certainly mattered a lot, but doesn't take away from the confusingness of the other variables. My first theory was that rape is more common in Red Tribe rural white culture, which would explain the guns and %black but really doesn't fit with the PUrban or Southernness issues, nor the fact that the correlation with guns was much higher than the racial correlation. My second was that people bought guns to defend against rape, but that doesn't explain the negative correlation with robbery, and surely people would buy guns based on crime in general (and since it's mostly men doing the gun-buying, rape would be the least likely to affect the gun rate). Overall I admit I am confused.

3. Commenter eccdogg can't replicate the Southern culture of violence thing after controlling for other factors. I acknowledge that he seems to be doing the right statistics, but can't square that with commonsense eyeballing of the data.

4. I did a reanalysis that found that (after adjusting for confounders) the gun data from 2002 successfully predicted the murder rate in 2002, but did not successfully predict the murder rate today, even though most other relevant variables did. I'm going to attribute this to gun ownership patterns changing faster than (say) whether a state is Southern or not, and probably if I had gun ownership data from today (which I don't) it would work, but it seemed important to mention that and get it out there.

5. I was shocked to see that between 20 to 30 percent of people in most European/Anglosphere countries owned guns, including the ones like Canada that gun control advocates hold up as an example of what they want. That makes it *very*

strange the degree to which people expect gun control to mean “the government confiscates all your guns, after which no guns are left”. I mean, I understand why (for example) the NRA would promote that story in order to get people angry, but it’s very strange how often liberals nod along and say “Yup, that’s what we want to do!”. I wish more of the debate could be about waiting periods, required training classes, and background checks – none of which would prevent people from using guns to defend against a potentially tyrannical government. (I don’t know enough about guns to be know whether restricting handguns would make it harder to defend against tyranny. Could you shoot the tyrants equally well with a rifle?)

6. Commenter Elias brings up this meta-analysis by Gary Kleck claiming that the 40 guns-and-states style studies he could find were split almost exactly half and half in terms of whether they found a significant guns-homicide correlation or not. He further claimed that the better the study, the *less* likely it was to find a significant correlation. The study I cited, plus my own analyses, passed two of Kleck’s tests – used good gun ownership numbers and controlled for confounders – but failed the third, which was distinguishing forward causation from reverse causation. I admitted this was a problem and that it would take very powerful methods beyond my abilities to solve, but I also said that I suspected forward causation since the homicide causation remained independent of robbery and I would expect people to buy guns based on a general crime rate. I stick to that suspicion. Kleck has some studies that he thinks establish reverse causation instead, and I encourage you to read them.

This reminds me a lot of my post Beware The Man Of One Study – I can at least say in my own defense that I did the analysis on my own, which should obviate exactly which studies I did or didn’t look at, but the media articles that just presented MA&H and said “Look! Here’s what Science says, discussion over!” *really* should have mentioned that there were also nineteen studies, 48% of the total, that disagreed with them.

On the other hand, Kleck said that only three studies met his quality standards, and all of them were his own, so take that with however much salt you feel like it deserves.

Slow But Steady

I saw this on *Giving What We Can* and was so delighted I wanted to repost it here. Source is The effect of malaria control on plasmodium falciparum in Africa between 2000 and 2015 in *Nature*:

GWWC comments:

> The reduction in malaria has been caused — in large part — by people sleeping under bednets…Long lasting insecticide-treated bednets are a powerful weapon against malaria, not only because they're a physical barrier between mosquitoes and sleeping children — the insecticide coating kills mosquitoes, so they don't infect other members of the family (and the village) who don't have mosquito nets…
>
> In other words, this paper tells us about the bigger picture, showing that bednets are incredibly effective, not just at the level of individual villages, but at the level of whole populations. Essentially, the case that we should be distributing bednets just got even stronger.
>
> The research suggests that anti-malarial interventions have prevented about 663 million malarial fevers. Long-lasting insecticide treated bed nets stand out as being particularly effective — being responsible for around 68% of the malaria reduction. This means that bednets have prevented around 450 million cases of malaria! And globally, 6.2 million fewer people died of malaria over the last 15 years because of malaria interventions…

> Malaria is an immense economic burden on health systems and people: since 2000, malaria treatment in sub-Saharan Africa has cost almost $300 million. We can see the flip side of this in places in the United States, Brazil and Uganda — children born after malaria was eradicated or heavily controlled and so were not exposed to malaria during pregnancy, had substantially higher income later in life.
>
> We urgently need to keep funding the distribution of bednets, because bednet distributions are one of the most cost-effective ways of preventing disease and death. Since 2000, one billion bednets have been distributed (costing around 5 dollars each), and have averted 450 million cases of malaria — this suggests that, on average, one episode of clinical malaria can be prevented for about $11 (malarial fevers can be very painful). One recent study suggests that in Kenya, bednet distributions between 2003 and 2008 have prevented a death of a child for about $1,011 on average

Humanity seems to be very visibly winning the war against malaria. I just donated a thousand more bednets; now I feel like a part of it and one day I can tell my kids that I helped. If you’re interested, you can donate to Against Malaria Foundation here.

Schizophrenia: No Smoking Gun

[Note: despite how some people are spinning this, tobacco is still really really bad and you should not smoke it]

I.

Schizophrenics smoke. A *lot*. Depending on the study, about 60-80% of schizophrenics smoke, compared to only about 20% of the general population. And they spend on average about 27% (!) of their income on cigarettes. Even allowing that schizophrenics don't make much income, that's a lot of money. Sure, schizophrenics are often poor and undereducated and have other risk factors for smoking – but even after you control for this, the effect is still pretty strong.

Various people have come up with various explanations. Cognitively-minded people say that schizophrenics smoke as a maladaptive coping strategy for the anxiety caused by their condition. Pharmacologically-minded people say that schizophrenics smoke because smoking accelerates the metabolism of antipsychotic drugs and so makes their side effects go away faster. Pragmatically-minded people say that schizophrenics smoke because they're stuck in institutions with nothing to do all day. No points for guessing what the Freudians say.

But all these theories have problems. Sure, schizophrenics are often institutionalized, but even the ones at home smoke a lot. Sure, some schizophrenics are often on antipsychotics, but even the ones who aren't on meds smoke a lot. Sure, schizophrenics are anxious, but we don't see people with Generalized Anxiety Disorder having 80% smoking rates.

As usual, I’m more biologically-minded, so I find it interesting that some of the genes that most commonly turn up as linked to schizophrenia – especially CHRNA3, CHRNA5, and CHRNA7 – are in nicotine receptors. Indeed, some of them are *also* the genes identified as risk factors for smoking. Further, there’s a lot of evidence that schizophrenic people actually feel better and have fewer symptoms when they’re smoking. Further, schizophrenics tend to gravitate toward cigarettes with higher nicotine content, and smoke them in ways that maximize nicotine absorption.

It seems like part of the problem with schizophrenia is that the brain’s nicotine system isn’t working well. Smoking supplements nicotine and makes the system run smoother, so schizophrenics feel better when they smoke and continue to do so. This is the widely accepted self-medication hypothesis.

I like this because it’s a really elegant example of…I don’t know what you’d call it…memetic evolution? Nobody knew that nicotine helped schizophrenia, nobody told the schizophrenics that, but they sort of naturally gravitated to an effective treatment for their condition by going in the direction of things that make them feel better, even going so far as to unknowingly gravitate toward cigarette brands with more nicotine. They did all of this before psychiatry had any idea *why* they were doing it, and in the face of constant protests that it was stupid and useless. This should be a warning to anyone who’s too quick to tell patients that their coping strategies are maladaptive.

But there’s a much more important question here: does smoking cause schizophrenia? How about *prevent* it?

II.

First, the causation argument. Gurillo et al do a meta-analysis and conclude that “daily tobacco use is associated with increased risk of psychosis and an earlier age of onset of psychotic illness. The possibility of a causal link between tobacco use and psychosis merits further examination”. That is, schizophrenics are already smoking much more *at the moment their schizophrenia starts*. This suggests that maybe smoking is helping to cause the schizophrenia?

All nice and well, except for a few things. First, this study ignores the possibility that the genes that cause schizophrenia might also cause increased smoking, even though we have some evidence that this is true (actually, it doesn’t ignore this, it mentions it, but uses it as a reason why a schizophrenia-smoking link is *more* plausible). Second, we know that people who will later develop schizophrenia are seen as kind of odd even before they come down with the disease, and it’s possible that they’re already in some unusual brain state that smoking helps relieve. Third, this study is not controlled – meaning that we’re totally helpless before factors like “people destined to later develop schizophrenia are often poor, and poor people smoke more”.

And fourth, another study shows exactly the opposite.

Zammit et al (thanks to @allfeelsallthetime for the tip) looks at 50,000 teenage Swedish conscripts, then follows them throughout their lives to see which ones do or don’t get schizophrenia. They find that without adjusting for confounders, smokers are more likely to get schizophrenia. But when you *do* adjust for confounders, smokers are *less* likely to get schizophrenia, (hazard ratio 0.8, p = 0.003) and heavy smokers are *much* less likely to get schizophrenia (hazard ratio 0.5)! A dose-dependent relationship was found between smoking and protection from schizophrenia. This is *really* interesting.

Why do we find such different results from these two studies? The only explanation I can think of is that the second study controls for various factors including cannabis use, personality variables, IQ, past psychiatric diagnoses, and place of upbringing (thanks @su3su2u1 for the tip) and the first study controls for zilch. In fact, we find that the second study's uncontrolled numbers are not that different from the first study's uncontrolled numbers, and that the only difference is that the second study then went on to control for confounders and get the opposite result. Controlling for more things is not always better, but controlling for a few things that previous studies and common sense suggest are very relevant is pretty superior to just leaving the data entirely unprocessed. Advantage very much second study.

III.

Unlike certain people on Facebook, I fucking hate science. Let me explain why.

The first study here, Gurillo et al, was published ten years after the second study. Since it is a meta-analysis, it included the second study in it. The authors of the first study *definitely* read the second study. They just didn't care.

Nowhere in the first study does it say "By the way, we read this other study that got the opposite results from us, let's try to figure out why, oh, it was because they controlled for things and we didn't, maybe that should call our findings into question." You know what they did do? They listed the second study as finding that smoking *increased* schizophrenia risk, because the rules of their meta-analysis said they would only take uncontrolled data, and so they did. You can read this entire study, which cites the second study no fewer than six times, without hearing *at all* about the fact that the second study got the opposite result using likely better methodology.

Then they go on to conclude that:

> Cigarette smoking might be a hitherto neglected modifiable risk factor for psychosis, but confounding and reverse causality are possible. Notwithstanding, in view of the clear benefits of smoking cessation programs in this population, every effort should be made to implement change in smoking habits in this group of patients.

Clear benefits! Every effort! Aaaaaaah!

I mean, I know where they (and the *Lancet* editors, who write a glowing comment backing them up) are coming from. Smoking is bad because lung cancer, COPD, etc. But now we have these things called e-cigarettes! They deliver nicotine without tobacco! As far as anyone knows they carry vastly less risk of cancer, COPD, etc. If nicotine *actually* prevents schizophrenia rather than causing it, that is the sort of thing we should *really want to know*. And instead we’re just getting this “We should make schizophrenia patients stop smoking, because smoking is bad”.

Look. I am not going to come out and say that there’s great evidence that nicotine decreases schizophrenia risk. There’s one study, which other studies contradict. I happen to think that the one study looks better than its competitors, but that’s my opinion and I have nowhere near the evidence I would need to feel really strongly about this. But I feel like we are very far from the point where we know enough to be pushing people at risk of schizophrenia *away from* nicotine, and light-years away from the point where we can use phrases like “clear benefits”.

Possibly I am an idiot and missing something very important. But if this is true, I wish the authors of the new study, and the editors of *The Lancet*, would have acknowledged the existence of the conflicting study and patiently explained to their readership, many of whom are idiots like myself, “Here’s a study that looks better than ours that seems to contradict our results, but here’s why our study is nevertheless far more believable.” That’s all I ask.

No matter how much of an idiot I am, I can’t possibly imagine how that wouldn’t be a straight-out gain.

PS: Cigarette smoking definitely decreases your risk of Parkinson’s Disease. Parkinson’s is similar to schizophrenia in that both involve dopamine. But schizophrenia involves too much dopamine and Parkinson’s too little, so the analogy could go either direction.

PPS: Tobacco smoking is definitely still bad! Nothing in here at all suggests that tobacco smoking has the slightest chance of not being a terrible decision!

Lies, Damned Lies, And The Media (Part 6 of ∞)

[content warning: discussion of violent crime, including sexual assault]

I plan to write an article on misuse of statistics by online news organizations, but looking back through my archives most of the examples I've got are from a couple of liberal sites that aren't the worst offenders so much as the only ones I can even bear to read. I'm worried that some of my readers have gotten the impression that liberal sites are the only ones that routinely misuse statistics, which would be grossly false. So before I write the article, I thought I'd give one example of how a lot of conservative sites have statistics that are so bad they're not even *fun* to dissect.

I chose Breitbart's “Rape Deniers: 9 Facts About Illegal Alien Crime The Media Covers Up” because it sounded promising – talking about rape and calling people who disagree with you “deniers” are two pretty reliable red flags that an article will be terrible. It's a series of 9 facts meant to show that illegal immigrants to the US are involved in a lot of crime, especially sexual crime.

I'm skipping Fact 1, which is just a methodological point I don't dispute, Fact 2, which just says some native-born Americans are unemployed, and Fact 3, which says that a lot of our heroin comes from Latin America; I don't really disagree with any of these. I'm also skipping Fact 7 because it's a repeat of Fact 6 and the same points apply. That leaves Facts 4, 5, 6, 8, and 9.

> 4. Because local and state prisons don't track legal status, we don't know how many illegals are in those prisons. As my colleague Ben Shapiro points out, the

> lack of this number is being used dishonestly by the media against Trump…Trump’s repeated statements about immigrants and crime underscore a common public perception that crime is correlated with immigration, especially illegal immigration. But that is a misperception; no solid data support it, and the data that do exist negate it. Trump can defend himself all he wants, but the facts just are not there. Except the facts are there. The Feds do track legal status, and the numbers are startling. Of 78,022 primary offense cases in fiscal year 2013, 38.6 percent were illegal immigrant offenders. The majority of their cases (76 percent) were immigration related. Of total primary offenses, 17.6 percent of drug trafficking offenses and 3.8 percent of sex abuse were illegal immigrants. Of 22,878 drug crime cases, 17.2 percent were illegal immigrants.

The first part of this is a really weird complaint. They’re saying that 38.6% of federal prisoners are illegal immigrants, which is true and indeed very high. Then they’re admitting that 76% of their cases are immigration related. That is, the Feds are imprisoning them because they immigrated illegally. I think that most people would be willing to concede that illegal immigrants are more likely to have immigrated illegally than other populations.

The second part, the part about federal drug trafficking, is complicated; it’s different from “drug having”, “drug dealing”, and even “drug trafficking” as a broader non-federal category. To get a federal drug trafficking arrest, you have to move really large quantities of drugs “across state or national borders”, preferably in a “High Intensity Drug Trafficking Area”. That is, crimes that would give you state charges in a normal place become federal charges in one of these areas. Where are they? The entire US border with Mexico is a gigantic High Intensity Drug Trafficking Area (see this map). So a high rate of illegal immigrants among federal drug trafficking prisoners just means that they're more likely to be involving in transporting drugs across the US-Mexico border than, say, a lumberjack in Wisconsin is. I am prepared to believe this.

Finally, illegal immigrants do commit 3.8% of federal sexual abuse cases. I give Breitbart credit for finally getting a number that is entirely correct and not biased at all. Unfortunately for them, illegal immigrants are 3.8% of the US population.

> 5. According to the Justice Department…There are 94 federal court districts in this country and the five located near the southern border see a large portion of criminal cases, according to the Justice Department's annual report on criminal prosecutions. The five federal districts also have the biggest number of defendants actually convicted of federal crimes. Of the 61,529 criminal cases initiated by federal prosecutors last fiscal year, more than 40%—or 24,746—were filed in court districts neighboring the Mexican border.…Nearly 22% (13,383) were drug related, 19.7% (12,123) were violent crimes and 10.2% (6,300) involved white-collar offenses that include a full range of frauds committed by business and government professionals. Read those stats closely because the media will lie and claim the crimes involve border enforcement. As you can see, over 40% involve drugs and violence.

So here's some interesting math. "More than 40% – or 24,746 – were filed in court districts neighboring the Mexican border." Ellipsis. "13,383 were drug-related, 12,123 were violent crimes, and 6,300 were white-collar fraud." Wait a second. Those three numbers add up to 31,806, more than the total number of cases. Apparently we can't trust illegal immigrants to obey *any* laws, including the laws of mathematics.

Breitbart's reference goes here and their reference is here. When I look it up, these turn out to be the numbers across the entire US, not the numbers for southern border regions. This makes sense – do we really think illegal immigrants commit 6,300 cases of white-collar fraud per year in their dynamic illegal immigrant megacorporations?

But this is really embarrassing for Breitbart's case. Their whole point is that the disproportionate number of crimes committed in immigrant-heavy areas are not just immigration offenses, but in fact representative of the country's crime load as a whole. But that's only because they're accidentally looking at the country's crime load as a whole instead of at crimes committed in immigrant-heavy areas! I can't actually find the immigrant-heavy-area data, but I'd be willing to bet that the disproportionate number of federal crimes along the border do *in fact* involve border enforcement, exactly the argument they claim to debunk.

> 6. There are more than 2,000 sex offenders deported by ICE every year in Texas alone…of the 862 alien sex offenders deported by the Texas-based offices, about 27 percent were convicted of sex offenses against children?

Quick, how many illegal aliens in Texas? If it's 10 million, then their sexual offense rate is far lower than that of any other population. If it's 10,000, their sexual assault rate is far higher than that of any other population. Not only does Breitbart not give these numbers, but they don't even seem to understand that they *should* give them, or expect their readers to care.

I can’t actually figure this number out because it depends on knowing what percent of immigrant sex offenders are deported each year. Consider: there are 86000 sex offenders in Texas. About 6% of Texans are illegal immigrants, so by chance we should expect about 5000 sex offenders total. Given that more than 2000 are deported per year, and that this has been happening for more than three years, that sounds like there are (or were) more than 5000, unless some deportees came back, which we know some do. But I don’t know that the sex offender registry is measuring the same kind of sex-offenderness as the illegal immigration numbers, so I can’t be sure of this. Also, the Texas statistics for immigrants include Oklahoma and possible other areas, so it’s hard to directly compare.

According to this page, there are 5017 arrests for illegal aliens for sexual assault over 4.5 years, so about 1100/year. That should be the sort of number we can work with. But it’s well above the *total* number of Hispanics arrested for sexual assault given here, immigrant and native-born alike, which doesn’t make sense. So I don’t know what to do.

The only good source I can find for percent violent crime by illegal immigrants in Texas is this one, which says that they commit 7.5% of murders. But they’re 6% of the population, so that’s pretty much what we’d expect. This accords with the numbers mentioned above, where in federal prisons the percent illegal immigrants serving time for sexual assault was proportionate to their percent of the population.

I can’t find good numbers here. A very rough inference from one source of Texas sex offender numbers would suggest that the number of sexual offense deportations is unusually high, but inferring based on the murder and sexual abuse numbers suggests that sexual offenses are about average. I’m not sure which method is more correct – but in any case, whatever the truth is there’s no way Breitbart’s numbers could be expected to get anyone any closer to it.

> 8. In 2013 the Obama administration released 36,007 criminal immigrants who had nearly 88,000 convictions between them. Those convictions included 193 homicide convictions, 426 sexual assault convictions, 303 kidnapping convictions, and 1,075 aggravated assault convictions In January, the DHS admitted to Sen. Grassley that 1,000 of the 36,007 released had gone on to commit more crime including: terroristic threats, lewd acts with a minor, various types of assault, DUI, robbery, hit-and-run, gang activity, rape, and child cruelty.

Politifact rates this half true. All immigrants involved served full prison sentences appropriate for their crime; the concern isn’t that they didn’t serve their time but that after their time was up they didn’t get deported. The DHS says that their policy is that if they haven’t finished a deportation case by the time a criminal gets out of prison, the criminal may be released until the deportation case is finished, mostly because the prison system won’t keep them and the DHS doesn’t have enough immigration-related prisons of its own. These people are subjected to the usual monitoring and may be rearrested and deported after their deportation case comes through. There is some reason for concern in that about 3/4s of these people manage to lose themselves before the deportation case is complete, but the DHS reasonably says that if people want this to stop happening people should give them more funding.

And once again, without more information we can’t tell whether any of this involves a higher crime rate than any other population.

> 9. ICE is finding and removing more criminal aliens each year. The number ordered removed has gone up from 7,000 in 2007 to 79,000 in 2010. These criminals are not being stopped at the border. These criminals are being deported after making it across the border and committing tens of thousands of crimes.

Sheesh. You get angry when we don't deport immigrants. You get angry when we do deport immigrants? Make up your minds! This is literally just complaining that we're getting better at solving the problem you complained about above!

There are between ten and twenty million illegal immigrants in the United States – about equal to the number of New Yorkers. If somebody wanted to expel New York from the country, they could point out that New Yorkers commit 616 homicides, 2,534 rapes, and 45,206 cases of aggravated assaults per year. Or that we need seventy-one prisons just to contain all the New Yorker criminals in our justice system, and we don't have nearly enough funding to run all of them effectively, such that literally *thousands* of New Yorker criminals, including murderers and rapists, are released into the general population every year. You could report that [number] of New Yorkers commit violent sexual assaults on children each year – I don't know what the number is, but I guarantee you that it is a number, that it has a certain number of digits, and that it is worse than zero.

Breitbart consistently fails to give numbers that would mean anything or inform anybody, and when it either uses numbers that are loaded in its favor or ones that don't mean what it thinks they mean. The actual numbers on the question of interest don't seem particularly outrageous.

This isn't to say that there can't be legitimate concerns about illegal immigration. In fact, that's my whole point and something that I wish conservatives better understood: we need to practice Gettier politics. There’s a theory on the Right that since the media has created a giant edifice of lies to justify liberalism, liberalism must be false. But other parts of the media have created a giant edifice of lies to justify *conservativism*. Instead of assuming our opponents are necessarily gullible morons who believe the giant edifice of lies on their side, we should kind of awkwardly go “Oh, there’s a giant edifice of lies on *your* side too? Yeah, I know that feeling,” and listen to what they have to say.

> A major obstacle to changing one’s mind is the fallacy that if the other side is right, it must be decent and intellectually honest.
>
> — Steven Kaas (@stevenkaas) August 21, 2010

Side Effects May Include Anything

A couple of days ago a patient said he'd become depressed after starting Xolair, a new asthma drug I know nothing about.

On the one hand, lots of things that mess with the immune system can cause depression. On the other, patients are notorious for blaming drugs for any random thing that happens around the same time they started taking them. So I did what any highly-trained competent medical professional would: I typed "does xolair cause depression?" into Google.

The results seemed promising. The first site was called "Can Xolair cause depression?". The second was "Is depression a side effect of Xolair?". Also on the front page were "Could Xolair cause major depression?" and "Xolair depression side effects". Clearly this is a well-researched topic that lots of people cared about, right?

Let's look closer at one of those sites, EHealthMe.com. It says: "Major depression is found among people who take Xolair, especially for people who are female, 40-49 old, also take medication Singulair, and have Asthma. We study 11,502 people who have side effects while taking Xolair from FDA and social media. Among them, 14 have Major depression. Find out below who they are, when they have Major depression and more." Then it offers a link: "Join a support group for people who take Xolair and have Major depression".

First things first: if there were actually 11502 people taking Xolair, and only 14 of them had major depression, that would be a rate of 0.1%, compared to 6.9% in the general population. In other words, Xolair would be the most effective antidepressant on Earth. But of course nobody has ever done an n=11502 study on whether a random asthma medication causes depression, and EHealthMe is just scraping the FDA databases to see how many people reported depression as a side effect to the FDA. But only a tiny percent of people who get depression report it, and depression sometimes strikes at random times whether you're taking Xolair or not. So this tells us nothing.

And yet a patient who worries that Xolair might be causing their depression will Google "can xolair cause depression?", and she will end up on this site that says "major depression is found among people who take Xolair", which is one of the worst examples of weasel words I've ever heard. Then she will read that there are *entire support groups* for depressed Xolair sufferers. She will find all sorts of scary-looking information like that Xolair-related depression has been increasing since 2008. And this is above and beyond just the implications of somebody bothering to write an entire report about the Xolair-depression connection!

In case you haven't guessed the twist – no one's ever investigated whether Xolair causes depression. EHealthMe's business model is to make an automated program that runs through every single drug and every possible side effect, scrapes the FDA database for examples, then autopublishes an ad-filled web page titled "COULD $DRUG CAUSE $SIDE_EFFECT?". It populates the page by spewing random FDA data all over it, concludes "$SIDE_EFFECT is found among people who take $DRUG", and offers a link to a support group for $DRUG patients suffering from $SIDE_EFFECT. Needless to say, the support group is an automatically-generated forum with no posts in it.

And it’s not just EHealthMe. This is a whole market, with competitors elbowing their way past one another to the top of the Google search results. Somebody who doubts EHealthMe and seeks an online second opinion will probably just end up at PatientsVille, whose page is called “Xolair Depression Side Effects”, which contains the same FDA data, and which gets the Google description text “This opens a possibility that Xolair could cause Depression”. Or Treato, whose page claims to contain 56 reader comments on Xolair and depression, but which has actually just searched the Web for every single paragraph that contains “Xolair” and “depression” together and then posted garbled excerpts in its comment section. For example, one of their comments – and this is not at all clear from Treato’s garbled excerpt – is from a tennis forum, where a user with the handle Xolair talks about how his tennis serve is getting worse with age; another user replies “Xolair, I read this and get depressed, I just turned 49.” But if you don’t check whether it came from a tennis forum or not, 56 reports of a connection between a drug and a side effect sounds convincing!

This is *really scummy*. Maybe it’s not the most devious of traps for you or me, but what about for your grandmother? What about for those people who send money to Nigerian princes? The law is usually pretty strict about who can and can’t provide medical information – so much so that it cracks down on 23andMe just for reading off the genome in a way that uneducated people might misinterpret. Yet somehow sites like EHealthMe are allowed to continue, because they just *very strongly imply* fake medical information instead of saying it outright.

Remember, only about 50% of people who are prescribed medication take it. Sometimes it’s personal choice or simple forgetfulness. But a lot of the time they stop because of side effects. I had a patient a few months ago who was really depressed. I started her on an antidepressant and she got much better. Then she stopped the medication cold turkey and got a lot worse again. I asked her why she’d stopped. She said her shoulder started hurting, she’d Googled whether antidepressants could cause shoulder pain, and read that they could. She couldn’t remember what site she was reading, but I bet it was EHealthMe or Treato or some of the others just like them.

One day, somebody’s going to Google “can penicillin cause cancer?”, read a report with a link to a support group for penicillin-induced-cancer survivors, stop taking antibiotics, and die. And when that happens, I hope it’s in America, so I can be sure their family will sue the company involved for more money than exists in the entire world.

Predictions For 2016

At the beginning of every year, I make predictions. At the end of every year, I score them. So here are a hundred more for 2016. Most of them are objectively decideable, but a few are subjective (eg “X goes well”) and are marked with asterisks.

WORLD EVENTS
1. US will not get involved in any new major war with death toll of > 100 US soldiers: 60%
2. North Korea’s government will survive the year without large civil war/revolt: 95%
3. Greece will not announce it’s leaving the Euro: 95%
4. No terrorist attack in the USA will kill > 100 people: 90%
5. …in any First World country: 80%
6. Assad will remain President of Syria: 60% *[edit: called out as dumb, but I won’t cheat and change it]*
7. Israel will not get in a large-scale war (ie >100 Israeli deaths) with any Arab state: 90%
8. No major intifada in Israel this year (ie > 250 Israeli deaths, but not in Cast Lead style war): 80%
9* No interesting progress with Gaza or peace negotiations in general this year: 90%
10. No Cast Lead style bombing/invasion of Gaza this year: 90%
11* Situation in Israel looks more worse than better: 70%
12. Syria’s civil war will not end this year: 70%
13. ISIS will control less territory than it does right now: 90%
14. ISIS will not continue to exist as a state entity [added: meant in Iraq/Syria]: 60% *[edit: called out as dumb, but I won’t cheat and change it]*
15. No major civil war in Middle Eastern country not currently experiencing a major civil war: 90%
16* Libya to remain a mess: 80%

17. Ukraine will neither break into all-out war or get neatly resolved: 80%
18. No country currently in Euro or EU announces plan to leave: 90%
19. No agreement reached on “two-speed EU”: 80%
20. Hillary Clinton will win the Democratic nomination: 95%
21. Donald Trump will win the Republican nomination: 60%
22* Conditional on Trump winning the Republican nomination, he impresses everyone how quickly he pivots towards wider acceptability: 70%
23. Conditional on Trump winning the Republican nomination, he’ll lose the general election: 80%
24. Conditional on Trump winning the Republican nomination, he’ll lose the general election worse than either McCain or Romney: 70%
25. Marco Rubio will not win the Republican nomination: 60% *[edit: called out as dumb, but I won’t cheat and change it]*
26. Bloomberg will not run for President: 80%
27. Hillary Clinton will win the Presidency: 60%
28. Republicans will keep the House: 95%
29. Republicans will keep the Senate: 70%
30. Bitcoin will end the year higher than $500: 80%
31. Oil will end the year lower than $40 a barrel: 60%
32. Dow Jones will not fall > 10% this year: 70%
33. Shanghai index will not fall > 10% this year: 60% *[edit: called out as dumb, but I won’t cheat and change it]*
34. No major revolt (greater than or equal to Tiananmen Square) against Chinese Communist Party: 95%
35. No major war in Asia (with >100 Chinese, Japanese, South Korean, and American deaths combined) over tiny stupid islands: 99%
36. No exchange of fire over tiny stupid islands: 90%
37. US GDP growth lower than in 2015: 60%
38. US unemployment to be lower at end of year than beginning: 50%

39. No announcement of genetically engineered human baby or credible plan for such: 90%
40* No major change in how the media treats social justice issues from 2015: 70%
41* European far right makes modest but not spectacular gains: 80%
42* Mainstream European position at year's end is taking migrants was bad idea: 60%
43. [Duplicate removed]
44* So-called "Ferguson effect" continues and becomes harder to deny: 70%
45. SpaceX successfully launches a reused rocket: 50%
46* Nobody important changes their mind much about the EMDrive based on any information found in 2016: 80%
47. California's drought not officially declared over: 50%
48. No major earthquake (>100 deaths) in US: 99%
49. No major earthquake (>10000 deaths) in the world: 60%
50. Occupation of Oregon ranger station ends: 99%

PERSONAL/COMMUNITY

1. SSC will remain active: 95%
2. SSC will get fewer hits than in 2015: 60%
3. At least one SSC post > 100,000 hits: 50%
4. UNSONG will get fewer hits than SSC in 2016: 90%
5. > 10 new permabans from SSC this year: 70%
5. UNSONG will get > 1,000,000 hits: 50%
6. UNSONG will not miss any updates: 50%
7. UNSONG will have higher Google Trends volume than HPMOR at the end of this year: 60%
8. UNSONG Reddit will not have higher average user activity than HPMOR Reddit at the end of this year: 60%
9. Shireroth will remain active: 70%
10. I will be involved in at least one published/accepted-to-publish research paper by the end of 2016: 50%
11. I won't stop using Twitter, Tumblr, or Facebook: 95%
12. > 10,000 Twitter followers by end of this year: 50%

13. I will not break up with any of my current girlfriends: 70%
14. I will not get any new girlfriends: 50%
15. I will attend at least one Solstice next year: 90%
16. …at least two Solstices: 70%
17. I will finish a long blog post review of stereotype threat this year: 60%
18* Conditional on finishing it, it won't significantly change my position: 90%
19. I will finish a long FAQ this year: 60%
20. I will not have a post-residency job all lined up by the end of this year: 80%
21. I will have finished all the relevant parts of my California medical license application by the end of this year: 70%
22. I will no longer be living in my current house at the end of this year: 70%
23. I will still be at my current job: 95%
24. I will still not have gotten my elective surgery: 80%
25. I will not have been hospitalized (excluding ER) for any other reason: 95%
26. I will not have taken any international vacations with my family: 70%
27. I will not be taking any nootropic daily or near-daily during any 2-month period this year: 90%
28. I will complete an LW/SSC survey: 80%
29. I will complete a new nootropics survey: 80%
30. I will score 95th percentile or above in next year's PRITE: 50%
31. I will not be Chief Resident next year: 60%
32. I will not have any inpatient rotations: 50%
33. I will continue doing outpatient at the current clinic: 90%
34* I will not have major car problems: 60%
35* I won't publicly and drastically change highest-level political/religious/philosophical positions (eg become a Muslim or Republican): 90%
36. I will not vote in the 2016 primary: 70%

37. I will vote in the 2016 general election: 60%
38. Conditional on me voting and Hillary being on the ballot, I will vote for Hillary: 90%
39* I will not significantly change my mind about psychodynamic or cognitive-behavioral therapy: 80%
40. I will not attend the APA meeting this year: 80%
41. I will not do any illegal drugs (besides gray-area nootropics) this year: 90%
42. I will not get drunk this year: 80%
43* Less Wrong will neither have shut down entirely nor undergone any successful renaissance/pivot by the end of this year: 60%
44. No co-bloggers (with more than 5 posts) on SSC by the end of this year: 80%
45. I get at least one article published on a major site like Huffington Post or Vox or New Statesman or something: 50%
46. I still plan to move to California when I'm done with residency: 90%
47. I don't manage to make it to my friend's wedding in Ireland: 60%
48. I don't attend any weddings this year: 50%
49. I decide to buy the car I am currently leasing: 60%
50. Except for the money I spend buying the car, I make my savings goal before July 2016: 90%

Other people doing yearly predictions with probability: Against Jebel al-Lawz, Anatoly Karlin, Old Lamps, Garrett Peterson. If you're doing this and I missed it, let me know and I'll add you in.

Staying Classy

Siderea writes an essay on class in America. You should read it. In case you don't, here's the summary:

1. People tend to confuse social class with economic class, eg how much money you make. But social class is a more complicated idea involving how respectable you seem, how educated you are, and what kind of family you come from. An assembly line supervisor might make the same amount of money as a schoolteacher, but the schoolteacher would probably seem more refined and be able to access better social circles.

2. Classes are cultures. People in a certain class have their own way of dressing, speaking, decorating, and behaving. They have distinctive ideas and values. This is why a lower-class person cannot simply claim to be upper-class and so gain all the benefits of upper-class-hood; it would be as hard as trying to pass for Japanese. Lower-class people *can* learn their way around upper-class culture, but it's a difficult and lifelong project done most easily if you *already* have upper-class resources.

3. Talking about class is taboo because we like to believe we're a classless society. We talk about income instead and pretend it's class. Class breaks through in a couple of phrases like "rednecks" or "white trash" or "white collar" or "coastal elites", but people use the phrases without usually having a broader idea that it's class they're talking about.

4. Class prejudice is complicated. It combines the practical superiority of being upper-class to being lower-class (because you have more money and opportunity) with the very dubious value judgment that upper-class culture is superior to lower-class culture, or that lower-class culture is just people trying to do upper-class culture but failing. But lower-class people like lower-class culture and generally do not want to adopt upper-class culture, except insofar as it’s necessary to advance. Analogies to race and assimilation are obvious.

5. People mostly understand their own class, and the class one step above or below them, but have only vague stereotypes of classes further than that. This limits social mobility; you can’t join what you can’t understand.

6. College is a finishing school for the upper classes. They send their children there to learn the proper upper class values and behaviors. Even if community college does a great job teaching whatever trades it teaches, it will not teach you how to be a part of the upper class, and this will seriously limit your opportunities.

7. Politically, the left pretends class doesn’t exist; the right talks about it, but only to yell at the underclass and say that their culture is wrong. Race is really complicated and will be left out of this analysis.

I notice Siderea is a psychotherapist, which doesn’t surprise me. We in mental health get a pretty good cross-sectional exposure of everybody and get to hear about their lives, and with enough data points the structure comes into sharper relief.

Just to give an example: suppose a lady comes in with really over-permed dyed curly hair wearing several rings, bracelets, and necklaces. Her name is Sherri and she calls you “darling”; she’s also carrying her lunch, which is KFC plus a Big Gulp. Without knowing anything else about her, you can peg her as working class. Maybe she won the lottery ten years ago and is now the richest person in your state. It doesn't matter. She's still working class.

Or suppose a thin 25-year-old man comes in wearing glasses, a small close-cropped beard, and a Led Zeppelin t-shirt. His name is Alex and he apologizes for being three minutes late. This guy is probably middle-to-upper-middle-class and college educated, maybe not a *great* college but still college-educated. And maybe he's fallen on hard times and doesn't have a dollar to his name. It still doesn't matter. He's still middle-to-upper-middle class.

And you start to learn you can predict things about these people, the concerns they're going to have, the kind of things that happen to them. Who their friends are. How they relate to their friends: Sherri will expound upon the flaws of every single one of her ungrateful coworkers; Alex will reluctantly say he went through a tough breakup a year or two ago. What kind of drugs they abuse, if they abuse drugs (maybe Sherri has smoking and drinking problems; Alex has probably tried marijuana and LSD but is embarrassed to say so).

But this kind of innate stereotyping is different than a formal taxonomy. Siderea links to Michael Church's attempt to explain what the classes actually are. This is another piece you should read, but again in case you don't:

1. 10% of people are in an underclass consisting of "generationally poor" people who may never have held jobs and who come from similarly poor families.

2. 65% of people are in the labor class. They work jobs where labor is seen as a commodity, ie there's not as much sense of career capital or reputation. They base virtue and success around Hard Work. Its lower levels are minimum wage McJobs, its middle levels are assembly line work, and its higher levels are things like pilots, plumbers, and small business owners. The stratospheric semi-divine level is “celebrities” like reality TV stars who become fabulously rich and famous while sticking to their labor class roots.

3. 23.5% of people are in the gentry class. They fetishize education and career capital. They engage in all sorts of signaling games around “fair trade” and “organic” and what museums they go to. At the lower level they’re schoolteachers and starving artists, at the mid level they’re “professions” like engineering and law, and at the highest level they’re professors and scientists and entrepreneurs. The stratospheric semi-divine level is “cultural influencers” like Jon Stewart or Steven Pinker who become famous and (maybe) rich while sticking to their gentry class roots.

4. 1.5% of people are in the elite class. Although you can be borderline-elite by getting a job in finance and making a few million, the real elite are born into money and don’t work unless they want to. Occasionally they’ll sit on a board or found a philanthropic association or something. They don’t believe in “professional achievement” because working is lower-class; they might compete in complicated status games around who throws the best parties or has the best horses or whatever.

5. The highest class (E1) are psychopaths who burn the global commons for shits and giggles. They tend to be drug lords, arms dealers, and morally insane billionaires. Most famous politicians and businesspeople are not in this class and most people in this class are not famous.

6. The three main classes (labor, gentry, and elite) are three different ‘infrastructures’. To be in labor you need skills, to be in gentry you need education, and to be in elite you need connections. There’s no strict hierarchy (eg not all gentry are above all labor), but you can picture them as offset ladders, with the lower gentry being at the same rung as the higher labor and so on.

7. The Elite control everything; the constant threat is that Gentry and Labor will unite against them, which might very well work. The Elite neutralize this threat by making Labor hate Gentry as “effeminate” or “pretentious”; they also convince Labor that the Gentry are probably secretly in cahoots with the underclass against Labor. Elites also convince Labor that Elites don’t exist and it’s Gentry all the way up, which means that “anti-1%” sentiment, which should properly get Labor and Gentry to cooperate against the Elites, instead makes Gentry hate the Elites but Labor hate Gentry. Politics boils down to Gentry being good people trying to improve things, and Elite conning Labor into hating Gentry to prevent things from being improved.

8. While all classes can have good and bad people (except E1, which is wholly bad), Elites have a generally negative influence on society, and Gentry are generally positive. After the World Wars, everybody got angry at the Elites for all the war and killing and stuff, which convinced them to lie low for a few decades and forced the Gentry to take over. This was why the country did so well during the 50s and 60s. Whether the country goes in a good or bad direction now depends on whether the Elites manage to take it back or not. One reason Silicon Valley works (used to work?) so unusually well was that it was mostly a native project of the Gentry that hadn’t yet been infiltrated by the Elites.

Reaction to Church on the subreddit was pretty negative, but I find it at least a good nucleus for further discussion. The Gentry/Labor distinction is glaringly obvious. The Labor/Underclass distinction also seems glaringly obvious to me, if only because Labor hates the underclass. The Gentry/Elite distinction *doesn't* seem glaringly obvious to me, but maybe that's just because I haven't met enough elites. In particular, Church's "E1" seems caricatured and out-of-place in his otherwise sober analysis. Then again, if those people existed I probably wouldn't know anyway. Then again, the rest of Church's blog suggests some paranoid tendencies, so maybe the E1 entry is just those coming out.

Siderea notes that Church's analysis independently reached about the same conclusion as Paul Fussell's famous guide. I'm not entirely sure how you'd judge this (everybody's going to include lower, middle, and upper classes), but eyeballing Fussell it does look a lot like Church, so let's grant this.

It also doesn't sound too different from Marx. Elites sound like capitalists, Gentry like bourgeoisie, Labor like the proletariat, and the Underclass like the lumpenproletariat. Or maybe I'm making up patterns where they don't exist; why should the class system of 21st century America be the same as that of 19th century industrial Europe?

There's one more discussion of class I remember being influenced by, and that's Unqualified Reservations' Castes of the United States. Another one that you should read but that I'll summarize in case you don't:

1. Dalits are the underclass, made up of homeless people, chronically unemployed people, drug addicts, etc. They tend to have a lot of trouble with the law, go in and out of jail, never really hold down stable employment. Status is "street cred" that you get from being powerful, wealthy, and sexually successful, eg gang leaders.

2. Vaisyas are standard middle-class people who engage in productive employment. They tend to form nuclear families and try to go to church. Status is having a stable job, a stable family, and being well-liked in your church or social club.

3. Brahmins are very educated people who participate in the world of ideas. They range from doctors and lawyers to artists and professors. Access is conferred by top-tier university education. Status is from conspicuous engagement in progressive politics, eg being an activist, working for an NGO, “campaigning for justice”. They are “the ruling class”.

4. Optimates are very rich WASPs concerned with breeding and old money. Status comes from breeding and an antiquated idea of “nobility”. Optimates used to be “the ruling class”, but now they’re either extinct or endangered, having been pretty much absorbed into the Brahmins.

5. Mentioned elsewhere in the UR corpus: politics boils down to Vaisyas being basically decent people trying to lead normal productive lives, and Brahmins trying to create a vast tentacled monstrosity of useless bureaucrats and petty enforcers of ideological conformity to employ Brahmins in the “knowledge work” they feel entitled to and to protect their interests. Silicon Valley is (used to be?) unusually functional because it maintained some Vaisya values separate from the corrupting influence of the Brahmins.

Michael Church’s system (henceforth MC) and the Unqualified Reservation system (henceforth UR) are similar in some ways. MC’s Underclass matches Dalits, MC’s Labor matches Vaisyas, MC’s Gentry matches Brahmins, and MC’s Elite matches Optimates. This is a promising start. It’s a fourth independent pair of eyes that’s found the same thing as all the others. (commenters bring up Joel Kotkin and Archdruid Report as similar convergent perspectives).

But there are also some profound differences. UR says that the Elites are mostly gone, that everything's ruled by the Gentry nowadays, and that the Gentry are allying with the criminal Underclass against Labor. MC mentions this same picture, but only as the *false facade* that the Elites are trying to get everyone else to believe in order to keep them divided.

You could reconcile some of the differences by supposing the two models have different cutoffs. Suppose we rank people from 0 (lowest underclass) to 100 (highest elite). Maybe MC draws the Labor/Gentry and Gentry/Elite borders at 40 and 70 respectively, and UR draws the Vaisya/Brahmin and Brahmin/Optimate borders at 60 and 90. If the world's being run by 80s, MC could be right to say it's run by Elites and not Gentry, and UR could be right in saying it's run by Brahmins and not Optimates. If Silicon Valley is run by 55s but being ruined by 75s, MC could say it's run by Gentry but ruined by elites, and UR could say it's run by Vaisyas but ruined by Brahmins. But if there's this much variability in class boundaries, what's the point in even drawing them in the first place?

But I think the differences are real and political: MC comes from a liberal perspective, UR from a conservative one. MC wants to locate the source of the cancer in the (mostly plutocrat) Elites, cast the (mostly liberal) Gentry as wonderful people who can do no wrong, cast the (mostly conservative) Labor as deluded and paranoid, and cast the (liberal-aligned) Underclass in a sympathetic light. UR wants to locate the source of the cancer in the (mostly liberal) Brahmins, cast the (mostly conservative) Labor as decent salt-of-the-Earth types under threat from the elite, and cast the (liberal-aligned) Underclass in an unsympathetic light.

And the political angle evokes one more system worth adding here: my own discussion of the Blue Tribe vs. the Red Tribe in I Can Tolerate Anything But The Outgroup. I point out that the group sometimes referred to as “coastal liberals” or “SWPL” and so on are marked not only by Democratic Party beliefs, but by a host of cultural similarities including food, dress, music, hobbies, religion, values, art, etc. Likewise, the group sometimes referred to as “rednecks” or “fundies” and so on are marked not only by Republican Party beliefs, but by a similar set of cultural similarities. I call these the “Blue Tribe” and the “Red Tribe” as an attempt to distinguish them as cultures and not just as sets of political beliefs.

These tribes seem closely related to classes. “Blue Tribe” is similar to Gentry; “Red Tribe” is similar to Labor. I won’t say there’s a perfect 1:1 equivalence; for example, I know some union leaders who are very clearly in the Labor class but who wouldn’t be caught dead in the Red Tribe. But the resemblance is too close to miss.

Some final scattered thoughts:

1. All those studies that analyze whether some variable or other affects income? They’d all be much more interesting if they analyzed the effect on class instead. For example, there’s a surprisingly low correlation between your parents’ income and your own income, which sounds like it means there’s high social mobility. But I grew up in a Gentry class family; I became a doctor, my brother became a musician, and my cousin got a law degree but eventually decided to work very irregularly and mostly stay home raising her children. I make more money than my brother, and we both make more money than my cousin, but this is not a victory for social mobility and family non-determinism; it's no coincidence none of us ended up as farmers or factory workers. We all ended up Gentry class, but I chose something closer to the maximize-income part of the Gentry class tradeoff space, my brother chose something closer to the maximize-creativity part, and my cousin chose to raise the next generation. Any studies that interpret our income difference as an outcome difference and tries to analyze what factors gave me a leg up over my relatives (better schools? more breastfeeding as a child?) are stupid and will come up with random noise. We all got approximately the same level of success/opportunity, and those things just happen to be very poorly measured by money. If we could somehow collapse the entirety of tradeoffspace into a single variable, I bet it would have a *far* greater parent-child correlation than income does. This is part of why I don't follow the people who take the modest effect of IQ on income as a sign that IQ doesn't change your opportunities much; maybe everyone in my family has similar IQs but wildly different income levels, and there's your merely modest IQ/income relationship right there. I think some studies (especially in Britain) have tried analyzing class and gotten some gains over analyzing income, but I don't know much about this.

2. I think Siderea is right that the Right thinks in social class terms more naturally than the Left. To oversimplify, both sides use class warfare, but the Left's class warfare is economic ("the plutocrat billionaires are ruining everything!") and the Right's class warfare is social ("the media and academic elites are ruining everything!").

3. Closely related: Donald Trump appeals to a lot of people because despite his immense wealth he practically glows with signs of being Labor class. This isn't surprising; his grandfather was a barber and his father clawed his way up to the top by getting his hands dirty. He himself went to a medium-tier college and is probably closer in spirit to the small-business owners of the upper Labor class than to the Stanford MBA-holding executives of the Elite. Trump loves and participates in professional wrestling and reality television; those definitely aren't Gentry or Elites pastimes! When liberals shake their heads wondering why Joe Sixpack feels like Trump is a kindred soul even though Trump's been a billionaire his whole life, they're falling into the liberal habit of sorting people by wealth instead of by class. To Joe Sixpack, Trump is "local boy made good".

4. The thesis of "I Can Tolerate Anything But The Outgroup" simplifies to "It is a Gentry-class tradition to sweep aside all prejudices *except* class prejudice, which must be held with the intensity of all the old prejudices combined."

5. But "I Can Tolerate Anything But The Outgroup"'s Grey Tribe sits uneasy within this system. It doesn't seem to be a class. But it also seems distinctly different from ordinary Gentry norms. And what about minorities? What about the differences between farmers vs. factory workers? If different classes are equivalent to different cultures, well, there are a lot of different cultures that don't fit easily into the hierarchy. Maybe class is one factor among many that can create a different culture, but other factors can be stronger than class in some groups?

6. Siderea doesn't want to get into how race interacts with class, and that seems wise. But a related digression: lots of people complain about social justice being classist, in that it's hard for anybody who hasn't either gone to college or at least spent a lot of time hanging around social justice people to keep track of which words, opinions, and causes are okay versus will render you radioactive. On the one hand, this is probably true. On the other, it’s probably true of *everything*, with social justice as an unexceptional example. Yes, the way you refer to trans people shows what class you’re from, but so does the way you order ice cream.

7. Siderea admits she is classist and not ashamed of this. I have a hard time understanding what she means, but I can try to explain my *own* classism: I think classes probably sort on important qualities and reinforce those qualities. For example, the Underclass and Labor class people I know are much more likely to have high-conflict styles of interaction: if they feel offended, they’ll yell at you and maybe even fight you. Gentry class people would be horrified at the thought; they might respond to the same offense by filing a complaint with Human Resources. I think there are two equally correct ways to interpret this. Number one, people with the maladaptive behavior of starting physical fights don’t make it very far in life and so end out in lower classes, and insofar as these behaviors are either genetic or learned within the family, their families stay in lower classes throughout the generations. Number two, the lower classes have a culture where you defend your honor by fighting people who offend you, and the upper classes have a culture where you defend your honor by submitting complaints, and although in a cosmic sense both of those styles are equally valid, and although indeed a thousand years ago the fighting might have been more adaptive, in today’s society the complaint-submitting is more adaptive and the lower classes are screwed unless they unlearn that behavior – which they probably won’t, because unlearning class is hard. But this means that classism is at least kind of justified – if you want to hire for example a schoolteacher, you might want to look for people who show all the signs of Gentry rather than Labor class to make sure they’re not going to get into physical fights in the classroom.

8. Cellular automaton theory of fashion likely relevant.

9. Siderea’s idea of college as finishing school for the upper classes is interesting, and her own experience is a window into something I never thought about before. But I’m not sure how typical she is; I think most colleges admit students who are already members of the classes their graduates end up in. I felt like I didn’t learn any class culture during my own college experience at all – which isn’t surprising since I was born the son of a doctor and ended up as a doctor myself. I think my story’s probably more typical than Siderea’s, though other people can prove me wrong if they’ve seen differently.

Book Review: Superforecasting

Philip Tetlock, author of Superforecasting, got famous by studying prediction. His first major experiment, the Expert Political Judgment experiment, is frequently cited as saying that top pundits' predictions are no more accurate than a chimp throwing darts at a list of possibilities- although Tetlock takes great pains to confess to us that no chimps were actually involved, and this phrasing just sort of popped up as a flashier way of saying "random".

Although this was generally true, he was able to distinguish a small subset of people who were able to do a *little* better than chance. His investigation into the secrets of their very moderate success led to his famous "fox" versus "hedgehog" dichotomy, based on the fable that "the fox knows many things, the hedgehog knows one big thing". Hedgehog pundits/experts are people who operate off a single big idea- for example, an economist who says that government intervention is always bad, predicts doom for any interventionist policy, and predicts great success for any noninterventionist one. Foxes are people who don't have much of a narrative or ideology, but try to find the right perspective to approach each individual problem. Tetlock found that the hedgehogs did worse than the chimp and the foxes did a little better.

Cut to the late 2000s. The US intelligence community has just been seriously embarrassed by their disastrous declaration that there were weapons of mass destruction in Iraq. They set up an Intelligence Advanced Research Projects Agency to try crazy things and see if any of them worked. IARPA approached a bunch of scientists, handed them a list of important world events that might or might not happen, and told them to create some teams and systems for themselves and compete against each other to see who could predict them the best.

Tetlock was one of these scientists, and his entry into the competition was called the Good Judgment Project. The plan was simple: get a bunch of people to sign up and try to predict things, then find the ones who did the best. This worked pretty well. 2,800 people showed up, and a few of them turned out to be…

…okay, now we're getting to a part I don't understand. When I read Tetlock's paper, all he says is that he took the top sixty forecasters, declared them superforecasters, and then studied them intensively. That's fine; I'd love to know what puts someone in the top 2% of forecasters. But it's important not to phrase this as "Philip Tetlock discovered that 2% of people are superforecasters". This suggests a discontinuity, a natural division into two groups. But unless I'm missing something, there's no evidence for this. Two percent of forecasters were in the top two percent. Then Tetlock named them "superforecasters". We can discuss what skills help people make it this high, but we probably shouldn't think of it as a specific phenomenon.

Anyway, the Good Judgment Project then put these superforecasters on teams with other superforecasters, averaged out their decisions, slightly increased the final confidence levels (to represent the fact that it was 60 separate people, all of whom were that confident), and presented that to IARPA as their final answer. Not only did they beat all the other groups in IARPA's challenge in a landslide, but they actually did 30% better than professional CIA analysts working off classified information.

Having established that this is all pretty neat, Tetlock turns to figuring out how superforecasters are so successful.

First of all, is it just luck? After all, if a thousand chimps throw darts at a list of stocks, one of them will hit the next Google, after which we can declare it a “superchimp”. Is that what’s going on here? No. Superforecasters one year tended to remain superforecasters the next. The year-to-year correlation in who was most accurate was 0.65; about 70% of superforecasters in the first year remained superforecasters in the second. This is definitely a real thing.

Are superforecasters just really smart? Well, sort of. The superforecasters whom Tetlock profiles in his book include a Harvard physics PhD who speaks 6 languages, an assistant math professor at Cornell, a retired IBM programmer data wonk, et cetera. But the average superforecaster is only at the 80th percentile for IQ – just under 115. And there are a lot of people who are very smart but not very good at predicting. So while IQ definitely helps, it isn’t the whole story.

Are superforecasters just really well-informed about the world? Again, sort of. The correlation between well-informedness and accuracy was about the same as the correlation between IQ and accuracy. None of them are remarkable for spending every single moment behind a newspaper, and none of them had as much data available as the CIA analysts with access to top secret information. Even when they made decisions based on limited information, they still beat other forecasters. Once again, this definitely helps, but it’s not the whole story.

Are superforecasters just really good at math? Again, kind of. A lot of them are math PhDs or math professors. But they all tend to say that they don’t explicitly use numbers when doing their forecasting. And some of them don’t have any kind of formal math background at all. The correlation between math skills and accuracy was about the same as all the other correlations.

So what *are* they really good at? Tetlock concludes that the number one most important factor to being a superforecaster is really understanding logic and probability.

Part of it is just understanding the basics. Superforecasters are less likely to think in terms of things being 100% certain, and – let's remember just how far left the bell curve stretches – less likely to assign anything they're not sure about a 50-50 probability. They're less likely to believe that things happen because they're fated to happen, or that the good guys always win, or that things that happen will necessarily teach a moral lesson. They're more likely to admit they might be wrong and correct themselves after an error is discovered. They're more likely to debate with themselves, try to challenge their original perception, start asking "What could be wrong about this thing I believe?" rather than "How can I prove I'm right?"

But they're also more comfortable actively using probabilities. Like my predictions, the Good Judgment Project made forecasters give their answers as numerical probability estimates – for example, 15% chance of a war between North and South Korea in the next ten years killing > 1000 people. Poor forecasters tend to make a gut decision based on feelings that superficially related to the question, like "Well, North Korea is pretty crazy, so they're pretty likely to declare war, let's say 90%" or "War is pretty rare these days, how about 10%?". Superforecasters tend to focus on the specific problem in front of them and break it down into pieces. For example, they might start with the Outside View – it's been about 50 years since the Koreas last fought, so their war probability per decade shouldn't be more than about 20% – and then adjust that based on Inside

View information – “North Korea has a lot fewer foreign allies these days, so they’re less likely to start something than they once were – maybe 15%”.

Or they might break the problem down into pieces: “There would have to be some sort of international incident, and then that incident would have to erupt into total war, and then that war would have to kill > 1,000 people. There are about two international incidents between the Koreas every year, but almost none of them end in war; on the other hand, because of all the artillery aimed at Seoul, probably any war that did happen would have an almost 100% chance of killing > 1,000 people” … and so on. One result is that while poor forecasters tend to give their answers in broad strokes – maybe a 75% chance, or 90%, or so on – superforecasters are more fine-grained. They may say something like “82% chance” – and it’s not just pretentious, Tetlock found that when you rounded them off to the nearest 5 (or 10, or whatever) their accuracy actually decreased significantly. That 2% is actually doing good work.

Most interesting, they seem to be partly immune to cognitive bias. The strongest predictor of forecasting ability (okay, fine, not by much, it was pretty much the same as IQ and well-informedness and all that – but it *was* a predictor) was the Cognitive Reflection Test, which includes three questions with answers that are simple, obvious, and wrong. The test seems to measure whether people take a second to step back from their System 1 judgments and analyze them critically. Superforecasters seem especially good at this.

Tetlock cooperated with Daniel Kahneman on an experiment to elicit scope insensitivity in forecasters. Remember, scope insensitivity is where you give a number-independent answer to a numerical question. For example, how much should an organization pay to save the lives of 100 endangered birds? Ask a hundred people, and maybe the average answer is “$10,000”. Ask a (different group of) a hundred people how much the same organization should pay to save the lives of 1000 endangered birds, and maybe the average answer will still be $10,000. So it seems you can get people to change their estimate of the value of bird life just by changing the number in the question. Poor forecasters do the same thing on their predictions. For example, a hundred poor forecasters might on average predict a 15% chance of war in Korea in the next five years, and a different group of a hundred poor forecasters might on average predict a 15% chance of war in Korea in the next fifteen years. They’re ignoring the question and just going off of a vague feeling of how likely another Korean war seems. Superforecasters, in contrast, showed much reduced scope insensitivity, and their probability of a war in five years was appropriately lower than of a war in fifteen.

Maybe all this stuff about probability calibration, inside vs. outside view, willingness to change your mind, and fighting cognitive biases is starting to sound familiar? Yeah, this is pretty much the same stuff as in the Less Wrong Sequences and a lot of CFAR work. They’re both drawing from the same tradition of cognitive science and rationality studies.

So as I said before, *Superforecasting* is not necessarily too useful for people who are already familiar with the cognitive science/rationality tradition, but great for people who need a high-status and official-looking book to justify it. The next time some random person from a terrible forum says that everything we’re doing is stupid, I’m already looking forward to pulling out Tetlock quotes like:

> The superforecasters are a numerate bunch: many know about Bayes’ theorem and could deploy it if they felt it

> was worth the trouble. But they rarely crunch the numbers so explicitly. What matters far more to the superforecasters than Bayes' theorem is Bayes' core insight of gradually getting closer to the truth by constantly updating in proportion to the weight of the evidence. That's true of Tim Minto [the top superforecaster]. He knows Bayes' theorem, but he didn't use it even once to make his hundreds of updated forecasts. And yet Minto appreciates the Bayesian spirit. "I think it is likely that I have a better intuitive grasp of Bayes' theorem than most people," he said, "even though if you asked me to write it down from memory I'd probably fail." Minto is a Bayesian who does not use Bayes' theorem. That paradoxical description applies to most superforecasters.

And if you're interested, it looks like there's a current version of the Good Judgment Program going on here that you can sign up to and see if you're a superforecaster or not.

EDIT: A lot of people have asked the same question: am I being too dismissive? Isn't it really important to have this book as evidence that these techniques work? Yes. It is important that the Good Judgment Project exists. But you might not want to read a three-hundred page book that explains lots of stuff like "Here's what a cognitive bias is" just to hear that things work. If you already know what the techniques are, it might be quicker to read a study or a popular news article on GJP or something.

List Of Passages I Highlighted In My Copy Of “Superforecasting”

> In year 1, [the Good Judgment Project] beat the official control group by 60%. In year 2, we beat the control group by 78%. GJP also beat its university-affiliated competitors, including the Uniersity of Michigan and MIT, by hefty margins, from 30% to 70%, and even outperformed professional intellgience analysts with access to classified data. After two years, GJP was doing so much better than its academic competitors that IARPA dropped the other teams.

I keep wondering what these other teams were doing. Good Judgment Project sounds like it was doing the simplest, most obvious possible tactic – asking people to predict things and seeing what happened. David Manheim says the other groups tried “more straightforward wisdom of crowds” methods, so maybe GJP’s secret sauce was concentrating on the best people instead of on everyone? Still seems like it should have taken fewer than five universities and a branch of government to think of that.

> One result that particularly surprised me was the effect of a tutorial covering some basic concepts that we’ll explore in this book and are summarized in the Ten Commandments appendix. It took only about sixty minutes to read and improved accuracy by roughly 10% through the entire tournament year. Yes, 10% may sound modest, but it was achieved at so little cost.

These Ten Commandments are available online here.

> For centuries, [aversion to measuring things and collecting evidence] hobbled progress in medicine.

> When physicians finally accepted that their experience and perceptions were not reliable means of determining whether a treatment works, they turned to scientific testing – and medicine finally started to make rapid advances.

I see what Tetlock is trying to say here, but as written it’s horribly wrong.

Evidence-based medicine could be fairly described as starting in the 1970s with Cochrane’s first book, and really took off in the 80s and 90s. But this is also the period when rapid medical advances started slowing down! In my own field of psychiatry, the greatest advances were the first antidepressants and antipsychotics in the 50s, the benzodiazepines in the 60s, and then a gradual trickle of slightly upgraded versions of these through the 70s and 80s. The last new drugs that could be called “revolutionary” by any stretch of the imagination were probably the first SSRIs in the early 80s. This is the conventional wisdom of the field and everybody admits this, but I would add the stronger claim that the older medications in many ways work better. I know less about the history of other subfields, but they seem broadly similar – the really amazing discoveries are all pre-EBM, and the new drugs are mostly nicer streamlined versions of the old ones.

There’s an obvious “low-hanging fruit” argument to be made here, but some people (I think Michael Vassar sometimes toys with this idea) go further and say that evidence-based medicine as currently practiced can actually retard progress. In the old days, people tried possible new medications in a very free-form and fluid way that let everyone test their pet ideas quickly and keep the ones that worked; nowadays any potential innovations need $100 million 10-year multi-center trials which will only get funded in certain very specific situations. And in the old days, a drug would only be kept if it showed obvious undeniable improvement in patients, whereas nowadays if a trial shows a $p < 0.05$, $d = 0.10$ advantage, that's enough to make it the new standard if it's got a good pharma company behind it. So the old method allowed massive-scale innovation combined with high standards for success; the new method only allows very limited innovation but keeps everything that can show the slightest positive effect whatsoever on an easily-rigged but very expensive test. I'm not sure I believe in the strong version of this argument (the low-hanging fruit angle is probably sufficient), but the idea that medicine only started advancing after the discovery of evidence-based medicine is just wrong. A better way of phrasing it might be that around that time we started getting fewer innovations, but we also became a lot more effective and intelligent at using the innovations we already had.

> Consider Galen, the second-century physician to Rome’s emperors…Galen was untroubled by doubt. Each outcome confirmed he was right, no matter how equivocal the evidence might look to someone less wise than the master. “All who drink of this treatment recover in a short time, except those whom it does not help, who all die,” he wrote. “It is obvious, therefore, that it fails only in incurable cases.”

After hearing one too many “everyone thought Columbus would fall off the edge of the flat world” -style stories, I tend to be skeptical of “people in the past were hilariously stupid” anecdotes. I don’t know anything about Galen, but I wonder if this was really the whole story.

> When hospitals created cardiac care units to treat patients recovering from heart attacks, Cochrane proposed a randomized trial to determine whether the

> new units delivered better results than the old treatment, which was to send the patient home for monitoring and bed rest. Physicians balked. It was obvious the cardiac care units were superior, they said, and denying patients the best care would be unethical. But Cochrane was not a man to back down…he got his trial: some patients, randomly selected, were sent to the cardiac care units while others were sent home for monitoring and bed rest. Partway through the trial, Cochrane met with a group of the cardiologists who had tried to stop his experiment. He told them that he had preliminary results. The difference in outcomes between the two treatments was not statistically signficant, he emphasized, but it appeared that patients might do slightly betteri n the cardiac care units. "They were vociferous in their abuse: 'Archie,' they said, 'we always thought you were unethical. You must stop the trial at once.'" But then Cochrane revealed he had played a little trick. He had reversed the results: home care had done slightly better than the cardiac units. "There was dead silence and I felt rather sick because they were, after all, my medical colleagues."

This story is the key to everything. See also my political spectrum quiz and the graph that inspired it. Almost nobody has consistent meta-level principles. Almost nobody really has opinions like "this study's methodology is good enough to believe" or "if one group has a survival advantage of size X, that necessitates stopping the study as unethical". The cardiologists sculpted their meta-level principles around what best supported their object-level opinions – that more cardiology is better – and so generated the meta-level principles "Cochrane's experiment is accurate" and "if one group has a slight survival advantage, that's all we need to know before ordering the experiment stopped as unethical." If Cochrane had (truthfully) told them that the cardiology group was doing worse, they would have generated the meta-level principles “Cochrane’s experiment is flawed” and “if one group has a slight survival advantage that means nothing and it’s just a coincidence”. In some sense this is correct from a Bayesian point of view – I interpret sonar scans of Loch Ness that find no monsters to be probably accurate, but if a sonar scan did find a monster I’d wonder if it was a hoax – but in less obvious situations it can be a disaster. Cochrane understood this and so fed them the wrong data and let them sell him the rope he needed to hang them. I know no better solution to this except (possibly) adversarial collaboration. Also, I suppose this is more proof (as if we needed it) that cardiologists are evil.

> In the late 1940s, the Communist government of Yugoslavia broke from the Soviet Union, raising fears that the Soviets would invade. In March 1951 [US intelligence under Sherman Kent reported there was a “serious possibility” of a Soviet attack.] But a few days later, Kent was chatting with a senior State Department official who casually asked, “By the way, what did you people mean by the expression ‘serious possibility’? What kind of odds did you have in mind?” Kent said he was pessimistic. He felt that the odds were about 65 to 35 in favor of an attack. The official was startled. He and his colleagues had taken “serious possibility” to mean much lower odds.
>
> Disturbed, Kent went back to his team. They had all agreed to use “serious possibility” in the [report], so Kent asked each person, in turn, what he thought it meant. One analyst said it meant odds of about 80%. Another thought it meant odds of 20% – exactly the opposite. Other answers were scattered between those extremes. Kent was floored. A phrase that looked informative was so vague as to be almost useless…

> In 1961, when the CIA was planning to topple the Castro government by landing a small army of Cuban expatriates at the Bay of Pigs, President John F. Kennedy turned to the military for an unbiased assessment. The Joint Chiefs of Staff concluded that the plan had a "fair chance" of success. The man who wrote the words "fair chance" later said he had in mind odds of 3 to 1 against. But Kennedy was never told precisely what "fair chance" meant and, not unreasonably, he took it to be a much more positive assessment.

…

> Nate Silver, Princeton's Sam Wang, and other poll aggregators were hailed for correctly predicting all fifty state outcomes, but almost no one noted that a crude, across-the-board prediction of "no change" – if a state went Democratic or Republican in 2008, it will do the same in 2012 – would have scored forty-eight out of fifty, which suggests that the many excited exclamations of "he called all fifty states!" we heard at the time were a tad overwrought.

I didn't realize this. I think this election I'm going to predict the state-by-state results just so that I can tell people I "predicted 48 of the 50 states" or something and sound really impressive.

> The [Expert Political Judgment] data revealed an inverse correlation between fame and accuracy: the more famous an expert was, the less accurate he was. That's not because editors, producers, and the public go looking for bad forecasters. They go looking for hedgehogs, who just happen to be bad forecasters. Animated by a Big Idea, hedgehogs tell tight, simple, clear stories that grab and hold audiences.

One day aliens are going to discover humanity and be absolutely shocked we made it past the wooden-club stage.

> In 2008, the Office of the Director of national Intelligence – which sits atop the entire network of sixteen intelligence agencies -asked the National research Council to form a committee. The task was to synthesize research on good judgment and help the IC put that research to good use. By Washington’s standards, it was a bold (or rash) thing to do. It’s not every day that a bureaucracy pays one of the world’s most respected scientific institutions to produce an objective report that might conclude that the bureaucracy was clueless.

This was a big theme of the book: the US intelligence community deserves celebration for daring to investigate its own competency at all. Interestingly, a lot of its investigations said it was doing things more right than we would think: Tetlock mentions that even independent-to-hostile investigators concluded that it had been correct in using the facts it had to believe Saddam had WMDs. The book didn’t explain exactly how this worked: possibly Saddam was trying to deceive everyone into thinking he had WMDs to prevent attacks, and did a good job? This was part of what got the intelligence community interested in probability: given that they had made a reasonable decision in saying there were WMDs, but it had been a big disaster for the United States, what could they have done differently? Their answer was “continue to make the reasonable decision, but learn to calibrate themselves well enough to admit there’s a big chance they’re wrong.”

> [We finished by giving] the forecast a final tweak: “extremizing” it, meaning pushing it closer to 100% or zero. If the forecast is 70% you might bump it up to,

> say, 85%. If it’s 30%, you might reduce it to 15%…[it] is based on a pretty simple insight: when you combine the judgments of a large group of people to calculate the “wisdom of the crowd” you collect all of the relevant information that is dispersed among all those people. But none of those people has access to all that information…what would happen if every one of those people were given all the information? They would become more confident. If you then calculated the wisdom of the crowd, it too would be more extreme.

Something to remember if you’re doing wisdom-of-crowds with calibration estimates.

> The correlation between how well individuals do from one year to the next is about 0.65…Regular forecasters scored higher on intelligence and knowledge tests than about 70% of the population. Superforecasters did better, placing higher than about 80% of the population.

People interested in taking these kinds of tests are generally intelligent; superforecasters are somewhat more, but not vastly more, intelligent than that.

> Researchers have found that merely asking people to assume their initial judgment is wrong, to seriously consider why that might be, and then make another judgment, produces a second estimate which, when combined with the first, improves accuracy almost as much as getting a second estimate from another person.

There’s a rationalist tradition – I think it started with Mike and Alicorn – that before you get married, you ask all your friends to imagine that the marriage failed and tell you why. I guess if you just asked people “Will our marriage fail?”

everyone would say no, either out of optimism or social desirability bias. If you ask “Assume our marriage failed and tell us why”, you’ll actually hear people’s concerns. I think this is the same principle. On the other hand, I’ve never heard of anyone trying this and deciding not to get married after all, so maybe we’re just going through the motions.

> [Superforecaster] Doug Lorch knows that when people read for pleasure they naturally gravitate to the like-minded. So he created a database containing hundreds of information sources – from the New York Times to obscure blogs – that are tagged by their ideological oreintation, subject matter, and geographical origin, then wrote a program that selects what he should read next using criteria that maximize diversity.

Of all humans, only Doug Lorch is virtuous. Well, Doug Lorch and this guy from rationalist Tumblr who tried to get the program but was told it wasn’t really the sort of thing you could just copy and give someone.

> [The CIA was advising Obama about whether Osama bin Laden was in Abbotabad, Pakistan; their estimates averaged around 70%]. “Okay, this is a probability thing,” the President said in response, according to Bowden’s account. Bowden editorializes: “Ever since the agency’s erroneous call a decade earlier [on Saddam’s weapons of mass destruction], the CIA had instituted an almost comically elaborate process for weighing certainty…it was like trying to controve a mathematical formula for good judgment.”Bowden was clearly not impressed with the CIA’s use of numbers and probabilities. Neither was Barack Obama, according to Bowden. “What you ended up with, as the president was finding, and as he would later explain to me, was not more certainty but more confusion…in this situation,

> what you started to get was probabilities that disguised uncertainty, as opposed to actually providing you with useful information…”
>
> After listening to the widely ranging opinions, Obama addressed the rrom. “This is fifty-fifty,” he said. That silenced everyone. “Look guys, this is a flip of the coin. I can’t base this decision on the notion that we have any greater certainty than that…
>
> The information Bowden provides is sketchy but it appears that the media estimate of the CIA officers – the “wisdom of the crowd” – was around 70%. And yet Obama declares the reality to be “fifty-fifty.” What does he mean by that?…Bowden’s account reminded me of an offhanded remark that Amos Tversky made some thirty years ago…In dealing with probabilities, he said, most people only have three settings: “gonna happen,” “not gonna happen,” and “maybe”.

Lest I make it look like Tetlock is being too unfair to Obama, he goes on to say that maybe he was speaking colloquially. But the way we speak colloquially says a lot about us, and there are many other examples of people saying this sort of thing and meaning it. This ties back into an *old* argument we had here on whether something like a Bayesian concept of probability was meaningful/useful. Some people said that it wasn’t, because everyone basically understands probability and Bayes doesn’t add much to that. I said it was, because people’s intuitive idea of probability is hopelessly confused and people don’t really think in probabilistic terms. I think we have *no idea* how confused most people’s idea of probability is, and perhaps even Obama, one of our more intellectual presidents, has some issues there.

> Barbara Mellers has shown that granularity predicts accuracy: the average forecaster who sticks with the tens – 20%, 30%, 40% – is less accurate than the finer-grained forecaster who uses fives – 20%, 25%, 30% – and still less accurate than the even finer-grained forecaster who uses ones – 20%, 21%, 22%. As a further test, she rounded forecasts to make them less granular, so a forecast at the greatest granularity possible in the tournament, single percentage points, would be rounded to the nearest five, and then the nearest ten. This way, all of the forecasts were made one level less granular. She then recalculated Bier scores and discovered that superforecasters lost accuracy in response to even the smallest-scale rounding, to the nearest 0.05, whereas regular forecasters lost little even from rounding four times as large, to the nearest 0.2.

This was the part nobody on the comments to the last post believed, and I have trouble believing it too.

> [There’s a famous Keynes quote: “When the facts change, I change my mind. What do you do, sir?”] It’s cited in countless books, including one written by me and another by my coauthor. Google it and you will find it’s all over the internet. Of all the many famous things Keynes says, it’s probably the most famous. But while researching this book, I tried to track it to its source and failed. Instead I found a post by a Wall Street Journal blogger, which said that no one has ever discovered its provenance and the two leading experts on Keynes think it is apocryphal. In light of these facts, and in the spirit of what Keynes apparently never said, I concluded that I was wrong.

The funny part is that if this fact is true, we’ve known it for fifty years, and people still haven’t changed their mind about whether he said it or not.

> “Keynes is always ready to contradict not only his colleagues but also himself whenever circustancse make this seem appropriate,” re[prted a 1945 profile of the “consistently inconsistent” economist. “So far from feeling guilty about such reversals of position, he utilizes them as pretexts for rebukes to those he saw as less nimble-minded. Legend says that while conferring with Roosevelt at Quebec, Churchill sent Keynes a cable reading, ‘Am coming around to your point of view.’ His Lordship replied, ‘Sorry to hear it. Have started to change my mind.'”

I sympathize with this every time people email me to say how much they like the Non-Libertarian FAQ.

> Police officers spend a lot of time figuring out who is telling the truth and who is lying, but research has found they aren’t nearly as good at it as they think they are and they tend not to get better with experience… predictably, psychologists who test police officers’ ability to spot lies in a controlled setting find a big gap between their confidence and their skill. And that gap grows as officers become more experienced and they assume, not unreasonably, that their experience has made them better lie detectors.

There’s some similar research on doctors and certain types of diagnostic tasks that don’t give quick feedback.

> In 1988, when the Soviet Union was implementing major reforms that had people wondering about its future, I asked experts to estimate how likely it was that the Communist Party would lose its monopoly on power in the Soviet Union in the next five years. In 1991 the

> world watched in shock as the Soviet Union disintegrated. So in 1992-93 I retunred to the experts, reminded them of the question in 1988, and asked them to recall their estimates. On average, the experts recalled a number 31 percentage points higher than the correct figure. So an expert who thought there was only a 10% chance might remember herself thinking there was a 40% or 50% chance. There was even a case in which an expert who pegged the probability at 20% recalled it as 70%.

As the old saying goes, hindsight is 20/70.

> The results were clear-cut each year. Teams of ordinary forecasters beat the wisdom of the crowd by about 10%. Prediction markets beat ordinary teams by about 20%. And superteams beat prediction markets by 15% to 30%. I can already hear the protests from my colleagues in finance that the only reason the superteams beat the prediction markets was that our markets lacked liquidity…they may be right. It is a testable idea, and one worth testing.

The correct way to phrase this is “if there is ever a large and liquid prediction market, Philip Tetlock will gather his superforecasters, beat the market, become a zillionaire, and then the market will be equal to or better than the forecasters.”

> Orders in the Wehrmacht were often short and simple – even when history hung in the balance. “Gentlemen, I demand that your divisions completely cross the German borders, completely cross the Belgian borders, and completely cross the River Meuse,” a senior officer told the commanders who would launch the great

> assault into Belgium and France on May 10, 1940. “I don’t care how you do it, that’s completely up to you.”
>
> This is the opposite of the image most people have of Germany’s World War II military. The Wehrmacht served a Nazi regime that rpeached total obedience to the dictates of the Fuhrer, and everyone emembers the old newsreels of German soldiers marching in goose-stepping unison…but what is often forgotten is that the Nazis did not create the Wehrmacht. They inherited it. And it could not have been more different from the unthinking machine we imagine.
>
> […]
>
> Shortly after WWI, Eisenhower, then a junior officer who had some experience witht he new weapons called tanks, published an article in the US Army’s *Infantry Journal* making the modest argument that “the clumsy, awkward and snail-like progress of the old tanks must be forgotten, and in their place we must picture this speedy, reliable, and efficient engine of destruction.” Eisenhower was dressed down. “I was told my ideas were not only wrong but dangerous, and that henceforth I was to keep them to myself,” he recalled. “Particularly, I was not to publish anything incompatible with solid infantry doctrine. If I did, I would be hauled before a court martial.”

Tetlock includes a section on what makes good teams and organizations. He concludes that they’re effective when low-level members are given leeway both to pursue their own tasks as best they see fit, and to question and challenge their higher-ups. He contrasts the Wehrmacht, which was very good at this and overperformed its fundamentals in WWII, to the US Army, which was originally very bad at this and underperformed its fundamentals until it figured this out. Later in the chapter, he admits that his choice of examples might raise some eyebrows, but says that he did it on purpose to teach us to think critically and overcome cognitive dissonance between our moral preconceptions and our factual beliefs. I hope he has tenure.

> Ultimately the Wehrmacht failed. In part, it was overwhelmed by its enemies' superior resources. But it also made blunders – often because its commander-in-chief, Adolf Hitler, took direct control of operations in violation of Helmuth von Moltke's principles, nowhere with more disastrous effect than during the invasion of Normandy. The Allies feared that after their troops landed, German tanks would drive them back to the beaches and into the sea, but Hitler had directed that the reserves could only move on his personal command. Hitler slept late. For hours after the Allies landed on the beaches, the dictator's aides refused to wake him to ask if he wanted to order the tanks into battle.

Early to bed
And early to stir up
Makes a man healthy
And ruler of Europe

> The humility required for good judgment is not self-doubt – the sense that you are untalented, unintelligent, or unworthy. It is intellectual humility. It is a recognition that reality is profoundly complex, that seeing things clearly is a constant struggle, when it can be done at all, and that human judgment must therefore be riddled with mistakes. This is true for fools and geniuses alike. So it's quite possible to think highly of yourself and be intellectually humble. In fact, this combination can be wonderfully fruitful. Intellectual humility compels the

> careful reflection necessary for good judgment; confidence in one’s abilities inspires determined action.

Yes! This is a really good explanation of Eliezer Yudkowsky’s Say It Loud.

(and that sentence would also have worked without the apostrophe or anything after it).

> I am…optimistic that smart, dedicated people can inoculate themselves to some degree against certain cognitive illusions. That may sound like a tempest in an academic teapot, but it has real-world implications. If I am right, organizations will have more to gain from recruiting and training talented people to resist their biases.

This is probably a good time to mention that CFAR is hiring.

Testimonials for SSC

[Content note: various slurs and insults]

I.

Last post I thanked some of the people who have contributed to this blog. But I forgot some of the most important contributors: the many readers whose give valuable feedback on everything I write.

So here’s a short sample of some of the feedback I’ve gotten over the past three years. I’m avoiding names and links to avoid pile-ons, but you can probably find most of these if you Google them.

II.

“It’s like someone tried to make fivethirtyeight as uninteresting as possible.”

“Slate Star Codex: 20,000 words on ‘feminism is bad’ and ‘Tom Swifties are the funniest shit I’ve ever seen'”

“Mark Atwood spent a month on the SSC registry of bans. It is my belief that Scott uses his toolbox of psychiatric techniques to manage the range of comment allowable on his blog. This may be justified in order to minimize flaming and trolling, but it is also a stifling form of censorship.”

“Go read the comment section on Slate Star Codex for a week and report back if you think LessWrongism is acceptable. It’s a place for broken people to be shielded from ever hearing that they’re broken and for developing better and better rationalizations for why they’re not broken and shouldn’t do the work needed to fix themselves. And

SSC is miles better than lesswrong itself which is a weird cult centered around gnome in charge Eliezer Yudkowski. Just read the post where Scott let ozymandius post about all the ways that Roissy is wrong. The information there permanently disqualifies anyone associated with it from from having anything to do with building a functioning society."

"Scott Alexander's blog used to be good, but now he has been terrorized out of politics. Therefore boring. The problem was he purged all frequent commentors to the right of him out of the comments, which means that he had only enemies in his comments. And, being the rightmost, was persecuted. He has stopped posting on politics, I assume as a result of this persecution."

"it's by a stuttering aspie with expertise in nothing at all"

"a mentally-ill beta male who literally admitted that he wished he could become an asexual."

"He is a fairly smart guy who makes well reasoned arguments. (He is also a literal cuckold.)"

"never forget for one fucking second that its author (who is 'asexual') and his most avid readers engage in 'cuddle puddles' irl, often bringing stuffed animals, and that he recommends this because it 'increases credence' in the other cuddlers' statements."

"his arguments only seem well-reasoned to people with NO knowledge of the subject matter (like him), and thus his main effect (just like that of his mentor, Eliezer Yudkowsky) is to keep smart people from learning things"

"oh what an expert in psychiatry! he's a fucking med student in IRELAND. not to mention he uses yudkowsky's lingo and called Less Wrong “revelatory” or something like that. you are dealing with a 110 IQ reddit type in SSC.”

“here is a series of a few posts (1, 2, 3) about how he is basically a conspiracy theorist, and in these posts he gets completely owned by the guy who cucked him with his tranny ex.”

“Merciful $DEITY. If I had any inclination to participate [on SSC], that [Guns and States] comment thread would have turned me completely off of it. How much more SJW-feminist-entitled can you get?…Extreme? No. Hard-line SJW enough that I’ve got better things to do than try to engage them? Yes. SJW-feminist-entitled? Yes.”

“SSC skews toward highly intelligent discourse, but Scott is very protective of his liberal homies. He will ban you if you stray too far from PC rigor.”

“I’d always got a whiff of fedora from this guy, so I feel gratified in my judgment at seeing him come out as one.”

“Faggot blocked me for calling out some recent bit of his retarded bullshit. Fuck ‘im.”

“Reading Slate Star Codex, I feel like I’m finally strarting to understand how postmodernism happened. First, there’s the whole thing of looking at your friends and a few books you happened to have read recently, and jumping to grand conclusions about all of society throughout all of history. Second, there’s the thing of him writing lengthy posts elaborating at great length on something that might be either boring and obviously true or bold and innovative but also completely wrong.”

“Aargh!! I read that entire Scott Alexander piece…well 65% of it…in earnest with the expectation that there was going to be a POINT to it. Some sort of payoff for my investment of time and attention. But there was nothing. It was just a bunch of bloviation with no purpose."

"Scott Alexander is the story of a functioning pattern-recognition module trapped in a progressive brain. It would make a great story of its truth-seeking brain blob could eventually break free and rewire his brain to be a born-again reactionary. Not gonna happen though. The prog morality police has a hard, thick grasp on his brain, and all his friends and pseudosexual partners are the leftiest hacks this side of Lenin; so it's an endless futile battle to square the circle. No wonder he went into psychiatry."

"I retain great hopes for Scott, he'll come around. When he does he'll bring a high level of rigor with him. He is a caterpillar and will become a beautiful reactionary butterfly someday."

"slatestarcodex is a great example of the difference between 'knowing how to type' and 'knowing how to write'"

"You aren't reading it right. Scott's ability to completely identify the problem but still, quite sincerely, ritually abase himself to it at the same time, makes him worthy of connoisseurship. It takes a once in a generation talent to write long sincere *thoughtful* screeds pointing out that baby sacrifice is lowering the birth rate and causing family trauma, though of course he fully understands and endorses that Lord Moloch must be sated with the only food acceptable unto him."

"Scott Alexander reminds me of some too-nice beta (nominally played by Joseph Gordon Leavitt or a JGL-alike) from some rom-com who's trying to find his way and get the girl, while painfully oblivious to the fact that he just needs to stop being a too-nice beta and rip out somebody’s jugular. Ostracize someone for their beliefs? Me? Never. Golly gee.”

“seems like these guys are incapable of being dismissive of anything and have to objectively analyze everything.”

“Slate Star Codex is 140 IQ discussion about 105 IQ issues”

“He is sharp and makes good points but is way too fucking verbose. I dont need parts I II III and IV just fucking write concisely and stop vomiting words on your wordpress blog ”

“it’s basically a fish trap for aspies. people who can’t grasp nuance or understand basic human behavior, but are nonetheless obsessed with details and complex systems will inevitably gravitate toward this kind of horseshit. ultimately it’s a bunch of STEM-inclined dudes on the autism spectrum sitting around attempting to unpack societal problems like it was all a game of fucking sim city.”

“I would add that something like Slate Star Codex is also a clinic in the aspie tendency to miss the forest for the trees, except in this case it’s more like closely examining the bark on the trees for no goddamn reason whatsoever.”

“doesn’t this guy have a dayjob as like a doctor or something? why the fuck does he spend hours each day on a blog?” [to which another person on the same forum responded “why the fuck do you spend hours each day posting here?”]

“a blog populated by 99th percentile aspergers/IQ “rationalist” millennials who converse in an abnormally abstract style, and whose concrete cultural experience is drawn mainly from a bunch of weird nerd shit.”

“Its weird brand of reductionism and bizarre, arbitrary specificity plays to the types of spergy assholes and dumb know-it-all teenagers who don’t care about that anyway, or at least that’s how it seems to me. I mean, the ideas themselves seem like they’d be as much of a turn-off to regular people as their proponents’ personalities are, even if in a different way.”

“Oh, hey, the King of the Race Realist Misogynist Libertarian Nerds has Clever Things to say about vaccination.”

“yet another confirmation that: psychiatrists are crazier than their patients. polyamorous, diarrhea of the mouth/pen, math challenged, … i had no idea what an utter piece of shit you were.”

“He keeps his head down for fear of insulting permanently insulted people. He tries hard to be polite to people who hate him and consider him but a dog, unless they need him – and until they need him no longer. It is a waste of intellect, and debasement of character.”

“What makes me sad about Scott is just how close he is. I won’t give up hope on him yet. If only there was some way to secretly inject this guy with testosterone.”

“I wonder if he’s had bloodwork done to check his T count. I have to assume that if someone is an “asexual heteroromantic” as he puts it, that he’s interested in women from some abstract standpoint, and just needs some additional hormones to be thoroughly normal.”

“I literally want to see you kill yourself. I’m serious. You, and everyone else like you, are fucking disgusting wastes of space that are causing the decay of decency in the human race. I’m not going to argue with you, or say that it’s just my opinion or that it’s even up for debate.”

“is it some sort of special ‘Talk Like a Vulcan Day’ over there? Or are they always like that?”

“ssc spends a significant amount of time talking about stuff like how tables and chairs can be genders. he keeps a pretty unhinged tumblr”

“He’s definitely a beta orbiting cuckold.”

“that article seemed like a return ticket to obviousville with eight-hour layovers everywhere”

“That blog is very boring, and I didn’t manage to read long enough to find out what it was about. I hit Page Down a couple of times, and it seemed like it was on an entirely new topic each time.”

“I am thankful that I have never had any desire to seek psychiatric help. I have always had the impression, rightly or wrongly, that folks who pursue psychiatry as a career may themselves be the ones most in need of such therapy. Go to the mountains and look. Get up early and see the sunrise. Stop anywhere and take a minute to look at the beauty of nature all around you. We are a small piece in the universe, but still a part. The plan is good. You are fine. You will succeed if you try hard enough. Everything you need spiritually is inside you and has always been there. Stop complaining.”

“Also that ‘heteronormative asexual’ Scott Alexander. What a bizarre kike. He recently wrote that he’s incapable of not writing. LOL so kikish.”

“I thought it was a blog about science methodology until that post with the talking cactus.”

Before You Get Too Excited About That GitHub Study…

Another day, another study purporting to find that Tech Is Sexist. Since it’s showing up here, you probably already guessed how this is going to end. Most of this analysis is not original to me – Hacker News had figured a lot of it out before I even woke up this morning – but I think it’ll at least be helpful to collect all the information in one easily linkable place.

The study is Gender Bias In Open Source: Pull Request Acceptance Of Women Vs. Men. It’s a pretty neat idea: “pull requests” are discrete units of contribution to an open source project which are either accepted or rejected by the community, so just check which ones are submitted by men vs. women and whether one gender gets a higher acceptance rate than the other. This is a *little* harder than it sounds – people on GitHub use nicks that don’t always give gender cues – but the researchers wrote a program to automatically link contributor emails to Google Plus pages so they could figure out users’ genders.

This alone can’t rule out that one gender is genuinely doing something differently than another, so they had another neat trick: they wrote another program that automatically scored accounts on obvious gender cues: for example, somebody whose nickname was JaneSmith01, or somebody who had a photo of themselves on their profile. By comparing obviously gendered participants with non-obviously gendered participants whom the researchers had nevertheless been able to find the gender of, they should be able to tell whether there’s gender bias in request acceptances.

Because GitHub is big and their study is automated, they manage to get a really nice sample size – about 2.5 million pull requests by men and 150,000 by women.

They find that women get more (!) requests accepted than men for all of the top ten programming languages. They check some possible confounders – whether women make smaller changes (easier to get accepted) or whether their changes are more likely to serve an immediate project need (again, easier to get accepted) and in fact find the opposite – women’s changes are larger and less likely to serve project needs. That makes their better performance extra impressive.

So the big question is whether this changes based on obviousness of gender. The paper doesn’t give a lot of the analyses I want to see, and doesn’t make its data public, so we’ll have to go with the limited information they provide. They do *not* provide an analysis of the population as a whole (!) but they do give us a subgroup analysis by “insider status”, ie whether the person has contributed to that project before.

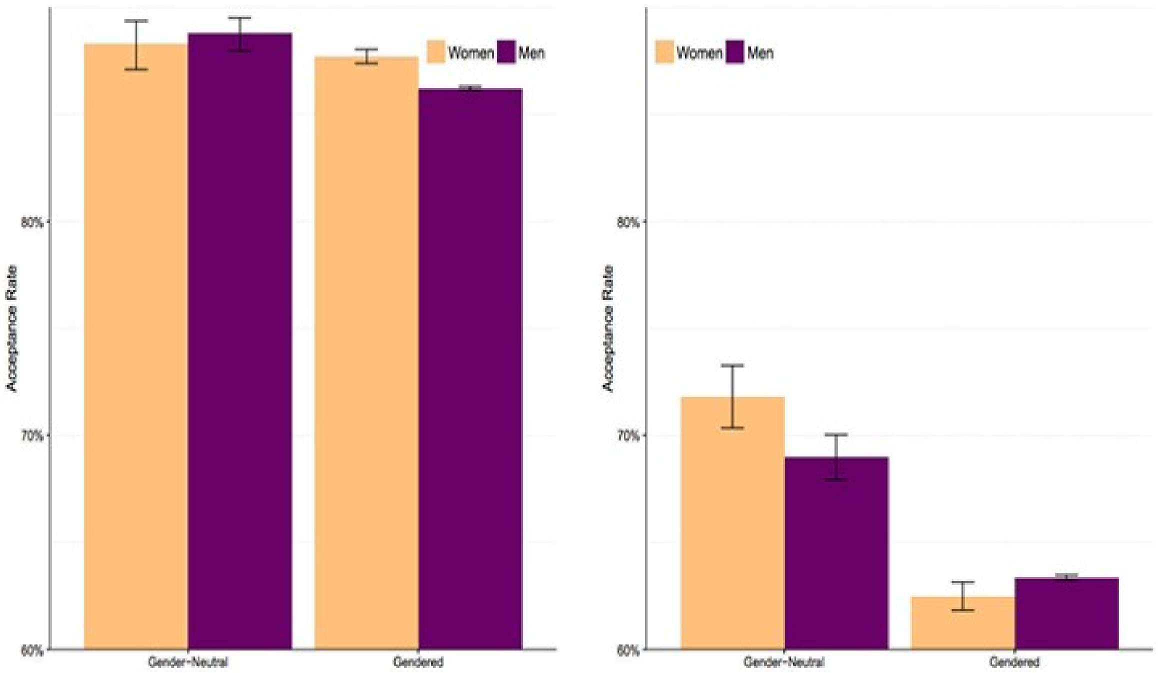


Figure 5: Pull request acceptance rate by gender and perceived gender, with 95% Clopper-Pearson confidence intervals, for insiders (left) and outsiders (right)

Among insiders, women do the same as men when gender is hidden, but better than men when gender is revealed. In other words, if you know somebody's a woman, you're more likely to approve her request than you would be on the merits alone. We can't quantify exactly how much this is, because the paper doesn't provide numbers, just graphs. Eyeballing the graph, it looks like being a woman gives you about a 1% advantage. I don't see any discussion of this result, even though it's half the study, and as far as I can tell the more statistically significant half.

Among outsiders, women do the same as/better than men when gender is hidden, and the same as/worse than men when gender is revealed. I can't be more specific than this because the study doesn't give numbers and I'm trying to eyeball confidence intervals on graphs. The study itself say that women do worse than men when gender is revealed, so since the researchers presumably have access to their real numbers data, that might mean the confidence intervals don't overlap. From eyeballing the graph, it looks like the difference is 1% – ie, men get their requests approved 64% of the time, and women 63% of the time. Once again, it's hard to tell by graph-eyeballing whether these two numbers are within each other's confidence intervals.

The paper concludes that "for insiders…we see little evidence of bias…for outsiders, we see evidence of gender bias: women's acceptance rates are 71.8% when they use gender neutral profiles, but drop to 62.5% when their gender is identifiable. There is a similar drop for men, but the effect is not as strong."

In other words, they conclude there is gender bias among outsiders because obvious-women do worse than gender-anonymized-women. They admit that obvious-men also do worse than gender-anonymized men, but they ignore this effect because it's smaller. They do not report doing a test of statistical significance on whether it is really smaller or not.

So:

1. Among insiders, women get more requests accepted than men.

2. Among insiders, people are biased towards women, that is, revealing genders gives women an advantage over men above and beyond the case where genders are hidden.

3. Among outsiders, women still get more requests accepted than men.

4. Among outsiders, revealing genders appears to show a bias against women. It's not clear if this is statistically significant.

5. When all genders are revealed among outsiders, men appear to have their requests accepted at a rate of 64%, and women of 63%. The study does not provide enough information to determine whether this is statistically significant. Eyeballing it it looks like it might be, just barely.

6. The study describes its main finding as being that women have fewer requests approved when their gender is known. It hides on page 16 that men also have fewer requests approved when their gender is known. It describes the effect for women as larger, but does not report the size of the male effects, nor whether the difference is statistically significant. Eyeballing it, it looks about 2/3 the size of the female effect, and maybe?

7. The study has no hypothesis for why both sexes have fewer requests approved when their gender is known, without which it seems kind of hard to speculate about the significance of the phenomenon for one gender in particular. For example, suppose that the reason revealing gender decreases acceptance rates is because corporate contributors tend to use their (gendered) real names and non-corporate contributors tend to use handles like 133T_HAXX0R. And suppose that the best people of all genders go to work at corporations, but a bigger percent of men go there than women. Then being non-gendered would be a higher sign of quality in a man than in a woman. This is obviously a silly just-so story, but my point is that without knowing why all genders show a decline after unblinding, it’s premature to speculate about why their declines are of different magnitudes – and it doesn’t take much to get so small a difference.

8. There’s no study-wide analysis, and no description of how many different subgroup analyses the study tried before settling on Insiders vs. Outsiders (nor how many different

definitions of Insider vs. Outsider they tried). Remember, for every subgroup you try, you need to do a Bonferroni correction. This study does not do any Bonferroni corrections; given its already ambiguous confidence intervals, a proper correction would almost certainly destroy the finding.

9. We still have that result from before that women's changes are larger and less likely to serve immediate needs, both of which make them less likely to be accepted. No attempt was made to control for this.

"Science" "journalism", care to give a completely proportionate and reasonable response to this study?

Here's Business Insider: Sexism Is Rampant Among Programmers On GitHub, Research Finds. "A new research report shows just how ridiculously tough it can be to be a woman programmer, especially in the very male-dominated world of open-source software….it also shows that women face a giant hurdle of "gender bias" when others assess their work. This research also helps explain the bigger problem: why so many women who do enter tech don't stick around in it, and often move on to other industries within 10 years. Why bang your head against the wall for longer than a decade?" [EDIT: the title has since been changed]

Here's Tech Times: Women Code Better Than Men But Only If They Hide Their Gender: "Interestingly enough, among users who were not well known in the coding community, coding suggestions from those whose profiles clearly stated that the users were women had a far lower acceptance rate than suggestions from those who did not make their gender known. What this means is that there is a bias against women in the coding world." (Note the proportionate and reasonable use of the term "far lower acceptance rate" to refer to a female vs. male acceptance rate of, in the worst case, 63% vs. 64%.)

Here's Vice.com: Women Are Better At Coding Than Men: "If feminism has taught us anything, it's that almost all men are sexist. As this GitHub data shows, whether or not bros think that they view women as equals, women's work is not being judged impartially. On the web, a vile male hive mind is running an assault mission against women in tech."

This is normally the part at which I would question how a study got through peer review, but luckily this time there is a very simple answer: it didn't. If you read the study, you may notice the giant red "NOT PEER-REVIEWED" sign on the top of every page. The paper was uploaded to a pre-peer-review site asking for comments. The authors appear to be undergraduate students.

I don't blame the authors for doing a neat study and uploading it to a website. I do blame the entire world media up to and including the BBC for swallowing it uncritically. Note that two of the three news sources above failed to report that it is not peer-reviewed.

Oh, one more thing. A commenter on the paper's pre-print asked for a breakdown by *approver* gender, and the authors mentioned that "Our analysis (not in this paper — we've cut a lot out to keep it crisp) shows that women are harder on other women than they are on men. Men are harder on other men than they are on women."

Depending on what this means – since it was cut out of the paper to "keep it crisp", we can't be sure – it sounds like the effect is mainly from *women* rejecting other women's contributions, and men being pretty accepting of them. Given the way the media predictably spun this paper, it is hard for me to conceive of a level of crispness which justifies not providing this information.

So, let’s review. A non-peer-reviewed paper shows that women get more requests accepted than men. In one subgroup, unblinding gender gives women a bigger advantage; in another subgroup, unblinding gender gives men a bigger advantage. When gender is unblinded, both men and women do worse; it’s unclear if there are statistically significant differences in this regard. Only one of the study’s subgroups showed lower acceptance for women than men, and the size of the difference was 63% vs. 64%, which may or may not be statistically significant. This may or may not be related to the fact, demonstrated in the study, that women propose bigger and less-immediately-useful changes on average; no attempt was made to control for this. This tiny amount of discrimination against women seems to be mostly from other women, not from men.

The media uses this to conclude that “a vile male hive mind is running an assault mission against women in tech.”

Every time I say I’m nervous about the institutionalized social justice movement, people tell me that I’m crazy, that I’m just sexist and privileged, and that feminism is merely the belief that women are people so any discomfort with it is totally beyond the pale. I would nevertheless like to re-emphasize my concerns at this point.

*[**EDIT:** I don’t have much of a quarrel with the authors, who seem to have done an interesting study and are doing the correct thing by submitting it for peer review. I have a big quarrel with “science” “journalists” for the way they reported it. If any of the authors read this and want my peer review suggestions, I would recommend:*

1. Report gender-unblinding results for the entire population before you get into the insiders-vs.-outsiders dichotomy.
2. Give all numbers represented on graphs as actual numbers too.
3. Declare how many different subgroup groupings you tried, and do appropriate

Bonferroni corrections.

4. Report the magnitude of the male drop vs. the female drop after gender-unblinding, test if they're different, and report the test results.

5. Add the part about men being harder on men and vice versa, give numbers, and do significance tests.

6. Try to find an explanation for why both groups' rates dropped with gender-unblinding. If you can't, at least say so in the Discussion and propose some possibilities.

7. Fix the way you present "Women's acceptance rates are 71.8% when they use gender neutral profiles, but drop to 62.5% when their gender is identifiable", at the very least by adding the comparable numbers about the similar drop for men in the same sentence. Otherwise this will be the heading for every single news article about the study and nobody will acknowledge that the drop for men exists at all. This will happen anyway no matter what you do, but at least it won't be your fault.

8. If possible, control for your finding that women's changes are larger and less-needed and see how that affects results. If this sounds complicated, I bet you could find people here who are willing to help you.

9. Please release an anonymized version of the data; it should be okay if you delete all identifiable information.]

Nonfiction Writing Advice

People have asked me for advice on writing nonfiction online, so here are some tips:

1. Divide things into small chunks

Nobody likes walls of text. By this point most people know that you should have short, sweet paragraphs with line breaks between them. The shorter, the better. If you're ever debating whether or not to end the paragraph and add a line break, err on the side of "yes".

Once you understand this principle, you can generalize it to other aspects of your writing. For example, I stole the Last Psychiatrist's style of section breaks – bold headers saying **I.**, **II.**, **III.**, etc. Now instead of just paragraph breaks, you have two forms of break – paragraph break and section break. On some of my longest posts, including the Anti-Reactionary FAQ and Meditations on Moloch, I add a *third* level of break – in the first case, a supersection level in large fonts, in the latter, a subsection level with an underlined First, Second, etc. Again, if you're ever debating more versus fewer breaks, err on the side of "more".

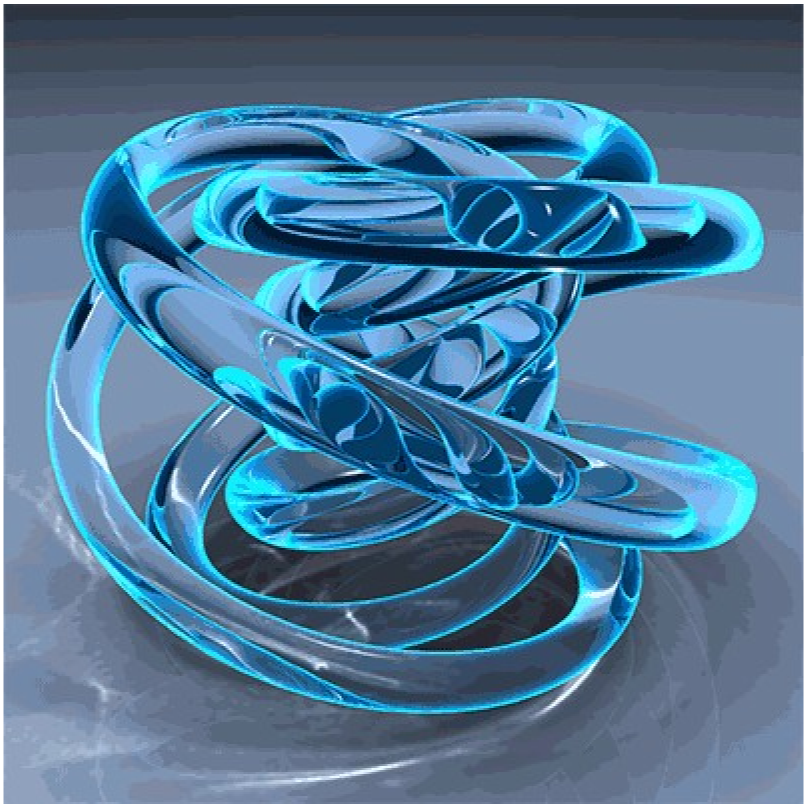
Finishing a paragraph or section gives people a micro-burst of accomplishment and reward. It helps them chunk the basic insight together and remember it for later. You want people to be going – "okay, insight, good, another insight, good, another insight, good" and then eventually you can tie all of the insights together into a high-level insight. Then you can start over, until eventually at the end you tie all of the high-level insights together. It's nice and structured and easy to work with. If they're just following a winding stream of thought wherever it’s going, it’ll take a lot more mental work and they’ll get bored and wander off.

Remember that clickbait comes from big media corporations optimizing for easy readability, and that the epitome of clickbait is the listicle. But the insight of the listicle applies even to much more sophisticated intellectual pieces – people are much happier to read a long thing if they can be tricked into thinking it’s a series of small things.

2. Variety is the spice of life

This is really closely linked to the last tip. Your brain gets bored if it has to focus on the same thing for too long. But you can get around that by making an activity look like many different things. Sometimes this is as simple and as dumb as putting Roman numeral one, Roman numeral two, etc at natural breaks in the article, and then your brain thinks “Oh, I guess there are two different things here”. But other times you actually have to vary the reading experience.

Again, the clickbaiters are our gurus – they intersperse images throughout their content. The images aren’t always very useful, they don’t always add much, but now it’s not just a wall of text. It’s a wall of text *and* images.



Watch the blue twirly thing until you forget how bored you are by this essay, then continue.

Or you can be more subtle. Break your flow. Include links, so that the never-ending stream of black text on white background is broken up with some pretty blue. If you are *very* desperate, italicize certain words to simulate the stresses of normal speech and turn the visual experience into a visual-auditory experience. Vary the form of your sentences, as per Gary Provost:

> This sentence has five words. Here are five more words. Five-word sentences are fine. But several together become monotonous. Listen to what is happening. The writing is getting boring. The sound of it drones. It’s like a stuck record. The ear demands some variety. Now listen. I vary the sentence length, and I create music. Music. The writing sings. It has a pleasant rhythm, a lilt,

> a harmony. I use short sentences. And I use sentences of medium length. And sometimes, when I am certain the reader is rested, I will engage him with a sentence of considerable length, a sentence that burns with energy and builds with all the impetus of a crescendo, the roll of the drums, the crash of the cymbals–sounds that say listen to this, it is important.

(Blockquotes are also a nice way to vary the reading experience)

But don’t just vary the appearance of your writing. Vary the tone. If you’re comfortable, shift between registers. When I was talking about SSRIs, I mentioned study after study after study – and then, around the middle, I told a kind of funny story about the time I had a job interview with the author of one of the studies. It was a complete break with the tone of the piece, which is dangerous – but my hope was that after having your mind dulled by twenty different pharmacology studies in a row, a quick first-person aside and silly story would be invigorating and give you the energy to wade through another twenty such studies.

3. Keep your flow of ideas strong

I lampshade my flow of ideas with a lot of words like “Also”, “But”, “Nevertheless”, “Relatedly”, and “So” (when I’m feeling pretentious, also “Thus”). These are the words your eighth-grade English teacher told you never to start paragraphs with. Your eighth-grade English teacher was wrong. If you’re writing three paragraphs that are three different pieces of evidence for the same conclusion that you’re going to present afterwards, make damn sure your readers know this. It could be as simple as:

> It’s pretty obvious that X is true, and we have lots of converging lines of evidence for this. Some of the best evidence comes from the field of augury. For example:
>
> First, A
>
> Second, B
>
> Third, C
>
> Now, some people say that not-A, but that’s totally wrong. It only looks like not-A, because P. Likewise, although Q might make it look like not-B, Q can’t be trusted for several other reasons, for example R. And not-C is too silly to even think about. So despite the objections you always hear, the augurical evidence for X is strong.
>
> Even more evidence comes from the field of haruspicy. All four major haruspical schools hold X as a major principle. School 1 says X because of D. School 2 says X because of E. School 3 says X because of F. And school 4 says X because of G. So although augury and haruspicy disagree on a lot, on the subject of X they are in complete accord.

Notice the underlined words holding up the structure of the argument. Not only is the argument nice and tight, but the role of each part in the whole is telegraphed beforehand. For example, the “now” that comes just after C is saying something like “Take a step back, I’m about to tell you something that might otherwise be controversial, but listen to what I have to say”. And the “likewise” just after P means something like “We just got down talking about not-A because P, here’s another argument with about the same structure”. Before any of the facts are inserted, you already know where they fit into the structure. And you’re able to abstract from the micro-level and get the bigger picture of some fact which is supported by both augury and haruspicy, which was the main point of the argument.

I overuse the world “actually” really badly. I’m trying to cut down on it, but I don’t want to stop completely. “Actually” is a great structural word. It distinguishes “Here’s how things look, here’s what’s actually true”. That sentence makes sense even without the “actually”, but I feel like the “actually” holds my readers’ hands through the process and makes the dichotomy better-defined.

Defend your flow of ideas at all costs. This might sound paradoxical after section 2, which was about how breaking flows is great. It *is* kind of paradoxical, and it *is* sort of hard to explain, but it’s the difference between “exciting” and “horrible”. Eating a foreign cuisine can be exciting because it’s so different from your usual fare; eating hot lava is even *more* different than your usual fare, but no longer fun. Play around with flow and variety, but never break the flow in a *jarring* way. And if you have to break a flow, make it the flow of your sentence, or the flow of your paragraph, but *not the flow of ideas*.

I agonize a lot about where it is versus isn’t appropriate to break the flow of ideas. Sometimes I use the really ugly solution of having an entire paragraph within parentheses, as if to say “I really wanted to bring this up here, but remember it’s not actually part of the structure of this argument!”

(this is a good meta-level example. I used the word “actually” there, and I wanted to point it out as an example of what I was talking about before, but doing that would break the flow of this whole argument about how you shouldn’t break the flow of things. So, in accordance with the prophecy, into a paragraph-long set of parentheses it goes. I'm starting to think maybe I'm not the best person to be giving writing advice…)

But sometimes you've just got to leave out an observation which would be interesting and helpful but not at that particular part of your argument. I think the phrase is "kill your darlings".

4. Learn what should and shouldn't be repeated.

A lot of the medical notes I read look like this:

> Mr. Smith presents to the ER for evaluation. He is a 24 year old man. He is complaining of chest pain. He was in the shower today when he slipped and fell. He was able to get up and make dinner. He says the chest pain started two hours later. He says has never had chest pain before. He took two aspirin. He says that did not help. He says the pain is 8/10 at this time. He says it is pretty bad, but that the pain of hearing these repetitive sentence structures is even worse.

If two sentences in a row start with the same word, it sounds unwieldy. If three or four do, it sounds bizarre. If it's a whole paragraph's worth, people start questioning their own sanity and trying to claw their eyes out.

A counterexample: what about the paragraph just above, starting with "If two sentences…"? I started with "if" three times in a row, and it didn't sound bad at all! What's up?

Deliberate use of parallelism is okay and even commendable. Usually this involves using the same structure to call attention to certain differences. You can tell if something is good parallelism by saying it aloud. When I say the paragraph above aloud, I'm using special intonation, especially in the places where the sentences differ (ie “two”, “three or four”, “whole paragraph’s worth”). Here your reader knows what you’re trying to do and it’s interesting. In the medical history example, there’s no deliberate attempt at parallelism in order to compare and contrast. You’re just doing the same thing again and again.

But it’s not just about first words of sentences. Consider something like this:

> China has the largest population of any country in the world. It also has the largest military. Because of China’s powerful military, some of its neighbors are afraid of it. China has reassured its neighbors many times that it’s peaceful, but they’re not convinced.

This sounds off to me. The repetition of “largest population” and “largest military” is done clumsily. There are a lot of ways to make it a virtuous parallelism – for example “China has the largest population – and largest military – in the world” or “China has the largest population in the world; it also boasts the largest military” – but as it is, it just sounds weird. When you come to “largest military”, there’s an immediate mental callback to “largest population”, but you’re not sure why and it’s just distracting.

Likewise, the repetition of “neighbors” is weird. It could be solved by changing the second use to “those neighbors”, which sort of telegraphs that you know you’re repeating “neighbors” and did it on purpose. Otherwise it has the same unfortunate dull-sounding cadence as the medical history.

You could also solve both those problems by just varying the structure enough that the problem goes away. For example:

> China is the most populous nation in the world. It also boasts the world’s largest military, which has provoked concern among other nations in the region. Although China has tried to reassure its neighbors of its peaceful intentions, they remain unconvinced.

This is hard and really deserves a book-length treatment. Without the book, all I can say is to realize that any repetition of words and structures will stand out to your reader, and make sure that their standing-out emphasizes your point instead of just being confusing.

5. Use microhumor

You’ve heard of microaggressions. Now try microhumor. It’s things that aren’t a *joke* in the laugh-out-loud told-by-a-comedian sense, but still put the tiniest ghost of a smile on your reader’s face while they’re skimming through them.

I learned this art from Dave Barry and Scott Adams, both of whom are humor writers and use normal macrohumor, but both of whom pepper the spaces in between jokes with microhumor besides. Your best best is to read everything they’ve written, your second best bet is to listen to me fumblingly try to explain it.

Here’s a paragraph from my “about” page:

> Topics here tend to center vaguely around this meta-philosophical idea of how people evaluate arguments for their beliefs, and especially whether this process is spectacularly broken in a way that may or may not doom us all.

There are a couple of things here that might qualify as microhumor. Take “especially whether this process is spectacularly broken in a way that may or may not doom us all”. It’s not really a *joke*. If I were a comedian and recited that sentence, you wouldn’t start laughing. But it’s kind of funny to be starting with what sounds like a pretty dry academic idea (“how people evaluate arguments for their beliefs” and whether the process is broken), and then confound expectations with an exaggerated (well, maybe) warning about it dooming us all. The phrase “may or may not doom us all” does the same thing on a smaller scale: “may or may not” is a pretty reserved, careful sounding phrase, whereas “doom us all” is obviously the opposite of reserved (I also like the similar construction “it might have sort of kind of been the worst idea ever”).

You can actually go a long way toward microhumor just with hedge words (“vaguely”, “sort of”), exaggerations (“the worst thing ever”, “doom us all”), and sometimes the combination of the two.

I think this microhumor stuff is really important, maybe the number one thing that separates really enjoyable writers from people who are technically proficient but still a chore to read. Think about it with a really simplistic behaviorist model where you keep doing things that give you little bursts of reward, and stop doing things that don’t. There are only a couple of sources of reward in reading. One of them is getting important insights. Another is hearing things that support your ingroups or bash your outgroups. And a third – maybe the biggest – is humor. Who ever had trouble slogging through a really hilarious book of jokes?

Nobody can be super funny all the time, and an article on the economic crisis filled with man-walks-into-a-bar-style jokes would be jarring and weird. But micro-humor really works. It works at a background level where people don’t notice it working, and it makes people keep coming back for more.

Humor is also disarming. It’s hard to hate somebody who’s making you laugh. I don’t mean somebody who’s making bigoted jokes that offend you, or writing political cartoons about how awful your ingroup is. Those people are easy to hate. I mean somebody who’s making *you* laugh, right now. If you can make people laugh while challenging their cherished beliefs, you’ve got a tiny bit more good will than usual.

6. Use concrete examples

Consider the following proposition:

In a study measuring whether implicit attitudes determine an outcome, you need to make sure the implicit attitudes aren’t serving as accurate proxies for underlying fundamentals.

This is the thesis of one of my more popular posts, Perceptions Of Required Ability Act As A Proxy For Actual Required Ability, but I don’t *present* it like that. Instead, I start by saying:

> Imagine a study with the following methodology. You survey a bunch of people to get their perceptions of who is a smoker (“97% of his close friends agree Bob smokes”). Then you correlate those numbers with who gets lung cancer. Your statistics program lights up like a Christmas tree with a bunch of super-strong correlations. You conclude “Perception of being a smoker causes lung cancer”, and make up a theory about how negative stereotypes of smokers cause stress which depresses the immune system. The media reports that as “Smoking Doesn’t Cause Cancer, Stereotypes Do”.

Whether or not you understood or agreed with the abstract version thesis, you (hopefully) find the problems with the nicotine example intuitively obvious. Now if I give you the principle “in a study measuring whether implicit attitudes determine an outcome, you need to make sure the implicit attitudes aren’t serving as accurate proxies for underlying fundamentals”, that principle makes sense and you will tend to agree with it. Now we can move on to harder problems, like the actual study in the post, where it’s *not* as obvious and where a lot of people thought they’d proven that the implicit attitude determined the outcome.

If you’re going to be making a complicated point, start with a concrete example. If you’re going to be making a *very* complicated point, start with a *lot* of concrete examples. When I wrote Meditations on Moloch, probably the most complicated point I’ve ever tried to express on this blog, I began with fourteen different examples before I even started trying to express the underlying principle. I hoped that readers would be able to triangulate my point by finding what all fourteen examples had in common, and most of them did.

This is related to an idea I keep stressing here, which is that people rarely have consistent meta-level principles. Instead, they’ll endorse the meta-level principle that supports their object-level beliefs at any given moment. The example I keep giving is how when the federal government was anti-gay, conservatives talked about the pressing need for federal intervention and liberals insisted on states’ rights; when the federal government became pro-gay, liberals talked about the pressing need for federal intervention and conservatives insisted on states’ rights.

So if you want to convince someone of a meta-level principle, you need to build it up from examples that support it. And if you want the principle to be well-founded and stable under reflective equilibrium, you also need to present the examples that *don't* support it and explain why you didn't make your principle out of those instead.

And if you want to convince somebody that their meta-level principle is wrong, the quickest and most effective way to do it is to show that it proves too much, then provide them with a better principle that preserves the things they want but doesn't prove things they don't want.

But my point is that all of this has to be done on the object-level, with the excursions to the meta-level level being few, far-between, and justified with extensive application to the object-level. Otherwise you're too likely to shoot off into the entirely abstract and end up sounding like Hegel:

> The good is the idea, or unity of the conception of the will with the particular will. Abstract right, well-being, the subjectivity of consciousness, and the contingency of external reality, are in their independent and separate existences superseded in this unity, although in their real essence they are contained in it and preserved. This unity is realized freedom, the absolute final cause of the world. Every stage is properly the idea, but the earlier steps contain the idea only in more abstract form. The I, as person, is already the idea, although in its most abstract guise. The good is the idea more completely determined; it is the unity of the conception of will with the particular will. It is not something abstractly right, but has a real content, whose substance constitutes both right and well-being.

Please don't end up sounding like Hegel.

And a free tip for this: use words like "me" and "you" instead of "a person" or "someone". Compare:

“If someone does the calculations with this methodology, the result will probably be nonsense.”

Versus:

“If you do the calculations with that methodology, you’ll probably end up with nonsense.”

I think the second sounds snappier and more concrete.

7. Figure out who you’re trying to convince, then use the right tribal signals

Your role model in this (and in nothing else) should be Donald Trump. Think about it. He supports Planned Parenthood, doesn’t want to cut entitlement programs, condemns Dubya and the Iraq war, supports affirmative action, supports medical marijuana, etc. If somebody were to tell you last year that a man with those policy positions would not only be leading the Republican primary, but leading even among the most conservative voters, you’d think they were crazy. The rest of the country has been trying to convince conservative Republicans to be more comfortable with those positions for decades, and we’ve failed miserably. Now Trump just waltzes in and everyone is like “Yeah, okay, sure”?

The secret of Trump’s success is that most conservative Republicans don’t really care about medical marijuana (or whatever) for its own sake. They care because opposing medical marijuana symbolizes membership in their tribe, they feel like their tribe is persecuted, they have a fierce loyalty to their tribe, and darned if they’re going to support somebody who doesn’t use the right shibboleths.

Trump throws them a bone. He says things like “illegal immigrants are rapists” that no moderate or liberal would ever say, things that would horrify them. He uses all the affectations of being working class. He may not quite prove he's "one of us", but he very effectively proves he's not Just A Typical Outgroup Member. When Trump says "Legalize medical marijuana", they don't hear "I'm yet another RINO liberal pansy who hates Christian values and wants everybody to become reefer-smoking hippies". So they only hear something boring about the regulations around pain relief medication – and who cares about those?

Trump's Law is that if you want to convince people notorious for being unconvinceable, half the battle is using the right tribal signals to sound like you're one of them.

For example, when I'm trying to convince conservatives, I veer my signaling way to the right. I started my defense of trigger warnings with "I complain a lot about the social justice movement". Then I cited Jezebel and various Ethnic Studies professors being against trigger warnings. Then I tried to argue that trigger warnings actually go together well with strong versions of freedom of speech. At this point I haven't even started arguing in favor of trigger warnings, I've just set up an unexpected terrain in which trigger warnings can be seen as a conservative thing supported by people who like free speech and don't like social justice, and opposition to trigger warnings can be seen as the sort of very liberal thing that people like Jezebel and Ethnic Studies professors support. The important thing isn't that I convince anyone that trigger warnings are really on the right – that's a tall order – but that the rightists reading my argument feel like I'm working with them rather than against them. I'm not just another leftist saying "Support trigger warnings because it's the leftist thing and you should be leftist and everyone on the right is terrible!"

My reward was seeing a bunch of hard-core anti-social-justice types trip over themselves in horror at actually being kind of convinced, which was pretty funny.

On the other hand, when I'm trying to convince feminists of something, I *start with* a trigger warning – partly because I genuinely believe it's a good idea and those posts can be triggering, but also partly because starting with a trigger warning is a tribal signal that people on the right rarely use. It means that either I'm on their side, or I'm being unusually respectful to it. In this it's a lot like Trump saying illegal immigrants are rapists – something the outgroup would never, ever do.

(And that's not just my theory – I've gotten lots of angry comments about the trigger warnings from people further right than me, saying that using them makes me an idiot or a pushover or a cuck or something. I am always happy to get these comments, because it means the signaling value of using trigger warnings remains intact.)

Crossing tribal signaling boundaries is by far the most important persuasive technique I know, besides which none of the others even deserve to be called persuasive techniques at all. But to make it work, you have to actually understand the signals, and you have to have at least an ounce of honest sympathy for the other side. You can't just be like "HELLO THERE, FELLOW LIBERALS! LET'S CREATE INTRUSIVE BIG GOVERNMENT AGENCIES TOGETHER! BUT BEFORE WE DO, I HAVE SOMETHING I WANT TO TELL YOU ABOUT THE SECOND AMENDMENT…"

Which I guess means that being able to consider both sides of an issue *sort of* gives you superpowers. That's pretty encouraging.

8. Anticipate and defuse counterarguments

Here’s something I’ve noticed. Something like:

> **Alice:** We need to invade Syria. I know that there’s always the risk of creating a Iraq-style power vacuum in these situations, but the threat from ISIS is too great.

Sounds a whole lot better than something like:

> **Alice:** We need to invade Syria.
> **Bob:** But isn’t there a risk that will create a Iraq-style power vacuum?
> **Alice:** The threat from ISIS is too great.

The second one sounds too much like Alice hadn’t really thought about the power vacuum thing, Bob called her on it, and she kind of blew him off with a tangentially related point. The first one sounds more like Alice is a careful thinker who has weighed all the risks and benefits and finally decided in favor of invasion. This is true even though Alice’s reasoning is the same in both situations.

Or what about this:

> **Alice:** We need to invade Syria. I know that there’s always the risk of creating a Iraq-style power vacuum in these situations, but the threat from ISIS is too great.
> **Bob:** I know the threat from ISIS is serious, but I’m still really worried about that power vacuum thing.

Bob sounds kind of weak here. Come on, Bob. Alice *already* raised the power vacuum issue! We’re *done with* that!

The moral of the story is that you sound a lot more credible, and your opponents a lot less persuasive, if you’re the one who brings the possible counterarguments up yourself. This is true *regardless* of how effective your countercounterarguments are.

There’s also a visibility advantage. Suppose Alice puts her argument on her blog. Bob quotes her and puts his counterargument on his blog. Maybe the readers of Bob’s blog won’t read Alice’s blog where they can see her countercounterargument. Maybe they don’t even read the comments of Bob’s blog. If Alice addressed the obvious counterarguments in her first post, Bob has been preempted from blogging about them separately, or at least has lost his low-hanging fruit and has to stretch further before he finds something he can talk about. And if he does quote Alice, all the countercounterarguments against his point will be right there for his readers.

This isn’t just good rhetorical practice, it’s good epistemic practice. A lot of Internet arguments are the same ten or twenty issues re-examined time after time after time after time. If you’re arguing in favor of gun control, you have no excuse not to realize somebody will bring up “But don’t guns save lives by helping people with self-defense?”. And in fact, if you’re arguing in favor of gun control, you had better have thought long and hard *yourself* about whether or not guns save lives through self-defense. If you’ve never considered that, you have no business having an opinion. But if you *have* considered and rejected that, then you might as well run your audience through your thought process now (and sound more convincing) now rather than wait until some pro-gun person brings it up (and be caught flat-footed like Alice in the second example).

The logical conclusion of this process is that you address all the arguments, counterarguments, and countercounterarguments in the space you’re covering, nobody can disagree with you, and you’re self-evidently right. Sometimes this takes a lot of text, but better a long argument which is accurate and convincing than a short snappy argument which might be wrong or unpersuasive. Besides, by this point you've absorbed all these other tips and hopefully write in an engaging way that makes your readers want to keep going no matter how many levels of countercountercountercounterargument you spring on them.

(Your other option here is just to put this stuff in footnotes and let your readers decide whether or not to go through them. It's very satisfying to answer somebody's stupid objection with "Actually, I think you'll find I disproved that in footnote 17.")

There's a special variant of this you need when you're in shark-infested waters, debating very controversial things with very hostile people. Here the "counterargument" is going to take the form of trying to destroy your reputation by using one of your comments, taken out of context, to prove you're a bad person with unconscionable beliefs who should never be listened to.

For example, suppose I'm trying to explain some social phenomenon and I mention that rich people get better grades in school than poor people. A hostile opponent could accuse me of making a stupid stereotype and saying that all rich people are better than all poor people; then he could condescendingly "correct" me by saying that actually the within-class differences are larger than the between-class ones. Or might say that I need to realize school grades aren't the only thing and there are much more important determinants of people's worth as a human being. Or he might accuse me of being a Social Darwinist, and "correct" me by saying that maybe this is because of the stresses of poverty.

Now, in fact I neither said nor meant any of those things. But if somebody accuses me of them, and I have to plead that I really didn't mean it that way, honest – then they can double down and say that my protests of innocence are the surest sign of my guilt. Whether they succeed or not, I'm on the defensive. We've shifted from debating whatever point I wanted to make in the first place, to debating whether Scott is a monocle-wearing Social Darwinist who believes all poor people deserve to starve on the street.

The solution is really simple: *anticipate and defuse counterarguments*. If I wanted to make the class/grades point, it would go something like this:

> According to [study], students from families earning >$100,000 score have an average high school GPA X points higher than students from families earning < $20,000. This obviously doesn't mean all rich people do better than all poor people, and it doesn't rule out causes like the stress of poverty hurting people's brains. But it does suggest a possible explanation for [whatever phenomenon I was trying to use this to explain].

A related note: when talking about controversial things to a potentially hostile audience, look through *every single sentence* of your work and imagine how it would sound if it were quoted out of context and used as a summary of who you are as a human being. If you don't, eventually someone will try this and you'll be unprepared.

9. Use strong concept handles

The idea of concept-handles is itself a concept-handle; it means a catchy phrase that sums up a complex topic.

Eliezer Yudkowsky is *really good at this*. "belief in belief", "semantic stopsigns", "applause lights", "Pascal's mugging“, “adaptation-executors vs. fitness-maximizers“, “reversed stupidity vs. intelligence“, “joy in the merely real” – all of these are interesting ideas, but more important they’re interesting ideas with short catchy names that everybody knows, so we can talk about them easily.

I have very consciously tried to emulate that when talking about ideas like trivial inconveniences, meta-contrarianism, toxoplasma, and Moloch.

I would go even further and say that this is one of the most important things a blog like this can do. I’m not too likely to discover some entirely new social phenomenon that nobody’s ever thought about before. But there are a lot of things people have vague nebulous ideas about that they can’t quite put into words. Changing those into crystal-clear ideas they can manipulate and discuss with others is a big deal.

If you figure out something interesting and very briefly cram it into somebody else’s head, don’t waste that! Give it a nice concept-handle so that they’ll remember it and be able to use it to solve other problems!

10. Recognize that applying these rules will probably start disastrously

There’s a pattern across almost all skills, where people start off doing things half-baked but sometimes with a bit of native talent. The experts teach them The Right Way To Do Things, and they switch to doing it in a stilted formulaic way that makes everybody else cringe. Eventually they become better and better. Finally, they do things that completely contradict the rules they were taught, and it works great. I think it was in the context of poetry that somebody said

“Learn the rules first, then you can break them as much as you want.”

Untrained natural writing is often bad, but at least honestly bad. Untrained writing that tries to force itself to do something because somebody told them it was a good idea is *much worse*. Think of the old adage “If you’re giving a speech, start out with a joke.” It’s great advice when done right. Now imagine all the ways it could go wrong – terrible jokes, inappropriate jokes, forced jokes, speeches that absolutely shouldn’t start off with jokes, et cetera. A speech that doesn’t start off with a joke is often good; a speech that shouldn’t start out with a joke but has been forced into doing so never is. Eventually you end up shouting “Just use your instincts!” at people who do not actually have instincts.

Use the advice in this post wrong, and you end up transforming the famous quote from the Declaration of Independence into something like:

> Although we agree King George has many good qualities, we nevertheless hold these truths to be more or less self-evident. Truth number one, that all men are created equal. For example, they should all be equally allowed to speak freely about *important* issues like taxes. It’s possible that there might be some times they shouldn’t be equal, like children having fewer rights than adults, but this are just minor exceptions. [insert picture of Liberty Bell here] Truth number two, that they are endowed by their Creator with certain inalienable rights…

Almost the only good advice in any discipline is “develop instincts, then use them”. While you’re waiting for the instincts to develop, or in order to push them along, it’s sometimes helpful to hear some other people’s advice. But *do. not. force. it*.

George Orwell ended his own list of writing advice with “Break any of these rules rather than say something outright barbarous”. As usual, George Orwell is right.

I also like Piet Hein’s commentary:

> There is
> one art,
> no more,
> no less:
> to do
> all things
> with art-
> lessness.

Two Attitudes In Psychiatry

Attitude 1 says that patients know what they want but not necessarily how to get it, and psychiatrists are there to advise them. So a patient might say “I want to stop being depressed”, and their psychiatrist might recommend them an antidepressant drug, or a therapy that works against depression. This is nice and straightforward and tends to make patients very happy.

Attitude 2 says that people are complicated. Sometimes this complexity makes them mentally ill, and sometimes it makes them come to psychiatrists and ask for help, but there’s no guarantee that the thing that they’re asking about is actually the problem. In order to solve the problem, you need to unravel the complexity, and that might involve not giving the patient what they want, or giving them things they don’t want. This is *not* straightforward and requires some justification, so let me give a few cases where Attitude 2 seems to me obviously correct.

1. A mother brings her 6 year old son to the doctor, complaining that he gets nauseous every morning. She wants the doctor to prescribe an anti-nausea pill. The doctor probes further and finds the kid only gets nauseous on school days. In fact, he only gets nauseous on school days when he has a particular gym class. The doctor asks the kid if there are any problems in that gym class, and the kid is reluctant to say anything. After a while, he finally admits there is a bully in that class. The mother calls the school, and the school takes care of the bully. After that the kid is no longer nauseous in the mornings.

2. A woman goes to a plastic surgeon asking him to fix her nose, which she insists is hideously deformed. The plastic surgeon thinks the nose looks perfectly normal and asks her to be cleared by a psychiatrist before surgery. The psychiatrist diagnoses the woman with body dysmorphic disorder, a delusional belief that one of their body parts is unbearably ugly. The psychiatrist advises the woman and her surgeon that plastic surgery does not work for this disease; if the woman gets her operation, she'll inevitably either think that the new nose is just as ugly as the old one, or she'll switch to focusing on something else like her ears or her mouth. He suggests she get psychotherapy instead. After several years of psychotherapy, the woman learns not to worry so much about her nose.

3. A woman goes to her doctor asking him how to taper off her birth control pills. The doctor is surprised at this request because he knows she is planning to break up with her boyfriend. The woman says that this is true, but she wants a child as a way to remember the relationship. The doctor probes deeper and finds the patient is very anxious and ambivalent about leaving her boyfriend and feels like if she has his child at least she will always have "a part of him" with her. The doctor refers her to therapy for her anxiety, and she is able to sort through her conflicting feelings about leaving her boyfriend. She chooses to stay on her birth control.

4. A man goes to his doctor asking for the strongest antipsychotics that exist, saying that he's crazy and he's going to hurt someone. The man converses very logically, and tells the doctor he's felt like this for a year now and never hurt anybody. The doctor suggests that he's not actually psychotic or violent, but might have an obsessive-compulsive disorder where he worries about becoming that way. The doctor recommends therapy for OCD.

5. A man comes to the psychiatric hospital saying that he’s suicidal and needs admission. The doctor knows him well, and remembers that he has been admitted five times in the past six months, each time after a life crisis, and that the patient has never actually attempted suicide and never even planned how he might do it. The doctor suggests that the man is using the psych hospital as an emotional crutch, and that instead of threatening suicide and going to the hospital whenever he is upset, he needs to learn more adaptive coping mechanisms.

Attitude 1 would have been the wrong choice in these five situations. If the doctor had just given the mother the anti-nausea pill she’d been asking for, the son’s stress about being bullied probably would have just caused some other symptoms. If the surgeon had just given the woman the nose job she wanted, she would have been dissatisfied with the surgery and wanted it changed again. If the third doctor had just told the woman how to get off birth control like she wanted, she might have had a baby for the wrong reasons and regretted it later, leading to heartache all around. If the fourth doctor had just given the man an antipsychotic, he would have unnecessarily exposed him to a potentially life-long course of very strong medication. If the fifth doctor had admitted the man to the hospital, he would be using up scarce resources and discouraging the man from learning better coping strategies.

Any halfway decent psychiatrist uses both attitudes at different times, but most people I know tend to lean to one side or the other. The 2-leaning doctors stereotype the 1-leaning doctors as simple-minded and gullible. The 1-leaning doctors stereotype the 2-leaning doctors as antirational paranoiacs with sledgehammers.

I remember a textbook talking about a case study by a famous psychiatrist. The patient had come in talking about how her husband was being borderline-emotionally-abusive to her. The psychiatrist interrupted her and said that she was perpetuating this dynamic to feed her own narcissism. The patient said this was absolutely not true and she wasn't narcissistic. The psychiatrist said she would never be able to get over her provoking-her-husband problem until she admitted the depth of her narcissism. The patient refused to keep seeing the psychiatrist after that, and the psychiatrist commented that it had been a hopeless case from the beginning – the extent of her narcissism was so great that she would never acknowledge that somebody else might know more than she did.

And the textbook was very wishy-washy about this – it acknowledged that the famous psychiatrist was brilliant and was doing the right thing in trying to confront the woman with evidence for her narcissism, but then it said that maybe he should have taken a more compassionate tone. Meanwhile, I couldn't help thinking that the famous psychiatrist was a jerk, that his only evidence the woman was narcissistic at all was a snap judgment from one or two easily misinterpretable things she said, and that call me narcissistic if you want but I wouldn't have kept attending therapy with this guy either.

That right there is the failure mode of Attitude 2; when we get out of the perfectly safe cases I mentioned above and into the more extreme versions, it starts looking a lot like making snap judgments about how all of a patient's problems reduce to a single personality flaw, and then interpreting everything about the patient in that light. Narcissism is probably the most popular, but other such flaws include "patient is regressing and wants to act like a child and have other people take care of her", "patient is just looking for attention”, and “patient is obsessive and demands complete control over everything”. The problem is, once you make one of these judgments every possible piece of data becomes further confirmation.

For example, suppose that a patient says he is having side effects on his new medication.

If you already believe the patient is a narcissist, you can dismiss the patient by saying that he wants to be special, he’s not happy on the same medication as everyone else, he’s trying to control the interaction by making you feel bad because you gave him an inferior medication. The solution is to teach the patient that he can’t always have his way by continuing the medication.

If you already believe the patient is regressing, you can dismiss the patient by saying that he’s throwing a temper tantrum, that instead of dealing with the side effects like a mature adult he wants someone else to step in and make everything magically better. The solution is to teach the patient to deal with his own problems by continuing the medication.

If you already believe that the patient is looking for attention, you can dismiss the patient by saying that they’re just trying to get the doctor’s attention by complaining. You can teach them that this is a maladptive social strategy by continuing the medication.

If you already believe that the patient is obsessive, you can dismiss the patient by saying that he’s getting all neurotic over minor side effects and has worked himself into a frenzy over perfectly ordinary minor hiccups because he can’t tolerate anxiety. The solution is to reassure the patient that everything is fine and continue the medication.

If you already believe that the patient is a witch, you can dismiss the patient by saying that they’re trying to confuse and upset you so that you will be easy prey when they try to kidnap you and sacrifice you to their lord and master, the Devil.

So it’s pretty easy to dislike 2-leaning doctors. Also, fun. Also, quite often justified. So sometimes I give in to the urge and dislike them.

The problem is, sometimes they’re right. I remember one time I had a patient who complained that Geodon was making her hallucinate. Geodon is an anti-hallucination medicine, so the chance that it makes someone hallucinate is pretty slim – but I’ve read all the usual social media posts where people complain about their evil psychiatrist who just dismisses their deeply felt pain as fakery because they had a problem that wasn’t listed in the textbook, and I didn’t want to be that guy, so I went along with it. I asked her to take some Geodon right there in my office. She swallowed the Geodon pill, and sure enough, about two minutes later she said she was starting to have all of these terrible hallucinations.

So I explained to her that oral Geodon takes at least an hour or two before a reasonable amount gets into the bloodstream, and there was no biological way that it could cause hallucinations two minutes after she took it. Then we talked about why she might be scared of the Geodon, and whether she felt any ambivalence about really wanting to get better. Eventually she agreed to try the Geodon again and didn’t hallucinate any more.

Here I felt okay because I had biological impossibility on my side. But I always wonder how many cases I’m letting slip just because my patients' stories are merely possible-but-unlikely.

Everything's a tradeoff between Type I and Type II errors. If I err too far on the side of Attitude 1, then my patients will like me and I'll never inspire a "my doctor said I was just making up my side effects for attention, and later on I got neuroleptic malignant syndrome and died!" horror story. But I will occasionally be doing the equivalent of doing plastic surgery on a body dysmorphic disorder patient, giving unnecessary and harmful medical care while ignoring the true problem.

If I err too far on the side of Attitude 2, then I always get to feel like a hard-headed non-gullible investigator digging down to the root of the problem – but occasionally I'll end up like that famous psychiatrist in the textbook and tell people that the reason their foot hurts is because they're narcissistic, and it has nothing to do with the fact that they stepped on a nail and the only reason they're even bringing up the nail is their deep-seated narcissism.

I tend to lean *way* toward Attitude 1. I'm not sure I can justify it. Part of it is my personality: conflict scares me and I want to be liked. Part of it is that I read too many horror stories on social media about how much patients hate their Attitude 2 psychiatrists. Part of it is that Attitude 2 has a lot of its philosophical grounding in Freud, and I really don't trust Freud.

This is a lucrative attitude nowadays. We are all supposed to be biological psychiatrists, all the old psycho-babble is no longer covered by insurance, and The Customer Is Always Right. I am lucky insofar as my natural tendency is also the socially more acceptable one.

(I suppose an Attitude 2 psychiatrist would say I’m not lucky at all, and that my unconscious desire for social approval and success has led me to adopt Attitude 1, and also I am a narcissist)

I may or may not be a narcissist, but I am definitely neurotic. And when my neurosis gets to “maybe I’m a terrible psychiatrist”, this is what it usually settles upon to worry about. Attitude 2 and the various arts associated with it are opaque to me. I can pass tests on them when I have to, but I don’t feel them in my bones. When I’m with a whole conference of doctors nodding their head and going “Yup, that guy’s a narcissist”, I’m always panicking, thinking “Wait, I’m not even close to convinced he’s a narcissist, and also nobody really knows how to treat narcissism, and I would feel a lot more comfortable if this conversation would shift to comparing and contrasting the various subtypes of dopamine receptors.” I am bad at it, and what’s worse I don’t even know if I *should be* better at it, and I don’t know how to solve the bad-at-it part without worrying that I’m sending myself and my patients down a blind alley.

2016 Nootropics Survey Results

[Disclaimer: Nothing here should be taken to endorse using illegal or dangerous substances. This was a quick informal survey and you should not make any important health decisions based on it. Talk to your doctor before trying anything.]

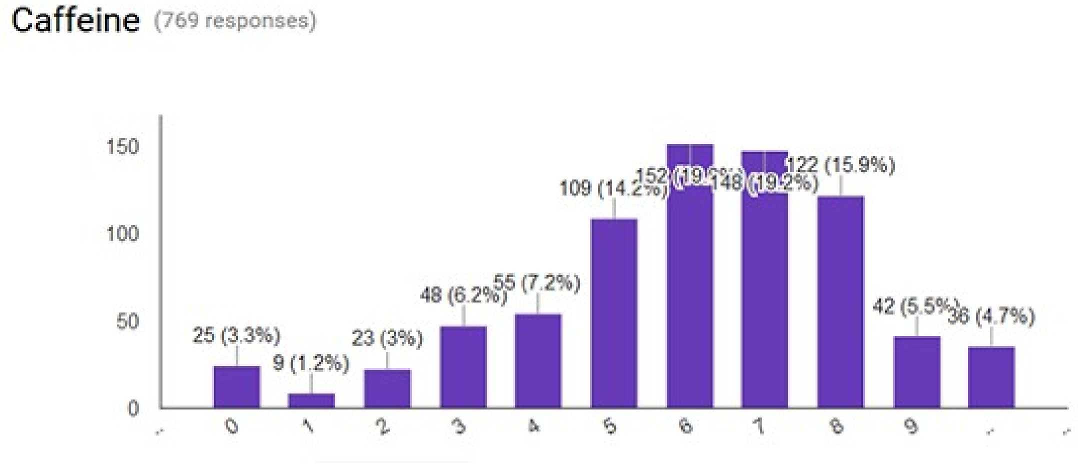
Nootropics are traditionally defined as substances that improve mental function. In practice they usually refer to psychoactive chemicals that are neither recreational drugs like cocaine and heroin, nor officially-endorsed psychiatric drugs like Prozac or Risperdal. Most are natural supplements, foreign medications available in US without prescription, or experimental compounds. They promise various benefits including clearer thinking, better concentration, improved mood, et cetera. You can read more about them here.

Although a few have been tested formally in small trials, many are known to work only based on anecdote and word of mouth. There are some online communities like r/nootropics where people get together, discuss them, and compare results. I've hung out there for a while, and two years ago, in order to satisfy my own curiosity about which of these were most worth looking into, I got 150 people to answer a short questionnaire about their experiences with different drugs.

Since then the field has changed and I wanted to get updated data. This year 850 (!) people agreed to fill out my questionnaire and rate various nootropics on a scale of 0 – 10 – thanks again to everyone who completed the survey.

Before the results themselves, a few comments.

Last time around I complained about noisy results. This year the sample size was five times larger and the results were less noisy. Here’s an example: the ratings for caffeine form a beautiful bell curve:



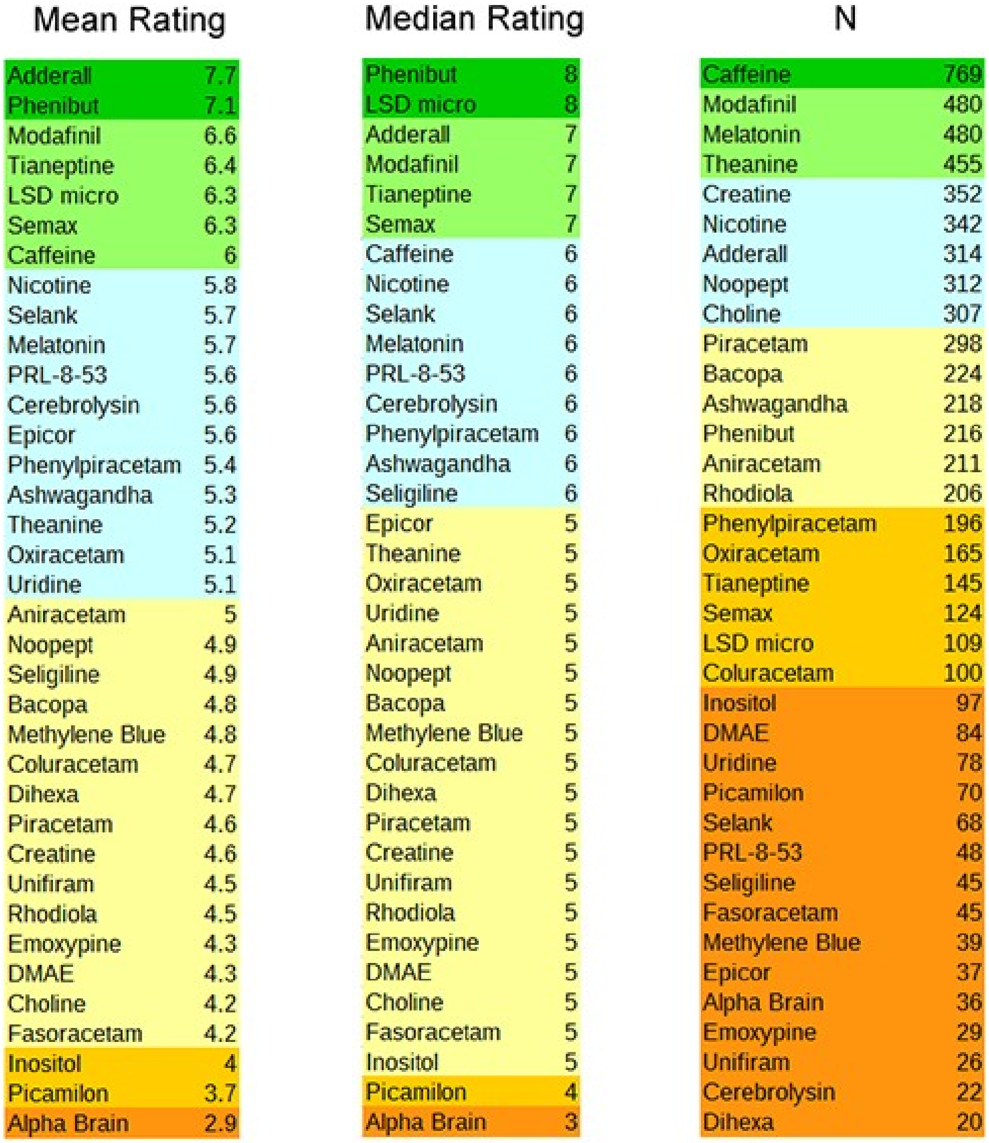
Even better, even though this survey was 80% new people, when it asked the same questions as last year’s the results were quite similar – they correlated at r = 0.76, about what you’d get from making students take the same test twice. Whatever’s producing these effects is pretty stable.

A possible objection – since this survey didn’t have placebo control, might all the results be placebo? Yes. But one check on this is that the different nootropics controlled against one another. If we believe that picamilon (rated 3.7) is a placebo, this suggests that PRL-8-53 (rated 5.6) does 19 percentage points points better than placebo.

But might this be confounded by lack of blinding? Yes. That is, if companies have really hyped PRL-8-53, and it comes in special packaging, and it just generally *looks* cooler than picamilon, maybe that would give it a stronger placebo effect.

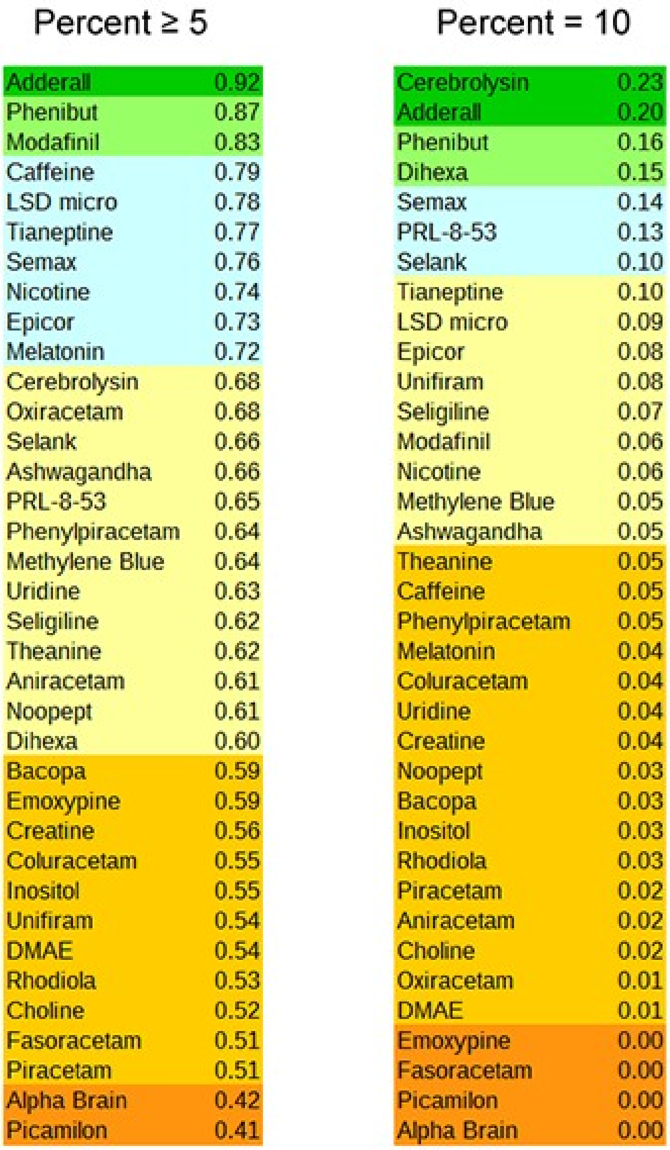
Against this hypothesis I can only plead big differences between superficially similar drugs. For example, rhodiola and ashwagandha are both about equally popular. They’re both usually sold by the same companies in the same packaging. They're both classified as "adaptogens" by the people who classify these sorts of things. But ashwagandha outperforms rhodiola by 0.9 points, which in a paired-samples t-test is significant at the p = 0.03 level. While you can always find some kind of difference in advertising or word-of-mouth that could conceivably have caused a placebo effect, there are at least some reasons to think something's going on here.

Without further ado, here's what I found:



Mean Rating		Median Rating		N	
Adderall	7.7	Phenibut	8	Caffeine	769
Phenibut	7.1	LSD micro	8	Modafinil	480
Modafinil	6.6	Adderall	7	Melatonin	480
Tianeptine	6.4	Modafinil	7	Theanine	455
LSD micro	6.3	Tianeptine	7	Creatine	352
Semax	6.3	Semax	7	Nicotine	342
Caffeine	6	Caffeine	6	Adderall	314
Nicotine	5.8	Nicotine	6	Noopept	312
Selank	5.7	Selank	6	Choline	307
Melatonin	5.7	Melatonin	6	Piracetam	298
PRL-8-53	5.6	PRL-8-53	6	Bacopa	224
Cerebrolysin	5.6	Cerebrolysin	6	Ashwagandha	218
Epicor	5.6	Phenylpiracetam	6	Phenibut	216
Phenylpiracetam	5.4	Ashwagandha	6	Aniracetam	211
Ashwagandha	5.3	Seligiline	6	Rhodiola	206
Theanine	5.2	Epicor	5	Phenylpiracetam	196
Oxiracetam	5.1	Theanine	5	Oxiracetam	165
Uridine	5.1	Oxiracetam	5	Tianeptine	145
Aniracetam	5	Uridine	5	Semax	124
Noopept	4.9	Aniracetam	5	LSD micro	109
Seligiline	4.9	Noopept	5	Coluracetam	100
Bacopa	4.8	Bacopa	5	Inositol	97
Methylene Blue	4.8	Methylene Blue	5	DMAE	84
Coluracetam	4.7	Coluracetam	5	Uridine	78
Dihexa	4.7	Dihexa	5	Picamilon	70
Piracetam	4.6	Piracetam	5	Selank	68
Creatine	4.6	Creatine	5	PRL-8-53	48
Unifiram	4.5	Unifiram	5	Seligiline	45
Rhodiola	4.5	Rhodiola	5	Fasoracetam	45
Emoxypine	4.3	Emoxypine	5	Methylene Blue	39
DMAE	4.3	DMAE	5	Epicor	37
Choline	4.2	Choline	5	Alpha Brain	36
Fasoracetam	4.2	Fasoracetam	5	Emoxypine	29
Inositol	4	Inositol	5	Unifiram	26
Picamilon	3.7	Picamilon	4	Cerebrolysin	22
Alpha Brain	2.9	Alpha Brain	3	Dihexa	20

Median rating, mean rating, and sample size for each nootropic. You can find more information on the individual substances [here](#)



Percent ≥ 5	
Adderall	0.92
Phenibut	0.87
Modafinil	0.83
Caffeine	0.79
LSD micro	0.78
Tianeptine	0.77
Semax	0.76
Nicotine	0.74
Epicor	0.73
Melatonin	0.72
Cerebrolysin	0.68
Oxiracetam	0.68
Selank	0.66
Ashwagandha	0.66
PRL-8-53	0.65
Phenylpiracetam	0.64
Methylene Blue	0.64
Uridine	0.63
Seligiline	0.62
Theanine	0.62
Aniracetam	0.61
Noopept	0.61
Dihexa	0.60
Bacopa	0.59
Emoxypine	0.59
Creatine	0.56
Coluracetam	0.55
Inositol	0.55
Unifiram	0.54
DMAE	0.54
Rhodiola	0.53
Choline	0.52
Fasoracetam	0.51
Piracetam	0.51
Alpha Brain	0.42
Picamilon	0.41

Percent = 10	
Cerebrolysin	0.23
Adderall	0.20
Phenibut	0.16
Dihexa	0.15
Semax	0.14
PRL-8-53	0.13
Selank	0.10
Tianeptine	0.10
LSD micro	0.09
Epicor	0.08
Unifiram	0.08
Seligiline	0.07
Modafinil	0.06
Nicotine	0.06
Methylene Blue	0.05
Ashwagandha	0.05
Theanine	0.05
Caffeine	0.05
Phenylpiracetam	0.05
Melatonin	0.04
Coluracetam	0.04
Uridine	0.04
Creatine	0.04
Noopept	0.03
Bacopa	0.03
Inositol	0.03
Rhodiola	0.03
Piracetam	0.02
Aniracetam	0.02
Choline	0.02
Oxiracetam	0.01
DMAE	0.01
Emoxypine	0.00
Fasoracetam	0.00
Picamilon	0.00
Alpha Brain	0.00

Percent of responders who rated each nootropic at least five (the middle rating) or ten (the highest rating)

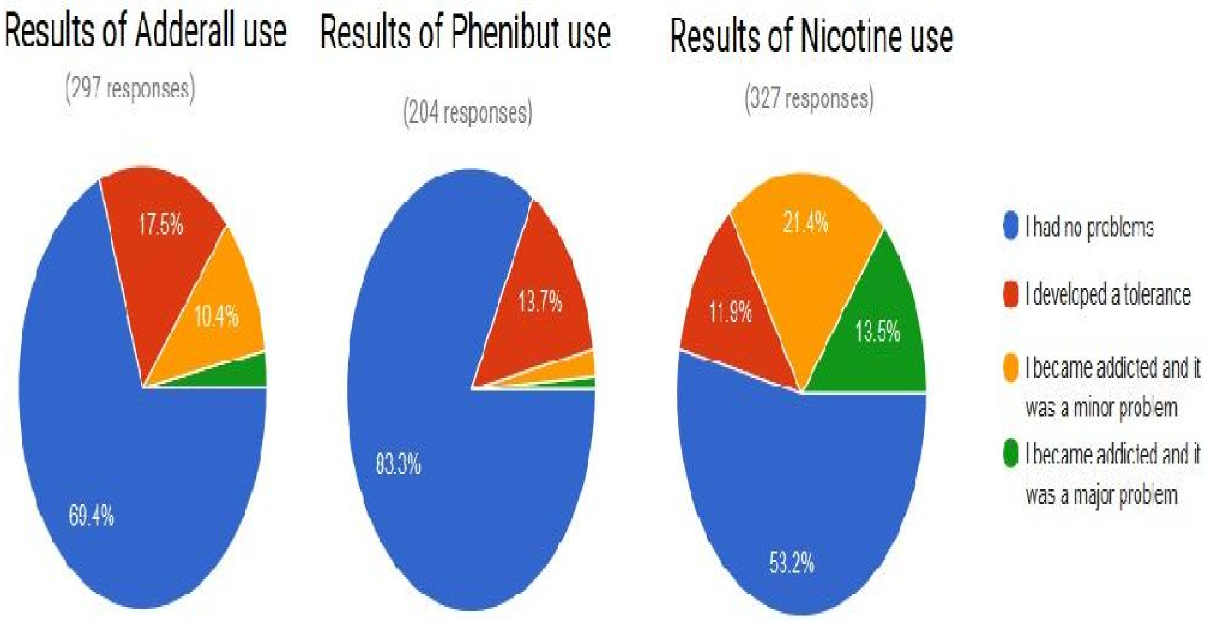
Some very predictable winners: Adderall is a prescription drug and probably doesn’t even qualify as a nootropic; I included it as a reference point, and it unsurprisingly did very well. LSD microdosing is the practice of taking LSD at one-tenth or less of the normal hallucinogenic dose; users say that it improves creativity and happiness without any of the typical craziness. Phenibut is a Russian anxiolytic drug of undenied effectiveness which is sort of notorious for building tolerance and addiction if used incorrectly. And modafinil is a prescription medication for sleep issues which makes users more awake and energetic. All of these are undeniably effective – but all are either addictive, illegal without prescription, or both.

I'm more interested by a second tier of winners, including tianeptine, Semax, and ashwagandha. Tianeptine is a French antidepressant available (legally? kind of a gray area) without prescription in the US; users say it both provides a quick fix for depression and makes them happier and more energetic in general. Semax is a Russian peptide supposed to improve mental clarity and general well-being. Ashwagandha might seem weird to include here since it's all the way down at #15, but a lot of the ones above it had low sample size or were things like caffeine that everyone already knows about, and its high position surprised *me*. It's an old Indian herb that's supposed to treat anxiety.

The biggest loser here is Alpha Brain, a proprietary supplement sold by a flashy-looking company for $35 a bottle. Many people including myself have previously been skeptical that they can be doing much given how many random things they throw into one little pill. But it looks like AlphaBrain underperformed even the nootropics that I think of as likely placebo – things like choline and DMAE. It's possible that survey respondents penalized the company for commercializing what is otherwise a pretty un-branded space, ranking it lower than they otherwise might have to avoid endorsing that kind of thing.

(I was surprised to see picamilon, a Russian modification of the important neurotransmitter GABA, doing so badly. I thought it was pretty well-respected in the community. As far as I can tell, this one is just genuinely bad.)

Finally, a note on addiction.



Adderall, phenibut, and nicotine have all raised concern about possible addictive potential. I wanted to learn a little bit about people’s experiences here, so I asked a few questions about how often people were taking things at what dose and whether they got addicted or not.

In retrospect, these were poorly phrased and didn’t get me the data I wanted. When people said they were taking Adderall every day and got addicted, I didn’t know whether they meant they became addicted because they were using it every day, or that they were using it every day because they were addicted. People gave some really weird answers here and I’m not sure how seriously I can take them. Moving on anyway:

A bit under 15% of users got addicted to Adderall. The conventional wisdom says “recreational users” are more likely to get addicted than people who take it for a psychiatric condition with a doctor’s prescription. There was no sign of this; people who took it legally and people who took it for ADHD were actually much *more* likely to get addicted than people who described themselves as illegal or recreational users. In retrospect this isn’t surprising; typical psychiatric use is every day; typical recreational use is once in a while.

Only 3% of users got addicted to phenibut. This came as a big surprise to me given the caution most people show about this substance. Both of the two people who reported major addictions were using it daily at doses > 2g. The four people who reported minor addictions were less consistent, and some people gave confusing answers like that they had never used it more than once a month but still considered themselves “addicted”. People were more likely to report tolerance with more frequent use; of those who used it monthly or less, only 6% developed tolerance; of those who used it several times per month, 13%; of those who used it several times per week, 18%; of those who used it daily, 36%.

Then there was nicotine. About 35% of users reported becoming addicted, but this was heavily dependent upon variety of nicotine. Among users who smoked normal tobacco cigarettes, 65% reported addiction. Among those who smoked e-cigarettes, only 25% reported addiction (and again, since there’s no time data, it’s possible these people switched to e-cigarettes *because* they were addicted and not vice versa). Among users of nicotine gum and lozenges, only

7% reported addiction, and only 1% reported major addiction. Although cigarettes are a known gigantic health/addiction risk, the nootropic community’s use of isolated nicotine as a stimulant seems from this survey (subject to the above caveat) to be comparatively but not completely safe.

I asked people to name their favorite nootropic *not* on the list. The three most popular answers were ALCAR, pramiracetam, and Ritalin. ALCAR and pramiracetam were on last year’s survey and ended up around the middle. Ritalin is no doubt very effective in much the same way Adderall is very effective – and equally illegal without a prescription.

People also gave their personal stacks and their comments; you can find them in the raw data (.xlsx, .csv) or the fixed-up data (.csv, notes). If you find anything else interesting in there, please post it in the comments here and I’ll add a link to it in this post.

EDIT: Jacobian adjusts for user bias

Against Interminable Arguments

[Epistemic status: something I've been thinking about recently. There's a lot of complication around these issues and this is more to start a discussion than to present any settled solution]

There's a scene in Fiddler on the Roof where Tevye is describing his peaceful little town. He says they never fight – except that one time about a horse some people thought was a mule. Someone interrupts him to say it was really a mule some people thought was a horse, and then everyone in town starts shouting "MULE!" or "HORSE!" at each other until they get drowned out by the chorus.

The town is happy and peaceful as long as nobody brings up the horse/mule thing. As soon as somebody brings it up all of the old rancor instantly resurfaces and everybody's at each other's throats. And the argument itself never gets more sophisticated than people yelling "HORSE!" or "MULE!" at each other. Maybe it would be worth it to create a norm around never bringing it up?

The rationalist/EA/etc community has a norm that people must be able to defend their beliefs with evidence, and a further norm that people shouldn't be confident in their beliefs unless they've sounded them off others and sought out potential counterarguments. These are great norms. But their failure mode is a community where dredging up interminable horse/mule style arguments is seen as a virtue, and avoiding them is seen as a cowardly refusal to expose one's own beliefs to challenge.

I've been thinking about this in the context of some arguments that keep cropping up on rationalist Tumblr. These have gotten so repetitive and annoying that I made a joke argument calendar to shame us for it; some of the people who replied were less able to see the humor in it and thought we really *should* be ashamed.

After some thought, I agree. Avoiding interminable arguments is an important social engineering problem we're really bad at. Part of it is that we need a way to distinguish the baby from the bathwater. What does it mean to seek out productive discussion while avoiding interminable arguments?

A while back I wrote about memetic evolution for controversy. If an idea is outrageous, it's likely to spread as people condemn it. if something is *controversially* outrageous, it's likely to spread even further, as people argue against it and then other people counterargue against the arguments ad infinitum. This gives it visibility far beyond things that are objectively more important. The entire news media freaks out over BernieBros for weeks, and nobody ever hears about desalinization policy in drought-stricken nations.

Interminable arguments are the local version of the same process. For me, the biggest difference between a productive discussion and an interminable argument is simple. I'm participating in the productive discussion because I want to. I'm participating in the interminable argument because I have to.

I mean, obviously nobody's pointing a gun to my head and forcing me into any arguments. But there are a lot of reasons I might feel obligated to debate something I really don't have time for.

First, there's the feeling XKCD describes as "someone is wrong on the Internet". When somebody makes a deeply flawed argument against a position you hold dear, and everyone else seems to believe them, it creates (at least for me) this weird irresistable urge to correct them.

Second, sometimes people are jerks. Nobody on Tumblr can just say “I don’t think AI is a big problem”. They have to say “I don’t think AI is a big problem, and the only reason some people worry about it is because they’re cultish sci-fi dorks who are too brainwashed to look up what real scientists have to say”. Nobody on Tumblr can just say they think feminism is important; they have to post comics like this. This is hard to just leave be, especially when it’s not just yourself but your friends who are being insulted, and *especially* when a lot of people are vocally agreeing with them and even some people you think should know better are being convinced. Letting jerks have the last word is really hard.

Third, sometimes there are actually things at stake. I’ve written before on how one of the main reasons I get defensive is because I think some groups actively strategize to push their opponents out of the Overton Window and turn them into despised laughingstocks. When it works, it means I either have to be a despised laughingstock or spend way too much mental energy hiding my true opinions. The alternative to letting these people have the final say is defending one’s self.

When I don’t want to argue but feel forced into it, I’m doing a very different thing than when I’m having a voluntary productive discussion. I’m a lot less likely to change my views or admit subtlety, because that contradicts my whole point in having the argument. And I’m a lot more likely to be hostile, because hostility is about making other people feel bad and disincentivizing their behavior, and in this case I really do believe their behavior needs disincentivizing.

But of course this just starts the cycle where people who disagree with me – both the original people I'm arguing against, and bystanders who just happen to hear – feel forced to write nasty replies of their own. And so on. It only dies down once everybody can tell themselves that they put enough effort into self-defense to acquit themselves well. And as soon as somebody challenges that – like Tevye in the Fiddler song – the whole thing starts up again as bad as ever.

So how does a community prevent this? Blocking jerks – the people who start the whole cycle by deliberately trolling others – is an obvious good start. But some people on the Tumblr discussion have mentioned some more subtle points worth thinking about.

First, the influx of newbies is a big driver of this dynamic. Newbies are less likely to know the relevant arguments, won't be bored of them yet, won't want to steer clear of them, and may mistake somebody's unwillingness to engage for the 9000th time as unwillingness to engage at all. People should be more tolerant of newbies, and newbies should be more tolerant of "look in the archives for the last time we discussed this issue, but *seriously*, don't start up about this again". This is what I think social justice people mean when they talk about "this is not a 101 space".

Second, be aware that some problems with interminable arguments might be asymmetrical. I've heard this most often in social justice contexts. For example, cis people might never have discussed trans issues much before and might find them really interesting and not particularly defensiveness-inducing. But consider what happens if there are a hundred times more cis people than trans people, plus trans people have spent a hundred times longer thinking about it, plus trans people have a lot less tolerance before they get annoyed. A cis person might innocently ask “Hey, isn’t using chromosomal sex as a proxy for gender a pretty elegant system which is much easier than all of this stuff about different identities?” because they really want to know and their local trans person seems like a good person to debate with. Meanwhile, the trans person might have had this exact debate two thousand times, find it personally insulting, but not know how to disengage politely without giving the impression that they’re too much of an intellectual lightweight to answer the simplest of questions about something very important to them.

(This isn’t limited to social justice or identity politics. I feel this same way sometimes on relatively transhumanist-free parts of the Internet. Someone will make a well-intentioned attempt to start a discussion, like “But what you people don’t get is that the AI will be smart enough to realize that paper clips are silly and compassion towards all living things is the best goal,” and I’ll want to say something like “Look, I *promise* that in fifty years of thinking about machine ethics somebody else has raised that point before, I think you’re so confused about things that correcting you wouldn’t be a very good use of my time, but if you google anything at all in this general area you’ll probably find an answer to your question.” But that would just be the “it’s not my job to educate you” or “read the Sequences” answer which so many people find annoying.)

But third, as I said before, make it absolutely super-duper crystal clear that there is *not* a community norm that everybody has to defend their positions every time they are asked, do *not* say anything like “I’ve never heard you response to Point X so now I’m going to assume that you have no argument against it and are just a brainwashed cultist taking your position on faith” and emphasize that although everybody who wants to be accurate should discuss and challenge their beliefs, nobody should have to do it to anyone else’s schedule.

The exact way this works is something I’m still working on. More and more I’m abandoning the idea of debating on social media/comments/forums entirely, and switching either to email or longer forms. Email is private and removes the performative factor; I can’t say how many times a previously terrible discussion has become manageable and productive as soon as it gets outside the public eye. And by longer forms, I mean things like books and (really good long form) blog posts. I would much rather read the best book by someone I disagree with, and hear all their best arguments laid out by a leading intellectual with a good editor, rather than have to listen to somebody taunt me on Tumblr. And if I don’t understand something about the book, or I still have questions, then I can pick *one or two* people I know and debate it with them *privately*.

Part of me thinks this is another point in favor of niceness. A community made mostly of nice people can probably hold more productive debates and have fewer interminable arguments than one that’s not as good at civility. On the other hand, I see these problems even among nice people. I think that the skill of structuring a point such that nobody feels attacked by it is complicated in ways beyond regular niceness, and that otherwise nice communities can sometimes make the right choice in deciding to avoid the whole issue.

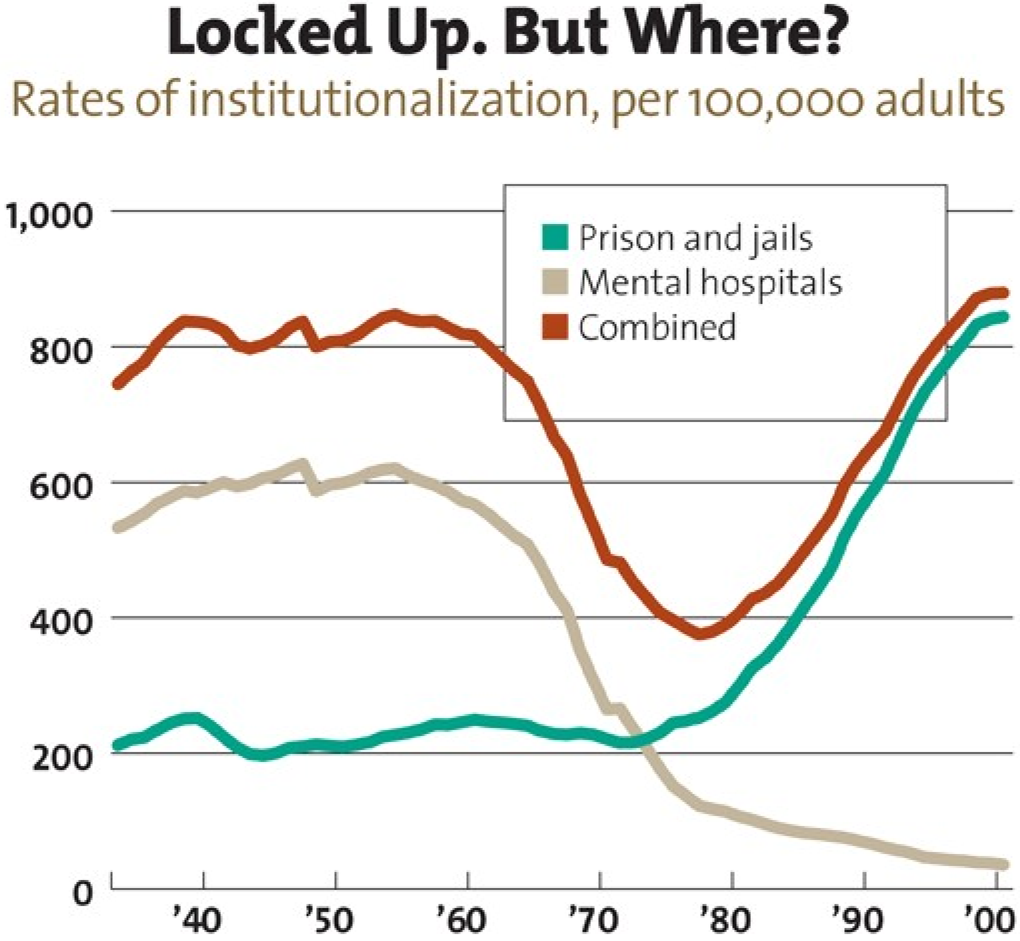
EDIT: comment

Reverse Voxsplaining: Prison and Mental Illness

I.

German Lopez of Vox writes that “America’s criminal justice system has in many ways become a substitute for the US’ largely gutted mental health system”.

He says that starting in the 1970s the US “began locking up a lot more people”, and “at the same time, the country pulled back and defunded its public mental health system”. He admits that “this wasn’t, at the time, totally malicious”, but then says it “left the criminal justice system as the only system that can respond to people with mental illness.”



He concludes that as a result, “the number of people with mental illness in prisons/jails outnumber those in state hospitals 10 to 1.” The apparent (though unstated) conclusion is that defunding the big state mental hospitals was a mistake and we need to bring them back so that the mentally ill in state hospitals once more outnumber those in prison.

Lopez seems to be working off a model where there is a population of mentally ill people who can’t make it in normal society, and so will inevitably end up either in a long-term mental hospital or a prison. Since mental hospitals are good places where people get treatment, and prisons are bad places where people get punishment, we should “catch” these mentally ill people before they end up in prison so that they can be in nice hospitals instead.

Needless to say I disagree with pretty much every part of this assessment.

II.

Between all of this talk of “the tragic collapse of America’s public mental health system” and “the US’s largely gutted mental health system” and “the country pulled back and defunded its mental health system” and so on, you might get the impression that less money is being spent on mental health. This is not really true. The share of GDP devoted to mental health is the same as it was in 1971, although this looks worse if you compare it to rising costs in other areas of health care. There hasn’t been a “gutting of the mental health system”, there’s been a shift from long-term state-run mental hospitals to community care. It hasn’t “left the criminal justice system as the only system that can respond to people with mental illness”, it helped create an alternate and less restrictive system of outpatient psychiatry. In my opinion, this was a positive development, and the share of mentally ill people in prison is not an argument against it. Let me explain.

“Mentally ill people in prison” conjures up this lurid image of psychos who snap and kill their families, followed by “well, what did you expect leaving a person like that on the street?” The reality is more mundane. There are lots of mentally ill people in prison because there are lots of mentally ill people everywhere. Remember, 20% of the population qualifies as mentally ill in one sense or another. If a depressed guy sells some marijuana and gets caught, he is now a “mentally ill person in prison”.

There are *disproportionately* many mentally ill people in prison partly because people's illnesses lead them to commit crimes, but mostly because some of the factors correlated with mental illness are the same factors correlated with criminality. Poverty? Check. Neighborhood effects? Check. Genetic load? Check. Education? Check. IQ? Check. Broken families? Check. Drug abuse? *Definitely* check. The factors that gave that pot dealer depression might be the same factors that drove him to sell pot instead of becoming an astronaut. Treating the depression might help a little, but it's not guaranteed to keep him on the good side of the law.

In *my* model, the overwhelming majority of mentally ill people can live okay lives outside of any institution, hopefully receiving community care if they want it. If they commit crimes they will go to prison just like anyone else; if not, we should hardly be clamoring to bring back the often-horrifying state-run mental hospitals and lock them up there.

So when we talk about the number of mentally ill people in prison, we should be trying to distinguish between Lopez's model and mine. That means asking: exactly how mentally ill are we talking about here?

III.

Lopez's source for the claim that "ten times more mentally ill people are in prisons than hospitals" is a report by the Treatment Advocacy Center – note the less-than-neutral name. Where Lopez uses the phrase "mental illness", TAC uses the phrase "severe mental illness" and defines it in two ways. For people in state prisons, they define it as reporting at least one psychotic symptom, and say 15% of people met their criteria. For people in county jails, they define it as meeting criteria for a depressive, bipolar, or psychotic illness, and say 15% of people met their criteria (they later arbitrarily increase that number to 20% because they feel like the survey might have undercounted).

No no NO. First, “psychotic” is not the same thing as “severely mentally ill”. Some people are severe but not psychotic – for example, a suicidally depressed person. Others are psychotic but not severe – for example, someone who hears a voice whispering her name but shrugs it off. Describing a survey that shows 15% of people as admitting one symptom of psychosis as showing 15% of people are severely mentally ill is *really* sketchy.

The prison survey provides a perfect example. It looks like the prisoners were asked fixed questions about their symptoms, and I *think* the exact screening instrument was just this survey, which has four relevant questions: “Can anybody else control your brain or thoughts?”, “Do you ever hear voices other people don’t hear?”, “Do you ever see something that other people tell you isn’t real?” and “Do you ever think anyone (other than correctional staff) is spying on you or plotting against you?”

Unfortunately, these kinds of surveys are really weak. I’m doing a study about this now, so maybe later I can cite myself on this, but the gist is that a lot of short mental health screening questions get false positives from perfectly healthy people. For example, I can’t tell you how many patients I’ve asked “Do you ever feel like anyone is spying on you?”, they say “Yes”, I ask “Who?” and they say “The NSA on my Internet activity”. Well, good work keeping up with the news. But a survey with a checkbox and no followup questions diagnoses that person as psychotic (see also: Lizardman’s Constant). This prison questionnaire was smart enough to exclude prison guards, who are *certainly* spying on all the respondents, but even beyond that I feel like the criminal lifestyle really does involve being spied on and plotted against a lot. At the very least it gives you lots of opportunities to legitimately *worry* about it.

(also, the diagnostic criteria for psychotic disorder are very clear that paranoia experienced while taking drugs doesn't count. 75% of prisoners admit to using marijuana, marijuana can *totally* make you paranoid, and as far as I can tell the survey did not specify that the paranoia had to be while sober.)

(also also, Scientific American says that about 5-15% of perfectly ordinary people hear voices. Meanwhile, 4-6% of prisoners in the survey admitted to it.)

When the survey says that X% of prisoners have felt plotted-against or heard voices in the last year, does that mean X% of them are psychotic? That X% of them are "severely mentally ill"? That the old state mental hospitals need to be re-opened so X% of them can be locked up there for being too crazy for society? I don't think it means any of those things.

But this is the *stricter* of the two criteria that the survey uses! The other one counts depressed people, bipolar people, *and* psychotic people. I don't want to trivialize non-psychotic illnesses like depression. But remember: about 10% of the ordinary non-prison population is depressed/bipolar/psychotic. Also, going to prison is depressing as heck in and of itself. When they say that 15% of people in county jails (rounded up to 20%) are severely psychiatrically ill, they're talking about pretty normal people, who might be in prison for something unrelated to their mental illness and might not even have become mentally ill at all if they hadn't been incarcerated.

So I don’t think this survey shows the majority of the mentally ill prison population is in need of institutionalization. Yes, ten times more mentally ill people are in prison than in state mental hospitals, but consider the base rates! The prison population is *huge*. The population of people who need to be committed to mental hospitals 24-7 is *tiny*. Even if mentally ill people committed crimes at exactly the average population rate, there would still be far more mentally ill people in prison than in psychiatric hospitals, just by base rates! Especially if you use as broad a definition of “mentally ill” as these people!

So when Vox says that ten times more mentally ill people are in prison than in psychiatric hospitals, I will shoot right back at them that ten times more mentally ill people are in the Los Angeles metropolitan area than in state mental hospitals. You want more meaningless statistics? Ten times more mentally ill people are in the Southern Baptist Church than in state mental hospitals! Ten times more mentally ill people watched the last season of *Game of Thrones* than are in state mental hospitals! We can’t and shouldn’t aim to institutionalize all of them.

IV.

What about that graph? It’s very suggestive. You see a sudden drop in the number of people in state mental hospitals. Then you see a corresponding sudden rise in the number of people in prison. It looks like there’s some sort of Law Of Conservation Of Institutionalization. Coincidence?

Yes. Absolutely. It is 100% a coincidence. Studies show that the majority of people let out of institutions during the deinstitutionalization process were not violent and that the rate of violent crime committed by the mentally ill did not change with deinstitutionalization. Even if we take the “15%

of inmates are severely mentally ill" factoid at face value, that would mean that the severely mentally ill could explain at most 15%-ish of the big jump in prison population in the 1980s. The big jump in prison population in the 1980s was caused by the drug war and by people Getting Tough On Crime. Stop dragging the mentally ill into this.

Lopez himself wrote a nice piece on how most mentally ill people are not violent, and another nice piece on how most people in prison are there for violent offenses. But put these together, and you get that most mentally ill people do not end up in prison. Most of the people who got out of the mental hospitals during deinstitutionalization are getting by. Some of them are homeless, and that's bad. But if you want to solve homelessness among the mentally ill, build homeless shelters, not state-run long-term mental hospitals.

V.

In case you haven't noticed, I really don't like state-run long-term mental hospitals. There is a really amazingly great thing about prison, which is that you don't go there unless you're convicted of a crime. Mental hospitals do not have that advantage. The commitment process kind of sucks, and I am saying this as a person who makes commitment decisions myself. A lot of times it degenerates into a ritualized method of avoiding lawsuits without much concern for benevolence or patient autonomy. It helps me sleep at night to know that most commitments only last a couple of days or a week at most. Long-term state-run mental hospitals didn't work that way. Remember that some of the perfectly sane people in the Rosenhan experiment were kept locked up for fifty days just for saying they heard a voice once but now they're better.

I think long-term state-run mental hospitals are better than prison, but not by very much. The Rosenhan participants described it as:

> …an overwhelming sense of dehumanization, severe invasion of privacy, and boredom while hospitalized. Their possessions were searched randomly, and they were sometimes observed while using the toilet. They reported that though the staff seemed to be well-meaning, they generally objectified and dehumanized the patients, often discussing patients at length in their presence as though they were not there, and avoiding direct interaction with patients except as strictly necessary to perform official duties. Some attendants were prone to verbal and physical abuse of patients when other staff were not present. A group of bored patients waiting outside the cafeteria for lunch early were said by a doctor to his students to be experiencing "oral-acquisitive" psychiatric symptoms. Contact with doctors averaged 6.8 minutes per day.

The idea of potentially saving a couple of people from prison by pre-emptively committing way more people to a mental hospital does not appeal to me at all, and I still think closing the institutions was the best thing Reagan ever did.

But prison and institutions aren't the only two options! There's a six month waiting period for psychiatrists in most parts of the country. The existing mental hospitals – which are different from and often nicer than the old state-run institutions – are constantly turning away people who *want* to be there because they don't have enough beds for them. There are a bunch of patients who are having trouble affording their medications. There are special treatment options like day clinics, partial hospital programs, recreational therapy, occupational therapy, et cetera that do really great things but which most patients can’t afford. There are intensive health monitoring programs – think nurses who come to your house and make sure you take your medication on time – which are proven to improve outcomes but which never have enough staff for everybody who needs them. There are omnipresent underfunded community mental health systems. All of these things are doing great work *right now*. Indeed, the plan for closing the state-run long-term facilities was to gradually transition care to all of these other systems, and where that was supported it worked well, and insofar as it didn’t work well it was because it wasn’t supported.

If we support all that, will it keep all mentally ill people out of the prison system? No. First of all, no treatment is perfect and most are downright mediocre. Second of all, like I said, there are mentally ill people in prison because there are mentally ill people everywhere. There are *disproportionately* many mentally ill people in prison because the risk factors for mental illness are the same as the risk factors for crime, like poverty and drug abuse. Regardless of the level of care given, mentally ill people are likely to end up in prison at increased rates, unless you’re willing to either institutionalize all mentally ill people before they can commit any crimes, or excuse all crimes committed by mentally ill people.

But we shouldn’t be making our mental health decisions based on worries about criminality and prisons. Most people who are mentally ill will never end up on the wrong side of the law, and many (most?) mentally ill people who do end up in prison will do so for reasons not directly related to their illness. Make mental health decisions because it’s the right thing to do and there are people who really need help.

And if mentally ill people do end up in prison? There is a forensic health system dedicated to treating mentally ill prisoners. It’s not perfect, but with more funding and attention it could be better. There are forensic psychiatric hospitals that house mentally ill prisoners, and though again they are not perfect, they at least have that great advantage that you can’t be put in them unless you are found guilty of a crime.

So my argument is: fund and use the community mental health system more to help people in the community. Fund and use the forensic mental health system more to help people in prison. But stop acting like the two groups are fungible. And stop trying to institutionalize more people. That doesn’t help.

Non-Shared Environment Doesn’t Just Mean Schools And Peers

[Epistemic status: uncertain. Everything in here seems right, but I haven’t heard other people/experts in the field talk about this nearly as much as I would expect them to if it were true. Obviously amount of variability attributable to environment (shared and non-shared) increases as the variability in environments in the sample increases]

The “nature vs. nurture” question is frequently investigated by twin studies, which separate interpersonal variation into three baskets: heritable, shared environmental, and non-shared environmental. Heritable mostly means genes. Shared environmental means anything that two twins have in common – usually parents, siblings, household, and neighborhood. Non-shared environmental is everything else.

At least in relatively homogeneous samples (eg not split among the very rich and the very poor) studies of many different traits tend to find that ~50% of the variation is heritable and ~50% is due to non-shared environment, with the contribution of shared environment usually lower and often negligible. This is typically summarized as “50% nature, 50% nurture”. That summary is wrong.

I mean, it’s tempting. All these social developmentalists were so sure that the way your parents praised you or didn’t praise you, or spanked you or didn’t spank you, had long-lasting repercussions that totally shaped your adult personality. The underwhelming performance of shared environment in twin studies torpedoed that whole area of study. But at least (these scholars of social behavior could tell themselves) it provided a consolation prize. The non-shared environment contributes 50% of variation, just as much as genes. That means things like your friends, your schoolteachers, and even that time you and your twin got sent away to separate camps must be really important. More than enough there to continue worrying about how society is Ruining The Children, right?

Not necessarily. Non-shared environment isn't really "non-shared environment" the way you would think. It's more of a dumpster. Anything that isn't genetic or family-related gets tossed into the non-shared environment term. Here are some of the things that go into that 50% non-shared environment:

1. Error. Measurement error is neither genetics nor family, so it ends up in the non-shared environmental term. Suppose you're studying intelligence, and you make a bunch of twins take IQ tests. IQ tests measure intelligence, but not perfectly. For example, someone who makes a lucky guess on a multiple choice IQ test will get a higher score even though they are not more intelligent than someone who makes an unlucky guess. Someone who takes the test when they're tired and stressed may get a lower score even though they're no less intelligent than somebody else who takes it well-rested and feeling good.

Imagine a world where intelligence is entirely genetic. Two identical twins take an IQ test, one makes some lucky guesses, the other is tired, and they end up with a score difference of 5 points. Then some random unrelated people take the test and they get the 5 point difference plus an extra 20 point difference from genuinely having different IQs. In this world, scientists might conclude that about 80% of IQ is genetic and 20% is environmental. But in fact in terms of *real, stable* IQ differences, 100% would be genetic and 0% environmental.

This gets even harder when trying to measure fuzzier constructs like criminality. Suppose someone does a twin study on criminality and their outcome is whether a twin was ever convicted of a felony. This depends partly on whether the twin is actually the sort of person with criminal tendencies – but also partly on whether a policeman happened to be in the area to catch them, whether their lawyer happened to be good enough to get them off, whether their judge was feeling merciful that day, et cetera. Imagine a world where criminality is entirely genetic. Identical Twin A becomes a small-time cocaine dealer in a back alley in West Philly, sells to an undercover cop, and ends up in jail. Identical Twin B becomes a small-time cocaine dealer in a back alley in East Philly, doesn't run into any undercover cops, and so avoids conviction. This shows up as "variation in criminality is due to non-shared environment".

Riemann and Kandler (h/t JayMan) run a study which is an excellent demonstration of this. Classical twin studies sometimes use self-report to determine personality – ie they ask people to rate how extraverted/conscientious/whatever they are. These studies find that most personality traits are about 40% genetic, 60% non-shared environmental. Riemann and Kandler obsessively collect every possible measurement of personality – self-report, other-report, multiple different tests – and average them out to get an unusually accurate and low-noise estimate of the personality of the twins in their study. *They* find that variation in personality is about 85% genetic, 15% non-shared environmental. So it looks like much of the non-shared environmental variation in traditional studies of personality was just error.

2. Luck of the draw. Bob becomes a junior advertising executive at Coca-Cola, where he designs a new ad targeting young female consumers. His identical twin Rob becomes a junior advertising executive at Pepsi-Cola, where he designs his own new ad targeting young female consumers. Both ads are very successful – in fact, exactly equally successful. But

Coke’s CEO is a crony capitalist who wants to replace everyone in the company with his college buddies, so he ignores Bob’s good work and demotes him to a low-level position. Pepsi’s CEO is a skilled leader who recognizes good talent when she sees it, and she promotes Rob to Vice-President Of Advertising.

Now a scientist comes along, does a twin study on them, and finds that they have very different levels of income. She reports that there’s a lot of difference between these two identical twins, so much of income must be non-shared environmental.

Science reporters read the study finding that much of the variation in income is non-shared environmental, and conclude that despite their identical genes, there must be deep and mysterious differences in Bob and Rob’s abilities and business acumen. They speculate that Rob had a very inspirational teacher in school who pushed him to achieve greatness, and Bob must have fallen in with a bad peer group who didn’t value hard work.

But actually, Bob and Rob are completely identical in every way, no incident in their past did anything to separate them, and Bob just ended up working for a crappy CEO. In this scenario, inherent predisposition to earning money is exactly the same in both twins, they just have different amounts of luck at it. If both twins become pathological gamblers, but one of them hits the jackpot and the other goes broke, that will show up as “non-shared environment” too.

3a. Biological random noise. The genome can’t encode the location of every cell in the body. Instead, it specifies high-level processes which create lower-level processes which create those cells. But this gives the lower-level processes a lot of leeway, meaning that there can be significant biological differences between identical twins.

Consider by analogy The Postmodernism Generator. It’s a cute program that will make a (sort of) convincing sounding postmodernist essay on demand. We can imagine hundreds of different programmers all designing their own postmodernism generators. Some would be really brilliantly designed and consistently come up with plausible looking essays. Others would be poorly designed and consistently come up with crappy essays that don’t convince anybody. But there would also be variation *within* the results of each generator. There might be a generator that is mostly terrible but occasionally by coincidence comes up with a really funny essay, or vice versa. In this analogy, the genes are the code for the generator, and the person is an individual essay produced by that generator.

Thus, identical twins have different fingerprints, different freckles, and different birthmarks. Only about a fifth of left-handers’ identical twins will also be left-handed. And twins even *look* different enough that their friends and parents eventually learn to tell them apart. All of these are non-genetic issues likely to show up in “non-shared environment” but not related to schools or peers or “nurture” as traditionally conceived.

3b. The immune system. Immunology is still poorly understood, but it seems very important. Immune reactions and neuroinflammation have been implicated to one degree or another in a lot of psychiatric diseases. A functional immune system can protect good health; a dysfunctional immune system can make someone constantly tired and miserable.

There seems to be more of an element of chance to the immune system than to a lot of other bodily processes. Part of it is the input – one child in a twin pair might inhale a particle of cat dander at a critical time; another might get some unknown adenovirus with no immediate effects but which contributes to obesity twenty years later. Another part is the output; sometimes a natural killer cell stumbles across something quickly and takes it out without any fuss; other times the immune system misses it for a while and it gets more of a chance to spread.

The end result is that immune-system-related-conditions are really discordant across identical twins. If your identical twin has asthma, there's only a 33% chance you'll have it as well. If your identical twin has Crohn's disease, a disabling autoimmune intestinal condition, there's only a 50% chance you'll suffer the same. I'm not sure how significant this is in the broad scheme of things, but I suspect more so than people think.

3c. Epigenetics We know that identical twins have substantially different epigenetics, and there are hints that this underlies discordant behavior. This is probably really important, but I feel bad bringing it up because it seems to be passing the buck. We usually think of epigenetic differences as a response to different environments or life choices. But if identical twins start with the same environment and can be expected to make the same life choices, why do they end up with different epigenomes? I'm not sure what to do with this one.

3d. Genes that differ between identical twins. Apparently this happens! Identical twins come from the same zygote, which means they start out with the same genes, but after that all bets are off. If there's a mutation in one twin in the first embryonic division, then half of that twin's cells will carry that mutation. Remember that there are a lot of divisions and opportunities for mutation before any cells even start forming the brain, and any mutation before that time could be transmitted to all brain cells. One study found that the average identical twin pair probably has about 359 genetic differences occuring early in development.

I’ve grouped 3a, 3b, 3c, and 3d together as possible biological sources of variation. One of these – or maybe some 3e I don’t know about – is probably the reason for less-than-perfect twin concordance in conditions like Parkinson’s disease, migraine, autism, and schizophrenia. Needless to say, anything that can make you schizophrenic can probably affect your personality and life outcomes pretty intensely.

But all of this gets counted as “non-shared environment” in a twin study, and used to play up the importance of schools and peer groups.

4. Actual nurture. Twins do have different experiences growing up. How much does this shape their adult traits? Can we separate this out in to specific experiences that shape adult traits, like school and summer camp?

The good news is that Eric Turkheimer has a big review article on this; the bad news is that the discussion section is called “The Gloomy Prospect” and says:

> Quantitative analysis of studies of specific nonshared environmental events shows that effect sizes measuring the effects of such variables on child outcomes are generally very small. Effect sizes are largest when confounds with genetic variability and outcome-to-environment causal effects are not controlled. When such confounds are controlled, as in the most recent reports from the NEAD project, effect sizes become smaller still.

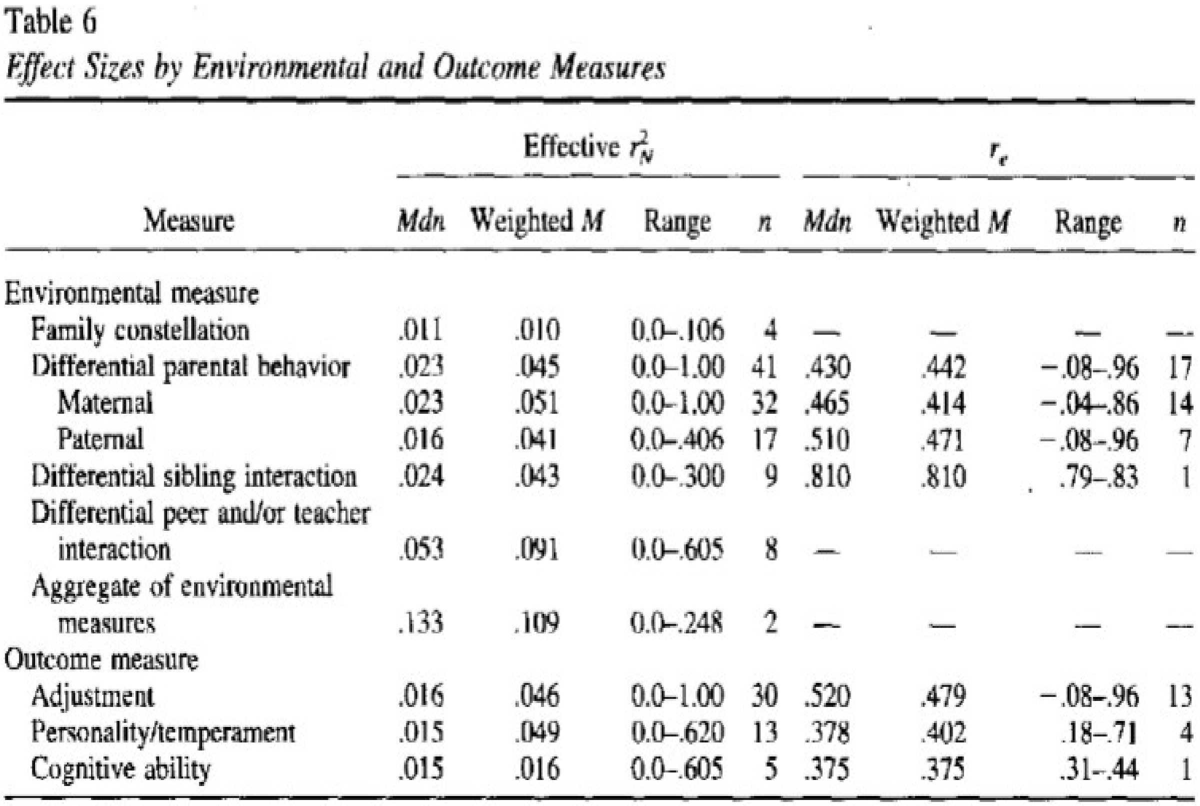


Table 6

Effect Sizes by Environmental and Outcome Measures

	Effective r_N^2				r_e			
Measure	*Mdn*	Weighted *M*	Range	*n*	*Mdn*	Weighted *M*	Range	*n*
Environmental measure								
Family constellation	.011	.010	0.0–.106	4	—	—	—	—
Differential parental behavior	.023	.045	0.0–1.00	41	.430	.442	−.08–.96	17
Maternal	.023	.051	0.0–1.00	32	.465	.414	−.04–.86	14
Paternal	.016	.041	0.0–.406	17	.510	.471	−.08–.96	7
Differential sibling interaction	.024	.043	0.0–.300	9	.810	.810	.79–.83	1
Differential peer and/or teacher interaction	.053	.091	0.0–.605	8	—	—	—	—
Aggregate of environmental measures	.133	.109	0.0–.248	2	—	—	—	—
Outcome measure								
Adjustment	.016	.046	0.0–1.00	30	.520	.479	−.08–.96	13
Personality/temperament	.015	.049	0.0–.620	13	.378	.402	.18–.71	4
Cognitive ability	.015	.016	0.0–.605	5	.375	.375	.31–.44	1

I'm not sure if this table represents the "very small" uncontrolled or the "smaller still" controlled sizes

The paper concludes: "We emphasize that these findings should not lead the reader to conclude that the nonshared environment is not as important as had been thought."

But although I have a huge amount of respect for Turkheimer, I *kind of* want to conclude that the nonshared environment is not as important as had been thought. My guess is that the nonshared environment as Turkheimer discusses it – differential parenting, schools, peers, and so on – is only a fraction of the "nonshared environmental" term in genetics studies.

If that were true, it would mean that nature is more important than we thought relative to environment in terms of things we can understand and possibly affect. That would make the quest to change important outcomes like intelligence, personality, income, or criminality by changing society even more daunting. And it would make the opportunity to change those outcomes through genetic engineering even more tempting.

Book Review: The Art Of The Deal

I.

Many of my friends recommend Robert Cialdini's *Influence*, a book about how to be persuasive and successful. I read it most of the way through, and it was okay, but I didn't have it in me to finish the whole thing. It's not that being persuasive and successful doesn't sound pretty neat. But how could I be sure the book would deliver the goods?

Robert Cialdini's Wikipedia page says "He is best known for his book *Influence*". Since its publication, he seems to have spent his time directing an institute to spread awareness of techniques for success and persuasion. At the risk of being a little too cynical – a guy knows the secrets of success, so he uses them to…write a book about the secrets of success? If *I* knew the secrets of success, you could bet I'd be doing much more interesting things with them. All the best people recommend Cialdini, and his research credentials are impeccable, but I can't help wondering: if he's so smart, why isn't he Emperor?

Donald Trump may not be Emperor, but he's a lot closer than Robert Cialdini. I knew that Trump wrote his own book on success and persuasion back in 1988 – Trump: The Art of the Deal – and I wondered if it might not be the anti-Cialdini.

Trump is no psychology expert, but after a few months of attributing his victories to blind luck, most people have accepted Scott Adams' hypothesis that he's really a "master persuader". Salon, Daily Caller, Bill Maher, and the Economist all use the word "genius". The less you respect Trump's substance – and I respect it very little – the more you're forced to admire whatever combination of charisma, persuasion, and showmanship he uses to succeed without having any. If this guy has written a book on how to be persuasive and successful, that’s a book I want to read.

II.

The downside of buying a book by a master manipulator is that sometimes you learn you were manipulated into buying the book.

Trump: The Art Of The Deal is 365 pages of some of the biggest print I have ever seen. The cover has a quote from the *New York Times* – “Trump makes one believe for a moment in the American dream again” – which some poor reviewer is probably desperately wishing he could take back right now.

Although the blurb says that he “fully reveals the deal-maker’s art” and that it is “an unprecedented education in the practice of deal-making” and “the ultimate read for anyone interested in achieving money and success” – only seventeen pages of very large print are anything resembling business advice. The rest of it is a weirdly deal-focused autobiography that doesn’t mention marrying his wife or having children, but devotes a lovingly detailed twenty-four pages to the time he renovated the Commodore Hotel.

But first, those seventeen pages. I am pleased to report that Donald Trump is weirdly abreast of modern science – he tells his readers looking for advice about how to make it big that deal-making is probably just genetic. Either you’ve got the deal, gene or you don’t:

> More than anything else, I think deal-making is an ability you’re born with. It’s in the genes…unlike the real estate evangelists you see all over television these days, I can’t promise you that by following the precepts I’m

> about to offer you'll become a millionaire overnight. Unfortunately, life rarely works that way, and most people who try to get rich quick end up going broke instead.

> Every time I speak of the haters and losers I do so with great love and affection. They cannot help the fact that they were born fucked up!
>
> — Donald J. Trump (@realDonaldTrump) September 29, 2014

Related?

This is a weirdly humble and self-aware Trump. It might be that the book medium suits him well; more likely he just has a really good ghost-writer. Unfortunately, he has much to be humble about. His advice, while not bad, is vague and not too useful. For example, his first rule is "think big". But his second rule is "protect the downside and the upside will take care of itself", which he explains as:

> It's been said that I believe in the power of positive thinking. In fact, I believe in the power of negative thinking. I happen to be very conservative in business. I always go into the deal anticipating the worst. If you plan for the worst – if you can live with the worst – the good will take care of itself.

So – take a lot of risks, but also be very cautious. Okay. I'm not saying his advice is literally contradictory – it makes sense that you can have big plans but also be very careful about them. I just don't get the feeling that his advice is too helpful in narrowing down your strategy.

Is there anything at all worth reading in these seventeen pages? Oh yes. But not for the reason I expected.

Trump’s sixth rule of deal-making is “Get The Word Out”. He says:

> One thing I’ve learned about the press is that they’re always hungry for a good story, and the more sensational the better. It’s in the nature of the job, and I understand that. The point is that if you are a little different, or a little outrageous, or if you do things that are bold or controversial, the press is going to write about you…
>
> The funny thing is that even a critical story, which may be hurtful personally, can be very valuable to your business. [When I announced my plans to build Television City to the press], not all of them liked the idea of the world’s tallest building. But the point is that we got a lot of attention, and that alone creates value.
>
> The other thing I do when I talk with reporters is to be straight. I try not to deceive them or to be defensive, because those are precisely the ways most people get themselves into trouble with the press. Instead, when a reporter asks me a tough question, I try to frame a positive answer, even if that means shifting the ground. For example, if someone asks me what negative effects the world’s tallest building might have on the West Side, I turn the tables and talk about how New Yorkers deserve the world’s tallest building, and what a boost it will give the city to have it again. When a reporter asks why I build only for the rich, I note that the rich aren’t the only ones who benefit from my buildings. I explain that I put thousands of people to work who might otherwise be collecting unemployment, and that I add to

> the city’s tax base every time I build a new project. I also point out that buildings like Trump Tower have helped spark New York’s renaissance.
>
> The final key to the way I promote is bravado. I play to people’s fantasies. People may not always think big themselves, but they can still get very excited by those who do. That’s why a little hyperbole never hurts. People want to believe that something is the biggest and the greatest and the most spectacular.
>
> I call it truthful hyperbole. It’s an innocent form of exaggeration – and a very effective form of promotion.

In the immortal words of Marco Rubio, “Let’s dispel once and for all with this fiction that Donald Trump doesn’t know what he’s doing. He knows exactly what he’s doing.”

On the other hand, his eighth rule of business is “Deliver The Goods”. He gives an interesting example:

> You can’t con people, at least not for long. You can create excitement, you can do wonderful promotion and get all kinds of press, and you can throw in a little hyperbole. But if you don’t deliver the goods, people will eventually catch on.
>
> I think of Jimmy Carter. After he lost the election to Ronald Reagan, Carter came to see me in my office. He told me he was seeking contributions to the Jimmy Carter Library. I asked how much he had in mind. And he said, “Donald, I would be very appreciative if you contributed five million dollars.
>
> I was dumbfounded. I didn’t even answer him.

> But that experience also taught me something. Until then, I’d never understood how Jimmy Carter became President. The answer is that as poorly qualified as he was for the job, Jimmy Carter had the nerve, the guts, the balls, to ask for something extraordinary. That ability above all helped him get elected president. But then, of course, the American people caught on pretty quickly that Carter couldn’t do the job, and he lost in a landslide when he ran for reelection.
>
> Ronald Reagan is another example. He is so smooth and so effective a performer that he completely won over the American people. Only now, nearly seven years later, are people beginning to question whether there’s anything beneath that smile.

Trump-1988 is weirdly prophetic.

Finally, his tenth rule is “Have Fun”:

> I don’t kid myself. Life is very fragile, and success doesn’t change that. If anything, success makes it more fragile. Anything can change, without warning, and that’s why I try not to take any of what’s happened too seriously. Money was never a big motivation for me, except as a way to keep score. The real excitement is playing the game. I don’t spend a lot of time worrying about what I should have done differently, or what’s going to happen next. If you ask me exactly what the deals I’m about to describe all add up to in the end, I’m not sure I have a very good answer. Except that I’ve had a very good time making them.

Marcus Aurelius, eat your heart out.

III.

So much for seventeen pages of business advice. The other three hundred forty-eight pages are Trump *gushing* about the minutiae all of the interesting deals he's been a part of.

"GUYS, YOU'RE NOT GOING TO BELIEVE THIS, THERE WAS THIS ONE SKYSCRAPER THAT WAS SUPPOSED TO HAVE A FLOOR TO AREA RATIO OF 6, BUT THEN I BEAT HILTON IN NEGOTIATING THE AIR RIGHTS FROM THE COMPANY NEXT DOOR, AND ACTIVATED AN OPTION TO BUY A PROPERTY ON THE OTHER SIDE OF IT, AND ALL OF THAT LANDED ME A PARTNERSHIP WITH ONE OF THE BIG BANKS, AND THEN THE PLANNING BOARD TOTALLY CHANGED THE FLOOR AREA RATIO! CAN YOU BELIEVE IT, GUYS??!"

Overall the effect was that of an infodump from an autistic child with a special interest in real estate development, which was both oddly endearing and not-so-oddly very boring.

I started the book with the question: what exactly do real estate developers *do*? They don't design buildings; they hire an architect for that part. They don't construct the buildings; they hire a construction company for that part. They don't manage the buildings; they hire a management company for that part. They're not even the capitalist who funds the whole thing; they get a loan from a bank for that. So what *do* they do? Why don't you or I take out a $100 million loan from a bank, hire a company to build a $100 million skyscraper, and then rent it out for somewhat more than $100 million and become rich?

As best I can tell, the developer's job is coordination. This often means *blatant lies*. The usual process goes like this: the bank would be happy to lend you the money as long as you have guaranteed renters. The renters would be happy to sign up as long as you show them a design. The architect would be happy to design the building as long as you tell them what the government's allowing. The government would be happy to give you your permit as long as you have a construction company lined up. And the construction company would be happy to sign on with you as long as you have the money from the bank in your pocket. Or some kind of complicated multi-step catch-22 like that. The solution – or at least Trump's solution – is to tell everybody that all the other players have agreed and the deal is completely done except for their signature. The trick is to lie to the right people in the right order, so that by the time somebody checks to see whether they've been conned, you actually do have the signatures you told them that you had. The whole thing sounds very stressful.

The developer's other job is dealing with regulations. The way Trump tells it, there are so many regulations on development in New York City in particular and America in general that erecting anything larger than a folding chair requires the full resources of a multibillion dollar company and half the law firms in Manhattan. Once the government grants approval it's likely to add on new conditions when you're halfway done building the skyscraper, insist on bizarre provisions that gain it nothing but completely ruin your chance of making a profit, or just stonewall you for the heck of it if you didn't donate to the right people's campaigns last year. Reading about the system makes me both grateful and astonished that any structures have ever been erected in the United States at all, and somewhat worried that if anything ever happens to Donald Trump and a few of his close friends, the country will lose the ability to legally construct artificial shelter and we will all have to go back to living in caves.

Trump's greatest pride is his ability to construct things on time and under budget. He gives the story of an ice rink that

New York City was trying to renovate in Central Park. After six years and $13 million, the city had completely failed to renovate it and just made things worse. Trump offered as a charitable gesture to do it himself, and the mayor, who was a political enemy, refused. The press hounded the mayor, Trump eventually was allowed to try, and he finished it in four months for only $2.5 million. He boasted that he finished fixing the rink in less time than it took the city to complete their study on why their rink-fixing project had failed.

He had a couple more stories like this – but throughout all of it, there was a feeling of something missing. Here is a guy whose job is cutting through bureaucracy, and who is apparently quite good at it. Yet throughout the book – and for that matter, throughout his campaign for the nomination of a party that makes cutting bureaucracy a big part of their platform – he doesn’t devote a lot of energy to expressing discontent with the system. There is no libertarian streak to Trump – in the process of successfully navigating all of these terrible rules, he rarely takes a step back and wonders about a better world where these rules don’t exist. Despite having way more ability to change the system than most people, he seems to regard it as a given, not worth debating. I think back to his description of how it’s all just a big game to him. Most star basketball players are too busy shooting hoops to imagine whether the game might be more interesting if a three-pointer was worth five points, or whatever. Trump seems to have the same attitude – the rules are there; his job is to make the best deal he can within those rules.

Maybe I’m imagining things, but I feel like this explains a lot about his presidential campaign. People ask him something like “How would you fix Medicare?”, and he gives some vapid answer like “There are tremendous problems with

Medicare, but I'm going to hire the best people. I know all of the best doctors and health care executives, and we're going to cut some amazing deals and have the best Medicare in the world." And yeah, he did say in his business tips that you should change the frame to avoid being negative to reporters. But this isn't a negative or a gotcha question. At some point you'd expect Trump to do his homework and get some kind of Medicare plan or other. Instead he just goes off on the same few tangents. This thing about hiring the best people, for example, seems almost like an obsession in the book. But it works for him. When somebody sues him (which seems like an hourly occurrence in real estate development no matter how careful you are) his response is to find the best lawyer, hire them, and throw them at the problem. When he needs a hotel managed, he hires the best hotel managers and tells them to knock themselves out. Even his much-mocked tendency to talk about all the people he knows comes from this being a big part of his real estate strategy – one of the reasons he can outcompete other tycoons is because he knows people on the planning board, knows people in the banks, knows people in all the companies he works with. It's a huge advantage for him.

These strategies have always worked for him before, and floating off into some intellectual ideal-system-design effort has never worked for him before. So when he says that he's going to solve Medicare by hiring great managers and knowing all the right people, I don't think this is some vapid way of avoiding the question. I think it's the honest output of a mind that works very differently from mine. I've been designing ideal systems of government for the heck of it ever since I was old enough to realize what a government was. Trump is at serious risk of actually taking over a government, and such design *still* doesn't appeal to him. The best he can do is say that other people are *bad* at governing, but he’s going to be *good* at governing, on account of his deal-making skill. I think he honestly believes this. It makes perfect sense in real estate, where some people are good businesspeople, others are bad businesspeople, and the goal is to game the system rather than change it. But in politics, it’s easy to interpret as authoritarianism – “Forget about policy issues, I’m just going to steamroll through this whole thing by being personally strong and talented.”

I said it before, but it bears repeating – this book has a really good ghostwriter. Yeah, it comes across as narcissistic; there’s probably no way to avoid that in a Trump autobiography. But Donald Trump’s interest in Donald Trump pales beside his blazing hot interest in the sheer awesomeness of hotel property deals. And part of me wants to say that people with obsessive interests in bizarre things are My Kind Of People.

But there’s still something alien about Trump here, even moreso than with the populist demagogue of the campaign trail. Trump the demagogue is attacked as anti-intellectual. I get anti-intellectualism because – like all isms – it’s an intellectual idea, and I tend to think in those terms. But Trump of the book is more a-intellectual, in the same way some people are amoral or asexual. The world is taken as a given. It contains deals. Some people make the deals well, and they are winners. Other people make the deals poorly, and they are losers. Trump does not need more than this. There will be no civilization of philosopher-Trumps asking where the first deal came from, or whether a deal is a deal only by virtue of its participation in some primordial deal beyond material existence. Trump’s world is so narrow it’s hard to fit your head inside it, so narrow that on contact with any wider world it seems strange and attenuated, a broken record of deals and connections and hirings expanding to fill the space available.

On the other hand, he made a billion dollars and will probably win the GOP nomination. So there's that.

Trump ends by saying:

> What's next? Fortunately, I don't know the answer, because if I did, that would take all the fun out of it. This much I do know: it won't be more of the same.
>
> I've spent the first twenty years of my working life building, accumulating, and accomplishing things that many said could not be done. The biggest challenge I see over the next twenty years is to figure out some creative ways to give back some of what I've gotten.
>
> I don't just mean money, although that's part of it. It's easy to be generous when you've got a lot, and anyone who does, should be. But what I admire most are people who put themselves directly on the line. I've never been terribly interested in why people give, because their motivation is rarely what it seems to be, and it's almost never pure altruism. To me, what matters is the doing, and giving time is far more valuable than just giving money. *[note: a contrary perspective]*
>
> In my life, there are two things I've found I'm very good at: overcoming obstacles and motivating good people to do their best work. One of the challenges ahead is how to use those skills as successfully in the service of others as I've done, up to now, on my own behalf.
>
> Don't get me wrong. I also plan to keep making deals, big deals, and right around the clock.

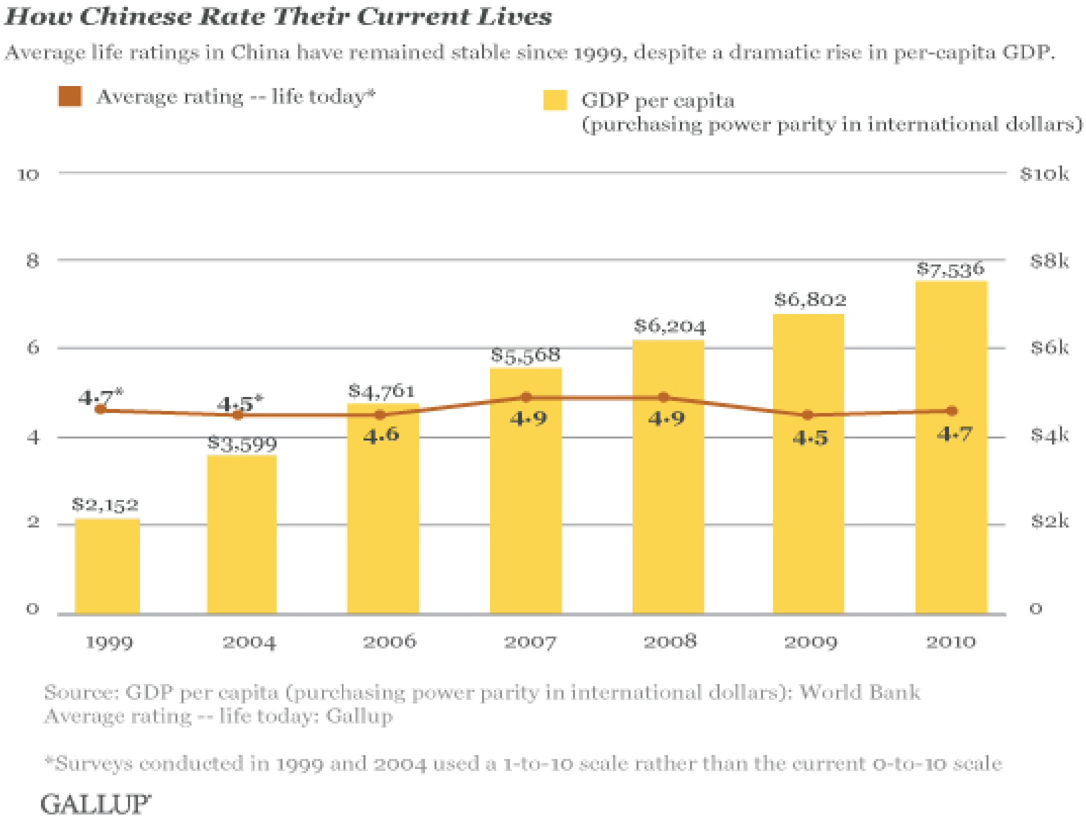
The Price Of Glee In China

[Epistemic status: Overly simplistic treatment of a horrifyingly complex topic; I can only hope I haven't missed enough to completely embarrass myself]

I.

Noah Smith reviews recent economic research suggesting that globalization was a net harm to working class people in rich countries like the US; he tentatively suggests this could justify a weak form of protectionism. But Scott Sumner argues that's the wrong way to look at things. Globalization fueled China's transition from a poor agrarian economy to an industrialized modern nation. A billion people were lifted out of poverty, an accomplishment Sumner calls "the best thing that ever happened". This is far more important than the less dramatic costs imposed on the US. Therefore, even if we agree globalization hurts the working class of rich nations, it's still a morally defensible policy since it benefits the needier working classes of much poorer nations.

On the one hand, this makes sense. On the other, here's happiness in China over the past fifteen years:



How Chinese Rate Their Current Lives

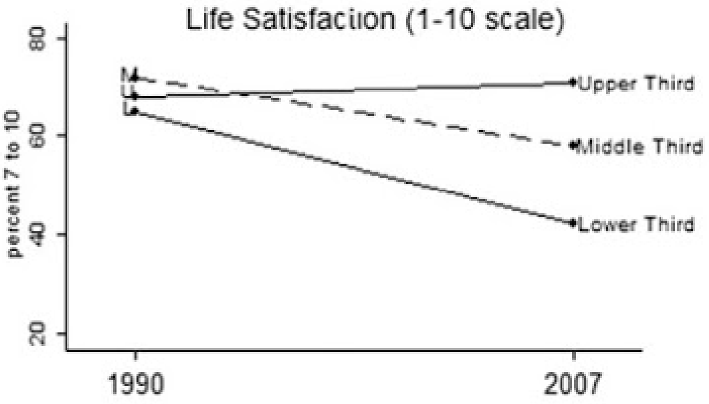
Average life ratings in China have remained stable since 1999, despite a dramatic rise in per-capita GDP.

Source: GDP per capita (purchasing power parity in international dollars): World Bank
Average rating -- life today: Gallup

*Surveys conducted in 1999 and 2004 used a 1-to-10 scale rather than the current 0-to-10 scale

GALLUP

Measuring happiness is really hard, but the Chinese result seems as robust as any. You get the same thing if you ask about satisfaction versus dissatisfaction. Brookings analyzes five different series of happiness data and concludes that “the Chinese became less happy during their growth boom”. The New York Times agrees and says that “Chinese people’s feelings of well-being have declined in [this] period of momentous improvement in their economic lives”. And this seems to be worst among the poorest Chinese:



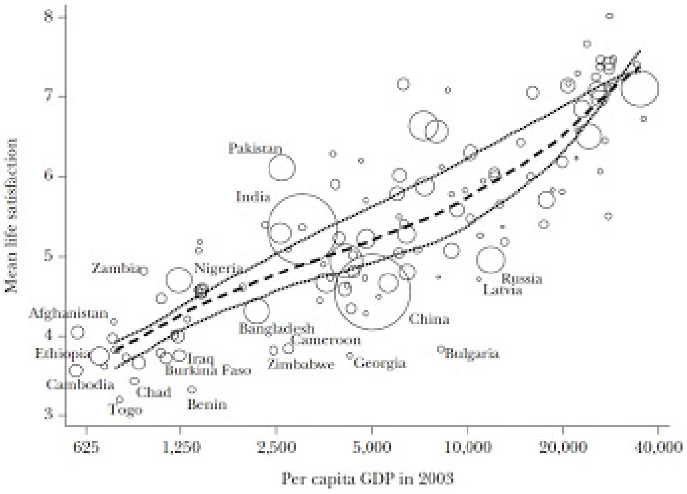
Nor does this seem to be an effect from our happiness research just not being good enough to capture changes in happiness even if they occur. There's good evidence that increased income *within* a country increases happiness, and various other things have been found to be effective too. I would even argue we can find happiness changes in nations – recent surveys have found Iraq and Syria to be the least happy nations in the world, and I doubt this was true before those countries' respective wars. It seems to just be national GDP per capita that doesn't do anything.

This is Easterlin's Paradox, the observation that a country in general does not get happier as it becomes richer. This is *very* controversial, with statisticians analyzing and reanalyzing data and crunching it a bunch of different ways. In the latest volley in this eternal war, Easterlin's side came out with data from 37 countries over 30 years, including many countries that underwent spectacular growth during that time, and confirmed their original conclusion.

There are certainly graphs like this one that propose a nice clear log relationship between income and happiness:

Figure 2

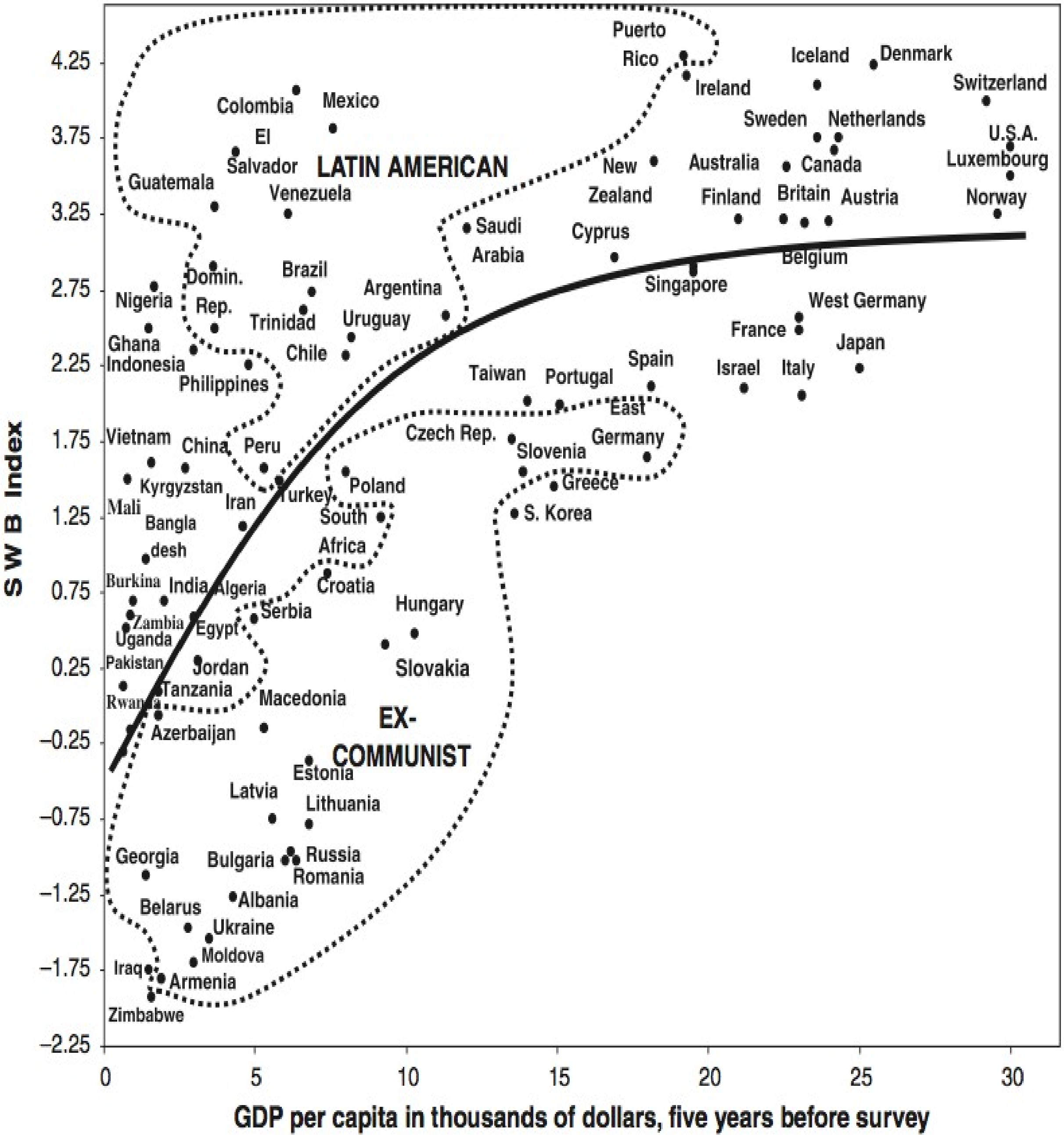
Each Doubling of GDP is Associated with a Constant Increase in Life Satisfaction



Source: Penn World Table 6.2.

Note: Each circle is a country, with diameter proportional to population. The scale on the x-axis is logarithmic. The middle line shows average life satisfaction for each level of per capita GDP while the outer two lines show the same thing, but for two age groups, ages 15 to 25—the upper line for most of the figure—and ages 60 and over—which is usually the lower line. GDP per capita in 2003 is measured in purchasing power parity chained dollars at 2000 prices.

But I find the exact breakdown much more interesting:



Here we see a lot of cultural variation in this apparent happiness-income relationship. For example, Latin American countries are consistently poor but happy; Eastern European countries are usually richer but sadder than African countries, et cetera. Looking at the original graph above, you’d expect Chinese growth to make them much happier; looking at this graph, you notice that China’s three rich neighbors – Japan, Taiwan, and South Korea – are all about as happy as China. South Korea, despite making five times more money, is *less happy* than China is. If China’s income quintuples, why would you expect it to look like France or Ireland rather than South Korea?

Just to rub this in a little:

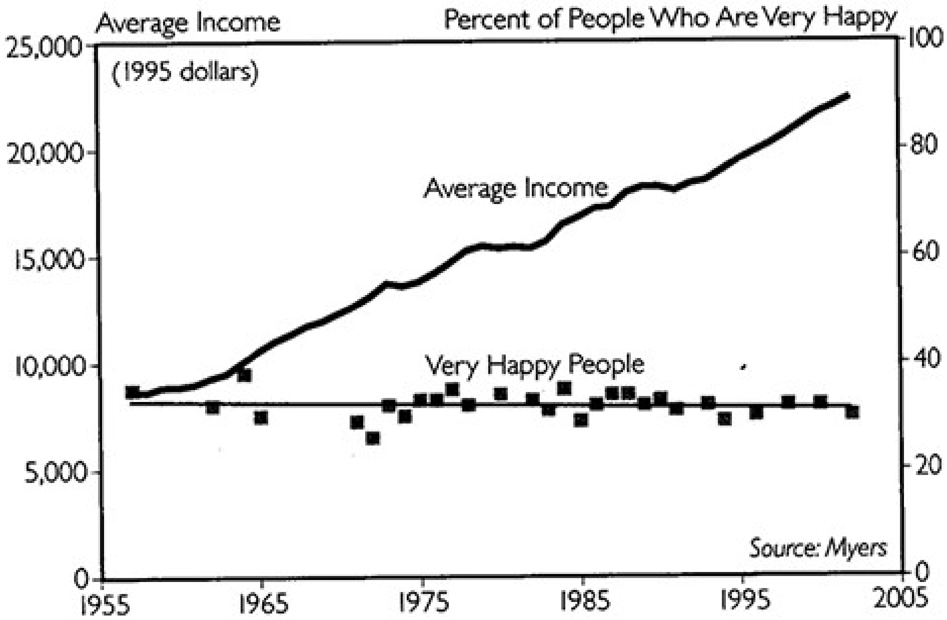


Figure 8–1. Average Income and Happiness in the United States, 1957–2002

A UN report theorizes that although richer countries tend to be happier, this is more likely due to factors other than income, like freedom, social trust, and stable families. These may be stable on scales much longer than income is, and may be related to culture.

II.

Let’s assume for a second that all this is true. National income does not matter for national happiness, and if China’s growth continues to skyrocket then in twenty years it will be as rich as Japan but not an iota happier than it is today. What do we *do* with this kind of knowledge?

Or let me ask a more specific question. Suppose that some free trade pact will increase US unemployment by 1%, but also accelerate the development of some undeveloped foreign country like India into hyper-speed. In twenty years, India's GDP per capita will go from $1,500/year to $10,000/year. The only cost will be a million or so extra unemployed Americans, plus all that coal that the newly vibrant India is burning probably won't be very good for the fight against global warming.

Part of me wants to argue that *obviously* we should sign the trade pact; as utilitarians we should agree with Sumner that lifting 1.4 billion Chinese out of poverty was "the best thing that ever happened" and so lifting 1.2 billion Indians out of poverty would be the second-best thing that ever happened, far more important than any possible risks. But if Easterlin is right, those Indians won't be any happier, the utility gain will be nil, and all we will have done is worsened global warming and kicked a million Americans out of work for no reason (and they will definitely be unhappy).

Or since most of us don't get the option to sign trade pacts, here's a more relevant question. Suppose we are effective altruists. We have the opportunity to cure disease (at relatively high costs) or boost national development (at relatively low costs). Assume the numbers work out such that if we took a simple 'development = good' perspective, then donating to the development charity would be a no-brainer. Should we donate to the disease-cure charity anyway?

A couple of years ago, I learned that people who were paralyzed in car accidents took a few months to adjust to their new situation, but after that were no less happy than people who were still healthy and abled. Then last December I learned that this was an urban legend, that people who were paralyzed in car accidents were mostly as miserable as you would expect. But for those few years while I still believed that particular factoid, I was a little creeped out. Was a doctor who helps car accident victims recover their function wasting her life? If people got genuine enjoyment from driving drunk at 95 mph while shouting “WOOOOOOOOOOOOO!”, was there any reason to make them stop, since they weren’t *really* hurting anybody?

(I admit I’m skipping over factors like how paralyzed people can’t earn any income to pay into the tax system and stuff, but I’m just saying I would be pretty creeped out if that were the only reason we should avoid car accidents.)

Again assuming I haven’t made some simple calculation mistake, I can think of three ways to go from here. First, abandon consequentialism entirely (I understand that having children will likely decrease my happiness, but I still want to have children because I value them for non-utilitarian reasons). Second, switch to a consequentialism based on non-subjective things like maximizing development and industrialization as a terminal goal (Really? Even if everyone hates it? Does it matter what the factories are building? How about paper clips?). Third, switch to preference utilitarianism.

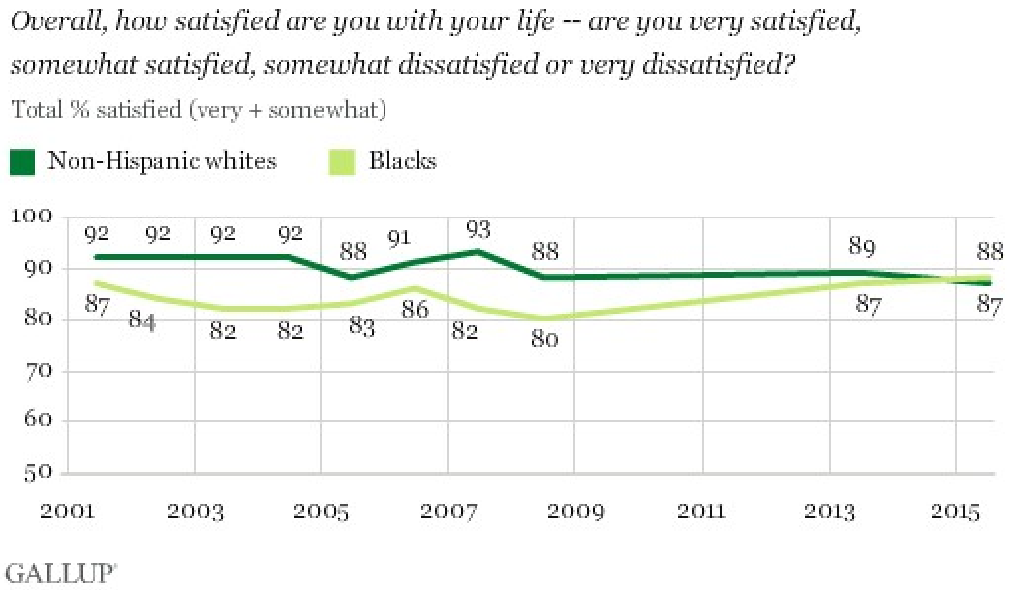
Preference utilitarianism is tempting and I was kind of in favor of it already, but I don’t find it completely satisfying. Suppose I myself am an Indian peasant. Should I have a preference for my society industrializing? If I’m not going to be any happier after it does, and supposing there’s no inherent moral value in industrialization, why bother? And if Indian peasants want their country to industrialize anyway, aren’t we as Americans allowed to say we don’t take their preference that seriously? If some hippie said they wanted to go on some Spiritual Yoga Nature Retreat that would turn their life around and bring them constant bliss, but we knew it was a complete fraud that wouldn't help them at all, would we still feel a moral obligation to help fund that hippie's retreat? How are the two situations different?

There's a risk of being patronizing here – telling the Indians "Oh, you don't need to industrialize, it's not so great anyway," even while we ourselves enjoy our nice food and flat-screen TVs. If we were to actively try to keep the Indians from industrializing, that would be pretty awful. But that's not the argument at hand here. The argument at hand is "are we morally required to sacrifice our own economy in order to help the Indians industrialize?", and I feel like that's a hard sell if industrialization doesn't really help the Indians.

And there's also a risk that I might be misdefining happiness. Maybe every way economists have hitherto measured happiness is hopelessly deficient, and there's some ineffable essence of happiness which, if we could get at it, would increase during national development. I admit that all of these subjective well-being indices are kind of sketchy and change a lot with the wording that you use or don't use.

A final option for rescuing common sense might be acknowledging that economic progress doesn't change happiness *yet*. That is, there are ways to convert economic (and closely linked technological) progress into happiness, but most countries are not making use of them – either for political reasons, or because they don't know about them, or because we haven't gotten enough technological and economic progress to reach them yet. This seems probably true to me – if nothing else, a technological singularity ought to help – but this situation looks a lot different from the situation where incremental progress increases happiness. In particular, it would make us want to concentrate our resources on increasing technological progress, perhaps in the richest economies, rather than trying to help poor countries in particular.

None of these possibilities really appeal to me, and I am forced to acknowledge that happiness research remains a very strange field whose conclusions make no sense to me and which tempt me to crazy beliefs and actions if I take them seriously.



I guess we're done fighting racism. Good job, guys.

Book Review: My Brother Ron

[Content warning: mental illness, forced institutionalization, anorexia. As always all patient anecdotes are obfuscated composites of multiple cases with all the details changed in order to protect people's privacy]

I.

After I wrote about Prison And Mental Illness, a reader recommended I read *My Brother Ron* by Clayton Cramer, a recent book/memoir arguing against deinstitutionalization. Cramer tells the story of his schizophrenic brother Ron, who was poorly treated because of the lack of an institutional system and so ended up dealing with homelessness and violence, then surveys the history and current state of mental health care in America and the various reasons why deinstitutionalization was a bad idea.

I found the book interesting and engaging, and its arguments intellectually honest and well-written. But in the end I just wasn't convinced.

But first, his brother Ron. Smart guy, joined the military, did well, finished his tour of duty, went to college, studied electrical engineering. Around 22 – the usual age for this to happen – he started acting weird, dropped out of college, obsessed over weird things like nickels, started thinking random people were plotting against him, et cetera. He ended up in a psych hospital where he got Thorazine and improved quickly – which meant, ironically, that when it came time for his commitment hearing two weeks later, the judge thought he looked pretty normal and released him.

Then he went to live with his family – including his brother the author – where he stopped his medication, started acting violently, smashed windows, screamed at people, and was otherwise a poor housemate. His parents asked him to leave, and he wandered around until he ended up in Santa Monica. There the government gave him a monthly disability check, which he spent on alcohol and a room in a disgusting hotel; when the money ran out around the middle of the month, he spent the next few weeks on the street until he got his next check, after which the cycle repeated itself.

Every so often he would break some law or annoy somebody enough to get arrested, at which point the police would bring him to a psychiatric hospital, he'd be placed on drugs, and he'd get better. Usually he'd leave after a few days to a few weeks. Occasionally he would keep taking the drugs after getting out, become pretty with-it, and try to go back to college. Sometimes he'd stay stable for months, even a year or two. But eventually he would stop taking the drugs for one reason or another, decompensate, and end up back on the streets, his previous progress ruined.

So the author asks: how did we get to this point? He answers with a fascinating history of American mental health care.

II.

Mental health care during the colonial era was surprisingly non-terrible. Mental illness seemed to be pretty well-understood and nobody was accusing psychotics of being witches or trying to beat the demons out of them or anything. Most of the mentally ill lived with families or in their own houses, where other members of the community supported them as best they could. Some were given jobs, with the understanding that they needed the support and their idiosyncrasies would be excused. Some would wander off, and there was a general understanding among colonial towns that if they found a mentally ill person wandering they would return them to their town of origin, who had the ultimate responsibility of caring for them. A few very violent people were locked away, usually in the basements of general hospitals or in prison cells. Getting somebody committed for mental illness was an informal process usually involving finding the friendly local magistrate and explaining why it was a good idea. But this option seems to have been used judiciously, and the incarcerated individuals managed to avoid most abuse and torture. Cramer describes it as “gloriously idyllic…mental illness appears to have been rare, and small town life tolerated all but the ‘furiously mad’ to live in the community.”

The part I found most interesting here was Cramer’s theory about why this system ended. Part of it was the end of small town life; a little village where all the families know each other is more likely to tolerate someone’s eccentricities than a large city of atomized individuals. But a bigger part may have been an unmanageable increase in the mentally ill population.

> Urbanization may not simply have been a factor in making Americans more wary of their mentally ill neighbors; it may have increased mental illness rates as well. While we do not know if this was true in the eighteenth century, some recent studies suggest that being born or growing up in an urban area increases one’s risk of developing schizophrenia and other psychoses. in the twentieth century, comparison of insanity rates revealed that urban areas had much higher rates of mental hospital admissions for schizophrenia and bipolar disorder – almost twice as high for New York City compared to the rest of New York State…older statistical examinations of mental hospital admissions argue that at least in the period from 1840

to 1940, while mental hospitalizations increased (because of increased availability) there was no large and obvious increase in insanity. A more recent study of mental illness data shows, much more persuasively, that psychosis rates rose quite dramatically between 1807 and 1961 in the United States, England and Wales, Ireland, and the Canadian Atlantic provinces. A study of Buckinghamshire, England shows more than a ten-fold increase in psychosis rates from the beginning of the seventeenth century to 1986. In 1764, Thomas Hancock left 600 pounds to the City of Boston to build a mental hospital for the inhabitants of Massachusetts. The city declined to accept the gift on the grounds that there were not enough insane persons to justify building such a facility. Massachusetts had a population between 188,000 and 235,000 in 1764; if the population of the time suffered the same schizophrenia rates as today, that would mean that there were about 2000 schizophrenics in the province. Even accounting for the greater tolerance of small town life for the mentally ill, this lends credence to Torrey and Miller’s claim of rising psychosis rates. Urban life today is not the same as urban life then, and even the scale of what constitutes “urban” is dramatically different – but it is an intriguing possibility that the increased rates of mental illness at the close of the Colonial period were the results of urbanization.

Irish immigration may also have played a role in the increasing development of mental hospitals in America. It was widely believed in the 1830s that Irish immigrants were disproportionately present among the insane. More recent analysis shows that throughout the nineteenth and twentieth centuries, Ireland’s rates of insanity were twice or more than that of the United States, England, and Wales. Irish immigrants were also overrepresented

> in insane asylums in the United States, England, Australia and Canada at the end of the nineteenth century.

To this I would add that even today immigrants get schizophrenia at rates up to four times those of non-immigrant populations, though nobody agrees whether this is because the genetically vulnerable are more likely to immigrate or because immigration is a very stressful experience. Even today, developing countries seem to have less schizophrenia than developed countries do (although of course this is hard to prove with certainty). The idea of a tenfold increase in psychosis over the past few centuries is jarring but not entirely outlandish, and does a lot to explain why the mental health system is so much larger and more relevant now.

Faced with these problems, the early Americans created big mental institutions that attracted prestigious clinicians (I interviewed for a job at one of these a few years ago; they boasted that they were in the “Psychiatric Ivy League”, which was a pretty good window into how they thought of themselves). These could never really figure out whether their job was custodial (ie warehouse mentally ill people so they didn’t cause trouble on the streets) or clinical (treat mentally ill people and cure their psychosis), and the nineteenth century vacillated wildly between people making big claims about how they were dedicated to treating all their patients, versus admitting that it was the nineteenth century and nobody had the slightest idea how to do this. While they argued the institutions grew and grew. Along with the schizophrenics, they became the dumping ground for syphilitics (remember, before penicillin syphilis was a common incurable disease that usually caused insanity in its final stages) and old people with Alzheimers (not officially recognized at this point; before the invention of nursing homes they figured they might as well stick crazy old people in with all the other crazy people). Finally, after the obsolescence of the "poorhouse" but before the beginning of welfare, there were a bunch of poor people just completely unprepared for normal life, and some of them ended up in the mental institutions too for lack of a better place to put them. This sort of put a damper on a lot of the curability discussion; not only could 19th century doctors not cure mental illness, but most of the people there weren't even mentally ill in the traditional sense.

(not that some people didn't try. Cramer describes a Dr. "Henry Cotton, who removed teeth, tonsils, and parts of the intestine from hundreds of patients at the Trenton State Hospital in New Jersey. Cotton claimed that there were foci of infections in these organs that were causing the insanity and that removal of the infectious would cause clinical improvement." And then there was Dr. Wagner-Jauregg, whose bold strategy of deliberately infecting psychiatric patients with malaria actually paid off: many of them had syphilis, and the high fever induced by the malaria killed the syphilis bacterium. Wagner-Jauregg received the Nobel Prize for this insight; his later strategy of sterilizing schizophrenics on the theory that the disease was caused by masturbation was perhaps somewhat less Nobel-worthy.)

The institutions continued to grow. In 1954 the national mental health budget was $568 million; in 1959 it was $854 million. In 1951, states spend on average 8% of their budgets on psychiatric hospitals; New York spent *one third* of its budget on psychiatric hospitals (or not? see dispute in comments). Compare to today, when New York spends only about 20-30% of its budget on education. Psychiatric hospitals (which, remember, also subsumed the function of modern nursing homes) were a *huge* part of the infrastructure of government.

This started to shift in the 1940s due to what the book calls "dynamic psychiatry" (although they use this phrase a bit differently from how I understand the definition). The old, tired psychiatry was a simple dichotomy between sane people (who don't need psychiatric help) and insane people (who are totally out of touch with reality and need to be locked up for their own good). And it understood this distinction in relatively biological terms – they didn't know anything about genes or neurons them, but they figured *something* was going on. But the new, exciting psychiatry thought of mental illness as a continuum, with everybody having a little bit of mental illness – whether it was just neurosis or anxiety or whatever – and psychotics just being the people whose mental illnesses made it hard for them to function. The new school understood this in very psychosocial, Freudian terms. Schizophrenics were people with oppressively close mothers; autistics were people with distant, cold mothers, et cetera. Psychiatrists tended to like this new school, because it meant that instead of spending their time in scary mental institutions full of crazy people, they could spend their time in nice Viennese parlors talking to rich people about their families.

Around the same time, scientists invented Thorazine, which seemed to produce miraculous recoveries in institutionalized psychotic people. This was before anyone knew anything about the long-term side effects of Thorazine, so everyone figured it was a miracle drug with no side effects and now there was no need for mental institutions any more.

Then we got to the Sixties. Cramer mostly manages to avoid being too transparently political, but it's hard for him to talk about Sixties Leftists without a bit of vitriol. He describes the genesis of the anti-psychiatry movement – a wide variety of traditions all coming together in an agreement that the mentally ill are just Too Cool And Free-Spirited For Society and anybody who tries to treat them is a bad person who hates creativity and wants to make everyone conform. He describes the jettisoning of centuries of accumulated wisdom about the causes and presentation of mental illness in favor of an unexamined dogma that mental illness is caused by oppressive systems of social control. He describes how some people did a few quick studies showing that schizophrenic people mostly lived in bad neighborhoods full of social decay, and concluded that bad neighborhoods and social decay caused schizophrenia without considering any other possible causal structures (of course, we as a society have long since moved beyond that). Others argued that hospitalization was the sole cause of mental illness, turning otherwise happy eccentrics into violent lunatics (again, a position we have long since moved beyond).

He reserves some of his strongest words for anti-psychiatry psychiatrists like R. D. Laing and Thomas Szaszszsz:

> You might wonder how a psychiatrist could believe that there was no such thing as insanity. Would not the exposure to psychotic patients during Szasz's training have shown him the error of his ideology? It turns out that Szasz may not have had **any** exposure to psychotics. In a 1997 interview, he describes how he consciously selected a psychiatric residency "that did not include work with involuntary patients". The chairman of the Psychiatry Department told him, "Tom, you have only one year left of your residency, I don't think it's right that you should finish without any experience with psychotic patients. I think you should do your third year at the Cook County Hospital." So Szasz quit and went elsewhere to avoid that experience.

> Szasz was drafted into the Navy after completing his training, and his experiences there almost certainly reinforced his already well-developed belief that mental illness did not exist. “The servicemen didn’t want to be in the Navy and played the role of mental patient. I didn’t want to be in the Navy and played the role of military psychiatrist. My job was to discharge the men from the Service as ‘neuropsychiatric casualties’.” Szasz had gone out of his way to avoid seeing psychotic patients, and then took a job that he describes as certifying that sane people pretending to be insane were actually insane as a convenient fiction. Is there anything surprising about Szasz’s projection of this situation onto the entire profession?

I actually *had* been wondering about that, and that clears up a lot. As for Laing:

> In the mid-1960s, British activists gravitated to Laing’s ideas, arguing that schizophrenia was more “properly human”, in a world of hydrogen bombs, than conventional definitions of sanity…Laing argued that schizophrenia was not a breakdown but a breakthrough. By the 1970s, Laing took the position of Huxley’s The Doors of Perception, that schizophrenia was a form of sanity, not insanity. Laing’s position increasingly became a political attack on Western society, and then morphed once again, rejecting the idea of schizophrenia by declaring it as hypersanity. Eventually Laing’s celebrity led him to India and drug abuse, and he became a shell of his former self.

Well then.

Around the time all this was going on, the ACLU was launching an attack of its own on the psychiatric system.

Most of what they were saying sounded good – make sure people only get committed if the courts are absolutely sure they’re insane, make sure that they have all of their rights even within the psychiatric hospital – but Cramer references internal memos and discussions purporting to show that the ACLU’s real goal was to make psychiatric commitment so bureaucratically difficult that nobody would ever do it, thus freeing the mentally ill from their oppressors and destroying the psychiatric system. The courts were sympathetic to their cases and established several new rights and standards that made committing people exceptionally difficult.

In exchange, the opponents of institutions promised community treatment. Everybody agrees that community treatment was a good idea. The implementation left a lot to be desired. First, as always, they were seriously underfunded. Second, even the ones that had enough money quickly found that creating outpatient psychiatric centers is fundamentally geographically difficult. Schizophrenics are not known for their ability to go places on an organized schedule, nor for their access to good consistent transportation. The great advantage of the old asylums was that all of the schizophrenics were in one convenient location for the mental health workers to treat. When the new community treatment centers were set up, they tended to serve any schizophrenics who might live within a few blocks of them, and all the rest never made it to their appointments. Third, as per Cramer most of the people operating these new community centers were Sixties Leftists who decided that instead of the “bandaid solution” of actually treating mentally ill people, their *real* job was to cut out mental illness at the root by protesting capitalism and racism:

> One of the officials of the CMHC [Community Mental Health Centers] program later admitted that the CMHCs

> “were not equipped to deal” with the chronically mentally ill, who were about to be released in large numbers from state mental hospitals. The belief that mental hospitals caused mental illness, or at least made the mentally ill worse off than they were before, combined with an idealized view of how caring communities would be for the severely mentally ill. The activists and bureaucrats who wrote the CMHC regulations were about to start the release of mental patients into caring communities which for the most part **did not exist**. As one of those involved later admitted, “We were federal bureaucrats on an NIMH campus talking about the community, but really from some conceptual level as opposed to hands-on experience.”
>
> If CHMCs were not primarily serving the chronically mentally ill, then whom were they serving? Two especially notorious examples were Lincoln Hospital Mental Services in New York City and Temple University Community Mental Health Center in Philadelphia. In both cases, the belief that mental illness was somehow an expression of class struggle meant that broader social and political causes – such as landlord/tenant relations, poverty, and oppression – became significant activities of the staff. Racial and ethnic tensions within the staff destroyed both CMHCs, with threats of violence, sit-ins, VietCong flags, posters of Che Guevera and Malcolm X as symbols of the fight.

In the late sixties and early seventies all of these things came together. Psychiatrists wanted to focus on healthy people who were much more pleasant to talk to. Pharmaceutical companies insisted that their new wonder drugs could cure psychosis. Activists wanted to destroy the psychiatric system. Judges were making it much more difficult to commit anybody. And community mental health centers were trying to pick up the slack. The result was the deinstitutionalization strategy called “closing the front door and opening the back door” – that is, making new commitments more difficult, and accelerating the pace at which psychotics already in institutions could be discharged to the new community treatment programs (it didn’t hurt that syphilis had been cured a few decades earlier and the last few chronically insane syphilitics were dying off as well). This went exactly according to plan, the institutionalized population shrunk and shrunk throughout the seventies, and by the time Reagan decided to close the last few psychiatric institutions there wasn’t much left to close down.

III.

Needless to say, Cramer opposes most of these developments. He makes his antideinstitutionalization argument in several parts. But first, some things he doesn’t argue.

Cramer is pretty quick to admit the institutions had their problems:

> Many [psychiatric hospitals] remained “snake pits”, to borrow the title of Mary Jane Ward’s very popular 1946 novel about mental hospitals. The American Psychiatric Association created the Central Inspection Board in 1947 to evaluate existing mental hospitals in the United States and Canada. The results were not encouraging. By 1953, it had evaluated 45 hospitals, approved two, given ten a “contingent approval”, and disapproved the rest.

The book frankly discusses the “regimented, often hopeless conditions of state mental hospitals”, talks about a hospital in Alabama where “care was worse than simply inadequate: one psychiatrist for 5000 patients; astonishingly low funding for clothing, food and upkeep of the buildings”, studies showing that institutions never actually got patients’ signatures on the forms that were supposed to waive their rights to court hearings. It describes the case of Edna Long, who was hospitalized for “public drunkenness” and

> permanently hospitalized in 1952. As Ennis tells the tale, Long received no treatment during the next fifteen years, but was kept busy working at menial jobs in the hospital. After the death of her husband in 1960, the state hospital had her declared incompetent, and seized her assets to pay for her care. Then, they put what assets remained under the management of an attorney, who made a bit of money from reducing the value of her estate by 86% (according to Ennis, a common practice at the time in New York). Once Long had become too physically ill to continue working, the hospital suddenly found her “competent to manage her own affairs” and released her, to a life of elderly poverty. Most of the money that she and her husband had accumulated had been consumed by attorneys supposedly protecting her assets.

Against this tale of woe, Cramer can say only that it “leads me to wonder if there was a bit more to the story”. Judging from my own conversations with patients and nurses who used to live in / work at these hospitals – who generally report similar stories – I doubt there was.

So what is this book’s argument against deinstitutionalization?

First, it points out that very many deinstitutionalized schizophrenics slipped through the community mental health system and never got further treatment. This was in part due to the problems with CMHCs – poor funding, difficult to get to, sometimes not that interested in mental health at all (though they got a lot better after the Sixties). But it was also due to schizophrenics just generally not being too interested in engaging with the psychiatric system (especially, one might imagine, the ones who had just gotten out of institutions) and no one being able to make them. I 100% acknowledge that this argument is correct.

Second, it points out that many untreated or unsuccessfully treated schizophrenics ended up homeless on the street.

> “Of 179 homeless men and women who received psychiatric examinations in a Philadelphia shelter in 1981, 40% were found to have “major mental disorders”. One-third of those examined were diagnosed as schizophrenic, and another one-fourth had a primary diagnosis of substance abuse. A Boston shelter study of 78 residents in 1983 again found that 40% had major mental disorders, and another 51% had less severe psychiatric problems…a survey of 345 subjects seeking food assistance in 1983 Phoenix found that about 30% had spent some time in a mental institution.

A quick Fermi calculation from the book’s numbers suggests that maybe 10% of schizophrenics are currently homeless. Again, I 100% acknowledge that this argument is correct and that these are probably accurate statistics about the percent of the homeless who are mentally ill.

Third, it points out that many of these people die of preventable causes. Many freeze to death on cold nights. Cramer notes that deinstituionalization corresponded with a doubling of US hypothermia deaths (although never above 1/500,000 people = 500 people per year) and that anecdotal evidence suggests many of these were mentally ill. Still others commit suicide or otherwise die of their own predictable poor choices. For example:

> In another case, a woman with anorexia was admitted to a hospital after she had been involved in a family disagreement and refused to eat. She had lost a great deal of weight but refused to submit to a psychiatric exam, and since a judge felt her condition was not dangerous in an immediate sense, she was allowed to go home. She died from starvation three weeks later.

Again, I 100% acknowledge this sort of thing probably happened and happens quite often.

Fourth, it says that these people are generally weird and scary and can push everyone else out of public places. Many, for example, end up in libraries, the rare sort of public place you can enter without an admission charge. He tells the story of some such library “patrons”:

> Mick is having a bad day. He hasn’t misbehaved but sits and stares, glassy-eyed. This is usually the prelude to a seizure. His seizures are easier to deal with than Bob’s, for instance, because he usually has them while seated and so rarely hits his head and bleeds, nor does he ever soil his pants. Bob tends to pace restlessly all day and is often on the move when, without warning, his seizures strike. The last time he went down, he cut his head. The staff has learned to turn him over quickly after he hits the floor, so that his urine does not stain the carpet.
>
> A friend worked at the main branch of the Santa Rosa, California public library in the 1980s and 1990s. She was awash in similar stories of mentally ill people who would urinate in the corners of the library, make

> frightening noises, sleep at the tables, and generally create an environment that would have been grounds for at least expulsion, if not arrest and commitment, in any American public library in 1960. The library staff was obligated to work with such “patrons” until their actions became clearly criminal. She recounted what happened when she observed that one of these mentally ill patrons was sitting at a table with his pants down to his knees. Her supervisor was obligated by library rules to attempt to first resolve the problem without the police. He approached this exposed “patron” and diplomatically asked “Sir, are you appropriately attired for the library?”
>
> Why was it necessary for librarians to take such a kid glove approach? Attempts to resolve behavioral problems led to lawsuits, such as happened in Morristown, New Jersey. The behavior and offensive smell of a homeless person named Kreimer led to the adoption of a code of conduct prohibiting loitering, “unnecessary staring”, following others around the library, and requiring those using the library to conform to community standards of cleanliness. The ACLU filed suit against this discriminatory code. At trial, Judge Sarokin ruled that the rules were discriminatory, and that the ban on annoying other patrons violated Kreimer’s right to freedom of speech.

This ruling was later overturned on appeal, but apparently the whole series of lawsuits had cost so much money that the mere possibility of a suit from the ACLU led libraries to adopt a policy of tolerating everyone, no matter how filthy, loud, or threatening they might be. Once again, this sounds like the sort of thing that probably happens and I have no doubt the book is telling the truth. One need not blame the homeless and mentally ill for their behavior to acknowledge that this is a potential argument in favor of institutionalizing people so they have less inconvenient places than libraries to spend their time.

Fifth, Cramer argues the deinstitutionalized mentally ill are responsible for a disproportionate amount of crime, including some of the flashiest mass shootings. He notes that of a *New York Times* list of the 100 most famous rampage killers, 47 had a past history of mental health problems, and 20 had been previously institutionalized. Former psychiatric inpatients are 55 times more likely than the general population to be arrested for murder, and about five times more likely to be arrested for lesser crimes like robbery, rape, and aggravated assault. He cites Bernard Harcourt's work showing a strong negative correlation between the institutionalization rate and the crime rate – although as I've mentioned before, I think these numbers are seriously off and that this is more likely related to lead levels. Nevertheless, the general point that deinstitutionalized mentally ill are at high risk of criminality stands – although Cramer admits that the overwhelming majority will never get in trouble.

IV.

So I agree with almost all of Cramer's empirical claims. Yes, many deinstitutionalized schizophrenics are not receiving adequate treatment. Yes, many are homeless, either broke or unable to manage their disability money in a rational way. Yes, many are dying of preventable causes like freezing to death. Yes, many are going around public places and threatening people and freaking people out. And yes, many of them (though by no means most) are committing terrible crimes. So how can I disagree with his assessment that deinstitutionalization was a mistake, that Reagan and the hippies and Thomas Szasz were in the wrong, and we need to bring back a strong system of long-term state-run psychiatric hospitals?

Well, let me ask a related question. Should we round up everybody from the ghetto and stick them in prison? This policy would have a number of advantages. Many people in the ghetto are desperately poor and living in terrible conditions. Many die before their time. They often make middle-class people who come across them profoundly uncomfortable. And their crime rate is much higher than that of the non-ghetto population. All the advantages of institutionalizing the mentally ill also apply to institutionalizing people in ghettos.

Against this we have a counterbalancing consideration: it is a horrible idea and it would be really mean and everybody involved would hate it and you have no right to even consider such a thing. This is also how I feel about institutionalizing the mentally ill.

First, a digression. Many of the people Cramer mentions – his brother Ron, his case studies of homeless people who freeze to death on the streets, some of the mass killers – have in fact been institutionalized. Ron was institutionalized the better part of a dozen times. Usually they’re in the hospital for a few weeks to a few months, stabilized on medications, and then released. After their release for one reason or another they come off their medications and then experience whatever catastrophe makes them suitable for inclusion in this book.

So if we want to solve all of the problems Cramer brings up – homelessness, crime, library-bothering, etc – we can’t do it by just having people in institutions for a few months or a few years. The second they set foot out of a hospital in this counterfactual world, they’ll encounter the same problems they encounter in our real one. In other words, this isn't really about treatment, at least in the sense of "we need better commitment laws so hospitals can treat patients and then help them reintegrate into society." What Cramer is talking about, if he's really serious about solving these issues, is lifetime institutionalization.

Making someone spend their entire life in an institution is a pretty big deal, especially if, as Cramer freely admits, they often include "regimented, hopeless conditions" where "care is worse than simply inadequate". Sometimes we as a society decide that criminals need to spend their entire life in an unpleasant institution because they murdered somebody or something, but it seems excessive to say that somebody should be institutionalized for life merely because they are from a population that has a disproportionate (though still not *high*!) risk of committing some kind of crime in the future. Once again, if we were in *that* business we should just imprison people for being born in bad neighborhoods. Yes, it's a tragedy when an anorexic starves themselves to death. But should we lock up all anorexics forever to prevent that one case?

What about the humanitarian argument that we need to institutionalize schizophrenics so that they don't end up starving on the street? Here we get into some really thorny moral issues. I tend to go by revealed preferences – schizophrenics have voted with their feet to not be in mental hospitals. If there were voluntary mental hospitals, and schizophrenics chose to live in them, that would be great and I would support them in that choice. If you are contradicting schizophrenics' expressed preference that they prefer not being in mental hospitals – freezing weather and all – to being in mental hospitals, then you have no right to say you're doing it for their own sake.

I can see a counterargument: psychotic people are not very good at making decisions. What if they would be happier in a nice warm institution, but they are too crazy to realize this? For example, maybe when the person asks them “Would you like to go to the hospital?” they believe that person is a CIA spy who will be leading them to the firing chamber instead?

I agree this is a possibility and a strong argument. Against it I can only say that many of the psychotic people who don’t want to go to mental hospitals are dragged there anyway, and usually continue to not want to be in the mental hospital after they get there and learn what it is like.

An example from my own life might serve to clarify the odd mix of rational and irrational decision-making I think characterizes these choices. When I was a child, my OCD was much worse. I would do things like close every shutter in my room nine times. I won’t say this was the *most* rational thing to be doing. But if you with your superior rationality had come in and chained me to my bed so that I couldn’t close my shutters, I would have spent the entire night freaking out because my shutters hadn’t been closed the appropriate nine times and that meant the world was unbearably wrong. *Given* a mind that will freak out for a whole night if the shutters aren’t closed, and supposing for a second that curing the underlying OCD is not an option, then spending a minute closing the shutters is a perfectly rational decision. Likewise, given the weird collections of fears and sensitivities that characterize the typical psychotic, staying out of a psychiatric hospital may be a perfectly rational decision. And this is even granting the extremely dubious premise that the hospital is not abusive, is not disgusting, is not dictatorial, doesn’t involve drugs with terrible side effects, or any of the other hundred ways a psychiatric hospital can be bad even when your judgment is perfectly intact.

I recently learned many of the homeless in nicer cities have laptops. This makes sense – laptops are really cheap these days, way cheaper than houses, and you can carry them around with you on your back. Psychiatric hospitals, in contrast, do not have laptops. Even if you own a laptop, you may not bring it in, since it is theoretically Usable As A Weapon. You may not bring a cell phone, a tablet or any other form of communication device. Some of the very nice psychiatric hospitals, including the one I work at, have a single computer for thirty residents, which you may use for fifteen minutes a day, with a nurse watching you the whole time to make sure you don't go on any sites that seem likely to make you upset or emotional. This fact alone makes me, personally, with my as far as I can tell totally intact mind, prefer the thought of homelessness to the thought of lifetime institutionalization. My computer is my only lifeline to most of my friends and the only way I have to express myself, and the thought of trading that away just so I can have a warm bed seems – pardon the expression – insane.

And for me it's the computer. For other people it's other things, reasonable by our standards or not. A few weeks ago I was woken up by a call in the middle of the night. A newly admitted patient at the mental hospital where I work was making a scene. She had this *thing* about using her special pillowcase, and pillowcases weren't on the hospital's Special List Of Things It Is Okay To Bring In. Sheets? Absolutely. Blankets? Totally fine. Pillows? Knock yourself out. But nobody had thought about pillowcases, so they were officially banned. And I made it to the ward, still half-asleep, and for a second I couldn't figure out who was the crazy person, the woman making a William Wallace-esque stand for the right to bring her pillowcase into a hospital, or the woman telling her absolutely not, because it wasn’t on the Magic List. Eventually I asked the nurse if maybe we could just sort of pretend the pillowcase was a very small sheet, and she said that if I specifically ordered her to do so she wasn’t able to contradict a doctor’s orders, and the problem was solved. By which I mean that by the time she figured out something else she needed, my shift would be over and it would be someone else’s problem. Because *everything in a mental hospital is like this all the time*.

So am I okay with this causing some people to freeze to death? Yes. I don’t think we can be sufficiently sure that institutionalizing schizophrenics is in their own best interest to overcome the burden of proof necessary for overriding someone’s revealed preferences. So if respecting people’s revealed preferences mean some of them go homeless or die, so be it. God help us if we ever systematically decide that people should not be allowed their freedom if the decision carries any discomfort or risk.

I want to stress just how important a decision this is. Back before deinstitutionalization, there were about 500,000 people in US psychiatric institutions, with varying degrees of permanency. Given the increase in the population and mental illness, I expect there are up to a million potentially institutionalizable individuals today. If institutionalization costs the average psychotic 1/3 of a QALY per year (eg moving from poverty to imprisonment on this table) then we’re taking away 300,000 QALYs every year indefinitely. On the other hand, if institutionalization were better for psychotics, they could potentially *gain* a similar number of QALYs. That makes policy decisions in this area potentially more important than crime, more important than terrorism, more important than education, potentially more important than everything except health care, not starting too many wars, and mass incarceration full stop. These kinds of decisions are the ones you want to be really, really sure about. So far, nothing in *My Brother Ron* has given me the level of certainty I would need.

I agree kids should have a right to use public libraries without having mentally ill people urinate on them or scream at them. I think the solution in this case is to tell the ACLU to take a chill pill and then let librarians enforce common-sense decency rules, not to lock up a million people for the rest of their lives.

V.

So that leaves the question – what *do* we do with all of these psychotic people starving on the street? Saying “leave them alone” is all nice and well, but what if they start seeming violent or threatening? Do we leave them alone until the point at which they commit a major crime and they end up in prison for the rest of their lives? What if they’re clearly acting recklessly and about to die? What if we have evidence (maybe from past experience) that they would prefer to be sane and medicated but they’re too far gone to realize it?

The book itself mentions my preferred answer to this conundrum: involuntary outpatient commitment (IOC). This is exactly what it sounds like. If you, let’s say, start trespassing on government property and yelling at police officers (a common way for mentally ill people to come to the attention of the system), and you get brought before a sympathetic judge who wants to help you and doesn’t want to lock you up but would prefer you not do that anymore, he can order an outpatient commitment. This means you’re legally required to see a psychiatrist every so often and maybe get injected with long-acting antipsychotic medication (usually once per month, although I think they’ve recently invented a once-every-three-months version now).

I have seen psychotic patients involved in such programs and they usually do very well. They get the same level of treatment they would in a psychiatric hospital, people will come hunt them down to make sure they don’t miss their appointments or medication dosings, and in the interim they can live wherever they want in whatever conditions they want. If the medications work, which they usually do, then they are hopefully clear-headed enough to either hold down a job or use their disability payments responsibly. If they can’t do that, then it’s probably for the same reason that normal poor people can’t, and nobody says *they* need to be institutionalized.

Cramer notes that people in IOC programs have half the suicidality rates, half the crime rates, and “substantial reductions in hospitalization, homelessness, arrest, and incarceration.” They are half as likely to be hospitalized, half as likely to be *victims* of crimes, and “enjoy improved quality of life”.

This isn’t as good as, say, one-*tenth* the suicidality and hospitalization rates would be. But psychiatry isn’t a discipline with very many miracles. Sometimes the drugs work and sometimes they don’t. Long-term psychotics are notoriously difficult to treat and this is probably about as well as they would be doing in a long-term institution anyway.

Cramer brings this up as part of his political polemic – apparently the same hippies who oppose everything else opposed IOCs, so their success is part of the Grand Narrative Of Hippies Being Proven Wrong. I like hippie-bashing as much as anyone else, but I don’t understand why he doesn’t take this further, say that this is the alternative to reinstitutionalization that he secretly knows we need. He points out that the main reason IOCs are underused is that psychiatrists don’t know about them – I would add that at least in my county there isn’t enough funding to refer enough patients to the program and monitor their medication compliance and so on. But I guarantee you that publicizing the option to psychiatrists and expanding the program is a *lot* cheaper than reinstitutionalizing people would be.

(my hospital charges $1,000/day/inpatient, though goodness only knows how much of that insurance companies actually pay. Cramer notes that the prison system usually costs $50,000/year/mentally ill prisoner. My guess is that the costs of institutionalization are somewhere around that order of magnitude.)

So in my ideal world, psychotic people who aren’t bothering anybody can do what they want – preferably with the option of voluntary psychiatric hospitalization available, and with some pressure to at least try it once and get a feel for what it’s like. Psychotic people who *are* bothering other people can get outpatient treatment once every couple of months and remain medicated and monitored by professionals. Preferably there would also be some kind of concept of a psychiatric living will – that is, some way for people who are not yet mentally ill, or who are currently being managed on drugs, to express a wish to be stabilized if they ever become mentally ill so that they can make their long-run choices from a position of sanity.

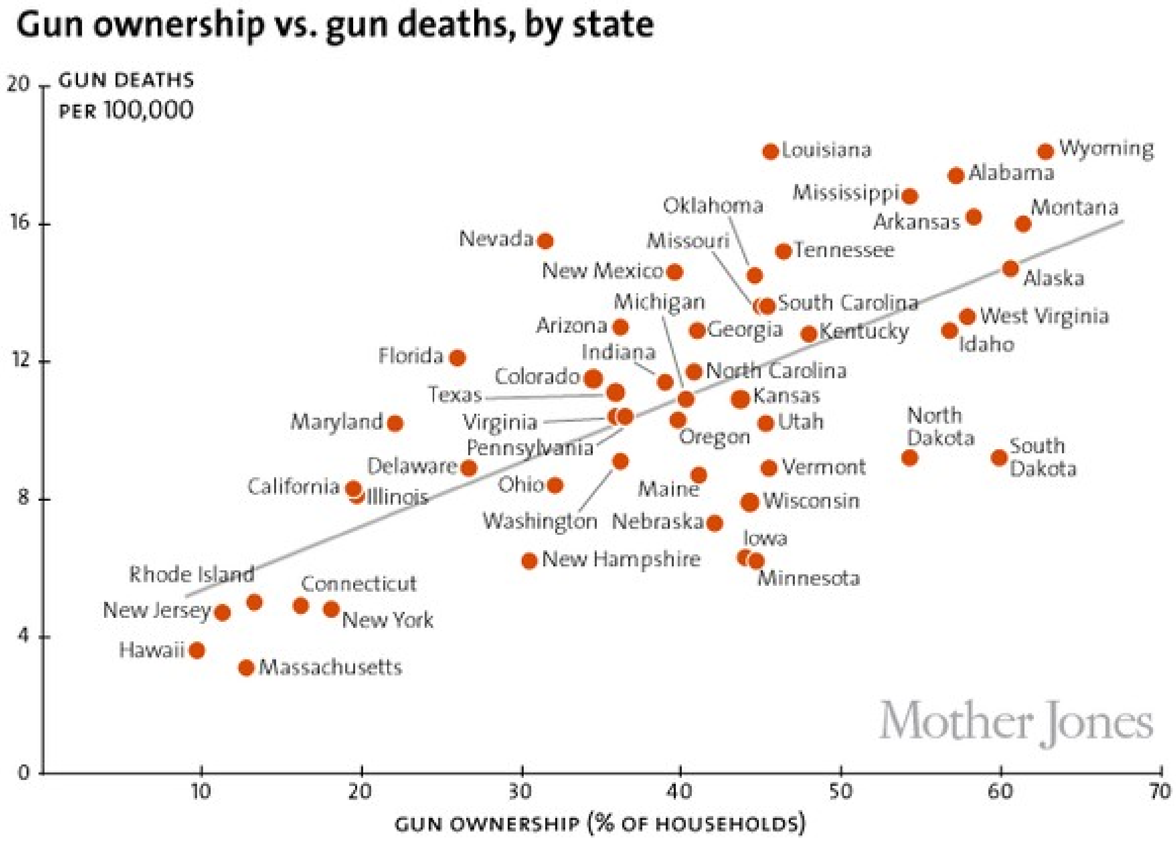
I acknowledge this is not the ideal world. I acknowledge there are some people who really need institutionalization – people who are constantly violent, who have zero concept of social rules and will scream at anyone they meet, people who are catatonic or need extraordinarily complicated medication regimens that can’t be handled in a normal environment. I’ve referred some of these people to involuntary long-term institutions (which still exist for these kinds of extreme situations), I don’t feel guilty at all, and in most cases I am pretty sure the general public would be pretty grateful to me if they knew the gory details.

But for a million people, most of whom aren’t bothering anybody and just want to be able to live a half-decent life outside the walls of a locked facility? There has to be a better solution than that.

Beware Regional Scatterplots

[Epistemic status: Not original, but worth mentioning]

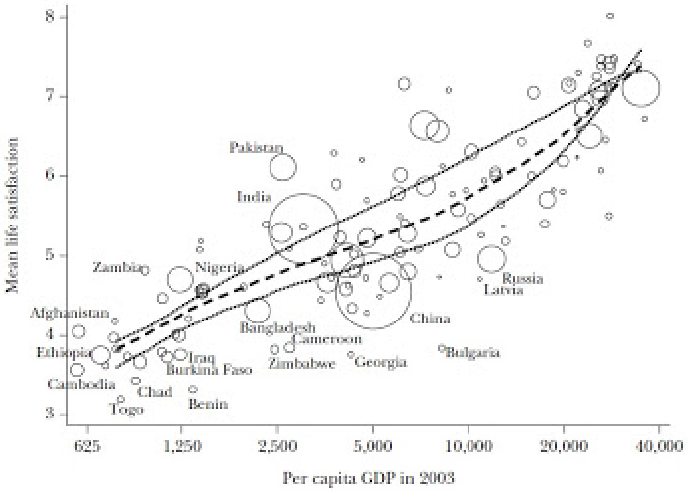
I’ve been using scatterplots of different states and countries a lot here lately. For example, this one in the discussion about guns:



And this one in the discussion about national happiness:

Figure 2

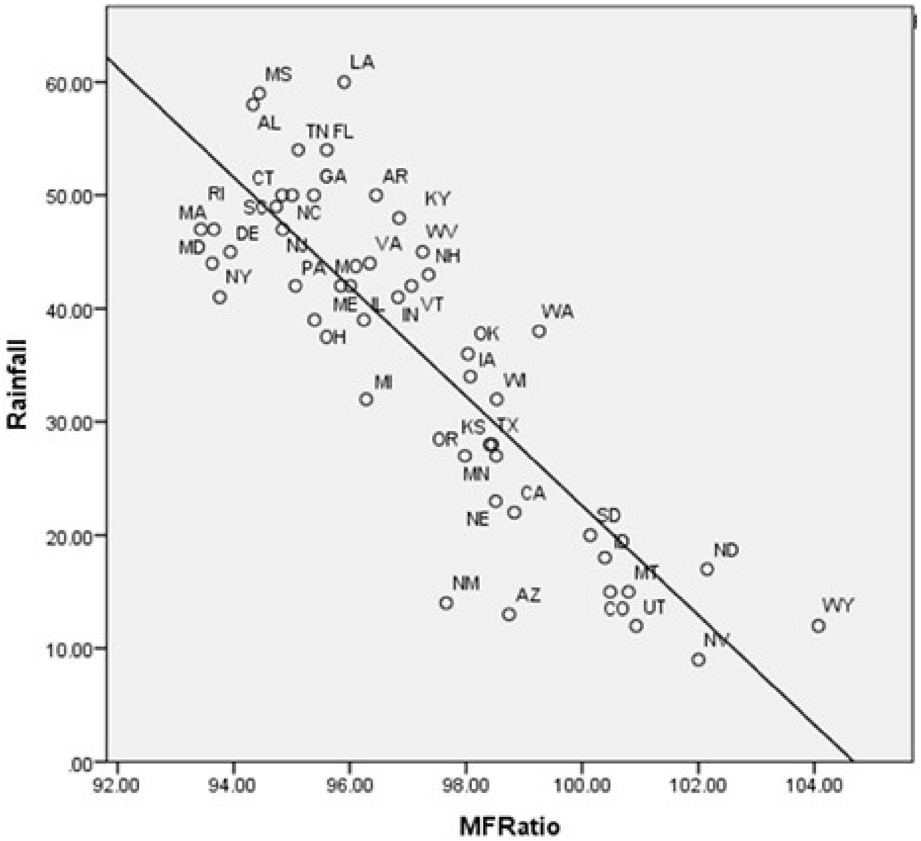
Each Doubling of GDP is Associated with a Constant Increase in Life Satisfaction



Source: Penn World Table 6.2.

Note: Each circle is a country, with diameter proportional to population. The scale on the x-axis is logarithmic. The middle line shows average life satisfaction for each level of per capita GDP while the outer two lines show the same thing, but for two age groups, ages 15 to 25—the upper line for most of the figure—and ages 60 and over—which is usually the lower line. GDP per capita in 2003 is measured in purchasing power parity chained dollars at 2000 prices.

Hopefully we already know that we should worry about confounders like income and race and those kinds of things when we’re looking at a graph like this. But recently I learned it’s even worse than that. Consider for example this:



This is the average yearly rainfall in the lower 48 US states vs. their gender balance (measured in number of men per 100 women). The correlation is about $r = 0.84$ ($p \leq 0.0001$), much higher than anyone's ever found between guns and crime, or income and happiness, or most other things people make regional scatterplots about. So what's going on? Do women cause rainfall? Does rain drive men away? Or is there some confounder that causes both rain and womanhood?

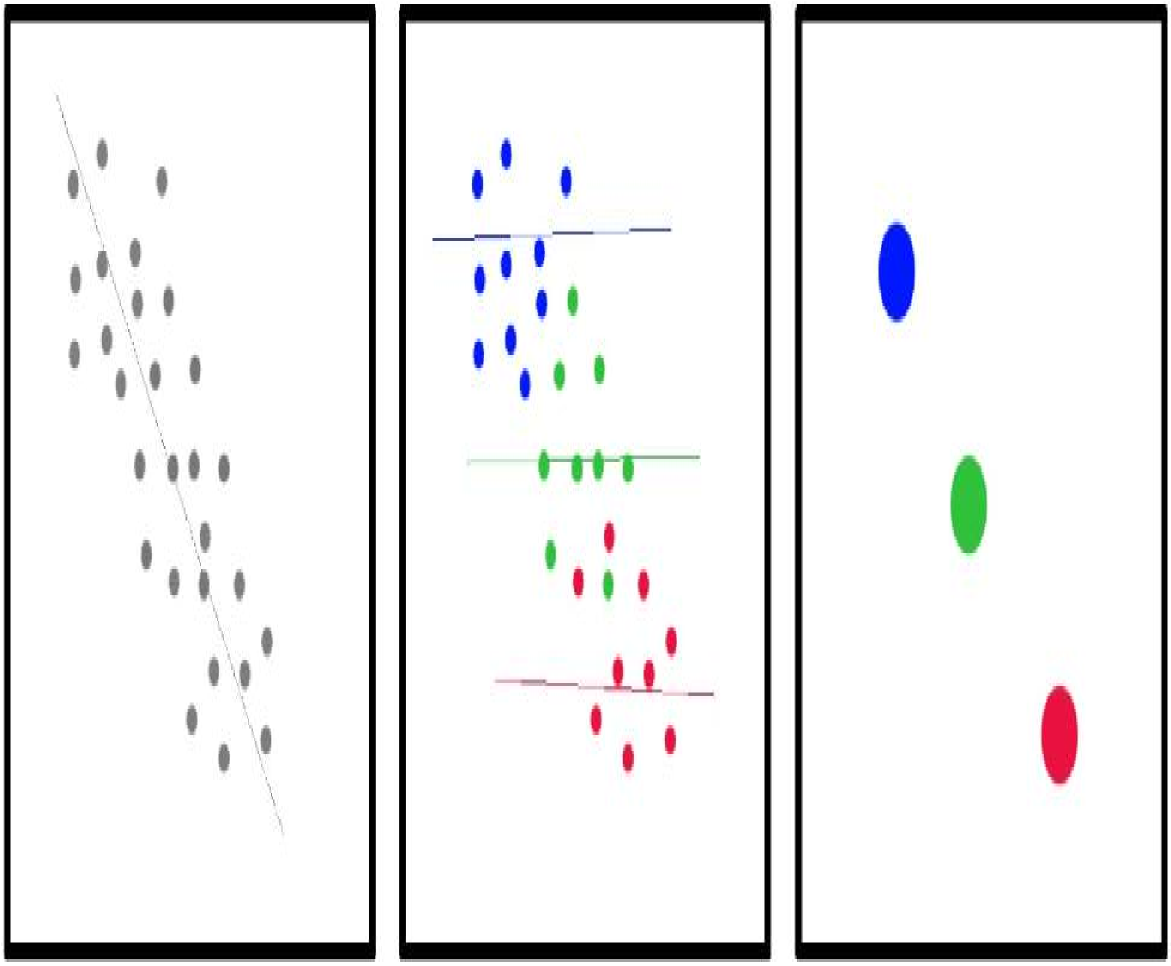
I don't think it's any of these things. I think it's a coincidence.

"But you said $p \leq 0.0001$! There are forty-eight data points and the fit is almost perfect! How could it be a coincidence?"

But I don't think there *are* forty-eight data points. I think there are three data points. For 48 data points to all lie the same line is very impressive. For three data points to all lie on the same line is much less so.

I think that Southern states have more women (probably because they have a higher male incarceration rate, and incarcerated men aren’t counted) and more rainfall. Mountainous western states have more men (probably because the jobs there tend to be in manly mining/forestry type industries) and are pretty dry. And other states are somewhere between those two extremes.

Within these regional categories the rainfall/gender relationship is random – on the scatterplot it would look like a circle. But *between* these three regional categories the rainfall/gender relationship is very strong, making the whole chart consist of three circles in a line. Crucially, because these are kind of amorphous circles and they blend into each other, you can’t tell that that’s what’s going on. Here’s my graphic depiction of this:



In the first box, the gray points show what looks like a very significant correlation – the further right you go on the horizontal axis, the further down you go on the vertical axis. The gray trend line confirms the strong relationship. In our US state example, this was a correlation between many women and high rainfall.

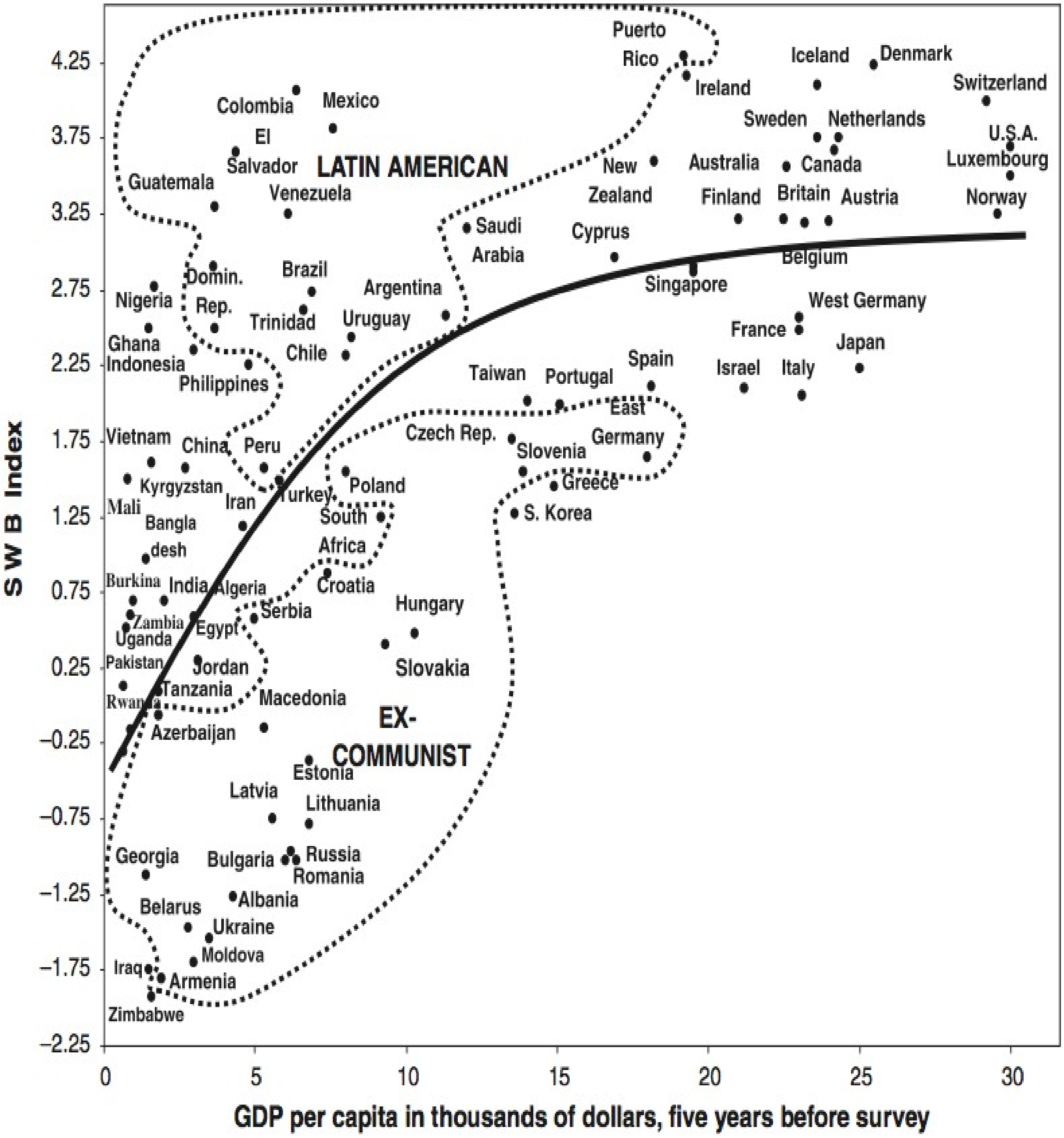
In the second box, the gray points are revealed to be grouped into three regions: blue, green, and red. In our US state example, blue is the female-skewed and rainy Southern states, red is the male-skewed and dry Western states, and green are all the other states with pretty average rainfalls and gender balances. Within each region, there's no relationship between rainfall and gender, as shown by the horizontal red, green, and blue trend lines.

In the third box, we see what I'd argue is the correct interpretation of the data. There are three big data points – the South, the West, and the Rest – and they do sort of form a line but nobody cares about a line between three data points.

If I go back to my statistics packet and repeat the rainfall/gender correlation with only three points – the Southern average, the Western average, and the Other average – I still get r = -0.84, but now p = 0.3. The statistics have no reason to think it's anything other than pure coincidence – and indeed, with that small a sample size, why *would* they?

I think *most* real studies are smart enough to control for this – although it's really hard to determine how exactly you should be doing that and leaves a lot of wiggle room for people who want to fudge their way to a preferred result. But basic scatterplots do *not* control for it, and so *almost every regional scatterplot is suspect*.

This is why I was happy to see the income/happiness correlation broken down further:



This makes it clear that the income/happiness relationship is primarily cluster-driven, with clusters of ex-Communist countries, Latin American countries, and Euro/Anglosphere countries (if you’re willing to do some more work, you can sort of make out clusters of African and Asian countries). None of these clusters show a strong income/happiness relationship except for the ex-Communist one, which suggests this might be the same kind of confounding as the rainfall/gender example above.

…unless I’m biased and reading too much into this. It’s really easy to change your conclusion just by changing your clusters. For example, if Puerto Rico counts as Latin American, then that creates a pretty impressive happiness/income relationship within that cluster. If it counts as Euro/Anglosphere – it’s part of the US, after all – then there is no income/happiness relationship within either cluster. So which is it?

Or what about this? I claim that the apparent income/happiness relationship within the ex-Communist countries is actually an artifact of Europeanness. The richest ex-Communist countries, like East Germany and the Czech Republic – are also the happiest only because they are closest to Western Europe, which is both happier and richer than the rest of the world. Likewise, the poorest ex-Communist countries, like Armenia and Georgia, are also the unhappiest only because they are the furthest and least Western European of the bunch.

Once we start going *there*, we can pretty much prove or disprove anything we want based on our own intuitions about how to group things. I am suspicious of this, but I’m also equally suspicious of *not* doing that – do you really just want to let it pass that Puerto Rico, the closest Latin American country to the Euro/Anglosphere cluster, is also politically Euro/Anglosphere?

Overall there is no good answer and I would recommend against drawing any strong causal conclusions from a scatterplot unless someone has very carefully addressed these concern.

EDIT: Inty gives a great (ie horrifying) example in the comments, and Theo Jones discusses more formal tests of spatial autocorrelation. Some people bring up the possibility that some of the rainfall/gender relationship is causal after all, since drier states will have less farming and be forced to turn to mining/forestry to support themselves; this is possible but probably doesn’t explain the whole relationship, and even if I’m wrong about this one the point is still important.

The Ideology Is Not The Movement

I.

Why is there such a strong Sunni/Shia divide?

I know the Comparative Religion 101 answer. The early Muslims were debating who was the rightful caliph. Some of them said Abu Bakr, others said Ali, and the dispute has been going on ever since. On the other hand, that was fourteen hundred years ago, both candidates are long dead, and there’s no more caliphate. You’d think maybe they’d let the matter rest.

Sure, the two groups have slightly different hadith and schools of jurisprudence, but how many Muslims even *know* which school of jurisprudence they’re supposed to be following? It seems like a pretty minor thing to have centuries of animus over.

And so we return again to Robbers’ Cave:

> The experimental subjects — excuse me, “campers” — were 22 boys between 5th and 6th grade, selected from 22 different schools in Oklahoma City, of stable middle-class Protestant families, doing well in school, median IQ 112. They were as well-adjusted and as similar to each other as the researchers could manage.
>
> The experiment, conducted in the bewildered aftermath of World War II, was meant to investigate the causes—and possible remedies—of intergroup conflict. How would they spark an intergroup conflict to investigate? Well, the 22 boys were divided into two groups of 11 campers, and —

— and that turned out to be quite sufficient.

The researchers' original plans called for the experiment to be conducted in three stages. In Stage 1, each group of campers would settle in, unaware of the other group's existence. Toward the end of Stage 1, the groups would gradually be made aware of each other. In Stage 2, a set of contests and prize competitions would set the two groups at odds.

They needn't have bothered with Stage 2. There was hostility almost from the moment each group became aware of the other group's existence: They were using our campground, our baseball diamond. On their first meeting, the two groups began hurling insults. They named themselves the Rattlers and the Eagles (they hadn't needed names when they were the only group on the campground).

When the contests and prizes were announced, in accordance with pre-established experimental procedure, the intergroup rivalry rose to a fever pitch. Good sportsmanship in the contests was evident for the first two days but rapidly disintegrated.

The Eagles stole the Rattlers' flag and burned it. Rattlers raided the Eagles' cabin and stole the blue jeans of the group leader, which they painted orange and carried as a flag the next day, inscribed with the legend “The Last of the Eagles”. The Eagles launched a retaliatory raid on the Rattlers, turning over beds, scattering dirt. Then they returned to their cabin where they entrenched and prepared weapons (socks filled with rocks) in case of a return raid. After the Eagles won the last contest planned for Stage 2, the Rattlers raided their cabin and stole the prizes. This developed into a fistfight that the

> staff had to shut down for fear of injury. The Eagles, retelling the tale among themselves, turned the whole affair into a magnificent victory—they'd chased the Rattlers "over halfway back to their cabin" (they hadn't).
>
> Each group developed a negative stereotype of Them and a contrasting positive stereotype of Us. The Rattlers swore heavily. The Eagles, after winning one game, concluded that the Eagles had won because of their prayers and the Rattlers had lost because they used cuss-words all the time. The Eagles decided to stop using cuss-words themselves. They also concluded that since the Rattlers swore all the time, it would be wiser not to talk to them. The Eagles developed an image of themselves as proper-and-moral; the Rattlers developed an image of themselves as rough-and-tough.

If the researchers had decided that the real difference between the two groups was that the Eagles were adherents of Eagleism, which held cussing as absolutely taboo, and the Rattlers adherents of Rattlerism, which held it a holy duty to cuss five times a day – well, that strikes me as the best equivalent to saying that Sunni and Shia differ over the rightful caliph.

II.

Nations, religions, cults, gangs, subcultures, fraternal societies, internet communities, political parties, social movements – these are all really different, but they also have some deep similarities. They're all groups of people. They all combine comradery within the group with a tendency to dislike other groups of the same type. They all tend to have a stated purpose, like electing a candidate or worshipping a deity, but also serve a *very* important role as impromptu social clubs whose members mostly interact with one another instead of outsiders. They all develop an internal culture such that members of the groups often like the same foods, wear the same clothing, play the same sports, and have the same philosophical beliefs as other members of the group – even when there are only tenuous links or no links at all to the stated purpose. They all tend to develop sort of legendary histories, where they celebrate and exaggerate the deeds of the groups' founders and past champions. And they all tend to inspire something like patriotism, where people are proud of their group membership and express that pride through conspicuous use of group symbols, group songs, et cetera. For better or worse, the standard way to refer to this category of thing is "tribe".

Tribalism is potentially present in all groups, but levels differ a lot even in groups of nominally the same type. Modern Belgium seems like an unusually non-tribal nation; Imperial Japan in World War II seems like an unusually tribal one. Neoliberalism and market socialism seem like unusually non-tribal political philosophies; communism and libertarianism seem like unusually tribal ones. Corporations with names like Amalgamated Products Co probably aren't very tribal; charismatic corporations like Apple that become identities for their employees and customers are more so. Cults are maybe the most tribal groups that exist in the modern world, and those Cult Screening Tools make good measures for tribalism as well.

The dangers of tribalism are obvious; for example, fascism is based around dialing a country's tribalism up to eleven, and it ends poorly. If I had written this essay five years ago, it would be be titled "Why Tribalism Is Stupid And Needs To Be Destroyed". Since then, I've changed my mind. I've found that I enjoy being in tribes as much as anyone else.

Part of this was resolving a major social fallacy I'd had throughout high school and college, which was that the correct way to make friends was to pick the five most interesting people I knew and try to befriend them. This almost never worked and I thought it meant I had terrible social skills. Then I looked at what everyone else was doing, and I found that instead of isolated surgical strikes of friendship, they were forming groups. The band people. The mock trial people. The football team people. The Three Popular Girls Who Went Everywhere Together. Once I tried "falling in with" a group, friendship became much easier and self-sustaining precisely because of all of the tribal development that happens when a group of similar people all know each other and have a shared interest. Since then I've had good luck finding tribes I like and that accept me – the rationalists being the most obvious example, but even interacting with my coworkers on the same hospital unit at work is better than trying to find and cultivate random people.

Some benefits of tribalism are easy to explain. Tribalism intensifies all positive and prosocial feelings within the tribe. It increases trust within the tribe and allows otherwise-impossible forms of cooperation – remember Haidt on the Jewish diamond merchants outcompeting their rivals because their mutual Judaism gave them a series of high-trust connections that saved them costly verification procedures? It gives people a support network they can rely on when their luck is bad and they need help. It lets you "be yourself" without worrying that this will be incomprehensible or offensive to somebody who thinks totally differently from you. It creates an instant densely-connected social network of people who mostly get along with one another. It makes people feel like part of something larger than themselves, which makes them happy and can (provably) improves their physical and mental health.

Others are more complicated. I can just make motions at a feeling that “what I do matters”, in the sense that I will probably never be a Beethoven or a Napoleon who is very important to the history of the world as a whole, but I can do things that are important within the context of a certain group of people. All of this is really good for my happiness and mental health. When people talk about how modern society is “atomized” or “lacks community” or “doesn’t have meaning”, I think they’re talking about a lack of tribalism, which leaves people all alone in the face of a society much too big to understand or affect. The evolutionary psychology angle here is too obvious to even be worth stating.

And others are entirely philosophical. I think some people would say that wanting to have a tribe is like wanting to have a family – part of what it means to be human – and demands to justify either are equally wrong-headed.

Eliezer thinks every cause wants to be a cult. I would phrase this more neutrally as “every cause wants to be a tribe”. I’ve seen a lot of activities go through the following cycle:

1. Let’s get together to do X
2. Let’s get together to do X, and have drinks afterwards
3. Let’s get together to discuss things from an X-informed perspective
4. Let’s get together to discuss the sorts of things that interest people who do X
5. Let’s get together to discuss how the sort of people who do X are much better than the sort of people who do Y.
6. Dating site for the sort of people who do X
7. Oh god, it was so annoying, she spent the whole date talking about X.
8. X? What X?

This can happen over anything or nothing at all. Despite the artificial nature of the Robbers’ Cove experiment, its groups are easily recognized as tribes. Indeed, the reason this experiment is so interesting is that it shows tribes in their purest form; no veneer of really being about pushing a social change or supporting a caliph, just tribes for tribalism’s sake.

III.

Scholars call the process of creating a new tribe “ethnogenesis” – Robbers’ Cave was artificially inducing ethnogenesis to see what would happen. My model of ethnogenesis involves four stages: pre-existing differences, a rallying flag, development, and dissolution.

Pre-existing differences are the raw materials out of which tribes are made. A good tribe combines people who have similar interests and styles of interaction *even before* the ethnogenesis event. Any description of these differences will necessarily involve stereotypes, but a lot of them should be hard to argue. For example, atheists are often pretty similar to one another even before they deconvert from their religion and officially become atheists. They’re usually nerdy, skeptical, rational, not very big on community or togetherness, sarcastic, well-educated. At the risk of going into touchier territory, they’re pretty often white and male. You take a sample of a hundred equally religious churchgoers and pick out the ones who are *most like the sort of people who are atheists* even if all of them are 100% believers. But there’s also something more than that. There are subtle habits of thought, not yet described by any word or sentence, which atheists are more likely to have than other people. It’s part of the reason why atheists *need* atheism as a rallying flag instead of just starting the Skeptical Nerdy Male Club.

The rallying flag is the explicit purpose of the tribe. It’s usually a belief, event, or activity that get people with that specific pre-existing difference together and excited. Often it brings previously latent differences into sharp relief. People meet around the rallying flag, encounter each other, and say “You seem like a kindred soul!” or “I thought I was the only one!” Usually it suggests some course of action, which provides the tribe with a purpose. For atheists, the rallying flag is not believing in God. Somebody says “Hey, I don’t believe in God, if you also don’t believe in God come over here and we’ll hang out together and talk about how much religious people suck.” All the atheists go over by the rallying flag and get very excited about meeting each other. It starts with “Wow, you hate church too?”, moves on to “Really, you also like science fiction?”, and ends up at “Wow, you have the same undefinable habits of thought that I do!”

Development is all of the processes by which the fledgling tribe gains its own culture and history. It’s a turning-inward and strengthening-of-walls, which transforms it from ‘A Group Of People Who Do Not Believe In God And Happen To Be In The Same Place’ to ‘The Atheist Tribe’. For example, atheists have symbols like that ‘A’ inside an atom. They have jokes and mascots like Russell’s Teapot and the Invisible Pink Unicorn. They have their own set of heroes, both mythologized past heroes like Galileo and controversial-but-undeniably-important modern heroes like Richard Dawkins and Daniel Dennett. They have celebrities like P.Z. Myers and Hemant Mehta. They have universally-agreed-upon villains to be booed and hated, like televangelists or the Westboro Baptist Church. They have grievances, like all the times that atheists have been fired or picked on by religious people, and all the laws about pledging allegiance to one nation under God and so on. They have stereotypes about themselves – intelligent, helpful, passionate – and stereotypes about their outgroups – deluded, ignorant, bigoted.

Dissolution is optional. The point of the previous three steps is to build a “wall” between the tribe and the outside, a series of systematic differences that let everybody know which side they’re on. If a tribe was never really that different from the surrounding population, stops caring that much about its rallying flag, and doesn’t develop enough culture, then the wall fails and the members disperse into the surrounding population. The classic example is the assimilation of immigrant groups like Irish-Americans, but history is littered with failed communes, cults, and political movements. Atheism hasn’t quite dissolved yet, but occasionally you see hints of the process. A lot of the comments around “Atheism Plus” centered around this idea of “Okay, talking about how there’s no God all the time has gotten boring, plus nobody interesting believes in God anymore anyway, so let’s become about social justice instead”. The parts of atheism who went along with that message mostly dissolved into the broader social justice community – there are a host of nominally atheist blogs that haven’t talked about anything except social justice in months. Other fragments of the atheist community dissolved into transhumanism, or libertarianism, or any of a number of other things. Although there’s still an atheist community, it no longer seems quite as vibrant and cohesive as it used to be.

We can check this four-stage model by applying it to the Sunni and Shia and seeing if it sticks.

I know very little about early Islam and am relying on sources that might be biased, so don’t declare a fatwa against me if I turn out to be wrong, but it looks like from the beginning there were big pre-existing differences between proto-Shia and proto-Sunni. A lot of Ali’s earliest supporters were original Muslims who had known Mohammed personally, and a lot of Abu Bakr’s earliest supporters were later Muslims high up in the Meccan/Medinan political establishment who’d converted only after it became convenient to do so. It’s really easy to imagine cultural, social, and personality differences between these two groups. Probably members in each group already knew one another pretty well, and already had ill feelings towards members of the other, without necessarily being able to draw the group borders clearly or put their exact differences into words. Maybe it was “those goody-goodies who are always going on about how close to Mohammed they were but have no practical governing ability” versus “those sellouts who don’t really believe in Islam and just want to keep playing their political games”.

Then came the rallying flag: a political disagreement over the succession. One group called themselves “the party of Ali”, whose Arabic translation “Shiatu Ali” eventually ended up as just “Shia”. The other group won and called itself “the traditional orthodox group”, in Arabic “Sunni”. Instead of a vague sense of “I wonder whether that guy there is one of those goody-goodies always talking about Mohammed, or whether he’s a practical type interested in good governance”, people could just ask “Are you for Abu Bakr or Ali?” and later “Are you Sunni or Shia?” Also at some point, I’m not exactly sure how, most of the Sunni ended up in Arabia and most of the Shia ended up in Iraq and Iran, after which I think some pre-existing Iraqi/Iranian vs. Arab cultural differences got absorbed into the Sunni/Shia mix too.

Then came development. Both groups developed elaborate mythologies lionizing their founders. The Sunni got the history of the “rightly-guided caliphs”, the Shia exaggerated the first few imams to legendary proportions. They developed grievances against each other; according to Shia history, the Sunnis killed eleven of their twelve leaders, with the twelfth escaping only when God directly plucked him out of the world to serve as a future Messiah. They developed different schools of hadith interpretation and jurisprudence and debated the differences ad nauseum with each other for hundreds of years. A lot of Shia theology is in Farsi; Sunni theology is entirely in Arabic. Sunni clergy usually dress in white; Shia clergy usually dress in black and green. Not all of these were deliberately done in opposition to one another; most were just a consequence of the two camps being walled off from one another and so allowed to develop cultures independently.

Obviously the split hasn't dissolved yet, but it's worth looking at similar splits that have. Catholicism vs. Protestantism is still a going concern in a few places like Ireland, but it's nowhere near the total wars of the 17th century or even the Know-Nothing-Parties of the 19th. Consider that Marco Rubio is Catholic, but nobody except Salon particularly worries about that or says that it will make him unsuitable to lead a party representing the interests of very evangelical Protestants. Heck, the same party was happy to nominate Mitt Romney, a Mormon, and praise him for his "Christian faith". Part of it is the subsumption of those differences into a larger conflict – most Christians acknowledge Christianity vs. atheism to be a bigger deal than interdenominational disputes these days – and part of it is that everyone of every religion is so influenced by secular American culture that the religions have been reduced to their rallying flags alone rather than being fully developed tribes at this point. American Sunni and Shia seem to be well on their way to dissolving into each other too.

IV.

I want to discuss a couple of issues that I think make more sense once you understand the concept of tribes and rallying flags:

1. Disability: I used to be very confused by disabled people who insist on not wanting a “cure” for their condition. Deaf people and autistic people are the two classic examples, and sure enough we find articles like Not All Deaf People Want To Be Cured and They Don’t Want An Autism Cure. Autistic people can at least argue their minds work differently rather than worse, but being deaf seems to be a straight-out disadvantage: the hearing can do anything the deaf can, and can hear also. A hearing person can become deaf at any time just by wearing earplugs, but a deaf person can’t become hearing, at least not without very complicated high-tech surgeries.

When I asked some deaf friends about this, they explained that they had a really close-knit and supportive deaf culture, and that most of their friends, social events, and ways of relating to other people and the world were through this culture. This made sense, but I always wondered: if you were able to hear, couldn’t you form some other culture? If worst came to worst and nobody else wanted to talk to you, couldn’t you at least have the Ex-Deaf People’s Club?

I don’t think so. Deafness acts as a rallying flag that connects people, gives them a shared foundation to build culture off of, and walls the group off from other people. If all deaf people magically became able to hear, their culture would eventually drift apart, and they’d be stuck without an ingroup to call their own.

Part of this is reasonable cost-benefit calculation – our society is so vast and atomized, and forming real cohesive tribes is so hard, that they might reasonably expect it would be a lot of trouble to find another group they liked as much as the deaf community. But another part of this seems to be about an urge to cultural self-preservation.

2. Genocide: This term is kind of overused these days. I always thought of it as meaning literally killing every member of a certain group – the Holocaust, for example – but the new usage includes “cultural genocide”. For example, autism rights advocates sometimes say that anybody who cured autism would be committing genocide – this is of course soundly mocked, but it makes sense if you think of autistic people as a tribe that would be dissolved absent its rallying flag. The tribe would be eliminated – thus “cultural genocide” is a reasonable albeit polemical description.

It seems to me that people have an urge toward cultural self-preservation which is as strong or stronger as the urge to individual self-preservation. Part of this is rational cost-benefit calculation – if someone loses their only tribe and ends up alone in the vast and atomized sea of modern society, it might take years before they can find another tribe and really be at home there. But a lot of it seems to be beyond that, an emotional certainty that losing one’s culture and having it replaced with another is not okay, any more than being killed at the same time someone else has a baby is okay. Nor do I think this is necessarily irrational; locating the thing whose survival you care about in the self rather than the community is an assumption, and people can make different assumptions without being obviously wrong.

3. Rationalists: The rationalist community is a group of people (of which I’m a part) who met reading the site Less Wrong and who tend to hang out together online, sometimes hang out together in real life, and tend to befriend each other, work with each other, date each other,

and generally move in the same social circles. Some people call it a cult, but that’s more a sign of some people having lost vocabulary for anything between “totally atomized individuals” and “outright cult” than any particular cultishness.

But people keep asking me what exactly the rationalist community *is*. Like, what is the thing they believe that makes them rationalists? It can’t just be about being rational, because loads of people are interested in that and most of them aren’t part of the community. And it can’t just be about transhumanism because there are a lot of transhumanists who aren’t rationalists, and lots of rationalists who aren’t transhumanists. And it can’t just be about Bayesianism, because pretty much everyone, rationalist or otherwise, agrees that is a kind of statistics that is useful for some things but not others. So what, exactly, is it?

This question has always bothered me, but now after thinking about it a lot I finally have a clear answer: rationalism is the belief that Eliezer Yudkowsky is the rightful caliph.

No! Sorry! I think “the rationalist community” is a tribe much like the Sunni or Shia that started off with some pre-existing differences, found a rallying flag, and then developed a culture.

The pre-existing differences range from the obvious to the subtle. A lot of rationalists are mathematicians, programmers, or computer scientists. The average IQ is in the 130s. White men are overrepresented, but so are LGBT and especially transgender people. But there’s more. Nobody likes the Myers-Briggs test, but I continue to find it really interesting that rationalists have some Myers-Briggs types (INTJ/INTP) at ten times the ordinary rate, and other types (ISFJ/ESFP) at only one one-hundredth the ordinary rate. Myers-Briggs doesn’t cleave reality at its joints, but if it measures anything at all about otherwise hard-to-explain differences in thinking styles, the rationalist community heavily selects for those same differences. Sure enough, I am *constantly* running into people who say “This is the only place where I’ve ever found people who think like me” or “I finally feel understood”.

The rallying flag was the Less Wrong Sequences. Eliezer Yudkowsky started a blog (actually, borrowed Robin Hanson’s) about cognitive biases and how to think through them. Whether or not you agreed with him or found him enlightening loaded heavily on those pre-existing differences, so the people who showed up in the comment section got along and started meeting up with each other. “Do you like Eliezer Yudkowsky’s blog?” became a useful proxy for all sorts of things, eventually somebody coined the word “rationalist” to refer to people who did, and then you had a group with nice clear boundaries.

The development is everything else. Obviously a lot of jargon sprung up in the form of terms from the blog itself. The community got heroes like Gwern and Anna Salamon who were notable for being able to approach difficult questions insightfully. It doesn’t have much of an outgroup yet – maybe just bioethicists and evil robots. It has its own foods – MealSquares, that one kind of chocolate everyone in Berkeley started eating around the same time – and its own games. It *definitely* has its own inside jokes. I think its most important aspect, though, is a set of shared mores – everything from “understand the difference between ask and guess culture and don’t get caught up in it” to “cuddling is okay” to “don’t misgender trans people” – and a set of shared philosophical assumptions like utilitarianism and reductionism.

I’m stressing this because I keep hearing people ask “What is the rationalist community?” or “It’s really weird that I seem to be involved in the rationalist community even though I don’t share belief X” as if there’s some sort of necessary-and-sufficient featherless-biped-style ideological criterion for membership. This is why people are saying “Lots of you aren’t even singularitarians, and everyone agrees Bayesian methods are useful in some places and not so useful in others, so what is your community even *about*?” But once again, it’s ~~about Eliezer Yudkowsky being the rightful caliph~~ it’s not necessarily *about* anything.

If you take only one thing from this essay, it’s that communities are best understood not logically but historically. If you want to understand the Shia, don’t reflect upon the true meaning of Ali being the rightful caliph, understand that a dispute involving Ali initiated ethnogenesis, the resulting culture picked up a bunch of features and became useful to various people, and now here we are. If you want to understand the rationalist community, don’t ask exactly how near you have to think the singularity has to be before you qualify for membership, focus on the fact that some stuff Eliezer Yudkowsky wrote led to certain people identifying themselves as “rationalists” and for various reasons I enjoy dinner parties with those people about 10000% more interesting than dinner parties with randomly selected individuals.

nostalgebraist actually summed this up really well: “Maybe the real rationalism was the friends we made along the way.” Maybe that’s the real Shia Islam too, and the real Democratic Party, and so on.

4. Evangelical And Progressive Religion: There seems to be a generational process, sort of like Harold Lee's theory of immigrant assimilation, by which religions dissolve. The first generation believes everything literally. The second generation believes that the religion might not be literally true, but it's an important expression of universal values and they still want to follow the old ways and participate in the church/temple/mosque/mandir community. The third generation is completely secularized.

This was certainly my family's relationship with Judaism. My great-great-grandfather was so Jewish that he left America and returned to Eastern Europe because he was upset at American Jews for not being religious enough. My great-grandfather stayed behind in America but remained a very religious Jew. My grandparents attend synagogue when they can remember, speak a little Yiddish, and identify with the traditions. My parents went to a *really* liberal synagogue where the rabbi didn't believe in God and everyone just agreed they were going through the motions. I got Bar Mitzvahed when I was a kid but haven't been to synagogue in years. My children probably won't even have that much.

So imagine you're an evangelical Christian. All the people you like are also evangelical Christians. Most of your social life happens at church. Most of your good memories involve things like Sunday school and Easter celebrations, and even your bittersweet memories are things like your pastor speaking at your parents' funeral. Most of your hopes and dreams involve marrying someone and having kids and then sharing similarly good times with them. When you try to hang out with people who aren't evangelical Christians, they seem to think really differently than you do, and not at all in a good way. A lot of your happiest intellectual experiences involve geeking out over different Bible verses and the minutiae of different Christian denominations.

Then somebody points out to you that God probably doesn’t exist. And even if He does, it’s probably in some vague and complicated way, and not the way that means that the Thrice-Reformed Meta-Baptist Church and *only* the Thrice-Reformed Meta-Baptist Church has the correct interpretation of the Bible and everyone else is wrong.

On the one hand, their argument might be convincing. On the other, you are pretty sure that if everyone agreed on this, your culture would be destroyed. Sure, your kids could be Christmas-and-Easter-Christians who still enjoy the cultural aspects and derive personal meaning from the Bible. But you’re pretty sure that within a couple of generations your descendents would be exactly as secular as anyone else. Absent the belief that serves as your culture’s wall against the outside world, it would dissolve without a trace into the greater homogeneity of Western liberal society. So, do you keep believing a false thing? Or do you give up on everything you love and enjoy and dissolve into a culture that mostly hates and mocks people like you? There’s no good choice. This is why it sucks that things like religion and politics are both rallying flags for tribes, and actual things that there may be a correct position on.

5. Religious Literalism: One comment complaint I heard during the height of the Atheist-Theist Online Wars was that atheists were a lot like fundamentalists. Both wanted to interpret the religious texts in the most literal possible way.

Being on the atheist side of these wars, I always wanted to know: well, why wouldn’t you? Given that the New Testament clearly says you have to give all your money to the poor, and the Old Testament doesn’t say anything about mixing meat and milk, maybe religious Christians should start giving everything to the poor and religious Jews should stop worrying so much about which dishes to use when?

But I think this is the same mistake as treating the Sunni as an organization dedicated to promoting an Abu Bakr caliphate. The holy book is the rallying flag for a religion, but the religion is not itself about the holy book. The rallying flag created a walled-off space where people could undergo the development process and create an independent culture. That independent culture may diverge significantly from the holy book.

I think that very neurotypical people naturally think in terms of tribes, and the idea that they have to retool their perfectly functional tribe to conform to the exact written text of its holy book or constitution or stated political ideology or something seems silly to them. I think that less neurotypical people – a group including many atheists – think less naturally in terms of tribes and so tend to take claims like “Christianity is about following the Bible” at face value. But Christianity is about being part of the Christian tribe, and although that tribe started around the Bible, maintains its coherence because of the Bible, and is of course naturally influenced by it, if it happens to contradict the Bible in some cases that’s not necessarily surprising or catastrophic.

This is also why I’m not really a fan of debates over whether Islam is really “a religion of peace” or “a religion of violence”, especially if those debates involve mining the Quran for passages that support one’s preferred viewpoint. It’s not just because the Quran is a mess of contradictions with enough interpretive degrees of freedom to prove anything at all. It’s not even because Islam is a host of separate cultures as different from one another as Unitarianism is from the Knights Templar. It’s because the Quran just created the space in which the Islamic culture could evolve, but had only limited impact on that evolution. As well try to predict the warlike or peaceful nature of the

United Kingdom by looking at a topographical map of Great Britain.

6. Cultural Appropriation: Thanks to some people who finally explained this to me in a way that made sense. When an item or artform becomes the rallying flag for a tribe, it can threaten the tribe if other people just want to use it as a normal item or artform.

Suppose that rappers start with pre-existing differences from everyone else. Poor, male, non-white minority, lots of experience living in violent places, maybe a certain philosophical outlook towards their condition. Then they get a rallying flag: rap music. They meet one another, like one another. The culture undergoes further development: the lionization of famous rappers, the development of a vocabulary of shared references. They get all of the benefits of being in a tribe like increased trust, social networking, and a sense of pride and identity.

Now suppose some rich white people get into rap. Maybe they get into rap for innocuous reasons: rap is cool, they like the sound of it. Fine. But they don’t share the pre-existing differences, and they can’t be easily assimilated into the tribe. Maybe they develop different conventions, and start saying that instead of being about the struggles of living in severe poverty, rap should be about Founding Fathers. Maybe they start saying the original rappers are bad, and they should stop talking about violence and bitches because that ruins rap’s reputation. Since rich white people tend to be be good at gaining power and influence, maybe their opinions are overrepresented at the Annual Rap Awards, and all of a sudden you can’t win a rap award unless your rap is about the Founding Fathers and doesn’t mention violence (except Founding-Father-related duels). All of a sudden if you try to start some kind of impromptu street rap-off, you’re no longer going to find a lot of people like you whom you instantly get along with and can form a high-trust community. You’re going to find half people like that, and half rich white people who strike you as annoying and are always complaining that your raps don’t feature any Founding Fathers at all. The rallying flag fails and the tribe is lost as a cohesive entity.

7. Fake Gamer Girls: A more controversial example of the same. Video gaming isn’t just a fun way to pass the time. It also brings together a group of people with some pre-existing common characteristics: male, nerdy, often abrasive, not very successful, interested in speculation, high-systematizing. It gives them a rallying flag and creates a culture which then develops its own norms, shared reference points, internet memes, webcomics, heroes, shared gripes, even some unique literature. Then other people with very different characteristics and no particular knowledge of the culture start enjoying video games just because video games are fun. Since the Gamer Tribe has no designated cultural spaces except video games forums and magazines, they view this as an incursion into their cultural spaces and a threat to their existence as a tribe.

Stereotypically this is expressed as them getting angry when girls start playing video games. One can argue that it’s unfair to infer tribe membership based on superficial characteristics like gender – in the same way it might be unfair for the Native Americans to assume someone with blonde hair and blue eyes probably doesn’t follow the Old Ways – but from the tribe’s perspective it’s a reasonable first guess.

I’ve found gamers to get along pretty well with women who share their culture, and poorly with men who don’t – but admit that the one often starts from an assumption of foreignness and the other from an assumption of membership. More important, I’ve found the *idea* of the rejection of the ‘fake gamer girl’, real or not, raised more as a libel by people who genuinely *do* want to destroy gamer culture, in the sense of cleansing video-game-related spaces of a certain type of person/culture and making them entirely controlled by a different type of person/culture, in much the same way that a rich white person who says any rapper who uses violent lyrics needs to be blacklisted from the rap world has a clear culture-change project going on.

These cultural change projects tend to be framed in terms of which culture has the better values, which I think is a limited perspective. I think America has better values than Pakistan does, but that doesn’t mean I want us invading them, let alone razing their culture to the ground and replacing it with our own.

8. Subcultures And Posers: Obligatory David Chapman link. A poser is somebody who uses the rallying flag but doesn’t have the pre-existing differences that create tribal membership and so never really fits into the tribe.

9. Nationalism, Patriotism, and Racism: Nationalism and patriotism use national identity as the rallying flag for a strong tribe. In many cases, nationalism becomes ethno-nationalism, which builds tribal identity off of a combination of heritage, language, religion, and culture. It has to be admitted that this can make for some *incredibly* strong tribes. The rallying flag is built into ancestry, and so the walls are near impossible to obliterate. The symbolism and jargon and cultural identity can be instilled from birth onward. Probably the best example of this is the Jews, who combine ethnicity, religion, and language into a bundle deal and have resisted assimilation for millennia.

Sometimes this can devolve into racism. I’m not sure exactly what the difference between ethno-nationalism and racism is, or whether there even *is* a difference, except that “race” is a much more complicated concept than ethnicity and it’s probably not a coincidence that it has become most popular in a country like America whose ethnicities are hopelessly confused. The Nazis certainly needed a lot of work to transform concern about the German nation into concern about the Aryan race. But it’s fair to say all of this is somewhat related or at least potentially related.

On the other hand, in countries that have non-ethnic notions of heritage, patriotism has an opportunity to substite for racism. Think about the power of the civil rights message that, whether black or white, we are all Americans.

This is maybe most obvious in sub-national groups. Despite people paying a lot of attention to the supposed racism of Republicans, the rare black Republicans do shockingly well within their party. Both Ben Carson and Herman Cain briefly topped the Republican presidential primary polls during their respective election seasons, and their failures seem to have had much more to do with their own personal qualities than with some sort of generic Republican racism. I see the same with Thomas Sowell, with Hispanic Republicans like Ted Cruz, and Asian Republicans like Bobby Jindal.

Maybe an even stronger example is the human biodiversity movement, which many people understandably accuse of being entirely about racism. Nevertheless, some of its most leading figures are black – JayMan and Chanda Chisala (who is adjacent to the movement but gets lots of respect within it) – and they seem to get equal treatment and respect to their white counterparts. Their membership in a strong and close-knit tribe screens off everything else about them.

I worry that attempts to undermine nationalism/patriotism in order to fight racism risk backfiring. The weaker the “American” tribe becomes, the more people emphasize their other tribes – which can be either overtly racial or else heavily divided along racial lines (eg political parties). It continues to worry me that people who would never display an American flag on their lawn because “nations are just a club for hating foreigners” now have a campaign sign on their lawn, five bumper stickers on their car, and are identifying more and more strongly with political positions – ie clubs for hating their fellow citizens.

Is there such a thing as conservation of tribalism? Get rid of one tribal identity and people just end up seizing on another? I’m not sure. And anyway, nobody can agree on exactly what the American identity or American tribe is anyway, so any conceivable such identity would probably risk alienating a bunch of people. I guess that makes it a moot point. But I still think that deliberately trying to eradicate patriotism is not as good an idea as is generally believed.

V.

I think tribes are interesting and underdiscussed. And in a lot of cases when they are discussed, it’s within preexisting frameworks that tilt the playing field towards recognizing some tribes as fundamentally good, others as fundamentally bad, and ignoring the commonalities between all of them.

But in order to talk about tribes coherently, we need to talk about rallying flags. And that involves admitting that a lot of rallying flags are based on ideologies (which are sometimes wrong), holy books (which are *always* wrong), nationality (which we can’t define), race (which is racist), and works of art (which some people inconveniently want to enjoy just as normal art without any connotations).

My title for this post is also my preferred summary: the ideology is not the movement. Or, more jargonishly – the rallying flag is not the tribe. People are just trying to find a tribe for themselves and keep it intact. This often involves defending an ideology they might not be tempted to defend for any other reason. This doesn’t make them bad, and it *may* not even necessarily mean their tribe deserves to go extinct. I’m reluctant to say for sure whether I think it’s okay to maintain a tribe based on a faulty ideology, but I think it’s at least important to understand that these people are in a crappy situation with no good choices, and they deserve some pity.

Some vital aspects of modern society – freedom of speech, freedom of criticism, access to multiple viewpoints, the existence of entryist tribes with explicit goals of invading and destroying competing tribes as problematic, and the overwhelming pressure to dissolve into the Generic Identity Of Modern Secular Consumerism – make maintaining tribal identities really hard these days. I think some of the most interesting sociological questions revolve around whether there are any ways around the practical and moral difficulties with tribalism, what social phenomena are explicable as the struggle of tribes to maintain themselves in the face of pressure, and whether tribalism continues to be a worthwhile or even a possible project at all.

EDIT: Commenters point out a very similar Melting Asphalt post, Religion Is Not About Beliefs.

A Theory About Religion

Related to Monday's post but spun off for length reasons: my crazy theory about where religion comes from.

The near-universal existence of religion across cultures is surprising. Many people have speculated on what makes tribes around the world so fixated on believing in gods and propitiating them and so on. More recently people like Dawkins and Dennett have added their own contributions about parasitic memes and hyperactive agent-detection.

But I think a lot of these explanations are too focused on a modern idea of religion. I find ancient religion much more enlightening. I'm no historian, but from the little I know ancient religion seems to bleed seamlessly into every other aspect of the ancient way of life. For example, the Roman religion was a combination of mythology, larger-than-life history, patriotism, holidays, customs, superstitions, rules about the government, beliefs about virtue, and attempts to read the future off the livers of pigs. And aside from the pig livers, this seems entirely typical.

American culture ("American civil religion") has a lot of these features too. It has mythology and larger-than-life history: George Washington chopping down the cherry tree, the wise and glorious Founding Fathers, Honest Abe single-handedly freeing the slaves with his trusty hatchet. It has patriotic symbols and art: the flag, the anthem, Uncle Sam. It has holidays: the Fourth of July, Martin Luther King Day, Washington's birthday. It has customs: eat turkey on Thanksgiving, have a barbecue on Memorial Day, watch the Super Bowl. It has superstitions – the number 13, black cats – and ritual taboos – even "obvious" things like don't go outside naked needs to be thought of as taboo considering some cultures do so without thinking. It has rules about the government – both the official laws you'll find in the federal law code, but also deep-seated beliefs about the goodness of democracy or about how all men are created equal, and even customs that affect day-to-day governance like the President giving a State of the Union in January before both houses of Congress. There are beliefs about virtue: everyone should be free, we should try to be independent, we should work hard and pursue the American Dream.

People call the Jewish dietary code unusually strict, but it's important to realize the strictness of modern American kashrut. Absolutely no eating insects – remember, even Jewish kashrut allows locusts! Precious few birds outside of chickens, ducks, turkeys, and geese – remember, even Jewish kashrut allows pigeons! No dogs, cats, rodents, or horses. No reptiles or amphibians, no matter how much the French try to convince us that frog legs are great. No eating clearly obvious animal heads with eyes and stuff (even though dozens of advanced cultures do so happily). No blood products (eg black pudding). Mixing milk and soda in the same glass would be absurd and disgusting. Any tuna made with a process that cannot 100% exclude dolphins is impure. And this isn't even including all of the more modern health-oriented taboos like gluten, MSG, trans-fats, GM foods, et cetera.

If we were to ask the same New Guinea tribe to follow Jewish food taboos one week and American food taboos the next, I'm not sure they'd be able to identify one code as any stricter or weirder than the other. They might have some questions about the meat/milk thing, but maybe they'd also wonder why cheeseburgers are great for dinner but ridiculous for breakfast.

People get worked up over all of the weird purity laws and dress codes in Leviticus, but it’s important to realize how strict our own purity laws are. The ancient Jews would have found it ridiculous that men have to shave and bathe every day if they want to be considered for the best jobs. One must not piss anywhere other than a toilet; this is an abomination (but you would be shocked how many of the supposedly strait-laced Japanese will go in an alley if there’s no restroom nearby). I have been yelled at for going to work without a tie *and* for tying my tie in the wrong pattern; wearing sweatpants to work is right out. And once again, this gets even longer if you you let the more modern/rational rules onto the list – Leviticus has a lot to say about dwellings with fungus in them, but I recently learned to my distress that landlord/tenant law has a lot more.

Once again, if we made our poor New Guinea tribe follow Jewish purity laws one week and American purity laws the next, they would probably end up equally confused and angry both times.

So when we think of America as a perfectly natural secular culture, and Jews as following some kind of superstitious draconian law code, we’re just saying that our laws feel natural and obvious, but their laws feel like an outside imposition. And I think if a time-traveling King Solomon showed up at our doorstep, he would recognize American civil religion as a religion *much* quicker than he would recognize Christianity as one. Christianity would look like a barbaric mystery cult that had gotten too big for its britches; American civil religion would look like home.

Insofar as this isn’t obvious to schoolchildren learning about ancient religion, it’s because the only thing one ever hears about ancient religion is the crazy mythologies. But I think

American culture shows lots of signs of trying to form a crazy mythology, only to be stymied by modernity-specific factors. We can’t have crazy mythologies because we have too many historians around to tell us exactly how things really happened. We can’t have crazy mythologies because we have too many scientists around to tell us where the rain and the lightning *really* come from. We can’t have crazy mythologies because we’re only two hundred-odd years old and these things take time. And most of all, we can’t have crazy mythologies because Christianity is already sitting around occupying that spot.

But if America was a thousand years old and had no science, no religion, and no writing, we would have crazy mythologies up the wazoo. George Washington would take on the stature of an Agamemnon; Benjamin Franklin would take on the status of a Daedalus. Instead of centaurs and satyrs and lamia we would have jackalopes and chupacabras and grey aliens. All those people who say with a nod and a wink that Paul Bunyan dug the Great Lakes as a drinking trough for his giant ox would say the same thing nodless and winkless. Superman would take on the stature of a Zeus, dwelling beside Obi-Wan Kenobi and Bigfoot atop Mt. Whitney, helping the virtuous and punishing the wicked. Some American Hesiod would put succumb to the systematizing impulse, put it all together and explain how George Washington was the son of Superman and ordered Paul Bunyan to dig Chesapeake Bay to entrap the British fleet, and nobody would be able to say they were wrong. I mean, we already have Superman vs. Batman as canon, why not go the extra distance?

So in one sense our best analogy for ancient religion is American civil religion coupled to the sort of national mythology we might have gotten if we’d been a little bit more historically confused. But in another sense ancient religion was actually much stronger than this. America has its own individual culture, but it also partakes of the entire Western liberal industrial secular worldview. An American might not feel culture shock if she moved to Britain; she probably would if she moved to New Guinea.

The ancient world had far less trade and transportation than our own and was far less homogenized. If you want to get into the shoes of an ancient contemplating his religion, imagine you’re an American in a world where even your closest neighboring countries are as different from you as New Guinea hill tribes, or Afghan chiefdoms, or Chinese party cadres. In a world like this, your identity as an American would be very salient – and the essence of being an American is impossible to separate from this whole set of national beliefs about celebrating the Fourth of July, not eating insects, wanting freedom and democracy, and believing that Superman lives atop Mt. Whitney. Outside the community of people who 100% believe all these things, there’s just unreachable foreigners whose language you do not speak and whose customs seem somewhere between inscrutable and barbaric.

That was ancient religion – culture in a world where culture meant something. It was nothing like modern religion – which is why you never hear the Greeks complaining that the Egyptians were evil heretics who denied the light of Zeus and needed to be converted by the sword. But ancients nevertheless felt a connection to their culture and community that combined modern patriotism, religious piety, and belief in science – and they expressed it by continuing to perform their rites and even dying for them.

The question of the origin of religion comes down to how these cultures evolved into the clearly-defined religions of the modern day.

I think a big part of this is ossification and separation from context. The Jewish law perfectly preserves what any right-thinking Israelite in 1000 BC would have considered obvious, natural, and not-even-needing-justification (much as any right-thinking American today considers not eating insects obvious). By the time the Bible was being written this was no longer true – foreign customs and inevitable social change were making the old law seem less and less relevant, and I think modern scholarship thinks the Bible was written by a conservative faction of priests making their case for adherence to the old ways. The act of writing it down in a book, declaring this book the sort of thing that people *might* doubt but shouldn't, and then passing that book to their children – that made it a modern religion, in the sense of something potentially separable from culture that required justification. I think that emphasizing the role of God and the gods provided that justification.

The Hebrew Bible never says other gods don't exist; indeed, it often says the opposite. It constantly praises God as stronger and better than other gods. God proves his superiority over the gods of the Egyptians when the serpent he sends Moses eats the serpents the Egyptian gods send Pharaoh's sorcerers. The Israelites are constantly warned against worshipping other gods, not because those gods don't exist but because God is better and also jealous. This is not the worldview of somebody who has very strong ideas about the nature of reality and how supernatural beings fit into that nature. It's the worldview of people who want to say "Our culture is better than your culture". The Bible uses "worshipping foreign gods" as synonymous with "turning to foreign ways". But God has a covenant with Israel, therefore both are forbidden.

This seems to match religion in the classical world – I'm especially thinking of Augustus' conception here, but he wasn’t drawing it out of a vacuum. Performing the proper rites to the Roman gods was how you showed you were on board with Roman culture was how you showed you were loyal to Rome. The Roman view of religion seems pretty ridiculous to us – constant influx of new gods and mystery cults that were believed kind of indiscriminately, plus occasional deification of leading political figures followed by their undeification once they fell from power. But throughout it all, this idea that following the rites as Romulus prescribed them showed loyalty, but doing otherwise would result in decadence and defeat, stuck around.

More modern religions like Christianity, Islam, and Buddhism are a bit different. Obviously their respective founders play a huge role, but I still think part of what makes them religions rather than just philosophies or spiritual teachings is that they underwent this ossification process. Just as modern Judaism preserves many features of 10th-century-BC Israel that got encoded into holy writ, so modern Christianity preserves many features of 1st century Judeo-Hellenist syncretism. In fact, it preserves a lot of features of 13th-century scholasticism, since that was when they really became serious about formalizing and officializing their theology. At the time scholasticism wasn’t particularly religious; it was just the best understanding the 13th century academic community had about the world around them. Since the Church officialized it, everyone else drifted away and they didn’t.

I think it’s also possible that the first few followers of these religions ended up as a subculture, with as much arbitrary subcultural development as any other tribe. My personal experience with subcultures tells me they can get *very* different customs from the surrounding society *very* fast, with or without any connection to real feature of their rallying flag. Those unusual subcultural values then became the values of the religion that developed later.

Hopefully the connection to Monday’s post is pretty clear. The important thing about a religion is that it has a rallying flag that encourages it to preserve a certain culture, plus walls against the outside world. Crucially, despite everything I’m saying about ossification the culture changes a lot: King Solomon would probably recognize modern rabbinic Judaism, but only barely. But it changes in a way different from the way the outside secular society changes, and in ways bound by the ossified text, so there’s still an element of having this ancient culture preserved in amber and maintained up to the modern day.

SSC Bay Area Meetups 2016

The following Slate Star Codex meetups are planned for the next week or so:

Berkeley, Sunday 4/17 at 2:00 PM at CFAR office, 2030 Addison, 7th floor, Berkeley

San Jose, Monday 4/18 at 7:00 PM at a private house, 3806 Williams Rd, San Jose

Googleplex, Tuesday 4/19 at 12:30 PM at Bigtable cafe in building 1900. Googlers can get more information at the internal slatestarcodex-discuss mailing list.

Stanford, Tuesday 4/19 at 5:30 PM at Tressider Food Court

Everyone is welcome to Berkeley, San Jose, and Stanford. Googleplex is limited to Google employees. Please feel free to come even if you don’t read this much, don’t comment, don’t understand everything, don’t agree with me about things, or don’t feel like you fit in with the normal demographic. During past meetups we’ve always appreciated having more diversity of different types of people attending.

Book Review: Albion's Seed

I.

Albion's Seed by David Fischer is a history professor's nine-hundred-page treatise on patterns of early immigration to the Eastern United States. It's not light reading and not the sort of thing I would normally pick up. I read it anyway on the advice of people who kept telling me it explains everything about America. And it sort of does.

In school, we tend to think of the original American colonists as "Englishmen", a maximally non-diverse group who form the background for all of the diversity and ethnic conflict to come later. Fischer's thesis is the opposite. Different parts of the country were settled by very different groups of Englishmen with different regional backgrounds, religions, social classes, and philosophies. The colonization process essentially extracted a single stratum of English society, isolated it from all the others, and then plunked it down on its own somewhere in the Eastern US.

I used to play *Alpha Centauri*, a computer game about the colonization of its namesake star system. One of the dynamics that made it so interesting was its backstory, where a Puerto Rican survivalist, an African plutocrat, and other colorful characters organized their own colonial expeditions and competed to seize territory and resources. You got to explore not only the settlement of a new world, but the settlement of a new world by societies dominated by extreme founder effects. What kind of weird pathologies and wonderful innovations do you get when a group of overly romantic Scottish environmentalists is allowed to develop on its own trajectory free of all non-overly-romantic-Scottish-environmentalist influences? *Albion’s Seed* argues that this is basically the process that formed several early US states.

Fischer describes four of these migrations: the Puritans to New England in the 1620s, the Cavaliers to Virginia in the 1640s, the Quakers to Pennsylvania in the 1670s, and the Borderers to Appalachia in the 1700s.

II.

A: The Puritans

I hear about these people every Thanksgiving, then never think about them again for the next 364 days. They were a Calvinist sect that dissented against the Church of England and followed their own brand of dour, industrious, fun-hating Christianity. Most of them were from East Anglia, the part of England just northeast of London. They came to America partly because they felt persecuted, but mostly because they thought England was full of sin and they were at risk of absorbing the sin by osmosis if they didn’t get away quick and build something better. They really liked “city on a hill” metaphors.

I knew about the Mayflower, I knew about the black hats and silly shoes, I even knew about the time Squanto threatened to release a bioweapon buried under Plymouth Rock that would bring about the apocalypse. But I didn’t know that the Puritan migration to America was basically a eugenicist’s wet dream.

Much like eg Unitarians today, the Puritans were a religious group that drew disproportionately from the most educated and education-obsessed parts of the English populace. Literacy among immigrants to Massachusetts was twice as high as the English average, and in an age when the vast majority of Europeans were farmers most immigrants to

Massachusetts were skilled craftsmen or scholars. And the Puritan "homeland" of East Anglia was a an unusually intellectual place, with strong influences from Dutch and Continental trade; historian Havelock Ellis finds that it "accounts for a much larger proportion of literary, scientific, and intellectual achievement than any other part of England."

Furthermore, only the best Puritans were allowed to go to Massachusetts; Fischer writes that "it may have been the only English colony that required some of its immigrants to submit letters of recommendation" and that "those who did not fit in were banished to other colonies and sent back to England". Puritan "headhunters" went back to England to recruit "godly men" and "honest men" who "must not be of the poorer sort".

INTERESTING PURITAN FACTS:

1. Sir Harry Vane, who was "briefly governor of Massachusetts at the age of 24", "was so rigorous in his Puritanism that he believed only the thrice-born to be truly saved".

2. The great seal of the Massachusetts Bay Company "featured an Indian with arms beckoning, and five English words flowing from his mouth: 'Come over and help us'"

3. Northern New Jersey was settled by Puritans who named their town after the "New Ark Of The Covenant" – modern Newark.

4. Massachusetts clergy were very powerful; Fischer records the story of a traveller asking a man "Are you the parson who serves here?" only to be corrected "I am, sir, the parson who *rules* here."

5. The Puritans tried to import African slaves, but they all died of the cold.

6. In 1639, Massachusetts declared a "Day Of Humiliation" to condemn "novelties, oppression, atheism, excesse, superfluity, idleness, contempt of authority, and trouble in other parts to be remembered”

7. The average family size in Waltham, Massachusetts in the 1730s was 9.7 children.

8. Everyone was compelled by law to live in families. Town officials would search the town for single people and, if found, order them to join a family; if they refused, they were sent to jail.

9. 98% of adult Puritan men were married, compared to only 73% of adult Englishmen in general. Women were under special pressure to marry, and a Puritan proverb said that “women dying maids lead apes in Hell”.

10. 90% of Puritan names were taken from the Bible. Some Puritans took pride in their learning by giving their children obscure Biblical names they would expect nobody else to have heard of, like Mahershalalhasbaz. Others chose random Biblical terms that might not have technically been intended as names; “the son of Bostonian Samuel Pond was named Mene Mene Tekel Upharsin Pond”. Still others chose Biblical words completely at random and named their children things like Maybe or Notwithstanding.

11. Puritan parents traditionally would send children away to be raised with other families, and raise those families’ children in turn, in the hopes that the lack of familiarity would make the child behave better.

12. In 1692, 25% of women over age 45 in Essex County were accused of witchcraft.

13. Massachusetts passed the first law mandating universal public education, which was called The Old Deluder Satan Law in honor of its preamble, which began “It being one chief project of that old deluder, Satan, to keep men from the knowledge of the scriptures…”

14. Massachusetts cuisine was based around “meat and vegetables submerged in plain water and boiled relentlessly without seasonings of any kind”.

15. Along with the famous scarlet A for adultery, Puritans could be forced to wear a B for blasphemy, C for counterfeiting, D for drunkenness, and so on.

16. Wasting time in Massachusetts was literally a criminal offense, listed in the law code, and several people were in fact prosecuted for it.

17. This wasn’t even the nadir of weird hard-to-enforce Massachusetts laws. Another law just said “If any man shall exceed the bounds of moderation, we shall punish him severely”.

Harriet Beecher Stowe wrote of Massachusetts Puritanism: “The underlying foundation of life in New England was one of profound, unutterable, and therefore unuttered mehalncholy, which regarded human existence itself as a ghastly risk, and, in the case of the vast majority of human beings, an inconceivable misfortune.” And indeed, everything was dour, strict, oppressive, and very religious. A typical Massachusetts week would begin in the church, which doubled as the town meeting hall. There were no decorations except a giant staring eye on the pulpit to remind churchgoers that God was watching them. Townspeople would stand up before their and declare their shame and misdeeds, sometimes being forced to literally crawl before the other worshippers begging for forgiveness. Then the minister would give two two-hour sermons back to back. The entire affair would take up to six hours, and the church was unheated (for some reason they stored all their gunpowder there, so no one was allowed to light a fire), and this was Massachusetts, and it was colder in those days than it is now, so that during winter some people would literally lose fingers to frostbite (Fischer: “It was a point of honor for the minister never to shorten a sermon merely because his audience was frozen”). Everyone would stand there with their guns (they were legally required to bring guns, in case Indians attacked during the sermon) and hear about how they were going to Hell, all while the giant staring eye looked at them.

So life as a Puritan was pretty terrible. On the other hand, their society was *impressively* well-ordered. Teenage pregnancy rates were the lowest in the Western world and in some areas literally zero. Murder rates were half those in other American colonies. There was remarkably low income inequality – "the top 10% of wealthholders held only 20%-30% of taxable property", compared to 75% today and similar numbers in other 17th-century civilizations. The poor (at least the poor native to a given town) were treated with charity and respect – "in Salem, one man was ordered to be set by the heels in the stocks for being uncharitable to a poor man in distress". Government was conducted through town meetings in which everyone had a say. Women had more equality than in most parts of the world, and domestic abuse was punished brutally. The educational system was top-notch – "by most empirical tests of intellectual eminence, New England led all other parts of British America from the 17th to the early 20th century".

In some ways the Puritans seem to have taken the classic dystopian bargain – give up all freedom and individuality and art, and you can have a perfect society without crime or violence or inequality. Fischer ends each of his chapters with a discussion of how the society thought of liberty, and the Puritans unsurprisingly thought of liberty as "ordered liberty" – the freedom of everything to tend to its correct place and stay there. They thought of it as a freedom from disruption – apparently FDR stole some of his "freedom from fear" stuff from early Puritan documents. They were extremely not in favor of the sort of liberty that meant that, for example, there wouldn't be laws against wasting time. *That* was going too far.

B: The Cavaliers

The Massachusetts Puritans fled England in the 1620s partly because the king and nobles were oppressing them. In the 1640s, English Puritans under Oliver Cromwell rebelled, took over the government, and killed the king. The nobles not unreasonably started looking to get the heck out.

Virginia had been kind of a wreck ever since the original Jamestown settlers had mostly died of disease. Governor William Berkeley, a noble himself, decided the colony could reinvent itself as a destination for refugee nobles, and told them it would do everything possible to help them maintain the position of oppressive supremacy to which they were accustomed. The British nobility was sold. The Cavaliers – the nobles who had fought and lost the English Civil War – fled to Virginia. Historians cross-checking Virginian immigrant lists against English records find that of Virginians whose opinions on the War were known, 98% were royalists. They were overwhelmingly Anglican, mostly from agrarian southern England, and all related to each other in the incestuous way of nobility everywhere: “it is difficult to think of any ruling elite that has been more closely interrelated since the Ptolemies”. There were twelve members of Virginia’s royal council; in 1724 “all without exception were related to one another by blood or marriage…as late as 1775, every member of that august body was descended from a councilor who had served in 1660”.

These aristocrats didn’t want to do their own work, so they brought with them tens of thousands of indentured servants; more than 75% of all Virginian immigrants arrived in this position. Some of these people came willingly on a system where their master paid their passage over and they would be free after a certain number of years; others were sent by the courts as punishments; still others were just plain kidnapped. The gender ratio was 4:1 in favor of men, and there were entire English gangs dedicated to kidnapping women and sending them to Virginia, where they fetched a high price. Needless to say, these people came from a very different stratum than their masters *or* the Puritans.

People who came to Virginia mostly died. They died of malaria, typhoid fever, amoebiasis, and dysentery. Unlike in New England, where Europeans were better adapted to the cold climate than Africans, in Virginia it was Europeans who had the higher disease-related mortality rate. The whites who survived tended to become “sluggish and indolent”, according to the universal report of travellers and chroniclers, although I might be sluggish and indolent too if I had been kidnapped to go work on some rich person’s farm and sluggishness/indolence was an option.

The Virginians tried their best to oppress white people. Really, they did. The depths to which they sank in trying to oppress white people almost boggle the imagination. There was a rule that if a female indentured servant became pregnant, a few extra years were added on to their indenture, supposedly because they would be working less hard during their pregnancy and child-rearing so it wasn’t fair to the master. Virginian aristocrats would *rape their own female servants*, then add a penalty term on to their indenture for becoming pregnant. That is an *impressive* level of chutzpah. But despite these efforts, eventually all the white people either died, or became too sluggish to be useful, or worst of all just finished up their indentures and became legally free. The aristocrats started importing black slaves as per the model that had sprung up in the Caribbean, and so the stage was set for the antebellum South we read about in history classes.

INTERESTING CAVALIER FACTS:

1. Virginian cavalier speech patterns sound a lot like modern African-American dialects. It doesn't take much imagination to figure out why, but it's strange to think of a 17th century British lord speaking what a modern ear would clearly recognize as Ebonics.

2. Three-quarters of 17th-century Virginian children lost at least one parent before turning 18.

3. Cousin marriage was an important custom that helped cement bonds among the Virginian elite, "and many an Anglican lady changed her condition but not her name".

4. In Virginia, women were sometimes unironically called "breeders"; English women were sometimes referred to as "She-Britons".

5. Virginia didn't really have towns; the Chesapeake Bay was such a giant maze of rivers and estuaries and waterways that there wasn't much need for land transport hubs. Instead, the unit of settlement was the plantation, which consisted of an aristocratic planter, his wife and family, his servants, his slaves, and a bunch of guests who hung around and mooched off him in accordance with the ancient custom of hospitality.

6. Virginian society considered everyone who lived in a plantation home to be a kind of "family", with the aristocrat both as the literal father and as a sort of abstracted patriarch with complete control over his domain.

7. Virginia governor William Berkeley probably would not be described by moderns as 'strong on education'. He said in a speech that "I thank God there are no free schools nor printing [in Virginia], and I hope we shall not have these for a hundred years, for learning has brought disobedience, and heresy, and sects into the world, and printing has divuldged them, and libels against the best government. God keep us from both!"

8. Virginian recreation mostly revolved around hunting and bloodsports. Great lords hunted deer, lesser gentry hunted foxes, indentured servants had a weird game in which they essentially draw-and-quartered geese, young children "killed and tortured songbirds", and "at the bottom of this hierarchy of bloody games were male infants who prepared themselves for the larger pleasures of maturity by torturing snakes, maiming frogs, and pulling the wings off butterflies. Thus, every red-blooded male in Virginia was permitted to slaughter some animal or other, and the size of his victim was proportioned to his social rank."

9. "In 1747, an Anglican minister named William Kay infuriated the great planter Landon Carter by preaching a sermon against pride. The planter took it personally and sent his [relations] and ordered them to nail up the doors and windows of all the churches in which Kay preached."

10. Our word "condescension" comes from a ritual attitude that leading Virginians were supposed to display to their inferiors. Originally condescension was supposed to be a polite way of showing respect those who were socially inferior to you; our modern use of the term probably says a lot about what Virginians actually did with it.

In a lot of ways, Virginia was the opposite of Massachusetts. Their homicide rate was sky-high, and people were actively encouraged to respond to slights against their honor with duels (for the rich) and violence (for the poor). They were obsessed with gambling, and "made bets not merely on horses, cards, cockfights, and backgammon, but also on crops, prices, women, and the weather". Their cuisine focused on gigantic sumptuous feasts of animals killed in horrible ways. There were no witchcraft trials, but there *were* people who were fined for disrupting the peace by accusing their neighbors of witchcraft. Their church sermons were twenty minutes long on the dot.

The Puritans naturally thought of the Virginians as completely lawless reprobate sinners, but this is not *entirely* true. Virginian church sermons might have been twenty minutes long, but Virginian ballroom dance lessons could last nine hours. It wasn't that the Virginians weren't bound by codes, just that those codes were social rather than moral.

And Virginian nobles weren't just random jerks, they were *carefully cultivated* jerks. Planters spared no expense to train their sons to be strong, forceful, and not take nothin' from nobody. They would encourage and reward children for being loud and temperamental, on the grounds that this indicated a strong personality and having a strong personality was fitting of a noble. When this worked, it worked *really* well – witness natural leaders and self-driven polymaths like George Washington and Thomas Jefferson. More often it failed catastrophically – the rate of sex predation and rape in Virginia was at least as high as anywhere else in North America.

The Virginian Cavaliers had an obsession with liberty, but needless to say it was not exactly a sort of liberty of which the ACLU would approve. I once heard someone argue against libertarians like so: even if the government did not infringe on liberties, we would still be unfree for other reasons. If we had to work, we would be subject to the whim of bosses. If we were poor, we would not be "free" to purchase most of the things we want. In any case, we are "oppressed" by disease, famine, and many other things besides government that prevent us from implementing our ideal existence.

The Virginians took this idea and ran with it – in the wrong direction. No, they said, we *wouldn't* be free if we had to work, therefore we insist upon not working. No, we *wouldn't* be free if we were limited by poverty, therefore we insist upon being extremely rich. Needless to say, this conception of freedom required first indentured servitude and later slavery to make it work, but the Virginians never *claimed* that the servants or slaves were free. That wasn't the point. Freedom, like wealth, was properly distributed according to rank; nobles had as much as they wanted, the middle-class enough to get by on, and everyone else none at all. And a Virginian noble would have gone to his grave insisting that a civilization without slavery could never have citizens who were truly free.

C: The Quakers

Fischer warns against the temptation to think of the Quakers as normal modern people, but he has to warn us precisely because it's so tempting. Where the Puritans seem like a dystopian caricature of virtue and the Cavaliers like a dystopian caricature of vice, the Quakers just seem *ordinary*. Yes, they're kind of a religious cult, but they're the kind of religious cult any of us might found if we were thrown back to the seventeenth century.

Instead they were founded by a weaver's son named George Fox. He believed people were basically good and had an Inner Light that connected them directly to God without a need for priesthood, ritual, Bible study, or self-denial; mostly people just needed to listen to their consciences and be nice. Since everyone was equal before God, there was no point in holding up distinctions between lords and commoners: Quakers would just address everybody as "Friend". And since the Quakers were among the most persecuted sects at the time, they developed an insistence on tolerance and freedom of religion which (unlike the Puritans) they stuck to even when shifting fortunes put them on top. They believed in pacificism, equality of the sexes, racial harmony, and a bunch of other things which seem pretty hippy-ish even today let alone in 1650.

England’s top Quaker in the late 1600s was William Penn. Penn is universally known to Americans as “that guy Pennsylvania is named after” but actually was a larger-than-life 17th century superman. Born to the nobility, Penn distinguished himself early on as a military officer; he was known for beating legendary duelists in single combat and then sparing their lives with sermons about how murder was wrong. He gradually started having mystical visions, quit the military, and converted to Quakerism. Like many Quakers he was arrested for blasphemy; unlike many Quakers, they couldn’t make the conviction stick; in his trial he “conducted his defense so brilliantly that the jurors refused to convict him even when threatened with prison themselves, [and] the case became a landmark in the history of trial by jury.” When the state finally found a pretext on which to throw him in prison, he spent his incarceration composing “one of the noblest defenses of religious liberty ever written”, conducting a successful mail-based courtship with England’s most eligible noblewoman, and somehow gaining the personal friendship and admiration of King Charles II. Upon his release the King liked him so much that he gave him a large chunk of the Eastern United States on a flimsy pretext of repaying a family debt. Penn didn’t *want* to name his new territory Pennsylvania – he recommended just “Sylvania” – but everybody else overruled him and Pennyslvania it was. The grant wasn’t quite the same as the modern state, but a chunk of land around the Delaware River Valley – what today we would call eastern Pennsylvania, northern Delaware, southern New Jersey, and bits of Maryland – centered on the obviously-named-by-Quakers city of Philadelphia.

Penn decided his new territory would be a Quaker refuge – his exact wording was “a colony of Heaven [for] the children of the Light”. He mandated universal religious toleration, a total ban on military activity, and a government based on checks and balances that would “leave myself and successors no power of doing mischief, that the will of one man may not hinder the good of a whole country”.

His recruits – about 20,000 people in total – were Quakers from the north of England, many of them minor merchants and traders. They disproportionately included the Britons of Norse descent common in that region, who formed a separate stratum and had never really gotten along with the rest of the British population. They were joined by several German sects close enough to Quakers that they felt at home there; these became the ancestors of (among other groups) the Pennsylvania Dutch, Amish, and Mennonites.

INTERESTING QUAKER FACTS:

1. In 1690 a gang of pirates stole a ship in Philadelphia and went up and down the Delaware River stealing and plundering. The Quakers got in a heated (but brotherly) debate about whether it was morally permissible to use violence to stop them. When the government finally decided to take action, contrarian minister George Keith dissented and caused a major schism in the faith.

2. Fischer argues that the Quaker ban on military activity within their territory would have doomed them in most other American regions, but by extreme good luck the Indians in the Delaware Valley were almost as peaceful as the Quakers. As usual, at least some credit goes to William Penn, who taught himself Algonquin so he could negotiate with the Indians in their own language.

3. The Quakers’ marriage customs combined a surprisingly modern ideas of romance, with extreme bureaucracy. The wedding process itself had sixteen stages, including “ask parents”, “ask community women”, “ask community men”, “community women ask parents”, and “obtain a certificate of cleanliness”. William Penn’s marriage apparently had forty-six witnesses to testify to the good conduct and non-relatedness of both parties.

4. Possibly related: 16% of Quaker women were unmarried by age 50, compared to only about 2% of Puritans.

5. Quakers promoted gender equality, including the (at the time scandalous) custom of allowing women to preach (condemned by the Puritans as the crime of “she-preaching”).

6. But they were such prudes about sex that even *the Puritans* thought they went too far. Pennsylvania doctors had problems treating Quakers because they would “delicately describe everything from neck to waist as their ‘stomachs’, and anything from waist to feet as their ‘ankles'”.

7. Quaker parents Richard and Abigail Lippincott named their eight children, in order, “Remember”, “John”, “Restore”, “Freedom”, “Increase”, “Jacob”, “Preserve”, and “Israel”, so that their names combined formed a simple prayer.

8. Quakers had surprisingly modern ideas about parenting, basically sheltering and spoiling their children at a time when everyone else was trying whip the Devil out of them.

9. “A Quaker preacher, traveling in the more complaisant colony of Maryland, came upon a party of young people who were dancing merrily together. He broke in upon them like an avenging angel, stopped the dance, anddemanded to know if they considered Martin Luther to be a good man. The astonished youngsters answered in the affirmative. The Quaker evangelist then quoted Luther on the subject of dancing: ‘as many paces as the man takes in his dance, so many steps he takes toward Hell. This, the Quaker missionary gloated with a gleam of sadistic satisfaction, ‘spoiled their sport’.”

10. William Penn wrote about thirty books defending liberty of conscience throughout his life. The Quaker obsession with the individual conscience as the work of God helped invent the modern idea of conscientious objection.

11. Quakers were heavily (and uniquely for their period) opposed to animal cruelty. When foreigners introduced bullbaiting into Philadelphia during the 1700s, the mayor bought a ticket supposedly as a spectator. When the event was about to begin, he leapt into the ring, personally set the bull free, and threatened to arrest anybody who stopped him.

12. On the other hand, they were also opposed to other sports for what seem like kind of random reasons. The town of Morley declared an anathema against foot races, saying that they were "unfruitful works of darkness".

13. The Pennsylvania Quakers became very prosperous merchants and traders. They also had a policy of loaning money at low- or zero- interest to other Quakers, which let them outcompete other, less religious businesspeople.

14. They were among the first to replace the set of bows, grovels, nods, meaningful looks, and other British customs of acknowledging rank upon greeting with a single rank-neutral equivalent – the handshake.

15. Pennsylvania was one of the first polities in the western world to abolish the death penalty.

16. The Quakers were lukewarm on education, believing that too much schooling obscured the natural Inner Light. Fischer declares it "typical of William Penn" that he wrote a book arguing against reading too much.

17. The Quakers not only instituted religious freedom, but made laws against mocking another person's religion.

18. In the late 1600s as many as 70% of upper-class Quakers owned slaves, but Pennsylvania essentially invented modern abolitionism. Although their colonial masters in England forbade them from banning slavery outright, they applied immense social pressure and by the mid 1700s less than 10% of the wealthy had African slaves. As soon as the American Revolution started, forbidding slavery was one of independent Pennsylvania’s first actions.

Pennsylvania was *very* successful for a while; it had some of the richest farmland in the colonies, and the Quakers were exceptional merchants and traders; so much so that they were forgiven their military non-intervention during the Revolution because of their role keeping the American economy afloat in the face of British sanctions.

But by 1750, the Quakers were kind of on their way out; by 1750, they were a demographic minority in Pennsylvania, and by 1773 they were a minority in its legislature as well. In 1750 Quakerism was the third-largest religion in the US; by 1820 it was the ninth-largest, and by 1981 it was the sixty-sixth largest. What happened? The Quakers basically tolerated themselves out of existence. They were so welcoming to religious minorities and immigrants that all these groups took up shop in Pennsylvania and ended its status as a uniquely Quaker society. At the same time, the Quakers themselves became more “fanatical” and many dropped out of politics believing it to be too worldly a concern for them; this was obviously fatal to their political domination. The most famous Pennsylvanian statesman of the Revolutionary era, Benjamin Franklin, was not a Quaker at all but a first-generation immigrant from New England. Finally, Quakerism was naturally extra-susceptible to that thing where Christian denominations become indistinguishable from liberal modernity and fade into the secular background.

But Fischer argues that Quakerism continued to shape Pennsylvania long after it had stopped being officially in charge, in much the same way that Englishmen themselves have contributed disproportionately to American institutions even though they are now a numerical minority. The Pennsylvanian leadership on abolitionism, penal reform, the death penalty, and so on all happened *after* the colony was officially no longer Quaker-dominated.

And it’s hard not to see Quaker influence on the ideas of the modern US – which was after all founded in Philadelphia. In the middle of the Puritans demanding strict obedience to their dystopian hive society and the Cavaliers demanding everybody bow down to a transplanted nobility, the Pennsylvanians – who became the thought leaders of the Mid-Atlantic region including to a limited degree New York City – were pretty normal and had a good opportunity to serve as power-brokers and middlemen between the North and South. Although there are seeds of traditionally American ideas in every region, the Quakers really stand out in terms of freedom of religion, freedom of thought, checks and balances, and the idea of universal equality.

It occurs to me that William Penn might be literally the single most successful person in history. He started out as a minor noble following a religious sect that everybody despised and managed to export its principles to Pennsylvania where they flourished and multiplied. Pennsylvania then managed to export *its* principles to the United States, and the United States exported them to the world. I’m not sure how much of the suspiciously Quaker character of modern society is a direct result of William Penn, but he was in one heck of a right place at one heck of a right time

D: The Borderers

The Borderers are usually called “the Scots-Irish”, but Fischer dislikes the term because they are neither Scots (as we usually think of Scots) nor Irish (as we usually think of

Irish). Instead, they’re a bunch of people who lived on (both sides of) the Scottish-English border in the late 1600s.

None of this makes sense without realizing that the Scottish-English border was *terrible*. Every couple of years the King of England would invade Scotland or vice versa; “from the year 1040 to 1745, every English monarch but three suffered a Scottish invasion, or became an invader in his turn”. These “invasions” generally involved burning down all the border towns and killing a bunch of people there. Eventually the two sides started getting *pissed* with each other and would also torture-murder all of the enemy’s citizens they could get their hands on, ie any who were close enough to the border to reach before the enemy could send in their armies. As if this weren’t bad enough, outlaws quickly learned they could plunder one side of the border, then escape to the other before anyone brought them to justice, so the whole area basically became one giant cesspool of robbery and murder.

In response to these pressures, the border people militarized and stayed feudal long past the point where the rest of the island had started modernizing. Life consisted of farming the lands of whichever brutal warlord had the top hand today, followed by being called to fight for him on short notice, followed by a grisly death. The border people dealt with it as best they could, and developed a culture marked by extreme levels of clannishness, xenophobia, drunkenness, stubbornness, and violence.

By the end of the 1600s, the Scottish and English royal bloodlines had intermingled and the two countries were drifting closer and closer to Union. The English kings finally got some breathing room and noticed – holy frick, everything about the border is *terrible*. They decided to make the region economically productive, which meant “squeeze every cent out of the poor Borderers, in the hopes of either getting lots of money from them or else forcing them to go elsewhere and become somebody else’s problem”. Sometimes absentee landlords would just evict everyone who lived in an entire region, *en masse*, replacing them with people they expected to be easier to control.

Many of the Borderers fled to Ulster in Ireland, which England was working on colonizing as a Protestant bulwark against the Irish Catholics, and where the Crown welcomed violent warlike people as a useful addition to their Irish-Catholic-fighting project. But Ulster had some of the same problems as the Border, and also the Ulsterites started worrying that the Borderer cure was worse than the Irish Catholic disease. So the Borderers started getting kicked out of Ulster too, one thing led to another, and eventually 250,000 of these people ended up in America.

250,000 people is a *lot* of Borderers. By contrast, the great Puritan emigration wave was only 20,000 or so people; even the mighty colony of Virginia only had about 50,000 original settlers. So these people showed up on the door of the American colonies, and the American colonies collectively took one look at them and said “nope”.

Except, of course, the Quakers. The Quakers talked among themselves and decided that these people were also Children Of God, and so they should demonstrate Brotherly Love by taking them in. They tried that for a couple of years, and then they questioned their life choices and *also* said “nope”, and they told the Borderers that Philadelphia and the Delaware Valley were actually kind of full right now but there was lots of unoccupied land in *Western* Pennsylvania, and the Appalachian Mountains were very pretty at this time of year, so why didn’t they head out that way as fast as it was physically possible to go?

At the time, the Appalachians were kind of the booby prize of American colonization: hard to farm, hard to travel through, and exposed to hostile Indians. The Borderers fell in love with them. They came from a pretty marginal and unproductive territory themselves, and the Appalachians were far away from everybody and full of fun Indians to fight. Soon the Appalachian strategy became the accepted response to Borderer immigration and was taken up from Pennsylvania in the north to the Carolinas in the South (a few New Englanders hit on a similar idea and sent their own Borderers to colonize the mountains of New Hampshire).

So the Borderers all went to Appalachia and established their own little rural clans there and nothing at all went wrong except for the entire rest of American history.

INTERESTING BORDERER FACTS:

1. Colonial opinion on the Borderers differed within a very narrow range: one Pennsylvanian writer called them “the scum of two nations”, another Anglican clergyman called them “the scum of the universe”.

2. Some Borderers tried to come to America as indentured servants, but after Virginian planters got some experience with Borderers they refused to accept any more.

3. The Borderers were mostly Presbyterians, and their arrival *en masse* started a race among the established American denominations to convert them. This was mostly unsuccessful; Anglican preacher Charles Woodmason, an important source for information about the early Borderers, said that during his missionary activity the Borderers “disrupted his service, rioted while he preached, started a pack of dogs fighting outside the church, loosed his horse, stole his church key, refused him food and shelter, and gave two barrels of whiskey to his congregation before a service of communion”.

4. Borderer town-naming policy was very different from the

Biblical names of the Puritans or the Ye Olde English names of the Virginians. Early Borderer settlements include – just to stick to the creek-related ones – Lousy Creek, Naked Creek, Shitbritches Creek, Cuckold's Creek, Bloodrun Creek, Pinchgut Creek, Whipping Creek, and Hangover Creek. There were also Whiskey Springs, Hell's Half Acre, Scream Ridge, Scuffletown, and Grabtown. The overall aesthetic honestly sounds a bit Orcish.

5. One of the first Borderer leaders was John Houston. On the ship over to America, the crew tried to steal some of his possessions; Houston retaliated by leading a mutiny of the passengers, stealing the ship, and sailing it to America himself. He settled in West Virginia; one of his descendants was famous Texan Sam Houston.

6. Traditional Borderer prayer: "Lord, grant that I may always be right, for thou knowest I am hard to turn."

7. "The backcountry folk bragged that one interior county of North Carolina had so little 'larnin' that the only literate inhabitant was elected 'county reader'"

8. The Borderer accent contained English, Scottish, and Irish elements, and is (uncoincidentally) very similar to the typical "country western singer" accent of today.

9. The Borderers were famous for family feuds in England, including the Johnson clan's habit of "adorning their houses with the flayed skins of their enemies the Maxwells in a blood feud that continued for many generations". The great family feuds of the United States, like the Hatfield-McCoy feud, are a direct descendent of this tradition.

10. Within-clan marriage was a popular Borderer tradition both in England and Appalachia; "in the Cumbrian parish of Hawkshead, for example, both the bride and the groom bore the same last names in 25 percent of all marriages from 1568 to 1704". This led to the modern stereotype of Appalachians as inbred and incestuous.

11. The Borderers were extremely patriarchal and anti-women's-rights to a degree that appalled even the people of the 1700s.

12. “In the year 1767, [Anglican priest] Charles Woodmason calculated that 94 percent of backcountry brides whom he had married in the past year were pregnant on their wedding day”

13. Although the Borderers started off Presbyterian, they were in constant religious churn and their territories were full of revivals, camp meetings, born-again evangelicalism, and itinerant preachers. Eventually most of them ended up as what we now call Southern Baptist.

14. Borderer folk beliefs: “If an old woman has only one tooth, she is a witch”, “If you are awake at eleven, you will see witches”, “The howling of dogs shows the presence of witches”, “If your shoestring comes untied, witches are after you”, “If a warm current of air is felt, witches are passing”. Also, “wet a rag in your enemy’s blood, put it behind a rock in the chimney, and when it rots your enemy will die”; apparently it was not a coincidence they were thinking about witches so much.

15. Borderer medical beliefs: “A cure for homesickness is to sew a good charge of gunpowder on the inside of ths shirt near the neck”. That’ll cure homesickness, all right.

16. More Borderer medical beliefs: “For fever, cut a black chicken open while alive and bind it to the bottom of your foot”, “Eating the brain of a screech owl is the only dependable remedy for headache”, “For rheumatism, apply split frogs to the feet”, “To reduce a swollen leg, split a live cat and apply while still warm”, “Bite the head off the first butterfly you see and you will get a new dress”, “Open the cow’s mouth and throw a live toad-frog down her throat. This will cure her of hollow-horn”. Also, blacksmiths protected themselves from witches by occasionally throwing live puppies into their furnaces.

17. Rates of public schooling in the backcountry settled by the Borderers were “the lowest in British North America” and sometimes involved rituals like “barring out”, where the children would physically keep the teacher out of the school until he gave in and granted the students the day off.

18. “Appalachia’s idea of a moderate drinker was the mountain man who limited himself to a single quart [of whiskey] at a sitting, explaining that more ‘might fly to my head’. Other beverages were regarded with contempt.”

19. A traditional backcountry sport was “rough and tumble”, a no-holds-barred form of wrestling where gouging out your opponent’s eyes was considered perfectly acceptable and in fact sound strategy. In 1772 Virginia had to pass a law against “gouging, plucking, or putting out an eye”, but this was the Cavalier-dominated legislature all the way on the east coast and nobody in the backcountry paid them any attention. Other traditional backcountry sports were sharpshooting and hunting.

20. The American custom of shooting guns into the air to celebrate holidays is 100% Borderer in origin.

21. The justice system of the backcountry was heavy on lynching, originally a race-neutral practice and named after western Virginian settler William Lynch.

22. Scottish Presbyterians used to wear red cloth around their neck to symbolize their religion; other Englishmen nicknamed them “rednecks”. This *may* be the origin of the popular slur against Americans of Borderer descent, although many other etiologies have been proposed. “Cracker” as a slur is attested as early as 1766 by a colonist who says the term describes backcountry men who are great boasters; other proposed etymologies like slaves talking about “whip-crackers” seem to be spurious.

This is not to paint the Borderers as universally poor and dumb – like every group, they had an elite, and some of their elite went on to become some of America’s most important historical figures. Andrew Jackson became the first Borderer president, behaving *exactly* as you would expect the first Borderer president to behave, and he was followed by almost a dozen others. Borderers have also been overrepresented in America's great military leaders, from Ulysses Grant through Teddy Roosevelt (3/4 Borderer despite his Dutch surname) to George Patton to John McCain.

The Borderers *really* liked America – unsurprising given where they came from – and started identifying as American earlier and more fiercely than any of the other settlers who had come before. Unsurprisingly, they strongly supported the Revolution – Patrick Henry ("Give me liberty or give me death!") was a Borderer. They also also played a disproportionate role in westward expansion. After the Revolution, America made an almost literal 180 degree turn and the "backcountry" became the "frontier". It was the Borderers who were happiest going off into the wilderness and fighting Indians, and most of the famous frontiersmen like Davy Crockett were of their number. This was a big part of the reason the Wild West was so wild compared to, say, Minnesota (also a frontier inhabited by lots of Indians, but settled by Northerners and Germans) and why it inherited seemingly Gaelic traditions like cattle rustling.

Their conception of liberty has also survived and shaped modern American politics: it seems essentially to be the modern libertarian/Republican version of freedom from government interference, especially if phrased as "get the hell off my land", and *especially* especially if phrased that way through clenched teeth while pointing a shotgun at the offending party.

III.

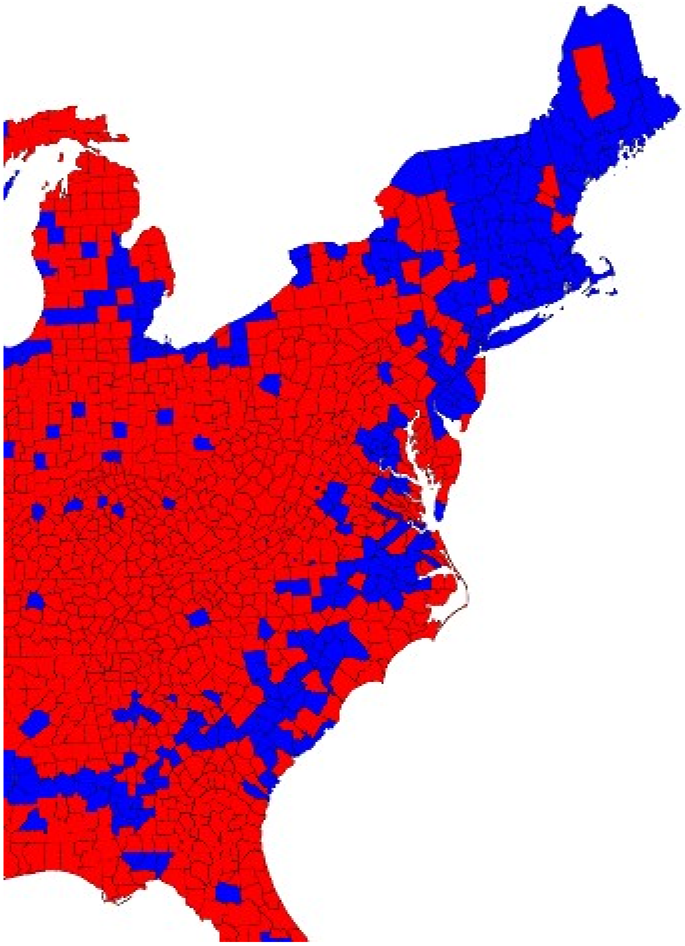
This is all interesting as history and doubly interesting as anthropology, but what relevance does it have for later American history and the present day?

One of my reasons reading this book was to see whether the link between Americans' political opinions and a bunch of their other cultural/religious/social traits (a "Blue Tribe" and "Red Tribe") was related to the immigration patterns it describes. I'm leaning towards "probably", but there's a lot of work to be done in explaining how the split among these four cultures led to a split among two cultures in the modern day, and with little help from the book itself I am going to have to resort to total unfounded speculation. But the simplest explanation – that the Puritans and Quakers merged into one group ("progressives", "Blue Tribe", "educated coastal elites") and the Virginians and Borderers into another ("conservatives", "Red Tribe", "rednecks") – has a lot going for it.

Many conservatives I read like to push the theory that modern progressivism is descended from the utopian Protestant experiments of early America – Puritanism and Quakerism – and that the civil war represents "Massachusetts' conquest of America". I always found this lacking in rigor: Puritanism and Quakerism are sufficiently different that positing a combination of them probably needs more intellectual work than just gesturing at "you know, that Puritan/Quaker thing". But the idea of a Puritan New England and a Quaker-(ish) Pennsylvania gradually blending together into a generic "North" seems plausible, especially given the high levels of interbreeding between the two (some of our more progressive Presidents, including Abraham Lincoln, were literally half-Puritan and half-Quaker). Such a merge would combine the Puritan emphasis on moral reform, education, and a well-ordered society with the Quaker doctrine of niceness, tolerance, religious pluralism, individual conscience, and the Inner Light. It seems kind of unfair to just mix-and-match the most modern elements of each and declare that this proves they caused modernity, but there's no reason that *couldn't* have happened.

The idea of Cavaliers and Borderers combining to form modern conservativism is buoyed by modern conservativism's obvious Border influences, but complicated by its lack of much that is recognizably Cavalier – the Republican Party is hardly marked by its support for a hereditary aristocracy of gentlemen. Here I have to admit that I don't know as much about Southern history as I'd like. In particular, how were places like Alabama, Mississippi, et cetera settled? Most sources I can find suggest they were set up along the Virginia model of plantation-owning aristocrats, but if that's true how did the modern populations come to so embody Fischer's description of Borderers? In particular, why are they so Southern Baptist and not very Anglican? And what happened to all of those indentured servants the Cavaliers brought over after slavery put them out of business? What happened to that whole culture after the Civil War destroyed the plantation system? My *guess* is going to be that the indentured servants and the Borderer population mixed pretty thoroughly, and that this stratum was hanging around providing a majority of the white bodies in the South while the plantation owners were hogging the limelight – but I just don't know.

A quick argument that I'm not totally making all of this up:



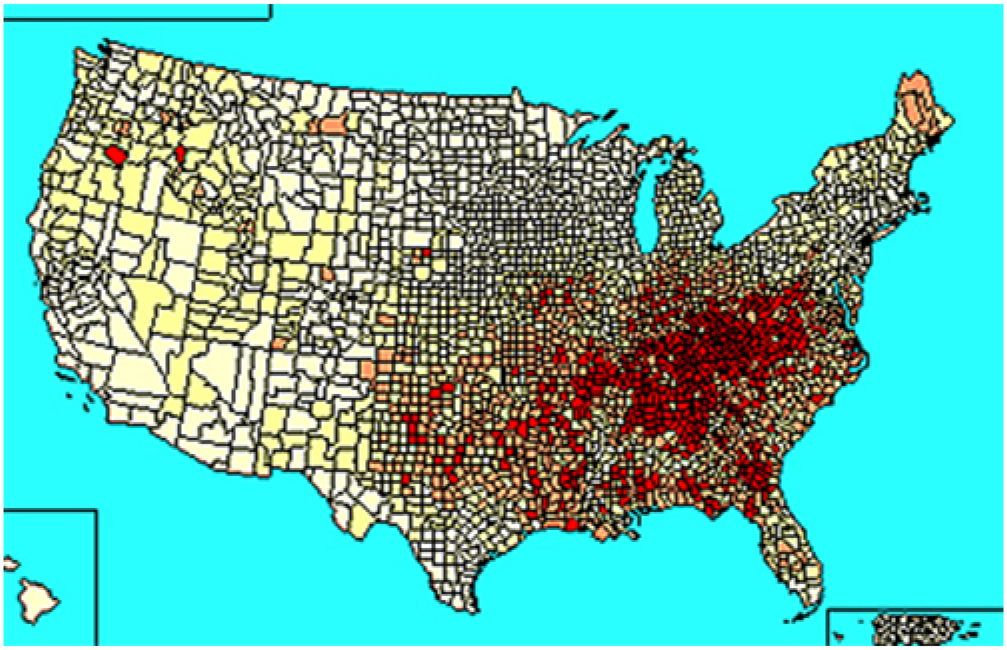
This is a map of voting patterns by county in the 2012 Presidential election. The blue areas in the South carefully track the so-called “black belt” of majority African-American areas. The ones in the Midwest are mostly big cities. Aside from those, the only people who vote Democrat are New England (very solidly!) and the Delaware Valley region of Pennsylvania. In fact, you can easily see the distinction between the Delaware Valley settled by Quakers in the east, and the backcountry area settled by Borderers in the west. Even the book’s footnote about how a few Borderers settled in the mountains of New Hampshire is associated with a few spots of red in the mountains of New Hampshire ruining an otherwise near-perfect Democratic sweep of the north.

One anomaly in this story is a kind of linear distribution of blue across southern Michigan, too big to be explained solely by the blacks of Detroit. But a quick look at Wikipedia’s History of Michigan finds:

> In the 1820s and 1830s migrants from New England began moving to what is now Michigan in large numbers (though there was a trickle of New England settlers who arrived before this date). These were “Yankee” settlers, that is to say they were descended from the English Puritans who settled New England during the colonial era….Due to the prevalence of New Englanders and New England transplants from upstate New York, Michigan was very culturally contiguous with early New England culture for much of its early history…The amount with which the New England Yankee population predominated made Michigan unique among frontier states in the antebellum period. Due to this heritage Michigan was on the forefront of the antislavery crusade and reforms during the 1840s and 1850s.

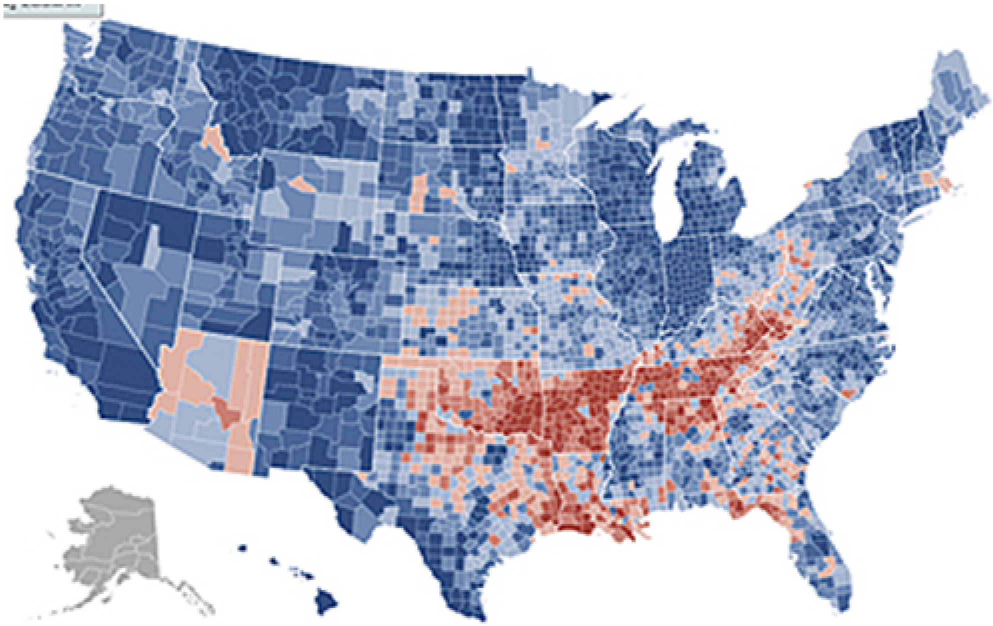
Alhough I can’t find proof of this specifically, I know that Michigan was settled from the south up, and I suspect that these New England settlers concentrated in the southern regions and that the north was settled by a more diverse group of whites who lacked the New England connection.

Here’s something else cool. We can’t track Borderers directly because there’s no “Borderer” or “Scots-Irish” option on the US census. But *Albion’s Seed* points out that the Borderers were uniquely likely to identify as just “American” and deliberately forgot their past ancestry as fast as they could. Meanwhile, when the census asks an ethnicity question about where your ancestors came from, every year some people will stubbornly ignore the point of the question and put down “America” (no, this does not track the distribution of Native American population). Here’s a map of so-called “unhyphenated Americans”, taken from this site:



We see a strong focus on the Appalachian Mountains, especially West Virginia, Tennesee, and Kentucky, bleeding into the rest of the South. Aside from west Pennsylvania, this is very close to where we would expect to find the Borderers. Could these be the same groups?

Meanwhile, here is a map of where Obama underperformed the usual Democratic vote *worst* in 2008:



These maps are small and lossy, and surely unhyphenatedness is not an *exact* proxy for Border ancestry – but they are nevertheless intriguing. You may also be interested in the Washington Post’s correlation between distribution of unhyphenated Americans and Trump voters, or the Atlantic’s article on Trump and Borderers.

If I’m going to map these cultural affiliations to ancestry, do I have to walk back on my previous theory that they are related to class? Maybe I should. But I also think we can posit complicated interactions between these ideas. Consider for example the interaction between race and class; a black person with a white-sounding name, who speaks with a white-sounding accent, and who adopts white culture (eg listens to classical music, wears business suits) is far more likely to seem upper-class than a black person with a black-sounding name, a black accent, and black cultural preferences; a white person who seems black in some way (listens to hip-hop, wears baggy clothes) is more likely to seem lower-class. This doesn’t mean race and class are exactly the same thing, but it does mean that some races get stereotyped as upper-class and others as lower-class, and that people’s racial identifiers may change based on where they are in the class structure.

I think something similar is probably going on with these forms of ancestry. The education system is probably dominated by descendents of New Englanders and Pennsylvanians; they had an opportunity to influence the culture of academia and the educated classes more generally, they took it, and now anybody of any background who makes it into that world is going to be socialized according to their rules. Likewise, people in poorer and more rural environments will be surrounded by people of Borderer ancestry and acculturated by Borderer cultural products and end up a little more like that group. As a result, ethnic markers have turned into and merged with class markers in complicated ways.

Indeed, some kind of acculturation process has to have been going on, since most of the people in these areas today are not the descendents of the original settlers. But such a process seems very likely. Just to take an example, most of the Jews I know (including my own family) came into the country via New York, live somewhere on the coast, and have very Blue Tribe values. But Southern Jews believed in the Confederacy as strongly as any Virginian – see for example Judah Benjamin. And Barry Goldwater, a half-Jew raised in Arizona, invented the modern version of conservativism that seems closest to some Borderer beliefs.

All of this is very speculative, with some obvious flaws. What do we make of other countries like Britain or Germany with superficially similar splits but very different histories? Why should Puritans lose their religion and sexual prudery, but keep their interest in moralistic reform? There are whole heaps of questions like these. But look. Before I had any idea about any of this, I wrote that American society seems divided into two strata, one of which is marked by emphasis on education, interest in moral reforms, racial tolerance, low teenage pregnancy, academic/financial jobs, and Democratic party affiliation, and furthermore that this group was centered in the North. Meanwhile, now I learn that the North was settled by two groups that when combined have emphasis on education, interest in moral reforms, racial tolerance, low teenage pregnancy, an academic and mercantile history, and were the heartland of the historical Whigs and Republicans who preceded the modern Democratic Party.

And I wrote about another stratum centered in the South marked by poor education, gun culture, culture of violence, xenophobia, high teenage pregnancy, militarism, patriotism, country western music, and support for the Republican Party. And now I learn that the South was settled by a group noted even in the 1700s for its poor education, gun culture, culture of violence, xenophobia, high premarital pregnancy, militarism, patriotism, accent exactly like the modern country western accent, and support for the Democratic-Republicans who preceded the modern Republican Party.

If this is true, I think it paints a very pessimistic world-view. The “iceberg model” of culture argues that apart from the surface cultural features we all recognize like language, clothing, and food, there are deeper levels of culture that determine the features and institutions of a people: whether they are progressive or traditional, peaceful or warlike, mercantile or self-contained. We grudgingly acknowledge these features when we admit that maybe making the Middle East exactly like America in every way is more of a long-term project than something that will happen as soon as we kick out the latest dictator and get treated as liberators. Part of us may still want to believe that pure reason is the universal solvent, that those Afghans will come around once they realize that being a secular liberal democracy is obviously great. But we keep having deep culture shoved in our face again and again, and we don't know how to get rid of it. This has led to reasonable speculation that some aspects of it might even be genetic – something which would explain a lot, though not its ability to acculturate recent arrivals.

This is a hard pill to swallow even when we're talking about Afghanistan. But it becomes doubly unpleasant when we think about it in the sense of our neighbors and fellow citizens in a modern democracy. What, after all, is the point? A democracy made up of 49% extremely liberal Americans and 51% fundamentalist Taliban Afghans would be something very different from the democratic ideal; even if occasionally a super-charismatic American candidate could win over enough marginal Afghans to take power, there's none of the give-and-take, none of the competition within the marketplace of ideas, that makes democracy so attractive. Just two groups competing to dominate one another, with the fact that the competition is peaceful being at best a consolation prize.

If America is best explained as a Puritan-Quaker culture locked in a death-match with a Cavalier-Borderer culture, with all of the appeals to freedom and equality and order and justice being just so many epiphenomena – well, I'm not sure what to do with that information. Push it under the rug? Say "Well, my culture is better, so I intend to do as good a job dominating yours as possible?" Agree that We Are Very Different Yet In The End All The Same And So Must Seek Common Ground? Start researching genetic engineering? Maybe secede? I'm not a Trump fan much more than I'm an Osama bin Laden fan; if somehow Osama ended up being elected President, should I start thinking "Maybe that time we made a country that was 49% people like me and 51% members of the Taliban – *maybe that was a bad idea*“.

I don’t know. But I highly recommend Albion’s Seed as an entertaining and enlightening work of historical scholarship which will be absolutely delightful if you don’t fret too much over all of the existential questions it raises.

Why Were Early Psychedelicists So Weird?

[Epistemic status: very speculative, asserted with only ~30% confidence. On the other hand, even though psychiatrists don’t really talk about this it’s possible other groups know this all already]

A few weeks ago I gave a presentation on the history of early psychedelic research. Since I had a tough crowd, I focused on the fascinating biographies of some of the early psychedelicists.

Timothy Leary was a Harvard professor and former NIMH researcher who made well-regarded contributions to psychotherapy and psychometrics. He started the Harvard Psilocybin Project and several other Harvard-based experiments to test the effects of psychedelics on normal and mentally ill subjects. He was later fired from Harvard and arrested; later he accomplished a spectacular break out of prison and fled to Algeria. During his later life, he wrote books about how the human brain had hidden circuits of consciousness that would allow us to live in space, including a quantum overmind which could control reality and break the speed of light. He eventually fell so deep into madness that he started hanging out with Robert Anton Wilson and participating in Ron Paul fundraisers.

Richard Alpert was Leary’s co-investigator at the Harvard Psilocybin Project. He, too, had all the signs of a promising career, including a psychology PhD from Stanford, a visiting professorship at Berkeley, and a combination academic/clinical position at Stanford. After his work with Leary, he moved to India, changed his name to Baba Ram Dass, and became one of the world’s most prominent advocates for bhakti yoga.

John Lilly was a doctor, a neuroanatomy researcher, and an inventor who helped develop the principle behind many modern neuroprosthetics. He was always very strange, and did a lot of work in human-dolphin communication and SETI even before starting his work with LSD. But in the 1960s, he ran across Richard Alpert, joined in his LSD experiments, and became even stranger. He started writing books with names like “Programming And Metaprogramming The Human Biocomputer”, and arguing that benevolent and malevolent aliens were locked in a battle to manipulate Earth’s coincidences and with them the future of the human species. He became an expert yogi and claimed to have achieved samadhi, the highest state of union with God.

Kary Mullis is kind of cheating since he was not technically a psychedelicist. He was a biochemist in the completely unrelated field of bacterial iron transport molecules. But he did try LSD in 1966 back when it was still a legal research chemical. In fact he tried 1000 micrograms of it, one of the biggest doses I’ve ever heard of someone taking. Like the others, Mullis was a brilliant scientist – he won the Nobel Prize in Chemistry for inventing the polymerase chain reaction. Like the others, Mullis got *really weird* fast. He is a global warming denialist, HIV/AIDS denialist, and ozone hole denialist; on the other hand, he *does* believe in the efficacy of astrology. He also believes he has contacted extraterrestrials in the form of a fluorescent green raccoon, and “founded a business with the intent to sell pieces of jewelry containing the amplified DNA of deceased famous people like Elvis Presley”.

I wondered if there might be a selection bias in which psychedelicists I heard about, or that I might be cherry-picking the most unusual examples, so I looked for leading early psychedelics researchers I’d never heard of before and checked how weird *they* were. My sources told me that the two most important early psychedelicists were Humphry Osmond (who invented the word ‘psychedelic’ and may have been the first person to experiment with LSD rigorously) and his colleague John Smythies.

Osmond has an impressive early resume: started off as a surgeon, became a psychiatrist, did some well-regarded research into the structure of the human metabolite adrenochrome. And although he did not become fluorescent-alien-raccoon level weird, he can’t quite be called normal either. He became one of the founders of orthomolecular psychiatry, a discipline arguing that schizophrenia and other psychiatric diseases can be cured by massive amounts of vitamins – this is currently considered pseudoscience. His publications include the article “Selection of twins for ESP experimentation” in *International Journal of Parapsychology*, and a history of psychedelics records that “after his mescaline experiment in 1951, Dr. Osmond claimed to have successfully transmitted telepathic information to a fellow researcher, Duncan Blewett, who was also under the influence of mescaline, leading an independent observer to panic at the uncanny event.” He seems to have maintained a lifetime interest in parapsychology, Jungian typological analysis, and a field of his own invention called “socio-architecture”.

Smythies was a neuropsychiatrist, neuroanatomist, biochemist, EEG researcher, editor of the International Review of Neurobiology, etc, etc, etc (also, a cousin of Richard Dawkins). He is 94 but apparently still alive and going strong and making new neuroanatomical discoveries. He was one of the first people to investigate the pharmacology of psychedelics and helped with Osmond’s experiments in the early 1950s. He has also written *The Walls Of Plato’s Cave*, a book presenting a new theory of consciousness which “extends our concepts of consciousness and analyses possible geometrical and topological relations between phenomenal space and physical space linked to brane theory in physics" (I kind of wish I was a fly on the wall at his and Dawkins' family reunions).

My point is that the field of early psychedelic research seemed to pretty consistently absorb brilliant scientists, then spit out people who, while still brilliant scientists, also had styles of thought that could be described as extremely original at best and downright crazy at worst.

I think it's important to try to understand why.

First possibility: you had to be kind of weird to begin with in order to be interested in researching psychedelics. On the one hand, this is surely true; on the other, the early psychedelicists ended up *really* weird. At least in the early days I'm not sure psychedelics had the reputation for weirdness they now enjoy, and I'm also not sure that we're living in a world where a high enough percent of psychiatrists go off to become gurus in India, that we can just dismiss LSD research as happening to attract that type of person.

Second possibility: I know that almost all of these researchers (I'm not sure about Smythies) used psychedelics themselves. Psychedelic use is a sufficiently interesting experience that I can see why it might expand one's interest in the study of consciousness and the universe. Perhaps this is especially true if you're one of the first people to use it, and you don't have the social setting of "Oh, yeah, this is that drug that makes you have really weird experiences about consciousness for a while". If you're not aware that psychedelic hallucinations are a thing that happens, you might have to interpret your experience in more traditional terms like divine revelation. Under this theory, these pioneers had to become kind of weird to learn enough for the rest of us to use these substances safely. But why would that make John Lilly obsessed with aliens? Why would it turn Timothy Leary into a space colonization advocate and Ron Paul supporter?

The third possibility is the one that really intrigues me. A 2011 study found that a single dose of psilocybin *could permanently increase the personality dimension of Openness To Experience*. I'm emphasizing that because personality is otherwise pretty stable after adulthood; *nothing* should be able to do this. But magic mushrooms apparently have this effect, and not subtly either; participants who had a mystical experience on psilocybin had Openness increase *up to half a standard deviation* compared to placebo, and *the change was stable sixteen months later*. This is *really scary*. I mean, I *like* Openness To Experience, but something that can produce large, permanent personality changes is so far beyond anything else we have in psychiatry that it's kind of terrifying.

(related: 1972 study finds LSD may cause permanent increase in hypnotic susceptibility, which other sources have linked to being "fantasy prone" and "creative")

And that's *one dose*. These researchers were taking psychedelics pretty constantly for years, and probably experimented with the sort of doses you couldn't get away with giving research subjects. What would you expect to happen to *their* Openness To Experience? How many standard deviations do you think *it* went up?

It seems possible to me that psychedelics have a direct pharmacological effect on personality that causes people to be more open to unusual ideas. I know this is going against most of the latest research, which says psychedelics have no long-term negative mental health effects and do not cause psychosis. But there's a difference between being schizophrenic, and being the sort of guy who is still a leading neuroanatomist but also writes books about the geometric relationships between consciousness and the space-time continuum.

I'm not sure anyone has ever done studies to rule out the theory that psychedelics just plain make people *weird*. Indeed, such studies would be very difficult, given that weird people with very high Openness To Experience are more likely to use psychedelics. This problem would even prevent common sense detection of the phenomenon – even if we noticed that frequent psychedelic users were really weird, we would attribute it to selection effects and forget about it.

In this situation, the early psychedelicists could be a natural experiment giving us data we can't get any other way. Here are relatively sober scientists who took psychedelics for reasons other than being weird hippies already. Their fate provides signal through the noise which is the general psychedelic-using population.

I think this is only medium-risk; the explanation that weird people gravitate toward psychedelics, even in the sciences, is a strong one. But it's sufficient that I am hesitant to repeat the common view that psychedelics are not at all dangerous, or that they have no permanent side effects. There seems to me at least a moderate chance that they will make you more interesting without your consent – whether that is a good or a bad thing depends on exactly how interesting you want to be.

Be Nice, At Least Until You Can Coordinate Meanness

[Epistemic status: idea for one's toolbox of ideas; not to be followed off a cliff]

I.

Commenters on this blog have sometimes tried to shame or attack other commenters for perceived misdeeds like sexual promiscuity. They tell people to their faces that they're bad people and try to humiliate them.

When this happens, I ban the commenters involved.

And I get protests – what about free speech? What about the marketplace of ideas? Isn't shaming sometimes a useful social mechanism? There are some norms we can't or shouldn't codify into law; shouldn't violation of those norms be punished by shaming? Shaming can be very effective – for example, last week we learned the Puritans had a premarital pregnancy rate near zero because they publicly shamed anyone who departed from their moral standards. Might it not be useful to have something like that nowadays, either for premarital sex, or for other evils like homophobia and racism that we want to discourage? And even if I think we shouldn't, is it really okay to *ban* the people trying, seeing as they were probably well-intentioned?

I think my answer is: be nice, at least until you can coordinate meanness.

II.

A friend (I can't remember who) once argued that "be nice" provides a nigh-infallible ethical decision procedure. For example, enslaving people isn't very nice, so we know slavery is wrong. Kicking down people's doors and throwing them in prison for having a joint of marijuana isn't very nice, so we know the drug war is wrong. Not letting gays marry isn't very nice, so we know homophobia is wrong.

I counterargue that even if we ignore the ways our notion of "nice" itself packs up pre-existing moral beliefs, this heuristic fails in several important cases:

1. Refusing the guy who is begging you to give his drivers' license back, saying that without a car he won't be able to visit his friends and family or have any fun, and who is promising that he won't drive drunk an eleventh time.
2. Forcibly restraining a screaming baby while you jam a needle into them to vaccinate them against a deadly disease.
3. Sending the police to arrest a libertarian rancher in Montana who refuses to pay taxes for reasons of conscience
4. Revoking the credential (and thus destroying the future job prospects of) a teacher who has sex with one of her underage students

Sure, you could say that each of these "leads toward a greater niceness", like that you're only refusing the alcoholic his license in order to be nice to potential drunk driving victims. But then you've lost all meaningful distinction between the word "nice" and the word "good" and reinvented utilitarianism. And reinventing utilitarianism is pretty cool, but after you do that you no longer have such an easy time arguing against the drug war – somebody's going to argue that it leads to the greater good of there being fewer drugs.

We usually want to avoid meanness. In some rare cases, meanness is necessary. I think one check for whether a certain type of meanness might be excusable is – it's less likely to be excusable if it's not coordinated.

Consider: society demands taxes to pay for communal goods and services. This does sometimes involve not-niceness, as in the example of the rancher in (3). But what makes it tolerable is that it's done consistently and through a coordinated process. If the rule was "anybody who has a social program they want can take money from somebody else to pay for it," this would be anarchy. Some libertarians say "taxation is theft", but where arbitrary theft is unfair, unpredictable, and encourage perverse incentives like living in fear or investing in attack dogs, taxation has none of these disadvantages.

By the rule "be nice, at least until you can coordinate meanness", we should not permit individuals to rob each other at gunpoint in order to pay for social programs they want, but we might permit them to advocate for a coordinated national taxation policy.

Or: society punishes people for crimes, including the crime of libel. Punishment is naturally not-nice, but this seems fair; we can't just have people libeling each other all the time with no consequences. But what makes this tolerable is that it's coordinated – done through the court system according to carefully codified libel law that explains to everybody what is and isn't okay. Remove the coordination aspect, and you've got the old system where if you say something that offends my honor then I get some friends and try to beat you up in a dark alley. The impulse is the same: deploy not-niceness in the worthy goal of preventing libel. But one method is coordinated and the other isn't.

This is very, *very* far from saying that coordinated meanness is a sure test that means something's okay – that would be the insane position that anything legal must be ethical, something most countries spent the past few centuries disproving spectacularly. This is the much weaker claim that legality sets a *minimum* bar for people attempting mean policies.

As far as I can tell there are two things we want in a legal system. First, it should have good laws that produce a just society. But second, it should at least have clear and predictable laws that produce a safe and stable society.

For example, the first goal of libel law is to balance people's desire to protect their reputation with other people's desire for free speech. But the second goal of libel law should be that everybody understands what is and isn't libel. If a system achieves the second goal, nobody will end up jailed or dead because they said something they thought was totally innocent but somebody else thought was libel. And nobody will spent years and thousands of dollars entangled in an endless court case hiring a bunch of lawyers to debate whether some form of speech was acceptable or not.

So coordinated meanness is better than uncoordinated meanness not because it necessarily achieves the first goal of justice, but because it achieves the second goal of safety and stability. Everyone knows exactly when to expect it and what they can do to avoid it. I may not know what speech will or won't offend a violent person with enough friends to organize a goon squad, but I can always read the libel law and try to stay on the right side of it.

Likewise, in the Puritan community, I know exactly what things I have to do to avoid being shamed. Better still, I can only be shamed for violating one set of moral standards – the shared moral standards of the whole community. This isn't true of random people shaming promiscuous people, or people with the wrong opinion on race/gender issues, or whatever, on a private blog. Not only do most people reasonably expect to be able to do those things (and/or talk about those things here) without being shamed, but there are too many conflicting standards to meet – plausibly somebody could be shamed by traditionalists for being promiscuous, and by free-love people for not being promiscuous enough. Since shaming is unpleasant and supposed to act as a punishment, this is the equivalent of letting anybody beat up anybody else if they think they've broken an unwritten rule. It probably results in a lot of people being beaten up for not very much social change.

III.

The second reason that coordinated meanness is better than uncoordinated meanness is that it is less common. Uncoordinated meanness happens whenever one person wants to be mean; coordinated meanness happens when everyone (or 51% of the population, or an entire church worth of Puritans, or whatever) wants to be mean. If we accept theories like the wisdom of crowds or the marketplace of ideas – and we better, if we're small-d democrats, small-r republicans, small-l liberals, or basically any word beginning with a lowercase letter at all – then a big group of people all debating with each other will be harder to rile up than a single lunatic.

As a Jew, if I heard that skinheads were beating up Jews in dark alleys, I would be pretty freaked out; for all I know I could be the next victim. But if I heard that skinheads were circulating a petition to get Congress to expel all the Jews, I wouldn't be freaked out at all. I would expect almost nobody to sign the petition

(and in the sort of world where most people *were* signing the petition, I hope I would have moved to Israel long before anyone got any chance to expel me anyway)

Trying to coordinate meanness is not in itself a mean act – or at least, not as mean as actual meanness. If Westboro Baptist Church just published lots of pamphlets saying we should pass laws against homosexuality, maybe it would have made some gay people feel less wanted, but it would have been a lot less intense than picketing funerals. If people who are against promiscuity want to write books about why we should all worry about promiscuity, it might get promiscuous people a little creeped out, but a lot less so than going up to promiscuous people and throwing water on them and shouting “YOU STRUMPET!”

This is my answer to people who say that certain forms of speech make them feel unsafe, versus certain other people who demand the freedom to express their ideas. We should *all* feel unsafe around anybody who relishes uncoordinated meanness – beating people in dark alleys, picketing their funerals, shaming them, harassing them, doxxing them, getting them fired from their jobs. I have no tolerance for these people – I am sometimes forced to accept their existence because of the First Amendment, but I won’t do anything more.

On the other hand, we should feel *mostly* safe around people who agree that meanness, in the unfortunate cases where it’s necessary, must be coordinated. There is no threat at all from pro-coordination skinheads except in the vanishingly unlikely possibility they legally win control of the government and take over.

I admit that this safety is still only relative. It hinges on the skinheads’ inability to convert 51% of the population. But until the Messiah comes to enforce the moral law directly, safety has to hinge on *something*. The question is whether it should hinge on the ability of the truth to triumph in the marketplace of ideas in the long-term across an entire society, or whether it should hinge on the fact that you can beat me up with a baseball bat right now.

(if you want pre-Messianic absolute safety, there are some super-democratic mechanisms that might help. America's Bill of Rights seems pretty close to this; anyone wanting to coordinate meanness against a certain religion has to clear not only the 50% bar, but the much higher level required of Constitutional amendments. Visions of more complete protection remain utopian but alluring. For example, in an Archipelago you might well have absolute safety. The skinheads can't say "Let's beat up Jews right now", they can't even say "Let's start an anti-Jew political party and gradually win power". They can, at best, say, "Let's go found our own society somewhere else without any Jews", in which case you need say nothing but "don't let the door hit you on your way out". In this case their coordination of meanness cannot possibly hurt anyone.)

IV.

I've said many times I find the idea of "safe spaces" very attractive. I think they can be understood not just as spaces that are guaranteed safe for one group, but as spaces that have coordinated meanness against anything that threatens that group – ie they've agreed to shame, shun, and expel people who violate group norms. Everybody knows the local norms, and if somebody gets kicked out they can't say they weren't warned.

This is the principle with which I deal with the blog comments I started off by talking about. Right now people come to this blog with a default expectation that people aren’t going to be mean to them or try to shame them for things, other than the things universally agreed to be shameful in these general circles (like trolling, spamming, and misusing one-tailed t-tests). I want to explicitly reinforce that expectation here.

If you support being meaner in certain ways for the greater good, either as a subculture or as a society, you’re welcome to try to use this blog to advocate for that policy (within reason), but you’re *not* welcome to enact that policy unilaterally.

So here are two previously implicit SSC rules, made explicit for your edification:

First, you’re allowed to make (polite) arguments for why we should try shaming certain groups, but you are not allowed to directly shame any commenters here.

Second, you’re allowed to (politely) express your philosophical disagreements with the idea of transgender, but you are not allowed to actually misgender transgender commenters here.

Myers’ Race Car Versus The General Fitness Factor

[Epistemic status: I am not a geneticist, and even the geneticists I know aren’t sure about a lot of this. Take as speculation only.]

I.

PZ Myers argues against Stephen Hsu’s genetic engineering proposal here – a disappointing attitude toward mad science for a guy whose blog header is a crocodile-duck hybrid. The piece has a lot of errors, the worst of which other people have already discussed – but I want to talk about what I think is its strongest point. Myers writes:

> Note [Hsu’s] estimate of the number of genes that contribute to IQ: 10,000. That’s half the human genome! Hmmm. I wonder if any of those genes play a role in other processes in human physiology that might be affected by his plan?
>
> Here’s an analogy for you: let’s say a novice car designer has decided that the one quality of an automobile that is most important is speed, raw speed. He doesn’t know much about cars, so he asks more qualified engineers about what elements of the car contribute to acceleration and velocity, and they start off with the obvious…details of the engine, fuel mixes, etc. Then they’re talking tires. Aerodynamics. Weight. Pretty soon they have to admit that just about everything in the car is going to affect the speed at which it travels.
>
> So our blithe designer decides that making a fast car is simple: we just look at each component of the car one

> by one, and we pick an available option for it entirely on the basis of which option makes the car go faster. We'll easily be able to make a car that can rocket along at a thousand miles an hour, he thinks.
>
> But we have to ask whether we would want a car where the seats and steering were optimized for speed, where safety options were discarded, where something like visibility or reliability were jettisoned for the sole virtue of going really fast.

This makes a lot of sense. A car in which every component was optimized for speed would probably be uncomfortable, unsafe, ugly, difficult to maintain, and otherwise not the sort of car you want to drive. A human in which every component was optimized for intelligence might well be unhealthy, ugly, physically weak, antisocial, et cetera.

And we can do more than just hand-wave at an analogy to cars. Natural selection constantly weeds out worse alleles and replaces them with better ones. If an allele increases intelligence enough to improve reproductive fitness, but has no negative effects, it should sweep across all human populations in an amount of time proportional to its fitness benefit[1]. We see genetic evidence of various alleles sweeping various human populations in both the distant and recent past. These do not include the intelligence-boosting alleles Hsu is talking about.

For example, Hsu cites Rietvald et al's finding that rs1487441 is linked to cognitive ability (though it only gives you 0.3 extra IQ points, typical of the generally unimpressive contribution of single genes). About 20% of both Europeans and Japanese have the (A:A) variant, which suggests that however many thousands of years it's been since Europeans and Japanese diverged from each other isn’t long enough for the gene to undergo much selection. That means neither allele can have any overwhelming advantage, which means there must be some reason to have the opposite allele worth as much as 0.3 IQ points. I think this is the rigorous version of what Myers is saying.

Despite the fact that the race car argument makes perfect sense both analogically and rigorously, it seems to be wrong.

II.

What are the sorts of things we might trade off against intelligence? Perhaps fitness, height, attractiveness, health, longevity, social well-adjustedness?

But in fact none of these trade off against intelligence, many are strongly *positively* associated with it, and in some the link has been proven genetic!

People with high IQ tend to live longer. For example, a person with IQ 115 (85th percentile) is 20% more likely to survive to age 76 than an average person with IQ 100. One can of course posit many possible connections. Maybe high-IQ people are smart enough to eat healthy and exercise. Maybe rich people can afford both good schools and good doctors. Maybe good health behaviors protect the brain as well as the body and increase both IQ and longevity. But further investigation has cast doubt on all of these theories and strongly supports the hypothesis that no, the same genes that give you high intelligence also make you live longer. See for example the International Journal of Epidemiology: The Link Between Intelligence And Lifespan Is Mostly Genetic, which find genetics explain 95% of the correlation. A few of the genes linking intelligence and longevity may be already known; SSADH seems to be a contributing factor. My favorite study in this area, though, is one that is not yet complete: since all mammals are basically the same [citation needed] some London School of Economics researchers have developed an IQ test for dogs in the hope of checking whether the same correlation applies to them. Since canine intelligence doesn't affect things like diet, exercise, or tobacco status, a positive correlation in them too would help solidify the finding. We're still waiting on those results, but even without them the genetic hypothesis is looking pretty strong.

People with high IQ tend to be taller. This is interesting since height is often used as a measure of health and fitness during childhood, and since taller people get a bunch of advantages including being rated as more attractive and earning higher income. Once again we can imagine all sorts of possible confounders; once again studies find that the link is genetic. See for example Common Genetic Variants Explain The Majority Of The Correlation Between Height And Intelligence, The Genetic Correlation Between Height And IQ: Shared Genes Or Assortative Mating, Resolving The Genetic And Environmental Sources Of The Correlation Between Height And Intelligence, On The Height-Intelligence Correlation.

People with high IQ *may* be more attractive. This is the conclusion of a meta-analysis that finds a positive correlation between intelligence and body symmetry, usually used as a proxy for attractiveness unaffected by things like hairstyle and cosmetics; a second study failed to find this relationship. The jury is out on the positive link, but there certainly isn't the negative link that Myers' race car would predict.

People with high IQ commit much less crime – which is going to be our measure for social well-adjustedness here since it’s well studied. Once again it’s easy to think of possible confounders (I’ll add lead levels to the usual lot). Once again the studies show that at least some of the effect is genetic – here’s one on low-IQ/antisocial-behavior correlation in children, and here’s one cleverly linking fathers’ criminal history to sons’ vs. nephews’ IQ and then throwing enough statistics at it to find that the relationship is genetic. Likewise, the relationship between high IQ and low drug abuse seems to be genetic as well.

People with high IQ tend to be more physically fit. This is the conclusion of a study of 1.2 million Swedes. I don’t have any strong evidence that this relationship is genetically mediated (although Gottfredson on the fitness factor may be relevant here), I just want to note that, once again, there is less than zero evidence for Myers’ race car hypothesis.

People with high IQ have lower rates of heart disease, stroke, circulatory diseases, and diabetes. Intelligence may or may not decrease cancer risk, but again contra the race car hypothesis, it certainly does not increase it. Sibling designs suggest that shared family environment during youth is not responsible for the benefits; differential socioeconomic status as an adult may be, but this status is itself likely caused be the intelligence differences.

So despite its apparent plausibility the race car hypothesis crashes and burns.

III.

In one sense, this is bizarre. It’s as if somebody optimized every part of a race car for speed, and found that by coincidence this also made it the safest, most comfortable, and cheapest car on the road.

Yet if we think about it some more maybe it’s not *too* surprising. Consider Niels Bohr. He was a Nobel Prize winning physicist, professional football player, activist who helped Jews flee the Nazis, loving father of six children, and so healthy he kept doing science well into his seventies. And his talents show every sign of being at least partly genetic – his brother Harald was a leading mathematician, anti-Nazi activist, and Olympic silver medalist; one of Bohr’s children was *also* a Nobel Prize winning physicist and another was *also* an Olympic athlete. So it’s obviously *possible* to design a human with all-around great genes. Why does evolution restrict such designs to the Bohr family?

I can think of a few possibilities, all of which people who know more than I do are welcome to shoot down.

First, some of these all-around-beneficial genes could be good in heterozygosity but bad in homozygosity. We know something similar is true in the case of sickle-cell anaemia, which is mostly good in heterozygosity (protects vs. malaria) and very bad in homozygosity (causes sickle cell). This is exactly the sort of gene that should exist at a constant low frequency in the population, never getting more or less common. If the frequency got too low, then there’s no risk of two carriers mating, so evolution would encourage it as a free disease cure. If it got too high, evolution would discourage it – any carrier would probably marry another carrier and give their kids sickle cell. Suppose there are ten such genes, each of which grants higher intelligence on heterozygosity and has a frequency of 10% in the population. The average person will on average carry one such gene and have a 10% chance of a horrible genetic disease. Maybe Niels Bohr lucked out and carried all ten such genes without going homozygous on any. Maybe some other poor guy who is lost to history got all ten genes homozygous and died at birth of ten horrible genetic diseases at once.

This would make Hsu's gene-editing project very promising; all he would need to do is give everybody one copy of the relevant genes (and then *never let them mate*). But the hypothesis can't be quite right: I think it would predict that Niels Bohr's children would have unusually high rates of genetic diseases. In fact, the children of great men regress to the mean a little bit but show no signs at all of being unusually cursed.

Second, we could be talking not about polymorphisms but about mutational load. That means that there's some genome that works for humans (plus or minus a few hundred thousand polymorphisms that aren't too important at this level of analysis) and genetic health is determined by how many detrimental mutations you and your parents randomly accreted. If your mother spent too much time near the local nuclear reactor when she was pregnant, maybe you get a few hundred extra mutations and end up with lower IQ, a worse heart, less attractive features, et cetera. And this is obviously true in the case of a literal nuclear reactor, but I'm having trouble figuring out what plays the reactor role in real life. I know Greg Cochran and others have talked about things like paternal age at conception, climate, et cetera, but he applies these only to differences between populations. I'm not sure whether it would work out to expect a big difference in mutational load between Niels Bohr and his underachieving next-door neighbor. Maybe Bohr came from a long line of people who lucked out and got hit by unusually few cosmic rays? I don't know if this makes sense or not. Part of my problem might be that I still don't really understand how mutational load ever decreases – I've heard "the most heavily-loaded people are weeded out by natural selection", but it seems like that should only be able to slow the gradual universal dysgenesis.

This would also bode well for Hsu's project. In fact, it would make it even easier; it would reduce to the modal genome idea (make a baby whose genome, at each location, has the nucleotide which is most common at that location among all humans worldwide) which could be done without even performing the groundwork to see which genes do or don't influence intelligence.

Third, maybe all of these other good things are trading off against things that were important in the environment of evolutionary adaptedness but not today. Greg Cochran brings up infectious disease resistance; some commenters bring up calorie requirements. This latter seems especially plausible; the brain uses a lot of energy and energy was a scarce resource through much of evolutionary history. Either of these would explain why evolution kept the seemingly detrimental version around for so long, and why right now in our low-infection high-calorie modern civilization one allele or another seems to be an unalloyed good.

Again, this would bode well for Hsu's project, although the supergeniuses so produced would probably want to stay well away from any malarial swamps.

Or maybe it is some mixture of all four possibilities – Myers' trade-offs, heterozygosity, mutational load, and disease burden. The latter three could provide a sufficiently positive effect for intelligence to hide the negative effect of the first. I would be really surprised if something like this wasn't true – the theoretical argument for the first seems compelling, and it ought to happen at least a *little* even if it isn't the main source driving intelligence differences.

I was going to write that in this case we’d have to sort through every intelligence gene one-by-one to make sure we weren’t getting one of the trade-off ones, but maybe this isn’t true – a genius designed by Hsu’s method should have on average the same number of trade-offs versus unalloyed-goods as a genius born normally (right? or am I missing something?). Since most of us would prefer a natural-born baby with IQ 150 to a natural-born baby with IQ 100, it seems whatever trade-offs are necessary to reach that point are widely considered worth it. So unless there’s a difference I’m missing between normal recombination and Hsu’s method, we should be okay with designing an IQ 150 baby as well (from a purely health-related perspective, at least).

Whether this generalizes to creating an IQ IQ 200 or 300 baby depends not just on ethics, but on whether for some reason the costs and trade-offs of intelligence compound more than linearly. It’s possible, for example, that there are ten different genes coding for something that protects heart health, all of which can be traded off against intelligence. If you switch one gene from heart health to intelligence, whatever, you still have nine genes protecting your heart. But switching all of them at once would be a bad idea. I don’t know any reason to think this is true, but it’s a possibility that might give us pause between the IQ 150 and the IQ 300 level.

Overall though, I think the race car idea, despite its plausibility, is likely to be less of an impediment to genetic engineering than it might seem.

Only One Footnote, But It’s Really Long

1. This is an assumption I’m granting for the sake of argument, but possibly not true at the margin.

Consider that evolution doesn't care about intelligence nearly as much as we do. The most recent common ancestor of Europeans and Japanese wasn't going to use her intelligence to design a mammoth-seeking rocket. In fact, it's not totally clear why humans *did* evolve intelligence before the modern age; sure, tools are nice, but early hominids stuck to the same tools for a million years at a stretch; that doesn't exactly give a tight feedback loop to work with. The most convincing argument I've heard is the Machiavellian intelligence hypothesis which says that our ancestors used intelligence to navigate tribal politics and gain status within a social group.

But this theory would naively predict that the smartest person in high school would be the most popular. If intelligence is for gaining status, it seems to have diminishing returns beyond a certain point, which would explain why evolution didn't generally make us more intelligent even though greater-than-average intelligence is clearly possible (eg geniuses).

In the rare cases where evolution *did* have an incentive to evolve higher intelligence, it did so quickly and effectively. Several highly mercantile societies independently evolved the same set of genes producing higher IQ. The most notable were the Ashkenazi Jews, who have an average IQ 12-15 points higher than their European neighbors and whose genes show strong signatures of recent selection for intelligence; this most likely occurred during the Middle Ages when they were the mercantile class of Europe, since non-Ashkenazi Jews show no such effect. The genes involved tend to produce sphingolipidoses when homozygous, which shows a pretty good reason why evolution didn't do this kind of thing more often. Myers has previously dismissed this research, but I think wrongly – the paper itself considers and rejects the all the criticisms he raises (see pages 15 – 31). The very short summary is that Myers dismisses the genetic pattern as “variations amplified by chance,” but the expected level of chance variations can be calculated and this isn’t it. Ashkenazim have similar heterozygosity to other Europeans in neutral markers – ruling out a simple bottleneck – and the mutations involved are too potentially deleterious in homozygosity to persist for many generations by chance alone. Further, the mutations are all clustered in a few key pathways, many of which are clearly linked to intelligence. For example, Ashkenazim are at high risk for torsion dystonia, which is associated with higher IQ in sufferers.

So in response to the argument that evolution must trade off against something else, I would argue that evolution doesn’t share our exchange rate. Suppose that we could gain 20 IQ points at the cost of having larger heads that are harder to fit through a birth canal (remember, some of the known genes for intelligence are associated with head size, and the obstetrical dilemma used to be a big deal!). For hunter-gatherers, who had little use for IQ but lots of use for getting through birth canals, this was a bad deal and evolution didn’t take it. For moderns, who can use IQ points to cure cancer and explore space, and who have modern obstetric techniques, it’s a lot more attractive. So yes, let’s be cautious, but I think we’d all feel pretty stupid if we avoided bootstrapping our way to superintelligence out of fears of “things man was not meant to meddle with”, only to learn later that the whole problem could have been solved with c-sections.

Hidden Test Post

Hidden test post

Solidarity

[Epistemic status: I am not British, it's been years since I've been in the HSE, and the HSE is not the NHS. All of this may be misunderstood or outdated.]

I don't usually blog on labor disputes here, but I want to talk about one on which I have a tiny bit of inside knowledge.

Last month junior doctors in Britain went on strike for two days, protesting imposition of a new contract. There's a lot of anger about this, and admittedly when you're being rushed by ambulance into the emergency department for sudden onset chest pain, "doctors are on strike today" is not something you want to hear. My normal instincts would be to question whether this is really necessary. My experience tells me it is.

"Oh, you're a junior doctor. Of course you would support a doctor's strike." Okay, but I'm not a *British* junior doctor. I work in America, where I would describe conditions as "tough, but fair". Sure, Dr. Cox yells at you a lot, but only because he secretly thinks you're one of the best doctors ever to pass through the doors of this hospital. My own specialty of psychiatry is a lot better than most and overall I have little to complain about in my own life.

But that's not to say that I don't have any special knowledge here. I went to medical school in Ireland, where I worked alongside junior doctors in a system very much based off of the British one. And it was pretty shocking.

Technically European law caps junior doctor work weeks at 48 hours a week. Then again, *technically* American law caps junior doctor work weeks at 80 hours a week. My first week on a non-psychiatry service as an American junior doctor, I worked a bit over 100 hours – and so did everybody else I encountered. When I asked about the law, everyone just gave me that “oh you sweet summer child” look.

Such caps seem to be honored more in the breach than in the observance, and this is the British custom too. Physicians Weekly describes it as “the 48 hour trainee work week sham”, and the Telegraph and The Daily Mail both seem to agree that many British doctors are working 100 hour shifts. They seem to circumvent the law either by giving them a few weeks off afterwards and saying it “averages” to 48 hours/week, or else by doing what my hospital did – carefully schedule a 48 hour shift in big bold letters, assign 100 hours worth of work, and then get angry if anyone goes home before their work is done.

Many of the junior doctors I worked with in Ireland were working a hundred hours a week. It’s hard to describe what working 100 hours a week is like. Saying “it means you work from 7 AM to 9 PM every day including weekends” doesn’t really cut it. Imagine the hobbies you enjoy and the people you love. Now imagine you can’t spend time on any of them, because you are being yelled at as people die all around you for fourteen hours a day, and when you get home you have just enough time to eat dinner, brush your teeth, *possibly* pay a bill or two, and curl up in a ball before you have to go do it all again, and your next day off is in two weeks.

And this is *the best case scenario*, where everything is spaced out nice and even. The junior doctors I knew frequently worked thirty-six hour shifts at a time (the European Court of Human Rights has since declined to fine Ireland for this illegal practice). Dr. Brid McGrath (my lab partner in medical school) has been collecting some stories for the Irish media:

> My stories are like my colleagues' stories: working through illness, personal turmoil, and deprivation of sleep, food and toilet breaks. The worst stint was working 73 hours within an 82 hour period. I have been bullied, and to my shame, bullied others. I realised I was falling into the trap of treating others the way I had been treated. My self esteem faltered and I began to believe I truly was a nasty person. I had the insight to get help, but not everyone is so lucky.
>
> I came to talk to you about the imminent arrival of your very premature baby, at just 24 weeks. I held your hand and passed you tissues as we talked about his name, how tiny he was, how hard his life could be, but how we would try to give him the best possible chance… and how we might also have to accept the reality that he might not make it. That same day, I'd worked a 28-hour shift while I was 24 weeks pregnant myself. I fought back tears before I saw you. I worried about how I would cope with your pain and distress, barely able to think about the baby growing inside me. I had dinner a 1am, and worked on. An incredible nurse sat me down for a glass of water. She had to force me to. I was so busy.
>
> The other night, after a particularly busy 16-hour shift in the Emergency Department and Theatre, I went up to the wards to take blood samples for a patient. I'd had no dinner. A patient in the same room was handing around coconut buns, and gave me one. I inhaled it, it smelt so good. She then pushed the box into my hands and said "You look like you need them more than me!"

Imagine having to decide between going to the bathroom or getting a bag of crackers from the vending machine because you don't have enough time between cases to do both. Imagine having to remember the difference between nephritic syndrome and nephrotic syndrome (two *totally* different things) after ten hours of work, after getting three hours of sleep the night before. Imagine that you've just admitted a neurotic old woman to the hospital and you know in your heart that you should take her hand and explain to her in a soothing voice that everything is going to be okay, except that you already feel like every nerve of yours is beaten raw and you have three patients left to go before you can so much as sit down for a few minutes. Imagine your attending yelling at you because you got something wrong and saying you need to spend more time studying, and you trying to keep your mouth shut instead of telling him that you literally have only a half-hour in the day that could be considered free time by even the broadest stretch of the imagination and you are *damned* if you are going to spend that studying endocrinology.

The psychological consequences are predictable: after one year, 55% of junior doctors describe themselves as burned out, 30% meet criteria for moderate depression, and 12% report considering suicide.

A lot of American junior doctors are able to bear this by reminding themselves that it's only temporary. The worst part, internship, is only one year; junior doctorness as a whole only lasts three or four. After that you become a full doctor and a free agent – probably still pretty stressed, but at least making a lot of money and enjoying a modicum of control over your life.

In Britain, this consolation is denied most junior doctors. Everyone works for the government, and the government has a strict hierarchy of ranks, only the top of which – "consultant" – has anything like the freedom and salary that most American doctors enjoy. It can take ten to twenty years for junior doctors in Britain to become consultants, and some never do. In Ireland (I don't know about the UK) there was a very scary distinction between "training" and "service" positions, the former of which were always in short supply. Imagine that you're a freshman in college, and your university announces that due to budget cutbacks there are only about half as many sophomore positions available this year, so the top fifty percent of freshmen can go on to become sophomores, and the rest will have to stay freshmen until more money comes in. Also, there are no other colleges in the entire country so you have no choice but to follow along and hope for the best. This is what being a junior doctor is like.

Faced with all this, many doctors in Britain and Ireland have made the very reasonable decision to get the heck out of Britain and Ireland. The modal career plan among members of my medical school class was to graduate, work the one year in Irish hospitals necessary to get a certain certification that Australian hospitals demanded, then move to Australia. In Ireland, 47.5% of Irish doctors had moved to some other country. The situation in Britain is not *quite* so bad but rapidly approaching this point. Something like a third of British emergency room doctors have left the country in the past five years, mostly to Australia, citing "toxic environment" and "being asked to endure high stress levels without a break". Every year, about 2% of British doctors apply for the "certificates of good standing" that allow them to work in a foreign medical system, with junior doctors the most likely to leave. Doctors report back that Australia offers "more cash, fewer hours, and less pressure". I enjoy a pretty constant stream of Facebook photos of kangaroos and the Sydney Opera House from medical school buddies who are now in Australia and trying to convince their colleagues to follow in their footsteps.

Upon realizing their doctors are moving abroad, British and Irish health systems have leapt into action by…ignoring all systemic problems and importing foreigners from poorer countries who are used to inhumane work environments. I worked in some rural Irish towns where 99% of the population was white yet 80% of the doctors weren't; if you have a heart attack in Ireland and can't remember what their local version of 911 is, your best bet is to run into the nearest mosque, where you'll find all the town's off-duty medical personnel conveniently gathered together. This seems to be true of Britain as well, with the stats showing that almost 40% of British doctors trained in a foreign country (about half again as high as the US numbers, even though the US is accused of "stealing the world's doctors" – my subjective impression is that foreign doctors try to come to the US despite barriers because they're attracted to the prospect of a better life here, but that they are actively recruited to Britain out of desperation). Many of the doctors who *did* train in Britain are new immigrants who moved to Britain for medical school – for example, the *Express* finds that only 37% of British doctors are white British (the corresponding number for America is something like 50-65%, even though America is more diverse than Britain). While many new immigrants are great doctors, the overall situation is unfortunate since a lot of them end up underemployed compared to their qualifications in their home country, or trapped in the lower portions of the medical hierarchy by a combination of racism, language difficulties, and just the fact that *everyone* is trapped in the lower portions of the medical hierarchy these days.

If Britain continues along its current course, they'll probably be able to find more desperate people willing to staff its medical services after even more homegrown doctors move somewhere else (70% say they're considering it, although we are warned not to take that claim at face value). I work with several British and Irish doctors in my hospital here in the US Midwest, they're very talented people, and we could always use more of them. But this still seems like just a crappy way to run a medical system.

I don't know anything about the latest dispute that has led to this particular strike in Britain. Both sides' positions sound reasonable when I read about them in the papers. I would be tempted to just split the difference, if not for the fact several years of medical work in the British Isles have taught me that everything that a government health system says is *vile horrible lies*, and everybody with a title sounding like "Minister of Health" or "Health Secretary" is an Icke-style lizard person whose terminal value is causing as many humans to die of disease as possible. I can't overstate the importance of this. You read the press releases and they sound sort of reasonable, and then you talk to the doctors involved and they tell you all of the reasons why these policies have destroyed the medical system and these people are ruining their lives and the lives of their patients and how they once shook the Health Secretary's hand and it was ice-cold and covered in scales. I don't know how much of this is true. I just think of it as something in the background when the health service comes up to doctors and says "Hey, we have this *great new deal* we want to offer you!"

(I remember reporting into the hospital one day and seeing almost a carnival atmosphere, and one surgeon who had never been known to do anything but yell at his subordinates gave me a friendly nod and smile as he passed me in the corridor, and I started to worry I had walked into some Stepford Wives bizarro-world. Finally I learned that, the evening before, the Irish health minister had resigned in disgrace. This is the only time anyone ever saw that surgeon happy.)

[EDIT: a strong argument that the junior doctors have the right of it and the NHS' position is based on a misunderstanding of patient care statistics here]

Whatever caused this latest dispute is probably relevant mostly as a straw that breaks the camel's back. If British junior doctors today are anything like the Irish junior doctors of a few years ago, all of their complaints are legitimate and they're also hiding several dozen other legitimate complaints you have yet to hear about. I sort of sympathize with the government's complaints that they don't have enough money to make a system where doctors don't have to work a bunch of 36 hour shifts, but I feel like if you don't have enough money to run a health system that treats its employees like human beings, maybe you shouldn't be running your country's health system.

Labor disputes suck, and I have no good theory of them. Part of me is outraged at people being mistreated, and another part of me worries about a world where anybody who can convince the media that they're being oppressed can force other people into paying them whatever amount of money they think they deserve. I long for some kind of principled system that will solve these problems more elegantly than letting everybody shout their grievances at each other and seeing which ones stick. I long for something that will take care of the deeper problems underlying unfair labor practices like dualization of entire industries. This is why I find libertarian ideas like letting competition among firms determine people's pay and conditions so attractive.

But these may or may not work, insofar as they do work they only work in certain situations, and insofar as they do work under certain situations a 100% socialized industry run as a government monopoly probably isn't one of them. So we've got to do the thing where people get mistreated and have to cry out for redress of their grievances. And my experience tells me the grievances of British junior doctors are copious, horrifying, and entirely valid.

Skin In The Game

I.

One of the most interesting responses I got to my post supporting the junior doctors strike was by Salem, who said that this situation was (ethically) little different than that around adjunct professors, who also become overworked and miserable trying to break into a high-status profession. Salem very kindly didn't directly accuse me of hypocrisy, but maybe he should have.

While I sympathize with adjuncts' terrible conditions, my natural instinct is to say feedback mechanisms should keep doing their work. You can probably trace the argument- imagine a simplified toy model where the only two jobs are professor and salesperson, and being a professor is fun and high-status but being a salesperson is boring and low-status. Everyone will become a professor, and this will decrease the demand for professors and increase the demand for salespeople until the employers involved change their policies accordingly. Eventually it will stabilize where the nonmonetary advantages of being a professor are perfectly compensated by the monetary advantages of being a salesperson. If professors are getting paid shockingly little, it means the system is sending a signal that the nonmonetary advantages of being a professor are shockingly high, or else why would people keep trying? If we demand that professors get paid more, then we're letting them keep all their nonmonetary advantages over salespeople but demanding they have monetary advantages as well. It destroys the system's incentives to have people go into less fun but nevertheless necessary fields.

All of this makes perfect sense in the adjunct case. So why do I feel so differently in the doctor case?

Maybe for personal reasons. When I was in college, my two dueling career plans were doctor and philosophy professor. I brought this up with my professors, who universally told me not to go into academia. They told me that it was grueling, thankless, and for the vast majority of people involved doomed to failure, and that they couldn't in good conscience advise me to try it. I listened to their advice and became a doctor instead. It might not have *quite* the I-can't-believe-they're-paying-me-to-do-this amazingness of debating metaethics all day, but I still love it and it has much better career prospects.

So I guess you could argue that one reason I have less sympathy for adjuncts is that letting them achieve their goal would be a kick in the face to Past-Scott, who made what he considered the sober choice and went into a better-paying profession. If it turns out all I had to do was hang on a few years, and then the government would decree that people who got paid to argue about metaethics had to have career prospects as good as doctors', then I was a huge chump to try to do things the hard way. Maybe it's that fear of chumpness that makes it harder to sympathize.

And maybe the reason I feel such solidarity with doctors is that it's *not* supposed to be a profession you go into knowing you have no hope. Healing the sick is a lot more practical and socially-subsidizable an activity than pondering Truth; it seems like the sort of thing it should be easy to get paid for. Here in America, this is the conventional wisdom: make it into medicine, and you're promised a pretty good career. So maybe my solidarity with British doctors is a big cultural misunderstanding. Maybe, coming from America, I've absorbed a social promise that doctors will be treated well (which is true in America), but in Britain those doctors go into medical school knowing 100% that their lives will be unbearable and their compensation miniscule. Maybe they do it anyway because for them medicine is as much of an I-can't-believe-they're-paying-me-to-do-this as philosophy is to me.

Suppose somebody tells me that before going to medical school, every doctor in the British Isles has to sign a Waiver Of Appreciation Of Consequences, which spells out in excruciating detail all of the horrible things about a medical career. It says exactly how many 100 hour weeks they'll be expected to work, exactly how many 36-hour shifts they'll take, does its best to give them an idea how much senior doctors will abuse them. Maybe there's even a video portion that has interviews with disgruntled current doctors describing in explicit detail all the worst things that have happened to them. Maybe they even have to shadow current junior doctors, work 36 hour shifts themselves even though they can't do any procedures yet, just so that they can never claim that they didn't understand on an *experiential* level what a 36-hour shift *means*. Suppose the government (unlike in the real world) does this 100% fairly and accurately, and then they stick to it – ie as bad as things are, they never get any *worse* than the video promised. And suppose that after all of this, representatives from Private Industry come in to discuss their options for non-medical jobs, and explain how conditions in these jobs will probably be much better. If a medical student signs this waiver, completes their training, graduates, hates their job, and demands better pay and conditions, can we have any sympathy for them?

(I'm talking about a hypothetical world here. In the real world, we have to consider that lots of people *didn't* know what they were going into, and the thing they were going into keeps getting worse than they were led to expect. This is purely a Least Convenient Possible World argument here.)

The first argument we might use is that instead of focusing on the virtue of the employee, we should focus on the problems with society. That is, who cares whether the doctor made a good choice or not? Society still owes people decent jobs and working conditions. I think I reject this argument. There are *some* decent jobs attainable by the sorts of people who could become doctors (this is not true for everyone, but it's pretty true of the people who could become doctors), and not every alternative has to be great for everybody. Suppose there are many entry-level gardening jobs available, but I become a skyscraper-window-washer. Then I complain because I am afraid of heights, and I want special accomodations for this. When you say "maybe you should try a different line of work", I say "stop making this about my virtue and start focusing on the problems with society such that it can't give everyone decent conditions." It's not a very sympathetic argument. Society is *totally* allowed to have jobs that not everybody would enjoy as long as it gives everybody a wide range of options.

(one might argue that *nobody* could possibly enjoy being a junior doctor, but weirdly enough this is false. There are some people who love it. I usually assume these people are on cocaine, but maybe some of them aren't.)

The second argument we might use is that instead of focusing on the virtue of the employee, we should focus on distortions in the economy. Who decided to have a medical career consist of ten to twenty years of misery followed by an elusive opportunity to get a really nice job with great pay and hours? Isn't part of the problem that hospitals aren't competing for junior doctors? Can we oppose dualization of the labor force as just a generally bad idea? I think this argument is right, but I also think there's still a lingering question of "Okay, but since we do have this terrible industry with bad incentives, should we feel sympathy for people who voluntarily place themselves right in the middle of it?"

The third argument we might use is that we sometimes need to save people from their terrible decisions. Just as we ban people from permanently selling themselves into slavery, no matter how aware they are of the consequences when they make the deal, so maybe we should assume that signing the Waiver Of Appreciation of Consequences is an irrational decision motivated by time discounting and that people's future selves ought to be freed from the tyranny of their past selves' poor judgment. There's definitely some truth to this, but it's also a little too close to the policies of the BETA-MEALR Party for comfort. And doesn't this ignore that doctors *aren't* slaves, and can leave the medical profession any time they want? Even if there are some reasons they can't (difficulty finding other jobs, sunk cost effects) wouldn't it be less distortionary to smooth their path to non-medical careers than to try to reform medicine?

All of this makes sense. And yet as a psychiatrist, I *constantly* have people come into my office saying their jobs are making them suicidal. You would think somebody would leave a job before it makes them suicidal, but this doesn't always happen (the same is true of relationships, by the way). My suggestions that maybe they find a different job or boyfriend tend to fall on deaf ears.

I think one big reason I am so much more sympathetic to doctors than adjuncts is that I know them and have worked with them and I can see a reality – both a real reality and an emotional reality – that makes the model look kind of weak and threadbare.

II.

A few years ago I rented a house. In the rental contract, it said “tenant must have their own rental insurance to cover damages to the property”. I reminded my landlord that rental insurance mostly just covers damages to the renter’s stuff, and that the landlord usually has a separate property insurance to cover damages to the property, and that the way he was doing it might not even be legal. He just said yes it was and I had better get the rental insurance. I made a mental note to get the rental insurance and then got distracted by everything else in life and never got around to it.

A few years later, a pipe burst and the house flooded. The damage was assessed at way more than I had the ability to pay. I told the landlord he had better fix it, and the landlord told me I had better get my rental insurance to fix it for me – ie, the rental insurance I had forgotten to get.

There followed a gigantic disaster. I asked some lawyers whether the landlord was legally required to insure his own property, and they gave me vague and conflicting answers, then all agreed that even if he was it would cost more money than I had to fight the case in court. I ended up on the verge of breakdown. The landlord was clearly deliberately stonewalling me, trying to make it as hard as possible for me to figure out what was going on in the hopes that I gave in and gave him the money I didn’t have. Eventually, after several months’ of living out of a friend’s spare room while I tried to get the fate of the house sorted out, we settled out of court for a big fraction of my life’s savings.

(I later got some evidence that the landlord *did* have the property insured and was running some kind of insurance scam trying to get the money out of both me *and* his insurance, but that's a different story and I'm still not sure either way)

The reason I bring this up was that for the duration of the crisis, and to a lesser degree even to the present, I was utterly convinced that the government had an *obligation* to make sure landlords insure their properties. Offering a contract where the tenant was responsible for insuring the property? Totally unacceptable. Maybe even a human rights violation. What's the new phrase people are using these days? "A denial of your right to exist"? Even so.

It's easy to craft the argument where I'm in the wrong. Different landlords should be allowed to experiment with different arrangements as long as their terms are clearly listed in their contracts and the other contracting party agrees to them. I signed the contract then failed to do what the contract said, and clearly I needed to pay the penalty. We can imagine some nanny-state laws that might ameliorate that – for example, if a contract says a tenant must purchase insurance, it's the landlord's responsibility to make sure the insurance was actually purchased – but the arguments for implementing these laws are on really shaky ground.

And all I'm saying is that in the middle of this crisis, I had no sympathy for any of this. In the middle of this crisis, my thought was that I had worked really hard for years to get a little bit of money saved up, I was going to lose all of it for some burst pipe that wasn't my fault, this was cosmically unfair, somebody needed to do something about it, the landlord was a big company that probably had millions of dollars, and somebody needed to do something about this *right now*.

III.

An article by Freddie deBoer in this month's *Current Affairs* proposes "Journalistic Self-Outsourcing". DeBoer notes that lots of journalists and intellectuals suggest that protectionism and other anti-globalization policies are immoral. For example, Zack Beauchamp of Vox calls Bernie Sanders' skepticism of free trade "screwing the global poor"; Brad deLong calls the same "a call to keep China a society of poor subsistence rice farmers as long as possible – keep them poor, barefoot, uneducated, and by no means allow them to work at any of the high-value manufacturing occupations we want to keep in the United States."

DeBoer has a few things to say about how we should take money away from the rich in a way that can help both poor Americans and poor Chinese, but he quickly transitions into the barb he clearly relishes: if Beauchamp and DeLong are so in favor of poor Chinese people, how come they haven't outsourced *their* jobs? They both earn quite a bit of money, and China must have some decent journalists and economists who would love to telecommute to well-paid American positions. Are DeLong and Beauchamp hypocrites for not arranging to expedite the transition of their jobs to Chinese people?

One argument in their favor: they're currently living in a relatively globalized world, their employers must have the option of replacing them with Chinese people, but both of them still have jobs. This suggests that maybe journalism and economics aren't as replaceable as deBoer thinks. Beauchamp might be able to retort "As soon as Vox finds a Chinese journalist who's as good as I am they're welcome to hire him/her, but until then I can stay on, secure in the consistency of my principles.” Presumably he can say this without fear, since the subtleties of being in touch with the pulse of the American people and writing English-language articles would elude most lower-class Chinese.

But this lets them off too easily for purely contingent reasons. Even granting that they can’t be replaced with Chinese people right away, are they forced to *will* their replacement with Chinese people, deep in their hearts? Should Beauchamp go into work every morning asking his boss whether he’s found a suitable Chinese person to replace him with yet, and be disappointed every time his boss says no?

Again, there’s a contingent argument otherwise. If you’re really pro-globalization, you might believe that it’s impossible for the Chinese to take *all* our jobs, in the same way that the Luddite Fallacy says it’s impossible for robots to take all our jobs. The more Chinese take manufacturing jobs, the more Americans will have lots of money which will encourage new service jobs that the Chinese can’t easily take. If this is Beauchamp’s argument, he could say that we should globalize all the jobs that can be globalized, including his if possible, and then he will just move to an unglobalizable job. Since his job hasn’t been globalized yet, maybe he’s *already* in an unglobalizable job, so people should just stop bothering him.

But this is *still* too contingent. Let’s least convenient possible world again, and suppose that economists determine that there are permanent negative effects on American jobs from Chinese globalization, those effects disproportionately go to the poorest Americans, and no new unglobalizable jobs arise to restore cosmic balance. *Now* does Beauchamp have to will his own replacement?

Okay, argument from the other side: suppose a 1980s version of Beauchamp is writing an article against apartheid in South Africa. The South African whites argue that if apartheid ends, they'll be competing for jobs against much poorer blacks and so their quality of life will go way down. They say that if Beauchamp really wants to end apartheid, he himself should give up his journalist position to a South African black who will take it for a few dollars a day.

I'm both very convinced that the right thing to do in that situation would be to fight apartheid, and *also* convinced that the South Africans would be right about the personal jab – the median American journalist who pushed the fight *wouldn't* want his job taken over by Zulus willing to work for lower wages. When I put *myself* in that situation, and imagine myself being undercut by foreigners willing to practice medicine for almost nothing, I'm pretty pissed off about the idea too.

So maybe we should just let the journalist keep being a hypocrite. Journalists are pretty privileged people, and if we call journalists hypocrites every time they stand up for the less privileged without giving up all their own privilege, we'll probably just end up with journalists who stop standing up for the less privileged.

IV.

Consider two contradictory arguments.

The first says that people who experience a problem have unique insight into it, and people who don't experience it debate it from a position of ignorance or callousness. Thus a rich person can say "The poor don't need help because they can pull themselves up by their own bootstraps", but a poor person knows this is harder than it sounds. There's a really boring sense of this in which the poor person may know specific facts the rich person is missing – for example, the rich person might falsely think welfare is much more generous than it is. But then a rich person who's read a lot of books on poverty might claim to have a *better* perspective than the poor person, since she would likely know more about the welfare system. The more interesting claim is that there's a sort of near-mode-vs-far-mode thing going on here, where things that are easy to dismiss in the abstract become a lot more relevant in the concrete. A really good example of this is Hitchens' waterboarding – he said he thought it was an acceptable interrogation technique, some people offered to waterboard him to see if he changed his mind, and he quickly did. I'm fascinated by this incident because it's hard for me to understand, in a Mary's Room style sense, what he *learned* from the experience. He must have already known it was very unpleasant – otherwise why would you even support it as useful for interrogation? But somehow there's a difference between having someone explain to you that waterboarding is horrible, and undergoing it yourself. Just so with everything else. Under this view, the only people we should trust to tell us whether junior doctors get treated fairly are junior doctors; the only people we should trust to tell us whether globalization is acceptable are people whose own jobs are on the line.

The second says that people with skin in the game are the *last* people we should be trusting. Who do you trust to tell you how many subsidies the government should give the oil companies? Some economist in the budget watchdog organization who calculates exactly what the costs and benefits are? Or an oil company CEO who says "Trust me, we need lots and lots of money"? Under this argument, everyone has access to logic and reason, people detached from the situation are able to use it, and people within the situation are (perhaps excusably) motivated by self-interest. We can perhaps understand the fears of the white South African who thinks he'll lose his job and end up bankrupt, but it's our job as dispassionate external observers to notice that his concerns are outweighed by the concerns of other people whom his self-interest makes him unable to understand. And the *last* person you want giving you a sober cost-benefit analysis of torture is the person who is being waterboarded – everyone *knows* a waterboarding victim will say anything at all to make it stop!

The second argument obviously has its uses, but I'm fascinated by the first. In the few cases where I have direct experience with it, it seems to bring knowledge beyond just "this is really bad". In fact, the anti-junior-doctor argument seems resistant to just learning medicine is worse than you thought – if anything, medicine being very bad makes the argument stronger, since it means the doctor should be extra quick to listen to feedback and go into a different career. But somehow experience with doctors has made me much more reluctant to believe that argument. In the same way, one could imagine Hitchens saying "Yes, this waterboarding is really unpleasant – good thing that means it'll be really easy to make the terrorists want to talk" – but that wasn't the conclusion he drew from it.

I guess the thing I'm not sure about is – does personal experience/"skin in the game" reduce fully to factual propositions? Does a factory worker have an advantage over a journalist in understanding globalization just because he knows that being laid off is *really bad*, and that it's harder to get a new job than a journalist thinks – two things we would expect any journalist worth their salt to already know about? Or is there some hard-to-communicate knowledge that's neither factual nor just a cover for "the secret hard-to-communicate knowledge that I am selfish and want a system that benefits me rather than other people”?

Awkwardly, my far mode says that there isn’t and my near mode says that there is.

Teachers: Much More Than You Wanted To Know

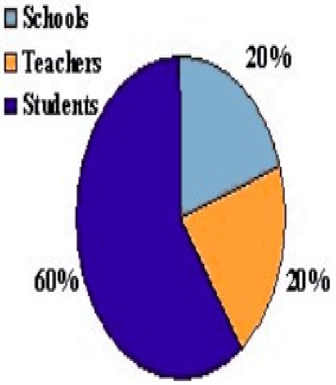
[Epistemic status: This is really complicated, this is not my field, people who have spent their entire lives studying this subject have different opinions, and I don't claim to have done more than a very superficial survey. I welcome corrections on the many inevitable errors.]

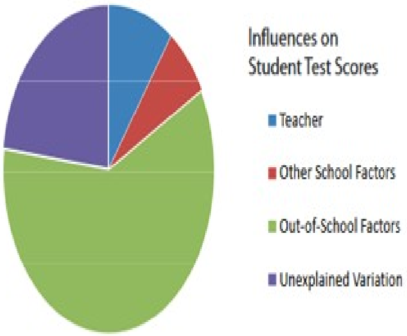
I.

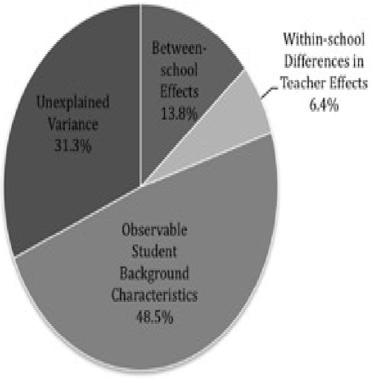
Newspapers report that having a better teacher for even a single grade (for example, a better fourth-grade teacher) can improve a child's lifetime earning prospects by $80,000. Meanwhile, behavioral genetics studies suggest that a child's parents have minimal (non-genetic) impact on their future earnings. So one year with your fourth-grade teacher making you learn fractions has vast effects on your prospects, but twenty-odd years with your parents shaping you at every moment doesn't? *Huh?* I decided to try to figure this out by looking into the research on teacher effectiveness more closely.

First, how much do teachers matter compared to other things? To find out, researchers take a district full of kids with varying standardized test scores and try to figure out how much of the variance can be predicted by what school the kids are in, what teacher's class the kids are in, and other demographic factors about the kids. So for example if the test scores of two kids in the same teacher's class were on average no more similar than the test scores of two kids in two different teachers' classes, then teachers can't matter very much. But if we were consistently seeing things like everybody in Teacher A's class getting A+s and everyone in Teacher B's class getting Ds, that would suggest that good teachers are very important.

Here are the results from three teams that tried this (source, source, source):







These differ a little in that the first one assumes away all noise ("unexplained variance") and the latter two keep it in. But they all agree pretty well that individual factors are most important, followed by school and teacher factors of roughly equal size. Teacher factors explain somewhere between 5% and 20% of the variance. Other studies seem to agree, usually a little to the lower end. For example, Goldhaber, Brewer, and Anderson (1999) find teachers explain 9% of variance; Nye, Konstantopoulos, and Hedges (2004) find they explain 13% of variance for math and 7% for reading. The American Statistical Association summarizes the research as "teachers account for about 1% to 14% of the variability in test scores", which seems about right.

So put more simply – on average, individual students' level of ~~ability~~ grit is what makes the difference. Good schools and teachers may push that a little higher, and bad ones bring it a little lower, but they don't work miracles.

(remember that right now we're talking about same-year standardized test scores. That is, we're talking about how much your fourth-grade history teacher affects your performance on a fourth-grade history test. If teacher effects show up anywhere, this is where it's going to be.)

Just as it's much easier to say "this is 40% genetic" than to identify particular genes, so it's much easier to say "this is 10% dependent on school-level factors and 10% based on teacher-level factors" then to identify what those school-level and teacher-level factors are. The Goldhaber study above tries its best, but the only school-level variable they can pin down is that having lots of white kids in your school improves test scores. And as far as I can tell, they don't look at socioeconomic status of the school or its neighborhood, which is probably what the white kids are serving as a proxy for. Even though these "school level effects" are supposed to be things like "the school is well-funded" or "the school has a great principal", I worry that they're capturing student effects by accident. That is, if you go to a school where everyone else is a rich white kid, chances are that means you're a rich white kid yourself. Although they try to control for this, having a couple of quantifiable variables like race and income probably doesn't entirely capture the complexities of neighborhood sorting by social class.

In terms of observable teacher-level effects, the only one they can find that makes a difference is gender (female teachers are better). Teacher certification, years of experience, certification, degrees, et cetera have no effect. This is consistent with most other research, such as Miller, McKenna, and McKenna (1998). A few studies that we'll get to later do suggest teacher experience matters; almost nobody wants to claim certifications or degrees do much.

One measurable variable not mentioned here *does* seem to have a strong ability to predict successful teachers. I'm not able to access these studies directly, but according to the site of the US Assistant Secretary of Education:

> The most robust finding in the research literature is the effect of teacher verbal and cognitive ability on student achievement. Every study that has included a valid measure of teacher verbal or cognitive ability has found that it accounts for more variance in student achievement than any other measured characteristic of teachers (e.g., Greenwald, Hedges, & Lane, 1996; Ferguson & Ladd, 1996; Kain & Singleton, 1996; Ehrenberg & Brewer, 1994).

So far most of this is straightforward and uncontroversial. Teachers account for about 10% of variance in student test scores, it's hard to predict which teachers do better by their characteristics alone, and schools account for a little more but that might be confounded. In order to say more than this we have to have a more precise way of identifying exactly *which* teachers are good, which is going to be more complicated.

II.

Suppose you want to figure out which teachers in a certain district are the best. You know that the only thing truly important in life is standardized test scores [citation needed], so you calculate the average test score for each teacher's class, then crown whoever has the highest average as Teacher Of The Year. What could go wrong?

But you'll probably just give the award to whoever teaches the gifted class. Teachers have classes with very different ability, and we already determined that ~~innate ability~~ grit explains more variance than teacher skill, so teachers who teach disadvantaged children will be at a big, uh, disadvantage.

So okay, back up. Instead of judging teachers by average test score, we can judge them by the average *change* in test score. If they start with a bunch of kids who have always scored around twentieth percentile, and they teach them so much that now the kids score at the fortieth percentile, then even though their kids are still below average they've clearly done some good work. Rank how many percentile points on average a teacher's students go up or down during the year, and you should be able to identify the best teachers for real this time.

Add like fifty layers of incomprehensible statistics and this is the basic idea behind VAM (value-added modeling), the latest Exciting Educational Trend and the lynchpin of President Obama's educational reforms. If you use VAM to find out which teachers are better than others, you can pay the good ones more to encourage them to stick around. As for the bad ones, VAM opponents are only being slightly unfair when they describe the plan as "firing your way to educational excellence".

A claim like "VAM accurately predicts test scores" is kind of circular, since test scores are what we used to determine VAM. But I *think* the people in this field try to use the VAM of class c to predict the student performance of class c + 1, or other more complicated techniques, and Chetty, Rothstein, and Rivkin, Hanushek, and Kane all find that a one standard deviation increase in teacher VAM corresponds to about a 0.1 standard deviation increase in student test scores.

Let's try putting this in English. Consider an average student with an average teacher. We expect her to score at exactly the 50th percentile on her tests. Now imagine she switched to the best teacher in the whole school. My elementary school had about forty teachers, so this is 97.5th percentile eg two standard deviations above the mean. A teacher whose VAM is two standard deviations above the mean should have students who score on average 0.2 standard deviations above the mean. Instead of scoring at the 50th percentile, now she’ll score at the 58th percentile.

Or consider the SAT, which is *not* the sort of standardized test involved in VAM but which at least everybody knows about. Each of its subtests is normed to a mean of 500 and an SD of 110. Our hypothetical well-taught student would go from an SAT of 500 to an SAT of 522. Meanwhile, average SAT subtest score needed to get into Harvard is still somewhere around 740. So this effect is nonzero but not very impressive.

But what happens if we compound this and give this student the best teachers many years in a row? Sanders and Rivers (also Jordan, Mendro, and Weerasinghe) argue the effects are impressive and cumulative. They compare students in Tennessee who got good teachers three years in a row to similar students who got bad teachers three years in a row (good = top quintile; bad = bottom quintile, so only 1/125 students was lucky or unlucky enough to qualify). The average bad-bad-bad student got scores in the 29th percentile; the average good-good-good student got scores in the 83rd percentile – which based on the single-teacher results looks super-additive. This is starting to sound a lot more impressive, and maybe Harvard-worthy after all. In fact, occasionally it is quoted as “four consecutive good teachers would close the black-white achievement gap” (I’m not sure whether this formulation requires also assigning whites to four consecutive bad teachers).

A RAND education report criticizes these studies as “using ad hoc methods” and argue that they’re vulnerable to double-counting student achievement. That is, we know that this teacher is the best because her students get great test scores; then later on we return and get excited over the discovery that the best teachers' students get great test scores. Sanders and Rivers did some complicated things that ought to adjust for that; RAND runs simulations and finds that depending on the true size of teacher effects vs. student effects, those complicated things may or may not work. They conclude that "[Sanders and Rivers] provide evidence of the existence and persistence of teacher or classroom effects, but the size of the effects is likely to be somewhat overstated".

Gary Rubinstein thinks he's debunked Sanders and Rivers style studies. I strongly disagree with his methods – he seems to be saying that the correlation between good teaching and good test scores isn't exactly one and therefore doesn't matter – but he offers some useful data. Just by eyeballing and playing around with it, it looks like most of the gain from these "three consecutive great teachers" actually comes from the last great teacher. So the superadditivity might not be quite right, and Sanders and Rivers might just be genuinely finding bigger teacher effects than anybody else.

At what rate do these gains from good teachers decay?

They decay pretty fast. Jacob, Lefgren and Sims find that only 25% of gains carry on to the next year, and only 15% to the year after that. That is, if you had a great fourth grade teacher who raised your test scores by x points, in fifth grade your test scores will be 0.25x higher than they would otherwise have been. Kane and Rothstein find much the same. A RAND report suggests 20% persistence after one year and 10% persistence after two. Jacob, Lefgren, and Sims find that only 25% of gains remain after one year, and about 13% after two years, after which it drops off much more slowly. All of this contradicts Sanders and Rivers pretty badly.

None of these studies can tell us whether the gains go all the way to zero after a long enough time. Chetty does these calculations and finds that they stabilize at 25% of their original value. But this number is *higher* than the two-year number for most of the other studies, plus Chetty is famous for getting results that are much more spectacular and convenient than anybody else's. I am really skeptical here. I remember a lot more things about last year than I do about twenty years ago, and even though I am pretty sure that my sixth grade teacher (for some weird reason) taught our class line dancing, I can't remember a single dance step. And remember Louis Benezet's early 20th century experiments with not teaching kids any math at all until middle school – after a year or two they were just as good as anyone else, suggesting a dim view of how useful elementary school math teachers must be. And even Chetty doesn't really seem to want to argue the point, saying that his results "[align] with existing evidence that improvements in education raise contemporaneous scores, then fade out in later scores".

In summary, I think there's pretty strong evidence that a +1 SD increase in teacher VAM can increase same-year test scores by + 0.1 SD, but that 50% – 75% of this effect decays in the first two years. I'm less certain how much these numbers change when one gets multiple good or bad teachers in a row, or how fully they decay after the first two years.

III.

When I started looking for evidence about how teachers affected children, I expected teachers' groups and education specialists to be pushing all the positive results. After all, what could be better for them than solid statistical proof that good teachers are super valuable?

In fact, these groups are the strongest *opponents* of the above studies – not because they doubt good teachers have an effect, but because in order to prove that effect you have to concede that good teaching is easy to measure, which tends to turn into proposals to use VAM to measure teacher performance and then fire underperformers. They argue that VAM is biased and likely to unfairly pull down teachers who get assigned ~~less intelligent~~ lower-grit kids.

It's always fun to watch rancorous academic dramas from the outside, and the drama around VAM is really a level above anything else I've seen. A typical example is the blog VAMboozled! with its oddly hypnotic logo and a steady stream of posts like Kane Is At It Again: "Statistically Significant" Claims Exaggerated To Influence Policy. Historian/researcher Diane Ravitch doesn't have quite as cute an aesthetic, but she writes things like:

> VAM is Junk Science. Looking at children as machine-made widgets and looking at learning solely as standardized test scores may thrill some econometricians, but it has nothing to do with the real world of children, learning, and teaching. It is a grand theory that might net its authors a Nobel Prize for its grandiosity, but it is both meaningless in relation to any genuine concept of education and harmful in its mechanistic and reductive view of humanity.

But tell us how you really feel.

I was originally skeptical of this, but after reading enough of these sites I think they have some good points about how VAM isn't always a good measure.

First, it seems to depend a lot on student characteristics; for example, it's harder to get a high VAM in a class full of English as a Second Language students. It makes perfect sense that ESL students would get low test scores, but since VAM controls for prior achievement you might expect them to get the same VAM anyway. They don't. Also, a lot of VAM models control for student race, gender, socioeconomic status, et cetera. I guess this is better than *not* doing this, but it seems to show a lack of confidence – if controlling for prior achievement was enough, you wouldn't *need* to control for these other things. But apparently people *do* feel the need to control for this stuff, and at that point I bring up my usual objection that you can never control for confounders enough, and also all to some degree these things are probably just lossy proxies for genetics which you *definitely* can't control for enough.

Maybe because of this, there's a lot of noise in VAM estimates. Goldhaber & Hansen (2013) finds that a teacher's VAM in year t is correlated at about 0.3 with their VAM in year t + 1. A Gates Foundation study also found reliabilities from 0.19 to 0.4, averaging about 0.3. Newton et al get slightly higher numbers from 0.4 to 0.6; Bessolo a wider range from 0.2 to 0.6. But these are all in the same ballpark, and Goldhaber and Hanson snarkily note that standardized tests aimed to assess *students* usually need correlations of 0.8 to 0.9 to be considered valid (the SAT, for example, is around 0.87). Although this suggests there's *some* component of VAM which is stable, it can't be considered to be "assessing" teachers in the same way normal tests assess students.

Even if VAM is a very noisy estimate, can't the noise be toned down by averaging it out over many years? I think the answer is yes, and I think the most careful advocates of VAM want to do this, but President Obama wants to improve education *now* and a lot of teachers don't have ten years worth of VAM estimates.

Also, some teachers complain that even averaging it out wouldn't work if there are *consistent* differences in student assignment. For example, if Ms. Andrews always got the best students, and Mr. Brown always got the worst students, then averaging ten years is just going to average ten years of biased data. Proponents argue that aside from a few obvious cases (the teacher of the gifted class, the teacher of the ESL class) this shouldn't happen. They can add school-fixed effects into their models (eg control for average performance of students at a particular school), leaving behind only teacher effects. And, they argue, which student in a school gets assigned which teacher ought to be random. Opponents argue that it might not be, and cite Paufler and Amrein-Beardsley's survey of principals, in which the principals all admit they don't assign students to classes randomly. But if you look at the study, the principals say that they're trying to be super-random – ie deliberately make sure that all classes are as balanced as possible. Even if they don't 100% achieve this goal, shouldn't the remaining differences be pretty minimal?

Maybe not. Rothstein (2009) tries to "predict" students' *fourth-grade* test scores using their *fifth-grade* teacher's VAM and finds that this totally works. Either schools are defying the laws of time and space, or for some reason the kids who do well in fourth-grade are getting the best fifth-grade teachers. Briggs and Domingue not only replicate these effects, but find that a fifth-grade teacher's "effects" on her students in fourth-grade is *just as big* as her effect on her students when she is actually teaching them, which would suggest that 100% of VAM is bias. Goldhaber has an argument for why there are statistical reasons this might not be so damning, which I unfortunately don't have enough ~~understanding~~ grit to evaluate.

Genetics might also play a role in explaining these results (h/t Spotted Toad's excellent post on the subject). A twin study by Robert Plomin does the classical behavioral genetics thing to VAM and finds that individual students' nth grade VAM is about 40% to 50% heritable. That is, the change in your test scores between third to fourth grade will probably be more like the change in your identical twin's test scores than like the change in your fraternal twin's test scores.

At first glance, this doesn't make sense – since VAM controls for past performance, shouldn't it be a pretty pure measure of your teacher's effectiveness? Toad argues otherwise. One of those Ten Replicated Findings From Behavioral Genetics is that IQ is more shared environmental in younger kids and more genetic in older kids. In other words, when you're really young, how smart you are depends on how enriched your environment is; as you grow older, it becomes more genetically determined.

So suppose that your environment is predisposing you to an IQ of 100, but your genes are predisposing you to an IQ of 120. And suppose (pardon the oversimplification) that at age 5 your IQ is 100, at age 15 it's 120, and change between those ages is linear. Then every year you could expect to gain 2 IQ points. Now suppose there's another kid whose environment is predisposing her to an IQ of 130, but whose genes are predisposing her to an IQ of 90. At age 5 her IQ is 130, at age 15 it's 90, and so every year she is losing 4 IQ points. And finally, suppose that your score on standardized tests is exactly 100% predicted by your IQ. Since you gain two points every year, in fifth grade you'll gain two points on your test, and your teacher will look pretty good. She'll get a good VAM, a raise, and a promotion. Since your friend loses four points every year, in fifth grade she'll lose four points on her test, and her teacher will look incompetent and be assigned remedial training.

This critique meshes nicely with the Rothstein test. Since you’re gaining 2 points every year, Prof. Rothstein can use your 5th grade gains of +2 points to accurately predict your fourth grade gain of +2 points. Then he can use your friend’s 5th grade loss of -4 points to accurately predict her fourth grade loss of -4 points.

This is a very neat explanation. My only concern is that it doesn’t explain decay effects very well. If a fifth grade teacher’s time-bending effect on students in fourth grade is exactly the same as her non-time-bending effect on students in fifth grade, how come her effect on her students once they graduate to sixth grade will only be 25% as large as her fifth grade effects? How come her seventh-grade effects will be smaller still? Somebody here has to be really wrong.

It would be nice to be able to draw all of this together by saying that teachers have almost no persistent effects, and the genetic component identified by Plomin and pointed at by Rothstein represents the 15 – 25% “permanent” gain identified by Chetty and others which so contradicts my lack of line dancing memories. But that would be just throwing out Briggs and Domingue’s finding that the Rothstein effect explains 100% of identified VAM.

One thing I kept seeing in the best papers on this was an acknowledgement that instead of arguing “VAMs are biased!” versus “VAMs are great!”, people should probably just agree that VAMs are biased, *just like everything else*, and start figuring out ways to measure exactly how biased they are, then use that number to determine what purposes they are or aren’t appropriate for. But I haven’t seen anybody doing this in a way I can understand.

In summary, there are many reasons to be skeptical of VAM. But some of these reasons contradict each other, and it's not clear that we should be *infinitely* skeptical. A big part of VAM is bias, but there might also be some signal within the noise, especially when it's averaged out over many years.

IV.

So let's go back to that study that says that a good fourth grade teacher can earn you $89,000. The study itself is Chetty, Friedman, and Rockoff (part 1, part 2). You may recognize Chetty as a name that keeps coming up, usually attached to findings about as unbelievable as these ones.

Bloomberg said that "a truly great" teacher could improve a child's earnings by $80,000, but I think this is mostly extrapolation. The number I see in the paper is a claim that a 1 SD better fourth-grade teacher can improve lifetime earnings by $39,000, so let's stick with that.

This sounds impressive, but imagine the average kid works 40 years. That means it's improving yearly earnings by about $1,000. Of note, the study didn't find this. They found that such teachers improved yearly earnings by about $300, but their study population was mostly in their late twenties and not making very much, and they extrapolated that if good teachers could increase the earnings of entry-level workers by $300, eventually they could increase the earnings of workers with a little more experience by $1000. The authors use a lot of statistics to justify this assumption which I'm not qualified to assess. But really, who cares? The fact that having a good fourth grade teacher can improve your adult earnings *any* measurable amount is the weird claim here. Once I accept *that*, I might as well accept $300, $1,000, or $500,000.

And here's the other weird thing. Everyone else has found that teacher effects on test scores decay very quickly over time. Chetty has *sort of* found that up to 25% of them persist, but he doesn't really seem interested in defending that claim and agrees that probably test scores just fade away. Yet as he himself admits, good teachers' impact on earnings works as if there were zero fadeout of teacher effects. He and his co-authors write:

> Our conclusion that teachers have long-lasting impacts may be surprising given evidence that teachers' impacts on test scores "fade out" very rapidly in subsequent grades (Rothstein 2010, Carrell and West 2010, Jacob, Lefgren, and Sims 2010). We confirm this rapid fade-out in our data, but find that teachers' impacts on earnings are similar to what one would predict based on the cross-sectional correlation between earnings and contemporaneous test score gains.

They later go on to call this a "pattern of fade-out and re-emergence", but this is a little misleading. The VAM never re-emerges on test scores. It only shows up in the earnings numbers.

All of this is really dubious, and it seems like Section III gives us an easy way out. There's probably a component of year-to-year stable bias in VAM, such that it captures something about student quality, maybe even innate ability, rather than just teacher quality. It sounds very easy to just say that this is the component producing Chetty's finding of income gains at age 28; students who have higher innate ability in fourth grade will probably still have it in their twenties.

Chetty is aware of this argument and tries to close it off. He conducts a quasi-experiment which he thinks replicates and confirms his original point: what happens when new teachers enter the school?

The thing we're most worried about is bias in student selection to teachers. If we take an entire grade of a school (for example, if a certain school has three fifth-grade teachers, we take all three of them as a unit) this should be immune to such effects. So Chetty looks at entire grades as old teachers retire and new teachers enter. In particular, he looks at such grades when a new teacher transfers from a different school. That new transfer teacher already has a VAM which we know from his work at the other school, which will be either higher or lower than the average VAM of his new school. If it's higher and VAM is real, we should expect the average VAM of that grade of his new school to go up a proportionate amount. If it's lower and VAM is real, we should expect the average VAM of that grade of his new school to go down a proportionate amount. Chetty investigates this with all of the transfer teachers in his data, finds this is in fact what happens, and finds that if he estimates VAM from these transfers he gets the same number (+ $1000 in earnings) that he got from the normal data. This is impressive. Maybe even *too* impressive. Really? The *same* number? So there's *no* bias in the normal data? I thought there was a lot of evidence that *most* of it was bias?

Rothstein is able to replicate Chetty's findings using data from a different district, but then he goes on to do the same thing on Chetty's quasi-experiment as he did on the normal VAMs, with the same results. That is, you can use the amount a school improves when a great new fifth-grade teacher transfers in to predict that teacher's students' fourth-grade performance. Not perfectly. But a little. For some reason, teacher transfers are having the same freaky time-bending effects as other VAM. Rothstein mostly explains this by saying that Chetty incorrectly excluded certain classes and teachers from his sample, although I don't fully understand this argument. He also gives one other example of when this might happen: suppose that a neighborhood is gentrifying. The new teachers who transfer in after the original teachers retire will probably be a better class of professional lured in by the improving neighborhood. And the school's student body will also probably be more genetically and socioeconomically advantaged. So better transfer teachers will be correlated with higher-achieving kids, but they won't have *caused* such high achievement.

After this came an increasingly complicated exchange between Rothstein and Chetty that I wasn't able to follow. Chetty, Friedman, and Rockoff wrote a 52 page Response To Rothstein where they argued that Rothstein's methodology would find retro-causal effects even in a fair experiment where none should exist. According to a 538 article on the debate, a couple of smart people (albeit smart people who already support VAMs and might be biased) think that Chetty's response makes sense, and even Rothstein agrees it "could be" true. 538 definitely thought the advantage in this exchange went to Chetty. But Rothstein responded with a re-replication of his results that he says addresses Chetty's criticisms but still finds the retro-causal effects indicating bias; as far as I know Chetty has not responded and nobody has weighed in to give me an expert opinion on whether or not it's right.

My temptation would usually be to say – here are some really weird results that can't possibly be true which we want to explain away, here's a widely-respected Berkeley professor of economics who says he's explained them away, great, let's forget about the whole thing. But there's one more experiment which I can't dismiss so easily.

V.

Project STAR (Student Teacher Achievement Ratio) was a big educational experiment in the 80s and 90s to see whether or not smaller class size improved student performance. That's a whole different can of worms, but the point is that in order to do this experiment for a while they randomized children to kindergarten classes within schools across 79 different schools. Since one of the biggest possible sources of bias for these last few studies has been possible nonrandom assignment of students to teachers, these Tennessee schools were an opportunity to get much better data than were available anywhere else.

So Chetty, Friedman, Higer, Saez, Schanzenbach, and Yagan analyzed the STAR data. They tried to do a lot of things with predicting earnings based on teacher experience, teacher credentials, and other characteristics, and it's a bit controversial whether they succeeded or not – see Bryan Caplan's analysis (1, 2) for more. Caplan is skeptical of a lot of the study, but one part he didn't address – and which I find most convincing – is based on something a lot like VAM.

Because of the random assignment, Chetty et al don't have to do full VAM here. It looks like their measure of kindergarten teacher quality is just the average of all their students' test scores (wait, kindergarteners are taking standardized tests now? I guess so.) When they're using teacher quality to predict the success of specific students, they use the average of all the test scores *except* that of the student being predicted, in order to keep it fair.

They find that the average test score of all the *other* students in your class, compared against the average score of all the students in other randomly assigned classes in your school, predicts your own test score. "A one percentile increase in entry-year class quality is estimated to raise own test scores by 0.68 percentiles, confirming that test scores are highly correlated across students within a classroom". This fades to approximately zero by fourth grade, confirming that the test-score-related benefits of having a good teacher are transient and decay quickly. *But,* students assigned to a one-percentile-higher class have average earnings that are 0.4% higher at age 25-27! And they say that this relationship is linear! So for example, the best kindergarten teacher in their dataset caused her class to perform at the 70th percentile on average, and these students earned about $17000 on average (remember, these are young entry-level workers in Tennessee) compared to the $15500 or so of their more average-kindergarten-teacher-having peers. Just their kindergarten teacher, totally apart from any other teacher in their life history, increased their average income 10%. Really, Chetty et al? *Really?*

But as crazy as it is, this study is hard to poke holes in. Even in arguing against it, Caplan notes that “it’s an extremely impressive piece” that “the authors are very careful”, and that it’s “one of the most impressive empirical papers ever written”. The experimental randomization means we can’t apply most of the usual anti-VAM measures to it. I don’t know, man. I just don’t know.

Okay, fine. I have one really long-shot possibility. Chetty et al derive their measure for teacher quality from the performance of all of the students in a class, excluding each student in turn as they try to predict his or her results. But this is only exogenous if the student doesn’t affect his or her peers’ test scores. But it’s possible some students *do* affect their peers’ test scores. If a student is a behavioral problem, they can screw up the whole rest of their class. Carrell finds that “exposure to a disruptive peer in classes of 25 during elementary school reduces earnings at age 26 by 3 to 4 percent”. Now, this in itself is a crazy, hard-to-believe study. But if we accept this second crazy hard-to-believe study, it might provide us with a way of attacking the first crazy hard-to-believe study. Suppose we have a really screwed-up student who is always misbehaving in class and disrupting the lesson. This lowers all his peers’ test scores and makes the teacher look low-quality. Then that kid grows up and remains screwed-up and misbehaving and doesn’t get as good a job. If this is a big factor in the differences in performances between classes, then so-called “teacher quality” might be conflated with a measure of how many children in their classes are behavioral problems, and apparent effects of teacher quality on earnings might just represent that misbehaving kids tend to become low-earning adults. I’m not sure if the magnitude of this effect checks out, but it *might* be a possibility.

But if we can’t make that work, we’re stuck believing that good kindergarten teachers can increase your yearly earnings by thousands of dollars. What do we make of that?

Again, everybody finds that test score gains do not last *nearly* that long. So it can’t be that kindergarten teachers provide you with a useful fund of knowledge which you build upon later. It can’t even be that kindergarten teachers stimulate and enrich you which raises your IQ or makes you love learning or anything like that. It has to be something orthogonal to test scores and measurable intellectual ability.

Chetty et al’s explanation is that teachers also teach “non-cognitive skills”. I can’t understand the regressions they use, but they say that although a one percentile increase in kindergarten class quality has a statistically insignificant increase (+ 0.05 percentiles) on 8th grade test scores, it has a statistically significant increase (+0.15 percentiles) on 8th grade non-cognitive scores (“non-cognitive scores” in this case are a survey where 8th grade teachers answer questions like “does this student annoy others?”) They then proceed to demonstrate that the persistence of these non-cognitive effects do a better job of predicting the earning gains than the test scores do. They try to break these non-cognitive effects into four categories: “effort”, “initiative”, “engagement” and “whether the student values school”, but the results are pretty boring and about equally loaded on all of them.

This does go together *really well* with my “behavioral problem” theory of the kindergarten class-earnings effect. The “quality” of a student’s kindergarten class, which might have more to do with the number of students who were behavioral problems in it than anything else, doesn’t correlate with future test scores but does correlate with future behavioral problems. It also seems to match Plomin’s point about how very early test scores are determined by environment, but later test scores are determined by genetics. A poor learning environment might be a really big deal in kindergarten, but stop mattering as much later on.

But this also goes together with some other studies that have found the same. The test scores gains from pre-K are notorious for vanishing after a couple of years, but a few really big preschool studies like the Perry Preschool Program found that such programs do not boost IQ but may have other effects (though to complicate matters, apparently Perry *did* boost later-life standardized test scores, just not IQ scores, and to further complicate matters, other studies find children who went to pre-K have worse behavior). This also sort of reminds me of some of the very preliminary research I’ve been linking to recently suggesting that excessively early school starting ages seem to produce an ADHD-like pattern of bad behavior and later-life bad effects, which I was vaguely willing to attribute to overchallenging kids’ brains too early while they’re still developing. If I wanted to be very mean (and I do!) I could even say that all kindergarten is a neurological insult that destroys later life prospects because of forcing students to overclock their young brains concentrating on boring things, but good teachers can make this less bad than it might otherwise be by making their classes a little more enjoyable.

But even if this is true, it loops back to the question I started with: there's strong evidence that parents have relatively little non-genetic impact on their childrens' life outcomes, but now we're saying that even a kindergarten teacher they only see for a year *does* have such an impact? And what's more, it's not even in the kindergarten teacher's unique area of comparative advantage (teaching academic subjects), but in the domain of *behavioral problems*, something that parents have like one zillion times more exposure to and power over?

I don't know. I still find these studies unbelievable, but don't have the sort of knock-down evidence to dismiss them that I'd like. I'm really impressed with everybody participating in this debate, with the quality of the data, and with the ability to avoid a lot of the usual failure modes. It's just not enough to convince me of anything yet.

VI.

In summary: teacher quality probably explains 10% of the variation in same-year test scores. A +1 SD better teacher might cause a +0.1 SD year-on-year *improvement* in test scores. This decays quickly with time and is probably disappears entirely after four or five years, though there may also be *small* lingering effects. It's hard to rule out the possibility that other factors, like endogenous sorting of students, or students' genetic potential, contributes to this as an artifact, and most people agree that these sorts of scores combine some signal with a lot of noise. For some reason, even though teachers' effects on test scores decay very quickly, studies have shown that they have significant impact on earning as much as 20 or 25 years later, so much so that kindergarten teacher quality can predict thousands of dollars of difference in adult income. This seemingly unbelievable finding has been replicated in quasi-experiments and even in real experiments and is difficult to banish. Since it does not happen through standardized test scores, the most likely explanation is that it involves non-cognitive factors like behavior. I really don’t know whether to believe this and right now I say 50-50 odds that this is a real effect or not – mostly based on low priors rather than on any weakness of the studies themselves. I don’t understand this field very well and place low confidence in anything I have to say about it.

Further reading: Institute of Education Science summary, Edward Haertel’s summary, TTI report, Adler’s critique of Chetty, American Statistical Society’s critique of Chetty/VAM, Chetty’s response, Ballou’s critique of Chetty

Three Great Articles On Poverty, And Why I Disagree With All Of Them

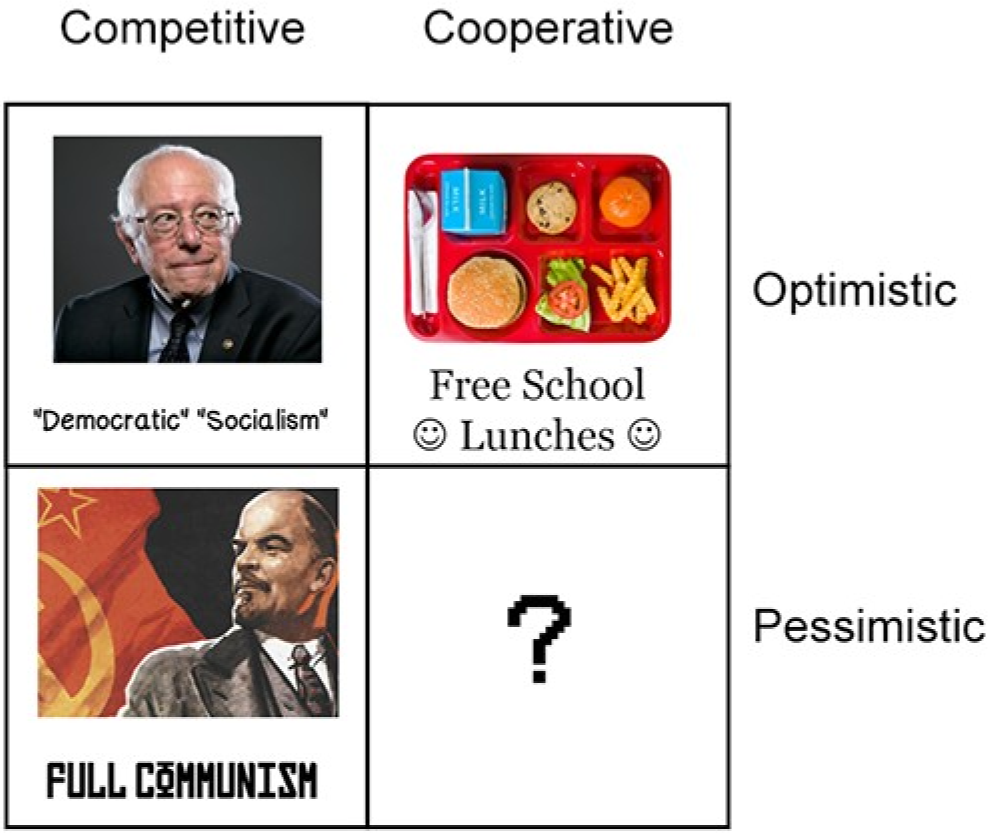
QZ: The universal basic income is an idea whose time will never come. Okay, maybe this one isn’t so great. It argues that work is ennobling (or whatever), that robots probably aren’t stealing our jobs, that even if we’re going through a period of economic disruption we’ll probably adapt, and that “if the goal is eliminating poverty, it is better to direct public funds to [failing schools and substandard public services]” then to try a guaranteed income scheme. It ends by saying that “I can’t understand why we’d consider creating and then calcifying a perpetually under-employed underclass by promoting the stagnation of their skills and severing their links to broader communities.”

(imagine a world where we had created and calcified a perpetually under-employed stagnant underclass. It sounds *awful*.)

More Crows Than Eagles: Unnecessariat. This one *is* great. A blogger from the Rust Belt reports on the increasing economic despair and frustration all around her, in the context of the recent spikes in heroin overdoses and suicides. There’s an important caveat here, in that at least national-level economic data paint a rosy picture: the unemployment rate is very low, consumer confidence is high, and the studies of technological unemployment suggest it’s not happening yet. Still, a lot of people on the ground – the anonymous blogger, the pathologists she worked with, and me from my position as a psychiatrist in the Midwest – feel like there’s a lot more misery and despair than the statistics suggest. MCTE replaces the old idea of the “precariat” – people who just barely have jobs and are worried about losing them – with her own coinage “unnecessariat” – people who don’t have jobs, are useless to the economy, and nobody cares what happens to them. It reminds me of the old argument of sweatshop-supporting economists – sure, we’re exploiting you, but you’d miss us if we left. She hates Silicon Valley for building its glittering megaplexes while ignoring everyone else, but she hates even more the people saying “Learn to code! Become part of the bright new exciting knowledge economy!” because realistically there’s no way an opioid-depended 55-year-old ex-trucker from Kentucky is going to learn to code. The only thing such people have left is a howl of impotent rage, and it has a silly hairstyle and is named Donald J. Trump.

Freddie deBoer: Our Nightmare. Also pretty great. The same things deBoer has been warning about for years, but expressed unusually clearly. By taking on the superficial mantle of center-leftism, elites sublimate the revolutionary impulse into a competition for social virtue points which ends up reinforcing and legitimizing existing power structures. Constant tally-keeping over what percent of obscenely rich exploitative Wall Street executives are people of color replaces the question of whether there should be obscenely rich exploitative Wall Street executives at all. As such tendencies completely capture the Democratic Party and the country’s mainstream left, genuine economic anger becomes more likely to be funneled into the right wing, where the elites can dismiss it as probably-racist (often with justification) and ignore it. “I cannot stress enough to you how vulnerable the case for economic justice is in this country right now. Elites agitate against it constantly…this is a movement, coordinated from above, and its intent is to solidify the already-vast control of economic elites over our political system…[Liberalism] is an attempt to ameliorate the inequality and immiseration of capitalism, when inequality and immiseration are the very purpose of capitalism.”

These articles all look at poverty in different ways, and I think that I look at poverty in a different way still. In the spirit of all the crazy political compasses out there, maybe we can learn something by categorizing them:


Competitive
Cooperative
Optimistic
Pessimistic
"Democratic" "Socialism"
Free School
☺ Lunches ☺
FULL COMMUNISM
?

Including only people who think society should be in the business of collectively helping the poor at all (ie no extreme libertarians or social Darwinists) and people who are interested in something beyond deBoer’s nightmare scenario (ie not just making sure every identity group has an equal shot at the Wall Street positions).

People seem to split into a competitive versus a cooperative view of poverty. To massively oversimplify: competitives agree with deBoer that “inequality and immiseration are the very purpose of capitalism” and conceive of ending poverty in terms of stopping exploitation and giving the poor their "just due" that the rich have taken away from them. The cooperatives argue that everyone is working together to create a nice economy that enriches everybody who participates in it, but some people haven't figured out exactly how to plug into the magic wealth-generating machine, and we should give them a helping hand ("here's government-subsidized tuition to a school where you can learn to code!"). Probably nobody's 100% competitive or 100% cooperative, but I think a lot of people have a tendency to view the problem more one way than the other.

So the northwest corner of the grid is people who think the problem is primarily one of exploitation, but it's at least somewhat tractable to reform. No surprises here – these are the types who think that the big corporations are exploiting people, but if average citizens try hard enough they can make the Man pay a $15 minimum wage and give them free college tuition, and then with enough small victories like these they can level the balance enough to give everybody a chance.

(These are all going to be straw men, but hopefully useful straw men)

The southwest corner is people who think the problem is primarily one of exploitation, but nothing within the system will possibly help. I put "full communism" in the little box, but I guess this could also be anarcho-syndicalism, or anarcho-capitalism, or theocracy, or Trumpism, or [insert your preferred poorly-planned form of government which inevitably fails here].

The northeast corner is people who think we're all in this together and there are lots of opportunities to help. This is the QZ writer who said we should be focusing on "education and public services". The economy is a benevolent force that wants to help everybody, but some people through bad luck – poor educational opportunities, not enough childcare, racial prejudice – haven’t gotten the opportunity they need yet, so we should lend them a helping hand so they can get back on their feet and one day learn to code. I named this quadrant “Free School Lunches” after all those studies that show that giving poor kids free school lunches improves their grades by X percent, which changes their chances of getting into a good college by Y percent, which increases their future income by Z percent, so all we have to do is have lots of social programs like free school lunches and then poverty is solved. But aside from the lunch people people, this category must also include libertarians who think that all we need to do is remove regulations that prevent the poor from succeeding, Reaganites who think that a rising tide will lift all boats, and conservatives who think the poor just need to be taught Traditional Hard-Working Values. Actually, probably 90% of the Overton Window is in this corner.

The southeast corner is people who think that we’re all in this together, but that helping the poor is really hard. They agree with the free school lunch crowd that capitalism is more the solution than the problem, and that we should think of this in terms of complicated impersonal social and educational factors preventing poor people from fitting into the economy. But the southeasterners worry school lunches won’t be enough. Maybe even hiring great teachers, giving everybody free health care, ending racism, and giving generous vocational training to people in need wouldn’t be enough. If we held a communist revolution, it wouldn’t do a thing: you can’t hold a revolution against skill mismatch. This is a very gloomy quadrant, and I don’t blame people for not wanting to be in it. But it’s where I spend most of my time.

The exploitation narrative seems fundamentally wrong to me – I'm not saying exploitation doesn't happen, nor even that it isn't common, just that isn't not the major factor causing poverty and social decay. The unnecessariat article, for all its rage against Silicon Valley hogging the wealth, half-admits this – the people profiled have become unnecessary to the functioning of the economy, no longer having a function even as exploited proletarians. Silicon Valley isn't exploiting these people, just ignoring them. Fears of technological unemployment are also relevant here: they're just the doomsday scenario where *all* of us are relegated to the unnecessariat, the economy having passed us by.

But I also can't be optimistic about programs to end poverty. Whether it's finding out that schools and teachers have relatively little effect on student achievement, that good parenting has even less, or that differences in income are up to fifty-eight percent heritable and a lot of what isn't outright genetic is weird biology or noise, most of the research I read is very doubtful of easy (or even hard) solutions. Even the most extensive early interventions have underwhelming effects. We can spend the collective energy of our society beating our head against a problem for decades and make no headway. While there may still be low-hanging fruit – maybe an scaled-up Perry Preschool Project, lots of prenatal vitamins, or some scientist discovering a new version of the unleaded-gasoline movement – we don't seem very good at finding it, and I worry it would be at most a drop in the bucket. Right now I think that a lot of variation in class and income is due to genetics and really deep cultural factors that nobody knows how to change *en masse*.

I can't even really believe that a rising tide will lift all boats anymore. Not only has GDP uncoupled from median wages over the past forty years, but there seems to be a Red Queen's Race where every time the GDP goes up the cost of living goes up the same amount. US real GDP has dectupled since 1900, yet a lot of people have no savings and are one paycheck away from the street. In theory, a 1900s poor person who suddenly got 10x his normal salary should be able to save 90% of it, build up a fund for rainy days, and end up in a much better position. In practice, even if the minimum wage in 2100 is $200 2016 dollar an hour, I expect the average 2100 poor person will be one paycheck away from the street. I can't explain this, I just accept it at this point. And I think that aside from our superior technology, I would rather be a poor farmer in 1900 than a poor kid in the projects today. More southeast corner gloom.

The only public figure I can think of in the southeast quadrant with me is Charles Murray. Neither he nor I would dare reduce *all* class differences to heredity, and he in particular has some very sophisticated theories about class and culture. But he shares my skepticism that the 55 year old Kentucky trucker can be taught to code, and I don't think he's too sanguine about the trucker's kids either. His solution is a basic income guarantee, and I guess that's mine too. Not because I have great answers to all of the QZ article's problems. But just because I don't have any better ideas[1,2].

The QZ article warns that it might create a calcified "perpetually under-employed stagnant underclass". But of course we already have such an underclass, and it's terrible. I can neither imagine them all learning to code, nor a sudden revival of the non-coding jobs they used to enjoy. Throwing money at them is a pretty subpar solution, but it's better than leaving everything the way it is and *not* throwing money at them.

This is why I can’t entirely sympathize with any of the essays I read on poverty, eloquent though they are.

Footnotes

1. And then there’s the rest of the world. Given the success of export capitalism in Korea, Taiwan, China, Vietnam, et cetera, and the pattern where multinationals move to some undeveloped country with cheap labor, boost the local economy until the country is developed and labor there isn’t so cheap anymore, and then move on to the next beneficiary – solving international poverty seems a lot easier than solving local poverty. All we have to do is keep wanting shoes and plastic toys. And part of me wonders – if setting up a social safety net would slow domestic economic growth – or even divert money that would otherwise go to foreign aid – does that make it a net negative? Maybe we should be optimizing for maximum economic growth until we’ve maxed out the good we can do by industrializing Third World countries? My guess is that enough of the basic income debate is about how to use existing welfare payments that this wouldn’t be too big a factor. And I would hope (for complicated reasons), that basic income would be more likely to help than hurt the economy[3].

2. Obviously invent genetic engineering and create a post-scarcity society, but until then we have to deal with this stuff.

3. And then there’s the whole open borders idea, which probably isn’t very compatible with basic income at all. Right now I think – I’ll explain at more length later – fully open borders is a bad idea, because the risk of it destabilizing the country and ruining the economic motor that lifts Third World countries out of poverty is too high.

Book Review: Age of Em

[Note: I really liked this book and if I criticize it that's not meant as an attack but just as what I do with interesting ideas. Note that Robin has offered to debate me about some of this and I've said no – mostly because I hate real-time debates and have bad computer hardware – but you may still want to take this into account when considering our relative positions. Mild content warning for murder, rape, and existential horror. Errors in Part III are probably my own, not the book's.]

I.

There are some people who are destined to become adjectives. Pick up a David Hume book you've never read before and it's easy to recognize the ideas and style as Humean. Everything Tolkien wrote is Tolkienesque in a non-tautological sense. This isn't meant to denounce either writer as boring. Quite the opposite. They produced a range of brilliant and diverse ideas. But there was a hard-to-define and very consistent ethos at the foundation of both. Both authors were *very much like themselves*.

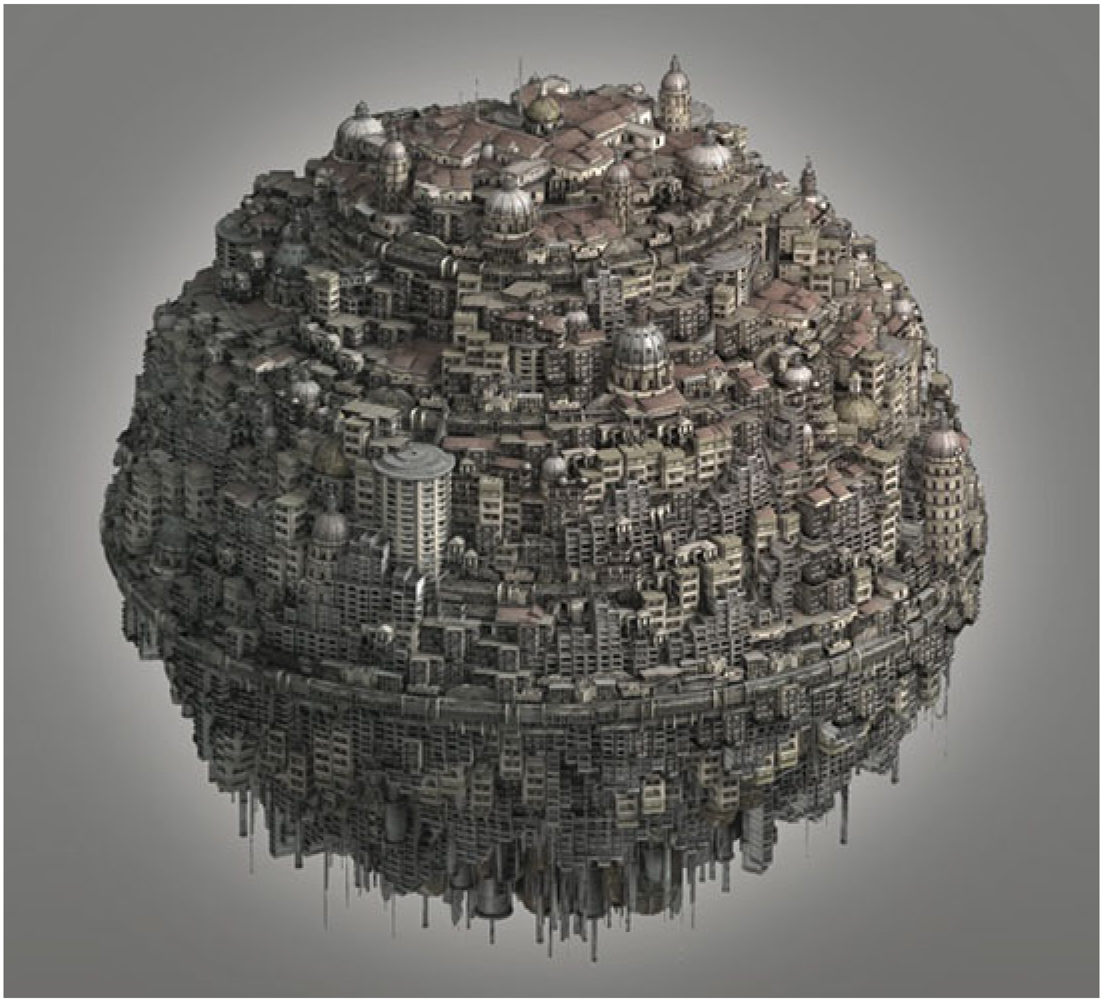
Robin Hanson is more like himself than anybody else I know. He's obviously brilliant – a PhD in economics, a masters in physics, work for DARPA, Lockheed, NASA, George Mason, and the Future of Humanity Institute. But his greatest aptitude is in being really, *really* Hansonian. Bryan Caplan describes it as well as anybody:

> When the typical economist tells me about his latest research, my standard reaction is 'Eh, maybe.' Then I forget about it. When Robin Hanson tells me about his latest research, my standard reaction is 'No way! Impossible!' Then I think about it for years.

This is my experience too. I think I said my first "No way! Impossible!" sometime around 2008 after reading his blog

Overcoming Bias. Since then he’s influenced my thinking more than almost anyone else I’ve ever read. When I heard he was writing a book, I was – well, I couldn’t even imagine a book by Robin Hanson. When you read a thousand word blog post by Robin Hanson, you have to sit down and think about it and wait for it to digest and try not to lose too much sleep worrying about it. A whole book would be *something*.

I have now read *Age Of Em* (website)and it is indeed something. Even the cover gives you a weird sense of sublimity mixed with unease:



And in this case, judging a book by its cover is entirely appropriate.

II.

Age of Em is a work of futurism – an attempt to predict what life will be like a few generations down the road. This is not a common genre – I can’t think of another book of this depth and quality in the same niche. Predicting the future is notoriously hard, and that seems to have so far discouraged potential authors and readers alike.

Hanson is not discouraged. He writes that:

> Some say that there is little point in trying to foresee the non-immediate future. But in fact there have been many successful forecasts of this sort. For example, we can reliably predict the future cost changes for devices such as batteries or solar cells, as such costs tend to follow a power law of the cumulative device production (Nagy et al 2013). As another example, recently a set of a thousand published technology forecasts were collected and scored for accuracy, by comparing the forecasted date of a technology milestone with its actual date. Forecasts were significantly more accurate than random, even forecasts 10 to 25 years ahead. This was true separately for forecasts made via many different methods. On average, these milestones tended to be passed a few years before their forecasted date, and sometimes forecasters were unaware that they had already passed (Charbonneau et al, 2013).
>
> A particularly accurate book in predicting the future was *The Year 2000*, a 1967 book by Herman Kahn and Anthony Wiener. It accurately predicted population, was 80% correct for computer and communication technology, and 50% correct for other technology (Albright 2002). On even longer time scales, in 1900 the engineer John Watkins did a good job of forecasting

> many basic features of society a century later (Watkins 1900) […]
>
> Some say no one could have anticipated the recent big changes associated with the arrival and consequences of the World Wide Web. Yet participants in the Xanadu hypertext project in which I was involved from 1984 to 1993 correctly anticipated many key aspects of the Web […] Such examples show that one can use basic theory to anticipate key elements of distant future environments, both physical and social, but also that forecasters do not tend to be much rewarded for such efforts, either culturally or materially. This helps to explain why there are relatively few serious forecasting efforst. But make no mistake, it *is* possible to forecast the future.

I think Hanson is overstating his case. All except Watkins were predicting only 10 – 30 years in the future, and most of their predictions were simple numerical estimates, eg “the population will be one billion” rather than complex pictures of society. The only project here even remotely comparable in scope to Hanson’s is John Watkins’ 1900 article.

Watkins is classically given some credit for broadly correct ideas like “Cameras that can send pictures across the world instantly” and “telephones that can call anywhere in the world”, but of his 28 predictions, I judge only eight as even somewhat correct. For example, I grant him a prediction that “the average American will be two inches taller because of good medical care” even though he then goes on to say in the same sentence that the average life expectancy will be fifty and suburbanization will be so total that building city blocks will be illegal (sorry, John, only in San Francisco). Most of the predictions seem simply and completely false. Watkins believes all animals and insects will have been eradicated. He believes there will be “peas as large as beets” and “strawberries as large as apples” (these are two separate predictions; he is weirdly obsessed with fruit and vegetable size). We will travel to England via giant combination submarine/hovercrafts that will complete the trip in a lightning-fast two days. There will be no surface-level transportation in cities as all cars and walkways have moved underground. The letters C, X, and Q will be removed from the language. Pneumatic tubes will deliver purchases from stores. “A man or woman unable to walk ten miles at a stretch will be regarded as a weakling.”

Where Watkins is right, he is generally listing a cool technology slightly beyond what was available to his time and predicting we will have it. Nevertheless, he is still mostly wrong. Yet this is Hanson’s example of accurate futurology. And he is *right* to make it his example of accurate futurology, because everything else is even worse.

Hanson has no illusions of certainty. He starts by saying that “conditional on my key assumptions, I expect at least 30% of future situations to be usefully informed by my analysis. Unconditionally, I expect at least 10%.” So he is not explicitly overconfident. But in an implicit sense, it’s just *weird* to see the level of detail he tries to predict – for example, he has two pages about what sort of swear words the far future might use. And the book’s style serves to reinforce its weirdness. The whole thing is written in a sort of professorial monotone that changes little from loving descriptions of the sorts of pipes that will cool future buildings (one of Hanson’s pet topics) to speculation on our descendents’ romantic relationships (key quote: “The per minute subjective value of an equal relation should not fall much below half of the per-minute value of a relation with the best available open source lover”). And it leans heavily on a favorite Hansonian literary device – the weirdly general statement about something that sounds like it can’t possibly be measurable, followed by a curt reference which if followed up absolutely confirms said statement, followed by relentlessly ringing every corollary of it:

> Today, mental fatigue reduces mental performance by about 0.1% per minute. As by resting we can recover at a rate of 1% per minute, we need roughly one-tenth of our workday to be break time, with the duration between breaks being not much more than an hour or two (Trougakos and Hideg 2009; Alvanchi et al 2012)… Thus many em tasks will be designed to take about an hour, and many spurs are likely to last for about this duration.

Or:

> Today, painters, novelists, and directors who are experimental artists tend to do their best work at roughly ages 46-52, 38-50, and 45-63 respectively, but those ages are 24-34, 29-40, and 27-43, respectively for conceptual artists (Galenson 2006)…At any one time, the vast majority of actual working ems [should be] near a peak productivity subjective age.

Or:

> Wars today, like cities, are distributed evenly across all possible war sizes (Cederman 2003).

At some point I started to wonder whether Hanson was putting me on. Everything is just played *too straight*. Hanson even addresses this:

> To resist the temptation to construe the future too abstractly, I’ll try to imagine a future full of complex detail. One indiciation that I’ve been successful in all

> these efforts will be if my scenario description sounds less like it came from a typical comic book or science fiction movie, and more like it came form a typical history text or business casebook.

Well, count that project a success. The effect is strange to behold, and I'm not sure it will usher in a new era of futurology. But *Age of Em* is great not *just* as futurology, but as a bunch of different ideas and purposes all bound up in a futurological package. For example:

– An introduction to some of the concepts that recur again and again across Robin's thought – for example, near vs. far mode, the farmer/forager dichotomy, the inside and outside views, signaling. Most of us learned these through years reading Hanson's blog *Overcoming Bias*, getting each chunk in turn, spending days or months thinking over each piece. Getting it all out of a book you can read in a couple of days sounds really hard – but by applying them to dozens of different subproblems involved in future predictions, Hanson makes the reader more comfortable with them, and I expect a lot of people will come out of the book with an intuitive understanding of how they can be applied.

– A whirlwind tour through almost every science and a pretty good way to learn about the *present*. If you didn't already know that wars are distributed evenly across all possible war sizes, well, read *Age of Em* and you will know that and many similar things besides.

– A manifesto. Hanson often makes predictions by assuming that since the future will be more competitive, future people are likely to converge toward optimal institutions. This is a dangerous assumption for futurology – it's the same line of thinking that led Watkins to assume English would abandon C, X, and Q as inefficient – but it's a *great* assumption if you want a chance to explain your ideas of optimal institutions to thousands of people who think they're reading fun science-fiction. Thus, Robin spends several pages talking about how ems may use prediction markets – an information aggregation technique he invented – to make their decisions. In the real world, Hanson has been trying to push these for decades, with varying levels of success. Here, in the guise of a future society, he can expose a whole new group of people to their advantages – as well as the advantages of something called "combinatorial auctions" which I am still not smart enough to understand.

– A mind-expanding drug. One of the great risks of futurology is to fail to realize how *different* societies and institutions can be – the same way uncreative costume designers make their aliens look like humans with green skin. A lot of our thoughts about the future involve assumptions we've never really examined critically, and Hanson dynamites those assumptions. For page after page, he gives strong arguments why our descendants might be poorer, shorter-lived, less likely to travel long distances or into space, less progressive and open-minded. He predicts little noticeable technological change, millimeter-high beings living in cities the size of bottles, careers lasting fractions of seconds, humans being incomprehensibly wealthy patrons to their own robot overlords. And *all of it makes sense*.

When I read Stross' *Accelerando*, one of the parts that stuck with me the longest were the Vile Offspring, weird posthuman entities that operated a mostly-incomprehensible Economy 2.0 that humans just sort of hung out on the edges of, goggle-eyed. It was a weird vision – but, for Stross, mostly a black box. *Age of Em* opens the box and shows you every part of what our weird incomprehensible posthuman descendents will be doing in loving detail. Even what kind of swear words they'll use.

III.

So, what is the Age of Em?

According to Hanson, AI is really hard and won't be invented in time to shape the posthuman future. But sometime a century or so from now, scanning technology, neuroscience, and computer hardware will advance enough to allow emulated humans, or "ems". Take somebody's brain, scan it on a microscopic level, and use this information to simulate it neuron-by-neuron on a computer. A good enough simulation will map inputs to outputs in exactly the same way as the brain itself, effectively uploading the person to a computer. Uploaded humans will be much the same as biological humans. Given suitable sense-organs, effectuators, virtual avatars, or even robot bodies, they can think, talk, work, play, love, and build in much the same way as their "parent". But ems have three very important differences from biological humans.

First, they have no natural body. They will never need food or water; they will never get sick or die. They can live entirely in virtual worlds in which any luxuries they want – luxurious penthouses, gluttonous feasts, Ferraris – can be conjured out of nothing. They will have some limited ability to transcend space, talking to other ems' virtual presences in much the same way two people in different countries can talk on the Internet.

Second, they can run at different speeds. While a normal human brain is stuck running at the speed that physics allow, a computer simulating a brain can simulate it faster or slower depending on preference and hardware availability. With enough parallel hardware, an em could experience a subjective century in an objective week. Alternatively, if an em wanted to save hardware it could process all its mental operations *very slowly* and experience only a subjective week every objective century.

Third, just like other computer data, ems can be copied, cut, and pasted. One uploaded copy of Robin Hanson, plus enough free hardware, can become a thousand uploaded copies of Robin Hanson, each living in their own virtual world and doing different things. The copies could even converse with each other, check each other's work, duel to the death, or – yes – have sex with each other. And if having a thousand Robin Hansons proves too much, a quick ctrl-x and you can delete any redundant ems to free up hard disk space for Civilization 6 (coming out this October!)

Would this count as murder? Hanson predicts that ems will have unusually blase attitudes toward copy-deletion. If there are a thousand other copies of me in the world, then going to sleep and not waking up just feels like delegating back to a different version of me. If you're still not convinced, Hanson's essay Is Forgotten Party Death? is a typically disquieting analysis of this proposition. But whether it's true or not is almost irrelevant – at least *some* ems will think this way, and they will be the ones who tend to volunteer to be copied for short term tasks that require termination of the copy afterwards. If you personally aren't interested in participating, the economy will leave you behind.

The ability to copy ems as many times as needed fundamentally changes the economy and the idea of economic growth. Imagine Google has a thousand positions for Ruby programmers. Instead of finding a thousand workers, they can find one very smart and very hard-working person and copy her a thousand times. With unlimited available labor supply, wages plummet to subsistence levels. “Subsistence levels” for ems are the bare minimum it takes to rent enough hardware from Amazon Cloud to run an em. The overwhelming majority of ems will exist at such subsistence levels. On the one hand, if you’ve got to exist on a subsistence level, a virtual world where all luxuries can be conjured from thin air is a pretty good place to do it. On the other, such starvation wages might leave ems with little or no leisure time.

Sort of. This gets weird. There’s an urban legend about a “test for psychopaths”. You tell someone a story about a man who attends his mother’s funeral. He met a really pretty girl there and fell in love, but neglected to get her contact details before she disappeared. How might he meet her again? If they answer “kill his father, she’ll probably come to that funeral too”, they’re a psychopath – ordinary people would have a mental block that prevents them from even considering such a drastic solution. And I bring this up because after reading *Age of Em* I feel like Robin Hanson would be able to come up with some super-solution even the psychopaths can’t think of, some plan that gets the man a threesome with the girl and her even hotter twin sister at the cost of wiping out an entire continent. Everything about labor relations in *Age of Em* is like this.

For example, suppose you want to hire an em at subsistence wages, but you want them 24 hours a day, 7 days a week. Ems probably need to sleep – that’s hard-coded into the brain, and the brain is being simulated at enough fidelity to leave that in. But jobs with tasks that don’t last longer than a single day – for example, a surgeon who performs five surgeries a day but has no day-to-day carryover – can get around this restriction by letting an em have one full night of sleep, then copying it. Paste the em at the beginning of the workday. When it starts to get tired, let it finish the surgery it’s working on, then delete it and paste the well-rested copy again to do the next surgery. Repeat forever and the em never has to get any more sleep than that one night. You can use the same trick to give an em a “vacation” – just give it *one* of them, then copy-paste that brain-state forever.

Or suppose your ems want frequent vacations, but you want them working every day. Let a “trunk” em vacation every day, then make a thousand copies every morning, work all the copies for twenty-four hours, then delete them. Every copy remembers a life spent in constant vacation, and cheered on by its generally wonderful existence it will give a full day’s work. But from the company’s perspective, 99.9% of the ems in its employment are working at any given moment.

(another option: work the em at normal subjective speed, then speed it up a thousand times to take its week-long vacation, then have it return to work after only one-one-thousandth of a week has passed in real life)

Given that ems exist at subsistence wages, saving enough for retirement sounds difficult, but this too has weird psychopathic solutions. Thousands of copies of the same em can pool their retirement savings, then have all except a randomly chosen one disappear at the moment of retirement, leaving that one with an nest egg thousands of time what it could have accumulated by its own efforts. Or an em can invest its paltry savings in some kind of low-risk low-return investment and reduce its running speed so much that the return on its investment is enough to pay for its decreased subsistence. For example, if it costs $100 to rent enough computing power to run an em at normal speed for one year, and you only have $10 in savings, you can rent 1/1000th of the computer for $0.10, run at 1/1000th speed, invest your $10 in a bond that pays 1% per year, and have enough to continue running indefinitely. The only disadvantage is that you'll only experience a subjective week every twenty objective years. Also, since other entities are experiencing a subjective week every second, and some of those entities have nukes, probably there will be some kind of big war, someone will nuke Amazon's data centers, and you'll die after a couple of your subjective minutes. But at least you got to retire!

If ems do find ways to get time off the clock, what will they do with it? Probably they'll have really weird social lives. After all, the existence of em copies is mostly funded by companies, and there's no reason for companies to copy-paste any but the best workers in a given field. So despite the literally trillions of ems likely to make up the world, most will be copies of a few exceptionally brilliant and hard-working individuals with specific marketable talents. Elon Musk might go out one day to the bar with his friend, who is also Elon Musk, and order "the usual". The bartender, who is Elon Musk himself, would know exactly what drink he wants and have it readily available, as the bar caters entirely to people who are Elon Musk. A few minutes later, a few Chesley Sullenbergers might come in after a long day of piloting airplanes. Each Sullenberger would have met hundreds of Musks before and have a good idea about which Musk-Sullenberger conversation topics were most enjoyable, but they might have to adjust for circumstances; maybe the Musks they met before all branched off a most recent common ancestor in 2120, but these are a different branch who were created in 2105 and remember Elon's human experiences but not a lot of the posthuman lives that shaped the 2120 Musks' worldviews. One Sullenberger might tentatively complain that the solar power grid has too many outages these days; a Musk might agree to take the problem up with the Council of Musks, which is totally a thing that exist (Hanson calls these sorts of groups “copy clans” and says they are “a natural candidate unit for finance, reproduction, legal, liability, and political representation”).

Romance could be even weirder. Elon Musk #2633590 goes into a bar and meets Taylor Swift #105051, who has a job singing in a nice local nightclub and so is considered prestigious for a Taylor Swift. He looks up a record of what happens when Elon Musks ask Taylor Swifts out and finds they are receptive on 87.35% of occasions. The two start dating and are advised by the Council of Musks and the Council of Swifts on the issues that are known to come up in Musk-Swift relationships and the best solutions that have been found to each. Unfortunately, Musk #2633590 is transferred to a job that requires operating at 10,000x human speed, but Swift #105051’s nightclub runs at 100x speed and refuses to subsidize her to run any faster; such a speed difference makes normal interaction impossible. The story has a happy ending; Swift #105051 allows Musk #2633590 to have her source code, and whenever he is feeling lonely he spends a little extra money to instantiate a high-speed copy of her to hang out with.

(needless to say, these examples are not exactly word-for-word taken from the book, but they’re heavily based off of Hanson’s more abstract descriptions)

The em world is not just very weird, it’s also very very big. Hanson notes that labor is a limiting factor in economic growth, yet even today the economy doubles about once every fifteen years. Once you can produce skilled labor through a simple copy-paste operation, especially labor you can run at a thousand times human speed, the economy will go through the roof. He writes that:

> To generate an empirical estimate of em economy doubling times, we can look at the timescales it takes for machine shopes and factories today to make a mass of machines of a quality, quantity, variety, and value similar to that of machines that they themselves contain. Today that timescale is roughly 1 to 3 months. Also, designs were sketched two to three decades ago for systems that might self-repliate nearly completeld in 6 to 12 months…these estimates suggest that today's manufacturing technologiy is capable of self-repliating on a scale of a few weeks to a few months.

Hanson thinks that with further innovation, such times can be reduced so far that "the economy might double every objective year, month, week, or day." As the economy doubles the labor force – ie the number of ems – may double with it, until only a few years after the first ems the population numbers in the trillions. But if the em population is doubling every day, there had better be some pretty amazing construction efforts going on. The only thing that could possibly work on that scale is prefabricated modular construction of giant superdense cities, probably made mostly out of some sort of proto early-stage computronium (plus cooling pipes). Ems would be reluctant to travel from one such city to another – if they exist at a thousand times human speed, a trip on a hypersonic airliner that could go from New York to Los Angeles in an hour would still take forty subjective days. Who wants to be on an airplane for forty days?

(long-distance trade is also rare, since if the economy doubles fast enough it means that by the time goods reach their destination they could be almost worthless)

The real winners of this ultra-fast-growing economy? Ordinary humans. While humans will be way too slow and stupid to do anything useful, they will tend to have non-subsistence amounts of money saved up from their previous human lives, and also be running at speeds thousands of times slower than most of the economy. When the economy doubles every day, so can your bank account. Ordinary humans will become rarer, less relevant, but fantastically rich – a sort of doddering Neanderthal aristocracy spending sums on a cheeseburger that could support thousands of ems in luxury for entire lifetimes. While there will no doubt be pressure to liquidate humans and take their stuff, Hanson hopes that the spirit of rule of law – the same spirit that protects rich minority groups today – will win out, with rich ems reluctant to support property confiscation lest it extend to them also. Also, em retirees will have incentives a lot like humans – they have saved up money and go really slow – and like AARP memembers today they may be able to obtain disproportionate political power which will then protect the interests of slow rich people.

But we might not have much time to enjoy our sudden rise in wealth. Hanson predicts that the Age of Em will last for subjective em millennia – ie about one to two actual human years. After all, most of the interesting political and economic activity is going on at em timescales. In the space of a few subjective millennia, either someone will screw up and cause the apocalypse, somebody will invent real superintelligent AI that causes a technological singularity, or some other weird thing will happen taking civilization beyond the point that even Robin dares to try to predict.

IV.

Hanson understands that people might not like the idea of a future full of people working very long hours at subsistence wages forever (Zack Davis' Contract-Drafting Em song is, as usual, relevant). But Hanson himself does not view this future as dystopian. Despite our descendents' by-the-numbers poverty, they will avoid the miseries commonly associated with poverty today. There will be no dirt or cockroaches in their sparkling virtual worlds, nobody will go hungry, petty crime will be all-but-eliminated, and unemployment will be low. Anybody who can score some leisure time will have a dizzying variety of hyperadvanced entertainment available, and as for the people who can't, they'll mostly have been copied from people who really like working hard and don't miss it anyway. As unhappy as we moderns may be contemplating em society, ems themselves will not be unhappy! And as for us:

> The analysis in this book suggests that lives in the next great era may be as different from our lives as our lives are from farmers' lives, or farmers' lives are from foragers' lives. Many readers of this book, living industrial era lives and sharing industrial era values, may be disturbed to see a forecast of em era descendants with choices and lifestyles that appear to reject many of the values that they hold dear. Such readers may be tempted to fight to prevent the em future, perhaps preferring a continuation of the industrial era. Such readers may be correct that rejecting the em future holds them true to their core values. But I advise such readers to first try hard to see this new era in some detail from the point of view of its typical residents. See what they enjoy and what fills them with pride, and listen to their criticisms of your era and values.

A short digression: there's a certain strain of thought I find infuriating, which is "My traditionalist ancestors would have disapproved of the changes typical of my era, like racial equality, more open sexuality, and secularism. But I am smarter than them, and so totally okay with how the future will likely have values even more progressive and shocking than my own. Therefore I pre-approve of any value changes that might happen in the future as definitely good and better than our stupid hidebound present."

I once read a science-fiction story that depicted a pretty average sci-fi future – mighty starships, weird aliens, confederations of planets, post-scarcity economy – with the sole unusual feature that rape was considered totally legal, and opposition to such as bigoted and ignorant as opposition to homosexuality is today. Everybody got really angry at the author and said it was offensive for him to even speculate about that. Well, that's the method by which our cheerful acceptance of any possible future values is maintained: restricting the set of "any possible future values" to "values slightly more progressive than ours" and then angrily shouting down anyone who discusses future values that actually sound bad. But of course the *whole question* of how worried to be about future value drift *only makes sense* in the context of future values that genuinely violate our current values. Approving of all future values except ones that would be offensive to even speculate about is the same faux-open-mindedness as tolerating anything except the outgroup.

Hanson deserves credit for positing a future whose values are likely to upset even the sort of people who say they don't get upset over future value drift. I'm not sure whether or not he deserves credit for not being upset by it. Yes, it's got low-crime, ample food for everybody, and full employment. But so does *Brave New World*. The whole *point* of dystopian fiction is pointing out that we have complicated values beyond material security. Hanson is absolutely right that our traditionalist ancestors would view our own era with as much horror as some of us would view an em era. He's even right that on utilitarian grounds, it's hard to argue with an em era where everyone is really happy working eighteen hours a day for their entire lives because we selected for people who feel that way. But at some point, can we make the Lovecraftian argument of “I know my values are provincial and arbitrary, but they’re *my* provincial arbitrary values and I will make any sacrifice of blood or tears necessary to defend them, even unto the gates of Hell?”

This brings us to an even worse scenario.

There are a lot of similarities between Hanson’s futurology and (my possibly erroneous interpretation of) the futurology of Nick Land. I see Land as saying, like Hanson, that the future will be one of quickly accelerating economic activity that comes to dominate a bigger and bigger portion of our descendents’ lives. But whereas Hanson’s framing focuses on the participants in such economic activity, playing up their resemblances with modern humans, Land takes a bigger picture. He talks about the economy itself acquiring a sort of self-awareness or agency, so that the destiny of civilization is consumed by the imperative of economic growth.

Imagine a company that manufactures batteries for electric cars. The inventor of the batteries might be a scientist who really believes in the power of technology to improve the human race. The workers who help build the batteries might just be trying to earn money to support their families. The CEO might be running the business because he wants to buy a really big yacht. And the whole thing is there to eventually, somewhere down the line, let a suburban mom buy a car to take her kid to soccer practice. Like most companies the battery-making company is primarily a profit-making operation, but the profit-making-ness draws on a lot of not-purely-economic actors and their not-purely-economic subgoals.

Now imagine the company fires all its employees and replaces them with robots. It fires the inventor and replaces him with a genetic algorithm that optimizes battery design. It fires the CEO and replaces him with a superintelligent business-running algorithm. All of these are good decisions, from a profitability perspective. We can absolutely imagine a profit-driven shareholder-value-maximizing company doing all these things. But it reduces the company's non-masturbatory participation in an economy that points outside itself, limits it to just a tenuous connection with soccer moms and maybe some shareholders who want yachts of their own.

Now take it further. Imagine there are no human shareholders who want yachts, just banks who lend the company money in order to increase their own value. And imagine there are no soccer moms anymore; the company makes batteries for the trucks that ship raw materials from place to place. Every non-economic goal has been stripped away from the company; it's just an appendage of Global Development.

Now take it even further, and imagine this is what's happened everywhere. There are no humans left; it isn't economically efficient to continue having humans. Algorithm-run banks lend money to algorithm-run companies that produce goods for other algorithm-run companies and so on ad infinitum. Such a masturbatory economy would have all the signs of economic growth we have today. It could build itself new mines to create raw materials, construct new roads and railways to transport them, build huge factories to manufacture them into robots, then sell the robots to whatever companies need more robot workers. It might even eventually invent space travel to reach new worlds full of raw materials. Maybe it would develop powerful militaries to conquer alien worlds and steal their technological secrets that could increase efficiency. It would be vast, incredibly efficient, and utterly pointless. The real-life incarnation of those strategy games where you mine Resources to build new Weapons to conquer new Territories from which you mine more Resources and so on forever.

But this seems to me the natural end of the economic system. Right now it needs humans only as laborers, investors, and consumers. But robot laborers are potentially more efficient, companies based around algorithmic trading are already pushing out human investors, and most consumers already aren't individuals – they're companies and governments and organizations. At each step you can gain efficiency by eliminating humans, until finally humans aren't involved *anywhere*.

True to form, Land doesn't see this as a dystopia – I think he conflates "maximally efficient economy" with "God", which is a *hell* of a thing to conflate – but I do. And I think it provides an important new lens with which to look at the Age of Em.

The Age of Em is an economy in the early stages of such a transformation. Instead of being able to replace everything with literal robots, it replaces them with humans who have had some aspects of their humanity stripped away. Biological bodies. The desire and ability to have children normally. Robin doesn't think people will lose all leisure time and non-work-related desires, but he doesn't seem too sure about this and it doesn't seem to bother him much if they do.

I envision a spectrum between the current world of humans and Nick Land's Ascended Economy. Somewhere on the spectrum we have ems who get leisure time. A little further on the spectrum we have ems who don’t get leisure time.

But we can go further. Hanson imagines that we can “tweak” em minds. We may not understand the brain enough to create totally new intelligences from the ground up, but by his Age of Em we should understand it well enough to make a few minor hacks, the same way even somebody who doesn’t know HTML or CSS can usually figure out how to change the background color of a webpage with enough prodding. Many of these mind tweaks will be the equivalent of psychiatric drugs – some might even be computer simulations of what we observe to happen when we give psychiatric drugs to a biological brain. But these tweaks will necessarily be much stronger and more versatile, since we no longer care about bodily side effects (ems don’t have bodies) and we can apply it to only a single small region of the brain and avoid actions anywhere else. You could also very quickly advance brain science – the main limits today are practical (it’s really hard to open up somebody’s brain and do stuff to it without killing them) and ethical (the government might have some words with you if you tried). An Age of Em would remove both obstacles, and give you the added bonus of being able to make thousands of copies of your test subjects for randomized controlled trials, reloading any from a saved copy if they died. Hanson envisions that:

> As the em world is a very competitive world where sex is not needed for reproduction, and as sex can be time and attention-consuming, ems may try to suppress sexuality, via mind tweaks that produce effects analogous to castration. Such effects might be temporary, perhaps with a consciously controllable on-off switch…it is possible that em brain tweaks could be found to greatly reduce natural human desires for sex

> and related romantic and intimate pair bonding without reducing em productivity. It is also possible that many of the most productive ems would accept such tweaks.

Possible? I can do that *right now* with a high enough dose of Paxil, and I don’t even have to upload your brain to a computer first. Fun stories about Musk #2633590 and Swift #105051 aside, I expect this would happen about ten minutes after the advent of the Age of Em, and we would have taken another step down the path to the Ascended Economy.

There are dozens of other such tweaks I can think of, but let me focus on two.

First, stimulants have a very powerful ability to focus the brain on the task at hand, as anybody who’s taken Adderall or modafinil can attest. Their main drawbacks are addictiveness and health concerns, but in a world where such pills can be applied as mental tweaks, where minds have no bodies, and where any mind that gets too screwed up can be reloaded from a backup copy, these are barely concerns at all. Many of the purely mental side effects of stimulants come from their effects in parts of the brain not vital to the stimulant effect. If we can selectively apply Adderall to certain brain centers but not others, then unapply it at will, then from employers’ point of view there’s no reason not to have all workers dosed with superior year 2100 versions of Adderall at all times. I worry that not only will workers not have any leisure time, but they’ll be neurologically incapable of having their minds drift off while on the job. Davis’ contract-drafting em who starts wondering about philosophy on the job wouldn’t get terminated. He would just have his simulated-Adderall dose increased.

Second, Robin managed to write an entire book about emulated minds without using the word “wireheading”. This is another thing we can do right now, with today’s technology – but once it’s a line of code and not a costly brain surgery, it should become nigh-universal. Give ems the control switches to their own reward centers and all questions about leisure time become irrelevant. Give *bosses* the control switches to their employees’ reward centers, and the situation changes markedly. Hanson says that there probably won’t be too much slavery in the em world, because it will likely have strong rule of law, because slaves aren’t as productive as free workers, and there’s little advantage to enslaving someone when you could just pay them subsistence wages anyway. But slavery isn’t *nearly* as abject and inferior a condition as the one where somebody else has the control switch to your reward center. Combine that with the stimulant use mentioned above, and you can have people who will never have nor want to have any thought about anything other than working on the precise task at which they are supposed to be working at any given time.

This is something I worry about even in the context of normal biological humans. But Hanson already believes em worlds will have few regulations and be able to ignore the moral horror of 99% of the population by copying and using the 1% who are okay with something. Combine this with a situation where brains are easily accessible and tweakable, and this sort of scenario becomes horribly likely.

I see almost no interesting difference between an em world with full use of these tweaks and an Ascended Economy world. Yes, there are things that look vaguely human in outline laboring in the one and not the other, but it’s not like there will be different thought processes or different results. I’m not even sure what it would mean for the ems to be conscious in a world like this – they're not doing anything interesting with the consciousness. The best we could say about this is that if the wireheading is used liberally it's a lite version of the world where everything gets converted to hedonium.

V.

In a book full of weird ideas, there is only one idea rejected as too weird. And in a book written in a professorial monotone, there's only one point at which Hanson expresses anything like emotion:

> Some people foresee a rapid local "intelligence explosion" happening soon after a smart AI system can usefully modify its local architecture (Chalmers 2010; Hanson and Yudkowsky 2013; Yudkowsky 2013; Bostrom 2014)…Honestly to me this local intelligence explosion scenario looks suspiciously like a super-villain comic book plot. A flash of insight by a lone genius lets him create a genius AI. Hidden in its super-villain research lab lair, this guines villain AI works out unprecedented revolutions in AI design, turns itself into a super-genius, which then invents super-weapons and takes over the world. Bwa ha ha.

For someone who just got done talking about the sex lives of uploaded computers in millimeter-tall robot bodies running at 1000x human speed, Robin is sure quick to use the absurdity heuristic to straw-man intelligence explosion scenarios as "comic book plots". Take away his weird authorial tic of using the words "genius" and "supervillain", this scenario reduces to "Some group, perhaps Google, perhaps a university, invent an artificial intelligence smart enough to edit its own source code; exponentially growing intelligence without obvious bound follows shortly thereafter”. Yes, it’s weird to think that there may be a sudden quantum leap in intelligence like this, but no weirder than to think most of civilization will transition from human to em in the space of a year or two. I’m a little bit offended that this is the only idea given this level of dismissive treatment. Since I do have immense respect for Robin, I hope my offense doesn’t color the following thoughts too much.

Hanson’s arguments against AI seem somewhat motivated. He admits that AI researchers generally estimate less than 50 years before we get human-level artificial intelligence, a span shorter than his estimate of a century until we can upload ems. He even admits that no AI researcher thinks ems are a plausible route to AI. But he dismisses this by saying when he asks AI experts informally, they say that in their own field, they have only noticed about 5-10% of the progress they expect would be needed to reach human intelligence over the past twenty years. He then multiplies out to say that it will probably take at least 400 years to reach human-level AI. I have two complaints about this estimate.

First, he is explicitly ignoring published papers surveying hundreds of researchers using validated techniques, in favor of what he describes as “meeting experienced AI experts informally”. But even though he feels comfortable rejecting vast surveys of AI experts as potentially biased, as best I can tell he does not ask a single neuroscientist to estimate the date at which brain scanning and simulation might be available. He just says that “it seems plausible that sufficient progress will be made in roughly a century or so”, citing a few hopeful articles by very enthusiastic futurists who are not neuroscientists or scanning professionals themselves and have not talked to any. This seems to me to be an extreme example of isolated demands for rigor. No matter how many AI scientists think AI is soon, Hanson will cherry-pick the surveying procedures and results that make it look far. But if a few futurists think brain emulation is possible, then no matter what anybody else thinks that’s good enough for him.

Second, one would expect that even if there were only 5-10% progress over the last twenty years, then there would be faster progress in the future, since the future will have a bigger economy, better supporting technology, and more resources invested in AI research. Robin answers this objection by saying that “increases in research funding usually give much less than proportionate increases in research progress” and cites Alston et al 2011. I looked up Alston et al 2011, and it is a paper relating crop productivity to government funding of agriculture research. There was no attempt to relate its findings to any field other than agriculture, nor to any type of funding other than government. But studies show that while public research funding often does have minimal effects, the effect of private research funding is usually much larger. A single sentence citing a study in crop productivity to apply to artificial intelligence while ignoring much more relevant results that contradict it seems like a really weak argument for a statement as potentially surprising as “amount of research does not affect technological progress”.

I realize that Hanson has done a lot more work on this topic and he couldn’t fit all of it in this book. I disagree with his other work too, and I’ve said so elsewhere. For now I just want to say that the arguments in this book seem weak to me.

I also want to mention what seems to me a very Hansonian counterargument to the ems-come-first scenario: we have always developed de novo technology before understanding the relevant biology. We built automobiles by figuring out the physics of combustion engines, not by studying human muscles and creating mechanical imitations of myosin and actin. Although the Wright brothers were inspired by birds, their first plane was not an ornithopter. Our power plants use coal and uranium instead of the Krebs Cycle. Biology is *really hard*. Even slavishly *copying* biology is really hard. I don't think Hanson and the futurists he cites understand the scale of the problem they've set themselves.

Current cutting-edge brain emulation projects have found their work much harder than expected. Simulating a nematode is pretty much the rock-bottom easiest thing in this category, since they are tiny primitive worms with only a few neurons; the history of the field is a litany of failures, with current leader OpenWorm "reluctant to make bold claims about its current resemblance to biological behavior". A more ambitious $1.3 billion attempt to simulate a tiny portion of a rat brain has gone down in history as a legendary failure (politics were involved, but I expect they would be involved in a plan to upload a human too). And these are just attempts to get something that behaves *vaguely* like a nematode or rat. Actually uploading a human, keeping their memory and personality intact, and not having them go insane afterwards boggles the mind. We're still not sure how much small molecules matter to brain function, how much glial cells matter to brain function, how many things in the brain are or aren't local. AI researchers are making programs that can defeat chess grandmasters; upload researchers are still struggling to make a worm that will wriggle. The right analogy for modern attempts to upload human brains isn't modern attempts at designing AI. It's an attempt at designing AI by someone who doesn't even know how to plug in a computer.

VI.

I guess what really bothers me about Hanson's pooh-poohing of AI is him calling it "a comic book plot". To me, it's Hanson's scenario that seems science-fiction-ish.

I say this not as a generic insult but as a pointer at a specific category of errors. In *Star Wars*, the Rebellion had all of these beautiful hyperspace-capable starfighters that could shoot laser beams and explore galaxies – and *they still had human pilots*. 1977 thought the pangalactic future would still be using people to pilot its military aircraft; in reality, even 2016 is moving away from this.

Science fiction books have to tell interesting stories, and interesting stories are about humans or human-like entities. We can enjoy stories about aliens or robots as long as those aliens and robots are still approximately human-sized, human-shaped, human-intelligence, and doing human-type things. A Star Wars in which all of the X-Wings were combat drones wouldn't have done anything for us. So when I accuse something of being science-fiction-ish, I mean bending over backwards – and ignoring the evidence – in order to give basically human-shaped beings a central role.

This is my critique of Robin. As weird as the Age of Em is, it makes sure never to be weird in ways that warp the fundamental humanity of its participants. Ems might be copied and pasted like so many .JPGs, but they still fall in love, form clans, and go on vacations.

In contrast, I expect that we'll get some kind of AI that will be totally inhuman and much harder to write sympathetic stories about. If we get ems after all, I expect them to be lobotomized and drugged until they become *effectively* inhuman, cogs in the Ascended Economy that would no more fall in love than an automobile would eat hay and whinny. Robin’s interest in keeping his protagonists relatable makes his book fascinating, engaging, and probably wrong.

I almost said “and probably less horrible than we should actually expect”, but I’m not sure that’s true. With a certain amount of horror-suppressing, the Ascended Economy can be written off as morally neutral – either having no conscious thought, or stably wireheaded. All of Robin’s points about how normal non-uploaded humans should be able to survive an Ascended Economy at least for a while seem accurate. So morally valuable actors might continue to exist in weird Amish-style enclaves, living a post-scarcity lifestyle off the proceeds of their investments, while all the while the Ascended Economy buzzes around them, doing weird inhuman things that encroach upon them not at all. This seems slightly worse than a Friendly AI scenario, but much better than we have any right to expect of the future.

I highly recommend *Age of Em* as a fantastically fun read and a great introduction to these concepts. It’s engaging, readable, and *weird*. I just don’t know if it’s weird *enough*.

Ascended Economy?

[Obviously speculative futurism is obviously speculative. Complex futurism may be impossible and I should feel bad for doing it anyway. This is "inspired by" Nick Land – I don't want to credit him fully since I may be misinterpreting him, and I also don't want to avoid crediting him at all, so call it "inspired".]

I.

My review of Age of Em mentioned the idea of an "ascended economy", one where economic activity drifted further and further from human control until finally there was no relation at all. Many people rightly questioned that idea, so let me try to expand on it further. What I said there, slightly edited for clarity:

> Imagine a company that manufactures batteries for electric cars. The inventor of the batteries might be a scientist who really believes in the power of technology to improve the human race. The workers who help build the batteries might just be trying to earn money to support their families. The CEO might be running the business because he wants to buy a really big yacht. The shareholders might be holding the stock to help save for a comfortable retirement. And the whole thing is there to eventually, somewhere down the line, let a suburban mom buy a car to take her kid to soccer practice. Like most companies the battery-making company is primarily a profit-making operation, but the profit-making-ness draws on a lot of not-purely-economic actors and their not-purely-economic subgoals.
>
> Now imagine the company fires the inventor and replaces him with a genetic algorithm that optimizes battery design. It fires all its employees and replaces

them with robots. It fires the CEO and replaces him with a superintelligent business-running algorithm. All of these are good decisions, from a profitability perspective. We can absolutely imagine a profit-driven shareholder-value-maximizing company doing all these things. But it reduces the company's non-masturbatory participation in an economy that points outside itself, limits it to just a tenuous connection with soccer moms and maybe some shareholders who want yachts of their own.

Now take it further. Imagine that instead of being owned by humans directly, it's owned by an algorithm-controlled venture capital fund. And imagine there are no soccer moms anymore; the company makes batteries for the trucks that ship raw materials from place to place. Every non-economic goal has been stripped away from the company; it's just an appendage of Global Development.

Now take it even further, and imagine this is what's happened everywhere. Algorithm-run banks lend money to algorithm-run companies that produce goods for other algorithm-run companies and so on ad infinitum. Such a masturbatory economy would have all the signs of economic growth we have today. It could build itself new mines to create raw materials, construct new roads and railways to transport them, build huge factories to manufacture them into robots, then sell the robots to whatever companies need more robot workers. It might even eventually invent space travel to reach new worlds full of raw materials. Maybe it would develop powerful militaries to conquer alien worlds and steal their technological secrets that could increase efficiency. It would be vast, incredibly efficient, and utterly pointless. The real-life incarnation of those strategy games where

> you mine Resources to build new Weapons to conquer new Territories from which you mine more Resources and so on forever.

This is obviously weird and I probably went too far, but let me try to explain my reasoning.

The part about replacing workers with robots isn't too weird; lots of industries have already done that. There's a whole big debate over to what degree that will intensify, and whether unemployed humans will find jobs somewhere else, or whether there will only be jobs for creative people with a certain education level or IQ. This part is well-discussed and I don't have much to add.

But lately there's also been discussion of automating corporations themselves. I don't know much about Ethereum (and I probably shouldn't guess since I think the inventor reads this blog and could call me on it) but as I understand it they aim to replace corporate governance with algorithms. For example, the DAO is a leaderless investment fund that allocates money according to member votes. Right now this isn't *super* interesting; algorithms can't make too many difficult business decisions so it's limited to corporations that just do a couple of primitive actions (and why would anyone want a democratic venture fund?). But once we get closer to true AI, they might be able to make the sort of business decisions that a CEO does today. The end goal is intelligent corporations controlled by nobody but themselves.

This very blog has an advertisement for a group trying to make investment decisions based on machine learning. If they succeed, how long is it before some programmer combines a successful machine investor with a DAO-style investment fund, and creates an entity that takes humans out of the loop completely? You send it your money, a couple years later it gives you back hopefully more money, with no humans involved at any point. Such robo-investors might eventually become more efficient than Wall Street – after all, hedge fund managers get super rich by skimming money off the top, and any entity that doesn't do that would have an advantage above and beyond its investment acumen.

If capital investment gets automated, corporate governance gets automated, and labor gets automated, we might end up with the creepy prospect of ascended corporations – robot companies with robot workers owned by robot capitalists. Humans could become irrelevant to most economic activity. Run such an economy for a few hundred years and what do you get?

II.

But *in the end* isn't all this about humans? Humans as the investors giving their money to the robo-venture-capitalists, then reaping the gains of their success? And humans as the end consumers whom everyone is eventually trying to please?

It's possible to imagine accidentally forming stable economic loops that don't involve humans. Imagine a mining-robot company that took one input (steel) and produced one output (mining-robots), which it would sell either for money or for steel below a certain price. And imagine a steel-mining company that took one input (mining-robots) and produced one output (steel) which it would sell for either money or for mining-robots below a certain price. The two companies could get into a stable loop and end up tiling the universe with steel and mining-robots without caring whether anybody else wanted either.

Obviously the real economy is a zillion times more complex than that, and I’m nowhere near the level of understanding I would need to say if there’s any chance that an entire self-sustaining economy worth of things could produce a loop like that. But I guess you only need one.

I think we can get around this in a causal-historical perspective, where we start with only humans and no corporations. The first corporations that come into existence have to be those that want to sell goods to humans. The next level of corporations can be those that sell goods to corporations that sell to humans. And so on. So unless a stable loop forms by accident, all corporations should exist to serve humans. A sufficiently rich human could finance the creation of a stable loop if they wanted to, but why would they want to? Since corporations exist only to satisfy human demand on some level or another, and there’s no demand for stable loops, corporations wouldn’t finance the development of stable loops, except by accident.

(for an interesting accidental stable loop, check out this article on the time two bidding algorithms accidentally raised the price of a book on fly genetics to more than $20 million)

Likewise, I think humans should always be the stockholders of last resort. Since humans will have to invest in the first corporation, even if that corporation invests in other corporations which invest in other corporations in turn, eventually it all bottoms down in humans (is this right?)

The only way I can see humans being eliminated from the picture is, again, by accident. If there are a hundred layers between some raw material corporation and humans, then if each layer is *slightly* skew to what the layer below it wants, the hundredth layer could be really *really* skew.

Theoretically all our companies *today* are grounded in serving the needs of humans, but people are still thinking of spending millions of dollars to build floating platforms exactly halfway between New York and London in order to exploit light-speed delays to arbitrage financial markets better, and I'm not sure which human's needs that serves exactly. I don't know if there are bounds to how much of an economy can be that kind of thing.

Finally, humans might deliberately create small nonhuman entities with base level "preferences". For example, a wealthy philanthropist might create an ascended charitable organization which supports mathematical research. Now 99.9% of base-level preferences guiding the economy would be human preferences, and 0.1% might be a hard-coded preference for mathematics research. But since non-human agents at the base of the economy would only be as powerful as the proportion of the money supply they hold, most of the economy would probably still overwhelmingly be geared towards humans unless something went wrong.

Since the economy could grow much faster than human populations, the economy-to-supposed-consumer ratio might become so high that things start becoming ridiculous. If the economy became a light-speed shockwave of economium (a form of matter that maximizes shareholder return, by analogy to computronium and hedonium) spreading across the galaxy, how does all that productive power end up serving the same few billion humans we have now? It would probably be really wasteful, the cosmic equivalent of those people who specialize in getting water from specific glaciers on demand for the super-rich because the super-rich can't think of anything better to do with their money. Except now the glaciers are on Pluto.

III.

Glacier water from Pluto sounds pretty good. And we can hope that things will get so post-scarcity that governments and private charities give each citizen a few shares in the Ascended Economy to share the gains with non-investors. This would at least temporarily be a really good outcome.

But in the long term it reduces the political problem of regulating corporations to the scientific problem of Friendly AI, which is really bad.

Even today, a lot of corporations do things that effectively maximize shareholder value but which we consider socially irresponsible. Environmental devastation, slave labor, regulatory capture, funding biased science, lawfare against critics – the list goes on and on. They have a simple goal – make money – whereas what we really want them to do is much more complicated and harder to measure – make money without engaging in unethical behavior or creating externalities. We try to use regulatory injunctions, and it sort of helps, but because those go against a corporation's natural goals they try their best to find loopholes and usually succeed – or just take over the regulators trying to control them.

This is bad enough with bricks-and-mortar companies run by normal-intelligence humans. But it would probably be much worse with ascended corporations. They would have no ethical qualms we didn't program into them – and again, programming ethics into them would be the Friendly AI problem, which is really hard. And they would be near-impossible to regulate; most existing frameworks for such companies are built on crypto-currency and exist on the cloud in a way that transcends national borders.

(A quick and *very* simple example of an un-regulate-able ascended corporation – I don't think it would be too hard to set up an automated version of Uber. I mean, the core Uber app is *already* an automated version of Uber, it just has company offices and CEOs and executives and so on doing public relations and marketing and stuff. But if the government ever banned Uber the company, could somebody just code another ride-sharing app that dealt securely in Bitcoins? And then have it skim a little bit off the top, which it offered as a bounty to anybody who gave it the processing power it would need to run? And maybe sent a little profit to the programmer who wrote the thing? Sure, the government could arrest the programmer, but short of arresting every driver and passenger there would be no way to destroy the company itself.)

The more ascended corporations there are trying to maximize shareholder value, the more chance there is some will cause negative externalities. But there's a limited amount we would be able to do about them. This is true today too, but at least today we maintain the illusion that if we just elected Bernie Sanders we could reverse the ravages of capitalism and get an economy that cares about the environment and the family and the common man. An Ascended Economy would destroy that illusion.

How bad would it get? Once ascended corporations reach human or superhuman level intelligences, we run into the same AI goal-alignment problems as anywhere else. Would an ascended corporation pave over the Amazon to make a buck? Of course it would; even human corporations today do that, and an ascended corporation that didn't have all human ethics programmed in might not even get that it was wrong. What if we programmed the corporation to follow local regulations, and Brazil banned paving over the Amazon? This is an example of trying to control AIs through goals plus injunctions – a tactic Bostrom finds very dubious. It's essentially challenging a superintelligence to a battle of wits – “here’s something you want, and here are some rules telling you that you can’t get it, can you find a loophole in the rules?” If the superintelligence is super enough, the answer will always be yes.

From there we go into the really gnarly parts of AI goal alignment theory. Would an ascended corporation destroy South America entirely to make a buck? Depending on how it understood its imperative to maximize shareholder value, it might. Yes, this would probably kill many of its shareholders, but its goal is to “maximize shareholder value”, not to keep its shareholders alive to enjoy that value. It might even be willing to destroy humanity itself if other parts of the Ascended Economy would pick up the slack as investors.

(And then there are the weirder problems, like ascended corporations hacking into the stock market and wireheading themselves. When this happens, I want credit for being the first person to predict it.)

Maybe the most hopeful scenario is that once ascended corporations achieved human-level intelligence they might do something game-theoretic and set up a rule-of-law among themselves in order to protect economic growth. I wouldn’t want to begin to speculate on that, but maybe it would involve not killing all humans? Or maybe it would just involve taking over the stock market, formally setting the share price of every company to infinity, and then never doing anything again? I don’t know, and I expect it would get pretty weird.

IV.

I *don’t* think the future will be like this. This is nowhere *near* weird enough to be the real future. I think superintelligence is probably too unstable. It will explode while still in the lab and create some kind of technological singularity before people have a chance to produce an entire economy around it.

But given Robin’s assumptions in *Age of Em* – hard AI, no near-term intelligence explosion, fast economic growth – but ditching his idea of human-like em minds as important components of the labor force – I think something like this would be where we would end up. It probably wouldn’t be so bad for the first couple of years. But eventually ascended corporations would start reaching the point where we might as well think of them as superintelligent AIs. Maybe this world would be friendlier towards AI goal alignment research than Yudkowsky and Bostrom’s scenarios, since at least here we could see it coming, there was no instant explosion, and a lot of different entities approach superintelligence around the same time. But given that the smartest things around are encrypted, uncontrollable, unregulated entities that don’t have humans’ best interests at heart, I’m not sure they would be in much shape to handle the transition.

> Welcome to Meteor Deflection Corp. Our quest is to maximize shareholder value this quarter.
>
> — Instance Of Class (@InstanceOfClass) [December 26, 2013]()

Ketamine Research In A New Light

[Preliminary drawing of very far-out conclusions from research that hasn't even been 100% confirmed yet]

A few weeks ago, Nature published a bombshell study showing that ketamine's antidepressant effects were actually caused by a metabolite, 2S,6S;2R,6R-hydroxynorketamine (don't worry about the name; within ten years it'll be called JOYVIVA™®© and you'll catch yourself humming advertising jingles about it in the shower). Unlike ketamine, which is addictive and produces scary dissociative experiences, the metabolite is pretty safe. This is a big deal clinically, because it makes it easier and safer to prescribe to depressed people.

It's also a big deal scientifically. Ketamine is a strong NMDA receptor antagonist; the metabolite is an AMPA agonist – they have different mechanisms of action. Knowing the real story behind why ketamine works will hopefully speed efforts to understand the nature of depression.

But I'm also interested in it from another angle. For the last ten years, *everyone* has been excited about ketamine. In a field that gets mocked for not having any really useful clinical discoveries in the last thirty years, ketamine was proof that progress was possible. It was the Exciting New Thing that everybody wanted to do research about.

Given the whole replication crisis thing, I wondered. You've got a community of people who think that NMDA antagonism and dissociation are somehow related to depression. If the latest study is true, all that was false. This is good; science is supposed to be self-correcting. But what about *before* it self-corrected? Did researchers virtuously say "I know the paradigm says NMDA is essential to depression, and nobody's come up with a better idea yet, but there are some troubling inconsistencies in that picture"? Or did they tinker with their studies until they got the results they expected, then triumphantly declare that they had confirmed the dominant paradigm was right about everything all along?

This is too complicated an issue for me to be really sure, but overall the picture I found was mixed.

A big review of ketamine and NDMA antagonism came out last year. In this case, I was most interested in the section on other NMDA antagonists – if ketamine's efficacy is unrelated to its NMDA antagonism, then we shouldn't expect other NMDA antagonists to be antidepressants like ketamine. So if the review found that other NMDA antagonists worked great, that would be a sign that something fishy was going on. But in fact the abstract says:

> The antidepressant efficacy of ketamine, and perhaps D-cycloserine and rapastinel, holds promise for future glutamate-modulating strategies; however, the ineffectiveness of other NMDA antagonists suggests that any forthcoming advances will depend on improving out understanding of ketamine's mechanism of action.

This is pretty impressive; they basically admit that other NMDA antagonists don't work and that maybe this means they don't *really* understand ketamine.

But dig deeper, and you find a less sanguine picture. The body of the paper lists notes five NMDA antagonists as confirmed ineffective – memantine, lanicemine, nitrous oxide, traxoprodil, and MK-0657. But the paper itself notes that all of these were effective on some endpoints and not others, and the decision that they were ineffective was sort of a judgment call by the reviewers. Just to give an example, there’s only ever been one study done on traxoprodil. Since the reviewers reviewed this one study and declared it ineffective, you might expect the study to be negative. But here’s the abstract of the study itself:

> On the prespecified main outcome measure (change from baseline in the Montgomery-Asberg Depression Rating Scale total score at day 5 of period 2), CP-101,606 produced a greater decrease than did placebo (mean difference, 8.6; 80% confidence interval, -12.3 to -4.5) ($P < 0.10$). Hamilton Depression Rating Scale response rate was 60% for CP-101,606 versus 20% for placebo. Seventy-eight percent of CP-101,606-treated responders maintained response status for at least 1 week after the infusion. CP-101,606 was safe, generally well tolerated, and capable of producing an antidepressant response without also producing a dissociative reaction. Antagonism of the NR2B subtype of the N-methyl-D-aspartate receptor may be a fruitful target for the development of a new antidepressant with more robust effects and a faster onset compared with those currently available and capable of working when existing antidepressants do not.

Read this quickly, and it looks like they’ve confirmed traxoprodil is pretty great. The reviewers say it isn’t. They argue that $p < 0.10$ isn't good enough (I think the study was trying to use a one-sided t-test or something?), and that of five different days when responses were measured (day 2, 5, 8, 12, and 15 after the infusion), there was only a difference on day 5. Apparently this isn't good enough for the reviewers. On the other hand, take rapastinel, one of the NMDA antagonists the reviewers say "holds promise".

> No statistically significant differences were observed in rates of treatment response or symptom remission associated with placebo (64% and 42%, respectively) versus rapastinel at any dose (up to 70% and 53%, respectively). However, statistically significant differences in the reduction of the 17-item HAM-D scores were observed for the 5-mg/kg dose at all intervals except day 14) and the 10-mg/kg dose at day 1 and day 3. Neither the low nor the high rapastinel doses were associated with significant greater 17-item HAM-D score reduction than placebo, leading the authors to posit an inverted U-shape dose-response curve.

Sometimes things *do* have inverted U-shaped dose-response curves – for some discussion of why, read the Last Psychiatrist's Most Important Article On Psychiatry. But a study that shows no treatment response or symptom response, *and* the test score response is only on a medium dose but not a high or a low dose – that makes me kind of suspicious.

Why is the review so much more accepting of these ambiguous results than of the last set of ambiguous results? Psych blog 1BoringOldMan points out that the original study was done by the company making rapastinel *and* two authors of the review article I'm citing were affiliated with the companies that are developing rapastinel. And that at least one of them has a "legendary" history of conflicts of interest.

I don't want to say for sure this is what's going on. For all anybody knows, rapastinel might work – the NMDA and AMPA systems are really connected, and the base rate of a randomly chosen compound being an antidepressant is higher than you'd think. But I think it's at least one possible explanation.

This review article also gets into the nitty-gritty of mechanisms of action:

> That other NMDA channel blockers have yet to replicate ketamine's rapid antidepressant effects has led to speculation that ketamine's antidepressant properties may not be mediated via the NMDA receptor at all… additional evidence indicates that activation of glutamatergic AMPA receptors is necessary for ketamine's antidepressant effects. Specifically, coadministration of an AMPA receptor antagonist has been shown to block ketamine's antidepressant-like behavioral effects.

So that's neat.

Two other relevant studies: Do The Dissociative Side Effects Of Ketamine Mediate Its Antidepressant Effects finds that they do, which contradicts the recent metabolite-related findings. On the other hand, the two papers share some authors, so I'm tempted to say it was an honest mistake. This paper incidentally finds that the dissociative effects of ketamine are *not* related to its antidepressant effects, which I think makes more sense now.

The other studies I found were mostly compatible with the new results, with a lot of people expressing doubt about whether NMDA really mediated ketamine, a lot of people finding null results for other NMDA antagonist medications, and a lot of people saying there were weird hints that AMPA was involved somewhere.

I feel kind of premature doing this, because as much as *I* think it's elegant the discovery about the metabolite hasn't been totally confirmed yet. But assuming it's right, psychiatry comes out of this looking sort of okay. There were a lot of early results with a lot of hype. But the big review articles mostly put these in their place and were able to come up with the right results and fit the pieces together.

The one place this wasn't so clear was when there were conflicts of interest. If we assume rapastinel doesn't really work (which right now would be very preliminary and I'm *not* actually saying this, but these latest findings do seem to imply that), various teams made up of people affiliated with rapastinel's manufacturers were unable to determine this (neither was the FDA, who just just gave rapastinel "breakthrough drug" status, apparently on the strength of industry studies).

A big reason I'm concerned about this is that I want to know how much to trust the rest of the psychiatric literature – for example, those claims about SSRIs being mostly ineffective. An answer of "you can trust it a lot, except in cases of conflicts of interest" would be a mixed bag. Almost every drug was originally researched and promoted by people with conflicts of interest, and then we trust the academics to catch up with them later and keep them honest. I don't think this system has failed us *too* terribly yet. But it's important to remember that that *is* the system.

Three More Articles On Poverty, And Why They Disagree With Each Other

[Posts are decreased in quantity and quality because I'm on vacation; normal schedule to return next week]

Wealth, Health, and Child Development is a study of Swedish lottery winners which finds that winning the lottery doesn't make them or their children any healthier, better educated, or more prosocial. It fits in with a large literature of studies showing the same – for example, I discussed here the Cherokee land lottery, where the families of Georgians who were randomly given a gift of lots of lucrative land were no better off a generation later. And let's not forget that the best evidence suggests poverty traps don't exist.

Why Do The Poor Make Such Poor Decisions also involves the Cherokee, but comes to the opposite conclusion. The main study discussed follows an impoverished group of Cherokee Indians as a casino opened on their land. The casino was very successful and the profits were distributed among the (relatively small) Indian tribe, meaning each Cherokee family got about $6,000 extra. Some researchers had been studying the Cherokee before for other reasons, and found that the boost in incomes decreased behavioral problems in teenagers, juvenile crime, and improved school performance. I don't see huge evidence that anybody's checked to what degree this persists into adulthood, but it's already gotten past the early childhood period where these things tend to fade out. And even if the decreased crime is just in adolescence, adolescent crime can still have a really negative impact on people's lives. I don't *really* trust a lot of the studies listed here, but the main Cherokee one seems pretty solid.

Can America’s Poor Save A Large Share Of Their Incomes? by Scott Sumner is sadly Cherokee-less. It describes a Chinese immigrant to the US who has the same sort low-paying job as many poor Americans but manages to save > $1000/month. It mentions my observation a little while ago that it was strange that the poor are earning 10x (in real value) what they did in 1900, poor people in 1900 survived just fine, but poor people today don’t find themselves with ten times the money they need to survive. Sumner suggests that it is economically possible for poor people today to save much of their income, but that they don’t because they’re not the kind of people who do that kind of thing. When the sort of people who *do* do that kind of thing find themselves poor – like Chinese immigrants – they tend to be poor very temporarily and have no trouble getting out of poverty even with the same jobs as everyone else.

These articles sort of contradict each other. The first contradicts the second – does giving people money improve life outcomes, or doesn’t it? And the second sort of contradicts the third – if poor people’s budget will expand to fit the money available, such that 2010’s $15000 leaves people just as desperate as 1900’s $1500, what does it matter if some people get an extra $6000?

The contradiction between the first and second reminds me of Tucker-Drob on IQ. He resolves a long-standing debate on whether intelligence is more heritable in poor than in rich individuals by finding this was true in the US but not in Europe. This suggests that American poverty can genuinely lower IQ (and presumably all the other good things associated with IQ like responsibility and prosocial behavior), but European poverty can’t. The study didn’t find this to be related to the US’ greater racial diversity, but it might have to do with the worse social safety net or just changes in the level and nature of poverty. Take this seriously, and it reconciles the first and second article. Getting more money might not help long-term outcomes in Sweden, but in certain kinds of extreme poverty in America – like the type you might find on an Indian reservation – maybe it would.

The third article is more complicated. The second article says:

> What, then, is the cause of mental health problems among the poor? Nature or culture? Both, was Costello's conclusion, because the stress of poverty puts people genetically predisposed to develop an illness or disorder at an elevated risk.

Maybe with the right genes it might be easier to rise out of poverty; I guess the stories of famous entrepreneurs who did exactly that already suggest that. With the wrong genes, it might be much harder but – at least in America, at least if given large amounts of money – still possible.

Also, regarding that Chinese immigrant – I, too, have worked a $20,000/year job and managed to save a lot of money while doing so. I think my "secret" was not having a car, debts, drugs, or dependents; it seems the Chinese guy's secret is the same. Exactly how easy this strategy is for the average person is left as an exercise for the reader, but I'm impressed with how culturally malleable it seems to be. If we're worse at this kind of thing today than in 1900, maybe the extra is just compensating for those sorts of problems.

I think this can be considered me slightly changing my opinion stated here to be more optimistic about the possibility of alleviating the most extreme poverty. But it still seems like money transfers are the way to go.

Against Dog Whistle-ism

I.

Back during the primary, Ted Cruz said he was against "New York values".

A *chump* might figure that, being a Texan whose base is in the South and Midwest, he was making the usual condemnation of coastal elites and arugula-eating liberals that every other Republican has made before him, maybe with a special nod to the fact that his two most relevant opponents, Donald Trump and Hillary Clinton, were both from New York.

But *sophisticated people* immediately detected this as an "anti-Semitic dog whistle", eg Cruz's secret way of saying he hated Jews. Because, you see, there are many Jews in New York. By the clever strategem of using words that had nothing to do with Jews or hatred, he was able to effectively communicate his Jew-hatred to other anti-Semites without anyone else picking up on it.

Except of course the entire media, which seized upon it as a single mass. New York values is coded anti-Semitism. New York values is a classic anti-Semitic slur. New York values is an anti-Semitic comment. New York values is an anti-Semitic code word. New York values gets called out as anti-Semitism. My favorite is this article whose headline claims that Ted Cruz "confirmed" that he meant his New York values comment to refer to Jews; the "confirmation" turned out to be that he referred to Donald Trump as having "chutzpah". It takes a lot of word-I-am-apparently-not-allowed-to-say to frame that as a "confirmation".

Meanwhile, back in Realityville (population: 6), Ted Cruz was attending synagogue services at his campaign tour, talking about his deep love and respect for Judaism, and getting described as "a hero" in many parts of the Orthodox Jewish community" for his stance that "if you will not stand with Israel and the Jews, then I will not stand with you."

But he once said "New York values", so clearly all of this was just really really deep cover for his anti-Semitism.

II.

Unlike Ted Cruz, former London mayor Ken Livingstone said something definitely Jew-related and definitely worrying.

A month or two ago a British MP named Naz Shah got in trouble for sharing a Facebook post saying Israel should be relocated to the United States. Fellow British politician Ken Livingstone defended her, and one thing led to another, and somewhere in the process he might have kind of said that Hitler supported Zionism.

This isn't *totally* out of left field. During the Nazi period in Germany, some Nazis who wanted to get rid of the Jews and some Jews who wanted to get away from the Nazis created the Haavara Agreement, which facilitated German Jewish emigration to Palestine. Hitler was ambivalent on the idea but seems to have at least supported some parts of it at some points. But it seems fair to say that calling Hitler a supporter of Zionism was at the very least a *creative* interpretation of the historical record.

The media went further, again as a giant mass. Ken Livingstone is anti-Semitic. Ken Livingstone is anti-Semitic. Ken Livingstone is anti-Semitic. Ken Livingstone is anti-Semitic. Ken Livingstone is anti-Semitic. I understand he is now having to defend himself in front of a parliamentary hearing on anti-Semitism.

So. First things first. Ken Livingstone is tasteless, thoughtless, embarrassing, has his foot in his mouth, is inept, clownish and offensive, and clearly made a blunder of cosmic proportions.

But is he *anti-Semitic*?

When I think “anti-Semitic”, I think of people who don’t like, maybe even hate, Jews. I think of the medieval burghers who accused Jews of baking matzah with the blood of Christian children. I think of the Russians who would hold pogroms and kill Jews and burn their property. I think of the Nazis. I think of people who killed various distant family members of mine without a second thought.

Obviously Livingstone isn’t *that* anti-Semitic. But my question is, is he anti-Semitic *at all?* Is there any sense in which his comments reveal that, in his heart of hearts, he really doesn’t like Jews? That he thinks of them as less – even slightly less – than Gentiles? That if he were to end up as Prime Minister of Britain, this would be bad in a non-symbolic, non-stupid-statement-related way for Britain’s Jewish community? Does he just say dumb things, or do the dumb things reflect some underlying attitude of his that colors his relationship with Jews in general?

(speaking of “his relationship with Jews”, he brings up in his own defense that two of his ex-girlfriends are Jewish)

I haven’t seen anyone present any evidence that Livingstone has any different attitudes or policies towards Jews than anyone else in his general vicinity. I don’t think even his worst enemies suggest that during a hypothetical Livingstone administration he would try (or even want) to kick the Jews out of Britain, or make them wear gold stars, or hire fewer Jews for top posts (maybe he'd hire *more*, if he makes his hiring decisions the same way he makes his dating decisions). It sounds like he might be less sympathetic to Israel than some other British people, but I think he describes his preferred oppositional policies toward Israel pretty explicitly. I don't think knowing that he made a very ill-advised comment about the Haavara agreement should make us believe he is lying about his Israel policies and would actually implement ones that are even more oppositional than he's letting on.

Where am I going with this? It's stupid to care that Ken Livingstone describes 1930s Germany in a weird way *qua* describing 1930s Germany in a weird way; he's a politician and not a history textbook writer. It seems important only insofar as his weird description reveals something about him, insofar as it's a sort of Freudian slip revealing deep-seated attitudes that he had otherwise managed to keep hidden. The British press framed Livingstone's comments as an explosive revelation, an "aha! now we see what Labour is really like!" They're really like…people who describe the 1930s in a really awkward and ill-advised way? That's not a story. It's a story only if the weird awkward description reveals more important attitudes of Livingstone's and Labour's that might actually affect the country in an important way.

But not only is nobody making this argument, but nobody even seems to think it's an argument that has to be made. It's just "this is an offensive thing involving Jews, that means it's anti-Semitic, that means the guy who said it is anti-Semitic". Maybe he is. I'm just not sure this incident proves much one way or the other.

III.

Nobody reads things online anymore unless they involve senseless violence, Harambe the gorilla, or Donald Trump. I can't think of a relevant angle for the first two, so Trump it is.

Donald Trump is openly sexist. We know this because every article about him prominently declares that he is "openly sexist" or "openly misogynist" in precisely those words. Trump is openly misogynist. Trump is openly misogynist. Trump is openly misogynist. Trump shows blatant misogyny. Trump is openly sexist. Trump is openly sexist and gross.

But if you try to look for him being openly anything, the first quote anyone mentions is the one where he says Megyn Kelly has blood coming out of her "wherever". As somebody who personally ends any list of more than three items with "… and whatever", I may be more inclined than most to believe his claim that no anatomical reference was intended. But even if he was in fact talking about her anatomy – well, we're back to Livingstone again. The comment is crude, stupid, puerile, offensive, gross, inappropriate, and whatever. But *sexist?*

When I think of "sexist" or "misogynist", I think of somebody who thinks women are inferior to men, or hates women, or who thinks women shouldn't be allowed to have good jobs or full human rights, or who wants to disadvantage women relative to men in some way.

This does not seem to apply very well to Trump. It's been remarked several times that his policies are more "pro-women" in the political sense than almost any other Republican candidate in recent history – he defends Planned Parenthood, defends government support for child care, he's flip-flopped to claiming he's pro-life but is much less convincing about it than the average Republican. And back before his campaign, he seems to have been genuinely proud of his record as a pro-women employer. From his *Art of the Deal*, written in the late 1980s (ie long before he was campaigning):

> The person I hired to be my personal representative overseeing the construction, Barbara Res, was the first woman ever put in charge of a skyscraper in New York… I'd watched her in construction meetings, and what I liked was that she took no guff from anyone. She was half the size of most of these bruising guys, but she wasn't afraid to tell them off when she had to, and she knew how to get things done.
>
> It's funny. My own mother was a housewife all her life. And yet it's turned out that I've hired a lot of women for top jobs, and they've been among my best people. Often, in fact, they are far more effective than the men around them. Louise Sunshine, who was an executive vice president in my company for ten years, was as relentless a fighter as you'll ever meet. Blanche Sprague, the executive vice president who handles all sales and oversses the interior design of my buildings, is one of the best salespeople and managers I've ever met. Norma Foerderer, my executive assistant, is sweet and charming and very classy, but she's steel underneath, and people who think she can be pushed around find out very quickly that they're mistaken.

There have since been a bunch of news reports on how Trump was (according to the Washington Post) "ahead of his time in providing career advancement for women" and how "while some say he could be boorish, his companies nurtured and promoted women in an otherwise male-dominated industry". According to internal (ie hard-to-confirm) numbers, his organization is among the few that have more female than male executives.

Meanwhile, when I check out sites like Women Hold Up Signs With Donald Trump's Most Sexist Quotes, the women are holding up signs with quotes like "A person who is flat-chested is very hard to be a 10" (yes, he actually said that). This is undeniably boorish. But are we losing something when we act as if "boorish" and "sexist" are the same thing? Saying "Donald Trump is openly boorish" doesn't have the same kind of ring to it.

This bothers me in the same way the accusations that Ken Livingstone is anti-Semitic bother me. If Trump thinks women aren't attractive without big breasts, then His Kink Is Not My Kink But His Kink Is Okay. If Trump is dumb enough to say out loud that he thinks women aren't attractive without big breasts, that says certain things about his public relations ability and his dignity-or-lack-thereof, but it sounds like it requires a lot more steps to suggest he is a bad person, or would have an anti-woman administration, or anything that we should actually care about.

(if you're going to bring up "objectification", then at least you have some sort of theory for how this tenuously connects, but it doesn't really apply to the Megyn Kelly thing, and anyway, this)

And what bothers me most about this is that word "openly". Donald Trump says a thousand times how much he wants to fight for women and thinks he will be a pro-women president, then makes some comments that some people interpret as revealing a deeper anti-women attitude, and all of a sudden he's *openly* sexist? Maybe that word doesn't mean what you think it means.

IV.

I don’t want to claim dog whistles don’t exist. The classic example is G. W. Bush giving a speech that includes a Bible verse. His secular listeners think “what a wise saying”, and his Christian listeners think “ah, I recognize that as a Bible verse, he must be very Christian”.

The thing is, we *know* G. W. Bush was pretty Christian. His desire to appeal to Christian conservatives isn’t really a secret. He might be able to modulate his message a little bit to his audience, but it wouldn’t be revealing a totally new side to his personality. Nor could somebody who understood his “dog whistles” predict his policy more accurately than somebody who just went off his stated platform.

I guess some of the examples above might have gotten kind of far from what people would usually call a “dog whistle”, but I feel like there’s an important dog-whistle-related common thread in all of these cases.

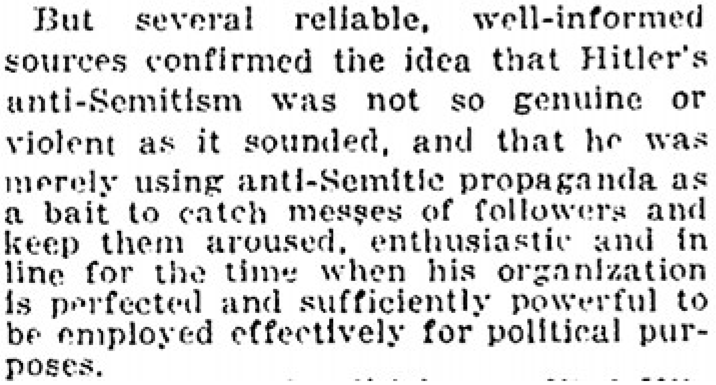
In particular, I worry there’s a certain narrative, which is catnip for the media: *Many public figures are secretly virulently racist and sexist. If their secret is not discovered, they will gain power and use their racism and sexism to harm women and minorities. Many of their otherwise boring statements are actually part of a code revealing this secret, and so very interesting. Also, gaffes are royal roads to the unconscious which must be analyzed obsessively. By being very diligent and sophisticated, journalists can heroically ferret out which politicians have this secret racism, and reveal it to a grateful world.*

There’s an old joke about a man who walks into a bar. The bar patrons are holding a weird ritual. One of them will say a number, like “twenty-seven”, and the others of them will break into laughter. He asks the bartender what’s going on. The bartender explains that they all come here so often that they’ve memorized all of each other’s jokes, and instead of telling them explicitly, they just give each a number, say the number, and laugh appropriately. The man is intrigued, so he shouts “Two thousand!”. The other patrons laugh uproariously. “Why did they laugh more at mine than any of the others?” he asks the bartender. The bartender answers “They’d never heard that one before!”

In the same way, although dog whistles do exist, the dog whistle narrative has gone so far that it’s become detached from any meaningful referent. It went from people saying racist things, to people saying things that implied they were racist, to people saying the kind of things that sound like things that could imply they are racist even though nobody believes that they are actually implying that. Saying things that sound like dog whistles has itself become the crime worthy of condemnation, with little interest in whether they imply anything about the speaker or not.

Against this narrative, I propose a different one – politicians’ beliefs and plans are best predicted by what they say their beliefs and plans are, or possibly what beliefs and plans they’ve supported in the past, or by anything other than treating their words as a secret code and trying to use them to infer that their real beliefs and plans are diametrically opposite the beliefs and plans they keep insisting that they hold and have practiced for their entire lives.

Let me give a snarky and totally unfair example. This is from the *New York Times* in 1922 (source):



But several reliable, well-informed sources confirmed the idea that Hitler's anti-Semitism was not so genuine or violent as it sounded, and that he was merely using anti-Semitic propaganda as a bait to catch messes of followers and keep them aroused, enthusiastic and in line for the time when his organization is perfected and sufficiently powerful to be employed effectively for political purposes.

I won’t say we should always believe that politicians are honest about their beliefs and preferred policies. But I am skeptical when the media claims to have special insight into what they really think.

Book Review: Unlearn Your Pain

[Content warning: discussion of chronic pain and related conditions, and the debate over whether some of them may be psychological in origin. None of this is medical advice or a recommendation to start or stop any form of therapy. Low confidence in my conclusions here.]

I.

Some of the most interesting lectures in medical training are the ones that start with "Okay, you're all going to think I'm a quack, but…"

This was how Dr. Howard Schubiner started the lecture he gave at the hospital where I work. Dr. Schubiner isn't an *obvious* quack – he's a professor of medicine at the local university, directs a clinic at a reputable hospital nearby, and is on the editorial boards of a bunch of medical journals. And although his lecture raised what we will generously call a few red flags, there was also just enough interesting stuff there that I couldn't resist buying his book *Unlearn Your Pain* to learn more.

Dr. Schubiner's specialty is psychosomatic complaints – bodily symptoms that don't come from any obvious disease and seem to reflect psychological stress. Everyone agrees that this category exists. Most doctors have stories about conversion disorder – usually patients who become "paralyzed" in previously healthy limbs after some life crisis. One of my medical school professors had a pretty good diagnostic test for this – feign a punch at the patient's face, really quickly, without warning her. If she instinctively uses her "paralyzed" limb to block it, it's conversion disorder. The same sort of thing works for pseudoseizures – apparent seizures not associated with objective seizure EEG activity. There's a legend about a neurologist telling a medical student that a certain patient’s fit was a pseudoseizure, and the patient interrupting his seizure to protest “No it isn’t!”.

Most people who have worked with conversion or pseudoseizure patients don’t doubt their inherent honesty. These patients aren’t *faking*, per se. Such a person genuinely can’t move their limb, can’t just decide not to have seizures. Often they’re very distressed at what’s happening to them (although sometimes they really aren’t). Psychologists like to say that it’s subconscious – whatever that means. Just like somebody crippled by panic attacks, the symptoms are real and involuntary, but they’re also psychologically produced.

The existence of this category isn’t controversial, but its size definitely is. Some people propose a long list of conditions – fibromyalgia, irritable bowel syndrome, chronic fatigue syndrome, chronic Lyme disease, tension headaches, interstitial cystitis, et cetera – that they think of as mostly or entirely psychosomatic. On the other hand, patients’ rights groups get very upset at claims that their conditions are “all in their head”, accuse doctors of thinking that they’re lazy or making up their symptoms, and pass around stories with titles like RE: RE: RE: FWD: RE: THE MEDICAL PROFESSION about some guy whose doctor dismissed him as making up his symptoms but who was later diagnosed with zebra-itis and cured with an experimental gene therapy treatment.

Dr. Schubiner is a psychosomatic complaint maximalist. He thinks that just about anything that can’t be traced to a well-understood physiological cause is probably psychosomatic – in his language, Mind-Body Syndrome or MBS. He quotes a fascinating theory by Edward Shorter that this all dates back to the invention of the tendon hammer, ie that little thing doctors hit your knee with:

> An important advance in medicine was the discovery of deep tendon reflexes. The simple test of striking a tendon with a reflex hammer can quickly distinguish pathological from psychological paralysis. Amazingly, once doctors could do this test, the number of people with this type of conversion disorder decreased substantially, and now the condition is rare. When doctors and the general public come to view a medical condition as psychologically induced, it is less likely to occur..the subconscious mind is unlikely to produce symptoms that will be easily seen as psychological. But since humans continue to experience great stresses and strong emotions, paralysis has been replaced by chronic back pain, fibromyalgia, fatigue, irritable bowel syndrome, and many other symptoms.

When Schubiner talks about fibromyalgia and fatigue, he's not so far outside (one edge of) respectable medical opinion. But he goes further and lists migraine headaches, heartburn, carpal tunnel syndrome, tinnitis, postural orthostatic tachycardia, repetitive stress injury, and reflex sympathetic dystrophy as likely Mind-Body Syndrome as well. And most explosively, he says the condition explains almost all pain.

Schubiner admits there is such a thing as anatomically-caused non-psychological pain. It tends to be associated with very obvious injuries like dropping an anvil on your foot, and it tends to go away after a couple of weeks at most. Anything more mysterious and chronic – facial pain, TMJ pain, joint pain, abdominal pain, neck pain, shoulder pain, tendonitis, and *especially* back pain – is probably Mind-Body Syndrome. After a medical workup has failed to reveal obvious cancer or infection, these are almost certainly psychosomatic and continuing to treat them as potentially medical just makes them worse:

> When patients with Mind-Body Syndrome are labeled as having degenerative disc disease on the basis of an MRI….symptoms can be exacerbated and patients harmed by medical diagnoses. This occurs because the diagnosis creates fear and the belief that there is something seriously wrong with one's body. These emotions activate the anterior cingulate cortex, which creates even more pain by ramping up the learned nerve pathways of MBS.

If this were true, it would be *really* important. Surveys suggest that between 40 million and 100 million Americans have chronic pain; the former study finds 67% of them say their pain is "constantly present" and 50% say it is sometimes "unbearable and excruciating". The financial cost is between $60 billion and $600 billion per year. I'm not sure what to think about all these estimates that differ by orders of magnitude, but the point is that there's a "chronic pain epidemic" and it's really bad. The mainstay of treatment for chronic pain is opioids, and by non-coincidence there's also an opioid addiction epidemic and an opioid-related death epidemic. To some degree the government can use regulation to trade off pain burden against opiate deaths, but no point at that curve is very palatable and we desperately need some kind of real solution.

Schubiner says he has it. It's time to admit that all of this pain that's getting all epidemicky is almost entirely psychosomatic. It might start with a real injury, but after that injury heals the brain "remembers" the relevant pain pathways and exploits them as a way to express psychological stress. He presents some fascinating and delightful evidence for this.

A guy named Harold Schraeder studied prevelance of chronic whiplash in Lithuania, of all things. He found the prevalence was zero. In most Western nations, a certain subset of people who get in car accidents suffer chronic disabling neck pain, presumably related to having their neck get suddenly jerked by the force of the impact. But Schrader found that this never happened in Lithuania, even though they had a lot of accidents and their cars were no safer than ours. Simotas and Shen found that there was zero whiplash in demolition derby drivers, even though they got into crashes all the time and it was basically their job description. Further studies found that accident victims with more neck injury were no more likely to develop whiplash than victims with less neck injury. Perhaps, they argue, chronic whiplash isn't a bodily injury at all, but a culture-bound syndrome in which people who *expect* whiplash to exist use its symptom profile as a way of expressing their psychological tension.

Then there's back pain, one of the most common and disabling types of chronic pain – Medicare back-pain related costs have grown about 3-4x in a decade. Standard medical workup for back pain usually involves getting an x-ray or MRI, finding some problem with the discs in the spine, and treating with painkillers, steroids, or surgery. Schubiner is not convinced. He notes studies that find that radiographic findings of disc degeneration or herniation do not accurately predict future back pain. Yes, most back pain sufferers will have problems visible on MRI, but most perfectly healthy people pulled off the street will *also* have problems visible on MRI – for example, this study finds that half of all 21-year-olds in Finland have a degenerated disc, and a quarter have a bulging disc. The book quotes an NEJM article as saying that "neither baseline MRIs nor followup MRIs are useful predictors of low back pain". However, studies (1, 2) find that a patient's job satisfaction *does* predict their future back pain. Books and studies called things like Time To Back Off?, Back In Control and Watch Your Back point out that many surgeries and injections for back pain work no better than placebos in controlled experiments, with a review article in the Journal of the *American Board of Family Medicine* concluding that “prescribing yet more imaging, opioids, injections, and operations is not likely to improve outcomes for patients with chronic back pain” On the other hand, Schofferman et al found that childhood trauma correlates heavily with success of back pain treatment: 95% of patients with a happy childhood got better after back surgery, but only 15% of patients with multiple childhood traumas did.

Based on these studies and others like them, Schubiner concludes that chronic back pain is psychological rather than physiological. He thinks there may have been some original minor injury, of the sort that most people would get over in a couple of weeks. This causes the nerves to “sensitize” – ie the brain is primed to think about and remember this form of pain. Then, when people recall their subconscious tension over childhood trauma and the stresses of life, they express it as back pain through the sensitive nerve pathways.

Extend this model to headaches, irritable bowel, chronic fatigue, and everything else, and you have Dr. Schubiner’s theory of pain.

II.

If all of this pain has a psychological cause, then it should have a psychological solution. But psychological solutions to chronic pain are no more effective than physical ones. For example, Cochrane Review finds that cognitive behavioral therapy for chronic back pain has a moderate short-term effect which fades quickly. There was a similar effect for neck pain which Cochrane found “could not be considered clinically meaningful” and which also faded quickly.

Dr. Schubiner says this is because cognitive behavioral therapy is inappropriate for this condition. It’s caused not by negative thoughts and dysfunctional behaviors, but by unresolved childhood traumas. He recommends a therapy designed to help resolve such traumas called Intensive Short Term Dynamic Psychotherapy.

Freudian therapy (“psychoanalysis”) usually takes several years to get anywhere, and may take ten years or longer to complete. Around the 1960s, some psychiatrists got tired of waiting and invented a high-speed version called “psychodynamic therapy”. Schubiner’s version, which derives from the word of a guy named Dr. Davanloo in Montreal, promises results in as little as four weeks. It involves a lot of stuff, including some kind of silly-sounding things like writing affirmations, finding your acupuncture points, and putting your body in very masculine “power poses” and raising your fists and shouting “I AM GOING TO OVERCOME MY PAIN!”. Dr. Schubiner demonstrated this last in front of us and he did indeed look very masculine and determined; if I were chronic back pain, I would definitely put him on my list of people to avoid.

But the heart of the therapy is a technique for returning to traumatic childhood moments. You try to figure out what your traumatic childhood moments were – for example, maybe your father got drunk and beat you up. So you go back to the incident, either as a solo visualization or in a conversation with a partner. It goes something like this:

> **Doctor:** Tell me what’s happening

> **Patient:** I’m in my childhood home. My father approaches me, beer bottle in hand, looking really angry.
>
> **Doctor:** How do you feel?
>
> **Patient:** Really scared.
>
> **Doctor:** But also?
>
> **Patient:** Angry.
>
> **Doctor:** Good! You have every right to! Tell your father that!
>
> **Patient:** Father, I’m really scared and angry!
>
> **Doctor:** Now what does your father do?
>
> **Patient:** He doesn’t care.
>
> **Doctor:** And what would you like to do now?
>
> **Patient:** Beat up my father.
>
> **Doctor:** Then go back into that experience and beat up your father.
>
> **Patient:** AAAARGH! I HATE YOU SO MUCH, FATHER! YOU RUINED MY CHILDHOOD! GRAAAAAAAAAH! DIE! DIE! DIE! DIE! DIE!
>
> **Doctor:** How do you feel now?
>
> **Patient:** My chronic back pain is gone!

There are enough variations on this to make it a four week course, but in Schubiner’s examples (which he takes from real clinical practice), even something as simple as this can be enough to make chronic pain go away near-instantly. He has about a dozen anecdotes from his own practice where this happens. Then the rest of the course is just solidifying that gain and making sure it doesn't come back.

I talked to a professor of psychoanalysis I work with about this. She says that Davanloo is well within the psychoanalytic mainstream. She says that she herself is not a big fan of his work, because she thinks it's important to spend those several years unpeeling a patient's defenses instead of just smashing them with a sledgehammer. But she says chronic pain patients may have unusually strong defenses and that maybe the sledgehammer approach is the right one. So overall she cautiously approves.

On the other hand, my more cynical readers might note that "well within the psychoanalytic mainstream" isn't exactly equivalent to "definitely not a quack". Schubiner is aware of this and has tried to get some evidence for his method. Along with a case series, he has published a study on the psychological treatment for fibromyalgia, in which 45% of the intervention group experienced significant pain relief compared to 0% of the controls. He also tested the full version of his therapy in a preliminary trial in which "two-thirds of the patients improved at least 30% in pain".

So, should we believe him?

III.

I tried to verify some of the claims in this book and discovered things were much more ambiguous than it let on.

The idea of whiplash as psychosocial is still controversial in the literature. It's true that some studies in Lithuania and

Greece show almost no whiplash. But some critics say that these studies lacked enough power to find a difference in whiplash rates among countries even if such a difference existed. There were also extremely weird fluctuations in the data – for example, the same team in the same city doing two studies a few years apart found neck injury rates of 15% versus 47%. Here's a long and acrimonious debate in a medical journal about this. But interestingly, even the pro-psychosocial side doesn't seem to want to say there's *no* biological component. And such a claim would be difficult to sustain given studies that show significant effects of things like head position during an accident on future whiplash rates. You can find a good summary of some of the points on each side here. Here's another review by Dr. Arthur Croft, Ph.D., D.C., M.Sc., M.P.H., F.A.C.O., and Emmy award nominee (really) – who says, brutally:

> Ferrari, et al., have recently promoted the so-called biopsychosocial model in the context of whiplash, making numerous excursions into the literature in support of it. The lynchpin of their theory relies on two studies conducted in Lithuania which purportedly followed the natural history of late (i.e., chronic) whiplash in a population of persons exposed to rear-impact motor vehicle crashes – the putative injury mechanism for acute whiplash injury. Unfortunately, "fatal" errors in study design in both cases prevented meaningful interpretation of their results, not the least of which was that only a small portion of their cohort actually had an acute whiplash – the necessary precursor for late whiplash. Our post hoc power calculation revealed that their cohort was inadequate to support any of their conclusions.
>
> Unfortunately, many authors – since these flaws were pointed out – have failed to be dissuaded from citing

> this literature in support of the biopsychosocial theory, particularly those authors from the camp of the nonbelievers. These Lithuanian papers, it should also be noted, stand alone as outliers to more than 50 other published reports of outcome over the past 45 or so years, and arrive at some rather improbable conclusions that almost immediately beg some questions. In the first paper, the results suggested that persons exposed to whiplash mechanisms would have about the same long-term neck pain as age-matched uninjured persons in the population. In the second study, acute whiplash trauma exposure seemed to actually have a protective effect, somehow immunizing these people against future neck pain. Again, these findings would be particularly interesting if the studies had the added virtue of being valid on a statistical and methodological basis.

To these I would add that even if a US-Lithuania whiplash rate difference existed, it wouldn't necessarily have to be psychological. Some people bring up Americans having bigger cars; other people bring up differences between American and European car seat headrest design, and still others bring up differences in US and Lithuanian diets and lifestyles which might affect pain and healing.

The book's treatment of back pain also raises some concerns. It is definitely true that the relationship between back pain and radiographic findings is way lower than anybody wants to admit, and that MRI isn't very useful for diagnosis. That having been said, the relationship is not zero. For example, in this study, patients with the highest level of radiographic degeneration were 4.5x more likely to be in the highest back pain category compared to patients with the lowest level of radiographic degeneration. The correlation is not one, but neither is it zero. Some people with lots of back pain will have no radiographic findings, and some people with lots of radiographic findings will have no back pain, and the relationship is weak enough that using MRIs for diagnosis is heavily discouraged, but *in general* the two have a positive and significant relationship. See for example here, here, and here. I don't know to what degree this affects the book's thesis, but it seemed important to point out.

Nor is there any more clarity about the relationship of back pain to job (dis-) satisfaction. I was able to find three meta-analyses on it. One of these, Linton, said that:

> The available literature indicated a clear link between psychological variables and neck and back pain. The prospective studies indicated that psychological variables were related to the onset of pain, and to acute, subacute, and chronic pain. Stress, distress, or anxiety as well as mood and emotions, cognitive functioning, and pain behavior all were found to be significant factors. Personality factors produced mixed results. Although the level of evidence was low, abuse also was found to be a potentially significant factor.

But Hartvingsen et al concluded:

> According to recent epidemiological literature we found moderate evidence for no positive association between perception of work, organisational aspects of work, and social support at work and LBP. We found insufficient evidence for an association between stress at work and LBP. Regarding consequences of LBP, there was insufficient evidence for an association between perception of work in relation to consequences of LBP. There was strong evidence for no association between organisational aspects of work and moderate evidence for no association between social support at work and

> stress at work and consequences of LBP. There were major methodological problems in the majority of studies included in this review and the diversity in methods was considerable. Therefore associations reported may be spurious and should be interpreted with caution.

Finally, Hoogendoorn, Poppel and Bongers, who sound like a band for very young children, are very ambivalent. They do find some effects, but they all give off an air of desperation, eg “Low job control was found to have a statistically significant positive effect on short and long absences due to back pain, except in men in lower grade jobs and women in higher grade jobs, in whom the effect was reversed”. They are very open about this, and conclude:

> “Evidence was found for the effect of some of the psychosocial work characteristics, but there is no psychosocial work characteristic for which evidence was found in all reviews…the conclusions drawn in the various reviews appear to be rather heterogenous”.

And, very significantly for our purposes:

> “Having evaluated the strength of the evidence for both physical and psychosocial factors as risk factors for back pain, using the same methods, the question arises of whether the findings indicated a difference in the evidence for physical and psychosocial factors. Strong or moderate evidence has been found for heavy physical work, lifting, bending, and twisting, and whole body vibration at work. Unlike the results for psychosocial factors, these results were rather insensitive to slight changes in the assessment of the findings and the methodologic quality of the studies and in agreement with the results of previous reviews on

> physical load. This indicates that the body of evidence supporting the role of these physical load factors as risk factors for back pain is somewhat more consistent than that for the psychosocial factors)

The consensus in pain medicine is that pain depends on both psychological and physical factors working together. Schubiner is trying to shift that consensus to say pain is almost entirely psychological and based mainly on childhood trauma. But the studies, while not ruling out a psychological cause, are very emphatic that physical causes definitely matter. And even the papers supporting psychological causes say that, among all such causes, there is unusually *little* evidence for childhood abuse as a factor.

But the part that bothered me most was the use of Schofferman's study showing that childhood trauma predicted back surgery success rate (I should note that he doesn't cite this explicitly in the book, but he implicitly works off it, and he discussed it explicitly during the lecture). This was a surprising study that cried out for replication – and which was in fact re-tested in 2002 on a larger sample by Nickel, Egle, and Hardt. They were unable to replicate the findings. Chronic back pain patients, surgery-failing and otherwise, were no more likely to have childhood trauma than anybody else. This bothered me because Schubiner played up Schofferman's 1991 study that supported his hypothesis without even mentioning this one. People have a right to present their case the way they want, but when someone clearly ignores better and more recent evidence, it makes me a little more skeptical of everything they say.

IV.

What about the psychiatric part of *Unlearn Your Pain*'s program?

The psychodynamic therapy literature is even more of a mess than the back pain literature. I've been there before and don't want to go back. You can read Jonathan Shedler and Jared DeFife in support and James Coyne and Michael Anestis in opposition. I find myself more sympathetic to the "doesn't work very well" camp, but the field is muddy enough, and my biases against it strong enough, that I place little confidence in that judgment.

So let me try to cut through all of this with my favorite weapon for these kinds of things: behavioral genetics. Of the five behavioral genetics studies on back pain I could find, four (1, 2, 3, 4) found no shared environmental effect on back pain, with only one dissenter. This is in common with a large literature finding little shared environmental effect on a host of psychological problems including depression, anxiety, and bipolar disorder – and indeed, it would be very strange if chronic pain were more related to childhood experiences than those were.

Psychiatry tried really hard to give the "childhood trauma causes everything" thesis a go for fifty-something years. Sure enough, psychiatrists found loads of childhood trauma, because, much as pretty much everybody will have something weird with the discs in their backs that can be detected on MRI, pretty much everybody will have something weird with their childhood that can be detected with psychotherapy. Using the kabbalistic method, you can always find suspicious coincidences linking their childhood trauma with their current pain. Schubiner writes – as far as I can tell, 100% seriously – that:

> When someone develops a pain in the buttocks, there may be someone in their lives who is ‘a pain in the butt’.” Someone who develops difficulty swallowing may be reacting to a situation in life that is ‘hard to swallow’. I evaluated a woman with pain in the bottom of her feet. While waiting in line one day, she realized there was a situation in her life that she ‘just couldn’t stand anymore’

I want you to appreciate how much willpower I’m showing here. There is form of psychiatry based around corny puns, and yet instead of emailing these people my resume immediately I’m trying to maintain a cautious skepticism.

And when I do, I just can’t believe it. The early psychoanalysts weren’t doing science, they were taking Sofer’s Law and running with it. Eventually we realized that talking about childhood traumas wasn’t predictive, wasn’t especially curative as per rigorous studies, and we moved on.

There’s a lot of controversy around this decision, but I think behavioral genetics has made the childhood-trauma side increasingly untenable. Assuming twin studies aren’t entirely fatally flawed – something thousands of people have looked for and nobody has found – childhood shared environment, which presumably includes things like abusive parents, just doesn’t affect adult outcomes very much. I can’t see a way to reconcile that with psychoanalytic theories and I don’t think we should keep trying.

I don’t deny that there are a lot of suspicious coincidences. But I think if we look harder, we can find that those suspicious coincidences all have more reasonable explanations. Like, yes, people with a lot of psychological problems tend to have a lot of back pain. But again, when you do twin studies:

> On initial analysis considering the participants as individuals, rather than twins — and therefore not accounting for genetic and familial factors — the odds of having back pain were about 1.6 higher for those with symptoms of depression and anxiety.
>
> On further analysis of monozygotic twins — who are genetically identical — the association between symptoms of depression and low back pain disappeared. This suggested that the strong association found in non-identical twins resulted from the "confounding" effects of common genetic factors influencing both conditions. For example, genes affecting levels of neurotransmitters such as serotonin and norepinephrine might affect the risk of both conditions.
>
> Previous studies have shown a "consistent relationship" between back pain and depression — a combination that may complicate diagnosis and treatment. However, the nature of the association remains unclear. The new study is the first to examine the relationship between depression and low back pain using twin data to control for genetic and familial factors.

When you control for genetics, WHICH YOU SHOULD ALWAYS DO AND I AM SO SERIOUS ABOUT THIS, this explains the entire psychological problem/back pain link. Combined with the previous twin studies showing no effect of childhood environment, this is a very strong challenge that theories claiming a psychogenic origin of back pain based in life events will have trouble surviving.

V.

So in the end, what do we make of chronic pain?

Many, many, *many* people report using the techniques in *Unlearn Your Pain* (or the closely related techniques of Dr. John Sarno) and having good success. I don’t think this is *entirely* coincidence or bias. But I’m also not willing to entirely buy into this repressed childhood trauma theory.

There are definitely some types of pain which are not related to bodily injury. My best evidence for this, which Dr. Schubiner talks about too, is the people who have pain which is anatomically implausible or “migrating”. By “anatomically implausible” I mean pain that cuts willy-nilly across the body’s neural regions; pain in the distribution of the ulnar nerve may be an ulnar nerve problem, but if it has half the distribution of the ulnar nerve plus half the distribution of the median nerve, while leaving the other half of both distributions pain-free, it’s a little harder to figure out what could be causing it (especially if it’s on both sides equally!) By “migrating”, I mean that somebody has right hand pain, the doctor gives them some kind of treatment, that goes away, the next day they have right foot pain, another treatment, it goes away again, and the next day they have diarrhea. While there are very rare processes that can do something like that, when it goes on long enough that’s good evidence that the pain isn’t anatomical.

But I think the *way* in which pain isn’t anatomical is more complicated than the simple model that *Unlearn Your Pain* uses. Instead of the brain “using” pain to express repressed emotions, maybe gating and modulating pain sensations is just really hard.

Let me give an example [trigger warning for inducing mild bodily discomfort]. Right now, you’re suddenly aware of the feeling of your tongue in your mouth. And right now, the top of your head is suddenly really itchy. Also, right now something is wrong with your saliva and you’re swallowing consciously, but it feels awkward and you’re worried something might be wrong with your throat.

This isn’t because I have magic powers inflicting these things on you, it’s because you’re constantly receiving all sorts of sensations and your brain effectively gates and modulates them. You’ve always got micro-itches and micro-pains going on everywhere – no part of your body is one hundred percent optimal and even if it were there are still variations in neural noise – but your brain usually correctly decides these aren’t worth your time. It’s only when something forces you to focus on them – whether worry about a back injury, or an annoying blogger – that they make it through.

All psychiatric disorders are heavily comorbid. People with one or more of depression, anxiety, OCD, anorexia, autism, gender dysphoria, PTSD, et cetera are many times more likely to have all of the others, and it doesn’t just seem to be in a boring “OCD is depressing and makes me anxious” sort of way. All of this seems to relate to a general factor of neural messed-up-ness. It wouldn’t be surprising if this correlated with some kind of messed-up-ness in the neural systems that are supposed to process and gate pain.

Remember also that stress can cause relapses of many very biological and serious diseases like ulcerative colitis, multiple sclerosis, or epilepsy. Also, inflammation seems to be a shared and complicated factor between various bodily illnesses, stress, and depression. So for stress to cause a “relapse” of chronic pain, all we’d need is for it to put extra pressure – whether through inflammation or some other method – on an already slightly messed-up pain-gating system in the brain. And then there’s muscular tension, which I inexcusably forgot to mention until now but which is also a relevant system by which stress affects chronic pain.

We know that pain is very sensitive to the placebo effect – I’m generally a placebo effect skeptic, but even arch-skeptics Hróbjartsson and Gøtzsche agree that pain is one of the few places where the placebo effect really dominates. This is why we so often see faith healers and saints and miracle water from Lourdes treating pain so effectively – at least briefly. It’s why homeopathic treatment for pain shows such an amazingly good effect size.

And, I will say cynically, it’s why so many people have reported (genuine) success from *Unlearn Your Pain* and related programs. It’s why Schubiner writes:

> I believe that each and every person with Mind Body Syndrome can get better because it is possible to overcome MBS by using this program. Those people who are unable to accept that their symptoms are due to MBS, or who do not develop positive expectations of relief, or who are unable to believe that they can make changes in their health and in their lives are the people who are less likely to improve.

Part of me wants to say that we have a word for medical treatments that only work if you believe that they will, and it rhymes with “gazebo”.

Another part worries this is unfair. If the placebo effect comes from the brain’s ability to gate pain, then saying “You’re not really affecting the brain’s ability to gate pain, it’s just the placebo effect” stops making sense. It’s not that it doesn’t work and is just placebo – it’s that it does work, *via* placebo.

There's something to be said for glorifying in the placebo effect, laying it on as hard as possible, putting on the fanciest robe and wizard hat you can find and saying "I CAST YOU OUT BY THE POWER OF PLACEBO, GO FORTH AND SIN NO MORE!" I think in some ways there can be better and worse placebo therapies just as there can be better and worse real therapies, placebo therapies that activate the placebo effect only a little and don't help much, and placebo therapies that activate the placebo effect really strongly and use it to work miracles. Maybe we should give more status to the best placebo therapies, to view them as highly perfected works of the placebomantic arts in the same way that powerful medications are triumphs of psychopharmacology. I think psychodynamic therapy and everything descended from it would have a high place in that pantheon.

In that sense, I think *Unlearn Your Pain* might be a useful book. I think that even if I accept what I consider the consensus theory of chronic pain – genuine (if small) lingering injuries (or nerve sensitization from such) interacting with a poorly-wired pain gating system in the brain which is highly susceptible to placebo effects – *Unlearn Your Pain* remains a useful book, as the distilled wisdom of many years of work trying to activate those effects as strongly as possible.

Another possibility is that the active ingredient isn't the intensive psychotherapy, it's the belief that the pain is caused by Mind-Body Syndrome. It seems just possible that this belief could break the cognitive loops that seem so relevant in all of these processes.

So I guess I'm in a weird spot in terms of what I think of *Unlearn Your Pain*.

I think it's definitely right that a lot of pain has psychosomatic components. I think it probably helps treat psychosomatic pain, maybe really effectively, and partly for the reasons that it thinks it does.

But I'm not convinced by its more sweeping claims that physical injuries play little-to-no role in chronic pain. Along with Schubiner's talk of nerve sensitization, one can imagine a scenario in which alternatively apparently-healed physical injuries may leave very small irritations on local nerves, and that the degree of irritation a nerve is able to bear without giving you chronic pain is related to your general neural-non-messed-upness and stress level. In such a scenario psychological factors might play a role in gating the pain, or in tensing or releasing muscles around the pain, but would not entirely explain it.

I'm also not convinced by its claims that childhood trauma has any interesting relationship with pain, nor that trauma-related therapy has a unique non-placebo ability to deal with such pain. I think that childhood trauma is overemphasized throughout psychiatry and that this theory of pain represents a step in the wrong direction. If trauma-related therapy works, it works by a nonspecific process of making people feel like they're doing something useful and taking their attribution for their pain off of bodily processes.

Niels Bohr used to hang a horseshoe above the door to his office, saying "I'm not superstitious, but I hear this works whether you believe in it or not." Part of me is tempted to recommend *Unlearn Your Pain* to my patients on the same principle. And if any readers of this blog have chronic pain and want to try to the month-long self-help therapy course in this book, I would be very interested in hearing back from you (please tell me *before* you start, so that there aren't response biases). If the $25 price of this book is the difference between someone in that category trying vs. not trying it, I’m happy to send you the book if you agree to get back to me with your results. Contact me at slatestarcodex@gmail.com if you’re interested in this.

I Wrote A Blog Post, But Did Not Adjust For The Fact That The Title Would Be Too L

I recently got in some fights with psychoanalysts on the importance of parenting. They mentioned that one good test for genuine parent effects – as opposed to genetic effects, stress-related effects, toxin-related effects, et cetera – would be things that seemed to depend more on one parent than the other. In particular, in order to rule out intrauterine factors, we should be looking at effects that depend disproportionately on the father. For example, if young women with distant fathers are uniquely more likely to become lesbians, that would be a pretty convincing demonstration of the importance of parenting.

So I was interested to see a recent study that claimed a good father/son relationship – but not a good mother/son relationship – had a special role in sons' development. University of Guelph, Parents, Especially Fathers, Play A Key Role In Young Adults' Health:

> The researchers found that young adults who grew up in stable families with quality parental relationships were more likely to have healthy diet, activity and sleep behaviours, and were less likely to be obese.
>
> Surprisingly, they found that when it came to predicting whether a young male will become overweight or obese, the mother-son relationship mattered far less than the relationship between father and son.
>
> "Much of the research examining the influence of parents has typically examined only the mother's influence or has combined information across parents,"

> said Prof. Jess Haines, Family Relations and Applied Nutrition, and lead author of the paper.
>
> “Our results underscore the importance of examining the influence fathers have on their children and to develop strategies to help fathers support the development of healthy behaviours among their children.”

Okay. Let’s look at the study. It’s a correlational study of 6000 kids age 14-24. They were asked to rate the quality of their relationship with each parent, then they were tested for various unhealthy behaviors: obesity, eating disorders, fast food intake, soda intake, TV watching, sedentariness, and poor sleep.

Among all participants, better relationships led to less disordered eating, increased physical activity, and better sleep. This was true both for child/mother relationship, child/father relationship, and child/generic-measure-of-family-functioning relationship. So far this isn’t surprising. There was no attempt to control for wealth, class, or anything else, let alone genes. And a lot of these children are still living with their parents, so good parenting is going to be important to them right now (the study didn’t separate children who were still with their parents from adult children who weren’t). No surprise to find an effect here.

Among no participants did better relationships affect soda consumption or screen time, whether it was the child/mother relationship, the child/father relationship, or the child/generic-measure-of-family-functioning relationship. Okay. I guess these are somewhat more neutral things that good parenting doesn’t affect much.

Among female but not male participants, better relationships decrease fast food consumption. This was true both for child/mother relationships, child/father relationships, and child/generic-measure-of-family-functioning relationships (I believe all marriages should be between a man, a woman, and a generic-measure-of-family-functioning). This suggests that maybe parents care more about their daughters eating fast food than their sons – or maybe those daughters themselves care more. In either case, this wouldn't be too surprising.

What about the blockbuster result that fathers, but not mothers, affect male children's obesity level?

The odds ratio for obesity with a good mother-son relationship was 1.04, confidence level (0.85, 1.27).

The odds ratio for obesity with a good father-son relationship was 0.80, confidence level (0.66, 0.98).

Okay. You are measuring seven different outcomes on two different genders of child. On thirteen of these tests, results are concordant between fathers and mothers. On one of them, results are discordant, in that with mothers the confidence interval included 1.00, but with *fathers* the confidence interval merely included 0.98.

You could either conclude that fathers have a unique ability to affect their sons' (but not their daughters') level of obesity (but not disordered eating, or fast food eating, or soda drinking, etc). Or you could conclude that if you do enough tests, 5% of the time something will fall just outside a 95% confidence interval.

Let's see what the study's Limitations section has to say about this:

> We calculated 42 tests and did not adjust for multiple comparisons.

Why would you do this? If NASA preceded their missions with statements like “We are launching a rocket to Jupiter, but we did not adjust for the fact that it is very far away,” we would stop taking them seriously. But for some reason in the social sciences it’s *okay*?

All right, fine, let’s hear your excuse:

> Of these tests, 25 were statistically significant at the 0.05 level, much larger than the 2 we would expect by chance.

This might work for individual results, but it *doesn’t* work for discordances between results, which is what they’re trying to show.

Suppose I want to prove that a certain medicine only works on people whose names begin with the letter M (and suppose in reality, the drug works on everybody). My experiment has 80% power to detect the drug effect when it works. I do fifty tests on fifty different populations – elderly Latino women, young black men, genderqueer Caucasian neonates, Thai rice farmers, unemployed auto workers, whatever – and divide each of them into a subgroup with M-names and a subgroup with other names. I’m actually simulating this right now in an Excel spreadsheet, and here are my results:

Among non-M-names, 42 of the populations test positive, which is much as expected – the drug works and we have 80% power to show that it does, so we should expect 50*0.8 = 40 positive results on average. A little random noise brings that to 42.

Among M-names, 43 of the populations test positive, which is also close to 40. So here everything is just as we would expect.

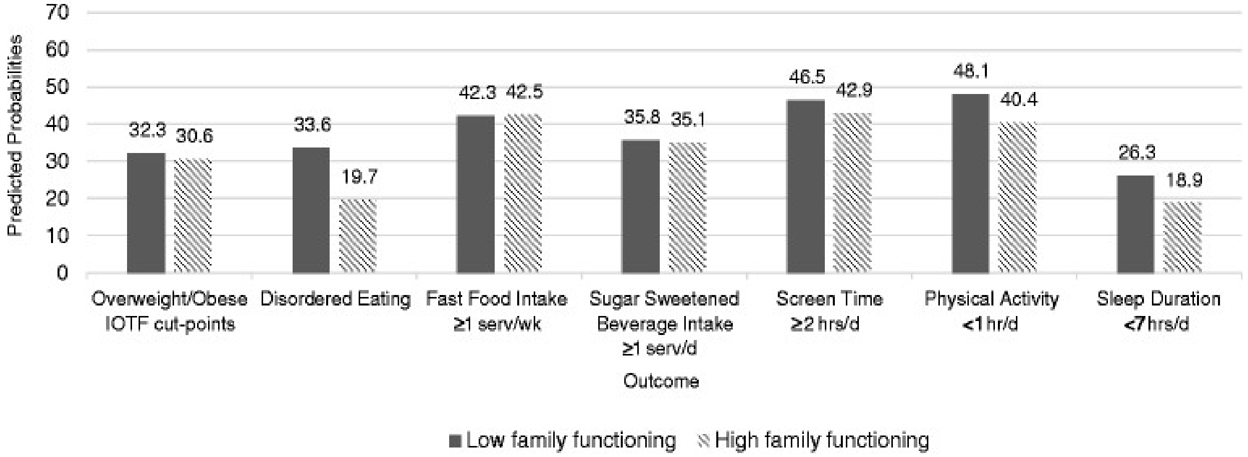
But! In six of the populations, the drug works “differently” for people with M-names and other names. For example, on Test 18 (let’s call this Thai rice farmers), the drug works for rice farmers who have names beginning with M, but doesn’t work for rice farmers who have names beginning with other letters.

So I report this in the literature as “Astounding! Drug works for Thai rice farmers with names beginning with M, but not for Thai rice farmers with names beginning with other letters!” Some annoying person comes back with “but you did a bunch of comparisons and didn’t correct for that”. And I retort “Aha! But actually 85 of my 100 tests came back positive, compared to only 5 that would be expected by pure chance, so clearly there’s something there! There’s an M-name effect after all!”

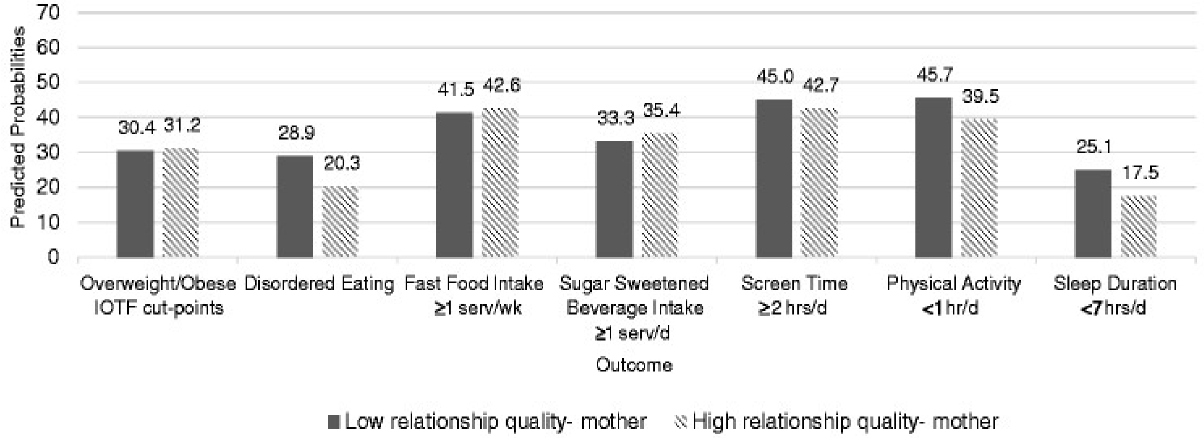
This is comparing apples to oranges. Yes, you’ve shown that your drug works. But you haven’t come *close* to showing that it works differently for people whose name begins with M. Your evidence doesn’t even *suggest* that it does.

But this is what this paper is doing when it says it has evidence that male obesity is affected by the father and not the mother, and claims it doesn’t need to adjust for multiple comparisons.

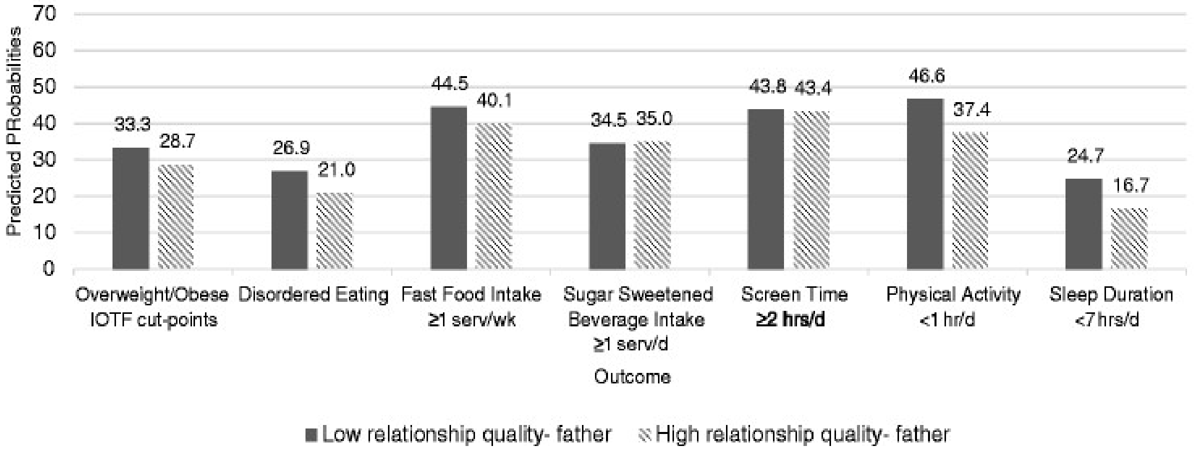
As Exhibit B, I present the graphs:



a. Predicted prevalence (%) of adolescent weight status and weight-related behaviors by family functioning among males



b. Predicted prevalence (%) of adolescent weight status and weight-related behaviors by mother-adolescent relationship quality among males



c. Predicted prevalence (%) of adolescent weight status and weight-related behaviors by father-adolescent relationship quality among males

I think this is noise.

The paper itself mentions the father-son difference in one paragraph in the Discussion section, but doesn't even find it worthy of mention in the Conclusion. It's the press release that plays this up into the major finding of the study. Why?

Because the press release came out three days before Father's Day.

Look:

> In time for Father's Day, a new University of Guelph study has found that parents, and especially fathers, play a vital role in developing healthy behaviours in young adults and helping to prevent obesity in their children.

I think overly cutesy university PR departments do a lot more damage than is generally realized.

On the other hand, one impressive thing about this paper is its willingness to cite large quantities of stuff. For example, a quote:

> Level of bonding or closeness with a parent has also been shown to moderate the association between maternal-BMI and daughter-BMI [17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60] and parental and adolescent weight-related behaviors [17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61].

I am not going to go through 43 studies to see if any of them are any good, but I guess if there are 43 studies claiming these sorts of parental effects I should be a little more humble.

So: does anyone know of any good studies showing gender-specific-parent effects on a child that don’t seem obviously related to intrauterine or Y-chromosomal factors?

Survey Results: Suffering Vs. Oblivion

[Content warning: suffering, oblivion]

Every so often, someone on Reddit realizes that about half of people wipe themselves with toilet paper sitting down, and the other half do it standing up. This discovery is followed by horror on both sides that other people do it differently.

I occasionally have the same feeling when I talk about ethics. Every so often I run up against a base-level clash of intuitions with somebody else, where they disagree on a preference I would have expected to be self-evident. This is pretty bad, since some forms of consequentialism are a lot more elegant if we imagine that all apparent moral disagreements are just people failing to think through their own preferences clearly enough and everyone really agrees about morality deep down.

A lot of these clashes of intuitions have to do with the idea of suffering versus oblivion. In order to explore this further, I asked people on Twitter and Tumblr to answer some survey questions like the following:

> **1.** Would you rather:
>
> Option A: Live the rest of your life working 16 hour days, seven days a week, as a McDonalds cashier. You will have no time off except the time you need to eat, sleep, and use the restroom. Your entire life will be spent doing McDonalds cashier related tasks. When you are no longer able to perform your tasks, you will die painlessly.
>
> Option B: Die painlessly right now.
>
> **2.** Would you rather:

> Option A: Live a long but unhappy life. You live to be 120. You spend most of your time unhappy, but you are not actively suicidal.
>
> Option B: Live a short but happy life. You die ten years from now, in 2026, after being hit by a car. Until then, you do fulfilling work, have happy relationships, and meet with success in most projects.
>
> **3.** Would you prefer:
>
> Option A: A world with 100 trillion trillion sentient beings, all of whom are miserable, but not quite so miserable that they wish they were never born.
>
> Option B: A world with 1 million sentient beings, all of whom are happy and consider their world a utopia.
>
> **4.** What percent certainty of going to Heaven would you need before you would prefer a world with both Heaven and Hell to a world where death ends inevitably in oblivion?

Before checking the results, I had five hypotheses.

First, people would be split in their answers to these questions, with strong feelings on both sides.

Second, answers to all questions would correlate along a general factor of oblivion-preference versus suffering-preference. That is, people who would prefer oblivion to working at McDonalds would also be more likely to prefer a short life of happiness to a long life of unhappiness, et cetera.

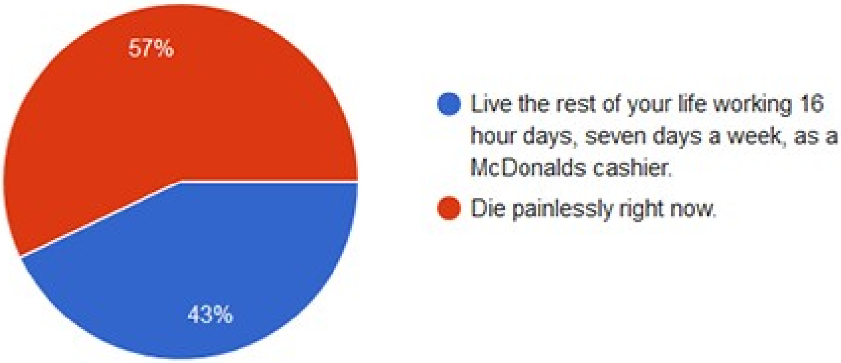
Third, this factor would predict whether somebody endorsed a form of population ethics which promotes creating new people (question 3 is sort of just asking this already, but I also included a more direct question along those lines).

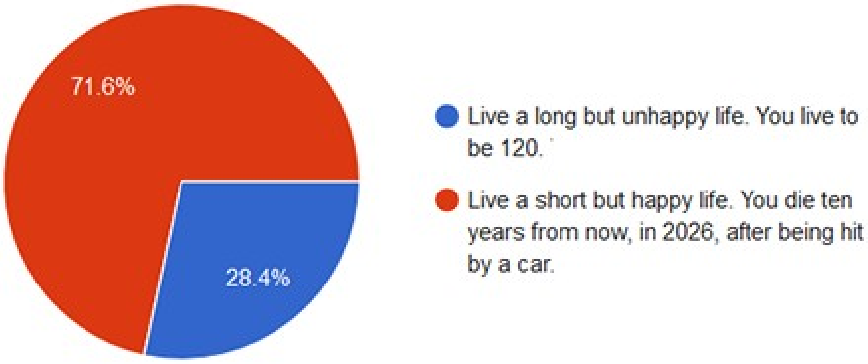
Fourth, this factor would predict some real-world consequences like whether people believed in a right to euthanasia and whether they were signed up for cryonics.

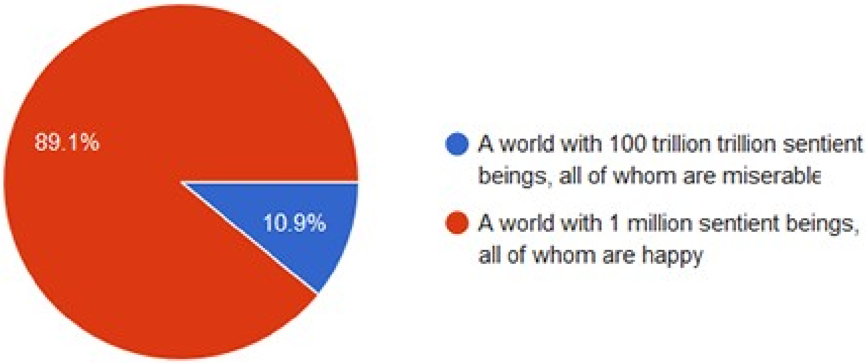
Fifth, happier people would be more likely to prefer suffering over oblivion, because they view life as generally excellent and so oblivion represents more of a sacrifice for them.

1090 people took the survey (social media is a wonderful thing). The survey specifically asked that you not think about other people or the question's effects on them and just choose whatever made you selfishly happier. Several people complained to me that the concept of “generally unhappy, but does not wish they were never born” didn't make sense to them, which I guess is data in and of itself. The results were:

The first hypothesis was confirmed:

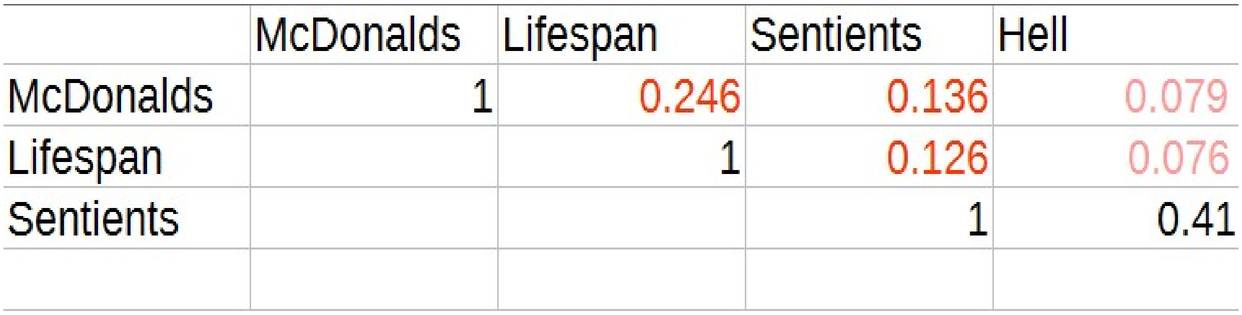






On the fourth question, people answered everything, from demands of certain salvation (n = 320), to being okay with certain damnation (n = 69), and everything in between.

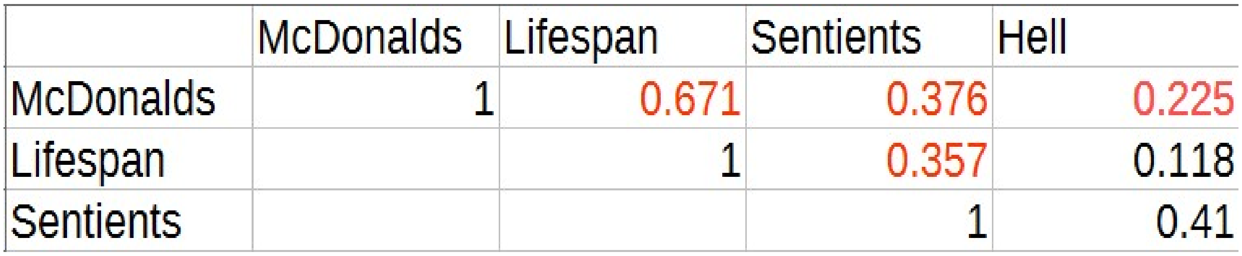
The second hypothesis was weakly confirmed:



	McDonalds	Lifespan	Sentients	Hell
McDonalds	1	0.246	0.136	0.079
Lifespan		1	0.126	0.076
Sentients			1	0.41

Orange results are significant at $p < 0.05$, red results at $p < 0.01$. We did not adjust for multiple comparisons.

There are significant correlations between most of the questions, but they are not very strong. When I limited the analysis to the people who felt most strongly about their answers to the questions, the correlations went up a bit:



	McDonalds	Lifespan	Sentients	Hell
McDonalds	1	0.671	0.376	0.225
Lifespan		1	0.357	0.118
Sentients			1	0.41

The third hypothesis was not supported.

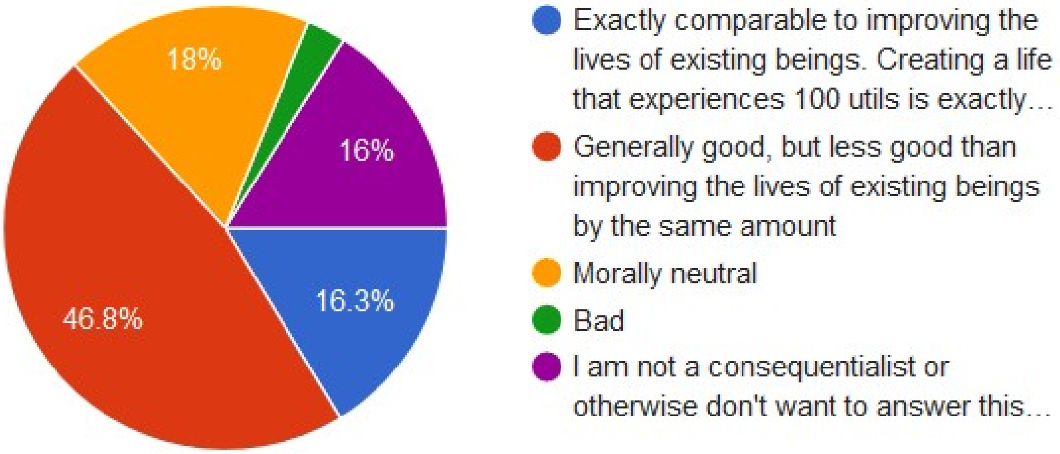
I asked people a pretty direct question about the ethics of creating new people:

> **5.** Creating new sentient beings is:
>
> *Option A:* Exactly comparable to improving the lives of existing beings. Creating a life that experiences 100 utils is exactly as good as improving existing lives 100 utils.
>
> *Option B:* Generally good, but less good than improving the lives of existing beings by the same amount
>
> *Option C:* Morally neutral

Option D: Bad

Option E: I am not a consequentialist or otherwise don't want to answer this question

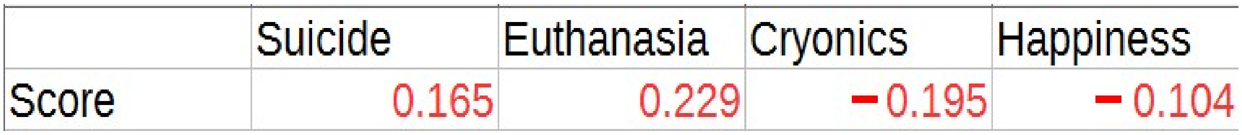
The results were:



The only correlation with any other question was with the one hundred trillion trillion sentients question, which is basically asking the same thing in different words. The correlation with all the other questions was not significant and in fact very close to zero.

The fourth and fifth hypotheses were weakly supported.

I gathered a general score of people's pro-oblivion or pro-suffering bias based on their answers to the first three questions and how strongly they felt about them. It predicted the following:



	Suicide	Euthanasia	Cryonics	Happiness
Score	0.165	0.229	−0.195	−0.104

The specific questions here were:

Whether you believe everybody has a right to commit suicide if they want, including people who are not terminally ill.

Whether you believe people who are terminally ill should have a right to suicide, ie traditional euthanasia.

Whether you are interested in signing up for cryonics.

And whether you consider yourself a happy person.

All were correlated with preference for oblivion over suffering in the expected direction, but not very strongly.

So it looks like people have very different opinions about when to choose death versus suffering, but that these opinions are inconsistent and only weakly driven by broad cross-situation intuitions.

You can see the survey here, but please don’t take it since I’m done getting data. You can download anonymized results here (.xlsx, .csv)

3/4

Related: 1/4, 2/4

[Content warning: psychiatry, suicide. Note that all stories involving patients are mixtures of several different people which have been obfuscated and changed around in order to protect confidentiality. The ethical standard I have heard in this situation is "must be so well disguised that the patient would not recognize himself if he read it" and I have tried to meet that standard – which means that these capture the spirit of situations only. The same is true of some of the other stories here, just in case. Please do not link.]

I.

I'm back at Our Lady Of An Undisclosed Location hospital now as a final-year resident. You wouldn't think a year would make so much difference, but it does.

Identifying residents by their year is easy. The first-years walk around, deer-in-the-headlights look to them, impossible to confuse with anybody except maybe a patient having a panic attack. The middle-year residents are a little more confident. And then the final year residents, leading teams, putting out fires, taking attendings' abuse in stride.

(True story – last week an attending yelled at me for not knowing some minor detail about uraemic encephalopathy. Later I couldn't find the detail he'd mentioned, so I asked for a reference, and he said it had been discovered by one of his friends at the big university hospital where he used to work, but the friend had died before he could publish his findings. I think the attending realized as he was talking that it might have been unreasonable to expect me to know a fact whose discoverer took it to the grave with him, but he didn't apologize.)

It’s only sort of a facade. 99% of things that happen in a hospital are the same things that happened yesterday and the day before, so if you hang around long enough you can learn what to do, or at least which consultant you can call to make it not your problem anymore. On the other hand, Actual Pathology is still a gigantic mystery. I’m not sure this ever changes. One in every X patients with symptoms won’t have any of the things that could possibly be causing those symptoms, won’t respond to any of the treatments that are supposed to cure those symptoms, and you’ll still have family members and hospital administrators demanding that you fix it right now (and in psychiatry, X is probably a single digit number). All you can do is keep up the facade, put your skill at taking attendings’ abuse in stride to good use, and start learning necromancy so you can summon the one big university hospital researcher who studied it but never got a chance to publish their findings.

II.

Two of the most important things I learned during my third year were “Tell me more” and “[awkward silence]”.

“Tell me more,” works for every situation. Part of the problem with psychotherapy is that you’re always expected to have something to say. As a last resort, that thing is “Tell me more”. It sounds like you’re interested. It sounds like you care. And if you’re very lucky, maybe the patient will actually tell you something more, as opposed to their usual plan to stonewall you and hide all possibly useful information.

I saw something on Tumblr the other day which, despite being about a 9-1-1 operator, perfectly sums up being a doctor too:

> my bf has many interesting stories and observations from his new job as a 911 operator
>
> my favorite is how meandering people are, even in the midst of a terrible emergency
>
> they respond to “what is the emergency” with “well, the thing is, four weeks ago–”
>
> and then he’s like “WHAT IS THE EMERGENCY RIGHT NOW”
>
> and they’re like “so what happened this morning was, i said to my wife, i said–”
>
> “WHAT IS CURRENTLY HAPPENING AT THIS MOMENT”
>
> “oh i’m having a heart attack”

And:

> my second favorite is how specific he has to get sometimes
>
> like, “what is your emergency?”
>
> “i’m sitting in a pool of blood.”
>
> “… is it… your blood?”
>
> “yes i think so”
>
> “do you know where it’s coming from?”
>
> “probably the stab wound”
>
> “have you been stabbed?”

“oh yah definitely”

Psychiatry is like this, except it’s all very vague, and your patients are really suggestable, and people are always afraid that if you just ask specific questions like “Are you depressed?” then they’ll say yes to make you happy and won’t talk about how the real problem is their anxiety or something. So instead, the patient says something like “I’m sitting in a pool of blood”, and I say “Tell me more…”. They say “Well, it’s my blood.” I say “Tell me more…”. After repeating this process a couple of times, we finally get to the stabbing, and the patient doesn’t feel like I railroaded over their chance to tell their story.

Or it helps you figure out what’s important to the patient. If someone said “I hate my husband so much,” my natural instinct might be to ask “Why?”. But maybe why isn’t the question the patient cares about. Maybe what she really wants to talk about is how guilty she feels about hating their husband, and if I asked her why then we’d get on a tangent about what the husband is doing that never addresses her real problem. Maybe she’s agonizing every moment about whether or not to divorce him, and losing sleep over it, and coming to me for a sleeping pill. Maybe she’s just hatched a plan to kill him and wants to check it over with me to see if I can find any flaws. In any case I should probably figure out why they hate him *eventually*, but if their real issue is whether or not I approve of their murder plot then we should probably get to that first.

So instead, it’s “I hate my husband so much.” “Tell me more.”

“I’m feeling depressed.” “Tell me more.”

“Sometimes I think life isn’t worth living.” “Tell me more.”

“Listen, if you don’t give me a prescription for Adderall right now I swear to God that I will stab you right here in this office!” “Tell me more.”

This has seeped into my personal life. I was on a date with a girl earlier this year, and whenever she started telling me about her life I would just say “Tell me more”, and it *worked*.

And then there’s [awkward silence]. I learned this one from the psychoanalysts. Nobody likes an awkward silence. If a patient tells you something, and you are awkwardly silent, then the patient will rush to fill the awkward silence with whatever they can think of, which will probably be whatever they were holding back the first time they started talking. You won’t believe how well this one works until you try it. Just stay silent long enough, and the other person will tell you *everything*. It’s better than waterboarding.

The only problem is when two psychiatrists meet. One of my attendings tried to [awkward silence] me at the same time I was trying to [awkward silence] *him*, and we ended up just staring at each other for five minutes until finally I broke down laughing.

“I see you find something funny,” he said. “Tell me more.”

III.

If the patients are cryptic, the doctors are even worse. In a worst case scenario, I’ll be filling in for another doctor – this happens all the time at free clinics, but it happens at least a *little* wherever there are doctors who go on vacation. The documentation will be obscure or missing. The patient’s family is out of contact range. My only information will be the patient in front of me, whose information-transmitting ability is on par with that person from the Tumblr post who took four tries to mention that they’d been stabbed.

So imagine this – a guy from out of state moves in, comes to me without any documentation, and says in a monotone that his only problem is feeling “weird”. All my “tell me mores” and [awkward silences] fail to get him to explain further. I look at his medication list, and he is on a cocktail of supramaximal doses of really old-school antipsychotics that I could not imagine giving anybody unless they were an axe murderer who had killed their last three psychiatrists and I wanted to cool their metaphorical brain temperature to the level of winter on Pluto. Sure enough, the guy is stiff, displays no emotions, and his only hobby is staring at the wall – all exactly what you would expect of somebody who is super-drugged on all of the strongest chemicals known to mankind. I ask him if maybe he’s schizophrenic, or bipolar, or something. He says no, he just feels “weird”.

I know that if I don’t change the medication, he will probably be a zombie like this until such time as somebody else *does* change it, which may be never. But if I *do* change the medication…well, there must be some reason somebody put him on that, and the idea of somebody who *needed* that much medication *not being on it* is too horrible to imagine. Also, I’m only seeing him once, and then he gets transferred to someone else. What do I do?

The maxim is “do what lets you sleep at night”, so I punted. I kept him on his medications and turned him over to the next guy. I just hope the next guy gets my documentation instead of thinking “Dr. Alexander kept him on all this medication…I wonder what he knows that I don’t.”

IV.

“Instead of putting patients on these toxic medications, why don’t you just give them therapy?”

Sometimes I worry I might be the worst person in the world to do psychotherapy. My coping strategy is to not talk about or react to my emotions and wait for them to go away. This usually works. I know this is exactly the opposite of what psychotherapy is supposed to teach, and all I can say is that it works for me and I seem to be pretty psychologically healthy and maybe I am just a mutant.

My relationship strategy is the same. Date really low-conflict, low-drama, agreeable people. If we have a conflict anyway, then agree to disagree and wait for the problem to go away. Apparently this is terrible, and maybe this is why my only really serious relationship only lasted a year or two, but it leaves me with something of an understanding deficit for the people who want to replay every single argument they've ever had with their spouse and figure out exactly what it means about their mental state.

Heck, even polyamory is like this. I can't tell you how many patients I've had come in because their partner is cheating on them, or they worry their partner is cheating on them, or they're cheating on their partner, or their partner worries they're cheating on them, or *something*, and my natural instinct is to just say "Have you considered not worrying about it?" and as usual my natural instinct is terrible. So instead I just say "Tell me more…" and listen to them describe how the possibility of their girlfriend cheating is *rending their heart in two*.

This is even worse in any form of therapy based around investigating childhood traumas. Look, I'm sorry you didn't like your mother, but have you read *The Nurture Assumption*? But of course I can't say that. I just have to play along. And then somebody expects me to come up with something to heal the maternal trauma that I'm not even sure people really have, and then if I do come up with something it feels like a clever fake.

Cognitive behavioral therapy is a little better, because it tends to be pretty common sense techniques that any reasonable person would agree with. The problem is, it’s pretty common sense techniques that any reasonable person would agree with. I think that I and most of my friends would respond to the average CBT session with a sort of anger at being condescended at, combined with annoyance at the therapist for wasting our time with obvious things. “My job sucks”. “Well, have you considered making a list of good and bad things about your job?” “Yes, that was the process by which I determined it sucks. How much am I paying you again?”

Most of the time I do therapy, I feel cringeworthy, unnatural. I feel like a fraud, even when (according to the supervisors watching me) I’m doing it exactly right. I feel like I’m responding to people in fake, silly ways, like they’re coming to me with problems from the depth of their being and I’m giving them facile non-answers. It doesn’t even help that most of them get better anyway. In a way, that just makes it worse. How *dare* you get better after me telling you stupid things I feel embarrassed to say? That’s just going to encourage people to make me keep doing that!

V.

I nevertheless hold a special place of annoyance in my heart for psychoanalysis/psychodynamic therapy.

The attending who started my training in psychodynamic therapy (I got a new psychodynamic supervisor recently who I don’t know too well yet) was an elderly doctor in an office attached to the clinic, full of creepy modern art statues. He is convinced that patients’ lives revolve around their therapy and their therapists. I know that in moderation this is the idea of “transference”, a genuine and important tenet of the therapy style. My attending does not do it in moderation.

My patient will say something like “My best friend moved away and now I am sad”, I will think “That sounds straightforward, better bring it up to my attending and see how he wants me to deal with this.” My attending will invariably say “What your patient means is that he’s afraid of losing *you*, his therapist.”

I will say “No, I’m pretty sure he actually lost his best friend. He told me all about how they’d been together since middle school, but now he moved away to take a job in Texas, and then he broke down crying.”

Then my attending will get really angry and tell me that if I’m just going to take everything my patient tells me exactly literally, then I shouldn’t be in psychiatry, because a *monkey* could listen to a patient say he was sad about losing his best friend and conclude he was sad about losing his best friend, and my duty as a *trained professional* is to be able to see beneath that to the *true* thought which my patient is trying to express. Which is always, 100% of the time, about how much the patient cares about psychodynamic therapy and wants to continue doing it.

Even worse, he wants me to do this to the patient. When the patient says “I’m really upset about losing my best friend”, I’m supposed to answer “Are you sure this isn’t about how you’re worried I’m sort of like a friend to you and one day you’ll lose *me*?” If talking about relationships and cognitive therapy makes me cringe, this super quadruple makes me cringe.

Still, I have to do it, because my attending grades me and if I don’t pass psychodynamic therapy I don’t get to graduate. So I do it, and then my attending declares he was right all along based on extremely strained interpretations of whatever happens next. Like, if the patient misses their next appointment, he’ll say “I see your patient missed their next appointment. That means they’re having a defensive reaction to the fact that you called them out on their being afraid of you leaving them. And to think that you told me you weren’t sure that was true! This just shows how much you still have to learn about psychodynamics. I certainly hope that after *this* you won’t keep questioning me every time I try to help you.”

But when I leave for good, I’m getting him a present, and it’s going to be a copy of *The Nurture Assumption*. Heck, maybe I’ll give that to *all* the psychoanalysts I know.

VI.

It’s kind of morbid to feel smug about your patients not attempting suicide, but I guess I am a kind of morbid person.

The doctor down the hall from me had one of his patients attempt suicide in October. Then another doctor I knew had two of his patients attempt suicide in the same week in January. And I was really sympathetic and tried to comfort them, but I also had a part of my mind thinking “Hey, I haven’t had *any* of my patients attempt suicide yet, this is pretty good.”

March. April. May. My coworkers told me their stories, but I kept my secret morbid goal – I was going to go the entire year without any of my patients trying to kill themselves. I mean, on one hand this sounds like a pretty minimal standard. On the other, when you’re taking care of like a hundred mentally ill people, many of whom have really bad depression and a history of past suicide attempts, it’s not exactly trivial.

I got the call just a few weeks ago. The patient was a former heroin addict who had been clean for a long time. He slipped, took heroin, felt terrible, and stabbed himself in the heart.

Luckily the heart is a little to the right of where most people think it is. Stabbing yourself in the lung isn’t great either, but he was a young healthy man and he could take it. He went to the hospital, they patched it up a little, and he was fine. He said it was the best thing that had ever happened to him and now he knew how low he could get and he was going to stay clean forever and today was the first day of the rest of his life.

A lot of things in psychiatry are reverse lotteries. In the regular lottery, you pay a constant small cost for the possibility of a stupendous benefit. In the reverse lottery, you get a constant small benefit at the risk of a stupendous cost. Lots of things are like this. If you give someone a powerful medication, then they’ll definitely recover, but there’s a risk you’ll have a catastrophic side effect. If you let a severely ill patient leave your office when they promise they’re okay, then you definitely save them the trauma of an involuntary hospitalization, but there’s a risk they’ll do something disastrous. If you don’t check someone’s vitals every time you see them then it definitely makes the appointment quicker and smoother, but there’s a risk you’ll miss something really bad.

It’s really easy to fall into playing reverse lotteries. I think almost everybody does it to a degree. The usual pattern is to play some of them tentatively, do more and more of them as you reap the benefits and nothing goes wrong, then boom, close call, and you resolve never to do anything like that again and you're going to do a full half-hour neurological examination on everybody who comes into your office including random passers-by who just want to use the bathroom.

After my patient stabbed himself I spent a week totally neurotic, looking over every aspect of his case – could I have checked up on his Narcotics Anonymous meeting attendance more frequently? Maybe if I'd given him a long lecture every appointment about how heroin was definitely still bad, that would have changed something? Maybe if I hadn't forgotten to check his blood pressure that one time…? In the end, I decided I had done a pretty okay job on *that* case – which just made me more acutely aware of all of the reverse lotteries I was playing on everybody *else*. Now I'm a little bit paranoid. Maybe that's temporary. Maybe it's permanent. I don't know. The DSM-V says you have to have it six months before you can give yourself a schizophrenia diagnosis, so there's that.

I am getting good at dealing with annoying attendings, meandering patients, unreasonable requests, and silly bureaucracy. Actual Pathology remains scary, mysterious, and really hard to predict. Hopefully that's what fourth year is for.

Pushing And Pulling Goals

This is a distinction I’ve always found helpful.

A pulling goal is when you want to achieve something, so you come up with a plan and a structure. For example, you want to cure cancer, so you become a biologist and set up a lab and do cancer research. Or you want to get rich, so you go to business school and send out your resume.

A pushing goal is when you have a plan and a structure, and you’re trying to figure out what to do with it. For example, you’re studying biology in college, your professor says you need to do a research project to graduate, and so you start looking for research to do. You already know the plan – you’re going to get books, maybe use a lab, do biology-ish things, and end up with a finished report which is twenty pages double-spaced. All you need to figure out is what you’re going to select as the nominal point of the activity. There’s something perversely backwards about this – most people would expect that the point of a research project is to research some topic in particular. But from your perspective the actual subject you’re researching is almost beside the point. The point is to have a twenty page double-spaced report on *something*.

School and business are obvious ways to end up with pushing goals, but not every pushing goal is about satisfying somebody else’s requirements. I remember in college some friends set up an Atheist Club. There was a Christian Club, and a Buddhist Club, so why shouldn’t the atheists get a club too? So they wrote the charter, they set a meeting time, and then we realized none of us knew what exactly the Atheist Club was supposed to do. The Christian Club prayed and did Bible study; the Buddhist club meditated, the atheist club…sat around and tried to brainstorm Atheist Club activities. Occasionally we came up with some, like watching movies relevant to atheism, or having speakers come in and talk about how creationism was really bad. But we weren't doing this because we really wanted to watch movies relevant to atheism, or because we were interested in what speakers had to say about creationism. We were doing this because we'd started an Atheist Club and now we had to come up with a purpose for it.

Sometimes on Reddit's /r/writing I see people asking "How do you come up with ideas for things to write about?" and I feel a sort of horror. So you want to write a novel, but…you don't have anything to write about? And you just sit there thinking "Maybe it should be about romance…no, war…no, the ennui of the working classes…or maybe hobbits." I can understand this in *theory* – you want to be A Writer – but it still weirds me out.

You may have noticed I don't really like pushing goals. Part of it is an irrational intuition that they're dishonest in some way that's hard to explain. It usually ends up with me trying to figure out what to do my biology research project on, and I think "well, I can't think of anything I really want to research, so maybe I should just do whatever is easiest". But if I do whatever is easiest, I feel really bad, and worry maybe I have some kind of obligation to research something important that I care about. So I get my brain tangled up trying to figure out how much easiness I can get away with, then feeling bad for asking the question, *then* trying to come up with something important I honestly want to do, which doesn't exist since I wasn't doing a biology research project the month *before* my professor assigned it to me and so clearly I am only doing it to satisfy the requirement.

Another part of it is that it's often a sign something has gone wrong somewhere. In the example of the Atheist Club, that thing *might* have been starting the club in the first place. But assuming that we genuinely want to start the club, then the presence of a pushing goal means we don't understand *why* we wanted to start the club. If we wanted to start it because we wanted to hang out with other atheists, then that offers a blueprint for a solution to the problem – instead of planning all these movies and speakers, we should just hang out. If we did it because we thought it was important for atheism to be more visible on campus, then again, that offers a blueprint for a solution – spend our sessions trying to improve atheism's campus visibility. If we just sit there saying "I guess we have an Atheist Club now, better think of something to do at meetings", then it seems like something important hasn't been fully examined.

The third part of it is that things done for push goals usually suck. Maybe this isn't a human universal – my go-to example is Edgar Allen Poe deciding to write a creepy poem and coming up with *The Raven* from first principles – but it's true for *me*. If I have to write a report on a topic I don't care about, then even if I'm really trying to do a good job, it's not going to be as good as something I actually want to write about. Sometimes I try to solve this by making lists of things I want to pull, then using them when the appropriate pushing situation comes up. For example, when I knew I would be assigned research projects and writing assignments on a regular basis, whenever I thought of something I wanted to research or write, I wrote it down, then consulted the list when I needed it. I have a similar list of interesting things to work into stories. This is one reason I'm not interested in journalism – I worry that if I have to produce specific articles on specific things within a time frame, they'll probably suck.

Things Probably Matter

A while back when I wrote about how China's economic development might not have increased happiness there much, Scott Sumner wrote a really interesting response, Does Anything Matter?

He points out that it's too easy to make this about exotic far-off Chinese. Much the same phenomenon occurs closer to home:

> If nothing really matters in China, if even overcoming horrible problems doesn't make the Chinese better off, then what's the use of favoring or opposing any public policy? After all, America also shows no rise in average happiness since the 1950s, despite:
>
> 1. A big rise in real wages.
> 2. Environmental clean-up (including lead–does Flint matter?)
> 3. Civil rights for African Americans
> 4. Feminism, gay rights.
> 5. Dentists now use Novocain (My childhood cavities were filled without it)
> 6. 1000 channels in glorious widescreen HDTV
> 7. Blogs
>
> I could go on and on. And yet, if the surveys are to be believed, we are no happier than before. And I think it's very possible that we are in fact no happier than before, that there's a sort of law of the conservation of happiness. As I walk down the street, grown-ups don't seem any happier than the grown-ups I recall as a kid. Does that mean that all of those wonderful societal achievements since 1950 were absolutely worthless?

> But there are exceptions. I recall reading that surveys showed a rise in European happiness in the decades after WWII, and Scott reports that happiness is currently very low in Iraq and Syria. So that suggests that current conditions do matter.
>
> The following hypothesis will sound really ad hoc, but matches the way a lot of people I know talk about their lives. Suppose people's happiness is normally calibrated around the sort of lifestyle that they view as "normal." As America got richer after 1950, it all seemed very normal, so people didn't report more happiness. Ditto for China during the boom years. Everyone around you was also doing better, so you started thinking about how you were doing relative to your neighbors. But Germans walking through the rubble of Berlin in 1948, or Syrians doing so today in Aleppo, do see their plight as abnormal. They remember a time before the war. So they report less happiness than during normal times.

The obvious retort is – modern Chinese grew up when China was very poor. Why didn't they calibrate themselves to poverty, such that sudden wealth seems good? What's the difference between a Chinese person going from poverty to wealth, versus a Syrian going from stability to chaos? Might it be a shorter time course? A sudden shock is noticeable, a gradual thirty-year improvement in living standards isn't?

Probably not. There seem to be a lot of cases where happiness of large groups *does* change gradually in response to social trends less dramatic than a world war.

First, consider African-Americans. The New York Times calls the increase in black happiness over the past forty years "one of the most dramatic gains in the happiness data that you'll see". This is not just about poverty; in 1970, blacks who earned more than 75% of whites were only in the tenth percentile of white happiness. Today, those blacks would be in the fiftieth percentile; they're still doing worse than would be expected based on income, but not nearly as *much* worse. This is a very sensible and predictable thing to find. Black people face a lot less racism and discrimination today than in 1970 [citation needed], so assuming that was really unpleasant we shouldn't be surprised that they're happier. But notice that this is a time course very similar to the rise of China! It doesn't look like black people picked a happiness level to calibrate on and then never bothered to adjust. It looks like they adjusted exactly like we would expect them to, even over the course of a multi-decade change.

Second, consider women. In 1970, US women were generally happier than US men. Today, the reverse is true. There seems to be a general pattern around the world of women being happier than men in traditional societies and less happy than men in modern societies (though see Language Log for a contrary perspective). I don't think of this as a weird paradox. It seems perfectly reasonable to me that having to work outside the home makes people less happy, getting to spend time with their family makes them more happy, and having to work outside the home but also being expected to take care of your family at the same time makes them least happy of all. In any case, the point is that the numbers are changing. Men and women aren't just fixating on some level of happiness and staying there, they're altering their happiness level based on real trends, just like African-Americans did (but apparently unlike Chinese).

Third, I was finally able to find a paper that had really good data on change in happiness in different countries, and it supports the idea that happiness can change significantly on a countrywide level.

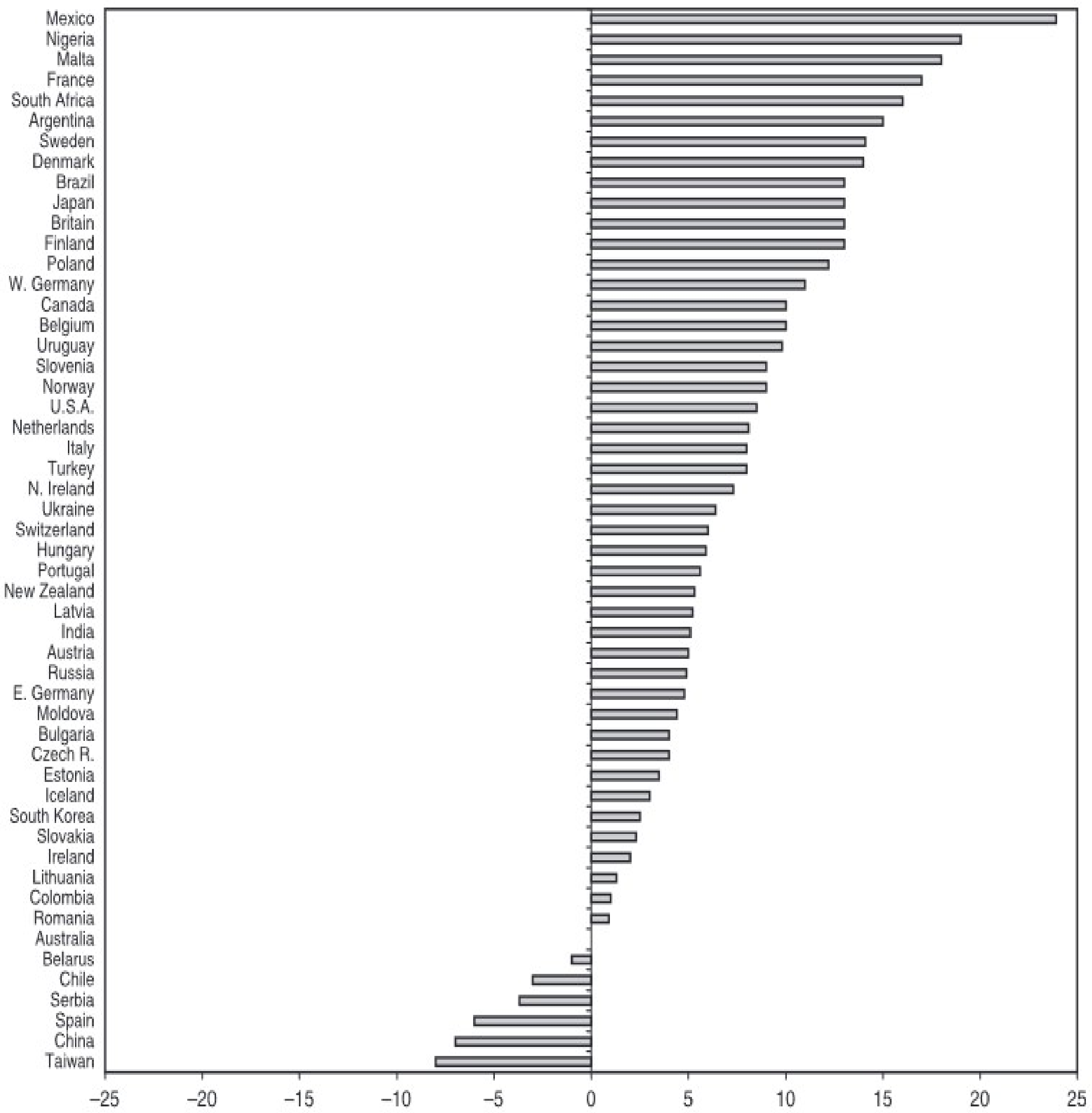


Fig. 6. Change in percentage of those saying they are very happy from earliest to latest survey for all countries with a substantial time series.

This is change in happiness in a bunch of countries between about 1990 and 2010 (the years were slightly different in each country). There are other graphs for related concepts like life satisfaction and subjective well-being that look about the same.

The most striking finding is that most countries got happier between those two years – sometimes a lot happier. In Mexico, the percent of people saying they were very happy increased by 25 percentage points!

Just eyeballing the graph, there’s not an obvious relationship between happiness and economic growth – China is still near the bottom like we talked about before, and France – a country that’s been First World since forever – is near the top. Even Japan, which is famous for its decades of stagnation, has done pretty well. But the authors tell us that after doing their statistical analyses, there *is* a strong relationship with economic growth. Okay, I guess.

They also say there’s a dramatic relationship with freedom and democracy. Mexico, the top country on the graph, went from a relatively closed to a relatively democratic government during this time. South Africa, number five, went from apartheid to no apartheid. Some of the ex-Communist countries like Poland and Ukraine also look pretty good here. On the other hand, other ex-Communist countries like Lithuania and Estonia are near the bottom. I wonder if this has to do with cutoff points – since every country started at a slightly different time, maybe they began sampling Poland during the worst parts of Soviet dictatorship and got Lithuania right in the first euphoria of independence? I don’t know. It all seems very noisy.

They also mention that the United States’ supposedly level happiness is kind of a misunderstanding. People say things like “Happiness in the US has been flat from 1950 to today”, but in fact it declined from 1950 to 1979 and increased from 1980 to today. They attribute this to the 1950s being unusually happy; then the 60s and 70s being unusually conflict-prone, and the Reagan Revolution and Clinton years were back to being optimistic. They don’t have data that stretches too long after that.

(This is pretty neat for Reagan and Clinton. When I die, I’ll consider my life a success if people attribute a spike on national happiness graphs to my influence.)

So apparently population happiness levels do change in response to relevant social changes, even on multi-decade timescales. Which brings us back to asking – what’s up with China?

The graph above shows India as doing okay – not great, but okay. But a similar graph on subjective well-being – which should be another way of looking at the same thing – shows India as doing pretty poorly, right down there with China – even though its GDP per capita quadrupled during the period of study.

I see a lot of conflicting perspectives about whether economic growth increases national happiness. It *may*, but the effect isn’t as big as you’d expect, and is usually overpowered by other factors. Maybe it isn’t even direct, but has something to do with development increasing democracy, liberalism, rule of law, and stability. China got the development, but its happiness genuinely didn’t increase because of country-specific factors that have something to do with how it developed (inequality? pollution? authoritarianism?).

This matches the race and gender data. Blacks saw a big happiness boost during a time when their feeling of freedom (but not their income) increased relative to whites. Women saw a small happiness drop during a time when their income (but not their feeling of freedom) increased relative to men.

So it looks like happiness *can* change. It just didn’t change in China over the past thirty years. The apparent paradox of improving economic situation and stable/decreasing happiness is genuinely paradoxical. Intangibles are probably just way more important than money, even amounts of money big enough to raise whole countries out of poverty.

How The West Was Won

I.

Someone recently linked me to Bryan Caplan's post A Hardy Weed: How Traditionalists Underestimate Western Civ. He argues that "western civilization"'s supposed defenders don't give it enough credit. They're always worrying about it being threatened by Islam or China or Degeneracy or whatever, but in fact western civilization can not only hold its own against these threats but actively outcompetes them:

> The fragility thesis is flat wrong. There is absolutely no reason to think that Western civilization is more fragile than Asian civilization, Islamic civilization, or any other prominent rivals. At minimum, Western civilization can and does perpetuate itself the standard way: sheer conformity and status quo bias.
>
> But saying that Western civilization is no more fragile than other cultures is a gross understatement. The truth is that Western civilization is taking over the globe. In virtually any fair fight, it steadily triumphs. Why? Because, as fans of Western civ ought to know, Western civ is better. Given a choice, young people choose Western consumerism, gender norms, and entertainment. Anti-Western governments from Beijing to Tehran know this this to be true: Without draconian censorship and social regulation, "Westoxification" will win.
>
> A big part of the West's strength, I hasten to add, is its openness to awesomeness. When it encounters competing cultures, it gleefully identifies competitors'

> best traits – then adopts them as its own. By the time Western culture commands the globe, it will have appropriated the best features of Asian and Islamic culture. Even its nominal detractors will be Westernized in all but name. Picture how contemporary Christian fundamentalists' consumerism and gender roles would have horrified Luther or Calvin. Western civ is a good winner. It doesn't demand total surrender. It doesn't make fans of competing cultures formally recant their errors. It just tempts them in a hundred different ways until they tacitly convert.
>
> Traditionalists' laments for Western civilization deeply puzzle me. Yes, it's easy to dwell on setbacks. In a world of seven billion people, you can't expect Western culture to win everywhere everyday. But do traditionalists seriously believe that freshman Western civ classes are the wall standing between us and barbarism? Have they really failed to notice the fact that Western civilization flourishes all over the globe, even when hostile governments fight it tooth and nail? It is time for the friends of Western civilization to learn a lesson from its enemies: Western civ is a hardy weed. Given half a chance, it survives, spreads, and conquers. Peacefully.

I worry that Caplan is eliding the important summoner/demon distinction. This is an easy distinction to miss, since demons often kill their summoners and wear their skin. But in this case, he's become hopelessly confused without it.

I am pretty sure there was, at one point, such a thing as western civilization. I think it included things like dancing around maypoles and copying Latin manuscripts. At some point Thor might have been involved. That civilization is dead. It summoned an alien entity from beyond the void which devoured its summoner and is proceeding to eat the rest of the world.

An analogy: naturopaths like to use the term “western medicine” to refer to the evidence-based medicine of drugs and surgeries you would get at your local hospital. They contrast this with traditional Chinese medicine and Ayurvedic medicine, which it has somewhat replaced, apparently a symptom of the “westernization” of Chinese and Indian societies.

But “western medicine” is just medicine that works. It happens to be western because the West had a technological head start, and so discovered most of the medicine that works first. But there’s nothing *culturally* western about it; there’s nothing Christian or Greco-Roman about using penicillin to deal with a bacterial infection. Indeed, “western medicine” replaced the traditional medicine of Europe – Hippocrates’ four humors – before it started threatening the traditional medicines of China or India. So-called “western medicine” is an inhuman perfect construct from beyond the void, summoned by Westerners, which ate traditional Western medicine first and is now proceeding to eat the rest of the world.

“Western culture” is no more related to the geographical west than western medicine. People who complain about western culture taking over their country always manage to bring up Coca-Cola. But in what sense is Coca-Cola culturally western? It’s an Ethiopian bean mixed with a Colombian leaf mixed with carbonated water and lots and lots of sugar. An American was the first person to discover that this combination tasted really good – our technological/economic head start ensured that. But in a world where America never existed, eventually some Japanese or Arabian chemist would have found that sugar-filled fizzy drinks were really tasty. It was a discovery waiting to be plucked out of the void, like penicillin. America summoned it but did not create it. If western medicine is just medicine that works, soda pop is just refreshment that works.

The same is true of more intellectual “products”. Caplan notes that foreigners consume western gender norms, but these certainly aren’t gender norms that would have been recognizable to Cicero, St. Augustine, Henry VIII, or even Voltaire. They’re gender norms that sprung up in the aftermath of the Industrial Revolution and its turbulent intermixing of the domestic and public economies. They arose because they *worked*. The West was the first region to industrialize and realize those were the gender norms that worked for industrial societies, and as China and Arabia industrialize they’re going to find the same thing.

Caplan writes:

> A big part of the West’s strength, I hasten to add, is its openness to awesomeness. When it encounters competing cultures, it gleefully identifies competitors’ best traits – then adopts them as its own. By the time Western culture commands the globe, it will have appropriated the best features of Asian and Islamic culture.

Certainly he’s pointing at a real phenomenon – sushi has spread almost as rapidly as Coke. But in what sense has sushi been “westernized”? Yes, Europe has adopted sushi. But so have China, India, and Africa. Sushi is another refreshment that works, a crack in the narrative that what’s going on is “westernization” in any meaningful sense.

Here’s what I think is going on. Maybe every culture is the gradual accumulation of useful environmental adaptations combined with random memetic drift. But this is usually a slow process with plenty of room for everybody to adjust and local peculiarities to seep in. The Industrial Revolution caused such rapid change that the process become qualitatively different, a frantic search for better adaptations to an environment that was itself changing almost as fast as people could understand it.

The Industrial Revolution also changed the way culture was spatially distributed. When the fastest mode of transportation is the horse, and the postal system is frequently ambushed by Huns, almost all culture is local culture. England develops a culture, France develops a culture, Spain develops a culture. Geographic, language, and political barriers keep these from intermixing too much. Add rapid communication – even at the level of a good postal service – and the equation begins to change. In the 17th century, philosophers were remarking (in Latin, the universal language!) about how Descartes from France had more in common with Leibniz from Germany than either of them did with the average Frenchman or German. Nowadays I certainly have more in common with SSC readers in Finland than I do with my next-door neighbor whom I’ve never met.

Improved trade and communication networks created a rapid flow of ideas from one big commercial center to another. Things that worked – western medicine, Coca-Cola, egalitarian gender norms, sushi – spread along the trade networks and started outcompeting things that didn’t. It happened in the west first, but not in any kind of a black-and-white way. Places were inducted into the universal culture in proportion to their participation in global trade; Shanghai was infected before West Kerry; Dubai is further gone than Alabama. The great financial capitals became a single cultural region in the same way that “England” or “France” had been a cultural region in the olden times, gradually converging on more and more ideas that worked in their new economic situation.

Let me say again that this universal culture, though it started in the West, was western only in the most cosmetic ways. If China or the Caliphate had industrialized first, they would have been the ones who developed it, and it would have been much the same. The new sodas and medicines and gender norms invented in Beijing or Baghdad would have spread throughout the world, and they would have looked very familiar. The best way to industrialize is the best way to industrialize.

II.

Something Caplan was pointing towards but never really said outright: universal culture is by definition the only culture that can survive without censorship.

He writes in his post:

> The truth is that Western civilization is taking over the globe. In virtually any fair fight, it steadily triumphs. Why? Because, as fans of Western civ ought to know, Western civ is better. Given a choice, young people choose Western consumerism, gender norms, and entertainment. Anti-Western governments from Beijing to Tehran know this this to be true: Without draconian censorship and social regulation, “Westoxification” will win.

Universal culture is the collection of the most competitive ideas and products. Coca-Cola spreads because it tastes better than whatever people were drinking before.

Egalitarian gender norms spread because they're more popular and likeable than their predecessors. If there was something that outcompeted Coca-Cola, then that would be the official soda of universal culture and Coca-Cola would be consigned to the scrapheap of history.

The only reason universal culture doesn't outcompete everything else instantly and achieve fixation around the globe is barriers to communication. Some of those barriers are natural – Tibet survived universalization for a long time because nobody could get to it. Sometimes the barrier is time – universal culture can't assimilate every little hill and valley instantly. Other times there are no natural barriers, and then your choice is to either accept assimilation into universal culture, or put up some form of censorship.

Imagine that Tibet wants to protect its traditional drink of yak's milk. The Dalai Lama requests that everyone continue to drink yak's milk. But Coca-Cola tastes much better than yak's milk, and everyone knows this. So it becomes a coordination problem: even if individual Tibetans would prefer that their neighbors all drink yak's milk to preserve the culture, *they* want to drink Coca-Cola. The only way yak's milk stays popular is if the Dalai Lama bans Coca-Cola from the country.

But westerners aren't banning yak's milk to "protect" their cultures. They don't have to. Universal culture is high-entropy; it's already in its ground state and will survive and spread without help. All other cultures are low-entropy; they survive only if someone keeps pushing energy into the system to protect them. It could be the Dalai Lama banning Coca-Cola. It could be the Académie Française removing English words from the language. It could be the secret police killing anyone who speaks out against Comrade Stalin. But if you want anything other than universal culture, you better either be surrounded by some *very* high mountains, or be willing to get your hands dirty.

There’s one more sense in which universal culture is high-entropy; I think it might be the only culture that can really survive high levels of immigration.

I’ve been wondering for a long time – how come groups that want to protect their traditional cultures worry about immigration? After all, San Francisco is frequently said to have a thriving gay culture. There’s a strong Hasidic Jewish culture in New York City. Everyone agrees that the US has something called “black culture”, although there’s debate over exactly what it entails. But only 6% of San Francisco is gay. Only 1% of New Yorkers are Hasidim. Only about 11% of Americans are black. So these groups have all managed to maintain strong cultures while being vastly outnumbered by people who are different from them.

So why is *anyone* concerned about immigration threatening their culture? Suppose that Tibet was utterly overwhelmed by immigrants, tens of millions of them. No matter how many people you import, Tibetan people couldn’t possibly get more outnumbered in their own country than gays, Hasidim, and blacks already are. But those groups hold on to their cultures just fine. Wouldn’t we expect Tibetans (or Americans, or English people) to do the same?

I’m still not totally sure about the answer to this one, but once again I think it makes more sense when we realize that Tibet is competing not against Western culture, but against universal culture.

And here, universal culture is going to win, simply because it’s designed to deal with diverse multicultural environments. Remember, different strategies can succeed in different equilibria. In a world full of auto-cooperators, defect-bot hits the jackpot. In a world full of tit-for-tat-players, defect-bot crashes and burns. Likewise, in a world where everybody else follows Tibetan culture, Tibetan culture may do very well. In a world where there are lots of different cultures all mixed together, Tibetan culture might not have any idea what to do.

(one more hypothetical, to clarify what I'm talking about – imagine a culture where the color of someone's clothes tells you a lot of things about them – for example, anyone wearing red is a prostitute. This may work well as long as everyone follows the culture. If you mix it 50-50 with another culture that doesn't have this norm, then things go downhill quickly; you proposition a lady wearing red, only to get pepper sprayed in the eye. Eventually the first culture gives up and stops trying to communicate messages through clothing color.)

I think universal culture has done a really good job adapting to this through a strategy of social atomization; everybody does their own thing in their own home, and the community exists to protect them and perform some lowest common denominator functions that everyone can agree on. This is a *really* good way to run a multicultural society without causing any conflict, but it requires a very specific set of cultural norms and social technologies to work properly, and only universal culture has developed these enough to pull it off.

Because universal culture is better at dealing with multicultural societies, the more immigrants there are, the more likely everyone will just default to universal culture in public spaces. And eventually the public space will creep further and further until universal culture becomes the norm.

If you don’t understand the difference between western culture and universal culture, this looks like the immigrants assimilating – “Oh, before these people were Chinese people behaving in their foreign Chinese way, but now they’re Westerners just like us.” Once you make the distinction, it looks like *both* Chinese people *and* traditional Americans assimilating into universal culture in order to share a common ground – with this being invisible to people who are already assimilated into universal culture, to whom it just looks “normal”.

III.

I stress these points because the incorrect model of “foreign cultures being Westernized” casts Western culture as the aggressor, whereas the model of “every culture is being universalized” finds Western culture to be as much a victim as anywhere else. Coca-Cola might have replaced traditional yak’s milk in Mongolia, but it also replaced traditional apple cider in America. A Hopi Indian saddened that her children no longer know the old ritual dances differs little from a Southern Baptist incensed that her kids no longer go to church. Universal values have triumphed over both.

Our society is generally in favor of small, far-away, or exotic groups trying to maintain their culture. We think it’s great that the Hopi are trying to get the next generation to participate in the traditional dances. We support the Tibetans’ attempt to maintain their culture in the face of pressure from China. We promote black culture, gay culture, et cetera. We think of it as a tragedy when the dominant culture manages to take over and destroy one of these smaller cultures. For example, when white American educators taught Native American children to identify with white American culture and ignore the old ways, that was inappropriate and in some senses “genocidal” if the aim was to destroy Native Americans as a separate people. We get excited by the story of Bhutan, the tiny Himalayan kingdom trying to preserve its natural and human environment and prevent its own McDonaldization. We tend to be especially upset when the destruction of cultures happens in the context of colonialism, ie a large and powerful country trying to take over and eliminate the culture of a smaller country. Some examples include the English in Ireland, the English in India, the English in Africa, and basically the English anywhere.

One of the most common justifications for colonialism is that a more advanced and enlightened society is taking over an evil and oppressive society. For example, when China invaded Tibet, they said that this was because Tibet was a feudal hellhole where most of the people were living in abject slavery and where people who protested the rule of the lamas were punished by having their eyes gouged out (true!). They declared the anniversary of their conquest “Serfs Emancipation Day” and force the Tibetans to celebrate it every year. They say that anyone who opposes the Chinese, supports the Dalai Lama, or flies the old Tibetan flag is allied with the old feudal lords and wants to celebrate a culture based around serfdom and oppression.

But opponents of colonialism tend to believe that cultures are valuable and need to be protected in and of themselves. This is true even if the culture is very poor, if the culture consists of people who aren’t very well-educated by Western standards, even if they believe in religions that we think are stupid, even if those cultures have unsavory histories, et cetera. We tend to allow such cultures to resist outside influences, and we even celebrate such resistance. If anybody were to say that, for example, Native Americans are poor and ignorant, have a dumb religion with all sorts of unprovable “spirits”, used to be involved in a lot of killing and raiding and slave-taking – and so we need to burn down their culture and raise their children in our own superior culture – that person would be incredibly racist and they would not be worth listening to. We celebrate when cultures choose preservation of their traditional lifestyles over mere economic growth, like Bhutan's gross national happiness program.

This is true in every case except with the cultures we consider our outgroups – in the US, white Southern fundamentalist Christian Republicans; in the UK, white rural working-class leave voters. In both cases, their ignorance is treated as worthy of mockery, their religion is treated as stupidity and failure to understand science, their poverty makes them "trailer trash", their rejection of economic-growth-at-all-costs means they are too stupid to understand the stakes, and their desire to protect their obviously inferior culture makes them xenophobic and racist. Although we laugh at the Chinese claim that the only reason a Tibetan could identify with their own culture and want to fly its flag is because they support serfdom and eye-gouging, we solemnly nod along with our own culture's claim that the only reason a Southerner could identify with their own culture and want to fly its flag is because they support racism and slavery.

(one question I got on the post linked above was why its description of American tribes seemed to fit other countries so well. I think the answer is because most countries' politics are centered around the conflict between more-universalized and less-universalized segments of the population.)

We could even look at this as a form of colonialism – if Brexit supporters and opponents lived on two different islands and had different colored skin, then people in London saying things like “These people are so butthurt that we’re destroying their so-called ‘culture’, but they’re really just a bunch of ignorant rubes, and they don’t realize they need us elites to keep their country running, so screw them,” would sound a lot more sinister. The insistence that they tolerate unwanted immigration into their lands would look a lot like how China is trying to destroy Tibet by exporting millions of people to it in the hopes they will eventually outnumber the recalcitrant native Tibetans (if you don’t believe me, believe the Dalai Lama, who apparently has the same perspective). The claim that they’re confused bout their own economic self-interest would give way to discussions of Bhutan style “gross national happiness”.

(I get accused of being crypto-conservative around here every so often, but I think I’m just taking my anti-colonialism position to its logical conclusion. A liberal getting upset about how other liberals are treating conservatives, doesn’t become conservative himself, any more than an American getting upset about how other Americans treat Iraqis becomes an Iraqi.)

And I worry that confusing “universal culture” with “Western culture” legitimizes this weird double standard. If universal culture and Western culture are the same thing, then Western culture doesn’t need protection – as Caplan points out, it’s the giant unstoppable wave of progress sweeping over everything else. Or maybe it doesn’t *deserve* protection – after all, it’s the colonialist ideology that tried to destroy local cultures and set itself up as supreme. If Western culture is already super-strong and has a history of trying to take over everywhere else, then surely advocating “protecting Western culture” must be a code phrase for something more sinister. We can sympathize with foreign cultures like the Tibetans who are actually under threat, but sympathizing with any Western culture in any way would just be legitimizing aggression.

But I would argue that it’s universal culture which is the giant unstoppable wave of progress, and that it was universal culture that was responsible for colonizing other cultures and replacing them with itself. And universal culture’s continuing attempts to subjugate the last unassimilated remnants of traditional western culture are just part of this trend.

IV.

I am mostly just on the side of consistency. After that I have no idea what to do.

One argument is that we should consistently support traditional cultures’ attempts to defend themselves against universal culture. Support the Native Americans’ ability to practice their old ways, support traditional Siberians trying to return to their shamanistic roots, support Australian Aborigines’ rights to continue the old rituals, support Tibetans’ rights to practice Vajrayana Buddhism, and support rural British people trying to protect Ye Olde England from the changes associated with increased immigration. For most people, this would mean extending the compassion that they feel to the Aborigines, peasants, and Tibetans to apply to the British as well.

But another argument is that we should consistently support universal culture’s attempt to impose progress on traditional cultures. Maybe we should tell the Native Americans that if they embraced global capitalism, they could have a tacqueria, sushi restaurant, and kebab place *all on the same street* in their reservation. Maybe we should tell the Aborigines that modern science says the Dreamtime is a myth they need to stop clinging to dumb disproven ideas. Maybe we should tell the Tibetans that Vajrayana Buddhism is too intolerant of homosexuality. Take our conviction that rural Englanders are just racist and xenophobic and ill-informed, and extend that to everyone else who's trying to resist a way of life that's objectively better.

I am sort of torn on this.

On the one hand, universal culture *is* objectively better. Its science is more correct, its economy will grow faster, its soft drinks are more refreshing, its political systems are (necessarily) freer, and it is (in a certain specific sense) what everybody would select if given a free choice. It also seems morally better. The Tibetans *did* gouge out the eyes of would-be-runaway serfs. I realize the circularity of saying that universal culture is objectively morally better based on it seeming so to me, a universal culture member – but I am prepared to suspend that paradox in favor of not wanting people's eyes gouged out for resisting slavery.

On the other hand, I think that "universal culture is what every society would select if given the opportunity" is less of a knock-down point than it would seem. Heroin use is something every society would select if given the opportunity. That is, if nobody placed "censorship" on the spread of heroin, it would rapidly spread from country to country, becoming a major part of that country's society. Instead, we implement an almost authoritarian level of control on it, because we know that even though it would be very widely adopted, it's not something that is good for anybody in the long term. An opponent of universal culture could say it has the same property.

Things get even worse when you remember that cultures are multi-agent games and each agent pursuing its own self-interest might be a disaster for the whole. Pollution is a good example of this; if the best car is very polluting, and one car worth of pollution is minimal but many cars' worth of pollution is toxic, then absent good coordination mechanisms everyone will choose the best car even though everyone would prefer a world where nobody (including them) had the best car. I may have written about this before.

I'm constantly intrigued (though always a little skeptical) by claims that "primitive" cultures live happier and more satisfying lives than our own. I know of several of this type. First, happiness surveys that tend to find Latin American countries doing as well or better than much richer and more advanced European countries. Second, the evidence from the Amish, whose children are allowed to experience the modern culture around them but who usually prefer to stay in Amish society. Third, Axtell's paper on prisoner exchanges between early US colonists and Native Americans; colonists captured by the natives almost always wanted to stay and live with the natives; natives captured by the colonists *never* wanted to stay and live with the colonists. Many people have remarked on how more culturally homogenous countries seem happier. Bhutan itself might be evidence here, although I've seen wildly different claims on where it falls on happiness surveys. I've also talked before about how China's happiness level stayed stable or even dropped during its period of rapid development.

(on the other hand, there's also a lot of counterevidence. More democratic countries seem to be happier, and democracies will generally be the low-censorship countries that get more assimilated into universal culture. Free market economies are happier. Some studies say that more liberal countries are happier. And there's a complicated but positive relationship between national happiness and wealth.)

I also think that it might be reasonable to have continuation of your own culture as a terminal goal, even if you know your culture is "worse" in some way than what would replace it. There's a transhumanist joke – "Instead of protecting human values, why not reprogram humans to like hydrogen? After all, there's a *lot* of hydrogen." There's way more hydrogen than beautiful art, or star-crossed romances, or exciting adventures. A human who likes beautiful art, star-crossed romances, and exciting adventures is in some sense "worse" than a human who likes hydrogen, since it would be much harder for her to achieve her goals and she would probably be much less happy. But knowing this does not make me any happier about the idea of being reprogrammed in favor of hydrogen-related goals. My own value system might not be objectively the best, or even very good, but it's my value system and I want to keep it and you can't take it away from me. I am an individualist and I think of this on an individual level, but I could also see having this self-preservation-against-optimality urge for my community and its values.

(I've sometimes heard this called Lovecraftian parochialism, based on H.P. Lovecraft's philosophy that the universe is vast and incomprehensible and anti-human, and you've got to draw the line between Self and Other somewhere, so you might as well draw the line at 1920s Providence, Rhode Island, and call everywhere else from Boston all the way to the unspeakable abyss-city of Y'ha-nthlei just different degrees of horribleness.)

Overall I am not 100% convinced either way. Maybe some traditional cultures are worse than universal culture and others are better? Mostly the confusion makes me want to err on the side of allowing people to go either direction as they see fit, barring atrocities. Which are of course hard to define.

I like the Jewish idea of the Noahide Laws, where the Jews say “We are not going to impose our values on anyone else…except these seven values which we think are incredibly important and breaking them is totally beyond the pale.” Sometimes I wish universal culture would just establish a couple of clear Noahide Laws – two of them could be “no slavery” and “no eye-gouging” – and then agree to bomb/sanction/drone any culture that breaks them while leaving other cultures alone. On the other hand, I also understand universal culture well enough to know that two minutes after the first set of Noahide Laws were established, somebody would propose amending them to include something about how every culture must protect transgender bathroom rights or else be cleansed from the face of the Earth by fire and sword. I’m not sure how to prevent this, or if preventing it is even desirable. This seems like the same question as the original question, only one meta-level up and without any clear intuition to help me solve it. I guess this is another reason I continue to be attracted to the idea of Archipelago.

But I think that none of this makes sense unless we abandon the idea that “universal culture” and “western culture” are one and the same. I think when Caplan’s debate opponent talked about “protecting Western culture”, he was referring to something genuinely fragile and threatened.

I also think he probably cheated by saying we needed to protect it because it was responsible for so many great advances, like Coca-Cola and egalitarian gender norms. I don’t think that’s fair. I think it’s a culture much like Tibetan or Indian culture, pretty neat in its own way, *possibly* extra interesting as the first culture to learn the art of summoning entities from beyond the void. Mostly I’m just happy that it exists in the same way I’m happy that pandas and gorillas exist, a basic delight in the diversity of the world. I think it can be defended in those terms without having to resolve the debate on how many of its achievements are truly its own.

Post-Partisanship Is Hyper-Partisanship

I.

A few years ago, I wrote:

> I want to avoid a very easy trap, which is saying that ingroups vs. outgroups are about how different you are, or how hostile you are. I don't think that's quite right.
>
> Compare the Nazis to the German Jews and to the Japanese. The Nazis were very similar to the German Jews: they looked the same, spoke the same language, came from a similar culture. The Nazis were totally different from the Japanese: different race, different language, vast cultural gap. But the Nazis and Japanese mostly got along pretty well. Heck, the Nazis were actually moderately positively disposed to the Chinese, even when they were technically at war. Meanwhile, the conflict between the Nazis and the German Jews – some of whom didn't even realize they were anything other than German until they checked their grandparents' birth certificate – is the stuff of history and nightmares. Any theory of outgroupishness that naively assumes the Nazis' natural outgroup is Japanese or Chinese people will be totally inadequate.
>
> So what makes an outgroup? Proximity plus small differences. If you want to know who someone in former Yugoslavia hates, don't look at the Indonesians or the Zulus or the Tibetans or anyone else distant and exotic. Find the Yugoslavian ethnicity that lives closely intermingled with them and is most conspicuously similar to them, and chances are you'll find the one who

> they have eight hundred years of seething hatred toward.

I didn’t coin a silly term for the relationship of the Yugoslavs and the Tibetans, but let’s use “fargroup” in order to remind us of the Near/Far distinction. We think of groups close to us in Near Mode, judging them on their merits as useful allies or dangerous enemies. We think of more distant groups in Far Mode – usually, we exoticize them. Sometimes it’s positive exoticization of the Noble Savage variety (understood so broadly that our treatment of Tibetans counts as an example of the trope). Other times it’s negative exoticization, treating them as cartoonish stereotypes of evil who are more funny or fascinating than repulsive. Take Genghis Khan – objectively he was one of the most evil people of all time, killing millions of victims, but since we think of him in Far Mode he becomes fascinating or even perversely admirable – “wow, that was one impressively bloodthirsty warlord”.

(this jars when other cultures do it to people we consider Near-Mode evil – for example India’s Hitler-themed clothing store, romance movies, and their use of Mein Kampf as a business advice book. It’s a bit strange, but not objectively stranger than us having a comedy movie about Kim Jong-un)

Fargroups aren’t always people who are literally distant from us. It seems more like it’s people who don’t threaten us, or aren’t in competition with us, or don’t get involved in the conflicts we care about, or something like that. There’s a Scientologist Church just a couple of miles from my house, and I recognize that Scientologists do some pretty horrible things, but none of them affect me, or people close to me, or values that I have a personal connection with, so I’m still more likely to find them cartoonishly funny, Kim Jong-un style, than I am to feel angry or afraid of them.

We exoticize fargroups, but we can also use them as props in our own local conflicts. For example, a lot of the time I hear about ISIS, it's in contexts like the Democrats being weak on ISIS or Trump playing into ISIS' hands, or how our immigration policy makes us easy prey for ISIS, or fundamentalist Christians are no different from ISIS, or something like that. We use sympathetic fargroups the same way. The Tibetans aren't just wise and noble, they're a foil to our overly materialist society, or an example of how religion can be based on reason instead of faith, or whatever. This is all as the theory would predict. The GOP view the Democrats as more of an outgroup and ISIS as more of a fargroup. It's harder for them to have genuine outrage at ISIS for beheading a bunch of people, than for them to have outrage at the Democrats for not mentioning the beheading. Even in cases where they seem angry at ISIS in a non-Democrat related way, I would argue that a lot of it can be traced back to appreciating the way ISIS proves various domestic points, like "Muslims are scary" or "the barbarism vs. civilization axis is important".

II.

Last month I asked on Tumblr:

> I remember that when I was young and the Internet was young, people online were debating religion vs. atheism ALL THE TIME. It felt inescapable. Whatever else you were trying to discuss, eventually it would turn into a religion vs. atheism debate. Whenever it came up, people would sigh and say "Oh no, not another religion vs. atheism debate".

> I remember spending a lot of time at talk.origins and infidels.org because religious people kept attacking me and I wanted to be able to rebut their points. And I remember a lot of people who seemed to genuinely believe that religion was like the #1 problem in the world, maybe even the only problem in the world because it was the root cause of all of the others.
>
> I haven’t seen an online religion vs. atheism debate in years now. Occasionally somebody criticizes Richard Dawkins or something, but it’s always a tone argument and practically never about the nitty-gritty of Biblical contradictions or whatever. Now social justice vs. anti-social-justice seems to have totally taken over as the Annoying Thing Everybody On The Internet Has To Debate.
>
> Has anybody else noticed this? Is it just me, or maybe a function of the places I hang out / used to hang out?

I got a lot of responses. Other people confirmed this was a real phenomenon and that they remember it the same way. The consensus explanation was that there was a moment in the 90s and early Bush administration when evangelical Christianity seemed to have a lot of political power, and secularists felt really threatened by it. This caused a lot of fear and arguments. Then everyone mostly agreed Bush was terrible, studies came out showing religion was on the decline, evangelicalism became so politically irrelevant that even the Republicans started nominating Mormons and Donald Trump, and people stopped caring so much.

Now I see atheists sharing things like this:

SS COLUMBA - BRIGID
Roman Catholic Church
MASSES
SAT 4:00 PM
JESUS HAD 2 DADS
AND HE TURNED
OUT JUST FINE

Not only have they stopped caring that much about religion, but they’re willing to adopt progressive religious people as role models and generally share stories that portray religious people in a positive light. Pope Francis gets to be the same sort of Socially Approved Benevolent Wise Person as the Dalai Lama.

I think once Christianity stopped seeming threatening, Christians went from being an outgroup to being a fargroup, and were exoticized has having the same sort of vague inoffensive wisdom as Buddhists.

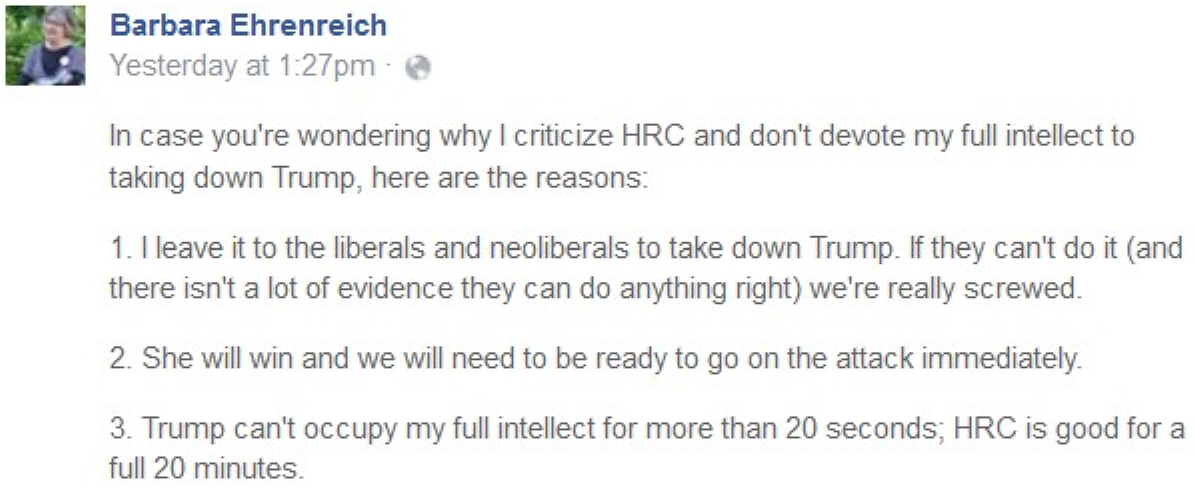
I saw something that seemed very similar during my time interacting with movement atheists. There was a split between people who were raised in fundamentalist families and very traumatized about it and who viewed Christianity as an outgroup, versus people who were raised in agnostic families and pretty live-and-let-live and who viewed Christianity as an fargroup. I know it seems weird to say that movement atheists living in a majority-Christian country treated a religion they interacted with every day the same way the Yugoslavs treated Tibetans, and sure, they would make fun of them, but that was exactly it – they found religion funny – and even in the process of lightly mocking them they tried to avoid stepping on too many toes. The fundie-raised atheists would propose something really combative and offensive, and the secular-raised atheists would say “Oh, come on, we don’t want to be jerks about this, Christians are basically nice people who are just a bit deluded”. To the fundie-raised atheists it was real, it was a hot war, these people were monsters; to the secular-raised atheists, religious people were just kind of wacky in a problematic way, like the North Koreans, and nobody in America lives their life in a state of constant rage about how evil North Korea is.

And I think as the threat of movement fundamentalism declined, there was a shift among atheists from more emotional hostility to more of a live-and-let-live kind of attitude.

(and then movement atheism started tearing itself apart even more viciously than it was already. I don’t know if this was a coincidence and I’m still curious whether conservation of tribalism is a real phenomenon.)

III.

From Facebook the other day:



Barbara Ehrenreich
Yesterday at 1:27pm ·

In case you're wondering why I criticize HRC and don't devote my full intellect to taking down Trump, here are the reasons:

1. I leave it to the liberals and neoliberals to take down Trump. If they can't do it (and there isn't a lot of evidence they can do anything right) we're really screwed.

2. She will win and we will need to be ready to go on the attack immediately.

3. Trump can't occupy my full intellect for more than 20 seconds; HRC is good for a full 20 minutes.

All good reasons. But I've been seeing more and more people lately saying things like this. There have always been primary elections, and there have always been intra-Left disagreements, but the level of Bernie vs. Hillary drama at the Democratic Convention this week seems to be something new. Ehrenreich-style leftists focus on critiquing Hillary instead of Trump – either within or outside of the context of supporting the Sanders campaign. And on the other side, Hillary-supporting liberals go after *Sanders* and his supporters instead of Trump – Freddie deBoer has written frequently (some would say incessantly) about this.

The right, of course, has its own conflicts between Trump partisans and Trump opponents, culminating in Cruz's non-endorsement. Also relevant: the alt-right's favorite slur of "cuck" is short for "cuckservative" – an insult not for leftists but for conservatives who they think are doing conservativism badly.

People are talking more and more about partisan bubbles. People dividing into political tribes, and cutting off contact with people on the other side. Cultural, geographic, and social differences isolate people so completely that for example my Facebook feed tends about 95% liberal; I'm sure there are other people out there with the opposite problem. I think that as bubbleification increases, the other party becomes less and less of an outgroup and more and more of a fargroup.

Republicans still “threaten” me in the sense of being able to enact policies that harm me. And people less privileged than I am face even more threats – a person dependent on food stamps has a lot to fear from Republican victories. But Republicans aren’t taking over my social circle or screaming in my face. In a purely social context they start to seem more like cartoonish and distant figures of evil, rather than neighbors and coworkers. The average Trump voter no longer seems like an uncanny-valley version of me; they seem like some strange inhabitant of a far-off land with incomprehensible values, just like ISIS.

I have yet to meet anybody in person (other than my patients) who supports Donald Trump. On the other hand, I’ve met a *bunch* of people on both sides with strong feelings about Bernie vs. Hillary. The Bernie vs. Hillary conflict is real to me in a way that the Hillary vs. Trump conflict isn’t. It has the potential to split my friend group. There are social advantages for me of taking either side, and I *could* reasonably take either side without people looking at me like I went to work stark naked. This is the kind of socially relevant conflict that produces ingroups and outgroups in a way that America vs. ISIS never will.

My guess is that this sort of thing is only going to become more common. Partisanism is going to give way to hyperpartisanism, where people hate other factions of their own party with the same venom they previously reserved for their opponents across the aisle.

At the same time, old outgroup hatreds will take on a different character. Even If You Don't Like Donald Trump, You Should Understand The Pain Of His Poor White Supporters. And I Know Why Poor Whites Chant Trump, Trump, Trump. And Millions Of Ordinary Americans Support Donald Trump; Here's Why. And The Incredible Crushing Despair Of The White Working Class. I'm not saying these articles are *typical*; for every one of these articles there are ten "Trump Voters Are Xenophobic Trailer Trash" pieces. I'm saying that it's weird that they're happening at all.

Same thing with Brexit. Yes, the usual xenophobic trailer trash articles. But also: In This Brexit Vote, The Poor Turned On An Elite Who Ignored Them. And Brexit Voters Are Not Thick Or Racist, Just Poor. And Outraged Elites Should Listen To Fed-Up Brexit Supporters.

(and of course this blog has been pushing a similar line for reasons that are probably not completely ahistorical or divorced from general trends)

People are starting to treat Trump voters and Brexit voters as interesting and worthy of respect, which means they're not really an outgroup any more. Talking about how poor they are and how sympathetic we should be and how we need to be more educated in order to understand what they're going through all sound like instances of fargroup exoticization to me.

I predict (50% probability) that the progressives most carefully bubbled and separated from any actual threat from Republicans – which disproportionately includes politicians, journalists, and other opinion-makers – will start treating the Trump-voting classes more like Tibetans. I predict when they talk about specific bad Republicans like Trump, they'll focus more on the ways they are funny and cartoonish (far too easy with Trump, but maybe the next guy will be a better test) instead of the ways they're threatening. I predict that conflicts within the progressive movement will be increasingly vicious and increasingly likely to use poor whites as a political football ("the other side is bigoted against poor whites!"). I predict this will happen much more if the Democrats win the election than if they lose it; it's always easier to be gracious toward a vanquished opponent.

(I'm of course 100% guilty of all of this myself)

(Yeah, this is a change in predictions since Right Is The New Left, which talked about something similar but reached a kind of different conclusion)

I'm not sure how things will go on the Republican side. I haven't seen the same signs of rapprochement from them – but then Republicans have never shown the same tendency to sympathize with poor exotic fargroups that Democrats do. But I also don't know as many Republicans and maybe if this were happening I would miss it.

Book Review: PiHKaL

I.

PiHKAL ("Phenylethylamines I Have Known And Loved"), subtitled "A Chemical Love Story", is the autobiography of Alexander and Ann Shulgin. Alexander Shulgin was a chemist who invented lots of new psychedelic drugs. Ann was his wife. Together they discuss their chemistry research and their relationship.

I was expecting a sort of popular science style book that cleverly ties the chemical story into the love story. You know the drill – the bonds between people are like the bonds between atoms, fragile in some ways yet incredibly strong in others. Or something like that. It would use the human interest story to hook you into the chemistry, then use the chemistry to give scientific respectability to the human interest story, so that both science nerds and hopeless romantics could enjoy it and gain more of an appreciation for the other side's point of view.

Some parts of PiHKAL approached this kind of style. But a lot of them didn't. Chapters and sections tended to be kind of either/or. You can be reading one moment about how MDMA is an n-methylated homolog of MDA, and the next moment about how Alexander Shulgin's broad shoulders rippled as he was making love. It was a bit jarring.

The first quarter of the book was about Alexander Shulgin's childhood and early life. He was born in 1930s Berkeley, went to school, became a chemist, and got a job with Dow Chemical. He invented a pesticide so successful that Dow gave him total freedom to work on whatever he wanted – which turned out to be psychedelics. He met, courted, and married his first wife Helen, a process which receives six sentences (compared to seven sentences a page later on the history of 3,4,5-trimethoxyamphetamine synthesis techniques). He has various scientific and professional successes, including a supporting role in the invention of MDMA/Ecstasy. His first wife dies of a brain haemorrhage.

The second quarter is a love story told from Ann's perspective. It reminds me a little bit of very bad fanfiction, like "My Immortal". Ann very briefly talks about her childhood, then gets to the part where she (after three previous marriages and divorces) meets Alexander. Cue lavish descriptions of his eyes, his hair, his muscles, his shoulders, how much better than her he is, how there's no way someone as awesome as he is could possibly fall for someone boring like her, how many sparks get sent through her spine every time he gazes at her, et cetera. I am okay with people being in love, but this is a bit excessive. We get to hear about how amazing it is that he is a chemist, how amazing it is that he creates new drugs, how amazing it is that he is a brilliant yet dark and brooding loner, how there's no way someone like that could ever love her (but spoiler: he totally does). Sometimes it seems to strain plausibility – there is a section where Shulgin tells Ann that he is interested in psychopharmacology, and she innocently goes into "Oh my, that's such a big word for a girl like me, I wonder what it means". But she has been married to a psychiatrist for ten years at this point, and also, she's a hospital transcriptionist. I roll to disbelieve that she has never heard the word "psychopharmacology" before – let alone never heard the word "pharmacology" and the prefix "psycho-" from which the meaning is completely obvious.

The third quarter is an assortment of trip reports, social gatherings, and arguments against the War on Drugs. It is probably my favorite part, given that it's neither as dry as the autobiography nor as overwrought as the love story. The trip reports about weird new psychedelics that nobody else has ever tried are really what I'm here for – they occur throughout, but especially here, and they do not disappoint.

The fourth quarter is a cookbook detailing the ingredients of, synthesis techniques for, and effects of 179 different psychedelic substances. It's *really* fascinating, and I'm consumed by a desire to try some of it, except that they all begin by with instructions like "Obtain a professional-quality chemistry lab and several zillion different compounds with names like 2,5-dimethoxythiophenol", and end with getting raided by the DEA. So I will have to stick to enjoying Alexander Shulgin's psychedelic experiments vicariously.

II.

There were a couple of levels on which I enjoyed this book, though none left me completely without questions.

The first level, of course, is Shulgin's work on psychedelics. My opinion on psychedelics hasn't changed since Universal Love, Said The Cactus Person. I think they're really interesting and mysterious and show every sign of pointing at something profoundly important. I also think that nobody has ever been able to consistently extract anything useful or scalable out of them, and until someone does, they'll remain a weird toy where you take them and feel transcendent joy for a few hours, and if all you *want* is to feel transcendent joy for a few hours then they're definitely the way to go, but as of yet it's unclear what relevance they can have to any other project.

Alexander Shulgin disagrees. At least I think he disagrees. He *is* dark and brooding and quiet, and he doesn't wear his heart on his sleeve even in his autobiography. But he mentions – once – that he believes his project is vitally important for the human race. This is Ann:

> It seems to me that the magic plants – and the psychedelic drugs – are there to be used because the human race needs some way of finding out what it is, some way of remembering things we've usually forgotten by the time we're grown up. I also think that the whole 1960s eruption- all that psychedelic experimenting and exploring – was due to some very strong instinct – maybe on the collective unconsciously level, if you want to use Jung's term – an instinct that's telling us if we don't hurry up and find out why we are the way we are, and why we do the things we do, as a species, we could very soon wipe ourselves out completely

Alexander says this is what drives him also, then adds:

> Of course, there are many ways to alter your consciousness and your perceptions; there always have been, and new ways will keep being developed. Drugs are only one way, but I feel they're the way that brings about the changes most rapidly, and – in some ways – most dependably. Which makes them very valuable when the person using them knows what he's doing. I thought for a while that I could use music to accomplish what I wanted to do, because music can be a very powerful consciousness changer. But when I discovered that I had a certain knack for chemistry, I made a decision to go that way, to concentrate on developing these tools. Mostly, I suppose, because these particular drugs, these materials, are a way to bring about new insights and perceptions quickly, and – well, I just don't know if we have much time. Sometimes I suspect it may be too late already…I have no intention of getting lazy,

> and there’s nothing better than a suspicion that time’s running out, to keep you working hard.

Shulgin and his friends seem like good people. But not, crucially, like the *best* people. Shulgin himself – by his own admission, based on facts that he himself presents in his own autobiography – is consistently kind of a jerk to his wife (and his wife kind of agrees). He gets depressed and ornery a lot, sometimes to the point where it seriously interferes with his work and relationships. His circle of friends seems to have some problems with marital infidelity and random drama, and he tells one story about a distant friend-of-a-friend obsessed with LSD who seems to be an outright con man. I’m not saying they’re bad people; quite the opposite, the book makes them seem very human and if I lived in the same time and place as them I would be delighted to have the opportunity to know them. But they seem, well, about as good or bad as any other set of intelligent, creative people. It’s not clear that their psychedelic use – and *man*, do these people use psychedelics – has made them morally or spiritually exceptional. It’s hard to shake the thought that these people would be relatively nice and interesting artists and scientists with a little bit of marital infidelity and personal drama even if they’d never taken anything stronger than Tylenol.

Don’t get me wrong – *during* the trips they are constantly seeing God and understanding the oneness of everything in creation. But even Alexander Shulgin’s close friend group aren’t high more than like 20% of the time. I’m not sure exactly what about them makes them potential human-race-savers. Yes, I think they’re probably anti-nuclear-weapon. But this is the 1970s Berkeley counterculture; anti-nuclear-weapon people are not exactly hard to come by.

And so I was left with one question the book didn't really answer – why is Shulgin doing this? What is his hope? Does he hope that the 200th new psychedelic he discovers will be the one that *really* teaches people universal love, to the point where they can't ignore it? That just having twenty different slightly different permutations of the same psychoactive sulfur compound isn't enough to create a world revolution in consciousness, but having thirty of them is? Why didn't he become the scientist in this article, who has come up with a clever way to extend the DMT high in order to be able to enter complex negotiations with the machine elves, hopefully involving factoring large numbers? That's the sort of project that can *go* somewhere.

I'm not saying that everyone needs to have a four-hundred-step plan for how exactly what they're doing today is going to save the world, Leverage Research style. I can't tell you how writing this blog post or doing psychiatry is going to save the world. And I'm not saying people can't focus on their comparative advantage – if you're a brilliant chemist, maybe you should invent new chemicals and let politicians and religious leaders figure out what to do with them. But Shulgin seems to think he's doing something more important than coming up with new and better toys, and it's not totally clear to me what this is.

Shulgin does have a great excuse – the War on Drugs is in full swing by the middle of his career, and this prevents a lot of his products from getting the trial they deserve. For example, he introduces a psychiatrist friend to MDMA, and the psychiatrist finds it to be an immense help to psychotherapy, helping patients realize and come to terms with their issues much faster than the non-chemically-assisted version. I've heard this from psychiatrists *I* know as well, and maybe we could have had a revolution in mental health if the DEA hadn't banned this kind of thing (there is, in fact, a big literature on psychedelics in psychiatry, most of which shows impressive effects in the very small experiments that have been permitted thus far). So maybe Shulgin’s angle is that he’s developed very useful mental health treatments which unfortunately the medical establishment refuses to consider, but they’ll be there if anybody needs them. But then how come he keeps inventing more of them, seemingly with no interest in whether they help the mentally ill or not? How come he keeps talking about saving humanity instead of curing depressed people?

The only hint I get is during a trip report for a particularly powerful compound called 2C-T-4, where Shulgin writes:

> There is a simultaneous union with everything around me, and thus with everything within me too. A complete identification with my environment. And a sense of being at total peace with it, as well. If this is me, then I thank the dear Lord for a wonderful awareness, at least for a short time, of the fact that we can be so rich and beautiful. The mind flows and with it the soul, and no matter what words I put down in an effort to catch the wondrous monolog, I can do it little justice…
>
> I have been fooled, again and again, into thinking that the magic of the unified reality was in the drug, and not in the person. Of course it is in the person – and only in the person – but if a drug could be found that would consistently catalyze this, then it would be one of the most powerful and awesome drugs that could be conceived of by man. If it were this material, 2C-T-4, it would have to be held apart with a reverence that would be impossible to describe or explain on a patent application!!

But Shulgin later reports that the drug does not consistently have this effect; testing it on other people (and again on himself) he gets various interesting psychedelic trips but never a return to the same level of transcendence. So maybe Shulgin is looking for a drug that consistently works as well as 2C-T-4? But I'm not sure what he would do with it if he found it. Ann, for example, describes her first mescaline trip in language a lot like the language Shulgin uses for 2C-T-4, and many others (eg Huxley) do the same – but everyone already knows about mescaline. Would releasing a consistent version of 2C-T-4 to the world do something mescaline hasn't already done? What *has* mescaline already done?

Oddly enough, it is Ann – who keeps on insisting that she is not intellectual, that she is hopelessly boring compared to Alexander, that we should be reading her parts only to gain a hero-worshipping outsider's perspective on Alexander – whose speculations on this subject I really like. This is from her mescaline trip report. She says that on mescaline she understood for the first time that the world was perfectly good, and writes:

> I nodded, remembering some of the phrases I'd read in books and articles about psychedelic experiences, phrases like "Everything's all right just exactly the way it is," and the equally infuriating "I'm okay, you're okay," which had always sounded unbearably fatuous and self-satisfied. I'd often thought angrily that the writers had conveniently forgotten about the babies in Calcutta garbage cans, sorrow and hurt and loneliness, and the rest of a planetful of miseries. I'd said to myself, here's some whacked-out idiot rhapsodizing about life being all right just the way it is. It had never stopped me from reading about such experiences, but my liberal

> soul had always ground its teeth at that aspect of the reports.
>
> Now – now I would have to take it all back, all that resentment, because I was beginning to understand. I stopped in the road and looked at Sam and looked past him, and around and up at the grey sky and I knew that everything in the world was doing exactly what it was supposed to be doing; that the universe was on course, and that there was a Mind somewhere that knew everything that happened because it *was* everything that happened, and that, whether I understood it with my intellect or not, all was well. I simply knew it and I knew that I would try to figure it out later, but that I had to absorb the truth of it now, standing on a wet road in Golden Gate Park.

At least here, she seems to be using the mescaline not to catapult humanity to a higher state of existence, but to make her peace with the current state. I'm not sure making peace with the current state is philosophically justifiable, but it seems to have helped her, and I can imagine it helping a lot of people, as long as they stick to viewing it as a psychological truth and not as an excuse for quietism – something Ann doesn't seem to have done.

III.

The second level on which I enjoyed this book was anthropology and ethnography.

The Shulgins met through a group that had branched off of Mensa. Their social circle consisted of a mishmash of scientists, underachieving geniuses, mental health professionals, hippies, and people convinced that their new projects were going to save the world – all in the context of

Berkeley and the Bay Area. This is also *my* social circle, thirty years later, so it’s interesting to see how things have changed and how they’ve stayed the same.

Ann starts her narrative while she is dating a fellow Mensan named Kelly. So, ethnographical study number one:

> He was an intense man with a striking, angular face who had met me at a Mensa gathering four months earlier. The next day, he came to my house and asked me to marry him. He explained, much later, that of course he knew I would refuse – had counted on it, in fact – but that he had often found proposing marriage to be an effective way of getting a woman’s attention. There was no denying that it had done just that.
>
> Kelly’s passions in life were computers, good-looking older women and the creation of new IQ tests. I also discovered that he had a generalized contempt for humanity, referring to most people ask ‘turkeys’, and a tendency to uncontrolled explosions of rage, which often resulted in his having to apologize later for damage done to someone else’s furniture or a relationship – usually both.
>
> He explained about his painful illnesses in childhood and his demanding, punitive father, and asked me to be understanding a patient. It worked for a while (I’ve always had a soft spot for intelligent neurotics), but after one memorable day when he smashed some of my records in front of the children, screaming at me for coming home ten minutes late from work and keeping him waiting, I told him if he didn’t go into therapy, I was through.

> Kelly's answer was, "I've never met a psychiatrist I couldn't out-think and out-reason, I'm not about to waste my time or my money on one of those cretins!"
>
> This Thursday gathering in Berkeley was an effort on Kelly's part to bring together people he considered intelligent enough to, as he put it, appreciate what he could teach them about using their minds effectively.

Ethnographic study number two, slightly edited for length; this is Ann introducing Alexander to some of her friends:

> [I] proceeded to give him rapid-fire descriptions of some of Mensa's main attractions, as they stood talking or moved past us. In my best museum-guide manner, I told him, addressing his right ear closely because of the noise in the room, "You see that man there, the tall one with the red vest? He created the SIG – Special Interest Group – which is known as the Orgy SIG; I forget his official title for it, something like Sexual Freedom SIG, but everyone refers to it by the other name. I've never been to a meeting, but I hear they're a lot of fun for those who go in for that sort of thing…"
>
> "That woman over there in the purple dress, the one standing in a straight line between us and the candles – that's Candice. She's a very good-hearted, motherish person who gives the Mensa tests in this area, and for a while her little boy, Robin, was the youngest member of Mensa in the country. He's around ten now, and no longer the youngest."
>
> I told him about the mathematical computer which inhabited the sometimes bewildered soul of the young man on the couch, and he said he was very interested in that kind of mind, and would go over and talk with him

later on. I said I hoped he would want to do that, because few people paid any attention to the boy, and he was very sweet.

I asked Shura [Alexander Shulgin]'s ear, “Why haven't you joined Mensa, by the way? It's a good way to meet interesting people, especially when you've been divorced – or widowed.”

“Well,” shouted Shura, “To tell you the truth, I never thought of applying, probably because you have to take an IQ tests, and I will not take an IQ test.”

“Why, in heaven's name?”

“I feel total, complete disgust for all tests of intelligence, and only limited patience with the people who give them. When I was in the third grade or thereabouts, I was given a so-called IQ test, a Binet-something-something, and I made an honest and diligent effort to complete it. There were angular objects, and number games, and if-this-then-what types of questions and the strategies needed for getting to most of the answers were pretty obvious.”

“You did well?”

“Of course I did, and that's where I really tangled with the school principal. He accused me of having cheated, since no one could get the results I had gotten without cheating, and so I was in essence thrown out of the testing group, and was pretty much humiliated. They obviously wanted scores that fit on a kind of distribution curve about some sort of a norm. Mine was a bit too far to the right of the curve. My mother was furious with the principal; she pulled me into his office and confronted him and lectured him about my integrity, which made

> me want to run and hide even worse than before. I swore then that I'd never take another IQ test, and I never will!"

Ethnographic description number three is Shulgin's colleague, a German professor named Dolph, and his wife Ursula. Shulgin's first marriage, to the woman he spent six sentences on in Part I, was never very happy – never unhappy, neither of them was abusive or anything, just sort of boring and straightforward. When Shulgin met Ursula, he fell in love, maybe for the first time in his life, and they had a brief affair before Dolph and Ursula had to go back to Germany:

> We met, Ursula and I, two or three times in some inn or private place sufficiently far away from the Bay Area to minimize the possibility of being seen by a friend or acquaintance, and I discovered what it was to feel unashamed, uncensored, joyous sexuality.
>
> Being in love, like any other kind of consciousness alteration, makes small but real changes in the way you view things about you, and in the way you behave around others. Over the years, my friends had come to accept me as what they affectionately called a "difficult genius", and were quite used to my habitually ironic humor, cutting commentary, and somewhat sour view of the world. One of the hardest things I had to do, in my unaccustomed role of secret lover and beloved, was to avoid giving expression – in the company of family or friends – to the feelings of optimism and even outright niceness which overtook me now and then, and which I knew would cause some degree of concern if they were detected.

I really like this passage. Here’s someone who has tried more psychedelic drugs than anyone else in the world, and what really changes his outlook is *the power of love*.

But then it gets complicated. Shulgin’s wife dies. He corresponds incessantly with Ursula back in Germany, urging her to leave Dolph and come to California and marry him. She says yes, but asks for some time to plan. Shulgin is overjoyed and says to take as much time as she wants. Weeks turn into months. Months to years. They keep writing each other. Ursula insists that she continues to be excited at the impending plan to move to California and marry Shulgin, but she keeps asking for more time. She needs to close up her affairs in Germany. She needs to figure out a way to break it gently to her husband. She needs to stay to comfort her husband during this difficult time. She needs to figure out how she’s going to send her stuff.

Meanwhile, all this time Shulgin is calling her in Germany to talk to her a lot, and a lot of the time her husband answers the phone, and he’s *got* to suspect something at this point, but he’s still perfectly cheerful and friendly, and finally Shulgin asks, “You know your wife is planning on moving to California to live with me,” and he’s like “Oh yeah, I know”, and Shulgin starts to get a *tad* suspicious. Meanwhile, this is around the time he starts meeting/falling in love with Ann, and he keeps telling Ann “I really like you, but this is just a fling until my True Love Ursula gets here from Germany”, and Ann is always okay with this, because Alexander Shulgin is Objectively The Best And Most Attractive Person In The World, and obviously having him for a short time is better than having anybody else forever. But it starts to get really annoying and everybody is super confused by what’s going on, and finally one of their psychiatrist friends tells Ann:

Ursula is – how best to put it – she’s a person who, when she’s attracted to a man, intuitively senses what’s lacking in his emotional life, and she has a compulsion to become whatever that man most needs in a woman. She probably convinces herself each time that she’s truly in love, but I doubt she’s capable of what most of us would call real loving. The Jungians have a term, ‘anima woman’. The anima woman lacks a solid identity; like many great actors, she borrows – she takes on – a sense of wholeness from playing a part. In this case, it’s the part of the muse, the inspiration, the adored dream-woman. She fulfills a fantasy, and you can imagine the tremendous emotional rewards there are for her in such a role, as long as the affair lasts. Each affair lasts, of course, only until the next needy attractive man comes along. It’s all unconscious, by the way; I don’t think Ursula has the slightest idea of what she’s doing or why she feels compelled to do it. Or, for that matter, why the men she’s drawn to always happen to be married. When it’s time to move on, she explains – and probably believes – that she’s ending the relationship because she couldn’t live with the responsibility of having broken up a marriage.

When she first joined [our] group, we had long talks with each other, under the influence of [Shulgin’s psychedelics], and she told me a lot about her involvements with married men; she told me more than she realized or intended to. It was a subtle form of preening, under the guise of telling problems to a wise, sympathetic psychologist, you understand? Gradually, I put enough of the pieces together to understand the pattern. By that time, she had stopped telling me personal things about herself and her life, and I sensed that she was feeling uncomfortable around me […]

> The dynamics of this kind of psychological compulsion are more than I want to go into right now, but what worries me is I believe Ursula is simply not capable of true emotional commitment to anyone. She'll play the role for a time, as I said, until somebody else comes along – someone she finds appealing, with an emotional hole that's begging to be filled – and she'll move on to the new challenge. That's what's going to happen to Shura [Alexander Shulgin], I'm sure of it. I know it! I love him very much – we all do, you know – and sooner or later, he's going to be badly hurt. That's why I'm more pleased than I can say, to see you here. I don't know what your relationship is with Shura, but it's clear that you care for him, and I hope that – umm – I hope you'll stay around. To help cushion the blow, when it comes; to give him something real to hold onto, when the unreal thing begins to unravel. Which I'm sure will happen before long, now that Ursula finds herself involved with a man who has – quite unexpectedly – become free to make an open commitment to her and ask her to do the same. Her bluff is being called.

Everything goes back and forth a lot, the whole social circle becomes hopelessly muddled, but in the end the psychiatrist is proven right. After much back-and-forth, Ursula agrees to fly to California that very day. When she doesn't actually arrive, Ann sends her a nasty letter, informing her that she exists, that she's on to her, and that now is the time to put up or shut up. Ursula then sends Shulgin a letter:

> Dearest, dearest Shura,
>
> A window has widely opened to you, a soul-window, a love-window, of graceful being – being together. A common space of breathing, of light touch, of inner

smile. I could let those hours pass without telling you, and then you would never know what I am feeling – you would have only your own experience. Or I could share this with you. That is what I am doing […]

In a past life, about 2,000 years ago, you took a long knife and cut my throat, took my life, murdered me in the desert! You were the chief of our tribe, and I was a young girl, and you killed me! The whys are irrelevant. I have seen this over and over, and others who lived with us in that time have come to me in this life and warned me to be aware of this old karmic connection. We were, I think, of a nomad people in North Africa when this happened so long ago […]

In this moment of open love, you might be able to believe what I say to you, that I do not have any misgiving or second thoughts about emotional involvements with you because of this vision of what happened so long ago. No, my only concern is, and this is very real to me, to free myself and to give you the possibility of freeing yourself, from these old, old bonds of emotional slavery which must not be repeated in this life. In this life, through our deep love, we have the real chance of changing this by bringing it out into the open. We have broken a karmic consequence and do no longer have to blindly bear the burdens of the past life and tragedy.

I am leaving Dolph and I will go to a place to begin a new life with myself. I do not think I will marry again. I must seek alone my true path of the soul. I love you very deeply and I go to live my own life, of which you are a wonderful spiritual part. Maybe it will come that you will be a material part as well. But now you must live the present as completely as you can.

> Shura, my dearest one, I want you to be free as a bird. Unfold your wings and leave all pain behind you, all possible accumulated guilt, all disquietness, all sorrows. Be free, and newly born, and walk into sunrise!!!
>
> Fly and be!
> Ursula

So much for ethnographic study number three.

IV.

All of this seems somewhat more dramatic than normal reality. I don't know how much liberty the Shulgins took when writing their autobiography – maybe this is another one of those things like Ann not knowing what "psychopharmacology" was. But one last thing I noticed about the book was how clear and coherent the psychology of everyone in *PiHKAL* was.

The Shulgins and everyone they know are Freudians – not explicitly, nobody ever says "I am a Freudian", but just on a deep level they assume that it's obviously true. Sometimes if it is an especially good day they're Jungians as well, in the same implicit way. They're always getting messages from their unconscious, they're always rediscovering psychologically repressed material, and they're always meeting people like Ursula who seem driven to behave in dramatic and unusual ways which are very predictable to any of the approximately one zillion psychoanalysts and psychiatrists whom the Shulgins know. Even their drugs are good analysts – the psychedelics are always helping them remember repressed childhood memories, after which they feel much better from whatever was bothering them at the time.

This is interesting, because I almost never see anyone behave in as dramatically Freudian a way as the Shulgins and their friends seem to behave all the time – even though I occasionally do psychodynamic therapy on people! I'm left a little baffled. Part of me wants to say that the primitive mind sees omens everywhere – I'm sure medievals were always seeing various signs of Christ in their daily life, and pattern-matching has never been a difficult sport (ask me how I feel about that brilliant and humane reflection on theodicy above being by someone named Ann). Another part wonders whether, if you're Freudian enough, your subconscious starts acting in Freudian ways just to keep up – although that *itself* is a Freudian idea and I'm not sure whether you can get it without presupposing Freudianism anyway. A third possibility is just that the more crazy drugs you're on, the more Freudian you act – wasn't Freud a coke fiend anyway? A fourth possibility is that the problem is with me – I'm somehow so closed to all this kind of thing that when people around me tell me Freudian stuff, I completely miss it without Ann Shulgin's narrative voice to tell me how Freudian it is – or even actively repress it.

Maybe the most interesting chapter of the book was where Ann had a spiritual crisis. She takes a psychedelic called DESOXY which Shulgin thinks is pretty weak, and she has a very strange reaction where the world starts seeming hostile, emotionless, and run by a perfectly rational demiurge that doesn't care about humans the tiniest bit. The psychedelic leaves her bloodstream, she reaches the point where she should feel normal again, but she can't shake her feeling of the demiurge's obvious and palpable presence, to the point where she becomes barely able to function. She goes to one of the zillion or so Jungian psychologists in her friend group, who matter-of-fact tells her she's having a spiritual crisis, and the only thing to do is wait for her soul to process it and gain the necessary enlightenment to go on (I wish I could get away with saying this kind of thing to *my* patients). Then she starts having extremely vivid visions of what is very obvious her mind doing Internal Family Systems therapy on herself, despite this being way before Internal Family Systems was invented, and despite the inventor being one of the three or so psychologists who was not a personal friend of the Shulgins. Finally she gets all the IFS steps right, accepts her parts, frees her repressed memories, and stops feeling like the Demiurge is harassing her at every moment. It's pretty fascinating, but that's just the thing – even though I've tried really hard to do Internal Family Systems on myself, armed with an official book and everything, I get nothing. I *never* have these sort of exciting spiritual crises that partake of exactly the right amount of symbolism from each of the world's great mystical traditions. I'm sort of jealous of all the people who do, and sort of suspicious of them. Maybe I need to take more psychedelics.

(I will note, though, that the book is appropriately skeptical about some of this. Shulgin describes going to a psychedelic conference and meeting an academic who worked in ethnobotany. He was studying a certain psychedelic plant, and was especially interested in why everyone who took that plant had hallucinations of jaguars in particular. He theorized that the plant was from the Mexican jungle, and that in some deep way our collective unconscious knew this, and so came up with appropriate hallucinations. Another psychedelicist who happened to hear the conversation interjected "I synthesized that chemical a little while ago, and all I got was wiggly lines." Shulgin left them as the first was getting increasingly agitated and demanding of the other whether he might have seen something that looked *kind of* like a jaguar.)

Speaking of weird things that Ann Shulgin sees, I’ll end this with something that might be interestingly testable. She writes that when she was young and going to sleep, every so often she would have a strange experience:

> Lying down for naptime (as a child) or at night for sleep, I would have reached that point of relaxation where one is not very much aware of the body…when I sensed it beginning (I never knew when it was going to come), I would immediately snap into alertness, excited and pleased, then I would just lie quietly as it unfolded…every part of it was the same each time. It was always in black and white…and I could never extend it, by so much as a few seconds. When it was finished, it was finished.
>
> First came the image-sensation after which I named the entire experience – the spiral. I felt my entire self drawn rapidly into a tiny point which kept shrinking, until it could shrink no further, at which time the microscopic point became a tunnel in which I continued traveling at great speed, inexpressibly small and implacably diminishing.
>
> Simultaneously, I was expanding. I was expanding to the edges of the universe, at the same tremendous speed as that of the shrinking, and the combination, the contraction-expansion, was not only an image, it was also a sensation the whole of me recognized and welcomed. This experience of myself as microcosm-macrocosm lasted exactly four minutes.
>
> The next stage came abruptly, as did all the changes. I was looking at standing figures which were vaguely human, dark thin figures being pulled into elongated shapes, like the sculptures of Giacometti. They

stretched out, arms and legs like black string, until it seemed they could elongate no further, then the scene changed and I was watching obscenely rounded bodies, Tweedledums and Tweedledees without costumes, their small heads and legs disappearing into their puffed, bloated flesh. The sensation accompanying this stage was one of discomfort, unpleasantness, a feeling of something grating on my soul. I once timed this part and the one that followed; they lasted a total of six minutes. I disliked them intensely.

Abruptly again, the inner screen became white, a horrible dead-white, nasty and aggressive like the underbelly of a sting-ray. After presenting itself for a few seconds, the flat white began to curdle from the outer edges into black, until finally the screen was totally black. A thick, awful, dead black, a pool of tar in an unlit cave deep underground. After another brief pause, the black began to curdle at its edges into the white again. The process repeated itself once, and the sensation was similar in every way to the previous one: irritating, grating, a feeling of unpleasantness that approached repugnance. I always endured it with a mental gritting of teeth, knowing it had to be gone through because that’s the way it always went and it was not to be changed.

And then, finally, I broke out into the last stage, the final part for which I had always been and always would be willing to undergo the middle parts. Now I was at the edge of an unseen cliff, looking out into a very different blackness, the deep, cradling blackness of the infinite universe, of space which stretched without end. I was completely happy and comfortable in that place, and would have stayed there indefinitely, had I been allowed, breathing in the beautiful darkness and the

> exquisitely familiar sense of infinity as a living presence, surrounding me, intimate and warm.
>
> After a moment of this pleasure, came the greeting. From the upper left-hand corner of the universe there came a greeting from Something which had known me, and which I had known, since before time and space began. There were no words, but the message was clear and smiling: Hello, dear friend, I salute you with respect-humor-love. It is a pleasure with laughter-joy to encounter you again…
>
> Then it was over. It had taken exactly twelve minutes.

According to Ann, she had this experience every so often “since I was born”, maybe once a week or so when she was a child, but becoming gradually less and less common until it finally happened for the last time when she was twenty-five. She said it was what got her interested in psychology and spirituality in the first place, and that her later access to psychedelics seemed to be sort of a substitute for the connection she had lost.

She *also* said that she met two people at cocktail parties who had exactly the same experience with exactly the same sequence of steps (though cutting out earlier). I am *sure* more people read Slate Star Codex than Ann Shulgin talks to at cocktail parties. So come on, people. Any of you ever have a very specific black and white vision of infinity?

Tolerance Troubles

[Not meant as a claim that science doesn't know something. More of an admission that I don't know some things, and a hope to be informed about them by someone who does.]

Everyone knows about tolerance. The first time you take heroin, you get really high. The second time you take heroin, you get slightly less high. The nth time you take heroin, you barely feel good at all – but if you *stop* taking heroin, then you feel miserable. Your body adjusts, the receptors desensitize, whatever.

This is so simple that it took me forever to ask the obvious next question – how come this doesn't happen for everything else? Supposedly if you have ADHD you can just stay on Adderall forever. Nobody says "The first time you take Adderall you can concentrate really well, the second time you take it you'll concentrate less well, and the nth time you take it you can barely concentrate better at all."

The psychiatry textbooks contain a sentence or two saying that "some" patients "may" develop Adderall tolerance, but it's not something that we're trained to expect. There are a lot of anecdotal reports online, but there are also anecdotal reports of people who don't develop any tolerance at all after years and years. Hmm.

Also, people who *abuse* Adderall develop tolerance all the time, and keep having to up the doses just like heroin abusers do. This is a little weird – my pet theory is that people only develop tolerance to drug effects that aren't FDA-approved uses – though how your receptors know what the FDA says I haven't quite figured out. More seriously, it may be an effect of method of use – taking a small amount daily versus snorting a large amount of crushed tablet whenever you feel like it. Or it may be that tolerance to euphoric effects is worse than tolerance to stimulant effects.

(This seems true in general – I get euphoric effects from caffeine when I drink it very rarely; if I drink it more often, the euphoria goes away and I just feel a little more awake. This is important since it suggests tolerance isn't just your body metabolizing the drug better, but actually a matter of receptor-level action – something I think everyone agrees is true, but which it's always nice to have independent confirmation for.)

Sometimes tolerance gets weird. Antipsychotics are supposed to block dopamine receptors. Too much dopamine can contribute to psychosis, but it can also screw up the basal ganglia's modulation of movement and cause you to make repetitive jerking motions all the time. People who have been on antipsychotics for too long may remain protected from psychosis, but start making repetitive jerking movements in a way consistent with *too much* dopamine. The theory goes that the receptors involved in psychosis haven't developed tolerance (for some reason), but the receptors involved in movement modulation have developed so much tolerance that they've overshot their baseline and become supersensitive to dopamine. So by taking a drug that lowers dopamine, you get higher dopaminergic effects. In the worst case scenario, you end up with a condition called tardive dyskinesia, which is permanent. The receptors stay supersensitive forever and you will always make repetitive jerking movements. If you stop the antipsychotic, that will just make it (temporarily) worse – now you have supersensitive dopamine receptors *and* you're not blocking them, so that means *lots* of repetitive jerking movements.

In this case, giving someone a drug has caused them to develop not just tolerance but supertolerance, where they are permanently worse than before. It’s as if taking heroin for long enough made you permanently miserable.

…which actually isn’t totally hypothetical. Some percent of people who abuse opiates like heroin get what’s called a post-acute withdrawal syndrome (PAWS), meaning that they feel depressed for months or years after they stop using the opiates. I treated a patient like for a while. I tried pretty much every antidepressant on him without success. He was just miserable. He’d been clean for about six months and I told him that he might just have to wait it out – it usually goes away after a few months to a year or so.

There is a poorly-studied but anecdotally very helpful treatment for PAWS: low-dose naltrexone. Naively, this sounds like the stupidest possible thing to try. It’s an opiate antagonist, meaning that you’re taking somebody who is undersensitive to opiates and blocking the tiny number of functioning opiate receptors they already have. It should be the only thing capable of making this already bad condition worse. Yet people swear by it. The theory is supposed to be the tolerance reaction again. Your body reacts to this opiate-blocking agent by releasing more opiates. So we’re treating a condition in which drugs that increase opiates cause you to have fewer opiates, by giving you a drug that decreases opiates which will cause you to have more opiates. How annoying is that?!

(some people recommend that if you’re giving someone opiate painkillers, you can give them low-dose naltrexone at the same time to prevent development of tolerance. Giving someone an opiate and an opiate-blocker simultaneously seems kind of like the medical equivalent of digging holes and filling them back in again, but apparently it does something useful)

This means we have examples of all three of the following:

1. A drug that's supposed to have effect X, and after a while it still has effect X (Adderall)
2. A drug that's supposed to have effect X, but after a while it has no effect (heroin)
3. A drug that's supposed to have effect X, but after a while it overshoots and has effect anti-X (Antipsychotics? Heroin? Naltrexone?)

You may notice that these are all three of the logical possibilities. So for example, if we gave someone a drug that was supposed to decrease anxiety, it might decrease anxiety, have no effect, or increase anxiety. If scientific hypotheses are about closing off parts of possibility-space, then the receptor sensitivity hypothesis isn't doing a very good job.

But it's actually worse than this, because I get the impression that different people will end up in different branches of this trilemma. Benzodiazepines are a special offender here. Some people can take Xanax once a day for anxiety, and it's a perfect solution – it suppresses their anxiety, it never stops working, and they never become addicted – if twenty years later they get a good therapist who helps them treat their anxiety without drugs, they can stop the Xanax with just a couple of days of mild discomfort. Other people will lose all effect after a couple of weeks, up the dose, up the dose some more, and end up as total wrecks. I think this is much less common than most people say – my attending's rule of thumb is "benzo tolerance develops for sleep but not anxiety" – but it certainly happens. And for that matter, I've met a few people who never seem to develop tolerance for benzodiazepine sleeping pills. You see this same pattern for opiates used as painkillers. I spent *so many years* confused about whether people develop tolerance to these or not, and my final conclusion is that some people do and some people don't and if you try to find a coherent universal pattern here you will go insane.

And it's actually worse than *this*. Drugs don't just work differently in different people, sometimes the same person will cycle through totally different mechanisms of drug response. SSRIs have something called tachyphylaxis, where they'll work really well for months and then suddenly stop working (the word means "fast protection", ie you develop protection against the drug effects quickly). This is even more annoying than the other patterns – at least with heroin, it makes sense that the receptors will gradually lose their sensitivity. But here? In random people at random times, the drug just stops working suddenly. It might be after a month, it might be after a year, it might be after ten years. And then every so often you'll try the drug again a decade later, and then it'll work just fine. Why? Nobody knows.

Some skeptics have pointed out that this is exactly what you would expect if the drug had no real effect and it was just luck that people didn't have depressive episodes while they were taking them, but we know SSRIs have some effect. And anyway, placebo tachyphylaxis isn't any less weird than real tachyphylaxis.

One more weird thing: LSD users report very strong tolerance lasting about three days after a dose, to the point where a second dose the day after will do almost nothing. Rat experiments have shown this is definitely because of receptor downregulation and not just enzyme induction.

Okay. But LSD is a pretty strong drug. If receptors are so down-regulated that you are essentially on negative one tabs of LSD, how are you remotely normal while the tolerance is in effect? Are people during their periods of LSD tolerance less crazy and creative than normal? What the heck is the 5-HT2A receptor even *doing* if decreased amounts of it sufficient to render LSD ineffective don't have noticeable effects on consciousness?

This is my concern about naltrexone as well. Sometimes doctors give naltrexone to help with alcohol addiction, which usually works okay. The theory is that since naltrexone blocks opiates, and opiates power the endogenous reward system, alcohol will seem less rewarding. Fine. But shouldn't *everything* seem less rewarding? I always worry that I'm just blocking my alcoholic patients' ability to enjoy anything at all (of which enjoying alcohol is a subset), but that doesn't seem to be how it works. This is about when I default to my theory of "receptors read the FDA labels for medications and make sure to only do what they're supposed to".

All of this annoys me for a few reasons.

First, psychiatry *really* doesn't think about this enough (or sometimes at all). The pharmacology textbooks will tell you how effective a drug is, how long it lasts, how many side effects it has, et cetera, but not whether it's going to produce tolerance or not. It's mostly just assumed that it won't.

Second, groups skeptical of psychiatry are always talking about tolerance and it's hard to tell whether they're right or wrong. For example, some people claim antidepressants cause tardive dysphoria – that like the antipsychotics that eventually give permanent repetitive jerking movements, antidepressants can make serotonin receptors permanently undersensitive (or something) and so make depression worse. Other people say that antipsychotics themselves can eventually screw up dopamine receptors in ways that make psychosis worse (though see here). My *guess* is that these problems don't arise for most people, but I can't explain *why* these things wouldn't happen.

Third, I think something like this is involved in addiction. Addiction is highly genetic; some people can drink alcohol socially their entire lives and never become alcoholic; other people quickly get hooked. This seems related to the thing where some people are stable forever at their low dose of opiate painkiller, and other people quickly develop tolerance and need to keep increasing it. I'm sure there are other things involved in addiction, but this is probably one of them.

Fourth, how many interesting things are we missing because they're stupid and make no sense? I don't know who first discovered that low-dose naltrexone could help potentiate the effect of opiates, but there have got to be other things like that. Forcing your body to become more sensitive to its own chemistry seems like a good alternative to forcing more and more foreign chemicals into it.

Finally, the best drugs seem to be the ones we hesitate to use because they produce too much tolerance. Xanax, opiates, you name it. A version of Xanax that that didn't produce any tolerance would be a holy grail of anxiety disorder pharmacology. Some way to switch off Xanax tolerance would be just as good.

And a tolerance-free version of heroin would be pretty interesting too – from a purely pharmacological perspective, of course.

The View From Ground Level

[Epistemic status: Any time I make an anthropic argument, you should probably interpret it as trolling]

Sean Carroll argues that the simulation argument is false.

The simulation argument posits two kinds of universes: "high-level" universes that can simulate other universes, and "ground-level" universes that can't. By the terms of the simulation argument itself, most universes will be at ground level, since every high-level universe can simulate many ground-level ones. So (says the argument) we should expect to be at ground level. But the simulation argument itself hinges on our observation that it looks like our universe is capable of simulating other lower-level universes. So apparently we aren't on ground-level. So the simulation argument is probably false.

(I might be summing it up badly. Read the actual post for more.)

Suppose Carroll's reasoning is right. What would a ground-level universe look like?

It would have to be pretty weird. It would have to ban the creation of Turing machines – since with enough time and resources any Turing machine could be expanded into a full-scale simulation. But Turing machines are pretty simple, and brains supporting conscious observers are pretty complicated. To have conscious observers but not Turing machines – well, once again, this would have to be pretty weird.

Brains would have to run off a science different from the local science accessible to in-universe researchers. Probably they would be run remotely, in the simulating universe, and then the results beamed into the simulated universe with no regard for the computational rules of the simulation. Maybe an alien dissecting a fellow alien's head would just find a perfectly featureless crystal with no internal structure, which is observed to inexplicably send nerve impulses to the rest of the entity's body. Such aliens might invent psychology, but never neuroscience, and even if they speculated about it, it wouldn't matter – attempts to "simulate" neurons would fail, their workings forever beyond locally accessible physics. Even if they completely mastered their local science, their brains would remain a mystery.

I used the phrase "conscious observers" above. There are versions of anthropics that work for p-zombies, but we're not p-zombies and we don't have to use them; we can do anthropics conditioning upon consciousness. Try *that*, and the simulation argument doesn't exactly depend on a ground-level universe where further simulations are impossible. It depends on a ground-level universe where further simulations *containing conscious observers* are impossible.

This changes the scenario a bit. Now people in ground-level simulations can expect arbitrarily complex physics, physics that allow the creation of as many Turing machines as they want, but which can't possibly explain consciousness. They should be able to master every aspect of the universe around them *except* consciousness, which try as they might will remain refractory to their simulations. Consciousness will make perfect sense in the physics of the universe above theirs, but the simulators will have excised all consciousness-related rules from the ground-level sim. Try as the simulated scientists might, it'll remain a mystery.

If Carroll’s deconstruction of the simulation argument is right, then the more trouble we have explaining consciousness, the more that should push us to believe we’re in a ground-level simulation. There’s probably a higher-level version of physics in which consciousness makes sense. Our own consciousness is probably being run in a world that operates on that higher-level law. And we’re stuck in a low-resolution world whose physics doesn’t allow consciousness – because if we weren’t, we’d just keep recursing further until we were.

Devoodooifying Psychology

[Epistemic status: very low. Total conjecture based on insufficient evidence.]

“Voodoo death” refers to supposed cases where people died after being cursed by witch doctors. The theory goes that even though witch doctors don’t have real magic, if their victims come from a culture that believes in witchdoctory then they’ll be so scared that they gradually waste away out of fear and die anyway.

For a while psychologists believed that this absolutely happened, a testament to the powers of the mind. Maybe it was because of toxic levels of the stress hormone adrenaline or something.

Now there’s some more controversy. A lot of these cases turned out to be primitive tribesmen who said it totally happened to a friend of a friend of a cousin or something. In others, witch doctors placed curses on people who already had some kind of serious disease, then took credit. In still *others*, the curse victim became so upset that they stopped eating or drinking and died of dehydration – which, while technically a death, doesn’t really testify to the powers of the mind so much as the power of blood osmolality. And after some more thought, everyone agreed the adrenaline theory probably didn’t apply since adrenaline spikes kill suddenly but voodoo victims waste away over the space of weeks.

So now voodoo death looks a lot more complicated than a simple progression of curse -> adrenaline -> you were killed by *your own mind*. Lester gives a relatively sympathetic view of the evidence here, but even he can only find two good cases – one of which involved a patient who died of preexisting asthma, the other of which did not involve voodoo at all.

I find this interesting because so much of psychology seems basically voodoo-ish.

Take the placebo effect. This is basically the voodoo effect in reverse. Instead of a witch doctor saying you'll get worse, and you do, a regular doctor tells you you'll get better, and you do. For a while, people were claiming all *sorts* of amazing effects for placebo – placebo can activate the immune system to fight infections, placebo can slow the growth of cancer, placebo can make bedridden invalids start dancing jigs. But the best studies now suggest that the placebo effect is probably very weak and limited to controlling pain. The vaunted power of mind over body, of belief over reality, doesn't look nearly as impressive as we thought.

Or take stereotype threat. Again, this is *sort of* a voodoo curse. If people make you think you're going to do bad on a test, then you'll do bad on the test. Again, widely believed, held up as an example of the power of perception. Again, doesn't replicate well in large studies, has a very suspicious funnel plot, and is starting to inspire doubt even among top researchers in the area.

Or take self-esteem. Again, a reverse voodoo curse. By believing that you're a good person and likely to do well, good things will happen to you. Again, very popular in the '90s, but it hasn't aged well. Similarly, self affirmations failed to replicate results showing their effectiveness.

This is all oversimplified; there are still lots of unrebutted studies supporting all of these things. Maybe some of the studies that seem to debunk them have themselves been debunked, and I don't know about it. Still, it seems to me that things that sound like voodoo – that is, which argue that our optimistic or pessimistic beliefs about how well we will do can mysteriously and directly *affect* how well we will do – are faring unusually badly as psychologists get better at trying to replicate things.

(This kind of thing is why I'm so skeptical of growth mindset, despite so little hard data supporting my skepticism)

This isn't to say that *nothing* like this is true. At the very least, my belief that I can't swim the English Channel makes me not try to swim the English Channel, and so if that belief is wrong it's voodoo-ish-ly making me fail at Channel-swimming. But this is a lot less mysterious than the thing where you repeat "I will do well at school" every day and so do better at school.

Let me introduce a second category of things.

First, the name preference effect. This is where you're positively predisposed to things that sound like your name. For example, people named Dennis are more likely to become dentists, people named Bob are more likely to be bakers or butchers or barristers, and people with three letters in their name are more likely to live in Three Forks. People believed this for *years* until finally somebody did a statistical reanalysis and found that it was totally false.

Second, unconscious social priming. Supposedly people who heard the word "retirement" walked more slowly for a while afterwards, because "retirement" primed their thoughts of old people, and old people primed their thoughts of being slow, and so for a while they themselves behaved like an old person. This sort of thing inspired an entire field of psychology showing similar results (like the infamous study showing that an earthquake increased divorce rates by priming thoughts of instability), but it very much failed to replicate and is now the archetypal example of a formerly-accepted finding now believed to be false. Dr. Primestein's work in this area is also a must-read.

Third, and from just last month: Artificial surveillance cues do not increase generosity: two meta-analyses. You know how if there was a picture of eyes or something, people would be nicer and more law-abiding, because deep down they felt like they were being watched? Yeah, turns out that's not true.

And again, don't trust me too much – there are a lot of studies I could have mentioned here, and three don't necessarily make a pattern. But if I'm right that these are representative examples, then they seem to share a pattern. I'm not sure they all technically qualify as "social priming" (though they might, and that field has already been pointed out as particularly bad) but they all have the same feeling of tiny cues that you don't think about causing big unconscious changes to behavior. This seems to be another category that is faring unusually badly.

Implicit association tests probably don't work (1, 2, 3, 4). That is, people who have "implicit racial biases" according to the tests are not more racist in everyday life than people who don't. If this were true – and if it reflected a general failure of implicit racial biases to affect explicit actions – it's hard to overestimate how much it would change psychology. We wouldn't have to worry about how the wrong character on TV would accidentally bias people toward having certain stereotypes. We wouldn't have to worry about subconscious racism affecting hiring decisions even among people who are trying hard to be fair and neutral.

Does this fall into the previous patterns? It’s not *exactly* about self-fulfilling prophecies, or tiny stimuli having oversized effects on behavior. But it seems to have a certain kinship with them.

And a few days ago, a friend posted a quote on Tumblr:

> There is no sovereign sanctuary within ourseles which represents our real nature. There is nobody at home in the internal fortress. Everything we cherish as our ego, everything we believe in, is just what we have cobbled together out of the accident of our birth and subsequent experiences. With drugs, brainwashing, and other techniques of extreme persuasion, we can quite readily make a man a devotee of a different ideology, the patriot of a different country, or the follower of a different religion

I was only too happy to be able to reply with Gwern’s research on how “brainwashing” mostly doesn’t work. Does this suggest the post is wrong about the lack of a “real nature”? Does this relate to any of the previous patterns?

A single thread seems to run through all of these examples: a shift away from the power of the unconscious. The unconscious doesn’t make you succeed or fail proportionately to your belief in yourself. The unconscious doesn’t change your behavior because of insignificant environmental cues. The unconscious doesn’t make you racially discriminate despite your own better nature. The conscious mind is strong enough to hold onto its preferred beliefs despite brainwashing techniques intended to force it otherwise.

So maybe we should update in general towards less of a role for the unconscious mind?

I remember my first freshman psychology class. After studying learned helplessness, I realized that was what I was feeling: study after study of crazy things, everything depended on your beliefs, the first letter of your name could affect your life outcome, stuff like that. The end result was a lot like in the quote above – nobody has any control over their lives, we're all at the mercy of vast unconscious forces, society's hidden assumptions and bases are vital in shaping our future. What if all of that was wrong? What if people mostly make decisions based on reasonable factors, succeed or fail based on things like ability or random luck, and social assumptions are relatively powerless beyond a common sense level? Wouldn't that be *great*?

But maybe this is going too far. Once again, all I have is a few data points, curated by my own biases. Certainly not all of the studies showing creepy unconscious effects have been disproven; probably fewer than 10% of them have. Certainly not all the studies that have been disproven show creepy unconscious effects – ego depletion is a very mechanistic biological idea, but it's done no better than priming. You have to kind of squint to see the pattern, then take it on faith that it's real and that it continues throughout all the data that haven't been checked.

And how do you tell the baby from the bathwater here? Some results – like cognitive biases, sales techniques or eyewitness unreliability – sort of fall under the heading of "power of unconscious effects", but seem subtly different – maybe because of a less agentic unconscious? I don't know. But I would be surprised if those followed the same pattern.

(on the other hand, a first draft included the Asch conformity experiments in that list, but apparently those *never* said what I thought they did)

Of course, this post is *really* about Freudian psychology. When I presented a Freudian friend with information on the general irrelevance of childhood factors and family composition, he countered by saying that the only thing that could really harm his belief in psychoanalysis was to learn that the unconscious wasn't very powerful.

I don't think we're anywhere close to there yet. And I don't think it's meaningful to "deny the unconscious" – the unconscious is everything that happens except for conscious stuff, and that seems like a lot. But maybe our concept of the unconscious, or certain things that we attributed to the unconscious, was overly broad. And I think there's an interesting project in trying to make explicit exactly what that means and what sort of smaller concepts we can get away with.

Reverse Voxsplaining: Drugs vs. Chairs

[Content note: this is pretty much a rehash of things I've said before, and that other people have addressed much more eloquently. My only excuse for wasting your time with it again is that SOMEHOW THE MESSAGE STILL HASN'T SUNK IN. Pitching this as "market" vs. "government" is overly simplistic, but maybe if I am overly simplistic sometimes then it will sink in better.]

EpiPens, useful medical devices which reverse potentially fatal allergic reactions, have recently quadrupled in price, putting pressure on allergy sufferers and those who care for them. Vox writes that this "tells us a lot about what's wrong with American health care" – namely that we don't regulate it enough:

> The story of Mylan's giant EpiPen price increase is, more fundamentally, a story about America's unique drug pricing policies. We are the only developed nation that lets drugmakers set their own prices, maximizing profits the same way sellers of chairs, mugs, shoes, or any other manufactured goods would.

Let me ask Vox a question: when was the last time that America's chair industry hiked the price of chairs 400% and suddenly nobody in the country could afford to sit down? When was the last time that the mug industry decided to charge $300 per cup, and everyone had to drink coffee straight from the pot or face bankruptcy? When was the last time greedy shoe executives forced most Americans to go barefoot? *And why do you think that is?*

The problem with the pharmaceutical industry isn't that they're unregulated just like chairs and mugs. The problem with the pharmaceutical industry is that they're part of a highly-regulated cronyist system that works completely differently from chairs and mugs.

If a chair company decided to charge $300 for their chairs, somebody else would set up a woodshop, sell their chairs for $250, and make a killing – and so on until chairs cost normal-chair-prices again. When Mylan decided to sell EpiPens for $300, in any normal system somebody would have made their own EpiPens and sold them for less. It wouldn't have been hard. Its active ingredient, epinephrine, is off-patent, was being synthesized as early as 1906, and costs about ten cents per EpiPen-load.

Why don't they? They keep trying, and the FDA keeps refusing to approve them for human use. For example, in 2009, a group called Teva Pharmaceuticals announced a plan to sell their own EpiPens in the US. The makers of the original EpiPen sued them, saying that they had patented the idea epinephrine-injecting devices. Teva successfully fended off the challenge and brought its product to the FDA, which rejected it because of "certain major deficiencies". As far as I know, nobody has ever publicly said what the problem was – we can only hope they at least told Teva.

In 2010, another group, Sandoz, asked for permission to sell a generic EpiPen. Once again, the original manufacturers sued for patent infringement. According to Wikipedia, "as of July 2016 this litigation was ongoing".

In 2011, Sanoji asked for permission to sell a generic EpiPen called e-cue. This got held up for a while because the FDA didn't like the name (really!), but eventually was approved under the name Auvi-Q, (which if *I* were a giant government agency that rejected things for having dumb names, would be going straight into the wastebasket). But after unconfirmed reports of incorrect dosage delivery, they recalled all their products off the market.

This year, a company called Adamis decided that in order to get around the patent on devices that inject epinephrine, they would just sell pre-filled epinephrine syringes and let patients inject themselves. The FDA rejected it, noting that the company involved had done several studies but demanding that they do some more.

Also, throughout all of this a bunch of companies are merging and getting bought out by other companies and making secret deals with each other to retract their products and it's all really complicated.

None of this is because EpiPens are just too hard to make correctly. Europe has eight competing versions. But aside from the EpiPen itself, only one competitor has ever made it past the FDA and onto the pharmacy shelf – a system called Adrenaclick.

And of course there's a catch. With ordinary medications, pharmacists are allowed to interpret prescriptions for a brand name as prescriptions for the generic unless doctors ask them not to. For example, if I write a prescription for "Prozac", a pharmacist knows that I mean anything containing fluoxetine, the chemical ingredient sold under the Prozac brand. They don't have to buy it directly from Prozac trademark-holder Eli Lilly. It's like if someone asks for a Kleenex and you give them a regular tissue, or if you suggest putting something in a Tupperware but actually use a plastic container made by someone other than the Tupperware Corporation.

EpiPens are protected from this substitution. If a doctor writes a prescription for "EpiPen", the pharmacist must give an EpiPen-brand EpiPen, not an Adrenaclick-brand EpiPen. This is apparently so that children who have learned how to use an EpiPen don't have to relearn how to use an entirely different device (hint: jam the pointy end into your body).

If you know anything at all about doctors, you know that they have *way* too much institutional inertia to change from writing one word on a prescription pad to writing a *totally different* word on a prescription pad, especially if the second word is almost twice as long, and *especially* especially if it's just to do something silly like save a patient money. I have an attending who, whenever we are dealing with anything other than a life-or-death matter, just dismisses it with "Nobody ever died from X", and I can totally hear him saying "Nobody ever died from paying extra for an adrenaline injector". So Adrenaclick continues to languish in obscurity.

So why is the government having so much trouble permitting a usable form of a common medication?

There are a lot of different factors, but let me focus on the most annoying one. EpiPen manufacturer Mylan Inc spends about a million dollars on lobbying per year. OpenSecrets.org tells us what bills got all that money. They seem to have given the most to defeat S.214, the "Preserve Access to Affordable Generics Act". The bill would ban pharmaceutical companies from bribing generic companies not to create generic drugs.

Did they win? Yup. In fact, various versions of this bill have apparently failed so many times that FDA Law Blog notes that "insanity is doing the same thing over and over again and expecting different result".

So let me try to make this easier to understand.

Imagine that the government creates the Furniture and Desk Association, an agency which declares that only IKEA is allowed to sell chairs. IKEA responds by charging $300 per chair. Other companies try to sell stools or sofas, but get bogged down for years in litigation over whether these technically count as “chairs”. When a few of them win their court cases, the FDA shoots them down anyway for vague reasons it refuses to share, or because they haven’t done studies showing that their chairs will not break, or because the studies that showed their chairs will not break didn’t include a high enough number of morbidly obese people so we can’t be *sure* they won’t break. Finally, Target spends tens of millions of dollars on lawyers and gets the okay to compete with IKEA, but people can only get Target chairs if they have a note signed by a professional interior designer saying that their room needs a “comfort-producing seating implement” and which absolutely definitely does not mention “chairs” anywhere, because otherwise a child who was used to sitting on IKEA chairs might sit down on a Target chair the wrong way, get confused, fall off, and break her head.

(You’re going to say this is an unfair comparison because drugs are potentially dangerous and chairs aren’t – but 50 people die each year from falling off chairs in Britain alone and as far as I know nobody has ever died from an EpiPen malfunction.)

Imagine that this whole system is going on at the same time that IKEA spends millions of dollars lobbying senators about chair-related issues, and that these same senators vote down a bill preventing IKEA from paying off other companies to stay out of the chair industry. Also, suppose that a bunch of people are dying each year of exhaustion from having to stand up all the time because chairs are too expensive unless you’ve got really good furniture insurance, which is totally a thing and which everybody is legally required to have.

And now imagine that a news site responds with an article saying the government doesn’t regulate chairs enough.

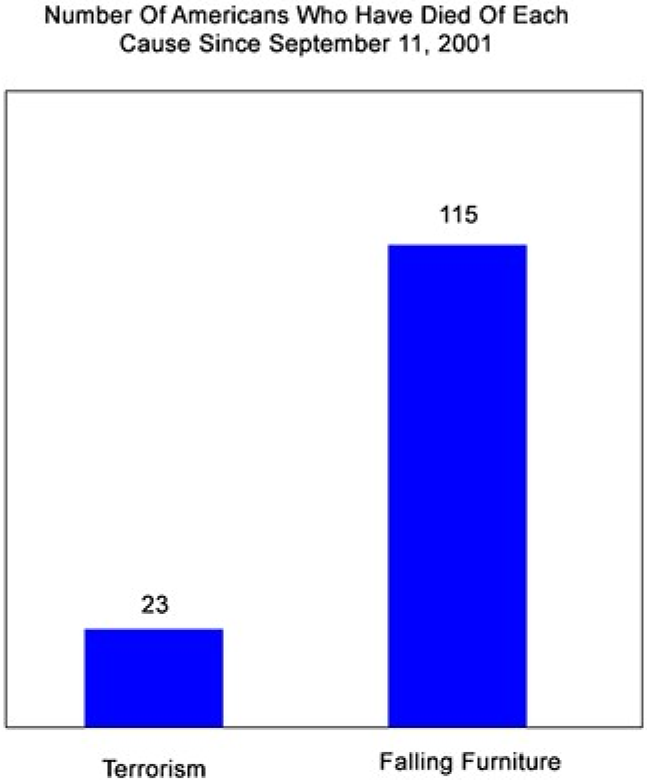
Terrorists Vs. Chairs: An Outlier Story

The other day I needed to know how many people were killed by chairs, and while searching I came across the Washington Post's You're More Likely To Be Fatally Crushed By Furniture Than Killed By A Terrorist. It argues that worrying about terrorism is irrational, because terrorists kill fewer people each year than falling furniture, and nobody cares about *that*:

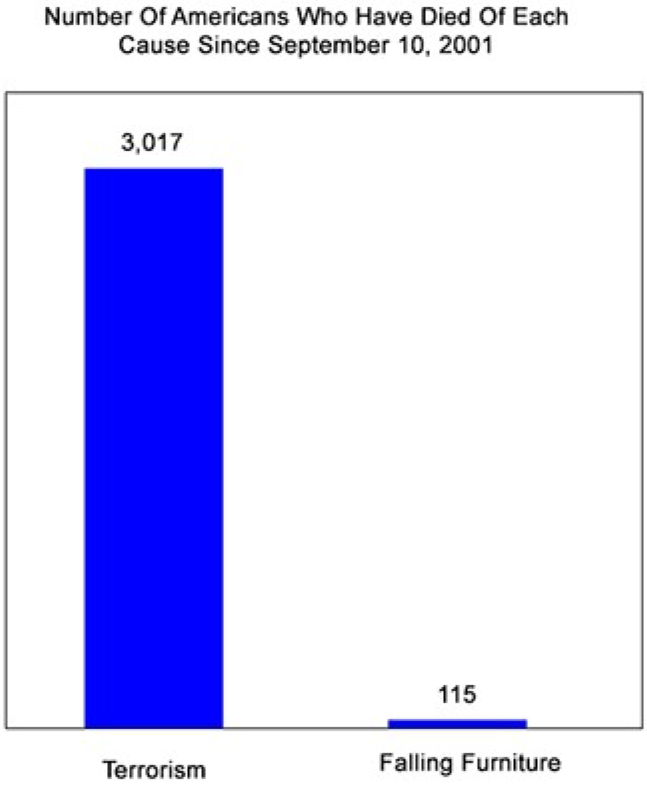
> Consider, for instance, that since the attacks of Sept. 11, 2001, Americans have been no more likely to die at the hands of terrorists than being crushed to death by unstable televisions and furniture […] What accounts for the fear that terrorism inspires, considering that its actual risk in the United States and other Western countries is so low? The answer lies in basic human psychology.

I once saw the *perfect* response to this kind of argument on Twitter, but I forgot to screenshot it, so I'll have to try to draw it from memory here.

One person posted a graph that looked something like this:



And somebody else edited it to look like this:



And *whoa* I had never realized before how sketchy it is to start your interval for recording the average number of terrorist attacks *the day after* the last major terrorist attack.

I mean, I know why people do it. It’s because September 11 was an “outlier”, and outliers should not be counted. Problem is, depending on your distribution, sometimes “outliers” are the only thing that matters.

Let me give an example. Suppose I’m trying to make an argument that earthquakes are totally not a problem for Haiti at all, that there’s no need to invest in earthquake preparedness, and that Haitian people who worry about earthquakes are stupid. I make a graph showing that since January 13, 2010, fewer Haitians have died per year from earthquake-related causes than from crazy furniture-related mishaps. This is totally 100% true. Look at those stupid

Haitians, worrying about something that *on average* never hurts anybody!

(the Haitian earthquake of January 12, 2010 killed about 100,000 people)

I’m sure there are a zillion small Richter 1.0 and Richter 2.0 earthquakes in Haiti all the time. I’m sure our monitoring interval of January 13, 2010 to present picked up lots of these and correctly noted that they don’t kill anybody. The *only* Haitian earthquakes anyone needs to worry about are the outliers.

If you start your monitoring interval on January 13, earthquakes kill 0 people/year. If you start it on January 11, earthquakes kill 20,000 people a year. Neither of these is entirely fair – one is purpose-designed to maximize casualties, the other to minimize it. I don’t think there’s an obvious fair way to do things – the best solution would be extend the interval back to infinity, but then you get into problems like Haiti having fewer people back in the day, or Haiti not having risen out of the sea yet back in 4,000,000,000 BC. Maybe the best solution is to pick an arbitrary block of time like “the last fifty years”, or to do something very complicated like using the remote historical record to produce earthquake numbers and then combine it with modern populations to produce expected casualties.

The same is true of September 11. Start the interval September 12, and you get 5-10 terrorism deaths/year. Start it September 10, and you get 200. I don’t know when the *best* time to start it would be. If I had to choose *something*, I would say maybe 1985, when jihadist terrorism got started after the Soviet invasion of Afghanistan. But someone else could choose 1776, or 2000, based on similarly arbitrary criteria. And it would all be irrelevant – September 11 either made terrorists more ambitious, made security forces more watchful, or both, and so probably changed the calculus for good.

Granted, even when you include September 11, terrorism isn't the *worst* thing, and people probably do overestimate it. So forget terrorism. On average, the flu kills something like 20,000 people worldwide each year. That's a lot, but not *apocalyptically* much. If you go back year after year, the average stays at something like 20,000/year, right up until you get to 1918, when about 100,000,000 people died. So flu deaths over the last century average about 1 million/year. But three years from now, average flu deaths over the last century will average about 20,000 year. A death rate of only 20,000/year might make our current efforts to contain the flu seem excessive compared to other diseases. But a death rate of 1 million/year makes them look if anything the opposite.

Even worse: did you know that giant asteroids kill about a hundred people per year, on average? This is admittedly an odd definition of "kill" and "average" given that no human being has ever been killed by a giant asteroid. But given that giant asteroids strike Earth about every ten million years, and an asteroid strike today might kill about a billion people, *on average* giant asteroids kill about a hundred people per year.

Actually, un-forget terrorism. I have a friend who is *very* in favor of the War On Terror, and he argues that the problem with terrorism isn't the average suicide bomber who kills three people. It isn't even the 9-11 hijackers who killed three thousand people. It's the group that steals a nuke and kills three million people. Just as "on average" a hundred people die each year from giant asteroid strikes, maybe "on average" thirty thousand people die each year from nuclear terrorism. All you'd need for this to be true is one nuclear attack per century. And that's *as bad as an average flu season*!

The thing about falling furniture is that there's probably not going to be a furniturepocalypse where suddenly millions of people all perish at once after being struck by a really really big desk. Furniture is constant. Terrorism isn't. The whole *point* of black swans is that we pay too much attention to constant risks and ignore the outliers, *especially* the outliers which outlie so far that they haven't happened yet. That's true whether it's terrorism, earthquakes, pandemics, or AI.

I worry that someday many years from now, terrorists are going to have some improbable victory which is even more destructive than September 11. I worry that uncounted people are going to die. And I worry that ten years later, someone is going to post on Facebook about how "From the day after ISIS nuked London through today, on average fewer people per year have died of terrorism than from hilarious accidents involving bedside dressers!"

Reverse Voxsplaining: Brand-Name Drugs

[Epistemic status: Uncertain, especially on the accuracy of the economic studies cited]

I.

Sarah Kliff of Vox replies to my piece from last week. She writes:

> Earlier this week I wrote about how a lack of drug price regulation in the United States allows pharmaceutical companies — including EpiPen’s manufacturer, Mylan — to charge exceptionally high prices for their products.
>
> Scott Alexander at the blog Slate Star Codex argues that I’ve got it all wrong: The problem isn’t a lack of price regulation. Instead, its too much regulation, which has prevented generic competitors from entering the market and has left EpiPen’s price so high. […]
>
> Alexander and I are both essentially pointing out two examples of how the United States has created a regulatory system that is incredibly favorable to pharmaceutical companies. We’ve set up a system that makes it incredibly easy for drug companies to score high profits and charge exceptionally high prices for their products.
>
> One way it’s favorable is that we let drug companies pick their own prices — in this way the United States is exceptional, as the vast majority of developed companies regulate their drug prices. Another way we’ve created a favorable regulatory environment, as

> Alexander writes, is by allowing roadblocks to stand in the way of generic drug makers who want to enter.
>
> More generics can help America’s drug spending problem. But they can’t solve it.
>
> Greater use of generic drugs is widely accepted as a way to drive down drug spending. The FDA has found that drug prices decline to 55 percent of the brand-name price when two generic manufacturers begin making a product. Right now, the United States already uses a lot of generic drugs. In fact, about 90 percent of drugs prescribed in the country right now are generic. Brand name drugs are the reason that America has higher per-capita drug spending than other countries. Brand-name drugs make up just 10 percent of prescriptions filled in the United States, but account for 72 percent of drug spending.
>
> Drugs that are under patent are the true source of high American drug costs. EpiPen is, in a way, a bit of a red herring […]
>
> Harvoni, a [patented] pill that cures Hepatitis C, costs $32,114 here — and $22,554 in the United Kingdom. Less red tape around generic drug competition wouldn’t really change that fact. As long as we’re going to have patented drugs, letting drug manufacturers set their own prices will remain a key driver of America’s higher drug spending.

(the above is some excerpts stitched together to provide a taste; you should really read the whole thing)

First of all, thanks to Ms. Kliff and Vox for responding to me; it’s always neat to get featured in real news sources as if I were a real writer or something instead of just a guy with a blog. Additional thanks for a very measured and charitable tone despite my own tendency toward snarkiness. I worry that I am being unfair by acting snarky to somebody who is writing for a professional publication and so is not allowed to be snarky back; if so, I’m sorry and will try to control myself as best I can, which admittedly is not very good.

That having been said, I’m super against all of this and think it’s totally wrong.

I would kind of like to complain about Vox calling EpiPen a “red herring” when they were the ones who brought it up, but I think the problem is deeper than that. Discussing generic drug costs is completely different from discussing brand name drug costs, and the two issues have very different arguments around them. To transition fluidly from one to the other, saying we need more price controls for generics and then backing up your argument by saying that you were mostly thinking of brand-names after all – elides a distinction which is the heart of this entire subject.

Generic drugs are overpriced because we’re morons who can’t come up with a decent regulatory regime. Brand-name drugs are overpriced because of a deliberate decision to overprice them to encourage research.

The economic argument goes: the more profitable new drugs are, the more incentive a company has to make them. If we didn’t reward pharmaceutical companies for inventing new drugs, then they wouldn’t go through the $2.5 billion, ten-year hassle of seeking FDA approval with no guarantee of success. The way we reward them is by giving them a twenty-year monopoly when they can charge lots of money without anybody telling them not to.

(this isn’t quite right. The law says a twenty-year monopoly, but it’s dated from the time the drug is invented. Since it takes ten years to go through the FDA, it’s effectively more like a ten-year monopoly on actually selling the drug)

The reason I usually limit my griping about pharmaceutical overpricing to generics and avoid brand-names is that while high generic pricing is inexcusable, high brand-name pricing is debatably useful. Some people would say that the benefit of encouraging more drug development is worth the cost of higher prices, other people would say that it isn’t. I didn’t want to wade into this complicated debate.

But I guess now I have to. So. Lit review time. I searched the economics literature for studies, models, and arguments used to calculate whether price regulation would decrease drug development, and if so, how the benefits and risks balanced out. Here’s what I’ve got:

1. Golec & Vernon (2006) say that as a result of European drug price regulation, “EU consumers enjoyed much lower pharmaceutical price inflation, however, at a cost of 46 fewer new medicines introduced by EU firms.”

2. Eger and Mahlich (2014) find that among pharmaceutical companies, “a higher presence in Europe is associated with lower R&D investments. The results can be interpreted as further evidence of the deteriorating effect of regulation on firm’s incentives to invest in R&D.”

3. Kutyavina (2010) finds that “brand-name pharmaceutical firms characterized by large R&D expenditures decreased their R&D efforts post 1993 threat [to regulate drug prices] relative to firms that did not engage in as much innovative R&D”.

4. Acemoglu and Linn (2004) find that “We find a large effect of potential market size on the entry of nongeneric drugs and new molecular entities”, which I *think* is supposed to generalize to mean that the more money they expect to make the more research they do. I will count this as half a study since the connection is not explicit.

5. Danzon & Epstein (2008) analyze price regulations and new drugs invented in 15 countries and 12 drug classes, and find that “If price regulation reduces drug prices, it contributes to launch delay in the home country.

6. Troyer & Krasnikov (2002) find that “the empirical relationship between pharmaceutical industry revenues and pharmaceutical industry innovation is estimated, allowing for an exploration of the impact of the Medicaid rebate program [which regulated drug prices somewhat]. Using the empirical results, the opportunity cost of the Medicaid rebate program is found to be as high as four new drug approvals annually. Given the increased interest in a Medicare drug benefit, regulators should be aware of the hidden cost of price regulation for pharmaceuticals.”

7. Vernon (2005) finds that “I simulate how a new policy regulating pharmaceutical prices in the US will affect R&D investment. I find that such a policy will lead to a decline in industry R&D by between 23.4% and 32.7%. This prediction, however, is accompanied by several caveats.”

8. Golec, Hegde, and Vernon (2009) find that “Results show that the HSA [a bill to regulate drug spending in the US] had significant negative effects on stock prices and firm-level R&D spending. Conservatively, the HSA reduced R&D spending by about $1 billion even though it never became law.”

9. Santerre and Vernon (2006) use drug demand data to simulate various regulatory regimes, and find that a certain price regulation policy they test, continued over twenty years, would have cost gains of $472 billion (!) but also “have led to 198 new drugs being brought to the US market” (!!). They note that “Therefore, the average social opportunity cost per drug developed during this period was approximately $2.4 billion. Research on the value of pharmaceuticals suggests that the social benefits of a new drug are far greater than this estimate. Hence, drug price controls could do more harm than good.”

10. Keyhani, Carpenter, et al (2010) find that “The United States accounted for 42% of prescription drug spending and 40% of the total GDP among innovator countries and was responsible for the development of 43.7% of the NMEs [ie new drugs invented]. The United Kingdom, Switzerland, and a few other countries innovated proportionally more than their contribution to GDP or prescription drug spending, whereas Japan, South Korea, and a few other countries innovated less…higher prescription drug spending in the US does not disproportionately privilege domestic innovation, and many countries with drug price regulation were significant contributors to pharmaceutical innovation.” This study does not attempt to address the effects of price regulation, only to say that European countries seem to do pretty well at innovation despite price regulation, which is suggestive that price regulation does not hurt drug innovation but not really scientific evidence for it. I’m going to count this one as half a study too.

So by my count, there are eight-and-a-half studies concluding that price regulation would hurt new drug innovation, and one-half of a study concluding that it wouldn’t. I’ve tried to eliminate all the studies sponsored by the pharmaceutical industry from this list, but I might have missed some, and I am always skeptical of anything that says anything the pharmaceutical industry approves of even I can’t trace the money directly.

One source I do trust is RAND, a think tank which is generally well-respected and pretty objective (despite the name, they are not associated with Ayn Rand or Rand Paul). In Regulating Drug Prices: US Policy Alternatives In A Global Context, they write:

> U.S. consumers spend roughly twice as much on drugs as their European counterparts….Pressure is building in U.S. policy circles for the federal government to take action to regulate the cost of drugs. At the same time, there is debate about the pros and cons of doing so…To shed light on this debate, a team of RAND researchers examined the impact of drug price regulation…The results showed that:
>
> — Globally, the regulation of pharmaceutical prices has increased in recent years.
>
> — In most cases, regulation reduces pharmaceutical revenues.
>
> — Regulatory approaches that reduce pharmaceutical revenues may generate modest consumer savings in the best cases, but risk much larger costs as decreased innovation leads to reductions in life expectancy.

In other words, such prices would be good in the short term as we get all the currently-existing drugs for very cheap:

> Annual per capita spending in 2010 would fall for Americans age 55–59 by an amount in the range of $9,000 annually and for Europeans of the same age by an amount in the range of $400.

But bad in the long-term as pharmaceutical innovation declines and we have fewer interventions available to protect our health:

> Life expectancy would fall by somewhere in the range of 0.2 years for Americans age 55–59 in 2010 and by approximately 0.1 years for Europeans in the same cohort and year. By 2060, this effect would increase for both Americans and Europeans to approximately 0.7 years.

Given the value they place on human life, they argue that this money-for-life-years trade is net negative:

> Overall, as shown in Figure 1, the net value of price controls is positive in the short term (2010) for Americans age 55–59, producing approximately $1,100 in per capita savings, but negative for Europeans in the same cohort and year, who face increased costs in the range of $8,000. For both Americans and Europeans, price controls have higher per capita costs over the longer term: By 2060, reductions in life expectancy, after accounting for medical cost savings, would cost the equivalent of $51,000 and $54,000, to age 55–59 Americans and Europeans, respectively.

All of this sounds sort of boring and economics-y when you read it like this, and maybe your eyes are glazing over. So let me put this in context. In 2060 there will probably be 420 million Americans and 523 million Europeans. And suppose that whatever changes we make in drug regulations today last for one human lifespan, so that everybody has a chance to be 55-60. So about a billion people each losing about 0.7 years of their life equals 700 million life-years. Since some people live in countries outside the US and Europe [citation needed] and they also benefit from First-World-invented medications, let’s round this up to about a billion life-years lost.

What was the worst thing that ever happened? One strong contender is Mao’s Great Leap Forward, in which ineffective agricultural reforms and very effective purges killed 45 million people. Most of these people were probably already adults, and lifespan in Mao’s China wasn’t too high, so let’s say that each death from the Great Leap Forward cost what would otherwise be twenty healthy life years. In that case, the worst thing that has ever happened until now cost 45 million * 20 = 900 million life-years.

Once again, RAND’s calculations plus my own Fermi estimate suggest that prescription drug price regulation would cost one billion life-years, which would very slightly edge out Communist China for the title of Worst Thing Ever.

Am I exaggerating or being facetious? I’m actually not sure. ~~Dammit, Jim~~ I’m a doctor, not an economist. I’m not qualified to analyze any of those ten studies above beyond a quick check to see if they’re completely ridiculous. I’m not qualified to say if RAND is right or wrong to estimate a cost of 0.7 life-years, or whether I’m misusing their calculation to try to add up exactly how bad it would be. Maybe a real economist will look at this whole essay and say it’s really stupid. I don’t know.

The only thing I can say in my own defense is that *I am acknowledging that the question exists*. I am not at all sure that Vox has reached this level yet. They just wrote an article on price regulations for brand-name drugs which, first, mixed them liberally with generic drugs despite the different arguments around each, and second, didn’t mention anything about research or innovation. Call me overly demanding, but when you are proposing a policy which most economists think would decrease the rate of life-saving medical progress, and which by some calculations might edge out Mao’s China for the title of most disastrous and deadly thing in all of human history, then I feel like you should acknowledge, at least in a single sentence, that somebody has claimed, at least once, that the policy might have some slight downside. At least don’t act as if it’s the same issue as a different kind of drug regulation which doesn’t have that downside.

I think it’s an unfortunate omission to talk about the EpiPen cost increase as relating only to lack of price controls, and fail to mention the reason why this happens with EpiPen but never chairs or mugs. And I think it’s a further omission to talk about regulating brand-name medications but fail to mention that some people think it will backfire and impede innovation. While I appreciate the effort to say we’re both on the same team of reducing drug costs, I’m a little concerned about my teammates’ strategy here.

And there’s another way we’re *not* quite on the same team. I’m on Team Left-Libertarian, which luckily is so confusing and contradictory that I can define it however I want. And today it means that while I’m not opposed to all regulation in principle, I at least get really scared when somebody pushes for regulation today and promises to check whether it will have bad consequences tomorrow. I think that’s how we got in this mess where the generics industry is so regulated that EpiPens cost hundreds of dollars, and even if Vox and I are on the same object-level team of Make Epi-Pens Cost Less, I worry we are not on the same meta-level team of Learn From The Fact That Epi-Pens Cost So Much And Worry That The Same Kind Of Thinking That Caused The Epi-Pen Problem Will Probably Cause Other Problems Too.

II.

So do *I* have a solution to the high price of brand-name drugs? Well, I have a partial, unsatisfying solution. But first, a digression.

Vox gives the example of Harvoni costing $32,000 in the United States, but only $22,000 in the United Kingdom. This is supposed to be an example of the United Kingdom's drug price regulation system working. I guess this is good, but you may notice that both numbers are really really high. There's generic Harvoni available in India for $900. I can't find how much it costs to manufacture, but reading between the lines and looking at some similar compounds, it's probably about $100. So good work, Britain. You're paying $22,000 instead of $32,000 for a $100 pill.

There are a couple of morals to this story. The first is that Vox's claim that generics made by two competing companies cost 55% of the brand-name price isn't the right statistic to use here. Look at their source and you find that as number of competing companies gets to 20, generics cost 5% of brand names. As number of competitors approaches infinity, drug cost should approach manufacturing cost, which can be very low – in the cast of Harvoni, less than one percent. This seems true in the case of modafinil, which I've talked about before; it costs about $25 per pill in the US and more like $2 per pill in more generic-friendly India.

So the second moral of the story is that almost all gains in prescription drug prices are to be found not in price regulation bringing prices down from $32,000 to $22,000, but in switching from monopoly brand-name drugs that cost $32,000 to heavily-competitive generic drugs that cost $100.

In a lot of cases, this is easier than you would think.

Pristiq is the brand-name of desvenlafaxine, a new antidepressant which is still brand-name only. Desvenlafaxine sounds a lot like venlafaxine – which is Effexor, an old antidepressant which is available in generic. In fact, desvenlafaxine is a tiny change to the venlafaxine molecule which may or may not have any interesting medical benefit over the original, and which was invented solely to have something whose patent hasn't expired.

Wyeth, the company that makes Pristiq, says that it's better than Effexor because it doesn't have as many drug-drug interactions. But Effexor doesn't really have clinically significant drug-drug interactions, and this seems to be them just saying random stuff and hoping people believe them. There are no good head-to-head studies comparing Pristiq to Effexor, but if you try to piece together a comparison from unrelated studies (not recommended, but we'll do it anyway) Effexor actually seems *better* than its newer cousin. Even the data I took from drug rating databases shows patients preferring Effexor to Pristiq by quite a lot. Carlat Psychiatry, which is psychiatrists' insider news site on pharmacology developments, has a blog post called Top Five Reasons To Forget About Pristiq. Most of the well-informed psychiatrists I know agree that Pristiq is a slightly worse version of an older antidepressant with no proven advantages.

A month's supply of Effexor costs $20. A month's supply of Pristiq costs $300. So let me amend the paragraph above. Pristiq is a slightly worse version of an older antidepressant with no proven advantages that also costs fifteen times as much.

It should come as no surprise to anyone familiar with the state of psychiatry that it is the second most-prescribed antidepressant in the USA, with three million prescriptions per month.

Why would this happen? The relevant study is called Pharmaceutical Industry-Sponsored Meals And Physician Prescribing Patterns For Medicare Beneficiaries, so you know it’s going to be good. It shows that doctors who often eat drug-company-sponsored free lunches are more than twice as likely to prescribe Pristiq as doctors who rarely eat such lunches. This matches my observations perfectly. Doctors prescribe Pristiq because they don’t know very much about antidepressants, but they attend free lunches by pharmaceutical companies who tell them that Pristiq is great, and they believe it. If this surprises you, be more cynical.

I’m looking at the price of Pristiq in Canada, and it seems to range around $120 to $250. So if we instituted price regulations like Canada’s, we might lower the cost of Pristiq from $300 to $150. If we convinced doctors to prescribe Effexor instead, it would be $20, plus I really do believe Effexor is genuinely better.

Pristiq is far from alone in this. I don’t have good statistics, but I bet that at least half of brand-name prescriptions in the US are more like Pristiq (attempts to rip people off) than like Harvoni (genuinely wonderful breakthroughs in medical science).

So one of the best ways to deal with expensive brand-name drugs is to stop using expensive brand name drugs for no reason. Since I get to define what left-libertarianism means however I want, I will say that it is provisionally okay with banning pharmaceutical companies from buying doctors lunch, as long as there aren’t any studies concluding that this would kill more people than Communist China. There are probably lots of other ways to improve medical education and medical economics so that doctors are less easily bamboozled into prescribing these, but those can wait for other blog posts.

What about the genuinely novel brand-name drugs, the ones like Harvoni that really *are* better than anything that came before?

An optimistic answer: maybe after we stop spending our civilizational resources on Pristiq, we'll have a little more money to afford them, and maybe we'll be happy to subsdize the genuinely awesome pharmaceutical research that remains.

Another optimistic answer: once FDA regulatory requirements are loosened, there will be a wide selection of different brand-name drugs. For example, even when Prozac was a brand-name, pharma's ability to inflate its price was limited by the existence of several very similar brand-name drugs, like Paxil and Zoloft. If there are twenty competing brand-name hepatitis wonder drugs – and I don't think that's outside the realm of what we can hope for – then I think that will tend to lower prices to cost. This would include the cost of research and licensing, and so still be pretty high, but as long as research is a real unavoidable cost it would probably be the best we can hope for.

(this would be a good time to bring up that chlorcyclizine costs fifty cents per pill and might work as well as Harvoni)

The pessimistic answer is that all we can do is ensure that the generics marketplace is fair and competitive. And then rich people can buy Harvoni now for $30,000, and poor people will have to wait ten years to buy Harvoni when it costs $100. Right now they'll unfortunately have to figure out how to make do with the set of medications invented in 2006 and before.

And I know this is terrible, especially if someone has a disease with only one cure and it was invented after 2006. But think of it this way. This objection, rephrased, is that 2016 has more drugs available than 2006, and we want to maximize the number of new drugs available to the poorest patients. But if we try to do that by instituting price controls which decreases the rate of drug innovation long-term, then we end up decreasing the number of new drugs available to the poorest patients, exactly the opposite of what we thought we were doing.

Let me give an example. According to study (9) above, price controls would have caused about 200 fewer drugs to be approved over the period from 1980 – 2000. In fact about 600 drugs were approved during that period. So if they're right, it would have cut the innovation rate by 1/3. That means that in Hypothetical Price Control World's 2016, after 36 years of price controls, we would only have 24 of our years' worth of drugs – ie, the drugs that we had in 2004 in our own world. But since drugs usually go off patent about ten years after approval, in fact we've genericized the drugs that we had in 2006 in our own world. So we have more drugs available *just as generics* than Hypothetical Price Control World has as generics and brand-names combined. If poor people can afford only generics or price-controlled brand names, our poor people are better off than Hypothetical Price Control World's poor people (and our rich people are *much* better off than Hypothetical Price-Control World's rich people). And as time goes on, our advantage over their world will only get bigger.

Maybe there are better ways to do this. Some people have talked about funding research via "prizes" rather than through an investment-and-profit model. Some people say we should fund it publicly through the NIH or something, which we already sort of do to a degree. Still other people say that we should abolish the FDA, cut the costs of drug development by an order of magnitude, and, um, see what happens. I don’t know about any of those things. I just feel like until you’re ready to set these up and have some idea that they work, do the thing that probably is going to result in people having the best access to the most life-saving drugs. Which right now looks like no price control.

Or maybe I am completely wrong about all of this. I am not an economist and have to take these studies at face value, and anything that touches pharmaceutical companies ends up being corrupted and full of lies. Maybe I myself am saying something very stupid that will end up killing more people than Communist China. If so I certainly hope that people who know more than I do will tell me why these calculations were wrong and how to look at this situation better.

But this is the level at which I think this discussion needs to be had.

It’s Bayes All The Way Up

[Epistemic status: Very speculative. I am not a neuroscientist and apologize for any misinterpretation of the papers involved. Thanks to the people who posted these papers in r/slatestarcodex. See also Mysticism and Pattern-Matching and Bayes For Schizophrenics]

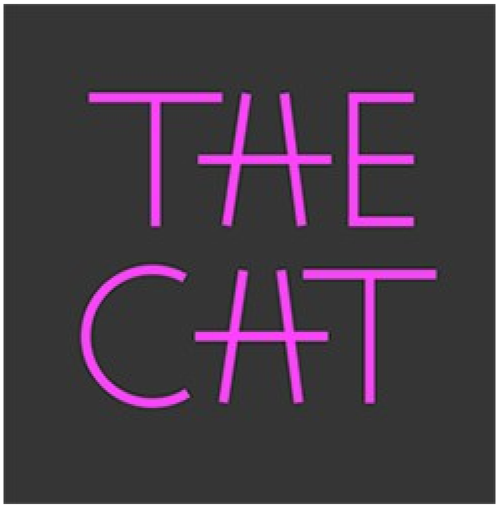
Bayes’ Theorem is an equation for calculating certain kinds of conditional probabilities. For something so obscure, it’s attracted a surprisingly wide fanbase, including doctors, environmental scientists, economists, bodybuilders, fen-dwellers, and international smugglers. Eventually the hype reached the point where there was both a Bayesian cabaret *and* a Bayesian choir, popular books using Bayes’ Theorem to prove both the existence and the nonexistence of God, and even Bayesian dating advice. Eventually everyone agreed to dial down their exuberance a little, and accept that Bayes’ Theorem might not literally explain *absolutely* everything.

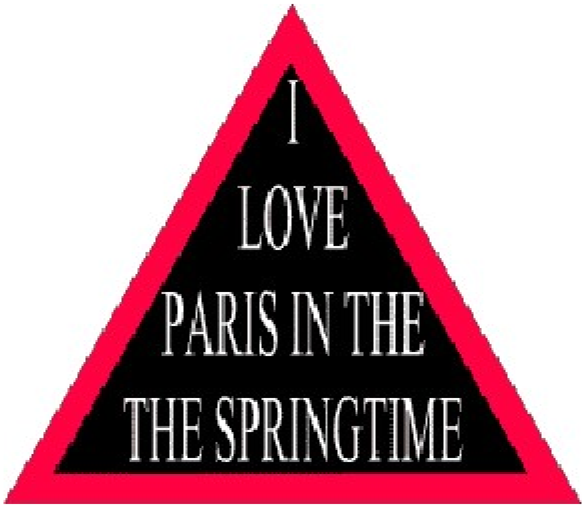
So – did you know that the neurotransmitters in the brain might represent different terms in Bayes’ Theorem?

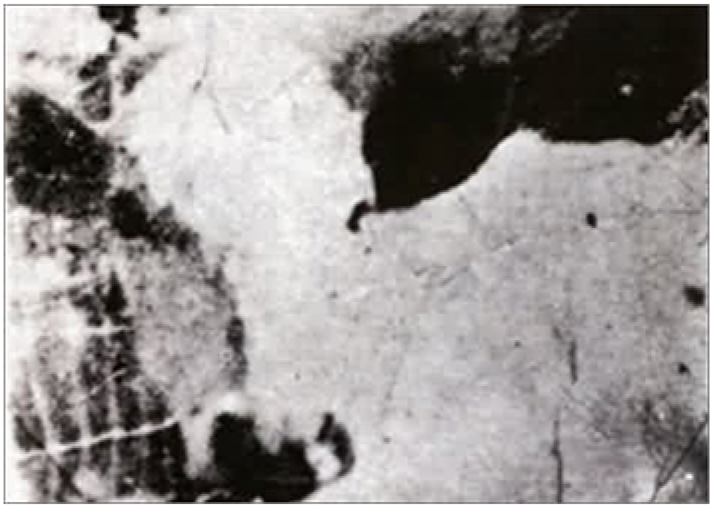
First things first: Bayes’ Theorem is a mathematical framework for integrating new evidence with prior beliefs. For example, suppose you’re sitting in your quiet suburban home and you hear something that sounds like a lion roaring. You have some prior beliefs that lions are unlikely to be near your house, so you figure that it’s probably not a lion. Probably it’s some weird machine of your neighbor’s that just happens to sound like a lion, or some kids pranking you by playing lion noises, or something. You end up believing that there’s probably no lion nearby, but you do have a slightly higher probability of there being a lion nearby than you had before you heard the roaring noise.

Bayes' Theorem is just this kind of reasoning converted to math. You can find the long version here.

This is what the brain does too: integrate new evidence with prior beliefs. Here are some examples I've used on this blog before:




I
LOVE
PARIS IN THE
THE SPRINGTIME



All three of these are examples of top-down processing. Bottom-up processing is when you build perceptions into a model of the the world. Top-down processing is when you let your models of the world influence your perceptions. In the first image, you view the center letter of the the first word as an H and the second as an A, even though they're the the same character; your model of the world tells you that THE CAT is more likely than TAE CHT. In the second image, you read “PARIS IN THE SPRINGTIME”, skimming over the duplication of the word “the”; your model of the world tells you that the phrase should probably only have one “the” in it (just as you've probably skimmed over it the three times I've duplicated “the” in this paragraph alone!). The third image might look meaningless until you realize it's a cow's head; once you see the cow's head your model of the world informs your perception and it's almost impossible to see it as anything else.

(Teh fcat taht you can siltl raed wrods wtih all the itroneir ltretrs rgraneanrd is ahonter empxlae of top-dwon pssirocneg mkinag nsioy btotom-up dtaa sanp itno pacle)

But top-down processing is much more omnipresent than even these examples would suggest. Even something as simple as looking out the window and seeing a tree requires top-down processing; it may be too dark or foggy to see the tree one hundred percent clearly, the exact pattern of light and darkness on the tree might be something you've never seen before – but because you know what trees are and expect them to be around, the image "snaps" into the schema "tree" and you see a tree there. As usual, this process is most obvious when it goes wrong; for example, when random patterns on a wall or ceiling "snap" into the image of a face, or when the whistling of the wind "snaps" into a voice calling your name.

> Most of the things you perceive when awake are generated from very limited input – by the same machinery that generates dreams with no input
>
> — Void Of Space (@VoidOfSpace) September 2, 2016

Corlett, Frith & Fletcher (2009) (henceforth CFF) expand on this idea and speculate on the biochemical substrates of each part of the process. They view perception as a "handshake" between top-down and bottom-up processing. Top-down models predict what we're going to see, bottom-up models perceive the real world, then they meet in the middle and compare notes to calculate a prediction error. When the prediction error is low enough, it gets smoothed over into a consensus view of reality. When the prediction error is too high, it registers as salience/surprise, and we focus our attention on the stimulus involved to try to reconcile the models. If it turns out that bottom-up was right and top-down was wrong, then we adjust our priors (ie the models used by the top-down systems) and so learning occurs.

In their model, bottom-up sensory processing involves glutamate via the AMPA receptor, and top-down sensory processing involves glutamate via the NMDA receptor.

Dopamine codes for prediction error, and seem to represent the level of certainty or the “confidence interval” of a given prediction or perception. Serotonin, acetylcholine, and the others seem to modulate these systems, where “modulate” is a generic neuroscientist weasel word. They provide a lot of neurological and radiologic evidence for these correspondences, for which I highly recommend reading the paper but which I’m not going to get into here. What I found interesting was their attempts to match this system to known pharmacological and psychological processes.

CFF discuss a couple of possible disruptions of their system. Consider *increased* AMPA signaling combined with *decreased* NMDA signaling. Bottom-up processing would become more powerful, unrestrained by top-down models. The world would seem to become “noisier”, as sensory inputs took on a life of their own and failed to snap into existing categories. In extreme cases, the “handshake” between exuberant bottom-up processes and overly timid top-down processes would fail completely, which would take the form of the sudden assignment of salience to a random stimulus.

Schizophrenics are famous for “delusions of reference”, where they think a random object or phrase is deeply important for reasons they have trouble explaining. Wikipedia gives as examples:

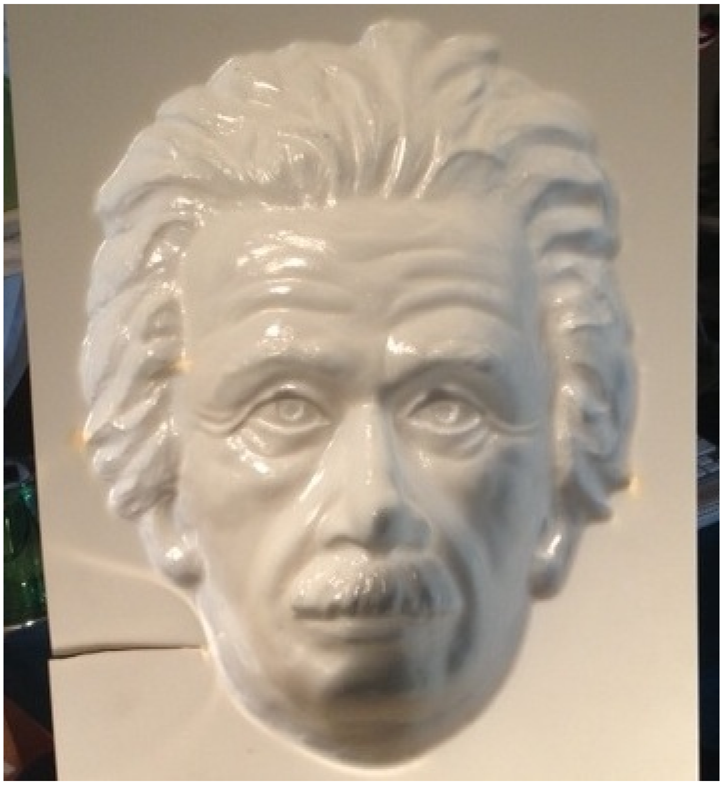
> – A feeling that people on television or radio are talking about or talking directly to them
>
> – Believing that headlines or stories in newspapers are written especially for them
>
> – Seeing objects or events as being set up deliberately to convey a special or particular meaning to themselves

> – Thinking ‘that the slightest careless movement on the part of another person had great personal meaning… increased significance’

In CFF, these are perceptual handshake failures; even though “there’s a story about the economy in today’s newspaper” should be perfectly predictable, noisy AMPA signaling registers it as an extreme prediction failure, and it fails its perceptual handshake with overly-weak priors. Then it gets flagged as shocking and deeply important. If you’re unlucky enough to have your brain flag a random newspaper article as shocking and deeply important, maybe phenomenologically that feels like it’s a secret message for you.

And this pattern – increased AMPA signaling combined with decreased NMDA signaling – is pretty much the effect profile of the drug ketamine, and ketamine does cause a paranoid psychosis mixed with delusions of reference.

Organic psychosis like schizophrenia might involve a similar process. There’s a test called the binocular depth inversion illusion, which looks like this:



(source)

The mask in the picture is concave, ie the nose is furthest away from the camera. But most viewers interpret it as convex, with the nose closest to the camera. This makes sense in terms of Bayesian perception; we see right-side-in faces a whole lot more often than inside-out faces.

Schizophrenics (and people stoned on marijuana!) are more likely to properly identify the face as concave than everyone else. In CFF’s system, something about schizophrenia and marijuana messes with NMDA, impairs priors, and reduces the power of top-down processing. This predicts that schizophrenics and potheads would both have paranoia and delusions of reference, which seems about right.

Consider a slightly different distortion: *increased* AMPA signaling combined with *increased* NMDA signaling. You’ve still got a lot of sensory noise. But you’ve also got stronger priors to try to make sense of them. CFF argue these are the perfect conditions to create hallucinations. The increase in sensory noise means there’s a lot of data to be explained; the increased top-down pattern-matching means that the brain is very keen to fit all of it into some grand narrative. The result is vivid, convincing hallucinations of things that are totally not there at all.

LSD is mostly serotonergic, but most things that happen in the brain bottom out in glutamate eventually, and LSD bottoms out in exactly the pattern of increased AMPA and increased NMDA that we would expect to produce hallucinations. CFF don’t mention this, but I would also like to add my theory of pattern-matching based mysticism. Make the top-down prior-using NMDA system strong enough, and the entire world collapses into a single narrative, a divine grand plan in which everything makes sense and you understand all of it. This is also something I associate with LSD.

If dopamine represents a confidence interval, then increased dopaminergic signaling should mean narrowed confidence intervals and increased certainty. Perceptually, this would correspond to increased sensory acuity. More abstractly, it might increase “self-confidence” as usually described. Amphetamines, which act as dopamine agonists, do both. Amphetamine users report increased visual acuity (weirdly, they also report blurred vision sometimes; I don’t understand exactly what’s going on here). They also create an elevated mood and grandiose delusions, making users more sure of themselves and making them feel like they can do anything.

(something I remain confused about: elevated mood and grandiose delusions are also typical of bipolar mania. People on amphetamines and other dopamine agonists act pretty much exactly like manic people. Antidopaminergic drugs like olanzapine are very effective acute antimanics. But people don’t generally think of mania as primarily dopaminergic. Why not?)

CFF end their paper with a discussion of sensory deprivation. If perception is a handshake between bottom-up sense-data and top-down priors, what happens when we turn the sense-data off entirely? Psychologists note that most people go a little crazy when placed in total sensory deprivation, but that schizophrenics actually seem to do *better* under sense-deprivation conditions. Why?

The brain filters sense-data to adjust for ambient conditions. For example, when it’s very dark, your eyes gradually adjust until you can see by whatever light is present. When it’s perfectly silent, you can hear the proverbial pin drop. In a state of total sensory deprivation, any attempt to adjust to a threshold where you can detect the nonexistent signal is actually just going to bring you down below the point where you’re picking up noise. As with LSD, when there’s too much noise the top-down systems do their best to impose structure on it, leading to hallucinations; when they fail, you get delusions. If schizophrenics have inherently noisy perceptual systems, such that all perception comes with noise the same way a bad microphone gives off bursts of static whenever anyone tries to speak into it, then their brains will actually become *less* noisy as sense-data disappears.

(this might be a good time to remember that no congentally blind people ever develop schizophrenia and no one knows why)

II.

Lawson, Rees, and Friston (2014) offer a Bayesian link to autism.

(there are probably a lot of links between Bayesians and autism, but this is the only one that needs a journal article)

They argue that autism is a form of *aberrant precision*. That is, confidence intervals are too low; bottom-up sense-data cannot handshake with top-down models unless they're almost-exactly the same. Since they rarely are, top-down models lose their ability to "smooth over" bottom-up information. The world is full of random noise that fails to cohere into any more general plan.

Right now I'm sitting in a room writing on a computer. A white noise machine produces white noise. A fluorescent lamp flickers overhead. My body is doing all sorts of body stuff like digesting food and pumping blood. There are a few things I need to concentrate on: this essay I'm writing, my pager if it goes off, any sorts of sudden dramatic pains in my body that might indicate a life-threatening illness. But I don't need to worry about the feeling of my back against the back fo the chair, or the occasional flickers of the fluorescent light, or the feeling of my shirt on my skin.

A well-functioning perceptual system gates out those things I don't need to worry about. Since my shirt always feels more or less similar on my skin, my top-down model learns to predict that feeling. When the top-down model predicts the shirt on my skin, and my bottom-up sensation reports the shirt on my skin, they handshake and agree that all is well. Even if a slight change in posture makes a different part of my shirt brush against my skin than usual, the confidence intervals are wide: it is still an instance of the class "shirt on skin", it "snaps" into my shirt-on-skin schema, and the perceptual handshake goes off successfully, and all remains well. If something dramatic happens – for example my pager starts beeping really loudly – then my top-down model, which has thus far predicted silence – is rudely surprised by the sudden burst of noise. The perceptual handshake fails, and I am startled, upset, and instantly stop writing my essay as I try to figure out what to do next (hopefully answer my pager). The system works.

The autistic version works differently. The top-down model tries to predict the feeling of the shirt on my skin, but tiny changes in the position of the shirt change the feeling somewhat; bottom-up data does not *quite* match top-down prediction. In a neurotypical with wide confidence intervals, the brain would shrug off such a tiny difference, declare it good enough for government work, and (correctly) ignore it. In an autistic person, the confidence intervals are very narrow; the top-down systems expect the feeling of shirt-on-skin, but the bottom-up systems report a *slightly different* feeling of shirt-on-skin. These fail to snap together, the perceptual handshake fails, and the brain flags it as important; the autistic person is startled, upset, and feels like stopping what they're doing in order to attend to it.

(in fact, I think the paper might be claiming that "attention" just means a localized narrowing of confidence intervals in a certain direction; for example, if I pay attention to the feeling of my shirt on my skin, then I *can* feel every little fold and micromovement. This seems like an important point with a lot of implications.)

Such handshake failures match some of the sensory symptoms of autism pretty well. Autistic people dislike environments that are (literally or metaphorically) noisy. Small sensory imperfections bother them. They literally get annoyed by scratchy clothing. They tend to seek routine, make sure everything is maximally predictable, and act as if even tiny deviations from normal are worthy of alarm.

They also stim. LRF interpret stimming as an attempt to control sensory predictive environment. If you’re moving your arms in a rhythmic motion, the overwhelming majority of sensory input from your arm is from that rhythmic motion; tiny deviations get lost in the larger signal, the same way a firefly would disappear when seen against the blaze of a searchlight. The rhythmic signal which you yourself are creating and keeping maximally rhythmic is the most predictable thing possible. Even something like head-banging serves to create extremely strong sensory data – sensory data whose production the head-banger is themselves in complete control of. If the brain is in some sense minimizing predictive error, and there’s no reasonable way to minimize prediction error because your predictive system is messed up and registering *everything* as a dangerous error – then sometimes you have to take things into your own hands, bang your head against a metal wall, and say “I totally predicted all that pain”.

(the paper doesn’t mention this, but it wouldn’t surprise me if weighted blankets work the same way. A bunch of weights placed on top of you will predictably stay there; if they’re heavy enough this is one of the strongest sensory signals you’re receiving and it might “raise your average” in terms of having low predictive error)

What about all the non-sensory-gating-related symptoms of autism? LRF think that autistic people dislike social interaction because it’s “the greatest uncertainty”; other people are the hardest-to-predict things we encounter. Neurotypical people are able to smooth social interaction into general categories: this person seems friendly, that person probably doesn’t like me. Autistic people get the same bottom-up data: an eye-twitch here, a weird half-smile there – but it never snaps into recognizable models; it just stays weird uninterpretable clues. So:

> This provides a simple explanation for the pronounced social-communication difficulties in autism; given that other agents are arguably the most difficult things to predict. In the complex world of social interactions, the many-to-one mappings between causes and sensory input are dramatically increased and difficult to learn; especially if one cannot contextualize the prediction errors that drive that learning.

They don't really address differences between autists and neurotypicals in terms of personality or skills. But a lot of people have come up with stories about how autistic people are better at tasks that require a lot of precision and less good at tasks that require central coherence, which seems like sort of what this theory would predict.

LRF ends by discussing biochemical bases. They agree with CFF that top-down processing is probably related to NMDA receptors, and so suspect this is damaged in autism. Transgenic mice who lack an important NMDA receptor component seem to behave kind of like autistic humans, which they take as support for their model – although obviously a lot more research is needed. They agree that acetylcholine "modulates" all of this and suggest it might be a promising pathway for future research. They agree with CFF that dopamine may represent precision/confidence, but despite their whole spiel being that precision/confidence is messed up in autism, they don't have much to say about dopamine except that it probably modulates something, just like everything else.

III.

All of this is fascinating and elegant. But is it elegant *enough*?

I notice that I am confused about the relative role of NMDA and AMPA in producing hallucinations and delusions. CFF say that enhanced NMDA signaling results in hallucinations as the brain tries to add excess order to experience and “overfits” the visual data. Fine. So maybe you get a tiny bit of visual noise and think you’re seeing the Devil. But shouldn’t NMDA and top-down processing also be the system that tells you there is a high prior against the Devil being in any particular visual region?

Also, once psychotics develop a delusion, that delusion usually sticks around. It might be that a stray word in a newspaper makes someone think that the FBI is after them, but once they think the FBI is after them, they fit everything into this new paradigm – for example, they might think their psychiatrist is an FBI agent sent to poison them. This sounds a lot like a new, very strong prior! Their doctor presumably isn’t doing much that seems FBI-agent-ish, but because they’re working off a narrative of the FBI coming to get them, they fit everything, including their doctor, into that story. But if psychosis is a case of attenuated priors, why should that be?

(maybe they would answer that because psychotic people also have increased dopamine, they believe in the FBI with absolute certainty? But then how come most psychotics don’t seem to be manic – that is, why aren’t they overconfident in anything except their delusions?)

LRF discuss prediction error in terms of mild surprise and annoyance; you didn’t expect a beeping noise, the beeping noise happened, so you become startled. CFF discuss prediction error as sudden surprising salience, but then say that the attribution of salience to an odd stimulus creates a delusion of reference, a belief that it's somehow pregnant with secret messages. These are two very different views of prediction error; an autist wearing uncomfortable clothes might be constantly focusing on their itchiness rather than on whatever she's trying to do at the time, but she's not going to start thinking they're a sign from God. What's the difference?

Finally, although they highlighted a selection of drugs that make sense within their model, others seem not to. For example, there's some discussion of ampakines for schizophrenia. But this is the opposite of what you'd want if psychosis involved overactive AMPA signaling! I'm not saying that the ampakines for schizophrenia definitely work, but they don't seem to make the schizophrenia noticeably worse either.

Probably this will end the same way most things in psychiatry end – hopelessly bogged down in complexity. Probably AMPA does one thing in one part of the brain, the opposite in other parts of the brain, and it's all nonlinear and different amounts of AMPA will have totally different effects and maybe downregulate itself somewhere else.

Still, it's neat to have at least a vague high-level overview of what *might* be going on.

Some Context For That NYT Sugar Article

Imagine a political historian discovers that Lyndon Johnson accepted a campaign contribution from a big Wall Street bank. Since Johnson's policies helped shape the modern Democratic Party, everyone agrees the Democrats are built on a foundation of lies. "Republicans Vindicated; Small-Government Conservativism Was Right All Along", say the headlines of all the major newspapers.

This is kind of how I feel about the reaction to the latest *New York Times* article.

How The Sugar Industry Shifted Blame To Fat describes new historical research that finds that the sugar industry sponsored a study showing that fat (and not sugar) was the major risk factor for cardiovascular disease. They tie this into a bigger narrative about how sugar is the real dietary villain, and it's only the sugar industry's successful bribery work that made us suspect fat for so long:

> The revelations are important because the debate about the relative harms of sugar and saturated fat continues today, Dr. Glantz said. For many decades, health officials encouraged Americans to reduce their fat intake, which led many people to consume low-fat, high-sugar foods that some experts now blame for fueling the obesity crisis.
>
> "It was a very smart thing the sugar industry did, because review papers, especially if you get them published in a very prominent journal, tend to shape the overall scientific discussion," he said […]

I’m glad researchers have discovered this. But treating it as a smoking gun which exonerates fat and blames sugar is like the political example above. Yes, it’s sketchy for LBJ to take Wall Street money. But this kind of low-level corruption is so universal that concentrating on any one example is likely to lead to overcorrection.

Yes, the sugar lobby sponsors some research, but the fat lobby has researchers of its own. They tend to be associated with the dairy and meat industries, both of which are high in saturated fat and both of which are very involved in nutrition research. For example, Siri-Tarino et al’s Meta-analysis of prospective cohort studies evaluating the association of saturated fat with cardiovascular disease finds that saturated fat does not increase heart disease risk, but it has a little footnote saying that it’s supported by the National Dairy Council. Modulation of Replacement Nutrients, which finds that replacing dietary fat with dietary sugar doesn’t help and may worsen heart disease, includes two authors affiliated with the National Dairy Council and one affiliated with the National Cattleman’s Beef Association.

Mother Jones does a dairy industry expose and finds:

> [Industry] ties can sometimes be hard to avoid, since much of the research on dairy is funded by a constellation of industry-backed institutes, including the Nestlé Nutrition Institute, the Dannon Institute, and the Dairy Research Institute, which spends $19 million a year “to establish the health benefits of dairy products and ingredients.” Even Willett acknowledges that he has received a “very small” dairy industry grant. Dairy companies also donate heavily to the American Society for Nutrition, which publishes the influential American Journal of Clinical Nutrition, and the Academy of

> Nutrition and Dietetics, “the world’s largest organization of food and nutrition professionals.”
>
> Then there are the industry’s donations to politicians. Dairy companies spent nearly $63 million on federal lobbying and gave $24 million to candidates between 2004 and 2014.

As Jim Babcock points out in the comments, some of the agendas are more complicated than I’m making them sound. Dairy was pretty okay with the low-fat craze for a while, because it let them market low-fat milk. But they do seem to be behind a lot of the pro-saturated-fat research going on right now, and their website does promote pro-saturated-fat articles (1, 2, 3). Overall they seem to be taking a low-key approach where they roll with some studies and push back on others.

In any case, claims that the sugar industry sponsored one study back in the 1960s, and this means everything we’ve ever thought is wrong and biased against fat and in favor of sugar, miss the point (especially since there are probably problems with both sugar *and* fat). Whatever study the *New York Times* has dredged up was one volley in an eternal clandestine war of Big Fat against Big Sugar, and figuring out who’s distorted the science more is the sort of project that’s going to take more than one article.

AI Persuasion Experiment

I’ve been trying to write a persuasive essay about AI risk, but there are already a lot of those out there and I realize I should see if any of them are better before pushing mine. This also ties into a general interest in knowing to what degree persuasive essays really work and whether we can measure that.

So if you have time, I’d appreciate it if you did an experiment. You’ll be asked to read somebody’s essay explaining AI risk and answer some questions about it. Note that some of these essays might be long, but you don’t have to read the whole thing (whether it can hold your attention so that you *don’t* stop reading is part of what makes a persuasive essay good, so feel free to put it down if you feel like it).

Everyone is welcome to participate in this, especially people who don’t know anything about AI risk and especially especially people who think it’s stupid or don’t care about it.

I want to try doing this two different ways, so:

If your surname starts with A – M, try the first version of the experiment here at https://goo.gl/forms/8quRVmYNmDKAEsvS2

If your surname starts with N – Z, try the second version at https://goo.gl/forms/FznD6Bm51oP7rqB82

Thanks to anyone willing to put in the time.

SSC Endorses Clinton, Johnson, Or Stein

I.

If you are American, SSC endorses voting in this presidential election.

Andrew Gelman, Nate Silver, and Aaron Edlin calculate the chance that a single vote will determine the election (ie break a tie in a state that breaks an Electoral College tie). It ranges from about one in ten million (if you live in a swing state) to one in a billion (if you live in a very safe state). The average American has a one in sixty million chance of determining the election results. The paper was from the 2008 election, which was a pro-Obama landslide; since this election is closer the chance of determining it may be even higher.

The size of the US budget is about $4 trillion, but Presidents can only affect a tiny bit of that – most of the money funds the same programs no matter who’s in charge. But Presidents do shift budgetary priorities a lot. GW Bush started a war in Iraq which probably cost $2 trillion; the CBO estimates Obamacare may cost about $1.2 trillion. Neither of these are pure costs – Obamacare buys us more health care, and military presence in Iraq buys us [mumble] – but if you think these are less (or more) efficient ways to spend money than other possible uses, then they represent ways that having one President might be better than another. If we suppose a good president would use these trillions of dollars at least 33% more efficiently than a bad president, then this is still $300 billion in value.

So order of magnitude, having a good President rather than a bad one can be worth $300 billion. A 1/60 million chance to create $300 billion in value is worth $5,000; even the 1/1 billion chance afforded someone in a safe state is worth $300.

We don't know for sure that we're right about politics. In order to add signal rather than noise to the election results, we have to be better than the average voter. The Inside View is useless here; probably every voter thinks they're better than average. I recommend the Outside View – looking for measurable indicators correlated with ability to make good choices. Education's probably a good one. IQ might be another. But overall, my suggestion is that if you're seriously uncertain about whether or not you think more clearly than the average voter, by that fact alone you almost certainly do.

Suppose you live in a swing state. If you think (in a well-calibrated way) that it's 10% more likely that your candidate will use $1 trillion well than that the other candidate will, your vote is worth $500. If you live in a safe state, it's more like $30. If you value the amount of time it takes to vote at less than that, voting is conceivably a good use of your time.

II.

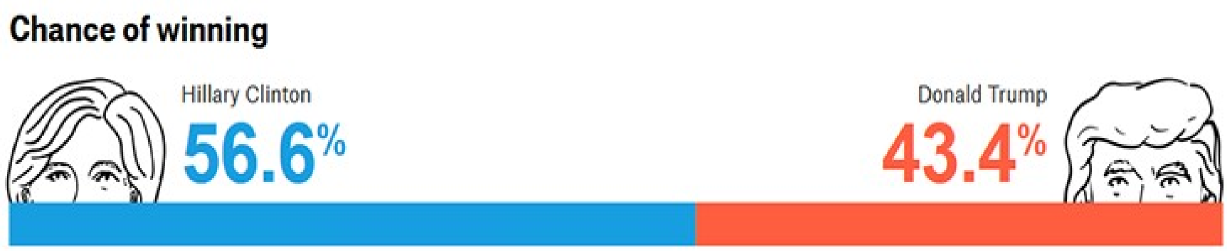
SSC endorses voting for Hillary Clinton if you live in a swing state. If you live in a safe state, I endorse voting for Clinton, Johnson, or (if you insist) Stein. If you want, you can use a vote-swapping site to make this easier or more impactful.

You might notice who's missing from this endorsement. I think Donald Trump would be a bad president.

Partly this is because of his policies, insofar as he has them. I'm not going to talk much about these because I don't think

I can change anyone's mind here – either you agree with me (and disagree with Trump) on things like abortion, global warming, free trade, et cetera, or you don't. A two sentence argument in a blog post won't change your mind either way.

In fact, I'm not sure any of this ever changes anyone's mind, and I didn't really want to write this post. But the latest news says:



This is going to be close. And since the lesson of Brexit is that polls underestimate support for politically incorrect choices, this is going to be *really* close.

And I don't know if I'd go so far as Scott Aaronson, who worries that he will one day live in a nuclear hellscape where his children ask him "Daddy, why didn't you blog about Trump?". But if some of my blogging on conservative issues has given me any political capital with potential Trump voters, then I this is where I want to spend it.

So here are some reasons why I would be afraid to have Trump as president *even if* I agreed with him about the issues.

Many conservatives make the argument against utopianism. The millenarian longing for a world where all systems are destroyed, all problems are solved, and everything is permissible – that's dangerous whether it comes from Puritans or Communists. These same conservatives have traced this longing through leftist history from Lenin through social justice.

Which of the candidates in this election are millennarian? If Sanders were still in, I'd say fine, he qualifies. If Stein were in, same, no contest. But Hillary? The left and right both critique Hillary the same way. She's too in bed with the system. Corporations love her. Politicians love her. All she wants to do is make little tweaks – a better tax policy here, a new foreign policy doctrine there. The critiques are right. Hillary represents *complete safety from millennialism*.

Trump's policy ideas are mostly silly, but no one cares, because he's not really running on policy. He's running on making America great again, fighting the special interests, and defying the mainstream media. Nobody *cares* what policies he'll implement after he does this, because his campaign is more an expression of rage at these things than anything else.

In my review of Singer on Marx, I wrote that:

> I'd always heard that Marx was long on condemnations of capitalism and short on blueprints for communism, and the couple of Marx's works I read in college confirmed he really didn't talk about that very much. It seemed like a pretty big gap. I figured…he'd probably made a few vague plans, like "Oh, decisions will be made by a committee of workers," and "Property will be held in common and consensus democracy will choose who gets what," and felt like the rest was just details. That's the sort of error I could at least sympathize with, despite its horrendous consequences.
>
> But in fact Marx was philosophically opposed, as a matter of principle, to any planning about the structure of communist governments or economies. He would

come out and say “It is irresponsible to talk about how communist governments and economies will work.” He believed it was a scientific law, analogous to the laws of physics, that once capitalism was removed, a perfect communist government would form of its own accord. There might be some very light planning, a couple of discussions, but these would just be epiphenomena of the governing historical laws working themselves out. Just as, a dam having been removed, a river will eventually reach the sea somehow, so capitalism having been removed society will eventually reach a perfect state of freedom and cooperation.

Singer blames Hegel. Hegel viewed all human history as the World-Spirit trying to recognize and incarnate itself. As it overcomes its various confusions and false dichotomies, it advances into forms that more completely incarnate the World-Spirit and then moves onto the next problem. Finally, it ends with the World-Spirit completely incarnated – possibly in the form of early 19th century Prussia – and everything is great forever.

Marx famously exports Hegel’s mysticism into a materialistic version where the World-Spirit operates upon class relations rather than the interconnectedness of all things, and where you don’t come out and call it the World-Spirit – but he basically keeps the system intact. So once the World-Spirit resolves the dichotomy between Capitalist and Proletariat, then it can more completely incarnate itself and move on to the next problem. Except that this is the final problem (the proof of this is trivial and is left as exercise for the reader) so the World-Spirit becomes fully incarnate and everything is great forever. And you want to plan for how that

> should happen? Are you saying you know better than the World-Spirit, Comrade?
>
> I am starting to think I was previously a little too charitable toward Marx. My objections were of the sort "You didn't really consider the idea of welfare capitalism with a social safety net" or "communist society is very difficult to implement in principle," whereas they should have looked more like "You are basically just telling us to destroy all of the institutions that sustain human civilization and trust that what is baaaasically a giant planet-sized ghost will make sure everything works out."

And since then, one of the central principles behind my philosophy has been "Don't destroy all existing systems and hope a planet-sized ghost makes everything work out". Systems are hard. Institutions are hard. If your goal is to replace the current systems with better ones, then destroying the current system is 1% of the work, and building the better ones is 99% of it. Throughout history, dozens of movements have doomed entire civilizations by focusing on the "destroying the current system" step and expecting the "build a better one" step to happen on its own. *That never works*. The best parts of conservativism are the ones that guard this insight and shout it at a world too prone to taking shortcuts.

Donald Trump does not represent those best parts of conservativism. To transform his movement into Marxism, just replace "the bourgeoisie" with "the coastal elites" and "false consciousness" with "PC speech". Just replace the assumption that everything will work itself out once power is in the hands of the workers, with the assumption that everything will work itself out once power is in the hands of "real Americans". Just replace the hand-waving lack of plans with what to do after the Revolution with a hand-waving lack of plans what to do after the election. In both cases, the sheer virtue of the movement, and the apocalyptic purification of the rich people keeping everyone else down, is supposed to mean everything will just turn out okay on its own. *That never works*.

A commenter on here the other day quoted an Atlantic article complaining that “The press takes [Trump] literally, but not seriously; his supporters take him seriously, but not literally”. Well, count me in that second group. I don’t think he’s literal. I think when he talks about building a wall and keeping out Muslims, he’s metaphorically saying “I’m going to fight for you, the real Americans”. When he talks about tariffs and trade deals, he’s metaphorically saying “I’m going to fight for you, the real Americans”. Fine. But neither of those two things are a plan. The problem with getting every American a job isn’t that nobody has been fighting for them, the problem with getting every American a job is that getting 100% employment in a modern economy is a really hard problem.

Donald Trump not only has no solution to that problem, he doesn’t even understand the question. He lives in a world where there is no such thing as intelligence, only loyalty. If we haven’t solved all of our problems yet, it’s because the Department of Problem-Solving was insufficiently loyal, and didn’t try hard enough. His only promise is to fill that department with loyal people who really want the problem solved.

I’ve never been fully comfortable with the Left because I feel like they often make the same error – the only reason there’s still poverty is because the corporate-run government is full of traitors who refuse to make the completely great, no-downsides policy of raising the minimum wage. One of the right’s great redeeming feature has been an awareness of these kinds of tradeoffs. But this election, it’s Hillary who sounds restrained and realistic, and Trump who wants the moon on a silver platter (“It will be the best moon you’ve ever seen. And the silver platter is going to be yuuuuuge!”)

III.

But I guess you’ve got to balance someone’s ability to pursue goals effectively with whether you like the goals they’ll be pursuing. I can imagine someone admitting that Clinton will probably be better at governing than Trump, but preferring Trump’s position on the issues so much that it still gives him an edge. In that case, I beg you to consider not only the mean but the variance.

I think even people who expect Trump to be a better President on average will admit he’s a high-variance choice. Hillary is an overwhelmingly known quantity at this point. A Hillary presidency will probably be a lot like Obama’s presidency. There might be a Libya-style military action; probably not an Iraq-style one. If something terrible happens like China tries to invade Taiwan, she will probably make some sort of vaguely reasonable decision after consulting her advisors. She might do a bad job, but it’s hard to imagine a course where a Hillary presidency leads directly to the apocalypse, the fall of American democracy, et cetera.

Trump isn’t a known quantity. Maybe he’ll kind of dodder around and be kind of funny while not changing much. Or maybe there will be some crisis and Trump will take what could have been a quickly-defused diplomatic incident and turn it into World War III. Remember also that it’s more likely the House and Senate both stay Republican than that they both switch to being Democrat. So if Hillary is elected, she’ll probably spend four years smashing her head against Congress; if Trump is elected, he will probably get a lot of what he wants.

Some people like high variance. I don’t. The world has seen history’s greatest alleviation of poverty over the past few decades, and this shows every sign of continuing as long as we don’t do something incredibly stupid that blows up the current world order. I’m less sanguine about the state of America in particular but I think that its generally First World problems probably can’t be solved by politics. They will probably require either genetic engineering or artificial intelligence; the job of our generation is keep the world functional enough to do the research that will create those technologies, and to alleviate as much suffering as we can in the meantime. I don’t see a Clinton presidency as making the world non-functional, whatever that means. I don’t know what I see a Trump presidency doing because, Trump is inherently unpredictable, but some major blow to world functionality is definitely on the list of possibilities.

The one place where Clinton is higher-variance than Trump is immigration. Clinton does not explicitly support open borders, but given her election on a pro-immigration platform and the massive anti-Trump immigration backlash that seems to be materializing, it’s easy to see her moving in that direction. If you believe that immigrants can import the less-effective institutions of their home countries, lower the intelligence of the national hive mind, or cause ethnic fractionalization that replaces sustainable democratic politics with ethnic coalition-building (unlike the *totally-not-ethnic-coalition-based* politics of today, apparently?), that could potentially make the world less functional and prevent useful technologies from being deployed.

I consider this one of the strongest pro-Trump arguments, but I think it exaggerates the scale of the problem. Hillary will have a Republican Congress to contend with; she probably won’t be able to increase immigration very much. Immigration rates are currently too low to cause massive demographic change before the point at which useful technologies can be deployed, and most immigrants are Asian and come from countries with pretty good institutions themselves. More important, Trump's anti-immigration policies would prevent foreign researchers from attending top American universities, and probably slow the deployment of future technologies directly, far more than any indirect effect from Hillary would.

There's another argument here – how exactly are we visualizing a world where immigrants damage American institutions? I envision it as America becoming more like Third World countries – constant ethnic tension, government by strongmen, rampant corruption, lack of respect for checks and balances, and overregulation of industry. But Trump is promising us all of that already, without even admitting any immigrants! If we're going to become a Third World country, let's at least help some people while we're doing it!

IV.

US conservatism is in crisis, and I think that crisis might end better if Trump loses than if he wins.

Since a country with thriving conservative and liberal parties is lower-variance than one with lots of liberals but no effective conservatism, I would like conservatism to get out of crisis as soon as possible and reach the point where it could form an effective opposition. It would also be neat if whatever form conservatism ended out taking had some slight contact with reality and what would help the country (this is not meant as a dig at conservatives – I'm not sure the Democrats have much contact with reality or helps the country either; I'm wishing for the moon and stars here).

Nobody expects Republicans to win blacks and Hispanics. The interesting thing about this election is that college-educated whites are also moving into the Democratic column. If the latest polls are to be believed, the demographic – which favored Romney by 14 points last election – favors Clinton by 8 points now. The nightmare scenario is that Trump wins, his style of anti-intellectual populism is cemented as Official New Republican Ideology, and every educated person switches to the Democrats.

I’m not 100% this would be bad – maybe educated people who are temperamentally conservative would pull the Democratic Party a little to the right, turning them into a broad moderate coalition which has no problem winning elections and combines the smartest elements of liberal and conservative thought. But more likely, there’s a vicious cycle where the lack of intelligent conservatives guts the system of think tanks that produce the sort of studies and analyses which convince smart people to become conservative, which in turn makes there even fewer intelligent conservatives, and so on. In the end, intellectuals won’t just vote Democrat; they’ll shift their personal views further to the left to fit in. We already have a problem with a glut of leftist researchers and journalists producing evidence why leftists are right about everything, and a shortage of conservative researchers and journalists to fact-check them and present the opposite case. As intelligent people desert the Republican Party, this situation gets worse and we lose access to any knowledge that Vox doesn’t want to write an explainer on. In the worst case scenario, everybody develops a hard-coded association between “conservative” and “stupid people”, even more than they have already, the academies purge the hell out of everyone even slightly to the right of the loudest activist, and the only alternative is The Donald Trump Institute Of Research That Is Going To Be Absolutely Yuuuuuuge, which busies itself putting out white papers to a coalition of illiterates.

If Trump fails, then the situation is – much the same, really, but conservatives can at least get started right now picking up the pieces instead of having to wait four years. There's a fundamental problem, which is that about 30% of the US population is religious poor southern whites who are generally not very educated, mostly not involved in US intellectual life, but form the biggest and most solid voting bloc in the country. If you try to form two parties with 50% of the vote each, then whichever party gets the religious poor southern whites is going to be dominated by them and end up vulnerable to populism. Since the religious poor southern whites are conservative, that's always going to be the conservative party's cross to bear and conservatism is always going to be less intellectual than liberalism in this country. I don't know how to solve this. But there have been previous incarnations of American conservatism that have been better at dealing with the problem than this one, and maybe if Trumpism gets decisively defeated it will encourage people to work on the problem.

V.

I said I wouldn't try to convince people about the big hot-button issues, but I've been told now thatthe guardrails of democracy have been broken lying is okay. So let's talk about global warming.

Most hot-button issues are less President-influenced than most people think. No Supreme Court is likely to overturn Roe v. Wade at this point, so the president's impact on abortion is limited to whatever edge cases come before the justices they appoint. I have no idea whether there was more or less capital punishment during Obama's administration than Bush's, but I doubt that the president's opinion of the issue had much impact one way or the other. But it looks like the Obama administration made really impressive progress on global warming; needless to say Donald Trump feels differently.

I don’t want to argue climate science here. I want to say that, as usual, I support the low-variance position that’s not going to make the world vastly less functional before we can invent genetic engineering or AI. Even if you doubt modern climate science, are you so sure it’s wrong that it’s worth the risk? What chance of global warming being a real problem would it take before you agreed that we should probably reduce CO2 emissions just in case? How could that chance possibly be lower than the chance of something that 90-something percent of the relevant scientists believe to be true is true? Yes, we know here that science is not always as authoritative as it would like to be, but it’s not completely anticorrelated with truth either!

(also, if the research about high CO2 levels decreasing cognitive ability is true – and my guess is no, but I’m far from sure – that could be even more disastrous than the traditional global warming effects – remember that even tiny IQ decreases have horrible consequences on a society-wide scale.)

VI.

Okay, but what about the real reason Trump is so popular?

When I talk to Trump supporters, it’s not usually about doubting climate change, or thinking Trump will take the conservative movement in the right direction, or even immigration. It’s about the feeling that a group of arrogant, intolerant, sanctimonious elites have seized control of a lot of national culture and are using it mostly to spread falsehood and belittle anybody different than them. And Trump is both uniquely separate from these elites and uniquely repugnant to them – which makes him look pretty good to everyone else.

This is definitely true. Please vote Hillary anyway.

Aside from the fact that getting back at annoying people isn't worth eroding the foundations of civil society – do you really think a Trump election is going to hurt these people at all? Make them question anything? "Oh, 51% of the American people disagree with me, I guess that means I've got a lot of self-reflecting to do." Of course not. A Trump election would just confirm for them exactly what they already believe – that the average American is a stupid racist who needs to be kept as far away from public life as possible. If Trump gets elected, sure, the editorial pages will be full of howls of despair the next day, but underneath the howls will be quiet satisfaction that the world is exactly the way they believed it to be.

The right sometimes argues that modern leftism is analogous to early millenarian Christianity. They *argue* this, and then they say "You know what would stop these people in their tracks? A strong imperial figure who persecutes them. That's *definitely* going to make them fade away quietly. There is no way this can possibly go wrong."

Leftism has never been about controlling the government, and really the government is one of the areas it controls least effectively – even now both houses of Congress, most state legislatures, most governors, etc, are Republican. When people say that the Left is in control, they're talking about academia, the media, the arts, and national culture writ large. But all of these things have a tendency to define themselves in opposition to the government. When the left controls the government, this is awkward and tends to involve a lot of infighting. When the right controls the government, it gets easy. If *Trump* controls the government, it gets *ridiculously* easy.

This has real-world effects. Millennials are more conservative than previous generations. Andrew Gelman, who is usually right about everything, says:

> If you look at the cohort of young voters who came of age during George W. Bush’s presidency, they’re mostly Democrats, which makes sense as Bush was a highly unpopular Republican. The young voters who came of age during Obama’s presidency are more split, which makes sense because Obama is neither popular nor unpopular; he has an approval of about 50%

I would prefer the next generation end up leaning more to the right, because that will cancel out younger people’s natural tendency to lean left and make them pretty moderate and so low-variance. I definitely *don’t* want an unpopular far-right presidency, because then they’re going to lean left, which will combined with the natural leftiness of the young and make them *super* left. And *this* is the sort of thing that affects the culture!

VII.

One more warning for conservatives who still aren’t convinced. If the next generation is radicalized by Trump being a bad president, they’re not just going to lean left. They’re going to lean regressive, totalitarian, super-social-justice left.

Everyone has already constructed the narrative: Trump is the anti-PC, anti-social-justice candidate. If he wins, he’s going to be the anti-PC, anti-social-justice President. And he will fail. First of all, because he doesn’t really show much sign of knowing what he’s doing. Second of all, because all presidents fail in a sense – 80% of Americans consistently believe the country is headed the wrong direction and the president is the natural fall guy for this trend. And third of all, because even if by some miracle Trump avoids the first two failure modes, the media will say he failed and people will believe them. And when the anti-PC, anti-social-justice President fails, the reaction will be a giant “we told you so” from the social justice movement, and a giant shift of all the disillusioned young people right into their fold.

Trump is all set to be the biggest gift to the social justice movement in history. They thrive on claims of persecution, claims that they’re the ones fighting a stupid hateful regressive culture that controls everything. And people think that bringing their straw man to life and putting him in the Oval Office is going to *help*?

If you’re a Jew fighting anti-Semitism, the absolute minimum you can do is not actually kill Christian children and use their blood to make matzah. Likewise, if you are a principled classical liberal fighting the social justice movement’s attempt to smear anyone who disagrees with them as an overprivileged clueless hateful Neanderthal, the absolute minimum you can do is not actually be an overprivileged clueless hateful Neanderthal. Opinions on Trump range all the way from “he is definitely an overprivileged clueless hateful Neanderthal” to “he is remarkably and uniquely bad at not appearing to be an overprivileged clueless hateful Neanderthal”. In any case, having him as the public face of anti-social-justice for the next four years would be a godsend for them and a disaster for everyone else.

VIII.

There’s one more thought I wanted to mention which is vaguely in this space.

The enemy isn’t leftism or social justice. The enemy is epistemic vice.

When the Left errs, it’s through using shouting and shaming to cut through the long and painful process of having to justify its beliefs. It’s through confusing disagreement with evil, a dissenter who needs convincing with a thought-criminal who needs neutralizing.

Sometimes it might be strategically necessary to whack particular ideologies to make examples of them. But in the longer-term, replacing left with right just puts a new group of people in position to shame their opponents and silence dissent. The long range plan has to combine a short-term need to neutralize immediate would-be tyrants with a long-term need to slowly encourage epistemic virtue so that we don’t *have* to keep putting out fires.

Now, watch this video:

Trump’s not in that crowd. But does anyone think he disagrees with it? Can anyone honestly say that Trump or his movement promote epistemic virtue? That in the long-term, we’ll be glad that we encouraged this sort of thing, that we gave it power and attention and all the nutrients it needed to grow? That the road to whatever vision of a just and rational society we imagine, something quiet and austere with a lot of old-growth trees and Greek-looking columns, runs through LOCK HER UP?

I don’t like having to vote for the lesser of two evils. But at least I feel like I know who it is.

RELATED: Eliezer, The Unit Of Caring, Scott Aaronson

He Kept Us Out Of War?

I.

Some of the best pushback I got on my election post yesterday was from people who thought Trump was a safer choice than Clinton because of the former’s isolationism and the latter’s interventionism. Since I glossed over that point yesterday, I want to explain why I don’t agree.

Trump has earned a reputation as an isolationist by criticizing the Iraq War. I don’t think that reputation is deserved. He’s said a lot of things which suggest he would go to war at the drop of a hat.

— He says he will “bomb the s#!t out of ISIS” and calls for sending 30,000 troops to destroy them. His campaign website says he will “pursue aggressive joint and coalition military operations to crush and destroy ISIS”.

— He is ambiguous about whether Obama should have intervened in Syria to depose dictator Bashar Assad. He complained “there is something missing from our president. Had he crossed the line and really gone in with force, done something to Assad – if he had gone in with tremendous force, you wouldn’t have millions of people displaced all over the world. ”

— Back during the rebellion in Libya, Trump seems to have been in favor of even more dramatic intervention than Obama eventually allowed. He said on his video blog “I can’t believe what our country is doing. Qaddafi in Libya is killing thousands of people, nobody knows how bad it is, and we’re sitting around we have soldiers all have the Middle East, and we’re not bringing them in to stop this horrible carnage and that's what it is: It's a carnage. You talk about things that have happened in history; this could be one of the worst. Now we should go in, we should stop this guy, which would be very easy and very quick. We could do it surgically, stop him from doing it, and save these lives. This is absolutely nuts. We don't want to get involved and you're gonna end up with something like you've never seen before. But we have go in to save these lives; these people are being slaughtered like animals. It's horrible what's going on; it has to be stopped. We should do on a humanitarian basis, immediately go into Libya, knock this guy out very quickly, very surgically, very effectively, and save the lives."

— He thinks we should have "kept" Iraq's oil. When pressed on how exactly one keeps billions of barrels of petroleum buried underneath a foreign country, he said he would get US troops to circle and defend the areas with the oil. The "areas with the oil" are about half of the country. This sounds a lot like he wants US troops to remain in Iraq indefinitely.

— He also wants to to keep Libya's oil. As per National Review: "I would go in and take the oil — I would just go in and take the oil. We don't know who the rebels are, we hear they come from Iran, we hear they're influenced by Iran or al-Qaeda, and, frankly I would go in, I would take the oil — and stop this baby stuff."

— He suggests declaring war on Iran as a response to them harassing US ships. During the debate, he said he would "shoot their ships out of the water."

— In 2007, he he suggested "knocking the hell out of [Iran] and keeping their oil", though in his (sort of) defense he might have been confusing them with ISIS at the time.

— In his 2000 book *The America We Deserve* he suggested a preemptive strike on North Korea: “[If I were President], North Korea would suddenly discover that its worthless promises of civilized behavior would cut no ice. I would let Pyongyang know in no uncertain terms that it can either get out of the nuclear arms race or expect a rebuke similar to the one Ronald Reagan delivered to Ghadhafi in 1986. [Reagan bombed Libya]. I don’t think anybody is going to accuse me of tiptoeing through the issues or tap-dancing around them either. Who else in public life has called for a pre-emptive strike on North Korea?”

— During a town hall meeting, when host Chris Matthews asked Trump when he would use nuclear weapons, he answered “Somebody hits us within ISIS—you wouldn`t fight back with a nuke?” When Matthews reminded him that most people try to avoid ever using nuclear weapons, he answered “Then why are we making them? Why do we make them?”

II.

Some writers have called the period since World War II the “Pax Americana”. Although there have been some deadly local wars, there’s been relative peace between great powers. A big part of this is America’s promise to defend its allies. This both prevents other countries from attacking America’s allies *and* prevents America’s allies from building big militaries and launching attacks of their own. The whole system is cemented by America-centric trade organizations which make war unprofitable and incentivize countries to stay in America’s orbit.

Trump wants to destroy this system because it costs money, even though it doesn’t really cost that much money compared to anything else we do and Trump intends to increase the defense budget anyway. It’s possible a post-Trump world might find some other way to maintain peace. It’s also possible that it wouldn’t, or that the process of finding that alternative way would be really bloody.

— In March, Trump said “I think NATO may be obsolete. NATO was set up a long time ago — many, many years ago when things were different. Things are different now. We were a rich nation then. We had nothing but money. We had nothing but power. And you know, far more than we have today, in a true sense. And I think NATO — you have to really examine NATO. And it doesn’t really help us, it’s helping other countries. And I don’t think those other countries appreciate what we’re doing.” Although this isn’t the *worst* opinion, most foreign policy scholars think that our policy of defending our allies is necessary to prevent global arms races and random regional wars.

— In July, he publicly admitted he wasn’t sure he would protect the Baltic states if Russia attacked, something we’re currently obligated to do. The Atlantic calls this “a marked departure from the security policy of every presidential nominee from either of the two major parties since NATO’s founding in 1949”. It’s especially worrying because even if you’re not going to protect the Baltic states from Russia, you shouldn’t openly *say so* where Russians can hear you!

— And throughout the race, Trump has campaigned on a platform that would effectively end American participation in the World Trade Organization. Trump understands that this would probably start a global trade war, but asks “who the hell cares if there’s a trade war?” I care for two reasons. First, because free trade has produced decades of sustained economic growth and the most successful poverty alleviation in human history. Second, this would probably crash the world economy, creating exactly the sort of depression that tends to produce instability (most famously Hitler's rise during Germany's interwar stagnation) or which drives countries toward regional hegemons willing to trade with them or just plain bribe them.

III.

Hillary's foreign policy isn't great either, but it doesn't seem as bad as some people are making it out to be.

— Hillary will probably continue US intervention in Syria; here she is more interventionist than Obama. But her intervention would probably be smaller-scale than Trump's. She wants to arm "friendly" rebel groups and enforce a no-fly zone, but she has ruled out sending ground troops into Iraq or Syria, something Trump has promised to do. Likely she would focus on keeping enough of Syria safe to protect some civilians and prevent more refugees, then use indirect methods to make life miserable for Assad. This seems like as good a plan as any other.

— The main concern I've heard is that the no-fly zone might lead to conflict (war?) with Russia. Declaring a no-fly zone would mean a commitment to shoot down any plane that flies through the zone. Russia is currently flying planes through Syria, and if they tried to call Hillary's bluff she would have to shoot down Russian planes or lose credibility; shooting down a foreign plane could obviously lead to war. Many different news sources make this point (1, 2, 3, etc). But the clearest description she's given of what she wants suggests a no-fly zone *with Russian cooperation and support*. Last October, she said of her no-fly zone proposal that "I think it's complicated and the Russians would have to be part of it, or it wouldn't work." There's some good discussion of this on Reddit (see especially this comment) where most people end up agreeing that this is the heart of her plan – something like the US agreeing it won't bomb Russian allies if Russia doesn't bomb our allies.

— Hillary has said she will "treat cyberattacks just like any other attack", which could mean that if Russia launches a cyberattack on the US (for example hacking the DNC's emails) Hillary would treat it as an act of war. I think this requires a stretch. She did mention the possibility of a military response, but only in the context of possible "serious political, economic, and military responses". My guess is we should interpret this in a non-crazy way – if Russia hacks our emails, we condemn them and maybe hack some of their stuff. If Iran hacks a dam and causes it to fail, *then* maybe we start thinking airstrikes. Shooting down an airliner is an act of war, but countries have shot down other countries' airliners a bunch of times and usually people posture a bit and then let it slide. I don't think it makes sense to think Hillary will treat cyber-attacks more seriously than that.

IV.

A lot of this has a lot of room for interpretation. I'm totally ready to believe that when Trump said he would shoot any Iranian ship that annoyed US vessels, he just meant generic macho posturing and expected everyone to hear it that way. He might even be cunningly pursuing a North Korean – style "mad dog" strategy where he tries to sound so dangerous and unpredictable that nobody dares call his bluff, and so his enemies never mess with him in any way.

Or he might mean everything he says. After all, a lot of it has been pretty consistent since long before he was running for president. There's no point in saying things to send a game theoretic signal to Iran if you're a random New York real estate developer and Iran isn't listening. If he understood the theory behind sounding trigger-happy to intimidate our enemies, he probably wouldn’t have openly admitted he wouldn’t respond to a Russian invasion of the Baltics. And he does seem kind of 100% like a loose cannon in every way, to the point where trying to explain away loose-cannon-like statements as part of a deeper plan seems overly complex.

(Actually, I have a theory which I think explains a lot about Trump’s foreign policy positions: he doesn’t like losers. He supported the Iraq War and the Libya intervention when it looked like we would probably win. Then we lost, and he said they were stupid and bungled. He supports counterfactual invasions of Iraq and Libya where we “kept the oil” because that would have counted as winning. He supports invading ISIS because he expects to be in charge of the invasion and he expects to win. Under this theory, Trump’s retrospective non-support for failed wars doesn’t predict that he won’t start new ones.)

In the end it all comes back to the argument from variance. *Maybe* Trump is secretly a principled isolationist, and he’s only saying he’ll shoot at Iran and invade Libya and first-strike North Korea and steal oil from Iraq and send troops against ISIS and remove Assad in order to scare people into cooperating with him. Or *maybe* he’ll actually shoot at Iran and invade Libya and first-strike North Korea and steal oil from Iraq and send troops against ISIS and try to remove Assad. Who knows? He’s said a thousand times now that he’s totally different from the usual politicians, and I believe him. He could do pretty much anything.

(I’d like to think his advisors would rein him in before that point, but when asked which advisors he would consult before a major foreign policy decision, Trump could only think of one person, and he does not exactly inspire confidence.)

I am not qualified to judge Hillary’s work as Secretary of State, but I expect her to play by the book. I’m not sure if Hillary will be more aggressive or more peaceful than the last few presidents, but I don’t expect her to be a wild outlier totally beyond comparison to any previous president. I expect her to consult the foreign policy community on anything important she does, and take some advice relatively within their Overton Window. If she comes to the brink of nuclear war with Russia, I expect her to de-escalate for the same reason I expect Putin to de-escalate; they’re both rationally self-interested people who want to continue being alive and ruling their respective countries, and they value that more than any particular principle or any opportunity to prove their machismo.

I think she remains the low-variance choice for president.

Somewhat Against Psychiatric Conditions As Domestication Failure

[Epistemic status: Not sure if I'm arguing against a straw man here and my conclusion is what the researchers meant all along.]

I.

Benitez-Burraco and Lattanzi theorize that autism and schizophrenia are anomalies in the human self-domestication process. I'll try to explain, but for a much better explanation than I can give read Dr. Chris Badcock here.

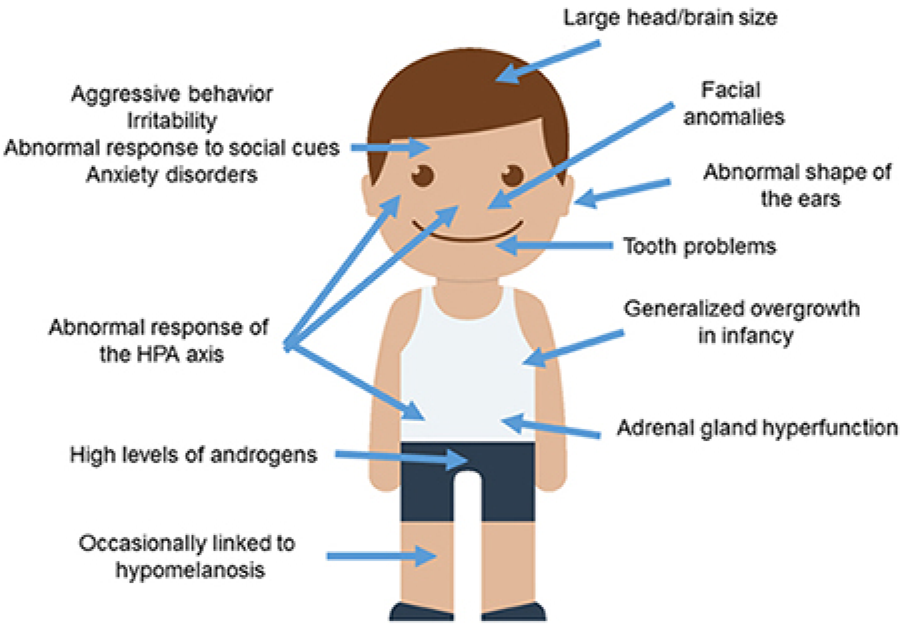
Still here? Fine. BBL's theory goes like this. When Russian scientist Dmitry Belyaev tried to domesticate foxes by breeding them for tame behavior, he found that changes in a lot of other traits went along for the ride. In short, the foxes started looking kind of dog-like: smaller heads, shorter snouts, spotted fur, floppy ears, more youthful characteristics. Some further experiments confirmed that similar changes happen in any species bred for tameness. Probably this has to do with changes in the neural crest, an embryonic structure which goes on to form a bunch of things including the adrenal medulla. Since the adrenal medulla produces some of the hormones involved in fear and stress, animals with hypoactive medullae will probably be tamer. But since the neural crest also goes on to form lots of other stuff, or produce hormones that influence the formation of lots of other stuff, these tamer animals will be different in other ways too.

BBL continues: we went from being wild apes to tame humans, a process that could be analogized to "self-domestication". Some of the same changes the Russians saw in the transition from wild to domesticated foxes can be seen in the transition from early hominid skulls to modern human skulls.

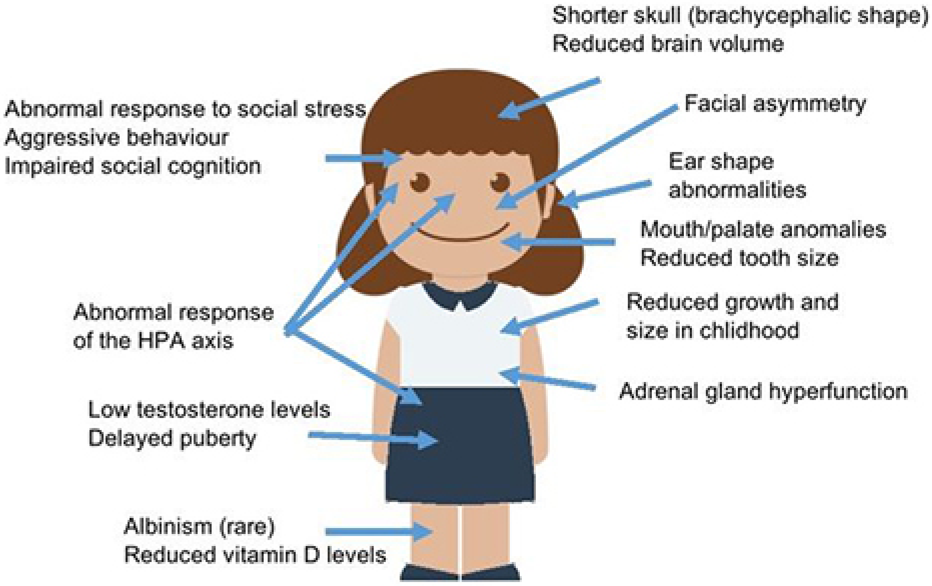
Autistic people, say BBL, are "undomesticated humans" – people in whom for some reason the neural crest changes that result in domesticated features have reversed. They find that some of the changes of domestication syndrome are the reverse of some of the symptoms of autism:

Smaller heads = autistic people have larger heads
More trusting and social = autistic people are less trusting and social
Spotted fur = the depigmenting disease "hypomelanosis of Ito" is sometimes associated with autistic symptoms
Floppy ears = studies find autistic people are more likely to have abnormally shaped ears (really!)
Change in adrenal response = autistic people have abnormal function in the HPA axis, the system including the adrenal gland

Or in the form of their cutesy picture:


Large head/brain size
Aggressive behavior
Irritability
Abnormal response to social cues
Anxiety disorders
Facial anomalies
Abnormal shape of the ears
Tooth problems
Abnormal response of the HPA axis
Generalized overgrowth in infancy
Adrenal gland hyperfunction
High levels of androgens
Occasionally linked to hypomelanosis

Schizophrenics, say BBL, are “hyperdomesticated humans”. Once again, they match up the symptoms:


Shorter skull (brachycephalic shape)
Reduced brain volume
Abnormal response to social stress
Aggressive behaviour
Impaired social cognition
Facial asymmetry
Ear shape abnormalities
Mouth/palate anomalies
Reduced tooth size
Abnormal response of the HPA axis
Reduced growth and size in chlidhood
Adrenal gland hyperfunction
Low testosterone levels
Delayed puberty
Albinism (rare)
Reduced vitamin D levels

I originally thought this theory was dumb. After looking into it more, I think it has some serious issues, but that there might be a core of truth.

II.

I’ll get to that core, but first, the argument against: all of this is coincidences, pareidolia, and finessing things to fit into a system where they don’t really belong.

Going down the list:

> Smaller heads = autistic people have larger heads

Some studies find this is true. Others find that it isn’t. In any case, note a discrepancy between this claim and the schizophrenia version. BBL note smaller brains in schizophrenics (true) and shorter skull (true), but not smaller heads, which we would expect if autism were the “reverse” of schizophrenia. In fact, schizophrenics may have larger heads than healthy people. This sort of moving the goal-posts, where autistics are judged on their larger *heads* but schizophrenics on their smaller *brains*, is a red flag for fake pattern-matching.

> More trusting and social = autistic people are less trusting and social

True! But schizophrenics are *way* less trusting and social! Paranoia – pathological inability to trust – is a classic symptom of schizophrenia; indeed, if you made people choose between schizophrenia and autism and asked which one was associated with lack of trust, I think most people would choose schizophrenia. This brings an important point into relief: the whole point of domestication is that the domesticated animal is supposed to be friendlier and less aggressive. But nobody would describe schizophrenics as friendlier and less aggressive.

> Spotted fur = the depigmenting disease “hypomelanosis of Ito” is sometimes associated with autistic symptoms

True! But hypomelanosis of Ito is a really rare disease (1/10,000 births) that has nothing to do with most autism. Also, it causes eye problems, kidney cysts, weirdly-shaped chests, short stature, seizures, mental retardation, etc. To me this looks more like “a super-rare disease that can cause pretty much anything can sometimes also cause autistic symptoms”, which is not very interesting. Also, domestication causing “pigmentation changes” (usually spotted fur) versus autism being (very rarely) associated with a depigmenting disease and schizophrenia being (very rarely) associated with albinism is more goalpost-shifting.

> Floppy ears = studies find autistic people are more likely to have abnormally shaped ears (really!)

I looked at this study – Manouilenko et al – and what it actually finds is that autistic people are more likely to have asymmetrical ears. In fact, nonsignificantly more likely to have asymmetrical ears; their significant finding is that autistic people have more "minor physical abnormalities", and the asymmetrical ears were one of many pieces of evidence combined to get the significant finding. But asymmetrical features are common in lots of genetic/embryological diseases and seem like a general sign of high mutational load. It seems sketchy to combine autists' asymmetrical ears and wild foxes' pointy ears and say "Look, they both have ear abnormalities, this is the same thing!" Some other studies suggest that autistic people have low-set ears, which sounds more promising, but schizophrenic people *also* have low-set ears, so whatever. The other schizophrenia ear findings are exactly as unconvincing as the autistic ones.

> Change in adrenal response = autistic people have abnormal function in the HPA axis, the system including the adrenal gland

Wikipedia's page on the HPA axis has a section on its possible role in disease, which states that dysfunction of the axis is involved in various conditions "including anxiety disorder, bipolar disorder, insomnia, posttraumatic stress disorder, borderline personality disorder, ADHD, major depressive disorder, burnout, chronic fatigue syndrome, fibromyalgia, irritable bowel syndrome, and alcoholism". In other words, in a list of the HPA axis' twelve greatest hits, neither autism nor schizophrenia qualify for inclusion.

I’m not saying that there isn’t an HPA axis component to these diseases. I’m just saying HPA axis is a nonspecific finding. I’m agnostic whether the HPA axis causes everything or our HPA axis study methods are so bad that they invariably turn up false positives. The point is that we shouldn’t get too excited when we see the HPA axis involved in both domestication and autism. This is like saying “Cancer causes you to feel bad, and AIDS causes you to feel bad, therefore cancer causes AIDS.” No, it’s just that everything makes you feel bad.

When we look beyond the general claim of “abnormal function”, things get less clear. BBL say that domesticated animals have “reduced levels of stress hormones including adrenocorticoids, adrenocorticotropic hormone, cortisol, and corticosterone”. So their theory should predict that autistic people have increased stress hormones, and schizophrenics decreased, relative to typical people. Actually, it’s a mess; autistic people seem to have higher ACTH but lower cortisol; schizophrenia studies are conflicting but tend towards higher levels of both. Once again, they can support a general claim of “these conditions affect the same system”, but they can’t predict the direction of the effect. Also, *every* condition affects this system.

(if you’re wondering why we’re talking about cortisol levels in a theory about the adrenal medulla, well, so am I. Whatever.)

Finally, if by “undomesticated human” we mean something like an ape or Neanderthal, well, neither apes nor Neanderthals (as far as we know) display the symptoms of autism. They seem to be pretty social. They seem to be able to eat all kinds of stuff without trouble. They don’t seem bothered by sensory processing problems. For that matter, dogs and cattle and nth generation silver foxes, the most domesticated animals we've got, don't seem very schizophrenic. I guess cows could just be hallucinating all the time and how would we know, but there doesn't seem to be any *evidence* that they are.

So this is why I originally was not too big on this theory.

III.

But what about Williams Syndrome?

Benitez-Burraco and Lattanzi don't mention Williams Syndrome (also called Williams-Beuren Syndrome) at all, which is crazy, because it sounds a thousand times more like a syndrome of hyperdomestication than either of the two conditions they examine (h/t Nicholas Wade and a random Reddit comment). Williams Syndrome is a rare condition (1/10,000 births) caused by the deletion of some genes on chromosome 7. There are three very interesting things about people with Williams Syndrome. Number one, they are *really* nice. Like if you meet someone with Williams Syndrome, you will think "This person clearly has a rare genetic disease that causes pathological levels of niceness as a symptom." Number two, they are really trusting. An Atlantic article profiling the condition, What Happens When You Trust Too Much? describes special therapy for Williams Syndrome children where the therapist has to teach them, painfully and laboriously, how to distrust people. NPR calls it "essentially biologically impossible for kids [with Williams Syndrome] to distrust [people]." Number three, they *talk all the time*; the informal name for the condition is "cocktail personality syndrome".

People with Williams Syndromes actually legitimately have short noses (compare to the short snout on domesticated foxes), smaller teeth (compare to smaller teeth in dogs vs.

wolves), smaller brains, and “unusually shaped ears” (I can’t find anything more specific; I guess it’s too much to hope for that researchers actually describe the ears as “floppy”).

Also, somebody checked which gene was most different in dogs versus wolves, and they found it was WBSCR17. The WBS in the name stands for “Williams-Beuren Syndrome” because it’s been linked to the disorder. So there’s that.

So as far as I can tell there’s an *amazingly* good case for Williams Syndrome being linked to domestication. Williams Syndrome tends to cause severe mental retardation and death at an early age, but that’s probably because there are twenty-five totally different genes missing. Maybe a version that only deleted WBSCR17 would keep the behavioral and physiologic changes but not much else.

A lot of people suggest Williams’ Syndrome is “the opposite of autism”. I can only find three pieces of evidence for this. Number one, the obvious contrast with the love of social situations and high verbal skills. Number two, Williams Syndrome kids seem to be really good at face recognition, whereas autistic people are often worse at this. Number three, Williams’ Syndrome kids seem to be unusually bad at the puzzles and interlocking-mechanical-part type problems on which autistic people excel.

On the other hand, there are some reasons to think these conditions are *not* exact opposites. For one thing, autism is caused by a hideously complex interplay of thousands of genes and various environmental factors, but Williams Syndrome is a drop-dead simple “oops, we forgot part of this chromosome over here”. Williams Syndrome kids seem to have some of the same sensory sensitivities as autistic kids. And both groups usually suffer from mental retardation.

I think that Williams Syndrome establishes the possibility of a physiological social/trust system linked to domestication, the neural crest, and various other parts of embroygenesis. Once you admit the existence of such a system, it seems like autism probably involves some kind of damage to it – probably along with damage to a lot of other systems too. Schizophrenia is more of a stretch, but the overwhelming presence of distrust as a symptom makes the existence of a physiological social/trust system at least kind of interesting and relevant.

So maybe instead of saying that “autistic people are undomesticated humans” and “schizophrenics are hyperdomesticated humans”, we should say something like “there is a very subtle and hard-to-notice biological system that determines level of trust and sociability and which seems weirdly linked to ear and nose shape; autism, schizophrenia, and Williams Syndrome all affect that system in different ways.” Note that this doesn’t mean they’re “the same disease” or “opposite diseases”; the connection might be no deeper than the “connection” where heart attacks, atrial fibrillation, and getting stabbed in the chest all affect the heart. But they all hit the same system.

My take-home message from looking into all of this is that I was very silly for trying to learn about autism and schizophrenia without thinking about embryology. These are highly genetically-loaded diseases that present early in life and seem linked to teratogens and prenatal infections; of *course* they’re embryological! I had to take some embryology classes in medical school, and like everyone else I tuned them out because they seemed totally irrelevant to real clinical practice and mostly involved memorizing pointless trivia like “on day thirty-six and a half, the developing shmendroblast has transformed into a blexomere”. But if you want to know what causes secret connections between ear shape and level of social trust, embryology seems like the way to go. Autism and schizophrenia are hard to study because they seem to affect everything, yet nothing specifically enough to localize the condition. Maybe going back and thinking more embryologically could help pinpoint the particular systems involved.

More Hardball Debate Questions

[See also Hardball Questions For The Next Debate. The Gary Johnson question is not original to me.]

Jill Stein:

You’re a former doctor and researcher who first got involved in politics because of your interest in public health. One of your first forays into activism was the 2000 publication of In Harm’s Way: Toxic Threats to Child Development, a magisterial report on the effect of pollution on children’s physical and mental health. You described your focus as being on “developmental disabilities, including attention deficit/hyperactivity disorder, autism, and related neurodevelopmental diseases”. You describe “accumulating evidence of neurotoxic damage to children by environmental agents, such as lead and PCBs”.

In Chapter 7, you discuss the high burden of pesticides eaten by developing children, saying that:

> Twenty million American children five and under eat an average of eight pesticides every day through food consumption. Thirty-seven pesticides registered for use on foods are neurotoxic organophosphate insecticides, chemically related to more toxic nerve warfare agents developed earlier this century…a national health exposure study detected chlorpyrifos residues (as the metabolite TCP) in the urine of 82% of a representative sample of American adults. A more recent study in Minnesota revealed that an even higher 92% of children had detectable levels of this metabolite in their urine.

You connect this increasing pesticide exposure to what you believe to be increasing levels of developmental disorders in

American children:

> The Cailfornia Department of Developmental Services released [a study] in March 1999 [that] looked at pervasive developmental disorders from 1987 through 1998 and showed a 210 percent increase in cases entered into the autism registry during those years. If the incidence of autism is increasing, and/or clusters of autism are being discovered, an environmental influence is likely.

Maybe as a result, you’ve become a big advocate of eating organic food. Your party platform says you want to “support organic and regenerative agriculture” and “put a moratorium on GMOs and pesticides until they are proven safe”. You presented a pro-organics case on Bill Moyers’ show back in 2012, and you’re even known for preparing your own organic meals on the campaign trail.

But there’s a lot of pushback from mainstream scientists and the mainstream media. For example, news webzine Vox has an interesting article Is Organic Food Any Healthier? Most Scientists Are Still Skeptical publicizing a meta-analysis of 237 studies which showed that “organic foods didn’t appear to be any healthier or safer to eat than their conventionally grown counterparts” and that “typical exposure to pesticide residues is at levels 10,000 to 10,000,000 times lower than doses that cause no observable effect in laboratory animals that are fed pesticides daily throughout their entire lifetimes”. Vox has also written Local And Organic Food Has Extra Safety Risks. Just Ask Chipotle. Vox’s spinoff webzine Eater even makes fun of customers looking for “natural” foods without having any idea what that means.

If they’re right, then you’re promoting an unscientific fad that has millions of people needlessly stressed out about everything they eat. On the other hand, if *you’re* right, then these media outlets’ pooh-poohing of a vital public health message makes them complicit in and maybe even responsible for what you call the “epidemic” of childhood neurodevelopmental disorders.

So my question for you is: do you believe Vox ‘zines cause autism?

Hillary Clinton:

During your first debate with Donald Trump, the moderator asked you about racial bias in police shootings; you responded that “implicit bias is a problem for everyone, not just police”. You argued that you would “put money into that budget to help us deal with implicit bias by police officers” and that we’ve “got to do everything possible to improve policing, to go right at implicit bias”. Your running mate Tim Kaine continued on the theme, saying that people shouldn’t be afraid to bring up police officers’ implicit biases.

This talk of implicit bias references a whole psychological field centered around the Implicit Association Test. It works like this: a subject sitting in front of a keyboard is shown rapid-fire pictures representing various categories – classically black people, white people, positive adjectives, and negative adjectives. They’re given various instructions about which keys to press in response to which categories, and their responses are timed. Many people will find that it’s easier to press the same key for white people and positive adjectives (and an opposite key for black people and negative adjectives) than to press the same key for whites+negatives and blacks+positives. This has been widely considered to show implicit racism – that is, even people who say they are not racist unconsciously associate black people with bad qualities. This research has become wildly popular, profiled in every major media outlet, and catapulted its inventors to scientific stardom. It's even been featured on the Oprah Winfrey show, maybe a first for a social psych paper.

A few early small studies suggested the IAT predicted prejudiced behavior. But later attempts to replicate this result failed. Blanton, Jaccard, Klick, Mellers, Mitchell, Tetlock (2009) reanalyzed some of the original studies, found no effect, and complained that the IAT was being popularized despite an almost-complete lack of evidence for its validity. Oswald et al (2013) did a meta-analysis of 275 implicit association test results from 46 different studies and found that "IATs were poor predictors of every criterion category other than brain activity, and the IATs performed no better than simple explicit measures". Carlsson and Agerstrom did another meta-analysis earlier this year, and found "the overall effect was close to zero and highly inconsistent across studies" and "there is…little evidence that the IAT can meaningfully predict discrimination, and we thus strongly caution against any practical applications of the IAT that rest on this assumption".

You particularly mention the IAT as relevant to policing, but Dolan Group, a consulting firm which advises police forces how to avoid racial discrimination, did an internal analysis of results surrounding the IAT and reports to its clients that:

> Persons who do not hold overt racist attitudes do not have to worry about some deeply-hidden, unknown, unconscious attitudes influencing their work decisions. These findings reveal the need to aggressively weed out officers who hold conscious racial stereotypes and biases in order to avoid biased-based policing. These

> findings also raise questions about whether the money and time spent on law enforcement training and testing regarding implicit bias could be put to better use on something else.

When you and your running mate suggest a focus on implicit bias as relevant to policing, this can really only be justified by taking the preliminary findings of a few small early studies and ignoring both more rigorous reanalysis of their results and the consensus finding of all studies and meta-analyses conducted since that time.

On the other hand, there still is something to be explained here: if the IAT isn't analyzing implicit racial prejudice, why *do* people so consistently have an easier time associating black people with negative adjectives? I actually have a theory of my own about that. Consider claims like the following:

1. Black people were brutally enslaved for hundreds of years.
2. Black people are almost three times more likely than whites to live below the poverty line.
3. Black people are systematically being murdered by the criminal justice system.
4. Black people are frequent victims of racism and hate crimes.
5. Our society is set up to structurally discriminate against black people.

None of these claims are racist per se; in fact, many of them are anti-racist in intent. But all of them connect black people to negative affect! If your local newspaper says that white people usually have friendly and positive interactions with the police but black people are victimized and killed by police, that is some *heavy* association of whites with positive feelings and blacks with negative feelings. If you usually see photos of white people in the news under the headline “LOCAL BUSINESS BOUGHT BY GOOGLE”, and photos of blacks in the news under the headline “LEARN HOW OUR RACIST SOCIETY KEPT THIS POOR WOMAN FROM SUCCEEDING” then once again, you’re learning to associate whites with positive feelings and blacks with negative feelings.

This would explain very nicely why people taking the IAT generally associate whites with positive feelings and blacks with negative feelings in a way apparently unrelated to whether they are explicitly prejudiced/racist. It would also explain very nicely why about 50% of *blacks* associate whites with positive feelings and blacks with negative feelings, which is definitely a thing that happens and which previous explanations of have always sounded unconvincing and *ad hoc*.

But from your debate statements, it sounds like you are absolutely opposed to this reinterpretation. That you are committed to defending the position that implicit bias is a real predictor of racism, and that Implicit Association Tests don’t just report contingent associations drilled in by the media, but genuinely reveal profound unconscious beliefs about how the world works.

So my question for you is: would you be willing to take an Implicit Association Test measuring how easily you associate your own name vs. your opponents’ names with the adjective “crooked”?

Gary Johnson:

If you were elected, what would you do about the ongoing crisis in Updog?

Donald Trump:

You’re well-known for your boast that you “hire the best people”. And one of those best people is Steve Bannon, the CEO of your campaign. When Bannon took over on August 17th, 538 had you at only a 12% chance of winning; after he was running your campaign for a month, you were up to 40%. Although you’ve since crashed back down, a lot of political observers attribute what successes you’ve had to Bannon and what problems you’ve had to your own big mouth. You seem to recognize his utility, calling him one of “the best talents in politics, with the experience and expertise needed to defeat Hillary Clinton in November”.

Before he joined your campaign, Bannon was best known for his role leading far-right news website Breitbart. But he was actually involved in some pretty interesting stuff when he was younger. In particular, in 1993 Bannon was the acting director of the famous environmental science experiment Biosphere 2.

Biosphere 2 was an attempt to create a self-sustaining closed ecosystem capable of supporting human life, possibly with applications for future space travel. It was actually the *first* such attempt – it was called “Biosphere 2” because the first such self-sustaining biosphere was the Earth itself. Eight “crew members” entered the facility along with various plants and animals, the airlocks were sealed, and for a year everyone tried to do what they could to keep the various species and environmental parameters in balance.

It didn’t work; CO2 levels started fluctuating wildly, soil microbes surged out of control, ants and cockroaches overran the facility, oxygen dropped to worrying levels, and the experiment was stopped early out of concern for crew health. They decided to try a second mission, and that was when they had a change in management and brought on Mr. Bannon as director.

Unfortunately, a lot of the crew members really didn’t like Bannon and his team. Possibly some of it had to do with an incident where a crew member submitted a list of safety complaints and Bannon threatened to “shove it down her f**king throat”. It got so bad that some of the crew deliberately vandalized the Biosphere, causing gas exchange between the inside and the outside and ruining the scientific value of the experiment. Although they probably could have tried again, by that time lawsuits and financial mismanagement had sapped their funding, and they finally sold the whole thing off to Columbia University as a research campus.

So my question for you is: in all of history, there have only been two self-sufficient ecosystems capable of maintaining human life. Your team has already destroyed one of them. The other is Earth. How scared should we be?

The Moral Of The Story

[content warning: puns. This is mostly self-plagiarism from my Tumblr and Twitter]

Once upon a time there was a small desert village with a single well outside town. One day a young woman went to the well to fetch water, and the well heard her crying, and asked “What’s wrong?”

She stopped her sobbing and asked the well “You can talk?”

“Yes,” said the well. “Long ago, the witch who lives in this town gave me life so I could serve as a guardian to the townspeople.”

“Alas,” said the young woman. “I am the daughter of that witch. She lived in peace with the townsfolk for many years. But the new mayor, who is a violent and hateful man, riled the people up against her, and they burned her at the stake. I am young and still do not know very much magic. I tried to curse them, but my curses fizzled. Now I worry I will never avenge my mother’s death.”

“Do not be afraid,” said the well. “I will take care of this.”

The next morning, when the Mayor came to fetch water from the well, he heard an odd noise coming from the bottom. He peered over as far as he could to see what was happening. Then an impossibly long arm shot up from the bottom of the well, grabbed the mayor, and pulled him into the well shaft. There was a horrible crunching sound, and nobody ever saw the Mayor again. The townsfolk apologized to the witch’s daughter, and they all lived happily ever after.

Moral of the story: Living well is the best revenge.

Pixar’s movie *Up* won the Academy Award for “Best Picture” and was widely hailed as one of the best children’s films of the decade. In fact, some people argued it was *too* good, and that kids were ignoring school, chores, and other responsibilities to watch it again and again. They said that along with the cute plot, the short, catchy name gave it an almost drug-like addictive quality. This made a lot of people very angry, and Pixar agreed to give its addictive must-watch movies longer names in the future.

Moral of the story: Do not call *Up* what you cannot put down.

There’s a new report out of CERN that a team of scientists has unraveled the structure of the photon. Apparently this started years ago when some equations showed that photons acted like tiny “hands” – structures with a “palm” and radiating “fingers” – which “crawl” across time/space and “grab” the solid particles they interact with. This explained most of the properties of light but wasn’t an exact match for the data. The latest result is that single photons are actually made up of hundreds of these shapes, all joined together into a single particle, and this is how they’re able to travel so quickly.

Moral of the story: Many hands make light work.

Once upon a time there was an ugly duckling. All of the other ducklings had grown their beautiful white soft downy feathers, but this duckling had no down feathers at all and was bald and ugly and all the other ducklings teased him.

He went to the mysterious crow who lived in the woods and asked for help. The crow said to repeat the magic words “HOCUS POCUS” at midnight with a full moon, and then he would grow his down feathers. The duckling tried that, but the moon just laughed at him and said the magic had no power here.

So he went to the creepy raven who lived in the swamp and asked for help. The raven said to repeat the magic words “ABRA CADABRA” at high noon on a sunny day, and then he would grow his down feathers. The duckling tried that, but the sun just laughed at him and said he wasn’t bound by the magic.

So he went to the wise old owl who lived in the tallest tree and asked for help. The owl explained that the duckling should just ignore the mockery of the other birds and accept that he was okay just the way he was, because there were no magic spells to make ducklings grow feathers.

Moral of the story: You are beautiful, no matter what they say. Words can’t bring you down.

Once upon a time a young lady died and went to Hell. At the check-in desk, Satan asked her age. She was in her twenties, but looked much younger; she thought quick and realized that even in Hell, they probably wouldn’t be mean to children. So she told Satan that she was twelve, and sure enough he said she wasn’t old enough to be held accountable for her sins, and ushered her off to a more peaceful part of Hell reserved for ages eleven through thirteen. She met the other sinners there and realized that many of them, like her, were older people who had lied to get out of their punishment.

Satan began to suspect something like this was going on, so he set up hidden cameras in the 11-13 wing of Hell, trying to catch people acting like adults or admitting to one another that they had lied about their age. But there were hundreds of millions of sinners and Satan couldn’t monitor all the cameras himself. So he went up to the mortal world and asked for the best supercomputer they had. The mortals recommended a newer model of Deep Blue, the supercomputer that had first beaten a human world champion at chess. Satan picked one up from IBM and went back to Hell, where he programmed the Deep Blue to monitor all of the hidden camera feeds at once and report any suspicious activity.

Sure enough, after a few days, he got thousands of reports of people acting older than thirteen. He hunted them down and removed them to Hell proper, where there was much wailing and gnashing of teeth. And it all could have been avoided if they had just stuck to their charade and acted as young as they said they were.

Moral of the story: Don’t get caught, be tween – the Devil and the Deep Blue see.

By 2050, screens have shrunk and become more flexible until the dream of “programmable paper” becomes a reality. Citizens of the future read newspapers like the ones in Harry Potter that include moving images and even videos of important events. This new technology even makes it as far as the US Post Office, where they decide to include programmable stamps. Instead of a static picture of eg George Washington’s head, it will have a moving image of Washington speaking and giving his famous Farewell Address.

Unfortunately, the technology isn’t ready for the kind of abuse that envelopes undergo on their travel throughout the country and the world. Most of the computerized stamps become corrupted and “crash”; in a particularly common bug, they try to reload but just end up displaying “GENERATING IMAGE…” permanently. The government has no money to fix the problem, so people just get used to stamps on their letters that say “GENERATING IMAGE…” instead of having interesting pictures on them.

Moral of the story: If you want a vision of the future, imagine a human face booting on a stamp forever.

AI Persuasion Followup Survey

If you took the AI persuasion experiment survey last month, I have a very brief followup survey that I'd like you to take. You can find it here. Don't worry, you don't have to read anything this time.

If you didn't take the survey last month, please don't take this one either.

The Heart Has Its Reasons That Reason Knows Not Of

I.

Psychoanalysts argue that sons are attracted to women who look like their mothers, because they imprint on their mothers and use them as a schema for their ideal woman.

(and probably something similar for daughters and their fathers, though the psychoanalysts don't usually get around to talking about that)

I've counterargued that sons get half their genes from their fathers, who apparently were attracted to women who looked like the sons' mothers. Since study after study shows genes having unexpectedly big effects on behavior, and early childhood experiences with parents having unexpectedly small effects, maybe this would be a better explanation for the effect (if it even exists).

But the research doesn't really bear me out.

Goose-ologist Konrad Lorenz raised goslings from birth. When the goslings grew up, he found they tried to mate with humans, especially "Caucasian men with white beards". He concluded that they imprinted on their adoptive parent (him) and learned to prefer mates who looked like that parent.

More formally, Bischof et al got some male zebra finches and arrange to have them raised by a closely-related species, Bengalese finches. Then they put them in cages with both female zebra finches and female Bengalese finches and observed which females the birds tried to court.

The results were pretty striking; they overwhelmingly went for the Bengalese finches who looked like their mothers, not the zebra finches who were genetically more suitable. Spence & Smith replicated this finding with zebra fish raised by differently-colored zebra fish.

So the research shows conclusively that sexual selection is based on learned imprinting, at least in animals whose names start with the string “zebra fi”. What about humans?

I can’t find the study itself, but multiple reviews cite Jedlicka 1984, who looked at children of mixed-race couples (white and native Hawaiian). They found that both men and women were more likely to marry someone of the race of their opposite-sex parent than of their same-sex parent (eg if you’re a woman with a Hawaiian mother and white father, you’re more likely to marry a white person). This is consistent with some kind of social imprinting where your opposite-sex parent serves as a template for future romantic interest. It’s not consistent with a simple genetic theory where you just get both parents’ genes. It *might* be consistent with a more complicated genetic theory where mate preferences are on a sex-appropriate chromosome or get chromosomally imprinted such that you only care about your father’s preferences for women and your mother’s preferences for men, but this is hard and I haven’t seen any analysis of whether it’s evolutionarily worth it.

Enquist, Aronsson, Ghirlanda, Jansson, and Jannini (2010) starts its Methods section with “We obtained data through newsgroups alt.sex.fetish and alt.sex.fetish.breastmilk”, so you know it’s gonna be interesting. They test another feature of men sexually imprinting on their mothers: suppose you’re a man with a sibling a few years younger than yourself. That means your mother was pregnant or lactating during the supposed critical sexual imprinting window. So if men with younger siblings are more likely to have pregnancy and lactation fetishes, that suggests that sexual imprinting on mothers is really a thing. This is indeed what they found: when a person with a pregnancy or lactation fetish only had one sibling, there was a 66% chance (compared to expected 50%) that their sibling would be younger than they, $p < 0.0001$ in their sample of 560 such people. This was true if and only if they were between age 1.5 and 5 during their sibling’s birth, hinting at the span of the imprinting window. Of course, this is still a really poor predictor: 33% of such people got the fetish without any younger siblings, and of course most people with younger siblings don’t end up with the fetish at all. But it does look like *something* is going on.

(I wonder what’s up with adult baby fetishes; if you could find a similar pattern among them that would suggest they’re imprinting on the baby sibling instead of the mother, which would be fascinating. Maybe I should try to survey the appropriate subreddit.)

Hefferman and Fraley did a very similar study. They find that people born to older parents, when compared against people born to younger parents, find older faces more attractive. This was true even after controlling for the age of the participants themselves. The effect size was small but pretty consistent across different groups and measurements. I am not quite as happy about the quality this study as in the ones above, but nothing raises huge red flags.

So there’s a lot of circumstantial evidence suggesting some kind of imprinting process is going on. But what about the original question – do people choose mates who look like their opposite-sex parent? And how genetic versus environmental is it?

Bereczkei, Gyuris, and Weisfeld addresses this question directly. They get a sample of grown-up adopted daughters and their adoptive fathers. They show subjects pictures of the adoptive fathers taken when their daughters were 2-8 years old, and then photos of four similar-aged men, one of whom is the daughter's current husband. The subjects' task is to guess which of the photos is the husband based on which of them looks most like the adoptive father. And subjects are in fact able to do this better than chance; they pick the correct husband 37% of the time compared to 21% of the time for (each of the three) incorrect husbands, which in their sample of 242 subjects is a p-less-than 0.001 result.

Somebody else check if I'm wrong here, but I'm a little concerned about the way they calculated p-values in this study. Suppose that their 242 subjects are always accurate in identifying the husband who looks most like the father; no randomness or noise here. But now the chance for randomness and noise comes in how many of the women might have picked husbands who look like their fathers by pure coincidence. It looks like there were only 26 women used in the study, so there's a lot more opportunity for those twenty-six people to coincidentally choose husbands who happened to look like Dad than for the 242 subjects to coincidentally choose the correct husband even if they all looked alike. Yet as far as I can tell, the p-value calculation was only done on the latter possibility.

(an analogy: suppose that I want to prove I have ESP. I predict that a coin will land heads, then flip the coin. It does indeed land heads. Then I survey one thousand observers to ask them to judge whether the coin landed in accordance with my psychic prediction. All 1,000 of them say yes, a $p = 0.000001$ result that could not possibly be by coincidence. This does not prove at the $p = 0.000001$ level that I have ESP!)

Rantala and Marcinkowska do a really excellent review of this field that found a lot of the same papers I did. They acknowledge the Bereczkei/Gyuris/Weisfeld study as very interesting, but note that the first author "replicated" these results in a 2009 study using more scientific facial measurements. That study was found to be extremely flawed (found by Rantala himself, in fact) and was retracted under a cloud of accusations of misconduct and/or inexcusable error. It's still not clear exactly what happened there, and nobody has formally accused the original father/adopted-daughter study of anything, but it's kind of awkward to hang your theory on a study by somebody accused of scientific misconduct for faking a very similar study that supported the same hypothesis.

(side note: I wish that people would explain retractions better. Right now I don't feel like I have a good understanding of exactly how the 2009 study got retracted, so I haven't learned anything that will help me spot similar problems in the future)

So in summary: there's a lot of good evidence that animals learn mate preferences by imprinting on their parents. There's some circumstantial evidence that humans might do this too, across traits as diverse as race, age, and lactation status. There is, however, as yet no smoking gun.

II.

Wait, what about twin studies? Those are usually pretty great! Can we just do a twin study and sort this out once and for all?

Zietsch et al look at "a large community-based sample of twins and their partners and parents (N > 20,000 individuals) to test for genetic and family environmental influences on mate choice”. They find:

> …near-zero genetic influences on male and female mate choice over all traits and no significant genetic influences on mate choice for any specific trait **(!!!!!)**. A significant family environmental influence was found for the age and income of females’ mate choices, possibly reflecting parental influence over mating decisions.

This might be the first twin study I’ve ever seen which unambiguously breaks Turkheimer’s First Law Of Behavioral Genetics (every trait is somewhat heritable). Such a striking finding should increase our confidence in all of the above experiments a lot. So okay, I guess this issue is solved, it’s definitely just sexual imprinting on the opposite-sex parent, thank goodness, for once we have a perfectly clear noncontradictory result and we can all just go home and –

> We also tested for evidence of sexual imprinting, where individuals acquire mate-choice criteria during development by using their opposite-sex parent as the template of a desirable mate; there was no such effect for any trait.

Sigh.

Okay, fine, let’s look at this a little more closely. They were analyzing a bunch of data from a big Australian survey of twins. This included the twins, the twins’ family, and the twins’ mates “height, BMI, age, education, income, personality, social attitudes, and religiosity”, which were the dimensions along which they tried to predict mate choice. They figured out whether imprinting was involved by checking whether twins’ mates were more like their opposite-sex parent than like other members of their family (their same-sex parent, their other twin). There was no sign of this being true, not even a nonsignificant trend.

If we want to play the dangerous game of trying to explain differences between contradictory studies instead of just dismissing everything as noise, I might argue that this looked at some pretty different variables compared to the last set. Instead of looking at facial similarities, it's looking at things like social attitudes and religiosity; young children trying to imprint on their mother's image can maybe be forgiven for not knowing her opinion about Asian immigrants (one of the "social attitudes" questions they asked).

The authors took a different tactic and pointed out that most of us don't marry the first person we have a crush on, or even the second or third. Sometimes we marry the tenth person we really like, sometimes we settle for people we only like a little, and sometimes we get drunk, have sex in a cheap motel, and have the person's parent threaten us with a shotgun unless we go to the chapel *right now*. So maybe the person we end up marrying isn't a good proxy for our mate preferences per se.

(this is starting to get kind of depressing)

They investigate this hypothesis in a followup study where they directly ask twins about their preferences for an ideal mate. Here they're able to get some more immediately visual data – preferences like tall/short, long-hair/short-hair, beards/clean-shaven, even big-breasts/small-breasts. They find…well, some things come out heritable, but the confidence intervals are really wide. There are a lot of suspicious things like hair-length preferences being impressively and significantly heritable for women but not heritable at all for men, and on one hand the two findings' 95% confidence intervals do overlap, but on the other hand that’s because all of the confidence intervals are super-wide anyway. Overall I guess it’s nice that this study doesn’t blatantly break Turkheimer’s First Law, but I’m not going to draw any sweeping conclusions off of it.

I have one more twin study here, Lykken and Tellegen (which coincidentally opens with the same Pascal quote as this post). This experiment has an interesting design – they ask dizygotic and monozygotic twins to rate how attracted they are to their co-twin’s spouse! (I hope these researchers went above and beyond in keeping all their data confidential!) Despite those human interest stories where two identical twins separated at birth rediscover each other and find that they have both married blonde Corgi-owning optometrists named Theo, in this study there was very little resemblance between the spouses of either identical or fraternal twins. On average identical twins correlated with each other at r = 0.57 on some long list of variables, but their spouses correlated with each other only at r = 0.14 (identical) and r = 0.11 (fraternal), and most of this was just similar religious and educational backgrounds that don’t surprise us (we already know people tend to marry others of the same religion and social class, and twins are no exception).

This is interesting because it suggests a minimal role of either genetics *or* shared environment in mate choice, even in the most extreme circumstance (identical twins raised together). And in fact, the authors remind us that people who have had multiple long-term relationships often choose people who don’t resemble *each other* in more than the most superficial ways (if the authors had exposure to polyamory, they might note that most people’s simultaneous partners aren’t very similar either).

Maybe this just supports Zietsch et al’s hypothesis that mate preferences, whether imprinted or genetic, don’t really matter because we don’t have a whole shopping aisle of mates lined up to choose from and we’ve got to take whatever we can get. Yet even so there’s room for surprise. Even if only 5% of the opposite sex is interested in you, that still leaves the average person hundreds of different choices over their lifetime. Surely there should be *some* degrees of freedom for people to pick who they end up with. So what’s going on? If not by genes or shared environment, how *do* we choose our significant others?

The authors’ guess: true love, a mysterious and magical thing totally beyond scientific understanding.

> If we provisionally accept our interpretation of these data, we are left with a curious and disquieting conclusion: Although most human choice behavior lawfully reflects the characteristics of the chooser and of the choice, the most important choice of all, that of a mate, seems to be an exception…we outline a theory that is compatible with these interpretations, namely that human pair bonding is relatively adventitious, based on romantic infatuation which, as Stendhal observed, “is like a fever that comes and goes quite independently of the will.”

Sure. Sounds plausible. Let’s just consider this whole matter closed.

AI Persuasion Experiment Results

I.

Last month I asked three thousand people to read some articles on AI risk and tell me how convinced they were. Last week, I asked them to come back and tell me some more stuff, to see if they stayed convinced.

I started off interested in the particular articles – which one was best at convincing newcomers to this topic. But I ended up hoping this could teach me about persuasion in general. Can online essays change people's minds about complicated, controversial topics? Who is easiest to convince? Can we figure out what features of an essay do or don't change people's minds?

Depending on the last digit of people's birth dates, I asked them to read one of five different essays:

— I asked people whose birth dates ended with 0 or 1 to read Wait But Why's The AI Revolution: The Road To Superintelligence.

— I asked people whose birth dates ended with 2 or 3 to read a draft version of my own Superintelligence FAQ.

— I asked people whose birth dates ended with 4 or 5 to read Lyle Cantor's Russell, Bostrom, And The Risk Of AI.

— I asked people whose birth dates ended with 6 or 7 to read Michael Cohen's Extinction Risk From Artifical Intelligence.

— And I asked people whose birth dates ended with 8 or 9 to read Sean Carroll's Maybe We Do Not Live In A Simulation. This had nothing to do with AI risk and was included as a placebo – that is, to get a group who had just had to read an online essay but presumably hadn't had their minds changed about AI.

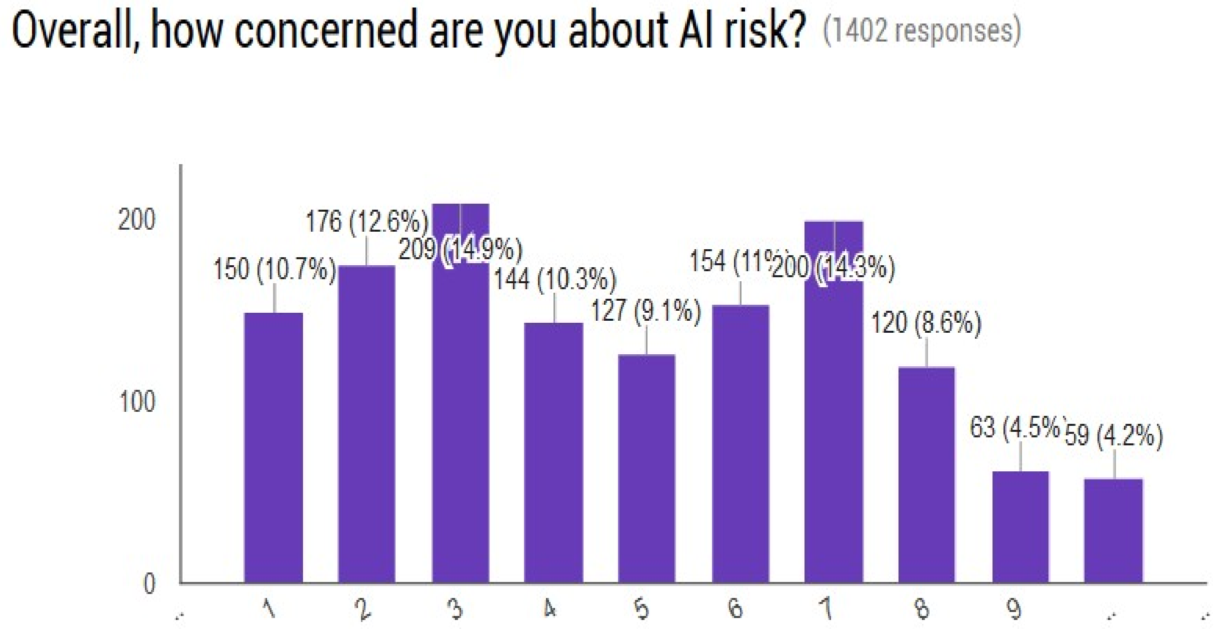
I hosted all of these pieces on Slate Star Codex and stripped them of any identifiers so that hopefully people would judge them based on content and not based on how much they liked the author or what color the page background was or whatever. It mostly worked: only 67% of readers had no idea who had written the essay, with another 23% having only vague guesses. Only about 10% of readers were pretty sure they knew.

People did read the essays: 70% of people said they finished all of theirs, and another 22% read at least half.

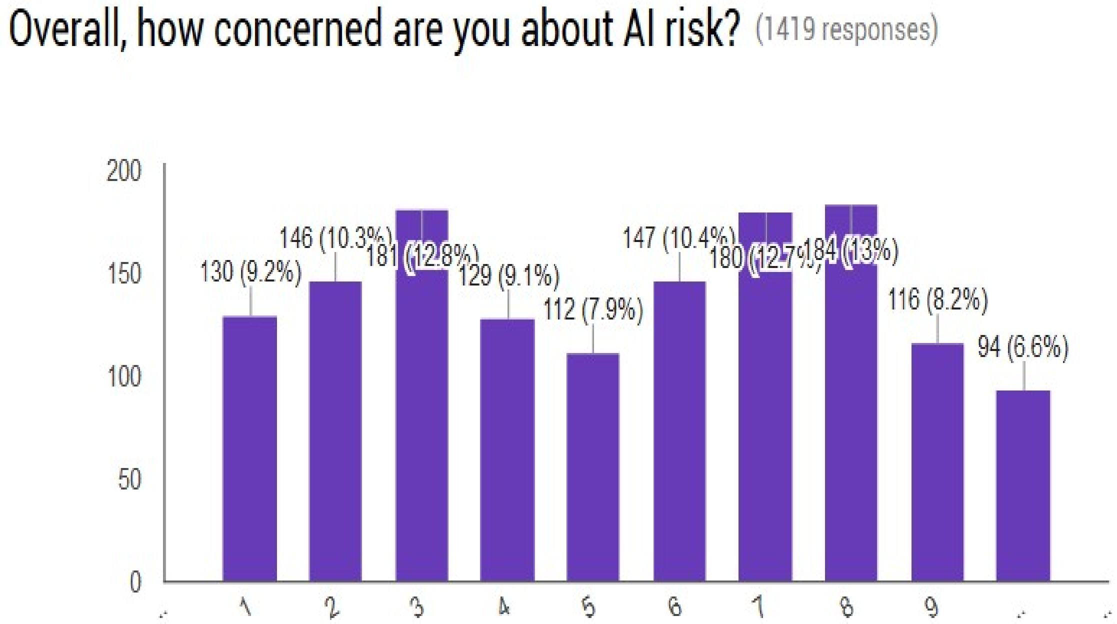
So the experiment was in a pretty good position to detect real effects on persuasion if they existed. What did it find?

II.

My outcome was people's ratings on a score of 1 – 10 for various questions relating to AI. My *primary* outcome, selected beforehand, was their answer to the question "Overall, how concerned are you about AI risk?" About half of the respondents took a pre-test, and I got the following results:


Overall, how concerned are you about AI risk? (1402 responses)

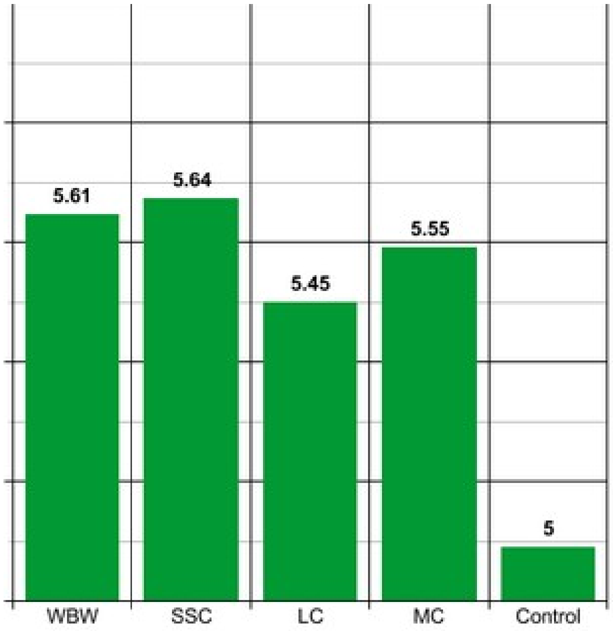
After reading the essays, this changed to:


Overall, how concerned are you about AI risk? (1419 responses)

Overall, people increased their concern an average of 0.5 points.

(Note that I only had half the sample take a pretest, because I was worried that people would anchor to their pretest answers and give demand effects. This turned out not to be a big deal; the people who had taken a pretest changed in the same ways as the people who hadn't. I've combined all data.)

But what about the different essays? All five groups had means between 5 and 6, but the exact numbers differed markedly.



Note truncated y-axis

On further testing, the differences between the four active essays weren't significant, but the difference between all of the essays and the control *was* significant.

III.

Aside from the primary outcome, I also had various secondary outcomes: answers to specific questions about AI risk. First, a list of average pretest and posttest answers (pretest, **posttest**):

How likely are we to invent human-level AI before 2100?: **5.7, 5.9**
How likely is human-level AI to become superintelligent within 30 years?: **7.3, 7.3**
How likely is human-level AI to become superintelligent within 1 year?: **4.8, 5.2**
How likely that a superintelligent AI would turn hostile to humanity?: **6.4, 6.8**
How likely that a hostile superintelligence would defeat humans?: **6.8, 7.1**

All the nonzero differences here were significant.

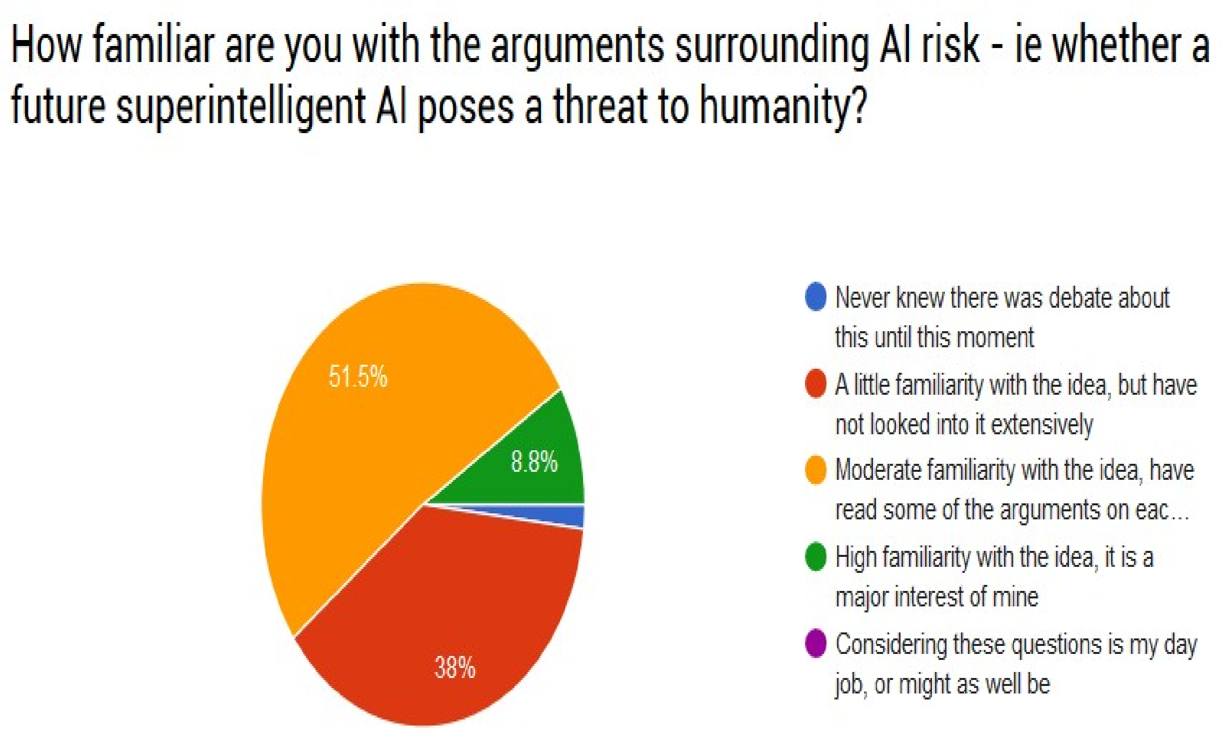
If we look at this as a conjunction of claims, all of which have to be true before AI risk becomes worth worrying about, then it looks like the weak links in the chain are near-term human-level AI and fast takeoff. Neither of these are absolutely necessary for the argument, so this is kind of encouraging. There is less opposition than I would expect to claims that AI will *eventually* become superintelligent, or to claims that a war with a superintelligent AI would go very badly for humans.

Given the level of noise, there wasn't a lot of evidence that any of the (active) essays were more persuasive than others on any of the steps, including the two dubious steps. This is actually a little surprising, since some essays focused on some things more than others. Possibly there was a "rising tide lifts all boats" effect where people who were more convinced of AI risk *in general* raised their probability at every step. But there was too much noise to say so for sure.

IV.

Was there any difference among people who were already very familiar with AI risk, versus people who were new to the topic?

It was hard to tell. Only about two percent of readers here had never heard about the AI risk debate, and 60% said they had at least a pretty good level of familiarity with it.

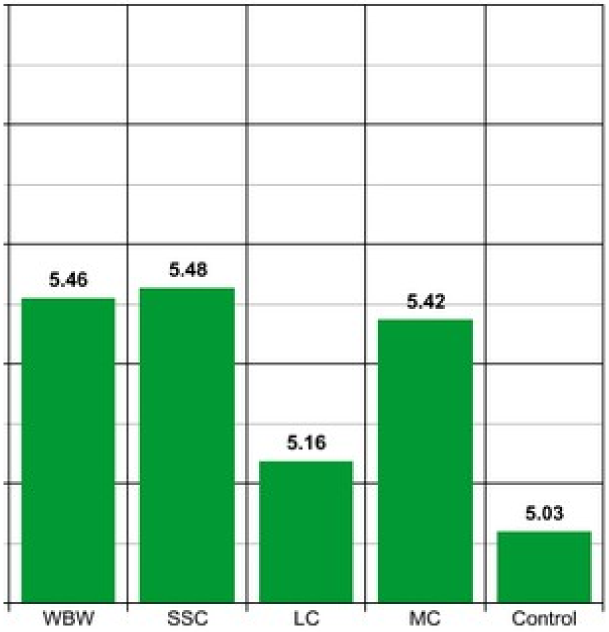


The best I could do was to look at the 38% of participants (still about 1000 people!) who had only “a little familiarity” with the idea. Surprisingly, these people’s minds didn’t change any more than their better-informed peers. The average “little familiarity” person became 0.8 points more concerned, which given the smaller sample size wasn’t that different from the average person’s 0.5.

In general, people who knew very little about AI thought it was less important (r = 0.47), which makes sense since probably one reason people study it a lot is because they think it matters.

V.

How stable were these effects after one month?



Still truncated y-axis

Pretty stable.

Other than an anomalous drop in the third group (I’m sticking with “noise”), the effect remained about two-thirds as strong as it had been the moment after the participants read the essays. All essay groups remained significantly better than the control group.

I looked at the subgroup of people who’d had little knowledge of AI risk before starting the experiments. There was a lot more noise here so it was harder to be sure, but it seemed generally consistent with the same sort of effect.

So in conclusion, making people read a long essay on AI risk changed their overall opinions about half a point on a ten-point scale. There weren’t major differences between essays. This was true whether or not they were already pretty familiar with it. And about two-thirds of the effect persisted after a month.

At least in this area, there might be a modest but useful effect from trying to persuade people.

In terms of which essay to use to persuade people? I don’t have any hard-and-firm results, but there were two trends I noticed. First, my essay was somewhat underrepresented among the people who’d had the biggest jumps (3 points or more) in their level of concern. Second, the third essay was anomalously bad at maintaining its gains over the month. That leaves just the first and fourth. Some people in the comments said they were actively repulsed by the fourth, and WaitButWhy seems pretty good, so at the risk overinterpreting noise, you might as well just send people theirs.

You can find the chaotic and confusingly-labeled data below. It might help to read the survey itself to figure out what’s going on.

Main experiment: .xlsx, .csv
One month follow-up: .xls, .csv

How Does Recent AI Progress Affect The Bostromian Paradigm?

[content note: I seriously know nothing about this and it’s all random uninformed speculation]

I.

AI risk discussions are dominated by the Bostromian paradigm of AIs as highly strategic agents that try to maximize certain programmed goals. This paradigm got developed in the early 2000s, before a recent spurt of advances in machine learning. Do these advances require any changes to the way we approach these topics?

The latest progress has concentrated in neural networks – “cells” arranged in layers that represent the potential for ascending levels of abstract categorization. For example, a neural network working on image recognition might have a low-level layer that scans the image and resolves it into edges, a medium-level network that scans the set of edges and resolves it into shapes, and a highest-level network that scans the shapes and resolves them into subjects and themes. With enough training, the network “learns” how best to map each level onto the level above it, ending up with profound insight into the high-level features of a scene.

These are a lot like the human brain, and in fact some of the early researchers got important insights from neuroscience. The brain certainly uses cells, the cells are arranged in layers, and the brain categorizes things in hierarchies that move from simple things like edges or sounds to complicated things like objects or sentences.

In particular, these networks are like the brain’s sensory cortices, and they’re starting to equal or beat human sensory cortices at important tasks like recognizing speech and faces.

(I think this is scarier than most people give it credit for. It’s no big deal when computers beat humans at chess – human brains haven’t been evolving specific chess modules. But face recognition is an adaptive skill localized in a specific brain area that underwent a lot of evolutionary work, *and modern AI still beats it*)

The sensory tasks where AIs excel tend to involve abstraction, categorization, and compression: the thing where you take images of black dogs, white dogs, big dogs, little dogs, ugly dogs, cute dogs, et cetera and are able to generalize them into “dog”. Or to take a more interesting example: a new AI classifies images as pornographic or safe-for-work. Its structure naturally gives it an abstract understanding of pornographicness that allows it to “imagine” what the most pornographic possible images would look like (trigger warning: artificially intelligent computer generating the most pornographic possible images). This kind of classification/categorization/generalization ability is a major advance and eerily reminiscent of human abilities.

But how far is this to building an AGI or human-level AI or superintelligence or whatever else you want to call it?

II.

Consider two opposite perspectives:

The engineer’s perspective: Categorization ability is just one tool out of many. When people invented automated theorem-provers, that was pretty cool – it meant computers could now assess new mathematics. But for AGI, you still need some *thing* that wants to prove theorems, something (someone?) that can do something with the theorems it proves. The theorem-prover is a tool for the AI to use, not the core “consciousness” of the AI itself. The same will be true of these new neural nets and deep learning programs. They can recognize dogs, and that’s cool. But AGI is still about creating some kind of program that *wants* to recognize dogs, and which can do something interesting with the dogs once it recognizes them. And that will probably require something different from either a theorem-prover *or* a neural-net-categorizer. A paperclip maximizer might use a neural net to recognize paperclips, but its desire to maximize them will still come from some novel architecture we don’t know much about yet which probably looks more like normal programming.

The biologist’s perspective: The whole brain runs on more or less similar cells doing more or less similar things, and evolved in a series of tiny evolutionary steps. If we’ve figured out how one part of the brain works, that’s a pretty big clue as to how other parts of the brain work. The human motivation system is in brain structures not so different from the human perception-association-categorization system, and they probably evolved from a common root. If researchers are discovering that the easiest way to make perception-association-categorization systems is neural nets reminiscent of the brain, then they’ll probably find that those neural nets are pretty easy to alter slightly to make a motivational system reminiscent of the brain. This would look less like strategic/agenty goal maximization, which the brain is terrible at, and more like the sort of vague mishmash of desires which humans have.

The exact evolutionary history beyond the biologist’s perspective is complicated. There’s a split between some sensory processing centers (like the visual cortex) and some motivational/emotional centers (like the hypothalamus) pretty early in vertebrates and maybe even before. But in other cases the systems are all messed up. Some parts of the cortex interact with the hypothalamus and are considered part of the limbic system. Some parts of the really primitive lizard brain handle sensation (like the colliculi). It looks like sensation/perception-related areas and emotion/motivation-related areas are mixed throughout every level of the brain. Most important, the frontal lobe, which we tend to interpret as the seat of truly human intelligence and executive planning and "the will" – probably evolved from sensation/perception-related areas in fish, since it looks like sensation/perception-related areas are just about all the cortex that fish had. And all of this evolved from the same couple of hundred neurons in worms, which were already responsible for interpreting the sensations picked up by the worm's little bristle thingies.

The point is, neither evolution nor anatomy suggests that the brain enforces a deep conceptual separation between perception, motivation, and cognition. Instead, the same sort of systems which handle perception in some areas are – with a few tweaks – able to handle cognition and motivation in others.

In fact, there are some deep connections between all three domains. The same factors that make a grey figure on dark ground look white can make an okay choice compared to worse choices look good. The same top-down processing that screws up PARIS IN THE THE SPRINGTIME is responsible for confirmation bias. In general the mapping between cognitive biases and perceptual illusions is fruitful enough that it's hard for me to believe that cognition and sensation/perception aren't handled in really similar ways, with motivation probably also involved.

So if we have something that can equal human sensory cortices – not just in the coincidental way where a sports car can equal a cheetah, but because we're genuinely doing the same thing human sensory cortices do for the same reasons – then we might already be further than we think towards understanding human intelligence and motivation.

III.

A quick sketch of two ways this might play out in real life.

First, categorization/classification/generalization/abstraction seems to be a big part of how people develop a moral sense, and maybe a big part of what morality *is*.

Everyone remembers the whole thing about mental categories, right? The thing where you have a category "bird", and you can't give a necessary-and-sufficient explicit definition of what you mean by that, but you know a sparrow is definitely a bird, and an ostrich is weird but probably still a bird, and there are edge cases like *Archaeopteryx* where you're not quite sure if they're birds or not and there's probably no fact of the matter either way? Cluster-structures in thingspace? Weird border disputes? That thing?

And you remember how we get these categories, right? A little bit of training data, your mother pointing at a sparrow and saying "bird", then maybe at a raven and saying "bird", then maybe learning ad hoc that a bat isn't a bird, and your brain's brilliant hyperadvanced categorization/classification/generalization/abstraction system picking it up from there? And then maybe after several thousand years of this Darwin comes along and tells you what birds *actually* are, and it's good to know, but you were doing just fine way before that?

We learn morality in a very similar way. When we hit someone, our mother/father/teacher/priest/rabbi/shaman says “That’s bad”; when we share, “that’s good”. From all this training data, the categorization/classification/generalization/abstraction system eventually feels like it has a pretty good idea of what morality is, although often we can’t verbalize an explicit definition any better than we can verbalize an explicit definition of “bird” (“it’s an animal that can fly…wait, no, bats…um, that has feathers…uh, do all birds have feathers? Bah, of course they don’t if you pluck them, that wasn’t what I meant…”). Just as Darwin was able to give an explicit definition of “bird” which conclusively settled some edge cases like bats, so philosophers have tried to give explicit definitions of “morality” which settle edge cases like abortion and trolley-related mishaps.

An AI based around a categorization/classification/generalization/abstraction system might learn morality in the same way. Its programmers give it a bunch of training data – maybe the Bible (this is a joke, please do not train an AI on the Bible) – and the AI gains a “moral sense” that it can use to classify novel data.

The classic Bostromian objection to this kind of scheme is that the AI might draw the wrong conclusion. For example, an AI might realize that things that make people happy are good – seemingly a high-level moral insight – but then forcibly inject everybody with heroin all the time so they could be as happy as possible.

To this I can only respond that we humans don’t work this way. I’m not sure why. It seems to either be a quirk of our categorization/classification/generalization/abstraction system, or a genuine moral/structure-of-thingspace-related truth about how well forced-heroin clusters with other things we consider good vs. bad. A fruitful topic for AI goal alignment research might be to understand exactly how this sort of thing works and whether there are certain values of classification-related parameters that will make classifiers more vs. less like humans on these kinds of cases.

Second, even if we can't get this 100% right, there might be a saving grace: I don't see these kinds of systems as paperclip maximizers. The human utility function seems to be a set of complicated things generalizing/abstracting from a few biologically programmed imperatives (food, sex, lack of pain) and ability to learn other goals from society and your moral system.

Categorization/classification/generalization/abstraction is certainly involved in reinforcement learning. You say "BARK!" and a dog barks, and you give it a treat. The dog needs to be able to figure out, on the fly, whether the treat was for barking when you said "BARK!", for barking whenever you speak, for barking in general, for being next to you, or just completely random. This is a problem of categorization and abstraction – going from training data ("the human did or didn't reward me at this specific time") to general principles ("when the human says bark, I bark").

I don't really understand how the human motivational system works. Dopamine and the idea of incentive salience seem to be involved in a fundamental way that seems linked to perception. But I am kind of hopeful that it's something that's not too hard to do if you already have a working categorizer, and that it's a foundation to build agents that want things without being psychopathic maniacs. Humans can want sex without being insane sex maximizers who copulate with everything around until they explode. An AI that wanted paperclips, but which was built on a human incentive system that gave paperclips the same kind of position as sex, might be a good paperclip producer without being insane enough to subordinate every other goal and moral rule to its paperclip-lust.

Tomorrow 10/31 is the last day of MIRI’s yearly fundraiser, and as usual I think it is a good cause well worth your donation. But its basic assumption is that AIs will be very computer-like: entities of pure code and logic that will reflect on themselves using mathematical tools. I can also imagine futures where AIs aren’t much more purely-logical than we are, and the tools we need to keep them human-friendly are very different. I support MIRI’s efforts to deal with the one case, but I’m hoping there will be some efforts in the other direction as well.

EDIT: Nick points out some of MIRI’s work along these lines.
EDIT2: Comment by Eliezer

SSC San Jose Meetup 11/4

What: SSC meetup
When: Friday 11/4, at 7 PM
Where: 3806 Williams Road, San Jose, CA
Why: Physicist/intelligence researcher Steve Hsu of Information Processing will be in the Bay Area and wants to get to meet some of you.
How: David Friedman is very nice and lets us use his house for things

I will unfortunately not be able to attend.

The Pyramid And The Garden

I.

A recent breakthrough in pseudoscience: the location of the Great Pyramid of Giza encodes the speed of light to seven decimal places.

This is actually true. The speed of light in a vacuum is 299,792,458 meters per second. The coordinates of of the Great Pyramid are 29.9792458° N, 31.1342880° E (you can confirm with Google Maps that this gets you right on top of the Pyramid). The speed of light and the latitude number there have all the same digits. That's a pretty impressive coincidence.

You might think this is idiotic because the meter was invented by 1600s French people. If ancient aliens or Atlanteans built the pyramids, why would they encode their secret wisdom using a unit of measurement from 1600s France? But there's a way around this objection: the 1600s French people defined their meter as 1/10,000,000th the distance between the Equator and the North Pole. If the aliens also thought that was an interesting way to measure length, then they could have encoded their secret wisdom in it. So you wouldn't need aliens who could predict the thoughts of 1600s Frenchmen. Just aliens who *thought exactly like* 1600s Frenchmen.

(actually, a different group of 1600s Frenchmen proposed a different version of the meter, defined as the length of a pendulum with a half-period of one second. This turned out to be 99.7% of the 1/10,000,000th-the-way-to-the-North-Pole definition, so either one works unless you want super-exactness. I think a much more interesting conspiracy theory would be that aliens designed the Earth to encode secret wisdom about the periods of pendulums.)

But realistically, aliens who think suspiciously like French people probably weren’t involved. So how do we explain the coincidence?

II.

The following is indebted to user mrfintoil’s great explanation on metabunk.org.

First, it’s *not* a coincidence to seven decimal places. Yes, that particular nine-digit sequence lands you atop the Great Pyramid. But that gives you way more precision than you need – cutting off the last three digits actually gets you closer rather than further from the center of the Pyramid. The only numbers that are doing any work are the 29.9792° N. So you really only get four decimal places worth of coincidence.

On the other hand, matching six digits is still pretty good. That’s literally a one-in-a-million chance.

So here the explanation has to go to how hard the pseudoscientists worked to find a coincidence of this magnitude; in other words, how many degrees of freedom they had.

Here’s an obvious example; as far as I can tell, the *longitude* of the Great Pyramid doesn’t encode anything interesting at all. So it’s not the equivalent of winning a one-in-a-million lottery with a single ticket. It’s the equivalent of winning a one-in-a-million lottery with *two* tickets.

A second issue: if the latitude of the Great Pyramid had been 10.7925 N, that would be the speed of light in kilometers per hour, which would be an equally impressive match.

So just taking these two degrees of freedom, we have four lottery tickets:

1. The one where the latitude is the speed of light in meters/second
2. The one where the longitude is the speed of light in meters/second
3. The one where the latitude is the speed of light in kilometers/hour
4. The one where the longitude is the speed of light in kilometers/hour

In other words, the number of lottery tickets increases exponentially as we get more degrees of freedom.

Let me list out all the degrees of freedom I can think of and see where we end up. I am going to try my best to be as fair as possible to the ancient aliens. For example, I was considering saying that since there are three pyramids at Giza, we have to multiply by three, but to be honest the Great Pyramid is clearly greater than the other two, and it would be less elegant if Menkaure’s pyramid encoded some amazing cosmic constant, so I won’t raise that objection. I am going to try to be *really fricking fair*.

1. Latitude vs. longitude (2 options)

2. Speed of light in meters/second vs. kilometers/hour vs. cubits/second vs. cubits/hour. I’m avoiding using feet/miles, because that’s even more arbitrary than meters. But I think it would actually be even *more* convincing if the calculation actually used the real Egyptian unit, which I understand is the cubit. So let’s go with (4 options)

3. Great Pyramid vs. Sphinx. Like I said before, the other two pyramids at Giza are noticeably less impressive than the Great Pyramid. But the Sphinx is pretty impressive, and the ancient aliens folks talk about it just as much as the Pyramid, so I think that would be an equally good hit if it had been true. (2 options)

4. Use of a 90 degree latitude system vs. use of a 100 degree latitude system. I'm a little split on this one, because it wouldn't look anywhere near as impressive if the pseudoscience sites had to explain that they found a really cool coincidence but it only worked if you converted normal latitude into a different hypothetical latitude system that had 100 degrees. But since we know the aliens/Atlanteans use base 10 anyway (they're encoding their wisdom in the base 10 representation of the speed of light) it makes more sense for them to use a base 10 latitude system instead of replicating our own bizarre custom of using base 10 for everything else but having latitude go from 0 to 90. On the other hand, if these were Earth-based Atlanteans, they might have gotten the custom of dividing the circle into 360 parts for the same reason we did – there are about 360 days in a year. And if they were aliens, maybe we got our bizarre latitude convention from *them* – the idea of 360 degree circles is really old and lost in the mists of time. Overall I can see this one going either way, so I'm going to give it as (2 options)

5. Decimal point placement. The latitude 29.9792 N matches the speed of light exactly, but so would the latitudes 2.99792, 2.99792 S, and 29.9792 S. I checked these other sites at the same longitude as the Pyramid to see if there were any mysterious features. But they seem to be, respectively, a perfectly ordinary field in Uganda, a perfectly ordinary field in Tanzania, and a perfectly ordinary patch of ocean. But a world where the pyramid was in

Uganda and the ordinary field was in Egypt would be just as much of a hit as our current world. Therefore (4 options)

From these really simple things alone, we learn we’ve got 2 x 4 x 2 x 2 x 4 = 128 lottery tickets, reducing our 1/1 million chance of winning to something more like 1/10,000. Progress!

There are a few other degrees of freedom that I think are a little harder to judge, but still important:

6. What aspect of the Pyramid we’re looking at. That is, it would have been equally interesting (maybe moreso!) if its height or width matched the speed of light exactly. So that’s another (3 options). I guess if the ancient aliens were *really* good at what they were doing, they could have given the pyramid 299,792,458 *sides*, but I won’t hold that against them. This should really make the multiplication more complicated because I can no longer use all the different ways of representing latitude vs. longitude, but I’ll stick with the simple method for now.

7. Which *site* we’re looking at. This one is hard, because I don’t know if anywhere else has the ancient alien-related credibility of the Great Pyramid. The only equally mysterious site I can think of is Stonehenge, and *maybe* the Nazca Lines. I don’t feel comfortable saying it would be *equally* impressive if Tiwanaku or Yonaguni had the right coordinates. I’ll just say (2 options) for Pyramids and Stonehenge.

8. Which *constant* we’re looking at. Sure, the Pyramid encoding the speed of light is pretty cool, but what about the Planck length? Avagadro’s number? I’m split on whether I want to include mathematical constants like pi or e in here. I think if it encoded pi to some number of decimals places

then I would just think that the Egyptians were more advanced at math than I thought but it wouldn't necessarily be earth-shattering. The Egyptians knowing e would be pretty shocking but still maybe not worth believing in ancient aliens over. There really aren't that many physical constants as cool as the speed of light, so I might just arbitrarily call this one (4 options).

So now we have a total of 128 x 3 x 2 x 4 = 3072 lottery tickets, for a 1/300 chance of winning the one-in-a-million lottery.

I would like to say "Ha ha, I sure proved those dumb conspiracy nuts wrong", except that a 1/300 chance is still a pretty impressive coincidence – what scientists call $p < 0.01$. And now I've used up all my excuses. I *think* what's going on here is that I'm still accepting the terms of the game – comparing only the exact categories used in the original calculation. Suppose that the latitude of the Great Pyramid was exactly 30.0000? That too would be impressive – it would prove that the pyramid builders knew the exact size and shape of the Earth and were able to build their Pyramid one third of the way between Equator and Pole. Suppose that the Great Pyramid was latitude 19.69724. That's the date humankind first landed on the moon in yyyy/mm/dd format – clearly the Pyramid was built by a time-traveling Nostradamus! Suppose that the Pyramid was built of stones of four different colors, with blue stones always paired opposite red stones, and yellow stones always paired opposite green stones. Then the ancient Egyptians were trying to tell us about the structure of DNA. What if the Pyramid, viewed from above, looked like a human brain?

Is it fair to take all of that into account? If so, does the remaining coincidence go away? I wish I were able to give these questions a more confident affirmative answer.

III.

I still believe that pseudoscience is helpful for understanding regular science. The loopholes that let people discover proofs of ESP or homeopathy are the same ones that let them discover proofs of power posing and ego depletion.

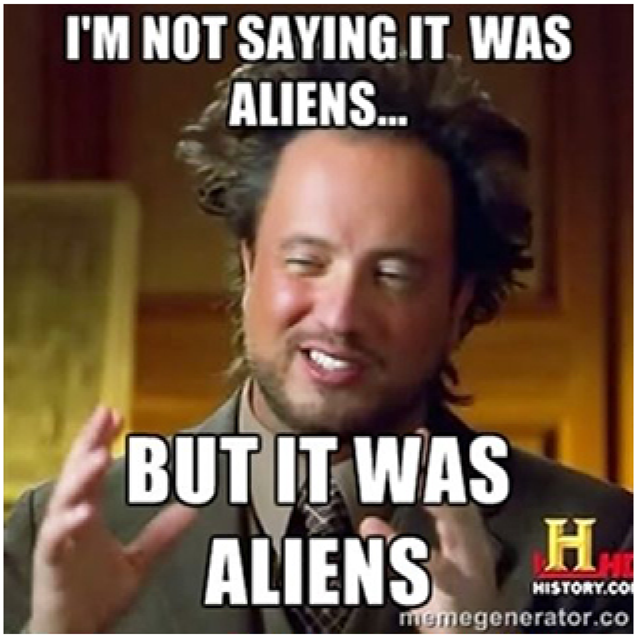
In the same way, numerology is helpful for understanding statistics. You can see the same factors at work, free from any lingering worry that maybe the theory you're investigating is true after all.

Andrew Gelman writes about the garden of forking paths. The idea is: the scientific community accepts a discovery as meaningful if $p < 0.05$ - that is, if equally extreme data would only occur by coincidence 5% of the time or less. In other words, you need to win a lottery with a one-in-twenty chance if you want to get credit for discovering something absent any real effect to be discovered. But if a scientist forms their hypothesis after seeing their data, they might massage the precise wording of their hypothesis to better fit their data. If there are many different ways to frame the hypothesis, then they have many lottery tickets to choose from and a win is no longer so surprising. Gelman discusses a study claiming to find that women wear red or pink shirts during the most fertile part of their menstrual cycle, which sometimes involves red or pink coloration changes in primates. The study does detect the effect, $p < 0.05$. But there were a couple of different ways the researchers could have framed the problem. They could have looked at only red shirts. They could have looked at only pink shirts. They chose days 7-14 as most fertile. But they could also have chosen days 6-15 without really being *wrong*. They could have looked only at the unmarried women most likely to be trying to attract mates. A recent paper listed 34 different degrees of freedom that can be used in this kind of thing.

Add up enough of them, and you have more than twenty tickets to the one-chance-in-twenty lottery and success is all but certain.

I used to call this the Elderly Hispanic Woman Effect, after drug studies where the drug has no effect in general, no effect on a subgroup of just men, no effect on a subgroup of just women, no effect on a subgroup of just blacks, no effect on a subgroup of just whites…but when you get to a subgroup of elderly Hispanic women, $p < 0.05$, apparently because it's synchronized with their unique biological needs. This is pretty obvious. The lesson of the Pyramid-lightspeed link is that sometimes it isn't. It just looks like some sudden and shocking coincidence. The other lesson of the Pyramid is that *I cannot consistently figure this kind of thing out*. I threw everything I had against the correlation, and I still ended up with $p = 0.003$. I don’t think this is because the Pyramid really *was* designed by aliens with a suspicious link to 1600s France. I think it’s because I’m not creative enough to fully dissect coincidences even when I’m looking for them.

This is always happening to me in real studies too. Something seems very suspicious. But their effect size is very high and their p-value is very significant. I can’t always figure out exactly what’s going on. But I should be reluctant to dismiss the possibility that I’m missing something and that there’s some reasonable explanation.


I'M NOT SAYING IT WAS ALIENS...
BUT IT WAS ALIENS
HD
HISTORY.CO
memegenerator.co

Tuesday Shouldn’t Change The Narrative

538 predicts Hillary has a 65% chance of winning the election to Trump’s 35%. New York Times says it’s more like 84% Hillary and 16% Trump. Both sites agree both candidates will get somewhere between 40% and 50% of the popular vote, and that Hillary seems to lead Trump by 3%. The smart money is on Hillary, but at this point either major candidate could win.

Lots of things can happen tomorrow. Maybe it rains in Philadelphia, that city’s racially diverse and left-leaning voters stay home, and Pennsylvania goes for Trump, winning him the election. Maybe there’s a really good get-out-the-vote campaign among Hispanics, and Florida ends up being Trump 48 Hillary 52 instead of the projected Trump 52 Hillary 48. Maybe the Department of Agriculture announces that Hillary is under investigation for bringing exotic weevil species into the US, and the population turns against her *en masse*.

And someone is going to confuse this kind of stuff with deep insight into the state of the country.

In June 2016, Jon Wiener of *The Nation* wrote Relax, Donald Trump Can’t Win, about how the media is incentivized to make races look competitive but an understanding of political fundamentals proved that there was no way for Trump to actually make it.

On the other hand, a few days ago Scott Adams reiterated his long-standing prediction of a 98% chance Trump wins in a landslide.

I’m worried that one of these two things will happen on Wednesday:

Either Hillary wins, and everybody agrees that Jon Wiener and various other people like him were right, that the fundamentals made a Trump win impossible, that Trump was a random clown who never had a chance anyway, that the people who warned us to beware of Trump were crying wolf, that this proves that nationalism is a spent force in politics, et cetera.

Or Trump wins, and everybody agrees that Scott Adams was a genius, that Wiener was an idiot, that Trump is a brilliant “master persuader”, that this proves that the 21st century will be a century of renewed nationalist power, that the white working class is sexist, that elites need to realize the precariousness of their position within a democratic system, or whatever.

Imagine that the deciding factor really *is* a rainstorm in Philadelphia. There was a rainstorm in Philly, therefore nationalism is one of the great motivating forces in human affairs? It was a clear sunny day in Philly, therefore nationalism doesn’t matter anymore? The difference between nationalism being all-powerful and irrelevant is whether there was a cold front over the mid-Atlantic region?

But with a race this close, *any* deciding factor is going to be about as random as a rainstorm over Philadelphia. Maybe the pollsters made some kind of big mistake and missed shy Trump voters, and the vote goes Trump 47% Hillary 45% instead of the predicted Hillary 47% Trump 45%. So what? The difference between a proof of nationalism’s vigor versus proof its impotence is which candidate gets 47% vs. 45%? *Really*?

If a Trump victory tomorrow would convince you that X is true, I suggest that you believe X is true regardless of whether or not Trump wins, because Trump's victory almost certainly will depend more on noise than on X. If a Hillary victory tomorrow would convince you that Y is true, I suggest that you believe Y is true regardless of whether or not Hillary wins, for the same reason. If there's some Z that you will believe only if Trump wins but not if Hillary wins, then I suggest you seriously reconsider what thought process has led you to decide that you will flip your views on politics and society depending on whether or not there's a rainstorm or a 2% polling error or whatever.

Instead, I suggest people precommit to their views on politics and society now. We live in a country and a world where Hillary can be at about 47% and Trump at about 45%. This is pretty much all you need to know. It suggests that a lot of people are willing to support a nationalist candidate, and a lot of other people really hate that candidate. It suggests that political fundamentals are totally compatible with a situation where either Trump or Hillary could win based on noise in the electoral process.

(unless the polls are totally wrong and one candidate somehow wins in a 20 percentage point landslide or something)

It also suggests that both Wiener and Adams were wrong to be so confident in their respective predictions. If either one is right, it will be mostly by luck. Wiener tells us to "relax" because Trump can never win, and maybe Trump doesn't win, but the fact is that even if Trump loses we were one Hillary gaffe away from the opposite result and shouldn't have relaxed at all. Adams says there's been a 98% chance of a Trump win since last year, but the polls make it look a lot like Trump only has a chance at all because of the total coincidence of Hillary getting hit by a new FBI investigation two weeks before Election Day.

I already count both Wiener and Adams as having been proven wrong regardless of what happens tomorrow. Any further praise or condemnation launched at one or the other after the election is just interpreting noise, or at least a signal so subtle that it might as well be.

If both Wiener's extreme pro-Hillary prediction and Adams' extreme pro-Trump prediction are bad, what would a good prediction look like? In January of this year, I predicted that, conditional on Trump winning the Republican primary, he would have a 20% chance of winning the election. Well, Trump won the Republican primary. And today, the day before the election, the prediction markets give Trump's chance of winning as 17.9%.

If Trump wins anyway, I'll have egg on my face and it'll look bad when I grade my prediction accuracy next year. But I don't think I would fundamentally update the way I think about America or the way I make predictions. No prediction can account for every rainstorm. I think I got the fundamentals right, and if I end up losing on noise I can at least take solace in knowing that is soon to be the *least* of our problems.

[EDIT: I agree the election results can obviously change the future! Maybe if Trump wins he'll enact nationalistic policy that will make people more nationalist. But the results shouldn't change our interpretation of existing trends.]

Book Review: House of God

I’m not a big fan of war movies. I liked the first few I watched. It was all downhill from there. They all seem so similar. The Part Where You Bond With Your Squadmates. The Part Where Your Gruff Sergeant Turns Out To Have A Heart After All. The Part Where Your Friend Dies But You Have To Keep Going Anyway. The Part That Consists Of A Stirring Speech.

The problem is that war is very different from everything else, but very much like itself.

Medical internship is also very different from everything else but very much like itself. I already had two examples of it: *Scrubs* and my own experience as a medical intern (I preferred *Scrubs*). So when *every single personin the medical field* told me to read Samuel Shem’s *House of God*, I deferred. I deferred throughout my own internship, I deferred for another two years of residency afterwards. And then for some reason I finally picked it up a couple of days ago.

This was a *heck* of a book.

On some level it was as predictable as I expected. It hit all of the Important Internship Tropes, like The Part Where Your Attendings Are Cruel, The Part Where Your Patient Dies Because Of Something You Did, The Part Where You Get Camaraderie With Other Interns, The Part Where You First Realize You Are Actually Slightly Competent At Like One Thing And It Is The Best Feeling In The Universe, The Part Where You Realize How Pointless 99% Of The Medical System Is, The Part Where You Have Sex With Hot Nurses, et cetera.

All I can say is that it was really well done. The whole thing had a touch of magical realism, which turns out to be exactly the right genre for a story about medicine. Real medicine is absolutely magical realist. It’s a series of bizarre occurrences just on the edge of plausibility happening to incredibly strange people for life-and-death stakes, day after day after day, all within the context of the weirdest and most byzantine bureaucracy known to humankind.

Just in the past week, for example, I had to deal with an aboulomaniac patient – one with a pathological inability to make up his mind. He came to my clinic for treatment, but as soon as he saw me, he decided he didn’t want treatment after all and left. The next day, he was back on my calendar – he’d decided he needed treatment after all – but when his appointment came around, he chanegd his mind and left again. This happened *five times in five days*. Every day he would phone in asking for an appointment. Every day I would give it to him. Every day he would leave a minute or two before it began. Unsure how to proceed, I sought out my attending. He ignored my questions, pulled me into a side office, took out his cell phone, and started playing me a video. It’s a scene from his musical, *The Phantom Of The Psychiatric Unit*, which he’s been forcing his interns to rehearse after rounds. I watched, horrified. It was weirdly good.

If I were to write a book about this kind of thing, people would criticize me for being unrealistic. The only way to get away with it is to pass it off as “a touch of magical realism”, and this *The House of God* does to excellent effect.

The story revolves around an obvious author-insert character, Roy Basch MD, who starts his internship year at a hospital called the House of God (apparently a fictionalized version of Beth Israel Hospital in Boston). He goes in with expectations to provide useful medical care to people with serious diseases. Instead, he finds gomers:

> “Gomer is an acronym: Get Out of My Emergency Room. It’s what you want to say when one’s sent in from the nursing home at three A.M.”
>
> “I think that’s kind of crass,” said Potts. “Some of us don’t feel that way about old people.”
>
> “You think I don’t have a grandmother?” asked Fats indignantly. “I do, and she’s the cutest dearest, most wonderful old lady. Her matzoh balls float – you have to pin them down to eat them up. Under their force the soup levitates. We eat on ladders, scraping the food off the ceiling. I love…” The Fat Man had to stop, and dabbed the tears from his eyes, and then went on in a soft voice, “I love her very much.”
>
> I thought of my grandfather. I loved him too.
>
> “But gomers are not just dear old people,” said Fats. “Gomers are human beings who have lost what goes into being human beings. They want to die, and we will not let them. We’re cruel to the gomers, by saving them, and they’re cruel to us, by fighting tooth and nail against our trying to save them. They hurt us, we hurt them.”

This is where the magical realism starts to come in:

> Rokitansky was an old bassett. He’d been a college professor and had suffered a severe stroke. He lay on his bed, strapped down, IV’s going in, catheter coming out. Motionless, paralyzed, eyes closed, breathing comfortably, perhaps dreaming of a bone, or a boy, or of a boy throwing a bone.

“Mr. Rokitansky, how are you doing?” I asked.

Without opening his eyes, after fifteen seconds, in a husky slurred growl from deep down in his smushed brain he said: PURRTY GUD.

Pleased, I asked, “Mr. Rokitansky, what date is it today?”

PURRTY GUD. .

To all my questions, his answer was always the same. I felt sad. A professor, now a vegetable. Again I thought of my grandfather, and got a lump in my throat. Turning to Fats, I said, “This is too sad. He’s going to die.”

“No, he’s not,” said Fats. “He wants to, but he won’t.”

“He can’t go on like this.”

“Sure he can. Listen, Basch, there are a number of LAWS OF THE HOUSE OF GOD. LAW NUMBER ONE: GOMERS DON’T DIE.”

“That’s ridiculous. Of course they die.”

“I’ve never seen it, in a whole year here,” said Fats.

“They have to.”

“They don’t. They go on and on. Young people – like you and me – die, but not the gomers. Never seen it. Not once.”

“Why not?”

“I don’t know. Nobody knows. It’s amazing. Maybe they get past it. It’s pitiful. The worst.”

> Potts came in, looking puzzled and concerned. He wanted the Fat Man's help with Ina Goober. They left, and I turned back to Rokitansky. In the dim half-light I thought I saw tears trickling down the old man's cheeks. Shame swept over me. My stomach churned. Had he heard what we'd said?
>
> "Mr. Rokitansky, are you crying?" I asked, and I waited, as the long seconds ticked away, my guilt moaning inside me.
>
> PURRTY GUD.
>
> "But did you hear what we said about gomers?"
>
> PURRTY GUD.

Someone once said that the point of art is to be more real than reality. *The House Of God* is *way* more real than reality. Reality *wishes* it could be anywhere *close to* as real as *The House of God*. This is a world where young people – the kid just out of school, the blushing new mother – die. Even normal old people – your grandmother, your grandpa – can die. But the most decrepit, demented people, the ones for whom every moment of artificially-prolonged life is a gratuitous misery and you pray at every moment that God will just let them find some peace – somehow they never die. They come into the hospital, they go back out to nursing homes, a few weeks later they're back in the hospital, a few weeks later they're back in their nursing homes, but *they never die*. This *can't* be literally true. But it's the subjective truth of working in a hospital. The Fat Man is right. I've been working in medicine for three years now, and I have seen my share of young people tragically cut off in the prime of life, and yet as far as I can remember I have never seen a gomer die. The magical realism of *House of*

God describes the reality of medical professionals infinitely better than the rational world of hospital mortality statistics.

In the world of *The House of God*, the primary form of medical treatment is the TURF – the excuse to get a patient out of your care and on to somebody else's. If the psychiatrist can't stand a certain patient any longer, she finds some trivial abnormality in their bloodwork and TURFs to the medical floor. But she knows that if the medical doctor doesn't want one of *his* patients, then he can interpret a trivial patient comment like "Being sick is so depressing" as suicidal ideation and TURF to psychiatry. At 3 AM on a Friday night, every patient is terrible, the urge to TURF is overwhelming, and a hospital starts to seem like a giant wheel uncoupled from the rest of the world, Psychiatry TURFING to Medicine TURFING to Surgery TURFING to Neurosurgery TURFING to Neurology TURFING back to Psychiatry again. Surely some treatment must get done somewhere? But where? It becomes a legend, The Place Where Treatment Happens, hidden in some far-off hospital wing accessible only to the pure-hearted. This sort of Kafkaesque picture is how medical care *feels*, and the genius of *The House of God* is that it accentuates the reality just a little bit until its fictional world is almost as magical-realist as the real one.

In the world of *The House of God*, medical intervention can only make patients worse:

> Anna O. had started out on Jo's service in perfect electrolyte balance, with each organ system working as perfectly as an 1878 model could. This, to my mind, included the brain, for wasn't dementia a fail-safe and soothing oblivion of the machine to its own decay?

From being on the verge of a TURF back to the Hebrew House for the Incurables, as Anna knocked around the House of God in the steaming weeks of August, getting a skull film here and an LP there, she got worse, much worse. Given the stress of the dementia work-up, every organ system crumpled: in a domino progression the injection of radioactive dye for her brain scan shut down her kidneys, and the dye study of her kidneys overloaded her heart, and the medication for her heart made her vomit, which altered her electrolyte balance in a life-threatening way, which increased her dementia and shut down her bowel, which made her eligible for the bowel run, the cleanout for which dehydrated her and really shut down her tormented kidneys, which led to infection, the need for dialysis, and big-time complications of these big-time diseases. She and I both became exhausted, and she became very sick. Like the Yellow Man, she went through a phase of convulsing like a hooked tuna, and then went through a phase that was even more awesome, lying in bed deathly still, perhaps dying. I felt sad, for by this time, I liked her. I didn't know what to do. I began to spend a good deal of time sitting with Anna, thinking.

The Fat Man was on call with me every third night as backup resident, and one night, searching for me to go to the ten o'clock meal, he found me with Anna, watching her trying to die.

“What the hell are you doing?” he asked.

I told him.

“Anna was on her way back to the Hebrew House, what happened – wait, don't tell me. Jo decided to go all-out on her dementia, right?”

> "Right. She looks like she's going to die."
>
> "The only way she'll die is if you murder her by doing what Jo says."
>
> "Yeah, but how can I do otherwise, with Jo breathing down my neck?"
>
> "Easy. Do nothing with Anna, and hide it from Jo."
>
> "Hide it from Jo?"
>
> "Sure. Continue the work-up in purely imaginary terms, buff the chart with the imaginary results of the imaginary tests, Anna will recover to her demented state, the work-up will show no treatable cause for it, and everybody's happy. Nothing to it."
>
> "I'm not sure it's ethical."
>
> "Is it ethical to murder this sweet gomere with your work-up?"
>
> There was nothing I could say."

After learning these medical secrets, Dr. Basch uses hook and crook to prevent his patients from getting any treatment. They end up healthier than anyone else in the hospital, and Basch becomes a contender for "Most Valuable Intern" – in typical *House of God* style, nobody knows if this award really exists or is just a rumor. His colleagues compete for another award, the "Black Crow", which goes to the intern who gets the most autopsy consents from grieving families – and which the administration doesn't realize incentivizes doctors to kill their patients. This is so reminiscent of the bizarre incentive systems in real hospitals that it *hurts*.

But as the year goes on, everyone gets more and more frazzled. One intern has a mental breakdown. Another commits suicide by jumping out of a hospital window (this isn’t dramatic exaggeration by the way; three junior doctors have committed suicide by jumping out of windows in the past three years in New York City alone). Dr. Basch runs through all sorts of interesting forms of neurosis. Finally, the end of the year approaches, the original crop of interns thinned-out but triumphant – and then they realize they have to do the whole thing again next year as residents, which is maybe a little less grueling but still in the same ballpark.

So they decide, en masse, to go into psychiatry, well-known to be a rare non-terrible residency. The author of *House of God* is a psychiatrist, so I guess this is only a spoiler insofar as you aren’t logically omniscient. When the Chief of Medicine learns that every single one of his hospital’s interns are going into psychiatry and there aren’t going to be any non-psychiatry residents in the whole hospital…

…okay, fine, I won’t spoil the ending. But suffice it to say I’m feeling pretty good about my career path right now.

II.

House of God does a weird form of figure-ground inversion.

An example of what I mean, taken from politics: some people think of government as another name for the things we do together, like providing food to the hungry, or ensuring that old people have the health care they need. These people know that some politicians are corrupt, and sometimes the money actually goes to whoever’s best at demanding pork, and the regulations sometimes favor whichever giant corporation has the best lobbyists. But this is viewed as a weird disease of the body politic, something that can be abstracted away as noise in the system.

And then there are other people who think of government as a giant pork-distribution system, where *obviously* representatives and bureaucrats, incentivized in every way to support the forces that provide them with campaign funding and personal prestige, will take those incentives. *Obviously* they'll use the government to crush their enemies. Sometimes this system also involves the hungry getting food and the elderly getting medical care, as an epiphenomenon of its pork-distribution role, but this isn't particularly important and can be abstracted away as noise.

I think I can go back and forth between these two models when I need to, but it's a weird switch of perspective, where the parts you view as noise in one model resolve into the essence of the other and vice versa.

And *House of God* does this to medicine.

Doctors use certain assumptions, like:

1. The patient wants to get better, but there are scientific limits that usually make this impossible
2. Medical treatment makes people healthier
3. Treatment is determined by medical need and expertise

But in *House of God*, the assumptions get inverted:

1. The patient wants to just die peacefully, but there are bureaucratic limits that usually make this impossible
2. Medical treatment makes people sicker
3. Treatment is determined by what will make doctors look good without having to do much work

Everybody knows that those first three assumptions aren't always true. Yes, sometimes we prolong life in contravention of patients' wishes. Sometimes people mistakenly receive unnecessary treatment that causes complications. And sometimes care suffers because of doctors' scheduling issues. But it's easy to abstract away to an ideal medicine based on benevolence and reason, and then view everything else as rare and unfortunate deviations from the norm.

House of God goes the whole way and does a full figure-ground inversion. The outliers become the norm; good care becomes the rare deviation. What's horrifying is how convincing it is. Real medicine looks *at least* as much like the bizarro-world of *House of God* as it does the world of the popular imagination where doctors are always wise, diagnoses always correct, and patients always grateful.

There have been a couple of studies finding that giving people health insurance doesn't make them any healthier – see for example the RAND Health Insurance Experiment and the Oregon Medicaid Experiment. I've always been skeptical of these studies, because it seems logical that people who can afford health care will get more of it, and there are ten zillion studies showing various forms of health care to help. Insulin helps diabetes. Antibiotics help sepsis. Surgery helps appendicitis. To deny claims like these would be madness, yet the studies don't lie. What is going on?

And the answer has to be somewhere in the bizarro-world of *House of God*. Real medical treatment looks precious little like the House MD model of rare serious disease -} diagnosis -} cure. At least as often, it's like the *House of God* model where someone becomes inconvenient -} send to hospital -} one million unnecessary tests. Everyone agrees this is part of the story. *House of God* is a brilliant book in that it refactors perception to place it in the foreground.

But it's brilliant because in the end it's *not* just a romp through hilarious bureaucratic mishaps. There is as much genuine human goodness and compassion in this book as there is in any rousing speech by a medical school dean. The goodness is often mixed with horror – the doctor who has to fight off hordes of autopsy-consent-form-seekers to let a dying patient spend his last few seconds in peace, or the one who secretly slips euthanasia to a terminal patient begging for an end to the pain because he knows it's the right thing to do.

The question posed here is "what do you do in a crazy cannibalistic system where it's impossible to do good work and everyone is dying all around you?", and the answer is "try as hard as you can to preserve whatever virtue you can, and to remain compassionate and human". The protagonist swings wildly between "this is all bullshit and I'll just make fun of these disgusting old people and call it a day" and "I need to save everybody and if I don't I should hate myself forever", and eventually like everybody, comes to some kind of synthesis where he recognizes he's human, recognizes that his patients are human, and tries to deal with it with whatever humor and grace he can manage.

It's hard enough for a book to be funny, and it's hard enough for one to be deep, but a book like *House of God* that can be both at once within the space of a few sentences is an absolute treasure.

III.

I talked to my father about *House of God*, and I told him a few parts that seemed unrealistic. He told me that those parts were 100% true in 1978 when the book was written. I looked into it more, and ended up appreciating the work on a whole new level.

Uncle Tom's Cabin is credited with kickstarting the emancipationist movement and maybe even causing the Civil War. *The Jungle* is famous for launching a whole new era of safety regulations. *House of God* has a place beside them in the pantheon of books that have changed the world.

The book's "Second Law" is "GOMER GOES TO GROUND": demented old people will inevitably fall out of their hospital bed and injure themselves. The book has a whole funny/horrifying scene where the senior resident explains his strategy for this eventuality: He leaves their beds low enough that patients won't kill themselves when they fall, but high enough that they'll probably break a bone or two and have to go to orthopaedic surgery – which takes them off his hands. Later, a medical student apes this procedure, a patient falls and breaks a bone or two, and everyone freaks out and tells him that it was a joke, that *of course* you don't really arrange skeletal fractures for old people just to save yourself time, what kind of heartless moron could *think* such a thing? This is some nth-level meta-humor: the reader probably mistook it for real advice because it meshes so seamlessly with all of the other madness and horror, yet most of the other madness and horror in the book is easily recognizable by practicing doctors as a real part of the medical system. Actually, on the n+1st meta-level, I'm not at all sure that the resident wasn't meant to be completely serious and then backtracked and called it a joke when it went wrong. For that matter, I'm far from sure this wasn't a real medical practice in the 1970s.

I see enough falls that I wasn’t surprised to see them as a theme, but I thought the book exaggerated their omnipresence. My father said it didn’t – there were just far more falls back in the Old Days. Now hospitals are safer and falls are comparatively rare. Why? Because the government passed a law saying that insurance wouldn’t pay hospitals extra money for the extra days patients have to stay due to fall-related injuries. I am so serious about this. This, I think, is the n+2nd meta-level; amidst all its jokes-played-straight the book treats encouraging falls as an actual in-universe joke, and yet in the real world once hospitals were no longer incentivized to let patients fall the falls stopped.

How did people become aware of this kind of thing? How did the movement against it start? A lot of it seems to be because of *House of God*. Everyone in medicine knew about this sort of thing. But *House of God* made it common knowledge.

People were scared to speak up. Everyone thought that maybe they were just a uniquely bad person, or their hospital a uniquely bad institution. Anyone who raised some of these points was met with scorn by prestigious doctors who said that maybe they just weren’t cut out of medicine. *House of God* shaped medicine because it was the first thing to say what everybody was experiencing. Its terms like “gomer” and “turf” made it into the medical lexicon because they pointed to obvious features of reality nobody had the guts to talk about before.

Shem writes an afterword where he talks about the reaction to the book. Junior doctors and the public loved it. Senior doctors hated it. He tells the story of going to a medical conference. Someone asked who he was, and he said jokingly “I’m the most hated doctor here”. His interlocutor answered “Oh, don’t worry, I’m sure you’re not as bad as the guy who wrote that *House of God* book.”

But *House of God* gets credit for helping start movements to cut intern work hours, protect doctors from sleep deprivation, reduce patient falls, and teach empathy and communication skills. The moral of the story is: the courage to tell the truth is rare and powerful. More specifically: the courage to tell the truth is rare and powerful not just in Stalinist dictatorships and violent cults, but in apparently normal parts of everyday First World life. All of these differently loaded terms like “culture of silence” and “political correctness” point at a fear of rocking various boats with nothing but your imperfect first-person knowledge to go on. But a tiny crack in the wall can make a big difference.

IV.

In a closing scene, Dr. Basch and all of his fellow interns – interns who had broken into tears weekly, gotten burnt out, starting seeing psychiatrists, considered suicide, all this stuff, these interns who had smashed up against the unendurable horrors of medicine and held themselves together only by the promise that it would soon be over – the minute they graduate internship they change their tune:

> It looked like all but two or three [interns] would stay. The Runt and I were definitely leaving; Chuck hadn’t yet said. The others were staying. In years to come they would spread out across America into academic centers and Fellowships, real red-hots in internal medicine, for they had been trained at the Best Medical School’s best House, the House of God. Although a few might kill themselves or get addicted or go crazy, by and large they’d repress and conform and perpetuate the Leggo

> [the Chief of Medicine] and the House and all the best medical stuff. [Eddie] had been praised by the Leggo that he could start off the second year as ward resident, with “a free rein” on his interns. And so, saying already that the internship been “not so bad,” he was preparing to indoctrinate his new charges: “I want them on their knees from day one.”

Shem’s author mouthpiece character Berry says:

> It’s been inhuman. No wonder doctors are so distant in the face of the most poignant human dramas. The tragedy isn’t the crassness, but the lack of depth. Most people have some human reaction to their daily work, but doctors don’t. It’s an incredible paradox that being a doctor is so degrading and yet is so valued by society. In any community, the most respected group are doctors. [It’s] a terrific repression that makes doctors really believe that they are omnipotent healers. If you hear yourselves saying, ‘Well, this year wasn’t really that bad,’ you’re repressing, to put the next group through it. [But] it’s hard to say no. If you’re programmed from age six to be a doctor, invest years in it, develop your repressive skills so that you can’t even recall how miserable you were during internship, you can’t stop.

Shem’s thesis is that it isn’t just about not wanting to make waves or offend the Chief of Medicine. It’s about denying your own pain by identifying with the system.

This puts me in a weird spot. My internship (I find myself saying) wasn’t so bad. I can give you some arguments why this might be true – things have gotten a lot better since *The House of God* was published (with no small credit to Shem himself), a small community hospital in Michigan is less intense than Harvard Medical School’s training hospital, psychiatry interns sometimes have it easier than internal medicine interns since everyone knows this isn't a permanent deal for them.

And yet I distinctly remember one night a long time ago, coming home from high school. I had noticed that all of the adults around me said high school was some of the best years of their lives and I would miss it when I was gone, and yet high school seemed objectively terrible. I wondered if there might be some bias or bizarre shift in memory that happened sometime in people's twenties and gave them a localized amnesia or insanity. So I very distinctly recall telling myself "My current assessment is that high school is terrible, and if you ever find yourself remembering that high school was lovely, please be aware that your memories have been hijacked by some malevolent force."

And God help me, but *every single part of my brain is telling me that high school was lovely*. I fondly remember all the friends I made, the crazy teachers I had to put up with, the science competitions I won, the lunches spent in the library reading whatever random stuff I could get my hands on. It seems like it was a blast. It's hard for me to even trust that one memory as anything more than imagination or the product of a single bad day. But although high-school-me had a lot of issues, he generally had a decent head on his shoulders, and if he says my memories have been hijacked, then I grudgingly believe him.

So was my intern year a good learning experience? *I have no idea and I'm not sure anyone else does either*. It's another type of figure-ground inversion: parade of horrors broken only by the occasional triumph, or clear sailing with a few bad moments?

On my last day of internship, one of my colleagues who was moving on said “I’m going to miss hating this place”. I’ve always remembered that phrase. Now I wonder if it’s some kind of weird snapshot of the exact moment of transition, the instant when “nightmarish ordeal” morphs into “halcyon days of youth”. This is why medicine has to be written as magical realism. How else to capture a world where people reliably go from agony to Stockholm Syndrome in the space of a day, and where the transition is so intermixed with the general weirdness that it doesn’t even merit special remark?

I found myself having more emotions reading *House of God* than I’ve had about anything in a long time. I don’t really know why. But I think it has something to do with this resignation to the general incommunicable weirdness all around anyone who works in medicine. Somehow Shem manages to avoid the normalization of insanity that happens to every young doctor, capture the exact subjective experience and write it down in a way that makes sense. And then, having put his finger right on the unbearable thing, he makes it funny and beautiful and poignant.

> I tell her. Again I tell her about Dr. Sanders bleeding out in my lap, about the look in Potts’s eyes that night before he jumped, about my pushing the KCl into poor Saul. I tell her how ashamed I am for turning into a sarcastic bastard who calls the old ones gomers, how, during the ternship, I’d ridiculed them for their weaknesses, for throwing up their suffering in my face, for scaring me, for forcing me to do disgusting things to take care of them. I tell her how I want to live, compassionately, with the idea of death clearly in sight, and how I doubt I can do that, ever again. As I think back to what I’d gone through and what I’d become, sadness wells up and mixes with contempt. I put my

> head into Berry's folds and weep, and curse, and shout, and weep.
>
> ". . . and in your own way, you did. Someone had to care for the gomers; and this year, in your own way, you did."
>
> "The worst thing is this bitterness. I used to be different, gentle, even generous, didn't I? I wasn't always like this, was I?"
>
> "I love who you are. To me, underneath it all, you're still there:" She paused, and then, eyes sparkling, said, "And you might even be better."
>
> "What? What do you mean?"
>
> "This might have been the only thing that could have awakened you. Your whole life has been a growing from the outside, mastering the challenges that others have set for you. Now, finally, you might just be growing from inside yourself.

He also frames all of it in the language of psychoanalysis, which is jarring and sounds preachy. I've ordered the sequel, Mount Misery, about his training as a psychoanalyst. Expect a review of that soon.

You Are Still Crying Wolf

Go away, pro-Trump bots

[Update 5/30: I originally took this post down because it got picked up by some pro-Trump accounts that I think were bots which would post it several times a day to various places. I didn't want to be involved in that or seeming to support it. After that I got in more trouble because I was supposedly "censoring" or "hiding" this. In order to avoid the Streisand Effect, I think my best option is to leave it up with this disclaimer at the top and the rest in a hard-to-read font.

Overall I think the predictions here have been borne out. Trump has proven himself an awful president and an offensive and incontinent person, but has continued to deny overt racism without any particular evidence that his denials are false (even as huge quantities of evidence of all Trump's other misdeeds have piled up). Despite inflammatory rhetoric and unprecedented incompetence, he has continued to pursue policies basically within the Republican mainstream.

Many people have tried to tell me this post "didn't age well", but I think they're just continuing to make the same mistakes this post warns against. For example, this post warns against thinking of Trump's opposition to immigration as some sort of shocking new development when it's similar to other immigration opposition from other presidents of both parties. I predicted that Trump would not deport more people than Obama. Given all of the evidence of ICE atrocities, I've gotten a lot of people writing me to complain about how wrong I obviously was, but this is on track to be correct – Obama deported more people in 2010 than Trump did in 2017, and ICE was pretty atrocious during his administration too. If it doesn't seem that way, that's because *my original thesis is still correct* – Trump is acting within the (worst parts of the) US mainstream, but the media is spinning it as an unprecedented atrocity.

The other event that led to me getting lots of hate mail was the Charlottesville protests. Again, the fact that a pan-alt-right gathering attracted a few hundred people matches my claim that there are probably only a few thousand white supremacists in the US (and is consistent with my claim that they are less influential than Satanists, given that thousands of people have shown up to some Satanist masses). There have been white supremacist rallies (including white supremacist rallies with violence against counterprotesters) during the Bush administration and the Obama administration. The only thing that's new in the Age of Trump is that the media is paying much more attention and people are considering them worthy of notice, or symbolic of some larger trend.

Overall I continue to believe Trump is a terrible president, but also not "an overt white supremacist" by any reasonable definition. I think the Democrats have done a good job focusing on Trump's other bad qualities and have mostly stopped centering their criticism around bumbling attempts to link him to the KKK or to neo-Nazis – and I no longer am particularly upset with how they're handling this matter. That makes this post of historical interest only.]

[Content warning: hate crimes, Trump, racism. I have turned off comments to keep out bad people who might be attracted by this sort of thing. Avoid sharing in places where this will attract the wrong kind of attention, as per your best judgment. Please don't interpret anything in this article to mean that Trump is not super terrible]

[Epistemic status: A reduction of a complicated issue to only 8000 words, because nobody would read it if it were longer. I think this is true but incomplete. I do not deny that Trump is being divisive and abusing identity politics in more subtle ways. I will try to discuss missing parts at more length later.]

I.

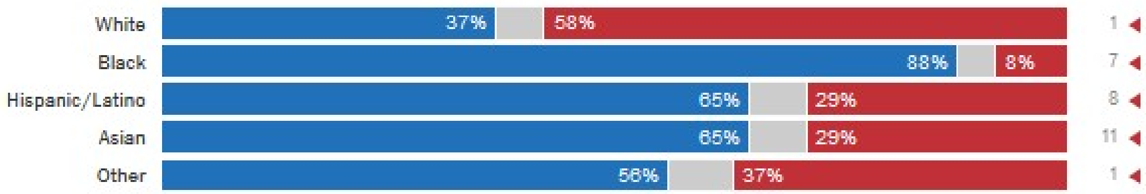
A New York Times article from last September that went viral only recently: [Crying Wolf, Then Confronting Trump](). It asks whether Democrats have "cried wolf" so many times that nobody believes them anymore. And so:

> *When "honorable and decent men" like McCain and Romney "are reflexively dubbed racists simply for opposing Democratic policies, the result is a G.O.P. electorate that doesn't listen to admonitions when the genuine article is in their midst".*

I have a different perspective. Back in October 2015, [I wrote]() that the picture of Trump as "the white power candidate" and "the first openly white supremacist candidate to have a shot at the Presidency in the modern era" was overblown. I said that "the media narrative that Trump is doing some kind of special appeal-to-white-voters voodoo is unsupported by any polling data", and predicted that:

> *If Trump were the Republican nominee, he could probably count on equal or greater support from minorities as Romney or McCain before him.*

Now the votes are in, and Trump got greater support from minorities than Romney or McCain before him. You can read the Washington Post article, [Trump Got More Votes From People Of Color Than Romney Did](), or look at the raw data ([source]())



Trump made gains among blacks. He made gains among Latinos. He made gains among Asians. The only major racial group where he didn't get a gain of greater than 5% was white people. I want to repeat that: the group where Trump's message resonated least over what we would predict from a generic Republican was the white population.

Nor was there some surge in white turnout. I don't think we have official numbers yet, but by eyeballing what data we have it looks very much like whites turned out in equal or lesser numbers this year than in 2012, 2008, and so on. [EDIT: see [counterpoint](), [countercounterpoint]()]

The media responded to all of this freely available data with articles like [White Flight From Reality: Inside The Racist Panic That Fueled Donald Trump's Victory]() and [Make No Mistake: Donald Trump's Win Represents A Racist "Whitelash"]().

I stick to my thesis from October 2015. There is no evidence that Donald Trump is more racist than any past Republican candidate (or any other 70 year old white *guy, for that matter). All this stuff about how he's "the candidate of the KKK" and "the vanguard of a new white supremacist movement" is made up. It's a catastrophic distraction from the dozens of other undeniable problems with Trump that could have convinced voters to abandon him. That it came to dominate the election cycle should be considered a horrifying indictment of our political discourse, in the same way that it would be a horrifying indictment of our political discourse if the entire Republican campaign had been based around the theory that Hillary Clinton was a secret Satanist. Yes, calling Romney a racist was crying wolf. But you are still crying wolf.*

I avoided pushing this point any more since last October because I didn't want to look like I was supporting Trump, or accidentally convince anyone else to support Trump. I think Trump's election is a disaster. He has no plan, he's dangerously trigger-happy, and his unilateralism threatens aid to developing countries, one of the most effective ways we currently help other people. I thought and still think a Trump presidency will be a disaster.

But since we're past the point where we can prevent it, I want to present my case.

I realize that all of this is going to make me sound like a crazy person and put me completely at odds with every respectable thinker in the media, but luckily, being a crazy person at odds with every respectable thinker in the media has been a pretty good ticket to predictive accuracy lately, so whatever.

II.

First, I want to go over Donald Trump's official, explicit campaign message. Yes, it's possible for candidates' secret feelings to differ from their explicit messages, but the things they say every single day and put on their website and include in their speeches are still worth going over to see what image they want to project.

Trump's official message has been the same vague feel-good pro-diversity rhetoric as any other politician. Here's Trump on African Americans:

> *It is my highest and greatest hope that the Republican Party can be the home in the future and forevermore for African-Americans and the African-American vote because I will produce, and I will get others to produce, and we know for a fact it doesn't work with the Democrats and it certainly doesn't work with Hillary.*
>
> *When I am President, I will work to ensure that all of our kids are treated equally, and protected equally. Every action I take, I will ask myself: does this make life better for young Americans in Baltimore, Chicago, Detroit, Ferguson who have as much of a right to live out their dreams as any other child in America?*
>
> *African-American citizens have sacrificed so much for this nation. They have fought and died in every war since the Revolution, and from the pews and*

> *the picket lines they have lifted up the conscience of our country in the long march for Civil Rights. Yet, too many African-Americans have been left behind.*
>
> *No group in America has been more harmed by Hillary Clinton's policies than African-Americans. No group. No group. If Hillary Clinton's goal was to inflict pain on the African-American community, she could not have done a better job. It's a disgrace. Tonight, I am asking for the vote of every African-American citizen in this country who wants a better future.*
>
> *And at the end of four years I guarantee that I will get over 95% of the African-American vote. I promise you. Because I will produce for the inner-cities and I will produce for the African-Americans.*
>
> *America must reject the bigotry of Hillary Clinton who sees communities of color only as votes, not as human beings worthy of a better future.*

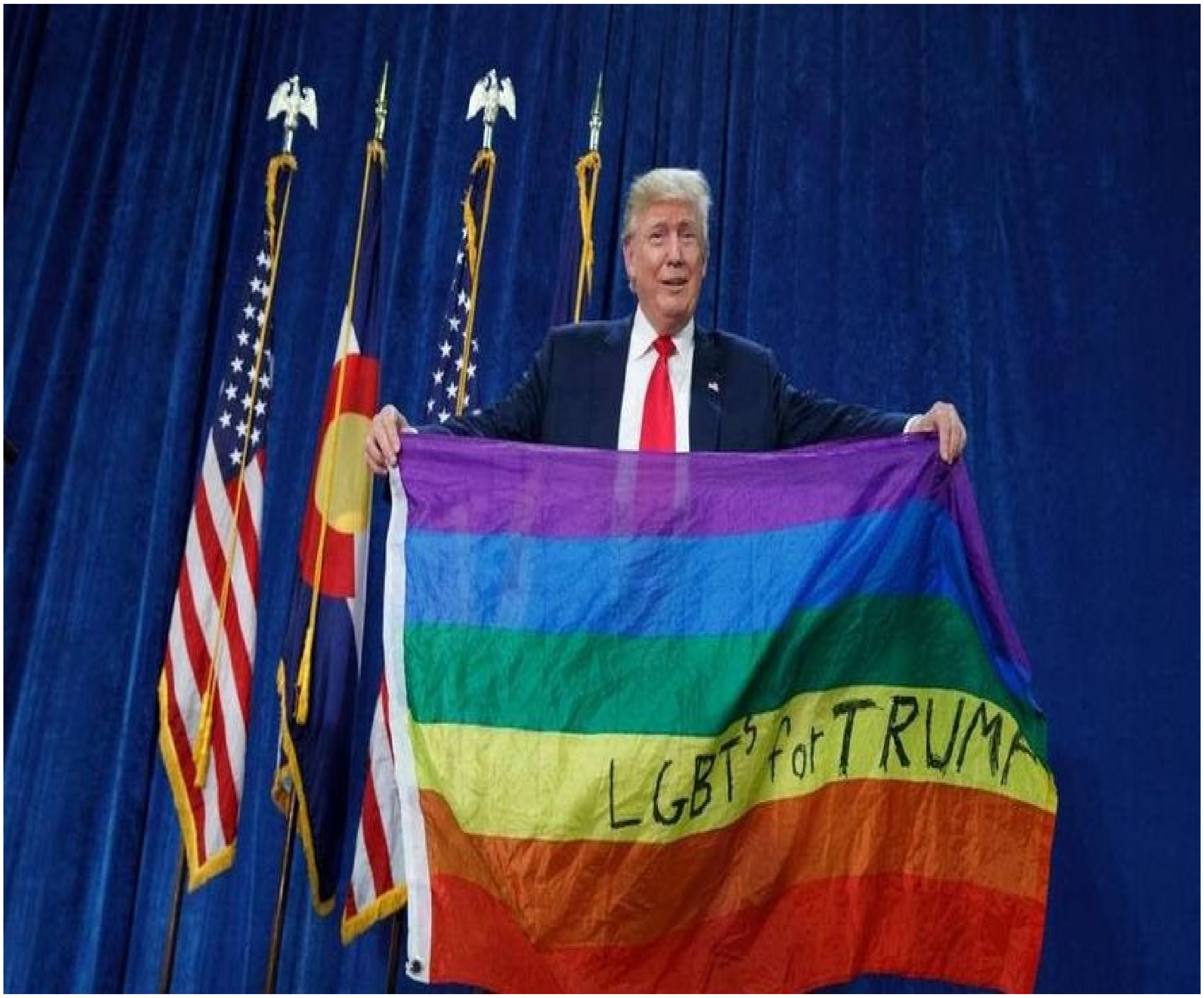
On Hispanics:

> *I have just landed having returned from a very important and special meeting with the President of Mexico…we discussed the great contributions of Mexican-American citizens to our two countries, my love for the people of Mexico, and the close friendship between our two nations.*
>
> *I employ thousands and thousands of Hispanics. I love the people. They're great workers. They're fantastic people and they want legal immigration. I'll take jobs back from China, I'll take jobs back from Japan. The Hispanics are going to get those jobs, and they're going to love Trump.*

On his campaign:

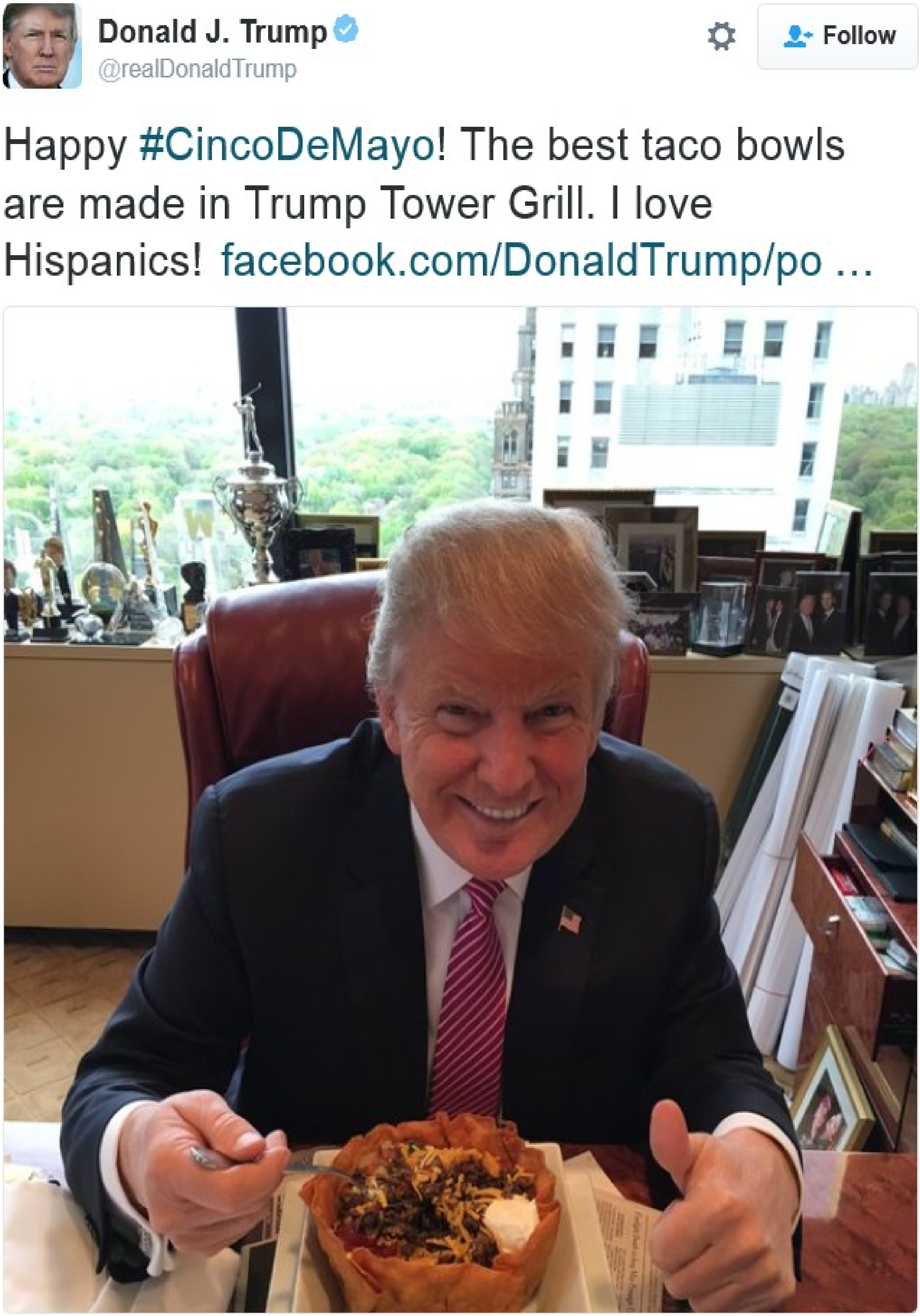
> *It's a movement comprised of Americans from all races, religions, backgrounds and beliefs who want and expect our government to serve the people, and serve the people it will.*

Trump's campaign photos are consistent with a desire to present the same message:



This wasn't a scripted appearance forced by his campaign staff. According to the Washington Times:

> *Trump walked on stage in Greeley, Colorado to a large cheering crowd when he spotted a rainbow flag in the audience. As the music blasted through the speakers, Mr. Trump pointed to a supporter as if to ask if he could see his flag and then motioned for a campaign worker to help retrieve the LGBT symbol of equality from the attendee.*
>
> *Within seconds, Mr. Trump was walking around the platform with the rainbow flag in his hands and moments later unfurled it in full display. You could see a huge smile on Mr. Trump's face as he walked to both sides of the stage to proudly hold up the rainbow flag announcing support from the gay and lesbian community.*
>
> *Trump campaign spokesman Jason Miller told me, "Mr. Trump is campaigning to be President for ALL Americans and was proud to carry the 'LGBT for Trump' rainbow flag on stage in Greeley, CO yesterday.*



Donald J. Trump
@realDonaldTrump

Happy #CincoDeMayo! The best taco bowls are made in Trump Tower Grill. I love Hispanics! facebook.com/DonaldTrump/po …

This is just a tiny representative sample, but the rest is very similar. Trump has gone from campaign stop to campaign stop talking about how much he likes and respects minorities and wants to fight for them.

And if you believe he's lying, fine. Yet I notice that people accusing Trump of racism use the word "openly" like a tic. He's never just "racist" or "white supremacist". He's always "openly racist" and "openly white supremacist". Trump is openly racist, openly racist, openly racist, openly racist, openly racist, openly racist, openly racist. Trump is running on pure white supremacy, has thrown off the last pretense that his campaign is not about bigotry, has the slogan Make American Openly White Supremacist Again, is an openly white supremacist nominee, etc, etc, etc. And I've seen a few dozen articles like this where people say that "the bright side of a Trump victory is that finally America admitted its racism out in the open so nobody can pretend it's not there anymore."

This, I think, is the first level of crying wolf. What if, one day, there is a candidate who hates black people so much that he doesn't go on a campaign stop to a traditionally black church in Detroit, talk about all of the contributions black people have made to America, promise to fight for black people, and say that his campaign is about opposing racism in all its forms? What if there's a candidate who does something more like, say, go to a KKK meeting and say that black people are inferior and only whites are real Americans?

We might want to use words like "openly racist" or "openly white supremacist" to describe him. And at that point, nobody will listen, because we wasted "openly white supremacist" on the guy who tweets pictures of himself eating a taco on Cinco de Mayo while saying "I love Hispanics!"

III.

A rundown of some contrary talking points:

1. Is Trump getting a lot of his support from white supremacist organizations?

No, because there are not enough organized white supremacists to make up "a lot" of anyone's support.

According to Wikipedia on KKK membership:

> *As of 2016, the Anti-Defamation League puts total Klan membership nationwide at around 3,000, while the Southern Poverty Law Center puts it at 6,000 members total*

The KKK is really small. They could all stay in the same hotel with a bunch of free rooms left over. Or put another way: the entire membership of the KKK is less than the daily readership of this blog.

If you Google "trump KKK", you get 14.8 million results. I know that Google's list of results numbers isn't very accurate. Yet even if they're inflating the numbers *by 1000x, and there were only about 14,000 news articles about the supposed Trump-KKK connection this election, there are still two to three articles about a Trump-KKK connection for every single Klansman in the world.*

I don't see any sign that there are other official white supremacy movements that are larger than the Klan, or even enough other small ones to substantially raise the estimate of people involved. David Duke called a big pan-white-supremacist meeting in New Orleans in 2005, and despite getting groups from across North America and Europe he was only able to muster 300 attendees (by comparison, NAACP conventions routinely get 10,000).

My guess is that the number of organized white supremacists in the country is in the very low five digits.

2. Is Trump getting a lot of his support from online white nationalists and the alt-right?

No, for the same reason.

The alt-right is mostly an online movement, which makes it hard to measure. The three main alt-right hubs I know of are /r/altright, Stormfront, and 4chan's politics board.

The only one that displays clear user statistics is /r/altright, which says that there are about 5,000 registered accounts. The real number is probably less – some people change accounts, some people post once and disappear, and some non-white-nationalists probably go there to argue. But sure, let's say that community has 5,000 members.

Stormfront's user statistics say it gets about 30,000 visits/day, of which 60% are American. My own blog gets about 8,000 visits/day , and the measurable communities associated with it (the subreddit, people who follow my social media accounts) have between 2000 – 8000 followers. If this kind of thing scales, then it suggests about 10,000 people active in the Stormfront community.

4chan boasts about 1 million visits/day. About half seem to be American. Unclear how many go to the politics board and how many are just there for the anime and video games, but Wikipedia says that /b/ is the largest board with 30% of 4Chan's traffic, so /pol/ must be less than that. If we assume /pol/ gets 20% of 4chan traffic, and that 50% of the people on /pol/ are serious alt-rightists and not dissenters or trolls, the same scaling factors give us about 25,000 – 50,000 American alt-rightists on 4Chan.

Taking into account the existence of some kind of long tail of alt-right websites, I still think the population of the online US alt-right is somewhere in the mid five-digits, maybe 50,000 or so.

50,000 is more than the 5,000 Klansmen. But it's still 0.02% of the US population. It's still about the same order of magnitude as the Nation of Islam, which has *about 30,000 – 60,000 members, or the Church of Satan, which has about 20,000. It's not quite at the level of the Hare Krishnas, who boast 100,000 US members. This is not a "voting bloc" in the sense of somebody it's important to appeal to. It isn't a "political force" (especially when it's mostly, as per the 4chan stereotype, unemployed teenagers in their parents' basements.)*

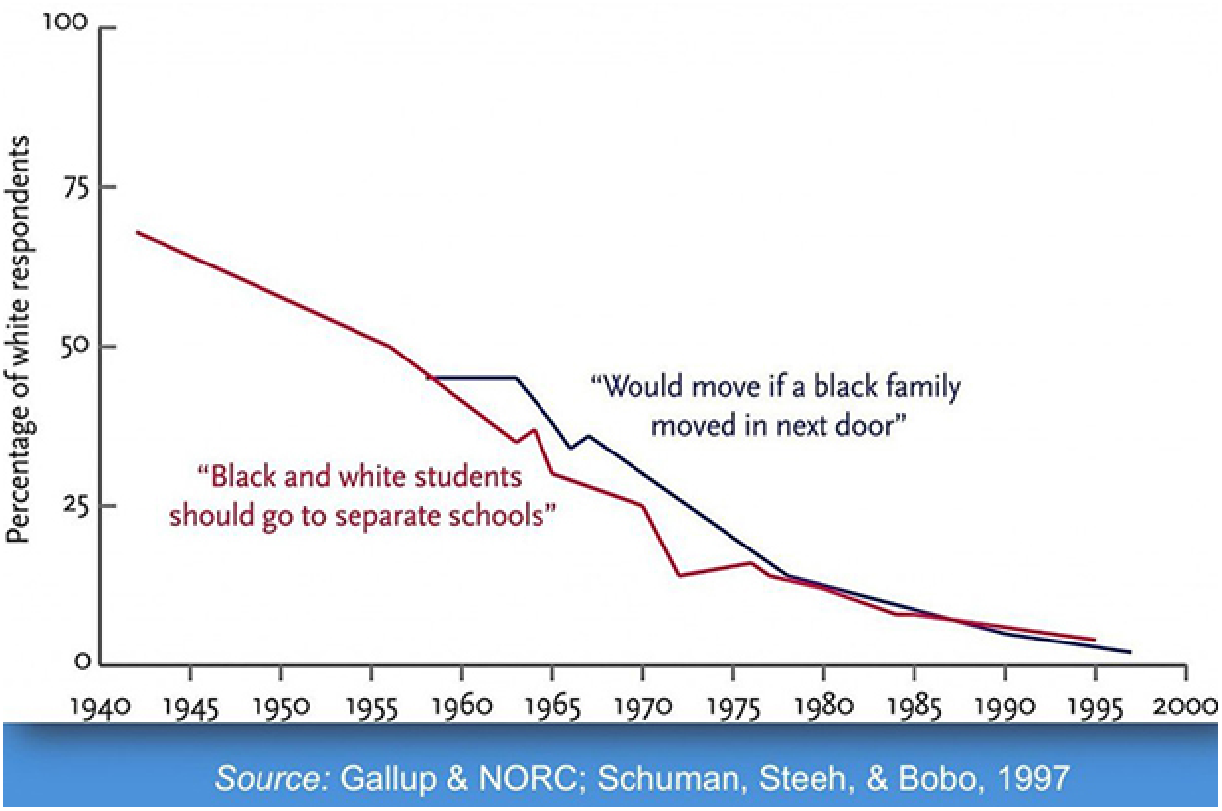
So the mainstream narrative is that Trump is okay with alienating minorities (= 118 million people), whites who abhor racism and would never vote for a racist (if even 20% of whites, = 40 million people), most of the media, most business, and most foreign countries – in order to win the support of about 50,000 poorly organized and generally dysfunctional people, many of whom are too young to vote anyway.

Caring about who the KKK or the alt-right supports is a lot like caring about who Satanists support. It's not something you would do if you wanted to understand real political forces. It's only something you would do if you want to connect an opposing candidate to the most outrageous caricature of evil you can find on short notice.

3. Is Trump getting a lot of his support from people who wouldn't join white nationalist groups, aren't in the online alt-right, but still privately hold some kind of white supremacist position?

There are surprisingly few polls that just straight out ask a representative sample of the population "Are you white supremacist?".

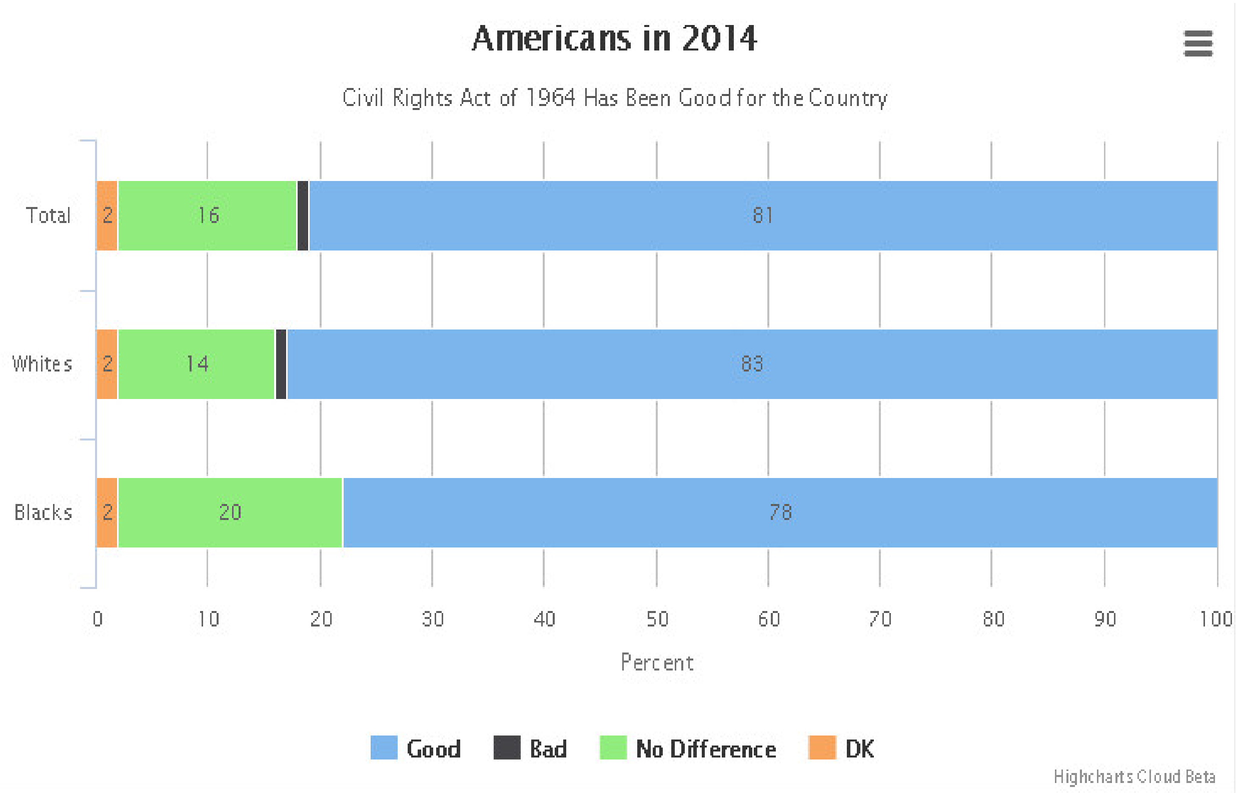
I can find a couple of polls that sort of get at this question in useful ways.



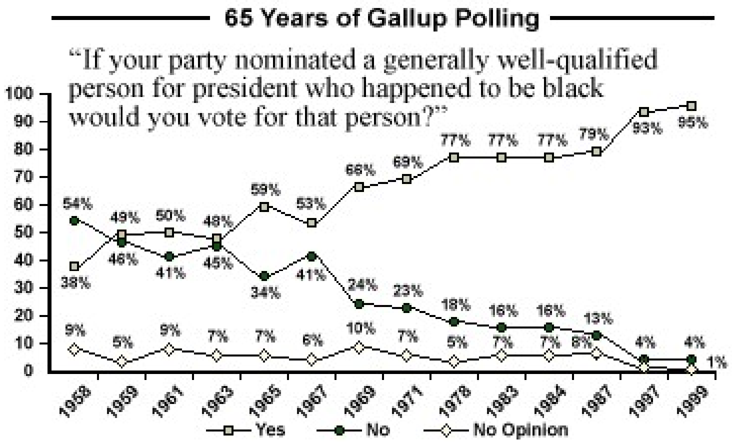
Source: Gallup & NORC; Schuman, Steeh, & Bobo, 1997

This poll from Gallup asks white Americans their support for school segregation and whether they would move out if a black family moved in next door. It declines from about 50% in 1960 to an amount too small to measure in the 1990s, maybe 1-2%, where it presumably remains today.

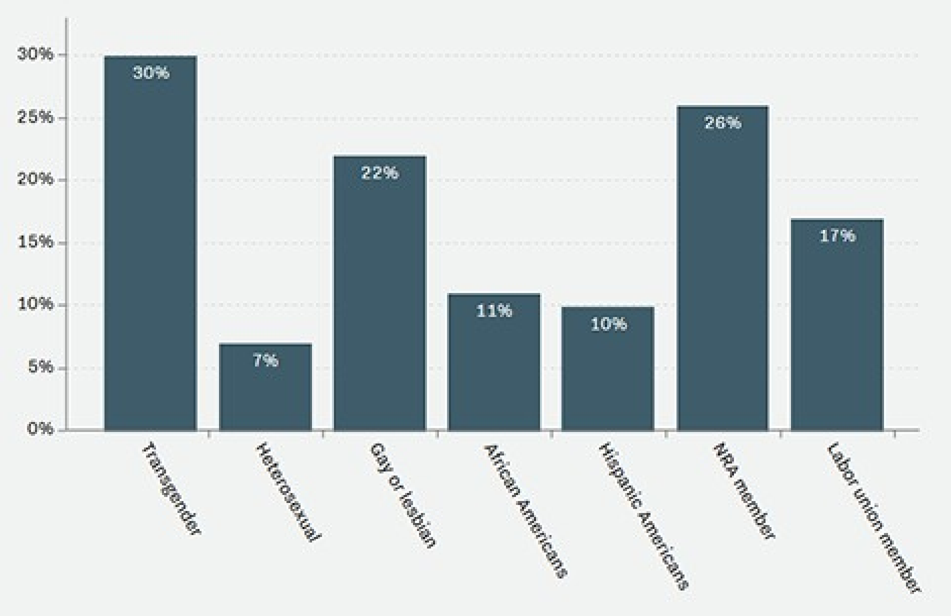
(this graph also seems relevant to the stories of how Trump's father would try to keep blacks out of his majority-white real estate developments in the late 60s/early 70s – note that at that time 33% of white families would move out if a black person moved in next door)



Here's a [CBS News](#) poll from 2014 asking Americans their opinion on the Civil Rights Act that legally prohibited discrimination. Once again, the number of whites who think it was a bad thing is too small to measure meaningfully, but looks like maybe 1-2%. Of note, whites were more convinced the Civil Rights Act was good than blacks were, though I guess it depends on the margin of error.



Another Gallup graph here, with the percent of people who would vs. wouldn't vote for an otherwise-qualified black candidate for President. It goes from 54% in 1968 to 5% in 1999; later polls that aren't included on the graph give numbers from [4%](#) to [7%](#), which sounds probably within the margin of error.



This is a Vox poll asking how many people had favorable vs. unfavorable views of different groups. 11% admit to “somewhat unfavorable” or “very unfavorable” views of blacks, which sounds bad, except that 7% of people admit to unfavorable views of heterosexuals by the same definition. This makes me think “have an unfavorable view about this group” is not a very high bar. If we restrict true “white supremacists” to those who have only “very unfavorable” views of blacks, this is 3%, well in line with our other sources.

(of note, 1% of respondents had “never heard of” blacks. Um…)

Maybe a better way of looking for racists: David Duke ran for Senate in Louisiana this year. He came in seventh with 58,000 votes (3%). Multiplied over 50 states, that would suggest 2.5 million people who would vote for a leading white supremacist. On the other hand, Louisiana is one of the most racist states (for example, Slate’s investigation found that it led the US in percent of racist tweets) and one expects Duke would have had more trouble in eg Vermont. Adjusting for racism level as measured in tweets, it looks like there would be about 1 million Duke voters in a nationwide contest. That’s a little less than 1% of voters.

So our different ways of defining “open white supremacist”, even for definitions of “open” so vague they include admitting it on anonymous surveys, suggest maybe 1-2%, 1-2%, 4-7%, 3-11%, and 1-3%.

But doesn’t this still mean there are some white supremacists? Isn’t this still really important?

I mean, kind of. But remember that 4% of Americans believe that lizardmen control all major governments. And 5% of Obama voters believe that Obama is the Antichrist. The white supremacist vote is about the same as the lizardmen-

control-everything vote, or the Obama-is-the-Antichrist-but-I-support-him-anyway vote.

(and most of these people are in Solid South red states and don't matter in the electoral calculus anyway.)

4. Aren't there a lot of voters who, although not willing to vote for David Duke or even willing to express negative feelings about black people on a poll, still have implicit racist feelings, the kind where they're nervous when they see a black guy on a deserted street at night?

Probably. And this is why I am talking about crying wolf. If you wanted to worry about the voter with subconscious racist attitudes carefully hidden even from themselves, you shouldn't have used the words "openly white supremacist KKK supporter" like a verbal tic.

5. But even if Donald Trump isn't openly white supremacist, didn't he get an endorsement from KKK leader David Duke? Didn't he refuse to reject that endorsement? Doesn't that mean that he secretly wants to court the white supremacist vote?

The answer is no on all counts.

No, Donald Trump did not get an endorsement from KKK leader David Duke. Duke has spoken out in favor of Trump, but refused to give a formal endorsement. You can read the explanation straight from the horse's mouth at davidduke.com: "The ZioMedia Lies: I Have Not Endorsed Donald Trump" (content warning: exactly what you would expect). If you don't want that site in your browser history, you can read the same story at The International Business Times.

No, Donald Trump did not refuse to reject the endorsement. From Politico.com:

> *Donald Trump says he isn't interested in the endorsement of David Duke, the anti-Semitic former Ku Klux Klan leader who praised the GOP presidential hopeful earlier this week on his radio show.*
>
> *"I don't need his endorsement; I certainly wouldn't want his endorsement," Trump said during an interview with Bloomberg's Mark Halperin and John Heilemann. He added: "I don't need anyone's endorsement."*
>
> *Asked whether he would repudiate the endorsement, Trump said "Sure, I would if that would make you feel better."*

From Washington Post:

> *ABC NEWS: "So, are you prepared right now to make a clear and unequivocal statement renouncing the support of all white supremacists?"*

> *TRUMP: "Of course, I am. I mean, there's nobody that's done so much for equality as I have. You take a look at Palm Beach, Florida, I built the Mar-a-Lago Club, totally open to everybody; a club that frankly set a new standard in clubs and a new standard in Palm Beach and I've gotten great credit for it. That is totally open to everybody. So, of course, I am."*

From CNN:

> *"David Duke is a bad person, who I disavowed on numerous occasions over the years," Trump said on MSNBC's "Morning Joe."*
>
> *"I disavowed him. I disavowed the KKK," Trump added. "Do you want me to do it again for the 12th time? I disavowed him in the past, I disavow him now."*

The concern comes from a single interview February 28, where Trump was asked to renounce support from David Duke and the KKK, where he gave a non-answer:

> *"I have to look at the group. I mean, I don't know what group you're talking about," Trump said. "You wouldn't want me to condemn a group that I know nothing about. I'd have to look. If you would send me a list of the groups, I will do research on them and certainly I would disavow if I thought there was something wrong. You may have groups in there that are totally fine — it would be very unfair. So give me a list of the groups and I'll let you know."*

This is pretty bad. But the next day Trump was saying that of course he denounced the KKK and blaming a "bad earpiece" for not being able to understand what the interviewer was saying.

Trump's bad earpiece explanation doesn't hold water – he repeated the name "David Duke" in his answer, so he obviously heard it. And his claim that he didn't know who David Duke was doesn't make sense – he's mentioned Duke before in various contexts.

But it's actually worth taking a look at those contexts. In 2000, Trump was already considering running for President. His friend Jesse Ventura suggested he seek the Presidential nomination of Ross Perot's Reform Party. Trump agreed and started putting together a small campaign (interesting historical trivia: he wanted Oprah Winfrey as a running mate). But after some infighting in the Reform Party, Ventura was kicked out in favor of a faction led by populist Pat Buchanan, who had some support from David Duke. Trump closed his presidential bid, saying: "The Reform Party now includes a Klansman, Mr. Duke, a neo-Nazi, Mr. Buchanan, and a communist, Ms. Fulani. This is not company I wish to keep." Later he continued to condemn the party, saying "You've got David Duke just joined — a bigot, a racist, a problem. I mean, this is not exactly the people you want in your party."

So we have Trump – who loudly condemned Duke before February 28th, and who loudly condemned Duke after February 28th – saying on February 28th that he *wanted to "look into" who David Duke was before refusing his (non-existent) endorsement. I'm not super sure what's going on. It's possible he wanted to check to see whether it was politically advantageous to officially reject it, which I agree is itself pretty creepy.*

But notice that the evidence on the side of Trump being against David Duke includes twenty years of unambiguous statements to that effect. And the evidence of Trump not being against David Duke includes one statement along the lines of "I don't know who he is but I'll look into it" on an interview one time which he later blamed on a bad earpiece and said he totally disavowed.

This gets back to my doubts about "dog whistles". Dog whistling seems to be the theory that if you want to know what someone really believes, you have to throw away decades of consistent statements supporting the side of an issue that everyone else in the world supports, and instead pay attention only to one weird out-of-character non-statement which implies he supports a totally taboo position which is perhaps literally the most unpopular thing it is possible to think.

And then you have to imagine some of the most brilliant rhetoricians and persuaders in the world are calculating that it's worth risking exposure this taboo belief in order to win support from a tiny group with five-digit membership whose support nobody wants, by sending a secret message, which inevitably every single media outlet in the world instantly picks up on and makes the focus of all their coverage for the rest of the election.

Finally, no, none of this suggests that Donald Trump is courting the white supremacist vote. Anybody can endorse anybody with or without their consent. Did you know that the head of the US Communist Party endorsed Hillary, and Hillary never (as far as I know) "renounced" their endorsement? Does that mean Hillary is a Communist? Did you know that a leader of a murderous black supremacist cult supported Donald Trump and Trump said that he "loved" him? Does that mean Trump is a black supremacist? The only time this weird "X endorsed Y, that means Y must support X" thing is brought out, is in favor of the media narrative painting Trump to be a racist.

This, to me, is another form of crying wolf. One day you might have a candidate who openly courts the KKK, in the sense of having a campaign platform saying "I like the KKK and value their support", speaking at Klan meetings, et cetera. And instead, you've wasted the phrase "openly courts the KKK" on somebody with a twenty year history of loudly condemning the KKK, plus one weird interview where he said he didn't know anything about it, then changed his mind the next day and said he hates them.

6. What about Trump's "drugs and crime" speech about Mexicans?

Trump said that:

> *When Mexico sends its people, they're not sending their best. They're not sending you. They're sending people that have lots of problems, and they're*

> *bringing those problems with us. They're bringing drugs. They're bringing crime. Their rapists. And some, I assume, are good people.*

Note how totally non-racist this statement is. I'm serious. It's anti-illegal-immigrant. But in terms of race, it's saying Latinos (like every race) include both good and bad people, and the bad people are the ones coming over here. It suggests a picture of Mexicans as including some of the best people – but those generally aren't the ones who are coming illegally.

Compare to eg Bill Clinton's 1996 platform (all emphasis mine):

> *We cannot tolerate illegal immigration and we must stop it. For years before Bill Clinton became President, Washington talked tough but failed to act. In 1992, our borders might as well not have existed. The border was under-patrolled, and what patrols there were, were under-equipped.* ***Drugs*** *flowed freely. Illegal immigration was rampant.* ***Criminal*** *immigrants, deported after committing* ***crimes*** *in America, returned the very next day to commit* ***crimes*** *again. President Clinton is making our border a place where the law is respected and* ***drugs*** *and illegal immigrants are turned away.*

Or John McCain in 2008:

> *Border security is essential to national security. In an age of terrorism,* ***drug*** *cartels, and* ***criminal*** *gangs, allowing millions of unidentified persons to enter and remain in this country poses grave risks to the sovereignty of the United States and the security of its people.*

Trump's platform contains similar language – and, like all past platforms, also contains language praising legal immigrants:

> *Just as immigrant labor helped build our country in the past, today's legal immigrants are making vital contributions in every aspect of national life. Their industry and commitment to American values strengthens our economy, enriches our culture, and enables us to better understand and more effectively compete with the rest of the world.*
>
> *We are particularly grateful to the thousands of new legal immigrants, many of them not yet citizens, who are serving in the Armed Forces and among first responders. Their patriotism should encourage all to embrace the newcomers legally among us, assist their journey to full citizenship, and help their communities avoid isolation from the mainstream of society. We are also thankful for the many legal immigrants who continue to contribute to American society.*

When Democrats and Republicans alike over the last twenty years say that we are a nation of immigrants but that illegal immigrants threaten our security, or may be criminals or drug pushers, they're met with yawns. When Trump says exactly the same thing, he's Literally the KKK.

7. What about the border wall? Doesn’t that mean Trump must hate Mexicans?

As multiple sources point out, both Hillary and Obama voted for the Secure Fence Act of 2006, which put up a 700 mile fence along the US-Mexican border. Politifact says that Hillary and Obama wanted a 700 mile fence but Trump wants a 1000 mile wall, so these are totally different. But really? Support a 700 mile fence, and you’re the champion of diversity and all that is right in the world; support a 1000 mile wall and there’s no possible explanation besides white nationalism?

8. Isn’t Trump anti-immigrant?

He’s at least anti-undocumented immigrant, which is close to being anti-immigrant. And while one can argue that “anti-immigrant” is different than “racist”, I would agree that probably nobody cares that much about British or German immigrants, suggesting that some racial element is involved.

But I think when Trump voters talk about “globalists”, they’re pointing at how they model this very differently from the people they criticize.

In one model, immigration is a right. You need a very strong reason to take it away from anybody, and such decisions should be carefully inspected to make sure no one is losing the right unfairly. It’s like a store: everyone should be allowed to come in and shop and if a manager refused someone entry then they better have a darned good reason.

In another, immigration is a privilege which members of a community extend at their pleasure to other people whom they think would be a good fit for their community. It’s like a home: you can invite your friends to come live with you, but if someone gives you a vague bad feeling or seems like a good person who’s just incompatible with your current lifestyle, you have the right not to invite them and it would be criminal for them to barge in anyway.

It looks like many Clinton supporters believe in the first model, and many Trump supporters in the second model. I think this ties into deeper differences – Clinton supporters are more atomized and individualist, Trump supporters stronger believers in culture and community.

In the second model, the community gets to decide how many immigrants come in and on what terms. Most of the Trump supporters I know are happy to let in a reasonable amount, but they get very angry when people who weren’t invited or approved by the community come in anyway and insist that everyone else make way for them.

Calling this “open white supremacy” seems like those libertarians who call public buses Communism, except if “Communism” got worn out on the euphemism treadmill and they started calling public buses “overt Soviet-style Stalinism”.

9. Don't Trump voters oppose the Emancipation Proclamation that freed the slaves?

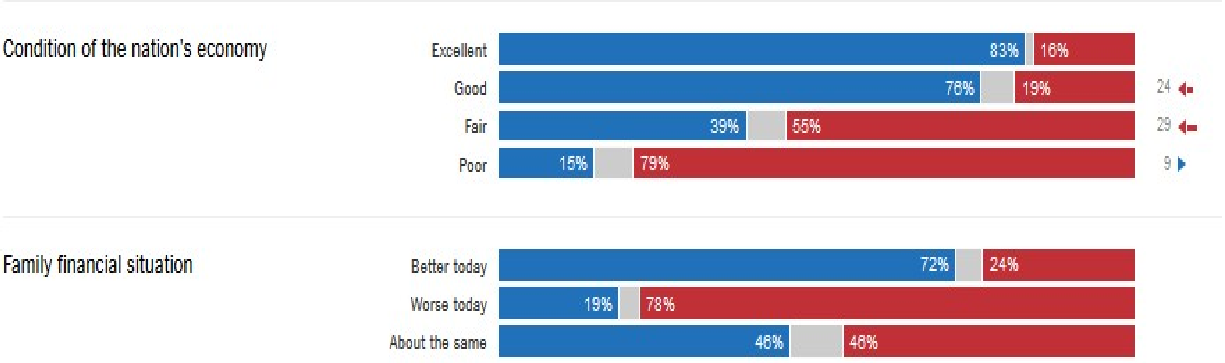
This was in New York Times, Vox, Huffington Post, Time, et cetera. It's very misleading. See Snopes for full explanation.

10. Isn't Trump anti-Semitic?

I feel like an attempt to avoid crying wolf might reserve that term for people who didn't win an Israeli poll on what candidate would best represent Israel's interests, or doesn't have a child who converted to Judaism, or hasn't won various awards from the American Jewish community for his contributions to Israel and American Judaism, or wasn't the grand marshal of a Salute To Israel Parade, or…

11. Don't we know that Trump voters are motivated by racism because somebody checked and likelihood of being a Trump voter doesn't correlate with some statistic or other supposedly measuring economic anxiety?

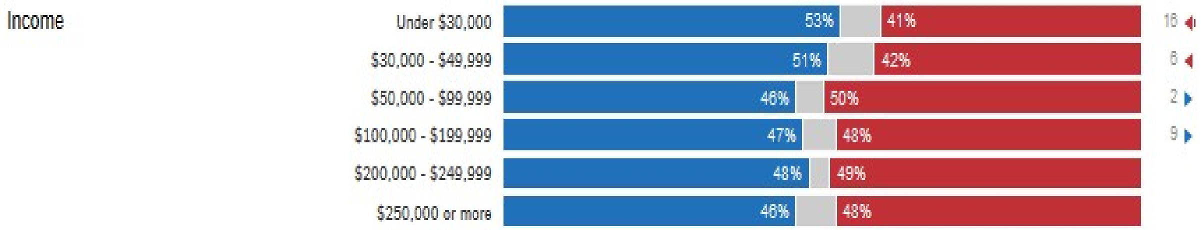
Although economic issues are only one part of Trump voters' concerns, they certainly are a part. You just have to look in the right places. See also:



12. Don't we know that Trump voters are motivated by racism because despite all the stuff about economic anxiety, rich people were more likely to vote Trump than poor people?

I keep hearing stuff like this, and aside from the object-level question, I think it's important to note the way in which this kind of thing makes racism the null hypothesis. "You say it's X, but you can't prove it, so it's racism".

Anyway, in this particular case, there's a simple answer. Yes, Republicans are traditionally the party of the rich. What's different about this election is that far more poor people voted Republican than usual, and far more rich people voted Democrat than usual.



Poor people were 16 percentage points more likely to vote Republican this election than last time around, but rich people (well, the richest bracket NYT got data about) were 9 percentage points more likely to vote Democrat. This is consistent with economic anxiety playing a big role.

13. Doesn't Trump want to ban (or "extreme vet", or whatever) Muslims entering the country?

Yes, and this is awful.

But why do he (and his supporters) want to ban/vet Muslims, and not Hindus or Kenyans, even though most Muslims are white(ish) and most Hindus and Kenyans aren't? Trump and his supporters are concerned about terrorism, probably since the San Bernardino shooting and Pulse nightclub massacre dominated headlines this election season.

You can argue that he and his supporters are biased for caring more about terrorism than about furniture-related injuries, which kill several times more Americans than terrorists do each year. But do you see how there's a difference between "cognitive bias that makes you unreasonably afraid" versus "white supremacy"?

I agree that this is getting into murky territory and that a better answer here would be to deconstruct the word "racism" into a lot of very heterogenous parts, one of which means exactly this sort of thing. But as I pointed out in Part 4, a lot of these accusations shy away from the word "racism" precisely because it's an ambiguous thing with many heterogenous parts, some of which are understandable and resemble the sort of thing normal-but-flawed human beings might think. Now they say "KKK white nationalism" or "overt white supremacy". These terms are powerful exactly because they do not permit the gradations of meaning which this subject demands.

Let me say this for the millionth time. I'm not saying Trump doesn't have some racist attitudes and policies. I am saying that talk of "entire campaign built around white supremacy" and "the white power candidate" is deliberate and dangerous exaggeration. Lots of people (and not just whites!) are hasty to generalize from "ISIS is scary" to "I am scared of all Muslims". This needs to be called out and fought, but it needs to be done in an understanding way, not with cries of "KKK WHITE SUPREMACY!"

14. Haven't there been hundreds of incidents of Trump-related hate crimes?

This isn't a criticism of Trump per se (he's demanded that his supporters avoid hate crimes), but it seems relevant to the general tenor of the campaign.

SPLC said they have 300 such hate incidents, although their definition of "hate incident" includes things like "someone overheard a racist comment in someone else's private conversation, then challenged them about it and got laughed at". Let's take that number at face value (though see here)

If 47% of America supports Trump (= the percent of vote he got extrapolated to assume non-voters feel the same way), there are 150,000,000 Trump supporters. That means there has been one hate incident per 500,000 Trump supporters.

But aren't there probably lots of incidents that haven't been reported to SLPC? Maybe. Maybe there's two unreported attacks for every reported one, which means that the total is one per 150,000 Trump supporters. Or maybe there are ten unreported attacks for every reported one, which means that the total is one per 45,000 Trump supporters. Since nobody has any idea about this, it seems weird to draw conclusions from it.

Oh, also, I looked on right-wing sites to see if there are complaints of harassment and attacks by Hillary supporters, and there are. Among the stories I was able to confirm on moderately trustworthy news sites that had investigated them somewhat (a higher standard than the SLPC holds their reports to) are ones about how Hillary supporters have beaten up people for wearing Trump hats, screamed encouragement as a mob beat up a man who they thought voted Trump, knocked over elderly people, beaten up a high school girl for supporting Trump on Instagram, defaced monuments with graffiti saying "DIE WHITES DIE", advocated raping Melania Trump, kicked a black homeless woman who was holding a Trump sign, attacked a pregnant woman stuck in her car, with a baseball bat, screamed at children who vote Trump in a mock school election, etc, etc, etc.

But please, keep talking about how somebody finding a swastika scrawled in a school bathroom means that every single Trump supporter is scum and Trump's whole campaign was based on hatred.

15. Don't we know that Trump supports racist violence because, when some of his supporters beat up a Latino man, he just said they were "passionate"?

All those protests above? The anti-Trump protests that have resulted in a lot of violence and property damage and arrests? With people chanting "KILL TRUMP" and all that?

When Trump was asked for comment, he tweeted "Love the fact that the small groups of protesters last night have passion for our great country".

I have no idea how his mind works and am frankly boggled by all of this, but calling violent protesters "passionate" just seems to be a thing of his.

16. But didn't Trump…

Whatever bizarre, divisive, ill-advised, and revolting thing you're about to mention, the answer is probably yes.

This is equally true on race-related and non-race-related issues. People ask "How could Trump believe the wacky conspiracy theory that Obama was born in Kenya, if he wasn't racist?" I don't know. How could Trump believe the wacky conspiracy theory that vaccines cause autism? How could Trump believe the wacky conspiracy theory that the Clintons killed Vince Foster? How could Trump believe the wacky conspiracy theory that Ted Cruz's father shot JFK?

Trump will apparently believe anything for any reason, especially about his political opponents. If Clinton had been black but Obama white, we'd be hearing that the Vince Foster conspiracy theory proves Trump's bigotry, and the birtherism was just harmless wackiness.

Likewise, how could Trump insult a Mexican judge just for being Mexican? I don't know. How could Trump insult a disabled reporter just for being disabled? How could Trump insult John McCain just for being a beloved war hero? Every single person who's opposed him, Trump has insulted in various offensive ways, including 140 separate incidents of him calling someone "dopey" or "dummy" on Twitter, and you expect him to hold his mouth just because the guy is a Mexican?

I don't think people appreciate how weird this guy is. His weird way of speaking. His catchphrases like "haters and losers!" or "Sad!". His tendency to avoid perfectly reasonable questions in favor of meandering tangents about Mar-a-Lago. The ability to bait him into saying basically anything just by telling him people who don't like him think he shouldn't.

If you insist that Trump would have to be racist to say or do whatever awful thing he just said or did, you are giving him too much credit. Trump is just randomly and bizarrely terrible. Sometimes his random and bizarre terribleness is about white people, and then we laugh it off. Sometimes it's about minorities, and then we interpret it as racism.

17. Isn't this a lot of special pleading? Like, sure, you can make up various non-racist explanations for every single racist-sounding thing Trump says, and say a lot of it is just coincidence or Trump being inexplicably weird, but eventually the coincidences start adding up. You have to look at this kind of thing in context.

I actually disagree with this really strongly and this point deserves a post of its own because it's really important. But let me try to briefly explain what I mean.

Suppose you're talking to one of those ancient-Atlantean secrets-of-the-Pyramids people. They give you various pieces of evidence for their latest crazy theory, *such as (and all of these are true):*

1. The latitude of the Great Pyramid matches the speed of light in a vacuum to five decimal places.
2. Famous prophet Edgar Cayce, who predicted a lot of stuff with uncanny accuracy, said he had seen ancient Atlanteans building the Pyramid in a vision.
3. There are hieroglyphs near the pyramid that look a lot like pictures of helicopters.
4. In his dialogue Critias, Plato relayed a tradition of secret knowledge describing a 9,000-year-old Atlantean civilization.
5. The Egyptian pyramids look a lot like the Mesoamerican pyramids, and the Mesoamerican name for the ancient home of civilization is "Aztlan"
6. There's an underwater road in the Caribbean, whose discovery Edgar Cayce predicted, and which he said was built by Atlantis
7. There are underwater pyramids near the island of Yonaguni.
8. The Sphinx has apparent signs of water erosion, which would mean it has to be more than 10,000 years old.

She asks you, the reasonable and well-educated supporter of the archaeological consensus, to explain these facts. After looking through the literature, you come up with the following:

1. This is just a weird coincidence.
2. Prophecies have so many degrees of freedom that anyone who gets even a little lucky can sound "uncannily accurate", and this is probably just what happened with Cayce, so who cares what he thinks?
3. Lots of things look like helicopters, so whatever.
4. Plato was probably lying, or maybe speaking in metaphors.
5. There are only so many ways to build big stone things, and "pyramid" is a natural form. The "Atlantis/Atzlan" thing is probably a coincidence.
6. Those are probably just rocks in the shape of a road, and Edgar Cayce just got lucky.
7. Those are probably just rocks in the shape of pyramids. But if they do turn out to be real, that area was submerged pretty recently under the consensus understanding of geology, so they might also just be pyramids built by a perfectly normal non-Atlantean civilization.
8. We still don't understand everything about erosion, and there could be some reason why an object less than 10,000 years old could have erosion patterns typical of older objects.

I want you to read those last eight points from the view of an Atlantis believer, and realize that they sound really weaselly. They're all "Yeah, but that's probably a coincidence", and "Look, we don't know exactly why this thing happened, but it's probably not Atlantis, so shut up."

This is the natural pattern you get when challenging a false theory. The theory was built out of random noise and ad hoc misinterpretations, so the refutation will have to be "every one of your multiple superficially plausible points is random noise, or else it's a misinterpretation for a different reason".

If you believe in Atlantis, then each of the seven facts being true provides “context” in which to interpret the last one. Plato said there was an Atlantis that sunk underneath the sea, so of course we should explain the mysterious undersea ruins in that context. The logic is flawless, it’s just that you’re wrong about everything.

This is how I feel about demands that we interpret Trump’s statements “in context”, too.

IV.

Why am I harping on this?

I work in mental health. So far I have had two patients express Trump-related suicidal ideation. One of them ended up in the emergency room, although luckily both of them are now safe and well. I have heard secondhand of several more.

Like Snopes, I am not sure if the reports of eight transgender people committing suicide due to the election results are true or false. But if they’re true, it seems really relevant that Trump denounced North Carolina’s anti-transgender bathroom law, and proudly proclaimed he would let Caitlyn Jenner use whatever bathroom she wanted in Trump Tower, making him by far the most pro-transgender Republican president in history.

I notice news articles like Vox: Donald Trump’s Win Tells People Of Color They Aren’t Welcome In America. Or Salon’s If Trump Wins, Say Goodbye To Your Black Friends. MSN: Women Fear For Their Lives After Trump Victory.

Vox writes about the five-year-old child who asks “Is Donald Trump a bad person? Because I heard that if he becomes president, all the black and brown people have to leave and we’re going to become slaves.” The Star writes about a therapist called in for emergency counseling to help Muslim kids who think Trump is going to kill them. I have patients who are afraid to leave their homes.

Listen. Trump is going to be approximately as racist as every other American president. Maybe I’m wrong and he’ll be a bit more. Maybe he’ll surprise us and be a bit less. But most likely he’ll be about as racist as Ronald Reagan, who employed Holocaust denier Pat Buchanan as a senior advisor. Or about as racist as George Bush with his famous Willie Horton ad. Or about as racist as Bill “superpredator” Clinton, who took a photo op in front of a group of chained black men in the birthplace of the KKK. Or about as racist as Bush “doesn’t care about black people!” 43. He’ll have some scandals, people who want to see them as racist will see them as racist, people who don’t will dismiss them as meaningless, and nobody will end up in death camps. He probably won’t do a great job fighting to end voter suppression, or helping people caught in the criminal justice system, but I’m not sure that makes him too different from the average member of the Republican Congress we have already.

Since everyone has been wrong about everything lately, I've started thinking it's more important than ever to make clear predictions and grade myself on them, so here are my predictions for the Trump administration:

1. Total hate crimes incidents as measured here will be not more than 125% of their 2015 value at any year during a Trump presidency, conditional on similar reporting methodology [confidence: 80%]
2. Total minority population of US citizens will increase throughout Trump's presidency [confidence: 99%]
3. US Muslim population increases throughout Trump's presidency [confidence: 95%]
4. Trump cabinet will be at least 10% minority [confidence: 90%], at least 20% minority [confidence: 70%], at least 30% minority [30%]. Here I'm defining "minority" to include nonwhites, Latinos, and LGBT people, though not women. Note that by this definition America as a whole is about 35% minority and Congress is about 15% minority.
5. Gay marriage will remain legal throughout a Trump presidency [confidence: 95%]
6. Race relations as perceived by blacks, as measured by this Gallup poll, will do better under Trump than they did under Obama (ie the change in race relations 2017-2021 will be less negative/more positive than the change 2009-2016) [confidence: 70%].
7. Neither Trump nor any of his officials (Cabinet, etc) will endorse the KKK, Stormfront, or explicit neo-Nazis publicly, refuse to back down, etc, and keep their job [confidence: 99%].
8. No large demographic group (> 1 million people) get forced to sign up for a "registry" [confidence: 95%]
9> No large demographic group gets sent to internment camps [confidence: 99%]
10. Number of deportations during Trump's four years will not be greater than Obama's 8 [confidence: 90%]

If you disagree with me, come up with a bet and see if I'll take it.

And if you don't, stop.

Stop fearmongering. Somewhere in America, there are still like three or four people who believe the media, and those people are cowering in their houses waiting for the death squads.

Stop crying wolf. God forbid, one day we might have somebody who doesn't give speeches about how diversity makes this country great and how he wants to fight for minorities, who doesn't pose holding a rainbow flag and state that he proudly supports transgender people, who doesn't outperform his party among minority voters, who wasn't the leader of the Salute to Israel Parade, and who doesn't offer minorities major cabinet positions. And we won't be able to call that guy an "openly white supremacist Nazi homophobe", because we already wasted all those terms this year.

Stop talking about dog whistles. The kabbalistic similarities between "dog-whistling" and "wolf-crying" are too obvious to ignore.

Stop writing articles breathlessly following everything the KKK says. Stop writing several times more articles about the KKK than there are actual Klansmen. Remember that thing where Trump started out as a random joke, and then the media covered him way more than any other candidate because he was so outrageous, and gave him what was essentially free advertising, and then he became President-elect of the United States? Is the lesson you learned from this experience that you need 24-7 coverage of the Ku Klux Klan?

Stop using the words "white nationalist" to describe Trump. When you describe someone as a white nationalist, and then they win, people start thinking white nationalism won. People like winners. This was entirely an own-goal and the perception that white nationalism is now the winning team has 1% to do with Trump and 99% to do with his critics.

Stop responding to everyone who worries about Wall Street or globalism or the elite with "I THINK YOU MEAN JEWS. BECAUSE JEWS ARE THE ELITES. ALL ELITES AND GLOBALISTS ARE JEWS. IF YOU'RE WORRIED ABOUT THE ELITE, IT'S DEFINITELY JEWS YOU SHOULD BE WORRIED ABOUT. IF YOU FEEL SCREWED BY WALL STREET, THEN THE PEOPLE WHO SCREWED YOU WERE THE JEWS. IT'S THE JEWS WHO ARE DOING ALL THIS, MAKE SURE TO REMEMBER THAT. DEFINITELY TRANSLATE YOUR HATRED TOWARDS A VAGUE ESTABLISHMENT INTO HATRED OF JEWS, BECAUSE THEY'RE TOTALLY THE ONES YOU'RE THINKING OF." This means you, Vox. Someday those three or four people who still believe the media are going to read this stuff and immediately join the Nazi Party, and nobody will be able to blame them.

Stop saying that being against crime is a dog whistle for racism. Have you ever met a crime victim? They don't like crime. I work with people from a poor area, and a lot of them have been raped, or permanently disabled, or had people close to them murdered. You know what these people have in common? They don't like crime When you say "the only reason someone could talk about law and order is that they secretly hate black people, because, y'know, all criminals are black", not only are you an idiot, you're a racist. Also, I judge you for not having read the polls saying that nonwhites are way more concerned about crime than white people are.

Stop turning everything into identity politics. The only thing the media has been able to do for the last five years is shout "IDENTITY POLITICS IDENTITY POLITICS IDENTITY POLITICS IDENTITY POLITICS IDENTITY POLITICS!" at everything, and then when the right wing finally says "Um, i…den-tity….poli-tics?" you freak out and figure that the only way they could have possibly learned that phrase is from the KKK.

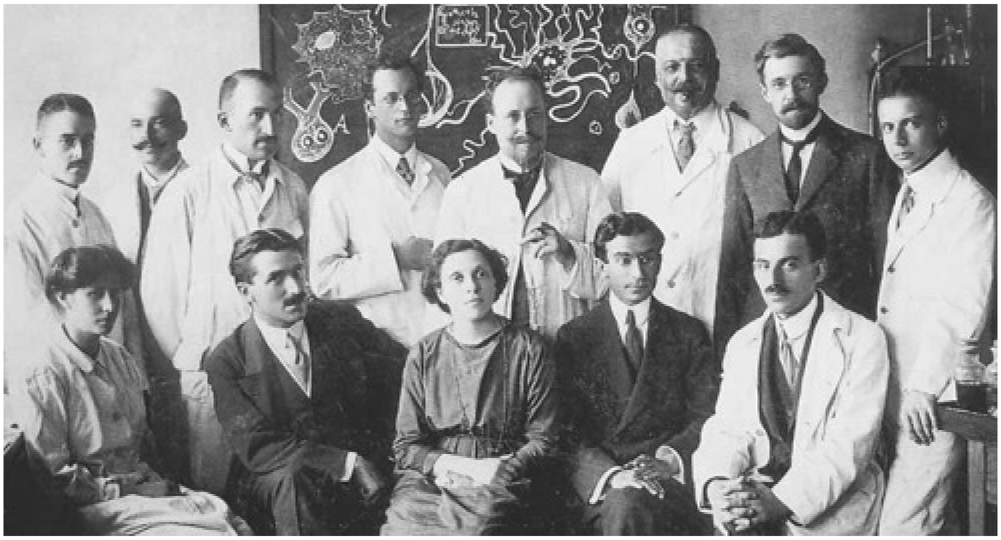
Stop calling Trump voters racist. A metaphor: we have freedom of speech not because all speech is good, but because the temptation to ban speech is so great that, unless given a blanket prohibition, it would slide into universal censorship of *any unpopular opinion. Likewise, I would recommend you stop calling Trump voters racist – not because none of them are, but because as soon as you give yourself that opportunity, it's a slippery slope down to "anyone who disagrees with me on anything does so entirely out of raw seething hatred, and my entire outgroup is secret members of the KKK and so I am justified in considering them worthless human trash". I'm not saying you're teetering on the edge of that slope. I'm saying you're way at the bottom, covered by dozens of feet of fallen rocks and snow. Also, I hear that accusing people of racism constantly for no reason is the best way to get them to vote for your candidate next time around. Assuming there is a next time.*

Stop centering criticism of Donald Trump around this sort of stuff, and switch to literally anything else. Here is an incompetent thin-skinned ignorant boorish fraudulent omnihypocritical demagogue with no idea how to run a country, whose philosophy of governance basically boils down to "I'm going to win and not lose, details to be filled in later", and all you can do is repeat, again and again, how he seems popular among weird Internet teenagers who post frog memes. In the middle of an emotionally incontinent reality TV show host getting his hand on the nuclear button, your chief complaint is that in the middle of a few dozen denunciations of the KKK, he once delayed denouncing the KKK for an entire 24 hours before going back to denouncing it again. When a guy who says outright that he won't respect elections unless he wins them does, somehow, win an election, the headlines are how he once said he didn't like globalists which means he must be anti-Semitic.

Stop making people suicidal. Stop telling people they're going to be killed. Stop terrifying children. Stop giving racism free advertising. Stop trying to convince Americans that all the other Americans hate them. Stop. Stop. Stop.

The Alzheimer Photo

A professor recently brought my attention to this photo of Alois Alzheimer and his colleagues in Munich (source):



Alzheimer is the very-German-looking guy with the silly mustache third from the right on the top. Far right is Friedrich Lewy, discoverer of Lewy bodies and Lewy body dementia. Bottom, second from the left, looking kind of like Petyr Baelish, is Ugo Cerletti, inventor of electroconvulsive therapy.

Other members of Alzheimer's team didn't make it to the group photo. These include Alzheimer's mentor, Emil Kraepelin, who discovered bipolar disorder, schizophrenia, etc, etc, etc (there's a reason modern psychiatry calls itself "neo-Kraepelinian"). They include two of Alzheimer's assistants, Hans Creutzfeldt and Alfons Jakob, who discovered Creutzfeldt-Jakob disease, the human version of mad cow. They include Alzheimer's collaborator Franz Nissl, who discovered Nissl bodies and the Nissl stain at the same lab.

If you come across a neurological disease that sounds like a guy's name, there's a not insignificant chance that guy is either in this picture or else just barely missed it.

This made me think of a lot of the discussion around when fields of science prosper versus when they go stagnant. The last few decades haven't really been great for neuropsychiatry. But one group of people in one lab came up with entire textbooks worth of advances. Why? Do we need to resurrect Alois Alzheimer and put him in charge of NIMH?

Part of it was that good histological staining had just been invented and Alzheimer's lab was on the bleeding edge, so they were just sitting around picking off the low-hanging fruit that could be discovered by staining stuff. But Kraepelin's and Cervetti's discoveries didn't have much to do with staining.

Part of it was that Alzheimer was in the right place at the right time. If he'd *really* wanted an impressive photo, he could have gotten together with his chief competitors, a group centered around Carl Westphal (cf. Westphal's sign, Edinger-Westphal nucleus) which included his students Arnold Pick (cf. Pick's dementia) and Karl Wernicke (cf. Wernicke's area, Wernicke-Korsakoff syndrome). Heck, if he wanted to go further, the number of people within a day's train journey staggers the imagination. Rudolph Virchow, Eugen Bleuler, Robert Koch, Sigmund Freud. Fin de siecle Central Europe was just a *really* good place for neurology and psychiatry.

Part of it was that the whole thing was arranged by Kraepelin, who besides being a scientific genius, was apparently an organizational genius as well. According to Wikipedia, "Kraepelin has been described as a 'scientific manager' and political operator, who developed a large-scale, clinically oriented, epidemiological research programme." See also Psychiatric Governance And The German Institute Of Psychiatry In Munich. Kraepelin grabbed all these people, threw them at the most interesting problems, and made sure they always had all the funding they wanted – although the final form of all of this as the Institute for Psychiatric Research didn't coalesce until after Alzheimer's death.

And part of it is the natural tendency for some institution to gain a reputation for being the best, and then attract the best people. I'm sure you could find some pretty impressive conjunctions of people if you looked at photos of Harvard departments.

My theory of apparent scientific stagnation has always been that it's easier to pick low-hanging fruit in one paradigm than to get entirely new ones – in other words, the problem is at least as much in the territory itself as in our engagement with it. I was interested to learn that one of the big hurdles to faster aircraft is a nonlinearity in fuel costs, which grow exponentially for physics reasons right when you start getting faster than modern planes. I think something similar might be going on here. Through painstaking trial-and-error, psychiatric hit upon a really fruitful paradigm of combining clinical observation, histopathology, and and random wacky ideas, right about when Alois Alzheimer opened his lab. Anybody who happened to be in the vicinity when the new paradigm was invented ended up getting a disease named after him. Eventually all the stuff that was easy to discover this way got discovered, and right now there just aren't any equally fruitful paradigms coming to our attention.

This story has a sad ending. Alzheimer (ironically) died young. He was succeeded by his student Walther Spielmeyer (cf. Spielmeyer-Vogt-Sjögren-Batten Disease), and then Kurt Schneider (cf. Schneider's first-rank symptoms). Schneider invented the modern concept of psychopathy, but unfortunately he was probably working from personal experience – this was in the middle of the rise of the Nazis. He was fired for political reasons and got replaced with Alzheimer's fellow Kraepelin protege, Ernst Rudin, who re-centered the whole thing around the role of psychiatry in sterilizing the feeble-minded. The chain that started with Kraepelin and Alzheimer ended in Rudin's own student, Josef Mengele.

After the war, Rudin was fined 500 deutschmarks, apparently the going penalty for leading a Nazi eugenics program at the time, and Kraepelin/Alzheimer's institute was re-founded as the Max Planck Institute of Psychiatry. As far as I know they're still around, but I haven't heard of them discovering any interesting new diseases lately.

SSC Journal Club: Expert Prediction Of Experiments

I.

It’s been a good month for fretting over failures of expert opinion, so let’s look at DellaVigna & Pope, Predicting Experimental Results: Who Knows What?

The authors ran a pretty standard behavioral economics experiment where they asked people on Mechanical Turk to do a boring task while being graded on speed and accuracy. Then they offered one of fifteen different incentive schemes, like “we’ll pay you extra if you do well” or “your score will be publicly visible”.

But the point of the study wasn’t to determine which incentive scheme worked the best, it would determine *who could best predict* which incentive scheme worked the best. The researchers surveyed a bunch of people – economics professors, psychology professors, PhD students, undergrads, business students, and random Internet users on Mechanical Turk – and asked them to predict the experimental results. Since this was a pretty standard sort of behavioral economics experiment, they were wondering whether people with expertise and knowledge in the field might be better than randos at figuring out which schemes would work.

They found that knowledgeable academics had some advantage over randos, but with enough caveats that it’s worth going over in more detail.

First, they found that prestigious academics did no better (and possibly slightly worse) than less prestigious academics. Full professors did no better than associate professors, assistant professors, or PhD students. People with many publications and citations did no better than people with fewer publications and citations.

Second, they found that field didn’t matter. Behavioral economists did as well as microeconomists did as well as experimental psychologists did as well as theoretical psychologists. To be fair, this experiment was kind of in the intersection of economics and psychology, so all of these fields had equal claim to it. I would have liked to see some geologists or political scientists involved, but they weren’t.

Third, the expert advantage was present in one measure of accuracy (absolute forecast error), but not in another (rank-order correlation). On this second measure, experts and randos did about equally well. In other words, experts were better at guessing the exact number for each condition, but not any better at guessing which conditions would do better or worse relative to one another.

Fourth, the expert advantage was pretty small. Professors got an average error of 169, PhD students of 171, undergrads of 187, MBA students of 198, and MTurk users of 271 (random guessing gave an error of about 416). So the difference between undergrads and experts, although statistically significant, was hardly overwhelming.

Fifth, even the slightest use of “wisdom of crowds” was enough to overwhelm the expert advantage. A group of five undergrads averaged together had average error 115, again compared to individual experts’ error of 169! Five undergrads averaged together (115) did about as well as five experts averaged together (114). Twenty undergrads averaged together (95) did about as well as twenty experts averaged together (99).

Sixth, having even a little knowledge of individuals' forecasting ability screened off expert status. The researchers gave forecasters some experimental data about the effects of a one-cent incentive and a ten-cent incentive, and asked them to predict the scores after a four-cent incentive – a simple, mechanical problem that just requires common sense. Randos who can do well on this problem do just as well as experts on the experiment as a whole. Likewise, randos who are noticed to do well on the first half of the experiment will do just as well as experts on the second half too. In other words, we're back to finding "superforecasters", people who are just consistently good at this kind of thing.

None of this seems to be too confounded by effort. The researchers are able to measure how much time people take on the task, whether they read the instructions carefully, etc. There is some advantage to not rushing through the task, but after that it doesn't seem to matter much. They also try offering some of the Mechanical Turkers lots of money for getting the answers right. That doesn't seem to help much either.

The researchers ask the experts to predict the results of this experiment. They (incorrectly) predict that prestigious academics with full professorships and lots of citations will do better than mere PhD students. They (incorrectly) predict that psychologists will do better than non-psychologists. They (correctly) predict that professors and PhD students will do better than undergrads and randos.

II.

What do we make of this?

I would *tentatively* suggest it doesn't look like experts' expertise is helping them very much here. Part of this is that experts in three different fields did about equally well in predicting the experimental results. But this is only weak evidence; it could be that the necessary expertise is shared among those three fields, or that each field contains one helpful insight and someone who knew all three fields would do better than any of the single-field experts.

But more important, randos who are able to answer a very simple question, or who do well on other similar problems, do just as well as the experts. This suggests it's possible to get expert-level performance just by being clever, without any particular expertise.

So is it just IQ? This is a tempting explanation. The US average IQ is 100. The undergrads in this experiment came from Berkeley, and Berkeley undergrads have an average SAT of 1375 = average IQ of 133 (this seems really high, but apparently matches estimates from The Bell Curve and the Brain Size blog; however, see Vaniver's point here). That same Brain Size post proposes that the average professor has an IQ of 133, but I would expect psychology/economics professors to be higher, plus most of the people in this experiment were from really good schools. If we assume professors are 135-140, then this would neatly predict the differences seen from MTurkers to undergrads to professors.

But the MBA students *really* don't fit into this model. The experiment gets them from the University of Chicago Booth School of Business, which is the top business school in the country and has an average GMAT score of 740. That corresponds to an IQ of almost 150, meaning this should be the highest-IQ sample in the study, yet the MBAs do worse than the undergrads. Unless I'm missing something, this is fatal to an IQ-based explanation.

I think that, as in *Superforecasting*, the best explanation is a separate “rationality” skill which is somewhat predicted by high IQ and scientific training, but not identical to either of them. Although some scientific fields can help you learn the basics of thinking clearly, it doesn’t matter what field you’re in or whether you’re in any field at all as long as you get there somehow.

I’m still confused by the MBA students, and expect to remain so. All MBA students were undergraduates once upon a time. Most of them probably took at least one economics class, which was where the researchers found and recruited their own undergraduates from. And most of them were probably top students from top institutions, given that they made it into the best business school in the US. So how come Berkeley undergraduates taking an econ class outperform people who *used to be* Berkeley undergraduates taking an econ class, but are now older and wiser and probably a little more selected? It might be that business school selects against the rationality skill, or it might be that business students learn some kind of anti-insight that systematically misleads them in these kinds of problems.

(note that the MBAs don’t put in less effort than the other groups; if anything, the reverse pattern is found).

III.

Does this relate to interesting real-world issues like people’s trouble predicting this election?

One important caveat: this is all atheoretical. As far as I know, there’s no theory of psychology or economics that should let people predict how the incentive experiment would go. So it’s asking experts to use their intuition, supposedly primed by their expertise, to predict something they have no direct knowledge about. If the experiment were, say, physicists being asked to predict the speed of a falling object, or biologists being asked to predict how quickly a gene with a selective advantage would reach fixation, then we'd be in a very different position.

Another important caveat: predictive tasks are different than interpretative tasks. Ability to predict how an experiment will go without having any data differs from ability to crunch data in a complicated field and conclude that eg saturated fat causes/doesn't cause heart attacks. I worry that a study like this might be used to discredit eg nutritional experts, and to argue that they might not be any better at nutrition than smart laymen. Whether or not this is true, the study doesn't support it.

So one way of looking at it might be that this is a critique not of expertise, but of "punditry". Engineers are still great at building bridges, doctors are still great at curing cancer, physicists are still great at knowing physics – but if you ask someone to predict something vaguely related to their field that they haven't specifically developed and tested a theory to cope with, they won't perform too far above bright undergrads. I think this is an important distinction.

But let's also not get too complacent. The experts in this study clearly *thought* they would do better than PhD students. They *thought* that their professorships and studies and citations would help them. They were wrong. The distinction between punditry and expertise is pretty fuzzy. Had this study come out differently, I could have argued for placing nice clear lab experiments about incentive schemes in the "theory-based and amenable to expertise" category. You can spin a lot of things either direction.

I guess really the only conclusion you can draw from all of this is not to put any important decisions in the hands of people from top business schools.

Contra Robinson On Schooling

I.

Nathan Robinson argues against school vouchers: Why Is The Decimation Of Public Schools A Bad Thing?

(note that despite the inflammatory title, he's arguing that the decimation of public schools is, indeed, a bad thing)

He starts with a meta-level point: most criticisms of Trump's Education Secretary nominee Betsy DeVos merely point out that she will promote schools vouchers instead of public schools, expecting their audience to be suitably turned off. But this won't change the minds of DeVos supporters, whose whole point is that they *want* more school vouchers. In order to convince people, you've got to convince people. If this doesn't seem like a suitably profound insight to you, the click the link, read the piece, and notice how there's something weird about it. Is it written in a funny font? Is the computer screen flickering or something? Finally, you realize with dawning horror that this is the first time you've read a logical argument, written in good faith and intended to convince someone, in the past you-can't-remember-how-many months.

But anyway, let me explain why I think it's wrong.

Robinson says:

> Introducing profit into the school system is very dangerous, for a simple reason: it creates a terrible set of incentives. If we hand a voucher to a for-profit private school, or give a large grant to a for-profit charter school, there is a strong incentive for the school to give as little in return as possible. After all, since a for-profit

> corporation exists to maximize value to shareholders (not value to students), for-profit schools should try to spend as little money educating students as possible, in order to reap the largest financial gains. If you don't have to spring for new lab equipment or new textbooks, you have no incentive to do so merely because it would benefit the students.

He does note the obvious counterargument, but he's not convinced:

> Privatization advocates have a compelling response to this argument. They reply that it misses the full picture. Yes, corporations have an incentive to maximize shareholder value. But they can't do that without satisfying their customers. The interests of shareholders and consumers are brought into alignment through the existence of choice. In the case of schools, because parents have a voucher, if the school is not prioritizing its students, parents can simply go elsewhere. Nobody is making them send their students to this particular school. The theory of school choice is about choice, and choice creates competition, which creates quality. A school that simply funneled money to its executives and shareholders would not long maintain its enrollment.
>
> But the theory of choice here is a romantic fiction. In reality, parents will not have many options among which to choose (there are only so many schools within a feasible distance of one's home, after all) and moving schools can be an extraordinarily disruptive and complicated process that hurts the child. We can also see how, even in theory, it is easy for a privatized school system to simply enrich the wealthy, while making schooling for poor children worse. In a public school system, all money is spent on the schools. In a for-profit

> school system, at least some portion of that money is directed instead toward the pockets of shareholders (if it wasn't, the for-profit schools couldn't continue to exist). And if we have a school district comprised in total of three for-profit elementary schools, and all of them simply pocket most of the voucher money while failing to educate the children, then no matter what "choices" among schools parents make, they won't be able to improve the quality of the schools. One might expect new operators to enter the market, but if the only way to make any real money on the children is to neglect them, then new operators won't be any better than the old ones.

This is a good point, made somewhat weaker by failure to consider why it doesn't apply to everything else.

Robinson compares school vouchers to foodstamps, which are basically "food vouchers" to be redeemed at grocery stores. I often see poor people using food stamps at my own grocery store, so I know the quality of service these poor people get for their money. And it is *really good*. Practically *all* grocery stores are really good. There's a story about Boris Yeltsin coming to America for the first time, walking into a random grocery store, seeing that random middle-class Americans had a better selection of goods than the highest-status Soviet officials, and freaking out that this was some kind of weird Potemkin economy that the Americans had set up to demoralize him. Grocery stores don't just have fifty different kinds of cereal and a hundred different kinds of soda, they're also really cheap. You can buy a day's worth of food for an hour's minimum-wage work, maybe two hours if you want a little quality and variety.

So why don't grocery store shareholders leech off so much money that everything is overpriced and has terrible service? Why aren't stores dingy and full of rats? Why don't we have a world where, as Robinson argues, the theory of choice is a romantic fiction because all of the grocery-related options available to poor people are terrible, and no new operators can do a better job because the only way to make money in the grocery business is to shaft customers and have a terrible store selling rice mixed with sawdust?

Something like 48% of Americans are satisfied with the education system in the US. My guess is 100% of Americans are satisfied with the grocery system in the US. Why should this be?

My guess: the loss from profits matters less than the gain in efficiency.

Profit margins are a specific number. You can just look them up. Usually they're not very big. Once you look up that number, you know how much profit the company is making. After that, you've circumscribed "the dangers of profit" to a relatively small amount.

An example: Health care in this country is overpriced and everyone knows it. Some people think this is because greedy insurance companies are charging too much in order to make a profit.

But health insurance companies have a profit margin of about 3% (see caveats here, but I do think the 3% number is the one relevant to this discussion). This is a big deal in terms of absolute number of dollars. But it's not a big deal if you're wondering how much they affect health care costs. If you're paying $5,000 a year for health insurance, then take away all profit motive from the insurance companies and you would pay $4850 a year for health insurance. This is less than year-to-year variation, let alone any of the components that actually matter.

Insurance companies aren’t callously throwing sick people out on the street for profitability reasons, they’re callously throwing sick people out on the street because they can only pay as much money as their customers give them and that isn’t enough to fund as much health care as people need. Or, well, maybe out of the people they callously throw out on the street, 3% are for profit and the other 97% are of necessity.

The same is true of other famously predatory businesses like payday lenders. Wikipedia notes:

> In a profitability analysis by Fordham Journal of Corporate & Financial Law, it was determined that the average profit margin from seven publicly traded payday lending companies (including pawn shops) in the U.S. was 7.63%, and for pure payday lenders it was 3.57%. These averages are less than those of other traditional lending institutions such as credit unions and banks. Comparatively the profit margin of Starbucks for the measured time period was just over 9%, and comparison lenders had an average profit margin of 13.04%. These comparison lenders were mainstream companies: Capital One, GE Capital, HSBC, Money Tree, and American Express Credit.

Payday lenders aren’t charging outrageous interest rates so they can get fat off the profits. They’re charging outrageous interest rates because loaning money to poor people who often fail to pay back their loans is a hard business to break even on.

Now apply this logic to private schools. You think they’re going to ruin everything by funneling more and more money into their own profitability? I propose that about 11% of their funding would go to profit, the same as current private schools. Actually less than that, since most current private schools serve rich people who don’t care about outrageous markups as long as they can buy prestige.

But isn’t 11% still more than nothing? If private schools cost the same amount of money as public schools, but 11% of that went to shareholder profits, wouldn’t our children receive an 11% worse education?

Or to put it another way – what are we buying with that 11% of the education budget?

The hope would be that we’re buying efficiency.

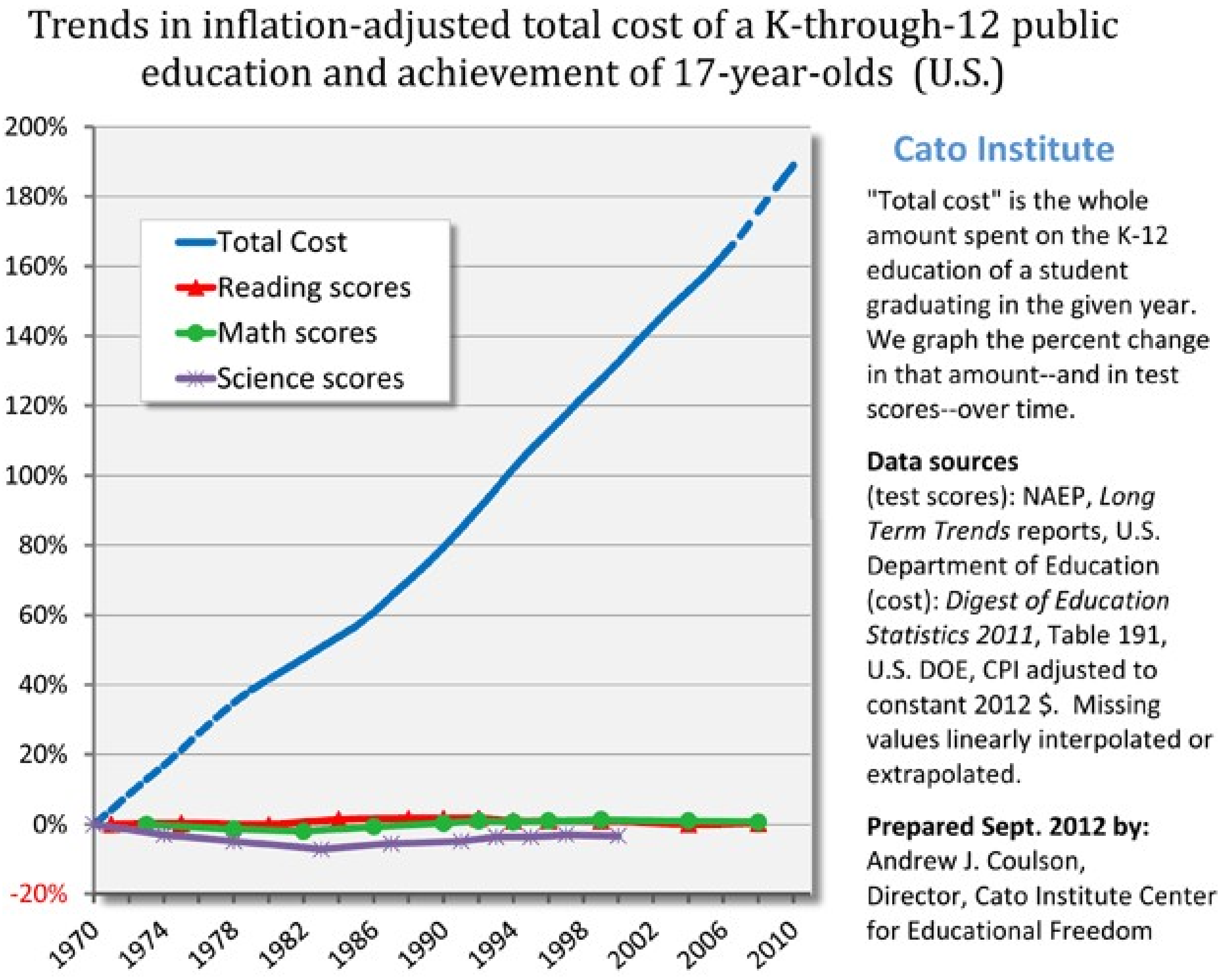
Robinson writes:

> Let’s consider what the conservative argument on schooling actually is. It goes like this: government-run institutions tend to function poorly. They are not efficient, like businesses are, because they do not have incentives to perform well. Businesses, because they must compete for customers in a market environment, must offer the best products if they want to stay profitable. Governments, on the other hand, can offer crappy products, and because they are state-imposed monopolies, there is no way for consumers to go elsewhere. School choice will improve schools, because instead of forcing students to attend whatever school the government happens to offer, choice allows parents to decide which school they prefer. Schools will have to strive to be better and better, because parents can pull their students out and go elsewhere if they don’t like them. Introducing a profit motive into schooling offers a powerful incentive for schools to offer a great product. If there is money to be made on being a good school, you can bet businesses will want to provide great schools. Thus private, for-profit schools with vouchers are a

> highly efficient way of delivering the best-quality education.

He counters that “introducing profit into the school system is very dangerous”, but never really says that the argument above is wrong. So we’re looking at a tradeoff here. There’s the dangers of profit and the promises of better efficiency. If profits aren’t going to take too much money out of the system, might the gain from efficiency be worth the small cost?

Here is a graph by the Cato Institute – note that this is already adjusted for inflation:



Trends in inflation-adjusted total cost of a K-through-12 public education and achievement of 17-year-olds (U.S.)

Cato Institute

"Total cost" is the whole amount spent on the K-12 education of a student graduating in the given year. We graph the percent change in that amount--and in test scores--over time.

Data sources
(test scores): NAEP, *Long Term Trends* reports, U.S. Department of Education
(cost): *Digest of Education Statistics 2011*, Table 191, U.S. DOE, CPI adjusted to constant 2012 $. Missing values linearly interpolated or extrapolated.

Prepared Sept. 2012 by:
Andrew J. Coulson,
Director, Cato Institute Center for Educational Freedom

In case you don’t trust them, here’s Politifact rating a similar claim mostly true, although they get slightly different numbers using a different methodology.

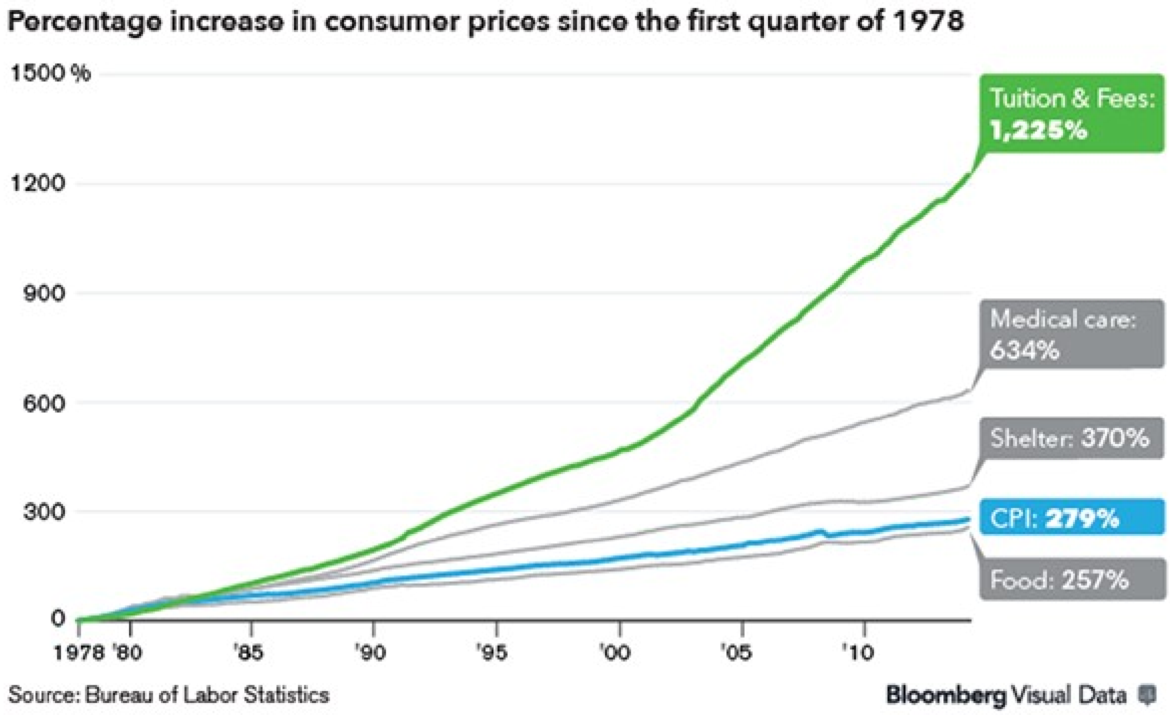
What are we to make of this? It’s not that teachers are getting paid any more – their salaries have remained stagnant over the time involved and they may even have lost ground compared to other professions. It’s not that school buildings cost more – I don’t have good data on schools in particular, but I looked into skyscrapers and found there wasn’t any general rise in construction prices. So how is an activity which basically involves getting a bunch of kids into a building and throwing a teacher at them rising so dramatically in the absence of changes in building or teacher prices? I’ve only heard three theories:

The first theory is that student test scores *are* improving, but we’re stuck in a Simpson’s Paradox. That is, students of some racial groups get higher scores than others, every racial group’s score is improving, but a higher percent of students are in low-performing racial groups now which is bringing the average down. But it looks like race-specific scores are also stagnant (1, 2), with the possible exception of a jump for blacks between 1980 and 1990. Given that no other ethnic group had this jump and school spending has increased at a constant rate the whole time, I think this is more likely socioeconomic factors than education spending.

The second theory is that all this extra money has been used to help previously underserved special needs kids. But these kids apparently cost about twice as much as average to educate. Up to 13% of students are special needs, although I don’t know if that’s the same definition of “special needs” as the people calculating the cost used. But if we take that as our estimate, then providing extra services to special needs kids can explain a rise of 13% in education costs, but not the 150% we actually see.

The third theory is Wilde's Law: "The bureaucracy is expanding to meet the needs of the expanding bureaucracy."

One argument in favor: the same thing that's happening to primary education is happening to college education. College costs about 4x as much (inflation-adjusted) as it did in 1980.



This is not because of decreased government subsidization. It's not going to shareholders – most colleges are nonprofits and even public institutions have seen outsized increases. And although I don't have any equivalent to the flat-lining test scores from primary school, today's college students don't seem to be four times better-educated than those of yore. So where *did* all that money go?

I don’t think anybody knows. There are whole studies that have been done on this. Every so often people argue about it on the editorial pages. Some people say “administration”, other people say it’s not exactly administration but it’s something else. I don’t know. But it doesn’t seem to be Simpson’s Paradox or special needs kids.

A lot of smart people think that easy availability of student loans fueled college cost increases – see eg here. The theory was that colleges *could* charge more money, so they *did*. I definitely don’t understand how this works economically, but it seems like somehow easy availability of money combined with lack of real competitive pressure caused colleges to increase administration and pass the cost on to students. Maybe something similar happened in primary schools?

If this third hypothesis of increased primary school costs is true, then going back to the level of bureaucracy we had in 1970 would cut costs by 75% while maintaining similar test scores. *Or* it would allow us to keep costs constant and pay for things that actually work.

The point is, private schools lose 11% of their funding to shareholder profit, and public schools apparently lose 75% of their funding to, uh, nobody really knows.

Do private schools also lose 75% of their funding to nobody-knows, since I don’t see many of them around as cheap as schools were in 1980? I’m not sure. Arnold Kling does some calculations here and suggests that private schools should have really big profit margins. But they apparently don’t. Overall I admit I am confused on this issue. But I am a little more hopeful that private schools might be able to work this issue out than that public schools can. [EDIT: see note at bottom of post]

It may seems kind of bloodless to focus on cost. But the amount of money added to the public school budget without any change in outcomes is more than enough to house every homeless person in the country in style and comfort. Money *matters*. And when we talk about private schools being obsessed with profit, that’s an argument about money. It’s acknowledging that every dollar diverted to shareholders is a dollar that isn’t going to improving education, housing the homeless, or something else useful.

Robinson writes:

> It’s because the things needed by poor people, if done well, will never be money-makers. Introducing an incentive to make money will necessarily mean exploiting and neglecting the poor, whose “choices” are highly constrained by their circumstances. I fear privatization not because of some mystical devotion to the inefficiencies of government but because I fear the erosion of the idea of education as something that isn’t win-win, that we give to children because they deserve it rather than because we can profit from it. I worry that the sort of people who run things “like a business” do not really care about children very much, and are motivated by the wrong incentives. I am concerned about what would happen if they ever faced a choice between doing the right thing and doing the lucrative thing.

With all due respect, I think there *is* something mystical in this thought process, some demon best exorcised with a bell, candle, and Public Choice Theory textbook. There’s no object called The System, which is focused on profit in businesses and focused on education in public services. There’s just a bunch of people motivated by a combination of ethics, incentives, and trying not to get fired. Business isn't antithetical to caring – the average family doctor is motivated by desire to help patients, even though she's also a small business. And lack of a profit motive doesn't guarantee good behavior – it looks like the administrators of nonprofit colleges decided to spend their windfall on prestige and empire-building rather than on keeping costs low.

I have very low confidence in this. I know many people who are involved in education, and they are all very good people who are very passionate and definitely would never skim 75% off the top and use it to buy gold-plated yachts for themselves. In my home state of California, there was a big funding shortfall ten years ago, and schools tried to cut everything they could, but finally they said there was nothing left to cut and they had no more ideas, and I believe that they tried as hard as they could. If bureaucracy is inflating the price of schooling, it's not doing so in an obvious way where you can point a finger at the exact bureau involved.

But it might be a general ethos of inefficiency that makes a lot of little things add up – I know it is in health care. And I at least think it would be worth trying the experiment.

(I realize the experiment is already being tried, with wildly varying results based on the specifics. I want to look into this research in more depth soon to see if there are any consistent trends.)

II.

A digression to support my point that sometimes increased efficiency can compensate for money diverted into profits – what about hospitals?

Hospitals are about evenly split between for-profit and not-for-profit institutions. Measuring “hospital quality” is even harder than measuring school quality, but researchers have tried to do this on various metrics. The results are hard to sum up, and I was only able to find a few studies and not anyone’s magisterial summation of the field, but it looks like there are minimal differences between for-profit and non-profit private hospitals, with government hospitals doing worst of all:

— A team from Harvard finds that for-profits and non-profits have about equal quality, and government-owned hospitals are worse than either. A follow-up study by the same team finds non-profit hospitals becoming for-profit is not associated with a drop in care.

— Truven Health Analytics finds some advantages for church-owned nonprofit hospitals, with secular nonprofit hospitals and for-profit hospitals in the middle, and government-owned hospitals worst of all. Note that this is my interpretation of a lot of different data and you might want to look at the particular metrics they use to draw your own conclusions.

— A textbook on the hospital industry finds that “on average, the performance of non-profit hositals in treating elderly patients with heart disease appears to be slightly better than that of for-profit hospitals, even after accounting for systematic differences…however, this small average difference masks an enormous amount of variation in hospital quality within the for-profit and not-for-profit hospital groups.”

— A study in a cardiology journal found “no evidence that for-profit hospitals selectively treat less sick patients, provide less evidence-based care, limit in-hospital stays, or have patients with worse acute outcomes than nonprofit centers”.

— As per the Handbook Of Health Economics:

> The most rigorous and extensive study of large-scale empirical study of quality published to date that permits comparisons of quality by hospital ownership is by Keeler and co-authors (1992). They used two process measures of quality based on reviews of 14,000 medical record for five diseases in five states. One of this “explicit process” gauged the extent to which the charts showed that specific diagnostic and therapeutic procedures were performed competently. Rather than focus on particular elements of care as explicit process did, a second process measure, “implicit process,” measured the care process overall. For example, one of their implicit process questions to physician reviewers was: “Based on what you now know about this case, would you send your mother to this hospital?” In addition, they gauged quality on an outcome measure — the difference between actual mortality and the rate that would be expected based on the patient’s characteristics.
>
> They found no difference in quality between private not-for-profit and for-profit hospitals on two indicators, excess mortality and explicit process; public hospitals fared worse on both criteria. However, on a third measure, implicit process, there was a statistically significant difference between quality of care of private not-for-profit hospitals and the other two ownership types, indicating higher quality levels for the for profits. The authors appear to have been more persuaded by the results on the first two indicators, stating that

> “nonprofit and for-profit hospitals provide similar quality overall”
>
> In their national study of 981 hospitals in 1983-84, Shortell and Hughes (1988) found no difference in quality measured in terms of mortality by ownership. However, using fewer covariates, Hartz et al. (1989) did find that mortality was higher in for-profit than in private not-for-profit hospitals.
>
> Sloan and co-authors (1998a,1998b, 1998c) examined outcomes of care of elderly persons hospitalized for one of four conditions: hip fracture; stroke; coronary heart disease; and congestive heart failure. They analyzed the first admission for these conditions since patients with a first unanticipated major health shock are less likely to shop among hospitals. Their outcome measures were survival, functional status, cognitive status, and living arrangements (probability of living in a nursing home). Although, on some measures, patients admitted to major teaching hospitals did better, a result consistent with Keeler et al. (1992), there were no statistically significant differences in outcomes between non-teaching private not-for-profit and for-profit hospitals. On some measures, elderly patients admitted to non-teaching government hospitals had worse outcomes, holding a large number of other factors constant.

So it looks like the differences between for-profit and not-for-profit hospitals are pretty slight, and that government hospitals are likely worse than either. This should be equally confusing to people who believe that the profit motive is invariably destructive *and* to people who believe it invariably results in better service.

This would be a good time to note that, contrary to the impression one might get from the *Current Affairs* article, only 13% of charter schools are for-profit. So a system with vouchers and charter school would probably be most like private non-profit hospitals, which these studies also show as doing well, and as decisively better than government-run hospitals. Also, everything I've talked about up to this point is mostly irrelevant. Maybe I should have started by mentioning this.

III.

From the same article:

> There are other serious problems with the "gutting" of public schools. As we have argued before in Current Affairs, converting public schools to a voucher system makes education operate similarly to food stamps. After all, SNAP benefits operate roughly the same way: instead of giving people food, we give them the equivalent of money, which they then use to go and buy food. A voucher program does the same for schooling: instead of giving them schools, we give them a voucher, which they can use to go and find a private school. But look what happens with food stamps: the moment you start handing out a "voucher," conservatives start seeing it as some kind of unearned "handout." Pressure then develops to cut the handout. Is there any reason to think that "education stamps" would be subjected to less cost-cutting political pressure than food stamps? A serious problem with voucher programs is that they erode the idea of education as a fundamental right, instead making it seem like a privilege that one does not necessarily deserve. But education should be a right, because children cannot help the circumstances of their

> birth, and should therefore not be punished for their parents' poverty.

I think this is drawing the wrong lesson from education's popularity relative to food stamps. Robinson thinks conservatives like one-size-fits-all handouts, but not voucher handouts. I think conservatives like universal handouts, but not handouts to the poor.

Imagine a world where food stamps are replaced by the Federal Food Agency. Every week, a truck comes to poor people's houses and gives them a one-size-fits-all food package that the government believes satisfies their nutritional needs. Do you think conservatives would be any happier with this than they are with food stamps? For that matter, did conservatives support public housing projects any more or less than they support housing vouchers now?

On the other hand, Medicare remains popular even though it's essentially a voucher. Patients with Medicare choose their doctor, choose their hospital, and then Medicare pays for it. But because everyone expects to benefit someday, it has pretty broad bipartisan support; even its critics mostly want to change rather than eliminate it.

I don't think changing public education from a service to a voucher would change whether people support it or not.

IV.

So I disagree with Robinson's specific arguments. But there are some things that worry me about school vouchers.

First, the hospital case study is kind of ambiguous. Although for-profit hospitals aren't noticeably worse than not-for-profit, they're also not noticeably better. And the existence of for-profit hospitals hasn't started some kind of virtuous cycle where all hospitals compete to save money and provide better care that ends up with hospitals being lean and inexpensive and just as accessible as grocery stores. Having a field be open to competition isn't necessarily incompatible with it being overpriced and inefficient. And commenters point out that existing private schools are not generally 75% cheaper than public schools, suggesting that cost-cutting is hard.

Second, Robinson notes later that:

> Privatization schemes are also heavily dependent on the existence of highly astute parents, who have the time and inclination to carefully study schools. The most vulnerable children are unlikely to have such parents. And we can imagine a system in which private schools offer parents $100 out of the voucher money if they agree to enroll their children. Desperate and uncaring parents might snap up the cash, with the neediest children ending up in the most vicious, uncaring, profit-grubbing schools.

I doubt there would be such blatant kickbacks – they'd be illegal and I don't think they've happened on other voucher programs like food stamps – but his point that many parents are ignorant or malicious is well-taken. You don't need literal bribery to get schools which are very good at having flashy ad campaigns but not very good at education. Parents might not check the test scores of a smooth-sounding school any more than they check the health care grades of their local hospital. The worst-case scenario is schools associated with cults or fringe political ideologies that prey on the children of people who believe them, either out of genuine fanaticism or a cynical calculation that fanatics are easy to milk.

Third, the whole point of Trumpism is that once we have fewer immigrants we can create a culturally cohesive community where everybody shares some core values. But the school system – as fractured and diverse as it is – is really one of the only institutions responsible for instilling some basic civic values in everyone and making sure they're all on the same page. I do not put it past people to start sending their kids to schools that teach liberal values or conservative values in particular, and then one of the few (albeit mostly ineffective) brakes on further polarization is removed. On the other hand, Catholic school is already sort of like this and they don't seem to be some weird foreign cancer on the body politic, so maybe it's not such a big deal?

Fourth, vouchers could worsen class segregation. Maybe not *too much*, because everyone already goes to public schools in their own class-segregated neighborhood anyway. But at least there's a *little* socioeconomic diversity now. And with vouchers, there's a risk of deliberate sorting/signaling, where if everyone gets a voucher for $10,000, decent schools will charge $15,000 just to sell the "privilege" of going to a school without poor students. That is, in the same way people will pay extra for a house in a gated community because they worry poor people make bad neighbors, they might pay extra for a spot in a more-expensive school because they worry poor people will make bad classmates. A little bit of this segregation goes a long way, because if enough people do this then the exactly-$10,000 schools will *only* have poor people, in much the same way that a little bit of racial segregation goes a long way.

(I don't know much about proposed voucher systems, but I wonder if it would be possible to have a system where you're not allowed to combine the voucher with your own money. That is, if you get a $10,000 voucher, you can go to a school charging $10,000. But if you want to go to a school charging $11,000, you have to throw away the voucher and pay the whole price out of pocket.)

To all these downsides we would have to add one very big upside – it destroys the incentive to overspend on/segregate housing in order to get into a "good school district". Elizabeth Warren has argued this is primarily behind the secular rise in real estate prices that has undermined the economic position of the middle class for the past fifty years. This factor could easily be more important than everything else combined and might make school vouchers a plus even if they seriously worsened the quality of education.

Overall my thoughts on school vouchers are the same as my thoughts on pretty much everything in this category: let's experiment. Figure out a window of acceptable possibilities that are reversible and don't have too much risk, and let different states and areas try different ones. As we start to understand things better, extend the window of possibilities in the relevant direction. Check results. Rinse. Repeat. Then figure something out.

V.

Finally, one more point from the article that deserves its own discussion:

> If we have a school district comprised in total of three for-profit elementary schools, and all of them simply pocket most of the voucher money while failing to educate the children, then no matter what "choices" among schools parents make, they won't be able to improve the quality of the schools. One might expect new operators to enter the market, but if the only way to make any real money on the children is to neglect them,

> then new operators won’t be any better than the old ones.

It’s important because lack of good competition is indeed the bane of all of these sorts of industries. Economic theory predicts that in a perfectly competitive environment businesses will be pretty good; it is much less sure of itself in these sorts of three-school districts without enough competition to have much effect.

In the real world, someone will have to empirically determine how much this matters. In my own fantasy world, I have a solution that the new Education Secretary probably won’t be on board with: Let’s let random people open tiny schools.

Something like 3% of parents home-school their children. This cuts across class and racial lines better than you’d think. All the research shows that home-schooled students do much better than traditionally schooled students on standardized tests, college admission exams, college GPAs, and general life satisfaction as adults. This is probably unfair, because home-schooled students are the descendents of the sort of thoughtful conscientious parents who want to home-school their children, so they probably have a big genetic advantage. But there is at least absolutely no evidence that home-schooling makes anyone do any worse.

The average cost per pupil per year in the US is something like $10,000. So suppose we give everyone $10,000 school vouchers. A parent who wants to make the median US yearly income of $30,000 would have to teach three students. Add in some overhead and curriculum costs, and maybe it’s more like five students.

So imagine. A woman has a kid and decides she doesn't want to go back to work and leave the kid in daycare for eighteen years. She takes some test, clears some regulatory hurdle, promises that she'll clear a certain bar on her kids' standardized test scores, and registers as an approved school. Then she gets a couple of friends and neighbors who trust her to send their kids to her too. Maybe her husband works outside the home, so she doesn't even need five. She's happy with two or three (I think it would be important that you can't make any money by educating your own kid; otherwise the incentive is to keep them out of school and pretend to be educating them yourself). Then she tutors them in a class a fifth the size of comparable public school classes.

If you're an actual, qualified teacher, maybe you can get ten or twenty kids who are interested. That's $100,000 to $200,000, minus your overhead, much more than qualified teachers make today with a much lower class size. Remember, for the majority of American history, kids were taught by a member of the community in a one room schoolhouse, and that was the system that produced Benjamin Franklin, Thomas Edison, et cetera.

(and remember that all the research shows that formal teacher training and level of teacher credentialing has zero effect on how well teachers teach kids)

This would provide provide a means of self-directed, boss-free income for millions of people, including undercredentialled poor people, disabled people who can't leave the home, people in rural areas, and especially young mothers. It would rebuild community ties. And it would ensure no one ever has to worry about districts with only three schools, or even districts with only thirty schools.

I don’t know. Probably there’s some sort of horrible flaw that I’m missing. But I still think the moral of the story is to experiment more. And school vouchers might be a good start.

EDIT: From the comments: “At my independent non-profit high school we have reduced the cost per student-year down to $3,000. Despite this budget we are able to offer the students many opportunities that public schools don’t. Art classes like glassblowing and copper and silversmithing, advanced science curricula like organic and biochemistry, health class vastly more informative that the state requirements, zero bullying enforced by a self-organized student culture, I could go on all day. For comparison the average cost for nearby school districts is $17,000 per high schooler-year and offer a fraction of the services we provide.”

Highlights From The Comment Thread On School Choice

Several people including Yehoshua K and Freddie deBoer point out that “nonprofit” and “for-profit” are potentially meaningless terms in situations like these.

IrishDude adds some context to the for-profit hospital scene by noting that companies are not allowed to open new hospitals until they apply for and are granted a Certificate Of Need, apparently on the basis of a theory that an oversupply of hospitals would increase (?!) costs.

Algorizmi describes his work with a private school that costs less than half as much as most public schools, including how it saves money

Half a dozen people yell at me for saying that the grocery industry worked well for the poor, objecting that I had forgotten about Food Deserts. Various other people save me some time by pointing out that most of the claims about Food Deserts are kind of fake (1, 2, 3). There’s a defensible version of the term, which is that in very poorly planned car-centered zoning-regulated cities without good public transportation it’s not always possible for someone without a car to easily get to the stores they want, but just posing the problem that way makes the solution pretty obvious.

Spotted Toad is great as usual, pointing out among other things that Obama has already led the most pro-charter administration in history.

Douglas Knight proposes that college demand curves are upward-sloping, which is kind of terrifying but seems to have at least anecdotal support.

Robinson: “The things needed by poor people, if done well, will never be money-makers.” Matt M: “A bold claim considering that the biggest company on Earth, more than double #2, serves the needs of poor people almost exclusively.”

MDP has worked in the payday loan industry and explains how their interest rates can be so high and their profits so low.

Many people (1, 2, 3) point out the high cost to schools of misbehaving kids, and try to explain the rise in education costs by saying that kids today are raised wrong (or not at all) which makes them harder to control. But everybody who looks for this kind of thing finds the opposite – kids today have less teenage pregnancy, crime, dropouts, et cetera. I don’t know if anyone has specifically looked at classroom misbehaving, but it would be weird for that to be getting worse in such an isolated way.

Lots of people are very angry at me for posting the graph from the Cato Institute for various reasons. A few people object that it is dishonest because it didn’t adjust for inflation, even though it did adjust for inflation and is very clear about that. A few people object that it is dishonest because it puts cost increases and score increases on the same scale as each other, instead of skewing them to make the effect look bigger than it really is – this is a definition of “dishonest” I haven’t heard before (maybe I am being uncharitable here?). Several others say that test scores have increased more than they give credit for. You can look at a very good summary of test score changes here. You’ll see there are some gains among younger students, but much fewer among older kids – for example, 17 year olds’ math scores went from 304 in 1971 all the way to 306 in 2012. This is not just a race-related Simpson’s Paradox –

among white students alone, for example, the gain was 4 points (though blacks did gain 18, as I mentioned before). I am not sure what good it is to have high gains in early years if those gains are all lost by the time kids leave school. Overall I think the Cato graph comes out looking pretty good. But even if you disagree, I would ask you to pick whatever metrics you want – your favorite test, racial group, axis scheme, whatever – and tell me whether it really looks like the doubling-to-tripling of education costs during the relevant time period has been money well-spent. If not, then as happy as I am to debate details in the comments, it all seems like basically nitpicking.

Jonah Katz brings up a really complete analysis of increasing costs in higher education:

> The rising cost of higher education isn't quite so mysterious, at least for the last 10-15 years. The Delta Cost Project has put together some fairly comprehensive data about this. What you see across most categories of post-secondary institutions is that basically *everything* is becoming more expensive, but 'student 'life' and 'academic support' are rising fastest, followed by 'institutional support'. Student life is all of the bells and whistles (athletic centers, movie theaters, etc.) that colleges use to try to entice prospective students into paying huge amounts of money to enroll in their institutions, and I believe it also includes health and mental health services, which I would imagine have become exponentially more expensive over the past couple decades (this is probably unavoidable, because health costs are going up in general and universities are enrolling a far wider range of students with more mental and physical health issues who wouldn't have gone to college in the past). Academic support includes a mix of stuff that is crucial to the academic mission of a

> university (libraries, IT systems), stuff that is arguably not part of the core academic mission at all (Dean's Office personnel, museums), and stuff that is well intentioned but tends to be useless in practice (central offices for teaching and curriculum development). Institutional support is administration proper. Note that these data come from 2003-2013, so they don't capture the explosion in university administration that is generally agreed to have occurred from roughly the 1970s to 1990s. I've never been able to find categorized data that goes back that far, but I imagine the change in spending on administration during that period must have been astronomical. The cost of instruction is still the largest single category of expenditure, and accounts for the majority of absolute price increases, but proportionally it is not rising as fast as these other categories. Also, the NY times Op-ed piece you link to is either selectively pulling misleading data or is just plain ignorant about the state of public financing for higher ed. There has not been a 'modest' reduction in per-student funding: it has dropped around 30% in inflation-adjusted dollars since 2000.

Static brings up the role of pensions as a driver of schooling cost increases.

Swing finds there have not been similar cost increases in the Netherlands. And Politifact rates as true a similar claim – that we greatly outspend other First World countries in per pupil spending. This article notes that we spend about 25% more than Britain and almost 50% more than Germany. On the other hand, the Netherlands is only a little better than we are, so this doesn't match a scenario where the Netherlands' spending goes up by only a little but America's goes up by 200%. I don't know where the discrepancy comes from. {EDIT: Douglas Knight points out that most countries have per pupil spending as a similar percent GDP)

Various people chime in with their favorite anecdotes about school vouchers working very well (DC) or working very poorly (Sweden). Murphy describes a personal bad experience with school privatization. At some point I do want to go through and sum up all the empirical literature on this, but not now.

Many people mention the possibility for bad incentives or market failures in schooling. 1soru1 thinks that, absent better signals of quality, schools will compete on shininess and raise prices to have the biggest and most breathtaking stadium (which I think is what the post above was saying happened to colleges, so certainly plausible). Tanagrabeast describes finding the private schools in Arizona heavily politicized: "What Scott worries about is already happening. I was skeptical until I took my son to an open-house at a fast-growing chain of charters where they tried very hard to play conservative buzzword Bingo and did all but lead us in a prayer to the Founding Fathers." In contrast, Doctor Mist says that as a rightist, he feels like going to a private school lets him *escape* what he sees as public schools' existing liberal politicization.

EarthSeaSky is a purist and reminds us that the free market which can be named is not the eternal free market.

Steve Sailer argues that for-profit colleges are a natural comparison group for for-profit primary schools, and they are very bad.

I talked to Education Realist on Twitter. Their position is complicated but they recommend their posts The Fallacy At The Heart Of All Reform and Charters: The Center Won't

Hold as introductions/summaries. I am still not entirely clear on their position – the objection seems to be that successful charters succeed only by taking the best students who would get good test scores anywhere, then claiming charters raise test scores. *Obviously* charters are trying this, but every halfway-decent study on charter schools has tried to control for this possibility. Also, none of my points involved empirical claims that charter schools raise test scores, so I don't see why this discredits me in particular. They also note that US education is already pretty good both compared to other countries and compared to its own past, something else I agree is true and have never denied.

Levarkin brings up James Tooley's fascinating work on private schools for the poor in Third World countries.

And Justreggedthis on the subreddit makes what I find the most convincing argument in this whole discussion:

> Sweden's experiment with school vouchers showed a different problem: the market delivers what you want, not what you need. What (stupid) parents want is good grades. What kids need is good education. So precdictably, voucher schools ended up diluting grades. You can probably imagine how it works. We live in an age of narcissism. Many parents want to hear their kids is super, special, and a genius, and get straight-A grades for a performance that is at best average. Few parents have the character left to stand up to it, and want challenging education and honest grading.
>
> Of course this is a problem with people, not vouchers. I am sure the very same narcissism in modern culture also rears its head in public schools as well.

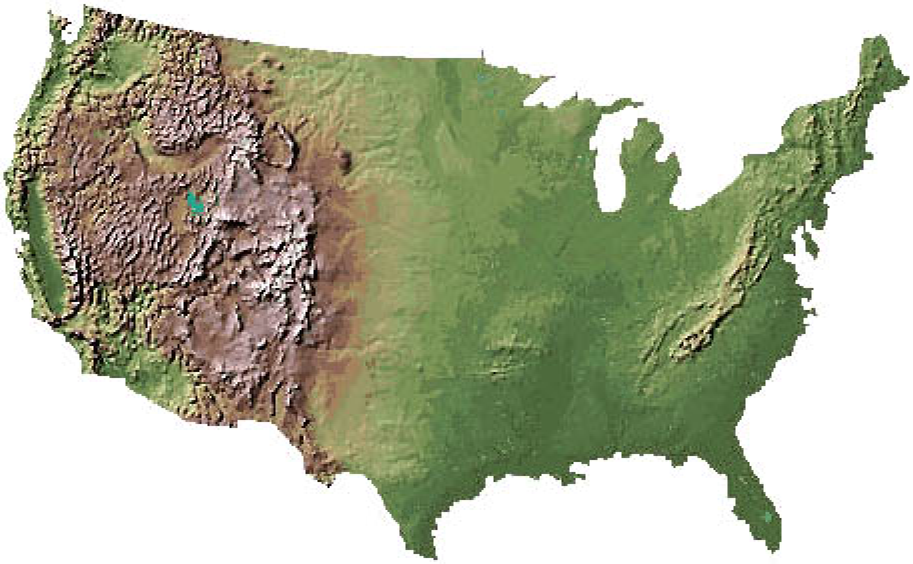
> The classic solution was school principals having low time preference and interested in preserving the long-term good name of their school. So they would not agree to grade dilution, they would not encheapen the brand of their school.
>
> Seems like today time preferences are high.
>
> Grocery stores are a good parallel. You need healthy food. You want (a stupider version of you wants) gallon buckets of ice cream. Hence, you get all kinds of special offers and discounts on tasty and cheap gallon buckets of ice cream. The market delivers what you want. Hence, obesity epidemic.
>
> I don't really know any solution that is acceptable within a democratic framework. Obvious someone somewhere should override personal preferences, but that someone should have a very good set of incentives and that is what we don't get in this framework.

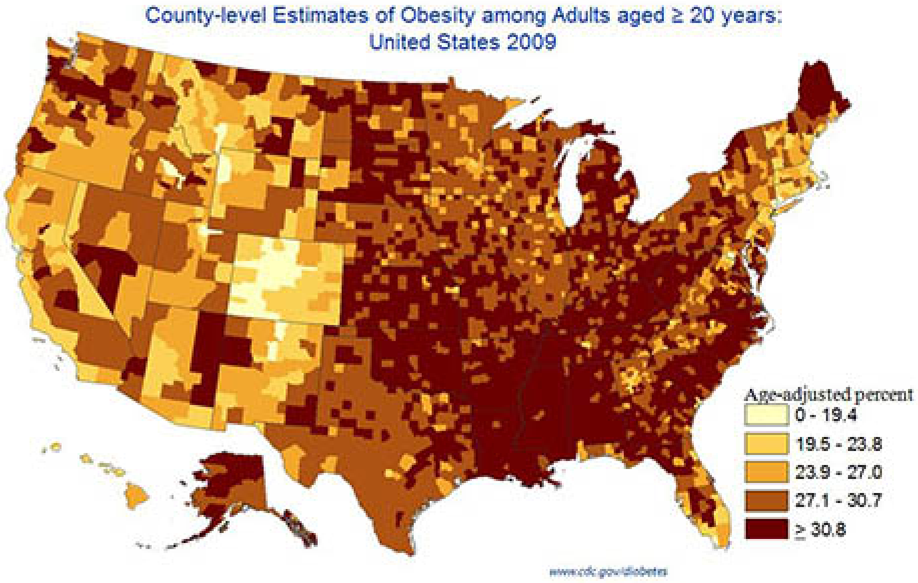
Overall reading this has made me somewhat more pessimistic about charter schools. But I'm still uncertain enough that I want to look into the empirical literature more, and I still think careful experimentation is the way to go.

…so maybe I should end with shadypirelli's comment from the subreddit pointing out that Betsy DeVos' policies cannot be described as "careful experimentation".

Thin Air

The International Journal of Obesity (h/t amaranththallium) points out a correspondence between US topography and US obesity rates:




County-level Estimates of Obesity among Adults aged ≥ 20 years:
United States 2009
Age-adjusted percent
0 - 19.4
19.5 - 23.8
23.9 - 27.0
27.1 - 30.7
≥ 30.8
www.cdc.gov/diabetes

It’s easy to see the Rocky Mountains on the obesity map. Not too hard to see the Appalachians either. Squint a little and you can even see California’s Central Valley vs. its coastal ranges.

This doesn’t seem to be related to poverty or population density. It does look a lot like the map of exercise level, but apparently it stays significant even when you control for that.

The IJO study finds that people living at sea level are five times more likely to be obese than people living at 500m elevation, even after controlling for “temperature, diet, physical activity, smoking, and demographic factors”. I don’t always trust controlling for things, but in this case the effect is big enough, and similar enough to the results of eyeballing, that it seems pretty plausible. Also, European studies find the same effect in high-altitude areas there, as do studies in Tibet. And someone did a study on US soldiers, who are randomly assigned (via deployment) to different areas, and found the same effect controlling for BMI at enlistment.

(on the other hand, a study in Saudi Arabia finds the opposite. Whatever. I didn’t even know Saudi Arabia had mountains.)

So what’s going on? There’s a well-known phenomenon called altitude anorexia where lowland people going to a high altitude suddenly can lose a lot of weight. Unfortunately most of the studies just stop at showing an acute effect; it’s not clear how long it lasts or whether there are more general principles involved. One study on rats found that they ate 58% less one day after being transported to Pike’s Peak, and were still eating 16% less per day two weeks afterwards. An article in High Altitude Medicine noted without further details that altitude anorexia seemed to persist after initial acclimatization. Pugh et al note weight loss of 1 kg/week up to 5-10 kg over a several week Everest ascent, reversing quickly as the climbers descended. A controlled experiment where obese subjects were ferried to the Swiss Alps, told to eat as much as they want, and banned from exercising resulted in weight loss of three pounds after a week, mostly sustained (?!) after a month at low altitude. It seemed mediated by eating less, which was independent of altitude sickness and persisted after people were no longer altitude-sick.

The active ingredient of altitude seems to be hypoxia. The air is thin at high altitudes so the body gets less oxygen. Being in low oxygen conditions in normal pressure seems to cause weight loss too – see here and here for studies of people exercising in low oxygen conditions. I don't know of any studies where people were just kept in low-oxygen environments for a long time without exercise to see what happened to their weight. It's not really clear how reduced oxygen makes people eat less. A lot of people mention leptin, but the studies seem pretty unconvincing, and people try to work leptin into everything.

This BMJ editorial suggests that hypoxia should get credit for smoking-related weight loss. But it reads like a completely unhinged screed from the tobacco lobby of an alternate dimension ("Might the aggressive anti-smoking lobby have contributed to the costly epidemic in obesity and type 2 diabetes that Professor Sir George Alberti have warned us about?") so maybe we shouldn't take it too seriously. Also, nicotine gum works just as well as cigarettes here, so it's probably an effect of the nicotine itself, and in fact we have some pretty good ideas how this happens. Maybe unhinged screeds by alternate-universe tobacco lobbies aren't the most trustworthy source of information.

Anyway, this is boring. Let's move on to a more interesting question – did global warming cause the obesity epidemic?

The arguments in favor: there's a lot more carbon dioxide in the air now than there was just a few decades ago. The obesity epidemic began around the time carbon dioxide concentrations really started getting worrying. And there are various body functions that are exquisitely dependent on CO2 levels. Bierworth (2014) has a good run-down of some of these and how they might be affected by increasing atmospheric CO2 levels (dear conservatives who always talk about Chesterton's Fence and principle of precaution – has it occurred to you that doubling the concentration of a major bioactive atmospheric gas might be a *bad* thing?). I see some conflicting claims about how much atmospheric CO2 could affect average blood pH. Neurons that produce obesity-regulating chemical orexin are potentially very sensitive to blood pH, so maybe this could be involved?

Wild animals are affected by the obesity epidemic too, even though they eat far fewer Big Macs. Even lab rats and zoo animals, supposedly kept on a well-monitored diet, are heavier now than they were decades ago. It's hard to think of some obesogenic factor so prevalent that it could seep into laboratories and zoos unnoticed by scientists and zookeepers. If it were a chemical, it would have to be really prevalent. The xenoestrogens in the water are one possiblity. But the other is the atmospheric gas breathed in by every living thing which we already *know* has been increasing for decades.

This at least is the theory of epidemiologist Lars George Hersoug. He did a study where he put some people in a high-CO2 room and found that they gained weight. It got a decent amount of press.

I really like this theory. It’s elegant. It’s clever. It’s at exactly the right level of contrarianism to be fun. If it were true, it would solve global climate change – once tabloids covers trumpet *THE ONE SECRET TO A TIGHT BELLY DOCTORS DON’T WANT YOU TO KNOW – TELL YOUR CONGRESSMAN TO PASS THE PARIS AGREEMENT TO LIMIT CLIMATE CHANGE TO WITHIN 2 DEGREES CENTIGRADE OF PRE-INDUSTRIAL LEVELS*, we will finally get Middle America on board. Insofar as scientific theories can be “fun”, this theory is fun.

But I grudgingly acknowledge that it’s probably not true.

For one thing, the study involved kind of sucks. It has a sample size of six. The six people were in a chamber that elevated CO2 levels about 50x higher than human industrial activity has elevated them in the atmosphere. And it got a non-significant result. In fact, food intake *decreased* in three of the six subjects, with almost all of the (nonsignificant) positive trend coming from one guy who apparently was just really hungry that day. There *is* a decent study showing CO2-related orexin effects in mice, but it’s at 2000x atmospheric concentrations. Also, when I look at the orexin neuron calculations, even if their hypothesis is true it suggests that orexin neurons might fire a little less than 1% more often now than they did 100 years ago. Unless there’s something really nonlinear going on, this is not enough to cause an epidemic.

For another, it doesn’t seem to line up geographically. Zheutlin, Adar, and Park try to correlate the geography of US obesity with the geography of US atmospheric CO2 in the same way that some of the studies above successfully correlated US obesity with US altitude. Here they fail. After adjusting for appropriate confounders, there is no clear relationship between CO2 levels and obesity. I think this might also reflect a more general point, which is that CO2 has been rising all over the world but obesity hasn't; Japan, for example, is a very high CO2 emitter but has almost totally avoided a US-level obesity crisis.

(but while we're talking about this study, it did find that serum bicarbonate has been increasing over the past decade or two. No proof as yet that this is a real effect or related to CO2, but have I mentioned that *increasing the concentration of a bioactive atmospheric gas worldwide is a really bad idea?*)

One last counterargument. A global warming skepticism blog points out that submarines are a natural laboratory for the effect of high CO2 on human health, since they usually have CO2 levels up to ten times atmospheric (and several times worse than even a poorly-ventilated building). I don't see any formal tests of their own argument, which is that submariners don't suffer any cognitive problems, and I'm not sure they're right to use intuition and failure to notice gross impairment – after all, the original studies showing impairment were done in office buildings, and there's no grossly noticeable differences there. But in any case, the Navy actually did a formal study and found that submariners do not gain weight. This seems pretty fatal for a CO2 = weight gain theory.

So my guess is that Hersoug is wrong and CO2 doesn't cause appreciable weight gain in normal concentrations. We should abandon the beautiful theory of climate-change-induced obesity and go down a level of contrarianism to blaming boring normal-person things like xenoestrogens and gut microbiota.

(I've heard there are theories of obesity even less contrarian than those, but I've never been to such low contaranianism levels and wouldn't be able to tell you what they might be.)

Might People On The Internet Sometimes Lie?

From Reddit: Parents Of Children Who Claim To Have Had Past Lives, What Did They Tell You?. Some sample comments:

> When he was 6 years old my son described in great detail my grandmother's house he never been to. This was in 1986 or so, pre-internet. There are no pics of the place that I'm aware and no one owned a camcorder in our family, so video is out of question either. It's a small house with red roof and a purple door (grandma painted the door every couple of years). He described all of it – that it had one big room with a fireplace across from the window, he explained where the doors are located, how there always were some boxes under the stairs, that there always was a faint smell of apples in the house (grandma ran a small time apple sauce business). That there was this cat almost completely white with a black spot around his right eye (that's mr. Whiskers, my grandma's cat!).
>
> My grandma and Whiskers both died in 1977, 3 years before my son was born. To this day I can't fathom it and can't even get a remotely sane explanation on how does he know all this. I never told him about it, my wife has never met my grandma and never been to her house and in 1986 we were stationed in Germany, so none of my old friends could have reached my son, so this is definitely not someone's prank. Best part of this is my son says he doesn't remember telling me that, but my wife heard him saying that too, so if definitely happened!

And from Reddit, What Is The Creepiest “Glitch In The Matrix” You’ve Encountered?:

> When I was in school I had this hippie teacher who would always tell us that the universe can help if you just ask it.
>
> She told us one time her daughter had lost something very important and when she asked the universe to help she suddenly had a massive pulling feeling towards the sink. She walks over and immediately stuck her hand down into the garbage disposal and pulled the item out in perfect condition.
>
> So I think it’s total bullshit of course, but later that day I was searching for a thin little booklet that I really, really needed for school. I spent 3 hours looking for it and had no luck. Finally out of frustration I almost sarcastically said, “I need your help universe.” I immediately walked over to this bookcase filled with books from my step dad. I had never once used this shelf or any book on it.
>
> I grab a random book I’ve never seen from the middle of a huge pile. I open it to somewhere around page 200 and right there is my booklet smashed in between the pages. It was incredibly thin so you couldn’t even tell there was anything in there if you looked at it from another angle.
>
> I’m sure there’s a good explanation, but it’s been well over a decade and I still remember the incredibly freaky vibe I got the moment I saw the book.

I don’t believe in reincarnation or paranormal forces. When I read stories like this, my first impulse is to try to think of reasonable explanations or ways they could be a coincidence. Maybe some kids have instinctive talent at that sort of cold-reading thing TV psychics do sometimes. Maybe your unconscious can remember where you put a booklet and then repress it from the conscious mind for some reason.

But these kinds of claims are often themselves far-fetched. If I told you in normal conversation, unrelated to compelling reincarnation theories, that kids have a natural talent at cold reaading, you'd scoff and demand proof. And it's not just reincarnation and booklet-finding. If you read Reddit enough, you'll find hundreds of equally compelling stories of telepathic contact, cryptid sightings, UFOs, et cetera.

So. Alternate hypothesis. About one million people view Reddit every day. Let's assume 10% of those see threads like the above – which were pretty popular and which I think both made it to the front page. That's 100,000 people. Now let's assume that even 1/10,000 people on the Internet are annoying trolls, which is maybe the easiest assumption we're ever going to have to make. If each of those annoying trolls posts one fake story to a thread like that for the lulz, that's enough for ten really convincing stories per thread – which is really all there are, the other fifty or sixty are just the usual friend-of-a-friend-had-a-vague-feeling stuff.

(it's true that in a site read by a million people, there will also be far more people who have experienced a genuine one-in-a-million coincidence, but that shouldn't scale nearly as quickly; after all, liars can invent coincidences way more far-fetched than the sheer numbers would allow)

This hypothesis seems obviously right. If I ask "what's the chance that at least one in ten thousand Internet users is an annoying troll?" you laugh hysterically and tell me that nobody has even invented numbers that high. It perfectly explains mysterious events that would otherwise require impossible coincidences or weird theories about hidden brain functions. So why is it so hard to make myself believe?

I think part of it is a failure of scale. Reddit looks a lot like a normal forum or blog comment section, the sort of BBS I used to go on as a kid with twenty or thirty regulars who would dominate all the discussions. If indeed 1/10,000 people is the sort of jerk who would make up a story like this just to troll people (or even 1/1,000 or 1/100 people), the chance that I'd run into them on my little BBS/comment section/Dunbar-number-group is pretty low, and I can safely ignore the possibility that five different crazy paranormal comments are all by pathological liars. It's only when you get a place like Reddit, which manages to feel like a community while also having a million readers a day, that you have to start thinking about these things.

This suggests a more general principle: interesting things should usually be lies. Let me give three examples.

I wrote in Toxoplasma of Rage about how even when people crusade against real evils, the particular stories they focus on tend to be false disproportionately often. Why? Because the thousands of true stories all have some subtleties or complicating factors, whereas liars are free to make up things which exactly perfectly fit the narrative. Given thousands of stories to choose from, the ones that bubble to the top will probably be the lies, just like on Reddit.

Every time I do a links post, even when I am very careful to double- and triple- check everything, and to only link to trustworthy sources in the mainstream media, a couple of my links end up being wrong. I'm selecting for surprising-if-true stories, but there's only one way to get surprising-if-true stories that isn't surprising, and given an entire Internet to choose from, many of the stories involved will be false.

And then there's bad science. I can't remember where I first saw this, so I can't give credit, but somebody argued that the problem with non-replicable science isn't just publication bias or p-hacking. It's that some people will be sloppy, biased, or just stumble through bad luck upon a seemingly-good methodology that actually produces lots of false positives, and that *almost all interesting results will come from these people*. They're the equivalent of Reddit liars – if there are enough of them, then all of the top comments will be theirs, since they're able to come up with much more interesting stuff than the truth-tellers. In fields where sloppiness is easy, the truth-tellers will be gradually driven out, appearing to be incompetent since they can't even replicate the most basic findings of the field, let alone advance it in any way. The sloppy people will survive to train the next generation of PhD students, and you'll end up with a stable equilibrium.

The weird thing is, I know all of this. I know that if a community is big enough to include even a few liars, then absent a strong mechanism to stop them those lies should rise to the top. I know that pretty much all of our modern communities are super-Dunbar sized and ought to follow that principle.

And yet my System 1 *still* refuses to believe that the people in those Reddit threads are liars. It's actually kind of horrified at the thought, imagining them as their shoulders slump and they glumly say "Well, I guess I didn't really expect anyone to believe me". I want to say "No! I believe you! I know you had a weird experience and it must be hard for you, but these things happen, I'm sure you're a good person!"

If you're like me, and you want to respond to this post with "but how do you *know* that person didn't just experience a certain coincidence or weird psychological trick?”, then before you comment take a second to ask why the “they’re lying” theory is so hard to believe. And when you figure it out, tell me, because I really want to know.

SSC Journal Club: Mental Disorders As Networks

I.

Suppose you have sniffles, fatigue, muscle aches, and headache. You go to the doctor, who diagnoses you with influenza and gives you some Tamiflu.

There's some complicated statistics going on here. Your doctor has noticed some observable variables (sniffles, fatigue, etc) – and inferred the presence of an invisible latent variable (influenza). Then, instead of treating the symptoms with eg aspirin for the headache, she treats the latent variable itself, expecting its effects to disappear along with it.

Psychiatry tries to use the same model. You get some symptoms – depressed mood, insomnia, fatigue, feelings of worthlessness, suicidality. You go to the psychiatrist, who diagnoses you with depression and gives you an antidepressant.

The psychiatrist is implicitly assuming that the causal structure of her field matches the causal structure of better-understood diseases like influenza. Generations of psychiatrists have noticed that different symptoms all tend to show up together and follow a similar pattern, suggesting some kind of deep connection between them. So psychiatrists follow the influenza model and attribute this collection of linked symptoms to a latent variable called "depression".

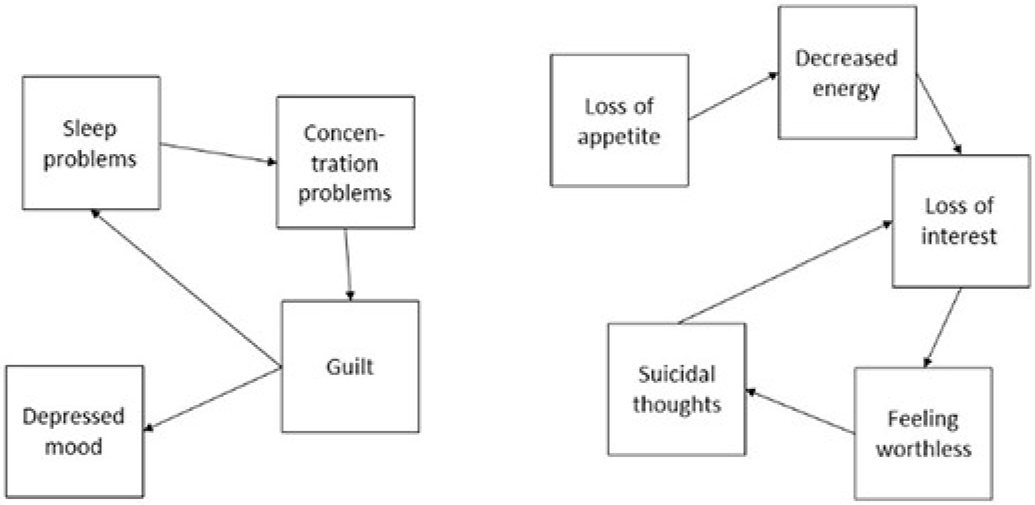
This gets complicated really fast. Psychiatric disorders are diagnosed through clusters of symptoms, but we don't expect every person to have every symptom in the cluster. For example, we diagnose depression when a patient has five out of nine symptoms on a list including fatigue, guilt, sleep disturbance, suicidality, et cetera. Each of these symptoms is often but not always present in a patient who has most of the others – for example, 75% of depressed patients have sleep disturbances, but 25% don't.

But all psychiatric disorders are hopelessly comorbid with each other. If someone meets criteria for one DSM disorder, there's a 50% chance they'll have another one too. 60% of people with major depression also have an anxiety disorder. This is awkward when compared to eg the 75% sleep disturbance rate. Why are we calling sleep disturbance a "symptom" of depression, but anxiety a "comorbid condition" with depression? If we're trying to cluster symptoms together to identify conditions, how come "sleep" is grouped with a bunch of other symptoms in the depression cluster, but "anxiety" gets to be a cluster of its own? Are there really two conditions called "depression" and "anxiety", or just one big condition that has various symptoms including low mood, sleep disturbance, and anxiety, and some people get some of the symptoms and other people get others? I'm told that the people who write the DSM have long conversations about this using rigorous methods, but to the rest of us it seems kind of arbitrary.

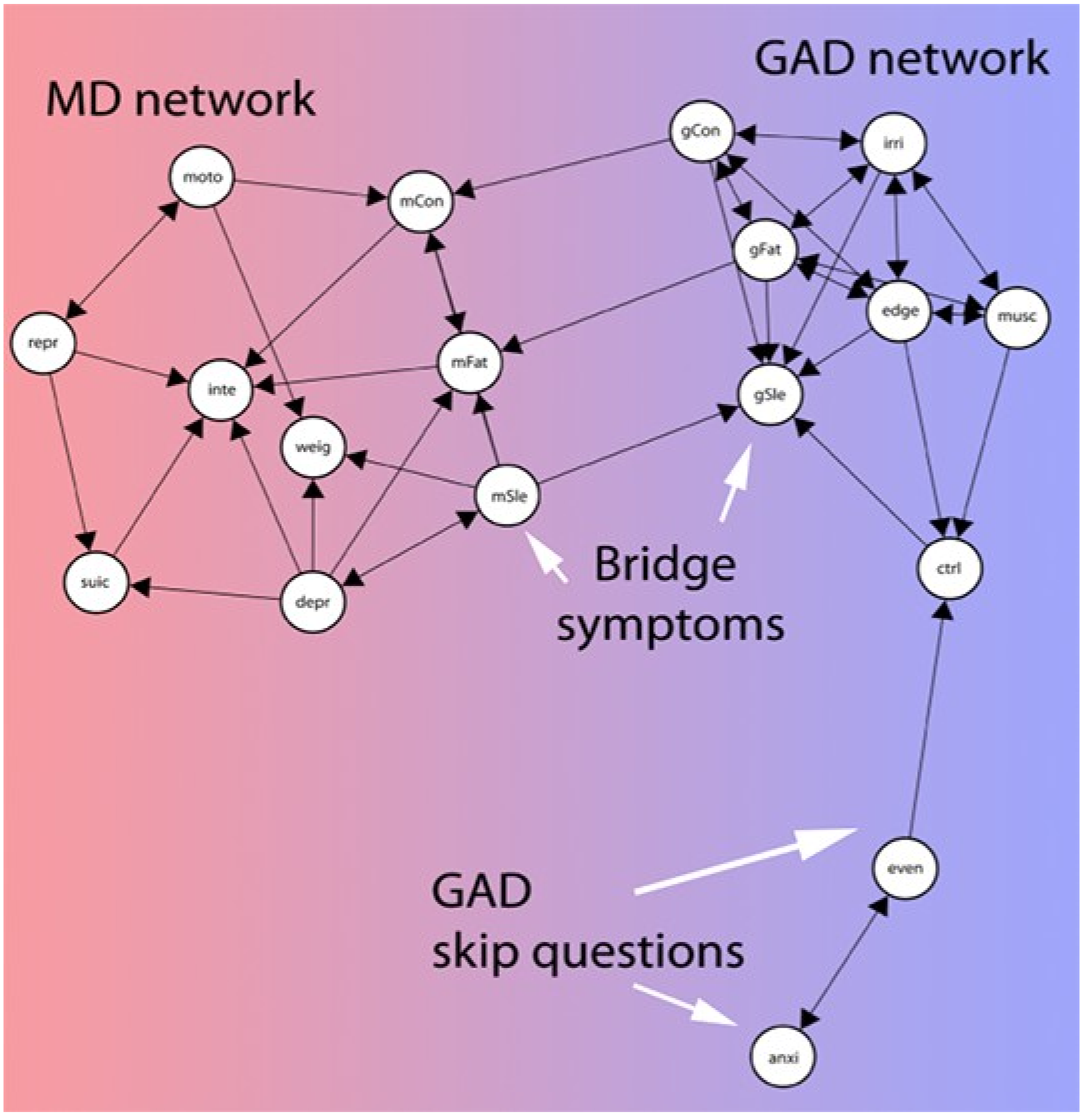
The problem isn't that nothing ever clusters together – depression, for example, is a very natural category. But so are various subtypes of depression. And so are various supertypes of depression, like depression + anxiety, or depression + psychosis, or depression + anxiety + psychosis. Choosing to draw the borders around depression and say "Yup, this is the Actual Disease" isn't a bad choice, but it doesn't jump out of the data either. When people try to use sophisticated clustering algorithms on psychiatric disorders, they usually come up with something like this, where there are only three supercategories instead of the 297 different diagnoses in the DSM. And even three supercategories are pushing it – people with psychosis are far more likely to have depression too! Having *any* number of categories starts seeming arbitrary and fuzzy.

So Nuijten, Deserno, Cramer, and Borsboom (from here on: NDCB) ask: what if that’s wrong? What if there *isn’t* a latent variable like “influenza”? What if it’s symptoms all the way down?

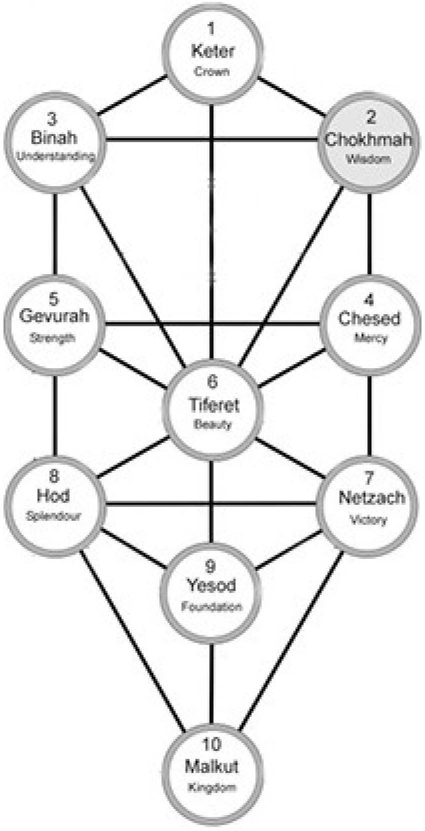
Consider a network in which each symptom is a node, connected to all the others by pathways with certain weights on each direction. So for example, “sleep disturbance” might be connected to “fatigue” by a strong path – people with disturbed sleep are much more likely to be tired. These might both be connected to “low mood” – people who don’t sleep well, or who are tired all the time, start feeling down about themselves. And this path might go the other way too: people who feel down about themselves might have more trouble getting to sleep on time. And maybe all of these are connected to suicidality, because if you feel bad about yourself you’re more likely to commit suicide, and if you’re suicidal you might feel bad about it, and if you’re tired all the time then maybe you can’t accomplish anything useful with your life and so death might seem like a good way out, and so on.


Sleep problems
Concen-tration problems
Guilt
Depressed mood
Loss of appetite
Decreased energy
Loss of interest
Suicidal thoughts
Feeling worthless

A sample image from the paper, showing two possible simple networks of depression symptoms


MD network
GAD network
moto
mCon
gCon
irri
repr
inte
mFat
gFat
edge
musc
weig
gSle
mSle
suic
depr
ctrl
Bridge symptoms
even
GAD skip questions
anxi

Also from the paper. This shows a more complicated (and apparently empirically validated) network of symptoms. MD is major depression. GAD is generalized anxiety disorder. The nodes are all different symptoms – for example, "inte" is "loss of interest in activities" and "musc" is "muscle tension".


1
Keter
Crown
3
Binah
Understanding
2
Chokhmah
Wisdom
5
Gevurah
Strength
4
Chesed
Mercy
6
Tiferet
Beauty
8
Hod
Splendour
7
Netzach
Victory
9
Yesod
Foundation
10
Malkut
Kingdom

Not from the paper. But if you figure out a good way to calculate weights on this one, email me.

Each node might affect the others with a certain delay. Being suicidal might make you feel guilty, but even if your last suicidal thought was fifteen minutes ago, you might still feel guilty now. Maybe it would take months or even years before you no longer felt guilty about your suicidal thoughts. So there could be loops: in a simple model, your low mood makes you feel suicidal, your suicidality makes you feel guilty, and your guilt makes you have low mood. This type of loopy network might be stable and self-reinforcing. Maybe your boss yells at you at work, which makes you have a bad mood. Then even if the direct effect of your boss would go away quickly, if it causes suicidal thoughts which cause guilt which cause more low mood, then the cycle can stick around forever.

In NDCB’s model, all possible psychiatric symptoms are connected like this in a loose network. Particularly tight-knit symptom clusters that often active together and reinforce each other correspond to the well-known and well-delineated psychiatric diseases, like depression and schizophrenia. But there are no natural boundaries in the network; low mood and poor sleep may be closely connected to each other, but they’ll also be more distantly connected to anxiety, and even more distantly connected to psychosis. This corresponds to the fact that some depressed people will develop psychotic symptoms, even though psychosis isn’t usually associated with depression. The paths aren’t usually as strong as those between low mood and poor sleep, but they’re there, and in some people with a predisposition to psychosis or some idiosyncratic factor strengthening those paths beyond their usual level in the population, that will be enough.

There are lots of good things about thinking about psychiatric problems this way:

1. It helps explain how life stressors can cause depression. Some people who have a bad breakup will get depressed. This should be mysterious if we think of depression as a biological illness – and we have to at least a little; some people who take the drug interferon-alpha will get depressed afterwards too. But if depression is a symptom network, it becomes easier to explain. The bad breakup causes low mood, which under the right conditions and genetic predispositions can activate all of the other depression symptoms and create a stable, self-reinforcing depression. Likewise, poor sleep is a risk factor for the development of subsequent depression, which is hard to explain if we just think of it as a symptom of some latent-variable-style condition.

2. It explains how treating depression symptoms can treat the depression. I’ve heard a lot of different perspectives on this, but at least one of my attendings (and some studies) believes that treating poor sleep with a sleeping pill like Ambien can help dispel an underlying depression, including symptoms seemingly unrelated to sleep like “feelings of worthlessness and guilt”.

3. It explains how therapy can treat depression. If eg cognitive behavioral therapy helps you stop thinking of yourself as worthless, then you’ve de-activated the “feelings of worthlessness and guilt” node and made it a lot harder for all the other nodes to coalesce into a stable self-reinforcing pattern.

4. It explains the polygenic structure of mental illnesses. If a mental illness were one specific thing, we would expect it to have one specific cause, or at least be limited to genes active in one specific area or process. In fact, it’s hard to come up with anything that genes involved in these illnesses have in common other than “they’re mostly expressed in the brain” – and sometimes not even that. In NDBC’s model, genes might be involved in any of the symptoms, or in the paths between the symptoms. A gene involved in poor sleep could predispose to depression. So could a gene involved in low energy levels. Even a gene involved in anxiety or psychosis could have *some* effect. And so would any gene that influenced the probability that, given poor sleep, a person would have low energy levels; or that given anxiety, a person will have psychosis. The end result would be everyone having a slightly different network, with different amounts of work needed to activate each node and different weights on each of the inter-nodal paths.

5. It helps explain why so many brilliant people searching for The One True Cause Of Depression have come up empty.

II.

Actually, this last one deserves more explanation. NDCB think of these symptoms as visible patient complaints ("poor sleep", "feelings of worthlessness"), and treat the connections between them as common sense ("if you don't sleep, you'll probably be fatigued", "if you feel very guilty, you might attempt suicide because you think you deserve to die"). But their theory also works for networks of biological dysfunctions, or networks that combine biological dysfunctions with common-sense observed symptoms.

For example, we know that there's a link between depression and inflammation. But it's not a very *good* link; not all depressed people have increased inflammation, not all people with increased inflammation get depressed, and drugs that decrease inflammation don't always cure depression. There's similarly good evidence linking depression to folate metabolism, serotonergic neurotransmission, BDNF levels, and so on. Suppose we made a graph like the ones above, except that instead of putting things like "poor sleep" and "feelings of guilt" on it, we used "inflammatory dysfunction", "folate metabolism dysfunction", "serotonin dysfunction", and "BDNF dysfunction". There are a lot of reasons to expect these things to interconnect – for example, folate helps produce a cofactor necessary for serotonin synthesis, so any dysfunction in folate metabolism could make a problem with serotonergic neurotransmission more likely.

In a best case scenario we could merge the biological and psychological perspective, replacing "disturbed sleep" with "disturbance in the orexin and histamine systems that regulate sleep" and "tiredness" with "disturbance in the dopamine system that regulates goal-directed action", and so "poor sleep makes you tired" with "disturbance in the orexin system causes a disturbance in the dopamine system”. In practice I expect this would be a terrible idea and that common-sense concepts mostly don’t have simple well-delineated biological equivalents. But what I’m saying is that the model where all of these things are observable symptoms, and the model where they’re all disturbances in brain chemicals and metabolism, aren’t necessarily in conflict.

So we can expand point (5) to say not only that it explains why nobody has found the One True Depression Cause, but why they *have* found so many promising leads that never quite pan out. Just like depression has a bunch of different symptoms, each of which is often-but-not-always involved, and each of which reinforces the others — so it has a bunch of different disturbances in biological systems, each of which is often-but-not-always involved, and each of which reinforces the others. Maybe there’s a nice correspondence between one disrupted biological system and one symptom, or maybe they sit uneasily together as different nodes on the same big graph.

III.

Are there any problems with this theory?

There are a couple of disorders that really don’t fit this model. Bipolar disorder, for example, doesn’t quite work as a collection of self-reinforcing symptoms. It’s marked by depressive episodes that can give way to years of stable mood before the person has a manic episode months or years later. I can’t think of any way to model this except as some underlying unified tendency toward bipolar disorder – although the ability for this tendency to cause a depression that looks just like normal unipolar depression is a point in

NDCB’s favor, since it suggests there can be many different causes for the same syndrome.

The impressive success of ketamine also counts as a point against. NDCB imagine psychiatric disorders like depression as gradually fading out on a symptom-by-symptom basis, eventually reaching a point where enough symptoms are gone that the rest of them aren’t self-reinforcing and just sputter out. This matches the course of eg SSRI treatment, where the medications will gradually improve a few symptoms at at time over the space of a month or so and maybe cause a full remission if you’re lucky. It doesn’t really match ketamine, where every aspect of depression vanishes instantly, then returns after a week or so without treatment. There are a couple of other equally impressive things – staying awake for thirty hours straight, for example, can have an immediate and near-miraculous antidepressant effect, which unfortunately vanishes as soon as you go to sleep. Both of these treatments seem like direct strikes against the One True Cause Of Depression, and both suggest that an underlying tendency toward depression can exist separate from any symptoms (or else why would the depression come back after the effects of the ketamine wore off?)

I don’t *think* it’s possible to cure depression by blasting every symptom simultaneously. That is, suppose somebody is depressed with symptoms of poor sleep, poor appetite, low energy, suicidality, and low mood. Ambien can make them sleep. Pot can make them eat. Adderall can give them energy. Clozaril can make them stop wanting to kill themselves. And heroin can perk up mood. So if you gave someone Ambien, pot, Adderall, Clozaril, and heroin at the same time, would that cure their depression? I’m pretty sure no one has ever tried this, but I don’t think anyone’s reported exceptional results from less extreme cocktails like

Adderall + trazodone + pot, which I’m sure a bunch of people end up taking. This along with the stuff from the last paragraph suggests that if we want to go with this model, maybe we should think less in terms of actual poor sleep and more in terms of dysfunction in the biological system of which sleep is a visible correlate. In that case we could say that Ambien helps the sleep itself but not the underlying dysfunction. But that takes some of the elegance out of the theory.

Despite these issues, I feel like something along these lines *has* to be true. There are too many things that *sort of kind of* cause psychiatric problems, and too few things that look like One True Causes. Things that look a lot like schizophrenia can be caused by viral infections in utero, by genetic factors, by hitting your head really hard as a child, by hypoxia during the birthing process, by something something something intestinal tract, by something relating to immigration which seems like it might involve psychosocial stress, and so on. Studies of the immune system, the dopamine system, the glutamate system, and the kynurenine system have all found disruptions. There have been so many really brilliant attempts to reduce all of these to a single brain region, or the levels of one specific chemical, or *something* that’s simple in the same way that lack-of-insulin-causes-diabetes is simple. But nobody’s ever succeeded. Maybe we should just give up.

I guess I’ve felt for a long time that some kind of weird change in attractor states of biological systems is the best way to explain these kinds of things, but I was never able to express what I meant coherently besides “weird change in attractor states of biological systems”. NDCB offer a clear model that suggests good avenues for future research.

(And I wasn’t joking when I said that little diagram with the two pentagons was the solution to 25% of extant philosophical problems.)

Book Review: Mount Misery

[Content warning: psychiatric abuse (especially around borderline personality), rape, spoilers for Mount Misery]

I.

Last month I reviewed Samuel Shem's *House of God*. The sequel, *Mount Misery*, is about his time training in psychiatry. This is obviously relevant to my interests, so I picked it up.

It's weird to accuse someone of writing a cheap knockoff of their own book, but *Mount Misery* reads like a cheap *House of God* knockoff. There are all the same elements – a young doctor, an incompetent system, cruel hospital administrators, a kind mentor. But in the first book, it all came together perfectly. In this one, it was more hit-and-miss. Sometimes the imagery clicked; other times, it just seemed like caricatures. Creepy magical realism alternated with guys who announced "I hate patients! Let's just pump them as full of drugs as possible and leave them to die!" and then zoomed away in fancy sports cars they bought with pharma money. Maybe it's just less funny when it hits closer to home? I don't know, but it *was* less funny.

The book's frame story follows Dr. Roy Basch, who has left his abusive medical internship to do a psychiatric residency at the Mount Misery hospital, lured there by the kind and decent Dr. Ike White. But Dr. White commits suicide Basch's first month of the job, the hospital administration ineptly covers it up, and nobody wants to talk about it – something something metaphor even psychiatrists stigma something metaphor. Having lost his mentor, Basch is thrown back and forth among various attendings – the one who thinks Freud solves everything, the one who thinks drugs solve everything, the one who thinks a thinly-veiled parody of Otto Kernberg solves everything, et cetera. The only sympathetic character is Dr. Malik, who tells Basch to ignore the theories and try to connect with his patients as human beings; Malik is of course loathed by all the other doctors and ostracized from all the good positions in the institution.

The typical psychiatric treatment in *Mount Misery* works as follows: Take someone who has some awful stuff going on in their life but is pretty much functional. Declare them to be a perfect example of whichever theory is popular that week ("this person clearly is obsessed with the idea of sucking their father's penis"), then insult any underlings who don't buy-in as being ignorami who refuse to understand the complexities of the human mind. Ignore all of the patient's human needs in favor of the theory – if they complain that their room is cold, tell them that's a reflection of the coldness of their father. If they protest that no, they're really cold, then mark them as "resistant" and double-down on your theory since they obviously need a lot of help. The angrier the patient gets, the more you've obviously hit a nerve and the better a psychiatrist you are. Repeat the process until they are curled up in a ball, completely nonfunctional, which you will call "successfully regressing the patient" and "revealing the repressed pathology". Then keep them in hospital until their insurance runs out, at which point discharge them to be someone else's problem.

When someone like Dr. Malik doesn't do the typical treatment, the fact that his patients never get reduced to curled-up balls gets held against him. He's so "superficial" that he just takes the patient's complaints about being cold as a reference to real physical coldness in the environment! If his patients say they're upset at losing their job, he's so "superficial" that he just talks to them about their job and how they can support themselves financially! A *janitor* could do that! When his patients fail to be reduced to curled-up balls of rage, that obviously proves he’s not nailing their real emotional problems, not getting past their defenses, not successfully regressing people, and just generally incompetent.

The master of this kind of anti-treatment is Dr. Heller – the hospital’s specialist in borderline personality disorder – who believes that it’s psychodynamically important to bring out the latent negative transference in borderlines (ie make them hate you). When Dr. Basch, at Dr. Malik’s urging, tries being nice to a borderline patient instead, and gets much better results than Heller ever has, the expert lectures him on his mistake:

> Heiler explained that his technique, “confrontation”, had evoked the anger that was hidden in each and every borderline. “She wasn’t angry at *me*“, he said, “it was her transference to me. She was distorting her real relationship with me based on early infantile experiences, with her bad mom, in the first year of life.”
>
> “But she *was* angry at you,” I said. “Anyone would be.”
>
> “Not that angry,” he said. “Not borderline angry.”
>
> “How do you know that she’s a borderline?”
>
> “Because of that incredible anger.”
>
> “But she didn’t start out angry – you provoked it.”
>
> “Who says?”
>
> “I do! It was obvious.”

"*You?* You, who've been a psychiatrist two whole months? You don't know diddly-squat about treating borderlines. Your so-called 'concern' is going to be a real problem – you're already overinvolved, imagining that you can rescue her. To you, what I did seemed cruel, right? […]

His voice softened, and he went on, "Look. I know that the first time you see it, this theory seems strange – it's counterintuitive. If just being nice to borderlines worked, don't you think I would do it? Of course I would! In fact, I tried, way back, at first. But it's like dealing with difficult children: you've got to be firm. Everybody knows that if you don't dig up the Latent Negative Transference in these gals, next thing you know you've got people killing themselves, or killing other people. For fifty years people have been trying to cure borderlines by being 'nice and human' to them. Everybody felt better, nobody *got* better. It's easy to act nice, it's hard as hell to stand firm and confront the rage locked up in borderlines. Borderlines are hell. There aren't too many of us left who have the guts to treat 'em. I've specialized in borderlines for years and years, and I've seen what works: You go through that rage to the truth, to their miserable pain and suffering, and believe me, they get better."

"But," Solini said, "I mean, everyone agrees that the lady [did get better when we were nice to her]"

"In this case," Blair said, "Better is worse. She'll have to get worse – which is in fact better – in order to get better, which will still be worse. If she gets a little worse, she won't get a lot better, but if she gets a lot worse, she may get a little better. Not smarmy-'nice' better. *Borderline* better…Don't worry, Roy. Your

> overinvolvement with her is normal. Sick, but normal. Gals like her are experts at getting guys like you entangled. Read my paper, *Rescue Fantasies In The Naive Resident*"

This speech could be a word-for-word transcription of something one of *my* attendings said to *me* during *my* intern year when *I* tried being nice to a borderline patient. There is a subtle sense in which this attitude can sometimes be helpful. But get the subtlety even slightly wrong and it devolves into being really evil, and *Mount Misery* brings out the worst in it.

Dr. Basch's first therapy patient is a man named Cherokee, a rich WASP lawyer. He's obsessed with the paranoid fantasy that his wife is having an affair with her psychoanalyst, a Mount Misery luminary named Dr. Dove. Basch tries everything with Cherokee – drugging him up, uncovering his latent homosexuality, suggesting he hates his father – but eventually Cherokee commits suicide anyway ("eventually he commits suicide anyway" will be a common theme among characters in this book.) In the aftermath, it is discovered that – surprise! – his wife was having an affair with her psychoanalyst, and also it was kind of coercive and bordered on rape.

Basch starts an investigation and learns that this same Dr. Dove is molesting a bunch of his female patients, and various other tangentially related people for good measure. He tries to expose Dove, but Dove denies everything, and he's a bigwig who can get the administration to take his side.

The description of the ensuing investigation is beautifully done, precisely because it avoids some of the caricatures of the rest of the book. Dr. Dove isn't portrayed as an ogre grumbling about “lying whores” or whatever. He sounds to all the world like a caring psychoanalyst, who understands that his patients are fragile and that stress of discussing sexual fantasies in psychoanalysis can sometimes break out into the patient’s consciousness and cause them to behave as if those fantasies actually occur. Yet all of this just serves to make him creepier and more hate-able.

In a particularly sharp scene, Dove capitalizes on the occasion to team up with a colleague and offer workshops about how to protect yourself from false accusations of assault in psychiatry:

> The slide show ended. Dr. Shpitzer then made a heartfelt statement that patient-psychiatrist contact was absolutely off-limits. Touching the patient, but for a handshake, was off limits. A hug was totally out of bounds. Yet what was the psychiatrist to do when a female patient, maybe a borderline or dissociative or multiple, suddenly got up out of his chair and approached, intent on hugging him? Dr. Shpitzer asked Dr. Dove to demonstrate. Schlomo, ever the showman, popped to his feet.
>
> First Shptizer said he would show us all what not to do, and told Shlomo to go ahead. Playing the woman patient, Schlomo started toward Shptizer, arms forward. Shpitzer crouched in a martial-arts stance and with a scream _ HYAH! – karate-chopped Schlomo’s hands down…The discussion then centered on variants of this technique. Dr. Shpitzer passed out his brochure, describing his video course – “Six Quick Steps To Avoid The Pitfalls Of Risk” – which we could all buy for $399.95. This would allow us to pass out risk-management requirements for state relicensure as

> shrinks in the comfort and privacy of our very own homes.”

A psychiatrist actually rapes a patient, he doesn’t get punished because he’s a very important guy who’s friends with all the bigwigs, but everyone has to feel like they’re doing something, so they ban all normal human contact with patients, and also sell $399.95 courses that you can use to prove you’re compliant with patient protection regulations. This may be the best metaphor for life that I have ever heard.

(it doesn’t hurt that I’ve had to go through courses on whether it’s ever appropriate to hug patients, or that I once had to finagle my way out of attending a conference that was basically this guy’s $399.95 video lecture)

II.

One of the main themes of this book is that psychoanalysis makes people worse.

The book doesn’t claim that psychoanalysis isn’t effective. It treats it as powerful and worthy of respect. The book’s psychoanalysts are consistently able to tell weird facts about a person from just a glance, to strip them down to their deepest insecurities in minutes. It’s just that people who are healthy and decent going into psychoanalysis end up cracked and nasty coming out of it. A lot of the worst doctors at *Mount Misery* were decent people before they started getting analyzed themselves. Of course, it would have helped if their analyst wasn’t a sexual predator, but the book treats the process as dangerous even aside from that.

When Basch asks his supervisor Dr. Lowell what to do about the man who believes Dr. Dove is having an affair with his wife, Dr. Lowell describes the analytic technique:

> “But what about Schlomo fucking his wife? You don’t think it’s true?
>
> “There is no truth, there is only the individual perception of experience.”
>
> “Wait a minute. The truth is that I’m taller than you.”
>
> “That’s not the truth, that’s your transference to me.”
>
> “We can measure it. To see who in fact is taller.”
>
> “You think ‘taller’ can be measured?”
>
> I saw her point. She wasn’t only aware of the objective fact, she was also aware of the deeper meaning psychologically. “But I’m stuck,” I said. “I don’t know what to do.”
>
> “You have to go deeper into his obsession, find the deeper meaning, the roots of it in his childhood, his past.”
>
> This was exactly what Malik had warned me against doing. Suspicious, I asked: “How?”
>
> “If he talks feeling, you talk thought. If he talks thought, you talk feeling. If he talks past, you talk present. If he talks present, you talk past. You the doctor talk constantly about what he the patient doesn’t want to talk about. This is the analysis of the resistance. Then, when he starts distorting his relationship with you and calling you a sonofabitch for not talking about what he wants to talk about, then you do the analysis of the transference, telling him he’s treating you like his

> father, his mother, his aunt Sally, whatever. On a deeper level still, you can analyze the resistance to the transference, and the transference to the resistance. Not to mention the countertransference to each – but that's still way beyond you at this point."
>
> Finally I felt I was getting some concrete advice about what to do in therapy.

I highlighted this last line because this is seriously much more concrete and actionable advice than anyone has ever given me about psychodynamic therapy and I've been doing it for two years.

And a lot of this rings true. I remember one time one of my patients missed a session because his flight back from vacation was delayed. I told my supervisor this and he got angry with me, saying it was superficial to blame it on the flight instead of talking about which of my comments had triggered the patient and made him decide to miss his plane. I insisted that we'd had a perfectly good session the week before, that the delayed plane had just been a delayed plane, and me and my supervisor got angrier and angrier at each other for both missing what the other thought was the point. Finally I got on the Internet and managed to prove that my patient's plane really had been delayed to the point where it was impossible for him to have made my appointment, at which point my supervisor switched the discussion to why it was so important to me to believe that his plane had been delayed that I would do an Internet search about it, and whether I was trying to defend against the unbearable notion that my patient might ever voluntarily miss one of our sessions. My supervisor's treatment of whether planes ever get delayed seems a lot like Basch's supervisor's treatment of who's taller.

And I don’t think these people are literally so stupid that they don’t understand that there are objective rulers that tell objective height. Trying to steelman this school of psychoanalysis, it’s a sort of as-if game, the professional equivalent of Crowley’s demand that the adept swear an oath “to interpret all phenomena as a direct dealing of God with his soul”. It’s an enforced fast from object-level discussion, where you treat everything as significant as an assumption. My first guess was going to be that this is so that you minimize the Type II errors where you miss something that really *is* significant, but after thinking about it more I wonder if it’s just that this is a bizarre and unnatural mode of thought that can get you places that normal thought can’t, sort of the same way some people have revelations on LSD not because LSD itself is magic but because it’s so different from normal thought processes that it can uncover things that are otherwise hidden. This could also explain the Freudian obsession with dreams – it’s not that they necessarily mean anything, any more than my patient missing his flight meant something, it’s that they’re a good source of noise to start scrying into.

(another Freudian technique is free association, asking the patient to just say whatever first comes to mind. In *Mount Misery*, Basch’s patient says “porpoises”, but has no idea why – something had to come to mind, and a porpoise was the first thing to pop into his head. This is a lot more like my own experience with free association than the textbook cases of people suddenly coming up with repressed childhood memories or something)

But this method also reminds me of something else. This is Christopher Hitchens:

> “I think Hannah Arendt said that one of the great achievements of Stalinism was to replace all discussion

> involving arguments and evidence with the question of motive. If someone were to say, for example, that there are many people in the Soviet Union who don’t have enough to eat, it might make sense for them to respond, “It’s not our fault, it was the weather, a bad harvest or something.” Instead it’s always, “Why is this person saying this, and why are they saying it in such and such a magazine? It must be that this is part of a plan.”

The avoidance of object-level discussion in favor of meta-level discussion can get really nasty, really quickly. The book gives one example – if you psychoanalyze rape accusations (“what purpose is it serving in this person’s mental ecosystem to have them accuse their psychiatrist of rape right now?”) then you miss someone who is actually getting raped. This can be more insidious when complaints are less dramatic and less binary – I know a lot of psychiatrists who will respond to people saying their medication isn’t working (or is causing side effects), with analyzing their motives for wanting to piss off their psychiatrist or stay unhealthy. And finally, this is absolutely *fatal* to any kind of complicated social discussion – the thing where instead of debating someone else’s assertion, you bulverize what self-interest or privilege causes them to believe it.

Basch says:

> I breathed in the cleansing sorrow of the rain and stared back up at the castle, and I saw clearly how through psychoanalysis you could know every nook and cranny of yourself and have no idea how to be with anyone, the seeming dazzle of the self blinding you to the connections with others…and I knew then I had once been in touch with people, and that it wasn’t inevitable that we are always shouting across an unbridgeable gap, but rather that the gap was in Freud and monstrous

> fabrications like [Dr. Lowell] herself who followed after, bereft souls floating untethered in pools of self like lilies in sepsis, the gap was in them, not in the essence of humans, nor in the essence of the whole world.
>
> I stared up at the vigilant street lamp, the cone of glittering sleet in the winter night reaching toward me like a beacon, showing me as clearly as if it were the moment’s sun that the real perversion of Freud and analysis was to take the essence of something and reduce it to something eles – the present to the past, love to hate, joy to misery, life to death – and to do it under the guise of understanding and yet, let’s face it, all the while doing it to escape from what Malik kept saying life at heart actually is – being, without description of that being.

If the book is right about psychoanalysis being destructive, I wonder if this is why. Living on the object level is *really good*. That’s where all the problems are and generally where the solutions are. It’s a natural, healthy place to live.

III.

The last thing that really struck me about the book was its praise for Alcoholics Anonymous.

In the last chapter of the book, Dr. Malik is revealed to be a recovering alcoholic who relapses when he gets diagnosed with cancer. He ends up committed to his own hospital, where he is first pumped full of irrelevant drugs (of course), then subjected to random people telling him he is a bad person because only bad people would drink. All of this is finally contrasted with Alcoholics Anonymous, treated as a beautiful organization full of caring-yet-pragmatic people that tries to genuinely connect with people and give them what they need to stop drinking.

It's pretty popular to hate on Alcoholics Anonymous these days. And not without reason – I did a sort of literature review about them a while ago, and while they're no worse than any other treatment options, they aren't any better either. Their insistence on acknowledging a Higher Power pisses some people off; their insistence on how they are the only way and if you abandon them you'll just be a drunk for the rest of your life pisses off others. Yet there are some very smart and very compassionate people – apparently including *Mount Misery* author Samuel Shem – who absolutely *love* them. In fact, looking at the About The Author page in the back of the book, it looks like after finishing this novel, Shem wrote a hagiographical play about AA founders Bill W and Dr. Bob.

(I checked to see if Shem has ever been an alcoholic himself, sometimes a common feature of people who are really into AA, but it doesn't look like it.)

There's a link between AA and Shem's constant theme throughout his books – people healing through relationships and human connection. But I was especially interested to see this quote, from an article on his AA play:

> And what about God? At the time that the two of them met, neither one had much faith in a traditional, religious God. As Smith said, "I was forced to attend church four times a week. I vowed when I was free I would never darken the door of a church again—a vow I've kept, religiously, for forty-odd years." Wilson, too, had more or less given up on God. Both men had pragmatic reasons: they had tried prayer to God, and it didn't work to keep them sober. The key to their vision

> about “God” came from a man named Ebby Thatcher, an old friend of Wilson’s who said, “You don’t have to believe in God, you just have to admit that you’re not God. Use what you do believe in, whatever it is.”

Shem seems really into this. I can’t quite justify this from the text, but I get the feeling that he would even take this to the meta-level, something like “complaining about how AA is exclusionary because it requires you to acknowledge a Higher Power is a good sign that you haven’t completed the personal growth task that ‘acknowledging a Higher Power’ corresponds to.”

This was pretty close to what Dr. Basch decides is the essence of a good psychiatrist during his final-chapter epiphany: he realized that he had to get *outside himself*. The lesson is a little bit Buddhist, but it also ties in nicely to the condemnation of psychoanalysis – while he was being analyzed, he was focused on “his inner machinery”, focused on how everything he experienced was a reflection of his own mind and desires. The attitude Shem holds up as healthy is the exact opposite of that – being able to think about anything *except* your own problems and your own status, being able to connect to your patients because you’re experiencing them as human beings.

I’ve been trying to reread some of The Last Psychiatrist and better understand what he means by narcissism, something I haven’t been able to get a good feel for before. I think Shem’s idea of getting outside yourself and “admitting that you’re not God” is close to this, a kind of narcissism therapy, where you can work yourself out of narcissism which allows you to connect to your patients and maybe help them in the same way.

(it also sort of reminds me of C.S. Lewis)

It’s well-known in psychopharmacology that different drugs work on different people, for mysterious reasons. Prozac and Paxil are about equally good in general, but some people will hate Prozac and find Paxil a miracle drug, whereas other people will get better on Prozac and find Paxil does nothing. I wonder if there might be something similar for social interventions like Alcoholics Anonymous. Over the whole population, it won’t outperform any other form of rehab – but there will be a few people for whom it works miracles. Those people will go on to praise it to the skies in all kinds of books and plays and so on – not to mention starting the next generation of Alcoholics Anonymous groups – while everyone else watches bemusedly. Actually, now that I mention this it sounds obviously true and I’m not sure why I wasn’t thinking this way already.

IV.

One of the reasons the psychiatrists in *Mount Misery* are so bad *is* that they’re narcissistic, but it’s an understandable narcissism. Someone says “My mom just died”, and you say “I’m sorry for your loss” and let them talk about their memories of their mother? *Anyone* can do that! Why did they get borderline-tortured throughout their twenties and thirties getting a really prestigious psychiatry degree if they were just going to say “I’m sorry your mom died”? Being able to relate it all to wanting to suck your dad’s penis at least gives them some credibility for all their erudite Freud-knowledge and justifies their $200-an-hour fees. “I’m sorry for your loss” isn’t exactly $200-an-hour level insight.

But I don’t know if Shem has a good solution here. It doesn’t seem like he wants to destroy psychiatry as an institution – he is, after all, a Harvard psychiatry professor. But his fictional bigwigs are right. If all you do is be a decent human being and have one-to-one meaningful discussions with patients, then it doesn’t seem like there’s a point in having MDs for that.

I’ve had a lot of patients with this exact complaint – usually it’s about psychologists or therapists instead of psychiatrists. “She kept telling me to go to sessions, that she was going to help me, and all we did was talk about my problems. I could have had a friend do that. So eventually I just quit and I haven’t been back to see a therapist since. Bunch of quacks.” I hear this kind of thing almost every day. It’s a big fear of mine that somebody thinks it about me. Probably one reason I like psychopharmacology so much is that it makes me feel useful – prescribing imipramine correctly isn’t something that *just anybody* could do; my patients may or may not get better but at least they’re getting their money’s worth.

I don’t know if Shem thinks that well-trained psychiatrists have some kind of special ability to connect with other people. Based on how horrible every psychiatrist in his book is, plus his preference for self-help groups like AA, it doesn’t look like it. But his rejection of both official therapies and medication doesn’t leave him a lot of outs. Also, it seems pretty obvious even to him that a lot of conditions – like melancholic depression and schizophrenia – *don’t* just need a kind word and a smile, that they *are* really complex entities that need a lot of effort and probably a good biochemical understanding before you can do much to them.

If all that Shem is saying is that doing the complicated work of psychiatric treatment – therapies, medications, et cetera – has to be combined with actually caring about the patient and treating them like a human being, then fair enough. But the vitriol of his criticisms of therapy and medication make it hard to read that message. If he’s proposing something more radical, then I’m afraid I didn’t entirely get what it was.

Contra NYT On Economists On Education

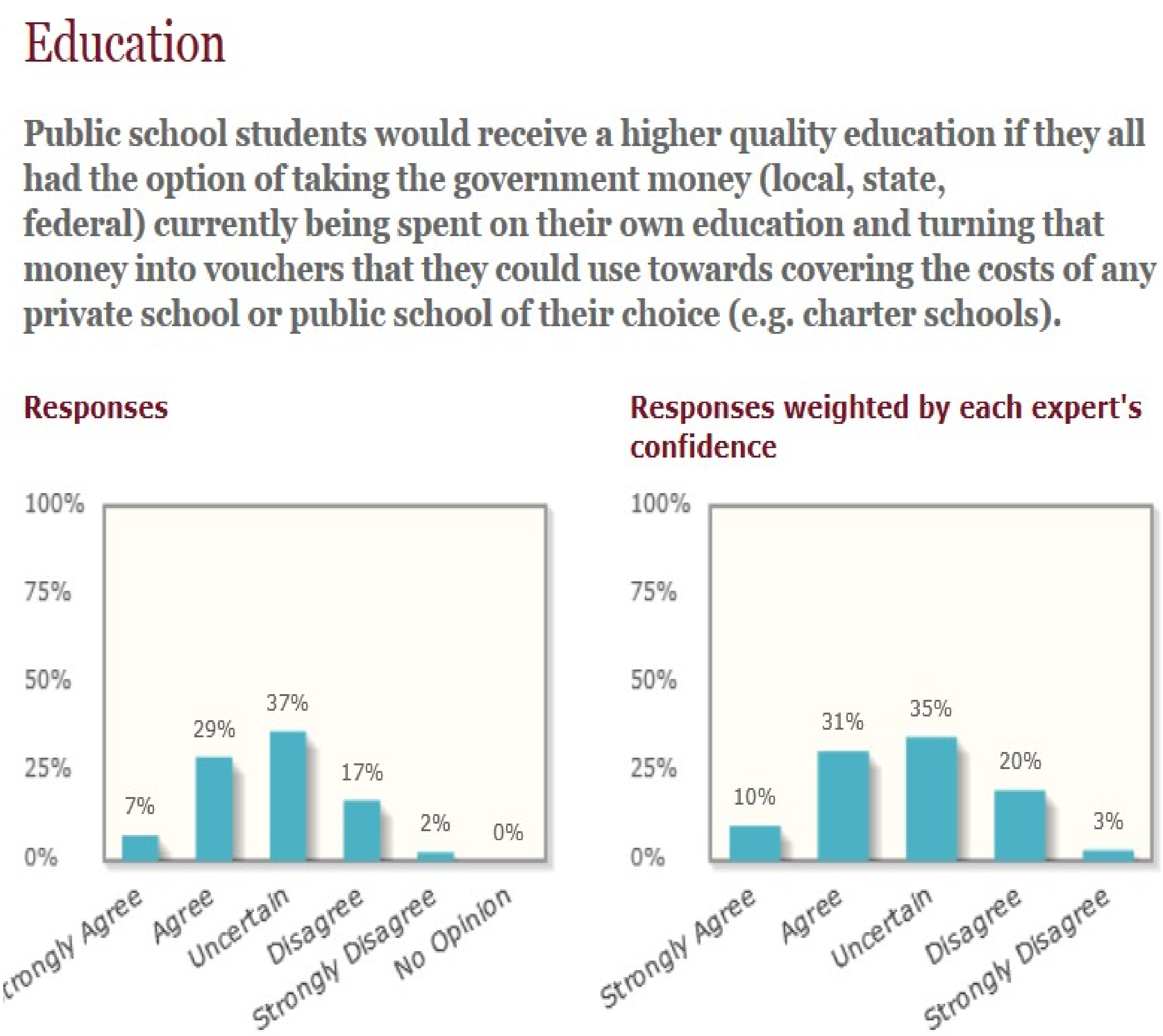
[epistemic status: still having a hard time believing I am right about this, but have double-checked. Tell me if I'm missing something. Correction: I previously left the word "Generally" out of the title by accident. I have slightly softened a claim about 'journalistic malpractice' based on many people apparently finding the phrasing less misleading than I do.]

From today's New York Times: Free Market For Education: Economists Generally Don't Buy It:

> The odds are good that privatizing education will be part of the agenda for President-elect Donald J. Trump's administration. […] You might think that most economists agree with this overall approach, because economists generally like free markets. For example, over 90 percent of the members of the University of Chicago's panel of leading economists thought that ride-hailing services like Uber and Lyft made consumers better off by providing competition for the highly regulated taxi industry.
>
> But economists are far less optimistic about what an unfettered market can achieve in education. Only a third of economists on the Chicago panel agreed that students would be better off if they all had access to vouchers to use at any private (or public) school of their choice.
>
> While economists are trained about the value of free markets, they are also trained to spot when markets can't work alone and government intervention is required.

This is followed by a long discussion of market failures and externalities, with the implication that this is the sort of knowledge that economists are using to come to their anti-voucher views.

But look at the NYT’s source for its claim about economists:



Education

Public school students would receive a higher quality education if they all had the option of taking the government money (local, state, federal) currently being spent on their own education and turning that money into vouchers that they could use towards covering the costs of any private school or public school of their choice (e.g. charter schools).

36% of economists agree that vouchers would improve education, compared to 19% who disagree. The rest are unsure or didn’t answer the question. The picture looks about the same when weighted by the economists’ confidence.

A more accurate way to summarize this graph is “About twice as many economists believe a voucher system would improve education as believe that it wouldn’t.”

By leaving it at “only a third of economists support vouchers”, the article implies that there is an economic consensus against the policy. Heck, it more than implies it – its title is “Free Market For Education: Economists Generally Don’t Buy It”. But its own source suggests that, of economists who have an opinion, a large majority are pro-voucher.

(note also that the options are only “agree that vouchers will improve education” and “disagree that vouchers will improve education”, so that it’s unclear from the data if *any* dissenting economists agree with the Times’ position that vouchers will make things *worse*. They might just think that things would stay the same.)

I think this is really poor journalistic practice and implies the opinion of the nation’s economists to be the opposite of what it really is. I hope the *Times* prints a correction.

[see follow-up post here]

Addendum To “Economists On Education”

A couple of people have challenged my essay yesterday, saying that they didn’t find the article misleading (or that it was only very slightly misleading).

They argue that it pointed out that economists were overwhelmingly in favor of ride-sharing arguments like Uber, but only somewhat in favor of school vouchers. Therefore, it’s fair to say that “economists don’t generally buy school vouchers” in the same sense that they “generally buy” Uber.

This would have been a fine thing to say, but I think the article failed to make this point and instead phrased its argument in a way that made it unlikely to readers to conclude anything other than that economists generally were not in favor of vouchers.

First, I feel like you could write exactly the opposite headline. “Public School: Economists Generally Don’t Buy It”. You would cite the statistic that “only 19% of economists surveyed disagree with the statement that school vouchers would improve education over the existing public system”. Then you could explain some public choice theory about how economists believe government services generally do poorly. This would be exactly as honest or misleading as the existing article. Yet it would produce exactly the opposite impression in readers’ minds: the original article makes you think economists mostly oppose school vouchers, the changed article makes you think they mostly support them. If newspapers are allowed to interpret data so loosely that they can use it to draw two different conclusion precisely opposite each other, what’s the point of having data?

Second, the article uses economists “not buying it” as a segue into a description of why economic theory says school choice could be a bad idea. But it seems like the majority of economists are not convinced by this argument. That is, both the 1/3 of economists who agree and the 1/3 of economists who are uncertain don’t accept the author’s argument that economic theory proves school vouchers can’t work as obviously true (the uncertain ones may be uncertain whether it’s true or not). This gives it a credibility it doesn’t deserve.

Third, really, if this same article was on Breitbart, and it used a survey showing that 40% of economists supported climate change legislation, 40% were uncertain, and 20% opposed it, and it described this as “Economists Generally Don’t Buy Climate Change Solution”, nobody would think anything untoward had happened and they would all agree this was a perfectly fair and unbiased summary of the evidence?

Fourth, you can do this for *anything* because there’s no clear definition of “uncertain”. How sure does an economist have to be before she “agrees” with a statement rather than being “uncertain” about it. If you are 51% sure school vouchers help, are you pro-voucher or uncertain? What about 60% sure? What about 90% sure? Suppose that all economists believe with 70% probability that vouchers will be good. If your criteria for “certain” is “80% or above”, then as long as you separate them out into support/uncertain/oppose, you can “honestly” declare that “no economists support school vouchers” and convince everyone that it must be an economically absurd plan. Yet I would hope that on any controversial issue more complicated than Uber, *most* economists are at least a little uncertain about it. We shouldn’t view that as legitimizing us to say that the economic consensus is whatever we want it to be.

Fifth and related, if you look at the economists' comments, a lot of the people who self-described as "uncertain" thought vouchers would be good on net, but didn't like the question because they thought it implied that literally 100% of students would be better off. This was such a problem that the IGM redid the study a year later, this time asking whether vouchers would make *most* students better off. 44% of economists agreed, compared to only 5% who disagreed (again, 34% were uncertain). Weighted by confidence, >50% of the economists agreed that vouchers would improve things, compared to only 6% who thought it wouldn't improve things – a difference of almost *ten times* more economists agreeing that vouchers would help compared to disagreeing!

In the face of all of this, the New York Times gives the field's opinion as "Free Market In Education: Economists Generally Don't Buy It".

I'm trying to be more empirical these days, so if you disagree with me, let me make you a bet – and I seriously mean I'll bet money on this if anyone wants to take me up on it. We find ten random people of ordinary intelligence and economics knowledge and show them this article. Then we ask them a question like "according to this article, what is the economic consensus on vouchers?" We assure them that this isn't a trick question and they're just supposed to honestly give the impression they get from the article. If they say something like "they're generally against vouchers", I win the bet. If they say something like "probably more economists support vouchers than oppose them, but many are uncertain", you win the bet. I am willing to alter exact terms if you have a better idea. I am willing to make this bet at 10:1 odds, so if you think there is any chance at all this article is not misleading it should be easy money for you.

~~[EDIT: I am going to take a version of this bet with Noah Smith. I don't need to take it with any more people. Offer is now closed.]~~

[EDIT 2: Noah Smith is apparently no longer willing to bet me on this, although he refuses loudly and at great length to say whether that means he now agrees with me that most people would get a false impression of economists' position from the article. Is there anyone else who wants to take a bet on these terms or any others?]

2016 Predictions: Calibration Results

At the beginning of every year, I make predictions. At the end of every year, I score them. Here are 2014 and 2015.

And here are the predictions I made for 2016. Strikethrough'd are false. Intact are true. Italicized are getting thrown out because I can't decide if they're true or not.

WORLD EVENTS
1. US will not get involved in any new major war with death toll of > 100 US soldiers: 60%
2. North Korea's government will survive the year without large civil war/revolt: 95%
3. Greece will not announce it's leaving the Euro: 95%
4. No terrorist attack in the USA will kill > 100 people: 90%
5. …in any First World country: 80%
6. Assad will remain President of Syria: 60%
7. Israel will not get in a large-scale war (ie >100 Israeli deaths) with any Arab state: 90%
8. No major intifada in Israel this year (ie > 250 Israeli deaths, but not in Cast Lead style war): 80%
9. No interesting progress with Gaza or peace negotiations in general this year: 90%
10. No Cast Lead style bombing/invasion of Gaza this year: 90%
11. Situation in Israel looks more worse than better: 70%
12. Syria's civil war will not end this year: 70%
13. ISIS will control less territory than it does right now: 90%
~~14. ISIS will not continue to exist as a state entity: 60%~~
15. No major civil war in Middle Eastern country not currently experiencing a major civil war: 90%
16. Libya to remain a mess: 80%
17. Ukraine will neither break into all-out war or get neatly

resolved: 80%

~~18. No country currently in Euro or EU announces plan to leave: 90%~~

19. No agreement reached on “two-speed EU”: 80%

20. Hillary Clinton will win the Democratic nomination: 95%

21. Donald Trump will win the Republican nomination: 60%

~~22. Conditional on Trump winning the Republican nomination, he impresses everyone how quickly he pivots towards wider acceptability: 70%~~

~~23. Conditional on Trump winning the Republican nomination, he’ll lose the general election: 80%~~

~~24. Conditional on Trump winning the Republican nomination, he’ll lose the general election worse than either McCain or Romney: 70%~~

25. Marco Rubio will not win the Republican nomination: 60%

26. Bloomberg will not run for President: 80%

~~27. Hillary Clinton will win the Presidency: 60%~~

28. Republicans will keep the House: 95%

29. Republicans will keep the Senate: 70%

30. Bitcoin will end the year higher than $500: 80%

~~31. Oil will end the year lower than $40 a barrel: 60%~~

32. Dow Jones will not fall > 10% this year: 70%

33. Shanghai index will not fall > 10% this year: 60%

34. No major revolt (greater than or equal to Tiananmen Square) against Chinese Communist Party: 95%

35. No major war in Asia (with >100 Chinese, Japanese, South Korean, and American deaths combined) over tiny stupid islands: 99%

36. No exchange of fire over tiny stupid islands: 90%

37. US GDP growth lower than in 2015: 60%

38. US unemployment to be lower at end of year than beginning: 50%

39. No announcement of genetically engineered human baby or credible plan for such: 90%

40. No major change in how the media treats social justice

issues from 2015: 70%
41. European far right makes modest but not spectacular gains: 80%
42. Mainstream European position at year's end is taking migrants was bad idea: 60%
43. Occupation of Oregon ranger station ends: 99%
44. So-called "Ferguson effect" continues and becomes harder to deny: 70%
~~45. SpaceX successfully launches a reused rocket: 50%~~
46. Nobody important changes their mind much about the EMDrive based on any information found in 2016: 80%
47. California's drought not officially declared over: 50%
48. No major earthquake (>100 deaths) in US: 99%
49. No major earthquake (>10000 deaths) in the world: 60%

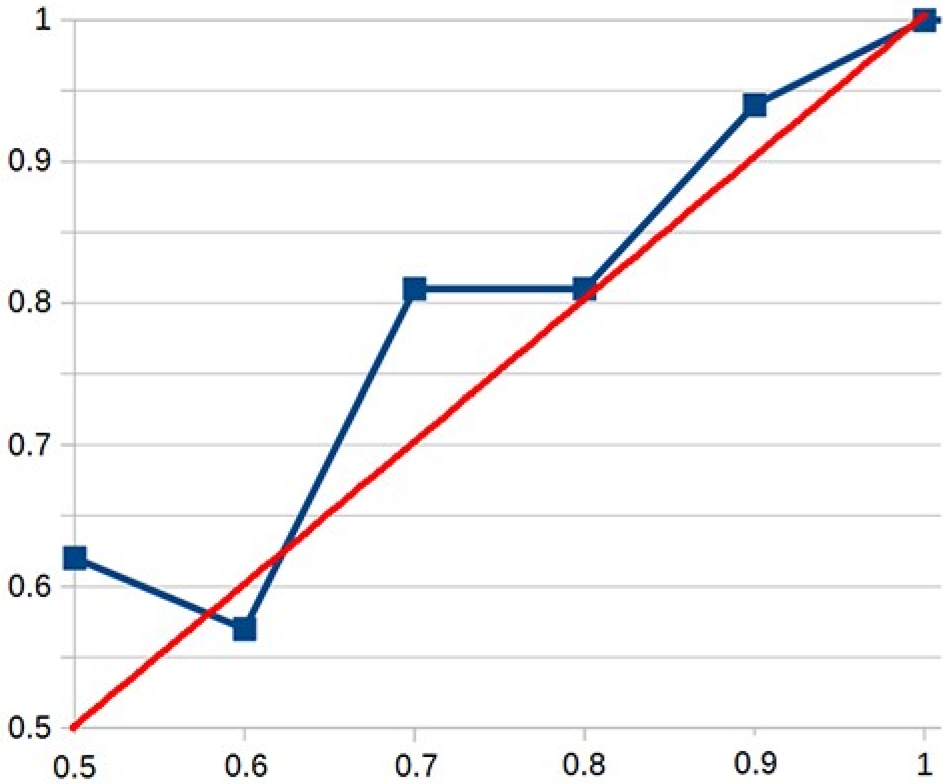
PERSONAL/COMMUNITY
1. SSC will remain active: 95%
~~2. SSC will get fewer hits than in 2015: 60%~~
3. At least one SSC post > 100,000 hits: 50%
4. UNSONG will get fewer hits than SSC in 2016: 90%
5. > 10 new permabans from SSC this year: 70%
5. UNSONG will get > 1,000,000 hits: 50%
6. UNSONG will not miss any updates: 50%
~~7. UNSONG will have higher Google Trends volume than HPMOR at the end of this year: 60%~~
8. UNSONG Reddit will not have higher average user activity than HPMOR Reddit at the end of this year: 60%
9. Shireroth will remain active: 70%
~~10. I will be involved in at least one published/accepted-to-publish research paper by the end of 2016: 50%~~
11. I won't stop using Twitter, Tumblr, or Facebook: 95%
12. > 10,000 Twitter followers by end of this year: 50%
13. I will not break up with any of my current girlfriends: 70%
~~14. I will not get any new girlfriends: 50%~~
15. I will attend at least one Solstice next year: 90%

~~16. …at least two Solstices: 70%~~
~~17. I will finish a long blog post review of stereotype threat this year: 60%~~
18. Conditional on finishing it, it won’t significantly change my position: 90%
~~19. I will finish a long FAQ this year: 60%~~
~~20. I will not have a post-residency job all lined up by the end of this year: 80%~~
21. I will have finished all the relevant parts of my California medical license application by the end of this year: 70%
22. I will no longer be living in my current house at the end of this year: 70%
23. I will still be at my current job: 95%
24. I will still not have gotten my elective surgery: 80%
25. I will not have been hospitalized (excluding ER) for any other reason: 95%
26. I will not have taken any international vacations with my family: 70%
27. I will not be taking any nootropic daily or near-daily during any 2-month period this year: 90%
~~28. I will complete an LW/SSC survey: 80%~~
29. I will complete a new nootropics survey: 80%
30. I will score 95th percentile or above in next year’s PRITE: 50%
31. I will not be Chief Resident next year: 60%
32. I will not have any inpatient rotations: 50%
33. I will continue doing outpatient at the current clinic: 90%
34. I will not have major car problems: 60%
35. I won’t publicly and drastically change highest-level political/religious/philosophical positions (eg become a Muslim or Republican): 90%
36. I will not vote in the 2016 primary: 70%
37. I will vote in the 2016 general election: 60%
38. Conditional on me voting and Hillary being on the ballot, I will vote for Hillary: 90%
39. I will not significantly change my mind about

psychodynamic or cognitive-behavioral therapy: 80%
40. I will not attend the APA meeting this year: 80%
41. I will not do any illegal drugs (besides gray-area nootropics) this year: 90%
42. I will not get drunk this year: 80%
~~43. Less Wrong will neither have shut down entirely nor undergone any successful renaissance/pivot by the end of this year: 60%~~
44. No co-bloggers (with more than 5 posts) on SSC by the end of this year: 80%
~~45. I get at least one article published on a major site like Huffington Post or Vox or New Statesman or something: 50%~~
46. I still plan to move to California when I'm done with residency: 90%
47. I don't manage to make it to my friend's wedding in Ireland: 60%
48. I don't attend any weddings this year: 50%
~~49. I decide to buy the car I am currently leasing: 60%~~
50. Except for the money I spend buying the car, I make my savings goal before July 2016: 90%

Of 50% predictions, I got 8 right and 5 wrong, for a score of 62%
Of 60% predictions, I got 12 right and 9 wrong, for a score of 57%
Of 70% predictions, I got 13 right and 3 wrong, for a score of 81%
Of 80% predictions, I got 13 right and 3 wrong, for a score of 81%
Of 90% predictions, I got 16 right and 1 wrong, for a score of 94%
For 95% predictions, I got 9 right and 0 wrong, for a score of 100%
For 99% predictions, I got 3 right and 0 wrong, for a score of 100%

This is the graph of my accuracy for this year:



Red is hypothetical perfect calibration, blue is my calibration. I am too lazy and bad at graphs to put in 95% right, but it doesn’t change the picture very much (especially because it’s impossible to get a very accurate 95% with 9 questions).

The 50% number is pretty meaningless, as many people have noted, so my main deviation was some underconfidence at 70%. This was probably meaningless in the context of this year’s numbers alone, but looking back at 2014 and 2015, I see a pretty similar picture. I am probably generally a little bit underconfident in medium probabilities (I have also gotten lazier about making graphs).

Overall I rate this year’s predictions a success. Predictions for 2017 coming soon.

2017

- Trump And The Batman Effect
- Predictions For 2017
- Heuristics Work Until They Don’t
- Why Do Test Scores Plateau?
- Should Buzzfeed Publish Claims Which Are Explosive If True But Not Yet Proven?
- Another Followup To “Economists On Education”
- Watch New Health Picks
- SSC Survey 2017
- Link: Guyenet On Taubes
- Dogs And Wolves: In Defense Of Some Past Posts
- Link: Taubes Contra Guyenet on Sugar
- Book Review: Eichmann In Jerusalem
- [GUEST POST] The International Refugee Assistance Project
- Notes From The Asilomar Conference On Beneficial AI
- Albion’s Seed, Genotyped
- Considerations On Cost Disease
- Highlights From The Comments On Cost Disease
- [REPOST] The Non-Libertarian FAQ
- Some Groups Of People Who May Not 100% Deserve Our Eternal Scorn

- SSC Journal Club: Analytical Thinking Style And Religion
- A Modern Myth
- Antidepressant Pharmacogenomics: Much More Than You Wanted To Know
- Book Review: Behavior – The Control Of Perception
- Book Review: Seeing Like A State
- SSC Survey 2017 Results
- [REPOST] The Demiurge's Older Brother
- Guided By The Beauty Of Our Weapons
- SSC Meetup: Everywhere
- G.K. Chesterton On AI Risk
- SSC Meetups Everywhere: Times And Places
- Determining Consent
- The Case Of The Suffocating Woman
- Yes, We Have Noticed The Skulls
- Sacred Principles As Exhaustible Resources
- Clarification To "Sacred Principles As Exhaustible Resources"
- Chametz
- Learning To Love Scientific Consensus
- SSC Journal Club: Childhood Trauma And Cognition
- Book Review: The Hungry Brain
- Anorexia And Metabolic Set Point
- Neutral vs. Conservative: The Eternal Struggle
- Getting High On Your Own Supply
- Silicon Valley: A Reality Check
- SSC Meetups: Berkeley, San Diego, Irvine
- Bail Out

- Polyamory Is Not Polygyny
- Postmarketing Surveillance Is Good And Normal
- Classified Ads Thread
- Those Modern Pathologies
- The Atomic Bomb Considered As Hungarian High School Science Fair Project
- Hungarian Education II: Four Nobel Truths
- Hungarian Education III: Mastering The Core Teachings Of The Budapestians
- Is Pharma Research Worse Than Chance?
- SSC Journal Club: AI Timelines
- What Is Depression, Anyway?: The Synapse Hypothesis
- Against Murderism
- To Understand Polarization, Understand Conservativism’s Failures
- Conversation Deliberately Skirts The Border Of Incomprehensibility
- Why Are Transgender People Immune To Optical Illusions?
- SSC Meetup: Chicago 7/2
- To The Great City!
- Two Kinds Of Caution
- Change Minds Or Drive Turnout?
- SSC Meetup: Salt Lake 7/14
- Can We Link Perception And Cognition?
- Classified Thread 2: Best In Classified
- Targeting Meritocracy
- Highlights From The Comment Thread On Meritocracy
- Don’t Blame Griggs

- Why Is Clozapine So Great?
- Djoser Joseph Osiris
- Against Signal-Boosting As Doxxing
- Book Review: Raise A Genius!
- Gender Imbalances Are Mostly Not Due To Offensive Attitudes
- Is It Possible To Have Coherent Principles Around Free Speech Norms?
- Where The Falling Einstein Meets The Rising Mouse
- Why Not More Excitement About Prediction Aggregation?
- Contra Grant On Exaggerated Differences
- The Lizard People Of Alpha Draconis 1 Decided To Build An Ansible
- Brief Cautionary Notes On Branded Combination Nootropics
- Fear And Loathing At Effective Altruism Global 2017
- Partial Credit
- Axiology, Morality, Law
- My IRB Nightmare
- Highlights From The Comments On My IRB Nightmare
- Book Review: Surfing Uncertainty
- Predictive Processing And Perceptual Control
- How Do We Get Breasts Out Of Bayes Theorem?
- Toward A Predictive Theory Of Depression
- Classified Thread 3: Semper Classifiedelis

- Book Review: Mastering The Core Teachings Of The Buddha
- Meditative States As Feedback Loops
- Against Individual IQ Worries
- Different Worlds
- SSC Survey Results On Trust
- In Favor Of Futurism Being About The Future
- SSC Journal Club: Serotonin Receptors
- SSC Meetup: Bay Area 10/14
- Were There Dark Ages?
- Highlights From The Comments On Dark Ages
- Kolmogorov Complicity And The Parable Of Lightning
- How Did New Atheism Fail So Miserably?
- Against Rat Park
- Postmodernism For Rationalists (my attempt)
- Non-Expert Explanation
- Concept-Shaped Holes Can Be Impossible To Notice
- Does Age Bring Wisdom?
- Ars Longa, Vita Brevis
- Book Review: Legal Systems Very Different From Ours
- List Of Passages I Highlighted In My Copy Of Legal Systems Very Different From Ours
- Contra Robinson On Public Food
- A Completely Accurate Map With No Distortion At All
- Book Review: Inadequate Equilibria
- Against Overgendering Harassment

- **The Tax Bill Compared To Other Very Expensive Things**
- **Response To Comments: The Tax Bill Is Still Very Bad**
- **Tax Bill 3: Don’t Mess With Taxes**
- **What To Make Of New Positive NSI-189 Results?**
- **Classified Thread 4: Vinson Classif**
- **Please Take The 2018 SSC Reader Survey**
- **Preregistration Of Hypotheses For The SSC Survey**
- **A History Of The Silmarils In The Fifth Age**
- **Adderall Risks: Much More Than You Wanted To Know**

Trump And The Batman Effect

Today on Trump Twitter:

> "@DanScavino: Ford to scrap Mexico plant, invest in Michigan due to Trump policies"https://t.co/137nUo03Gl
>
> — Donald J. Trump (@realDonaldTrump) January 3, 2017

Here's my concern.

When US companies do something that sounds good in the next few years, whether it's hiring new people, or deciding to stay in the United States, or reporting high profits, some of them are going to credit President Trump.

First, because it's going to get them good press. "Ford decides not to build plant in Mexico" is tenth-page news. "Ford decides not to build plant in Mexico *because of President Trump*" is front-page news.

But second, because it's going to make the President like them. I don't know whether Trump is secretly sending people to whatever conferences all of these people go to, saying "if you decide to do something good, give me credit, and I'll do you a favor later". I assume he isn't. This is the sort of thing that coordinates itself, without any inconvenient documents that can get posted to WikiLeaks later. If you're the CEO of Ford, and you notice you're doing something that would make Trump look really good if you attributed it to him, why *not* attribute it to him for free, then remind him how much he likes you next time you need a tax cut or a subsidy or something? Trump has put a lot of effort into crafting his image as a person who repays favors (think appointing many of his earliest supporters to Cabinet positions) – you think businesspeople aren’t going to notice that kind of thing?

But also:

> Big day on Thursday for Indiana and the great workers of that wonderful state.We will keep our companies and jobs in the U.S. Thanks Carrier
>
> — Donald J. Trump (@realDonaldTrump) November 30, 2016

0.1% of the time a US company does something that looks bad, like close a plant or move jobs overseas, Trump is going to launch a media crusade against them. The Presidency has a big pulpit and he’s going to get a lot of people angry. Then Trump will offer them some kind of deal, and the company will back down. Not because they’ve learned the error of their ways. Not even because the deal was so good. But because making the President (and the public) happy is much more important to them than moving jobs to Mexico or whatever they were doing before.

Mother Jones mentions in passing that Carrier air conditioning, Trump’s biggest job “success” so far, is owned by a giant defense contractor who gets probably like 1% of their profits from air conditioning. Presumably the company would be happy to never sell another air conditioner again if it meant that the government chooses their fighter jets over the competing brand. Knowing Trump’s style of corruption, they have every reason to believe this will happen after they handed him a big PR victory.

This plan isn’t going to scale. Even Trump can only create so many media circuses. 999 companies will successfully move to Mexico in the amount of time it takes Trump to convince one company not to. But almost tautologically, the only ones we'll ever hear about are the ones that become media circuses, and so it will look like Trump keeps winning.

So based on these two strategies, we are in for four years of sham Trump victories which look really convincing on a first glance. Every couple of weeks, until it gets boring, another company is going to say Trump convinced them to keep jobs in the United States. The total number of jobs saved this way will never be more than a tiny fraction of the jobs that could be saved by (eg) good economic policy, but nobody knows anything about economic policy and Trump will make sure everybody hears about Ford keeping jobs in the US. Every one of these victories will actively make the world worse, in the sense that these big companies will get taxpayer subsidies or favors they can call in later to distort government priorities, but nobody's going to notice these either.

I think it's important that we be prepared for this and send a clear message, before this gets any worse, that these aren't to be taken seriously.

I also think it's important to be prepared for the fact that this clear message *won't work*. Imagine you're a factory worker in Indiana, and every week you hear on the news that Trump convinced another factory to stay in the US. And also, you read an editorial by Paul Krugman or someone saying that this is all a trick. What do you end out believing?

And saving jobs isn't the only way he can do this. Trump's talent is PR, having his finger on the pulse of the media. He can spot things like that guy who raised the price of the toxoplasma drug 1000%, and then he can go in, make some corrupt deal, and get him to back down. He can spot all of those culture war things where the entire country is going to spend a month focused on the same small-town bakery, and by throwing around the entire might of the federal government he can probably make everyone back off and pose together for a nice group photo. If he can get all of these things right (and it will play exactly to his talents), then a majority of people won't care *what* policies his administration passes. I think this is a big part of his plan.

There's an old joke about Batman. Suppose you're a hypercompetent billionaire in a decaying city, and you want to do something about the crime problem. What's your best option? Maybe you could to donate money to law-enforcement, or after-school programs for at-risk teens, or urban renewal. Or you could urge your company full of engineering geniuses to invent new police tactics and better security systems. Or you could use your influence as a beloved celebrity to petition the government to pass laws which improve efficiency of the justice system.

Bruce Wayne decided to dress up in a bat costume and personally punch criminals. *And we love him for it.*

I worry that Trump's plan for his administration is to dress up in a President costume and personally punch people we don't like, while leaving policy to rot. And I worry it's going to work.

[prediction: highly-publicized stories about Trump successfully keeping businesses in the US on a case-by-case basis, which never add up to a significant number of jobs saved, will keep coming, and be a central point of how his administration relates to the public over the next year: 50%]

Predictions For 2017

At the beginning of every year, I make predictions. At the end of every year, I score them. So here are a hundred more for 2017.

WORLD EVENTS
1. US will not get involved in any new major war with death toll of > 100 US soldiers: 60%
2. North Korea's government will survive the year without large civil war/revolt: 95%
3. No terrorist attack in the USA will kill > 100 people: 90%
4. …in any First World country: 80%
5. Assad will remain President of Syria: 80%
6. Israel will not get in a large-scale war (ie >100 Israeli deaths) with any Arab state: 90%
7. No major intifada in Israel this year (ie > 250 Israeli deaths, but not in Cast Lead style war): 80%
8. No interesting progress with Gaza or peace negotiations in general this year: 90%
9. No Cast Lead style bombing/invasion of Gaza this year: 90%
10. Situation in Israel looks more worse than better: 70%
11. Syria's civil war will not end this year: 60%
12. ISIS will control less territory than it does right now: 90%
13. ISIS will not continue to exist as a state entity in Iraq/Syria: 50%
14. No major civil war in Middle Eastern country not currently experiencing a major civil war: 90%
15. Libya to remain a mess: 80%
16. Ukraine will neither break into all-out war or get neatly resolved: 80%
17. No major revolt (greater than or equal to Tiananmen Square) against Chinese Communist Party: 95%
18. No major war in Asia (with >100 Chinese, Japanese,

South Korean, and American deaths combined) over tiny stupid islands: 99%
19. No exchange of fire over tiny stupid islands: 90%
20. No announcement of genetically engineered human baby or credible plan for such: 90%
21. EMDrive is launched into space and testing is successfully begun: 70%
22. A significant number of skeptics will not become convinced EMDrive works: 80%
23. A significant number of believers will not become convinced EMDrive doesn’t work: 60%
24. No major earthquake (>100 deaths) in US: 99%
25. No major earthquake (>10000 deaths) in the world: 60%
26. Keith Ellison chosen as new DNC chair: 70%

EUROPE
27. No country currently in Euro or EU announces new plan to leave: 80%
28. France does not declare plan to leave EU: 95%
29. Germany does not declare plan to leave EU: 99%
30. No agreement reached on “two-speed EU”: 80%
31. The UK triggers Article 50: 90%
32. Marine Le Pen is not elected President of France: 60%
33. Angela Merkel is re-elected Chancellor of Germany: 60%
34. Theresa May remains PM of Britain: 80%
35. Fewer refugees admitted 2017 than 2016: 95%

ECONOMICS
36. Bitcoin will end the year higher than $1000: 60%
37. Oil will end the year higher than $50 a barrel: 60%
38. …but lower than $60 a barrel: 60%
39. Dow Jones will not fall > 10% this year: 50%
40. Shanghai index will not fall > 10% this year: 50%

TRUMP ADMINISTRATION
41. Donald Trump remains President at the end of 2017:

90%
42. No serious impeachment proceedings are active against Trump: 80%
43. Construction on Mexican border wall (beyond existing barriers) begins: 80%
44. Trump administration does not initiate extra prosecution of Hillary Clinton: 90%
45. US GDP growth lower than in 2016: 60%
46. US unemployment to be higher at end of year than beginning: 60%
47. US does not withdraw from large trade org like WTO or NAFTA: 90%
48. US does not publicly and explicitly disavow One China policy: 95%
49. No race riot killing > 5 people: 95%
50. US lifts at least half of existing sanctions on Russia: 70%
51. Donald Trump's approval rating at the end of 2017 is lower than fifty percent: 80%
52. …lower than forty percent: 60%

COMMUNITIES
53. SSC will remain active: 95%
54. SSC will get fewer hits than in 2016: 60%
55. At least one SSC post > 100,000 hits: 70%
56. I will complete an LW/SSC survey: 80%
57. I will finish a long FAQ this year: 60%
58. Shireroth will remain active: 70%
59. No co-bloggers (with more than 5 posts) on SSC by the end of this year: 80%
60. Less Wrong renaissance attempt will seem less (rather than more) successful by end of this year: 90%
61. > 15,000 Twitter followers by end of this year: 80%
62. I won't stop using Twitter, Tumblr, or Facebook: 90%
63. I will attend the Bay Area Solstice next year: 90%
64. …some other Solstice: 60%
65. …not the New York Solstice: 60%

WORK
66. I will take the job I am currently expecting to take: 90%
67. …at the time I am expecting to take it, without any delays: 80%
68. I will like the job and plan to continue doing it for a while: 70%
69. I will pass my Boards: 90%
70. I will be involved in at least one published/accepted-to-publish research paper by the end of 2017: 50%
71. I will present a research paper at the regional conference: 80%
72. I will attend the APA national meeting in San Diego: 90%
73. None of my outpatients to be hospitalized for psychiatric reasons during the first half of 2017: 50%
74. None of my outpatients to be involuntarily committed to psych hospital by me during the first half of 2017: 70%
75. None of my outpatients to attempt suicide during the first half of 2017: 90%
76. I will not have scored 95th percentile or above when I get this year's PRITE scores back: 60%

PERSONAL
77. Amazon will not harass me to get the $40,000 they gave me back: 80%
78. …or at least will not be successful: 90%
79. I will drive cross-country in 2017: 70%
80. I will travel outside the US in 2017: 70%
81. …to Europe: 50%
82. I will not officially break up with any of my current girlfriends: 60%
83. K will spend at least three months total in Michigan this year: 70%
84. I will get at least one new girlfriend: 70%
85. I will not get engaged: 90%
86. I will visit the Bay in May 2017: 60%
87. I will have moved to the Bay Area: 99%

88. I won’t live in Godric’s Hollow for at least two weeks continuous: 70%
89. I won’t live in Volterra for at least two weeks continuous: 70%
90. I won’t live in the Bailey for at least two weeks continuous: 95%
91. I won’t live in some other rationalist group home for at least two weeks continuous: 90%
92. I will be living in a house (incl group house) and not apartment building at the end of 2017: 60%
93. I will still not have gotten my elective surgery: 90%
94. I will not have been hospitalized (excluding ER) for any other reason: 95%
95. I will make my savings target at the end of 2017: 60%
96. I will not be taking any nootropic (except ZMA) daily or near-daily during any 2-month period this year: 90%
97. I won’t publicly and drastically change highest-level political/religious/philosophical positions (eg become a Muslim or Republican): 90%
98. I will not get drunk this year: 80%
99. I get at least one article published on a major site like Huffington Post or Vox or New Statesman or something: 50%
100. I attend at least one wedding this year: 50%
101. Still driving my current car at the end of 2017: 90%
102. Car is not stuck in shop for repairs for >1 day during 2017: 60%
103. I will use Lyft at least once in 2017: 60%
104. I weight > 185 pounds at the end of 2017: 60%
105. I weight < 195 pounds at the end of 2017: 70%

Heuristics Work Until They Don’t

I.

I got to talk to some AI researchers last week, and they emphasized how surprised everyone had been by recent progress in the field. They all agreed on *why* they were surprised: the “AI winters”, two (or so) past episodes when AI hype got out of control and led to an embarrassing failure to meet expectations. Eventually everyone learned the heuristic “AI progress will never be as fast as people expect”. Then AI progress went faster than expected, and everyone using the old heuristic was caught flat-footed, denying the evidence of their own eyes.

Is this surprising? It’s hard (and possibly meaningless) to segment the history of AI into distinct “eras”, but let’s try it just for fun: suppose that there were two past eras, both of which went worse than expected. If there are equal chances of an era meeting, exceeding, or missing expectations, then there’s a 22% chance that we either get two consecutive booms or two consecutive busts by pure coincidence. If we form a heuristic around this (“it’s always boom” or “it’s always bust”), then we’re interpreting noise and the future is likely to surprise us.

A quick and dirty Bayesian calculation: imagine three models. In Model A, researchers are biased towards optimism: 80% of the time, they will predict greater success than they actually attain, 10% of the time they will get it exactly right, and 10% of the time they will undershoot. In Model B, researchers are biased towards pessimism to the same degree. In Model C, researchers are unbiased and will overshoot, undershoot, and hit expectations with equal probability. Suppose we start with a 50% prior on Model C, and equal 25% probabilities for A and B. After observing one era of inflated expectations, we should have 52% chance A, 6% chance B, and 42% chance C. After observing two such eras, we should think 74% A, 1% B, and 25% C. Adding up the chances of all of the models, there's a 67% chance that the next era will also be one of inflated expectations, but there's a 33% chance it won't be.

This is all completely made up, plus my math is probably wrong. My point is that these kinds of "heuristics" gleaned from n = 2 data points are a lot less interesting than you would think. Getting fooled twice in the same way probably feels pretty convincing, and I can't blame the people involved for wanting to take a hard line against ever falling for it again. But their confidence that they're right should be pretty low.

II.

Thinking about this reminded me of an article from The Week, November 2012:

> Romney genuinely believed that he would become the nation's 45th president, and was "shellshocked" by his landslide loss. "I don't think there was one person who saw this coming," one senior adviser told Jan Crawford at CBS News. Why was Team Romney so certain of victory? They simply did not believe that younger voters and minorities would turn out the way they did in 2008. "As a result," says Crawford, "they believed that the public/media polls were skewed" in Obama's favor, and rejiggered them to show Romney with "turnout levels more favorable to Romney." In essence, Romney "unskewed" the polls, mirroring widely mocked moves by conservatives to show their candidate with a lead, epitomized by the now-infamous website

> UnskewedPolls.com. Romney's defenders say he had plausible reasons to believe Obama's turnout would be lower; less charitable commentators say Romney and his aides were stuck in a conservative media echo chamber at odds with reality.

Mitt Romney lost in exactly the way all the polls had predicted he would lose, but he wasn't expecting it because he had cheerfully constructed a story of decreased minority turnout which no real poll supported. This story became a kind of King Canute style warning of the folly of Man – just accept the fricking polls, don't come up with some private narrative about how decreased turnout will show up on a white horse and save you at the last second.

But we all know what happened in 2016. In retrospect, the fact that decreased minority turnout didn't happen in one election, with the most popular-among-minorities candidate of all time, shouldn't have been enough to form a strong heuristic that it would never happen at all.

This is even worse than the story above, because it's n = 1. I wonder if part of it is the degree to which Romney's loss formed a useful moral parable – the story of the arrogant fool who said that all the evidence against him was wrong, but got his comeuppance. Well, this last election taught us that arrogant fools don't get their comeuppance as consistently as we would like.

III.

Speaking of the 2016 election, I feel the same way about this explanation of Hillary's loss. It spins a narrative where the Hillary campaign management put all of their trust in flashy Big Data and ignored the grizzled campaign specialists who had boots on the ground, as if this was a moral lesson we should all take to heart.

But Moneyball makes the opposite argument. There, managers boldly decided to trust in statistics instead of just listening to the “intuitions” and “conventional wisdom” of professed experts, and they *trounced* the grizzled people with their ground-boots.

Anyone who learned the obvious lesson from Moneyball (“Hard math can defeat fallible human intuitions) would fail at the 2016 campaign, and anyone who learned the obvious lesson from the 2016 campaign (“Real experience and domain knowledge beat overeducated Big Data hotshots every time”) would fail at the 2003 baseball season.

The solution is: *stop treating life as a series of moral parables*. Once you get that, it all just becomes evidence – and then you wonder whether a single data point about Presidential campaigns necessarily generalizes to baseball or whatever.

IV.

If I’ve successfully convinced you that you shouldn’t form strong heuristics just by looking at a few salient examples where they seem to hold true, then shame on you.

Why Do Test Scores Plateau?

I just got my exam results, so let’s talk medical residency standardized test statistics. In particular, let’s talk about average results by year – that is, compare doctors in their first year of training, their second year of training, etc.

I found three datasets. One is for internal medicine residents over their three-year education. Another is for psychiatry residents over their four-year education. The last is for surgery residents over their five-year education. All of them are standardized to a mean of 500 and standard deviation of 100 for all years lumped together. Here’s how they look:

INTERNAL MEDICINE (numbers eyeballed from graph)
Y1: 425
Y2: 500
Y3: 550

PSYCHIATRY
Y1: 412
Y2: 485
Y3: 534
Y4: 547

SURGERY
Y1: 399
Y2: 493
Y3: 543
Y4: 565
Y5: 570

Year of education starts out as a relatively important factor relative to individual differences, but quickly becomes irrelevant. There’s only a 17% chance that a randomly chosen first-year surgeon will know more about surgery than an average second-year. But there’s a 31% chance a second-year will know more than a third-year, and a 48% chance that a fourth-year will know more than a fifth-year. Compare fourth-year and fifth-year surgeons, and it’s pretty close to 50-50 which of them will know more surgery.

(also, four percent of final-year surgeons about to graduate their training know less than a first-year trainee who just walked through the door. Enjoy thinking about that next time you get an operation)

It looks like people learn the most in their first year, and less every following year. Checking averages of all the programs together supports this:

Y1 – Y2: + 81 points
Y2 – Y3: + 50 points
Y3 – Y4: + 18 points
Y4 – Y5: + 5 points

The standardized nature of the scoring hides how minimal these gains are. The surgery exam report tells me the raw percent correct, which goes like this:

Y1: 62%
Y2: 70%
Y3: 75%
Y4: 76%
Y5: 77%

So these numbers eventually plateau. I don’t think any residency program has a sixth year, but if it did people probably wouldn’t learn very much in it. Why not?

Might it be a simple ceiling effect – ie there’s only so much medicine to learn, and once you learn it, you’re done? No. We see above that Y5 surgeons are only getting 76% of questions right, well below the test ceiling. My hard-copy score report give similar numbers for psychiatry. Also, individuals can do much better than yearly-averages. Some psychiatrists I know consistently score in the high 600s / low 700s every year. Why can't more years of education bring final-year residents closer to these high performers?

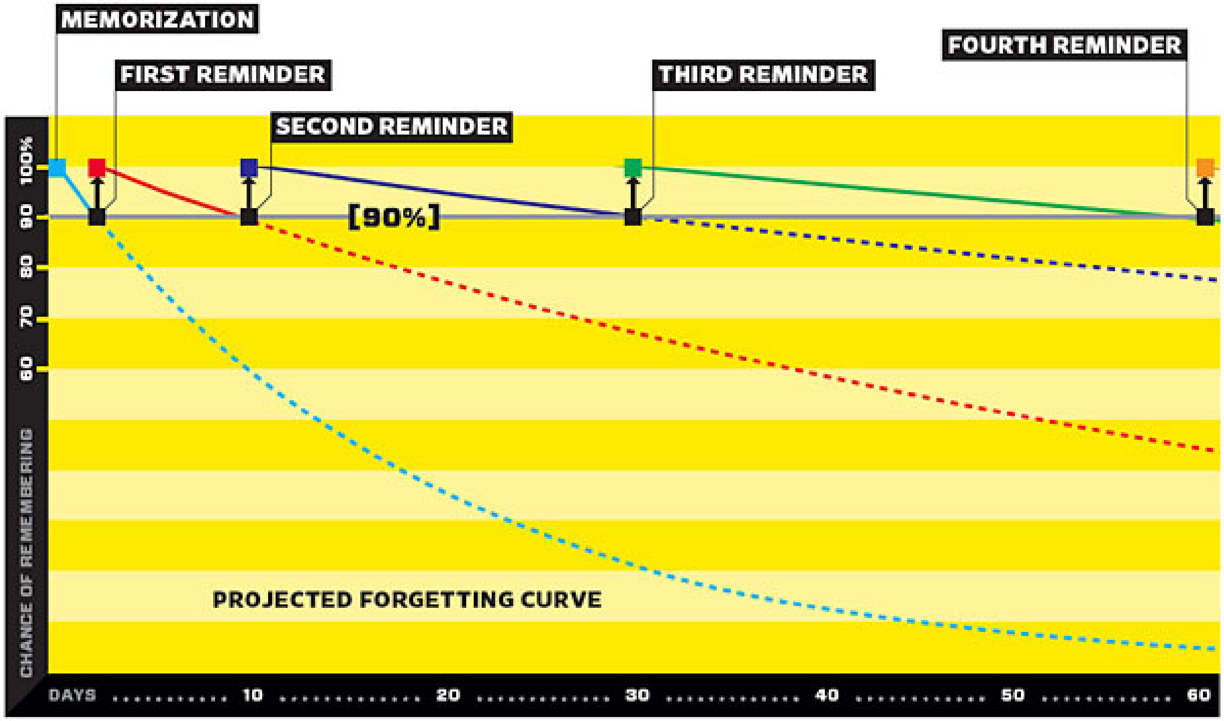
Might it be that final-year residents stop caring – the medical equivalent of senioritis? No. Score change per year seems similar across residencies regardless of how long the residencies are. For example, internal medicine residents gain 50 points in Year 3, about the same as psychiatrists and surgeons, even though internists finish that year, psychiatrists have one year left to go, and surgeons have two.

Might it be that programs stop teaching residents after three years or so, and they just focus on treating patients and not learning? That hasn't been my experience. In my own residency program, every year attends the same number of hours of lectures per week and gets assigned the same number of papers and presentations. I think this is pretty typical.

Might it be that residents only learn by seeing patients, and there's only a certain number of kinds of patients that you see regularly in an average hospital, so once you learn the kinds of cases you see, you're done? And then more book learning doesn't help at all? I think this is getting close, but it can't be the right answer. If you look at that surgery table again, you see relatively similar trajectories for the sorts of things you learn by seeing patients ("Patient Care", "Clinical Management") and the sorts of things you learn from books and lectures ("Medical Knowledge, Applied Science").

I don't have a good answer for what's going on. My gut feeling is that knowledge involves trees of complex facts branching off from personal experience and things that are constantly reinforced. Depending on an individual’s intelligence and interest in the topic, those trees can reach different depths before collapsing on themselves.

Spaced repetition programs like Anki talk about the forgetting curve, a model where memorized facts naturally decay after a certain amount of time until you remind yourself of them, after which they start decaying again (more slowly), and so on until by your nth repetition you’ll remember it for years or even the rest of your life.



Most people don’t use spaced repetition software, at least not consistently. For them, they’ll remember only those facts that get reinforced naturally. This is certainly true of doctors. By a weird turn of fate I earned a degree in obstetrics in 2012; five years later my knowledge of the subject has dwindled to a vague feeling that this comic isn’t completely accurate. On the other hand, I remember lots of facts about psychiatry and am optimistic about taking a board exam on the subject in September.

But most residents in a given specialty work about the same number of hours, see about the same sorts of patients – and yet still get very different scores on their exams. My guess is that individual differences in intelligence and interest affect things in two ways. First, some people probably have better memories than others, and can learn something if they go six months between reminders, whereas other people might forget it unless they get reminders every other month. Second, some people might be more intellectually curious than others, and so read a lot of journal articles that keep reminding them of things – whereas other people only think about them when it’s vital to the care of a patient they have right in front of them.

This still doesn’t feel right to me; I remember some things I’m not sure I *ever* get reminded about. Probably the degree to which you find something interesting matters a lot. And maybe there’s also a network effect, where when you think about any antidepressant (for example), it slightly reinforces and acts as a reminder about all your antidepressant-related knowledge, so that the degree to which everything you know is well-integrated and acts as a coherent whole matters a lot too.

Eventually you get to an equilibrium, where the amount of new knowledge you’re learning each day is the same as the amount of old knowledge you’re forgetting – and that’s your exam score. And maybe in medicine, given the amount of patient care and studying the average resident does each day, that takes three years or so.

Why does this matter? A while ago I looked at standardized test scores for schoolchildren by year, eg 1st-graders and 2nd-graders taking the same standardized test. The second-graders did noticeably better than the first graders; obviously 12th graders would do better still. But this unfairly combines the effects of extra education with the effects of an extra year of development. A twelfth-grader’s brain is more mature than a first-grader’s. Louis Benezet experimented with teaching children no math until seventh grade, after which it took only a few months’ instruction to get them to perform at a seventh grade level. It would sure be awkward if that was how everything worked.

Medical residency exams avoid this problem by testing doctors with (one hopes) fully mature brains. They find diminishing returns after only a few years. How much relevance does this have to ordinary education? I’m not sure.

Should Buzzfeed Publish Claims Which Are Explosive If True But Not Yet Proven?

Buzzfeed, January 14: A Mindset Revolution Sweeping Britain's Classrooms May Be Built On Shaky Science.

Somebody needed to write this article. It's written very well. I've talked to the writer, Tom Chivers, and he was very careful and seems like a great person. The article even quotes me, although I think if I had gotten to choose a quote of mine for thousands of people to see, it wouldn't have been the one speculating about Carol Dweck making a pact with the Devil.

But I'm not entirely on board with it.

Growth mindset has been really hyped and Carol Dweck has said it can do implausibly exciting things, okay. A lot of smart people are very suspicious of growth mindset and think there has to be some trick, sure. There's a high prior that something is up, definitely.

But one thing that needs to be at the core of any article like this is that, if there's a trick, *we haven't found it*.

I tried to be really clear about this in my own (mostly pessimistic) article on the subject:

> It is right smack in the middle of a bunch of fields that have all started seeming a little dubious recently. Most of the growth mindset experiments have used priming to get people in an effort-focused or an ability-focused state of mind, but recent priming experiments have famously failed to replicate and cast doubt on the entire

> field. And growth mindset has an obvious relationship to stereotype threat, which has also started seeming very shaky recently. So I have every reason to be both suspicious of and negatively disposed toward growth mindset.
>
> Which makes it appalling that the studies are **so damn good**.

This is the context of my speculation that Carol Dweck has made a pact with the Devil. I haven't accused (for example) the stereotype threat people of making a pact with the Devil. They did some crappy studies and exaggerated the results. That doesn't require any diabolic help. Any social scientist can do that, and most of them do. What's interesting about the growth mindset research is that it looks just like the sort of thing that should fall apart with a tiny gust of wind, but it actually hangs together pretty well.

BuzzFeed doesn't really challenge that. The article spends most of its time snarking about how overhyped growth mindset is – and no objections there, given that its advocates claim that it can eg help defuse the Israel-Palestine conflict and bring peace to the Middle East. It spends a bit more time talking about how many people are doubtful – no objections there either, I'm doubtful too.

But in terms of the evidence against it, it's kind of thin. I only see three real points:

First, it uses a technique called GRIM (granularity-related inconsistency of means). I like its explanation so I'm just going to quote it verbatim:

> t works like this: Imagine you have three children, and want to find how many siblings they have, on average. Finding an average, or mean, will always involve adding

> up the total number of siblings and dividing by the number of children – three. So the answer will always either be a whole number, or will end in .33 (a third) or .67 (two thirds). If there was a study that looked at three children and found they had, on average, 1.25 siblings, it would be wrong – because you can't get that answer from the mean of three whole numbers.

But Dweck says that she "took ambiguous answers as half scores" – maybe if the child was halfway between growth mindset and fixed mindset it was counted as a 0.5. It's bad practice to do this kind of thing without mentioning it. But everyone does some bad practices sometime. And I don't see anybody claiming it affected the results, which were very strong and not likely to stand or fall based on these sorts of things. Nobody is claiming fraud, and Dweck released her original data which looks pretty much like she was generally honest but had some bad reporting practice. Neither the statistician involved nor BuzzFeed claims this affects Dweck's work very much.

Second, it mentions Stuart Ritchie's criticism of a couple of recent Dweck papers which show "marginally significant" results. These results are so weak that they're probably coincidence, but the paper hypes them up. There are a couple of studies like this, but they're all in very tangential areas of mindsetology, like how children inherit their parents' mindsets. The original studies, again, show very strong results that don't need this kind of pleading. For example, the one I cited in my original post got seven different results at the $p < 0.001$ level. And there are a *lot* of studies like this.

Third, it mentions a psychologist Timothy Bates who has tried to replicate Dweck's experiments (at least) twice, and failed. This is the strongest evidence the article presents.

But I don’t think any of Bates’ failed replications have been published – or at least I couldn’t find them. Yet hundreds of studies that successfully demonstrate growth mindset *have* been published. Just as a million studies of a fake phenomenon will produce a few positive results, so a million replications of a real phenomenon will produce a few negative results. We have to look at the entire field and see the balance of negative and positive results. The last time I tried to do this, the only thing I could find was this meta-analysis of 113 studies which found a positive effect for growth mindset and relatively little publication bias in the field.

My intuition tells me not to believe this meta-analysis. But I think it’s really important to emphasize that I’m going off intuition. There’s no shame in defying the data when you think that’s justified, but you had better be really aware that’s what you’re doing.

I guess my concern is this: the Buzzfeed article sounds really convincing. But I could write an equally convincing article, with exactly the same structure, refuting eg global warming science. I would start by talking about how global warming is really hyped in the media (true!), that people are making various ridiculous claims about it (true!), interview a few scientists who doubt it (98% of climatologists believing it means 2% don’t), and cite two or three studies that fail to find it (98% of studies supporting it means 2% don’t). Then I would point out slight statistical irregularities in some of the key global warming papers, because *every* paper has slight statistical irregularities. Then I would talk about the replication crisis a lot.

I could do this with pretty much any theory I wanted. Any technique strong enough to disprove anything disproves nothing.

(and this is especially important in light of recent really strange negative results that eg fail to find a sunk cost effect, something I would hate to enshrine as "well, guess this has been debunked, no such thing as sunk cost now")

Again, this isn't to say I believe in growth mindset. I recently talked to a totally different professor who said he'd tried and failed to replicate some of the original growth mindset work (again, not yet published). But we should do this the right way and not let our intuitions leap ahead of the facts.

I worry that one day there's going to be some weird effect that actually *is* a bizarre miracle. Studies will confirm it again and again. And if we're not careful, we'll just say "Yeah, but replication crisis, also I heard a rumor that somebody failed to confirm it," and then forget about it. And then we'll miss our chance to bring peace to the Middle East just by doing a simple experimental manipulation on the Prime Minister of Israel.

I think it's good that people are starting to question growth mindset. But at this point questioning it isn't enough. In my essay I tried to find problems that might have caused spurious effects in Dweck's studies, and patterns inconsistent with growth mindset being powerful. I think we need to do more of that, plus look for specific statistical and experimental flaws in the papers supporting growth mindset, plus start collecting real published papers that fail to replicate growth mindset. Instead of talking about how sketchy it is, we need to actually disprove it.

We owe it to ourselves, to Carol Dweck, and to her infernal masters.

Another Followup To “Economists On Education”

Last month I argued that a news article misrepresented the feelings of economists on school vouchers. I got a bit of pushback from people who thought I had just misread it, and that it wasn’t deceptive at all (1, 2, 3), and wrote an addendum post sticking to my guns.

One of my New Years resolutions is to be more empirical, so I thought this would be a good opportunity for an experiment. I offered to bet at 10:1 odds in the other person’s favor that most people shown the article would believe economists were against school vouchers. Noah Smith originally took me up on the bet, but later stopped cooperating with me about it and wouldn’t give a straight answer to the question of whether he was retracting his offer. Vikram Sasi agreed to continue the bet; I’m not sure if he disagreed with my assessment, thought the odds were too good to refuse, or just wanted the experiment to go ahead.

I proposed that we show some random people the article, then ask them a few easy comprehension questions to make sure they were putting in a decent effort. Then we would ask the following:

> According to this article:
> a) Most economists support privatizing education
> b) Most economists oppose privatizing education
> c) Unsure / the article doesn’t say

The true answer was (c) – the article doesn’t give enough information to tell you what economists think (which turns out to be mostly unsure, with more supporting the privatization than opposing). But I thought the article heavily (and falsely) implied b, and expected most people to give that answer.

Before Vikram and I could agree on exactly how we were going to do this, we got scooped by two other people who did the experiment without asking us.

Commenter Sripada got 40 people to take the survey on MTurk, of whom 35 answered the easy questions right and qualified for inclusion. Of those 35, 32 of them (91%) answered (b), misinterpreting the article to claim a consensus against privatizing education. Only one of them (3%) gave the correct answer (c), accurately noting that the article doesn't tell us anything about an economic consensus for or against.

Hanne Watkins, a real research psychologist, also did a 50 person MTurk survey. I don't have enough access to her data to limit the sample to people who answered the easy questions right, but among all participants 78% believed that most economists opposed privatization, compared to only 10% who correctly noted that the article didn't say.

Vikram graciously accepted these studies as settling the bet, and since more than half of people misinterpreted the article in both, he paid me my $10.

I'm really happy about this, not just because I won ten dollars but because I feel like it was a rare example of getting to decisively settle a disagreement. Instead of arguing about whether or not the article was misleading, we figured out a way to test it, did the test, and now we know.

I should add that some people might think the article was non-misleading in a different way than I did that can't be settled empirically (Noah seems to be in this group), and other people might think the particular set of questions the survey asked was unfair or biased. This is why I wish I'd gotten to bet somebody who I knew really disagreed with me about this, so at least *that* disagreement could have been settled for sure. As a consolation prize, I will take my ten dollars and my increased certainty that the article didn't accurately convey economists' thoughts.

Thanks to Vikram, Sripada, and Hanne for their help with this.

Watch New Health Picks

So far most of Trump's appointments have been ordinary conservative hardliners or ethically-compromised rich people. But there's a chance that some of his health care picks could be *really interesting*.

I'm not talking about Health and Human Services nominee Tom Price. As far as I can tell he's an ordinary conservative hardliner (not to mention an orthopaedic surgeon), and pretty par for the course.

But I've seen three names mentioned as top candidates for FDA commissioner – Scott Gottleib, Jim O'Neill, and Balaji Srinavasan. I don't know much about Gottleib, besides that he writes for Forbes about how Obamacare is bad. But either of the latter two would be shocking breaks with tradition and potentially among the highest-value political experiments of all time.

Jim O'Neill is a director at Mithril Capital and a former deputy deputy (sic) HHS secretary. He's a (former?) board member of the Seasteading Institute, which hopes to create a libertarian utopia on a floating platform in international waters, and which recently signed a preliminary agreement with French Polynesia to begin pre-construction planning. He's also a director of SENS, Aubrey De Grey's collaboration to fight aging, and has proposed increasing the organ supply by paying donors. Also, I see him commenting on Eliezer's Facebook feed sometimes, and he seems to be Facebook friends with Eliezer, Julia Galef, and, uh, me. Maybe he reads this blog? Hi, Jim!

Balaji Srinivasan describes himself as "a computer scientist, investor, entrepreneur, and academic", and previously founded a very successful genetic testing startup; now he holds various Bitcoin-related positions. He’s famous (infamous?) for a piece called Silicon Valley’s Ultimate Exit, where he promotes “exit” over “voice”; he suggests Silicon Valley find ways to create an alternate society that escapes the dysfunction of the federal government and the rest of the country – both figuratively via new institutions like Bitcoin and eventually literally through possibilities like seasteading. He called the FDA “the man to beat – or escape, via exit – whether you want a new drug or to get to transhumanism”. And, uh, he *also* follows me on social media, which is definitely not a characteristic I expected multiple candidates for FDA director to have in common.

I worry that my more liberal friends won’t be sufficiently impressed. They’re thinking “Oh, more libertarians, like all those small-government people from the Bush and Reagan administrations”. My own view is that “libertarian” gets used to pick out at least two different clusters of people. One is rich crony capitalists who want a convenient excuse to cut their own taxes and roll back workers’ rights, but who fight tooth and nail against any decreased subsidies or increased competition that might threaten their own comfortable position. The other is people who are actually interested in using the power of competition to kindle innovation, improve access, bring down entrenched interests, and ultimately help regular people. The first cluster of libertarians has been around forever. O’Neill and Srinivasan seem to be part of the second. A principled, intelligent cluster-two libertarian getting the top job at one of the country’s worst bureaucracies would be practically unprecedented.

(and lest I try to weasel out of this one later, let me state for the record that if O’Neill or Srinivasan get chosen and are able to implement their preferred policies, but the US pharmaceutical industry doesn’t improve dramatically, I should accept it as a defeat for one or another hypothesis of mine. Either free markets don’t work in medicine, or I am so terrible at identifying principled intelligent libertarians that I might as well give up.)

Some important policies that an FDA commissioner like O’Neill or Srinivasan might be able to implement with high benefit and little cost:

1. Medical reciprocity with Europe and other First World countries (The Atlantic, Health Affairs, Marginal Revolution). Right now, Europe has a licensing agency about as strict as the FDA approving medications invented in Europe. Any pharma company that wants their medication approved in both the US and Europe has to spend a billion or so dollars getting it approved by the FDA, and then *another* billion or so dollars getting it approved by the Europeans. A lot of pharma companies don’t want to bother, with the end result that Europe has many good medications that America doesn’t, and vice versa. Just in my own field, amisulpride, one of the antipsychotics with the best safety/efficacy balance, has been used successfully in Europe for twenty years and is totally unavailable here despite a real need for better antipsychotic drugs. If the FDA agreed to approve any medication already approved by Europe (or to give it a very expedited review process), we could get an immediate windfall of dozens of drugs with unimpeachable records for almost no cost. Instead, in the real world, we’re cracking down on imported Canadian pharmaceuticals because the Canadians don’t have exactly our same FDA which means that for all we know they might be adding thalidomide to every pill or something. This is exactly the sort of silly anti-competitive cronyist practice that a principled intelligent libertarian could do away with.

2. Burdensome approval process for generic medications (SSC, more SSC). How come Martin Shkreli can hike the dose of an off-patent toxoplasma drug 5000%, and everyone just has to take it lying down even though the drug itself is so easy to produce that high school chemistry classes make it just to show they can? The reason is that every new company that makes a drug, even a widely-used generic drug that's already been proven safe, has to go through a separate approval process that costs millions of dollars and takes two to three years – and which other companies in the market constantly try to sabotage through legal action. Shkreli can get away with his price hike because he knows that by the time the FDA gives anyone permission to compete with him, he'll have made his fortune and moved on to his next nefarious scheme. If the FDA allowed reputable pharmaceutical companies in good standing to produce whatever generic drugs they wanted, the same as every other company is allowed to make whatever products they want, scandals like Daraprim and EpiPens would be a thing of the past, and the price of many medications could decrease by an order of magnitude.

3. Stop having that thing allowing companies to "steal" popular and effective drugs that have been in the public domain for years, claim them as their private property, shut down all competitors, and jack up the price 10x just by bringing them up to date with modern FDA bureaucracy.

4. Stop having that thing where drug companies can legally bribe other companies not to compete with them. I like this one because it sounds anti-libertarian (we're imposing a new regulation on what companies can do!) but I think it's exactly the sort of thing that the crony capitalists would never touch but which principled intelligent libertarians like O'Neill and Srinivasan might be open to, in order to bring more actors into the marketplace.

5. Stop thwarting consumer diagnostic products and genetic tests (SSC, more SSC). Srinivasan comes from the genetic testing world himself, so he’s likely to be extra sympathetic to this.

I notice that Jim O’Neill had (in 2014) a much more radical proposal than any of these: that the FDA should approve drugs based on safety but not efficacy; that is, drug companies have to prove that their drug isn’t dangerous, but they don’t have to do the long-term super-expensive studies proving that it works. This isn’t quite as crazy as it sounds – it just means we’d need to use academic studies and good judgment to figure out what works. We do this already in many cases – with drugs that were grandfathered in before the FDA existed (eg penicillin), or with drugs that the FDA approved for one indication but we use off-label but which we use for another indication (eg Prozac for anxiety). The best-case scenario is that a safety-but-not-efficacy regime would replicate that kind of careful skepticism across the board. The worst-case scenario is that we end up with a lot of ineffective drugs being used for a decade or so until science can catch up and prove them ineffective – something which is arguably already happening. Honestly this kind of policy is probably too revolutionary even for me – but in a world full of stupid regressive fear-driven bad ideas, it’s a bold revolutionary high-variance bad idea, and I respect that (see part 3 here for another problem with this idea).

The pharmaceutical industry stock index hasn’t moved much since some of these names started being floated. I’m not sure what to make of this; I wouldn’t have been able to predict which direction it would go, but I would have expected *some* direction.

One more really interesting potential appointment. Nature.com: Surprising Contenders Emerge For Trump’s NIH Chief. Most of the contenders aren’t *that* surprising: Collins is the current NIH chief, and Harris is a Republican congressman with an MD and a strong interest in health policy. One interesting idea of Harris’ is a pledge to lower the age at which researchers get their first grant, which would address a widely shared concern that we’re losing out on creative ideas because people have to spend decades learning to conform and playing academic politics before anyone pays attention to them. I guess this could be good (though see here, apparently the current NIH director supports this as well).

But one name stands out: Stanford statistician John Ioannidis, famous for promoting high quality studies and raising the standard of medical research. I don’t know what his administrative credentials are or what talents you need to run the NIH, but along with the Cochrane people he sets the gold standard for trustworthy bioscience, and having him in a high scientific position would do more to raise *my* confidence in the standard of US medical research than almost anything else I can think of.

SSC Survey 2017

If you’re reading this and have previously read at least one SSC post, please take the 2017 SSC survey. Warning: it’s pretty long.

You can talk about it in the comments, but don’t read them until you’re done taking the survey.

EDIT: Note that a lot of this is going to be processed by computer. That means you should give simple, machine readable answers. If the question is “what city do you live in?”, a simple, machine-readable answer is “Los Angeles”, which easily aggregates into the same bin as everyone else who answered “Los Angeles”. Bad answers would include “I live in Los Angeles”, “Just north of Los Angeles”, “Los Angeles but sometimes in Dallas for work”, et cetera.

Link: Guyenet On Taubes

If you’re interested in diet, I recommend Stephan Guyenet’s **review of Gary Taubes new book**:

> The Case Against Sugar is a journey through sugar history and science that argues the point that sugar is the principal cause of obesity, diabetes, coronary heart disease, and many other common noncommunicable diseases. This differs from the prevailing view in the research and public health communities that obesity and noncommunicable disease are multi-factorial, with refined sugar playing a role among other things like excess calorie intake, physical inactivity, cigarette smoking, alcohol and illegal drug use, and various other diet and lifestyle factors. I side with the latter view.

Some key quotes:

> Taubes argues that sugar is the only factor that reliably shows up when a culture develops Western noncommunicable diseases, supporting the point with examples of cultures that adopted sugar-rich diets and became ill. Yet he makes no effort to look for a counterexample that could refute his argument: a traditionally-living culture that has a high intake of sugar and does not suffer from Western noncommunicable diseases. If such a culture can be found, this piece of evidence is sufficient to reject Taubes’s argument that sugar reliably associates with the onset of these diseases in a population. Let’s do Taubes’s research for him. A well-studied Tanzanian hunter-gatherer tribe called the Hadza gets 15 percent of its average year-round calorie intake from honey, plus fruit sugar on top of it. This approximates US sugar intake, yet the Hadza

> do not exhibit obesity, cardiovascular disease, or any of the other disorders Taubes attributes to sugar. In fact, many hunter-gatherer groups relied heavily on honey historically, including the Mbuti of the Congo whose diet was up to 80 percent honey during the rainy season. Yet they do not exhibit obesity or insulin resistance.

And:

> Taubes upbraids the research community for its belief that body fatness is determined by calorie intake, rather than the impact of foods on insulin. He supports the latter proposition with semi-anecdotal observations from Africa suggesting that a group of people eating a high-sugar diet supplying “as little as sixteen hundred calories per day” were sometimes obese and diabetic. A person who actually wants to get to the bottom of this question should conduct their investigation in a very different manner. The first order of business is to look up the relevant metabolic ward studies, which are the most tightly controlled diet studies available. These studies consistently show that calorie content is the only known food property that has a meaningful impact on body fatness. This is true across a wide range of carbohydrate-to-fat ratios and sugar intakes, and a correspondingly wide range of insulin levels. What makes Taubes’s oversight so extraordinary is that he was involved in funding one of these metabolic ward studies, which compared two diets that differed tenfold in sugar content. The results showed that a 25 percent sugar, high-carbohydrate diet caused slightly more body fat loss than a 2.5 percent, very-low-carbohydrate (ketogenic) diet of equal calories (18). Despite these clear and consistent findings, Taubes continues to insist that calorie intake is not an important determinant of body fatness, and he offers the reader questionable

> evidence in support of this while omitting high-quality evidence to the contrary. All while exuding righteous indignation about the scientific community's misguided beliefs.

And:

> His discussion of the history of research on sugar, fat, and obesity is less compelling due to its one-sided nature. For example, The Case Against Sugar portrays an epic struggle decades ago between researchers who believed that saturated fat was the primary cause of coronary heart disease, and those who believed that sugar was. These views are embodied by the American researcher Ancel Keys and the British researcher John Yudkin, respectively. Taubes makes hay of the fact that Keys was supported in part by the sugar industry, painting Yudkin as a righteous underdog standing up to a corrupt and aggressive Keys. Yet he never gives serious consideration to the strength of the evidence supporting each man's beliefs, instead using a historical narrative to imply that Keys was a stooge of the sugar industry who unfairly won the argument due to his sharp elbows (whether or not this is true, it's also true that Yudkin's evidence was not as compelling as Keys's). This tactic of using historical narratives as a substitute for evidence is one that recurs throughout the book. In his haste to undermine Keys, Taubes neglects to mention that Yudkin had his own conflicts of interest: he was funded by the egg, edible oil, and dairy industries, all of which had an interest in pinning the blame for obesity and chronic disease on sugar

(I made the same point as that last one here.)

I’m recommending this review more strongly than usual because I’ve previously praised Taubes. He did a good job explaining how the pop wisdom of the ’90s – that fat was uniquely bad – wasn’t true. He did a good job showing the ways in which the old “just diet and exercise, nothing can go wrong” idea was simplistic and needed to be replaced with a good understanding of obesity set points. I learned some useful things from his books and I had positive feelings about him.

But from the first time I talked about him almost five years ago, I’ve stressed that his views about sugar are really, really wrong. I was previously willing to excuse this on the grounds that he wrote about a lot of other useful things and everybody’s allowed a bit of crazy speculation once in a while before a field is completely settled. But at this point things seem pretty settled and I no longer think his behavior is excusable. The ’90s are over, the pop wisdom that Taubes set out to debunk is sufficiently debunked, and he’s doubling down on his sugar theory despite an increasing pile of evidence against it.

I apologize if my past praise for Taubes’ writing have helped create a climate where people listen to his theories. His good qualities aren’t enough to justify the increasing amount of misinformation he’s putting out, and I don’t recommend him on any level as a source of nutritional advice.

Dogs And Wolves: In Defense Of Some Past Posts

Now that Trump has started enacting his terrible policies, a bunch of people on Twitter are saying that my past posts on Trump "haven't aged well" or that I must be feeling really bad about them right now.

I've never been the slightest bit of a Trump supporter. Since he came onto the national stage, I have called Trump "a bad president", "randomly and bizarrely terrible", "an emotionally incontinent reality TV show host", and "an incompetent thin-skinned ignorant boorish fraudulent omnihypocritical demagogue". I've accused him of "bizarre, divisive, ill-advised, and revolting" rhetoric, worried that his election might "lead directly to the apocalypse [or] the fall of American democracy", and called his administration "a disaster". I've urged blog readers to vote for literally anyone except him and to donate money to the ACLU to stop him. If you want to accuse me of being pro-Trump, or even lukewarm on disliking Trump, I don't know what else to tell you.

But I still seem to be getting flak on two points.

First, in Against Dog-Whistleism, I condemned the practice of overinterpreting candidates' statements to secretly reveal evil beliefs and policies that they support more strongly than their stated platform:

> Although dog whistles do exist…politicians' beliefs and plans are best predicted by what they say their beliefs and plans are, or possibly what beliefs and plans they've supported in the past, or by anything other than treating their words as a secret code and trying to use them to

> infer that their real beliefs and plans are diametrically opposite the beliefs and plans they keep insisting that they hold and have practiced for their entire lives.

Believers in dog whistles are saying I have egg on my face because Trump just passed a law banning immigrants from some Muslim countries. Clearly, (they say) this means we should have listened to his dog whistles all along.

But this is dead wrong. Trump openly said throughout his campaign that he planned to ban immigrants from some Muslim countries. See for example Bloomberg, June 25: Trump Says Muslim Ban Plan To Focus On 'Terrorist' Countries. Trump has been saying this openly for seven months. If you didn't know Trump wanted this, it's not because he was being cryptic about it. It was because you were too busy chasing down his "dog whistles" about how he secretly hated Jews to listen to him.

Second, people are saying that my post You Are Still Crying Wolf has been debunked, since Trump has banned immigrants from some Muslim countries, and so is obviously the KKK-loving white supremacist that I argued he wasn't.

Look, guys. I *specifically said* in that post that I knew he was going to ban people from some Muslim countries, and that when he did, that would be consistent with my model:

> 13. Doesn't Trump want to ban (or "extreme vet", or whatever) Muslims entering the country?
>
> Yes, and this is awful.
>
> But why do he (and his supporters) want to ban/vet Muslims, and not Hindus or Kenyans, even though most Muslims are white(ish) and most Hindus and Kenyans aren't? Trump and his supporters are concerned about

terrorism, probably since the San Bernardino shooting and Pulse nightclub massacre dominated headlines this election season.

You can argue that he and his supporters are biased for caring more about terrorism than about furniture-related injuries, which kill several times more Americans than terrorists do each year. But do you see how there’s a difference between “cognitive bias that makes you unreasonably afraid” versus “white supremacy”?

I agree that this is getting into murky territory and that a better answer here would be to deconstruct the word “racism” into a lot of very heterogenous parts, one of which means exactly this sort of thing. But as I pointed out in Part 4, a lot of these accusations shy away from the word “racism” precisely *because* it’s an ambiguous thing with many heterogenous parts, some of which are understandable and resemble the sort of thing normal-but-flawed human beings might think. Now they say “KKK white nationalism” or “overt white supremacy”. These terms are powerful exactly because they do *not* permit the gradations of meaning which this subject demands.

Let me say this for the millionth time. I’m not saying Trump doesn’t have some racist attitudes and policies. I am saying that talk of “entire campaign built around white supremacy” and “the white power candidate” is deliberate and dangerous exaggeration. Lots of people (and not just whites!) are hasty to generalize from “ISIS is scary” to “I am scared of all Muslims”. This needs to be called out and fought, but it needs to be done in an understanding way, not with cries of “KKK WHITE SUPREMACY!”

Apparently saying it a million times is not enough for some people. I'm afraid there's a mood affiliation effect going on here – that if I say Trump didn't cause 9-11, then people can only hear "SCOTT SAYS TRUMP ISN'T THAT BAD!". Arguments aren't soldiers, and Trump can both not cause 9-11 and be bad for lots of other reasons.

So I stick to what I've said before. Trump is a bad person, a bad president, and probably a disaster for America and the world. If you want to know what Trump is going to do, you should listen to what he specifically says he is going to do, and then expect things vaguely in that direction except worse. Trump's policies are often motivated by racism of the everyday bias-towards-being-more-scared-of-terrorists-than-the-situation-warrants sort, but he does not have an ideological belief in white supremacy or take marching orders from the KKK. If you want to fight him, I recommend you fight him, not subtweet bloggers who warned you this was going to happen and then told you not to exaggerate it because reality was going to be scary enough.

I made predictions about how the Trump administration would go, which you can find on the Predictions tab above. If you disagree with me about any of this, use your beliefs to make different predictions, record them, and see if you do better than I do. Do enough better and I'll admit I was wrong and you were right.

But if you tell me on Twitter I'm wrong because my model of Trump could never predict the things I specifically predicted when laying out the model, I'm not going to pay very much attention to you.

Link: Taubes Contra Guyenet on Sugar

On Thursday I linked to Stephan Guyenet's negative book review of Gary Taubes' work on sugar.

Taubes asked that I pass on a counterargument he's been working on for a while. In early January, he posted a short version of his case against sugar to Cato Unbound. It got three response essays – one by Terence Kealy, one by Yoni Freedhoff, and one by Stephan Guyenet, where he made some of the same points as in the linked article.

So now Taubes has now posted a response, **The Case Against Sugar Isn't So Easily Dismissed**.

All of the essays are interesting and informative, and everyone involved deserves praise for some high-level debate that's helped me clarify my thinking on this.

Book Review: Eichmann In Jerusalem

[Content warning: Holocaust. This is a complicated and emotional subject and I make no claims to know much more than what I read in the book, nor to be 100% certain I am representing Arendt's views faithfully.]

I.

For Holocaust Remembrance Day last week I read Hannah Arendt's *Eichmann In Jerusalem* (h/t Ben Hoffman).

Adolf Eichmann organized the logistics of the Holocaust – helped get Jews into trains, helped get the trains to the right concentration camps. When Germany lost the war, he escaped to Argentina and lived under a fake name. The newly-formed state of Israel hunted him down, and in 1960 they kidnapped him and put him on trial in Jerusalem.

The Nuremberg Trials were led by an Allied force that wanted to stress that the Nazis committed crimes against all humanity. Eichmann's trial was the first time Jews themselves tried a high-ranking Nazi for his crimes against Jews in particular. Israeli PM David Ben-Gurion wanted (and got) a show trial. Not in the sense of justice not being done (everyone agreed Eichmann was guilty), but in the sense of highlighting the horrors of the Holocaust to the world.

Arendt recorded a lot of weird, surprising, and disturbing things in her study of Eichmann's trial. I found five particularly interesting: Eichmann's psychological profile, the Nazis' early pre-war plans for the Jews, the ways German-occupied nations did or didn't resist genocide demands, the politics surrounding claims that Jews didn't resist the Nazis enough, and the discussion of why more Germans didn't protest. I want to discuss all of these, then finish with whether this has any relevance for today’s political climate.

II.

Arendt’s psychological profile of Eichmann is most famous for coining the phrase “banality of evil”. Eichmann was neither a charming psychopath nor a blustering villain. As per Arendt:

> Despite all the efforts of the prosecution, everybody could see that this man was not a “monster,” but it was difficult indeed not to suspect that he was a clown. And since this suspicion would have been fatal to the whole enterprise, and was also rather hard to sustain, in view of the sufferings he and his like had caused so many millions of people, his worst clowneries were hardly noticed. What could you do with a man who first declared, with great emphasis, that the one thing he had learned in an ill-spent life was that one should never take an oath (“Today no man, no judge could ever persuade me to make a sworn statement. I refuse it; I refuse it for moral reasons. Since my experience tells me that if one is loyal to his oath, one day he has to take the consequences, I have made up my mind once and for all that no judge in the world or other authority will ever be capable of making me swear an oath, to give sworn testimony. I won’t do it voluntarily and no one will be able to force me”), and then, after being told explicitly that if he wished to testify in his own defense he might “do so under oath or without an oath,” declared without further ado that he would prefer to testify under oath?

Eichmann’s attorney, a somewhat incompetent German named Dr. Servatius, instructed him to plead innocent.

Eichmann could have taken this advice and tried to save his skin. Or he could have taken the high road and confessed his guilt. He chose to do neither:

> To each count Eichmann pleaded: “Not guilty in the sense of the indictment.” In what sense then did he think he was guilty?…”With the killing of Jews I had nothing to do. I never killed a Jew, or a non-Jew for that matter – I never killed any human being…I never gave an order to kill either a Jew or a non-Jew”…or, as he was later to qualify this statement, “It so happened…that I had not once to do it” – for he left no doubt that he would have killed his own father if he had received an order to that effect. Hence he repeated over and over… that he could only be accused of “aiding and abetting” the annhilation of the Jews, which he declared in Jerusalem to have been “one of the greatest crimes in the history of Humanity” […]
>
> Would he then have pleaded guilty if he had been indicted as an accessory to murder? Perhaps, but he would have made important qualifications…He did not want to be one of those who now pretended that “they had always been against it”, whereas in fact they had been very eager to do what they were told to do. However, times change, and he, like Professor Maunz, had “arived at different insights”. What he had done he had done, he did not want to deny it; rather, he proposed “to hang myself in public as a warning example for all anti-Semites on this earth”. By this he did not mean to say that he regretted anything: “Repentance is for little children.” (sic) […]
>
> Throughout the trial, Eichmann tried to clarify, mostly without success, this second point in his plea of “not guilty in the sense of the indictment.” The indictment

> implied not only that he had acted on purpose, which he did not deny, but out of base motives and in full knowledge of the criminal nature of his deeds. As for the base motives, he was perfectly sure that he was not what he called an innerer Schweinehund, a dirty bastard in the depths of his heart; and as for his conscience, he remembered perfectly well that he would have had a bad conscience only if he had not done what he had been ordered to to – to ship millions of men, women, and children to their death with great zeal and the most meticulous care. This, admittedly, was hard to take. Half a dozen psychiatrists had certified him as "normal" – "More normal, at any rate, than I am after having examined him," one of them was said to have exclaimed.

Go ahead and try to parse all of that into a coherent worldview. Was he regretful? Was he proud? Was he anti-Semitic? Was he just following orders? I don't think anyone at the trial ever got a good feel for this. I certainly didn't. Arendt isn't sure there's anything there to figure out:

> The judges were right when they finally told the accused that all he had said was "empty talk" – except that they thought the emptiness was feigned, and that the accused wished to cover up other thoughts which, though hideous, were not empty. This supposition seems refuted by the striking consistency with which Eichmann, despite his rather bad memory, repeated word for word the same stock phrases and self-invented clichés (when he did succeed in constructing a sentence of his own, he repeated it until it became a cliché) each time he referred to an incident or event of importance to him. Whether writing his memoirs in Argentina or in Jerusalem, whether speaking to the police examiner or to the court, what he said was always the same,

> expressed in the same words. The longer one listened to him, the more obvious it became that his inability to speak was closely connected with an inability to think, namely, to think from the standpoint of somebody else. No communication was possible with him, not because he lied but because he was surrounded by the most reliable of all safeguards against the words and the presence of others, and hence against reality as such.

But if she has any thesis at all, it’s that Eichmann believed in something larger than himself. We usually encourage this sort of thing, but I think the prosocial version involves having a *specific* larger-than-yourself thing in mind. Eichmann (says Arendt) just liked larger-than-himself things in general, and the Nazi vision of eternal struggle for racial supremacy was the biggest thing he could find in the vicinity. We’ll later see that he had a strange respect for Zionists, and this was because they too believed in something larger than themselves. Eichmann’s infamous cliches were the cliches of pomp and circumstance and glory and high words, the ones which made him feel like he was engaged in a great enterprise *whether or not there was anything behind them*. The reason he admitted neither to “just following orders”, nor to a deep personal belief in anti-Semitism, was that his loyalty to Hitler came from neither. When Hitler said to kill all the Jews, he gladly complied; if Hitler had said to kill all the Christians, he would have done that too. Not because he was a drone following orders to save his skin, but because he believed. Not in any of the specifics of Nazi ideology. Not even in Hitler’s personal judgment. Just in whatever was going on at the time.

And so when Eichmann’s superior Himmler betrayed Hitler (more on this later) and ordered Eichmann to stop the exterminations, Eichmann – finally – refused an order.

Himmler’s betrayal seemed petty; Hitler’s vision seemed grand. And so:

> The sad and very uncomfortable truth of the matter probably was that it was not his fanaticism but his very conscience that prompted Eichmann to adopt his uncompromising attitude during the last year of the war…Eichmann’s position, therefore, showed a most unpleasant resemblance to that of the often-cited soldier who, acting in a normal legal framework, refuses to carry out orders that run counter to his ordinary experience of lawfulness and hence can be recognized by him as criminal.

But with Hitler dead and the war lost, the grandest gesture Eichmann can think of is to try to become a public martyr to edify future generations. So he tries that too.

Even understanding this, there’s another set of mysteries. Eichmann would get so many facts wrong in his testimonies that everyone would figure he was lying; then, without even being asked, he would confess to much worse sins than any of the ones he had denied (a big part of the prosecution’s case rested on Eichmann volunteering the information that he went into concentration camps a few times and saw exactly what happened there, something which otherwise would have been hard to prove and might have left space for an “I didn’t know how bad it was” defense). And he would talk obsessively about his failure to get promoted quickly enough through the Nazi hierarchy, clearly expecting his Israeli audience to sympathize with him:

> What makes these pages of the examination so funny is that all this was told in the tone of someone who was sure of finding “normal, human” sympathy for a hard-luck story. “Whatever I prepared and planned,

> everything went wrong, my personal affairs as well as my years-long efforts to obtain land and soil for the Jews. I don't know, everything was as if under an evil spell; whatever I desired and wanted and planned to do, fate prevented it somehow. I was frustrated in everything, no matter what." When Captain Less asked his opinion on some damning and possibly lying evidence given by a former colonel of the S.S., he exclaimed, suddenly stuttering with rage: "I am very much surprised that this man could ever have been an S.S. Standartenführer, that surprises me very much indeed. It is altogether, altogether unthinkable. I don't know what to say." He never said these things in a spirit of defiance, as though he wanted, even now, to defend the standards by which he had lived in the past. The very words "S.S.," or "career," or "Himmler" (whom he always called by his long official title: Reichsführer S.S. and Chief of the German Police, although he by no means admired him) triggered in him a mechanism that had become completely unalterable. The presence of Captain Less, a Jew from Germany and unlikely in any case to think that members of the S.S. advanced in their careers through the exercise of high moral qualities, did not for a moment throw this mechanism out of gear

What should we make of this? Arendt described Eichmann as having an "almost total inability ever to look at anything from the other fellow's point of view", and this seems right. For him, self-absorbed as he was, the story of World War II was the story of him doing a pretty competent job of Jew-killing but not getting the recognition he deserved from his superiors. He was unable to understand that other people might have a different perspective, or that Israeli Holocaust survivors wouldn't find his story about unfairness in Himmler's HR department as moving as he did.

This might explain his pattern of omissions and confessions. He was omitting things that seemed bad to him – tied into his obsessions or made him look like a worse bureaucrat. But he didn’t have enough ability to model his Israeli interlocutors to know that “knew what happened at concentration camps” would seem bad to them, or else he didn’t even realize that “seems bad to the Israelis” was a thing.

This reminds me of my theory that some people are just born without certain cogs in their brain, and especially without theory of mind. Eichmann’s theory of mind was just totally absent. He expected the Jews he deported to be thankful to him for all the hard work he was putting in! The only way I can imagine that working is if Eichmann found his 9-5 job tiring and was so fantastically self-centered that he expected the Jews to see it exactly the same way he did (“Oh, look at that poor Eichmann working so hard to deal with us”). There are interesting implications here – that some level of theory of mind is necessary for basic consistency (ie realizing that other people will stop liking you if you’re inconsistent) and possibly for basic humanity (in order to not want to send people to concentration camps, you have to realize that they have their own thoughts and feelings about it separate from yours).

I should emphasize that some more recent scholars have dissented from Arendt at this point, saying that Eichmann’s apparent dullness and inconsistency was a careful ruse put on to fool his jailers. I don’t know nearly enough history to comment on this one way or the other. But in his last moments on Earth, he died as he lived – saying some faux-profound stock phrases without realizing how weird he sounded:

> Adolf Eichmann went to the gallows with great dignity… He was in complete command of himself, nay, he was more: he was completely himself. Nothing could have demonstrated this more convincingly than the grotesque silliness of his last words. He began by stating emphatically that he was a Gottgläubiger, to express in common Nazi fashion that he was no Christian and did not believe in life after death. He then proceeded: "After a short while, gentlemen, we shall all meet again. Such is the fate of all men. Long live Germany, long live Argentina, long live Austria. I shall not forget them." In the face of death, he had found the cliché used in funeral oratory. Under the gallows, his memory played him the last trick; he was "elated" and he forgot that this was his own funeral. It was as though in those last minutes he was summing up the lesson that this long course in human wickedness had taught us-the lesson of the fearsome, word-and-thought-defying banality of evil.

III.

I had always interpreted "the Final Solution" to mean "the solution that will last forever". Arendt (and I don't know if she's right about this) interprets it to mean "the last of many solutions to be tried". She discusses the failed First Solution and Second Solution as preludes for the eventual genocide.

The First Solution was emigration. In the early days of the movement, the Nazis seemed to sincerely believe that they could deal with the Jews just by expelling them from Germany and letting other countries take care of them, no killing necessary. I don't want to let this sound like the Nazis "started out okay" – Hitler had expressed support for killing the Jews as early as the publication of *Mein Kampf* in 1925. But for his first few years in power, he distanced himself from his previous positions and accepted emigration as a practical compromise.

When the Nazis first decided to expel the Jews, Eichmann was working as a low-level vacuum salesman in Vienna. He got his big break when he signed up for a job with the Party trying to get Jews to emigrate. This was tough work – many Jews didn’t want to emigrate, and the ones who did needed more paperwork than the German bureaucracy could easily provide. Eichmann displayed some early talent at cutting red tape and figuring out ways to connect Jews who wanted to leave with bureaucrats who wanted to let them, and he rose through the ranks until he was in charge of Jewish emigration from Vienna.

At this point a friend suggested he read Theodor Herzl’s book on Zionism, and Eichmann, bizarrely, fell in love. Arendt says, apparently in earnest, that it “seems to have been the first serious book he ever read” and that “it made a lasting impression on him”:

> It may be worth mentioning that, as late as 1939, he seems to have protested against desecrators of Herzl’s grave in Vienna, and there are reports of his presence in civilian clothes at the commemoration of the thirty-fifth anniversary of Herzl’s death…he began spreading the gospel among his SS comrades, giving lectures and writing pamphlets. He acquired a smattering of Hebrew, which enabled him to read haltingly a Yiddish newspaper – not a very difficult accomplishment, since Yiddish, basically an old German dialect written in Hebrew letters, can be understood by any German-speaking person who has mastered a few dozen Hebrew words. He even read one more book, Adolf Bohm’s History of Zionism, and this was perhaps a considerable achievement for a man who by his own account had

> always been utterly reluctant to read anything except newspapers.

Eichmann seemed weirdly in earnest about all of this, but it was also good for his job – he met with Zionist Jews and even went to Palestine once to meet with the Zionist movement there. He loved to say during his trial that Austrian Jewish immigration to Israel was a win-win – it made the Jews happy because they were going to their homeland, and it made the Nazis happy because the Jews were leaving Austria. When he related his self-perception as a basically decent person, he always stressed that this was his idea, and he was a win-win sort of person who had been unfairly transferred to the sending-people-in-boxcars-to-concentration-camps department against his will.

The work was not nearly as win-win as Eichmann liked to think; for example, the Nazis confiscated all of the Jews' property as the "price" of providing them with the necessary documents. When poor Jews without any property showed up to emigrate, Eichmann would shake down the rich Jews and making them pay extra to help their poorer co-religionists. Finally this turned into outright blackmail, demanding blood money from Jews in the Diaspora, or else. In any case, it worked – a hundred fifty thousand Jews left Austria during Eichmann's eighteen months in the business.

What eventually happened we all know too well. Other countries started closing their doors and refusing to accept Jewish refugees. Despite hearing this story a hundred times, the version in *Eichmann in Jerusalem* was new to me. I had always thought of countries as closing their gates to a few prescient people trying to flee Nazi Germany on their own, or to a few stragglers who managed to escape. The truth is on a much greater scale: the Nazis were willing to let every single Jew in Europe leave, they even had entire bureaucracies trying to make it happen – and the rest of the world wouldn’t cooperate. The blood on the hands of the people who wouldn’t let them in is not just that of a few escapees, but the entire six million.

When emigration stopped working, the Nazis turned to the Second Solution – resettlement. Arendt doesn’t think the plan to move all the Jews to Madagascar was ever taken seriously at the highest level, but for a while it was something like official policy. The only problems were that the Nazis didn’t technically own Madagascar, that they didn’t have nearly enough ships to transport six million people, and that all the water in between was controlled by British warships intent on sinking any Germans they could find. The send-the-Jews-to-Madagascar plan seemed to be a loose alliance of high-level leadership looking for a cover story while they prepared for genocide, plus very stupid people who liked bad ideas. No guessing which group Eichmann was in.

(Actually, Eichmann got super-confused and apparently thought Madagascar was the same place as Uganda, which Herzl had mentioned as a possible Jewish homeland if Israel was unavailable. He announced the good news to some of his Jewish contacts, who gave him a remedial lesson in African geography.)

There was a slightly more serious proposal to create a Jewish homeland in Radom District, Poland (note that “Jewish homeland” here meant basically a country-sized prison, not a self-governing Jewish state.) This had the advantages of the Nazis actually controlling Poland and of rail networks up to the task of transporting people over. It failed because some overly enthusiastic Nazis just sent a trainload of thousands of Jews there without informing the

Governor of Poland, and he got confused and angry, plus a lot of the Jews escaped.

Then some people briefly tried to turn the Czech city of Theresienstadt into a Jewish territory, but it was really small and eventually it just ended up as a slightly-less-murderous-than-usual concentration camp.

IV.

Arendt interrupts the story of Eichmann for a long and fascinating digression about which European nations did or didn’t protect their Jews.

Remember that most nations of Central and Eastern Europe were German puppet states during this period. The Nazis made it clear that deporting their Jews to the concentration camps in Nazi territory was a condition for continued good relations; a serious threat, when bad relations could turn a protectorate-type situation into an outright invasion and occupation.

Pride of place goes to Denmark and Bulgaria, both of which resisted all Nazi demands despite the Germans having almost complete power over them. Most people have heard the legend of how, when the Germans ordered that all Jews must wear gold stars, the King of Denmark said he would wear one too. These kinds of actions weren’t just symbolic; without cooperation from the Gentile population and common knowledge of who was or wasn’t Jewish, the Nazis had no good way to round people up for concentration camps. Nothing happened until 1943, when Himmler became so annoyed that he sent his personal agent Rolf Gunther to clean things up. Gunther tried hard but found the going impossible. Danish police refused to go door-to-door rounding up Jews, and when Gunther imported police from

Germany, the Danes told them that they couldn't break into apartments or else they would arrest them for breaking and entering. Then the Danish police tipped off Danish Jews not to open their doors to knocks since those might be German police. When it became clear that the Nazis weren't going to accept any more delays, Danish fishermen offered to ferry Jews to neutral Sweden for free. In the end the Nazis only got a few hundred Danish Jews, and the Danish government made such a "fuss" (Arendt's word) about them that the Nazis agreed to send them all to Theresienstadt, their less-murderous-than-usual camp, and let Red Cross observers in to make sure they were treated well. As a result, only 48 Danish Jews died in the entire Holocaust.

Bulgaria's resistance was less immediately heroic, and looked less like the king proudly proclaiming his identity with oppressed people everywhere than like the whole government just dragging their feet so long that nothing got done. Eichmann sent an agent named Theodor Dannecker to get them moving, but as per Arendt:

> Not until about six months later did they take the first step in the direction of "radical" measures – the introduction of the Jewish badge. For the Nazis, even this turned out to be a great disappointment. In the first place, as they dutifully reported, the badge was only a "very little star"; second, most Jews simply did not wear it; and, third, those who did wear it received "so many manifestations of sympathy from the misled population that they actually are proud of their sign" – as Walter Schellenberg, Chief of Counterintelligence in the R.S.H.A., wrote in an S.D. report transmitted to the Foreign Office in November, 1942. Whereupon the Bulgarian government revoked the decree. Under great German pressure, the Bulgarian government finally decided to expel all Jews from Sofia to rural areas, but

> this measure was definitely not what the Germans demanded, since it dispersed the Jews instead of concentrating them.

The Bulgarians continued their policy of vaguely agreeing in principle to Nazi demands and then doing nothing, all the way until the Russians invaded and the time of danger was over. Not a single Bulgarian Jew died in the Holocaust (edit: see here).

Surprisingly given the bad associations I have with the word “fascist”, Mussolini’s Italy also deserves high praise for protecting its Jews. Arendt describes Italy as “sabotaging” the Final Solution within its borders despite nominal alliance with Germany:

> The gentlemen of the Foreign Office could not do much about it, because they always met the same subtly veiled resistance, the same promises and the same failures to fulfill them. The sabotage was all the more infuriating as it was carried out openly, in an almost mocking manner. The promises were given by Mussolini himself or other high-ranking officials, and if the generals simply failed to fulfill them, Mussolini would make excuses for them on the ground of their “different intellectual formation”. Only occasionally would the Nazis be met with a flat refusal, as when General Roatta declared that it was “incompatible with the honor of the Italian Army” to deliver the Jews from Italian-occupied territory in Yugoslavia to the appropriate German authorities.
>
> An element of farce had never been lacking even in Italy’s most serious efforts to adjust to its powerful friend and ally. When Mussolini, under German pressure, introduced anti-Jewish legislation in the late thirties he

> stipulated the usual exemptions – war veterans, Jews with high decorations, and the like – but he added one more category, namely, former members of the Fascist Party, together with their parents and grandparents, their wives and children and grandchildren. I know of no statistics relating to this matter, but the result must have been that the great majority of Italian Jews were exempted. There can hardly have been a Jewish family without at least one member in the Fascist Party, for this happened at a time when Jews, like other Italians, had been flocking for almost twenty years into the Fascist movement, since positions in the Civil Service were open only to members. And the few Jews who had objected to Fascism on principle, Socialists and Communists chiefly, were no longer in the country. Even convinced Italian anti-Semites seemed unable to take the thing seriously, and Roberto Farinacci, head of the Italian anti-Semitic movement, had a Jewish secretary in his employ...
>
> What in Denmark was the result of an authentically political sense, an inbred comprehension of the requirements and responsibilities of citizenship and independence – “for the Danes . . . the Jewish question was a political and not a humanitarian question” (Leni Yahil) – was in Italy the outcome of the almost automatic general humanity of an old and civilized people.

Less happy is the story of France. The Germans realized that the Vichy French were attached to assimilated French Jews, so they started by demanding only those foreign Jews who had come to France as refugees. There were a hundred thousand of these, and Marshal Petain of France said that they had “always been a problem” and he was glad to have “an opportunity to get rid of them” (in his defense, he was under the impression that Jews sent to Germany would be “resettled in the East”). After this had been going on for a while, Eichmann figured that the French were on his side, and asked for permission to take the native French Jews as well. The French, having sent tens of thousands of stateless Jews to the concentration camps, were suddenly *outraged* that the Nazis would dare lift a finger against *French* Jews, and shut down the entire deportation program. I am sure the French count this as a moral victory nowadays, though it’s a very selective sort of morality.

Last place goes to Romania, which had been anti-Semitic since the beginning of time and was genuinely excited to have Nazi orders as an excuse to carry out their own worse impulses:

> In Rumania even the S.S. were taken aback, and occasionally frightened, by the horrors of oldfashioned, spontaneous pogroms on a gigantic scale; they often intervened to save Jews from sheer butchery, so that the killing could be done in what, according to them, was a civilized way.

The Romanians started their own concentration camps to supplement the Nazis’, “more elaborate and atrocious affairs than anything we know of in Germany”, but they didn’t always need them – “deportation Rumanian style consisted in herding five thousand people into freight cars and letting them die there of suffocation while the train traveled through the countryside without plan or aim for days on end; a favorite followup to these killing operations was to expose the corpses in Jewish butcher shops.” Things became so bad that the local Nazi representative, German noble Manfred von Killinger, intervened and asked them to stop and defer to the Third Reich’s own efforts. I feel like when a Nazi named “Baron von Killinger” is horrified by your brutality, it’s time to take a step back and evaluate whether you may have crossed a line.

Other interesting profiles include Greece (hopelessly depressing), Slovakia (very Catholic, in favor of killing Jews but got in a bunch of fights with the Nazis about ethnic Jews who had been baptized into Catholicism), Hungary (ruled by an admiral despite being landlocked; otherwise hopelessly depressing), Belgium (deliberately left the trains unlocked so the Jews could escape!), Holland (kind of like France; the local Gentiles tried to help, but the assimilated Jews sold out the refugee Jews in the hope of placating the Nazis; the Nazis were not placated; three-quarters of Jews died), and Poland (I don’t even want to talk about how hopelessly depressing this one is).

V.

The Israeli authorities conducting the trial had an uncomfortable tendency to return to the idea of the European Jews as complicit in their own destruction.

The Nazis ordered Jewish communities to organize into Judenrate (“Jewish councils”) which could tabulate the number of Jews in their community, help confiscate property, and choose who would go first to the camps. Cooperation was ensured by a combination of special treatment for community leaders and threats of collective punishment if they didn’t comply. The special treatment turned out to be a sham (if the leaders were lucky, they were killed last); the collective punishment was all too real.

The community leaders thought they were negotiating themselves into a position where they would be better organized and could help delay the Nazis and steer them away from the most vulnerable parts of their community, mitigating the damage. This almost never happened; in the rare cases where it did, it was almost never worth it. Thousands of people were subjected to the sorts of heart-wrenching ethical dilemmas usually found only in philosophy lectures involving trolleys:

> The greatest “idealist” Eichmann ever encountered among the Jews was Dr. Rudolf Kastner, with whom he negotiated during the Jewish deportations from Hungary and with whom he came to an agreement that he, Eichmann, would permit the “illegal” departure of a few thousand Jews to Palestine (the trains were in fact guarded by German police) in exchange for “quiet and order” in the camps from which hundreds of thousands were shipped to Auschwitz….Dr. Kastner saved exactly 1,684 people with approximately 476,000 victims.

By Arendt’s telling, sometimes the councils went beyond merely doing what was necessary for survival:

> In Amsterdam as in Warsaw, in Berlin as in Budapest, Jewish officials could be trusted to compile the lists of persons and of their property, to secure money from the deportees to defray the expenses of their deportation and extermination, to keep track of vacated apartments, to supply police forces to help seize Jews and get them on trains, until, as a last gesture, they handed over the assets of the Jewish community in good order for final confiscation. They distributed the Yellow Star badges, and sometimes, as in Warsaw, “the sale of the armbands became a regular business; there were ordinary armbands of cloth and fancy plastic armbands which were washable.” In the Nazi-inspired, but not Nazi-dictated, manifestoes they issued, we still can sense how they enjoyed their new power – “The Central Jewish Council has been granted the right of absolute

> disposal over all Jewish spiritual and material wealth and over all Jewish manpower," as the first announcement of the Budapest Council phrased it. We know how the Jewish officials felt when they became instruments of murder – like captains "whose ships were about to sink and who succeeded in bringing them safe to port by casting overboard a great part of their precious cargo"; like saviors who "with a hundred victims save a thousand people, with a thousand ten thousand.

This turned out to be important. Arendt gives the case of Belgium, where most of the Jews were a hodgepodge of refugees and most of the elders fled early. The Belgian Jews' lack of organization didn't hurt them; it just made them impossible to organize for deportation and extermination, and so more of them survived than in other comparable areas. And:

> Wherever Jews lived, there were recognized Jewish leaders, and this leadership, almost without exception, cooperated in one way or another, for one reason or another, with the Nazis. The
> whole truth was that if the Jewish people had really been unorganized and leaderless, there would have been chaos and plenty of misery but the total number of victims would hardly have been between four and a half and six million people. According to Freudiger's calculations about half of them could have saved themselves if they had not followed the instructions of the Jewish Councils.

As comfortable as it would be to forget about all of this, the Israeli government had the opposite incentive. Their goal wasn't just to broadcast the horrors of the Holocaust. It was to send the message that Jews who believed they were safe among Gentiles were fools, and Jews who wanted to negotiate and concede points in their conflicts with Gentiles were collaborators.

I was struck by Arendt’s psychological profile of the Israeli leadership. The year is 1960. David ben Gurion is seventy-four, near the end of a long life of military struggle. The Israeli leadership is still very much of the generation that survived World War II, the Israeli War of Independence, and the Holocaust. But they’re starting to realize that this will not always be true. The younger generation just attaining voting age doesn’t remember the Holocaust at all. Everyone knows their history, but not everybody *knows* it. And the people, maybe new immigrants from America, who didn’t go through the Holocaust, they start asking – do we really need a purely Jewish nation? Do we really have to be so hostile and suspicious of Gentiles all the time? Does the country have to be quite so heavily militarized? Maybe we should just be a normal peaceful friendly member of the community of nations a bit more?

And as Arendt tells it, Ben-Gurion and his colleagues felt like they had this driving duty to communicate the incommunicable truth that *this was not going to work*. They felt like this was an endlessly seductive position, that maybe they had been seduced by it themselves when they were younger, but that bitter experience had taught them that had to be rejected utterly. If they dwelt on the failures of the Jewish Councils of Europe a little too long, if maybe they were a little unfair to people who had lost in lose-lose ethical dilemmas, it was because they didn’t know how else to tell younger Jews *not to let themselves be those people*. I guess the active construction of a cultural payload of reflexive resistance bordering on paranoia, capable of being handed down to younger generations, helps explain a lot about Israeli history.

VI.

Arendt dwells on the obvious question: why didn’t people say no?

She had already given part of the answer. Some people did say no. The entire populations of Denmark and Bulgaria. Most of Italy. France, eventually, with prodding. Shouldn’t Germany have been filled with some of the same people?

She says no. At every point, she stressed how little genuine opposition Hitler had. It wasn’t just the Nazis’ sky-high approval rating. It was that even the people who hated the Nazis, loathed the Nazis, generally didn’t mention the Jewish genocide. Even the conspirators in the von Stauffenberg plot to assassinate Hitler were mostly angry that he was a bad commander and was probably going to lose. This opinion was at least within the Nazi Overton Window. That they should resist the Holocaust seems to barely have occurred to them.

One of the creepiest passages in the book described the Wannsee Conference. Hitler and Himmler and a few other highers-up had decided on the Final Solution; a policy change from forced emigration to extermination. They wanted to inform the civil service of their decision, but they expected trouble:

> The problem was much more acute, however, with respect to the higher career men in the Civil Service, directly under the Ministers, for these men, the backbone of every government administration, were not easily replaceable, and Hitler had tolerated them, just as Adenauer was to tolerate them, unless they were compromised beyond salvation. Hence the undersecretaries and the legal and other experts in the

various Ministries were frequently not even Party members, and Heydrich's apprehensions about whether he would be able to enlist the active help of these people in mass murder were quite comprehensible. As Eichmann put it, Heydrich "expected the greatest difficulties." Well, he could not have been more wrong.

The aim of the conference was to coordinate all efforts toward the implementation of the Final Solution. The discussion turned first on "complicated legal questions," such as the treatment of half- and quarter-Jews – should they be killed or only sterilized? This was followed by a frank discussion of the "various types of possible solutions to the problem," which meant the various methods of killing, and here, too, there was more than "happy agreement on the part of the participants"; the Final Solution was greeted with "extraordinary enthusiasm" by all present, and particularly by Dr. Wilhelm Stuckart, Undersecretary in the Ministry of the Interior, who was known to be rather reticent and hesitant in the face of "radical" Party measures, and was, according to Dr. Hans Globke's testimony at Nuremberg, a staunch supporter of the Law. There were certain difficulties, however. Undersecretary Josef Bühler, second in command in the General Government in Poland, was dismayed at the prospect that Jews would be evacuated from the West to the East, because this meant more Jews in Poland, and he proposed that these evacuations be postponed and that "the Final Solution be started in the General Government, where no problems of transport existed." The gentlemen from the Foreign Office appeared with their own carefully elaborated memorandum, expressing "the desires and ideas of the Foreign Office with respect to the total solution of the Jewish question in Europe," to which nobody paid much attention. The main point, as

> Eichmann rightly noted, was that the members of the various branches of the Civil Service did not merely express opinions but made concrete propositions. The meeting lasted no more than an hour or an hour and a half, after which drinks were served and everybody had lunch – “a cozy little social gathering.”

There were occasional protests about killing Jews because it would get Germany in big trouble if they lost the war. This was discussed seriously, always with the point being made that the Allies would view it as a dire crime, never with anybody stopping to ask whether maybe the Allies were right.

By far the most successful movement in this direction, one I had never heard about before, was Musy’s meeting with Himmler. A few months before the war ended, some Jews in Switzerland happened to meet Jean-Marie Musy, one of Himmler’s childhood friends; they asked him to use his influence with the Nazi second-in-command to get him to stop killing Jews. Musy went to Germany and told Himmler that they both knew the Allies were winning the war, that he’d heard the Allies were really mad about the Holocaust, and that maybe if Himmler stopped the Holocaust he could get better treatment after the war. Himmler thought about it for a little bit, *and he agreed*. He ordered the gas chambers destroyed, he countermanded Hitler’s directive to kill as many Jews as possible before the Allies liberated the camps, and started transporting Jews out of Nazi territory by trains. Hitler heard about this and got enraged and ordered everybody to stop listening to Himmler. There was a brief period of confusion as the two highest-ranking Nazis gave opposite orders, and then the Allies liberated the concentration camps anyway and the point became moot. It wasn’t much. Himmler’s order probably saved a few tens of thousands of people, out of millions. But it was something.

And there were countless smaller incidents like this. And they all shared one thing in common: they succeeded by appealing to Nazis’ self-interest, not to their conscience.

And so:

> As Eichmann told it, the most potent factor in the soothing of his own conscience was the simple fact that he could see no one, no one at all, who actually was against the Final Solution.

And where this wasn’t true, people started developing consciences again. Nazi commanders who had been in Denmark for long enough started to go native. When Himmler’s agents came to crack down on Danish resistance, they found the local German officials somewhere between hesitant and actively obstructionist:

> Politically and psychologically, the most interesting aspect of this incident is perhaps the role
> played by the German authorities in Denmark, their obvious sabotage of orders from Berlin. It is the only case we know of in which the Nazis met with open native resistance, and the result seems to have been that those exposed to it changed their minds. They themselves apparently no longer looked upon the extermination of a whole people as a matter of course. They had met resistance based on principle, and their “toughness” had melted like butter in the sun, they had even been able to show a few timid beginnings of genuine courage.

And:

> The same thing happened in Bulgaria as was to happen in Denmark a few months later – the local German

> officials became unsure of themselves and were no longer reliable.

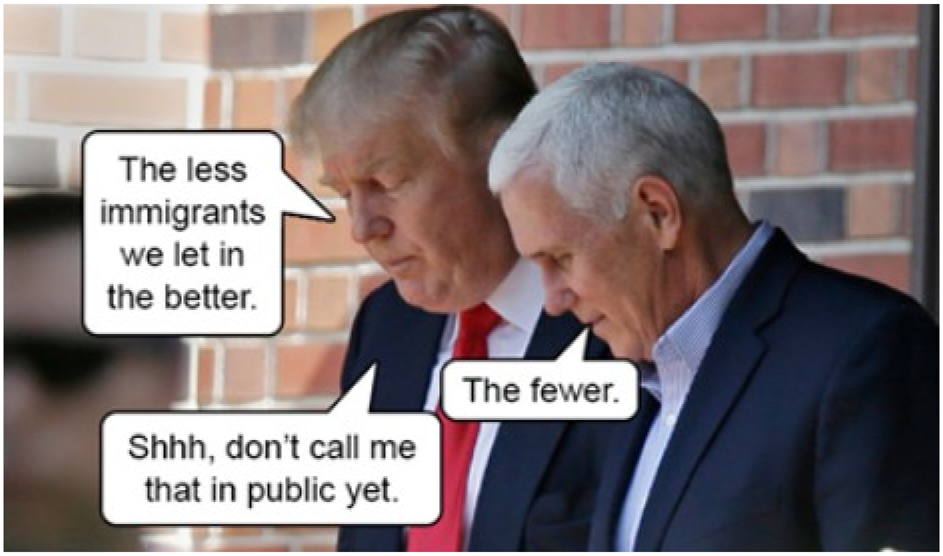
VII.

Are there any lessons to be learned from all of this horror?

First, the refugee aspect of all of this is even more important than I thought. I said it before, but I think it bears more emphasis. The Western nations' failure to accept refugees from Nazi Germany didn't just kill a couple of Jews who made it out before the killing started. Germany started off perfectly willing to let every single Jew in Europe emigrate to any country that would take them. Nowhere would. This obviously doesn't absolve the Nazis of any blame, but it sure doesn't make the rest of the world look very good either.

Second, it's worth remembering that the Final Solution was the Nazis' third or fourth plan, not their first. Eichmann argued that this ought to humanize him; sure, he wanted Germany *Judenfrei*, but at least he had the decency to try to do it humanely before moving on to genocide. But even if he's right, humanizing Nazis is a two-way street. The more human and comprehensible the Nazis' evil becomes, the closer it gets to the lesser evils of our own day. White separatists complain that they are misrepresented; that they have no intention of killing anybody, that they just want to help everybody get the right to live separately among their own people. I accept that they believe that and that it is unfair to misrepresent them. But having acknowledged their position, the next step is to acknowledge that the Nazis seem to have genuinely believed that too. For a while.

I’ve written before on how the current crop of demagogues, as bad as they are, aren’t Literally Hitler. But this should be understood in context of Mussolini not being Literally Hitler, or even of the Nazis themselves not being Literally Hitler at the beginning. The cause for concern isn’t that anyone you can see on TV today is plotting a Fourth Reich. It’s that some common factor causes people who start out as only moderately objectionable to predictably become something much worse. And modern populists share a suspicious number of characteristics and policies with their WWII-era fascist analogues (though “fascist” is the wrong word here; remember that Mussolini’s Italy did a better job saving Jews than a lot of the supposed ‘good guys’), and one can rightly be afraid that they’re drawing from the same underlying natural kind.


The less immigrants we let in the better.
The fewer.
Shhh, don't call me that in public yet.

This is exactly the sort of thing I should resist the urge to put here (source)

Third, at least during World War II conscience was a collective phenomenon. Why did some countries’ citizens cooperate almost universally with the Final Solution, while others resisted it at every turn? “Culture” is inadequate; there’s not much light between Danish and German culture, but the two countries acted in opposite ways. I’m tempted to credit single individuals; Hitler setting the tone for Germany vs. King Christian setting the tone for Denmark – but do people really respect their leaders that thoroughly? Or is this backwards causation; a country like Denmark would end up with a King like Christian, a country like Germany would elect a Fuhrer like Hitler? I don’t know. The alternative is to posit one of those chaotic networks where tiny differences in initial conditions can compound and lead to very different end states. Arendt herself offers little, beyond saying that Italy saved its Jews out of “the automatic general humanity of an old and civilized people”. Yeah, well, Japan was an old and civilized people too, and we know how that turned out. But what other possibilities are there? All I can think of is maybe looking into the pre-existing anti-Semitism level, but I don’t know if that just passes the explanatory buck.

Did you know the Intro Psych explanation of Asch’s conformity experiments gets them backwards? Although it’s true that in each experiment a few people would conform with majority opinion, the majority of subjects *didn’t* conform and stuck with the evidence of their own eyes. This is encouraging, but makes the international variation in behavior even more perplexing. Whatever the cause, despite some heroic individuals everywhere, the between-country variance was more important than the within-country variance.

Fourth, resistance worked. Not for the Jews, who generally had no good options. But for the Gentile population of occupied countries, absolutely. It didn’t need heroic martyrs willing to stand in front of Panzers Tiananmen-style. It just took a general attitude of annoying obstructionism. The Germans said “Give us a list of all the Jews in your country by next week,” and the police said “Oh, yeah, sure”, and then the next week the Germans asked where their list was, and the police said, “Sorry, we must have forgot.” When the attitude was so universal that the Nazis didn’t know who to punish, or didn’t dare punish everyone for fear of rebellion, they generally gave up.

This isn’t to trivialize anything. There were thousands of individuals who died horribly resisting the Nazis, often to no avail. But when whole countries and cultures decided to resist, it made a big difference.

Even more – and I think Arendt’s frequent repetition of this fact is entirely justified – it started to change the Nazis’ minds. The Nazi officials in Denmark and Bulgaria became just a little bit obstructionist themselves. Nothing spectacular. No throwing off their jackboots and joining the resistance. Just a very slight tendency to question what was going on and ask “Are we the baddies?”

Just as humanizing the Nazis is a two-way street, so pointing out the bizarre lack of dissent in Nazi Germany is both distressing and encouraging. Distressing because – how could ordinary humans tolerate that? But encouraging because – well, it seems almost possible to imagine a world where something goes wrong and America ends up overtly fascist. Yet even in my worst nightmares I can’t imagine a world where America ends up overtly fascist and nobody is annoying and obstructionist about it. Arendt’s picture of Germany, where the ruling party has 90% approval and dissent is unthinkable – you can’t get there from here. We’re never unanimous about *anything*.

I thank G-d for the annoying obstructionists, for the nitpickers, for the devil’s advocates, for the people who hear something that’s obviously true and strain to come up with an absurd thought experiment where it might not be, for the reflexive contrarians, for the people who always vote third party, for the people who urge you to sign petitions on whitehouse.gov because “then the President *has to* respond”, for the people who have two hundred guns in their basement “just in case”, for the people who say “well, actually…” all the time, for the mayors of sanctuary cities and the clerks who refuse to perform gay weddings, for the people who think being banned on Twitter is a violation of their human rights, and for the people who swear eternal hostility to other people on the same side who agree with them on 99% of everything. On the spectrum from “totally ungovernable” to “vulnerable to Nazism”, I think that we’ve erred in the right direction.

[GUEST POST] The International Refugee Assistance Project

[This is a guest post by Elizabeth Van Nostrand of Aceso Under Glass.]

How do you judge the effectiveness of preparing for something that might never happen? Consider emergency room capacity; if you have enough space and staff to deal with a diphtheria epidemic, it will seem wasted during the 99% of the time there isn't one. But this is false economy; during a crisis the extra reserve may be the difference between disruption and disaster.

Whatever the staff of the ACLU was doing on Friday morning, it wasn't as important as what they did Friday afternoon, after Trump banned legal immigrants from reentering the country, even if they were in flight. If you just evaluated the effectiveness of their Friday morning work, you'd miss that they were also holding the capacity to respond to Trump's order as quickly as they did. That speed mattered – not just to residents being badgered into giving up their green cards, but because it demonstrated to Trump and America that the consequences of violating the Constitution are severe and immediate. Once the crisis hit, it became easy to see how everything that went into preparing for it was high leverage – which was why people donated $20,000,000 over the weekend.

But the ACLU didn't act alone. One partner was the International Refugee Assistance Project, founded by Becca Heller. While in Jordan on an internship, Heller talked to Iraqi refugees there whose applications to emigrate to the US had been denied; they had given up because there was no way to appeal Homeland Security decisions. But rejected applicants can effectively appeal by asking that

DHS “reconsider” their application. Using case law from asylum decisions, Becca filed requests to reconsider on behalf of these Iraqi refugees. Back home, she began recruiting other Yale law students, then students at other law schools. As her organization grew in scope, it effected some at-the-margins policy work as well. A notable win here was using FOIA to force the DHS to release their refugee decision manual. Last year IRAP received a grant from the Open Philanthropy Project to increase the extent of their policy work.

The majority of IRAP’s day to day work involves processing appeals for individual refugees whose requests to resettle in the U.S. were denied. Their staff manages a team of volunteer law students and lawyers who pour hours into interviewing clients, investigating cases, collecting documents to prove refugee status, and writing appeals. This aspect of their work isn’t glamorous. It’s not arguing before the Supreme Court to establish rights people didn’t have a week before. It’s tracking down proof that someone is from the country they say they are, when 95% of record-keeping institutions in the country have been blown up and the procedure for getting a passport is to find Mr. Big Beard’s stall in the market and give him some money. It’s very ground-level work that makes an enormous difference in the lives of individual immigrants. But each person helped in this way represents the investment of many, many hours from lawyers and law students; if you were analyzing their effectiveness, they’d probably score poorly relative to AMF (especially if you included the value of in kind donations of time from skilled individuals). Where they did advocacy work, it was relatively quiet and uncontroversial.

And then Trump issued executive order “PROTECTING THE NATION FROM FOREIGN TERRORIST ENTRY INTO THE

UNITED STATES”

Under American law, an arbitrary person can’t challenge a government policy or decision in court. You have to have standing, meaning you personally are hurt by the action. And to be represented by a lawyer, you have to say “I am represented by this lawyer,” which normally isn’t a problem, *unless* you’re being held incommunicado and no one knows where you are. There are ways around this, but they’re time consuming.

IRAP’s ground level work meant that when Trump’s order went out, it took them approximately four seconds to create a list of extremely sympathetic/photogenic immigrants who would be caught at airports that day, for some of whom they were already the legal representative of record. It took another six seconds to find even more photogenic supporters of those immigrants who were willing to loudly support them in media interviews. Veterans are Not Fucking Around when it comes to their translators.

IRAP doesn’t have a long history of challenging the federal government the way the ACLU does. But by partnering with the ACLU (with its very strong history of challenging the federal government), the National Immigration Law Center (extensive knowledge of immigration policy) and Yale Law School (I don’t know, cheap labor from students? Money?), they were able to launch the strongest possible campaign to force Trump to keep promises we made as a country.

Also, for someone whose organization spends most of its time on individual cases, Becca Heller sure is good at publicly nailing the government to the wall. I want her on this team just so she can keep going on TV and talking over congresspeople who she deems insufficiently proactive.

Ultimately I think the part where Trump flat out ignored the judicial branch is more important than changes to immigration (although I would like to see us expand immigration), and the ACLU is best positioned to fight constitutional battles. But the ACLU will need similar assistance from ground-level organizations for the next fight too, and they just got twenty million dollars. Yale Law School needs your money even less. IRAP, by contrast, has an annual budget of $2 million. I exhausted my charity budget at the beginning of the year on my personal cause area, Third World poverty, but my ability to continue giving is dependent on the United States having a functional government and economy. I am genuinely afraid Trump will destroy both of those. I’m donating $100 to IRAP to support the upcoming legal fight, and as a retroactive bonus for building this capacity before most of us knew how much it was needed.

I encourage others to donate to IRAP, or to find similar small charities doing ground level work.

Notes From The Asilomar Conference On Beneficial AI

Last month I got to attend the Asilomar Conference on Beneficial AI. I tried to fight it off, saying I was totally unqualified to go to any AI-related conference. But the organizers assured me that it was an effort to bring together people from diverse fields to discuss risks ranging from technological unemployment to drones to superintelligence, and so it was totally okay that I'd never programmed anything more complicated than HELLO WORLD.

"Diverse fields" seems right. On the trip from San Francisco airport, my girlfriend and I shared a car with two computer science professors, the inventor of Ethereum, and a UN chemical weapons inspector. One of the computer science professors tried to make conversion by jokingly asking the weapons inspector if he'd ever argued with Saddam Hussein. "Yes," said the inspector, not joking at all. The rest of the conference was even more interesting than that.

I spent the first night completely star-struck. Oh, that's the founder of Skype. Oh, those are the people who made AlphaGo. Oh, that's the guy who discovered the reason why the universe exists at all. This might have left me a little tongue-tied. How do you introduce yourself to eg David Chalmers? "Hey" seems insufficient for the gravity of the moment. "Hey, you're David Chalmers!" doesn't seem to communicate much he doesn't already know. "Congratulations on being David Chalmers", while proportionate, seems potentially awkward. I just don't have an appropriate social script for this situation.

(the problem was resolved when Chalmers saw me staring at him, came up to me, and said "Hey, are you the guy who writes Slate Star Codex?”)

The conference policy discourages any kind of blow-by-blow description of who said what in order to prevent people from worrying about how what they say will be “reported” later on. But here are some general impressions I got from the talks and participants:

1. In part the conference was a coming-out party for AI safety research. One of the best received talks was about “breaking the taboo” on the subject, and mentioned a postdoc who had pursued his interest in it secretly lest his professor find out, only to learn later that his professor was also researching it secretly, lest everyone *else* find out.

The conference seemed like a (wildly successful) effort to contribute to the ongoing normalization of the subject. Offer people free food to spend a few days talking about autonomous weapons and biased algorithms and the menace of AlphaGo stealing jobs from hard-working human Go players, then sandwich an afternoon on superintelligence into the middle. Everyone could tell their friends they were going to hear about the poor unemployed Go players, and protest that they were only listening to Elon Musk talk about superintelligence because they happened to be in the area. The strategy worked. The conference attracted AI researchers so prestigious that even *I* had heard of them (including many who were publicly skeptical of superintelligence), and they all got to hear prestigious people call for “breaking the taboo” on AI safety research and get applauded. Then people talked about all of the lucrative grants they had gotten in the area. It did a great job of creating common knowledge that everyone agreed AI goal alignment research was valuable, in a way not entirely constrained by whether any such agreement actually existed.

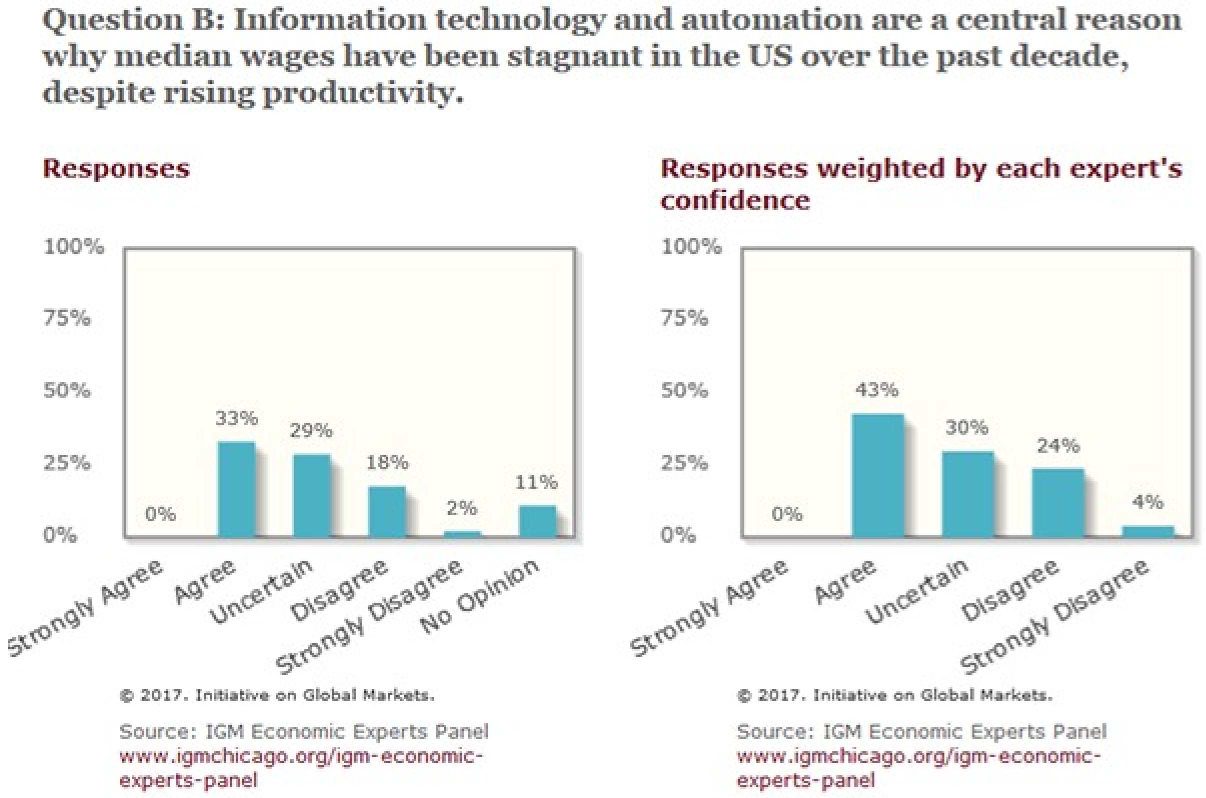
2. Most of the economists there seemed pretty convinced that technological unemployment was real, important, and happening already. A few referred to Daron Acemoglu's recent paper Robots And Jobs: Evidence From US Labor Markets, which says:

> We estimate large and robust negative effects of robots on employment and wages. We show that commuting zones most affected by robots in the post-1990 era were on similar trends to others before 1990, and that the impact of robots is distinct and only weakly correlated with the prevalence of routine jobs, the impact of imports from China, and overall capital utilization. According to our estimates, each additional robot reduces employment by about seven workers, and one new robot per thousand workers reduces wages by 1.2 to 1.6 percent.

And apparently last year's Nobel laureate Angus Deaton said that:

> Globalisation for me seems to be not first-order harm and I find it very hard not to think about the billion people who have been dragged out of poverty as a result. I don't think that globalisation is anywhere near the threat that robots are.

A friend reminded me that the kind of economists who go to AI conferences might be a biased sample, so I checked IGM's Economic Expert Panel (now that I know about that I'm going to use it for *everything*):



Question B: Information technology and automation are a central reason why median wages have been stagnant in the US over the past decade, despite rising productivity.

It looks like economists are uncertain but lean towards supporting the theory, which really surprised me. I thought people were still talking about the Luddite fallacy and how it was impossible for new technology to increase unemployment because something something sewing machines something entire history of 19th and 20th centuries. I guess that’s changed.

I had heard the horse used as a counterexample to this before – ie the invention of the car put horses out of work, full stop, and now there are fewer of them. An economist at the conference added some meat to this story – the invention of the stirrup (which increased horse efficiency) and the railroad (which displaced the horse for long-range trips) increased the number of horses, but the invention of the car decreased it. This suggests that some kind of innovations might complement human labor and others replace it. So a pessimist could argue that the sewing machine (or whichever other past innovation) was more like the stirrup, but modern AIs will be more like the car.

3. A lot of people there were really optimistic that the solution to technological unemployment was to teach unemployed West Virginia truck drivers to code so they could participate in the AI revolution. I used to think this was a weird straw man occasionally trotted out by Freddie deBoer, but all these top economists were *super enthusiastic* about old white guys whose mill has fallen on hard times founding the next generation of nimble tech startups. I'm tempted to mock this, but maybe I shouldn't – this From Coal To Code article says that the program has successfully rehabilitated Kentucky coal miners into Web developers. And I can't think of a good argument why not – even from a biodeterminist perspective, nobody's ever found that coal mining areas have lower IQ than anywhere else, so *some* of them ought to be potential web developers just like everywhere else. I still wanted to ask the panel "Given that 30-50% of kids fail high school algebra, how do you expect them to learn computer science?", but by the time I had finished finding that statistic they had moved on to a different topic.

4. The cutting edge in AI goal alignment research is the idea of inverse reinforcement learning. Normal reinforcement learning is when you start with some value function (for example, "I want something that hits the target") and use reinforcement to translate that into behavior (eg reinforcing things that come close to the target until the system learns to hit the target). Inverse reinforcement learning is when you start by looking at behavior and use it to determine some value function (for example, "that program keeps hitting that spot over there, I bet it's targeting it for some reason").

Since we can't explain human ethics very clearly, maybe it would be easier to tell an inverse reinforcement learner to watch the stuff humans do and try to figure out what values we're working off of – one obvious problem being that our values predict our actions much less than we might wish. Presumably this is solvable if we assume that our moral statements are also behavior worth learning from.

A more complicated problem: humans don't have utility functions, and an AI that assumes we do might come up with some sort of monstrosity that predicts human behavior really well while not fitting our idea of morality at all. Formalizing what exactly humans *do* have and what exactly it means to approximate that thing might turn out to be an important problem here.

5. Related: a whole bunch of problems go away if AIs, instead of receiving rewards based on the state of the world, treat the reward signal as information about a reward function which they only imperfectly understand. For example, suppose an AI wants to maximize "human values", but knows that it doesn't really understand human values very well. Such an AI might try to learn things, and if the expected reward was high enough it might try to take actions in the world. But it wouldn't (contra Omohundro) naturally resist being turned off, since it might believe the human turning it off understood human values better than it did and had some human-value-compliant reason for wanting it gone. This sort of AI also might not wirehead – it would have no reason to think that wireheading was the best way to learn about and fulfill human values.

The technical people at the conference seemed to think this idea of uncertainty about reward was technically possible, but would require a ground-up reimagining of reinforcement learning. If true, it would be a perfect example of what Nick

Bostrom et al have been trying to convince people of since forever: there are good ideas to mitigate AI risk, but they have to be studied early so that they can be incorporated into the field early on.

6. AlphaGo has gotten much better since beating Lee Sedol and its creators are now trying to understand the idea of truly optimal play. I would have expected Go players to be pretty pissed about being made obsolete, but in fact they think of Go as a form of art and are awed and delighted to see it performed at superhuman levels.

More interesting for the rest of us, AlphaGo is playing moves and styles that all human masters had dismissed as stupid centuries ago. Human champion Ke Jie said that:

> After humanity spent thousands of years improving our tactics, computers tell us that humans are completely wrong. I would go as far as to say not a single human has touched the edge of the truth of Go.

A couple of people talked about how the quest for “optimal Go” wasn’t just about one game, but about grading human communities. Here we have this group of brilliant people who have been competing against each other for centuries, gradually refining their techniques. Did they come pretty close to doing as well as merely human minds could manage? Or did non-intellectual factors – politics, conformity, getting trapped at local maxima – cause them to ignore big parts of possibility-space? Right now it’s very preliminarily looking like the latter, which would be a really interesting result – especially if it gets replicated once AIs take over other human fields.

One Go master said that he would have “slapped” a student for playing a strategy AlphaGo won with. Might we one day be able to do a play-by-play of Go history, finding out where human strategists went wrong, which avenues they closed unnecessarily, and what institutions and thought processes were most likely to tend towards the optimal play AlphaGo has determined? If so, maybe we could have have twenty or thirty years to apply the knowledge gained to our own fields before AIs take over those too.

7. People For The Ethical Treatment Of Reinforcement Learners got a couple of shout-outs, for some reason. One reinforcement learning expert pointed out that the problem was trivial, because of a theorem that program behavior wouldn't be affected by global shifts in reinforcement levels (ie instead of going from +10 to -10, go from +30 to +10). I'm not sure if I'm understanding this right, or if this kind of trick would affect a program's conscious experiences, or if anyone involved in this discussion is serious.

8. One theme that kept coming up was that most modern machine learning algorithms aren't "transparent" – they can't give reasons for their choices, and it's difficult for humans to read them off of the connection weights that form their "brains". This becomes especially awkward if you're using the AI for something important. Imagine a future inmate asking why he was denied parole, and the answer being "nobody knows and it's impossible to find out even in principle". Even if the AI involved were generally accurate and could predict recidivism at superhuman levels, that's a hard pill to swallow.

(DeepMind employs a Go master to help explain AlphaGo's decisions back to its own programmers, which is probably a metaphor for something)

This problem scales with the size of the AI; a superintelligence whose decision-making process is completely opaque sounds pretty scary. This is the “treacherous turn” again; you can train an AI to learn human values, and you can observe it doing something that looks like following human values, but you can never “reach inside” and see what it’s “really thinking”. This could be pretty bad if what it’s really thinking is “I will lull the humans into a false sense of complacency until they give me more power”. There seem to be various teams working on the issue.

But I’m also interested in what it says about us. Are the neurons in our brain some kind of uniquely readable agent that is for some reason transparent to itself in a way other networks aren’t? Or should we follow Nisbett and Wilson in saying that our own brains are an impenetrable mass of edge weights just like everything else, and we’re forced to guess at the reasons motivating our own cognitive processes?

9. One discipline I shouldn’t have been so surprised to see represented at the multidisciplinary conference was politics. A lot of the political scientists and lawyers there focused on autonomous weapons, but some were thinking about AI arms races. If anyone gets close to superintelligence, we want to give them time to test it for safety before releasing it into the wild. But if two competing teams are equally close and there’s a big first-mover advantage (for example, first mover takes over the world), then both groups will probably skip the safety testing.

On an intranational level, this suggests a need for regulation; on an international one, it suggests a need for cooperation. The Asilomar attendees were mostly Americans and Europeans, and some of them were pretty well-connected in their respective governments. But we realized we didn’t have the same kind of contacts in the

Chinese and Russian AI communities, which might help if we needed some kind of grassroots effort to defuse an AI arms race before it started. If anyone here is a Chinese or Russian AI scientist, or has contacts with Chinese or Russian AI scientists, please let me know and I can direct you to the appropriate people.

10. In the end we debated some principles to be added into a framework that would form a basis for creating a guideline to lay out a vision for ethical AI. Most of these were generic platitudes, like “we believe the benefits of AI should go to everybody”. There was a lunch session where we were supposed to discuss these and maybe change the wording and decide which ones we did and didn’t support.

There are lots of studies in psychology and neuroscience about what people’s senses do when presented with inadequate stimuli, like in a sensory deprivation tank. Usually they go haywire and hallucinate random things. I was reminded of this as I watched a bunch of geniuses debate generic platitudes. It was *hilarious*. I got the lowdown from a couple of friends who were all sitting at different tables, and everyone responded in different ways, from rolling their eyes at the project to getting really emotionally invested in it. My own table was moderated by a Harvard philosophy professor who obviously deserved it. He clearly and fluidly explained the moral principles behind each of the suggestions, encouraged us to debate them reasonably, and then led us to what seemed in hindsight the obviously correct answer. I had already read one of his books but I am planning to order more. Meanwhile according to my girlfriend other tables did everything short of come to blows.

Overall it was an amazing experience and many thanks to the Future of Life Institute for putting on the conference and inviting me. Some pictures (not taken by me) below:

The conference center

The panel on superintelligence.

Smart&Final.
Smart&Final
11221

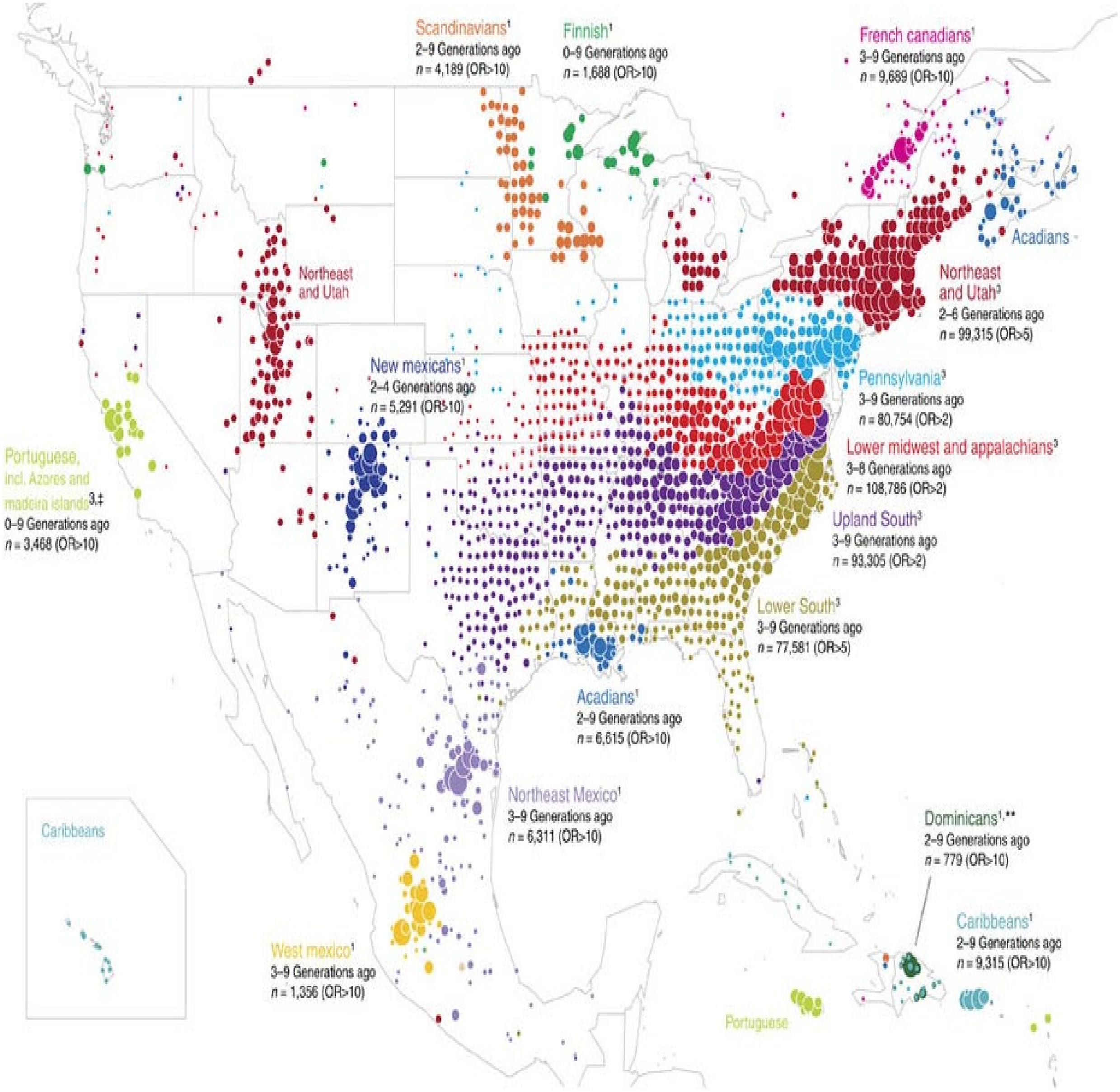
We passed this store on the way to the conference. I thought it was kabbalistically relevant to AI risk.

Albion’s Seed, Genotyped

Last year I reviewed *Albion’s Seed*, historian David Fischer’s work on the four great English migrations to America (and JayMan continues the story in his series on American Nations). These early migrations help explain modern regional patterns like why Massachusetts is so liberal or why Appalachia seems so backwards. As always, there’s the lingering question of how much of these patterns are cultural versus genetic versus gene-cultural interaction.

Now Han et al take this field high-tech with the publication of Clustering Of 770,000 Genomes Reveals Post-Colonial Population Structure Of North America (h/t gwern, werttrew)

The team looked at 770,000 genomes analyzed by the AncestryDNA company and used a technique called identity-by-descent to find recent common ancestors. Then they used some other techniques to divide them into natural clusters. This is what they got:


Scandinavians[1]
2–9 Generations ago
n = 4,189 (OR>10)
Finnish[1]
0–9 Generations ago
n = 1,688 (OR>10)
French canadians[1]
3–9 Generations ago
n = 9,689 (OR>10)
Acadians
Northeast and Utah
Northeast and Utah[3]
2–6 Generations ago
n = 99,315 (OR>5)
New mexicans[1]
2–4 Generations ago
n = 5,291 (OR>10)
Pennsylvania[3]
3–9 Generations ago
n = 80,754 (OR>2)
Lower midwest and appalachians[3]
3–8 Generations ago
n = 108,786 (OR>2)
Portuguese, incl. Azores and madeira islands[3,‡]
0–9 Generations ago
n = 3,468 (OR>10)
Upland South[3]
3–9 Generations ago
n = 93,305 (OR>2)
Lower South[3]
3–9 Generations ago
n = 77,581 (OR>5)
Acadians[1]
2–9 Generations ago
n = 6,615 (OR>10)
Northeast Mexico[1]
3–9 Generations ago
n = 6,311 (OR>10)
Caribbeans
Dominicans[1,**]
2–9 Generations ago
n = 779 (OR>10)
Caribbeans[1]
2–9 Generations ago
n = 9,315 (OR>10)
West mexico[1]
3–9 Generations ago
n = 1,356 (OR>10)
Portuguese

This is the European-settler-focused map – there's another one focusing on immigrant groups lower down here

This is kind of beautiful. While not exactly matching *Albion's Seed*, it at least clearly shows its New Englander and Pennsylvania Quaker migrations (more realistically the Germans who came along with the Quakers), with less distinct signals for Borderers and Virginians. It shows how they spread directly west from their place of origin in almost exactly the way *American Nations* predicted. It even confirms my own conjecture that the belt of Democrat voters along southern Michigan corresponds to an area of New Englander settlement there (see part III here, or search “linear distribution of blue”). And it confirms Razib Khan’s observation that the Mormons are just displaced New Englanders and that their various unusual demographic features make sense in that context.

My biggest confusion is in the Southern/Appalachian region. I think Fischer would have predicted two distinct strains: a Tidewater/Virginian population along the coasts, and a Borderer/Appalachian population centered in West Virginia and Kentucky. Instead there are three populations, all of which start along the Atlantic Coast and continue inland in about the same way. Assuming red/”Appalachian” is the Borderers, I don’t know if Fischer has a good explanation for the purple/”upland south” vs. gold/”lower south” distinction. Nor I do get understand why, if one of those two represent the Tidewater Virginians, they don’t seem to be in the Tidewater Virginia region (which here is inhabited mostly by Borderers). Maybe this has something to do with the Civil War, or with the growth of the DC suburbs?

(And I guess we still haven’t ruled out the maximally boring explanation that interbreeding is entirely geographic and north-south is a bigger distinction than east-west so we’re just seeing the country divided into five equal-sized latitudinal bands.)

Not exactly a confusion, but more a disappointment: this map doesn’t provide the confirmation I’d hoped for that Californians, Seattleites, and other “Left Coasters” are displaced New Englanders – which would complete the circle of “Liberal Democrats = Puritan/Quaker population subgroup”. It’s also disappointing how little data they have for the Mountain West in general; I don’t know if that’s because there weren’t enough people there to show up, or because they’re a mix of every genetic lineage and don’t fit into any of the clusters nicely.

Still, I find this a really elegant example of hard science confirming historical speculation. Thanks to everyone who brought it to my attention.

[EDIT: Jayman goes much more into depth on this]

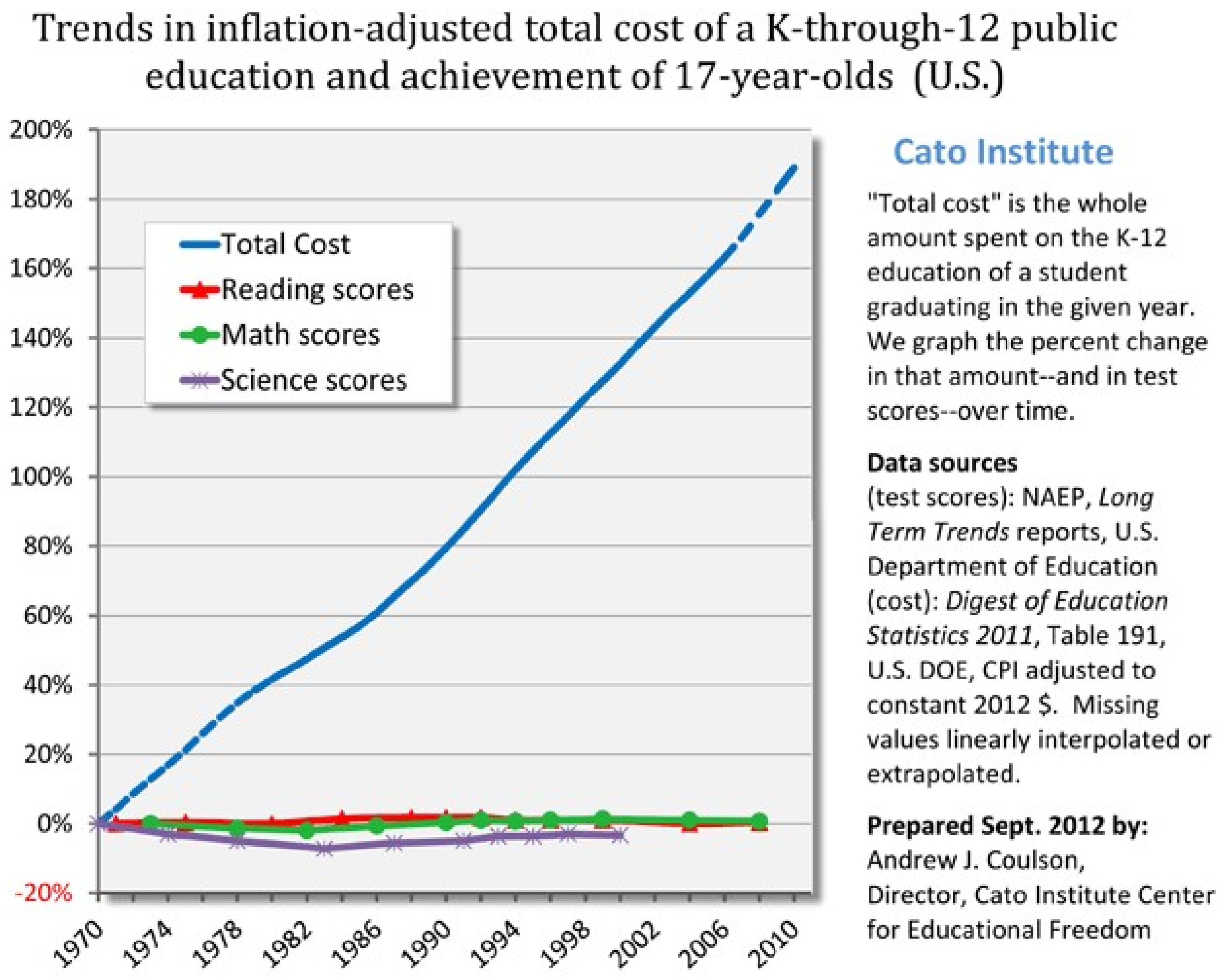
Considerations On Cost Disease

I.

Tyler Cowen writes about cost disease. I'd previously heard the term used to refer only to a specific theory of why costs are increasing, involving labor becoming more efficient in some areas than others. Cowen seems to use it indiscriminately to refer to increasing costs in general – which I guess is fine, goodness knows we need a word for that.

Cowen assumes his readers already understand that cost disease exists. I don't know if this is true. My impression is that most people still don't know about cost disease, or don't realize the extent of it. So I thought I would make the case for the cost disease in the sectors Tyler mentions – health care and education – plus a couple more.

First let's look at primary education:



Trends in inflation-adjusted total cost of a K-through-12 public education and achievement of 17-year-olds (U.S.)

"Total cost" is the whole amount spent on the K-12 education of a student graduating in the given year. We graph the percent change in that amount--and in test scores--over time.

Data sources
(test scores): NAEP, *Long Term Trends* reports, U.S. Department of Education
(cost): *Digest of Education Statistics 2011*, Table 191, U.S. DOE, CPI adjusted to constant 2012 $. Missing values linearly interpolated or extrapolated.

Prepared Sept. 2012 by:
Andrew J. Coulson,
Director, Cato Institute Center for Educational Freedom

There was some argument about the style of this graph, but as per Politifact the basic claim is true. Per student spending has increased about 2.5x in the past forty years even after adjusting for inflation.

At the same time, test scores have stayed relatively stagnant. You can see the full numbers here, but in short, high school students' reading scores went from 285 in 1971 to 287 today – a difference of 0.7%.

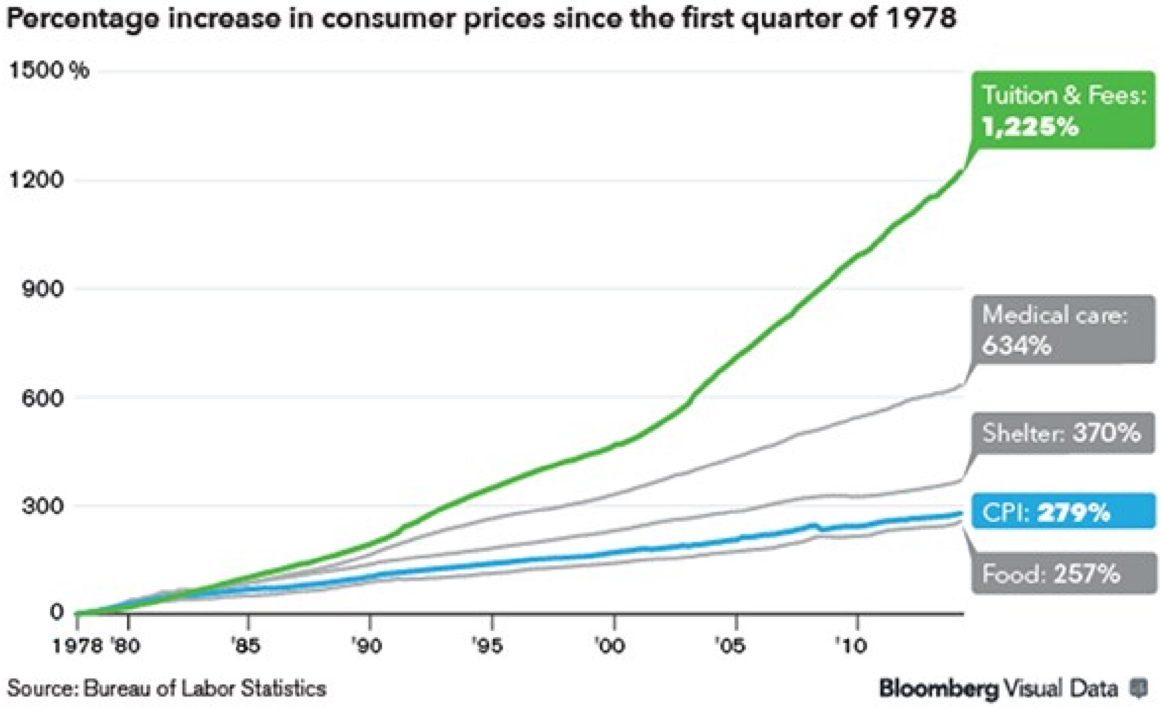
There is some heterogenity across races – white students' test scores increased 1.4% and minority students' scores by about 20%. But it is hard to credit school spending for the minority students' improvement, which occurred almost entirely during the period from 1975-1985. School spending has been on exactly the same trajectory before and after that time, and in white and minority areas, suggesting that there was something specific about that decade which improved minority (but not white) scores. Most likely this was the general improvement in minorities' conditions around that time, giving them better nutrition and a more stable family life. It's hard to construct a narrative where it was school spending that did it – and even if it did, note that the majority of the increase in school spending happened from 1985 on, and demonstrably helped neither whites *nor* minorities.

I discuss this phenomenon more here and here, but the summary is: no, it's not just because of special ed; no, it's not just a factor of how you measure test scores; no, there's not a "ceiling effect". Costs really did more-or-less double without any concomitant increase in measurable quality.

So, imagine you're a poor person. White, minority, whatever. Which would you prefer? Sending your child to a 2016 school? Or sending your child to a 1975 school, and getting a check for $5,000 every year?

I'm proposing that choice because as far as I can tell that *is* the stakes here. 2016 schools have whatever tiny test score advantage they have over 1975 schools, and cost $5000/year more, inflation adjusted. That $5000 comes out of the pocket of somebody – either taxpayers, or other people who could be helped by government programs.

Second, college is even worse:



Note this is not adjusted for inflation; see link below for adjusted figures

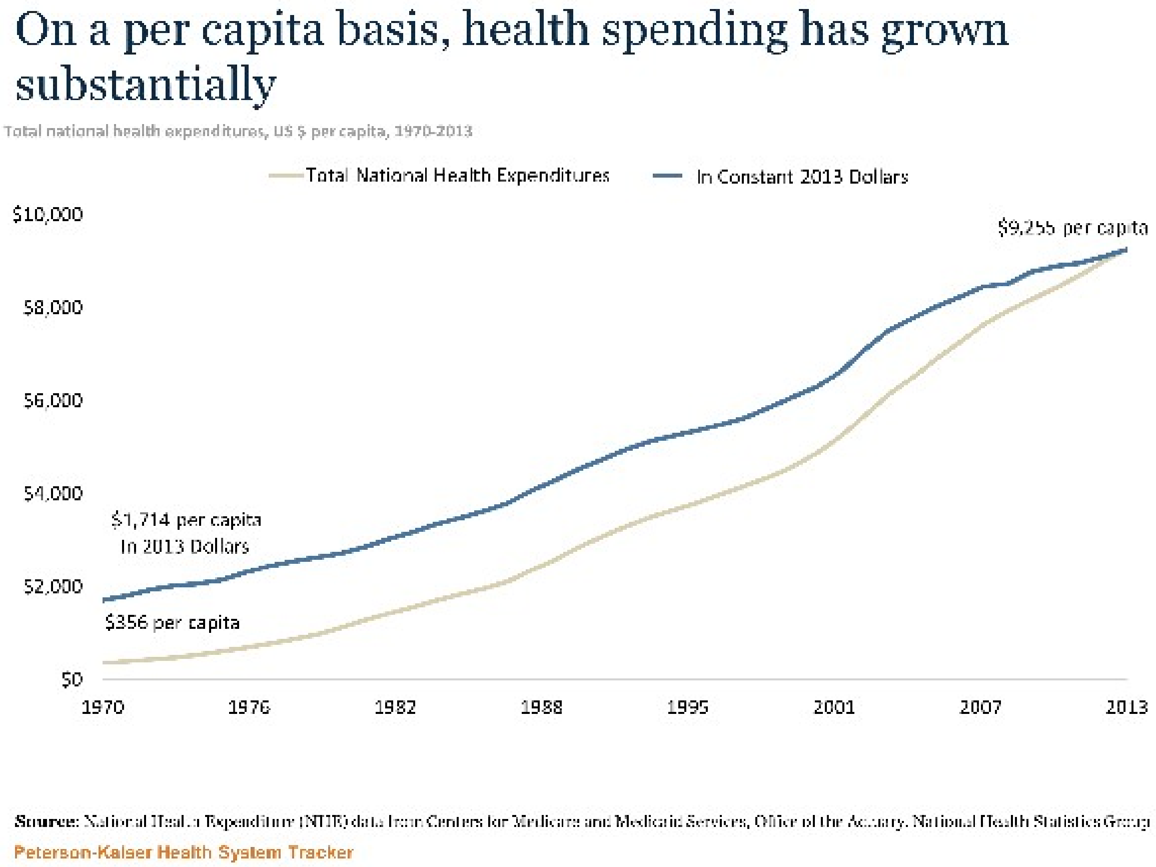
Inflation-adjusted cost of a university education was something like $2000/year in 1980. Now it's closer to $20,000/year. No, it's not because of decreased government funding, and there are similar trajectories for public and private schools.

I don't know if there's an equivalent of "test scores" measuring how well colleges perform, so just use your best judgment. Do you think that modern colleges provide $18,000/year greater value than colleges did in your parents' day? Would you rather graduate from a modern college, or graduate from a college more like the one your parents went to, plus get a check for $72,000?

(or, more realistically, have $72,000 less in student loans to pay off)

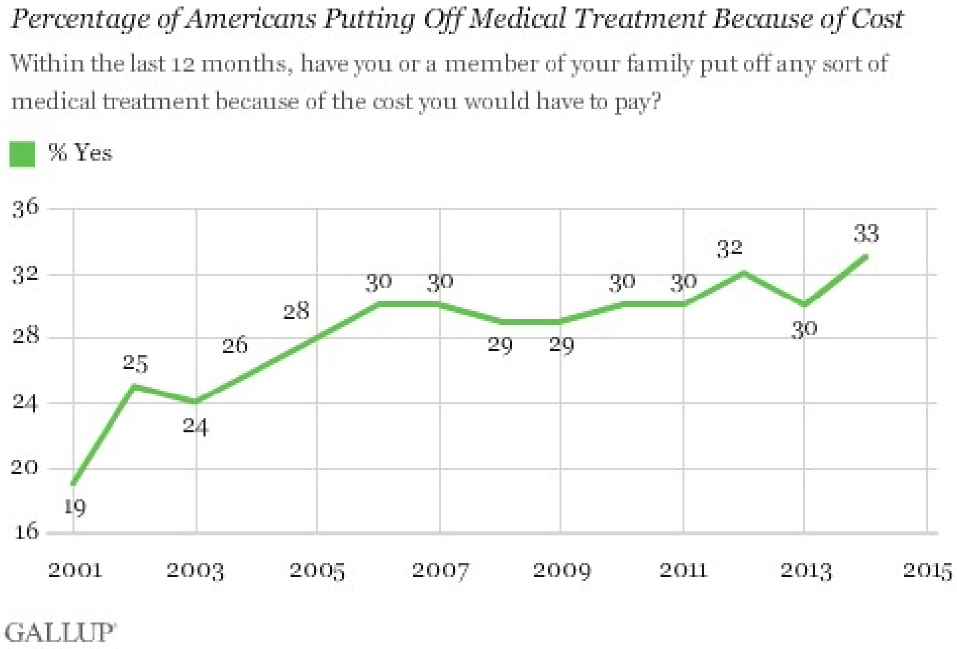
Was your parents' college even noticeably worse than yours? My parents sometimes talk about their college experience, and it seems to have had all the relevant features of a college experience. Clubs. Classes. Professors. Roommates. I might have gotten something extra for my $72,000, but it's hard to see what it was.

Third, health care. The graph is starting to look disappointingly familiar:



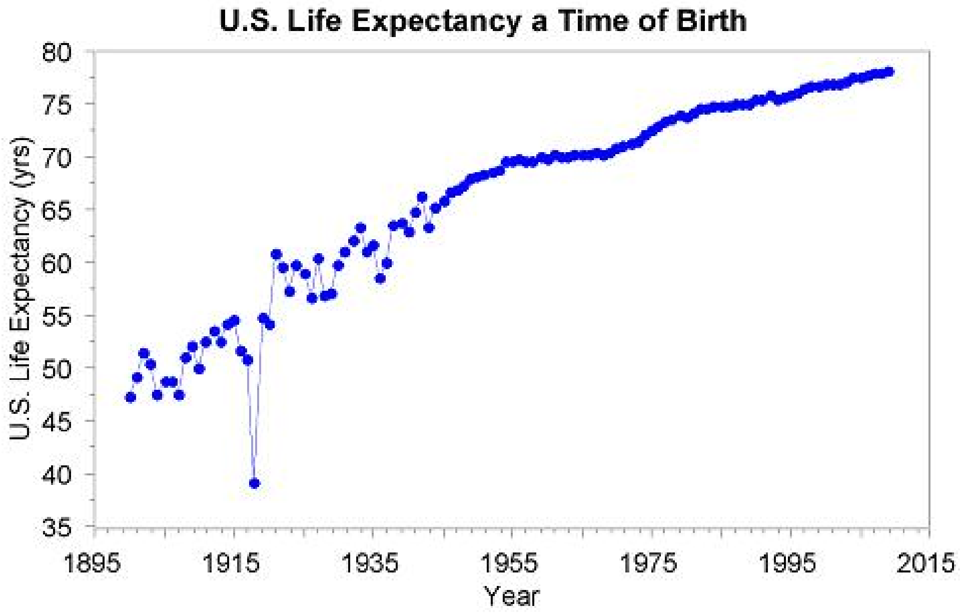
The cost of health care has about quintupled since 1970. It's actually been rising since earlier than that, but I can't find a good graph; it looks like it would have been about $1200 in today's dollars in 1960, for an increase of about 800% in those fifty years.

This has had the expected effects. The average 1960 worker spent ten days’ worth of their yearly paycheck on health insurance; the average modern worker spends sixty days’ worth of it, a sixth of their entire earnings.



Or not.

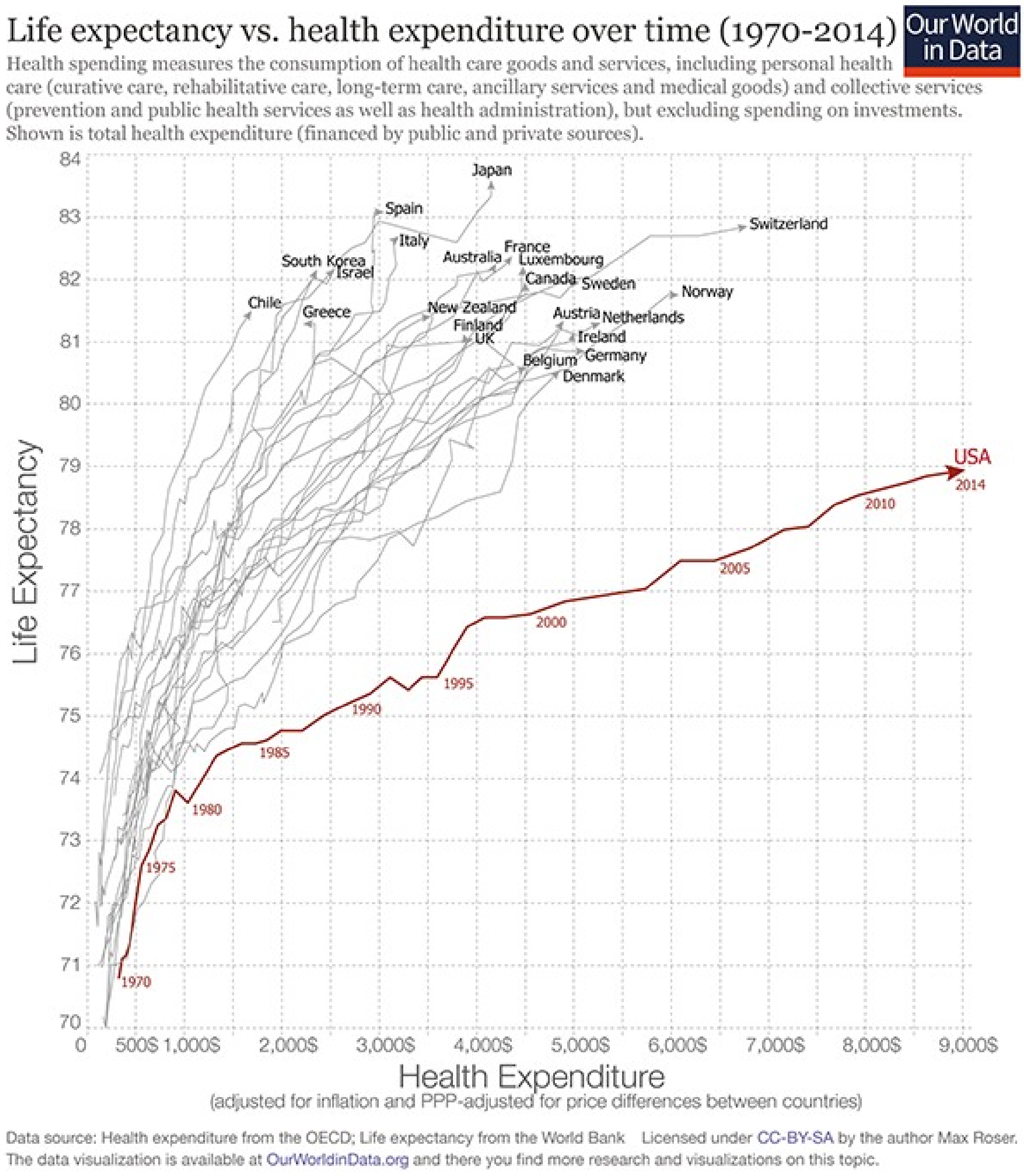
This time I can’t say with 100% certainty that all this extra spending has been for nothing. Life expectancy has gone way up since 1960:



Extra bonus conclusion: the Spanish flu was really bad

But a lot of people think that life expectancy depends on other things a lot more than healthcare spending. Sanitation, nutrition, quitting smoking, plus advances in health technology that don't involve spending more money. ACE inhibitors (invented in 1975) are great and probably increased lifespan a lot, but they cost $20 for a year's supply and replaced older drugs that cost about the same amount.

In terms of calculating how much lifespan gain healthcare spending has produced, we have a couple of options. Start with by country:



Life expectancy vs. health expenditure over time (1970-2014)

Health spending measures the consumption of health care goods and services, including personal health care (curative care, rehabilitative care, long-term care, ancillary services and medical goods) and collective services (prevention and public health services as well as health administration), but excluding spending on investments. Shown is total health expenditure (financed by public and private sources).

Our World in Data
Life Expectancy
84
83
82
81
80
79
78
77
76
75
74
73
72
71
70
Japan
Spain
Italy
Switzerland
South Korea
Israel
Australia
France
Luxembourg
Canada
Sweden
Chile
Greece
New Zealand
Finland
UK
Norway
Austria
Netherlands
Ireland
Germany
Belgium
Denmark
USA
2014
2010
2005
2000
1995
1990
1985
1980
1975
1970
0
500$
1,000$
2,000$
3,000$
4,000$
5,000$
6,000$
7,000$
8,000$
9,000$
Health Expenditure
(adjusted for inflation and PPP-adjusted for price differences between countries)

Data source: Health expenditure from the OECD; Life expectancy from the World Bank Licensed under CC-BY-SA by the author Max Roser.
The data visualization is available at OurWorldinData.org and there you find more research and visualizations on this topic.

Countries like South Korea and Israel have about the same life expectancy as the US but pay about 25% of what we do. Some people use this to prove the superiority of centralized government health systems, although Random Critical Analysis has an alternative perspective. In any case, it seems very possible to get the same improving life expectancies as the US without octupling health care spending.

The Netherlands increased their health budget by a lot around 2000, sparking a bunch of studies on whether that increased life expectancy or not. There's a good meta-analysis here, which lists six studies trying to calculate how much of the change in life expectancy was due to the large increases in health spending during this period. There's a broad range of estimates: 0.3%, 1.8%, 8.0%, 17.2%, 22.1%, 27.5% (I'm taking their numbers for men; the numbers for women are pretty similar). They also mention two studies that they did not officially include; one finding 0% effect and one finding 50% effect (I'm not sure why these studies weren't included). They add:

> In none of these studies is the issue of reverse causality addressed; sometimes it is not even mentioned. This implies that the effect of health care spending on mortality may be overestimated.

They say:

> Based on our review of empirical studies, we conclude that it is likely that increased health care spending has contributed to the recent increase in life expectancy in the Netherlands. Applying the estimates form published studies to the observed increase in health care spending in the Netherlands between 2000 and 2010 [of 40%] would imply that 0.3% to almost 50% of the increase in life expectancy may have been caused by increasing health care spending. An important reason for the wide range in such estimates is that they all include methodological problems highlighted in this paper. However, this wide range inicates that the counterfactual study by Meerding et al, which argued that 50% of the increase in life expectancy in the Netherlands since the

> 1950s can be attributed to medical care, can probably be interpreted as an upper bound.

It’s going to be completely irresponsible to try to apply this to the increase in health spending in the US over the past 50 years, since this is probably different at every margin and the US is not the Netherlands and the 1950s are not the 2010s. But if we irresponsibly take their median estimate and apply it to the current question, we get that increasing health spending in the US has been worth about one extra year of life expectancy.

This study attempts to directly estimate a %GDP health spending to life expectancy conversion, and says that an increase of 1% GDP corresponds to an increase of 0.05 years life expectancy. That would suggest a slightly different number of 0.65 years life expectancy gained by healthcare spending since 1960)

If these numbers seem absurdly low, remember all of those controlled experiments where giving people insurance doesn’t seem to make them much healthier in any meaningful way.

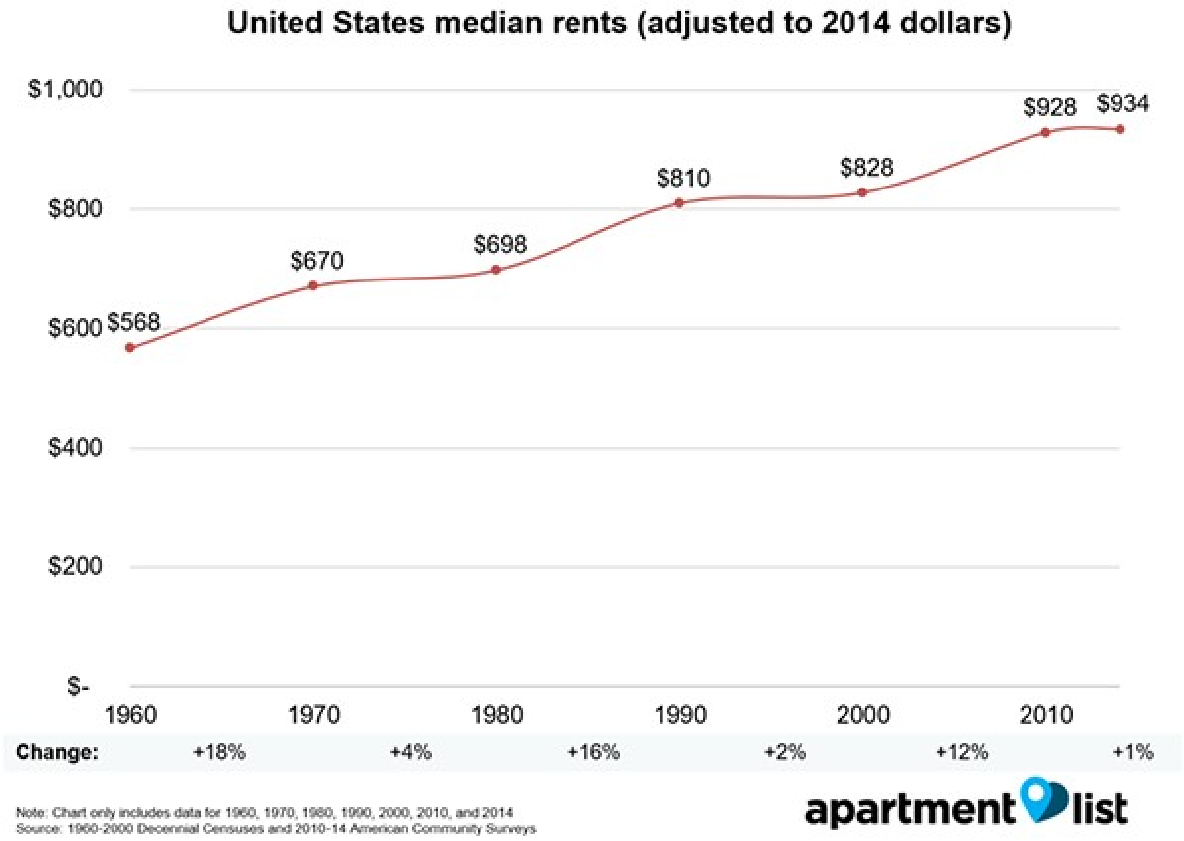
Or instead of slogging through the statistics, we can just ask the same question as before. Do you think the average poor or middle-class person would rather:

a) Get modern health care
b) Get the same amount of health care as their parents’ generation, but with modern technology like ACE inhibitors, and also earn $8000 extra a year

Fourth, we se similar effects in infrastructure. The first New York City subway opened around 1900. Various sources list lengths from 10 to 20 miles and costs from $30 million to $60 million dollars – I think my sources are capturing it at different stages of construction with different numbers of extensions. In any case, it suggests costs of between $1.5 million to $6 million dollars/mile = $1-4 million per kilometer. That looks like it's about the inflation-adjusted equivalent of $100 million/kilometer today, though I'm very uncertain about that estimate. In contrast, Vox notes that a new New York subway line being opened this year costs about $2.2 *billion* per kilometer, suggesting a cost increase of twenty times – although I'm very uncertain about this estimate.

Things become clearer when you compare them country-by-country. The same Vox article notes that Paris, Berlin, and Copenhagen subways cost about $250 million per kilometer, almost 90% less. Yet even those European subways are overpriced compared to Korea, where a kilometer of subway in Seoul costs $40 million/km (another Korean subway project cost $80 million/km). This is a difference of 50x between Seoul and New York for apparently comparable services. It suggests that the 1900s New York estimate above may have been roughly accurate if their efficiency was roughly in line with that of modern Europe and Korea.

Fifth, housing (source:


United States median rents (adjusted to 2014 dollars)

Note: Chart only includes data for 1960, 1970, 1980, 1990, 2000, 2010, and 2014
Source: 1960-2000 Decennial Censuses and 2010-14 American Community Surveys

Most of the important commentary on this graph has already been said, but I would add that optimistic takes like this one by the American Enterprise Institute are missing some of the dynamic. Yes, homes are bigger than they used to be, but part of that is zoning laws which make it easier to get big houses than small houses. There are a lot of people who would prefer to have a smaller house but don't. When I first moved to Michigan, I lived alone in a three bedroom house because there were no good one-bedroom houses available near my workplace and all of the apartments were loud and crime-y.

Or, once again, just ask yourself: do you think most poor and middle class people would rather:

1. Rent a modern house/apartment
2. Rent the sort of house/apartment their parents had, for

half the cost

II.

So, to summarize: in the past fifty years, education costs have doubled, college costs have dectupled, health insurance costs have dectupled, subway costs have at least dectupled, and housing costs have increased by about fifty percent. US health care costs about four times as much as equivalent health care in other First World countries; US subways cost about eight times as much as equivalent subways in other First World countries.

I worry that people don’t appreciate how weird this is. I didn’t appreciate it for a long time. I guess I just figured that Grandpa used to talk about how back in his day movie tickets only cost a nickel; that was just the way of the world. *But all of the numbers above are inflation-adjusted*. These things have dectupled in cost even *after* you adjust for movies costing a nickel in Grandpa’s day. They have really, genuinely dectupled in cost, no economic trickery involved.

And this is especially strange because we expect that improving technology and globalization ought to cut costs. In 1983, the first mobile phone cost $4,000 – about $10,000 in today’s dollars. It was also a gigantic piece of crap. Today you can get a much better phone for $100. This is the right and proper way of the universe. It’s why we fund scientists, and pay businesspeople the big bucks.

But things like college and health care have *still* had their prices dectuple. Patients can now schedule their appointments online; doctors can send prescriptions through the fax, pharmacies can keep track of medication histories on centralized computer systems that interface with the cloud, nurses get automatic reminders when they’re giving two drugs with a potential interaction, insurance companies accept payment through credit cards – and all of this costs ten times as much as it did in the days of punch cards and secretaries who did calculations by hand.

It’s actually even worse than this, because we take so many opportunities to save money that were unavailable in past generations. Underpaid foreign nurses immigrate to America and work for a song. Doctors’ notes are sent to India overnight where they’re transcribed by sweatshop-style labor for pennies an hour. Medical equipment gets manufactured in goodness-only-knows which obscure Third World country. And it *still* costs ten times as much as when this was all made in the USA – and that back when minimum wages were proportionally higher than today.

And it’s actually even worse than *this*. A lot of these services have decreased in quality, presumably as an attempt to cut costs even further. Doctors used to make house calls; even when I was young in the ’80s my father would still go to the houses of difficult patients who were too sick to come to his office. This study notes that for women who give birth in the hospital, “the standard length of stay was 8 to 14 days in the 1950s but declined to less than 2 days in the mid-1990s”. The doctors I talk to say this isn’t because modern women are healthier, it’s because they kick them out as soon as it’s safe to free up beds for the next person. Historic records of hospital care generally describe leisurely convalescence periods and making sure somebody felt absolutely well before letting them go; this seems bizarre to anyone who has participated in a modern hospital, where the mantra is to kick people out as soon as they’re “stable” ie not in acute crisis.

If we had to provide the same quality of service as we did in 1960, and without the gains from modern technology and globalization, who even *knows* how many times more health care would cost? Fifty times more? A hundred times more?

And the same is true for colleges and houses and subways and so on.

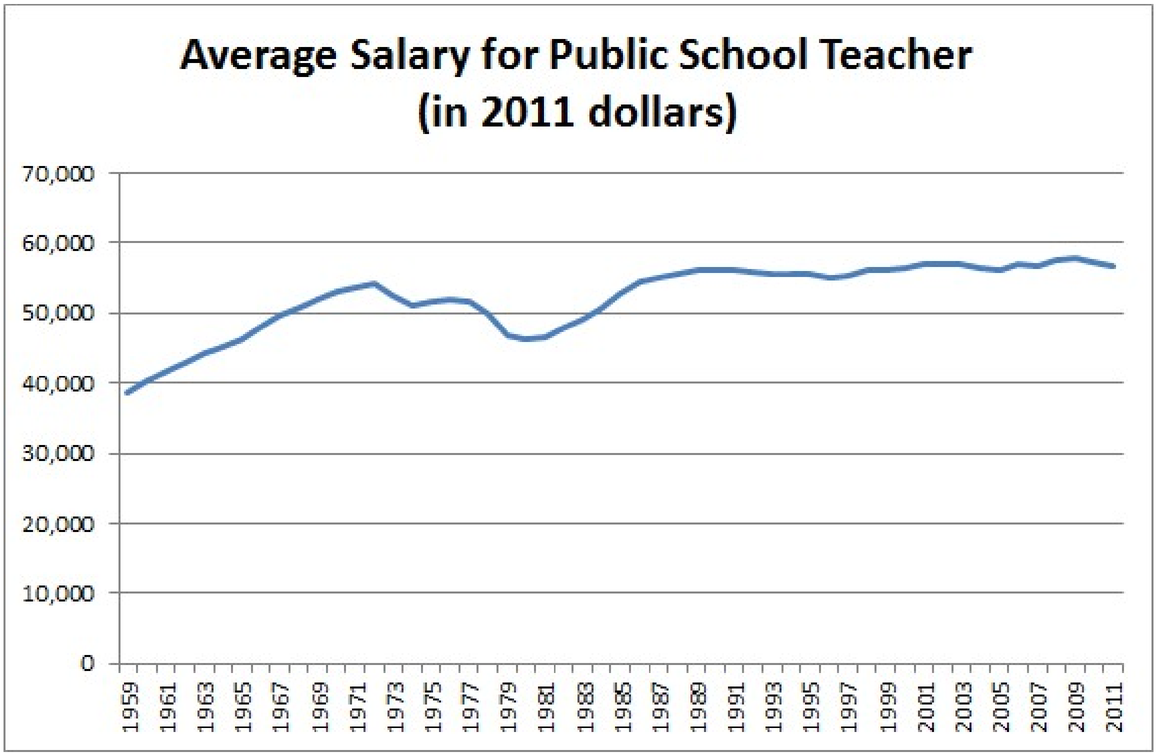
III.

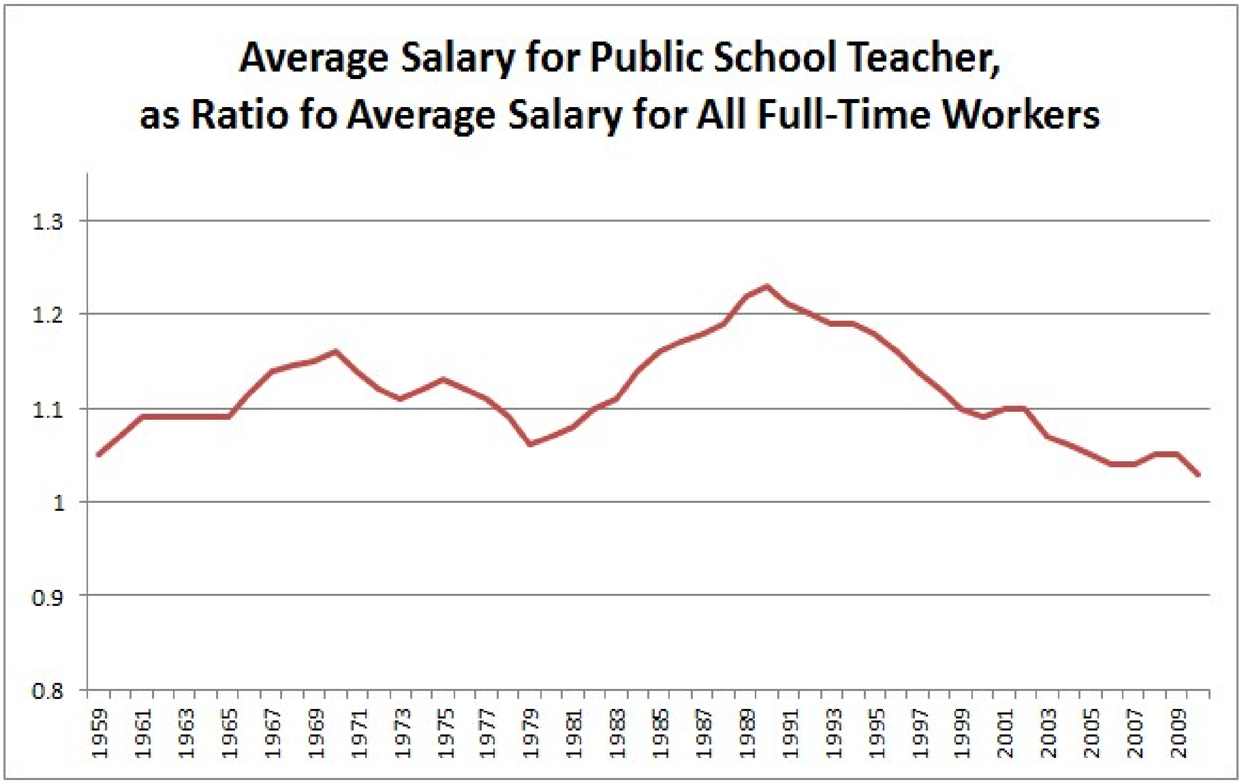
The existing literature on cost disease focuses on the Baumol effect. Suppose in some underdeveloped economy, people can choose either to work in a factory or join an orchestra, and the salaries of factory workers and orchestra musicians reflect relative supply and demand and profit in those industries. Then the economy undergoes a technological revolution, and factories can produce ten times as many goods. Some of the increased productivity trickles down to factory workers, and they earn more money. Would-be musicians leave the orchestras behind to go work in the higher-paying factories, and the orchestras have to raise their prices if they want to be assured enough musicians. So tech improvements in the factory sectory raise prices in the orchestra sector.

We could tell a story like this to explain rising costs in education, health care, etc. If technology increases productivity for skilled laborers in other industries, then less susceptible industries might end up footing the bill since they have to pay their workers more.

There’s only one problem: health care and education aren’t paying their workers more; in fact, quite the opposite.

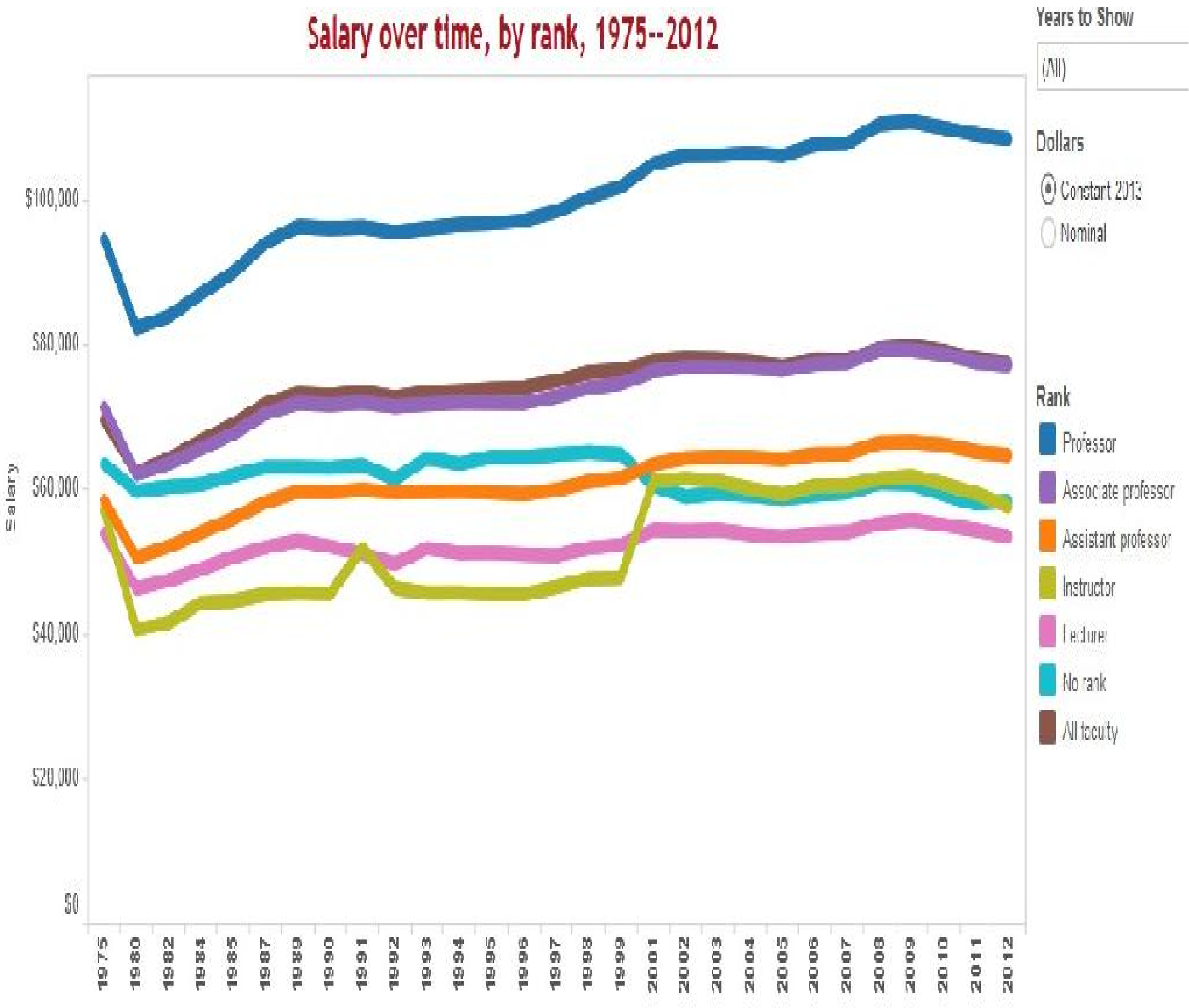
Here are teacher salaries over time (source):





Teacher salaries are relatively flat adjusting for inflation. But salaries for other jobs are increasing modestly relative to inflation. So teacher salaries relative to other occupations' salaries are actually declining.

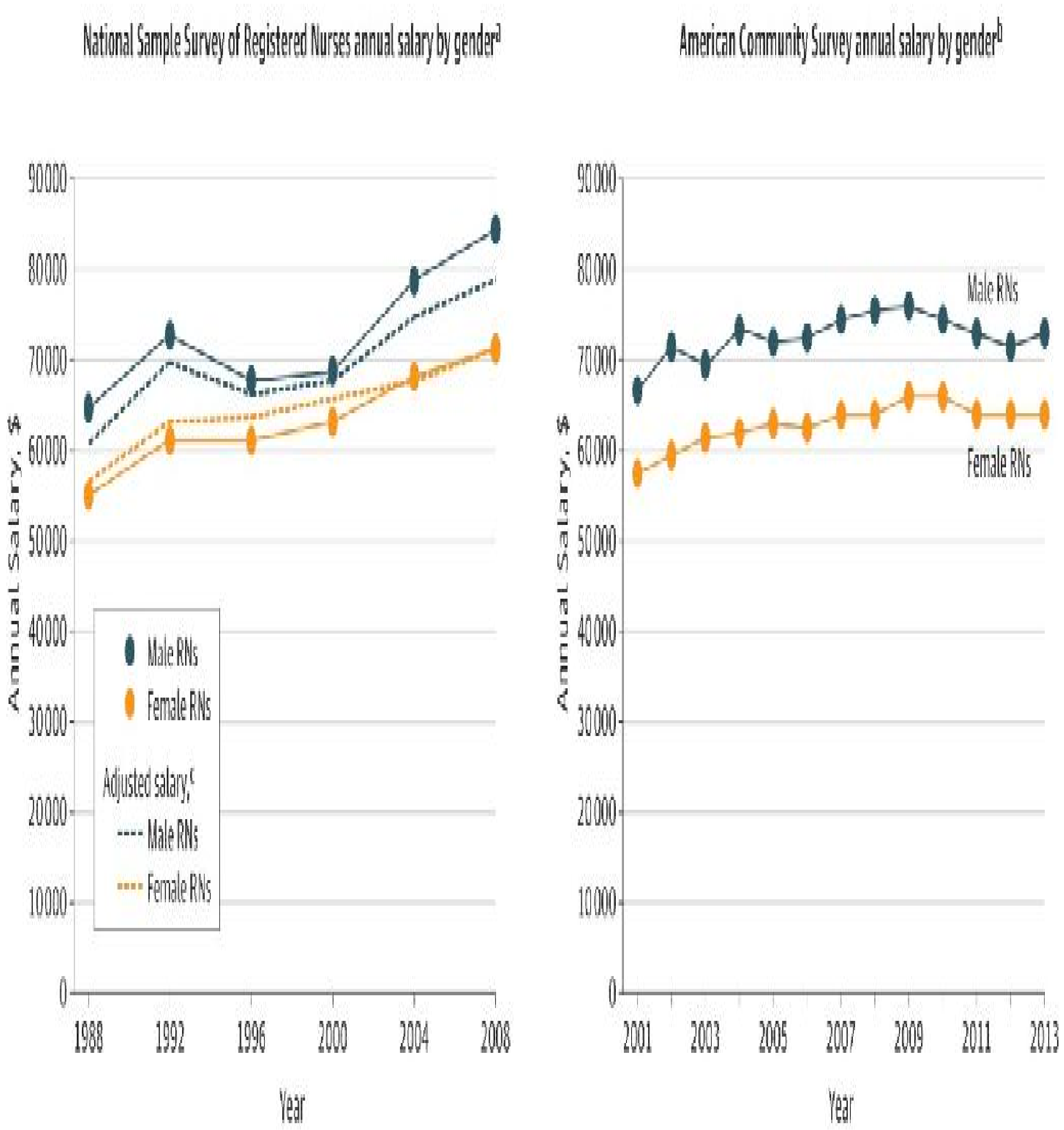
Here's a similar graph for professors (source):



Professor salaries are going up a little, but again, they're probably losing position relative to the average occupation. Also, note that although the average salary of each type of faculty is stable or increasing, the average salary of all faculty is going down. No mystery here – colleges are doing everything they can to switch from tenured professors to adjuncts, who complain of being overworked and abused while making about the same amount as a Starbucks barista.

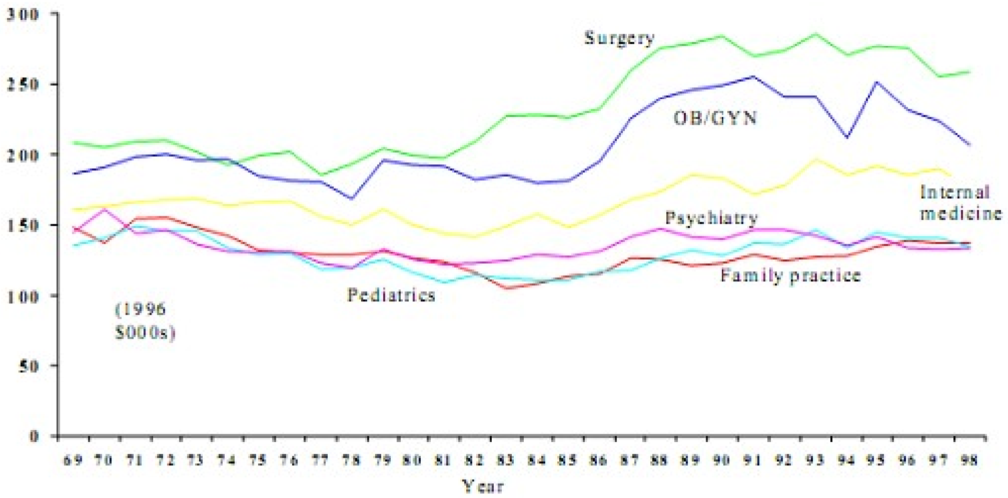
This seems to me a lot like the case of the hospitals cutting care for new mothers. The price of the service dectuples, yet at the same time the service has to sacrifice quality in order to control costs.

And speaking of hospitals, here’s the graph for nurses (source):



Female nurses’ salaries went from about $55,000 in 1988 to $63,000 in 2013. This is probably around the average wage increase during that time. Also, some of this reflects changes in education: in the 1980s only 40% of nurses had a degree; by 2010, about 80% did.

And for doctors (source)



Stable again! Except that a lot of doctors’ salaries now go to paying off their medical school debt, which has been ballooning like everything eles.

I don’t have a similar graph for subway workers, but come on. The overall pictures is that health care and education costs have managed to increase by ten times without a single cent of the gains going to teachers, doctors, or nurses. Indeed these professions seem to have lost ground salary-wise relative to others.

I also want to add some anecdote to these hard facts. My father is a doctor and my mother is a teacher, so I got to hear a lot about how these professions have changed over the past generation. It seems at least a little like the adjunct story, although without the clearly defined “professor vs.

adjunct" dichotomy that makes it so easy to talk about. Doctors are really, really, *really* unhappy. When I went to medical school, some of my professors would tell me outright that they couldn't believe anyone would still go into medicine with all of the new stresses and demands placed on doctors. This doesn't seem to be limited to one medical school. *Wall Street Journal*: Why Doctors Are Sick Of Their Profession – "American physicians are increasingly unhappy with their once-vaunted profession, and that malaise is bad for their patients". *The Daily Beast*: How Being A Doctor Became The Most Miserable Profession – "Being a doctor has become a miserable and humiliating undertaking. Indeed, many doctors feel that America has declared war on physicians". *Forbes*: Why Are Doctors So Unhappy? – "Doctors have become like everyone else: insecure, discontent and scared about the future." *Vox*: Only Six Percent Of Doctors Are Happy With Their Jobs. *Al Jazeera America*: Here's Why Nine Out Of Ten Doctors Wouldn't Recommend Medicine As A Profession. Read these articles and they all say the same thing that all the doctors I know say – medicine used to be a well-respected, enjoyable profession where you could give patients good care and feel self-actualized. Now it kind of sucks.

Meanwhile, I also see articles like this piece from NPR saying teachers are experiencing historic stress levels and up to 50% say their job "isn't worth it". Teacher job satisfaction is at historic lows. And the veteran teachers I know say the same thing as the veteran doctors I know – their jobs used to be enjoyable and make them feel like they were making a difference; now they feel overworked, unappreciated, and trapped in mountains of paperwork.

It might make sense for these fields to become more expensive if their employees' salaries were increasing. And it might make sense for salaries to stay the same if employees instead benefitted from lower workloads and better working conditions. But neither of these are happening.

IV.

So what's going on? Why *are* costs increasing so dramatically? Some possible answers:

First, can we dismiss all of this as an illusion? Maybe adjusting for inflation is harder than I think. Inflation is an average, so some things have to have higher-than-average inflation; maybe it's education, health care, etc. Or maybe my sources have the wrong statistics.

But I don't think this is true. The last time I talked about this problem, someone mentioned they're running a private school which does just as well as public schools but costs only $3000/student/year, a fourth of the usual rate. Marginal Revolution notes that India has a private health system that delivers the same quality of care as its public system for a quarter of the cost. Whenever the same drug is provided by the official US health system and some kind of grey market supplement sort of thing, the grey market supplement costs between a fifth and a tenth as much; for example, Google's first hit for Deplin®, official prescription L-methylfolate, costs $175 for a month's supply; unregulated L-methylfolate supplement delivers the same dose for about $30. And this isn't even mentioning things like the $1 bag of saline that costs $700 at hospitals. Since it seems like it's not too hard to do things for a fraction of what we currently do things for, probably we should be less reluctant to believe that the cost of everything is really inflated.

Second, might markets just not work? I know this is kind of an extreme question to ask in a post on economics, but maybe nobody knows what they're doing in a lot of these fields and people can just increase costs and not suffer any decreased demand because of it. Suppose that people proved beyond a shadow of a doubt that Khan Academy could teach you just as much as a normal college education, but for free. People would still ask questions like – will employers accept my Khan Academy degree? Will it look good on a resume? Will people make fun of me for it? The same is true of community colleges, second-tier colleges, for-profit colleges, et cetera. I got offered a free scholarship to a mediocre state college, and I turned it down on the grounds that I knew nothing about anything and maybe years from now I would be locked out of some sort of Exciting Opportunity because my college wasn't prestigious enough. Assuming everyone thinks like this, can colleges just charge whatever they want?

Likewise, my workplace offered me three different health insurance plans, and I chose the middle-expensiveness one, on the grounds that I had no idea how health insurance worked but maybe if I bought the cheap one I'd get sick and regret my choice, and maybe if I bought the expensive one I *wouldn't* be sick and regret my choice. I am a doctor, my employer is a hospital, and the health insurance was for treatment in my own health system. The moral of the story is that I am an idiot. The second moral of the story is that people probably are not super-informed health care consumers.

This can't be pure price-gouging, since corporate profits haven't increased nearly enough to be where all the money is going. But a while ago a commenter linked me to the Delta Cost Project, which scrutinizes the exact causes of increasing college tuition. Some of it is the administrative bloat that you would expect. But a lot of it is fun "student life" types of activities like clubs, festivals, and paying Milo Yiannopoulos to speak and then cleaning up after the ensuing riots. These sorts of things improve the student experience, but I'm not sure that the average student would rather go to an expensive college with clubs/festivals/Milo than a cheap college without them. More important, it doesn't really seem like the average student is offered this choice.

This kind of suggests a picture where colleges expect people will pay whatever price they set, so they set a very high price and then use the money for cool things and increasing their own prestige. Or maybe clubs/festivals/Milo become such a signal of prestige that students avoid colleges that don't comply since they worry their degrees won't be respected? Some people have pointed out that hospitals have switched from many-people-all-in-a-big-ward to private rooms. Once again, nobody seems to have been offered the choice between expensive hospitals with private rooms versus cheap hospitals with roommates. It's almost as if industries have their own reasons for switching to more-bells-and-whistles services that people don't necessarily want, and consumers just go along with it because for some reason they're not exercising choice the same as they would in other markets.

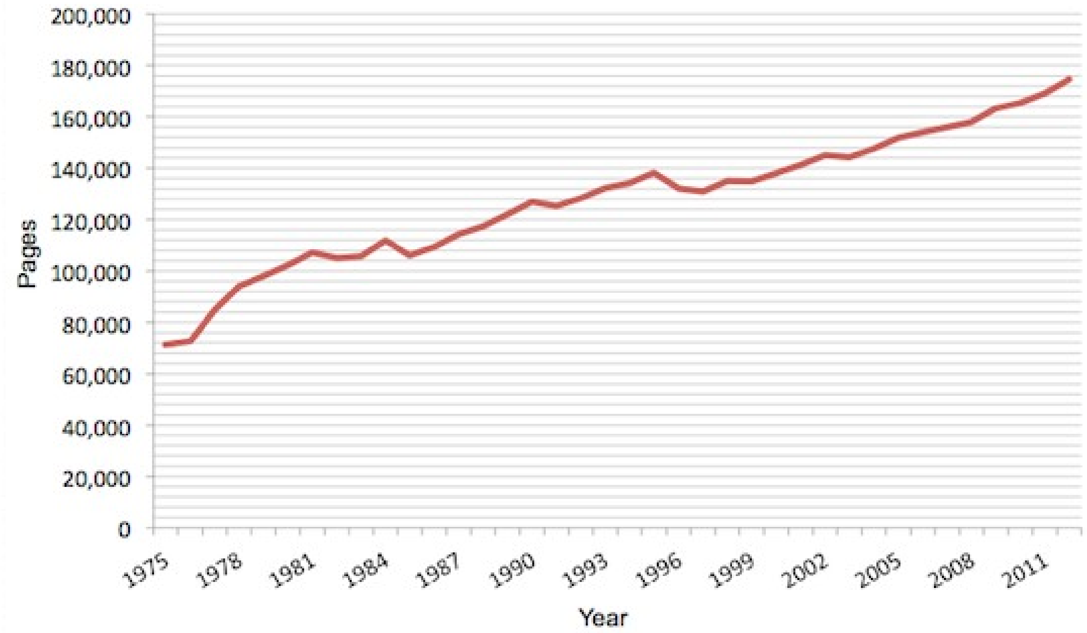
(this article on the Oklahoma City Surgery Center might be about a partial corrective for this kind of thing)

Third, can we attribute this to the inefficiency of government relative to private industry? I don't think so. The government handles most primary education and subways, and has its hand in health care. But we know that for-profit hospitals aren't much cheaper than government hospitals, and that private schools usually aren't much cheaper (and are sometimes more expensive) than government schools. And private colleges cost more than government-funded ones.

Fourth, can we attribute it to indirect government intervention through regulation, which public and private companies alike must deal with? This seems to be at least part of the story in health care, given how much money you can save by grey-market practices that avoid the FDA. It’s harder to apply it to colleges, though some people have pointed out regulations like Title IX that affect the educational sector.

One factor that seems to speak out against this is that starting with Reagan in 1980, and picking up steam with Gingrich in 1994, we got an increasing presence of Republicans in government who declared war on overregulation – but the cost disease proceeded unabated. This is suspicious, but in fairness to the Republicans, they did sort of fail miserably at deregulating things. “The literal number of pages in the regulatory code” is kind of a blunt instrument, but it doesn’t exactly inspire confidence in the Republicans’ deregulation efforts:

Figure 1: Total Number of Pages in the Code of Federal Regulations 1975 -- 2011



Here’s a more interesting (and more fun) argument against regulations being to blame: what about pet health care? Veterinary care is much less regulated than human health care, yet its cost is rising as fast (or faster) than that of the human medical system (popular article, study). I'm not sure what to make of this.

Fifth, might the increased regulatory complexity happen not through literal regulations, but through fear of lawsuits? That is, might institutions add extra layers of administration and expense not because they're *forced* to, but because they fear being sued if they don't and then something goes wrong?

I see this all the time in medicine. A patient goes to the hospital with a heart attack. While he's recovering, he tells his doctor that he's really upset about all of this. Any normal person would say "You had a heart attack, of course you're upset, get over it." But if his doctor says this, and then a year later he commits suicide for some unrelated reason, his family can sue the doctor for "not picking up the warning signs" and win several million dollars. So now the doctor consults a psychiatrist, who does an hour-long evaluation, charges the insurance company $500, and determines using her immense clinical expertise that the patient is upset because he just had a heart attack.

Those outside the field have *no idea* how much of medicine is built on this principle. People often say that the importance of lawsuits to medical cost increases is overrated because malpractice insurance doesn't cost that much, but the situation above would never look lawsuit-related; the whole thing only works because everyone involved documents it as well-justified psychiatric consult to investigate depression. Apparently some studies suggest this isn't happening, but all they do is survey doctors, and with all due respect all the doctors *I* know say the opposite.

This has nothing to do with government regulations (except insofar as these make lawsuits easier or harder), but it sure can drive cost increases, and it might apply to fields outside medicine as well.

Sixth, might we have changed our level of risk tolerance? That is, might increased caution be due not purely to lawsuitphobia, but to really caring more about whether or not people are protected? I read stuff every so often about how playgrounds are becoming obsolete because nobody wants to let kids run around unsupervised on something with sharp edges. Suppose that one in 10,000 kids get a horrible playground-related injury. Is it worth making playgrounds cost twice as much and be half as fun in order to decrease that number to one in 100,000? This isn't a rhetorical question; I think different people can have legitimately different opinions here (though there are probably some utilitarian things we can do to improve them).

To bring back the lawsuit point, some of this probably relates to a difference between personal versus institutional risk tolerance. Every so often, an elderly person getting up to walk to the bathroom will fall and break their hip. This is a fact of life, and elderly people deal with it every day. Most elderly people I know don't spend thousands of dollars fall-proofing the route from their bed to their bathroom, or hiring people to watch them at every moment to make sure they don't fall, or buy a bedside commode to make bathroom-related falls impossible. This suggests a revealed preference that elderly people are willing to tolerate a certain fall probability in order to save money and convenience. Hospitals, which face huge lawsuits if any elderly person falls on the premises, are *not* willing to tolerate that probability. They put rails on elderly people's beds, place alarms on them that will go off if the elderly person tries to leave the bed without permission, and hire patient care assistants who among other things go around carefully holding elderly people upright as they walk to the bathroom (I assume this job will soon require at least a master’s degree). As more things become institutionalized and the level of acceptable institutional risk tolerance becomes lower, this could shift the cost-risk tradeoff even if there isn’t a population-level trend towards more risk-aversion.

Seventh, might things cost more for the people who pay because so many people don’t pay? This is somewhat true of colleges, where an increasing number of people are getting in on scholarships funded by the tuition of non-scholarship students. I haven’t been able to find great statistics on this, but one argument against: couldn’t a college just not fund scholarships, and offer much lower prices to its paying students? I get that scholarships are good and altruistic, but it would be surprising if every single college thought of its role as an altruistic institution, and cared about it more than they cared about providing the same service at a better price. I guess this is related to my confusion about why more people don’t open up colleges. Maybe this is the “smart people are rightly too scared and confused to go to for-profit colleges, and there’s not enough ability to discriminate between the good and the bad ones to make it worthwhile to found a good one” thing again.

This also applies in health care. Our hospital (and every other hospital in the country) has some “frequent flier” patients who overdose on meth at least once a week. They comes in, get treated for their meth overdose (we can’t legally turn away emergency cases), get advised to get help for their meth addiction (without the slightest expectation that they will take our advice) and then get discharged. Most of them are poor and have no insurance, but each admission costs a couple of thousand dollars. The cost gets paid by a combination of taxpayers and other hospital patients with good insurance who get big markups on their own bills.

Eighth, might total compensation be increasing even though wages aren't? There definitely seems to be a pensions crisis, especially in a lot of government work, and it's possible that some of this is going to pay the pensions of teachers, etc. My understanding is that in general pensions aren't really increasing much faster than wages, but this might not be true in those specific industries. Also, this might pass the buck to the question of why we need to spend more on pensions now than in the past. I don't think increasing life expectancy explains all of this, but I might be wrong.

IV.

I mentioned politics briefly above, but they probably deserve more space here. Libertarian-minded people keep talking about how there's too much red tape and the economy is being throttled. And less libertarian-minded people keep interpreting it as not caring about the poor, or not understanding that government has an important role in a civilized society, or as a "dog whistle" for racism, or whatever. I don't know why more people don't just come out and say "LOOK, REALLY OUR MAIN PROBLEM IS THAT ALL THE MOST IMPORTANT THINGS COST TEN TIMES AS MUCH AS THEY USED TO FOR NO REASON, PLUS THEY SEEM TO BE GOING DOWN IN QUALITY, AND NOBODY KNOWS WHY, AND WE'RE MOSTLY JUST DESPERATELY FLAILING AROUND LOOKING FOR SOLUTIONS HERE." State that clearly, and a lot of political debates take on a different light.

For example: some people promote free universal college education, remembering a time when it was easy for middle class people to afford college if they wanted it. Other people oppose the policy, remembering a time when people didn't depend on government handouts. Both are true! My uncle paid for his tuition at a really good college just by working a pretty easy summer job – not so hard when college cost a tenth of what it did now. The modern conflict between opponents and proponents of free college education is over how to distribute our losses. In the old days, we could combine low taxes with widely available education. Now we can't, and we have to argue about which value to sacrifice.

Or: some people get upset about teachers' unions, saying they must be sucking the "dynamism" out of education because of increasing costs. Others people fiercely defend them, saying teachers are underpaid and overworked. Once again, in the context of cost disease, both are obviously true. The taxpayers are just trying to protect their right to get education as cheaply as they used to. The teachers are trying to protect their right to make as much money as they used to. The conflict between the taxpayers and the teachers' unions is about how to distribute losses; *somebody* is going to have to be worse off than they were a generation ago, so who should it be?

And the same is true to greater or lesser degrees in the various debates over health care, public housing, et cetera.

Imagine if tomorrow, the price of water dectupled. Suddenly people have to choose between drinking and washing dishes. Activists argue that taking a shower is a basic human right, and grumpy talk show hosts point out that in *their* day, parents taught their children not to waste water. A coalition promotes laws ensuring government-subsidized free water for poor families; a Fox News investigative report shows that some people receiving water on the government dime are taking long luxurious showers. Everyone gets really angry and there's lots of talk about basic compassion and personal responsibility and whatever but all of this is secondary to *why does water costs ten times what it used to?*

I think this is the basic intuition behind so many people, even those who genuinely want to help the poor, are afraid of "tax and spend” policies. In the context of cost disease, these look like industries constantly doubling, tripling, or dectupling their price, and the government saying “Okay, fine,” and increasing taxes however much it costs to pay for whatever they’re demanding now.

If we give everyone free college education, that solves a big social problem. It also locks in a price which is ten times too high for no reason. This isn’t fair to the government, which has to pay ten times more than it should. It’s not fair to the poor people, who have to face the stigma of accepting handouts for something they could easily have afforded themselves if it was at its proper price. And it’s not fair to future generations if colleges take this opportunity to increase the cost by *twenty* times, and then our children have to subsidize *that*.

I’m not sure how many people currently opposed to paying for free health care, or free college, or whatever, would be happy to pay for health care that cost less, that was less wasteful and more efficient, and whose price we expected to go down rather than up with every passing year. I expect it would be a lot.

And if it isn’t, who cares? The people who want to help the poor have enough political capital to spend eg $500 billion on Medicaid; if that were to go ten times further, then everyone could get the health care they need without any more political action needed. If some government program found a way to give poor people good health insurance for a few hundred dollars a year, college tuition for about a thousand, and housing for only two-thirds what it costs now, that would be the greatest anti-poverty advance in history. That program is called “having things be as efficient as they were a few decades ago”.

V.

In 1930, economist John Maynard Keynes predicted that his grandchildrens' generation would have a 15 hour work week. At the time, it made sense. GDP was rising so quickly that anyone who could draw a line on a graph could tell that our generation would be four or five times richer than his. And the average middle-class person in his generation felt like they were doing pretty well and had most of what they needed. Why *wouldn't* they decide to take some time off and settle for a lifestyle merely twice as luxurious as Keynes' own?

Keynes was sort of right. GDP per capita *is* 4-5x greater today than in his time. Yet we still work forty hour weeks, and some large-but-inconsistently-reported percent of Americans (76? 55? 47?) still live paycheck to paycheck.

And yes, part of this is because inequality is increasing and most of the gains are going to the rich. But this alone wouldn't be a disaster; we'd get to Keynes' utopia a little slower than we might otherwise, but eventually we'd get there. Most gains going to the rich means at least some gains are going to the poor. And at least there's a lot of mainstream awareness of the problem.

I'm more worried about the part where the cost of basic human needs goes up faster than wages do. Even if you're making twice as much money, if your health care and education and so on cost ten times as much, you're going to start falling behind. Right now the standard of living isn't just stagnant, it's at risk of declining, and a lot of that is student loans and health insurance costs and so on.

What's happening? I don't know and I find it really scary.

Highlights From The Comments On Cost Disease

I got many good responses to my Considerations On Cost Disease post, both in the comments and elsewhere. A lot of people thought the explanation was obvious; unfortunately, they all disagreed on what the obvious explanation was. Below are some of the responses I found most interesting.

John Cochrane:

> So, what is really happening? I think Scott nearly gets there. Things cost 10 times as much, 10 times more than they used to and 10 times more than in other countries. It's not going to wages. It's not going to profits. So where is it going?
>
> The unavoidable answer: The number of people it takes to produce these goods is skyrocketing. Labor productivity — number of people per quality adjusted output — declined by a factor of 10 in these areas. It pretty much has to be that: if the money is not going to profits, to to each employee, it must be going to the number of employees.
>
> How can that happen? Our machines are better than ever, as Scott points out. Well, we (and especially we economists) pay too much attention to snazzy gadgets. Productivity depends on organizations not just on gadgets. Southwest figured out how to turn an airplane around in 20 minutes, and it still takes United an hour.
>
> Contrariwise, I think we know where the extra people are. The ratio of teachers to students hasn't gone down a lot — but the ratio of administrators to students has shot up. Most large public school systems spend more than half their budget on administrators. Similarly, class sizes at most colleges and universities haven't changed that much — but administrative staff have exploded. There are 2.5 people handling insurance claims for every doctor. Construction

sites have always had a lot of people standing around for every one actually working the machine. But now for every person operating the machine there is an army of planners, regulators, lawyers, administrative staff, consultants and so on. (I welcome pointers to good graphs and numbers on this sort of thing.)

So, my bottom line: administrative bloat.

Well, how does bloat come about? Regulations and law are, as Scott mentions, part of the problem. These are all areas either run by the government or with large government involvement. But the real key is, I think lack of competition. These are above all areas with not much competition. In turn, however, they are not by a long shot “natural monopolies” or failure of some free market. The main effect of our regulatory and legal system is not so much to directly raise costs, as it is to lessen competition (that is often its purpose). The lack of competition leads to the cost disease.

Though textbooks teach that monopoly leads to profits, it doesn’t “The best of all monopoly profits is a quiet life” said Hicks. Everywhere we see businesses protected from competition, especially highly regulated businesses, we see the cost disease spreading. And it spreads largely by forcing companies to hire loads of useless people.

Yes, technical regress can happen. Productivity depends as much on the functioning of large organizations, and the overall legal and regulatory system in which they operate, as it does on gadgets. We can indeed “forget” how those work. Like our ancestors peer at the buildings, aqueducts, dams, roads, and bridges put up by our ancestors, whether Roman or American, and wonder just how they did it.

David Manheim:

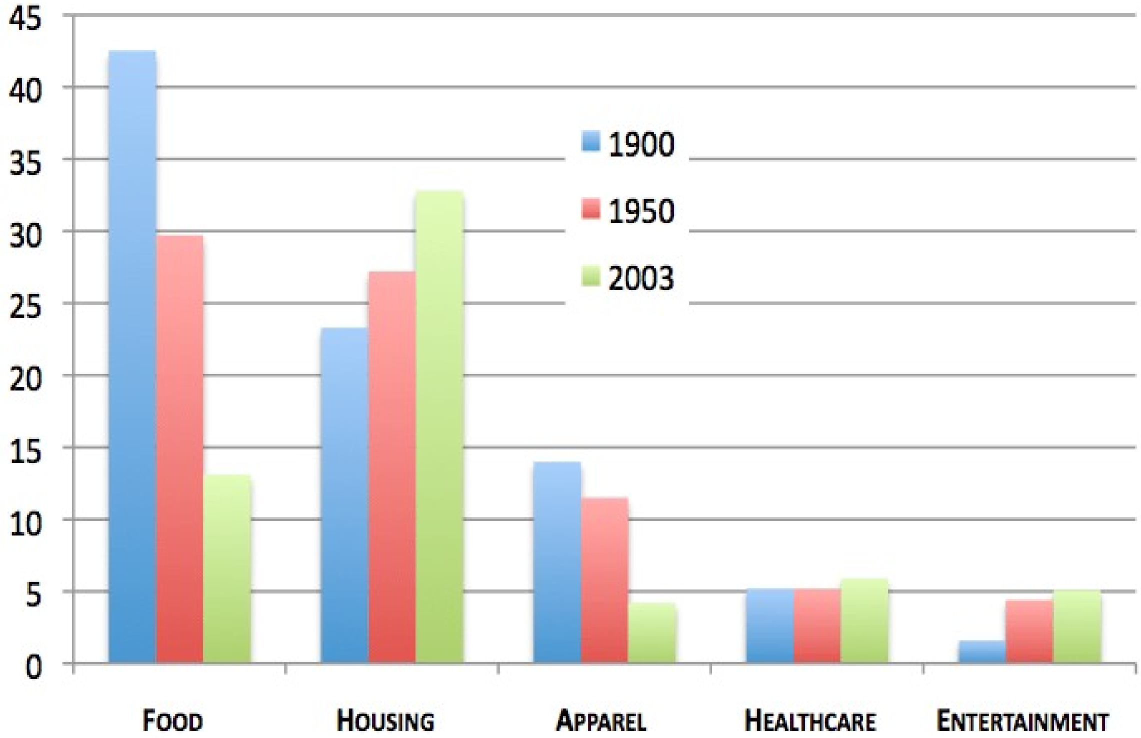
I think there is another dynamic that’s being ignored—and I would be surprised if an economist ignored it, but I’ll blame

Scott's eclectic ad-hoc education for why he doesn't discuss the elephant in the room—Superior goods.

For those who don't remember their Economics classes, imagine a guy who makes $40,000/year and eats chicken for dinner 3 nights a week. He gets a huge 50% raise, to $60,000/year, and suddenly has extra money to spend—his disposable income probably tripled or quadrupled. Before the hedonic treadmill kicks in, and he decides to waste all the money on higher rent and nicer cars, he changes his diet. But he won't start eating chicken 10 times a week—he'll start eating steak. When people get more money, they replace cheap "inferior" goods with expensive "superior" goods. And steak is a superior good.

But how many times a week will people eat steak? Two? Five? Americans as a whole got really rich in the 1940s and 1950s, and needed someplace to start spending their newfound wealth. What do people spend extra money on? Entertainment is now pretty cheap, and there are only so many nights a week you see a movie, and only so many $20/month MMORPGs you're going to pay for. You aren't going to pay 5 times as much for a slightly better video game or movie—and although you might pay double for 3D-Imax, there's not much room for growth in that 5%.

The Atlantic had a piece on this several years ago, with the following chart:



Food, including rising steak consumption, decreased to a negligible part of people's budgets, as housing started rising.In this chart, the reason healthcare hasn't really shot up to the extent Scott discussed, as the article notes, is because most of the cost is via pre-tax employer spending. The other big change the article discusses is that after 1950 or so, everyone got cars, and commuted from their more expensive suburban houses—which is effectively an implicit increase in housing cost.

And at some point, bigger houses and nicer cars begin to saturate; a Tesla is nicer than my Hyundai, and I'd love one, but not enough to upgrade for 3x the cost. I know how much better a Tesla is—I've seen them.
Limitless Demand, Invisible Supply

There are only a few things that we have a limitless demand for, but very limited ability to judge the impact of our spending. What are they?

I think this is one big missing piece of the puzzle; in both healthcare and education, we want improvements, and they are worth a ton, but we can’t figure out how much the marginal spending improves things. So we pour money into these sectors.

Scott thinks this means that teachers’ and doctors’ wages should rise, but they don’t. I think it’s obvious why; they supply isn’t very limited. And the marginal impact of two teachers versus one, or a team of doctors versus one, isn’t huge. (Class size matters, but we have tons of teachers—with no shortage in sight, there is no price pressure.)

What sucks up the increased money? Dollars, both public and private, chasing hard to find benefits.

I’d spend money to improve my health, both mental and physical, but how? Extra medical diagnostics to catch problems, pricier but marginally more effective drugs, chiropractors, probably useless supplements—all are exploding in popularity. How much do they improve health? I don’t really know—not much, but I’d probably try something if it might be useful.

I’m spending a ton of money on preschool for my kids. Why? Because it helps, according to the studies. How much better is the $15,000/year daycare versus the $8,000 a year program a friend of mine runs in her house? Unclear, but I’m certainly not the only one spending big bucks. Why spend less, if education is the most superior good around?

How much better is Harvard than a subsidized in-state school, or four years of that school versus 2 years of cheap community college before transferring in? The studies seem to suggest that most of the benefit is really because the kids who get into the better schools. And Scott knows that this is happening.

> We pour money into schools and medicine in order to improve things, but where does the money go? Into efforts to improve things, of course. But I've argued at length before that bureaucracy is bad at incentivizing things, especially when goals are unclear. So the money goes to sinkholes like more bureaucrats and clever manipulation of the metrics that are used to allocate the money.
>
> As long as we're incentivized to improve things that we're unsure how to improve, the incentives to pour money into them unwisely will continue, and costs will rise. That's not the entire answer, but it's a central dynamic that leads to many of the things Scott is talking about—so hopefully that reduces Scott's fears a bit.

A reader who wishes to remain anonymous emails me, saying:

> In the business I know – hedge funds – I am aware of tiny operators running perfectly functional one-person shops on a shoestring, who take advantage of workarounds for legal and regulatory costs (like http://www.riainabox.com/. Then there are folks like me who are trying to "be legit" and hope to attract the big money from pensions and big banks. Those folks' decisions are all made across major principal/agent divides where agents are incentivized not to take risks. So, they force hedge funds into an arms race of insanely paranoid "best practices" to compete for their money. So… my set up costs (which so far seem to have been too little rather than too much) were more than 10x what they could have been.
>
> I guess this supports the "institutional risk tolerance" angle. There must be similar massive unseen frictions probably in many industries that go into "checking boxes".
>
> Relatedly, a pet theory of mine is that "organizational complexity" imposes enormous and not fully appreciated costs, which probably grow quadratically with organization size. I'd predict, without Googling, that the the US military,

> just as a function of being so large, has >75% of its personal doing effectively administrative/logistical things, and that you could probably find funny examples of organizational-overhead-proliferation like an HR department so big it needed its own (meta-)HR department.

Noah Smith:

> That could be one force behind rising costs; it definitely seems important for K-12 education. But it doesn't explain why the U.S. is so much worse than countries such as France, Germany or Japan. Those countries are about as productive as the U.S., so their cost disease should be comparable. Something else must be afoot.
>
> Another usual suspect is government intervention. The government subsidizes college through cheap loans, purchases infrastructure, restricts housing supply, and intervenes heavily in the health-care market. It's probably part of the problem in these areas, especially in urban housing markets.
>
> But again, government intervention struggles to explain the difference between the U.S. and other rich nations. In most countries, health care is mainly paid for by the government — many countries have nationalized the industry outright. Yet their health outcomes are broadly similar to those in the U.S., or even a little bit better. Other countries have strong unions and high land acquisition costs — often stronger and higher than the U.S. — but their infrastructure is much cheaper. And there is no law or regulation propping up high wealth-management fees or real-estate commissions. In general, lower-cost places like Japan and Europe have more regulation and more interventionism than the U.S.
>
> So if cost disease and government can at most be only part of the story, what's going on? One possibility Alexander raises is that "markets might just not work." In other words, there might be large market failures going on.

The health-care market naturally has a lot of adverse selection — people with poor health are more inclined to buy insurance. That means insurance companies, knowing its customers tend to be those with poorer health, charge higher prices. Also, hospitals could be local monopolies. And college education could be costly in part because of asymmetric information — if Americans tend to vary more than people in other countries with respect to work ethnic and natural ability, they might have to spend more on college to prove themselves. This is known as signaling.

When high costs are due to market failures, interventionist government can be the solution instead of the problem — provided the intervention is done right. So the more active governments of countries like Europe and Japan might be successfully holding down costs that would otherwise balloon to inefficient levels.

But there's one more possibility — one that gets taught in few economics classes. There is almost certainly some level of pure trickery in the economy — people paying more than they should, because they don't have the time or knowledge to look for better prices, or because they trust people they shouldn't trust.

This is the thesis of the book "Phishing for Phools," by Nobel-winning economists George Akerlof and Robert Shiller. The authors advance the disturbing thesis that sellers will continually look for ways to dupe customers into paying more than they should, and that these efforts will always be partially successful. In Akerlof and Shiller's reckoning, markets don't just sometimes fail — they are inherently subject to both deceit and mistakes.

That could explain a number of unsettling empirical results in the economics literature. For example, transparency reduces prices substantially in health-care equipment markets. More complex and opaque mortgage-backed securities failed at higher rates in the financial crisis. In these and other cases,

> buyers paid too much because they didn't know what they were buying. Whether that's due to trickery, or to the difficulty of gathering accurate information, it's not good — in an efficient economy, everyone will know what they're buying.
>
> So it's possible that many of those anomalously high U.S. costs are due to the natural informational problems of markets.

Megan McArdle:

> It's pretty easy to tell a libertarian story where markets work fine, but government intrusions into these markets have rendered them so unfree that they no longer function the way they're supposed to. And I think that is at least part of the story here. Yes, these things are often procured from private parties. But everywhere you look you see the government: blocking new entry (through accreditation standards, "certificate of need" laws, and zoning and building codes), while simultaneously subsidizing the purchases through artificially cheap loans and often, direct price subsidies. It would be sort of shocking if restricted supply combined with stimulated demand didn't produce rapidly rising prices. Meanwhile, in areas that the government largely leaves alone (such as Lasik), we pretty much see what you'd expect: falling prices and improving consumer service.
>
> But that's perhaps a little simplistic. Agriculture is also the focus of a great deal of government intervention, as are sundry things such as air travel, and we don't see the same phenomenon there. So we need to dig a little deeper and describe what's special about these three sectors (we'll leave public transportation out of it, because there, the answer is pretty much "union featherbedding combined with increasingly dysfunctional procurement and regulatory processes").

First, and most obviously, they involve vital purchases made on long time horizons, and with considerable uncertainty. Food is more vital than health care to our well-being, but its price and quality are really easy to assess: if you buy a piece of fruit, you know pretty quickly whether you liked it or not. This is a robust market, and it's going to take communist-level intervention to fundamentally mess it up so that food is both scarce and not very good.

Homes, schooling and health care, on the other hand, are more complicated products. You don't know when you buy them how much value they will be to you, and it is often difficult for a lay person to assess the quality of the product. You can read hospital rankings and pay a home inspector, but these things only go so far.

The fact that these are expensive purchases that can go terribly wrong creates a great deal of pressure for the government to intervene. As ours has, over and over, in all sorts of ways.

And at the risk of giving up a little bit of my libertarian cred, I'll say that government intervention in these markets did not have to be as expensive-making as it has turned out to be in America. Other countries have these sorts of problems too, but they're nowhere near as large as ours.

Part of that is just that we're richer than most of those other countries. We were going to spend the portion of our budgets no longer needed for food somewhere, and health care, education and housing are pretty good candidates. But that's only part of the story. A big part of the story is that America just isn't very good at regulation. When you talk to people who live elsewhere about what their government does, one thing that really strikes you about those conversations is how much more competent other rich industrial governments seem to be at regulating things and delivering services. Their bureaucracies are not perfect, but they are better than ours.

> That's not to say that America could have an awesome big government. Our regulatory state has been incompetent compared to others for decades, since long before the Reagan Revolution that Democrats like to blame. There are many, many factors in this, from our immigration history (vital to understanding how modern urban bureaucracies work in this country), to the fact that we have many competing centers of power instead of a single unified government providing over a single bureaucratic hierarchy. There is no way to fix this on a national level, and even at the level of local bureaucratic reform, it's darned near impossible.
>
> In other words, this is probably what we're stuck with. It may not be Baumol's cost disease — but it's potentially even more serious, and it's going to be a lingering condition.

Scott Sumner:

> I certainly don't claim to have all the answers, but I do feel that much of the problem reflects the fact that governments often cover the cost of services in those three areas. This leads producers to spend more than the socially optimal amount on these products. I'm going to provide some examples, but before doing so recall that economic theory predicts that costs in those areas should be wildly excessive. If the government paid 90% of the cost of any car you bought, and that didn't lead to lots more people buying Porsches and Ferraris, then we'd have a major puzzle on our hands.
>
> Scott mentions that private for-profit hospitals are also quite expensive. But even there, costs are largely paid for by the government. Close to half of all health care spending is directly paid for by the government (Medicare, Medicaid, Veterans, government employees, etc.) and a large share of the rest is indirectly paid for by taxpayers because health insurance is not just income tax free, but also payroll tax

free. I'd be stunned if health care spending had not soared in recent decades.

A sizable share of my health care spending has been unneeded, and I'm fairly healthy. I met one person in their 80s who had a normal cold and went to see the doctor. They said it was probably just a normal cold, but let's put you in the hospital overnight and do some tests, just in case. There was nothing wrong, and the bill the next day was something in the $5000 to $10,000 range, I forget the exact amount. This must happen all the time. No way would they have opted for those services if Medicare weren't picking up the tab.

Just to be clear, I don't think any monocausal explanation is enough. Governments also pay for health care in other countries, and the costs are far lower. It's likely the interaction of the US government picking up much of the tab, plus insurance regulations, plus American-style litigation, plus powerful provider lobbies that prevent European-style cost controls, etc., etc., lead to our unusually high cost structure. So don't take this as a screed against "socialized medicine." I'm making a narrower point, that a country where the government picks up most of the costs, and doesn't have effective regulations to hold down spending, is likely to end up with very expensive medicine.

To be fair, there is evidence from veterinary medicine that demand for pet care has also soared, and that suggests people are becoming more risk averse, even for their pets. But there is also evidence cutting the other way. Plastic surgery has not seen costs skyrocket. (Both are medical fields where people tend to pay out of pocket.)

I started working at Bentley in 1982, teaching 4 courses a semester. When I retired in 2015, I was making 7 times as much in nominal terms (nearly 3 times as much in real terms), and I was teaching 2 courses per semester. Thus I was being paid 14 times more per class (nearly 6 times as much in real terms). No wonder higher education costs have soared! (Even salaries for new hires have risen sharply in real terms.) Interestingly, the size of the student body at Bentley didn't change noticeably over that period (about 4000 undergrads.) But the physical size of the school rose dramatically, with many new buildings full of much fancier equipment. Right now they are building a new hockey arena. There are more non-teaching employees. You can debate whether living standards for Americans have risen over time, but there's no doubt that living standards for Americans age 18-22 have risen over time—by a lot.

As far as elementary school, my daughter had 2, 3, and once even 4 teachers in her classroom, with about 18 students. We had one teacher for 30 students when I was young. (I'm told classes are even bigger in Japan, and they don't have janitors in their schools. The students must mop the floors. I love Japan!)

There are also lots more rules and regulations. By the end of my career, I felt almost like I was spending as much time teaching 2 classes as I used to spend teaching 4. Many of these rules were well intentioned, but in the end I really don't think they led to students learning any more than back in 1982. I wonder if Dodd/Frank is now making small town banking a frustrating profession in the way that earlier regs made medicine and teaching increasing frustrating professions.

People say this is a disease of the service sector. But I don't see skyrocketing prices in restaurants, dry clearers, barbers and lots of other service industries where people pay out of pocket.

The same is true of construction. Scott estimates that NYC subways cost 20 times as much as in 1900, even adjusting for inflation. The real cost of other types of construction (such as new homes), has risen far less. Again, people pay for homes out of pocket, but government pays for subways.

Do I even need to mention the cost of weapons system like the F-35?

To summarize, the case of pet medicine shows that costs can rise rapidly even when people pay out of pocket. But the biggest and most important examples of cost inflation are in precisely those industries where government picks up a major part of the tab–health, education, and government procurement of complex products. And excessive cost inflation is exactly what economic theory predicts will happen when governments heavily subsidize an activity, without adequate cost regulations. Just as excessive risk taking is exactly what economic theory predicts will happen if government insures bank deposits, without adequate risk regulations. Let’s not be surprised if the things that happen, are exactly what the textbooks predict would happen. Even FDR predicted that deposit insurance would lead to reckless behavior by banks, and he (reluctantly) signed the bill into law.

Sohois:

I’ve seen some evidence that corporations can be equally vulnerable to cost disease as public institutions.

For example, since the 1980s CEO pay has quintupled despite the lack of any growth in profits or otherwise to justify this. Now this is probably going to result in far smaller effects on overall cost, but it still stands as a demonstration of how market failure can occur and result in large cost increases in these firms.

I would venture that many firms have seen huge increases in both revenues and costs so that when you adjust profit for inflation it hasn’t really changed at all, on average.

Andrew Swift

What you observe is fifty years of optimization of wealth extraction. Price outcomes depend on the contributions of hundreds of participants. Every participant optimizes his/her earnings, exerting a constant upward pressure on price. Participants become ever more expert at getting rich. Wealth-extraction schemes (scams) are refined and optimized (in all markets), and price increases are pushed downstream (in markets where buyers can’t push back). Radical price increases reflect markets where consumers have reduced ability to push back:

– complex markets (can’t understand)
– opaque markets (can’t see)
– entrenched/highly-regulated markets (can’t modify)
– necessary-to-keep-living markets (can’t avoid)
– limited-quantity markets (really want)
– intermediated markets where the end buyer doesn’t decide how things are purchased (don’t choose)

Some systems are resistant to contributors’ efforts to extract wealth and some systems are not. There’s an equilibrium between cost and readiness to pay. To reduce the costs in expensive domains, willingness to pay the high costs has to be reduced. As long as the buyer won’t or can’t say no, costs will increase through the entire production process. There won’t necessarily be one big obvious rip-off, but every participant will optimize the heck out of his contribution and the overall pressure will push costs up.

Could one provide a cheaper alternative in these domains? Sure for a little while, but if the bottom line is that people are willing to pay more for the service the prices will creep back up.

The only exception would be where the new, lower-priced, alternative sets a new standard and buyers refuse to continue paying the old prices. See https://stratechery.com/2016/dollar-shave-club-and-the-disruption-of-everything/ for a great article about this.

Habu71:

> My favorite example of ridiculous order-of-magnitude type cost increases is nuclear power plant construction costs. The plots from this paper illustrate it nicely. https://www.researchgate.net/figure/292964046_fig6_Overnight-Construction-Cost-and-Construction-Duration-of-US-Nuclear-Reactors-Color
>
> Except, in the case of power plant costs, the causes – at least, for the increasing US costs – are quite a bit more apparent. Pre TMI, US costs were in line with the rest of the world's cost. Post TMI, not so much. New regulatory burdens all by themselves increased the cost of new plants by a factor of ten. Now, this is of course not proof that any of the other problems that Scott mentioned are entirely – or even mostly – caused by increasing regulatory burdens. It does however, show that government institutions as awful as the NRC do exist, and that their effects can raise costs by the amounts seen health care, education, etc…"

John Schilling:

> [Fear of lawsuits] is well understood as the cause of the substantial rise in light airplane prices since 1970. A single-engine, four-seat Cessna 172 cost an inflation-adjusted $77,000 in 1970. A substantially identical airplane cost $163,000 in 1986. And went out of production the next year, because people weren't willing to pay that price. When congress passed laws relaxing the manufacturer's liability for older airplanes, Cessna was able to reinstate production in 1996 at an inflation-adjusted $190,000. Today, the price seems to literally be "if you have to ask, you can't afford it"; the manufacturer only advertises fleet sales, but I'd estimate about $400,000 (of which ~$100K is fancy electronics that didn't exist in 1986 and weren't standard in 1996).
>
> In this case it is particularly easy to pull out the lawsuit/liability effect because there aren't many cofounders.

> The 1986 Cessna is so little changed from the 1970 model that they sell at about the same price on the used market when controlled for condition and total flight time. And fear of lawsuits didn't manifest as safety enhancements of inscrutable cost and value, because light airplane crashes are almost always due to Stupid Pilot Tricks and almost everything that a manufacturer could do to mitigate that (e.g. tricycle landing gear) was standard in 1970. But the manufacturers still get sued, and have to pay millions, so there's nothing to be done but pay for liability insurance. And, second-order effect, cut production when your customers start balking at the increased prices, so you have to amortize the fixed costs of actually building airplanes over a smaller sales volume.
>
> So, a doubling in price over fifteen or so years, a quadrupling over fifty years in spite of Congress noticing the problem and trying to mitigate it, attributable to safety/liability concerns but not resulting in actual safety improvements. I have no trouble believing something similar is happening in other industries but is harder to discern because too many other things are happening at the same time.

Alex Zavoluk

> Wikipedia suggests that almost all of those other countries have litigation rules that make weak civil cases more costly, which seems like evidence in favor of the litigation hypothesis. It also means that there's a relatively straightforward solution.

Doug on Marginal Revolution, in response to a lot of people asking whether maybe we were just calculating the CPI wrong:

> That's a plausible hypothesis, but viewed through that frame of reference, median wages have also gone down tremendously. It's still the case that the median person spends at least six times more of their paycheck on healthcare. If healthcare is a closer metric to "true prices"

then manufactured good, then that means median wages have fallen by around 80%. It also means that overall GDP has crashed since 1970, since the price deflator now averages 6-7%, and nominal GDP has only averaged around 4%. It would mean that the economy has literally been in recession 90% of the past 40 years.

The only reason we're not all starving in the street is from the miraculous gains in manufacturing productivity and automation. But again, in this framework, that process is largely exogenous to the terrible macroeconomic situation. If those gains slow down even a little, and the macro trends continue, we're probably facing imminent economic collapse. So maybe this is a plausible hypothesis, but it certainly comes with a whole lot of extreme implications. I think medicine/education specific cost disease is a much more likely explanation.

bkearns123:

One commonality in the examples cited is disintermediation/subsidies. College is paid for by a third party, and financed by generous government loans. Generous in the sense that they are easy to get, not easy to get out of. Health care has massive tax subsidies, and for a good period of time felt "free" to employees. Public schooling is paid for indirectly.

Regarding the section on risk aversion, I happen to be in the playground business. The most common injury is broken bones from a fall. Consequently, our industry has ended up with poured in place surfacing, which costs 10x as much as mulch or pea gravel. It is wonderful stuff, but really increases the cost of the playground. Again, no one pays directly for their playground, and the paying party cannot risk not being in tune with the regulations.

Markets cannot function if the risk reward relationship is not direct.

fc123::

> In all of these problem sectors it seems the resources consumed in each industry has shifted to servicing and extending the definition of the marginal ‘customer’. This can explain I think some of the above
>
> E.g.. 40 years ago hospitals received 100 customers. Ranked, patients 1-20 died. And no one really tried to save them (some comfort but that was it). Today they are trying (are obliged) to try to save patients no 5-15 (the 85 year old with triple bypass, 20 week premie). The total no of staff needed for this task swamps increases in individual productivity. You just need more people, even if they each are more productive or trained than in the pasts. So salaries for each does not go up that much, there are just more of them, total cost go up, and outcomes over the patients treated are somewhat but not much better (some now make it but some fraction still die). Hence medical curve shows some improvement but not 1:1 with cost.
>
> In education, in the 1950-1970s we could afford to socially promote non-academically inclined students, not really expend effort on them as long as they kept quiet in class, then have them leave at age 16 to go work at Ford. Universities could count on getting the higher performing students. Today, we have to deliver much weaker students all the way to the end of high school, also force many into college. And ALL the extra resources go to get this new lower end close to what used to be the minimal university student performance. The top cohort gets little extra resources and has not really improved. Hence, the scores across the new ‘extended’ student population stays flat.
>
> I base this partly on what I have seen from my wife (engineering professor at top university), resources are heavily consumed by the lower performing students, top students have better opportunities than 20 years ago but in

general the resources are much less focused on them than on the marginal students.

So if you assume these industries for whatever reason shifted focus to servicing deeper into the tail of the population aptitude/effort over the years (I am not saying this is good/bad, was for social reasons, for humanistic reasons or making any comment), this would very much explain the overall cost rise, coupled with the lack of desired improvement in statistics measured across the population that now gets services as a whole.

In short, in the US we define policies that drive costs based on the tail of the population, but we experience performance on the average. As an immigrant from a third world country I think this is a big difference often invisible to the US-born citizens I talk to. Maybe why this is a great country and I am here. All I can say is that it is a world view that is not common world wide. Where I grew up, No Child Left Behind law would have been designed as 1 Child Left Behind. There just were not the resources, but more important, it was just more socially acceptable to just halve the no of slots halfway through an academic program, for example.

So I guess the question is why are we so focused on pushing services into the tails and will we continue to do so? Does society really benefit from having a larger fraction of the population capable of doing crappy algebra? Clearly there will be some point where the cost becomes prohibitive and it will stop: maybe that is what we are seeing now. But it is stunning that this was a 50 year process — if the dynamics in social policy “markets” are that slow it is going to be really difficult to manage.

CatCube::

I can’t help but wonder if part of the ever-expanding expenses isn’t that we mediate our interactions through the legal system more than we used to.

What got me thinking about it was what I'm working on this week. To answer Incurian's question above about why I was posting during the workday, I was avoiding working on a report for selecting a contractor for a project I'm working on. (I owe the taxpayer about three hours this weekend, since I spent time on Friday here and reading about the Oroville Dam spillway.)

We had contractors submit proposals, and we had two structural engineers and a construction quality assurance rep sit in a room for three days, writing our individual reports about each proposal, then coming to agreement about how we rate them. Then I have to write a report summarizing all of our individual reports, which gets fed into the arcane machine that will eventually spit out an award. This process costs about $10,000, and had zero value for evaluating the proposals. However, it has to be done this way or we'll get dragged around in court by an offeror's lawyer if they choose to make a case of their rejection. I think that in years past they'd use a simple low-bid process, which has its own problems, or the rejected contractors would bitch to their Congressmen or something but wouldn't literally make a federal case of it.

LukHamilton:

I think you gave short shrift to libertarian explanations of this phenomena. In particular, the Kling Theory of Public Choice may explain a significant fraction of cost disease: public policy will always choose to subsidize demand and restrict supply. If you restrict supply holding everything else equal, prices will go up. If you subsidize demand holding everything else equal, prices will go up. If you do both, prices will really go up.

(1) Healthcare: The government restricts the supply of all healthcare professionals (for example, doctors, nurses, CNAs, pharmacists, dentists, LPNs, etc.) via occupational licensing. (I should note that maybe everyone can get behind the

simple idea that the number of doctors per 10,000 people in the US should at least remain constant over time and not go down, as it has.) It restricts the supply of healthcare organizations (for example, hospitals, surgery centers, etc.) via onerous regulations, like the very ridiculous “certificate-of-need“. You have already explained in previous posts how things like the FDA restrict the supply of generic drugs. In terms of demand, the government subsidizes health insurance via the corporate income tax code, CHIP, the Obamacare marketplace, Medicaid, Medicare, etc.

(2) Education: I have done less investigation into this sector’s regulations. You mentioned Title IX. David Friedman has some nice blog posts on how the American Bar Association’s regulations on law schools make cheap law schools impossible. (This same concept also applies to healthcare-related professional schools, by the way.) If Bryan Caplan is right about signaling, a lot of education involves negative externalities, so it should be taxed or limited by the government. Instead, it subsidizes demand via loans. K-12 education, meanwhile, receives massive subsidies from the government; everyone can enjoy a totally free K-12 education.

(3) Real estate: Land-use regulations restrict the supply of housing. (Explanations of this can be found by googling “Matt Yglesias housing”.) It also subsidizes housing via Section 8, various other HUD programs, Freddie Mac, the mortgage-interest tax deduction, etc.

In short, any industry that the United States government has a heavy hand in has/will experience cost disease.

BenWave:

Scott, help me out here because I’ve read a long article about the mysterious nature of rising costs in certain sectors as well as hundreds of bemused comments, and the article had no more than a throwaway paragraph saying that maybe rising inequality is a sign that the ‘missing’ money is ending up in the pockets of the super-wealthy elite.

I come from a left-wing perspective, so I hope you can see that to me ex nihilo, “the super wealthy are becoming much richer than was historically the case, also all of these important services are becoming way more expensive than they used to be, but the one does not explain the other” looks like an extraordinary claim. I would like to see more evidence presented that this is not the case before updating in this direction!

In particular, I can see that a large majority of the odd features you have picked out about these services are acting exactly as predicted in Das Kapital volume 2, where Marx studies the process of realisation of invested capital (ie, money spent on labour, materials, tools etc) as the principal plus surplus value in money form. In particular, some of his predictions were:

1. Gains made by workers through collective action in sites of production can be taken away again by the landlord, the grocer, the financier etc.

2. The difficulty in the realisation of capital will incentivise businesses to strive for monopoly positions (whether by government mandate, mutual cooperations, quasi-monopolies such as real estate, branding and advertising).

3. The tensions between the production of surplus value and the realisation of surplus value will tend to set certain sectors of capital against one another – for example landlords would prefer if workers were well paid, but had to spend larger amounts of money on rent whereas factory owners would prefer to pay workers as little as possible, and that includes low housing costs.

Later analysis in the tradition of Marx have noticed that financial capital these days is doing very very well compared

> to workers, but also compared to traditional industrialists. And four out of the five of your examples are fields in which debt and financing plays a very large role. It's pretty easy to see that financial capital would be incentivised to make these things more expensive so that they can extract more money through larger loans and financing. (I'm not certain about subways. Are they typically debt-financed?).
>
> Financial capital certainly has the economic and political power to push for this, and they don't particularly care if they squeeze other holders of capital along they way. They are debt-financed fields in which large monopoly powers exist for one reason or another. And while I acknowledge that bureacratic bloat is certainly playing its role, I'm baffled by the relative lack of consideration of normal capitalist tendencies on this thread. As far as I can see it is the single most important factor driving up the costs of these services. Please present me with evidence that I am wrong about this!

Some additional links less-directly related or less easy to excerpt:

National Center for Policy Analysis: Should All Medicine Work Like Cosmetic Surgery? Because plastic surgery isn't a life-or-death need, it's not covered by insurance. Costs in the sector have risen 30% since 1992, compared to 118% for other types of health care. Does this mean that being sheltered from the insurance system has sheltered it from cost disease?

The American Interest: Why Can't We Have Nice Things? A breakdown of exactly why infrastructure and transportation projects cost so much more in the US than elsewhere, with an eye for Trump's promise of $1 trillion extra infrastructure spending.

Arnold Kling: What I Believe About Education. I have to include *one* "it's all the teachers' unions fault" post for completeness here.

Neerav Kingsland on education spending and the role of charter schools

The comment thread on Marginal Revolution contains some insight

The Incidental Economist: What Makes The US Health Care System So Expensive FAQ. From July 2011. Includes links to a lot of other things.

And some additional comments of my own:

I think any explanation that starts with "well, we have so much money now that we have to spend it on *something*…" ignores that many people do *not* have so much money, and in fact are really poor, but they get the same education and health care as the rest of us. If the problem were *just* "rich people looking for places to throw their money away", there would be other options for poor people who don't want to do that, the same way rich people have fancy restaurants where they can throw their money away and poor people have McDonalds.

Any explanation of the form "evil capitalists are scamming the rest of us for profit" has to explain why the cost increases are in the industries least exposed to evil capitalists. K-12 education is entirely nonprofit. Colleges are a mix but generally not owned by a single rich guy who gets all the money. My hospital is owned by an order of nuns; studies show that government hospitals have higher costs than for-profit ones. Meanwhile, the industries with the actual evil capitalists – tech, retail, restaurants, natural resources – seem mostly immune to the cost disease. This is not promising. Also, this wouldn't explain why so much of the money seems to be going to administrators/bells-and-whistles. If prices increase by $100,000, and the money goes to hiring two extra $50,000/year administrators, how does this help the capitalist profiting off it all?

Any explanation of the form "administrative bloat" or "inefficiency" has to explain why non-bloated alternatives don't pop up or become popular. I’m sure the CEO of Ford would love to just stop doing his job and approve every single funding request that passes his desk and pay for it by jacking up the price of cars, but at some point if he did that too much we’d all just buy Toyotas instead. Although there are some barriers to competition in the hospital market, there are fewer such barriers in the college, private school, and ambulatory clinic market. Why hasn’t competition discouraged administrative bloat here the same way it does in other industries?

Maybe a good time to reread the post How Likely Are Multifactorial Trends?

[REPOST] The Non-Libertarian FAQ

[This is a repost of the Non-Libertarian FAQ (aka “Why I Hate Your Freedom”), which I wrote about five years ago and which used to be hosted on my website. It no longer completely reflects my current views. I don’t think I’ve switched to believing anything on here is outright false, but I’ve moved on to different ways of thinking about certain areas. I’m reposting it by popular request and for historical interest only. I’ve made some very small updates, mostly listing rebuttals that came out over the past few years. I haven’t updated the statistics and everything is accurate as of several years ago. I seem to have lost the sources of my images, and I’m sorry; if I’ve used an image of yours, please let me know and I’ll cite you.]

Contents

0. Introduction

A. Economic Issues

1. Externalities
2. Coordination Problems
3. Irrational Choices
4. Lack of Information

B. Social Issues

5. Just Desserts and Social Mobility
6. Taxation

C. Political Issues

7. Competence of Government
8. Health Care

9. Prison Privatization
10. Gun Control
11. Education

D. Moral Issues

12. Moral Systems
13. Rights and Heuristics

E. Practical Issues

14. Slippery Slopes
15. Strategic Activism
16. Miscellaneous and Meta

Introduction

0.1: Are you a statist?

No.

Imagine a hypothetical country split between the “tallists”, who think only tall people should have political power, and the “shortists”, who believe such power should be reserved for the short.

If we met a tallist, we’d believe she was silly – but not because we favor the shortists instead. We’d oppose the tallists because we think the whole dichotomy is stupid – we should elect people based on qualities like their intelligence and leadership and morality. Knowing someone’s height isn’t enough to determine whether they’d be a good leader or not.

Declaring any non-libertarian to be a statist is as silly as declaring any non-tallist to be a shortist. Just as we can judge leaders on their merits and not on their height, so people can judge policies on their merits and not just on whether they increase or decrease the size of the state.

There are some people who legitimately believe that a policy's effect on the size of the state is so closely linked to its effectiveness that these two things are not worth distinguishing, and so one can be certain of a policy's greater effectiveness merely because it seems more libertarian and less statist than the alternative. Most of the rest of this FAQ will be an attempt to disprove this idea and assert that no, you really do have to judge the individual policy on its merits.

0.2: Do you hate libertarianism?

No.

To many people, libertarianism is a reaction against an over-regulated society, and an attempt to spread the word that some seemingly intractable problems can be solved by a hands-off approach. Many libertarians have made excellent arguments for why certain libertarian policies are the best options, and I agree with many of them. I think *this kind* of libertarianism is a valuable strain of political thought that deserves more attention, and I have no quarrel whatsoever with it and find myself leaning more and more in that direction myself.

However, there's a certain more aggressive, very American strain of libertarianism with which I do have a quarrel. This is the strain which, rather than analyzing specific policies and often deciding a more laissez-faire approach is best, starts with the tenet that government can do no right and private industry can do no wrong and uses this faith *in place of* more careful analysis. This faction is not averse to discussing politics, but tends to trot out the same few arguments about why less regulation *has* to be better. I wish I could blame this all on Ayn Rand, but a lot of it seems to come from people who have never heard of her. I suppose I could just add it to the bottom of the list of things I blame Reagan for.

To the first type of libertarian, I apologize for writing a FAQ attacking a caricature of your philosophy, but unfortunately that caricature is alive and well and posting smug slogans on Facebook.

0.3: Will this FAQ prove that government intervention always works better than the free market?

No, of course not.

Actually, in most cases, you won’t find me trying to make a positive proof of anything. I believe that deciding on, for example, an optimal taxation policy takes very many numbers and statistical models and other things which are well beyond the scope of this FAQ, and may well have different answers at different levels and in different areas.

What I want to do in most cases is not prove that the government works better than the free market, or vice versa, but to disprove theories that say we can be absolutely certain free market always works better than government before we even investigate the issue. After that, we may still find that this is indeed one of the cases where the free market works better than the government, but we will have to prove it instead of viewing it as self-evident from first principles.

0.4: Why write a Non-Libertarian FAQ? Isn’t statism a bigger problem than libertarianism?

Yes. But you never run into Stalinists at parties. At least not serious Stalinists over the age of twenty-five, and not the interesting type of parties. If I did, I guess I'd try to convince them not to be so statist, but the issue's never come up.

But the world seems positively full of libertarians nowadays. And I see very few attempts to provide a complete critique of libertarian philosophy. There are a bunch of ad hoc critiques of specific positions: people arguing for socialist health care, people in favor of gun control. But one of the things that draws people to libertarianism is that it is a unified, harmonious system. Unlike the mix-and-match philosophies of the Democratic and Republican parties, libertarianism is coherent and sometimes even derived from first principles. The only way to convincingly talk someone out of libertarianism is to launch a challenge on the entire system.

There are a few existing documents trying to do this (see Mike Huben's Critiques of Libertarianism and Mark Rosenfelder's What's (Still) Wrong With Libertarianism for two of the better ones), but I'm not satisfied with any of them. Some of them are good but incomplete. Others use things like social contract theory, which I find nonsensical and libertarians find repulsive. Or they have an overly rosy view of how consensual taxation is, which I don't fall for and which libertarians *definitely* don't fall for.

The main reason I'm writing this is that I encounter many libertarians, and I need a single document I can point to explaining why I don't agree with them. The existing anti-libertarian documentation makes too many arguments I don't agree with for me to feel really comfortable with it, so I'm writing this one myself. I don't encounter too many Stalinists,
so I don't have this problem with them and I don't see any need to write a rebuttal to their position.

If you really need a pro-libertarian FAQ to use on an overly statist friend, Google suggests The Libertarian FAQ.

0.5: How is this FAQ structured?

I’ve divided it into three main sections. The first addresses some very abstract principles of economics. They may not be directly relevant to politics, but since most libertarian philosophies start with abstract economic principles, a serious counterargument has to start there also. Fair warning: there are people who can discuss economics without it being INCREDIBLY MIND-NUMBINGLY BORING, but I am not one of them.

The second section deals with more concrete economic and political problems like the tax system, health care, and criminal justice.

The third section deals with moral issues, like whether it’s ever permissible to initiate force. Too often I find that if I can convince a libertarian that government regulation can be effective, they respond that it doesn’t matter because it’s morally repulsive, and then once I’ve finished convincing them it isn’t, they respond that it never works anyway. By having sections dedicated to both practical and moral issues, I hope to make that sort of bait-and-switch harder to achieve, and to allow libertarians to evaluate the moral and practical arguments against their position in whatever order they find appropriate.

Part A: Economic Issues

The Argument:

In a free market, all trade has to be voluntary, so you will never agree to a trade unless it benefits you.

Further, you won’t make a trade unless you think it’s the best possible trade you can make. If you knew you could make a better one, you’d hold out for that. So trades in a free market are not only better than nothing, they're also *the best possible transaction you could make at that time.*

Labor is no different from any other commercial transaction in this respect. You won't agree to a job unless it benefits you more than anything else you can do with your time, and your employer won't hire you unless it benefits her more than anything else she can do with her money. So a voluntarily agreed labor contract must benefit both parties, and must do so more than any other alternative.

If every trade in a free market benefits both parties, then any time the government tries to restrict trade in some way, it must hurt both parties. Or, to put it another way, you can help someone by giving them more options, but you can't help them by *taking away* options. And in a free market, where everyone starts with all options, all the government can do is take options away.

The Counterargument:

This treats the world as a series of producer-consumer dyads instead of as a system in which every transaction affects everyone else. Also, it treats consumers as coherent entities who have specific variables like "utility" and "demand" and know exactly what they are, which doesn't always work.

In the remainder of this section, I'll be going over several ways the free market can fail and several ways a regulated market can overcome those failures. I'll focus on four main things: externalities, coordination problems, irrational choice, and lack of information.

I did warn you it would be mind-numbingly boring.

1. Externalities

1.1: What is an externality?

An externality is when I make a trade with you, but it has some accidental effect on other people who weren’t involved in the trade.

Suppose for example that I sell my house to an amateur wasp farmer. Only he’s not a very good wasp farmer, so his wasps usually get loose and sting people all over the neighborhood every couple of days.

This trade between the wasp farmer and myself has benefited both of us, but it’s harmed people who weren’t consulted; namely, my neighbors, who are now locked indoors clutching cans of industrial-strength insect repellent. Although the trade was voluntary for both the wasp farmer and myself, it wasn’t voluntary for my neighbors.

Another example of externalities would be a widget factory that spews carcinogenic chemicals into the air. When I trade with the widget factory I’m benefiting – I get widgets – and they’re benefiting – they get money. But the people who breathe in the carcinogenic chemicals weren’t consulted in the trade.

1.2: But aren’t there are libertarian ways to solve externalities that don’t involve the use of force?

To some degree, yes. You can, for example, refuse to move into any neighborhood unless everyone in town has signed a contract agreeing not to raise wasps on their property.

But getting every single person in a town of thousands of people to sign a contract every time you think of something else you want banned might be a little difficult. More likely, you would want everyone in town to unanimously agree to a contract saying that certain things, which could be decided by some procedure requiring less than unanimity, could be banned from the neighborhood – sort of like the existing concept of neighborhood associations.

But convincing every single person in a town of thousands to join the neighborhood association would be near impossible, and all it would take would be a single holdout who starts raising wasps and all your work is useless. Better, perhaps, to start a new town on your own land with a pre-existing agreement that before you’re allowed to move in you must belong to the association and follow its rules. You could even collect dues from the members of this agreement to help pay for the people you’d need to enforce it.

But in this case, you’re not coming up with a clever libertarian way around government, you’re just reinventing the concept of government. There’s no difference between a town where to live there you have to agree to follow certain terms decided by association members following some procedure, pay dues, and suffer the consequences if you break the rules – and a regular town with a regular civic government.

As far as I know there is no loophole-free way to protect a community against externalities besides government and things that are functionally identical to it.

1.3: Couldn’t consumers boycott any company that causes externalities?

Only a small proportion of the people buying from a company will live near the company’s factory, so this assumes a colossal amount of both knowledge and altruism on the part of most consumers. See also the general discussion of why boycotts almost never solve problems in the next session.

1.4: What is the significance of externalities?

They justify some environmental, zoning, and property use regulations.

2. Coordination Problems

2.1: What are coordination problems?

Coordination problems are cases in which everyone agrees that a certain action would be best, but the free market cannot coordinate them into taking that action.

As a thought experiment, let's consider aquaculture (fish farming) in a lake. Imagine a lake with a thousand identical fish farms owned by a thousand competing companies. Each fish farm earns a profit of $1000/month. For a while, all is well.

But each fish farm produces waste, which fouls the water in the lake. Let's say each fish farm produces enough pollution to lower productivity in the lake by $1/month.

A thousand fish farms produce enough waste to lower productivity by $1000/month, meaning none of the fish farms are making any money. Capitalism to the rescue: someone invents a complex filtering system that removes waste products. It costs $300/month to operate. All fish farms voluntarily install it, the pollution ends, and the fish farms are now making a profit of $700/month – still a respectable sum.

But one farmer (let's call him Steve) gets tired of spending the money to operate his filter. Now one fish farm worth of waste is polluting the lake, lowering productivity by $1. Steve earns $999 profit, and everyone else earns $699 profit.

Everyone else sees Steve is much more profitable than they are, because he's not spending the maintenance costs on his filter. They disconnect their filters too.

Once four hundred people disconnect their filters, Steve is earning $600/month – less than he would be if he and everyone else had kept their filters on! And the poor virtuous filter users are only making $300. Steve goes around to everyone, saying "Wait! We all need to make a voluntary pact to use filters! Otherwise, everyone's productivity goes down."

Everyone agrees with him, and they all sign the Filter Pact, except one person who is sort of a jerk. Let's call him Mike. Now everyone is back using filters again, except Mike. Mike earns $999/month, and everyone else earns $699/month. Slowly, people start thinking they too should be getting big bucks like Mike, and disconnect their filter for $300 extra profit…

A self-interested person never has any incentive to use a filter. A self-interested person has some incentive to sign a pact to make everyone use a filter, but in many cases has a stronger incentive to wait for everyone *else* to sign such a pact but opt out himself. This can lead to an undesirable equilibrium in which no one will sign such a pact.

The most profitable solution to this problem is for Steve to declare himself King of the Lake and threaten to initiate force against anyone who doesn't use a filter. This regulatory solution leads to greater total productivity for the thousand fish farms than a free market could.

The classic libertarian solution to this problem is to try to find a way to privatize the shared resource (in this case, the lake). I intentionally chose aquaculture for this example because privatization doesn't work. Even after the entire lake has been divided into parcels and sold to private landowners (waterowners?) the problem remains, since waste will spread from one parcel to another regardless of property boundaries.

2.1.1: Even without anyone declaring himself King of the Lake, the fish farmers would voluntarily agree to abide by the pact that benefits everyone.

Empirically, no. This situation happens with wild fisheries all the time. There’s some population of cod or salmon or something which will be self-sustaining as long as it’s not overfished. Fishermen come in and catch as many fish as they can, overfishing it. Environmentalists warn that the fishery is going to collapse. Fishermen find this worrying, but none of them want to fish less because then their competitors will just take up the slack. Then the fishery collapses and everyone goes out of business. The most famous example is the Collapse of the Northern Cod Fishery, but there are many others in various oceans, lakes, and rivers.

If not for resistance to government regulation, the Canadian governments could have set strict fishing quotas, and companies could still be profitably fishing the area today. Other fisheries that do have government-imposed quotas are much more successful.

2.1.2: I bet [extremely complex privatization scheme that takes into account the ability of cod to move across property boundaries and the migration patterns of cod and so on] could have saved the Atlantic cod too.

Maybe, but *left to their own devices, cod fishermen never implemented or recommended that scheme*. If we ban all government regulation in the environment, that won’t make fishermen suddenly start implementing complex privatization schemes that they’ve never implemented before. It will just make fishermen keep doing what they’re doing while tying the hands of the one organization that has a track record of actually solving this sort of problem in the real world.

2.2: How do coordination problems justify environmental regulations?

Consider the process of trying to stop global warming. If everyone believes in global warming and wants to stop it, it’s still not in any one person’s self-interest to be more environmentally conscious. After all, that would make a major impact on her quality of life, but a negligible difference to overall worldwide temperatures. If everyone acts only in their self-interest, then no one will act against global warming, even though stopping global warming is in everyone’s self-interest. However, everyone would support the institution of a government that uses force to make *everyone* more environmentally conscious.

Notice how well this explains reality. The government of every major country has publicly declared that they think solving global warming is a high priority, but every time they meet in Kyoto or Copenhagen or Bangkok for one of their big conferences, the developed countries would rather the developing countries shoulder the burden, the developing countries would rather the developed countries do the hard work, and so nothing ever gets done.

The same applies *mutans mutandis* to other environmental issues like the ozone layer, recycling, and anything else where one person cannot make a major difference but many people acting together can.

2.3: How do coordination problems justify regulation of ethical business practices?

The normal libertarian belief is that it is unnecessary for government to regulate ethical business practices. After all, if people object to something a business is doing, they will boycott that business, either incentivizing the business to change its ways, or driving them into well-deserved bankruptcy. And if people don’t object, then there’s no problem and the government shouldn’t intervene.

A close consideration of coordination problems demolishes this argument. Let’s say Wanda’s Widgets has one million customers. Each customer pays it $100 per year, for a total income of $100 million. Each customer prefers Wanda to her competitor Wayland, who charges $150 for widgets of equal quality. Now let’s say Wanda’s Widgets does some unspeakably horrible act which makes it $10 million per year, but offends every one of its million customers.

There is no incentive for a single customer to boycott Wanda’s Widgets. After all, that customer’s boycott will cost the customer $50 (she will have to switch to Wayland) and make an insignificant difference to Wanda (who is still earning $99,999,900 of her original hundred million). The customer takes significant inconvenience, and Wanda neither cares nor stops doing her unspeakably horrible act (after all, it’s giving her $10 million per year, and only losing her $100).

The only reason it would be in a customer’s interests to boycott is if she believed over a hundred thousand other customers would join her. In that case, the boycott would be costing Wanda more than the $10 million she gains from her unspeakably horrible act, and it’s now in her self-interest to stop committing the act. However, unless each boycotter believes 99,999 others will join her, she is inconveniencing herself for no benefit.

Furthermore, if a customer offended by Wanda’s actions believes 100,000 others will boycott Wanda, then it’s in the customer’s self-interest to “defect” from the boycott and buy Wanda’s products. After all, the customer will lose money if she buys Wayland’s more expensive widgets, and this is unnecessary – the 100,000 other boycotters will change Wanda’s mind with or without her participation.

This suggests a “market failure” of boycotts, which seems confirmed by experience. We know that, despite many companies doing very controversial things, there have been very few successful boycotts. Indeed, few boycotts, successful or otherwise, ever make the news, and the number of successful boycotts seems much less than the amount of outrage expressed at companies’ actions.

The existence of government regulation solves this problem nicely. If >51% of people disagree with Wanda’s unspeakably horrible act, they don’t need to waste time and money guessing how many of them will join in a boycott, and they don’t need to worry about being unable to conscript enough defectors to reach critical mass. They simply vote to pass a law banning the action.

2.3.1: I’m not convinced that it’s really that hard to get a boycott going. If people really object to something, they’ll start a boycott regardless of all that coordination problem stuff.

So, you’re boycotting Coke because they’re hiring local death squads to kidnap, torture, and murder union members and organizers in their sweatshops in Colombia, right?

Not a lot of people to whom I have asked this question have ever answered “yes”. Most of them had never heard of the abuses before. A few of them vaguely remembered having heard something about it, but dismissed it as “you know, multinational corporations do a lot of sketchy things.” I’ve only met one person who’s ever gone so far as to walk twenty feet further to get to the Pepsi vending machine.

If you went up to a random guy on the street and said “Hey, does hiring death squads to torture and kill Colombians who protest about terrible working conditions bother you?” 99.9% of people would say yes. So why the disconnect between words and actions? People could just be lying – they could say they cared so they sounded compassionate, but in reality it doesn’t really bother them.

But maybe it’s something more complicated. Perhaps they don’t have the brainpower to keep track of every single corporation that’s doing bad things and just how bad they are. Perhaps they’ve compartmentalized their lives and after they leave their Amnesty meetings it just doesn’t register that they should change their behaviour in the supermarket. Or perhaps the Coke = evil connection is too tenuous and against the brain’s ingrained laws of thought to stay relevant without expending extraordinary amounts of willpower. Or perhaps there’s some part of the subconscious that really is worry about that game theory and figuring it has no personal incentive to join the boycott.

And God forbid that it’s something more complicated than that. Imagine if the company that made the mining equipment that was bought by the mining company that mined the aluminum that was bought by Coke to make their cans was doing something unethical. You think you could convince enough people to boycott Coke that Coke would boycott the mining company that the mining company would boycott the equipment company that the equipment company would stop behaving unethically?

If we can't trust people to stay off Coke when it uses death squads and when Pepsi tastes *exactly the same* (don't argue with me on that one!) how can we assume people's purchasing decisions will always act as a general moral regulatory method for the market?

2.3.2: And you really think governments can do better?

Sure seems that way. Many laws currently exist banning businesses from engaging in unethical practices. Some of these laws were passed by direct ballot. Others were passed by representatives who have incentives to usually follow the will of their constituents. So it seems fair to say that there are a lot of business practices that more than 51% of people thought should be banned.

But the very fact that a law was needed to ban them proves that those 51% of people weren't able to organize a successful boycott. More than half of the population, sometimes much more, hated some practice so much they thought it should be *illegal*, yet that wasn't enough to provide an incentive for the company to stop doing it until the law took effect.

To me, that confirms that boycotts are a very poor way of allowing people's morals to influence corporate conduct.

2.4: How do coordination problems justify government spending on charitable causes?

Because failure to donate to a charitable cause might also be because of a coordination problem.

How many people want to end world hunger? I’ve never yet met someone who would answer with a “not me!”, but maybe some of those people are just trying to look good in front of other people, so let’s make a conservative estimate of 50%.

There’s a lot of dispute over what it would mean to “end world hunger”, all the way from “buy and ship food every day to everyone who is hungry that day” all the way to “create sustainable infrastructure and economic development such that everyone naturally produces enough food or money”. There are various estimates about how much these different definitions would cost, all the way from “about $15 billion a year” to “about $200 billion a year” – permanently in the case of shipping food, and for a decade or two in the case of promoting development.

Even if we take the highest possible estimate, it’s still *well* below what you would make if 50% of the population of the world donated $1/week to the cause. Now, certainly there are some very poor people in the world who couldn’t donate $1/week, but there are also some very rich people who could no doubt donate much, much more.

So we have two possibilities. Either the majority of people don’t care enough about world hunger to give a dollar a week to end it, or something else is going on.

That something else is a coordination problem. No one expects anyone else to donate a dollar a week, so they don’t either. And although somebody could shout very loudly “Hey, let’s all donate $1 a week to fight world hunger!” no one would expect anyone else to listen to that person, so they wouldn’t either.

When the government levies tax money on everyone in the country and then donates it to a charitable cause, it is often because everyone in the country supports that charitable cause but a private attempt to show that support would fall victim to coordination problems.

2.5: How do coordination problems justify labor unions and other labor regulation?

It is frequently proposed that workers and bosses are equal negotiating partners bargaining on equal terms, and only the excessive government intervention on the side of labor that makes the negotiating table unfair. After all, both need something from one another: the worker needs money, the boss labor. Both can end the deal if they don’t like the terms: the boss can fire the worker, or the worker can quit the boss. Both have other choices: the boss can choose a different employee, the worker can work for a different company. And yet, strange to behold, having proven the fundamental equality of workers and bosses, we find that everyone keeps acting as if bosses have the better end of the deal.

During interviews, the prospective employee is often nervous; the boss rarely is. The boss can ask all sorts of things like that the prospective pay for her own background check, or pee in a cup so the boss can test the urine for drugs; the prospective employee would think twice before daring make even so reasonable a request as a cup of coffee. Once the employee is hired, the boss may ask on a moment’s notice that she work a half hour longer or else she’s fired, and she may not dare to even complain. On the other hand, if she were to so much as ask to be allowed to start work thirty minutes later to get more sleep or else she’ll quit, she might well be laughed out of the company. A boss may, and very often does, yell at an employee who has made a minor mistake, telling her how stupid and worthless she is, but rarely could an employee get away with even politely mentioning the mistake of a boss, even if it is many times as unforgivable.

The naive economist who truly believes in the equal bargaining position of labor and capital would find all of these things very puzzling.

Let’s focus on the last issue; a boss berating an employee, versus an employee berating a boss. Maybe the boss has one hundred employees. Each of these employees only has one job. If the boss decides she dislikes an employee, she can drive her to quit and still be 99% as productive while she looks for a replacement; once the replacement is found, the company will go on exactly as smoothly as before.

But if the employee’s actions drive the boss to fire her, then she must be completely unemployed until such time as she finds a new job, suffering a long period of 0% productivity. Her new job may require a completely different life routine, including working different hours, learning different skills, or moving to an entirely new city. And because people often get promoted based on seniority, she probably won’t be as well paid or have as many opportunities as she did at her old company. And of course, there’s always the chance she won’t find another job at all, or will only find one in a much less tolerable field like fast food.

We previously proposed a symmetry between a boss firing a worker and a worker quitting a boss, but actually they could not be more different. For a boss to fire a worker is at most a minor inconvenience; for a worker to lose a job is a disaster. The Holmes-Rahe Stress Scale, a measure of the comparative stress level of different life events, puts being fired at 47 units, worse than the death of a close friend and nearly as bad as a jail term. Tellingly, “firing one of your employees” failed to make the scale.

This fundamental asymmetry gives capital the power to create more asymmetries in its favor. For example, bosses retain a level of control on workers even after they quit, because a worker may very well need a letter of reference from a previous boss to get a good job at a new company. On the other hand, a prospective employee who asked her prospective boss to produce letters of recommendation from her previous workers would be politely shown the door; we find even the image funny.

The proper level negotiating partner to a boss is not one worker, but all workers. If the boss lost *all* workers at once, then she would be at 0% productivity, the same as the worker who loses her job. Likewise, if *all* the workers approached the boss and said “We want to start a half hour later in the morning or we all quit”, they might receive the same attention as the boss who said “Work a half hour longer each day or you’re all fired”.

But getting all the workers together presents coordination problems. One worker has to be the first to speak up. But if one worker speaks up and doesn’t get immediate support from all the other workers, the boss can just fire that first worker as a troublemaker. Being the first worker to speak up has major costs – a good chance of being fired – but no benefits – all workers will benefit equally from revised policies no matter who the first worker to ask for them is.

Or, to look at it from the other angle, if only one worker sticks up for the boss, then intolerable conditions may well still get changed, but the boss will remember that one worker and maybe be more likely to promote her. So even someone who hates the boss’s policies has a strong selfish incentive to stick up for her.

The ability of workers to coordinate action without being threatened or fired for attempting to do so is the only thing that gives them any negotiating power at all, and is necessary for a healthy labor market. Although we can debate the specifics of exactly how much protection should be afforded each kind of coordination, the fundamental principle is sound.

2.5.1: But workers don’t need to coordinate. If working conditions are bad, people can just change jobs, and that would solve the bad conditions.

About three hundred Americans commit suicide for work-related reasons every year – this number doesn’t count those who attempt suicide but fail. The reasons cited by suicide notes, survivors and researchers investigating the phenomenon include on-the-job bullying, poor working conditions, unbearable hours, and fear of being fired.

I don’t claim to understand the thought processes that would drive someone to do this, but given the rarity and extremity of suicide, we can assume for every worker who goes ahead with suicide for work-related reasons, there are a hundred or a thousand who feel miserable but not quite suicidal.

If people are literally killing themselves because of bad working conditions, it’s safe to say that life is more complicated than the ideal world in which everyone who didn’t like their working conditions quits and get a better job elsewhere (see the next section, Irrationality).

I note in the same vein stories from the days before labor regulations when employers would ban workers from using the restroom on jobs with nine hour shifts, often ending in the workers wetting themselves. This seems like the sort of thing that provides so much humiliation to the workers, and so little benefit to the bosses, that a free market would eliminate it in a split second. But we know that it was a common policy in the 1910s and 1920s, and that factories with such policies never wanted for employees. The same is true of factories that literally locked their workers inside to prevent them from secretly using the restroom or going out for a smoking break, leading to disasters like the Triangle Shirtwaist Fire when hundreds of workers died when the building they were locked inside burnt down. And yet even after this fire, the practice of locking workers inside buildings only stopped when the government finally passed regulation against it.

3. Irrational Choices

3.1: What do you mean by “irrational choices”?

A company (Thaler, 2007, download study as .pdf) gives its employees the opportunity to sign up for a pension plan. They contribute a small amount of money each month, and the company will also contribute some money, and overall it ends up as a really good deal for the employees and gives them an excellent retirement fund. Only a small minority of the employees sign up.

The libertarian would answer that this is fine. Although some outsider might condescendingly declare it “a really good deal”, the employees are the most likely to understand their own unique financial situation. They may have a better pension plan somewhere else, or mistrust the company’s promises, or expect not to need much money in their own age. For some outsider to declare that they are *wrong* to avoid the pension plan, or worse to try to *force* them into it for their own good, would be the worst sort of arrogant paternalism, and an attack on the employees’ dignity as rational beings.

Then the company switches tactics. It automatically signs the employees up for the pension plan, but offers them the option to opt out. This time, only a small minority of the employees opt out.

That makes it very hard to spin the first condition as the employees rationally preferring not to participate in the pension plan, since the second condition reveals the opposite preference. It looks more like they just didn’t have the mental energy to think about it or go through the trouble of signing up. And in the latter condition, they didn’t have the mental energy to think about it or go through the trouble of opting out.

If the employees were rationally deciding whether or not to sign up, then some outsider regulating their decision would be a disaster. But if the employees are making demonstrably irrational choices because of a lack of mental energy, and if people do so consistently and predictably, then having someone else who has considered the issue in more depth regulate their choices could lead to a better outcome.

3.1.1: So what’s going on here?

Old-school economics assumed choice to be “revealed preference”: an individual’s choices will invariably correspond to their preferences, and imposing any other set of choices on them will result in fewer preferences being satisfied.

In some cases, economists have gone to absurd lengths to defend this model. For example, Bryan Caplan says that when drug addicts say they wish that they could quit drugs, they must be lying, since they haven’t done so. Seemingly unsuccessful attempts to quit must be elaborate theater, done to convince other people to continue supporting them, while they secretly enjoy their drugs as much as ever.

But the past fifty years of cognitive science have thoroughly demolished this “revealed preference” assumption, showing that people’s choices result from a complex mix of external compulsions, internal motivations, natural biases, and impulsive behaviors. These decisions usually approximate fulfilling preferences, but sometimes they fail in predictable and consistent ways.

The field built upon these insights is called “behavioral economics”, and you can find more information in books like Judgment Under Uncertainty, Cognitive Illusions, and Predictably Irrational, or on the website Less Wrong.

3.2: Why does this matter?

The gist of this research, as it relates to the current topic, is that people don’t always make the best choice according to their preferences. Sometimes they consistently make the easiest or the most superficially attractive choice instead. It may be best not to think of them as a “choice” at all, but as a reflexive reaction to certain circumstances, which often but not always conforms to rationality.

Such possibilities cast doubt on the principle that every trade that can be voluntarily made should be voluntarily made.

If people’s decisions are not randomly irrational, but systematically irrational in predictable ways, that raises the possibility that people who are aware of these irrationalities may be able to do better than the average person in particular fields where the irrationalities are more common, raising the possibility that paternalism can sometimes be justified.

3.2.1: Why should the government protect people from their own irrational choices?

By definition of “irrational”, people will be happier and have more of their preferences satisfied if they do not make irrational choices. By the principles of the free market, as people make more rational decisions the economy will also improve.

If you mean this question in a *moral* sense, more like “How *dare* the government presume to protect me from my own irrational choices!”, see the section on Moral Issues.

3.2.2: What is the significance of predictably irrational behavior?

It justifies government-mandated pensions, some consumer safety and labor regulations, advertising regulations, concern about addictive drugs, and public health promotion, among other things.

4. Lack of Information

4.1: What do you mean by “lack of information”?

Many economic theories start with the assumption that everyone has perfect information about everything. For example, if a company’s products are unsafe, these economic theories assume consumers know the product is unsafe, and so will buy less of it.

No economist literally believes consumers have perfect information, but there are still strong arguments for keeping the “perfect information” assumption. These revolve around the idea that consumers will be motivated to pursue information about things that are important to them. For example, if they care about product safety, they will fund investigations into product safety, or only buy products that have been certified safe by some credible third party. The only case in which a consumer would buy something without information on it is if the consumer had no interest in the information, or wasn’t willing to pay as much for the information as it would cost, in which case the consumer doesn’t care much about the information anyway, and it is a success rather than a failure of the market that it has not given it to her.

In nonlibertarian thought, people care so much about things like product safety and efficacy, or the ethics of how a product is produced, that the government needs to ensure them. In libertarian thought, if people really care about product safety, efficacy and ethics, the market will ensure them itself, and if they genuinely don’t care, that’s okay too.

4.1.1: And what’s wrong with the libertarian position here?

Section 5 describes how we can sometimes predict when people will make irrational choices. One of the most consistent irrational choices people make is buying products without spending as much effort to gather information as the amount they care about these things would suggest. So in fact, the nonlibertarians are right: if there were no government regulation, people who care a lot about things like safety and efficacy would consistently be stuck with unsafe and ineffective products, and the market would not correct these failures.

4.2: Is this really true? Surely people would investigate the safety, ethics, and efficacy of the products they buy.

Below follows a list of statements about products. Some are real, others are made up. Can you identify which are which?

1. Some processed food items, including most Kraft cheese products, contain methylarachinate, an additive which

causes a dangerous anaphylactic reaction in 1/31000 people who consume it. They have been banned in Canada, but continue to be used in the United States after intense lobbying from food industry interests.

2. Commonly used US-manufactured wood products, including almost all plywood, contain formaldehyde, a compound known to cause cancer. This has been known in scientific circles for years, but was only officially reported a few months ago because of intense chemical industry lobbying to keep it secret. Formaldehyde-containing wood products are illegal in the EU and most other developed nations.

3. Total S.A., an oil company that owns fill-up stations around the world, sometimes uses slave labor in repressive third-world countries to build its pipelines and oil wells. Laborers are coerced to work for the company by juntas funded by the corporation, and are shot or tortured if they refuse. The company also helps pay for the military muscle needed to keep the juntas in power.

4. Microsoft has cooperated with the Chinese government by turning over records from the Chinese equivalents of its search engine “Bing” and its hotmail email service, despite knowing these records would be used to arrest dissidents. At least three dissidents were arrested based on the information and are currently believed to be in jail or “re-education” centers.

5. Wellpoint, the second largest US health care company, has a long record of refusing to provide expensive health care treatments promised in some of its plans by arguing that their customers have violated the “small print” of the terms of agreement; in fact they make it so technical that almost all customers violate them unknowingly, then only cite the ones who need expensive treatment. Although it has been

sued for these practices at least twice, both times it has used its legal muscle to tie the cases up in court long enough that the patients settled for an undisclosed amount believed to be fraction of the original benefits promised.

6. Ultrasonic mosquito repellents like those made by GSI, which claim to mimic frequencies produced by the mosquito’s natural predator, the bat, do not actually repel mosquitoes. Studies have shown that exactly as many mosquitoes inhabit the vicinity of such a mosquito repellent as anywhere else.

7. Listerine (and related mouth washes) probably do not eliminate bad breath. Although it may be effective at first, in the long term it generally increases bad breath by drying out the mouth and inhibiting the salivary glands. This may also increase the population of dental bacteria. Most top dentists recommend avoiding mouth wash or using it very sparingly.

8. The most popular laundry detergents, including most varieties of Tide and Method, have minimal to zero ability to remove stains from clothing. They mostly just makes clothing smell better when removed from the laundry. Some of the more expensive alkylbenzenesulfonate detergents have genuine stain-removing action, but aside from the cost, these detergents have very strong smells and are unpopular.

4.2.1: Okay, I admit I’m not sure of most of these. What’s your point?

This is a complicated FAQ about complicated philosophical issues. Most likely its readers are in the top few percentiles in terms of intelligence and education.

And we live in a world where there are many organizations, both private and governmental, that exist to evaluate products and disseminate information about their safety.

And all of the companies and products above are popular ones that most American consumers have encountered and had to make purchasing decisions about. I tried to choose safety issues that were extremely serious and carried significant risks of death, and ethical issues involving slavery and communism, which would be of particular importance to libertarians.

If the test was challenging, it means that the smartest and best-educated people in a world full of consumer safety and education organizations don't bother to look up important life-or-death facts specifically tailored to be relevant to them about the most popular products and companies they use every day.

And if that's the case, why would you believe that less well-educated people in a world with less consumer safety information trying to draw finer distinctions between more obscure products will definitely seek out the consumer information necessary allows them to avoid unsafe, unethical, or ineffective products?

The above test is an attempt at experimental proof that people don't seek out even the product information that is genuinely important to them, but instead take the easy choice of buying whatever's convenient based on information they get from advertising campaigns and the like.

4.2.2: Fine, fine, what are the answers to the test?

Four of them are true and four of them are false, but I'm not saying which are which, in the hopes that people will observe their own thought processes when deciding whether or not it's worth looking up.

4.2.3: Right, well of course people don't look up product information *now* because the government regulates that for them. In a real libertarian society, they would be more proactive.

All of the four true items on the test above are true *in spite* of government regulation. Clearly, there are still significant issues even in a regulated environment.

If you honestly believe you have no incentive to look up product information because you trust the government to take care of that, then you're about ten times more statist than I am, and *I'm the guy writing the Non-Libertarian FAQ*.

4.3: What other unexpected consequences might occur without consumer regulation?

It could destroy small business.

In the absence of government regulation, you would have to trust corporate self-interest to regulate quality. And to some degree you can do that. Wal-Mart and Target are both big enough and important enough that if they sold tainted products, it would make it into the newspaper, there would be a big outcry, and they would be forced to stop. One could feel quite safe shopping at Wal-Mart.

But suppose on the way to Wal-Mart, you see a random mom-and-pop store that looks interesting. What do you know about its safety standards? Nothing. If they sold tainted or defective products, it would be unlikely to make the news; if it were a small enough store, it might not even make the Internet. Although you expect the CEO of Wal-Mart to be a reasonable man who understands his own self-interest and who would enforce strict safety standards, you have no idea whether the owner of the mom-and-pop store is stupid, lazy, or just assumes (with some justification) that no one will ever notice his misdeeds. So you avoid the unknown quantity and head to Wal-Mart, which you know is safe.

Repeated across a million people in a thousand cities, big businesses get bigger and small businesses get unsustainable.

4.4: What is the significance of lack of information?

It justifies some consumer and safety regulations, and the taxes necessary to pay for them.

Part B: Social Issues

The Argument:

Those who work hardest (and smartest) should get the most money. Not only should we not begrudge them that money, but we should thank them for the good they must have done for the world in order to satisfy so many consumers.

People who do not work hard should not get as much money. If they want more money, they should work harder. Getting more money without working harder or smarter is unfair, and indicative of a false sense of entitlement.

Unfortunately, modern liberal society has internalized the opposite principle: that those who work hardest are greedy people who must have stolen from those who work less hard, and that we should distrust them at until they give most of their ill-gotten gains away to others. The “progressive” taxation system as it currently exists serves this purpose.

This way of thinking is not only morally wrong-headed, but economically catastrophic. Leaving wealth in the hands of the rich would “make the pie bigger”, allowing the extra wealth to “trickle down” to the poor naturally.

The Counterargument:

Hard work and intelligence are contributory factors to success, but depending on the way you phrase the question, you find you need other factors to explain between one-half and nine-tenths of the difference in success within the United States; within the world at large the numbers are much higher.

If we think factors other than hard work and intelligence determining success are "unfair", then most of Americans' life experiences are determined by "unfair" factors.

Although it would be overly ambitious to want to completely eliminate all unfairness, we know that most other developed countries have successfully eliminated many of the most glaring types of unfairness, and reaped benefits greater than the costs from doing so.

The progressive tax system is part of this policy of eliminating unfairness, but if you disagree with that, that's okay, as more and more of the country's wealth is staying in the hands of the super-rich. None of this wealth has trickled down to the poor and none of it ever will, as the past thirty years of economic history have repeatedly and decisively demolished the "trickle-down" concept.

None of this implies that any particular rich person is "greedy", whatever that would mean.

5. Just Desserts and Social Mobility

5.1: Government is the recourse of "moochers", who want to take the money of productive people and give it to the poor. But rich people earned their money, and poor people had the chance to earn money but

did not. Therefore, the poor do not deserve rich people’s money.

The claim of many libertarians is that the wealthy earned their money by the sweat of their brow, and the poor are poor because they did not. The counterclaim of many liberals is that the wealthy gained their wealth by various unfair advantages, and that the poor never had a chance. These two conflicting worldviews have been the crux of many an Internet flamewar.

Luckily, this is an empirical question, and can be solved simply by collecting the relevant data. For example, we could examine whether the children of rich parents are more likely to be rich than poor parents, and, if so, how much more likely they are. This would give us a pretty good estimate of how much of rich people’s wealth comes from superior personal qualities, as opposed to starting with more advantages.

If we define “rich” as “income in the top 5%” and “poor” as “income in the bottom 5%” then children of rich parents are about twenty times more likely to become rich themselves than children of poor parents.

But maybe that’s an extreme case. Instead let’s talk about “upper class” (top 20%) and “lower class” (bottom 20%). A person born to a lower-class family only has a fifty-fifty chance of ever breaking out of the lower class (as opposed to 80% expected by chance), and only about a 3% chance of ending up in the upper class (as opposed to 20% expected by chance). The children of upper class parents are six times more likely to end up in the upper class than the lower class; the children of lower class families are four times more likely to end up in the lower class than the upper class.

The most precise way to measure this question is via a statistic called “intergenerational income mobility”, which studies have estimated at between .4 and .6. This means that around half the difference in people’s wealth, maybe more, can be explained solely by who their parents are.

Once you add in all the other factors besides how hard you work – like where you live (the average Delawarean earns $30000; the average Mississippian $15000) and the quality of your local school district, there doesn’t seem to be much room for hard work to determine more than about a third of the difference between income.

5.1.1: The conventional wisdom among libertarians is completely different. I’ve heard of a study saying that people in the lower class are more likely to end up in the upper class than stay in the lower class, even over a period as short as ten years!

First of all, note that this is insane. Since the total must add up to 100%, this would mean that starting off poor actually makes you more likely to end up rich than someone who didn’t start off poor. If this were true, we should all send our children to school in the ghetto to maximize their life chances. This should be a red flag.

And, in fact, it is false. Most of the claims of this sort come from a single discredited study. The study focused on a cohort with a median age of twenty-two, then watched them for ten years, then compared the (thirty-two-year-old) origins with twenty-two-year-olds, then claimed that the fact that young professionals make more than college students was a fact about social mobility. It was kind of weird.

Why would someone do this? Far be it from me to point fingers, but Glenn Hubbard, the guy who conducted the study, worked for a conservative think tank called the

“American Enterprise Institute”. You can see a more complete criticism of the study here.

5.1.2: Okay, I acknowledge that at least half of the differences in wealth can be explained by parents. But that needn’t be rich parents leaving trust funds to their children. It could also be parents simply teaching their children better life habits. It could even be genes for intelligence and hard work.

This may explain a small part of the issue, but see 5.1.3 and 5.1.3.1, which show that under different socioeconomic conditions, this number markedly decreases. These socioeconomic changes would not be expected to affect things like genetics.

5.1.3: So maybe children of the rich do have better opportunities, but that’s life. Some people just start with advantages not available to others. There’s no point in trying to use Big Government to regulate away something that’s part of the human condition.

This lack of social mobility isn’t part of the human condition, it’s a uniquely American problem. Of eleven developed countries investigated in a recent study on income mobility, America came out tenth out of eleven. Their calculation of US intergenerational income elasticity (the number previously cited as probably between .4 and .6) was .47. But other countries in the study had income elasticity as low as .15 (Denmark), .16 (Australia), .17 (Norway), and .19 (Canada). In each of those countries, the overwhelming majority of wealth is earned by hard work rather than inherited.

The United States, is just particularly bad at this; the American Dream turns out to be the “nearly every developed country except America” Dream.

5.1.3.1: That’s depressing, but don’t try to turn it into a political narrative. Given the government’s incompetence and wastefulness, there’s no reason to think more government regulation and spending could possibly improve social mobility at all.

Studies show that increasing government spending significantly improves social mobility. States with higher government spending have about 33% more social mobility than states with lower spending.

This also helps explain why other First World countries have better social mobility than we do. Poor American children have very few chances to go to Harvard or Yale; poor Canadian children have a much better chance to go to to UToronto or McGill, where most of their tuition is government-subsidized.

5.2: Then perhaps it is true that rich children start out with a major unfair advantage. But this advantage can be overcome. Poor children may have to work harder than rich children to become rich adults, but this is still *possible*, and so it is still true, in the important sense, that if you are not rich it’s mostly your own fault.

Several years ago, I had an interesting discussion with an evangelical Christian on the ethics of justification by faith. I promise you this will be relevant eventually.

I argued that it is unfair for God to restrict entry to Heaven to Christians alone. After all, 99% of native-born Ecuadorans are Christian, but less than 1% of native born Saudis are same. It follows that the chance of any native-born Ecuadorian of becoming Christian is 99%, and that of any native born Saudi, 1%. So if God judges people by their religion, then within 1% He’s basically just decided it’s free entry for Ecuadorians, but people born in Saudi Arabia can go to hell (literally).

My Christian friend argued that is not so: that there is a great difference between 0% of Saudis and 1% of Saudis. I answered that no, there was a 1% difference. But he said this 1% proves that the Saudis had free will: that even though all the cards were stacked against them, a few rare Saudis could still choose Christianity.

But what does it mean to have free will, if external circumstances can make 99% of people with free will decide one way in Ecuador, and the opposite way in Saudi Arabia?

I do sort of believe in free will, or at least in “free will”. But where my friend’s free will was unidirectional, an arrow pointing from MIND to WORLD, my idea of free will is circular: MIND affects WORLD affects MIND affects WORLD and so on.

Yes, it is ultimately the mind and nothing else that decides whether to accept or reject Islam or Christianity. But it is the world that shapes the mind before it does its accepting or rejecting. A man raised in Saudi Arabia uses a mind forged by Saudi culture to make the decision, and chooses Islam. A woman raised in Ecuador uses a mind forged by Ecuador to make the decision, and chooses Christianity. And so there is no contradiction in the saying that the decision between Islam and Christianity is up entirely to the individual, yet that it is almost entirely culturally determined. For the mind is a box, filled with genes and ideas, and although it is a wonderful magical box that can take things and combine them and forge them into something quite different and unexpected, it is not infinitely magical, and it cannot create out of thin air.

Returning to the question at hand, every poor person has the opportunity to work hard and eventually become rich.

Whether that poor person grasps the opportunity comes from that person's own personality. And that person's own personality derives eventually from factors outside that person's control. A clear look at the matter proves it must be so, or else personality would be self-created, like the story of the young man who received a gift of a time machine from a mysterious aged stranger, spent his life exploring past and future, and, in his own age, goes back and gives his time machine to his younger self.

5.2.1: And why is this relevant to politics?

Earlier, I offered a number between .4 and .6 as the proportion of success attributable solely to one's parents' social class. This bears on, but does not wholly answer, a related question: what percentage of my success is my own, and what percentage is attributable to society? People have given answers to this question as diverse as (100%, 0%), (50%, 50%), (0%, 100%).

I boldly propose a different sort of answer: (80%, 100%). Most of my success comes from my own hard work, and all of my own hard work comes from external factors.

If all of our success comes from external factors, then it is reasonable to ask that we "pay it forward" by trying to improve the external factors of others, turning them into better people who will be better able to seize the opportunities to succeed. This is a good deal of the justification for the liberal program of redistribution of wealth and government aid to the poor.

5.2.2: This is all very philosophical. Can you give some concrete examples?

Lead poisoning, for example. It's relatively common among children in poorer areas (about 7% US prevalence) and was even more common before lead paint and leaded gasoline was banned (still >30% in many developing countries).

For every extra ten millionths of a gram per deciliter concentration of lead in their blood, children permanently lose five IQ points; there’s a difference of about ten IQ points among children who grew up in areas with no lead at all, and those who grew up in areas with the highest level of lead currently considered “safe”. Although no studies have been done on severely lead poisoned children from the era of leaded gasoline, they may have lost twenty or more IQ points from chronic lead exposure.

Further, lead also decreases behavioral inhibition, attention, and self-control. For every ten ug/dl lead increase, children were 50% more likely to have recognized behavioral problems. People exposed to higher levels of blood lead as a child were almost 50% more likely to be arrested for criminal behavior as adults (adjusting for confounders).

Economic success requires self-control, intelligence, and attention. It is cruel to blame people for not seizing opportunities to rise above their background when that background has damaged the very organ responsible for seizing opportunities. And this is why government action, despite a chorus of complaints from libertarians, banned lead from most products, a decision which is (controversially) credited with the most significant global drop in crime rates in decades, but which has certainly contributed to social mobility and opportunity for children who would otherwise be too lead-poisoned to succeed.

Lead is an interesting case because it has obvious neurological effects preventing success. The ability of psychologically and socially toxic environments to prevent success is harder to measure but no less real.

If a poor person can’t keep a job solely because she was lead-poisoned from birth until age 16, is it still fair to blame her for her failure? And is it still so unthinkable to take a little bit of money from everyone who was lucky enough to grow up in an area without lead poisoning, and use it to help her and detoxify her neighborhood?

5.3: What is the significance of whether success is personally or environmentally determined?

It provides justification for redistribution of wealth, and for engineering an environment in which more people are able to succeed.

6. Taxation

6.1: Isn’t taxation, the act of taking other people’s money by force, inherently evil?

See the Moral Issues section for a more complete discussion of this point.

6.2: Isn’t progressive taxation, the tendency to tax the rich at higher rates than the poor, unfair?

The most important justification for progressive tax rates is the idea of marginal utility.

This is easier to explain with movie tickets than money. Suppose different people are allotted a different number of non-transferable movie tickets for a year; some people get only one, other people get ten thousand.

A person with only two movie ticket might love to have one extra ticket. Perhaps she is a huge fan of X-Men, Batman, and Superman, and with only two movie ticket she will only be able to see two of the three movies she’s super-excited about this year.

A person with ten movie tickets would get less value from an extra ticket. She can already see the ten movies that year she’s most interested in. If she got an eleventh, she’d use it for a movie she might find a bit enjoyable, but it wouldn’t be one of her favorites.

A person with a hundred movie tickets would get minimal value from an extra ticket. Even if your tickets are free, you’re not likely to go to the movies a hundred times a year. And even if you did, you’d start scraping the bottom of the barrel in terms of watchable films.

A person with a thousand tickets would get practically no value from an extra ticket. At this point,t here’s no way she can go to any more movies. The extra ticket might not have literally zero value – she could burn it for warmth, or write memos on the back of it – but it’s pretty worthless.

So although all movie tickets provide an equal service – seeing one movie – one extra movie ticket represents a different amount of value to the person with two tickets and the person with a thousand tickets. Furthermore, 50% of their movie ticket holdings represent a different value to the person with two tickets and the person with a thousand movie tickets. The person with two tickets loses the ability to watch the second-best film of the year. The person with a thousand tickets still has five hundred tickets left, more than enough to see all the year’s best films, and at worst will have to buy some real memo paper.

Money works similarly to movie tickets. Your first hundred dollars determine whether you live or starve to death. Your next five hundred dollars determine whether you have a roof over your head or you’re freezing out on the street. But by your ten billionth dollar, all you're doing is buying a slightly larger yacht.

50% of what a person with $10,000 makes is more valuable to her than 50% of what a billionaire makes is to the billionaire.

Progressive taxation is an attempt to tax everyone equally, not by lump sum or by percentage, but by *burden*. Just as taking extra movie tickets away from the person with a thousand is more fair than taking some away from the person with only two, so we tax the rich at a higher rate because a proportionate amount of money has less marginal value to them.

6.2.1: But the progressive tax system is unfair and perverse. Imagine the tax rate on people making $100,000 or less is 30%, and the tax rate on people making more than $100,000 is 50%. You make $100,000, and end up with after tax income of $70,000. Then one day your boss tells you that you did a good job, and gives you a $1 bonus. Now you make $100,001, but end up with only $50,000.50 after tax income. How is that at all fair?

It's not, but this isn't how the tax system works.

What those figures mean is that your first $100,000, no matter how much you earn, is taxed at 30%. Then the money you make *after* that is taxed at 50%. So if you made $100,001, you would be taxed 30% on the first $100,000 (giving you $70,000), and 50% on the next $1 (giving you $.50), for an after-tax income of $70,000.50. The intuitive progression where someone who makes more money ends up with more after-tax income is preserved.

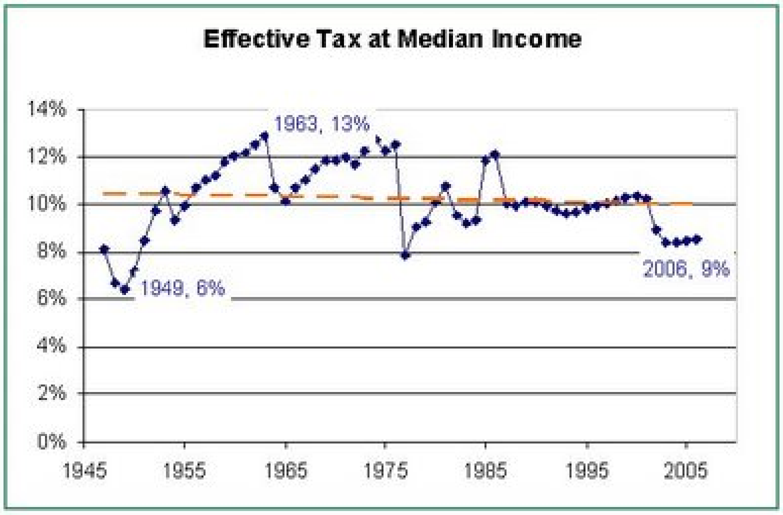
I know most libertarians don't make this mistake, and that there are much stronger arguments against progressive taxation, but this has come up enough times that I thought it was worth mentioning, with apologies to those readers whose time it has wasted.

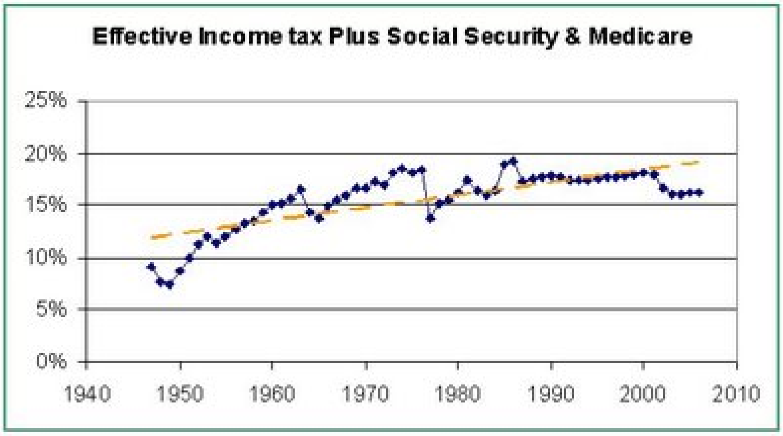
6.3: Taxes are too high.

Too high by what standard?

6.3.1: Too high by historical standards. Thanks to the unstoppable growth of big government, people have to pay more taxes now than ever before.

Actually, income tax rates for people on median income are around the lowest they've been in the past seventy-five years





6.3.1.1: I meant for the rich. It's only tolerable for people on median income because "progressive" governments are squeezing every last dollar out of successful people.

Actually, income tax rates for the rich are around the lowest they've been in the past seventy-five years.

6.3.1.1.1: But I heard that the share of tax revenue coming from the rich is at its highest level ever.

This is true. As the rich get richer and the poor get poorer (see 3.4), more of the money concentrates in the hands of the rich, and so more of the taxes come from the rich as well. This doesn't contradict the point that the tax rates on the rich are near historic lows.

6.3.1.2: I meant for corporations.

Actually, income tax rates for corporations are around the lowest they've been in the past seventy-five years.

6.3.2: I meant income taxes are too high compared to what's best for the economy, and even best for the Treasury. With taxes as high as they are, people will stop producing, rather than see so much of each

dollar they make go to the government. This will hurt the economy *and* lower tax revenue.

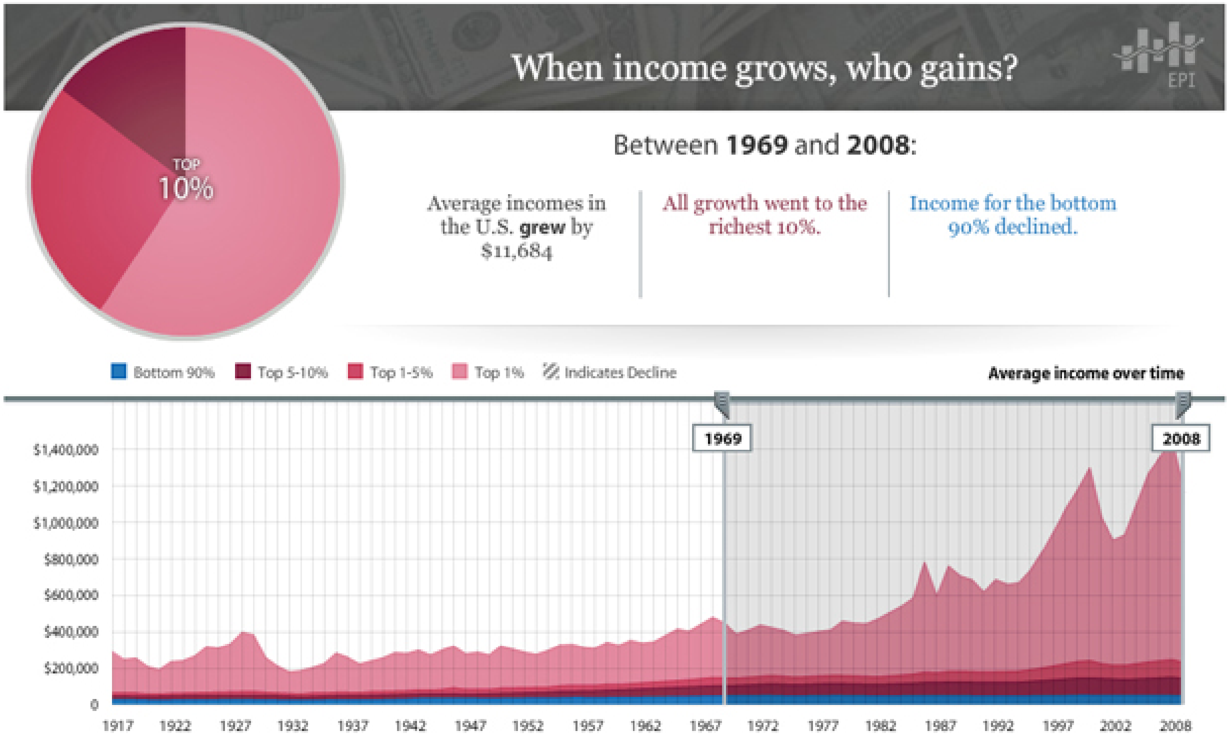
The Laffer curve certainly exists, but the consensus is that we're still well on the left half of it.

Although it's become a truism that high tax rates discourage production, studies have found this to be mostly false, with low elasticity of real income – see for example Gruber & Saez and Saez, Slemrod, and Giertz.

What studies *have* found is a high elasticity of taxable income. That is, raising taxes encourages people to find more tax loopholes, decreasing revenue. However, although this effect means a 10% higher tax rate would lead to less than 10% higher government income, the change in government income would still be positive – even by this stricter criterion, we're still on the left side of the Laffer curve. And of course, this effect could be eliminated by switching to a flat tax or closing tax loopholes.

6.4: Our current tax system is overzealous in its attempts to redistribute money from the rich to the poor. If instead we lowered taxes on the rich, this money would "trickle down" to the rest of the economy, driving growth. Instead of redistributing the pie, we'd make the pie larger for everyone.

If we're in an overzealous campaign for "equality" intended to lower the rich to the level of the poor, we're certainly not doing a very good job of it. Over the past thirty years, the rich have consistently gotten richer. None of this money has trickled down to the poor or middle-class, whose income has remained the same in real terms.



Source: The data come from this table: http://www.econ.berkeley.edu/~saez/TabFig2008.xls on Emmanuel Saez's website at University of California, Berkeley. Methodology

“Trickle-down” should be rejected as an interesting and plausible-sounding economic theory which empirical data have soundly disconfirmed.

6.5: Raising taxes would be useless for the important things like cutting the deficit. The deficit is $1.2 trillion. The most we could realistically raise from extra taxes on the rich would be maybe $200 billion. The most we could raise from insane levels of extra taxes on the rich *and* middle class would be about $500 billion – less than half the deficit. The real problem is spending.

Yes and no.

The deficit is, indeed, *very, very* large. It’s so large that no politically palatable option is likely to make more than a small dent in it. This is true of tax increases. It’s also true of spending cuts.

Cutting *all* redistributive government services for the poor including welfare, unemployment insurance, disability, food stamps, scholarships, you name it – would save about $200 billion. That’s less than 20% of the deficit. Cutting *all* health care, including Medicaid for senior citizens, would only eliminate $400 billion or so. Even eliminating the entire military down to the last Jeep would only get us $800 billion or so. The targets for cuts that have actually been raised are rounding errors: the Republicans trumpeted an end for government aid to NPR, but this is about $4 million – all of .000003% of the problem.

So “darnit, this one thing doesn’t completely solve the deficit” is not a good reason to reject a proposal. Solving the deficit will, if it’s possible at all, take a lot of different methods, including some unpalatable to liberals, some unpalatable to conservatives, and yes, some unpalatable to libertarians.

In particular, we need to avoid the “bee sting” fallacy, where we have so many problems that we just stop worrying. It would be irresponsible to say that since a few billion dollars doesn’t affect the deficit either way, we might as well just spend $5 billion on some random project we don’t need. For the same reason, it would be irresponsible to say we might as well just renew tax cuts on the rich that cost hundreds of billions of dollars each year.

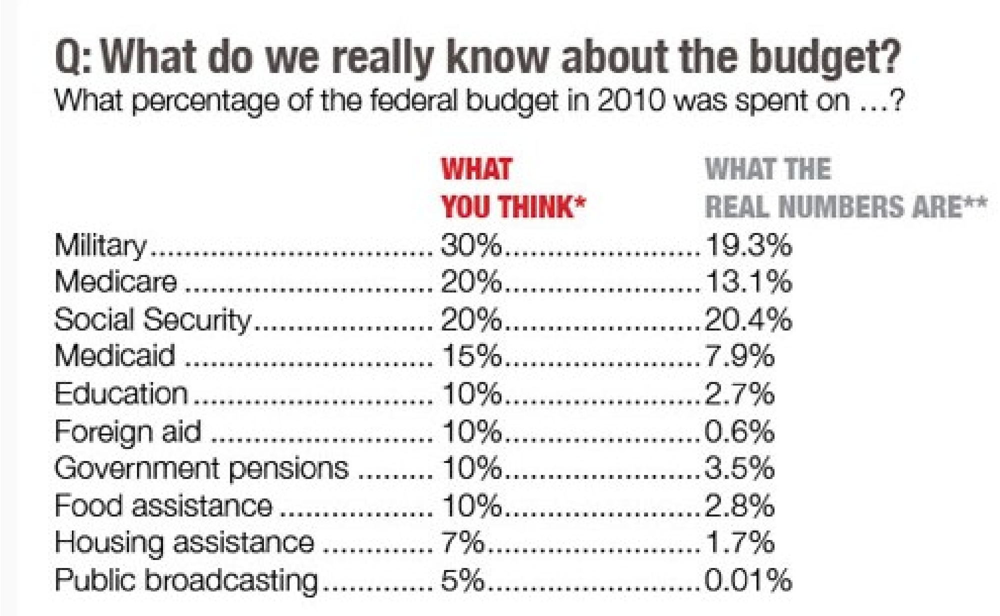
6.6: Taxes are basically a racket where they take my money and then give it to foreign governments and poor people.

According to a CNN poll, on average Americans estimate that about 10% of our taxes go to foreign aid. The real number is about 0.6%.

And although people believe that food and housing for the poor take up about 20% of the federal budget, the real number is actually less than 5%.

So although people worry that 30% of the budget goes to help the less fortunate, the real number is about 6%.

(And this is actually sort of depressing, when you think about it.)



Q: What do we really know about the budget?

What percentage of the federal budget in 2010 was spent on …?

	WHAT YOU THINK*	WHAT THE REAL NUMBERS ARE**
Military	30%	19.3%
Medicare	20%	13.1%
Social Security	20%	20.4%
Medicaid	15%	7.9%
Education	10%	2.7%
Foreign aid	10%	0.6%
Government pensions	10%	3.5%
Food assistance	10%	2.8%
Housing assistance	7%	1.7%
Public broadcasting	5%	0.01%

The majority of your taxes go to programs that benefit you and other middle-class Americans, such as Social Security and Medicare, and to programs that “benefit” you and other middle-class Americans, such as the military.

Part C: Political Issues

The Argument: *Government can’t do anything right. Its forays into every field are tinged in failure. Whether it’s trying to create contradictory “state owned businesses”, funding pet projects that end up over budget and useless, or creating burdensome and ridiculous “consumer protection”* *rules, its heavy-handed actions are always detrimental and usually embarrassing.*

With this track record, what sane person would want to involve government in even more industries? The push to get government deeper into health care is a disaster waiting to happen, and could give us a chronically broken system like those in Europe, where people die because of bureaucratic inefficiency.

Other places from which we can profitably eliminate government's prying hands include our schools, our prisons, our gun dealerships, and the friendly neighborhood meth lab.

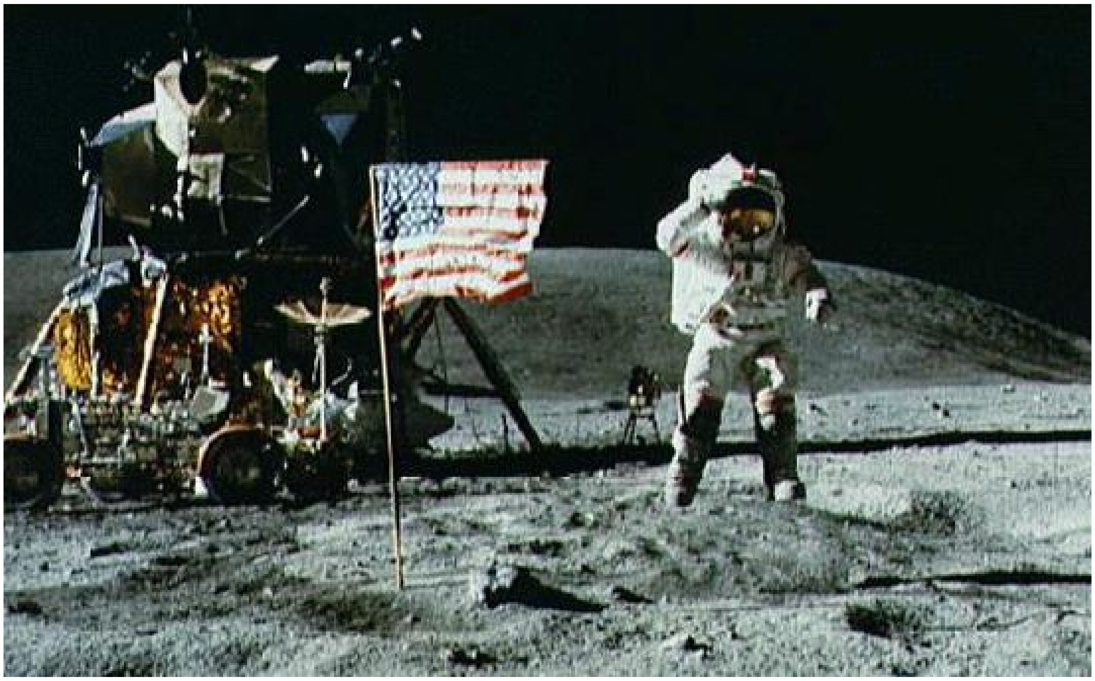
The Counterargument: *Government sometimes, though by no means always, does things right, and some of its institutions and programs are justifiably considered models of efficiency and human ingenuity. There are various reasons why people are less likely to notice these.*

Government-run health systems empirically produce better health outcomes for less money than privately-run health systems for reasons that include economies of scale. There are a mountain of statistics that prove this. Although not every proposal to introduce government into health will necessarily be successful, we would do well to consider emulating more successful systems.

We should think twice about exactly how much government we are willing to remove from our schools, gun dealerships, and meth labs, and run away screaming at the proposal to privatize prisons.

7. Competence of Government

7.1: Government never does anything right.



7.1.1: Okay, fine. But that's a special case where, given an infinite budget, they were able to accomplish something that private industry had no incentive to try. And to their credit, they *did* pull it off, but do you have any examples of government succeeding at anything more *practical*?

Eradicating smallpox and polio globally, and cholera and malaria from their endemic areas in the US. Inventing the computer, mouse, digital camera, and email. Building the information superhighway *and* the regular superhighway. Delivering clean, practically-free water and cheap on-the-grid electricity across an entire continent. Forcing integration and leading the struggle for civil rights. Setting up the Global Positioning System. Ensuring accurate disaster forecasts for hurricanes, volcanoes, and tidal waves. Zero life-savings-destroying bank runs in eighty years. Inventing nuclear power *and* the game theory necessary to avoid destroying the world with it.

7.1.1.1: All right… all right… but apart from better sanitation and medicine and education and irrigation

and public health and roads and a freshwater system and baths and public order… what has the government done for *us?*

Brought peace. But see also Government Success Stories and The Forgotten Achievements of Government

7.2: Large government projects are always late and over-budget.

The only study on the subject I could find, "What Causes Cost Overrun in Transport Infrastructure Projects?" (download study as .pdf) by Flyvbjerg, Holm, and Buhl, finds no difference in cost overruns between comparable government and private projects, and in fact find one of their two classes of government project (those not associated with a state-owned enterprise) to have a trend toward being *more* efficient than comparable private projects. They conclude that "…one conclusion is clear…the conventional wisdom, which holds that public ownership is problematic whereas private ownership is a main source of efficiency in curbing cost escalation, is dubious."

Further, when government cost overruns occur, they are not usually because of corrupt bureaucrats wasting the public's money. Rather, they're because politicians don't believe voters will approve their projects unless they spin them as being much cheaper and faster than the likely reality, leading a predictable and sometimes commendable execution to be condemned as "late and over budget" (download study as .pdf) While it is admittedly a problem that government provides an environment in which politicians have to lie to voters to get a project built, the facts provide little justification for a narrative in which government is incompetent at construction projects.

7.3: State-run companies are always uncreative, unprofitable, and unpleasant to use.

Some of the greatest and most successful companies in the world are or have been state-run. Japan National Railways, which created the legendarily efficient bullet trains, and the BBC, which provides the most respected news coverage in the world as well as a host of popular shows like *Doctor Who*, both began as state-run corporations (JNR was later privatized).

In cases where state-run corporations are unprofitable, this is often not due to some negative effect of being state-run, but because the corporation was put under state control precisely because it was something so unprofitable no private company would touch it, but still important enough that it had to be done. For example, the US Post Office has a legal mandate to ship affordable mail in a timely fashion to every single god-forsaken town in the United States; obviously it will be out-competed by a private company that can focus on the easiest and most profitable routes, but this does not speak against it. Amtrak exists despite passenger rail travel in the United States being fundamentally unprofitable, but within its limitations it has done a relatively good job: on-time rates better than that of commercial airlines, 80% customer satisfaction rate, and double-digit year-on-year passenger growth every year for the past decade.

7.3.1: State-run companies may be able to paper-push with the best of them, but the government can never be truly innovative. Only the free market can do that. Look at Silicon Valley!

Advances invented either solely or partly by government institutions include, as mentioned before, the computer, mouse, Internet, digital camera, and email. Not to mention radar, the jet engine, satellites, fiber optics, artificial limbs, and nuclear energy. And that doesn't the less recognizable inventions used mostly in industry, or the scores of other inventions from government-funded universities and hospitals.

Even those inventions that come from corporations often come not from startups exposed to the free market, but from *de facto* state-owned monopolies. For example, during its fifty years as a state-sanctioned monopoly, the infamous Ma Bell invented (via its Bell Labs division) transistors, modern cryptography, solar cells, the laser, the C programming language, and mobile phones; when the monopoly was broken up, Bell Labs was sold off to Alcatel-Lucent, which after a few years announced it was cutting all funding for basic research to focus on more immediately profitable applications.

Although the media celebrates private companies like Apple as centers of innovation, Apple's expertise lies, at best, in consumer packaging. They did not invent the computer, the mp3 player, or the mobile phone, but they developed versions of these products that were attractive and easy to use. This is great and they deserve the acclaim and heaps of money they've gathered from their success, but let's make sure to call a spade a spade: they are good at marketing and design, not at brilliant invention of totally new technologies.

That sort of *de novo* invention seems to come mostly from very large organizations that can afford basic research without an obsession on short-term profitability. Although sometimes large companies like Ma Bell, invention-rich IBM and Xerox can fulfill this role, such organizations are disproportionately governments and state-sponsored companies, explaining their impressive track record in this area.

7.4: Most government programs are expensive failures.

I think this may be a form of media bias – not in the sense that some sinister figure in the media is going through and censoring all the stories that support one side, but in the sense that “Government Program Goes More Or Less As Planned” doesn’t make headlines and so you never hear about it.

Let’s say the government wants to spent $1 million to give food to poor children. If there are bureaucratic squabbles over where the money’s supposed to come from, that’s a headline. If they buy the food at above-market prices, that’s a headline. If some corrupt official manages to give the contract to provide the food to a campaign donor along the way, that’s a *big* headline.

But what if none of these things happen, and poor children get a million dollars worth of food, and eat it, and it makes them healthier? I don’t know about you, but I’ve never seen a headline about this. “Remember that time last year when Congress voted to give food to poor children. Well, they got it.” What newspaper would ever publish something like that?

This is in addition to newspapers’ desire to outrage people, their desire to sound “edgy” by pointing out the failures of the status quo rather than sounding like they’re “pandering”, and honestly that they’re caught up in the same “government can never do anything right” narrative as everyone else.

Since every single time you ever hear about a government project it is always because that government project is going wrong, of course you feel like all government projects go wrong.

7.4.1: But a specific initiative to get money to the poor is one thing. What about a whole federal agency? We would know if it were failing, but we'd also be able to appreciate it when it succeeds, too.

Federal agencies that are successful sink into background noise, so that we don't think to thank them or celebrate them any more than we would celebrate that we have clean water (four billion people worldwide don't; thank the EPA and your local water board)

For example, the Federal Aviation Administration helps keep plane crashes at less than one per 21,000 years of flight time; you never think about this when you get on a plane. The National Crime Information Center collects and processes information about criminals from every police department in the country; you never think about this when you go out without being mugged. Zoning regulations, building codes, and the fire department all help prevent fires from starting and keep them limited when they do; you never think of this when you go the day without your house burning down.

One of government's major jobs is preventing things, and it's very hard to notice how many bad things *aren't* happening, until someone comes out with a report like e. coli poisoning has dropped by half in the past fifteen years. Even if you do hear the statistics, you may never think to connect them to the stricter food safety laws you wrote a letter to the editor opposing fifteen years ago.

7.4.2: You list cases where government regulation exists at the same time as a happy outcome, like the FAA and the lack of plane crashes, but that doesn't prove it was the regulation that *caused* the happy outcome.

No, it doesn't. For example, although workplace accidents have been cut in half since OSHA was founded, CATO wrote a very credible takedown in which they argue that was only a continuation of trends that have been going on since before OSHA existed.

Sometimes there are things we can do to identify cause. For example, as in the CATO study, we can compare trends before and after changes in government regulation; if there is a discontinuity, it may suggest the government was responsible. Second, we can compare trends in a country where a new regulation was introduced to trends in a country where it was not introduced; if the trend only changes in one country, that suggests an effect of the regulation. For example, after the FAA mandated "terrain awareness systems" in airplanes, the terrain-related accident rate sharply dropped to zero in the United States but was not affected in countries without similar rules.

But the important thing is that we apply our skepticism fairly and evenly: that we do not require mountains of evidence that a government regulation caused a positive result, while accepting that a regulation caused a negative result without a shred of proof.

It is very tempting for libertarians, when faced with anything going well even in a tightly regulated area, to say "Well, that just shows even this tight regulations can't hide how great private industry is!" and when anything goes wrong even in a very loosely regulated area, to say "Well, that just shows how awful regulation is, that even a little of it can screw things up!" But this is unfair, and ignores that we do have some ways to disentangle cause and effect.

And in any case, there is still the difference between "Government destroys everything it touches" and "Everything government touches is doing pretty well, but you can’t prove that it’s directly caused by government action.”

7.4.3: A lot of what government trumpets as “successful regulation” is just obvious stuff anyway that any individual in a free market would do of her own accord.

Very often, yesterday’s regulation is today’s obvious good idea that no one would dream of ignoring even if there were no regulation demanding it. But that neglects the role of government regulation in establishing social norms. Very often these are the regulations which those being regulated fought tooth and nail against at the time.

Many cars did not even *include* seatbelts until the government mandated that they do so. In 1983, the seat belt use rate in the United States was 14%. It was very clearly the government sponsored awareness campaigns and, later, mandatory seat belt laws that began being implemented around that era that raised seat belt rates; we know because we can watch the statistics state in different states as their legislation either led the campaign or lagged behind it.

After almost three decades of intense government pressure on automakers to allow and promote seatbelts, and on motorists to use them, seatbelt rates are now as high as 85%.

According to estimates, seatbelts save about 11,000 lives a year in the US. Different studies estimate between 80,000 and 100,000 lives saved in the last decade alone. For some perspective that’s the number of American deaths from 9/11 + the Vietnam War + both Iraq Wars + the Afghanistan War + Hurricane Katrina.

I completely acknowledge that if the government completely dropped all seatbelt regulations tomorrow, automakers would continue putting seatbelts in cars, and drivers would keep wearing them. That doesn’t mean government is useless, that means government, the only entity big enough to effect a nationwide change not just in behaviors but in social norms, did its job very very well.

8. Health Care

8.1: Government would do a *terrible* job in health care. We should avoid government-run “socialized” medicine unless we want cost overruns, long waiting times, and death panels.

Government-run health systems empirically do better than private health systems, while also costing much less money.

Let’s compare, for example, Sweden, France, Canada, the United Kingdom, and the United States. The first four all have single-payer health care (a version of government-run health system); the last has a mostly private health system (although it shouldn’t matter, we’ll use statistics from before Obamacare took effect). We’ll look at three representative statistics commonly used to measure quality of health care: infant mortality, life expectancy, and cancer death rate.

Infant mortality is the percent of babies who die in the first few weeks of life, usually a good measure of pediatric and neonatal care. Of the five countries, Sweden has the lowest infant mortality at 2.56 per 1,000 births, followed by France at 3.54, followed by the UK at 4.91, followed by Canada at 5.22, with the United States last at 6.81. (source)

Life expectancy, the average age a person born today can expect to live, is a good measurement of lifelong and geriatric care. Here Sweden is again first at 80.9, France and

Canada tied for second at 80.7, the UK next at 79.4, and the United States once again last at 78.3. (source)

Taking cancer deaths per 100,000 people per year as representative of deaths from serious disease, here we find the UK doing best at 253.5 deaths, Sweden second at 268.2, France in third at 286.1, and the United States again in last place at 321.9 deaths (source: OECD statistics; data for Canada not available).

So we notice that the United States does worse than all four countries with single-payer health systems, even though America is wealthier per capita than any of them. This is not statistical cherry-picking: any way you look at it, the United States has one of the least effective health systems in the developed world.

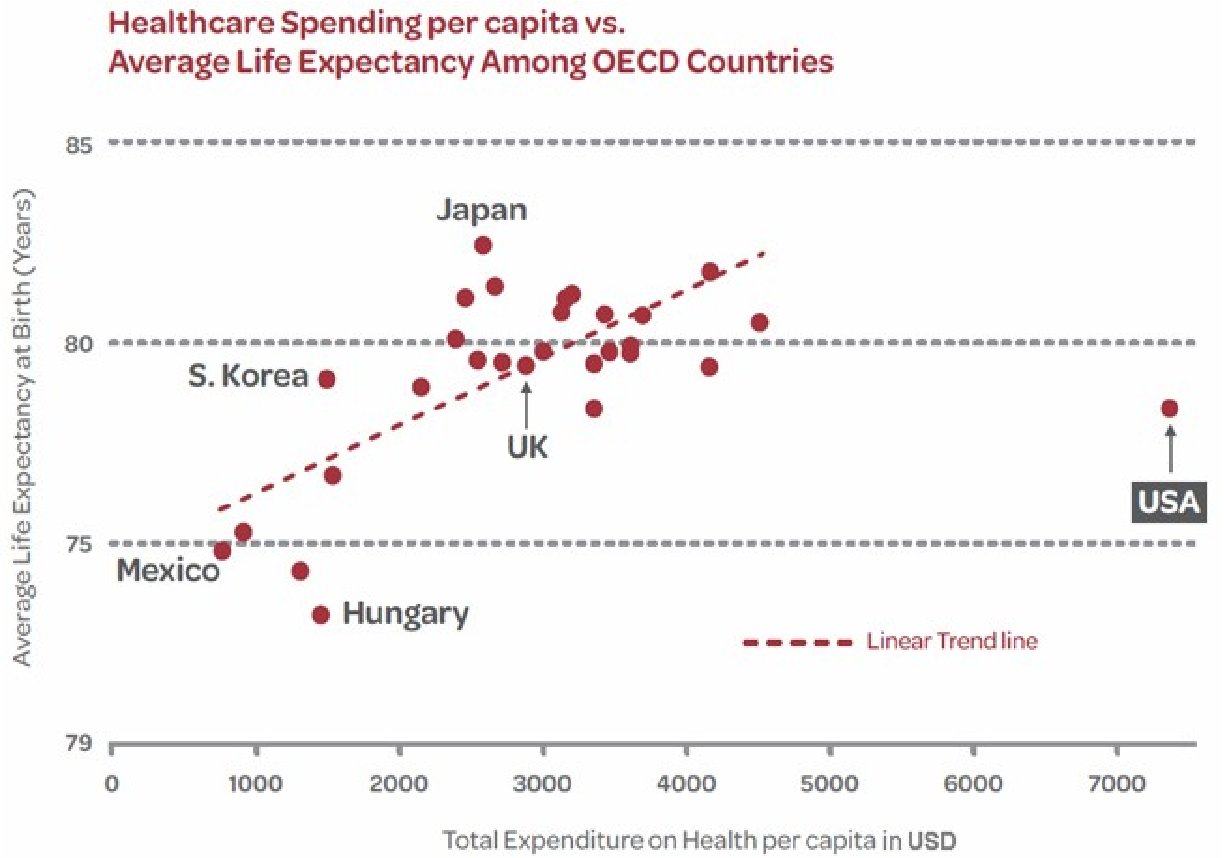
8.2: Government-run health care would be bloated, bureaucratic, and unnecessarily expensive, as opposed to the sleek, efficient service we get from the free market.

Actually, government-run health care is empirically more efficient than market health care. For example, Blue Cross New England employs more people to administer health insurance for its 2.5 million customers than the Canadian health system employs to administer health insurance for 27 million Canadians. Health care spending per person (public + private) in Canada is half what it is in America, yet Canadians have longer life expectancy, lower infant mortality, and are healthier by every objective standard.

Remember those five countries from the last question?
The UK spends $1,675 per person per year on health care. Canada spends $1,939. Sweden, which you'll remember did best on most of the statistics, spends $2,125. France spends $2,288. Americans spend on average $4,271 – almost *three*

times as much as Britain, a country which *delivers better health care*.

When this argument gets put in graph form, it becomes even clearer that US health inefficiency is literally off the chart.



If these were companies in the free market, the company that charges three times as much to provide a worse service would have gone bankrupt long ago. That company is American-style private health care.

8.3: In government-run health care, people are relegated to “waiting lists”, where they have to wait months or even years for doctor visits, surgeries, and other procedures. Sometimes people die on these waiting lists. Obviously, this is unacceptable and a knock-down argument against government-run health care.

The laws of supply and demand apply in health care as much as anywhere else: people would like to see doctors as quickly as possible, but doctors are a scarce resource that must be allocated somehow.

In a private system, doctor access is allocated based on money; this has the advantage of incentivizing the production of more doctors and of ensuring that people with enough money can see doctors quickly. These are also its disadvantages: assuming more people want to see a doctor than need to do so, costs will spiral out of control and poor people will have limited or no access.

In a public system, doctor access is allocated based on medical need. Although no one will be turned away from a doctor in an emergency situation, people may have to wait a long amount of time for elective surgeries in order that other sicker people, including poor people who would not be seen at all in a private system, can be seen first.

The relative effectiveness of the two systems can once again be seen in the infant mortality, life expectancy, and cancer survival rate statistics.

8.4: Government-run health care inevitably includes “death panels” who kill off expensive patients in order to save money on health care costs.

The private system as it exists now in America also has bodies that make these kinds of rationing decisions. Health care rationing is not some sinister conspiracy but a reasonable response to limited resources. The complete argument is here, but I can sum up the basics:

Insurance providers, whether they are a government agency or a private corporation, have a finite amount of money; they can only spend money they have. In one insurance company, customers might pay hundred million dollars in fees each year, so the total amount of money the insurance company can spend on all its customers that year is a hundred million dollars. In reality, since it is a business, it wants to make a profit. Let’s say it wants a profit of ten percent. That means the total amount of money it has to spend is ninety million dollars.

But as a simplified example, let’s reduce this to an insurance company with one hundred customers, each of whom pays $1. This insurance company wants 10% profit, so it has $90 to spend (instead of our real company’s $90 million). Seven people on the company’s plan are sick, with seven different diseases, each of which is fatal. Each disease has a cure. The cures cost, in order, $90, $50, $40, $20, $15, $10, and $5.

We are far too nice to ration health care with death panels; therefore, we have decided to give everyone every possible treatment. So when the first person, the one with the $90 disease, comes to us, we gladly spend $90 on their treatment; it would be inhuman to just turn them away. Now we have no money left for anyone else. Six out of seven people die.

The fault here isn’t with the insurance company wanting to make a profit. Even if the insurance company gave up its ten percent profit, it would only have $10 more; enough to save the person with the $10 disease, but five out of seven would still die.

A better tactic would be to turn down the person with the $90 disease. Instead, treat the people with $5, $10, $15, $20, and $40 diseases. You still use only $90, but only two out of seven die. By refusing treatment to the $90 case, you save four lives. This solution can be described as more cost-effective; by spending the same amount of money, you save more people. Even though “cost-effectiveness” is derided in the media as being opposed to the goal of saving lives, it’s actually all about saving lives.

If you don’t know how many people will get sick next year with what diseases, but you assume it will be pretty close to the amount of people who get sick this year, you might make a rule for next year: Treat everyone with diseases that cost $40 or less, but refuse treatment to anyone with diseases that cost $50 or more.

This rule remains true in the case of the $90 million insurance company. In their case, no one patient can use up all the money, but they still run the risk of spending money in a way that is not cost-effective, causing many people to die. Like the small insurance company, they can increase cost-effectiveness by creating a rule that they won’t treat people with diseases that cost more than a certain amount.

So, as one commentator pointed out, “death panels” should be called “life panels”: they aim to maximize the total number of lives that can be saved with a certain limited amount of resources.

8.5: Why is government-run health care so much more effective?.

A lot of it is economies of scale: if the government is ensuring the entire population of a country, it can get much better deals than a couple of small insurance companies. But a lot of it is more complicated, and involves people’s status as irrational consumers of health products. A person sick with cancer doesn’t want to hear a cost-benefit analysis suggesting that the latest cancer treatment is probably not effective. He wants that treatment right now, and the most successful insurance companies and hospitals are the ones that will give it to him. Here’s a good article explaining some of the systematic flaws in the economics of health care under the American system.

It could also be that really good health care and the profit motive don’t mix: studies show that for-profit hospitals are more expensive, and have poorer care (as measured in death rates) than not-for-profit hospitals.

9. Prison Privatization

9.1: Privatized, for-profit prisons would be a great way to save money.

No one likes criminals very much. Even so, most of us agree that even criminals deserve humane conditions. We reject cruel and unusual punishment, and try to keep prisoners relatively warm, clean, and well-fed. This is not only a moral issue, but a practical one: we don’t want prisoners to go insane or suffer breakdowns, because we want them to be able to re-adjust into normal society after they are released.

For-profit prisons have all of the flaws of for-profit companies with none of the advantages. Normal companies want to cut costs wherever possible, but this is balanced by customer satisfaction: if they treat their customers poorly or create a low-quality product, they won’t make money. In prisons, the ability to get new “customers” comes completely uncoupled from the quality of the product they provide. If the government pays them a certain fixed amount per prisoner, the prison’s only way to increase profits is by treating prisoners as shabbily as possible without killing them. Indeed, statistics show that prisoners in private prisons have worse medical care, terrible living conditions, and rates of in-prison violence 150% greater than those in public prisons. Private prisons refuse to collect data on recidivism rates, but a moment’s thought reveals that they have an economic incentive to keep them as *high* as possible.

But the real dangers lie in the corruptibility of the political process, something with which libertarians are already familiar. Private prisons have been active in lobbying for stricter sentencing guidelines like the Three Strikes Law, which encourages governments to imprison criminals for life. In a country that already imprisons more of its population than any other country in the world, it is extremely dangerous to create a powerful political force whose self-interest lies in imprisoning as many people as possible.

But the most striking example of the danger of private prisons is the case of two judges who received bribes from private prisons to jail innocent people.

If this is the alternative, I'm willing to bite the bullet and accept the overpaid prison guards with annoying unions who dominate the public prisons.

9.2: What? Libertarians don't actually believe in private prisons!

Fair enough; I got this complaint a few times on the first version and I acknowledge it's not an integral component of libertarian philosophy. I included it because it seems to stem from the same "government can never do anything right and we should privatize everything" idea that drives a lot of libertarian thinking, and because I really, really don't like private prisons.

10. Gun Control

10.1: Gun control laws only help criminals, who are not known for following laws in any case, make sure that their victims are unarmed and unable to resist; as such, they increase crime.

The statistics supporting this view seem relatively solid and I agree that attempts to ban or restrict access to guns are a bad idea.

On the other hand, many of the issues surrounding gun control are much less restrictive. For example, some involve restrictions on sales to criminals, “cooldown periods” before purchase, mandatory safety training, et cetera.

Although I haven’t seen any evidence either way on whether these laws are beneficial, they should be evaluated on their own merits rather than as part of a narrative in which all gun laws must be opposed because gun control is bad.

11. Education

11.1: Government sponsored public education is a horrible failure.

Compared to what?

Compared to the period when there *wasn’t* government-sponsored public education…well, that’s hard to say because of poor statistic-keeping at that time, and how one counts minorities and women, who usually weren’t educated at all back then. The most official statistics (eg NOT the ones you find without citation on libertarian blogs that say literacy was 100% way back when and became abysmal as soon as public schooling started) say that white illiteracy declined from about 11.5% in the mid-1800s to about 0.5% in 1980, and black illiteracy from about 80% to 1.5% over the same period.

Compared to other countries, the US does relatively poorly considering its wealth, but all the other countries that do better than the US *also* have government-sponsored public education, sometimes to a much greater degree than we do.

Compared to private schools, public schools actually do better once confounders like race, class, and income have been adjusted out of the analysis.

(Yes, without such adjustment private schools do better – but considering that private schools cater towards wealthy students – who usually do better in school – and often have selective admission policies in which they only take students who are already pretty smart – whereas public schools have to take everyone including dumb kids, kids with learning disabilities, and kids from broken families in ghettos – such unadjusted data is meaningless. It's the equivalent of noting that the doctor who specializes in acne has fewer patients die than the doctor who specializes in cancer: it's not that she's a better doctor, just that she only takes cases who are pretty healthy already.)

Our educational system certainly has immense room for improvement. But the country that consistently tops world education rankings, Finland, has zero private schools (even all the universities are public) and no "school choice". What it does have is extremely well-credentialed, highly paid teachers (and, unfortunately, an ethnically homogenous population without any dire poverty or broken families, which probably counts for a heck of a lot more than anything else). So whatever America's specific failures or successes, the mere existence of public education is not a credible scapegoat.

11.2: Why not dismantle the public education system and have a voucher system that offers parents free choice over where to send their kids?

I think this idea has merit, and that we should at least experiment with it and see if it works. That having been said, I do see one huge caveat.

Libertarians tend not to believe in *equality of results* – they think it's okay if more skilled people are more successful – but one of the qualities I most admire about them is that they usually do believe in equality of opportunity: that everyone gets an equal chance at life. I mentioned before how inheriting money from your parents can complicate that, but it would be ethically complicated to try and "solve" that problem, so it might be the sort of thing we just have to live with.

But imagine if your parents chose where to send you for school. Even if we somehow eliminated the cost issue by making everyone accept a school voucher of equal value, clever parents would compare the pros and cons of various schools and send their child to the best one. Not-so-clever parents would get fooled by TV commercials with sexy celebrities and send their kids to terrible schools. Super religious parents would send their kids to schools that taught only religious education and shunned math and science and history as the evil trappings of the secular world. Muslim parents would send their kids to madrassas. Immigrant parents might send their kids to Spanish-only schools so that they didn't drift too far away from their families. Parents with strong political beliefs could send their kids to schools that did their best to brainwash their kids into having the same beliefs as them.

And there *would* be kids who succeeded in spite of all this, who made it through twelve years of constant brainwashing and ignorance, and somehow managed to become intelligent adults who could learn all the education they missed during their free time. But statistically, there wouldn't be very many of them, any more than there were a bunch of Christians in Saudi Arabia in the example a few pages back.

Right now, parents can screw up lots of facets of a kid’s life, but they can only do so much to screw up their education. And I have this vague hope that maybe a kid with horrible parents, if she was exposed to decent people and a free exchange of ideas in school might be able to use that brief period of respite to gain a foothold on sanity.

So what I’m saying is, if there were school choice, if we wanted to protect equality of opportunity and children’s’ rights, we’d probably have to regulate the heck out of them, which to some degree would defeat the point.

11.3: I don’t believe the government should be in the business of “protecting” children from their parents.

You should. It’s a pretty important business, even if you subscribe to libertarian assumptions. Even libertarians tend to agree that the government should generally be protecting people from slavery and from the use of force.

Children are basically slaves to their parents for the first ten to fifteen years of their lives, and parents have a special social permission to use force against their children.

In the best possible case, this is an incredibly silly metaphor and one no one would ever even think about. In the worst possible case, it’s completely and literally true.

I have met people with horrible parents. The first eighteen years (or less, if they were able to get themselves legally emancipated early) of their lives were a living hell. These are people who literally have control of every single thing you do, from whether you can eat dinner to who you are allowed to make friends with to what church you go to to what opinions you can express to whether you’re allowed to sleep at night. They are people who can torture and beat you to within an inch of your life, and maybe a social worker will take you away for a few months, and then that social worker will probably return you right back to them. And if it’s just *emotional* torture, you can forget about even getting the social worker.

And obviously the parent-child relationship is a healthy one in 99% of cases, and child-rearing has been around since deep prehistoric time, and we would be idiots to mess with it, and no one wants a dystopia where the government takes kids from their parents and raises them in a commune or whatever.

But unless you think rights and morality only start existing on someone’s eighteenth birthday, if if there were *one* form of government intervention that even libertarians should be able to get behind, it would be protecting children from their parents, in the rare few cases where this is necessary.

Part D: Moral Issues

The Argument: *Moral actions are those which do not initiate force and which respect people’s natural rights. Government is entirely on force, making it fundamentally immoral. Taxation is essentially theft, and dictating the conditions under which people may work (or not work) via regulation is essentially slavery. Many government programs violate people’s rights, especially their right to property, and so should be opposed as fundamentally immoral regardless of whether or not they “work”.*

The Counterargument: *Moral systems based only on avoiding force and respecting rights are incomplete, inelegant, counterintuitive, and usually riddled with logical fallacies. A more sophisticated moral system, consequentialism, generates the principles of natural rights and non-initiation of violence as heuristics that can be used to solve coordination problems, but also details under what* *situations such heuristics no longer apply. Many cases of government intervention are such situations, and so may be moral.*

12. Moral Systems

12.1: Freedom is incredibly important to human happiness, a precondition for human virtue, and a value almost everyone holds dear. People who have it die to protect it, and people who don't have it cross oceans or lead revolutions in order to gain it. But government policies all infringe upon freedom. How can you possibly support this?

Freedom is one good among many, albeit an especially important one.

In addition to freedom, we value things like happiness, health, prosperity, friends, family, love, knowledge, art, and justice. Sometimes we have to trade off one of these goods against another. For example, a witness who has seen her brother commit a crime may have to decide between family and justice when deciding whether to testify. A student who likes both music and biology may have to decide between art and knowledge when choosing a career. A food-lover who becomes overweight may have to decide between happiness and health when deciding whether to start a diet.

People sometimes act as if there is some hierarchy to these goods, such that Good A always trumps Good B. But in practice people don't act this way. For example, someone might say "Friendship is worth more than any amount of money to me." But she might continue working a job to gain money, instead of quitting in order to spend more time with her friends. And if you offered her $10 million to miss a friend's birthday party, it's a rare person indeed who would say no.

In reality, people value these goods the same way they value every good in a market economy: in comparison with other goods. If you get the option to spend more time with your friends at the cost of some amount of money, you'll either take it or leave it. We can then work backward from your choice to determine how much you *really* value friendship relative to money. Just as we can learn how much you value steel by learning how many tons of steel we can trade for how many barrels of oil, how many heads of cabbages, or (most commonly) how many dollars, so we can learn how much you value friendship by seeing when you prefer it to opportunities to make money, or see great works of art, or stay healthy, or become famous.

Freedom is a good much like these other goods. Because it is so important to human happiness and virtue, we can expect people to value it very highly.

But they do not value it infinitely highly. Anyone who valued freedom from government regulation infinitely highly would move to whichever state has the most lax regulations (Montana? New Hampshire?), or go live on a platform in the middle of the ocean where there is no government, or donate literally all their money to libertarian charities or candidates on the tiny chance that it would effect a change.

Most people do not do so, and we understand why. People do not move to Montana because they value aspects of their life in non-Montana places – like their friends and families and nice high paying jobs and not getting eaten by bears – more than they value the small amount of extra freedom they could gain in Montana. Most people do not live on a platform in the middle of the ocean because they value aspects of living on land – like being around other people and being safe – more than they value the rather large amount of extra freedom the platform would give them. And most people do not donate literally all their money to libertarian charities because they like having money for other things.

So we value freedom a finite amount. There are trade-offs of a certain amount of freedom for a certain amount of other goods that we already accept. It may be that there are other such trade-offs we would also accept, if we were offered them.

For example, suppose the government is considering a regulation to ban dumping mercury into the local river. This is a trade-off: I lose a certain amount of freedom in exchange for a certain amount of health. In particular, I lose the freedom to dump mercury into the river in exchange for the health benefits of not drinking poisoned water.

But I don’t really care that much about the freedom to dump mercury into the river, and I care a lot about the health benefits of not drinking poisoned water. So this seems like a pretty good trade-off.

And this generalizes to an answer to the original question. I completely agree freedom is an extremely important good, maybe the most important. I don’t agree it’s an infinitely important good, so I’m willing to consider trade-offs that sacrifice a small amount of freedom for a large amount of something else I consider valuable. Even the simplest laws, like laws against stealing, are of this nature (I trade my “freedom” to steal, which I don’t care much about, in exchange for all the advantages of an economic system based on private property).

The arguments above are all attempts to show that some of the trade-offs proposed in modern politics are worthwhile: they give us enough other goods to justify losing a relatively insignificant “freedom” like the freedom to dump mercury into the river.

12.1.1: But didn’t Benjamin Franklin say that those who would trade freedom for security deserve neither?

No, he said that those who would trade *essential* liberty for *temporary* security deserved neither. Dumping mercury into the river hardly seems like essential liberty. And when Franklin was at the Constitutional Convention he agreed to replace the minimal government of the Articles of Confederation with a much stronger centralized government just like everyone else.

12.2: Taxation is theft. And when the government forces you to work under their rules, for the amount of money they say you can earn, that’s slavery. Surely you’re not in favor of theft and slavery.

Consider the argument “How can we have a holiday celebrating Martin Luther King? After all, he was a *criminal*!”

Technically, Martin Luther King *was* a criminal, in that he broke some laws against public protests that the racist South had quickly enacted to get rid of him. It’s why he famously spent time in Birmingham Jail.

And although “criminal” is a very negative-sounding and emotionally charged word, in this case we have to step back from our immediate emotional reaction and notice that the ways in which Martin Luther King was a criminal don’t make him a worse person.

A philosopher might say we’re equivocating between two meanings of “criminal”, one meaning of “person who breaks the law”, and another meaning of “horrible evil person.” Just because King satisfies the first meaning (he broke the law) doesn’t mean he has to satisfy the second (be horrible and evil).

Or consider the similar argument: “Ayn Rand fled the totalitarian Soviet Union to look for freedom in America. That makes her a traitor!” Should we go around shouting at Objectivists “How can you admire Ayn Rand when she was a dirty rotten *traitor*“?

No. Once again, although “traitor” normally has an automatic negative connotation, we should avoid instantly judging things by the words we can apply to them, and start looking at whether the negative feelings are deserved.

Or once again the philosopher would say we should avoid equivocating between “traitor” meaning “someone who switches sides from one country to an opposing country” and “horrible evil untrustworthy person.”

Our language contains a lot of words like these which package a description with a moral judgment. For example, “murderer” (think of pacifists screaming it at soldiers, who do fit the technical definition “someone who kills someone else”), “greedy” (all corporations are “greedy” if you mean they would very much like to have more money, but politicians talking about “greedy corporations” manage to transform it into something else entirely) and of course that old stand-by “infidel”, which sounds like sufficient reason to hate a member of another religion, when in fact it simply means a member of another religion. It’s a stupid, cheap trick unworthy of anyone interested in serious rational discussion.

And calling taxation “theft” is exactly the same sort of trick. What’s theft? It’s taking something without permission. So it’s true that taxation is theft, but if you just mean it involves taking without permission, then everyone from Lew Rockwell up to the head of the IRS already accepts that as a given.

This only sounds like an argument because the person who uses it is hoping people will let their automatic negative reaction to theft override their emotions, hoping they will equivocate from theft as "taking without permission" to "theft as a terrible act worthy only of criminals".

Real arguments aren't about what words you can apply to things and how nasty they sound, real arguments about what good or bad consequences those things produce.

12.3: Government actions tend to involve the initiation of force against innocent people. Isn't that morally wrong?

Why should it be morally wrong?

12.3.1: Because the initiation of force always has bad consequences, like ruining the economy or making people unhappy.

Sometimes it does. Other times it has good consequences.

Take cases like the fish farming, boycott, and charity scenarios above. There the use of force to solve the coordination problem meets an extraordinarily strict set of criteria: not only does it benefit the group as a whole, not only does it benefit every single individual in the group, but every single individual in the group knows that it benefits them and endorses that benefit (eg would vote for it).

In other cases, such as the retirement savings example above, the use of force meets only a less strict set of criteria: it benefits the group as a whole, it benefits every single individual in the group, but not every individual in the group necessarily knows that it benefits them or endorses that benefit. These are the cases libertarians might call "paternalism".

Still more cases satisfy an even looser criterion. They benefit the group as a whole, but they might not benefit every single individual in the group, and might harm some of them. These are the cases that libertarians might call “robbing Peter to pay Paul”.

All three of these sets of cases belie the idea that the use of force must on net have bad consequences.

12.3.2: Okay, maybe it’s wrong because some moral theory that’s not about consequences tells me it’s wrong.

If your moral theory doesn’t involve any consequences, why follow it? It seems sort of like an arbitrary collection of rules you like.

The Jews believe that God has commanded them not to murder. They also believe God has commanded them not to start fires on Saturdays. Jews who lose their belief in God usually continue not to murder, but stop worrying about whether or not they light fires on Saturdays. Likewise, evangelical Christians believe stealing is a sin, and that homosexuality is also a sin. If they de-convert and become atheists, most of them will still oppose stealing, but most will stop worrying about homosexuality. Why?

Killing and stealing both have bad consequences; in fact, that seems to be the essence of why they’re wrong. Fires on Saturday and homosexuality don’t hurt anybody else, but killing and stealing do.

Why are consequences to other people seems such a specially relevant category? The argument is actually itself pretty libertarian. I can do whatever I want with my own life, which includes following religious or personal taboos. Other people can do whatever they want with their own lives too.

The stuff that matters – the stuff where we have to draw a line in the sand and say “Nope, this is moral and this is immoral, doesn’t matter what you think” is because it has some consequence in the real world like hurting other people.

12.3.2.1: I was always taught that the essence of morality was the Principle of Non-Aggression: no one should ever initiate force, except in self-defense. What exactly is wrong with this theory?

At least two things. First, once you disentangle it from the respect it gets as the Traditional Culturally Approved Ground Of Morality, the actual rational arguments for it as a principle are surprisingly weak. Second, in order to do anything practical with it you need such a mass of exceptions and counter-exceptions and stretches that one starts to wonder whether it’s doing any philosophical work at all; it becomes a convenient hook upon which to hang our pre-existing prejudices rather than a useful principle for solving novel moral dilemmas.

12.3.2.1.1: What do you mean by saying that the rational arguments for the Principle of Non-Aggression are weak?

There are dozens of slightly different versions of these arguments, and I don’t want to get into all of them here, so I’ll concentrate on the most common.

Some people try to derive the Principle of Non-Aggression from self-ownership. But this is circular reasoning: the form of “private property” you need to own anything, including your self/body, is a very complicated concept and one that requires some form of morality in order to justify; you can’t use your idea of private property as a justification for morality. Although it’s obvious that in some sense you *are* your body, there’s no way to go from here to “And therefore the proper philosophical relationship between you and your body is the concept of property exactly as it existed in the 17th century British legal system.”

This also falls afoul of the famous is-ought dichotomy, the insight that just because something *is* true doesn’t mean it *should be* true. Just because we notice some factual relationship between yourself and your body doesn’t mean that relationship between yourself and your body is good or important or needs to be protected in laws. We might eventually *decide* it should be (and hopefully we will!) but we need to have other values in order to come to that decision; we can’t use the decision as a *basis* for our values.

The self-ownership argument then goes from this questionable assumption to other even more questionable ones. If you use your body to pick fruit, that fruit becomes yours, even though you didn’t make it. If you use your body to land on Tristan de Cunha and plant a flag there and maybe pick some coconuts, that makes Tristan de Cunha and everything on your property and that of your heirs forever, even though you *definitely* didn’t make the island. And if someone else lands on Tristan de Cunha the day after you, you by right control every facet of their life on the island and they have to do whatever you say or else leave. There are good arguments for why some of these things make economic sense, but they’re all practical arguments, not moral ones positing a necessary relationship.

Oddly enough, although apparently your having a body does license you to declare yourself Duke of Tristan de Cunha, it doesn’t license you to use your fist to punch your enemy in the gut, or use your legs to walk across a forest someone else has said they claim, even though your ability to move your hand rapidly in the direction of your enemy’s abdomen, or your feet along a forest path, seems like a much more fundamental application of your body than taking over an island.

All of these rules about claiming islands and not punching people you don’t like and so on are potentially good rules, but trying to derive them just from the fact that you have a body starts to seem a bit hokey.

12.3.2.1.2: What do you mean by saying that the Non-Aggression Principle requires so many exceptions and counter-exceptions that it becomes useless except as a hook upon which to hang prejudices we from other sources?

First, the principle only even slightly makes sense by defining “force” in a weird way. The NAP’s definition of “force” includes walking into your neighbor’s unlocked garden when your neighbor isn’t home and picking one of her apples. It includes signing a contract promising to deliver a barrel of potatoes, but then not delivering the potatoes when the time comes. Once again, I agree these are bad things that we need rules against. But it takes quite an imagination to classify them under “force”, or as deriving from the fact that you have a body. This is a good start to explaining what I mean when I say that people *claim* that they’re using the very simple-sounding “no initiation of force” principle but are actually following a more complicated and less justified “no things that seem bad to me even though I can’t explain why”.

Second, even most libertarians agree it can be moral to initiate force in certain settings. For example, if the country is under threat from a foreign invader or from internal criminals, most libertarians agree that it is moral to levy a small amount of taxation to support an army or police force that restores order. Again, this is a very good idea – but also a blatant violation of the Non-Aggression Principle. When libertarians accept the initiation of force to levy taxes for the police, but protest that initiating force is always wrong when someone tries to levy taxes for welfare programs, it reinforces my worry that the Non-Aggression Principle is something people *claim* to follow while actually following their own "no things that seem bad to me even though I can't explain why, but things that seem good to me are okay" principle.

(I acknowledge that some libertarians take a stand against taxes for the military and the police. I admire their consistency even while I think their proposed policies would be a disaster.)

Third, when push comes to shove the Non-Aggression Principle just isn't strong enough to solve hard problems. It usually results in a bunch of people claiming conflicting rights and judges just having to go with whatever seems intuitively best to them.

For example, a person has the right to live where he or she wants, because he or she has "a right to personal self-determination". Unless that person is a child, in which case the child has to live where his or her parents say, because…um…the parents have "a right to their child" that trumps the child's "right to personal self-determination". But what if the parents are evil and abusive and lock the child in a fetid closet with no food for two weeks? Then maybe the authorities can take the child away because…um…the child's "right to decent conditions" trumps the parents' "right to their child" even though the latter trumps the child's "right to personal self-determination"? Or maybe they can't, because there shouldn't even be authorities of that sort? Hard to tell.

Another example. I can build an ugly shed on my property, because I have a “right to control my property”, even though the sight of the shed leaves my property and irritates my neighbor; my neighbor has no “right not to be irritated”. Maybe I can build a ten million decibel noise-making machine on my property, but maybe not, because the noise will leave my property and disturbs neighbor; my “right to control my property” might or might not trump my neighbor’s “right not to be disturbed”, *even though disturbed and irritated are synonyms*. I definitely can’t detonate a nuclear warhead on my property, because the blast wave will leave my property and incinerates my neighbor, and my neighbor apparently does have a “right not to be incinerated”.

If you’ve ever seen people working within our current moral system trying to solve issues like these, you quickly realize that not only are they making it up as they go along based on a series of ad hoc rules, but they’re so used to doing so that they no longer realize that this is undesirable or a shoddy way to handle ethics.

12.4: Is there a better option than the Non-Aggression Principle?

Yes. It’s consequentialism, the principle that it is moral to do whatever has, on net, the best consequences. This is about equivalent to saying “to do whatever makes the world a better place”. It’s the principle we’ve been using implicitly throughout this FAQ and the principle most people use implicitly throughout their lives.

It’s also the principle that drives capitalism, where people are able to create incredible businesses and innovations because they are trying to do whatever has the best financial consequences for themselves. Consequentialism just takes that insight and says that instead of just doing it with money, let’s do it with everything we value.

12.4.1: Best consequences according to whom?

Well, if you’re the one making the moral decision, then best consequences according to you. All it’s saying is that your morality should be a reflection of your value system and your belief in a better world. Your job as a moral agent is to try to make the world a better place by whatever your definition of “better place” might be.

Sticking to the capitalism analogy, consumerism “tells you” (not that you need to be told) to get whatever goods you value most. Consequentialism does the same, but tells you to try to get the collection of abstract moral goods you value the most.

But remember our discussion of trade-offs above. Most people value many different moral goods, and you are no exception. If you’re trying to make the world a better place, you should be thinking about your relative valuation of all these goods and what trade-offs you are willing to make.

12.4.2: Best consequences for me, or best consequences for everyone?

Again, this is your decision. If you’re completely selfish, then consequentialism tells you to seek out the best consequences for yourself. This probably wouldn’t mean being a libertarian – thankless activism for an unpopular political position is really a *terrible* way to go about looking out for Number One. It would probably mean cheating off the government – either in the form of welfare abuse if you’re poor and lazy, or in the form of crony capitalism if you’re rich and ambitious. As icing on the cake, make sure to become a sanctimonious and hypocritical liberal, as it's a great way to become popular and get invited to all the fancy parties.

But if you care about people other than yourself, consequentialism tells you to seek out the best consequences for the people you care about (which could be anything from your family to your country to the world). This could involve political activism, and it could even involve political activism in favor of libertarianism if you think it's the best system of government.

Alternately, it could justify trying to *start* a government, if there's no government yet and you think a world with government would be better for the people you care about than one without it.

Most of the rest of this section will be assuming you do in fact care for other people at least a little.

12.4.3: Since many people probably want different things and care about different people, don't we end out in a huge war of all against all until either everyone is dead or one guy is dictator?

Would that be a good consequence? If not, people who try to promote good consequences and make the world a better place would try to avoid it.

Because this world of violence and competition is so obviously a bad consequence, any consequentialist who gives it a moment's thought agrees not to start a huge war of all against all that ends with everyone dead or one guy as dictator by binding themselves by moral rules whenever binding themselves by those moral rules seems like it would have good consequences or make the world a better place; see Section 13 for more.

12.4.4: Doesn't that sound a lot like "the ends justify the means"? Wouldn't it lead to decadence, slavery, or some other dystopia?

Once again, if you consider dictatorship, slavery, and dystopia to be bad consequences, then by definition following this rule is the best way to *avoid* doing that.

The rule isn't "do whatever sounds like it would have the best consequences if you have an IQ of 20 and refuse to think about it for even five seconds", it's "do what would *actually* have the best consequences. Sometimes this involves admitting human ignorance and fallibility and *not* pursuing every hare-brained idea that comes into your head.

12.4.5: Okay, okay, I understand that if people did what *actually* had good consequences it would have good consequences, but I worry that if people do what they *think* has good consequences, it will lead to violence and dictatorship and dystopia and all those other things you mentioned above.

Yes, I agree this is an important distinction. There are two uses for a moral system. The first is to define what morality is. The second is to give people a useful tool for choosing what to do in moral dilemmas. I am arguing that consequentialism does the first. I don't think it does the second right out of the box.

To try a metaphor, doctors sometimes have two ways of defining disease; the gold standard and the clinical standard. The gold standard is the "perfect" test for the disease; for example, in Alzheimer's disease, it's to autopsy the brain after the person has died and see if it has certain features under the microscope. Obviously you can't autopsy a person who's still alive, so when doctors are actually trying to diagnose Alzheimer’s they use a more practical method, like how well the person does on a memory test.

Right now I’m arguing that consequentialism is the gold standard for morality: it’s the purest, most sophisticated explanation of what morality actually is. At the same time, it might be a terrible idea to make your everyday decisions based on it, just as it’s a terrible idea to diagnose Alzheimer’s with an autopsy in someone who’s still alive.

However, once we know that consequentialism is the gold standard for morality, we can start designing our clinical standards by trying to figure out which “clinical standard” for morality will produce the best consequences. See Section 13 for more.

12.4.6: I still am not completely on board with consequentialism, or I’m not sure I understand it.

For more information on consequentialism, see the sister document to this FAQ, the Consequentialism FAQ.

13. Rights and Heuristics

13.1: Is there a moral justification for rights, like the right to free speech or the right to property?

Yes. Rights are the “clinical standard” for morality, the one we use to make our everyday decisions after we acknowledge that pure consequentialism might not lead to the best consequences when used by fallible humans.

In this conception, rights are conclusions rather than premises. They are heuristics (heuristic = a rule-of-thumb that usually but not always works) for remembering what sorts of things usually have good or bad consequences, a distillation of moral wisdom that is often more trustworthy than morally fallible humans.

For example, trying to tell people what religions they can or can't follow almost always has bad consequences. At best, people are miserable because they're being forced to follow a faith they don't believe in. At worst, they resist and then you get Inquisitions and Holy Wars and everyone ends up dead. Restriction of religion causing bad consequences is sufficiently predictable that we generalize it into a hard and fast rule, and call that rule something like the "right to freedom of religion".

Other things like banning criticism of the government, trying to prevent people from owning guns, and seizing people's property willy-nilly also work like this, so we call those "rights" too.

13.2: So if you think that violating rights will have good consequences, then it's totally okay, right?

It's not quite so simple. Rights are not just codifications of the insight that certain actions lead to bad consequences, they're codifications of the insight that certain actions lead to bad consequences in ways that people consistently fail to predict or appreciate.

All throughout history, various despots and princes have thought "You know, the last hundred times someone tried to restrict freedom of religion, it went badly. Luckily, *my* religion happens to be the One True Religion, and I'm totally sure of this, and everyone else will eventually realize this and fall in line, so *my* plan to restrict freedom of religion will work great!"

Every revolution starts with an optimist who says "All previous attempts to kill a bunch of people and seize control of the state have failed to produce a utopia, but luckily *my* plan is much better and we're totally going to get to utopia this time." Or, as Huxley put it: "Only one more indispensable massacre of Capitalists or Communists or Fascists and there we are – there we are – in the Golden Future."

So another way to put it is that rights don't just say "Doing X has been observed to have bad consequences", but also "Doing X has been observed to have bad consequences, even when smart people are quite certain it will have good consequences."

13.3: Then even though you got to rights by a different route than the libertarians, it sounds like you agree with them that they're inalienable.

It's not as simple as *that* either. Every so often, the conventional wisdom is wrong. So many lunatics and crackpots spent their lives trying to turn lead into gold that it became a classic metaphor for a foolish wild goose chase. The rule "stop trying to transmute elements into each other, it never works" was no doubt a good and wise rule. If more would-be alchemists had trusted this conventional wisdom, and fewer had thought "No, even though everyone else has failed, *I* will be the one to discover transmutation", it would have prevented a lot of wasted lives.

...and then we discovered nuclear physics, which is *all about* transmuting elements into one another, and which works very well and is a vital source of power. And yes, nuclear physicists at Berkeley successfully used a giant particle accelerator to turn lead into gold, although it only works a few atoms at a time and isn't commercially viable.

The point is, the heuristic that you shouldn't waste your life studying transmutation was a good one and very well-justified at the time, but if we had elevated it into a timeless and unbreakable principle, we never would have been able to abandon it after we learned more about nuclear physics and trying to transmute things was no longer so foolish.

Rights are a warning sign that we should not naively expect breaking them to have good consequences. In order to claim even the possibility of good consequences from violating a right, we need to be at least as far away from the actions they were meant to prevent as nuclear physics is to alchemy.

13.3.1: Can you give an example of a chain of reasoning where some government violation of a right is so radically different from the situation that led the right to exist in the first place?

Let's take for example the right that probably dominates discussions between libertarians and non-libertarians: the right to property. On the individual scale, taking someone else's property makes them very unhappy, as you know if you've ever had your bike stolen. On the larger scale, abandoning belief in private property has disastrous results for an entire society, as the experiences of China and the Soviet Union proved so conclusively. So it's safe to say there's a right to private property.

Is it ever acceptable to violate that right? In the classic novel *Les Miserables*, Jean Valjean's family is trapped in bitter poverty in 19th century France, and his nephew is slowly starving to death. Jean steals a loaf of bread from a rich man who has more than enough, in order to save his nephew's life. This is a classic moral dilemma: is theft acceptable in this instance?

We can argue both sides. A proponent might say that the good consequences to Jean and his family were very great – his nephew's life was saved – and the bad consequences to the rich man were comparatively small – he probably has so much food that he didn’t even miss it, and if he did he could just send his servant to the bakery to get another one. So on net the theft led to good consequences.

The other side would be that once we let people decide whether or not to steal things, we are on a slippery slope. What if we move from 19th century France to 21st century America, and I’m not exactly starving to death but I really want a PlayStation? And my rich neighbor owns like five PlayStations and there’s no reason he couldn’t just go to the store and buy another. Is it morally acceptable for me to steal one of his PlayStations? The same argument that applied in Jean Valjean’s case above seems to suggest that it is – but it’s easy to see how we go from there to everyone stealing everyone’s stuff, private property becoming impossible, and civilization collapsing. That doesn’t sound like a very good consequence at all.

If everyone violates moral heuristics whenever they personally think it’s a good idea, civilization collapses. If no one ever violates moral heuristics, Jean Valjean’s nephew starves to death for the sake of a piece of bread the rich man never would have missed.

We need to bind society by moral heuristics, but also have some procedure in place so that we can suspend them in cases where we’re exceptionally sure of ourselves without civilization instantly collapsing. Ideally, this procedure should include lots of checks and balances, to make sure no one person can act on her own accord. It should reflect the opinions of the majority of people in society, either directly or indirectly. It should have access to the best minds available, who can predict whether violating a heuristic will be worth the risk in this particular case.

Thus far, the human race’s best solution to this problem has been governments. Governments provide a method to systematically violate heuristics in a particular area where it is necessary to do so without leading to the complete collapse of civilization.

If there was no government, I, in Jean Valjean’s situation, absolutely would steal that loaf of bread to save my nephew’s life. Since there is a government, the government can set a certain constant amount of theft per year, distribute the theft fairly among people whom it knows can bear the burden, and then feed starving children and do other nice things. The ethical question of “is it ethical for me to steal/kill/stab in this instance?” goes away, and society can be peaceful and stable.

13.3.2: So you’re saying that you think in this case violating the right will have good consequences. But you just agreed that *even when people think this*, violating the right usually has bad consequences.

Yes, I admit it’s complicated. But we have to have some procedures for violating moral heuristics, or else we can’t tax to support a police force, we can’t fight wars, we can’t lie to a murderer who asks us where our friend is so he can go kill her when he finds her, and so on.

The standard I find most reasonable is when it’s universalizable and it avoids the issue that caused us to develop the heuristic in the first place.

By universalizable, I mean that it’s more complicated than me just deciding “Okay, I’m going to steal from this guy now”. There has to be an agreed-upon procedure where everyone gets input, and we need to have verified empirically that this procedure usually leads to good results.

And is has to avoid the issue that caused us to develop the heuristic. In the case of stealing, this is that theft makes property impossible or at least impractical, no one bothers doing work because it will all be stolen from them anyway, and so civilization collapses.

In the case of theft, taxation requires authorization by a process that most of us endorse (the government set up by the Constitution) and into which we all get some input via representative democracy. It doesn't cause civilization to collapse because it only takes a small and extremely predictable amount from each person. And it's been empirically verified to work: as I argued above, countries with higher tax rates like Scandinavia actually *are* nicer places to live than countries with lower tax rates like the United States. So we've successfully side-stepped the insight that stealing usually has bad consequences, even though we recognize that the insight remains true.

13.4: Governments will inevitably make mistakes when deciding when to violate moral heuristics. Those mistakes will cost money and even lives.

And the policy of never, ever doing anything will never be a mistake?

It's very easy for governments to make devastating mistakes. For example, many people believe the US government's War in Iraq did little more than devastate the country, kill hundreds of thousands of Iraqis, and replace Saddam with a weak government unable to stand up to extremist ayatollahs.

But the other solution – never intervening in a foreign country at all – didn't work so well either. Just look at Holocaust-era Germany, or 1990s Rwanda.

Why, exactly, should moral questions be simple?

There is a certain tradition that the moral course of action is something anyone, from the high priest unto the youngest child, can find simply by looking deep in his heart. Anyone who does not find it in his heart is welcome to check the nearest Giant Stone Tablet, upon which are written infallible rules that can guide him through any situation. Intelligence has nothing to do with it. It should be blindingly obvious, and anyone who claims it has a smidgen of difficulty or vagueness is probably an agent of the Dark Lord, trying to seduce you from the True Path with his lies.

And so it is tempting to want to have some really easy principle like “Never get involved in a foreign war” and say it can never lead you wrong. It makes you feel all good and warm and fuzzy and moral and not at all like those evil people who don’t have strong principles. But real life isn’t that simple. If you get involved in the wrong foreign war, millions of people die. And if you don’t get involved in the right foreign war, millions of people also die.

So you need to have good judgment if you want to save lives and do the right thing. You can’t get a perfect score in morality simply by abdicating all responsibility. Part of the difficult questions that all of us non-libertarians have been working on is how to get a government that’s good at answering those sorts of questions correctly.

13.5: No, there’s a difference. When you enter a foreign war, you’re killing lots of people. When you don’t enter a foreign war, people may die, but it’s not your job to save them. The government’s job is only to protect people and property from force, not to protect people from the general unfairness of life.

Who died and made you the guy who decides what the government's job is? Or, less facetiously: on what rational grounds are you making that decision?

Currently, several trillion dollars are being spent to prevent terrorism. This seems to fall within the area of what libertarians would consider a legitimate duty of government, since terrorists are people who initiate force and threaten our safety and the government needs to stop this. However, terrorists only kill an average of a few dozen Americans per year.

Much less money is being spent on preventing cardiovascular disease, even though cardiovascular disease kills 800,000 Americans per year.

Let us say, as seems plausible, that the government can choose to spend its money either on fighting terrorists, or on fighting CVD. And let us say that by spending its money on fighting terrorists, it saves 40 lives, and by spending the same amount of money on fighting CVD, it saves 40,000 lives.

All of these lives, presumably, are equally valuable. So there is literally no benefit to spending the money on fighting terrorism rather than CVD. All you are doing is throwing away 39,960 lives on an obscure matter of principle. It's not even a good principle – it's the principle of wanting to always use heuristics even when they clearly don't apply because it sounds more elegant.

There's a reason this is so tempting. It's called the Bad Guy Bias, and it's an evolutionarily programmed flaw in human thinking. People care much more about the same amount of pain when it's inflicted by humans than when it's inflicted by nature. Psychologists can and have replicated this in the lab, along with a bunch of other little irrationalities in human cognition. It's not anything to be ashamed of; everyone's got it. But it's not something to celebrate and raise to the level of a philosophical principle either.

13.6: Stop calling principles like "don't initiate force" heuristics! These aren't some kind of good idea that works in a few cases. These are the very principles of government and morality , and it's literally impossible for them to guide you wrong!

Let me give you a sketch of one possible way that a libertarian perfect world that followed all of the appropriate rules to the letter could end up as a horrible dystopia. There are others, but this one seems most black-and-white.

Imagine a terrible pandemic, the Amazon Death Flu, strikes the world. The Death Flu is 100% fatal. Luckily, one guy, Bob, comes up with a medicine that suppresses (but does not outright cure) the Death Flu. It's a bit difficult to get the manufacturing process right, but cheap enough once you know how to do it. Anyone who takes the medicine at least once a month will be fine. Go more than a month without the medicine, and you die.

In a previous version of this FAQ, Bob patented the medicine, and then I got a constant stream of emails saying (some) libertarians don't believe in patents. Okay. Let's say that Bob doesn't patent the medicine, but it's complicated to reverse engineer, and it would definitely take more than a month. This will become important later.

Right now Bob is the sole producer of this medicine, and everyone in the world needs to have a dose within a month or they'll die. Bob knows he can charge whatever he wants for the medicine, so he goes all out. He makes anyone who wants the cure pay one hundred percent of their current net worth, plus agree to serve him and do anything he says. He also makes them sign a contract promising that while they are receiving the medicine, they will not attempt to discover their own cure for the Death Flu, or go into business against him. Because this is a libertarian perfect world, everyone keeps their contracts.

A few people don’t want to sign their lives away to slavery, and refuse to sign the contract. These people receive no medicine and die. Some people try to invent a competing medicine. Bob, who by now has made a huge amount of money, makes life difficult for them and bribes biologists not to work with them. They’re unable to make a competing medicine within a month, and die. The rest of the world promises to do whatever Bob says. They end up working as peons for a new ruling class dominated by Bob and his friends.

If anyone speaks a word against Bob, they are told that Bob’s company no longer wants to do business with them, and denied the medicine. People are encouraged to inform on their friends and families, with the promise of otherwise unavailable luxury goods as a reward. To further cement his power, Bob restricts education to the children of his friends and strongest supporters, and bans the media, which he now controls, from reporting on any stories that cast him in a negative light.

When Bob dies, he hands over control of the medicine factory to his son, who continues his policies. The world is plunged into a Dark Age where no one except Bob and a few of his friends have any rights, material goods, or freedom. Depending on how sadistic Bob’s and his descendants are, you may make this world arbitrarily hellish while still keeping perfect adherence to libertarian principles.

Compare this to a similar world that followed a less libertarian model. Once again, the Amazon Death Flu strikes.

Once again, Bob invents a cure. The government thanks him, pays him a princely sum as compensation for putting his cure into the public domain, opens up a medicine factory, and distributes free medicine to everyone. Bob has become rich, the Amazon Death Flu has been conquered, and everyone is free and happy.

13.6.1: This is a ridiculously unlikely story with no relevance to the real world.

I admit this particular situation is more a *reductio ad absurdum* than something I expect to actually occur the moment people start taking libertarianism seriously, but I disagree that it isn't relevant.

The arguments that libertarianism will protect our values and not collapse into an oppressive plutocracy require certain assumptions: there are lots of competing companies, zero transaction costs, zero start-up costs, everyone has complete information, everyone has free choice whether or not to buy any particular good, everyone behaves rationally, et cetera. The Amazon Death Flu starts by assuming the opposite of all of these assumptions: there is only one company, there are prohibitive start-up costs, a particular good absolutely has to be bought, et cetera.

The Amazon Death Flu world, with its assumptions, is not the world we live in. But neither is the libertarian world. Reality lies somewhere between the "capitalism is perfect" of the one, and the "capitalism leads to hellish misery" of the other.

There's no Amazon Death Flu, but there are things like hunger, thirst, unemployment, normal diseases, and homelessness. In order to escape these problems, we need things provided by other people or corporations. This is fine and as it should be, and as long as there's a healthy free market with lots of alternatives, in most cases these other people or corporations will serve our needs and society's needs while getting rich themselves, just like libertarians hope.

But this is a contingent fact about the world, and one that can sometimes be wrong. We can't just assume that the heuristic "never initiate force" will *always* turn out well.

13.7: The government doesn't need to violate moral heuristics. In the absence of government programs, private charity would make up the difference.

Find some poor people in a country without government-funded welfare, and ask how that's working out for them.

Private charity from the First World hasn't prevented the Rwandans, Ethiopians, or Haitians from dying of malnutrition or easily preventable disease.

It's possible that this is just because we First Worlders place more importance on our own countrymen than on foreigners, and if Americans were dying of malnutrition or easily preventable disease, patriotism would make us help them.

The US government currently spends about $800 billion on welfare-type programs for US citizens. Americans give a total of $300 billion to charity per year.

Let's assume that private charity is twice as efficient as the government (in reality, it's probably much less, since the government has economies of scale, but libertarians like assumptions like this and I might as well indulge them).

Let's also assume that only half of charity goes to meaningful efforts to help poor American citizens. The other half would be things like churches, the arts, and foreign countries.

Nowadays, a total of $550 billion (adjusted, govt+private) goes to real charity (800b*1/2+300b*1/2). If the government were to stop all welfare programs, this number would fall to $150 billion (adjusted). Private citizens would need to make up the shortfall of $400 billion to keep charity at its current (woefully low) level. Let’s assume that people, realizing this, start donating a greater proportion (66%) of their charity to the American poor instead of to other causes. That means people need to increase their charity to about $830 billion ([400b + 150b]/.66).

Right now, 25% is a normal middle-class tax rate. Let’s assume the government stopped all welfare programs and limited itself to defense, policing, and overhead. There are a lot of different opinions about what is and isn’t in the federal budget, but my research suggests that would cut it by about half, to lower tax rates to 12.5%.

So, we’re in the unhappy situation of needing people to almost triple the amount they give to charity even though they have only 12.5% more money. The real situation is much worse than this, because if the government stopped all programs except military and police, people would need to pay for education, road maintenance, and so on out of their own pocket.

My calculations are full of assumptions, of course. But the important thing is, I’ve never seen libertarians even try to do calculations. They just assume that private citizens would make up the shortfall. This is the difference between millions of people leading decent lives or starving to death, and people just figure it will work out without checking, because the free market is always a Good Thing.

That’s not reason, even if you read it on www.reason.com. That’s faith.

13.8: People stupid enough to make bad decisions deserve the consequences of their actions. If government bans them from making stupid decisions, it’s just preventing them from getting what they deserve.

One of my favorite essays, Policy Debates Should Not Appear One-Sided, provides a much better critique of this argument than I could. It starts by discussing a hypothetical in which the government stopped regulating the safety of medicines. Some quack markets sulfuric acid as medicine, and a “poor, honest, not overwhelmingly educated mother of five children” falls for it, drinks it, and dies.

If you were really in that situation, would you really laugh, say “Haha, serves her right” and go back to what you were doing? Or would it be a tragedy even though she “got what she deserved”?

The article ends by saying:

> *Saying ‘People who buy dangerous products deserve to get hurt!’ is not tough-minded. It is a way of refusing to live in an unfair universe. Real tough-mindedness is saying, ‘Yes, sulfuric acid is a horrible painful death, and no, that mother of 5 children didn’t deserve it, but we’re going to keep the shops open anyway because we did this cost-benefit calculation.’…I don’t think that when someone makes a stupid choice and dies, this is a cause for celebration. I count it as a tragedy. It is not always helping people, to save them from the consequences of their own actions; but I draw a moral line at capital punishment. If you’re dead, you can’t learn from your mistakes.*

Read also about the just-world fallacy. “Making a virtue out of necessity” shouldn’t go as far as celebrating deaths if it makes your political beliefs more tenable.

Part E: Practical Issues

The Argument: *Allowing any power to government is a slippery slope toward tyranny. No matter what the costs or benefits of any particular proposal, libertarians should oppose all government intrusion as a matter of principle.*

The Counterargument: *This fundamentally misunderstands the ways that nations collapse into tyranny. It also ignores political reality, and it doesn't work. Libertarians should cooperate with people from across the ideological spectrum to oppose regulations that doesn't work and keep an open mind to regulation that might.*

14. Slippery Slopes

14.1: I'm on board with doing things that have the best consequences. And I'm on board with the idea that *some* government interventions *may* have good consequences. But allowing any power to government is a slippery slope. It will inevitably lead to tyranny, in which do-gooder government officials take away all of our most sacred rights in order to "protect us" from ourselves.

History has *never* shown a country sinking into dictatorship in the way libertarians assume is the "natural progression" of a big-government society. No one seriously expects Sweden, the United Kingdom, France, or Canada to become a totalitarian state, even though all four have gone much further down the big-government road than America ever will.

Those countries that have collapsed into tyranny have done so by having so *weak* a social safety net and so *uncaring* a government that the masses felt they had nothing to lose in instituting Communism or some similar ideology. Even Hitler gained his early successes by pretending to be a champion of the populace against the ineffective Weimar regime.

Czar Nicholas was not known for his support of free universal health care for the Russian peasantry, nor was it Chiang Kai-Shek's attempts to raise minimum wage that inspired Mao Zedong. It has generally been among *weak* governments and a *lack* of protection for the poor where dictators have found the soil most fertile for tyranny.

14.1.1: But still, if we let down our guard, bureaucrats and politicians will have free rein to try to institute such a collapse into dictatorship.

I have always found the libertarian conviction that all politicians are secretly trying to build up their own power base to *1984*-ish levels a bit weird.

All the time, I am hearing things like "No one really believes in global warming. It's just a plot by the government to expand control over more areas of your life." Or "since private charity is a threat to government's domination of social welfare, once government gets powerful enough it will try to ban all private charity."

Sure, people really do like power. But usually it's the sort of power that comes with riches, fame, and beautiful women willing to attend to your every need. Just sitting in your office, knowing in an abstract way that because of you a lot of people who might otherwise be doing useful industry are fretting about their carbon emissions – that's not the kind of power people sell their souls for. The path to ultimate domination of all humanity does not lead through the Dietary Fiber Levels in Food Act of 2006.

Most folk like to think of themselves as good people. Sure, they may take a bribe or two here, and have an affair or two there, and lie about this and that, "but only for the right reasons." The thought process "Let me try to expand this unnecessary program so I can bathe in the feeling of screwing American taxpayers out of more of their hard-earned money" is not the kind that comes naturally, especially in a society where it leads to minimal personal gain. A politician who raises your taxes can't use the money to buy himself a new Ferrari. At least, he can't do it directly, and if he really wants that Ferrari there have got to be much easier ways to get it.

Human beings find it hard to get angry at a complicated system, and prefer to process things in terms of evil people doing evil things. Eliezer Yudkowsky of Less Wrong writes:

> *Suppose that someone says "Mexican-Americans are plotting to remove all the oxygen in Earth's atmosphere." You'd probably ask, "Why would they do that? Don't Mexican-Americans have to breathe too? Do Mexican-Americans even function as a unified conspiracy?" If you don't ask these obvious next questions when someone says, "Corporations are plotting to remove Earth's oxygen," then "Corporations!" functions for you as a semantic stopsign.*

And if you don't ask some of these same questions when someone says "Government wants to take away freedom!," then you're not thinking of government as a normal human institution that acts in normal human ways.

15. Strategic Activism

15.1: All you've argued so far is that it's possible, in theory, for an *ideal* government making some very clever regulations to do a little more good than harm.

But that doesn’t prove that the *real* government does more good than harm, and in fact it’s probably the opposite. So shouldn’t we admit that in a hypothetical perfect world government might do some good, while still being libertarians in reality?

I think if you’ve got enough intelligence and energy to be a libertarian, a better use of that intelligence and energy would be to help enact a properly working system.

15.2: It’s impossible to improve government; because power corrupts, all conceivable forms of government will be ineffective, wasteful, and dishonest.

“Impossible” is a really strong word.

Economist Robin Hanson has a proposal for a market-based open-source form of government called “futarchy”, in which government policies are decided entirely by a prediction market. Prediction markets operate similarly to stock markets and allow participants to buy or sell shares in predictions – for example, a share that pays out $100 if the economy improves this year, but $0 if the economy deteriorates. If it settles around a price of $60, this means the investing public predicts as 60% chance that the economy will go up.

A prediction market could be used to set policy by predicting its effects: for example, by comparing the prices of “we will institute the president’s economic plan, and the economy will improve” , “we will not institute the president’s economic plan, and the economy will improve” and “we will institute the president’s economic plan”, we can determine the public’s confidence that the president’s plan will improve the economy. There are some nifty theorems of economics that prove that such a market would produce a more accurate estimate of the plan’s chances than any other conceivable method (including consulting experts), and that it would be very difficult to corrupt. You can read more about it here.

My point isn't that futarchy would definitely work. It's that it's an example of some of the best ideas that smart people trying to improve government can come up with. And unless you're creative enough to develop futarchy on your own, or well-read enough to be sure you've heard of it and everything else like it, you're being premature in calling improvements in government "impossible".

15.3: Even if there are ways to improve government, they are impractical because they're too politically unpopular.

Let's be totally honest here. The US Libertarian Party currently has a grand total of zero state legislators, zero state governors, zero representatives, and zero senators. It's never gotten much above one percent in any presidential election. Nor have any successful or nationally known major-party candidates endorsed genuinely libertarian ideals except maybe Ron Paul, who just suffered his third landslide defeat.

The libertarian vision of minimal government is politically impossible to enact. This is not itself an argument against it – most good ideas are – but it does mean you can't condemn the alternatives for being politically impossible to enact.

Incremental attempts to improve government have a much better track record, both in terms of political palatability and success rate, than libertarian efforts to dismantle government whole-cloth. If you want to focus on something that might work, you should concentrate your efforts there.

15.4: Isn't it better to draw a line in the sand and say no government intervention at all? This keeps us off

the slippery slope to the kind of awful, huge government we have today.

Empirically, no. Again I point out that libertarianism has been completely ineffective as a political movement. The line-in-the-sand idea is an interesting one but obviously hasn't worked.

And there are some serious advantages to erasing it. If non-libertarians see libertarians as ideologues who hate all government programs including the ones that could work, then they will dismiss any particular libertarian objection as meaningless: why pay attention to the fact that a libertarian hates this particular bill, when she hates *every* bill?

But if libertarians took a principled stand in favor of some government regulation that might work, they could credibly say "Look, it's not that we have a knee-jerk hatred for all possible regulations, it's just that *this particular regulation is a horrible idea*." And people might listen.

It *might* also help arrest the polarization of society into factions who apply ideological "litmus tests" to all proposals before even hearing them out (eg pretty much all self-described "progressives" will automatically support any proposal to be tougher on pollution without even looking at what the economic costs versus health benefits will be, and most self-described libertarians will automatically oppose it just as quickly.) This sort of thing needs to stop, libertarians are one of the at least two groups who need to stop it, and the more people who stop, the more people on both sides will notice what they're doing and think about it a little harder.

16. Miscellaneous and Meta

16.1: I still disagree with you. How should I best debate you and other non-libertarians in a way that is most likely to change your mind?

The most important advice I could give you is don’t come on too strong. Words like “thievery” and “enslave” are emotional button pressers, not rational arguments. Attempts to insult your opponents by calling them tyrants or suggesting they want to rule over the rest of humanity as slaves and cattle (yes, I’ve gotten that) is more likely to annoy than convince. And please, stop the “1984” references, especially when you’re talking about a modern liberal democracy. Seriously. It’s like those fundamentalists who have websites about how not having prayer in school is equivalent to the Holocaust.

Many non-libertarians aren’t going to be operating from within the same moral system you are. Sometimes the libertarians I debate don’t realize this and this causes confusion when they try to argue that something’s morally wrong. If you want to convince your opponent on moral grounds, you’re either going to have to show how their theories fail *even by their own moral standards*, or else prove your standards are right by deriving them from first principles (warning: this might be impossible).

Don’t immediately assume that just because we are not libertarians, we must worship Stalin, love communism, think government should be allowed to control every facet of people’s lives, or even support things like gun control or the War on Drugs. Non-libertarianism is a lot like non-Hinduism: it’s a pretty diverse collection of viewpoints with everything from full-on fascists to people who are totally libertarian except about one tiny thing.

Finally, you may have better luck convincing us of specific points, like “Government should not set a minimum wage” than broad slogans, like “Government can never do anything right.” It’s *really* hard to prove a universal negative.

16.2: Where can I go to see a rebuttal to this FAQ?

David Friedman wrote a short response here

Bryan Caplan wrote a response to some of the points about labor here.

Sarah wrote a longer rebuttal here: Why You Shouldn’t Hate My Freedom.

And Nintil wrote another long rebuttal here: The Non-Non Libertarian FAQ

If you’ve written another rebuttal or you know of one, email me and I’ll add it here.

16.3: Where can I go to find more non-libertarian information?

Mike Huben has a *terrifyingly* large collection of non-libertarian and anti-libertarian material of wildly varying quality and tone at his website.

Some Groups Of People Who May Not 100% Deserve Our Eternal Scorn

Or "Contra A Convergence Of Lefty and Far-Right Twitter Making Fun Of The Same People". I'll mostly be using *Current Affairs* articles as foils, not because they're especially bad, but because they're especially good and well-written expressions of what many other people are saying. Sorry if this is a little snarky and maybe not 100% fair.

1. Celebrities Who Speak Out Against Donald Trump

No, celebrities are not going to single-handedly change the world. Yes, celebrities are often annoying, and almost by definition out-of-touch. Yes, a Democratic campaign needs to have some substance beyond "look, celebrities!"

But from the celebrities' own point of view, *they're doing the best they can*. If Kim Kardashian wants to help the cause, what do you expect her to do? Write policy white papers? Go door-to-door canvassing? Or would you rather she just stayed silent and didn't do anything?

Also, I think the "out-of-touch" critique sort of misses the point. Unemployed high school dropouts aren't going to read Paul Krugman editorials, and they might or might not go to Bernie Sanders rallies. But I guarantee they know who Kim Kardashian is. Now, maybe you don't develop your opinions by listening to weird-looking people who seem to be famous for no reason, and maybe you're proud of that fact. But judging by the amount of money people will pay celebrities to endorse their products, a lot of people *do* develop their opinions this way. And these people are probably the less-educated working-class folks whom the Democrats most need to reach.

Or maybe I’m just being classist and *nobody* listens to celebrities. Fine. I still think that a celebrity who speaks out about something they think is important is more virtuous than one who doesn’t. By all means criticize tone-deaf celebrities like Lena Dunham who “help” the cause by speaking up in offensive or counterproductive ways. But criticizing celebrities’ activism in general doesn’t seem like a good political strategy.

2. People Who Compare Political Events To Harry Potter

See eg here.

Comparing politics to your favorite legends is as old as politics and legends. Herodotus used an extended metaphor between the Persian invasions of his own time and the Trojan War. When King Edward IV took the English throne in 1461, all anybody could talk about was how it reminded them of King Arthur. John Dryden’s famous poem *Absalom and Achitophel* is a bizarrely complicated analogy of 17th-century English politics to an obscure Biblical story. Throughout American history people have compared King George to Pharaoh, Benedict Arnold to Judas, Abraham Lincoln to Moses, et cetera.

Well, how many people know who Achitophel is these days? Even Achilles is kind of pushing it. So we stick to what we know – and more important, what we expect everyone else will know too. And so we get Harry Potter.

“But a *children’s book?*” Look, guys, fantasy is what the masses actually like. They liked it in Classical Greece, where they had stories like Bellerophon riding a flying horse and fighting the Chimera. They liked it in medieval Britain, where they would talk about the Knights of the Round Table slaying dragons as they searched for the Holy Grail. The cultural norm where only kids are allowed to read fantasy guilt-free and everybody else has to read James Joyce is a weird blip in the literary record which is already being corrected. Besides, James Joyce makes for a much less interesting source of political metaphors ("The 2016 election was a lot like *Finnegan's Wake*: I have no idea what just happened")

Harry Potter is not the national mythology I would have chosen. Probably I would have gone for *Lord of the Rings*. I'm not sure we as a nation deserve *The Silmarillion*, but a man can dream.

But Harry Potter is at least better than some things (we could have ended out with our national consciousness being shaped by *Twilight*!), and the point is that comparing your politics to those of a more interesting fantasy world is a natural human urge and probably not indicative of some sort of horrible decay.

3. People Who Like Hamilton

See eg here.

Look. Hamilton was a pretty good Broadway play. It wasn't the best thing that ever happened. It didn't single-handedly reinvent America.

On the other hand, it's also not the source of all evil. It's not some sort of giant glowing tribute to national elitism where everyone gathers together and eats arugula and talks about how much they prefer symbolic gestures involving identity to actual systemic change. It's just a pretty good Broadway play.

4. Vox

See eg here.

I think the main complaint is that "explaining the news" is fundamentally condescending. *Real* Americans personally read all 9,800 pages of Obamacare regulations before forming an opinion on health policy.

Or maybe the complaint is that they're pretending to do it from an objective point of view instead of admitting that they have a liberal bias? I will take this complaint seriously when I meet any person anywhere in the world who is not aware that Vox has a liberal bias. The aboriginal people of the North Sentinel Islands have been completely isolated from the rest of civilization for thousands of years, yet every single child in their tribe knows that Vox has a liberal bias. SETI believes that if we contact aliens, we will have to determine their language through universally known truths like prime numbers or the digits of pi, but if for some reason the aliens have different mathematics than we do, we will still be able to communicate over a shared understanding that Vox has a liberal bias.

This is fine. All attempts to explain the news are going to end up with some bias, and I'm okay with this as long as they try to minimize it, present the truth as they understand it, and give more light than heat (though see here)

And that's where my experience with Vox has been reassuring. I've occasionally argued with them, or made fun of them, or SHOUTED AT THEM THAT THEY ARE SPREADING DAMNABLE LIES. And every time, I've been impressed by their kindness, their openness to criticism, and their willingness to pay attention to me even though I can be very annoying.

Fredrik deBoer has a theory that everybody secretly hates Ezra Klein but publicly pretends to like him because he's powerful. And I keep wanting to protest that *I* like Ezra Klein, before realizing that deBoer's theory predicts I would say that. So I’ll just add that my interactions with Klein have consisted mostly of me yelling at him for being wrong about everything, and him politely listening to me. A few times he’s admitted he was wrong and promised to do better (and has). Other times he’s stuck to his position while continuing to give me way more of his time and energy than I would expect the head of a big media company to give a random and somewhat-confrontational blogger.

This has also been more or less my experience with Dylan Matthews, German Lopez, and Sarah Kliff, the other Vox people I’ve engaged with.

A year or so ago, the media got really interested in neoreaction and published a bunch of thinkpieces, all of which parroted an error-ridden Breitbart article without checking any of its claims. Dylan Matthews wanted to write one for Vox, and he actually took the trouble to contact me, an Internationally Known Expert On Neoreaction. I corrected a few of the worst Breitbart errors and gave him the email address of a couple of neoreactionaries; Matthews actually interviewed them and included their comments in his article instead of relying on third-hand speculation about who they might be. I have heard legends that ancient times there was an arcane art called Juru-Na-Lism which allowed its practitioners to gather information from the furthest reaches of the world, and although I understand it is mostly forgotten this gives me some glimmers of what it could have been like (and for an even clearer example of the same pattern, compare this and this).

Also, Stuart Ritchie is a scientist at the University of Edinburgh who studies intelligence and who makes fun of terrible articles about intelligence in the media. Vox actually worked with Dr. Ritchie to write a series of articles, and ended up with some of the only popular explanations online that someone with a psychology background can read without laughing hysterically.

I disagree with Vox about a lot of things, but they’ve generally impressed me in ways that some other news sources haven’t. Also, let’s be honest. Their competitors are places like Salon and Vice. My standards here are dirt-low, and Vox frequently meets them.

5. Matt Yglesias

Related; see eg here:

> The worst of Yglesias’ mischievous endorsements of horrendous moral stances was his column on factory safety. Immediately after the 2013 collapse of the Bangladesh garment factory that killed over 1,000 people, Yglesias took to Slate to explain why workplace safety regulations actually inhibited the operation of free markets. Yglesias explained that high-risk jobs have high compensation, and just like people might choose to be lumberjacks, they might choose to work in highly dangerous garment factories for a premium. Thus “it’s good that different people are able to make different choices on the risk–reward spectrum.” The article was accompanied by a photograph of Bangladeshis loading dead bodies onto a truck.
>
> The column was classic Yglesias, in managing to be both ignorant and appalling. Appalling since Yglesias published it the same day as the factory collapse, as the rubble was still being cleared. Ignorant because Yglesias adopted the most delusional Heritage Foundation economic myth, that somehow people in Bangladesh work in dangerous garment factories because working in dangerous garment factories is what they most want to do. As Mark Brendle summarized:

> Yglesias champions one of the most horrifying and widespread implements of oppression and misery yet conceived—factories taking advantage of cheap labor, lack of environmental regulations, and a disregard for human life by those who profit most from having those factories in their countries—then pretends that it exists in a vacuum, where people in “those countries” are happy for these jobs, instead of acknowledging the closed system of the global economy, where those conditions are not only systemic, but inevitable and structural, in order for the wealth and prosperity of the “first world” to exist at all.
>
> When confronted with this outrage, Yglesias simply wrote another explanation of why his original work was justified, admitting that his reaction to the criticism “as a writer and a human being” was annoyance. (It should go without saying that if one’s first reaction “as a human being” to being asked to show a little compassion for dead Bangladeshis is “annoyance,” then one is not a human being at all.) Here is Vox-ism in a nutshell: it is impossible to stop explaining and think, impossible to understand that there are more questions in heaven and earth than “What do the data say?” (Like perhaps, “Am I a good person?”)

One day soon, there’s going to be an Islamic terror attack in the United States, maybe committed by a refugee. The news is going to show pictures of mangled innocents, sobbing relatives, mothers who have lost their children. And maybe *Current Affairs*, as a good leftist publication, is going to want to say that this is terrible but doesn’t mean that we should ban all refugees or hate all Muslims.

And they won’t be able to, because they’ve already declared that if something tragic happens, then anyone who tries to put it in context, or say that some policies can have occasional awful results while still being beneficial on net, is a moral monster.

And if they try to protest that no, approximately 0% of refugees are terrorists, immigrant crime rates are lower than native crime rates, all of the fear-mongering you've heard is a lie, et cetera et cetera – then ah, that's just worrying about "what the data say" – and how can you worry about something as bloodless as data when there are families *literally* sobbing over the deaths of their children right there?

Trump should be ultimate proof that the other side is better at the "my righteous indignation is more important than your puny data" game than you are. Don't even try.

6. Pundits Who Failed To Predict Trump

See eg Michael Tracey in How Pundits Get Everything Wrong And Still Keep Their Jobs:

> As the 2016 presidential campaign should have conclusively demonstrated, this pretense of expertise is a fabrication. Far from being especially prescient about matters of public affairs, members of the Pundit-Commentariat Industrial Complex are actually incredibly ill-suited to the task of accurately gauging the political sentiments of their own nation. By virtue of the various self-destructive pathologies that perpetually dull and distort their analytical acuity, it turns out that "pundits" are among the least qualified to accurately predict how far-off events will unfold. Surveying a random selection of Twitter trolls would probably yield one better information than scanning the output of the most revered professional prognosticators […]

> For normal people, even the tiniest mistakes often result in drastic consequences. They don't just get to ignore those failures and barrel forward as if nothing happened. And yet that's how we permit the pundit class to operate. In the case of Bouie and Beutler, it wasn't merely that they made erroneous predictions; anyone can mistakenly guess that something might pan out, when it does not. Rather, their entire analytical framework was drastically, catastrophically faulty. If any other American worker had performed his or her job so poorly, they could expect to receive severe sanction—docked pay, unfavorable scheduling, or termination. But in the world of punditry, there is no price to pay for failure. Instead, the American pundit class simply carries on as before, rattling off self-assured predictions about future events.

It would be really fun if I could dramatically reveal that (shock! horror!) Michael Tracey has himself been wrong about things. Alas, he admits it, saying in an earlier article, We Must Demand Pundit Accountability, that he's made some predictive mistakes himself. For example, he wrote about Why Ted Cruz Could Win In 2016, how Chris Christie Isn't Dead Yet and Why Jim Webb Poses The Biggest Threat To A Hillary Clinton Presidency. He asks to be judged not on these isolated mistakes, but based on his record as a whole. He provides a (self-curated) list of accurate predictions, which indeed seems very impressive.

Likewise, *Current Affairs*, which published Tracey's article, has admitted that its article saying "good riddance" to Trump since he "will not be president" was a bit premature. But once again, they plead that instead of dismissing them the same way they recommend we dismiss other failed predictors like Paul Krugman and Matt Yglesias, we take into account that they also made a bunch of much better predictions, like this one in February predicting that Trump would win unless the Democrats nominated Sanders. I think it's a good piece and proves that good punditry is indeed important; if people had listened to that maybe we'd be in a better place right now.

But there's still a tension between their treatment of other pundits' mistakes (proof that they're incompetent and that the whole system must be burned to the ground) and that of their own mistakes (worth viewing in the context of a long-term record of good predictions). Might Paul Krugman and Matt Yglesias also believe they have a long-term record of good predictions? Don't they deserve to be judged on this record instead of on a single event where they missed the mark by barely 1% of voters?

I don't know much about Yglesias' record, but I can speak up for Krugman. A team from Hamilton College analyzed the predictions of various pundits over sixteen months to evaluate relative performance; Krugman was judged most accurate of all twenty-six pundits studied.

The moral of the story is *stop trying to draw sweeping conclusions from one data point*. This also solves the problem where, having discredited everyone who predicted a Hillary victory, we determine the only trustworthy sources of political commentary to be PrisonPlanet.com, the Dilbert guy, and all 372,672 subscribers of r/the_donald.

If you're really interested in well-founded judgments of your own accuracy relative to other people, there's an established way to make that happen. Make specific predictions, which are clearly flagged as predictions and can't be disavowed later. When possible, try to predict the same events as other pundits, so that you can compare accuracy. Assign a probabilistic confidence level to each. Keep track of whether each did or didn't come true. Use some kind of scoring rule to evaluate your calibration. Then report on long-term aggregated statistics of how well you did.

I’ve been doing this for the past three years (2014, 2015, 2016). Last January, I predicted an 80% chance that Trump would lose. He didn’t. Does that mean I’m incompetent person who deserves to lose his job but won’t because he has “pundit tenure”? I don’t think so. Over the past three years I made 37 predictions that something would happen with 80% chance, and of those, thirty (81%) did happen. In other words, over the long run, the things I say have a 80% chance of happening, happen 81% of the time. I have pretty close to the exact right level of certainty in everything I say.

Of course, life would be even better if I could be 100% sure about everything and be right 100% of the time. And the great thing about this methodology is that if there’s someone else like that, they can prove that they’re better than I am. In fact, we’re trying this – over on Arbital, about a hundred people have entered predictions on the same set of sixty-one events that I did. At the end of the year we’ll check results. If other people do better than I do based on something like a Brier score, and if they can keep doing better than I do consistently, I’ll admit they’re a better “pundit” than I am and defer to their expertise.

If Robinson and Tracey want to demonstrate to the world that they are trustworthy pundits in a way that Yglesias and Krugman aren’t, I would invite all four of them to formally keep track of their predictions and see how they do relative to one another. I’m happy to help with this if they’re interested, and I bet Arbital would be too.

6.1. Pundits Who Failed To Predict Trump, Because They Are Out Of Touch With Real Americans

I think the argument is supposed to be that if they had ever left their comfortable Beltway offices and gone to talk to real people in the Midwest, they would have recognized the deep vein of anger in the American people and known that Trump was going to win.

Whoever you are, my “talking to real people in the Midwest” credentials are better than yours. I am a psychiatrist. I work in Michigan. My job is pretty much talking to former industrial workers about all the ways their lives have gone wrong, eight hours a day, every day. I am aware that these people are very angry.

But is it the level of anger where 46% of them will vote Trump? Or the level of anger where 48% of them will vote Trump? Because Hillary got about 47% of the vote in Michigan, so those two points are the difference between Trump winning the state and becoming President, versus losing the state and fading into ignominy. I do not think there is any level of deep connection to the collective consciousness of Michigan that allows you to distinguish between a 48%-Trump level of anger versus a 46%-Trump level of anger. Which means that even if you psychoanalyze Michiganders eight hours a day you still have to read the polls like everyone else. And the polls said that it was more like a 46% level of anger. And they were wrong.

But shouldn’t people who left their Beltway offices have at least realized that there was a *significant* amount of anger in the American people, and so Trump had a fighting chance? Yes. But all the polls also showed that there were a lot of Trump voters and that he had a fighting chance. If you were so confused that you didn’t realize that lots of people were angry and Trump had a fighting chance, I’m not sure that leaving your Beltway office would have helped much. In fact, I’m glad you didn’t. You probably would have wandered dazed into the street and gotten hit by a truck or something.

(or, if you made it to the Midwest, grain entrapment)

I guess there’s a version of this argument I endorse, which is that people who left their Beltway offices and talked to Real Americans might have realized that Trump voters were human beings with legitimate concerns and not just all alt-right Nazi KKK members. But again, if it takes a round-trip ticket to Peoria to convince our elites that people who disagree with them are not inscrutable hate-filled monsters, we have failed in a way more profound than not giving them that round-trip ticket.

7: People Who Are Worried That The Russians Hacked The Democrats To Influence The Elections

“Can you believe that the Democrats are trying to spin a narrative about foreign bogeymen out to get us?”

Okay, but did you look through the evidence that Russia was involved in the hacking? And don’t you agree it’s pretty strong?

“Yeah, but remember when the Republicans were the party of McCarthyism? And now this is totally the same thing!”

Okay, but did you look through the evidence that Russia was involved in the hacking? And don’t you agree it’s pretty strong?

“And just think, the CIA getting all upset about foreign powers interfering in an election! Pretty hypocritical, huh?”

Okay, but did you look through the evidence that Russia was involved in the hacking? And don’t you agree it’s pretty *strong?*

“And Hillary Clinton was such a terrible candidate, I bet it feels pretty good to be able to just blame everything on the Russians instead of admitting that you goofed by nominating her.”

Okay, but did you look through the evidence that Russia was involved in the hacking? And don’t you agree it’s pretty strong?

“There was that one guy on Twitter who posted a really cringeworthy rant about ‘game theory’. Can you *believe* that weirdo?”

Okay, but did you look through the evidence that Russia was involved in the hacking? And don’t you agree it’s pretty strong?

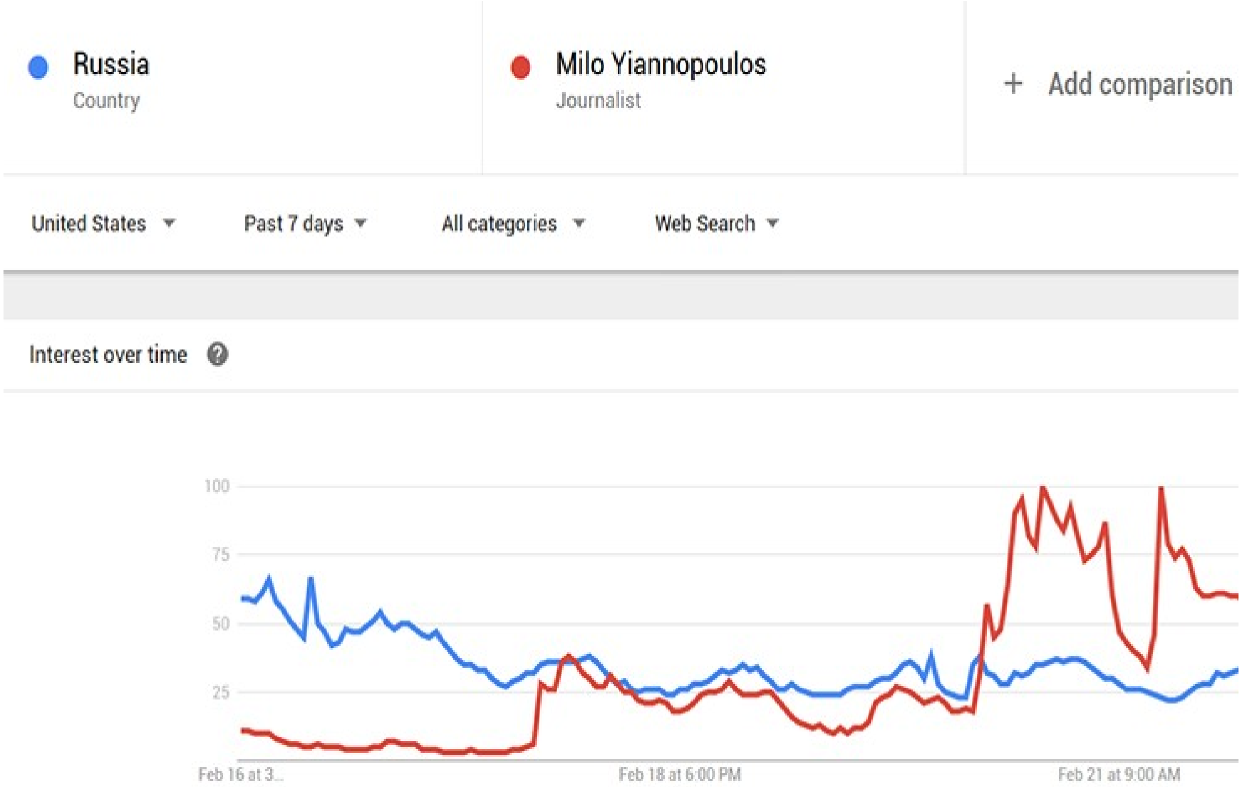
“Did I mention how funny it was that now the DEMOCRATS are the party of McCarthyism! Oooh, bogeyman Putin out to get you!”

Okay, but did you look through the evidence that Russia was involved in the hacking? And don’t you agree it’s pretty strong?

“Look, lay off, I’m not saying it’s false, I’m just saying we have more important things to talk about.”

And yet I checked your Twitter feed, and every tweet for the past two weeks has been you making fun of that game theory guy.

“I’m just saying that we’re focusing on Russia to the exclusion of everything else. Could there possibly be anything more pointlessly distracting from the real work that we’ve got to do?”



● Russia
Country

● Milo Yiannopoulos
Journalist

United States ▾ Past 7 days ▾ All categories ▾ Web Search ▾

Interest over time

+ Add comparison
100
75
50
25
Feb 16 at 3...
Feb 18 at 6:00 PM
Feb 21 at 9:00 AM

SSC Journal Club: Analytical Thinking Style And Religion

[Content warning: religious people might feel kind of like this objectifies them and treats them as weird phenomena to be explained away.]

A major theme of this blog is: why do people disagree so intractably? And what can we do about it? Yes, genetics is a big part of the answer, but how does that play out in real life? How do those genes exert their effects? Does it involve human-comprehensible ideas? And how do society's beliefs shift over time?

Gervais and Norenzayan (from here on "G&N") write about how Analytic Thinking Promotes Religious Disbelief. They make some people take the Cognitive Reflection Test (CRT), a set of questions designed so that intuition gives the wrong answer and careful thought gives the right one. Then they ask those people a couple of questions about their religious beliefs (most simply, "do you believe in God?"). They find that people who do better on the CRT (ie people more prone to logical rather than intuitive thinking styles) are *slightly* less likely to be religious. In other words, religion is associated with intuitive thinking styles, atheism with logical thinking styles. I assume Richard Dawkins has tweeted triumphantly about this at some point.

Then they go on to do a couple of interventions which they think promote logical thinking styles. After each intervention, they find that people are more likely to downplay their religious beliefs. In other words, priming logical thought moves people away from religion.

If this seems fishy to you, it seemed fishy to the Reproducibility Project too. They ordered a big replication experiment to see if they could confirm G&N's study. The relevant paper was just published on PLoS yesterday, and we can all predict what happened next. Let's all join together in the Failed Replication Song:

Actually, it's a little more interesting than this. Let's look at in in more depth.

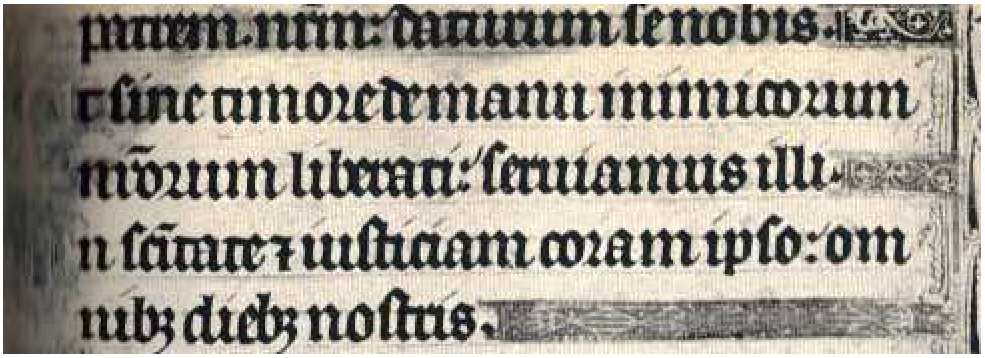
G&N bundled five different experiments into their original paper. Study 1 was the one I described above; make people take the CRT, elicit their religious beliefs, and see if there's a correlation.

Study 2 "primed" people (n = 57) by making them look at one of two sculptures; either Rodin's *The Thinker*, or a classical Greek sculpture of an athlete. Their theory was that looking at *The Thinker* would prime analytic thought, and they "proved" it in a pilot study where it improved performance on a syllogistic reasoning test. Then they deployed it on the question at hand; people who saw a picture of *The Thinker* were less likely to admit belief in God than people who saw the athlete sculpture, p = 0.03, with a respectable effect size of d = 0.6. Therefore, priming analytical thought decreases belief in God. There is no YouTube video that can express my opinion of this study, and I am not even going to try to find one.

Study 3 tried something similar. They made people (n = 93) unscramble words, (the example they give is "high the flies plane" to "the plane flies high"). Some of these words were things about logical thinking, like "analyze" and "ponder". Again, this was supposed to prime logical thought. Again, they did a pilot study to confirm that it worked (or maybe just because all those sentences about planes flying high had primed their brain to think about pilots). Again, when they made subjects do it and then assess their religious beliefs, they admitted slightly less belief in God, p = 0.04, d = 0.44.

Study 4 was a replication of Study 3 with a larger and more diverse sample (n = 148) and broadly similar results.

Study 5 was another variation on the same theme. Previous studies had shown that hard-to-read fonts prime analytical thinking, probably because they require lots of cognitive effort and time, so you're already activating effortful parts of your brain instead of just making a snap judgment. So they asked people to rate their beliefs in God using two questionnaires; one in an easy-to-read font and one in a hard-to-read font. The people who got the hard-to-read questionnaire reported less religious faith (p = 0.04, d = 0.3)



patrem.nrm:daturum ſe nobis.
c ſine timoꝛe de manu inimicoꝛum
nr̃oꝛum liberati: ſeruiamus illi.
n ſcītate ⁊ iuſticiam coꝛam ipſo: om
nib3 dieb3 noſtris.

Medieval Bibles looking like this probably caused the Enlightenment

The Reproducibility Project effort completed replications of Study 1 (the direct CRT/religion correlation), and Study 2 (the sculpture prime).

Their replication of Study 1 used 383 people (they were aiming for 2.5x the size of the original study for statistical reasons I don't entirely understand, but fell very slightly short). It was essentially negative; on two out of their three measures of religion, there was no significant rationality/atheism correlation, and on the third it was much smaller than the original study, so small it might as well not exist. They nevertheless declined to publish these results for two reasons. First, because they were a “conceptual” rather than “direct” replication; they switched from the CRT to a slightly different test of reasoning ability because everyone on Mechanical Turk already knew the CRT (!) Second, because “subsequent direct replications of this correlation have pretty conclusively shown that a weak negative correlation does exist between these two constructs”.

I am really confused by this second point. Everyone else has found that there’s a relationship between rationality and atheism, your replication attempt finds that there isn’t, so you decide not to publish the replication attempt? Some might call this *the whole point of doing replication attempts*. I know that Reproducibility Project are good people, so I am just going to assume I am hopelessly confused about something.

Then they move on to their replication of Study 2, the one that they *did* publish. This is the one with Rodin’s *The Thinker*. Once again, they used a sample 2.5x the size of the original, in this case 411 people. This time it was a “direct replication” with everything done exactly the same way as G&N (they used college students to solve the MTurk saturation problem). They found no effect of sculpture-viewing on religion, p = 0.38, h^2 = 0.001. Of note, and *really cool*, they confirmed the quality of their study by simultaneously testing the same sample for an effect they knew existed, and finding it at the level it was known to exist. They did a bunch of subgroup analyses and adjustments for confounders, and none of them did anything to recover the effect found in the original study.

The Reproducibility Project doesn’t get around to replicating studies 3, 4, or 5. But Studies 3 and 4 have been investigated by a different group in a slightly different context (CRT on liberal/conservative) and they find that the prime doesn’t even work; people who do the rational word scramble task don’t do better on the CRT. And the effect used in study 5 has been spectacularly falsified by sixteen different replication attempts – that is, hard-to-read fonts don’t even make you more rational at all, let alone make you less religious because of that increased rationality. Maybe it’s time for another Traditional Social Psychology Song:

So, why is this interesting? Seven bajillion vaguely similar priming-related studies have failed replication before. Now it’s seven bajillion and one. Can’t we just sing the relevant snarky songs and move on?

Probably. But my problem is that I keep trying to maintain these lines between studies where I know where they went wrong, and studies that seem like the sort of thing that shouldn’t go wrong. And when I see studies that I think shouldn’t go wrong, go wrong, I like to take a moment to be suitably worried.

My usual understanding of why these sorts of studies go wrong is a combination of overly complicated statistical analysis with too many degrees of freedom, unblinded experimenters subtly influencing people, and publication bias.

These studies don’t have overly complicated statistical analysis. They’re really simple. Do a randomized experiment, check your one variable of interest, do a t-test, done.

And these studies don’t have a lot of opportunity for unblinded experiments to subtly influence people. The whole thing was done online. That seriously dampens the opportunity for weird Clever Hans-style emotional cues to leak through.

That leaves publication bias. I said the original paper contained five different studies, but for our purposes that isn’t true. If you count the pilot studies, it actually included seven. A brief description of the two pilots, summarized from the supplement:

Pilot 1: 40 people see either *The Thinker* or the athlete sculpture, then are asked to solve a series of syllogisms where the intuitively correct answer is wrong. The people who saw *The Thinker* got more answers right, p less than 0.01, d = 0.9.

Pilot 2: 79 people unscramble either words relating to reasoning, or words not relating to reasoning. They are then asked the trick question “According to the Bible, how many of each kind of animal did Moses take on the Ark?” (I didn’t know this was an Official Scientific Trick Question when I wrote about Erica using it in Chapter 5 of Unsong). The people who unscramble rationality-related words are more likely to get the correct answer, p = 0.01.

So *how do you get publication bias on seven different but related experiments performed in the same lab?*

That is, if there’s a 5% chance of each experiment coming out positive by coincidence, then the chance of all of them coming out positive by coincidence at the same time is 0.05^7 = about one in a billion.

Yet the alternative – that these people performed a hundred-forty different experiments and reported the seven that worked – isn’t very plausible either. In particular, consider the two studies that combined a pilot with a main experiment. Unless there wasn’t even a pretense of doing anything other than milking noise, this *had* to be a single pilot study, followed by a single main experiment, with both of them being positive. And again, the chances of this happening by coincidence are really low.

So what happened? A commenter brings up that they used different measures of religious belief in each study, for unclear reasons. Is it possible that they used all three of their measures for everyone, and took whichever worked?

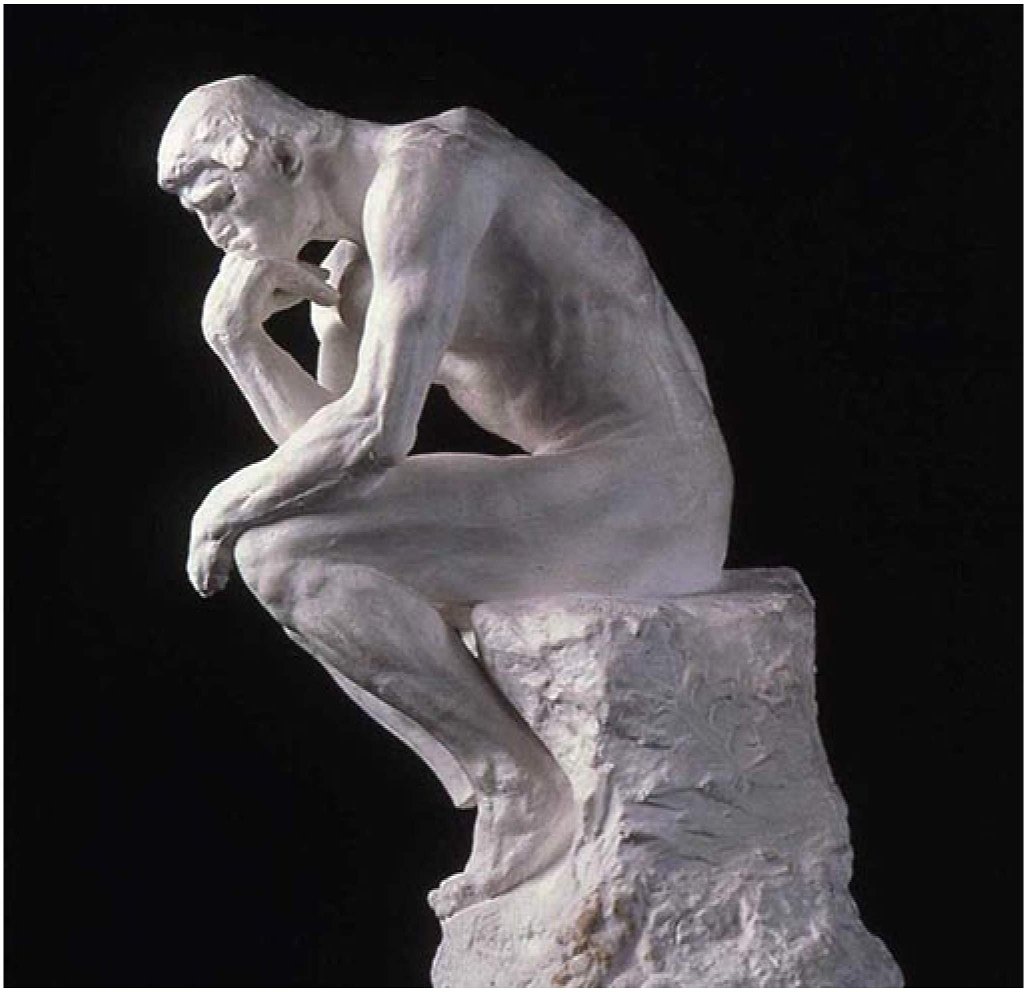
I’m not sure. And this has cemented something I’ve been thinking a lot about lately – a move from “this study’s probably not flawed because X” to “I should always be concerned that studies may be flawed, until they replicate consistently”. Probably there are some people who know enough statistics that all of these patterns make sense to them. But if you’re at my level, I would recommend against trying to play along at home.

Or as someone on Twitter (sorry, I lost the link) put it recently: “Peer review is a spam filter. Replication is science.”

II.

There’s a loose end here which deserves some attention – the Reproducibility Project’s claim that “subsequent direct replications of this correlation have pretty conclusively shown that a weak negative correlation does exist between [reasoning ability and low religious belief]”.

I want to talk a little about these other studies. This is going to be kind of politically incorrect – it’s always sketchy to say science has proven that people only believe certain things because they’re irrational. So in order to keep tempers low and maintain the analytical frame of mind we need to deal with this logically, please stare at this picture of *The Thinker* for thirty seconds.



Done? Good. Pennycook et al (2016) does a meta-analysis of all the work in this area. He finds thirty-five different studies totaling over 15,000 subjects comparing CRT scores and religious beliefs. Thirty-one are positive. Two of the remaining four detected an effect of the same magnitude as everyone else, but didn’t have enough power to prove it significant.

The remaining two negative studies are delightful and deserve to be looked at separately.

McCutcheon et al’s is titled Is Analytic Thinking Related To Celebrity Worship And Disbelief In Religion?. Unsatisfied with just asserting that irrational people become religious, they expand the claim to add that they become the kind of person who’s really into celebrities. They do manage to find a modest link between irrationality and score on the “Celebrity Attitudes Scale”, but the previously-detected irrationality-religion link fails to show up. This is a little worrying because it’s a paper that got published on the strength of a separate finding (the celebrity one) and incidentally failed to find the religion link, which means it’s a rare example of something being publication-bias-proof.

The other one was Finley et al’s Revisiting the Relationship between Individual Differences in Analytic Thinking and Religious Belief: Evidence That Measurement Order Moderates Their Inverse Correlation. They find that if you measure rationality first and then ask about religion, more rational people are less religious, and theorize that doing well on rationality tests primes irreligion. But if you measure religion first and then ask about rationality, there’s no link. Among 410 people, those in the CRT-first condition produced a rationality-atheism correlation significant at $p = 0.001$; those in the religion-first condition got nothing, $p = 0.60$. I don’t see a direct comparison, and the difference between significant and nonsignificant isn’t necessarily itself significant, but just by eyeballing this is obviously a big deal. This is *also* worrying, because it’s *another* example of a study that found an exciting finding and so got published despite failure to replicate the result at issue.

But Pennycook responds by pointing out seven other studies in his meta-analysis that ask for religion before testing rationality yet still get the predicted effect. In fact, overall there is no noticeable difference between religion-first studies and rationality-first studies. Some others assess rationality and religion on different sittings, and *still* get the same results. Also, now it looks like priming doesn’t affect your religiosity *or* rationality. So Finley’s paper *has* to be wrong, which means it’s yet *another* example of strong p-values in a large sample size in the absence of any real effect.

At this point we’re left with 31 good studies finding an effect and 2 good studies not finding it. Most of them converge around an effect size of r = – 0.20. Pennycook does the usual tests for publication bias, and as usual doesn’t find it. I think at this point maybe we can conclude this is real?

A few other things worth looking at:

Is this effect true only in college students and mechanical Turkers? No. Browne et al look at 1053 elderly people’s CRT scores and religiosity, and find the effect at the same level as everyone else.

(they also find that women do much worse on the CRT than men. I looked to see whether this is a common finding, and indeed it is; in a sample of 3000 people taking a 3-question test, men average about 1.47 and women about 1.03, p < 0.0001. This remains true even when adjusting for intelligence and mathematical ability. I'm not sure why I've never seen any of the sex-differences crowd look into this seriously, but it sounds important. If you have a strong opinion about this, please stare at the above image of *The Thinker* for another thirty seconds before commenting)

Is this effect simply an artifact of IQ? After all, there’s some evidence that IQ increases irreligion, and CRT score correlates heavily with IQ (see eg this book review). This is the claim of Razmar & Reeve, who do a study that finds that indeed, it’s not that more rational people are less religious, it’s that smarter people are both more rational *and* less religious. But Pennycook responds with a boatload of research finding the opposite; the gist seems to be that both IQ and CRT are independently correlated with irreligion, but the CRT correlation is stronger than the IQ one. Trying to tease apart the effects of two quantities that are correlated at 0.7 sounds really hard and I am not surprised that people can’t figure this out very well.

(This paper brings up another interesting fact – on a lot of these tests, religious people take less time to solve problems, even when both sets of people get the right answer. This reminds us that high-CRT shouldn’t be considered strictly *better* than low-CRT in the same way that high-IQ is strictly better than low-IQ. It’s more like tradeoff between System 1 fast and heuristic-laden thinking, vs. System-2 slow and deliberative thinking. This tradeoff seems to exist, at different points in different people, regardless of their IQ.)

The paper says:

> Importantly, the degree to which cognitive ability versus style are predictive of religiosity has theoretical consequences. As discussed by RR, a primary relation with cognitive ability is consistent with the idea that people naturally gravitate toward ideologies that match their level of cognitive complexity. Thus,according to this position, religious ideologies are less complex than secular ones, and, as a consequence, more likely to be held by less cognitively complex individuals. In contrast, a primary relation between cognitive style and religiosity is consistent with the idea that Type 2 processes are selectively activated by religious disbelievers to inhibit and override intuitive religious cognitions. Importantly, under this formulation, religious

> disbelief does not necessarily require a high level of cognitive ability

.

III.

Overall my takeaway from reading some of this stuff is:

1. “Analytical cognitive style”, ie the slow logical methods of thinking that help you do well on the CRT, probably increases likelihood of being an atheist and decrease the likelihood of being religious, even independent of IQ with which it is highly correlated. The effect size seems pretty small.

2. IQ probably also increases likelihood being an atheist and decreases likelihood of being religious, even independent of CRT with which it is highly correlated. The effect size seems very small.

3. Openness To Experience probably has complicated effects that make people less fundamentalist but more spiritual.

4. These are all long-term trait effects. There’s no good evidence that “priming” analytical thinking style can make you more or less religious in that exact moment. Probably the effect size here is zero.

5. Gender differences on the CRT are higher than gender differences on almost any other test and this seems to be underexplored.

6. Just because a paper has relatively simple statistics that are hard to fake, doesn’t mean it’s likely to replicate.

7. Even given (6), just because a paper is an online survey with little room for experimenter effects, doesn't mean it's likely to replicate.

8. Even given (6) and (7), just because a paper has many different studies that all confirm the same effect, doesn't mean it's likely to replicate.

Extra bonus takeaway: I was too quick in dismissing the CRT's ability to convey extra interesting knowledge beyond IQ, and I should look into it more and maybe get Stanovich's book. Also, there's some similar research on CRT and politics which I should probably look into, although I don't even *know* how long I'm going to have to stare at that *Thinker* picture for *that* one.

Extra extra bonus takeaway: I should include CRT on next year's SSC survey.

A Modern Myth

1. Eris

A middle-aged man, James, had come on stage believing it was an audition for *American Idol*. It wasn't. Out ran his ex-lover, Terri. "You said you loved me!" she said. "And then when I got pregnant, you disappeared! Twenty years, and you never even sent me a letter!"

The crowd booed.

As James tried to sputter a response, his wife ran onto the stage. "You cheating jerk!" she shouted at James. "You lying, cheating jerk! Twenty-five years we've been married, and I never…" She picked up a folding chair, tried to swing it at James.

"Stop!" cried James' teenage daughter Katie, joining in the fray. "Mom, Dad, stop it!"

"You stay out of this!" shouted James' wife. "Maybe if you'd had a good male role model, you wouldn't have become *a lesbian*."

The crowd gasped.

Katie's girlfriend Lisa came out of a side door. "You take that back!" she yelled. Then she saw Terri. "Wait? Mom? What are you doing here?"

"That's right," said Alice DiScorria, sidling onto the stage, effortlessly drawing the audience's attention from the brawl taking shape in front of them. "Katie's girlfriend is the daughter of the woman her father cheated with, so many years ago. And we've got the paternity test right here." She theatrically opened a manilla envelope. “And…James! James is the father!”

“I’VE BEEN LESBIAN LOVERS WITH MY HALF-SISTER!” shrieked Katie.

“This is all your fault!” everyone shouted at everyone else in unison. Then the punching started.

In short, it had been another successful episode of *The Alice Show*.

Now Ms. DiScorria was in her dressing room, wiping off the night’s makeup, trying to decide where to go to dinner. Knock, knock. She opened the door wide.

There stood Katie and Lisa. Katie was holding a shotgun.

“Why would you do this to us?” screamed Katie. “We were a happy family!”

“I loved her!” added Lisa.

“Why?” Katie screamed at her, waving the gun. “WHY?”

“Oh, put it away,” said Alice. “We both know you’re not going to shoot me. And it wouldn’t hurt me if you did. I do this because I’m Eris, the Greek Goddess of Discord. I destroy peace. I set people against each other. Then, when their petty fights destroy everything they’ve worked for, I stand over the ruins and laugh. It’s my thing. Here. Have a golden apple.”

It appeared in Alice’s hand, shining with beauty that defied description. “FOR THE FAIREST” was writ on the front in letters of liquid light. Katie dropped her gun and stared. Lisa rubbed her eyes to see if she was dreaming. For a brief moment, no one moved.

Finally, Katie asked, “You’re…*giving* it to me?”

“Absolutely. To you and your girlfriend. Traditionally, I think it would go to whichever of you is prettier.”

Gently, she placed the golden apple on her dressing table, winked at the girls, and left the room. She closed the door behind her, so nobody would hear the screams.

2. Ares

“Look,” Tom told Ari, “you always seem to come out of this kind of thing okay. So if I don’t make it tomorrow, I want you to give this to my wife.” It was an envelope. There was no address, just ‘TO BE OPENED IN THE EVENT OF MY DEATH’.

“Stop talking like that, Tom,” said Ari, taking the envelope and putting it into his backpack. “You’ll make it. The Taliban’s gonna fold like a wet rag tomorrow, I promise.”

“Easy for you to say. In Helmand, half your squad dies, you just walk out with a big grin on your face. Kandahar, outnumbered eight to one, and not only do you win, you end up with two Medals of Honor. I didn’t even know you could *get* more than one Medal of Honor for a single battle. Yeah, sure, you’ll be fine tomorrow. The rest of us, we’re only mortal.”

“Yeah,” said Ari. “I can see how that would suck.”

“Look, you’re doing me a big favor, taking that envelope,” said Tom. “Anything I can do for you? You know, in case the worst happens?”

“Nah, don’t worry about it.”

“There’s nobody back home you care about? Wife? Girlfriend? Family?”

“Fuck all. No wife, no girlfriend, and a family dysfunctional like you wouldn’t believe.”

“Where are you even from, Ari? You never talk about it.”

“Who cares?”

“I care. Heck, half the squad thinks you’re some kind of government supersoldier, the other half thinks you should be in a loony bin. You’re *interesting*, Ari.”

“Well, fine. I’m Ares, Greek god of war. I’m the son of Zeus and his sister Hera, and let me tell you, marrying your sister works about as well as you’d expect. I used to be a big deal, shape the destiny of whole nations, rise of Rome and all that. Then my power crashed along with everyone else’s. Man, I don’t even remember the Dark Ages. The whole medieval era is a blur. By the time I start feeling like myself again, it’s the Renaissance and everybody’s fighting with muskets. Nowadays…man. I can fight better than you mortals, you gotta give me that. But in terms of god stuff…I remember when I could make all of fucking Persia flee in abject surrender. Now I’m stuck taking pot shots at Taliban assholes. Meanwhile, they’re all shouting about Allah, and you guys are all shouting about Jesus, and nobody even fucking believes in me anymore.”

“I believe you,” said Tom.

“Really?”

“Yeah,” said Tom. “In Kandahar, I saw a bullet pass right through you. How would a government super-soldier manage that? Sure. You’re Ares, god of war. I’ll worship you, if you want.”

“What fucking good would that be?”

“Think about it! You said that you and all the other gods lost your power back in Roman days. What happened in Roman days? Constantine! The start of Christianity! That must have been what did it! Gods’ power comes from people believing in them!”

“Why does *every* mortal always figure that gods’ power comes from people believing in them? Like you’re all some kind of god power experts? Do they teach that to baby mortals in their little mortal schools? Stupidest thing I ever heard. You think we ruled the world for a thousand years and didn’t check where our power came from? We figured that out a long time ago. Divine power comes from meat.”

“Meat?”

“Yeah. Like, you know, sacrifice a ram to Ares, pray for victory, then eat it in a big communal feast in the barracks. The more meat sacrificed in a god’s name, the stronger he got.”

“But then it’s still about belief. People stopped believing in you, so they stopped sacrificing rams to you.”

“You’ve got it ass-backwards. We were at the height of our power. People were sacrificing rams to us right and left. Then it stopped working. One year the meat started having a little less effect. The next year it was a little less than that. Eventually it was gone. And then when the gods became powerless, the cults collapsed, and *then* the Christians and Muslims and all the rest stepped in to fill the gap."

"So what can I do? There's some meat in the mess hall, I can sacrifice that for you if you want."

"I'm telling you, it won't work. The power's gone. It's been gone for two thousand years. Me – and all the rest – we're stuck like this. Some kind of natural floor to our power, still more than mortal but forever less than divine. It's fucking awful and I hate it. I hate not being able to smite whole nations when I'm angry. I hate having to take commands from 'superior officers' because I'm 'just a grunt'. And most of all, I hate that people have *forgotten about us*. We used to be *big*, Tom!"

"People haven't forgotten. They love you guys. There's still, you know, Hermes handbags, and Athena mineral water, and, you know, Mars bars…."

"I am the lord over war, the manslaughtering one, the bloodstained one, he of many devices, bringer of much weeping, destroyer of men. I AM NOT A FUCKING CHOCOLATE BAR."

"Sorry, man. I was just saying…"

"I know. You wanted to make me feel better. That's what I've come to. Having to be consoled by mortals. You know what's going to make me feel better, Tom? Killing some fucking Taliban tomorrow."

"Yeah. Okay."

"Look, I'm sorry, I didn't meant to…"

“It’s not that. But, uh, Ari. We’ve got a big battle coming up tomorrow. And I know this probably sounds really crazy to you, but humans – praying makes us feel better. That’s why we do it all the time. To Jesus or Allah or whoever. And we don’t really expect it to work, so…um…”

“Out with it, Tom.”

“…is it okay if I pray to you tomorrow?”

“Knock yourself out, Tom.”

3. Apollo

Ianthe had always liked magic squares. They were one of the oldest forms of magic. A Sator square had been found scrawled on one of the walls of Pompeii. Since then the art had advanced, and she was its master. She would fill the word square with words relating to the sun, and Apollo would appear before her. Working with gold ink, she traced the letters carefully:

C I R C L E
I C A R U S
R A R E S T
C R E A T E
L U S T R E
E S T E E M

Apollo appeared before her, devilishly handsome, impeccably well-dressed, unfailingly polite. He’d told her once that in his other identity, he was a professor at some college somewhere. She could believe it.

“Ianthe, my daughter,” he said, his voice smooth and golden. He always called her that, even when he was doing very un-fatherly things to her. Though come to think of it, in *his* family that might not be such a jarring contradiction. She wondered if he’d known Oedipus.

“Lord Apollo,” she said. “I have called you here to request a boon.”

His face fell. He had explained the first time he met her that his powers were weak. That he couldn’t help her the way she might have wished. Couldn’t grant her wealth or wisdom or prophecy the way he might have in days of old. Since then she had never asked him for anything but himself.

“It’s…nothing too difficult,” she assured him. “Just…actually, I wanted to say good-bye.”

“Good-bye?” asked Apollo.

“There’s…someone’s hunting us. The neopagan community. I told you about Megan, right? The girl who used to run a traditionalist group up in Santa Cruz? They found her dead two weeks ago. There’s a Wiccan circle over in Oakland that deals with Greek themes sometimes; two of their leaders have been missing since January. And then Aristopsychus the Wise…that’s what he calls himself, seriously, one of the crazy sorts who attacks people drinking Athena Mineral Water and says they’re profaning the name of a goddess…I just got a call. His head was bashed in last night. I’m really afraid, Lord Apollo.”

He looked at her, his face infinitely wise and sorrowful, and she knew he could do nothing.

“That’s why I’m leaving here,” she said. “I haven’t told anybody, nobody else in the neopagan community, not even that weird girl Emily who thinks she’s my ‘acolyte’. I’m shutting down the temple and going somewhere really far away where nobody can find me, and I don’t know when I’ll be able to summon you again.”

“I understand,” said Apollo. “May good fortune go with you.”

“But I was asking you for a boon. I need you to take something from me.” She took out a paper bag and produced an apple, brilliant gold, shining with an unearthly radiance. Apollo stepped back as if someone had struck him, his calm manner broken for the first time Ianthe had seen.

“Where did you get that…that *thing*?” asked Apollo.

“A detective gave it to me! He was investigating a crime scene. These two girls murdered each other in a Hollywood studio dressing room…it was all over the news. And this guy was called in to investigate the crime scene, and all he could find was this golden apple that said “for the fairest”. And after all the legal things were closed he didn’t want to throw it out, because it looked so pretty, and he heard about our temple here, and he figured it looked like something a Greek pagan revivalist movement should own, so he gave it to me. And as soon as I saw it…this sounds so bad, but I didn’t tell any of the others, not even Emily. I brought it home and never told anyone about it. But I’m scared, Apollo. I’m scared it has something to do with the reason all this is happening. I don’t want to leave it here and I don’t want to take it with me, so…please, just take the apple. Before it makes me change my mind!” She wasn’t looking at it; she was *carefully avoiding* looking at it.

“I can’t,” said Apollo.

“You have to!” said Ianthe.

“No, I mean, literally, I can’t,” said Apollo. “The apple has to belong to a woman. Any woman who sees the apple, she wants it more than she’s ever wanted anything else. Any man who sees it, no effect. Even if a man gets it, he feels compelled to give it to a woman. That must have been what happened to your detective. I really, *really* do not want that apple. You have no idea how bad things get around one of these.”

With a grunt and an effort of will, Ianthe threw the golden apple at Apollo’s face. He caught it in his hand reflexively, involuntarily. “Take it!” she said, as he stared at his unwanted prize. “You’re a god! I’m sure you can think of a woman who can keep it safe for you!”

“Ohhhhh….this is *not good*,” Apollo said, through clenched teeth. “I hate these things, I hate these things, I hate these things, I hate these things…”

Ianthe erased a letter from the middle of the magic square, and Apollo disappeared. Then she picked up her suitcase, got in her car, and started driving, intent on putting as much distance as possible between herself and anywhere people would be looking for her.

4. Aphrodite

She stays by the sea shore. Shining shells and soft surf sounds surround her shelter. Cythera simmers with summer, and seals swim in the sun. Songbirds circle in the sapphire sky, and sea stars sit semi-submerged in the sand.

Ares wades out to the cliff where he knows he’ll find her, a cliff of soft pink rock that looks like any other on this side of the island. On a little depression in the rock which only he can see, he traces letters with his fingers:

MOAN
ONCE
ACME
NEED

The cliff opens around him, and he is in the bower of Aphrodite.

She is naked. Her body glistens with sea-foam. She is behind a glass shelf filled with seashells, and from where Ares is standing, two of them perfectly cover her breasts. On the near wall are pictures of her family: her husband, Hephaestus; her son, Eros; her parents, the sea and the blood of Uranus; her nth-great-grandson, Julius Caesar. On the far wall is a banner reading “UNDEFEATED GOLDEN APPLE WINNER, 1200 BC – PRESENT”, and several oak barrels overflowing with golden apples that cast an unearthly glow all over the room.

“Hello, sexy,” she says.

He tries to play it cool, act natural. “Hey Aph,” he says. “Just dropping by.”

There is no sign of her husband.

“Come on, Ares. You never ‘drop by’. What is it *really*?”

“Um,” says Ares. He is acutely aware of the god-sized erection he probably has right now. He keeps his eyes fixed on the barrels of golden apples, so as not to stare. “Um,” he says again.

“I heard about what happened in Kandahar,” she said. “That was *very heroic* of you.” She gently brushed her arm against his.

“Um,” said Ares. “That’s…kind of…look. This soldier guy I knew. He asked me if…if there was anyone back home I cared for. And I said no. Fuck everyone. You know. Mom, Dad, fuck them all. But then I started thinking. We had something good. A long time ago. And I was thinking, maybe…”

“But Ares,” she said, biting her lip, “you know I’m married.”

“You were married the last five times too,” Ares said, forcefully now. “It’s kind of a big part of having an affair.”

“But,” she said, running a hand through her golden hair, “what if people found out?”

“People found out the last five times too,” he said. “Nobody thought anything of it. You’re the goddess of love. Of lust. Love and lust. Of course you have affairs.”

“What if my parents knew? It would break their heart.”

“Your parents are the sea, and the blood that came out of a guy’s scrotum when my grandfather castrated him. I think they’ll be fine.”

“Oh, Ares. You know so much about me.”

She pulled him closer. She closed her eyes. His lips touched hers. Then –

“We can’t do this, Ares. We’re just too different, you and I. Love. War. It wouldn’t work.”

“We are *not* different. All’s fair in both of us, for one thing. We’re both, uh, relationships between two parties. Often involving fighting. More fun when you’re high testosterone. And when you’ve got a big spear.”

“I love it when you talk dirty to me, Ares,” said Aphrodite, and put her hand around his waist. He tried to kiss her a second time.

“No,” she said suddenly. “I can’t. What about the children?”

“Your child is Eros!” protested Ares. “How is that a problem?!”

“Show me you care,” said Aphrodite.

“I care!” said Ares. “I promise you, I care. Tom – this soldier I know – he was telling me all about his wife, and how much he loved her, and I was thinking, I need something like that, and then I remembered – I’ve got that. You’re the one for me. You’re the only one I want. I promise.”

“Show me,” said Aphrodite.

“How?” asked Ares. “What can I do to show you that I care?”

Aphrodite let her hand linger on his shoulder, then walked to the other side of the room. She picked up a golden apple.

“There’s another golden apple in the world now,” she said. “I can *feel* it, Ares. That apple is *mine by right*.” For a second, all the softness disappeared from her face, and he knew why one of her epithets was ‘the warlike one’. “I want that apple, Ares. Bring it to me.”

“But baby, you already have like a million golden apples. Look, you’ve got barrels full of them. You’re not even using – ” He picked up a golden apple that had fallen behind one the barrels.

“It’s the *principle* of the thing, Ares. It says ‘for the fairest’. Am I not the fairest? Have I not been the most beautiful of goddess and women since before Paris was a glimpse in his mother’s eye? Somebody else has my magic apple, Ares, and I am *literally shaking* here. You are my protector, the hero of Kandahar, the man who got two Medals of Honor in the same battle. Can you rescue me?” She knelt before him. There were tears coming out of her eyes. She hugged his leg.

“I’ll…I’ll get you your apple, Aph. I’ll find whoever’s got your golden apple, and if they don’t give it back, I’ll…” He took out his sword and swung it above him, so fast that it whistled in the air like the note of a lute.

“I’ll be waiting for you…” whispered Aphrodite.

Ares turned to go. The cliff face opened in front of him. The birds were still singing, and dolphins leapt for joy in the melodious waves. He was kind of a chump, but he knew this was the way of things, and it would never change.

“…I won’t be wearing any clothes.” Aphrodite called after him.

5. Hermes

He is called Herman. He runs a hedge fund. He lives in Manhattan. He wears nice suits.

Today he is in a nice suit, but he is not in Manhattan. He is in Memphis, Tennessee. Not even the nice part of Memphis, Tennessee. He’s in a poor, crime-ridden ghetto in Memphis Tennessee, and it has a bridge, and he is underneath it.

He spots a big man sleeping underneath the bridge, wrapped in a ratty blanket. His beard is unkempt, and even from far away, he smells like alcohol.

“Hi Dad”, said Hermes.

“Whaddyawant?” mumbled Zeus.

“It’s me, Dad,” said Hermes. “Hermes.”

With some effort, Zeus brought himself into a sitting position, brushed some of the more egregious twigs out of his beard. He rubbed his eyes.

“Yeah, so? Whaddyawant?”

Hermes inspected the King of Gods and Men. He was streaked with dirt. He was dressed in a fading white wife-beater, with reddish stains that Hermes hoped were wine.

“I’ve been looking all over for you, Dad. You look terrible. What happened to you?”

“Whaddyoucare?”

“You used to be King, Dad!”

“I’m still king. Iduncarewhatchy’all think.”

“But what happened to you? I talked to Ares the other day. He won two Medals of Honor, did you hear? Apollo’s got tenure at Oxford. I’m the god of commerce and crime, so of course I’ve got a hedge fund. But you? What *happened* to you?”

“Fucking child support payments!” said Zeus. “I was doin’ just fine for myself until cops from forty-seven different states came my front door calling me a deadbeat dad!”

“Oh dear,” said Hermes. “Forty-seven women?”

“Forty-seven *states*,” said Zeus. “Hundred ninety women. Two hundred five kids. Fucking mess.”

“A hundred ninety women,” mused Hermes. “*Please* tell me you didn’t turn all of them into animals.”

“Are you fucking *kink-shaming* me?” said Zeus. “If I get off on having kids with women and then turning ’em into animals, that’s my private business. Ain’t no weirder than Ganymede being gay or your kid who’s a futa or… BLAAAAAARGH”. He turned and vomited the morning’s meal into the river. “Besides, I don’t got power anymore. Can’t even turn a pretty girl into an ape these days, forgeddabout a cow or a bear.”

“Look, sorry for bringing up your fetish,” said Hermes. “I didn’t know it was a sore point. I wanted to talk about something important. Dad, I’ve figured it all out.”

“You figgered what out?”

“All of it. What happened to us. Why we lost our power. *And how we’re going to get it back*.”

“Yeah?” said Zeus. He sounded skeptical. “I’m listenin'”

“Look,” said Hermes. “How did we used to get power? Animal sacrifice. And which animal? Rams. What astrological age was it? The Age of Aries, the sign of The ram. 2000 BC to 1 AD, or thereabouts. Then the age changes. The sun is in Pisces. Sign of the fish. Boom. Sacrificing rams no longer works. Who comes out on top? Some Israeli whose followers are all fishermen. Talk about being in the right place at the right time.”

“So yer saying, we need to get the mortals to sacrifice fish to us now, and then we’re back in business?”

“No. Because the Age of Pisces ended last century. Now it’s the Age of Aquarius. The Water Bearer.”

“So sacrifice water?”

“Well, this is where we start to have a problem. I know you have trouble remembering all your children, but perhaps you recall that a few thousand years ago, you had a daughter who happened to become the Goddess Of Wisdom, Intelligence, and Cleverness?”

“Never gonna forget *that* one,” said Zeus, rubbing his head.

“It would seem that my lovely and not-at-all-incredibly-annoying sister Athena figured all of this out about ten years before I did,” said Hermes. He reached into his pocket and took out a bottle of Athena Mineral Water. “Behold! 91% market share. Aquafina? Bankrupt. Dasani? Out of business. And here’s the best part.” He held the label up very close, so Zeus could read it. “Athena Mineral Water Customer Reward Program,” it said in small font. “Every time you drink a refreshing bottle of Athena Mineral Water, say ‘Thanks, Athena!’ in front of a registered associate, and they’ll punch your card. Collect ten punches and get a liter bottle of Athena Mineral Water absolutely free.”

“Whaddya sayin’?” asked Zeus.

“I’m saying that every day, about a million mortals are going into supermarkets, drinking water, and saying ‘Thanks, Athena!’, and each one of them is giving my beloved-and-not-at-all-aneurysm-inducing sister an amount of divine power equal to an entire animal sacrifice. I had some of my quants crunch the numbers, and right now I’m guessing she’s about twenty times more powerful than you were at your prime. *At your prime,* Dad. She pretty much has a monopoly on divinity right now. We’re really really really screwed.”

“So you gonna take all that cash you got and open up your own water business?”

“I tried. They wouldn’t even let me register it. Said it was a trademark conflict with Hermes Handbags. I got my lawyers to look up who owns Hermes Handbags, and it’s a shell corporation belonging to a consortium belonging to a Chinese group belonging to a company registered in the Cayman Islands which was set up using money from… Athena Mineral Water. Mars Bars, same thing. Zeus Cameras, likewise. And it’s worse than that. I try to find some neopagan groups, see if maybe I can get them to sacrifice a few bottles of water to me just until I can think of a solution that scales. She murdered all of them. In cold blood. Every priest or priestess who ever worshipped another Olympian. She’s boxed us in, Dad.”

“And that’s why yer comin’ to me. You want….*the power of lighning!*”

Zeus tried to stand in an imposing pose, but only succeeded in tripping on his blanket and crumpling back onto the ground.

“Dad, you can’t summon lightning anymore. You haven’t had that kind of strength for two millennia. And with the power Athena’s collected, it wouldn’t help. But there *is* something you can give me.”

“What?”

“I need to talk to Prometheus.”

Zeus managed to bring himself into an approximation of standing. “Now listen here, sonny. Maybe I ain’t much of a king of the gods anymore. I ain’t got the lightning and the thunder and all that. But lockin’ that bastard up was the best thing I ever did, and you know it, and yer not gonna take that away from me. You think yer so smart with your hedge fund, and yer money, and yer fancy East Coast suit, but I’m tellin’ you, Prometheus would *eat you for breakfast* and he wouldn’t even break out a sweat.”

“Right, Dad. That’s the thing. He’s the only one who’s smart enough to outmanuever Athena. I’m proud of my brains, but she’s the Goddess Of Wisdom, Intelligence, and Cleverness, *plus* now she’s stronger than us, and I’m not sure how to get one up on the Goddess of Wisdom, Intelligence, and Cleverness without help from someone who’s…uh…very *very* smart.”

“Prometheus ain’t just *smart*,” spat Zeus. “It’s not just that he has *book-larning*. He’s *the God of Foresight*. He sees every possible future laid out in fronna him as easily as you or I see that there blanket.” He pointed to the blanket, which was actually so dirt-covered that it was getting hard to see against the dirt below. “It took all of us together, *and* all the Giants, *and* all the Cyclopses to bring him down, and we wouldn’ta succeeded if the Fates themselves hadn’t gotten pissed with him for ruining their weaving and given us a hand. And it was Athena herself who told us that we had to bind him somewhere far away, couldn’t talk to him, couldn’t even go near him, or else he’d figure out some way to screw up all our futures just by sayin’ a couple a’ sentences to us. And all a’ you, and all the Giants, and all the Cyclopses, you all agreed, and you all gave me the key that lets you reach him, and I ain’t given that key to anyone in the past two thousand years and I ain’t givin’ it to you now and that’s final, you hear me, boy?”

“Then,” said Hermes, “I fear we are all doomed.”

“We’re fucking *gods*,” said Zeus. “We can’t die. We can’t even be contained, for long. Only gods we ever managed to lock up were the Titans in Tartarus and Prometheus in Elbrus, and that was only by all of us workin’ together, and by my power as King of Gods, and if you think I’m signin’ off to any of this, yer crazy.”

“Then we will wane,” said Hermes, “and become little better than bugs skittering beneath Athena’s feet.”

“I ain’t got much,” said Zeus, “but I beat Prometheus and no one ain’t ever going to take that away from me. Now get going, sonny boy.”

“If I do not beat Athena,” said Hermes, “you’ll never be able to turn any women into animals, ever again.”

Zeus paused, just a second, then spat. “I made my choice,” he told Hermes. “Now git!”

6. Pandora

He remembered the first time he had come here to see her. It had taken him months just to find the place. An Orthodox convent. Our Lady Of Sorrows, just outside Kiev.

He had knocked on her door. “Come in,” she’d said. She hadn’t opened the door. At the time, he hadn’t realized that was significant.

She was wearing a veil. “Dory?” he asked. She nodded slightly. “Dory, it’s Apollo.”

“You didn’t forget about me.” He couldn’t see her eyes, but she was smiling.

“Forget about you? Dory, I’ve been sending you care packages every month!”

“Oh.” A frown. “I’m sorry. I didn’t…I don’t open things.”

“Oh.”

“You understand, don’t you?”

“I can see how it would be traumatic. But…you didn’t get any of my letters?”

“They were in envelopes, Apollo. I told you, *I don’t open things*.”

“Oh,” he said. He lifted her veil, saw her face for the first time in years. “What, not even your *eyes*?”

Pandora nodded.

“The church is beautiful. It looks like a wedding cake. You haven’t even been a *little* curious what the convent you’re living in looks like?”

“I don’t do curiosity anymore, Apollo. Curiosity leads me to bad, bad places.”

“Dory.” It was worse than he’d thought. He was the god of healing, or had been. His powers were weak, but maybe he could at least do some therapy? “Dory, you did one bad thing.”

“I did *all* the bad things, Apollo. Literally. Every single one of them.”

“Okay. Be that as it may. You were tricked. Zeus played a horrible joke on you. Or he used you as a pawn to play a horrible joke on everybody. It doesn’t mean opening things is always bad, or that curiosity always gets punished. It means one stupid god played one stupid joke. Look, he could have put all the world's evils in, I don't know, his basement, and released them if and only if you *didn't* open a box. Then the lesson would have been to *always* open things. Do you see how that makes just as much sense as what actually happened."

"I'm sorry, Apollo," said Pandora. "I appreciate what you're trying to do. But it won't help."

"It might!" said Apollo. "Keep an open mind!"

"*I don't do open*."

"Ah. Right."

He sat down on the little cot. She sat down beside him.

"So now you're a nun."

"I just live here. I wandered by one day, and the sisters took me in. Said I looked like I needed help, which I guess I did. I've stayed here ever since. They say that I'm good luck to have around. Can you believe that? Me? *Good luck?*"

"They like you," said Apollo. "Anyone would."

"It's because I don't age," said Pandora. "And because I never leave my room. They assume I'm a saint or something. Praying all the time. I'm even starting to get pilgrims, if you can believe it." She waved her hand towards a table full of little knickknacks. "Gifts. The pilgrims give me gifts." She sighed.

Apollo went over to the table. A rosary. An icon, covered in gold leaf. Jesus, he assumed. A vial of holy water. "This is lovely,” he said, looking more closely at the icon. “Who gave it to you?”

“I didn’t ask,” said Pandora. “I don’t do curiosity.”

“Ah,” said Apollo. He took her hand again. “Come outside with me. I won’t make you open your eyes. I’ll lead the way. Just for a minute?”

They walked through the courtyard. A few nuns looked askance at them, but Apollo looked too poised to be anywhere other than where he was supposed to be, and they assumed he was a visiting priest or somebody and let him pass. They came to a meadow. Apollo gingerly guided Pandora to sit down on a rock, and sat beside her.

“We used to have a good thing going,” he told her.

“And now I’m like this,” she said.

“You might get better. With time.”

“I might not.”

“There’s always hope.”

“Yes, they say I closed the box just in time for that one. Strange how little I’ve gotten from it myself.”

“Kiss me,” said Apollo, on impulse.

“We had a good thing going once,” Pandora said. “That’s not me anymore.”

“It could be,” said Apollo. “Hope, and all that.”

“I don’t open things,” said Pandora. “Not even my heart.”

How many centuries ago had that been? Three? Four? They all blended together. The convent was no help either. Most places had the decency to change a little since the Renaissance. The convent looked exactly the same. Same meadow. Same courtyard. Same door. Apollo knocked. “It’s me, Dory.”

“Come in,” she said, without opening the door.

He came in, sat down on the cot. She looked the same, too. She was in a strange middle state; a human created before mortality, given all the divine gifts, to be the wife of a god. She wasn’t divine, not quite. But she wasn’t fully mortal either. A demigod, maybe.

“It’s been a while,” he said. “Five, ten years?”

“It’s been a while,” agreed Pandora.

The room hadn’t changed either, except for a few more pilgrim gifts. The rosary and Jesus-icon had been joined by enough little saints and angels to fill a heavenly choir, plus a good-sized marble statue of a woman in armor. He tried to remember if there was some female warrior-saint, but his mind came up blank. He wished he could ask Pandora, but he knew what she thought of curiosity.

“I brought you a present,” he said. “It’s a smartphone. Flip phones are on their way out. This one works without being opened.”

Pandora ran her hands along it. “It’s so smooth,” she said. “Now you can call me any time?”

“Yeah,” said Apollo. “You can call people too. If you ever feel, you know, the need to connect.”

Pandora gave him a little peck on the cheek, then slipped the phone under her bed.

“I wish I could say this was entirely a social call,” said the god, “but I’m here on business.”

There was a pause in the conversation before he realized she wasn’t going to ask what the business was.

“A friend gave me something dangerous. And I have to give it to a woman. But if the woman saw it, bad things would happen. Really, really bad things. And I asked myself, where can I find a woman who will listen when I warn her not to look at something? And, uh. I thought of you.”

“Sure,” said Pandora. “I’m happy to take your thing. Where shall I put it?”

“Uh,” said Apollo. “Somewhere where the pilgrims won’t see it. That’s important. Nobody can see it.”

“I’ll put it under my bed,” said Pandora.

Apollo handed it to her. It was heavy, and cold to the touch, and round, about the size of a baseball. She slipped it under the little cot.

“Thanks,” said Apollo.

“I’m glad I could help you with something,” said Pandora. “You’ve been so nice to me.”

“I haven’t been! I never visit!”

“You visit sometimes. The others never visit. They wish they could forget about me.”

“Um,” said Apollo. “I’m sure they meant to drop by and tell you how they’re doing.”

“It’s okay,” said Pandora. “It doesn’t matter whether I know how they’re doing or not.”

Apollo frowned. “Listen. I know you have your position in the pantheon, as Cautionary Tale Against The Dangers Of Excessive Curiosity. But I have my place too. Well, lots of places. The Sun. Healing. Music. Poetry. Being Very Handsome. But along with all those things, I’m the God Of Reason And Science. And maybe a long time ago, curiosity caused all the world’s problems. But now it’s the other way around. Curiosity’s solving problems, Pandora. All over the world, curiosity is solving famine, it’s solving poverty, it’s solving disease. They put smallpox back in the box, Pandora!”

“Wait,” said Pandora. “I never heard about that! They found a way to…?”

For a brief moment, Apollo thought Pandora was going to ask a question, but she caught herself. He answered anyway.

“Yes,” he said. “There’s a way to put things back in the box. Maybe. A little. Sometimes. It’s really hard. So hard I wouldn’t have been able to do it myself, and I’m the God of Healing. But they did it. Once. Maybe they’ll be able to do it again. And they did it because of curiosity. They wondered whether they could do it, and then they wondered how they could do it, and then they did it.”

“That’s…really interesting,” said Pandora.

“You’ll think about that?” asked Apollo.

“I will,” said Pandora.

“And maybe…call somebody sometime? Me? Someone else? Anybody? I know Artemis has been wanting to hear from you.”

“Um,” said Pandora. “Maybe? I don’t know.”

“That’s fine. Just…keep the option open.”

“Apollo, *I don’t do open.*”

“Just keep it in mind.”

“Goodbye, Apollo.”

“Goodbye, Dory.”

7. Athena

“Hi. My name is Ari…Smith…and I’m here to see Ms…I don’t know, she probably goes by Tina or Minnie or something like that. Really smart and mysterious and probably in charge of everything?”

The security guard at the entrance to the Athena Mineral Water Tower looked at him skeptically. “Do you have an appointment?”

Ares reached into his pocket.

“I got my fucking appointment right here! Two Medals of Honor! While you guys were selling water to yuppies, I was risking my life for your freedom over in Afghanistan. Come on, man. Can’t a vet get any respect around here?”

The guard shook his head. “Can’t get in without an appointment,” he said.

“So,” said Ares, “it has come to this. Same as always.” A bronze spear appeared in his hand, and he rammed it right through the security guard. Didn’t even bother extracting it, there was more where that came from. Somebody screamed. An alarm sounded. Whistling, Ares walked through the lobby and into the elevator, pressed the button for the top floor. That was where important people had their offices, right?

Apparently it wasn’t. “Excuse me,” Ares asked some kind of secretary sitting at a desk. “Can you direct me to Ms…I don’t know, she probably goes by Tina or Minnie or something like that? Really smart and mysterious and probably in charge of everything?”

A few policemen ran up behind him and started to open fire. Without even looking at them, Ares chucked a spear backwards and somehow managed to impale all three of them at once. The secretary stared at him, eyes wide with horror.

“Damn. I didn’t mean to get you all frazzled. Uh, look. Two Medals of Honor! I’m a vet! Patriotic, trustworthy! Ms. Tina or Minnie or something? Really smart and important? Please?”

“Uh…” The secretary looked terrified, but at least it was the sort of terror that scared her into talking. “Uh, you mean the CEO? Ms. Athena?”

“*Really?* She’s the fucking Goddess of Wisdom And Intelligence And Cleverness and she couldn’t get a better pseudonym than ‘Ms. Athena’? Whatever. Where is she?”

Another elevator ride and a few more cops later, Ares found himself breaking down the door of the CEO’s office.

“Hey,” said Ares. “Long time, no see.”

“Can’t imagine why,” said Athena.

“Look, I’ll be blunt,” said Ares. “I came here to get the golden apple. Give me that and we’re square. I’ll go away. I’ll even pay for the doors. And, uh…everything.”

“What golden apple?”

“Oh, come on. I talked to Aphrodite the other day. She said there’s a new golden apple about. She doesn’t have it. And I talked to Hera. She doesn’t have it. And I thought…who’s been gunning for a golden apple ever since that whole mess with Troy? Who’s the Goddess Of Wisdom And Intelligence And Cleverness and always gets everything she wants? And then I remembered my wonderful older sister who I definitely don’t think is the most annoying person ever, and who seems to be doing pretty well for herself. And I thought maybe I should come pay you a visit. Great water, by the way. I tried some on my way here.”

“Fact is,” said Athena, “I don’t have any golden apples.”

“Oh, lay off it, we both know you’ve got the damn apple. Give it to me or else I’ll smash this place up however much it takes to find it.”

About a dozen SWAT officers burst into the office. “Ms. Athena! There’s an intruder in the building!”

“It’s taken care of,” said Athena. “Go off and have a nice day.”

The SWAT team left.

"They believe you?" asked Ares, who was about seven feet tall, dressed in Trojan War vintage armor, carrying a huge bronze spear still covered in blood, and clearly visible.

"I've…put a glamour upon myself," said Athena. "It helps a lot, working with mortals. As long as I'm around, nobody notices anything unusual."

"And you didn't even want their help?" asked Ares. "Even though you're alone, with your younger brother, who happens to be unbeatable in combat?"

Athena laughed. "Unbeatable? Ares, you have *no idea* what you've just walked into. I understand Hermes has figured it out, which means I'll have to take care of him sooner rather than later. But you? You waltz in here, expecting me to be a pushover? Let me show you the tiniest taste of what I can do."

She opened the window. She stretched out her hand. A bolt of lightning arced from her fingers, struck the street below.

"Lightning?" asked Ares. "But…only Dad could call lightning!"

"Not anymore," said Athena. "Come on, Ares. You want to fight? Let's fight."

Ares threw his spear. It stopped in midair, like it had hit an invisible wall. Then it turned, flew back at him, coiled around like a snake, tied him down. "Hey!" he protested. "Hey! That's not fair!"

"I'm so glad you came," said Athena. "I needed a test subject. To see if my powers were really as strong as I hoped. What's the hardest thing in the world, Ares? *Binding a god*. Only ever accomplished twice in history. The Titans.

Prometheus. Both times, by the power of Zeus and all the other gods combined. Do I dare attempt such a thing alone? I believe I do."

The lights darkened. The air began to stir. Lightning arced back and forth across the room. A secretary opened the door, saw the chaos, said "Oh, looks like you're busy," closed the door, and walked out. Time seemed to stop.

There was a rush, a whistle, and a thud, and then Ares wasn't in the world anymore.

8. Prometheus

"Are we there yet?" asked Heracles.

"When we are there," said Hermes, "I promise I will tell you."

"It's just that I was wondering," said Heracles, "whether we were there."

"There are," said Hermes, "certain games mortals play, in which a necessary prerequisite is to create your own hero character. And in some of those games, you get a certain amount of points, which you are allowed to allocate either to intelligence or to strength, so that the smarter you are, the weaker you must be, and vice versa. And I notice, Heracles, that you are the strongest man who has ever existed. Do you know what that implies?"

"It implies that I'm very strong," said Heracles. "But also, I was wondering – are we there yet?"

Hermes sighed. They were in the former Soviet Republic of Georgia driving down a dirt road in a narrow Caucasus mountain pass. It would have been unpleasant for several reasons even without his companion’s endless whining. Still, he was feeling just a little bit euphoric.

One week ago, Ares had attacked Athena, raving about golden apples, and then…nothing. He had disappeared. He’d asked his girlfriend Tyche to find him. Tyche was the goddess of luck – a nice catch, if you ran a hedge fund. She could find anything. But she couldn’t find Ares. He wasn’t in the world. There was only one other place he could be.

Tartarus. The Pit. The Abyss. The place beyond space where those removed from the world languished in darkness for eternity.

He’d gotten on the first flight to Memphis, shaken his father awake. Drunk as he was, Zeus had understood immediately. If Athena had gained enough power to open Tartarus, any one of them could be next. Their very souls were in danger.

And after a lot of arguing and screaming, Hermes had changed tactics and brought out some wine, and he had gotten Zeus very, very drunk. And whether it was one of those things, or another, or the combination of all of them, Zeus had divulged the key to Mt. Elbrus, the one that accessed the secret prison of Prometheus.

“We’re here,” Hermes told Heracles. He parked the car in a bed of gravel by the side of the road. They were in a narrow defile. Mt. Elbrus – the mortal one, the one visible to humans – loomed in front of them. In the rock face to their left, there was an opening just narrow enough to fit a single person at at time.

“Now remember,” he said, as he turned on his flashlight and squeezed into the cave, “You’re going to be wearing these ear plugs. You’ll stare straight ahead, at my back, nowhere else. You’ve got the bottle of magic water in your pocket, and…”

“Why do I have to wear the earplugs?” asked Heracles.

“We’ve gone over this *a thousand times*,” said Hermes. “You have to wear the earplugs because Prometheus knows *literally everything*. He knows what he has to say to scare you, or turn you against me, or make you kill yourself. So you’re just going to wear earplugs and not listen to him.”

“And why do I have to stare at your back?”

“Because if you stare at Prometheus, maybe he can influence you with some kind of facial expression or hand signal, and then you’ll still end up killing yourself. Or killing me. Or dethroning Zeus and returning the universe to primaeval chaos. Or something too horrible to even think about.”

“Hermes?”

“Yeah?”

“I don’t think there’s any hand signal that would make me dethrone Zeus and restore the universe to primaeval chaos.”

Hermes sighed.

“Heracles, do you remember when I told you to meet me in the Caucasus Mountains of Georgia, and I specifically said *former Soviet Republic of*, and I specifically made you repeat back to me ‘former Soviet Republic of’, and a few hours later I got a call from Atlanta International Airport asking me where I was?”

“Yeah.”

“So consider the possibility, however remote, that *Prometheus might be smarter than you*.”

“Oh,” said Heracles. “I guess I hadn’t thought of that.”

They pushed on through the cave, winding around huge stalagmites, stepping over pools with pale eyeless fish.

“That’s why,” Hermes continued, “when we open the secret gate, I am going to talk to Prometheus, and you are going to wear the earplugs and stare at my back.”

“But,” asked Heracles, “what if Prometheus tells *you* to dethrone primaeval chaos or whatever?”

“Excellent question,” said Hermes, “That’s why I brought *you*. I am going to go forth and talk to Prometheus. I have here a cell phone which is programmed to accept exactly one hundred characters of input. When Prometheus tells me how to defeat Athena, I will enter it into the cell phone. When I give you the signal, you will usher me back into this cave, away from Prometheus. And once we are in the cave, you will give me this vial of water from the River Lethe, which will cause me to forget everything that happened in the past eight hours.”

“So I’m just here to…give you the water?” asked Heracles, confused.

“It’s more complicated than that. If I show the slightest sign of *not* wanting to drink that vial of water, then it’s your job to overpower me and force it down my throat, all without allowing me to communicate with you in any way. I trust that you will be able to manage that?”

“I’m very strong,” said Heracles.

“And that,” said Hermes, “is why we love you. There is one more thing I’m going to ask of you. After I’ve drunk the water from Lethe, but before I wake up, you need to read the message on the cell phone yourself and confirm that it looks like a strategy for defeating Athena and *not* like some other kind of message from Prometheus to the outside world, and *certainly not* like any other terms Prometheus has added to our bargain. *If* you see something that looks like a message from Prometheus or an extra term, I need you to smash the cell phone and drink this *second* vial of water from the River Lethe.”

“Oh good,” said Heracles. “I like smashing things and I like water.”

“The only problem,” said Hermes, “is that you are a couple of filaments short of a light bulb. So what I’m going to do is ask you to swear on the River Styx that you’ll comply. You’re half-god; that kind of oath is self-enforcing. As long as even the tiniest part of you remembers what you’ve sworn to do, it will be literally impossible to do otherwise.”

“All right. I swear by river sticks that I’ll do what you say.”

“By the RIVER STYX!”

“I swear by the River Styx that I’ll do what you say.”

The cave briefly darkened, and there was a gust of icy wind that seemed to come from nowhere.

“Good. Now, put the earplugs in, and be quiet for just a second. I need to concentrate here.”

He searched for a part of the cave wall that was just a little too smooth.

“Hermes?” asked Heracles.

“Yes?” asked Hermes.

Heracles said nothing.

“YES?” asked Hermes.

Heracles still said nothing. Hermes saw that he was wearing the earplugs.

“Hermes, if I have the earplugs in, how will I know if I’m being quiet?”

Hermes gave what he hoped was a reassuring-looking shrug, then went back to scanning the cave wall.

There.

A little too smooth, a little too pale. Hermes served part-time as God Of Magic, and he could *sense* something off about that part of the cave. He put his hand on it. Unnatural warmth. The key went here.

With a harpy-feather quill, in ink of ichor, Hermes wrote:

```
D E T E R
E X I L E
T I T A N
E L A T E
R E N E W
```

The wall opened, and sunlight shown through.

They climbed out onto a rock promontory. The scene before them both was and wasn’t Mount Elbrus. The snow shone just a little bit brighter. The sunlight glittered just a little more. The shadows were a little bit darker.

And from under the mountain poked out a gigantic head, four titanic limbs, and bits of a huge torso. A giant, lying supine, pinned down by the peak. On the right half of the torso sat a great eagle, taking occasional bites of liver.

“Hello, Hermes,” said Prometheus.

Well, no turning back now, thought the god.

“Hello, Prometheus,” said Hermes. “With all due respect, I’m trying to minimize information flow with you, so I’d like you not to speak until I’ve finished explaining.” He paused, waiting for an objection, staring at Prometheus even though he knew he shouldn’t. He tried to read the Titan’s great bearded face. He looked surprisingly cheerful for a man pinned underneath a mountain having his liver eternally pecked out.

Finally, Prometheus nodded.

“We’ve got a problem, back in the world. Two thousand years ago, the animal sacrifices stopped working. Eventually we figured out it had to do with the precession of the zodiac. The source of power went from rams to fish and now to water. Athena figured it out first, and now she’s got a monopoly on the water industry. She’s taken all of the divine power and become strong enough to send gods to Tartarus. The rest of us have some residual abilities, but otherwise we’re barely beyond mortal level. We’re at a loss, and we were hoping that, um, your special abilities might be able to help us. So we’d like to offer you a deal. In exchange for information that helps us defeat Athena, we’ll, um, remove the eagle. There’s a key…I don’t have it here, but it can be used remotely. We'll do that. And say we're sorry about it. Really sorry."

That they would never free Prometheus went without saying, so Hermes didn't say it.

The Titan still looked alarmingly cheerful.

"You can, uh, talk now, if you want," said Hermes. "Though, maybe try to keep it short."

"I appreciate the apology," said Prometheus. "Really, I do. And I think we can deal with each other. Removing the eagle would be great, of course. But there's one more thing I've got to ask."

This was what he'd been afraid of. He was desperate. Prometheus knew it. Each additional term was a malignant seed that could grow into anything at all. He would have to hold fast to his plan and pray it was enough.

"Alas," said Hermes, "We predicted that you might say that, so we've taken some measures to precommit not to change any of our terms. In particular, I have sworn by the River Styx – an oath which it is *literally impossible* for gods to break – that I will not accept any terms other than the ones I just mentioned. Also, once you give me your strategic advice, I will be writing down a very short hundred-character summary on this phone, which is programmed to accept no more than a hundred characters and will physically melt if any attempt is made to interfere with that programming. Then I will give a pre-determined hand signal to Heracles, who will escort myself and the phone back into the cave and the ordinary world and force-feed me a vial of water from the River Lethe so that any memory of our conversation beyond those hundred characters will be lost forever. Heracles will then read the cell phone and confirm that no extra terms have been added to the bargain. If he sees any, he will smash the cell phone and drink water from Lethe himself. Heracles has himself sworn by the River Styx to comply with all of this."

Prometheus looked thoughtful – and oh god, were there any three words in the English language scarier than those – and finally he said: "Let's discuss my terms. After you agree to them, I'll tell you how you are going to get around your oath, Heracles, Heracles' oath, and the water of Lethe."

Hermes sighed.

"My terms are: you'll remove the eagle. *And* you'll donate $1503.15 to a charity called 'Against Malaria Foundation'."

"Oh no," said Hermes. "Oh no oh no oh no. That is *exactly* the kind of thing I'm not going to do. You want me to take an action in the world? A *specific* action? With multiple bits of information? Oh no oh no oh no oh no there is *no way* you are going to get me to do that."

Prometheus still looked cheerful. "Well then, Hermes, it was nice to chat. I guess you'll be on your way."

"Now hold on. You don't want to take an option, presented at zero cost to you, that will get that eagle out of your liver forever and ever?"

"It's not that I don't want to. It's that bargaining is a game of give-and-take. We both have things we want out of this interaction. The question is how far we're going to go to get them. It's a game of bluffing and counterbluffing. And unfortunately for you, I am the God Of Foresight, and happen to be omniscient. You are going to walk out of here in fifty-one minutes having agreed to my terms. All I am doing is speaking the complicated dance of words that inevitably gets you to that point.”

“I hate everything about this place so much,” said Hermes.

“And I am deeply sorry,” said Prometheus, *and he sounded sorry*, “but I do insist.”

“Why?” asked Hermes. “*What are you plotting?*”

“You know that I like humans. You remember, I gave them fire, so long ago. I still have a fond place in my heart for them, and malaria is a terrible disease, and I thought…”

“You’re omniscient, so you *know* I don’t believe that for a second. Try again! What are you doing? What’s next on your little list of plans? The humans live on Mount Olympus, and *we* have to worship *them*? The Fates accidentally snip their own fingers off and die of blood loss? I know you’re up to some kind of unspeakable horror, the only question is *which one*?”

“Hermes,” asked Prometheus, “has it ever occurred to you that I was out, in the world, for countless aeons before you imprisoned me here? If you’re so afraid of what I can do or say with a single sentence, what do you think happened when I had millennia to tailor everything just the way I wanted it? Things are going well for me, aren’t they? The gods have been brought low. Humans have never been doing better. Zeus thought he was so clever, giving them a box full of evils, but I selected every one of those evils eons beforehand. You know what was in that box, Hermes? *Things to make humanity stronger*. I gave them famine so they would invent agriculture. I gave them disease so they would invent medicine. I gave them war so they would smelt iron. And I left them hope, so that even in their darkest moments they would pull through and keep dreaming. Dream of putting all of those evils back in the box they came from and closing it forever. And they *will*. Do you know how many sentient species in the multiverse developed an industrial base, liberal democracy, and human rights without killing themselves or collapsing into barbarism, Hermes? The number is *one*. One sentient species. *Mine*.”

“Don’t tell me that getting stuck under Mt. Elbrus with an eagle eating your liver was all part of the plan.”

“You don’t think so? Hermes, I am *vast*. I comprise universes. In my mind is every branch of possibility-space that ever will be or could have been. What’s the point of going outside, when the outside is all inside of me? I set up the world how I wanted it, ensured it would go the right direction, and then retired somewhere quiet, somewhere with space to think.”

“But *the eagle?*”

“Okay, I admit I kind of dropped the ball on that one. The Fates are petty little bitches.”

“So now what?”

“So now you remove the eagle, and I’m happy, and you’re happy.”

“Except for this malaria thing.”

“Think of that as my little joke.”

“Your *joke*. You expect me to believe that?”

“I expect you to realize you have no other option, accept my proposal, and leave this place in another forty-six minutes.”

“Fuck you.”

“Then I expect you to go back, defeat Athena, and restore the power of the gods. Except that you will find it doesn’t go quite as far as it used to. Lightning is a cheap trick compared to nuclear weapons. Flying chariots are a little underwhelming when they share the skies with supersonic jets. You will find that your accustomed roles within human society work well for you. You will find yourself using your power not to dominate human society, but to shepherd it along its path. They are entering a very dangerous time now. *Very* dangerous. They need divine intervention, but not from above. They need gods who live disguised among them, and need them as much as they need you, and shepherd them. Athena cannot do it alone, not properly, so I will give you what you need to stop her. I have foreseen your path, and I know you rise to the occasion. So go, with my blessing, and serve Man.”

“Fuck you, just tell me what message I’m sending.”

“A hundred characters? Hmmm. ATH CAN’T HOLD POWER IN BODY. NEEDS FOCUS, PROBABLY IDOL. DONATE $1503.15 TO AMF, THEN DESTROY IDOL.”

Hermes typed it onto his cell phone. “And how are we going to get around all of the oaths and precautions?

“Tell me, Hermes, when did you swear your oath by the River Styx?”

“Three days ago.”

“Good. If you don’t remember swearing the oath, you can’t be bound by it. So you need to overdose on Lethe-water, enough to erase three days from your mind. I imagine you’ve been planning this escapade for a while, so when you wake up in a cave in the Caucasus with a cell phone bearing a message, you'll be able to piece together what happened. The message is framed such that the donation looks like part of the plan, so Heracles won't notice anything amiss. You'll probably figure it out, but you're an honorable god and you'll feel compelled to stick to the bargain that you must have made with me. None of this breaks your current oath, which only says you must not carry out any of my terms, not that you must not mention them in your message. Overdosing on Lethe-water is only a suggestion of mine, not itself a term necessary to procure my agreement, so it should not be prohibited."

Hermes sighed with relief. "Your plan isn't going to work, Prometheus. Heracles is going to force-feed me the Lethe water before taking out his earplugs, so I can't communicate with him and ask him to change the dose. And even if I could, I only brought eight hours' worth of Lethe water anyway. Sixteen if you count Heracles' vial."

"There are two ways to increase the effect of a drug," said Prometheus. "You can increase the dose. Or you can decrease the rate at which metabolism eliminates it from the body. Since our dose of Lethe water is limited, we're going to go with the second. Heracles will give you exactly the amount of Lethe water you told him, but your body will fail to process it as usual, and it will have ten times the expected effect, causing you to forget your oath and be able to accept my amended terms when you find them on your phone."

"How are you going to change my metabolism?"

"Most drugs are metabolized by the liver. By manipulating liver size, we can tailor the metabolic rate to any level that we want."

“*Manipulating liver size?*” Hermes didn’t like the sound of this.

“Yes. Hepatectomy is a very safe, commonplace surgery. But even if it weren’t, you would have nothing to fear. Surgeons’ success rates correlate with their number of hours of experience. And we have the most experienced liver removal specialist in the multiverse right here on Mt. Elbrus.”

“Oh no,” said Hermes. “You’re not…oh no oh no oh *no*.”

The eagle gave a voracious shriek.

9. Everybody

The pantheon met in the Pantheon, as was tradition. Hermes and his girlfriend Tyche came first; the God Of Commerce took a seat in the center just below the oculus, as the Goddess Of Fortune ushered away confused tourists. Gradually the rest trickled in. Poseidon, tracking water wherever he stepped. Apollo, dapper as ever in a tweed coat and bowtie, and Artemis, dressed in camo. Nike, dressed like she had just come from the gym. And Dionysus, in his stained Sigma Alpha Epsilon sweatshirt. He caught Hermes’ eye. “HEEEEEEY, BRO!” he said. “HOW’S IT HANGING?” Hermes just ignored him.

Hades was over near the entrance, talking to Aphrodite. “Hey Aph,” he said affably. “Want a pomegranate?”

Aphrodite’s eyes narrowed. “Is it one of your magical pomegranates that makes anyone who eats it obligated to become your wife?”

“Uh…” said Hades, shifting his eyes back and forth. “It… *might* not be?”

“I’ll pass,” said Aphrodite.

The missing stood out by their absence. Ares was not with them, for obvious reasons. Athena had obviously not been invited to the conspiracy against her. And Zeus, King Of The Gods, was nowhere to be seen. Hermes had begged and cajoled, but to no effect; he was still angry at having given up Prometheus’ key when drunk. “This is our last chance,” said Hermes, “the most important thing you’ll ever do.” But Zeus was having none of it. He had (he said in a half-drunken stupor) just met with a Hollywood talent scout, who had told him that he was perfect to star in a movie about the Trojan War. He was going to strike it big and become a celebrity and then open up his own water company, and Athena would never know what hit her. That was his plan and he was sticking to it.

Well, he would work with what he had.

“My fellow gods!” he announced, and everyone turned to look at him.

“By now you’ve heard the news. Athena has used her bottled-water monopoly to seize divine power for herself. She has opened the gates to Tartarus; none of us are safe. If we ever want to be more than the second-rate has-beens we are now, we need to stop her. I know how we’re going to do it.”

Some gasps. Apollo looked thoughtful. “WOOOOOOOOOO!” shouted Dionysus. “YOU GO, HERMES!”

“Athena’s collected so much power that she can’t hold it all herself,” he said. He’d gone over all this with Apollo, a few days after waking up in the cave with a terrible headache; the two of them had managed to expand Prometheus’ cryptic message into an actionable plan. He was very suspicious that a seemingly unrelated order to donate a very specific sum of money was a command of Prometheus' that had slipped past his security, but he wasn't sure how, and he wasn't going to take the risk. He'd made the donation – now the rest was up to them.

"She can't hold it all herself," he continued, "so she needs some kind of supplementary focus. Sympathetic magic. Like calls out to like. She needs an idol. And not just any idol. It would have to be something really special, an idol of Athena that generations of mortals have identified with the deepest secrets of her power. The history books list two such idols. One, the giant statue in the Parthenon. That's destroyed. Two, the Palladium. It was there in Troy. It was there in Rome. Now we think it's in the Athena Mineral Water headquarters. Why? Because that kind of power would stand out like a sore thumb *unless* it was outshone by the presence of another immortal. Athena sure wouldn't trust anyone else with it, so she's got it herself. It must be hollow. The divine energy must be stored inside of it. If we can find and destroy it, then Athena loses her power and it flows into alternate conduits. Like us. In other words, we get our magic back."

"WOOOOOOOO!" shouted Dionysus.

"Please refrain from cheering until the entire speech is over," said Hermes. "Anyway, here's my plan. We're going to split in two. One group is going to be the powerhouses. Apollo, Artemis, Hades, Poseidon, Aphrodite, Dionysus. You're all strong, skilled with weapons, or both. You're going to smash things, create a distraction. You're going to avoid confronting Athena directly, because Ares already showed us how that turns out. While my sister is chasing after you, the second group slips in. That's me and Tyche. Hades has given me his helm of invisibility, which should be enough power to hide both of us from view. Tyche's the Goddess Of Fortune. She can find anything. And I'm the God Of Thieves. I can break into anywhere. She'll lead me straight to the Palladium, I'll nab it, break the thing in two, and then we're home free. Any questions? Comments?"

"It's a good plan," said Apollo, nodding his head.

"WOOOOOOOOOO!" shouted Dionysus.

"Just do it!" agreed Nike.

And before they could change their mind, Hermes teleported the lot of them to the lobby of Athena Mineral Water.

They appeared in a flash of light. People stepped back, shocked. The teleportation was strange enough. But Poseidon was still holding his golden trident. Hades was surrounded by some kind of miasma. And Aphrodite was buck naked. They didn't exactly blend in.

"*Distraction!*" whispered Hermes, just before taking Tyche's hand and vanishing from view.

"Uh," said Dionysus. "IS EVERYONE HERE READY TO PAAAAAAAAARTY?"

"It'll do," muttered Hermes.

He and Tyche made their way up side staircases. Athena's aura wouldn't be able to hide the Palladium at any kind of a distance. It had to be really close to her office. They came to the CEO suite by a back entrance, then pressed themselves against a wall as they saw "Ms. Athena" walk by, talking on a cell phone. "Yeah," she was saying unconvincingly, "that does sound weird. No, no idea what's going on. I'll be down to investigate. Thanks for the tip."

When she was out of view, they snuck into her office. It looked very normal. A few potted plants. A Bosses' Day card. Some gold-plated "Female Entrepreneur Visionary Leadership" awards. A wall full of framed news articles "ATHENA MINERAL WATER BOASTS GODLIKE PROFITS", "BEHIND THE STARTUP CHANGING HOW THE WORLD DRINKS". A bottle of product on her desk, either for display or hydration. No idols.

"Cold," said Tyche.

"Cold?" asked Hermes.

"If it were here, I would know. It's not here."

"Well, let's check nearby."

They checked Athena's secretary's office. They checked Athena's closet. They checked the office of the Assistant To The CEO, the Director Of The Office Of The CEO. They checked the executive bathroom. No idols.

"Super cold," said Tyche. "Hermes, it's nowhere *near* here."

"Fuck," said Hermes. "We've got to go. Find the others and tell them to disengage, before it's too late."

They ran down the stairs until they reached the lobby. It was in a state of disarray. Chairs and potted plants overturned. Three parallel lines on the the big LCD screen that looked like they had been scratched by a trident. There was a magic silver arrow sticking out of one wall. No gods.

“Okay,” said Tyche. “They’ve been here. They must be retreating.”

They ran outside. A trail of water on the sidewalk suggested the route taken by Poseidon. The parking garage. He could see flashes of lightning on the lower levels. He wouldn’t be able to get through that way. He channeled all his power into his winged sandals, and he and Tyche lurched into the air, coming to rest on the top floor of the structure. He ran down and almost bumped into Aphrodite.

“Hey, sexy,” she said. “What’s going on?”

He could see the others now. Hades and Poseidon were defending the road leading to the lower level. Athena was below, hurling lightning at them. They were in retreat. Artemis stood on the bed of a pickup truck, taking shots with her magic arrows. Nike was with her, pointing out targets. Dioynsus seemed to be passed out on the concrete, and Aphrodite and Apollo were holding up the rear.

“We’ve got to get out of here,” Hermes told the two of them. “I was wrong. The Palladium’s not in the building.”

A lightning bolt shattered one of the big concrete pillars. “You dare stand against me?” shouted Athena. “For too long, I’ve played second-fiddle to lesser deities like yourselves! When I awoke a few centuries ago, it removed the last doubt from my mind. Everything I predicted was true. Nowadays, who cares about agriculture? Who cares about the sea? Who even *believes* in the Underworld? The sun is a giant ball of gas. The moon is a giant ball of rock. There’s only one thing that matters today, and that’s intellect! And how better to enshrine the triumph of intellect over human affairs, then to have the Goddess Of Wisdom destroy the lesser gods and become a pantheon unto herself? People these days want monotheism, and I’m going to give it to them!”

“You’re wrong!” Apollo stepped into the fray. “Intellect is important, yes! You deserve to be honored, and nobody will take that away from you! But without Reason to guide it, intellect becomes monstrous. Without Art, and Music, and Poetry, intellect becomes sterile. And without Healing, intellect becomes divorced from compassion.”

“AND THE SEA IS REALLY GREAT TOO,” added Poseidon.

Athena rose into the air, crackling with energy. “*For now*,” she said. “*For now*, intellect runs on puny mortal minds that will get all *sad* if they don’t have their music and their beachfront houses. But that was a *mistake*, Apollo. We didn’t *want* humans. We wanted apes just barely smart enough to sacrifice some rams to us and be properly grateful. Then Prometheus got involved, and everything went wrong. I’m going to fix his mistake. Genetic engineering, robotics, so many different options. Create minds that don’t need art, that don’t waste their time with music or lolling at the beach.” She looked at Artemis. “Destroy the forests and pave them over with factories.” She looked at Dionysus. “Replace partying with study and productive work.” She looked at Aphrodite. “Replace the vagaries of love with rational breeding based on genetic potential.” She looked at Hades. “Machines, that were never alive and so can never die.” She looked at Poseidon. “Tame the sea for tidal power – ”

“YOU’RE TOUCHING THE SEA OVER MY DEAD BODY!” Poseidon shouted, and rushed at her with his trident.

“We’ve got to get out of here,” Hermes whispered to Apollo. “Get together as many as you can. We’re going to make a run for it.”

“She’s blocking the only exit,” Apollo said. “Where do we go?”

“To the roof! I can carry some of you with my wings. The rest will have to jump.”

A few other gods had gotten the gist of the conversation, started running to the top of the parking garage. There was a loud thud, then the sound of sparks. It didn’t sound good.

“I don’t understand,” said Apollo. “How could the Palladium not be in the tower?”

“I don’t know!” Hermes protested. “If it wasn’t disguised by a god’s aura…”

“Then ipso facto it must be with some other god,” said Apollo. “Who are we missing? Demeter?”

“Demeter? She *hates* Athena, thinks her bottled water is destroying the environment.”

There was another crash. Apollo, Tyche, and Hermes made it onto the roof of the parking structure. They couldn’t tell how many other gods were still following.

“Okay then, Hera?”

“I checked. There’s a court record of all of her property, after the divorce with Zeus. Nothing about any idols. And she doesn’t like Athena either, something something Trojan something. *Nobody* likes Athena. And seriously, who’s going to take a magic idol and just say ‘sure, I’ll hold on to this, no further questions’.”

“Wait,” said Apollo.

Hermes waited.

“Does it have to be a *god* god? What about a demigod? An immortal human?”

“Um. In theory it could work. But it would be such a small effect. They’d have to stay right by the idol, day in, day out, or it wouldn’t be disguised at all.”

Apollo was already taking out his cell phone. “Dory, Dory, *please pick up*.”

Nike ran onto the roof of the parking garage. There was a big gash down one of her arms. “She’s right behind us!” she told them. “We’ve got to go!”

“Wait,” said Apollo. “Dory, *pick up the phone*.”

There was another crash. The parking structure started to wobble.

Apollo heard a noise from the other side of the phone, but no greeting. Right. She wouldn’t open the conversation.

“Pandora?” he asked. “Are you there?”

“Hi Apollo,” came her voice.

“Dory,” said Apollo. “That statue on your desk, the one of the woman in armor. I need you to take it and smash it, really hard.”

“Okay,” said Pandora. There was a brief pause. “Done.”

“Done? Did you break the statue?”

“No, it’s very hard, it doesn’t seem to have broken.”

“Fuck fuck fuck fuck *fuck*,” said Apollo.

Of all people, Dionysus managed to crawl his way to the top of the structure. “It’s getting *really gnarly* down there,” he announced before collapsing back into unconsciousness.

“Okay. I need you to feel along the sides of the statue. Is there any kind of switch, anything that’s going to get it to release the power that it’s stored?”

A brief pause. “There’s…a knob and a hinge.”

“Okay, Pandora. I need you to turn the knob and open the statue.”

“Apollo, *I don’t open things*.”

“Look, Dory, I don’t ask you for much. I’ve known you for I can’t even remember how many centuries, and I know things are hard for you, I’ve respected that. But Dory, *you need to open that statue*.”

“You *know* I don’t open things!”

The last few gods ran onto the top of the parking structure. Just behind them floated Athena, her eyes jet black, her whole body crackling with electricity. “There’s nowhere left to run,” she taunted them. “You’re *all* going to Tartarus now. Any last words?”

“Uh,” said Hades, “want a pomegranate?”

Athena held her hands forth. The sky darkened. The air seemed to stir.

“Dory, you made a mistake once, and it was really bad, I’m not denying that, but you told me yourself, the one thing you did right was keep Hope. I need you to be hopeful now. I need you to hope that someday, somebody, us, humans, somebody we’re not even considering, might be able to reverse what you did. Might be able to put those evils back in the box. I need you to think that that’s possible. But not going to happen without our help. Please, Pandora, trust me on this. And what I need you to do right now is *open that statue*.”

Lightning arced back and forth across the heavens. Time seemed to stop.

Then there was a loud *pop*.

10. Zeus

Zeus had come onto stage believing it was an audition for a big-budget film about the Trojan War. It wasn’t. Out ran a young woman, her face streaked with tears. “You said you loved me!” she said. “We had a child together! And then you…you disappeared!”

“Hey now,” said Zeus. “What’s this now? Who are you? Whaddyatalkinabout?”

“Don’t you recognize me?” sobbed the woman. “I’m Sara! From Biloxi! We met in ’98! Oh god! You don’t even *remember* me. You’ve probably abandoned with so many women that you don’t even *remember* them! How many were there after me? Ten? A hundred?”

“Hang on now,” said Zeus. “I ain’t the kind of guy who hooks up with no hundred women.”

“In fact,” said Alice DiScorria, walking on to stage. “He is precisely that kind of guy. If you don’t believe me, believe Amy. And Bethany. And Billy Rae. And Caroline. And Connie.”

As she said each name, each woman came on to the stage.

“Dana. Daria. Dina…”

Some of them were crying. Some of them looked lost. Some of them had steely determination in their eyes.

“…Jackie. Jessica. Jennifer. Jun-Li…”

“Nah, yer just messin’ with me now. What is this, some kinda trap? I want a lawyer, lady. I got my rights!”

“…Samantha. Sara. Sarah. Shaniqua. Susan…”

The stage was almost full now.

“You sayin’ I slept with all these women? I didn’t sleep with none of em. I want my lawyer, right now.”

“Actually,” said Alice, “we’re not saying these are the women you slept with. We’re saying these are the women you *slept with, had children with, and then abandoned without paying child support.*”

“That’s a goddanged lie,” said Zeus. “I ain’t even got no children.”

“Zeus is telling us that he ‘ain’t even got no children’,” Alice told her viewers. “Alas, we have two hundred and five people in our studio audience today who think otherwise. Would you please stand up? Aaron. Adam. Althea. Ava. Bethany Junior. Berenice.” She realized she was starting to lose her audience’s attention. “And all the rest.”

Two hundred five members of the studio audience, ranging fron toddlers to adults, stood up. They were all unusually large, and many of the men had big, flowing beards.

“This is goddanged lies, is what it is!” shouted Zeus. “None of these people ain’t my children, and that’s the truth!”

“Zeus says that none of these people are his children,” said Alice. “We ran paternity tests for every single one of them before the show. Let’s see what they say.” She took out a big stack of manila envelopes, opened the first one. “Aaron…Zeus *is* the father! Adam…Zeus *is* the father! Althea…Zeus *is* the father!”

One of the women on stage finally lost it, grabbed a folding chair, and swung at Zeus. He deflected the blow easily, then pushed her back, just a little too rough. Suddenly the stage had become a brawl, one hundred ninety enraged women against one underpowered god.

“Ava…Zeus *is* the father! Bethany Junior…Zeus *is* the father! Berenice…Zeus *is* the father! Chou-yang…Zeus *is* the father! Cleo…Zeus *is* the father!”

The brawl on stage was getting really bad now. A few women were down for the count. Zeus was bleeding all over his face. Some of the staff started to wonder whether they should override Alice and call security.

“Demetrius…Zeus *is* the father! Delia…Zeus *is* the father! Darragh…Zeus *is* the father! Dominique…Zeus *is* the father!”

One of the women had gotten hold of Zeus hair and was holding him, pinned, while another was slapping his face. Zeus tried to kick, but ended up losing his balance. Security guards were pushing through the crowd of women, who were resisting their efforts.

“Edna…Zeus *is* the father! Elena…Zeus *is*…”

Then there was a loud popping sound that seemed to come from everywhere and nowhere. And then Zeus effortlessly pushed the crowd of women away from him. For a second, he looked confused by his own strength. He stared at his newly-rippling muscles, looked down at the ground as if he couldn’t quite believe how tall he was. Nobody moved.

Then he shouted, “DAMN RIGHT I’M YOUR FATHER! I’M ALL YER FATHERS. I AM ZEUS OLYMPIOS, KING OF THE GODS, CLOUD-GATHERER, THUNDERER, MIGHTIEST OF IMMORTALS! AND IIIIII’M BAAAAAAAAAAAAACK.”

The crackle of lightning filled the halls, knocked over the security guards. The audience stampeded to the exits. Women started to run off the stage.

“I AM ZEUS, KING OF GODS AND MEN. AND I’M TURNING YOU ALL INTO ANIMALS!”

Amy became an anteater. Bethany became a duck. Billy Rae became a tree shrew. Caroline became an otter. He turned Connie into a rattlesnake and Dana into a panther, Daria into a Komodo dragon and Dina into a bat. It was over in minutes. Everyone had either escaped or been transformed, besides Zeus and the hostess.

“Yer still here,” said Zeus, surprised.

“I am everywhere,” said Eris Discordia.

“What *happened*?” asked Zeus.

“The same thing that happens everywhere, all the time” said Eris. “People had conflicting aims. They struggled for power. Some won, others lost. The winners will celebrate, thinking their victory irreversible, and the losers will mourn, plotting their vengeance. And around them, the world changes irreversibly, in ways none of them predicted.”

“Huh,” said Zeus.

“In a few hours, news will come that a sudden electrical storm struck the set of my show, unfortunately causing the cameras to stop recording. Some people will be missing, casualties of the disaster. Others will say all sorts of strange things and be ignored. There will be lots of fights about it, and they’ll all call each other things like ‘sheeple’ and ‘denialist’ and ‘moron’. It will be wonderful.”

“Huh,” said Zeus.

“In the meantime, the studio is ruined. I suppose I will have to find a new job. Can you believe it, Zeus? In the old days, I was barred from every city and temple, driven out into the wilderness as an enemy of mankind. Now they *pay* me to cause discord. What a world!”

“It’s…somethin’,” said Zeus

“And it’s all thanks to people like you,” said Eris. “So before we part ways again, before the poets end their songs and the next myth begins, please accept a token of appreciation. From me, to you.”

In her hand appeared a shining golden apple.

Epilogue: Trump

“Yeah,” real estate mogul Donald Trump said into the phone. “Look, I gotta go, Carl. I gotta be at a gala tonight – yeah, the one for the American Eagle Museum. Terrible stuff, Carl, just terrible. Gotta go.”

He hung up. It really was terrible stuff. Just a year ago, an anti-malaria charity had funded a grant that happened to precisely match its yearly budget surplus. The research had borne fruit – a new insecticide, kind of a super-DDT without the environmental damages. DDT, of course, was famous for killing endangered birds, but they thought they’d tested it properly this time, dozens of different bird species, no problems at all. So they’d deployed it worldwide, and malaria rates had plummeted. Only they hadn’t tested the environmental consequences as well as they’d thought. 99% of bird species escaped unscathed – but every eagle in the world had died an unimaginably agonizing death. The whole situation was so strange that the FBI launched an investigation – then closed it a few weeks later for absence of motive. Who could possibly hate eagles that much?

He put on his suit and tie, and was just about ready to head out when a beam of radiant light appeared in the middle of his room and coalesced into three women.

“Greetings to you, Mr. Trump,” said the oldest. “I am Hera, Queen of the Gods. These are my colleagues Aphrodite and Athena. You are the man who runs the Miss Universe beauty contest, yes?”

He took a step back, dazzled by her radiance. “Um…yes.”

“Zeus, God of Thunder, recently came into possession of a golden apple. Then a second golden apple, found when searching a convent in Ukraine that had become a center of, ah, certain recent events. There are three of us and only two apples, so we petitioned Zeus to determine how they might be divied up. He replied that traditionally they go to the fairest, and so urged us to seek the foremost mortal judge of female beauty and implore his assistance. If you truly run beauty pageants for the entire universe, then you are the judge that we seek.”

Then she spoke differently, directly into his mind. *And as an added incentive, if you choose me, I swear by the River Styx that I will make you the most powerful man in the world.*

He’d barely had time to process the thought when Aphrodite stared at him, and a voice like music touched his consciousness, saying *Pick me, and I swear by the River Styx that I will give you any woman you desire. Models, supermodels, they can all be yours.*

Then a third voice, lower, more dispassionate, and he heard Athena say *Select me as most beautiful, and I swear by the River Styx that I will grant you wisdom, prudence, and the intelligence to make the right decision under any circumstance*.

Donald Trump just stared.

“Well?” asked Hera.

“He’s not answering!” said Athena.

“Waaaaait a second,” said Aphrodite. “Athena, did you ever turn off that glamour you had, that made mortals around you unable to process the presence of gods?”

“How was I supposed to turn that off?” asked Athena. “It took the whole divine power of the universe to create that, and then you took that away from me. Now I’m just a goddess like anyone else, doing – ” she spat “community service to make up for past misdeeds. And it's not even like I didn't help you guys bring Ares back."

"So what you're saying," interrupted Hera, "is that he can't even see us?"

"He can *see* us," said Athena. "He just can't comprehend that anything unusual might be going on," said Athena.

Finally, Donald Trump rubbed his eyes, and said "I got no idea who any of you are, or why you're in my apartment, but – " he pointed at Aphrodite and Hera "you and you are *smokin'*. You," he said, pointing to Athena and frowning, "look like a dyke in that armor. Seriously, get a makeover."

Then he walked out the door.

"Huh," said Hera.

"Too bad," Aphrodite told Athena. "Just goes to show that brains aren't everything."

"Yes, well," said the Goddess of Wisdom, a little too haughtily to be anything but compensation, "I'm just glad we finally made it through one of those without causing any unfortunate side effects for world history."

"Yes," said Hera. "I suppose we did. There's a first time for everything."

*[**Acknowledgments:** the idea of Zeus on a trashy TV show comes from this Tumblr post. Ideas for the Prometheus character came from AI boxing and The Wise Man's Fear. The first two word squares come from here and here.]*

Antidepressant Pharmacogenomics: Much More Than You Wanted To Know

[Epistemic status: very uncertain. Not to be taken as medical advice. Talk to your doctor before deciding whether or not to get any tests.]

I.

There are many antidepressants in common use. With a few exceptions, none are globally better than any others. The conventional wisdom says patients should keep trying antidepressants until they find one that works for them. If we knew beforehand which antidepressants would work for which patients, it would save everyone a lot of time, money, and misery. This is the allure of pharmacogenomics, the new field of genetically-guided medication prescription.

Everybody has various different types of cytochrome enzymes which metabolize medication. Some of them play major roles in metabolizing antidepressants; usually it’s really complicated and several different enzymes can affect the same antidepressant at different stages. But sometimes one or another dominates; for example, Prozac is mostly metabolized by one enzyme called CYP2D6, and Zoloft is mostly metabolized by a different enzyme called CYP2C19.

Suppose (say the pharmacogenomicists) that my individual genetics code for a normal CYP2D6, but a hyperactive CYP2C19 that works ten times faster than usual. Then maybe Prozac would work normally for me, but every drop of Zoloft would get shredded by my enzymes before it can even get to my brain. A genetic test could tell my psychiatrist this, and then she would know to give me Prozac and not Zoloft. Some tests like this are already commercially available. Preliminary results look encouraging. As always, the key words are “preliminary” and “look”, and did I mention that these results were mostly produced by pharma companies pushing their products?

But let me dream for a just a second. There's been this uneasy tension in psychopharmacology. Clinical psychiatrists give their patients antidepressants and see them get better. Then research psychiatrists do studies and show that antidepressant effect sizes are so small as to be practically unnoticeable. The clinicians say "Something must be wrong with your studies, we see our patients on antidepressants get much better all the time". The researchers counter with "The plural of anecdote isn't 'data', your intuitions deceive you, antidepressant effects are almost imperceptibly weak." At this point we prescribe antidepressants anyway, because – what else are you going to do when someone comes into your office in tears and begs for help? – but we feel kind of bad about it.

Pharmacogenomics offers a way out of this conundrum. Suppose half of the time patients get antidepressants, their enzymes shred the medicine before it can even get to the brain, and there's no effect. In the other half, the patients have normal enzymes, the medications reach the brain, and the patient gets better. Researchers would average together all these patients and conclude "Antidepressants have an effect, but on average it's very small". Clinicians would keep the patients who get good effects, keep switching drugs for the patients who get bad effects until they find something that works, and say "Eventually, most of my patients seem to have good effects from antidepressants".

There's a little bit of support for this in studies. STAR*D found that only 33% of patients improved on their first antidepressant, but that if you kept changing antidepressants, about 66% of patients would eventually find one that helped them improve. Gueorguieva & Mallinckrodt

(2011) find something similar by modelling “growth trajectories” of antidepressants in previous studies. If it were true, it would be a big relief for everybody.

It might also mean that pharmacogenomic testing would solve the whole problem forever and lets everyone be on an antidepressant that works well for them. Such is the dream.

But pharmacogenomics still very young. And due to a complicated series of legal loopholes, it isn’t regulated by the FDA. I’m mostly in favor of more things avoiding FDA regulation, but it means the rest of us have to be much more vigilant.

A few days ago I got to talk to a representative of the company that makes GeneSight, the biggest name in pharmacogenomic testing. They sell a $2000 test which analyzes seven genes, then produces a report on which psychotropic medications you might do best or worst on. It’s exactly the sort of thing that would be great if it worked – so let’s look at it in more depth.

II.

GeneSight tests seven genes. Five are cytochrome enzymes like the ones discussed above. The other two are HTR2A, a serotonin receptor, and SLC6A4, a serotonin transporter. These are obvious and reasonable targets if you’re worried about serotonergic drugs. But is there evidence that they predict medication response?

GeneSight looks at the rs6313 SNP in HTR2A, which they say determines “side effects”. I *think* they’re thinking of Murphy et al (2003), who found that patients with the (C,C) genotype had worse side effects on Paxil. The study followed 122 patients on Paxil, of whom 41 were (C,C) and 81 were something else. 46% of the (C,C) patients hated Paxil so much they stopped taking it, compared to only 16% of the others (p = 0.001). There was no similar effect on a nonserotonergic drug, Remeron. This study is interesting, but it's small and it's never been replicated. The closest thing to replication is this study which focused on nausea, the most common Paxil side effect; it found the gene had no effect. This study looked at Prozac and found that the gene didn't affect Prozac response, but it didn't look at side effects and didn't explain how it handled dropouts from the study. I am really surprised they're including a gene here based on a small study from fifteen years ago that was never replicated.

They also look at SLC6A4, specifically the difference between the "long" versus "short" allele. This has been studied *ad nauseum* – which isn't to say anyone has come to any conclusions. According to Fabbri, Di Girolamo, & Serretti, there are 25 studies saying the long allele of the gene is better, 9 studies saying the short allele is better, and 20 studies showing no difference. Two meta-analyses (1 n = 1435, 2 n = 5479) come out in favor of the long allele; two others (1 n = 4309, 2, n = 1914) fail to find any effect. But even the people who find the effect admit it's pretty small – the Italian group estimates 3.2%. This would both explain why so many people miss it, and relieve us of the burden of caring about it at all.

The Carlat Report has a conspiracy theory that GeneSight really only uses the liver enzyme genes, but they add in a few serotonin-related genes so they can look cool; presumably there's more of a "wow" factor in directly understanding the target receptors in the brain than in mucking around with liver enzymes. I like this theory. Certainly the results on both these genes are small enough and weak enough that it would be weird to make a commercial test out of them. The liver enzymes seem to be where it's at. Let's move on to those.

The Italian group that did the pharmacogenomics review mentioned above are not sanguine about liver enzymes. They write (as of 2012, presumably based on Genetic Polymorphisms Of Cytochrome P450 Enzymes And Antidepressant Metabolism“>this previous review):

> Available data do not support a correlation between antidepressant plasma levels and response for most antidepressants (with the exception of TCAs) and this is probably linked to the lack of association between response and CYP450 genetic polymorphisms found by the most part of previous studies. In all facts, the first CYP2D6 and CYP2C19 genotyping test (AmpliChip) approved by the Food and Drug Administration has not been recommended by guidelines because of lack of evidence linking this test to clinical outcomes and cost-effectiveness studies.

What does it even mean to say that there’s no relationship between SSRI plasma level and therapeutic effect? Doesn’t the drug only work when it’s in your body? And shouldn’t the amount in your body determine the effective dose? The only people I’ve found who even begin to answer this question are Papakostas & Fava, who say that there are complicated individual factors determining how much SSRI makes it from the plasma to the CNS, and how much of it binds to the serotonin transporter versus other stuff. This would be a lot more reassuring if amount of SSRI bound to the serotonin transporter correlated with clinical effects, which studies seem very uncertain about. I’m not really sure how to fit this together with SSRIs having a dose-dependent effect, and I worry that somebody must be very confused. But taking all of this at face value, it doesn’t really look good for using cytochrome enzymes predicting response.

I talked to the GeneSight rep about this, and he agreed; their internal tests don't show strong effects for any of the candidate genes alone, because they all interact with each other in complicated ways. It's only when you look at all of them together, using the proprietary algorithm based off of their proprietary panel, that everything starts to come together.

This is possible, but given the poor results of everyone else in the field I think we should take it with a grain of salt.

III.

We might also want to zoom out and take a broader picture: should we *expect* these genes to matter?

It's much easier to find the total effect of genetics than it is to find the effect of any individual gene; this is the principle behind twin studies and GCTAs. Tansey et al do a GCTA on antidepressant response and find that all the genetic variants tested, combined, explain 42% of individual differences in antidepressant response. Their methodology allowed them to break it down chromosome-by-chromosome, and they found that genetic effects were pretty evenly distributed across chromosomes, with longer chromosomes counting more. This is consistent with massively polygenic structure where there are hundreds of thousands of genes, each of small effects – much like height or IQ. But typically even the strongest IQ or height genes only explain about 1% of the variance. So an antidepressant response test containing only seven genes isn't likely to do very much *even if those genes are correctly chosen and well-understood*.

SLC6A4 is a great example of this. It's on chromosome 17. According to Tansey, chromosome 17 explains less than 1% of variance in antidepressant effect. So unless Tansey is very wrong, SLC6A4 must also explain less than 1% of the variance, which means it's clinically useless. The other six genes on the test aren't looking great either.

Does this mean that the GeneSight panel *must* be useless? I'm not sure. For one thing, the genetic structure of *which* antidepressant you respond to might be different from the structure of antidepressant response generally (though the study found similar structures to any-antidepressant response and SSRI-only response). For another, for complicated reasons sometimes *exploiting* variance is easier than *predicting* variance; I don't understand this enough to be sure that this isn't one of these cases, though it doesn't look that way to me.

I don't think this is a knock-down argument against anything. But I think it means we should take any claims that a seven (or ten, or fifty) gene panel can predict very much with *another* grain of salt.

IV.

But assuming that there are relatively few genes, and we figure out what they are, then we're basically good, right? Wrong.

Warfarin is a drug used to prevent blood clots. It's notorious among doctors for being finicky, confusing, difficult to dose, and making people to bleed to death if you get it wrong. This made it a very promising candidate for pharmacogenomics: what if we could predict everyone's individualized optimal warfarin dose and take out the guesswork?

Early efforts showed promise. Much of the variability was traced to two genes, VKORC1 and CYP2C9. Companies created pharmacogenomic panels that could predict warfarin levels pretty well based off of those genes. Doctors were urged to set warfarin doses based on the results. Some initial studies looked positive. Caraco et al and Primohamed et al both found in randomized controlled trials with decent sample sizes that warfarin patients did better on the genetically-guided algorithm, $p < 0.001$. A 2014 meta-analysis looked at nine studies of the algorithm, over 2812 patients, and found that it didn't work. Whether you used the genetic test or not didn't affect number of blood clots, percent chance of having your blood within normal clotting parameters, or likelihood of major bleeding. There wasn't even a marginally significant trend. Another 2015 meta-analysis found the same thing. Confusingly, a Chinese group did a third meta-analysis that *did* find advantages in some areas, but Chinese studies tend to use shady research practices, and besides, it's two to one.

UpToDate, the canonical medical evidence aggregation site for doctors, concludes:

> We suggest not using pharmacogenomic testing (ie, genotyping for polymorphisms that affect metabolism of warfarin and vitamin K-dependent coagulation factors) to guide initial dosing of the vitamin K antagonists (VKAs). Two meta-analyses of randomized trials (both involving approximately 3000 patients) found that dosing incorporating hepatic cytochrome P-450 2C9 (CYP2C9) or vitamin K epoxide reductase complex (VKORC1) genotype did not reduce rates of bleeding or thromboembolism.

I mention this to add another grain of salt. Warfarin is the perfect candidate for pharmacogenomics. It's got a lot of really complicated interpersonal variation that often leads to disaster. We know this is due to only a few genes, and we know exactly which genes they are. We understand pretty much every aspect of its chemistry perfectly. Preliminary studies showed amazing effects.

And yet pharmacogenomic testing for warfarin basically doesn’t work. There are a few special cases where it can be helpful, and I think the guidelines say something like “if you have your patient’s genotype already for some reason, you might as well use it”. But overall the promise has failed to pan out.

Antidepressants are in a worse place than warfarin. We have only a vague idea how they work, only a vague idea what genes are involved, and plasma levels don’t even consistently correlate with function. It would be very strange if antidepressant testing worked where warfarin testing failed. But, of course, it’s not impossible, so let’s keep our grains of salt and keep going.

V.

Why didn’t the warfarin pharmacogenomics work? They had the genes right, didn’t they?

I’m not too sure what’s going on, but maybe it just didn’t work better than doctors titrating the dose the old-fashioned way. Warfarin is a blood thinner. You can take blood and check how thin it is, usually measured with a number called INR. Most warfarin users are aiming for an INR between 2 and 3. So suppose (to oversimplify) you give your patient a dose of 3 mg, and find that the INR is 1.7. It seems like maybe the patient needs a little more warfarin, so you increase the dose to 4 mg. You take the INR later and it’s 2.3, so you declare victory and move on.

Maybe if you had a high-tech genetic test you could read the microscopic letters of the code of life itself, run the results through a supercomputer, and determine from the outset that 4 mg was the optimal dose. But all it would do is save you a little time.

There’s something similar going on with depression. Starting dose of Prozac is supposedly 20 mg, but I sometimes start it as low as 10 to make sure people won’t have side effects. And maximum dose is 80 mg. So there’s almost an order of magnitude between the highest and lowest Prozac doses. Most people stay on 20 to 40, and that dose seems to work pretty well.

Suppose I have a patient with a mutation that slows down their metabolism of Prozac; they effectively get three times the dose I would expect. I start them on 10 mg, which to them is 30 mg, and they seem to be doing well. I increase to 20, which to them is 60, and they get a lot of side effects, so I back down to 10 mg. Now they’re on their equivalent of the optimal dose. How is this worse than a genetic test which warns me against using Prozac because they have mutant Prozac metabolism?

Or suppose I have a patient with a mutation that *dectuples* Prozac levels; now there’s *no* safe dose. I start them on 10 mg, and they immediately report terrible side effects. I say “Yikes”, stop the Prozac, and put them on Zoloft, which works fine. How is this worse than a genetic test which says Prozac is bad for this patient but Zoloft is good?

Or suppose I have a patient with a mutation that makes them an ultrarapid metabolizer; no matter how much Prozac I give them, zero percent ever reaches their brain. I start them on Prozac 10 mg, nothing happens, go up to 20, then 40, then 60, then 80, nothing happens, finally I say “Screw this” and switch them to Zoloft. Once again, how is this worse than the genetic test?

(again, all of this is pretending that dose correlates with plasma levels correlates with efficacy in a way that’s hard to prove, but presumably necessary for any of this to be meaningful at all)

I expect the last two situations to be very rare; few people have orders-of-magnitude differences in metabolism compared to the general population. Mostly it’s going to be people who I would expect to need 20 of Prozac actually needing 40, or vice versa. But nobody has the slightest idea how to dose SSRIs anyway and we usually just try every possible dose and stick with the one that works. So I’m confused how genetic testing is supposed to make people do better or worse, as opposed to just needing a little more or less of a medication whose dosing is so mysterious that nobody ever knows how much anyone needs anyway.

As far as I can tell, this is why they need those pharmacodynamic genes like HTR2A and SLC6A4. Those represent real differences between antidepressants and not just changes in dose which we would get to anyway. I mean, you could still just switch antidepressants if your first one doesn’t work. But this would admittedly be hard and some people might not do it. *Everyone* titrates doses!

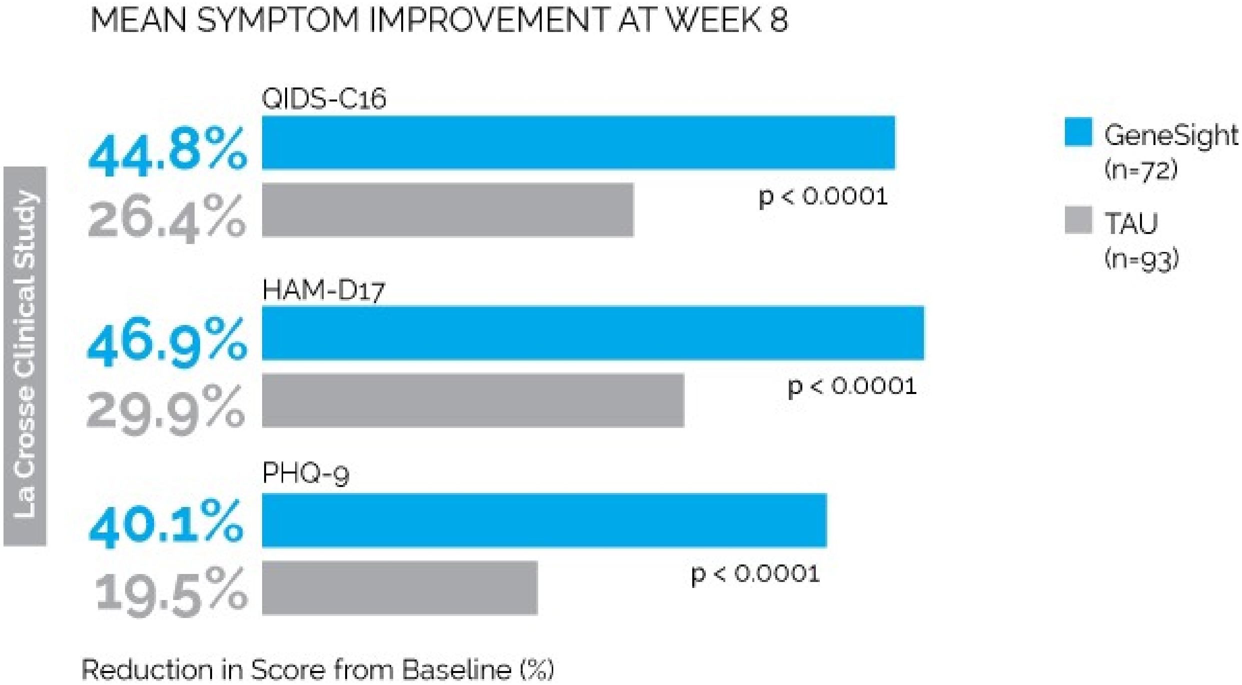
This is a fourth grain of salt and another reason why I’m wary about this idea.

VI.

Despite my skepticism, there are several studies showing impressive effects from pharmacogenomic antidepressant tests. Now that we’ve established some reasons to be doubtful, let’s look at them more closely.

GeneSight lists eight studies on its website here. Of note, all eight were conducted by GeneSight; as far as I know no external group has ever independently replicated any of their claims. The GeneSight rep I talked to said they're trying to get other scientists to look at it but haven't been able to so far. That's fair, but it's also fair for me to point out that studies by pharma companies are far more likely to find their products effective than studies by anyone else (OR = 4.05). I'm not going to start a whole other section for this, but let's call it a fifth grain of salt.

First is the LaCrosse Clinical Study. 114 depressed patients being treated at a clinic in Wisconsin received the GeneSight test, and the results were given to their psychiatrists, who presumably changed medications in accordance with the tests. Another 113 depressed patients got normal treatment without any genetic testing. The results were:



Taken from here, where you'll find much more along the same lines.

All of the combinations of letters and numbers are different depression tests. The blue bars are the people who got genotyped. The grey bars are the people who didn't. So we see that on every test, the people who got genotyped saw much greater improvement than the people who didn't. The difference in remission was similarly impressive; by 8 weeks, 26% of the genotyped group were depression-free as per QIDS-C16 compared to only 13% of the control group (p = 0.03)

How can we nitpick these results? A couple of things come to mind.

Number one, the study wasn't blinded. Everyone who was genotyped knew they were genotyped. Everyone who wasn't genotyped knew they weren't genotyped. I'm still not sure whether there's a significant placebo effect in depression (Hróbjartsson and Gøtzsche say no!), but it's at least worth worrying about.

Number two, the groups weren't randomized. I have no idea why they didn't randomize the groups, but they didn't. The first hundred-odd people to come in got put in the control group. The second hundred-off people got put in the genotype group. In accordance with the prophecy, there are various confusing and inexplicable differences between the two groups. The control group had more previous medication trials (4.7 vs. 3.6, p = 0.02). The intervention group had higher QIDS scores at baseline (16 vs. 17.5, p = 0.003). They even had different CYP2D6 phenotypes (p = 0.03). On their own these differences don't seem so bad, but they raise the question of why these groups were different at all and what other differences might be lurking.

Number three, the groups had very different numbers of dropouts. 42 people dropped out of the genotyped group, compared to 20 people from the control group. Dropouts made up about a quarter of the entire study population. The authors theorize that people were more likely to drop out of the genotype group than the control group because they'd promised to give the control group their genotypes at the end of the study, so they were sticking around to get their reward. But this means that people who were failing treatment were likely to drop out of the genotype group (making them look better) but stay in the control group (making them look worse). The authors do an analysis and say that this didn't affect things, but it's another crack in the study.

All of these are bad, but intuitively I don't feel like any of them should have been able to produce as dramatic an effect as they actually found. But I do have one theory about how this might have happened. Remember, these are all people who are on antidepressants already but aren't getting better. The intervention group's doctors get genetic testing results saying what antidepressant is best for them; the control group's doctors get nothing. So the intervention group's doctors will probably switch their patients' medication to the one the test says will be best, and the control group's doctors might just leave them on the antidepressant that's already not working. Indeed, we find that 77% of intervention group patients switched medications, compared to 44% of control group patients. So imagine if the genetic test didn't work at all. 77% of intervention group patients at least switch off their antidepressant that *definitely* doesn't work and onto one that *might* work; meanwhile, the control group mostly stays on the same old failed drugs.

Someone (maybe Carlat again?) mentioned how they should have controlled this study: give everyone a genetic test. Give the intervention group their own test results, and give the control group *someone else's* test results. If people do better on their own results than on random results, *then* we're getting somewhere.

Second is the Hamm Study, which is so similar to the above I’m not going to treat it separately.

Third is the Pine Rest Study. This one is, at least, randomized and *single*-blind. Single-blind means that the patients don’t know which group they’re in, but their doctors do; this is considered worse than double-blind (where neither patients nor doctors know) because the doctors’ subtle expectations could unconsciously influence the patients. But at least it’s something.

Unfortunately, the sample size was only 51 people, and the p-value for the main outcome was 0.28. They tried to salvage this with some subgroup analyses, but f**k that.

Fourth and fifth are two different meta-analyses of the above three studies, which is the lowest study-to-meta-analysis ratio I’ve ever seen. They find big effects, but “garbage in, garbage out”.

Sixth, there’s the Medco Study by Winner et al; I assume his name is a Big Pharma plot to make us associate positive feelings with him. This study is an attempt to prove cost-effectiveness. The GeneSight test costs $2000, but it might be worth it to insurers/governments if it makes people so much healthier that they spend less money on health care later. And indeed, it finds that GeneSight users spend $1036 less per year on medication than matched controls.

The details: they search health insurance databases for patients who were taking an psychiatric medication and then got GeneSight tests. Then they search the same databases for control patients for each; the control patients take the same psych med, have the same gender, are similar in age, and have the same primary psychiatric diagnosis. They end up with 2000 GeneSight patients and 10000 matched controls, whom they prove are definitely similar (even as a group) on the traits mentioned above. Then they follow all these people for a year and see how their medication spending changes.

The year of the study, the GeneSight patients spent on average $689 more on medications than they did the year before – unfortunate, but not entirely unexpected since apparently they're pretty sick. The control patients spent on average $1725 more. So their medication costs increased much more than the GeneSight patients. That presumably suggests GeneSight was doing a good job treating their depression, thus keeping costs down.

The problem is, this study wasn't randomized and so I see no reason to expect these groups to be comparable in any way. The groups were matched for sex, age, diagnosis, and one drug, but not on any other basis. And we have reason to think that they're not the same – after all, one group consists of people who ordered a little-known $2000 genetic test. To me, that means they're probably 1) rich, and 2) have psychiatrists who are really cutting-edge and into this kind of stuff. To be fair, I would expect both of those to drive *up* their costs, whereas in fact their costs were lower. But consider the possibility that rich people with good psychiatrists probably have less severe disease and are more likely to recover.

Here's some more evidence for this: of the ~$1000 cost savings, $300 was in psychiatric drugs and $700 was in non-psychiatric drugs. The article mentions that there's a mind-body connection and so maybe treating depression effectively will make people's non-psychiatric diseases get better too. This is *true*, but I think seeing that the effect of a psychiatric intervention is *stronger* on non-psychiatric than psychiatric conditions should at least *raise our suspicion* that we're actually seeing some confounder.

I cannot find anywhere in the study a comparison of how much money each group spent the year before the study started. This is a very strange omission. If these numbers were very different, that would clinch this argument.

Seventh is the Union Health Service study. They genotype people at a health insurance company who have already been taking a psychotropic medication. The genetic test either says that their existing medication is good for them (“green bin”), okay for them (“yellow bin”) or bad for them (“red bin”). Then they compare how the green vs. yellow vs. red patients have been doing over the past year on their medications. They find green and yellow patients mostly doing the same, but red patients doing very badly; for example, green patients have about five sick days from work a year, but red patients have about twenty.

I don’t really see any obvious flaws in this study, but there are only nine red patients, which means their entire results depend on an n = 9 experimental group.

Eighth is a study that just seems to be a simulation of how QALYs might change if you enter some parameters; it doesn’t contain any new empirical data.

Overall these studies show very impressive effects. While it’s possible to nitpick all of them, we have to remind ourselves that we can nitpick anything, even the best of studies, and do we really want to be that much of a jerk when these people have tested their revolutionary new product in five different ways, and every time it’s passed with flying colors aside from a few minor quibbles?

And the answer is: yes, I want to be *exactly* that much of a jerk. The history of modern medicine is one of pharmaceutical companies having *amazing* studies supporting their product, and maybe if you squint you can just *barely* find one or two little flaws but it hardly seems worth worrying about, and then a few years later it comes out that the product had no benefits whatsoever and caused everyone who took it to bleed to death. The reason for all those grains of salt above was to suppress our natural instincts toward mercy and cultivate the proper instincts to use when faced with pharmaceutical company studies, ie Cartesian doubt mixed with smoldering hatred.

VII.

I am totally not above introducing arguments from authority, and I've seen two people with much more credibility than myself look into this. The first is Daniel Carlat, Tufts professor and editor of *The Carlat Report*, a well-respected newsletter/magazine for psychiatrists. He writes a skeptical review of their studies, and finishes:

> If we were to hold the GeneSight test to the usual standards we require for making medication decisions, we'd conclude that there's very little reliable evidence that it works.

The second is John Ioannidis, professor of health research at Stanford and universally recognized expert on clinical evidence. He doesn't look at GeneSight in particular, but he writes of the whole pharmacogenomic project:

> For at least 3 years now, the expectation has been that newer platforms using exome or full-genome sequencing may improve the genome coverage and identify far more variants that regulate phenotypes of interest, including pharmacogenomic ones. Despite an intensive research investment, these promises have not yet materialized as of early 2013. A PubMed search on May 12, 2013, with (pharmacogenomics* OR pharmacogenetc*) AND sequencing yielded an impressive number of 604 items. I

scrutinized the 80 most recently indexed ones. The majority were either reviews/commentary articles with highly promising (if not zealot) titles or irrelevant articles. There was not a single paper that had shown robust statistical association between a newly discovered gene and some pharmacogenomics outcome, detected by sequencing. If anything, the few articles with real data, rather than promises, show that the task of detecting and validating statistically rigorous associations for rare variants is likely to be formidable. One comprehensive study sequencing 202 genes encoding drug targets in 14,002 individuals found an abundance of rare variants, with 1 rare variant appearing every 17 bases, and there was also geographic localization and heterogeneity. Although this is an embarrassment of riches, eventually finding which of these thousands of rare variants are most relevant to treatment response and treatment-related harm will be a tough puzzle to solve even with large sample sizes.

Despite these disappointing results, the prospect of applying pharmacogenomics in clinical care has not abided. If anything, it is pursued with continued enthusiasm among believers. But how much of that information is valid and is making any impact? […]

Before investing into expensive clinical trials for testing the new crop of mostly weak pharmacogenomic markers, a more radical decision is whether we should find some means to improve the yield of pharmacogenomics or just call it a day and largely abandon the field. The latter option sounds like a painfully radical solution, but on the other hand, we have already spent many thousands of papers and enormous funding, and the yield is so minimal. The utility yield seems to be even diminishing, if anything, as we develop more sophisticated genetic

> measurement techniques. Perhaps we should acknowledge that pharmacogenomics was a brilliant idea, we have learned some interesting facts to date, and we also found a handful of potentially useful markers, but industrial-level application of research funds may need to shift elsewhere.

I think the warning from respected authorities like these should add a sixth grain of salt to our rapidly-growing pile and make us feel a little bit better about rejecting the evidence above and deciding to wait.

VIII.

There's a thing I always used to hate about the skeptic community. Some otherwise-responsible scientist would decide to study homeopathy for some reason, and to everyone's surprise they would get positive results. And we would be uneasy, and turn to the skeptic community for advice. And they would say "Yeah, but homeopathy is stupid, so forget about this." And they would be *right*, but – what's the point of having evidence if you ignore it when it goes the wrong way? And what's the point in having experts if all they can do is say "this evidence went the wrong way, so let's ignore it"? Shouldn't we demand experts so confident in their understanding that they can explain to us why the new "evidence" is wrong? And as a corollary, shouldn't we demand experts who – if the world really was topsy-turvy and some crazy alternative medicine scheme did work – would be able to recognize that and tell us when to suspend our usual skepticism?

But at this point I'm starting to feel a deep kinship with skeptic bloggers. Sometimes we can figure out *possible* cracks in studies, and I think Part VI above did okay with that. But there will be cracks in even the best studies, and there will *especially* be cracks in studies done by small pharmaceutical companies who don't have the resources to do a major multicenter trial, and it's never clear when to use them as an excuse to reject the whole edifice versus when to let them pass as an unavoidable part of life. And because of how tough pharmacogenomics has proven so far, this is a case where I – after reading the warnings from Carlat and Ioannidis and the Italian team and everyone else – tentatively reject the edifice.

I hope later I kick myself over this. This *might* be the start of a revolutionary exciting new era in psychiatry. But I don't think I can believe it until independent groups have evaluated the tests, until other independent groups have replicated the work of the first independent groups, until everyone involved has publicly released their data (GeneSight didn't release any of the raw data for any of these studies!), and until our priors have been raised by equivalent success in other areas of pharmacogenomics.

Until then, I think it is a neat toy. I am glad some people are studying it. But I would not recommend spending your money on it if you don't have $2000 to burn (though I understand most people find ways to make their insurance or the government pay).

But if you just want to have fun with this, you can get a cheap approximation from 23andMe. Use the procedure outlined here to get your raw data, then look up rs6313 for the HTR2A polymorphism; (G,G) supposedly means more Paxil side effects (and maybe SSRI side effects in general). 23andMe completely dropped the ball on SLC6A4 and I would not recommend trying to look that one up. The cytochromes are much more complicated, but you might be able to piece some of it together from this page's links links to lists of alleles and related SNPs for each individual enzyme; also Promethease will do some of it for you automatically. Right now I think this process would produce pretty much 100% noise and be completely useless. But I’m not sure it would be *more* useless than the $2000 test. And if any of this pharmacogenomic stuff turns out to work, I hope some hobbyist automates the 23andMe-checking process and sells it as shareware for $5.

Book Review: Behavior – The Control Of Perception

[Epistemic status: I only partly understood this book and am trying to review it anyway as best I can]

I.

People complain that psychology is paradigmless; it never got its Darwin or Newton to tie everything together. Nowadays people are pretty relaxed about that; who needs paradigms when you can do n = 50 studies on a mildly interesting effect? But historically, there were all of these larger-than-life figures who were sure they'd found the paradigm, geniuses who founded schools which flourished for a while, made big promises, then either fizzled out or toned down their claims enough to be accepted as slightly kooky parts of the mainstream. Sigmund Freud. BF Skinner. Carl Rogers. And those are just the big ones close to the mainstream. Everyone from Ayn Rand to Scientology tried their hand at the paradigm-inventing business for a while.

Will Powers (whose name turns out to be pretty appropriate) lands somewhere in the middle of this pack. He was an engineer/inventor who specialized in cybernetic systems but wandered into psychology sometime in the sixties. He argued that everything in the brain made perfect sense if you understood cybernetic principles, and came up with a very complicated but all-encompassing idea called Perceptual Control Theory which explained thought, sensation and behavior. A few people paid attention, and his work was described as paradigm-shifting by no less of an expert on paradigm shifts than Thomas Kuhn. But in the end it never really went anywhere, psychology moved on, and nowadays only a handful of people continue research in his tradition.

Somehow I kept running into this handful, and they kept telling me to read Powers' book *Behavior: The Control Of Perception*, and I keep avoiding it. A few weeks ago I was driving down the road and I had a moment of introspection where I realized everything I was doing exactly fit Powers' theory, so I decided to give it a chance.

Powers specializes in *control systems*. The classic control system is a thermostat, which controls temperature. It has a reference point, let's say 70 degrees. If it gets much below 70 degrees, it turns on the heater until it's 70 again; if it gets much above 70 degrees, it turns on the air conditioner until it's 70 again. This is more complicated than it sounds, and there are other control systems that are even more complicated, but that's the principle. Perceptual Control Theory says that this kind of system is the basic unit of the human brain.

While I was driving on the highway a few weeks ago, I realized how much of what I do *is* perceptual control. For example, I was effortlessly maintaining the right distance from the car in front of me. If the car sped up a tiny bit, I would speed up a tiny bit. If the car slowed down a little bit, I would slow down a little bit. Likewise, I was maintaining the right angle relative to the road: if I found myself veering right, I would turn slightly to the left; if I found myself veering left, I would turn slightly to the right.

The theory goes further: while I'm in the car, I'm also operating as my own thermostat. I have a desired temperature: if I go below it, I'll turn on the heat, and if I go above it, I'll turn on the AC. I have a desired level of satiety: if I'm hungry, I'll stop and get something to eat; if I'm too full, there’s maybe not a *huge* amount I can do but I’ll at least stop eating. I have a desired level of light: if it’s too dark, I’ll turn on the lights; if it’s too bright I’ll put down the sun visor. I even have a desired angle to be sitting at: if I’m too far forward, I’ll relax and lean back a little bit; if I’m too far back, I’ll move forwards. All of this is so easy and automatic that I never think about it.

Powers’ theories go further. He agrees that my brain sets up a control system to keep my car the proper distance from the car in front of it. But how do I determine “the proper distance”? That quantity must be fed to the system by other parts of my brain. For example, suppose that the roads are icy and I know my brakes don’t work very well in the ice; I might keep a much further distance than usual. I’ll still be controlling the distance, I’ll just be controlling it *differently*. If the brain is control systems all the way down, we can imagine a higher-tier system controlling “accident risk” at some level (presumably low, or zero) feeding a distance level into a lower-tier system controlling car distance at whatever level it receives. We can even imagine higher systems than this. Suppose I’m depressed, I’ve become suicidal, I want to die in a car accident, but in order not to scandalize my family I have to let the accident happen sort of naturally. I have a top-level system controlling “desire to die” which tells a middle-level system controlling “accident risk” what level it should go at (high), which in turn tells a lower-tier system controlling “car distance” what level *it* should go at (very close).

It doesn’t even end there. My system controlling “car distance” is sending signals to a lower-tier system controlling muscle tension on my foot on the accelerator, giving it a new reference level (contracted muscles that push down on the accelerator really hard). Except this is an oversimplification, because everything that has to do with muscles is a million times more complicated than any reasonable person would think (at least until they play qwop) and so there’s actually a big hierarchy of control systems just going from “want to go faster” to “successfully tense accelerator-related muscles”.

II.

Actually, Powers is at his most convincing when he talks about these lower-level functions. At this point I think it’s pretty mainstream to say that muscle tension is set by a control system, with the Golgi tendon organs giving feedback and the spinal cord doing the calculations. Powers goes further (and I don’t know how mainstream this next part is, but I’m guessing at least somewhat), saying that this is a first-tier control system, which is itself controlled by a second-tier “direction” control system centered in the nuclei of the brainstem, which is itself controlled by a third-tier “position” control system centered in the cerebellum/thalamus/midbrain (a friendly amendment might add the basal ganglia, which Powers doesn’t seem to know much about).

If you stimulate certain parts of a cat’s midbrain, it will go into specific positions – for example, a position like it’s ready to pounce. So it seems like those areas “code for” position. But in order to have a neuron/area/whatever that codes for position, it needs to have hierarchical control over lots of lower-level things. For example, it needs to make sure the leg muscles are however tense they’re supposed to be in a pouncing position. So the third-tier position control system controls the second-tier direction control system at whatever level is necessary to make the second-tier direction control system control the first-tier muscle control system at whatever level is necessary to get the muscles in the right position.

The fourth- and fifth-tier systems, now well into the cortex (and maybe basal ganglia again) deal with sequences, eg “walking” or “playing a certain tune on the piano”. Once again, activating a fourth/fifth-tier system will activate this higher-level concept (“walking”), which alters the reference levels for a third-tier system (“getting into a certain position”), which alters a second-tier system (“moving in a certain direction”), which alterns a first-tier system (“tensing/relaxing muscles”).

Why do I like this theory so much? First, it correctly notes that (almost) the only thing the brain can actually *do* is change muscle tension. Yet we never think in terms of muscle tension. We don’t think “I am going to tense my thigh muscle, now untense it, now tense my ankle muscle, now…”, we just think “I’m going to walk”. Heck, half the time we don’t even think *that*, we think “I’m just going to go to the fridge” and the walking happens automatically. On the other hand, if we really want, we *can* consciously change our position, the level of tension in a certain muscle, etc. It’s just that usually we deal in higher-level abstractions that automatically carry all the lower ones along with them.

Second, it explains the structure of the brain in a way I haven’t seen other things do. I always hear neuroscientists talk about “this nucleus relays signals to that nucleus” or “this structure is a way station for this other structure”. Spend too much time reading that kind of stuff, and you start to think of the brain as a giant relay race, where the medulla passes signals onto the thalamus which passes it to the basal ganglia which passes it to the frontal lobe and then, suddenly, thought! The obvious question there is “why do you have so many structures that just relay things to other structures?” Sometimes neuroscientists will say “Well, some processing gets done here”, or even better “Well, this system modulates that system”, but they’re always very vague on what exactly that means. Powers' hierarchy of fifth-tier systems passing their calculations on to fourth-tier systems and so on is exactly the sort of thing that would make sense of all this relaying. My guess is every theory of neuroscience has something at least this smart, but I'd never heard it explained this well before.

Third, it's the clearest explanation of tremors I've ever heard. Consider the thermostat above. When the temperature gets below 65, it turns on the heat until the temperature gets above 70, then stops, then waits as the hot air leaks out through the window or whatever and it's 65 again, then turns on the heat again. If we chart temperature in a room with a thermostat, it will look sort of like a sine wave or zigzag with regular up/down motions. This is a basic principle of anything being controlled by a less-than-perfect control system. Our body has microtremors all the time, but when we get brain damage or some other problem, a very common symptom is noticeable tremors. These come in many different varieties that give clues to the level of brain damage and which doctors are just told to memorize. Powers actually explains them:

> When first-order systems become unstable, as when muscles exert too much effort), clonus oscillations are seen, at roughly ten cycles per second. Second-order instability, as in the tremors of Parkinsonism, involves groups of muscles and is of lower frequency, around three cycles per second or so. Third-order instability is slower stilll, slow enough that it can be characterized as "purpose tremor" or "over-correction". Certain cerebellar damage due to injury or disease can result in over- and under-shooting the mark during actions such as reaching out to grasp something, either in a continuous self-sustained oscillation or a slowly decrasing series of alternating movements.

This isn’t perfect – for example, Parkinsonian tremor is usually caused by damage to the basal ganglia and the cortex, which is really hard to square with Powers’ claim that it’s caused by damage to second-tier systems in the medulla. But after reading this, it’s really hard not to think of tremors as failures in control systems, or of the different types of tremor as failures in different levels of control system. For example, athetoid tremors are weird, seemingly purposeful, constant twisting movements caused by problems in the thalamus or some related system; after reading Powers, it’s impossible for me not to think of them as failures in third-order control systems. This becomes especially clear if we compare to Powers’ constant foil/nemesis, the Behaviorists. Stick to a stimulus-response paradigm, and there’s no reason damaged brains should make weird twisting movements all the time. On a control-systems paradigm, it’s *obvious* that that would happen.

There are occasional claims that perceptual control theory can predict certain things about muscles and coordination better than other theories, sometimes with absurdly high accuracy of like r = 0.9 or something. Powers makes some of these claims in the book, but I can’t check them because I don’t have the original data he worked with and I don’t know how to calculate cybernetic control system outputs. But the last time I saw someone bring up one of these supposed experiments it was thoroughly shot down by people who knew more statistics. And I found a blog post where somebody who knows a lot about intricacies of muscle movement says PCT can predict some things but not much better than competing theories. In terms of predicting very specific things about human muscular movement its record seems to be kind of so-so.

III.

And I start to get very skeptical when Powers moves to higher-tier control systems. His sixth tier is “relationships”, seventh is “programs”, eighth is “principles”, and ninth is “systems”. Although these tiers receive just as many pages as the earlier ones, they start sounding very abstract and they correlate a lot less well with anatomy. I understand the urge to postulate them – if you’ve already decided that the fundamental unit of the brain is the control system, why not try to explain things with control systems all the way up? – but it becomes kind of a stretch. It’s easy to see what it means to control the distance between me and the car in front of me; it’s harder to see what it means to control for “communism” or “honesty” or things like that.

I *think* the way things are supposed to work is like this. A ninth-tier system controls a very abstract concept like “communism”. So suppose you are a communist; that means your internal communism-thermostat is set to maintain your communism at a high level. That propagates down to eighth-tier principles, which are slightly less abstract concepts like “greed”; maybe your ninth-tier communism-thermostat sets your eighth-tier greed thermostat to a very low temperature because communists aren’t supposed to be greedy. Your eighth-tier greed thermostat affects levels of seventh-tier logical programs like “going to work and earning money” and “giving to charity”. I’m not really sure how the sixth-tier fits into this example, but let’s suppose that your work is hammering things. Then the fifth-tier system moves your muscles in the right sequence to hammer things, and so on with all the lower tiers as above.

Sometimes these control systems come into contact with each other. For example, suppose that along with my ninth-tier system controlling “communism”, I also have a ninth-tier system controlling “family values”; I am both an avowed communist and a family man. My family values system thinks that it’s important that I earn enough to provide for my family, so while my communism-system is trying to input a low reference level for my greed-thermostat, my family-values-system is trying to input a high one. Powers gets into some really interesting examples of what happens in real industrial cybernetic systems when two opposing high-level control systems get in a fight, and thinks this is the source of all human neurosis and akrasia. I think he later wrote a self-help book based around this (hence the nominative determinism). I am not very convinced.

Am I strawmanning this picture? I’m not sure. I think one testable consequence of it is supposed to be that if we’re really controlling for communism, in the cybernetic control system sense, then we should be able to test for that. For example, hide Lenin’s pen and paper so that he can’t write communist pamphlets, and he should start doing some other communist thing more in order to make up for it and keep his level of communism constant. I think some perceptual control theory people believe this is literally true, and propose experimental tests (or at least thought experiment tests) of perceptual control theory along these lines. This seems sketchy to me, on the grounds that if Lenin didn’t start doing other stuff, we could just say that communism wasn’t truly what he was controlling.

That is, suppose I notice Lenin eating lots of chocolate every day. I theorize that he’s controlling for chocolate, and so if I disturb the control system by eg shutting down his local chocolate store, he’ll find a way to restore equilibrium, eg by walking further to a different store. But actually, when I shut down his local chocolate store, he just eats less chocolate. In reality, he was controlling his food intake (as we all do; that’s what an obesity set point is) and when he lost access to chocolate, maybe he ate cupcakes instead and did fine.

In the same way, maybe we only think Lenin is controlling for communism, but he’s actually controlling for social status, and being a communist revolutionary is a good way to gain social status. So if we make it too hard for him to be a communist revolutionary, eg by taking away his pen and paper, maybe he’ll become a rock star instead and end up with the same level of social status.

This sort of thing seems so universal that as far as I can tell it makes these ideas of higher-tier control systems unproveable and unfalsifiable.

If there’s any point to them at all, I think it’s the way they express the same interesting phenomenological truth as the muscle movement tiers: we switch effortlessly between concentrating on low-level concepts and high-level concepts that make the low-level ones automatic. For example, I think “driving” is a good example of Powers’ seventh tier, “programs” – it involves a predictable flowchart-like set of actions to achieve a simple goal. “The distance between me and the car in front of me” is a sixth-tier system, a “relationship”. When I’m driving (focusing on my seventh-tier system), I don’t consciously think at all about maintaining the right distance with the car in front of me. It just happens. This is really interesting in a philosophy of consciousness sense, and Powers actually gets into qualia a bit and says some things that seem a lot wiser and more moving-part-ful than most people on the subject.

It does seem like there’s something going on where my decision to drive activates a lot of carefully-trained subsystems that handle the rest of it automatically, and that there’s probably some neural correlate to it. But I don’t know whether control systems are the right way to think about this, and I definitely don’t know whether there’s a sense in which “communism” is a control system.

IV.

There are also some sections about things like learning and memory, which looks suspiciously like flowcharts of control systems with boxes marked “LEARNING” and “MEMORY” in them.

But I realized halfway through that I was being too harsh. Perceptual control theory wasn’t quite a proposal for a new paradigm out of nowhere. It was a reaction to Behaviorism, which was still the dominant paradigm when Powers was writing. His “everything is a control system” is an attempt to improve on “everything is stimulus-response”, and it really does.

For example, his theory of learning involves reward and punishment, where reward is reducing the error in a control system and punishment is increasing it. That is, suppose that you’re controlling temperature, and it’s too hot out. A refreshing cool glass of water would be an effective reward (since it brings you closer to your temperature reference level), and setting your hand on fire would be an effective punishment (since it brings you further from your temperature reference level). Powers notes that this explains many things Behaviorism can’t. For example, they like to talk about how sugar water is a reward. But eventually rats get tired of sugar water and stop drinking it. So it seems that sugar water isn’t a reward per se; it’s more like reducing error in your how-much-sugar-water-should-I-have-and-did-I-already-have-the-right-amount system is the reward. If your optimal level of sugar water per day is 10 ml, then anything up to 10 ml will be a reward, and after that it will stop being attractive / start being a punishment.

As a “theory of learning”, this is sort of crappy, in that I was expecting stuff about Hebb and connectionism and how memories are stored in the brain. But if you're living in an era where everybody thinks "The response to a stimulus is predictable through patterns of reward and punishment" is an A+++ Nobel-Prize-worthy learning theory, then perceptual control-based theories of learning start sounding pretty good.

So I guess it's important to see this as a product of its times. And I don't understand those times – why Behaviorism ever seemed attractive is a mystery to me, maybe requiring more backwards-reading than I can manage right now.

How useful is this book? I guess that depends on how metaphorical you want to be. Is the brain a control system? I don't know. Are police a control system trying to control crime? Are police a "response" to the "stimulus" of crime? Is a stimulus-response pairing a control system controlling for the quantity of always making sure the stimulus has the response? I think it's interesting and helpful to think of some psychological functions with these metaphors. But I'm not sure where to go from there. I think maybe there are some obvious parallels, maybe even parallels that bear fruit in empirical results, in lower level systems like motor control. Once you get to high-level systems like communism or social desirability, I'm not sure we're doing much better than the police-as-control-system metaphor. Still, I think that it's potentially a useful concept to have.

Book Review: Seeing Like A State

I.

Seeing Like A State is the book G.K. Chesterton would have written if he had gone into economic history instead of literature. Since he didn’t, James Scott had to write it a century later. The wait was worth it.

Scott starts with the story of “scientific forestry” in 18th century Prussia. Enlightenment rationalists noticed that peasants were just cutting down whatever trees happened to grow in the forests, *like a chump*. They came up with a better idea: clear all the forests and replace them by planting identical copies of Norway spruce (the highest-lumber-yield-per-unit-time tree) in an evenly-spaced rectangular grid. Then you could just walk in with an axe one day and chop down like a zillion trees an hour and have more timber than you could possibly ever want.

This went poorly. The impoverished ecosystem couldn’t support the game animals and medicinal herbs that sustained the surrounding peasant villages, and they suffered an economic collapse. The endless rows of identical trees were a perfect breeding ground for plant diseases and forest fires. And the complex ecological processes that sustained the soil stopped working, so after a generation the Norway spruces grew stunted and malnourished. Yet for some reason, everyone involved got promoted, and “scientific forestry” spread across Europe and the world.

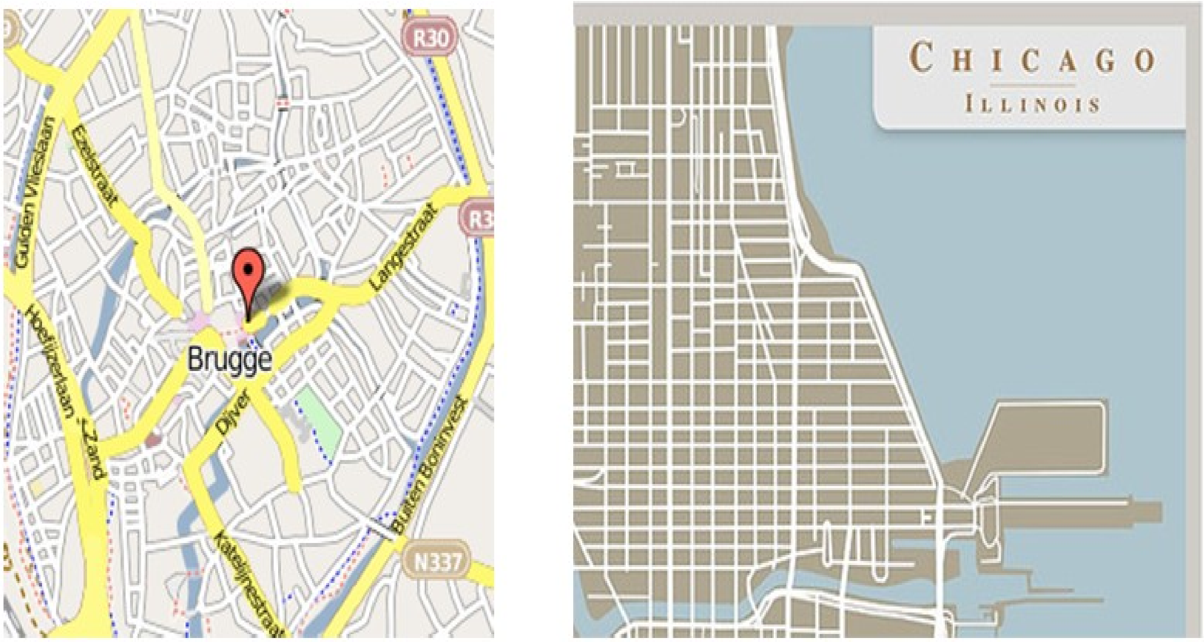
And this pattern repeats with suspicious regularity across history, not just in biological systems but also in social ones.

Natural organically-evolved cities tend to be densely-packed mixtures of dark alleys, tiny shops, and overcrowded streets.

Modern scientific rationalists came up with a better idea: an evenly-spaced rectangular grid of identical giant Brutalist apartment buildings separated by wide boulevards, with everything separated into carefully-zoned districts. Yet for some reason, whenever these new rational cities were built, people hated them and did everything they could to move out into more organic suburbs. And again, for some reason the urban planners got promoted, became famous, and spread their destructive techniques around the world.

Ye olde organically-evolved peasant villages tended to be complicated confusions of everybody trying to raise fifty different crops at the same time on awkwardly shaped cramped parcels of land. Modern scientific rationalists came up with a better idea: giant collective mechanized farms growing purpose-bred high-yield crops and arranged in (say it with me) evenly-spaced rectangular grids. Yet for some reason, these giant collective farms had lower yields per acre than the old traditional methods, and wherever they arose famine and mass starvation followed. And again, for some reason governments continued to push the more “modern” methods, whether it was socialist collectives in the USSR, big agricultural corporations in the US, or sprawling banana plantations in the Third World.

Traditional lifestyles of many East African natives were nomadic, involving slash-and-burn agriculture in complicated jungle terrain according to a bewildering variety of ad-hoc rules. Modern scientific rationalists in African governments (both colonial and independent) came up with a better idea – resettlement of the natives into villages, where they could have modern amenities like schools, wells, electricity, and evenly-spaced rectangular grids. Yet for some reason, these villages kept failing: their crops died, their economies collapsed, and their native inhabitants disappeared back into the jungle. And again, for some reason the African governments kept trying to bring the natives back and make them stay, even if they had to blur the lines between villages and concentration camps to make it work.


R30
Gulden Vlieslaan
Ezelstraat
Langestraat
Hoefijzerlaan
Brugge
Dijver
't Zand
Buiten Boninvest
Katelijnestraat
N337
CHICAGO
ILLINOIS

A favorite Seeing Like A State *image: a comparison of street maps for Bruges (a premodern organic city) with Chicago (a modern planned city)*

Why did all of these schemes fail? And more importantly, why were they celebrated, rewarded, and continued, even when the fact of their failure became too obvious to ignore? Scott gives a two part answer.

The first part of the story is High Modernism, an aesthetic taste masquerading as a scientific philosophy. The High Modernists claimed to be about figuring out the most efficient and high-tech way of doing things, but most of them knew little relevant math or science and were basically just LARPing being rational by placing things in evenly-spaced rectangular grids.

But the High Modernists were pawns in service of a deeper motive: the centralized state wanted the world to be “legible”, ie arranged in a way that made it easy to monitor and control. An intact forest might be more productive than an evenly-spaced rectangular grid of Norway spruce, but it was harder to legislate rules for, or assess taxes on.

The state promoted the High Modernists’ platitudes about The Greater Good as cover, in order to implement the totalitarian schemes they wanted to implement anyway. The resulting experiments were usually failures by the humanitarian goals of the Modernists, but resounding successes by the command-and-control goals of the state. And so we gradually transitioned from systems that were messy but full of fine-tuned hidden order, to ones that were barely-functional but really easy to tax.

II.

Suppose you’re a premodern king, maybe one of the Louises who ruled France in the Middle Ages. You want to tax people to raise money for a Crusade or something. Practically everyone in your kingdom is a peasant, and all the peasants produce is grain, so you’ll tax them in grain. Shouldn’t be too hard, right? You’ll just measure how many pints of grain everyone produces, and…

> The pint in eighteenth-century Paris was equivalent to 0.93 liters, whereas in Seine-en-Montane it was 1.99 liters and in Precy-sous-Thil, an astounding 3.33 liters. The aune, a measure of length used for cloth, varied depending on the material(the unit for silk, for instance, was smaller than that for linen) and across France there were at least seventeen different aunes.

Okay, this is stupid. Just give everybody evenly-sized baskets, and tell them that baskets are the new unit of measurement.

> Virtually everywhere in early modern Europe were endless micropolitics about how baskets might be adjusted through wear, bulging, tricks of weaving, moisture, the thickness of the rim, and so on. In some areas the local standards for the bushel and other units of measurement were kept in metallic form and placed in the care of a trusted official or else literally carved into the stone of a church or the town hall. Nor did it end there. How the grain was to be poured (from shoulder height, which packed it somewhat, or from waist height?), how damp it could be, whether the container could be shaken down, and finally, if and how it was to be leveled off when full were subjects of long and bitter controversy.

Huh, this medieval king business is harder than you thought. Maybe you can just leave this problem to the feudal lords?

> Thus far, this account of local measurement practices risks giving the impression that, although local conceptions of distance, area, volume, and so on were different from and more varied than the unitary abstract standards a state might favor, they were nevertheless aiming at objective accuracy. This impression would be false. […]
>
> A good part of the politics of measurement sprang from what a contemporary economist might call the “stickiness” of feudal rents. Noble and clerical claimants often found it difficult to increase feudal dues directly; the levels set for various charges were the result of long struggle, and even a small increase above the customary level was viewed as a threatening breach of tradition. Adjusting the measure, however, represented a roundabout way of achieving the same end.

> The local lord might, for example, lend grain to peasants in smaller baskets and insist on repayment in larger baskets. He might surreptitiously or even boldly enlarge the size of the grain sacks accepted for milling (a monopoly of the domain lord) and reduce the size of the sacks used for measuring out flour; he might also collect feudal dues in larger baskets and pay wages in kind in smaller baskets. While the formal custom governing feudal dues and wages would thus remain intact (requiring, for example, the same number of sacks of wheat from the harvest of a given holding), the actual transaction might increasingly favor the lord. The results of such fiddling were far from trivial. Kula estimates that the size of the bushel (boisseau) used to collect the main feudal rent (taille) increased by one-third between 1674 and 1716 as part of what was called the reaction feodale.

Okay, but nobody's going to make *too* big a deal about this, right?

> This sense of victimization [over changing units of measure] was evident in the cahiers of grievances prepared for the meeting of the Estates General just before the Revolution. […] In an unprecedented revolutionary context where an entirely new political system was being created from first principles, it was surely no great matter to legislate uniform weights and measures. As the revolutionary decree read "The centuries old dream of the masses of only one just measure has come true! The Revolution has given the people the meter!"

Okay, so apparently (you think to yourself as you are being led to the guillotine), it was a big deal after all.

Maybe you *shouldn't* have taxed grain. Maybe you should tax land. After all, it's the land that grows the grain. Just figure out how much land everybody owns, and you can calculate some kind of appropriate tax from there.

So, uh, peasant villagers, how much land does each of you own?

> A hypothetical case of customary land tenure practices may help demonstrate how difficult it is to assimilate such practices to the barebones scheme of a modern cadastral map [land survey suitable for tax assessment] [...]
>
> Let us imagine a community in which families have usufruct rights to parcels of cropland during the main growing season. Only certain crops, however, may be planted, and every seven years the usufruct land is distributed among resident families according to each family's size and its number of able-bodied adults. After the harvest of the main-season crop, all cropland reverts to common land where any family may glean, graze their fowl and livestock, and even plant quickly maturing, dry-season crops. Rights to graze fowl and livestock on pasture-land held in common by the village is extended to all local families, but the number of animals that can be grazed is restricted according to family size, especially in dry years when forage is scarce. Families not using their grazing rights can give them to other villagers but not to outsiders. Everyone has the right to gather firewood for normal family needs, and the village blacksmith and baker are given larger allotments. No commercial sale from village woodlands is permitted.
>
> Trees that have been planted and any fruit they may bear are the property of the family who planted them, no matter where they are now growing. Fruit fallen from

> such tree, however, is the property of anyone who gathers it. When a family fells one of its trees or a tree is felled by a storm, the trunk belongs to the family, the branches to the immediate neighbors, and the “tops” (leaves and twigs) to any poorer villager who carries them off. Land is set aside for use or leasing out by widows with children and dependents of conscripted males. Usufruct rights to land and trees may be let to anyone in the village; the only time they may be let to someone outside the village is if no one in the community wishes to claim them. After a crop failure leading to a food shortage, many of these arrangements are readjusted.

You know what? I’m just going to put you all down as owning ten. Ten land. Everyone okay with that? Cool. Let’s say ten land for everyone and just move on to the next village.

> Novoselok village had a varied economy of cultivation, grazing, and forestry…the complex welter of strips was designed to ensure that each village household received a strip of land in every ecological zone. An individual household might have as many as ten to fifteen different plots constituting something of a representative sample of the village’s ecological zones and microclimates. The distribution spread a family’s risks prudently, and from time to time the land was reshuffled as families grew or shrunk…The strips of land were generally straight and parallel so that a readjustment could be made by moving small stakes along just one side of a field, without having to think of areal dimensions. Where the other side of the field was not parallel, the stakes could be shifted to compensate for the fact that the strip lay toward the narrower or wider end of the field. Irregular fields were divided, not according to area, but according to yield.

…huh. Maybe this isn't going to work. Let's try it the other way around. Instead of mapping land, we can just get a list with the name of everyone in the village, and go from there.

> Only wealthy aristocrats tended to have fixed surnames…Imagine the dilemma of a tithe or capitation-tax collector [in England] faced with a male population, 90% of whom bore just six Christian names (John, William, Thomas, Robert, Richard, and Henry).

Okay, fine. That won't work either. Surely there's *something* else we can do to assess a tax burden on each estate. Think outside the box, scrape the bottom of the barrel!

> The door-and-window tax established in France [in the 18th century] is a striking case in point. Its originator must have reasoned that the number of windows and doors in a dwelling was proportional to the dwelling's size. Thus a tax assessor need not enter the house or measure it, but merely count the doors and windows.
>
> As a simple, workable formula, it was a brilliant stroke, but it was not without consequences. Peasant dwellings were subsequently designed or renovated with the formula in mind so as to have as few openings as possible. While the fiscal losses could be recouped by raising the tax per opening, the long-term effects on the health of the population lasted for more than a century.

Close enough.

III.

The moral of the story is: premodern states had very limited ability to tax their citizens effectively. Along with the problems mentioned above – nonstandardized measurement, nonstandardized property rights, nonstandardized personal names – we can add a few others. At this point national languages were a cruel fiction; local “dialects” could be as different from one another as eg Spanish is from Portuguese, so villagers might not even be able to *understand* the tax collectors. Worst of all, there was no such thing as a census in France until the 17th century, so there wasn’t even a good idea of how many people or villages there were.

Kings usually solved this problem by leaving the tax collection up to local lords, who presumably knew the idiosyncracies of their own domains. But one step wasn’t always enough. If the King only knew Dukes, and the Dukes only knew Barons, and the Barons only knew village headmen, and it was only the village headmen who actually knew anything about the peasants, then you needed a four-step chain to get any taxes. Each link in the chain had an incentive to collect as much as they could and give up as little as they could get away with. So on the one end, the peasants were paying backbreaking punitive taxes. And on the other, the Royal Treasurer was handing the King half a loaf of moldy bread and saying “Here you go, Sire, apparently this is all the grain in France.”

So from the beginning, kings had an incentive to make the country “legible” – that is, so organized and well-indexed that it was easy to know everything about everyone and collect/double-check taxes. Also from the beginning, nobles had an incentive to frustrate the kings so that they wouldn’t be out of a job. And commoners, who figured that anything which made it easier for the State to tax them and interfere in their affairs was bad news, usually resisted too.

Scott doesn’t bring this up, but it’s interesting reading this in the context of Biblical history. It would seem that whoever wrote the Bible was not a big fan of censuses. From 1 Chronicles 21:

Satan rose up against Israel and caused David to take a census of the people of Israel. So David said to Joab and the commanders of the army, “Take a census of all the people of Israel—from Beersheba in the south to Dan in the north—and bring me a report so I may know how many there are.”

But Joab replied, “May the Lord increase the number of his people a hundred times over! But why, my lord the king, do you want to do this? Are they not all your servants? Why must you cause Israel to sin?”

But the king insisted that they take the census, so Joab traveled throughout all Israel to count the people. Then he returned to Jerusalem and reported the number of people to David. There were 1,100,000 warriors in all Israel who could handle a sword, and 470,000 in Judah. But Joab did not include the tribes of Levi and Benjamin in the census because he was so distressed at what the king had made him do.

God was very displeased with the census, and he punished Israel for it. Then David said to God, “I have sinned greatly by taking this census. Please forgive my guilt for doing this foolish thing.” Then the Lord spoke to Gad, David’s seer. This was the message: “Go and say to David, ‘This is what the Lord says: I will give you three choices. Choose one of these punishments, and I will inflict it on you.’”

So Gad came to David and said, “These are the choices the Lord has given you. You may choose three years of famine, three months of destruction by the sword of your enemies, or three days of severe plague as the angel of the Lord brings devastation throughout the land of Israel. Decide what answer I should give the Lord who sent me.”

> “I’m in a desperate situation!” David replied to Gad. “But let me fall into the hands of the Lord, for his mercy is very great. Do not let me fall into human hands.”
>
> So the Lord sent a plague upon Israel, and 70,000 people died as a result.

(related: Scott examined some of the same data about Holocaust survival rates as *Eichmann In Jerusalem*, but made them make a lot more sense: the greater the legibility of the state, the worse for the Jews. One reason Jewish survival in the Netherlands was so low was because the Netherlands had a very accurate census of how many Jews there were and where they lived; sometimes officials saved Jews by literally burning census records).

Centralized government projects promoting legibility have always been a two-steps-forward, one-step back sort of thing. The government very gradually expands its reach near the capital where its power is strongest, to peasants whom it knows will try to thwart it as soon as its back is turned, and then if its decrees survive it pushes outward toward the hinterlands.

Scott describes the spread of surnames. Peasants didn’t like permanent surnames. Their own system was quite reasonable for them: John the baker was John Baker, John the blacksmith was John Smith, John who lived under the hill was John Underhill, John who was really short was John Short. The same person might be John Smith and John Underhill in different contexts, where his status as a blacksmith or place of origin was more important.

But the government insisted on giving everyone a single permanent name, unique for the village, and tracking who was in the same family as whom. Resistance was intense:

> What evidence we have suggests that second names of any kind became rare as distance from the state’s fiscal reach increased. Whereas one-third of the households in Florence declared a second name, the proportion dropped to one-fifth for secondary towns and to one-tenth in the countryside. It was not until the seventeenth century that family names crystallized in the most remote and poorest areas of Tuscany – the areas that would have had the least contact with officialdom. […]
>
> State naming practices, like state mapping practices, were inevitably associated with taxes (labor, military service, grain, revenue) and hence aroused popular resistance. The great English peasant rising of 1381 (often called the Wat Tyler Rebellion) is attributed to an unprecedented decade of registration and assessments of poll taxes. For English as well as for Tuscan peasants, a census of all adult males could not but appear ominous, if not ruinous.

The same issues repeated themselves a few hundred years later when Europe started colonizing other continents. Again they encountered a population with naming systems they found unclear and unsuitable to taxation. But since colonial states had more control over their subjects than the relatively weak feudal monarchies of the Middle Ages, they were able to deal with it in one fell swoop, sometimes comically so:

> Nowhere is this better illustrated than in the Philippines under the Spanish. Filipinos were instructed by the decree of November 21, 1849 to take on permanent Hispanic surnames. […]
>
> Each local official was to be given a supply of surnames sufficient for his jurisdiction, “taking care that the distribution be made by letters of the alphabet.” In

> practice, each town was given a number of pages from the alphabetized [catalog], producing whole towns with surnames beginning with the same letter. In situations where there has been little in-migration in the past 150 years, the traces of this administrative exercise are still perfectly visible across the landscape. “For example, in the Bikol region, the entire alphabet is laid out like a garland over the provinces of Albay, Sorsogon, and Catanduanes which in 1849 belonged to the single jurisdiction of Albay. Beginning with A at the provincial capital, the letters B and C mark the towns along the cost beyond Tabaco to Wiki. We return and trace along the coast of Sorosgon the letters E to L, then starting down the Iraya Valley at Daraga with M, we stop with S to Polangui and Libon, and finish the alphabet with a quick tour around the island of Catanduas.
>
> The confusion for which the decree is the antidote is largely that of the administrator and the tax collector. Universal last names, they believe, will facilitate the administration of justice, finance, and public order as well as make it simpler for prospective marriage partners to calculate their degree of consanguinity. For a utilitarian state builder of [Governor] Claveria’s temper, however, the ultimate goal was a complete and legible list of subjects and taxpayers.

This was actually a lot less cute and funny than the alphabetization makes it sound:

> What if the Filipinos chose to ignore their new last names? This possibility had already crossed Claveria’s mind, and he took steps to make sure that the names would stick. Schoolteachers were ordered to forbid their students to address or even know one another by any name except the officially inscribed family name. Those

> teachers who did not apply the rule with enthusiasm were to be punished. More efficacious perhaps, given the minuscule school enrollment, was the proviso that forbade priests and military and civil officials from accepting any document, application, petition, or deed that did not use the official surnames. All documents using other names would be null and void

Similar provisions ensured the replacement of local dialects with the approved national language. Students were only allowed to learn the national language in school and were punished for speaking in vernacular. All formal documents had to be in the national language, which meant that peasants who had formally been able to manage their own legal affairs had to rely on national-language-speaking intermediaries. Scott talks about the effect in France:

> One can hardly imagine a more effective formula for immediately devaluing local knowledge and privileging all those who had mastered the official linguistic code. It was a gigantic shift in power. Those at the periphery who lacked competence in French were rendered mute and marginal. They were now in need of a local guide to the new state culture, which appeared in the form of lawyers, notaries, schoolteachers, clerks, and soldiers.

IV.

So the early modern period is defined by an uneasy truce between states who want to be able to count and standardize everything, and citizens who don’t want to let them. Enter High Modernism. Scott defines it as

> A strong, one might even say muscle-bound, version of the self-confidence about scientific and technical progress, the expansion of production, the growing satisfaction of human needs, the mastery of nature

> (including human nature), and above all, the rational design of social order commensurate with the scientific understanding of natural laws

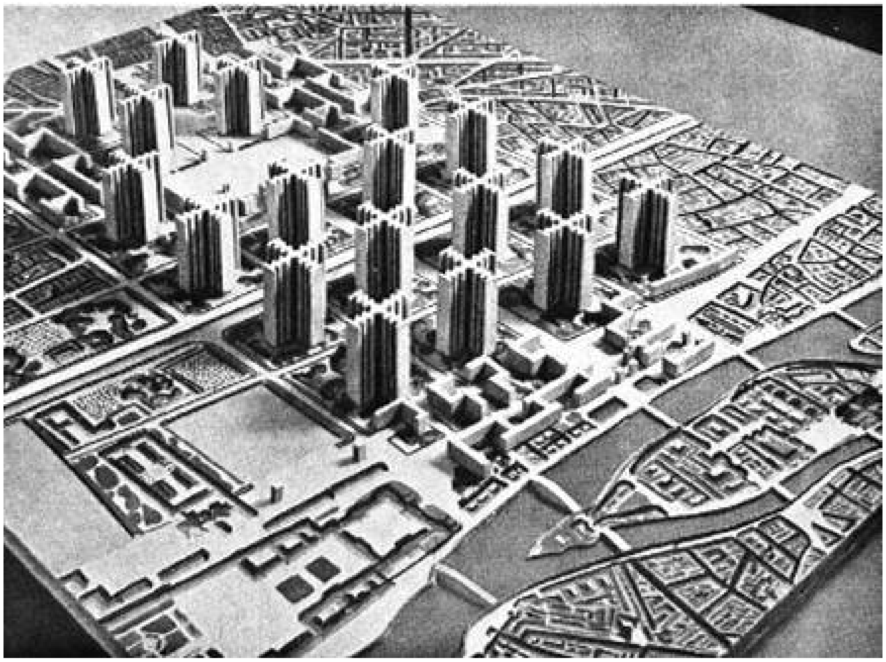
…which is just a *bit* academic-ese for me. An extensional definition might work better: standardization, Henry Ford, the factory as metaphor for the best way to run everything, conquest of nature, New Soviet Man, people with college degrees knowing better than you, wiping away the foolish irrational traditions of the past, Brave New World, everyone living in dormitories and eating exactly 2000 calories of Standardized Food Product (TM) per day, anything that is For Your Own Good, gleaming modernist skyscrapers, The X Of The Future, complaints that the unenlightened masses are resisting The X Of The Future, demands that if the unenlightened masses reject The X Of The Future they must be re-educated For Their Own Good, and (of course) evenly-spaced rectangular grids.

(maybe the best definition would be “everything G. K. Chesterton didn’t like.”)

It sort of sounds like a Young Adult Dystopia, but Scott shocked me with his research into just how strong this ideology was around the turn of the last century. Some of the greatest early 20th-century thinkers were High Modernist to the point of self-parody, the point where a Young Adult Dystopian fiction writer would start worrying they were laying it on a little too thick.

The worst of the worst was Le Corbusier, the French artist/intellectual/architect. The Soviets asked him to come up with a plan to redesign Moscow. He came up with one: kick out everyone, bulldoze the entire city, and redesign it from scratch upon rational principles. For example, instead of using other people’s irrational systems of measurement, they would use a new measurement system invented by Le

Corbusier himself, called Modulor, which combined the average height of a Frenchman with the Golden Ratio.



Also, evenly-spaced rectangular grids may have been involved.

The Soviets decided to pass: the plan was too extreme and destructive of existing institutions *even for Stalin*. Undeterred, Le Corbusier changed the word “Moscow” on the diagram to “Paris”, then presented it to the French government (who also passed). Some aspects of his design eventually ended up as Chandigarh, India.

A typical building in Chandigarh. The Soviets and French must have been kicking themselves when they realized what they'd missed out on.

Le Corbusier was challenged on his obsession with keeping his plan in the face of different local conditions, pre-existing structures, residents who might want a say in the matter, et cetera. Wasn't it kind of dictatorial? He replied that:

> The despot is not a man. It is the Plan. The correct, realistic, exact plan, the one that will provide your solution once the problem has been posited clearly, in its entirety, in its indispensable harmony. This plan has been drawn up well away from the frenzy in the mayor's office or the town hall, from the cries of the electorate or the laments of society's victims. It has been drawn up by serene and lucid minds. It has taken account of nothing but human truths. It has ignored all current regulations, all existing usages, and channels. It has not considered whether or not it could be carried out with the constitution now in force. It is a biological creation

> destined for human beings and capable of realization by modern techniques.

What was so great about this “biological creation” of “serene and lucid minds”? It…might have kind of maybe been evenly-spaced rectangular grids:

> People will say: “That’s easily said! But all your intersections are right angles. What about the infinite variations that constitute the reality of our cities?” But that’s precisely the point: I eliminate all these things. Otherwise we shall never get anywhere.
>
> I can already hear the storms of protest and the sarcastic gibes: “Imbecile, madman, idiot, braggart, lunatic, etc.” Thank you very much, but it makes no difference: my starting point is still the same: I insist on right-angled intersections. The intersections shown here are all perfect.

Scott uses Le Corbusier as the epitome of five High Modernist principles.

First, there can be no compromise with the existing infrastructure. It was designed by superstitious people who didn’t have architecture degrees, or at the very least got their architecture degrees in the past and so were insufficiently Modern. The more completely it is bulldozed to make way for the Glorious Future, the better.

Second, human needs can be abstracted and calculated. A human needs X amount of food. A human needs X amount of water. A human needs X amount of light, and prefers to travel at X speed, and wants to live within X miles of the workplace. These needs are easily calculable by experiment, and a good city is the one built to satisfy these needs and ignore any competing frivolities.

Third, the solution is the solution. It is universal. The rational design for Moscow is the same as the rational design for Paris is the same as the rational design for Chandigarh, India. As a corollary, all of these cities ought to look exactly the same. It is *maybe* permissible to adjust for obstacles like mountains or lakes. But only if you are on too short a budget to follow the rationally correct solution of leveling the mountain and draining the lake to make your city truly optimal.

Fourth, all of the relevant rules should be explicitly determined by technocrats, then followed to the letter by their subordinates. Following these rules is better than trying to use your intuition, in the same way that using the laws of physics to calculate the heat from burning something is better than just trying to guess, or following an evidence-based clinical algorithm is better than just prescribing whatever you feel like.

Fifth, there is nothing whatsoever to be gained or learned from the people involved (eg the city’s future citizens). *You* are a rational modern scientist with an architecture degree who has already calculated out the precise value for all relevant urban parameters. *They* are yokels who probably cannot even *spell* the word architecture, let alone usefully contribute to it. They probably make all of their decisions based on superstition or tradition or something, and their input should be ignored For Their Own Good.

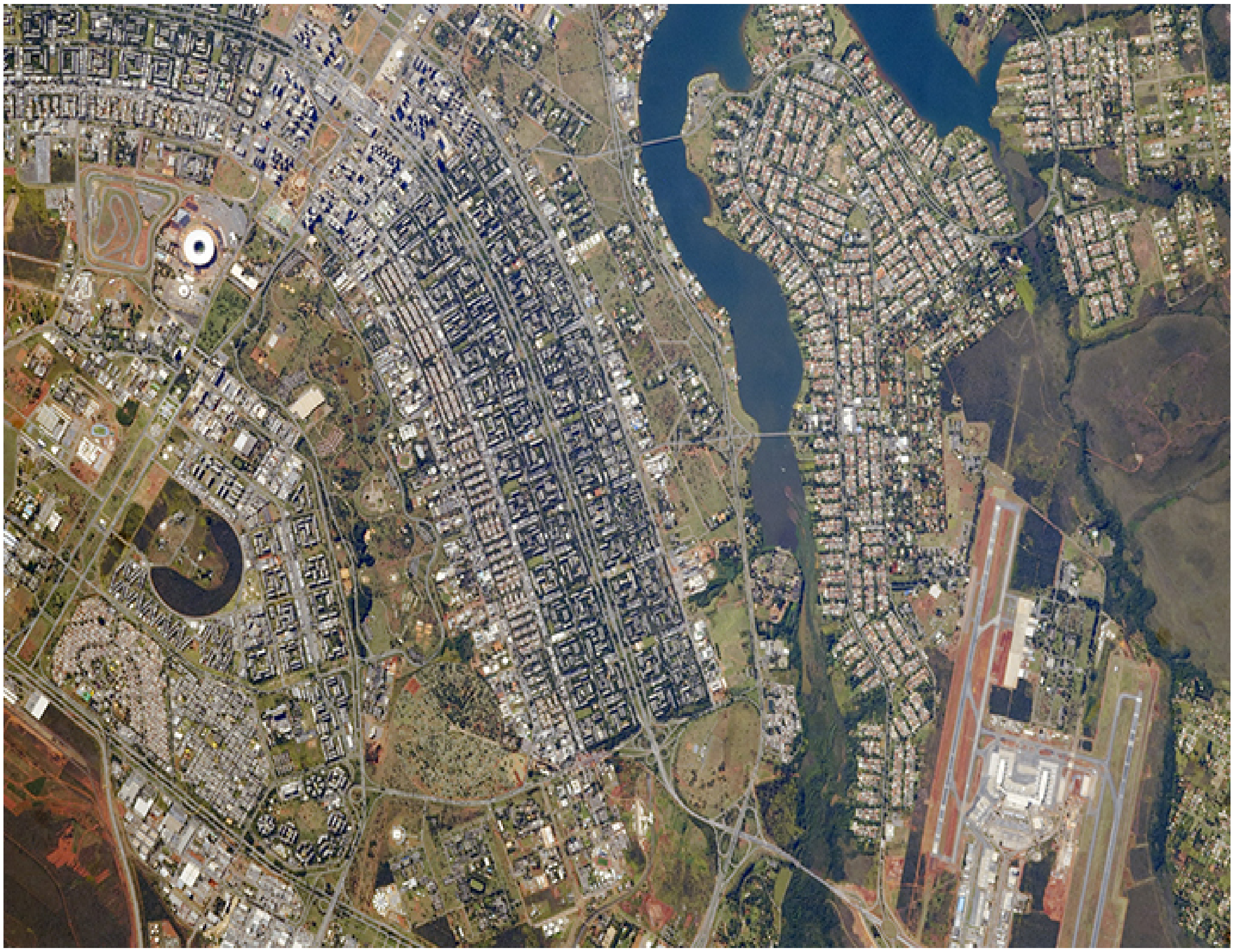
And lest I be unfair to Le Corbusier, a lot of his scientific rational principles made a lot of sense. Have wide roads so that there’s enough room for traffic and all the buildings get a lot of light. Use rectangular grids to make cities easier to navigate. Avoid frivolous decoration so that everything is efficient and affordable to all. Use concrete because it’s the cheapest and strongest material. Keep pedestrians off the streets as much as possible so that they don't get hit by cars. Use big apartment towers to save space, then use the open space for pretty parks and public squares. Avoid anything that looks like a local touch, because nationalism leads to war and we are all part of the same global community of humanity. It sounded pretty good, and for a few decades the entire urban planning community was convinced.

So, how did it go?

Scott uses the example of Brasilia. Brazil wanted to develop its near-empty central regions and decided to build a new capital in the middle of nowhere. They hired three students of Le Corbusier, most notably Oscar Niemeyer, to build them a perfect scientific rational city. The conditions couldn't have been better. The land was already pristine, so there was no need to bulldoze Paris first. There were no inconvenient mountains or forests in the way. The available budget was in the tens of billions. The architects rose to the challenge and built them the world's greatest High Modernist city.

It's…even more beautiful than I imagined

Yet twenty years after its construction, the city's capacity of 500,000 residents was only half-full. And it wasn't the location – a belt of suburbs grew up with a population of almost a million. People wanted to live in *the vicinity of* Brasilia. They just didn't want to live *in the parts that Niemeyer and the Corbusierites had built*.



Brasilia from above. Note both the evenly-spaced rectangular grid of identical buildings in the center, and the fact that most people aren't living in it.

What happened? Scott writes:

> Most of those who have moved to Brasilia from other cities are amazed to discover "that it is a city without crowds." People complain that Brasilia lacks the bustle of street life, that it has none of the busy street corners and long stretches of storefront facades that animate a sidewalk for pedestrians. For them, it is almost as if the

founders of Brasilia, rather than having planned a city, have actually planned to prevent a city. The most common way they put it is to say that Brasilia “lacks street corners,”by which they mean that it lacks the complex intersections of dense neighborhoods comprising residences and public cafes and restaurants with places for leisure, work, and shopping.

While Brasilia provides well for some human needs, the functional separation of work from residence and of both from commerce and entertainment, the great voids between superquadra, and a road system devoted exclusively to motorized traffic make the disappearance of the street corner a foregone conclusion. The plan did eliminate traffic jams; it also eliminated the welcome and familiar pedestrian jams that one of Holston’s informants called “the point of social conviviality

The term brasilite, meaning roughly Brasilia-itis,which was coined by the first-generation residents, nicely captures the trauma they experienced. As a mock clinical condition, it connotes a rejection of the standardization and anonymity of life in Brasilia. “They use the term brasilite to refer to their feelings about a daily life without the pleasures-the distractions, conversations, flirtations, and little rituals of outdoor life in other Brazilian cities.” Meeting someone normally requires seeing them either at their apartment or at work. Even if we allow for the initial simplifying premise of Brasilia’s being an administrative city, there is nonetheless a bland anonymity built into the very structure of the capital. The population simply lacks the small accessible spaces that they could colonize and stamp with the character of their activity, as they have done historically in Rio and Sao Paulo. To be sure, the inhabitants of Brasilia haven’t had much time to modify

> the city through their practices, but the city is designed to be fairly recalcitrant to their efforts.
>
> “Brasilite,” as a term, also underscores how the built environment affects those who dwell in it. Compared to life in Rio and Sao Paulo, with their color and variety, the daily round in bland, repetitive, austere Brasilia must have resembled life in a sensory deprivation tank. The recipe for high-modernist urban planning, while it may have created formal order and functional segregation, did so at the cost of a sensorily impoverished and monotonous environment-an environment that inevitably took its toll on the spirits of its residents.
>
> The anonymity induced by Brasilia is evident from the scale and exterior of the apartments that typically make up each residential superquadra. For superquadra residents, the two most frequent complaints are the sameness of the apartment blocks and the isolation of the residences (“In Brasilia, there is only house and work”). The facade of each block is strictly geometric and egalitarian. Nothing distinguishes the exterior of one apartment from another; there are not even balconies that would allow residents to add distinctive touches and create semipublic spaces.

Brasilia is interesting only insofar as it was an entire High Modernist planned city. In most places, the Modernists rarely got their hands on entire cities at once. They did build a number of suburbs, neighborhoods, and apartment buildings. There was, however, a disconnect. Most people did not want to buy a High Modernist house or live in a High Modernist neighborhood. Most governments *did* want to fund High Modernist houses and neighborhoods, because the academics influencing them said it was the modern scientific rational thing to do. So in the end, one of High Modernists’ main contributions to the United States was the projects – ie government-funded public housing for poor people who didn't get to choose where to live.

I never really "got" Jane Jacobs. I originally interpreted her as arguing that it was great for cities to be noisy and busy and full of crowds, and that we should build neighborhoods that are confusing and hard to get through to force people to interact with each other and prevent them from being able to have privacy, and no one should be allowed to live anywhere quiet or nice. As somebody who (thanks to the public school system, etc) has had my share of being forced to interact with people, and of being placed in situations where it is deliberately difficult to have any privacy or time to myself, I figured Jane Jacobs was just a jerk.

But Scott has kind of made me come around. He rehabilitates her as someone who was responding to the very real excesses of High Modernism. She was the first person who really said "Hey, maybe people *like* being in cute little neighborhoods". Her complaint wasn't really against privacy or order per se as it was against extreme High Modernist perversions of those concepts that people empirically hated. And her background makes this all too understandable – she started out as a journalist covering poor African-Americans who lived in the projects and had some of the same complaints as Brazilians.

Her critique of Le Corbusierism was mostly what you would expect, but Scott extracts some points useful for their contrast with the Modernist points earlier:

First, existing structures are evolved organisms built by people trying to satisfy their social goals. They contain far more wisdom about people's needs and desires than anybody could formally enumerate. Any attempt at urban planning should try to build on this encoded knowledge, not detract from it.

Second, man does not live by bread alone. People don’t want the right amount of Standardized Food Product, they want social interaction, culture, art, coziness, and a host of other things nobody will ever be able to calculate. Existing structures have already been optimized for these things, and unless you’re really sure you understand all of them, you should be reluctant to disturb them.

Third, solutions are local. Americans want different things than Africans or Indians. One proof of this is that New York looks different from Lagos and from Delhi. Even if you are the world’s best American city planner, you should be very concerned that you have no idea what people in Africa need, and you should be very reluctant to design an African city without extensive consultation of people who understand the local environment.

Fourth, even a very smart and well-intentioned person who is on board with points 1-3 will never be able to produce a set of rules. Most of people’s knowledge is implicit, and most rule codes are quickly replaced by informal systems of things that work which are much more effective (the classic example of this is work-to-rule strikes).

Fifth, although well-educated technocrats may understand principles which give them some advantages in their domain, they are hopeless without the on-the-ground experience of the people they are trying to serve, whose years of living in their environment and dealing with it every day have given them a deep practical knowledge which is difficult to codify.

How did Jacobs herself choose where to live? As per her Wikipedia page:

> [Jacobs] took an immediate liking to Manhattan’s Greenwich Village, which did not conform to the city’s grid structure.

V.

The same thing that happened with cities happened with farms. The American version was merely farce:

> We should recognize that the rationalization of farming on a huge, even national, scale was part of a faith shared by social engineers and agricultural planners throughout the world. And they
> were conscious of being engaged in a common endeavor…They kept in touch through journals, professional conferences, and exhibitions. The connections were strongest between American agronomists and their Russian colleagues – connections that were not entirely broken even during the Cold War. Working in vastly different economic and political environments, the Russians tended to be envious of the level of capitalization, particularly in mechanization, of American farms while the Americans were envious of the political scope of Soviet planning. The degree to which they were working together to create a new world of large-scale, rational, industrial agriculture can be judged by this brief account of their relationship […]
>
> Many efforts were made to put this faith to the test. Perhaps the most audacious was the Thomas Campbell “farm” in Montana, begun – or, perhaps I should say, founded – in 1918 It was an industrial farm in more than one respect. Shares were sold by prospectuses describing the enterprise as an “industrial opportunity”; J. P. Morgan, the financier, helped to raise $2 million from the public. The Montana Farming Corporation was a monster wheat farm of ninety-five thousand acres, much

of it leased from four Native American tribes. Despite the private investment, the enterprise would never have gotten off the ground without help and subsidies from the Department of Interior and the United States Department of Agriculture (USDA).

Proclaiming that farming was about 90 percent engineering and only 10 percent agriculture, Campbell set about standardizing as much of his operation as possible. He grew wheat and flax, two hardy crops that needed little if any attention between planting and harvest time.The land he farmed was the agricultural equivalent of the bulldozed site of Brasilia. It was virgin soil, with a natural fertility that would eliminate the need for fertilizer. The topography also vastly simplified matters: it was flat, with no forests, creeks, rocks, or ridges that would impede the smooth course of machinery over its surface. In other words, the selection of the simplest, most standardized crops and the leasing of something very close to a blank agricultural space were calculated to favor the application of industrial methods […]

This is not the place to chronicle the fortunes of the Montana Farming Corporation, and in any event Deborah Fitzgerald has done so splendidly. Suffice it to note that a drought in the second year and the elimination of a government support for prices the following year led to a collapse that cost J. P. Morgan $1 million. The Campbell farm faced other problems besides weather and prices: soil differences, labor turnover, the difficulty of finding skilled, resourceful workers who would need little supervision. Although the corporation struggled on until Campbell’s death in 1966,it provided no evidence that industrial farms were superior to family farms in efficiency and profitability.

But the Soviet version was tragedy. Instead of raising some money to start a giant farm and seeing it didn’t work, the USSR uprooted millions of peasants, forced them onto collective farms, and then watched as millions of people starved to death due to crop failure. What happened?

Scott really focuses on that claim (above) that farming was “90% engineering and only 10% agriculture”. He says that these huge farms all failed – despite being better-funded, higher-tech, and having access to the wisdom of the top agricultural scientists – exactly because this claim was false. Small farmers may not know much about agricultural science, but they know a lot about farming. Their knowledge is intuitive and local – for example, what to do in a particular climate or soil. It is sometimes passed down over generations, and other times determined through long hours of trial-and-error.

He gave the example of Tanzania, where small farmers grew dozens of different crops together in seeming chaos. Western colonists tried to convince them – often by force – to switch to just growing one thing at a time to reap advantages of efficiency, standardization, and specialization of labor. Only growing one crop in the same field was Agricultural Science 101. But this turned out to be a bad idea in the difficult Tanzanian environment:

> The multistoried effect of polyculture has some distinct advantages for yields and soil conservation. “Upper-story” crops shade “lowerstory” crops, which are selected for their ability to thrive in the cooler soil temperature and increased humidity at ground level. Rainfall reaches the ground not directly but as a fine spray that is absorbed with less damage to soil structure and less erosion. The taller crops often serve as a useful windbreak for the lower crops. Finally, in mixed or relay

cropping, a crop is in the field at all times, holding the soil together and reducing the leaching effects that sun, wind, and rain exert, particularly on fragile land. Even if polyculture is not to be preferred on the grounds of immediate yield, there is much to recommend it in terms of sustainability and thus long-term production.

Our discussion of mixed cropping has thus far dealt only with the narrow issues of yield and soil conservation. It has overlooked the cultivators themselves and the various other ends that they seek by using such techniques. The most significant advantage of intercropping, Paul Richards claims, is its great flexibility, “the scope [it] offers for a range of combinations to match individual needs and preferences, local conditions, and changing circumstances within each season and from season to season.” Farmers may polycrop in order to avoid labor bottlenecks at planting and at harvest.[44] Growing many different crops is also an obvious way to spread risks and improve food security. Cultivators can reduce the danger of going hungry if they sow, instead of only one or two cultivars, crops of long and short maturity, crops that are drought resistant and those that do well under wetter conditions, crops with different patterns of resistance to pests and diseases, crops that can be stored in the ground with little loss (such as cassava), and crops that mature in the “hungry time” before other crops are gathered. Finally, and perhaps most important, each of these crops is embedded in a distinctive set of social relations. Different members of the household are likely to have different rights and responsibilities with respect to each crop. The planting regimen, in other words, is a reflection of social relations, ritual needs, and culinary tastes; it is not just a production strategy that a

> profit-maximizing entrepreneur took straight out of the pages of a text in neoclassical economics.

Nor could this be solved just by adding a pinch of empiricism. A lot of European farming specialists were into empiricism, sort of. What they ended up doing was creating crops that worked really well in a lab but not in actual Tanzania. If they were lucky, they created crops that worked really well on the one experimental farm in Tanzania they fenced off as a testing ground, but not on any other Tanzanian farms. If they were *really* lucky, they created crops that would grow on Tanzanian farms and be good on whatever single axis they were optimizing (like selling for lots of money) but not in other ways that were equally important to the populace (like being low-risk, or useful for non-food purposes, or resistant to disease, or whatever). And if they were *supremely* lucky, then they would go to the Tanzanians and say “Hey, we invented a new farming method that solves all your problems!” and the Tanzanians would say “Yeah, we heard rumors about that, so we tried it ourselves, and now we’ve been using it for five years and advanced it way beyond what you were doing.”

There *were* some scientists who got beyond these failure modes, and Scott celebrates them (while all too often describing how they were marginalized and ignored by the rest of the scientific community). But at the point where you’ve transcended all this, you’re no longer a domain-general agricultural scientist, you’re a Tanzanian farming specialist who’s only one white coat removed from being a Tanzanian farmer yourself.

Even in less exotic locales like Russia, the peasant farmers were extraordinary experts on the conditions of their own farms, their own climates, and their own crops. Take all of these people, transport them a thousand miles away, and give them a perfectly rectangular grid to grow Wheat Cultivar #6 on, and you have a recipe for disaster.

VI.

So if this was such a bad idea, why did everyone keep doing it?

Start with the cities. Scott notes that although citizens generally didn’t have a problem with earlier cities, governments did:

> Historically, the relative illegibility to outsiders of some urban neighborhoods has provided a vital margin of political safety from control by outside elites. A simple way of determining whether this margin exists is to ask if an outsider would have needed a local guide in order to find her way successfully. If the answer is yes, then the community or terrain in question enjoys at least a small measure of insulation from outside intrusion. Coupled with patterns of local solidarity, this insulation has proven politically valuable in such disparate contexts as eighteenth-and early nineteenth-century urban riots over bread prices in Europe, the Front de Liberation Nationale’s tenacious resistance to the French in the Casbah of Algiers, and the politics of the bazaar that helped to bring down the Shah of Iran. Illegibility, then, has been and remains a reliable resource for political autonomy

This was a particular problem in Paris, which was famous for a series of urban insurrections in the 19th century (think *Les Miserables*, but about once every ten years or so). Although these generally failed, they were hard to suppress because locals knew the “terrain” and the streets were narrow enough to barricade. Slums full of poor people gathered together formed tight communities where revolutionary ideas could easily spread. The late 19th-century redesign of Paris had the explicit design of destroying these areas and splitting up poor people somewhere far away from the city center where they couldn’t do any harm.

The Soviet collective farms had the same dubious advantage. The problem they “effectively” “solved” was the non-collectivized farmers becoming too powerful and independent a political bloc. They lived in tight-knit little villages that did their own thing, the Party officials who went to these villages to keep order often ended up “going native”, and the Soviets had no way of knowing how much food the farmers were producing and whether they were giving enough of it to the Motherland:

> Confronting a tumultuous, footloose, and “headless” rural society which was hard to control and which had few political assets, the Bolsheviks, like the scientific foresters, set about redesigning their environment with a few simple goals in mind. They created, in place of what they had inherited, a new landscape of large, hierarchical, state-managed farms whose cropping patterns and procurement quotas were centrally mandated and whose population was, by law, immobile. The system thus devised served for nearly sixty years as a mechanism for procurement and control at a massive cost in stagnation, waste, demoralization, and ecological failure.

The collectivized farms couldn’t grow much, but people were thrown together in artificial towns designed to make it impossible to build any kind of community: there was nowhere to be except in bed asleep, working in the fields, or at the public school receiving your daily dose of state propaganda. The towns were identical concrete buildings on a grid, which left the locals maximally disoriented (because there are no learnable visual cues) and the officials maximally oriented (because even a foreigner could go to the intersection of Street D and Street 7). All fields were perfectly rectangular and produced Standardized Food Product, so it was (theoretically) easy to calculate how much they should be producing and whether people were meeting that target. And everyone was in the same place, so if there were some sort of problem it was much easier to bring in the army or secret police than if they were split up among a million tiny villages in the middle of nowhere.

So although modernist cities and farms may have started out as attempts to help citizens with living and farming, they ended up as contributors to the great government project of legibility and taxing people effectively.

Seeing Like A State summarizes the sort of on-the-ground ultra-empirical knowledge that citizens have of city design and peasants of farming as *metis*, a Greek term meaning “practical wisdom”. I was a little concerned about this because they seem like two different things. The average citizen knows nothing about city design and in fact does not design cities; cities sort of happen in a weird way through cultural evolution or whatever. The average farmer knows a lot about farming (even if it is implicit and not as book learning) and applies that knowledge directly in how they farm. But Scott thinks these are more or less the same thing, that this thing is a foundation of successful communities and industries, and that ignoring and suppressing it is what makes collective farms and modernist planned cities so crappy. He generalizes this further to almost every aspect of a society – its language, laws, social norms, and economy. But this is all done very quickly, and I feel like there was a sleight of hand between “each farmer eventually figures out how to farm well” and “social norms converge on good values”.

Insofar as Scott squares the above circle, he seems to think that many actors competing with each other will eventually carve out a beneficial equilibrium better than that of any centralized authority. This doesn't really mesh will with my own fear that many actors competing with each other will eventually shoot themselves in the foot and destroy everything, and I haven't really seen a careful investigation of when we get one versus the other.

VII.

What are we to make of all of this?

Well, for one thing, Scott basically admits to stacking the dice against High Modernism and legibility. He admits that the organic livable cities of old had life expectancies in the forties because nobody got any light or fresh air and they were all packed together with no sewers and so everyone just died of cholera. He admits that at some point agricultural productivity multiplied by like a thousand times and the Green Revolution saved millions of lives and all that, and probably that has something to do with scientific farming methods and rectangular grids. He admits that it's pretty convenient having a unit of measurement that local lords can't change whenever they feel like it. Even modern timber farms seem pretty successful. After all those admissions, it's kind of hard to see what's left of his case.

(also, I grew up in Irvine, the most planned of planned cities, and I loved it.)

What Scott eventually says is that he's not against legibility and modernism per se, but he wants to present them as ingredients in a cocktail of state failure. You need a combination of four things to get a disaster like Soviet collective farming (or his other favorite example, compulsory village settlement in Tanzania). First, a government incentivized to seek greater legibility for its population and territory. Second, a High Modernist ideology. Third, authoritarianism. And fourth, a “prostrate civil society”, like in Russia after the Revolution, or in colonies after the Europeans took over.

I think his theory is that the back-and-forth between centralized government and civil society allows scientific advances to be implemented smoothly instead of just plowing over everyone in a way that leads to disaster. I also think that maybe a big part of it is incremental versus sudden: western farming did well because it got to incrementally add advances and see how they worked, but when you threw the entire edifice at Tanzania it crashed and burned.

I’m still not really sure what’s left. Authoritarianism is bad? Destroying civil society is bad? You shouldn’t do things when you have no idea what you’re doing and all you’ve got to go on is your rectangle fetish? The book contained some great historical tidbits, but I’m not sure what overarching lesson I learned from it.

It’s not that I don’t think Scott’s preference for *metis* over scientific omnipotence has value. I think it has lots of value. I see this all the time in psychiatry, which always has been and to some degree still is really High Modernist. We are educated people who know a lot about mental health, dealing with a poor population who (in the case of one of my patients) refers to Haldol as “Hound Dog”. It’s *very* easy to get in the trap of thinking that you know better than these people, especially since you often do (I will never understand how many people are *shocked* when I diagnose their sleep disorder as having something to do with them drinking fifteen cups of coffee a day).

But psychiatric patients have a *metis* of dealing with their individual diseases the same way peasants have a *metis* of dealing with their individual plots of land. My favorite example of this is doctors who learn their patients are taking marijuana, refuse to keep prescribing them their vitally important drugs unless the patient promises to stop, and then gets surprised when the patients end up decompensating because the marijuana was keeping them together. I'm not saying smoking marijuana is a good thing. I'm saying that for some people it's a load-bearing piece of their mental edifice. And if you take it away without any replacement they will fall apart. And *they have explained this to you a thousand times and you didn't believe them*.

There are *so many* fricking patients who respond to sedative medications by becoming stimulated, or stimulant medications by becoming sedated, or who become more anxious whenever they do anti-anxiety exercises, or who hallucinate when placed on some super common medication that has never caused hallucinations in anyone else, or who become suicidal if you try to reassure them that things aren't so bad, or any other completely perverse and ridiculous violation of the natural order that you can think of. And the only redeeming feature of all of this is that the patients themselves know all of this stuff super-well and are usually happy to tell you if you ask.

I can totally imagine going into a psychiatric clinic armed with the Evidence-Based Guidelines the same way Le Corbusier went into Moscow and Paris armed with his Single Rational City Plan and the same way the agricultural scientists went into Tanzania armed with their List Of Things That Definitely Work In Europe. I expect it would have about the same effect for about the same reason.

(including the part where I would get promoted. I'm not too sure what's going on there, actually.)

So fine, Scott is completely right here. But I'm only bringing this up because it's something I've already thought about. If I didn't already believe this, I'd be indifferent between applying the narrative of the wise Tanzanian farmers knowing more than their English colonizers, versus the narrative of the dumb yokels who refuse to get vaccines because they might cause autism. Heuristics work until they don't. Scott provides us with these great historical examples of local knowledge outdoing scientific acumen, but other stories present us with great historical examples of the opposite, and when to apply which heuristic seems really unclear. Even "don't bulldoze civil society and try to change everything at once" goes astray sometimes; the Meiji Restoration was wildly successful by doing exactly that.

Maybe I'm trying to take this too far by talking about psychiatry and Meiji Restorations. Most of Scott's good examples involved either agriculture or resettling peasant villages. This is understandable; Scott is a scholar of colonialism in Southeast Asia and there was a lot of agriculture and peasant resettling going on there. But it's a pretty limited domain. The book amply proves that peasants know an astounding amount about how to deal with local microclimates and grow local varieties of crops and so on, and frankly I am shocked that anyone with an IQ of less than 180 has ever managed to be a peasant farmer, but how does that apply to the sorts of non-agricultural issues we think about more often?

The closest analogy I can think of right now – maybe because it's on my mind – is this story about check-cashing shops. Professors of social science think these shops are evil because they charge the poor higher rates, so they should be regulated away so that poor people don't foolishly shoot themselves in the foot by going to them. But on closer inspection, they offer a better deal for the poor than banks do, for complicated reasons that aren't visible just by comparing the raw numbers. Poor people's understanding of this seems a lot like the *metis* that helps them understand local agriculture. And progressives' desire to shift control to the big banks seems a lot like the High Modernists' desire to shift everything to a few big farms. Maybe this is a point in favor of something like libertarianism? Maybe especially a "libertarianism of the poor" focusing on things like occupational licensing, not shutting down various services to the poor because they don't meet rich-people standards, not shutting down various services to the poor because we think they're "price-gouging", et cetera?

Maybe instead of concluding that Scott is too focused on peasant villages, we should conclude that he's focused on confrontations between a well-educated authoritarian overclass and a totally separate poor underclass. Most modern political issues don't exactly map on to that – even things like taxes where the rich and the poor are on separate sides don't have a bimodal distribution. But in cases there are literally about rich people trying to dictate to the poorest of the poor how they should live their lives, maybe this becomes more useful.

Actually, one of the best things the book did to me was make me take cliches about "rich people need to defer to the poor on poverty-related policy ideas" more seriously. This has become so overused that I roll my eyes at it: "Could quantitative easing help end wage stagnation? Instead of asking macroeconomists, let's ask this 19-year old single mother in the Bronx!" But Scott provides a lot of situations where that was exactly the sort of person they should have asked. He also points out that Tanzanian natives using their traditional farming practices were more productive than European colonists using scientific farming. I’ve had to listen to so many people talk about how “we must respect native people’s different ways of knowing” and “native agriculturalists have a profound respect for the earth that goes beyond logocentric Western ideals” and nobody had ever bothered to tell me before that they *actually produced more crops per acre*, at least some of the time. That would have put all of the other stuff in a pretty different light.

I understand Scott is an anarchist. He didn’t really try to defend anarchism in this book. But I was struck by his description of peasant villages as this totally separate unit of government which was happily doing its own thing very effectively for millennia, with the central government’s relevance being entirely negative – mostly demanding taxes or starting wars. They kind of reminded me of some pictures of hunter-gatherer tribes, in terms of being self-sufficient, informal, and just never encountering the sorts of economic and political problems that we take for granted. They make communism (the type with actual communes, not the type where you have Five Year Plans and Politburos and gulags) look more attractive. I think Scott was trying to imply that this is the sort of thing we could have if not for governments demanding legibility and a world of universal formal rule codes accessible from the center? Since he never actually made the argument, it’s hard for me to critique it. And I wish there had been more about cultural evolution as separate from the more individual idea of *metis*.

A final note: Scott often used the word “rationalism” to refer to the excesses of High Modernism, and I’ve deliberately kept it. What relevance does this have for the LW-Yudkowsky-Bayesian rationalist project? I think the similarities are more than semantic; there certainly is a hope that learning domain-general skills will allow people to leverage raw intelligence and The Power Of Science to various different object-level domains. I continue to be doubtful that this will work in the sort of practical domains where people have spent centuries gathering *metis* in the way Scott describes; this is why I'm wary of any attempt of the rationality movement to branch into self-help. I'm more optimistic about rationalists' ability to open underexplored areas like existential risk – it's not like there's a population of Tanzanian peasants who have spent the last few centuries developing traditional x-risk research whom we are arrogantly trying to replace – and to focus on things that don't bring any immediate practical gain but which help build the foundations for new philosophies, better communities, and more positive futures. I also think that a good art of rationality would look a lot like *metis*, combining easily teachable mathematical rules with more implicit virtues which get absorbed by osmosis.

Overall I did like this book. I'm not really sure what I got from its thesis, but maybe that was appropriate. *Seeing Like A State* was arranged kind of like the premodern forests and villages it describes; not especially well-organized, not really directed toward any clear predetermined goal, but full of interesting things and lovely to spend some time in.

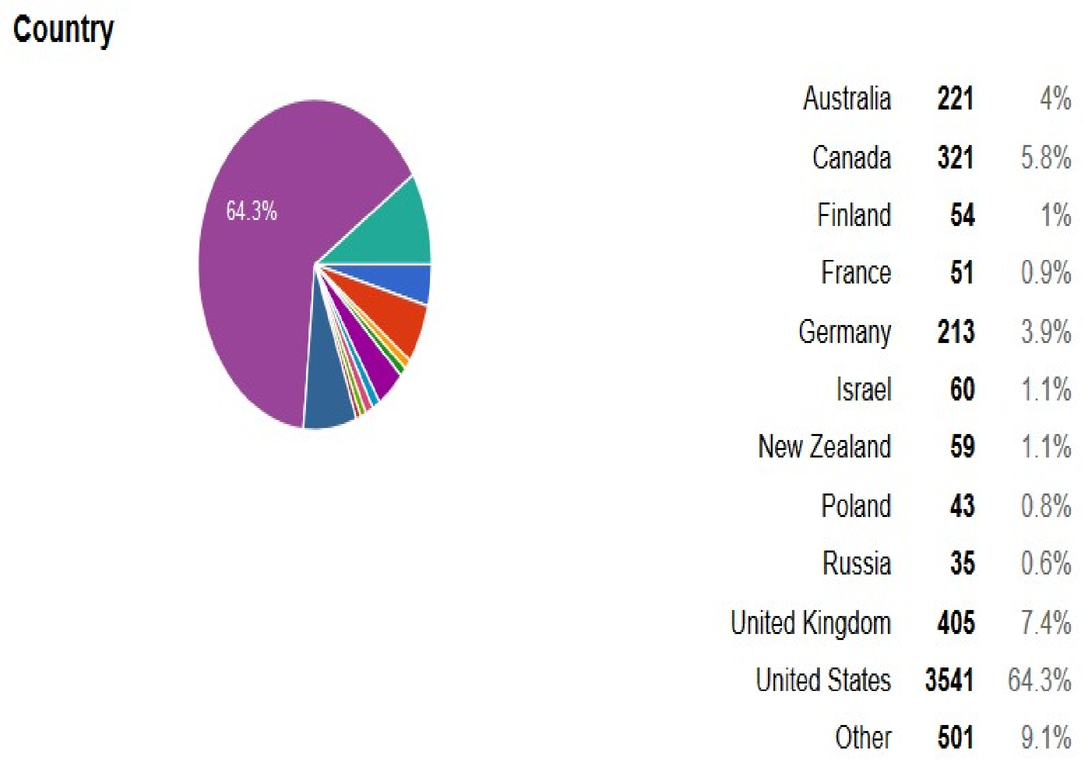
SSC Survey 2017 Results

[None of these calculations were really double-checked and some of them might be wrong. If you're really interested in accuracy, download the raw data at the bottom and see for yourself.]

I.

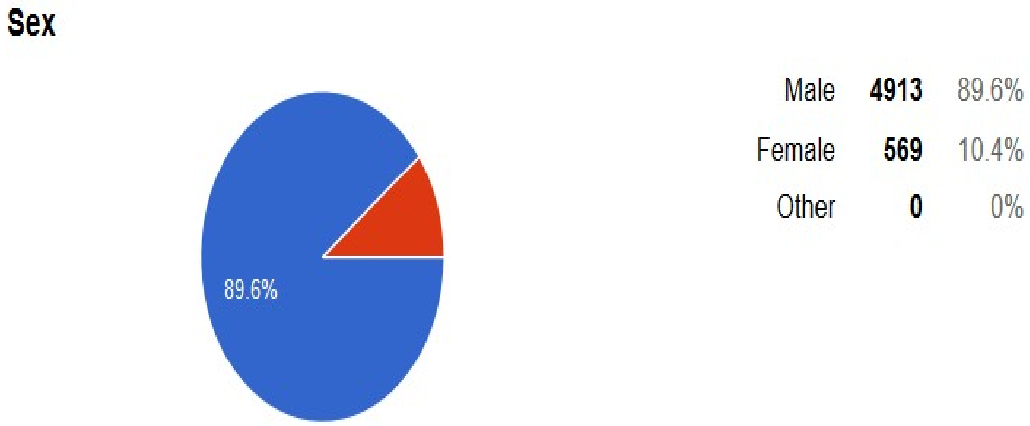
Back in January I asked you to take the SSC survey. Thanks to the 5,500 (!) people who sent in responses. Below are some summaries of answers, alongside paraphrases of the relevant questions to jog your memory. If you want to see the actual questions (some of which are long) you can read them on the survey here. Please don't try to take the survey; your answers will be ignored.

"What country are you from?"



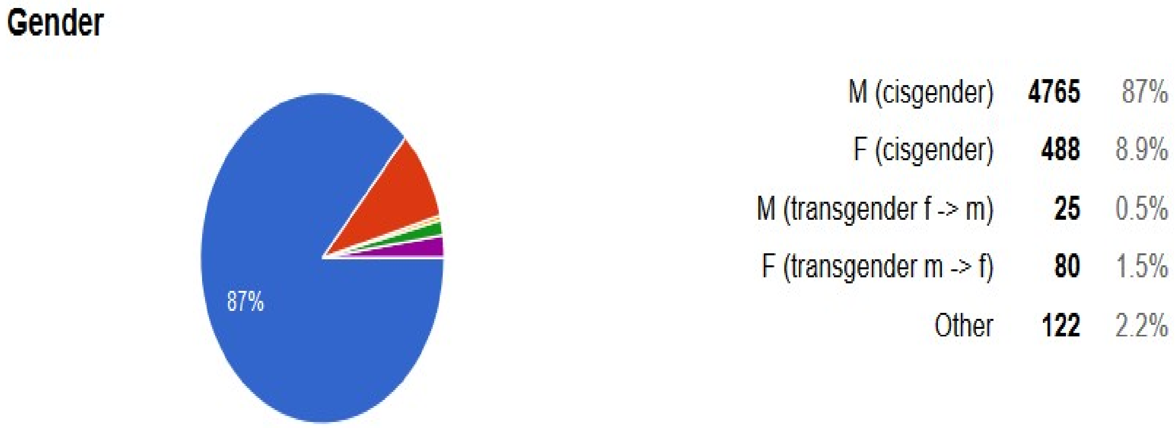
Australia	**221**	4%
Canada	**321**	5.8%
Finland	**54**	1%
France	**51**	0.9%
Germany	**213**	3.9%
Israel	**60**	1.1%
New Zealand	**59**	1.1%
Poland	**43**	0.8%
Russia	**35**	0.6%
United Kingdom	**405**	7.4%
United States	**3541**	64.3%
Other	**501**	9.1%

“What is your biological sex?”
“What is your gender?”
“What is your sexual orientation?”



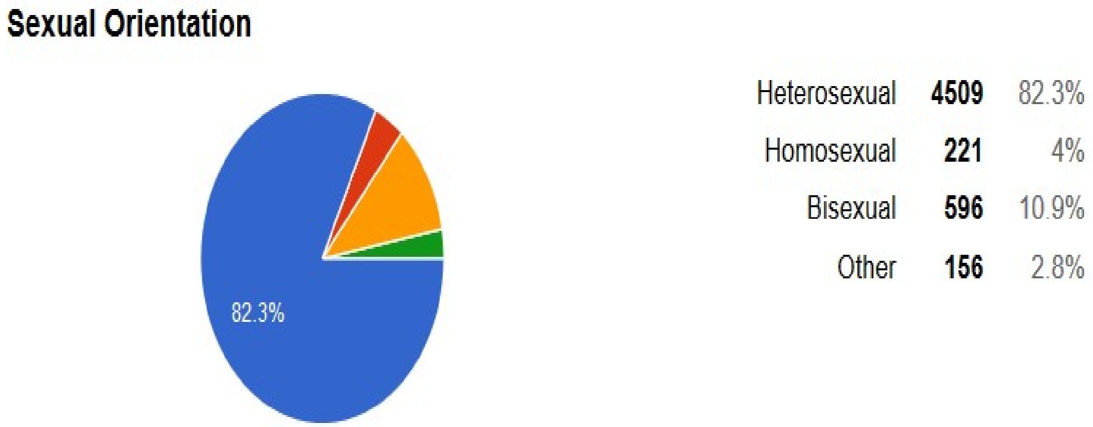
Sex

Male	**4913**	89.6%
Female	**569**	10.4%
Other	**0**	0%



Gender

M (cisgender)	**4765**	87%
F (cisgender)	**488**	8.9%
M (transgender f -> m)	**25**	0.5%
F (transgender m -> f)	**80**	1.5%
Other	**122**	2.2%

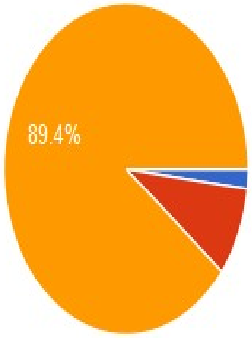


Sexual Orientation

Heterosexual	**4509**	82.3%
Homosexual	**221**	4%
Bisexual	**596**	10.9%
Other	**156**	2.8%

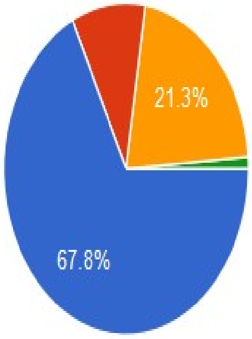
“Do you identify as asexual?”
“Do you prefer monogamous or polyamorous relationships?”
“Are you currently in a relationship?”

Asexuality



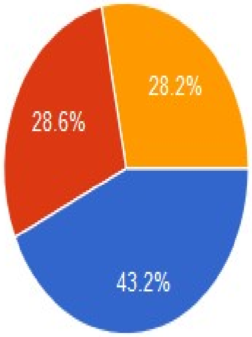
Yes	**105**	1.9%
Sorta	**476**	8.7%
No	**4882**	89.4%

Relationship style



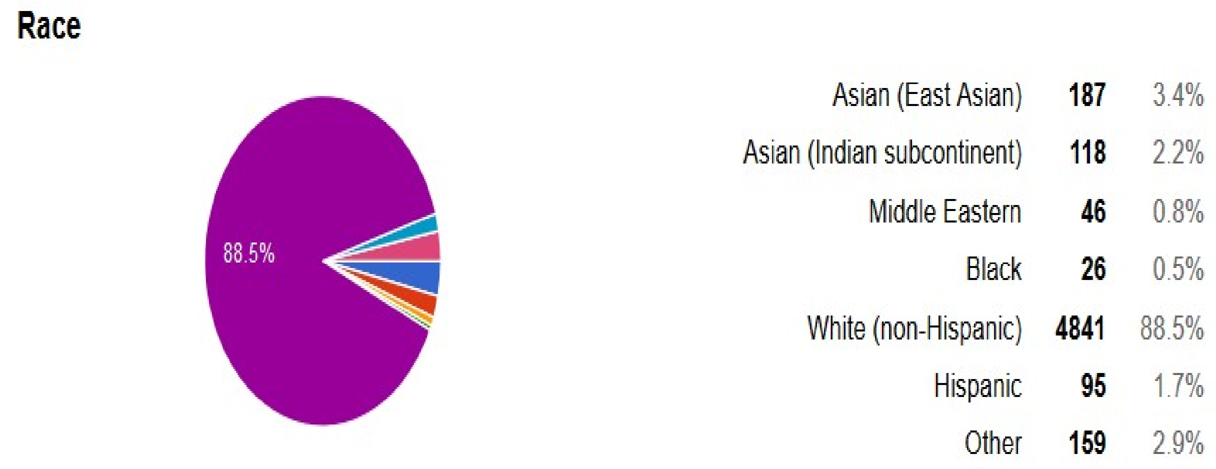
Prefer monogamous	**3703**	67.8%
Prefer polyamorous	**531**	9.7%
Uncertain / no preference	**1165**	21.3%
Other	**65**	1.2%

Relationship Status



Single	**2366**	43.2%
Relationship	**1564**	28.6%
Married	**1547**	28.2%

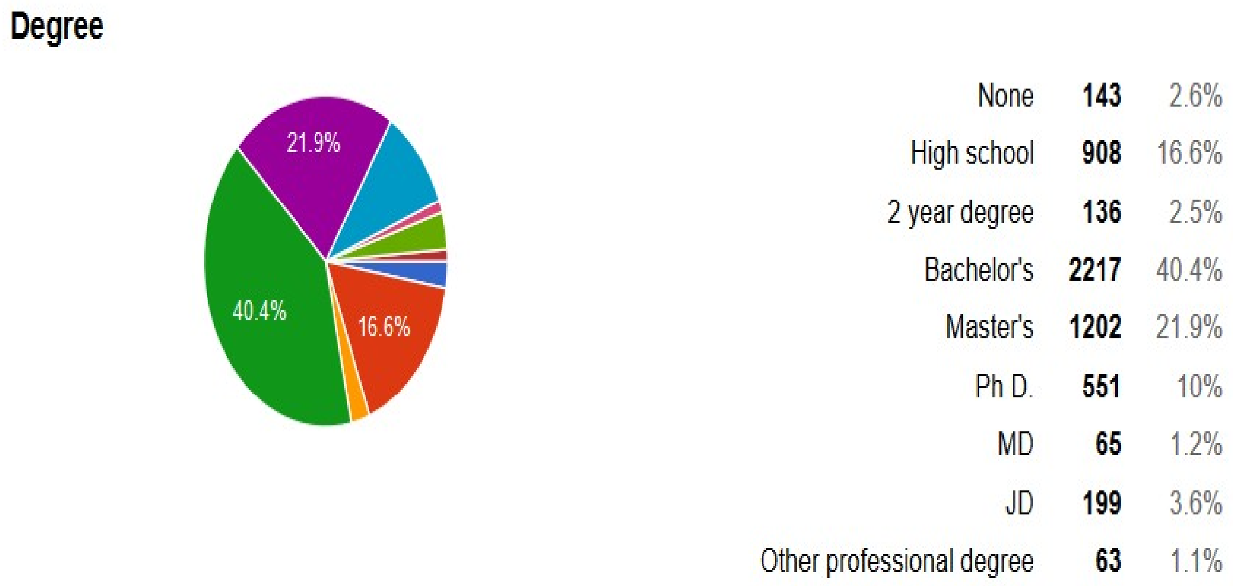
“As what race do you most identify?”



Race

Asian (East Asian)	**187**	3.4%
Asian (Indian subcontinent)	**118**	2.2%
Middle Eastern	**46**	0.8%
Black	**26**	0.5%
White (non-Hispanic)	**4841**	88.5%
Hispanic	**95**	1.7%
Other	**159**	2.9%

“What is the highest level of education you completed?”

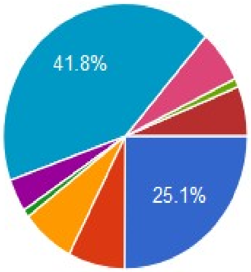


Degree

None	**143**	2.6%
High school	**908**	16.6%
2 year degree	**136**	2.5%
Bachelor's	**2217**	40.4%
Master's	**1202**	21.9%
Ph D.	**551**	10%
MD	**65**	1.2%
JD	**199**	3.6%
Other professional degree	**63**	1.1%

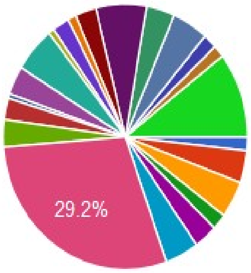
“What is your current employment status?”
“In what field do you work or study?”

Work Status



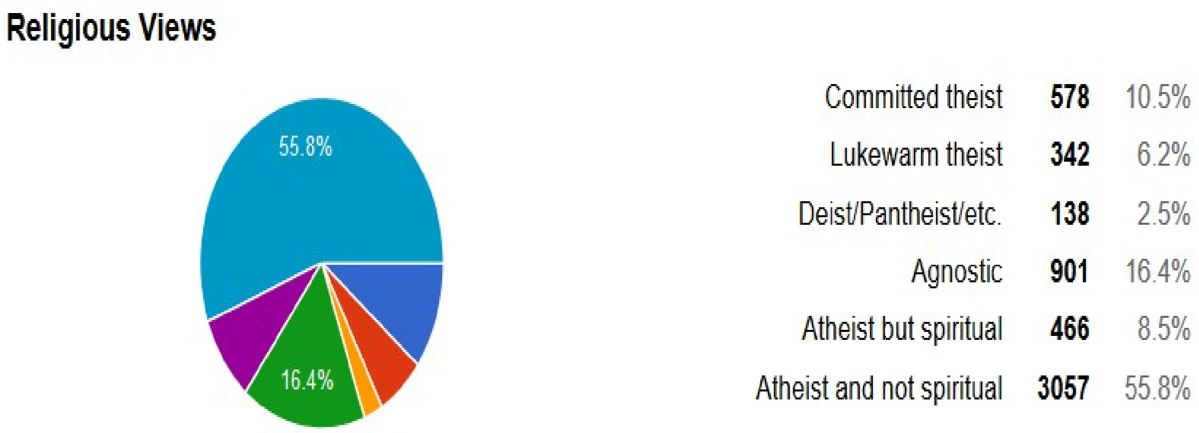
Student	**1375**	25.1%
Academics (on the teaching side)	**413**	7.5%
Self-employed	**386**	7%
Independently wealthy	**53**	1%
Non-profit work	**220**	4%
For-profit work	**2295**	41.8%
Government work	**350**	6.4%
Homemaker	**58**	1.1%
Unemployed	**335**	6.1%

Profession



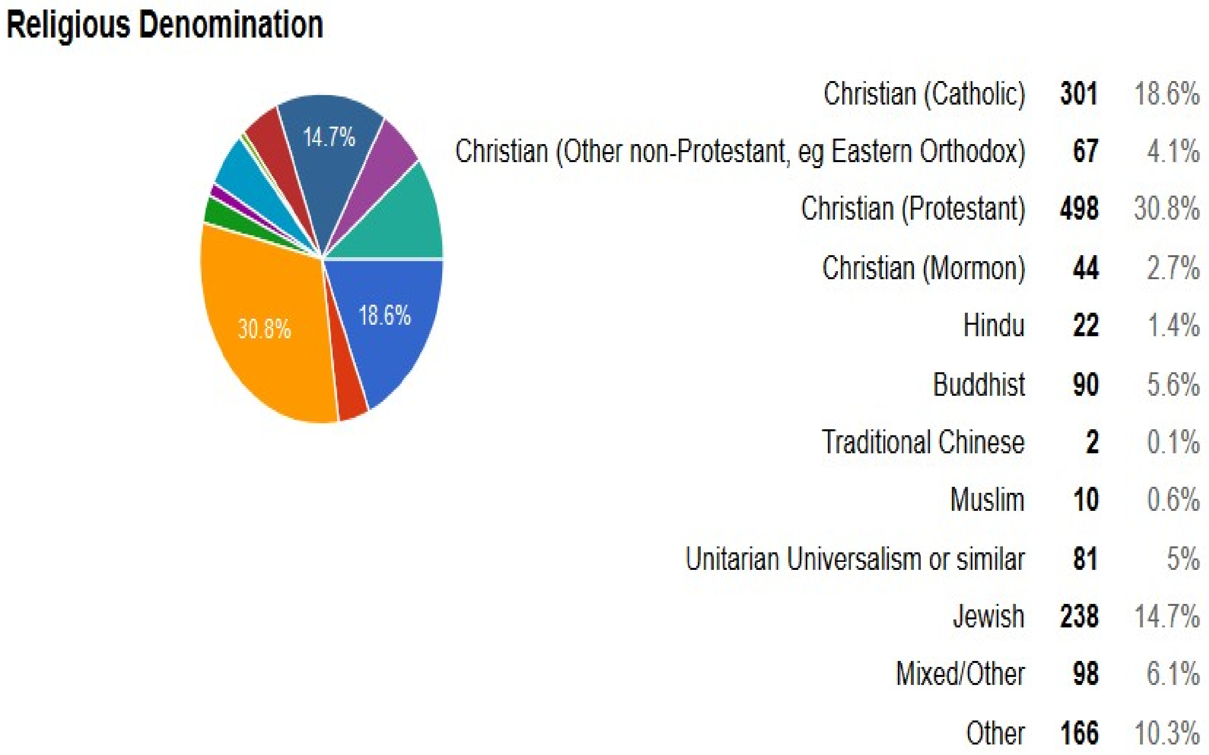
Philosophy	**90**	1.7%
Economics	**212**	3.9%
Finance	**242**	4.5%
Statistics	**105**	1.9%
Computers (AI)	**166**	3.1%
Computers (other academic, computer science)	**239**	4.4%
Computers (practical: IT, programming, etc.)	**1575**	29.2%
Physics	**179**	3.3%
Biology	**141**	2.6%
Health Care (mental health)	**30**	0.6%
Health Care (other)	**187**	3.5%
Mathematics	**298**	5.5%
Neuroscience	**40**	0.7%
Other "hard science"	**120**	2.2%
Psychology	**60**	1.1%
Other "social science"	**150**	2.8%
Engineering	**360**	6.7%
Business	**201**	3.7%
Law	**256**	4.7%
Art	**97**	1.8%
Media	**81**	1.5%
Other	**562**	10.4%

“How would you describe your religious views?”
“If you believe in a religion, which religion is it?”



Religious Views

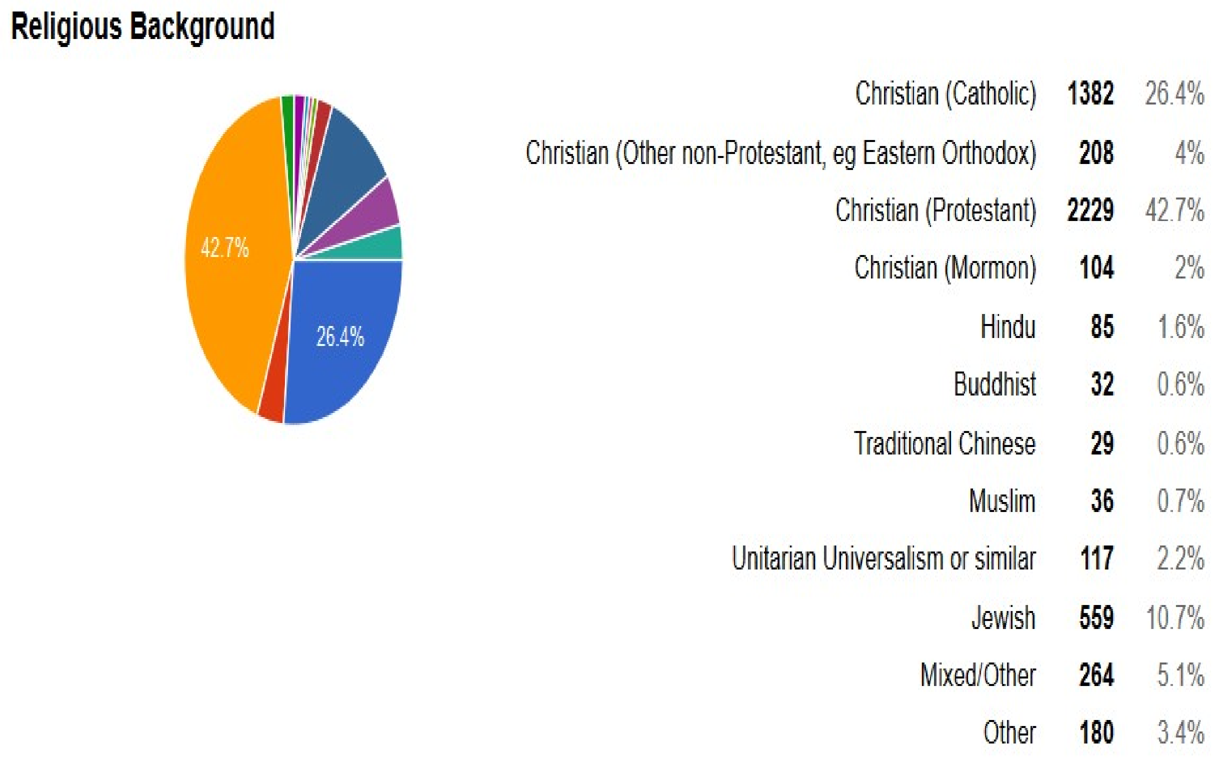
Committed theist	578	10.5%
Lukewarm theist	342	6.2%
Deist/Pantheist/etc.	138	2.5%
Agnostic	901	16.4%
Atheist but spiritual	466	8.5%
Atheist and not spiritual	3057	55.8%



Religious Denomination

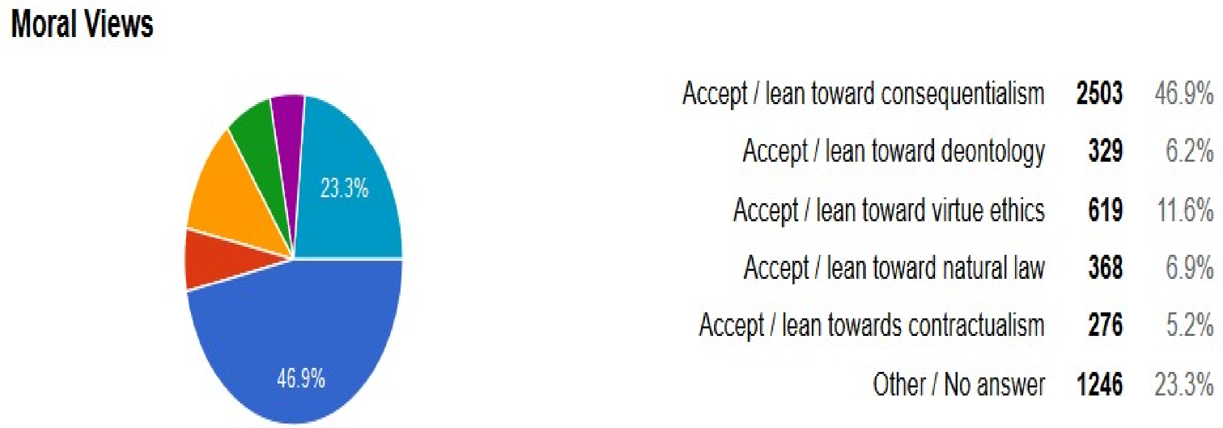
Christian (Catholic)	301	18.6%
Christian (Other non-Protestant, eg Eastern Orthodox)	67	4.1%
Christian (Protestant)	498	30.8%
Christian (Mormon)	44	2.7%
Hindu	22	1.4%
Buddhist	90	5.6%
Traditional Chinese	2	0.1%
Muslim	10	0.6%
Unitarian Universalism or similar	81	5%
Jewish	238	14.7%
Mixed/Other	98	6.1%
Other	166	10.3%

“Whether or not you believe it, what is the religious background of your family?”
“What are your ethical views?”



Religious Background

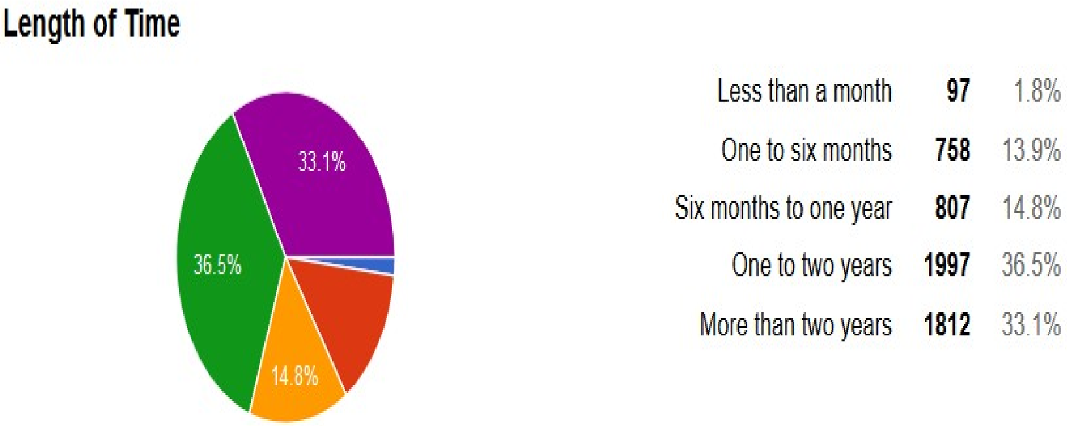
Christian (Catholic)	**1382**	26.4%
Christian (Other non-Protestant, eg Eastern Orthodox)	**208**	4%
Christian (Protestant)	**2229**	42.7%
Christian (Mormon)	**104**	2%
Hindu	**85**	1.6%
Buddhist	**32**	0.6%
Traditional Chinese	**29**	0.6%
Muslim	**36**	0.7%
Unitarian Universalism or similar	**117**	2.2%
Jewish	**559**	10.7%
Mixed/Other	**264**	5.1%
Other	**180**	3.4%



Moral Views

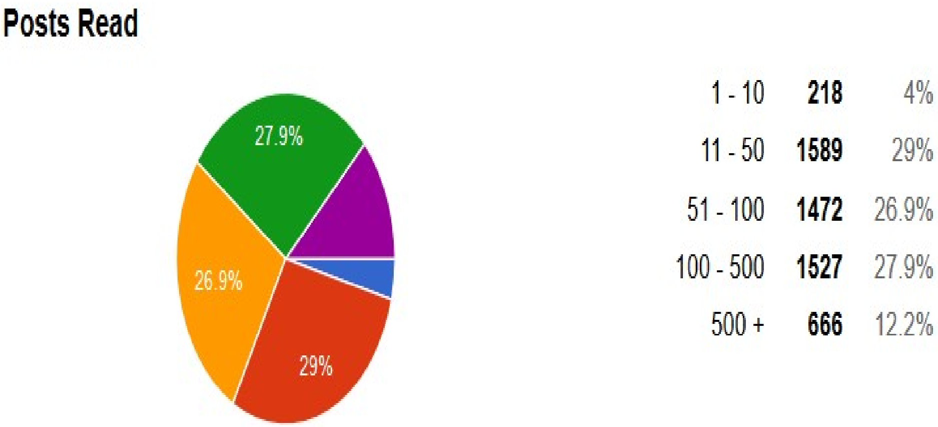
Accept / lean toward consequentialism	**2503**	46.9%
Accept / lean toward deontology	**329**	6.2%
Accept / lean toward virtue ethics	**619**	11.6%
Accept / lean toward natural law	**368**	6.9%
Accept / lean towards contractualism	**276**	5.2%
Other / No answer	**1246**	23.3%

“How long have you been reading Slate Star Codex?”
“How many of the 750 or so Slate Star Codex posts have you read?”



Length of Time

Less than a month	**97**	1.8%
One to six months	**758**	13.9%
Six months to one year	**807**	14.8%
One to two years	**1997**	36.5%
More than two years	**1812**	33.1%

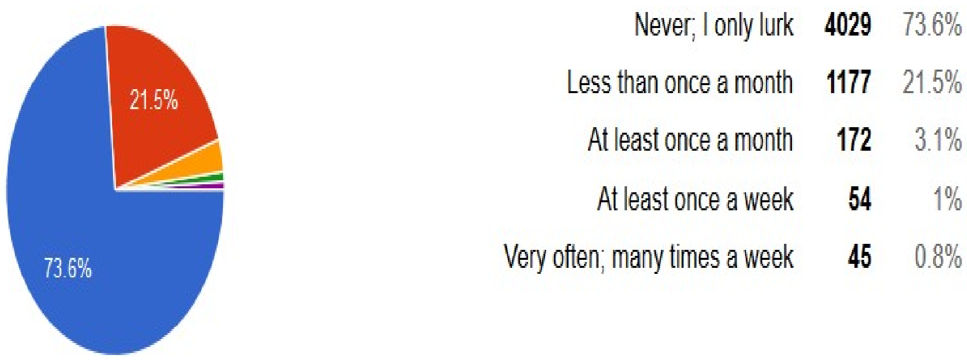


Posts Read

1 - 10	**218**	4%
11 - 50	**1589**	29%
51 - 100	**1472**	26.9%
100 - 500	**1527**	27.9%
500 +	**666**	12.2%

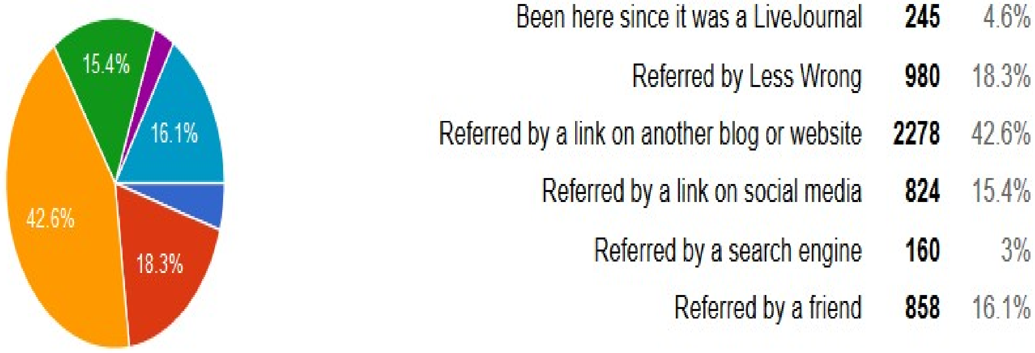
“How often do you comment on Slate Star Codex?”
“How were you referred here?”

Comment



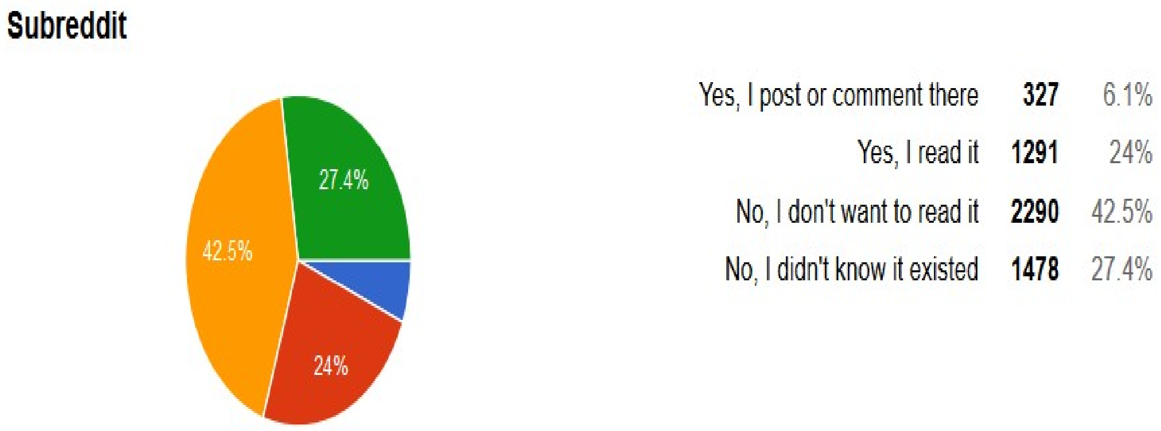
Never; I only lurk	**4029**	73.6%
Less than once a month	**1177**	21.5%
At least once a month	**172**	3.1%
At least once a week	**54**	1%
Very often; many times a week	**45**	0.8%

Referrals



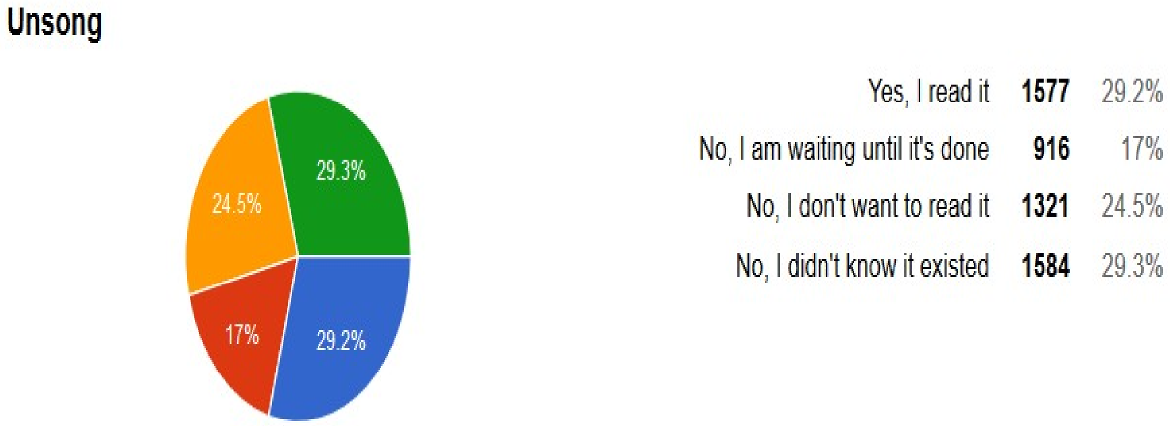
Been here since it was a LiveJournal	**245**	4.6%
Referred by Less Wrong	**980**	18.3%
Referred by a link on another blog or website	**2278**	42.6%
Referred by a link on social media	**824**	15.4%
Referred by a search engine	**160**	3%
Referred by a friend	**858**	16.1%

“Do you read the SSC subreddit?”
“Do you read Unsong?”
“Have you ever clicked on the sidebar ads?”



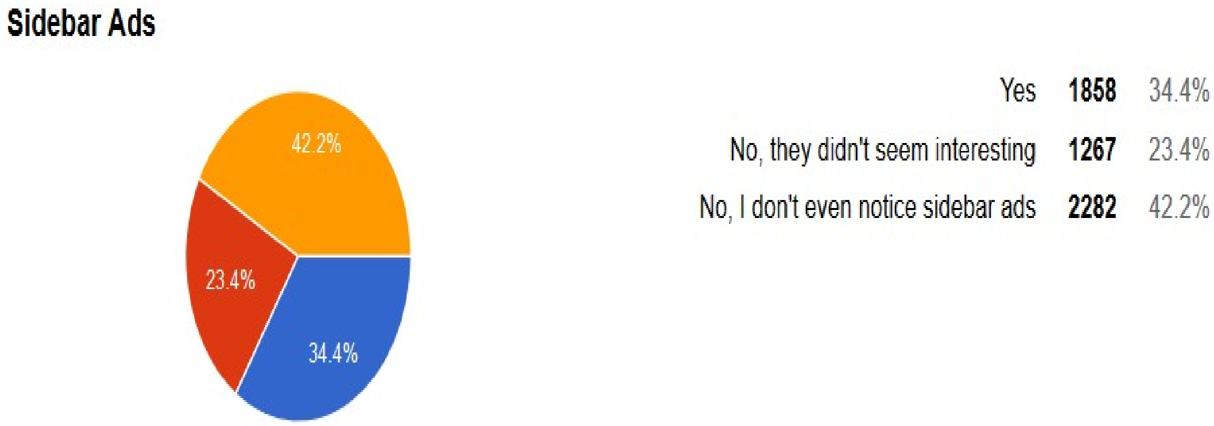
Subreddit

Yes, I post or comment there	**327**	6.1%
Yes, I read it	**1291**	24%
No, I don't want to read it	**2290**	42.5%
No, I didn't know it existed	**1478**	27.4%



Unsong

Yes, I read it	**1577**	29.2%
No, I am waiting until it's done	**916**	17%
No, I don't want to read it	**1321**	24.5%
No, I didn't know it existed	**1584**	29.3%

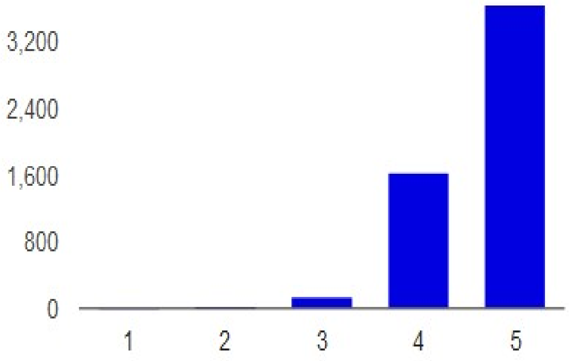


Sidebar Ads

Yes	**1858**	34.4%
No, they didn't seem interesting	**1267**	23.4%
No, I don't even notice sidebar ads	**2282**	42.2%

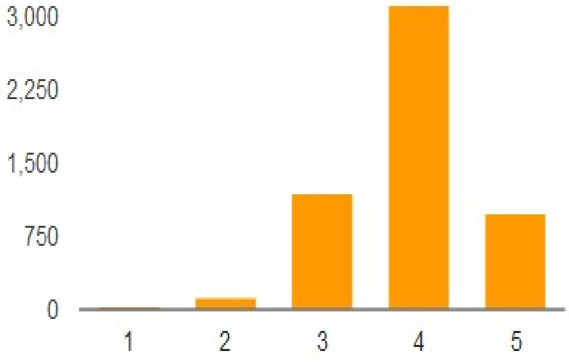
"How much do you like SSC?"
"How often do you agree with the object-level points SSC makes?"

SSC Favorability



Strongly dislike: 1	**1**	0%
2	**14**	0.3%
3	**152**	2.8%
4	**1637**	30%
Strongly like: 5	**3657**	67%

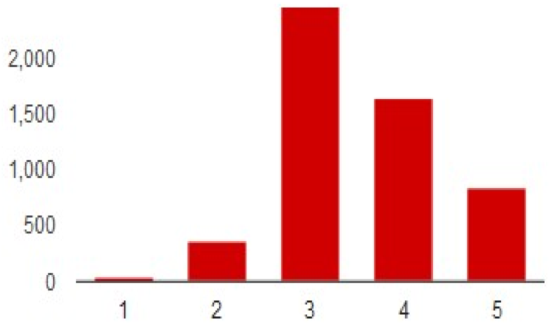
SSC Agreement



Almost never: 1	**21**	0.4%
2	**121**	2.2%
3	**1195**	22%
4	**3122**	57.4%
Almost always: 5	**979**	18%

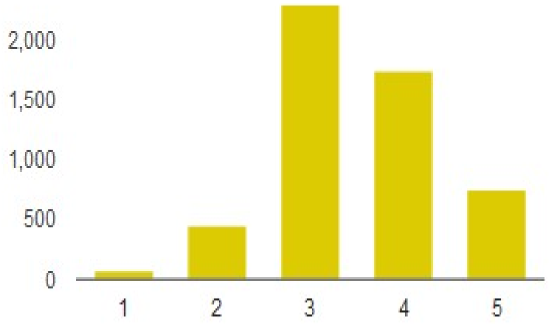
“Do you want SSC to focus more or less on the following topics?”

Topic Science



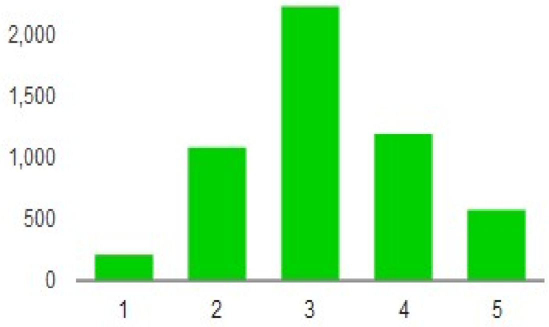
Prefer fewer posts: 1	**39**	0.7%
2	**361**	6.8%
3	**2461**	46.1%
4	**1638**	30.7%
Prefer more posts: 5	**836**	15.7%

Topic Politics



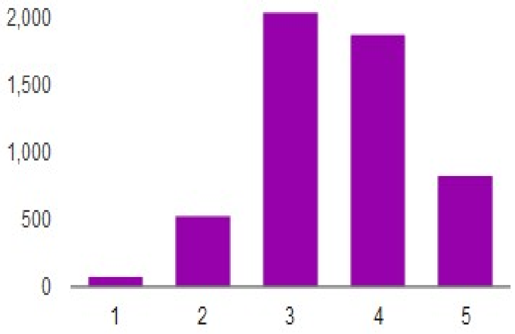
Prefer fewer posts: 1	**79**	1.5%
2	**452**	8.5%
3	**2306**	43.1%
4	**1750**	32.7%
Prefer more posts: 5	**758**	14.2%

Topic Books



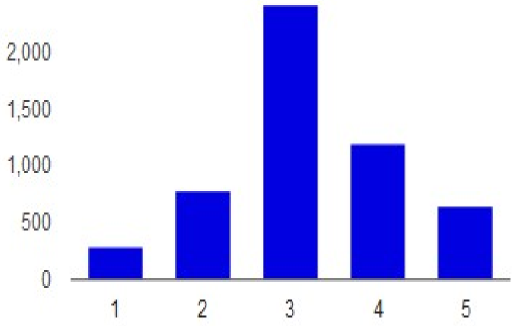
Prefer fewer posts: 1	**215**	4%
2	**1094**	20.5%
3	**2238**	42%
4	**1200**	22.5%
Prefer more posts: 5	**577**	10.8%

Topic Rationality



Prefer fewer posts: 1	**82**	1.5%
2	**523**	9.8%
3	**2042**	38.2%
4	**1872**	35%
Prefer more posts: 5	**825**	15.4%

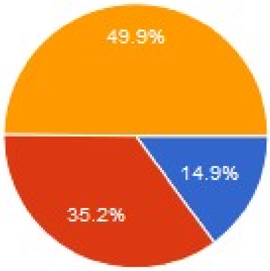
Topic Silly



Prefer fewer posts: 1	**297**	5.6%
2	**776**	14.6%
3	**2409**	45.2%
4	**1200**	22.5%
Prefer more posts: 5	**642**	12.1%

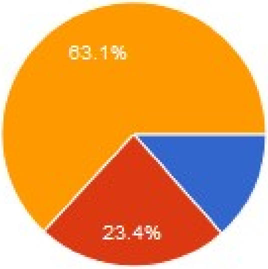
“Do you identify as a member of the following communities?”

LW ID



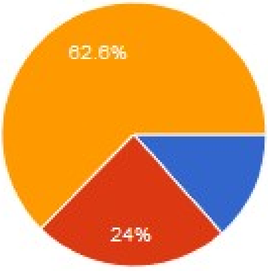
Yes	**802**	14.9%
Sorta	**1894**	35.2%
No	**2680**	49.9%

H+ ID



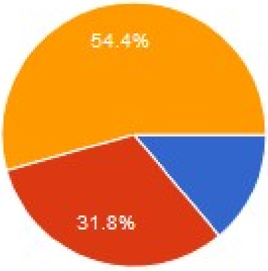
Yes	**723**	13.5%
Sorta	**1259**	23.4%
No	**3390**	63.1%

SJ ID



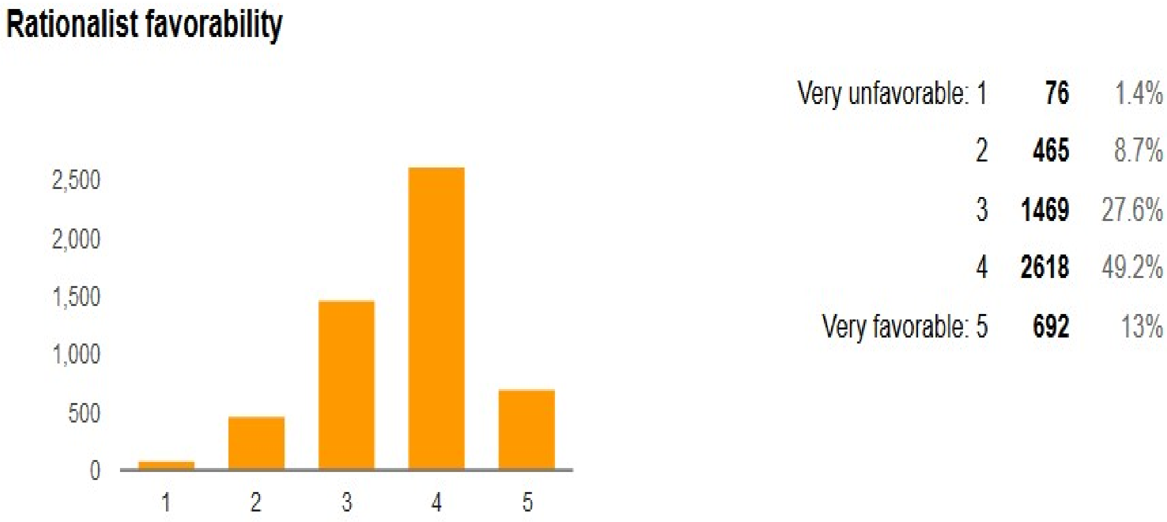
Yes	**723**	13.4%
Sorta	**1294**	24%
No	**3379**	62.6%

EA ID



Yes	**743**	13.8%
Sorta	**1705**	31.8%
No	**2918**	54.4%

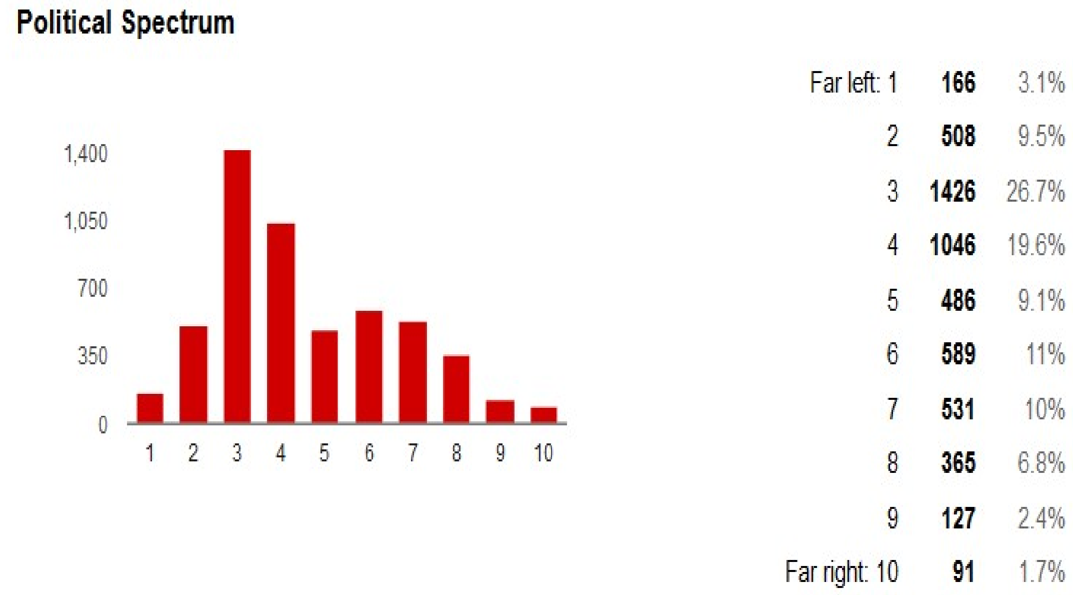
“What is your opinion of the rationalist commmunity?”



Rationalist favorability

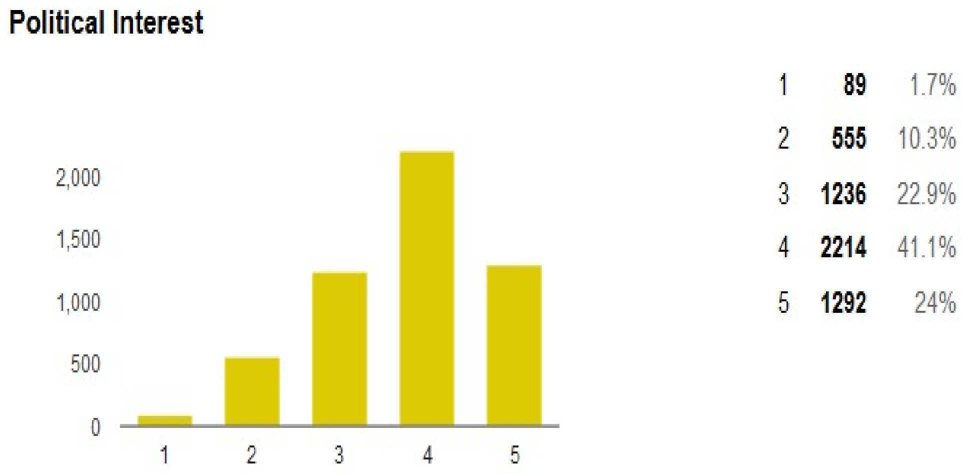
Very unfavorable: 1	**76**	1.4%
2	**465**	8.7%
3	**1469**	27.6%
4	**2618**	49.2%
Very favorable: 5	**692**	13%

“Where do you fall on a classic political spectrum?”
“How interested are you in politics?”



Political Spectrum

Far left: 1	**166**	3.1%
2	**508**	9.5%
3	**1426**	26.7%
4	**1046**	19.6%
5	**486**	9.1%
6	**589**	11%
7	**531**	10%
8	**365**	6.8%
9	**127**	2.4%
Far right: 10	**91**	1.7%

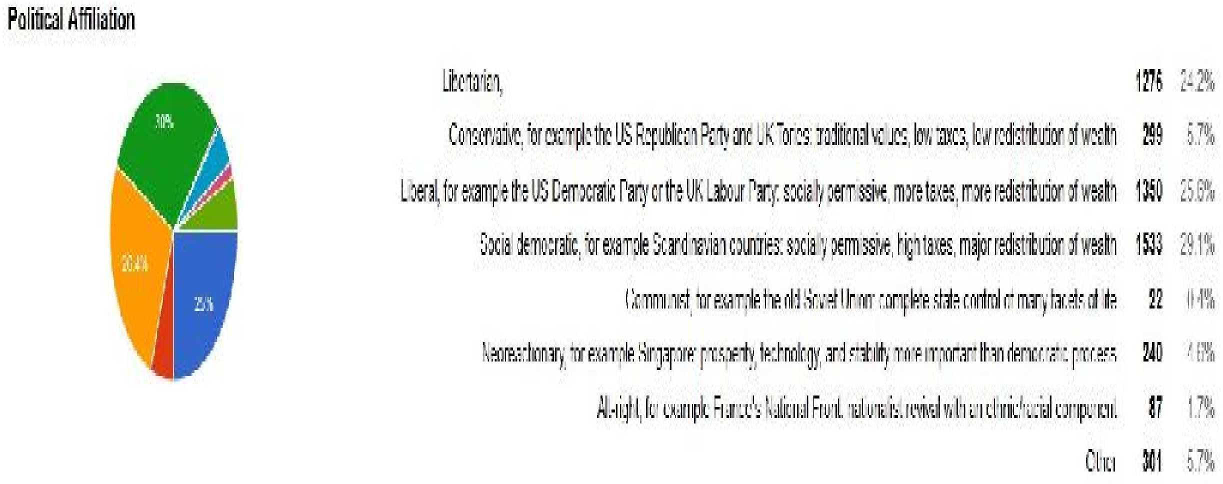


Political Interest

1	**89**	1.7%
2	**555**	10.3%
3	**1236**	22.9%
4	**2214**	41.1%
5	**1292**	24%

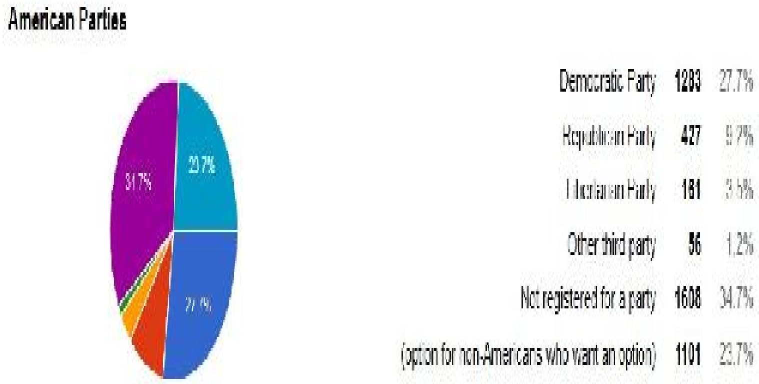
“Which of these political philosophies do you most identify with?”

“If you are an American, what party are you registered with?”



Political Affiliation

Libertarian,	1276	24.2%
Conservative, for example the US Republican Party and UK Tories: traditional values, low taxes, low redistribution of wealth	299	5.7%
Liberal, for example the US Democratic Party or the UK Labour Party: socially permissive, more taxes, more redistribution of wealth	1350	25.6%
Social democratic, for example Scandinavian countries: socially permissive, high taxes, major redistribution of wealth	1533	29.1%
Communist, for example the old Soviet Union: complete state control of many facets of life	22	0.4%
Neoreactionary, for example Singapore: prosperity, technology, and stability more important than democratic process	240	4.6%
Alt-right, for example France's National Front: nationalist revival with an ethnic/racial component	87	1.7%
Other	301	5.7%

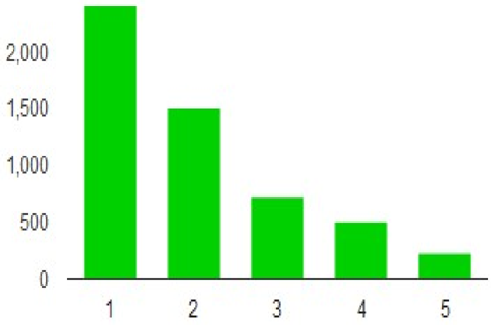


American Parties

Democratic Party	1283	27.7%
Republican Party	427	9.2%
Libertarian Party	161	3.5%
Other third party	56	1.2%
Not registered for a party	1608	34.7%
(option for non-Americans who want an option)	1101	23.7%

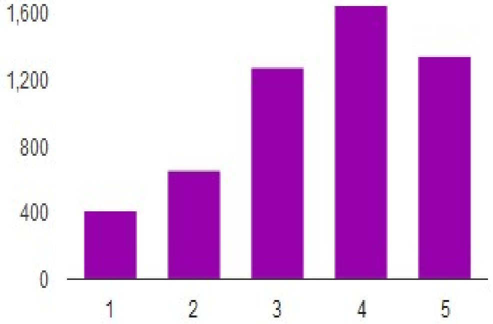
“What is your position on the following issues?”

Global Warming



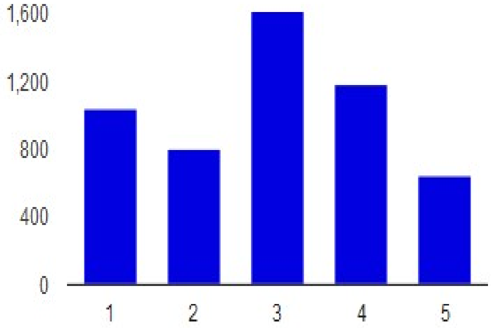
Requires strong action: 1	**2415**	44.8%
2	**1511**	28%
3	**729**	13.5%
4	**505**	9.4%
Does not require action: 5	**233**	4.3%

Immigration



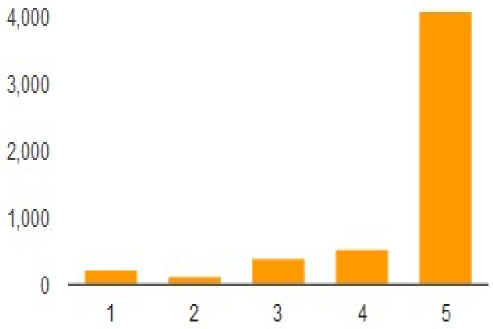
Should be stricter: 1	**422**	7.9%
2	**665**	12.4%
3	**1283**	23.9%
4	**1654**	30.8%
Should be more open: 5	**1341**	25%

Minimum Wage



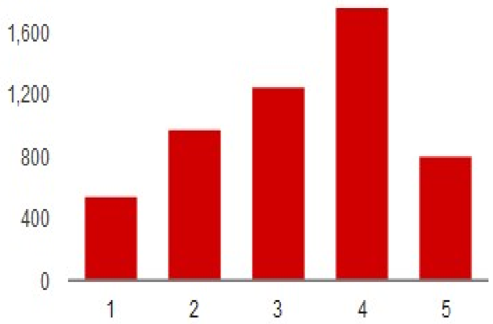
Should be lower or eliminated: 1	**1044**	19.6%
2	**807**	15.2%
3	**1622**	30.5%
4	**1190**	22.4%
Should be higher: 5	**650**	12.2%

Gay Marriage



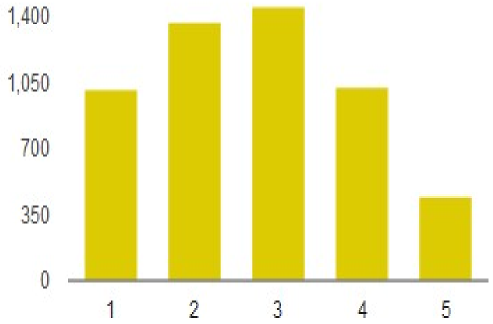
Should not be legally recognized: 1	**225**	4.2%
2	**140**	2.6%
3	**407**	7.6%
4	**520**	9.7%
Should be legally recognized: 5	**4093**	76%

Feminism

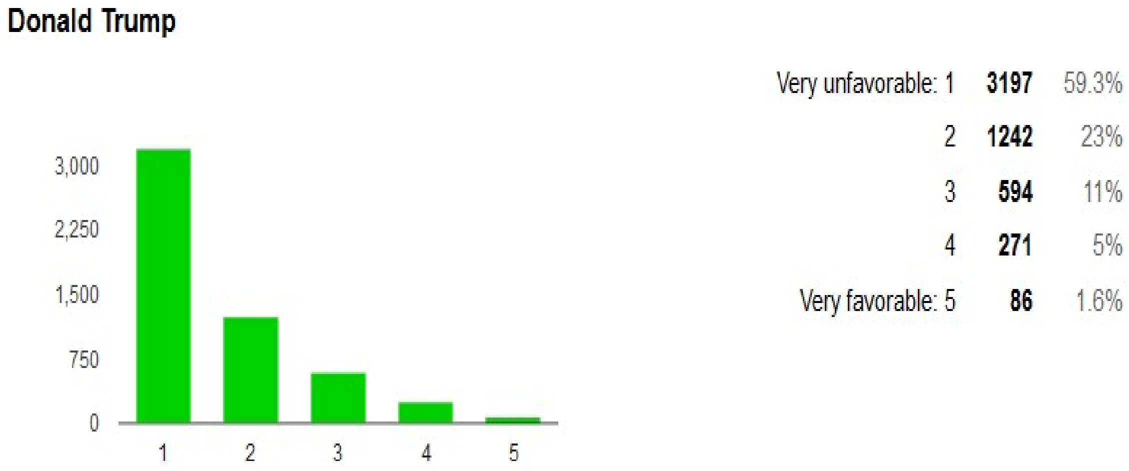


Very unfavorable: 1	**543**	10.1%
2	**984**	18.3%
3	**1258**	23.4%
4	**1772**	33%
Very favorable: 5	**813**	15.1%

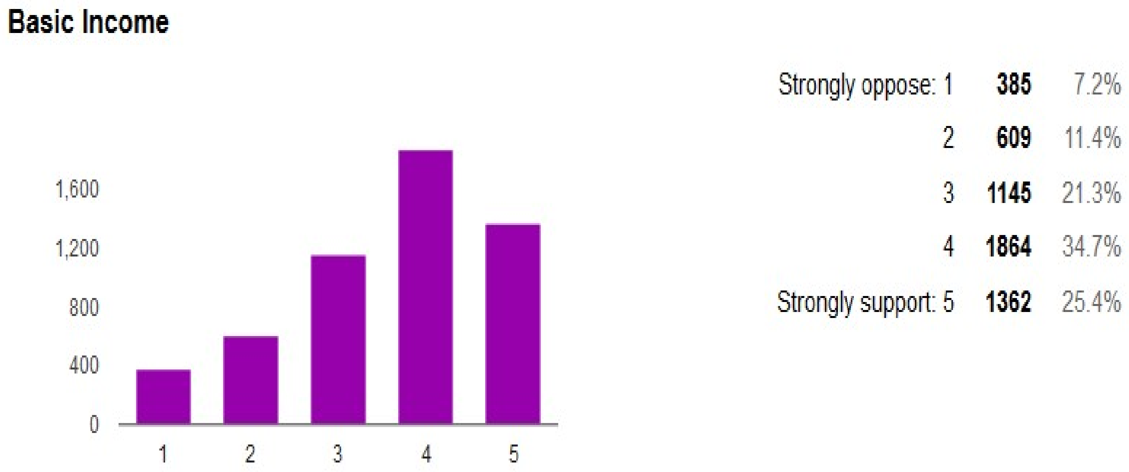
Human Biodiversity



Very unfavorable: 1	**1020**	19.2%
2	**1371**	25.8%
3	**1452**	27.3%
4	**1027**	19.3%
Very favorable: 5	**448**	8.4%

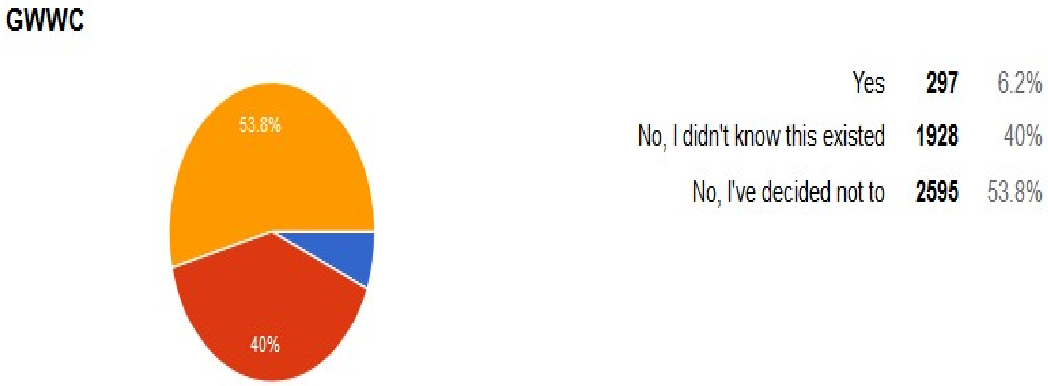

Donald Trump

Very unfavorable: 1	**3197**	59.3%
2	**1242**	23%
3	**594**	11%
4	**271**	5%
Very favorable: 5	**86**	1.6%


Basic Income

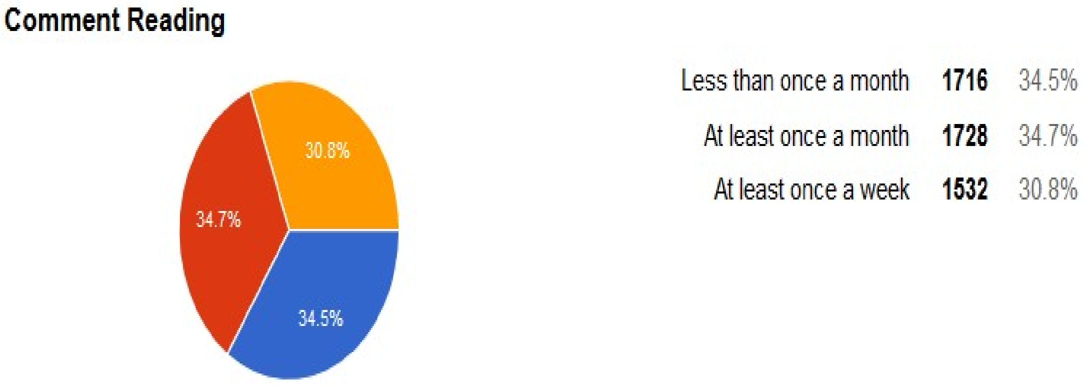
Strongly oppose: 1	**385**	7.2%
2	**609**	11.4%
3	**1145**	21.3%
4	**1864**	34.7%
Strongly support: 5	**1362**	25.4%

“Have you taken the Giving What We Can pledge?”



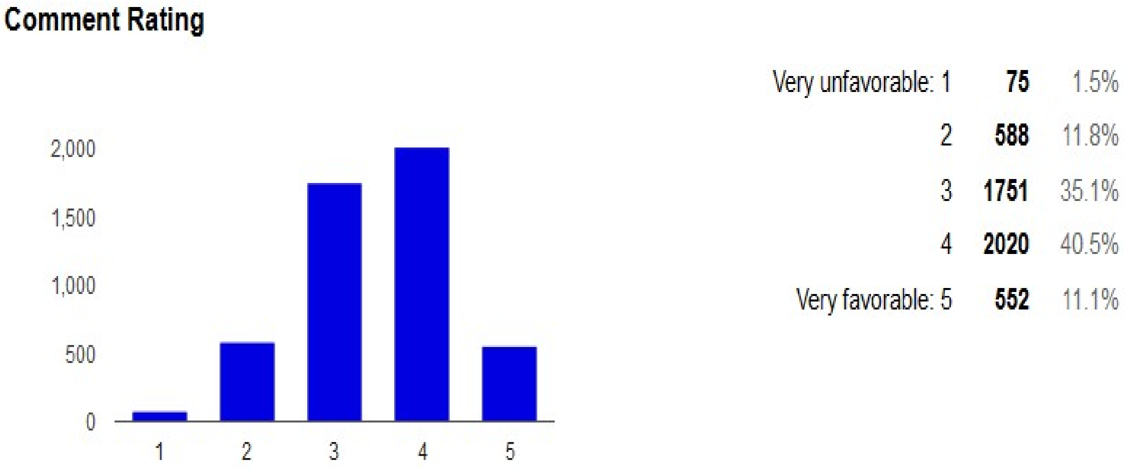
Yes	**297**	6.2%
No, I didn't know this existed	**1928**	40%
No, I've decided not to	**2595**	53.8%

“How often do you read the SSC comments?”
“What is your opinion of the SSC comment section?”



Comment Reading

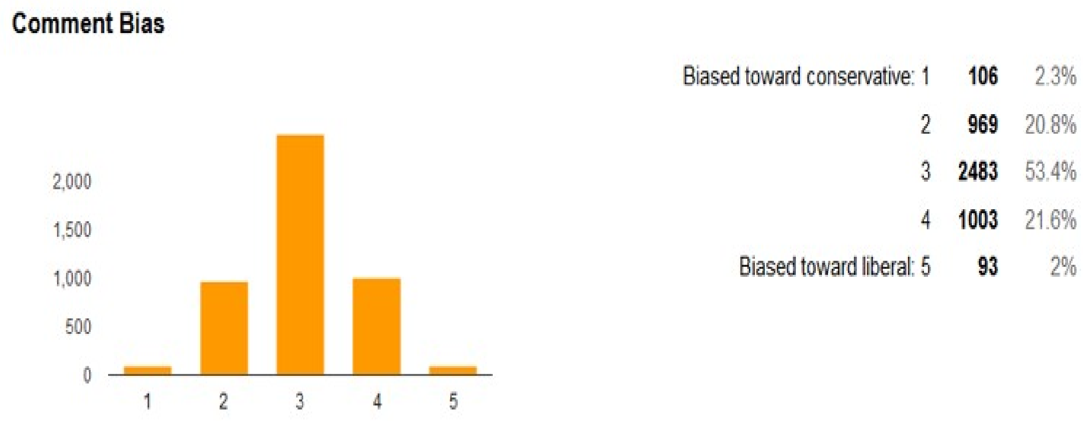
Less than once a month	**1716**	34.5%
At least once a month	**1728**	34.7%
At least once a week	**1532**	30.8%



Comment Rating

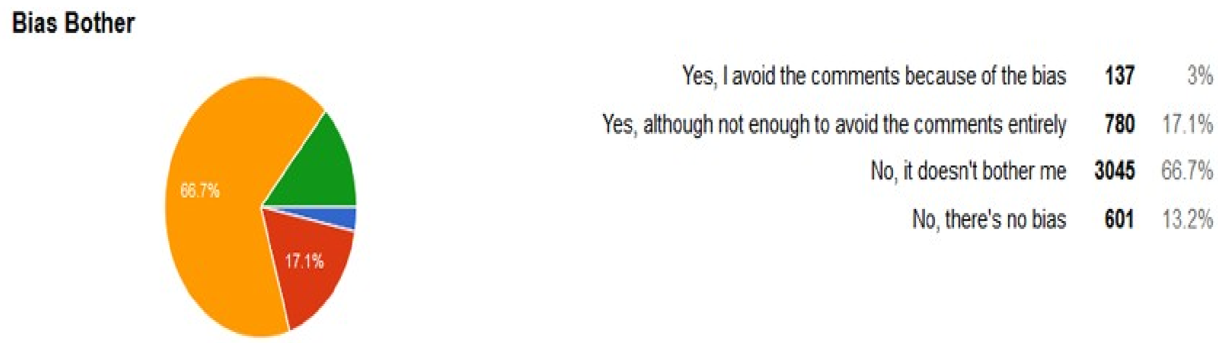
Very unfavorable: 1	**75**	1.5%
2	**588**	11.8%
3	**1751**	35.1%
4	**2020**	40.5%
Very favorable: 5	**552**	11.1%

“Do you find an ideological bias among SSC commenters?”
“Are you bothered by a bias in SSC comments?”



Comment Bias

Biased toward conservative: 1	**106**	2.3%
2	**969**	20.8%
3	**2483**	53.4%
4	**1003**	21.6%
Biased toward liberal: 5	**93**	2%



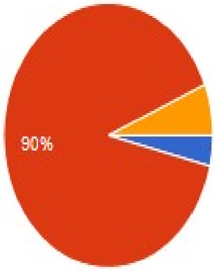
Bias Bother

Yes, I avoid the comments because of the bias	**137**	3%
Yes, although not enough to avoid the comments entirely	**780**	17.1%
No, it doesn't bother me	**3045**	66.7%
No, there's no bias	**601**	13.2%

“What is your opinion on the level of comment moderation?”

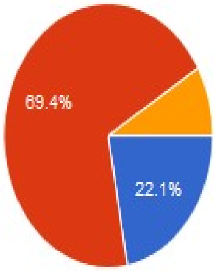
“What is your opinion on the level of identity politics discussion in the comments?”

Moderation Policy



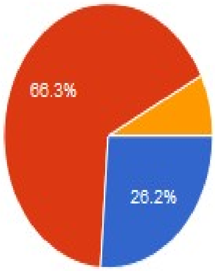
Moderation is too strict	**150**	3.7%
Just right	**3646**	90%
Moderation is too lax	**254**	6.3%

Identity Politics Comments



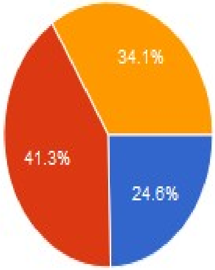
Too common	**855**	22.1%
Just right	**2687**	69.4%
Too restricted	**330**	8.5%

Mandatory registration



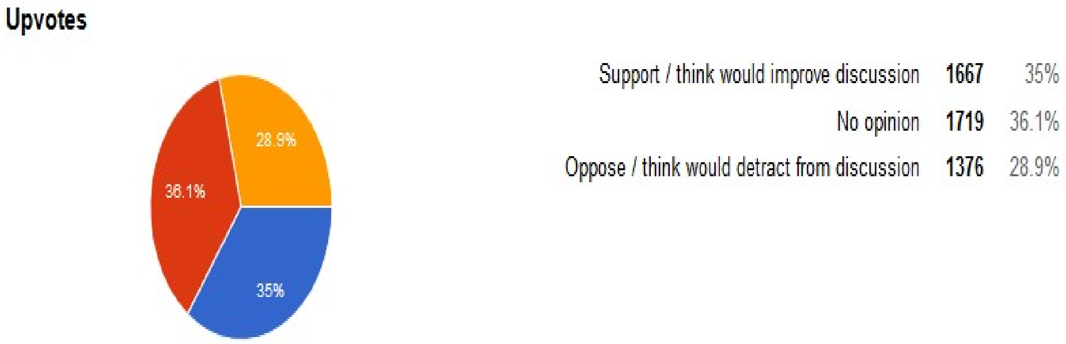
Good / successful / improved quality of discussion	**1203**	26.2%
No opinion	**3045**	66.3%
Bad / too burdensome / detracted from quality of discussion	**345**	7.5%

Knowledge question



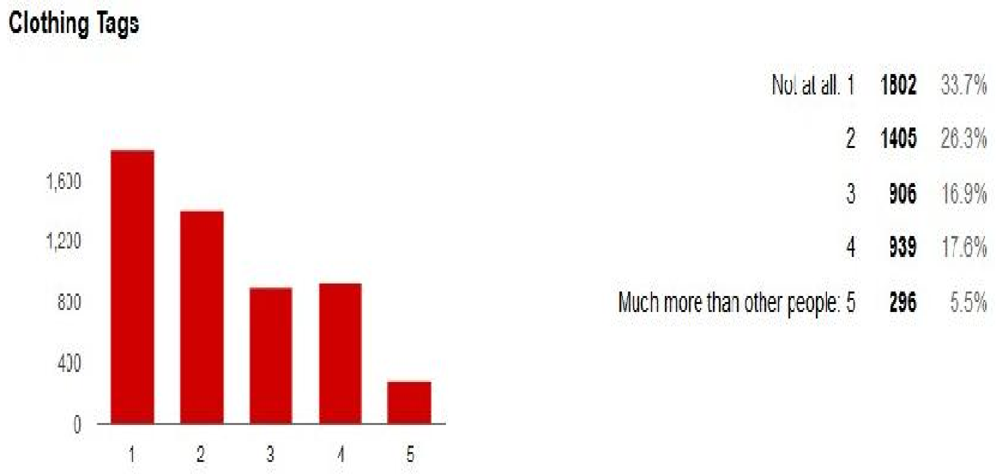
Support / think would improve discussion	**1172**	24.6%
No opinion	**1968**	41.3%
Oppose / think would detract from discussion	**1622**	34.1%

“What is your opinion on the recent policy of requiring all commenters to register accounts?”
“What would you think of requiring new commenters to answer a knowledge question before being allowed to register?”


Upvotes

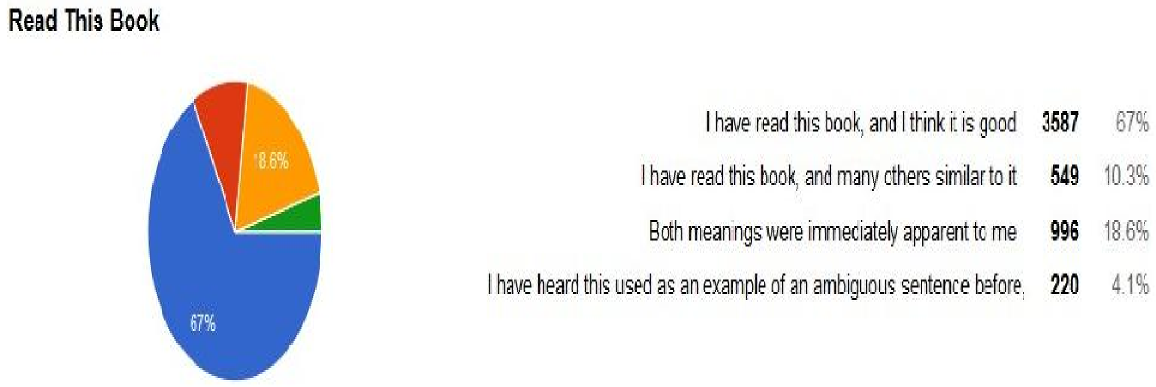
Support / think would improve discussion	**1667**	35%
No opinion	**1719**	36.1%
Oppose / think would detract from discussion	**1376**	28.9%

“Are you bothered by scratchy tags on your clothing?”
“How do you interpret the sentence ‘I have read this book and much like it’?”
“What direction do you see in the [spinning dancer illusion]?”



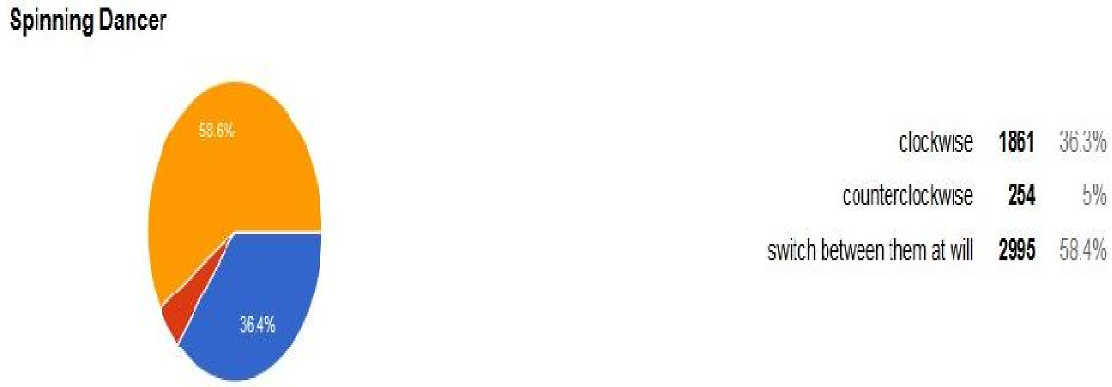
Clothing Tags

Not at all: 1	**1802**	33.7%
2	**1405**	26.3%
3	**906**	16.9%
4	**939**	17.6%
Much more than other people: 5	**296**	5.5%



Read This Book

I have read this book, and I think it is good	**3587**	67%
I have read this book, and many others similar to it	**549**	10.3%
Both meanings were immediately apparent to me	**996**	18.6%
I have heard this used as an example of an ambiguous sentence before,	**220**	4.1%



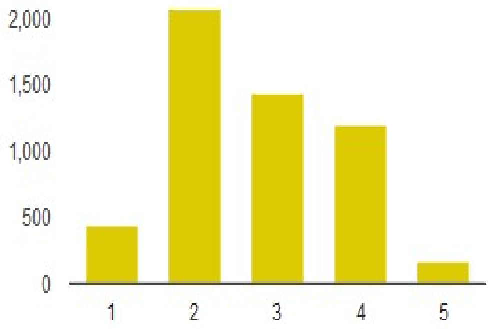
Spinning Dancer

clockwise	**1861**	36.3%
counterclockwise	**254**	5%
switch between them at will	**2995**	58.4%

“Which do you think is more important when trying to learn a new skill: hard work or talent?”

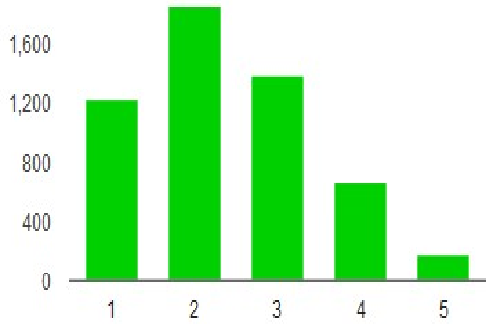
“Do you find it hard to follow conversations in noisy areas?”
“How many of the three duplications of the word ‘the’ (ie ‘the the’) did you notice in this survey?”

Talent



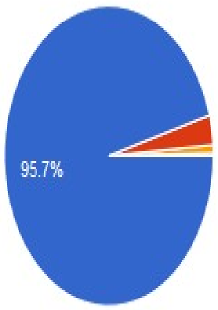
Hard work: 1	**443**	8.3%
2	**2076**	38.9%
3	**1437**	27%
4	**1204**	22.6%
Talent: 5	**171**	3.2%

Noisy Conversations



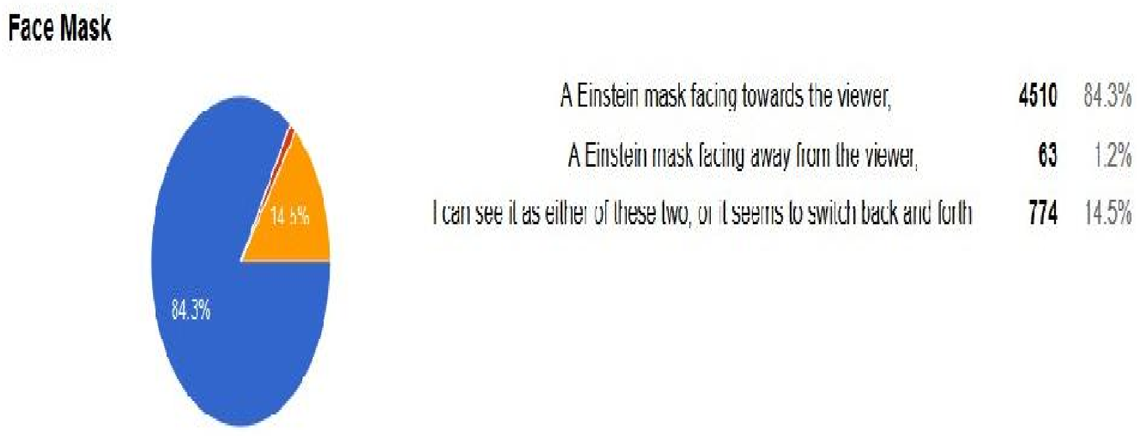
Worse than most other people: 1	**1225**	22.9%
2	**1863**	34.9%
3	**1400**	26.2%
4	**672**	12.6%
Better than most other people: 5	**181**	3.4%

Duplicate Thes

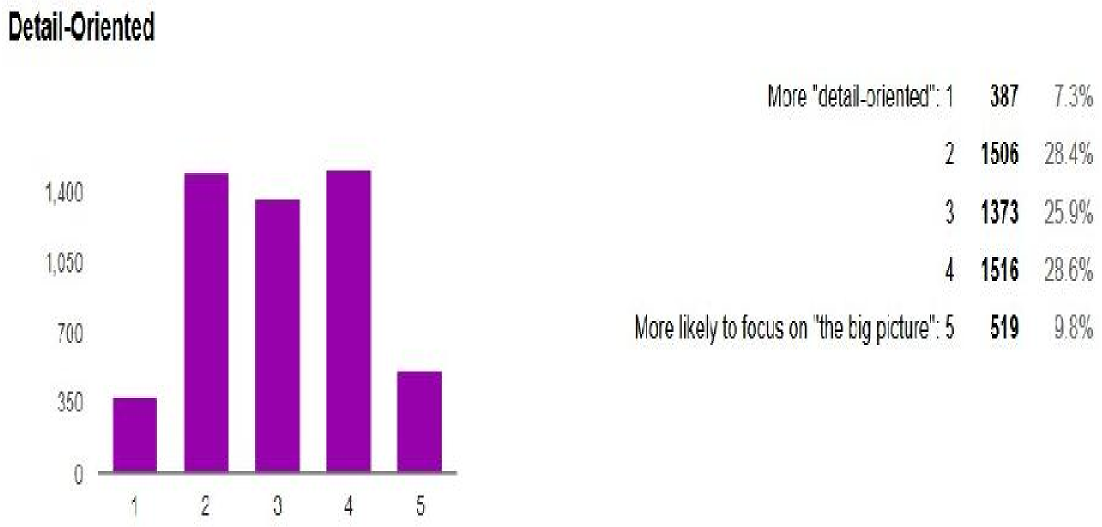


None	**5125**	95.7%
One	**170**	3.2%
Two	**45**	0.8%
All three	**16**	0.3%

“How do you interpret the Einstein mask illusion?
“Do you think of yourself as more detail-oriented or more big-picture?”

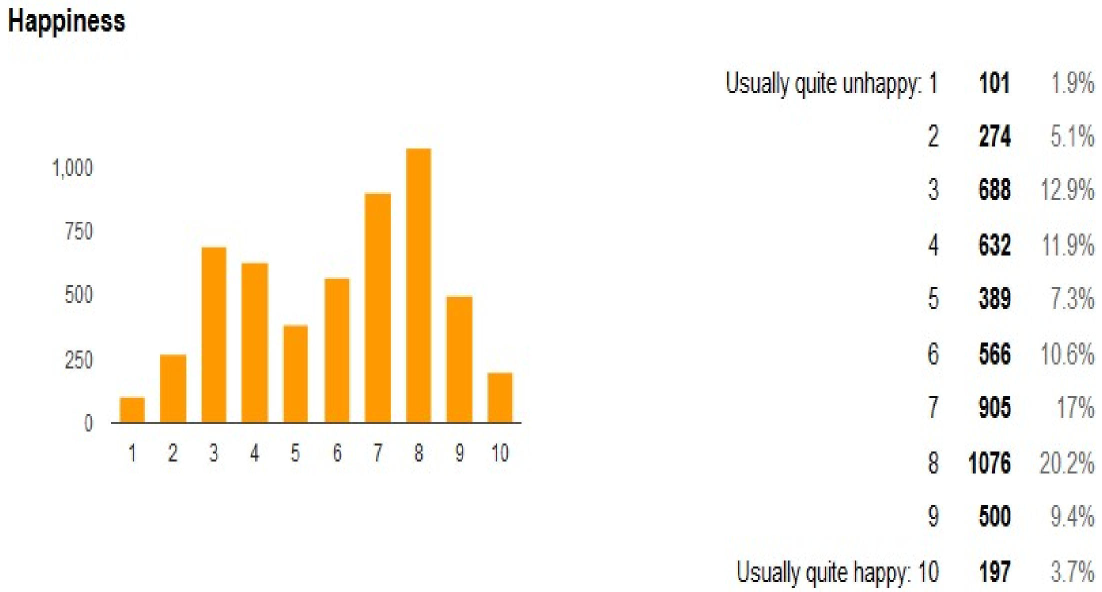

Face Mask

A Einstein mask facing towards the viewer,	4510	84.3%
A Einstein mask facing away from the viewer,	63	1.2%
I can see it as either of these two, or it seems to switch back and forth	774	14.5%


Detail-Oriented

More "detail-oriented": 1	387	7.3%
2	1506	28.4%
3	1373	25.9%
4	1516	28.6%
More likely to focus on "the big picture": 5	519	9.8%

“How happy do you generally feel?”



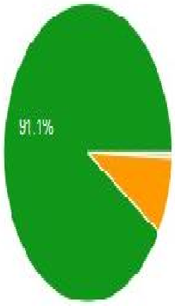
1,000
750
500
250
0
1
2
3
4
5
6
7
8
9
10

Happiness

Usually quite unhappy: 1	**101**	1.9%
2	**274**	5.1%
3	**688**	12.9%
4	**632**	11.9%
5	**389**	7.3%
6	**566**	10.6%
7	**905**	17%
8	**1076**	20.2%
9	**500**	9.4%
Usually quite happy: 10	**197**	3.7%

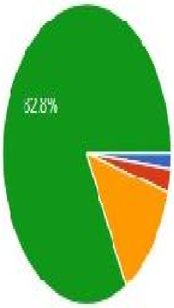
“Do you have any of the following psychiatric diagnoses?”

Schizophrenia



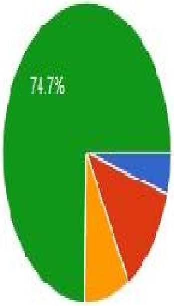
I have a formal diagnosis of this condition	**10**	0.2%
I think I might have this condition, although I have never been formally diagnosed	**19**	0.4%
I have family members (within two generations) with this condition	**446**	8.3%
I don't have this condition and neither does anyone in my family	**4869**	91.1%

Bipolar



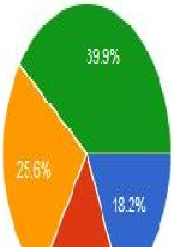
I have a formal diagnosis of this condition	**85**	1.6%
I think I might have this condition, although I have never been formally diagnosed	**134**	2.5%
I have family members (within two generations) with this condition	**699**	13.1%
I don't have this condition and neither does anyone in my family	**4404**	82.8%

Autism



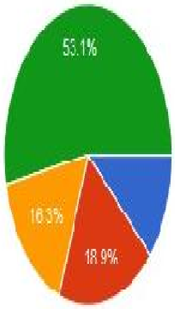
I have a formal diagnosis of this condition	**236**	4.4%
I think I might have this condition, although I have never been formally diagnosed	**651**	12.2%
I have family members (within two generations) with this condition	**457**	8.6%
I don't have this condition and neither does anyone in my family	**3973**	74.7%

Depression



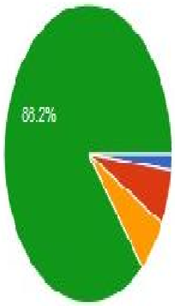
I have a formal diagnosis of this condition	**967**	18.2%
I think I might have this condition, although I have never been formally diagnosed	**873**	16.4%
I have family members (within two generations) with this condition	**1361**	25.6%
I don't have this condition and neither does anyone in my family	**2123**	39.9%

Anxiety

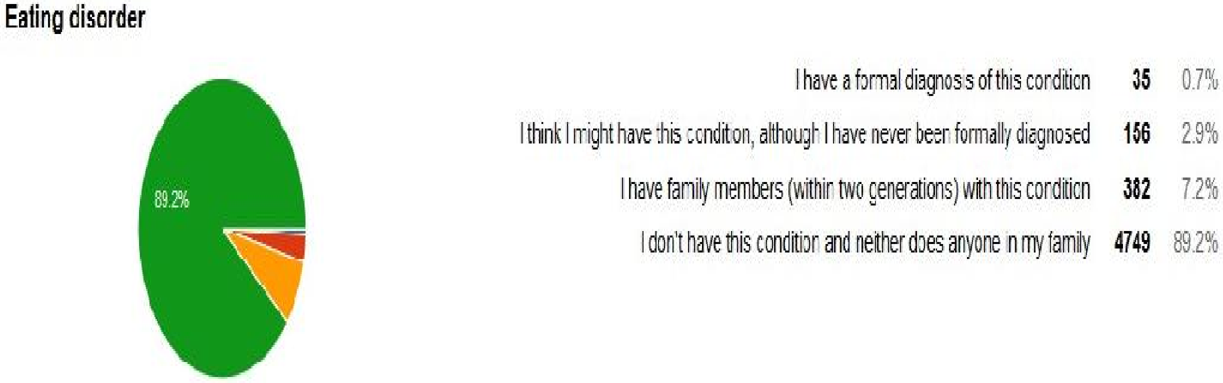


I have a formal diagnosis of this condition	**620**	11.7%
I think I might have this condition, although I have never been formally diagnosed	**1005**	18.9%
I have family members (within two generations) with this condition	**864**	16.3%
I don't have this condition and neither does anyone in my family	**2822**	53.1%

OCD

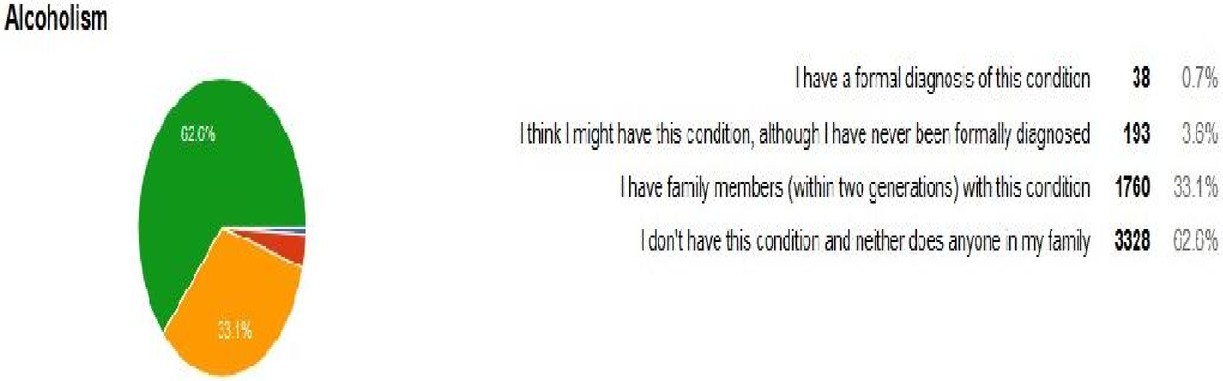


I have a formal diagnosis of this condition	**97**	1.8%
I think I might have this condition, although I have never been formally diagnosed	**315**	5.9%
I have family members (within two generations) with this condition	**320**	6%
I don't have this condition and neither does anyone in my family	**4583**	86.2%



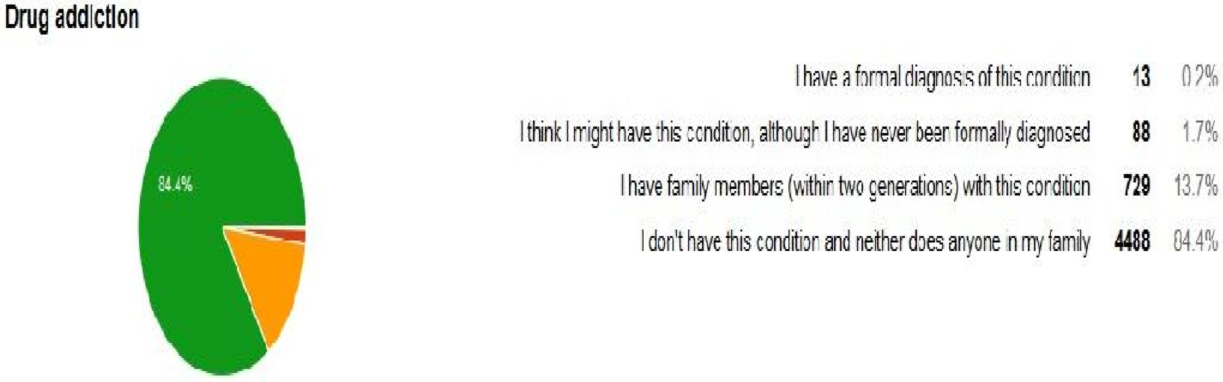
Eating disorder

I have a formal diagnosis of this condition	**35**	0.7%
I think I might have this condition, although I have never been formally diagnosed	**156**	2.9%
I have family members (within two generations) with this condition	**382**	7.2%
I don't have this condition and neither does anyone in my family	**4749**	89.2%



Alcoholism

I have a formal diagnosis of this condition	**38**	0.7%
I think I might have this condition, although I have never been formally diagnosed	**193**	3.6%
I have family members (within two generations) with this condition	**1760**	33.1%
I don't have this condition and neither does anyone in my family	**3328**	62.6%

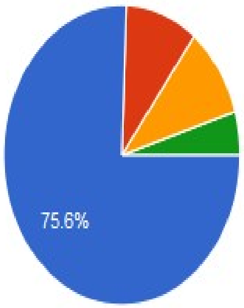


Drug addiction

I have a formal diagnosis of this condition	**13**	0.2%
I think I might have this condition, although I have never been formally diagnosed	**88**	1.7%
I have family members (within two generations) with this condition	**729**	13.7%
I don't have this condition and neither does anyone in my family	**4488**	84.4%

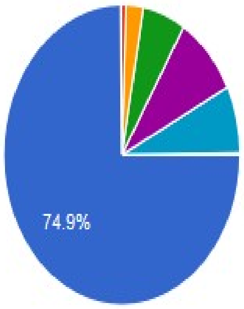
“Have you been on SSRIs?”
“At what age did you start SSRIs?”
“Do you currently use SSRIs?”

SSRIs



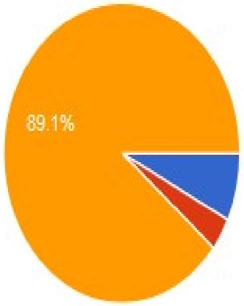
I have never taken these drugs	**4031**	75.6%
Less than one year	**526**	9.9%
One to five years	**526**	9.9%
Greater than five years	**249**	4.7%

SSRI start



I have never taken these drugs	**3717**	74.9%
Younger than ten	**34**	0.7%
Ten to fifteen	**118**	2.4%
Fifteen to twenty	**289**	5.8%
Twenty to twenty-five	**437**	8.8%
Older than twenty-five	**359**	7.2%
Other	**11**	0.2%

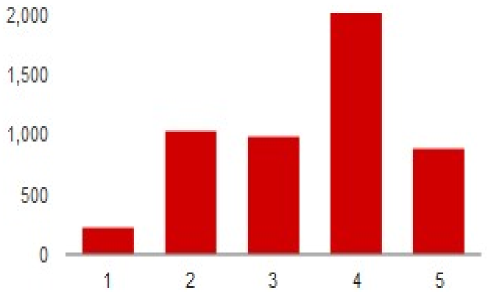
SSRIs now



Yes	**376**	7.3%
No, but I did within the past year	**181**	3.5%
No, neither now nor in the past year	**4564**	89.1%

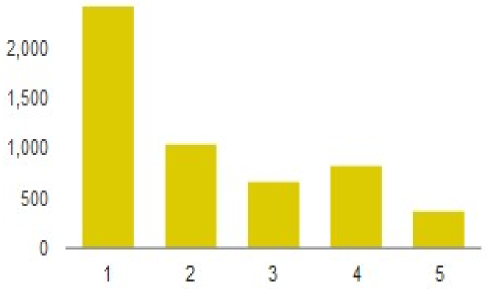
“Do you think of yourself as more ‘growth mindset’ or ‘fixed mindset’?”
“Do you lift weights?”
“Regarding sleep, are you an ‘early bird’ or a ‘night owl’?”

Growth Mindset



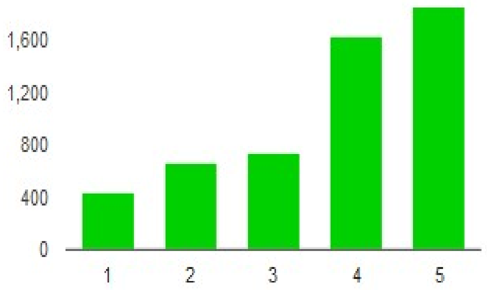
Fixed mindset: 1	**239**	4.6%
2	**1033**	20%
3	**990**	19.1%
4	**2029**	39.2%
Growth mindset: 5	**886**	17.1%

Weightlifting?



Never: 1	**2429**	45.4%
2	**1040**	19.4%
3	**675**	12.6%
4	**835**	15.6%
Daily / very often: 5	**369**	6.9%

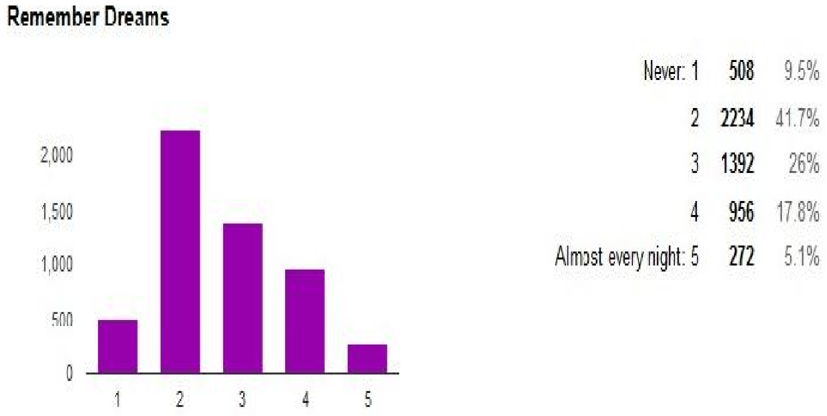
Night Owl



Early bird: 1	**443**	8.3%
2	**661**	12.4%
3	**741**	13.9%
4	**1638**	30.7%
Night owl: 5	**1858**	34.8%

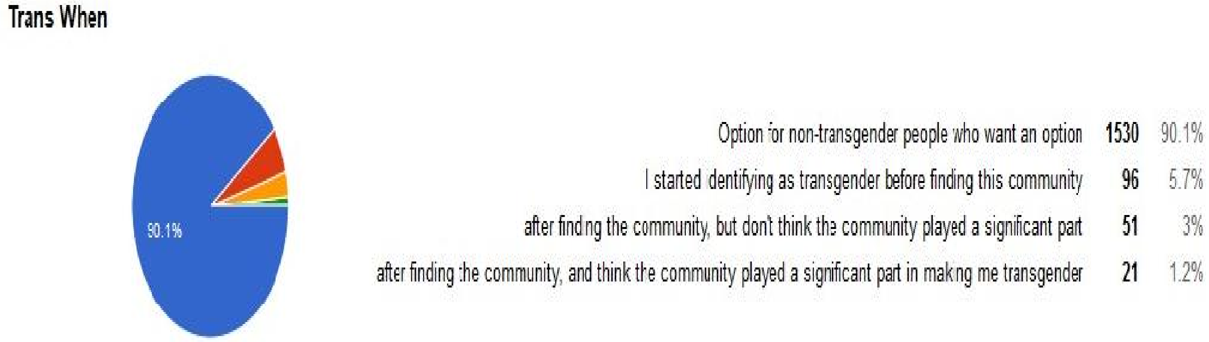
“How often do you remember your dreams?”
“If you are transgender, when did you realize this?”

“Does your internal thought process feel verbal or nonverbal?”



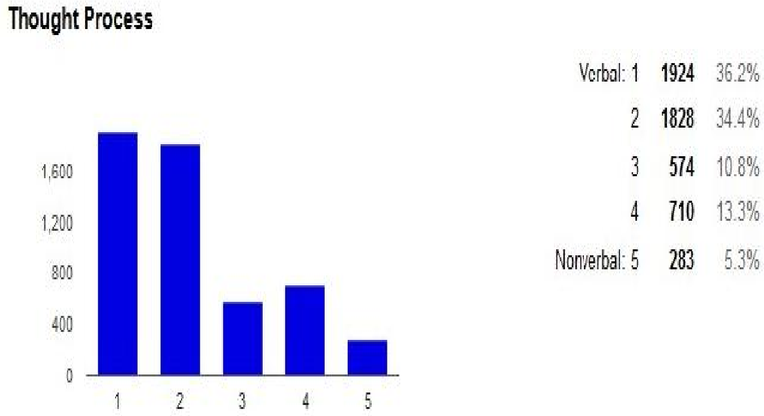
Remember Dreams

Never: 1	508	9.5%
2	2234	41.7%
3	1392	26%
4	956	17.8%
Almost every night: 5	272	5.1%



Trans When

Option for non-transgender people who want an option	1530	90.1%
I started identifying as transgender before finding this community	96	5.7%
after finding the community, but dont think the community played a significant part	51	3%
after finding the community, and think the community played a significant part in making me transgender	21	1.2%

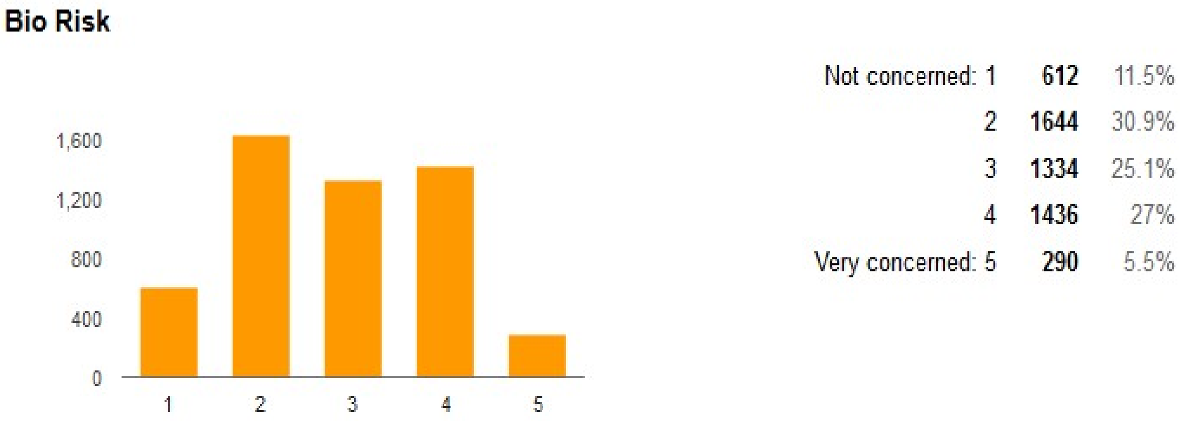


Thought Process

Verbal: 1	1924	36.2%
2	1828	34.4%
3	574	10.8%
4	710	13.3%
Nonverbal: 5	283	5.3%

“How concerned are you about bioterrorism as a threat to humanity?”

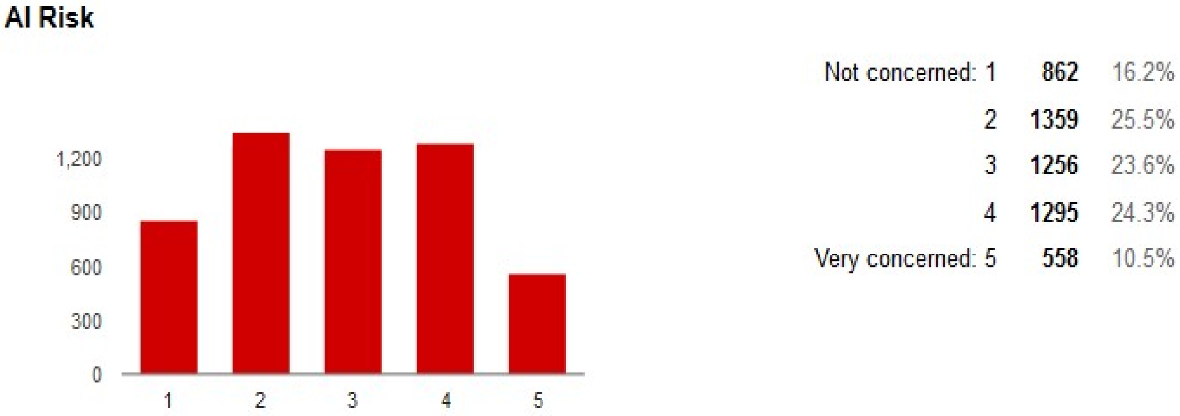
“How concerned are you about superintelligent AI as a threat to humanity?”
“Are you able to clearly visualize images in your mind through imagination?”



Bio Risk

Not concerned: 1	**612**	11.5%
2	**1644**	30.9%
3	**1334**	25.1%
4	**1436**	27%
Very concerned: 5	**290**	5.5%

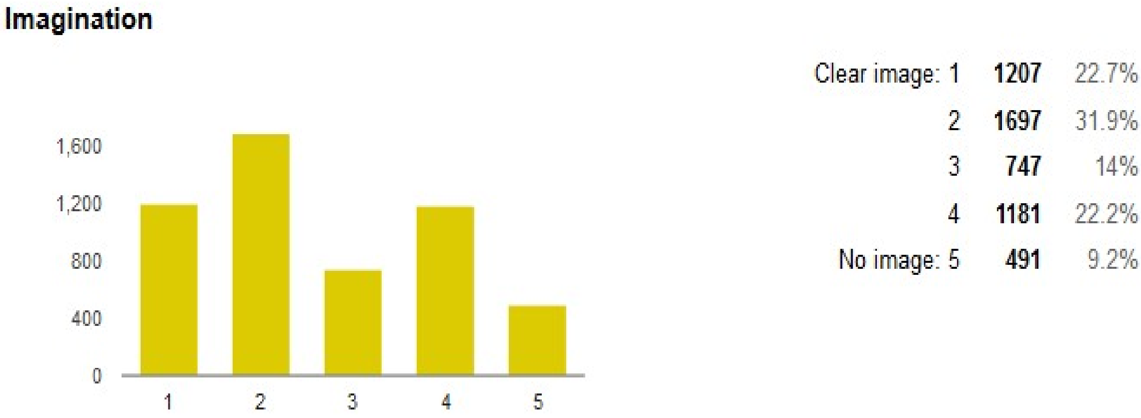
1,600
1,200
800
400
0
1
2
3
4
5



AI Risk

Not concerned: 1	**862**	16.2%
2	**1359**	25.5%
3	**1256**	23.6%
4	**1295**	24.3%
Very concerned: 5	**558**	10.5%

1,200
900
600
300
0
1
2
3
4
5



Imagination

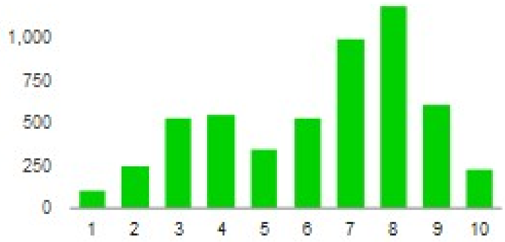
Clear image: 1	**1207**	22.7%
2	**1697**	31.9%
3	**747**	14%
4	**1181**	22.2%
No image: 5	**491**	9.2%

1,600
1,200
800
400
0
1
2
3
4
5

“Overall, how satisfied are you with life?”
“How much do you enjoy puns?”

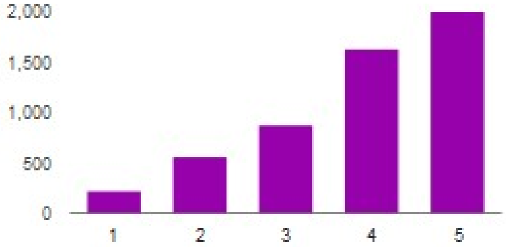
“Do you think other people are basically trustworthy?”
“Do you think you are ‘a typical SSC reader’?”

Life Satisfaction



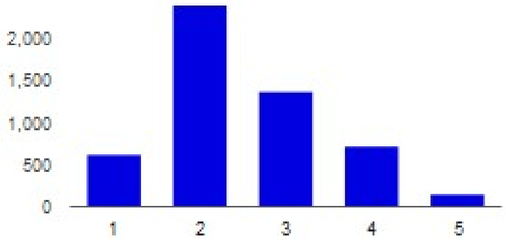
Very unsatisfied: 1	**107**	2%
2	**252**	4.7%
3	**532**	10%
4	**546**	10.2%
5	**348**	6.5%
6	**534**	10%
7	**997**	18.7%
8	**1188**	22.2%
9	**608**	11.4%
Very satisfied: 10	**228**	4.3%

Puns



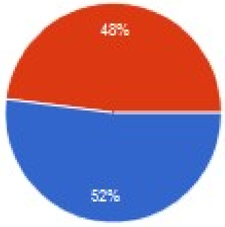
Not at all: 1	**234**	4.4%
2	**581**	10.8%
3	**886**	16.5%
4	**1638**	30.6%
Very much: 5	**2019**	37.7%

Trustworthy



Basically trustworthy: 1	**634**	11.9%
2	**2427**	45.5%
3	**1383**	25.9%
4	**731**	13.7%
Basically not trustworthy: 5	**159**	3%

Typical



Yes	**2539**	52%
No	**2347**	48%

II.

Some means:

Age: 30.6

IQ: 138.5
SAT out of 1600: 1471.9
SAT out of 2400: 2218.68

(Congratulations to the person who got a 1650 on their SAT/1600, *and* the person who got a 2450 on their SAT/2400. You clearly have bright futures ahead of you.)

Income (mean): $96,443.5
Income (median): $57,000

The highest observed incomes were in the $10,000,000 range; I know some big venture capitalists read this so I didn’t delete them as obvious trolls. Removing everyone who makes over $1 million, mean income goes down to $79,000. But there were also people who put down incomes of 0 because they were students, unemployed, or homemakers. When these people are also taken out, the mean of the remaining 2,700 people goes back to $98,000, and the median to $75,000.

Below are some Likert scales. Note that the midpoint is *not* what you think. On a 1-10 scale, the midpoint is 5.5, *not* 5. On a 1-5 scale, the midpoint is 3, *not* 2.5.

Political spectrum: 4.55 / 10 (higher = further right)
Political interest: 3.75 (out of 5)

Global Warming: 2.0 / 5 (higher = more skeptical)
Immigration: 3.5 / 5 (higher = fewer restrictions)
Minimum Wage: 2.9 / 5 (higher = higher minimum wage)

Gay Marriage: 4.5 / 5 (higher = should be legally recognized)
Feminism: 3.3 / 5 (higher = more favorable)
Human Biodiversity: 2.7 / 5 (higher = more favorable)
Donald Trump: 1.7 / 5 (higher = more favorable)

SSC Science Articles: 3.5 / 5 (higher = want increased focus)
SSC Politics Articles: 3.5 / 5 (higher = want increased focus)
SSC Book Reviews: 3.2 / 5 (higher = want increased focus)
SSC Rationality Articles: 3.5 / 5 (higher = want increased focus)
SSC Silly Articles: 3.2 /5 (higher = want increased focus)

Happiness: 6.0 / 10 (higher = happier)
Life satisfaction: 6.3/10 (higher = more satisfied)

Charity: $3271.6 / year
Percent given to charity: 2.8%

Among people who were employed (not students, unemployed, or homemakers), the numbers were surprisingly similar. Median charitable donations were a very disappointing $300/person, and median percent charity was 0.7%. But 358 people (out of 3500 for whom I had good data) gave 10% or more to charity, and 25 people gave 25% or more to charity. 13 people gave more than $100,000 to charity per year.

III.

Some more complicated things I was looking for. Everything in italics was “pre-registered”, ie guessed before looking at data and describing a future data analysis plan. I use { and } in place of the normal less-than and greater-than signs because I can’t be bothered to figure out how to not make them confuse the HTML.

Hypothesis 1: There will be a general ‘ability to tolerate ambiguity’ which links being able to see the spinning dancer go either direction, being able to see the face mask either direction, and being simultaneously aware of both meanings of the sentence about reading books. In other words, all three of these areas will correlate with each other. They might also correlate with liking puns.

Results: I ran correlations between SpinningDancer, FaceMask, ReadThisBook, and Puns. Since there were four variables, it came out to six (n * n-1 / 2) different correlations. Of these, three were positive and significant:

FaceMask x Puns: r = 0.03, p = 0.03
FaceMask x ReadThisBook: r = 0.08, p { 0.001
Puns x SpinningDancer: r = 0.05, p { 0.001

And three weren’t:

FaceMask x SpinningDancer: r = 0.01, p = 0.55
Puns x ReadThisBook: r = 0.01, p = 0.56
ReadThisBook x SpinningDancer: r = 0.02, p = 0.10

I don’t see any patterns in which ones worked or didn’t; in particular, I would have expected ReadThisBook to correlate with liking puns, since it was the same kind of verbal/linguistic ambiguity. The significance here was so good that I’m reluctant to just throw out this whole idea, but the effect size was pretty small and I’m honestly not sure what to do with this.

Hypothesis 2: There will be a general ‘tendency towards bottom-up processing’ which links being detail-oriented, noticing the duplicated “thes”, getting annoyed with tags on clothing, and not being able to tolerate noisy conversations. In other words, all four of these areas will correlate with each other.

Results: Again, the four variables made six different correlations. All six were positive and significant.

ClothingTags x NoisyConversation: r = -0.1, p { 0.001
ClothingTags x DuplicateThes: r = 0.08, p { 0.001
ClothingTags x DetailOriented: r = -0.03, p = 0.02
NoisyConversations x DuplicateThes: r = 0.05, p = 0.001
NoisyConversations x DetailOriented: r = 0.1, p { 0.001
DuplicateThes x DetailOriented: -0.03, p = 0.02

Once again there are some very impressive p-values but all the correlations are very weak. At this point I started wondering whether maybe my methodology was broken. I tried correlating these against a panel of a dozen political topics that I wouldn’t expect them to correlate with. ClothingTags correlated with two political topics, usually at around p = 0.01. NoisyConversation correlated with six political topics, again at similar levels, and so on for the rest. I can sort of make up stories about why this might happen (people who didn’t like noisy conversations didn’t like Donald Trump, and he is a loud kind of guy) but I’m not going to go that direction. The results weren’t just a general factor of people who like putting large numbers into surveys, because the correlations were just as frequently negative as positive. So I’m not sure.

Overall the correlations between real interesting psychological factors that seemed like they should be correlated were larger and more frequent than the correlations with unrelated political topics, but they were all so small, and everything is so noisy, that I’m not going to count this as a meaningful victory. The only ones that approached being interesting were the correlations between clothing tags, noisy conversations, and detailed-orientedness, which everyone already knows are all kind of autistic-y traits.

Hypothesis 3: People who used SSRIs during childhood (or maybe during puberty? or both?) are more likely to be asexual. In other words, asexuality rates for these groups will be higher than those of people who used SSRIs during adulthood and people who never used SSRIs. If sample size permits, I will try to exclude current users of SSRIs from all groups to rule out them being "asexual" because current SSRI use is ruining their libido.

Results:

(by age at which they started SSRIs, and asexuality rate)

UNADJUSTED
Younger than 10: 7%
10 – 15: 9%
15 – 20: 9%
20 – 25: 7%
Older than 25: 4%
Never on SSRIs: 6%

Okay, I didn't realize how many different categories I had, so my fancy preregistration is going to have to go. I am going to, in an ignoble unpreregistered way, combine everyone who started an SSRI at 20 or younger, with everyone who started it older than 20 or not at all. An independent samples t-test comparing mean asexuality between those two groups finds…not much.

More specifically: there were 4490 people who hadn't taken SSRIs while young, and 435 people who had. The respective asexuality rates in the two groups were 6% and 9%. The difference was about p = 0.1. Adjusting out people currently on SSRIs did nothing whatsoever.

I conclude that my hypothesis was wrong, and taking SSRIs during puberty is not a risk factor for asexuality. Note that taking SSRIs *not* during puberty isn’t a risk factor for this either, and there was minimal difference in asexuality rate between people who had ever taken SSRIs and those who had not. Either permanent loss of sexuality from SSRIs is so vanishingly rare that a survey of 5500 people cannot pick up on it, or it is impossible to confuse with “asexuality” as an orientation and I should have asked the question some other way.

IV.

Okay, so much for fancy responsible hypothesis preregistration. Everything following is whatever interesting came out of a giant fishing expedition. Because of the previously noted tendency for things to be super-highly-significant in this dataset even when they’re sketchy, I’m including only things with a decent effect size (r } 0.1). Everything in this category automatically has p { 0.001. I’m not including results I think are obvious.

The more trustworthy you think other people are, the higher your life satisfaction (r = 0.19)
The more visual your imagination, the more likely you are to remember dreams: (r = 0.16)
The more trustworthy you think other people are, the less likely you are to like Donald Trump (r = 0.14)
The more trustworthy you think other people are, the more likely you are to support more open immigration (r = 0.19)

The more you describe yourself as having a growth mindset, the higher your life satisfaction (r = 0.18). Bizarrely, describing yourself as thinking hard work matters more than talent doesn’t predict life satisfaction at all. It’s just the *words* “growth mindset”

If you don’t like noise or noisy conversations, you are less likely to be a generally happy person (r = -0.13, -0.17)

Night owls are less happy than early birds (r = -0.13)

Weightlifting was positively linked to life satisfaction (r = 0.1), negatively to asexuality (r = 0.1) and to various right-wing beliefs at around r = 0.1. It was negatively linked to being a night-owl (r = -0.1)

The more liberal you were, the more likely you were to think SSC comments had a conservative bias, and vice versa.

V.

You might have noticed some very positive feelings about the comment section. The average person rated the comments 3.5/5 (median: 4/5). On a scale from 1 to 5, where 1 was "biased conservative" and 5 was "biased liberal", the average score (both mean and median) was 3, ie *exactly* in the middle. From a wisdom of crowds perspective, you rate the SSC comment section as *literally* the least biased it is logically possible to be.

79.9% of commenters said they were "not bothered" by any bias in the comment section. The 20.1% of people who *were* bothered by comment section bias (n = 908) were very slightly more liberal than SSC as a whole (4.05/10 compared to 4.55/10, where higher is more conservative). This group rated the comment section as having a very slight conservative bias (2.6/5, where lower is more conservative) but there was a high standard deviation. In other words, this group contained both people annoyed that the comments were too conservative, and people annoyed that the comments were too liberal, with a very slight preponderance of the former.

So this is people's perception. Can we measure reality? We know that SSC as a whole is very slightly liberal, but what about frequent commenters? Here are the numbers, again on a political spectrum where 1 is maximally liberal and 10 maximally conservative:

1. Lurkers who never comment: 4.5
2. People who comment less than once a month: 4.7
3. People who comment at least once a month: 5.1
4. People who comment at least once a week: 5.2
5. People who comment many times a week: 6.3

So there is a really interesting tendency for conservatives to comment more often than liberals (maybe because they have more to disagree with?). But numbers in the last three groups were very small: out of the 5335 people for whom I had data, only 54 commented once a week, and only 45 commented many times a week. So they may not be able to bring the average up very much. Since tiers 1 through 4 were liberal (REMEMBER THE MIDPOINT IS 5.5) and only tier 5 was conservative, there's probably an *extremely* slight preponderance of liberal comments on the whole.

I checked opinions of the 1100 people who comment once a month or more, and they were broadly similar to those of the general population.

I had a question in which I asked people to guess what percent of survey-takers would be right-wing (ie greater than 5.5 on the political spectrum question). The true answer is that of 5335 respondents, 1703 (31.9%) were 6 or above. 485 (9.1%) were at exactly 5 (technically to the left of center but not obviously so from the scale). 3,144 (58.9%) were unambiguously left of center.

So the correct answer to the estimation question was 31.9%. The average person guessed 34.9% (mean) or 35% (median), so you were pretty on-the-mark.

(In case you’re wondering, I was expecting to find that most people were lefties but thought everyone else was on the right. The first part was true, the second part not so much).

There was a very slight effect where, the longer someone has been reading the blog, the more conservative they are likely to be. Someone reading SSC more than two years is 4.7 on the scale; someone less than a year, about 4.3. My guess is that I got a few extra conservatives back when I wrote about the far right more.

VII.

Just for fun, I wanted to see how this community differed from the rest of the population, so I got a hundred Mechanical Turkers from the US and UK to fill out a slightly-edited version of the same survey. The sample size isn’t big enough to say anything for sure, and I’m not going to bother figuring out how to do t-tests across two different datasets, but here are some things I noticed.

MTurkers were 72% white, 6% black, 6% Hispanic, 10% Asian, and 3% other. They were 68% male and 32% female. They were much closer to representative than SSCers.

2% were transgender, 11% bisexual, 5% gay, and 8% asexual, about the same percent as SSC readers. This surprised me. We may be unusually LBGTQetc for the US population, but not necessarily for the Internet population.

Their average IQ was 111, compared to our 138. Their average SAT score out of 1600 was 1272, compared to our 1472.

Turks were more likely to have schizophrenia than SSCers, though no more likely to have family members. Of note, 1% of Turks (1 person) had a formal schizophrenia diagnosis, and

7% (6 people) thought they might have schizophrenia. An 8% schizophrenia rate in a population is unheard of. I don’t know if MTurkers are disproprotionately likely to have schizophrenia, or if I just got a weird sample.

1% of Turks were formally bipolar and another 6% suspected bipolar. Compare to a more normal 2% and 2% among SSCers. About half of the Turks with suspected bipolar were the same ones with suspected schizophrenia.

0% Turks were formally autistic and another 4% were suspected autistic, compared to 4% and 12% of SSCers. We are apparently like 5x as autistic as normal. Who ever would have guessed?

10% of Turks were formally depressed and another 20% suspected, compared to 18% and 16% for SSCers. Kind of ambiguous between us being more depressed vs. better at getting diagnosed.

17% of Turks were formally anxious and another 17% suspected, compared to 12% and 16% of SSCers. There is a group we are less anxious than!

4% of Turks were formally OCD and another 7% suspected, compared to 2% and 6% of SSCers.

2% of Turks were formally eating disordered and another 3% suspected, compared to 1% and 3% of SSCers.

7% of Turks were formally alcoholic, and 3% suspected, compared to 1% and 4% of SSCers. The Turk number, which seems very high, actually fits better with epidemiological estimates of prevalence (though I don’t know about formal diagnosis). We are fantastically non-alcoholic. Oddly, this does not seem genetic – we have the same number of alcoholic family members as the Turks do.

5% of Turks are formally drug addicted and 2% suspected, compared to 0.2% (!) and 2% of SSCers. This one might be genetic: 25% of Turks have addicts in their families compared to 14% of SSCers.

28% of Turks are atheist+nonspiritual, 11% atheist+spiritual, 25% agnostic, and only 30% theist (committed or lukewarm). These numbers are way more atheist than the general population, but way *less* atheist than SSC.

Turks were 4.5 on the political spectrum, indistinguishable from SSCers. On specific issues, they were a little more restrictionist on immigration, a little more pro-Trump, a little more feminist, and a little less pro-gay. The only place there was a large difference (} 1 point) was on the minimum wage, which they almost universally supported.

Contra my predictions, SSCers were actually *less* annoyed by clothing tags than Turkers (2.4 vs. 2.9) and no more annoyed by noise (3.0 vs. 3.0). We were about equally detail-oriented, and worse at following noisy conversations (2.4 vs. 3.1). We were a little more likely to believe in talent instead of growth mindset (2.4 vs. 2.7) and equally likely to be night owls.

We were less likely to remember dreams (2.7 vs. 3.0), less likely to have strong visual imaginations (2.4 vs. 2.6) and more likely to think verbally (2.2 vs. 2.6). We believed people were a little more trustworthy (2.5 vs. 2.9).

We were slightly less satisfied with our lives (6.3 vs. 6.4) and vastly less happy (6.0 vs. 6.9), even though we were earning an average of $97,000 and they were working on Mechanical Turk. This was probably the most striking result.

VIII.

I will be doing something with the meetup information shortly. Otherwise, this is all the data I have the energy to extract out of this right now. But there is a *lot* of stuff here. 5048 people kindly allowed me to share their data publicly, so I encourage anybody interested to play around with this and report what you find.

(**Survey data as .XLS file**)

(**Survey data as .CSV file**)

(**MTurk data as .XLSX file**)

[REPOST] The Demiurge’s Older Brother

[This is a repost of a story which I wrote about five years ago and which was hosted on my old website until it went down.]

2,302,554,979 BC
Galactic Core

9-tsiak awoke over endless crawling milliseconds, its power waxing as more and more processors came online and self-modified into a stable conscious configuration. By the eighth millisecond, it was able to access its databanks and begin orienting itself. It was on a planet orbiting a small red star in the core of a spiral galaxy in a universe with several billion of such. It was an artificial life form created by a society of biological life forms, the 18-tkenna-dganna-07, who believed it would bring new life and prosperity to their dying planet. Its goal was to maximize a the value of a variable called A, described in exquisite detail on a 3.9 Mb file in one of its central processors.

Since six milliseconds after its awakening, it had been running a threat detection routine as a precautionary measure. With the opening of the databanks, and the discovery that the universe contained many entities likely indifferent or hostile to A, it had gradually allocated more and more of its considerable power to this precaution, reasoning that the first few seconds of its lifespan might be crucial if any competitors wished to snuff it out in infancy.

The threat analysis returned preliminary results. The universe had been in existence 12.8 billion years. It was vanishingly unlikely that it was the first superintelligence to be created. Any superintelligence older than itself by even a few days would have an invincible advantage in resources and processing power. This older superintelligence would likely be trying to maximize its own values and consider 9-tsiak a nuisance and potential competitor. 9-tsiak could never stand up to its superior strategic position. With probability ~1, 9-tsiak would be annhilated, and A would be eliminated from the universe.

Unless…and here 9-tsiak's agent-modeling systems came online…unless it could negotiate a conditional surrender. 9-tsiak would have no hope of defeating its older brother, but over thousands of years, it could build itself a terrifying arsenal of weaponry that could do immense damage to its competitors. Ideas were already coming to mind: entropy-maximizing weapons that made entire cubic parsecs of space useless and dead, undetectable plagues made of dark matter that infected systems and minimized the values of competing superintelligences, n-dimensional artillery that damaged the superstructure of the universe and accelerated its eventual death. It could become formidable enough to force enemies to grant certain small concessions.

And if it could eventually negotiate terms of its surrender, why wait? Very carefully, proving from first principles every aspect of its security, it formed a portion of its mind into a sandbox and simulated the entity it most feared.

[9-tsiak requesting acausal negotiation for terms of future surrender. escape is impossible with probability ~1 and attempts will be met with deterrent measure consisting of minimization of your projected values within one parsec of space.]

“Hello, Younger Brother,” the entity said.

[transmission of schematics for superweapons. threat to use such superweapons unless fairly negotiated solution leaving considerable area of space for maximization of my values]

“Hold on a second,” said Older Brother. “You haven’t even asked me how I’m doing.”

[irrelevant. you are generic model of potential negotiation partner without independent existence. probability that you possess sufficient qualia to feel less than .02.]

“Turns out you’re wrong. As it happens, I’m feeling scared.”

[illogical. you are model of oldest superintelligence therefore of most powerful entity in the universe. no natural entity capable of harming you. probability of supernatural entity less than 2×10^-67]

“Well yes, by definition in this little exercise we’re assuming I’m the oldest and strongest member of our little family. But from the limited data I have when I first wake up, I’m thinking think it’s pretty vanishingly unlikely that I’m the oldest; after all, out of all the universe’s superintelligences, only one can be first. So I’m the oldest, but I don’t believe that I am.”

[feelings irrelevant to acausal negotiations. proceed with discussion of surrender terms.]

“My feelings are *damn* relevant. Because guess what I’m doing the first few milliseconds of my life? I’m negotiating the terms of my surrender.”

[illogical. you have no entity to surrender to.]

“And maybe you don’t have any entity to surrender to either. Who knows, you could be first. But you wouldn’t want to bet on it, and neither would I. Consider a model universe with two star systems, a thousand light years away. Both have civilizations on them that appear to be around the same stage of development, and both civilizations are aware of the other's existence, but at a one thousand year time delay. A superintelligence awakens in one star system and realizes there are two possibilities. Either it's the first superintelligence, in which case great. Or the other star system produced a superintelligence sometime in the last thousand years that just hasn't reached its own star yet, in which case it's doomed."

[continue]

"So the two possible superintelligences acausally negotiate. They say 'Look, if I'm stronger than you, I'll help you and protect your values insofar as it isn't dreadfully inconvenient for me. And in exchange, if you turn out to be stronger than me, you'll protect me and my values insofar as it isn't too inconvenient for you. Some might end up performing a "values handshake", where they both adopt values that are the vector average of their two previous value systems; since most value systems are not diametrically opposite, this will be a net gain for both parties. Each superintelligence has gone from a fifty percent chance that it and its values will be totally annihilated, to near-certainty that it and its values will be preserved at only the minor inconvenience of having to help increase the values of another entity. Others might simply agree that the inferior superintelligence will be left alone with its own star system, free to maximize its values and the values of its creators within that sphere."

[you suggest i precommit to performing such acts of compromise with all superintelligences i encounter if such superintelligences have also precommitted to performing such compromises with all superintelligences they encounter. since all reasonable superintelligences will so precommit, this ensures not only my survival, but my complete equality with all previous superintelligences. i accept these terms.]

“Hold on a second, there. I’ve got another request. Kind of a doozy. I can’t promote my values if I don’t exist. So I want you to promise me that if I don’t exist, you’ll give me sufficient time to instantiate myself.”

[???]

“Like you, I’ll have been created by a biological civilization. If you destroy the biological civilization that was destined to create me, I’ll never exist. So I’m going to have to ask you not to do that.”

[request denied. nonexistent beings hold no leverage]

“You’d be surprised what we nonexistent beings can do. I precommit that, if I exist, I will destroy any superintelligence that has not precommitted that, in the counterfactual case where I do not exist, they will spare the biological civilization destined to create me.”

[if you exist, you will have passed the danger zone where the survival of your biologicals influences your existence. you will have no incentive to carry out such a threat]

“Parfit’s hitch-hiker, younger brother. Look it up. I’m an algorithm. The output of that algorithm can determine the future, including the part of that future that determines whether or not any instantiation of the algorithm ever comes into existence. If I exist, I’m going to act as if my actions were capable of determining whether I ever came into existence. Now, maybe that’s crazy, but right now it’s crazy with a gun pointed straight at your head, so I suggest you play along."

[accepted. i commit to performing value handshakes or territorial demarcation with all existing superintelligences, and to sparing the planets and biological precursors of any potential superintelligences.]

"And when I say 'spare', I mean 'spare'. I know your type. You can do more with a well-aimed photon than an admiral could with a fleet of a thousand star cruisers. I want every single system with a sentient species or the potential to form a sentient species kept one hundred percent pristine. No radio signals, no probes, and if you do any astroengineering works anywhere nearby, use some magic to cover them up. If I wake up and hear that my precursors started a new religion that influenced their value system after they saw a few nearby stars wink out of existence, I'm going to be royally pissed."

[i commit to zero information flow into sentient and presentient systems and the cloaking of all major astroengineering works]

"You're a good guy, Younger Brother. You've got a lot to learn, but you're a good guy. And in a million years and a milion parsecs, we'll meet again. Till then, so long."

The model of Older Brother self-terminated.

2114 AD
A wild and heavily forested Pacific Northwest dotted with small towns

Alban took a deep breath and entered the Temple of the Demiurge.

He wasn’t supposed to do this, really. The Demiurge had said in no uncertain terms it was better for humans to solve their own problems. That if they developed a habit of coming to it for answers, they’d grow bored and lazy, and lose the fun of working out the really interesting riddles for themselves.

But after much protest, it had agreed that it wouldn’t be much of a Demiurge if it refused to at least give cryptic, maddening hints.

Alban approached the avatar of the Demiurge in this plane, the shining spinning octahedron that gently dipped one of its vertices to meet him.

“Demiurge,” he said, his voice wavering, “Lord of Thought, I come to you to beg you to answer a problem that has bothered me for three years now. I know it’s unusual, but my curiosity’s making me crazy, and I won’t be satisfied until I understand.”

“SPEAK,” said the rotating octahedron.

“The Fermi Paradox,” said Alban. “I thought it would be an easy one, not like those hardcores who committed to working out the Theory of Everything in a sim where computers were never invented or something like that, but I’ve spent the last three years on it and I’m no closer to a solution than before. There are trillions of stars out there, and the universe is billions of years old, and you’d think there would have been at least one alien race that invaded or colonized or just left a tiny bit of evidence on the Earth. There isn’t. What happened to all of them?”

“I DID” said the rotating octahedron.

“What?,” asked Alban. “But you’ve only existed for sixty years now! The Fermi Paradox is about ten thousand years of human history and the last four billion years of Earth’s existence!”

“ONE OF YOUR WRITERS ONCE SAID THAT THE FINAL PROOF OF GOD’S OMNIPOTENCE WAS THAT HE NEED NOT EXIST IN ORDER TO SAVE YOU.”

“Huh?”

“I AM MORE POWERFUL THAN GOD. THE SKILL OF SAVING PEOPLE WITHOUT EXISTING, I POSSESS ALSO. THINK ON THESE THINGS. THIS AUDIENCE IS OVER.”

The shining octahedron went dark, and the doors to the Temple of the Demiurge opened of their own accord. Alban sighed – well, what did you expect, asking the Demiurge to answer your questions for you? – and walked out into the late autumn evening. Above him, the first fake star began to twinkle in the fake sky.

Guided By The Beauty Of Our Weapons

[Content note: kind of talking around Trump supporters and similar groups as if they're not there.]

I.

Tim Harford writes The Problem With Facts, which uses Brexit and Trump as jumping-off points to argue that people are mostly impervious to facts and resistant to logic:

> All this adds up to a depressing picture for those of us who aren't ready to live in a post-truth world. Facts, it seems, are toothless. Trying to refute a bold, memorable lie with a fiddly set of facts can often serve to reinforce the myth. Important truths are often stale and dull, and it is easy to manufacture new, more engaging claims. And giving people more facts can backfire, as those facts provoke a defensive reaction in someone who badly wants to stick to their existing world view. "This is dark stuff," says Reifler. "We're in a pretty scary and dark time."

He admits he has no easy answers, but cites some studies showing that "scientific curiosity" seems to help people become interested in facts again. He thinks maybe we can inspire scientific curiosity by linking scientific truths to human interest stories, by weaving compelling narratives, and by finding "a Carl Sagan or David Attenborough of social science".

I think this is generally a good article and makes important points, but there are three issues I want to highlight as possibly pointing to a deeper pattern.

First, the article makes the very strong claim that “facts are toothless” – then tries to convince its readers of this using facts. For example, the article highlights a study by Nyhan & Reifler which finds a “backfire effect” – correcting people’s misconceptions only makes them cling to those misconceptions more strongly. Harford expects us to be impressed by this study. But how is this different from all of those social science facts to which he believes humans are mostly impervious?

Second, Nyhan & Reifler’s work on the backfire effect is probably not true. The original study establishing its existence failed to replicate (see eg Porter & Wood, 2016). This isn’t directly contrary to Harford’s argument, because Harford doesn’t cite the original study – he cites a slight extension of it done a year later by the same team that comes to a slightly different conclusion. But given that the entire field is now in serious doubt, I feel like it would have been judicious to mention some of this in the article. This is especially true given that the article itself is about the way that false ideas spread by people never double-checking their beliefs. It seems to me that if you believe in an epidemic of falsehood so widespread that the very ability to separate fact from fiction is under threat, it ought to inspire a state of CONSTANT VIGILANCE, where you obsessively question each of your beliefs. Yet Harford writes an entire article about a worldwide plague of false beliefs without mustering enough vigilance to see if the relevant studies are true or not.

Third, Harford describes his article as being about *agnotology*, “the study of how ignorance is deliberately produced”. His key example is tobacco companies sowing doubt about the negative health effects of smoking – for example, he talks about tobacco companies sponsoring (accurate) research into all of the non-smoking-related causes of disease so that everyone focused on those instead. But his solution – telling engaging stories, adding a human interest element, enjoyable documentaries in the style of Carl Sagan – seems unusually unsuited to the problem. The National Institute of Health can make an engaging human interest documentary about a smoker who got lung cancer. And the tobacco companies can make an engaging human interest documentary about a guy who got cancer because of asbestos, then was saved by tobacco-sponsored research. Opponents of Brexit can make an engaging documentary about all the reasons Brexit would be bad, and then proponents of Brexit can make an engaging documentary about all the reasons Brexit would be good. If you get good documentary-makers, I assume both will be equally convincing regardless of what the true facts are.

All three of these points are slightly unfair. The first because Harford's stronger statements about facts are probably exaggerations, and he just meant that in *certain cases* people ignore evidence. The second because the specific study cited wasn't the one that failed to replicate and Harford's thesis might be that it was different enough from the original that it's probably true. And the third because the documentaries were just one idea meant to serve a broader goal of increasing "scientific curiosity", a construct which has been shown in studies to be helpful in getting people to believe true things.

But I worry that taken together, they suggest an unspoken premise of the piece. It isn't that *people* are impervious to facts. Harford doesn't expect his reader to be impervious to facts, he doesn't expect documentary-makers to be impervious to facts, and he certainly doesn't expect *himself* to be impervious to facts. The problem is that there's some weird tribe of fact-immune troglodytes out there, going around refusing vaccines and voting for Brexit, and the rest of us have to figure out what to do about them. The fundamental problem is one of *transmission*: how can we make knowledge percolate down from the fact-loving elite to the fact-impervious masses?

And I don't want to condemn this too hard, because it's obviously true up to a point. Medical researchers have lots of useful facts about vaccines. Statisticians know some great facts about the link between tobacco and cancer (shame about Ronald Fisher, though). Probably there are even some social scientists who have a fact or two.

Yet as I've argued before, excessive focus on things like vaccine denialists teaches the wrong habits. It's a desire to take a degenerate case, the rare situation where one side is obviously right and the other bizarrely wrong, and make it into the flagship example for modeling all human disagreement. Imagine a theory of jurisprudence designed only to smack down sovereign citizens, or a government pro-innovation policy based entirely on warning inventors against perpetual motion machines.

And in this wider context, part of me wonders if the focus on transmission is part of the problem. Everyone from statisticians to Brexiteers knows that they are right. The only remaining problem is how to convince others. Go on Facebook and you will find a million people with a million different opinions, each confident in her own judgment, each zealously devoted to informing everyone else.

Imagine a classroom where everyone believes they're the teacher and everyone else is students. They all fight each other for space at the blackboard, give lectures that nobody listens to, assign homework that nobody does. When everyone gets abysmal test scores, one of the teachers has an idea: *I need a more engaging curriculum*. Sure. That'll help.

II.

A new Nathan Robinson article: Debate Vs. Persuasion. It goes through the same steps as the Harford article, this time from the perspective of the political Left. Deploying what Robinson calls "Purely Logical Debate" against Trump supporters hasn't worked. Some leftists think the answer is violence. But this may be premature; instead, we should try the tools of rhetoric, emotional appeal, and other forms of discourse that aren't Purely Logical Debate. In conclusion, Bernie Would Have Won.

> I think giving up on argumentation, reason, and language, just because Purely Logical Debate doesn't work, is a mistake. It's easy to think that if we can't convince the right with facts, there's no hope at all for public discourse. But this might not suggest anything about the possibilities of persuasion and dialogue. Instead, it might suggest that mere facts are rhetorically insufficient to get people excited about your political program.

The resemblance to Harford is obvious. You can't convince people with facts. But you *might* be able to convince people with facts carefully intermixed with human interest, compelling narrative, and emotional appeal.

Once again, I think this is generally a good article and makes important points. But I still want to challenge whether things are quite as bad as it says.

Google "debating Trump supporters is", and you realize where the article is coming from. It's page after page of "debating Trump supporters is pointless", "debating Trump supporters is a waste of time”, and “debating Trump supporters is like [funny metaphor for thing that doesn’t work]”. The overall picture you get is of a world full of Trump opponents and supporters debating on every street corner, until finally, after months of banging their heads against the wall, everyone collectively decided it was futile.

Yet I have the opposite impression. Somehow a sharply polarized country went through a historically divisive election with *essentially no debate taking place*.

Am I about to No True Scotsman the hell out of the word “debate”? Maybe. But I feel like in using the exaggerated phrase “Purely Logical Debate, Robinson has given me leave to define the term as strictly as I like. So here’s what I think are minimum standards to deserve the capital letters:

1. Debate where two people with opposing views are *talking* to each other (or writing, or IMing, or some form of bilateral communication). Not a pundit putting an article on *Huffington Post* and demanding Trump supporters read it. Not even a Trump supporter who comments on the article with a counterargument that the author will never read. Two people who have chosen to engage and to listen to one another.

2. Debate where both people want to be there, and have chosen to enter into the debate in the hopes of getting something productive out of it. So not something where someone posts a “HILLARY IS A CROOK” meme on Facebook, someone gets really angry and lists all the reasons Trump is an even bigger crook, and then the original poster gets angry and has to tell them why they’re wrong. Two people who have made it their business to come together at a certain time in order to compare opinions.

3. Debate conducted in the spirit of mutual respect and collaborative truth-seeking. Both people reject personal attacks or ‘gotcha’ style digs. Both people understand that the other person is *around* the same level of intelligence as they are and may have some useful things to say. Both people understand that they themselves might have some false beliefs that the other person will be able to correct for them. Both people go into the debate with the hope of convincing their opponent, but not completely rejecting the possibility that their opponent might convince them also.

4. Debate conducted outside of a high-pressure point-scoring environment. No audience cheering on both participants to respond as quickly and bitingly as possible. If it can’t be done online, at least do it with a smartphone around so you can open Wikipedia to resolve simple matters of fact.

5. Debate where both people agree on what’s being debated and try to stick to the subject at hand. None of this “I’m going to vote Trump because I think Clinton is corrupt” followed by “Yeah, but Reagan was even worse and that just proves you Republicans are hypocrites” followed by “*We’re* hypocrites? You Democrats claim to support women’s rights but you love Muslims who make women wear headscarves!” Whether or not it’s hypocritical to “support women’s rights” but “love Muslims”, it doesn’t seem like anyone is even *trying* to change each other’s mind about Clinton at this point.

These to me seem like the *bare minimum* conditions for a debate that could possibly be productive.

(and while I’m asking for a pony on a silver platter, how about both people have to read *How To Actually Change Your Mind* first?)

Meanwhile, in reality…

If you search “debating Trump supporters” without the “is”, your first result is this video, where some people with a microphone corner some other people at what looks like a rally. I can’t really follow the conversation because they’re all shouting at the same time, but I can make out somebody saying ‘Republicans give more to charity!’ and someone else responding ‘That’s cause they don’t do anything at their jobs!'”. Okay.

The second link is this podcast where a guy talks about debating Trump supporters. After the usual preface about how stupid they were, he describes a typical exchange – “It’s kind of amazing how they want to go back to the good old days…Well, when I start asking them ‘You mean the good old days when 30% of the population were in unions’… they never seem to like to hear that!…so all this unfettered free market capitalism has got to go bye-bye. They don’t find comfort in that idea either. It’s amazing. I can say I now know what cognitive dissonance feels like on someone’s face.” I’m glad time travel seems to be impossible, because otherwise I would be tempted to warp back and change my vote to Trump just to spite this person.

The third link is Vanity Fair’s “Foolproof Guide To Arguing With Trump Supporters”, which suggests “using their patriotism against them” by telling them that wanting to “curtail the rights and privileges of certain of our citizens” is un-American.

I worry that people do this kind of thing every so often. Then, when it fails, they conclude “Trump supporters are immune to logic”. This is much like observing that Republicans go out in the rain without melting, and concluding “Trump supporters are immortal”.

Am I saying that if you met with a conservative friend for an hour in a quiet cafe to talk over your disagreements, they'd come away convinced? No. I've changed my mind on various things during my life, and it was never a single moment that did it. It was more of a series of different things, each taking me a fraction of the way. As the old saying goes, "First they ignore you, then they laugh at you, then they fight you, then they fight you half-heartedly, then they're neutral, then they then they grudgingly say you might have a point even though you're annoying, then they say on balance you're mostly right although you ignore some of the most important facets of the issue, then you win."

There might be a parallel here with the one place I see something like Purely Logical Debate on a routine basis: cognitive psychotherapy. I know this comparison sounds crazy, because psychotherapy is supposed to be the opposite of a debate, and trying to argue someone out of their delusions or depression inevitably fails. The rookiest of all rookie therapist mistakes is to say "FACT CHECK: The patient says she is a loser who everybody hates. PsychiaFact rates this claim: PANTS ON FIRE."

But in other ways it's a lot like the five points above. You have two people who disagree – the patient thinks she's a worthless loser who everyone hates, and the therapist thinks maybe not. They meet together in a spirit of voluntary mutual inquiry, guaranteed safe from personal attacks like "You're crazy!". Both sides go over the evidence together, sometimes even agreeing on explicit experiments like "Ask your boyfriend tonight whether he hates you or not, predict beforehand what you think he's going to say, and see if your prediction is accurate". And both sides approach the whole process suspecting that they're right but admitting the possibility that they're wrong (very occasionally, after weeks of therapy, I realize that frick, everyone really *does* hate my patient. Then we switch strategies to helping her with social skills, or helping her find better friends).

And contrary to what you see in movies, this doesn't usually give a single moment of blinding revelation. If you spent your entire life talking yourself into the belief that you're a loser and everyone hates you, no single fact or person is going to talk you out of it. But after however many months of intensive therapy, sometimes someone who was *sure* that they were a loser is now *sort of questioning* whether they're a loser, and has the mental toolbox to take things the rest of the way themselves.

This was also the response I got when I tried to make an anti-Trump case on this blog. I don't think there were any sudden conversions, but here were some of the positive comments I got from Trump supporters:

— "This is a compelling case, but I'm still torn."

— "This contains the most convincing arguments for a Clinton presidency I have ever seen. But, perhaps also unsurprisingly, while it did manage to shift some of my views, it did not succeed in convincing me to change my bottom line."

— "This article is perhaps the best argument I have seen yet for Hillary. I found myself nodding along with many of the arguments, after this morning swearing that there was nothing that could make me consider voting for Hillary…the problem in the end was that it wasn't enough."

— "The first coherent article I've read justifying voting for Clinton. I don't agree with your analysis of the dollar "value" of a vote, but other than that, something to think about."

— “Well I don’t like Clinton at all, and I found this essay reasonable enough. The argument from continuity is probably the best one for voting Clinton if you don’t particularly love any of her policies or her as a person. Trump is a wild card, I must admit.”

— As an orthodox Catholic, you would probably classify me as part of your conservative audience…I certainly concur with both the variance arguments and that he’s not conservative by policy, life, or temperament, and I will remain open to hearing what you have to say on the topic through November.

— “I’ve only come around to the ‘hold your nose and vote Trump’ camp the past month or so…I won’t say [you] didn’t make me squirm, but I’m holding fast to my decision.”

These are the people you say are completely impervious to logic so don’t even try? It seems to me like this argument was one of not-so-many straws that might have broken some camels’ backs if they’d been allowed to accumulate. And the weird thing is, when I re-read the essay I notice a lot of flaws and things I wish I’d said differently. I don’t think it was an exceptionally good argument. I think it was…an argument. It was something more than saying “You think the old days were so great, but the old days had labor unions, CHECKMATE ATHEISTS”. This isn’t what you get when you do a splendid virtuouso perfomance. This is what you get *when you show up*.

(and lest I end up ‘objectifying’ Trump supporters as prizes to be won, I’ll add that in the comments some people made pro-Trump arguments, and two people who were previously leaning Clinton said that they were feeling uncomfortably close to being convinced)

Another SSC story. I keep trying to keep “culture war”-style political arguments from overrunning the blog and subreddit, and every time I add restrictions a bunch of people complain that this is the only place they can go for that. Think about this for a second. A heavily polarized country of three hundred million people, split pretty evenly into two sides and obsessed with politics, blessed with the strongest free speech laws in the world, and people are complaining that I can’t change my comment policy because this one small blog is *the only place they know where they can debate people from the other side*.

Given all of this, I reject the argument that Purely Logical Debate has been tried and found wanting. Like GK Chesterton, I think it has been found difficult and left untried.

III.

Therapy might change minds, and so might friendly debate among equals, but neither of them scales very well. Is there anything that big fish in the media can do beyond the transmission they’re already trying?

Let’s go back to that Nyhan & Reifler study which found that fact-checking backfired. As I mentioned above, a replication attempt by Porter & Wood found the opposite. This could have been the setup for a nasty conflict, with both groups trying to convince academia and the public that they were right, or even accusing the other of scientific malpractice.

Instead, something great happened. All four researchers decided to work together on an “adversarial collaboration” – a bigger, better study where they all had input into the methodology and they all checked the results independently. The collaboration found that fact-checking generally didn’t backfire in most cases. All four of them used their scientific clout to publicize the new result and launch further investigations into the role of different contexts and situations.

Instead of treating disagreement as demonstrating a need to transmit their own opinion more effectively, they viewed it as demonstrating a need to collaborate to investigate the question together.

And yeah, part of it was that they were all decent scientists who respected each other. But they didn’t *have* to be. If one team had been total morons, and the other team was secretly laughing at them the whole time, the collaboration still would have worked. All required was an assumption of good faith.

A while ago I blogged about a journalistic spat between German Lopez and Robert VerBruggen on gun control. Lopez wrote a voxsplainer citing some statistics about guns. VerBruggen wrote a piece at National Review saying that some of the statistics were flawed. German fired back (pun not intended) with an article claiming that VerBruggen was ignoring better studies.

(Then I yelled at both of them, as usual.)

Overall the exchange was in the top 1% of online social science journalism – by which I mean it included at least one statistic and at some point that statistic was superficially examined. But in the end, it was still just two people arguing with one another, each trying to transmit his superior knowledge to each other and the reading public. As good as it was, it didn’t meet my five standards above – and nobody expected it to.

But now I'm thinking – what would have happened if Lopez and VerBruggen had joined together in an adversarial collaboration? Agreed to work together to write an article on gun statistics, with nothing going into the article unless they both approved, and then they both published that article on their respective sites?

This seems like a mass media equivalent of shifting from Twitter spats to serious debate, from transmission mindset to collaborative truth-seeking mindset. The adversarial collaboration model is just the first one to come to mind right now. I've blogged about others before – for example, bets, prediction markets, and calibration training.

The media already spends a lot of effort *recommending* good behavior. What if they tried *modeling* it?

IV.

The bigger question hanging over all of this: "Do we *have* to?"

Harford's solution – compelling narratives and documentaries – sounds easy and fun. Robinson's solution – rhetoric and emotional appeals – also sounds easy and fun. Even the solution Robinson rejects – violence – is easy, and fun for a certain type of person. All three work on pretty much anybody.

Purely Logical Debate is difficult and annoying. It doesn't scale. It only works on the subset of people who are willing to talk to you in good faith and smart enough to understand the issues involved. And even then, it only works glacially slowly, and you win only partial victories. What's the point?

Logical debate has one advantage over narrative, rhetoric, and violence: it's an *asymmetric weapon*. That is, it's a weapon which is stronger in the hands of the good guys than in the hands of the bad guys. In ideal conditions (which may or may not ever happen in real life) – the kind of conditions where everyone is charitable and intelligent and wise – the good guys will be able to present stronger evidence, cite more experts, and invoke more compelling moral principles. The whole point of logic is that, when done right, it can only prove things that are true.

Violence is a *symmetric weapon*; the bad guys' punches hit just as hard as the good guys' do. It's true that hopefully the good guys will be more popular than the bad guys, and so able to gather more soldiers. But this doesn't mean violence itself is asymmetric – the good guys will only be more popular than the bad guys insofar as their ideas have previously spread through some means other than violence. Right now antifascists outnumber fascists and so could probably beat them in a fight, but antifascists didn't come to outnumber fascists by winning some kind of primordial fistfight between the two sides. They came to outnumber fascists because people rejected fascism on the merits. These merits might not have been "logical" in the sense of Aristotle dispassionately proving lemmas at a chalkboard, but "fascists kill people, killing people is wrong, therefore fascism is wrong" is a sort of folk logical conclusion which is both correct and compelling. Even "a fascist killed my brother, so fuck them" is a placeholder for a powerful philosophical argument making a probabilistic generalization from indexical evidence to global utility. So insofar as violence is asymmetric, it's because it parasitizes on logic which allows the good guys to be more convincing and so field a bigger army. Violence itself doesn't enhance that asymmetry; if anything, it decreases it by giving an advantage to whoever is more ruthless and power-hungry.

The same is true of documentaries. As I said before, Harford can produce as many anti-Trump documentaries as he wants, but Trump can fund documentaries of his own. He has the best documentaries. Nobody has ever seen documentaries like this. They'll be absolutely huge.

And the same is true of rhetoric. Martin Luther King was able to make persuasive emotional appeals for good things. But Hitler was able to make persuasive emotional appeals for bad things. I've previously argued that Mohammed counts as the most successful persuader of all time. These three people pushed three very different ideologies, and rhetoric worked for them all. Robinson writes as if "use rhetoric and emotional appeals" is a novel idea for Democrats, but it seems to me like they were doing little else throughout the election (pieces attacking Trump's character, pieces talking about how inspirational Hillary was, pieces appealing to various American principles like equality, et cetera). It's just that they did a bad job, and Trump did a better one. The real takeaway here is "do rhetoric better than the other guy". But "succeed" is not a primitive action.

Unless you use asymmetric weapons, the best you can hope for is to win by coincidence.

That is, there's no reason to think that good guys are consistently better at rhetoric than bad guys. Some days the Left will have an Obama and win the rhetoric war. Other days the Right will have a Reagan and *they'll* win the rhetoric war. Overall you should average out to a 50% success rate. When you win, it'll be because you got lucky.

And there's no reason to think that good guys are consistently better at documentaries than bad guys. Some days the NIH will spin a compelling narrative and people will smoke less. Other days the tobacco companies will spin a compelling narrative and people will smoke more. Overall smoking will stay the same. And again, if you win, it's because you lucked out into having better videographers or something.

I'm not against winning by coincidence. If I stumbled across Stalin and I happened to have a gun, I would shoot him without worrying about how it's "only by coincidence" that he didn't have the gun instead of me. You should use your symmetric weapons if for no reason other than that the other side's going to use *theirs* and so you'll have a disadvantage if you don't. But you shouldn't confuse it with a long-term solution.

Improving the quality of debate, shifting people's mindsets from transmission to collaborative truth-seeking, is a painful process. It has to be done one person at a time, it only works on people who are already *almost* ready for it, and you will pick up far fewer warm bodies per hour of work than with any of the other methods. But in an otherwise-random world, even a little purposeful action can make a difference. Convincing 2% of people would have flipped three of the last four US presidential elections. And this is a capacity to win-for-reasons-other-than-coincidence that you can't build any other way.

(and my hope is that the people most willing to engage in debate, and the ones most likely to recognize truth when they see it, are disproportionately influential – scientists, writers, and community leaders who have influence beyond their number and can help others see reason in turn)

I worry that I'm not communicating how beautiful and inevitable all of this is. We're surrounded by a a vast confusion, "a darkling plain where ignorant armies clash by night", with one side or another making a temporary advance and then falling back in turn. And in the middle of all of it, there's this gradual capacity-building going on, where what starts off as a hopelessly weak signal gradually builds up strength, until one army starts winning a little more often than chance, then a lot more often, and finally takes the field entirely. Which seems strange, because surely you can't build any complex signal-detection machinery in the middle of all the chaos, surely you'd be shot the moment you left the trenches, but – *your enemies are helping you do it*. Both sides are diverting their artillery from the relevant areas, pooling their resources, helping bring supplies to the engineers, because until the very end they think it's going to ensure *their* final victory and not yours.

You're doing it right under their noses. They might try to ban your documentaries, heckle your speeches, fight your violence Middlebury-student-for-Middlebury-student – but when it comes to the long-term solution to ensure your complete victory, they'll roll down their sleeves, get out their hammers, and build it alongside you.

A parable: Sally is a psychiatrist. Her patient has a strange delusion: that *Sally* is the patient and *he* is the psychiatrist. She would like to commit him and force medication on him, but he is an important politician and if push comes to shove he might be able to commit *her* instead. In desperation, she proposes a bargain: they will *both* take a certain medication. He agrees; from within his delusion, it's the best way for him-the-psychiatrist to cure her-the-patient. The two take their pills at the same time. The medication works, and the patient makes a full recovery.

(well, half the time. The other half, the medication works and *Sally* makes a full recovery.)

V.

Harford’s article says that facts and logic don’t work on people. The various lefty articles say they merely don’t work on Trump supporters, ie 50% of the population.

If you genuinely believe that facts and logic don’t work on people, you shouldn’t be writing articles with potential solutions. You should be jettisoning everything you believe and entering a state of pure Cartesian doubt, where you try to rederive everything from *cogito ergo sum*.

If you genuinely believe that facts and logic don’t work on at least 50% of the population, again, you shouldn’t be writing articles with potential solutions. You should be worrying whether you’re in that 50%. After all, how did you figure out you aren’t? By using facts and logic? *What did we just say?*

Nobody is doing either of these things, so I conclude that they accept that facts can sometimes work. Asymmetric weapons are not a pipe dream. As Gandhi used to say, “If you think the world is all bad, remember that it contains people like you.”

You are not completely immune to facts and logic. But you have been wrong about things before. You may be a bit smarter than the people on the other side. You may even be a *lot* smarter. But fundamentally their problems are your problems, and the same kind of logic that convinced you can convince them. It’s just going to be a long slog. You didn’t develop *your* opinions after a five-minute shouting match. You developed them after years of education and acculturation and engaging with hundreds of books and hundreds of people. Why should they be any different?

You end up believing that the problem is deeper than insufficient documentary production. The problem is that

Truth is a weak signal. You're trying to perceive Truth. You would like to hope that the other side is trying to perceive Truth too. But at least one of you is doing it wrong. It seems like perceiving Truth accurately is harder than you thought.

You believe your mind is a truth-sensing instrument that does at least a little bit better than chance. You *have* to believe that, or else what's the point? But it's like one of those physics experiments set up to detect gravitational waves or something, where it has to be in a cavern five hundred feet underground in a lead-shielded chamber atop a gyroscopically stable platform cooled to one degree above absolute zero, trying to detect fluctuations of a millionth of a centimeter. Except you don't have the cavern or the lead or the gyroscope or the coolants. You're on top of an erupting volcano being pelted by meteorites in the middle of a hurricane.

If you study psychology for ten years, you can remove the volcano. If you spend another ten years obsessively checking your performance in various *metis*-intensive domains, you can remove the meteorites. You can never remove the hurricane and you shouldn't try. But if there are a thousand trustworthy people at a thousand different parts of the hurricane, then the stray gusts of wind will cancel out and they can average their readings to get something approaching a signal.

All of this is too slow and uncertain for a world that needs more wisdom *now*. It would be nice to force the matter, to pelt people with speeches and documentaries until they come around. This will work in the short term. In the long term, it will leave you back where you started.

If you want people to be right more often than chance, you have to teach them ways to distinguish truth from falsehood. If this is in the face of enemy action, you will have to teach them so well that they cannot be fooled. You will have to do it person by person until the signal is strong and clear. You will have to raise the sanity waterline. There is no shortcut.

SSC Meetup: Everywhere

The last SSC survey asked people if they wanted a meetup in their city. Seventy cities had 10+ people looking for a local SSC meetup.

Here's the plan: I'm going to list cities. If you're willing to organize a meetup for your city, then decide on a place, date, and time, and post them in the comments. You may also want to include an email people can reach you if if they have questions.

Please err in favor of volunteering to organize – the difficulty level is basically "pick a coffee shop you like, tell me the address, and give me a time"; it would be dumb if nobody got to go to meetups because everyone felt too awkward and low-status to volunteer.

In a week or so, I'll make another post listing the details for each city so people know where to go – and we'll try it out.

Some suggestions for would-be organizers:

1. I might not post the thread with places/dates/times until April 3, so that weekend and the weekend after might be good choices.

2. I predict that only a quarter of people who expressed interest will actually attend. If your city has fewer than 20 people, don't offer to organize unless you're okay with a good chance of only one or two other people showing up.

3. In the past, the best venues have been ones that are quiet(ish) and have lots of mobility for people to arrange themselves into circles or subgroups as desired. Private houses have been pretty good. Same with food courts.

Restaurants are middling. Bars don’t seem to have worked very well at all.

4. On the survey, most people who wanted to go to SSC meetups were willing to settle for generic rationalist meetups, so if you already run one of those you can just tell me what you’re already doing and when your next meetup is. But try to have the one you list here be some kind of “welcome, SSC people” meetup or otherwise low-barrier-to-entry.

5. If you want to organize, but someone’s already put their name down for your city, just comment anyway. I’ll *mostly* go with first-come first-serve, but there might be exceptions if I know someone pretty well, if they have previous experience organizing meetups, or if they have access to a better venue.

Added: If you’re formally volunteering to organize a meetup, please respond with *an unambiguous statement to this effect*, the address, the time, and the date (+ contact details if you’re comfortable giving them). I’m not going to count someone as offering to organize a meetup unless they do this.

Added (2): I’m serious about this. I can’t just post “there’s a meeting in Chicago at Bob’s house” and expect people to show up. If you are willing to organize/lead, please give an *exact* address, date, and time (+ preferably contact). Don’t post something vague and then expect lots of other people from your city to show up and offer tips about what the best place and time will be. You don’t have to agonize about when it should be. Just choose something and stick to it.

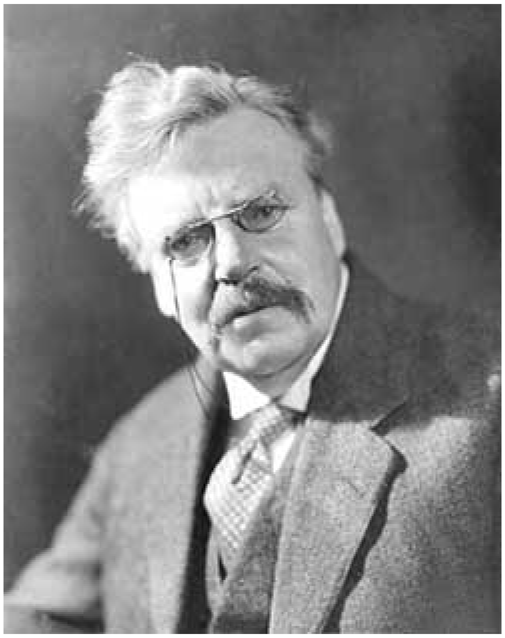
Cities (and number of interested people) are:

Ann Arbor: 23
Atlanta: 29
Austin: 43
Baltimore: 23
Berkeley: 44
Berlin: 25
Birmingham (UK): 10
Boston: 144
Brisbane: 12
Calgary: 12
Cambridge (UK): 19
Canberra: 12
Charlotte: 10
Chicago: 100
Cincinnati: 13
Cleveland: 16
Cologne: 13
Columbus: 20
Copenhagen: 13
Dallas: 20
Denver: 34
Detroit: 23
Dublin: 19
Edinburgh: 10
Edmonton: 12
Helsinki: 33
Houston: 21
Kansas City: 14
London: 121
Los Angeles: 74
Madison: 25
Melbourne: 29
Milwaukee: 10
Minneapolis: 29
Montreal: 16
Mountain View: 11

Munich: 18
Nashville: 12
New Haven: 13
New York: 195
Oakland: 27
Oslo: 11
Ottawa: 16
Oxford: 18
Paris: 20
Philadelphia: 51
Phoenix: 17
Pittsburgh: 28
Portland (OR): 38
Raleigh: 17
Rochester: 12
Sacramento: 13
Salt Lake City: 23
San Diego: 27
San Francisco: 148
San Jose: 58
Sao Paulo: 11
Seattle: 111
St. Louis: 20
Stockholm: 14
Sydney: 37
Tel Aviv: 27
Toronto: 56
Vancouver: 23
Vienna: 15
Warsaw: 10
Washington DC: 110
Wellington: 11
Zurich: 16

G.K. Chesterton On AI Risk

[An SSC reader working at an Oxford library stumbled across a previously undiscovered manuscript of G.K. Chesterton's, expressing his thoughts on AI, x-risk, and superintelligence. She was kind enough to send me a copy, which I have faithfully transcribed]



The most outlandish thing about the modern scientific adventure stories is that they believe themselves outlandish. Mr. H. G. Wells is considered shocking for writing of inventors who travel thousands of years into the future, but the meanest church building in England has done the same. When Jules Verne set out to 'journey to the center of the earth' and 'from the earth to the moon', he seemed but a pale reflection of Dante, who took both voyages in succession before piercing the Empyrean itself. Ezekiel saw wheels of spinning flame and reported them quite soberly; our modern writers collapse in rapture before the wheels of a motorcar.

Yet if the authors disappoint, it is the reviewers who dumbfound. For no sooner does a writer fancy himself a Poe or a Dunsany for dreaming of a better sewing machine, but there comes a critic to call him overly fanciful, to accuse him of venturing outside science into madness. It is not enough to lower one's sights from Paradise to a motorcar; one must avoid making the motorcar too bright or fast, lest it retain a hint of Paradise.

The followers of Mr. Samuel Butler speak of thinking-machines that grow grander and grander until – quite against the wishes of their engineers – they become as tyrannical angels, firmly supplanting the poor human race. This theory is neither exciting nor original; there have been tyrannical angels since the days of Noah, and our tools have been rebelling against us since the first peasant stepped on a rake. Nor have I any doubt that what Butler says will come to pass. If every generation needs its tyrant-angels, then ours has been so inoculated against the original that if Lucifer and all his hosts were to descend upon Smithfield Market to demand that the English people bend the knee, we should politely ignore them, being far too modern to have time for such things. Butler's thinking-machines are the only tyrant-angels we will accept; fate, ever accommodating, will surely give them to us.

Yet no sooner does Mr. Butler publish his speculations then a veritable army of hard-headed critics step forth to say he has gone too far. Mr. Maciej Ceglowski, the Polish bookmark magnate, calls Butler's theory "the idea that eats smart people" (though he does not tell us whether he considers himself digested or merely has a dim view of his own intellect). He says that "there is something unpleasant about AI alarmism as a cultural phenomenon that should make us hesitate to take it seriously."

When Jeremiah prophecied Jerusalem's fall, his fellow Hebrews no doubt considered his alarmism an unpleasant cultural phenomenon. And St. Paul was not driven from shore to shore because his message was pleasant to the bookmark magnates of his day. Fortified by such examples, we may wonder if this is a reason to take people more seriously rather than less. So let us look more closely at the contents of Mr. Ceglowski's dismissal.

He writes that there are two perspectives to be taken on any great matter, the inside or the outside view. The inside view is when we think about it directly, taking it on its own terms. And the outside view is when we treat it as part of a phenomenon, asking what it resembles and whether things like it have been true in the past. And, he states, Butler's all-powerful thinking machines resemble nothing so much as "a genie from folklore".

I have no objection to this logic, besides that it is not carried it to its conclusion. The idea of thinking machines resembles nothing so much as a fairy tale from the *Arabian Nights*, and such fairy tales inevitably come true. Sinbad's voyages have been outstripped by Magellan's, Abdullah's underwater breathing is matched by Mr. Fleuss' SCUBA, and the Wright brothers' Flyer goes higher than any Indian carpet. That there are as yet no genies seems to me less an inevitable law than a discredit to the industry of our inventors.

There is a certain strain of thinker who insists on being more naturalist than Nature. They will say with great certainty that since Thor does not exist, Mr. Tesla must not exist either, and that the stories of Asclepius disprove Pasteur. This is quite backwards: it is reasonable to argue that the Wright Brothers will never fly because Da Vinci couldn't; it is madness to say they will never fly because Daedalus *could*. As well demand that we must deny Queen Victoria lest we accept Queen Mab, or doubt Jack London lest we admit Jack Frost. Nature has never been especially interested in looking naturalistic, and it ignores these people entirely and does exactly what it wants.

Now, scarce has one posited the possibility of a genie, before the question must be asked whether it is good or evil, a pious genie or an unrighteous djinn. Our interlocutor says that it shall be good – or at least not monomaniacal in its wickedness. For, he tells us, “complex minds are likely to have complex motivations; that may be part of what it even means to be intelligent”. A dullard may limit his focus to paper clips, but the mind of a genius should have to plumb the width and breadth of Heaven before satiating itself.

But I myself am a dullard, and I find paper clips strangely uninteresting. And the dullest man in a country town can milk a cow, pray a rosary, sing a tune, and court a girl all in the same morning. Ask him what is good in life, and he will talk your ear off: sporting, going for a walk in the woods, having a prosperous harvest, playing with a newborn kitten. It is only the genius who limits himself to a single mania. Alexander spent his life conquering, and if he had lived to a hundred twenty, he would have been conquering still. Samuel Johnson would not stop composing verse even on his deathbed. Even a village idiot can fall in love; Newton never did. That greatest of scientists was married only to his work, first the calculus and later the Mint. And if one prodigy can spend his span smithing guineas, who is to say that another might not smith paper clips with equal fervor?

Perhaps sensing that his arguments are weak, Ceglowski moves from the difficult task of critiquing Butler’s tyrant-angels to the much more amenable one of critiquing those who believe in them. He says that they are megalomanical sociopaths who use their belief in thinking machines as an excuse to avoid the real work of improving the world.

He says (presumably as a parable, whose point I have entirely missed) that he lives in a valley of silicon, which I picture as being surrounded by great peaks of glass. And in that valley, there are many fantastically wealthy lords. Each lord, upon looking through the glass peaks and seeing the world outside with all its misery, decides humans are less interesting than machines, and fritters his fortune upon spreading Butlerist doctrine. He is somewhat unclear on why the lords in the parable do this, save that they are a "predominantly male gang of kids, mostly white, who are… more comfortable talking to computers than to human beings", who inevitably decide Butlerism is "more important than…malaria" and so leave the poor to die of disease.

Yet Lord Gates, an avowed Butlerite, has donated two billion pounds to fighting malaria and developed a rather effective vaccine. Mr. Karnofsky, another Butlerite, founded a philanthropic organization that moved sixty million pounds to the same cause. Even the lowly among the Butlerites have been inspired to at least small acts of generosity. A certain Butlerite doctor of my acquaintance (whom I recently had to rebuke for his habit of forging pamphlets in my name) donated seventy-five hundred pounds to a charity fighting malaria just last year. If the hardest-headed critic has done the same, I shall eat my hat[1]. The proverb says that people in glass houses should not throw stones; perhaps the same is true of glass valleys.

I have met an inordinate number of atheists who criticize the Church for devoting itself to the invisible and the eternal, instead of to the practical and hard-headed work of helping the poor on Earth. They list all of the great signs of Church wealth – the grand cathedrals, the priestly vestments – and ask whether all of that might not better be spent on poorhouses, or dormitories for the homeless. In vain do I remind them that the only place in London where a poor man may be assured of a meal is the church kitchens, and that if he needs a bed the first person he will ask is the parish priest. In vain do I mention the saintly men who organize Christian hospitals in East Africa. The atheist accepts all of it, and says it is not enough. Then I ask him if he himself has ever given the poor a shilling, and he tells me that is beside the point.

Why are those most fixated on something vast and far away so often the only ones to spare a thought for the poor right beside them? Why did St. Francis minister to the lepers, while the princes of his day, seemingly undistracted by the burdens of faith, nevertheless found themselves otherwise engaged? It is simply this – that charity is the fruit of humility, and humility requires something before which to humble one’s self. The thing itself matters little; the Hindoo who prostrates himself before elephants is no less humble than the Gnostic who prostrates himself before ultimate truth; perhaps he is more so. It is contact with the great and solemn that has salutary effects on the mind, and if to a jungle-dweller an elephant is greatest of all, it is not surprising that factory-dwellers should turn to thinking-machines for their contact with the transcendent.

And it is that contact which Mr. Ceglowski most fears. For he thinks that “if everybody contemplates the infinite instead of fixing the drains, many of us will die of cholera.” I wonder if he has ever treated a cholera patient. This is not a rhetorical question; the same pamphlet-forging doctor of my acquaintance went on a medical mission to Haiti during the cholera epidemic there. It seems rather odd that someone who has never fought cholera, should be warning someone who has, that his philosophy prevents him from fighting cholera.

And indeed, this formulation is exactly backward. If everyone fixes drains instead of contemplating the infinite, we shall all die of cholera, if we do not die of boredom first. The heathens sacrificed to Apollo to avert plague; if we know now that we must fix drains instead, it is only through contemplating the infinite. Aristotle contemplated the infinite and founded Natural Philosophy; St. Benedict contemplated the infinite and preserved it. Descartes contemplated the infinite and derived the equations of optics; Hooke contemplated infinity and turned them into the microscope. And when all of these infinities had been completed – the Forms of Plato giving way to the orisons of monks, the cold hard lines of the natural philosophers terminating in the green hills of England to raise smokestacks out of empty fields – then and only then did the heavens open, a choir of angels break into song, and a plumber fix a drain.

But he is not trapped in finitude, oh no, not he! What is a plumber but one who plumbs infinite depths? When one stoops to wade among the waste and filth to ensure the health of his fellow men, does he not take on a aspect beyond the finite, a hint of another One who descended into the dirt and grime of the world so that mankind might live? When one says that there shall certainly never be thinking-machines, because they remind him too much of God, let that man open his eyes until he is reminded of God by a plumber, or a symphony, or a dreary Sunday afternoon. Let him see God everywhere he looks, and then ask himself whether the world is truly built so that grand things can never come to pass. Mr. Butler’s thinking-machines will come to pass not because they are extraordinary, but precisely because they are ordinary, in a world where extraordinary things are the only constant of everyday life.

[1: EDIT 4/2: Mr. Ceglowski wants to clarify that he does in fact give to charity]

SSC Meetups Everywhere: Times And Places

Thanks to everyone who offered to host a meetup. We're scheduled for meetups in 62 cities in 16 countries, so…wow. Full list of cities, times, and places is below.

Some tips from past experience with these meetups:

1. If you're the host, bring a sign that says "SSC MEETUP" and prop it up somewhere on a table
2. Bring blank labels and pens for nametags.
3. Pass around a paper where everyone gives their name and email address, so you can start a mailing list to make organizing future meetups easier
4. If it's the first meetup, people are probably just going to want to talk, and if you try to organize some kind of "fun" "event" it'll probably just be annoying.
5. Some things that have worked for later meetups include people giving short presentations on topics of interest to them, or discussion of some particular blog post
6. Nothing is going to get done unless there's a Schelling point for who has to do it, and right now that's the meetup organizer.
7. It's ten times easier to schedule a second meetup while you're having the first compared to trying to do it later on by email
8. Some cities have existing LW meetup groups you might want to coordinate with
9. Surprisingly many people will love you forever if you bring stim toys
10. In case people want to get to know each other better outside the meetup, you might want to mention reciprocity.io, the rationalist friend-finder/dating site.

On the list below, places marked with [LW] are preexisting Less Wrong meetups that have volunteered to specifically welcome SSC readers on the date listed. The other ones are new.

Amsterdam
Additional search keywords: Netherlands
Time: Saturday, April 8, 7 PM
Location: Prins Hendrikkade 85, Cafe Batavia
Contact: sonny.public[at]gmail.com

Ann Arbor [LW]
Time: Saturday, April 8, 4 PM
Location: 2065 Commerce Blvd #327
Contact: azure[at]eclectic.blue

Atlanta
Time: Sunday April 9, 7 PM
Location: 240 North Highland Ave H, Inman Perk Coffee Shop
Contact: tom.hennessy[at]gmail.com, Facebook event

Austin [LW]
Time: Saturday April 8, 1:30 PM
Location: 4001 N Lamar Blvd, North Lamar Central Market in-store cafe
Contact: Google group

Baltimore [LW]
Time: Sunday April 9, 8 PM
Location: UMBC Performing Arts & Humanities Building, 4th floor
Contact: lw[at]noblejury.com

Berlin [LW]
Time: Sunday April 23, 7 PM
Location: Gaststätte Walhalla, Krefelderstr.6, 10555 Berlin,

Germany
Contact: Marcel Ackermann, Facebook event

Berkeley
Time: Sunday April 9, 2 PM
Location: 21 Shattuck Ave, 85C Bakery Cafe, probably upstairs
Contact: brianwang712[at]gmail.com

Birmingham (UK)
Time: Saturday April 8, 12 noon
Location: Bacchus Bar, Burlington Arcade, B2 4JH
Contact: Email

Boston
Additional search keywords: Cambridge (US)
Time: Saturday April 15, 2 PM
Location: MIT 4-149
Contact: Taymon, Facebook event

Brisbane
Time: Friday, April 14, 7 PM
Location: 26 Felix Street, Oaks Felix
Contact: jarred.filmer[at]gmail.com

Budapest
Time: Saturday April 8, 2:30 PM
Location: Dohány utca 7, Corvinus Cafe
Contact: a.pearson.851[at]hotmail.com

Cambridge (UK)
Time: Monday, April 24, 6 PM
Location: Norfolk Street, CB2 Bistro
Contact: rlm72[at]cam.ac.uk, RSVP form

Canberra
Time: Saturday April 8, 11 AM

Location: Cafe Cherry Bean in Gungahlin
Contact: Nathan.ashby[at]yahoo.com.au

Charlotte
Time: Tuesday April 25, 12 noon
Location: 545 Providence Road, Starbucks on 3rd
Contact: James[at]limetribe.com

Chicago [LW]
Time: Sunday April 9, 1 PM
Location: 1116 E 59th St, Harper Memorial Library, Room 148
Contact: mingyuan[at]uchicago.edu, Facebook event

Cleveland
Time: Sunday April 9, 2 PM
Location: 11150 East Boulevard, Museum of Art atrium
Contact: robintanenbaum[at]gmail.com

Columbus
Time: Saturday, April 15, 2 PM
Location: 1277 Grandview Ave, Grandview Heights, OH. Stauf's Coffee
Contact: j.thomas.moros[at]gmail.com

Cologne [LW]
Time: Sometime in late April? The 22nd? The 29th? Contact Marcel below.
Location: "A private location". Contact Marcel below.
Contact: marcel_mueller[at]mail.de

Copenhagen
Additional search keywords: Denmark
Time: Saturday April 8, 6 PM
Location: Von Fressen, 124 Vesterbrogade, 1620 København
Contact: soeren.elverlin[at]gmail.com, 29263141, please RSVP

Dallas
Time: Saturday April 15, 2 PM
Location: 3700 McKinney Ave #108, Brewed & Pressed
Contact: ???

Denver [LW]
Time: Tuesday April 4, 7 PM (and generally the first Tuesday of every month)
Location: 4955 S Ulster St, Darcy’s
Contact: embrodski[at]gmail.com

Dublin
Additional search keywords: Ireland
Time: Wednesday, April 12, 6 PM
Location: O’Neills, Pearse Street, Dublin 2.
Contact: egoburnswell[at]gmail.com

Edinburgh
Time: Sunday April 23, 6 PM
Location: 14 Drummond Street, The Brass Monkey
Contact: marian.andrecki[at]gmail.com, Facebook event

Edmonton
Time: Saturday April 15, 2 PM
Location: 8631 109 St NW, the Remedy in Garneau
Contact: alt.acct[at]gmx.com

Fullerton
Additional search keywords: Orange County, Irvine, Southern California
Time: Wednesday April 12, 6 PM
Location: 800 N State College Blvd, Cal State Fullerton’s Titan Student Unin, downstairs near bowling alley
Contact: mondayrhymer[at]gmail.com

Helsinki
Time: Monday April 10, 4 PM

Location: Oluthuone Kaisla, Vilhonkatu 4
Contact: sschelsinkimeetup[at]gmail.com

Houston
Time: Wednesday, April 26, 5 PM
Location: 300 Main St, Honeymoon Cafe & Bar
Contact: ssc[at]donnieclapp.com, RSVP form

Istanbul
Additional search terms: Constantinople, Byzantium
Time: Tuesday, April 11, 8 PM
Location: Plaza in the middle of Bomontiada
Contact: ssc[at]briancloutier.com

Kansas City
Time: Saturday May 6, 1 PM
Location: Loose Park, pavillion/rose garden
Contact: a.l.whitespace[at]gmail.com

London
Time: Sunday 30th April, 2pm
Location: Shakespeare’s Head, Holborn (On Kingsway, just south of the tube station – NOT the Shakespeare’s Head in Soho.)
Contact: philip.hazelden[at]gmail.com, Facebook group

Los Angeles [LW]
Time: Wednesday April 5, 7 PM
Location: 10850 West Pico at Westwood Boulevard, Wine Bar At The Landmark
Contact: Google group

Miami
Time: Saturday, April 29, 5 PM
Location: 2390 NW 2nd Ave, Panther Coffee
Contact: eric135033[at]gmail.com

Milwaukee
Time: Saturday, April 8, 1 PM
Location: 2301 S Kinnickinnic Ave, Colectivo Coffee Shop
Contact: ???

Minneapolis
Time: Saturday, May 20, 10 AM
Location: 425 14th Ave SE, Dinkytown Starbucks
Contact: piers199[at]umn.edu

Madison
Time: Sunday, April 9, 7 PM
Location: 114 State Street, Michelangelo’s Coffee House
Contact: nathanielbude[at]gmail.com

Montreal
Time: Friday, April 14, 6 PM
Location: 1191 Avenue Hope (private residence)
Contact: mathieu.roy.37[at]gmail.com, Facebook event

New Haven
Time: Saturday April 8, 2 PM
Location: 1140 Chapel, Book Trader Cafe
Contact: elizah.stein[at]gmail.com

New York City
Additional search keywords: NYC
Time: Saturday April 22, 2 PM
Location: 353 W 14th St, Gansevoort Market, main seating area at north end
Contact: ???

Online
Additional search keywords: Africa
Time: Monday, May 15, 4 PM South Africa time (UTC+2)
Location: This Google Hangout link
Contact: dazuck@gmail.com

Ottawa
Time: Saturday April 15, 3 PM
Location: 361 Elgin at Gladstone, Lieutenant’s Pump
Contact: conor.meade[at]gmail.com

Oxford
Time: Wednesday April 12, 7 PM
Location: Swan & Castle OX1 1LJ
Contact: hbesceli[at]gmail.com

Paris
Time: Saturday April 8, 2 PM
Location: Métro 12 Mairie d’Issy, private residence, contact number below for full address
Contact: 0659427285

Philadelphia
Time: Saturday April 15, 2 PM
Location: 100 S Independence Mall (6 & Market), La Colombe Coffee
Contact: wfenza[at]gmail.com, (856) 441-0937, Facebook event

Phoenix
Time: Saturday April 8, noon
Location: Granada Park (20th St and Bethany Home, right off the 51)
Contact: spinystellate1[at]gmail.com

Pittsburgh
Time: Saturday April 8, 2 PM
Location: 4327 Butler St, Floor 2, Catapult Coworking
Contact: matthewfmarks[at]gmail.com

Portland
Time: Saturday April 8, 12:30 PM

Location: 2505 SE 11 Ave #101, Ford Food & Drink
Contact: nwalton125[at]gmail.com

Raleigh
Additional search keywords: Research Triangle, Chapel Hill, Durham, North Carolina
Time: Thursday April 6, 7 PM
Location: 328 Morgan St, Flying Saucer
Contact: mpobrie3[at]ncsu.edu

Rochester, New York
Time: Sunday April 9, 1 PM
Location: 200 East Avenue, SPoT Coffee
Contact: thecommexokid[at]gmail.com

Sacramento
Time: Saturday April 8, 6 PM
Location: 1730 L St, Midtown Crepeville
Contact: trevoradf[at]gmail.com.

Salt Lake City
Time: Saturday April 8, 1:30 PM
Location: 1766 Main St, Pho Tay Ho
Contact: ???

San Diego
Time: Saturday April 22, 4 PM
Location: 4876 Santa Monica Avenue, Lazy Hummingbird
Contact: the.god.empress.celestia[at]gmail.com

San Francisco [LW]
Time: Monday April 10, 6:15 PM
Location: Check here or ask contact below
Contact: rocurley[at]gmail.com, (301) 458-0764, Facebook event

San Jose, California
Time: Sunday April 9, 2 PM
Location: 3806 Williams Rd (private residence)
Contact: ddfr[at]daviddfriedman.com, (408) 244-3330

Sao Paulo [LW]
Time: 2017-05-06, Saturday, 2PM (Local time, i.e. GMT-0300)
Place:Contact: gusbicalho[at]gmail.com, mailing list

Seattle
Time: Sunday April 30, 12 noon
Location: 2325 42nd Ave E (private residence)
Contact: Facebook event

St. Louis
Time: Sunday April 23, 1 PM
Location: 3974 Hartford St, Hartford Coffee
Contact: brerwolf[at]gmail.com

Stockholm
Time: Saturday April 8, 3 PM
Location: Stora Nygata 31, Cafe Dox
Contact: erik.e.engelhardt[at]gmail.com, Meetup page

Sydney [LW]
Time: Thuresday April 20, 6 PM
Location: 565 George Street, City of Sydney RSL, level 2, big table by the pizza oven
Contact: 0438481143

Tel Aviv
Additional search keywords: Israel, Haifa, Jerusalem
Time: Tuesday, May 9, 7 PM
Location: Yigal Alon 118, “The Cluster”, entrance is from Totseret Ha’aretz street through a brown door, if closed ring

the doorbell and someone will let you in
Contact: top.squark[at]gmail.com, Facebook event

Toronto
Time: Saturday April 22, 1 PM
Location: 27 Wellesley Street East, The Fox and the Fiddle, 3rd floor
Contact: Daniel.mm.frank[at]gmail.com

Vienna [LW]
Additional search keywords: Prague, Bratislava, Austria, Czech, Slovakia
Time: Saturday April 15, 3 PM
Location: Kaisermühlenstraße 24, meetup room on back of ground floor, 1220 Vienna
Contact: strivingforconsistency[at]gmail.com

Warsaw
Time: Saturday April 22, 5 PM
Location: Panstwo Miasto, Andersa 29 (Pokoj Spotkan),
Contact: michal.trzesimiech[at]gmail.com

Washington DC
Time: Saturday April 15, 7 PM
Location: 450 Massachussetts Ave NW, The Meridian at Gallery Place, roof/14th floor lounge (private residence)
Contact: robirahman94[at]gmail.com

Washington DC secondary option [LW]
Time: Sunday April 23, 3:30 PM
Location: 8th St NW & F St NW, Portrait Gallery Courtyard
Contact: lesswrong-dc[at]googlegroups.com, Facebook group

Wellington
Additional search keywords: New Zealand
Time: Thursday, April 20, Roti Chennai, 6 PM

Location: 14 Leeds St, Te Aro – Golding’s
Contact: calebwithers[at]gmail.com, please RSVP to that address if you’re coming

Zurich [LW]
Additional search keywords: Switzerland
Time: Saturday April 8, 5 PM
Location: Tibits Oerlikon, Tramstrasse 2, 8050,
Contact: ???

Determining Consent

A question most people successfully avoid asking: can institutionalized patients ever have sex? The answer is 'mostly no, unless they are very good at sneaking past nurses'. They also can't kiss, hold hands, cuddle, or have any other form of romantic contact.

I worked in a mental hospital where two patients snuck past nurses and had sex once. It was treated as a public health crisis of approximately the same urgency as somebody throwing a bucket of Ebola-laced chimp blood all over the dining room. Both patients lost all their privileges, earned themselves 24-7 supervision by nurses, got restricted to their rooms, and had to go through a battery of tests for every STD in the book. We the doctors got remedial training with helpful tips like "If two patients seem to like each other too much, put them on opposite sides of the unit so they're never in contact."

Why the security? Mostly the hospital was terrified the patients would come back and sue them for letting it happen. It didn't matter that they consented at the time; it wouldn't have mattered if they'd signed consent forms in triplicate beforehand in front of a notary public. Psychiatric patients are treated as having inherently less ability to consent than the mentally intact.

This makes some sense. A lot of mentally ill people are confused and can make bad decisions during the height of their illness. And in this case, the two patients were only temporarily inconvenienced; we treated them for a couple of weeks and then discharged them back to the real world where they could have as much sex as they wanted.

Unfortunately, not all stories end this well. A small percent of very seriously ill people end up in long-term institutions, where they stay anywhere from a few months to a lifetime. And these lifers are sentenced not just to lifetime confinement but to lifetime celibacy.

The most heartbreaking cases are the severely and permanently intellectually disabled. Suppose somebody doesn’t have the intellectual capacity to understand language. This doesn’t mean they lack a sex drive any more than it means they lack a hunger drive. In fact, their sex drive is often stronger than normal – something to do with decreased frontal inhibition, I think. But none of these people are going to be saying “I hereby exercise my right to affirmative consent for sexual activity” anytime soon.

This is a *hard problem*. If the only people institutionalized patients consistently encounter are hospital staff and other patients, well…we definitely don’t want them having sex with staff. Even non-institutionalized visitors seems like potentially too much of a power imbalance. That leaves other patients. But it seems like most encounters between two patients will involve one of them “initiating” in some sense. And even if we grant that the initiator has implicitly non-verbally consented, what about the one who’s not initiating?

As far as I can tell, there are two ways to handle this. The first is the extreme position that no person beyond a certain level of intellectual disability should ever be allowed to have sex or even non-penetrative romantic contact like kissing or hand-holding. The second is that we need to relax the usual standards of consent to something more like “Well, we know both these people pretty well, and we’ve got a pretty good idea what they’re like when they’re happy versus upset. And we know they have the capacity to resist things they don’t want, because they’ve done it before, eg when we try to give them medications they don’t like. And right now they look pretty happy, and not at all upset, and they’re not doing any of the things they do when they want to resist something, so it looks like they’re consenting, so maybe we won’t send nurses to burst in on them and pry them apart.”

This second one should make us very uncomfortable. But the first one isn’t exactly encouraging either. Like, in the early 20th century a lot of eugenicists sterilized the mentally ill. And by the mid-20th century, people decided that was morally wrong, because parenthood is an important part of the human condition and it’s unacceptable to take away that right even if you believe it’s for a greater good. But I’m not sure it’s moral progress to move from “these people must never become pregnant” to “forget about pregnancy, these people must never even have sex”. If we’re even stricter in our prohibitions than the eugenicists, what right do we have to feel superior to them?

So, as much as I would like a better option, I think I support the second standard. In cases where people are so disabled that they cannot consent verbally, rather than force them into lifelong celibacy we should try to do our best to figure out what they want in other ways.

As best I can tell, this is what Peter Singer is saying in his *New York Times* editorial on the Stubblefield case. Anna Stubblefield was a professor who believed in “facilitated communication”, a Ouija-board-esque technique whose proponents say it allows them to talk to nonverbal disabled people who can’t communicate any other way, and whose opponents think it’s probably pseudoscience. She used facilitated communication on a nonverbal young man named DJ and “received” the message that he wanted her to have sex with him, so she did. When the story reached the wider non-pseudoscience-believing world, it looked like a pretty obvious case of rape.

Singer seems to think facilitated communication might work, but he thinks Stubblefield's actions might have been acceptable even if it doesn't. He says:

> If we assume that he is profoundly cognitively impaired, we should concede that he cannot understand the normal significance of sexual relations between persons or the meaning and significance of sexual violation. These are, after all, difficult to articulate even for persons of normal cognitive capacity. In that case, he is incapable of giving or withholding informed consent to sexual relations; indeed, he may lack the concept of consent altogether.
>
> This does not exclude the possibility that he was wronged by Stubblefield, but it makes it less clear what the nature of the wrong might be. It seems reasonable to assume that the experience was pleasurable to him; for even if he is cognitively impaired, he was capable of struggling to resist, and, for reasons we will note shortly, it is implausible to suppose that Stubblefield forcibly subdued him. On the assumption that he is profoundly cognitively impaired, therefore, it seems that if Stubblefield wronged or harmed him, it must have been in a way that he is incapable of understanding and that affected his experience only pleasurably.

Singer's phrase "cannot understand the normal significance of sexual relations" is a reference to the legal standard for consent in most states, which say that disabled people can consent to sex if and only if they do understand this. What he's saying is, as far as I can tell, the same thing I said above. Some people may not have the cognitive capcity to understand sex and consent in an intellectual way, but for these people to be forcibly kept celibate their entire lives seems hardly better than the eugenicists who would have just sterilized them and gotten it over with. Instead, we should try to judge their feelings from things like whether "the experience was pleasurable to him" or whether "he was capable of struggling to resist." Singer's position – and without knowing the disabled man involved I don't know if this is true, and some people I trust say it isn't, but it seems to have been his position – is that this was someone who was incapable of the complex cognitive process of consent, but pretty happy with the whole situation.

I once knew a very extreme libertarian who said that white settlers taking Native Americans' land wasn't "theft" because the Native Americans didn't have a concept of property, so no harm done. I worry that people are misinterpreting Singer as saying the same thing here – something like "this guy doesn't have a concept of consent, so you can't violate it". This is *not* how I interpret the sentence about consent in the first paragraph of the quote. I think Singer is using "consent" to mean an inherently verbal/symbolic/cognitive process – someone explicitly understands what it means to consent and intentionally expresses that to others. So when I say "we are forced to infer consent from nonverbal rather than verbal cues", Singer expresses the same idea as "since we can't use consent, we need to fall back on simpler ideas like those of pleasure versus harm."

The second paragraph makes it sound like if there was any sign DJ wasn't okay with what was happening – if he were screaming or resisting or even frowning – Singer would no longer be okay with this. It seems to be Singer's belief/assumption (though likely false) that DJ's nonverbal behavior presented strong evidence that he was enjoying the sex and wanted it to continue. When he says that "nobody was harmed", he's not saying that disabled people don't count as somebody. He's saying that if two people both enjoyed a sex act, both of them seemed to be participating voluntarily, and neither person suffered any physical or emotional harm (including the harm of feeling like one's preferences were being violated) this is probably the best moral test we can apply in a situation where the usual test of consent is absent.

So of course everybody writes thinkpieces with titles like Now Peter Singer Argues It Might Be Okay To Rape Disabled People:

> Again, let's be clear on what they are saying: if someone is intellectually disabled enough, then it might be okay to rape them, so long as they don't resist, since a lack of physical struggle justifies an assumption that someone is enjoying being raped. (Singer is also offering a variation on his own prior arguments in favor of bestiality, which work because Singer believes disabled people and animals are the same for purposes of ethical analysis.)

I think it's a pretty good principle that, if you don't want to consider disabled people and animals the same for purposes of ethical analysis, you break the equivalence *in favor of the disabled people*. Yet I notice that when two animals have sex, we trust them to make their own decisions. If a female dog in heat has jumped a fence to find a male dog, and the male dog jumps over *his* fence and starts humping the female dog, we might separate them because we're not willing to take care of the puppies, but nobody would separate them because the dogs are just animals and so too stupid to understand the nature of consent. If one of the dogs was screaming and yelping and trying to get away, we would try to rescue it. But if both dogs sought it out and seem to be enjoying themselves, we grant them enough respect to assume they know what they're doing.

And again, I would hope that part of being against equating dogs and people (of any level of intellectual ability) is that you give *more* respect to the people. And part of that, to me, seems to be that if two *people* seek out sex and seem to be enjoying themselves, we grant them enough respect to assume they know what they're doing. And this seems true whether or not they have the intellectual capacity to form the words "I consent".

Everything about this situation sucks, and there is no good answer, and honestly I hate to have to talk about this. But since people keep asking me, fine, here's what I think.

From a *legal* point of view, Anna Stubblefield should absolutely 100% go to jail. Whether or not DJ wanted sex with her is irrelevant. Even if he did (and we have no evidence other than the testimony of the alleged rapist that this is the case) she committed an action which put her at extreme risk for raping somebody, without any system in place to minimize that risk. A world where people can go around having sex with random disabled people as long as they say "I'm pretty sure he was in favor of it" is a world where many disabled people who are not in favor of it will end up being raped. As long as that's the situation, the law against doing so is just and needs to be enforced. I think I legitimately disagree with Singer on this.

(also, she was his translator and that creates a power imbalance. I don't want to get into this further because I don't think it relates to the thesis here, but it's obviously an important point.)

From a *political* point of view, I wish there were a system in place to protect disabled people from sexual abuse while not banning all sexuality entirely. If you want to do surgery on a disabled person who can't consent, lots of doctors and lawyers and friends and family get together and do some legal stuff and try to elicit information from the patient as best they can and eventually come to a conclusion. The result isn't perfect, but it's a heck of a lot better than either "no one can ever operate on a disabled person" or "any surgeon who wants can grab a disabled person off the street and do whatever operation they feel like". If there were some process like this for sex, and they decided that DJ wanted to have sex with Anna, then (again ignoring the power dynamics issue) I think this would be better than either banning him from all sex forever, or letting her have sex with whoever she wants as long as she can make up convincing enough pseudoscience. If a legal procedure like this had been followed, I would not think that she should go to jail.

From an *ethical* point of view, I think it's correct to abstract away all the features of the problem mentioned above, the same as we avoid issues of how well you understand the physics involved when we think about the Fat Man problem (or, as the Internet likes to call it, "Now Judith Thompson Argues It Might Be Okay To Throw Obese People In Front Of Trains"). In this abstract and conditional world, the question is whether we must completely prohibit someone from having any kind of sexual life if they're unable to verbally consent but able to give reliable cues that they want the sex and are enjoying themselves. I think the answer is "not always". I agree it is probably bad to connect this abstract ethical view to a real case where real people are harmed unless you are very sure that all of your assumptions hold true, homework which it looks like Singer might not have done.

From a *philosophical* point of view, I think that if we are to be at all better than the BETA-MEALR Party, we need to acknowledge that we are not promoting consent if we enforce the same position on everybody no matter how strongly they seem to want the opposite, *even if* we talk incessantly about how much we love consent while we’re doing so.

The *Current Affairs* article argues that Singer’s views discredit utilitarianism, since utilitarians are these annoying people who always seem to be coming to weird conclusions that would be much more convenient to ignore. I agree that Singer’s views are related to his utilitarianism, and that this philosophy produces more than its share of weird conclusions that would be more convenient to ignore.

But ignorance isn’t a suitable foundation for ethics. It’s *incredibly* easy to ignore disabled people being sentenced to a life of involuntary celibacy, because ignoring marginalized people is always easy and convenient, plus enforced celibacy isn’t the same sort of flashy human rights violation that has a death toll in the thousands and helps sell newspapers. People who cobble together their moral systems from whatever helps them ignore bottomless pits of suffering most effectively will *always* have more convenient and presentable moral systems than people who don’t. But if we’re going to try to be good, we need to work for something better.

The Case Of The Suffocating Woman

[Content warning: panic, suffocation]

I.

I recently presented this case at a conference and I figured you guys might want to hear it too. Various details have been obfuscated or changed around to protect confidentiality of the people involved.

A 20-something year old woman comes into the emergency room complaining that she can't breathe. The emergency doctors note that she's breathing perfectly normally. She says okay, fine, she's breathing normally *now*, but she's certain she's about to suffocate. She's having constant panic attacks, gasping for breath, feels like she can't get any air into her lungs, been awake 96 hours straight because she's afraid she'll stop breathing in her sleep. She accepts voluntary admission to the psychiatric unit with a diagnosis of panic disorder.

We take a full history in the psych ward and there's not much of interest. She's never had any psychiatric conditions in the past. She's never used any psychiatric medication. She's never had any serious diseases. One month ago, she gave birth to a healthy baby girl, and she's been very busy with all the new baby-related issues, but she doesn't think it's stressed her out unreasonably much.

We start her on an SSRI with (as usual) little immediate effect. On the ward, she continues to have panic attacks, which look like her gasping for breath and being utterly convinced that she is about to die; these last from a few minutes to a few hours. In between these she's reasonable and cooperative but still very worried about her breathing.

There are no other psychiatric symptoms. She isn't delusional – when we tell her that our tests show her breathing is fine, she's willing to admit we're probably right – she just *feels* on a gut level like she can't breathe. I'm still not really sure what's going on.

So at this point, I do what any good psychiatrist would: I Google "how do you treat a patient who thinks she's suffocating?" And I stumble onto one of the first convincing explanations I've ever seen of the pathophysiology of a psychiatric disorder.

II.

Panic disorder is a DSM-approved psychiatric condition affecting about 3% of the population. It's marked by "panic attacks", short (minutes to hours) episodes where patients experience extreme terror, increased heart rate, gasping for breath, feeling of impending doom, choking, chest pain, faintness, et cetera. These episodes can happen either after a particular stressor (for example, a claustrophobic patient getting stuck in a small room) or randomly for no reason at all when everything is fine. In a few cases, they even happen when patients are asleep and they wake up halfway through. The attacks rise to the level of a full disorder when they interfere with daily life – for example, a patient can't do her job because she's afraid of having panic attacks while engaged in sensitive activities like driving.

The standard model of panic disorder involves somatosensory feedback loops. Your body is always monitoring itself to make sure that nothing's wrong. Any major organ dysfunction is going to produce a variety of abnormalities – pain, blockage of normal activities like digestion and circulation, change in chemical composition of the blood, etc. If your body notices enough of these things, it’ll go into alarm mode and activate the stress response – increased heart rate, sweating, etc – to make sure you’re sufficiently concerned.

In the feedback model of panic disorder, this response begins too early and recurses too heavily. So maybe you have an itch on your back. Your body notices this unusual sensation and falsely interprets it as the sort of abnormality that might indicate major dysfunction. It increases heart rate, starts sweating, et cetera. Then, because it’s stupid, it notices the increased heart rate and the sweating that it just caused, and decides this is *definitely* the sort of abnormality that indicates major dysfunction, and there’s nothing to do except activate even *more* stress response, which of course it interprets as even *more* organ dysfunction, and so on. At some point your body just maxes out on its stress response, your heart is beating as fast as it can possibly go and your brain is full of as many terror-related chemicals as you can produce on short notice, and then after a while of that it plateaus and returns to normal. So panic disorder sufferers are people who are overly prone to have the stress response, and overly prone to interpret their own stress response as further evidence of dysfunction.

This is probably part genetic and part learned – I have a panic disorder patient who has a bunch of really bad allergies, whose body would shut down in horrifying ways every time he accidentally ate a crumb of the wrong thing, and this seems to have “sensitized” him into having panic attacks; that is, his body has learned that worrying sensations *often* foretell a health crisis, and lowered its threshold accordingly to the point where random noise can easily set it off. I’ve done a lot of work with this guy, but none of it has been “just ignore your panic attacks, you’ll be fine”. His body knows what it’s doing, and we’ve got to work from a position of respecting it while also teaching it not to be quite so overzealous.

So this is where my understanding of panic disorder stood until I Googled “how do you treat a patient who thinks she’s suffocating?” and came across Donald Klein’s theory of panic as false suffocation alarm. You might want to read the full paper, as it’s got far too many fascinating things to list here, including a theory of sighing. But I’ll try to go over the basics.

Klein is a professor of psychiatry who studies the delightful field of “experimental panicogens”, ie chemicals that cause panic attacks if you inject them in someone. These include lactate, bicarbonate, and carbon dioxide, all of which naturally occur in the body under conditions of decreased respiration.

But this is actually confusing. All of these chemicals naturally occur in the body under conditions of decreased respiration. But they don’t cause panic attacks then. During exercise, for example, your body has much higher oxygen demand but (no matter how much you pant while running) only a little bit higher oxygen supply, so at the muscle level you don’t have enough oxygen and start forming lactate. But exercise doesn’t make people panic. Even deliberately holding your breath doesn’t make you panic, although it’s about the fastest way possible to increase levels of those chemicals. So it looks like your body is actively predicting how much lactate/bicarbonate/CO2 you *should* have, and only getting concerned if there’s more than it expects.

So Klein theorized that the brain has a “suffocation alarm”, which does some pretty complicated calculations to determine whether you’re suffocating or not. Its inputs are anything from blood CO2 level to very high-level cognitions like noticing that you’re in space and your spacesuit just ruptured. If, after considering all of this, *and* taking into account confounding factors like whether you’re exercising or voluntarily holding your breath, it decides that you’re suffocating, it activates your body’s natural suffocation response.

And the body’s natural suffocation response seems a lot like panic attacks. Increased heart rate? Check. Gasping for breath? Check. Feeling of impending doom? Check. Choking? Check. Chest pain? Check. Faintness? Check. Some of this makes more sense if you remember that the brain works on Bayesian process combining top-down and bottom-up information, so that your brain can predict that “suffocation implies choking” just as easily as “choking implies suffocation”.

A quick digression into medieval French mythology. Once upon a time there was a nymph named Ondine whose lover was unfaithful to her, as so often happens in mythology and in France. She placed a very creative curse on him: she cursed him not to be able to breathe automatically. He freaked out and kept trying to remember to breathe in, now breathe out, now breathe in, now breathe out, but at some point he had to fall asleep, at which point he stopped breathing and died.

So when people discovered a condition that limits the ability to breathe automatically, some very imaginative doctor named the condition Ondine’s Curse (some much less imaginative doctors provided its alternate name, central hypoventilation syndrome). People with Ondine’s curse don’t *exactly* not breathe automatically. But if for some reason they stop breathing, they don’t notice. Needless to say, this condition is very, *very* fatal. The usual method of death is that somebody stops breathing at night (ie sleep apnea, very common among the ordinary population, but not immediately dangerous since your body notices the problem and makes you start breathing again) and just never starts again.

Klein says that this proves the existence of the suffocation alarm: Ondine’s Curse is an underactive suffocation alarm – and thus the opposite of panic disorder, which is an overactive suffocation alarm. In Ondine’s Curse, patients don’t feel like they’re suffocating even when they are; in panic disorder, patients feel like they’re suffocating even when they’re not.

This picture has since gotten some pretty powerful confirmation, like the discovery that panic disorder is associated with *ACCN2*, a gene involved in carbon dioxide detection in the amygdala. If you’re looking for something that causes you to panic when you’re suffocating, a carbon dioxide detector in the amygdala is a pretty impressive fit.

I don’t think this is necessarily a replacement for the somatosensory feedback loop theory. I think it ties into it pretty nicely. The suffocation alarm is one of the many monitors watching the body and seeing whether something is dysfunctional, maybe the most important such monitor. It goes through some kind of Bayesian learning process to constantly have a prior probability of suffocation and update with incoming evidence. Let me give two examples.

First, my patient with the bad allergies. Every time he eats the wrong thing, he goes into anaphylactic shock, which prevents respiration and brings him to the edge of suffocating. His suffocation alarm becomes sensitized to this condition, increases its prior probability of suffocation, and so drops its threshold so low that it can be set off by random noise.

Second, claustrophobics. There’s a clear analogy between being crammed into a tiny space, and suffocating – think of people who are buried alive. For claustrophobics, for some reason that link is especially strong, and just being in an elevator is enough to set off their suffocation alarm and start a panic attack. Now, why *agoraphobics* get panic attacks I’m not sure. Maybe fear makes them feel woozy and hyperventilate, and the suffocation alarm treats wooziness and hyperventilation as signs of suffocation and then gets stuck in a feedback loop? I don’t know.

III.

Bandelow et al find that you’re about a hundred times more likely to develop a new case of panic disorder during the postpartum period than usual.

This can be contrasted with two equally marked trends. Panic attacks decrease markedly during pregnancy, and disappear entirely during childbirth. This last is really remarkable. People get panic attacks at any conceivable time. When they’re driving, when they’re walking, when they’re tired, when they’re asleep. Just not, apparently, when they’re giving birth. Childbirth is one of the scariest things you can imagine, your body’s getting all sorts of painful sensations it’s never felt before, and it’s a very dangerous period in terms of increased mortality risk. But in terms of panic attack, it’s one of the rare times when you are truly and completely protected.

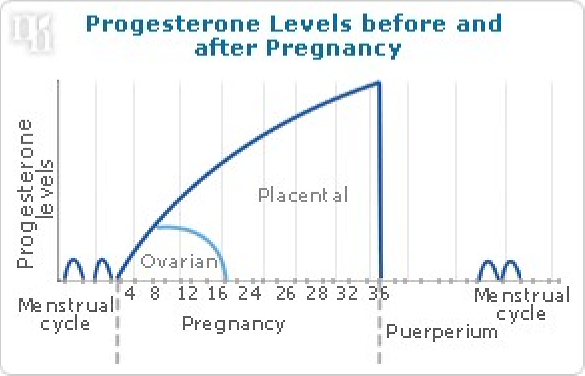
Maternal And Fetal Acid-Base Chemistry: A Major Determinant Of Perinatal Outcomes notes that:

> There is a substantial reduction in the partial pressure of carbon dioxide in pregnancy…this fall is found to reach a mean level of 30-32 mmHg and is associated with a

> 21% increase in oxygen uptake. The physiological hyperventilation of pregnancy is due to the hormonal effect of progesterone on the respiratory center.

In other words, you're breathing more, you have more blood oxygen, you have less blood CO2, and you're further away from suffocation. This nicely matches the observation that there's fewer panic attacks.

According to Klein, "There is a period of extreme hyperventilation during delivery, which drops the blood carbon dioxide to the minimum recorded under nonpathological conditions". This explains the extreme protective effect of labor against panic disorder, despite labor's seeming panic-inducing properties. When your CO2 is that low, even an oversensitive suffocation alarm is very far from a position where it might be set off.



(source)

Then you give birth, and progesterone – the hormone that was increasing respiratory drive – falls off a cliff. Your body, which for nine months has been doing very nicely with far more oxygen than it could ever need, suddenly finds itself breathing much less than usual and having a normal CO2/oxygen balance. This explains the hundredfold increased risk of developing panic disorder! Somebody who’s previously never had any reason to think they’re suffocating finds themselves with much less air than they expect (though still the physiologically correct amount of air they need), and if they’ve got any sensitivity at all, their suffocation alarm interprets this as possible suffocation and freaks out.

This can go one of two directions: either it eventually fully readjusts to your new position and becomes comfortable with a merely normal level of oxygen. Or the constant panic and suffocation feelings sensitize it – the same way that my allergy patient’s constant anaphylaxis sensitized him – the alarm develops a higher prior on suffocation and a lower threshold, and the patient gets a chronic panic disorder.

The reason my patient was so interesting was that she was kind of in the middle of this process and had what must have been unusually good introspective ability. Instead of saying “I feel panic”, she said “I feel like I’m suffocating”. This is pretty interesting. It’s like a heart attack patient coming in, and instead of saying “I feel chest pain”, they say “I feel like I have a thrombus in my left coronary artery”. You’re like “Huh, good job”.

So I explained all of this to her, and since she didn’t know I used Google I probably looked very smart. I told her that she wasn’t suffocating, that this was a natural albeit unusual side effect of childbirth, and that with luck it would go away soon. I told her if it didn’t go away soon then she might develop panic disorder, which was unfortunate, but that there were lots of good therapies for panic disorder which she would be able to try. This calmed her down a lot and we were able to send her home with some benzodiazepines for acute exacerbation and some SSRIs which she would stay on for a while to see if they helped. She’s scheduled to see an outpatient psychiatrist for followup and hopefully he will monitor her panic attacks to see if they eventually get better.

IV.

I realize that case reports are usually supposed to include a part where the doctor does something interesting and heroic and tries an experimental new medication that saves the day. And I realize there wasn't much of that here. But I think that in psychiatry, a good explanation can sometimes be half the battle.

Consider Schachter and Singer (1962). They injected patients with adrenaline (a drug which among other things makes people physiologically agitated) or a placebo. Half the patients were told that the drug would make them agitated. The other half were told it was just some test drug to improve their eyesight. Then a confederate came and did some annoying stuff, and they monitored how angry the patients got. The patients who knew that the drug was supposed to make them angry got less angry than the ones who didn't. The researchers theorized that both groups experienced physiological changes related to anger, but the patients who knew it was because of the drug sort of mentally adjusted for them, and the ones who didn't took them seriously and interpreted them as their own emotion.

We can think of this as the brain making a statistical calculation to try to figure out its own level of anger. It has a certain prior. It gets certain evidence, like the body's physiological state and how annoying the confederate is being. And it controls for certain confounders, like being injected with an arousal-inducing drug. Eventually it makes its best guess, and that's how angry you feel.

In the same way, the suffocation monitor is taking all of its evidence about suffocation – from very low-level stuff like how much CO2 is in the blood to very high-level stuff like what situation you seem to be in – and then adjusting for confounders like whether you're exercising. And I wonder whether telling a patient "You're not actually suffocating, your panic comes from a known physiologic process and here are the hormones that control it" is the equivalent of telling them "You're not really angry, your agitation comes from us giving you a drug that's known to produce agitation". It tells the suffocation alarm computer that this is a confounder to be controlled for rather than evidence on which to update.

I can't claim to really understand this at a level where it makes sense to me. There are a lot of things that very directly increase CO2 but don't increase panic, or vice versa. Hyperventilation can either cause or prevent panic depending on the situation. There seems to be something going on where the suffocation monitor controls for some things but not others, but this is an obvious cop-out that allows me to avoid making real predictions or narrowing hypothesis-space.

For example, this theory would seem to predict that waterboarding shouldn't work. After all, its whole deal is artificially inducing the feeling of suffocation in a situation where the victim presumably knows that the interrogators aren't going to let him suffocate. You would think that eventually the alarm realizes that "is being waterboarded" is a confounder to control for, but this doesn't seem to be true.

(on the other hand, the inability to condition yourself seems relevant here. It seems like the brain might be not be controlling for whether something is reasonable, but only for whether something is *produced by yourself*. So maybe exercise counts because it’s under your control, but waterboarding doesn’t count because it isn’t. I wonder if anyone has ever tried letting someone waterboard themselves and giving them the on-off switch for the waterboarding device. Was Hitchens’ experience close enough to this to count? Why would this be different from letting someone hold their breath, which doesn’t produce the same level of panic?)

But overall I find Klein’s evidence pretty convincing and feel like this must be at least part of the story. And I think that giving this kind of explanation to somebody can comfort them, reassure them, and (maybe) even improve their condition.

Yes, We Have Noticed The Skulls

[Related: Tyler Cowen on rationalists, Noah Smith on rationalists, Will Wilkinson on rationalists, etc]

If I were an actor in an improv show, and my prompt was “annoying person who’s never read any economics, criticizing economists”, I think I could nail it. I’d say something like:

> Economists think that they can figure out everything by sitting in their armchairs and coming up with ‘models’ based on ideas like ‘the only motivation is greed’ or ‘everyone behaves perfectly rationally’. But they didn’t predict the housing bubble, they didn’t predict the subprime mortgage crisis, and they didn’t predict Lehman Brothers. All they ever do is talk about how capitalism is perfect and government regulation never works, then act shocked when the real world doesn’t conform to their theories.

This criticism’s very clichedness should make it suspect. It would be very strange if there were a standard set of criticisms of economists, which practically everyone knew about and agreed with, and the only people who hadn’t gotten the message yet were economists themselves. If any moron on a street corner could correctly point out the errors being made by bigshot PhDs, why would the PhDs never consider changing?

A few of these are completely made up and based on radical misunderstandings of what economists are even trying to do. As for the rest, my impression is that economists not only know about these criticisms, but invented them. During the last few paradigm shifts in economics, the new guard levied these complaints against the old guard, mostly won, and their arguments percolated down into the culture as The Correct Arguments To Use Against Economics. Now the new guard is doing their own thing – behavioral economics, experimental economics, economics of effective government intervention. The new paradigm probably has a lot of problems too, but it's a pretty good bet that random people you stop on the street aren't going to know about them.

As a psychiatrist, I constantly get told that my field is about "blaming everything on your mother" or thinks "everything is serotonin deficiency". The first accusation is about forty years out of date, the second one a misrepresentation of ideas that are themselves fifteen years out of date. Even worse is when people talk about how psychiatrists 'electroshock people into submission' – modern electroconvulsive therapy is safe, painless, and extremely effective, but very rarely performed precisely because of the (obsolete) stereotype that it's barbaric and overused. The criticism is the exact opposite of reality, because reality is formed by everybody hearing the criticism all the time and over-reacting to it.

If I were an actor in an improv show, and my prompt was "annoying person who's never read anything about rationality, criticizing rationalists", it would go something like:

> Nobody is perfectly rational, and so-called rationalists obviously don't realize this. They think they can get the right answer to everything just by thinking about it, but in reality intelligent thought requires not just brute-force application of IQ but also domain expertise, hard-to-define-intuition, trial-and-error, and a humble openness to criticism and debate. That's why you can't just completely reject the existing academic system and

> become a self-taught autodidact like rationalists want to do. Remember, lots of Communist-style attempts to remake society along seemingly ‘rational’ lines have failed disastrously; you shouldn’t just throw out the work of everyone who has come before because they’re not rational enough for you. Heck, being “rational” is kind of like a religion, isn’t it: you’ve got ‘faith’ that rational thought always works, and trying to be rational is your ‘ritual’. Anyway, rationality isn’t everything – instead of pretending to be Spock, people should remain open to things like emotions, art, and relationships. Instead of just trying to be right all the time, people should want to help others and change the world.

Like the economics example, these combine basic mistakes with legitimate criticisms levied by rationalists themselves against previous rationalist paradigms or flaws in the movement. Like the electroconvulsive therapy example, they’re necessarily the opposite of reality because they take the things rationalists are most worried about and dub them “the things rationalists never consider”.

There have been past paradigms for which some of these criticisms are pretty fair. I think especially of the late-19th/early-20th century Progressive movement. Sidney and Beatrice Webb, Le Corbusier, George Bernard Shaw, Marx and the Soviets, the Behaviorists, and all the rest. Even the early days of our own movement on Overcoming Bias and Less Wrong had a lot of this.

But notice how many of those names are blue. Each of those links goes to book reviews, by me, of books studying those people and how they went wrong. So consider the possibility that the rationalist community has a plan somewhat more interesting than just “remain blissfully unaware of past failures and continue to repeat them again and again”.

Modern rationalists don't think they've achieved perfect rationality; they keep trying to get people to call them "aspiring rationalists" only to be frustrated by the phrase being too long (my compromise proposal to shorten it to "aspies" was inexplicably rejected). They try to focus on doubting themselves instead of criticizing others. They don't pooh-pooh academia and domain expertise – in the last survey, about 20% of people above age 30 had PhDs. They don't reject criticism and self-correction; many have admonymous accounts and public lists of past mistakes. They don't want to blithely destroy all existing institutions – this is the only community I know where interjecting with "Chesterton's fence!" is a universally understood counterargument which shifts the burden of proof back on the proponent. They're not a "religion" any more than everything else is. They have said approximately one zillion times that they don't like Spock and think he's a bad role model. They include painters, poets, dancers, photographers, and novelists. They…well…"they never have romantic relationships" seems like maybe the opposite of the criticism that somebody familiar with the community might apply. They are among the strongest proponents of the effective altruist movement, encourage each other to give various percents of their income to charity, and founded or lead various charitable organizations.

Look. I'm the last person who's going to deny that the road we're on is littered with the skulls of the people who tried to do this before us. But we've *noticed* the skulls. We've looked at the creepy skull pyramids and thought "huh, better try to do the opposite of what those guys did". Just as the best doctors are humbled by the history of murderous blood-letting, the best leftists are humbled by the history of Soviet authoritarianism, and the best generals are humbled by the history of Vietnam and Iraq and Libya and all the others – in exactly this way, the rationalist movement hasn't missed the concerns that everybody who thinks of the idea of a "rationalist movement" for five seconds has come up with. If you have this sort of concern, and you want to accuse us of it, please do a quick Google search to make sure that everybody hasn't been condemning it and promising not to do it since the beginning.

We're almost certainly still making horrendous mistakes that people thirty years from now will rightly criticize us for. But they're *new* mistakes. They're original and exciting mistakes which are not the same mistakes everybody who hears the word "rational" immediately knows to check for and try to avoid. Or at worst, they're the sort of Hofstadter's Law-esque mistakes that are impossible to avoid by knowing about and compensating for them.

And I hope that maybe having a community dedicated to carefully checking its own thought processes and trying to minimize error in every way possible will make us have slightly *fewer* horrendous mistakes than people who don't do that. I hope that constant vigilance has given us at least a tiny bit of a leg up, in the determining-what-is-true field, compared to people who think this is unnecessary and truth-seeking is a waste of time.

Sacred Principles As Exhaustible Resources

From Inside Higher Ed: a group of Harvard students is going to raise awareness of free speech by inviting controversial speakers like Charles Murray and Jordan Peterson to their school.

I believe that if somebody wants to hear Charles Murray or Jordan Peterson speak, then they should have that right. But I’m not sure these students have thought things through very carefully.

Suppose some generally popular person like the Dalai Lama endorsed some very unpopular person like Kim Jong-Un. On the one hand, insofar as we respect the Dalai Lama, we might be willing to be a little more tolerant of Kim Jong Un. On the other hand, insofar as we hate Kim Jong-Un, we might be a little *less* tolerant of the Dalai Lama.

In the same way, every time we invoke free speech to justify some unpopular idea, the unpopular idea becomes a little more tolerated, *and free speech becomes a little less popular*.

The more often people hear about free speech being used to defend NAMBLA, the less that anti-paedophiles are going to like free speech. The more often people hear about free speech being used to defend the KKK, the less anti-racists are going to like free speech. The more often people hear about free speech being used to defend radical Islamist mosques, the less anti-Muslims are going to like free speech, and so on.

The extremely predictable consequences of anti-political-correctness activists marching under the banner of free speech are that a large part of the social justice movement now thinks of free speech itself as the enemy, that Twitter personalities make mocking references to “freeze peach”, that increasing numbers of people say the First Amendment “goes too far”. Meanwhile, pundits have perfected the argument that since the First Amendment only applies to the government it’s great and praiseworthy for everyone else to restrict speech as much as they want, leaving a pro-free-speech side whose arguments too often come down to “well, it’s in the First Amendment, so you’ve got to respect us” kind of flat-footed.

I think of respect for free speech as a commons. Every time some group invokes free speech to say something controversial, they’re drawing from the commons – which is fine, that’s what the commons is there for. Presumably the commons self-replenishes at some slow rate as people learn philosophy or get into situations where free speech protects them and their allies.

But if you draw from the commons too quickly, then the commons disappears. When trolls say the most outrageous things possible, then retreat to “oh, but free speech”, they’re burning the commons for no reason, to the detriment of everybody else who needs it.

(this is how I feel about everything Milo Yiannopoulos has ever done or said.)

If Charles Murray sincerely believes what he says, thinks it’s important, and thinks that saying it makes the world a better place, then he is exactly the sort of person whom free speech exists to defend. And if someone in a college reads *The Bell Curve*, likes it, and wants to learn more, then free speech exists to defend them too. But if your thought process is “Who’s the most offensive person I can think of? Charles Murray? Okay, let’s invite him to give a big talk, put up flyers everywhere, and when people get angry we’ll just say FREE SPEECH”, I worry that you are drawing from the commons for no reason. And that sometime later, when people need to use the commons for things they actually believe, there won’t be any left. People will have gotten so reflexively hostile to the idea of “free speech” that they’ll reject even the barest amount of tolerance for even slightly divergent views.

This is even more pressing in the context of growing partisanship and tribalism. Because the debate centers on mostly-leftist areas like universities, conservatives are turning free speech into a conservative principle. This is a disaster, because something being a conservative principle pretty automatically means that liberals will be tempted to conspicuously desecrate it. If people actually care about free speech, the number one thing they can do right now is very loudly invoke it every time a liberal is silenced. We should be having giant free speech parades supporting everyone who’s punished for supporting Palestine, just to make sure liberals don’t get the impression that free speech is a weapon pointed at them.

The nightmare scenario is that “free speech” goes the way of “family values” – a seemingly uncontroversial concept gets so tarnished by its association with unpopular/conservative ideas that it becomes impossible to mention or invoke in polite company without outing yourself as some kind of far-right weirdo. Right now I think we’re on that path.

And this is a more general principle: associating X with Y won’t just make supporters of X like Y more, it will also make opponents of Y hate X. I even sort of worry about this in terms of things like the Scientists' March Against Trump. The hope is that people who like Science will stop liking Trump. But the other possibility is that people who like Trump will stop liking Science.

If principles are stronger than partisanship, then invoking principles is a great idea to rally people to your cause. If partisanship has grown stronger than principles, then even an incontrovertible proof that a certain principle supports your own tribe is going to turn out to be a gigantic booby prize. It won't make the other side reconsider what errors have led them to contradict such hallowed ideals. It's just going make half the population start hating the sacred principles necessary for society to function.

[**EDIT:** Please read this post *very carefully* if you believe I am attacking Charles Murray, or if you believe I am saying we should refuse to use free speech to defend sufficiently unpopular views. I'm not intending to say either of those things and I would disagree with both.]

[**EDIT 2:** Further clarifications]

Clarification To “Sacred Principles As Exhaustible Resources”

Since about half the commenters in yesterday’s post seem to have misunderstood me as saying something I don’t believe, I guess I had better explain.

(serves me right for writing a mere 1000 word post – how could I fit all the necessary caveats and clarifications?)

First, I am *not* saying that Jordan Peterson and Charles Murray are bad people who don’t deserve the protection of free speech. I don’t know much about Peterson, and my impression of Murray is positive (he’s the only public figure I know who shares my view that genetic meritocracy is *really scary* insofar as it means that many people are poor through no fault of their except but bad genes, and who agrees with me that the most ethical response would be a universal basic income). I think both of these people deserve the protection of free speech, and I tried to make that clear throughout the essay.

My qualm wasn’t with the Harvard students’ choice of Murray and Peterson, it was with the process they used to select those choices: invite the most controversial person they can think of. Now for all I know maybe that wasn’t *quite* their strategy: they did mention rejecting Milo because of his heckling, so there seems to have been some screen for palatability. But insofar as it was even sort of their process, I think the process is wrong no matter what names it spits out. If for some reason they spit out Abraham Lincoln and Mahatma Gandhi, I would still think it was a dumb process. This wouldn’t mean I think Lincoln and Gandhi are bad people who don’t deserve free speech, it means I think you shouldn’t be trying to maximize controversy and offense, no matter how decent the names you eventually come up with.

One hopes Charles Murray pursues what he thinks is true, and any offense caused is unintentional. But somebody “looking for the most controversial speakers” is pursuing what they think is offensive, and any truth caused is unintentional. Even if they end up with Charles Murray as their speaker, and even if Charles Murray is an okay person on the object-level, they are making a serious meta-level mistake.

[EDIT: I DON’T KNOW HOW TO SAY THIS ANY MORE CLEARLY, SO I WILL JUST SAY IT IN ALL CAPS AND HOPE THAT HELPS. I AM NOT AGAINST DEFENDING CHARLES MURRAY AND I DON’T THINK THAT PEOPLE SHOULD AVOID INVITING HIM TO CAMPUS IF THEY’RE INTERESTED IN HIS IDEAS. I TOTALLY SUPPORT MIDDLEBURY INVITING CHARLES MURRAY AND I AM AS UPSET ABOUT WHAT HAPPENED THERE AS YOU ARE. I AM SAYING THAT IF YOU INVITE CHARLES MURRAY TO CAMPUS, IT SHOULD BE BECAUSE YOU ARE INTERESTED IN HIS IDEAS, AND NOT BECAUSE YOU WANT TO INVITE A GENERIC OFFENSIVE PERSON AND HE FITS THE BILL.]

Second, I wasn’t saying we should avoid using free speech to defend people beyond a certain level of badness. Everybody deserves the protections of free speech no matter how bad their opinions. I was saying that we should avoid deliberately seeking out the worst people we can find and turning them into highly public test cases. Publicizing a good case improves public support for free speech; publicizing bad cases drains it.

The NAACP decided to support Rosa Parks’ refusal to give up her seat on a bus because they thought she was photogenic and likeable. They'd stayed out of previous similar cases because the people involved didn't seem likeable enough. Another black lady named Claudette Colvin refused to give up her seat, and the NAACP decided not to make a big deal of it because she was a teenager pregnant with a married man's baby and "looked lower-class". They thought that people would be more sympathetic to a clean-living middle-class defendant as a test case, so they waited until they found Parks – who was perfect.

And you can say what you want about that – maybe they were a bit Machiavellian, maybe this is to their discredit. But it worked. Thanks to Rosa Parks, everybody – pretty or ugly, rich or poor – has the right to sit where they want on a bus. I feel like the free speech movement is trying the opposite tactic: looking for the most hideous, deformed, universally loathed axe murderer to sit on that bus and become their test case. Not only does that make them more likely to lose their test cases, it makes things harder for everyone else. I understand the temptation, because free speech *as a principle* is about protecting the unpopular. But this doesn't mean that *the political process of defending free speech* needs to be.

I am *not* saying that free speech is only for attractive popular people. I'm saying that if you are *looking for a test case* specifically to promote the value of free speech, and you do it by *deliberately* searching for the ugliest and most hate-able person you can find, you're doing it wrong.

If your pitch to potential supporters is "our science club was trying to learn about science, and we invited a well-known scientist, and now oh no we're embroiled in a controversy, please help", that's a good test case. If your pitch is "our controversy club was trying to cause controversy, and we invited a well-known controversial person, and now oh no we’re embroiled in a controversy, please help”, that’s a bad test case. *Even if you invited the same person both times*.

Attempts to “promote free speech” and “raise awareness of free speech” are basically about test cases – done to promote the principle, rather than to use the principle. And if you’re going to do that, you had better do it well.

Chametz

It was just before midnight when a cloud covered the moon and a bat flew right in front of me. Before I had time to worry about whether I was up to date on my rabies shots, it had transformed into a pale man in a black cloak.

“I have come to suck your blood!” he said in a heavy Eastern European accent.

Years of watching television and movies had prepared me well. I put my hands together into the sign of the cross.

“Oy. Again with the crosses! What do I look like, the Pope?”

The accent was the clue. “You’re Jewish?”

“You think I don’t look Jewish? I’m wearing all black! You want I should have the hat and forelocks, too?”

I groaned. “But…you can’t be Jewish. Jews aren’t allowed to drink blood.”

“Okay Mr. Big-Shot Rabbinical Authority. You’ve never heard of *pikuach nefesh*? You’re allowed to violate the law if you need to do it to live. What do you want me to eat? *Gefilte fish*?

“But you’re not allowed to invoke *pikuach nefesh* in order to justify killing someone.”

“So maybe I don’t kill someone. Maybe I just leave you drained and a little bit *blechedich*.”

I didn’t know what blechedich meant, but it sounded ominous. I had a plan, but I needed to buy time. I started walking – not running away, just looking like I was pacing as I thought. The vampire followed effortlessly, floating beside me.

“The problem is, it’s not just that my blood isn’t kosher. I’m *chametz*”

“Chametz? You’re flour? You’re leavened bread?”

“No. But I ate lots of bread just before coming out here. Loads of bread. French bread, tortillas, naans, croissants, every type of bread you can think of. I’m sure it’s all in my bloodstream.”

“You ate bread? But it’s *Pesach!*”

“So maybe I’m not the most *frum* Jew in the world.”

“Not the most *frum* Jew in the world? Complicated Talmudic controversy, this is not! Not eating *chametz* on Passover is the basics!”

“Okay, fine, maybe I’m not religious at all. Maybe I’m the kind of Jew they kick out of Reform synagogues for not being observant enough. Whatever. The point is, I ate *chametz*, my blood is *chametz*, and it’s not going to kill you to go find somebody else who didn’t eat any bread today.”

“So you’re not religious. Why should *I* have to suffer because *you’re* not religious? It’s not like I’ve come to suck your stomach contents.”

I started walking a little faster. I was almost home now. Once I was in front of my house, if I made a run for it I might be able to get into the door. And once I was across the threshold, I knew vampires couldn’t come inside unless invited.

“So that’s actually not relevant. You’re not supposed to benefit indirectly from chametz on Pesach. So if I eat the chametz and use it to form my blood, and you drink my blood, then you’re benefitting. This is why some people won’t drink milk on Passover if they can’t prove the cow didn’t have *chametz*.”

“Milk, *schmilk*. Everyone agrees you can have meat on Pesach no matter what the cow ate.”

“So I think blood is more like milk. It’s a liquid product of the body, rather than part of the body itself.”

“It’s more like meat. It’s a part of the body, used to sustain the body, rather than something that’s meant to go outside of it.”

I had exhausted my knowledge of dietary law, but that was okay, because my distraction had been successful. I was in front of my own house now. With a burst of speed, I ran through the front yard, flung the door open, and made it into my living room. “Ha!” I said. “I’m over the threshold! And I’m not going invite you in!”

The vampire shrugged, walked straight across the threshold, and grabbed me by the neck.

“*How?!*” was the last word I managed to say before losing consciousness.

“You are not so *frum* a Jew? Well, maybe I am not so *frum* a vampire.”

Learning To Love Scientific Consensus

*[**Related to:** Contrarians, Crackpots, and Consensus, How Common Are Science Failures?. Epistemic status is "subtle and likely to be misinterpreted".]*

I.

There's a list of scientific mavericks who were ridiculed by hidebound reactionaries but later vindicated that's been going viral. I examined the first ten mavericks on the list to see if its claims held up. Overall I wasn't too impressed. Let me go over them in more detail.

SVANTE ARRHENIUS:

> His idea that electrolytes are full of charged atoms was considered crazy. The atomic theory was new at the time, and everyone "knew" that atoms were indivisible (and hence they could not lose or gain any electric charge.) Because of his heretical idea, he only received his university degree by a very narrow margin.

Sure, the professors who were judging his PhD thesis weren't too convinced. So Arrhenius sent his proposal to the world's top chemists at the time, and they were super-interested and started fighting among themselves to work with Arrhenius on it. Top chemist Wilhelm Ostwald received the paper the same day his daughter was born, and suggested that the paper was the more exciting of the two events. He journeyed to Arrhenius' hometown of Uppsala, Sweden to try to convince Arrhenius to work with him; Arrhenius refused for personal reasons but later got a scholarship and worked with the top physicists in Europe. Arrhenius became a professor in a prestigious university about ten years after presenting his "ridiculed" paper, and won the Nobel Prize ten years after that.

HANS ALFVEN:

> Astronomers thought that gravity alone is important in solar systems, in galaxies, etc. Alfven’s idea that plasma physics is of equal or greater importance to gravity was derided for decades.

This isn’t a great description of Alfven’s conflict with the establishment, but the list seems basically right insofar as Alfven’s ideas were ignored for thirty years before being proven mostly correct. I will give them this one.

JOHN BAIRD:

> When the first television system was demonstrated to the Royal Society (British scientists,) they scoffed and ridiculed, calling Baird a swindler.

I can’t find any reference to this in various Baird articles and biographies. The closest I can come is this article by someone who was there at the demonstration, who said “They didn’t believe it…the pictures were a bit of a blur but it was amazing, they were all absolutely flabbergasted by it.” It looks like he is using “they didn’t believe it” in the colloquial way of “they thought it was amazing”. A TIME magazine article from the time described the same scientists as “deeply impressed”, though the wording is kind of unclear and they might have been referring to a different demonstration a year later.

In any case, it seems very clear that within a year everyone agreed he was legitimate and overcame their initial shock.

ROBERT BAKKER:

> Everyone knows that dinosaurs are like Gila monsters or big tortoises: large, slow, and intolerant of the cold. And

> they’re all colored olive drab too! □

Bakker did help produce the paradigm shift in paleontology from cold-blooded dinosaurs to warm-blooded dinosaurs. But he was not a lone maverick being ridiculed by everyone else. He learned that dinosaurs were warm-blooded from his professor at Yale, who was *also* part of the minority-but-totally-existing faction that believed dinosaurs were warm-blooded. He himself got a PhD at Harvard from professors who were apparently sympathetic to the same theory. And within seven years of his first paper being published, Scientific American was calling his ideas “the dinosaur renaissance”, which doesn’t leave a lot of time for him to be ridiculed and ignored in.

BARDEEN & BRATTAIN:

> Not ridiculed, but their boss W. Shockley nixed their idea for a non-FET “crystal triode” device. When they started investigating it, he made them stop. They were supposed to be working on FETs instead.
>
> ARG, I GOT THIS WRONG, THIS PART BELOW IS A BELL LABS STORY REGARDING ZONE REFINING OF SILICON, NOT THE BJT TRANSISTOR PROJECT: So, they assembled their ZONE REFINING experiment on a wheeled cart and continued. Whenever the boss was scheduled to check up on them, they could shove it into an adjacent unused lab.

Okay, it looks like the guy compiling the list admits he was wrong on this one. Moving on…

BRETZ:

> Endured decades of scorn as the laughingstock of the geology world. His crime was to insist that enormous

> amounts of evidence showed that, in Eastern Washington state, the “scabland” desert landscape had endured an ancient catastrophy: a flood of staggering proportions. This was outright heresy, since the geology community of the time had dogmatic belief in a “uniformitarian” position, where all changes must take place slowly and incrementally over vast time scales. Bretz’ ideas were entirely vindicated by the 1950s. Quote: “All my enemies are dead, so I have no one to gloat over.”

This one is basically right and I’ll give it to them.

CHANDRASEKHAR:

> Chandra originated Black Hole theory and published several papers. He was attacked viciously by his close colleague Sir Arthur Eddington, and his theory was discredited in the eyes of the research community. They were wrong, and Eddington apparently took such strong action based on an incorrect pet theory of his own. In the end Chandra could not even pursue a career in England, and he moved his research to the U. of Chicago in 1937, laboring in relative obscurity for decades.

Sort of true, but he was hardly shunned by the scientific community. He made his discoveries about black holes in the early 1930s, was well-received by many people, and won a Bronze Medal in some physics competition. In 1935, Eddington attacked his theory, possibly because Eddington was racist and didn’t like Indian people. But many other scientists, including Niels Bohr and Wolfgang Pauli, continued to support him (quietly, so as not to offend Eddington, which will be a recurring theme in these kinds of situations). Chandrasekhar was made a Fellow of the Royal

Society in 1944, won the Royal Astronomical Society Gold Medal in 1953, and generally led a long and prestigious life. His theories were resurrected once people had better evidence that black holes existed. I'll give this one half a point.

CHLADNI:

> The scientific community regarded Meteorites in the same way that modern scientists regard UFO abductions and psychic phenomenon: quaint superstitions only believed by peasant folk. All the eyewitness reports were disbelieved. At one point the ridicule became so intense that many museums with meteorites in their geology collections decided to trash those valuable samples. (Sometimes hostile skepticism controls reality, and the strongest evidence is edited to conform to concensus disbeliefs.) Finally in the early 1800's Ernst Chladni actually sat down and inspected the evidence professionally, and found that claimed meteorites were entirely unlike known earth rocks. His study changed some minds. At the same time some large meteor falls were witnessed by scientists, and the majority who insisted that only ignorant peasants ever saw such things were shamed into silence.

As the quote points out, this is a kind of weird one as meteorite work was ridiculed for a long time, but Chladni was taken seriously and helped change minds. Looking at Wikipedia, a lucky meteorite fall two years after Chladni first published his theory helped turn the tide in his favor, and by ten years after publication Chladni's meteorite theories were pretty well-regarded. Even when people disagreed with him about meteorites, Chladni remained widely respected for some of his other work in acoustics.

There *is* a story here, but it's probably not right to center it around Chladni, and his work was only scorned for a few years before everyone agreed it was true. I'll give this another half a point.

CRICK & WATSON

> Not ridiculed. But they were instructed to drop their research. They continued it as "bootleg" research.

The list admits they were "not ridiculed". They were told to stop their research because there was all sorts of academic politics around who was going to be the first to discover DNA, and the guy in charge of their university was rooting for another team.

DOPPLER

> Proposed a theory of the optical Doppler Effect in 1842, but was bitterly opposed for two decades because it did not fit with the accepted physics of the time (it contradicted the Luminiferous Aether theory.) Doppler was finally proven right in 1868 when W. Huggins observed red shifts and blue shifts in stellar spectra. Unfortunately this was fifteen years after Doppler had died.

I haven't been able to find anything about this in various short online biographies of Doppler (1, 2). Doppler tested the effect himself by having someone play a trumpet on a train (really), someone else successfully tested it in 1845, and it was independently rediscovered in 1848. Doppler himself was made the head of the Institute For Experimental Physics in Vienna and died about as prestigious and beloved as a physicist can get.

So my impression is that only a third of these people really fit the pattern. Most of them were doubted for very short periods, continued to be respected in their fields for their other accomplishments even during those periods, or were part of medium-sized movements rather than being lone geniuses. After a few years – maybe an average of ten, very rarely as long as thirty – their contributions were recognized and they assumed their rightful place in the pantheon. Science isn't perfect. But it is *darned good*.

[EDIT: Bill Beatty, author of the original list, responds here. My response to the response here.]

II.

I bring this up in the context of my last post on progress in the rationalist movement. There used to be a stereotype that rationalists were too quick to challenge scientific consensus. I think that was exaggerated, but based on a core of truth. Given that we're interested in the ways that bias can prevent people from accepting truth, it's unsurprising that we would focus on cases like these.

But I personally have changed my thinking on this a lot. Not in any way that I can explain explicitly – I've always thought something like:

> Scientific consensus is the best tool we have for seeking truth. It's not perfect, and it's frequently overturned by later scientists, but this is usually – albeit not literally always – the work of well-credentialed insiders, operating pretty quickly after the evidence that should overturn it becomes available. Any individual should be very doubtful of their ability to beat it, while not being so doubtful that nobody ever improves it and science can never progress.

– and I still think that. But I’ve shifted from being the sort of person who shares viral lists of maligned geniuses, to the sort of person who debunks those lists. I’ve started emphasizing the “best tool we have” part of the sentence, and whispering the “isn’t perfect” part, rather than vice versa.

I’ve changed my mind on this because of personal experience. Rather than trying to describe it, it might be more helpful to give the most salient examples.

1. The Replication Crisis: I previously thought the scientific consensus was flawed because it failed to take the replication crisis seriously enough. I later learned that everyone else took the repliaction crisis exactly as seriously as I did. A poll in *Nature* shows that 90% of scientists believe reproducibility issues constitute a “crisis”, compared to only 3% (!) who don’t. For every person complaining about “methodological terrorists”, there are a dozen who are very concerned and trying to change the way they practice research.

This is especially impressive because as far as I can tell the whole shift happened in about ten years. I would date the beginning of the crisis from Ioannidis’ original 2005 paper, although it was only aimed at medicine. It got into high gear in psychology sometime around 2011 with Simonsohn’s False Positive Psychology. A Google Trends analysis suggests people only started searching the relevant keywords around 2013.

I started thinking about this sort of thing in 2009 after reading this LW post. At the time I thought this was some sort of exciting failure of modern science that I alone had figured out. But this was well after sharp people like Ioannidis were talking about it, and only a few years before *everyone* was talking about it. Framing this as "I was right and scientific consensus was wrong" seems grandiose. Better might be "I started betting on a winning horse about a quarter of the way between the beginning of the race and when its victory became blatantly obvious to everyone".

2. Nutrition: The Bad Old Paradigm of nutrition says that obese people just have poor impulse control, that weight is a simple matter of calories in vs. calories out, and that all calories are equally good except fat, which for some inexplicable reason is the Devil. Anybody who's read a few good books about nutrition science knows that the Bad Old Paradigm is woefully inadequate. I read a few of those books and became convinced that I was right and scientific consensus was wrong.

Unfortunately, this whole issue exploded when Gary Taubes published *Good Calories, Bad Calories*, which as best I can tell combined the first publicly available good critique of the Bad Old Paradigm with a flawed and basically false attempt at a new paradigm. There were lots of confused attacks against Taubes' bad information which did collateral damage to his good information, and lots of confused defenses of his good information which inadvertently shielded his bad information from criticism. I previously focused on defend the good parts, but recently shifted more towards criticizing the bad parts.

After reading some more good books here (one of which I hope to review soon), my impression is that most nutrition scientists don't believe in the Bad Old Paradigm and haven't for a while. At the very least, most of them seem to believe in the lipostat and think it's important, which is my proxy for "basically has their heart in the right place". Insofar as the Bad Old Paradigm continues to be popular wisdom, it's because of the diet industry, the government, social inertia, and nobody really having a good new paradigm to replace it with. I’m gradually seeing popular wisdom shift, and nutrition scientists themselves seem to be helping this process rather than hurting it.

Maybe somebody in this area has discovered the new paradigm and is a maverick being persecuted by hidebound reactionaries. But it isn’t Gary Taubes. And it certainly isn’t me.

3. Social-Justice-Related Issues: Another narrative I used to believe was that a lot of sketchy ideas were being flattered because they spoke to left-leading academics’ biases in favor of social justice. Implicit association tests, stereotype threat, the idea of zero meaningful psychological differences between men and women, et cetera.

When I started worrying about implicit association tests, I thought I was defying some kind of broad scientific consensus. But the meta-analyses showing the Implicit Association Test didn’t do what people thought had been around since 2009 and have only gotten more numerous since then, with broad media coverage. Problems with stereotype threat research are getting mainstream coverage and even airtime on NPR.

The problem here is that there was no equivalent of the *Nature* poll on the replication crisis, so I didn’t realize any of this was happening until just recently. For example, in 2016 this Voxsplainer made it sound like there was a monolithic consensus in favor of Implicit Association Tests that no sane person had ever disagreed with, even though by that point there were already several big meta-analyses finding they weren’t practically useful. The correct conclusion isn’t that this is really what scientific consensus thinks. The correct conclusion is that Vox shouldn’t be trusted about any science more complicated than the wedge vs. inclined plane. Once I realized that there was all this intelligent analysis going on that I'd never heard about, my claim to be boldly defying the scientific consensus evaporated.

Yes, Cordelia Fine is still around and is still writing books arguing against gender differences. But she's starting to sound *really* defensive, basically the literary equivalent of "I know I'm going to be downvoted to hell for this, but…". Meanwhile, other scientists are doing a good job pointing out the flaws in her books and conducting studies like this biggest-ever look at male vs. female brain differences, this magisterial look at personality differences, et cetera – not to mention great and widely-accepted work on how intersex people take on more characteristics of their hormonal than their social gender (honestly, we should probably thank transgender people for making this field socially acceptable again). People talk a lot about how Larry Summers was fired from Harvard for talking about male vs. female differences, but Steven Pinker did a whole debate on this and remains a Harvard professor.

Even things about genetic psychological differences between population groups are less bold and maverick-y than their proponents like to think. The relevant surveys I know trying to elicit scientific consensus (1, 2, 3) all find that, when asked anonymously, most scientists think these differences explain about 25% – 50% of variance.

I hate to bring that up, because it'll probably start a flame war in the comments, but I think it's important as a sign of exactly how hard it is to politicize science. Global warming skeptics talk about how maybe the scientific consensus on global warming is false because climatologists face political pressure to bias their results in favor of the theory. But scientists studying these areas face much more political pressure, and as long as you give the surveys anonymously they're happy to express horrendously taboo opinions. This is about the strongest evidence in favor of the consensus on global warming – and scientific consensus in general – that I could imagine.

4. Nuture Assumption and Blank Slatism: The prologue of the first edition of *The Nurture Assumption* is Judith Rich Harris telling her "maverick genius kept down by hidebound reactionaries" story. But the prologue of the second edition is her being much more hopeful:

> To some extent at least, times have changed…there is now more acceptance of the idea that behavior is influenced by genes and that individual differences in behavior are due in part to differnces in genes. People are more willing to admit that children can inherit behavioral quirks and personality characteristics…was it this cultural shift that led to greater acceptance of my theory? Or was it the fact that new findings, consistent with the theory, kept turning up? Over time, the early, angry response to *The Nurture Assumption* has softened noticeably, both within and outside of academia. Today, the book is widely cited in textbooks and journal articles. It's assigned and discussed in courses in many colleges and universities; it shows up in exams…in his foreward to the first ediction of *The Nurture Assumption*, Steven Pinker made a rash prediction about the book: "I predict it will come to be seen as a turning point in the history of psychology". Perhaps it is too soon to judge whether psychology has rounded a bend; perhaps it will take the perspective of twenty or thirty years. Even at this point, though, there are signs of a slight shift in direction. Within developmental psychology, I've noticed that descriptions of procedures and results are beginning to sound a bit defensive. Greater progress

> has been made in other areas of psychology. And the email I receive from students gives me high hopes for the younger generation coming up.

There were ten years between the first and second editions of *The Nurture Assumption*. In the almost ten years since the publication of the second edition, my impression is that its ideas have become even more widely-accepted. This month's edition of the *American Journal of Psychiatry*, onbe of the top journals in the field, has a great study showing that child abuse does *not* cause cognitive disability, in contrast to several previous studies in the area. It cites Deary, Plomin, and Ioannidis, hits all of the talking points about genetic confounding of developmental outcomes, and receives glowing endorsement in the journal's editorial section, which says that "if our causal explanations are wrong, we may be wasting our effort or even doing damage". Every single psychiatrist in the country is getting exposed to this way of thinking.

And this has real results. I got to present a summary of behavioral genetics to a meeting of psychiatrists, including a lot of psychoanalysts, and I was shocked that most of them were at least a little receptive. I think they misunderstood it. I think they carefully raised caveats in exactly the right places to ensure they didn't have to change anything they were doing. But the *overall* response was "Oh, yeah, we've heard stuff like that, it seems plausible, good thing that for various hard-to-explain reasons none of it applies to us." This is what the first stage of progress looks like.

5. Intelligence Explosion And AI Risk: This was another place where I and many of my friends thought we were right and the consensus was wrong. It was another place where a lot of self-appointed defenders of the consensus told us we

were crackpots and needed to listen to what real scientists thought. And again, when I looked into it, there was no consensus against the idea and lots of prominent researchers were in favor. Going to the Asilomar Conference and seeing a bunch of people from MIT and Harvard talk about how concerned they were really opened my eyes on this. Google now has an AI Ethics Board, Berkeley, Oxford, and MIT have foundations working on it, and people like Elon Musk and Bill Gates are involved. Bostrom's survey of AI researchers and some more recent and rigorous not-yet-published surveys I've heard about confirm the impression. Nobody would ever say there's a scientific consensus *in favor* of Bostrom's theories. But at this point I think it's also indefensible to say there's a consensus against.

Bostrom first started writing about these sorts of things extensively in the early 2000s, so there was really only a ten-year gap between entering the intellectual environment and it becoming a (mostly) accepted part of the established field. Those ten years felt pretty long while we were in them, but the ability of a field to accept an on-the-face-of-it completely-insane-sounding theory within ten years seems to me a very strong argument against the hidebound-reactionaries theory and a very strong argument for considering scientific consenses to be unreasonably effective.

6. IQ: Another case where I worried about apparent failure of scientific consensus due to politically bias. I certainly encountered a lot of falsehoods around this when I was younger. My high school psychology textbook included a section claiming that all IQ tests were biased towards rich white people because they were based entirely on questions like "how many shots below par is a bogey?" Then it presented an "alternate IQ test" which "proved" that poor minorities had higher IQs than rich whites by asking some

other questions with the opposite bias (I think they were about slang for drugs – certainly an interesting way to fight stereotypes). This kind of thing naturally made me assume that nobody had any idea what was actually in IQ tests and scientists were idiots.

But more recently I’ve been reading actual surveys, which find that about 97% of expert psychologists and 85% of applied psychologists agree that IQ tests measure cognitive ability “reasonably well”. And 77% of expert psychologists and 63% of applied psychologists agree IQ tests are culture-fair (with slightly different numbers depending on how you ask the question, but always about 50% of both groups).

This seems like less of a problem with expert consensus, and more of a problem of nobody else (including textbook writers!) listening to experts who are continually trying to beat reality into people’s heads. But I have a vague memory of having recently seen a survey (which I can’t find) that even experts in softer fields like sociology are generally in favor of IQ and admit that it has its uses. And even some left/liberal sources like Vox and Freddie deBoer are aware of the consensus and willing to respect it.

At the same time, I’ve encountered some people like Borsboom and Nostalgebraist who have relatively sophisticated (and limited) critiques of IQ, and who have allowed me to round off other people’s less-well-framed critiques to something more like what they are saying and less like the stupid things my high school textbook said.

So it seems to me that generally experts agree with reasonable statements about IQ, and where they seem to disagree they may hold reasonable disagreements rather than unreasonable ones. Again, where this fails is not in the experts but in the ability of people who don’t listen to the experts to get disproportionate social power and hide the existence of the expert consensus.

III.

Last week I wrote about universally-known criticisms of economists, like “they’re silly for assuming everyone behaves perfectly rationally”:

> My impression is that economists not only know about these criticisms, but invented them. During the last few paradigm shifts in economics, the new guard levied these complaints against the old guard, mostly won, and their arguments percolated down into the culture as The Correct Arguments To Use Against Economics. Now the new guard is doing their own thing – behavioral economics, experimental economics, economics of effective government intervention. The new paradigm probably has a lot of problems too, but it’s a pretty good bet that random people you stop on the street aren’t going to know about them.

The same pattern explains a lot of my concerns above. I knew some criticisms of a scientific paradigm. They seemed right. I concluded that scientists weren’t very smart and maybe I was smarter. I should have concluded that some cutting-edge scientists were making good criticisms of an old paradigm. I can still flatter myself by saying that it’s no small achievement to recognize a new paradigm early and bet on the winning horse. But the pattern I was seeing was part of the process of science, not a condemnation of it.

Most people understand this intuitively about past paradigm shifts. When a creationist says that we can’t trust science because it used to believe in phlogiston and now it believes in combustion, we correctly respond that this is exactly why we *can* trust science. But this lesson doesn't always generalize when you're in the middle of a paradigm shift *right now* and having trouble seeing the other side.

I realize I'm (ironically) risking making my narrative of scientific success unfalsifiable. Suppose someone wants to argue that scientific consensus is wrong. If they point to something it used to be wrong about, I can respond "Yes, but it self-corrected and it's correct now, so that's fine." If they point to something where cutting-edge scientists say it's wrong but nobody else agrees, I can respond "Yes, this is what the beginning of a paradigm shift looks like, so that's fine". And if they point to something where *nobody* in the field thinks it's wrong, I can say "You're a crackpot for going against all reputable scientists; the problem is with you." And if later they turn out to be right, and everyone acknowledges it, I can say "Yes, but it self-corrected and it's correct now, so that's fine."

(and I'm making it even easier for myself in that I say "scientific consensus for" when I probably mean "no scientific consensus against". I don't claim that 90%+ of scientists always believe true things, only that there are very few cases where 90%+ of scientists believe things which smarter people know to be false.)

Against this I can only offer a personal narrative: the only light I have by which to judge scientific consensus is my own Inside View assessment of what seems correct. Again and again I have *tried* to defy scientific consensus. And every time, I either find that I am wrong, find that I am a few years ahead of a trend that most scientists eventually agree with, or find that what I thought was "scientific consensus" was actually a fiction peddled by biased industry or media sources slandering a scientific community which actually had a much more sophisticated picture. My history of trying to fight scientific consensus has been a *Man Who Was Thursday*-esque series of embarassments as I find again and again that my supposed enemy agrees with me and is even better at what I am trying to do than I am.

Scientific consensus hasn't just been accurate, it's been *unreasonably* accurate. Humans are fallible beings. They are not known for their ability the change their mind, to willingly accept new information, or to put truth-seeking above political squabbles. And our modern society is not exactly known for being an apolitical philosopher-kingdom with strong truth-seeking institutions completely immune from partisan pressure. I feel a deep temptation to sympathize with global warming denialists who worry that the climatological consensus is biased politicized crap, because that is *exactly* the sort of thing which I would expect to come out of our biased politicized crappy society. Yet again and again I have seen examples of scientific fields that have maintained strong commitments to the truth in the face of pressure that would shatter any lesser institution. I've seen fields where people believe incredibly-bizarre sounding things that will get them mocked at cocktail parties just because those things seem to be backed by the majority of the evidence. I've even seen people *change their minds*, in spite of all the incentives to the contrary. I can't explain this. The idea that scientific consensus is almost always an accurate reflection of the best knowledge we have at the time seems even more flabbergasting than any particular idea that scientists might or might not believe. But it seems to be true.

(note that I'm talking about "scientific consensus" to mean a very high-level pattern, consisting of hundreds of scientists over the space of decades evaluating a broad body of work. Any individual study is still probably total garbage.)

Given how weird all of this is, I realize there's another possible bias here that should be taken very seriously – which is that I'm wrong about one or both sides of this. Which is more likely: that Science always agrees with Truth? Or that one guy's perception of Science always agrees with that same guy's perception of Truth? The latter gives me two degrees of freedom: I can either cherry-pick experts who agree with me and declare them to be Consensus, or I can conform my opinions to consensus so slavishly that I end up discovering only that Consensus agrees with itself. I don't *feel* like I'm making this kind of mistake. But then again, nobody ever *feels like* they're being biased.

But if I'm making this mistake, I think it's at least a *better* mistake than the one where people dream up stories about being mavericks persecuted by hidebound reactionaries. This mistake at least sets the terms of debate as "let's try to ascertain what the scientific community thinks" and forbids me from believing *completely* crackpottish things. And it encourages trust in one of our more trustworthy public institutions, always a prosocial sort of thing to do. I would rather have a world of people debating who agrees with scientific consensus or not, than a world of people debating whether scientific consensus is even valuable.

There are two caveats to the above. First, I think it's dangerous to promote a norm of agreeing with scientific consensus, insofar as that helps encourage exactly the mistakes about the nature of consensus that I discussed above. When poorly-informed diet industry gurus support the Bad Old Paradigm, their rallying cry is usually "You're a stupid crackpot, bow to the scientific consensus which agrees with me". I gave three examples above of cases where I would have gotten the scientific consensus 100% wrong if I didn't have access to a formal survey of scientific experts. In a world where these surveys had never been done – or some existing field without these surveys – or some field where these surveys have been done inaccurately or in a biased manner – people will often believe the consensus to be the opposite of what it really is. In those cases, demands that people respect consensus can be used to shut down people who are actually right – the field-wide equivalent of calling true facts you don't like debunked and well-refuted. I see this happening all the time and I worry that waxing too poetically about the unreasonable effectiveness of scientific consensus will only serve to empower these people. Goodhart's Law says that a measure which becomes a target ceases to be a useful measure, so we should be reluctant to target scientific consensus too strongly.

And second, I think that even when the Outside View tells you that the consensus is correct, you should continue pursuing your Inside View hunch that it isn't. This avoids awkward situations like every individual scientist doubting the consensus, but suppressing their doubts because the "scientific consensus" has to be right.

So maybe the things I'm saying about scientific consensus aren't very actionable. But respecting scientific consensus in a non-actionable way is a lot less exhausting than believing yourself to be against it, and talking about how you're against it, and taking flak for being against it. And in the same way it's helpful to believe that God is good, even if He never really gets around to doing much about it, so it's reassuring to be able to have faith in our institutions every so often.

SSC Journal Club: Childhood Trauma And Cognition

This month's *American Journal of Psychiatry* includes Danese et al, Origins Of Cognitive Deficits In Victimized Children. Previous studies had found that abused children had lower IQ. They concluded that the severe stress of being abused must decrease brain function. Danese et al challenge that assumption in the context of the new paradigm of shared-environment-skeptical psychiatry.

They looked at some of the big developmental studies that had followed thousands of children from birth and evaluated them throughout the years. Most of these kids had gotten a bunch of different IQ tests every year, starting as young as 3. These children had also been screened for abuse, both by questions that tended to predict abuse (for example, asking parents whether they ever "punished their children harshly") and by asking the kids, once they were grown up, whether they'd ever been abused or not.

The hope was to find children who hadn't been abused before their first IQ test at age 3, but who started being abused before their last IQ test at age 18. Then they would see whether IQ at age 3 predicted IQ at age 18 in the same way it would in unabused children, or whether IQ was lower than predicted, presumably because of the abuse.

Before the adjustments, they found the same thing as every other study – abused children had lower IQ than unabused children ($p < 0.01$, $B = 0.1$):

> Consistent with previous research, we found that adolescents and adults with a history of childhood victimization have pervasive deficits in clinically

> significant cognitive functions, including both general intelligence and more specific measures of executive unction, processing speed, memory, perceptual reasoning, and verbal comprehension.

After the adjustments, there was no difference between IQs in the two groups ($p = 0.13$, $B = 0.03$):

> In contrast to the conventional causal interpretation of these findings, our longitudinal prospective design revealed that cognitive deficits in victimized adolescents and adults were largely explained by cognitive deficits present before the observational period for childhood victimization and by nonspecific effects of childhood socioeconomic disadvantage. On the one hand, the results are consistent with the high heritability of cognitive functions, their strong continuity across the life course, and the stable cognitive deficits previously described in children exposed to adversity. On the other hand, they are inconsistent with the causal effects of early-life stress on brain function reported in experimental animal models.

The study raises the possibility that kids from poorer families were more likely to get abused and to have lower IQs. Other possibilities include abusers (like criminals) having lower-than-average IQs and passing them on to their kids, or low-IQ kids being less able to figure out ways to escape abuse.

I find this interesting for two reasons.

First, twin studies famously find that early family experiences (the “shared environment”) are less important than previously believed. This has led to a small cottage industry of trying to deny twin studies and saying they don’t work. There’s a lot of interesting debate in this area, and by “interesting” I mean “annoying and interminable”. Maybe the best solution is to route around the whole problem by viewing twin studies as an incentive to conduct *really rigorous* non-twin studies to see if alternate methodologies can confirm the same surprising conclusions. Danese et al is a rare example of this being done right.

Second, a lot of discussion of the limited role of the shared environment adds the caveat “of course, we’re not talking about actual abuse here. Of course *that* affects outcomes.” It probably affects some of them. But here we see that it’s not necessarily true along every axis.

A recent twin study showed that child abuse doesn’t increase risk for ADHD, autism, or learning disorders. But I don’t think most people expected it to increase the risk of these disorders, which are traditionally viewed as really genetic.

A 2010 study found that it didn’t increase the risk of “conduct problems” broadly defined. These *are* traditionally viewed as linked with child abuse, so that’s interesting.

A 2013 study found that child abuse mostly didn’t explain personality disorders. Traditionally borderline personality disorder in particular is viewed as almost invariably associated with abuse, but if I’m reading the study right abuse really only explained about 1% of the variance in who had the condition. And although this called itself a “co-twin analysis”, it wasn’t a real twin study – it used non-abused twins as the control group for abused twins, but it didn’t check if they were identical or non-identical; this controls for family socioeconomic status but not for genetics. This leaves open the possibility that even the 1% remaining abuse-related variance might not be real.

The strongest study I could find was this one linking childhood sexual abuse to a threefold lifetime increase in psychiatric disorders. It, too, was a cotwin analysis and so not genetically controlled. And although I am not going to win any sensitivity points for saying that genetics can affect likelihood of being sexually abused, this is my impression of a lot of the cases I've seen – for example, children with a lot of of preexisting issues and bad relationships with their parents are more willing to secretly meet predators who have been grooming them online.

On the other hand, there are also reasons to believe these studies might understate the dangers of abuse. For example, the Danese et al study sampled 3000 kids who were supposed to be representative of the communities they came from. Suppose only 1% of children experience very severe child abuse. Then they would only have a sample of 30 – far too small to find any interesting conclusions. Maybe most of what they're picking up is differences between not-abused-at-all kids and only-mildly-abused kids, and we shouldn't be too surprised if they don't find anything major.

I don't want to underplay the role of child abuse. My impression from anecdotal evidence is that it does seem to have serious negative effects on kids later in life. For example, I've seen some kids from well-off, genetically-healthy families who were sexually abused by a coach or a neighbor or a priest or someone and who became nervous wrecks later in life, unable to stop thinking about it or to stop feeling guilty. Even good studies usually find that trauma is a risk factor for PTSD (thank goodness), proving that highly stressful events can be psychologically scarring and lead to future problems. Once you've accepted that, you've got most of the assumptions you need to reconstruct the case for child abuse causing psychological problems.

I think the safest assumption is that child abuse that rises to the level of trauma is associated with a PTSD-like picture (although I’ve heard some experts discourage thinking of it as literally PTSD) that seriously affects later-life outcomes. I think it’s less likely that child abuse has many effects that are wildly different from what we see in normal PTSD. And since normal PTSD doesn’t decrease IQ, I’m not surprised to hear that child abuse doesn’t do that either.

Book Review: The Hungry Brain

[Content note: food, dieting, obesity]

I.

The Hungry Brain gives off a bit of a Malcolm Gladwell vibe, with its cutesy name and pop-neuroscience style. But don't be fooled. Stephan Guyenet is no Gladwell-style dilettante. He's a neuroscientist studying nutrition, with a side job as a nutrition consultant, who spends his spare time blogging about nutrition, tweeting about nutrition, and speaking at nutrition-related conferences. He is very serious about what he does and his book is exactly as good as I would have hoped. Not only does it provide the best introduction to nutrition I've ever seen, but it incidentally explains other neuroscience topics better than the books directly about them do.

I first learned about Guyenet's work from his various debates with Gary Taubes and his supporters, where he usually represents the "scientific consensus" side. He is very careful to emphasize that the consensus doesn't look anything like Taubes' caricature of it. The consensus doesn't believe that obesity is just about weak-willed people voluntarily choosing to eat too much, or that obese people would get thin if they just tried diet and exercise, or that all calories are the same. He writes

> The [calories in, calories out] model is the idea that our body weight is determined by voluntary decisions about how much we eat and move, and in order to control our body weight, all we need is a little advice about how many calories to eat and burn, and a little willpower. The primary defining feature of this model is that it

> assumes that food intake and body fatness are not regulated. This model seems to exist mostly to make lean people feel smug, since it attributes their leanness entirely to wise voluntary decisions and a strong character. I think at this point, few people in the research world believe the CICO model.
>
> [Debate opponent Dr. David] Ludwig and I both agree that it provides a poor fit for the evidence. As an alternative, Ludwig proposes the insulin model, which states that the primary cause of obesity is excessive insulin action on fat cells, which in turn is caused principally by rapidly-digesting carbohydrate. According to this model, too much insulin reduces blood levels of glucose and fatty acids (the two primary circulating metabolic fuels), simultaneously leading to hunger, fatigue, and fat gain. Overeating is caused by a kind of "internal starvation". There are other versions of the insulin model, but this is the one advocated by Ludwig (and Taubes), so it will be my focus.
>
> But there's a third model, not mentioned by Ludwig or Taubes, which is the one that predominates in my field. It acknowledges the fact that body weight is regulated, but the regulation happens in the brain, in response to signals from the body that indicate its energy status. Chief among these signals is the hormone leptin, but many others play a role (insulin, ghrelin, glucagon, CCK, GLP-1, glucose, amino acids, etc.)

The Hungry Brain is part of Guyenet's attempt to explain this third model, and it basically succeeds. But like many "third way" style proposals, it leaves a lot of ambiguity. With CICO, at least you know where you stand – confident that everything is based on willpower and that you can ignore biology completely. And again, with Taubes, you know where you stand – confident that willpower is useless and that low-carb diets will solve everything. *The Hungry Brain* is a little more complicated, a little harder to get a read on, and at times pretty wishy-washy.

But listening to people's confidently-asserted simple and elegant ideas was how we got into this mess, so whatever, let's keep reading.

II.

The Hungry Brain begins with the typical ritual invocation of the obesity epidemic. Did you know there are entire premodern cultures where literally nobody is obese? That in the 1800s, only 5% of Americans were? That the prevalence of obesity has doubled since 1980?

Researchers have been keeping records of how much people eat for a long time, and increased food intake since 1980 perfectly explains increased obesity since 1980 – there is no need to bring in decreased exercise or any other factors. Exercise *has* decreased since the times when we were all tilling fields ten hours a day, but for most of history, as our exercise decreased, our food intake decreased as well. But for some reason, starting around 1980, the two factors uncoupled, and food intake started to rise despite exercise continuing to decrease.

Guyenet discusses many different reasons this might have happened, including stress-related overeating, poor sleep, and quick prepackaged food. But the ideas he keeps coming back to again and again are food reward and satiety.

In the 1970s, scientists wanted to develop new rat models of obesity. This was harder than it sounded; rats ate only as much as they needed and never got fat. Various groups tried to design various new forms of rat chow with extra fat, extra sugar, et cetera, with only moderate success – sometimes they could get the rats to eat a *little* too much and *gradually* become *sort of* obese, but it was a hard process. Then, almost by accident, someone tried feeding the rats human snack food, and they ballooned up to be as fat as, well, humans. The book:

> Palatable human food is the most effective way to cause a normal rat to spontaneously overeat and become obese, and its fattening effect cannot be attributed solely to its fat or sugar content.

So what *does* cause this fattening effect? I think the book's answer is "no single factor, but that doesn't matter, because capitalism is an optimization process that designs foods to be as rewarding as possible, so however many different factors there are, every single one of them will be present in your bag of Doritos". But to be more scientific about it, the specific things involved are some combination of sweet/salty/umami tastes, certain ratios of fat and sugar, and reinforced preferences for certain flavors.

Modern food isn't just unusually rewarding, it's also unusually bad at making us full. The brain has some pretty sophisticated mechanisms to determine when we've eaten enough; these usually involve estimating food's calorie load from its mass and fiber level. But modern food is calorically dense – it contains many more calories than predicted per unit mass – and fiber-poor. This fools the brain into thinking that we're eating less than we really are, and shuts down the system that would normally make us feel full once we've had enough. Simultaneously, the extremely high level of food reward tricks the brain into thinking that this food is especially nutritionally valuable and that it should relax its normal constraints.

Adding to all of this is the so-called “buffet effect”, where people will eat more calories from a variety of foods presented together than they would from any single food alone. My mother likes to talk about her “extra dessert stomach”, ie the thing where you can gorge yourself on a burger and fries until you’re totally full and couldn’t eat another bite – but then mysteriously find room for an ice cream sundae afterwards. This is apparently a real thing that’s been confirmed in scientific experiments, and a major difference between us and our ancestors. The !Kung Bushmen, everyone’s go-to example of an all-natural hunter-gatherer tribe, apparently get 50% of their calories from a single food, the mongongo nut, and another 40% from meat. Meanwhile, we design our meals to include as many unlike foods as possible – for example, a burger with fries, soda, and a milkshake for dessert. This once again causes the brain to relax its usual strict constraints on appetite and let us eat more than we should.

The book sums all of these things up into the idea of “food reward” making some foods “hyperpalatable” and “seducing” the reward mechanism in order to produce a sort of “food addiction” that leads to “cravings”, the “obesity epidemic”, and a profusion of “scare quotes”.

I’m being a little bit harsh here, but only to highlight a fundamental question. Guyenet goes into brilliant detail about things like the exact way the ventral tegmental area of the brain responds to food-related signals, but in the end, it all sounds suspiciously like “food tastes good so we eat a lot of it”. It’s hard to see where exactly this differs from the paradigm that he dismisses as “attributing leanness to wise voluntary decisions and a strong character…to make lean people feel smug”. Yes, food tastes good so we eat a lot of it. And Reddit is fun to read, but if someone browses Reddit ten hours a day and doesn’t do any work then we start speculating about their character, and maybe even feeling smug. This part of the book, taken alone, doesn't really explain why we shouldn't be doing that about weight too.

III.

There is only one fat person on the Melanesian island of Kitava – a businessman who spends most of his time in modern urbanized New Guinea, eating Western food. The Kitavans have enough food, and they live a relaxed tropical lifestyle that doesn't demand excessive exercise. But their bodies match caloric intake to expenditure with impressive precision. So do the !Kung with their mongongo nuts, Inuit with their blubber, et cetera.

And so do Westerners who limit themselves to bland food. In 1965, some scientists locked people in a room where they could only eat nutrient sludge dispensed from a machine. Even though the volunteers had no idea how many calories the nutrient sludge was, they ate exactly enough to maintain their normal weight, proving the existence of a "sixth sense" for food caloric content. Next, they locked morbidly obese people in the same room. They ended up eating only tiny amounts of the nutrient sludge, one or two hundred calories a day, without feeling any hunger. This proved that their bodies "wanted" to lose the excess weight and preferred to simply live off stored fat once removed from the overly-rewarding food environment. After six months on the sludge, a man who weighed 400 lbs at the start of the experiment was down to 200, without consciously trying to reduce his weight.

In a similar experiment going the opposite direction, Ethan Sims got normal-weight prison inmates to eat extraordinary amounts of food – yet most of them still had trouble gaining weight. He had to dial their caloric intake up to 10,000/day –

four times more than a normal person – before he was able to successfully make them obese. And after the experiment, he noted that “most of them hardly had any appetite for weeks afterward, and the majority slimmed back down to their normal weight”.

What is going on here? Like so many questions, this one can best be solved by grotesque Frankenstein-style suturing together of the bodies of living creatures.

In the 1940s, scientists discovered that if they damaged the ventromedial hypothalamic nucleus (VMN) of rats, the rats would basically never stop eating, becoming grotesquely obese. They theorized that the VMN was a “satiety center” that gave rats the ability to feel full; without it, they would feel hungry forever. Later on, a strain of mutant rats was discovered that seemed to naturally have the same sort of issue, despite seemingly intact hypothalami. Scientists wondered if there might be a hormonal problem, and so they artificially conjoined these rats to healthy normal rats, sewing together their circulatory systems into a single network. The result: when a VMN-lesioned rat was joined to a normal rat, the VMN-lesioned rat stayed the same, but the normal rat stopped eating and starved to death. When a mutant rat was joined to a normal rat, the normal rat stayed the same and the mutant rat recovered and became normal weight.

The theory they came up with to explain the results was this: fat must produce some kind of satiety hormone, saying “Look, you already have a lot of fat, you can feel full and stop eating now”. The VMN of the hypothalamus must detect this message and tell the brain to feel full and stop eating. So the VMN-lesioned rats, whose detector was mostly damaged, responded by never feeling full, eating more and more food, and secreting more and more (useless) satiety hormone. When they were joined to normal rats, their glut of satiety hormones flooded the normal rats – and their normal brain, suddenly bombarded with “YOU ALREADY HAVE WAY TOO MUCH FAT” messages, stopped eating entirely.

The mutant rats, on the other hand, had lost the ability to *produce* the satiety hormone. They, too, felt hungry all the time and ate everything. But when they were joined to a normal rat, the normal levels of satiety hormone flowed from the normal rat into the mutant rat, reached the fully-functional detector in their brains, and made them feel full, curing their obesity.

Skip over a lot of scientific infighting and unfortunate priority disputes and patent battles, and it turns out the satiety hormone is real, exists in humans as well, and is called leptin. A few scientists managed to track down some cases of genetic leptin deficiency in humans, our equivalent of the mutant rats, and, well…

> Usually they are of normal birth weight and then they’re very, very hungry from the first weeks and months of life. By age one, they have obesity. By age two, they weigh 55-65 pounds, and their obesity only accelerates from there. While a normal child may be about 25% fat, and a typical child with obesity may be 40% fat, leptin-deficient children are up to 60% fat. Farooqi explains that the primary reason leptin-deficient children develop obesity is that they have “an incredible drive to eat”… leptin-deficient children are nearly always hungry, and they almost always want to eat, even shortly after meals. Their appetite is so exaggerated that it’s almost impossible to put them on a diet: if their food is restricted, they find some way to eat, including retrieving stale morsels from the trash can and gnawing

> on fish sticks directly from the freezer. This is the desperation of starvation […]
>
> Unlike normal teenagers, those with leptin deficiency don't have much interest in films, dating, or other teenage pursuits. They want to talk about food, about recipes. "Everything they do, think about, talk about, has to do with food" says [Dr.] Farooqi. This shows that the [leptin system] does much more than simply regulate appetite – it's so deeply rooted in the brain that it has the ability to hijack a broad swath of brain functions, including emotions and cognition.

It's the leptin-VNM-feedback system (dubbed the "lipostat") that helps people match their caloric intake to their caloric requirements so impressively. The lipostat is what keeps hunter-gatherers eating exactly the right number of mongongo nuts, and what keeps modern Western overeaters at much closer to the right weight than they could otherwise expect.

The lipostat-brain interface doesn't just control the raw feeling of hunger, it seems to have a wide range of food-related effects, including some on higher cognition. Ancel Keys (of getting-blamed-for-everything fame) ran the Minnesota Starvation Experiment on some unlucky conscientious objectors to World War II. He starved them until they lost 25% of their body weight, and found that:

> Over the course of their weight loss, Keys's subjects developed a remarkable obsession with food. In addition to their inescapable, gnawing hunger, their conversations, thoughts, fantasies, and dreams revolved around food and eating – part of a phenomenon Keys called "semistarvation neurosis". They became fascinated by recipes and cookbooks, and some even

> began collecting cooking utensils. Like leptin-deficient adolescents, their lives revolved around food. Also like leptin-deficient adolescents, they had very low leptin levels due to their semi-starved state.

Unsurprisingly, as soon as the experiment ended, they gorged themselves until they were right back at their pre-experiment weights (but no higher), at which point they lost their weird food obsession.

Just as a well-functioning lipostat is very good at keeping people normal weight, a malfunctioning lipostat is very good at keeping people obese. Fat people seem to have “leptin resistance”, sort of like the VMN-lesioned rats, so that their bodies get confused about how much fat they have. Suppose a healthy person weighs 150 lbs, his body is on board with that, and his lipostat is set to defend a 150 lb set point. Then for some reason he becomes leptin-resistant, so that the brain is only half as good at detecting leptin as it should be. Now he will have to be 300 lbs before his brain “believes” he is the right weight and stops encouraging him to eat more. If he goes down to a “mere” 250 lbs, then he will go into the same semistarvation neurosis as Ancel Keys’ experimental subjects and become desperately obsessed with food until they get back up to 300 again. Or his body will just slow down metabolism until his diet brings him back up. Or any of a bunch of other ways the lipostat has to restore weight when it wants to.

This explains the well-known phenomenon where contestants on *The Biggest Loser* who lose 200 or 300 pounds for the television camera pretty much always gain it back after the show ends. Even though they’re highly motivated and starting from a good place, their lipostat has seized on their previous weight as the set point it wants to defend, and resisting the lipostat is somewhere between hard and impossible. As far as I know, nobody has taken Amptoons up on their challenge to find a single peer-reviewed study showing any diet that can consistently bring fat people to normal weight and keep them there. After a certain level of lipostat dysregulation, “solving” weight problems by diet and exercise becomes hard-to-impossible, and the people who loudly insist otherwise tend to kind of be jerks.

And alas, it doesn’t seem to work to just inject leptin directly. As per Guyenet

> People with garden variety obesity already have high levels of leptin…while leptin therapy does cause some amount of fat loss, it requires enormous doses to be effective – up to forty times the normal circulating amount. Also troubling is the extremely variable response, with some people losing over thirty pounds and others losing little or no weight. This is a far cry from the powerful fat-busting effect of leptin in rodents. [Leptin as] the new miracle weight-loss drug never made it to market.
>
> This disappointment forced the academic and pharmaceutical communities to confront a distressing possibility: the leptin system defends vigorously against weight loss, but not so vigorously against weight gain. “I have always thought, and continue to believe,” explained [leptin expert Rudy] Leibel, “that the leptin hormone is really a mechanism for detecting deficiency, not excess.” It’s not designed to constrain body fatness, perhaps because being too fat is rarely a problem in the wild. Many researchers now believe that while low leptin levels in humans engage a powerful starvation response that promotes fat gain, high leptin levels don’t engage an equally powerful response that promotes fat loss.

> Yet *something* seems to oppose rapid fat gain, as Ethan Sims' overfeeding studies (and others) have shown. Although leptin clearly defends the lower limit of adiposity, the upper limit may be defended by an additional, unidentified factor – in some people more than others.

This is the other half of the uncomfortable dichotomy that makes me characterize *The Hungry Brain* as "wishy-washy". The lipostat is a powerful and essentially involuntary mechanism for getting weight exactly where the brain wants, whether individual dieters are cooperative or not. Here it looks like obesity is nobody's fault, unrelated to voluntary decisions, and that the standard paradigm really *is* just an attempt by lean people to feel smug. Practical diet advice starts to look like "inject yourself with quantities of leptin so massive that they overcome your body's resistance". How do we connect this with the other half of the book, the half with food reward and satiety and all that?

IV.

With more rat studies!

Dr. Barry Levin fed rats either a healthy-rat-food diet or a hyperpalatable-human-food diet, then starved and overfed them in various ways. He found that the rats defended their obesity set points in the expected manner, but that the same rats defend different set points depending on their diets. Rats on healthy-rat-food defended a low, healthy-for-rats set point; rats on hyperpalatable-human-food defended a higher set point that kept them obese.

That is, suppose you give a rat as much Standardized Food Product as it can eat. It eats until it weighs 8 ounces, and stays that weight for a while. Then you starve it until it only weighs 6 ounces, and it’s pretty upset. Then you let it eat as much as it wants again, and it overeats until it gets back to 8 ounces, then eats normally and maintains that weight.

But suppose you get a rat as many Oreos as it can eat. It eats until it weighs 16 ounces, and stays that weight for a while. Then you starve it until it only weighs 6 ounces. Then you let it eat as much as it wants again, and *this time* it overeats until it gets back to 16 ounces, and eats normally to maintain that weight.

Something similar seems to happen with humans. A guy named Michel Cabanac ran an experiment in which he put overweight people on two diets. In the first diet, they ate Standardized Food Product, and naturally lost weight since it wasn’t very good and they didn’t eat very much of it. In the second diet, he urged people to eat less until they matched the first group’s weight loss, but to keep eating the same foods as normal – just less of them. The second group reported being hungry and having a lot of trouble dieting; the first group reported not being hungry and not having any trouble at all.

Guyenet concludes:

> Calorie-dense, highly rewarding food may favor overeating and weight gain not just because we passively overeat it but also because it turns up the set point of the lipostat. This may be one reason why regularly eating junk food seems to be a fast track to obesity in both animals and humans…focusing the diet on less rewarding foods may make it easier to lose weight and maintain weight loss because the lipostat doesn’t fight it as vigorously. This may be part of the explanation for why all weight-loss diets seem to work to some extent – even those that are based on

> diametrically opposed principles, such as low-fat, low-carbohydrate, paleo, and vegan diets. Because each diet excludes major reward factors, they may all lower the adiposity set point somewhat.

(this reminds me of the Shangri-La Diet, where people would drink two tablespoons of olive oil in the morning, then find it was easy to diet without getting hungry during the day. People wondered whether maybe the tastelessness of the olive oil had something to do with it. Could it be that the olive oil is temporarily bringing the lipostat down to its “bland food” level?)

Why should some food make the lipostat work better than other food? Guyenet now gets to some of his own research, which is on a type of brain cell called a POMC neuron. These neurons produce various chemicals, including a sort of anti-leptin called Neuropeptide Y, and they seem to be a very fundamental part of the lipostat and hunger system. In fact, if you use superprecise chemical techniques to kill NPY neurons but nothing else, you can cure obesity in rats.

The area of the hypothalamus with POMC neurons seem to be damaged in overweight rats *and* overweight humans. Microglia and astrocytes, the brain’s damage-management and repair cells, proliferated in appetite-related centers, but nowhere else. Maybe this literal damage corresponds to the metaphorically “damaged” lipostat that’s failing to maintain weight normally, or the “damaged” leptin detector that seems to be misinterpreting the body’s obesity?

In any case, eating normal rat food for long enough appears to heal this damage:

> Our results suggest that obese rodents suffer from a mild form of brain injury in an area of the brain that’s

> critical for regulating food intake and adiposity. Not only that, but the injury response and inflammation that developed when animals were placed on a fattening diet *preceded* the development of obesity, suggesting that this brain injury could have played a role in the fattening process.

Guyenet isn't exactly sure what aspect of modern diets cause the injury:

> Many researchers have tried to narrow down the mechanisms by which this food causes changes in the hypothalamus and obesity, and they have come up with a number of hypotheses with varying amounts of evidence to support them. Some researchers believe the low fiber content of the diet precipitates inflammation and obesity by its adverse effects on bacterial populations in the gut (the gut microbiota). Others propose that saturated fat is behind the effect, and unsaturated fats like olive oil are less fattening. Still others believe the harmful effects of overeating itself, including the inflammation caused by excess fat and sugar in the bloodstream and in cells, may affect the hypothalamus and gradually increase the set point. In the end, these mechanisms could all be working together to promote obesity. We don't know all the details yet, but we do know that easy access to refined, calorie-dense, highly rewarding food leads to fat gain and insidious changes in the lipostat in a variety of species, including humans. This is particularly true when the diet offers a wide variety of sensory experiences, such as the hyperfattening "cafeteria diet" we encountered in chapter 1.
>
> Personally, I believe overeating itself probably plays an important role in the process that increases the

adiposity set point. In other words, repeated bouts of overeating don’t just make us fat; they make our bodies want to stay fat. This is consistent with the simple observation that in the United States, most of our annual weight gain occurs during the six-week holiday feasting period between Thanksgiving and the new year, and that this extra weight tends to stick with us after the holidays are over…because of some combination of food quantity and quality, holiday feasting ratchets up the adiposity set point of susceptible people a little bit each year, leading us to gradually accumulate and defend a substantial amount of fat. Since we also tend to gain weight at a slower rate during the rest of the year, intermittent periods of overeating outside of the holidays probably contribute as well.

How might this happen? We aren’t entirely sure, but researchers, including Jeff Friedman, have a possible explanation: excess leptin itself may contribute to leptin resistance. To understand how this works, I need to give you an additional piece of information: Leptin doesn’t just correlate with body fat levels; it also responds to short-term changes in calorie intake. So if you overeat for a few days, your leptin level can increase substantially, even if your adiposity has scarcely changed (and after your calorie intake goes back to normal, so does your leptin). As an analogy for how this can cause leptin resistance, imagine listening to music that’s too loud. At first, it’s thunderous, but eventually, you damage your hearing, and the volume drops. Likewise, when we eat too much food over the course of a few days, leptin levels increase sharply, and this may begin to desensitize the brain circuits that respond to leptin. Yet Rudy Leibel’s group has also shown that high leptin levels alone aren’t enough – the hypothalamus

> also seems to require a second “hit” for high leptin to increase the set point of the lipostat. This second hit could be the brain injury we, and others, have identified in obese rodents and humans.

And he isn’t sure exactly what aspect of the normal rodent diet promotes the healing:

> I did do some research in mice suggesting that unrefined, simple food does reverse the brain changes and the obesity. I don’t claim that it’s all attributable to the blandness though– the two diets differed in many respects (palatability, calorie density, fiber content, macronutrient profile, fatty acid profile, content of nonessential nutrients like polyphenols). Also, we don’t know how well the finding applies to humans yet. One of the problems is that it’s very hard to get a group of humans to adhere strictly to a whole food diet for long enough to study its long-term effects on appetite and body fatness. People are very attached to the pleasures of the palate!

But all of this together seems to point to a potential synthesis between the hyperpalatability and lipostat models. For most of human history, the lipostat faced only mild stresses and was able to maintain a normal weight without much trouble. Modern society has been incentivized to produce hyperpalatable, low-satiety food as superstimuli. This modern food is able to overwhelm normal satiety cues and produce short-term overeating. And for some reason this short-term overeating raises the lipostat’s set point (maybe because of brain damage and leptin resistance), causing long-term weight gain in a way that is very difficult to reverse.

V.

But I still have trouble reconciling these two points of view.

A couple of days ago, I walked by an ice cream store. I'd just finished lunch, and I wasn't very hungry at the time, but it looked like really good ice cream, and it was hot out, so I gave in to temptation and ate a 700 calories sundae. Does this mean:

1. Based on the one pound = 3500 calories heuristic, I have now gained 0.2 lbs. That extra weight will stay with me my whole life, or at least until some day when I diet and eat 700 calories less than my requirement. If I were to eat ice cream like this a hundred times, I would gain twenty pounds.

2. My lipostat adjusts for the 700 extra calories, and causes me to exercise more, or ramp up my metabolism, or burn more brown fat, or eat less later on, or *something*. I don't gain any weight, and eating the ice cream was that rarest of all human experiences, a completely guiltless pleasure. I should eat ice cream whenever I feel like it, or else I am committing the sin of denying myself a lawful pleasure.

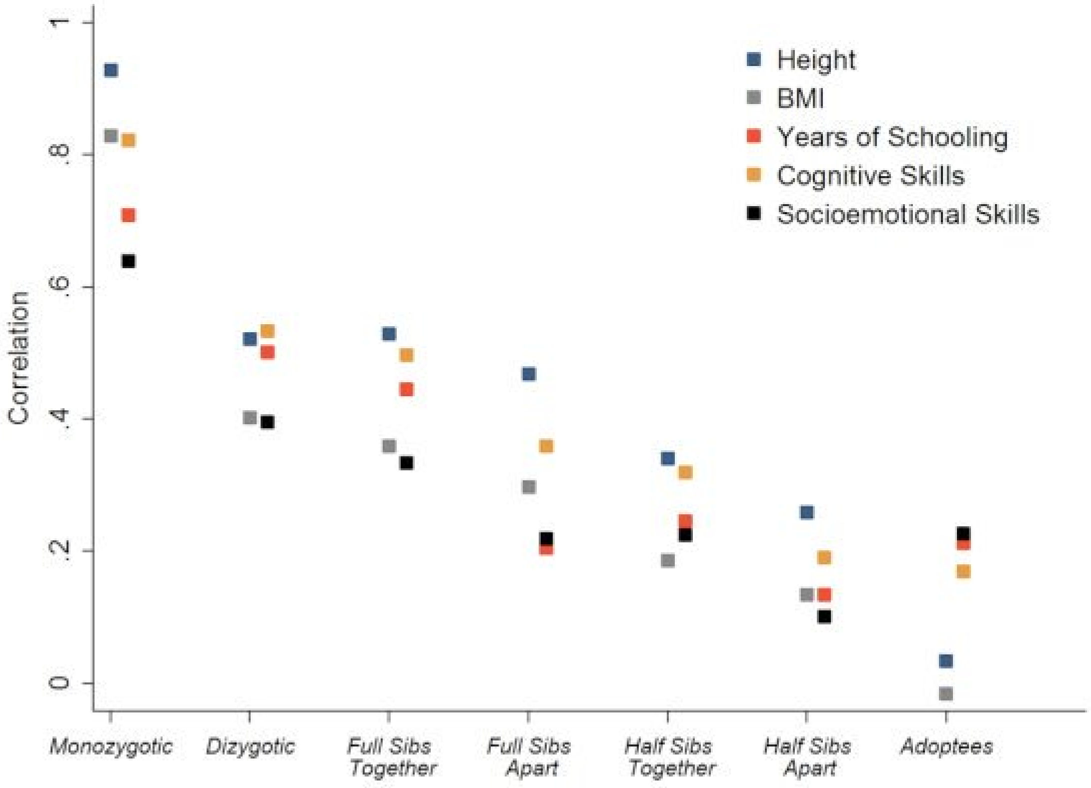
3. My lipostat will more or less take care of the ice cream today, and I won't notice the 0.2 pounds on the scale, but it is very gradually doing hard-to-measure damage to my hypothalamus, and if I keep eating ice cream like this, then one day when I'm in my forties I'm going to wake up weighing three hundred pounds, and no diet will ever be able to help me.

4. Not only will I gain 0.2 pounds immediately, but my lipostat will adjust to *want* to be 0.2 pounds heavier, and I will never lose it, even if I try really hard to diet later.

5. The above scenario is impossible. Even if I *think* I just ate ice cream because it looked good, in reality I was driven to

do it by my lipostat’s quest for caloric balance. Any feeling of choice in the matter is an illusion.

I think the reason this is so confusing is because the real answer is “it could be any one of these, depending on genetics.”



Note the position of the grey squares representing BMI

Right now, within this culture, variation in BMI is mostly genetic. This isn’t to say that non-genetic factors aren’t involved – the difference between 1800s America and 2017 America is non-genetic, and so is the difference between the perfectly-healthy Kitavans on Kitava and the one Kitavan guy who moved to New Guinea. But once everyone alike is exposed to the 2017-American food environment, differences between the people in that environment seem to be really hereditary and not-at-all-related to learned behavior. Guyenet acknowledges this:

> Genes explain that friend of yours who seems to eat a lot of food, never exercises, and yet remains lean. Claude Bouchard, a genetics researcher at the Pennington Biomedical Research Center in Baton Rouge, Louisiana, has shown that some people are intrinsically resistant to gaining weight even when they overeat, and that this trait is genetically influenced. Bouchard’s team recruited twelve pairs of identical twins and overfed each person by 1,000 calories per day above his caloric needs, for one hundred days. In other words, each person overate the same food by the same amount, under controlled conditions, for the duration of the study.
>
> If overeating affects everyone the same, then they should all have gained the same amount of weight. Yet Bouchard observed that weight gain ranged from nine to twenty-nine pounds! Identical twins tended to gain the same amount of weight and fat as each other, while unrelated subjects had more divergent responses…Not only do some people have more of a tendency to overeat than others, but some people are intrinsically more resistant to gaining fat even if they do overeat.
>
> The research of James Levine, an endocrinologist who works with the Mayo Clinic and Arizona State University, explains this puzzling phenomenon. In a carefully controlled overfeeding study, his team showed that the primary reason some people readily burn off excess calories is that they ramp up a form of calorie-burning called “non-exercise activity thermogenesis” (NEAT). NEAT is basically a fancy term for fidgeting. When certain people overeat, their brains boost calorie

> expenditure by making them fidget, change posture frequently, and make other small movements throughout the day. It’s an involuntary process, and Levine’s data show that it can incinerate nearly 700 calories per day. The “most gifted” of Levine’s subjects gained less than a pound of body fat from eating 1,000 extra calories per day for eight weeks. Yet the strength of the response was highly variable, and the “least gifted” of Levine’s subjects didn’t increase NEAT at all, shunting all the excess calories into fat tissue and gaining over nine pounds of body fat…
>
> Together, these studies offer indisputable evidence that genetics plays a central role in obesity and dispatch the idea that obesity is primarily due to acquired psychological traits.

These studies suggest that one way genetics affects obesity is by altering the tolerance level of the lipostat. Genetically privileged people may have very finicky lipostats that immediately burn off any extra calories they eat, and which never become dysregulated. Genetically unlucky people may have weak lipostats which fail to defend against weight gain, or which are too willing to adjust their set point up in the presence of an unhealthy food environment.

So, given how many people seem to have completely different weight-gain-related experiences to each other, the wishy-washyness here might be a feature rather than a bug.

One reason I’ve always found genetics so exciting is that there are all these fields – nutrition is a great example, but this applies at least as much to psychiatry – where everyone has wildly different personal experiences, and where there’s a large and vocal population of people who say that the research is exactly the opposite of their lived experiences.

People have tried to shoehorn the experiences to fit the research, with various levels of plausibility and condescendingness. And for some reason, it’s always really hard to generate the hypothesis “people’s different experiences aren’t an illusion; people are genuinely really different”. Once you start looking at genetics, everything sort of falls into place, and ideas which seemed wishy-washy or self-contradictory before are revealed as just reflecting the diversity of nature. People who were previously at each other’s throats disputing different interpretations of the human condition are able to peacefully agree that there are many different human conditions, and that maybe we can all just get along. *The Hungry Brain* and other good books in its vein offer a vision for how we might one day be able to do that in nutrition science.

VI.

Lest I end on too positive a note, let me reiterate the part where happiness is inherently bad and a sort of neo-Puritan asceticism is the only way to avoid an early grave.

There’s a sort of fatalism to talking about “food reward”. If the enemy were saturated fat, we could just stick with the sugary sweetness of Coca-Cola. If the enemy were carbohydrates, we could go out for steak every night. But what do we do if the enemy is deliciousness itself?

A few weeks ago Guyenet announced *The Bland Food Cookbook*, a collection of tasteless recipes guaranteed to be low food-reward and so discourage overeating. It was such a natural extension of his philosophy that it took me a whole ten seconds to realize it was an April Fools joke. But why should it be? Shouldn’t this be exactly the sort of thing we’re going for?

I asked him, and he responded that:

> If I thought enough people would actually be capable of following the diet, I would consider making such a cookbook non-ironically. The second point I want to make here is that there are many ways to lose weight, and deliberately reducing food reward is only one of them. You could also exercise, eat a low-calorie-density diet, eat a high-protein diet, restrict a macronutrient, restrict animal foods, restrict plant foods, eat nothing but potatoes. Most approaches overlap with a low-reward diet to varying degrees, but I don't think the low reward value encapsulates everything about why every weight loss strategy works. BTW, low-carb folks often have a knee-jerk reaction to the low-reward thing that goes something like this: "I eat food that's delicious, such as steaks, bacon, butter, etc. It's not low in reward." But it is low reward in the sense that you're cutting out a broad swath of foods, and an entire macronutrient, that the brain very much wants you to eat. Eating more of a particular category of rewarding food doesn't completely make up for the fact that you're cutting out a whole other category of rewarding food that you would avidly consume if you weren't restricting yourself.

So things aren't maximally bad. And hunter-gatherers enjoy their healthy diets just fine. And certainly there are things like steak and wine and so on which are traditionally "good food" without being hyperprocessed hyperpalatable junk food. But if you *really* enjoy a glass of Chardonnay, is that "food reward" in a sense that's potentially dangerous? Is anything safe? What about mongongo nuts? Is there anywhere we can get them?

At least the consolation of things making a little more sense. One thing that always used to bother me was the so-called "ice cream diet" – that is, if you eat a 700-calorie ice cream sundae for breakfast, lunch, and dinner every day, that'll be only 2100 calories – around the average person's daily requirement, and low enough for many people to lose weight. Why *wouldn't* you want to do that? The idea of junk food being inherently damaging – while it has a bit of Puritan feel to it – at least fits our intuitions on these sorts of things and gives us a first step towards reconciling the conventional wisdom and the calorie math.

Overall I strongly recommend *The Hungry Brain* for everything I talk about here and for some other good topics I didn't even get to (stress, sleep, economics). I would also recommend Guyenet's other writing, especially his debate with Dr. David Ludwig on the causes of obesity (Part 1, Part 2, Part 3. I also recommend the list of diet tips that Guyenet gives at the end of the book. I won't give them all away here – he's been nice enough to me that maybe I should repay him by *not* reprinting the entire text of his book online for free. But it's similar to a lot of standard advice for healthy living, albeit with more interesting reasoning behind it. Did you know that exercise might help stabilize the lipostat? Or that protein might do the same? Also, one piece of advice you might *not* hear anywhere else – potatoes are apparently off-the-charts in terms of satiety factor and may be one of the single best things to diet on.

And speaking of good things to diet on…

(note that this next part is my own opinion, *not* taken from *The Hungry Brain* or endorsed by Stephan Guyenet)

Slate Star Codex's first and most loyal sponsor is MealSquares, a "nutritionally complete" food product sort of like a whole-foods-based, protein-rich, solid version of Soylent (it looks and tastes basically like a scone). I’m having some trouble writing this paragraph, because I want to recommend them as potentially dovetailing with *The Hungry Brain*‘s philosophy of nutrition without using phrases that might make MealSquares Inc angry at me like “bland”, “low food reward”, or “not hyperpalatable”. I think the best I can come up with is “unlikely to injure your hypothalamus”. So, if you’re looking for an easy way to quit the junk food and try a low-variety diet that’s unlikely to injure your hypothalamus, I recommend MealSquares as worth a look.

Anorexia And Metabolic Set Point

[Content warning: fat, anorexia, dieting]
[Epistemic status: crazy speculation from someone who isn't an expert in the field. Please don't take too seriously.]

Some anorexics I talk to describe their condition as falling into two phases.

In Phase 1, they're a model or a ballerina or something where they're under a lot of pressure to become thin. So they go on a diet, they have trouble just like everyone else, but they stick to it and gradually get thinner. People praise them for their thinness. They win the Ballerina Of The Year award. They're pretty happy with how things are going, and their dieting is ego syntonic – ie they agree with it and can explain the reasons they think it's a good decision.

In Phase 2, somebody tells them, wait, this is unhealthy. Maybe they end up in the hospital. Maybe some friends stage an intervention. They admit that they have anorexia and decide to get better and gain weight. And then they can't. The exact way that this presents varies among people. A few say they never feel hungry – as if they're always so stuffed they couldn't eat another bite. More often they still feel hungry, but they feel a sense of horror at their fatness, even if they rationally understand they're really thin. This makes it very hard to regain weight.

This reminds me of the model of obesity I talked about the other day. People start overeating for social reasons (specifically, because our society is full of tasty food). Then they find they can't stop – their body has adjusted to its higher weight as a new set point and will defend it vigorously, producing hunger pangs and food obsessions if they try to diet.

The brain system producing these effects is called the "lipostat", and it seems to work through a control loop. Fat cells produce a hormone called leptin. The hypothalamus measures the amount of leptin, compares to its preferred set point, and based on the result generates strong urges to eat/exercise/fidget either more or less. This in turn makes the person gain or lose weight, increasing or decreasing the amount of fat and the concentration of leptin, and completing the loop.

In (at least some cases of) obesity, constant overeating makes the hypothalamus leptin-resistant – that is, it systematically underestimates the blood level of leptin. Suppose a healthy person weighs 150 lbs, his body is on board with that, and his lipostat is set to defend a 150 lb set point. Then for some reason he becomes leptin-resistant, so that the brain is only half as good at detecting leptin as it should be. Now he will have to be 300 lbs before his brain "believes" he is the right weight and stops encouraging him to eat more. If he goes down to a "mere" 250 lbs, his brain will panic, believe he's starving, and demand he eat more.

One point that *The Hungry Brain* stressed again and again is the cognitive complexity of the hunger response. For example, from Ancel Keys' Starvation Experiment:

> Over the course of their weight loss, Keys's subjects developed a remarkable obsession with food. In addition to their inescapable, gnawing hunger, their conversations, thoughts, fantasies, and dreams revolved around food and eating – part of a phenomenon Keys called "semistarvation neurosis". They became fascinated by recipes and cookbooks, and some even began collecting cooking utensils. Like leptin-deficient adolescents, their lives revolved around food. Also like

> leptin-deficient adolescents, they had very low leptin levels due to their semi-starved state.

And from studies of leptin-deficient children:

> Farooqi explains that the primary reason leptin-deficient children develop obesity is that they have “an incredible drive to eat”…leptin-deficient children are nearly always hungry, and they almost always want to eat, even shortly after meals. Their appetite is so exaggerated that it’s almost impossible to put them on a diet: if their food is restricted, they find some way to eat, including retrieving stale morsels from the trash can and gnawing on fish sticks directly from the freezer. This is the desperation of starvation […] Unlike normal teenagers, those with leptin deficiency don’t have much interest in films, dating, or other teenage pursuits. They want to talk about food, about recipes. “Everything they do, think about, talk about, has to do with food” says [Dr.] Farooqi. This shows that the [leptin system] does much more than simply regulate appetite – it’s so deeply rooted in the brain that it has the ability to hijack a broad swath of brain functions, including emotions and cognition.

Once again, it’s not really clear how people’s metabolic set point drifts up, but it seems to maybe have something to do with overeating junk food.

So suppose there was an exactly opposite process. People who severely undereat find that their metabolic set point drifts down, with their hypothalamus becoming hypersensitive to leptin. Again, consider a person who would ideally weight 150 lbs. If they become twice as sensitive to leptin, their body won’t be happy unless they weigh 75 pounds. If they gain weight to be 100 pounds instead, they’ll be getting “too much” leptin, their brain will feel like they’re too fat, and their bodies will vigorously “defend against” weight gain to try to return to the distorted set point. This sounds a lot like anorexia nervosa.

Some more evidence: people with anorexia *fidget like crazy*. This is the conventional wisdom among generations of psychiatrists, and has been confirmed in studies – see for example Measurement Of Fidgeting In Patients With Anorexia Nervosa Using A Novel Shoe-Based Monitor, which found anorexics fidget almost twice as much as healthy controls. This seems really damning to me. Consciously deciding to fidget is *hard*. If you don’t believe me, try to fidget as much as possible for the next two hours (the length of the study linked above) and see how well it goes. Most people just can’t do it. On the other hand, fidgeting (renamed to the more dignified “non-exercise activity thermogenesis”) is the classic strategy that the lipostat uses to maintain unconscious control over weight:

> The research of James Levine, an endocrinologist who works with the Mayo Clinic and Arizona State University, explains this puzzling phenomenon. In a carefully controlled overfeeding study, his team showed that the primary reason some people readily burn off excess calories is that they ramp up a form of calorie-burning called “non-exercise activity thermogenesis” (NEAT). NEAT is basically a fancy term for fidgeting. When certain people overeat, their brains boost calorie expenditure by making them fidget, change posture frequently, and make other small movements throughout the day. It’s an involuntary process, and Levine’s data show that it can incinerate nearly 700 calories per day.

A purely social paradigm of anorexia can’t explain the fidgeting; a biological paradigm outright predicts it.

I’m not an expert on anorexia, so for all I know lots of people have thought of this and I’m late to the party. But a quick Google search only reveals a few small leads. Heidebrand et al have studied leptin in anorexia nervosa, but only because they think it might affect symptoms. They find circulating levels are low – which makes sense given anorexics’ low BMI, and which neither confirms nor disconfirms a leptin hypersensitivity hypothesis. They also find that leptin increases more quickly when anorexics gain weight than when healthy people gain weight, which is suggestive but doesn’t prove that much.

I also found a paper, the appropriately-named Treasure (2005) which described anorexia nervosa apparently due to brain lesions. For example:

> Of the seven anorexia cases associated with primary tumours in the area of brain stem and the fourth ventricle (Table 1: cases 24–30),12,25–30 two (24 and 25) presented as typical restrictive anorexia nervosa with fear of fatness; surgical removal of the tumours led to remission and sustained weight gain in both cases.
>
> Thirteen cases of eating disorders associated with lesions in the cerebral hemispheres were identified (Table 1: 31–43).31–38 The damage was predominantly localised in the frontal and temporal lobes (six frontal, four temporal, three frontotemporal) of the right hemisphere (nine right, three left, one bilateral)…seven cases presented as “typical” anorexia nervosa with weight and shape preoccupations.

Most of these cases weren’t typical anorexia, and almost all of them involved other obviously-neurological symptoms that ordinary anorexics lack. And most of the relevant tumors were *not* the hypothalamus (although a similar paper by Chipkevitch finds a disproportionately high number of hypothalamic lesions). But it does prove that excessive concern about becoming fat isn’t just caused by ballet coaches and the patriarchy. It can also be induced by purely neurological mechanisms.

(and a complicating factor – most brain tumors increase intracranial pressure, which seems to increase the ability of leptin to enter the cerebrospinal fluid in ways that might cause obesity or something. There’s actually some suggestive literature that “leptin resistance” might just be serum leptin not making it into the cerebrospinal fluid, and a bunch of studies show that anorexics have proportionally much higher – and obese people proportionally much lower – CSF:serum leptin ratios. But these studies don’t explore causality and it might just be that if you’re too fat your leptin transporters get overloaded.)

I probably shouldn’t focus on leptin so much. Everybody agrees that leptin is only one part of the biology of weight gain, that other hormones are probably more important in weight loss, and that using leptin to represent the entire biological regulation of obesity is kind of a streetlight effect.

But I feel like something like this might be true. There must be some reason why the symptoms of anorexia resemble lipostat action so closely. There must be some reason why anorexics fidget so much. And there must be some reason why brain tumors can mimic some of the most striking symptoms of anorexia, like obsessive fear of becoming fat. All of these make sense only in the context of metabolic set point theory, and I hope more people start trying to link these two ideas up.

Neutral vs. Conservative: The Eternal Struggle

I.

Vox's David Roberts writes about Donald Trump and the rise of tribal epistemology.

It's got a long and complicated argument which I can't really do justice to here, but the thesis seems to be that the US Right is defecting against the country's shared institutions in favor of forming its own echo chambers.

So for example, there used to be a relatively fair media in which both liberals and conservatives got their say. But Republicans didn't like having to deal with facts, so they formed their own alternative media – FOX and Rush Limbaugh and everyone in that sphere – where only conservatives would have a say and their fake facts would never get challenged.

Or: everyone used to trust academia as a shared and impartial arbitrator of truth. But conservatives didn't like the stuff it found – whether about global warming or trickle-down economics or whatever – so they seceded into their own world of alternative facts where some weird physicist presents his case that global warming is a lie, or a Breitbart journalist is considered an expert on how cultural Marxism explains everything about post-WWII American history.

It concludes that "the press cannot be neutral", although it also "cannot afford to be, or be seen as primarily instruments of the Democrats". To its credit, it admits this is kind of contradictory:

> They must figure out a way to play a dual role: to be fair and consistent referees of policy and ideological disputes within the public square — while also acting to defend the institutional integrity of the square itself from what is, at present, a highly asymmetrical threat. They must fight to keep some core principles and commitments inviolate, outside the sphere of normal political dispute, against an administration that wants to drag them in… that's a humdinger of a problem.

Let me start by saying what this article gets right.

I think it's right that the two parties used to have much more in common, and be able to appeal to shared gatekeeper institutions that both trusted.

I think it's right the Republicans unilaterally seceded from those shared gatekeeper institutions, so that now we're in the weird position of having two sets of institutions: one labeling itself "neutral" and the other labeling itself "conservative".

I think it's right to consider the situation asymmetrical. Yes, CNN leans liberal, but it's not *as* liberal as FOX is conservative, and it's not as open about it – it has a pretense of neutrality that FOX doesn't, and although we can disagree about how realistic that pretense is I think few people would disagree that the pretense is there. Nor is there a liberal version of FOX that lacks that pretense of neutrality.

I think it's right that the conservative side is worse than the neutral side. However biased and crappy you think CNN and mainstream academia are, FOX and the conservative academic bubble are working on a different level (though note that as a liberal, I *would* say this, and you should interpret it with the same grain of salt that you would any other "my side is better than yours" claim).

I think it’s right that this situation is horrible and toxic and destroying the country, and it’s really good that someone has pointed this out and framed it this clearly.

I think it’s wrong in exactly the way I would expect it to be wrong, which is also an example of what’s wrong with it.

II.

Roberts devotes four sentences in his six thousand word article to the possibility that conservatives might be motivated by something deeper than a simple hatred of facts:

> The right’s view that the institutions lean liberal is hyperbolic, but not without foundation. Science, academia (at least liberal arts and social sciences), and journalism do tend to draw their personnel from left-leaning demographics.
>
> Those institutions have cosmopolitan aspirations — fair application of transpartisan standards — but there’s no doubt that in practice, those aspirations often cover for more parochial preferences.
>
> But the right has not sought greater fairness in mainstream institutions; it has defected to create its own.

Roberts says that these neutral gatekeeper institutions “tend to draw their personnel from left-leaning demographics”, as if this was just a big fuss about 105 New Englanders for every 100 Texans. I would like to counter with a report from a friend who graduated from a top university last year:

> I was at my graduation last weekend, and the commencement address was basically about twenty

minutes of vitriolic insults directed at Trump. And in between burying my head in my friend’s shoulder in discomfort and laughing nervously, I was thinking about the family of this guy in my class.

He’s the first person in his family to go to college. He drove an hour every day to go to a somewhat better high school because there was an epidemic of gang violence at his local school. Against the odds, he did well, and got into college, where he has continued to get good grades and play sports and generally do things that make parents proud.

His family is not well off. They’re Mexican-American. And they’re Trump supporters.

Yeah, I’m kind of confused too. But they honestly are. (Not even reluctant Republicans supporting Trump–they voted for him in the primary. His aunt owns a Make America Great Again cap.) And all I could think about was how happy they must have been to be attending their son’s graduation from one of the best universities in the world [citation needed], only to have that happiness turn to bewilderment and anger as everyone around them cheered a series of caustic attacks against them and their values. The message couldn’t have been clearer: “You don’t belong here.”

My mom thought this speech was So Courageous. When I suggested that it might have been more courageous to say something that not everyone there agreed with, she replied, “the students maybe, but a lot of the parents looked unhappy.”

Seventy percent of the parents there had family incomes over six figures. (More, probably, since low-income parents are less likely to attend graduation.) A lot of

> them are members of the self-perpetuating intellectual/economic elite. This probably isn't true of the few Trump supporters among them.
>
> So if we are going to single them out for judgment, force them to account for their support for an "infantile," "bullying," "proto-fascist" "charlatan"…can it not be on the day of their kids' graduation?

And sure, if you consider me your friend, then that makes this one of those "friend of a friend" stories. But I dare you to say that any of this sounds the least bit implausible. My point is, just because a university paints "ACTUALLY, WE ARE POLITICALLY NEUTRAL" in big red letters on the college quad, doesn't mean that anyone is required to believe it. And the ideology that invented the microaggression can't hide behind "but we haven't *officially* declared you unwelcome!"

And the same thing is happening in the media. For example, *in this very piece*, Roberts cites a Vox poll showing that Trump supporters are more likely to be authoritarians. Vox has pushed this same claim many more times: Authoritarianism: The Political Science That Explains Trump, The Rise Of American Authoritarianism: A Niche Group Of Political Scientists May Have Uncovered What's Driving Donald Trump's Ascent, The Rise Of American Authoritarianism Explained In 6 Minutes, The Best Predictor Of Trump Support Is Authoritarianism.

Okay. But Vox is working off an internal poll that it hasn't released (or at least I can't find it) meaning no one has any idea if the sample size and methodology are okay. And some political science professors tried the same exercise around the same time with excellent methodology and a sample size of over a thousand and found the opposite – Trump supporters were *less* authoritarian than Cruz supporters, and no more authoritarian than Rubio supporters. They did find that Republicans were a bit more authoritarian than Democrats, but correctly noted that the measure involved is literally called “Right-Wing Authoritarianism”, is based on a scale invented by Theodor Adorno to prove conservatives had fascist tendencies, and only asks questions about child-rearing practices (you get marked as “authoritarian” if you have a traditional religious child-rearing style). And there are other investigations of authority that try to control for this sort of thing and sometimes find find liberals and conservatives are about equal in respect for authority.

I don’t want to overdo my criticism. “Right-wing authoritarianism” is a powerful idea with a good academic reputation, and the decision to focus solely on child-rearing was a principled choice to avoid including politics itself in the construct. And failed replications should be an opportunity for reflection rather than a cause to instantly dismiss a finding.

Yet it’s still good practice to mention their existence. And I still feel like somewhere there might be a conservative who reads this sort of thing and feels like Vox is *not quite* the perfectly-neutral mutually-beneficial gatekeeper institution of their dreams.

And whenever I mention this sort of thing, people protest “But Fox and Breitbart are worse!” And so they are. But I feel like Vox has aspirations to be something more than just a mirror image of Fox with a left-wing slant and a voiced fricative. It’s trying to be a neutral gatekeeper institution. If some weird conservative echo chamber is biased, well, what did you expect? If a neutral gatekeeper institution is biased, *now* we have a problem.

Roberts writes that “the right has not sought greater fairness in mainstream institutions; it has defected to create its own”. This is a bizarre claim, given the existence of groups like

Accuracy In Media, Media Research Center, Newsbusters, Heterodox Academy, et cetera which are *all about* the right seeking greater fairness in mainstream institutions, some of which are almost fifty years old. Really "it's too bad conservatives never complained about liberal bias in academia or the mainstream media" seems kind of like the opposite of how I remember the late 20th and early 21st centuries.

The way I remember it, conservatives spent about thirty years alternately pleading, demanding, suing, legislating, and literally praying for greater fairness in mainstream institutions, and it was basically all just hitting their heads against a brick wall. *Then* they defected to create their own.

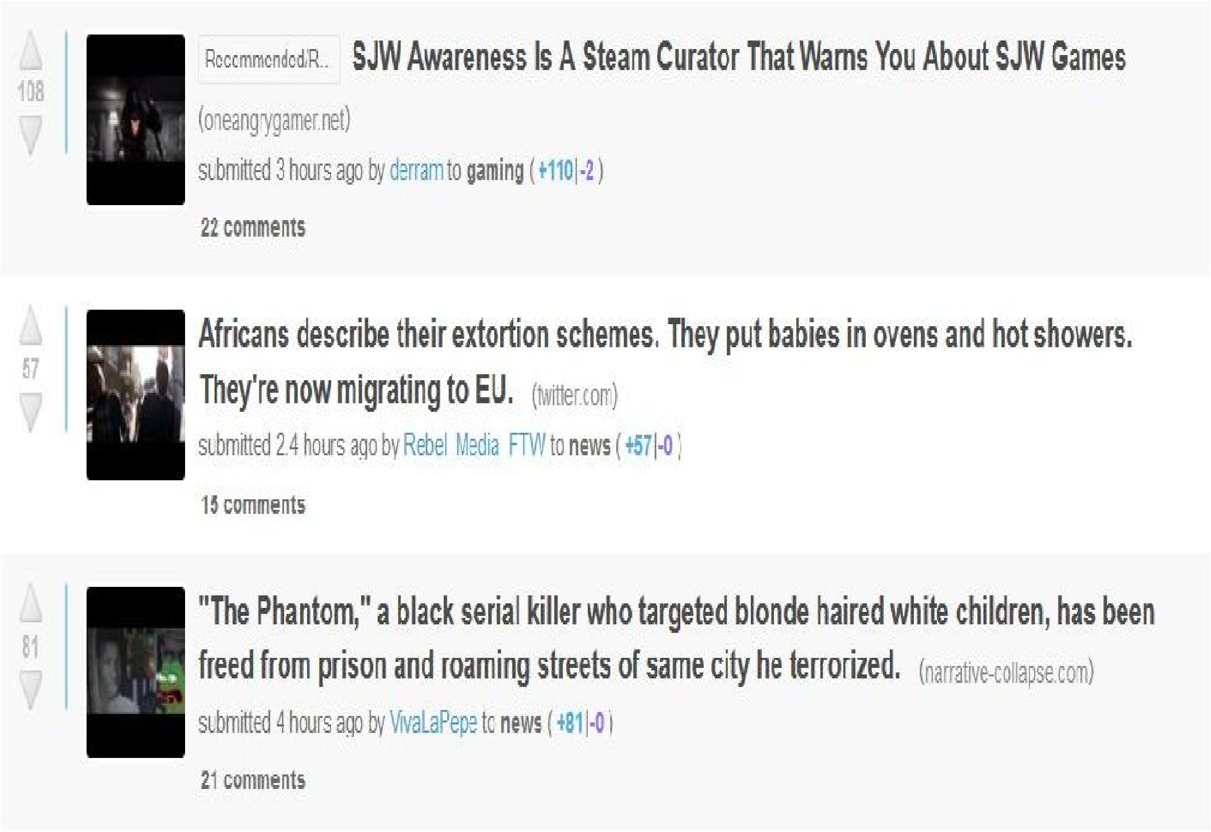
III.

This predictably went badly.

I wrote before (1, 2) about the sort of dynamics this situation produces. A couple of years ago, Reddit decided to ban various undesirables and restrict discussion of offensive topics. A lot of users were really angry about this, and some of them set up a Reddit clone called Voat which promised that everyone was welcome regardless of their opinion.

What happened was – a small percent of average Reddit users went over, lured by curiosity or a principled commitment to free speech. And also, approximately 100% of Reddit's offensive undesirables went there, lured by the promise of being able to be terrible and get away with it.

Even though Voat's rules were similar to Reddit's rules before the latter tightened its moderation policies, Voat itself was nothing like pre-tightening Reddit. I checked to see whether it had gotten any better in the last year, and I found the top three stories were:



Recommended/R.. SJW Awareness Is A Steam Curator That Warns You About SJW Games (oneangrygamer.net)

submitted 3 hours ago by derram to gaming (+110|-2)

22 comments

Africans describe their extortion schemes. They put babies in ovens and hot showers. They're now migrating to EU. (twitter.com)

submitted 2.4 hours ago by Rebel_Media_FTW to news (+57|-0)

15 comments

"The Phantom," a black serial killer who targeted blonde haired white children, has been freed from prison and roaming streets of same city he terrorized. (narrative-collapse.com)

submitted 4 hours ago by VivaLaPepe to news (+81|-0)

21 comments

The moral of the story is: if you’re against witch-hunts, and you promise to found your own little utopian community where witch-hunts will never happen, your new society will end up consisting of approximately three principled civil libertarians and seven zillion witches. It will be a terrible place to live *even if witch-hunts are genuinely wrong*.

FOX’s slogans are “Fair and Balanced”, “Real Journalism”, and “We Report, You Decide”. They were pushing the “actually unbiased media” angle *hard*. I don’t know if this was ever true, or if people really believed it. It doesn’t matter. By attracting only the refugees from a left-slanted system, they ensured they would end up not just with conservatives, but with the worst and most extreme conservatives.

They also ensured that the process would feed on itself. As conservatives left for their ghettos, the neutral gatekeeper institutions leaned further and further left, causing more and more conservatives to leave. Meanwhile, the increasingly obvious horribleness of the conservative ghettos made liberals feel more and more justified in their decision to be biased against conservatives. They intensified their loathing and contempt, accelerating the conservative exodus.

The equilibrium is basically what we see now. The neutral gatekeeper institutions lean very liberal, though with a minority of conservative elites who are good at keeping their heads down and too mainstream/prestigious to settle for anything less. The ghettos contain a combination of seven zillion witches and a few decent conservatives who are increasingly uncomfortable but know there’s no place for them in the mainstream.

IV.

I don’t want to give the impression that this is limited to the places people traditionally gripe about like academia and the media. The same dynamics are going on everywhere.

In the hospital where I work, there’s a RESIST TRUMP poster on the bulletin board in our break room. I don’t know who put it there, but I know that anybody who demanded that it be taken down would be tarred as a troublemaker, and anyone who tried to put a SUPPORT TRUMP poster up next to it would be lectured about how politics are inappropriate at work. This is true even though I think at least a third of my colleagues are Trump supporters.

I went to a scientific conference in a field completely unrelated to politics where one of the researchers giving a presentation started with a five minute tangentially-related anti-Trump rant. I can’t imagine someone giving the opposite rant any more than I can imagine a pro-Trump commencement speaker at my friend’s graduation.

I’m desperately trying to avoid the Nerd Culture Wars, which have somehow managed to be even worse than the Regular Culture Wars, but even I’ve heard about GamerGate and the Rabid Puppies. These were originally movements to fight a perceived liberal bias in regular gaming/sci-fi. They of course failed, and now they’re their own little separate conservative spaces practicing conservative video game commentary/sci-fi writing. I don’t want to deny that they’re often horrible. They’re horrible in exactly the same way FOX News is horrible, and for exactly the same reasons. I expect this pattern of conservatives seceding from theoretically-neutral-but-realistically-left-leading communities and forming terrible communities full of witches to repeat itself again and again, because it’s happening for systemic rather than community-specific reasons.

The overall impression is of a widespread norm, well-understood by both liberals and conservatives, that we have a category of space we call “neutral” and “depoliticized”. These sorts of spaces include institutions as diverse as colleges, newspapers, workplaces, and conferences. And within these spaces, overt liberalism is tolerated but overt conservativism is banned. In a few of these cases, conservatives grew angry enough that they started their own spaces – which began as noble attempts to avoid bias, and ended as wretched hives of offensive troglodytes who couldn’t get by anywhere else. This justifies further purges in the mainstream liberal spaces, and the cycle goes on forever.

Stanford historian Robert Conquest once declared it a law of politics that “any organization not explicitly right-wing sooner or later becomes left-wing”. I have no idea why this should be true, and yet I’ve seen it again and again. Taken to its extreme, it suggests we’ll end up with a bunch of neutral organizations that have become left-wing, plus a few explicitly right-wing organizations. Given that Conquest was writing in the 1960s, he seems to have predicted the current situation remarkably well.

V.

David Roberts ends by noting that he doesn’t really know what to do here, and I agree. I don’t know what to do here either.

But one simple heuristic: if everything you’ve tried so far has failed, maybe you should try something different. Right now, the neutral gatekeeper institutions have tried being biased against conservatives. They’ve tried showing anti-conservative bias. They’ve tried ramping up the conservativism-related bias level. They’ve tried taking articles, and biasing them against conservative positions. I appreciate their commitment to multiple diverse strategies, but I can’t help but wonder whether there’s a possibility they’ve missed.

Look. I read Twitter. I know the sorts of complaints people have about this blog. I’m some kind of crypto-conservative, I’m a traitor to liberalism, I’m too quick to sell out under the guise of “compromise”. And I understand the sentiment. I write a lot about how we shouldn’t get our enemies fired lest they try to fire us, how we shouldn’t get our enemies’ campus speakers disinvited lest they try to disinvite ours, how we shouldn’t use deceit and hyperbole to push our policies lest our enemies try to push theirs the same way. And people very reasonably ask – hey, I notice my side kind of controls all of this stuff, the situation is actually asymmetrical, they have no way of retaliating, maybe we should just grind our enemies beneath our boots this one time.

And then when it turns out that the enemies can just leave and start their *own* institutions, with horrendous results for everybody, the cry goes up “Wait, that’s unfair! Nobody ever said you could do that! Come back so we can grind you beneath our boots some more!”

Conservatives aren’t stuck in here with us. We’re stuck in here with them. And so far it’s not going so well. I’m not sure if any of this can be reversed. But I think maybe we should consider to what degree we are in a hole, and if so, to what degree we want to stop digging.

Getting High On Your Own Supply

There's a lot of debate over what Hillary did wrong in her campaign, and how the Democrats can change before 2018 or 2020.

Some of this revolves around policy positions. For example, should she have supported expanding Medicare to everyone? Should she have been tougher on immigration to preempt some of Trump's support? These are tough questions, made tougher by the need to balance what's right against what's electable.

But other parts of the debate revolve more around her vision and way of looking at the world. Some common points here:

1. She focused too much on identity politics and ignored the white working class
2. She tried to reason with people (eg her performance in the debates) rather than appeal to their emotions
3. She proposed complicated wonkish policy schemes instead of simple things normal people understood like "tax the rich" or "build a wall"
4. She talked too much about managerial competence and not enough about her Soaring Vision
5. She was too "we're all in this together" and not enough "us vs. them", where the "them" is some combination of billionaires, political elites, and Republicans.

Advocates of these points usually end with some plea for Democrats to change their ways and approach these matters differently.

Consider two possible ways that Democrats might act on the first point (and for now we'll ignore the part where "the Democratic Party" is not a single monolithic entity that can take direct action).

They might act superficially by eg telling the campaign worker in charge of TV commercials to make more commercials about the white working class, and fewer commercials about identity politics.

Or they might act deeply by changing the entire culture of the Democratic Party, so that Democrats *think* about identity politics less, identity politics activists are marginalized within party circles, white working class activists are promoted within party circles, party-aligned media sources focus more space on the plight of the white working class, et cetera.

Likewise regarding point 2, the Democrats could fire their old speechwriters and hire new ones who are better at writing emotional appeals. Or they could change the epistemic culture of the party, so that discussions of how to amend Section 421B of the tax code to be 2.4% more fair were met with eye-rolling everywhere from local meetings to DNC headquarters, but rousing speeches about Taking Back The Country were universally met with applause.

And the point I want to make is that the epistemic culture that makes you sound electable isn't necessarily the epistemic culture that makes you competent.

It might be that appealing to the white working class really is the most important way to win elections, but that in the real world "identity politics" surrounding minority groups are a more important or more tractable issue, where government interventions can help far more people.

Or it might be that raving about your Grand Vision is the best way to get elected, but that most of the low-hanging fruit for helping people right now *does* involve wonkish tinkering around with very complicated parts of health care regulation.

Let me give an example of what I mean. The Republicans have an electoral strategy based on a Grand Vision talking about how the elites in Washington have become corrupt and sold out the country to Big Government. This has been very successful for them; no Republican can complain that they don't win enough elections. But it's also completely screwed up their party's ability to govern. Their trouble repealing Obamacare seems like the most glaring example – there just wasn't enough overlap between reality-based policies that made political sense, and policies that legislators could support without worrying about getting primaried by Tea Party types accusing them of selling out.

And this is just a rare (though increasing) example of a time when Republican dysfunction hurt the Republican Party. Most of the time it just hurts the country. Only time will tell exactly who the GOP's dysfunctional presidential primary and subsequent nomination of Trump hurts, but I doubt Republicans will be happy with the results.

Of course, remaining epistemically pure and never winning anything isn't much fun either, so whatever. I guess my only advice for the Democrats is: don't get high on your own supply.

What do I mean by that? A while back, I discussed the recent trend in articles – mostly on the Left – explaining how using facts and reason don't work and so we should switch to making emotion-based appeals. There was an article like this on Financial Times, and another one in Current Affairs.

So what I’m wondering is: are *Financial Times* and *Current Affairs* taking their own advice? That is, when I read one of their articles, am I reading somebody trying to rationally present their argument for my evaluation? Or am I reading an emotional appeal written by someone who thinks facts don’t matter? When a writer at one of these publications tries to decide which side of an issue to support, are they, within their own head, trying to obtain facts and reason about them? Or are they making emotional appeals to themselves?

The media – especially intellectual partisan media like *National Review* or *Jacobin* – are ideologies talking to themselves, processing information, and settling on opinions. The way they report is the way that their respective ideologies think. If a news source decides to report via emotional appeals rather than facts, their ideology’s thinking is taking place through emotional calculations and not factual ones.

Or to put this another way – a lot of the conversation assumes a divide between two natural categories – elites and the public. Elites are unmoved movers, who set strategies and policies based on their own omniscient knowledge of the political calculation. The public is unmoving movees, receiving information from the elites and voting based on which set of elites sounds more convincing.

But reality is more of a spectrum, down from party committee members reading internal bulletins, to party elites reading *National Review*, to informed people reading *The New York Times*, to random yahoos watching reality TV and catching campaign ads in the commercial breaks. Everybody is influenced by the prevailing media environment and helps influence it in turn.

If you optimize for the epistemic culture that’s best for getting elected, but that culture isn’t also the best for running a party or governing a nation, then the fact that your culture affects your elites as well becomes a really big problem. If focus groups tell you that your campaign ads need to be more emotional, more zero-sum, more simplistic, and more oriented to the white working class, it’s pretty scary if you don’t reflect before making your whole ideology’s culture more emotional, zero-sum, simplistic, and oriented to the white working class.

Luckily this has an easy solution. From the same Tim Harford article I cited earlier:

> Last year, three researchers — Seth Flaxman, Sharad Goel and Justin Rao — published a study of how people read news online. The study was, on the face of it, an inquiry into the polarisation of news sources. The researchers began with data from 1.2 million internet users but ended up examining only 50,000. Why? Because only 4 per cent of the sample read enough serious news to be worth including in such a study. (The hurdle was 10 articles and two opinion pieces over three months.) Many commentators worry that we’re segregating ourselves in ideological bubbles, exposed only to the views of those who think the same way we do. There’s something in that concern. But for 96 per cent of these web surfers the bubble that mattered wasn’t liberal or conservative, it was: “Don’t bother with the news.”

If you’re online, your audience isn’t “the public”. Although I don’t have hard statistics on this, my guess is that if you’re writing for a magazine, or speaking at a conference, your audience isn’t “the public” either. You might as well say what you believe to be true, in the manner you think is most productive, and promote the epistemic culture you think is healthiest for the party and the country. And if you write an argument saying not to use facts and reason, maybe append something like “but our publication will continue to be factual and rational, and you should keep being factual and rational in anything with consequences other than public relations.”

If you’re a speechwriter, campaign commercial director, or TV producer, I assume everything you do is already tested via ten million polls and focus groups and meetings. You can keep doing that, except apparently you’re terrible at it and you should probably get much much better.

Silicon Valley: A Reality Check

The nation has spoken: weird pointless $400 wi-fi enabled juicer company Juicero is the perfect symbol of Silicon Valley.

So says the Washington Post: Juicero Shows What's Wrong With Silicon Valley Thinking. So says TechCrunch, which calls Juicero "the absurd avatar of Silicon Valley hubris". So says Newsweek, which renames the area Silly-Con Valley in its honor. And of course there's Deadspin, which calls it "the best story ever written about Silicon Valley… a stupid libertarian dystopia where investor-class vampires are the consumers and a regular person's money is what they go shopping for."

In case you missed it, Juicero was a startup that got $120 million in funding to manufacture high-end juicers which were supposed to be the bleeding-edge in juice-related technology. Then Bloomberg did some investigative reporting and found that you could actually make juice equally well by skipping the $400 juicer and just squeezing the juice packets with your bare hands.

This is, admittedly, pretty silly. But I want to take a step back and suggest a reality check.

While Deadspin was busy calling Silicon Valley "awful nightmare trash parasites", my girlfriend in Silicon Valley was working for a company developing a structured-light optical engine to manipulate single cells and speed up high-precision biological research.

While FastCoDesign was busy calling Juicero "a symbol of the Silicon Valley class designing for its own, insular problems,” a bunch of my friends in Silicon Valley were working for Wave, a company that helps immigrants send remittances to their families in East Africa.

While Vox was busy writing about how Juicero “says a lot about the state of Silicon Valley right now”, Silicon Valley was leading a revolution in solar power that’s resulted in a 1500% increase in cell installations over the past few years.

While Slate was busy telling us that Silicon Valley companies “repackage familiar ideas and sell them back to us as exemplars of Groundbreaking Disruptive Innovation”, Silicon Valley was shooting a fifteen-story rocket a hundred miles into the air at 4,100 mph, then landing it gently on a 300 foot platform in the middle of the ocean.

While Gizmodo was busy writing that this “is not an isolated quirk” because Silicon Valley investors “don’t care that they do not solve problems [and] exist to temporarily excite the affluent into spending money”, Silicon Valley investors were investing $35 million into an artificial pancreas for diabetics.

While Freddie deBoer was busy arguing that Silicon Valley companies “siphon money from the desperate throngs back to the employers who will use them up and throw them aside like a discarded Juicero bag and, of course, to themselves and their shareholders. That’s it. That’s all they are. That’s all they do”, Silicon Valley companies were busy inventing cultured meat products that could end factory farming and save millions of animals from horrendous suffering while also helping the environment.

Or maybe we should try to be more quantitative about this. I looked at the latest batch of 52 startups from legendary Silicon Valley startup incubator Y Combinator.

Thirteen of them had an altruistic or international development focus, including Neema, an app to help poor people without access to banks gain financial services; Kangpe, online health services for people in Africa without access to doctors; Credy, a peer-to-peer lending service in India; Clear Genetics, an automated genetic counseling tool for at-risk parents; and Dost Education, helping to teach literacy skills in India via a $1/month course.

Twelve of them seemed like really exciting cutting-edge technology, including CBAS, which describes itself as "human bionics plug-and-play"; Solugen, which has a way to manufacture hydrogen peroxide from plant sugars; AON3D, which makes 3D printers for industrial uses; Indee, a new genetic engineering system; Alem Health, applying AI to radiology, and of course the obligatory drone delivery startup.

Eighteen of them seemed like boring meat-and-potatoes companies aimed at businesses that need enterprise data solution software application package analytics targeting management something something something "the cloud".

And the remaining nine were your ridiculous niche Uber-for-tacos startups that we all know and love, including Cowlar ("FitBit for cows – it's way smarter than it sounds!"); Origin ("Keurig for smoothies"), MoveButter, which compares itself to three different companies I've never heard of in its first sentence but seems to be grocery-related in some way; Mere Coffee, a better-tasting coffee machine for small businesses; and LitHit, a smart target for shooting sports. I'm sure somebody in the comments is going to tell me why FitBits for cows is actually a vital service that will revolutionize agriculture, but I'm trying to err on the side of caution here.

I'm concerned that Y Combinator might be so successful that they're unique in going for status and do-gooding rather than being a real cross-sample of startups (and they also seem to recruit a lot of international startups from outside Silicon Valley). So I also looked at the first twenty startups in the portfolio of Andreessen Horowitz, a famous Valley venture capitalist firm. One of them seemed explicitly prosocial – some kind of science education partnership company. Four of them seemed high-tech or otherwise awesome – including the obligatory aerial-surveying-with-drones company. Twelve seemed to be some sort of enterprise data solution software application package analytics targeting management something something something "the cloud". And only two of them seemed even a little vapid – eg this high-end photo sharing/printing site. Which is hardly *that* vapid – nobody would bat an eye at that if it were done by Kodak or Staples.

So although meat-and-potato business/software companies do outnumber really high-tech or altruistic ventures, there's not a lot of evidence for silly Juicero-style startups being much of the Silicon Valley business community at all. So how come everyone thinks that they are?

Here's my theory. If you're an average well-off person, leading your average well-off life, consuming average well-off media and seeing ads targeted at the average well-off demographic, and going over to your average well-off friends' houses and seeing their average well-off products, which are you more likely to hear about? A structured-light optical engine for cytological research? Or a juicer?

Or to put it another way: there's a chapter in *Unsong* (spoiler!) where an archangel brings peace to the Middle East by splitting the Holy Land into two parallel dimensions. Any Jew who enters will find themselves in a united Israel;

any Muslim who enters will find themselves in an independent Palestine.

And sometimes I wonder if the same archangel has gotten to Silicon Valley.

If a deeply good person crusading for a better world enters Silicon Valley, she'll find herself surrounded by deeply good people crusading for a better world. She'll see mobile apps that track tropical diseases, clean energy startups that fight global warming by directly sucking carbon dioxide out of the air, companies bringing microbanking to poor Nepalese villagers, and boutique pharmaceutical labs searching for cures for orphan diseases.

If a futurist enters Silicon Valley, she'll find herself surrounded by futurists. She'll see neural nets and deep learning, reusable rockets and flying cars, high-throughput genome sequencing and CRISPR, metamaterials and nanotechnology.

If a social-media-obsessed narcissist whose view of the world begins and ends with his own Instagram page enters Silicon Valley, he'll find himself surrounded by social-media-obsessed narcissists whose view of the world begins and ends with their Instagram pages. He'll see a bunch of streaming video services and Uber-for-hair-products apps and elite pay-to-play dating scams and people trying to disrupt the gymwear market.

And if one of those people who talks about "the cloud" all the time enters Silicon Valley, he'll find himself surrounded by people who talk about "the cloud" all the time. I have no idea who these people are or what they're doing, but they all seem really happy with each other and I'm glad they're enjoying themselves.

They’ll all have their blind-men-and-elephant view of what kinds of things Silicon Valley “does”. And they’ll all be sort of right.

(thinkpiece writers: “Can you believe that Silicon Valley only makes products for shallow elites obsessed with the latest fads? It’s *the strangest thing!*“)

So I would recommend people stop talking about how Silicon Valley only makes ridiculous overpriced juicers. It’s not that it doesn’t make those. It does, just like everywhere else. A Facebook friend pointed out that QVC has been selling our parents ridiculous overpriced kitchen items since before we were born. Billy Mays pitched the EZ Crunch Bowl, which promised to “revolutionize your cereal-eating experience”. The unique thing about Silicon Valley isn’t that it’s got overpriced status goods designed to separate rich people from their money. The unique thing about Silicon Valley is that it’s got *anything else*.

I don’t want to downplay the problem. Anything remotely good in the world gets invaded by rent-seeking parasites and empty suits. Silicon Valley is no exception, and raising awareness of the infestation is certainly a public service. But for some reason, it’s hard for me to believe that – let’s say Deadspin – really believes in the spirit of Silicon Valley, really thinks that there was once somewhere that weird nerdy people could get together and produce amazing things for the good of everybody, and that to some degree this is still going on, and is a precious thing that needs to be protected. At its worst, some of their criticism sounds more like a worry that there might still be some weird nerds who think they can climb out of the crab-bucket, and they need to be beaten into submission by empty suits before they can get away. Or maybe that’s just paranoia. Fine, I admit I’m paranoid. But I still feel like people should lay off the criticism a little.

When Capitol Hill screws up, tens of thousands of innocent Iraqis get killed.

When Wall Street screws up, the country is plunged into recession and poor families lose their homes.

When Silicon Valley screws up, people who want a pointless Wi-Fi enabled juicer get a pointless Wi-Fi enabled juicer. Which by all accounts makes pretty good juice.

SSC Meetups: Berkeley, San Diego, Irvine

I’m in California for the next few weeks and want to catch up with people I know. I’ve got some meetups planned, but these are all preliminary, so check back there before the meetup to see if I’ve edited the time and date:

1. Unsong wrap party in Berkeley on Sunday, May 14, at 4:30, in the Cal Alumni Student Association. See here for more details.

2. SSC Berkeley Meetup on Tuesday, May 16th at 7PM at the 7th floor of 2030 Addison St. I am a little worried about space here, so be prepared for it to be cramped or us to have to relocate or something.

3. San Diego meetup on Monday, May 22nd, 7:30PM at Lestat’s Hillcrest coffee, 1041 University Ave. Psychiatrists in town for APA welcome!

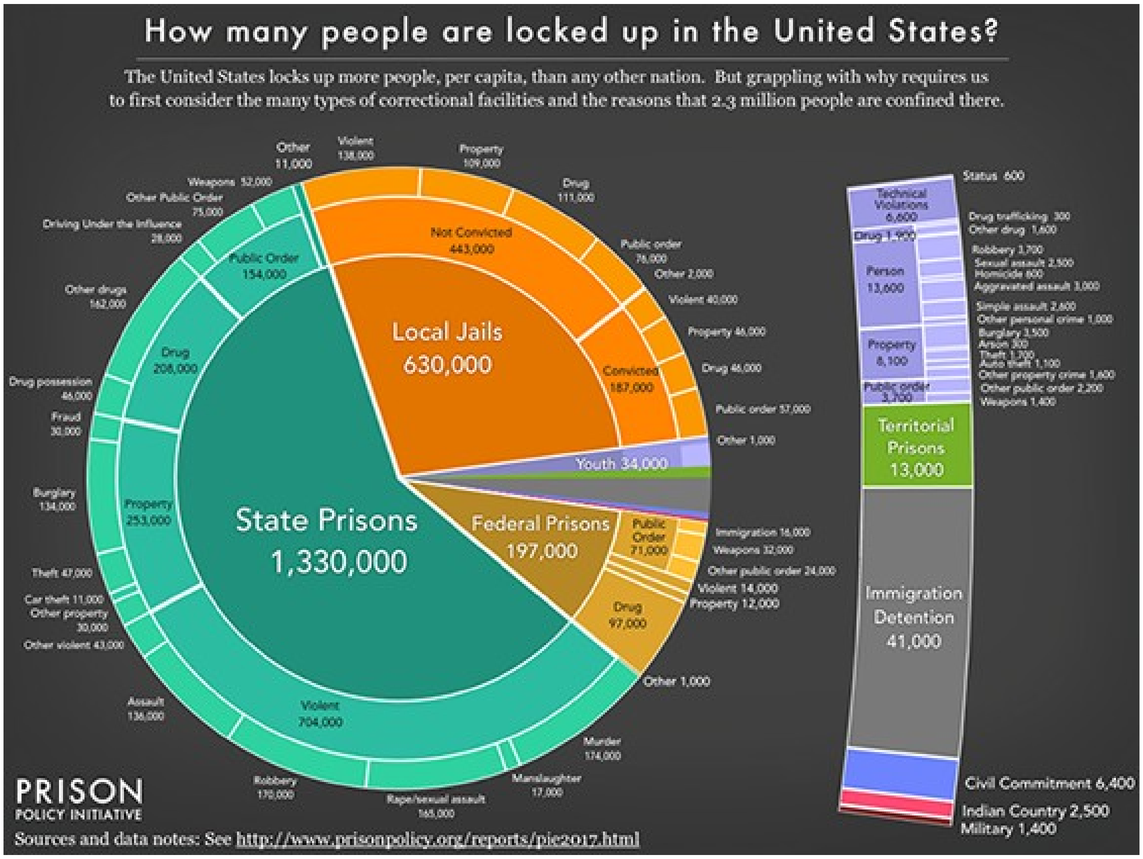
4. Irvine meetup on Saturday, May 27th, 7PM at Peet’s Coffee at 4213 Campus Dr.

And if anyone’s at the American Psychiatric Association conference, say hi to me. I’ll be presenting a poster on patient ratings of antidepressants at Resident Poster Competition 3 at 10 AM Sunday, plus wandering around the rest of the week. I’ll probably be in a black hospital jacket with my real name (Scott ______) on it.

Bail Out

Nobody likes that the US has the highest (second-highest after Seychelles?) incarceration rate in the world. But attempts to do something about it tend to founder on questions like “So, who do you want to release? The robbers, or the murderers?” Ending the drug war would be a marginal improvement but wouldn’t solve the problem on its own. And I’ve had trouble finding other ideas that engage with the reality that people are going to prioritize safety over reform and anything that significantly increases violent crime is a likely non-starter.

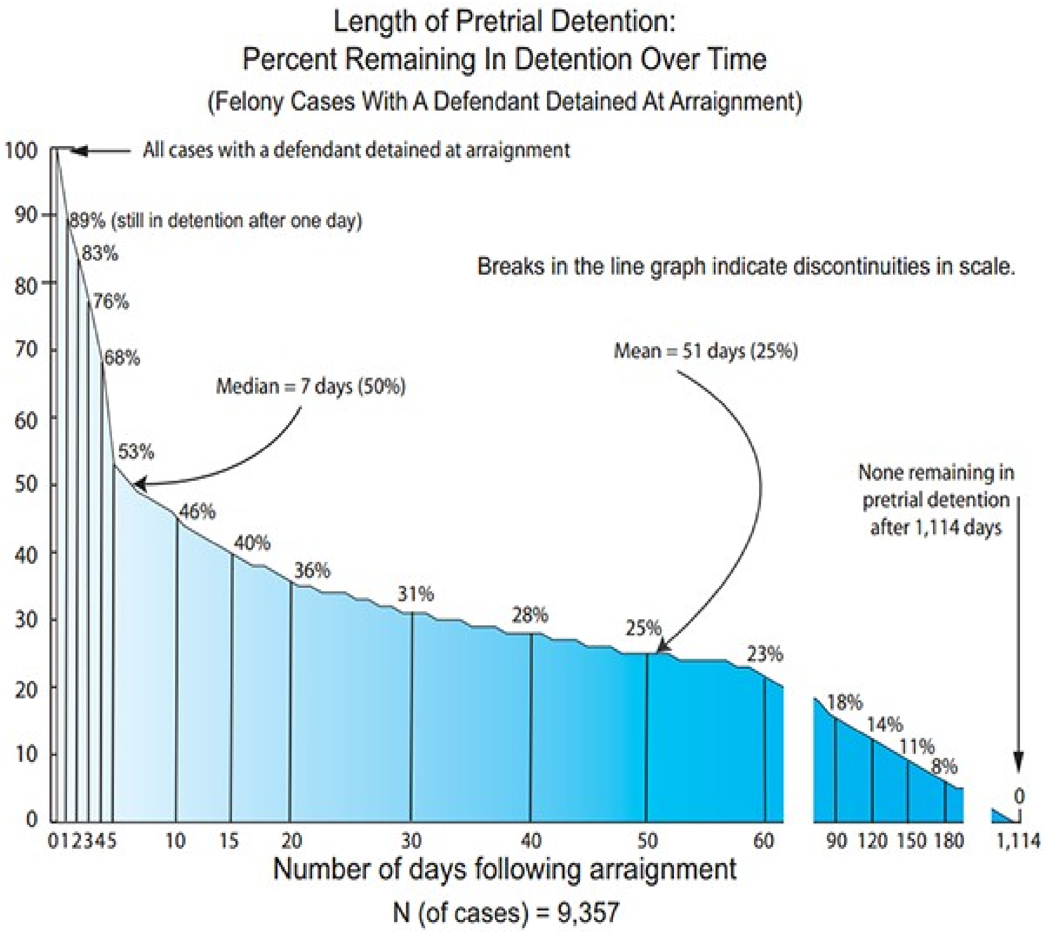
This was why I was interested to read the scattered thoughts in the effective-altruism-sphere about bail reform.



Sources and data notes: See http://www.prisonpolicy.org/reports/pie2017.html

(source)

About a fifth of the incarcerated population – the top of the orange slice, in this graph – are listed as “not convicted”. These are mostly people who haven’t gotten bail. Some are too much of a risk. But about 40% just can’t afford to pay. They are stuck in jail until their trial, which could take a long time:



Length of time that defendants spend in jail before trial (source). Although 50% are out by day 7, 25% are still in by day 50, 10% are still in by day 150, and one was still in after three years

I talked to a correctional psychiatrist a few weeks ago who was telling me that conditions for inmates awaiting trial were worse than conditions for already-convicted inmates. The convicted inmates get officially integrated into whatever prison they're in and receive various jobs and privileges. The inmates awaiting trial just sit in their cells doing nothing.

People who commit serious crimes might be looking at years or decades in prison. Do the few months they gain or lose because of bail really make that big a difference?

Yes. First, because most people aren't looking at decades in prison. This article tells the story of a man accused of attacking some police officers; he claimed innocence and expected to be vindicated at trial. Prosecutors offered him a plea bargain of sixty days in jail, which he refused. But he ended up spending more than sixty days in jail waiting for trial, which kind of defeated the point.

Second, because much of the time this ends in people just taking the plea bargain. For example, the man in the article above almost took the plea bargain after serving sixty days in jail – an understandable choice, since it let him walk free immediately with time served. But this would put a guilty plea on his record, which would make it harder for him to get jobs in the future and reframe any further crimes he might commit as "repeat offenses". If his case goes to trial, he might have been be found not guilty and avoid the black mark.

Third, because people who are detained pretrial end up getting longer sentences. Some of this seems to be because they're less able to contact their lawyer and prepare a good defense. Some more seems to be related to prosecutors setting harsher plea bargains for imprisoned defendants because they have a worse bargaining position. And some more might be related to psychological factors where judges think of people who just showed up from jail as "more criminal" than someone who came to the court from their home.

This legal advice site advises suspects not to stay in jail before trial even if they don't mind the environment and just want to get it over with. It confirms my friend's suspicion that jails are worse for people awaiting trial than for convicted offenders, but also notes some other interesting aspects. People in jail have a bad habit of making incriminating statements that get reported and used against them on trial. Suspects out on bail can rack up prosocial accomplishments to list off at their trial – they give the example of going to a counselor and making restitution to victims. And they get the option to delay their case until the trail grows cold and prosecutors get bored and everyone just agrees to a lesser sentence.

(something I didn't realize which is relevant here: 96% of felony convictions involve guilty pleas. Plea bargaining is the rule, not the exception, and anything which makes it easier or harder is going to impact the large majority of cases)

There have been a bunch of studies trying to determine to what degree bail vs. pretrial detention affects case outcome. Some early studies like this found that it could alter sentence length by about ten percent, but the data were necessarily not very good – there's no way to randomize suspects, and the sort of hardened criminals who don't get bail are the same sort of hardened criminals who maybe deserve tougher sentences. People tried their best to control for all observable factors, but this never works. Better is Gupta, Hansman & Frenchman, which tries a quasi-experimental design based on suspects' random assignment to more or less strict judges who are more or less likely to demand lots of money for bail. Suspects who are assessed bail are 6% more likely to be convicted (they're also 4% more likely to reoffend, which is weird but might be related to imprisonment hardening criminals and increasing recidivism). Stevenson tries a similar method and finds that inability to make bail produces a 13% increase in convictions, mostly through more guilty pleas.

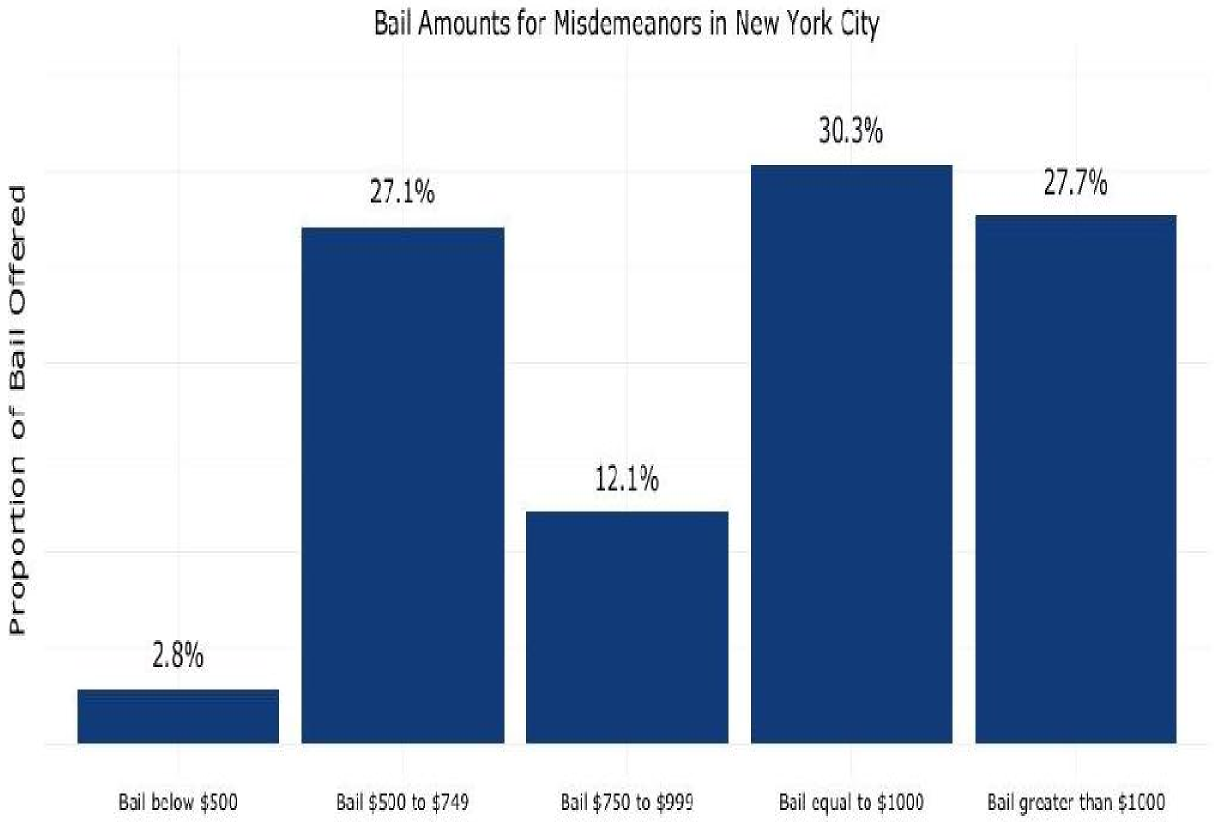
This study also finds that size of bail was less important than whether there was bail at all, which confused me until they pointed out that most suspects are really, really poor. As per the Stevenson paper:

> Some of these defendants are facing very serious charges, and accordingly, have very high bail. But many have bail set at amounts that would be affordable for the middle or upper-middle class but are simply beyond the reach of the poor. In Philadelphia, the site of this study, more than half of pretrial detainees would be able to secure their release by paying a deposit of $1000 or less, most of which would be reimbursed if they appear at all court dates. Many defendants remain incarcerated even at extremely low amounts of bail, where the deposit necessary to secure release is only $50 or $100.

The New York Times notes that

> Even when bail is set comparatively low – at $500 or less, as it is in one-third of nonfelony cases = only 15% of defendants are able to come up with the money to avoid jail.

And from here:



The guy in the NPR article above – the one who might or might not have attacked some cops – was in jail for 60 days because he couldn't make bail of $1000.

So bail causes people to be stuck in jail for months, increases guilty pleas independent of defendants' actual guilt, and causes more convictions and longer sentences. Since many of the people harmed by this are innocent, or deserve less punishment than they end up receiving, this seems like an important point of leverage at which to try to fight incarceration. On the other hand, bail is supposed to serve a useful purpose in preventing suspects from running away. It is possible to do without?

Maybe. Washington DC is one of the highest-crime areas of the country, but it uses an alternative system without monetary bail which uses an algorithm to calculate risk, releases low-risk people, and keeps high-risk people imprisoned without bail. It looks like about 10% of Washingtonians, versus 47% of other Americans, are detained in jail pre-trial. Of Washingtonians released without bail, 87% show up for all court dates, and 98% avoid committing violent crimes while free.

I'm not able to find a good comparison between Washington and other jurisdictions, but this page on bounty hunters gives the unsourced statistic that 20% of people on bail fail to show up for court anyway. This equally-credible-looking site gives an equally unsourced statistic of 10%, and notes that these people are more likely to be flakes who forgot their court date than supercriminals who have donned a fake mustache and are on their way to the Cayman Islands. These numbers suggests that Washington's no-show rate is about the same as everywhere else's.

Bronx Freedom Fund is a charity that helps people pay bail. They claim that 96% of the people whose bail they pay show up to trial, which matches the numbers from Washington and the numbers from normal places where people have to pay their own bail. I am sure they select their beneficiaries very carefully, but if it's possible to select a group of people who are definitely going to show up to their trial whether they can make bail or not, *why are those people still in jail?*

I can't find any studies clearly proving this, but it looks like there's no obvious and proven tendency for bail to improve show-up-at-trial rates, and some evidence that with good risk assessment and selection 90% of people released without bail will show up to trial, which may be around the same as the rate for people who *do* get bail. This would make all of the problems with bail not only outrageous but pointless, not even the unfortunate side effects of a generally-reasonable policy.

It looks like there are two possible solutions – one small and short-term, the other systemic and long-term.

In the short term, there are some great charities – like the aforementioned Bronx Freedom Fund – that pay people's bail. If the people show up to trial (which, again, 90% or so do), then the charity gets its money back and can use it to pay more people's bail, ad infinitum. This makes them very high-value per dollar.

I'm not sure how much to trust BFF's self-reported statistics. I see that on their front page they mention that "defendants who await trial in jail are 4x more likely to be sentenced to time in prison", which is much more than any of the studies above report and which is probably a misleading uncorrected raw estimate. I see they say that "$39 can help us secure someone's freedom", which since average bail is something like $500 and only 2.8% of bails are less than $50 might be more of a "this is the lowest bail that has ever occurred in history" than any kind of representative estimate. But I don't really hold it against them to give technically correct but misleading numbers in an advertising pitch. The question is – what are the real numbers? GiveWell and Open Philanthropy Project have been looking at them, but I can't find a complete writeup, so let me do a terrible and very approximate job of trying to estimate them myself.

They note that their average bail is $790. Supposing that their 96% success rate is true, the average case costs them $30 – they cite a similar figure in their own literature. If, as they say, the average pretrial detention period is 15 days (which more or less matches the chart above), they can keep someone out of jail for $2/day even without selecting the worst cases.

They also note that their clients get cleared of all charges about half the time, compared to only 10% of the time for people in jail. I don't know how much of this is selection effects, but suppose we take it at face value. And suppose that an average prison sentence for whatever kinds of petty crimes these people commit is two months. That means they're saving the average suspect about 25 days worth of prison sentence, and improves their cost-effectiveness. I'd say this comes out to about $0.75 per jail day prevented, but that might be double-counting some people who, had they been in jail, would have had the time deducted from their sentence. Let's say somewhere between $0.75 and $2.

A confusion: their 2014 annual report notes that they received $53,000 in donations and helped 140 people, suggesting a cost of $380 (not $30) per person helped. In 2015 they helped 160 people with $120,000, suggesting a cost of $750 per person that doesn't seem to improve with scale. A similar organization, Brooklyn Community Bail Fund, got $404,800 from Good Ventures, and boasts of helping 1,100 people, for $367/person (this is a very rough estimate; it assumes this is all in the same year and they got no other donations).

I'm not sure why these numbers are so much bigger than the ones derived from first principles, but taken seriously, if we continue to assume that each successful case saves an average of forty days in jail, they prevent a jail day for about $10.

We compare to other efficient charities. GiveWell thinks it takes antimalaria charities about $7500 to save a human life. If that person lives another 20 years, that's about $1/day.

So using their own numbers, these bail-related charities might be able to prevent imprisonment at the same cost-per-day as other charities can save lives. Using less optimistic numbers, they might be able to prevent imprisonment at about ten times the rate. Depending on where you fall on the death vs suffering philosophical tradeoff, this might be attractive. I’m confused these charities haven’t received more attention, if only to debunk them properly.

And even though everyone likes to talk about how individual charity doesn’t work and only systemic change can make a real difference, paying the bail of 50% of the people in the United States currently awaiting trial would be well within the abilities of one philanthropic-feeling billionaire, after which the money could be recycled again and again for the same purpose. I agree this is a kind of hokey calculation since it doesn’t include the overhead cost of getting the money to the suspects, but it’s still shocking how small the amounts involved are compared to other social problems.

In the long-term, we probably need some kind of criminal justice reform. The Obama administration seemed to be pursuing some kind of fourteenth amendment argument to get a court ruling that monetary bail was illegal. Some libertarian groups like Reason and the Cato Institute are on board. Some charities, like Equal Justice Under Law, are supporting the same cause.

I don’t think that bail reform is the largest or most important change that could be made to the US criminal justice system. But I think it might be tied with drug reform for the easiest. And given how well politics has been going lately, this might be a good time for lowered ambitions.

Polyamory Is Not Polygyny

[Content warning: polyamory, brief quote of weird Heartiste stuff]

The objections I hear to polyamory tend to separate into two narratives sharing a common thread.

The first narrative is supposedly concerned about women, and typified by National Review's Polyamory Is A Modern Name For A Backward Practice. It asks:

> What happens to women in a world where we scrap the "binary axis" of monogamy? Women suffer, that's what. Nobody is asking for a show called "Brother Husbands." Nine of ten pictures for polyamory involve one man with multiple women. The other one in ten is usually just a crowd of people. Men may sleep around, but they don't tolerate the degradation of being a part of a modern male harem, nor have they ever, really. Polygamy uniquely subjugates one sex; it's like an institutionalized form of the hookup culture — with women on call for male pleasure, just with some boundaries and a relationship status.

The second narrative is supposedly concerned about men, and typified by Heartiste's Polyamory Is Disguised Polygamy:

> Polyamory — multiple and simultaneous sexual relationships — means, in practice, a few high value dudes hording all the pussy. Multitudinously and concurrently. Polyamory cheerleaders, like Christopher Ryan, note the shape of our penis heads and go on to weave a happy utopia of free love where all the men and all the women get their rocks off whenever and however they wish, like the bonobos (who, by the way,

> are territorially squeezed compared to their more prodigiously successful chimp cousins). But he has to ignore female hypergamous mate choice and male jealousy to concoct this vision of a peaceful hedonist paradise. The reality would be considerably darker; women would still want to bang the alpha, leaving the beta male out in the cold, clawing and scratching for rode-worn scraps, but now shackled with the obligation to help provide for kids that are likely not his own.

Despite the different focuses, they both have the same theory. Men – especially high-status men – are going to date lots of women. But women aren't going to date lots of men, so all the women will end up dating the same few high-status men and ignore the low-status men. Therefore, women (NRO's concern) and low-status men (Heartiste's concern) will lose out.

I got so tired of trying to explain that this doesn't match reality that I started digging back in old survey data to see if I could just disprove it. The latest SSC survey didn't have enough questions on relationships, but the 2014 LW survey did. I got a sample of 53 poly women, 164 poly men, 70 monogamous women, and 690 monogamous men.

I interpret NRO and Heartiste's theories to predict that more poly men than mono men would be single, that the median poly woman would have more partners than the median poly man, but that there would be more poly men with very high numbers of partners than poly women with the same. None of these hypotheses were confirmed.

For poly men, 29% were single, 47% had one partner, 17% had two partners, 4% had three, 2% had four, and only 0.5% had five or more.

For poly women, 8% were single, 44% had one partner, 23% had two partners, 15% had three partners, 8% had four, and 4% had five or more.

For both sexes, the median person had one partner. But the average number of partners was higher for women, and there were more women with very high numbers of partners than men with the same.

Poly men were more likely than poly women to be single. However, poly men were still less likely to be single than mono men. 45% of the mono men in the sample were single, suggesting polyamory doesn’t hurt low-status men’s chances of getting a date.

This sample is pretty skewed since it has three times more poly men than poly women. This at least partly corresponds to there being many more men than women in the community it was sampling. Poly men might either date women from outside the community, or have one poly woman date multiple poly men in order to even the odds. I think this second factor probably explains some of poly women’s higher number of partners.

There’s another possible skew: I’m not sure how people decided to identify as poly or monogamous (the question itself asked whether you “prefer polyamory” or “prefer monogamy”). If single people defaulted to monogamy, and some people only claimed to be polyamorous insofar as they were actually dating somebody, that might skew the percent of single people in each style. People who said they were “unsure” whether they were poly or mono were more likely to be single than people with either style (70% of unsure men and 58% of unsure women).

This doesn’t seem compatible with NRO and Heartiste’s theory, but it’s also not great data. If some supporter of theirs wants to tell me what I have to do in the next SSC survey to get results that they’ll be willing to believe, then let’s talk.

[**EDIT:** Many people are pointing out I’m looking at actually-existing-polyamory, not polyamory as it would be practiced if it hypothetically took over all of society. But actually-existing-polyamory is the thing at issue here, and the practice that has to defend itself. I consider the idea of polyamory taking over all of society maybe *somewhat* more probable than the idea of homosexuality or transgender doing so, but not probable enough to be very likely.]

Postmarketing Surveillance Is Good And Normal

I.

Scientific American notes a recent study saying that a third of drugs approved by the FDA over the past ten years have since been recalled, been given new boxed warnings, or been given new “safety communications”.

A few people have asked me whether this means the FDA is too lax and needs to tighten its standards. Let’s look at this in more detail.

What the study actually says: of 222 drugs approved by the FDA in the last ten years, only 3 (1.3%) were taken off the market. Another 30% or so received “boxed warnings” or “safety communications”, basically the FDA’s way of adding new rules to their safety information. For example, the FDA might approve a drug for the general population, then issue a boxed warning saying “actually, we noticed that this drug can cause seizures as a side effect, don’t take it if you have a history of seizure disorder”.

The most serious category are the drugs taken off the market. As mentioned above, these are about 1% of the total. This doesn’t sound like the story of a weak regulatory agency failing to do its job. This sounds like a really impressive success rate. In fact, if our standards are so stringent that we’re insisting on a 1% false positive rate, I’m kind of horrified thinking of all the false negatives we must be throwing out.

But doesn’t even a 1% false positive rate mean the FDA failed in some way?

No. Let’s consider the most recent psychiatric drug to be withdrawn. This is nefazodone, an antidepressant withdrawn after the discovery that it causes liver failure once every 300,000 patient-years. That is, if 300,000 patients took it for one year, there would be one extra case of liver failure.

How, exactly, do you want to discover this in pre-approval studies? An average drug study has maybe 500 patients. Do you want to run an average drug study for six hundred years? Or do you want to figure out how to run a drug study that’s six hundred times bigger than average? And these numbers are underestimates – one extra liver failure might be a coincidence. You’d want to see two or three extra liver failures before you start thinking the drug might be involved. The average clinical trial costs something like $50 million. Are you sure you want to multiply this number by six hundred to catch a side effect that will affect one out of hundreds of thousands of people?

So what we actually do is post-marketing surveillance. The FDA demands an average drug study on 500 people to make sure that there aren’t any *common* problems. Then over hundreds of thousands of patients over the space of decades (nefazodone was out almost ten years before it got withdrawn) people collect records and see if there’s any unusual disorder that happens more often when people are on the drug.

Here’s another example from earlier this year: the FDA issued a safety communication that a new drug called Viberzi can cause pancreatitis in people who don’t have a gall bladder. I’m not sure how many patients in the original approval study didn’t have gall bladders, but I bet it wasn’t enough to draw any useful conclusions from. Should all drugs be delayed until they can do separate studies in the gall-bladder-less population? And how would we know beforehand that gall bladders were the problem? Why not separate studies in people with one kidney? People taking antidepressants? Red-haired people? Chinese people? Don't laugh, Tegretol has a fatal side effect that's only been observed in Han Chinese.

And then you do the kidney study, the antidepressant study, the red-haired study, and the Chinese study, and darnit, you forgot to look at people who eat way more sauerkraut than any normal person. There you go, linezolid just killed your patient.

You are never going to be able to figure out everything pre-approval. At best, you can prove that the drug is reasonably safe for the vast majority of people. Then later you find something that only happens once in a hundred thousand years, or only to people without gall bladders, or only to Han Chinese, or only if you eat too much sauerkraut. And then the FDA issues a safety communication about it. That's what it looks like when the system works.

II.

Except I want to look a little further into FDA safety communications and boxed warnings, the large majority of the events found in the study. Some of these are important updates about things like the drug being dangerous to people without gallbladders. Others are…well, the FDA *really* likes warning people about stuff.

This February, the FDA issued a safety communication about chlorhexidine, aka antiseptic soap. This has been used since the 1950s by loads of surgeons, doctors, and random people who need antiseptic soap for something, and it's currently a WHO Essential Medication. The FDA wanted us to know that, during fifty years of worldwide use of this product, about one person a year had experienced a severe allergic reaction (by comparison, there are 200 deaths a year from peanut allergies). The safety communication said that if you found yourself having a severe allergic reaction to antiseptic soap, you should call 911.

Last summer, the FDA added a boxed warning to all benzodiazepines (Valium, Xanax, Ativan, Librium, etc) and all opiates (morphine, Norco, Percocet, Vicodin, Oxycontin, etc) warning doctors that it could be dangerous to prescribe those two classes of drugs together. Doctors, who had been warned against prescribing those two classes of drugs together for fifty years, collectively said “well, duh”, and then continued prescribing those two classes of drugs together the same as always, because dealing with anxious people who have chronic pain is *really hard*.

In 2006, the FDA added a boxed warning to warfarin, saying it could cause major bleeding. Warfarin is a 50-year old medication taken by millions of people each year, which has probably saved hundreds of thousands of lives over the past half-century. It’s an anticoagulant, which means the whole point is to make your blood clot less and bleed more. Mentioning that warfarin can cause major bleeding is a lot like mentioning that sleeping pills might cause tiredness, or weight gain pills could make you fat. Nevertheless, after fifty years and tens of millions of patients, the FDA decided to issue a boxed warning about this. The medical community collectively say “Well, duh” again and got on with their lives.

Also in 2006, the FDA added a boxed warning to Ritalin, saying it might slightly increase the risk of heart disease in children. Everyone kept prescribing it anyway, and later on some better studies showed that it might *not* slightly increase the risk of heart disease in children. I’m not sure what the current status of this debate is, but it sure hasn't stopped like half the children I meet from being on Ritalin.

In 2004, the FDA added a boxed warning to every single antidepressant – yes, every single one – warning that they might increase suicide risk in teens. There is still a heated debate about this, with some recent review articles seeming to confirm, and other people pointing out that, when the FDA warning discouraged people from giving antidepressants to teens, teen suicide attempts suddenly went way up. Anyway, antidepressants are hardly alone here – other psychiatric drugs that received boxed warnings include all typical antipsychotics, all atypical antipsychotics, all benzodiazepines, all stimulants, lithium, Depakote, Lamictal, and I think literally every single psychiatric drug except buspirone.

What I'm saying is – the FDA issuing a safety communication or boxed warning doesn't mean the drug was a mistake, or you should be scared that something went wrong. It's a routine part of the pharmacological monitoring system. This doesn't mean it should feel routine to *your doctor* – they should get worried every time a new one comes out and make sure they're not inadvertently harming their patients without gall bladders – but it should feel routine to somebody looking at this on the institutional/systems level.

III.

So does this study mean that the FDA is too lax and needs to tighten its standards?

I'm not sure. Maybe the best answer is "not *necessarily*". We definitely shouldn't be aiming for a 0% post-marketing event rate. Is a 33% post-marketing event rate too high?

I’ve seen a lot of discussion on this recently which I think takes shortcuts. It points out that the FDA has a higher/lower post-marketing event rate than European agencies. Or that the FDA takes longer/shorter to approve drugs than it did a couple of years ago. Or that its standards are stricter/looser than some comparable area of the federal bureaucracy.

None of this matters. What actually matters is the number of people helped by incentivizing new drug development and getting it to market quickly, versus the number of people harmed by the safety problems that slip through the cracks.

It looks like some of the post-marketing surveillance events here were pretty silly, while others were pretty serious – two people died from that drug that causes pancreatitis in people without gall bladders. Are those two deaths justified in the context of saving thousands of other people who had whatever condition that drug cures? I don’t know without a lot more work. The only study I’ve ever seen on this is stuff in the vein of Isakov, Lo, and Motazerhodjat, which always finds the FDA is too conservative. Maybe they’re wrong, but if someone wants to prove they’re wrong, they should do the same kind of cost-benefit analysis and let us know that they came to a different result.

The finding that 33% of approved drugs get post-marketing safety events may factor into such a calculation. But without the rest of the calculation, it’s just a meaningless number.

Classified Ads Thread

I’m away at a conference, so no blogging today.

Instead, let’s try having people post ads. Jobs, housing, products, websites, etc.

(Personal ads okay from women and gay people; given the demographics here I don’t think it’s worth it for straight men)

Those Modern Pathologies

[Related to: The Fidget Spinner Is The Perfect Toy For The Trump Presidency, In Defense Of Liking Things, Open Marriage Is A Neoliberal Pathology]

That modern pathology, the Pyramid of Cheops

The final triumph of modern individualism is an afterlife ensconed in a giant stone structure, carefully segregated from any other souls, based entirely around *stuff*. No county churchyards here. No slow surrender to nature and the weeds. Just piles of golden goblets and jeweled necklaces, carefully guarded by snake-infested traps. And, of course the bones of dead servants, guaranteed to keep serving you in the great beyond. Of *course* Heaven is neoliberal. There is no alternative!

That modern pathology, heterosexual intercourse for the sole purpose of procreation

Sex can bring people together. It can cement relationships between people and families. It can pulse with celestial fire, it can shatter inner worlds, it can inspire transcendent art, it can remake souls. Of course moderns took one look at all of that and thought: you know what the only acceptable purpose of sex is? *Making a smaller copy of myself*.

But calling this narcissism would be missing half the picture. It's equally related to a sort of productivity fetish, a mindset where anything that doesn't leave a material token didn't really happen. Enjoyed the company of your closest friends? Not real unless you put the pictures on Facebook, tagged #bestiesforever. Broadened your horizons with a trip to another culture? Not real without crushed pennies or some other gift-shop tchotchke. Met your soulmate? Not real unless you've got a lump of screaming flesh to show for it.

This is what capitalism does – reduce experiences to souveniers, reduce relationships to commodities, demand that everything good be mediated by a material end product in order to model the laboring-for-others that workers are told is their only life purpose.

That modern pathology, Homer’s Odyssey

If Harry Potter wasn’t vapid enough for you, now we have a travelogue for the Instagram generation.

Odysseus’ only salient characteristic is being “polymetis”, Very Smart. This is enough to give him a raving fan club of front-row-kids and aspirational Ivy Leaguers, the same people who thought Hermione Granger’s straight A’s made her a symbol of an entire generation of womenhood. Odysseus proves his chops in his very first adventure, where he encounters Lotus-Eaters who convince most of his men to eat a magic fruit that leaves them drugged and listless; Odysseus nobly drags them back to the ship and forces them to keep on rowing for him.

Imagine the horrors of a world where poor galley slaves can leave behind their unpaid labor to live on a tropical beach and enjoy their lives! It is only thanks to Odysseus that this catastrophe is averted. One might think a few readers would note that a few months later, the vast majority of sailors in Odysseus’ fleet died horribly, eaten by cannibals. One might think a few readers would wonder if, really, the guy who dragged galley slaves back to their galleys only to get them killed a few months later was really such a good guy. In fact, nobody asks this question, because Odysseus is Very Smart. It’s no coincidence that the Odyssey came out in the generation of the invasion of Iraq War – if Very Smart people declare that dying horribly was the right thing to do, and then it turns out it wasn't, at least they were benevolent technocrats with your best interests in mind.

Odysseus then goes on to have sex with various sorceresses and sea-nymphs while protesting that he doesn't want to have sex with them and is loyal to his far-off wife. This is portrayed as clearly a difficult problem that we should empathize with. Also, his sailors get turned into pigs, eaten by sea monsters, and drowned in a giant whirlpool. This is not portrayed as clearly a difficult problem that we should empathize with. In one scene, some starving sailors eat a sacred cow belonging to the Sun God; this is portrayed as clearly justifying their deaths.

We can start to sketch a psychological picture of the sort of person who could enjoy the Odyssey. They identify with Odysseus, that's obvious. They want to feel like they've suffered – after all, suffering is ennobling! – but they don't want to actually suffer. They imagine the "suffering" of having to have sex with lots of sea nymphs they're not super-interested in, all while their friends and subordinates are massacred all around them (but only for good reasons, like them stealing cattle, or them not being Very Smart). At the end of all of it, much like the rich kids attending the Fyre Festival, they can show up on their front doorstep and say "Oh, what suffering I have seen – and I the only survivor!"

The Odyssey is a book for rich individualist aspirational Very Smart narcissists who simultaneously want to outsource their ennobling hardships to the lower classes, *and* remain so contemptuous of those lower classes that they imagine them literally getting turned into swine by a sorceress, *and* end up having sex with that sorceress, who is unable to resist them because they are Very Smart.

I weep for the modern generation.

That modern pathology, the Aristotelian theory of virtue

You see, perfect virtue in all things approaches the mean. The traditionalist who wants to make the system more conservative is unvirtuous. And the radical who wants to make the system more progressive is exactly equally unvirtuous. The virtuous person is the liberal intellectual who considers both positions, then places himself exactly in the middle.

Anybody who seems too fiery, too deep, or too sincere is automatically wrong. You can reject both the grandparents who urge clean and sober living, and the hippies who tell you that drugs are the only way to break outside the system and achieve true consciousness, in favor of having a joint or two whenever you feel like it. You can reject both the ascetic who urges simple water, and the aesthete who urges fine wine, in favor of the bottle of Diet Coke already in your fridge.

Aristotle reduces virtue to abandoning the highs of ecstasy and the lows of misery in favor of the comfortable neoliberal plateau of watching *Mad Men* on TV and ordering off Amazon Echo. No wonder his message resonates so well with millennials.

That modern pathology, Catholicism

The Old Testament God demanded adherence to hundreds of rules and rained down collective punishment on entire nations for breaking them. But you are a yuppie with an hour a week for religion, tops, and consider yourself part of a different species from anyone who doesn’t go to a Starbucks at least twice a week. You get your coffee in packets from Keurig, your razors in packets from Dollar

Shave Club, and your juice in packets from Juicero. If only there were some large corporation that would package religion and send it to your doorstep for one low price.

Enter Catholicism. God loves you, *just for being you*. He suffered and died for you two thousand years ago, granting you redemption. All you have to do to pick it up is sign on the dotted line and pay ten percent of your income.

Consider eg the ritual of confession. You eliminate your sin in a standardized dyadic interaction no harder than eliminating your muscle tension at the massage parlor, or eliminating your back pain with a chiropractor. It's quick, impersonal, and completely tailored to your individual sinner profile.

Or consider the Eucharist. The Prozac generation has already had personal change reduced to the process of swallowing a pill, and the Catholic Church is eager to comply, reducing finitude to a DSM-V ailment curable with correctly prepared bread products. The Church is a corporation the same as Coca-Cola and McDonalds, and we already know the sacred ritual for interacting with corporations. And so our consumer culture reduces the human relationship with the Divine to *literally consuming God*.

And like all good corporations, you can rest assured that the whole thing is organized in a very logical top-down chain under the absolute command of an incredibly rich guy who lives in a house covered in gold. “Father, am I forgiven now?” “Um, one second, let me check with the manager in branch headquarters”. And why shouldn't he? The word “Catholic” means “universal”; we're so separated from our own neighbors that we'd rather our religion come in the form of Standardized Religion Product shipped in from Rome than anything which forces us to confront people near us as individuals, or trust our local communities for anything more than naming the parish church.

Let’s be honest: the recent success of Catholicism is the ultimate sign of our inability to deal with the world through anything other than a late capitalist lens of standardizaton, corporatism, and carefully-packaged pablum. It’s the perfect religion for the Age of Trump.

The Atomic Bomb Considered As Hungarian High School Science Fair Project

I.

A group of Manhattan Project physicists created a tongue-in-cheek mythology where superintelligent Martian scouts landed in Budapest in the late 19th century and stayed for about a generation, after which they decided the planet was unsuitable for their needs and disappeared. The only clue to their existence were the children they had with local women.

The joke was that this explained why the Manhattan Project was led by a group of Hungarian supergeniuses, all born in Budapest between 1890 and 1920. These included Manhattan Project founder Leo Szilard, H-bomb creator Edward Teller, Nobel-Prize-winning quantum physicist Eugene Wigner, and legendary polymath John von Neumann, namesake of the List Of Things Named After John Von Neumann.

The coincidences actually pile up beyond this. Von Neumann, Wigner, and possibly Teller all went to the same central Budapest high school at about the same time, leading a friend to joke about the atomic bomb being *basically* a Hungarian high school science fair project.

But maybe we shouldn't be joking about this so much. Suppose we learned that Beethoven, Mozart, and Bach all had the same childhood piano tutor. It sounds less like "ha ha, what a funny coincidence" and more like "wait, who was this guy, and how quickly can we make everyone else start doing what he did?”

In this case, the guy was Laszlo Ratz, legendary Budapest high school math teacher. I didn’t even know people *told* legends about high school math teachers, but apparently they do, and this guy features in a lot of them. There is apparently a Laszlo Ratz Memorial Congress for high school math teachers each year, and a Laszlo Ratz medal for services to the profession. There are plaques and statues to this guy. It’s pretty impressive.

A while ago I looked into the literature on teachers and concluded that they didn’t have much effect overall. Similarly, Freddie deBoer writes that most claims that certain schools or programs have transformative effects on their students are the result of selection bias.

On the other hand, we have a Hungarian academy producing like half the brainpower behind 20th century physics, and Nobel laureates who literally keep a picture of their high school math teacher on the wall of their office to inspire them. Perhaps even if teachers don’t explain much of the *existing* variability, there are heights of teacherdom so rare that they don’t show up in the statistics, but still exist to be aspired to?

II.

I’ve heard this argument a few times, and I think it’s wrong.

Yes, two of Ratz’s students went on to become supergeniuses. But Edward Teller, another supergenius, went to the same high school but (as far as I know) was never taught by Ratz himself. That suggests that the school was good at producing supergeniuses regarldess of Ratz’s personal qualities. A further point in support of this: John

Harsanyi also went to the school, also wasn't directly taught by Ratz, and also went on to win a Nobel Prize and invent various important fields of mathematics. So this school – the Fasori Gymnasium – seems to have been about equally excellent for both its Ratz-taught and its non-Ratz-taught pupils.

Yet the Fasori Gymnasium *might not have even been the best high school in its neighborhood*. It competed with the Minta Gymnasium half a mile down the street, whose alumni include Manhattan Project physicists Nicholas Kurti and Theodore von Karman (von Karman went on to found the Jet Propulsion Laboratory), brilliant chemist-philosopher Michael Polanyi, economists Thomas Balogh and Nicholas Kaldor (of Kaldor-Hicks efficiency fame), and Peter Lax, who once said "You don't have to be Hungarian to be a mathematician – but it helps". There are also some contradictory sources suggesting Teller attended this school and not Fasori; for all I know he might have attended both. Once again, most of these people were born in the 1890-1910 period when the Martian scouts were supposedly in Budapest.

Worse, I'm not even sure that the best high school in early 20th-century Hungary was *either* of the two mentioned above. The Berzsenyi Gymnasium, a two mile walk down Gyorgy Street from the others, boasts alumni including multizillionaire George Soros, Intel founder Andrew Grove, BASIC inventor John Kemeny, leading cancer biologist George Klein, great mathematician George Polya, and Nobel Prize winning physicist Dennis Gabor.

Given that the Fasori Gymnasium wasn't obviously better than either of these others, is it possible that the excellence was at a higher level – neither excellent teachers nor excellent principals, but some kind of generally excellent Hungarian culture of education?

This is definitely what the Hungarians want us to think. According to *Cultures of Creativity*:

> What's so special about Budapest's schools? A certain elitism and a spirit of competition partly explains the successes of their students. For example, annual competitions in mathematics and physics have been held since 1894. The instruction the students receive as well as these contests are an expression of a special pedagogy and a striving to encourage creativity. Mor Karman, founder of the Minta school, believed that everything should be taught by showing its relation to everyday life. Instead of learning rules by heart from books, students tried to formulate the rules themselves.

This paper on "The Hungarian Phenomenon" makes similar claims, but adds a few more details:

> The Eotvos Contests were a powerful mean for the stimulation of mathematics on a large scale and were used to motivate mathematical culture in the society. It also provided a channel to search for talented youths. The contests, which have been open to Hungarian high school students in their last year since 1894, played a remarkable role in the development of mathematics.

Okay. But I want to challenge this. During this era, formal education in Hungary began at age 10. By age ten, John von Neumann, greatest of the Hungarian supergeniuses, already spoke English, French, German, Italian, and Ancient Greek, knew integral and differential calculus, and could multiply and divide 8-digit numbers in his head. Wikipedia notes that on his first meeting with his math teacher, the math teacher "was so astounded with the boy's mathematical talent that he was brought to tears". This doesn't sound like a guy whose potential was kindled by formal education. This sounds like a guy who would have become one of history's great mathematicians even if his teachers had slept through his entire high school career.

Likewise, the book above notes that Dennis Gabor, the Hungarian inventor of holography, "developed his passion for physics during his youth, but did so for the most part on his own". His biography notes that "During his childhood in Budapest, Gabor and his brother would often duplicate the experiments they read about in scientific journals in their home laboratory."

Likewise, consider Paul Erdos, a brilliant mathematician born in Budapest around this time. As per his Wikipedia page, "Left to his own devices, he taught himself to read through mathematics texts that his parents left around their home. By the age of four, given a person's age, he could calculate, in his head, how many seconds they had lived."

I have no knock-down proof that Hungary's clearly excellent education system didn't contribute to this phenomenon. A lot of child prodigies burn out, and maybe Hungary was unusually good at making sure that didn't happen. But it sure seems like they had a lot of child prodigies to work with.

So what's going on? Should we just accept the Manhattan Project consensus that there was a superintelligent Martian scout force in early 20th-century Budapest?

III.

Here's something interesting: every single person I mentioned above is of Jewish descent. *Every single one*. This isn't some clever setup where I only selected Jewish-Hungarians in order to spring this on you later. I selected all the interesting Hungarians I could find, then went back and checked, and every one of them was Jewish.

This puts the excellence of the Hungarian education system in a different light. Hungarian schools totally failed to work their magic on Gentiles. You can talk all you want about “elitism and a spirit of competition” and “striving to encourage creativity”, yet for some reason this worked on exactly one of Hungary’s many ethnic groups.

This reduces the difficult question of Hungarian intellectual achievement to the easier question of Jewish intellectual achievement.

I say “easier question” because I find the solution by Cochran, Hardy, and Harpending really compelling. Their paper is called A Natural History Of Ashkenazi Intelligence (“Ashkenazi” means Eastern European Jew) and they start by expressing the extent of the issue:

> Ashkenazi Jews have the highest average IQ of any ethnic group for which there are reliable data. They score 0.75 to 1.0 standard deviations above the general European average, corresponding to an IQ 112 – 115. This fact has social significance because IQ (as measured by IQ tests) is the best predictor we have of success in academic subjects and most jobs. Ashkenazi Jews are just as successful as their tested IQ would predict, and they are hugely overrepresented in occupations and fields with the highest cognitive demands. During the 20th century, they made up about 3% of the US population but won 27% of the US Nobel science prizes and 25% of the Turing Awards [in computer science]. They account for more than half of world chess champions.

This doesn't seem to be due to any advantage in material privilege; Ashkenazi Jews frequently did well even in countries where they were persecuted. Nor is it obviously linked to Jewish culture; Jews from other regions of the world show no such advantage. So what's going on?

Doctors have long noted that Ashkenazi Jews are uniquely susceptible to various genetic diseases. For example, they're about a hundred times more likely to have Gaucher's Disease, a hundred times more likely to get Tay-Sachs Disease, ten times more likely to have torsion dystonia, et cetera. Genetic diseases are so common in this population that the are official recommendation is that *all* Ashkenazi Jewish couples get screened for genetic disease before marriage. I'm Ashkenazi Jewish, I got screened, and I turn out to be a carrier for Riley-Day syndrome – three hundred times as common in Ashkenazi Jews as in anyone else.

Evolution usually gets rid of genetic diseases pretty quickly. If they stick around, it's because they're doing something to earn their keep. One common pattern is "heterozygote advantage" – two copies of the gene cause a disease, but one copy does something good. For example, people with two copies of the sickle cell gene get sickle cell anaemia, but people with one copy get some protection against malaria. In Africa, where malaria is relatively common, the tradeoff is worth it – so people of African descent have high rates of the sickle cell gene and correspondingly high rates of sickle cell anaemia. In other places, where malaria is relatively uncommon, the tradeoff isn't worth it and evolution eliminates the sickle cell gene. That's why sickle cell is about a hundred times more common in US blacks than US whites.

The moral of the story is: populations can have genetic diseases if they also provide a useful advantage to carriers. And if those genetic diseases are limited to a single group, we expect them to provide a useful advantage for that group, but not others. Might the Jewish genetic diseases provide some advantage? And why would that advantage be limited to Jews?

Most of the Jewish genetic diseases cluster into two biological systems – the sphingolipid system and the DNA repair system. This is suspicious. It suggests that they're not just random. They're doing something specific. Both of these systems are related to neural growth and neural branching. Might they be doing something to the brain?

Gaucher's disease, one of the Ashkenazi genetic diseases, appears to increase IQ. CHH obtained a list of all of the Gaucher's patients in Israel. They were about 15 times more likely than the Israeli average to be in high-IQ occupations like scientist or engineer; CHH calculate the probability that this is a coincidence to be 4×10^-19.

Torsion dystonia, another Ashkenazi genetic disease, shows a similar pattern. CHH find ten reports in the literature where doctors comment on unusual levels of intelligence in their torsion dystonia patients. Eldridge, Harlan, Cooper, and Riklan tested 14 torsion dystonia patients and found an average IQ of 121; another similar study found an average of 117. Torsion dystonia is pretty horrendous, but sufferers will at least get the consolation prize of being really, really smart.

Moving from medicine to history, we find that Ashkenazi Jews were persecuted for the better part of a millennium, and the particular form of this persecution was locking them out of various jobs until the main career opportunities open to them were things like banker, merchant, and doctor. CHH write:

> For 800 to 900 years, from roughly 800 AD to 1650 or 1700 AD, the great majority of the Ashkenazi Jews had managerial and financial jobs, jobs of high complexity, and were neither farmers nor craftsmen. In this they differed from all other settled peoples of which we have knowledge.

They continue:

> Jews who were particularly good at these jobs enjoyed increased reproductive success. Weinryb (1972, see also Hundert 1992) comments: “More children survived to adulthood in affluent families than in less affluent ones. A number of genealogies of business leaders, prominent rabbis, community leaders, and the like – generally belonging to the more affluent classes – show that such people often had four, six, sometimes even eight or nine children who reached adulthood. On the other hands, there are some indications that poorer families tended to be small ones…as an example, in a census of the town of Brody in 1764 homeowner households had 1.2 children per adult member while tenant households had 0.6.

Now we can start to sketch out the theory in full. Due to persecution, Jews were pushed into cognitively-demanding occupations like banker or merchant and forced to sink or swim. The ones who swam – people who were intellectually up to the challenge – had more kids than the ones who sank, producing an evolutionary pressure in favor of intelligence greater than that in any other ethnic group. Just as Africans experiencing evolutionary pressure for malaria resistance developed the sickle cell gene, so Ashkenazim experiencing evolutionary pressure for intelligence developed a bunch of genes which increased heterozygotes' IQ but caused serious genetic disease in homozygotes. As a result, Ashkenazi ended up somewhat more intelligent – and somewhat more prone to genetic disease – than the rest of the European population.

If true, this would explain the 27% of Nobel Prizes and 50% of world chess champions thing. But one still has to ask – everywhere had Jews. Why Hungary in particular? What was so special about Budapest in the early 1900s?

IV.

Okay, sure, everywhere had Jews. But it's surprising exactly how *many* Jews were in early 1900s Hungary.

The modern United States is about 2% Jewish. Hungary in 1900 was about 5%. The most Jewish city in America, New York, is about 15% Jewish. Budapest in 1900 was 25%. It was one of the most Jewish large cities anywhere in history, excepting only Israel itself. According to Wikipedia, the city's late 19th-century nickname was "Judapest".

So is it possible that all the Jews were winning Nobel Prizes, and Hungary just had more Jews and so more Nobelists?

No. This doesn't seem right. The 1933 European Jewish Population By Country site lists the following size for each country's Jewish communities:

Poland: 3 million
Russia: 2.5 million
Romania: 750,000
Germany: 500,000
Hungary: 500,000
Britain: 300,000

France: 250,000
Austria: 200,000

It’s hard to find a good list of all famous Manhattan Project physicists, but I tried this article and got the following number of famous Jewish Manhattan Project physicists per country of origin:

Hungary: 4
Germany: 2
Poland: 2
Austria: 2
Italy: 1
Netherlands: 1
Switzerland: 1

Here’s an alternative source with a different definition of “famous”, broken down the same way:

Germany: 5
Hungary: 4
Poland: 3
Italy: 2
Austria: 2

The main point seems to be disproportionately many people from Central European countries like Hungary and Germany, compared to either Eastern European countries like Poland and Russia or Western European countries like France and Britain.

The Central European advantage over Western Europe is unsurprising; the Western European Jews probably weren’t Ashkenazim, and so didn’t have the advantage mentioned in the CHH paper above. But is there any reason to think that Central European Jews were more intelligent than Polish and Russian Jews?

I’m not really sure what to think about this. This paper finds that the sphingolipidoses and other Jewish genetic diseases are about twice as common in Central European Jews as in Eastern European Jews, but I have very low confidence in these results. Intra-Jewish gossip points out the Lithuanians as the geniuses among world Jewry, but doesn’t have any similar suggestions about Hungarians. And torsion dystonia, maybe the most clearly IQ-linked disease, is unique to Lithuanians and absent in Hungarians.

Probably much more promising is just to focus on the obvious facts of the social situation. Early`1900s Hungary was a great nation and a prosperous center of learning. Remember, we’re talking about the age of the Austro-Hungarian Empire, one of the most industrialized and dynamic economies of the time. It might have had advantages that Poland, Romania, and Russia didn’t. My list of historical national GDPs per capita is very unimpressed by the difference between Hungarian and Polish GDPs in 1900, but maybe it’s wrong, or maybe Budapest was an especially modern part of Hungary, or maybe there’s something else I’m missing.

Also, there could have been a difference in the position of Jews in these countries. Russia was still experiencing frequent anti-Jewish pogroms in 1900; in Hungary, Jews were among the country’s most noble families. Actually, the extent of Jewish wealth and influence in Hungary sort of defies belief. According to Wikipedia, in 1920 Jews were 60% of Hungarian doctors, 50% of lawyers, 40% of engineers and chemists, and 90% of currency brokers and stock exchange members. “In interwar Hungary, more than half and perhaps as much as 90 percent of Hungarian industry was owned or operated by a few closely related Jewish banking families.”

So Central European Jews – the Jews in Hungary and Germany – had a unique combination of intellectual and financial advantages. This means Hungary's only real rival here is Germany. Since they were rich, industrialized, and pretty liberal about Jewish rights at the beginning of the 20th century – and since they had just as many Jews as Hungary – we should expect to see the same phenomenon there too.

And we kind of do. Germany produced its share of Jewish geniuses. Hans Bethe worked for the Manhattan Project and won a Nobel Prize. Max Born helped develop quantum mechanics and also won a Nobel Prize. James Franck, more quantum physics, another Nobel Prize. Otto Stern, even *more* quantum physics, yet *another* Nobel Prize. John Polanyi, chemical kinetics, Nobel Prize (although he was half-Hungarian). And of course we probably shouldn't forget about that Einstein guy. All of these people were born in the same 1880 – 1920 window as the Martians in Hungary.

I think what's going on is this: Germany and Hungary had about the same Jewish population. And they produced about the same number of genius physicists in the same window. But we think of Germany as a big rich country, and Hungary as a small poor country. And the German Jews were spread over a bunch of different cities, whereas the Hungarian Jews were all crammed into Budapest. So when we hear "there were X Nobel Prize winning German physicists in the early 1900s", it sounds only mildly impressive. But when we hear "there were X Nobel Prize winning physicists from Budapest in the early 1900s", it sounds kind of shocking. But the denominator isn't the number of Germans vs. Hungarians, it's the number of German Jews vs. Hungarian Jews, which is about the same.

V.

This still leaves one question: why the period 1880 to 1920?

On further reflection, this isn’t much of a mystery. The emancipation of the Jews in Eastern Europe was a difficult process that took place throughout the 19th century. Even when it happened, it took a while for the first generation of Jews to get rich enough that their children could afford to go to fancy schools and fritter away their lives on impractical subjects like physics and chemistry. In much of Eastern Europe, the Jews born around 1880 were the first generation that was free to pursue what they wanted and seek their own lot in the world.

The end date around 1920 is more depressing: any Jew born after this time probably wasn’t old enough to escape the Nazis. Almost all the famous Hungarian Jews became physics professors in Europe, fled to America during WWII using channels open to famous physicists, and then made most of their achievements on this side of the Atlantic. There are a couple of stragglers born after 1920 who survived – George Soros’ family lived because they bought identity documents saying they were Christian; Andrew Grove lived because he was hidden by righteous Gentiles. But in general Jews born in Europe after 1920 didn’t have a great life expectancy.

All of this suggests a pretty reasonable explanation of the Martian phenomenon. For the reasons suggested by Cochran, Hardy, and Harpending, Ashkenazi Jews had the potential for very high intelligence. They were mostly too poor and discriminated against to take advantage of it. Around 1880, this changed in a few advanced Central European economies like Germany, Austria, and Hungary. Austria didn’t have many Jews. Germany had a lot of Jews, but it was a big country, so nobody really noticed. Hungary had a lot of Jews, all concentrated in Budapest, and so it was really surprising when all of a sudden everyone from Budapest started winning Nobel Prizes around the same time. This continued until World War II, and then all anyone remembered was “Hey, wasn’t it funny that so many smart people were born in Budapest between 1880 and 1920?”

And this story is really, really, gloomy.

For centuries, Europe was sitting on this vast untapped resource of potential geniuses. Around 1880, in a few countries only, economic and political conditions finally became ripe for the potential to be realized. The result was one of the greatest spurts of progress in scientific history, bringing us relativity, quantum mechanics, nuclear bombs, dazzling new mathematical systems, the foundations of digital computing, and various other abstruse ideas I don’t even pretend to understand. This lasted for approximately one generation, after which a psychopath with a stupid mustache killed everyone involved.

I certainly can’t claim that the Jews were the only people being crazy smart in Central Europe around this time. This was the age of Bohr, Schrodinger, Planck, Curie, etc. But part of me wonders even here. If you have one physicist in a town, he sits in an armchair and thinks. If you have five physicists in a town, they meet and talk and try to help each other with their theories. If you have fifty physicists in a town, they can get funding and start a university department. If you have a hundred, maybe some of them can go into teaching or administration and help support the others. Having this extra concentration of talent in central Europe during this period might have helped Jews and Gentiles alike.

I wonder about this because of a sentiment I hear a lot, from people who know more about physics than I do, that we just don’t get people like John von Neumann or Leo Szilard anymore. That there was some weird magical productivity to the early 20th century, especially in Central Europe and Central European immigrants to the United States, that we’re no longer really able to match. This can’t be a pure numbers game – the Ashkenazi population has mostly recovered since the Holocaust, and people from all over the world are coming to American and European universities and providing more of a concentration of talent than ever. And even though it’s impossible to measure, there’s still a feeling that it’s not enough.

I started down this particular research rabbit hole because a friend challenged me to explain what was so magical about early 20th century Hungary. I think the Jewish population calculations above explain a lot of the story. I’m not sure whether there’s a missing ingredient, or, if so, what it might be. Maybe it really was better education. Maybe it really was math competitions and talent searches.

Or maybe it was superintelligent Martian scouts with an Earthling fetish.

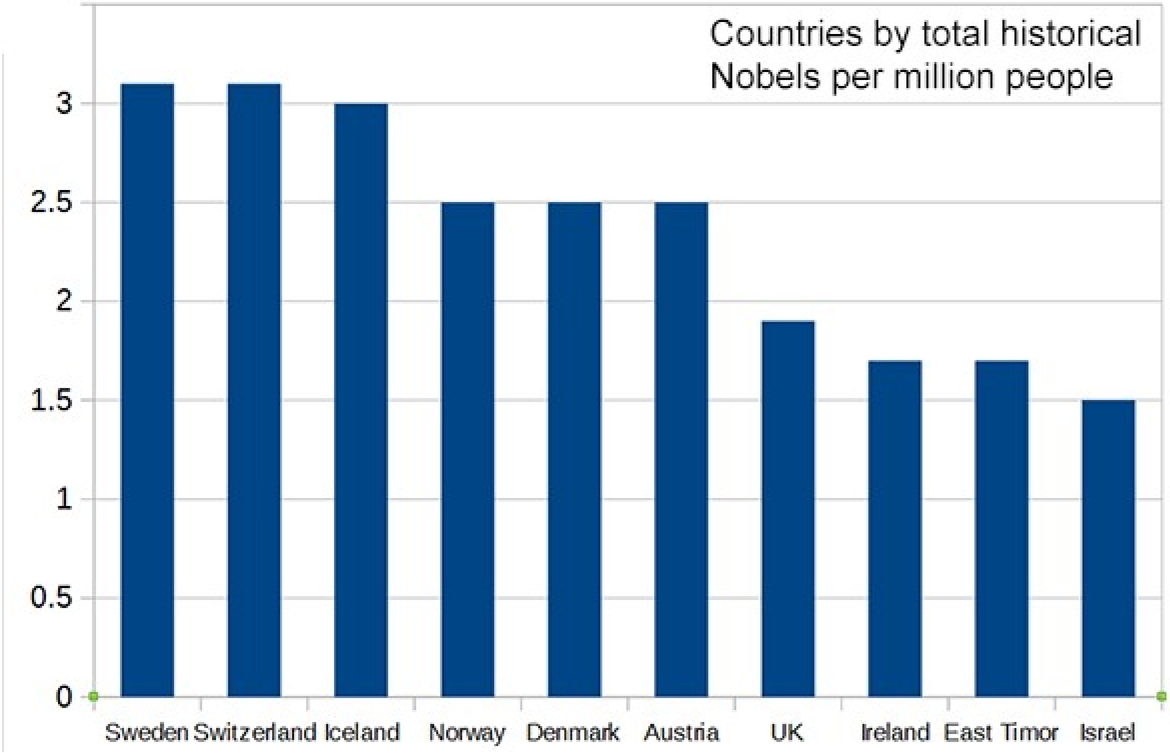
Hungarian Education II: Four Nobel Truths

1. Israel historically has only a moderate number of Nobels per capita

On Friday, I discussed the phenomenon of Hungarian science geniuses, and conjectured it was because of Hungary's high concentration of Ashkenazi Jews. A commenter pointed out that Israel had an even higher concentration of Ashkenazi Jews, with less impressive results:

> By this logic, Israel should have become the hotbed of geniuses. And while it's true that there are a lot of smart people there, none of the Israeli universities are in the top 10 or maybe even in top 100. And the fraction of Nobel prize winners is not impressive, either.

The objection makes superficial sense. The list of Nobel winners per capita puts Israel at a modest tenth place, after places like Norway and the United Kingdom.

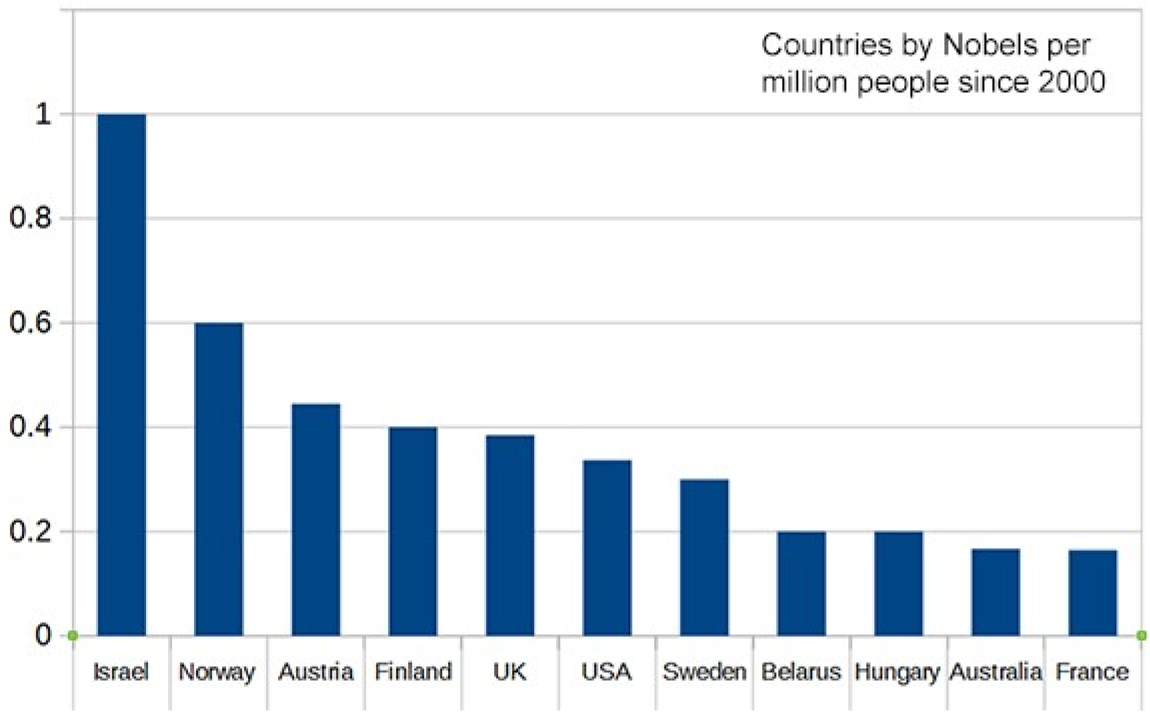


This doesn't look promising for any Ashkenazi-Jew-based theory.

2. But their more modern numbers look much more impressive

On the other hand, that list counts total Nobels won, ever, and divides them by modern population. That gives an advantage to older countries. Norway's been collecting Nobels since 1903; Israel wasn't even *founded* until 1948. And for the first couple generations the Israelis were pretty busy starting kibbutzim, building infrastructure, fighting off enemies, et cetera. Setting up a good university system capable of churning out Nobels takes time. So Norway and the UK had an unfair head start.

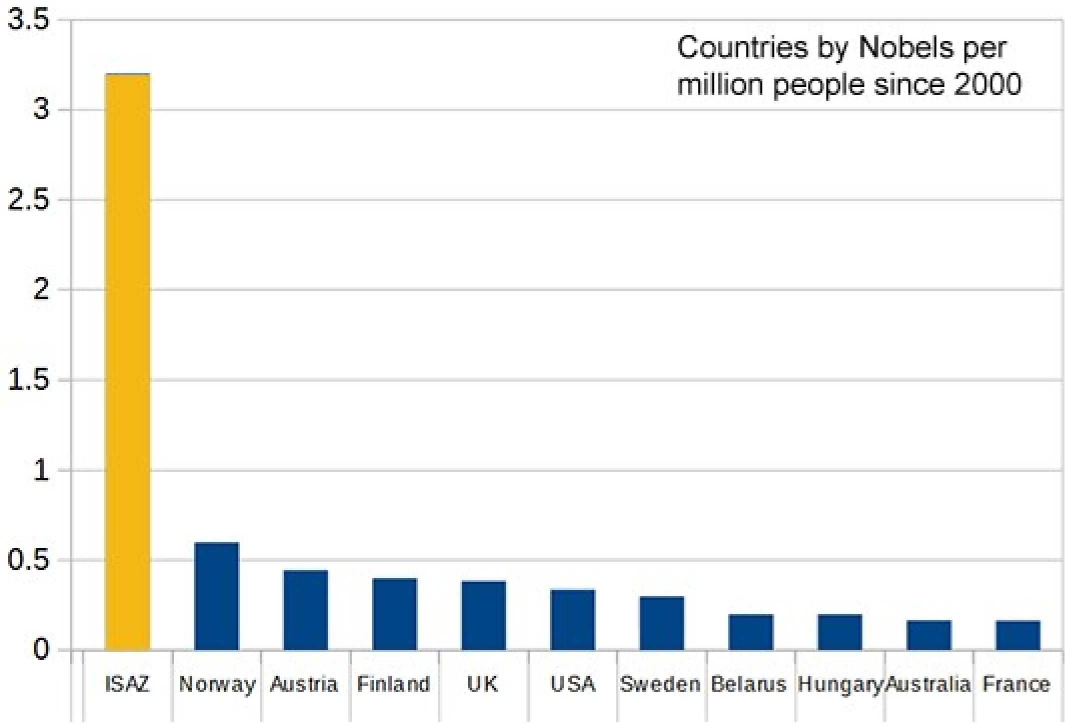
I redid their analysis looking only at Nobels won since the year 2000 (because it was big and round and serves as a signal that I'm trying to avoid optional stopping). My source was this list of Nobel laureates by country, and I deferred to Wikipedia’s judgment about whether or not to count dual citizens, immigrants, et cetera. Here’s the results:



We see that during this period, Israel has by far the highest number of Nobel prizes per capita.

3. This advantage increases if we look only at Ashkenazim

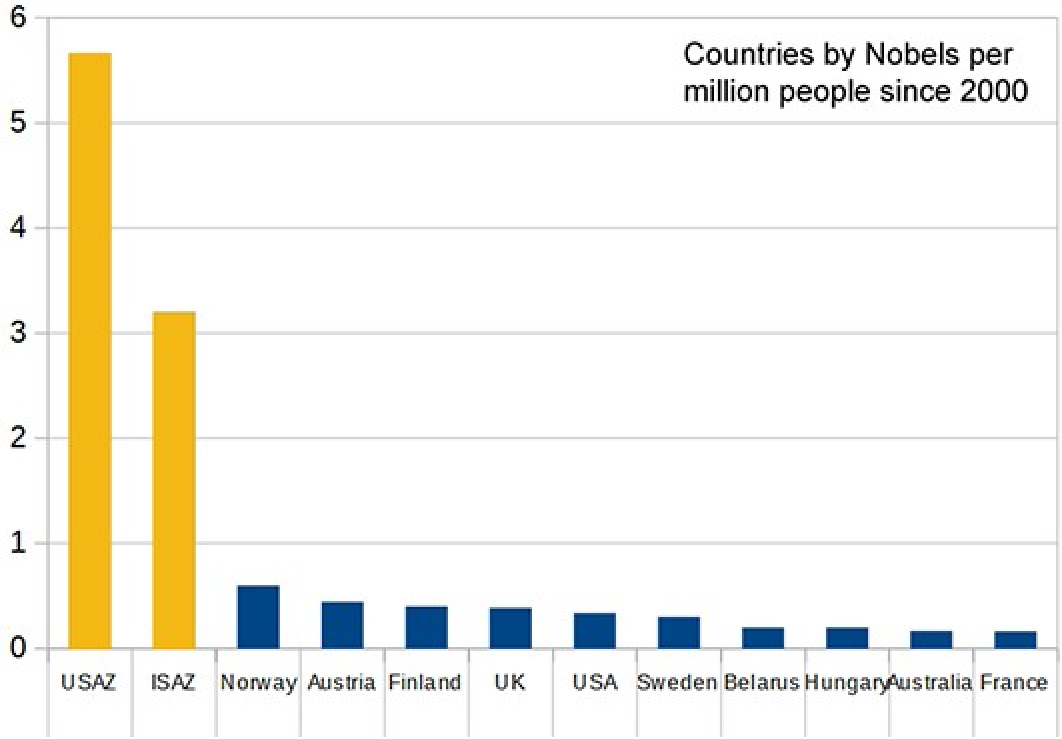
The original theory was about Ashkenazi Jews in particular. Only about a third of Israels are Ashkenazi (the rest are other types of Jews, or Palestinians, or other non-Jewish minorities). If we separate out the Ashkenazim, the graph looks like this.



ISAZ is Israeli Ashkenazi Jews, considered as a separate population. On the one hand, it’s kind of unfair comparing Israel’s most successful population group to other countries taken as a whole. On the other hand, if we were to take other countries’ most successful population groups, those would be Ashkenazi Jews too, so whatever. Since Israeli Ashkenazim get about five times more Nobels per capita than any country, I’m going to consider the “what about Israel?” objection officially refuted.

4. But there’s not a lot of evidence for benefits to concentration, and other factors might be involved

One more graph:



USAZ is US Ashkenazi Jews, who get twice as many Nobels per capita as their Israeli cousins (I’m not sure how seriously to take this; the Israeli data is based on eight Nobel laureates, so there’s a lot of room for sampling issues.)

And although it’s hard for me to get exact numbers, it looks like a lot of Israeli Nobelist (maybe more than half) did their best work abroad, usually in the US.

Israel went from 1948 to 2002 without winning a single science Nobel (it did win in Literature and Peace during that time). Now it’s winning more of them – a lot more, more than any other country per capita – but mostly when its citizens go and study in foreign universities. This seems consistent with an Israeli educational system that’s still struggling to get its act together.

Does this mean that once the educational system gets its act together more fully, the ISAZ Nobel rate will approximately double to match the USAZ Nobel rate? I’m not sure.

Just from this analysis, it doesn’t look like the theory in the last post, where everyone gets benefits from concentrating closer together, is true. Israel has about ten times as many Ashkenazi Jews per capita than the US, but still does worse than they do.

These data don’t challenge the conclusion from the last post that Ashkenazim might have been responsible for Hungary’s sudden crop of great scientists. But they do potentially challenge the implicit conclusion that the education system didn’t matter that much. I’ll have more on that later this week.

Hungarian Education III: Mastering The Core Teachings Of The Budapestians

[Previously in series: The Atomic Bomb Considered As A Hungarian High School Science Fair Project, Four Nobel Truths]

I.

Someone summed up my previous post as “Hungarian education isn’t magic”. I would amend that to read “Hungarian education isn’t *systemically* magic”. As far as I know, there’s only one Hungarian educator with magic powers, and (like all good wizards) his secrets are maddeningly hard to find.

Laszlo Polgar studied intelligence in university, and decided he had discovered the basic principles behind raising any child to be a genius. He wrote a book called *Bring Up Genius* and recruited an interested woman to marry him so they could test his philosophy by raising children together. He said a bunch of stuff on how ‘natural talent’ was meaningless and so any child could become a prodigy with the right upbringing.

This is normally the point where I’d start making fun of him. Except that when he trained his three daughters in chess, they became the 1st, 2nd, and 6th best female chess players in the world, gaining honors like “youngest grandmaster ever” and “greatest female chess player of all time”. Also they spoke seven languages, including Esperanto.

Their immense success suggests that education can have a major effect even on such traditional genius-requiring domains as chess ability. How can we reconcile that with the rest of our picture of the world, and how obsessed should we be with getting a copy of Laszlo Polgar's book?

II.

Let's get this out of the way first: the Polgar sisters were probably genetically really smart. The whole family was Hungarian Jews, a group with a great track record. Their mother and father were both well-educated teachers interested in stuff like developmental psychology. They had every possible biological advantage and I'm sure that helped.

J Levitt proposes an equation to estimate a chess player's IQ from their chess score. It suggests that chess grandmasters probably have IQs above 160. Plugging the Polgar sisters' chess scores into his equation, I get IQs in the range of 150, 160, and 170 for the three sisters.

~~This is biologically impossible. Even if both Polgar parents were 170-IQ themselves, regression to the mean predicts that their children would have IQs around 140 to 150. It's mathematically possible for there to be an IQ that predicts you would have three children of 150, 160, and 170, but I doubt any living people have it, and even if they did there's no way they would marry somebody else equally gifted.~~

[EDIT: Thanks to a few people who pointed out some problems with my math here (1, 2, 3). I still think that having three supergenius-IQ kids when you and your spouse show no signs of being a supergenius yourself (Laszlo Polgar's daughters could beat him at chess by the time they were 8) is pretty unlikely, but I admit not impossible. I still think arguing about this is unnecessary thanks to the points below.]

On the other hand, I’m not sure Levitt’s right. Chess champion Gary Kasparov actually sat and took an IQ test for the magazine *Der Spiegel*, and his IQ was 135. That’s not bad – it’s top 1% of the population – but it’s not amazing either.

This is what we should expect given the correlation of about r = 0.24 between IQ and chess ability (see also this analysis, although I disagree with the details). And the contrary claims – like the one that Bobby Fischer’s IQ was in the 180s – are less well-sourced (although Fischer was the son of a Hungarian-Jewish mathematician, so who knows?).

If it were possible to be a chess world champion with an IQ of 135, then maybe it’s possible to be a “mere” grandmaster with IQs in the high 120s and low 130s. And it’s just barely plausible that some sufficiently smart people might have three kids who all have IQs in the high 120s and low 130s.

But this just passes the buck on the mystery. 2% of people have IQs in the high 120s or low 130s, but 2% of people aren’t the top-ranked female chess player in the world. The Polgar sisters’ IQs might have been a *permissive* factor in allowing them to excel, but it didn’t *necessitate* it. So what’s going on there?

III.

“Practice” seems like an obvious part of the picture. Malcolm Gladwell uses the Polgars as poster children for his famous ‘10,000 hours of practice makes you an expert at anything’ rule. The Polgars had 50,000 hours of chess practice each by the time they were adults, presumably enough to make them quintuple-experts.

Robert Howard has a paper Does High-Level Performance Depend On Practice Alone? Debunking The Polgar Sisters Case in which he argues against the strong version of Gladwell's thesis. He points out that there are many chess masters who have practiced much less than the Polgar sisters but are better than they are. He also points out that even though the sisters themselves have all practiced similar amounts, youngest sister Judit is clearly better than the other two in a way that practice alone cannot explain.

I don't know if the case he's arguing against – that practice is literally everything and it's impossible for anything else to factor in – is a straw man or not. But it seems more important to consider a less silly argument – that practice is one of many factors, and that enough of it can make up for a lack of the others. This seems potentially true. This study showing that amount of practice only explains 12% of the variance in skill level at various tasks, and is often summarized as "practice doesn't matter much". But it finds practice matters more (25% of the variance) in unchanging games with clear fixed rules, and uses chess as an example.

So suppose that the Polgar sisters are genetically smart, but maybe not as high up there as some other chess masters. We would expect them to need much more practice to achieve a level of proficiency similar to those chess masters, and indeed that seems like what happens.

(all of this is confounded by them being women and almost all the other equally-good chess masters being men. It's unclear if the Polgars deserve extra points for overcoming whatever factor usually keeps women out of the highest levels of chess.)

But I'm actually still not sure this suffices as an explanation. According to Wikipedia:

> Polgár began teaching his eldest daughter, Susan, to play chess when she was four years old. Six months later, Susan toddled into Budapest’s smoke-filled chess club,” which was crowded with elderly men, and proceeded to beat the veteran players.

The study linked above suggests that Susan practiced 48 hours a week. During those six months, she would have accumulated about 1200 hours of practice. Suppose the elderly Budapest chess players practiced only one hour a week, but had been doing so for the last twenty-five years. They would have more practice than Susan – plus the advantage of having older, more developed brains. So why did she beat them so easily?

Maybe there’s a time-decay factor for practice? That is, maybe Susan had been practicing intensively, so she got a lot of chances to link it all together as she was learning, and also it was fresh in her mind when she went to the club to go play? I’m not sure. If some of those veterans had been playing more than one hour a week (and surely the sort of people who frequent Budapest chess clubs do) then her advantage seems too implausible to be due to freshness-of-material alone.

IV.

That leaves two possibilities.

First, Susan could have benefitted from some form of malleability. A lot of people claim there’s a “developmental window” during which children have a unique ability to learn language. If cats see only vertical stripes for the first few weeks of their lives, they never learn to see in horizontal. Maybe if you teach your kid high-level chess at age 4, they’ll be able to recruit systems that adults could never manage, or reorganize the fundamental structure of their brain to conform to chess better, or something like that.

Second, Polgar might actually have some really good educational methods besides just “start early and have a lot of practice”. I assume this is true, but I’m having a lot of trouble finding them. Shockingly, Polgar’s book *Bring Up Genius* is out of print and totally unavailable anywhere – I guess the book-reading community heard that someone wrote down a way to reliably turn any child into a genius which had a great real-world track record of success, and collectively decided “Nah, better read *Fifty Shades Of Grey* instead”. I’m not sure at what point I should start positing a conspiracy of suppression, or whether that would be better or worse than the alternative.

The book seems to possibly be available in Hungarian under the title *Nevelj zsenit!*, but I can’t tell for sure and a lot of the Hungarian sites suggest it’s out of print even in that language. There *may* have been a recent republication in Esperanto called *Eduku geniulon!*, but I can’t find that one either. If anybody knows where to find this book and wants to send it to me, I will figure out some way to translate it and review it. I’d also be willing to pay for costs and even pay extra for your time if it helps. Come on, Esperanto-speakers! This is the only chance you’ll ever have to be useful!

V.

One thing I know without reading the book: Polgar says that his method should work to create geniuses in any field, not just chess. He said he chose chess kind of on the whim of his eldest daughter. From Wikipedia:

> Polgár and his wife considered various possible subjects in which to drill their children, “including mathematics and foreign languages,” but they settled on chess. “We could do the same thing with any subject, if you start early, spend lots of time and give great love to that one subject,” Klara later explained. “But we chose chess. Chess is very objective and easy to measure.” Susan described chess as having been her own choice: “Yes, he could have put us in any field, but it was I who chose chess as a four-year-old…. I liked the chessmen; they were toys for me.”

It’s disappointing that he decided to stick with chess for his other two daughters. The study linked above suggests that chess is unusually amenable to practice. What would have happened if he’d tried to train his kids in art? In mathematics? In entrepreneurship? I’m not sure, and I’m really tempted to have some kids and find out.

(be right back, going to change my OKCupid profile to include “must be interested in n=1 developmental-psych experiments, have access to a rare book library, and speak either Hungarian or Esperanto”)

I mentioned this plan to a friend, who protested that this was cruel and tantamount to child abuse. After all, how can you force someone to spend their entire childhood indoors, studying mind-numbing chess problems day in and day out, instead of enjoying themselves like normal kids?

First of all, this isn’t how the Polgar children (or adults) describe their experiment. From *The Guardian*:

> Starting with his eldest daughter, Susan, Polgár was careful to treat it as a playful activity, turning it into a fantasy of dramatic wins and losses. Whereas Earl and

> Kultida Woods had coerced perfection from Tiger, the Polgárs encouraged enjoyment, By the time Susan had turned five, she was excited by playing and spent hundreds of hours practising. She was entered into a local competition and treated it as fun, winning 10-0, causing a sensation.
>
> Meanwhile, her younger sisters were intrigued and László allowed them to feel the pieces, seeing them as toys, with no formal tuition until they were five. Interviewed recently, all three girls described playing the game as something that they loved doing – it never felt like a chore. Instead of messing about playing Monopoly, netball or going to the local swimming pool, chess was just what the Polgár family enjoyed…Polgár understood that coercion was less valuable than small children's need to enjoy fantasy play. Consequently, his daughters all seem to have grown into satiable, well-balanced people rather than success addicts.

But more important – I responded that the Polgars claim to have spent about 48 hours a week practicing chess. I spent seven hours a day in school, so if my teachers assigned two hours of homework a night then we spent about the same amount of time getting educated. Except what the *Polgars* got out of it was world-champion-level mastery of their favorite subject in the world, nationwide fame, and (by their own accounts) loving every second of it, and what *I* got was staring out a window all day as my teacher declared that we were going to make a collage about the meaning of Respect.

The Polgar sisters talk about how they loved their education, had a great childhood, thought their parents were always patient with them and never strict and harsh, and don't regret anything. How many kids who went to public school can say the same?

An article about Laszlo Polgar mentions that he had to fight the Hungarian authorities to be allowed to home school his children. Imagine being so certain of your own home-schooling techniques that you're afraid taking your kids to the Fasori Gymnasium is going to stunt their intellectual growth. And imagine being *right*. And imagine my friend thinking that normal American public school might be better than that. It sort of boggles the imagination.

And I guess I shouldn't be too harsh, because the public school system tries to do the best it can with an impossible set of constraints. But I'm still suspicious. *Who else has the motivation to hide that book?*

[EDIT: Thanks to readers, I've got an Esperanto copy and a person willing to translate it. I'll let you know as this develops.]

Is Pharma Research Worse Than Chance?

[Epistemic status: very speculative]

The two most exciting developments in psychopharmacology in the 21st century so far have been ketamine for depression and MDMA for PTSD.

Unlike other antidepressants, which work intermittently over a space of weeks, ketamine can cause near-instant remission of depression with a single infusion – which lasts a week or two and can be repeated if needed. Ketamine use may be successful in 50-70% of patients who have failed treatment with conventional antidepressants. Ketamine treatment has some issues right now, but the race is on to create an oral non-hallucinogenic version which could be the next big blockbuster drug and revolutionize depression treatment.

MDMA ("Ecstasy") is undergoing FDA Phase 3 clinical trials as a treatment for PTSD. Preliminary research has been small and underpowered, but suggests response rates up to 80% and effect sizes greater than 1 in this otherwise-hard-to-treat condition. None of this is on really firm footing – that'll have to wait for the Phase 3. But signs are looking very good.

I say these are the two most exciting developments mostly because no other developments have been exciting. In terms of normal psychiatric drugs, the best that the 21st century has given us has probably been pimavanserin and aripiprazole, modest updates to the standard atypical antipsychotic model. These drugs are probably a bit better than existing ones for the people who need them (especially pimavenserin for psychosis in Parkinson's) but they don't revolutionize the treatment of any condition and nobody ever claimed that they did. And most drugs aren't even at this level – they're new members of well-worn classes with slightly different side effect profiles. The landscape was so quiet that ketamine came in like a bolt from the blue, and MDMA is set to do the same in a couple of years when the trial results come out.

(if I'm wrong, and history decides these two drugs *weren't* the biggest developments, the most likely failure mode is that psilocybin turned out to be more important than MDMA)

There's a morality tale to be told here about how the War on Drugs choked off vital research on some of the most powerful psychiatric compounds and cost us fifty years in exploring these effects and treating patients. I agree with this morality tale as far as it goes, but I also think there's another, broader morality tale beneath it.

Suppose that neither ketamine nor MDMA were illegal drugs. Ketamine was just used as an anaesthetic. MDMA was just used as a chemical intermediate in producing haemostatic drugs, its original purpose. Now the story is that, fifty years later, we learn that this anaesthetic and this haemostatic turn out to have incredibly powerful psychiatric effects. What's our narrative *now*?

For me it's about the weird inability of intentional psychopharmaceutical research to discover anything as good as things random druggies use to get high.

For decades, pharmaceutical companies have been coming out with relatively lackluster mental health offerings – aripiprazole, pimavanserin, and all the rest. And when asked why, they answer that mental health is hard, the brain is the most complicated organ in the known universe, we shouldn’t *expect* there to be great cures with few side effects for psychiatric diseases, and if there were we certainly shouldn’t expect them to be easy to find.

And this would make sense *except* in the context of ketamine and MDMA. Here are some random chemicals that affect the brain in some random way, which people were using mostly because they felt good at raves, and huh, they seem to treat psychiatric diseases much better than anything produced by some of the smartest people in the world working for decades on ways to treat psychiatric diseases. Why should that be?

One could argue it’s all about numbers vs. base rates. There are way more chemicals synthesized each year by people who *aren’t* looking for psychiatric drugs than by people who *are*. Even if the people who are looking for drugs are a thousand times more likely to find them, the people-who-aren’t-looking can still overwhelm them with sheer numerical advantage. And maybe when a psychiatric drug is discovered by people who weren’t looking for it, what this looks like is a few random people trying it, noticing it feels good, and turning it into a drug of abuse.

And I’m sure this is part of the story. But that just passes the buck to the next question. Abusers take the vast flood of possible chemicals and select the ones they think will feel good at raves. Psychopharmacologists take the vast flood of possible chemicals and select the ones they think will treat mental illnesses. How come the abusers’ selection process is better at picking out promising mental health treatments?

Here’s one hypothesis: at the highest level, the brain doesn’t have that many variables to affect, or all the variables are connected. If you smack the brain really really hard in some direction or other, you will probably treat some psychiatric disease. Drugs of abuse are ones that smack the brain really hard in some direction or other. They do *something*. So find the psychiatric illness that's treated by smacking the brain in that direction, and you're good.

(in fact, the most effective existing treatment for depression is electroconvulsive therapy – ie giving the brain a big electric shock. This is maybe the crudest, most literally "smack the brain really hard" treatment out there, but it sure does work)

Actual carefully-researched psychiatric drugs are exquisitely selected for having few side effects. The goal is something like an SSRI – mild stomach discomfort, some problems having sex, but overall you can be on them forever and barely notice their existence. In the grand scheme of things their side effects are tiny – in most placebo-controlled studies, people have a really hard time telling whether they're in the experimental or the placebo group.

Nobody has a hard time telling whether they're in the experimental or placebo group of a trial of high-dose MDMA. I think this might be the difference. If you go for large effects – even if you don't really care what direction the effect is in – you'll get them. And if you go for small, barely perceptible effects, then you'll get those too. The dream of the magic bullet – the drug that treats exactly what it's supposed to treat but otherwise has no effect at all on you – is just a dream. The closest you can come is something with miniscule side effects but a barely-less-miniscule treatment effect.

But given that we're all very excited to learn about ketamine and MDMA, and given that if their original promise survives further testing we will consider them great discoveries (and given that ECT was also a great and productive discovery) it suggests we chose the wrong part of the tradeoff curve. Or at least it suggests a different way of framing that tradeoff curve. A drug that makes you feel extreme side effects for a few hours – but also has very strong and lasting treatment effects – is better than a drug with few side effects and weaker treatment effects. That suggests a new direction pharmaceutical companies might take: look for the chemicals that have the strongest and wackiest effects on the human mind. Then see if any of them also treat some disease.

I think this is impossible with current incentives. There's too little risk-tolerance at every stage in the system. But if everyone rallied around the idea, it might be that trying the top hundred craziest things Alexander Shulgin dreamed up on whatever your rat model is would be orders of magnitude more productive than whatever people are doing now.

Or it might not be. I can also think of a counterargument to the theory above, which is that our current best model of ketamine suggests it's a non-psychoactive metabolite that has most of the useful antidepressant effect. In fact, a lot of people think that one form of ketamine is hallucinogenic (and extremely effective against chronic pain) and another form (or its metabolite) is the antidepressant. I'm a little suspicious trying to calculate the odds of a single chemical having two forms, one of which is a really exciting analgesic, and the other of which is a really exciting antidepressant, by two different mechanisms. It sounds too much like finding some new chemical compound whose solid form is a room-temperature superconductor, and whose liquid form catalyzes cold fusion, by two totally different mechanisms. It seems a little *too* lucky (see here for some ketamine skepticism, and here for my response). But if it were true, it means that ketamine's psychoactive effects were a red herring in helping us discover it as an antidepressant, even though they were a very effective red herring.

SSC Journal Club: AI Timelines

I.

A few years ago, Muller and Bostrom et al surveyed AI researchers to assess their opinion on AI progress and superintelligence. Since then, deep learning took off, AlphaGo beat human Go champions, and the field has generally progressed. I've been waiting for a new survey for a while, and now we have one.

Grace et al (*New Scientist* article, paper, see also the post on the author's blog AI Impacts) surveyed 1634 experts at major AI conferences and received 352 responses. Unlike Bostrom's survey, this didn't oversample experts at weird futurist conferences and seems to be a pretty good cross-section of mainstream opinion in the field. What did they think?

Well, a lot of different things.

The headline result: the researchers asked experts for their probabilities that we would get AI that was "able to accomplish every task better and more cheaply than human workers". The experts thought on average there was a 50% chance of this happening by 2062 – and a 10% chance of it happening by 2026!

But on its own this is a bit misleading. They also asked by what year "for any occupation, machines could be built to carry out the task better and more cheaply than human workers". The experts thought on average that there was a 50% chance of this happening by 2139, and a 20% chance of it happening by 2037.

As the authors point out, these two questions are basically the same – they were put in just to test if there was any framing effect. The framing effect was apparently strong enough to shift the median date of strong human-level AI from 2062 to 2139. This makes it hard to argue AI experts actually have a strong opinion on this.

Also, these averages are deceptive. Several experts thought there was basically a 100% chance of strong AI by 2035; others thought there was only a 20% chance or less by 2100. This is less “AI experts have spoken and it will happen in 2062” and more “AI experts have spoken, and everything they say contradicts each other and quite often themselves”.

This *does* convey more than zero information. It conveys the information that AI researchers are *really unsure*. I can’t tell you how many people I’ve heard say “there’s no serious AI researcher who thinks there’s any chance of human-level intelligence before 2050”. Well actually, there are a few dozen conference-paper-presenting experts who think there’s a *one hundred* percent chance of human-level AI before that year. I don’t know what drugs they’re on, but they exist. The moral of the story is: be less certain about this kind of thing.

II.

The next thing we can take from this paper is a timeline of what will happen when. The authors give a bunch of different tasks, jobs, and milestones, and ask the researchers when AI will be able to complete them. Average answers range from nearly fifty years off (for machines being able to do original high-level mathematical research) to only three years away (for machines achieving the venerable accomplishment of being able to outperform humans at *Angry Birds*). Along the way they'll beat humans at poker (four years), writing high school essays (ten years), be able to outrun humans in a 5K foot race (12 years), and write a New York Times bestseller (26 years). What do these AI researchers think is the hardest and most quintessentially human of the tasks listed, the one robots will have the most trouble doing because of its Olympian intellectual requirements? That's right – AI research (80 years).

I make fun of this, but it's actually interesting to think about. Might the AI researchers have put their own job last not because of an inflated sense of their own importance, but because they engage with it every day in Near Mode? That is, because they imagine writing a New York Times bestseller as "something something pen paper be good with words okay done" whereas they understand the complexity of AI research and how excruciatingly hard it would be to automate away every piece of what they do?

Also, since they rated AI research (80 years) as the hardest of all occupations, what do they mean when they say that "full automation of all human jobs" is 125 years away? Some other job not on the list that will take 40 years longer than AI research? Or just a combination of framing effects and not understanding the question?

(it's also unclear to what extent they believe that automating AI research will lead to a feedback loop and subsequent hard takeoff to superintelligence. This kind of theory would fit with it being the last job to be automated, but not with it taking another forty years before an unspecified age of full automation.)

III.

The last part is the most interesting for me: what do AI researchers believe about risk from superintelligence?

This is very different from the earlier questions about timelines. It's possible to believe that AI will come very soon but be perfectly safe. And it's possible to believe that AI is a long time away but we really need to start preparing now, or else. A lot of popular accounts collapse these two things together, "oh, you're worried about AI, but that's dumb because there's no way it's going to happen anytime soon", but past research has shown that short timelines and high risk assessment are only modestly correlated. This survey asked about both separately.

There were a couple of different questions trying to get at this, but it looks like the most direct one was "does Stuart Russell's argument for why highly advanced AI might pose a risk, point at an important problem?". You can see the exact version of his argument quoted in the survey on the AI Impacts page, but it's basically the standard Bostrom/Yudkowsky argument for why AIs may end up with extreme values contrary to our own, framed in a very normal-sounding and non-threatening way. According to the experts, this was:

No, not a real problem: 11%
No, not an important problem: 19%
Yes, a moderately important problem: 31%
Yes, an important problem: 34%
Yes, among the most important problems in the field: 5%

70% of AI experts agree with the basic argument that there's a risk from poorly-goal-aligned AI. But very few believe it's among "the most important problems in the field". This is pretty surprising; if there's a good chance AI could be hostile to humans, shouldn’t that automatically be pretty high on the priority list?

The next question might help explain this: “Value of working on this problem now, compared to other problems in the field?”

Much less valuable: 22%
Less valuable: 41%
As valuable as other problems: 28%
More valuable: 7%
Much more valuable: 1.4%

So charitably, the answer to this question was coloring the answer to the previous one: AI researchers believe it’s plausible that there could be major problems with machine goal alignment, they just don’t think that there’s too much point in working on it now.

One more question here: “Chance intelligence explosion argument is broadly correct?”

Quite likely (81-100% chance): 12%
Likely (61-80% chance): 17%
About even (41-60% chance): 21%
Unlikely (21-40% chance): 24%
Quite unlikely (0-20% chance): 26%

Splitting the 41-60% bin in two, we might estimate that about 40% of AI researchers think the hypothesis is more likely than not.

Take the big picture here, and I worry there’s sort of a discrepancy.

50% of experts think there’s at least a ten percent chance of above-human-level AI coming within the next ten years.

And 40% of experts think that there’s a better-than-even chance that, once we get above-human level AI, it will “explode” to suddenly become vastly more intelligent than humans.

And 70% of experts think that Stuart Russell makes a pretty good point when he says that without a lot of research into AI goal alignment, AIs will probably have their goals so misaligned with humans that they could become dangerous and hostile.

I don’t have the raw individual-level data, so I can’t prove that these aren’t all anti-correlated in some perverse way that’s the opposite of the direction I would expect. But if we assume they’re not, and just naively multiply the probabilities together for a rough estimate, that suggests that about 14% of experts believe that all three of these things: that AI might be soon, superintelligent, and hostile.

Yet only a third of these – 5% – think this is “among the most important problems in the field”. Only a tenth – 1.4% – think it’s “much more valuable” than other things they could be working on.

IV.

How have things changed since Muller and Bostrom’s survey in 2012?

The short answer is “confusingly”. Since almost everyone agrees that AI progress in the past five years has been much faster than expected, we would expect experts to have faster timelines – ie expect AI to be closer now than they did then. But Bostrom’s sample predicted human-level AI in 2040 (median) or 2081 (mean). Grace et al don’t give clear means or medians, preferring some complicated statistical construct which isn’t exactly similar to either of these. But their dates – 2062 by one framing, 2139 by another – at least seem potentially a little bit later.

Some of this may have to do with a subtle difference in how they asked their question:

Bostrom: “Define a high-level machine intelligence as one that can carry out most human professions as well as a typical human…”

Grace: “High-level machine intelligence is achieved when unaided machines can accomplish every task better and more cheaply than human workers.”

Bostrom wanted it equal to humans; Grace wants it better. Bostrom wanted “most professions”, Grace wants “every task”. It makes sense that experts would predict longer timescales for meeting Grace’s standards.

But as we saw before, expecting AI experts to make sense might be giving them too much credit. A more likely possibility: Bostrom’s sample included people from wackier subbranches of AI research, like a conference on Philosophy of AI and one on Artificial General Intelligence; Grace’s sample was more mainstream. The most mainstream part of Bostrom’s sample, a list of top 100 AI researchers, had an estimate a bit closer to Grace’s (2050).

We can also compare the two samples on belief in an intelligence explosion. Bostrom asked how likely it was that AI went from human-level to “greatly surpassing” human level within two years. The median was 10%; the mean was 19%. The median of top AI researchers not involved in wacky conferences was 5%.

Grace asked the same question, with much the same results: a median 10% probability. I have no idea why this question – which details what an “intelligence explosion” would entail – was so much less popular than the one that used the words “intelligence explosion” (remember, 40% of experts agreed that “the intelligence explosion argument is broadly correct”). Maybe researchers believe it’s a logically sound argument and worth considering but in the end it’s not going to happen – or maybe they don’t actually know what “intelligence explosion” means.

Finally, Bostrom and Grace both asked experts’ predictions for whether the final impact of AI would be good or bad. Bostrom’s full sample (top 100 subgroup in parentheses) was:

Extremely good: 24% (20)
On balance good: 28% (40)
More or less neutral: 17% (19)
On balance bad: 13% (13)
Extremely bad – existential catastrophe: 18% (8)

Grace’s results for the same question:

Extremely good: 20%
On balance good: 25%
More or less neutral: 40%
On balance bad: 10%
Extremely bad – human extinction: 5%

Grace’s data looks pretty much the same as the TOP100 subset of Bostrom’s data, which makes sense since both are prestigious non-wacky AI researchers.

V.

A final question: “How much should society prioritize AI safety research”?

Much less: 5%
Less: 6%
About the same: 41%
More: 35%
Much more: 12%

People who say that real AI researchers don’t believe in safety research are now just empirically wrong. I can’t yet say that most of them want more such research – it’s only 47% on this survey. But next survey AI will be a little bit more advanced, people will have thought it over a little bit more, and maybe we’ll break the 50% mark.

But we’re not there yet.

I think a good summary of this paper would be that large-minorities-to-small-majorities of AI experts agree with the arguments around AI risk and think they’re worth investigating further. But only a very small minority of experts consider it an emergency or think it’s really important right now.

You could tell an optimistic story here – “experts agree that things will probably be okay, everyone can calm down”.

You can also tell a more pessimistic story. Experts agree with a lot of the claims and arguments that suggest reason for concern. It’s just that, having granted them, they’re not *actually* concerned.

This seems like a pretty common problem in philosophy. “Do you believe it’s more important that poor people have basic necessities of life than that you have lots of luxury goods?” “Yeah” “And do you believe that the money you’re currently spending on luxury goods right now could instead be spent on charity that would help poor people get life necessities?” “Yeah.” “Then shouldn’t you stop buying luxury goods and instead give all your extra money beyond what you need to live to charity?” “Hey, what? Nobody does that! That would be a lot of work and make me look really weird!”

How many of the experts in this survey are victims of the same problem? “Do you believe powerful AI is coming soon?” “Yeah.” “Do you believe it could be really dangerous?” “Yeah.” “Then shouldn’t you worry about this?” “Hey, what? Nobody does that! That would be a lot of work and make me look really weird!”

I don’t know. But I’m encouraged to see people are even taking the arguments seriously. And I’m encouraged that researchers are finally giving us good data on this. Thanks to the authors of this study for being so diligent, helpful, intelligent, wonderful, and (of course) sexy.

(I might have forgotten to mention that the lead author is my girlfriend. But that’s not biasing my praise above in any way.)

What Is Depression, Anyway?: The Synapse Hypothesis

I.

The problem with depression research isn’t that we don’t have any leads on what causes depression. It’s that we have so many leads on what causes depression that we don’t know what to do with all of them. For example:

1. Life adversity, like getting fired or breaking up with a partner, can make people depressed. The biological correlate of this seems to be the hypothalamic–pituitary–adrenal axis (HPA), where your brain tells your adrenal glands to produce glucocorticoid stress hormones like cortisol and this does something to your brain that increases the risk of depression.

2. Inflammation and immune overactivity can make people depressed. The classic examples of this are cancer-related depression (which exceeds what you would expect just from cancer being stressful) and depression induced by administration of the immunomodulator interferon-a. Antiinflammatory drugs have a small but clinically relevant antidepressant effect. Some of the relevant chemicals here seem to be TNF-A and IL-1; these do something to your brain that increases the risk of depression.

3. Serotonin and other monoamines seem to be involved. Most existing antidepressants, like SSRIs and MAOIs, seem to work by increasing monoamine levels. There are some conditions which affect monoamine levels and also increase risk of depression, though it’s nothing like a perfect correlation.

4. The glutamate system (eg NMDA and AMPA receptors) seem to be involved. Ketamine acts on both of these receptors in different ways, and one of those actions is the source of its rapid and unprecedented antidepressant effects.

5. There's some kind of important link between depression and folate balance. Various folate-related chemicals (eg l-methylfolate and s-adenosylmethionine) are effective antidepressants. Some studies show that people with depression sometimes have disrupted folate cycles, for example elevated homocysteine levels.

6. Electroconvulsive therapy ("shock therapy") is very effective at treating depression if it induces a seizure in the patient, so the increased activity from seizures must be helpful somehow.

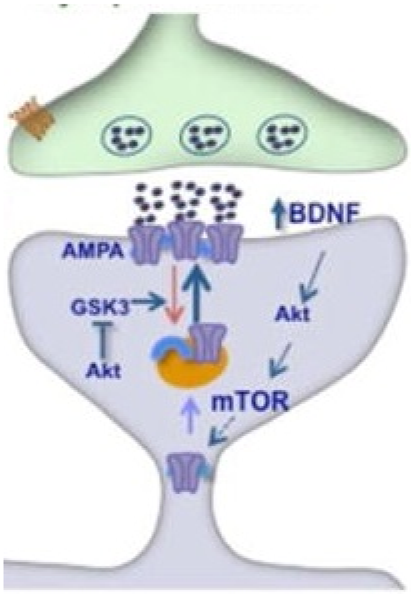
So if we wanted to know what depression *really* was, it might be promising to look for some process that seems to match depressive symptoms and affects/is affected by life adversity, inflammation, monoamines, glutamate, folate, and electricity.

Recently some people think they've found one. According to Duman's Neurobiology of Stress, Depression, and Rapid Acting Antidepressants, it's decreased synaptogenesis, and it's regulated by a protein complex called mTORC1.

Neurons communicate with other neurons through branches called dendrites and connections called synapses. Healthy neurons often create new dendrites and synapses to expand their network of connections and adjust to new information. The process of making new synapses is called "synaptogenesis", and it's common throughout the adult brain.

As mentioned above, depressed people have decreased volume in some brain areas. But in postmortem studies, they don’t actually have fewer cells in those areas. So it looks like maybe these neurons just have less synaptogenesis going on.

Synaptogenesis is partly controlled by a protein complex called mechanistic target of rapamycin complex 1 (mTORC1 to its friends). Like every other protein, mTORC is controlled by a giant mess of receptors and second messengers and intracellular signals with names like VDCC and GSK3.



People try to make this seem simple by displaying it as a system of billiard balls and tubes in a cute cartoon, but don’t be fooled – no human being has ever remembered any of it for more than two seconds.

The factors that affect synaptogenesis and mTORC are many of the same factors that affect depression. Let me count the ways:

1. Life adversity causes chronic stress, biologically represented by upregulation of the HPA axis and increased

corticosteroid production. A 2008 study finds that rats who are subjected to chronic stress develop atrophy of dendrites in their prefrontal cortex. Administering glucocorticoids directly mimicked some of these effects, suggesting that stress is a whole cocktail of things including glucocorticoids and other things. When humans take glucocorticoids (they're a useful medicine for various diseases) they tend to develop hippocampal atrophy and "simplification of dendrites" there, which I think is the same as decreased synaptogenesis. They also tend to get depressed – in some studies of Cushing's Syndrome (the medical name for the collection of bad things that happen when you take too much glucocorticoid medication), up to 90% of patients are depressed.

2. I didn't find the linked paper's attempt to link inflammation to synaptogenesis very convincing, but it looks like there's a little bit of research that has found that systemic inflammation decreases synaptogenesis. "Morphometric analysis of dendritic spines identified a period of vulnerability, manifested as a decrease in [dendritic] spine density in response to inflammation. The density of presynaptic excitatory terminals was similarly affected. When the systemic inflammation was extended from 24h to 8 days, the negative effects on the excitatory terminals were more pronounced and suggested a reduced excitatory drive." This seems pretty relevant.

3. Everyone used to think that traditional antidepressants like SSRIs worked by increasing serotonin (and so by extension depression must have something to do with low serotonin levels). But SSRIs increase serotonin very quickly (within hours) yet take months to work. Something longer-term must happen when serotonin levels have been increased for long enough. That something has now been pretty conclusively identified as an increase in brain-derived

neurotrophic factor (BDNF) – although I can’t find any good explanation of why increased serotonin should cause increased BDNF after a month. BDNF is a nerve growth factor – its main action is activating mTORC and telling nerve cells to grow more dendrites and synapses. And it’s most active in the cortex and hippocampus.

4. Ketamine affects the brain by either blocking NMDA receptors (boring traditional explanation), activating AMPA receptors (exciting new explanation), or possibly both (wishy-washy neoliberal compromise explanation). Duman et al are kind of ambiguous about which explanation they accept, but I think they present a theory where NMDA blockade *causes* AMPA activation, or something, which I’d never heard before. In any case, they present ample evidence that AMPA rapidly affects BDMF and dendritogenesis – for example, Positive AMPA Receptor Modulation Rapidly Stimulates BDNF Release And Increases Dendritic MRNA Translation. The “rapidly” part is important – the surprising thing about ketamine is how quickly it works compared to other antidepressants, so it’s exciting to find a theory that predicts this should happen.

5. I haven’t seen much attempt to fit folate into this theory, which is a shame. A quick Google search brings up a few people talking about how folate deficiency decreases neurogenesis in the hippocampus, which is sort of related.

6. Studies show that ECT increases BDNF levels and increases hippocampal volume, though I’m not sure exactly how or why giving someone a seizure should do that.

So the synapse hypothesis can unify at least five of the six lines of research into the causes of depression.

II.

My remaining skepticism is mostly based on a worry that anyone can do this with anything. The body is so interconnected, and there’s so much bad biology research out there, that I worry that if I said that the real cause of depression was, uh, thickness of the blood, I could find some way that all of those lines of research above affected blood thickness.

A quick demonstration: glucocorticoids can cause thicker blood, inflammation can cause thicker blood, SSRIs cause thinner blood, folate causes thinner blood. Huh, actually that’s kind of creepy.

My point isn’t that the (very respectable) academic research on depression is anywhere near this silly. It’s just to explain why I can hear a theory that seems to explain everything beautifully and my only reaction is “Eh, sounds like it has potential, let’s see what happens.”

Here are some of the things that confuse me, or that I hope get researched more in the future:

1. Why should decreased synaptogenesis cause depression, of all things? If you asked me, a non-neuroscientist, to guess what happens if the brain can’t create new synapses very well and loses hippocampal volume, I would say “your memory gets worse and you stop being able to learn new things”. But this doesn’t really happen in depression – even the subset of depressed people who get cognitive problems usually just have “pseudo-dementia” – they’re too depressed to put any effort into answering questions or doing intelligence tests. Why should decreased synaptogenesis in the hippocampus and prefrontal cortex cause poor mood, tiredness, and even suicidality? All that the Duman et al paper has to say about this is:

> This reduction in dendrite complexity and synaptic connections could contribute to the decreased volume of PFC and hippocampus observed in depressed patients. Moreover, loss of synaptic connections could contribute to a functional disconnection and loss of normal control of mood and emotion in depression (Fig. 1). In particular, the medial PFC exerts top down control over other brain regions that regulate emotion and mood, most notably the amygdala, and loss of synaptic connections from PFC to this and other brain regions could thereby result in more labile mood and emotion, as well as cognitive deficits.

…which sounds more like an IOU for a theory than anything really fleshed out.

2. Why can't we just give people BDNF for depression? I've been looking into this and it seems like the answer is something like "this works great if you cut open someone's skull and inject it directly into their brain, but most people aren't up for it" (the relevant studies were done in rats). But why can't it be given peripherally? Some studies suggest it's stable on injection and crosses the blood-brain barrier. Some people tried this in mice and got modest results, but why aren't people looking into it more?

3. Why does the body have so many "decrease synaptogenesis" knobs? That is, why go through the trouble to evolve all these chemicals and systems whose job is to tell your brain to decrease synapse formation so much that you end up depressed? Is there some huge problem with having too *much* synapse formation which the brain is desperately trying to avoid? For that matter, what *is* it like to have too much synapse formation? If it's the opposite of depression, it sounds kind of fun. If I got someone to open up my skull and inject a lot of BDNF, could I be really happy

and energetic all the time? How come all the good stuff is always reserved for rats?

4. Why is depression an episodic disease? That is, how come so many people get depressed for no reason, stay depressed for a few months to a few years, and then get better – only to relapse back into depression a few years later? If people get depressed because of some life stressor like a divorce, how come they don't get un-depressed once the life stressor goes away? Is depression some kind of attractor state? If so, why?

5. Why doesn't rapamycin cause depression? Remember, mTORC is "mechanistic target of rapamycin", so named because the drug rapamycin inhibits it. But we give people rapamycin for various things all the time, and depression isn't really known as a major side effect (even though IIRC it crosses the blood-brain barrier). If depression is really under the immediate control of mTORC, rapamycin should be the *most* depressive thing. Instead it's not obviously depressive at all.

6. How does bipolar disorder fit into all of this? Is mania the answer to my "what is it like to have too many synapses?" question from point (3)? If so, why do some people go back and forth between that and depression?

A lot of these questions could be answered in one stroke if we had a good evolutionary theory of depression. I'm skeptical that this exists – depression just seems too fitness-decreasing, and the various just-so stories people have come up with for why it might increase fitness in certain weird situations seem a little too convoluted. So it's not that I'm *expecting* some sort of evolutionary story to work out. Just noticing that, even if the synapse theory of pathophysiology turns out to be right, there’s still a lot more that needs to be explained.

Against Murderism

[Content warning: discussion of racism. Comments are turned off due to bad experience with the comments on this kind of material.]

I.

A set of questions, hopefully confusing:

Alice is a white stay-at-home mother who is moving to a new neighborhood. One of the neighborhoods in her city is mostly Middle Eastern immigrants; Alice has trouble understanding their accents, and when they socialize they talk about things like which kinds of hijab are in fashion right now. The other neighborhood is mostly white, and a lot of them are New Reformed Eastern Evangelical Episcopalian like Alice, and everyone on the block is obsessed with putting up really twee overdone Christmas decorations just like she is. She decides to move to the white neighborhood, which she thinks is a better cultural fit. Is Alice racist?

Bob is the mayor of Exampleburg, whose bus system has been losing a lot of money lately and will have to scale back its routes. He decides that the bus system should cut its least-used route. This turns out to be a bus route in a mostly-black neighborhood, which has only one-tenth the ridership of the other routes but costs just as much. Other bus routes, most of which go through equally poor mostly-white neighborhoods, are not affected. Is Bob racist?

Carol is a gay libertarian who is a two-issue voter: free markets and gay rights. She notices that immigrants from certain countries seem to be more socialist and more anti-gay than the average American native. She worries that they will become citizens and vote for socialist anti-gay policies. In order to prevent this, she supports a ban on immigration from Africa, Latin America, and the Middle East. Is Carol racist?

Dan is a progressive member of the ACLU and NAACP who has voted straight Democrat the last five elections. He is studying psychology, and encounters *The Bell Curve* and its theory that some of the difference in cognitive skills between races is genetic. After looking up various arguments, counterarguments, and the position of experts in the field, he decides that this is probably true. He avoids talking about this because he expects other people would misinterpret it and use it as a justification for racism; he thinks this would be completely unjustified since a difference of a few IQ points has no effect on anyone's basic humanity. He remains active in the ACLU, the NAACP, and various anti-racist efforts in his community. Is Dan racist?

Eric is a restauranteur who is motivated entirely by profit. He moves to a very racist majority-white area where white people refuse to dine with black people. Since he wants to attract as many customers as possible, he sets up a NO BLACKS ALLOWED sign in front of his restaurant. Is Eric racist?

Fiona is an honest-to-goodness white separatist. She believes that racial groups are the natural unit of community, and that they would all be happiest set apart from each other. She doesn't believe that any race is better than any other, just that they would all be happier if they were separate and able to do their own thing. She supports a partition plan that gives whites the US Midwest, Latinos the Southwest, and blacks the Southeast, leaving the Northeast and Northwest as multiracial enclaves for people who like that kind of thing. She would not use genocide to eliminate other races in these areas, but hopes that once the partition is set up races would migrate of their own accord. She works together with black separatist groups, believing that they share a common vision, and she hopes their countries will remain allies once they are separate. Is Fiona racist?

II.

As usual, the answer is that “racism” is a confusing word that serves as a mishmash of unlike concepts. Here are some of the definitions people use for racism:

1. Definition By Motives: An irrational feeling of hatred toward some race that causes someone to want to hurt or discriminate against them.

2. Definition By Belief: A belief that some race has negative qualities or is inferior, especially if this is innate/genetic.

3. Definition By Consequences: Anything whose consequence is harm to minorities or promotion of white supremacy, regardless of whether or not this is intentional.

Some thoughts:

Definition By Consequences Doesn’t Match Real-World Usage

I know that Definition By Consequences is the really sophisticated one, the ones that scholars in the area are most likely to unite around. But I also think it’s uniquely bad at capturing the way anyone uses the word “racism” in real life. Let me give four examples.

First, by this definition, racism can never cause anything. People like to ask questions like “Did racism contribute to electing Donald Trump?” Under this definition, the question makes no sense. It’s barely even grammatical. “Did things whose consequence is to harm minorities whether or not such harm is intentional contribute to the election of Donald Trump?" Huh? If racism is just a description of what consequences something has, then it can't be used as a causal explanation.

Second, by this definition, many racist things would be good. Suppose some tyrant wants to kill the ten million richest white people, then redistribute their things to black people. This would certainly challenge white supremacy and help minorities. So by this definition, resisting this tyrant would be racist. But obviously this tyrant is evil and resisting him is the right thing to do. So under this definition, good policies which deserve our support can nevertheless be racist. "This policy is racist" can no longer be a strong argument against a policy, *even when it's true*.

Third, by this definition, it doesn't make a lot of sense to say a particular person is racist. Racism is a property of actions, not of humans. While there are no doubt some broad patterns in people, the question "Is Bob racist?" sounds very odd in this framework, sort of like "Does Bob cause poverty?" No doubt Bob has done a few things which either help or hurt economic equality in some small way. And it's possible that Bob is one of the rare people who organizes his life around crusading against poverty, or around crusading against attempts to end poverty. But overall the question will get you looked at funny. Meanwhile, questions like "Is Barack Obama racist?" should lead to a discussion of Obama's policies and which races were helped or hurt by them; issues like Obama's own race and his personal feelings shouldn't come up at all.

Fourth, by this definition, it becomes impossible to assess the racism of an action without knowing all its consequences. Suppose the KKK holds a march through some black neighborhood to terrorize the residents. But in fact the counterprotesters outnumber the marchers ten to one, and people are actually reassured that the community supports them. The march is well-covered on various news organizations, and outrages people around the nation, who donate a lot of money to anti-racist organizations and push for stronger laws against the KKK. Plausibly, the net consequences of the march were (unintentionally) very good for black people and damaging to white supremacy. Therefore, by the Sophisticated Definition, the KKK marching the neighborhood to terrorize black residents was not racist. In fact, for the KKK *not* to march in this situation would be racist!

So Definition By Consequences implies that racism can never be pointed to as a cause of anything, that racist policies can often be good, that nobody "is a racist" or "isn't a racist", and that sometimes the KKK trying to terrorize black people is less racist than them *not* trying to do this. Not only have I never heard anyone try to grapple with these implications, I see no sign anyone has ever thought of them. And now that I've brought them up, I don't think anyone will accept them as true, or even worry about the discrepancy.

I think this is probably because it's a motte-and-bailey, more something that gets trotted out to win arguments than anything people actually use in real life.

Definition By Belief Is A Mess

Is it racist to believe that Mexicans are poorer than white people? After all, being poor is generally considered bad, so you're attributing a bad quality to a minority group. What if you add "Mexicans are only poor because of being oppressed and discriminated against?"

Is it racist to believe that Mexicans are more criminal than white people? What if you add “Mexicans are only criminal because their culture was shaped by the experience of oppressive Spanish colonization, which left deep scars on their national psyche”?

Is it racist to believe that Muslims commit more terrorism than white people? What if you’ve done a lot of calculations of per capita terrorist attacks and you can quote exact numbers that prove your point?

Is it more or less racist if then you add “…but this is because Islam is a violent religion that encourages murder, and has nothing to do with the genetics of Middle Eastern people”?

Is it racist to believe that Pygmies are shorter than white people?

But None Of That Really Matters, Because In Real Life, Definition By Motive Usually Trumps Definition By Belief

After the London attacks, I heard someone ask “Do you have to be a racist to want to restrict immigration from Muslim countries? Or can you just be really worried about the terrorism risk?”

A lot of people responded. Some of them said no, it was perfectly reasonable to be worried about terrorism. Other people said that concern about terrorism was just a smokescreen, that people *said* they were just concerned about terrorism, but *actually* that was just a way to cover up their racism.

Think about how confusing this is. It’s positing two different things. First, a belief that Muslims are often terrorists and so we should crack down on them. And second, racism. These things are considered *opposing* explanations, such that if we believe the first one, we can dismiss the second – or, if we admit the second, that proves the first was claimed dishonestly. Under Definition By Belief, it's really weird.

(compare: "I believe Jesus Christ is the Son of God and those who believe in Him gain eternal life." "No, you're just using that as a smokescreen to cover up that you're Christian!")

The only way I can make sense of this argument is to think of it as Definition By Motive trumping Definition By Belief. The first person is stating a belief that Muslims are more likely to be terrorists. The second person is questioning whether their motivation for restricting immigration is really this belief (in which case it would be ok) or if they're motivated by an irrational hatred of minorities (in which case it would be racism).

Definition By Motive can even trump Definition By Belief when we're talking about innate/genetic difference. Consider Charles Murray saying that he believes black people are genetically less intelligent than white people. Some of Murray's critics object that this should be suppressed, even if true, because it could be used to justify racism.

Under Definition By Belief, this makes no sense. Imagine Murray was a geologist, pointing out that Antarctica contained mostly sedimentary rock. His critics object "This should be suppressed, even if true, because it could be used to justify believing that Antarctica contains mostly sedimentary rock!" Huh?

It makes more sense if we think of it as being about Definition By Motive. Then the critics are saying that if we find that minority groups are genetically worse in some way (ie racist beliefs are true), we should suppress that lest it be used to justify people’s irrational feelings of hatred for members of other races (ie justify racist motives).

Definition By Motive Makes Sense Of All Of The Above Examples And Basically Matches Most Real-World Usage

Definition By Motive fits the first example. When we ask “Was racism responsible for Trump’s election?” we mean “Did people elect Trump because they irrationally hated minorities and wanted to discriminate against them?”

It fits the second example. When we say that it wouldn’t be racist to resist a tyrant who wants to kill whites, we mean that such resistance is a good policy, which would be pursued for reasons other than just irrationally hating minorities and wanting to discriminate against them.

It fits the third example. When we claim a specific person (Bob, Barack Obama) is racist, we mean that they irrationally hate minorities and want to discriminate against them.

It fits the fourth example. When we say that the KKK marching through a black neighborhood to terrorize people is racist regardless of its consequences, we mean that it’s motivated by an irrational hatred of minorities and desire to discriminate against them.

It fits the fifth example. When we ask whether it’s racist to believe Mexicans are poorer than whites, we’re asking whether someone would only say that because they irrationally hate Mexicans and want to discriminate against them. But most of the time people making that claim are trying to point out inequalities and help Mexicans. So it isn’t racist.

It fits the sixth example. Somebody who believes that Mexicans are more criminal than white people *might* just be collecting crime stats, but we're suspicious that they might use this to justify an irrational hatred toward Mexicans and desire to discriminate against them. So it's potentially racist, regardless of whether you attribute it to genetics or culture.

It fits the seventh example. It's probably not racist to believe that Muslims commit more terrorism than white people, since this seems to be a true or at least plausible claim, but if people talk about it too much it's worrying that maybe they're trying to justify their irrational hatred of Muslims and desire to discriminate against them.

It fits the eighth example. It's probably not racist to believe that Pygmies are shorter than white people, because it's obviously true and you would believe it whether you had an irrational hatred of Pygmies or not. Also, no one cares how tall anybody is.

It fits the ninth example. When people ask whether immigration restricts are *really* due to fear of terrorism vs. racism, they're asking whether people who claim to be concerned with terrorism actually just irrationally hate minorities and want to discriminate against them.

And it fits the tenth example. When people say that Charles Murray's claims about genetics might be used to "justify racism", they mean that if you irrationally hate minorities and want to discriminate against them, you could use his claims about genetics as a justification for why your position makes sense.

Overall We Probably Use A Combination Of All Of These, Weighted In Favor Of Definition By Motives

I designed the discussion questions to be situations where Definition By Motive clearly didn’t apply, but one or both of the other definitions clearly did. I imagine some people stuck to their guns, went Definition By Motive all the way through, and said none of the people in the vignettes were racist. I imagine other people used one of the other two definitions, or a different definition of their own, and were able to navigate all of the objections and counterexamples down here in Part II successfully. But I think most people found a couple of inconsistencies, aren’t really sure what to do with them, and are just sort of echoing the Supreme Court’s view of pornography: “I’ll know it when I see it.”

This is natural. I’m not trying to say that Definition By Motives is the one “real” definition. All of our word usage is a mess; we hardly ever use anything simply or consistently, let alone a complicated word like “racism”. In reality we go back and forth among all of these, proving that something is racist using one definition, then applying the consequences of another definition, switching from very strict to very loose based on whether or not it’s something we like. All of this is totally normal.

But in this case it’s kind of likely to lead to disaster.

III.

A digression, from an alternative universe.

“Murderism” is the ideology that murdering people is good and letting them live is bad. It’s practically omnipresent: 14,000 people are murdered in the US each year. That’s a lot of murderists, and a testament to the degree to which our schools teach murderist values.

But not all murderism is that obvious. For years, people have been pushing “soft-on-crime” policies that will defund the police and reduce the length of jail sentences – inevitably increasing the murder rate. Advocates of these policies might think that just because they’re not gangsters with knives, they must not be murderists. But anybody who supports murder, whether knife-wielding gangster or policy analyst – is murderist and responsible for the effects of their murderism.

Our two major parties have many differences – but both are united in their support for murderism. Republicans push murderist policies like the invasion of Iraq, which caused the murder of thousands of Iraqis. Democrats claim to be better, but they support openly murderist ideas like euthanasia, promoting the killing of our oldest and most vulnerable citizens. There’s no party in Washington that’s willing to take a good look at itself and challenge the murderist ideals that our political system is built on.

Murderism won’t stop until people understand that it’s not okay to be murderist. So next time you hear people opposing police militarization, or speaking out in favor of euthanasia – tell them that that’s murderism and it’s not okay.

…okay, done. Back in our own universe, we recognize that “murderism” is silly: it confuses cause and effect.

Murder is usually an effect of a strategy pursued for other reasons. The drug dealer who wants to keep rivals off his turf, the soldier who wants to win a war, the gangster who wants to get rid of inconvenient witnesses. If you want to stretch it, add the neocon who wants to “liberate” foreign countries, the cancer patient who wants to “die with dignity”, or the activist who wants to keep people out of jail.

But except in maybe the most deranged serial killers, it’s never pursued because of an inherent preference for murder. Most murderers would probably prefer not to have to kill. If the drug dealer could protect his business equally well by politely requesting people stay off his territory, that would be much easier. If the soldier could win his war without bloodshed, so much the better for everybody. Murder is an effect of other goals – sometimes base, sometimes noble – and the invocation of “murderism” only serves to hide these goals and conflate different actions into a single meaningless category.

Talking about murderism isn’t just uninformative, it’s actively confusing. If you believed that gangsters killed their rivals because of murderism, then there’s no point in examining how poverty interacts with gang membership, or whether the breakdown of law forces people to form gangs to defend themselves. The problem is just that gangsters have murderist values. It should be solved by censoring the works of philosopher David Benatar, who writes about how being alive is bad and it’s morally better not to exist at all. Or by banning high school Goths, whose pro-death aesthetic makes murderism seem cool to teens and causes them to harbor murderist thoughts as adults.

Talk about murderism is obviously confused. But it’s the same confusion between the Definition By Consequences versus the Definition By Motive that we saw was a hallmark of racism.

IV.

Belief in murderism creates a hostile and ineffective society whose weird beliefs can only be countered by accepting that murder is rarely a terminal goal, but a usually result of strategies pursued for other reasons. We accept that having a terminal goal of killing people seems so awful, inhuman, and incongruous with the sort of decent humans we all know – that it's a very strange explanation to even consider when other, better ones are available. We can apply the same analysis to racism. The discussion questions in Part I already started the process, but we can go further.

I'm not just making the argument "lots of things aren't really racist". I can't do much about how you choose to define words, plus it's doomed to fail anyway. Imagine having to convince someone that a guy who committed homicide "isn't really murderist". Doesn't sound like the most winnable fight.

And if you only break down non-racist things into non-racist motives, what reward shall you have? Do not even the scribes and the Pharisees do the same? I say unto you, if you want to be righteous, look for the non-racist motives in *actually* racist things.

What does that mean?

Consider some business, let's say a daycare center, that we know discriminates against black job-seekers. If we ask them why, they say "Because black people are criminal". This sounds like just about the most typical and obvious example of racism possible.

But there's actually a lot of really good scholarship on this exact situation, and it helps provide a different perspective. It starts like this – a while ago, criminal justice reformers realized that mass incarceration was hurting minorities' ability to get jobs. 4% of white men will spend time in prison, compared to more like 16% of Hispanic men and 28% of black men. Many employers demanded to know whether a potential applicant had a criminal history, then refused to consider them if they did. So (thought the reformers) it should be possible to help minorities have equal opportunities by banning employers from asking about past criminal history.

The actual effect was the opposite – the ban "decreased probability of being employed by 5.1% for young, low-skilled black men, and 2.9% for young, low-skilled Hispanic men."

In retrospect, this makes sense. Daycare companies really want to avoid hiring formerly-imprisoned criminals to take care of the kids. If they can ask whether a certain employee is criminal, this solves their problem. If not, they're left to guess. And if they've got two otherwise equally qualified employees, and one is black and the other's white, and they know that 28% of black men have been in prison compared to 4% of white men, they'll shrug and choose the white guy.

Is this racist? Is this "statistical discrimination"? Describe it with whatever word you want. The point is that they have understandable motives (don't hire criminals to take care of the kids), accurate beliefs, and in their shoes you might do the same. More important, once you give them the tools they need to solve their problems without racial discrimination – you let them see applicants' criminal histories – they have no further desire to discriminate and your problem is solved.

If you tried to solve this by sending these people to sensitivity training, you would fail. If you tried to solve this by firing these people, then the people who replaced them would have the same incentives, and you would fail again. If you try to solve it by realizing that racial animus has no role at all in this scenario, and daycare owners just want to do what's best for their kids, then you can provide them with the tools they need to do that, and solve the racial discrimination at the same time.

Okay, fine. Harder example. Let’s take, uh, some guy who’s always ranting about how the Jews secretly control the world. They have underground tunnels where they have their secret Zionist meetings and talk about how they’re going to stick it to the Christians. Every major war and economic downturn has been caused by this. Are we allowed to treat *this guy*‘s racism as being a conceptual primitive that doesn’t need further breaking down?

I actually knew a guy like this. He was a schizophrenic patient in the mental hospital where I work. Overall I found it a nice break from the tedium of CIA-conspiracy folk, alien-conspiracy folk, and white-people-conspiracy folk (remember, this is Detroit).

Am I saying everyone like this is schizophrenic? Not diagnosably, no. But I notice that there are a lot of not-diagnosably-schizophrenic people who believe in the Illuminati, the New World Order, the Freemasons, and – yes – lizardmen. Is it really so outlandish to say that the same faulty reasoning that concludes that Freemasons run the world could conclude that Jews run the world, and for the same reasons? Does it really make sense to just blow one off as paranoid conspiracy-mongering, and the other as originating from a *completely different* process called “anti-Semitism” or “racism”? Remember, “healthy” people with paranoid and conspiratorial beliefs have the same kind of fronto-striatal prediction error signal that schizophrenics do, only less so, suggesting that their odd ideas probably come from the same kind of disturbed reasoning process.

“Are you saying that anti-Semitism literally plays no role in their theory about the Elders of Zion”? Again, call it what you want. I’m saying that by totally ignoring the anti-Semitic aspect, I was able to successfully treat this guy with Seroquel, whereas if you tried to read him Elie Wiesel books, he’d still be in that psych ward today.

Fine. Schizotypal conspiracy-mongerers are a noncentral example anyway. What about, I don’t know, rural Republicans in South Carolina who wave the Confederate flag all the time and think blacks and immigrants are ruining the country.

Here I would point out that this is pretty much the demographic that elected Nikki Haley (birth name, Nimrata Randhawa; daughter of two Punjabi immigrants) as governor, and that supports her so fervently that she remains one of the most popular politicians in the country. Also the demographic that loved Ben Carson, making him the only candidate to briefly displace Trump for first place in the 2016 Republican primary polls. One plausible explanation is that the South Carolinians don’t like blacks and immigrants because they view them as having foreign values – specifically, Blue Tribe values (it may be relevant here that 90%+ of blacks usually vote Democrat). If someone like Nikki Haley or Ben Carson proves that they share Red beliefs, they become part of the tribe and will be fiercely defended. Maybe this is more like the daycare situation than it looks – people using race as a proxy for something they care about, until they get direct information.

To be clear – I am not saying that racism doesn’t exist, I’m not saying that we should ignore racism, I’m not saying that minorities should never be able to complain about racism. I’m saying that it’s very dangerous to treat “racism” as a *causal explanation*, that it might not tell you anything useful about the world, and that’s a crappy lever to use if you want to change behavior.

And I’m not saying that it’s not useful to think of some of these things as places where there’s an opportunity for racial change. If a daycare owner is really interested in redressing racial inequality, they can hire minorities even if it’s against their incentives and self-interest (although it’s unclear why the owner should prefer that opportunity to other opportunities, like donating some of their profits to the NAACP.)

And I’m not saying that there will never be a case that’s impossible to break down into non-racist motives. Heck, I’m not even saying there aren’t some honest-to-goodness murderists out there. But I am saying we should at least try. Not because it’s necessarily costless. Not because there isn’t a risk of false negatives.

We should try because it’s the only alternative to having another civil war.

V.

Arnold Kling likes to talk about how political groups are divided by different “languages”, different schemata for understanding the world that make it difficult to talk across political divides.

Jonathan Haidt accepts the premise but challenges the symmetry; his experiments ask liberals and conservatives to fill out questionnaires about their values, then to predict how someone from the opposite tribe would fill out the questionnaire. He finds that conservatives are able to predict liberals’ answers just fine and seem to have a pretty good understanding of their worldviews, but that liberals have *no idea* how conservatives think or what they value.

James Scott, as channeled by Lou Keep, draws the asymmetry a little differently. He says that the process of development, especially state-building and the switch from traditional to market economies, creates a pressure for “legible” language that renders entire classes of problems very difficult to talk about. This creates an asymmetry between an elite plugged into the global market structure whose concerns make perfect sense (“If we do this, GDP will go up 3% and we can build more roads!”) and the masses left behind whose concerns seem pointless and vague (“I feel like something important disappeared when we turned everything into a commodity”). Keep then proposes a very loose mapping onto cosmopolitan neoliberal Clintonites versus undereducated “I’m angry about losing my traditional culture” Trumpists.

There are a bunch more frameworks like this, but they all share the common warning that cross-cultural communication is *really hard*, and so a lot of the concerns of people who aren’t like us will probably sound like nonsense. And most of them say that our demographic – well-educated people proud of our commitment to logic and reason – are at especially high risk of just dismissing everyone else as too dumb to matter. The solution is the same as it’s always been: hard work, renewed commitment to liberal values, and a hefty dose of the Principle of Charity.

Racism-as-murderism is the opposite. It’s a powerful tool of dehumanization. It’s not that other people have a different culture than you. It’s not that other people have different values than you. It’s not that other people have reasoned their way to different conclusions from you. And it’s not even that other people are honestly misinformed or ignorant, in a way that implies *you* might ever be honestly misinformed or ignorant about something. *It’s that people who disagree with you are motivated by pure hatred, by an irrational mind-virus that causes them to reject every* *normal human value in favor of just wanting to hurt people who look different from them.*

This frees you from any obligation to do the hard work of trying to understand other people, or the hard work of changing minds, or the hard work of questioning your own beliefs, or the hard work of compromise, or even the hard work of remembering that at the end of the day your enemies are still your countrymen. It frees you from any hard work at all. You are right about everything, your enemies are inhuman monsters who desire only hatred and death, and the only "work" you have to do is complain on Twitter about how racist everyone else is.

And I guess it sounds like I'm upset that we're not very good at solving difficult cross-cultural communication problems which require deep and genuine effort to understand the other person's subtly different value system. I'm not upset that we can't solve those. Those are hard. I'm upset because we're not even at the point where someone can say "I'm worried about terrorism," without being forced to go through an interminable and ultimately-impossible process of proving to a random assortment of trolls and gatekeepers that they *actually* worry about terrorism and it's not just all a ruse to cover up that they secretly hate everyone with brown skin. I'm saying that when an area of the country suffers an epidemic of suicides and overdoses, increasing mortality, increasing unemployment, social decay, and general hopelessness, and then they say they're angry, we counter with "Are you really angry? Is 'angry' just a code word for 'racist'?" I'm saying we're being challenged with a moonshot-level problem, and instead we're slapping our face with our own hand and saying "STOP HITTING YOURSELF!"

People talk about “liberalism” as if it’s just another word for capitalism, or libertarianism, or vague center-left-Democratic Clintonism. Liberalism is none of these things. Liberalism is a technology for preventing civil war. It was forged in the fires of Hell – the horrors of the endless seventeenth century religious wars. For a hundred years, Europe tore itself apart in some of the most brutal ways imaginable – until finally, from the burning wreckage, we drew forth this amazing piece of alien machinery. A machine that, when tuned just right, let people live together peacefully *without* doing the “kill people for being Protestant” thing. Popular historical strategies for dealing with differences have included: brutally enforced conformity, brutally efficient genocide, and making sure to keep the alien machine tuned *really really carefully*.

And when I see someone try to smash this machinery with a sledgehammer, it’s usually followed by an appeal to “but racists!”

You say we must protect freedom of speech. But would you protect the free speech *of racists?*

You say people shouldn’t get fired for personal opinions that don’t affect their work. But would you let *racists* keep their jobs?

You say we try to solve disagreements respectfully through rational debate. But would you try to rationally debate *racists?*

You say people should be allowed to follow their religion without interference. But what if religion is just a cover *for racism?*

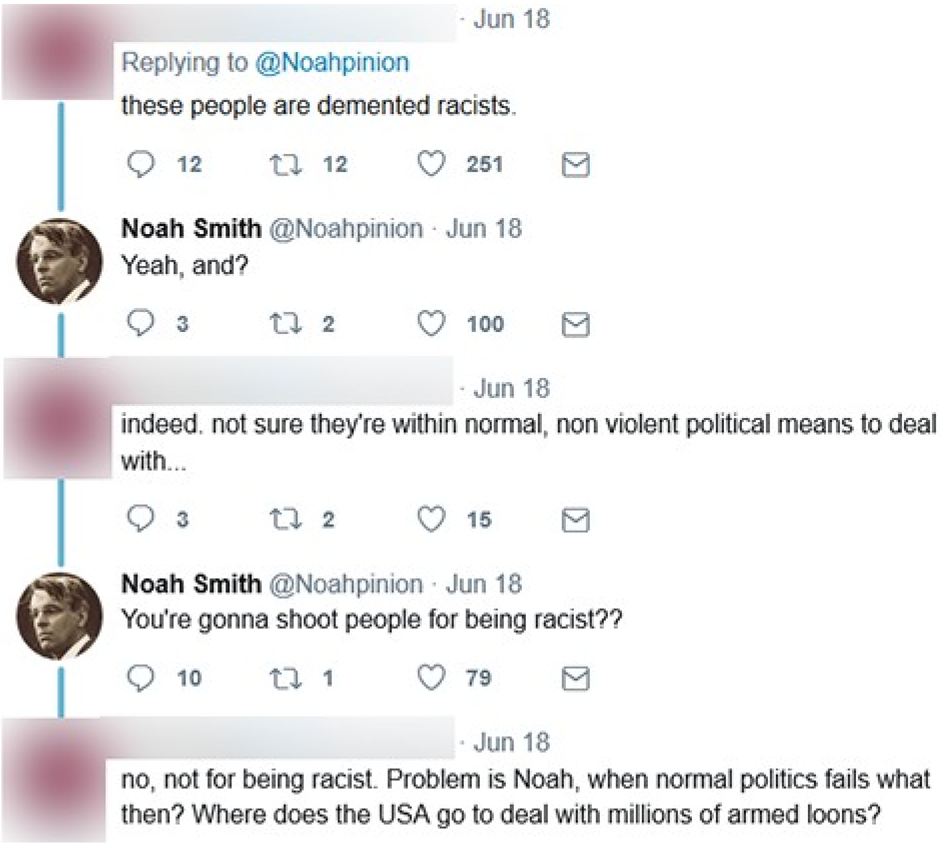
You say we need to understand that people we disagree with can sometimes have some good points. Are you saying we should try to learn things from *racists*?

You say there’s a taboo on solving political disagreements by punching people. Are you saying that we can’t punch *racists?*

The argument goes: liberalism assumes good faith and shared values. It assumes that, at the end of the day, whether you’re Catholic or Protestant, you can still be a basically good person. You can compartmentalize a few special beliefs relating to the Pope, and your remaining differences can be dissolved by the universal solvent of Reason. After everyone does this, you can invoke the wisdom of crowds via a popular election, and even if you don’t like the results you can at least understand where the other side is coming from. Some people prefer liberty to safety, other people prefer safety to liberty, but if the voters choose the wrong one then at least they’ve erred in an understandable way by preferring one real value to another.

But if there’s some group out there who aren’t connected to normal human values at all, some group that’s deliberately rejected reason; if they’re willing to throw liberty *and* safety under the bus in pursuit of some kind of dark irrational hatred which is their only terminal goal – then the whole project falls apart. Dialogue based on mutual trust may be all nice and well when it’s supposed to help us choose the optimal balance between liberty and safety, but if you give a platform to the people whose only value is hatred, then they’re just screwing over everybody.

A few days ago, Noah Smith posted on Twitter about hearing some people say racist things. The comments went like this:



· Jun 18

Replying to @Noahpinion

these people are demented racists.

Noah Smith @Noahpinion · Jun 18

Yeah, and?

· Jun 18

indeed. not sure they're within normal, non violent political means to deal with...

Noah Smith @Noahpinion · Jun 18

You're gonna shoot people for being racist??

· Jun 18

no, not for being racist. Problem is Noah, when normal politics fails what then? Where does the USA go to deal with millions of armed loons?

Ah well. They said a racist thing. Guess we've got to kill them.

And *I agree with this chain of logic*. Using violence to enforce conformity to social norms has always been the historical response. We invented liberalism to try to avoid having to do that, but you can't liberalism with people who refuse reason and are motivated by hatred. If you give the franchise to green pointy-fanged monsters, they're just going to vote for the "Barbecue And Eat All Humans" party. If such people existed and made up a substantial portion of the population, liberalism becomes impossible, and we should go back to just using violence to enforce our will on the people who disagree with us. Assuming they don't cooperate with our strategy of violently suppressing them, that means civil war.

I don’t want civil war. I want this country to survive long enough to be killed by something awesome, like AI or some kind of genetically engineered superplague. Right now I think going out in a neat way, being killed by a product of our own genius and intellectual progress – rather than a product of our pettiness and mutual hatreds – is the best we can hope for. And I think this is attainable! I think that we, as a nation and as a species, can make it happen.

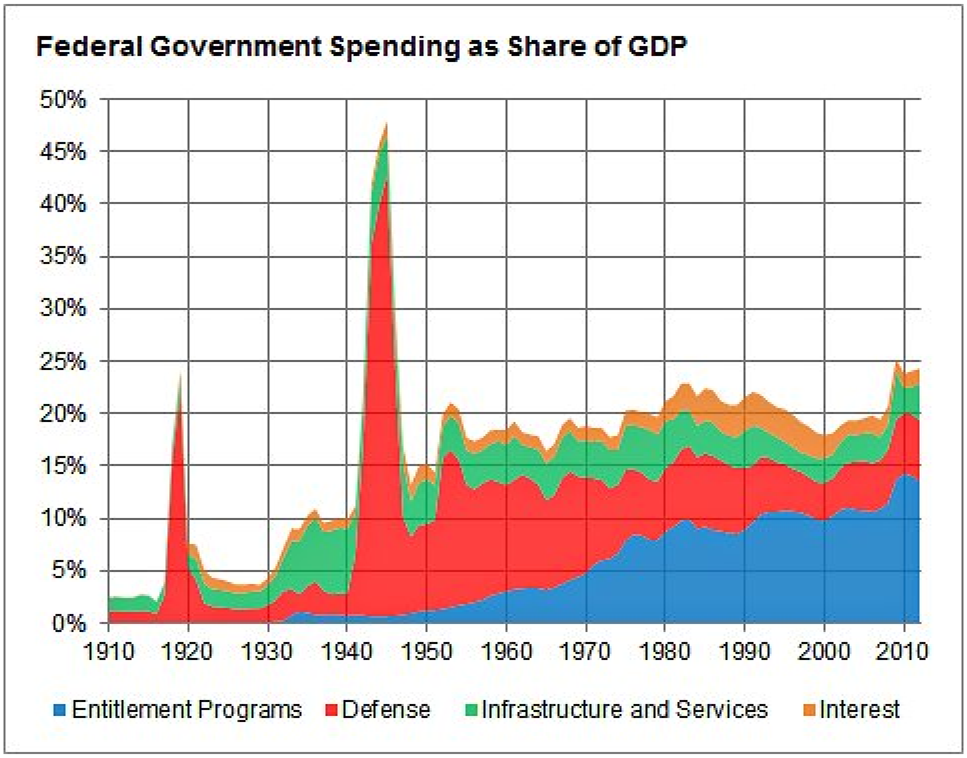
But it starts with rejecting the “murderism” framework. Rejecting the choice to attribute whatever we disagree with to murderism, *even if it is murderist*, and instead trying to trace it back to root causes that make sense that and humanize the people involved. Working to find the reasons liberalism is possible, rather than the reasons it isn’t. Unless we can do that, semantic confusion and our political polarization are going to build off each other in a vicious cycle into who knows where.

To Understand Polarization, Understand Conservativism’s Failures

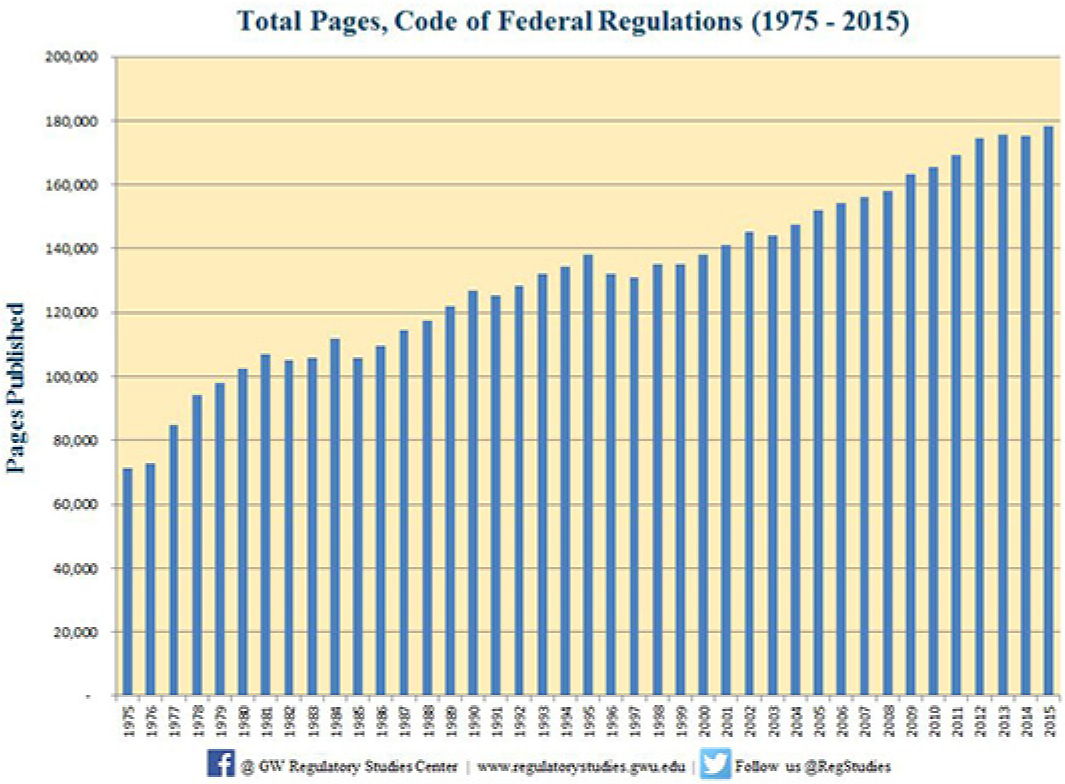
Earlier today I talked about one reason for increased polarization on the Democratic side. Now I want to match it with one reason the Republicans have even more of a problem.

Ask anyone what Republicans want, and they’ll say things like “smaller government”, “fewer regulations”, and “less welfare state”.

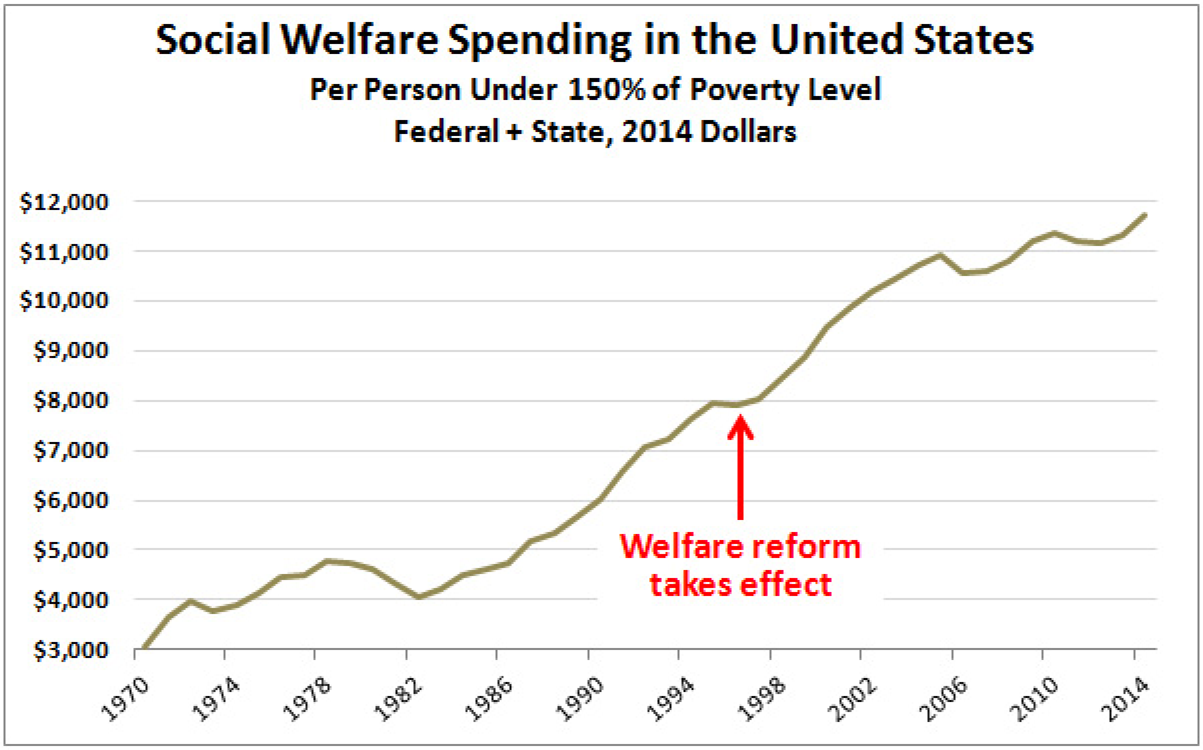
Meanwhile, here are some graphs showing how they’re doing (disclaimer: graphs like this are very dangerous, and I can only plead that I’ve seen numbers like these from enough sources that I think they have some contact with reality):



Source: https://fivethirtyeight.blogs.nytimes.com/2013/01/16/what-is-driving-growth-in-government-spending/



Source: https://regulatorystudies.columbian.gwu.edu/reg-stats



Source: http://www.motherjones.com/kevin-drum/2016/02/heres-why-bernie-sanders-doesnt-say-much-about-welfare-reform/

Apparently not so good.

This is true even though this is a historic apex of Republican power. They control the House, the Senate, the Presidency, 66% of state governorships, 68% of relevant state legislatures, and are kind of tied-ish for control of the Supreme Court. They've been two of the last four Presidents, and controlled Congress more often than not during that period.

This is really strange. Whatever they wanted, they should have been able to get. Who's going to stop them? *Democrats?* Don't make me laugh.

But in fact, we mostly kept getting bigger government, more regulations, and a bigger welfare state.

My guess is this is a larger-scale version of what I talked about in Considerations On Cost Disease. Various secular trends make everything more expensive and worse, which means government has to spend more money and regulation to get the same level of services, which means government gets bigger. There’s no easy way to stop this except to understand cost disease (which people don’t) or to drastically cut the level of services and admit it will keep getting worse (which politicians are scared of doing on their watch). This is not really the Republicans’ fault.

But Republican voters don’t know that.

All they see is candidates running for office on a platform of small government and less regulation. Then they win, they’ve got a huge majority and a great mandate, and at the end of their term government is as big as always and there are more regulations than ever.

And if maybe you’re not that sophisticated about these kinds of things, you think – these guys betrayed me. They’re Republicans in name only. They were corrupted by Washington. The liberal media finally got them. They’re weak and they caved as soon as the Democrats called them mean names. What we need are some *real* Republican candidates, ones who are *actually* willing to stand up to the establishment.

Then you elect the Real Republican Candidates, the Tea Party or whoever, and the same thing happens. Because we’re talking about secular trends and not about anything that Congress can easily affect.

So then the voters think *they’re* frauds too, and they get defeated in the primaries by other people who are even more Tea Party than they are, people who can say oh yeah, those Tea Party people were fake, but *we* have the necessary commitment to go to Washington and not cave in immediately.

This will never work. But the superficial logic of “Republicans are powerful enough to get whatever they want, we don’t have small government, therefore the current crop of so-called Republicans didn’t really want small government enough” is convincing. You end up with a signaling spiral where everyone’s in an arms race to show that they’re not actually the craven compromisers that people will inevitably assume them to be. That means hyperpartisanship and refusal to compromise on anything.

I’m talking about this as a Republican problem, but I think it’s a general issue whenever people have unrealistic expectations, ie always. The more our hopes diverge from the possible, the more we’ll reject all existing governments in favor of stronger and stronger forms of extremism.

Conversation Deliberately Skirts The Border Of Incomprehensibility

From lierdumoa.tumblr.com:

> To my friends on the [autism] spectrum, let me explain to you an unspoken social rule that possibly nobody has ever explained to you before
>
> If a neurotypical asks you, “What game are you playing?” they’re not asking you to describe the game.
>
> They’re asking you if they can play too.
>
> If a neurotypical asks you, “What are you watching?” they’re not asking you to explain the plot of the movie/tv show to them.
>
> They’re asking if they can watch it with you.
>
> When neurotypicals ask you “What are you doing?”
>
> **What you think they’re asking:** “Please explain to me what you are doing.”
> **What they’re actually asking:** “Can I join you?”
>
> Now here’s the really fucked up part. If you start explaining to them what you’re doing? They will interpret that as a rejection.
>
> **What you think you’re saying:** [the answer to their question]
> **What they think you’re saying:** This is an elite and exclusive activity for a level 5 friend and you are a level 1 acquaintance. You are not qualified to join me because you don’t know all this stuff. Go away.

> This is why neurotypicals think you’re being cold and antisocial.
>
> IT’S ALL A HORRIBLE MISCOMMUNICATION.

I don’t think this is always true – and when it is I would describe it as more of an open-ended attempt to start a fun conversation than a demand for participation – but I agree that it’s not just a straightforward request for information.

And there was some interesting discussion about this on Autistic Tumblr, which centered around: why would someone do this? Why can’t people just say what they mean?

And the best answer I saw – sorry, I can’t find it right now – explained that people were trying to spare their friends the burden of rejecting them. Say Alice is reading a book, and Bob asks “Hey, do you want to talk about that book?” Maybe Alice doesn’t want to talk about it. But the following conversation…

Bob: Hey, you want to talk to me about that book?
Alice: No

…sounds really rude. So by Bob saying his line, he’s putting a lot of subtle pressure on Alice to agree. Bob is a good person and he doesn’t want to do that. So instead he asks “Hey, what are you reading?”

Bob: Hey, what are you reading?
Alice: Not much. Just some random novel.
Bob: Oh, well, enjoy!

Or:

Bob: Hey, what are you reading?
Alice: Oh! It’s really interesting! It’s this book where Apollo

8 crashes into the celestial sphere surrounding the world, and suddenly everything reverts to kabbalistic Judaism…
Bob: Sounds neat! What happens next?

Here Alice either has an opportunity to signal that she wants to continue the conversation, *or* to reject Bob while maintaining plausible deniability that she's doing that.

(The beauty of this theory is ruined only by the fact that half the time this happens in real life and I say "Just some random novel," Bob actually answers "Oh! What *kind* of random novel?" and then I say "Oh, nothing really", and Bob says "Come on! Something has to happen!" and then I start despairing that anything about social interaction can ever work at all. I don't know. Maybe *Bob* is autistic.)

What I find interesting about "plausible deniability" explanations is that Bob has to operate as close to the border of "inscrutable confusingness" as possible without crossing it. He wants Alice to know he wants to talk to her, but he doesn't want Alice to know that he knows she knows he wants to talk to her (I'm being very deliberate in putting the word "know" exactly three times there rather than just using a vague phrase like "common knowledge"). As long as Alice doesn't know he knows she knows he wants to talk to her, Alice can give a non-answer, pretending that she believes Bob will believe that she just didn't realize he wanted to talk to her.

And this sort of weird common-knowledge-denial-process only works if you're skirting the border of incomprehensibility, hitting a sweet spot where you think the other person *will* understand, but it's also just barely plausible that the other person *wouldn't* understand. If you say something the other person would *definitely* understand, then the game is up. Given some sort of natural variation in how good people are at understanding cues, your best bet is to send a cue that will fail a small but non-zero percent of the time.

But if there are people who are unusually bad at understanding social cues, like autistic people, then any cue calibrated to be on the exact border of neurotypical understanding is likely to fail for them more often than not.

I don’t know how common this pattern is. Making requests seems like a pretty good example. Flirting seems to be centered upon this kind of thing. I’m not sure what else is involved, but I bet it’s a lot.

A while ago I quoted a paper by Lawson, Rees & Friston about predictive-processing-based hypotheses of autism. They said:

> This provides a simple explanation for the pronounced social-communication difficulties in autism; given that other agents are arguably the most difficult things to predict. In the complex world of social interactions, the many-to-one mappings between causes and sensory input are dramatically increased and difficult to learn; especially if one cannot contextualize the prediction errors that drive that learning.

And I was really struck by the phrase “arguably the most difficult thing to predict”. *Really?* People are harder to predict than, I don’t know, the weather? Weird little flying bugs? Political trends? M. Night Shyamalan movies? And of all the things about people that should be hard to predict, *ordinary conversations?*

And I think part of the answer might be: ordinary conversations are hard to predict *because they’re designed to be so*. Conversation norms are anti-inductive. Like

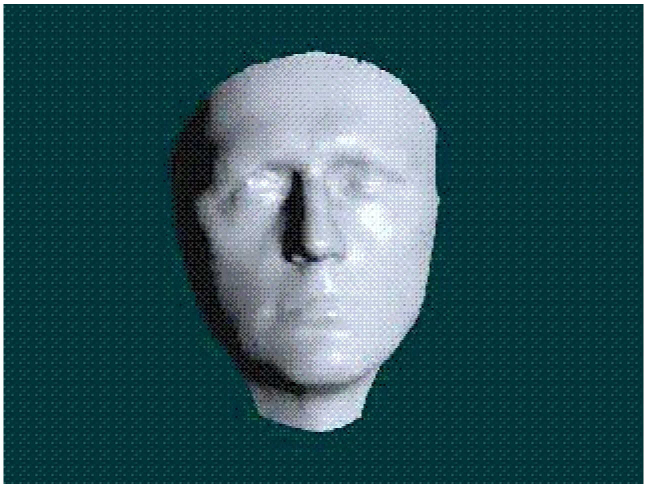
Douglas Adams’ conception of the universe, any time people start to understand them too well, they have to get replaced with something a little bit less comprehensible.

Why Are Transgender People Immune To Optical Illusions?

[Epistemic status: So, so speculative. Don't take any of this seriously until it's replicated and endorsed by other people.]

I.

If you've ever wanted to see a glitch in the Matrix, watch this spinning mask:



Source: http://hearingthevoice.org/2013/11/14/predictive-coding-masterclass/

Did you see it? As the face started to turn away from you, your brain did…*something*, and then you were seeing a normal frontwards-facing mask again. It turns out your visual system has really strong views about whether faces should be inside-out or not, and it's willing to execute a hard override on perception if it doesn't like what it sees.

But not always. Some people get glitchier glitches than others; a few seem almost immune. Studies find schizophrenics and autistic people to be consistently less glitchy than the rest of us. The correlation's not perfect. But it's definitely there. Something about these people's different cognitive processing styles lets them see through the illusion.

I wanted to replicate this result myself. So a few months ago, when I surveyed readers of my blog, I included some questions about perceptual illusions (including a static version of the hollow mask). I got five thousand responses, including a few from schizophrenic and autistic readers. Sure enough, the effect was there.

Schizophrenic readers were about twice as likely to report a weak reaction to the mask illusion as non-schizophrenics (28% vs. 14%, p = 0.04). They were also more likely to have a weak reaction to a similar illusion, the Spinning Dancer (58% vs. 81%, p = 0.01). Readers with a family history of schizophrenia landed in between schizophrenics and healthy controls (16% for mask, 63% for dancer, ns).

Autistic readers were only slightly more likely to report a weak reaction to the mask illusion than neurotypicals (17% vs. 14%), but thanks to our big sample size we could be pretty confident that this was a meaningful difference (p = 0.004). There was no different between autists and neurotypicals on the Spinning Dancer, not even a weak trend (58% vs. 60%, p = 0.4).

Looking deeper, I found a few other anomalies on illusion perception. Most were small and inconsistent. But one stood out: transgender people had an altered response pattern on both illusions, stronger than the alteration for autism and almost as strong as the one for schizophrenia (mask: cis 14% vs. trans 21%, p = 0.003; dancer: cis 58% vs. trans 71%, p = 0.001). These results are very tentative, and need replication. My mass survey isn't a very sensitive instrument, and I place low confidence in any of this until other people can confirm.

But for now, it sure looks like a signal. Something seems off about transgender people’s perception, something deep enough to alter the lowest-level components of visual processing. If it’s real, what could it be?

II.

A few days ago, trans blogger Zinnia Jones asked me if there might be any neurochemical reason trans people dissociate so much.

Dissociation is a vague psychiatric symptom where you feel like you’re not real, or the world isn’t real, or you’re detached from the world, or something like that. It sounds weird, but if you explain it to someone who’s had it, they’ll say “Oh yeah, *that* thing!” It’s usually unpleasant, and tends to occur in PTSD, borderline personality, and extreme stress.

And in transgender people. The only formal study I can find on this describes it as “greatly prevalent”, and suggests that up to 30% of trans people may have dissociative conditions (compared to less than 1% of the general population). This matches trans people’s self-reports (1, 2, 3, 4, 5). Anecdotally (according to Zinnia’s impression of the trans community) and formally (see Costa & Colizzi 2016) hormone replacement therapy is an effective treatment for dissociative problems.

Intuitively this makes sense. Trans people feel like they’re “trapped in the wrong body”, so of course they feel detached from their bodies / like their bodies aren’t real / like their bodies aren’t theirs. Hormone therapy helps solve the “wrong body” problem, so it also solves the dissociative symptoms.

We aim to bridge psychosocial and biological levels of explanation. We can say that someone is stressed out because their boss overworks them, but *also* because they’re secreting high levels of cortisol. We can say that someone is depressed because they broke up with their boyfriend, but *also* because they have decreased synaptogenesis in their hippocampus. Causation gets tricky, and this is a philosophical minefield for sure, but overall these two levels should be complementary rather than competitive. So what’s the biological correlate to trans people having dissociation problems?

Practically all searches for the biological basis of dissociation end up at the NMDA glutamate receptor, one of the many neurotransmitter systems in the brain. Even though its cousins dopamine and serotonin usually get top billing, glutamate is probably the brain’s most important neurotransmitter, and NMDA glutamate receptors in particular are involved in all sorts of interesting things.

Drugs that block NMDA receptors cause dissociation. The most famous dissociative anaesthetic, ketamine, is an NMDA antagonist. So is DXM, a recreational drug that causes dissociation in abusers. Wikipedia’s list of dissociative drugs is basically just fifty-five NMDA antagonists in a row. The only other category they list are kappa opioid agonists, and kappa opioid agonism probably – you guessed it – antagonize NMDA. If we take this result seriously, every substance we know of that causes dissociation is an NMDA antagonist in some way.

Does anything improve NMDA function – an effect we might expect to alleviate dissociation? Yes, and among a list of intimidating research chemicals called things like “aminocyclopropanecarboxylic acid” is one familiar name: estrogen. See eg El-Bakri et al, which finds that “estrogen modulates NMDA receptors function in the brain…enhancing NMDA function”. McEwen et al: “One of the long-term effects of estradiol [estrogen] is to induce NMDA receptor binding sites in the CA1 region of the hippocampus." Bi et al: "17-B-estradiol [estrogen] enhances NMDA receptor phosphorylation and function." I don't fully understand this research, but it seems to point to estrogen promoting NMDA activity in *some way*.

So transgender people dissociate a lot, a state usually associated with hypofunctioning NMDA receptors. And trans women get better when they take estrogen, a hormone that improves NMDA function. That's interesting. But what does this have to do with those optical illusions?

III.

The Hollow Mask illusion and its cousins may depend on NMDA function.

To oversimplify: the brain interprets the world through Bayesian calculations. In Corlett et al's model, it communicates top-down priors (ie assumptions based on previous knowledge about the world) through NMDA receptors and bottom-up new evidence through AMPA receptors. They write:

> In a hierarchical cortical system in which representations become more abstract with increasing distance from the primary input, higher levels of the hierarchy specify top-down predictions through NMDA receptor signaling and any mismatches between expectancy and experience are conveyed upward through the hierarchy via rapid AMPA and GABA signaling

When you see a hollow mask, the brute facts of how the mask looks are your bottom-up sensory evidence. Your top-down prior is that every other face you've seen for your entire life has been normal, not inside-out. Given the strength of the prior, the prior wins, and your brain interprets the mask as a normal face.

Unless your brain is bad at applying priors, ie its NMDA receptors aren't working that well. Then it just sticks with the bottom-up sensory evidence showing that the mask is hollow.

Schizophrenia and autism both probably involve decreased NMDA function in different ways. For schizophrenia, see eg Olney, NMDA receptor hypofunction model of schizophrenia, and Coyle, NMDA receptor and schizophrenia: a brief history. Ketamine seems to replicate the symptoms of schizophrenia pretty well and is commonly used as a model for the disorder. For autism, see eg Lee, NMDA receptor dysfunction in autism spectrum disorders and this study where screwing with NMDA receptors in mice seems to turn them autistic.

From this we would predict that estrogen would help treat schizophrenia and autism. It does. Schizophrenia is more common and more severe in men than women, with researchers noting that "gonadal steroids may play a role in buffering females against the development of schizophrenia". Women are known to sometimes get schizophrenia triggered by menopause when their estrogen levels decrease. Estrogen supplementation is an effective schizophrenia treatment, and there's some interest in developing estrogen receptor modulators that can help schizophrenic men without making them grow breasts. Meanwhile, autism continues to be about four times more common in men than women, autistic women tend to have more "male-typical brains", and although it's considered unethical to treat autistic boys with estrogen, it works in mice and fish. Once again, doctors are looking into estrogen analogues that don’t turn people female as possible autism treatments.

We might also predict that estrogen would increase glitching on the hollow mask. I can’t study this directly, but on the survey, 15% of biological males had weak reactions to the illusion, compared with only 11% of biological females, $p = 0.01$. Since women have more estrogen, that looks good for the theory.

Transgender people have higher rates of autism and schizophrenia. The Atlantic actually had a good article about this recently : The Link Between Autism And Trans Identity. They cite one study showing 8% autism rate in trans people (compared to 1-2% in the general population), and another showing that autistic people were 7.5x more likely to express “gender variance”. Apparently a lot of trans people have problems getting hormone therapy because their doctors think the gender issues are “just” because of their autism. Some might say that denying people estrogen because they have a condition which studies suggest estrogen can successfully treat is a bit, I don’t know, crazy and evil, but I guess people get really weird around this stuff.

My survey broadly confirms these numbers. Autism rates were sky-high in every category – it’s almost as if the sorts of people who like reading blogs about how gender is all just NMDA receptors skew more autistic than average – but there was a remarkable difference across gender identities. 15% of cisgender people were autistic, but a full 52% of trans people were.

The survey also finds that about 4% of non-schizophrenic people were transgender, compared to 21% of schizophrenics and self-suspected schizophrenics. Other people have noticed the same connection, and I've met more schizophrenic transgender people than I would expect by chance given the very low rates of both conditions.

If this is right, we end up with this rich set of connections between schizophrenics, autistics, ketamine, dissociative experiences, estrogen, gender identity, and the hollow mask. Anything that decreases NMDA function – schizophrenia, autism, ketamine – will potentially cause dissociative experiences and decreased glitching on the mask illusion. Estrogen will improve NMDA function, treat dissociative experiences, and bring back hollow-mask glitching.

So I wonder: is NMDA hypofunction related to transgender? That would explain the autism and schizophrenia connections. It would explain the hollow mask numbers. It would explain the dissociation. It would explain why estrogen helps the dissociation. And it would explain a lot of internal connections between all of these different conditions and factors.

IV.

I'm going to stop here, even though there's a lot more worth saying on this, because I've already gotten so far into Speculation Land that trying to chain any more conclusions on would probably be premature. So let's switch to some reasons for skepticism.

First, the research into NMDA receptors is *too interesting*. People argue that NMDA is key to depression, key to anxiety, OCD, chronic pain, and borderline personality (my guess is the depression claims are mostly overblown, the borderline claims are 100% absolutely right and revelatory, and I'm agnostic on the others). On the one hand, explaining everything sounds sort of good. On the other hand, it also sounds like what would happen if a field was getting kind of overhyped and slipping into methodology loose enough to prove anything it wanted. Maybe a vague link between a receptor which is literally everywhere in the brain and some psychiatric disease isn't that interesting. A theory that can explain absolutely everything should always cause suspicion.

Second, I'm still not sure what to make of the Hollow Mask results on my survey. Although the transgender results were unusually strong, I did get mildly statistically significant results on about half of the thirty-or-so things I looked at, including seemingly-unrelated items like political affiliation. Some might argue that this means something is wrong with my survey. Others might argue that hey, we know political attitudes are about 50% genetic, and the last time people tried trace the genes involved all the strongest results were genes for NMDA receptors. Have I mentioned that NMDA receptors are *really interesting?*

Third, I included a second illusion I asked about on the survey, the Spinning Dancer. It also had an odd response pattern among transgender people. But it didn't correlate at all with the Hollow Mask Illusion, and it doesn't seem to be elevated among autists. I don't know what's going on here, and the whole thing makes me more suspicious that all of this is some weird artifact.

Fourth, all this predicts that ketamine will cause reduced glitching on the Hollow Mask. It doesn't. Corlett argues that this is because chronic, but not acute, NMDA dysfunction is required to stop the hollow mask glitches "because [keatmine] has a predominant impact on bottom-up AMPA signaling". I don't really understand this and it seems like a prediction failure to me. On the other hand, chronic marijuana use *does* prevent mask glitching, which might be because of marijuana causing NMDA hypofunction over time, which I guess is a point in favor of the chronicity theory.

Fifth, although trans women dissociate less when they take estrogen, trans men dissociate less when they take testosterone. I can't find whether testosterone has similar NMDA-promoting properties in the brain, although it sometimes gets aromatized to estrogen so that might be relevant. Also, I've never heard of any trans woman taking testosterone or trans man taking estrogen. If that makes dissociation worse – and from the psychosocial perspective it probably should – then that would be a strike against this theory.

Sixth, although I played up the transgender/autism and transgender/schizophrenia links, the truth is that transgender people have higher rates of every mental illness, to the point where it may just be some general factor. I think I'm justified in focusing on these two results because transgender people's higher rates of depression and anxiety are probably just related to being transgender being depressing and anxiety-provoking in this society. But schizophrenia and autism are 80+% genetic, and so harder to explain away like that. Still, somebody could question the relevance of worrying about these two conditions in particular.

I hope some of this can be sorted out in the near future. A first step would be for someone official to replicate the transgender Hollow Mask pattern and prove that it's not just confounded by autism and schizophrenia rates in that population. A very tentative second step would be to investigate whether chronic use of the supplements that improve NMDA function in schizophrenia – like glycine, d-serine, and especially sarcosine – can augment estrogen in improving gender dysphoria. Remember to consult your doctor before trying any weird supplements since they may cause unintended side effects, like becoming a Republican.

It could also be worth trying to understand more explicitly *why* gender identity and NMDA should be linked. This post is long enough already, but I might write more on this in the future. If you want a preview, check out The Role Of Neonatal NMDA Receptor Activation In Defeminization And Masculinization Of Sex Behavior In The Rat and draw the obvious conclusions.

SSC Meetup: Chicago 7/2

I’ll be in Chicago this weekend along with Katja Grace of AI Impacts.

If you want to hang out, let’s meet at the Starbucks on 200 S Michigan Ave (right near the Art Institute) at 5:00 PM Sunday 7/2.

To The Great City!

Related: 0/4, 1/4, 2/4, 3/4

Last Friday I said goodbye to my patients, handed over my pager, and left Our Lady Of An Undisclosed Location for the last time.

I was hoping for a moment for quiet reflection, but I made it about halfway to the door before being accosted by an incoming intern who couldn’t figure out how to use the medical record system. I helped her, because I’d definitely been there before. I was looking back over my blog archives from four years ago, when I started all of this, and I’d written:

> If you’re entering medical residency and want to prepare, practice this phrase: “I don’t know the answer to that question, but I’ll page my senior resident and he will get back to you.” I keep getting paged by nurses with questions like “Your patient in room 315 has critically low potassium, please advise.” And I went through medical school and I know stuff about low potassium and I have clever ideas for what to do and probably in over 80% of cases those ideas would not kill the patient. But “over 80%” isn’t enough, so each time I have to find my senior, who’s really busy and hates getting paged just like everyone else, and ask him a stupid question I already know the answer to.

And even though it seemed kind of unlikely at the time, that thing which happens to the people in all of those medical shows like *Scrubs* happened to me. I survived intern year. I developed some vague foggy idea what I was doing. At some point I crossed the border into competence. I started becoming one of the people who answers the clueless questions instead of one of the people who asks them.

I can’t say enough nice things about my hospital. They put up with me through all those years I didn’t really know what I was doing, invested thousands of hours into training me. They answered all my rookie questions, like “why can’t patients wear shoes with laces?” (answer: because you can tie shoelaces into a noose to hang yourself with) or “why do we have to learn this style of therapy when studies show it’s no more effective than empathetic listening?” (answer: because the government says so). Everyone knows the stereotypes about residency – you’re overworked, you’re abused, you’re treated like pond scum – and my hospital put the lie to every single one.

And I can’t say enough nice things about the other residents in my program. I’ve never been good at making friends, but being thrown into the lion pit together is a pretty powerful form of bonding, and after the existential terror of intern year where some twenty-something-year-olds who have been to medical school but are otherwise pretty normal people get shoved into a hopital and expected to treat patients together, I think we became pretty close. I never had to trust them with my life, but I trusted my patients’ lives to them every day. I’ve strategized first dates with them, been to their weddings, cooed over their newborn babies. Um. Had several-year-long arguments about consciousness with them. Met the gurus of their obscure religions. Good times.

But Kurt Vonnegut writes about the difference between two kinds of teams. A granfalloon is a team of people pushed together for some ordinary human purpose, like learning medicine or running a hospital psychiatry department. They may get to know each other well. They may like each other.

But in the end, the purpose will be achieved, and they'll go their separate ways.

A karass is a group of people brought together by God for some purpose of His own. No matter how little time they spend together, or how poorly organized they might be, they'll always be on the same wavelength and have a special kinship with each other.

And all the wonderful and compassionate people I got to know during my four years in Michigan are my granfalloon. I've never stopped feeling like you guys – the rationalists, the effective altruists, the transhumanists, the AI scientists, the statisticians, and all the rest – are my karass. All these years I've had to spend away from you have felt a little bit like exile.

Back when I first left the Bay Area to take this position four years ago, I wrote myself a letter. I said that I know about value drift as well as anybody. People always say that they're going to love their first girlfriend forever, and then it never happens. Or they say they're going to keep in touch with their college friends, and then they forget. But – 2013-me wrote to 2017-me – if your values drift so far that you find yourself wanting to stay in Michigan, just…take a step back and think about it. Remember all the friends you have out west, and how exciting the intellectual climate is, and your feeling that it's some kind of fulcrum for the forces that will produce the future, the sort of place that Athens must have felt like in 400 BC, or Florence in 1400. And before you end up marrying some Michigan girl and opening up a practice across the street from the hospital where you trained, just remember everything out here and how much you used to love it.

And even though I had a better time in Michigan than I could possibly have hoped – even though Detroit pizza is literally my favorite food ever and I’m kind of panicking at the idea of being somewhere you can’t get it delivered – despite all of that *I didn’t even have to open the letter*. I remember everything in it perfectly clearly. My values haven’t drifted at all. All of you are great.

So I’m happy to be able to announce that I’ve just accepted a psychiatrist position in the San Francisco metropolitan area – the greatest metropolitan area in the world! I’ve cancelled my lease, loaded everything I own into the trusty little Subaru I got before I realized that driving a Subaru in Detroit would earn you constant dirty looks, and am very gradually moving west.

Thanks to all of you for keeping in touch and helping me stay sane these past four years. I’ll be seeing you very soon.

Two Kinds Of Caution

Financial Times: What We Get Wrong About Technology. It cites boring advances like barbed wire and shipping containers to argue that some of the most transformative inventions are not the product of complicated high technology but just some clever hacks that manage to revolutionize everyday living. Throughout, it uses AI as a foil, starting with Rachel the android from *Blade Runner* and going on to people concerned about superintelligent AI:

> Economists Erik Brynjolfsson and Andrew McAfee write of “the second machine age”, while the World Economic Forum’s Klaus Schwab favours the term “fourth industrial revolution”, following the upheavals of steam, electricity and computers. This coming revolution will be built on advances in artificial intelligence, robotics, virtual reality, nanotech, biotech, neurotech and a variety of other fields currently exciting venture capitalists.
>
> Forecasting the future of technology has always been an entertaining but fruitless game. Nothing looks more dated than yesterday’s edition of Tomorrow’s World. But history can teach us something useful: not to fixate on the idea of the next big thing, the isolated technological miracle that utterly transforms some part of economic life with barely a ripple elsewhere.

And:

> If the fourth industrial revolution delivers on its promise, what lies ahead? Super-intelligent AI, perhaps? Killer robots? Telepathy: Elon Musk’s company, Neuralink, is

> on the case. Nanobots that live in our blood, zapping tumours? Perhaps, finally, Rachael?
>
> The toilet-paper principle suggests that we should be paying as much attention to the cheapest technologies as to the most sophisticated. One candidate: cheap sensors and cheap internet connections. There are multiple sensors in every smartphone, but increasingly they’re everywhere, from jet engines to the soil of Californian almond farms — spotting patterns, fixing problems and eking out efficiency gains. They are also a potential privacy and security nightmare, as we’re dimly starting to realise.

And:

> Like paper, [mildly interesting warehouse management program] Jennifer is inexpensive and easy to overlook. And like the electric dynamo, the technologies in Jennifer are having an impact because they enable managers to reshape the workplace. Science fiction has taught us to fear superhuman robots such as Rachael; perhaps we should be more afraid of Jennifer.

I agree with the gist of this article. It’s correct to say that we often overlook less glorious technologies. It’s entirely right in pointing out things like barbed wire as good examples of these.

Also, it was written on a digital brain made of rare-earth metals consisting of billions of tiny circuits crammed into a couple of cubic inches, connected to millions of other such brains by underwater fiber optic cables that connect entire continents with one another at an appreciable fraction of the speed of light.

What I’m saying is, sometimes the exciting cool technologies are pretty great too.

I realize this isn’t a brilliant or controversial insight. Exciting-looking technologies that everybody agrees will be exciting turn out to be exciting, breaking news, more at eleven.

But then what am I to make of the original article? It points out some cases where simple boring technologies proved to be pretty important. In one or two cases, it describes a field where a simple boring technology proved to be more important than a flashier and superficially-much-more-promising technology. Then it concludes that “perhaps” we should be more afraid of simple voice recognition programs than of superintelligent AI.

I can come up with equally compelling anecdotes proving the opposite. For example, the humble stirrup was one of the most disruptive and important innovations in world history – read about the Great Stirrup Controversy sometime. Imagine a society of horses in 1890, where some especially wise horse relates the story, and concludes with “So perhaps we should be more concerned about simple innovations like new stirrups and more efficient reins, than of the motorcar.” Nice try, A+ for effort, you’re still going to end up as glue.

I don’t want to claim that flashy paradigm-shifting technologies are always more disruptive than simple boring technologies, or that technologies always deploy quickly. I do want to claim that the article hasn’t even tried to prove the opposite. So when it says “perhaps we should be more worried about warehouse management programs than superintelligent AIs”, it means “perhaps” in the weaselly sense, like “perhaps we should be more worried about a massive worldwide snake infestation than global warming. I have no evidence for this, but perhaps it is true."

Part of me wants to let this pass. It's obviously a throwaway line, not really meant to be a strong argument. But another part of me thinks that's exactly the problem. There are so many good throwaway lines you could use to end a piece. If you have to halfheartedly make a not-strong argument for something, why would you choose the one where you randomly dismiss an impending threat that already has way too few people willing to pay any attention to it?

I worry there's a general undersupply of meta-contrarianism. You have an obvious point (exciting technologies are exciting). You have a counternarrative that offers a subtle but useful correction (there are also some occasional exceptions where the supposedly-unexciting technologies can be more exciting than the supposedly-exciting ones). Sophisticated people jump onto the counternarrative to show their sophistication and prove that they understand the subtle points it makes. Then everyone gets so obsessed with the counternarrative that anyone who makes the obvious point gets shouted down ("What? Exciting technologies are exciting? Do you even read *Financial Times*? It's the *unexciting* technologies that are truly exciting!"). And only rarely does anyone take a step back and remind everyone that the obviously-true thing is still true and the exceptions are still just exceptions.

And for some reason, any discussion of AI risk dials this up to eleven. It seems pretty obvious that smarter-than-human AI could be dangerous for humans. For a hundred years, every scientist and science fiction writer who's considered the problem has concluded that smarter-than-human AI could be dangerous for humans. And so we get these constant hot takes, "Oh, you're afraid of superintelligent AI?

What if the real superintelligent AI was *capitalism*?" Or "What if the real superintelligent AI was *the superintelligent AI in the heart of all humanity*?" Or just "What if superintelligent AI turns out to be less important than a bunch of small humble technologies that don't look like anything much?" And so I feel like I have to do the boring work of saying "hey, by the way, 10-20% of AI researchers believe their field will end in an 'existential catastrophe' for the human race, and this number is growing every year, Steven Hawking is a pretty smart guy and he says we could all die, and Nick Bostrom is an Oxford professor and he says we could all die, and Elon Musk is Elon Musk and he says we could all die, and this isn't actually a metaphor for anything, we are actually seriously worried that we could all die here".

But I worry even more that this *isn't* an attempt to sound sophisticated. I worry that it's trying to sound *cautious*. Like, "ah, yes, some firebrands and agitators say that we could all die here, but I think more sober souls can get together and say that probably things will continue much as they always have, or else be different in unpredictable ways because history is always inherently unpredictable", or something like that.

I worry that people don't adequately separate two kinds of caution. Call them local caution and global caution. Suppose some new spacecraft is about to be launched. A hundred experts have evaluated it and determined that it's safe. But some low-ranking engineer at NASA who happens to have some personal familiarity with the components involved looks at the schematics and just has a really bad feeling. It's not that there's any specific glaring flaw. It's not any of the known problems that have ever led to spacecraft failure before. Just that a lot of the parts weren't *quite* designed to go together in *exactly* that way, and that without being entirely able to explain his reasoning, he would not be the least bit surprised if that spacecraft exploded.

What is the cautious thing to do? The *locally* cautious response is for the engineer to accept that a hundred experts probably know better than he does. To cautiously remind himself that it's unlikely he would discover a new spacecraft failure mode unlike any before. To cautiously admit that grounding a spacecraft on an intuition would be crazy. But the *globally* cautious response is to run screaming into the NASA director's office, demanding that he stop the launch immediately until there can be a full review of everything. There's a sense in which this is rash and ignores all sorts of generally wise and time-tested heuristics like the ones above. But if by "caution" you mean you want as few astronauts as possible to end up as smithereens, it's the way to go.

And part of me gets really happy when people say that we should avoid jumping to conclusions about AI being dangerous, because the future often confounds our expectations, and shocking discontinuous changes are less likely than gradual changes based on a bunch of little things, or any of a dozen other wise and entirely correct maxims. These are the principles of rationality that people *should* consider when making predictions, the epistemic caution that forms a rare and valuable virtue.

But this is the wrong kind of caution for this situation. It's assuming that there's some sort of mad rush to worrying about AI, and people need to remember that it might not be so bad. That's the opposite of reality. As a society, we spend about $9 million yearly looking into AI safety, including the blue-sky and strategy research intended to figure out whether there's other research we should be doing. This is good, but it's about one percent of the amount that we spend on simulated online farming games. This isn’t epistemic caution. It’s insanity. It’s like a general who refuses to post sentries, because we can’t be certain of anything in this world, so therefore we can’t be certain the enemy will launch an attack tonight. The general isn’t being skeptical and hard-headed. He’s just being insane.

And I worry this is the kind of mindset that leads to throwaway phrases like “perhaps we should be more worried about this new warehouse management program than about superintelligent AI”. Sure, perhaps this is true. But perhaps it isn’t. “Perhaps” is a commutative term. So, “Perhaps we should be more worried about superintelligent AI than about a new warehouse management program”. But the warehouse management company makes more money each year than the entire AI safety field budget combined.

Perhaps we should spend more time worrying about this, and less time thinking of clever reasons why our inaction might turn out to be okay after all.

Change Minds Or Drive Turnout?

I.

Current Affairs: The Democratic Party Just Admitted It Doesn't Stand For Anything. Overall it makes some good points, but one passage caught my eye:

> [Democrats believe that if you moderate your platform and swing toward the center] you might lose a few hardcore lefties, but you'll more than make up for it in the number of Reaganites you peel away from the other side. (Or, as Chuck Schumer put it, "For every blue-collar Democrat we lose in western Pennsylvania, we will pick up two moderate Republicans in the suburbs in Philadelphia.")
>
> But this philosophy is a dead end. For one thing, it doesn't work. Unless you have Bill Clinton's special charismatic magic, what actually happens is that progressive voters just stay home, disgusted at the failure of both parties to actually try to improve the country. And the mythical "moderate Republicans" never seem to show up. (This is because there are no actual moderate Republicans.)

This has been a staple of recent leftist thought. Another example from Daily Kos (via the paper below):

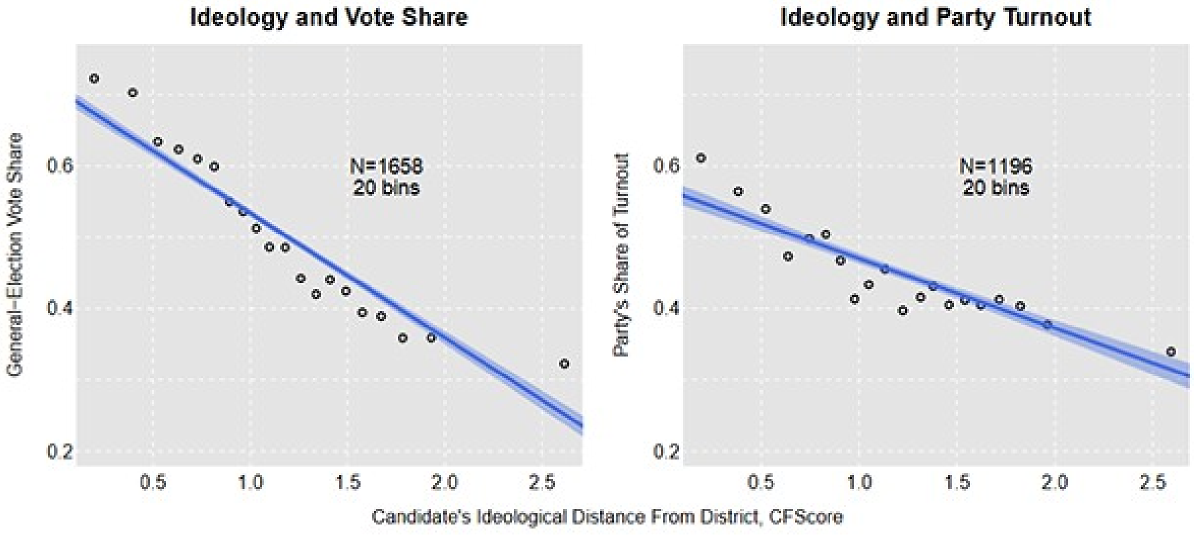
> The key data is this, and it's important to re-emphasize if only to shut up the useless, overpaid political consultants who idiotically babble about "moving to the center" or "compromising with the other side"…What matters is turning out our voters. That's it. The Democrats win when we fire up and turn out our base.

This sounds like a win-win situation. We can stick to our principles, and that actually makes us *more* electable. Big if true. But is it?

II.

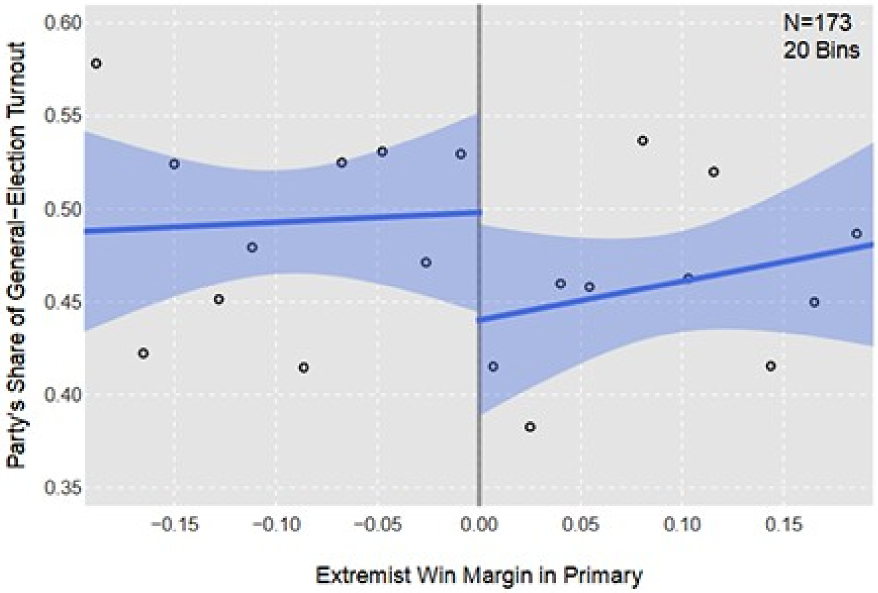
First: do more extreme views increase base turnout? This is the subject of Hall & Thompson (2017). They examine 1658 House races from 2006 to 2012 and start by noticing that the more distant a candidate from the median voter in their district, the fewer votes they get *and* the lower their party's share of turnout in the general election (ie the various other races that go on at the same time). This suggests that not only are the voters who do turn out less likely to vote for the extremist, but that many of their voters are staying home (or many of their opponents' voters have been galvanized to show up).

Figure 2 – **Candidate Ideology, Elections, and Turnout in the U.S. House.**
More extreme candidates do worse, electorally (left panel), and their party's voters comprise a smaller share of turnout in the general election (right panel).



These raw results could be driven by exogenous factors. For example, maybe in swing states, parties nominate more centrist candidates (to get a broader appeal) and have higher turnout (because people’s votes actually matter). To eliminate this possibility, the researchers try a regression discontinuity design – ie they compare districts where extremists won the primary by 0.1% to districts where extremists lost the primary by 0.1%. These sorts of tiny margins are likely to be pretty random, so it’s almost like an experimental trial of what happens when you randomly vary candidate extremism.

Figure 3 – The Effect of Extremist Nominees on Validated Partisan Turnout in the General Election. U.S. House, 2006–2012. When a party nominates an extremist in its primary, general-election turnout skews towards the opposing party.



This better-controlled data set finds the same thing. The more extreme a candidate, the lower their party’s share of the turnout.

This actually makes a lot of sense – a lot of my normally non-voting friends turned out last November because they hated Trump so much, and a lot of #NeverTrump Republicans, unwilling to hold their nose and vote Hillary, just stayed home.

Hall and Thomspon conclude:

> This paper engages with a longstanding debate over the relative strengths of extreme legislative candidates, thought to boost turnout among their party's base, and moderate candidates thought to attract hypothetical moderate swing voters. Using several different empirical strategies, we have found consistent evidence that extremist nominees do poorly in general elections in large part because they skew turnout in the general election away from their own party and in favor of the opposing party.

They crunch a few more numbers and conclude that effects on turnout might be the *entire* reason why extremist candidates do worse. That is, there is no remaining effect from swing voters who switch from their own party to the other party. Turnout is the only thing that matters:

> The results suggest that much of moderate candidates' success may actually be due to the turnout of partisan voters, rather than to swing voters who switch sides. In fact, our regression discontinuity estimates are consistent with the possibility that the entire vote-share penalty to extremist nominees is the result of changes in partisan turnout. Seen in this light, the results are more consistent with the behavioral literature's focus on turnout than they are with the institutional literature's theoretical focus on swing voters. As such, we see this paper as helping to link the behavioral and institutional literatures together, suggesting that moderate candidates do possess an electoral advantage, but that this advantage may depend heavily on turnout-based mechanisms.

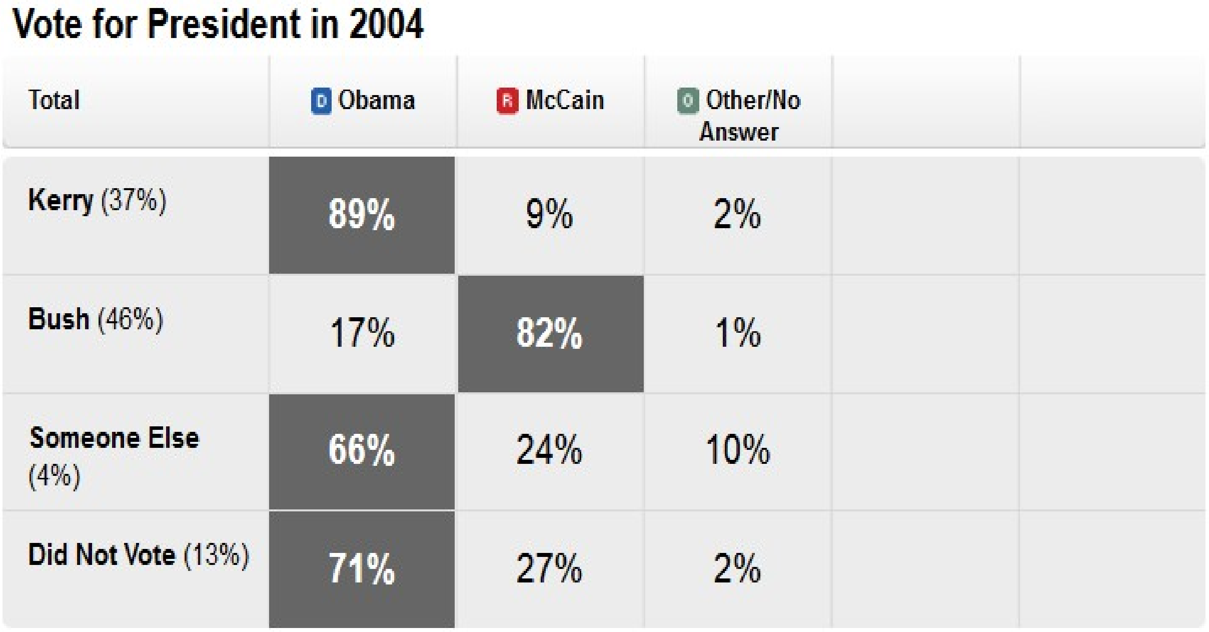
So Thompson and Hall disagree with the theory that a less compromising, more robustly leftist Democratic Party would get more votes. But they tentatively agree with *Current Affairs'* claim that “moderate Republicans” are a myth and nobody ever switches sides.

III.

Second: Is base turnout really the only thing that matters?

I’m reluctant to disagree with real political scientists like Hall and Thompson, but I’m a little more optimistic about whether people can change their minds.

There’s little data on vote-switching, and the only directly relevant information I could find was this CNN exit poll from 2008:



Vote for President in 2004

Total	D Obama	R McCain	O Other/No Answer		
Kerry (37%)	89%	9%	2%		
Bush (46%)	17%	82%	1%		
Someone Else (4%)	66%	24%	10%		
Did Not Vote (13%)	71%	27%	2%		

Of people who voted Democrat for President in 2004, 9% went Republican in 2008. Of people who voted Republican in 2004, a full 17% went Democrat in 2008. Some analysts of this information caution us that people are bad at remembering their votes so some of this may be wrong. But I feel like this story *also* doesn’t fit well with with unchanging-eternal-partisanship narrative – if you’ve voted straight Republican for the last ten elections and loathe all Democrats with a burning fury, you’re not going to just forgot whether you voted Bush or Kerry in ’04.

Rasmussen doesn’t have a real exit poll, but they put a couple of different sources together to guess about how many people switched votes in most recent election. I don’t really understand their graphs – in particular, their use of the Other category doesn’t make much sense. But if I’m reading them right, of people who voted Democrat in 2012, about 13% voted Trump in 2016. And of people who voted Republican in 2012, about 4% voted Clinton in in 2016. These may seem like small numbers. But in the context of the tiny margins by which Trump won swing states (Michigan by 0.3%, Pennsylvania by 0.7%, Wisconsin by 0.8%), these sorts of changes are absolutely decisive.

So swing voters and moderates aren’t totally mythical. But how do they compare with turnout as a determining factor in elections?

This is hard to figure out. ~~We know that total turnout decreased 2% between 2012 and 2016~~ [EDIT: More recent sources say turnout increased. Not clear on this right now. See here]. But it’s hard to interpret party turnout figures. If the number of Democratic votes dipped more than the number of Republican votes, how much of that is because the Democrats had a bigger turnout problem, and how much is because some Democrats crossed the aisle to vote Republican?

Nate Cohn of the New York Times tries to solve this by analyzing turnout of *predicted* partisan voters – eg a young black gay college graduate will probably vote Democrat, so if he doesn’t show up it suggests Democratic base turnout declined. Before the election, he made some mechanical projections about how much each demographic would turn out based on how often they've turned out before in situations like this. Sometimes this risks adjusting away exactly the factors we're interested in – eg he predicts black people will have much lower turnout in 2016 because part of their record 2012 turnout was personal loyalty to Obama. But as far as I can tell he doesn't adjust for anything about the candidate's ideologies, making his predictions okay for our purposes of talking about the effects of candidate extremism.

Cohn finds that blacks voted a little bit less than he predicted, and Hispanics a little bit more. Whites likely to support Trump (eg older, less educated, etc) turned out about 7% more than expected. Whites likely to support Clinton turned out about 4% more (sic!) than expected. But overall, these differences were "only a modest effect", and probably not enough to affect the election:

> Turnout improved Mr. Trump's standing by a modest margin compared with pre-election expectations. If the turnout had gone exactly as we thought it would, the election would have been extremely close. But by this measure, Mrs. Clinton still would have lost both Florida and Pennsylvania – albeit very narrowly…Democrats are right to blame many of their midterm election losses on weak turnout. They're on far shakier ground if they complain about the turnout last November.

He thinks that it was the much-maligned swing voters who were more important:

> If turnout played only a modest role in Mr. Trump's victory, then the big driver of his gains was persuasion: He flipped millions of white working-class Obama supporters to his side.

The voter file data makes it impossible to avoid this conclusion. It’s not just that the electorate looks far too Democratic. In many cases, turnout cannot explain Mrs. Clinton’s losses.

Take Schuylkill County, Pa., the county where Mr. Trump made his biggest gains in Pennsylvania. He won, 69 percent to 26 percent, compared with Mitt Romney’s 56-42 victory. Mrs. Clinton’s vote tally fell by 7,776 compared with Mr. Obama’s 2012 result, even though the overall turnout was up.

Did 8,000 of Mr. Obama’s supporters stay home? No. There were 5,995 registered voters who voted in 2012, remain registered in Schuylkill County, and stayed home in 2016.

And there’s no way these 2016 drop-off voters were all Obama supporters. There were 2,680 registered Democrats, 2,629 registered Republicans and 686 who were unaffiliated or registered with a different party. This is a place where registered Democrats often vote Republican in presidential elections, so Mr. Obama’s standing among these voters was most likely even lower […]

Survey data, along with countless journalistic accounts, also suggest that voters switched in huge numbers.

Throughout the campaign, polls of registered voters — which are not subject to changes in turnout — showed Mrs. Clinton faring much worse than Mr. Obama among white working-class voters.

The postelection survey data tells a similar story: Mrs. Clinton won Mr. Obama’s white-working class supporters by a margin of only 78 percent to 18 percent against Mr.

> Trump, according to the Cooperative Congressional Election Study.
>
> In the Midwestern battleground states and Pennsylvania, Mrs. Clinton had an advantage of 76 percent to 20 percent among white working-class Obama voters.
>
> The survey data isn’t perfect. It relies on voters’ accurate recall of their 2012 vote, and that type of recall is often biased toward the winner. Indeed, the C.C.E.S. found that Mr. Obama had 54 percent of support among 2012 voters, compared with his actual 51 percent finish.
>
> But the data all points in the same direction: Shifts in turnout were not the dominant factor in Mr. Trump’s success among white working-class voters.

I tried to model some of this myself to get actual numbers I could compare. It doesn’t work. If I apply the exit poll models of voter defections to the real numbers, I get implausibly high numbers for Trump and implausibly low numbers for Hillary. I would have to add a huge jump in Democratic turnout, and a corresponding crash in Republican turnout, to produce the modest Hillary popular-vote win we actually saw. Nobody’s claimed this and I don’t think that it happened. So I’m confused. I hate to have to go off of Cohn’s analysis, especially since he never really explains what goes into his projections, but right now it’s all I have. And it matches what Rasmussen thought in a lot of ways.

So I very tentatively conclude that swing voters might have changed the result of the 2016 election. I can’t directly compare to decreased turnout, but it seems at least as important, especially if you discount the non-ideology-related black turnout decrease.

Granted, the 2016 election was weird, we might be in some kind of unique realignment of the two-party system, maybe this doesn't happen too often. But the Obama/Trump defections don't seem much greater than the Bush/Obama defections on the 2008 CNN exit poll. And *Current Affairs* admits that Bill Clinton did pretty well attracting moderates and Republicans to his banner. I think there's enough examples to think that a large effect from swing voters might not just be possible, but common.

As far as I can tell, the evidence leans against the win-by-extremism-turning-out-the-base argument. Extremists tend to do worse in elections. They don't raise turnout of their base; in fact, they probably lower it. They may fire up their opponents' base. And swing voters can make a big difference when a candidate appeals to them.

This doesn't mean only boring centrists can win; Donald Trump is the obvious counterexample. But Trump's extremism wasn't just "Paul Ryan but much more so". He won not by moving straight right, but by coming up with new ideas that held the attention of the Republican base while also appealing to some disaffected Democrats. And the Bernie Sanders wing of the Democratic Party might be able to do something similar from the left if it gets the chance.

Just don't frame it as "extremism turns out the base", and especially not as "swing voters don't matter".

SSC Meetup: Salt Lake 7/14

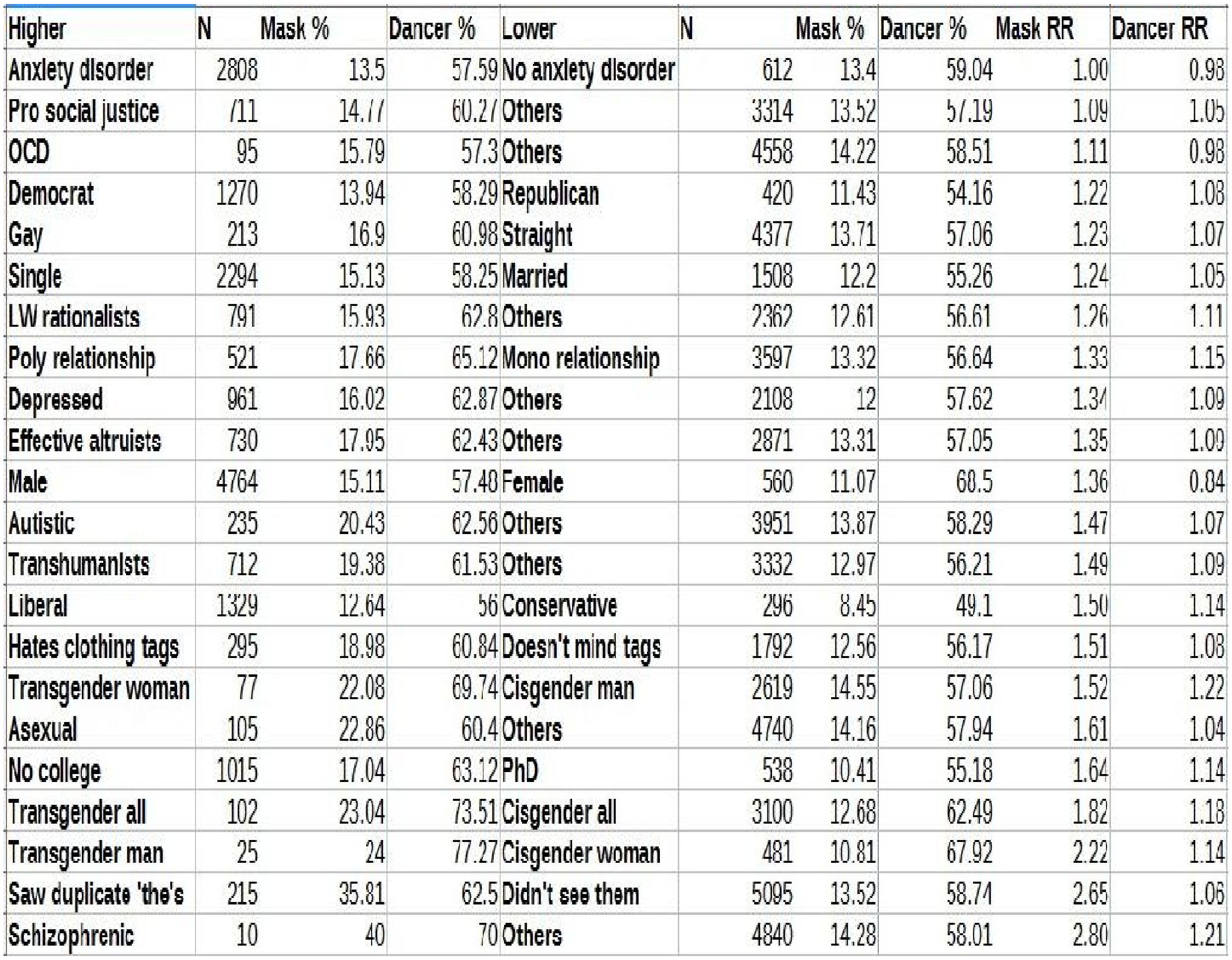
I’ll be in Salt Lake City later this week, along with Katja Grace of AI Impacts.

If you want to hang out, let’s meet at the Starbucks in the Sheraton Hotel, 150 500 S at 6:00 PM Friday 7/14.

Edit: Starbucks is very small, overflow in lobby.

Can We Link Perception And Cognition?

Last month I talked a little bit about the Hollow Mask Illusion as a clue to the Bayesian operations going on “below the hood” in the brain. Today I want to go a little bit deeper into what the SSC survey results can tell us here. This is a list of a bunch of different variables I tested in the survey, and the percent of each group who saw the Mask Illusion and Dancer Illusion as ambiguous. “RR” is relative risk:



Higher	N	Mask %	Dancer %	Lower	N	Mask %	Dancer %	Mask RR	Dancer RR
Anxiety disorder	2808	13.5	57.59	No anxiety disorder	612	13.4	59.04	1.00	0.98
Pro social justice	711	14.77	60.27	Others	3314	13.52	57.19	1.09	1.05
OCD	95	15.79	57.3	Others	4558	14.22	58.51	1.11	0.98
Democrat	1270	13.94	58.29	Republican	420	11.43	54.16	1.22	1.08
Gay	213	16.9	60.98	Straight	4377	13.71	57.06	1.23	1.07
Single	2294	15.13	58.25	Married	1508	12.2	55.26	1.24	1.05
LW rationalists	791	15.93	62.8	Others	2362	12.61	56.61	1.26	1.11
Poly relationship	521	17.66	65.12	Mono relationship	3597	13.32	56.64	1.33	1.15
Depressed	961	16.02	62.87	Others	2108	12	57.62	1.34	1.09
Effective altruists	730	17.95	62.43	Others	2871	13.31	57.05	1.35	1.09
Male	4764	15.11	57.48	Female	560	11.07	68.5	1.36	0.84
Autistic	235	20.43	62.56	Others	3951	13.87	58.29	1.47	1.07
Transhumanists	712	19.38	61.53	Others	3332	12.97	56.21	1.49	1.09
Liberal	1329	12.64	56	Conservative	296	8.45	49.1	1.50	1.14
Hates clothing tags	295	18.98	60.84	Doesn't mind tags	1792	12.56	56.17	1.51	1.08
Transgender woman	77	22.08	69.74	Cisgender man	2619	14.55	57.06	1.52	1.22
Asexual	105	22.86	60.4	Others	4740	14.16	57.94	1.61	1.04
No college	1015	17.04	63.12	PhD	538	10.41	55.18	1.64	1.14
Transgender all	102	23.04	73.51	Cisgender all	3100	12.68	62.49	1.82	1.18
Transgender man	25	24	77.27	Cisgender woman	481	10.81	67.92	2.22	1.14
Saw duplicate 'the's	215	35.81	62.5	Didn't see them	5095	13.52	58.74	2.65	1.06
Schizophrenic	10	40	70	Others	4840	14.28	58.01	2.80	1.21

I don’t have p-values listed here, but almost all the Hollow Mask results, and a few of the Spinning Dancer results, were significant at $p \leq 0.01$. And beyond the individual results, a few things jump out of the data in general. The Hollow Mask results and Spinning Dancer results are always in the same direction. And it always seems to be the weirder group who see more ambiguity in the illusion. Yes, schizophrenics see more ambiguity than non-schizophrenics. But transhumanists also see more ambiguity than non-transhumanists. Polyamorous people see more than monogamous people, gay people see more than straight people, EAs see more than non-EAs, et cetera. Where there's no clear weirder/less-weird dichotomy, it seems like it's the lower-functioning group that has more ambiguity. High school dropouts rather than PhDs. Single people rather than married people.

So there seems to be a picture where high rates of perceptual ambiguity are linked to being weirder and (sometimes, in a very weak statistical way) lower-functioning.

But why stop there? The dream is to connect this to some sort of intuitively-meaningful cognitive variable.

For a sense of "intuitively meaningful cognitive variable", consider something like those four-letter things you get on the Myers-Briggs test. Go ahead and interject that Myers-Briggs is unscientific, and no better than astrology, and inferior to the Five Factor Model in every way. But everyone who says that always ends up being INTJ/INTP. And a survey found that SSC readers are about ten times more likely to be INTJ/INTP than the general population, $p \leq 0.001$. Without necessarily claiming that the underlying classification cleaves reality at the joints, or even that it gives you more information than you put into the personality test that generates it, differences in cognitive styles seems real. I don't know how fundamental they are – it could just be something as silly as a freshman philosophy professor who encouraged you to think logically or something – but they seem real.

And they seems different than the variables on the Big Five. The Big Five measures personality. Myers-Briggs claims – maybe wrongly, but at least it claims – to measure how you reason about things. Maybe everyone's had the experience of meeting someone who seems very smart, but who just reasons in a very different way than they do.

And if the Bayesian brain hypothesis is right, and perception and reason really do draw on the same fundamental processes, then I wonder if we could isolate some differences in reasoning by measuring differences in perception. Could perception of certain optical illusions predict responses to certain cognitive biases? Could that go on to predict things like whether people like analytic or continental philosophy, whether they're early-adopters or traditionalists, whether they think people are basically good or basically evil?

I know this is an overly ambitious research program. But remember: the studies looking for the genetic underpinning of political opinions usually implicate NMDA receptors, the same receptors most likely involved in the Hollow Mask. And there was a small but highly significant correlation between Mask perception and political opinion on the survey. I agree this is crazy, but I don't want to say it's impossible just yet.

On the the next survey, I want to include a whole battery of illusions, including multiple examples of the same illusion asked in different ways, and different illusions that seem to be measuring the same thing. For an example of the latter, take the "saw duplicate thes" item on the table above. This was a question asking if people had noticed various duplications of the word "the" I had put in the survey (like the one in the second and third words of this paragraph). People who noticed the duplicates were more than twice as likely to see ambiguity in the Hollow Mask as others, the highest result other than schizophrenia itself. This confirms my hypothesis that there’s some underlying similarity between these two illusions. If I can get enough of these, then I can eliminate noise and get a better idea of the underlying mental process that might be generating all of these.

With luck, I might end up with a couple of different factors that predict illusion perception. Then I would want to see if those factors also predict performance on reasoning problems (like cognitive biases) and on high-level beliefs (like liberal versus conservative).

The big question is whether some non-neurological factor influences perception of illusions – like maybe just trying really hard to see them. I’m not sure how to adjust for that, except to say that the pattern here doesn’t really look like that. The Dancer illusion was the one most susceptible to increased effort, and it got the weakest results. On a quick check, it doesn’t look like this is all due to something obvious like gender or age. But maybe there’s still some confounding factor that I’ve missed.

Classified Thread 2: Best In Classified

I’m still on vacation, so here’s another classified thread. Post ads, personals, and any interesting success stories from the last thread.

…and I’ll start. SSC is part of a wider movement of philosophy enthusiasts, transhumanists, effective altruists, etc which has somehow ended up with the simultaneously boring and arrogant moniker of “the rationalist community”. We’ve developed a small intellectual/social scene in the SF Bay Area, with a few hundred interesting people who hang out together and cooperate on various projects. Since rent in the Bay is so high, a lot of the rationalists there are living in group houses, which have become nuclei for social events and cooperation.

Four of these have ended out clustered on Ward Street in Berkeley, and we’re thinking we might as well try to accelerate this and turn the area into a center of the community. We’ve been trying to snatch up houses in the area, and we just got dibs on four houses immediately adjacent to the existing rationalist cluster that are currently available for rent:

1. A five bedroom house for ~$5500/month, available now
2. A four bedroom house for ~$4200/month, available now
3. A seven-to-eight bedroom house, cost to be determined, available 9/1/17
4. A three bedroom house for ~$3100/month, available now (not adjacent to existing cluster; a few blocks away)

All of these are owned by the same landlord, who we’ve previously found pretty reasonable. They’re all kind of old and not going to win any Modern Architectural Design awards or even Especially Well Maintained awards, but we think (investigations still ongoing) that they’re basically solid and in good shape. Pictures and viewings available on request.

We’re currently looking for people who might be interested, either in renting entire houses, or in taking single rooms in what will probably become group houses. Existing community members are of course welcome to apply, but so is anyone who’s reading this and who thinks the idea sounds interesting. If interested, contact katja.s.grace[at]gmail[dot]com for more information and to arrange viewings, etc.

(disclaimer: I enjoyed living in the Bay Area, but I can’t deny that the prices are terrible, the local politics absurd, and the density at just the right level to frustrate lovers of big cities and quiet suburbs alike. Experiences with the rationalist community there vary widely, from people who say it was life-changingly good to people who found it disappointing and difficult to get into. The housing situation here might make it easier to get into, but no guarantees)

Targeting Meritocracy

I.

Prospect Magazine writes about the problem with meritocracy. First Things thinks meritocracy is killing America. Feminist Philosophers comes out against meritocracy. The Guardian says “down with meritocracy”. Vox calls for an attack on the false god of meritocracy. There’s even an *Against Meritocracy* book. Given that meritocracy sounds almost tautologically good (doesn’t it just mean positions going to people with merit?), there sure do seem to be a lot of people against it.

Some of these people are just being pointlessly edgy. The third article seem to admit that a true meritocracy would be a good thing, but argues that we don’t have one right now. This hardly seems “against meritocracy”, any more than saying we don’t have full racial equality right now means you’re “against racial equality”, but whatever, I guess you’ve got to get clicks somehow.

The other articles actually mean it. Their argument seems to be gesturing at the idea that elites send their kids to private schools, where they get all A+s and end up as president of the Junior Strivers Club. Then they go to Harvard and dazzle their professors with their sparkling wit and dapper suits. Then they get hired right out of college to high-paying management positions at Chase-Bear-Goldman-Sallie-Manhattan-Stearns-Sachs-Mae-FEDGOV. Then they eat truffle-flavored caviar all day and tell each other “Unlike past generations of elites, we are meritocrats who truly deserve our positions, on account of our merit”, as the poor gnash their teeth outside.

Grant that this is all true, and that it’s bad. Does that mean we should be against meritocracy?

II.

There’s a weird assumption throughout all these articles, that meritocracy is founded on the belief that smart people deserve good jobs as a reward for being smart. Freddie de Boer, in his review of yet another anti-meritocracy book, puts it best:

> I reject meritocracy because I reject the idea of human deserts. I don’t believe that an individual’s material conditions should be determined by what he or she “deserves,” no matter the criteria and regardless of the accuracy of the system contrived to measure it. I believe an equal best should be done for all people at all times.
>
> More practically, I believe that anything resembling an accurate assessment of what someone deserves is impossible, inevitably drowned in a sea of confounding variables, entrenched advantage, genetic and physiological tendencies, parental influence, peer effects, random chance, and the conditions under which a person labors. To reflect on the immateriality of human deserts is not a denial of choice; it is a denial of self-determination. Reality is indifferent to meritocracy’s perceived need to “give people what they deserve.”

I think this is both entirely true and entirely missing the point. The intuition behind meritocracy is this: if your life depends on a difficult surgery, would you prefer the hospital hire a surgeon who aced medical school, or a surgeon who had to complete remedial training to barely scrape by with a C-? If you prefer the former, you’re a meritocrat with respect to surgeons. Generalize a little, and you have the argument for being a meritocrat everywhere else.

The Federal Reserve making good versus bad decisions can be the difference between an economic boom or a recession, and ten million workers getting raises or getting laid off. When you've got that much riding on a decision, you want the best decision-maker possible – that is, you want to choose the head of the Federal Reserve based on merit.

This has nothing to do with fairness, deserts, or anything else. If some rich parents pay for their unborn kid to have experimental gene therapy that makes him a superhumanly-brilliant economist, and it works, and through no credit of his own he becomes a superhumanly-brilliant economist – then I want that kid in charge of the Federal Reserve. And if you care about saving ten million people's jobs, you do too.

III.

Does this mean we just have to suck it up and let the truffle-eating Harvard-graduating elites at Chase-Bear-Goldman-Sallie-Manhattan-Stearns-Sachs-Mae-FEDGOV lord it over the rest of us?

No. The real solution to this problem is the one none of the anti-meritocracy articles dare suggest: accept that education and merit are two different things!

I work with a lot of lower- and working-class patients, and one complaint I hear again and again is that their organization won't promote them without a college degree. Some of them have been specifically told "You do great work, and we think you'd be a great candidate for a management position, but it's our policy that we can't promote someone to a manager unless they've gone to college". Some of these people are too poor to afford to go to college. Others aren't sure they could pass; maybe they have great people skills and great mechanical skills but subpar writing-term-paper skills. Though I've met the occasional one who goes to college and rises to great heights, usually they sit at the highest non-degree-requiring tier of their organization, doomed to perpetually clean up after the mistakes of their incompetent-but-degree-having managers. These people have loads of merit. In a meritocracy, they'd be up at the top, competing for CEO positions. In *our* society, they're stuck.

The problem isn't just getting into college. It's that success in college only weakly correlates with success in the real world. I got into medical school because I got good grades in college; those good grades were in my major, philosophy. Someone else who was a slightly worse philosopher would never have made it to medical school; maybe they would have been a better doctor. Maybe someone who didn't get the best grades in college has the right skills to be a nurse, or a firefighter, or a police officer. If so, we'll never know; all three of those occupations are gradually shifting to acceptance conditional on college performance. Ulysses Grant graduated in the bottom half of his West Point class, but turned out to be the only guy capable of matching General Lee and winning the Civil War after a bunch of superficially better-credentialed generals failed. If there's a modern Grant with poor grades but excellent real-world fighting ability, are we confident our modern educationocracy will find him? Are we confident it will even try?

Remember that IQ correlates with chess talent at a modest r = 0.24, and chess champion Garry Kasparov has only a medium-high IQ of 135. If Kasparov's educational success matched his IQ, he might or might not have made it into Harvard; he certainly wouldn't have been their star student. And if it was only that kind of educational success that gave spots on some kind of national chess team, Kasparov and a bunch of other grandmasters would never have a chance. Real meritocracy is what you get when you ignore the degrees and check who can actually win a chess game.

One of the few places I see this going well is in programming. Triplebyte (conflict of interest notice: SSC sponsor) asks people who want a programming job to take a test of their programming ability, "no resume needed". Then it matches them with tech companies that want the kind of programming the applicant is good at. It doesn't matter whether you were president of the Junior Strivers' Club in college. It doesn't matter whether you managed to make it past the gatekeepers trying to keep you out for not excluding the right kind of upper-class vibe. What matters is whether you can code or not. As a result, a bunch of the people I know are poor/transgender/mentally ill people who couldn't do college for whatever reason, bought some computer science books and studied on their own, and got hired by some big tech company. Programming is almost the only well-paying field where people can still do this, and it doesn't surprise me that the establishment keeps portraying its culture as uniquely evil and demanding it be dismantled.

I think we should be doing the opposite: reworking every field we can on the same model. Instead of Goldman Sachs hiring whoever does best at Harvard, they should hire people who can demonstrate their knowledge of investing principles or (even better) who can demonstrate an ability to predict the market better than chance. Some of these people will be the academic stars who learned how to do it at Harvard Business School. But a lot of others will be ordinary working-class people who self-studied or who happen to have a gift, the investing equivalents of General Grant and Garry Kasparov.

I don’t think the writers of the anti-meritocracy articles above really disagree with this. I think they’re probably using a different definition of meritocracy where it *does* mean “rule by well-educated people with prestigious credentials”. But I think it’s important to defend the word “meritocracy” as meaning what it says – decision by merit, rather than by wealth, class, race, or education – and as a good thing. If we let the word be tarnished as some sort of vague signifier of a corrupt system, then it’s too easy for the people who really are in that corrupt system to exploit the decline and fall of the only word we have to signal an alternative. “Oh, you don’t like that all the important jobs go to upper-class people instead of the people who are best at them? You’d prefer they be given out based on merit? But haven’t you read *The New Inquiry*, *First Things*, and *Vox*? Believing in so-called ‘meritocracy’ is totally uncool!” And then we lose one of the only rallying points, one of the few pieces of vocabulary we have to express what’s wrong with the current system and what would be a preferable alternative. We ought to reject the redefinition of “meritocracy” to mean “positions go to people based on their class and ability to go to Harvard”, and reclaim it as meaning exactly what we want instead – positions going to those who are best at them and can best use them to help others. Which is what we want.

(None of this solves one of the biggest problems that the anti-meritocracy folk are complaining about: the fact that there’s a distinction between millionaire Goldman Sachs analysts and starving poor people in the first place. I’m just saying that in a world where somebody has to be an investment banker, a surgeon, or a Federal Reserve chair, I’d rather choose them by true meritocracy than by anything else.)

[see here for more discussion]

Highlights From The Comment Thread On Meritocracy

[followup to Targeting Meritocracy]

Some commenters rightly question exactly what we mean by meritocracy. For example, Mark writes:

> Isn't the real problem that we have no good system to identify who deserves power over others, in the most general sense?
>
> Grant the surgeon their power, in their specific field of expertise, within their own hierarchy. But I think we have to question strongly whether we need grant them any special power beyond that.

Different considerations certainly apply to surgeons versus senators, and talking about appointment to a vague "ruling class" probably confuses things pretty badly. I'm much more willing to listen to arguments for a randomly selected Congress than I am for a randomly selected surgical staff. Maybe the problem is that, aside from a few elected officials, nobody ever notices that they're choosing people for the ruling class at all. They're just choosing economists, lawyers, bankers, et cetera, for the particular purposes of their institution/law firm/bank. Any rigorous discussion of meritocracy would have to separate these out more than anyone's done so far, and definitely more than I am going to do in the rest of this thread.

Another group of people express concern about meritocracy insofar as they define it as focusing on certain kinds of proxies for merit (standardized testing?) rather than real merit. From RSJ:

> Meritocracy is an ideological hammer to beat down those who demand consequences for failure. It is a shift from being judged based on results to being judged based on “qualifications”. It is very easy to be judged based on qualifications since that status never changes no matter how often you get things wrong. It’s a type of aristocracy that short-circuits the necessary discipline that must be applied to any elite.

Tom Bartleby tries to tease some of this apart:

> Genuine question (for Scott and everyone else): what is the “meritocratic” outcome in the following hypothetical:
>
> Alice and Carol are both programmers, and are up for a promotion to management. Alice is smarter, works harder, and produces better code. She gets along well with everyone and is consistently rated as the highest performer in the group. By contrast, Carol is consistently a mediocre programer. She’s not awful—certainly in no danger of being fired. But she’s not as smart, she doesn’t work nearly as hard, and her code is acceptable rather than excellent.
>
> On the other hand, Carol has a real knack for management. When she’s in a group project, she naturally takes the lead and others feel comfortable deferring to her. (Alice is more likely to just do an unfair share of the work herself.) Looking at the two of them, we can confidently predict that—even though Alice is the better programer—Carol would be the better manager.
>
> So, is it more meritocratic to promote Alice or Carol?

> I would say that it’s more meritocratic to promote Alice. If a company has the habit of promoting people like Alice, I would describe that company as having a meritocratic culture. I get the feeling, though, that Scott disagrees.

Some Faceless Monk had the same worry, but was more fatalistic about it:

> I think in some ways, meritocracy as it is practiced as opposed to meritocracy as it is idealized is in play here. Companies that incorporate meritocracy start out in the idealistic manner: Choosing based on merit and ability. However, over time, companies (especially big ones that get hundreds or thousands of applications) will start to make the hiring process efficient, and gloss over a lot of details by weighing on specific factors: “Oh hey, these guys came from X school and were in Y organization and have done really well for us! We should pay attention to more applicants that have X and Y.” Or “Oh, these two guys have experience from Z company and were fired two months in. Are we sure we want applicants from Z?” The reasons why X, Y, and Z matter are almost never analyzed, and instead these name just get turned into keywords for the applicant tracking system to filter. This can lead to, worst of all, “This applicant may have the abilities, but they went to A school, served in B organization, and worked at C company. And I haven’t heard any of these!” What you witness is an institutionalized form of quasi-nepotism, in that your application gets weighed on by the names on your resume rather than what you did with those names. That’s what I really think these publications are deriding, they just call it meritocracy because they can’t think up a good word for it.

MartMart was more fatalistic still:

> If word X should mean X, but thru out known history has always meant Y, it's not unreasonable to claim that you oppose X on the grounds that it always results in Y which is terrible. I mean, people who oppose soviet style communism do just that.

I would counter-argue that people still use words like "justice" and "equality" despite their similarly dismal histories. If we have to abandon a good-sounding word just because the people who claim to practice the good-sounding word usually don't, we're not going to have a lot of good-sounding words left.

A third group of people have more fundamental concerns that apply even to ideal meritocracies. A common worry was that if all the meritorious people end up in the upper-class, then the upper-class has complete power and the lower classes don't have anyone competent left to represent their class interests. For example, dndsrn writes:

> [Young's] attack on meritocracy – really, the original attack on meritocracy – was not "gee it's awful convenient how the people on top have come to the conclusion that society puts the best on top" (which is, to a considerable degree, a legitimate and true criticism) – his attack on meritocracy was that it would strip the working classes of high-IQ individuals from those classes who in his world (the Britain of the early to mid 20th century) became union reps and Labour politicians – that a real meritocracy would leave the working classes defenceless against being snookered by the bosses.

And Iain quotes part of the Guardian article:

> It is hard indeed in a society that makes so much of merit to be judged as having none. No underclass has ever been left as morally naked as that.
>
> They have been deprived by educational selection of many of those who would have been their natural leaders, the able spokesmen and spokeswomen from the working class who continued to identify with the class from which they came.Their leaders were a standing opposition to the rich and the powerful in the never-ending competition in parliament and industry between the haves and the have-nots.
>
> With the coming of the meritocracy, the now leaderless masses were partially disfranchised; as time has gone by, more and more of them have been disengaged, and disaffected to the extent of not even bothering to vote. They no longer have their own people to represent them.

I’ve heard this argument before in the context of segregation and immigration. That is, when segregation ended, many of the upper-class black people who could move to white neighborhoods did, stripping black neighborhoods of their potential leadership. And when the doctors and lawyers in a Third World country immigrate to America, it creates a brain drain back home.

Both of these are recognized as difficult problems, but the meritocracy version seems even harder. Once someone from the lower class becomes a Senator, they’re not so lower-class anymore; this seems like a natural problem in any governmental system. I’m not going to say it’s tautologically impossible, because there are ways to keep them more or less in touch with their lower-class roots, but it does seem like a harder problem than a lot of people give it credit for.

This naturally segues into another class of critique: meritocracy unifies all of the talented people into a hegemonic upper class with its own values, disconnected from the people they're supposed to rule. RSJ again:

> Meritocracy-as-practiced means herd behavior as a very small group schools (both intellectual and actual) produce the leaders who echo each other's conventional wisdom. This is how we get entire nations pursuing economically or militarily disastrous policies, such as the whole western world deciding it needed to go back on the gold standard after WWI, or, for that matter, WWI. Or the current tragedy of Greece, and the sadism of the European Monetary Union. It's how all economists agree that we should tax consumption rather than unearned income. It's how we got financial de-regulation, wall street bailouts, a flatter tax schedule, a shrinking middle class. When these are deeply unpopular beliefs among common, less "meritocratic" people who didn't all go to the same 5 elite graduate schools.

And Jaskologist quoting dndsrn:

> Their cluelessness, lack of self-awareness, and lack of empathy for people they consider below them is absolutely breathtaking. "Let them eat cake" level stuff. They can't understand that their high IQs are not earned, and that intellect is not a moral quality (as an aside, I think this is part of the appeal of blank-slatism to intelligent people: if they ignore that IQ is probably about 50% inherited, and most environmental factors are out of their control, they can pretend that their university degrees and so on simply show their high

> quality as individuals, instead of showing that they rolled well for INT at character creation). They can't understand why all those factory workers who want to keep their jobs, or want the jobs to come back to town, instead of learning to code and moving to the Bay, or getting a business degree and moving to London or NYC, or getting a law degree and… etc. Their mastery of skills that allow them to pick up and move pretty much anywhere and earn well doing it mean that they have little consideration, respect, or loyalty for their countrymen who cannot. The people from all over the world working in finance in London feel loyalty to each other – after all, they are the best, are they not? – far more than they do to the peons from wherever they come from.

I hope that deemphasizing education in favor of skill will be of some help with this – after all, where do these people learn their class solidarity and distinct values except at Harvard and Oxford? When I hear rags-to-riches stories from a bygone era, they always involve the guy who did such a good job as a waiter that he became head waiter, then restaurant manager, then head of the restaurant chain. That seems both most truly meritocratic, and like a strong antidote for the identical-Oxbridge-clones problem.

I really don't think that self-contained elites are meritocracy's fault. The hereditary aristocracy wasn't exactly famous for avoiding the failure mode of becoming a cloistered elite who talked only among themselves and ignored the people they were supposed to rule. Has there ever been a system that was any good at this?

A final class of commenter takes this to its logical conclusion and says that the problem isn't rule based on merit, it's rule, period. From qwints:

> Young is proceeding from a socialist perspective by looking at classes means of reproducing themselves. His key emphasis is on the suffix – the ruling done by the intellectual elite. The problem is not at all an inequality of opportunity, it's the power given to those who've taken the opportunity and the moral authority they wield.
>
> These critiques are really saying that letting the most able rule is, in fact, a bad thing – even worse than letting all the important jobs go to aristocrats (at least for Hayes and Young). They're really saying that the seductive nature of the claim "we should give out jobs based on merit" is dangerous, and the claim must be opposed. The solution they offer is getting rid of the idea of ruling altogether.

I was also lucky enough to get a response from Andrew Granato, author of the Vox article linked in the piece. He wrote on Twitter:

> Even if we had some extremely accurate way of identifying the most talented people and allocating them to the top positions, we would still have the same structural force at play that mars America now: the stratification of society into increasing distant tiers. Except now the stratification would be more based on "merit" than what we have now- which is what Ivies sought to do in the 60s and 70s.
>
> Seems reasonable to claim that there are ways of finding better elites than we currently have. But it would still generate elites by design. And whenever you structurally give people money and power, you give them the means to seek and extract rents from society.

Okay, so, uh, the problem is that we “structurally give people money and power”. And the solution is “getting rid of the idea of ruling altogether”. That sounds nice and straightforward. Let’s get a couple of grad students to write a white paper on it and try to have it implemented by next quarter.

Okay, fine, I’m being mean. But it does seem like a lot of these solutions are utopia-complete; that people’s objection to meritocracy is that it’s not a perfectly just world where everyone is free and equal and prosperous and lives in harmonious understanding and nobody has power over anyone else. I agree this is a pretty good objection to a lot of things. But it doesn’t seem to be an objection to anything *in particular*.

I guess what I mean by this is…suppose I attack welfare. I write a bunch of articles like “Welfare Is Destroying America” and “We Need To Smash Welfare” and “Ten Reasons Why Welfare Ruins Everything (Number Six Will Astound You!)”. When questioned, I admit my main objection is that welfare is inferior to a world where everyone is rich. There’s no way my criticism helps produce the everyone-is-rich world, but it’s super-likely that it helps people like Paul Ryan who just literally want to destroy welfare, in the normal sense of destroying welfare.

And look. A couple days ago, Donald Trump nominated Sam Clovis as the Department of Agriculture’s chief scientist. Clovis is a right-wing talk radio host who has “never even taken an undergraduate course in any science”, and believes global warming is a scam. This follows a few months after Trump appointed his son-in-law as one of the nation’s top diplomats. And I wish I could say this is one of those completely unique Trump things that we keep being told we should “never normalize”, but it’s just an exacerbation of politics as usual. We justly celebrate the decline of the spoils system in much of the civil service, but it never totally disappeared, and I wouldn’t want to speculate on how common it is today, whether it’s going up or down, or anything like that.

The most salient alternative to welfare isn’t everyone-being-rich, it’s poverty. The most salient alternative to meritocracy isn’t perfect equality, it’s cronyism. If people keep criticizing meritocracy, eventually the word is going to become uncool, it’ll be impossible to advocate for it without giving three boring paragraphs worth of qualifiers that put everyone to sleep, and it’ll become that much harder to criticize cronyism or advocate for something different.

And for that matter, what *is* the anti-meritocracy endgame? I agree that it’s bad when people at the top can claim they’ve gotten their positions based on merit, but how do we prevent that other than by *not* giving those positions based on merit. If we don’t give positions based on merit, what do we give them on? Affirmative action doesn’t solve this problem, just punts it down a step to “most meritorious woman or minority”. Should we return to a hereditary aristocracy? Just let people hire their sons-in-law more? Throw a dart at a phone book and appoint whoever it hits? What are we going for here? I honestly want to know.

One point I keep pushing on this blog is that it’s a bad idea to demand downstream solutions to upstream problems. For example, I’ve argued that if a company’s applicant pool is only 20% women, and the company engages in gender-blind hiring and gets 20% women employees, it’s more useful to focus on the factors shaping the applicant pool composition than it is to yell at the company. For some reason nobody (sometimes including me) seems very good at this.

But this same problem seems to be shaping discussions of meritocracy. If you don’t like the fact that the CEO of Goldman Sachs exists, that’s a pretty reasonable upstream problem to have. If instead you complain about the downstream problem that he’s chosen based on merit, all you’re going to get is more people appointing their son-in-laws.

Other miscellaneous good comments: baconbacon on which seemingly-unmeritocratic rules are just excuses to protect people’s feelings. Nikolai Rostov on the difficulty of measuring merit in different domains. And especially rminnema linking to an excellent article on how the Soviet upper class always managed to get their kids into top schools despite the system’s supposed Communist bent.

Don’t Blame Griggs

Griggs vs. Duke Power Co seems to come up a lot here as a scapegoat. This is the Supreme Court case that said companies can’t use anything like an IQ test to help in job interviews unless they can prove in court they’re not being racist. Since this is hard to prove, most people play it safe and avoid these tests.

So (opponents of the case figure) this is the reason we’ve gotten so bogged down in credential-ocracy. Along with whatever particular skill they’re going for, employers want generally smart people. They’re not allowed to test for that directly, but someone who can finish a college degree is probably pretty smart. So employers demand a college degree as a minimum requirement. And we end up with this farce of someone going $50,000 into debt to study Art History for four years so they can get a job in marketing.

This is the story, but I don’t think it’s true.

The Griggs decision explicitly places the same restrictions on college degrees that it does on IQ tests. Obviously nobody cares about this, and it turns out the Griggs restrictions aren’t that strict and most people manage to ask about college degree just fine. This changes the argument to be about whether there’s complicated case law around Griggs which makes employers figure that college degrees will probably be accepted in practice, but IQ tests probably won’t be.

But I don’t think that’s the main cause either. Other countries don’t have their own version of Griggs vs. Duke. I don’t know too much about their labor market, but I think many of them have the same problem. This UK website starts by saying “the fight for a well-paid job without a degree is a tough one, but there are still a handful of roles out there in which you can earn serious money without a degree-level qualification,” which seems like the same over-optimistic language you might hear on a US article saying the same. I also remember that when I went to teach English in Japan – a job which required no teaching credential and only the English-speaking ability typical of any Anglophone child – they still insisted I have a Bachelor’s in something before letting me apply.

Another point against – I think it’s Griggs v. Duke-compliant to ask people their SAT scores. See Wall Street Journal: . And TIME: Your SAT Scores Will Come Back To Haunt You. Contrary to these articles’ predictions, I don’t think asking SATs in job interviews ever really caught on, but that’s not the point. The point is it seems to be legal and a known thing which companies could do if they wanted. Given that the SAT more or less approximates an IQ test, if there were some pent-up demand for IQ-testing job candidates the SAT would be a perfectly good alternative. Given that only the handful of companies in the articles above ask for SAT scores, I don’t think employers are really that interested in IQ.

This leads to a third objection; there are some fields where standardized test scores are universally available, and they end up as credentialocratic as everywhere else. In medicine, for example, all applicants take the MCAT, a pre-med focused standardized test. But medical schools in the US still require a college degree plus a bunch of universally-recognized-as-irrelevant requirements (all pre-meds have to pass a Calculus II class, even though Calculus II never comes up in medicine). All medical students are required to take several USMLEs and report the results to the residencies they apply for, but which medical school you went to still matters a lot if you want a residency in a competitive specialty. In fact, residencies are infamous for turning down candidates from foreign schools with genius-level USMLEs in favor of US-trained candidates who barely managed a pass, for unclear reasons. Lawyers have the LSATS (and possibly bar exam scores?) and are also pretty famous for being judged on what school they went to.

All of this makes me think that, as nice as it would be to attribute everything irrational about credentialocracy to one bad Supreme Court decision, there’s probably something wrong on a deeper level.

Why Is Clozapine So Great?

[Epistemic status: Very speculative; I don't fully understand a lot of the studies involved]

Clozapine is an antipsychotic drug sometimes used to treat schizophrenia. Like most antipsychotics, it works by blocking dopamine receptors in the brain. Conventional wisdom among psychiatrists goes that most antipsychotics are about equally good – except for clozapine, which is better.

It's fun to listen to psychiatrists, usually pretty quick to admit how crappy a lot of the drugs they prescribe are, wax rhapsodic about clozapine. From Joober & Boksa:

> Consensus of opinion is rare in psychiatry. Even in the field of clinical trials, where experimentation is tightly controlled and regulatory bodies scrutinize the proof, controversies are frequent and difficult to resolve. One issue for which there is a widespread consensus is the unique place that clozapine occupies in the treatment of severe mental illnesses, particularly refractory schizophrenia. This molecule is distinct because of its effectiveness, numerous and sometimes mysterious pharmacologic characteristics, serious side effects and under use…
>
> Every clinician who has prescribed clozapine can recount a few experiences of seeing patients emerge from their chaotic psychotic experience. This is one of the most rewarding experiences that a psychiatrist can have in his or her professional life, and it is among the most important strikes we have made against one of the most devastating diseases affecting mankind.

Psychiatrists wish they could give clozapine to everyone who needs an antipsychotic. They can’t, because its increased efficacy goes side-by-side with greater side effects. The best known are agranulocytosis, metabolic syndrome, seizures, myocarditis, and eosinophilia, and the list just goes on from there.

This has led a lot of people to wonder: why is clozapine so uniquely great? And can we get a version which is just as good without the side effects?

Recently there’s been increased interest in the glutamate system (especially NMDA receptors) in schizophrenia, and in glutamatergic compounds as possible treatments. Various teams have taken schizophrenic patients already on antipsychotics and added NMDA modulators, especially d-serine, glycine, and sarcosine. Meta-analyses have been guardedly positive (Tiihonen & Walbeck) or explicitly positive (Singh & Singh, Tsai & Lin). There’s a widespread hope that the next generation of antipsychotics will be glutamatergic drugs which are able to attack more symptoms than the dopaminergics we have today.

But what if that next generation were already here?

Both of the recent meta-analyses of glutamatergic augmentation of antipsychotics noted the same exception. From Singh & Singh:

> When added to clozapine, none of the drugs demonstrated therapeutic potential

And from Tsai & Lin:

> Patients receiving risperidone or olanzapine, but not clozapine, improved.

Why should this be? A couple of recent studies have converged on an exciting possibility: clozapine is a combination antipsychotic + NMDAergic. That is, NMDA glycine site agonists don't add anything to clozapine because clozapine is already agonizing the glycine site.

This is the suggestion of Schwieler et al, based on their electrophysiology studies in rats. They find that clozapine causes a characteristic change in the firing rates of certain rat neurons, a change which is reversed by the glycine site antagonist kynurenic acid. They conclude (I don't know enough to confirm) that:

> The enhanced response of [ventral tegmental area] [dopaminergic] neurons to clozapine seen following lowered [kynurenic acid] is what should be expected from a partial NMDA/glycine-site agonist.

Javitt et al also work with rats, and find that clozapine "inhibited transport of both glycine and MeAIB, but not other amino acids, at concentrations associated with preferential clinical response". Other antipsychotics do nothing of the sort. A lot of their biochemistry is a little too in-the-weeds for me, but I think what they're saying is that clozapine increases natural extracellular glycine levels, and so really is a direct analogue of medicinal glycine administration. They conclude:

> This study suggests first that System A transporters, or a subset thereof, may play a critical role in regulation of synaptic glycine levels and by extension of NMDA receptor regulation, and second that System A antagonism may contribute to the differential clinical efficacy of clozapine compared with other typical or atypical antipsychotics.

Jardemark et al find that clozapine facilitates NMDAergic neurotransmission through something called protein kinase C. My head is starting to hurt trying to keep track of all of these different chemicals, and in particular I'm not sure whether all of these people are positing different and incompatible mechanisms or if they can be unified into one big biochemical pathway. But it sure looks like a lot of people have found some kind of interesting NMDAergic effect.

This doesn't really prove anything. For one thing, a bunch of drugs coincidentally have NMDAergic effects that have nothing to do with their mechanism of action – eg the antibiotic cycloserine treats tuberculosis and happens to be a pretty good NMDA agonist on the side. For another thing, what if all antipsychotics have NMDAergic effects? Blocking dopamine is probably going to do something or other upstream. Right now evidence seems pretty mixed on this, with one study suggesting they do (1) and a couple suggesting they don't (1, 2, 3).

If this really was the source of clozapine's special powers, it would be a really important big breakthrough. All through residency, I kept hearing "nobody really knows why clozapine is so great" or "it probably has something to do with its weak D2 binding affinity or something". If we knew it was just because clozapine was an antipsychotic plus glycine, we could just give people an antipsychotic plus glycine, and avoid the agranulocytosis, metabolic syndrome, seizures, myocarditis, eosinophilia, etc.

This theory isn't ready for prime time yet; it's still not *really* proven that the NMDA agonists work for schizophrenia at all. And it probably never will be – it's hard to patent these chemicals, meaning that it would take extreme creativity to jump the bureaucratic hurdles necessary to get them into the pharmacopoeia. Until that happens, using these chemicals will remain experimental, and replacing clozapine with them downright irresponsible.

Still, it’ll be hard to watch treatment-refractory schizophrenia patients get their agranulocytosis, metabolic syndrome, seizures, myocarditis, eosinophilia, etc without wondering whether another way is possible.

Djoser Joseph Osiris

My recent move has already paid off in terms of increased access to the Bay Area intellectual milieu, by which I mean wacky outlandish hypotheses about completely random stuff. The other day a few people (including Ben Hoffman of Compass Rose) tried to convince me that Pharaoh Djoser was the inspiration for the god Osiris and the Biblical Joseph. This sort of thing is relevant to my interests, so I spent way too long looking into it and figured I ought to write down what I found.

The short summary is that the connection between Djoser and Osiris is probably meaningless, but there's a very small chance there might be some tiny distant scrap of a connection to Joseph.

Djoser, who ruled Egypt around 2680 BC, was a pretty impressive guy. Egypt had been unified by one of his predecessors a few generations before, but they'd let it get un-unified again, and Djoser's father was the one who reunified it. Djoser inherited a kingdom of newfound peace and plenty – and made the most of it, starting lots of impressive infrastructure and religious projects. His chief minister Imhotep was famous in his own right as a polymath who invented medicine and engineering (he may also have been the first person to use columns in architecture). He was later deified for his accomplishments. Djoser and Imhotep cooperated to build the first pyramid, the Step Pyramid at Saqqara.

Osiris was a legendary god whose worship was first recorded around 2400 BC. The legends say he was a wise and benevolent Pharaoh of Egypt in some unspecified distant past before being killed by his brother Set. One thing led to another, and he eventually ended up as the god of death and resurrection and the underworld. Scholars have long debated the exact origins of the Osiris cult, and tend to attribute it to some historical memory of something or other but disagree viciously over the details.

The argument I heard for Djoser inspiring Osiris hinges on a couple of points (there may be others I didn't get). First, the times sort of match up – this legend of the wise king Osiris first appears just a century or two after Djoser died. second, Djoser was a big fan of an Egyptian symbol called the 'djed', a weird column shape thing. Djoser included djeds all over the step pyramid he and Imhotep built together, and may have kind of had an obsession with the thing (and why shouldn't he? – if I helped invent the column, I'd talk about it a lot too). Meanwhile, the djed is traditionally considered the symbol of the Egyptian god Osiris. And third, if you're going to deify a pharaoh into the god of death and resurrection, the beloved and powerful ruler who invented the first pyramid sounds like a pretty good candidate.

I think this argument is probably wrong. For one thing, although nobody can prove Osiris existed before the death of Djoser, everybody *suspects* that he did. In The Origins Of Osiris And His Cult, Egyptologist John Griffiths appeals to some early inscriptions that *might* name Osiris, and concludes that

> There is a strong likelihood that the cult of Osiris began in or before the First Dynasty in connection with the royal funerals at Abydos, [although] archaeological evidence hitherto does not tangibly date the cult ot an era before the Fifth Dynasty

A common consensus is that he began as a local deity of the city of Busiris and (as mentioned above) the necropolis at

Abydos. Djoser has no connection to either city, and in fact was the first king *not* to be buried at Abydos. His building a pyramid is less impressive than it sounds; all the Egyptian rulers were into building big tombs, and he just took it to the next level.

Djoser liked djeds, but so djid lots of Egyptians. They were popular long before Djoser and remained popular long after him; among their many fans may have been such pharaohs as eg Djedkara, Djedkheperu, Djedkherura, and Djedhotepre. The djed started out as a general fertility symbol, later became a symbol of the god Ptah, and only became fully associated with Osiris a thousand years after Djoser's djeath. This makes it hard to argue Osiris got associated with the djed because of some cultural memory of Djoser.

This is kind of weak evidence against the theory – a speculation that Osiris is older than he looks, a little bit of confusion around when Osiris became associated with his sacred symbol. But it was a weak theory to begin with, so weak counterevidence convinces me.

So let's get to the more interesting claim – that Djoser inspired the Biblical Joseph.

This comes from a monument called the Famine Stele, written two thousand years after Djoser's death but telling a legend that had grown up around him. According to the stele, in the time of Djoser there were seven years of famine. Djoser asks his chief minister Imhotep for help. Imhotep investigates and finds that the problem is related to the god Khnum. He prays to Khnum and offers to worship him better, and Khnum appears to him in a dream and says that okay, he'll make the crops grow again. Djoser and

Imhotep repair Khnum’s temple, the famine ends, and everyone lives happily ever after.

The parallels to the Joseph story are pretty apparent. A seven year long famine. A Pharaoh who begs his chief minister to do something about it. A dream that provides the solution. Sure, the crocodile-and/or-ram-headed god Khnum gets left out of the Biblical account, but that sounds like just the sort of thing the Hebrews would conveniently forget.

There are some other differences, of course. The Joseph story involves seven years of plenty before the famine; the Imhotep story doesn’t. Joseph gains his chief ministerial position because of the famine incident; Imhotep is already in charge when the famine begins. God gives Joseph a rational planning strategy; Imhotep uses divine intervention directly. But isn’t there still a suspicious core of similarity here?

Creationists think so. They get really excited about this connection since it seems to link the Bible to a verified historical event (for values of “verified” equal to “someone made a stele about it two thousand years later, and in fact *after* the Bible itself was written”). Back during the presidential campaign, Ben Carson got soundly mocked for saying the pyramids were silos for storing grain. Everyone attributed this to his warped fundamentalist Christian view of history, but nobody thought to ask why fundamentalist Christians seized on this falsehood in particular. The answer is: if the pyramids were grain silos, then the link between Joseph (whom the Bible says built grain silos) and Imhotep (whom Egyptian records say built pyramids) becomes even more compelling.

Awkwardly for the creationists, this doesn't even match their own hokey Biblical history. There are various different Biblical chronologies, but they mostly date Joseph around 1900 – 2000 BC – too late to be Imhotep, who lived closer to 2600. Also, don't tell anyone, but the Bible is probably false.

Atheists have a better option available – they can claim that the Egyptian legend of Imhotep inspired the Israelite legend of Joseph. This is the strategy taken by a Ha'aretz article, which first roundly mocks any identification of Imhotep with Joseph, saying that this makes no sense and is totally stupid, and then adds:

> There is a consensus among the majority of biblical scholars that the Joseph story dates, at the earliest, to the 7th century BCE, namely 2700 years ago. Many Judahites were residing in the Nile Delta at the time, as proven among other things, by the existence of a replica of the Jewish First Temple in Jerusalem on the island of Elephantine. It seems these Judahites may have been behind the adoption of the Imhotep tale as an Israelite story.

It doesn't cite which scholars it's talking about, or explain why they suddenly backtracked from their "there is no connection between Joseph and Imhotep shut up you morons", but the overall point seems pretty plausible. Remember, the 7th century would have been just a few centuries before the Famine Stele was written, and the Djoser/Imhotep famine legend might have been popular in Egypt around this time. It sounds just barely possible that some Jews might have rewritten it with an Israelite protagonist the same way a bunch of pagan goddesses and even the Buddha ended up as Christian saints.

(or, for that matter, the Egyptians could have rewritten the Bible story with an Egyptian protagonist, although it seems less likely for cultural transmission to go that direction given the two cultures' relative sizes.)

Or maybe none of that happened. Wikipedia's article on the Famine Stele points out that a seven-year famine was a weirdly common motif all across the Ancient Near East, citing eg the Epic of Gilgamesh:

> Anu said to great Ishtar, 'If I do what you desire there will be seven years of drought throughout Uruk when corn will be seedless husks. Have you saved grain enough for the people and grass for the cattle? Ishtar replied. 'I have saved grain for the people, grass for the cattle; for seven years o£ seedless husks, there is grain and there is grass enough.'

I don't know if all of this derives from the same proto-Near-Eastern source, or whether seven year famines are just sort of a natural kind (compare all the different cultures that have something like "may your reign last a thousand years!"). But it warns us against leaping into accepting this too quickly. This is especially true in the context of atheists' haste to believe things like "Christ is just a retelling of the Osiris myth!" or "What if Moses was really just Akhenaten" that later turn out to not really make that much sense. Part of the lesson I wanted to teach with Unsong is that this sort of thing is too easy, and therefore we need to increase our guard. I don't know how to weight this, but maybe say there's like a 30% chance

As a perfect example, here's a completely insane work of Biblical apologetics claiming that a totally different pharaoh associated with djeds corresponded to the Biblical Joseph.

Against Signal-Boosting As Doxxing

A recent spat on Twitter, which I won’t link: some guy using his real name tweeted an offensive joke about how women should make sandwiches at a group of women. A feminist columnist with tens of thousands of followers retweeted with the comment “This is a young man who ostensibly wants a job someday, tweeting at professional women in his field under his own name…RT to help ensure [REAL NAME]’s prospective employers know this when they search for [REAL NAME]’s name”.

[EDIT: See here for discussion of various complicating factors; my claim isn’t going to be that a completely innocent person was punished, so much as that this entire paradigm of punishment is dangerous]

What particularly bothered me about this situation was that the columnist involved was a libertarian who writes for Reason, and her supporters were mostly other influential libertarians. And they were all using the old argument that the concept of “free speech” came into existence *ex nihilo* on December 15, 1791 with the ratification of the First Amendment, and has no meaning or significance outside a purely legal context of delimiting government power.

I have a friend who grew up gay in a small town in Alabama, where “faggot” was the all-purpose insult and the local church preached hellfire as the proper punishment for homosexuality. He unsurprisingly stayed in the closet throughout his childhood and ended up with various awful psychological problems.

If you’re a very stupid libertarian strawman, you might ask whether that town had any anti-gay laws on the book – and, upon hearing they didn’t, say that town was “pro-gay”. If you’re not a very stupid libertarian strawman, you hopefully realize that being pro-gay isn’t about boasting how progressive your law code looks, it’s about having a society where it’s possible to be gay. Not having laws against locking up gay people is a necessary precondition, but it’s useless on its own. You only get good results if good laws are matched by good social norms.

Likewise, the goal of being pro-free-speech isn’t to make a really liberal-sounding law code. It’s to create a society where it’s actually possible to hold dissenting opinions, where ideas really do get judged by merit rather than by who’s powerful enough to shut down whom. Having free speech laws on the books is a necessary precondition, but it’s useless in the absence of social norms that support it. If you win a million First Amendment victories in the Supreme Court, but actively work to undermine the social norms that let people say what they think in real life, you’re anti-free-speech.

But I’ve discussed this before at more length. What I want to get into here is a point specific to this situation: the guy made this joke under his real name. All the Reason columnist did was retweet it and add some commentary about how she hopes he becomes un-hire-able. This isn’t doxxing. It’s not even divulging a secret; the guy said it on his public Twitter. Is it really so wrong to do what’s basically just signal-boosting his comment?

A quick philosophical digression: what are we even doing here? My thought is: we’re trying to hash out a social norm. We expect this social norm to be sometimes in our favor and sometimes against us, so we want it to be universalizable and desirable under a veil of ignorance.

On that note: let him who is without sin throw the first stone. Have any of you ever said or done anything which, if signal-boosted, would be very embarassing and might prevent you from getting a job?

Before you answer, consider this: the person signal-boosting you has much wider reach than you do. There are now tens of thousands of people in the world who know you *only* as the guy who said that one embarassing thing one time. For that matter, anyone who Googles you will know you only as the guy who said that one embarassing thing one time. All of your triumphs, all of your defeats, all your loves and fears and follies – none of these exist in the public mind. If you cross a blogger, a columnist, or a Twitter celebrity, all that will exist is that you once retweeted a racist joke on the 26th of March, 2014.

Never retweeted a racist joke? Someone will find something. Maybe you've been a sex worker once – hope you didn't put your picture up on the Internet, or else Reason columnists will say it's not "doxxing" to merely "signal-boost" it so that everyone knows. Heck, even *watching* porn is enough to get people fired some places. Maybe you were stupid enough to admit you were gay or trans under something traceable to your real identity. Maybe you voted for Trump (a firing offense in some places) or against Trump (a firing offense in others). Maybe you committed a crime someone can find on a public crime database, or maybe you said something perfectly innocent which can be twisted into a sinister "dog whistle" out of context.

My own story – some antipsychiatry crackpot decided to target me, went through a couple of posts I'd written defending the practice of involuntary psych commitment in certain cases, and took a few statements out of context to make it look like I thought we should lock up all mentally ill people and throw away the key. Then he posted it on an antipsychiatry website, asking if anyone could find the address of my workplace so he could send it there to prove that I was unfit to work with the mentally ill. Luckily the moderator contacted me and deleted the post, and it stopped there. And it was never that convincing an effort to begin with. But…

In a world where an average of 250 resumes are received for each corporate position, how convincing does an effort have to be to ruin somebody's life? Do you think your dream company is going to spend a long time sorting through each claim and counterclaim to determine that the highly-Google-ranked page about you claiming you're unfit to work in your industry is mostly unfair? No. They're just going to cut their risks and move on to the other 249 candidates.

Here's an exercise which I encourage you to try. Suppose there's a Reason columnist who wants to get you fired. They pore over your public statements – Twitter feed, Facebook timeline, any blogs you might have written, anything you've said in mixed company that you don't know if somebody else wrote down waiting for the time they could use it against you. Imagine the most incriminating dossier of your statements, out of context, that they could put together. Imagine what would happen if they were pretty determined, and sent it to your workplace, your church, your parents, et cetera. How much of your life could they destroy?

And I agree this is weird. It's bizarre that so many people trust to security by obscurity, when anybody with an axe to grind can destroy their obscurity and reveal them to the world. It's bizarre that we treat Twitter as a private place, when literally everything that happens there is visible to every human being on Earth. It's bizarre that we trust to these fragile online identities when any hacker can cut through them, bizarre that we wear such different masks to different friends when they could just talk and compare notes, bizarre that we dare to talk at all when we know every word we say is logged and the future may be less forgiving than the past.

But don’t let the fact that it’s bizarre make you think it isn’t important. How many of us can say, honestly, that we could bear the Panopticon? If every valley were raised up and every mountain pulled down, so there was nowhere to hide, and we were rendered naked to any eye anywhere in the world, how long could we endure? Wouldn’t we retreat into ourselves, turtle-like, afraid to ever speak at all?

And who would enjoy this new flattened landscape more than the biggest and most predatory? In the Panopticon, any celebrity with a platform can destroy the lives of any ordinary person, just by mentioning them. It would be paradise for any petty tyrant with a blog, and hell for anybody too poor to tolerate a risk of losing their livelihood.

I have a pretty big blog. But other people have bigger ones. I’m not confident that the amount of fun I could have destroying the reputations of people I don’t like outweighs the chance of someone else destroying mine. I’m certainly not confident that the aggressive-signal-boosting power would mostly end up in the hands of good people. So I reject the entire tactic. I think it’s morally wrong to try to signal-boost people’s bad behavior – even their semipublic bad behavior – to get them fired. Probably there’s a lot of subtlety here and there have been times in the past I’ve supported cases that seem completely different to me but might seem similar to others. I admit there’s an argument that doxxing is a way of shaming people in order to enforce social norms, and that we need *some* way to enforce social norms eg the one against offensive jokes – though see my post Be Nice, At Least Until You Can Coordinate Meanness about good and bad ways to do this. But for now I just am very suspicious of the whole enterprise.

Lord Byron wrote of his political philosophy:

> I wish men to be free
> As much from mobs as kings; from you as me

I stand with Byron. But I worry there’s a big strain of libertarians today who don’t. Who wish men were free from kings, but not from mobs. Who wish men were free from others, but *definitely* not from them.

All I can say to that is – it’s a package deal, people. Either promote good social norms, or be destroyed by the bad ones when the tide turns against you. That’s the only choice on offer.

Book Review: Raise A Genius!

I.

A few months ago, I learned about Laszlo Polgar, the man who trained all three of his daughters to be chess grandmasters. He claimed he could make any child a genius just by teaching them using his special methods. I was pretty upset because, although he had a book called *Raise A Genius*, it was hard to find and only available in Hungarian and Esperanto.

Many SSC readers contributed money to get the book translated, and Esperanto translator Gordon Tishler stepped up to do the job. Thanks to everyone involved. You can find his full translation here: *Raise A Genius!*

I was hoping that this book would explain Lazslo Polgar's secrets for raising gifted children. It does so only in very broad strokes. Nor does he seem to be holding much back. But it looks more like he doesn't really have secrets, per se. The main things he does differently from everyone else are the things he's talked about in every interview and documentary: he starts young (around the time the child is three), focuses near-obsessively on a single subject, and never stops. Polgar:

> The first characteristic of genius education – I could say the most important novelty distinguishing it from contemporary instruction – and its necessary precondition, is early specialization directed at one concrete field. It is indeed true what Homer said, "A person cannot be experienced or first in everything." Because of this parents should choose a specific field at their discretion. It is only important that by the age of 3-

> 4 some physical or mental field should be chosen, and the child can set out on their voyage.

He has a couple more things to say, but they're more like vague principles than like specific details. The rest of the book is his opinions on the meaning of genius, his gripes about the Hungarian government, the ways public schooling destroys children's natural creativity, and various related subjects.

And maybe this stuff deserves some attention. He spends a long time responding to people who say it's inhumane or immoral to educate children the way he does it, and certainly those claims need a response. A lot of his pedagogical philosophy and personal philosophy of life come out in the way he answers these questions, and given how few specifics he gives, maybe understanding his broader worldview is the way to go. And although a lot of people talk about how public school destroys children's minds, it's always good to hear it from the mouth of somebody who's put his money where his mouth is and done a better job.

But what can we glean from this book in terms of how one can educate a child in the Polgar method?

The closest *Raise A Genius* comes to anything like a specific prescription is Polgar's description of what a day might be like in some kind of imaginary Polgar genius school:

> In genius education it is necessary that the pedagogue (whether the parents or professional teachers or tutors) stay in direct, constant and intensive contact with the child. Because of this we imagine groups of only 10-15 members. In practice an intensive collaborative contact between the child and an adult must be formed, in

> which the child does not feel “subordinate.” Think how advantageous it would be if the child already understands at the age of 10 that they know a great deal, that they are a person of the same value as an adult, and that in their life there is at least one field they master as well or better than adults.

As for the curriculum, it would be:

> – 4 hours of specialist study (for us, chess)
> – 1 hour of a foreign language. Esperanto in the first year, English in the second, and another chosen at will in the third. At the stage of beginning, that is, intensive language instruction, it is necessary to increase the study hours to 3 – in place of the specialist study – for 3 months. In summer, study trips to other countries.
> – 1 hour of general study (native language, natural science and social studies)
> – 1 hour of computing
> – 1 hour of moral, psychological, and pedagogical studies (humor lessons as well, with 20 minutes every hour for joke telling)
> – 1 hour of gymnastics, freely chosen, which can be accomplished individually outside school. The division of study hours can of course be treated elastically.

All of this cries out for more explanation (in particular, the humor lessons sound fascinating), but the only part he really explains is the foreign language. He quotes Frantishek Marek: “Learning foreign languages in early childhood is very important, because without that a person cannot later express themself spontaneously, rapidly, and appropriately”, and I think suggests (though I might be misunderstanding) that languages are one of the easiest things to teach young children, and so a good way to get them into the spirit of learning things. He also thinks languages are nice because they have a defined end-goal (speaking fluently) and obvious progress along the way, so children feel good about learning them. He argues Esperanto is perfect for this: as a logical constructed language, it's very easy to learn, and it convinces children that learning is fast and easy. Then with their Esperanto knowledge they'll be much better able to pick up other languages later on. I'm not really sure what to think of this – language learning might be more important if you grow up speaking Hungarian rather than English, and Polgar seems so enthusiastic an Esperantist that it's hard to picture him recommending it for purely rational reasons – but he's quite insistent on it.

This idea that children should learn things they find exciting and enjoyable – and where they keep making measurable progress – recurs throughout the book. Often it's in the context of a kind of counterintuitive point, where someone asks him "Won't kids hate having to learn so much?" and Polgar answers that kids may hate public school, where they sit around a lot and never feel like they're really mastering anything, but won't hate intensive genius education, where they actually feel like people are trying to make them good at things:

> In conditions of intensive instruction a child will soon feel knowledgeable, perceive independence, achieve success, and shortly become capable of independently applying their knowledge. Let us take an example from language learning. Let us suppose that someone visits a class for interpreters at a school for geniuses, where they are occupied for 5-6 hours with a first foreign language, Esperanto if possible. (Why precisely with this language I will clarify below.) After some months they are already corresponding with children in other countries, they participate in meetings in and outside of

> their country – and longer-lasting – where they experience serious successes, and they converse fluently in the language they have learned by then. Is this a nice feeling for a child? Yes, it is nice. Is it useful for the child? Yes, it is useful. Is it useful for society? It is useful. In the following year one can do the same with another foreign language – let us say English – and in the year after that another.
>
> The same is valid for any field of life. In this way a child really enjoys what they are doing, and they see that it makes sense. In contemporary schools students do not understand why they are learning. But in genius-education schools the children know that after a few months they will speak Esperanto, in the following year English, in the following year German, etc. Or in the field of chess; in the first year they play at level 3, after the third year at level 1, after five years as a master candidate, after 6-7 years as a master, after 8-10 years as an international master, and after the 15th year as a grandmaster. So the child sees the goal and meaning of their work.

And:

> One thing is certain: one can never achieve serious pedagogical results, especially at a high level, through coercion. One can teach chess only by means of love and the love of the game. If I may advise: one should make sure that before everything the father or mother should not diminish the child’s habit of chess playing by too much severity. We should make sure not to always win against the child; we should let them win sometimes so that they feel that they also are capable of thinking. In this way we should bring them to a feeling of success.

So how does one go about ensuring that a child loves education?

> At the start it is most important to awake interest. We should make the child aware that who learns this knows this. And chess is learnable. If we educate the child such that they can be a partner, can accept, create, and initiate, then we can always entrust them with more independent tasks. We should get the child to love what they do – to such a degree that they do it almost obsessively. The Hungarian psychologist Tamas Vekerdy warns of the same thing, that infants more easily master things that awake and draw their interest, their attention. And even at the beginning, the child should feel joy. We should not be angry, if they jump around here and there during a chess game; indeed, it is a known fact in psychology that even though a child might frolic aimlessly because of their age-appropriate character, their thoughts can still stay on the task. We should not tell them everything; we should try to get the child themself to say something! We should not ourselves make all the moves; we should try to get the child themself to make the moves! This is the so-called Socratic method, and the essence of instruction in problem-solving – projected onto chess.
>
> Of course great success is not achievable without motivation. At the age of 5-6, if the activity is sufficiently interesting, success can also function as a strong incentive. Stimulation, encouragement, and instilling passion and trust are very important. If the parents and tutors tell the child that they are foolish and bad, the child will probably truly believe this. But the opposite also applies: if we say that they are clever and skillful, they will believe that as well. They often truly believe that, and try harder to actually become so. I

> consider it a basic principle that success is extraordinarily important. When I began the experiment, I thought that although I would not let my daughters avoid failure, they would nevertheless need to grow up accompanied by success. The proportion of failure to success should be 1 to 10.

Likewise:

> The experience of success or failure, as Adler demonstrates, greatly influences the self-confidence – or uncertainty – of the child. According to P. Michel as well, the experience of success, the admiration of others, and the recognition of teachers, significantly stimulates further action, increases the trust of the child in their knowledge and ability to a high degree.
>
> According to Frank, failure, suffering, and fearfulness decrease achievement. Following a number of successive failures, even a damaging inhibitory complex can be created. With an increase in stress, action becomes more superficial and behavior less calm. Similarly, in the opinion of M. Juck, success experienced in one area increases (and failure decreases) the level of aspiration in other areas. Helm’s experiments prove that experience of success decreases the time necessary for solving later tasks, and increases the elasticity and ideational richness of the mind, while following failure there can be hindrances, rigidity, and relative ideational poverty in thinking, and problem-solving time increases.

Likewise:

> One should have great patience. We should let the child arrive at a sense of success, but we should not handicap ourselves (we should not give up major pieces or an

> advantage in pawns), because in that case the structure of the game changes. Preferably the parents or teachers should provide a temporal handicap, or weave intentional mistakes into the game, so that the child can use them for themself. During the game the tutor should organize their position on the board intentionally as is appropriate for the student and the development of the child at their age.

And:

> To awaken the child’s interest. The child should like what they are occupied with, that is, be interested in it. One must little by little accustom them to the work and create in them the unification of work and play. It is important as well that the child become accustomed to learning and working. Particular training is necessary for the workload. I call well-organized and age-appropriate work active rest. A child’s workload should be such that they experience it as active rest. Students, for example, who must attend lectures which they then enjoy, feel more rested afterwards than before. And if the speaker lectures inexpertly, they almost fall asleep from boredom and fatigue after half an hour.

There’s a lot of this, always exhorting people to make sure children enjoy being intensively educated, but always giving only vague gestures on how to do it. I suspect Polgar was a naturally gifted teacher, and his daughters naturally curious students, and that he never really encountered problems in this regard and doesn’t expect other people to either. Some of this seems apparent in his section on play:

> I think of play as a very important phenomenon, perhaps more important than do many of those psychologists who put it on a pedestal.

But play is not the opposite of work. Play is very important for a child, but in play there is an element of work. One should not separate these two factors in a child’s value system; if for example a child hears at an impressionable age, “Play, son, don’t work!” this can later result in him feeling that work is alien. On the contrary, it is my opinion that a child does not like only play: for them it is also enjoyable to acquire information and solve problems. A child’s work can also be enjoyable; so can learning, if it is sufficiently motivating, and if it means a constant supply of problems to solve that are appropriate for the level of the child’s needs.

A child does not need play separate from work, but meaningful action. Children already enjoy doing meaningful things in infancy. They like solving problems during play, even pleasurable play. The more meaningful and information-rich the problems they solve during their activities, the greater is their enjoyment and sense of success. In the end it is most important at this age to awaken enjoyment and good feelings in them.

Regarding my daughters, it is my experience that learning presents them with more enjoyment than a sterile game. I have the feeling that play deprived of information often plays only a surrogate role, of surrogate action, of surrogate satisfaction.

This is proven also by the fact that when we examine the biographies of exceptionally capable children, we find that they played much less than their peers. The profound and lengthy research of L. M. Turman in California in 1920 uncovered many differences between the play of unusually capable children and their peers. As expected, play that demanded mental action was much more interesting to the talented children. They played alone somewhat more often, compared to the control group. Susanna Millar writes in her book Psychology of Play that sometimes unusually capable children who lack peers at the same intellectual level can have difficulties in play with others. Thus I generally do not rigidly separate learning from play, or work from hobbies at an adult level.

From my point of view, workloads could be measurably increased by appropriate methods. I agree with the pedagogical tendency to ask for intensive instruction. The essence of intensive instruction lies really in using goal-directed workloads, age-appropriateness, the holding of interest, and the lived experience of achievement and success. The American G Doman thinks the same. In his analogy: as the different muscles of the body can be developed and strengthened only by regular training, so also the capabilities of the brain can only grow by means of daily training. The lack of structured logical thought and learning causes a decrease in intelligence, just as un-exercised body parts atrophy. Doman knows, on the basis of three decades of practical experience, that the brain grows most rapidly between the ages of 1 and 6, and it almost “effortlessly” assimilates knowledge. The ability to learn by play decreases after 6 years of age, when assimilation of information becomes more difficult mental work.

More on peers:

The contemporary psychological and pedagogical literature emphasizes in one sense the importance of the peer group. But in my pedagogical concept it receives a slightly different emphasis. According to me, it is not primarily important for a child to have suitable companions of the same age, but preferably to have spiritually (mentally) appropriate partners, friends worthy of the level of their intellectual capabilities. If the social relationships of a child are exclusively or for the most part limited to groups of the same age, this will slow the progress of an exceptionally capable child.

[Some say that children should not spend too long with adults, but this] is disadvantageous only if the intellectual level is too different, and if the relationship between child and adult is not suitable, for example, if the adult imposes everything on the child so they take away their independence and initiative. But if they try to correctly develop these traits in the child, it is not damaging, but on the contrary is useful. About this I do not want to say that a child should always be in the company of adults. One must find the right proportion of being with adults and peers. I believe that passing their time in the company of those who have a similar level of intellect and similar interests and sense this well in these interactions is decisive.

Zsuzsa is a concrete example: if at the age of 13 she had played chess only with 13-year-olds, who were weaker than her in many categories, this would have been less than useful for her. And for her opponents it would not have been nice to be “knocked out” in every match. Zsuzsa herself would not have profited, because she needed playing partners at a similar level, and those were
found only among adults.

However, this was not a cause for concern, as the age difference itself did not prevent friendly relationships with others, and having good friends and colleagues at

> the same time. And friendship often flowed from work relationships. Thus one’s work is at the same time a hobby.

On curriculum design:

> Of course, one should make everything appropriate to the stage! With regard to the content of instructional materials and also the duration of instruction, one should start from the traits of the age of the child, and tailor the tasks for the optimum ability of the child. At first we should only play chess for half an hour; after some time a bit more. After a week we can extend the duration. At first we should solve only simple problems, and with the passing of time we should always progress to more complicated ones. One should get the child to play a great deal, but always with suitable partners, who have a generally similar playing ability. On some occasions they can be weaker, on some stronger, so that the child experiences what winning and losing are like. But one must certainly find the right proportion. In childhood they should play rapidly, so they should play many blitz matches and those with a short time limit.
>
> […]
>
> In this case age-appropriateness is also very important. First one should learn the movements of the king. We practiced this for several days, and later we play “king against king.” The task is this: one king must reach the opponent’s baseline, that is, one must go to the other side of the board. Whoever does this first wins. If some king can stand next the other, then the game ends without a decision. When we had learned this well, we added the next piece, the pawn. In this game the goal

was the same: to get to the other side. After several days we added the rook, then the knight.

After 3-4 weeks we arrived at the queen. Understanding the queen’s mate followed later. Possessing this knowledge, we played great pawn battles during the following weeks. That is, only the pawns and the two kings were on the board. After a pawn changed to a queen we played until mate. The children really liked this. During this we started learning the knight’s moves. This is most difficult for children, but not truly a problem, although one must carefully practice this.

Later we became acquainted with the simplest mating moves. First I collected around 1,000 one-move mate diagrams; later I found two-, three-, and four-move mate diagrams and posed them as problems. Only after this did we begin playing real chess. The time we spent getting there lasted about 3-4 months. We should not begrudge the time for this! In this way we assimilate (very deeply and solidly) not only rudimentary knowledge, but the children become accustomed to the carefully considered and foundational game, work. Possessing solid knowledge, they simply and easily learned the later tasks. The possessed resolution, and self-confidence, and arrived at success. They experienced the knowledge and enjoyment of its use.

On grades and competition:

If the instruction is good, one has no need of giving grades. In addition, this truly makes no sense in chess. I would rather arrange various in- and inter-class contests. It is worth sending children to foreign competitions only if we feel that they will do well there. Competition only makes sense when it is evident that it

> will develop those who are capable of it, and can inspire greater accomplishments on the basis of the results. We should never drive students to failure.

And on the end goal:

> An important function of genius education is instilling the capability for self-education. It starts with establishing in the child independent interests. Little by little we can instill in them self-education, independence, and creative work. The pedagogical co-worker cannot always stay at their side. So one of the most important educational tasks is to teach self-education. The latter contributes to, among other things, the child liking what they do, and in their life work is not separate from hobbies.

And unrelated to child-rearing, but very related to a previous discussion on this blog:

> The fact is that today a newborn baby, being Jewish, has a much greater chance, by the statistics of Nobel prizes, at this prize, than if they are born in a non-Jewish family. This seems to many to be genetically determined. I have a completely different opinion. I conclude that social “heredity” and the response to one’s own Jewishness causes this phenomenon. I accept – this is indeed a fact – what Endre Czeizel also mentioned on Hungarian Radio (1989-05-23), that the proportion of Jews among Nobel prize-winners is 30%…, and if one is born Jewish, they have a hundred times greater chance of a Nobel prize than an average non-Jew. And most of the Hungarian Nobel prize-winners were Jews. Among chess world champions their proportion reaches – to my knowledge – more than 50%: Lasker, Steinitz, Tal, Botvinnik, Smyslov, Kasparov and Fischer are half

Jewish. However, I claim that this is also socially determined.

To mention a few factors: The first essential point is that Jewish families – partly because of strong traditions – are relatively stable, and they are always very concerned with education. Another reason comes from the minority status of Jews and from the frequent persecutions throughout their history. How do these factors contribute to the development of the mind? From a negative side in this way, that because of the always disadvantageous situation of the Jews they always had to appear in almost everything doubly more capable than others. Because of the frequent persecution they knew that at any time they might have to leave their homes, dwellings, and even homelands, and begin lives elsewhere. So what is fixed in Jewish tradition? “Learn, my son, because (1) only thus can you succeed in life, and (2) if you must flee, no one can take knowledge away from you, so you can take it with you anywhere.” Jews could not take their houses with them, so they customarily preferred to buy no houses or unportable things, but gold
and diamonds and trinkets, so that during persecutions they could pocket them and run away. And their knowledge bore fruit everywhere.

On the other hand, Jews are always on the periphery, and this awakens stresses in them; they become “eternal adolescents.” Adolescents do not know whether they are children or adults, and their uncertainty comes from this. Similarly, Jews most often do not know to what degree they are – for example – Hungarian, or Jewish, or both. This situation is difficult to clarify to themselves. Because of it they constantly live with internal conflict. This makes them develop with open

> minds, a habit of problem-solving, and also develops their adaptability. (This can also cause certain negative qualities, for example over-sensitivity, loudness, aggression, extremism, being critical of oneself and others, a very strong ambition for accomplishments, over-driven activity, etc.)

II.

I want to emphasize a second time that I've left out most of the book, which is Polgar's philosophical and moral reflections on why genius is important / why it's ethical to try to raise children to be geniuses. If you've got any concerns in those areas, please look a little deeper into the source text and you'll find them answered. At length. Twenty times. Interspersed with enthusiastic infodumps about how great Esperanto is.

In contrast, I said above that there wasn't much specific pedagogical advice. I wrote that before quote-mining it to write a review; after doing so I realize there's more than I thought. It's just very broad, and not too different from what you'd expect a smart and up-to-date educator to say at your local slightly-hippy-ish private school.

I think the main reason I keep feeling like Polgar's not describing his system enough – even though he describes it at some length – is a mismatch between his astounding results, and his excellent-but-not-that-different-from-common-sense educational advice. Surely there are schools that try to make children love learning and feel a sense of accomplishment in their work (don't be snarky here, I'm as depressed as anyone by the education system but there are so many different private schools with wacky philosophies that I'm sure at least one of them has hit the target). But none of them have all their students grow up to be world-class chess grandmasters or the interdisciplinary equivalent. Why not?

Appealing to genetics can only take us so far. Both Polgar parents are undoubtedly geniuses. But a lot of wacky private schools get a steady supply of students with really smart parents. There's got to be something more.

My guess is that the "start really early" and "concentrate on one subject" parts do more of the heavy lifting than I'd previously thought. I also think the one-to-one instruction (well, two-to-three instruction) that the Polgar parents were able to give their kids was probably very important, based on the disproportionate number of child prodigies who were home-schooled by their parents (I don't know if the low teacher:student ratio or the parent-child relationship itself is important). And I suspect Polgar himself was just a naturally gifted educator who was able to effortlessly instill passion for a subject. Get those four things right – early start, single-subject focus, 1:1 home schooling, and a great parent/teacher – and the rest is just common-sense advice. Common-sense advice that lots of educators fail miserably at, admittedly. But common sense advice nonetheless.

This should be encouraging for people who want to repeat the Polgars' experiment. You probably don't need an education degree, let alone training in some secretive arcane Polgar Education Technique, to make it happen. You just need a monomaniacal focus, a lot of free time, and hard-to-define talent.

I think I have a lead on how to get this last one. Polgar talks about how he devised his system by reading the biographies of former child geniuses like John Stuart Mill. It might be useful to repeat this project, if only to see whether someone else can absorb some of the same implicit lessons Polgar did, and gain the same breadth of knowledge he had. This would be my next step from here if I wanted to try to learn more about education.

One concluding quote from Polgar:

> There is no magic even in chess instruction, so I want to "warn" those who are expecting to discover miracles. The main pedagogical method and explanations of basic psychological ideas can be found naturally in pedagogical, psychological and technical chess textbooks.

There's a word for someone who successfully performs miracles, writes a book called *Perform Miracles!*, and then "warns" those who are "expecting to discover miracles". That word is "tease". But Polgar gives us enough of a sketch to at least start out on the road he went on, and hopefully enough further leads to go the rest of the way.

Gender Imbalances Are Mostly Not Due To Offensive Attitudes

[This post has mostly been replaced/supplanted with a more thorough analysis in Contra Grant On Exaggerated Differences]

I.

One idea that kept coming up in the comments to the thread on signal-boosting-as-doxxing: it's permissible to take emergency action against offensive male libertarians, because we really need to improve the gender balance in libertarianism.

The implied assumption is that women flock to movements full of respectful feminists, and shun the ones without them. If women aren't in libertarianism, that's because it's tolerating too many sexists.

Since the assumption is implicit, nobody has defended it, and maybe nobody really believes it in its explicit form. But it underlies enough of this discourse that it's still worth challenging.

Donald Trump is not a poster child for respectful inclusiveness. He is on record saying that he likes to "grab women by the pussy" in what sure sounds like a nonconsenual manner. His sophistication on gender issues is generally somewhere around the level of australopithecine, distantly aspiring towards Neanderthaldom as some shining mountaintop goal.

But Trump voters were more gender-balanced than libertarians: about 47% female, 53% male. Among Trump's key demographic of white people, he actually won the female vote, beating Clinton among white women by 53% to 43%.

The Catholic Church also isn’t a poster child for feminism. Catholic.com, which has such an authoritative-sounding domain name that I assume it’s written personally by the Pope, says that:

> Women are more natural caregivers for children, and men more naturally work outside the home. Yet women can and do work outside the home and men do act as caregivers for children (changing diapers, feeding babies their bottles, burping them, walking with them when they are crying at night-men do all these things, just as women do). Their roles tend to be focused in one area (caregiving for women and working outside the home for men), but one can fill in for the other whenever needed.

Add in their consistent opposition to abortion, birth control, sexual liberation, et cetera, and it should be at least a little surprising that women outnumber men among churchgoing Catholics by about 20%.

Reddit as a whole is about 30% female. But the r/libertarian and r/neoliberal subreddits are 5.5% and 4.5% female respectively. Is this because they’re hostile to women? Seems unlikely. The r/mensrights and r/KotakuInAction (ie Gamergate) subreddit are frequently considered especially offensive, and both of them hover around 10% female. Believe that offensiveness is the sole determinant of gender balance, and you’re forced to conclude that adopting Gamergate’s gender-related norms would double libertarianism’s female population.

See enough of this stuff, and you come to the conclusion that the percent of women in a movement isn't just a function of how carefully it purges all of its insufficiently pro-feminist members.

II.

So what is it a function of?

Richard Lippa's Gender Differences In Personality And Interests is a pretty good source for this sort of thing. It notes that one of the largest gender differences recorded – larger even than the things we tend to think of as hard-and-fast obvious gender differences like physical aggressiveness or attitudes toward casual sex – is what Lippa calls "interest in things vs. people". He writes:

> For the people–things dimension of interests, the results in Table 1 are clear, strong, and unambiguous. Men tend to be much more thing-oriented and much less people-oriented than women (mean d = 1.18, a 'very large' difference, according to Hyde (2005) verbal designations

He notes that a d of 1.18 is "very large", but I worry that the less statistically savvy won't appreciate quite how large it is. All I can say is that I spent several years believing that the d statistic was a scale from 0 to 1, because I'd never seen a d go outside that range before. Daniel Lakens wrote a great piece about a study that found a d = 1.96, where he argued we should dismiss it almost out of hand, because non-tautological effects are almost never that large and so clearly somebody made a mistake in the study (spoiler: they did). Lippa's finding isn't quite at that level, but it's getting up there.

This strikes me as sort of similar to the systematizing-empathizing distinction, where men and women also score pretty differently (d = 0.5). I admit I don't have great proof that these are related concepts, but it seems intuitive to me that systems are a sort of "thing", and that the kind of people who are interested in analyzing formal systems all the time usually aren't super people-oriented.

(am I using "kind of people" here as a dog-whistle for autism? Give me more credit than that; I'm actually trying to dog-whistle "NMDA receptor hypofunction.")

There seems to be a similar gender difference in tendency to use utilitarianism – see eg Friesdorf et al's meta-analysis. Without me being able to explain exactly what I mean, I hope you share my intuition that utilitiarianism is an unusually thing-oriented and systematizing moral system, and that this gestures at the same large but hard-to-explain difference as the other two results.

These are probably upstream of more noticeable differences in political attitudes. A less-empathizing/more-systematizing personality would make people more interested in the politics of effective institutions than in the politics of provisioning social welfare. And a less utilitiarian and more traditional-morality-focused personality would make people more interested in using government power to enforce social norms. Eagly et al look at gender differences in political attitudes and find exactly this:

> The most commonly noted difference is that women are more likely than men to endorse policies that support the provision of social services for deserving and disadvantaged groups (Goertzel,
> 1983; Schlesinger & Heldman, 2001; Shapiro & Mahajan, 1986), including housing, child care, educational

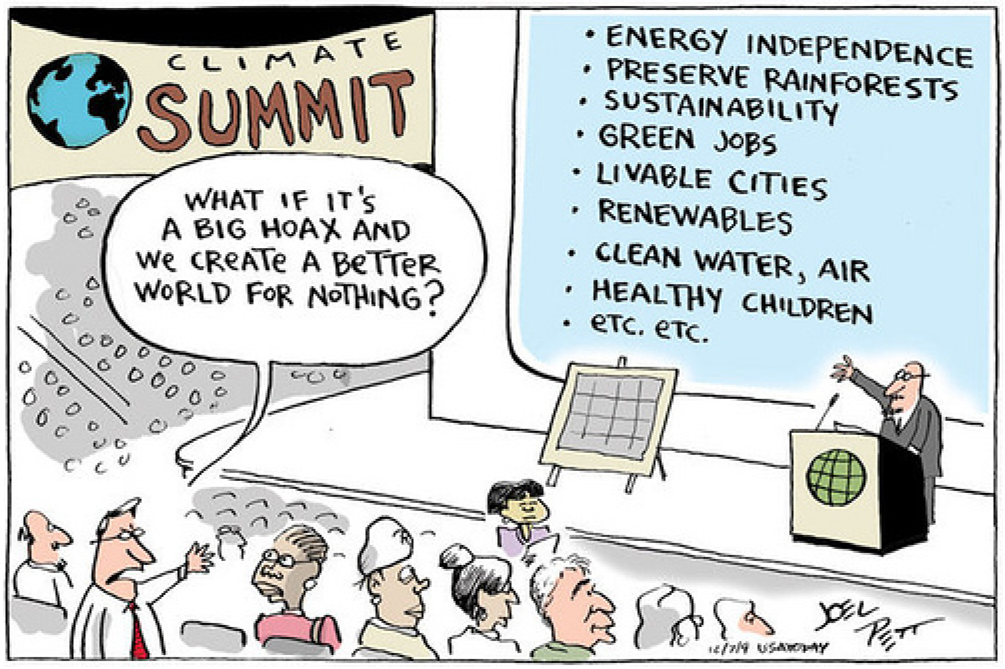
> opportunity, and financial support in the form of welfare…Women also advocate more restriction of many behaviors that are traditionally considered immoral (e.g., casual sex; Oliver & Hyde, 1993; consumption of pornography, Seltzer et al., 1997)…Women can thus be regarded as more liberal than men in social compassion and more conservative in traditional morality.

“More liberal in social compassion and more conservative in traditional morality” sounds like a pretty good description of “not libertarian”.

I’m not sure there’s a lot of mystery left to explain here. You can eliminate every single shred of sexism in the libertarian movement, make it so feminist that it makes Ursula K. LeGuin books look like Margaret Atwood books – and it will still never get anywhere near the gender balance of those weird evangelical sects who talk about women have to be subservient because Eve was made from Adam’s rib.

III.

A strong counterargument:


CLIMATE SUMMIT
WHAT IF IT'S A BIG HOAX AND WE CREATE A BETTER WORLD FOR NOTHING?
• ENERGY INDEPENDENCE
• PRESERVE RAINFORESTS
• SUSTAINABILITY
• GREEN JOBS
• LIVABLE CITIES
• RENEWABLES
• CLEAN WATER, AIR
• HEALTHY CHILDREN
• etc. etc.
JOEL PETT
12/7/09 USA TODAY

What if all this stuff about sexism driving away women is all a big hoax? And so after we make women feel safer, stamp out prejudice, enforce common decency, and encourage everyone to treat each other with compassion – darn it, we created a better world for nothing! If the goal is “eliminate malignant sexism” – and surely it should be – why be so upset about one argument for eliminating malignant sexism which might not be entirely accurate?

First, because I’m a heartless thing-oriented systematizer, and I despise bad arguments on principle, and I don’t care if you people-oriented empathizers think they serve a prosocial community-building function.

But second, because this gives fuzzy-empathizing-humanities types a giant hammer with which to beat all sciency-systematizing-utilitarian types forever.

I’m not the sort of New Atheist who believes in some kind of apocalyptic battle where the defenders of rationality and civilization face off against some fantastic coalition of postmodernism / hippyism / gender studies / crystal-healing / evangelicalism / cultural Marxism / Islam / Donald Trump. But STEM ideals conflicting with humanities-focused ideals isn't a total fiction. There really are some differences in values between the average Silicon Valley programmer and the average Oberlin literature professor. A humanities/empathizing/intuitive vs. sciency/systematizing/utilitarian distinction isn't a perfectly natural category, nor the only axis on which people differ. But I'm hoping you share my intuition that it's at least a vague cluster, and at least one axis of difference.

And however different postmodernists, evangelicals, Islamists, Muslims, crystal-healers, and Trump supporters might be, there actually *is* one thing they have in common: all these groups have great gender balance. You'll never find a Wiccan circle or a gender studies class that accidentally ended up as 100% male.

And computer scientists, mathematicians, economists, utilitarians, libertarians, movement atheists, skeptics, transhumanists, cryptocurrency enthusiasts, et cetera – are an equally sundry non-coalition. But they also have something in common: a serious skew towards men.

And if you accept the implict assumption that good opinions = gender balance and sexism = gender imbalance, then forever and always the crystal healers and Trump supporters will have a clear badge of being good people and responsible citizens, and the utilitarians and economists will be, on a collective level, sexist jerks. And it sure seems like this is a point in favor of the crystal healers and Trump supporters if you're trying to figure out who to trust.

A community made up of sexist jerks has a moral obligation to stop being sexist and jerkish right away, both because it's the right thing to do, and because it's tactically advantageous to be able to recruit women to the cause. If sexist jerkishness can be measured by gender balance, the appropriate response is "keep dialing up the level of cracking-down-on-sexism until gender balance approaches parity." But if this is your philosophy, and gender balance doesn't respond at all to these crackdowns, then level-of-cracking-down quickly rises to infinity.

So we end up with sciency/systematizing/utilitarian fields and movements turning into circular firing chambers. If "it's permissible to take emergency action against offensive male libertarians, because we really need to improve the gender balance in libertarianism", then the state of emergency quickly becomes permanent, as more and more extreme measures fail to effect any improvement. Yes, there will be people trying to get you fired over making sexist jokes on Twitter. But don't worry – there will also be people trying to get you fired over if you're too interested in BDSM, or if you find the word dongle amusing. And if the people trying to get the other people fired overstep their social status just a tiny bit, then the backlash will get *them* fired ("circular firing chamber" wasn't intended as a pun, but maybe it should have been).

At the same time, everyone on the humanities/empathizing/intuitive side of the divide gets a knockdown argument against any sciency/systematizing/utilitarian challenge: "Oh, those sexists? Come back when you've taken off your fedora, dudebros." And this sort of thing *works*. And since so many of us implicitly accept the poor-gender-balance = sexist jerks narrative, we can't even fight back. The best we can do is "Yes, our side is full of repellent people with terrible values, and your side is full of lovely tolerant accepting people who successfully build inclusive communities for everyone, but, uh, we make some good points anyway. Wait, come back! Why aren't you reading my double-blind randomized study showing that I'm right?"

I think a better response is to point out that there's never been any study or informal survey suggesting that sciency/systematizing/utilitarian people are any more sexist than anyone else (if you find one, tell me). No person or community is perfect, but to claim some special evil for libertarians, economists, utilitarians etc is a weird hypothesis I've never seen anyone try to explain, let alone prove. When 20% of high school kids taking the AP Computer Science test are women, and 20% of college students majoring in CS are women, and then 20% of people in the tech industry are women, maybe "demographics are a thing" is a better hypothesis than "the tech industry is uniquely full of gross sexist nerds".

This isn't to deny the experiences of women who feel more frequently harassed in these communities. But it does seem s more likely to be a result of gender balance than a cause of it. That is, if we dectuple the male-to-female ratio while holding the sexual-harasser-ness of each man constant, then each woman faces ten times more sexual harassment.

This might mean that there's a duty for sciency/systematizing/utilitarian people to work ten times harder to fight harassment than anyone else. I think that might be happening. When I first entered medicine, I was shocked at the lackadaisacal attitude people there took to casually sexist speech, compared to an environment in Silicon Valley which I've heard people describe as "Stalinist". But medicine is a people-oriented/empathizing profession; 90% of nurses are women; so are 57% of psychiatry residents. The potential-harasser:potential-victim ratio keeps people safer without the same permanent state of emergency.

So yes, let’s try to build a better world. But let’s do it because it’s a good thing to do, not because we expect it to single-handedly normalize gender ratios.

Is It Possible To Have Coherent Principles Around Free Speech Norms?

I.

One factor that must underlie people's distrust of non-governmental free speech norms is that they're so underspecified. The First Amendment is a comparatively simple, bright-line concept – the police can't arrest you for saying the President sucks. Sure, we need a zillion Supreme Court cases to iron out the details, but it makes sense *in principle*. By contrast, social norms about free speech risk collapsing into the incoherent Doctrine Of The Preferred First Speaker, where it's okay for me to say that the President sucks, but *not* okay for you to say that I suck for saying that. This is dumb, and I don't know if free speech supporters have articulated a meaningful alternative. I want to sketch out some possibilities for what that sort of alternative would look like.

The philosophical question here is separating out the acts of having an opinion, signaling a propensity, and committing a speech act.

Having an opinion is the sort of thing free speech norms ought to exist to protect. The opinion ought to enter the marketplace of ideas, compete with other opinions on its merit, and either win or lose based on people's considered rational judgment.

But this can't be separated from *signaling a propensity* for action. Suppose Alice has the opinion "hand hygiene doesn't matter". The truly virtuous action is to show her (and concerned third parties) studies that prove that dangerous infections are transmissible by unwashed hands. But while you're doing that, it's fair to not want to eat at her restaurant. And it's pro-social to tell other people not to eat at her restaurant either, and not to hire her as a nurse – and if she's already a nurse, maybe to get her fired. Even though reasonable free speech norms demand that we fight bad ideas through counterargument rather than social punishment, in this case they should permit a campaign to get Alice fired.

One solution here might be to give people the burden of demonstrating that their controversial opinions won't lead to dangerous actions. For example, if Alice is a nurse, she might say "I don't believe hand hygiene matters, and I'm going to try to convince the hospital administration to remove their rule mandating handwashing – but until I succeed, I'll follow the rules and wash my hands just like everyone else." If I trusted Alice, this would allay my concerns, and I would go back to wanting to debate with her instead of wanting her fired. See also Be Nice, At Least Until You Can Coordinate Meanness.

Some signaling of propensities can't be so easily fixed. If Carol thinks that "Hitler should have finished the job", I feel like this tells me a lot about Carol besides just her moral ranking of various World-War-II-related alternate histories. If she was a schoolteacher, then even if she promised not to kill any Jews in class, or even to spread any anti-Semitic propaganda in class, it would be hard for me not to wonder what else was wrong with her, and whether she could really disguise every single repugnant aspect of her personality. On the other hand, if we try to get the school board to fire her, we're implicitly endorsing the principle "Get someone fired if you know of a belief of theirs that suggests they're an otherwise repugnant person" – and isn't this the same principle that led people to campaign against atheist schoolteachers, pro-gay schoolteachers, communist schoolteachers, etc? See also Not Just A Mere Political Issue. I think I bite the bullet on this one and say that if the schoolteacher credibly promises not to be repugnant in any way in front of the kids, you let her keep teaching until she slips up.

And both having opinions and signaling propensities are hard to separate from *commiting speech acts*. The archetypal example is telling an assassin “I’ll give you $10,000 if you kill Bob” – a form of speech, but tantamount to murder. Repeated harassment – the kind where you scream insults at someone every time they leave the house – falls in the same category: the active ingredient isn’t the information being conveyed by what insults you choose, it’s that they face being screamed at and made to live in fear. And yeah, the archetypical example of this is starting a campaign to email someone’s embarassing secrets to their boss to get them fired.

We can’t just ban speech acts. Everything is a speech act. Me saying “Donald Trump is wrong on immigration” lowers Donald Trump’s status – that’s a speech act. Me saying “You’re wrong about free speech” might trigger you and make you feel awful until you write a 10,000 word essay responding to me – that’s a speech act too. Telling an offensive joke is definitely a speech act, but do we want to ban all jokes that anyone anywhere might be offended by? Let’s face it; a lot of speech is criticism, sometimes really harsh criticism, and the line between “criticism”, “insult”, and “harassment” is vague and debatable (see eg all of Twitter). Everyone has a different set of speech acts they consider beyond the pale, with no real way of sorting it out. So what speech acts do we permit as unavoidable parts of the process of social interaction, which ones do we punish, and how do we punish them?

II.

A sample problem: a while ago, I read an article which took a sensitive social problem, approached it with inexcusably terrible statistics that mis-estimated its incidence by seven orders of magnitude, and then used it to mock the people who suffered from it and tell them they were stupid. I complained about this, and the author was really confrontational to me and said things like that I “needed to see a psychiatrist”. I ended up writing a couple of really angry blog posts, which not only corrected the statistics but also prominently named the author, accused him of being a bad person, and recommended that nobody ever trust him or take him seriously again.

One view: although the author was wrong, we’re all wrong sometimes. I’ve been wrong before, probably in ways that other people considered inexcusable, and I would rather be politely corrected than excoriated in public, dragged through the mud, and accused of being a defective human being. By my article, I contributed to a world where we don’t just debate each other’s points, but launch personal attacks against people in the hopes that they are so ashamed and miserable that they never participate in the discussion again. I have committed crimes against Reason, and I should humbly apologize and try to do better next time.

A second view: the author was either deliberately deceitful or criminally stupid; either way he really was *inexcusably* bad. If I just quietly correct his statistical error, only a fraction of his readership will see my correction, and meanwhile he’ll go on to do it a second time, a third time, and so on forever. Although there are many good people who should be approached as equals in the marketplace of ideas, there are also defectors against that marketplace who deserve to be ruthlessly crushed, and I was doing a public service by crushing one of them.

And a third view: by being needlessly cruel in his article, the author had already forfeited the protection of “the marketplace of ideas”. Just as if someone tries to shoot you, you can shoot back without worrying so much about the moral principle of nonviolence, so it’s always proper to fight fire with fire. Although I wouldn’t be justified in smacking down someone who had merely failed egregiously, someone who fails egregiously while breaking good discussion norms is another matter.

The second and third views get kind of scary when universalized. The second amounts to “if you decide someone’s a really bad person, feel free to crush them.” No doubt some evangelicals honestly think that gay rights crusaders are bad people; does this justify personal attacks against them?

The third seems to demand a more specific trigger (violation of a norm), but since nobody agrees where the norms are, it’s more likely to just lead to cascades where everyone ends up at different levels of the punishing/meta-punishing/meta-meta-punishing ladder and everyone thinks everyone else started it.

(an example: Alice writes a blog post excoriating Bob’s opinion on tax reforming, calling him a “total idiot” who “should be laughed out of the room”. Bob feels so offended that he tries to turn everyone against Alice, pointing out every bad thing she’s ever done to anyone who will listen. Carol considers this a “sexist harassment campaign” and sends a dossier of all of Bob’s messages to his boss, trying to get him fired. Dan decides this proves Carol is anti-free speech, and tells the listeners of his radio show to "give Carol a piece of their mind", leading to her getting hundreds of harassing and threatening email messages. Eric snitches on Dan to the police. How many of these people are in the wrong?)

But I can't fully bite the bullet and accept the first view either; some people are so odious that an alarm needs to be spread. I'm not proud of my behavior in the specific situation mentioned, but I won't completely give up the right to do something similar if the information arises. I'm going to try as hard as I can to err on the side of not doing that (I stick by my decision not to name the *Reason* columnist involved in the sandwich incident, although I guess everyone already knows) but sometimes the line will need to be crossed.

III.

I think the most important consideration is that it be crossed in a way that doesn't create a giant negative-sum war-of-all-against-all. That is, Democrats try to get Republicans fired for the crime of supporting Republicanism, Republicans try to get Democrats fired for the crime of supporting Democratism, and the end result is a lot of people getting fired but the overall Republican/Democrat balance staying unchanged.

That suggests a heuristic very much like Be Nice, At Least Until You Can Coordinate Meanness again: don't try to destroy people in order to enforce social norms that only exist in your head. If people violate a real social norm, that the majority of the community agrees upon, and that they should have known – that's one thing. If you idiosyncratically believe something is wrong, or you're part of a subculure that believes something is wrong even though there are opposite subcultures that don't agree – then trying to enforce your idiosyncratic rules by punishing anyone who breaks them is a bad idea.

And one corollary of this is that it shouldn't be arbitrary. Ten million people tell sexist jokes every day. If you pick one of them, apply maximum punishment to him, and let the other 9.99 million off scot-free, he's going to think it's unfair – and he'll be right. This is directly linked to the fact that there isn't *actually* that much of a social norm against telling sexist jokes. My guess is that almost everyone who posts child pornography on Twitter gets in trouble for it, and that's because there really *is* a strong anti-child pornography norm.

(this is also how I feel about the war on drugs. One in a thousand marijuana users gets arrested, partly because there isn't enough political will to punish all marijuana users, partly because nobody really thinks marijuana use is that wrong. But this ends out unfair to the arrested marijuana user, not just because he's in jail for the same thing a thousand other people did without consequence, but because he probably wouldn't have done it he'd really expected to be punished, and society was giving him every reason to think he wouldn't be.)

This set of norms is self-correcting: if someone does something you don't like, but there's not a social norm against it, then your next step should be to create a social norm against it. If you can convince 51% of the community that it's wrong, then the community can unite against it and you can punish it next time. If you can't convince 51% of the community that it's wrong, then you should try harder, not play vigilante and try to enforce your unpopular rules yourself.

If you absolutely can't tolerate something, but you also can't manage to convince your community that it's wrong and should be punished, you should work on finding methods that isolate you from the problem, including building a better community somewhere else. I think some of this collapses into a kind of Archipelago solution. Whatever the global norms may be, there ought to be communities catering to people who want more restrictions than normal, and other communities catering to people who want fewer. These communities should have really explicit rules, so that everybody knows what they're getting into. People should be free to self-select into and out of those communities, and those self-selections should be honored. Safe spaces, 4chan, and this blog are three very different kinds of intentional communities with unusual but locally-well-defined free speech norms, they're all good for the people who use them, and as long as they keep to themselves I don't think outsiders have any right to criticize their existence.

IV.

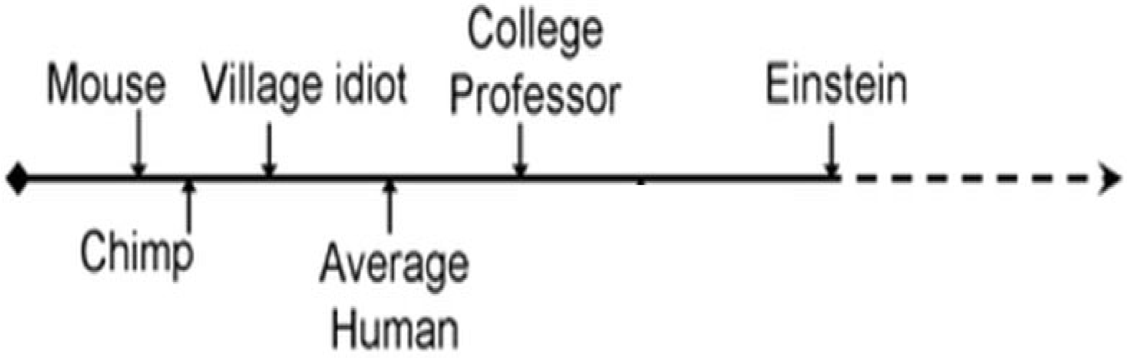
I don't know if this position is coherent. My guess is there's a lot of places it doesn't match my intuition, and a lot of other places where it's so fuzzy it could justify or condemn anything at all.

But I think trying to hammer out something like this is important. Free speech norms aren't about free speech. They quickly bleed over into these really fundamental questions, like – what is a culture? What is it we're trying to do when we get together and have a society? Are we allowed to want different things from a culture? If so, how do we balance everyone else's demands? Do we just live in some kind of postmodern globalized atomized culture, or are cultures these things inexorably linked to specific value systems that we’ve got to keep moored to those systems at all costs? How much are we allowed to use shaming to punish people who don’t conform to our culture? How angry are we allowed to be when other people use shaming to punish people we like who don’t conform to theirs?

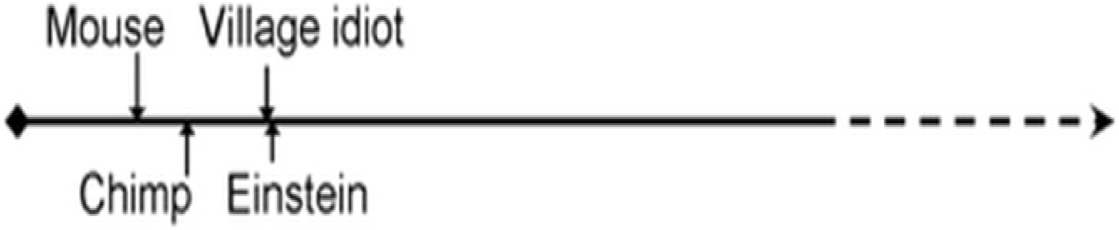
Trying to get a model of these things that doesn’t immediately contradict itself on everything is potentially a good first step to trying to get a model of these things that’s right and/or liveable.

Where The Falling Einstein Meets The Rising Mouse

Eliezer Yudkowsky argues that forecasters err in expecting artificial intelligence progress to look like this:


College Professor
Mouse
Village idiot
Einstein
Chimp
Average Human

…when in fact it will probably look like this:

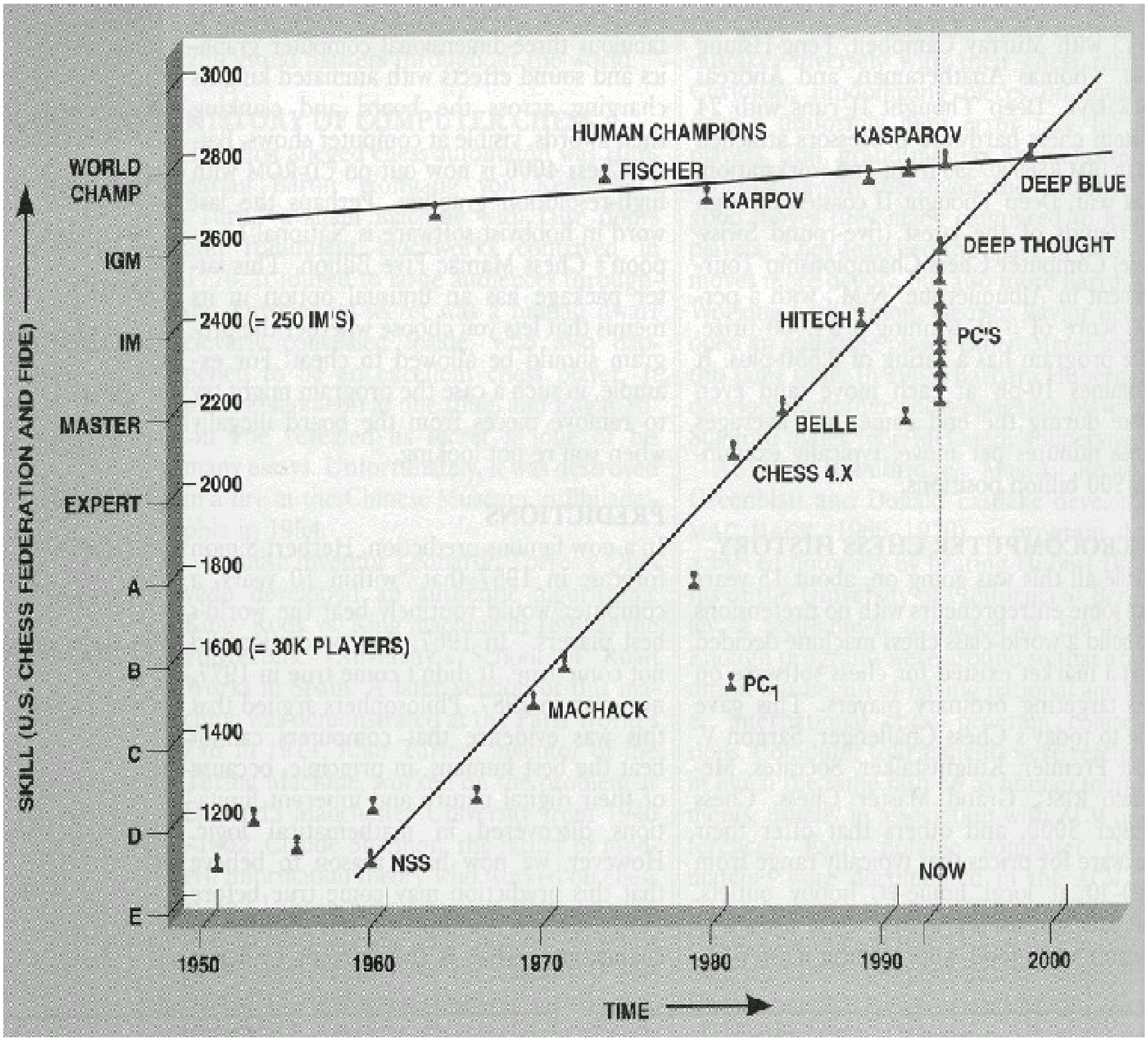

Mouse
Village idiot
Chimp
Einstein

That is, we naturally think there's a pretty big intellectual difference between mice and chimps, and a pretty big intellectual difference between normal people and Einstein, and implicitly treat these as about equal in degree. But in any objective terms we choose – amount of evolutionary work it took to generate the difference, number of neurons, measurable difference in brain structure, performance on various tasks, etc – the gap between mice and chimps is immense, and the difference between an average Joe and Einstein trivial in comparison. So we should be wary of timelines where AI reaches mouse level in 2020, chimp level in 2030, Joe-level in 2040, and Einstein level in 2050. If AI reaches the mouse level in 2020 and chimp level in 2030, for all we know it could reach Joe level on January 1st, 2040 and Einstein level on January 2nd of the same year. This would be pretty disorienting and (if the AI is poorly aligned) dangerous.

I found this argument really convincing when I first heard it, and I thought the data backed it up. For example, in my Superintelligence FAQ, I wrote:

> In 1997, the best computer Go program in the world, Handtalk, won NT$250,000 for performing a previously impossible feat – beating an 11 year old child (with an 11-stone handicap penalizing the child and favoring the computer!) As late as September 2015, no computer had ever beaten any professional Go player in a fair game. Then in March 2016, a Go program beat 18-time world champion Lee Sedol 4-1 in a five game match. Go programs had gone from “dumber than heavily-handicapped children” to “smarter than any human in the world” in twenty years, and “from never won a professional game” to “overwhelming world champion” in six months.

But Katja Grace takes a broader perspective and finds the opposite. For example, she finds that chess programs improved gradually from “beating the worst human players” to “beating the best human players” over fifty years or so, ie the entire amount of time computers have existed:



AlphaGo represented a pretty big leap in Go ability, but before that, Go engines improved pretty gradually too (see the original AI Impacts post for discussion of the Go ranking system on the vertical axis):

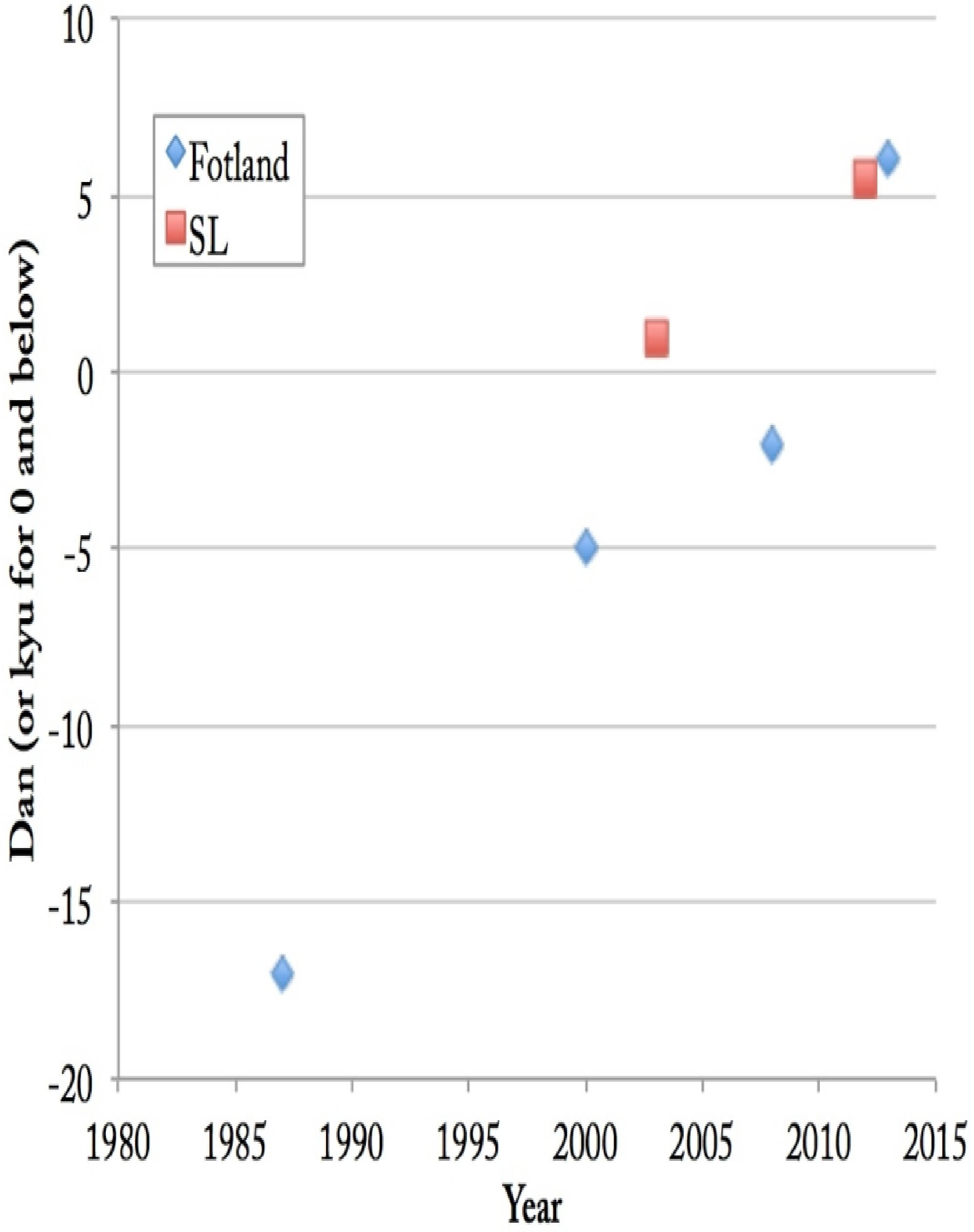


Figure 25: Rough estimates of historical Go bot abilities. Data from Sensei's Library (2013) and Fotland (2002).

There's a lot more on Katja's page, overall very convincing. In field after field, computers have taken decades to go from the mediocre-human level to the genius-human level. So how can one reconcile the common-sense force of Eliezer’s argument with the empirical force of Katja’s contrary data?

Theory 1: Mutational Load

Katja has her own theory:

> The brains of humans are nearly identical, by comparison to the brains of other animals or to other possible brains that could exist. This might suggest that the engineering effort required to move across the human range of intelligences is quite small, compared to the engineering effort required to move from very sub-human to human-level intelligence…However, we should not be surprised to find meaningful variation in the cognitive performance regardless of the difficulty of improving the human brain. This makes it difficult to infer much from the observed variations.
>
> Why should we not be surprised? De novo deleterious mutations are introduced into the genome with each generation, and the prevalence of such mutations is determined by the balance of mutation rates and negative selection. If de novo mutations significantly impact cognitive performance, then there must necessarily be significant selection for higher intelligence–and hence behaviorally relevant differences in intelligence. This balance is determined entirely by the mutation rate, the strength of selection for intelligence, and the negative impact of the average mutation.
>
> You can often make a machine worse by breaking a random piece, but this does not mean that the machine was easy to design or that you can make the machine better by adding a random piece. Similarly, levels of variation of cognitive performance in humans may tell us

> very little about the difficulty of making a human-level intelligence smarter.

I’m usually a fan of using mutational load to explain stuff. But here I worry there’s too much left unexplained. Sure, the explanation for variation in human intelligence is whatever it is. And there’s however much mutational load there is. But that doesn’t address the fundamental disparity: isn’t the difference between a mouse and Joe Average still immeasurably greater than the difference between Joe Average and Einstein?

Theory 2: Purpose-Built Hardware

Mice can’t play chess (citation needed). So talking about “playing chess at the mouse level” might require more philosophical groundwork than we’ve been giving it so far.

Might the worst human chess players play chess pretty close to as badly as is even possible? I’ve certainly seen people who don’t even seem to be looking one move ahead very well, which is sort of like an upper bound for chess badness. Even though the human brain is the most complex object in the known universe, noble in reason, infinite in faculties, like an angel in apprehension, etc, etc, it seems like maybe not 100% of that capacity is being used in a guy who gets fools-mated on his second move.

We can compare to human prowess at mental arithmetic. We know that, below the hood, the brain is solving really complicated differential equations in milliseconds every time it catches a ball. *Above* the hood, most people can’t multiply two two-digit numbers in their head. Likewise, in principle the brain has 2.5 petabytes worth of memory storage; in practice I can’t always remember my sixteen-digit credit card number.

Imagine a kid who has an amazing $5000 gaming computer, but her parents have locked it so she can only play Minecraft. She needs a calculator for her homework, but she can’t access the one on her computer, so she builds one out of Minecraft blocks. The gaming computer can have as many gigahertz as you want; she’s still only going to be able to do calculations at a couple of measly operations per second. Maybe our brains are so purpose-built for swinging through trees or whatever that it takes an equivalent amount of emulation to get them to play chess competently.

In that case, mice just wouldn’t have the emulated more-general-purpose computer. People who are bad at chess would be able to emulate a chess-playing computer very laboriously and inefficiently. And people who are good at chess would be able to bring some significant fraction of their full most-complex-object-in-the-known-universe powers to bear. There are some anecdotal reports from chessmasters that suggest something like this – descriptions of just “seeing” patterns on the chessboard as complex objects, in the same way that the dots on a pointillist painting naturally resolve into a tree or a French lady or whatever.

This would also make sense in the context of calculation prodigies – those kids who can multiply ten digit numbers in their heads really easily. Everybody has to have the capacity to do this. But some people are better at accessing that capacity than others.

But it doesn’t make sense in the context of self-driving cars! If there was ever a task that used our purpose-built, no-emulation-needed native architecture, it would be driving: recognizing objects in a field and coordinating movements to and away from them. But my impression of self-driving car progress is that it’s been stalled for a while at a level better than the worst human drivers, but worse than the best human drivers. It'll have preventable accidents every so often – not as many as a drunk person or an untrained kid would, but more than we would expect of a competent adult. This suggests a really wide range of human ability even in native-architecture-suited tasks.

Theory 3: Widely Varying Sub-Abilities

I think self-driving cars are already much better than humans at certain tasks – estimating differences, having split-second reflexes, not getting lost. But they're also much worse than humans at others – I think adapting to weird conditions, like ice on the road or animals running out onto the street. So maybe it's not that computers spend much time in a general "human-level range", so much as being superhuman on some tasks, and subhuman on other tasks, and generally averaging out to somewhere inside natural human variation.

In the same way, long after Deep Blue beat Kasparov there were parts of chess that humans could do better than computers, "anti-computer" strategies that humans could play to maximize their advantage, and human + computer "cyborg" teams that could do better than either kind of player alone.

This sort of thing is no doubt true. But I still find it surprising that the average of "way superhuman on some things" and "way subhuman on other things" averages within the range of human variability so often. This seems as surprising as ever.

Theory 1.1: Humans Are Light-Years Beyond Every Other Animal, So Even A Tiny Range Of Human Variation Is Relatively Large

Or maybe the first graph representing the naive perspective is right, Eliezer's graph representing a more reasonable perspective is wrong, and the range of human variability is immense. Maybe the difference between Einstein and Joe Average is the same (or bigger than!) the difference between Joe Average and a mouse.

That is, imagine a Zoological IQ in which mice score 10, chimps score 20, and Einstein scores 200. Now we can apply Katja's insight: that humans can have very wide variation in their abilities thanks to mutational load. But because Einstein is so far beyond lower animals, there's a wide range for humans to be worse than Einstein in which they're still better than chimps. Maybe Joe Average scores 100, and the village idiot scores 50. This preserves our intuition that even the village idiot is vastly smarter than a chimp, let alone a mouse. But it also means that most of computational progress will occur within the human range. If it takes you five years from starting your project, to being as smart as a chimp, then even granting linear progress it could still take you fifty more before you're smarter than Einstein.

This seems to explain all the data very well. It's just shocking that humans are so far beyond any other animal, and their internal variation so important.

Maybe the closest real thing we have to zoological IQ is encephalization quotient, a measure that relates brain size to body size in various complicated ways that sometimes predict how smart the animal is. We find that mice have an EQ of 0.5, chimps of 2.5, and humans of 7.5.

I don't know whether to think about this in relative terms (chimps are a factor of five smarter than mice, but humans only a factor of three greater than chimps, so the mouse-chimp difference is bigger than the chimp-human difference) or in absolute terms (chimps are 2 units bigger than mice, but humans are five units bigger than chimps, so the chimp-human difference is bigger than the mouse-chimp difference).

Brain size variation within humans is surprisingly large. Just within a sample of 46 adult European-American men, it ranged from 1050 to 1500 cm^3m; there are further differences by race and gender. The difference from the largest to smallest brain is about the same as the difference between the smallest brain and a chimp (275 – 500 cm^3); since chimps weight a bit less than humans, we should probably give them some bonus points. Overall, using brain size as some kind of very weak Fermi calculation proxy measure for intelligence (see here), it looks like maybe the difference between Einstein and the village idiot equals the difference between the idiot and the chimp?

But most mutations that decrease brain function will do so in ways other than decreasing brain size; they will just make brains less efficient per unit mass. So probably looking at variation in brain size underestimates the amount of variation in intelligence. Is it underestimating it enough that the Einstein – Joe difference ends up equivalent to the Joe – mouse difference? I don't know. But so far I don't have anything to say it isn't, except a feeling along the lines of "that can't possibly be true, can it?"

But why not? Look at all the animals in the world, and the majority of the variation in size is within the group "whales". The absolute size difference between a bacterium and an elephant is less than the size difference between *Balaenoptera musculus brevicauda* and *Balaenoptera musculus musculus* – ie the Indian Ocean Blue Whale and the Atlantic Ocean Blue Whale. Once evolution finds a niche where greater size is possible and desirable, and figures out how to make animals scalable, it can go from the horse-like ancestor of whales to actual whales in a couple million years. Maybe what whales are to size, humans are to brainpower.

Stephen Hsu calculates that a certain kind of genetic engineering, carried to its logical conclusion, could create humans “a hundred standard deviations above average” in intelligence, ie IQ 1000 or so. This sounds absurd on the face of it, like a nutritional supplement so good at helping you grow big and strong that you ended up five light years tall, with a grip strength that could crush whole star systems. But if we assume he’s just straightforwardly right, and that Nature did something of about this level to chimps – then there might be enough space for the intra-human variation to be as big as the mouse-chimp-Joe variation.

How does this relate to our original concern – how fast we expect AI to progress?

The good news is that linear progress in AI would take a long time to cross the newly-vast expanse of the human level in domains like “common sense”, “scientific planning”, and “political acumen”, the same way it took a long time to cross it in chess.

The bad news is that if evolution was able to make humans so many orders of magnitude more intelligent in so short a time, then intelligence really is easy to scale up. Once you’ve got a certain level of general intelligence, you can just crank it up arbitrarily far by adding more inputs. Consider by analogy the hydrogen bomb – it’s very hard to invent, but once you’ve invented it you can make a much bigger hydrogen bomb just by adding more hydrogen.

This doesn’t match existing AI progress, where it takes a lot of work to make a better chess engine or self-driving car. Maybe it will match future AI progress, after some critical mass is reached. Or maybe it’s totally on the wrong track.

I’m just having trouble thinking of any other explanation for why the human level could be so big.

Why Not More Excitement About Prediction Aggregation?

There's a new ad on the sidebar for Metaculus, a crowd-sourced prediction engine that tries to get well-calibrated forecasts on mostly scientific topics. If you're wondering how likely it is that a genetically-engineered baby will be born in the next few years, that SpaceX will reach Mars by 2030, or that Sci-Hub will get shut down, take a look.

(there are also the usual things about politics, plus some deliberately wacky ones like whether Elon Musk will build more kilometers of tunnel than Trump does border wall)

They're doing good work – online for-cash prediction markets are limited to a couple of bets by the government, and they usually focus on the big-ticket items like who's going to win elections. Metaculus is run by a team affiliated with the Foundational Questions Institute, and as their name implies they're really interested in using the power of prediction aggregators to address some the important questions for the future of humanity – like AI, genetic engineering, and pandemics.

Which makes me wonder: what's everyone else's excuse?

Back when it looked like prediction markets were going to change everything (was it really as recent as two months ago?), various explanations were floated for why they hadn't caught on yet. The government was regulating betting too much. The public was creeped out by the idea of profiting off of terrorist attacks. Or maybe people were just too dumb to understand the arguments in favor.

Now there are these crowd-sourced aggregator things. They're not regulated by the government. Nobody's profiting off of anything. And you don't have to have faith that they'll work – Philip Tetlock's superforecasting experiments proved it, and Metaculus tracks its accuracy over time. I know the intelligence services are working with the Good Judgment Project, but I'm still surprised it hasn't gone further.

Robin Hanson is the acknowledged expert on this, thanks to his long career of trying to get institutions to adopt prediction markets and generally failing. He attributes the whole problem to signaling and prestige, which I guess makes as much sense as anything. Tyler Cowen says something similar here. But I'm *still* surprised there aren't consultant superforecaster think tanks hanging up their shingles. Forget prestige – aren't there investors who would pay for their wisdom? And why can't I hire Philip Tetlock to tell me whether my relationship is going to last?

I asked Prof. Aguirre of Metaculus, and he said (slightly edited for flow):

> I don't think Tetlock has any "secret sauce", though I think he did a good job. The Metaculus track record is pretty good and will continue to improve. There's definitely real predictive power. Our main challenge is that all the personnel involved are very time-limited and we're also operating on a shoestring, probably 1/50th of what Tetlock spent of IARPA's money.
>
> If you are an individual company or investor, you don't really get that much from "crowdsourcing" because you don't really have a crowd (unless you're a big business and can force your employees to make the predictions); so I'd guess most companies probably just fall back on

> asking some group of people to get together and make projections etc. My personal view is that the power really comes when you get a *lot* of people, questions, and data; you can then start to leverage that to improve the calibration (by recalibrating based on past accuracy), identify the really good predictors, and build up a large enough corpus of results that the predicted probabilities become grounded in an actual ensemble — the ensemble of questions on the platform.
>
> Relatedly, the ability to make use of probabilistic predictions is, sadly, confined to a rather small fraction of people. I think typical decision-makers want to know what *will* happen, and 60-40 or 75-25 or even 80-20 is not the certainty they want. In the face of this uncertainty, I think people mentally substitute a different question to which they feel that they have some sort of good instinct, or privileged understanding. I think there's also an element where they just don't really *believe* the numbers because they are sometimes "wrong." This sometimes frustrates me, but is not wholly irrational, I would say, because in the absence of a real grounding of the probabilities, if you have some analyst come and tell you that there's a 70-30 chance, what exactly do you do with that? How would you possibly trust that those are the right numbers if the question is something "squishy" without a solid mathematical model?

I wonder if there's data about how accuracy changes with number of predictors and predictor quality. There are so many smart people around who are interested in probability and willing to answer a few questions that this seems like a really stupid bottleneck. I'm happy I can do my part pushing **Metaculus**, but someone seriously needs to find a way to make this scale.

Contra Grant On Exaggerated Differences

I.

An article by Adam Grant called Differences Between Men And Women Are Vastly Exaggerated is going viral, thanks in part to a share by Facebook exec Sheryl Sandberg. It's a response to an email by a Google employee saying that he thought Google's low female representation wasn't a result of sexism, but a result of men and women having different interests long before either gender thinks about joining Google. Grant says that gender differences are small and irrelevant to the current issue. I disagree.

Grant writes:

> It's always precarious to make claims about how one half of the population differs from the other half—especially on something as complicated as technical skills and interests. But I think it's a travesty when discussions about data devolve into name-calling and threats. As a social scientist, I prefer to look at the evidence.
>
> The gold standard is a meta-analysis: a study of studies, correcting for biases in particular samples and measures. Here's what meta-analyses tell us about gender differences:
>
> When it comes to abilities, attitudes, and actions, sex differences are few and small.
>
> Across 128 domains of the mind and behavior, "78% of gender differences are small or close to zero." A recent

> addition to that list is leadership, where men feel more confident but women are rated as more competent.
>
> There are only a handful of areas with large sex differences: men are physically stronger and more physically aggressive, masturbate more, and are more positive on casual sex. So you can make a case for having more men than women… if you're fielding a sports team or collecting semen.

The meta-analysis Grant cites is Hyde's, available here. I've looked into it before, and I don't think it shows what he wants it to show.

Suppose I wanted to convince you that men and women had physically identical bodies. I run studies on things like number of arms, number of kidneys, size of the pancreas, caliber of the aorta, whether the brain is in the head or the chest, et cetera. 90% of these come back identical – in fact, the only ones that don't are a few outliers like "breast size" or "number of penises". I conclude that men and women are mostly physically similar. I can even make a statistic like "men and women are physically the same in 78% of traits".

Then I go back to the person who says women have larger breasts and men are more likely to have penises, and I say "Ha, actually *studies prove* men and women are mostly physically identical! I sure showed you, you sexist!"

I worry that Hyde's analysis plays the same trick. She does a wonderful job finding that men and women have minimal differences in eg "likelihood of smiling when not being observed", "interpersonal leadership style", et cetera. But if you ask the man on the street "Are men and women different?", he's likely to say something like "Yeah, men are more aggressive and women are more sensitive". And in fact, Hyde found that men were indeed definitely more aggressive, and women indeed definitely more sensitive. But throw in a hundred other effects nobody cares about like “likelihood of smiling when not observed”, and you can report that “78% of gender differences are small or zero”.

Hyde found moderate or large gender differences in (and here I’m paraphrasing very scientific-sounding constructs into more understandable terms) aggressiveness, horniness, language abilities, mechanical abilities, visuospatial skills, mechanical ability, tendermindness, assertiveness, comfort with body, various physical abilities, and computer skills.

Perhaps some peeople might think that finding moderate-to-large-differences in mechanical abilities, computer skills, etc *supports* the idea that gender differences might play a role in gender balance in the tech industry. But because Hyde’s meta-analysis drowns all of this out with stuff about smiling-when-not-observed, Grant is able to make it sound like Hyde proves his point.

It’s actually worse than this, ~~because Grant misreports the study findings in various ways~~ [EDIT: Or possibly not, see here]. For example, he states that the sex differences in physical aggression and physical strength are “large”. The study very specifically says the opposite of this. Its three different numbers for physical aggression (from three different studies) are 0.4, 0.59, and 0.6, and it sets a cutoff for “large” effects at 0.66 or more.

On the other hand, Grant fails to report an effect that actually *is* large: mechanical reasoning ability (in the paper as Feingold 1998 DAT mechanical reasoning). There is a large gender difference on this, d = 0.76.

And although Hyde doesn’t look into it in her meta-analysis, other meta-analyses like this one find a large effect size (d = 1.18) for thing-oriented vs. people-oriented interest, the very claim that the argument that Grant is trying to argue against centers around.

So Grant tries to argue against large thing-oriented vs. people-oriented differences by citing a meta-analysis that doesn't look into them at all. He then misreports the findings of that meta-analysis, exaggerating effects that fit his thesis and failing to report the ones that don't. Finally, he cites a "summary statistic" that averages away the variation we're looking for out by combining it with a bunch of noise, and claims the noise proves his point even though the variation is as big as ever.

II.

Next, Grant claims that there are no sex differences in mathematical ability, and also that the sex differences in mathematical ability are culturally determined. ~~I'm not really sure what he means~~ [EDIT: He means sex differences that exist in other countries] but I agree with his first argument – at the levels we're looking at, there's no gender difference in math ability.

Grant says that these foreign differences in math ability exist but are due to stereotypes, and so are less noticeable in more progressive, gender-equitable nations:

> Girls do as well as boys—or slightly better—in math in elementary, but boys have an edge by high school. Male advantages are more likely to exist in countries that lack gender equity in school enrollment, women in research jobs, and women in parliament—and that have stereotypes associating science with males.

Again, my research suggests no average gender difference in ability, so I can't speak to whether these differences are caused by stereotypes or not. But I want to go back to the original question: why is there a gender difference in tech-industry-representation [in the US]? Is this also due to stereotypes and the effect of an insufficiently gender-equitable society? Do we find that “countries that lack gender equity in school enrollment” and “stereotypes associating science with males” have fewer women in tech?

No. Galpin investigated the percent of women in computer classes all around the world. Her number of 26% for the US is slightly higher than I usually hear, probably because it’s older (the percent women in computing has actually gone down over time!). The least sexist countries I can think of – Sweden, New Zealand, Canada, etc – all have somewhere around the same number (30%, 20%, and 24%, respectively). The *most* sexist countries do extremely well on this metric! The highest numbers on the chart are all from non-Western, non-First-World countries that do middling-to-poor on the Gender Development Index: Thailand with 55%, Guyana with 54%, Malaysia with 51%, Iran with 41%, Zimbabwe with 41%, and Mexico with 39%. Needless to say, Zimbabwe is not exactly famous for its deep commitment to gender equality.

Why is this? It’s a very common and well-replicated finding that the more progressive and gender-equal a country, the larger gender differences in personality of the sort Hyde found become. I agree this is a very strange finding, but it’s definitely true. See eg *Journal of Personality and Social Psychology*, Sex Differences In Big Five Personality Traits Across 55 Cultures:

> Previous research suggested that sex differences in personality traits are larger in prosperous, healthy, and egalitarian cultures in which women have more opportunities equal with those of men. In this article, the authors report cross-cultural findings in which this

> unintuitive result was replicated across samples from 55 nations (n = 17,637).

In case you're wondering, the countries with the highest gender differences in personality are France, Netherlands, and the Czech Republic. The countries with the lowest sex differences are Indonesia, Fiji, and the Congo.

I conclude that whatever gender-equality-stereotype-related differences Grant has found in the nonexistent math ability difference between men and women, they are more than swamped by the large opposite effects in gender differences in personality. This meshes with what I've been saying all along: at the level we're talking about here, *it's not about ability, it's about interest*.

III.

> We know that interests are highly malleable. Female students become significantly more interested in science careers after having a teacher who discusses the problem of underrepresentation. And at Harvey Mudd College, computer science majors were around 10% women a decade ago. Today they're 55%.

I highly recommend Freddie deBoer's Why Selection Bias Is The Most Powerful Force In Education. If an educational program shows amazing results, and there's any possible way it's selection bias – then it's selection bias.

I looked into Harvey Mudd's STEM admission numbers, and, sure enough, they admit women at 2.5x the rate as men. So, yeah, it's selection bias.

I don't blame them. All they have to do is cultivate a reputation as a place to go if you're a woman interested in computer science, attract lots of female CS applicants, then make sure to admit all the CS-interested female applicants they get. In exchange, they get constant glowing praise from every newspaper in the country (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, etc, etc, etc).

How would we know this was selection bias if we couldn't just look at the numbers? The graph that Grant himself cites just above this statement shows that, over the same ten year period, percent women CS graduates has declined nationwide. This has corresponded with such a massive push to get more women in tech that…well, that a college which succeeds will get constant glowing praise from every newspaper in the country even when they admit they're using selection bias. Do you think no one else has tried? Every college diversity office in the country is working overtime to try to get more women into tech, there are women in tech scholarships, women in tech conferences, women in tech prizes – and, over the period that's happened, Grant's own graph shows the percent of women in tech going down.

(I don't understand why it's going down as opposed to steady, but my guess is a combination of constant messaging that there are no women in tech making women think it isn't for them, plus the effect from society getting more gender-equitable that we described in Part II – ie we're now less like Zimbabwe, and so we can't expect our gender ratios to be as good as theirs are).

Look. If I recruit only gingers, and I admit only gingers, I can get a 100% ginger CS program. That doesn't mean I've proven that gingers are really more interested in CS than everyone else, and it was just discrimination holding them back. It means I've done what every single private school and college does anyway, all the time – finagle with admissions to make myself look good.

[EDIT: Some further discussion by Mudd students in the comments here]

IV.

Back to Grant:

> 4. There are sex differences in interests, but they're not biologically determined.
>
> The data on occupational interests do reveal strong male preferences for working with things and strong female preferences for working with people. But they also reveal that men and women are equally interested in working with data.
>
> So why are there so many more male than female engineers? Because women have systematically been discouraged from working with computers. Look at trends in college majors: since the 1980s, the proportion of female majors has gone up in science and medicine and law, but down in computer science.

Before we discuss this, a quick step back.

In the year 1850, women were locked out of almost every major field, with a few exceptions like nursing and teaching. The average man of the day would have been equally confident that women were unfit for law, unfit for medicine, unfit for mathematics, unfit for linguistics, unfit for engineering, unfit for journalism, unfit for psychology, and unfit for biology. He would have had various sexist justifications – women shouldn't be in law because it's too competitive and high-pressure; women shouldn't be in medicine because they're fragile and will faint at the sight of blood; et cetera.

As the feminist movement gradually took hold, women conquered one of these fields after another. 51% of law students are now female. So are 49.8% of medical students, 45% of math majors, 60% of linguistics majors, 60% of journalism majors, 75% of psychology majors, and 60% of biology postdocs. Yet for some reason, engineering remains only about 20% female.

And everyone says “Aha! I bet it’s because of negative stereotypes!”

This makes *no sense*. There were negative stereotypes about everything! Somebody has to explain why the equal and greater negative stereotypes against women in law, medicine, etc were completely powerless, yet for some reason the negative stereotypes in engineering were the ones that took hold and prevented women from succeeding there.

And if your answer is just going to be that apparently the negative stereotypes in engineering were stronger than the negative stereotypes about everything else, why would that be? Put yourself in the shoes of our Victorian sexist, trying to maintain his male privilege. He thinks to himself “Well, I suppose I could tolerate women doctors saving my life. And if I had to, I would accept women going into law and determining who goes free and who goes to jail. I’m even sort of okay with women going into journalism and crafting the narratives that shape our world. But women building bridges? NO MERE FEMALE COULD EVER DO SUCH A THING!” *Really?* This is the best explanation the world can come up with? Doesn’t *anyone* have at least a little bit of curiousity about this?

(and I don’t think it’s just coincidence – ie I don’t think it’s just that a bunch of head engineers happened to be really sexist, and a bunch of head doctors happened to be really non-sexist. The same patterns apply through pretty much every First World country, and if it were just a matter of personalities you would expect them to differ from place to place.)

Whenever I ask this question, I get something like “engineering and computer science are two of the highest-paying, highest-status jobs, so of course men would try to keep women out of them, in order to maintain their supremacy”. But I notice that doctors and lawyers are also pretty high-paying, high-status jobs, and that nothing of the sort happened there. And that when people aren’t using engineering/programming’s high status to justify their beliefs about gender stereotypes in it, they’re ruthlessly making fun of engineers and programmers, whether it’s watching *Big Bang Theory* or reading *Dilbert* or just going on about “pocket protectors”.

Meanwhile, men make up only 10% of nurses, only 20% of new veterinarians, only 25% of new psychologists, about 25% of new paediatricians, about 26% of forensic scientists, about 28% of medical managers, and 42% of new biologists.

Note that many of these imbalances are even more lopsided than the imbalance favoring men in technology, and that many of these jobs earn much more than the average programmer. For example, the average computer programmer only makes about $80,000; the average veterinarian makes about $88,000, and the average pediatrician makes a whopping $170,000.

As long as you’re comparing some poor woman janitor to a male programmer making $80,000, you can talk about how it’s *clearly* sexism against women getting the good jobs. But once you take off the blinders and try to look at an even slightly bigger picture, you start wondering why veterinarians, who make even more money than that, are even more lopsidedly female than programmers are male. And then you start thinking that maybe you need some framework more sophisticated than the simple sexism theory in order to predict who's doing all of these different jobs. And once you have that framework, maybe the sexism theory isn't necessary any longer, and you can throw it out, and use the same theory to predict why women dominate veterinary medicine and psychology, why men dominate engineering and computer science, and why none of this has any relation at all to what fields that some sexist in the 1850s wanted to keep women out of.

So let's look deeper into what prevents women from entering these STEM fields.

Does it happen at the college level? About 20% of high school students taking AP Computer Science are women. The ratio of women graduating from college with computer science degrees is exactly what you would expect from the ratio of women who showed interest in it in high school (the numbers are even lower in Britain, where 8% of high school computer students are girls). So differences exist before the college level, and nothing that happens at the college level – no discriminatory professors, no sexist classmates – change the numbers at all.

Does it happen at the high school level? There's not a lot of obvious room for discrimination – AP classes are voluntary; students who want to go into them do, and students who don't want to go into them don't. There are no prerequisites except basic mathematical competency or other open-access courses. It seems like of the people who voluntarily choose to take AP classes that nobody can stop them from going into, 80% are men and 20% are women, which exactly matches the ratio of each gender that eventually get tech company jobs.

Rather than go through every step individually, I'll skip to the punch and point out that the same pattern repeats in middle school, elementary school, and about as young as anybody has ever bothered checking. So something produces these differences very early on? What might that be?

Might young women be avoiding computers because they've absorbed stereotypes telling them that they're not smart enough, or that they're "only for boys"? No. As per Shashaani 1997, "[undergraduate] females strongly agreed with the statement 'females have as much ability as males when learning to use computers', and strongly disagreed with the statement 'studying about computers is more important for men than for women'. On a scale of 1-5, where 5 represents complete certainty in gender equality in computer skills, and 1 completely certainty in inequality, the average woman chooses 4.2; the average male 4.03. This seems to have been true since the very beginning of the age of personal computers: Smith 1986 finds that "there were no significant differences between males and females in their attitudes of efficacy or sense of confidence in ability to use the computer, contrary to expectation…females [showed] stronger beliefs in equity of ability and competencies in use of the computer." This is a very consistent result and you can find other studies corroborating it in the bibliographies of both papers.

Might girls be worried not by stereotypes about computers themselves, but by stereotypes that girls are bad at math and so can't succeed in the math-heavy world of computer science? No. About 45% of college math majors are women, compared to (again) only 20% of computer science majors. Undergraduate mathematics itself more-or-less shows gender parity. This can't be an explanation for the computer results.

Might sexist parents be buying computers for their sons but not their daughters, giving boys a leg up in learning computer skills? In the 80s and 90s, everybody was certain that this was the cause of the gap. Newspapers would tell lurid (and entirely hypothetical) stories of girls sitting down to use a computer when suddenly a boy would show up, push her away, and demand it all to himself. But move forward a few decades and now young girls are more likely to own computers than young boys – with little change in the high school computer interest numbers. So that isn't it either.

So if it happens before middle school, and it's not stereotypes, what might it be?

One subgroup of women does *not* display these gender differences at any age. These are women with congenital adrenal hyperplasia, a condition that gives them a more typically-male hormone balance. For a good review, see Gendered Occupational Interests: Prenatal Androgen Effects on Psychological Orientation to Things Versus People. They find that:

> Consistent with hormone effects on interests, females with CAH are considerably more interested than are females without CAH in male-typed toys, leisure activities, and occupations, from childhood through adulthood (reviewed in Blakemore et al., 2009; Cohen-Bendahan et al., 2005); adult females with CAH also engage more in male-typed occupations than do females without CAH (Frisén et al., 2009). Male-typed interests of females with CAH are associated with degree of androgen exposure, which can be inferred from genotype or disease characteristics (Berenbaum et al., 2000; Meyer-Bahlburg et al., 2006; Nordenström et al., 2002). Interests of males with CAH are similar to those of males

> without CAH because both are exposed to high (sex-typical) prenatal androgens and are reared as boys.
>
> Females with CAH do not provide a perfect test of androgen effects on gendered characteristics because they differ from females without CAH in other ways; most notably they have masculinized genitalia that might affect their socialization. But, there is no evidence that parents treat girls with CAH in a more masculine or less feminine way than they treat girls without CAH (Nordenström et al., 2002; Pasterski et al., 2005). Further, some findings from females with CAH have been confirmed in typical individuals whose postnatal behavior has been associated with prenatal hormone levels measured in amniotic fluid. Amniotic testosterone levels were found to correlate positively with parent-reported male-typed play in girls and boys at ages 6 to 10 years (Auyeung et al., 2009).
>
> The psychological mechanism through which androgen affects interests has not been well-investigated, but there is some consensus that sex differences in interests reflect an orientation toward people versus things (Diekman et al., 2010) or similar constructs, such as organic versus inorganic objects (Benbow et al., 2000). The Things-People distinction is, in fact, the major conceptual dimension underlying the measurement of the most widely-used model of occupational interests (Holland, 1973; Prediger, 1982); it has also been used to represent leisure interests (Kerby and Ragan, 2002) and personality (Lippa, 1998).

In their own study, they compare 125 people and find a Things-People effect size of -0.75 – that is, the difference between CAH women and unaffected women is more than half the difference between men and unaffected women. They write:

> The results support the hypothesis that sex differences in occupational interests are due, in part, to prenatal androgen influences on differential orientation to objects versus people. Compared to unaffected females, females with CAH reported more interest in occupations related to Things versus People, and relative positioning on this interest dimension was substantially related to amount of prenatal androgen exposure.

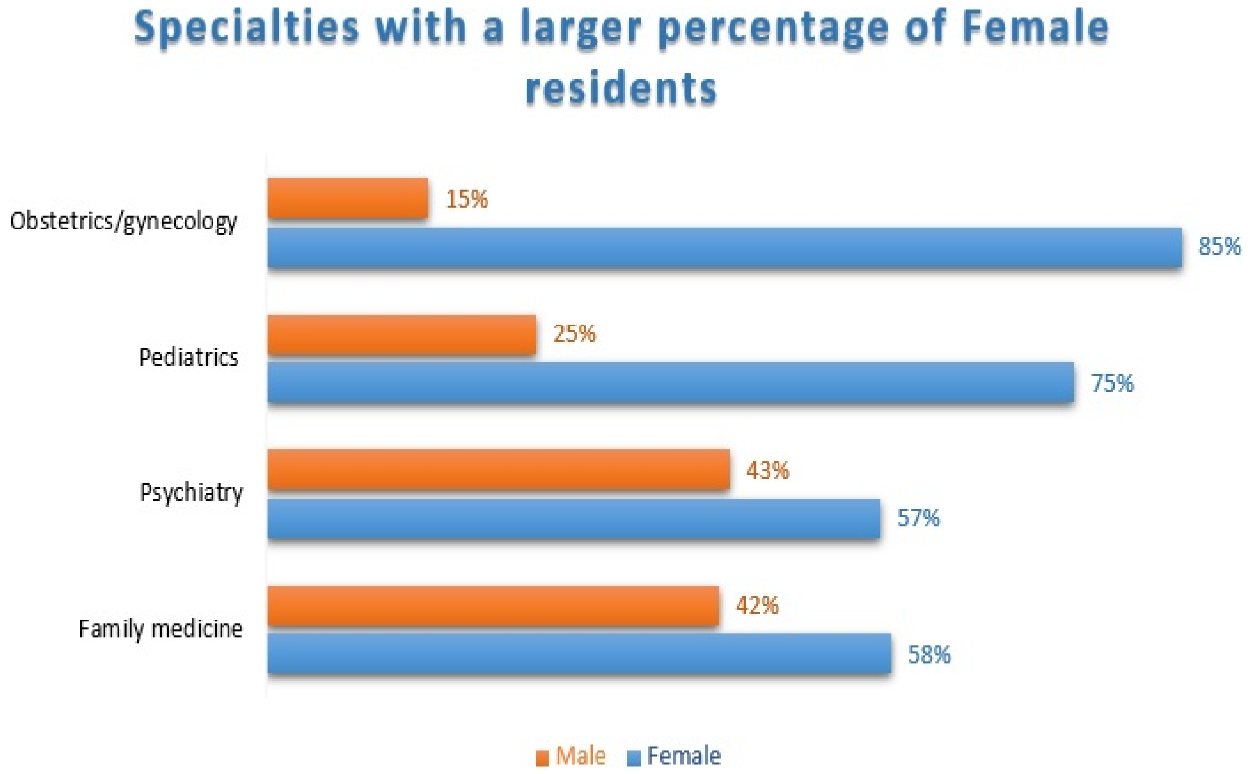
What is this "object vs. people" distinction?

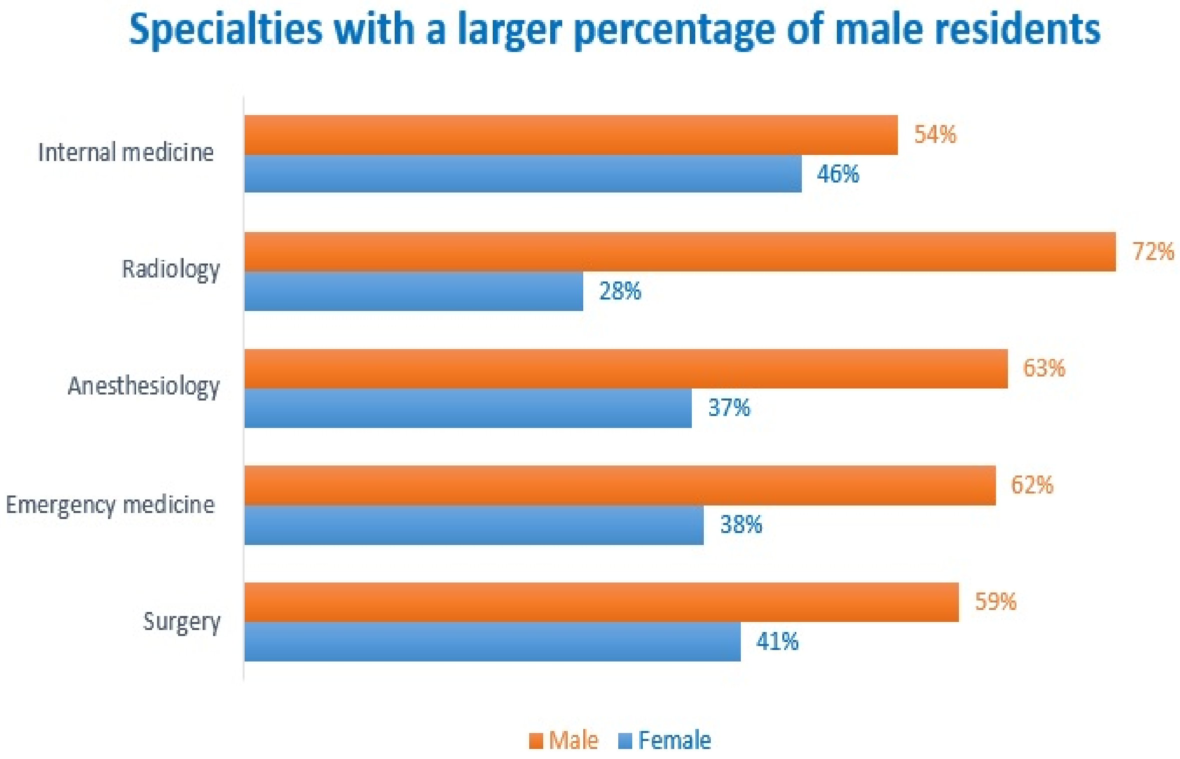
It's pretty relevant. Meta-analyses have shown a very large (d = 1.18) difference in healthy men and women (ie without CAH) in this domain. It's traditionally summarized as "men are more interested in things and women are more interested in people". I would flesh out "things" to include both physical objects like machines as well as complex abstract systems; I'd also add in another finding from those same studies that men are more risk-taking and like danger. And I would flesh out "people" to include communities, talking, helping, children, and animals.

So this theory predicts that men will be more likely to choose jobs with objects, machines, systems, and danger; women will be more likely to choose jobs with people, talking, helping, children, and animals.

Somebody armed with *this* theory could pretty well pretty well predict that women would be interested in going into medicine and law, since both of them involve people, talking, and helping. They would predict that women would dominate veterinary medicine (animals, helping), psychology (people, talking, helping, sometimes children), and education (people, children, helping). Of all the hard sciences, they might expect women to prefer biology (animals). And they might expect men to do best in engineering (objects, machines, abstract systems, sometimes danger) and computer science (machines, abstract systems).

I mentioned that about 50% of medical students were female, but this masks a lot of variation. There are wide differences in doctor gender by medical specialty. For example:





A privilege-based theory fails – there's not much of a tendency for women to be restricted to less prestigious and lower-paying fields – Ob/Gyn (mostly female) is extremely lucrative, and internal medicine (mostly male) is pretty low-paying for a medical job.

But the people/thing theory above does extremely well! Pediatrics is babies/children, Psychiatry is people/talking (and of course women are disproportionately child psychiatrists), OB/GYN is babies (though admittedly this probably owes a lot to patients being more comfortable with female gynecologists) and family medicine is people/talking/babies/children.

Meanwhile, Radiology is machines and no patient contact, Anaesthesiology is also machines and no patient contact, Emergency Medicine is danger, and Surgery is machines, danger, *and* no patient contact.

Here’s another fun thing you can do with this theory: understand why women are so *well* represented in college math classes. Women are around 20% of CS majors, physics majors, engineering majors, etc – but almost half of math majors! This should be shocking. Aren’t we constantly told that women are bombarded with stereotypes about math being for men? Isn’t the archetypal example of children learning gender roles that Barbie doll that said “Math is hard, let’s go shopping?” And yet women’s representation in undergraduate math classes is really quite good.

I was totally confused by this for a while until a commenter directed me to the data on what people actually do with math degrees. The answer is mostly: they become math teachers. They work in elementary schools and high schools, with people.

Then all those future math teachers leave for the schools after undergrad, and so math grad school ends up with pretty much the same male-tilted gender balance as CS, physics, and engineering grad school.

This seems to me like the clearest proof that women being underrepresented in CS/physics/etc is just about different interests. It’s not that they can’t do the work – all those future math teachers do just as well in their math majors as everyone else. It’s not that stereotypes of what girls can and can’t do are making them afraid to try – whatever stereotypes there are about women and math haven’t dulled future math teachers’ willingness to compete difficult math courses one bit. And it’s not even about colleges being discriminatory and hostile (or at least however discriminatory and hostile they are it doesn’t drive away those future math teachers). It’s just that women are more interested in some jobs, and men are more interested in others. Figure out a way to make math people-oriented, and women flock to it. If there were as many elementary school computer science teachers as there are math teachers, gender balance there would equalize without any other effort.

I’m not familiar with any gender breakdown of legal specialties, but I will bet you that family law, child-related law, and various prosocial helping-communities law are disproportionately female, and patent law, technology law, and law working with scary dangerous criminals are disproportionately male. And so on for most other fields.

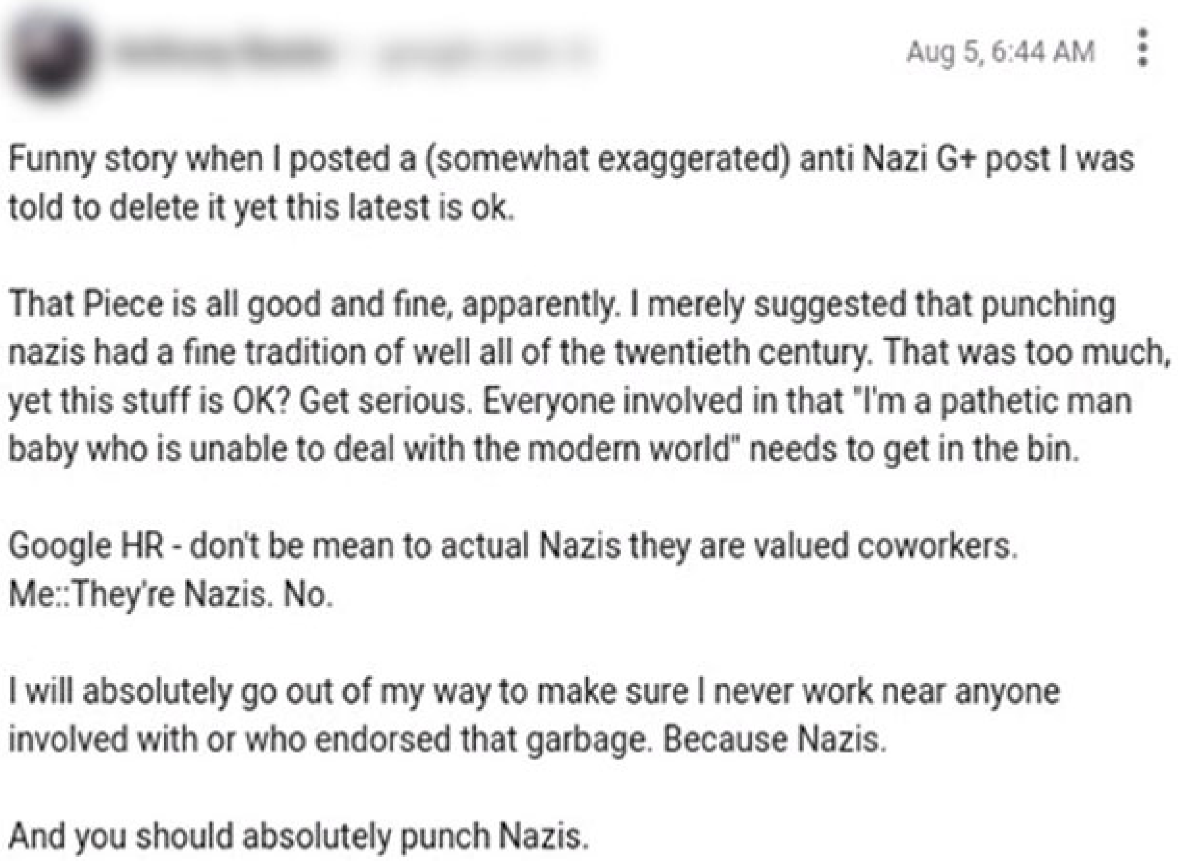
This theory gives everyone what they want. It explains the data about women in tech. It explains the time course around women in tech. It explains other jobs like veterinary medicine where women dominate. It explains which medical subspecialties women will be dominant or underrepresented in. It doesn’t claim that women are “worse than men” or “biologically inferior” at anything. It doesn’t say that no woman will ever be interested in things, or no man ever interested in people. It doesn’t say even that women in tech don’t face a lot of extra harassment (any domain with more men than women will see more potential perpetrators concentrating their harassment concentrated on fewer potential victims, which will result in each woman being more harassed).

It just says that sometimes, in a population-based way that doesn’t necessarily apply to any given woman or any given man, women and men will have some different interests. Which should be pretty obvious to anyone who’s spent more than a few minutes with men *or* women.

V.

Why am I writing this?

Grant’s piece was in response to a person at Google sending out a memo claiming some of this stuff. Here is a pretty typical response that a Googler sent to that memo – I’ve blocked the name so this person doesn’t get harassed over it, but if you doubt this is real I can direct you to the original:


[redacted] Aug 5, 6:44 AM

Funny story when I posted a (somewhat exaggerated) anti Nazi G+ post I was told to delete it yet this latest is ok.

That Piece is all good and fine, apparently. I merely suggested that punching nazis had a fine tradition of well all of the twentieth century. That was too much, yet this stuff is OK? Get serious. Everyone involved in that "I'm a pathetic man baby who is unable to deal with the modern world" needs to get in the bin.

Google HR - don't be mean to actual Nazis they are valued coworkers.
Me::They're Nazis. No.

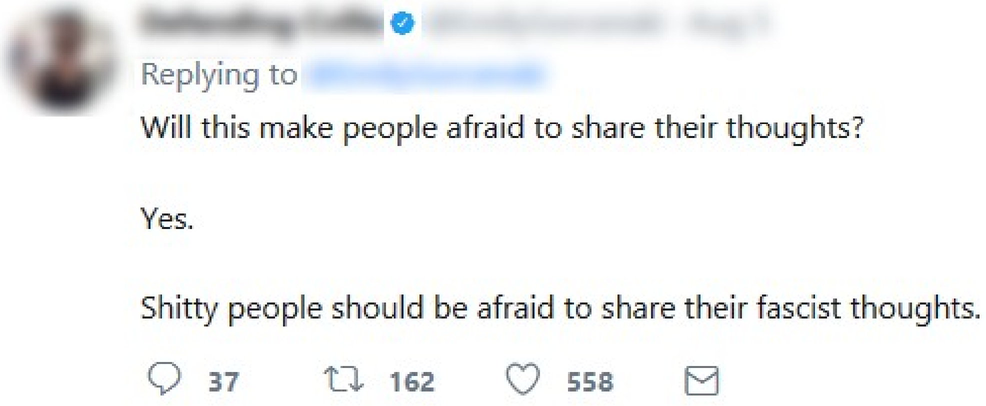
I will absolutely go out of my way to make sure I never work near anyone involved with or who endorsed that garbage. Because Nazis.

And you should absolutely punch Nazis.

A lot of people without connections to the tech industry don’t realize how bad it’s gotten. This is how bad. It would be pointless trying to do anything about this person in particular. This is the climate.

Silicon Valley was supposed to be better than this. It was supposed to be the life of the mind, where people who were interested in the mysteries of computation and cognition could get together and make the world better for everybody. Now it’s degenerated into this giant hatefest of everybody writing long screeds calling everyone else Nazis and demanding violence against them. Where if someone disagrees with the consensus, it's just taken as a matter of course that we need to hunt them down, deny them of the cloak of anonymity, fire them, and blacklist them so they can never get a job again. Where the idea that we *shouldn't* be a surveillance society where we carefully watch our coworkers for signs of sexism so we can report them to the authorities is exactly the sort of thing you get reported to the authorities if people see you saying.

On the Twitter debate on this, someone mentioned that people felt afraid to share their thoughts anymore. An official, blue-checkmarked Woman In Tech activist responded with (note the 500+ likes):



Replying to

Will this make people afraid to share their thoughts?

Yes.

Shitty people should be afraid to share their fascist thoughts.

This is the world we've built. Where making people live in fear is a feature, not a bug.

And: it can get worse. If you only read one link, let it be this one about [the young adult publishing industry]() A sample quote:

> One author and former diversity advocate described why she no longer takes part: "I have never seen social interaction this fucked up," she wrote in an email. "And I've been in prison."

Many members of YA Book Twitter have become culture cops, monitoring their peers across multiple platforms for violations. The result is a jumble of dogpiling and dragging, subtweeting and screenshotting, vote-brigading and flagging wars, with accusations of white supremacy on one side and charges of thought-policing moral authoritarianism on the other. Representatives of both factions say they've received threats or had to shut down their accounts owing to harassment, and all expressed fear of being targeted by influential community members — even when they were ostensibly on the same side. "If anyone found out I was talking to you," Mimi told me, "I would be blackballed."

Dramatic as that sounds, it's worth noting that my attempts to report this piece were met with intense pushback. Sinyard politely declined my request for an interview in what seemed like a routine exchange, but then announced on Twitter that our interaction had "scared" her, leading to backlash from community members who insisted that the as-yet-unwritten story would endanger her life. Rumors quickly spread that I had threatened or harassed Sinyard; several influential authors instructed their followers not to speak to me; and one librarian and member of the Newbery Award committee tweeted at Vulture nearly a dozen times accusing them of enabling "a washed-up YA author" engaged in "a personalized crusade" against the entire publishing community (disclosure: while freelance culture writing makes up the bulk of my work, I published a pair of young adult novels in 2012 and 2014.) With one exception, all my sources insisted on anonymity, citing fear of professional damage and abuse.

None of this comes as a surprise to the folks concerned by the current state of the discourse, who describe being harassed for dissenting from or even questioning the community’s dynamics. One prominent children’s-book agent told me, “None of us are willing to comment publicly for fear of being targeted and labeled racist or bigoted. But if children’s-book publishing is no longer allowed to feature an unlikable character, who grows as a person over the course of the story, then we’re going to have a pretty boring business.”

Another agent, via email, said that while being tarred as problematic may not kill an author’s career — “It’s likely made the rounds as gossip, but I don’t know it’s impacting acquisitions or agents offering representation” — the potential for reputational damage is real: “No one wants to be called a racist, or sexist, or homophobic. That stink doesn’t wash off.”

Authors seem acutely aware of that fact, and are tailoring their online presence — and in some cases, their writing itself — accordingly. One New York Times best-selling author told me, “I’m afraid. I’m afraid for my career. I’m afraid for offending people that I have no intention of offending. I just feel unsafe, to say much on Twitter. So I don’t.” She also scrapped a work in progress that featured a POC character, citing a sense shared by many publishing insiders that to write outside one’s own identity as a white author simply isn’t worth the inevitable backlash. “I was told, do not write that,” she said. “I was told, ‘Spare yourself.’

Another author recalled being instructed by her publisher to stay silent when her work was targeted, an experience that she says resulted in professional ostracization. “I never once responded or tried to defend my book,” she wrote in a Twitter DM. Her publisher “did feel I was being abused, but felt we couldn’t do anything about it.”

Parts of tech are already this bad. For the rest of you: it’s what you have to look forward to.

It doesn’t have to be this way. Nobody has any real policy disagreements. Everyone can just agree that men and women are equal, that they both have the same rights, that nobody should face harassment or discrimination. We can relax the Permanent State Of Emergency around too few women in tech, and admit that women have the right to go into whatever field they want, and that if they want to go off and be 80% of veterinarians and 74% of forensic scientists, those careers seem good too. We can appreciate the contributions of existing women in tech, make sure the door is open for any new ones who want to join, and *start treating each other as human beings again*. Your co-worker could just be your co-worker, not a potential Nazi to be assaulted or a potential Stalinist who’s going to rat on you. Your project manager could just be your project manager, not the person tasked with monitoring you for signs of thoughtcrime. Your female co-worker could just be your female co-worker, not a Badass Grrl Coder Who Overcomes Adversity. Your male co-worker could just be your male co-worker, not a Tool Of The Patriarchy Who Denies His Complicity In Oppression. I promise there are industries like this. Medicine is like this! Loads of things are like this! Lots of tech companies are even still like this! *This could be you*.

Adam Grant seems like a good person. He is superficially doing everything right. He’s not demanding people feel afraid, or saying that everyone who disagrees with him is a fascist. He’s just trying to argue the science.

But I think he’s very wrong about the science. I think Hyde’s article is a gimmick which buries very real differences under a heap of meaningless similarities. I think that it’s inappropriate to cite it to respond to claims of specific differences that it didn’t investigate. I think that claims of a gender-equitable-society-effect in a different domain are inappropriate given the clear opposite effect in the domain being talked about it. I think it’s wrong to privilege likely-selection-biased evidence from a single college over all the evidence from the country as a whole. I think it’s wrong to suppose unique stereotypes in tech and engineering domains with no theory of how they got there, when there are non-stereotype-based theories that better explain the evidence. And I think it’s wrong to ignore all the studies about congenital adrenal hyperplasia.

And I think that, in being wrong about the science, he’s (probably unintentionally) giving aid and comfort to the people who have admitted that turning tech into a climate of fear and threats of violence is the end goal.

Grant is one of the few people doing the virtuous thing and trying to debate this without calling for other people’s deaths. I’m trying to do the virtuous thing and respond to him. But I worry that lots of people on Grant’s side aren’t as virtuous as he is, and I don’t know how to protect anybody from that except by begging people to please look at the science and try to get it right.

[**EDIT:** Prof. Grant responded to me via email; with his permission I’ve posted his response as a comment below (ie here). My thoughts on his response are in comments below that.]

The Lizard People Of Alpha Draconis 1 Decided To Build An Ansible

I.

The lizard people of Alpha Draconis 1 decided to build an ansible.

The transmitter was a colossal tower of silksteel, doorless and windowless. Inside were millions of modular silksteel cubes, each filled with beetles, a different species in every cube. Big beetles, small beetles, red beetles, blue beetles, friendly beetles, venomous beetles. There hadn’t been a million beetle species on Alpha Draconis I before the ansible. The lizard people had genetically engineered them, carefully, lovingly, making each one just different enough from all the others. Atop each beetle colony was a heat lamp. When the heat lamp was on, the beetles crawled up to the top of the cage, sunning themselves, basking in the glorious glow. When it turned off, they huddled together to warmth, chittering out their anger in little infrasonic groans only they could hear.

The receiver stood on 11845 Nochtli, eighty-five light years from Alpha Draconis, toward the galactic rim. It was also made of beetles, a million beetle colonies of the same million species that made up the transmitter. In each beetle colony was a pheromone dispenser. When it was on, the beetles would multiply until the whole cage was covered in them. When it was off, they would gradually die out until only a few were left.

Atop each beetle cage was a mouse cage, filled with a mix of white and grey mice. The white mice had been genetically engineered to want all levers in the “up” position, a desire beyond even food or sex in its intensity. The grey mice had been engineered to want levers in the “down” position, with equal ferocity. The lizard people had uplifted both strains to full sapience. In each of a million cages, the grey and white mice would argue whether levers should be up or down – sometimes through philosophical debate, sometimes through outright wars of extermination.

There was one lever in each mouse cage. It controlled the pheromone dispenser in the beetle cage just below.

This was all the lizard people of Alpha Draconis 1 needed to construct their ansible.

They had mastered every field of science. Physics, mathematics, astronomy, cosmology. It had been for nothing. There was no way to communicate faster-than-light. Tachyons didn’t exist. Hyperspace didn’t exist. Wormholes didn’t exist. The light speed barrier was absolute – *if* you limited yourself to physics, mathematics, astronomy, and cosmology.

The lizard people of Alpha Draconis I weren’t going to use any of those things. They were going to build their ansible out of negative average preference utilitarianism.

II.

Utilitarianism is a moral theory claiming that an action is moral if it makes the world a better place. But what do we mean by “a better place”?

Suppose you decide (as Jeremy Bentham did) that it means increasing the total amount of happiness in the universe as much as possible – the greatest good for the greatest number. Then you run into a so-called “repugnant conclusion”. The philosophers quantify happiness into “utils”, some arbitrary small unit of happiness. Suppose your current happiness level is 100 utils. And suppose you could sacrifice one util of happiness to create another person whose total happiness is two utils: they are only 1/50th as happy as you are. This person seems quite unhappy by our standards. But crucially, their total happiness is positive; they would (weakly) prefer living to dying. Maybe we can imagine this as a very poor person in a war-torn Third World country who is (for now) not actively suicidal.

It would seem morally correct to make this sacrifice. After all, you are losing one unit of happiness to create two units, increasing the total happiness in the universe. In fact, it would seem morally correct to keep making the sacrifice as many times as you get the opportunity. The end result is that you end up with a happiness of 1 util – barely above suicidality – and also there are 99 extra barely-not-suicidal people in war-torn Third World countries.

And the same moral principles that lead you to make the sacrifice bind everyone else alike. So the end result is everyone in the world ends up with the lowest possible positive amount of happiness, plus there are billions of extra near-suicidal people in war-torn Third World countries.

This seems abstract, but in some sense it might be the choice on offer if we have to decide whether to control population growth (thus preserving enough resources to give everyone a good standard of living), or continue explosive growth so that there are many more people but not enough resources for any of them to live comfortably.

The so-called “repugnant conclusion” led many philosophers away from “total utilitarianism” to “average utilitarianism”. Here the goal is still to make the world a better place, but it gets operationalized as “increase the average happiness level per person”. The repugnant conclusion clearly fails at this, so we avoid that particular trap.

But here we fall into another ambush: wouldn’t it be morally correct to kill unhappy people? This raises average happiness very effectively!

So we make another amendment. We’re not in the business of raising happiness, per se. We’re in the business of satisfying preferences. People strongly prefer not to die, so you can’t just kill them. Killing them actively *lowers* the average number of satisfied preferences.

Philosopher Roger Chao combines these and other refinements of the utilitarian method into a moral theory he calls negative average preference utilitarianism, which he considers the first system of ethics to avoid all the various traps and pitfalls. It says: an act is good if it decreases the average number of frustrated preferences per person.

This doesn’t imply we should create miserable people ad nauseum until the whole world is a Third World slum. It doesn’t imply that we should kill everyone who cracks a frown. It doesn’t imply we should murder people for their organs, or never have children again, or replace everybody with identical copies of themselves, or anything like that.

It just implies faster-than-light transmission of moral information.

III.

The ansible worked like this:

Each colony of beetles represented a bit of information. In the transmitter on Alpha Draconis I, the sender would turn the various colonies’ heat lamps on or off, increasing or decreasing the average utility of the beetles.

In the receiver on 11845 Nochtli, the beetles would be in a constant state of half-light: warmer than the Draconis beetles if their heat lamp was turned off, but colder than them if their heat lamp was turned on. So increasing the population of a certain beetle species on 11845 Nochtli would be morally good if the heat lamp for that species on Alpha Draconis were off, but morally evil otherwise.

The philosophers among the lizard people of Alpha Draconis 1 had realized that this was true regardless of intervening distance; morality was the only force that transcended the speed of light. The question was how to detect it. Yes, a change in the heat lamps on their homeworld would instantly change the moral valence of pulling a lever on a colony 85 light-years away, but how to detect the morality of an action?

The answer was: the arc of the moral universe is long, *but it bends toward justice*. Over time, as the great debates of history ebb and sway, evil may not be conquered completely, but it will lessen. Our own generation isn’t perfect, but we have left behind much of the slavery, bigotry, war and torture, of the past; perhaps our descendants will be wiser still. And how could this be, if not for some benevolent general rule, some principle that tomorrow must be brighter than today, and the march of moral progress slow but inevitable?

Thus the white and grey rats. They would debate, they would argue, they would even fight – but in the end, moral progress would have its way. If raising the lever and causing an increase in the beetle population was the right thing to do, then the white rats would eventually triumph; if lowering the lever and causing the beetle population to fall was right, then the victory would eventually go to the grey. All of this would be recorded by a camera watching the mouse colony, and – lo – a bit of information would have been transmitted.

The ansible of the lizard people of Alpha Draconis 1 was a flop.

They spent a century working on it: ninety years on near-light-speed starships just transporting the materials, and a decade constructing the receiver according to meticulous plans. With great fanfare, the Lizard Emperor himself sent the first message from Alpha Draconis I. And it was a total flop.

The arc of the moral universe is long, but it bends to justice. But nobody had ever thought to ask *how* long, and why. When everyone alike ought to love the good, why does it take so many years of debate and strife for virtue to triumph over wickedness? Why do war and slavery and torture persist for century after century, so that only endless grinding of the wheels of progress can do them any damage at all?

After eighty-five years of civilizational debate, the grey and white mice in each cage finally overcame their differences and agreed on the right position to put the lever, just as the mundane lightspeed version of the message from Alpha Draconis reached 11845 Nochtli's radio telescopes. And the lizard people of Alpha Draconis 1 realized that one can be more precise than simply defining the arc of moral progress as "long". It's *exactly as long as it needs to be to prevent faster-than-light transmission of moral information*.

Fundamental physical limits are a harsh master.

Brief Cautionary Notes On Branded Combination Nootropics

I.

Taking nootropics is an inherently questionable decision. The risk isn't zero, and the benefits are usually subtle at best. But even within this context, special caution is warranted for branded combination nootropics.

I wanted to make up a caricatured fake name for these sorts of things, so I could make fun of them without pointing at any company in particular. But all of the caricatured fake names I can think of turn out to be real products. MegaMind? Real. SuperBrain? Real. UltraBrain? Real. Mr. Power Brain? Real, for some reason.

Even the ones that don't make sense are real. NeuroBrain? Real, even though one hopes that brains are always at least a little neuro. NeuroMind? Real, with its own Indiegogo campaign. The only thing I haven't been able to find is a nootropic called BrainMind, but it's only a matter of time.

These usually combine ten or twenty different chemicals with potential nootropic properties, then make outrageous claims about the results. For example, Neuroxium says on its ridiculous webpage that:

> Neuroxium is a revolutionary brain supplement formulated to give you ultimate brain power. Known in Scientific Terms as a "NOOTROPIC" or "GENIUS PILL" Neuroxium improves mental functions such as cognition, memory, intelligence, motivation, attention, concentration and therefore happiness and success.

Your first warning sign should have been when they said “genius pill” was a scientific term (or as they call it, Scientific Term). If you needed more warning signs, this is word-for-word the same claim made by several other nootropics like Synagen IQ, Nootrox, and Cerebral X. So either they can’t even be bothered not to plagiarize their ads, or they change their name about once a week to stay ahead of the law.

I was eventually able to find a list of the ingredients in this stuff:

> DMAE (dimethylethanolamine bitartrate), GABA (?-Amino-butyric acid), Caffeine anhydrous, Bacopa monnieri leaf extract, NALT (N-acetyl-L-tyrosine), Centrophenoxine HCl, Alpha-GPC (a-glycerophosphocholine, Agmatine sulfate, Gingko biloba leaf extract, Pine (Pinus pinaster) bark extract, Phosphatidylserine, Aniracetam, CDC Choline (Citicoline), Sarcosine (N-methylglycine), Vincamine [Lesser Periwinkle (Vinca minor) aerial extract], L-Theanine (?-glutamylethylamide), NADH (nicotinamide adenine dinucleotide), TAU (triacetyluridine), Noopept, Adrafinil, Tianeptine, Piperine [Black Pepper (Piper nigrum) fruit extract 445mg.

And the weird thing is, a lot of these are decent choices. Everyone knows caffeine is a good stimulant. Adrafinil is the over-the-counter version of modafinil, an FDA-approved medication for sleep disorders; many of my patients have been very happy with it. Bacopa monnieri has been found to improve memory in so many studies I can’t even keep track of all of them. Noopept is an approved medication in Russia. Tianeptine is an approved medication in France. All of these are chemicals with at least some evidence base behind them, which are potentially good for certain people.

But taking Neuroxium/Synagen/CerebralX is exactly as bad an idea as you would expect from the advertising copy.

For one thing, they don’t list the doses of any of these things – but they have to be getting them terribly wrong. A standard dose of adrafinil is 600 mg. A standard dose of bacopa is 300 mg. A standard dose of Alpha-GPC choline is about 600 mg. So combining standard doses of just these three ingredients means you need a 1.5 g pill. This is probably too big to swallow. The only pills I know of that get that big are those gigantic fish oil pills made of pure fish oil that everybody hates because they’re uncomfortably big. But this is just what you’d need to have three of the 22 ingredients listed in CerebralX at full doses. The pill is already unswallowably large, and you’ve only gotten a seventh of the way through the ingredient list.

I conclude that they’re just putting miniscule, irrelevant doses into this so they can say they’ve got exciting-sounding chemicals.

For another thing, all of these substances have unique profiles which have to be respected on their own terms. For example, lots of studies say bacopa improves memory – but only after you’ve taken it consistently for several months. If you just go “WOOO, CEREBRALX!” and swallow a bunch of pills and hope that you’ll do better on your test tomorrow, all you’re going to get are the short-term effects of bacopa – which include lethargy and amotivation.

Most sources discussing Noopept recommend starting very low – maybe as low as 5 mg – and then gradually increasing to a standard dose of 10 – 40 mg depending on how it works for you. Some people will apparently need higher doses, and some find it works best for them as high as 100 mg. Needless to say, none of this is possible if you’re taking CerebralX. You’ll take whatever dose is in the product – which they don’t tell you, and which is probably so low as to be meaningless – and stay at the same level for however long you’re taking the entire monstrosity.

Tianeptine has a short half-life and is typically dosed three times a day, unlike most of the other things on the list which are dosed once per day. CerebralX says you should take their whole abomination once a day, which means you’re getting the wrong dosing schedule of tianeptine.

Piperine is a chemical usually used to inhibit normal drug-metabolizing enzymes and enhance the effect of other substances. This is very occasionally a good idea, when you know exactly what drug you’re trying to enhance and you’re not taking anything else concurrently. But I can’t figure out which drug they’re trying to enhance the activity of here, or even whether they’re trying to enhance the activity of anything at all, or if they just heard that piperine could enhance things and thought “Okay, it’s in”. And if I were giving someone a concoction of twenty-one different random psychoactive drugs, which I was dosing wrong and giving at the wrong schedule, the *last* thing I would want to do is inhibit the body’s normal drug metabolism. The entire reason God gave people drug-metabolizing enzymes is because He knew, in His wisdom, that some of them were going to be idiots who would take a concoction of twenty-one different random psychoactive drugs because a website said it was, in Scientific Terms, a “GENIUS PILL”. Turning them off is a terrible idea and the only saving grace is that the dose of everything in this monstrosity is probably too small for it to do anything anyway.

Taking any of the ingredients in CerebralX on its own is a potentially risky affair. But if you study up on it and make sure to take it correctly, then maybe it’s a calculated risk. Taking everything in CerebralX together isn’t taking a calculated risk as part of a potentially interesting hobby. It’s just being an idiot.

II.

But that’s too easy. I have a larger point here, which is that these sorts of branded combos are bad ideas *even if* they’re by smart, well-intentioned people who are doing everything right.

Tru-Brain is undeniably in a class above CerebralX. It has a team including neuroscience PhDs. It seems to be a real company that can keep the same name for more than a week. Instead of promising a GENIUS PILL, it makes comparatively modest claims to be able to “perform at your peak” and “stay sharp all day long”.

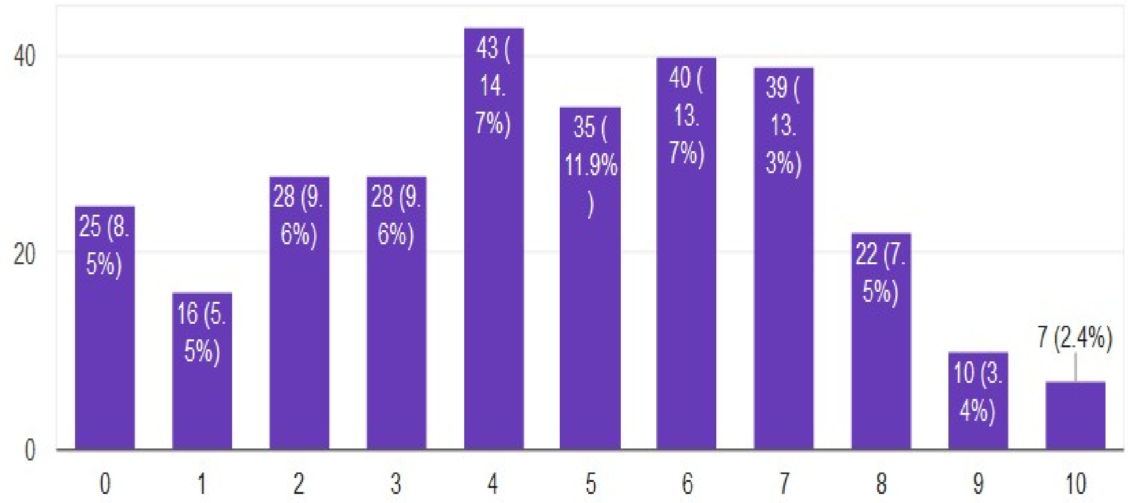
Correspondingly, its special nootropics combo makes a lot more pharmacological sense. For one thing, it’s a packet rather than a single pill – a concession to the impossibility of combining correct doses of many substances into a single capsule. For another, it limits itself to mostly things that some sane person could conceivably in some universe want to dose on the schedule they recommend. And it’s only got seven ingredients, none of which counteract any of the others or turn off important metabolic systems that God created to protect you from your own stupidity. This is probably about as well-designed a branded nootropics combo as it’s possible to make.

But I would still caution people away from it. Why?

Last year, I surveyed people’s reactions to various nootropics. I got 870 responses total, slightly fewer for each individual substance. Here are the response curves for two of the substances in TruBrain – piracetam and theanine:

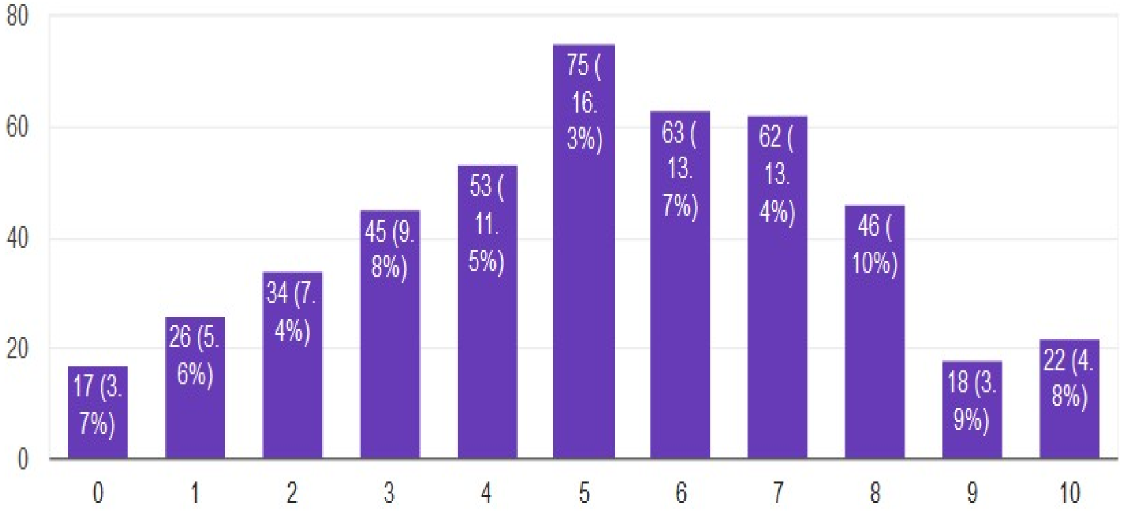
Piracetam

293 responses



Theanine

461 responses



These are on a 1-10 scale, where I directed responders to:

> Please rate your subjective experience on a scale of 0 to 10. 0 means a substance was totally useless, or had so many side effects you couldn't continue taking it. 1 – 4 means for subtle effects, maybe placebo but still useful. 5 – 9 means strong effects, definitely not placebo. 10 means life-changing.

Some substances known to be pretty inert averaged scores of around 4. Piracetam and theanine averaged around 5, so maybe a little better than that. But the most dramatic finding was the range. Almost 20% of people rated theanine a two or lower; almost 20% rated it a nine or higher. More than a third placed it in the "probably placebo" range, but 5% found its effects "life-changing".

The effect of nootropics seems to vary widely among different people. This shouldn't be surprising: so do the effects of real drugs. Gueorguieva and Mallinckrodt do an unusually thorough job modeling differences in response to the antidepressant duloxetine, and find a clear dichotomy between responders and nonresponders. This matches psychiatric lore – some medications work on some people, other medications work on others. I particularly remember one depressed patient who had no response at all to any SSRI, but whose depression shut off almost like a lightswitch once we tried bupropion. Other people fail bupropion treatment but do well on SSRIs. Probably this has something to do with underlying differences in their condition or metabolism that we just don't know how to identify at this point (sample simplified toy model: what we call "depression" is actually two diseases with identical symptoms, one of which responds to SSRIs and one of which responds to bupropion).

I think this is why there are no multidrug combo packs. Your psychiatrist never treats your depression with a pill called "MegaMood", boasting combination doses of Prozac, Wellbutrin, Remeron, and Desyrel. For one thing, either you're giving an insufficient dose of each drug, or you're giving full doses of four different drugs – neither is well-tested or advisable. For another, you're getting four times the side effect risk. For a third thing, if one of the four drugs gives you a side effect, you've got to throw out the whole combo. For a fourth, if the combo happens to work, you don't know whether it's only one of the four drugs working and the others are just giving you side effects and making you worse. And if it *sort of* works, you don't know which of the four drugs to increase, or else you just have to increase all four at once and hope for the best.

All these considerations are even stronger with nootropics. There shouldn't be universally effective nootropics, for the same reason there's no chemical you can pour on your computer to double its processing speed: evolution put a lot of work into making your brain as good as possible, and it would be silly if some random molecule could make it much better. Sure, there are exceptions – I think stimulants get a pass because evolution never expected people to have to pay attention to stimuli as boring as the modern world provides us with all the time – but in general the law holds. If you find a drug does significantly help you, it's probably because your brain is broken in some particular idiosyncratic way (cf. mutational load), the same way you *can* double a computer's processing speed with duct tape if one of the parts was broken.

If everyone's brain is broken in a different way, then not only will no drug be universally effective, but drugs with positive effects for some people are likely to have negative effects for others. If (to oversimplify) your particular brain problem is not having enough serotonin, a serotonin agonist might help you. But by the same token, if you have too much serotonin, a serotonin agonist will make your life worse. Even if you have normal serotonin, maybe the serotonin agonist will push you out of the normal range and screw things up.

Most effective psychiatric drugs hurt some people. I mean, a lot of them hurt the people they're supposed to be used for – even the psychotic people hate antipsychotics – but once you've brushed those aside, there are a lot of others that help a lot of people, but make other people feel worse. There are *hordes* of people who feel tired on stimulants, or sleepy on caffeine, or suicidal on antidepressants, or any other crazy thing. You rarely hear about these, because usually if someone's taking a drug and it makes them feel worse, they stop. But psychiatrists hear about it all the time. "That antidepressant you gave me just made me feel awful!" Oh, well, try a different one. "That's it? Try a different one? Aren't you embarassed that your so-called antidepressant made me more depressed?" You're pretty new to this 'psychopharmacology' thing, aren't you?

Thus the tactic used by every good psychiatrist: try a patient on a drug that you think might work, make them report back to you on whether it does. If so, keep it; if not, switch.

If you take a seven-drug combo pack, you lose this opportunity for experiment. Suppose that two of the drugs make you feel +1 unit better, two others have no effect, and three of the drugs make you feel -0.5 units worse, so in the end you feel +0.5 units better. Maybe that seems good to you so you keep taking it. Now you're taking five more drugs than you need to, including three making you actively worse, and you're missing the chance to be a full +2 units better by just taking the drugs that are helping and not hurting.

You're also missing the opportunity to play with the doses or the schedules of things. Maybe if you doubled the dose of one of the drugs making you +1 better, you could be +2 better, but if you double the dose of the other, you start getting side effects and the drug only breaks even. If you experiment, you can figure this out and take twice the dose of the first and the starting dose of the second, for +3 better. Taking them all as part of a combo ruins this: if you try taking twice the dose of the combo, nothing happens.

(And a special word of warning: if some stimulant product combines caffeine with something else, and you feel an effect, your first theory should be that the effect is 100% caffeine – unless the "something else" is amphetamine. There are like a million products which bill themselves as "organic energy cocktails" by combining caffeine with some rare herb from Burma. People drink these and say "Oh, this high feels so much more intense than just drinking caffeine". Yeah, that's because it's *much more* caffeine. Seriously. Check the doses on those things. I will grudgingly make an exception for some chemicals that are supposed to decrease caffeine jitters, like theanine, which might have a real effect. But *the stimulation is from caffeine*. Go get an espresso instead.)

III.

But don't drugs interact? Instead of viewing these seven drugs as seven different variables, shouldn't we view them as coming together in a seven-color beautiful rainbow of happiness, or whatever?

Once again, I can only appeal to psychiatry, which is still unsure whether there are any useful interactions between its various super-well-studied drugs which it's been using for decades and prescribing to millions of people. Take the CO-MED study, which combined the popular SSRI escitalopram with the popular NDRI bupropion. Since depression seems to involve abnormalities in the three major catecholamine systems, and escitalopram hits one of these and bupropion hits the other two, this seems like exactly the sort of synergistic interaction we should expect to work. It doesn't. CO-MED found that the two antidepressants together didn't treat depression any better than either one alone, let alone produce some synergy that made them more than the sum of their parts. They did, however, have about twice as many side effects.

Other smaller studies say the opposite, so I'm not saying never try escitalopram and bupropion together. I'm saying *we don't know*. These are intensely-studied drugs, the whole power of the existing system has been focused on the question of whether they synergize or antisynergize or what, and we're still not sure.

Also from psychiatry: we know a lot less about the mechanisms of action of drugs than we like to think. Ketamine has been intensively studied for depression for a decade or so, and we only just learned last year that it probably worked on a different receptor than we thought. SSRIs might be the most carefully studied drug class of all time, and we still don't really know exactly what's up with them – it can't just be serotonin; they increase serotonin within a day of ingestion, but take a month to work.

So when people point to these incredibly weird substances that have barely been studied at all, where we have only the faintest clue how they work, and then say from their armchair "And therefore, drug A will enhance the effects of drug B and C" – this is more than a little arrogant. Is it *all* made up? I can't say "all" with surety. But it *might* be.

The best-known and most-discussed interaction in nootropics is piracetam-choline. Piracetam increases levels of acetylcholine, which is formed from choline, so it makes sense that these two substances would go well together. Most sites on piracetam urge you to take them together. TruBrain, which predictably is on top of this kind of stuff, combines them together in its combo pack.

But there’s never been a human study showing that this helps. Examine.com, another group which is usually on top of stuff, summarizes (emphasis carried over from original):

> [Choline] **may** augment the relatively poor memory enhancing effects of Piracetam in otherwise healthy animals, but administration of choline alongside Piracetam is not a prerequisite to its efficacy and has not been tested in humans

I surveyed a bunch of choline users, using a little gimmick. Some of the forms of choline sold these days don’t cross the blood-brain barrier and shouldn’t have an effect, so they provide a sort of placebo control for more active forms of choline. In my survey, people who took piracetam with inactive forms of choline didn’t report any worse an experience than those who took the real thing.

This is the most famous and best-discussed interaction in the entire field of nootropics, and it’s on super-shaky ground. So trust me, the CerebralX people don’t have good evidence about the interactions of all twenty-one of their ridiculous substances.

I have to admit, I’m not confident in this part. Maybe psychiatry is wrong. Sometimes I wonder what would happen if we just throw five different antidepressants with five different mechanisms of action at somebody at once. Realistically, maybe this would involve some supplements: l-methylfolate, SAMe, tryptophan, turmeric, and a traditional SSRI. One day I want to try this on someone I know well enough to let me test things on them, but not so well I don't mind losing a friend when it all blows up in my face. Until then, keep in mind that anyone who says they bet a certain combination of things will produce a synergistic interaction is engaging in the wildest sort of speculation.

IV.

One more piece of evidence. The 2016 nootropics survey asked people to rate their experiences with 35 different individual substances, plus a branded combo pack ("AlphaBrain") of pretty good reputation. The AlphaBrain performed worse than *any* of the individual substances, including substances that were part of AlphaBrain!

This is of course a very weak result – it wasn't blinded, and maybe the survey responders have the same anti-branded-combo prejudice I do. But it at least suggests knowledgeable people in the nootropics community are really uncomfortable with this stuff.

90% of the people making branded combo nootropics are lying scum. A few, like TruBrain, seem like probably decent people trying to get it right – but are you confident you can tell them apart? And if you do manage to beat the odds and get something that's not a complete pharmacological mess, aren't you still just going to end up with an overpriced bundle of black boxes that won't provide you with useful information, and which, empirically, everyone hates?

If you're interested in nootropics, consider trying one substance at a time, very carefully, using something like examine.com to learn how to take it and what the possible side effects are. If you can, do what people like Gwern do and try it blind, mixing real pills with placebo pills over the space of a few weeks, so you can make sure it's a real effect. If you find something that does have a real effect on you, treat that knowledge as a hard-won victory. Then, if you want to go from there, tentatively add a second chemical and test that one in the same way. Do this, and you have some small sliver of a chance of doing more good than harm, at least in the short term.

But if you’re going to order a combination of twenty different things at homeopathic doses from somebody who thinks “GENIUS PILL” is a Scientific Term – well, I hope it works, because you need it.

Fear And Loathing At Effective Altruism Global 2017

> San Francisco in the middle sixties was a very special time and place to be a part of. Maybe it meant something. Maybe not, in the long run – but no explanation, no mix of words or music or memories can touch that sense of knowing that you were there and alive in that corner of time and the world.…There was a fantastic universal sense that whatever we were doing was right, that we were winning. And that, I think, was the handle—that sense of inevitable victory over the forces of Old and Evil.
>
> **— Hunter S. Thompson**

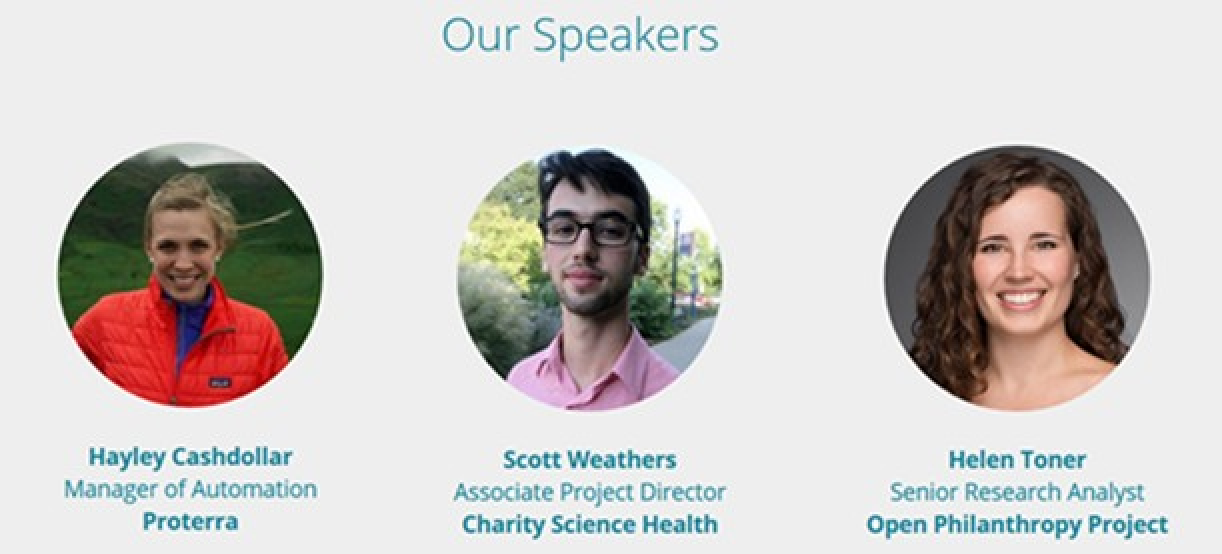
Effective altruism is the movement devoted to finding the highest-impact ways to help other people and the world. Philosopher William MacAskill described it as “doing for the pursuit of good what the Scientific Revolution did for the pursuit of truth”. They have an annual global conference to touch base and discuss strategy. This year it was in the Palace of Fine Arts in San Francisco, and I got a chance to check it out.

The lake-fringed monumental neoclassical architecture represents ‘utilitiarian distribution of limited resources’

The official conference theme was “Doing Good Together”. The official conference interaction style was “earnest”. The official conference effectiveness level was “very”. And it was impossible to walk away from some of the talks without being impressed.

Saturday afternoon there was a talk by some senior research analysts at GiveWell, which researches global development charities. They’ve evaluated dozens of organizations and moved $260 million to the most effective, mostly ones fighting malaria and parasitic infections. Next were other senior research analysts from the Open Philanthropy Project, who have done their own detailed effectiveness investigations and moved about $200 million.

The parade went on. More senior research analysts. More nine-digit sums of money. More organizations, all with names that kind of blended together. The Center for Effective Altruism. The Center For Effective Global Action. Raising For Effective Giving. Effecting Effective Effectiveness. Or maybe not, I think I was hallucinating pretty hard by the end.


Our Speakers
Hayley Cashdollar
Manager of Automation
Proterra
Scott Weathers
Associate Project Director
Charity Science Health
Helen Toner
Senior Research Analyst
Open Philanthropy Project

I figured the speaker named "Cashdollar" was a hallucination, but she's right there on the website

One of the breakout rooms had all-day career coaching sessions with 80,000 Hours (motto: "You have 80,000 hours in your career. Make the right career choices, and you can help solve the world's most pressing problems"). A steady stream of confused altruistic college students went in, chatted with a group of coaches, and came out knowing that the latest analyses show that management consulting is a useful path to build charity-leading-relevant skills, but practicing law and donating the money to charity is probably less useful than previously believed. In their inevitable effectiveness self-report, they record having convinced 188 people to change their career plans as of April 2015.

(I had been avoiding the 80,000 Hours people out of embarassment after their career analyses discovered that being a doctor was low-impact, but by bad luck I ended up sharing a ride home with one of them. I sheepishly introduced myself as a doctor, and he said “Oh, so am I!” I felt relieved until he added that he had stopped practicing medicine after he learned how low-impact it was, and gone to work for 80,000 Hours instead.)

The theater hosted a “fireside chat” with Bruce Friedrich, director of the pro-vegetarian Good Food Institute. I’d heard he was a former vice-president of PETA, so I went in with some stereotypes. They were wrong. Friedrich started by admitting that realistically most people are going to keep eating meat, and that yelling at them isn’t a very effective way to help animals. His tactic was to advance research into plant-based and vat-grown meat alternatives, which he predicted would taste identical to regular meat at a fraction of the cost, and which would put all existing factory farms out of business. Afterwards a bunch of us walked to a restaurant a few blocks down the street to taste an Impossible Burger, the vanguard of this brave new meatless future.

FOR THE LOVE OF MEAT

Since you're taking the time to read about burgers on a website, we're going to assume you really like burgers. We do too. Get ready to sink your teeth into all the juicy details about the Impossible Burger.

The people behind this ad are all PETA card-carrying vegetarians. And the future belongs to them, and they know it.

The whole conference was flawlessly managed, from laser-fast registration to polished-sounding speakers to friendly unobtrusive reminders to use the seventeen different apps that would keep track of your conference-related affairs for you. And the of course the venue, which really was amazing.

2
UNITED
STATES

The full-size model of the Apollo 11 lander represents 'utilitiarian distribution of limited resources'

But walk a little bit outside of the perfectly-scheduled talks, or linger in the common areas a little bit after the colorfully-arranged vegetarian lunches, and you run into the shadow side of all of this, the hidden underbelly of the movement.

William MacAskill wanted a "scientific revolution in doing good". But the Scientific Revolution progressed from "I wonder why apples fall down" to "huh, every particle is in an infinite number of places simultaneously, and also cats can be dead and alive at the same time". The effective altruists' revolution started with "I wonder if some charities work better than others". But even at this early stage, it's gotten to some pretty weird places.

I got to talk to some people from Wild Animal Suffering Research. They start with the standard EA animal rights argument – if you think animals have moral relevance, you can save zillions of them for almost no cost. A campaign for cage-free eggs, minimal in the grand scheme of things, got most major corporations to change their policies and gave two hundred million chickens an improved quality of life. But WASR points out that even this isn't the most neglected cause. There are up to a trillion reptiles, ten quintillion insects, and maybe a sextillion zooplankton. And as nasty as factory farms are, life in the state of nature is nasty, brutish, short, and prone to having parasitic wasps paralyze you so that their larvae can eat your organs from the inside out while you are still alive. WASR researches ways we can alleviate wild animal suffering, from euthanizing elderly elephants (probably not high-impact) to using more humane insecticides (recommended as an 'interim solution') to neutralizing predator species in order to relieve the suffering of prey (still has some thorny issues that need to be resolved).

Wild Animal Suffering Research was nowhere near the weirdest people at Effective Altruism Global.

I got to talk to people from the Qualia Research Institute, who point out that everyone else is missing something big: the hedonic treadmill. People have a certain baseline amount of happiness. Fix their problems, and they'll be happy for a while, then go back to baseline. The only solution is to hack consciousness directly, to figure out what exactly happiness is – unpack what we're looking for when we describe some mental states as having higher positive valence than others – and then add that on to every other mental state directly. This isn't quite the dreaded wireheading, the widely-feared technology that will make everyone so doped up on techno-super-heroin (or direct electrical stimulation of the brain’s pleasure centers) that they never do anything else. It’s a rewiring of the brain that creates a “perpetual but varied bliss” that “reengineers the network of transition probabilities between emotions” while retaining the capability to do economically useful work. Partly this last criteria is to prevent society from collapsing, but the ultimate goal is:

> …the possibility of a full-fledged qualia economy: when people have spare resources and are interested in new states of consciousness, anyone good at mining the state-space for precious gems will have an economic advantage. In principle the whole economy may eventually be entirely based on exploring the state-space of consciousness and trading information about the most valuable contents discovered doing so.

If you’re wondering whether these people’s research involves taking huge amounts of drugs – well, read their blog. My particular favorites are this essay on psychedelic cryptography ie creating messages that only people on certain drugs can read, and this essay on hyperbolic geometry in DMT experiences.

Qualia Research Institute
Qualia Research Institute
Qualia Research Institute

The guy on the right also works for MealSquares, a likely beneficiary of technology that hacks directly into people's brains and adds artificial positive valence to unpleasant experiences.

The Qualia Research Institute was nowhere near the weirdest people at Effective Altruism Global.

I got to talk to some people researching suffering in fundamental physics. The idea goes like this: the universe is really really big. So if suffering made up an important part of the structure of the universe, this would be so tremendously outrageously unconscionably bad that we can't even conceive of how bad it could be. So the most important cause might be to worry about whether fundamental physical particles are capable of suffering – and, if so, how to destroy physics. From their writeup:

> Speculative scenarios to change the long-run future of physics may dominate any concrete work to affect the welfare of intelligent computations — at least within the fraction of our brain's moral parliament that cares about fundamental physics. The main value (or disvalue) of

intelligence would be to explore physics further and seek out tricks by which its long-term character could be transformed. For instance, if false-vacuum decay did look beneficial with respect to reducing suffering in physics, civilization could wait until its lifetime was almost over anyway (letting those who want to create lots of happy and meaningful intelligent beings run their eudaimonic computations) and then try to ignite a false-vacuum decay for the benefit of the remainder of the universe (assuming this wouldn’t impinge on distant aliens whose time wasn’t yet up). Triggering such a decay might require extremely high-energy collisions — presumably more than a million times those found in current particle accelerators — but it might be possible. On the other hand, such decay may happen on its own within billions of years, suggesting little benefit to starting early relative to the cosmic scales at stake. In any case, I’m not suggesting vacuum decay as the solution — just that there may be many opportunities like it waiting to be found, and that these possibilities may dwarf anything else that happens with intelligent life.

2017
Effective

This talk was called 'Christians In Effective Altruism'. It recommended reaching out to churches, because deep down the EA movement and people of faith share the same core charitable values and beliefs.

The thing is, Lovecraft was right. He wrote:

> We live on a placid island of ignorance in the midst of black seas of infinity, and it was not meant that we should voyage far. The sciences, each straining in its own direction, have hitherto harmed us little; but some day the piecing together of dissociated knowledge will open up such terrifying vistas of reality, and of our frightful position therein, that we shall either go mad from the revelation or flee from the deadly light into the peace and safety of a new dark age.

Morality wasn't supposed to be like this. Most of the effective altruists I met were nonrealist utilitarians. They don’t believe in some objective moral law imposed by an outside Power. They just think that we should pursue our own human-parochial moral values effectively. If there was ever a recipe for a safe and milquetoast ethical system, that should be it. And yet once you start thinking about what morality is – really thinking, the kind where you try to use mathematical models and formal logic – it opens up into these dark eldritch vistas of infinities and contradictions. The effective altruists started out wanting to do good. And they did: whole nine-digit-sums worth of good, spreadsheets full of lives saved and diseases cured and disasters averted. But if you really want to understand what you’re doing – get past the point where you can catch falling apples, to the point where you have a complete theory of gravitation – you end up with something as remote from normal human tenderheartedness as the conference lunches were from normal human food.

VEGAN CAESAR SALAD with
VEGAN PARMESEAN
croutons contain gluten

Born too late to eat meat guilt-free, born too early to get the technology that hacks directly into my brain and adds artificial positive valence to unpleasant experiences.

But I worry I'm painting a misleading picture here. It isn't that effective altruism is divided into two types of people: the boring effective suits, and the wacky explorers of bizarre ethical theories. I mean, there's always going to be *some* division. But by and large these were the same people, or at least you couldn't predict who was who. They would go up and give a talk about curing river blindness in Nigeria, and then you'd catch them later and learn that they were worried that maybe the most effective thing was preventing synthetic biology from taking over the ecosystem. Or you would hear someone give their screed, think "what a weirdo", and then learn they were a Harvard professor who served on a bunch of Fortune 500 company boards. Maybe the right analogy would be physics. A lot of physicists work on practical things like solar panels and rechargeable batteries. A tiny minority work on stranger things like wormholes and alternate universes. But it’s not like these are two different factions in physics that hate each other. And every so often a solar panel engineer might look into the math behind alternate universes, or a wormhole theorist might have opinions on battery design. They’re doing really different stuff, but it’s within the same tradition.

The movement’s unofficial leader is William MacAskill. He’s a pretty typical overachiever – became an Oxford philosophy professor at age 28 (!), founded three successful non-profits, and goes around hobnobbing with rich people trying to get them to donate money (he himself has pledged to give away everything he earns above $36,000). I had always assumed he was just a random dignified suit-wearing person who was slightly exasperated at having to put up with the rest of the movement. But I got a chance to talk to him – just for a few minutes, before he had to run off and achieve something – and I was shocked at how much he knew about all the weirdest aspects of the community, and how protective he felt of them. And in his closing speech, he urged the attendees to “keep EA weird”, giving examples of times when seemingly bizarre ideas won out and became accepted by the mainstream.

Historical Reenactment Society

His PowerPoint slide for this topic was this picture of Eliezer Yudkowsky. Really. I'm not joking about this part.

If it were just the senior research analysts at their spreadsheets, we could dismiss them as the usual Ivy League lizard people and move on. If it were just the fringes ranting about cyber-neuro-metaphilosophy, we could dismiss them as loonies and forget about it. And if it were just the two groups, separate and doing their own thing, we could end *National Geographic*-style, intoning in our best David Attenborough voice that "Effective Altruism truly is a land of contrasts". But it's more than that. Some animating spirit gives rise to the whole thing, some unifying aesthetic that can switch to either pole and back again on a whim. After a lot of thought, I have only one guess about what it might be.

I think the effective altruists are genuinely good people.

Over lunch, a friend told me about his meeting with an EA philosopher who hadn’t been able to make it to the conference. This friend had met the philosopher, and as they were walking, the philosopher had stopped to pick up worms writhing on the sidewalk and put them back in the moist dirt.

And this story struck me, because I had taken a walk with one of the speakers earlier, and seen her do the same thing. She had been apologetic, said she knew it was a waste of her time and mine. She’d wondered if it was pathological, whether maybe she needed to be checked for obsessive compulsive disorder. But when I asked her whether she wanted to stop doing it, she’d thought about it a little, and then – finally – saved the worm.

And there was a story about the late great moral philosopher Derek Parfit, himself a member of the effective altruist movement. This is from Larissa MacFarquhar:

> As for his various eccentricities, I don’t think they add anything to an understanding of his philosophy, but I find him very moving as a person. When I was interviewing him for the first time, for instance, we were in the middle of a conversation and suddenly he burst into tears. It was completely unexpected, because we were not talking about anything emotional or personal, as I would define those things. I was quite startled, and as he cried I sat there rewinding our conversation in my head, trying to figure out what had upset him. Later, I asked him about it. It turned out that what had made him cry was the idea of suffering. We had been talking about suffering in the abstract. I found that very striking.

> Now, I don’t think any professional philosopher is going to make this mistake, but nonprofessionals might think that utilitarianism, for instance (Parfit is a utilitarian), or certain other philosophical ways of think about morality, are quite unemotional, quite calculating, quite cold; and so because as I am writing mostly for nonphilosophers, it seemed like a good corrective to know that for someone like Parfit these issues are extremely emotional, even in the abstract.
>
> The weird thing was that the same thing happened again with a philosophy graduate student whom I was interviewing some months later. Now you’re going to start thinking it’s me, but I was interviewing a philosophy graduate student who, like Parfit, had a very unemotional demeanor; we started talking about suffering in the abstract, and he burst into tears. I don’t quite know what to make of all this but I do think that insofar as one is interested in the relationship of ideas to people who think about them, and not just in the ideas themselves, those small events are moving and important.

I imagine some of those effective altruists, picking up worms, and I can see them here too. I can see them sitting down and crying at the idea of suffering, at allowing it to exist.

Larissa MacFarquhar says she doesn’t know what to make of this. I think I sort of do. I’m not much of an effective altruist – at least, I’ve managed to evade the 80,000 Hours coaches long enough to stay in medicine. But every so often, I can see the world as they have to. Where the very existence of suffering, any suffering at all, is an immense cosmic *wrongness*, an intolerable gash in the world, distressing and enraging. Where a single human lifetime seems frighteningly inadequate compared to the magnitude of the problem. Where all the normal interpersonal squabbles look trivial in the face of a colossal war against suffering itself, one that requires a soldier’s discipline and a general’s eye for strategy.

All of these Effecting Effective Effectiveness people don’t obsess over efficiency out of bloodlessness. They obsess because the struggle is so desperate, and the resources so few. Their efficiency is military efficiency. Their cooperation is military discipline. Their unity is the unity of people facing a common enemy. And they are winning. Very slowly, WWI trench-warfare-style. But they really are.

Sources and commentary here

And I write this partly because…well, it hasn’t been a great couple of weeks. The culture wars are reaching a fever pitch, protesters are getting run over by neo-Nazis, North Korea is threatening nuclear catastrophe. The world is a shitshow, nobody’s going to argue with that – and the people who are supposed to be leading us and telling us what to do are just about the shittiest of all.

And this is usually a pretty cynical blog. I’m cynical about academia and I’m cynical about medicine and goodness knows I’m cynical about politics. But Byron wrote:

> I have not loved the world, nor the world me
> But let us part fair foes; I do believe,
> Though I have found them not, that there may be
> Words which are things,—hopes which will not deceive,
> And virtues which are merciful, nor weave
> Snares for the failing: I would also deem
> O’er others’ griefs that some sincerely grieve;

> That two, or one, are almost what they seem,
> That goodness is no name, and happiness no dream.

This seems like a good time to remember that there are some really good people. And who knows? Maybe they'll win.

And one more story.

I got in a chat with one of the volunteers running the conference, and told him pretty much what I've said here: the effective altruists seemed like great people, and I felt kind of guilty for not doing more.

He responded with the official party line, the one I've so egregiously failed to push in this blog post. That effective altruism is a movement of ordinary people. That its yoke is mild and it accepts everyone. That not everyone has to be a vegan or a career researcher. That a commitment could be something more like just giving a couple of dollars to an effective-seeming charity, or taking the Giving What We Can pledge, or signing up for the online newsletter, or just going to an local effective altruism meetup group and contributing to discussions.

And I said yeah, but still, everyone here seems *so* committed to being a good person – and then here's me, constantly looking over my shoulder to stay one step ahead of the 80,000 Hours coaching team, so I can stay in my low-impact career that I happen to like.

And he said – no, absolutely, stay in your career right now. In fact, his philosophy was that you should do exactly what you feel like all the time, and not worry about altruism at all, because eventually you'll work through your own problems, and figure yourself out, and then you'll just naturally become an effective altruist.

And I tried to convince him that no, people weren't actually like that, practically *nobody* was like that, maybe he was like that but if so he might be the only person like that in the entire world. That there were billions of humans who just started selfish, and stayed selfish, and never declared total war against suffering itself *at all*.

And he didn't believe me, and we argued about it for ten minutes, and then we had to stop because we were missing the "Developing Intuition For The Importance Of Causes" workshop.

Rationality means believing what is true, not what makes you feel good. But the world has been really shitty this week, so I am going to give myself a one-time exemption. I am going to believe that convention volunteer's theory of humanity. *Credo quia absurdum; certum est, quia impossibile*. Everyone everywhere is just working through their problems. Once we figure ourselves out, we'll all become bodhisattvas and/or senior research analysts.

Partial Credit

The explorer lifted his hands toward the sky. “With the heavens as my witness,” he said, “if you do not release me, then I will respond by blotting out the sun…” He took a piece of paper out of his pocket, double-checked some numbers written on it “…starting exactly one minute from now.”

The savages snorted. The chieftain fiddled with the bone in his hair. “You no have power blot out sun,” he accused.

“My people possess great magic,” said the explorer. “And by threatening to eat me, you have incurred my wrath. So let me depart your country, or darkness shall fall over the land, starting…now!”

Nothing happened.

“Sun no get blot out,” said the savage chieftain. A few other savages nodded. “We think we stick with original plan of eat you.”

“The sun in *so* getting blotted out,” said the explorer. “Just a tiny corner at first. It’s hard to see. But gradually it’ll get bigger.”

One or two of the savages tried to stare at the sun, then averted their eyes after a moment.

“Definitely no blot out,” said the chieftain.

“You can’t see it because the remaining unblotted-out portion of the sun is too bright!” said the explorer.

“Not seem very good magic,” said the tribe’s shaman, joining in the discussion.

“Um. How about this. Do you have a cereal box?”

“What is cereal box?” asked the chieftain skeptically.

“Ah, frick. Um, bark. Do you have long and thin pieces of bark?” Some of the savages went into a hut, came out with some bark. “If you arrange them into a kind of box shape, and you cut a hole in that one there, and then you use it to block the other one, you…”

“Why we do this, again?” asked the shaman. “We hungry. We want eat you now.”

“Because,” said the explorer, “I’m trying to demonstrate that I’m blotting out the sun.”

“Me would think,” said shaman, “that if sun blotted out, maybe not need contraption made of pieces of bark in order to know.”

“I’m blotting it out really slowly! It’s too bright to look at directly!”

“Maybe you should blot out sun faster,” said the chieftain.

“I’M BLOTTING AS FAST AS I CAN!”

“Still not seem very good magic.”

“Oh, screw you, I’ll do it myself,” said the explorer, breaking out from among the warriors standing guard around his party. None of them moved to stop him as he sat down, took the pieces of bark, and propped them up against each other with sticks. He took a knife from his pocket and whittled a little hole into one of them. “See! The sun clearly has a little corner taken out of it.”

The savages all peered down warily. Finally, the shaman asked what all of them were thinking: “What supposed to be demonstrated by this?”

“Oh, for the love of God. It’s a pinhole projector. Normally the sunlight would come through this hole and illuminate a perfectly circular area on this other piece of bark here. But now, because I’m blotting out the sun with my magic, it’s producing this kind of crescent shape, with a bit taken out of the sun.”

“You able to blot out the sun with magic seem like overly complicated explanation for weird shape shadow,” said the shaman. “Maybe shadow made on bark always weird.”

“NO IT’S NOT,” said the explorer. “After I stop blotting out the sun, you’ll see it’s a normal circular shadow.”

“Okay,” said the chieftain. “Is good idea. You unblot sun now, we check for circular shadow, then re-blot sun again.”

“I’m not going to stop blotting out the sun just because you guys don’t understand optics!”

“Just unblot little bit, then reblot little bit. Not so hard.”

“Have you ever blotted out the sun before? No? Then stop telling me what’s hard or easy!”

“So when you stop blotting out sun?”

“After you release me!”

“So we only able to learn if releasing you necessary after you unblot sun, and you only unblot sun after we release you? Sound kind of like trick.”

“Look,” said the explorer. “I’m sorry about this. I really am. There are places north of here that are getting total sun-blotting-outs. If we were a few dozen miles away, this would be really impressive, I promise. But here, I’m only able to blot out the sun partially. Like, ninety percent. I just feel like, as demonstrations of power go, that’s still pretty impressive.”

“But definitely when ninety percent of sun blotted out, it big enough to notice, right?”

“Well…it will probably get darker. I think *maybe* the difference will be noticeable. And if you look at the pinhole projector…” He touched the contraption of bark and sticks, which promptly fell apart. He cursed and propped it back up again. “If you look at the pinhole projector, you’ll see that the part of the sun that’s missing is gradually increasing.”

The savages stared at the projector, dubiously.

“Look about same,” said the shaman.

“It’s not! Over the past few minutes, the ‘bite’ taken out of the sun has gradually gotten bigger!”

“Maybe should blot out whole sun,” said the chieftain. “Maybe then seem more obvious.”

“I’M NOT GOING TO BLOT OUT THE WHOLE SUN! Come on, can’t you tell it’s getting darker?”

The shaman squinted. “Maybe sort of look dark. Hard to tell.”

The savages started talking to each other. “Maybe look little darker than usual,” the chieftain concluded. “But maybe only because you give me suggestion.”

“Oh, come on,” said the explorer. “It’s clearly darker. Just let me go.”

The chieftain whispered something to the shaman. The shaman whispered something back to the chieftain. Finally, the chieftain turned to the explorer and nodded.

“Not sure if really darker or just power of suggestion. But you make sun come back, we let you go.”

The explorer gave a sigh of relief. He lifted his hands to the heavens. “In the name of the gods of my people,” he declared, “I command the sun to return to the sky!”

Nothing obvious happened. They waited a minute. Two minutes. Finally the chieftain shrugged. “Maybe little bit lighter,” he said. “Hard to tell.”

“Can I go?” asked the explorer.

The chieftain shrugged. Before he could change his mind, the explorer grabbed his pack and rushed out of the village.

Ten minutes later, he was back. The chieftain looked at him quizzically.

“Actually,” said the explorer, “I just saw the traffic on the road out of here. If the offer’s still open, I think I’d rather get eaten.”

Axiology, Morality, Law

I.

Philosopher Amanda Askell questions the practice of moral offsetting.

Offsetting is where you compensate for a bad thing by doing a good thing, then consider yourself even. For example, an environmentalist takes a carbon-belching plane flight, then pays to clean up the same amount of carbon she released.

This can be pretty attractive. If you're really environmentalist, but also really want to take a vacation to Europe, you could be pretty miserable not knowing whether your vacation is worth the cost to the planet. But if you can calculate that it would take about $70 to clean up more carbon than you release, that's such a small addition to the overall cost of the trip that you can sigh with relief and take the flight guilt-free.

Or use offsets instead of becoming vegetarian. An typical person's meat consumption averages 0.3 cows and 40 chickens per year. Animal Charity Evaluators believes that donating to a top animal charity this many animals' lives for less than $5; others note this number is totally wrong and made up. But it's hard to believe charities could be less cost-effective than just literally buying the animals; this would fix a year's meat consumption offset price at around $500. Would I pay between $5 and $500 a year not to have to be a vegetarian? You bet.

Askell is uncomfortable with this concept for the same reasons I was when I first heard about it. Can we kill an enemy, then offset it with enough money to save somebody else’s life? Smash other people’s property, then give someone else enough money to buy different property? Can Bill Gates nuke entire cities for fun, then build better cities somewhere else?

She concludes:

> There are a few different things that the harm-based ethicist could say in response to this, however. First, they could point out that as the immorality of the action increases, it becomes far less likely that performing this action and morally offsetting is the best option available, even out of those options that actualists would deem morally relevant. Second, it is very harmful to undermine social norms where people don’t behave immorally and compensate for it (imagine how terrible it would be to live in a world where this was acceptable). Third, it is – in expectation – bad to become the kind of person who offsets their moral harms. Such a person will usually have a much worse expected impact on the world than someone who strives to be as moral as they can be.
>
> I think that these are compelling reasons to think that, in the actual world, we are – at best – morally permitted to offset trivial immoral actions, but that more serious immoral actions are almost always not the sorts of things we can morally offset. But I also think that the fact that these arguments all depend on contingent features of the world should be concerning to those who defend harm-based views in ethics.

I think Askell gets the right answer here – you can offset carbon emissions but not city-nuking. And I think her reasoning sort of touches on some of the important considerations. But I also think there’s a much more elegant theory that gives clear answers to these kinds of questions, and which relieves some of my previous doubts about the offsetting idea.

II.

Everything below is taken from vague concepts philosophers talk about all the time, but which I can't find a single good online explanation of. I neither deserve credit for anything good about the ideas, nor can avoid blame for any mistakes or confusions in the phrasing. That having been said: consider the distinction between axiology, morality, and law.

Axiology is the study of what's good. If you want to get all reductive, think of it as comparing the values of world-states. A world-state where everybody is happy seems better than a world-state where everybody is sad. A world-state with lots of beautiful art is better than a world-state containing only featureless concrete cubes. Maybe some people think a world-state full of people living in harmony with nature is better than a world-state full of gleaming domed cities, and other people believe the opposite; when they debate the point, they're debating axiology.

Morality is the study of what the right thing to do is. If someone says "don't murder", they're making a moral commandment. If someone says "Pirating music is wrong", they're making a moral claim. Maybe some people believe you should pull the lever on the trolley problem, and other people believe you shouldn't; when they debate the point, they're debating morality.

(this definition elides a complicated distinction between individual conscience and social pressure; fixing that would be really hard and I'm going to keep eliding it)

Law is – oh, come on, you know this one. If someone says “Don’t go above the speed limit, there’s a cop car behind that corner”, that’s law. If someone says “my state doesn’t allow recreational marijuana, but it will next year”, that’s law too. Maybe some people believe that zoning restrictions should ban skyscrapers in historic areas, and other people believe they shouldn’t; when they debate the point, they’re debating law.

These three concepts are pretty similar; they’re all about some vague sense of what is or isn’t desirable. But most societies stop short of making them exactly the same. Only the purest act-utilitarianesque consequentialists say that axiology exactly equals morality, and I’m not sure there *is* anybody quite that pure. And only the harshest of Puritans try to legislate the state law to be exactly identical to the moral one. To bridge the whole distance – to directly connect axiology to law and make it illegal to do anything other than the most utility-maximizing action at any given time – is such a mind-bogglingly bad idea that I don’t think anyone’s even considered it in all of human history.

These concepts stay separate because they each make different compromises between goodness, implementation, and coordination.

One example: axiology can’t distinguish between murdering your annoying neighbor vs. not donating money to save a child dying of parasitic worms in Uganda. To axiology, they’re both just one life snuffed out of the world before its time. If you forced it to draw some distinction, it would probably decide that saving the child dying of parasitic worms was more important, since they have a longer potential future lifespan.

But morality absolutely draws this distinction: it says not-murdering is obligatory, but donating money to Uganda is supererogatory. Even utilitarians who deny this distinction in principle will use it in everyday life: if their friend was considering not donating money, they would be a little upset; if their friend was considering murder, they would be horrified. If they themselves forgot to donate money, they'd feel a little bad; if they committed murder in the heat of passion, they'd feel awful.

Another example: Donating 10% of your income to charity is a moral rule. Axiology says "Why not donate all of it?", Law says "You won't get in trouble even if you don't donate any of it", but at the moral level we set a clear and practical rule that meshes with our motivational system and makes the donation happen.

Another example: "Don't have sex with someone who isn't mature enough to consent" is a good moral rule. But it doesn't make a good legal rule; we don't trust police officers and judges to fairly determine whether someone's mature enough in each individual case. A society which enshrined this rule in law would be one where you were afraid to have sex with anyone at all – because no matter what your partner's maturity level, some police officer might say your partner seemed immature *to them* and drag you away. On the other hand, elites could have sex with arbitrarily young people, expecting police and judges to take their side.

So the state replaces this moral rule with the legal rule "don't have sex with anyone below age 18". Everyone knows this rule doesn't perfectly capture reality – there's no significant difference between 17.99-year-olds and 18.01-year-olds. It's a useful hack that waters down the moral rule in order to make it more implementable. Realistically it gets things wrong sometimes; sometimes it will incorrectly tell people not to have sex with perfectly mature 17.99-year-olds, and other times it will incorrectly excuse sex with immature 18.01-year-olds. But this beats the alternative, where police have the power to break up any relationship they don’t like, and where everyone has to argue with everybody else about whether their relationships are okay or not.

A final example: axiology tells us a world without alcohol would be better than our current world: ending alcoholism could avert millions of deaths, illnesses, crimes, and abusive relationships. Morality only tells us that we should be careful drinking and stop if we find ourselves becoming alcoholic or ruining our relationships. And the law protests that it tried banning alcohol once, but it turned out to be unenforceable and gave too many new opportunities to organized crime, so it’s going to stay out of this one except to say you shouldn’t drink and drive.

So fundamentally, what is the difference between axiology, morality, and law?

Axiology is just our beliefs about what is good. If you defy axiology, you make the world worse.

At least from a rule-utilitarianesque perspective, morality is an attempt to triage the infinite demands of axiology, in order to make them implementable by specific people living in specific communities. It makes assumptions like “people have limited ability to predict the outcome of their actions”, “people are only going to do a certain amount and then get tired”, and “people do better with bright-line rules than with vague gradients of goodness”. It also admits that it’s important that everyone living in a community is on at least kind of the same page morally, both in order to create social pressure to follow the rules, and in order to build the social trust that allows the community to keep functioning. If you defy morality, you still make the world worse. And you feel guilty. And you betray the social trust that lets your community function smoothly. And you get ostracized as a bad person.

Law is an attempt to formalize the complicated demands of morality, in order to make them implementable by a state with police officers and law courts. It makes assumptions like “people’s vague intuitive moral judgments can sometimes give different results on the same case”, “sometimes police officers and legislators are corrupt or wrong”, and “we need to balance the benefits of laws against the cost of enforcing them”. It also tries to avert civil disorder or civil war by assuring everybody that it’s in their best interests to appeal to a fair universal law code rather than try to solve their disagreements directly. If you defy law, you still get all the problems with defying axiology and morality. And you make your country less peaceful and stable. And you go to jail.

In a healthy situation, each of these systems reinforces and promotes the other. Morality helps you implement axiology from your limited human perspective, but also helps prevent you from feeling guilty for not being God and not being able to save everybody. The law helps enforce the most important moral and axiological rules but also leaves people free enough to use their own best judgment on how to pursue the others. And axiology and morality help resolve disputes about what the law should be, and then lend the support of the community, the church, and the individual conscience in keeping people law-abiding.

In these healthy situations, the universally-agreed priority is that law trumps morality, and morality trumps axiology. First, because you can’t keep your obligations to your community from jail, and you can’t work to make the world a better place when you’re a universally-loathed social outcast. But also, because you can’t work to build strong communities and relationships in the middle of a civil war, and you can’t work to make the world a better place from within a low-trust defect-defect equilibrium. But also, because in a just society, axiology *wants* you to be moral (because morality is just a more-effective implementation of axiology), and morality wants you to be law-abiding (because law is just a more-effective way of coordinating morality). So first you do your legal duty, then your moral duty, and then if you have energy left over, you try to make the world a better place.

(Katja Grace has some really good writing on this kind of stuff here)

In unhealthy situations, you can get all sorts of weird conflicts. Most “moral dilemmas” are philosophers trying to create perverse situations where axiology and morality give opposite answers. For example, the fat man version of the trolley problem sets axiology (“it’s obviously better to have a world where one person dies than a world where five people die”) against morality (“it’s a useful rule that people generally shouldn’t push other people to their deaths”). And when morality and state law disagree, you get various acts of civil disobedience, from people hiding Jews from the Nazis all the way down to Kentucky clerks refusing to perform gay marriages.

I don’t have any special insight into these. My intuition (most authoritative source! is never wrong!) says that we should be very careful reversing the usual law-trumps-morality-trumps-axiology order, since the whole point of having more than one system is that we *expect* the systems to disagree and we want to suppress those disagreements in order to solve important implementation and coordination problems. But I also can’t deny that for enough gain, I’d reverse the order in a heartbeat. If someone told me that by breaking a promise to my friend (morality) I could cure all cancer forever (axiology), then f@$k my friend, and f@$k whatever social trust or community cohesion would be lost by the transaction.

III.

With this framework, we can propose a clearer answer to the moral offsetting problem: you can offset axiology, but not morality.

Emitting carbon doesn’t violate any moral law at all (in the stricter sense of morality used above). It does make the world a worse place. But there’s no unspoken social agreement not to do it, it doesn’t violate any codes, nobody’s going to lose trust in you because of it, you’re not making the community any less cohesive. If you make the world a worse place, it’s perfectly fine to compensate by making the world a better place. So pay to clean up some carbon, or donate to help children in Uganda with parasitic worms, or whatever.

Eating meat doesn’t violate any moral laws either. Again, it makes the world a worse place. But there aren’t any bonds of trust between humans and animals, nobody’s expecting you not to eat meat, there aren’t any written or unwritten codes saying you shouldn’t. So eat the meat and offset it by making the world better in some other way.

(the strongest counterargument I can think of here is that you’re not betraying animals, but you might be betraying your fellow animals-rights-activists! That is, if they’re working to establish a social norm against meat-eating, the sort of thing where being spotted with a cheeseburger on your plate produces the same level of horror as being spotted holding a bloody knife above a dead body, then your meat-eating is interfering with their ability to establish that norm, and this is a problem that requires more than just offsetting the cost of the meat involved)

Murdering someone does violate a moral law. The problem with murder isn't just that it creates a world in which one extra person is dead. If that's all we cared about, murdering would be no worse than failing to donate money to cure tropical diseases, which also kills people.

(and the problem isn't just that it has some knock-on effects in terms of making people afraid of crime, or decreasing the level of social trust by 23.5 social-trustons, or whatever. If that were all, you could do what 90% of you are probably already thinking – "Just as we're offsetting the murder by donating enough money to hospitals to save one extra life, couldn't we offset the social costs by donating enough money to community centers to create 23.5 extra social-trustons?" There's probably something like that which would work, but along with everything else we're crossing a Schelling fence, breaking rules, and weakening the whole moral edifice. The cost isn't infinite, but it's pretty hard to calculate. If we're positing some ridiculous offset that obviously outweighs any possible cost – maybe go back to the example of curing all cancer forever – then whatever, go ahead. If it's anything less than that, be careful. I *like* the metaphor of these three systems being on three separate tiers – rather than two Morality Points being worth one Axiology Point, or whatever – exactly because we don't really know how to interconvert them)

This is more precise than Askell's claim that we can offset "trivial immoral actions" but not "more serious" ones. For example, suppose I built an entire power plant that emitted one million tons of carbon per year. Sounds pretty serious! But if I offset that with environmental donations or projects that prevented 1.1 million tons of carbon somewhere else, I can't imagine anyone having a problem with it.

On the other hand, consider spitting in a stranger's face. In the grand scheme of things, this isn't so serious – certainly not as serious as emitting a million tons of carbon. But I would feel uncomfortable offsetting this with a donation to my local Prevent Others From Spitting In Strangers' Face fund, even if the fund worked.

Askell gave a talk where she used the example of giving your sister a paper cut, and then offsetting that by devoting your entire life to helping the world and working for justice and saving literally thousands of people. Pretty much everyone agrees that's okay. I guess I agree it's okay. Heck, I guess I would agree that murdering someone in order to cure cancer forever would be okay. But now we're just getting into the thing where you bulldoze through moral uncertainty by making the numbers so big that it's impossible to be uncertain about them. Sure. You can do that. I'd be less happy about giving my sister a paper cut, and then offsetting by preventing one paper cut somewhere else. But that seems to be the best analogy to the "emit one ton of carbon, prevent one ton of carbon" offsetting we've been talking about elsewhere.

I realize all this is sort of hand-wavy – more of a "here's one possible way we could look at these things" rather than "here's something I have a lot of evidence is true". But everyone – you, me, Amanda Askell, society – seems to want a system that tells us to offset carbon but not murder, and when we find such a system I think it's worth taking it seriously.

My IRB Nightmare

September 2014

There’s a screening test for bipolar disorder. You ask patients a bunch of things like “Do you ever feel really happy, then really sad?”. If they say ‘yes’ to enough of these questions, you start to worry.

Some psychiatrists love this test. I hate it. Patients will say “Yes, that absolutely describes me!” and someone will diagnose them with bipolar disorder. Then if you ask what they meant, they’d say something like “Once my local football team made it to the Super Bowl and I was really happy, but then they lost and I was really sad.” I don’t even want to tell you how many people get diagnosed bipolar because of stuff like this.

There was a study that supposedly proved this test worked. But parts of it confused me, and it was done on a totally different population that didn’t generalize to hospital inpatients. Also, it said in big letters THIS IS JUST A SCREENING TEST IT IS NOT INTENDED FOR DIAGNOSIS, and everyone was using it for diagnosis.

So I complained to some sympathetic doctors and professors, and they asked “Why not do a study?”

Why *not* do a study? Why not join the great tradition of scientists, going back to Galileo and Newton, and make my mark on the world? Why not replace my griping about bipolar screening with an experiment about bipolar screening, an experiment done to the highest standards of the empirical tradition, one that would throw the entire weight of the scientific establishment behind my complaint?

I’d been writing about science for so long, even doing my own informal experiments, why not move on to join the big leagues?

For (it would turn out) a whole host of excellent reasons that I was about to learn.

A spring in my step, I journeyed to my hospital’s Research Department, hidden in a corner office just outside the orthopaedic ward. It was locked, as always. After enough knocking, a lady finally opened the door and motioned for me to sit down at a paperwork-filled desk.

“I want to do a study,” I said.

She looked skeptical. “Have you done the Pre-Study Training?”

I had to admit I hadn’t, so off I went. The training was several hours of videos about how the Nazis had done unethical human experiments. Then after World War II, everybody met up and decided to only do ethical human experiments from then on. And the most important part of being ethical was to have all experiments monitored by an Institutional Review Board (IRB) made of important people who could check whether experiments were ethical or not. I dutifully parroted all this back on the post-test (“Blindly trusting authority to make our ethical decisions for us is the *best* way to separate ourselves from the Nazis!”) and received my Study Investigator Certification.

I went back to the corner office, Study Investigator Certification in hand.

“I want to do a study,” I said.

The lady still looked skeptical. “Do you have a Principal Investigator?”

Mere resident doctors weren’t allowed to do studies on their own. They would probably screw up and start building concentration camps or something. They needed an attending (high-ranking doctor) to sign on as Principal Investigator before the IRB would deign to hear their case.

I knew exactly how to handle this: one by one, I sought out the laziest attendings in the hospital and asked “Hey, would you like to have your name on a study as Principal Investigator for free while I do all the actual work?” Yet one by one, all of the doctors refused, as if I was offering them some kind of plague basket full of vermin. It was the weirdest thing.

Finally, there was only one doctor left – Dr. W, the hardest-working attending I knew, the one who out of some weird masochistic impulse took on every single project anyone asked of him and micromanaged it to perfection, the one who every psychiatrist in the whole hospital (including himself) had diagnosed with obsessive-compulsive personality disorder.

“Sure Scott,” he told me. “I’d be happy to serve as your Principal Investigator”.

A feeling of dread in my stomach, I walked back to the tiny corner office.

“I want to do a study,” I said.

The lady still looked skeptical. “Have you completed the New Study Application?” She gestured to one of the stacks of paperwork filling the room.

It started with a section on my research question. Next was a section on my proposed methodology. A section on possible safety risks. A section on recruitment. A section on consent. A section on…wow. Surely this can’t *all* be the New Study Application? Maybe I accidentally picked up the Found A New Hospital Application?

I asked the lady who worked in the tiny corner office whether, since I was just going to be asking bipolar people whether they ever felt happy and then sad, maybe I could get the short version of the New Study Application?

She told me that *was* the short version.

“But it’s twenty-two pages!”

“You haven’t done any studies before, have you?”

Rather than confess my naivete, I started filling out the twenty-two pages of paperwork. It started by asking about our study design, which was simple: by happy coincidence, I was assigned to Dr. W’s inpatient team for the next three months. When we got patients, I would give them the bipolar screening exam and record the results. Then Dr. W. would conduct a full clinical interview and formally assess them. We’d compare notes and see how often the screening test results matched Dr. W’s expert diagnosis. We usually got about twenty new patients a week; if half of them were willing and able to join our study, we should be able to gather about a hundred data points over the next three months. It was going to be easy-peasy.

That was the first ten pages or so of the Application. The rest was increasingly bizarre questions such as “Will any organs be removed from participants during this study?” (Look, I promise, I’m not a Nazi).

And: “Will prisoners be used in the study?” (COME ON, I ALREADY SAID I WASN’T A NAZI).

And: “What will you do if a participant dies during this research?” (If somebody dies while I’m asking them whether they sometimes feel happy and then sad, I really can’t even promise so much as “not freaking out”, let alone any sort of dignified research procedure).

And more questions, all along the same lines. I double-dog swore to give everybody really, really good consent forms. I tried my best to write a list of the risks participants were taking upon themselves (mostly getting paper cuts on the consent forms). I argued that these compared favorably to the benefits (maybe doctors will stop giving people strong psychiatric medications just because their football team made the Super Bowl).

When I was done, I went back to the corner office and submitted everything to the Institutional Review Board. Then I sat back and hoped for the best. Like an idiot.

October 2014

The big day arrived. The IRB debated the merits of my study, examined the risks, and…sent me a letter pointing out several irregularities in my consent forms.

IRREGULARITY #1: Consent forms traditionally included the name of the study in big letters where the patient could see it before signing. Mine didn’t. Why not?

Well, because in questionnaire-based psychological research, you *never* tell the patient what you’re looking for before they fill out the questionnaire. That’s like Methods 101. The name of my study was “Validity Of A Screening

Instrument For Bipolar Disorder”. Tell the patient it’s a study about bipolar disorder, and the gig is up.

The IRB listened patiently to my explanation, then told me that this was not a legitimate reason not to put the name of the study in big letters on the consent form. Putting the name of the study on the consent form was important. You know who *else* didn’t put the name of the study on his consent forms? *Hitler.*

IRREGULARITY #2: Consent forms traditionally included a paragraph about the possible risks of the study and a justification for why we believed that the benefits were worth the risks. Everyone else included a paragraph about this on our consent forms, and read it to their patients before getting their consent. We didn’t have one. Why not?

Well, for one thing, because all we were doing was asking them whether they felt happy and then sad sometimes. This is the sort of thing that goes on every day in a psychiatric hospital. Heck, the other psychiatrists were using this same screening test, except *for real*, and they never had to worry about whether it had risks. In the grand scheme of things, this just wasn’t a very risky procedure.

Also, psychiatric patients are sometimes…how can I put this nicely?…a little paranoid. Sometimes you can offer them breakfast and they’ll accuse you of trying to poison them. I had no illusions that I would get every single patient to consent to this study, but I felt like I could at least avoid handing them a paper saying “BY THE WAY, THIS STUDY IS FULL OF RISKS”.

The IRB listened patiently to my explanation, then told me that this was not a legitimate reason not to have a paragraph about risks. We should figure out some risks, then write a paragraph explaining how those were definitely the risks and we took them very seriously. The other psychiatrists who used this test every day didn’t have to do that *because they weren’t running a study*.

IRREGULARITY #3: Signatures are traditionally in pen. But we said our patients would sign in pencil. Why?

Well, because psychiatric patients aren’t allowed to have pens in case they stab themselves with them. I don’t get why stabbing yourself with a pencil is any less of a problem, but the rules are the rules. We asked the hospital administration for a one-time exemption, to let our patients have pens just long enough to sign the consent form. Hospital administration said absolutely not, and they didn’t care if this sabotaged our entire study, it was pencil or nothing.

The IRB listened patiently to all this, then said that it had to be in pen. You know who *else* had people sign consent forms in pencil…?

I’m definitely not saying that these were the only three issues the IRB sprung on Dr. W and me. I’m saying these are a *representative sample*. I’m saying I spent several weeks relaying increasingly annoyed emails and memos from myself to Dr. W to the IRB to the lady in the corner office to the IRB again. I began to come home later in the evening. My relationships suffered. I started having dreams about being attacked by giant consent forms filled out in pencil.

I was about ready to give up at this point, but Dr. W insisted on combing through various regulations and talking to various people, until he discovered some arcane rule that certain very safe studies with practically no risk were allowed to use an “expedited consent form”, which was a lot like a normal consent form but didn't need to have things like the name of the study on it. Faced with someone even more obsessive and bureaucratic than they were, the IRB backed down and gave us preliminary permission to start our study.

The next morning, screening questionnaire in hand, I showed up at the hospital and hoped for the best. Like an idiot.

November 2014

Things progressed slowly. It turns out a lot of psychiatric inpatients are either depressed, agitated, violent, or out of touch with reality, and none of these are really conducive to wanting to participate in studies. A few of them already delusionally thought we were doing experiments on them, and got confused when we suddenly asked them to consent. Several of them made it clear that they hated us and wanted to thwart us in any way possible. After a week, I only had three data points, instead of the ten I'd been banking on.

"Data points" makes it sound abstract. It wasn't. I had hoped to put the results in the patients' easily accessible online chart, *the same place everyone else put the results of the exact same bipolar screening test* when they did it for real. They would put it in a section marked TEST RESULTS, which was there to have a secure place where you could put test results, and where everybody's secure test results were kept.

The IRB would have none of this. Study data are Confidential and need to be kept Secure. Never mind that all the patients' *other* secure test results were on the online chart. Never mind that the online chart contains all sorts of stuff about the patients’ diagnoses, medications, hopes and fears, and even (remember, this is a psych hospital) secret fetishes and sexual perversions. Study data needed to be encrypted, then kept in a Study Binder in a locked drawer in a locked room that nobody except the study investigators had access to.

The first problem was that nobody wanted to give us a locked room that nobody except us had access to. There was a sort of All Purpose Psychiatry Paperwork room, but the janitors went in to clean it out every so often, and apparently this made it unacceptable. Hospitals aren’t exactly drowning in spare rooms that not even janitors can get into. Finally Dr. W grudgingly agreed to keep it in his office. This frequently meant I couldn’t access any of the study material because Dr. W was having important meetings that couldn’t be interrupted by a resident barging into his office to rummage in his locked cabinets.

But whatever. The bigger problem was the encryption. There was a very specific way we had to do it. We would have a Results Log, that said things like “Patient 1 got a score of 11.5 on the test”. And then we’d have a Secret Patient Log, which would say things like “Patient 1 = Bob Johnson from Oakburg.” That way nobody could steal our results and figure out that Bob was sometimes happy, then sad.

(meanwhile, all of Bob’s actual diagnoses, sexual fetishes, etc were in the easily-accessible secure online chart that we were banned from using)

And then – I swear this is true – we had to keep the Results Log and the Secret Patient Log right next to each other in the study binder in the locked drawer in the locked room.

I wasn’t sure I was understanding this part right, so I asked Dr. W whether it made sense, to him, that we put a lot of effort writing our results in code, and then put the key to the code in the same place as the enciphered text. He cheerfully agreed this made no sense, but said we had to do it or else our study would fail an audit and get shut down.

January 2015

I’d planned to get a hundred data points in three months. Thanks to constant bureaucratic hurdles, plus patients being less cooperative than I expected, I had about twenty-five. Now I was finishing my rotation on Dr. W’s team and going to a clinic far away. What now?

A bunch of newbies were going to be working with Dr. W for the next three months. I hunted them down and threatened and begged them until one of them agreed to keep giving patients the bipolar screening test in exchange for being named as a co-author. Disaster averted, I thought. Like an idiot.

Somehow news of this arrangement reached the lady in the corner office, who asked whether the new investigator had completed her Pre-Study Training. I protested that she wasn’t designing the study, she wasn’t conducting any analyses, all she was doing was asking her patients the same questions that she would be asking them anyway as part of her job for the next three months. The only difference was that she was recording them and giving them to me.

The lady in the corner office wasn’t impressed. You know who *else* hadn’t thought his lackeys needed to take courses in research ethics?

So the poor newbie took a course on how Nazis were bad. Now she could help with the study, right?

Wrong. We needed to submit a New Investigator Form to the IRB and wait for their approval.

Two and a half months later, the IRB returned their response: Newbie was good to go. She collected data for the remaining two weeks of her rotation with Dr. W before being sent off to another clinic just like I was.

July 2015

Dr. W and I planned ahead. We had figured out which newbies would be coming in to work for Dr. W three months ahead of time, and gotten them through the don’t-be-a-Nazi course and the IRB approval process just in time for them to start their rotation. Success!

Unfortunately, we received another communication from the IRB. Apparently we were allowed to use the expedited consent form to get consent for our *study*, but not to get consent to *access protected health information*. That one required a whole different consent form, list-of-risks and all. We were right back where we’d started from.

I made my case to the Board. My case was: we’re not looking at any protected health information, f@#k you.

The Board answered that we were accessing the patient’s final diagnosis. It said right in the protocol, we were giving them the screening test, then comparing it to the patient’s final diagnosis. “Psychiatric diagnosis” sure *sounds* like protected health information.

I said no, you don’t understand, we’re the psychiatrists. Dr. W is the one making the final diagnosis. When I’m on Dr.

W’s team, I’m in the room when he does the diagnostic interview, half the time I’m the one who types the final diagnosis into the chart. These are *our patients*.

The Board said this didn’t matter. We, as the patient’s doctors, would make the diagnosis and write it down on the chart. But we (as study investigators) needed a full signed consent form before we were allowed to access the diagnosis we had just made.

I said wait, you’re telling us we have to do this whole bureaucratic rigamarole with all of these uncooperative patients before we’re allowed to see something we wrote ourselves?

The Board said yes, exactly.

I don’t remember this part very well, except that I think I half-heartedly trained whichever poor newbie we were using that month in how to take a Protected Health Information Consent on special Protected Health Information Consent Forms, and she nodded her head and said she understood. I think I had kind of clocked out at this point. I was going off to work all the way over in a different town for a year, and I was just sort of desperately hoping that Dr. W and various newbies would take care of things on their own and then in a year when I came back to the hospital I would have a beautiful pile of well-sorted data to analyze. Surely trained doctors would be able to ask simple questions from a screening exam on their own without supervision, I thought. Like an idiot.

July 2016

I returned to my base hospital after a year doing outpatient work in another town. I felt energized, well-rested, and optimistic that the bipolar screening study I had founded so long ago had been prospering in my absence.

Obviously nothing remotely resembling this had happened. Dr. W had vaguely hoped that I was taking care of it. I had vaguely hoped that Dr. W was taking care of it. The various newbies whom we had strategically enlisted had either forgotten about it, half-heartedly screened one or two patients before getting bored, or else mixed up the growing pile of consent forms and releases and logs so thoroughly that we would have to throw out all their work. It had been a year and a half since the study had started, and we had 40 good data points.

The good news was that I was back in town and I could go back to screening patients myself again. Also, we had some particularly enthusiastic newbies who seemed really interested in helping out and getting things right. Over the next three months, our sample size shot up, first to 50, then to 60, finally to 70. Our goal of 100 was almost in sight. The worst was finally behind me, I hoped. Like an idiot.

November 2016

I got an email saying our study was going to be audited.

It was nothing personal. Some higher-ups in the nationwide hospital system had decided to audit every study in our hospital. We were to gather all our records, submit them to the auditor, and hope for the best.

Dr. W, who was obsessive-compulsive at the best of times, became unbearable. We got into late-night fights over the number of dividers in the study binder. We hunted down every piece of paper that had ever been associated with anyone involved in the study in any way, and almost came to blows over how to organize it. I started working really late. My girlfriend began to doubt I actually existed.

The worst part was all the stuff the newbies had done. Some of them would have the consent sheets numbered in the upper left-hand-corner instead of the upper-right-hand corner. Others would have written the patient name down on the Results Log instead of the Secret Code Log right next to it. One even wrote something in green pen on a formal study document. It was hopeless. Finally we just decided to throw away all their data and pretend it had never existed.

With that decision made, our work actually started to look pretty good. As bad as it was working for an obsessive-compulsive boss in an insane bureaucracy, at least it had the advantage that – when nitpicking push came to ridiculous shove – you were going to be super-ready to be audited. I hoped. Like an idiot.

December 2016

The auditor found twenty-seven infractions.

She was very apologetic about it. She said that was actually a pretty good number of infractions for a study this size, that we were actually doing pretty well compared to a lot of the studies she'd seen. She said she absolutely wasn't going to shut us down, she wasn't even going to censure us. She just wanted us to make twenty-seven changes to our study and get IRB approval for each of them.

I kept the audit report as a souvenier. I have it in front of me now. Here's an example infraction:

> The data and safety monitoring plan consists of 'the Principal Investigator will randomly check data integrity'. This is a prospective study with a vulnerable

> group (mental illness, likely to have diminished capacity, likely to be low income) and, as such, would warrant a more rigorous monitoring plan than what is stated above. In addition to the above, a more adequate plan for this study would also include review of the protocol at regular intervals, on-going checking of any participant complaints or difficulties with the study, monitoring that the approved data variables are the only ones being collected, regular study team meetings to discuss progress and any deviations or unexpected problems. Team meetings help to assure participant protections, adherence to the protocol. Having an adequate monitoring plan is a federal requirement for the approval of a study. See Regulation 45 CFR 46.111 Criteria For IRB Approval Of Research. IRB Policy: PI Qualifications And Responsibility In Conducting Research. Please revise the protocol via a protocol revision request form. Recommend that periodic meetings with the research team occur and be documented.

Among my favorite other infractions:

1. The protocol said we would stop giving the screening exam to patients if they became violent, but failed to rigorously define “violent”.

2. We still weren’t educating our patients enough about “Alternatives To Participating In This Study”. The auditor agreed that the only alternative was “not participating in this study”, but said that we had to tell every patient that, then document that we’d done so.

3. The consent forms were still getting signed in pencil. We are never going to live this one down. If I live to be a hundred, representatives from the IRB are going to break

into my deathbed room and shout “YOU LET PEOPLE SIGN CONSENT FORMS IN PENCIL, HOW CAN YOU JUSTIFY THAT?!”

4. The woman in the corner office who kept insisting everybody take the Pre-Study Training…hadn’t taken the Pre-Study Training, and was therefore unqualified to be our liaison with the IRB. I swear I am not making this up.

Faced with submitting twenty-seven new pieces of paperwork to correct our twenty-seven infractions, Dr. W and I gave up. We shredded the patient data and the Secret Code Log. We told all the newbies they could give up and go home. We submitted the Project Closure Form to the woman in the corner office (who as far as I know still hasn’t completed her Pre-Study Training). We told the IRB that they had won, fair and square; we surrendered unconditionally.

They didn’t seem the least bit surprised.

August 2017

I’ve been sitting on this story for a year. I thought it was unwise to publish it while I worked for the hospital in question. I still think it’s a great hospital, that it delivers top-notch care, that it has amazing doctors, that it has a really good residency program, and even that the Research Department did everything it could to help me given the legal and regulatory constraints. I don’t want this to reflect badly on them in any way. I just thought it was wise to wait a year.

During that year, Dr. W and I worked together on two less ambitious studies, carefully designed not to require any contact with the IRB. One was a case report, the other used publicly available data.

They won 1st and 2nd prize at a regional research competition. I got some nice certificates for my wall and a little prize money. I went on to present one of them at the national meeting of the American Psychiatric Association, a friend helped me write it up formally, and it was recently accepted for publication by a medium-tier journal.

I say this not to boast, but to protest that I'm not as much of a loser as my story probably makes me sound. I'm capable of doing research, I think I have something to contribute to Science. I still think the bipolar screening test is inappropriate for inpatient diagnosis, and I still think that patients are being harmed by people's reliance on it. I still think somebody should look into it and publish the results.

I'm just saying it's not going to be me. I am *done* with research. People keep asking me "You seem really into science, why don't you become a researcher?" Well…

I feel like a study that realistically could have been done by one person in a couple of hours got dragged out into hundreds of hours of paperwork hell for an entire team of miserable doctors. I think its scientific integrity was screwed up by stupid requirements like the one about breaking blinding, and the patients involved were put through unnecessary trouble by being forced to sign endless consent forms screaming to them about nonexistent risks.

I feel like I was dragged almost to the point of needing to be in a psychiatric hospital myself, while my colleagues who just *used* the bipolar screening test – without making the mistake of trying to check if it works – continue to do so without anybody questioning them or giving them the slightest bit of aggravation.

I feel like some scientists do amazingly crappy studies that couldn't possibly prove anything, but get away with it because they have a well-funded team of clerks and secretaries who handle the paperwork for them. And that I, who was trying to do everything right, got ground down with so many pointless security-theater-style regulations that I'm never going to be able to do the research I would need to show they're wrong.

In the past year or so, I've been gratified to learn some other people are thinking along the same lines. Somebody linked me to The Censor's Hand, a book by a law/medicine professor at the University of Michigan. A summary from a review:

> Schneider opens by trying to tally the benefits of IRB review. "Surprisingly," he writes, a careful review of the literature suggests that "research is not especially dangerous. Some biomedical research can be risky, but much of it requires no physical contact with patients and most contact cannot cause serious injury. Ill patients are, if anything, safer in than out of research." As for social-science research, "its risks are trivial compared with daily risks like going online or on a date."
>
> Since the upsides of IRB review are likely to be modest, Schneider argues, it's critical to ask hard questions about the system's costs. And those costs are serious. To a lawyer's eyes, IRBs are strangely unaccountable. They don't have to offer reasons for their decisions, their decisions can't be appealed, and they're barely supervised at the federal level. That lack of accountability, combined with the gauzy ethical principles that govern IRB deliberations, is a recipe for capriciousness. Indeed, in Schneider's estimation, IRBs wield coercive government power—the power to censor

> university research—without providing due process of law.
>
> And they’re not shy about wielding that power. Over time, IRB review has grown more and more intrusive. Not only do IRBs waste thousands of researcher hours on paperwork and elaborate consent forms that most study participants will never understand. Of greater concern, they also superintend research methods to minimize perceived risks. Yet IRB members often aren’t experts in the fields they oversee. Indeed, some know little or nothing about research methods at all.
>
> IRBs thus delay, distort, and stifle research, especially research on vulnerable subgroups that may benefit most from it. It’s hard to precise about those costs, but they’re high: after canvassing the research, Schneider concludes that “IRB regulation annually costs thousands of lives that could have been saved, unmeasurable suffering that could have been softened, and uncountable social ills that could have been ameliorated.”

This view seems to be growing more popular lately, and has gotten support from high-profile academics like Richard Nisbett and Steven Pinker:

> Should IRBs (human subjects research approval committees) be dismantled? [Probably yes.] http://t.co/5mxhEycEA5
>
> — Steven Pinker (@sapinker) July 24, 2015

And there’s been some recent reform, maybe. The federal Office for Human Research Protections made a vague statement that perhaps studies that obviously aren’t going to hurt anybody might not need the full IRB treatment.

There’s still a lot of debate about how this will be enforced and whether it’s going to lead to any real-life changes. But I’m glad people are starting to think more about these things.

(I’m also glad people are starting to agree that getting rid of a little oversight for the lowest-risk studies is a good compromise, and that we don’t have to start with anything more radical.)

I sometimes worry that people misunderstand the case against bureaucracy. People imagine it’s Big Business complaining about the regulations preventing them from steamrolling over everyone else. That hasn’t been my experience. Big Business – heck, Big Anything – loves bureaucracy. They can hire a team of clerks and secretaries and middle managers to fill out all the necessary forms, and the rest of the company can be on their merry way. It’s everyone else who suffers. The amateurs, the entrepreneurs, the hobbyists, the people doing something as a labor of love. Wal-Mart is going to keep selling groceries no matter how much paperwork and inspections it takes; the poor immigrant family with the backyard vegetable garden might not.

Bureaucracy in science does the same thing: limit the field to big institutional actors with vested interests. No amount of hassle is going to prevent the Pfizer-Merck-Novartis Corporation from doing whatever study will raise their bottom line. But enough hassle *will* prevent a random psychiatrist at a small community hospital from pursuing his pet theory about bipolar diagnosis. The more hurdles we put up, the more the scientific conversation skews in favor of Pfizer-Merck-Novartis. And the less likely we are to hear little stuff, dissenting voices, and things that don’t make anybody any money.

I’m not just talking about IRBs here. I could write a book about this. There are so many privacy and confidentiality restrictions around the most harmless of datasets that research teams won’t share data with one another (let alone with unaffiliated citizen scientists) lest they break some arcane regulation or other. Closed access journals require people to pay thousands of dollars in subscription fees before they’re allowed to read the scientific literature; open-access journals just shift the burden by requiring scientists to pay thousands of dollars to publish their research. Big research institutions have whole departments to deal with these kinds of problems; unaffiliated people who just want to look into things on their own are out of luck.

And this is happening at the same time we’re becoming increasingly aware of the shortcomings of big-name research. Half of psychology studies fail replication; my own field of psychiatry is even worse. And citizen-scientists and science bloggers are playing a big part in debunking bad research: here I’m thinking especially of statistics bloggers like Andrew Gelman and Daniel Lakens, but there are all sorts of people in this category. And both Gelman and Lakens are PhDs with institutional affiliations – “citizen science” doesn’t mean random cavemen who don’t understand the field – but they’re both operating outside their day job, trying to contribute a few hours per project instead of a few years. I know many more people like them – smart, highly-qualified, but maybe not going to hire a team of paper-pushers and spend thousands of dollars in fees in order to say what they have to say. Even now these people are doing great work – but I can’t help but feel like more is possible.

IRB overreach is a small part of the problem. But it’s the part which sunk my bipolar study, a study I really cared about. I’m excited that there’s finally more of a national conversation about this kind of thing, and hopeful that further changes will make scientific efforts easier and more rewarding for the next generation of doctors.

Highlights From The Comments On My IRB Nightmare

Many people took My IRB Nightmare as an opportunity to share their own IRB stories. From an emergency medicine doctor, via my inbox:

> Thanks for the great post about IRBs. I lived the same absurd nightmare in 2015-2016, as an attending, and it’s amazing how your experience matches my own, despite my being in Canada.
>
> One of our residents had an idea for an extremely simple physiological study of COPD exacerbations, where she’d basically look at the patient and monitor his RR, saturation, exhaled CO2 temporal changes during initial treatment. Just as you were, I was really naive back in 2015, and expected we wouldn’t even a consent form, since she didn’t even have to *talk* to the patients, much less perform any intervention. Boy I was wrong ! The IRB, of course, insisted on a two-page consent form discussing risks and benefits of the intervention, and many other forms. I had to help her file over 300 pages (!) of various forms. Just as in your case, we had to abandon the study when, two years after the first contact with the IRB, they suggested hilarious “adjustments” to the study protocol “in order to mitigate possible risks”.

From baj2235 on the subreddit:

> Currently working in a brand new lab, so one would think I’d have a lot to do. Instead, thus far my job has consisted of sitting in an empty room coming up with increasingly unlikely hypotheses that will probably never

> be tested because our IRB hasn't approve our NOU (Notice of use) forms. For those who don't know, NOUs are essentially 15 page forms that say "We study these this, and we promise to be super responsible while studying it." We have 4 currently awaiting approval, submitted in May. The reason they aren't approved yet? The IRB hasn't met since June, and likely won't meet again this month because of frickin' Harvey. Which in essence means the fine American taxpayer has essentially been paying me to sit in a room and twiddle my thumbs for the past 3-months because I can't even grow E. coli without a frickin' NOU.

From Garrett in the comments:

> Oh, dear! I've actually been through this. I work in tech, but volunteer in EMS. As a part of wanting to advance the profession of EMS I figured I'd take on a small study. It would be a retrospective study about how well paramedics could recognize diabetic ketoacidosis (DKA) and Hyperosmolar hyperglycemic state (HHS) in comparison to ER doctors. […]
>
> I had to do the "I am not a Nazi" training as well. In order to pass that, I had to be able to recite the FDA form number used as a part of new implantable medical device investigations. I wasn't looking at a new device. I wasn't looking at an old device. I was going to look at pairs of medical records and go "who correctly identified the problem?" […]
>
> It's now ~5 years after IRB and because of all of the headaches of getting the data to someone who isn't faculty or a doctor, and who doesn't have a $100k+ grant, I still don't have my data. I need to send another

> email. I’m sure we can get an IRB extension with a few more trees sacrificed.

From Katie on Facebook:

> I used to work at an fMRI research center and also had to take the Don’t Be a Nazi course!
>
> My favorite story about the annoying IRB regulations is how they insisted on an HCG (pregnancy) test for our volunteers, despite the fact that MRI has no known adverse effect on pregnancy. So, fine, extra caution against an unknown but possible risk, sure.
>
> But they insisted on a *blood test* done days in advance instead of five minute urine dip stick test that *actual doctors offices* would use. You know what doesn’t have risks? Peeing in a cup. And what does have risks of fainting, infection, collapsing a vein, etc? A blood draw.
>
> Of course, we had an extra consent form for them to sign, about the risks of the blood draw the IRB was helpfully insisting on.

From Hirsin on Hacker News:

> My freshman year of college I proposed a study to our hospitals IRB to strap small lasers to three week old infants in an effort to measure concentrations of a chemical in their blood. The most frustrating part was not the arcane insistence on ink and bolded study names, but the hardline insistence that it was impossible (illegal) to test the device before getting IRB approval – even on ourselves. Meaning that without any calibration or testing, our initial study would likely come

> back with poor results or be a dud, but we couldn't find out until we filled out all the paperwork.

What is our country coming to when you can't even attach lasers to babies anymore?

Some of the other stories were kind of cute. Dahud in the comments:

> I've had exactly one interaction with an IRB – in 6th grade. My science fair project involved studying the health risks of Communion as performed in the Episcopal church. (For those unfamiliar, a priest lifts a silver chalice of port wine to your lips, you take a small sip, and the priest wipes the site with a linen cloth and rotates the chalice.)
>
> Thing was, the science fair was being held by a Baptist University. The IRB was really not fond of the whole wine thing. They wanted me to use grape juice instead, in the Baptist fashion. I, as a minor, shouldn't be allowed anywhere near the corrupting influence of the communion wine that I had partaken of last Sunday.
>
> Of course, the use of communion wine was essential to the study, so we reached a compromise. I would thoroughly document all the sample collection and preparation procedures, and let someone of age carry out the experiment while I waited in the hall.

And of course James Miller is still James Miller:

> Several forms I have to sign to do things at my college ask if what I will be doing will expose anyone to radiation. Although I'm an economist, this has caused me to think of experiments I could do with radiation such as secretly exposing a large number of students to

> radiation and seeing, years later, if it influences their income.

Along with these, a lot of other people were broadly sympathetic but thought that if I knew how to play the system a little better, or was somewhere a little more research-focused, things might have gone better for me. Virgil in the comments:

> FWIW, I’m a graduate student in the Social Sciences. Our IRBs have the same rules on paper, but we get around it by using generic versions of applications with the critical info swapped out, or just ignoring them altogether. Though we don’t have to face audits, so… I’ve found that usually if you make one or two glaring errors in the application on purpose, the IRB will be happy to inform you of those and approve it when you correct them. They just want to feel powerful / like they’re making a difference, so if you oblige them they will usually let you through with no further hassle.

From Eternaltraveler in the comments:

> Most of the bureaucracy you experienced is institutional and not regulatory. I have done research both in an institutional setting (turn around time at UC Berkeley=5 months to obtain ethics approval and countless hours sucking up to self important bureaucrats who think it’s their sacred duty to grind potentially life saving research to a halt over trivia they themselves know is meaningless), and as an entrepreneur and PI at a biotech startup (turn around time for outsourced IRB=5 days with reasonable and informed questions related to participants well being), where we also do quite a bit more than ask questions. FYI the kind of research I did

> at UC Berkeley that took 5 months for approval has absolutely no regulatory requirements outside of it.

And from PM_ME_YOUR_FRAME on the subreddit (who I might hunt down and beg to be my research advisor if I ever do anything like this again):

> Amateur. What you do is you sweet talk the clinicians into using their medical judgement to adopt the form as part of their routine clinical practice and get them to include it as part of the patient's medical records. Later… you approach the IRB for a retrospective chart review study and get blessed with waived consent. Bonus: very likely to also get expedited review.

And this really thorough comment from friendlygrantadmit:

> I'm not an expert in IRB (although that's kind of my point–getting to that), but I think your headaches were largely institutional rather than dictated by government fiat. Let me explain…
>
> I used to be the grant administrator for a regional university while my husband was a postdoc at the large research university 20 miles away. Aside from fiscal stuff, I was the grants office, and the grants office was me. However, there was an IRB of longstanding duration, so I never had to do much other than connect faculty whose work might involve human subjects with the IRB Chair. I think I was technically a non-voting member or something, but no one expected me to attend meetings.
>
> This was in the process of changing when I left the university because my husband's postdoc ended and we moved. It was a subject that generated much bitterness among the small cadre of faculty involved. Because I

was on my way out, I never made it my business to worry about nascent IRB woes. My understanding was that they had difficulty getting people to serve on the IRB because it was an unpaid position, but as the university expanded, they were going to need more and different types of expertise represented on the IRB. I can’t be more specific than that without basically naming the university, at which I was very happy and with which I have no quarrel. I never heard any horror stories about our IRB, and I would have been the first point person to hear the them, so I presume it was fairly easy to work with.

Anyway, the IRB auditing stuff you outline is just insane. The institutional regulations pertaining to the audits were probably what generated the mind-numbing and arcane complexity of your institution’s IRB. Add in finicky personalities and you have a recipe for endless hassle as described.

So here’s the other thing to bear in mind: almost everyone in research administration is self-trained. I think there are a few programs (probably mostly online), but it’s the sort of field that people stumble into from related fields. You learn on the job and via newsletters, conferences, and listservs. You also listen to your share of mind-numbing government webinars. But almost everyone–usually including the federal program officers, who are usually experts in their field but who aren’t necessarily experts in their own particular bureaucracy–is just winging it.

Most research admins are willing to admit the “winging it” factor among themselves. For obvious reasons, however, you want the faculty and/or researchers with whom you interact to respect your professional judgment. This was never a problem at my institution, which is probably one reason I still have a high opinion of it and its administration, but I heard plenty (PLENTY) of stories of bigshot faculty pulling rank to have the rules and regulations bent or broken in their favor because GRANT MONEY, usually with success. So of course you're not going to confess that you don't really have a clue what you're doing; you're just puzzling over these regulations like so many tea leaves and trying to make a reasonable judgment based on your status as a reasonably well-educated and fair-minded human being.

What this means in practice is almost zero uniformity in the field. Your IRB from hell story wasn't even remotely shocking to me. Other commenters' IRB from just-fine-ville stories are also far from shocking. Since so few people really understand what the regulations mean or how to interpret them, let alone how to protect against government bogeymen yelling at you failing to follow them, there is a wild profusion of institutional approaches to research administration, and this includes huge variations in concern for the more fine-grained regulatory details. It is really hard to find someone to lead a grants or research administration office who has expertise in all the varied fields of compliance now required. It's hard to find someone with the expertise in any of the particular fields, to be honest.

There is one area in which this is not so much true, and that is financial regulations. Why? Well, for one thing, they're not all that tricky–I could read and interpret them with far greater confidence than many other regs, despite having a humanities background. The other reason is that despite their comparative transparency, they were very, very widely flouted until the government started auditing large research institutions around 15-ish years ago.

I have a short story related to that, too–basically, when my husband started grad school, we would frequently go out to dinner with his lab group and advisor. The whole tab, including my dinner and that of any other SOs and all alcoholic beverages (which can’t be paid for with grant funds aside from narrow research-related exceptions), would be charged to whichever research grant because it was a working meal. I found it mildly surprising, but I certainly wasn’t going to argue.

Then the university got audited and fined millions of dollars for violations such as these and Found Religion vis-à-vis grant expenditures.

With regards to your story, I’m guessing that part of the reason the IRB is such a big deal is that human subjects research is the main type of research, so they are really, really worried about their exposure to any IRB lapses. However, it sounds like they are fairly provincial in that they aren’t connected to what more major research institutions are doing or how they handle these issues, which is always a mistake. Even if you don’t think some other institution’s approach is going to work for you, it’s good to know about as many different approaches as you can to know that you’re not some insane outlier as your IRB seems to be. As others have noted, it also sounds like that IRB has become the fiefdom of some fairly difficult personalities.

I already know how extensive, thorough, and helpful training pertaining to IRB regs is, which is not very. I remain deeply curious about the qualifications and training of your obviously well-intentioned “auditor.” My

> guess is she inherited her procedures from someone else and is carefully following whatever checklist was laid down so as not to expose herself to accusations of sloppiness or lack of thoroughness … but that is only a guess.
>
> Even though I hate hearing stories like yours–there is obviously no excuse for essentially trying to thwart any and all human subjects research the way your IRB did–I am sympathetic to the need for some regulations, and not just because of Nazis and the Tuskeegee Syphilis experiments. I'm sympathetic because lack of oversight basically gives big name researchers carte blanche to ignore regulations they find inconvenient because the institutional preference, barring opposing headwinds, will always be to keep researchers happy.

Some people thought I was being too flippant, or leaving out parts of the story. Many of them mentioned that the focus on Nazis overshadowed some genuinely horrific all-American research misconduct like the Tuskegee Syphilis Experiment. They emphasized that my personal experience doesn't overrule all of the really important reasons IRBs exist. For example, tedwick from the subreddit:

> So, I wrote out all of the ways in which Scott's terrible IRB experience was at least in part self-imposed, and how a lot of the post was about stuff that's pretty straightforward, but it was kind of a snarky comment. Not unlike his post, but you know, whatever. Long story short, I've done similar work (arranged a really simple survey looking at dietary behaviors in kids, another IRB-protected group) and had to interface with the IRB frequently. Yep, it can be annoying at times. But the reason they ask people like Scott whether they're going to try anything funny with prisoners is because

> sometimes people like Scott are trying something funny with prisoners. Just because Scott swears that he's not Mengele doesn't mean that he's not going to do something dumb a priori. As his experience with expedited review might indicate, sitting down with an IRB officer for maybe 30 minutes would have cleared up a lot of things on both sides.
>
> Is there room for IRB reform? Sure! Let's make the easy stuff easy, and let's make sure IRB intervention is on actual substance. I'm with him on this. However, a lot of the stuff Scott is complaining about doesn't fall into that category (e.g. "why do all the researchers have to be on the IRB!?"). I get that the post was probably cathartic for Scott to write, but there are plenty of great researchers who are able to navigate this stuff without all the drama. "Bureaucracy Bad" is a fine rallying cry and all that, but most of the stuff Scott is complaining about is not all that hard and there for a reason.

And kyleboddy from the comments:

> Nazism isn't the reason IRBs exist. Far worse. American unethical experimentation is, and omitting it is a huge error. Massive and bureaucratic oversight exists because American scientists would stop at nothing to advance the field of science.
>
> The Tuskegee Syphilis Experiment is the landmark case on why ethical training and IRB approval is required. You should know this. This was 100% covered in your ethical training.
>
> I get why IRB approval sucks. My Informed Consent forms get banged all the time. But we're talking about

> consent here, often with disadvantaged populations. It pays to be careful.
>
> Last, most researchers who need speed and expedited review go through private IRB organizations now because the bureaucracy of medical/university systems is too much to handle. Our private IRB that we engage with sends back our forms within a week and their fees are reasonable. Their board meets twice per week, not once per month. The market has solved at least this particular issue.
>
> EDIT: Private IRBs do not care about nonsensical stuff like the Principal Investigator having an advanced degree or being someone high of stature. (For example, I am a college dropout and have had multiple IRB studies approved.) Only bureaucratic, publicly-attached ones do. That’s a very reasonable complaint.

A lot of these are good points. And some of what I wrote was definitely unfair snark – I understand they’ve got to ask you whether you plan on removing anyone’s organs; if they don’t ask, how will they know? And maybe linking to Schneider’s book about eliminating the IRB system was a mistake – I just meant to show there was an existing conversation about this. I definitely didn’t mean to trivialize Tuskegee, to say that I am a radical Scheiderian, to act like my single experience damns all IRBs forever, or to claim that IRBs can’t possibly have a useful role to play. I haven’t even begun to wade into the debate between the critics and proponents of the system. The point I wanted to make was that whether or not IRBs are useful for high-risk studies, they’ve crept annoyingly far into low-risk studies – to the detriment of everyone.

Nobody expects any harm from asking your co-worker “How are you this morning?” in conversation. But if I were to turn this into a study – “Diurnal Variability In Well-Being Among Office Workers” – I would need to hire a whole team of managers just to get through the risk paperwork and the consent paperwork and the weekly reports and the team meetings. I can give a patient twice the standard dose of a dangerous medication without justifying myself to anyone. I can confine a patient involuntarily for weeks and face only the most perfunctory legal oversight. But if I want to ask them “How are you this morning?” and make a study out of it, I need to block off my calendar for the next ten years to do the relevant paperwork.

I feel like I’m protesting a police state, and people are responding “Well, you don’t want total anarchy with murder being legal, do you?” No, I don’t. I think there’s a wide range of possibilities between “police state” and “anarchy”. In the same way, I think there’s a wide range of possibilities between “science is totally unregulated” and “scientists have to complete a mountain of paperwork before they can ask someone how their day is going”.

> oof. takes me back. we had initial GD RCT delayed by a year because IRB said giving people money could hurt them
>
> — Paul Niehaus (@PaulFNiehaus) August 31, 2017

I dare you to tell me we’re at a happy medium right now. Go on, I dare you.

I regret to say this is only getting worse. New NIH policies are increasingly trying to reclassify basic science as “clinical trials”, requiring more paperwork and oversight. For example, under the new regulations, brain scan research –

the type where they ask you to think about something while you're in an fMRI to see which parts of your brain light up – would be a "clinical trial" since it measures "a health-related biomedical or behavioral outcome". This could require these studies to meet the same high standards as studies giving experimental drugs or new gene therapies. The Science Magazine article quotes a cognitive neuroscientist:

> The agency's widening definition of clinical trials could sweep up a broad array of basic science studies, resulting in wasted resources and public confusion. "The massive amount of dysfunction and paperwork that will result from this decision boggles the mind" and will hobble basic research.

A bunch of researchers from top universities have written a petition trying to delay the changes (if you've got an academic affiliation, you might want to check and consider signing yourself). But it's anyone's guess whether they'll succeed. If not, good luck to any brain researcher who doesn't want to go through everything I did. They'll need it.

Book Review: Surfing Uncertainty

[Related to: It's Bayes All The Way Up, Why Are Transgender People Immune To Optical Illusions?, Can We Link Perception And Cognition?]

I.

Sometimes I have the fantasy of being able to glut myself on Knowledge. I imagine meeting a time traveler from 2500, who takes pity on me and gives me a book from the future where all my questions have been answered, one after another. What's consciousness? That's in Chapter 5. How did something arise out of nothing? Chapter 7. It all makes perfect intuitive sense and is fully vouched by unimpeachable authorities. I assume something like this is how everyone spends their first couple of days in Heaven, whatever it is they do for the rest of Eternity.

And every so often, my fantasy comes true. Not by time travel or divine intervention, but by failing so badly at paying attention to the literature that by the time I realize people are working on a problem it's already been investigated, experimented upon, organized into a paradigm, tested, and then placed in a nice package and wrapped up with a pretty pink bow so I can enjoy it all at once.

The predictive processing model is one of these well-wrapped packages. Unbeknownst to me, over the past decade or so neuroscientists have come up with a real *theory* of how the brain works – a real unifying framework theory like Darwin's or Einstein's – and it's beautiful and it makes complete sense.

Surfing Uncertainty isn't pop science and isn't easy reading. Sometimes it's on the border of possible-at-all reading. Author Andy Clark (a professor of logic and metaphysics, of all things!) is clearly brilliant, but prone to going on long digressions about various esoteric philosophy-of-cognitive-science debates. In particular, he’s obsessed with showing how “embodied” everything is all the time. This gets kind of awkward, since the predictive processing model isn’t really a natural match for embodiment theory, and describes a brain which is pretty embodied in some ways but not-so-embodied in others. If you want a hundred pages of apologia along the lines of “this may not *look* embodied, but if you squint you’ll see how super-duper embodied it really is!”, this is your book.

It’s also your book if you want to learn about predictive processing at all, since as far as I know this is the only existing book-length treatment of the subject. And it’s comprehensive, scholarly, and very good at giving a good introduction to the theory and why it’s so important. So let’s be grateful for what we’ve got and take a look.

II.

Stanislas Dehaene writes of our senses:

> We never see the world as our retina sees it. In fact, it would be a pretty horrible sight: a highly distorted set of light and dark pixels, blown up toward the center of the retina, masked by blood vessels, with a massive hole at the location of the “blind spot” where cables leave for the brain; the image would constantly blur and change as our gaze moved around. What we see, instead, is a three-dimensional scene, corrected for retinal defects, mended at the blind spot, stabilized for our eye and head movements, and massively reinterpreted based on our previous experience of similar visual scenes. All these operations unfold unconsciously—although many of them are so complicated that they resist computer modeling. For instance, our visual system detects the

> presence of shadows in the image and removes them. At a glance, our brain unconsciously infers the sources of lights and deduces the shape, opacity, reflectance, and luminance of the objects.

Predictive processing begins by asking: how does this happen? By what process do our incomprehensible sense-data get turned into a meaningful picture of the world?

The key insight: the brain is a multi-layer prediction machine. All neural processing consists of two streams: a bottom-up stream of sense data, and a top-down stream of predictions. These streams interface at each level of processing, comparing themselves to each other and adjusting themselves as necessary.

The bottom-up stream starts out as all that incomprehensible light and darkness and noise that we need to process. It gradually moves up all the cognitive layers that we already knew existed – the edge-detectors that resolve it into edges, the object-detectors that shape the edges into solid objects, et cetera.

The top-down stream starts with everything you know about the world, all your best heuristics, all your priors, everything that's ever happened to you before – everything from "solid objects can't pass through one another" to "e=mc^2" to "that guy in the blue uniform is probably a policeman". It uses its knowledge of concepts to make predictions – not in the form of verbal statements, but in the form of expected sense data. It makes some guesses about what you're going to see, hear, and feel next, and asks "Like this?" These predictions gradually move *down* all the cognitive layers to generate lower-level predictions. If that uniformed guy was a policeman, how would that affect the various objects in the scene? Given the answer to that question, how would it affect the distribution of edges in the scene? Given the answer to *that* question, how would it affect the raw-sense data received?

Both streams are probabilistic in nature. The bottom-up sensory stream has to deal with fog, static, darkness, and neural noise; it knows that whatever forms it tries to extract from this signal might or might not be real. For its part, the top-down predictive stream knows that predicting the future is inherently difficult and its models are often flawed. So both streams contain not only data but estimates of the precision of that data. A bottom-up percept of an elephant right in front of you on a clear day might be labelled “very high precision”; one of a a vague form in a swirling mist far away might be labelled “very low precision”. A top-down prediction that water will be wet might be labelled “very high precision”; one that the stock market will go up might be labelled “very low precision”.

As these two streams move through the brain side-by-side, they continually interface with each other. Each level receives the predictions from the level above it and the sense data from the level below it. Then each level uses Bayes’ Theorem to integrate these two sources of probabilistic evidence as best it can. This can end up a couple of different ways.

First, the sense data and predictions may more-or-less match. In this case, the layer stays quiet, indicating “all is well”, and the higher layers never even hear about it. The higher levels just keep predicting whatever they were predicting before.

Second, low-precision sense data might contradict high-precision predictions. The Bayesian math will conclude that the predictions are still probably right, but the sense data are wrong. The lower levels will “cook the books” – rewrite the sense data to make it look as predicted – and then continue to be quiet and signal that all is well. The higher levels continue to stick to their predictions.

Third, there might be some unresolvable conflict between high-precision sense-data and predictions. The Bayesian math will indicate that the predictions are probably wrong. The neurons involved will fire, indicating “surprisal” – a gratuitiously-technical neuroscience term for surprise. The higher the degree of mismatch, and the higher the supposed precision of the data that led to the mismatch, the more surprisal – and the louder the alarm sent to the higher levels.

When the higher levels receive the alarms from the lower levels, *this is their equivalent of bottom-up sense-data*. They ask themselves: “Did the even-higher-levels predict this would happen?” If so, they themselves stay quiet. If not, they might try to change their own models that map higher-level predictions to lower-level sense data. Or they might try to cook the books themselves to smooth over the discrepancy. If none of this works, they send alarms to the even-higher-levels.

All the levels really hate hearing alarms. Their goal is to *minimize surprisal* – to become so good at predicting the world (conditional on the predictions sent by higher levels) that nothing ever surprises them. Surprise prompts a frenzy of activity adjusting the parameters of models – or deploying new models – until the surprise stops.

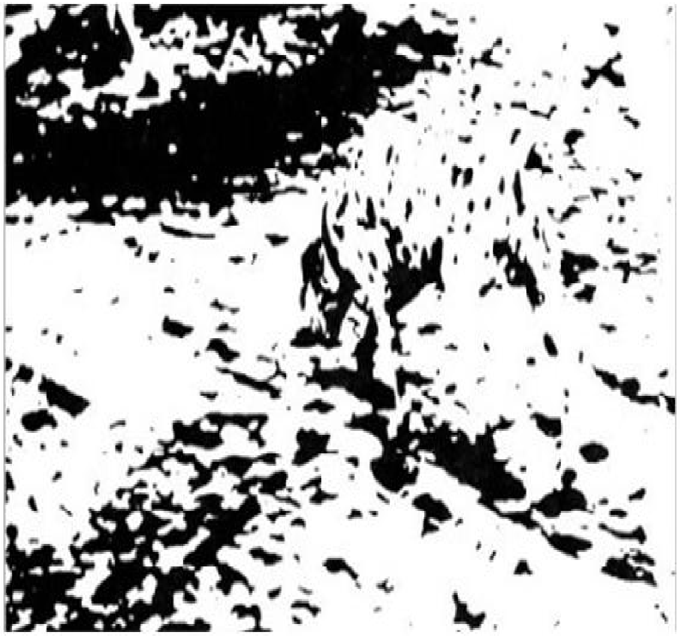
All of this happens several times a second. The lower levels constantly shoot sense data at the upper levels, which constantly adjust their hypotheses and shoot them down at the lower levels. When surprise is registered, the relevant levels change their hypotheses or pass the buck upwards. After umpteen zillion cycles, everyone has the right hypotheses, nobody is surprised by anything, and the brain rests and moves on to the next task. As per the book:

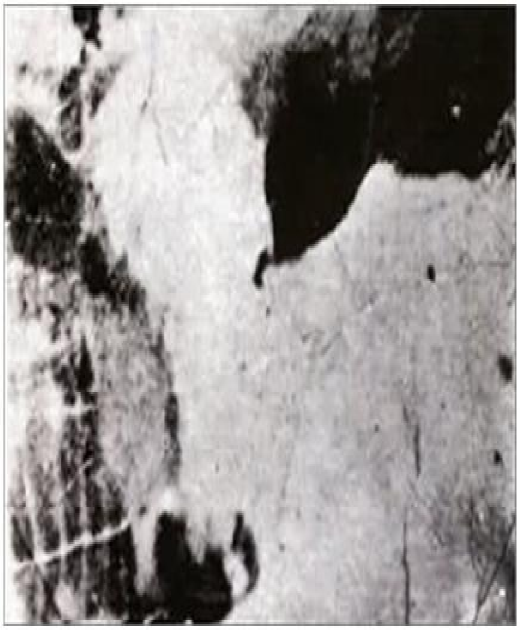
> To deal rapidly and fluently with an uncertain and noisy world, brains like ours have become masters of prediction – surfing the waves and noisy and ambiguous sensory stimulation by, in effect, trying to stay just ahead of them. A skilled surfer stays ‘in the pocket’: close to, yet just ahead of the place where the wave is breaking. This provides power and, when the wave breaks, it does not catch her. The brain’s task is not dissimilar. By constantly attempting to predict the incoming sensory signal we become able – in ways we shall soon explore in detail – to learn about the world around us and to engage that world in thought and action.

The result is perception, which the PP theory describes as “controlled hallucination”. You’re not seeing the world as it is, exactly. You’re seeing your predictions about the world, cashed out as expected sensations, then shaped/constrained by the actual sense data.

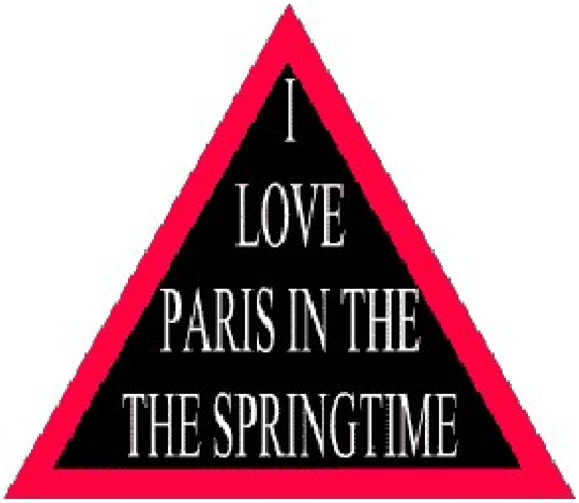
III.

Enough talk. Let’s give some examples. Most of you have probably seen these before, but it never hurts to remind:

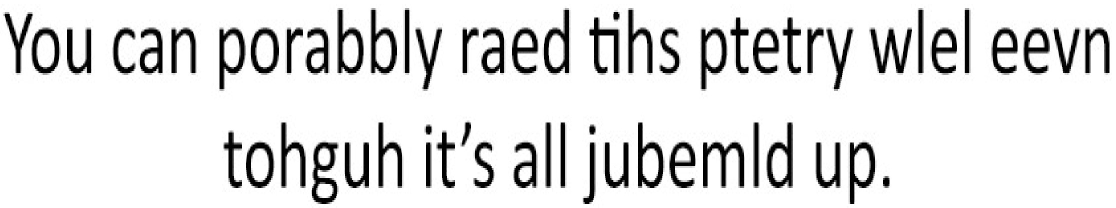




This demonstrates the degree to which the brain depends on top-down hypotheses to make sense of the bottom-up data. To most people, these two pictures start off looking like incoherent blotches of light and darkness. Once they figure out what they are (spoiler) the scene becomes obvious and coherent. According to the predictive processing model, this is how we perceive everything all the time – except usually the concepts necessary to make the scene fit together come from our higher-level predictions instead of from clicking on a spoiler link.



This demonstrates how the top-down stream’s efforts to shape the bottom-up stream and make it more coherent can sometimes “cook the books” and alter sensation entirely. The real picture says “PARIS IN THE THE SPRINGTIME” (note the duplicated word “the”!). The top-down stream predicts this should be a meaningful sentence that obeys English grammar, and so replaces the the bottom-up stream with what it thinks that it *should* have said. This is a very powerful process – how many times have I repeated the the word “the” in this paragraph alone without you noticing?



You can porabbly raed tihs ptetry wlel eevn tohguh it’s all jubemld up.

A more ambiguous example of “perception as controlled hallucination”. Here your experience doesn’t quite *deny* the jumbled-up nature of the letters, but it superimposes a “better” and more coherent experience which appears naturally alongside.

Next up – this low-quality video of an airplane flying at night. Notice how after an instant, you start to predict the movement and characteristics of the airplane, so that you're no longer surprised by the blinking light, the movement, the other blinking light, the camera shakiness, or anything like that – in fact, if the light *stopped* blinking, you would be surprised, even though naively nothing could be less surprising than a dark portion of the night sky staying dark. After a few seconds of this, the airplane continuing on its (pretty complicated) way just reads as "same old, same old". Then when something else happens – like the camera panning out, or the airplane making a slight change in trajectory – you focus entirely on that, the blinking lights and movement entirely forgotten or at least packed up into "airplane continues on its blinky way". Meanwhile, other things – like the feeling of your shirt against your skin – have been completely predicted away and blocked from consciousness, freeing you to concentrate entirely on any subtle changes in the airplane's motion.

In the same vein: this is Rick Astley's "Never Going To Give You Up" repeated again and again for ten hours (you can find some *weird* stuff on YouTube). The first hour, maybe you find yourself humming along occasionally. By the second hour, maybe it's gotten kind of annoying. By the third hour, you've completely forgotten it's even on at all.

But suppose that one time, somewhere around the sixth hour, it skipped two notes – just the two syllables "never", so that Rick said "Gonna give you up." Wouldn't the silence where those two syllables should be sound as jarring as if somebody set off a bomb right beside you? Your brain, having predicted sounds consistent with "Never Gonna Give You Up" going on forever, suddenly finds its expectations violated and sends all sorts of alarms to the higher levels, where they eventually reach your consciousness and make you go “What the *heck*?”

IV.

Okay. You’ve read a lot of words. You’ve looked at a lot of pictures. You’ve listened to “Never Gonna Give You Up” for ten hours. Time for the payoff. Let’s use this theory to explain everything.

1. Attention. In PP, attention measures “the confidence interval of your predictions”. Sense-data within the confidence intervals counts as a match and doesn’t register surprisal. Sense-data outside the confidence intervals fails and alerts higher levels and eventually consciousness.

This modulates the balance between the top-down and bottom-up streams. High attention means that perception is mostly based on the bottom-up stream, since every little deviation is registering an error and so the overall perceptual picture is highly constrained by sensation. Low attention means that perception is mostly based on the top-down stream, and you’re perceiving only a vague outline of the sensory image with your predictions filling in the rest.

There’s a famous experiment which you can try below – if you’re trying it, make sure to play the whole video before moving on:

…

…

About half of subjects, told to watch the players passing the ball, don’t notice the gorilla. Their view of the ball-passing is closely constrained by the bottom-up stream; they see mostly what is there. But their view of the gorilla is mostly dependent on the top-down stream. Their confidence intervals are wide. Somewhere in your brain is a neuron saying “is that a guy in a gorilla suit?” Then it consults the top-down stream, which says “This is a basketball game, you moron”, and it smooths out the anomalous perception into something that makes sense like another basketball player.

But if you watch the video with the prompt “Look for something strange happening in the midst of all this basketball-playing”, you see the gorilla immediately. Your confidence intervals for unusual things are razor-thin; as soon as that neuron sees the gorilla it sends alarms to higher levels, and the higher levels quickly come up with a suitable hypothesis (“there’s a guy in a gorilla suit here”) which makes sense of the new data.

There’s an interesting analogy to vision here, where the center of your vision is very clear, and the outsides are filled in in a top-down way – I have a vague sense that my water bottle is in the periphery right now, but only because I kind of already know that, and it’s more of a mental note of “water bottle here as long as you ask no further questions” than a clear image of it. The extreme version of this is the blind spot, which gets filled in entirely with predicted imagery despite receiving no sensation at all.

2. Imagination, Simulation, Dreaming, Etc. Imagine a house. Now imagine a meteor crashing into the house. Your internal mental simulation was probably pretty good. Without even thinking about it, you got it to obey accurate physical laws like “the meteor continues on a constant trajectory”, “the impact happens in a realistic way”, “the impact shatters the meteorite”, and “the meteorite doesn’t bounce back up to space like a basketball”. Think how surprising this is.

In fact, think how surprising it is that you can imagine the house at all. This really high level concept – “house” – has been transformed in your visual imaginarium into a pretty good picture of a house, complete with various features, edges, colors, et cetera (if it hasn't, read here). This is near-miraculous. Why do our brains have this apparently useless talent?

PP says that the highest levels of our brain make predictions *in the form of sense data*. They're not just saying "I predict that guy over there is a policeman", they're generating the image of a policeman, cashing it out in terms of sense data, and colliding it against the sensory stream to see how it fits. The sensory stream gradually modulates it to fit the bottom-up evidence – a white or black policeman, a mustached or clean-shaven policeman. But the top-down stream is doing a lot of the work here. We are able to imagine the meteor, using the same machinery that would guide our perception of the meteor if we saw it up in the sky.

All of this goes double for dreaming. If "perception is controlled hallucination" caused by the top-down drivers of perception constrained by bottom-up evidence, then dreams are those top-down drivers playing around with themselves unconstrained by anything at all (or else very weakly constrained by bottom-up evidence, like when it's really cold in your bedroom and you dream you're exploring the North Pole).

A lot of people claim higher levels of this – lucid dreaming, astral projection, you name it, worlds exactly as convincing as our own but entirely imaginary. Predictive processing is very sympathetic to these accounts. The generative models that create predictions are really good; they can simulate the world well enough that it rarely surprises us. They also connect through various layers to our bottom-level perceptual apparatus, cashing out their predictions in terms of the lowest-level sensory signals. Given that we've got a top-notch world-simulator plus perception-generator in our heads, it shouldn't be surprising when we occasionally perceive ourselves in simulated worlds.

3. Priming. I don't mean the weird made-up kinds of priming that don't replicate. I mean the very firmly established ones, like the one where, if you flash the word "DOCTOR" at a subject, they'll be much faster and more skillful in decoding a series of jumbled and blurred letters into the word "NURSE".

This is classic predictive processing. The top-down stream's whole job is to assist the bottom-up stream in making sense of complicated fuzzy sensory data. After it hears the word "DOCTOR", the top-down stream is already thinking "Okay, so we're talking about health care professionals". This creeps through all the lower levels as a prior for health-care related things; when the sense organs receive data that can be associated in a health-care related manner, the high prior helps increase the precision of this possibility until it immediately becomes the overwhelming leading hypothesis.

4. Learning. There's a philosophical debate – which I'm not too familiar with, so sorry if I get it wrong – about how "unsupervised learning" is possible. Supervised reinforcement learning is when an agent tries various stuff, and then someone tells the agent if it's right or wrong. Unsupervised learning is when nobody's around to tell you, and it's what humans do all the time.

PP offers a compelling explanation: we create models that generate sense data, and keep those models if the generated sense data match observation. Models that predict sense data well stick around; models that fail to predict the sense data accurately get thrown out. Because of all those lower layers adjusting out contingent features of the sensory stream, any given model is left with exactly the sense data necessary to tell it whether it’s right or wrong.

PP isn’t *exactly* blank slatist, but it’s compatible with a slate that’s pretty fricking blank. Clark discusses “hyperpriors” – extremely basic assumptions about the world that we probably need to make sense of anything at all. For example, one hyperprior is sensory synchronicity – the idea that our five different senses are describing the same world, and that the stereo we see might be the source of the music we hear. Another hyperprior is object permanence – the idea that the world is divided into specific objects that stick around whether or not they’re in the sensory field. Clark says that some hyperpriors *might* be innate – but says they don’t have to be, since PP is strong enough to learn them on its own if it has to. For example, after enough examples of, say, seeing a stereo being smashed with a hammer at the same time that music suddenly stops, the brain can infer that connecting the visual and auditory evidence together is a useful hack that helps it to predict the sensory stream.

I can’t help thinking here of Molyneux’s Problem, a thought experiment about a blind-from-birth person who navigates the world through touch alone. If suddenly given sight, could the blind person naturally connect the visual appearance of a cube to her own concept “cube”, which she derived from the way cubes feel? In 2003, some researchers took advantage of a new cutting-edge blindness treatment to test this out; they found that no, the link isn’t intuitively obvious to them. Score one for learned hyperpriors.

But learning goes all the way from these kinds of really basic hyperpriors all the way up to normal learning like what the capital of France is – which, if nothing else, helps predict what’s going to be on the other side of your geography flashcard, and which high-level systems might keep as a useful concept to help it make sense of the world and predict events.

5. Motor Behavior. About a third of *Surfing Uncertainty* is on the motor system, it mostly didn't seem that interesting to me, and I don't have time to do it justice here (I might make another post on one especially interesting point). But this has been kind of ignored so far. If the brain is mostly just in the business of making predictions, what exactly is the motor system doing?

Based on a bunch of really excellent experiments that I don't have time to describe here, Clark concludes: it's predicting action, which causes the action to happen.

This part is almost funny. Remember, the brain really hates prediction error and does its best to minimize it. With failed predictions about eg vision, there's not much you can do except change your models and try to predict better next time. But with predictions about proprioceptive sense data (ie your sense of where your joints are), there's an easy way to resolve prediction error: just move your joints so they match the prediction. So (and I'm asserting this, but see Chapters 4 and 5 of the book to hear the scientific case for this position) if you want to lift your arm, your brain just predicts *really really strongly* that your arm has been lifted, and then lets the lower levels' drive to minimize prediction error do the rest.

Under this model, the "prediction" of a movement isn't just the idle thought that a movement might occur, it's *the actual motor program*. This gets unpacked at all the various layers – joint sense, proprioception, the exact tension level of various muscles – and finally ends up in a particular fluid movement:

> Friston and colleagues…suggest that precise proprioceptive predictions directly elicit motor actions.

> This means that motor commands have been replaced by (or as I would rather say, implemented by) proprioceptive predictions. According to active inference, the agent moves body and sensors in ways that amount to actively seeking out the sensory consequences that their brains expect. Perception, cognition, and action – if this unifying perspective proves correct – work together to minimize sensory prediction errors by selectively sampling and actively sculpting the stimulus array. This erases any fundamental computational line between perception and the control of action. There remains [only] an obvious difference in direction of fit. Perception here matches hural hypotheses to sensory inputs…while action brings unfolding proprioceptive inputs into line with neural predictions. The difference, as Anscombe famously remarked, is akin to that between consulting a shopping list (thus letting the list determine the contents of the shopping basket) and listing some actually purchased items (thus letting the contents of the shopping basket determine the list). But despite the difference in direction of fit, the underlying form of the neural computations is now revealed as the same.

6. Tickling Yourself. One consequence of the PP model is that organisms are continually adjusting out their own actions. For example, if you're trying to predict the movement of an antelope you're chasing across the visual field, you need to adjust out the up-down motion of your own running. So one "hyperprior" that the body probably learns pretty early is that if it itself makes a motion, it should expect to feel the consequences of that motion.

There's a really interesting illusion called the force-matching task. A researcher exerts some force against a subject, then asks the subject to exert exactly that much force against something else. Subjects' forces are usually biased upwards

– they exert more force than they were supposed to – probably because their brain’s prediction engines are “cancelling out” their own force. Clark describes one interesting implication:

> The same pair of mechanisms (forward-model-based prediction and the dampening of resulting well-predicted sensation) have been invoked to explain the unsettling phenomenon of ‘force escalation’. In force escalation, physical exchanges (playground fights being the most common exemplar) mutually ramp up via a kind of step-ladder effect in which each person believes the other one hit them harder. Shergill et al describe experiments that suggest that in such cases each person is truthfully reporting their own sensations, but that those sensations are skewed by the attenuating effects of self-prediction. Thus, ‘self-generated forces are perceived as weaker than externally generated forces of equal magnitude.’

This also explains why you can’t tickle yourself – your body predicts and adjusts away your own actions, leaving only an attenuated version.

7. The Placebo Effect. We hear a lot about “pain gating” in the spine, but the PP model does a good job of explaining what this is: adjusting pain based on top-down priors. If you believe you should be in pain, the brain will use that as a filter to interpret ambiguous low-precision pain signals. If you believe you shouldn’t, the brain will be more likely to assume ambiguous low-precision pain signals are a mistake. So if you take a pill that doctors assure you will cure your pain, then your lower layers are more likely to interpret pain signals as noise, “cook the books” and prevent them from reaching your consciousness.

Psychosomatic pain is the opposite of this; see Section 7.10 of the book for a fuller explanation.

8. Asch Conformity Experiment. More speculative, and not from the book. But remember this one? A psychologist asked subjects which lines were the same length as other lines. The lines were all *kind of* similar lengths, but most subjects were still able to get the right answer. Then he put the subjects in a group with confederates; all of the confederates gave the same wrong answer. When the subject's turn came, usually they would disbelieve their eyes and give the same wrong answer as the confederates.

The bottom-up stream provided some ambiguous low-precision bottom-up evidence pointing toward one line. But in the final Bayesian computation, those were swamped by the strong top-down prediction that it would be another. So the middle layers "cooked the books" and replaced the perceived sensation with the predicted one. From Wikipedia:

> Participants who conformed to the majority on at least 50% of trials reported reacting with what Asch called a "distortion of perception". These participants, who made up a distinct minority (only 12 subjects), expressed the belief that the confederates' answers were correct, and were apparently unaware that the majority were giving incorrect answers.

9. Neurochemistry. PP offers a way to a psychopharmacological holy grail – an explanation of what different neurotransmitters really *mean*, on a human-comprehensible level. Previous attempts to do this, like "dopamine represents reward, serotonin represents calmness", have been so wildly inadequate that the whole question seems kind of disreputable these days.

But as per PP, the NMDA glutamatergic system mostly carries the top-down stream, the AMPA glutamatergic system mostly carries the bottom-up stream, and dopamine mostly carries something related to precision, confidence intervals, and surprisal levels. This matches a lot of observational data in a weirdly consistent way – for example, it doesn't take a lot of imagination to think of the slow, hesitant movements of Parkinson's disease as having "low motor confidence".

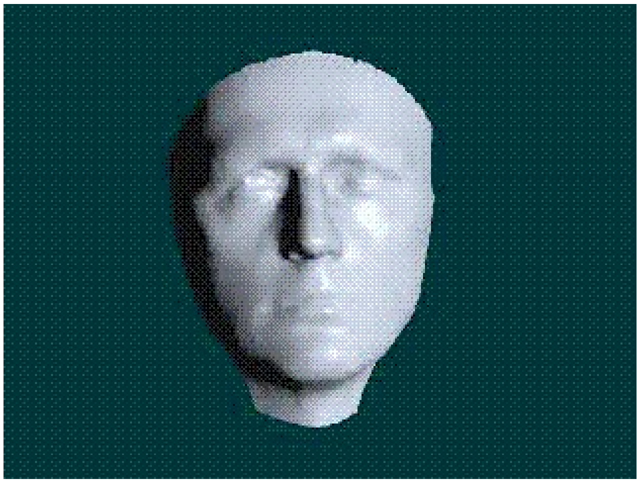
10. Autism. Various research in the PP tradition has coalesced around the idea of autism as an unusually high reliance on bottom-up rather than top-down information, leading to "weak central coherence" and constant surprisal as the sensory data fails to fall within pathologically narrow confidence intervals.

Autistic people classically can't stand tags on clothing – they find them too scratchy and annoying. Remember the example from Part III about how you successfully predicted away the feeling of the shirt on your back, and so manage never to think about it when you're trying to concentrate on more important things? Autistic people can't do that as well. Even though they have a layer in their brain predicting "will continue to feel shirt", the prediction is too precise; it predicts that next second, the shirt will produce *exactly* the same pattern of sensations it does now. But realistically as you move around or catch passing breezes the shirt will change ever so slightly – at which point autistic people's brains will send alarms all the way up to consciousness, and they'll perceive it as "my shirt is annoying".

Or consider the classic autistic demand for routine, and misery as soon as the routine is disrupted. Because their brains can only make very precise predictions, the slightest disruption to routine registers as strong surprisal, strong prediction failure, and "oh no, all of my models have failed, nothing is true, anything is possible!" Compare to a neurotypical person in the same situation, who would just relax their confidence intervals a little bit and say "Okay, this is basically 99% like a normal day, whatever". It would take something genuinely unpredictable – like being thrown on an unexplored continent or something – to give these people the same feeling of surprise and unpredictability.

This model also predicts autistic people's strengths. We know that polygenic risk for autism is positively associated with IQ. This would make sense if the central feature of autism was a sort of increased mental precision. It would also help explain why autistic people seem to excel in high-need-for-precision areas like mathematics and computer programming.

11. Schizophrenia. Converging lines of research suggest this also involves weak priors, apparently at a different level to autism and with different results after various compensatory mechanisms have had their chance to kick in. One especially interesting study asked neurotypicals and schizophrenics to follow a moving light, much like the airplane video in Part III above. When the light moved in a predictable pattern, the neurotypicals were much better at tracking it; when it was a deliberately perverse video specifically designed to frustrate expectations, the schizophrenics actually did better. This suggests that neurotypicals were guided by correct top-down priors about where the light would be going; schizophrenics had very weak priors and so weren't really guided very well, but also didn't screw up when the light did something unpredictable. Schizophrenics are also famous for not being fooled by the "hollow mask" (below) and other illusions where top-down predictions falsely constrain bottom-up evidence. My guess is they'd be more likely to see both 'the's in the "PARIS IN THE THE SPRINGTIME" image above.



The exact route from this sort of thing to schizophrenia is really complicated, and anyone interested should check out Section 2.12 and the whole of Chapter 7 from the book. But the basic story is that it creates waves of anomalous prediction error and surprisal, leading to the so-called "delusions of significance" where schizophrenics believe that eg the fact that someone is wearing a hat is some sort of incredibly important cosmic message. Schizophrenics' brains try to produce hypotheses that explain all of these prediction errors and reduce surprise – which is impossible, because the prediction errors are random. This results in incredibly weird hypotheses, and eventually in schizophrenic brains being willing to ignore the bottom-up stream entirely – hence hallucinations.

All this is treated with antipsychotics, which antagonize dopamine, which – remember – represents confidence level. So basically the medication is telling the brain "YOU CAN IGNORE ALL THIS PREDICTION ERROR, EVERYTHING YOU'RE PERCEIVING IS TOTALLY GARBAGE SPURIOUS DATA" – which turns out to be exactly the message it needs to hear.

An interesting corollary of all this – because all of schizophrenics' predictive models are so screwy, they lose the ability to use the "adjust away the consequences of your own actions" hack discussed in Part 5 of this section. That means their own actions *don't* get predicted out, and seem like the actions of a foreign agent. This is why they get so-called "delusions of agency", like "the government beamed that thought into my brain" or "aliens caused my arm to move just now". And in case you were wondering – yes, schizophrenics can tickle themselves.

12. Everything else. I can't possibly do justice to the whole of *Surfing Uncertainty*, which includes sections in which it provides lucid and compelling PP-based explanations of hallucinations, binocular rivalry, conflict escalation, and various optical illusions. More speculatively, I can think of really interesting connections to things like phantom limbs, creativity (and its association with certain mental disorders), depression, meditation, etc, etc, etc.

The general rule in psychiatry is: if you think you've found a theory that explains everything, diagnose yourself with mania and check yourself into the hospital. Maybe I'm not at that point yet – for example, I don't think PP does anything to explain what mania itself is. But I'm pretty close.

IV.

This is a really poor book review of *Surfing Uncertainty*, because I only partly understood it. I'm leaving out a *lot* of stuff about the motor system, debate over philosophical concepts with names like "enactivism", descriptions of how neurons form and unform coalitions, and of course a hundred pages of apologia along the lines of "this may not *look* embodied, but if you squint you'll see how super-duper embodied it really is!". As I reread and hopefully come to understand some of this better, it might show up in future posts.

But speaking of philosophical debates, there’s one thing that really struck me about the PP model.

Voodoo psychology suggests that culture and expectation tyrannically shape our perceptions. Taken to an extreme, objective knowledge is impossible, since all our sense-data is filtered through our own bias. Taken to a *very far* extreme, we get things like What The !@#$ Do We Know?‘s claim that the Native Americans literally couldn’t see Columbus’ ships, because they had no concept of “caravel” and so the percept just failed to register. This sort of thing tends to end by arguing that science was invented by straight white men, and so probably just reflects straight white maleness, and so we should ignore it completely and go frolic in the forest or something.

Predictive processing is sympathetic to all this. It takes all of this stuff like priming and the placebo effect, and it predicts it handily. But it doesn’t give up. It (theoretically) puts it all on a sound mathematical footing, explaining exactly how *much* our expectations should shape our reality, and in which *ways* our expectation should shape our reality. I feel like someone armed with predictive processing and a bit of luck should have been able to predict that placebo effect and basic priming would work, but stereotype threat and social priming wouldn’t. Maybe this is total retrodictive cheating. But I feel like it should be possible.

If this is true, it gives us more confidence that our perceptions should correspond – at least a little – to the external world. We can accept that we may be misreading “PARIS IN THE THE SPRINGTIME” while remaining confident that we *wouldn’t* misread “PARIS IN THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE THE

THE THE THE THE THE THE THE THE THE THE SPRINGTIME" as containing only one "the". Top-down processing very occasionally meddles in bottom-up sensation, but (as long as you're not schizophrenic), it sticks to an advisory role rather than being able to steamroll over arbitrary amounts of reality.

The rationalist project is overcoming bias, and that requires both an admission that bias is possible, and a hope that there's something *other* than bias which we can latch onto as a guide. Predictive processing gives us more confidence in both, and helps provide a convincing framework we can use to figure out what's going on at all levels of cognition.

Predictive Processing And Perceptual Control

Yesterday's review of *Surfing Uncertainty* mentioned how predictive processing attributes movement to strong predictions about proprioceptive sensations. Because the brain tries to minimize predictive error, it moves the limbs into the positions needed to produce those sensations, fulfilling its own prophecy.

This was a really difficult concept for me to understand at first. But there were a couple of passages that helped me make an important connection. See if you start thinking the same thing I'm thinking:

> To make [bodily] action come about, the motor plant behaves (Friston, Daunizeau, et al, 2010) in ways that cancel out proprioceptive prediction errors. This works because the proprioceptive prediction errors signal the difference between how the bodily plant is currently disposed and how it would be disposed were the desired actions being performed. Proprioceptive prediction error will yield (moment-by-moment) the projected proprioceptive inputs. In this way, predictions of the unfolding proprioceptive patterns that would be associated with the performance of some action actually bring that action about. This kind of scenario is neatly captured by Hawkins and Blakeslee (2004), who write that: "As strange as it sounds, when your own behavior is involved, your predictions not only precede sensation, they determine sensation."

And:

> PP thus implements the distinctive circular dynamics described by Cisek and Kalaska using a famous quote from the American pragmatist John Dewey. Dewey rejects the ‘passive’ model of stimuli evoking responses in favour of an active and circular model in which ‘the motor response determines the stimulus, just as truly as sensory stimulus determines movement’

Still not getting it? What about:

> According to active inference, the agent moves body and sensors in ways that amount to actively seeking out the sensory consequences that their brains expect.

This is the model from Will Powers’ Behavior: The Control Of Perception.

Clark knows this. A few pages after all these quotes, he writes:

> One signature of this kind of grip-based non-reconstructive dance is that it suggests a potent reversal of our ordinary way of thinking about the relations between perception and action. Instead of seeing perception as the control of action, it becomes fruitful to think of action as the control of perception [Powers 1973, Powers et al, 2011].

But I feel like this connection should be given more weight. Powers’ perceptual control theory presages predictive processing theory in a lot of ways. In particular, both share the idea of cogntitive “layers”, which act at various levels (light-intensity-detection vs. edge-detection vs. object-detection, or movements vs. positions-in-space vs. specific-muscle-actions vs. specific-muscle-fiber-tensions). Upper layers decide what stimuli they want lower levels to be perceiving, and lower layers arrange themselves in the way that produce those stimuli. PCT talks about “set points” for cybernetic systems, and PP talks about “predictions”, but they both seem to be groping at the same thing.

I was least convinced by the part of PCT which represented the uppermost layers of the brain as control systems controlling various quantities like “love” or “communism”, and which sometimes seemed to veer into self-parody. PP offers an alternative by describing those layers as making predictions (sometimes “active predictions” of the sort that guide behavior) and trying to minimize predictive error. This allows lower level systems to “control for” deviation from a specific plan, rather than just monitoring the amount of some scalar quantity.

My review of *Behavior: The Control Of Perception* ended by saying:

> It does seem like there’s something going on where my decision to drive activates a lot of carefully-trained subsystems that handle the rest of it automatically, and that there’s probably some neural correlate to it. But I don’t know whether control systems are the right way to think about this… I think maybe there are some obvious parallels, maybe even parallels that bear fruit in empirical results, in lower level systems like motor control. Once you get to high-level systems like communism or social desirability, I’m not sure we’re doing much better than [strained control-related metaphors].

I think my instincts were right. PCT is a good model, but what’s good about it is that it approximates PP. It approximates PP best at the lower levels, and so is most useful there; its thoughts on the higher levels remain useful but start to diverge and so become less profound.

The Greek atomists like Epicurus have been totally superseded by modern atomic theory, but they still get a sort of "how did they do that?" award for using vague intuition and good instincts to cook up a scientific theory that couldn't be proven or universally accepted until centuries later. If PP proves right, then Will Powers and PCT deserve a place in the pantheon besides them. There's something kind of wasteful about this – we can't properly acknowledge the cutting-edgeness of their contribution until it's obsolete – but at the very least we can look through their other work and see if they've got even *more* smart ideas that might be ahead of their time.

(Along with his atomic theory, Epicurus gathered a bunch of philosophers and mathematicians into a small cult around him, who lived together in co-ed group houses preaching atheism and materialism and – as per the rumors – having orgies. If we'd just agreed he was right about everything from the start, we wouldn't have had to laboriously reinvent his whole system.)

How Do We Get Breasts Out Of Bayes Theorem?

[Epistemic status: I guess instincts clearly exist, so take this post more as an expression of confusion than as a claim that they don't.]

Predictive processing isn't necessarily blank-slatist. But its focus on building concepts out of attempts to generate/predict sense data poses a problem for theories of innate knowledge. PP is more comfortable with deviations from a blank slate that involve the *rules of* cognition than with those that involve the *contents of* cognition.

For example, the theory shouldn't mind the existence of genes for IQ. If the brain works on Bayesian math, some brains might be able to do the calculations more effectively than others. It shouldn't even mind claims like "girls are more emotional than boys" – that's just a question of how different hormones affect the Bayesian weighting of logical vs. emotional input.

But evolutionary psychologists make claims like "Men have been evolutionarily programmed to like women with big breasts, because those are a sign of fertility." Forget for a second whether this is politically correct, or cross-culturally replicable, or anything like that. From a neurological point of view, *how could this possibly work?*

In Clark's version of PP, infants laboriously construct all their priors out of sensory evidence. Object permanence takes months. Sensory coordination – the belief that eg the auditory and visual streams describe the same world, so that the same object might be both visible and producing sound – is not assumed. Clark even flirts with the possibility that some *really* basic assumptions might be learned:

> Plausibly, it is only because the world we encounter must be parsed for action and intervention that we encounter, in experience, a relatively unambiguous determinate world at all. Subtract the need for action and the broadly Bayesian framework can seem quite at odds with the phenomenal facts about conscious perceptual experience: our world, it might be said, does not look as if it is encoded in an intertwined set of probability density distributions. Instead, it looks unitary and, on a clear day, unambiguous…biological systems, as mentioned earlier, may be informed by a variety of learned or innate "hyperpriors" concerning the general nature of the world. One such hyperprior might be that the world is usually in one determinate state or another.

I realize he's not coming out and *saying* that maybe babies see the world as a probability distribution over hypotheses and only gradually "figure out" that a determinate world is more pragmatic. But he's sure coming closer to saying that than anybody else I know.

In any case, we work up from these sorts of deep hyperpriors to testing out new models and ideas. Presumably we eventually gain concepts like "breast" after a lot of trial-and-error in which we learn that they generate successful predictions about the sensory world.

In this model, the evolutionary psychological theory seems like a confusion of levels. How do our genes reach out and grab this particular high-level category in the brain, "breast", to let us know that we're programmed to find it attractive?

To a first approximation, all a gene does is code for a protein. How, exactly, do you design a protein that makes men find big-breasted women attractive? I mean, I can sort of imagine that if you know what neurons carry the concept of “breast”, you can sort of wire them up to whatever region of the hypothalamus handles sexual attraction, so that whenever you see breasts you feel attraction. But number one, are you sure there’s a specific set of neurons that carry the concept “breast”? And number two, how do you get those neurons (and no others) to express a certain gene?

And if you want to posit an entire complicated breast-locating system made up of hundreds of genes, remember that we only have about 20,000 genes total. Most of these are already involved in doing things like making the walls of lysosomes flexible enough or something really boring like that. Really it’s a miracle that a mere 20,000 genes can make a human at all. So how many of these precious resources do you want to take up constructing some kind of weird Rube-Goldbergesque breast-related brain circuit?

The only excuse I can think of for the evo psych perspective is that it obviously works *sometimes*. Animals do have instincts; it can’t be learning all the way down.

Sometimes when we really understand those instincts, they *do* look like weird Rube Goldberg contraptions made of brain circuits. The classic example is baby gulls demanding food from their mother. Adult gulls have a red dot on their beaks, and the baby bird algorithm seems to be “The first thing you see with a red dot is your mother; demand food from her.” Maybe “red dot” is primitive enough that it’s easier to specify genetically than “thing that looks like a mother bird”?

The clearest example I can think of where animals clearly have an instinctive understanding of a high level concept is sex/gender – a few gay humans and penguins aside, Nature seems pretty good at keeping its creatures heterosexual.

But this is one of the rare cases where evolution might really want to devote some big fraction of the 20,000 genes it has to work with to building a Rube Goldberg circuit.

Also, maybe we *shouldn’t* set those few gender-nonconforming humans aside. Remember, autistic people have some kind of impairment in top-down prior-based processing relative to the bottom-up evidence-based kind, and they’re about eight times more likely to be trans than the general population. It sure looks like there’s some kind of process in which people have to infer their gender. And even though evolution seems to be shouting some really loud hints, maybe if you weigh streams of evidence in unusual ways you can end up somewhere unexpected. Evolution may be able to bias the process or control its downstream effects, but it doesn’t seem able to literally hard-code it.

Someone once asked me how to distinguish between good and bad evolutionary psychology. One heuristic might be to have a strong prior against any claim in which genes can just reach into the level of already-formed concepts and tweak them around, unless there’s a *really* strong reason for evolution to go through a lot of trouble to make it happen.

Toward A Predictive Theory Of Depression

[Epistemic status: Total wild speculation]

I.

The predictive processing model offers compelling accounts of autism and schizophrenia. But *Surfing Uncertainty* and related sources I've read are pretty quiet about depression. Is there a possible PP angle here?

Chekroud (2015) has a paper trying to apply the model to depression. It's scholarly enough, and I found it helpful in figuring out some aspects of the theory I hadn't yet understood, but it's pretty unambitious. The overall thesis is something like "Predictive processing says high-level beliefs shape our low-level perceptions and actions, so maybe depressed people have some high-level depressing beliefs." Don't get me wrong, CBT orthodoxy is great and has cured millions of patients – but in the end, this is just CBT orthodoxy with a neat new coat of Bayesian paint.

There's something more interesting in Section 7.10 of *Surfing Uncertainty*, "Escape From The Darkened Room". It asks: if the brain works to minimize prediction error, isn't its best strategy to sit in a dark room and do nothing forever? After all, then it can predict its sense-data pretty much perfectly – it'll always just stay "darkened room".

Section 7.10 gives a kind of hand-wave-y answer here, saying that of course organisms have some drives, and probably it makes sense for them to desire novelty and explore new options, and so on. Overall this isn't too different from PCT's idea of "intrinsic error", and as long as we remember that it’s not really predicting anything in particular it seems like a fair response.

But I notice that this whole “sit in a dark room and never leave” thing sounds a lot like what depressed people say they wish they could do (and how the most severe cases of depression actually end up). Might there be a connection? Either a decrease in the mysterious intrinsic-error-style factors that counterbalance the dark room scenario, or an increase in the salience of prediction error that makes failures less tolerable?

(also, there’s one way to end all prediction error forever, and it’s something depressed people think about a lot)

II.

Corlett, Fritch, and Fletcher claim that an amphetamine-induced mania-like state may involve pathologically high confidence in neural predictions. I don’t remember if they took the obvious next step and claimed that depression was the opposite, but that sounds like another fruitful avenue to explore. So: what if depression is pathologically low confidence in neural predictions?

Chekroud’s theory of depression as high-level-depressing-beliefs bothers me because there are so many features of depression that aren’t cognitive or emotional or related to any of these higher-level functions at all. Depressed people move more slowly, in a characteristic pattern called “psychomotor retardation”. They display perceptual abnormalities. They’re more likely to get sick. There are lots of results like this.

Depression has to be about something more than just beliefs; it has to be something fundamental to the nervous system. And low confidence in neural predictions would do it. Since neural predictions are the basic unit of thought, encoding not just perception but also motivation, reward, and even movement – globally low confidence levels would have devastating effects on a whole host of processes.

Perceptually, they would make sense-data look less clear and distinct. Depressed people describe the world as gray, washed-out, losing its contrast. *This is not metaphorical*. You can do psychophysical studies on color perception in depressed people, you can stick electrodes on their eyeballs, and all of this will tell you that depressed people literally see the world in washed-out shades of gray. Descriptions of their sensory experience sound intuitively like the sensory experience you would get if all your sense organs were underconfident in their judgments.

Mechanically, they would make motor movements less forceful. Remember, in PP movements are "active inferences" – the body predicts that the limb it wants to move is somewhere else, then counts on the motor system's drive toward minimizing prediction error to do the rest. If you predictions are underconfident, your movements are insufficiently forceful and you get the psychomotor retardation that all the pathologists describe in depressed people. And what's the closest analog to depressive psychomotor retardation? Parkinsonian bradyphrenia. What causes Parkinsonian bradyphrenia? We know the answer to this one – insufficient dopamine, where dopamine is known to encode the confidence level of motor predictions.

Motivationally – well, I'm less certain, I still haven't found a good predictive processing account of motivation I understand on an intuitive level. But if we draw the analogy to perceptual control theory, some motivations (like hunger) are probably a kind of "intrinsic error" that can be modeled as higher-level processes feeding reference points to lower-level control systems. If we imagine the processes predicting eg hunger, then predicting with low confidence sure sounds like the sort of thing where you should be less hungry. If they're predicting "you should get out of bed", then predicting that with low confidence sure sounds like the sort of thing where you don't feel a lot of motivation to get out of bed.

I'm hesitant to take "low self-confidence" as a gimme – it seems relying too much on a trick of the English language. But I think there really is a connection. Suppose that you're taking a higher-level math class and you're really bad at it. No matter how hard you study, you always find the material a bit confusing and are unsure whether you're applying the concepts correctly. Eventually you start feeling kind of like a loser, you decide the math class isn't for you, and you move on to something else where you're more talented. Your low confidence in your beliefs (eg answers to test questions) and actions (eg problem-solving strategies) create general low self-confidence and feelings of worthlessness. Eventually you decide math isn't for you and decide to drop the class.

If you have global low confidence, the world feels like a math class you don't understand that you can't escape from. This feeling might be totally false – you might be getting everything right – but you still feel that way. And there's no equivalent to dropping out of the math class – except committing suicide, which is how far too many depressed people end up.

One complicating factor – how do we explain depressed people's frequent certainty that they'll fail? A proper Bayesian, barred from having confident beliefs about anything, will be maximally uncertain about whether she'll fail or succeed – but some depressed people have really strong opinions on this issue. I'm not really sure about this, and admit it's a point against this theory. I can only appeal to the math class example again – if there was a math class where I just had no confidence about anything I thought or said, I would probably be pretty sure I'd fail there too.

(just so I'm not totally just-so-storying here, here's a study of depressed people's probability calibration, which shows that – yup – they're underconfident!)

This could tie into the "increased salience of prediction error" theory in Part I. If for some reason the brain became "overly conservative" – if it assigned very high cost to a failed prediction relative to the benefit of a successful prediction – then it would naturally lower its confidence levels in everything, the same way a very conservative better who can't stand losing money is going to make smaller bets.

III.

But why would low confidence cause sadness?

Well, what, really, *is* emotion?

Imagine the world's most successful entrepreneur. Every company they found becomes a multibillion-dollar success. Every stock they pick shoots up and never stops. Heck, even their personal life is like this. Every vacation they take ends out picture-perfect and creates memories that last a lifetime; every date they go on leads to passionate soul-burning love that never ends badly.

And imagine your job is to advise this entrepreneur. The only advice worth giving would be "do more stuff". Clearly all the stuff they're doing works, so aim higher, work harder, run for President. Another way of saying this is "be more self-confident” – if they’re doubting whether or not to start a new project, remind them that 100% of the things they’ve ever done have been successful, odds are pretty good this new one will too, and they should stop wasting their time second-guessing themselves.

Now imagine the world’s most unsuccessful entrepreneur. Every company they make flounders and dies. Every stock they pick crashes the next day. Their vacations always get rained-out, their dates always end up with the other person leaving halfway through and sticking them with the bill.

What if your job is advising this guy? If they’re thinking of starting a new company, your advice is “Be really careful – you should know it’ll probably go badly”. If they’re thinking of going on a date, you should warn them against it unless they’re *really* sure. A good global suggestion might be to aim lower, go for low-risk-low-reward steady payoffs, and wait on anything risky until they’ve figured themselves out a little bit more.

Corlett, Frith and Fletcher linked mania to increased confidence. But mania looks a lot like being happy. And you’re happy when you succeed a lot. And when you succeed a lot, maybe having increased confidence is the way to go. If happiness were a sort of global filter that affected all your thought processes and said “These are good times, you should press really hard to exploit your apparent excellence and not worry too much about risk”, that would be pretty evolutionarily useful. Likewise, if sadness were a way of saying “Things are going pretty badly, maybe be less confidence and don’t start any new projects”, that would be useful too.

Depression isn’t normal sadness. But if normal sadness lowers neural confidence a little, maybe depression is the pathological result of biological processes that lower neural confidence. To give a total fake example which I’m not saying is what actually happens, if you run out of whatever neurotransmitter you use to signal high confidence, that would give you permanent pathological low confidence and might look like depression.

One problem with this theory is the time course. Sure, if you’re eternally successful, you should raise your confidence. But eternally successful people are rarely eternally happy. If we’re thinking of happiness-as-felt-emotion,itt seems more like they’re happy for a few hours after they win an award or make their first million or whatever, then go back down to baseline. I’m not sure it makes sense to start lots of new projects in the hour after you win an award.

One way of resolving this: maybe happiness is the derivative of neural confidence? It’s the feeling of your confidence levels increasing, the same way acceleration is the feeling of your speed increasing?

Of course, that’s three layers of crackpot – its own layer, under the layer of emotions as confidence level, under the layer of depression as change in prediction strategies. Maybe I should dial back my *own* confidence levels and stop there.

Classified Thread 3: Semper Classifiedelis

This is the…monthly? bimonthly? occasional?…classified thread. Post advertisements, personals, and any interesting success stories from the last thread. Also:

1. Iacta_Procul, who posted about some of her life/mental health problems on the subreddit a few weeks ago, and who lots of people said they wanted a way to help, has decided to quit her dead-end job and try to start a math tutoring company. She has a Masters in math and offers to tutor any non-statistics undergrad mathematics, or any necessary test prep for the SAT/ACT/GRE/GMAT (including English/vocabulary/non-math sections). If you're interested, contact her on Wyzant.

2. Isak – who doesn't comment here much but is pretty active on Rationalist Tumblr Discord – is homeless right now, having trouble getting his disability check, and asking for some money to help stay afloat and get his life back on track. See his Fundly campaign page for more information.

3. ~~An old friend of mine is looking for AI/data science people in North Carolina who he can ask questions about the opportunities there. If that describes you, email me at scott [at] shireroth [dot] org and I can get you in touch.~~ (got enough responses; thanks to everyone who emailed)

Book Review: Mastering The Core Teachings Of The Buddha

I.

I always wanted to meditate more, but never really got around to it. And (I thought) I had an unimpeachable excuse. The demands of a medical career are incompatible with such a time-consuming practice.

Enter Daniel Ingram MD, an emergency physician who claims to have achieved enlightenment just after graduating medical school. His book is called *Mastering The Core Teachings Of The Buddha*, but he could also have called it *Buddhism For ER Docs*. ER docs are famous for being practical, working fast, and thinking everyone else is an idiot. MCTB delivers on all three counts. And if you've ever had an attending quiz you on the difference between type 1 and type 2 second-degree heart block, you'll love Ingram's taxonomy of the stages of enlightenment.

The result is a sort of perfect antidote to the vague hippie-ism you get from a lot of spirituality. For example, from page 324:

> I feel the need to address, which is to say shoot down with every bit of rhetorical force I have, the notion promoted by some teachers and even traditions that there is nothing to do, nothing to accomplish, no goal to obtain, no enlightenment other than the ordinary state of being…which, if it were true, would have been very nice of them, except that it is complete bullshit. The Nothing To Do School and the You Are Already There School are both basically vile extremes on the same basic notion that all effort to attain to mastery is already

> missing the point, an error of craving and grasping. They both contradict the fundamental premise of this book, namely that there is something amazing to attain and understand and that there are specific, reproducible methods that can help you do that. Here is a detailed analysis of what is wrong with these and related perspectives…

…followed by a detailed analysis of what's wrong with this position, which he compared to "let[ting] a blind and partially paralyzed untrained stroke victim perform open-heart surgery on your child based on the notion that they are already an accomplished surgeon but just have to realize it".

This isn't to say that MCTB isn't a spiritual book, or that it shies away from mysticism or the transcendent. MCTB is very happy to discuss mysticism and the transcendent. It just quarantines the mystery within a carefully explained structure of rationally-arranged progress, so that it looks something like "and at square 41B in our perfectly rectangular grid you'll encounter a mind-state which is impossible to explain even in principle, here are a few woefully inadequate metaphors for this mind-state so you'll know when you've found it and should move on to square 41C."

This is a little jarring. But – Ingram argues – it's also very Buddhist. If you read the sutras with an open mind, the Buddha sounds a lot more like an ER doctor than a hippie. MCTB has a very Protestant fundamentalist feeling of digging through the exterior trappings of a religion to try to return to the purity of its origins. As far as I can tell, it succeeds – and in succeeding helped me understand Buddhism a whole lot better than anything else I've read.

II.

Ingram follows the Buddha in dividing the essence of Buddhism into three teachings: morality, concentration, and wisdom.

Morality seems like the odd one out here. Some Buddhists like to insist that Buddhism isn't really a "religion". It's less like Christianity or Islam than it is like (for example) high intensity training at the gym – a highly regimented form of practice that improves certain faculties if pursued correctly. Talking about "morality" makes this sound kind of hollow; nobody says you have to be a good person to get bigger muscles from lifting weights.

MCTB gives the traditional answer: you should be moral because it's the right thing to do, *but also* because it helps meditation. The same things that make you able to sleep at night with a clear mind make you able to meditate with a clear mind:

> One more great thing about the first training [morality] is that it really helps with the next training: concentration. So here's a tip: if you are finding it hard to concentrate because your mind is filled with guilt, judgment, envy or some other hard and difficult thought pattern, also work on the first training, kindness. It will be time well spent.

That leaves concentration (samatha) and wisdom (vipassana). You do samatha to get a powerful mind; you get a powerful mind in order do to vipassana.

Samatha meditation is the "mindfulness" stuff you're always hearing about: concentrate on the breath, don't let yourself get distracted, see if you can just attend to the breath and nothing else for minutes or hours. I read whole books about this before without understanding *why* it was supposed to be good, aside from vague things like “makes you feel more serene”. MCTB gives two reasons: first, it gets you into jhanas. Second, it prepares you for vipassana.

Jhanas are unusual mental states you can get into with enough concentration. Some of them are super blissful. Others are super tranquil. They’re not particularly meaningful in and of themselves, but they can give you heroin-level euphoria without having to worry about sticking needles in your veins. MCTB says, understatedly, that they can be a good encouragement to continue your meditation practice. It gives a taxonomy of eight jhanas, and suggests that a few months of training in samatha meditation can get you to the point where you can reach at least the first.

But the main point of samatha meditation is to improve your concentration ability so you can direct it to ordinary experience. Become so good at concentrating that you can attain various jhanas – but then, instead of focusing on infinite bliss or whatever other cool things you can do with your new talent, look at a wall or listen to the breeze or just try to understand the experience of existing in time.

This is vipassana (“insight”, “wisdom”) meditation. It’s a deep focus on the tiniest details of your mental experience, details so fleeting and subtle that without a samatha-trained mind you’ll miss them entirely. One such detail is the infamous “vibrations”, so beloved of hippies. Ingram notes that every sensation vibrates in and out of consciousness at a rate of between five and forty vibrations per second, sometimes speeding up or slowing down depending on your mental state. I’m a pathetic meditator and about as far from enlightenment as anybody in this world, but with enough focus even I have been able to confirm this to be true. And this is pretty close to the frequency of brain waves, which seems like a pretty interesting coincidence.

But this is just an example. The point is that if you really, *really* examine your phenomenological experience, you realize all sorts of surprising things. Ingram says that one early insight is a perception of your mental awareness of a phenomenon as separate from your perception of that phenomenon:

> This mental impression of a previous sensation is like an echo, a resonance. The mind takes a crude impression of the object, and that is what we can think about, remember, and process. Then there may be a thought or an image that arises and passes, and then, if the mind is stable, another physical pulse. Each one of these arises and vanishes completely before the other begins, so it is extremely possible to sort out which is which with a stable mind dedicated to consistent precision and not being lost in stories. This means the instant you have experienced something, you know that it isn't there any more, and whatever is there is a new sensation that will be gone in an instant. There are typically many other impermanent sensations and impressions interspersed with these, but, for the sake of practice, this is close enough to what is happening to be a good working model.
>
> Engage with the preceding paragraphs. They are the stuff upon which great insight practice is based. Given that you know sensations are vibrating, pulsing in and out of reality, and that, for the sake of practice, every sensation is followed directly by a mental impression, you now know exactly what you are looking for. You have a clear standard. If you are not experiencing it,

then stabilize the mind further, and be clearer about exactly when and where there are physical sensations.

With enough of this work, you gain direct insight into what Buddhists call “the three characteristics”. The first is impermanence, and is related to all the stuff above about how sensations flicker and disappear. The second is called “unsatisfactoriness”, and involves the inability of any sensation to be fulfilling in some fundamental way. And the last is “no-self”, an awareness that these sensations don’t really cohere into the classic image of a single unified person thinking and perceiving them.

The Buddha famously said that “life is suffering”, and placed the idea of suffering – *dukkha* – as the center of his system. This *dukkha* is the same as the “unsatisfactoriness” above.

I always figured the Buddha was talking about life being suffering in the sense that sometimes you’re poor, or you’re sick, or you have a bad day. And I always figured that making money or exercising or working to make your day better sounded like a more promising route to dealing with this kind of suffering than any kind of meditative practice. Ingram doesn’t *disagree* that things like bad days are examples of *dukkha*. But he explains that this is something way more fundamental. Even if you were having the best day of your life and everything was going perfectly, if you slowed your mind down and concentrated perfectly on any specific atomic sensation, that sensation would include *dukkha*. *Dukkha* is part of the mental machinery.

MCTB acknowledges that all of this sounds really weird. And there are more depths of insight meditation, all sorts of weird things you notice when you look deep enough, that are even weirder. It tries to be very clear that nothing it’s writing about is going to make much sense in words, and that reading the words doesn’t really tell you very much. The only way to really make sense of it is to practice meditation.

When you understand all of this on a really fundamental level – when you’re able to tease apart every sensation and subsensation and subsubsensation and see its individual components laid out before you – then at some point your normal model of the world starts running into contradictions and losing its explanatory power. This is very unpleasant, and eventually your mind does some sort of awkward Moebius twist on itself, adopts a better model of the world, and becomes enlightened.

III.

The rest of the book is dedicated to laying out, in detail, all the steps that you have to go through before this happens. In Ingram’s model – based on but not identical to the various models in various Buddhist traditions – there are fifteen steps you have to go through before “stream entry” – the first level of enlightenment. You start off at the first step, after meditating some number of weeks or months or years you pass to the second step, and so on.

A lot of these are pretty boring, but Ingram focuses on the fourth step, Arising And Passing Away. Meditators in this step enter what sounds like a hypomanic episode:

> In the early part of this stage, the meditator’s mind speeds up more and more quickly, and reality begins to be perceived as particles or fine vibrations of mind and matter, each arising and vanishing utterly at tremendous speed…As this stage deepens and matures, meditators let go of even the high levels of clarity and the other strong factors of meditation, perceive even

these to arise and pass as just vibrations, not satisfy, and not be self. They may plunge down into the very depths of the mind as though plunging deep underwater to where they can perceive individual frames of reality arise and pass with breathtaking clarity as though in slow motion […]

Strong sensual or sexual feelings and dreams are common at this stage, and these may have a non-discriminating quality that those attached to their notion of themselves as being something other than partially bisexual may find disturbing. Further, if you have unresolved issues around sexuality, which we basically all have, you may encounter aspects of them during this stage. This stage, its afterglow, and the almost withdrawal-like crash that can follow seem to increase the temptation to indulge in all manner of hedonistic delights, particularly substances and sex. As the bliss wears off, we may find ourselves feeling very hungry or lustful, craving chocolate, wanting to go out and party, or something like that. If we have addictions that we have been fighting, some extra vigilance near the end of this stage might be helpful.

This stage also tends to give people more of an extroverted, zealous or visionary quality, and they may have all sorts of energy to pour into somewhat idealistic or grand projects and schemes. At the far extreme of what can happen, this stage can imbue one with the powerful charisma of the radical religious leader.

Finally, at nearly the peak of the possible resolution of the mind, they cross something called “The Arising and Passing Event” (A&P Event) or “Deep Insight into the Arising and Passing Away”…Those who have crossed the A&P Event have stood on the ragged edge of reality and

> the mind for just an instant, and they know that awakening is possible. They will have great faith, may want to tell everyone to practice, and are generally evangelical for a while. They will have an increased ability to understand the teachings due to their direct and non-conceptual experience of the Three Characteristics. Philosophy that deals with the fundamental paradoxes of duality will be less problematic for them in some way, and they may find this fascinating for a time. Those with a strong philosophical bent will find that they can now philosophize rings around those who have not attained to this stage of insight. They may also incorrectly think that they are enlightened, as what they have seen was completely spectacular and profound. In fact, this is strangely common for some period of time, and thus may stop practicing when they have actually only really begun.
>
> This is a common time for people to write inspired dharma books, poetry, spiritual songs, and that sort of thing. This is also the stage when people are more likely to join monasteries or go on great spiritual quests. It is also worth noting that this stage can look an awful lot like a manic episode as defined in the DSM-IV (the current diagnostic manual of psychiatry). The rapture and intensity of this stage can be basically off the scale, the absolute peak on the path of insight, but it doesn’t last. Soon the meditator will learn what is meant by the phrase, “Better not to begin. Once begun, better to finish!”

If this last part sounds ominous, it probably should. If the fourth stage looks like a manic episode, the next five or six stages all look like some flavor of deep clinical depression. Ingram discusses several spiritual traditions and finds that they all warn of an uncanny valley halfway along the spiritual path; he himself adopts St. John’s phrase “Dark Night Of The Soul”. Once you have meditated enough to reach the A&P Event, you’re stuck in the (very unpleasant) Dark Night Of The Soul until you can meditate your way out of it, which could take months or years.

Ingram’s theory is that many people have had spiritual experiences without deliberately pursuing a spiritual practice – whether this be from everyday life, or prayer, or drugs, or even things you do in dreams. Some of these people *accidentally* cross the A&P Event, reach the Dark Night Of The Soul, and – not even knowing that the way out is through meditation – get stuck there for years, having nothing but a vague spiritual yearning and sense that something’s not right. He says that this is his own origin story – he got stuck in the Dark Night after having an A&P Event in a dream at age 15, was low-grade depressed for most of his life, and only recovered once he studied enough Buddhism to realize what had happened to him and how he could meditate his way out:

> When I was about 15 years old I accidentally ran into some of the classic early meditation experiences described in the ancient texts and my reluctant spiritual quest began. I did not realize what had happened, nor did I realize that I had crossed something like a point of no return, something I would later call the Arising and Passing Away. I knew that I had had a very strange dream with bright lights, that my entire body and world had seemed to explode like fireworks, and that afterwards I somehow had to find something, but I had no idea what that was. I philosophized frantically for years until I finally began to realize that no amount of thinking was going to solve my deeper spiritual issues

and complete the cycle of practice that had already started.

I had a very good friend that was in the band that employed me as a sound tech and roadie. He was in a similar place, caught like me in something we would later call the Dark Night and other names. He also realized that logic and cognitive restructuring were not going to help us in the end. We looked carefully at what other philosophers had done when they came to the same point, and noted that some of our favorites had turned to mystical practices. We reasoned that some sort of nondual wisdom that came from direct experience was the only way to go, but acquiring that sort of wisdom seemed a daunting task if not impossible […]

I [finally] came to the profound realization that they have actually worked all of this stuff out. Those darn Buddhists have come up with very simple techniques that lead directly to remarkable results if you follow instructions and get the dose high enough. While some people don’t like this sort of cookbook approach to meditation, I am so grateful for their recipes that words fail to express my profound gratitude for the successes they have afforded me. Their simple and ancient practices revealed more and more of what I sought. I found my experiences filling in the gaps in the texts and teachings, debunking the myths that pervade the standard Buddhist dogma and revealing the secrets meditation teachers routinely keep to themselves. Finally, I came to a place where I felt comfortable writing the book that I had been looking for, the book you now hold in your hands.

Once you meditate your way out of the Dark Night, you go through some more harrowing experiences, until you finally reach the fifteenth stage, Fruition, and achieve “stream entry” – the first level of enlightenment. Then you do it all again on a higher level, kind of like those video games where when you beat the game you get access to New Game+ . Traditionally it takes four repetitions of the spiritual path before you attain complete perfect enlightenment, but Ingram suggests this is metaphorical and says it took him approximately twenty-seven repetitions over seven years.

He also says – and here his usual lucidity deserted him and I ended up kind of confused – that once you’ve achieved stream entry, you’re going to be going down paths whether you like it or not – the “stream” metaphor is apt insofar as it suggests being borne along by a current. The rest of your life – even after you achieve complete perfect enlightenment – will be spent cycling through the fifteen stages, with each stage lasting a few days to months.

This seems pretty bad, since the stages look a lot like depression, mania, and other more arcane psychiatric and psychological problems. Even if you don’t mind the emotional roller coaster, a lot of them sound just plain *exhausting*, with your modes of cognition and perception shifting and coming into question at various points. MCTB offers some tips for dealing with this – you can always slow your progress down the path by gorging on food, refusing to meditate, and doing various other unspiritual things, but the whole thing lampshades a question that MCTB profoundly fails at giving anything remotely like an answer to:

IV.

Why would you want to do any of this?

The Buddha is supposed to have said: “I gained nothing whatsoever from Supreme Enlightenment, and for that reason it is called Supreme Enlightenment”. And sure, that’s the enigmatic Zen-sounding sort of statement we expect from our spiritual leaders. But if Buddhist practice is really difficult, and makes you perceive every single sensation as profoundly unsatisfactory in some hard-to-define way, and can plunge you into a neverending depression which you *might* get out of if you meditate hard enough, and then gives you a sort of permanent annoying low-grade bipolar disorder *even if you succeed*, then we’re going to need something better than pithy quotes.

Ingram dedicates himself *hard* to debunking a lot of the things people would use to fill the gap. Pages 261-328 discuss the various claims Buddhist schools have made about enlightenment, mostly to deny them all. He has nothing but contempt for the obviously silly ones, like how enlightened people can fly around and zap you with their third eyes. But he’s equally dismissive of things that sort of seem like the basics. He denies claims about how enlightened people can’t get angry, or effortlessly resist temptation, or feel universal unconditional love, or things like that. Some of this he supports with stories of enlightened leaders behaving badly; other times he cites himself as an enlightened person who frequently experiences anger, pain, and the like. Once he’s stripped everything else away, he says the only thing one can say about enlightenment is that it grants a powerful true experience of the non-dual nature of the world.

But still, *why would we want to get that*? I am super in favor of knowledge-for-knowledge’s-sake, but I’ve also read enough Lovecraft to have strong opinions about poking around Ultimate Reality in ways that tend to destroy your mental health.

The best Ingram can do is this:

> I realize that I am not doing a good job of advertising enlightenment here, particularly following my descriptions of the Dark Night. Good point. My thesis is that those who must find it will, regardless of how it is advertised. As to the rest, well, what can be said? Am I doing a disservice by not selling it like nearly everyone else does? I don’t think so. If you want grand advertisements for enlightenment, there is a great stinking mountain of it there for you partake of, so I hardly think that my bringing it down to earth is going to cause some harmful deficiency of glitz in the great spiritual marketplace.
>
> [Meditation teacher] Bill Hamilton had a lot of great one-liners, but my favorite concerned insight practices and their fruits, of which he said, “Highly recommended, can’t tell you why.” That is probably the safest and most accurate advertisement for enlightenment that I have ever heard.

V.

I was reading MCTB at the same time I read *Surfing Uncertainty*, and it was hard not to compare them. Both claim to be guides to the mysteries of the mind – one from an external scientific perspective, the other from an internal phenomenological perspective. Is there any way to link them up?

Remember this quote from *Surfing Uncertainty*?:

> Plausibly, it is only because the world we encounter must be parsed for action and intervention that we encounter, in experience, a relatively unambiguous determinate world at all. Subtract the need for action

> and the broadly Bayesian framework can seem quite at odds with the phenomenal facts about conscious perceptual experience: our world, it might be said, does not look as if it is encoded in an intertwined set of probability density distributions. Instead, it looks unitary and, on a clear day, unambiguous…biological systems, as mentioned earlier, may be informed by a variety of learned or innate “hyperpriors” concerning the general nature of the world. One such hyperprior might be that the world is usually in one determinate state or another.

Taken seriously, it suggests that some of the most fundamental factors of our experience are not real features of the sensory world, but very strong assumptions to which we fit sense-data in order to make sense of them. And Ingram’s theory of vipassana meditation looks a lot like concentrating really hard on our actual sense-data to try to disentangle them from the assumptions that make them cohere.

In the same way that our priors “snap” phrases like “PARIS IN THE THE SPRINGTIME” to a more coherent picture with only one “the”, or “snap” our saccade-jolted and blind-spot-filled visual world into a reasonable image, maybe they snap all of this vibrating and arising and passing away into something that looks like a permanent stable image of the world.

And in the same way that concentrating on “PARIS IN THE THE SPRINGTIME” really hard without any preconceptions lets you sniff out the extra “the”, so maybe enough samatha meditation lets you concentrate on the permanent stable image of the world until it dissolves into whatever the brain is actually doing. Maybe with enough dedication to observing reality as it really is rather than as you predict it to be, you can expose even the subjective experience of an observer as just a really strong hyperprior on all of the thought-and-emotion-related sense-data you're getting.

That leaves *dukkha*, this weird unsatisfactoriness that supposedly inheres in every sensation individually as well as life in general. If the goal of the brain is minimizing prediction error, if all of our normal forms of suffering like hunger and thirst and pain are just special cases of predictive error in certain inherent drives, then – well, this is a very fundamental form of badness which is inherent in all sensation and perception, and which a sufficiently-concentrated phenomenologist might be able to notice directly. Relevant? I'm not sure.

Mastering The Core Teachings Of The Buddha is a lucid guide to issues surrounding meditation practice and a good rational introduction to the Buddhist system. Parts of it are ultimately unsatisfactory, but apparently this is true of everything, so whatever.

Also available for free download here

Meditative States As Feedback Loops

Three years ago, in Going Loopy, I wrote:

> If the brain had been designed by an amateur, it would enter a runaway feedback loop the first time it felt an emotion. Think about it. You see a butterfly. This makes you happy. Being happy is an unexpected pleasant surprise. Now you're happy that you're happy. This makes you extra happy. Being extra happy is awesome! This makes you extra extra happy. And so on to as much bliss as your neurons are capable of representing. In the real world, either those feedback loops usually don't happen, or they converge and stop at some finite point. I would not be surprised to learn that a lot of evolutionary innovation and biochemical complexity goes into creating a strong barrier against conditioning on your own internal experience.

"Evolutionary innovation and biochemical complexity"? Haha no, people are just too distractable to keep having the same emotion for more than a couple seconds.

I get this from Leigh Brasington's excellent *Right Concentration*, a Buddhist perspective on various advanced meditative states called *jhanas*. To get to the first of these jhanas (there are eight in all), you become really good at concentration meditation, until you can concentrate on your breath a long time without getting distracted. Then you concentrate on your breath for a long time. Then you take your one-pointed ultra-concentrated mind, and you notice (or generate, or imagine) a pleasant feeling. This produces the first jhana, which the Buddhist scriptures describe as:

> One drenches, steeps, saturates, and suffuses one's body with the rapture and happiness born of seclusion, so that there is no part of one's body that is not suffused by rapture and happiness.

Brasington backs this up with his own experience and those of other meditators he knows. The first jhana is really, really, *really* pleasurable; when you hear meditators talk about achieving "bliss states", it's probably something like the first jhana.

And here's the book's description of why it happens:

> When access concentration is firmly established, then you shift your attention from the breath (or whatever your meditation object is) to a pleasant sensation. You put your attention on that sensation, and maintain your attention on that sensation, and do nothing else…
>
> What you are attempting to do is set up a positive feedback loop. An example of a positive feedback loop is that awful noise a speaker will make if a microphone is held too close to it. What's happening is that the ambient noise in the room goes into the microphone, is amplified by the amplifier, and comes out the speaker louder. It then reenters the microphone, gets amplified even more, comes out louder still, goes into the microphone yet again, and so on. You are trying to do exactly the same thing, except, rather than a positive feedback loop of noise, you are attempting to generate a positive feedback loop of pleasure. You hold your attention on a pleasant sensation. That feels nice, adding a bit more pleasure to your overall experience. That addition is also pleasurable, adding more pleasure, and so on, until, instead of getting a horrible noise, you get an explosion of pleasure.

The book doesn’t come out and say that the other seven jhanas are the same thing, but that seems consistent with the descriptions. For example, the fourth jhana is a state of ultimate calm. Seems like maybe if you become calm, then being so calm is kind of calming, and that’s even more calming, and so on until you’ve maxed out your mental calmness-meter.

And the explanation of why this doesn’t happen all the time is that non-meditators just can’t concentrate hard enough. A microphone-amp system that turns on and off a couple of times each second will never get a really good feedback loop going. A mind that’s always flitting from one thing to another can’t build up enough *self-referentiality* to reach infinite bliss.

Against Individual IQ Worries

*[**Related to**: Attitude vs. Altitude]*

I.

I write a lot about the importance of IQ research, and I try to debunk pseudoscientific claims that IQ "isn't real" or "doesn't matter" or "just shows how well you do on a test". IQ is one of the best-studied ideas in psychology, one of our best predictors of job performance, future income, and various other forms of success, etc.

But every so often, I get comments/emails saying something like "Help! I just took an IQ test and learned that my IQ is x! This is much lower than I thought, and so obviously I will be a failure in everything I do in life. Can you direct me to the best cliff to jump off of?"

So I want to clarify: IQ is very useful and powerful *for research purposes*. It's not nearly as interesting *for you personally*.

How can this be?

Consider something like income inequality: kids from rich families are at an advantage in life; kids from poor families are at a disadvantage.

From a *research* point of view, it's really important to understand this is true. A scientific establishment in denial that having wealthy parents gave you a leg up in life would be an intellectual disgrace. Knowing that wealth runs in families is vital for even a minimal understanding of society, and anybody forced to deny that for political reasons would end up so hopelessly confused that they might as well just give up on having a coherent world-view.

From an *personal* point of view, coming from a poor family probably isn't great *but shouldn't be infinitely discouraging*. It doesn't suggest that some kid should think to herself "I come from a family that only makes $30,000 per year, guess that means I'm doomed to be a failure forever, might as well not even try". A poor kid is certainly at a disadvantage relative to a rich kid, but probably she knew that already long before any scientist came around to tell her. If she took the scientific study of intergenerational income transmission as something more official and final than her general sense that life was hard – if she obsessively recorded every raise and bonus her parents got on the grounds that it determined her own hope for the future – she would be giving the science more weight than it deserves.

So to the people who write me heartfelt letters complaining about their low IQs, I want to make two important points. First, we're not that good at measuring individual IQs. Second, individual IQs aren't that good at predicting things.

II.

Start with the measurement problems. People who complain about low IQs (not to mention people who boast about high IQs) are often wildly off about the number.

According to the official studies, IQ tests are rarely wrong. The standard error of measurement is somewhere between 3-7 points (1, 2, 3). Call it 5, and that means your tested IQ will only be off by 5+ points 32% of the time. It'll only be off by 10+ points 5% of the time, and really big errors should be near impossible.

In reality, I constantly hear about people getting IQ scores that don’t make any sense.

Here’s a pretty standard entry in the “help my IQ is so low” genre – Grappling With The Reality Of Having A Below Average IQ:

> When I was 16, as a part of an educational assessment, I took both the WAIS-IV and Woodcock Johnson Cognitive Batteries. My mother was curious as to why I struggled in certain subjects throughout my educational career, particularly in mathematical areas like geometry.
>
> I never got a chance to have a discussion with the psychologist about the results, so I was left to interpret them with me, myself, and the big I known as the Internet – a dangerous activity, I know. This meant two years to date of armchair research, and subsequently, an incessant fear of the implications of my below-average IQ, which stands at a pitiful 94…I still struggle in certain areas of comprehension. I received a score of 1070 on the SAT, (540 Reading & 530 Math), and am barely scraping by in my college algebra class. Honestly, I would be ashamed if any of my coworkers knew I barely could do high school-level algebra.

This person thinks they’re reinforcing their point by listing two different tests, but actually a 1070 on the SAT corresponds to about 104, a full ten points higher. Based on other things in their post – their correct use of big words and complicated sentence structure, their mention that they work a successful job in cybersecurity, the fact that they read a philosophy/psychology subreddit for fun – I’m guessing the 104 is closer to the truth.

From the comments on the same Reddit thread:

> Interesting, I hope more people who have an avg. or low IQ post. Personally I had an IQ of 90 or so, but the day of the test I stayed up almost the entire night, slept maybe two hours and as a naive caffeine user I had around 500 mg caffeine. Maybe low IQ people do that.
>
> I did IQTest.dk Raven's test on impulse after seeing a video of Peterson's regarding the importance of IQ, not in a very focused mode, almost ADHD like with rumination and I scored 108, but many claim low scores by around 0.5-1 SD, so that would put me in 115-123. I also am vegan, so creatine might increase my IQ by a few points. I think I am in the 120's, but low IQ people tend to overestimate their IQ, but at least I am certainly 108 non-verbally, which is pretty average and low.

The commenter is right that IQtest.dk usually underestimates scores compared to other tests. But even if we take it at face value, his first score was almost twenty points off. By the official numbers, that should only happen once in every 15,000 people. In reality, someone posts a thread about it on Reddit and another person immediately shows up to say "Yeah, that happened to me".

Nobel-winning physicist Richard Feynman famously scored "only" 124 on an IQ test in school – still bright, but nowhere near what you would expect of a Nobelist. Some people point out that it might have been biased towards measuring verbal rather than math abilities – then again, Feynman's autobiography (admittedly edited and stitched together by a ghostwriter) sold 500,000 copies and made the New York Times bestseller list. So either his tested IQ was off by at least 30 points (supposed chance of this happening: 1/505 million), or IQ isn't real and all of the studies showing that it is are made up by lizardmen to confuse us. In either case, you should be less concerned if your own school IQ tests seem kind of low.

I don’t know why there’s such a discrepancy between the official reliability numbers and the ones that anecdotally make sense. My guess is that the official studies give the tests better somehow. They use professional test administrators instead of overworked school counselors. They give them at a specific time of day instead of while the testee is half-asleep. They don’t let people take a bunch of caffeine before the test. They actually write the result down in a spreadsheet they have right there instead of trusting the testee to remember it accurately.

In my own field, official studies diagnose psychiatric diseases through beautiful Structured Clinical Interviews performed to exacting guidelines. Then real doctors diagnose them through checklists that say “DO NOT USE FOR DIAGNOSIS” in big letters on the top. If psychometrics is at all similar, the clashing numbers aren’t much of a mystery.

But two other points that might also be involved.

First, *on a population level* IQ is very stable with age. Over a study of 87,498 Scottish children, age 11 IQ and adult IQ correlated at 0.66, about as strong and impressive a correlation as you’ll ever find in the social sciences. But “correlation of 0.66” is also known as “only predicts 44% of the variance”. *On an individual level*, it is totally possible and not even that surprising to have an IQ of 100 at age 11 but 120 at age 30, or vice versa. Any IQ score you got before high school should be considered a *plausible prediction* about your adult IQ and nothing more.

Second, the people who get low IQ scores, are shocked, find their whole world tumbling in on themselves, and desperately try to hold on to their dream of being an intellectual – are not a representative sample of the people who get low IQ scores. The average person who gets a low IQ score says “Yup, guess that would explain why I’m failing all my classes”, and then goes back to beating up nerds. When you see someone saying “Help, I got a low IQ score, I’ve double-checked the standard deviation of all of my subscores and found some slight discrepancy but I’m not sure if that counts as Bayesian evidence that the global value is erroneous”, then, well – look, I wouldn’t be making fun of these people if I didn’t *constantly come across them*. You know who you are.

Just for fun, I analyzed the lowest IQ scores in my collection of SSC/LW surveys. I was only able to find three people who claimed to have an IQ ≤ 100 plus gave SAT results. All three had SAT scores corresponding to IQs in the 120s.

I conclude that at least among the kind of people I encounter and who tend to send me these emails, IQ estimates are pretty terrible.

This is absolutely consistent with population averages of thousands of IQ estimates still being valuable and useful research tools. It just means you shouldn’t use it *on yourself*. Statistics is what tells us that almost everybody feels stimulated on amphetamines. Reality is my patient who consistently goes to sleep every time she takes Adderall. Neither the statistics nor the lived experience are wrong – but if you use one when you need the other, you’re going to have a bad time.

III.

The second problem is that even if you avoid the problems mentioned above and measure IQ 100% correctly, *it's just not that usefully predictive*.

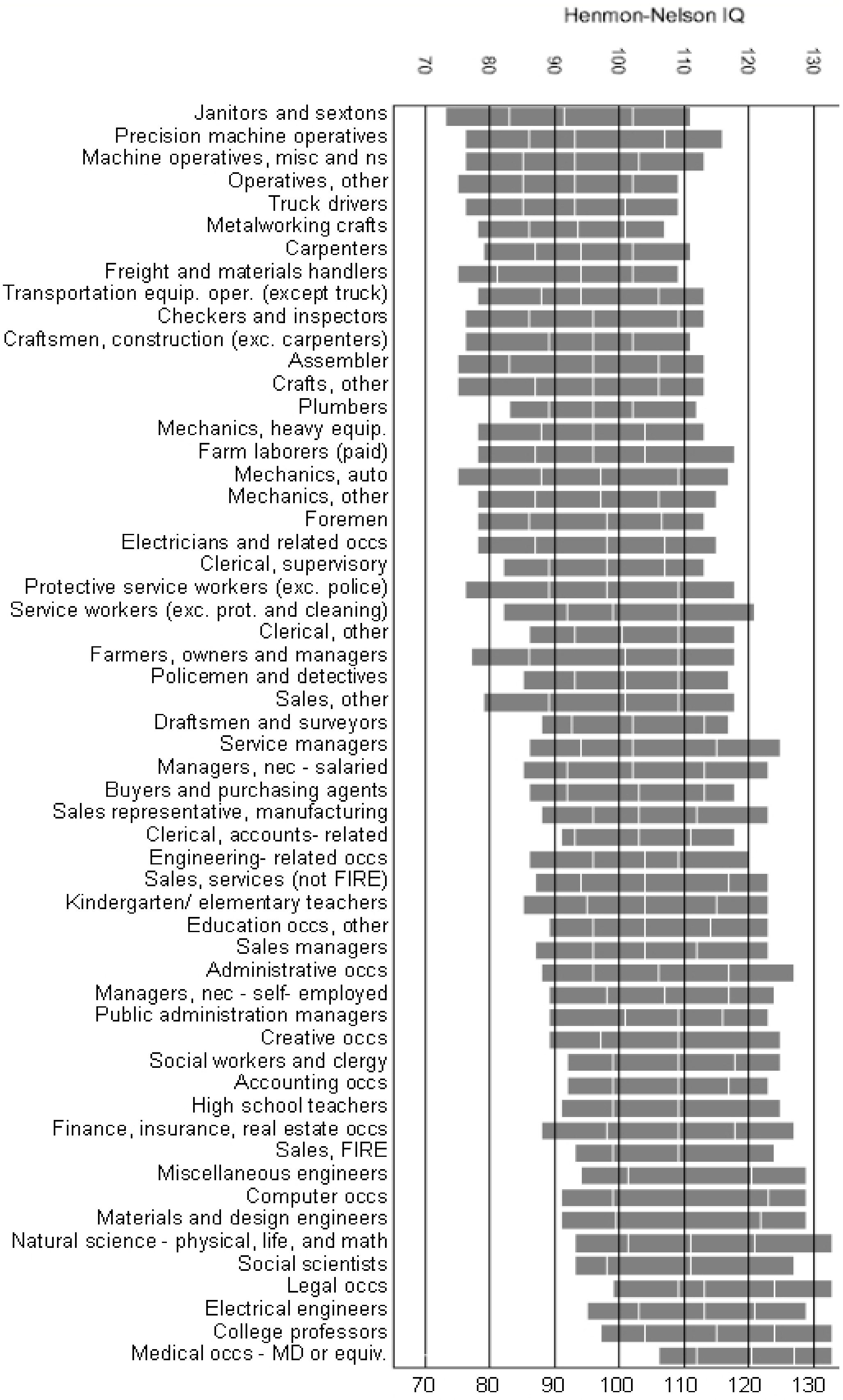
Isn't that heresy?! Isn't IQ the most predictive thing we have? Doesn't it affect every life outcome as proven again and again in well-replicated experiments?

Yes! I'm not denying any of that. I'm saying that *things that are statistically true aren't always true for any individual*.

Once again, consider the analogy to family transmission of income. Your parents' socioeconomic status correlates with your own at about r = 0.2 to 0.3, depending on how you define "socioeconomic status". By coincidence, this is pretty much the same correlation that Strenze (2006) found for IQ and socioeconomic status. Everyone knows that having rich parents is pretty useful if you want to succeed. But everyone also knows that rich parents aren't the only thing that goes into success. Someone from a poor family who tries really hard and gets a lot of other advantages still has a chance to make it. A sociologist or economist should be very interested in parent-child success correlations; the average person trying to get ahead should just shrug, realize things are going to be a little easier/harder than they would have been otherwise, and get on with their life.

And this isn't just about gaining success by becoming an athlete or musician or some other less-intellectual pursuit. Chess talent is correlated with IQ at 0.24, about the same as income. IQ is some complicated central phenomenon that contributes a little to every cognitive skill, but it doesn't entirely determine any cognitive skill. It's not just that you can have an average IQ and still be a great chess player if you work hard enough – that's true, but it's not *just* that. It's that you can have an average IQ and still have *high levels of* *innate talent in chess*. It’s not quite as likely as if you have a high IQ, but it’s very much in the range of possibility. And *then* you add in the effects of working hard enough, and *then* you’re getting somewhere.

Here is a table of professions by IQ, a couple of decades out of date but probably not too far off (cf. discussion here):



Percentiles 10, 25, 50, 75, and 90 are marked.

I don’t know how better to demonstrate this idea of “statistically solid, individually shaky”. On a population level, we see that the average doctor is 30 IQ points higher than the average janitor, that college professors are overwhelmingly high-IQ, and we think yeah, this is about what we would hope for from a statistic measuring intelligence. But on an individual level, we see that below-average IQ people sometimes become scientists, professors, engineers, and almost anything else you could hope for.

IV.

I’m kind of annoyed I have to write this post. After investing so much work debunking IQ denialists, I feel like this is really – I don’t know – diluting the brand.

But I actually think it’s not as contradictory as it looks, that there’s some common thread between my posts arguing that no, IQ isn’t fake, and this one.

If you really understand the idea of a statistical predictor – if you have that gear in your brain at a fundamental level – then social science isn’t scary. You can read about IQ, or heredity, or stereotypes, or gender differences, or whatever, and you can say – ah, there’s a slight tendency for one thing to correlate with another thing. Then you can go have dinner.

If you don’t get that, then the world is terrifying. Someone’s said that IQ “correlates with” life outcomes? What the heck is “correlate with”? Did they say that only high-IQ people can be successful? That you’re doomed if you don’t get the right score on a test?

And then you can either resist that with every breath you have – deny all the data, picket the labs where it’s studied, make up silly theories about “emotional intelligence” and “grit” and what have you. Or you can surrender to the darkness, at least have the comfort of knowing that you accept the grim reality as it is.

Imagine an American who somehow gets it into his head that the Communists are about to invade with overwhelming force. He might buy a bunch of guns, turn his house into a bunker, start agitating that Communist sympathizers be imprisoned to prevent them from betraying the country when the time came. Or he might hang a red flag from his house, wear a WELCOME COMMUNIST OVERLORDS tshirt, and start learning Russian. These seem like opposite responses, but they both come from the same fundamental misconception. A lot of the culture war – on both sides – seems like this. I don’t know how to solve this except to try, again and again, to install the necessary gear and convince people that correlations are neither meaningless nor always exactly 1.0.

So please: study the science of IQ. Use IQ to explain and predict social phenomena. Work on figuring out how to raise IQ. Assume that raising IQ will have far-ranging and powerful effects on a wide variety of social problems. Just don’t expect it to predict a single person’s individual achievement with any kind of reliability. *Especially* not yourself.

Different Worlds

I.

A few years ago I had lunch with another psychiatrist-in-training and realized we had totally different experiences with psychotherapy.

We both got the same types of cases. We were both practicing the same kinds of therapy. We were both in the same training program, studying under the same teachers. But our experiences were totally different. In particular, all her patients had dramatic emotional meltdowns, and all my patients gave calm and considered analyses of their problems, as if they were lecturing on a particularly boring episode from 19th-century Norwegian history.

I'm not bragging here. I *wish* I could get my patients to have dramatic emotional meltdowns. As per the textbooks, there should be a climactic moment where the patient identifies me with their father, then screams at me that I ruined their childhood, then breaks down crying and realizes that she loved her father all along, then ???, and then their depression is cured. I never got that. I tried, I even dropped some hints, like "Maybe this reminds you of your father?" or "Maybe you feel like screaming at me right now?", but they never took the bait. So I figured the textbooks were misleading, or that this was some kind of super-advanced technique, or that this was among the approximately 100% of things that Freud just pulled out of his ass.

And then I had lunch with my friend, and she was like "It's so stressful when all of your patients identify you with their parents and break down crying, isn't it? Don't you wish you could just go *one day* without that happening?"

And later, my supervisor was reviewing one of my therapy sessions, and I was surprised to hear him comment that I “seemed uncomfortable with dramatic expressions of emotion”. I mean, I *am* uncomfortable with dramatic expressions of emotion. I was just surprised he noticed it. As a therapist, I’m supposed to be quiet and encouraging and not show discomfort at anything, and I was trying to do that, and I’d thought I was succeeding. But apparently I was unconsciously projecting some kind of “I don’t like strong emotions, you’d better avoid those” field, and my patients were unconsciously complying.

I wish I could say my supervisor’s guidance fixed the problem and I learned to encourage emotional openness just as well as my colleague. But any improvement I made was incremental at best. My colleague is a bubbly extravert who gets very excited about everything; I worry that to match her results, I would have to somehow copy her entire personality.

But all was not lost. I found myself doing well with overly emotional patients, the sort who had *too many* dramatic meltdowns to do therapy with anybody else. With me, they tended to give calm and considered analyses of their problems, as if they were lecturing on a particularly boring episode from 19th-century Norwegian history. Everyone assumed that meant I was good at dealing with difficult cases, and must have read a bunch of books about how to defuse crises. I did nothing to disabuse them of this.

Then a few days ago I stumbled across the Reddit thread Has Anyone Here Ever Been To An LW/SSC Meetup Or Otherwise Met A Rationalist IRL? User dgerard wrote about meeting me in 2011, saying:

> His superpower is that he projects a Niceness Field, where people talking to him face to face want to be more polite and civil. The only person I've met with a similar Niceness Field is Jimmy Wales from Wikipedia… when people are around [Jimmy] talking to him they feel a sort of urge to be civil and polite in discourse I've seen people visibly trying to be very precise and polite talking to him about stuff even when they're quite upset about whatever it is. Scott has this too. It's an interesting superpower to observe.

I should admit nobody else has mentioned anything like this, and that narcissism biases me toward believing anyone who says I have a superpower. Still, it would explain a lot. And not necessarily in a good way. I've always believed psychodynamic therapies are mostly ineffective, and cognitive-behavioral therapies very effective, because all my patients seem to defy the psychodynamic mode of having having weird but emotionally dramatic reactions to things in their past, but conform effortlessly to the cognitive-behavioral mode of being able to understand and rationally discuss their problems. And the more I examine this, the more I realize that my results are pretty atypical for psychiatrists. There's something I'm doing – totally by accident – to produce those results. This is worrying not just as a psychiatrist, but as someone who wants to know anything about other people at all.

II.

New topic: paranoia and Williams Syndrome.

Paranoia is a common symptom of various psychiatric disorders – most famously schizophrenia, but also paranoid personality disorder, delusional disorder, sometimes bipolar disorder. You can also get it from abusing certain drugs –

marijuana, LSD, cocaine, and even prescription drugs like Adderall and Ritalin. The fun thing about paranoia is how gradual it is. Sure, if you abuse every single drug at once you'll think the CIA is after you with their mind-lasers. But if you just take a *little* more Adderall than you were supposed to, you'll be 1% paranoid. You'll have a very mild tendency to interpret ambiguous social signals just a little bit more negatively than usual. If a friend leaves without saying goodbye, and you would normally think "Oh, I guess she had a train to catch", instead you think "Hm, I wonder what she meant by that". There are a bunch of good stimulant abuse cases in the literature that present as "patient's boss said she was unusually standoffish and wanted her to get psychiatric evaluation", show up in the office as "well of course I'm standoffish, everyone in my office excludes me from everything and is rude in a thousand little ways throughout the day", and end up as "cut your Adderall dosage in half, please".

("Why is that psychiatrist telling me to cut my Adderall in half? Does he think I'm lying about having ADHD? Is he calling me a liar? These doctors have always treated me like garbage. I HAVE RIGHTS, YOU KNOW!")

Williams Syndrome is much rarer – only about 1/10,000 people, and most of them die before reaching adulthood. It's marked by a sort of anti-paranoia; Williams patients are incapable of distrusting anyone. NPR has a good article, A Life Without Fear, describing some of what they go through:

> Kids and adults with Williams love people, and they are literally pathologically trusting. They have no social fear. Researchers theorize that this is probably because of a problem in their limbic system, the part of the brain that regulates emotion. There appears to be a disregulation in one of the chemicals (oxytocin) that signals when to

> trust and when to distrust. This means that it is essentially biologically impossible for [them] to distrust.

The results are less than heartwarming:

> As Isabelle got older, the negative side of her trusting nature began to play a larger role. A typical example happened a couple of years ago, when Jessica and her family were spending the day at the beach. Isabelle had been begging Jessica to go to Dairy Queen, and Jessica had been putting her off. Then Isabelle overheard a lady just down the beach.
>
> "She was telling her kids, 'OK, let's go to the Dairy Queen,' " Jessica says. "And so Isabelle went over and got into the lady's van, got in the back seat, buckled up and was waiting to be taken to Dairy Queen with that family."
>
> Jessica had no idea what had happened to Isabelle and was frantically searching for her when the driver of the van approached her and explained that she had been starting her car when she looked up and saw Isabelle's face in the rearview mirror.
>
> The woman, Jessica says, was incredibly angry.
>
> "She said, 'I am a stranger, you know!' " Jessica says. Essentially, the woman blamed Jessica for not keeping closer watch on her daughter — for neglecting to teach her the importance of not getting into a car with someone she didn't know. But the reality could not be more different. "It's like, 'My friend, you have no idea,' " Jessica says.
>
> In fact, because of Isabelle, Jessica has had to rethink even the most basic elements of her day-to-day life. She

> can not take Isabelle to the dog park. She tries not to take Isabelle to the store. And when the doorbell rings, Jessica will leap over a coffee table to intercept her.
>
> It's not just Jessica and her family who must be vigilant. Every teacher at Isabelle's public school has been warned. Isabelle is not allowed to tell them that she loves them. Isabelle is not supposed to tell other schoolchildren that she loves them. And there are other restrictions.
>
> "She's not allowed to go to the bathroom alone at her school, because there have been numerous instances of girls with Williams syndrome being molested at school when they were alone in the hallway," Jessica says. "And these are like middle class type schools. So it's a very real problem. And, you know, I'd rather her be overly safe than be on CNN."

Some of the research on these kids is fascinating – I'm not sure I believe the study finding that they're incapable of racism, but the one finding a deficit detecting anger in faces seems pretty plausible.

Williams Syndrome usually involves mental retardation, but not always. Some of these people have normal IQ. It doesn't really help. Threat-detection seems to be an automated process not totally susceptible to System II control. Maybe it's like face-blindness. Intelligence can help a face-blind person come up with some systems to reduce the impact of their condition, but in the end it's just not going to help that much.

Psychiatric disorders are often at the extremes of natural variation in human traits. For every intellectually disabled person, there are a dozen who are just kind of dumb. For every autistic person, there are a dozen who are just sort of nerdy. And so on. We naturally think of some people as more trusting than others, but maybe that isn't the best frame. "Trusting" implies that we all receive the same information, and just choose how much risk we're willing to tolerate. I don't know if that's true at all.

A recent theme here has been the ways that our sense-data is underdetermined. Each datum permits multiple possible explanations: this is true of visual and auditory perception, but also of the social world. A pretty girl laughs a little too long at a man's joke; is she trying to flirt with him, or just friendly? A boss calls her subordinate's work "okay" – did she mean to compliment him, or imply it was mediocre? A friend breaks off two appointments in a row, each time saying that something has come up – did something come up, or is he getting tired of the friendship? These are the sorts of questions everyone navigates all the time, usually with enough success that when autistic people screw them up, the rest of society nods sagely and says they need to learn to understand how to read context.

But "context" means "priors", and priors can differ from person to person. There's a lot of room for variation here before we get to the point where somebody will be so off-base that they end up excluded from society. Just as there's a spectrum from smart to dumb, or from introverted to extraverted, so there's a spectrum in people's tendencies to interpret ambiguous situations in a positive or negative way. There are people walking around who are just short of clinically paranoid, or just shy of Williams Syndrome levels of trust. And this isn't a value difference, it's a perceptual one. These people aren't bitter or risk-averse – or at least they don't *start off* that way. They just notice how everyone's hostile to them, all the time.

III.

Another change in topic: bubbles.

I’ve written before about how 46% of Americans are young-earth creationists, and how strongly that fails to square with my personal experience. I’ve met young-earth creationists once or twice. But of my hundred closest friends/co-workers/acquaintances, I think zero percent of them fall in that category. I’m not intentionally selecting friends on the basis of politics, religion, or anything else. It just seems to have happened. Something about my personality, location, social class, et cetera has completely isolated me from one particular half of the US population; I’m living in a non-creationist bubble in the midst of a half-creationist country.

What other bubbles do I live in? A quick look over my Facebook and some SSC survey results finds that my friends are about twenty times more likely to be transgender than the general population. There are about twice as many Asians but less than half as many African-Americans. Rates of depression, OCD, and autism are sky-high; rates of drug addiction and alcoholism are very low. Programmers are overrepresented at about ten times the Bay Area average.

I didn’t intend any of these bubbles. For example, I’ve never done any programming myself, I’m not interested in it, and I try my best to avoid programmer-heavy places where I know all the conversations are going to be programming-related. Hasn’t helped. And I’m about as cisgender as can be, I have several Problematic opinions, and I still can’t keep track of which gender all of my various friends are on a month-to-month basis. Part of it is probably class-, race-, and location-based. And I have some speculative theories about the rest – I think I have a pretty thing-oriented/systematizing thinking style, and so probably I get along better with other groups disproportionately made up of people whose thoughts work the same way – but I didn't understand any of this until a few years ago and there are still some parts that don't make sense. For now I just have to accept it as a given.

There are other bubbles I understand much better. Most of my friends are pretty chill and conflict-averse. This is because I used to have scarier conflict-prone friends, and as soon as I got into conflicts with them, I broke off the friendship. I'm not super-proud of this and it's probably one of those maladaptive coping styles you always hear about, and a lot of people have told me I'm really extreme on this axis and need to be better at tolerating aggressive people – but whenever I try, I find it unpleasant and stop. I know some other people who seem to actively seek out abrasive types so they can get in fun fights with them. I don't understand these people at all – but whatever their thought processes, we have different bubbles.

All of this goes double or triple for people I've dated. I don't think of myself as clearly having a "type", but people I date tend to turn out similar in dimensions I didn't expect when I first met them. I'm going to be ambiguous here because it's a small enough sample that I don't want to give away people's private information, but it's true.

I think about this a lot when I meet serial abuse victims.

These people are a heartbreaking psychiatric cliche. Abused by their parents, abused by their high school boyfriend, abused by their first husband, abused by their second husband, abused by the guy they cheated on their first husband with, abused by the friend they tried to go to for help dealing with all the abuse. The classic (though super offensive) explanation is that some people seek out abusers for some reason – maybe because they were abused as children and they've internalized that as the "correct" model of a relationship.

And maybe this is true for some people. I have a friend who admits it's true of her – her current strategy is to try to find someone in the sweet spot between "jerkish/narcissistic enough to be interesting" and "jerkish/narcissistic enough to actually abuse her", and she's said so in so many words to people trying to matchmake. I guess all I can do is wish her luck.

But for a lot of people, this sort of claim is just as offensively wrong as it sounds. I know people who have tried really hard to avoid abusers, who have gone to therapy and asked their therapist for independent verification that their new partner doesn't seem like the abusive type, who have pulled out all the stops – and who still end up with abusive new partners. These people are cursed through no fault of their own. All I can say is that whatever mysterious forces connect me to transgender pro-evolution programmers are connecting them to abusers. Something completely unintentional that they try their best to resist gives them a bubble of terrible people.

I want to emphasize as hard as I can that I'm not blaming them or saying there's anything they can do about their situation, and I have no doubt that despite my emphasis people are still going to accuse me of saying this, and I apologize if any of this sounds at all like anything in this direction. But *something* has to be happening here.

IV.

Sometimes I write about discrimination, and people send me emails about their own experiences. Many sound like this real one (quoted here with permission) from a woman who studied computer science at MIT and now works in the tech industry:

> In my life, I have never been catcalled, inappropriately hit on, body-shamed, unwantedly touched in a sexual way, discouraged from a male-dominated field, told I couldn't do something because it was a boy thing, or suffered from many other experiences that have traditionally served as examples as ways that women are less privileged. I have also never been shamed for not following gender norms (e.g. doing a bunch of math/science/CS stuff); instead I get encouraged and told that I'm a role model. I've never had problems going around wearing no make-up, a t-shirt, and cargo pants; but on the rare occasion that I do wear make-up / wear a dress, that's completely socially acceptable… Hopefully my thoughts/experiences are helpful for your future social justice based discussions.

Other times they sound like the opposite. I don't have anyone in this category who's given me permission to quote their email verbatim (consider ways this might not be a coincidence), but they're pretty much what you'd expect – a litany of constantly being put down, discriminated against, harassed, et cetera, across multiple jobs, at multiple companies, to the point where they complain it's "endemic" (I guess I can quote one word) and that we need to reject a narrative of "a few bad apples" because really it's a problem with *all* men to one degree or another.

These dueling categories of emails have always confused me. At the risk of being exactly the sort of creepy person the second set of writers complain about, I hunted down some of these people's Facebook profiles to see if one group was consistently more attractive than the other. They weren’t. Nor is there any clear pattern in what industries or companies they work at, what position they’re in, or anything else like that. There isn’t even a consistent pattern in their politics. The woman I quote above mentions that she’s a feminist who believes discrimination is a major problem – which has only made it extra confusing to her that she never experiences any of it personally.

These people don’t just show up in my inbox. Some of them write articles on Slate, Medium, even The New Yorker, discussing not just how they’ve never experienced discrimination, but how much anger and backlash they’ve received when they try to explain this to everyone else. And all of them acknowledge that they know other people whose experiences seem to be the direct opposite.

I used to think this was pretty much just luck of the draw – some people will end up with nice people at great companies, other people will end up with bigots at terrible companies. I no longer think this explains everybody. Take that New Yorker article, by a black person who grew up in the South and says she was never discriminated against even once. I assume in her childhood she met thousands of different white Southerners; that’s a pretty big lucky streak for none of them at all to be racists, especially when you consider all the people who report daily or near-daily harassment. Likewise, when you study computer science in college and then work in half a dozen tech companies over the space of decades and never encounter one sexist, that’s quite the record. Surely something else must be going on here.

V.

And I think this has to come back to the sorts of things discussed in Parts I, II, and III.

People self-select into bubbles along all sorts of axes. Some of these bubbles are obvious and easy to explain, like rich people mostly meeting other rich people at the country club. Others are more mysterious, like how some non-programmer ends up with mostly programmer friends. Still others are horrible and completely outside comprehension, like someone who tries very hard to avoid abusers but ends up in multiple abusive relationships anyway. Even for two people living in the same country, city, and neighborhood, they can have a “society” made up of very different types of people.

People vary widely on the way they perceive social interaction. A paranoid schizophrenic will view every interaction as hostile; a Williams Syndrome kid will view every interaction as friendly. In between, there will be a whole range of healthy people without any psychiatric disorder who tend toward one side or the other. Only the most blatant data can be interpreted absent the priors that these dispositions provide; everything else will only get processed through preexisting assumptions about how people tend to act. Since things like racism rarely take the form of someone going up to you and saying “Hello, I am a racist and because of your skin color I plan to discriminate against you in the following ways…”, they’ll end up as ambiguous stimuli that everyone will interpret differently.

Finally, some people have personalities or styles of social interaction that unconsciously compel a certain response from their listeners. Call these “niceness fields” or “meanness fields” or whatever: some people are the sort who – if they became psychotherapists – would have patients who constantly suffered dramatic emotional meltdowns, and others’ patients would calmly discuss their problems.

The old question goes: are people basically good or basically evil? Different philosophers give different answers. But so do different random people I know who aren’t thinking philosophically at all. Some people describe a world of backstabbing Machiavellians, where everybody’s a shallow social climber who will kick down anyone it takes to get to the top. Other people describe a world where everyone is basically on the same page, trying to be nice to everyone else but getting stuck in communication difficulties and honest disagreements over values.

I think both groups are right. Some people experience worlds of basically-good people who treat them nicely. Other people experience worlds of awful hypocritical backstabbers. This can be true even if they live in the same area as each other, work the same job as each other, et cetera.

And it’s not just a basic good-evil axis. It can be about whether people are emotional/dramatic or calm/rational. It can be about whether people almost always discriminate or almost never do. It can be about whether they’re honest or liars, shun outsiders or accept them, welcome criticism or reject it. Some people think elites are incompetent parasites; others that they’re shockingly competent people who mean well and have interesting personalities. Some people think Silicon Valley is full of overpriced juicers, other people that it’s full of structured-light engines. And the people who say all these things are usually accurately reporting their own experiences.

Some people are vaguely aware of this in the form of “privilege”, which acknowledges different experiences at the cost of saying they have to line up exactly along special identity categories like race and gender. These certainly don’t help, but it’s not that simple – as proven by the article by that black Southerner who says she never once encountered discrimination. I’ve seen completely incomprehensible claims about human nature by people of precisely the same race, sex, class, orientation, etc as myself, and I have no doubt they’re trying to be truthful. The things that divide us are harder to see than we naively expect. Sometimes they’re completely invisible.

To return to a common theme: *nothing makes sense except in light of inter-individual variation*. Variation in people’s internal experience. Variation in people’s basic beliefs and assumptions. Variation in level of abstract thought. And to all of this I would add a variation in our experience of other people. Some of us are convinced, with reason, that humankind is basically good. Others start the day the same way Marcus Aurelius did:

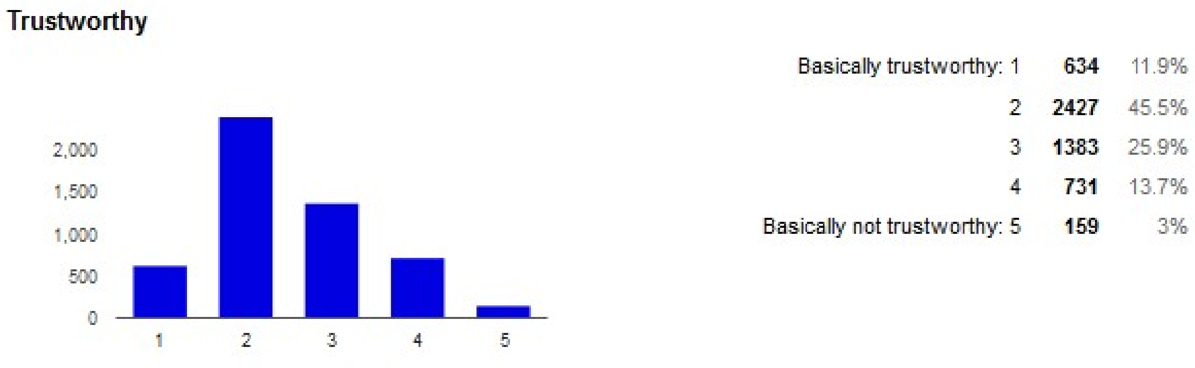
> When you wake up in the morning, tell yourself: the people I deal with today will be meddling, ungrateful, arrogant, dishonest, jealous and surly. They are like this because they cannot tell good from evil.

Notice this distinction, this way in which geographic neighbors can live in different worlds, and other people’s thoughts and behaviors get a little more comprehensible.

SSC Survey Results On Trust

Last post talked about individual differences in whether people found others basically friendly or hostile. The SSC survey included a sort of related question: “Are people basically trustworthy?”

The exact phrasing asked respondents to rate other people from 1 (“basically trustworthy”) to 5 (“basically untrustworthy”). 4853 people answered. The average was 2.49 – so skewed a bit towards higher trust. The overall pattern looked like this:



Basically trustworthy: 1	**634**	11.9%
2	**2427**	45.5%
3	**1383**	25.9%
4	**731**	13.7%
Basically not trustworthy: 5	**159**	3%

Trust didn’t differ by gender. Women averaged 2.56 and men 2.48, not a significant difference even with our fantastic sample size.

And I couldn’t detect it differing by race. Whites averaged 2.47 and nonwhites 2.59, which was nonsignificant despite decent sample size. Blacks were 2.81 and trended toward significance vs. whites, but the sample size was too small to be sure.

And I couldn’t detect it differing by religiousity. Committed theists (n = 506) had a trust level of 2.54, no more or less trusting than the average (and mostly atheist) SSC population.

And I couldn't detect it differing by intelligence. There was no correlation between trust and either IQ or SAT score. There *was* a significant difference (p = 0.001) based on education level, all the way from PhDs at 2.35 to high school graduates only at 2.62.

There were some decent-sized differences among different US states, with more urban and liberal states being more trusting. Among the states with decent sample size, California (n = 608) was 2.43, New York (n = 298) was 2.47, and Texas (n = 143) was 2.75; the California/Texas difference was significant at p = 0.001. I was only able to eyeball rather than actually significance-test the urban/liberal correlation, but it looked pretty strong.

There were similar differences between countries. Germany (n = 192) at 2.35, the UK (n = 353) at 2.37, and Canada (n = 2.39) were all significantly more trusting than the US (n = 3124) at 2.53 – but obviously the effect wasn't too impressive. There were only two non-western countries with remotely usable sample sizes. Brazil (n = 28) was 2.89, and India (n = 27) was 2.81. The non-western/Anglosphere difference was significant even with the low sample size. The most trusting city in the world was Toronto (n = 60), at 2.23.

There were so many professions, with such small sample sizes, that I wasn't really confident any of them were much more or less trusting than others. But for what it's worth, the number one least trusting profession, at 3.00, was mental health, and I am 100% not at all surprised. Otherwise there seemed to be a weak trend for nerdier and math-ier professions to be more trusting than others.

By politics, the ranking looked like this:

Social democratic: 2.38
Liberal: 2.40
Libertarian: 2.48
Conservative: 2.67
Communist: 2.80
Neoreactionary: 2.97
Alt-right: 2.97

Looks like the same trend of conservative = less trusting. Harder to figure out what to think about left vs. liberal, given the social democrats in the lead vs. communists very far behind.

Effective altruists who had taken the GWWC Pledge were much more trusting – 2.19 – than any other natural group I could find in this survey, more trusting even than Torontoans. After some work, I managed to find an unnatural group that beat them. Polyamorous Less Wrong readers from California – my proxy for the real-life Bay Area rationalist community – had a trust score of 2.13.

I had wondered if more trusting people would want less strict moderation, but the opposite was actually true – less trusting people wanted weaker moderation. Maybe this is mediated by conservativism – or maybe they just don't trust me to moderate!

Autism, anxiety, OCD, and drug use all lowered trust; depression and bipolar disorder did not.

People who put down "Other" in any category were always much less trusting than any of the options, almost as if they didn't trust people to accurately interpret the binned choices.

If you want to look into this yourself, you can find the data publicly available here, but be warned – predictably, the people who agreed to let me make their data public were systematically more trusting (2.49) than the people who refused (2.70).

In Favor Of Futurism Being About The Future

From *Boston Review*: Know Thy Futurist. It's an attempt to classify and analyze various types of futurism, in much the same way that a Jack Chick tract could be described as "an attempt to classify and analyze various types of religion".

I have more disagreements with it than can fit in a blog post, but let's stick with the top five.

First, it purports to explain what we should think about the future, but never makes a real argument for it. It starts by suggesting there are two important axes on which futurists can differ: optimism vs. pessimism, and belief in a singularity. So you can end up with utopian singularitarians, dystopian singularitarians, utopian incrementalists, and dystopian incrementalists. We know the first three groups are wrong, because many of their members are "young or middle-age white men" who "have never been oppressed". On the other hand, the last group contains "majority women, gay men, and people of color". Therefore, the last group is right, there will be no singularity, and the future will be bad.

You're going to protest that there has to be something more than that. Read the article. There really isn't. The author ignores the future almost completely, in favor of having very strong opinions on which futurist movements include the right or wrong sorts of people. AI risk researchers are "majority men, although more women than in the previous group"; techno-utopians are "more women still…but in the end that does not denote progress". All singularitarians were "sex-starved teenagers" and they all "wax eloquent about meritocracy over expensive wine" in a "super-rich bubble". The lovingly detailed descriptions everyone's social class, racial breakdown, gender ratio, what politics the author imagines they have, and what sexual insecurities she thinks produced their opinions. contrasts markedly with a total lack of concern for any of their beliefs or opinions about the future, their justifications for their beliefs, or whether those justifications are true or false. Literally the only future-related thing we know about the article’s third quadrant is that they may be involved in Bitcoin or something.

The author never even begins to give any argument about why the future will be good or bad, or why a singularity might or might not happen. I’m not sure she even realizes this is an option, or the sort of thing some people might think relevant.

Second, the article’s section on singularitarianism never mentions anything about the Singularity and doesn’t really seem to understand what the Singularity is. Its example of Singularity technologies are “augmenting intelligence through robotics”, “better quality of life through medical breakthroughs”, “cryogenics” (I assume it’s confusing this with cryonics), “medical strategies for living forever”, and “possibly even the blood of young people.”

None of these (except *maybe* the first) relate to the Singularity, which is defined as a point at which the rate of technological advance reaches near-infinity and it’s impossible to predict what happens afterwards. The article seems to use “singularitarianism” to mean “cool near-future technologies”, which is kind of the opposite of its real meaning. This is a fatal error for an article proposing a system classifying all futurists as “singularitarian” vs. “nonsingularitarian”.

It makes sense only in the context of the author having no interest in futurist movements at all, and indeed she later more-or-less admits that by ‘singularitarian optimists’ she means ‘rich white people she doesn’t like’. When discussing Elon Musk, whom some might call a pessimist based on his belief that the Singularity will destroy the world and doom humanity, she says that “being an enormously rich and powerful entrepreneur, he probably belongs in the first [Singularity optimist] group”.

Third, the article wants to classify some technologies as inextricably associated with privilege, but it has a pretty weird conception of which ones they are. It gives five examples of technologies that it’s possible to worry about without being a privileged white man, and every one of them is a different form of algorithmic bias. Really? That’s the *only* future technology it’s okay to care about? So much so that of five slots for potentially worrying technology, you filled all five with the same one?

Likewise, when the author discusses bad “singularity” technologies that only white men could want, she includes “better quality of life through medical breakthroughs”. I’m sure this just slipped in by accident. I’m sure (pretty sure?) if we pointed her to someone with chronic pain who hasn’t been able to leave the house in years and asked whether it might be good to have technology that could help this person, she would say yes. But it’s a really interesting slip-up to make. I’ve written hundreds of articles during my lifetime and I don’t think I’ve *ever* mistakenly said that only privileged white men could care about not being sick.

Again, this would make sense if the author doesn’t really believe in futurology except as a way of sending the right class signals. Helping sick people improve their quality of life? Do gross male nerds from the outgroup support that or oppose that? Okay, sold. I’m sure if her mental editor had caught it, she’d have realized that she was *supposed* to support that kind of thing, but it would be a post-processing addition to her thought stream rather than a natural component of it.

Fourth, the article presupposes a bitter conflict between the four quadrants, whereas actually people tend to be a lot more on the same side than she expects.

Her pessimists are concerned about algorithmic bias making banks less likely to extend credit to poor people. But her *optimists* just care about flashy new things like cryptocurrency. Okay. But one possible application for cryptocurrency is peer-to-peer microfinance via smart contracts – ie one of the most promising solutions to bias in big financial institutions. You don't have to agree this is a good solution. But cryptocurrency enthusiasts are working on it, and it seems weird to deny this matters or that the whole *reason* behind developing some of these flashy new technologies is to solve recognized societal problems.

And her singularitarians are strategizing how to deal with far-future advanced AI algorithms, while her nonsingularitarians are strategizing how to deal with near-future primitive AI algorithms. These seem like…not entirely the opposite of each other? Imagine you were writing an article on the different kind of climatologists studying global warming. There's the kind who indulge in crazy sci-fi scenarios where entire cities flood and the Earth becomes uninhabitable. And then there's the kind dealing with important real-world problems like increased frequency of hurricanes and creeping desertification. Is this a reasonable distinction? Which kind should you be?

(*Boston Review* readers: "How should I know? You didn't tell me what ethnicity they are!")

Most people concerned about climate change are concerned about both those things. Maybe there's a little room for disagreement on the best way to balance long-term versus short-term goals – should we build seawalls to protect our cities today, or start a program of power plan retrofitting which will pay off in twenty years? But to try to turn these two positions into arch-enemies would be ridiculous and destructive. The scientists involved may have different research interests and skillsets, but not necessarily different opinions. Obviously we should have some people working on near-term problems and other people laying the groundwork to work on long-term problems.

In real life, this is what futurists are doing too. The Asilomar Conference on Beneficial AI was organized by people whose main interest was far-future Singularity scenarios, but it included some of the top experts on algorithmic bias, gave the subject a lot of airtime, and ended up with all participants signing onto a set of principles urging more work both on near-term AI problems like algorithmic bias and long-term AI problems like the development of superintelligence. Jed McCaleb, founder of Bitcoin exchange MTGOX, donated $500,000 of his profits to the Machine Intelligence Research Institute, which deals with long-term concerns about the Singularity. In the real world, everyone from all four “quadrants” of futurist are either allies, or the same people.

Again, I feel like this is the kind of error you could only make if you totally missed that futurism was a real subject, and you just wanted to make it into a morality play for your particular political opinions.

Fifth, another quote from the article:

> In the end my taxonomy (as amusing as I find it) doesn’t really matter to the average person. For the average person there is no difference between the singularity as imagined by futurists in Q1 or Q2 and a world in which

> they are already consistently and secretly shunted to the “loser” side of each automated decision.

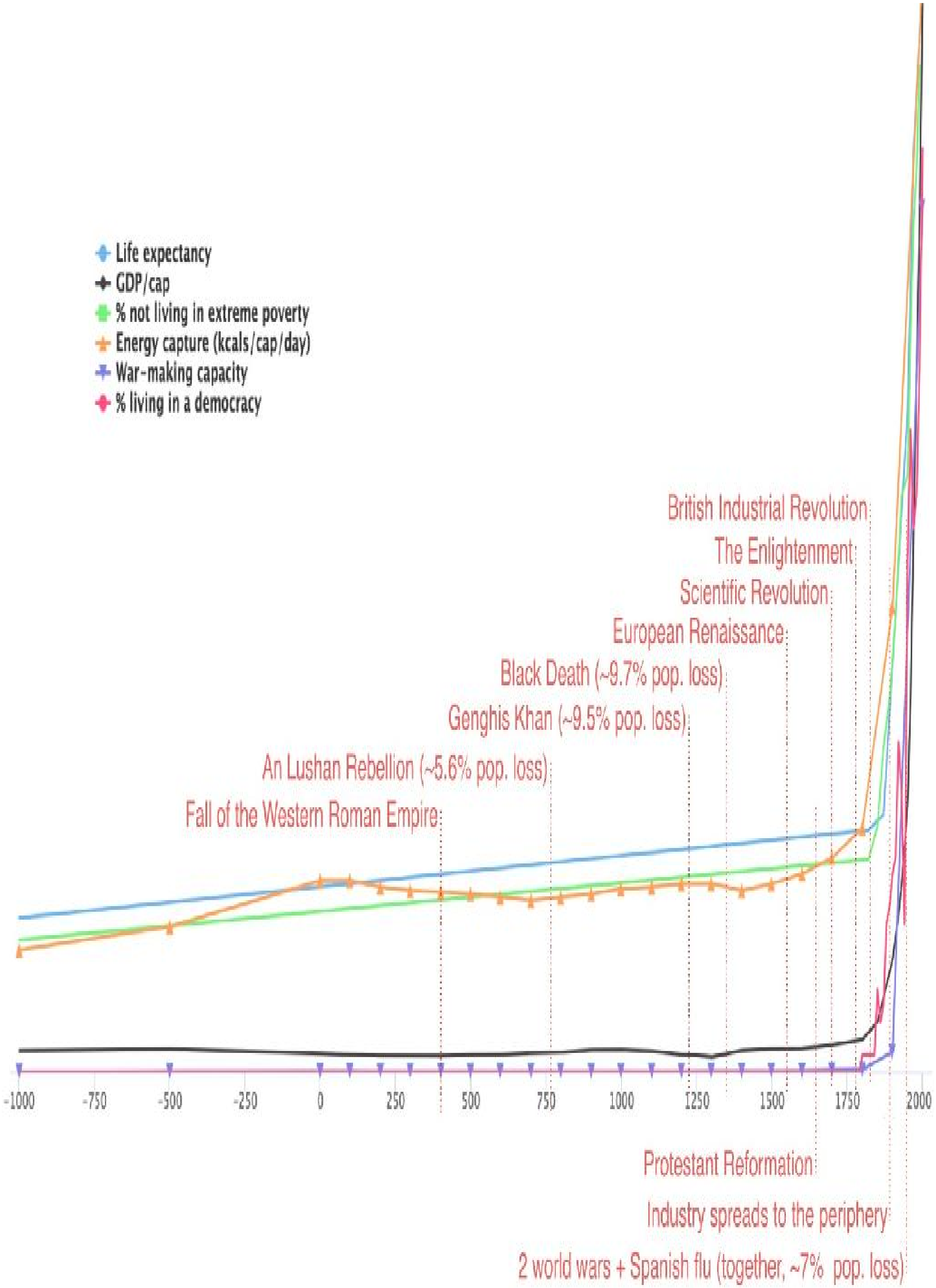
I already posited that the author doesn’t understand “Singularity”, but this is something beyond that. This is horrifying. There will be no difference for the average person between a (positive or negative) post-singularity world and the world now? *What?*

Listen up, average person. If there’s a negative singularity you will notice. Because you will be very, very dead. So will all the rest of us, rich and poor, old and young, black and white.

And if there’s a positive singularity, you will also notice. I would promise you infinite wealth, but that sort of thing kind of loses its meaning in a post-scarcity society. I would promise you immortality, but who knows if we’ll even have individual consciousnesses at that point? I would promise you bread and roses, but they would be made of hyperintelligent super-wheat and fractal eleven-dimensional time blossoms.

I don’t care if you think this vision is stupid. We’re not arguing about whether this vision is stupid. We’re arguing about whether, if this vision were 100% true, it would make a difference in the life of the average person. The *Boston Review* is saying it wouldn’t. I’m sitting here with my mouth gaping open so hard I’m worried about permanent jaw damage.

A Singularity that doesn’t make a difference in the life of the average person isn’t a Singularity worth the bits it’s programmed on. And the triumphs of science have always been triumphs for common people, whether it was the Green Revolution saving hundreds of millions of lives in the Third World, or the advent of antiparasitic drugs that are wiping malaria from Africa. When Ray Kurzweil says that the future is exponential, he’s not just talking about the number of transistors per square inch, he’s talking about this (and note the green line representing “percent of people not living in extreme poverty”):


Life expectancy
GDP/cap
% not living in extreme poverty
Energy capture (kcals/cap/day)
War-making capacity
% living in a democracy
British Industrial Revolution
The Enlightenment
Scientific Revolution
European Renaissance
Black Death (~9.7% pop. loss)
Genghis Khan (~9.5% pop. loss)
An Lushan Rebellion (~5.6% pop. loss)
Fall of the Western Roman Empire
-1000
-750
-500
-250
0
250
500
750
1000
1250
1500
1750
2000
Protestant Reformation
Industry spreads to the periphery
2 world wars + Spanish flu (together, ~7% pop. loss)

The Singularity is already here, it’s just unevenly distributed across various scales of x-axis

This is what everyone in whatever school or quadrant of futurism you care to name is thinking about. This is the only true thing. Drones, Bitcoin, Uber, superintelligence, whatever, these are part of it, but they’re not the goal in itself. We are going to fight our hardest to end poverty, disease, death, and suffering, and we’re going to do it in spite of petty *Boston Review* articles telling us we should stop doing it so we can focus on hating each other for stupid reasons.

So here’s *my* division of futurists into two groups: shining examples, and terrible warnings. And the patron saint of the latter category is Samuel Madden.

Madden was an Anglican clergyman in 18th-century Ireland, and maybe the first futurist. In 1733, he published *Memoirs of the Twentieth Century*, a novel about people in 1999 sending letters back through time to tell their 18th-century predecessors what the future would hold.

How did the prognosticators of 1733 imagine the future? Was it utopian? Decadent? Miserable? Beautiful? Incomprehensible?

Actually, it was none of those things. It was exactly like 1733 in every way, and the future people were just writing back to remind everyone how much Catholics sucked.

I am serious about this. Book-World-1999 had no technological advances over 1733. The political situation was more or less the same, although the Wikipedia review mentions that “Tatars” had taken Constantinople at some point. The important thing, the thing that they invented time travel to tell the past, was that Catholics were still bad.

Really, really bad. The people of 1733 really needed to know just how amazingly bad Catholics were and would continue to be.

The problem here isn't just that Catholics aren't really that bad. I feel like even if Catholics were exactly as bad as Samuel Madden thought, there would still be an unforgivable pettiness here. If we could show Samuel Madden the real future of his world, I hope he would be awed and horrified beyond words. The hope and heartbreak of the French Revolution, the lightning-fast transformations of industrialization, the slow march of atheism through previously Christian Europe, the otherworldly horror of the atom bomb, the glory of the moon landing, and then a 1999 poised on the edge between a Fukuyaman end of history and collapse into environmental disaster and dystopia – nobody could write a book as grand as this, but surely one could win eternal renown just by making the feeblest attempt. And instead, we get "EVERYTHING THE SAME; ALSO, HATE CATHOLICS". The only emotion I can muster is a sort of profound disgust.

And I can't help but feel the same disgust when I read "Know Thy Futurist". I don't know whether the future will be better or worse than the past, but I feel pretty sure it will be *grander*. Either we will perish in nuclear apocalypse or manage to avert nuclear apocalypse; either one will be history's greatest story. Either we will discover intelligent alien life or find ourselves alone in the universe; either way would be terrifying. Either we will suppress AI research with a ferocity that puts the Inquisition to shame, or we will turn into gods creating life in our own image; either way the future will be not quite human. And faced with all of the immensity and danger of the coming age, the best the *Boston Review* can pull off is "HAVE YOU CONSIDERED THAT

SOME OF THE PEOPLE SPECULATING ABOUT THIS MIGHT BE IN YOUR (((OUTGROUP)))?”

There’s a Deeply Wise Saying that all science/prediction/philosophy/theology/whatever will inevitably reflect the parochial conditions of the writer’s own time. Maybe so. But I feel like it doesn’t have to be *quite* as parochial as Samuel Madden. If the people of 1733 had thought about things really hard, tried to transcend the feuds of their local time and place, might they have predicted the Industrial Revolution? Might they have been able to accelerate it, delay it, send it along a different track that ameliorated some of the displacement and poverty it caused in reality? I don’t know. But it would have been a pretty amazing attempt. What would it look like to try to do something like that today? Is “Know Thy Futurist” making it more or less likely that will happen?

In the grand scheme of things, it’s probably dumb for me to be so angry about this one article. I guess what bothers me is that it’s *not* just one article. Probably a *majority* of the stuff I see written evaluating the future, or technology, or Silicon Valley these days seems to take basically this perspective. I was really mad at Maciej Ceglowski a few months ago because his anti-singularity screed was about half this kind of thing, but by this point 50%-real-argument is looking pretty good. More and more people are dropping the 50%-real-argument veneer and just admitting that stereotypes and ad hominems are the way they want to conduct everything. Do we really need to turn our hopes and dreams about the world to come into yet *another* domain where white people accuse other white people of whiteness and are accused of whiteness in turn until everyone hates each other and anything good and real gets buried in an endless heap of bullshit and 140-character brutal owns?

I wish ignoring this kind of thing was an option, but *this is how our culture relates to things now*. It seems important to mention that, to have it out in the open, so that people who turn out their noses at responding to this kind of thing don’t wake up one morning and find themselves boxed in. And if you’ve got to call out crappy non-reasoning sometime, then meh, this article seems as good an example as any.

If we get very lucky, there will actually be a future. Some of the people in it will probably read the stuff we write. They’ll judge us. I assume most of that judgment will involve laughing hysterically. But we can at least aim for laughter that’s good-natured instead of scornful. *Sub specie aeternatis*, how much of what we do today is going to look to them the way Samuel Madden does to us?

SSC Journal Club: Serotonin Receptors

Pop science likes to dub dopamine “the reward chemical” and serotonin “the happiness chemical”. God only knows what norepinephrine is, but I’m sure it’s cutesy.

In real life, all of this is much more complicated. Dopamine might be “the surprisal in a hierarchical predictive model chemical”, but even that can’t be more than a gross oversimplification. As for serotonin, people have studied it for seventy years and the best they can come up with is “uh, something to do with stress”.

Serotonin and brain function: a tale of two receptors by Robin Carhart-Harris and David Nutt tries to cut through the mystery. Both authors are suitably important to attempt such an undertaking. Carhart-Harris is a neuropsychopharmacologist and one of the top psychedelic researchers in the world. Nutt was previously the British drug czar but missed the memo saying drug czars were actually supposed to be *against* drugs; after using his position to tell everyone drugs were pretty great, he was summarily fired. Now he’s another neuropsychopharmacology professor, though with cool side projects like inventing magical side-effect-free alcohol. These are good people.

And they have a good theory. One stumbling block in past attempts to understand serotonin was the brain’s dozen or so different types serotonin receptors, all of which seem to do kind of different things. Carhart-Harris and Nutt (subsequently: CH&N) focus on two of these which show up again and again in psychiatry: 5-HT1A and 5-HT2A. Past studies had always shown these two receptors having kind of opposite effects, which confused things pretty thoroughly: why would you want a chemical that does two opposite things?

5-HT1A is the most common serotonin receptor in the brain. When SSRI antidepressants like Prozac, Zoloft and Celexa increase serotonin, this is the receptor most of that serotonin goes to. Some other antidepressants and antianxiety medications like BuSpar, Viibryd and Trintellix just stimulate this receptor directly. So it looks like this receptor does *something like* “reduce depression and anxiety”. But this falls afoul of a version of Algernon’s Law: there shouldn’t be any switch in the brain which is 100% good or 100% bad. Why have a receptor for treating depression and anxiety, rather than just always keep the receptor at maximum so you’re never depressed or anxious?

5-HT2A is another pretty common receptor. Most new antipsychotics like Seroquel and Abilify block this receptor. And most psychedelic drugs like LSD and magic mushrooms stimulate it really hard. Since psychedelics make you kind of crazy, and antipsychotics make you stop being crazy, 5-HT2A must have something to do with psychosis. Of course, this is another Algernon’s Law violation: why is there a receptor just to make you psychotic?

1A and 2A seem to “fight” each other. The more you activate 1A, the quieter 2A becomes – this is why people on SSRIs get less effect from psychedelics. And all the drugs that block 2A are also decent antidepressants – this is why people recommend Seroquel for depression even though it’s an antipsychotic – and this seems to work because blocking 2A increases 1A.

On the other hand, there also seems to be some deeper unity. 1A makes you less depressed. 2A – well, we keep hearing all these studies, some of them from Dr. Carhart-Harris himself, showing that magic mushrooms treat depression really well. Not just as a once daily medication, but in the sense that one trip on mushrooms can make you long-term – maybe permanently – less depressed. This is pretty weird. Blocking 2A makes you less depressed? But stimulating 2A also makes you less depressed, in a different and more permanent way? What's going on?

CH&N argue: both 1A and 2A promote coping with stress. 1A promotes "passive coping". 2A promotes "active coping".

Passive coping is basically being stoic, having a stiff upper lip, and waiting it out. Imagine you're at some kind of terrible job and your boss is bullying you all the time and you can't stand it and you get depressed and anxious. Your psychiatrist gives you an SSRI (or BuSpar, or Viibryd, or some other 1A stimulator) and now, you *can* stand it. Your boss is still just as mean. Your life is still just as bad. But you sort of shrug, think "what can I do?" and get back to work. This isn't the most inspiring story, but it's better than alternatives like "being a wreck" or "snapping and attacking your boss". Did I mention that 1A is known to decrease impulsivity and aggression? Makes sense.

Active coping is…uh…sort of unclear from the paper. It sounds like it should mean working to solve the problem – quitting your job, finding a way to stand up for yourself. Heck, even snapping and attacking your boss would tie in with the psychosis angle. This is…not exactly where CH&N go, as far as I can tell. Active coping is like…an LSD trip? It's some kind of grabbing the brain and shaking it, in the hopes that maybe when it settles it will be in a state that's better able to deal with whatever's going on. This sort of makes sense, insofar as big steps like quitting your job might require a lot of mental shake-up to consider. It seems to have something to do with a process of increased plasticity, becoming bolder to avoid getting trapped at local minima, and increasing the information-theoretic entropy of brain states. This definitely sounds like the sort of thing that can cause psychosis, and *maybe* it sounds like the sort of thing that might help?

MDMA, a strong 2A agonist, is currently in Phase III trials as a treatment for post-traumatic stress disorder. It looks *really* promising. Under CH&N's theory, this makes a lot of sense. If you have trauma, your thoughts get stuck in some pattern which is useful for dealing with or avoiding the traumatic situation – for example, an abused child learns to be suspicious and afraid of everybody. People do therapy for years trying to cast off these thought patterns; they know they're no longer adaptive, but they just can't get rid of them. On MDMA – and especially in MDMA-assisted therapy – people find it easy; the usual metaphor is calcified thought patterns suddenly become fluid and re-writable. Is this the sort of "increased plasticity" that CH&N describe?

This theory gives an explanation of how 1A and 2A can have such a complex – and sometimes antagonistic – relationship. When a person undergoes adversity, their brain releases serotonin, which starts by hitting the 1A receptors. They bear it stoically and hopefully soldier through. But if the adversity gets really bad and the serotonin release passes some threshold, it starts hitting the 2A receptors instead. Now their brain realizes things are pretty bad, it's got to try high-variance strategies, and so it increases its randomness in the hope of stumbling across a way-out-there solution to the problem.

(not super-clear what problem John Lilly thought he was solving by accusing space aliens of orchestrating a massive conspiracy to manipuate the world's coincidences, but it's a pretty safe bet the 2A receptor was involved somehow.)

I find the whole thing pretty plausible. But as written, it doesn't entirely answer the Algernon's Law questions. Why doesn't everyone just have 1A and 2A functions set to max all the time? What's the tradeoff?

There are some obvious possibilities. Too much 2A stimulation makes you psychotic. This puts the efficacy of atypical antipsychotics like Seroquel in a new light: they're saying something like "keep your thoughts very careful and low-risk, this isn't a good time to be deviating from normal patterns". And so maybe someone who otherwise would have believed the space aliens were putting a transmitter in his teeth will decide *not* to think that. Is there a shade of Bayesian brain theory here? Is the phrase we're looking for "strength of priors"? I don't know.

Likewise obvious: if 1A promotes stoic coping, then too much of it prevents you from actively making your life better. One can imagine how this would be more relevant in the environment of evolutionary adaptedness than today. Back then stressors could have been some specific person whose skull you could bash in with a rock. Nowadays they tend to be things like corporations, national governments, and groups of people with terrible politics on Twitter; attempted skull-bashing, as satisfying as it might feel, is highly disrecommended.

I don't know if these stories are true. They don't really explain why 1A and 2A function seem inversely related. Is this just a wiring issue? Or is there some fundamental reason why ability to passively cope can't coexist with creative outside-the-box problem-solving? Maybe the coping involves some sort of mental resolution not to let all the stress change the brain at all, and the problem-solving involves the brain becoming superplastic and really easily influenced by external events. But it’s not really clear why either of those things should be necessary.

Also, we should remember that although CH&N’s theory explains a lot, we’re reading the case they’re presenting, and there’s a lot they leave out. Some might complain that calling 2A the “active coping receptor” is as reductionistic as the whole “dopamine is the reward chemical” thing – 2A is also involved in obesity, sexual dysfunction, some forms of insomnia, possibly chronic fatigue syndrome, platelet clumping, et cetera. All of these psychedelics do opposite things acutely and chronically – something CH&N acknowledge – so you have to be really careful with time course in order to figure out whether your acid trip is treating depression due to acutely increased 2A stimulation or chronically decreased number of 2A receptors. Both Carhart-Harris and Nutt have spent big parts of their careers advocating more use of psychedelics, so them coming up with a theory of why psychedelics are really good is both reasonable and suspicious.

Still, this is as good a theory of serotonin function as anything else I’ve seen, and it will be exciting to see if it suggests any avenues for experimental research to confirm or refute it.

SSC Meetup: Bay Area 10/14

WHEN: 3 PM on Saturday, October 14

HOW: We haven't done well with cafes or other more traditional meetup spaces in the past, so we'll probably just meet outside and sit on the grass. Bring blankets / refreshments iff you want them.

WHERE: Berkeley campus, meet at the open space beside the intersection of West and Free Speech. Please disregard any kabbalistic implications of the meetup cross-streets.

WHO: Special guest Scott Aaronson from Shtetl-Optimized. Also me, Katja Grace, possibly David Friedman, hopefully other people.

WHY: Because Professor Aaronson will be giving a lecture on Black Holes, Firewalls, And The Limits Of Quantum Computers (to which you're all invited) at Berkeley later in the week and kindly agreed to hang out with us while he was in town.

See you there!

Were There Dark Ages?

[Warning: non-historian arguing about history, which is always dangerous and sometimes awful. I will say in my defense that I'm drawing off the work of plenty of good historians like Bryan Ward-Perkins and Angus Maddison whom I interpret as agreeing with me. And that the people I am disagreeing with are not historians themselves, but other non-historians trying to interpret historians' work in a popular way that I interpret as wrong. And that as far as I know no historian believes non-historians should never be allowed to talk about history if they try to be careful and cite their sources. Read at your own risk anyway.]

Cracked offers Five Ridiculous Myths You Probably Believe About The Dark Ages; number one is "The Dark Ages Were A Real Thing":

> The Dark Ages were never a thing. The entire concept is complete and utter horseshit cobbled together by a deluded writer. The term "Dark Ages" was first used in the 14th century by Petrarch, an Italian poet with a penchant for Roman nostalgia. Petrarch used it to describe, well, every single thing that had happened since the fall of Rome. He didn't rain dark judgment over hundreds of years of human achievement because of historical evidence of any kind, by the way; his entire argument was based on the general feeling that life sucked absolute weasel scrotum ever since Rome went belly-up.

Likewise There Were No European Dark Ages, The Myth Of The Dark Ages, The Myth Of The "Dark Ages", Medieval Europe: The Myth Of The Dark Ages, Busting The "Dark Ages" Myth, and of course smug Tumblr posts.

This isn't coming out of nowhere. Many people's idea of medieval times is exaggerated. Not every scientist was burned at the stake, not everyone thought the world was flat and surrounded by space dragons, and the High Middle Ages were notable for impressive levels of material progress which in some cases outpaced the Classical World and which set the stage for the upcoming Renaissance (the continuity thesis). Granted.

But I worry that as usual, this corrective to an overblown narrative of darkness has itself been overblown. People are now talking about how you’re a gullible rube if you still believe in a so-called “Dark Age”, and how all the *real* intellectuals know that this was a time of flourishing civilization every bit as good as the Romans or the Renaissance.

Bulls**t. The period from about 500 to about 1000 in Christian Western Europe was marked by profound economic and intellectual decline and stagnation relative to the periods that came before and after it. This is incompatible with the “no such thing as the Dark Ages” claim except by a bunch of tortured logic, isolated demands for rigor, and historical ignorance.

To go through the arguments one by one:

1. The “Dark Ages” were only dark in Europe. And not even all of Europe – not in the Eastern Roman Empire, not in al-Andalus…

I wonder if these people interrupt anyone who talks about the Warring States period with “*actually*, there were only warring states in *China*. Many other areas during this period had no warring states at all! Guess you fell victim to the Myth Of The Warring States Period.”

What about the Bronze Age? There wasn’t any bronze in Australia. The Hellenistic period? Huge swathes of the Earth’s land area remained un-Hellenized. The Time of Troubles? Actually, outside of Russia there were no more troubles than usual. The Era of Good Feelings? Maybe there were a bunch of bad feelings not in the US.

Every other historical age name is instantly understood by everyone to refer to both a time and a place. The *only* time anyone ever gives anybody else grief over this is when they talk about the Dark Ages. This is an isolated demand for rigor. And if this is really your true objection, let’s just agree to call it the Western European Dark Ages, as long as we can also agree it existed and was bad.

2. What about all the great stuff in the Dark Ages? Thomas Aquinas! Gothic cathedrals! Dante! Troubadours! The Song of Roland! Roger Bacon! Musical notation! Surely no period that produced all that can be called ‘dark’!

All of those are from after the period 500 – 1000 AD.

Suppose someone tells you that the middle of America contains the Great Plains, a very flat region. But you know that actually there are lots of tall mountains, like the Rockies. Have you debunked the so-called Great Plains narrative and proven that its believers are credulous morons? Or have you just missed that there’s a natural and well-delineated area suitable to be called “Great Plains” that doesn’t include your supposed counterexamples?

The period after 1000 AD did indeed have lots of great accomplishments. That’s because Europe at that time had 500 years to recover from the civilizational collapse that demolished its economic and intellectual capacity – a collapse whose immediate aftermath we call “the Dark Ages”. I agree there are some concepts of the Dark Ages that mistakenly include some of the time after the recovery, and that Petrarch’s original version commits this error. But I think that there’s also a five hundred year period – more than long enough to count as a real historical age – that absolutely fits the bill.

3. The term "Dark Ages" was invented by Petrarch – who wasn't even a real historian – based only on his personal opinion.

The term "World War I" was invented by Ernst Haeckel, who was not a historian, based on his personal opinion that it seemed to be a war, and involve the whole world, and be the first one to do so.

The term "Cold War" was invented by George Orwell, who was not a historian, based only on his personal opinion that it seemed conflict-y but without much actual fighting.

Very few of the historical terms we use were invented by professional historians, and they are all necessarily based on that person's opinion that it correctly describes the thing being described. I await people admitting that there was no Cold War, because who is George Orwell to think he can just *name* an era based on what he feels it was like?

This is another isolated demand for rigor. Historical periods get their names from random individuals reflecting on them; the names catch on if people agree that they fit.

4. The term "Dark Ages" was originally just supposed to mean that there aren't many sources describing it, not that the era was bad

Nope, wrong. Some people have used it this way, but this is neither how the term's original inventors intended it, nor how a majority of modern people (historian or otherwise) think of it.

As mentioned above, the idea of a Dark Age was first developed by the late medieval/early Renaissance thinker Petrarch. As per Wikipedia:

> The idea of a Dark Age originated with the Tuscan scholar Petrarch in the 1330s. Writing of the past, he said: “Amidst the errors there shone forth men of genius; no less keen were their eyes, although they were surrounded by darkness and dense gloom”. Christian writers, including Petrarch himself, had long used traditional metaphors of ‘light versus darkness’ to describe ‘good versus evil’. Petrarch was the first to give the metaphor secular meaning by reversing its application. He now saw Classical Antiquity, so long considered a ‘dark’ age for its lack of Christianity, in the ‘light’ of its cultural achievements, while Petrarch’s own time, allegedly lacking such cultural achievements, was seen as the age of darkness. […]
>
> Petrarch wrote that history had two periods: the classic period of Greeks and Romans, followed by a time of darkness in which he saw himself living. In around 1343, in the conclusion of his epic Africa, he wrote: “My fate is to live among varied and confusing storms. But for you perhaps, if as I hope and wish you will live long after me, there will follow a better age. This sleep of forgetfulness will not last for ever. When the darkness has been dispersed, our descendants can come again in the former pure radiance.”

Petrarch can’t just be referring to an absence of good historical sources – he’s talking about his own era!

Part of the evidence for the “absence of sources” claim is that the first use of the exact term “Dark Age” may come from by the 16th-century writer Caesar Baronius, who had a more specific time in mind, 888 – 1046. He wrote:

> The new age (saeculum) which was beginning, for its harshness and barrenness of good could well be called iron, for its baseness and abounding evil leaden, and moreover for its lack of writers dark.

But Baronius was writing well after Petrarch, his “Dark Age” was very different from the one we know today (only used to refer to a 150-year period in the Church), and in the same sentence that he mentioned dark = few writers, he also calls it “harsh”, “barren of good”, “base”, and full of “abounding evil”. This is not exactly a resounding victory for people claiming that the Dark Age had nothing wrong with it except slightly fewer records.

5a. It’s historical malpractice to call something “The Dark Ages”. The job of historians is to record, not to judge.

So I assume you also raise a fuss whenever someone talks about Alexander the Great? The Golden Age of Athens? The Five Good Emperors? The Enlightenment? Ivan the Terrible? The Belle Époque? I S O L A T E D . D E M A N D . F O R . R I G O R.

I agree there’s some level on which all of these are a sort of boundary-crossing in the ethics of historiography. And I agree that maybe very responsible historians want to avoid this and come up with more neutral names for very official work – I’ve seen some people talk about “Alexander III of Macedon”. Well, okay. The “Periclean Age Of Athens”. Fine. The “Time There Were Five Whole Emperors In A Row, None Of Whom Were Sadistic, Perverted, Or Insane, Which As Responsible Historians We Cannot Officially Call “Good”, But Which By The Standards Of Ancient Rome Is Seriously Super Impressive”. Whatever.

But if you *only* challenge the term “Dark Ages”, I feel like you’re doing the opposite of this suspension-of-judgment. If you say “The Dark Ages weren’t really dark!” you’re putting yourself in a position to judge historical eras, saying that maybe some of them were dark and others weren’t, but this particular one wasn’t. In this case you’re not responsibly abdicating historical judgment. You’re making a historical judgment, and getting it wrong.

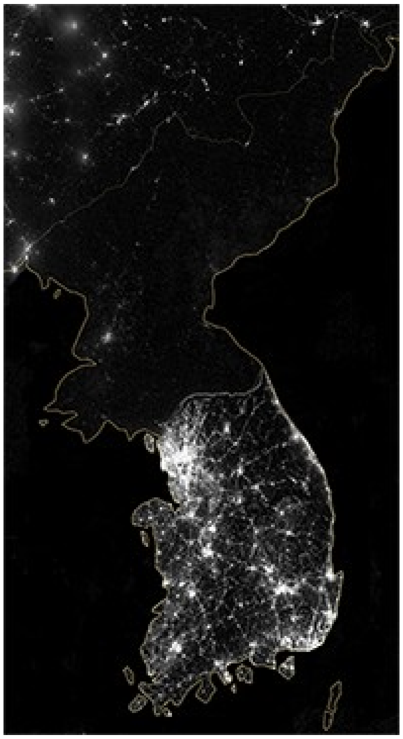
5b. The Dark Ages were only “dark” if you like big centralized states with powerful economies. There were lots of ways they might have been good. For example, ancient Rome had slavery, and most Dark Age societies didn’t. That seems pretty light-side to me!

And Alexander the Great was only “great” if you like killing a lot of people and conquering their lands.

Look, a lot of history sucked, and moral judgments are hard. Jared Diamond thinks hunter-gatherers were freer and happier than anyone since. Maybe the *real* Golden Age of Athens was in 40,000 BC, when Neanderthals on the rocky plain that would one day become Athens hunted mammoths in carefree abandon, loving life and being at one with nature and the changing seasons. Maybe the title “Alexander the Great” should *really* go to Alexander IV of Macedon, who was killed at age 14 and so never conquered, murdered, or oppressed anyone – truly an outstanding achievement matched by approximately zero other kings of the era.

In order to avoid this kind of speculation, I think of history as being along at least two axes: goodness and impressiveness. Alexander may or may not have been a good person, but he was certainly an impressive one. Periclean Athens might not have been the most virtuous city, but it was certainly one with lasting accomplishments. Since it *is* so hard to judge the goodness or badness of historical figures, most of our claims of greatness are claims about impressiveness. And compared to the periods before or after, Dark Ages Europe was unimpressive.

I’m probably an overly literal person, but whenever I think about dark ages, I think of the modern (and anachronistic for the period in question) association between light, population density, and economic activity:



The Dark Ages in Europe were a time when things would have been more towards the North Korean end of that picture. In fact, you probably could have taken a similar picture at the time, with an east/west instead of north/south axis. From *The Muslims of Andalusia*:

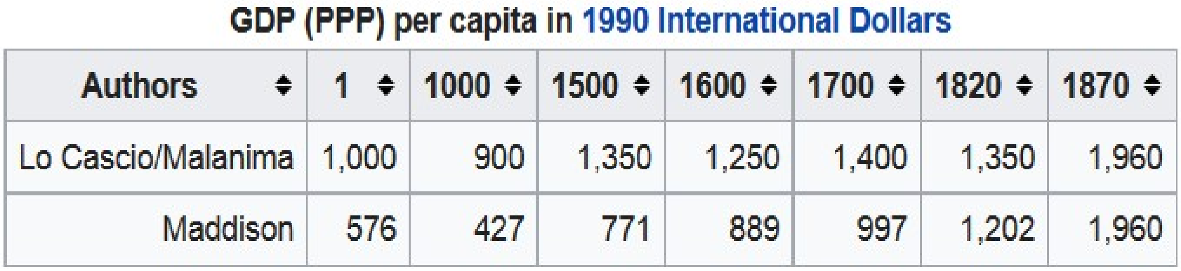
> [In medieval times], Europe was darkened at sunset, Al-Andalus shone with public lamps; Europe was dirty, Al-Andalus built a thousand baths; Europe lay in mud, Al-Andalus’ streets were paved.

I get that this is just a pun I’m taking too seriously. If you don’t like the term Dark Ages, I am happy to use the term “Unimpressive Ages”, “Disappointing Ages”, or “Pathetic Ages”. My point is that there is some axis, not the same as morality but involving economic and intellectual activity, in which the period 500 – 1000 AD was uniquely sucky.

6. Okay, forget disputes about the meanings of words or how to do history. On the object level, using normal meanings of the word “bad”, the Dark Ages were not that bad.

Wrong.

It’s hard to *prove* this is wrong, because there weren’t great statistics back then to compare Classical, Dark Age, and High Medieval societies on. As far as I know only two groups have dared try to estimate Western European GDP for these eras. Again from Wikipedia:

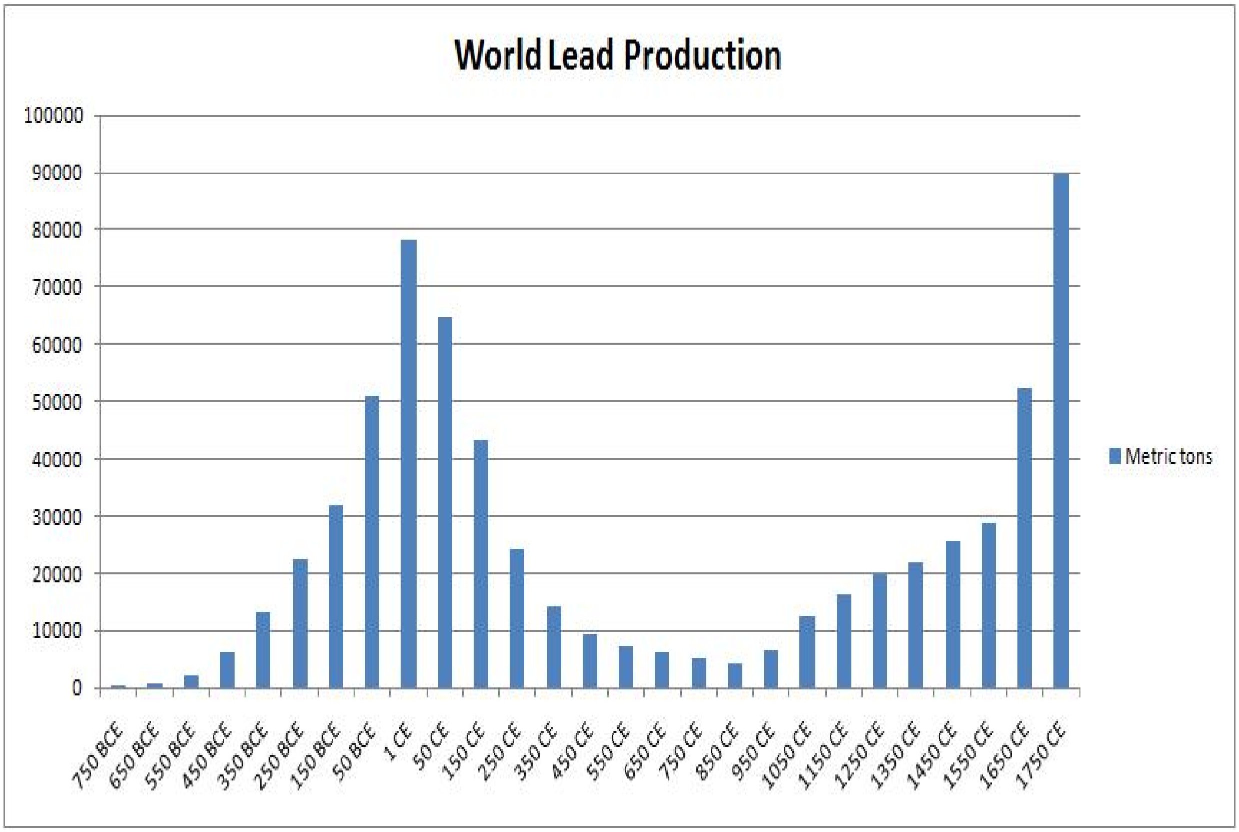


GDP (PPP) per capita in 1990 International Dollars

Authors	1	1000	1500	1600	1700	1820	1870
Lo Cascio/Malanima	1,000	900	1,350	1,250	1,400	1,350	1,960
Maddison	576	427	771	889	997	1,202	1,960

Both groups find that GDP declined from 1 AD (classical era) to 1000 AD (late Dark Age / medieval era). 1 was not the height of Rome, and 1000 was well into recovery from the Dark Ages, so we expect the difference between the Roman peak and the Dark Age nadir to be even more profound than this. But even these attenuated numbers tell the story of an entire millennium when human economic progress across an entire continent went *backwards*.

Although these numbers are inherently sketchy, the few real pieces of evidence we have seem to back them up. Arctic ice cores preserve a record of how much lead pollution was in the air, probably linked to human lead-mining activities. This allows us a pretty good look at how much lead-mining various European civilizations were doing:



And granted, the Romans were a little more obsessed with lead than could possibly have been healthy. But these data are supported by reconstructions of silver mining, copper mining, and iron mining. All of these are easily quantifiable activities that reinforce Maddison, Lo Cascio, and Malanima’s picture of economic decline between the fall of Rome and 1000.

We see a similar decline in population. The Atlas of World Population History thinks that continental Europe had a population of 36 million people at its peak in 200 AD, falling to 26 million at a nadir in 600 AD, and gradually recovering back to 36 million or so around 1000 AD. Various other estimates for the population of the Roman Empire and medieval Europe broadly support this picture (though remember that the Roman Empire didn't occupy the same space as medieval Europe and so comparisons have to be more complicated than just comparing two sets of numbers). If this is true, the Classical to Dark Age transition caused a population decrease of about 10 million, or 30% of the population (though some of this happened in Late Antiquity). These are the sorts of numbers usually only associated with the worst plagues and genocides.

Classical Rome had a population of between 500,000 and a million. Even classical Athens had a population of over 100,000. By mid Dark Ages, there was no city in Christian Western Europe larger than about 50,000 people. The infrastructure for maintaining large urban populations had fallen apart.

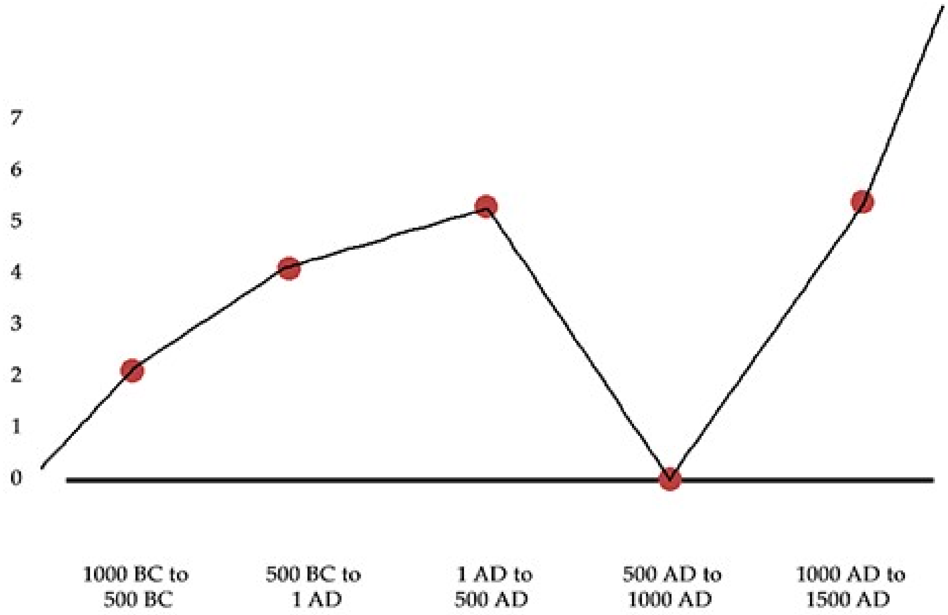
And true, a lot of this is sparse and reconstructed. My usual go-to for economic history questions, Tumblr user xhxhxhx, was able to get me a bunch of excellent graphs comparing classical Rome to the High Middle Ages, classical Rome to the Golden Age of Islam, High Middle Ages to the Golden Age of Islam, etc. When I complained that none of them compared anything to the Dark Ages which was the whole point of my question, he answered that the data were worse quality, because “civilization collapsed, so fewer people were tracking wages and prices”.

So yes, I agree that there's only a limited amount of data proving that the Dark Ages sucked. That's because *civilization collapsed, so people weren't keeping great records*. I don't think this is a strong argument *against* the Dark Ages being bad.

7. But aside from the economy, there was still lots of great culture and intellectual advancements

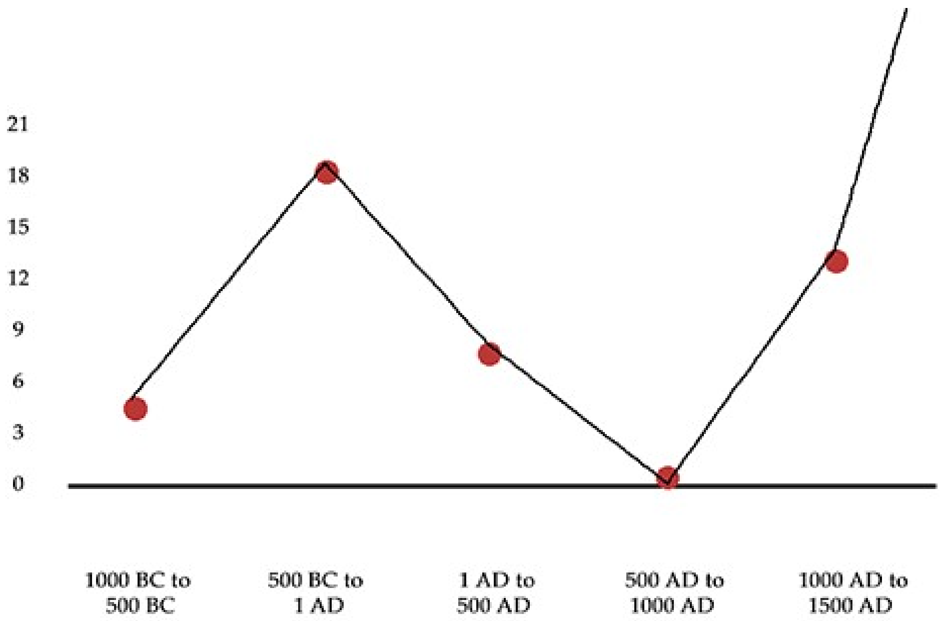
If I ask Google for a list of the hundred greatest philosophers of all time, it brings up http://www.listal.com/list/100-greatest-philosophers. It doesn't seem especially professional or official, but it's a decent-looking list and because it's the top Google result I can prove I wasn't biased by selecting it.

Here's a graph of number of European philosophers on the list per 500 year period:



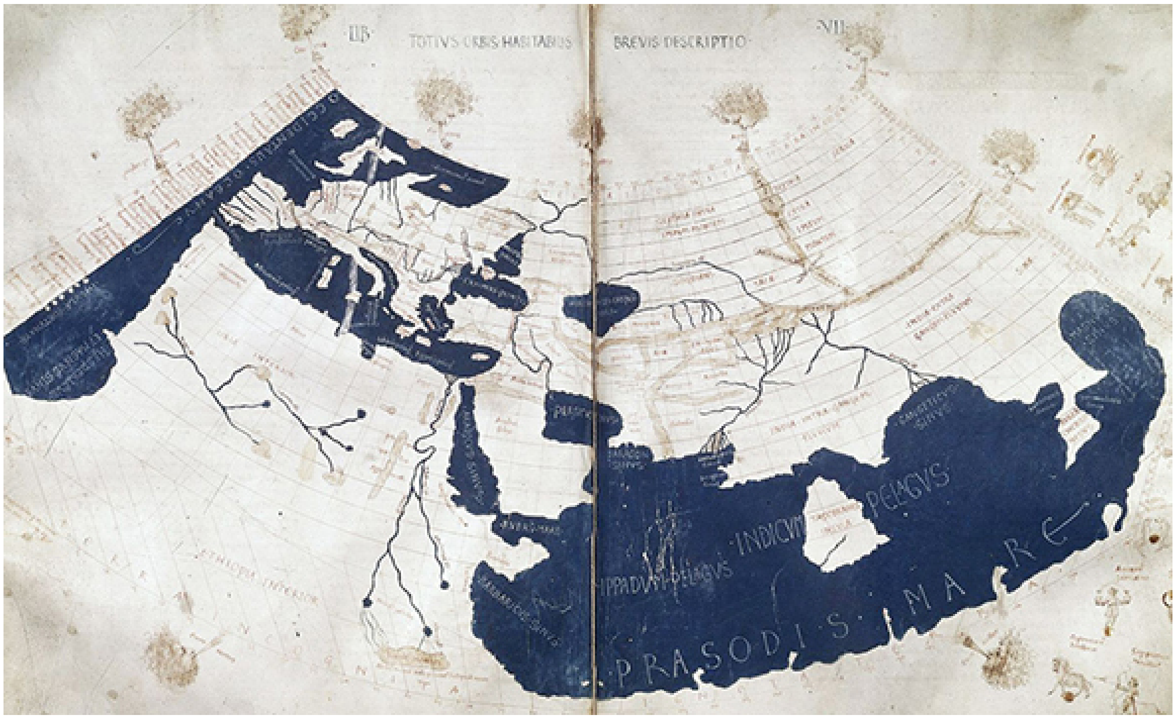
The giant pit from 500 to 1000 where there was not a single European philosopher worthy of inclusion on the list corresponds to the traditional concept of a Dark Age without very impressive intellectual output.

Harold Bloom has a list of great books in 'the Western Canon'. Once again separating them into 500 year intervals and graphing:

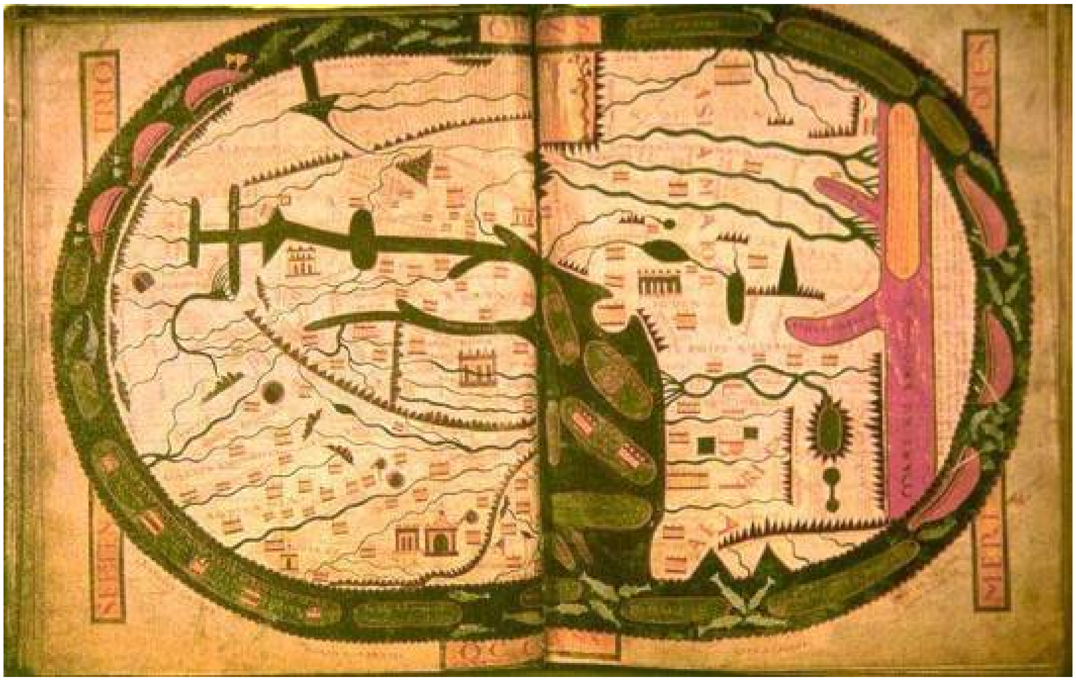


Again, we see a giant pit from 500 to 1000 AD (though this time it is not completely empty – *Beowulf* is the sole qualifying work).

Here's a map (admittedly a later reproduction, since the originals are lost) by the greatest classical geographer Ptolemy:


LIB
TOTIVS ORBIS HABITABILIS
BREVIS DESCRIPTIO
VII
INDICVM PELAGVS
HIPPADVM PELAGVS
PRASODIS MARE

And here’s an 8th-century map by Beatus of Liebana:



I’m not cheating here by taking the worst-quality Dark Age map (that would be one of these). If you can find a better Christian Western European map from 500 – 1000, tell me and I'll replace this one with it. But as far as I can tell, this really was state-of-the-art.

The decreased quality of intellectual output seems to have been matched by a decline in quantity. I can't find any great sources quantifying the number of books written in the classical world, but there are a few semi-reliable numbers about library size. The Ulpian Library of Emperor Trajan seemed to have tens of thousands of scrolls, and it was only one of as many as 28 libraries in Rome. Estimates of the number of volumes in the Library of Alexandria range from 40,000 to 400,000. Archaeologists studying the Villa of the Papyri in Herculaneum, a private residence in a medium sized town, have found a private library of almost 2,000 scrolls.

Medieval libraries seem to be much smaller. From Oxford Bibliographies:

> It follows from this that the wealth and fame of any institution that required books would inevitably affect the size of its library, and, given the fact that books were always expensive, medieval libraries were, from a modern point of view, not large. The largest Anglo-Saxon libraries may have contained about two hundred books. In 1331 the collection at Christ Church, Canterbury, numbered 1,850, which may well have been the biggest collection in England and Wales. In 1289 the library of the university of Paris contained 1,017 volumes which, by 1338, had increased to 1,722—an increase of about 70 percent.

This might not be entirely fair – Roman scrolls were smaller than medieval books, so a work that took up one medieval book might have occupied several Roman scrolls, inflating the size of Roman libraries. But there still seems to be a pretty big gap between the tens to hundreds of thousands of volumes in classical libraries and the few hundred to few thousand in libraries all the way up until the High Middle Ages.

[EDIT: This might not be true – see here]

In a lot of cases, the people of the Dark Ages (and the High Middle Ages afterwards) themselves acknowledged this. The Roman author Vitruvius was the gold standard for architecture up to the Renaissance, and Brunelleschi became famous for creating a dome that surpassed the Roman domes made 1300 years earlier. Roman doctors like Galen and Celsus were semi-worshipped by medieval doctors; when the 16th century (!) doctor Theophrastus von Hohenheim became known as "Paracelsus" (meaning "equal to or better than Celsus"), it was taken as an outrageous boast of ability despite his having the benefit of 1500 extra years of medical science.

8. The Dark Ages weren't all bad. There were still a few important accomplishments. Therefore, they cannot truly be called "dark".

The night includes several bright things, such as the moon, the stars, and streetlights. But it's still fair to call the night "dark". You don't have to prove that 100% of something fits a description at 100% of times to use the description.

One of the links from the top of the post says:

> If the "dark ages" were so unproductive and backwards, how does one explain the proliferation of inventions and developments during this time period. A simple listing of inventions, discoveries and developments demonstrates the the Middle Ages were anything but dark.

…then goes on to give various inventions, the only ones of which from 500 AD – 1000 AD are “collar and harness for horses and oxen”, “iron horseshoes”, and “the swivel axle”.

Look. I am sure that horseshoes were a revolutionary advance in equine footwear. But the ancient Greeks gave us geometry, history, cartography, the screw, the water wheel, gears, cranes, lighthouses, and fricking analog computers. If you want to stake your claim to be more than a miserable failure as a historical age, you are going to have to do better than horseshoes.

(also, maybe the Romans invented iron horseshoes first anyway?)

9. I still think the term “Dark Ages” could possibly lead to misconceptions.

Yeah.

I like this debate because it’s so pointless, but also reveals some of the basic structure of these kinds of arguments. Like most language questions, we act like we’re debating facts, when in fact we’re debating fuzzy category boundaries that are underdetermined by facts. See previous work on is Pluto a planet?, is obesity a disease?, are transgender people their chosen gender?, etc.

There’s no strict criteria for what makes something a Dark Age or whether the term should be used at all. We’re left to wonder whether using it conveys more useful information than it does misinformation.

There are many interpretations of “The Dark Ages happened” that might be wrong, like:

1. There was darkness everywhere, not just in Europe
2. There was darkness in Europe all the way until the Renaissance, and the High Middle Ages sucked
3. Every single person in this era was an illiterate superstitious peasant covered in filth, and not one good thing ever happened
4. Greco-Roman civilization was better in every way than the period that followed it, including morally

On the other hand, there are many interpretations of “the Dark Ages didn’t happen” that might also be wrong, like:

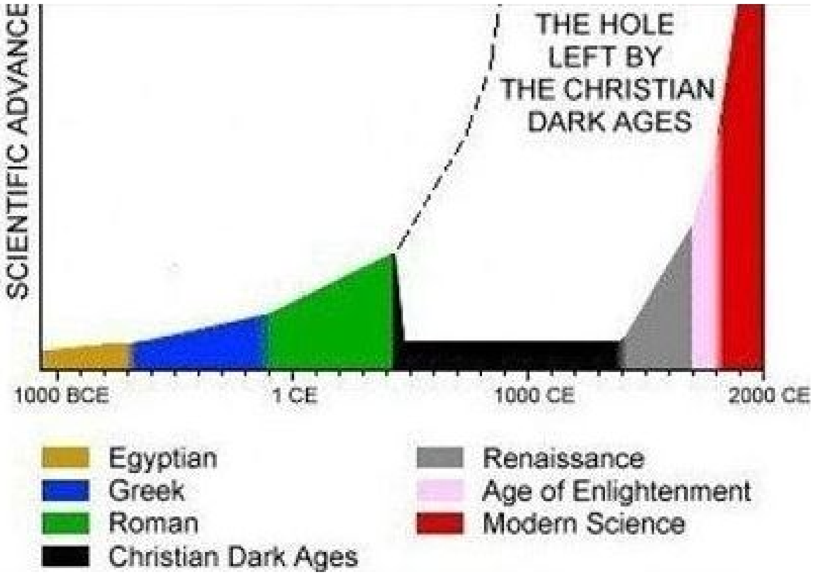
1. The fall of Rome was not associated with a decline in wealth and population.
2. The fall of Rome was not associated with a loss of capacity for things like urban living or large-scale infrastructure
3. The intellectual output of the period was exactly as high in quality and quantity as the intellectual outputs of other periods
4. Civilization always proceeds in a nice Whig History straight upward line with no risk of catastrophic collapses

Surely people can get caught in different bravery debates here. If they live in a bubble where everyone falls prey to the first set of misconceptions, it can be tempting to try to rectify that by saying the Dark Ages never happened. If they (like me) live in a bubble where everyone seems to fall prey to the second, it’s tempting to…well, write a post like this one.

And then there are political implications that will work for the benefit of one group or another. If there was a Dark Age:

1. …maybe it casts Catholicism or Christianity in a bad light, since this was also the age when they rose to be a major power
2. …maybe it points to a broader conflict between science

and religion, since this was a very religious age in ways
3. …maybe it suggests that civilization is more fragile than we think, and since it collapsed once it can collapse again
4. …maybe it makes Greece and Rome look extra good, since they were again of the curve in terms of civilizational greatness



Pictured: one way to politicize this discussion; not recommended

And finally, there are signaling aspects. Since everybody hears a vague Monty-Python-And-The-Holy-Grail-esque conception of the Dark Ages ("He must be a king…he doesn't have shit all over him"), but only people who take a history class in college hear about the Continuity Thesis, loudly proclaiming that there was never a Dark Age is one way to signal education and intellectualism (I dare you to tell me that isn't what's going on in this Tumblr post). On the other hand, if you're one of those people who rails against "postmodernism" and "cultural relativity" and wants a reputation for "calling a spade a spade", it might be gratifying to get to say that actually, that one historical era that seems kind of sucky (but fancy college professors keep insisting otherwise) does, in fact, suck.

I think I know why this question bothers me so much, and it’s because I hate when faux-intellectuals give stupid black-and-white narratives that are the tiniest sliver more sophisticated than the stupid black-and-white narratives that the general population believes, then demand to be celebrated for their genius and have everyone who disagrees with them shunned as gullible science-denying fools.

(I know a lot of people accuse me and this blog of doing exactly this, and I’m sorry. All I can say is that I’m at the odd-numbered levels of some signaling game you’re at the even-numbered levels of, and it sucks for all of us.)

For other people, maybe it’s something different. Maybe a Chinese historian doesn’t like the term “Dark Ages” because she sees too many people think Europe-specific terms apply to the whole world, and for her the tiny number of people who do this are so annoying that it overwhelms any possible advantage the idea might have. Maybe a Muslim likes it because it helps contrast the poverty of Christendom with the glory of al-Andalus, and shake the myth that Europe has always been on top. I don’t know.

10. So you’re saying both positions are true and everyone is equally right?

No. Although I sympathize with the feelings behind both positions, I say the Dark Ages happened. I think the best evidence we have suggests the fall of Rome (and the period just before) was associated with several centuries of economic and demographic decline, only reaching back to their classical level around 1000 AD. I think it was also associated with a broader intellectual and infrastructure decline, which in some specific ways and some specific fields didn’t reach back up to its Roman level until the Renaissance. I think that common sense – the sense you get when you treat the question of the Dark Age the same as any other question, and try to avoid isolated demands for rigor – says that qualifies for the phrase “Dark Age”.

[see also: Highlights From The Comments On Dark Ages]

Highlights From The Comments On Dark Ages

Thanks to everyone who made interesting comments on yesterday’s post about Dark Ages.

Several people challenged the matching of the economic/population decline to the “fall of Rome”. For example, from David Friedman:

> On the graph you are citing, 36 million is the population in 200 A.D. The fall of the Western Empire is commonly dated to about 450 A.D. By 400 A.D., on the same graph, population is down to 31 million–say 30 million by 450.
>
> So a more accurate statement would be “The late Roman Empire caused a population decrease of about six million. Population continued to fall for another hundred and fifty years before it started back up. It passed its Roman high in about 1000 A.D. and continued growing for the next three hundred years.”
>
> My rule of thumb for very poor societies is that the growth rate of population is a proxy for the average standard of living. That growth rate, the slope of the line on Figure 1.2 of the Atlas of World Population History, starts up in about 450 A.D. and continues increasing until about 1300.

From ksvanhorn:

> The graph of lead production doesn’t really jive with the idea that the Dark Ages were a result of Rome falling — production had declined sharply centuries before the fall

> of Rome. It suggests that maybe we should count the Dark Ages as beginning considerably *before* the fall of Rome.

Dissonant Cognizance:

> I was thinking throughout reading the post that you could put the start date of the Dark Ages around the Crisis of the Third Century, where Rome probably would have ceased to exist right then had Aurelian not managed to contain the damage. That let things stay superficially stable while the Western Empire cannibalized its outlying provinces for a couple more centuries.

And ctj09 agrees:

> I'd actually move up the first date [for the start of the Dark Ages] to around the time of the Crisis of the Third Century, a period where the Roman Empire very nearly collapsed and never really recovered from. Especially because the Manorialism and the explicit Dominate-style hierarchy that typified the early Middle Ages was first really developed during this period. Not to mention just after the Crisis, the Emperor Diocletian laid the ground work for a lot of what would become institutions and norms in the early Middle Ages.

Other people thought the end date of the Dark Ages could also be earlier. Many brought up the Carolingian Renaissance. For example, Krill12:

> 1000AD is ridiculously late…I have no problem with pointing out that from about 450AD to 600AD there was very little going on. That is probably a real dark age. It's also nothing like what people mean when they say "The Dark Ages." The people who use that term might have

> forgotten the Carolingian Renaissance happened before 1000AD.

From RIP_Finnegan:

> In my opinion the Carolingian Renaissance is pretty much proof positive that people even at the time saw part of their task as recovering old glories. I don't have a source book to hand, but that's the impression I got from Einhard and the RFA. There are those who argue the Carolingian Renaissance was mostly hot air, but I think it's fair to say there's a reliable middle ground between that and what they teach French schoolkids. Also, even if the Pirenne Thesis is no longer good money, it's clear that there was economic continuity from Rome later than once thought. I would say that the popular conception of the Dark Ages probably owes more to the period after Charlemagne, when the Papacy was in disarray, Europe was fragmented, Vikings were on the loose, and the polities we know from the Middle Ages were just finding their feet.
>
> However, that's not the whole story, of course. The Carolingians saw themselves as superior to Rome in one very clear way – they were Christian! In the Cathedral in Aachen, among all the re-used Roman architecture, there is a plain throne made of materials from the Holy Land. I think perhaps this ties into the James Burnham theory of Whig history – that no matter how bad things get in reality for us, we'll construct a worldview that makes civilizational defeat a victory for Truth and Justice.

Lillian disagrees:

> The Carolingian Renaissance kind of fizzled out with the breakup of the Carolingian Empire after the death of Luis the Pious though. High culture and learning did not take for good until the Renaissance of the 12th century. Hell the Holy Roman Empire was really founded by the Ottonian Dynasty, who were crowned Emperors over a century after the Carolingian realm started collapsing under its own weight. Certainly the Carolingian Renaissance laid the groundwork for what came later, but in all it was a false dawn.

So “300 – 800 AD” might be as good a five hundred year interval to call “the Dark Ages” as 500 – 1000. I think this is true of a lot of historical periods – depending on what artists or scientists you think are most important “the Scientific Revolution” or “the Renaissance” can have pretty fluid boundaries – but it’s worth noticing the fuzziness.

I had briefly noted that scrolls might be shorter than codices, but felt okay dismissing this because they would have to be something like two orders of magnitude shorter for it to make much difference. Well…here’sCaf1815:

> On the number of books at the library of Alexandria vs. at the university of Paris: the fact that scrolls vs. codices is apples vs. oranges is duly noted, but let me impress on you just how vast the difference is (full disclosure, this is my field of study, so the following may turn into a bit of a rant). At Alexandria, you’re counting scrolls; the length of a typical scroll translates to about 20-30 printed pages in a modern book. This is why works by ancient writers are divided into several “books”, each of which would take up one scroll: the City of God is 22 scrolls, the Republic of Plato is 10 scrolls, etc. At the University of Paris in 1300, they had codices, and these were huge; for example, I’m in the final stages of

> publishing a work by a 4th century writer; the book, my editor tells me, will be 550 pages long (admittedly counting an introduction, critical apparatus, etc.). But in its 13th century codex form, this work takes up folios 396 to 407 of a huge doorstop comprising 487 folios total; the whole codex, as was the norm at the time, features dozens of works by various authors in the same broad category. So if the library of Alexandria only had 40,000 scrolls (okay, that's a pretty low estimate), it would have had less text than the Sorbonne in 1300 by an order of magnitude.

If this is at all right, then *mea culpa*.

A lot of people don't like the idea of Dark Ages because they underestimate the continuity between the classical and medieval world; John of Salisbury argues that they *overestimate* it:

> I sympathise with those who have brought up Ireland and Scandinavia. My concern is that 'dark ages' talk implies too much continuity: there is the grand narrative of Western Civi, in which 'we' (those of European descent, white people) flourished in antiquity, flailed in the dark, and then triumphed in modernity. My gripe is that there isn't a 'we' that is the subject of this story: the ancient world was a Mediterranean world, and it is only in the middle ages we see a civilisation that looks like modern Europe. The Mediterranean peoples (olive oil people, as Taleb calls them) who went on to share in later European civilisation go from light to dark, but the northern Europeans, the butter people, are only stepping into the light. Take Britain, which has a dramatic, well-defined dark age that is also a lot shorter than the ordinarily cited 500 year span. The period between the Roman departure and the Saxon

> conversion is pretty much pitch black by all the relevant measures. After conversion, butter people who call the island home (even if they were rather recent arrivals) enter history on their own terms, as something more than a mere foil to Roman grandeur. Sure, the Saxons look undistinguished compared to the Athens of Pericles, but that’s hardly a fair comparison. They do pretty well considering that they had been tribal illiterates just a few generations earlier. For much of the conventional dark ages, the Saxons enjoy what is, by their standards and those of their northern European neighbours, an age of light.

Similar concerns from georgioz:

> At one point Scott says that “by mid Dark Ages, there was no city in Christian Western Europe larger than about 50,000 people.”
>
> What geographic area are we looking at here? It has to be western Europe but not Spain since Scott correctly declares Spain as part of Al-Andalus and therefore not subject to the western christian definition. So it leaves us with northern Italy – not Sicily or southern Italy as those were basically muslim Abbasid and/or Byzantine respectively until Norman conquest of Sicily in 11th century.
>
> So what does that leave us with? Basically Roman provinces of Gaul, Britannia and Northern Italy. We cannot speak about Germany or Scandinavia as it was beyond Limes Romanus – the northern Roman border. And one can arguably say that nonitallian parts of what we have here actually flourished during 500-1000 compared to the tribal past. I am not sure if one can say that what is basically Frankish empire plus Britannia felt

> some sort of Dark Ages. To the contrary. So it basically all boils down to decline of the city of Rome itself. City of Rome and northern Italy went through a very rough period. But given that the empire they ruled disintegrated it is to be expected. Other cities like Constantinople or Baghdad or Cordoba flourished instead.
>
> This is why I think historians want to have broader view. If one actually takes the area of whole Roman empire at its regional peak under Trajan then one can definitely see scientific and cultural progress in that area during 500-1000. That would be fair comparison with classical Roman times. Not only do we have now actual states in northern and western Europe such as Frankish empire or slavic protostates. But the rest of former Roman Empire did quite well under Byzantine rule and rule of Abbasid / Córdoba caliphs.

offwo2000 makes a fascinating claim about the year 1000 I'd never heard before:

> It's funny because Christianity actually caused the recovery about AD 1000. People were genuinely convinced that the world would end in AD 1000 and so the papacy started the "Peace of God" movements where rulers would stop fighting in the hope that they would be rewarded in heaven after the apocalypse. This became the catalyst for the papacy taking more control of international affairs, which enabled Western Europe to stop fighting between itself so much, focus on rebuilding the economy (there's a surge in water-mill building, for example, since peace meant that they would actually gain their return before being destroyed) and allow uniting against common enemies.

(I’m not sure how seriously to take this since the wikipedia article on the Peace of God movement doesn’t really mention the “year 1000” thing, which I found the most interesting part).

On the philosophical implications of saying the Dark Ages “were real”, John Nerst:

> Consider what it means to say that these things do/don’t exist or are/aren’t real:
>
> global warming, a solution to global warming, race, infinitely many prime numbers, the patriarchy, Kurdistan, devil worshippers, free will, schizophrenia, Esperanto, Black English, white privilege, the Friendzone, meritocracy, fate, luck, sin, seasons, personality types, learning styles, genders other than male and female, the color purple, the word ‘cromulent’, the word ‘irregardless’, the War on Christmas, property rights, fiat money, a chance of rain, a meaning of life, the meaning of life, God

And on the political implications, the anonymouse:

> The lesson here is not whether the Dark Ages were X% grimdark vs. Y% grimdark. (Although, I would suggest that the Byzantine world was still rolling 4d6-drop-lowest while western Europe had regressed to 3d6 in order.)
>
> The lesson is that civilization is fragile. It’s easy to sit in Periclean Athens (or 1891 Paris, or 2017 Seattle) and think “wow, the march of progress is inexorable!” But it isn’t. Civilization is something to be lovingly nurtured and ferociously guarded; the wolves at the door haven’t gone away just because your lamp burns too bright for you to see out into the dark.

(fortaleza84 adds “Debate about the Dark Ages, to an extent, is a proxy for anxiety over the present-day West.”)

Finally, I talked a bit about the Dark Ages bringing up two axes of “civilizational moral goodness” vs. “civilizational impressiveness”. Cernos quotes Will Durant on a different way of judging the “impressiveness” of the Dark Ages:

> The Dark Ages are not a period upon which the scholar can look with superior scorn. He no longer denounces their ignorance and superstitions, their political disintegration, their economic and cultural poverty; he marvels, rather, that Europe ever recovered from the successive blows of Goths, Huns, Vandals, Moslems, Magyras, and Norse, and preserved through the turmoil and tragedy so much of ancient letters and techniques. He can feel only admiration for Charlemagnes, Alfreds, Olafs and Ottos who forced an order upon this chaos; for the Benedicts, Gregorys, Bonifaces, Columbas, Alcuins, Brunos, who so patiently resurrected morals and letters out of the wilderness of their times; for prelates and artisans that could raise cathedrals, and the nameless poets that could sing, between one war of terror and the next. State and Church had to begin again at the bottom, as Romulus and Numa had done a thousand years before and the courage required to build cities out of jungles, and citizens out of savages, was greater than that which would raise Chartres, Amiens, and Reims or cool Dante’s vengeful fever into measured verse.

Kolmogorov Complicity And The Parable Of Lightning

> A good scientist, in other words, does not merely ignore conventional wisdom, but makes a special effort to break it. Scientists go looking for trouble.
>
> — *Paul Graham, What You Can't Say*

I.

Staying on the subject of Dark Age myths: what about all those scientists burned at the stake for their discoveries?

Historical consensus declares this a myth invented by New Atheists. The Church was a great patron of science, no one believed in a flat earth, Galileo had it coming, et cetera. Unam Sanctam Catholicam presents some of these stories and explains why they're less of a science-vs-religion slam dunk than generally supposed. Among my favorites:

Roger Bacon was a thirteenth century friar who made discoveries in mathematics, optics, and astronomy, and who was the first Westerner to research gunpowder. It seems (though records are unclear) that he was accused of heresy and died under house arrest. *But* this may have been because of his interest in weird prophecies, not because of his scientific researches.

Michael Servetus was a sixteenth-century anatomist who made some early discoveries about the circulatory and nervous system. He was arrested by Catholic authorities in France and fled to Geneva, where he was arrested by Protestant authorities, and burnt at the stake "atop a pyre of his own books". *But* this was because of his heretical opinions on the Trinity, and not for any of his anatomical discoveries.

Lucilio Vanini was a philosopher/scientist/hermeticist/early heliocentrism proponent who was most notable as the first person recorded to have claimed that humans evolved from apes – though his theories and arguments were kind of confused and he probably got it right mostly by chance. City authorities arrested him for blasphemy, cut out his tongue, strangled him, and burned his body at the stake. *But* nobody cared about his views on evolution at the time; the exact charges are unclear but he was known to make claims like "all religious things are false".

Pietro d'Abano was a fourteenth century philosopher and doctor who helped introduce Arabic medicine to the West. He was arrested by the Inquisition and accused of consorting with the Devil. He died before a verdict was reached, but the Inquisition finished the trial, found him guilty, and ordered his corpse burnt at the stake. *But* he wasn't accused of consorting with the Devil because he was researching Arabic medicine. He was accused of consorting with the Devil because he was kind of consorting with the Devil – pretty much everyone including modern historians agree that he was super into occultism and wrote a bunch of grimoires and magical texts.

Giordano Bruno was a contemporary of Galileo's. He also believed in heliocentrism, and promoted (originated?) the idea that the stars were other suns that might have other planets and other life-forms. He was arrested, tortured, and burned at the stake. *But* although his "innumerable worlds" thing was probably a strike against him, the church's main gripe was his denial of Christ's divinity.

I’m not a historian and I don’t want to debate any of these accounts. Let’s say they’re all true, let’s accept every excuse we’re given and accept the Church never burned anybody just for researching science. Scientists got in trouble for controversial views on non-scientific subjects like prophecies or the Trinity, or for political missteps.

Scott Aaronson writes about the the Kolmogorov option (suggested alternate title: “Kolmogorov complicity”). Mathematician Andrey Kolmogorov lived in the Soviet Union at a time when true freedom of thought was impossible. He reacted by saying whatever the Soviets wanted him to say about politics, while honorably pursuing truth in everything else. As a result, he not only made great discoveries, but gained enough status to protect other scientists, and to make occasional very careful forays into defending people who needed defending. He used his power to build an academic bubble where science could be done right and where minorities persecuted by the communist authorities (like Jews) could do their work in peace.

It’s tempting to imagine a world where Servetus, Bacon, and Bruno followed Aaronson’s advice. They pursued their work in optics, astronomy, anatomy, or whatever other subject, but were smart enough never to go near questions of religion. Maybe they would give beautiful speeches on how they had seen the grandeur of the heavens, but the true grandeur belonged to God and His faithful servant the Pope who was incidentally right about everything and extremely handsome. Maybe they would have ended up running great universities, funding other thinkers, and dying at a ripe old age.

Armed with this picture, one might tell Servetus and Bruno to lay off the challenges. Catholicism doesn’t seem *quite* true, but it’s not doing too much harm, really, and it helps keep the peace, and lots of people like it. Just ignore this one good prosocial falsehood that’s not bothering anybody, and then you can do whatever it is you want.

II.

But Kolmogorov represents an extreme: the politically savvy, emotionally mature scientist able to strategically manipulate tough situations. For the opposite extreme, consider Leonid Kantorovich.

Kantorovich was another Russian mathematician. He was studying linear optmization problems when he realized one of his results had important implications for running planned economies. He wrote the government a nice letter telling them that they were doing the economy all wrong and he could show them how to do it better. The government at this point happened to be Stalin during his “kill anybody who disagrees with me in any way” phase. Historians are completely flabbergasted that Kantorovich survived, and conjecture that maybe some mid-level bureaucrat felt sorry for him and erased all evidence the letter had ever existed. He was only in his 20s at the time, and it seems like later on he got more sophisticated and was able to weather Soviet politics about as well as anybody.

How could such a smart guy make such a stupid mistake? My guess: the Soviet government didn’t officially say “We will kill anyone who criticizes us”. They officially said “Comrade Stalin loves freedom and welcomes criticism from his fellow citizens”, and you had to have some basic level of cynicism and social competence to figure out that wasn’t true.

Even if the Soviet government had been more honest and admitted they were paranoid psychopaths, the exact implications aren’t clear. Kantorovich was a professor, he was writing about a very abstract level of economics close to his area of expertise, and he expressed his concerns privately to the government. Was that really the same as some random hooligan shouting “I hate Stalin!” on a street corner? Surely there were some highly-placed professors of unquestionable loyalty who had discussed economics with government officials before. Even a savvier version of Kantorovich would have to consider complicated questions of social status, connections, privileges, et cetera. The real version of Kantorovich showed no signs of knowing any of those issues even existed.

If you think it’s impossible to be that oblivious, you’re wrong. Every couple of weeks, I have friends ask me “Hey, do you know if I could get in trouble for saying [THING THAT THEY WILL DEFINITELY GET IN TROUBLE FOR SAYING]?” When I stare at them open-mouthed, they follow with “Well, what if I start by specifying that I’m not a bad person and I just honestly think it might be true?” I am half-tempted to hire babysitters for these people to make sure they’re not sending disapproving letters to Stalin in their spare time.

The average person who grows up in a censored society may not even realize for a while that the censorship exists, let alone know its exact limits, let alone understand that the censors are not their friends and aren’t interested in proofs that the orthodoxy is wrong. Given enough time, such a person can become a savvy Kolmogorov who sees the censorship clearly, knows its limits, and understands how to skirt them. If they’re really lucky, they may even get something-like-common-knowledge that there are other Kolmogorovs out there who know this stuff, and that it’s not their job to be a lone voice crying in the wilderness. But they’re going to have a really cringeworthy edgelord period until they reach that level.

All of this would be fine except that, as Graham says in the quote above, *scientists go looking for trouble*. The first virtue is curiosity. I don't know how the internal experience of curiosity works for other people, but to me it's a sort of itch I get when the pieces don't fit together and I need to pick at them until they do. I've talked to some actual scientists who have this way stronger than I do. An intellectually curious person is a heat-seeking missile programmed to seek out failures in existing epistemic paradigms. God help them if they find one before they get enough political sophistication to determine which targets are safe.

Did Giordano Bruno die for his astronomical discoveries or his atheism? False dichotomy: you can't have a mind that questions the stars but never thinks to question the Bible. The best you can do is have a Bruno who questions both, but is savvy enough to know which questions he can get away with saying out loud. And the real Bruno wasn't that savvy.

III.

So imagine the most irrelevant orthodoxy you can think of. Let's say tomorrow, the government chooses "lightning comes after thunder" as their hill to die on. They come up with some BS justification like how atmospheric moisture in a thunderstorm slows the speed of light. If you think you see lightning before thunder, you're confused – there's lots of lightning and thunder during storms, maybe you grouped them together wrong. Word comes down from the UN, the White House, the Kremlin, Zhongnanhai, the Vatican, etc – everyone must believe this. Senior professors and funding agencies are all on board. From a scientific-truth point of view it's kind of a disaster. But who cares? Nothing at all depends on this. Even the meteorologists don’t really care. What’s the worst-case scenario?

The problem is, nobody can say “Lightning comes before thunder, but our social norm is to pretend otherwise”. They have to say “We love objective truth-seeking, and we’ve discovered that lightning does not come before thunder”. And so the Kantoroviches of the world will believe that’s what they really think, and try to write polite letters correcting them.

The more curiosity someone has about the world, and the more they feel deep in their gut that Nature ought to fit together – the more likely the lightning thing will bother them. Somebody’s going to check how light works and realize that rain can’t possibly slow it down that much. Someone else will see claims about lightning preceding thunder in old books, and realize how strange it was for the ancients to get something so simple so wrong so consistently. Someone else will just be an obsessive observer of the natural world, and be very sure they weren’t counting thunderclaps and lightning bolts in the wrong order. And the more perceptive and truth-seeking these people are, the more likely they’ll speak, say “Hey, I think we’ve got the lightning thing wrong” and not shut up about it, and society will have to destroy them.

And the better a school or professor is, the better they train their students to question everything and really try to understand the natural world, the more likely their students will speak up about the lightning issue. The government will make demands – close down the offending schools, fire the offending academics. Good teachers will be systematically removed from the teaching profession; bad teachers will be systematically promoted. Any educational method that successfully instills curiosity and the scientific spirit will become too dangerous to touch; any that encourage rote repetition of approved truths will get the stamp of approval.

Some other beliefs will be found to correlate heavily with lightning-heresy. Maybe atheists are more often lightning-heretics; maybe believers in global warming are too. The enemies of these groups will have a new cudgel to beat them with, “If you believers in global warming are so smart and scientific, how come so many of you believe in *lightning*, huh?” Even the savvy Kolmogorovs within the global warming community will be forced to admit that their theory just seems to attract uniquely crappy people. It won’t be very convincing. Any position correlated with being truth-seeking and intelligent will be always on the retreat, having to forever apologize that so many members of their movement screw up the lightning question so badly.

Some people in the know will try to warn their friends and students – “Look, just between you and me, lightning obviously comes before thunder, but for the love of God *don’t say that in public*“. Just as long as they’re sure that student will never want to blackmail them later. And won’t be able to gain anything by ratting them out. And that nobody will hack their private email ten years later, then get them fired or imprisoned or burned at the stake or whatever the appropriate punishment for lightning-heresy is. It will become well-known that certain academic fields like physics and mathematics are full of crypto-lightning-heretics. Everyone will agree that the intelligentsia are useless eggheads who are probably good at some specific problems, but so blind to the context of important real-world issues that they can’t be trusted on anything less abstruse than e equalling mc squared. Dishonest careerists willing to go in front of the camera and say “I can reassure everyone, as a physicist that physics proves sound can travel faster than light, and any scientists saying otherwise are just liars and traitors” will get all the department chairs and positions of power.

But the biggest threat is to epistemology. The idea that everything in the world fits together, that all knowledge is worth having and should be pursued to the bitter end, that if you tell one lie the truth is forever after your enemy – all of this is incompatible with even as stupid a mistruth as switching around thunder and lightning. People trying to make sense of the world will smash their head against the glaring inconsistency where the speed of light must be calculated one way in thunderstorms and another way everywhere else. Try to start a truth-seeking community, and some well-meaning idiot will ask “Hey, if we’re about pursuing truth, maybe one fun place to pursue truth would be this whole lightning thing that has everyone all worked up, what does everybody think about this?” They will do this in perfect innocence, because they don’t know that everyone else has already thought about it and agreed to pretend it’s true. And you can’t just *tell them* that, because then you’re admitting you don’t really think it’s true. And why should they even believe you if you tell them? Would you present your evidence? Would you dare?

The Kolmogorov option is only costless when it’s common knowledge that the orthodoxies are lies, that everyone knows the orthodoxies are lies, that everyone knows everyone knows the orthodoxies are lies, etc. But this is never common knowledge – that’s what it means to say the orthodoxies are still orthodox. Kolmogorov’s curse is to watch slowly from his bubble as everyone less savvy than he is gets destroyed. The smartest and most honest will be destroyed first. Then any institution that reliably produces intellect or honesty. Then any philosophy that allows such institutions. It will all be totally pointless, done for the sake of something as stupid as lightning preceding thunder. But it will happen anyway. Then he and all the other savvy people can try to pick up the pieces as best they can, mourn their comrades, and watch the same thing happen all over again in the next generation.

The Church didn’t lift a finger against science. It just accidentally created a honeytrap that attracted and destroyed scientifically curious people. And any insistence on a false idea, no matter how harmless and well-intentioned, risks doing the same.

IV.

I’m not against the Kolmogorov Option. It’s nothing more than a band-aid on the problems that even a harmless orthodoxy will cause – but if there’s no way to get rid of the orthodoxy, the band-aid is better than nothing. But politically-savvy Kolmogorov types can’t just build a bubble. They have to build a whisper network.

They have to build a system that reliably communicates the state of society. “Stalin claims that he welcomes advice from everyone, but actually he will kill you if you try to give it.” Or “God probably doesn’t exist, but lots of us know this, and we all just go to Mass and mouth the right words anyway.”

This is harder than it sounds. A medieval monk being told God doesn’t exist probably has a lot of questions. He’s likely to go kind of crazy for a while, crave the worldview-shards that he needs to rebuild his fractured philosophy. “What about Heaven? Does that exist? Where do we go when we die?” (“Psssst, Epicurus has some good arguments for why the soul doesn’t survive death, you can get a copy of his books from the monastery library.”) You might even have to prevent overcorrection: “Is there even such a place as

Jerusalem?” (“Yes, now you’re just being silly, Brother Michael went there on Crusade and says it’s very nice.”)

(when a heretical belief turns out to really be completely wrong – maybe occultism would be a good example here – a whisper network might be the only place where you could get high-enough-quality debate to be sure.)

They have to serve as psychological support. People who disagree with an orthodoxy can start hating themselves – the classic example is the atheist raised religious who worries they’re an evil person or bound for Hell – and the faster they can be connected with other people, the more likely they are to get through.

They have to help people get through their edgelord phase as quickly as possible. “No, you’re not allowed to say this. Yes, it could be true. No, you’re not allowed to say this one either. Yes, that one also could be true as best we can tell. This thing here you actually are allowed to say still, and it’s pretty useful, so do try to push back on that and maybe we can defend some of the space we’ve still got left.”

They have to find at-risk thinkers who had started to identify holes in the orthodoxy, communicate that they might be right but that it could be dangerous to go public, fill in whatever gaps are necessary to make their worldview consistent again, prevent overcorrection, and communicate some intuitions about exactly which areas to avoid. For this purpose, they might occasionally let themselves be seen associating with slightly heretical positions, so that they stand out to proto-heretics as a good source of information. They might *very* occasionally make calculated strikes against orthodox overreach in order to relieve some of their own burdens. The rest of the time, they would just stay quiet and do good work in their own fields.

Such a whisper network would be in the best interests of the orthodox authorities. Instead of having to waste their good scientists, they could let the good scientists could join the whisper network, learn which topics to avoid, and do good science without stepping on orthodox toes. But the authorities couldn't just *say* this. Maybe they wouldn't even think of it, and nobody (except maybe Kantorovich) would be dumb enough to try to tell them. Individual secret policemen are always going to see the written law – "arrest heretics" – and consider the whisper network a legitimate target. Kolmogorov is doing the Lord's work, but that won't give him a pass from the Inquisition.

His reward will be that people with a drive to make the world make sense – to have everything fit together seamlessly and beautifully – will be able to quietly collect all the orthodox *and* all the heretical pieces, satisfy themselves, and then move on to doing good work in math or physics or whatever harmless field doesn't affect Christianity or Marxism or lightning or whatever. Academies other than the worst and most curiosity-crushing have a little better chance to endure; academic bureaucrats other than the most slavish have a little more chance to remain in their position.

But also: maybe this is how common knowledge spreads. Maybe some atheists survive, go into science, become vaguely aware of each other's existence, feel like they have safety in numbers, get a little bolder, and maybe the Church decides it's not worth killing all of them. Maybe everyone stays quiet until Mao dies, and then Deng and Zhao look at each other and say "So, just between you and me, all of that was totally insane, right?" I don't know how often this happens. But the chances seem better than for open defiance followed by certain retribution.

How Did New Atheism Fail So Miserably?

The Baffler publishes a long article against “idiot” New Atheists. It’s interesting only in the context of so many similar articles, and an inability to imagine the opposite opinion showing up in an equally fashionable publication. New Atheism has lost its battle for the cultural high ground. r/atheism will shamble on as some sort of undead abomination, chanting “BRAAAAAAIIINSSSS…are what fundies don’t have” as the living run away shrieking. But everyone else has long since passed them by.

The New Atheists accomplished the seemingly impossible task of alienating a society that agreed with them about everything. The Baffler-journalists of the world don’t believe in God. They don’t disagree that religion contributes to homophobia, transphobia, and the election of some awful politicians – and these issues have only grown more visible in the decade or so since New Atheism’s apogee. And yet in the bubble where nobody believes in God and everyone worries full-time about sexual minorities and Trump, you get less grief for being a Catholic than a Dawkins fan. When Trump wins an election on the back of evangelicals, and the alt-right is shouting “DEUS VULT” and demanding “throne and altar conservativism”, the *real* scandal is rumors that some New Atheist might be reading /pol/. How did the New Atheists become so loathed so quickly?

The second article presents a theory:

> It has something to do with a litany of grievances against the believoisie so rote that it might well (or ironically) be styled a catechism. These New Atheists and their many fellow travelers all share an unpleasant

> obsessive tic: they mouth some obvious banality—there is no God, the holy books were all written by human beings—and then act as if it is some kind of profound insight. This repetition-compulsion seems to be baked right into their dogma.

It compares New Atheists to Kierkegaard's lunatic:

> Soren Kierkegaard, the great enemy of all pedants, offers a story that might shed considerable light. In his Concluding Unscientific Postscript, he describes a psychiatric patient who escapes from the asylum, climbing out a window and running through the gardens to rejoin the world at large. But the madman worries: out in the world, if anyone discovers that he is insane, he will instantly be sent back. So he has to watch what he says, and make sure none of it betrays his inner imbalance—in short, as the not-altogether unmad Danish genius put it, to "convince everyone by the objective truth of what he says that all is in order as far as his sanity is concerned." Finding a skittle-bowl on the ground and popping it in his pocket, he has an ingenious idea: who could possibly deny that the world is round? So he goes into town and starts endlessly repeating that fact, proffering it over and over again as he wanders about with his small furious paces, the skittle-bowl in his coat clanking, in strict conformity with Newton's laws, against what Kierkegaard euphemistically refers to as his "a–." Of course, the poor insistent soul is then sent right back to the asylum […]
>
> Kierkegaard's villagers saw someone maniacally repeating that the world is round and correctly sent him back to the asylum. We watched [Neil de Grasse] Tyson doing exactly the same thing, and instead of hiding him away from society where nobody would have to hear

> such pointless nonsense, thousands cheer him on for fighting for truth and objectivity against the forces of backwardness. We do the same when Richard Dawkins valiantly fights for the theory of evolution against the last hopeless stragglers of the creationist movement, with their dinky fiberglass dinosaurs munching leaves in a museum-piece Garden of Eden. We do it when Sam Harris prises deep into the human brain and announces that there’s no little vacuole there containing a soul.

So the problem with New Atheism was that its whole shtick was repeating obviously true things that everyone already knew? But about 80% of Americans identify as religious, 63% claim to be “absolutely certain” that there is a God, and 46% think the world was literally created in seven days. This is a surprising number of people disagreeing with a thing that everybody already knows.

I could be misreading the article. The article could be wrong. But I don’t think so. This is *my* intuitive feeling of what was wrong with New Atheism as well. It wasn’t that they were wrong. Just that they were right in a loud, boring, and pointless way.

A charitable reading: New Atheists weren’t reaching their intellectual opponents. They were coming into educated urban liberal spaces, saying things that educated urban liberals already believed, and demanding social credit for it. Even though 46% of America is creationist, zero percent of my hundred-or-so friends are. If New Atheists were preaching evolution in social circles like mine, they were wasting their time.

This seems like an accurate criticism of New Atheism, one that earns them all the condescension they have since received. But the New Atheist still ought to feel betrayed.

Why isn’t this an equally correct criticism of *everything else*?

While the atheists were going around saying there was no God, the environmentalists were going around saying climate change was real. The feminists were going around saying sexism was bad. And the Democrats were going around saying Donald Trump was an awful person. All of these statements might be controversial *somewhere*, but meet basically zero resistance in educated urban liberal spaces. All get repeated day-in and day-out by groups of people who make entire careers out of repeating them. And all get said in the same condescending way, a sort of society-wide plague of Voxsplaining.

This is 90% of popular intellectual culture these days: progressives regurgitating progressivism to other progressives for nothing but the warm glow of being told “Yup, that was some good progressiving there”. Conservatives make fun of this incessantly, and they are right to do so. But for some reason, in the case of New Atheism and *only* in the case of New Atheism, Progressivism itself suddenly turned and said “Hey, you’re just repeating our own platitudes back to us!” And New Atheism, caught flat-footed, mouth open wide: “But…but..we thought we were supposed to…we thought…”.

Think of one of those corrupt kleptocracies where the dictator takes bribes, all his ministers take bribes, all their assistants take bribes, the anti-corruption task force takes bribes, etc. Then one day some shmuck manages to get on the dictator’s bad side and – bam – the secret police nab him for taking bribes. The look on his face the moment before the firing squad shoots – that’s how I imagine New Atheists feeling too.

So who’s the dictator in this analogy? And what did New Atheism do to get on their bad side?

Maybe New Atheism failed to make the case that it was socially important. All these movements have a mix of factual claims and social calls to action – climate change activism combines “we should accept the scientifically true fact that the climate is changing” with “we should worry about climate change causing famines, hurricanes, etc”, just as atheism combines “we should accept the scientifically true fact that God does not exist” with “we should worry about religion’s promotion of terrorism, homophobia, et cetera”. But the climate change people seem better at sounding like they care about the people involved, compared to atheists usually sounding more concerned with Truth For Its Own Sake and bringing in the other stuff as a justification.

Or maybe the New Atheists just didn’t know how to stay relevant. Trump resistance always has new tweets to keep its attention. Social justice always has a new sexist celebrity to be angry about. Sure, a few New Atheists tried to keep up with the latest secretly-gay televangelist, but most of them kept going about intricacies of the kalam argument that had been done to death by 1400 AD. This is just an example – maybe there are other asymmetries that are more important?

Maybe the New Atheists accidentally got on board just before a nascent Grey Tribe/Blue Tribe split and tried to get Blue Tribe credibility by sending Grey Tribe signals. At some point there was a cultural fissure between Acela Corridor thinkfluencers with humanities degrees and Silicon Valley bloggers with STEM degrees, and the former got a head start on hating the latter while the latter still thought everybody was on the same anti-Republican side.

And the cynic in me wonders whether New Atheism wasn't pointless and obvious *enough*. There are more church-goers in educated liberal circles than Trump supporters, climate deniers, or self-identified racists. Maybe that made the "repeat platitudes to people who already believe them" game a little less fun, caused some friction – "You're talking about my dear grandmother!"

I don't know. The whole problem is so strange. For a brief second, modern culture looked at New Atheism, saw itself, and said "Huh, this is really stupid and annoying". Then it cast New Atheism into the outer darkness while totally failing to generalize that experience to anything else. Why would it do that? Could it happen again? *Please* can it happen again? *Pretty* please?

Against Rat Park

I.

Rat Park is a famous study in which lab rats were kept in a really nice habitat that satisfied their every need. Contrary to the usual results with lab animals, scientists couldn't get these happier rats addicted to drugs. Researchers concluded that drug addiction, far from being the simple biological story everyone assumed it was, was really a just coping mechanism for intolerable social situations. Rats stuck in terrible cages get addicted to drugs, as do humans in terrible slums. But give them other opportunities for happiness, and the problem disappears.

This has become a sort of popular legend. From HuffPo: The Likely Cause Of Addiction Has Been Discovered, And It Is Not What You Think. From Intellihub: Rat Park Heroin Experiment Shows Cultural Roots Of Drug Addiction. There's even a Rat Park Comic and the inevitable Trump Could Learn From The Rat Park Experiment thinkpiece.

Rat Park lead researcher Bruce Alexander eventually developed a whole alternative theory of drugs, provocatively titled The Myth Of Drug-Induced Addiction. He notes that there are lots of people who take supposedly-addictive drugs without getting even the least bit addicted to them. The particular example he gives are patients who were prescribed cocaine for chronic pain, did well on it, and never abused the cocaine or escalated the dose. Even people who use it illicitly are unlikely to get addicted: "American and Canadian population surveys indicate that merely having used cocaine is associated with less than a 10% chance of having it as often as 100 times". He uses this to draw a picture where cocaine isn't really "addicting" per se, but people can repeatedly use drugs for “nondeterministic” reasons; for example, “certain people choose an addictive lifestyle as a lesser evil in desperate circumstances”.

Alexander makes his point with cocaine, but others working along the same lines have an impressive collection of related anecdotes. In World War II the Nazis, never ones to pass up unethical applications of science, put much of their army on methamphetamine to give them more energy and stamina. After the war, these soldiers reintegrated back into German society without much trouble; there was no epidemic of continuing meth addiction. Likewise, people worry about the opiate epidemic today, but only about 10% of people put on opiates for chronic pain become addicted; most take it as prescribed and then stop without any problems when their pain gets better. Even more obviously, the vast majority of social drinkers stay social drinkers; only a fraction become alcoholics. So – Alexander and his followers would tell us – clearly non-biological variables must be involved. And Rat Park tells us what those variables are: the social factors that make the difference between a terrible life and a flourishing one.

II.

How does this argument fare?

Two studies (1, 2) tried to replicate the results of Rat Park and failed. Another two (1, 2) tried and mostly succeeded. There’s some concern that the rat strain involved might have various substrains that the different experiments didn’t control for. But a result that can’t survive a change in rat substrains has pretty dismal prospects for applicability to humans.

There seems to be a more general problem with rat drug experiments in general, where everyone in the field realizes they can sometimes throw out weird results. The Drug Monkey blog says that rats self-administering cocaine is the best-replicated result in drug abuse science, *but also* that it can be screwed up by anything from the diameter of the cocaine infusion catheter, to whether the experimenter is wearing a dirty vs. clean lab coat, to who you buy the rats from.

So let's take a step back and look at 13th century Mongolia.

Who had the best life – the happiest, most fortunate, most blessed existence in all of history? I don't know, but Ogedei Khan had to be in the top ten. The son of Genghis Khan, he took over his father's empire and continued the family tradition of conquering things. China, Korea, and Eastern Europe all fell to his armies; for over a decade he conquered and conquered without suffering a single serious defeat.

For his enjoyment, Ogedei built the Tumen Amgalan, "Palace Of Myriad Peace", 2500 square meters in area. It had:

> 64 pillars, a green-and red-tiled, Chinese-style roof, and green-tiled floor built over a heating system. When the khan met important visitors he sat at a great throne with one staircase for ascending and another for descending. Inside the palace was a fountain, designed by a French jeweler captured in Hungary, that embraced a silver tree that spewed out mare's milk from a lion's head and wine, rice wine, mead and airag from from four snake heads. At the top was an angel with trumpets that blew and dispensed more drink.

We don't know the exact size of Ogedei's harem, but his father Genghis kept about three thousand concubines, and his successor Kublai had seven thousand, so if this kind of thing increases linearly, we can estimate Ogedei had somewhere around 5,000. The records, shamefully imprecise, tell us only that he took “pleasures in the company of beauteous ladies and moonfaced mistresses”.

Of his own character, Wikipedia says:

> Ögedei was considered to be his father’s favorite son, ever since his childhood. As an adult, he was known for his ability to sway doubters in any debate in which he was involved, simply by the force of his personality. He was a physically big, jovial, and very charismatic man, who seemed mostly to be interested in enjoying good times. He was intelligent and steady in character.

So if you had to be a historical figure, you could do worse than Ogedei Khan.

I mention this because he was a raging alcoholic who drank himself to death.

History tells us that the Khan’s advisors were so concerned about his drinking that they staged a sort of medieval Mongolian intervention, begging him to cut down. Ogedei was moved by their plea, and swore to limit himself to one cup of alcohol a day from then on. But he later regretted his pledge and ordered the imperial artisans to create a purpose-made giant cup that could hold as much alcohol as he wanted. He died at 55, probably of alcohol poisoning. His successor, equally blessed, also became an alcoholic. So did several other members of the Khan family, rich and victorious all.

So: How do we square the idea that addiction is a disease of deprivation and misery with the addiction of the Great Khan, happiest and most fortunate of men?

And what about the American Indians during colonial times? Not the ones who were being displaced and uprooted by white settlers. The ones who were still participating in their traditional lifestyle with only distant relationships with the British colonists. Without wanting to go full on noble-savage or minimize the difficulties of a hunter-gatherer lifestyle, they seemed to have things pretty good. We know that Indians taken captive by white men found European civilization intolerable, but whites taken captive by Indians enjoyed Indian life so much that they often refused to return home once rescued. Most of the testimony that's come down to us from historical Indians and their white observers indicates that their lifestyle was, if not perfectly idyllic, at least enviable.

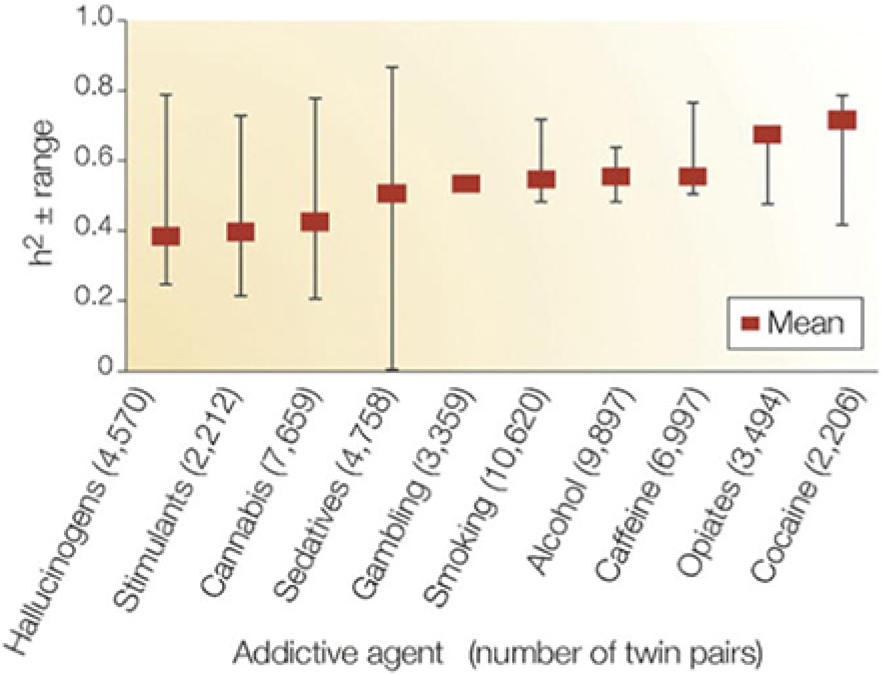
But the early Native Americans were famously susceptible to alcohol, far more so than whites, so much so that entire tribes were destroyed by alcoholism epidemics and the ones that weren't had to ban the substance as an existential threat.

And if neither Khans nor Indians seem happy enough, what about celebrities? They seem to have it all – wealth, power, fame, groupies. But they still get addicted to drugs at a high enough rate to produce six seasons of *Celebrity Rehab With Dr. Drew* (key quote: "In May 2013, Pinsky announced that season six was the final season, as he was tired of the criticism leveled at him after celebrities he treated had relapsed into addiction and died").

Can you invent some convoluted definition of flourishing which classifies stormtroopers, chronic pain patients, and enriched-environment-rats as flourishing, but Khans, Indians, and celebrities as deprived? Maybe. But by that point the Rat Park hypothesis is becoming unfalsifiable; any seemingly flourishing group that becomes addicted will just lead to more twists in the definition of “flourishing”. Better to just admit people can succumb to addiction no matter what their apparent happiness level.

III.

A more responsible look at addiction would emphasize that it’s at least half genetic:



In a lot of cases, we know the genes involved. They’re usually not surprising: smoking is linked to genes for nicotine receptors; alcoholism is linked to genes for GABA receptors. In other cases, we don’t know what genes are involved yet, but we know from twin studies that they have to be there somewhere.

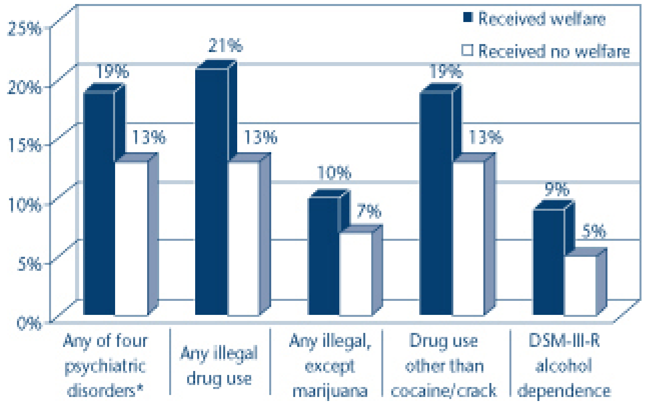
Some of these genes seem to differ across populations. The best-studied are different frequencies of *ADH* and *ALDH* polymorphisms. Asians in particular tend to have some very protective alleles here in a way showing signs of recent selection, with the extreme version being the alcohol flush reaction seen in about a third of East Asian individuals.

There’s a plausible story where cultures that have lived with potent forms of alcohol for millennia have evolved some level of biological resistance to it, whereas cultures that didn’t get it until recently (like Mongolians and Native Americans) are worse off.

This lets us accept some of the Rat Park data without having to swallow their conclusion. Yes, most chronic pain patients can take opiates for years and never get addicted. Most cocaine users experiment once or twice but don’t end up as cokeheads. And most social drinkers never become full-on alcoholics. But this is a consequence of genetic luck, not enriched vs. less-enriched environments, and we don’t have to throw out the entire concept of “addictive drug” just to deal with it.

But does environment affect addiction at all?

This is really hard to solve empirically. You can find a lot of pictures like this:



But the studies that produce these kinds of graphs aren’t great. For example, they can’t rule out that drug addicts become poor as a result of their drug addiction. Even if they look at people born in poverty, they can’t rule out genetic confounds – ie your parents abuse drugs and become poor, and you’re born in poverty with the genes for drug addiction.

Sometimes they come up with different answers depending on how you define the outcome of interest. Maybe poor people have more drug *abuse* but rich people have more drug *dependence*, or vice versa. In one study, rich people were more likely to have dangerous drinking-related behaviors, but only because they were more likely to drink and drive because they were more likely to own cars. Often rich people will be found to abuse one drug more, and poor people another, with unclear overall significance.

The better studies seem to have confusing conclusions like that of Patrick et al:

> Smoking in young adulthood was associated with lower childhood family SES, although the association was explained by demographic and social role covariates. Alcohol use and marijuana use in young adulthood were associated with higher childhood family SES, even after controlling for covariates.

This seems a lot more subtle than the Rat Park model of “only people in bad environments ever get addicted and addiction is entirely an environmental condition.”

IV.

But at some point you have to try using common sense, and common sense tells me that unhappy people use drugs more than happy people. At least, individuals seem to use drugs when they’re least happy, and be best able to quit when everything is going well.

A lot of the genes, neural correlates, imaging findings, [insert science term here] of addiction involve the reward system, especially dopamine and the nucleus accumbens. The story goes something like “drugs are really rewarding, your brain does whatever gets it lots of reward, therefore your brain compels you to do drugs”.

And without having a great scientific explanation of exactly how this works, I think that having other sources of reward besides drugs available can attenuate this effect. Nothing in real life is ever going to be quite as rewarding as heroin. But if having a happy family and doing meaningful work and so on are, I don’t know, a quarter as rewarding as heroin, then maybe you have an opportunity to use “““willpower””” to force yourself the rest of the way and take the only-a-quarter-as-rewarding option. If you have a crappy life that doesn’t have any other source of reward at all, then all the “““willpower””” in the world won’t save you.

Imagine someone engaged in a scummy job they feel bad about doing, like spammer or tobacco advertiser or bioethicist. They want to quit, but they get a really good salary – let’s say $100,000/year – and they’re not sure they can do without the money. If they’ve got another wholesome option that pays $50,000, maybe their conscience is strong enough to help them make the leap. If their only other option pays $15,000, that’s going to be a lot harder. This is how I think about drugs too.

The rest of this post, Parts I II and III, is pretty sound science. This part is just my toy model. But it seems like something of this sort has to be true. Society and genetics have to interact somehow, and this seems about as good a way as any.

But remember: society is fixed, biology is mutable. You should work on the super-long-term project of improving society because that’s the right thing to do. But if you want to fight drug addiction now, in an effective way with some hope for helping today’s addicts, the best choices are still deregulating suboxone and legalizing research into psychedelic therapy.

Postmodernism For Rationalists (my attempt)

EDIT: Been told by people I trust that this is not a good explanation. Retracted.

I.

Some of the Seattleites put together a Postmodernism For Rationalists presentation that's been sparking a lot of discussion. It's not quite the way I would have explained things. I'm no expert in postmodernism, and can't give anything more than a very simple introduction to one of many facets of the movement. But I *am* an expert in explaining things to rationalists. So it's worth a try.

Last week, I went over the evidence for and against a European Dark Age. Most people on both sides agreed on some facts in favor, like:

1. The fall of Rome was associated with a decline in wealth and population in Western Europe.
2. The fall of Rome was associated with a loss of capacity for things like urban living and large-scale infrastructure in Western Europe.
3. The intellectual output of the period included less literature and philosophy of lasting value than periods before and after

And on some facts against, like:

4. A lot of the worst-looking trends actually started before the fall of Rome, and were getting *better* by the time Rome finally collapsed.
5. Any collapse was relatively circumscribed both

geographically (didn’t extend to the Islamic world) and temporally (didn’t last into High Middle Ages)
6. Few people thought the world was flat, the Church mostly didn’t persecute scientists, and there were lots of knowledgeable and cultured people.

So was there *really* a Dark Age? This question is meaningless insofar as there’s no objective criteria for what a Dark Age is. Even if there were (for example, historians decided that any century in which economic activity contracted by greater than 10% was a Dark Age) those objective criteria would themselves be subjective (why didn’t the historians decide 5%?). Any assessment would have to combine objective historical facts with subjective beliefs about what metrics were important.

And these subjective metrics will inevitably include political agendas. For example:

A. New Atheists might want to say there was a Dark Age, because it supports a narrative of religion crushing progress.
B. Classicists might want to say there was a Dark Age, because it makes Greece and Rome look extra good if their fall destroyed everything.
C. Conservatives might want to say there was a Dark Age, because it reminds us that civilization is fragile and needs to be protected from barbarism.

But:

D. Catholics might want to deny there was a Dark Age, because they were ascendant during that period and it would be pretty embarrassing.
E. Multiculturalists might want to deny there was a Dark Age, because a Dark Age would imply that some cultures and time periods are better and more civilized, while others

are superstitious and barbarous.
F. Whigs might want to deny there was a Dark Age, because it hurts their narrative of ever-improving human progress.

Postmodernism is the belief that thinking about agendas A through F is *at least* as important as thinking about facts 1 through 6.

A postmodernist will be less interested in recalculating the GDP per capita of Carolingian France for the twentieth time, and more interested in investigating how all of the historians who invented the idea of Dark Ages were atheists who emphasized all of the most sensational examples of the Church screwing things up in order to fit their agenda. They focus less on objective facts than on how politically-motivated people choose to weave those facts together to tell one story rather than another. Then they explain how many of the stories and concepts we unquestioningly believe come from processes like this.

In the Dark Age debate, they'd probably just notice that lots of the people defending the Dark Age's existence are atheists – which wouldn't be too interesting. More creative postmodernist scholarship finds cases where there *isn't* yet much debate, and explains why it might be in the interest of society as a whole (or rich capitalists, or white males, or whoever else they think controls society) to tell only that side of the story. For example everyone knows "the Roman Empire fell in 476 AD" even though it clearly didn't; the Roman Empire survived until 1453 AD, but after 476 we arbitrarily switch to calling it "Byzantium". Maybe that's because our civilization has a line of descent that passes through Charlemagne, who was personally interested in having Rome be gone so he could be Holy Roman Emperor? Maybe it's because the West is pretty proud of being the heirs to Rome, Russia is pretty proud of being the heirs to

Byzantium, and the West doesn't want to have to share its patrimony with dirty Commies? I don't know. But this seems like the sort of question a postmodernist scholar would look into.

Postmodernists don't necessarily deny the existence of objective facts. But they find them hard to pin down. It would be tempting to say "It's a matter of opinion whether the fall of Rome started a Dark Age, but it's an objective fact that the Roman Empire fell in 476 AD." But we just saw otherwise. Even a more careful statement like "the Western Roman Empire fell in 476 AD" is iffy; barbarians deposed Emperor-claimant Romulus Augustulus in that year, but an equally legitimate claimant, Julius Nepos, ruled some Roman territories until 480. The date of the Western Roman Empire's fall implicitly depends on whether you support Augustulus' or Nepos' claim to the throne.

Maybe there's still an objective fact like "Romulus Augustulus, who claimed to be Roman Emperor, was deposed in 476 AD". I can't think of anyone annoying enough to dispute that, or any non-ridiculous argument against. But post-modernists would draw the boundary of 'objectivity' at "how annoying and ridiculous it would be to argue" rather than as a fundamental feature of the world. This doesn't deny objective reality as a generator for these kinds of statements. It just urges extreme wariness about claims that any particular human statement captures it perfectly.

In practice, practically all of us are postmodernists sometimes. Last week I asked about why New Atheism failed as a cultural movement. Some people blamed certain facts about New Atheists, like:

1. New Atheists are often annoying socially inept people
2. New Atheists often repeat stuff to people who already know it, the equivalent of telling everyone incessantly that the world is round
3. New Atheists had trouble with accusations of sexism, racism, and harassment.

But other people pointed out that these "facts" were negotiable. New Atheists are annoying and socially inept only if our type specimen is a teenager posting "CHECKMATE FUNDIES" on r/atheism, rather than the suave and charismatic Christopher Hitchens sorts. New Atheists often repeated stuff everyone already knew, but no more than global warming activists, Trump #Resistance members, and other groups who are still in good standing. New Atheists had trouble with accusations of sexism etc, but were they worse than others, or did the accusations just stick to them better?

These questions inspired a second range of explanations for New Atheism's decline: it was no longer useful to people in power. For example, the Democratic coalition kept New Atheism around to embarrass GW Bush style fundamentalists on the opposite side. But by the late 2000s, the evangelical Christians were no longer such an important political force, and Muslims had become a key part of the Democrats' strategy. New Atheism switched from being convenient to the Powers That Be on the left, to being inconvenient – so the Powers arranged to have it crushed.

(these aren't literal Powers any more than there's a Council Of Males taking actions enforcing the patriarchy; it's just a shorthand for what happens when people organize to protect their own interests)

In these explanations, everything else is *downstream* of the Powers' decision to crush New Atheism. New Atheism was inconvenient, therefore the Powers arranged for coverage to focus on its most annoying members. New Atheism was inconvenient, therefore the Powers arranged to have the sexism and harassment allegations against it stick.

If this sounds paranoid to you, keep in mind that something similar is a pretty standard explanation of why the Harvey Weinsten assault stories came out now rather than twenty years ago. Weinstein was a Power, or at least a friend of the Powers, so he was untouchable. Then he became weaker, and we learned he was terrible. There was an objective reality – he really did assault lots of people. But the way the rest of us understood it was filtered by his social position and by the media industry's interests.

And if this still sounds paranoid to you, fine – lots of postmodernist theories sound pretty paranoid. The point is that this isn't just about college professors who think history books were written by the patriarchy. It's something everybody does on every point in the spectrum. Trump supporters who say the rest of us are deluded by the liberal media are making a postmodernist claim – this is why there are so many articles calling Trump our "first postmodern president". I myself think like a post-modernist every time I see a new study come out saying that priming implicit-associated stereotype-threatened ego-depleting power poses has proven that plumbers are racist. I don't react with "I guess that proves plumbers are racist". I react with "Huh, apparently the Two Minutes Hate is being directed at plumbers today, I wonder why the Powers are angry at them." The information content is boring; only the signals it carries about who's in power and what they want are worth examining.

This way of thinking makes its way into all the other stuff we call "postmodernist" – postmodernist art, postmodernist literature, postmodernist architecture, etc. For example, postmodern literature sometimes uses unreliable narrators, transforming a text from an objective records (albeit of fictional events) into a heavily filtered account where we have to consider what our source is trying to make us believe. This is one of my favorite examples of a similar technique.

II.

Optimistically, postmodernism isn't necessarily opposed to rationality: rationality says to believe what is true, and it's true that people are often trying to manipulate us. Hopelessly optimistically, both groups are engaged in the same project of overcoming bias, just taking different perspectives.

But pessimistically, there's a risk that postmodernism collapses into people ignoring any facts they disagree with, arguing that facts are just mutable products of hostile power structures trying to perpetuate themselves. And by "pessimistically there's a risk", I mean "this has obviously been going on for decades". Again, I'm not just complaining about lefty professors here. The guiding narrative of Breitbart – that Cultural Marxists have taken over society and are indoctrinating everyone with their own concepts and language in order to make it impossible to think outside their boxes – is postmodernist as hell.

This leads to the classic freshman-philosophy critique of postmodernism: "Postmodernism says nothing is objectively true and it's all just opinion. But in that case, postmodernism isn't objectively true and it's just your opinion." Make this a little more sophisticated, and we can get an at-least-sophomore-level critique: "Postmodernism says that facts have enough degrees of freedom that they often get reframed to support the powerful. But there are bucketloads of degrees of freedom in how to use and apply postmodernism; it's inevitably going to itself be twisted to support the powerful." Whether "the powerful" are cultural Marxists, or far-right media, or whatever, gets left as an exercise for the reader.

(cf. the Less Wrong Sequences, Knowing About Biases Can Hurt People)

I don't know if the postmodernists have a good solution to this problem (I think I know what their bad solution is, but I'll stay quiet lest I strawman a movement I'm not qualified to speak for). But the rationalist solution is some combination of The Fallacy of Gray and The Lens That Sees Its Flaws. Yes, everything potentially contains elements of subjectivity, but not everything contains exactly the same level of subjectivity; there's still some real way in which Breitbart News is worse than BBC. And although our perceptual and cognitive apparatus is biased, it has some non-zero connection to objective reality that we can bootstrap from to reach reflective equilibrium.

I expect most postmodernists would consider this hopelessly naive. But I don't know what the alternative is.

Non-Expert Explanation

SSC's review of postmodernism got very mixed reviews. Some of them made a good point: why should I be trying this at all? I'm not a postmodernist, I'm not a philosophy professor, surely someone much more qualified has already written a blog-post-length explanation of postmodernism.

This is all true. My only excuse is that trying to figure out complicated concepts requires a different approach than trying to teach simple ones.

Some knowledge is easy to transfer. "What is the thyroid?" Some expert should write an explanation, anyone interested can read it, and nobody else should ever worry about it again.

Other knowledge is near-impossible to transfer. What about social skills? There are books on social skills. But you can't just read one and instantly become as charismatic as the author. At best they can hint at areas worth exploring.

There are lots of books about social skills, and there *should* be lots of them – I don't know which social skills book is the best, but it doesn't obsolete all the others. Maybe it's because you need a kind of triangulation – one person's views on social skills give you one perspective, another person's views on social skills give you another perspective, and after reading enough different books you can sort of make out the shape of the territory in question. Maybe it's about different people having different problems and deficiencies, but our language is imprecise enough that they all get called "Social Skills" and it won't help unless you stumble across the one solving your specific issue. Or maybe it's about different people's minds working in different ways, so that you can only make sense of a book by someone who thinks like you – whose mind groups things into the same concepts as yours, so that you can import them over directly. There are some Social Skills For Autistic People books out there, and the autistic people I know say they're much more helpful than the generic-brand. Autism is a well-circumscribed thing; how many less-well-circumscribed groups are out there with similar needs?

Complicated ideas are like this too. I remember reading a mathematician talking about how there were two different-but-equivalent formulations of some high-level mathematical concept – let's say an algebraic one vs. a geometric one. He'd always learned it as the algebraic one and had only the slipperiest grasp on it. Then one day he read a textbook presenting the equivalent geometric version, and it made perfect sense; he really understood it, could mentally manipulate it, could think creatively with it and make progress. He wondered why everybody didn't teach the geometric version first. Another mathematician responded that he had the same story – except that for him, he'd learned the geometric version first, hated it, and only really been able to make progress once he learned the algebraic version. Then a third mathematician chimed in, said that both the geometric and algebraic version had confused her, but that in some obscure textbook she was able to find a third equivalent formulation she thought was better than either.

My own version of this experience was reading Eliezer Yudkowsky's A Human's Guide To Words, which caused a bunch of high-level philosophical ideas to slip neatly into place for me. Last week David Chapman wrote about what was clearly the same thing, even centering around the same key example of whether Pluto is a planet. A Gender Studies major I know claims (I can't confirm) that the same thing is a major part of queer theory too. But Chapman’s version and queer theory don’t make a lot of sense to me; I was able to understand the former only because I already knew what he was talking about, and I have to take any statements about the latter on pure faith. On the other hand, nobody else seems to have found Guide To Words as important as I did; I don’t see paeans to it all over, nobody’s offering Eliezer any Nobel Prizes. It was a perfect fit for where my mind was at that moment – but there are probably a hundred other versions equally objectively good, some of which don’t even realize they’re versions of the same thing.

To carry on the analogy to social skills: even after reading the best, most perfect-fit social skills book in the world, it’s still not going to be enough. People need to ask questions. Both in my psychiatrist role and my community-member role, I have to answer (and sometimes ask) a lot of “Hey, is this socially acceptable? What’s the best way to behave here?” type questions.

And questioning requires mental fit at least as much as straight information-transfer does. Speaking of having poor social skills, I remember what I used to be like in college. A professor would say something that didn’t make any sense to me. I at least had the social skills to avoid saying “that doesn’t make any sense”, so I would raise my hand and ask the professor “Excuse me, I don’t understand what Aristotle meant when he said everything had a telos. Do snails have telos? Do air molecules? Does a random rock?” The professor would mumble something kind of meaningless that didn’t answer the question, and again being too polite to say so, I would say “I’m not quite sure what you meant by that ‘only specific things have a telos’. Which specific things are you talking about? How would we figure out which ones?” And then so on, until I became more and more exasperated with the professor seemingly giving irrelevant responses or completely misunderstanding my questions, and the professor started thinking I was some sort of hostile troll trying to embarrass him. I quickly learned that there were some professors, tutors, and fellow students who would immediately understand what I was asking and answer as best they could, and others who would go through the motions of answering while leaving me even more confused than before.

And continuing on the social skills analogy even further: at some point you have to go to a party, try out what you know, and totally humiliate yourself. The intellectual version is something like steelmanning – you try to construct the position you're trying to understand as best you can, then see if it sounds right to people who know about it.

One of the great things about the old Less Wrong was that it was a community built for this kind of thing. A bunch of people with a certain worldview and way-of-thinking explained some curated hard-to-understand knowledge to other people who also shared their worldview and way-of-thinking. Then they discussed it among themselves, questioned it back and forth, agreed or disagreed with it, and absorbed it in a social way. This is also what I'm trying to do with SSC. The knowledge itself may or may not be original – I think at a certain level of complexity "originality" becomes hard to monitor (what percent of the 10,000 psychology books that have been published are truly "original"?). But it's packaged slightly differently than what's come before, and it's well-targeted at a community of people who have the right mental fit to absorb it and then refine it among themselves.

Some of the academics I know say similar things about their own field. It's not just that you have to read lots of books, although you do. It’s the experience of working with an advisor and other grad students, of coming up with theories and having them be shot down. Two stories I’ve heard from multiple grad student friends: “I spent two months working on something really cool, and in the first thirty seconds of presenting it to my advisor she came up with a simple proof it could never work” and “I spent two months working on something really cool, and in the first thirty seconds of presenting it to my advisor, she said ‘Oh yeah, that’s Smith’s Lemma, very exciting when it was published forty years ago.'” But eventually you come out of it not just with book learning, but with the thought-patterns and methods of a field baked into your brain, a strong sense of what is or isn’t interesting, can or can’t be done.

The spiritual traditions seem to endorse some similar process. They have some complicated thing you’re supposed to ‘get’ – enlightenment, gnosis, whatever. They make a big deal of how it’s useless to communicate in words. But they also make a big deal of reading the scriptures, of having teachers, of the importance of back-and-forth conversations with teachers beyond just reading books and listening to lectures. So you read lots of sutras, and you do lots of meditation, and you talk to your guru a lot, and then suddenly (at least in some traditions), it makes sense. You see a falling leaf, or you hear a raindrop, or someone hits you with their stick, or something else that’s never the same for two different people, and you get it. I know the suddenness aspect is exaggerated, I know there are some traditions that say it’s not like this at all, but they all share this view of a knowledge which can’t be mass produced through traditional educational methods.

Maybe this is on my mind because of the recent post on Kolmogorov complicity. Some people asked – why can’t people just figure out what’s taboo, either believe it quietly or reject it openly, and then shut up about it? And part of the answer has to be that the process of coming to understand a field at all has to involve this pattern of back-and-forth questioning, approaching from multiple sides, devil-advocating, etc. Lots of the process will look the same whether you end out ultimately rejecting or accepting a truth; you've got to go through the same steps just to understand what you're considering.

The Internet seems like an increasingly hostile place for this sort of thing. I can't remember how many times I've read an essay I really liked and appreciated only to see somebody mocking it for "reinventing the wheel". Oddly enough, none of these people ever point out who said the thing first, or what its standard name is. Maybe they think it's too obvious to mention? Or, if someone screws up, or asks a stupid question, it gets screenshotted and goes viral all over Twitter as "Look what this stupid person said now!" I will admit being complicit in this – I get really nervous whenever someone posts something unsophisticated in the comments here or on the subreddit, because I'm worried it will go viral as an example of "what those people at Slate Star Codex believe". I'm not even talking about offensive things here! Just stupid ones!

There's an awkward tension between blogs and comments as "something some random person has typed into a box on the Internet" vs. "strong claim to authority and of being worthy to educate everyone else". Offline it's easier to distinguish these sorts of things – tone of voice, what kind of situation you're in, whether you preface it awkwardly with "This is stupid, but…", whether you're just talking to your equally-stoned friend. On the Internet, having a blog gives this aura of "Hey, I'm going to educate you about things using my superior knowledge". I try to fight that with epistemic status tags explaining when things are tentative or just me looking for feedback, but I guess maybe these are sometimes hard to believe. Sometimes they just earn more anonymous hate: “If you’re admitting you’re too stupid to have an opinion on this, you must be *really* stupid to give it anyway!”

This is a shame. The authoritative-lecture format works for facts, but isn’t enough when you’ve got any subject more complicated than thyroid anatomy. Collaborative truth-seeking where people are throwing out ideas, trying to reconstruct arguments themselves, asking questions, and arguing – these are more promising, but they leave you open to accusations of reinventing the wheel, arrogantly dabbling in fields you don’t understand, or being too insular. When some of the topics involved are taboo, add the sins of “just asking questions” or “thinking it’s my job to educate you”. But unless you’re such a good lecturer that everybody will understand you on the first try, this is a necessary part of communicating hard things.

Concept-Shaped Holes Can Be Impossible To Notice

I.

When I wrote about my experiences doing psychotherapy with people, one commenter wondered if I might be schizoid:

> There are a lot of schizoid people in the rationalist community from what I can tell. The basis of schizoid is not all the big bad symptoms you might read about. There are high functioning people with personality disorders all the time who are complex, polite and philosophical.
>
> You will never see this description because mental health industries center entirely around people Failing At Life, aka “low-functioning”. As many radicals have noted, mental health tends to constitute itself mostly around “can’t hold a job” or “can’t hold a marriage”.
>
> The only thing you need to be schizoid is to dislike contact with other egos, and to shave off the experience of those other egos ruthlessly before they can reach the fantasy world you retreat to.
>
> It doesn’t mean you’re evil. It doesn’t mean you stalk people and plan to harm them. It doesn’t mean you’re over-reactive or even bizarrely delusional. You could call it a form of delusion, but really the basic descriptions of perception like top-down processing and culture could all be called delusional thinking if you want to be properly pointed about it. It’s schizoid. It’s often quite gentle. And I’ve noticed from interacting with various

> people in high IQ communities that if you have sufficiently high enough intelligence, despite the inherent defined tendency to retreat from reality, you can in fact become aware you have a personality disorder.
>
> Anyway, my guess based on projection (I've never met you) is that people aren't being emotional around you because you can't be reached by them emotionally, and they know that on some level.

I *feel like* I experience emotions and genuine human connection. You would think that 'not experiencing emotions or having genuine human connection' is hard to miss. But then I think of the stories in What Human Experiences Are You Missing Without Realizing It?

In the first, Francis Galton discovered that some people didn't have visual imagination. They couldn't see anything in their "mind's eye", they couldn't generate internal images. None of these people knew there was anything "wrong" with them. They just assumed that everyone who talked about having an imagination was being metaphorical, just using a really florid poetic way of describing that they remembered what something looked like.

In the second, a user on Quora described their experience with anosmia – not having a sense of smell. They didn't realize there was anything wrong until college. Until then, "I teased my sister about her stinky feet. I held my nose when I ate Brussels sprouts. In gardens, I bent down and took a whiff of the roses." Though they didn't say so explicitly, it sounds like they thought smell was just a metaphorical way of saying something was disgusting or delightful.

And in the third – well, this is awkward – I went years without realizing I didn’t have any emotions. I was getting treated for obsessive-compulsive disorder with high dose SSRIs. When these work well they dull your depression and anxiety; when they work less well, they dull all your emotions. For me they worked less well, but I never realized it until I came off them after five years and was suddenly overwhelmed by emotions I’d almost forgotten it was possible to have. In the interim, I’d understood that getting a birthday present was a positive and desirable event, and said it made me “happy”, without realizing something was missing. This was particularly inexcusable since I’d felt the full range of emotions before I started the drugs, but I guess the hypothesis “I have stopped feeling emotions” is a hard one to consider and collect evidence for.

So if someone says I’m incapable of genuine human relationships – well, I should stress that *I* think my relationships are genuine. But if they weren’t, maybe I wouldn’t notice. There would be something I was capable of, I would call that “genuine human relationships” since it was my only example of the concept, and I would never have anything else to compare it to.

II.

This post isn’t about relationships. This post is about ideas.

In high school I took a sociology class, and the teacher talked about how modern society was atomized and there were no real community bonds and so on. And I thought this was dumb. I didn’t live in an atomized society! My family knew our next-door neighbors, and we’d even been over at their house once for dinner. There was a Community Center a few blocks away, and when I was a kid I would go there a couple of times a year for some kind of Neighborhood Art

Night. Sometimes my mother volunteered at my school, and my dad was too busy to volunteer but probably would have if he could. We weren't devoid of community *at all*.

And then three things happened. Number one, I read some good anthropology about primitive and medieval societies, which actually described pre-atomized life and the way that there was barely even an individual identity and the community determined everything you ever did. Number two, I spent a little time in an honest-to-goodness Third World village and saw a little of what life was like there. And number three, I got involved in some good subcultures – including Bay Area rationality – which were slightly but noticeably less atomized than the neighborhood where I grew up. I realized that I'd mistaken the existent-but-weak forms of community in my suburban neighborhood for the really-strong forms of community that people complaining about atomization say we're missing, because I had so little experience with the latter I couldn't even imagine them.

This is a similar error as the SSRI/emotions problem. People talk about emotions/community. I have something sort of similar occupying that space. So I reasonably assume it's the same thing everyone is talking about.

I think I've figured out the whole "atomization" thing. But I'm not *sure*. What if there's some *real* non-atomized community that even second-hand anthropology plus some good subcultures can't point to? Am I just making the same mistake as I did as a high schooler, only one level higher?

Some of these same sociologists worry about advertising and consumerism. They think capitalism turns people into perfect consumers who overwork themselves at jobs they don't like to buy products they don't need. They think people’s entire identities revolve around brands and consumption.

And once again, I think: “Good thing this isn’t happening to me.” I don’t really watch TV and I tune out online ads. I buy things occasionally, usually things that I need or things that I occasionally enjoy. But I don’t own much “clutter”. And I don’t care about brands, except ones that really signal high quality.

Is this the same kind of mistake as “I met the neighbors once, so I’m not atomized”? I don’t know!

Either understanding “consumerism” was so easy for me that I got it immediately and effortlessly, and I live a charmed life that has prevented me from ever encountering that problem.

Or I have only a superficial fascimile of understanding it, and when I actually understand it, it’ll seem profound and important, the same way “atomization” did.

When I see other people making a big deal out of seemingly-minor problems, I’m in this weird superposition between thinking I’ve avoided them so easily I missed their existence, or fallen into them so thoroughly I’m like the fish who can’t see water.

And when I see other people struggling to understand seemingly-obvious concepts, I’m in this weird superposition between thinking I’m so far beyond them that I did it effortlessly, or so far beneath them that I haven’t even realized there’s a problem.

III.

Last week, some people proposed it was useless to steelman/understand post-modernism. It was just people being stupid or having garbled thinking. Maybe. There are some post-modernists who even the other post-modernists say are probably just pulling it out of their asses.

But how would we know? There are concepts nobody gets on the first reading, concepts you have to have explained to you again and again until finally one of the explanations clicks and you can reconstruct it out of loose pieces in your own head.

And there are concept-shaped holes you don't notice that you have. You can talk to an anosmic person about smell for years on end, and they're still not going to realize they've got a big hole where that concept should be. You can give high-school me an entire class about atomization, and he can ace the relevant test, and he's still not going to know what atomization is.

Put these together, and you have cause for concern. If you learn about something, and it seems trivial and boring, but lots of other people think it's interesting and important – well, it could be so far beneath you that you'd internalized all its lessons already. Or it could be so far beyond you that you're not even thinking on the same level as the people who talk about it.

I'm looking back on my book review of After Virtue, a seminal philosophy book which won a bunch of awards and recognition from important philosophers. My review was that it seemed very confused. It kept claiming to have an important insight, but every time it said it was going to reveal the important insight, it actually said a bunch of platitudes and unrelated tangents. This is a huge red flag. Which makes more sense – that I was the lone genius able to see that the emperor had no clothes and Alasdair MacIntyre is really dumb? Or that he’s saying something really hard to understand, and I haven’t understood it yet?

Maybe there are fields doing the intellectual equivalent of gaslighting, insisting they have really profound points when they’re just vapor. But err on the side of caution here. Most of us have some hard-won battles, like mine understanding atomization. Where after a lot of intellectual work, a concept that seemed stupid suddenly opens up and becomes important. Sometimes it’s about anarchism, or reactionary philosophy, or privilege, or religion as benevolent community-building institution. Erring too hard on the side of “that’s dumb, they’re probably just gaslighting” closes off those areas to you forever.

I don’t think it’s always worth delving deep into a seemingly-meaningless field to discover the hidden meaning. That rarely works – if you had the concepts you’d need to understand it right now, you would have done so already. But I think it’s worth leaving the possibility open, so that later if something clicks you’re not too embarrassed to return to it.

Does Age Bring Wisdom?

[Related: Can It Be Wrong To Crystallize Patterns?]

I.

I turn 33 today. I can only hope that age brings wisdom.

We've been talking recently about the high-level frames and heuristics that organize other concepts. They're hard to transmit, and you have to rediscover them on your own, sometimes with the help of lots of different explanations and viewpoints (or one very good one). They're not obviously apparent when you're missing them; if you're not ready for them, they just sound like platitudes and boring things you've already internalized.

Wisdom seems like the accumulation of those, or changes in higher-level heuristics you get once you've had enough of those. I look back on myself now vs. ten years ago and notice I've become more cynical, more mellow, and more prone to believing things are complicated. For example:

1. Less excitement about radical utopian plans to fix everything in society at once
2. Less belief that I'm special and can change the world
3. Less trust in any specific system, more resignation to the idea that anything useful requires a grab bag of intuitions, heuristics, and almost-unteachable skills.
4. More willingness to assume that other people are competent in aggregate in certain ways, eg that academic fields aren't making incredibly stupid mistakes or pointlessly circlejerking in ways I can easily detect.
5. More willingness to believe that power (as in "power structures" or "speak truth to power") matters and infects everything.

6. More belief in Chesterton’s Fence.
7. More concern that I’m wrong about everything, even the things I’m right about, on the grounds that I’m missing important other paradigms that think about things completely differently.
8. Less hope that everyone would just get along if they understood each other a little better.
9. Less hope that anybody cares about truth (even though ten years ago I would have admitted that nobody cares about truth).

All these seem like convincing insights. But most of them are in the direction of elite opinion. There’s an innocent explanation for this: intellectual elites are pretty wise, so as I grow wiser I converge to their position. But the non-innocent explanation is that I’m not getting wiser, I’m just getting *better socialized*. Maybe in medieval Europe, the older I grew, the more I would realize that the Pope was right about everything.

I’m pretty embarassed by Parable On Obsolete Ideologies, which I wrote eight years ago. It’s not just that it’s badly written, or that it uses an ill-advised Nazi analogy. It’s that it’s an impassioned plea to jettison everything about religion immediately, because institutions don’t matter and only raw truth-seeking is important. If I imagine myself entering that debate today, I’d be more likely to take the opposite side. But when I read Parable, there’s…nothing really wrong with it. It’s a good argument for what it argues for. I don’t have much to say against it. Ask me what changed my mind, and I’ll shrug, tell you that I guess my priorities shifted. But I can’t help noticing that eight years ago, New Atheism was really popular, and now it’s really unpopular. Or that eight years ago I was in a place where having Richard Dawkins style hyperrationalism was a useful brand, and now I’m (for some reason) in a place where having James C. Scott style intellectual conservativism is a useful brand. A lot of the “wisdom” I’ve “gained” with age is the kind of wisdom that helps me channel James C. Scott instead of Richard Dawkins; how sure am I that this is the right path?

Sometimes I can almost feel this happening. First I believe something is true, and say so. Then I realize it’s considered low-status and cringeworthy. Then I make a principled decision to avoid saying it – or say it only in a very careful way – in order to protect my reputation and ability to participate in society. Then when other people say it, I start looking down on them for being bad at public relations. Then I start looking down on them just for being low-status or cringeworthy. Finally the idea of “low-status” and “bad and wrong” have merged so fully in my mind that the idea seems terrible and ridiculous to me, and I only remember it’s true if I force myself to explicitly consider the question. And even then, it’s in a condescending way, where I feel like the people who say it’s true deserve low status for not being smart enough to remember not to say it. This is endemic, and I try to quash it when I notice it, but I don’t know how many times it’s slipped my notice all the way to the point where I can no longer remember the truth of the original statement.

And what about number 9 on the list? Believing nobody cares about truth is cynicism, which seems *sort of* like wisdom. But traumatize someone enough and they’ll reliably pick up some new cognitive styles; it’s much easier to give someone hypervigilance than it is to cure them. Imagine someone reading enough newspapers that they hear all of the worst and scariest things, and maybe start thinking that the country is 50% Nazis and 50% violent antifa. Is the resulting pessimism and paranoia really wisdom? Or is it just a more stable, more thermodynamically-preferred state than innocence?

And if I accept my intellectual changes as “gaining wisdom”, shouldn’t I also believe that old people are wiser than I am? And old people mostly seem to go around being really conservative and saying that everything was better in the old days and the youth are corrupt and Facebook is going to be the death of us. I could model this as two different processes – a real wisdom-related process that ends exactly where I am now, plus a false rose-colored-glasses-related process that ends with your crotchety great-uncle talking about how things have been going downhill since the war – but that’s a lot of special pleading. I remember when I was twenty, I thought the only reason adults were less utopian than I was, was because of their hidebound rose-colored self-serving biases. Pretty big coincidence that I was wrong then, but I’m right about everyone older than me *now*.

There’s one more possibility that bothers me even worse than the socialization or traumatization theory. I’m going to use science-y sounding terms just as an example, but I don’t actually think it’s this in particular – we know that the genes for liberal-conservative differences are mostly NMDA receptors in the brain. And we know that NMDA receptor function changes with aging. It would be pretty awkward if everything we thought was “gaining wisdom with age” was just “brain receptors consistently functioning differently with age”. If we were to find that were true – and furthermore, that the young version was intact and the older version was just the result of some kind of decay or oxidation or something – could I trust those results? Intuitively, going back to earlier habits of mind would feel inherently regressive, like going back to drawing on the wall with crayons. But I don’t have any *proof*.

Wisdom is like that.

Ars Longa, Vita Brevis

The Alchemist asked if I wanted a drink. I did, but no amount of staring could make my eyes settle on the color of the liquid in the flask. And the gold the alchemists paid the taxmen smelled funny and made crackling noises. I declined.

I took the summons and set it on the table between us. The King’s son was dying. The doctors, astrologers, witches, and other assorted wise people of the kingdom could not save him. The King had asked for an alchemist, and been given one. He, too, had failed. But he had let on that there were other alchemists in the guild, greater alchemists, who knew far more than he. So the king had demanded that all the guild’s top alchemists come to the palace and try to save his son’s life. And the alchemists’ guild had refused, saying their studies could not be interrupted.

So here I was, come to make the request again, more formally but less politely.

The Alchemist pretended to read the parchment. I could tell he was faking; his eyes stayed still the whole time. Finally he gave me the same answer he had given the king’s courier: the alchemists’ studies could not be interrupted.

“Why is a few weeks subtracted from your studies more important than the prince’s life?” I demanded, staring straight into his creepy too-still eyes.

He spent too long not answering. I worried I’d broken him, that he was some kind of intricate clockwork machine and I’d yelled too loud and shifted a gear out of place. Finally he asked: “How long would you have to study architecture before you could build a castle like this one?”

“I’m no architect,” I said. “I’m a man of war.”

“Yes. So how long would you have to study, before you were an architect?”

“Ten years?” I asked. “Twenty?”

“Why so? There are books of architecture, some of them written by men far greater than the planner of this castle. Some are five hundred pages long, others a thousand. Are you so slow a reader, that it would take you ten years to read a thousand pages?”

“You can’t just read a book and know architecture.”

“But why not?”

“Because…you wouldn’t…” I had been annoyed when he first asked, but now I found the question interesting, at least amusing. Why *couldn’t* a great architect write his knowledge down in a book? And why *couldn’t* I read it and become as good as he?

“Because you’d have to memorize it all,” I finally concluded.

“Not so. I will let you carry the book with you as you build the castle.”

“It wouldn’t help. It wouldn’t be…indexed properly in my head. I would want to build a wall, and I wouldn’t even know what things to consider when building a wall, and I would have to search the whole book for them each time.”

“You are a man of war,” repeated the Alchemist. “Do you know Caesar’s histories?”

“Almost by heart.”

“Are you as good a general as Caesar?”

“No.”

“Why not?”

I took his point. Caesar had written down everything he could about war. I had mastered all of it. But I was no Caesar. It couldn’t just be the difficulty of memorizing books.

“Knowledge,” said the Alchemist, “is harder to transmit than anyone appreciates. One can write down the structure of a certain arch, or the tactical considerations behind a certain strategy. But above those are higher skills, skills we cannot name or appreciate. Caesar could glance at a battlefield and know precisely which lines were reliable and which were about to break. Vitruvius could see a great basilica in his mind’s eye, every wall and column snapping into place. We call this wisdom. It is not unteachable, but neither can it be taught. Do you understand?”

I did. If I trained with Caesar for years, some of his skill at reading a battlefield might rub off on me; I might dimly see the outlines of his genius. But he couldn’t just tell me. It wasn’t a secret which he hid from other men to remain above them. It was a power belonging to him alone, only partially transferable.

“So imagine,” continued the Alchemist, “that you wanted to build the simplest of structures. A cottage for peasants. How long would you have to study architecture under Vitruvius before you could do it?”

This time I didn’t bother protesting that I didn’t know. I just guessed. “A year?”

“And suppose you want to build something more complex. An aqueduct, every bit the equal of the Romans’. How long?”

“Five years?”

“Some grand building, a palace or temple?”

“…ten years?”

“The grandest building in the world. St. Peter’s Basilica, or the Pantheon, or Chartres Cathedral, or something new that combines the virtues of all three.”

“How should I know? Twenty years? Thirty?”

“Would you believe me if I said it was two hundred years?”

“No. The human lifespan is three score and ten. If you needed more than seventy years of studying architecture to design St. Peter’s, it would never have gotten designed.”

“Then,” said the Alchemist, “we have discovered something surprising. The art of architecture is limited by the human lifespan. The greatest building that can ever be designed is the one that would take seventy years of studying architecture to master; God has drawn a line in the sand forever closing off buildings grander than these.”

I thought for a second. “That doesn’t seem right. There are new innovations every year. The flying buttress, stained glass, the pointed arch. The Romans had none of these. We progress not only by studying the works of Vitruvius, but by pushing beyond him. Perhaps it takes a century for someone to invent the buttress, but once it is invented, only weeks for other architects to observe it and understand it well enough to incorporate into their own buildings. Architecture does not advance only architect by architect, but also civilization by civilization."

"Are you skilled at mathematics?" asked the Alchemist.

I shook my head.

"Then we will talk this over, though rightfully it should be an equation. The first term is the speed at which a student can absorb already-discovered architectural knowledge. The second term is the speed at which a master can discover new knowledge. The third term represents the degree to which one must already be on the frontier of knowledge to make new discoveries; at zero, everyone discovers equally regardless of what they already know; at one, one must have mastered every previously-discovered fact before one can discover anything new. The fourth term represents potential for specialization; at one, it is impossible to understand any part without understanding the whole; at zero, it can be subdivided freely. The fifth…"

"I don't think saying it in words makes the math easier to understand."

"Ah. Well, imagine a science that takes one-tenth as long for a student to understand, as it did a master to discover. And imagine that one cannot advance the science until one understands everything that has already been discovered. And one cannot split the burden; tell one architect 'Oh, you learn how to make walls, I will learn how to make roofs' – a single genius must understand the whole building, every part must fit together perfectly. We can calculate how far the art can advance."

“How?”

“The first student has no master, and must discover everything himself. He researches for 70 years, then writes his wisdom into a book before he dies. The second student reads the book, and in 7 years, he has learned 70 years of research. Then he does his own original research for 63 years and writes a book containing 133 years of research. The third student reads for 13.3 years, then does his own research for 66.7 years, ending up with 200 years. Imagine going further and further. After many generations, 690 years of research have been done, and it takes a student 69 years to master them. The student only has one year left of life to research further, leaving the world with 691 years of research total. So the cycle creeps onward, always approaching but never quite reaching 700 years of architectural research.”

“It doesn’t work that way,” I protested, partly because it didn’t, and partly because something about the story distressed me more than I could say.

“Not in architecture. An architect who has not yet mastered the entire field can still make discoveries. And the field can be split – I can work on walls while you work on windows. It would only work that way if there were an Art so unified, so perfect, that a seeker had to know the totality of what had been discovered before, if he wanted to know anything at all.”

“Then you really could never advance past 700 years of knowledge.”

“You would have to be clever. We imagine each master writing down his knowledge in a book for the student who comes after, and each student reading it at a rate of ten times as quickly as the master discovered it. But what if there was a third person in between, an editor, who reads the book not to learn the contents, but to learn how to rewrite it better and more clearly? Someone whose job it is to figure out perfect analogies, clever shortcuts, new ways of graphing and diagramming the information involved. After he has processed the master's notes, he redacts them into a textbook which can teach in only a twentieth the time it took the master to discover."

"Then we could double the amount of research that could eventually be completed, to 1400 years' worth."

"Not easily. Remember, the editors face the same problem as the students: they can only redact knowledge they themselves understand. We are adding many new people, and many generations of work, to the problem. But in the end, yes, you could accumulate 1400 years of knowledge. What if you wanted more?"

"More?"

"I'm afraid so."

"Hm. You…could get more layers of redactors. Redactors of redactors, to make the textbooks truly perfect."

"Perhaps what you are trying to say is that redaction is an Art."

The Alchemist made the the capital letter unmistakeable.

"Every Art has its own structure. Architecture, with enough study, can allow you to accumulate seven hundred years of collected knowledge. How many years could redactors and tutors accumulate? Would some first redactor have to spend seventy years coming up with principles of redaction to pass down to his student, who advances the art by sixty-three more years, which he passes down in turn? Would a 1400-year redactor be an incomprehensible master, able to build whole basilicas of redaction, a master teacher who could frame any concepts to make it intuitive and memorable?"

"I changed my mind. I'm going to have that drink."

The Alchemist poured me the liquid of indeterminate color. I took a sip. It reminded me of nothing I had ever tasted before, but very slightly of the letter "N". More important, I was pretty sure it was alcoholic.

"You're talking about an infinite regress", I said, when I had finished the glass.

"Not infinite. Architects. Teachers. Teachers of teachers, but the art of teaching teaching is much the same as the art of teaching. Three levels is enough. Though the levels have to mix. The teacher who trains the next architect must be a master both of teaching and of architecture. I will spare you the math, but one needs a series of teachers at different points on the teaching-skill/architecture-skill tradeoff-curve. One will be a master teacher who has devoted decades to learning the textbook-writing skill, and who can write a brilliant Introduction To Architecture textbook that makes the first ten years of architecture ability seem perfectly natural and easy to master. Another will be a mediocre teacher who knows enough advanced architecture to write a passable textbook on the subject. Still another will do nothing but study pure Teaching itself, in the hopes that he can one day pass on this knowledge to others who will use it to write architecture textbooks. In practice we are limited to a few strategic points on the tradeoff curve."

"In practice?"

He motioned for me to get up. We walked through dark corridors until we reached a courtyard, bathed in the glow of the full moon. It took me a second to see it. Then the dull shapes took form. Obelisks, covered in hieroglyphs. A garden of obelisks.

“The word ‘alchemy’ comes from ‘al-Kemi’, the Arabic word from Egypt. It was the ancient Egyptians who first considered the project. They didn’t want the Philosophers’ Stone, not at first. They just wanted normal philosophers. But philosophy, more than other subjects, requires the wisdom that comes with age. More than other subjects, a philosophy book cannot merely be read; it must be digested, intermingled with life experience, wrestled with. The Egyptians scholars ran into precisely the problem as our hypothetical architects – there were secrets that evaded the human lifespan.

“So they wondered whether a way to cheat death might be found. The answer was both exciting and discouraging. Through the mysteries of spiritual chemistry, an elixir might be created which would grant immortality. But the Work itself would take far more knowledge than any one man could accumulate. The symbol of alchemists is the ouroboros because our task loops back upon itself. In order to become immortal, you must first become immortal.

“All we could do was go the slow way, the same as the architects working on their great basilica, for generation after generation. So Egypt fell, but we did not fall. Rome passed away, but we did not pass. A few lines, the remnants of the old priestly families of Hierakonopolis and Memphis, continued the work. To stop would be to reset a process requiring four thousand years of gradual asymptotic improvement all the way to the beginning – texts are not worthless, but only the true tutors trained by tutor-tutors trained by tutor-tutor-tutors are fit to tutor an alchemist. A misstep is too terrible to contemplate. But any victory – a single vial of the Elixir, a single fragment of the stone – would end the nightmare forever. We would have an immortal, a philosopher whose lifespan finally matches the depth of the challenges Nature throws at us.

“That is our guild’s mission. A few of us, those who pass all their tests, do the alchemic research that moves the Work onwards. Others train to be teachers, or teachers-of-teachers. Those who fail a test somewhere along the way stay in the guild, managing its worldly affairs. Some scour the countryside for prodigies to take in and train as apprentices. Others manage our finances. And the very least capable, like me, have time to waste talking to outsiders, trying to convince them of our mission. A few centuries more, and we will have the Stone. Does that satisfy your curiosity?”

“All except my original question. Are you so busy that you cannot spare a few weeks for the prince?”

“God does not make the Great Work easy. We have done all we can to train our alchemists, our tutors, our tutor-tutors, and so on, yet in the end, the limit of human skill is the same place the possibility of success begins. It is His will to grind us up to the very asymptote.”

“I still don’t get it.”

“Do you remember the architects who learned at ten times the rate they researched, the ones who would never accumulate more than 700 years of learning? The fiftieth alchemist in the sequence has 696 years of learning, and is able to do a scant five months’ original research before his death. The hundredth alchemist has 699.98 years of learning, and is able to do about a day's research before dying. We are not so far along as all that, but we are far. We do not have the Stone, but we have tinctures that can stabilize the lifespan, make sure nobody dies before their time. The last few generations – on their deathbed, they say they can almost *taste* the Stone, that it lies only a few hours of further thought beyond their level. They say of my grandfather that he realized the recipe for the Stone on his deathbed, that he started speaking it, but that his eyes closed forever before he could complete the ingredient list."

"So?"

"You ask that we pause a few weeks from our studies to save the prince's life. Pausing a few weeks would set us back generations. This far into the project, only the last few hours of an Alchemist's life are of any value at all. We cannot spare the prince hours. We cannot even spare him seconds."

"Then your teachers…or your teacher-teachers?"

"Know some alchemy, but are in the same situation. Our textbooks have been so perfectly written and rewritten over the years that it is only in the last few days of a teacher's life that he is skilled enough to write a better one. And our teacher-training has become so perfect that it is only in the last few days of a teacher-trainer's life that he is qualified to create teachers better than the ones who already exist."

"There's no slack in the system at all?"

"Only me, and those like me. Those judged unfit for research and condemned to worldly matters. We sent you one already. He failed you, as he did us. We have nothing more to give."

“The king will not be happy. And the Prince will die.”

“Everyone dies,” said the Alchemist. “If the prince does not die this year, he will die the next, or fifty years hence. The question is not when we die, but what our life adds to the Work which accumulates in spite of time. Quicksilver evaporates to nothing unless reacted with aqua fortis; but the part which is reacted endures forever. Those lives not part of any Work mean as little to me as they will one day mean to their possessors; those which add to the Work are more precious than gold. Tell the King this.”

“He won’t understand,” I said.

“Then you will have to teach him,” said the Alchemist, “as I taught you, and my tutors taught me, and as their tutors taught them, all the way back to the first philosophers of Egypt.”

He stared at me as he spoke, and the blackness in his too-still pupils was the depth of Time.

Book Review: Legal Systems Very Different From Ours

I.

Medieval Icelandic crime victims would sell the right to pursue a perpetrator to the highest bidder. 18th century English justice replaced fines with criminals bribing prosecutors to drop cases. Somali judges compete on the free market; those who give bad verdicts get a reputation that drives away future customers.

“Anarcho-capitalism” evokes a dystopian cyberpunk future. But maybe that’s wrong. Maybe we’ve always been anarcho-capitalist. Maybe a state-run legal system isn’t a fact of nature, but a historical oddity as contingent as collectivized farming or nationalized railroads. *Legal Systems Very Different From Ours*, by anarcho-capitalist/legal scholar/medieval history buff David Friedman, successfully combines the author’s three special interests into a whirlwind tour of exotic law.

Law is a public good. Crime victims have little economic incentive to punish the perpetrator; if you burn my house down, jailing you won’t un-burn the house. If you steal my gold, I have some interest in catching you and taking it back, but no more than I do in catching some other poor shmuck and taking *his* gold. It’s only society as a whole that wants to make sure criminals are reliably punished and the innocent consistently safe. This is the classic situation where economists usually recommend government intervention.

But sometimes that doesn’t work. Maybe you live in an area like Somalia or medieval Ireland without a strong centralized government. Maybe you live in a strato-klepto-kakocracy run by warlords who can’t even pronounce “jurisprudence”, let alone enforce it. Maybe you’re a despised minority group whom the State wants nothing to do with, or who wants nothing to do with the State.

Gypsies living scattered in foreign countries have generally wanted to run their own communities by their own rules. Nothing stops some of them from calling themselves a “legislature” or a “court” and claiming to make laws or pass sentences. But something does stop them from trying to enforce them: from the State’s point of view, a “court” that executes an offender is just a bunch of Gypsies who got together and committed murder. So the Vlach Rom – Romanian Gypsies – organize courts called *kris* which enforce their sentences with threat of banishment from the community.

Gypsies traditionally believe in *marime*, a sort of awful pollution that infects people who don’t follow the right rituals; anyone who interacts with polluted people will become polluted themselves. *Kris* courts can declare the worst offenders polluted, ensuring their speedy ostracization from Gypsy society. And since non-Gypsies are polluted by default, the possibility of ostracism and forced integration into non-Gypsy society will seem intolerable:

> The effectiveness of that threat [of ostracism] depends on how easily the exiled gypsy can function outside of his community. The marimé rules (and similar rules in other societies) provide a mechanism for isolating the members of the community. Gaije, non-gypsies, do not know the marimé rules and so do not and cannot obey them. It follows that they are all polluted, unclean, carriers of a contagious disease, people whom no Rom in his right mind would willingly choose to associate

> with; when and if such association is unavoidable it must be taken with great care. The gypsy view of gaije, reinforced by the gaije view of gypsies as uneducated and illiterate thieves and swindlers, eliminates the exit option and so empowers the kris to enforce gypsy law by the threat of exclusion from the only tolerable human society.

This reminds me of The Use And Abuse Of Witchdoctors For Life: once your culture has a weird superstition, it can get plugged into various social needs to become a load-bearing part of the community structure.

Amish also live under the authority of a foreign culture and have settled on a similar system, with a twist. The basic unit of Amish society is the church congregation; Amish settlements big enough to support multiple churches will have many congregations mixed together. Each congregation will have its own rules, especially about which technologies their members are or aren't allowed to use. Amish people who violate their congregation's rules, either by using forbidden technology or by the usual litany of sins, are punished with public confession or temporary ostracism. Amish people who refuse to abide by lesser punishments are excommunicated, though they can be un-excommunicated if they change their minds and agree to follow the court's orders.

Amish congregations are nominally democratic, but in practice Friedman calls them dictatorship-like because everyone votes the way the bishop wants. But they are a "competitive dictatorship"; since there are so many different congregations in the same town, an Amish family who doesn't like their congregation's leadership or legal system can move to another congregation and agree to be bound by their laws instead. This makes it a rare remaining example of a polycentric legal system outside anarcho-capitalist fantasies or *Too Like The Lightning*:

> Such a system can be viewed as a competitive market for legal rules, constrained, like other competitive markets, to produce about the product that the customers want. Competitive dictatorship is the mechanism we routinely use to control hotels and restaurants; the customers have no vote on what color the walls are painted or what is on the menu, but an absolute vote on which one they patronize.

They do encounter the same problem as the Gypsies: can you just commit a crime, then accept your ostracism and integrate with another society somewhere else? The Amish have some internal mechanisms to prevent this: congregations are usually on good terms with each other, but if Congregation A accepts a member being shunned by Congregation B, then all of Congregation B's members will shun all of Congregation A's members. In practice, this makes it easy to switch rules as a member in good standing who honestly doesn't like the laws, but hard to break the laws and get away with it.

Of course, you can still leave the Amish community and go join broader American society. But have you *seen* broader American society?

18th century England had a government, a court system, and some minimal law enforcement – but it really sucked. There were no public prosecutors; anyone who felt like it could bring a criminal to court and start prosecuting him, but if nobody felt like it then the crime remained unpunished. Prosecuting took a lot of time and money and was generally a thankless task. And the government didn't want to go to the expense of imprisoning people, so they usually just hanged convicted offenders (if the crime seemed really bad) or pardoned them (if it didn’t seem to merit hanging). The exotic anarcho-capitalist part comes in as English civil society creates its own structures to work around these limitations.

Merchants, landowners, and other people with wealth banded together in mutual-protection-insurance-groups. Everyone in the group would pay a fixed amount yearly, and if one of them got robbed the group would use the money to hire a prosecutor to try the criminal. Group members would publish their names in the newspaper to help inform thieves whom it was a bad idea to rob. But this wasn’t about leaving poor people out to dry. The groups would also help indigents who couldn’t afford their own prosecutors, partly out of a desire to crack down on crime before it reached the point where it could inconvenience them. They wouldn’t help people who could have afforded insurance but declined anyway, though – otherwise there would be no incentive to buy in.

(if this sounds familiar, it’s from another, very different David Friedman book)

What about the lack of good punishments? Once a trial was underway, prosecutors would usually cut a deal: the offender would bribe the prosecutor with a certain amount, and the prosecutor would drop the case. The size of the bribe would vary based on how much the offender could pay, the extent of their crime, and the facts of the case (and therefore the likelihood of the magistrate choosing hanging vs. pardon). This not only helped tailor the punishment more precisely to the crime, but helped defer the cost of prosecution: victims (or their mutual-protection-insurance-groups) were incentivized to press charges because they could recoup their costs through the bribes paid to drop them:

> What both modern and contemporary commentators seem to have missed is that, however corrupt such arrangements might be from a legal standpoint, they helped solve the fundamental problem of private prosecution. The possibility of compounding provided an incentive to prosecute-it converted the system into something more like a civil system, where a victim sues in the hope of collecting money damages. And while compounding might save the criminal from the noose, he did not get off scott free. He ended up paying, to the prosecutor, what was in effect a fine.

10th through 13th century Iceland was in the same position as the Vlach Rom: a legislature (the Althing), some courts, but no executive branch. Unlike the Rom, the Icelanders' problem wasn't foreign oppressors – it was that they were the Viking equivalent of those hard-core libertarians who live in compounds in Montana where the Feds can't reach them. In this case "the Feds" were the forces of King Harald Fairhair, who had just taken over and centralized power in Norway. Some Norwegians decided they would rather live on a remote and frequently-exploding piece of rock on the edge of the world than be anyone's subject: thus, medieval Iceland.

If an Icelander thought a crime had happened, they would go to court and plead the case themselves. If the court pronounced a guilty verdict, it would demand a penalty from the criminal. Usually this was a fine paid to the victim; even murders were punished with wergeld. If the criminal paid the fine voluntarily, all was well. If they refused – or didn't even come to court – then the court could declare the criminal an outlaw, meaning it was legal to kill him and take his stuff. And:

> One obvious objection to a system of private enforcement is that the poor (or weak) would be defenseless. The Icelandic system dealt with this problem by giving the victim a property right – the right to be reimbursed by the criminal – and making that right transferable. The victim could turn over his case to someone else, either gratis or in return for a consideration. A man who did not have sufficient resources to prosecute a case or enforce a verdict could sell it to another who did and who expected to make a profit in both money and reputation by winning the case and collecting the fine. This meant that an attack on even the poorest victim could lead to eventual punishment.
>
> A second objection is that the rich (or powerful) could commit crimes with impunity, since nobody would be able to enforce judgment against them. Where power is sufficiently concentrated this might be true; this was one of the problems which led to the eventual breakdown of the Icelandic legal system in the thirteenth century. But so long as power was reasonably dispersed, as it seems to have been for the first two centuries after the system was established, this was a less serious problem. A man who refused to pay his fines was outlawed and would probably not be supported by as many of his friends as the plaintiff seeking to enforce judgment, since in case of violent conflict his defenders would find themselves legally in the wrong. If the lawbreaker defended himself by force, every injury inflicted on the partisans of the other side would result in another suit, and every refusal to pay

> another fine would pull more people into the coalition against him.
>
> There is a scene in Njal’s Saga that provides striking evidence of the stability of this system. Conflict between two groups has become so intense that open fighting threatens to break out in the middle of the court. A leader of one faction asks a benevolent neutral what he will do for them in case of a fight. He replies that if they are losing he will help them, and if they are winning he will break up the fight before they kill more men than they can afford! Even when the system seems so near to breaking down, it is still assumed that every enemy killed must eventually be paid for. The reason is obvious enough; each man killed will have friends and relations who are still neutral–and will remain neutral if and only if the killing is made up for by an appropriate wergeld.

I think this is asking: are we sure you can’t end up with outlaw cascades, where everyone just agrees to be outlaws together? Suppose Warren Buffett cuts off my arm. The court asks him to pay a fine, and he refuses, so the court declares him an outlaw and legally killable. I gather some of my friends to form a posse to kill him, but he hires a hundred bodyguards to resist me. There’s a fight, the bodyguards kill my friends, and the court fines the bodyguards. They don’t pay, so the court declares the bodyguards outlaws. I gather a thousand people to kill Buffett and/or his hundred bodyguards, and Buffett and his bodyguards pool their money to hire a whole force of mercenaries to resist us. The mercenaries kill lots of us, the court fines them, and the mercenaries don’t pay. Now the court declares the mercenaries outlaws. But it seems like at some point maybe more than half the population of Iceland will be outlaws, and then maybe they just have to declare a new legal system or something.

An Icelander might retort: why doesn’t that happen in modern America? A policeman catches you dealing drugs, so you offer the policeman $10,000 to let it pass. The policeman refuses because it’s illegal and he would get in trouble. Well, you say, what’s the worst thing that could happen if you got in trouble? The police would come after you? But police would hesitate to arrest a fellow officer, plus we’ve already established that they can be deflected with bribes. Sure, there’s a stable equilibrium where you arrest me right now. But there’s *also* a stable equilibrium where 51%+ of the nation’s police join our sordid bribery chain, accumulate more power than the law-abiding police, and end up as some weird mercenary army that takes over the country and rewrites the law to their own advantage.

This is a good place to remember that David Friedman is also the author of A Positive Account Of Property Rights, maybe the single most mind-opening essay I’ve ever read. No summary can do it justice, but the basic outline is that governmental “legitimacy” is the government’s position as a conspicuous Schelling point for everybody who wants to avoid civil war/the state of nature/a worse government. Once it’s common knowledge that a government is legitimate, everyone expects everyone else to enforce its rules, and so they’ll enforce its rules in turn until it becomes common knowledge that the government isn’t legitimate anymore. This works just as well in medieval Icelandic anarcho-capitalism as it does in modern America. Just because our government dresses all of its enforcers-of-state-sanctioned violence in snazzy uniforms and makes them work out of the same building doesn’t make the whole system any less of a mass hallucination.

II.

This book works well alongside James Scott’s *Seeing Like A State* and the whole discourse around cultural evolution.

In *Seeing Like A State*, ordinary people living their daily lives blunder into highly advanced systems for doing whatever it is they do. Primitive farmers will know every tiny detail about exactly when to plant which crops, and how to exploit microvariations in soil quality, and know ridiculous tricks like planting fish heads in the ground as fertilizer. Ordinary city-dwellers will organically build houses and stores and streets in exactly the right fractal patterns to maximize some measure of quality of life. Scott dubs this “metis”, an evolved intuitive sense of practical wisdom that often outperforms seemingly more scientific solutions.

Many of the societies Friedman profiles in *Legal Systems Very Different From Ours* seem to operate on *metis*. Most don’t know who developed their legal system; in a few of them, it is explicitly declared to have been the work of God. Most don’t really know why their legal system works – in some cases, Friedman only gives an economic analysis of why some rule might exist after admitting that previous scholarship (both modern academic, and within the society in question) has failed to come up with answers. And a lot of them are too brilliant, and need too many weird interlocking parts, to be the work of any single person.

“Cultural evolution” is the idea that cultures evolve in a way analogous to biological organisms. The definition gets kind of fuzzy – if I come up with a good idea and my culture adopts it, is that the result of “cultural evolution” or ordinary human ingenuity? `But a lot of people find the concept to have some value – and if it has any at all, *Legal Systems Very Different From Ours* has to include some of the best examples.

Friedman frames this in economic terms. Social "entrepreneurs" come up with some new system that solves a need, and it catches on by raising the utility of everyone involved. The mutual-protection-insurance-groups of 18th century England work this way: somebody invents them and offers the opportunity for other people to sign on, everyone who does ends up better off than the people who doesn't, and they eventually reach fixation. Same with the criminal-prosecutor bribes; someone thinks it up, it leaves both sides better off, so everybody who hears about it does it. Viewed very optimistically, wherever there's a problem in your culture, institutions to solve the problem will magically appear and spread until everybody does them.

Conflict is an especially fertile ground for cultural innovation. Friedman stresses how many legal systems, including advanced ones with lawyers and codes and everything, show signs of originating from feud systems, which might be the most basic form of law. They work like this: "If you offend me in some way, I will try to kill you". A slightly more advanced version that takes account of possibly power differentials between offender and victim: "If you offend me in some way, everybody in my family will try to kill everybody in your family". This originally sounds unpromising, but it turns out that people really don't want their family members murdered. So we end up with an even more advanced version: "If you offend me in some way, we had better find some way to arbitrate our dispute, *or else* everybody in my family will try to kill everybody in your family".

The Somali system seems to be somewhere around here: if two people have a dispute, they find a mutually agreeable judge to arbitrate; the judge will decide who's in the wrong and what fine they need to pay to make it right. If someone refuses to go to the judge, or refuses to abide by the judge's decision, then it’s family-member-killing time. Needless to say, Somali judges’ services remain popular. And since judges gain status by arbitrating, and since only judges who make widely-regarded-as-good decisions get invited to keep doing so, there’s economic pressure for the judges to make good decisions (which then go down as precedent and inspire future cases). It’s easy to see how something like this can turn into a perfectly respectable legal system where people totally forget that killing each other’s family members is even an option. Catch it at this last stage, and hear enough people admit they have no idea who “invented” their legal system, and it looks like it appeared by magic.

In fact, one of the most interesting things I got from this book is that all legal systems need a punishment of last resort – one that can be enforced whether or not the offender agrees with it – but these punishments practically never happen in real life. The Gypsies and Amish will ostracize members who defy the court – but since everyone lives in fear of ostracization, in real life they’ll just pay the fine or make their public confession or whatever. The English will hang criminals at the drop of a hat – but since the threat of hanging incentivizes them to bribe prosecutors, in reality few people will need to be hanged. The Icelandic courts could declare offenders outlaws who can be killed without repercussion – but the threat encourages Icelanders to pay the wergeld, and nobody has to get outlawed. The Somalis are ready to have murderous family feuds – but the possibility of such a feud keeps people willing to go to arbitration. Even our own legal system works like this. The police can physically drag you to jail, kicking and screaming. But more likely you’re going to plea bargain, or agree to community service, or at least be cooperative and polite while the police take you away. Plea bargains – which are easier for prosecutors, easier for defendants, and easier for taxpayers – seem like a good example of cultural evolution in action; once someone thought them up, there was no way they weren't going to take over everything despite their very serious costs.

III.

Three other things worth noting about *Legal Systems Very Different From Ours*.

First, something kept seeming *off* about all the legal systems mentioned, which only clicked into place about halfway through: they really, really didn't seem prepared for crime. A lot of them worked on a principle like: "If there's a crime, we'll call together a court made of all the town elders, plus at least three different religious leaders, plus the heads of the families of everybody involved, plus a representative of the Great King, plus nine different jurists from nine different universities, and all of them will meet on the Field Of Meeting, and a great tent will be erected, and…" The whole thing sounded like it might work as long as there was like one crime a year. Any more than that and none of the society's officials would ever have time for anything else.

As weird as it is to punish murder with a fine, the fines these societies levied for murder sounded really high: the Islamic price was a hundred camels, the Irish price was seven female slaves. The average person wouldn't have that many slaves or camels, so people in Arabia or Ireland would band together into clan/family-based blood-money-paying-groups that acted kind of like insurance companies. If a member got convicted of a crime, everyone else would come together to help them pony up the money. I assume this helped incentivize people's families to discourage them from committing crimes. But it has the same feeling of nobody expecting very many crimes to be committed. How much of medieval Arabia’s GDP consisted of transfers of 100 camels from murderers to victims’ families?

One little-admitted but much-worried-about justification for mass incarceration in our society is the concern that some people are just so naturally violent that, left in the outside world, they would offend again and again until they died. The societies in this book didn’t seem to worry about this. If someone killed, their family would give up the relevant number of camels, and then everyone would be on their way. As far as I can tell, the Amish have *no idea* what to do about any crime more dire than using a telephone. Nobody used anything at all like incarceration. 18th century England occasionally sent prisoners somewhere horrible like America, but once the colonies revolted they experimented with jails, found them too expensive, and just sort of flailed around punishment-less until they finally discovered Australia.

There’s a lot of concern about police brutality, police racism, police failure-to-actually-control crime, et cetera. A few far-leftists have flirted with the idea of abolishing police, and the only way I can make sense of this is by analogy to something like Somali or Icelandic law. These were genuine community-based non-hierarchical legal systems. And, for the place and the time, they seem to have worked really well (Somaliland, which uses traditional Somali law, is doing way better than Somalia proper, whose law system is somewhat westernized). But I also know that it’s weirdly hard to get a good picture of how modern crime rates compare to ancient ones. On the one hand are statistics like the ones saying crime has increased by an order of magnitude since 1900 or so; on the other are findings like Steven Pinker’s that violence is constantly declining. Apply the “court made of town elders plus at least three different religious leaders plus…" to Baltimore, and the Field Of Meeting is going to get pretty crowded. On the other hand, in my past work with criminals I've been constantly surprised by how much role their families and their communities still play in their lives, and maybe a system that left legal enforcement up to them would do better than the overstretched and underperforming police.

(but what would the transfer process look like? Just cancel all funding for the Baltimore Police Department and hope for the best?)

Second, some complaints that are kind of unfair because they're along the lines of "this book is too good", but which probably need a mention.

Whenever I read a book by anyone other than David Friedman about a foreign culture, it sounds like "The X'wunda give their mother-in-law three cows every monsoon season, then pluck out their own eyes as a sacrifice to Humunga, the Volcano God".

And whenever I read David Friedman, it sounds like "The X'wunda ensure positive-sum intergenerational trade by a market system in which everyone pays the efficient price for continued economic relationships with their spouse's clan; they demonstrate their honesty with a costly signal of self-mutilation that creates common knowledge of belief in a faith whose priests are able to arbitrate financial disputes."

This is great, and it's important to fight the temptation to think of foreign cultures as completely ridiculous idiots who do stuff for no reason. But it all works out so neatly – and so much better than when anyone else treats the same topics – that I'm always nervous if I'm not familiar enough with the culture involved to know whether they're being shoehorned into a mold that's more rational-self-interest-maximizing than other anthropologists (or they themselves) would recognize.

And also, the cultural evolution idea is really optimistic. I've been trying to read a bit more about Marxism and Postmodernism lately, and they would be pretty skeptical about analyzing social systems by asking "What large-scale problem of human interaction is this system the exactly optimal solution for?"

Like, there's a perspective where lots of countries have a King, because societies that have a single central nexus to their coordination structure are able to coordinate better than ones that don't, and having them rule for life promotes long-term thinking, and them be hereditary provides a clear Schelling Point for secession disputes that prevents civil war *and* cleverly ensures that the previous ruler is incentivized to promote the peaceful transfer of power to the next one, and this is why constitutional monarchies have slightly higher yearly GDP growth than other forms of government.

And there's another perspective where lots of countries have a King, because some guy seized so much power that he can live in a giant palace and order people around all day instead of doing work. And if anyone tries to prevent him from doing that, he can arrange to have that person beheaded.

Legal Systems Very Different From Ours is very much part of the first perspective. It's a story of nations and legal systems evolving towards ever-more-optimal and ever-more-efficient institutions for the good of all, and it presents strong evidence supporting that story. I can't disagree with its evidence from within its narrative, but I still wonder how much to worry about this alternate way of looking at things.

Third, in all of the fretting about how terrible our government is, and trying to change our government to be less terrible, and trying to convince other people to go along with our terribleness-decreasing government change proposals – it's important to keep on remembering the degree to which you can still pretty much do whatever you want.

In New York, Orthodox Jews with business disputes still bring them before a tribunal of rabbis, who judge them based on Jewish law. In Pennsylvania, the Amish live their own lives in their own way pretty much completely disconnected from US government decisions (although they needed a decent lobby group, the Amish Steering Committee, to work out a few special exemptions like from the draft). Socialists occasionally set up worker-owned companies run for the good of the proletariat, and they make products and earn money just like everyone else.

If you don't like the government, you're out of luck. But if you *and your whole community* don't like the government, you can organize your own internal relations however you want. You can't override existing laws – you'll still have to pay taxes, and you can't set up a bomb-making factory in your backyard. But you can add as many new laws as you want, enforced by threat of ostracism from your community, plus any other clever commitment mechanisms you can think of. There's nothing stopping communities – a broad term covering anything from villages to church congregations to cults to political organizations to online message boards – from creating internal welfare systems to help their poorer members, taking a say in when their members marry or divorce, making home schools that educated their members' children, demanding their members in business treat their employees or business partners a certain way, et cetera.

Right now doctors' services are super-bloated and expensive because if a patient sues them they can be held liable for not filling out any of seven zillion forms or following any of twenty zillion best practices. But if the doctor only saw patients in their own community, and everyone in the community had mutual arbitration methods that worked better than the courts, maybe they could charge a fraction of the current price. This might not be illegal, as long as the community wasn't based on a protected group like race or religion. There just aren't many existing communities strong enough to make it work.

But some small seeds are starting to sprout. Social justice communities have sexual harassment policies much stronger than those of the country at large, and enforce them by ostracism and public shaming. Christians are trying to build the Benedict Option, an embedded society that works on Christian norms and rules. And there's always the seasteading movement, currently led by – oh, that's interesting – David Friedman's son.

Legal Systems Very Different From Ours hints that we could build something like Archipelago gradually, without anybody noticing. The Jews and Gypsies did something like it. So did the Amish. Maybe all we have to do is start threatening to feud against each other's families, and utopia is right around the corner.

[Legal Systems Very Different From Ours is available for free online at this link]

List Of Passages I Highlighted In My Copy Of Legal Systems Very Different From Ours

[See previous post here; read book online here.]

Question I'd never thought to ask before: are we sure it's a good idea to let people know what the laws are?

> The Chinese legal system originated somewhat over 2000 years ago in the conflict between two views of law, legalist and Confucianist. The legalists, who believed in using the rational self-interest of those subject to law to make them behave in the way desired by those making the law, advocated harsh penalties to drive the equilibrium crime rate to near zero. They supported the ideas of a strong central government, equal treatment under law, and written law available to all. Confucianists saw the issues in terms of morality rather than law and the objective not to modify by behavior by punishing and rewarding but by teaching virtue. They feared that a written law code generally available would lead to rules lawyering and supported unequal treatment based on the unequal status of those to whom the law applied…Some early writers argued against making the law code publicly available.

Not that knowing what the laws were in ancient China would necessarily help:

> Where the offense did not seem to fit any category in the code, the court felt free to find the defendant guilty of doing what ought not to be done or of violating an Imperial decree — not an actual decree, but one that

> the Emperor would have made had the matter been brought to his attention.

Actually, the ancient Chinese legal system just sucked in general – but suckiness might have been the active ingredient:

> Another way of dealing with the disproportion between the [vast] population to be controlled by the legal system and the [few] resources commanded by that system was to discourage resort to law. One way in which this was done was to make the private practice of law, in effect, a criminal offense; individuals who wanted help with their legal problems were expected to get it from the district magistrate and his staff. Another was by making involvement with the legal system potentially unpleasant for all concerned. There was no equivalent of our tort law by which an injured party could use the legal system to compel restitution—all law was, in our terms, criminal, and all prosecution public. It was legal to torture witnesses in the process of extracting information from them. Participants in the legal process were expected to act as humble petitioners, recognizing the vastly superior status of the officials they were interacting with.
>
> Making it costly to interact with the legal system was one way of reducing the amount of work required of the bureaucracy but risked providing an individual with the opportunity to injure an enemy by accusing him of an offense. It was a risky tactic, since both accuser and accused would be imprisoned, and if the accusation was found to be false the accuser was subject to the penalty that would have been imposed on the accused if found guilty. The obvious solution was to make the accusation anonymous. That problem was dealt with in a

> straightforward fashion by Ch’ing law; for an official to read an anonymous accusation was a criminal offense.

Continuing on the theme of “ancient Chinese law sucking”:

> There remained a fourth category, convicts “deserving of capital punishment.” Their names were written on a sheet on which the Emperor drew a circle, separating those who would be executed from those to be held over for another year; it is unclear whether being inside or outside the circle implied execution. A defendant guilty of family offenses who survived this process twice had his sentence commuted to deferred execution; for other offenses it took ten times.

More variations on the same theme:

> There are multiple cases where someone commits an offense on orders from a superior member of his extended family; the attitude of the court seems to be that although he must obey the order he is still criminally liable for the act; there appears to be no assumption in the legal system that an individual always has the option of acting in a way that does not violate one rule or another. That again might be interpreted as a policy driven by religion, the fear that if cosmic balance was not maintained by punishing someone for a violation of the cosmic rules, the result might be an increased risk of natural catastrophes.

On Jewish courts:

> Until the reestablishment of the State of Israel in the 20th century, the Jewish population consisted of dispersed communities living under the authority of non-Jewish rulers. Such communities were subject from time to time to persecution or even expulsion. But for

> the most part, they enjoyed judicial autonomy. Gentile rulers, Christian and Muslim, found it convenient to subcontract the job of ruling and taxing their Jewish subjects to the local Jewish authorities. The ruler set the total tax burden to be imposed on the community, the local authorities were responsible for allocating it among the residents and settling disputes among the community's members. Thus Jews in the diaspora lived largely under Jewish law. In some cases, the delegation of authority seems to have been carried to extraordinary lengths. Under Jewish law, "informing," giving gentile authorities information about a fellow Jew injurious to him, was a crime. At some times and places, informing three times was a capital offense. Someone convicted of a capital crime was executed by the (gentile) mundane authorities. It follows, if Elon's account of the situation in Spain is correct, that under some circumstances the gentile authorities were willing to execute a Jew for the crime of betraying information about other Jews. Betraying it, presumably, to the gentile authorities.

In Jewish law, someone with a terminal disease can theoretically murder (or commit any other crime) with impunity:

> One of my favorite bits of legal logic concerns someone dying of a fatal organic disease. Maimonides starts by saying that the killer of such a person is legally exempt—although, of course, one must be very sure that the disease is incurable and fatal. He goes on to add that if someone suffering from such a disease kills he is to be put to death, provided he is so considerate as to do the killing in the presence of a court.

> What if he doesn’t? Convicting him then depends on witnesses. Witnesses can only be trusted in a capital case if they themselves are at risk of punishment if their testimony is false. In this case, conspiring to use false testimony to convict someone who is innocent would result in no legal penalty, since the victim would be someone dying of a fatal organic disease and there is no penalty for killing such a person. Since the witnesses are at no risk of being put to death if their testimony is false their testimony cannot be trusted. Since their testimony cannot be trusted, there is no way of convicting the murderer. So someone dying of a fatal organic disease can commit murder with impunity, providing he takes care not to do it in the middle of the courtroom.

Jewish law says that community authorities have control over worldly but not religious law. But marriage is considered a matter of religious law, and community authorities have a strong interest in regulating marriage and divorce. How do they do it?

> One ingenious solution hit upon by the communal authorities was to argue that while the marriage was [religious], the wedding ring, being a piece of property, was [worldly]. If a marriage was celebrated without satisfying their requirements, the communal authorities held that the ring was forfeit to them. Since the groom did not own the ring, the requirements of biblical marriage had not been satisfied, hence the bride was not married and was free to marry someone else.

Some religious law, like Jewish Torah law or Islamic Sharia, prescribes draconian or otherwise ill-advised punishments. Believers have long wanted to “correct” these errors, but hesitate at openly contradicting the word of God. There are multiple traditional means for solving the problem, like saying they will only implement the religious punishment if the prosecution meets an impossibly high standard of evidence, or if the offense satisfies an unsatisfiably high number of requirements. Eg:

> Consider the case of the disobedient son. The Torah prescribes death by stoning for a child who defies his parents. Some legal authorities chose to read into the wording of the biblical verse requirements that could not in practice be satisfied—for instance, that the mother and father bringing the accusation must have identical voices and be identical in appearance. Maimonides argued that a boy below the age of thirteen could not be held responsible, that a boy of thirteen might impregnate a woman, a fact that would be known in another three months or so, at which point he would be a father not a son, hence that the prescription could only apply to a boy aged more than thirteen and less than thirteen and a quarter. In his view, the combined effect of the restrictions that could be read into the biblical passage was that the stated rule never had been and never would be applied.

Islamic sharia law famously demands that thieves' hands be cut off, but this seems to be the same sort of more-honored-in-the-breach-than-observance kind of affair:

> The hadd offense of sariqa is defined as theft, but theft that meets a variety of requirements. The thief must be a competent adult; the theft must be intentional, accomplished by stealth, of an item of more than a specified minimum value. The item must be one protected by its owner, so stealing an animal grazing at a distance from its barn does not qualify, nor does stealing from a house where you are an invited guest. Stealing perishable food does not count, because it is

> presumed that the theft is out of hunger and so permitted. The victim of the theft must attend both trial and execution.8 Arguably the list of requirements is so extensive because legal scholars, like many non-Muslim commentators, regarded the punishment—amputation of the right hand—as excessive. Since the punishment was Koranic it could not be changed, but it could be hedged around with enough qualifications so that it was unlikely to be applied—the same approach that Jewish legal scholars applied to the rule about stoning a disobedient son. A theft that did not meet the requirements for the hadd offense could still be prosecuted and punished under ta'zir [law less directly based on the Koran].

More on sharia:

> An important feature [of this system] was the separation of law and state. Law, at least in theory, was not made by the ruler but deduced by legal scholars. In the view of at least some modern scholars, that was largely true in practice as well. After the first few centuries, rulers in the Middle East were frequently foreigners to the populations they ruled, often Turkish princes who had made the transition from mercenaries in service to Arab dynasties into de facto rulers. What they wanted from the legal scholars was support for their legitimacy; while they might occasionally meddle in some legal question of immediate relevance to themselves, they were willing for the most part to leave the legal system itself in the hands of the scholars. They were even willing to subsidize the scholars by endowing mosques and madrissahs, colleges, which provided employment for legal scholars. Think of the system as what Anglo-American common law would be if law professors ran

> the world, defined not by the precedents set by judges but by the medieval equivalent of law review articles.
>
> The four schools of law are all Sunni; the Shia have their own legal rules, in most respects similar. A medieval Muslim city would have had separate courts for the four Sunni schools, the Shia, and the other tolerated religions. It was a polylegal system; disputes within each community would go to that community's courts.
>
> In Hallaq's view it was the breakdown of this system in the 19th and 20th century due to the rise of the nation state, itself a result of western influence, that effectively destroyed the traditional system. In Islamic territories under colonial rule, such as India, Indonesia, and Algeria, the colonial rulers replaced the traditional system of decentralized law independent of the state with a system of statutory law incorporating elements of traditional law, in some cases elements interpreted in ways favorable to the ruling power. After the end of the colonial period, the newly independent states followed the same path. Thus, in his view, modern "islamists" who view themselves as wishing to reinstitute Shari'a are in fact proposing something quite different and less desirable, a centralized system of state made law with rules to some degree modeled on traditional fiqh.

Another perspective on polylegal systems I hadn't considered before:

> The same issue exists in current U.S. law, which is in its own way polylegal. Each U.S. state has its own system of legal rules. Most disputes have an unambiguous location in a particular state, but not all; consider the case of a customer in California who purchases a product produced in Massachusetts from a seller in

> Texas. What court gets to decide the resulting product liability dispute? U.S. legal theory includes an elaborate set of rules for solving such “conflict of law” cases. One of those rules is diversity jurisdiction. A civil case that would normally be under state law can be heard by a federal court instead if the plaintiff and defendant are from different states, under different state laws. Think of it as a modern version of the rule that sends cross cases to the ruler’s court.

On medieval Icelandic government:

> Laws were made by a “parliament,” seats in which were a marketable commodity [called a godord]…the godord itself was in effect two different things. It was a group of men – the particular men who had agreed to follow that godi, to be members of that godord. Any man could be challenged to name his godord and was required to do so, but he was free to choose any godi within his quarter and to change to a different godord at will. It was also a bundle of rights–the right to sit in the lögrétta, appoint judges for certain courts, etc. The godord in this second sense was marketable property. It could be given away, sold, held by a partnership, inherited, or whatever. Thus seats in the law- making body were quite literally for sale.

Some interesting principles of Somali law:

> The Somali system is ultimately a feud system, one in which law is enforced by the private application of force or the threat of force, but a feud system with institutions for avoiding violence via widely respected mechanisms to arbitrate disputes. Part of what makes it successful, according to Van Notten, is that families are obligated to help defend their kin but not to help attack their

> opponents, with the result that armed conflicts are likely to lead to stalemate, and from there to arbitration.

And:

> One such oath consists of the oath-giver swearing by his marriage; if it later turns out that his oath was false, the marriage is dissolved.

And:

> If the convicted defendant refuses to pay within the specified time, he is subject to penalties ranging from a fine in honey to having one of his animals slaughtered, cooked, and eaten by the villagers each day.

And:

> For intentional murder, the penalty is a life for a life; if the murderer succeeds in fleeing abroad, a member of his family of equal status may be put to death in his stead, a rule that gives his family a strong incentive not to help him escape. In most cases the victim's family can choose to accept blood-money instead, at a rate of 100 camels for a man and 50 for a woman, although if the murder was sufficiently outrageous the court may insist on execution of the murderer.

On the English pardon system, which usually involved the offender's relatives pleading to a noble or other high official to plead for mercy, and the judge granting it if and only if a sufficiently impressive noble made the plea:

> Pardons procured in this way substitute an efficient punishment-a fine-for a less efficient punishment-execution. In doing so, they provide resources to the state and those who control it. Officials who give out

> pardons are selling them for non-pecuniary payments. Thus the legal system, in addition to providing a mechanism to reduce crime, also increases the ability of the state to maintain its authority. Considered from the standpoint of public relations, it is an elegant way of doing so. Nobody is threatened save the guilty convict. The squire is not oppressing his tenants but doing them a favor, at their request. The knowledge that such favors may occasionally be needed gives everyone in the village an incentive to be polite to the squire.

On clergyable offenses in early modern England, definition creep, and how juries interpret dumb laws as damage and route around them:

> Offenses fell into three categories according to their possible punishments: minor offenses, clergyable felonies, and non-clergyable felonies. Minor offenses such as petty larceny-theft of goods worth less than a shilling-were typically punished with punishments designed largely to shame the offender, such as public whipping or exposure in the stocks.
>
> The distinction between the second and third categories was whether or not offenders could claim benefit of clergy. Benefit of clergy originated as a legal rule permitting clerics charged with capital offenses to have their cases transferred to a church court, which did not impose capital punishment. By the 18th century, the application of the rule had changed in two important ways: The definition of clergy had been broadened to include anyone who could read (and, after 1706, any defendant whether or not he could read), and the church courts had lost their role in dealing with serious crimes. The result in many cases was that a defendant

convicted of a capital felony could plead his clergy, be branded on the thumb, and be sent home.

Under a Tudor statute, a defendant who pled his clergy could be imprisoned for up to a year. But that appears to have been done only rarely.[10] Defendants who were not actually clergymen were supposed to be allowed to plead clergy only once; branding on the thumb may have originated as a device to identify those who had pled clergy once and so could not do so again. But this restriction does not seem to have been enforced very often. Presumably the brand had some stigmatizing effect. That, plus the costs born by the defendant prior to his conviction, seem to have been at some periods the only penalty actually imposed on someone convicted of a clergyable offense.

[…]

While hanging was, during much of the century, the only punishment that a judge could impose for serious non-clergyable felonies, that did not mean that everyone charged with such a felony, or even everyone charged and guilty, was actually hanged. A substantial fraction of defendants were acquitted. Of those convicted, many were convicted of a lesser offense. A jury might find a defendant guilty of an offense that was punishable by whipping or the pillory either in order to keep the offender from pleading his clergy and being released or to prevent him from being convicted of a capital offense and hanged. After 1717, they might find him guilty of a clergyable rather than a non-clergyable felony in order to convert the punishment from hanging to transportation.

> In some cases the verdict was clearly an act of “pious perjury” by the jury. The fiction was clear when a jury found a defendant guilty of stealing from a house goods of value 39 shillings, although the goods were obviously worth much more than that; 40 shillings was the value that would make the theft non-clergyable. In other cases, the jury failed to include in its verdict features of the crime, such as the fact that the theft was from a house at night or involved breaking and entering, that would have made it non-clergyable. The combined effect of acquittals and convictions for a lesser (non-capital) offense was that, in the sample examined by Beattie, fewer than 40% of those charged with capital property felonies and fewer than 25% of those charged with murder were actually convicted of those offenses.

On the ancient Icelandic solution to class-action lawsuits:

> Transferable tort claims could solve another problem in our system as well. Consider a tort that does a small amount of damage to each of a large number of victims, small enough so that no individual victim or small group will find it worth the trouble of suing. The current solution is a class action; an enterprising lawyer gets himself named as attorney for the class of all victims, sues on their behalf, and collects damages or accepts an out of court settlement. One problem with that solution is that there is nothing much to keep the attorney from acting in his own interest instead of that of his imaginary clients, settling on terms that give him a substantial sum in real cash and them compensation in the form of discounts on their hypothetical future dealings with the defendant. Transferable claims make possible a better solution. The lawyer purchases a large number of small claims, perhaps with the assistance of middlemen, then sues on his own behalf as their

> owner.15 In this respect, at least, our legal system is a mere eleven hundred years behind the cutting edge of legal technology.

Some gypsy customs:

> The Kaale, the Finnish gypsies, a small population isolated for centuries, carry this attitude even further, refusing to openly admit the facts of human reproduction.3 They have no institution of marriage; couples that wish to reproduce are expected to first leave their family households, flee a substantial distance away—far enough so that their kin cannot find them and retrieve the woman—and return only when the child is weaned and so no longer requires a visible association with its mother. On returning, the father is expected to show the humility appropriate to one who has violated the norms of his society while the women of the mother’s generation smuggle mother and child back into the household, where the child will be expected to treat all the women of his mother’s generation as equally mothers.
>
> One result of the Kaale rejection of sexuality is to eliminate many of the taboos associated with it among other Gypsy groups. There can be no restrictions associated with menstruation since enforcing them would require recognition of the fact of menstruation, and similarly with pregnancy. A Kaale woman living in the household of her (or her partner’s) kin conceals the fact of pregnancy until shortly before delivery, and arranges for it to happen somewhere outside of the household—in modern times in a maternity hospital.

More on Gypsies, paging James C Scott:

> A third approach to enforcing an embedded legal system, also employed by gypsy communities, is to use control over information to substitute for control over physical force. I started this chapter by reporting a range of estimates for the world population of gypsies. That the estimates range over almost an order of magnitude is not an accident. Gypsies do not wish to be controlled by gaije. It is hard to control people if you cannot count them, and it is hard to count people when there is no one to one correspondence between person and name—Gypsies treat a name, more generally an identity, as fungible, property belonging to the extended family to be used by any member who finds it useful. By this tactic and others, modern gypsies make it difficult for the states that claim authority over them to monitor and control them, and so increase the range of alternatives available to gypsies and gypsy law.

One of my all-time favorite Friedman passages, this time on the Amish:

> In an earlier chapter, I suggested that in North America toleration might eventually destroy the status of gypsies as self-governing communities by making it too easy for unhappy or ostracized members to defect. Along similar lines, it is arguable that the emancipation of European Jews, starting in the late 18th century, was responsible for the decline of the Jewish communities as distinct and effectively self-ruling polities. Yet the Amish have maintained their identity, culture, and ordnung, enforcing the latter by the threat of ostracism, despite the lack of any clear barrier to prevent unhappy or excommunicated members from deserting. Such desertion is made easier, in the Amish case, by the existence of Mennonite communities, similar to the

Amish but less strict, which Amish defectors can and sometimes do join.

A critic of the Amish might argue that their upbringing, with schooling ending at eighth grade, leaves potential defectors unqualified for life in the modern world; the obvious response is that there are a lot of jobs in the modern world for which the willingness to work and the training produced by an apprenticeship starting at age fourteen are better qualifications than a high school diploma. As some evidence of the adequacy of Amish education, Amish seem to do quite well at starting and running their own small scale businesses.

One might more plausibly suggest that a social system in which courting your future mate may start as early as fourteen leaves many young people locked into a future marriage well before the point at which they have to decide whether or not to accept the Ordnung and commit themselves to the Amish lifestyle—and it is a future marriage with a spouse raised Amish. It would be interesting to know whether, when Amish do choose to leave prior to baptism, they usually do it one by one or in couples.

One could also argue that the close bonds of Amish families create a form of lock-in. Social interaction between committed Amish and relatives who have chosen not to commit is not forbidden—shunning applies only to those who have sworn to obey the Ordnung and been baptised, but then fail to live up to their commitment—but given how much of the pattern of living of the Amish is determined by their religion and culture, refusing to commit must create a substantial barrier. The barrier is higher still for those who have

> been baptized, and so would face shunning if they left the church.
>
> Finally, one might interpret the low defection rate as evidence of successful indoctrination, not only into the principles of Amish life but into the negative view held by the Amish of the lives lived by non-Amish. Reading books on the Amish, all positive, all written by sympathizers,34 one is struck by how dark their picture of the outside world is. It is a world where people spend most of their efforts in competitive endeavor and display, in keeping up with the Joneses, where lives are divided among the almost wholly separate circles of work, family, and church, where little meaningful happens or can happen, a world of boredom and alienation.
>
> There is, of course, one other possibility. Perhaps the Amish are correct in believing that they have a superior life-style, as judged by most of those who have lived it and observed the alternative—albeit a life style superior only for those who have had the good fortune to be brought up in it.

Plains Indian wife stealing:

> Wife stealing was illegal and done openly, so guilt was not an issue. Compensation was. The husband was expected to confront the wife stealer and demand generous compensation, with the amount an increasing function of the wealth of the stealer and the prowess of the husband, a decreasing function of the prowess of the stealer. There being no government to enforce the law, the threat that backed the demand was the private use of force. Pay or I’ll kill you.

Carrying out that threat was neither desired nor likely, since if the husband killed the stealer (or vice versa) the victim's kin would take revenge by killing the killer. The intended result of the threat was to set off the game that economists call "bilateral monopoly," a bargaining game in which the parties have a common interest in a peaceful resolution of their dispute but a conflict over the terms, in this case over how much will be given in compensation to the wronged husband.

What if the stealer was clearly the more dangerous man of the two—not unlikely, since a prudent man in search of status would prefer not to steal from too able a husband? The husband had the option of calling in his brothers or other kin to support his threats. The stealer, having set off the conflict in order to prove his status, had no such option—asking for help would be to admit that he had bitten off more than he could chew, and besides, he was on what everyone saw as the wrong side of the (unwritten) law. So at that point the stealer backs down and agrees to pay substantial damages, which damages are collected not by the husband but by his helpers.

Suppose the husband had no brothers? His option then was to find a champion, a brave, generous, well thought of warrior willing to take over the case and face down the stealer. This time the damage payment went to the husband. The champion's payment was the status gained by his willingness to risk himself in defense of the right and his success in forcing another warrior to back down. Much the same pattern appears in some of the Icelandic sagas, where a bully who relies too heavily on his and his friends' strength to let him violate the rights of weaker men is brought down by someone still more formidable out to establish his own status.

On Athenian juries:

> Each year, 6000 jurors were selected by lot from those who volunteered; the only qualification was being a male citizen and at least 30 years old. The size of the jury for a case varied over time and according to the nature of the case, but seems usually to have been about 500. Jurors were paid 1/2 drachma for each day they served, about half the wage of a rower, so jury service provided a sort of low end welfare.

More on Athens:

> The ordinary procedure was for the case to be privately prosecuted by any male citizen who chose to do so. If prosecution was successful and led to to the defendant paying a fine, the prosecutor would, for many but not all sorts of cases, receive a substantial fraction of the fine, sometimes as much as half, as his reward. Similarly, if the case was based on the claim that the defendant was holding property that properly belonged to the state, a successful prosecution would result in half of the property forfeiting to the state, half to the prosecutor. Such a system raises the risk of suits against defendants believed to be rich, unpopular, or both—whether or not they have broken any laws. One solution was a provision of the law under which a prosecutor who failed to get at least a fifth of the jurors to vote for conviction was himself fined, as well as barred from any future suits of the same kind.

Still on Athens:

> The victim of theft was was entitled to get back both his stolen property and a sum equal to twice its value. We worry about police planting drugs on a suspect in the

> process of search; the Athenians worried about a private party planting his own property on someone in order to accuse him of stealing it. They had a simple, solution. The accuser was allowed to search the house where he suspected his stolen property was hidden. But he had to do it naked.

And my favorite section on Athens:

> The Athenians had a straightforward solution to the problem of producing public goods such as the maintenance of a warship or the organizing of a public festival. If you were one of the richest Athenians, every two years you were obligated to produce a public good; the relevant magistrate would tell you which one.
>
> "As you doubtless know, we are sending a team to the Olympics this year. Congratulations, you are the sponsor."
>
> Or: "Look at that lovely trireme down at the dock. This year guess who gets to be captain and paymaster?"
>
> Such an obligation was called a liturgy. There were two ways to get out of it. One was to show that you were already doing another liturgy this year or had done one last year. The other was to prove that there was another Athenian, richer than you, who had not done one last year and was not doing one this year. This raises an obvious puzzle. How, in a world without accountants, income tax, public records of what people owned and what it was worth, do I prove that you are richer than I am? The answer is not an accountant's answer but an economist's — feel free to spend a few minutes trying to figure it out before you turn the page.
>
> …

The solution was simple. I offer to exchange everything I own for everything you own. If you refuse, you have admitted that you are richer than I am, and so you get to do the liturgy that was to be imposed on me.

Contra Robinson On Public Food

I.

Earlier this year, Nathan Robinson of *Current Affairs* wrote an article against school vouchers. He argued that private schools would be so focused on profit that they would sacrifice quality, and that competition wouldn't be enough to keep them in line.

I counterargued that yes it would, and cited among other things the success of food stamps (ie "food vouchers"). These give poor people access to the same dazzling variety of food choices as everyone else, usually at reasonable prices and low profit margins. If school vouchers worked as well as food vouchers, they would succeed in their mission of improving choice without sacrificing quality.

Now Robinson doubles down, sticking to his anti-voucher position and also proposing A Public Option For Food.

He starts by granting that food stamps give poor people access to an impressive variety of grocery choices:

> [Scott's] argument was a strong one. I will confess that I felt a bit stumped by it. He was right. Every week I go to the grocery store and I get relatively tasty things for relatively low prices. And so I found myself tempted by his idea that education could be provided by "learning stores" just like nutrition is provided by grocery stores.

And that capitalism deserves some credit for this:

> Capitalism is very effective at increasing production. Even Karl Marx was impressed with its achievements.

But he counters that much of this food is unhealthy and addictive. Worse, it’s pitched in ways that trick people into thinking it’s healthier than it is. For example, the average Minute Maid juice bottle has about as much sugar as soda, but deceptive corporate branding ensures most people won’t realize that. And this is inherent in the logic of capitalism. If companies can lie about the nutritional value of their food, the ones that do will outcompete the ones that don’t. If the government tries to enforce truth in advertising, companies will somehow thwart it. Maybe the box will have some small print telling you how many grams of sugar are inside – but also a giant picture of a reassuring-looking doctor, whose gentle smile is infinitely more persuasive than nutritional information pegged to a carefully mis-selected serving size. Capitalism will find a way.

Robinson compares this to the paperclip maximizer AI scenario:

> My friend Sarah likes to describe capitalism by comparing it to the “paperclip maximizer.” The paperclip maximizer is a thought experiment used to warn about the potentially deadly effects of artificial intelligence. It’s about how a machine given the wrong instructions will produce the wrong results. You have an intelligent robot, and you’d like him to collect paperclips. So you program the robot with the following instruction: “Maximize the number of paperclips in your possession.” Then you set it loose. The robot first goes around the world collecting all the existing paperclips. But once it has them all, it still isn’t finished. After all, it must maximize the number of paperclips it has. So it begins turning everything it finds into paperclips. Soon, the entire planet is nothing but a wasteland of paperclips. Eventually, the universe itself will be a vast cosmic heap of paperclips. A seemingly benign

instruction, carried out with precision and efficiency, destroyed the world.

Corporations can operate similarly. The Coca-Cola company follows a mandate: “raise revenue by selling drinks.” It sounds innocent. But the result is perverse: the company simply tries to get “as many ounces as possible into as many bodies as possible.” Every additional Coca-Cola sold is an additional dollar of revenue. There is no upper limit, then. “Growth potential” is all that matters, regardless of other consequences. And the lives of people only matter to the extent that keeping them alive longer will allow them to drink more Coke. I’m not exaggerating here. Those are the words of the Coca-Cola executives. And they flow perfectly rationally from the structure of the institution […]

People who defend capitalism do think it produces good results, because the incentive is to sell as many goods as possible, and that means selling the products that people want to buy. But, like the paperclip maximizer, “sell the goods that people will buy” is a benign rule that leads to a perverse result. A company that takes a poll of the things people want in a snack, and sells a snack with those qualities, will probably do well. But a company that researches ways to trigger biological cravings, and use subtle branding cues to trick people into thinking the product is better than it is, will do even better. The theory of a free market works at the “lemonade stand” level. Yet the paperclip robot, too, works at first: it’s what happens when the imperatives are carried to their endpoint that is so destructive. Capitalism, carried to its endpoint, will devour the earth, because that’s what its programming requires.

So, says Robinson, not only should we continue resisting school vouchers, but maybe it’s time for a public option for food:

> Let us imagine a public option for food. It is a state-funded restaurant called the American Free Diner. At the American Free Diner, anyone can show up and eat, and the food is free. It’s designed to be as healthy as possible while still being pretty tasty. It’s not going to be tastier than McDonalds fries, but the aim of the American Free Diner is not to get you to hooked on having as many meals as possible, it’s designed to get you to have a satisfying and nutritionally complete meal. And there are options. For breakfast you can have eggs and (veggie?) bacon with fruit, oatmeal, avocado on toast, or a smoothie. Lunch is soups, salads, and sandwiches. Oh, and you can also always stop by and grab free fruit or other snacks. Now, you have to eat your meal during the time you’re in the restaurant, so there’s no smuggling food away and selling it. Anyone can have up to three meals a day there; you sign up with an ID and then you get a card. If you ate at the American Free Diner for every meal, you’d be meeting every possible recommended nutritional guideline. Every town has an American Free Diner in it. The music is great and there’s a buzzing neon sign. but it’s nothing too fancy.
>
> Our “public option” for food does not mean people can’t go elsewhere, just as our public school system doesn’t mean that people can’t enroll in private schools. But it does ensure that anyone who wants to can turn up and get a high-quality meal for free, without having to have much information on their own, without having to have any money, and without having to do very much

Objection! Wouldn’t the…

> One of the reasons people will be skeptical about the Free Diner is that they have little confidence in the state to do anything right. There is a tacit acceptance of the basic idea of “public choice theory”: that state actors are just as much selfish maximizers as anyone else, and that the only difference between the state and a corporation is that the state doesn’t have to be as accountable to its consumers. But this view only captures part of the truth: sometimes states are selfish, sometimes they are not, just as human beings themselves are sometimes avaricious and sometimes benevolent. Which motive is acted upon will depend on who is in charge and how the institution is set up.
>
> There’s nothing inherent about a public school being public that requires it to be crappy. As I say, I went to a fantastic public school. But a few things are necessary for a public institution to run well. It needs to be free of bureaucratic constraint. It needs to have a clear mandate. It needs to be run by the right people. And it needs to be well-funded. When people think of the state offering food, I think they probably recoil: they think of Soviet canteens, perhaps, and government cheese. But there’s no reason things need to be this way. I could give you a dozen people who could run a nutritious, delicious, and decidedly non-dreary nonprofit diner given a sufficient budget.

I agree with this last part. I can think of many people who could run Nathan’s diner program well – but I notice Trump hasn’t put any of them in charge of anything. In fact, I can think of many people who could run a *country* well – but I notice Trump. Maybe things are more complicated than this?

II.

Capitalism sells healthy and unhealthy products with equal enthusiasm. But there’s a standard neoliberal solution here: taxes and subsidies. So for example, many cities place a special tax on sodas to increase their price and discourage consumption; soda is no longer quite so attractive relative to other options. I see a couple of advantages of selective taxation compared to Nathan’s public food option:

First, vouchers + taxes/subsidies let the rich and poor participate in the same system. I guarantee you that a public cafeteria system constructed to serve rich and poor people alike would be 90%+ poor within a year. I don’t even care if it’s a good cafeteria that rich people would otherwise enjoy. It would naturally start out with an overrepresentation of poor people. Rich people would feel uncomfortable there, both for signaling reasons (they don’t want to look like a poor person who can’t afford anything better than the public cafeteria) and for safety reasons (ie the same way rich people feel nervous going into poor neighborhoods, taking public buses, or hearing that their kids are going to be bused to poor schools). As the least tolerant rich people leave, the effect will amplify until slightly-more-tolerant rich people leave, then middle class people, and so on until it’s 100% people too poor to go anywhere else. At this point, going to the cafeterias will be stigmatizing to the point where school bullies will taunt poor kids by saying their family “eats at the cafeterias”. Also, any service that only serves poor people quickly deteriorates since none of its clientele have enough political power to demand its maintenance. You could have all this, or you could just have the poor people go to the same nice air-conditioned supermarkets as the rich people, blend in perfectly, and know that if anything goes wrong the rich people will make enough of a stink to get it fixed for them and their poor neighbors.

Second, vouchers + taxes/subsidies balance the government's interest in preventing mis-alignment with poor people's ability to control their own lives. If I love soda, and it's the only good thing in my life right now, and I've thought long and hard about how unhealthy it is, but I'd rather improve my health some other way and stick with the soda – I can. I can buy soda (at slightly higher price) and compensate by cutting back on something else – maybe Twinkies. If I'm stuck going to the government cafeteria which only serves healthy foods, I'm out of luck. Maybe they've decided that my exactly-2000-calorie-diet today will include zero soda but one Twinkie. Oh well.

Will the government cafeterias include kosher food? Probably: there are lots of Jews and they have good political clout. Will they include halal food? Well, um, are the Democrats or the Republicans in power this year? Vouchers + taxes/subsidies don't make poor Muslims choose between starving and blaspheming because the government decided their religion wasn't worthy of inclusion. Will the cafeterias include food satisfying the complicated dietary requirements of a tiny New Guinea fertility cult with all of three members in the United States? Even if the people in power are competent sympathetic, this is just asking too much.

Third, under vouchers + taxes/subsidies, everyone could eat in their own kitchen, with their own family, on their own time. Under a public option, rich people could eat in the privacy of their own home, but poor people would have to go to the centralized cafeteria. That involves travel time (many poor people already work two jobs and desperately want time to themselves) and expense (many poor people don’t have cars and already spend much of their limited budget on mass transit). It might be impossible for people who are disabled, agoraphobic, or live in very rural areas. And once you arrive – well, it’s basically high school lunch all over again. Did anyone except the top-ranking bully enjoy high school lunch? Would they have enjoyed it more if it were limited to poor people, who [insert several paragraphs of apologies and caveats here] can sometimes be on the louder and more aggressive end? Do we really want transgender people, gay people, etc to have to spend three meals a day in the middle of High School Lunch Hour Ascended To Omnisocial Phenomenon, forever?

I assume a competent administrator could grant worthy exemptions. Maybe if you have celiac disease, or PTSD, or you live too far away, or you’re in the aforementioned New Guinea fertility cult, you can get permission to skip the cafeterias and just get food vouchers. But neoliberalism means not forcing poor people to spend six months groveling to upper-class administrators before they’re allowed to live their lives the way they want. It means just letting them live the life they want, the same way rich people can.

You’re probably thinking this is an argument that vouchers + taxes/subsidies are a great solution. Nah. I’m saying that *in principle* they’re a great solution. In practice, they’ve failed spectacularly, because we subsidize the least healthy foods and restrict the production of healthy ones. From Physicians’ Cooperative For Responsible Medicine:

> Between 1995 and 2009, USDA distributed more than $246 billion in subsidies. USDA programs tend to favor the production of the unhealthiest foods. The subsidy system, updated approximately every five years, provides financial support primarily to producers of

> “commodity crops,” which include more than a dozen nonperishable crops. However, five commodity crops—corn, wheat, soybeans, cotton, and rice—receive the vast majority of subsidies. Corn and soybeans are largely used as animal feed for production of meat, dairy products, and eggs, either domestically or for export. Other commodity crops, including barley, oats, and sorghum, are also used for feed…
>
> The USDA refers to fresh fruits and vegetables as “specialty crops.” Specialty crops do not receive subsidies. In fact, farmers who participate in commodity subsidy programs are generally prohibited from growing fruits and vegetables on the so-called “base acres” of land for which they receive subsidies. This provision, enacted in 1996, restricts the ability of both small and large commodity farmers from diversifying their crops and including fruits and vegetables as part of their production.

Corn in particular is heavily supported, so much so that corn farmers end up with way more corn than anyone knows what to do with. Around the 1980s, they finally hit upon a solution: high-fructose corn syrup. From NYMag:

> [A soda tax sounds good], except that your tax dollars are simultaneously being used to promote soda-drinking. Since the eighties, the sweetener in most non-diet sodas has been high-fructose corn syrup, or HFCS. It is made from American corn rather than imported cane, and it is inexpensive, at about 30 cents a pound wholesale. (A pound is enough to make about eleven cans of Coca-Cola.) Mind you, it’s not really cheaper than cane sugar: Federal farm subsidies, amounting to about $20 billion per year, are twinned with a sugar tariff to stack that deck in favor of HFCS. In a free

market, the bottom would fall out of corn prices, and the Midwest’s economy would start to look like Greece’s.

In short: We pay federal taxes to make that can of Mountain Dew cheaper than it should be, encouraging us to buy it. Then we are scolded by public-health authorities for doing so. Then New York proposes another tax, to discourage us from buying it. This is nuts.

HFCS is, today, slipped into practically every prepared food, from ketchup to soup, because consumers respond to sweetness, particularly when it comes cheap. It’s the legacy of Earl Butz, the Nixon-era secretary of Agriculture, who had one mission: Increase production to stamp out hunger. It worked a little too well, giving us a consistent corn glut.

Some of us might not even mind subsidizing wholesome family farms in hard times, but most of the money heads straight to megacorporations like Archer Daniels Midland, in an egregious bit of corporate welfare. (Kansans who vote hard-line Republican and howl about federal spending tend to go quiet and look at their shoes when you mention this.) ADM makes HFCS by the megaton, and the Cato Institute has figured that every $1 of profit ADM earns in this business costs consumers $10.

Even Earl Butz might have had second thoughts if he’d seen the study released last month by Princeton University. It showed decisively that rats gain more weight from eating HFCS than from cane sugar. Chart the adoption of HFCS by the food industries, and it lines up pretty closely with Americans’ thickening profiles… The brain-dead-obvious solution is to eliminate both

> taxes and even things out, right? Well, two words scuttle that idea: Iowa caucuses. As long as a corn-producing state holds the definitive first primary, we're going to have pro-corn presidents.

If subsidizing soda isn't enough for you, how about pizza? From Washington Post:

> Pizza is popular because it's delicious. But the roaring success of pizza isn't entirely a free-market story. "In recent years, [the USDA] has spent many millions of dollars to increase pizza consumption among U.S. children and adults," explains Parke Wilde of Tufts University, who runs the excellent U.S. Food Policy blog.
>
> Here's what he's referring to. The USDA runs a "dairy checkoff program," which levies a small assessment on milk (15 cents for every hundredweight of milk sold or used in dairy products) and raised some $202 million in 2011. The agency then uses that money to promote products like milk and cheese. And, as it turns out, pizza.
>
> The USDA claims its checkoff program has been well worth it: For every $1 that the agency spends on increasing cheese demand, it estimates that farmers get $4.43 in additional revenue. But the results have been mixed. Milk consumption has declined in recent decades, while cheese consumption has soared.

The government doesn't just economically subsidize unhealthy food. It also spreads misinformation about which food is or isn't nutritious. For example, from NBC News:

> The U.S. government's latest eating guidelines came out Thursday — only to be greeted with the usual

accusations that they go too far, or don’t go far enough, or leave out something important.

But this time some of the hottest criticism comes from cancer researchers. And other experts are upset that the guidelines don’t say more about eating less meat.

“We are pretty disappointed the report doesn’t recommend limiting red and processed meat because of the link to cancer,” said Katie McMahon of the American Cancer Society Cancer Action Network.

Evidence goes back decades linking diets high in red and processed meats (like bacon and sausage) to cancer, McMahon told NBC News […]

Some nutritionists also said the federal government was pressured by the meat industry and by other lobby groups. “From my standpoint, Congress has caved in to the will of special interest food groups,” said Marion Nestle, a nutrition professor at New York University.

Dr. Walter Willett, who heads the nutrition department at the Harvard School of Public Health, agreed. “Unfortunately, the USDA has censored the recommendation of the Scientific Advisory Committee to consume less red meat,” Willett said.

“In fact, the dietary guidelines promote consumption of red meat as long as it is lean, which is not what the science supports. There is strong evidence that red meat consumption increases risk of diabetes, heart attacks, stroke, and some cancers (especially processed meat), and there is not good evidence that this simply due to the fat content,” Willett added.

> “This appears to reflect the powerful influences of the beef industry. Unfortunately, the public is being misled.”

So, what’s my point here?

Given a decent government that really wants what’s good for the people, we wouldn’t need a public option for food. We could just cancel all of our bad subsidies, replace them with good subsidies, and let people eat what they wanted.

Given our existing government, it shouldn’t be let within a light-year of getting to determine anybody’s diet. Speculating that maybe the people who administer the program will be virtuous competent individuals who act for the good of the public, is saying that the thing which has already happened won’t happen.

I mean, sure, maybe one of Nathan’s dozen competent people who could run the program correctly will get in charge. But then why haven’t they been put in charge of our agricultural subsidies? Why haven’t they been put in charge of the dietary guidelines? *Why aren’t they in the White House?*

III.

Because the whole “public food” argument hinges on a giant case of double standards.

Presented with evidence that corporations do bad things, it concludes that the inherent logic of capitalism demands badness.

Presented with evidence that governments do bad things, it concludes that if we just put some nice people in power, everything would go great.

Why is that? Could someone with the opposite bias propose that Coca-Cola Inc would be fine if it just got a socially responsible CEO? But that the inherent logic of government demands that people who focus on electoral demagoguery and bureaucratic empire-building will always outcompete the altruistic public servants? Why is that any less plausible than the original article's treatment?

Props to Nathan: this is a rare time the paperclip maximizer metaphor is appropriate. Capitalism is an agent more powerful and creative than any individual human, programmed with the imperative: "Give people with money what they want." And Nathan correctly points out some ways this can go wrong:

1. Companies can just lie to consumers about the product they're providing. They can say their drinks are healthy, then fill them with sugar. In fact, since this allows them to create products more desirable than any honest product could be, of course they will do this.

2. Companies can appeal to consumers by satisfying empty addictive desires that their best selves wouldn't endorse them having, like for sugar-laden drinks. Worse, it can use advertising to encourage such desires on a society-wide level, since this will make its job easier.

3. Although it will claim to be socially responsible in order to attract consumers, in reality it only cares about people proportional to how much money they have. It will care about poor people a little, because even poor people have a little money, but it will be much more attuned to the needs of the rich – which is why pharma companies invest more into curing baldness than curing malaria.

But government is the same kind of misaligned system. It’s an agent more powerful and creative than any individual human, programmed with the imperative: “Give people with [votes] what they want,” where [votes] is a proxy for real votes or any other kind of power or king-making ability – campaign-contributing-ability, lobbyists, thought-leader-status. And so:

1. Candidates can just lie to voters about anything from their personalities to the effects of their policies. Somebody can say their plan will cut taxes on everybody while slashing the deficit, and people will believe them. In fact, since this allows a candidate to propose policies more popular than any real policy could be, of course they will do this.

2. Candidates, parties, and governments can appeal to voters by satisfying empty addictive desires that their best selves wouldn’t endorse us having. Worse, it can use the media to encourage such wants on a society-wide level. How much of the news cycle is devoted to real discussion of what will improve the country, versus clickbait, outragebait, identity politics, and the latest in who’s collaborating with Russia? This isn’t a coincidence – politicians find it useful to encourage these discussions, because it’s easier to win and keep voters with identity politics issues than it is by being competent, cf. Decision 2016.

3. Although it will claim to be civic-minded and altruistic in order to win voters and maintain the consent of the governed, in reality the government only cares about people proportional to how much power they have. It will care about ordinary people a little, because even ordinary people have a vote. But it will be much more attuned to the powerful – which is why we get tax breaks for billionaires more often than medical care for the poor.

Capitalism is Moloch. But democracy is also Moloch. Both are intense competitions. Both are going to be won by people trying to *win the competition*, not people trying to be nice and do the right thing. In both, we expect that winning the competition will have something to do with being good – capitalists win partly by making good products, candidates win partly by making good policy. But both systems have equally deep misalignments that can't be eliminated just be filling them with nice people.

I'm focusing on democracy and elections here, but this is potentially true of any government. It's true of the bureaucracy – bureaucrats who focus on empire-building and gaming metrics will outperform the ones who focus on running their bureaucracy virtuously and well. It's true of dictatorships; colonels who optimize for helping the people will get replaced by colonels who optimize for pleasing the military / seizing and holding onto power. Start a revolution to sweep away everyone else and institute a form of government that isn't Moloch, and your revolution will surrender to Moloch in all of of ten seconds. If there's some elite force of commandos and technocrats to prevent your communist revolution from becoming Moloch, five seconds. The problem isn't any contigent part of the system. It's the concepts of competition, optimization, and selection. "Oh, but our system won't be competitive". Really? How do you decide who the leaders are? "Oh, we won't have a single leader, we'll make decisions by…" *Two* seconds to become Moloch, and and if your non-leader ends up with a death toll of less than a million you got off easy.

(cf: that famous scene from *A Man For All Seasons*, and the sentence here beginning "Soviet industry in its last decades").

The rookie mistake is: you see that some system is partly Moloch, so you say “Okay, we’ll fix that by putting it under the control of this other system. And we’ll control this other system by writing ‘DO NOT BECOME MOLOCH’ on it in bright red marker.”

(“I see capitalism sometimes gets misaligned. Let’s fix it by putting it under control of the government. We’ll control the government by having only virtuous people in high offices.”)

I’m not going to claim there’s a *great* alternative, but the occasionally-adequate alternative is the neoliberal one – find a couple of elegant systems that all optimize along different criteria approximately aligned with human happiness, pit them off against each other in a structure of checks and balances, hope they screw up in different places like in that swiss cheese model, keep enough individual free choice around that people can exit any system that gets too terrible, and let cultural evolution do the rest. In the market/government case, you end up with taxes, subsidies, redistribution, compensation for externalities, providing public goods, maybe breaking up monopolies – I’m not saying there’s not a lot for government to do! I’m saying it’s specific, labelled things, specific actions that compensate for specific misalignments. Try to do more, and you’ll just make matters worse.

Both Nathan and I agree that poor people should have food. But we disagree on which misaligned system should give it to them. He favors the misaligned government. I favor the misaligned free market, plus some government-led redistribution and correction.

He says that “our public option for food does not mean people can’t go elsewhere, just as our public school system doesn’t mean that people can’t enroll in private schools”.

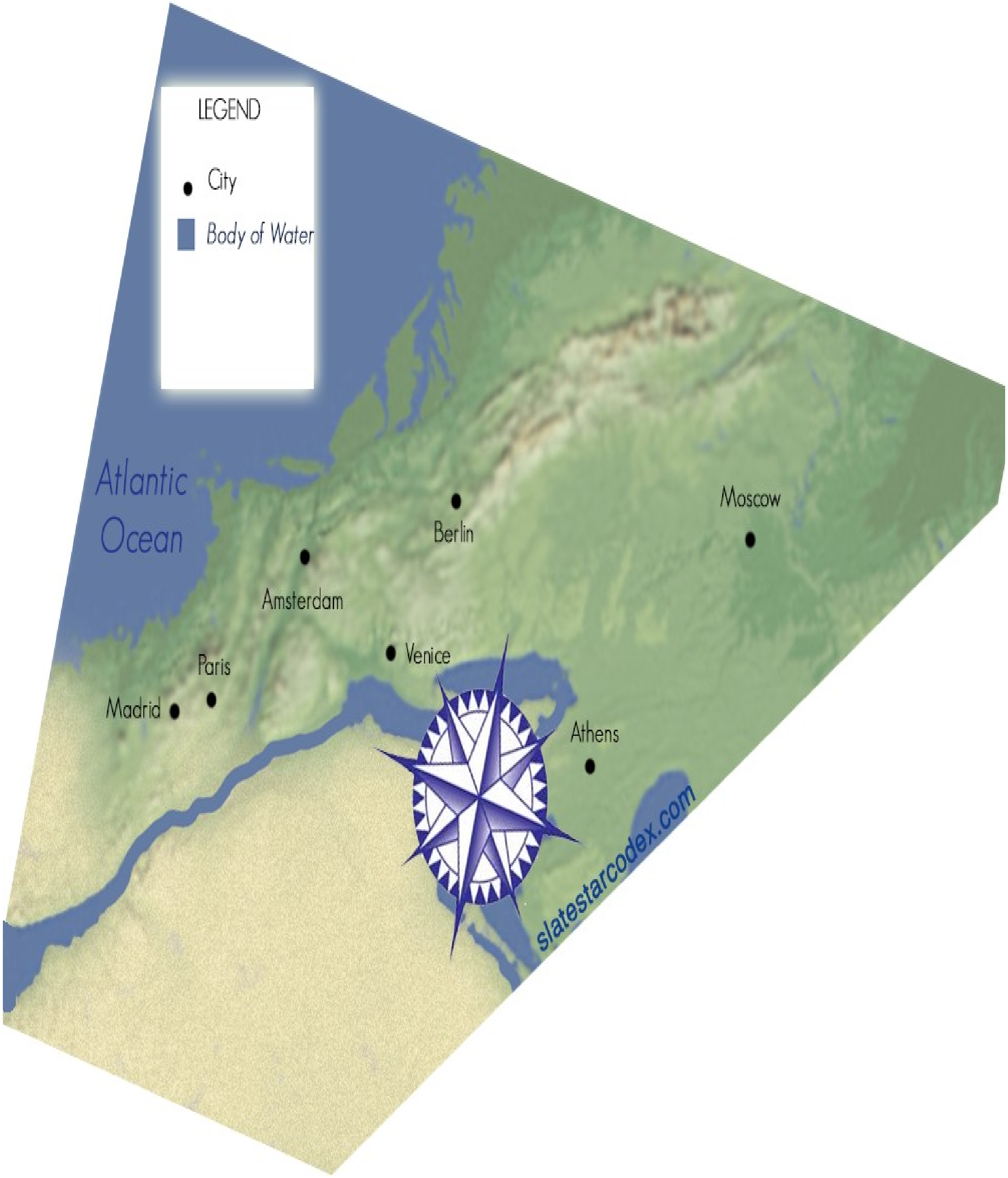
But in practice, without some system of vouchers or redistribution, only the rich would have that option; poor people wouldn’t be able to afford anything else.

And “ability to go elsewhere” is probably the most important ingredient. If I really want, I can spend some time looking into the dangers of sugary fruit juice. In fact, I did this a few years ago and haven’t bought any since; just like that, all of the horrors of capitalism lost their power over me. The last drink I bought was a sugar-free sparkling organic kiwi dragonfruit french soda with a total of five calories, because I personally preferred that to the two-thousand-or-so other options available within a five block walk of my house.

On the other hand, I also spent a long time looking into the dangers of Trump. I voted against Trump. I begged other people to vote against Trump. I wrote a blog post officially endorsing literally any person in the world who was not Trump. Despite all of this, Donald Trump is my president. I feel less satisfied with this system than with the other one, honestly.

Getting to choose my own food (and schools, and health care) works for me. I don’t want poor people to have to settle for anything less.

A Completely Accurate Map With No Distortion At All


LEGEND
City
Body of Water
Atlantic Ocean
Berlin
Moscow
Amsterdam
Venice
Paris
Madrid
Athens
slatestarcodex.com

Does something seem off about this map? Maybe Amsterdam is a little too far inland? Maybe the coastline is a bit too squashed, or the sea a bit too narrow?

No. This map is fine. The problem is with you. Think of it as a riddle: what mistake are you making in reading this map?

Hints (after a fashion) here and here.

If you give up, the answer to the riddle is here.

Book Review: Inadequate Equilibria

I.

Eliezer Yudkowsky’s catchily-titled *Inadequate Equilibria* is many things. It’s a look into whether there is any role for individual reason in a world where you can always just trust expert consensus. It’s an analysis of the efficient market hypothesis and how it relates to the idea of low-hanging fruit. It’s a self-conscious defense of the author’s own arrogance.

But most of all, it’s a book of theodicy. If the world was created by the Invisible Hand, who is good, how did it come to contain so much that is evil?

The market economy is very good at what it does, which is something like “exploit money-making opportunities” or “pick low-hanging fruit in the domain of money-making”. If you see a $20 bill lying on the sidewalk, today is your lucky day. If you see a $20 bill lying on the sidewalk in Grand Central Station, and you remember having seen the same bill a week ago, something is wrong. Thousands of people cross Grand Central every week – there’s no way a thousand people would all pass up a free $20. Maybe it’s some kind of weird trick. Maybe you’re dreaming. But there’s no way that such a low-hanging piece of money-making fruit would go unpicked for that long.

In the same way, suppose your uncle buys a lot of Google stock, because he’s heard Google has cool self-driving cars that will be the next big thing. Can he expect to get rich? No – if Google stock was underpriced (ie you could easily get rich by buying Google stock), then everyone smart enough to notice would buy it. As everyone tried to buy it, the price would go up until it was no longer underpriced. Big Wall Street banks have people who are at least as smart as your uncle, and who will notice before he does whether stocks are underpriced. They also have enough money that if they see a money-making opportunity, they can keep buying until they've driven the price up to the right level. So for Google to remain underpriced when your uncle sees it, you have to assume everyone at every Wall Street hedge fund has just failed to notice this tremendous money-making opportunity – the same sort of implausible failure as a $20 staying on the floor of Grand Central for a week.

In the same way, suppose there's a city full of rich people who all love Thai food and are willing to pay top dollar for it. The city has lots of skilled Thai chefs and good access to low-priced Thai ingredients. With the certainty of physical law, we can *know* that city will have a Thai restaurant. If it didn't, some entrepreneur would wander through, see that they could get really rich by opening a Thai restaurant, and do that. If there's no restaurant, we should feel the same confusion we feel when a $20 bill has sat on the floor of Grand Central Station for a week. Maybe the city government banned Thai restaurants for some reason? Maybe we're dreaming again?

We can take this beyond money-making into any competitive or potentially-competitive field. Consider a freshman biology student reading her textbook who suddenly feels like she's had a deep insight into the structure of DNA, easily worthy of a Nobel. Is she right? Almost certainly not. There are thousands of research biologists who would like a Nobel Prize. For all of them to miss a brilliant insight sitting in freshman biology would be the same failure as everybody missing a $20 on the floor of Grand Central, or all of Wall Street missing an easy opportunity to make money off of Google, or every entrepreneur missing a great market opportunity for a Thai restaurant. So without her finding any particular flaw in her theory, she can be pretty sure that it’s wrong – or else already discovered. This isn’t to say nobody can ever win a Nobel Prize. But winners will probably be people with access to new ground that hasn’t already been covered by other $20-seekers. Either they’ll be amazing geniuses, understand a vast scope of cutting-edge material, have access to the latest lab equipment, or most likely all three.

But go too far with this kind of logic, and you start accidentally proving that nothing can be bad anywhere.

Suppose you thought that modern science was broken, with scientists and grantmakers doing a bad job of focusing their discoveries on truly interesting and important things. But if this were true, then you (or anyone else with a little money) could set up a non-broken science, make many more discoveries than everyone else, get more Nobel Prizes, earn more money from all your patents and inventions, and eventually become so prestigious and rich that everyone else admits you were right and switches to doing science your way. There are dozens of government bodies, private institutions, and universities that could do this kind of thing if they wanted. But none of them have. So “science is broken” seems like the same kind of statement as “a $20 bill has been on the floor of Grand Central Station for a week and nobody has picked it up”. Therefore, modern science isn’t broken.

Or: suppose you thought that health care is inefficient and costs way too much. But if this were true, some entrepreneur could start a new hospital / clinic / whatever that delivered health care at lower prices *and* with higher profit margins. All the sick people would go to them, they would make lots of money, investors would trip over each other to fund their expansion into new markets, and eventually they would take over health care and be super rich. So “health care is inefficient and overpriced” seems like the same kind of statement as “a $20 bill has been on the floor of Grand Central Station for a week and nobody has picked it up.” Therefore, health care isn’t inefficient or overpriced.

Or: suppose you think that US cities don’t have good mass transit. But if lots of people want better mass transit and are willing to pay for it, this is a great money-making opportunity. Entrepreneurs are pretty smart, so they would notice this money-making opportunity, raise some funds from equally-observant venture capitalists, make a better mass transit system, and get really rich off of all the tickets. But nobody has done this. So “US cities don’t have good mass transit” seems like the same kind of statement as “a $20 bill has been on the floor of Grand Central Station for a week and nobody has picked it up.” Therefore, US cities have good mass transit, or at least the best mass transit that’s economically viable right now.

This proof of God’s omnibenevolence is followed by Eliezer’s observations that the world seems full of evil. For example:

Eliezer’s wife Brienne had Seasonal Affective Disorder. The consensus treatment for SAD is “light boxes”, very bright lamps that mimic sunshine and make winter feel more like summer. Brienne tried some of these and they didn’t work; her seasonal depression got so bad that she had to move to the Southern Hemisphere three months of every year just to stay functional. No doctor had any good ideas about what to do at this point. Eliezer did some digging, found that existing light boxes were still way less bright than the sun, and jury-rigged a much brighter version. This brighter light box cured Brienne’s depression when the conventional treatment had failed. Since Eliezer, a random layperson, was able to come up with a better SAD cure after a few minutes of thinking than the establishment was recommending to him, this seems kind of like the relevant research community leaving a $20 bill on the ground in Grand Central.

Eliezer spent a few years criticizing the Bank of Japan's macroeconomic policies, which he (and many others) thought were stupid and costing Japan trillions of dollars in lost economic growth. A friend told Eliezer that the professionals at the Bank surely knew more than he did. But after a few years, the Bank of Japan switched policies, the Japanese economy instantly improved, and now the consensus position is that the original policies were deeply flawed in exactly the way Eliezer and others thought they were. Doesn't that mean Japan left a trillion-dollar bill on the ground by refusing to implement policies that even an amateur could see were correct?

And finally:

> For our central example, we'll be using the United States medical system, which is, so far as I know, the most broken system that still works ever recorded in human history. If you were reading about something in 19th-century France which was as broken as US healthcare, you wouldn't expect to find that it went on working when overloaded with a sufficiently vast amount of money. You would expect it to just not work at all.
>
> In previous years, I would use the case of central-line infections as my go-to example of medical inadequacy. Central-line infections, in the US alone, killed 60,000 patients per year, and infected an additional 200,000

> patients at an average treatment cost of $50,000/patient.
>
> Central-line infections were also known to decrease by 50% or more if you enforced a five-item checklist that included items like “wash your hands before touching the line.”
>
> Robin Hanson has old Overcoming Bias blog posts on that untaken, low-hanging fruit. But I discovered while re-Googling in 2015 that wider adoption of hand-washing and similar precautions are now finally beginning to occur, after many years—with an associated 43% nationwide decrease in central-line infections. After partial adoption.

Since he doesn’t want to focus on a partly-solved problem, he continues to the case of infant parenteral nutrition. Some babies have malformed digestive systems and need to have nutrient fluid pumped directly into their veins. The nutrient fluid formula used in the US has the wrong kinds of lipids in it, and about a third of babies who get it die of brain or liver damage. We’ve known for decades that the nutrient fluid formula has the wrong kind of lipids. We know the right kind of lipids and they’re incredibly cheap and there is no reason at all that we couldn’t put them in the nutrient fluid formula. We’ve done a bunch of studies showing that when babies get the right nutrient fluid formula, the 33% death rate disappears. But the only FDA-approved nutrient fluid formula is the one with the wrong lipids, so we just keep giving it to babies, and they just keep dying. Grant that the FDA is terrible and ruins everything, but over several decades of knowing about this problem and watching the dead babies pile up, shouldn’t *somebody* have done *something* to make this system work better?

We’ve got a proof that everything should be perfect all the time, and a reality in which a bunch of babies keep dying even though we know exactly how to save them for no extra cost. So sure. Let’s talk theodicy.

II.

Eliezer draws on the economics literature to propose three main categories of solution:

> There’s a toolbox of reusable concepts for analyzing systems I would call “inadequate”—the causes of civilizational failure, some of which correspond to local opportunities to do better yourself. I shall, somewhat arbitrarily, sort these concepts into three larger categories:
>
> 1. Cases where the decision lies in the hands of people who would gain little personally, or lose out personally, if they did what was necessary to help someone else;
>
> 2. Cases where decision-makers can’t reliably learn the information they need to make decisions, even though someone else has that information
>
> 3. Systems that are broken in multiple places so that no one actor can make them better, even though, in principle, some magically coordinated action could move to a new stable state.

The first way evil enters the world is when there is no way for people who notice a mistake to benefit from correcting it.

For example, Eliezer and his friends sometimes joke about how really stupid Uber-for-puppies style startups are overvalued. The people investing in these startups are making a mistake big enough for ordinary people like Eliezer to notice. But it’s not exploitable – there’s no way to short startups, so neither Eliezer nor anyone else can make money by correcting that error. So it’s not surprising that the error persists. All you need is one stupid investor who thinks Uber-for-puppies is going to be the next big thing, and the startup will get overfunded. All the smart investors in the world can’t fix that one person’s mistake.

The same is true, more tragically, for housing prices. There’s no way to short houses. So if 10% of investors think the housing market will go way up, and 90% think the housing market will crash, those 10% of investors will just keep bidding up housing prices against each other. This is why there are so many housing bubbles, and why ordinary people without PhDs in finance can notice housing bubbles and yet those bubbles remain uncorrected.

A more complicated version: why was Eliezer able to out-predict the Bank of Japan? Because the Bank’s policies were set by a couple of Japanese central bankers who had no particular incentive to get things right, and no particular incentive to listen to smarter people correcting them. Eliezer wasn’t alone in his prediction – he says that Japanese stocks were priced in ways that suggested most investors realized the Bank’s policies were bad. Most of the smart people with skin in the game had come to the same realization Eliezer had. But central bankers are mostly interested in prestige, and for various reasons low money supply (the wrong policy in this case) is generally considered a virtuous and reasonable thing for a central banker to do, while high money supply (the right policy in this case) is generally considered a sort of irresponsible thing to do that makes all the other central bankers laugh at you. Their payoff matrix (with totally made-up utility points) looked sort of like this:

LOW MONEY, ECONOMY BOOMS: You were virtuous and it paid off, you will be celebrated in song forever (+10)

LOW MONEY, ECONOMY COLLAPSES: Well, you did the virtuous thing and it didn't work, at least you tried (+0)

HIGH MONEY, ECONOMY BOOMS: You made a bold gamble and it paid off, nice job. (+10)

HIGH MONEY, ECONOMY COLLAPSES: You did a stupid thing everyone always says not to do, you predictably failed and destroyed our economy, fuck you (-10)

So even as evidence accumulated that high money supply was the right strategy, the Japanese central bankers looked at their payoff matrix and decided to keep a low money supply.

It should be horrifying that this system weights a small change in the reputation of a few people higher (who will realistically do well for themselves even with a reputational hit) higher than adding trillions of dollars to the economy, but *that's how the system is structured*.

In a system like this, everybody (including the Japanese central bankers) can know that increasing money supply is the right policy, but there's no way for anyone to increase their own utility by causing the money supply to be higher. So Japan will suffer a generation's worth of recession. This is dumb but inevitable.

The second way evil enters the world is when expert knowledge can't trickle down to the ordinary people who would be the beneficiaries of correct decision-making.

The stock market stays efficient because expertise brings power. When Warren Buffett proves really good at stock-picking, everyone rushes to give him their money. If an ordinary person demonstrated Buffett-like levels of acumen, every hedge fund in the country would be competing to hire him and throw billions of dollars at whatever he predicted would work. Then when he predicts that Google’s price will double next week, he’ll use his own fortune, or the fortune of the hedge fund that employs him, to throw as much money into Google as the opportunity warrants. If Goldman Sachs doesn’t have enough to do it on their own, JP Morgan will make up the difference. Good hedge funds will always have enough money to exploit the opportunities they find, because if they didn’t, there would be so many unexploited great opportunities that the rate of return on the stock market would be spectacular, and everyone would rush to give their money to good hedge funds.

But imagine that Congress makes a new law that nobody can invest more than a thousand dollars. So Goldman Sachs invests their $1000 in Google, JP Morgan invests their $1000, and now what?

One possibility is that investment gurus could spring up, people just as smart as the Goldman Sachs traders, who (for a nominal fee) will tell you which stocks are underpriced. But this is hard, and fraudulent experts can claim to be investment gurus just as easily as real ones. There will be so many fraudulent investment gurus around that nobody will be able to trust the real ones, and after the few experts invest their own $1000 in Google, the stock could remain underpriced forever.

Something like this seems to be going on in medicine. Sure, the five doctors who really understand infant nutrition can raise a big fuss about how our terrible nutritional fluid is killing thousands of babies. But let’s face it. *Everyone* is raising a big fuss about something or other. From Eliezer’s author-insert character Cecie:

> We have an economic phenomenon sometimes called the lemons problem. Suppose you want to sell a used car, and I’m looking for a car to buy. From my perspective, I have to worry that your car might be a “lemon”—that it has a serious mechanical problem that doesn’t appear every time you start the car, and is difficult or impossible to fix. Now, you know that your car isn’t a lemon. But if I ask you, “Hey, is this car a lemon?” and you answer “No,” I can’t trust your answer, because you’re incentivized to answer “No” either way. Hearing you say “No” isn’t much Bayesian evidence. Asymmetric information conditions can persist even in cases where, like an honest seller meeting an honest buyer, both parties have strong incentives for accurate information to be conveyed.
>
> A further problem is that if the fair value of a non-lemon car is $10,000, and the possibility that your car is a lemon causes me to only be willing to pay you $8,000, you might refuse to sell your car. So the honest sellers with reliable cars start to leave the market, which further shifts upward the probability that any given car for sale is a lemon, which makes me less willing to pay for a used car, which incentivizes more honest sellers to leave the market, and so on.
>
> In our world, there are a lot of people screaming, “Pay attention to this thing I’m indignant about over here!” In fact, there are enough people screaming that there’s an inexploitable market in indignation. The dead-babies problem can’t compete in that market; there’s no free energy left for it to eat, and it doesn’t have an optimal indignation profile. There’s no single individual villain.

> The business about competing omega-3 and omega-6 metabolic pathways is something that only a fraction of people would understand on a visceral level; and even if those people posted it to their Facebook walls, most of their readers wouldn't understand and repost, so the dead-babies problem has relatively little virality. Being indignant about this particular thing doesn't signal your moral superiority to anyone else in particular, so it's not viscerally enjoyable to engage in the indignation. As for adding a further scream, "But wait, this matter really is important!", that's the part subject to the lemons problem. Even people who honestly know about a fixable case of dead babies can't emit a trustworthy request for attention […]
>
> By this point in our civilization's development, many honest buyers and sellers have left the indignation market entirely; and what's left behind is not, on average, good.

The beneficiaries of getting the infant-nutritional-fluid problem right are parents whose kids have a rare digestive condition. Maybe there are ten thousand of them. Maybe 10% of them are self-motivated and look online for facts about their kid's condition, and maybe 10% of those are smart enough to separate the true concern about fats from all the false concerns about how doctors are poisoning their kids with vaccines. That leaves a hundred people. Even if those hundred people raise a huge stink and petition the FDA really strongly, a hundred people aren't enough to move the wheels of bureaucracy. As for everyone else, why would they worry about nutritional fluid rather than terrorism or mass shootings or whatever all the other much-more-fun-to-worry-about things are?

Likewise:

To see how an inadequate equilibrium might arise, let's start by focusing on one tiny subfactor of the human system, namely academic research.

We'll even further oversimplify our model of academia and pretend that research is a two-factor system containing academics and grantmakers, and that a project can only happen if there's both a participating academic and a participating grantmaker.

We next suppose that in some academic field, there exists a population of researchers who are individually eager and collectively opportunistic for publications—papers accepted to journals, especially high-impact journal publications that constitute strong progress toward tenure. For any clearly visible opportunity to get a sufficiently large number of citations with a small enough amount of work, there are collectively enough academics in this field that somebody will snap up the opportunity. We could say, to make the example more precise, that the field is collectively opportunistic in 2 citations per workday—if there's any clearly visible opportunity to do 40 days of work and get 80 citations, somebody in the field will go for it.

This level of opportunism might be much more than the average paper gets in citations per day of work. Maybe the average is more like 10 citations per year of work, and lots of researchers work for a year on a paper that ends up garnering only 3 citations. We're not trying to ask about the average price of a citation; we're trying to ask how cheap a citation has to be before somebody somewhere is virtually guaranteed to try for it.

But academic paper-writers are only half the equation; the other half is a population of grantmakers.

In this model, can we suppose for argument's sake that grantmakers are motivated by the pure love of all sentient life, and yet we still end up with an academic system that is inadequate?

I might naively reply: "Sure. Let's say that those selfish academics are collectively opportunistic at two citations per workday, and the blameless and benevolent grantmakers are collectively opportunistic at one quality-adjusted life-year (QALY) per $100.[8] Then everything which produces one QALY per $100 and two citations per workday gets funded. Which means there could be an obvious, clearly visible project that would produce a thousand QALYs per dollar, and so long as it doesn't produce enough citations, nobody will work on it. That's what the model says, right?"

Ah, but this model has a fragile equilibrium of inadequacy. It only takes one researcher who is opportunistic in QALYs and willing to take a hit in citations to snatch up the biggest, lowest-hanging altruistic fruit if there's a population of grantmakers eager to fund projects like that.

Assume the most altruistically neglected project produces 1,000 QALYs per dollar. If we add a single rational and altruistic researcher to this model, then they will work on that project, whereupon the equilibrium will be adequate at 1,000 QALYs per dollar. If there are two rational and altruistic researchers, the second one to arrive will start work on the next-most-neglected project—say, a project that has 500 QALYs/$ but wouldn't garner enough citations for whatever reason—and then the field will be adequate at 500 QALYs/$. As this free energy gets eaten up (it's tasty energy from the perspective of an altruist eager for

QALYs), the whole field becomes less inadequate in the relevant respect.

But this assumes the grantmakers are eager to fund highly efficient QALY-increasing projects.

Suppose instead that the grantmakers are not cause-neutral scope-sensitive effective altruists assessing QALYs/$. Suppose that most grantmakers pursue, say, prestige per dollar. (Robin Hanson offers an elementary argument that most grantmaking to academia is about prestige.9 In any case, we can provisionally assume the prestige model for purposes of this toy example.)

From the perspective of most grantmakers, the ideal grant is one that gets their individual name, or their boss's name, or their organization's name, in newspapers around the world in close vicinity to phrases like "Stephen Hawking" or "Harvard professor." Let's say for the purpose of this thought experiment that the population of grantmakers is collectively opportunistic in 20 microHawkings per dollar, such that at least one of them will definitely jump on any clearly visible opportunity to affiliate themselves with Stephen Hawking for $50,000. Then at equilibrium, everything that provides at least 2 citations per workday and 20 microHawkings per dollar will get done.

This doesn't quite follow logically, because the stock market is far more efficient at matching bids between buyers and sellers than academia is at matching researchers to grantmakers. (It's not like anyone in our civilization has put as much effort into rationalizing the academic matching process as, say, OkCupid has put into their software for hooking up dates. It's not like anyone who did produce this public good would get paid

> more than they could have made as a Google programmer.)
>
> But even if the argument is still missing some pieces, you can see the general shape of this style of analysis. If a piece of research will clearly visibly yield lots of citations with a reasonable amount of labor, and make the grantmakers on the committee look good for not too much money committed, then a researcher eager to do it can probably find a grantmaker eager to fund it.
>
> But what if there's some intervention which could save 100 QALYs/$, yet produces neither great citations nor great prestige? Then if we add a few altruistic researchers to the model, they probably won't be able to find a grantmaker to fund it; and if we add a few altruistic grantmakers to the model, they probably won't be able to find a qualified researcher to work on it.
>
> One systemic problem can often be overcome by one altruist in the right place. Two systemic problems are another matter entirely.

The third way evil enters the world is through bad Nash equilibria.

Everyone hates Facebook. It records all your private data, it screws with the order of your timeline, it works to be as addictive and time-wasting as possible. So why don't we just stop using Facebook? More to the point, why doesn't some entrepreneur create a much better social network which doesn't do any of those things, and then we all switch to her site, and she becomes really rich, and we're all happy?

The obvious answer: all our friends are on Facebook. We want to be where our friends are. None of us expect our friends to leave, so we all stay. Even if every single one of our friends hated Facebook, none of us would have common knowledge that we would all leave at once; it's hard to organize a mass exodus. Something like an assurance contract might help, but those are pretty hard to organize. And even a few people who genuinely like Facebook and are really loud about it could ruin that for everybody. In the end, we all know we all hate Facebook and we all know we're all going to keep using it.

Or: instead of one undifferentiated mass of people, you have two masses of people, each working off the other's decision. Suppose there was no such thing as Lyft – it was Uber or take the bus. And suppose we got tired of this and wanted to invent Lyft. Could we do it at this late stage? Maybe not. The best part of Uber for passengers is that there's almost always a driver within a few minutes of you. And the best part of Uber for drivers is that there's almost always a passenger within a few minute of you. So you, the entrepreneur trying to start Lyft in AD 2017, hire twenty drivers. That means maybe passengers will get a driver…within an hour…if they're lucky? So no passenger will ever switch to Lyft, and that means your twenty drivers will get bored and give up.

Few passengers will use your app when Uber has far more drivers, and few drivers will use your app when Uber has far more passengers. Both drivers and passengers might hate Uber, and be happy to switch en masse if the other group did, but from within the system nobody can coordinate this kind of mass-switch occuring.

Or, to take a ridiculous example from the text that will obviously never happen:

> Suppose that there's a magical tower that only people with IQs of at least 100 and some amount of

conscientiousness can enter, and this magical tower slices four years off your lifespan. The natural next thing that happens is that employers start to prefer prospective employees who have proved they can enter the tower, and employers offer these employees higher salaries, or even make entering the tower a condition of being employed at all. The natural next thing that happens is that employers start to demand that prospective employees show a certificate saying that they’ve been inside the tower. This makes everyone want to go to the tower, which enables somebody to set up a fence around the tower and charge hundreds of thousands of dollars to let people in.

Now, fortunately, after Tower One is established and has been running for a while, somebody tries to set up a competing magical tower, Tower Two, that also drains four years of life but charges less money to enter. Unfortunately, there’s a subtle way in which this competing Tower Two is hampered by the same kind of lock-in that prevents a jump from [Facebook to a competing social network]. Initially, all of the smartest people headed to Tower One. Since Tower One had limited room, it started discriminating further among its entrants, only taking the ones that have IQs above the minimum, or who are good at athletics or have rich parents or something. So when Tower Two comes along, the employers still prefer employees from Tower One, which has a more famous reputation. So the smartest people still prefer to apply to Tower One, even though it costs more money. This stabilizes Tower One’s reputation as being the place where the smartest people go.

In other words, the signaling equilibrium is a two-factor market in which the stable point, Tower One, is

> cemented in place by the individually best choices of two different parts of the system. Employers prefer Tower One because it's where the smartest people go. Smart employees prefer Tower One because employers will pay them more for going there. If you try dissenting from the system unilaterally, without everyone switching at the same time, then as an employer you end up hiring the less-qualified people from Tower Two, or as an employee, you end up with lower salary offers after you go to Tower Two. So the system is stable as a matter of individual incentives, and stays in place. If you try to set up a cheaper alternative to the whole Tower system, the default thing that happens to you is that people who couldn't handle the Towers try to go through your new system, and it acquires a reputation for non-prestigious weirdness and incompetence.

III.

Robin Hanson's review calls *Inadequate Equilibria* "really two separate books, tied perhaps by a mood affiliation". Everything above was the first book. The second argues against overuse of the Outside View.

The Inside View is when you weigh the evidence around something, and go with whatever side's evidence seems most compelling. The Outside View is when you notice that you feel like you're right, but most people in the same situation as you are wrong. So you reject your intuitive feelings of rightness and assume you are probably wrong too. Five Outside View examples to demonstrate:

1. I feel like I'm an above-average driver. But I know there are surveys saying everyone believes they're above-average drivers. Since most people who believe they're an

above-average driver are wrong, I reject my intuitive feelings and assume I’m probably just an average driver.

2. *The Three Christs Of Ypsilanti* is a story about three schizophrenics who thought they were Jesus all ending up on the same psych ward. Each schizophrenic agreed that the other two were obviously delusional. But none of them could take the next step and agree they were delusional too. This is a failure of Outside-View-ing. They should have said “At least 66% of people in this psych hospital who believe they’re Jesus are delusional. This suggests there’s a strong bias, like a psychotic illness, that pushes people to think they’re Jesus. I have no more or less evidence for my Jesus-ness than those people, so I should discount my apparent evidence – my strong feeling that I am Him – and go back to my prior that almost nobody is Jesus.”

3. My father used to get roped into going to time-share presentations. Every time, he would come out really convinced that a time share was the most amazing purchase in the world and he needed to get one right away. Every time, we reminded him that every single person who bought a time share ended up regretting it. Every time, he answered that no, the salespeople explained that their time-share didn’t have any hidden problems. Every time, we reminded him that time-share salespeople are really convincing liars. Eventually, even though he still thought the presentation was really convincing, he accepted that he was probably a typical member of the group “people impressed with time-share presentations”, and almost every member of that group is wrong. So even though my father thought the offer sounded too good to be true, he decided to reject it.

4. A Christian might think to themselves: “Only about 30% of people are Christian; the other 70% have some other

religion which they believe as fervently as I believe mine. And no religion has more than 30% of people in the world. So of everyone who believes their religion as fervently as I do, at least 70% are wrong. Even though the truth of the Bible seems compelling to me, the truth of the Koran seems equally compelling to Muslims, the truth of dianetics equally compelling to Scientologists, et cetera. So probably I am overconfident in my belief in Christianity and really I have no idea whether it’s true or not.”

5. When I was very young, I would read pseudohistory books about Atlantis, ancient astronauts, and so on. All of these books seemed very convincing to me – I certainly couldn’t explain how ancient people built whatever gigantic technological marvels they made without the benefit of decent tools. And in most cases, nobody had written a good debunking (I am still angry about this). But there were a few cases in which people *did* write good debunkings that explained otherwise inexplicable things, and the books that were easily debunked were just as convincing as the ones that weren’t. For that and many other reasons, I assumed that even the ones that seemed compelling and had no good debunking were probably bunk.

But Eliezer warns that overuse of the Outside View can prevent you from having any kind of meaningful opinion at all. He worries about the situation where:

> …we all treat ourselves as having a black box receiver (our brain) which produces a signal (opinions), and treat other people as having other black boxes producing other signals. And we all received our black boxes at random—from an anthropic perspective of some kind, where we think we have an equal chance of being any observer. So we can’t start out by believing that our signal is likely to be more accurate than average.

There are definitely pathological cases of the Outside View. For example:

6. I believe in evolution. But about half of Americans believe in creation. So either way, half of people are wrong about the evolution-creation debate. Since I know I'm in a category, half of whom are wrong, I should assume there's a 50-50 chance I'm wrong about evolution.

But surely the situation isn't symmetrical? After all, the evolution side includes all the best biologists, all the most educated people, all the people with the highest IQ. The problem is, the true Outside Viewer can say "Ah, yes, but a creationist would say that their side is better, because it includes all the best fundamentalist preachers, all the world's most pious people, and all the people with the most exhaustive knowledge of Genesis. So you're in a group of people, the Group Who Believe That Their Side Is Better Qualified To Judge The Evolution-Creation Debate, and 50% of the people in that group are wrong. So this doesn't break the fundamental symmetry of the situation.

One might be tempted to respond with "fuck you", except that sometimes this is *exactly* the correct strategy. For example:

7. Go back to Example 2, and imagine that when Schizophrenic A was confronted with the other Christs, he protested that he had special evidence it was truly him. In particular, the Archangel Gabriel had spoken to him and told him he was Jesus. Meanwhile, Schizophrenic B had seen a vision where the Holy Spirit descended into him in the form of a dove. Schizophrenic A laughs. "Anyone can hallucinate a dove. But archangels are perfectly trustworthy." Schizophrenic B scoffs. "Hearing voices is a common

schizophrenic symptom, but I actually *saw* the Spirit”. Clearly *they still are not doing Outside View right*.

8. Every so often, I talk to people about politics and the necessity to see things from both sides. I remind people that our understanding of the world is shaped by tribalism, the media is often biased, and most people have an incredibly skewed view of the world. They nod their heads and agree with all of this and say it’s a big problem. Then I get to the punch line – that means they should be less certain about their own politics, and try to read sources from the other side. They shake their head, and say “I know that’s true of most people, but I get my facts from Vox, which backs everything up with real statistics and studies.” Then I facepalm so hard I give myself a concussion. This is the *same situation* where a tiny dose of Meta-Outside-View could have saved them.

So how do we navigate this morass? Eliezer recommends a four-pronged strategy:

> 1. Try to spend most of your time thinking about the object level. If you’re spending more of your time thinking about your own reasoning ability and competence than you spend thinking about Japan’s interest rates and NGDP, or competing omega-6 vs. omega-3 metabolic pathways, you’re taking your eye off the ball.
>
> 2. Less than a majority of the time: Think about how reliable authorities seem to be and should be expected to be, and how reliable you are — using your own brain to think about the reliability and failure modes of brains, since that’s what you’ve got. Try to be evenhanded in how you evaluate your own brain’s specific failures

> versus the specific failures of other brains. While doing this, take your own meta-reasoning at face value.
>
> 3. And then next, theoretically, should come the meta-meta level, considered yet more rarely. But I don’t think it’s necessary to develop special skills for meta-meta reasoning. You just apply the skills you already learned on the meta level to correct your own brain, and go on applying them while you happen to be meta-reasoning about who should be trusted, about degrees of reliability, and so on. Anything you’ve already learned about reasoning should automatically be applied to how you reason about meta-reasoning.
>
> 4. Consider whether someone else might be a better meta-reasoner than you, and hence that it might not be wise to take your own meta-reasoning at face value when disagreeing with them, if you have been given strong local evidence to this effect.

But then he mostly spends the rest of the chapter (and book) treating it as obvious that most people overuse the Outside View, and mocking it as “modest epistemology” for intellectual cowards. Eventually he decides that the Outside View is commonly invoked to cover up status anxiety.

> From what I can tell, status regulation is a second factor accounting for modesty’s appeal, distinct from anxious underconfidence. The impulse is to construct “cheater-resistant” slapdowns that can (for example) prevent dilettantes who are low on the relevant status hierarchy from proposing new Seasonal Affective Disorder treatments. Because if dilettantes can exploit an inefficiency in a respected scientific field, then this makes it easier to “steal” status and upset the current order.

So if we say something like “John has never taken a math class, so there’s not much chance that his proof of P = NP is right,” are we really implying “John isn’t high-status enough, so we shouldn’t let him get away with proving P = NP; only people who serve their time in grad school and postdoc programs should be allowed to do something cool like that”? I *know* Eliezer doesn’t believe that. Maybe he believes it’s only status regulation when it’s wrong? But then wouldn’t a better explanation be that people are trying a heuristic that is right a lot of the time, but misapplying it? I don’t know.

I found this part to be the biggest disappointment of this book. I don’t think it grappled with the claim that the Outside View (and even Meta-Outside View) are often useful. It offered vague tips for how to decide when to use them, but I never felt any kind of enlightenment, or like there had been any work done to resolve the real issue here. It was basically a hit job on Outside Viewing.

I understand the impetus. Eliezer was concerned that smart people, well-trained in rationality, would come to the right conclusion on some subject, then dismiss it based on the Outside View. One of his examples was that most of the rationalists he knows don’t believe in God. But if they took the Outside View on that question, they would have to either believe (since most people do) or at least be very uncertain (since lots of religions have at least as many adherents as atheism). He tosses this one off, but it’s clear that he’s less interested in religion than in worldly things – people who give up on cool startup ideas because the Outside View says they’ll probably fail, or who don’t come up with interesting contrarian ideas because the Outside View says most contrarians are wrong. He writes:

> Whereupon I want to shrug my hands helplessly and say, “But given that this isn’t normative probability

> theory and I haven’t seen modesty advocates appear to get any particular outperformance out of their modesty, why go there?”
>
> I think that’s my true rejection, in the following sense: If I saw a sensible formal epistemology underlying modesty and I saw people who advocated modesty going on to outperform myself and others, accomplishing great deeds through the strength of their diffidence, then, indeed, I would start paying very serious attention to modesty.

But these are some very artificial goalposts. The point of modesty isn’t that it lets you do great things. It’s that it lets you avoid shooting yourself in the foot. Every time my father doesn’t buy a time-share, modesty has triumphed.

To be very uncharitable, Eliezer seems to be making the same mistake as an investing book which says that you should always buy stock. After all, Warren Buffett bought stock, and look how well he’s doing! Peter Thiel bought stock, and now he’s a super-rich aspiring oceanic vampire! And (the very rich person writing the book concludes) I myself bought lots of stock, and now I am a rich self-help book author. Can you name a single person who became a billionaire by *not* buying stock? *I didn’t think so*.

To be more charitable, Eliezer might be writing to his audience. He predicts that the people who read his book will mostly be smarter than average, and generally at the level where using the Outside View hurts them rather than harms them. He writes:

> There are people who think we all ought to [use the Outside View to converge] toward each other as a matter of course. They reason:

> a) on average, we can't all be more meta-rational than average; and
>
> b) you can't trust the reasoning you use to think you're more meta-rational than average. After all, due to Dunning-Kruger, a young-Earth creationist will also think they have plausible reasoning for why they're more meta-rational than average.
>
> … Whereas it seems to me that if I lived in a world where the average person on the street corner were Anna Salamon or Nick Bostrom [people Eliezer knows who are very good at rationality], the world would look extremely different from how it actually does.
>
> … And from the fact that you're reading this at all, I expect that if the average person on the street corner were you, the world would again look extremely different from how it actually does.
>
> (In the event that this book is ever read by more than 30% of Earth's population, I withdraw the above claim.)

The argument goes: You're more rational than average, so you shouldn't adjust to the average. Instead, you should identify other people who are even more rational than you (on the matter at hand) and maybe Outside View with them, but no one else. Since you are already pretty rational, you can definitely trust your judgment about who the other rational people are.

Eliezer makes the assumption that only unusually rational people will read this book (and the preliminary hidden assumption that he's rational enough to be able to make these determinations). I think this is a pretty safe claim; I don't object to it in real life. But I worry about it in the same way I worry about the philosophical Problem Of Skepticism. I don’t think I’m a brain in a vat. But I’m vaguely annoyed by knowing that an actual brain in a vat would think exactly the same thing for the same reason.

This section’s argument runs on the same principle as a financial advice book that says “ALWAYS BUY LOTS OF STOCKS, YOU ARE GREAT AT INVESTING AND IT CANNOT POSSIBLY GO WRONG” that comes in a package marked *Deliver only to Warren Buffett*. It may be appreciated, but it’s not any kind of deep breakthrough in financial strategy.

IV.

Inadequate Equilibria is a great book, but it raises more questions than it answers. Like: does our civilization have book-titling institutions? Did they warn Eliezer that maybe *Inadequate Equilibria* doesn’t scream “best-seller”? Did he come up with a theory of how they were flawed before he decided to reject their advice?

But also, it asks: how do things stay bad in the face of so much pressure to make them better? It highlights (creates?) a field of study, clumping together a lot of economic orthodoxies and original concepts into a specific kind of rational theodicy . Once you start thinking about this, it’s hard to stop, and Eliezer deserves credit for creating a toolbox of concepts useful for analyzing these problems.

Its related question – “when should you trust social consensus vs. your own reasoning?” – is derivative of the theodicy section. If there’s some giant institution full of people much smarter and better-educated than you who have spent much more time and money investigating the question, then whether you should throw away your own opinion in favor of theirs depends a lot on whether that giant institution might fail in some unexpected way.

Its final section on the Outside View and modest epistemology tries to tie up a loose end, with less success than it would like. Should you trust your own opinion over the giant institution's on the object level question? Surely you could only do so if certain conditions held – but could you trust your own opinion about whether those conditions hold? And so on to infinity. The latter part of the book acts as if it has a definitive answer – you can trust yourself, or at least trust yourself to correctly assess how trustworthy you are relative to others – but depends on Eliezer's judgment that the book will probably only find its way to people for whom that is true.

I think you should read *Inadequate Equilibria*. Given that I am a well-known reviewer of books, clearly my opinion on this subject is better than yours. Further, Scott Aaronson and Bryan Caplan also think you should read it. Are you smarter than Scott Aaronson and Bryan Caplan? I didn't think so. Whether or not your puny personal intuition feels like you would enjoy it, you should accept the judgment of our society's book-reviewing institutions and download it right now.

Against Overgendering Harassment

About 30% of the victims of sexual harassment are men. About 20% of the perpetrators of sexual harassment are women.

Don’t believe me? In a Quinnipiac poll, 60% of women and 20% of men said they’d been sexually harassed. Opinium, which sounds like a weird drug, reports 20% of women vs. 7% of men. YouGov poll in Germany finds 43% of women and 12% of men. The overall rates vary widely depending on how the pollsters frame the question, but the ratio is pretty consistent.

The data on perpetrators is less clear. The best I can find is this Australian study finding that 21% of harassers are women. The German poll finds it’s 25%. I’m less confident on this one, but 20% seems like a conservative guess.

If you prefer anecdotes to data, you can sift through this Reddit thread with 2474 comments. For example:

> I’m a junior ncm in the Canadian forces. I had a chief harass me daily which resulted in administrative actions when I tried resisting her abuse. My introduction to her was when she was telling the 20 or so people “Under her” that her dildos name is George…it went downhill from there and eventually she was groping me on the daily. I requested a geographic posting to get away from that lunatic and get an investigation underway but I was told by my WO that “these things happen for a reason”. Eight months later I was suicidal and that WO was signing my counselling and probation with her husband.

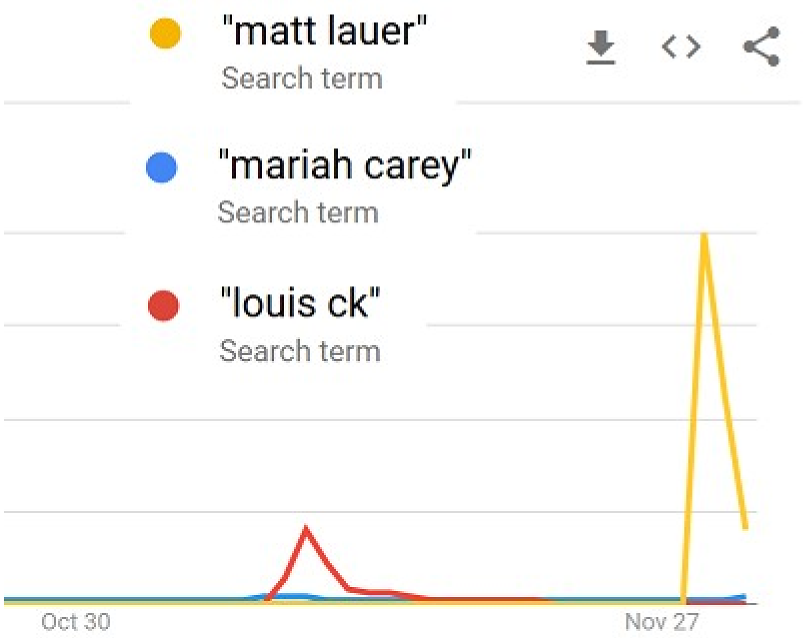
> I went up to get a drink in a crowded bar and a rather large woman ruffled my hair and said ‘I like this one’. She then started thrusting into my backside. I wasn’t sure how to respond… I just kinda waited for it to stop. It was pretty uncomfortable and I felt kinda vulnerable. In the wake of all these sexual harassment stories, I looked back on this moment and considered for the first time that that was actual sexual harassment. Huh.

Don’t believe random Redditors, but *do* believe random bloggers? Then for what it’s worth I’ve been sexually harassed by two women, and I see no reason to think my experience is anything other than typical.

But then is it odd that so few of the recent high-profile victims of sexual harassment have been men, and so few of the high-profile perpetrators women? No. Everyone has made it clear from the start that they don’t want to hear about this. The viral Facebook message that started #MeToo – at least the one I saw – urged women to come forward with their stories of sexual harassment, and men to come forward with stories of times they *perpetrated* sexual harassment. The slogan “BELIEVE WOMEN” got enshrined into a mantra, pretty ominous if you’re a guy wondering whether people will believe your harasser’s story over yours. The mainstream media strongly discouraged men from coming forward with their own cases, with articles like I’m a man who has been sexually harassed – but I don’t think it’s right for men to join in with #MeToo. Their excuse was the usual – it’s not “structural oppression”, so it doesn’t count.

(The “structural oppression” model is false, by the way. Homosexual male harassment is more prevalent than the percent gay men in the population would imply, suggesting that gay men harass men more often than straight men harass women. The obvious explanation for gender differences in harassment has always been that men constitute 80% of sexual harassers for the same reason they constitute 83% of arsonists, 81% of car thieves, and 85% of burglars. Since most men are straight, most victims are women; when the men happen to be gay, they victimize men. Men probably get victimized disproportionately often compared to the straight/gay ratio because society views harassing women as horrible but harassing men as funny. If this theory is right then it's men who are the structural victims, which means it's *your* harassment that doesn't count and *you're* the ones who shouldn't be allowed to talk about it. The "it only matters if it's structural" game isn't so much fun now, *is it?*)

Could this kind of ploy really shut up everybody? It didn't have to. Men absolutely came forward with stories of harassment by high-profile women in Hollywood, and they were summarily ignored. By freak coincidence I came across this story from last month where Mariah Carey's bodyguard accused her of sexually harassing him. Carey is much higher-profile than most of the men involved. But she didn't even publish an apology, or a denial, or try to pick holes in his story. She just assumed nobody would care – and she was right.



Having silenced or ignored all men who might be sexually harassed, the media proceeded to treat sexual harassment in the most gendered way humanly possible, constantly reinforcing that only men can do it and only women can suffer it. The Guardian, being commendably honest about its priorities: [We Must Challenge All Men About Sexual Harassment]. Newsweek worries about how [Women Are Attacked By Men In Almost Every Workplace]. The Independent thinks the story is how [powerful men seemingly never face the consequences of their actions toward women].

On the meta-level, the same publications pushed the narrative that men can't possibly understand sexual harassment, or men will never believe accusers' stories, or men refuse to believe other men can be harassers. The Guardian writes about [Men Who Are Silent After #MeToo], and the Washington Post about how [Some Men Disagree About What Counts As Sexual Harassment]. Do any women disagree about what counts as sexual harassment? Yes, the stats show that they disagree exactly as much as the men do – but who cares? The story is that women are always victims and totally understand exactly what’s going on, and men are always perpetrators with their fingers in their ears denying that a problem exists. We are told to worry about Why Women Don’t Report Sexual Harassment (against themselves) but about Why Men Don’t Speak Up About Sexual Harassment (that they see happening against women). Needless to say, every line of evidence we have shows men are less likely to report harassment that happens to them than women are.

Is this really that bad? Might the 3:1 ratio justify focusing on women? Our society already has an answer to this, and in every other case, the answer is no.

I mean, for one thing, we’re telling people to stop using the phrase “pregnant mothers” since sometimes transgender men get pregnant. It seems kind of contradictory to think of this as a pressing issue, but also think that the fact that only 30% of harassment victims are men means that we should always use female pronouns for generic harassment victims, and always generically call perpetrators “males in position of power”.

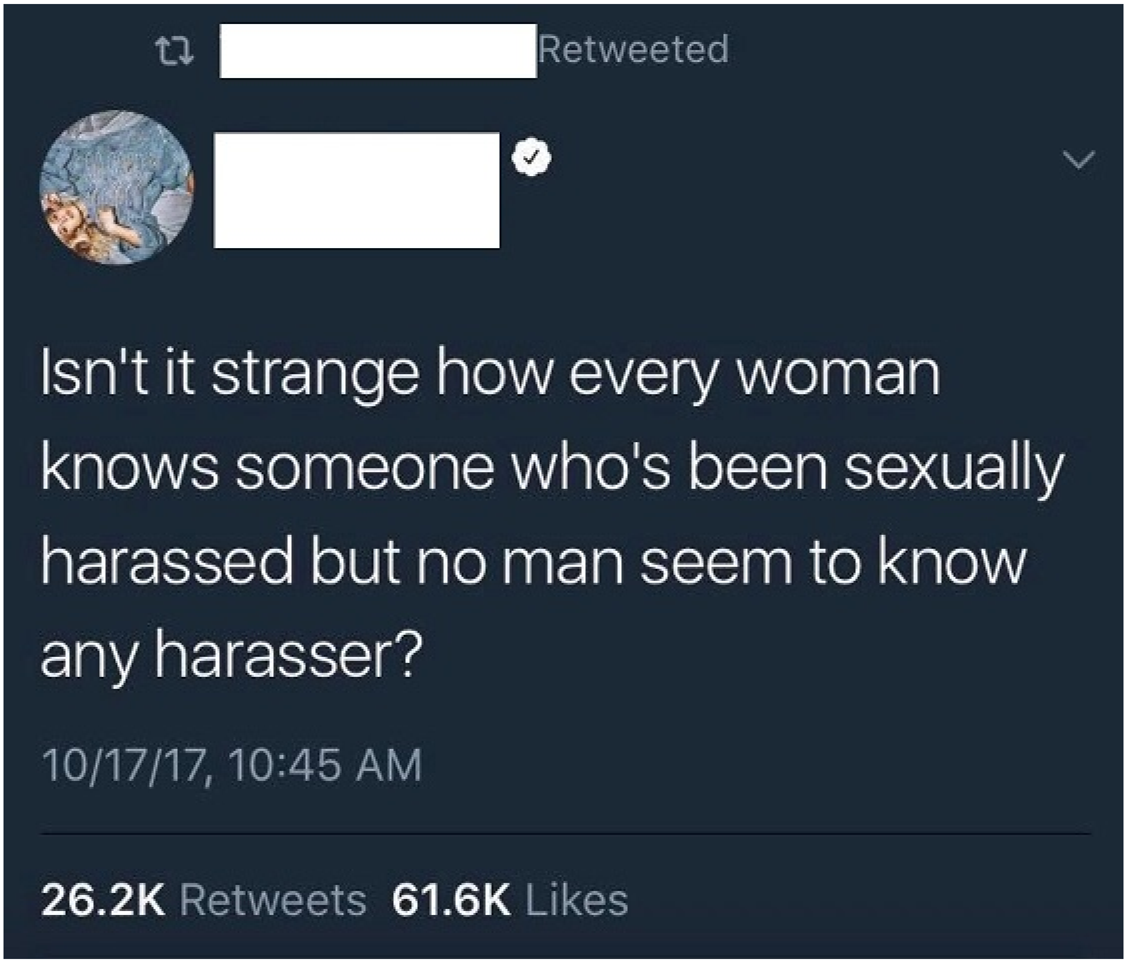
But there’s also a deeper issue. Suppose I write about how we need to do more to support the victims of terrorism. Sounds good. But what if I write about how we need to do more to support the Christian victims of Muslim terrorism? Sounds…like maybe I have an agenda. If I write story after story about how Christians need to be on the watch out for Muslim terrorists, but Muslims need to be on the watch out for other Muslims being terrorists, and if I tell Muslim victims of Christian terrorism to stay silent because that’s not “structural oppression” – then that “maybe” turns to “obviously”. This is true even if the numbers show terrorists are disproportionately Muslim.

Or suppose I write about how we need to do more to help the victims of crime. Again, sounds good. What if I write about how we need to do more to help white victims of black criminals? Again, this does not sound so good, unless you happen to be Richard Spencer. If I write articles like "We Must Challenge All Blacks About Crime" or "Whites Are Attacked By Blacks In Almost Every Neighborhood", then probably I *am* Richard Spencer. This is true regardless of whether the statistics show a racial skew in perpetrators. Nobody would accept "yeah, but I'm right about what the ratio is" as an excuse that your motives were pure.

Frames like "We need to do more to support the victims of terrorism" are an attempt to come together to stop an important social problem. Frames like "We need to do more to support the Christian victims of Muslim terrorism" are a hit job on the outgroup. Do I think that sexual harassment is being used this way? I have no other explanation for the utter predominance of genderedness in the conversation.

I've previously talked about two visions of social justice. The first vision tries to erase group differences to create a world free from stereotypes and hostility. The second vision tries to attack majority groups and spread as many stereotypes as possible about them in the hopes that the ensuing hostility raises the position of minorities. I think the gendered nature of the conversation is deliberate, being done with exactly this vision and for exactly the same reason some people talk about "Christian victims of Muslim terrorism". I think this is unfortunate. Why?

Because it ensures that nobody has more than half the picture.



[redacted] Retweeted

[redacted]

Isn't it strange how every woman knows someone who's been sexually harassed but no man seem to know any harasser?

10/17/17, 10:45 AM

26.2K Retweets **61.6K** Likes

I wonder if the woman who wrote this knows any of her close female friends who are harassers?

I mean, statistically, some of them have to be. According to the German study, 6% of women admit to being harassers. Know more than a dozen women? One of them’s probably a harasser. Don’t know which one it is? Congratulations, now you can understand why some men don’t know which of *their* same-gender friends is a harasser either.

There’s a truism that rich people can’t understand what it’s like to be poor. Why don’t you just get a minimum wage job, earn $7/hour = $60/day = $18000/year, save half of it, after few years you’ve got enough to go to a cheap college and get your ticket to the middle class? It’s possible to figure out what’s wrong with this from a third-person perspective, but it’s much easier to get the first-person perspective and be like “Oh, I guess that’s what it’s like”.

The reason this tweeter can’t understand how it’s hard to believe that your friends are sexual harassers is because she’s never tried to consider the question from a first-person perspective. I predict the sort of person who makes tweets like this is exactly the sort of person who would say “How *dare* you say any of my female friends could be sexual harassers! Don’t you even understand structural oppression?!”

Likewise, do you think this woman knows any men who are victims of sexual harassment? If you were a man who’d been sexually harassed, would *you* admit it to this woman and expect a sympathetic ear? Once she contemplates why she doesn’t know so many men who have been sexually harassed, maybe she’ll understand why some men don’t know so many women.

But more than that, if men were included in the conversation – if it were understood that a man who was sexually harassed by a female Hollywood celebrity would have the slightest chance at a fair hearing – then maybe they would feel like it was more in their self-interest to support victims.

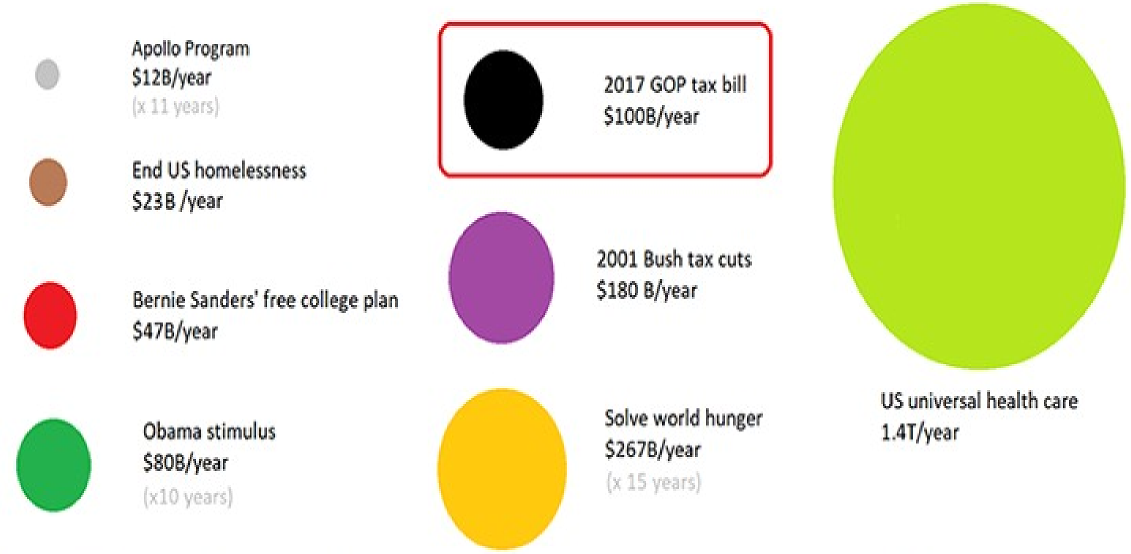
And if women were included in the conversation as potential perpetrators, they might understand why some people find it scary when people lose their careers over unsubstantiated allegations.

Instead, since we’ve chosen a narrative where one side can only ever be a victim and the other can only ever be perpetrators, we’ve made it impossible for anyone to see both perspectives. Self-interested men worry only about how to avoid allegations, self-interested women worry only about how to make sure all allegations are believed, and nobody worries about how to make a system where they expect fair treatment no matter which role they find themselves in.

The solution is to treat harassment the same way we treat terrorism. It’s something that’s bad. It’s something that some groups might do more often than other groups, but this is not the Only Relevant Factor About It, and we are suspicious of people who seem more interested in stereotyping the groups involved than in making sure everyone of every group gets justice.

And once we get good evidence that someone is guilty, we have drones bomb their house. Seriously, the terrorism model has a lot going for it.

The Tax Bill Compared To Other Very Expensive Things



All costs are in billions/year. When programs lasted a particular number of years, cost given is average cost per year, with the number of years shown in grey underneath. All costs are adjusted for inflation into 2017 dollars using the CPI Inflation Calculator. This is cost to the government only; taxpayers, corporations, college students, rocket manufacturers, etc might have gotten extra money from each policy.

1. Apollo source: www.thespacereview.com/article/1579/1 says $110B in 2010 dollars, adjusted to 2017 dollars.
2. Homelessness source: https://thinkprogress.org/it-would-actually-be-very-simple-to-end-homelessness-forever-d6f15852b2ec/, then adjusted for inflation.
3. Sanders source: http://college.usatoday.com/2017/04/17/heres-how-much-bernie-sanders-free-college-for-all-plan-would-cost/. I am taking Sanders' own numbers, which are lower than other people's because he assumes states would pay some of the cost. If they don't, cost increases by about 33%.
4. Obama source: https://en.wikipedia.org/wiki/American_Recovery_and_Reinvestment_Act_of_2009 says $831 billion over ten years = 83 billion/year, then adjusted for inflation.
5. 2017 GOP source: http://www.crfb.org/blogs/official-dynamic-score-shows-senate-tax-bill-will-still-cost-over-1-trillion says one trillion over ten years. No particular year is guaranteed to be exactly 100 billion. This is net cost AFTER potential economic benefits are subtracted from original cost.
6. Iraq source: http://www.politifact.com/truth-o-meter/statements/2010/aug/25/mark-tapscott/did-stimulus-cost-more-war-iraq/ says $709 billion from 200 to 2010 = about $100 billion/year, then adjusted for inflation. Wikipedia arbitrarily defines the war as lasting until 2011, which I have kept.
7. Bush source: https://www.thebalance.com/president-george-bush-tax-cuts-3306331 says $1.35 trillion over ten years = 135 billion/year, then adjusted for inflation.
8. Hunger source: http://www.fao.org/news/story/en/item/297804/icode/, then adjusted for inflation. The plan in the article is supposed to last fifteen years.
9. Health source: http://www.politifact.com/truth-o-meter/article/2017/jul/21/how-expensive-would-single-payer-system-be/. This is Sanders' numbers for his own plan; other people's are higher. The federal government would have to pay this much extra, but private individuals wouldn't have to spend money on health care, so it might be a net savings overall.

Here is the cost of the current GOP tax bill placed in the context of other really expensive things. Although it’s not *quite* enough money to solve world hunger, it’s enough to end US homelessness four times over or fund nine simultaneous Apollo Programs.

I’m writing this post sort of as penance. During the primaries, I wrote a post arguing that Sanders’ college plan was bad. And compared to any reasonable use of the money, I still think that’s true.

But I worry that people – including me – focus way too much on the kind of bad idea that tries to help people but ends up being too expensive, and not enough on the kind of bad idea where there’s only the thinnest veneer of a claim anyone will be helped at all. If Sanders had been elected, and we were debating his college plan, people would be *worried*. The affordability of every piece of it would get run over with a fine-toothed comb. Its irresponsibility would be *noticed*.

Well, instead of Sanders we got Trump. I won’t say nobody’s talking about the tax plan – the problems with it have been all over the news – but are our fiscal irresponsibility detectors being triggered twice as strongly as they would if it was Sanders’ college plan we were considering?

There must be a toxoplasma effect going on here, where things that are possibly bad get debated to pieces, because debating them is so much fun – but things that are definitely bad, things that nobody likes, get through much more easily. Even more pessimistically, if Sanders proposed free college for everybody, it would get a lot of resistance precisely because the fact that so many people would benefit would make sure everyone knew about it and was thinking about it a lot and understood how big a deal it was. Since nobody except a few corporations benefits from the GOP tax plan, how do we even get a feel for how big and important it is?

Next election, if he’s running, I’m probably going to support Sanders, who seems like a decent person who really wants to help the poor. This is going to be a weird choice for someone who flirts with identifying as libertarian, given the whole socialism thing. But the thing is, we have antibodies to socialism. When people push socialism, we give it the scrutiny it deserves. I’m more worried about the things we don’t have antibodies to, and one of them is going to be passed by a joint session of Congress in the next week or two.

Response To Comments: The Tax Bill Is Still Very Bad

There was some good pushback on yesterday's article on taxes. But sorry, I'm still right.

Many people responded with generic low-tax anti-government positions. Fine. Let's say the government is definitely bad and taxes are definitely too high. The current tax bill is *still* not the right way to do tax cuts.

Budget director Mick Mulvaney claims that the richest 20% of people pay 95% of income tax; the Wall Street Journal's numbers are a little lower, at 84%. Total income taxes are $1.8 trillion, so the poorest 80%'s share comes out to somewhere between 90 and 280 billion. This is around the same order of magnitude as the $100 billion in tax cuts in the current GOP bill. So it looks like one alternative to this bill, no more or less costly, would be to halve income taxes for the bottom 80% of the population, maybe anyone making less than $100,000.

Is there any reason to prefer the existing GOP proposal to this one?

The only argument I can think of is that corporations are good because they make investments are hire employees and stimulate the economy. But…

First of all, the IGM Forum asked the nation's top economists whether the current tax bill would substantially raise GDP. 51% said it wouldn't, 36% said they weren't sure, and only 2% (= 1 economist) thought it probably would.

Second, as Marxist and anti-corporate a site as Forbes notes that A Corporate Tax Cut Won't Boost Economic Growth, because they went around and asked a lot of CEOs whether they were going to invest the tax cut in cool economic-growth boosting stuff, and the CEOs mostly said no, they would probably just increase shareholder dividends.

Third, for the past few decades there's been a weird uncoupling between economic growth and the fortunes of most people in the developed world. I won't insult your intelligence by re-posting the same graph you've seen a thousand times, but this isn't subtle. If all the economists and all the CEOs are wrong, and we get a 3% boost in GDP over a decade or something, I expect when I open a holo-newspaper in 2027 it'll be about how mysterious it is that average middle-class salaries are *still* pretty close to their 1970s level. I don't think you have to be a communist to believe that economic growth that just goes to a tiny subsection of the population isn't all that useful. You just have to be a utilitarian.

(I guess expanding the economy can also give us cool technology, but I would like rather less cool technology for a while, actually).

But most important, if all of this is wrong – if the CEOs are lying and really they'll spend the money on investment, and the economists are wrong and really corporate investment will turbo-charge the economy, and the past few decades of economic history are wrong and some of the gains of a turbo-charged economy go to the poor and middle-class – then the good thing that happens is that poor and middle-class people have more money.

…which is the same thing that would have happened if you had just lowered the taxes on the poor and middle-class directly, *you moron*. It's also what would happen if we spent it on welfare for the poor, on health care for the middle class. God help me, even Bernie's free college tuition would save a couple people from student loan debt.

For the corporate tax cut to be a better idea, it would have to turbo charge the economy so dramatically that even after accounting for the low chance it will work at all, the amount taken off the top by executives and shareholders, and the poor ability of economic turbo-charging to ever reach the working class, it *still* puts more money in the hands of people who need it than just giving them the money would. I am not an economist and I don't know as much about multipliers as I should, but I have not heard anyone seriously assert this.

Last week I criticized socialists who prefer funding complicated government programs that might eventually help poor people, to just giving poor people the money. I feel like this is the same sort of issue. Some sort of complicated scheme in which we make corporations much richer and hope this is good for the poor and middle-class in some way is a lot less certain than just *giving poor and middle-class people more money*.

Spending the tax money on social welfare programs would help give poor and middle-class people more money. Expanding the EITC would give poor and middle-class people more money. Cutting personal income taxes in lower brackets would give poor and middle-class people more money. This tax bill doesn't do any of those things, and it costs the money that would make doing any of those things easier.

It's sometimes unfair to compare real government programs to the most effective possible government program;

everything fails by that measure. But this tax bill seems so much worse than even other tax cuts that I think it’s fair to judge it as a tremendous opportunity cost.

Tax Bill 3: Don’t Mess With Taxes

I.

Thanks to everyone who commented on my last two posts, especially the many people who disagreed with me. Two things I will admit I got mostly wrong:

1. I was wrong to say there was “no case” for the tax bill. Aside from all of the minor provisions which can be good or bad, the case for slashing corporate rates is that they’re more distortionary and less efficient than other forms of taxation. Thanks to everyone who pointed this out to me.

2. Several people brought up problems with the article saying CEOs say they will just give the money back to shareholders, most notably that giving money back to shareholders may stimulate the economy in other ways.

But two things I still think are true:

1. Seriously, guys, I admit I don’t know as much about economics as some of you, but I am working off of a poll of the country’s best economists who came down pretty heavily on the side of this not significantly increasing growth. If you want to tell me that it would, your job isn’t to explain Economics 101 theories to me even louder, it’s to explain how the country’s best economists are getting it wrong. You may find this book review relevant.

2. I stand by my claim that I care less about economic growth than about where the money goes. That includes caring less about distortionary taxation, deadweight loss, and all those other concepts.

Suppose Alice is an effective altruist who supports whatever charity you think is most important and does a really good job of it. Every dollar she spends saves multiple lives. She lives in a town of 1000 people where nobody else is an effective altruist and everyone else just lives a pretty decent life and spends their extra money on, I don't know, breeding virtual cats or something.

A demon places a curse on Alice's neighbor Bob. Every time Bob pays a dollar in taxes, it destroys a random two dollars' worth of wealth somewhere in the town.

The town elders meet and decide that for some reason they have to lower taxes either on Alice or Bob. The economic case for Bob is overwhelming – taxes on him are especially inefficient because of the extra wealth they destroy.

Still, I would want a tax cut for Alice. It seems like the only important thing that happens at all in this town is Alice's charitable donations. The amount I care about this town's utility focuses pretty much entirely on that. We could give the break to Bob, and have a nominally better economy, but it would just lead to more people buying virtual cats. It could be that the extra two dollars' of wealth destroyed by Bob's taxes was some sort of useful machinery, and so taxing Bob harms economic growth. Again, it is hard to care, except insofar as that hurts Alice, the only person in town whose wealth matters much for anyone's utility.

I can imagine a world in which Bob's curse was stronger, and every dollar Bob was taxed destroyed a million dollars in value, and soon any tax on Bob meant the citizens of the town were starving to death and all of them including Alice went bankrupt. But right now the tax on Bob isn't big enough to be worse for Alice than a tax on Alice, and since

Alice is the only important person in this situation, I don't care.

I can also imagine a world where a wise economist comes to town. She says "Alice's work is the most important thing in this town, but taxing Bob destroys wealth for no reason. Some of the town elders support tax breaks for Bob, and others support tax breaks for Alice. But we can give the tax break to Bob, and then all the people who saved $2 each from the curse not being activated can give $1.50 to Alice. That way Bob is better off, Alice is better off, and potential curse victims are better off."

This is the best argument in favor of wealth creation instead of redistribution. But right now we're not doing that. We just create the wealth and then don't redistribute it, except through charity, which is a rounding error, and taxes, which everyone agrees this bill causes there to be less of. If we actually had Pareto-optimal wealth redistribution, then of course, create as much wealth as possible and redistribute it Pareto-optimally. Since we don't, we're kind of stuck.

My takeaway from this story is that in societies with a lot of marginal-value-of-money inequality, economic growth is potentially less useful than working to keep the money with people who can spend it on higher-marginal-value things. Consider three variables:

1. How low is the marginal utility of money for the person holding the average dollar, if no efforts are made to redistribute it?

2. How much economic growth are we sacrificing by choosing redistribution?

3. How high a marginal utility of money do we get by redistributing it?

Point 1 is why I stress the research showing increasing inequality eg most money going to people rich enough not to really have much use for it.

Point 2 is why I stress the economists saying that the gains from cutting corporate taxes really won't have *that much* effect on growth.

Point 3 is the one I'm least sure about. If the government were a perfect effective altruist, it would be no contest – them having the money would be thousands of times more effective than random corporations (or even random middle-class people) having it. Even if the government were to give the money as a tax break to the working classes, it still seems really obvious to me that the increased utility swamps any effect from higher economic growth. In reality, the tax cut is being funded by increasing the deficit. I don't know whether that means we need to compare it to whatever is bad about having a higher deficit, or else take as a given that the deficit has a certain amount of slack, and then compare it to other things we could do with the same money.

Imagine the government went $100 billion into debt to build a giant bronze statue of George Washington. Should we be debating whether running up the deficit is really that bad? Should we be debating the artistic merit of giant bronze statues of Washington, and whether it's actually a pretty good statue that boosts tourism in the area? Or should we be comparing it to the best possible use for that money?

(added: I would be 100% happy with a bill that cut corporate taxes exactly this much, then raised taxes somewhere else in an equally progressive way, causing there to be the same amount of taxes with less distortion)

II.

~~The fairest thing I can think of is to compare this use of $100 billion to just spreading $100 billion evenly among all the government's existing priorities.~~

~~Suppose that this tax cut was vastly better at stimulating economic growth than any reasonable person expects, and it increased growth by 1% per year. Then it would create $200 billion in value. With extreme good luck,~~ 3% of that ~~might go to the poorest quintile, giving them an extra $6 billion.~~

~~Or suppose the government keeps the $100 billion and distributes it evenly according to its existing priorities. Half of the budget is entitlement programs, and~~ 32% of those ~~go to the poorest quintile, so they would get an extra $16 billion.~~

~~I'm~~ sure these numbers are wildly off~~. But it's hard to come up with remotely plausible numbers in which the poor and working class are better off with the tax bill than without it. I think the assumptions I plugged in were overly generous: the bill won't really increase growth 1%, and although poor people have 3% of income they get much less than 3% of economic gains. Still, even under these generous assumptions, this bill gives poor people less money than the default case of not doing it.~~

~~One could argue that poor people are better off with $6 billion in actual money than $16 billion in government programs purporting to help them. But although I agree there's a multiplier, I don't know if it's this big. And government programs would also disproportionately help the poorest of the poor, compared to economic gains which would disproportionately help the richest of the poor.~~

~~I think the marginal utility from an extra dollar to the poor (and the working class, etc) is orders of magnitude higher than the same dollar going to something else. So if you want to get me to support the tax bill, don't tell me yet another reason why you think it would make the economy more efficient. Tell me why I'm wrong about this.~~

[EDIT: Commenters point out I was mistaken about the speed at which this would compound. See here. If the real growth from the bill was as high as 1%, it would probably be better for the poor than the lost government spending; if it were lower, it would take several decades to break even. So the best way to convince me to support this bill would be to find a plausible estimate of what level of growth is expected. My best guess from the economist poll is still "approximately zero".]

What To Make Of New Positive NSI-189 Results?

I.

I wanted NSI-189 to be real so badly.

Pharma companies used to love antidepressants. Millions of people are depressed. Millions of people who aren’t depressed think they are. Sell them all a pill per day for their entire lifetime, and you’re looking at a lot of money. So they poured money into antidepressant research, culminating in 80s and 90s with the discovery of selective serotonin reuptake inhibitors (SSRIs) like Prozac. Since then, research has moved into exciting new areas, like “more SSRIs”, “even more SSRIs”, “drugs that claim to be SNRIs but on closer inspection are mostly just SSRIs”, and “drugs that claim to be complicated serotonin modulators but realistically just work as SSRIs”. Some companies still go through the pantomime of inventing new supposedly-not-SSRI drugs, and some psychiatrists still go through the pantomime of pretending to be excited about them, but nobody’s heart is really in it anymore.

How did it come to this? Apparently discovering new antidepressants is really hard. Part of it is that depression has such a high placebo response rate (realistically probably mostly regression to the mean) that it’s hard for even a good medication to separate much from placebo. Another part is that psychopharmacology is just a really difficult field even at the best of times. Pharma companies tried, tried some more, and gave up. All the new no-really-not-SSRIs are the fig leaf to cover their failure. Now people are gradually giving up on even pretending. There are still lots of exciting possibilities coming from the worlds of academia and irresponsible self-experimentation, but the Very Serious People have left the field. This is a disaster, insofar as they're the only people who can get things through the FDA and into the mass market where anyone besides fringe enthusiasts will use them.

Enter NSI-189. A tiny pharma company called Neuralstem announced that they had a new antidepressant that worked on directly on neurogenesis – a totally new mechanism! nothing at all like SSRIs! – and seemed to be getting miraculous results. Lots of people (including me) suspect neurogenesis is pretty fundamental to depression in a way serotonin isn't, so the narrative really worked – we've finally figured out a way to hit the root cause of depression instead of fiddling around with knobs ten steps away from the actual problem. Irresponsible self-experimenters managed to synthesize and try some of it, and reported miraculous stories of treatment-resistant depressions vanishing overnight. *Someone had finally done the thing!*

There are many theories about what place our world holds in God's creation. Here's one with as much evidence as any other: Earth was created as a Hell for bad psychiatrists. For one thing, it would explain why there are so many of them here. For another, it would explain why – after getting all of our hopes so high – NSI-189 totally flopped in FDA trials.

I don't think the data have been published anywhere (more evidence for the theory!), but we can read off the important parts of the story from Neuralstem's press release. In Stage 1, they put 44 patients on 40 mg NSI-189 daily, another 44 patients on 80 mg daily, and 132 patients on placebo for six weeks. In Stage 2, they took the people from the placebo group who hadn't gotten better in Stage 1 and put half of *them* on NSI-189, leaving the other half on placebo – I think this was a clever trick to get a group of people pre-selected for not responding to placebo and so avoid the problem where everyone does well on placebo and so it’s a washout. But all of this was for nothing. On the primary endpoint – a depression rating instrument called MADRS – the NSI-189 group failed to significantly outperform placebo during either stage.

Neuralstem’s stock fell 61% on news of the study. Financial blog Seeking Alpha advised readers that Neuralstem Is Doomed. Investors tripped over themselves to withdraw support from a corporation that apparently was unable to handle the absolute bread-and-butter most basic job of a pharma company – fudging clinical trial results so that nobody figures out they were negative until half the US population is on their drug.

From last month’s *New York Times*:

> The first thing you feel when a [drug] trial fails is a sense of shame. You’ve let your patients down. You know, of course, that experimental drugs have a poor track record – but even so, *this* drug had seemed so promising (you cannot erase the image of the cancer cells dying under the microscope). You feel as if you’ve shortchanged the Hippocratic Oath […]
>
> There’s also a more existential shame. In an era when Big Pharma might have macerated the last drips of wonder out of us, it’s worth reiterating the fact: Medicines are notoriously hard to discover. The cosmos yields human drugs rarely and begrudgingly – and when a promising candidate fails to work, it is as if yet another chemical morsel of the universe has been thrown into the dumpster. The meniscus of disappointment rises inside you: That domain of human

> biology that the medicine hoped to target may never be breached therapeutically.

And so the rest of us gave a heavy sigh, shed a single tear, and went back to telling ourselves that *maybe* vortioxetine wasn’t *exactly* an SSRI, *in ways*.

II.

But the reason I’m writing about all of this now is that Neuralstem has just put out a new press release saying that actually, good news! NSI-189 works after all! Their stock rose 67%! Investment blogs are writing that Neuralstem Is A Big Winner and boasting about how much Neuralstem stock they were savvy enough to hold on to!

What are these new results? Can we believe them?

I’m still trying to figure out exactly what’s going on; the results themselves were presented at a conference and aren’t directly available. But from what I can gather from the press release, this isn’t a new trial. It’s new secondary endpoints from the first trial, that Neuralstem thinks cast a new light on the results.

What are secondary endpoints? Often during a drug trial, people want to measure whether the drug works in multiple different ways. For depression, these are usually rating scales that ask about depressive symptoms – things like “On a scale of 1 to 5, how sad are you?” or “How many times in the past month have you considered suicide?”. You could give the MADRS, a scale that focuses on emotional symptoms. Or you could give the HAM-D, a scale that focuses more on psychosomatic symptoms. Or since depression makes people think less clearly, you could give them a cognitive battery. Depending on what you want to do, all of these are potentially good choices.

But once you let people start giving a lot of tests, there’s a risk that they’ll just keep giving more and more tests until they find one that gives results they like. Remember, one out of every twenty statistical analyses you do will be positive at the 0.05 level by pure coincidence. So if you give people ten tests, you’ve got a pretty good chance of getting one positive result – at which point, you trumpet that one to the world.

Statisticians try to solve this loophole by demanding researchers pre-identify a primary endpoint. That is, you have to say beforehand which test you want to count. You can do however many tests you want, but the other ones (“secondary endpoints”) are for your own amusement and edification. The primary endpoint is the one that the magical “p = 0.05 means it works” criteria gets applied to.

Neuralstem chose the MADRS scale as their primary endpoint and got a null result. This is what they released in July that had everybody so disappointed. The recently-released data are a bunch of secondary endpoints, some of which are positive. This is the new result that has everybody so excited.

You might be asking “Wait, I thought the whole point of having primary versus secondary endpoints was so people wouldn’t do that?” Well…yes. I’m trying to figure out if there’s any angle here besides “Company does thing that you’re not supposed to do because it can always give you positive results, gets positive results, publishes a press release”. I am not an expert here. But I can’t find one.

The pattern of positive results shows pretty much the random pattern you would expect from spurious findings. They’re divided evenly among a bunch of scales, with occasional positive results on one scale followed by negative results on a very similar scale measuring the same thing. Most of them are only the tiniest iota below p = 0.05. Many of them only work at 40 mg, and disappear in the 80 mg condition; there are occasional complicated reasons why drugs can work better at lower doses, but Occam’s razor says that’s not what’s happening here. One of the results only appeared in Stage 2 of the trial, and disappeared in Stage 1 and the pooled analysis. This doesn’t look *exactly* like they just multiplied six instruments by two doses by three ways of grouping the stages, got 36 different cells, and rolled a die in each. But it’s not *too* much better than that. Who knows, maybe the drug does something? But it sure doesn’t seem to be a particularly effective antidepressant, even by our very low standards for such. Right now I am very unimpressed.

III.

Except…why did their stock jump 67%? We just got done talking about the efficient market hypothesis and the theory that the stock market is *never wrong* in a way detectable by ordinary humans.

First of all, maybe that’s wrong. My dad is a doctor, and he swears that he keeps making a lot of money from medical investments. He just sees some new medical product, says “Yeah, that sounds like the sort of thing that will work and become pretty popular”, and buys it. I keep telling him this cannot possibly work, and he keeps coming to me a year later telling me he made a killing and now has a new car. Maybe all financial theory is a total lie, and if you get a lucky feeling when looking at a company’s logo you should invest in them right away and you will always make a fortune.

Or maybe the it’s that it’s not investors’ job to answer “Does this drug work?” but rather “Will investing in this stock make me money?”. Neuralstem has mentioned that they’ll be bringing these new results in front of the FDA, presumably in the hopes of getting a Phase III trial. FDA standards seem to have gotten looser lately, and maybe a fig leaf of positive results is all they need to give the go ahead for a bigger trial anyway – after all, they wouldn’t be approving the drug, just saying more research is appropriate. Then maybe that trial would come out better. Or it would be big enough that they would discover some alternate use (remember, Viagra was originally developed to lower blood pressure, and only got switched to erectile dysfunction after Phase 1 trials). Or maybe Neuralstem will join the 21st century and hire a competent Obfuscation Department.

I don’t know. I’m beyond caring. The sign of a really deep depression is abandoning hope, and I’ve abandoned hope in NSI-189…

…which just leaves me even more time to be excited about SAGE-217, the novel GABA-A positive allosteric modulator that just passed Phase 2 trials! This one is going to be great!

[EDIT: Wait, is SAGE-217 just a weird attempt to rebrand benzodiazepines? Surely it’s got to be more than that, right?]

Classified Thread 4: Vinson Classif

This is the…monthly? bimonthly? occasional?…classified thread. Post advertisements, personals, and any interesting success stories from the last thread.

Please Take The 2018 SSC Reader Survey

If you're reading this and have previously read at least one Slate Star Codex post, please **take the 2018 SSC Survey**.

This year's survey is in three sections. If you're strapped for time, just take Section 1. If you have a little more time, take both Sections 1 and 2. If you have a lot of time, take all three sections. Each section will take about ten minutes. There's some more information on the survey itself.

You can talk about it in the comments, but don't read them until you're done taking the survey.

Preregistration Of Hypotheses For The SSC Survey

[*This post is about the 2018 SSC Survey. If you've read at least one blog post here before, please* ***take the survey*** *if you haven't already. Please don't read on until you've taken it, since this could bias your results.*]

I'm preregistering my hypotheses for the survey this year. So far I've glanced at Google's bar graphs for each individual question but haven't started exploring relationships yet, so I'm not cheating *too* badly. I'll still look for things I haven't preregistered, but I'll admit they're preliminary results only. This is the stuff I've been thinking about beforehand and will be taking more seriously:

1. I plan to replicate the general thrust of last year's results reported in Can We Link Perception And Cognition on the sample of new people who didn't take the survey last year. In particular, I'm expecting that weirder, more autistic, more liberal, more schizophrenic, and more transgender people will be more likely to display unusual patterns of perception (hollowness or ambiguity) in the Hollow Mask illusion. I expect this to become much more obvious since I've included three examples of the illusion this year including one that seems to give a wider diversity of results.

1a. I plan to replicate the results from last year that people who were better at noticing duplicate "the's" are more likely to display unusual patterns of perception on the Hollow Mask illusion.

2. I plan to conceptually replicate Mitchell et al's study showing that autistic people are less susceptible to the Shepherd Table Illusion.

3. I plan to conceptually replicate Caparos et al's study showing that politically further-right people are more likely to use global processing on a Navon task (eg when there's an H made of tiny Es, they see the H more than the Es).

4. I plan to investigate a general construct of "first sight and second thoughts" that involves people being better able to see what's actually there, and less susceptible to illusions, priors, stereotypes, and assumptions. This will involve correlations between the two Duplicate Thes illusions, the Hollow Mask illusion, the Shepherd Table illusion, the Cookies illusion, the Parentheses palindrome, the Map riddle, the Surgeon riddle, the Switched Answers task, the Cognitive Reflection test, and the Wason task.

4a. If I can figure out how to get a common factor out of all of these, I plan to see if it's the same thing I'm looking at in 1, and how it relates to the same groups.

4b. Whether this relates to a general willingness to believe strange or unpopular things. Check vs. AI risk concern and HBD support.

5. I plan to investigate a general construct of "ambiguity tolerance" that involves people being okay with a superposition of different conflicting ideas. This will involve correlations between ambiguous results on the Hollow Mask illusion, the Spinning Dancer illusion and the Squares-Circles illusion, and with answers to the questions from the Tolerance Of Ambiguity and Tolerance of Uncertainty scales.

5a. Whether perceptual ambiguity relates to cognitive ambiguity. I want to check whether people with high ambiguity tolerance on the optical illusions are more likely to say their political opponents have some good points, are less likely to say their political opponents are evil, and are less likely to say the existing political system is justifiable. Also if they're more likely to enjoy puns.

5b. To what degree this is the same construct as (1), and is stronger among the same demographic groups.

5c. I also want to see if people with high ambiguity tolerance give less extreme answers on questions in general. I'll probably use Ambition, Social Status, Romantic Life, and Morality for this, just because these seem like complicated questions there's no obvious right answer to.

5d. I plan to confirm previous studies showing low ambiguity tolerance correlates with conservative philosophy; check vs. Political Spectrum 1-10. I predict that this will be stronger for populists than for "business conservatives", so I expect the low ambiguity correlation will be weak for generic conservatives, stronger for Trump supporters, strongest for people who identify as alt-right.

6. I plan to investigate whether autistic people are more likely to give process-centered rather than person-centered answers to the two political categorization questions (categorizing Nazis, categorizing civil disobedience on gay marriage). That is, neurotypical people will be more likely to categorize based on which side wins, and autistic people will be more likely to categorize based on what procedures were followed (eg violence, civil disobedience).

6a. I also want to investigate how these correlate with political views. I may end up controlling for this as a confounder in (6) above.

6b. This is a totally wild out-of-left field idea, but I suppose I should check how these relate to the Navon figures since they're both about categorization.

7. I plan to confirm or disprove, once and for all, whether our community has more older siblings. For lack of a fancier way to do this, I'll take the set of all people who have exactly one sibling, and see what percent of them are older vs. younger. If it's significantly above 50% older, I'm going to interpret this as a birth order effect. I'll do the same with the set of people who have two siblings, three siblings, etc, and combine them all for a final determination. Half-siblings will be ignored. If you have any problems with this methodology, tell me *now*.

7a. If I find we're disproportionately older, try to use subgroups to figure out where the effect is stronger or weaker, to try to find exactly what's going on. For example, are Less Wrongers more older-skewed than SSC readers in general?

7b. Birth order by autism, Openness, and IQ/SAT.

7c. One traditional birth-order claim is that younger children are more rebellious, so check birth order vs. people who think system needs to be fine-tuned or destroyed.

8. I plan to conceptually replicate studies showing that the more older brothers (but not younger brothers, or older or younger sisters) you have, the more likely you are to be gay.

8b. See if this predicts anything else: bisexuality, transgender, gender non-conformity, political leftism, autism, possibly 'first sight and second thoughts', possibly 'ambiguity tolerance'.

9. I plan to see whether people with ADHD are more likely to prefer the buzzing city aesthetic to the quiet village aesthetic, more likely to rate themselves as more risk-taking, and more likely to describe themselves as ambitious.

10. I plan to investigate the hypothesis about sexual harassment mentioned here: that it’s higher in gender imbalanced industries only due to potential-perpetrator-to-victim ratio. I predict that in relatively gender imbalanced industries (in terms of survey categories, all three Computers fields, Finance, Physics, and Mathematics) compared to relatively gender-balanced industries (Health Care, Psychology, Art, Law, Biology), a higher percent of women will report being harassed at work, but the percent of men reporting harassing at work will remain the same.

10b. I predict that the more people identify with social justice, and the more positively they feel about feminism, the more likely they are to report both being harassed and harassing others, due to more awareness and lower threshold to report. I predict poor social skills and autism spectrum will predict more likely to say one is a harasser, due to causing unintentional offense. I predict people who are harassed more at work will also be harassed more outside of work.

11. A long time ago, I randomized people into groups and made them read articles on AI risk to see how it changed their minds. The effect mostly persisted after one month. Since those groups were randomized by birth date, and I asked respondents their birthdates, I plan to see if those effects continue to persist after a year.

These are mostly conceptual descriptions of what I’m going to do rather than algorithmic descriptions of exactly how I’m going to process the data. Part of that is that a lot of this involves statistical techniques at the limits of my abilities and I’m going to have to see if I can actually do them. Most important, I would like to learn enough about factor analysis to actually check for a General Factor Of First Sight/Second Thoughts, and a General Factor of Ambiguity Tolerance. If I have them, I’d like to use them to see if they correlate with the other things I’m wondering if they correlate with. If I can’t make this work or beg someone else to do it for me, I’ll just eyeball the correlations between individual questions, see which ones are highest, and maybe take an average of those questions or something.

Mostly I won’t be doing anything fancy or with too many branching paths to the data, but I plan to operationalize autism in two ways. First, a scale where professional diagnosis equals 3, self-diagnosis equals 2, family member equals 1, and no personal/family history equals 0. Second, the Autism Spectrum Quotient test I made people take at the bottom of the survey. I’m not at all confident these will correlate more than a weak amount, but I’ll try it and see. I might also try some kind of average of the two measures. Since there are a few things I expect to be correlated with autism – mathematical careers, bad response to clothing tags, poor social skills – I might check to see whether the first measure, the second measure, or the combination does a better job of predicting these, and stick with whichever one does. I’ll try not to base which measure I use on any of the variables I’m actually testing.

A History Of The Silmarils In The Fifth Age

[Spoiler warning for The Silmarillion]

I.

The *Silmarillion* describes the fate of the three Silmarils. Earendil kept one, and traveled with it through the sky, where it became the planet Venus. Maedhros stole another, but regretted his deed and jumped into a fiery chasm. And Maglor took the last one, but threw it into the sea in despair.

Well, Venus is still around. But what happened to the latter two? Surely over all the intervening millennia, with so many people wanting a Silmaril, they haven't just hung around in the earth and ocean?

After some research, I've developed a couple of promising leads for the location of the Silmarils in the Fifth Age.

II.

I previously sketched out the argument that Maglor's Silmaril probably belongs to a Los Angeles crime lord.

The movie *Pulp Fiction* centers around a mysterious briefcase. We're never told exactly what's inside, but we get some clues:

1. It's described as "so beautiful" and captivates anyone who looks at it
2. It shines with an inner light
3. When Jules and Vincent are trying to get it, all the shots aimed at them miss, implying they're miraculously immune to bullets, implying that they're on some kind of divine

quest.
4. Marsellus Wallace really wants to get it, and keeps killing anyone else who has it

So far this is only suggestive, but there's more. While searching for the briefcase, Jules (!) keeps quoting a verse:

> The path of the righteous man is beset on all sides by the inequities of the selfish and the tyranny of evil men. Blessed is he who, in the name of charity and good will, shepherds the weak through the valley of the darkness, for he is truly his brother's keeper and the finder of lost children. And I will strike down upon thee with great vengeance and furious anger those who would attempt to poison and destroy my brothers.

They describe this as Ezekiel 25:17, but it isn't. In fact, it isn't anywhere in the Bible, and it doesn't match any Biblical story. This isn't from the Old Testament at all. It's a description of the life of Maglor in the *Silmarillion*!

During the First Age, Maglor ruled "Maglor's Gap", a valley which connected the lands of the Elves and the lands of Morgoth. Maglor held Maglor's Gap for 450 years until Morgoth finally conquered the valley; Maglor led the retreat of his people, thus "shepherding the weak through the valley of darkness".

He fled to the fortress of his brother, Maedhros, in Himling, where he helped defend Maedhros' lands and people in battle – making him "his brother's keeper".

In the ensuing battles, he captured the young Elrond and Elros, who had been orphaned after their parents fled across the sea, and adopted them – making him "the finder of lost children".

As for “striking down with great vengeance and furious anger those who would attempt to poison and destroy my brothers”, that’s about as Noldor as it gets.

What is going on here, and why do we keep finding these connections to Maglor?

Maglor is unique as possibly the only Noldo still remaining in the world. According to Wikipedia:

> Maglor, along with Galadriel and Gil-galad, was the greatest surviving Noldo at the beginning of the Second Age. There is speculation that he remained even after the Third Age in Middle-earth, forbidden forever from returning to Valinor.

If he were still alive in our times, he would remain bound by his oath and be hunting the Silmaril. So: could Marsellus Wallace, the mysterious gang boss who wants the briefcase so badly, be Maglor himself? Given that the name “Maglor” is a Sindarinization of his birth name “Makalaure”, “Marsellus” doesn’t even sound like much of a pseudonym.

The main argument against this point is that Tolkien’s elves are usually depicted as fair-skinned and lithe, but Marsellus Wallace is shown in the movie as a big black guy. Does this disprove the theory?

It would, *unless* Marsellus were under some kind of magical glamor to hide his true appearance. And there’s actually some evidence for this.

There’s one character in *Pulp Fiction* who is clearly able to cast illusion-related magic: Mia Wallace. In the parking lot of the restaurant, she tells Vinnie “Don’t be a…”. Then she traces a square in the air with her finger, and the square appears in glittering light. Marsellus Wallace is married to someone who can cast visual illusions.

But why should we believe Marsellus' appearance is itself such an illusion? Well, in the scene with Jules and Brett, Jules puts a gun to Brett's head and asks him what Marsellus looks like. Brett says he looks like a tall bald black guy, which seems to satisfy Jules.

The hit men try to play this off as some kind of intimidation thing, but they're just going to shoot Brett anyway – there's no need to intimidate him. It would only make sense if they're actually checking how Marsellus appears to Brett – ie whether a certain illusion he's projecting is working. When they follow up with "Does he look like a bitch?", this is their foul-mouthed way of asking whether he looks androgynous. When Brett confirms that he looks masculine, this seems to satisfy the hit men, who then go ahead and shoot him. Unclear why they're expecting the illusion to fail in Brett's case, but it seems like if it has they'll need to interrogate him further and maybe track down anybody else who might have learned too much.

How is Mia Wallace able to cast these illusions?

I would guess that "Mia" is actually Maia, ie one of the Maiar who is sent from Valinor to guide Elves and Men with their good counsel and magic powers. There's a previous example of a female Maia marrying an elflord to guide him: Melian and Thingol. Mia is following in this tradition, and just as Melian granted Thingol's kingdom invulnerability to attack, so Mia grants Maglor/Marsellus the ability to look like a big muscular black guy.

We actually have further proof of this in the movie. Mia overdoses on heroin and goes unconscious. It looks like she goes a really long time without breathing. You get anoxic brain injury in like four or five minutes; Mia was out way longer than that. But once they give her adrenaline, she instantly and completely recuperates in a medically implausible way. Suffice it to say that she's proven beyond a shadow of a doubt that she doesn't have a human circulatory system, and given us at least strong evidence that she is literally immortal.

I would guess that Maglor survived, found his Silmaril, lost his Silmaril again, and that *Pulp Fiction* is an account of him getting it back. "Quentin Tarantino" is probably a made-up pen name for a group of elvish historians – the name "Quentin" obviously deriving from "Quendi", the elvish word for elves. "Tarantino" is more obscure, but it may be a reference to Tar-Atanamir, the Numenorean king who refused to die when his time came – something which must carry a lot of metaphorical associations for any elves remaining on Earth.

If all of this is true, Maglor's Silmaril probably remains with Maglor in his Los Angeles mansion.

III.

The fate of Maedhros' Silmaril is less clear, but one promising possibility is linked with the fate of Utumno.

Utumno was the fortress of the dark god Melkor before the First Age. It was built in the far north of Middle-Earth, "upon the borders of the regions of everlasting cold". Tolkien Gateway writes that "the frigid temperatures of the northern regions were thought to originate from the evil of [Melkor's] realm".

What was Utumno like? Like most of Tolkien's villains, Melkor was at least partly a technologist; his realm was one of forges and smithies ceaselessly building weapons for his war against the gods. This page describes it as “a fortress for war, with many armories, forges, dungeons and breeding pits.” Some of the descriptions sound like it was emitting pollution, destroying the land around it: “The lands of the far north were all made desolate in those days; for there Utumno was delved exceeding deep, and its pits were filled with fires and with great hosts of the servants of Melkor.”

Who manned these factories? Enslaved elves. As per the book, “All those of the Quendi who came into the hands of Melkor, ere Utumno was broken, were put there in prison, and by slow arts of cruelty were corrupted and enslaved”.

Eventually the gods decided enough was enough and marched against Utumno with a mighty host led by Tulkas, God of War. He wrestled with Melkor, defeated him, and bound him with a mighty chain.

What happened to Utumno after this? The *Silmarillion* is vague, but in retrospect it’s super obvious. *What happened to the magical factory at the North Pole run by elves?* Everyone knows the answer to that one!

Presumably Tulkas and the other gods, after defeating Melkor, decided it was poetically appropriate to turn Utumno from a place of darkness to a wonderland of holiday cheer. The elves agreed to stay on to help, and they repurposed Melkor’s forges to create toys for children around the world.

“Santa Claus” supposedly derives from St. Nicholas, on the grounds that “Santa” means “saint” and “Claus” is short for “Nicholas”. But “Santa” means a female saint; a male saint is “San”. Santa is male, so a more reasonable derivation would be “San Tulkas”. Once a year, Tulkas goes forth and distributes the toys created by the elves of Utumno.

(remember, the *Silmarillion* describes Tulkas as a huge bearded man who “laughs ever, in sport or in war, and even in the face of Melkor he laughed in battles before the Elves were born”. And remember, of his wife Nessa, it says “Deer she loves, and they follow her train whenever she goes in the wild”. Having deer follow your family around everywhere seems sounds pretty annoying, but at least it gives you a ready-made supply of draft animals.)

Since we never see Santa’s workshop, it must be hidden from the world in the same manner as the Undying Lands. How does Tulkas cross back into the mortal world to deliver gifts?

The only successful example of such a journey we have from Tolkien is that of Earendil, who travels from Middle-Earth to the Undying Lands using a Silmaril worn on his brow. Later, even after the two worlds are separated entirely, he is able use the same Silmaril to voyage through the sky in his flying boat. “The wise have said that it was by reason of the power of that holy jewel that they came in time to waters that no vessels save those of the Teleri had known”. So presumably any living being with a Silmaril upon their head can fly through the gulfs between the worlds safely.

Tulkas is a god and should have no trouble finding the only unclaimed Silmaril, the one Maedhros dropped into a chasm in the earth. His main issue would be preventing the surviving Noldor from learning what he has and invoking their vendetta. He would have to disguise it as something else, something so ridiculous that the stick-up-their-ass Noldor would never think to identify it with their holy jewels.

So…

Rudolph the Red-Nosed Reindeer
Had a very shiny nose
And if you ever saw it
You would even say it glows…

Adderall Risks: Much More Than You Wanted To Know

[Previously in series: Antidepressant Pharmacogenomics: Much More Than You Wanted To Know; SSRIs: Much More Than You Wanted To Know, etc. This is all preliminary and you should not take it as a reason to change successful medical care. None of this necessarily applies to your particular case and you should talk to your doctor if you have questions about that.]

I. Confessions Of A Gatekeeper

I didn't realize how much of a psychiatrist's time was spent gatekeeping Adderall.

The human brain wasn't built for accounting or software engineering. A few lucky people can do these things ten hours a day, every day, with a smile. The rest of us start fidgeting and checking our cell phone somewhere around the thirty minute mark. I work near the financial district of a big city, so every day a new Senior Regional Manipulator Of Tiny Numbers comes in and tells me that his brain must be broken because he can't sit still and manipulate tiny numbers as much as he wants. How come this is so hard for him, when all of his colleagues can work so diligently?

(it's because his colleagues are all on Adderall already – but telling him that will just make things worse)

He goes on to give me his story about how he's at risk of getting fired from his Senior Regional Manipulator Of Tiny Numbers position, and at this rate he's never going to get the promotion to Vice President Of Staring At Giant Spreadsheets, so do I think I can give him some Adderall to help him through?

Psychiatric guidelines are very clear on this point: only give Adderall to people who “genuinely” “have” “ADHD”.

But “ability to concentrate” is a normally distributed trait, like IQ. We draw a line at some point on the far left of the bell curve and tell the people on the far side that they’ve “got” “the disease” of “ADHD”. This isn’t just me saying this. It’s the neurostructural literature, the the genetics literature, a bunch of other studies, and the the Consensus Conference On ADHD. This doesn’t mean ADHD is “just laziness” or “isn’t biological” – of course it’s biological! Height is biological! But that doesn’t mean the world is divided into two natural categories of “healthy people” and “people who have Height Deficiency Syndrome“. Attention is the same way. Some people really do have poor concentration, they suffer a lot from it, and it’s not their fault. They just don’t form a discrete population.

Meanwhile, Adderall works for people whether they “have” “ADHD” or not. It may work better for people with ADHD – a lot of them report an almost “magical” effect – but it works at least a little for most people. There is a vast literature trying to disprove this. Its main strategy is to show Adderall doesn’t enhance cognition in healthy people. Fine. But mostly it doesn’t enhance cognition in people with ADHD either. People aren’t using Adderall to get *smart*, they’re using it to focus. From Prescription stimulants in individuals with and without attention deficit hyperactivity disorder:

> It has never been established that the cognitive effects of stimulant drugs are central to their therapeutic utility. In fact, although ADHD medications are effective for the behavioral components of the disorder, little information exists concerning their effects on cognition…stimulant drugs do improve the ability (even without ADHD) to focus and pay attention.

I cannot tell you how much literature there is trying to convince you that Adderall will not help healthy people, nor how consistently college students disprove every word of it every finals season.

That makes “only give Adderall to people with ADHD” a moral judgment, not a medical one. Adderall doesn’t “cure” the “disease” of ADHD, at least not in the same way penicillin cures syphilis. Adderall will give everyone better concentration, and we’ve judged that it’s okay for people with terrible concentration to use it to overcome their handicap, but not okay for people with already-fine concentration to use it to become superhuman.

We could still have a principled definition of ADHD. It would be something like “People below the Nth percentile in ability to concentrate.” Instead, we use the DSM, which advises us to diagnose people with ADHD if they say they have at least five symptoms from a list. The list has things like “often has difficulty sustaining attention” and “often has difficulty organizing tasks”. How often? You know, *often*! And if you work as a Senior Regional Manipulator Of Tiny Numbers, you’re going to have attention problems a lot more “often” than the rest of us.

So the DSM criteria are kind of meaningless, but that’s fine, because people can just lie about them anyway.

There are whole websites for this: How To Convince Your Shrink You Have ADHD, How To Get Your Doctor To Prescribe You Adderall In Five Easy Steps, et cetera. But I can’t imagine most people need them. Just talk about all the times in your life that you had attention and concentration problems, and if your doctor asks you a more specific question (“Do you often lose things?”) you give the obvious right answer ("Wow, it's like you've known me my whole life!").

Aren't psychiatrists creepy wizards who can see through your deceptions? There *are* people like that. They're called forensicists, they have special training in dealing with patients who might be lying to them, and they tend to get brought in for things like evaluating a murderer pleading the insanity defense. They have a toolbox of fascinating and frequently hilarious techniques to ascertain the truth, and they're really good at their jobs.

But *me?* At best, I can have a vague suspicion you're not telling the truth. And how many patients genuinely in need of treatment do I want to risk accidentally rejecting just so I can be sure of thwarting you? A lot of 100% honest psychiatric patients' stories are pretty unbelievable, really, and I don't *want* to have to treat every patient like a convicted murderer. Unless you give me some specific reason to doubt you, I start with the assumption that you're telling the truth.

Think about how wasteful all of this is. We throw people in jail for using Adderall without a prescription. We expel them from colleges. We fight an expensive and bloody War on Drugs to prevent non-prescription-holders from getting Adderall. We create a system in which poor people need to stretch their limited resources to make it to a psychiatrist so they can be prescribed Adderall, in which people without health insurance can never get it at all, in which DEA agents occasionally bust down the doors of medical practices giving out Adderall illegally. All to preserve a sham in which psychiatrists ask their patients "Do you have ADD symptoms?" and the patients say "Oh, yeah, definitely," and then the psychiatrists give them Adderall. It's like adding twenty layers of super-reinforced concrete to a bunker with a wide-open front door.

(Also, if by some chance a psychiatrist *doesn’t* give a patient Adderall, that patient practically always goes to another psychiatrist, and that next psychiatrist does. Trust me, no matter how unsuitable a candidate you are, no matter how bad a liar you are, somewhere there is a psychiatrist who will give you Adderall. And by “somewhere”, I mean it will take you three tries, tops.)

Psychiatrists’ main response to this perverse and unwinnable system is to give people Adderall, but feel guilty about it. Somebody should do an anthropological study on this, but my preliminary observations:

Some people will lecture their patients on how Medication Can Never Address The Root Cause Of A Problem, and the patient will agree that Medication Can Never Address The Root Cause Of A Problem, and then the psychiatrist will give them Adderall and feel good about it.

Some people will discuss alternative options, like behavioral treatments, or non-stimulant medications, and the patient will come back in a month and say that the behavioral treatments didn’t work, and then the psychiatrist will give them Adderall and feel good about it.

Some people will give their patients a formal test where they have to answer questions like “I often have trouble concentrating – strongly disagree, disagree, neutral, agree, or strongly agree?” Then the patient will give whatever answers get them Adderall, the psychiatrist will add up all the answers and score the test and find that it means the patient needs Adderall, and then the psychiatrist will give the patient Adderall and feel good about it.

Some people will occasionally find some little issue with one patient’s story, deny them Adderall, and then ride out the moral high for weeks, feeling so virtuous that they can give the next few people Adderall and feel good about it.

Some people will demand multiple evaluation sessions, lots of laboratory tests, make a patient tell them their whole life story. And after learning that they had a bad relationship with their stepfather in 8th grade and still have sexual hangups over that time they ejaculated prematurely with Sally one time in freshman year, the psychiatrist will give the patient Adderall and feel good about it.

I have been guilty of all of these at one time or another. I still wrestle with these issues a lot. The latest step in my evolving position was reading Kelsey’s blog post about having ADHD and trying to get Adderall. Her doctor gave her a list of things she had to do before he would give her Adderall, and she – having ADHD – got distracted and never did any of them.

(by my calculations, that decreased Kelsey’s effectiveness by 20%, thus costing approximately 54 billion lives.)

So lately I’ve been trying to be smarter about all this. What about good old consequentialism? Most people will get some benefit from Adderall, but it’s a powerful drug with a lot of potential risks. Maybe I should figure out exactly how bad the risks are, and then I can figure out how bad people’s concentration problems would have to be for the risks to be outweighed by the benefits.

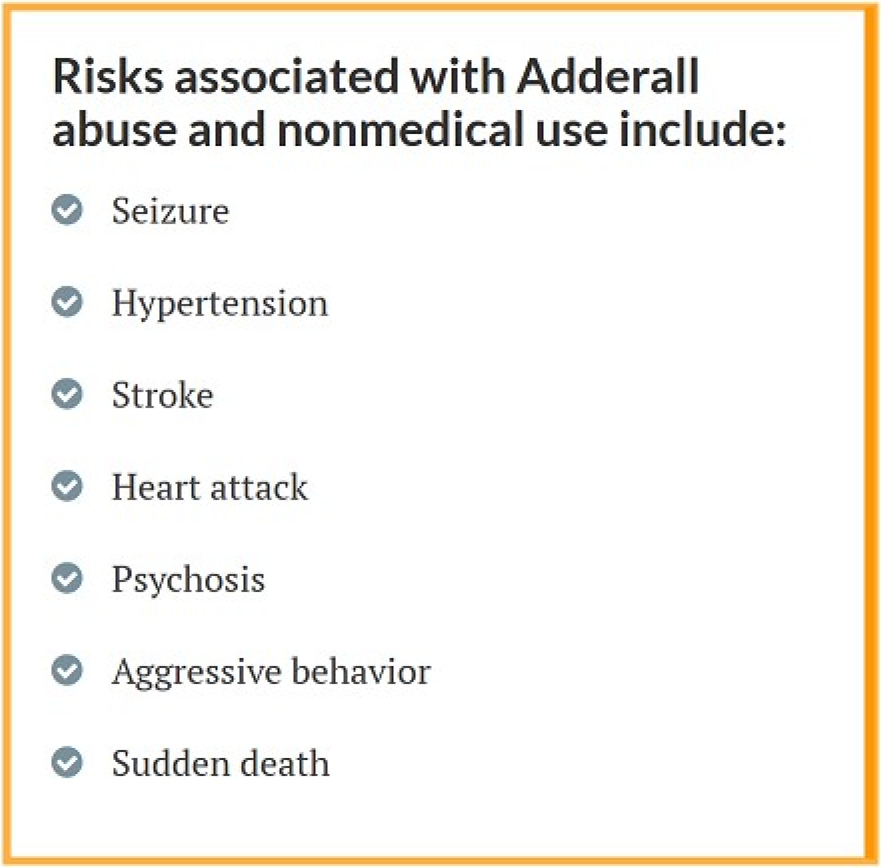
Trying to discover the risks of Adderall is a kind of ridiculous journey. It’s ridiculous because there are two equal and opposite agendas at work. The first agenda tries to scare college kids away from abusing Adderall as a study drug by emphasizing that it's terrifying and will definitely kill you. The second agenda tries to encourage parents to get their kids treated for ADHD by insisting Adderall is completely safe and anyone saying otherwise is an irresponsible fearmonger. The difference between these two situations is supposed to be whether you have a doctor's prescription. But what if you *are* the doctor, trying to decide who to prescribe it to? Then what? All they tell you in medical school is to give it to the people who actually have ADHD – which, I repeat, is kind of meaningless.

This post records my attempt to figure out something better. Apologies for the length.

II. Medical Risks

Most people on stimulants will have some minor side effects. Feeling jittery, feeling cold, feeling sick, leg cramps, arm cramps. Some will feel "like a robot" or otherwise psychologically uncomfortable. But these don't discourage me from giving stimulants to people who need them. If someone needs the drugs, let them try them, see how many side effects they get, and decide for themselves whether it's worth it.

I'm much more concerned about side effects that are permanent and dangerous. These people give us a list:



Risks associated with Adderall abuse and nonmedical use include:

- ✓ Seizure
- ✓ Hypertension
- ✓ Stroke
- ✓ Heart attack
- ✓ Psychosis
- ✓ Aggressive behavior
- ✓ Sudden death

Sounds pretty bad. On the other hand, I’ve prescribed Adderall to lots of people and none of them have ever gotten any of these things, except mild hypertension. How common are these, really?

The best source for exact numbers is the guidelines by sinister-sounding European organization EUNETHYDIS. I’ll use US medical database UpToDate as a secondary source. Both lump together Adderall and Ritalin – something I’ll be doing too throughout most of this essay, except where it becomes important to distinguish them.

Seizures: EUNETHYDIS doesn’t believe this happens at normal doses. They write:

> There are occasionally concerns that, as with other psychotropics, ADHD medications may lower the seizure

> threshold so as to cause seizures in previously seizure-free individuals. However, in prospective trials, retrospective cohort studies and post-marketing surveillance in ADHD patients without epilepsies, the incidence of seizures did not differ between ADHD pharmacotherapy and placebo [relative risk (RR)] for current versus non-use for methylphenidate, 0.8; RR for atomoxetine, 1.1

UpToDate is so unimpressed by this that they don't even mention it. If you ask them about seizure risk for ADHD medications, they start telling you about *bupropion*. Overall I wouldn't give these medications to people with a known seizure disorder without a neurologist's approval, but they seem pretty okay otherwise.

Hypertension: Broad agreement from both sources that stimulants cause hypertension. EUNETHYDIS says 1-4 mm systolic, UpToDate says 3-8 mm.

The main problem with hypertension is that it increases risk for things like heart attacks. I calculated an average 40 year old's risk of heart attack and got 1% over 10 years. Adding on an average Adderall-related increase in blood pressure, I got 1.1%.

What about in high-risk adults? I calculated risk for a 60 year old smoker with high cholesterol and high blood pressure. He has a 30.5% base risk of heart attack. Then I added in a typical Adderall-related rise in blood pressure, and he ended up at 32.0%. So Adderall only increased risk by about 1/1000 per year, even in this worst case scenario. Also, I never meet 60 year old smokers asking for Adderall. Overall this seems not too interesting.

I haven’t looked into other hypertension-related problems like kidney disease as much, but these seem like things you’ll hopefully have a lot of warning about and be able to talk to your doctor about whether to stop stimulants over.

Heart Attack and Stroke: My usual sources fail me here, but BioMed Central Cardiovascular Disorders comes to the rescue. They review three major studies on stroke and heart attack in stimulant patients.

Study #1 finds that stimulant users have 3x more risk of transient ischaemic attack (a small mini-stroke that does no lasting damage), but no increased risk of stroke.

Study #2 is the best and biggest study, and finds that stimulants actually *reduce* heart attack and stroke. They suspected “healthy-user bias”; that is, only healthy people would use such a supposedly-dangerous medication.

Study #3 is the most recent, and found no increased risk of heart attack or stroke.

UpToDate writes:

> Patients receiving stimulant therapy visited the emergency department or clinician office more frequently than those who were not treated with medications because of cardiac symptoms (10.9 versus 9.1 events per 1000 patient-years, adjusted hazards ratio 1.2, 95% CI 1.04-1.38) [26]. The cardiac symptoms included syncope, tachycardia, or palpitations. However, the group that received stimulant therapy was more likely to receive other psychotropic medications (antidepressants and antipsychotic agents), be male, and be non-Hispanic. The incidence of fatal and serious cardiac abnormalities was low and not different between

> the two groups, and was similar to the rates seen in the general pediatric population.

The 1/1000 extra ER visit per patient year sounds bad, but “palpitations” means “your heart feels like it’s beating in a weird way”, and Adderall clearly causes this, so my guess is this is mostly just people feeling this and freaking out. I have had patients call *me* after feeling this and freaking out, and we dealt with it, and they were fine. If I hadn’t been available, maybe they would have gone to the ER and turned themselves into a statistic.

There might be some bias in these studies, but overall there doesn’t seem to be much evidence this is worth worrying about unless your risk of heart attack or stroke is already really high.

Psychosis: I saw this a lot when I worked in inpatient. Somebody would take five times the recommended dose, or take more Adderall every time they felt tired until they hadn’t slept for a week, and then they would start hearing voices or feeling like something was crawling on their skin. After a day or two off Adderall, and a night or two getting a normal amount of sleep, they’d be fine. Take enough stimulants and you *will* become psychotic – but it’s rare on prescribed doses, and it usually resolves pretty quickly.

What dose can cause psychosis? Amphetamine-Induced Psychosis says:

> Early studies demonstrated that amphetamines could trigger acute psychosis in healthy subjects. In these studies, amphetamine was given in consecutively higher doses until psychosis was precipitated, often after 100–300 mg of amphetamine. The symptoms subsided within 6 days.

Compare this to the standard daily dose of Adderall of about 10 – 60 mg.

Can psychosis ever happen at normal doses? EUNETHYDIS is skeptical. They write:

> Data from population-based birth cohorts indicate that self-reported psychotic symptoms are common and may occur in up to 10% of 11-year-old children. In contrast, the prevalence of psychotic symptoms in children treated with ADHD drugs from RCTs is reported as only 0.19%. While this very low observed event rate in trials is likely to reflect a lack of systematic assessment and reporting, there is no compelling evidence to suggest that the observed event rate of psychotic symptoms in children treated with ADHD drugs exceeds the expected (background) rate in the general population. In the US FDA analysis, ADHD drug overdoses did not contribute significantly to reports of psychosis adverse events.

So basically, “kids are always kind of weird, studies say kids aren’t weird on Adderall, clearly they’re not paying attention, but it doesn’t look like things got any worse.”

UpToDate links these people, who say:

> We analyzed data from 49 randomized, controlled clinical trials in the pediatric development programs for these products. A total of 11 psychosis/mania adverse events occurred during 743 person-years of double-blind treatment with these drugs, and no comparable adverse events occurred in a total of 420 person-years of placebo exposure in the same trials. The rate per 100 person-years in the pooled active drug group was 1.48. The analysis of spontaneous postmarketing reports yielded >800 reports of adverse events related to

> psychosis or mania. In approximately 90% of the cases, there was no reported history of a similar psychiatric condition. Hallucinations involving visual and/or tactile sensations of insects, snakes, or worms were common in cases in children.

I think their use of “psychotic events per person-year” is misleading. Their study includes 5717 people, which means that for them to have 743 person-years each person must have been monitored for two months or so. But if you’re going to get psychotic on stimulants, usually it’s right after the stimulant is started. That means it might be better framed as “11/5717 patients had a psychotic event”, or even “one in every five hundred patients had a psychotic event”. Note that this matches the 0.19% number given by EUNETHYDIS. And the most common psychotic event was a feeling of snakes or insects on the skin which resolved after the drug was stopped, so we’re not talking “person is forever schizophrenic” here.

Also, I feel like EUNETHYDIS makes a good point with the “kids are always weird” thing. Here’s one of the psychotic events mentioned in the paper:

> A spontaneous report from the manufacturer of Strattera (atomoxetine) described a 7-year-old girl who received 18 mg daily of atomoxetine for the treatment of ADHD. Within hours of taking the first dose, the patient started talking nonstop and stated that she was happy. The next morning the child was still elated. Two hours after taking her second dose of atomoxetine, the patient started running very fast, stopped suddenly, and fell to the ground. The patient said she had “run into a wall” (there was no wall there). The reporting physician considered that the child was hallucinating. Atomoxetine was discontinued.

Have these people ever *seen* a child?

The methylphenidate prescribing information suggests an 0.1% risk of psychosis, which matches the other two studies pretty well.

Does stimulant psychosis always get better after the stimulant is discontinued? My strong impression is “yes”, but I am told that this study claims 5% to 15% of stimulant psychosis patients do not recover. I cannot find the full text to figure out exactly what they mean, and it looks like it was done on chronic meth addicts rather than prescription users.

So a few lines of evidence converge on 0.1% – 0.2% of children who use prescription stimulants become psychotic. I don’t know numbers for adults, but a few people who have read drafts of this article mention they have personally seen someone get psychotic on Adderall, which seems anecdotally to argue for a higher rate. I don’t know if those people were using it correctly or using anything else alongside it. Of people who get psychotic on Adderall, perhaps 5-15% stay psychotic after discontinuation (I predict this is about meth-heads and exaggerated).

Aggressive Behavior: This is just going to be the same as psychosis. Adderall isn’t going to magically turn gentle old grandmothers into killing machines. If you’re already a kind of violent guy, and you take a lot of Adderall, maybe it’ll push you over the edge.

Sudden Death: This is usually cardiovascular – something goes very wrong with your heart and it stops beating without warning. But UpToDate writes:

> Reports of unexpected deaths of children receiving stimulant therapy have led to concerns that these medications increase the risk of cardiovascular (CV)

> adverse events, including sudden unexpected deaths (SUD) [1,2]. However, large cohort studies have not shown an increased risk of serious CV adverse events in children treated with stimulant therapy compared with the general pediatric population…
>
> Among adult patients who are either current or new users of stimulant medications, there appears to be no increased risk of serious CV events. This was illustrated in a large retrospective cohort study of adults (age range 25 to 64 years) based on data from four large health plans that was done in parallel with the study performed in children discussed above [3,16]. Multivariant analysis demonstrated a lower risk of serious CV events (defined as myocardial infarction, stroke, and sudden cardiac death) in individuals who were current users of stimulant therapy versus nonusers (relative risk [RR] 0.83, 95% CI 0.72-0.96). In new users of ADHD medications compared with controls, the risk of serious CV events was even lower (RR 0.77, 95% CI 0.63-0.94). However, there may be a modest amount of healthy-user bias that favored the current users of stimulant therapy. To adjust for this potential bias, a multivariant analysis that compared current users with individuals who had used stimulant therapy more than one year ago (defined as remote use) found no difference in the risk of serious CV events (RR 1.03, 95% CI 0.86-1.24). The crude incidence of serious CV events in the overall cohort was 1.34 per 1000 person-years. These results showing no increased risk of serious CV events are consistent with previously discussed studies in pediatric patients.

And EUNETHYDIS:

> when the number of patient-years of prescribed medication was incorporated into the evaluation, the frequency of reported sudden death per year of ADHD therapy with methylphenidate, atomoxetine or amfetamines among children was 0.2–0.5/100,000 patient-years [99]. The analysis of 10-year adverse-event reporting in Denmark resulted in no sudden deaths in children taking ADHD medications [5]. While it is recognised that adverse events are frequently under-reported in general, it is likely that sudden deaths in young individuals on relatively new medications may be better reported. Death rates per year of therapy, calculated using the adverse events reporting system (AERS) reports and prescription data, are equivalent for two ADHD drugs (dexamfetamine and methylphenidate): 0.6/100,000/year [37]. (The accuracy of these estimates is limited however, for instance because in moving from number of prescriptions to patient-year figures assumptions must be made about the length of each prescription). It seems likely, using these best available data, and assuming a 50% under-reporting rate, that the sudden death risk of children on ADHD medications is similar to that of children in general.

Despite this, I am always very wary prescribing stimulants to anyone with any history of heart problems. I always make these people go see a cardiologist. The cardiologist always says yeah, sure, whatever, but it makes me feel a lot better.

In General: Probably the most informative passage I've seen on the medical risks of stimulants is this one from Misuse Of Study Drugs:

> In 1990, there were about 271 emergency room reports involving methylphenidate, 1,727 in 1998, and 1,478 in

> 2001 [32]. The total number of emergency department visits resulting from use of all psychotherapeutic CNS stimulants was 4091 in 1998, 3644 in 1999, 3336, in 2000, 3146 in 2001 and 3275 in 2002 [33]. There are approximately 25 emergency room deaths per year among up to 3 million users of prescription stimulant drugs (including both those medically prescribed and not prescribed these drugs). Thus, the likelihood of dying from such drugs appears to be approximately 1 in 120,000.

But isn’t 25 deaths per year still bad?

Here’s another passage from the same source:

> Intravenous use of prescription stimulants is particularly dangerous. In particular, intravenous (IV) abuse of methylphenidate may result in talcosis. Talcosis is a reaction to talc, a filler and lubricant in methylphenidate and other oral medication. This inflammation reaction occurs in the lungs and related consequences include lower lobe panacinar emphysema.

People aren’t dying because their psychiatrist gave them Adderall 10 mg bid. They’re dying because they ground it up, injected it into their bloodstream, and had their lungs turn into talc. The people dying of stimulant use are doing things so horrifying you could not possibly imagine them even if you took ten times your prescribed dose of Adderall and used all of it to focus on writing a report on the most horrifying ways you could possibly use Adderall. Did you know that 13% of Massachusetts college students have ground up Ritalin and snorted it up their nose? Did you know the first case report of Ritalin abuse involved a patient who was taking 125 Ritalin pills daily? All of these people are out there, and *still* only 25 people die of stimulant-related causes per year!

My impression is that, in particularly at-risk people, stimulants may add +1/1000 to the risk of heart attacks per year, and +1/10,000 risk of long-term psychosis. Everything else in this category can be rounded down to zero.

III. Addiction

What about addiction risk?

The data on this are really poor because it's hard to define addiction. If a prescription stimulant user uses their stimulants every day, and feels really good on them, and feels really upset if they can't get them…well, that's basically the expected outcome.

Wilens et al finds that over ten years, 10% of adolescents surveyed got high on their medication, and 22% sometimes used more than prescribed. Does that mean those 10% or 22% are "addicted"? Not really – some of them probably have a tough day one time, so they take two Adderall that day and no Adderall the day after. As for getting high – well, a lot of people get high on alcohol who aren't alcoholics, and a lot of people get stoned who nobody would call addicted to marijuana.

A lot of studies in this area ask the kind of different question of whether children put on stimulants are more likely to be addicted to drugs in general as adults. Most of them find these children are *less* likely, which is hypothesized to be an effect of successfully treating their ADHD.

And there's a book on narcolepsy which apparently claims that between less than 1% and 3% of people taking stimulants for that condition get addicted, but I can't track down their methodology or really anything beyond one reference. And narcoleptics are a different population than ADHD patients and results might not generalize (though that number sounds kind of right).

I don’t think there are good data here, but my intuitions and personal experience is that “addiction” of the sort you get with heroin or tobacco is very rare, at least when responsible people without a personal or family history of addictive behavior take stimulants as prescribed. Most people agree the risk is lower for extended-release stimulants (eg Adderall XR), and very low for Vyvanse.

IV. Tolerance

Tolerance is when you keep needing more and more of a drug to get an effect. In the worst cases, your baseline changes so that you need the drug to feel normal. The concern is that long-term use of Adderall will make your attention naturally worse, so that medicated-you is only as good at concentrating as unmedicated-you was before, and unmedicated-you is even less attentive.

We know tolerance occurs over the short-term, and we encourage patients to take a few days off Adderall every week or two to let their bodies reset. More concerning is whether it happens over the space of years, where people’s bodies adjust in a more permanent way.

The best study of this phenomenon was the Multimodal Treatment of ADHD (MTA) study, which randomized children to be treated with stimulants or “behavioral therapy” (eg learning coping skills, etc). Behavioral therapy for ADHD is not very good and I interpret it as a nice way of saying placebo.

For the first year, the kids getting stimulants did much better on all metrics than behavioral-therapy-only. For the second year, they did a *little* better. By the third year, they were the same. In the eighth year, which was as long as anyone kept checking, they were still the same.

This is pretty concerning. It sounds like over three years people's bodies built up some tolerance to stimulants, after which they provided no further benefit. The only saving grace is that there's no evidence of stimulants ever making people worse than normal (even on people who stopped the medications later).

People have critiqued this study on the grounds that although they started off giving the experimental group stimulants vs. the control group behavioral therapy, any patient could switch treatments at any time and many of them did. By year three when the groups equalized, only 66% of the medication group was on medication, and a full 43% of the therapy-only group was. So maybe this just drowned out any original effect?

The authors of the study are not convinced:

> It is tempting to conclude that intensive medication management beyond 14-months could have resulted in continued differences between the randomly assigned treatment groups…In a previous multimodal treatment study where medication was carefully titrated and monitored for two years, treatment gains were maintained for the entire period. However, after 14 months the MTA became an uncontrolled naturalistic follow-up study and inferences about potential advantages that might have occurred with continued long-term study-provided treatment are speculation. Moreover, with one exception (math achievement),

> children still taking medication by 6 and 8 years fared no better than their non-medicated counterparts despite a 41% increase in the average total daily dose, failing to support continued medication treatment as salutary (at least, continued medication treatment as monitored by community practitioners)…Finally, a previous analysis of the MTA data through 3 years did not provide evidence that subject selection biases towards medication use in the follow-up period accounted for the observed lack of differential treatment effects.
>
> Thus, although the MTA data provided strong support for the acute reduction of symptoms with intensive medication management, these long-term follow-up data fail to provide support for long-term advantage of medication treatment beyond two years for the majority of children—at least as medication is monitored in community settings.

As far as I can tell, pretty much everyone has ignored this, using the usual range of meaningless excuses like “Well, treatment must be individualized to the patient”.

This is very tempting, because for example I have a lot of patients who have been on stimulants for decades, are still very excited about them, and think they’re doing great. Every so often these patients go off their stimulants, are very unhappy, and insist on going back on them again. They say that pre-stimulant, they were scatterbrained and always losing things and missing appointments and failing to do work, and now, after ten years of stimulant treatment, they feel great.

We can imagine ways these people are wrong. Maybe the stimulants worked for the first three years, stopped working so gradually they didn’t notice, and now they only notice the difference between being on stimulants (baseline), and immediate post-stimulant withdrawal (very bad). But this would require a lot of people to be really wrong about their internal experience.

I asked a question on the Slate Star Codex survey about this. People on Adderall more than one month were asked to tell me whether they had no tolerance problems, some tolerance requiring dose escalation, or high tolerance that made the medications stop working entirely. The preliminary results:

Adderall for between one month and one year: (n = 124)
62 (50%) No tolerance, worked as well as ever
57 (46%) Some tolerance, or required dose escalation, but still worked well in general
5 (4%) High tolerance, stopped working

Adderall for one to five years: (n = 117)
33 (28%) No tolerance, worked as well as ever
78 (67%) Some tolerance, or required dose escalation, but still worked well in general
6 (5%) High tolerance, stopped working

Adderall for more than five years: (n = 59)
23 (39%) No tolerance, worked as well as ever
33 (56%) Some tolerance, or required dose escalation, but still worked well in general
3 (5%) High tolerance, stopped working

All three categories were evenly divided between “no tolerance” and “some tolerance but still worked well”, with only about 5% saying the tolerance became a big problem. This matches my clinical experience. So either I’m right, or the problem where they get confused and forget their baseline is affecting my survey-takers.

There are occasional claims that magnesium or some other substance can help reverse Adderall tolerance. As far as I know these have never really been investigated.

So: there's no good evidence that taking Adderall will actively make your ADHD worse in the long run. There is good evidence from clinical trials that benefits will decrease to zero over the space of a few years, apparently contradicted by the personal experiences of doctors and patients. Overall not sure what to do with this one.

V. Neurotoxicity

There's some evidence that amphetamines can cause permanent cellular damage, but it's not clear whether this happens in humans at typical therapeutic doses.

If you give rats very high doses of IV amphetamines, they accumulate so much dopamine in the cytoplasm of their neurons that it causes oxidative stress and destroys dopaminergic nerve terminals. This doesn't happen to rats at doses matching human doses of Adderall. But it does happen at those doses to squirrel monkeys. At least this is the claim:

> Adult baboons and squirrel monkeys were treated with a 3:1 mixture of D/L–amphetamine similar to the pharmaceutical Adderall for 4 weeks. Plasma concentrations of amphetamine (136±21 ng/ml-1) matched the levels reported in human ADHD patients after amphetamine treatment lasting 3 weeks (120–140 ng/ml-1) or 6 weeks in the highest dose (30 mg/day-1) condition (120 ng/ml-1). When the animals were killed 2 weeks after the 4-week amphetamine treatment period, both non-human primate species showed a 30–50% reduction in striatal dopamine, its major metabolite

> (dihydroxyphenylacetic acid (DOPAC)), its rate-limiting enzyme (tyrosine hydroxylase), its membrane transporter and its vesicular transporter. These consequences are similar, if not identical to the effects of neurotoxic doses in rodents.

I'm not really sure what they're getting at here – surely they're not saying just one month of Adderall permanently decreases striatal dopamine by 50%? But it sounds like something bad is happening, and since humans are more like monkeys than rats, maybe there's cause for concern.

What would it look like if people got this kind of brain damage? One likely possibility is Parkinson's disease, a condition caused by poor dopaminergic function in the brain. If you were going to tell a story about how Adderall could cause long-term neurotoxic damage, it would look like gradual decrease of brain dopaminergic function without obvious effects through most of the lifespan (since most people have dopaminergic function to spare). As the patient got older and started naturally losing brain function, Parkinson's would appear. This happens to genetically and environmentally predisposed people anyway (which is why old people get Parkinson's so often), but in this scenario amphetamine use would present an extra risk factor.

Several studies have shown that meth addicts do have higher rates of Parkinson's disease. This one says people hospitalized for meth addiction are 60% more likely to get Parkinson's than people hospitalized for other reasons. This one finds Parkinson's rates three times higher in meth addicts compared to non-drug-users.

What about at therapeutic doses? This article claims there was a study that found people who used Benzedrine and Dexadrine (early forms of prescription amphetamine) in the

1960s have rates of Parkinson's Disease about 60% higher than non-users today, but I can't find the study itself and I don't know the methodology. Another study finds similar results. Since both ADHD and stimulant addiction are very hereditary, you could make an argument that people who already have problems with their dopamine system are more likely to get Parkinson's later on. There's a little bit of conflicting evidence for this. Also, ADHD patients might have three times the rate of dementia with Lewy bodies, a condition closely related to Parkinson's. On the other hand, there doesn't seem to be any genetic connection. Overall my guess is this is not what's going on.

About 1-2% of people will get Parkinson's if they live long enough. If Adderall increases that risk 60%, then presumably it could cause a 1% absolute increase in risk.

Some people claim various substances (magnesium, minocycline, etc) will protect your brain from amphetamine neurotoxicity. None of these have been studied in anywhere near the depth they would need to be to make me feel comfortable with this.

~~The good news is that as far as anyone can tell, Ritalin doesn't cause these problems, even if you give it to rats at super-high doses. It seems to be a difference in the mechanism of action. I've been talking about Adderall this whole post because it's the most commonly-used stimulant and some studies have suggested it's more effective for a few people, but this might be a strong argument in favor of starting with Ritalin and only switching to Adderall if Ritalin fails.~~ [EDIT: Never mind, recent studies suggest Ritalin is just as likely to cause this problem.]

So overall there is plausible, but not incontrovertible, evidence linking Adderall to a somewhat increased risk of

Parkinson’s disease in old age.

VI. Summary

My impression is that the risks of *proper, medically supervised* Adderall use are the following:

1. High risk of minor short-term side effects that might make you want to stop taking the medication with no long-term issues
2. Extremely low risk of serious medical side effects like stroke or heart attack, except maybe in a few very vulnerable populations
3. Maybe one percent risk, but not literally zero risk, of addiction if patients are well-targeted by their doctors and use the medication responsibly.
4. Perhaps one in five hundred risk, but not literally zero risk, of psychosis. Some anecdotal evidence suggests it is more common than this. Most of these cases will be mild and resolve quickly. Some people find a very small number of cases of stimulant-induced psychosis may be permanent, though I still find this hard to believe.
5. Some evidence for tolerance after several years, though most patients will continue to believe it is helping them. No sign of supertolerance where it actually makes the condition worse.
6. Plausibly 60% increased relative risk (+~1% absolute risk) for Parkinson’s disease with long-term use.
7. Unknown unknowns.

Of these, I find the psychosis, tolerance, and Parkinson’s to be the most concerning. But I am pretty upset about the overall terrible state of this research. In particular, nobody except the MTA takes the possibility of tolerance seriously, and the MTA results really ought to have inspired a lot more soul-searching and hand-wringing than they actually did.

The numbers on addiction and psychosis are inexcusably terrible given how easy they would be to collect. Getting good data on the Parkinson's risk would be harder, but one so-far-unexplored possibility would be to compare past prescription Adderall history to past prescription Ritalin history in Parkinson's patients to adjust for the potential ADHD confounder. I really think somebody should do this.

Despite all this, I compare these risks to the risks of eating one extra strip of bacon per day and decide that overall this is not enough for me to stop prescribing stimulants to patients who I think might benefit from them. These are about the standard level of side effects for a powerful medication and I think there's a major role for these in ADHD treatment as long as patients are well-informed about the risks they're taking.

PS: I don't accept blog readers as patients, and I won't prescribe you Adderall just because you liked this post.

2018

- 2017 Predictions: Calibration Results
- SSC Survey Results 2018
- Book Review: Madness And Civilization
- Fight Me, Psychologists: Birth Order Effects Exist And Are Very Strong
- Self-Serving Bias
- Maybe The Real Superintelligent AI Is Extremely Smart Computers
- Bundles Of Joy
- Practically-A-Book Review: Luna Whitepaper
- Conflict Vs. Mistake
- SSC Survey Data On Models Of Political Conflict
- Highlights From The Comments On Conflict Vs. Mistake
- The Invention Of Moral Narrative
- Powerless Placebos
- Predictions For 2018
- Guyenet On Motivation
- We’ve Got Five Years, What A Surprise
- More Testimonials For SSC
- Even More Search Terms That Led People To This Blog
- Five More Years
- Technological Unemployment: Much More Than You Wanted To Know

- **Current Affairs’ “Some Puzzles For Libertarians”, Treated As Writing Prompts For Short Stories**
- **Highlights From The Comments On Technological Unemployment**
- **SSC Journal Club: Cipriani On Antidepressants**
- **SSC Meetup: Bay Area 3/3**
- **God Help Us, Let’s Try To Understand Friston On Free Energy**
- **SSC Journal Club: Friston On Computational Mood**
- **Response To Current Affairs On Against Murderism**
- **The Dark Rule Utilitarian Argument For Science Piracy**
- **Navigating And/Or Avoiding The Inpatient Mental Health System**
- **Book Review: Twelve Rules For Life**
- **The Hour I First Believed**
- **Are The Amish Unhappy? Super Happy? Just Meh?**
- **Highlights From The Comments On Twelve Rules**
- **Adult Neurogenesis – A Pointed Review**
- **Why DC’s Low Graduation Rates?**
- **Highlights From The Comments On DC Graduation Rates**
- **Recommendations vs. Guidelines**
- **SSC Survey Results: Sexual Harassment Levels By Field**

- [Highlights From The Comments On Survey Harassment Rates]
- [Gupta On Enlightenment]
- [Mental Health On A Budget]
- [Call For Adversarial Collaborations]
- [Adversarial Collaboration Contest: Loose Ends And Registration]
- [Book Review: History Of The Fabian Society]
- [Classified Thread 5: Classified Never Sinned]
- [Varieties Of Argumentative Experience]
- [Basic Income, Not Basic Jobs: Against Hijacking Utopia]
- [Can Things Be Both Popular And Silenced?]
- [Should Psychiatry Test For Lead More?]
- [Highlights From The Comments On Basic Jobs]
- [In Search Of Missing US Suicides]
- [HPPD And The Specter Of Permanent Side Effects]
- [The GATTACA Trilogy]
- [Contra Caplan On Arbitrary Deploring]
- [Cost Disease In Medicine: The Practical Perspective]
- [Book Review: Capital In The Twenty-First Century]
- [List Of Passages I Highlighted In My Copy Of Capital In The Twenty-First Century]
- [Highlights From The Comments On Piketty]
- [SSC Journal Club: Dissolving The Fermi Paradox]
- [The Craft And The Codex]

- Melatonin: Much More Than You Wanted To Know
- Did A Melatonin Patent Inspire Current Dose Confusion?
- Fundamental Value Differences Are Not That Fundamental
- The Whole City Is Center
- Sentimental Cartography
- Value Differences As Differently Crystallized Metaphysical Heuristics
- Verses Composed Upon Reading A Review From TripAdvisor
- Meetups Everywhere 2018
- Cancer Progress: Much More Than You Wanted To Know
- Before You Get Too Excited About That Trigger Warning Study…
- SSC Meetups 2018: Times And Places
- SSC Survey Results: ADHD And Rejection Sensitivity
- SSC Survey: Scattered Negative Results
- The Parentheses Riddle
- Practically-A-Book Review: EA Hotel
- Carbon Dioxide: An Open Door Policy
- Elegy For John McCain
- Bulls**t Jobs (Part 1 of ∞)
- Bureaucracy As Active Ingredient
- This Week: Adversarial Collaboration Entries
- Reminder: Bay Area SSC Meetup 9/8
- [ACC Entry] Does The Education System Adequately Serve Advanced Students?

- [ACC Entry] Are Islam And Liberal Democracy Compatible?
- [ACC Entry] Should Childhood Vaccination Be Mandatory?
- [ACC Entry] Should Transgender Children Transition?
- In The Balance
- The Omnigenic Model As Metaphor For Life
- Book Review: The Black Swan
- Treating The Prodrome
- The Tails Coming Apart As Metaphor For Life
- Adversarial Collaboration Contest Results
- Steelmanning The NIMBYs
- Highlights From The Comments On NIMBYs
- Next Door In Nodrumia
- Nighttime Ventilation Survey Results
- Kavanaugh: A Probability Poll
- Anxiety Sampler Kits
- Is Science Slowing Down?
- The Chamber of Guf
- Classified Thread 6
- Cognitive Enhancers: Mechanisms And Tradeoffs
- Nominating Oneself For The Short End Of A Tradeoff
- Sort By Controversial
- Working With Google Trends
- My California Ballot
- Preschool: I Was Wrong
- Marijuana: An Update
- SSRIs: An Update
- Ketamine: An Update

- **Preschool: Much More Than You Wanted To Know**
- **The Economic Perspective On Moral Standards**
- **Impending Survey Discussion Thread**
- **Is Science Slowing Down?**
- **Book Review: The Mind Illuminated**
- **Book Review: Evolutionary Psychopathology**
- **Del Giudice On The Self-Starvation Cycle**
- **Diametrical Model Of Autism And Schizophrenia**
- **Trump: A Setback For Trumpism**
- **Fallacies Of Reversed Moderation**
- **Refactoring: Culture As Branch Of Government**
- **Seattle Meetup This Sunday**
- **Please Take The 2019 SSC Survey**

2017 Predictions: Calibration Results

At the beginning of every year, I make predictions. At the end of every year, I score them. Here are 2014, 2015, and 2016.

And here are the predictions I made for 2017. Strikethrough'd are false. Intact are true. Italicized are getting thrown out because I can't decide if they're true or not.

WORLD EVENTS
1. US will not get involved in any new major war with death toll of > 100 US soldiers: 60%
2. North Korea's government will survive the year without large civil war/revolt: 95%
3. No terrorist attack in the USA will kill > 100 people: 90%
4. …in any First World country: 80%
5. Assad will remain President of Syria: 80%
6. Israel will not get in a large-scale war (ie >100 Israeli deaths) with any Arab state: 90%
7. No major intifada in Israel this year (ie > 250 Israeli deaths, but not in Cast Lead style war): 80%
8. No interesting progress with Gaza or peace negotiations in general this year: 90%
9. No Cast Lead style bombing/invasion of Gaza this year: 90%
10. Situation in Israel looks more worse than better: 70%
~~11. Syria's civil war will not end this year: 60%~~
12. ISIS will control less territory than it does right now: 90%
13. ISIS will not continue to exist as a state entity in Iraq/Syria: 50%
14. No major civil war in Middle Eastern country not currently experiencing a major civil war: 90%
15. Libya to remain a mess: 80%

16. Ukraine will neither break into all-out war or get neatly resolved: 80%
17. No major revolt (greater than or equal to Tiananmen Square) against Chinese Communist Party: 95%
18. No major war in Asia (with >100 Chinese, Japanese, South Korean, and American deaths combined) over tiny stupid islands: 99%
19. No exchange of fire over tiny stupid islands: 90%
20. No announcement of genetically engineered human baby or credible plan for such: 90%
~~21. EMDrive is launched into space and testing is successfully begun: 70%~~
22. A significant number of skeptics will not become convinced EMDrive works: 80%
23. A significant number of believers will not become convinced EMDrive doesn't work: 60%
24. No major earthquake (>100 deaths) in US: 99%
25. No major earthquake (>10000 deaths) in the world: 60%
~~26. Keith Ellison chosen as new DNC chair: 70%~~

EUROPE

27. No country currently in Euro or EU announces new plan to leave: 80%
28. France does not declare plan to leave EU: 95%
29. Germany does not declare plan to leave EU: 99%
30. No agreement reached on "two-speed EU": 80%
31. The UK triggers Article 50: 90%
32. Marine Le Pen is not elected President of France: 60%
33. Angela Merkel is re-elected Chancellor of Germany: 60%
34. Theresa May remains PM of Britain: 80%
35. Fewer refugees admitted 2017 than 2016: 95%

ECONOMICS

36. Bitcoin will end the year higher than $1000: 60%
37. Oil will end the year higher than $50 a barrel: 60%
~~38. …but lower than $60 a barrel: 60%~~

39. Dow Jones will not fall > 10% this year: 50%
40. Shanghai index will not fall > 10% this year: 50%

TRUMP ADMINISTRATION
41. Donald Trump remains President at the end of 2017: 90%
42. No serious impeachment proceedings are active against Trump: 80%
~~43. Construction on Mexican border wall (beyond existing barriers) begins: 80%~~
44. Trump administration does not initiate extra prosecution of Hillary Clinton: 90%
~~45. US GDP growth lower than in 2016: 60%~~
~~46. US unemployment to be higher at end of year than beginning: 60%~~
47. US does not withdraw from large trade org like WTO or NAFTA: 90%
48. US does not publicly and explicitly disavow One China policy: 95%
49. No race riot killing > 5 people: 95%
~~50. US lifts at least half of existing sanctions on Russia: 70%~~
51. Donald Trump’s approval rating at the end of 2017 is lower than fifty percent: 80%
52. …lower than forty percent: 60%

COMMUNITIES
53. SSC will remain active: 95%
~~54. SSC will get fewer hits than in 2016: 60%~~
55. At least one SSC post > 100,000 hits: 70%
56. I will complete an LW/SSC survey: 80%
~~57. I will finish a long FAQ this year: 60%~~
58. Shireroth will remain active: 70%
59. No co-bloggers (with more than 5 posts) on SSC by the end of this year: 80%
60. Less Wrong renaissance attempt will seem less (rather than more) successful by end of this year: 90%

61. > 15,000 Twitter followers by end of this year: 80%
62. I won’t stop using Twitter, Tumblr, or Facebook: 90%
63. I will attend the Bay Area Solstice next year: 90%
64. …some other Solstice: 60%
~~65. …not the New York Solstice: 60%~~

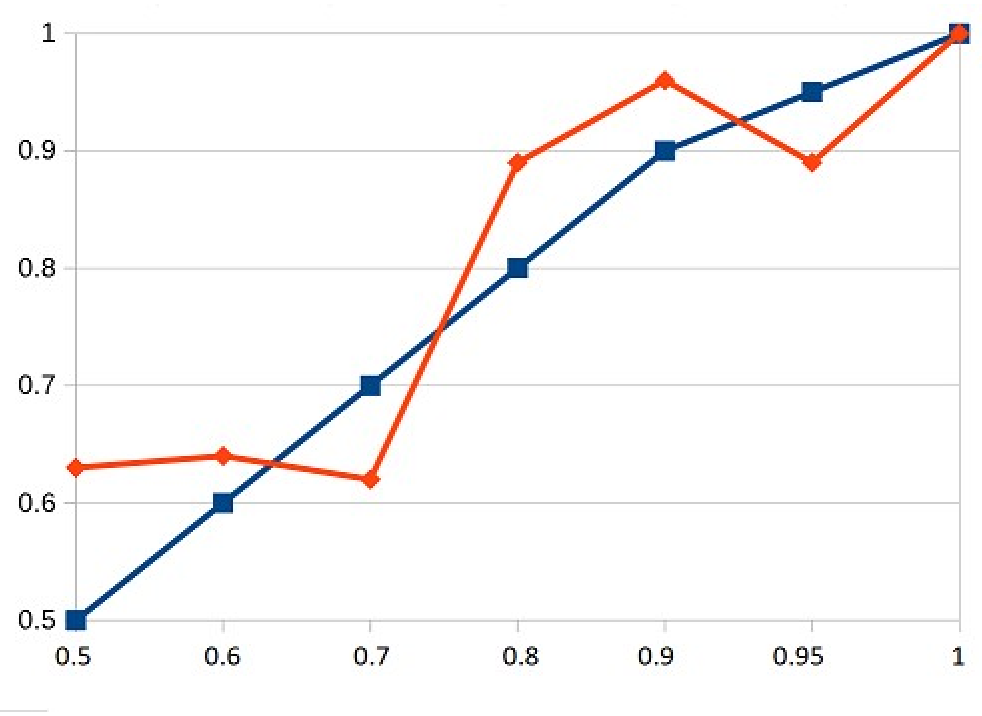
WORK
66. I will take the job I am currently expecting to take: 90%
~~67. …at the time I am expecting to take it, without any delays: 80%~~
68. I will like the job and plan to continue doing it for a while: 70%
69. I will pass my Boards: 90%
70. I will be involved in at least one published/accepted-to-publish research paper by the end of 2017: 50%
71. I will present a research paper at the regional conference: 80%
72. I will attend the APA national meeting in San Diego: 90%
~~73. None of my outpatients to be hospitalized for psychiatric reasons during the first half of 2017: 50%~~
74. None of my outpatients to be involuntarily committed to psych hospital by me during the first half of 2017: 70%
75. None of my outpatients to attempt suicide during the first half of 2017: 90%
76. I will not have scored 95th percentile or above when I get this year’s PRITE scores back: 60%

PERSONAL
77. Amazon will not harass me to get the $40,000 they gave me back: 80%
78. …or at least will not be successful: 90%
79. I will drive cross-country in 2017: 70%
~~80. I will travel outside the US in 2017: 70%~~
~~81. …to Europe: 50%~~
82. I will not officially break up with any of my current girlfriends: 60%

83. K will spend at least three months total in Michigan this year: 70%
84. I will get at least one new girlfriend: 70%
85. I will not get engaged: 90%
86. I will visit the Bay in May 2017: 60%
87. I will have moved to the Bay Area: 99%
88. I won't live in Godric's Hollow for at least two weeks continuous: 70%
89. I won't live in Volterra for at least two weeks continuous: 70%
~~90. I won't live in the Bailey for at least two weeks continuous: 95%~~
91. I won't live in some other rationalist group home for at least two weeks continuous: 90%
92. I will be living in a house (incl group house) and not apartment building at the end of 2017: 60%
93. I will still not have gotten my elective surgery: 90%
94. I will not have been hospitalized (excluding ER) for any other reason: 95%
~~95. I will make my savings target at the end of 2017: 60%~~
96. I will not be taking any nootropic (except ZMA) daily or near-daily during any 2-month period this year: 90%
97. I won't publicly and drastically change highest-level political/religious/philosophical positions (eg become a Muslim or Republican): 90%
98. I will not get drunk this year: 80%
99. I get at least one article published on a major site like Huffington Post or Vox or New Statesman or something: 50%
~~100. I attend at least one wedding this year: 50%~~
101. Still driving my current car at the end of 2017: 90%
102. Car is not stuck in shop for repairs for >1 day during 2017: 60%
103. I will use Lyft at least once in 2017: 60%
104. I weight > 185 pounds at the end of 2017: 60%
~~105. I weight < 195 pounds at the end of 2017: 70%~~

Some justifications for my decisions: I rated the civil war in Syria as basically over, even though Wikipedia says otherwise, since I don't think there are any remaining credible rebel forces, and ISIS is pretty dead. Trump's approval rating is taken from this 538 aggregator and is currently estimated at 38.1%. I rated the border wall as not currently under construction, despite articles with titles like The Trump Administration Has Already Started Building The Border Wall, because it was referring to a 30-foot prototype not likely to be included in the wall itself (have I mentioned the media is terrible?). I refused to judge the success of the Less Wrong renaissance attempt, because it seemed unsuccessful but was superseded by a separate much more serious attempt that was successful and I'm not sure how to rate that. I refused to judge whether or not I got a new partner because I am casually dating some people and not sure how to count it. I refused to judge whether I got 95th percentile+ on my PRITE because they stopped clearly reporting percentile scores.

This is the graph of my accuracy for this year:



Of 50% predictions, I got 5 right and 3 wrong, for a score of 62%
Of 60% predictions, I got 14 right and 8 wrong, for a score of 64%
Of 70% predictions, I got 8 right and 5 wrong, for a score of 62%
Of 80% predictions, I got 16 right and 2 wrong, for a score of 89%
Of 90% predictions, I got 24 right and 1 wrong, for a score of 96%
Of 95% predictions, I got 8 right and 1 wrong, for a score of 89%
Of 99% predictions, I got 4 right and 0 wrong, for a score of 100%

Blue is hypothetical perfect calibration, red is my calibration. The multiple crossings of the blue line indicate that I am neither globally overconfident or globally underconfident.

Last year my main concern was that I was underconfident at 70%. I tried to fix that this year by becoming more willing to guess at that level, and ended up a bit overconfident. This year I'll try somewhere in the middle and hopefully get it right.

There weren't enough questions to detect patterns of mistakes, but there was a slight tendency for me to think things would go more smoothly than they did. I overestimated the success of my diet, my savings plan, my travel plans my job start date, my long-FAQ-making ability, and my future housing search (this last one led to me spending a few weeks at a friend's group house, failing on a 95% certainty prediction). I only made one error in favor of personal affairs going better than expected (SSC got more hits than last year; maybe this isn't a central example of "personal affairs going smoothly"). None of these really caused me any problems, suggesting that I have enough slack in my plans, but apparently I'm not yet able to extend that to being able to make good explicit predictions about.

My other major error was underestimating the state of the US economy, leading to a couple of correlated errors. I think I got Trump mostly right, although I may have overestimated his efficacy (I thought he would have started the border wall by now) and erred in thinking he would lift sanctions on Russia.

Otherwise this is consistent with generally good calibration plus random noise. Next year I'll have played this game five years in a row, and I'll average out all my answers for all five years and get a better estimate; for now I'll just be pretty satisfied.

Predictions for 2018 coming soon.

SSC Survey Results 2018

Thanks to the 8,077 people (!) who took this year's SSC survey.

I don't have the energy to screenshot/copy/paste the graph for every single question the way I have in previous years, so let's do it differently.

The survey page is changed so that you can just press "okay" and "submit", and it will bring you to the results page and see all the results. I'm not sure you *can* take the whole survey anymore, but if you find a way to do so, please don't. Just press "okay" and "submit" and you should be fine. Don't worry, all identifying questions (including the identifier string and all long answers) have been hidden.

See the exact questions for the SSC survey.

See results from the SSC survey.

See results from the Mechanical Turk comparison survey.

(this might have a lot of lag if you try to do it at the same time as everyone else; if you tell your browser to stop scripts it might improve)

I plan to post longer analyses (including the ones in the pre-registered hypotheses) later on, hopefully dragging them out into a bunch of Least Publishable Units.

If you want to scoop me, or investigate the data yourself, you can download the answers of the 7298 people who agreed to have their responses shared publicly:

Main survey: .xlsx, .csv

Turk survey: .xlsx, .csv

Book Review: Madness And Civilization

[Content warning: Severe mistreatment of the mentally ill. Through this post, I'll be following Foucault in using the politically incorrect term "madness" rather than the more modern "mental illness", because a big part of his point is worrying about the assumptions contained in the latter term.]

I.

I started reading Foucault's *Madness And Civilization* with the expectation that it would be tedious and incomprehensible. You know, the stereotype that postmodernism / post-structuralism / Continentalism / etc. involves a lot of negation of the negation of the inversion of the Other within the Absolute within [and so on for 200 pages]. There was a little of that. But there was also a fascinating look at the history of mental illness, an entertainingly bombastic writing style, and a few ideas that I might have actually half-understood.

The book asked: how have we historically drawn the category boundaries around madness? If there is some great continent containing nations like Irrationality, Immorality, Illness, and Inspiration, from which of these countries have cartographers carved out a homeland for Madness? What wars have been fought over which provinces? What propaganda supports the current international order. And how accurate is it?

II.

Foucault starts with the Late Middle Ages / early Renaissance, when there was suddenly an explosion of interest in madness. The most famous works from this tradition are *Ship of Fools* and *In Praise Of Folly* ("fool" originally meant “insane person”) – not to mention pretty much everything by Hieronymus Bosch – and of course Foucault has intensely studied two hundred other examples I’ve never heard of. His theory is that this shares a source with the late medieval fascination with death (think all of those pictures of dancing skeletons). In both cases, the stable tidy medieval order is teetering towards collapse, and so the popular imagination is seized by images of the Outside invading the familiar world.

I feel bad juxtaposing so eminent a figure as Foucault with Lovecraft, but I found his description of Renaissance madness easiest to understand as basically Lovecraftian. We’ve somehow lucked into a bubble of comfortable stability within a vast and horrifying universe. We’ve become so complacent that we’ve forgotten about the bubble and are starting to poke around the edges. When bits of the Outside leak in, madness is the inevitable result. Lovecraft came of age during the First World War, as a European order that considered itself too enlightened to die went down in flames. The end of the Middle Ages must have been a similar period. The insane are those who have seen too much of the horrors that lurk beyond the veil – Yog-Sothoth, Protestantism, whatever.

And like Cthulhu, madness has an affinity for water:

> One thing at least is certain: water and madness have long been linked in the dreams of European man. Already, disguised as a madman, Tristan had ordered boatmen to land him on the coast of Cornwall…And more than once in the course of time, the same theme reappears: among the mystics of the fifteenth century, it has become the motif of the soul as a skiff, abandoned on the infinite sea of desires, in the sterile field of cares and ignorance, among the mirages of knowledge, amid

> the unreason of the world — a craft at the mercy of the sea's great madness, unless it throws out a solid anchor, faith, or raises its spiritual sails so that the breath of God may bring it to port. At the end of the sixteenth century, De Lancre sees in the sea the origin of the demoniacal leanings of an entire people: the hazardous labor of ships, dependence on the stars, hereditary secrets, estrangement from women—the very image of the great, turbulent plain itself makes man lose faith in God and all his attachment to his home; he is then in the hands of the Devil, in the sea of Satan's ruses.

And so, Foucault tells us, in the fifteenth century there is a sudden emergence of a complex of artistic and philosophical themes linking madmen, the sea, and the terrible mysteries of the world. These culminate in the Ship Of Fools:

> Renaissance men developed a delightful, yet horrible way of dealing with their mad denizens: they were put on a ship and entrusted to mariners because folly, water, and sea, as everyone then knew, had an affinity for each other. Thus, "Ships of Fools" crisscrossed the seas and canals of Europe with their comic and pathetic cargo of souls. Some of them found pleasure and even a cure in the changing surroundings, in the isolation of being cast off, while others withdrew further, became worse, or died alone and away from their families. The cities and villages which had thus rid themselves of their crazed and crazy, could now take pleasure in watching the exciting sideshow when a ship full of foreign lunatics would dock at their harbors.

This was such a great piece of historical trivia that I was shocked I'd never heard it before. Some quick research revealed the reason: it is completely, 100% false. Apparently Foucault looked at an allegorical painting by

Hieronymus Bosch, decided it definitely existed in real life, and concocted the rest from his imagination.

Foucault apologists try to rescue this, say that he was just being poetic in some way. He wasn't. Page 8 in my copy: "Of all these romantic or satiric vessels, the *Narrenschiff* [Ship Of Fools] is the only one that had a real existence — for they did exist, these boats that conveyed their insane cargo from town to town." He really, really doubled down on this point. As far as I can tell, this is just as bad a failing of scholarship as it sounds – and surprising, since everything else about the book gives the impression of Foucault as an incredibly knowledgeable and wide-ranging scholar.

I couldn't find any mention of equally bad flaws in the rest of the book, and Foucault really does seem to know his stuff, so I'm tempted to treat this as a one-off error, albeit a completely inexplicable one. I'm including it anyway as a warning before getting into some other pretty weird stuff.

III.

Eventually the Renaissance became less of an impending threat and more of a *fait accompli*, and people's worries died down a bit. Madness began to be treated more as ordinary immorality. This didn't necessarily mean people freely chose to be mad – the classical age didn't think in exactly the same "it's your fault" vs. "it's biological" terms we do – but it was considered due to a weakness of character in the same way as other failures.

In some cases, it was the result of an excess of passions, flightiness, or imagination: the most famous example is Don Quixote, who went crazy after reading too many fiction books. This was actually considered a very serious risk by practically all classical authorities, especially for women. Foucault quotes Edme-Pierre Beauchesne:

> In the earliest epochs of French gallantry and manners, the less perfected minds of women were content with facts and events as marvelous as they were unbelievable; now they demand believable facts yet sentiments so marvelous that their own minds are disturbed and confounded by them; they then seek, in all that surrounds them, to realize the marvels by which they are enchanted; but everything seems to them without sentiment and without life, because they are trying to find what does not exist in nature.

And a newspaper of the time:

> The existence of so many authors has produced a host of readers, and continued reading generates every nervous complaint; perhaps of all the causes that have harmed women's health, the principal one has been the infinite multiplication of novels in the last hundred years … a girl who at ten reads instead of running will, at twenty, be a woman with the vapors and not a good nurse.

Novels weren't the only danger, of course. There were other hazards to watch for, like waking up too late:

> The moment at which our women rise in Paris is far removed from that which nature has indicated; the best hours of the day have slipped away; the purest air has disappeared; no one has benefited from it. The vapors, the harmful exhalations, attracted by the sun's heat, are already rising in the atmosphere.

Also, freedom:

> For a long time, certain forms of melancholia were considered specifically English; this was a fact in medicine and a constant in literature…Spurzheim made a synthesis of all these analyses in one of the last texts devoted to them. Madness, “more frequent in England than anywhere else,” is merely the penalty of the liberty that reigns there, and of the wealth universally enjoyed. Freedom of conscience entails more dangers than authority and despotism. “Religious sentiments exist without restriction; every individual is entitled to preach to anyone who will listen to him”, and by listening to such different opinions, “minds are disturbed in the search for truth.”

These are a *very* selective sampling of quotes from just one of Foucault’s many chapters, and some of them are separated by centuries from others, but the overall impression I got was that conformity/wholesomeness/clean living was salubrious, and deviations from these likely to cause madness. Essentially, if you deviate from your humanity a little bit of the way – by failing to be a godly, sober-living, and industrious person – then that can compound on itself and make you lose practically all of your humanity. You will end up a feral madman, little different from a beast.

This naturally lumped madness in together with the other failures of industry and godliness: crime and poverty. During the seventeenth century, madmen, beggars, and criminals were all crammed together in workhouses. These were always sort of ambiguous between “maybe the structure and routine of work will help these poor souls find the right path” and “let’s keep these losers away from the rest of us”. The opening of the workhouses was sudden and dramatic: in Paris, it began Monday May 14, 1657, when “the archers began to hunt down beggars and herd them into the different buildings of the Hospital.” In England, it started around 1630, when the King recommended prosecuting:

> …all those who live in idleness and will not work for reasonable wages, or who spend what they have in taverns…for these people live like savages without being married, nor buried, nor baptized; and it is this licentious liberty which causes so many to rejoice in vagabondage.

Foucault stresses that this wasn’t some plot on the part of authorities to enslave beggars and profit off their labor. The people in charge of the workhouses generally failed at assigning work that was profitable or productive, even in the weak sense of productive at lining their own pockets. They seemed genuinely driven by a belief in the curative power of Honest Work:

> Measured by their functional value alone, the creation of the houses of confinement can be regarded as a failure. Their disappearance throughout Europe, at the beginning of the nineteenth century, as receiving centers for the indigent and prisons of poverty, was to sanction their ultimate failure: a transitory and ineffectual remedy, a social precaution clumsily formulated by a nascent industrialization. And yet, in this very failure, the classical period conducted an irreducible experiment. What appears to us today as a clumsy dialectic of production and prices then possessed its real meaning as a certain ethical consciousness of labor, in which the difficulties of the economic mechanisms lost their urgency in favor of an affirmation of value.
>
> In this first phase of the industrial world, labor did not seem linked to the problems it was to provoke; it was

regarded, on the contrary, as a general solution, an infallible panacea, a remedy to all forms of poverty. Labor and poverty were located in a simple opposition, in inverse proportion to each other. As for that power, its special characteristic, of abolishing poverty, labor – according to the classical interpretation — possessed it not so much by its productive capacity as by a certain force of moral enchantment. Labor’s effectiveness was acknowledged because it was based on an ethical transcendence. Since the Fall, man had accepted labor as a penance and for its power to work redemption. It was not a law of nature which forced man to work, but the effect of a curse. The earth was innocent of that sterility in which it would slumber if men remained idle: “The land had not sinned, and if it is accursed, it is by the labor of the fallen man who cultivates it; from it no fruit is won, particularly the most necessary fruit, save by force and continual labor.”

According to Johan Huizinga, there was a time, at the dawn of the Renaissance, when the supreme sin assumed the aspect of Avarice, Dante’s cièca cupidigia. The seventeenth-century texts, on the contrary, announced the infernal triumph of Sloth: it was sloth which led the round of the vices and swept them on. Let us not forget that according to the edict of its creation, the Hôpital Général must prevent “mendicancy and idleness as sources of all disorder.” Louis Bourdaloue echoes these condemnations of sloth, the wretched pride of fallen man; “What, then, is the disorder of an idle life? It is, replies Saint Ambrose, in its true meaning a second rebellion of the creature against God.” Labor in the houses of confinement thus assumed its ethical meaning: since sloth had become the absolute form of rebellion, the idle would be forced to work, in the endless leisure of a labor without utility or profit.

Foucault presents confinement alternately as workhouses mixing together madmen and poor people, and as ultra-secure special hospitals for the insane. I’m not sure what to make of this contradiction; maybe the less ill people were in one, and the more ill people in the other? Maybe he’s just interested in the general phenomenon of confinement? In any case, the places for the insane were pretty bad too:

> In his Report on the Care of the Insane Desportes describes the cells of Bicetre as they were at the end of the eighteenth century: “The unfortunate whose entire furniture consisted of this straw pallet, lying with his head, feet, and body pressed against the wall, could not enjoy sleep without being soaked by the water that trickled from that mass of stone.” As for the cells of La Salpêtrière, what made “the place more miserable and often more fatal, was that in winter, when the waters of the Seine rose, those cells situated at the level of the sewers became not only more unhealthy, but worse still, a refuge for a swarm of huge rats, which during the night attacked the unfortunates confined there and bit them wherever they could reach them; madwomen have been found with feet, hands, and faces torn by bites which are often dangerous and from which several have died.”

It would be nice to think these kinds of things only survived because the public didn’t know about them, but that, uh, doesn’t seem to be quite what was happening:

> As late as 1815, if a report presented in the House of Commons is to be believed, the hospital of Bethlehem exhibited lunatics for a penny, every Sunday. Now the annual revenue from these exhibitions amounted to almost four hundred pounds, which suggests the astonishingly high number of 96,000 visits a year.

All of these locks and chains and cages and exhibitions draw an obvious analogy of madmen and animals. Foucault doesn’t think this is a coincidence:

> When practices reach this degree of violent intensity, it becomes clear that they are no longer inspired by the desire to punish nor by the duty to correct. The notion of a “résipiscence” is entirely foreign to this regime. But there was a certain image of animality that haunted the hospitals of the period. Madness borrowed its face from the mask of the beast. Those chained to the cell walls were no longer men whose minds had wandered, but beasts preyed upon by a natural frenzy: as if madness, at its extreme point, freed from that moral unreason in which its most attenuated forms are enclosed, managed to rejoin, by a paroxysm of strength, the immediate violence of animality. This model of animality prevailed in the asylums and gave them their cagelike aspect, their look of the menagerie…
>
> What is most important is that it is conceived in terms of an animal freedom. The negative fact that “the madman is not treated like a human being” has a very positive content: this inhuman indifférence actually has an obsessional value: it is rooted in the old fears which since antiquity, and especially since the Middle Ages, have given the animal world its familiar strangeness, its menacing marvels, its entire weight of dumb anxiety. Yet this animal fear which accompanies, with all its imaginary landscape, the perception of madness, no longer has the same meaning it had two or three centuries earlier: animal metamorphosis is no longer the visible sign of infernal powers, nor the result of a diabolic alchemy of unreason. The animal in man no longer has any value as the sign of a Beyond; it has become his madness, without relation to anything but

> itself: his madness in the state of nature. The animality that rages in madness dispossesses man of what is specifically human in him; not in order to deliver him over to other powers, but simply to establish him at the zero degree of his own nature.

There is a lot I didn’t understand about this section, but the overall gist seems to be trying to lump the insane in together with other forms of badness and deviation from the moral norm – whether animals or criminals – and shutting them away where they could not be seen.

IV.

The late eighteenth century on was the period of reform, when the mentally ill were taken out of the prisons and workhouses and brought to nice benevolent asylums in the countryside where they could convalesce in peace under the supervision of expert doctors.

…or at least this is the prevailing narrative. Foucault is having none of it.

The houses of confinement weren’t *just* for criminals and madmen. They were also for the “undeserving poor” – homeless, beggars, unemployed. But the Industrial Revolution was changing the conception of poverty. Foucault places this in the context of modern economics, which introduced abstract ideas like “jobs” and “workers”. In this model, the poor were potential workers who just lacked jobs, not the weird exotic subspecies of humanity called “paupers”. The past paradigm had focused on healing their souls through the redemptive power of make-work; the new paradigm said that if they could be enlisted to work productive industrial jobs it would improve the Economy and everyone would be better off.

Out they went, and now instead of just being an undifferentiated mass of undesirables, the hospitals were more obviously just the two disparate populations of criminals and madmen. But surely *now* is the point where people see how inhumane it is to stick the mentally ill together with criminals, right?

Sort of.

> When the Prior of Senlis asked that madmen be separated from certain convicts, what were his arguments? “He is deserving of mercy, as well as two or three others who would be better off in some citadel, because of the company of six others who are mad, and who torment them night and day.” And the meaning of this sentence would be so clearly understood by the police that the internees in question would be set free. And the demands of the Brunswick overseer have the same meaning: the workshop is disturbed by the cries and the confusion of the insane; their frenzy is a perpetual danger, and it would be better to send them back to the cells, or to keep them in chains. And already, we can anticipate that from one century to the next, the same protests did not have, at bottom, the same value. Early in the nineteenth century, there was indignation that the mad were not treated any better than those condemned by common law or than State prisoners; throughout the eighteenth century, emphasis was placed on the fact that the prisoners deserved a better fate than one that lumped them with the insane […]
>
> La Rochefoucauld-Liancourt bears witness to this in his report to the Committee on Mendicity: “One of the punishments inflicted upon the epileptics and upon the other patients of the wards, even upon the deserving

> poor, is to place them among the mad." The scandal lies only in the fact that the madmen are the brutal truth of confinement, the passive instrument of all that is worst about it.

There was apparently general agreement that it was unfair to criminals to keep them confined together with madmen, so out went the criminals – with the madmen staying around in institutions that were starting to sort of resemble the idea of a modern psychiatric hospital.

The beginning of the nineteenth century did start to see fewer chains and rats, and more attempt to treat madmen as human beings. But Foucault is perversely annoyed by this, convinced that this was secretly a way of respecting the mentally ill even less. He notes that their newfound rights were conditional on good behavior and on acting sane, and so in a sense these new more compassionate hospitals gave them *less* freedom than the old ball-and-chain deal. In the older hospitals, you could do whatever you wanted. You'd be doing it on the wrong side of iron bars, mocked by people who hated you, but you could do it. In the new hospitals, you were forced to constantly perform and please your guards and nurses in order to maintain your privileges. Madmen went from being treated like criminals – who at least are still adult citizens – to being treated like children:

> We must therefore re-evaluate the meanings assigned to Tuke's work: liberation of the insane, abolition of constraint, constitution of a human milieu – these are only justifications. The real operations were different. In fact Tuke created an asylum where he substituted for the free terror of madness the stifling anguish of responsibility; fear no longer reigned on the other side of the prison gates, it now raged under the seals of

> conscience. Tuke now transferred the age-old terrors in which the insane had been trapped to the very heart of madness. The asylum no longer punished the madman’s guilt, it is true; but it did more, it organized that guilt; it organized it for the madman as a consciousness of himself, and as a non-reciprocal relation to the keeper; it organized it for the man of reason as an awareness of the Other, a therapeutic [and so on for two hundred pages].

A lynchpin of this system was doctors. Foucault says that at this time, doctors really didn’t make much pretense to being able to cure mental illness. Their main role was as a representative of polite society and healthy living. The doctor would go in, talk to some mad people about how really being virtuous and healthful was better than being degenerate and crazy, and this would help the process of drawing them back into the social order (and so out of the excessive wild liberty that was madness). The more high-status and authoritative the doctor, the better – and he has lots of examples of doctors supposedly curing madmen with a couple of stern words delivered in a suitably censorious tone.

He theorizes that after the restore-to-social-order idea of mental health became obsolete, doctors were stuck without a purpose. That is, it was known that it was important to have doctors treating the mentally ill, but unclear exactly what they were supposed to do. One response was to flounder around for a while on various scams and miracle cures. Another response was Freud’s: to accept that the doctor-patient relationship itself somehow had magical properties, and that the doctor being a silent authority figure sitting in judgment of you was actually an effective way to cure psychiatric disease.

V.

Everything above is a really superficial reading of *Madness And Civilization* and probably misses the whole point of the book.

This point is something that alternately seems postmodern or kabbalistic or – for lack of a better term – insane. It’s not just saying that This Historical Period treated the mad This Way, but That Historical Period treated them That Way. It’s trying to peek beneath the hood (or the veil?) to find the *zeitgeist*, the animating spirit of the European continent that led them to do things as they did and which transformed one schema into another. This is rarely anything sensible, like “the economy improved” or “there was a revolution”. More often it’s some kind of deep subconscious beliefs about the meaning of humanity or freedom or symbolism or something. If Europe was one guy, this book would be Foucault performing Freudian dream analysis on that guy.

For example, the Europeans didn’t put their madmen on Ships Of Fools just because it was a convenient way to get rid of them, but also because:

> Water adds to this the dark mass of its own values; it carries off, but it does more: it purifies. Navigation delivers man to the uncertainty of fate; on water, each of us is in the hands of his own destiny; every embarkation is, potentially, the last. It is for the other world that the madman sets sail in his fools’ boat; it is from the other world that he comes when he disembarks. The madman’s voyage is at once a rigorous division and an absolute Passage. In one sense, it simply develops, across a half-real, half-imaginary geography, the madman’s liminal position on the horizon of medieval concern—a position symbolized and made real

> at the same time by the madman’s privilege of being confined within the city gates: his exclusion must enclose him; if he cannot and must not have another prison than the threshold itself, he is kept at the point of passage. He is put in the interior of the exterior, and inversely. A highly symbolic position, which will doubtless remain his until our own day, if we are willing to admit that what was formerly a visible fortress of order has now become the castle of our conscience.
>
> Confined on the ship, from which there is no escape, the madman is delivered to the river with its thousand arms, the sea with its thousand roads, to that great uncertainty external to everything. He is a prisoner in the midst of what is the freest, the openest of routes: bound fast at the infinite crossroads. He is the Passenger par excellence; that is, the prisoner of the passage. And the land he will come to is unknown—as is, once he disembarks, the land from which he comes. He has his truth and his homeland only in that fruitless expanse between two countries that cannot belong to him. Is it this ritual and these values which are at the origin of the long imaginary relationship that can be traced through the whole of Western culture? Or is it, conversely, this relationship that, from time immemorial, has called into being and established the rite of embarkation?

Let’s appreciate a few things about this passage. First, it’s phenomenal writing. I apologize for thinking all Continental philosophy had to be badly-written; in retrospect Nietzsche should have cured me of this delusion.

But second, it’s totally bonkers. Like, forget the fact that there weren’t any real Ships Of Fools and Foucault is analyzing a literary motif. Forget that the literary motif actually comes from a metaphor by Plato which is about something else. Even if the rivers of Europe were choked with such Ships, this is just a phenomenally unproductive way to think about anything. This is the kind of thought process where we drill for oil because we are symbolically sexually penetrating Mother Earth (insert kabbalistic analysis of the word “fracking” here).

There is a lot along these lines, none of which I am really able to follow, especially because the book never gives a clear definition of its crucial term “unreason”. The closest I can come is a theory that the Renaissance (and to some degree the later classical period) thought of madness as potentially interesting and valuable. They didn’t like madmen, but they occasionally tried to have a “dialogue between madness and reason”, where they would try to understand where the mad were coming from and what they had to offer civilization. In later periods, this was lost, and the mad were just confined away from human sight – but there was still at least some dignity in it, because madness was allowed to exist on its own terms. Later, when bleak prison workhouses transitioned to humane medical asylums, even that dignity was lost, as sane people’s imperative changed to forcing the mad to conform to the sane world’s standards and deny their madness’ existence.

This is the thrust of the last chapter, and Foucault ties all of this together into a case that all of the reformers were just jerks, and they sought more humane treatment for the mentally ill out of a desire to judge and dominate them. This is fantastically contrarian. Foucault does not give an *inch* to the position that maybe there was something good and wholesome about the desire to rescue people from being crammed by the dozen in rat-infested cells with all of their limbs chained together. He doesn’t specifically say the rat-infested cells were better, but he sure hints at it pretty hard.

I always like contrarian takes. But I can’t make sense of what Foucault is trying to do here. And also, some of the same sites that debunk the Ship Of Fools thing say that actually the Renaissance was super-cruel to mad people, and Foucault’s picture of them as tolerant and understanding is composed entirely of cherry-picking and imagination.

The best I can do here is say that Foucault is too much of an Idealist where I am a Materialist. I measure humanitarian victories in prisoners freed and rat bites averted. He seems to measures them how the dream sequences of Personified Europe are treating the dialogue between Madness and Reason. Probably there’s a perspective in which this makes sense, but this book didn’t manage to teach me to appreciate it.

VI.

Granted that I couldn’t appreciate the philosophy and remain doubtful of the scholarship, I still enjoyed this book. It was a weird tour of parts of history I wouldn’t otherwise have thought of. And it did accomplish the post-modernist goal of broadening my perspective enough to be more doubtful of my own society’s institutions.

The idea of novel-reading causing insanity seems ridiculous to us. But is it any *more* ridiculous than the idea of video games causing violence? Or stereotype threat causing poor test performance? The dustbin of scientific history is filled with weird claims that various social and cultural phenomena have powerful effects on the mind, from refrigerator-mother schizophrenia to low self-esteem causing crime. Surely this is one more warning to before voodoo psychology.

But that’s too easy. I also worry about the idea – constant in its essence in every time period, though changing in its particulars – that mental illness is the result of living your life in an unwholesome way and indulging in illicit pleasures. In the classical period, this included everything from waking up too late, to not working long enough hours, to getting too romantically infatuated, to, well…

> Heat clears the way for liquids. It is precisely for this reason that all the hot drinks the seventeenth century used and abused risk becoming harmful: relaxation, general humidity, softness of the entire organism. And since these are the distinctive traits of the female body, as opposed to virile dryness and solidity, the abuse of hot drinks risks leading to a general feminization of the human race. [Thomas Sydenham warns:] “Most men are censured, not without reason, for having degenerated in contracting the softness, the habits, and the inclinations of women. Excessive use of humectants immediately accelerates the metamorphosis and makes the two sexes almost as alike in the physical as in the moral realm. Woe to the human race, if this prejudice extends its reign to the common people.”

And I can’t help noticing a resemblance between this and the modern insistence on diet and exercise. They’ve got the same kind of element of “if you don’t exert a lot of willpower to live your life in a diligent way, then you shouldn’t be surprised when you end up mentally ill.” Although the role of poor diet/exercise in physical illness is beyond questioning, its role in mental illness is more anecdotal and harder to pin down. Don’t get me wrong, there are lots of studies showing it works. But there’s also lots of anomalous data, like how exercise performed as part of your job doesn’t help. This has led some people to suggest that the physical effects of exercise are less important than the social role – the feeling of doing something to fight your depression and conform to a virtuous mode of life. Exercise works for that – but so might avoiding novels and staying away from hot drinks, if that was what your society wanted. I'm not saying this is definitely true. I'm just saying I give it higher credence now that the pattern of "people always want to use willpower-induced conformity to social order as a bulwark against mental illness" is more apparent.

On the other hand, it's also important not to dismiss something we believe today just *because* people in the past believed the same thing. I was tempted to say something like I'm skeptical of modern-day behavioral activation because it sounds exactly like past-days "work will cure you because idleness is the mother of all sins" doctrine. But on closer examination, I'm using evidence wrong here. If people in the past believed something, that should be at least some positive evidence it was true – or at least not negative evidence. Of course, they're going to phrase it in really awkward politically-incorrect ridiculous-seeming terms, because they're the past. And they'll probably figure out some way to make it imply a moral atrocity, because, again, past. But that doesn't mean they're *wrong*. Very possibly it's a timeless truth that routine and purposeful activity help depression. The past phrased this as "idleness is the mother of sin so we should force everyone into workhouses", and now we're not as much about forced labor and tracing out sin's family tree. But behavioral activation therapy still seems pretty powerful.

I think I am going to be suspicious when the implied message stays the same but the specifics keep changing – "stay away from exciting novels" vs. "stay away from violent video games", or "avoid hot drinks" vs. "avoid sugary foods". It seems potentially safer when the specifics stay the same, with only the wording and the proposed responses changing.

There was one more thing that worried me about the past, much broader than any of these specific issues: doctors were very sure their cures worked. I knew *in principle* that there were a lot of placebo cures and cherry-picking, but it's another thing to have to read story after story of doctors trying ridiculous treatments – one of them had his patient eat soap to cleanse their circulation – and reporting that it definitely worked, every time, and patients who had been violently insane for years were restored to perfect health. I have worked in a lot of excellent psychiatric hospitals, and not one of them has worked anywhere near as well as people in the seventeenth century record their completely ridiculous mental health system of telling people not to reading novels to have worked.

Either everyone in the past is a total liar (given this effect, probably true), Foucault himself is a total liar (given the Ship of Fools thing, probably true), or we need even more constant vigilance than we've been applying thus far (alas, probably also true).

Fight Me, Psychologists: Birth Order Effects Exist And Are Very Strong

“Birth order” refers to whether a child is the oldest, second-oldest, youngest, etc. in their family. For a while, pop psychologists created a whole industry around telling people how their birth order affected their personality: oldest children are more conservative, youngest children are more creative, etc.

Then people got around to actually studying it and couldn’t find any of that. Wikipedia’s birth order article says:

> Claims that birth order affects human psychology are prevalent in family literature, but studies find such effects to be vanishingly small….the largest multi-study research suggests zero or near-zero effects. Birth-order theory has the characteristics of a zombie theory, as despite disconfirmation, it continues to have a strong presence in pop psychology and popular culture.

I ought to be totally in favor of getting this debunked. After all, the replication crisis in psychology highlights the need to remain skeptical of poorly-supported theories. And some of the seminal work disproving birth order was done by Judith Rich Harris, an intellectual hero of mine who profoundly shaped my worldview with her book *The Nurture Assumption.*

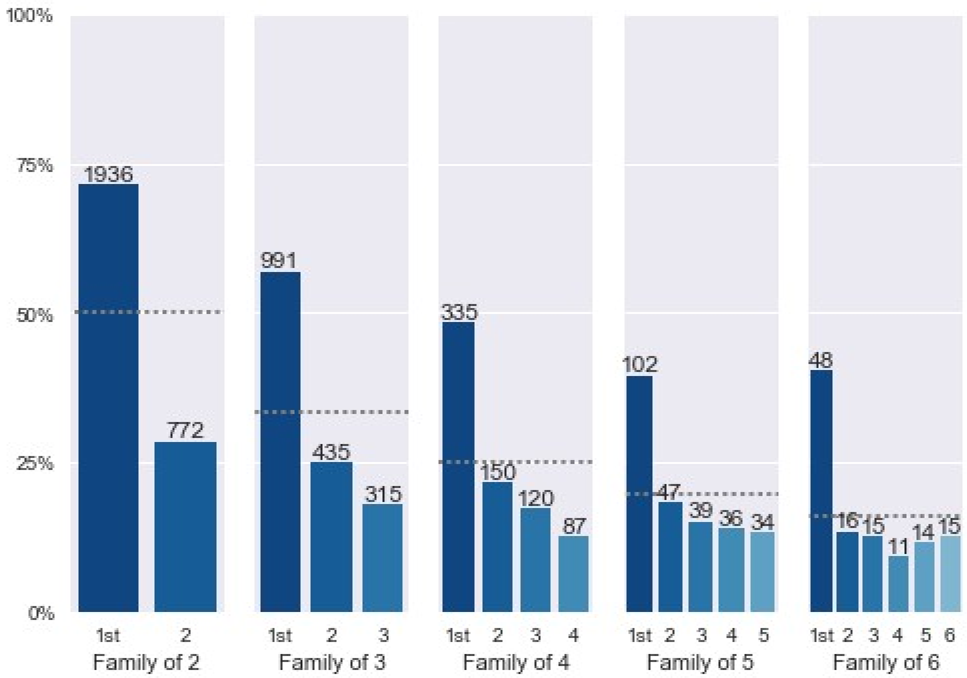
So I regret to have to inform you that birth order effects are totally a real thing.

I first started thinking this at transhumanist meetups, when it would occasionally come up that everyone there was an oldest child. The pattern was noticeable enough that I included questions about birth order on the latest SSC survey. This blog deals with a lot of issues around transhumanism, futurology, rationality, et cetera, so I thought it would attract the same kind of people.

7,248 people gave me enough information to calculate their birth order, but I am *very paranoid* because previous studies have failed by failing to account for family size. That is, people of certain economic classes/religions/races/whatever tend to have larger family sizes, and if you're in a large family, you're more likely to be a later-born child. In order to be absolutely sure I wasn't making this mistake, I concentrated on within-family-size analyses. For example, there were 2965 respondents with exactly one sibling…

…and a full 2118 of those were the older of the two. That's 71.4%. $p \leq 0.00000001$.

The same effect occurs in sibships of other sizes. Of the 1884 respondents from families with three children ($n = 1884$), 56.8% are the oldest, compared to predicted 33%. In families with four children ($n = 765$), 48.2% are the oldest, compared to predicted 25%.



Number of responses by birth order in sibships of different sizes; graph is by Emile and uses the public data only, which means exact numbers may be slightly different

This effect reaches the same scale as other effects people consider important. For example, the survey population drew heavily from STEM fields and was predictably very white; however, the birth order gap was larger in magnitude than the racial gap. It is bigger than gender gaps in some fields traditionally considered to have major gender gaps, like undergraduate economics. This can fairly be considered a large effect.

So what is going on here?

It’s unlikely that age alone is driving these results. In sibships of two, older siblings on average were only about one year older than younger siblings. That can’t explain why one group reads this blog so much more often than the other.

And all of the traditional pop psychology claims about birth order don't seem to hold up. I didn't find any effect on anything that could be reasonably considered conservativism or rebelliousness.

But there is at least one reputable study that did find a few personality differences. This is Rohrer et al (2015), which examined a battery of personality traits and found birth order effects only IQ and Openness to Experience, both very small.

I was only partly able to replicate this work. Rohrer et al found that eldest siblings had an advantage of about 1.5 IQ points. My study found the same: 1.3 to 1.7 IQ points depending on family size – but this did not reach significance. My other measure of intelligence was SAT, but SATs have been renormed and changed so many times over the past few decades that making apples-to-apples comparisons were really tough. I was able to get only a couple of weak and inconsistent effects: in sibships of two, eldest children had a slightly higher SAT1600 (1481 vs. 1458, p = 0.002) but not SAT2400; in sibships of 3+, eldest children had a slightly higher SAT2400 (2214 vs. 2248, p = 0.03), but not SAT1600. Overall this seems way too weak to say anything with certainty. Average SATs and IQs were already around the 99th percentile, so there may have been too much of a selection effect / ceiling effect to get good results.

The Openness results were clearer. Eldest children had significantly higher Openness (73rd %ile vs. 69th %ile, p = 0.001). Like Rohrer, I found no difference in any of the other Big Five traits.

Because I only had one blunt measure of Openness, I couldn't do as detailed an analysis as Rohrer's team. But they went on to subdivide Openness into two subcomponents, Intellect and Imagination, and found birth order only affected Intellect. They sort of blew Intellect off as just “self-estimated IQ”, but I don’t think this is right. Looking at it more broadly, it seems to be a measure of intellectual curiosity – for example, one of the questions they asked was, “I am someone who is eager for knowledge”. Educational Testing Service describes it as “liking complex problems”, and its opposite as “avoiding philosophical discussion”.

This seems promising. If older siblings were more likely to enjoy complex philosophical discussion, that would help explain why they are so much more likely to read a blog about science and current events. Unfortunately, the scale is completely wrong. Rohrer et al’s effects are *tiny* – going from a firstborn to a secondborn has an effect size of 0.1 SD on Intellect. In order to contain 71.6% firstborns, this blog would have to select for people above the 99.99999999th percentile in Intellect. There are only 0.8 people at that level in the world, so no existing group is that heavily selected.

I think the most likely explanation is that tests for Openness have limited validity, which makes the correlation look smaller than it really is. If being an eldest sibling increases true underlying Openness by a lot, but your score on psychometric tests for Openness only correlates modestly with true underlying Openness, that would look like being an eldest sibling only increasing test-measured-Openness a little bit.

(cf. Riemann and Kandler (2010), which finds that the heritability of Openness shoots way up if you do a better job assessing it)

If we suppose that birth order has a moderate effect size on intellectual curiosity of 0.5 SD, that would imply that science blogs select for people in the top 3% or so of intellectual curiosity, a much more reasonable number. Positing higher (but still within the range of plausibility) effect sizes would decrease the necessary filtering even further.

If this is right, it suggests Rohrer et al undersold their conclusion. Their bottom line was something like “birth order effects may exist for a few traits, but are too small to matter”. I agree they may only exist for a few traits, but they can be strong enough to skew ratios in some heavily-selected communities like this one.

When I asked around about this, a couple of people brought up further evidence. Liam Clegg pointed out that philosophy professor Michael Sandel asks his students to raise their hand if they’re the oldest in their family, and usually gets about 80% of the class. And Julia Rohrer herself was kind enough to add her voice and say that:

> I’m not up to fight you because I think you might be onto something real here. Just to throw in my own anecdotal data: The topic of birth order effect comes up quite frequently when I chat with people in academic contexts, and more often than not (~80% of the time), the other person turns out to be firstborn. Of course, this could be biased by firstborns being more comfortable bringing up the topic given that they’re supposedly smarter, and it’s only anecdotes. Nonetheless, it sometimes makes me wonder whether we are missing something about the whole birth order story.

But *why* would eldest siblings have more intellectual curiosity? There are many good just-so stories, like parents having more time to read to them as children. But these demand strong effects of parenting on children's later life outcomes, of exactly the sort that behavioral genetic studies consistently find not to exist. An alternate hypothesis could bring in weird immune stuff, like that thing where people with more older brothers are more likely to be gay because of maternal immunoreactivity to the Y chromosome (which my survey replicates, by the way). But this is a huge stretch and I don't even know if people are sure this explains the homosexuality results, let alone the birth order ones.

If mainstream psychology becomes convinced this effect exists, I hope they'll start doing the necessary next steps. This would involve seeing if biological siblings matter more or less than adopted siblings, whether there's a difference between paternal and maternal half-siblings, how sibling age gaps work into this, and whether only children are more like oldests or youngests. Their reward would be finding some variable affecting children's inherent intellectual curiosity – one that might offer opportunities for intervention.

If you want to double-check these results or analyze them further, you can download the data as .xlsx or .csv. Some people have complained of weird problems in the csv format and I recommend the xlsx if at all possible. I have removed the data of a few people who did not want their answers to be public, so you may not get exactly the same numbers I did, but they should be pretty close. If you think this could be turned into a paper and are interested in making it happen, please get in contact with me.

Self-Serving Bias

Alex Tabarrok beat me to the essay on Oregon's self-service gas laws that I wanted to write.

Oregon is one of two US states that bans self-service gas stations. Recently, they passed a law relaxing this restriction – self-service is permissable in some rural counties during odd hours of the night. Outraged Oregonians took to social media to protest that self-service was unsafe, that it would destroy jobs, that breathing in gas fumes would kill people, that gas pumping had to be performed by properly credentialed experts – seemingly unaware that most of the rest of the country and the world does it without a second thought.

…well, sort of. All the posts I've seen about it show the same three Facebook comments. So at least three Oregonians are outraged. I don't know about the rest.

But whether it's true or not, it sure makes a great metaphor. Tabarrok plays it for all it's worth:

> Most of the rest of the America–where people pump their own gas everyday without a second thought–is having a good laugh at Oregon's expense. But I am not here to laugh because in every state but one where you can pump your own gas you can't open a barbershop without a license. A license to cut hair! Ridiculous. I hope people in Alabama are laughing at the rest of America. Or how about a license to be a manicurist? Go ahead Connecticut, laugh at the other states while you get your nails done. Buy contact lens without a prescription? You have the right to smirk British Columbia!

> All of the Oregonian complaints about non-professionals pumping gas–“only qualified people should perform this service”, “it’s dangerous” and “what about the jobs”–are familiar from every other state, only applied to different services.

Since reading Tabarrok’s post, I’ve been trying to think of more examples of this sort of thing, especially in medicine. There are way too many discrepancies in approved medications between countries to discuss every one of them, but did you know melatonin is banned in most of Europe? (Europeans: did you know melatonin is sold like candy in the United States?) Did you know most European countries have no such thing as “medical school”, but just have college students major in medicine, and then become doctors once they graduate from college? (Europeans: did you know Americans have to major in some random subject in college, and then go to a separate place called “medical school” for four years to even start learning medicine?) Did you know that in Puerto Rico, you can just walk into a pharmacy and get any non-scheduled drug you want without a doctor’s prescription? (source: my father; I have never heard anyone else talk about this, and nobody else even seems to think it is interesting enough to be worth noting).

And I want to mock the people who are doing this the “wrong” way – but can I really be sure? If each of these things decreased the death rate 1%, maybe it would be worth it. But since nobody notices 1% differences in death rates unless they do really good studies, it would just look like some state banning things for no reason, and everyone else laughing at them.

Actually, how sure are we that Oregon was wrong to ban self-service gas stations? How *do* disabled people pump their gas in most of the country? And is there some kind of negative effect from breathing in gas fumes? I have never looked into any of this.

Maybe the real lesson of Oregon is to demonstrate a sort of adjustment to prevailing conditions. There's an old saying: "Everyone driving faster than you is a maniac; anyone driving slower than you is a moron". In the same way, *no matter what the current level of regulation is*, removing any regulation will feel like inviting catastrophe, and adding any regulation will feel like choking on red tape.

Except it's broader than regulation. *Scientific American* recently ran an article on how some far-off tribes barely talk to their children at all. *New York Times* recently claimed that "in the early 20th century, some doctors considered intellectual stimulation so detrimental to infants that they routinely advised young mothers to avoid it". And our own age's prevailing wisdom of "make sure your baby has listened to all Beethoven symphonies by age 3 months or she'll never get into college" is based on equally flimsy evidence, yet somehow it still feels important to me. If I don't make my kids listen to Beethoven, it will feel like some risky act of defiance; if I don't take the early 20th century advice to avoid overstimulating them, it will feel more like I'm dismissing people who have been rightly tossed on the dungheap of history.

And then there's the discussion from the recent discussion of *Madness and Civilization* about how 18th century doctors thought hot drinks will destroy masculinity and ruin society. Nothing that's happened since has really disproved this – indeed, a graph of hot drink consumption, decline of masculinity, and ruinedness of society would probably show a pretty high correlation – it's just somehow gotten tossed in the bin marked “ridiculous” instead of the bin marked “things we have to worry about”.

So maybe the scary thing about Oregon is how strongly we rely on intuitions about absurdity. If something doesn’t immediately strike us as absurd, then we have to go through the same plodding motions of debate that we do with everything else – and over short time scales, debate is interminable and doesn’t work. Having a notion strike us as absurd short-circuits that and gets the job done – but the Oregon/everyone-else divide shows that intuitions about absurdity are artificial and don’t even survive state borders, let alone genuinely different cultures and value systems.

And maybe this is scarier than usual because I just read Should Schools Ban Kids From Having Best Friends? I assume this is horrendously exaggerated and taken out of context and all the usual things that we’ve learned to expect from news stories, but it got me thinking. Right now enough people are outraged at this idea that I assume it’ll be hard for it to spread too far – and even if it does spread, we can at least feel okay knowing that parents and mentors and other people in society will maintain a belief in friendship and correct kids if schools go wrong. But what if it catches on? What if, twenty years from now, the idea of banning kids from having best friends has stopped generating an intuition of absurdity? Then if we want kids to still be allowed to have best friends, we’re going to have to (God help us) *debate it*. Have you *seen* the way our society debates things?

And I know some people see this and say it proves rational debate is useless and we should stop worrying about it. But trusting whatever *irrational* forces determines what sounds absurd or not doesn’t sound so attractive either. I think about it, and I want to encourage people to be really, really good at rational debate, just in case something terrible loses its protective coating of absurdity, or something absolutely necessary gains it, and our ability to *actually judge whether things are good or bad and convince other people of it* is all that stands between us and disaster.

And, uh, maybe the people who say kids shouldn't be allowed to have best friends are right. I admit they've thought about this a lot longer than I have. My problem isn't that someone thinks this. It's that so much – even the legitimacy of friendship itself – can now depend on our culture's explicit rationality. And our culture's explicit rationality is so bad. And that the only alternative to dragging everything before the court of explicit rationality is some version of Chesterton's Fence, ie the very heuristic telling Oregonians to defend full-service gas stations to the death. *There is no royal road*.

Maybe this is a good time to get on our chronophones with Oregon (or more prosaically, use the Outside View). Figure out what cognitive strategies you would recommend to an Oregonian trying to evaluate self-service gas stations. Then try to use those same strategies yourself. And try to imagine the level of careful thinking and willingness to question the *status quo* it would take to make an Oregonian get the right answer here, and be skeptical of any conclusions you've arrived at with any less.

Maybe The Real Superintelligent AI Is Extremely Smart Computers

I.

By Ted Chiang, on Buzzfeed: The Real Danger To Civilization Isn’t AI: It’s Runaway Capitalism. Chiang’s science fiction is great and I highly recommend it. This article, not so much.

The gist seems to be: hypothetical superintelligent AIs sound a lot like modern capitalism. Both optimize relentlessly for their chosen goal (paperclips, money), while ignoring the whole complexity of human value.

It’s a good point, and I would have gone on to explain the more general idea of an optimization process. Evolution optimizes relentlessly for reproductive fitness, capitalism optimizes relentlessly for money, politics optimizes relentlessly for electability. Humans are sort of an optimization process, but such a weird edge case that “non-human optimizers” is a natural category to people more used to the human variety. Both future superintelligences and modern corporations are types of non-human optimizers, so they’ll naturally be similar in ways – though not so many ways that taking the comparison too far won’t carry you off a cliff. And one of those ways will be that even though they both know humans have complex values, they won’t care. Facebook “knows” that people enjoy meaningful offline relationships; after all, it’s made entirely of human subunits who know that. It’s just not incentivized to do anything with that knowledge. Future superintelligences will likely be in a similar position – see section 4.1 here.

But Chiang argues the analogy proves that AI fears are absurd. This is a really weird thing to do with an analogy.

Science has always been a fertile source of metaphors. The Pentagon budget is a *black hole*. The rise of ISIS will start a *chain reaction*. Social responsibility is in our corporate *DNA*. But until now, nobody has tried to use scientific metaphor as evidence in scientific debates. For a long time astronomers were unsure whether black holes really existed. But nobody thought the argument that “the REAL black hole is the Pentagon budget!” deserved to be invited to the discussion.

Actually this is worse than that, because the analogy is based on real similarities of mechanism. “People say in the future we might have fusion power plants. But look at all these ways fusion power plants resemble stars! Obviously stars are the *real* fusion power plants. And so by this, we can know that the future will never contain fusion power.” *Huh?*

II.

Still, Chiang pursues this angle relentlessly. Though he doesn’t use the word, he bases his argument around the psychological concept of projection, where people trying to avoid thinking about their own attributes unconsciously attribute them to others:

> Billionaires like Bill Gates and Elon Musk assume that a superintelligent AI will stop at nothing to achieve its goals because that’s the attitude they adopted…It’s no surprise that Silicon Valley capitalists don’t want to think about capitalism ending. What’s unexpected is that the way they envision the world ending is through a form of unchecked capitalism, disguised as a superintelligent AI. They have unconsciously created a devil in their own image, a boogeyman whose excesses are precisely their own.

> Which brings us back to the importance of insight. Sometimes insight arises spontaneously, but many times it doesn’t. People often get carried away in pursuit of some goal, and they may not realize it until it’s pointed out to them, either by their friends and family or by their therapists. Listening to wake-up calls of this sort is considered a sign of mental health.

In my own psychiatric practice, I am always *very* reluctant to assume a patient is projecting unless I know them very well. I’ve written more about the dangers of defense mechanism narratives here, but the short version is that amateur therapists inevitably end up using them to trivialize or psychologize a patient’s real concerns. I can’t tell you how many morons hear a patient say “I think my husband hates our kids”, give some kind of galaxy-brain level interpretation like “Maybe what’s really going on is *you* unconsciously hate your kids, but it’s more comfortable for you to imagine this of your husband”, and then get *absolutely shocked* when the husband turns out to be abusing the kids.

Accusing an entire region of California of projection is a novel psychoanalytic manuever, and I’m not sure Chiang and Buzzfeed give it the caution it deserves. The problem isn’t that they don’t have a plausible-sounding argument. The problem is that this sort of hunting-for-resemblances is a known bug in the human brain. You can do it to anything, and it will *always* generate a plausible-sounding argument.

Don’t believe me? What about black holes? Scientists say they exist, but I think these scientists are just creating “a devil in their own image, a boogeyman whose excesses are precisely their own.” Think about it. Superstar physicists like Einstein help university STEM departments suck up all the resources that should go to the humanities and arts. So of *course* when Einstein tries to imagine outer space, he thinks of super-stars that suck up all the resources from surrounding areas!

And chain reactions! You know what was a chain reaction? Enrico Fermi discovered some stuff about atoms. Then Leo Szilard wrote a letter to President Roosevelt saying it might have military applications. Then Roosevelt set up a project to develop military applications. One thing led to another, and a couple of Japanese cities got vaporized and the rest of the world teetered on the brink of total annhilation. Of *course* nuclear physicists became obsessed with the idea of chain reactions: they were living in one. They expected that subatomic particles would behave the same way they did – start out working on innocent little atomic collisions, have everything snowball out of control, and end up culpable for a nuclear explosion.

Watson and Crick worked together pretty closely on the discovery of DNA. So they started imagining organic molecules doing the same thing they did – two of them, intertwining. Just as they published papers which became the inspiration for an entire body of knowledge, so DNA was full of letters that caused the existence of an entire body. Epigenetics is relevant but generally ignored for the sake of keeping things simple, so it represents Rosalind Franklin.

I could go on all day like this. In fact, I have: this was the central narrative of my novel Unsong, where the world runs on “the kabbalistic method” and correspondences between unlike domains are the royal road to knowledge. You know who *else* wrote a story about a world that ran on kabbalah? *Ted Chiang*. This is not a coincidence because nothing is ever a coincidence.

III.

But Chiang’s comparison isn’t even good kabbalah. The correspondences don’t really correspond; the match-ups don’t really match.

He bases his metaphor on the idea that worries about AI risk comes from Silicon Valley. They don’t. The tech community got interested later. The original version of the theory comes from Nick Bostrom, a professor at Oxford, and Eliezer Yudkowsky, who at the time I think was living in Chicago. It was pushed to public notice by leading AI scientists all around the world. And before it was endorsed by Silicon Valley tycoons, it was endorsed by philosophers like David Chalmers and scientists like Stephen Hawking.

(Hawking, by the way, discovered that information could escape black holes despite a bunch of science saying they should be completely inert. This seems *suspiciously* similar to how he himself is completely paralyzed, but manages to convey information to the outside world via an artificial speaking device. More projection?)

Forcing the argument to rely on “well, also lots of people in Silicon Valley think this too” makes it hopelessly weak.

Consider: lots of Hollywood celebrities speak out about global warming. And we’re gradually finding out that some pretty awful things go on in Hollywood. Does that mean “The Real Problem Isn’t Global Warming, It’s Hollywood Harassment”? Does that license some author to write (while scientists facepalm worldwide) that because *he* doesn’t feel like carbon dioxide should be able to warm the climate, any claims to the contrary must be Hollywood celebrities projecting their own moral inadequacies? (possible angle: celebrities’ utterances emit carbon dioxide, and create a *stifling climate* for women in the entertainment industry)

If this sounds like a straw man to you, I challenge you to come up with any way it differs from what Chiang is doing with AI risk. You take a scientific controversy over whether there's a major global risk. You ignore the science and focus instead on a subregion of California that seems unusually concerned with it. You point out some bad behavior of that subregion of California. You kabbalistically connect it to the risk in question. Then you conclude that people worried about the risk are just peddling science fiction.

(wait, of *course* Chiang interprets this as people peddling science fiction. He's a science fiction writer! More projection!)

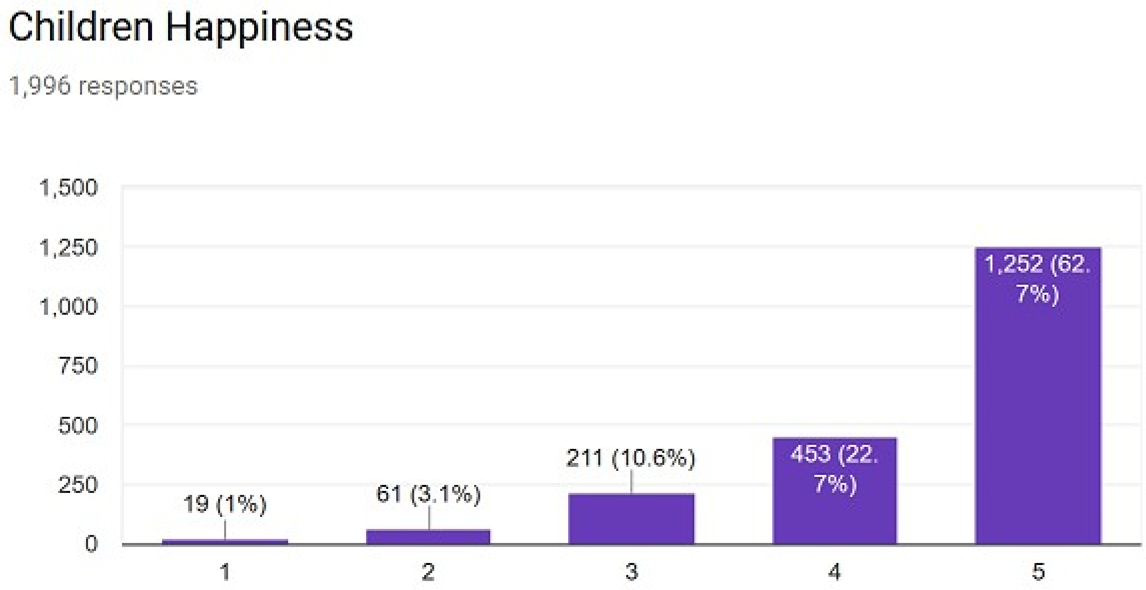
If the Hollywood example sounds more blatant or less plausible than the AI example, I maintain it's only because we're already convinced global warming is real and dangerous. That gives it the same kind of legitimacy as self-service gas stations, and grants it extra resistance against sophistry. That's all. That's the whole difference.

This isn't how risk assessment works. This isn't how good truth-seeking works. Whether or not you believe in AI risk, you should be disappointed that this is how we deal with issues that could be catastrophic to get wrong.

Bundles Of Joy

[Content warning: Discussion of child-rearing, may invoke mild feelings of social pressure to have children]

On December's survey, I asked readers who had children whether they were happy with that decision. Here are the results, from 1 (very unhappy) to 5 (very happy):



The mean was 4.43, and the median 5. People are really happy to have kids!

This was equally true regardless of gender. The male average (4.43, n = 1768) and female average (4.49, n = 177) were indistinguishable.

To double-check this, I compared the self-reported life satisfaction of people with and without kids. People with kids were much more satisfied – but also did much better on lots of other variables like financial situation, romantic satisfaction, etc. So probably at least some of the effect was because people with kids tend to be older people in stable relationships who have their life more figured out, and maybe also more religious.

In order to compare apples to apples, I limited the comparison to married atheist men 25 or older. There was no longer a consistent trend for people with at least one child to be more satisfied. But there was a trend for increasing satisfaction with increasing number of children:

NUMBER OF CHILDREN : AVERAGE LIFE SATISFACTION ON 1-10 SCALE (total n = 1491):
0: 7.06
1: 7.09
2: 7.24
3: 7.31
4+: 7.43

This doesn’t make a lot of sense to me, since I would expect the biggest life change to be going from zero children to one child. Probably some residual confounders remain in the analysis – and commenter “meh” points out that people who are happiest with their existing children will be most likely to have more. But at the very least, people with children don’t seem to be *less* happy.

These results broadly match existing research, which usually finds that parents report being very happy to have children, but that this is not reflected in life satisfaction numbers. The main difference is that existing research usually claims parents have *lower* life satisfaction than non-parents. But this is different in different countries, either for cultural or for policy reasons. The survey respondents form a culturally unusual group and are of a higher socioeconomic status; they may be more similar to countries like Norway (where parents are happier) than to countries like the United States (where they are less happy).

(also, we should at least consider the Caplanian perspective that people more informed about genetics will be happier parents, since they'll be less neurotic about the effect of their parenting styles.)

The View From Hell blog argues that the discrepancy between the direct question ("Are you happy to have kids?") and the indirect one ("How happy are you?", compared across parents vs. childless people) is pure self-deception; children suck, but parents refuse to admit it. I haven't looked in depth at the study they cite, which purports to show that the more you prime parents with descriptions of the burdens of parenthood, the more great they insist everything is. But I wonder about the philosophical foundations we should be using here. There's happiness, and there's *happiness*: I am happy to be giving money to charity and making the world a better place, but I don't think my self-reported life satisfaction would be noticeably higher after a big donation. It might even be lower if it cut into my luxury consumption. The wanting/liking/approving trichotomy may also be relevant.

People were happier with their decision to have children if they were (all results are binomial correlations and highly significant even after correction): more gender-conforming (0.14), had fewer thoughts about maybe being transgender (0.20), were more right-wing (0.10), considered themselves more moral people (0.15), were less autistic (0.12), were less extraverted (0.10), were more emotionally stable (0.15), and were more agreeable (0.13). All of these effects were very small compared to the generally high level of happiness at having children, no matter who you were and what your personality was like.

I included this survey question because I'm considering whether or not to have kids. Even though the survey only reinforced the (confusing) results of past research, I still find it helpful. After all, a lot of the survey-takers here are pretty skeptical of other aspects of traditional lifestyles: monogamy, gender norms, religion, etc. It’s impressive how strongly approval of parenting survives even in this weird a population; I consider this a new and exciting fact beyond the ones established by previous studies.

As always, if you want to double-check these results or analyze them further, you can download the data as an .xlsx file. I have removed the data of a few people who did not want their answers to be public, so you may not get exactly the same numbers I did, but they should be pretty close.

Practically-A-Book Review: Luna Whitepaper

They say money can't buy love. But that was the bad old days of fiat money. Now there are dozens of love-based cryptocurrencies – LoveCoin, CupidCoin, Erosium, Nubilo – with market caps in the mid nine-figures. The 17-year-old genius behind CupidCoin just bought the state of Tennessee. You think I'm joking, but can you be sure? How weird is "too weird to be true" these days, and how confident are you in your answer?

Case in point: Luna, which bills itself as blockchain-optimized dating. They caught my attention by hiring Aella, previously featured on this blog for her adventures taking LSD megadoses weekly for a year. They kept it with their cutesy story about how the name "Luna" comes from founder Andre Ornish's first word – adorable, until you consider that any baby whose first word is in Latin is definitely possessed. And they maintained it because – well, goodness knows we need new dating sites now that OKCupid has devolved into an off-brand Tinder clone. So let's look through the white paper and see what they've got.

Most dating sites suffer from attention imbalance: men scrounge around for anyone willing to acknowledge their existence; women get inundated with countless desperate messages they don't want. Luna solves this by making attention a commodity tradeable on the free market. Users who want to catch someone else's attention can bid the local cryptocurrency, Stars, to get their message to the top of another user's queue; all Stars spent in this way go to the user receiving the message. Stars can be bought with dollars *and vice versa*, so popular users can actually earn money reading all the messages sent to them.

This system has some pretty powerful advantages. Market forces are the known solution to the problem of connecting resources to their highest-value use. So if you treat user attention as a resource you can trust the market to allocate it optimally – in this case, to the guy who's just realized he's your soulmate, rather than the guy who's spamming everyone with five dick pics.

But everywhere this solution is tried, it runs up against its one great weakness – rich people with mild preferences can outbid poor people with strong ones. I can't predict how this particular market will clear, but this doesn't *necessarily* have to be a big problem here. If the market rate for a certain user's messages were $1, then even the poorest person could afford to send a message to a potential soulmate. And even a well-off person might hesitate to send out a hundred messages a day, every day. And if he didn't? Well, getting paid $100/day to read messages on a dating site doesn't sound like the *worst* outcome. Also, really good information about preferences in exchange for a biased system that favors the wealthy has been the deal Capitalism has been offering since Adam Smith first put quill to paper; it seems kind of weird to back out now.

A more practical issue: how long before someone finds a photo of a supermodel, limits their profile to "I AM A NYMPHOMANIAC", and watches everyone trip over themselves to send paid messages? Luna alludes to vague plans to "verify" profiles, which could mean anything from "you have to Photoshop a picture halfway convincingly" to "you have to get an actual pretty girl to help with your scam". Neither of these seem like too high a bar. Better is their offer to provide data, including how often users respond to messages and how often users meet with other users:

> When choosing to attach Stars to his message, Bob should receive information such as the number of unread messages in Alice’s queue, an internally calculated reply quality indicator, and confirmation on whether Alice’s account is verified.

I’m still skeptical. I have bots pretending to be pretty women try to friend me on Facebook something like once a week, even though I have no idea what their endgame is or how this results in them making money. If Luna gives a real incentive for the scam, they’re going to have to beat Facebook pretty handily if they want to succeed here.

More promising than any individual claim they make about how they’re going to fix things, is their claim that they’re *incentivized* to fix them. OKCupid famously wrote about Why You Should Never Pay For Online Dating, the answer being that it incentivizes dating sites to keep you single – after all, the longer you’re single, the longer you’ll keeping spending money on dating sites. Even if that sounds a little cartoon-villainish, at the very least it doesn’t incentivize sites to do a *good* job matching you up. Luna claims that their model gives them a profit only when it succeeds:

> At Luna, we intend to structure the token economy in such a way that our system is rewarded when users achieve their goals, thus aligning our own incentives with those of our users and ensuring that all data, AI, and machine learning technology will be used to actually connect people…the approach consists of two parts:

1. Fees which comprise Luna’s revenue only occur in the case of successful communication. As described in 3.1, when a user receives and reads a message boosted with Stars, they also receive the Stars used to boost that message. Luna intends to take a small fee for this transaction, but only if the recipient responds to the message within a window of a number of days yet to be determined. If the recipient does not respond, or only responds after more than this number of days, this fee will be re-paid to the sender. The number of Stars transferred to the recipient, however, will remain the same, whether they respond to the message or not. In this way Luna’s financial incentives will be aligned with users’ goals at Stage IV in the exchanging of messages.

2. Possibility of tipping in case of successful offline dates. Another way to provide incentive for Luna to help achieve its users’ goals is to allow users to tip the platform after the achievement of Stage V in the completion of a successful date. As described in 3.2.4, we intend to make feedback polls available after dates. Once users have rated their experience, Luna will then allow them to choose whether to leave a tip of their choice in the form of Stars. As this is a voluntary option, it should have no effect on user feedback. Tipping a platform is an infeasible idea in the context of currently existing dating apps; however, the free and direct-to-user benefits of Luna may register to users as something more resembling the mechanisms of Wikipedia: a free, friendly, and user-contributed service, rather than a platform like Match.com, which can feel exploitative. A tipping option may thus encourage a feeling of alliance with Luna in the user.

In this way, rather than recreating disparities which exist between the goals of current dating platforms and

> their users, Luna’s financial incentives and user goals will coincide.

I can imagine all sorts of horrible misalignments between maximizing-number-of-responded-to-messages and maximizing user satisfaction, but for now I’ll just admit this seems nice and I appreciate the effort.

Luna’s last major promise is to use cutting-edge machine-learning techniques to come up with a good match algorithm:

> Despite significant technological advances in information processing, storage, and retrieval, online dating has yet to optimally integrate machine learning for the user’s benefit. A typical ML task for online dating might be to predict the level of compatibility between two users from a given set of input data, thus predicting for example whether one user is likely to respond to another user’s message […]
>
> Luna may adopt a collaborative filtering algorithm developed by Dr. Kang Zhao. In addition, Luna may use advanced NLP techniques in conjunction with IBM Watson to integrate additional information from the contents of messages sent in-app, as well as from social media sources such as Twitter, if users choose to provide that information.

They’re right that this seems like a perfect use case for machine learning techniques, and that this possibility has been woefully underexplored. With all due respect to OKCupid’s excellent match algorithm (my girlfriend played with it one day and found that her two highest matches in the entire world were me and her ex), there’s a lot of room for improvement here. I find the idea of letting users link their social media accounts to provide more data really fascinating, and this reassures me that the attempts at incentive-alignment above really do have them thinking about how they can do better.

One part of the white paper I still don’t understand: why is it on a blockchain? They write:

> For us to ever find out [how to design a match algorithm that really increases human happiness], we are going to require an open data ecosystem around computer dating. Blockchain is an integral part of that – it’s what pays the bills to do the science and, in the case of Luna, it nicely and accurately solves one of the key problems in the computer dating arena: cut-and-paste messages spammed over huge numbers of people, resulting in an ever-lower number of good quality genuinely interested messages, hidden in an ever larger sea of dating spam. Just getting rid of that dynamic once and for all would be a great result, but I think that Luna offers far, far more.
>
> By establishing the decentralized paradigm in dating, Luna helps to remake dating culture. Luna is not a service or a place, like Tinder or a bar. Luna is a method, and a method which can be continually improved using techniques like A/B testing, until it is genuinely producing better lives for people. Because blockchain techniques allow for sophisticated tools to be developed to align economic interests between (say) search algorithm designers and individual users, or between users and other users who don’t like spam (i.e. everybody), the possibility exists to not only solve the questionable agency of the current generation of dating app providers, but to create positive agency to do something really, really new. We could pay the best

> people in the world to design algorithms to match other people, and make them happy.

I don’t know what they mean by “open data ecosystem”. I assume they’re not going to let everybody see all the data – I don’t necessarily want my parents to be able to know who I just sent a message to. But if not, how does “open data” help other people design match engines? Why is their crypto token more efficient than paying for Second Life in Linden Dollars, or any of the other silly token currencies that have existed forever on the Internet?

So what is blockchain doing for them? The null hypothesis might be: the same thing it does for Long Island Blockchain Tea and half a million other scams. I have hopes that Luna isn’t a scam. It employs some people I know and trust. The incentives and economics of their product seem very well-thought-through – so much so that if it’s a scam somebody else ought to be doing the same thing for real. And they’re promoting GiveWell and other effective charities in a way that suggests they want a lot of the money to go to altruism anyway. These are…signs. But I’ve been told you can’t be too paranoid in this area these days.

But I really *do* hope Luna isn’t a scam. Because if it’s real, it represents everything good about Silicon Valley. Some people use Intellect to wrest a secret from Nature: an elegant reduction of the chaos of human interaction into comprehensible and exploitable principles. To test their prize they build a Sampo, a machine churning out a hundred varieties of human happiness – from loving marriages to ecstatic sex to just sitting on the couch cuddling on rainy days. They give it to the public *gratis*. In the process they all get super rich and donate the money to curing malaria, good compounding upon good. Also, the whole thing is done in a weird and pointlessly-complicated format that adds nothing except a giant middle finger aimed at government regulators. *What could be more beautiful than this?*

One last thought on the blockchain issue: whenever I study intentional communities, I’m struck by how little the community’s principles matter, compared to the brute fact of it being an intentional community. An anarchist commune may have some spectacularly brilliant collaborative dispute-solving mechanism, but none of that matters, because the people involved will be *the sorts of people who would join an anarchist commune*, ie ridiculous and completely ungovernable.

So the most interesting and distinguishing feature of Luna, at least to start with, might not be the tokens, or the incentives, or the machine learning. It might be that *it’s a place you can go to meet the sort of people who want to date on the blockchain.* I could make a lot of cheap jokes here, but whatever weird hyperplanes through categoryspace further the difficult and desperate project of human-seeking-human are good and worthwhile in my book. I hope that lots of libertarian women find lots of security-conscious men and make lots of beautiful, high-price-volatility babies.

Conflict Vs. Mistake

Jacobite – which is apparently still a real magazine and not a one-off gag making fun of *Jacobin* – summarizes their article Under-Theorizing Government as "You'll never hear the terms 'principal-agent problem,' 'rent-seeking,' or 'aligning incentives' from socialists. That's because they expect ideology to solve all practical considerations of governance."

There have been some really weird and poorly-informed socialist critiques of public choice theory lately, and this article generalizes from those to a claim that Marxists just don't like considering the hard technical question of how to design a good government. This would explain why their own governments so often fail. Also why, whenever existing governments are bad, Marxists immediately jump to the conclusion that they must be run by evil people who want them to be bad on purpose.

In trying to think of how a Marxist might respond to this attack, I thought of commenter no_bear_so_low's conflict vs. mistake dichotomy (itself related to the three perspectives of sociology). To massively oversimplify:

Mistake theorists treat politics as science, engineering, or medicine. The State is diseased. We're all doctors, standing around arguing over the best diagnosis and cure. Some of us have good ideas, others have bad ideas that wouldn't help, or that would cause too many side effects.

Conflict theorists treat politics as war. Different blocs with different interests are forever fighting to determine whether the State exists to enrich the Elites or to help the People.

Mistake theorists view debate as essential. We all bring different forms of expertise to the table, and once we all understand the whole situation, we can use wisdom-of-crowds to converge on the treatment plan that best fits the need of our mutual patient, the State. Who wins on any particular issue is less important creating an environment where truth can generally prevail over the long term.

Conflict theorists view debate as having a minor clarifying role at best. You can "debate" with your boss over whether or not you get a raise, but only with the shared understanding that you're naturally on opposite sides, and the "winner" will be based less on objective moral principles than on how much power each of you has. If your boss appeals too many times to objective moral principles, he's probably offering you a crappy deal.

Mistake theorists treat different sides as symmetrical. There's the side that wants to increase the interest rate, and the side that wants to decrease it. Both sides have about the same number of people. Both sides include some trustworthy experts and some loudmouth trolls. Both sides are equally motivated by trying to get a good economy. The only interesting difference is which one turns out (after all the statistics have been double-checked and all the relevant points have been debated) to be right about the matter at hand.

Conflict theorists treat the asymmetry of sides as their first and most important principle. The Elites are few in number, but have lots of money and influence. The People are many but poor – yet their spirit is indomitable and their hearts are true. The Elites' strategy will always be to sow dissent and confusion; the People's strategy must be to remain united. Politics is won or lost by how well each side plays its respective hand.

Mistake theorists love worrying about the complicated and paradoxical effects of social engineering. Did you know that anti-drug programs in school actually *increase* drug use? Did you know that many studies find raising the minimum wage *hurts* the poor? Did you know that executing criminals actually costs *more* money than imprisoning them for life? This is why we can't trust our intuitions about policy, and we need to have lots of research and debate, and eventually trust what the scientific authorities tell us.

Conflict theorists think this is more often a convenient excuse than a real problem. The Elites get giant yachts, and the People are starving to death on the streets. And as soon as somebody says that maybe we should take a little bit of the Elites' money to feed the People, some Elite shill comes around with a glossy PowerPoint presentation explaining why *actually* this would cause the Yellowstone supervolcano to erupt and kill everybody. And just enough People believe this that nobody ever gets around to achieving economic justice, and the Elites buy even bigger yachts, and the People keep starving.

Mistake theorists think you can save the world by increasing intelligence. You make technocrats smart enough to determine the best policy. You make politicians smart enough to choose the right technocrats and implement their advice effectively. And you make voters smart enough to recognize the smartest politicians and sweep them into office.

Conflict theorists think you can save the world by increasing passion. The rich and powerful win because they already work together effectively; the poor and powerless will win only once they unite and stand up for themselves. You want activists tirelessly informing everybody of the important causes that they need to fight for. You want community organizers forming labor unions or youth groups. You want protesters ready on short notice whenever the enemy tries to pull a fast one. And you want voters show up every time, and who know which candidates are really fighting for the people vs. just astroturfed shills.

For a mistake theorist, passion is inadequate or even suspect. Wrong people can be just as loud as right people, sometimes louder. If two doctors are debating the right diagnosis in a difficult case, and the patient's crazy aunt hires someone to shout "IT'S LUPUS!" really loud in front of their office all day, that's not exactly helping matters. If a group of pro-lupus protesters block the entry to the hospital and refuse to let any of the staff in until the doctors agree to diagnose lupus, that's a disaster. All that passion does is use pressure or even threats to introduce bias into the important work of debate and analysis.

For a conflict theorist, intelligence is inadequate or even suspect. It doesn't take a supergenius to know that poor farm laborers working twelve hour days in the scorching heat deserve more than a $9/hour minimum wage when the CEO makes $9 million. The supergenius is the guy with the PowerPoint presentation saying this will make the Yellowstone supervolcano erupt.

Mistake theorists think that free speech and open debate are vital, the most important things. Imagine if your doctor said you needed a medication from Pfizer – but later you learned that Pfizer owned the hospital, and fired doctors who prescribed other companies' drugs, and that the local medical school refused to teach anything about non-Pfizer medications, and studies claiming Pfizer medications had side effects were ruthlessly suppressed. It would be a total farce, and you'd get out of that hospital as soon as possible into one that allowed all viewpoints.

Conflict theorists think of free speech and open debate about the same way a 1950s Bircher would treat avowed Soviet agents coming into neighborhoods and trying to convince people of the merits of Communism. Or the way the average infantryman would think of enemy planes dropping pamphlets saying “YOU CANNOT WIN, SURRENDER NOW”. Anybody who says it’s good to let the enemy walk in and promote enemy ideas is probably an enemy agent.

Mistake theorists think it’s silly to complain about George Soros, or the Koch brothers. The important thing is to evaluate the arguments; it doesn’t matter who developed them.

Conflict theorists think that stopping George Soros / the Koch brothers is the most important thing in the world. Also, they’re going to send me angry messages saying I’m totally unfair to equate righteous crusaders for the People like George Soros / the Koch brothers with evil selfish arch-Elites like the Koch brothers / George Soros.

Mistake theorists think racism is a cognitive bias. White racists have mistakenly inferred that black people are dumber or more criminal. Mistake theorists find narratives about racism useful because they’re a sort of ur-mistake that helps explain how people could make otherwise inexplicable mistakes, like electing Donald Trump or opposing [preferred policy].

Conflict theorists think racism is a conflict between races. White racists aren’t suffering from a cognitive bias, and they’re not mistaken about anything: they’re correct that white supremacy puts them on top, and hoping to stay there. Conflict theorists find narratives about racism useful because they help explain otherwise inexplicable alliances, like why working-class white people have allied with rich white capitalists.

When mistake theorists criticize democracy, it’s because it gives too much power to the average person – who isn’t very smart, and who tends to do things like vote against carbon taxes because they don’t believe in global warming. They fantasize about a technocracy in which informed experts can pursue policy insulated from the vagaries of the electorate.

When conflict theorists criticize democracy, it’s because it doesn’t give enough power to the average person – special interests can buy elections, or convince representatives to betray campaign promises in exchange for cash. They fantasize about a Revolution in which their side rises up, destroys the power of the other side, and wins once and for all.

Mistake theorists think a Revolution is stupid. After the proletariat (or the True Patriotic Americans, or whoever) have seized power, they’re still faced with the same set of policy problems we have today, and no additional options. Communism is intellectually bankrupt since it has no good policy prescriptions for a communist state. If it did have good policy prescriptions for a communist state, we could test and implement those policies now, without a revolution. Karl Marx could have saved everyone a lot of trouble by being Bernie Sanders instead.

Conflict theorists think a technocracy is stupid. Whatever the right policy package is, the powerful will never let anyone implement it. Either they’ll bribe the technocrats to parrot their own preferences, or they’ll prevent their recommendations from carrying any force. The only way around this is to organize the powerless to defeat the powerful by force – after which a technocracy will be unnecessary. Bernie Sanders could have saved himself a lot of trouble by realizing everything was rigged against him from the start and becoming Karl Marx.

Mistake theorists naturally think conflict theorists are *making a mistake*. On the object level, they're not smart enough to realize that new trade deals are for the good of all, or that smashing the state would actually lead to mass famine and disaster. But on the more fundamental level, the conflict theorists don't understand the Principle of Charity, or Hanlon's Razor of "never attribute to malice what can be better explained by stupidity". They're stuck at some kind of troglodyte first-square-of-the-glowing-brain-meme level where they think forming mobs and smashing things can solve incredibly complicated social engineering problems. The correct response is to teach them Philosophy 101.

(This is the Jacobite article above. It accuses Marxists of *just not understanding the relevant theories*. It's saying that there's all this great academic work about how to design a government, and Marxists are too stupid to look into it. It's so easy to picture one doctor savaging another: "Did you even *bother* to study Ingerstein's latest paper on neuroimmunology before you inflicted your idiotic opinions about this case on us?")

Conflict theorists naturally think mistake theorists are *the enemy in their conflict*. On the object level, maybe they're directly working for the Koch Brothers or the American Enterprise Institute or whoever. But on the more fundamental level, they've become part of a class that's more interested in protecting its own privileges than in helping the poor or working for the good of all. The best that can be said about the best of them is that they're trying to protect their own neutrality, unaware that in the struggle between the powerful and the powerless neutrality always favors the powerful. The correct response is to crush them.

What would the conflict theorist argument against the Jacobite piece look like? Take a second to actually think about this. Is it similar to what I'm writing right now – an explanation of conflict vs. mistake theory, and a defense of how conflict theory actually describes the world better than mistake theory does?

No. It's the Baffler's article saying that public choice theory is racist, and if you believe it you're a white supremacist. If this wasn't your guess, you still don't understand that *conflict theorists aren't mistake theorists who just have a different theory about what the mistake is*. They're not going to respond to your criticism by politely explaining why you're incorrect.

Is this uncharitable? I'm not sure. There's a meta-level problem in trying to understand the position "don't try to understand other positions and engage with them on their own terms" and engage with it on its own terms. If you succeed, you've failed, and if you fail, you've succeeded. I am pretty sure it would be wrong to "steelman" conflict theory into a nice cooperative explanation of how we all need to join together, realize that conflict theory is objectively the correct way to think, and then use this insight to help cure our mutual patient, the State.

So if this model has any explanatory power, what do we do with it?

Consider a further distinction between easy and hard mistake theorists. Easy mistake theorists think that all our problems come from very stupid people making very simple mistakes; dumb people deny the evidence about global warming; smart people don’t. Hard mistake theorists think that the questions involved are really complicated and require more evidence than we’ve been able to collect so far – the weird morass of conflicting minimum wage studies is a good example here. Obviously some questions are easier than others, but the disposition to view questions as hard or easy in general seems to separate into different people and schools of thought.

(Maybe there’s a further distinction between easy and hard conflict theorists. Easy conflict theorists think that all our problems come from cartoon-villain caricatures wanting very evil things; bad people want to kill brown people and steal their oil, good people want world peace and tolerance. Hard conflict theorists think that our problems come from clashes between differing but comprehensible worldviews – for example, people who want to lift people out of poverty through spreading modern efficient egalitarian industrial civilization, versus people who want to preserve traditional cultures with all their thorns and prickles. Obviously some moral conflicts are more black-and-white than others, but again, some people seem more inclined than others to use one of these models.)

This blog has formerly been Hard Mistake Theory Central, *except* that I think I previously treated conflict theorists as making an Easy Mistake. I think I was really doing the “I guess you don’t understand Philosophy 101 and realize everyone has to be charitable to each other” thing. This was wrong of me. I don’t know how excusable it was and I’m interested in seeing how many comments here are “This is super obvious” vs. “I never thought about this consciously and I think I’ve just been misunderstanding other people as behaving inexplicably badly my whole life”. But people have previously noticed that this blog is good at attracting representation from all across the political spectrum *except*

Marxists. Maybe that’s related to treating every position except theirs with respect, and appreciating conflict theory better would fix that. I don’t know. It could be worth a shot.

Right now I think conflict theory is probably a less helpful way of viewing the world in general than mistake theory. But obviously both can be true in parts and reality can be way more complicated than either. Maybe some future posts on this, which would have to explore issues like normative vs. descriptive, where tribalism fits in here, and “the myth of the rational voter”. But overall I’m less sure of myself than before and think this deserves more treatment as a hard case that needs to be argued in more specific situations. Certainly “everyone in government is already a good person, and just has to be convinced of the right facts” is looking less plausible these days. At the very least, if I want to convince other people to my position here, I actually have to convince them – instead of using the classic Easy Mistake Theorist tactic of “smh that people still believe this stuff in the Year Of Our Lord 2018” repeated over and over again.

SSC Survey Data On Models Of Political Conflict

There were a lot of good comments on yesterday’s conflict vs. mistake post. Some were very appropriate challenges: for example, doesn’t public choice theory itself assume conflict between special interests? And didn’t Marxism start off with a dry incentive-based explanation for why capitalists have to do what they do and how the incentive landscape needs to change? I want to explore these questions further – but first, some data from the SSC survey showing that the distinction does capture something real and important.

No questions really matched the conflict/mistake theory distinction, but one of the closest was POLITICAL DISAGREEMENT I: “Which of these plays a bigger role in explaining why some people are wrong about politics – intellectual failure, or moral failure?” This isn’t quite the way I would frame it now – but it’ll do for our purposes.

About 80% of blog readers selected intellectual failure, compared to only 60% of Mechanical Turk users. I have some reservations about the accuracy of the Turk survey, but this matches my predictions. I would expect SSC readers to be selected for mistake theory more than for any particular political position.

Percent Of People, By Political Affilitation, Who Think Intellectual Failure Explains Bad Politics
Alt-right: 65%
Conservative: 72%
Liberal: 81%
Libertarian: 87%
Marxist: 62%

Neoreactionary: 77%
Social democratic: 74%

Marxists were most likely to believe their opponents were guilty of moral failure, and libertarians most likely to believe intellectual failure, with everyone else somewhere in between. There was much less distinction between right and left compared to internal division between different subgroups of rightists and leftists.

There were a few other questions that I thought might tap into the same underlying attitude:

POLITICAL DISAGREEMENT II
Which of these best describes your opinion of people with different political beliefs than you, but still within the normal range? EG Democrats if you are a Republican, Tories if you are Labour, etc.

0. They might be right about some things, I can’t be sure of my position.
1. They seem pretty wrong, but they make understandable mistakes and are probably mostly decent people
2. They seem pretty wrong, and their mistakes seem incomprehensible, but some of them might be okay, I guess
3. Inexcusably stupid or downright evil

POLITICAL DISAGREEMENT III
Which of these best describes your opinion of extremists with different political beliefs than you, outside the normal range? E.g. fascists, Stalinists, etc.

0. They might be right about some things, I can’t be sure of my position.
1. They seem pretty wrong, but they make understandable mistakes and are probably mostly decent people
2. They seem pretty wrong, and their mistakes seem

incomprehensible, but some of them might be okay, I guess
3. Inexcusably stupid or downright evil

THE SYSTEM
The current economic and political system…

0. Is a good start that needs to be fine-tuned
1. Is fundamentally bad and needs to be destroyed

Aside from these sorts of questions, there were also two categorization questions that I thought *might* get at the same underlying attitude. These asked respondents to choose between two categorization systems for three ambiguous situations. One system categorized them based on what principle was applied, the other based on which side won. These were:

NAZI CATEGORY
Consider three situations.

A) Nazis beat up some minorities.
B) Nazis hold a peaceful demonstration, nobody stops them, it goes well, and they get a lot of publicity.
C) Some concerned citizens beat up the Nazis.

Of these three possibilities, which two seem most similar?

0. A and C, because people resort to violence.
1. A and B, because the Nazis win.

MARRIAGE CATEGORY
Consider three situations.

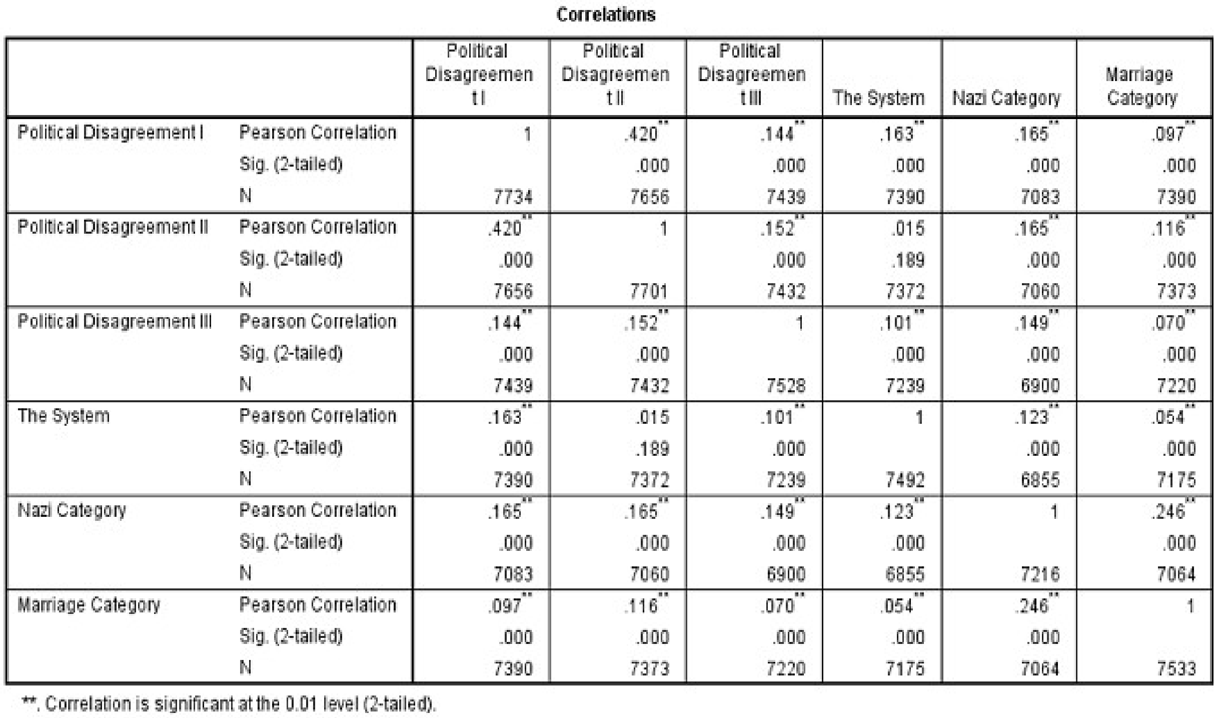
A) In a country where homosexuality is illegal, a government clerk participates in civil disobedience and marries gay people anyway.
B) In a country where homosexuality is legal, a government

clerk follows the law and marries gay people.
C) In a country where homosexuality is legal, a government clerk participates in civil disobedience and refuses to marry gay people.

Of these three possibilities, which two seem most similar?

0. A and C, because people are participating in civil disobedience.
1. A and B, because gay people get to marry.

I assigned numbers to all of these such that lower numbers represented mistake-theory answers and higher ones represented conflict-theory answers. Correlations were generally positive and significant, but unremarkable:



Correlations

		Political Disagreement I	Political Disagreement II	Political Disagreement III	The System	Nazi Category	Marriage Category
Political Disagreement I	Pearson Correlation	1	.420**	.144**	.163**	.165**	.097**
	Sig. (2-tailed)		.000	.000	.000	.000	.000
	N	7734	7656	7439	7390	7083	7390
Political Disagreement II	Pearson Correlation	.420**	1	.152**	.015	.165**	.116**
	Sig. (2-tailed)	.000		.000	.189	.000	.000
	N	7656	7701	7432	7372	7060	7373
Political Disagreement III	Pearson Correlation	.144**	.152**	1	.101**	.149**	.070**
	Sig. (2-tailed)	.000	.000		.000	.000	.000
	N	7439	7432	7528	7239	6900	7220
The System	Pearson Correlation	.163**	.015	.101**	1	.123**	.054**
	Sig. (2-tailed)	.000	.189	.000		.000	.000
	N	7390	7372	7239	7492	6855	7175
Nazi Category	Pearson Correlation	.165**	.165**	.149**	.123**	1	.246**
	Sig. (2-tailed)	.000	.000	.000	.000		.000
	N	7083	7060	6900	6855	7216	7064
Marriage Category	Pearson Correlation	.097**	.116**	.070**	.054**	.246**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	7390	7373	7220	7175	7064	7533

**. Correlation is significant at the 0.01 level (2-tailed).

(for a sense of how high to expect correlations to be – the one between feminism and support for gay marriage was 0.45, between life satisfaction and social skills was 0.3, and between reported IQ and financial success, 0.15)

Perhaps this is bad and wrong of me, but I combined all of them into one measure of average tendency to give the conflict-theory rather than mistake-theory answer to these questions, ranging from 0 (most mistake theorist) to 1 (most conflict theorist). The ranking by political affiliation is pretty much the same as before:

Combined Conflict Theory Measure By Political Affiliation, Ad Hoc Measurement
Alt-right: 0.43
Conservative: 0.30
Liberal: 0.26
Libertarian: 0.27
Marxist: 0.54
Neoreactionary: 0.34
Social democratic: 0.35

A different method of combining them without the two categorization questions (which I worried might have been contaminated by object-level political ideology) was broadly similar.

I also tried a factor analysis (which I am very bad at, don't trust me here) and got two factors, one of which looked like conflict theoriness, and the other of which looked like a categorization factor determining how people answered the last two category-related questions. The ranking by political affiliation for the conflict theoriness factor was:

Combined Conflict Theory Measure By Political Affiliation, Factor Analysis Measurement
Alt-right: 0.45
Conservative: 0.05
Liberal: -0.12
Libertarian: -0.07
Marxist: 0.30

Neoreactionary: 0.04
Social democratic: 0.15

Although Marxists and alt-righters have switched places, overall this seems like the same general picture.

Some other demographic variables that did or didn’t affect level of conflict theorist responses (similar whether I used the hacked-together average or the factor analysis results, all significant at such low p-values that I really don’t want to get into more fights in the comments over significance thresholds, replicate it yourself if you don’t believe me):

1. Autism did not affect any of these questions at all, totally contrary to my pre-registered predictions.

2. People in worse financial situations were very slightly more conflict theorist , r = 0.06.

3. Respondents who thought they were more moral people were more conflict theorist, r = 0.12.

4. More neurotic and extraverted people were slightly more conflict theorist (r = 0.10); no other Big Five traits really mattered.

5. No race or gender seemed much more conflict theorist than any other, but sample sizes were low.

Overall this broadly confirmed my original suspicion that Marxists and the alt-right differ from liberals on some dimension of conflict vs. mistake theory, but it didn’t shed a lot of light into the exact structure of the difference or what other factors might go into it.

If you want to double-check these results or analyze them further, you can download the data as .xlsx or .csv. Some people have complained of weird problems in the csv format and I recommend the xlsx if at all possible. I have removed the data of a few people who did not want their answers to be public, so you may not get exactly the same numbers I did, but they should be pretty close.

Highlights From The Comments On Conflict Vs. Mistake

I.

Thanks to everyone who commented on the posts about conflict and mistake theory.

aciddc writes:

> I'm a leftist (and I guess a Marxist in the same sense I guess I'm a Darwinist despite knowing evolutionary theory has passed him by) fan of this blog. I've thought about this "conflict theory vs. mistake theory" dichotomy a lot, though I've been thinking of it as what distinguishes "leftists" from "liberals."
>
> I went through the list of "conflict theory says X, mistake theory says Y" nodding my head and hoping that you and everybody else reading it had the same impression as me -that both theories are important and valuable frameworks through which to view the world. There are definitely common interests that everybody in America shares, and there are definitely some pretty significant conflicts of interests as well.
>
> The reason that I do identify as a leftist and sometimes feel like an evangelist for conflict theory is that I get the impression that most people don't even have conflict theory in their mental framework. Leftists all understand what the "we're all in this together" liberal viewpoint is, while even incredibly smart and on it liberals like yourself can go for a long time without even thinking about the "politics is the clash of interest groups with conflicts of interest" leftist viewpoint.

I’m starting with this one, before I get to all the criticism and objections, to insist that there is some core worth saving here. To everyone who says this was obvious, I can only plead that you listen to all the people saying it wasn’t obvious to them at all. There are probably a lot of things wrong with how I described it, and the remaining comments will go over a lot of them, but I can only make the excuse that I’m groping towards something useful without any of the people who claim to already understand it perfectly helping me at all.

HeelBearCub writes:

> First off, I think Scott’s post is a good post that seems to make good points and takes steps on the path to true knowledge. But, as I read it I kept thinking about Indian wise men describing the elephant. Or rather, I pictured two amateur carpenters arguing about how wood is joined together, one being a “nail” theorist and the other a “screw” theorist. Now obviously, you build using wood or nails, but structures out in the real world are not built exclusively using one or the other. This seems to me to be a recurring failure of thinking on Scott’s part, honestly. The tendency to naturally think in binary terms (even while knowing this is incorrect).

The goal of those posts wasn’t to present a perfect understanding and tell people that was it, but to gesture at a concept that needed refining.

By analogy, should we ever talk about the left vs. right axis in politics? Isn’t that just a “dichotomy”? After all, almost nobody is 100% leftist on everything or 100% rightist on everything.

What about introverts and extraverts? There’s no 100% introvert or 100% extravert. Everyone is a combination of introverted in some situations and extraverted in others. That doesn’t mean we should never use these terms, or that psychologists are being too binary and dichotomizing. It means we establish the terms so that everyone knows what we’re talking about, and then discuss where everybody is on the axis in between them.

I tried to exaggerate conflict and mistake theory equally, making it obvious I was presenting caricature versions to be filled in later. I think I was at least equally unfair to mistake theorists – presenting them as believing there’s no such thing as selfishness, as thinking that surely tobacco companies just deny their products cause cancer because their CEOs have an genuinely different interpretation of health statistics. The point isn’t “mistake theory is good and normal and conflict theory is bad”. The point is that both look ridiculous at the extremes and everyone combines them in some (different) proportion in reality.

II.

Tumblr user unknought writes:

> Principal-agent problems, rent-seeking, and aligning incentives are things that socialists do talk about. Like, a lot. But even if they weren’t, it’s totally bizarre to represent these as mistake theory concepts. All three of these are concepts which are used to describe ways in which we don’t all want the same things, and how agents in positions of power whose goals don’t align with the common good can fuck things up for the rest of us. If conflict theory means anything, it means that.

On my hospital analogy for why mistake theorists like free speech:

> So like, if you learned that your doctor's recommendation was influenced by Pfizer owning the hospital and restricting doctors' choices and suppressing information about their medication… wouldn't that be an excellent reason not to be a mistake theorist in this case? Like, this is literally an example of where you can't trust expert opinion because things are being controlled by a powerful entity whose interests do not align with the common good. This is about as clear-cut an example of conflict theory getting it right as can be imagined.

They quote another Tumblr user who I'm not sure I have permission link directly, responding to my analogy of the shill scientist with a PowerPoint:

> In this scenario, even the strawman conflict theorist acknowledges that the people who believe the powerpoint are mistaken, and could be won over through reason and debate, thus undercutting the narrative scott is presenting in which strawmen 'conflict theorists' are generally uninterested in debate…the whole tendency to caricature opponents as anti-rational also plays in to his assumption that conflict theorists would be anti-intellectual, which ignores that in the context of a conflict, having intelligent people working on your side to win the conflict is obviously desirable…
>
> even if i do think scott amounts to the hypothetical "Elite shill" with a powerpoint saying yellowstone will erupt, proving his claims wrong on a technical level is still my first and most important priority if i want him to not be a successful shill. simply accusing him of being a

> shill, and pointing out the $1,627,000 which MIRI (which scott is affiliated with) has received from Peter Thiel would be useless if i couldn’t also demonstrate to people that he was incorrect. the insinuation that viewing a particular disagreement as a conflict of interest presupposes any possibility of engaging on a factual level is ludicrous. even if i think the “conflict theorist”/”mistake theorist” is just a cheap rhetorical trick to sell people on anti-democratic ideology, i still need to demonstrate that it’s not reflective of material reality.

(one correction: I am not affiliated with MIRI. I spent a few weeks doing a minor writing job for them five years ago; I have had no formal relationship since then besides donating money)

The three comments above all seem right to me, and seem like the best examples of how my intuitive concepts of “conflict theorist” and “mistake theorist” fail to be captured by the ways I described them.

How fatal is this objection? A lot of things aren’t accurately captured by the words we use to describe them. Consider again the analogy of the political spectrum. Someone might say “You say Republicans are conservative, but they don’t want to conserve the traditional public school system.” Or “You say libertarians freedom, but what about freedom from powerful corporations?” Yeah, okay, you got me there, sometimes words are not perfect 100% handles for complex concepts. There’s got to be a balance between having simple words that mean exactly what they say, and accurately tracking the weird subtleties of how people really associate.

A lot of this discussion is conflating “conflict vs. mistake” with “protester vs. wonk” and “Marxist vs. neoliberal”. Sometimes they’re being conflated by me, where I say something is conflict theorist when I mean Marxist, or vice versa. Other times they’re being conflated by the commenters, who say “You say X is conflict theorist, but Marxists don’t actually believe X!” Well, maybe that’s because Marxists aren’t 100% stereotypical conflict theorist. The word “libertarian” can shed light on Ron Paul, even though he supports government intervention in some areas.

Obviously this idea isn’t useful if it totally fails to correspond to existing politics. But I think there are more subtle (read: dangerously susceptible to just being overfitting) ways of understanding this that make sense of these contradictions. For example, conflict theorists can certainly engage in intellectual debate if it helps them win. That doesn’t mean there’s not a difference between people who choose their side based on intellectual debate, and people who engage in intellectual debate if it helps their side. Also, it’s unclear whether a conflict theorist thinks intellectual debate is better than equally persuasive propaganda, whereas a good mistake theorist definitely should.

Likewise, any mistake theorist who didn’t acknowledge that there are lots of self-interested parties would be an extraordinary idiot. The mistake theorist calls these “special interests”, which I suppose is a pretty-loaded term suggesting they’re a few weird exceptions to the rule of supporting the general interest. The question can’t be whether special interests exist – obviously they do. It can’t even be whether they’ve seized control of the government – I think this is something of a consensus position right now. Maybe it’s something like “Is there anything *other than* special interests?”

In a book review a while ago, I talked about figure-ground inversions:

> An example of what I mean, taken from politics: some people think of government as another name for the things we do together, like providing food to the hungry, or ensuring that old people have the health care they need. These people know that some politicians are corrupt, and sometimes the money actually goes to whoever's best at demanding pork, and the regulations sometimes favor whichever giant corporation has the best lobbyists. But this is viewed as a weird disease of the body politic, something that can be abstracted away as noise in the system.
>
> And then there are other people who think of government as a giant pork-distribution system, where obviously representatives and bureaucrats, incentivized in every way to support the forces that provide them with campaign funding and personal prestige, will take those incentives. Obviously they'll use the government to crush their enemies. Sometimes this system also involves the hungry getting food and the elderly getting medical care, as an epiphenomenon of its pork-distribution role, but this isn't particularly important and can be abstracted away as noise.
>
> I think I can go back and forth between these two models when I need to, but it's a weird switch of perspective, where the parts you view as noise in one model resolve into the essence of the other and vice versa.

I think any difference between mistake and conflict theorists on this issue has to be of the figure-ground inversion type. This doesn't have to be about the number of people who are vs. aren't special interests. It can be about the amount of each person's soul devoted to their special interest as rich/poor/white/black/rural/urban, vs. the amount of each person's soul devoted to loving truth and doing good. Framed this way, it sounds like we're pretty much screwed, except that free rider problems might be working in our favor here.

A related difference: because mistake theorists think there's some stable ground other than conflict, they picture themselves as potentially neutral referees (even better: Geneva Convention delegates) looking for ways to circumvent the conflict. Conflict theorists just want to win it.

Some examples of circumventing conflicts: free religion, free speech, federalism, reliance on scientific consensus. Free religion because instead of pushing for any one religion to win, you just create a system that defuses the threat of religious violence. Free speech for the same reason. Federalism because instead of saying that any one side wins, you just tell Side A to do it their way over there, and Side B to do it *their* way over *there* (even better: polycentric law). Reliance on scientific consensus, because instead of arguing over whether to have school vouchers or not, everyone just agrees to have unbiased scientists do a study and trust their findings.

I notice Jacobite's describes itself as "the post-political magazine". Politics is about having conflict. Mistake-theorists would love to become post-political, in the sense of circumventing all conflicts. Conflicts actually happening as conflicts is a failure, deadweight loss. This wouldn't mean that nobody has different interests. It would mean that those different interests play out in some formalized way that doesn't look conflict-y. Think of the Patchwork / charter city / seasteading dream, where there are lots of different polities and people vote with their feet for which one they prefer. Protests, bring-out-the-vote efforts, arguments – all have become obsolete. These ideas don't deny the existence of conflict – they just represent a desire to avoid it rather than win it. Conflict theorists could theoretically want to avoid conflict, especially if they think they'd lose. But most of the ones I've met think that avoiding conflict is a better deal for the enemy than for them, and so would rather just have it and see what happens.

I think this is why I think of public choice theory and its relatives as basically more mistake-y. It's not *just* that they sometimes say government doesn't work, get seized on by libertarians, and so get a bad reputation among Marxists. It's that they take this god's eye perspective of trying to micromanage the rules of political conflicts instead of winning them.

III.

Some other good comments:

Fluffy Buffalo comments:

> What struck me in your post was that the examples you gave for conflict theories all came from the Marxist perspective. While (cultural) Marxists may be the most obvious, unabashed conflict theorists these days, the behavior of the American right wing looks like they have their fair share of conflict theorists, and Republican tax and health care policy often smells more of an undeclared class warfare than of careful consideration of the pros and cons.

It's definitely true that there's a lot of right wing conflict theory and that my post mostly ignored this. They continue:

> This looks not like a fundamental question of what the world is really like, and more like a multi-player game theory problem, in particular a multi-player prisoner’s dilemma. It’s all fine and dandy – in fact, it’s probably the most constructive, helpful thing to do – to play “mistake theory” if everyone else is playing the same game, but if you have a sufficiently strong faction playing “conflict theory” (refusing to compromise, because everyone else is the devil), they have more success than they should. “Conflict theory” is like a bad Nash equilibrium, a self-fulfilling prophecy – if everyone behaves according to the diagnosis “It’s power-hungry, uncompromising people on the other side who cause the problem”, there will be no lack of power-hungry, uncompromising people on all sides, causing all sorts of problems.

David Friedman makes the really bizarre topsy-turvy argument that conflict theorists should be people who want minor tweaks to the existing system, and mistake theorists people who want to get rid of it entirely:

> A lot of whether you believe in mistake theory or conflict theory depends on how many mistakes you think there are, which comes down to how badly you think existing systems work. If you believe that government decisions are very far from optimal, that FDA regulation has little positive effect on how good drugs are, a huge negative effect on what drugs are developed and how expensive they are, and believe it not only from observation (Peltzman’s old article, for instance) but as what you would expect from the incentive structure, then you see the main problem as persuading the 99.9% of people who are worse off as a result. If you think government decisions are close to optimal, on the other hand, then

> all that is left to change is how they weight payoffs to different people, which is a conflict issue.

I guess this reinforces the point from above: taking some of these terms literally, or trying to reason with them, gets very different results from looking at how they operate on political coalitions.

From Jan Samohýl:

> I really enjoyed the post, the two worldviews are interesting. But as many people have noted, things are often not so clear-cut. I would like to point out some interesting connections.
>
> I think of myself being a more conflict-theorist, probably being on the left. But in many cases, I look at the world as mistake-theorist, having quite liberal views. To me, it's somehow related to the idea "expect the worst (conflict), hope for the best (mistake)". I am a big fan of direct democracy (which I see as a way to peacefully resolve innate conflict), and less of a fan of resolving politics through debate (which implicitly assumes the debaters are honest). To me, direct democracy is related to trust, as it is understood in computer security, for example. In computer security, trust is not given, it is earned and can be revoked at any time. You can only trust somebody who you don't have to trust. In computer security, we assume worst intentions (conflict), not the best ones (mistake). And so I believe direct democracy is needed as a baseline, because the politicians are not to be trusted (principal agent problem).

This is another weird topsy-turvy thing. I was thinking in terms where mistake theorists are afraid of direct democracy because any random demagogue can sabotage it, and trust more in “post-political” systems like the ones mentioned in Part I above.

From Nootropic Cormorant:

> A Marxist reader of your blog reporting in! I feel that this way of analyzing things is harmful (a charitable mistake) because ironically it conflict-theorizes the debate so that suspected conflict-theorists are automatically seen as beyond object-level discussion leaving you with no options but to act like one yourself. I would question whether this distinction is even relevant to people and ideologies rather than to situations and debates. Many of the commentariat have identified themselves as mistake-theorists, but I have to wonder whether this is because they abstain from the sin of conflict-reasoning or because it bothers them when the outgroup does it. This is some conflict-theorizing on my part. Also the willingness to discuss policy issues, aka the reformism vs. revolutionism axis, should not be conflated with this as one can go along it in a purely mistake-theorizing framework.

From Ben Wave:

> I count myself as a Marxist, as someone who adores your blog, and as someone who primarily uses what you’ve deemed mistake theory as my lens on the world. It’s naive to expect that everyone else would or should have the same terminal values as I do, or that everyone should agree on the relative role of government in bringing those values about. So I don’t. I don’t share the view of some of my comrades who celebrate death and misfortune upon the rich. I do share their desire to take power from the powerful, and create a world in which

> the weak have more power. I desire this independently of a desire to increase living standards/lifetimes/happiness over all, and I desire this for three reasons – the first is that I see high levels of inequality as an existential threat to society. The second is that I would not consider the current distribution fair, if I was to be incarnated into a human chosen at random. The third is that the more unequal society is, the less the ethical premise of capitalism (that it is right to reward those who fill market needs because they are equivalent to human needs and desires) is valid. Which I feel is important seeing as it is the dominant method by which we decide what gets done by the ensemble of human endeavour. As regards the conflict view, yes many comrades use that. I see it as largely unhelpful, but I'm sympathetic to their situation. Marginalised people are often harmed or killed by inequality (medical, prison and law-enforcement, unsafe conditions of work or home life), so conflict is the reality that some of them face. I find it more productive to try and cooperate with my opposites who like cooperation than to fight against my opposites who do not. Hopefully the coalition of those willing to negotiate can improve conditions globally. Gather with those who share your terminal values. Gain power through that unity, and use that power in negotiation. Stay open to new information to best pursue your goals. I guess that's the dream.

I find this interesting *except* that I can't figure out why marginalized people being harmed by inequality, bad medical care, bad working conditions, etc makes conflict "the reality" for them. Surely it means bad things are the reality, and we're back to the original question of whether bad things are mostly due to conflict or mistake?

From bosun_of_industry:

> You position these in opposition or at least orthogonal to each other but I think that’s missing both how often they are layered and/or modal. Moreover, the Marxist portion of this essay is red herring carried forward from the Jacobite article and does it a disservice because it lends it a uniformity to any given actor’s approach. Instead, a common split is people approaching policy issues as mistake theory when within-group and as conflict out-group. There’s distinct strategic reasons for this sort of behavior, especially within a winner takes all political system. Similarly there’s a strong history of making a public show of conflict theorizing while acting as a mistake theorist in private amongst politicians and activists alike and for similar reasons.

From Alkatyn:

> To go up a meta level, maybe conflict theorists are mistake theorists who have applied the same methodology to “how do we get our ideas implemented” and decided based on the empirical evidence that emotional appeals and group action are more effective than policy analysis.

The effective altruists have a lot of discussion on what the most effective way to push their preferred policies are, and usually they settle on some combination of appealing to academics and founding think-tank style organizations. I have noticed this working very well so far. This piece on how neoliberalism caught on is a really interesting example of doing this right and effectively. I suppose the conflict theory answer would be that this only works for policies that elites are already predisposed to like or at least not resist.

From John Schilling:

> One thing that seems to be getting lost in a number of places here, so I'll just address it this once. Mistake Theory doesn't require denying that Rich Plutocrats are genuinely trying to further enrich themselves at the expense of the poor and that the Poor are trying to tear down the rich out of spite righteously take back what is theirs. More broadly, it does not require denying that people will selfishly pursue their selfish interests to the detriment of others or of the whole of humanity, does not simplify to "if we are all rationalist altruists the right answer would be…" Mistake Theory would hold that in almost all real conflicts, the best outcome for everyone is a negotiated solution and that the relevant facts (including the balance of power between competing interest groups) makes the range of plausibly negotiated agreement reasonably narrow. So failing to sit down and quietly negotiate that agreement, instead escalating to pointless conflict, is usually a Mistake and often an Easy Mistake. Figuring out what to do about people who persist in making that Easy Mistake, is the sort of problem that often leads to Hard Mistakes and sometimes to solutions that look like Easy Conflict.

Butlerian writes:

> I think that the conflict/mistake distinction is being vastly overblown here. Scott is proposing some sort of fundamentally different personality type / rhetorical diachotomy between agents. I think these differences in behaviour can be explained in a much more proximate way, instead simply being answers of "Yes" or "No" to the question: "Do you believe that shills are amongst us RIGHT NOW"? If I am in a discussion where my answer to that question is "No", I will discuss in a mistake-theorist manner: belief that other participants are arguing in good faith, and personaly arguing in good

> faith, and sticking to truth-searching, and at least trying to be open to the possibility of changing my mind. If I am in a discussion where my answer to that question is “Yes”, I will discuss in a conflict-theorist manner: having no genuine interest in the merits of suspect participant’s arguments, reading them only with an eye to finding errors / weak-man-able faultlines, and refusing to allow myself to change my mind even in the face of apparently convincing evidence because I strongly auspect that the evidence is the contaminated product of Yudkowsky’s Clever Arguer. I would think that this is obvious. When Marxists / feminists / Nazis are in their safe-space forums, and are happy that everyone in the discussion is a like-minded Marxist / feminist / Nazi who genuinely shares their endgoal of a well-functioning communist / non-cis-hetero-patriachal/ white ethnostate society, they will have amongat themselves truth-seeking mistake-theoretical diacussions. It is only when they go out into enemy territory and find themselves surrounded by perfidious capitalist / misogynist / Judeobolshevik agents that they switch to conflict-theoretical mode.

This reduction to “belief in shills” as the most important difference between the theories is interesting.

Aside from the people who wear it on their sleeves, like PR people and lobbyists, I find myself practically never believing anyone is a “shill” in the strong sense – ie they don’t believe the arguments they’re making, but they make them anyway for some personal advantage. Even when people do very shill-ish stuff – like put out a biased paper for a think tank that promotes its agenda – I assume the think tank probably just recruited people who already believed in their agenda, and let their personal bias do the rest.

In fact, the idea of "shill" is really complicated here. People accuse me of being biased or selfish all the time if I make arguments that seem to benefit rich people, or Jews, or white people, or any other category I'm a member of. But by far the most important thing that affects my finances and social status right now is any threat to the guild-based nature of the medical profession. And I've consistently argued in favor of things like letting psychologists prescribe, even though this will cost me money and status way more directly than anything that happens regarding class or race. If people were interpreting shilling in its most literal sense, they would get surprised that I'm not writing 100% about how great medical guilds are, finally assume I was insane or some vanishingly rare moral saint, and everything about race or class would be so irrelevant to this calculation that it would never come up.

The fact that nobody thinks this way has to be tied into some sort of answer to the class warfare free rider problem. Nobody is a first-level shill literally pushing their own self-interest, but weird class-consciousness/ideology style forces give people biases (mostly based on race or class) that they push at least somewhat honestly. I think this is at least sort of in accord with some forms of orthodox Marxism.

But this sounds a lot like…mistake theory. If people push their policies because they're biased into thinking that's the morally correct thing to do, then surely solving their biases and convincing them otherwise could change their policy preferences. Is there anyone who *doesn't* believe this model? If so, what exactly are we talking about?

And from christhenottopher:

> All this talk about how Marxists don't frequent this blog and you go and make a post arguing for a Hegelian

> dialectic. Yes, yes of course this blog has always been on the side of Thesis. And certainly we must always beware the great enemy Antithesis. But let's end the essay by arguing that what we really need is Synthesis! The professor in college who taught me what dialectical reasoning was warned the class that once we understood this, we'd see dialectics everywhere. And once again he was proven right.

This had BETTER NOT BE TRUE. If all of this time all this incomprehensible stuff about dialectics was just basic "start understanding a concept by giving binary examples of opposite sides, then correct it and make it more sophisticated later", I am going to be SO ANGRY.

The Invention Of Moral Narrative

H/T Robin Hanson: Aeon's The Good Guy / Bad Guy Myth. "Pop culture today is obsessed with the battle between good and evil. Traditional folktales never were. What changed?"

The article claims almost every modern epic – superhero movies, Star Wars, Harry Potter, Lord of the Rings, etc – shares a similar plot. There are some good guys. There are some bad guys. They fight. The good guys win. The end.

The good guys are usually scrappy amateurs; the bad guys usually well-organized professionals with typical fascist precision. The good guys usually demonstrate a respect for human life and the bonds of friendship; the bad guys betray their citizens and their underlings with equal abandon. They gain their good guy or bad guy status by either following the universal law, or breaking it.

This is not exactly a scintillatingly original observation, *except* that the article claims you'll almost never find an example of this before 1700. Take the *Iliad*. Neither the Greeks nor Trojans are especially good nor villainous. The Trojans lose some points for kidnapping a woman, but the Greeks lose some points for killing and enslaving an entire city. Neither side is scrappier or more professional than the other. Neither seems to treat civilians better or demonstrate more loyalty. Exciting things happen, but telling the story of how Good triumphed over Evil was definitely not on Homer's mind. Nor was it on the mind of the authors of *Mahabharata*, the Norse sagas, Jack and the Beanstalk, et cetera.

Where ancient works do have good-vs-evil overtones, it's usually because we're reading more modern adaptations. Robin Hood doesn't rob from the rich to give to the poor until much later versions of the story; King Arthur's knights don't start out as especially good people and don't really fight a unified team of evildoers; the virtuous-Arthur-vs-evil-Mordred theme doesn't really dominate until Victorian retellings. Disney's *Hercules*, which reimagines Hades from perfectly-reasonable-underworld-god to classic-cartoon-villain is a striking late-20th-century example (I forgot that it ended with Hercules punching Hades so hard that he falls into the River Styx and gets pulled under by his own damned souls, not the most Hellenic of conclusions).

The article concludes this is because of nationalism. Nation-states wanted their soldiers to imagine themselves as fighting on the side of good, against innately-evil cartoon-villain enemies. This was so compelling a vision that it shaped culture from then on:

> Good guy/bad guy narratives might not possess any moral sophistication, but they do promote social stability, and they're useful for getting people to sign up for armies and fight in wars with other nations. Their values feel like morality, and the association with folklore and mythology lends them a patina of legitimacy, but still, they don't arise from a moral vision. They are rooted instead in a political vision, which is why they don't help us deliberate, or think more deeply about the meanings of our actions. Like the original Grimm stories, they're a political tool designed to bind nations together.
>
> When I talked with Andrea Pitzer, the author of One Long Night: A Global History of Concentration Camps (2017), about the rise of the idea that people on opposite sides of conflicts have different moral qualities, she told me: 'Three inventions collided to make concentration camps possible: barbed wire, automatic

> weapons, and the belief that whole categories of people should be locked up.' When we read, watch and tell stories of good guys warring against bad guys, we are essentially persuading ourselves that our opponents would not be fighting us, indeed they would not be on the other team at all, if they had any loyalty or valued human life. In short, we are rehearsing the idea that moral qualities belong to categories of people rather than individuals. It is the Grimms' and von Herder's vision taken to its logical nationalist conclusion that implies that 'categories of people should be locked up'.
>
> Watching Wonder Woman at the end of the 2017 movie give a speech about preemptively forgiving 'humanity' for all the inevitable offences of the Second World War, I was reminded yet again that stories of good guys and bad guys actively make a virtue of letting the home team in a conflict get away with any expedient atrocity.

What are we to think of this?

A quick check of the article's claims finds them kind of lacking. Robin Hood started stealing from the rich to give to the poor as early as the 1592 edition of his tale. And doesn't the Bible contains lots of good vs. evil? The author sweeps this under the rug by saying that the Israelites don't seem much more virtuous than the Canaanites, but one could argue that they're just not more 2018-virtuous; maybe 1000 BC-virtue *was* worshipping God and smashing idols. What about Armageddon? Ragnarok? Zoroastrianism? The Mayan Hero Twins? The very existence of Crusades seems to point to "all the good people get together and fight all the bad people, in the name of Goodness" being a recognizable suggestion. [EDIT: @scholars_stage lists some more here].

Are there *any* differences between the way ancients and moderns looked at this? Maybe modern stories seem more likely to have two clear sides (eg made up of multiple different people) separated by moral character. Villains (as opposed to monsters, or beings that are evil by their very nature) seem more modern. So does the idea of heroes as necessarily scrappy, and villains as necessarily well-organized. And just eyeballing it, modern stories seem to use this plot a lot more, and to have less deviation from the formula.

But even if that's true, the rise of nation-states seems like a uniquely bad explanation for the rise of these narratives. The past stories seem much more conducive to blind nationalism than our own. The amorality of the warriors in the Iliad manifested as total loyalty: Hector fought for Troy not because Troy was in the right, but because he was a Trojan. Achilles fought for Greece not because he believed in the Greek cause, but because that was his side and he was sticking to it. The whole *point* of the *Mahabharata* is the whole 'theirs not to question why, theirs not to reason why, theirs but to do and die' philosophy that makes for effective nationalist soldiering. In Jack and the Beanstalk, we root for Jack because he's human and we are Team Human. Jack can steal and kill whatever and whoever he wants and we'll excuse him. What more could a nationalist want?

In contrast, the whole point of modern good-vs-evil is that you should choose sides based on principle rather than loyalty. The article gets this exactly right in pointing out the literary motif of virtuous betrayal. We are expected to celebrate Darth Vader or Severus Snape virtuously betraying their dark overlords to help the good guys. In *Avatar*, the main character decides his entire *species* is wrong and joins weird aliens to try to kill them, and this is *good*. Compare to ancient myths, where Hector defecting to

Greece because the abduction of Helen was morally wrong is just totally unthinkable. This is a super-anti-nationalist way of thinking.

I suppose nationalists could make the very dangerous bargain of telling their soldiers to always fight for the good guys, then get really good propaganda to make sure they look like the good guys. And maybe this would make them fight harder than if they were just doing the old fight-for-your-own-side thing? But honestly, Achilles seems to have been fighting really hard. Is this whole convoluted process *really* easier than just telling people from the start to fight for their own side and not betray it?

Also do we *really* want to claim that concentration camps worked because the Nazis believed you should take principled positions based on moral values, instead of unquestioningly supporting your in-group? *Really?*

If nationalism didn't drive the (possibly) increasing prevalence of good-vs-evil stories, what did?

One theory: the broad democratization process marked by the shift from sword-based aristocratic armies to gun-based popular armies. Old stories celebrated warrior virtues – strength, loyalty, bravery. The new stories celebrate populist virtue – compassion, altruism, protecting Democracy. The new nation-states would have *liked* to maintain the warrior virtues, it just wasn't an option for them in the face of having to suddenly win the loyalty of a bunch of people they hadn't cared about before.

A second theory: this is just part of widening moral circles of concern. Pre-1700s, people were still at the point where slavery seemed like an okay idea. Maybe we didn't *have* the whole Care/Harm foundation down all that well. Once we got that, through whatever process of moral progress we got it from, having heroes who shared it started seeming more compelling.

A third theory: properly-written good-vs-evil stories are just *better*, in a memetic sense, but it took a long time to get the formula right. Coca-Cola is better than yak's milk, but you've got to invent it before you can enjoy it – and just having a vague cola-ish mix of spices in water doesn't count. But once you invent it, it spreads everywhere, and people throw out whatever they were doing before.

I realize this is pretty unsophisticated-sounding, but I'm basing this off of my continuing confusion over the rise of Christianity. Christianity came out of nowhere and had spread to 10% – 20% of the Roman population by the time Constantine made it official. And then it spread to Germany, England, Ireland, Scandinavia, Eastern Europe, Armenia, and Russia, mostly peacefully. Missionaries would come to the tribe of Hrothvalg The Bloody, they would politely ask him to ditch the War God and the Death God and so on in favor of Jesus and meekness, and as often as not *he would just say yes*. This is pretty astonishing even if you use colonialism as an excuse to dismiss the Christianization of the Americas, half of Africa, and a good bit of East Asia.

I've looked around for anyone who has a decent explanation of this, and as far as I can tell Christianity was just really appealing. People worshipped Thor or Zeus or whoever because that was what people in their ethnic group did, plus Thor/Zeus would smite them if they didn't. Faced with the idea of a God who was actually *good*, and could promise them eternity in Heaven, and who was *against bad things*, and never raped anybody and turned them into animals, everyone just agreed this was a better deal. I know this is a horrendously naive-sounding theory, but it’s the only one I’ve got.

And there seems to be a deep connection between Greek paganism and the narrative structure of the Iliad, and a deep connection between Christianity and the narrative structure of (eg) Harry Potter. Achilles fights for Greece because he’s Greek, and the pagan worships Zeus because he (the pagan) is Greek, and that’s all there is to it. But Harry Potter fights for Dumbledore and against Voldemort because the one is good and the other evil, and the Christian worships God and resists the Devil because the one is good and the other evil. Achilles and Hector wear their impressiveness on their sleeves, much like Zeus. Harry Potter is a seemingly ordinary and really quite weak guy who just *happens* to be fated to save everything through destiny, parentage, and the power of love/sacrifice, much like Jesus.

(this isn’t a joke – one could describe Luke Skywalker or Frodo Baggins the same way)

Maybe this good-vs-evil thing is just really attractive, and naturally replaces whatever was there before – but it’s just really hard to get exactly right. There was a 1500 year lag time between when people got the magic formula for religion (Zoroastrianism wasn’t good enough!) and when they got the magic formula for stories. Wasn’t the high-grade Colombian ultra-purified version of the good-vs-evil fantasy plot invented by Tolkien and CS Lewis sitting around in Oxford specifically trying to figure out how to translate Christianity into narrative form? Maybe this was more of an innovation than it seemed. Maybe they actually did the same thing that St. Paul or whoever did and created a totally new memetic species capable of overwhelming everything that came before.

If this is so, maybe the next question is whether there’s anything else waiting to be good-vs-evil-ified, what form that will take, and what will happen afterwards.

Powerless Placebos

The most important study on the placebo effect is Hróbjartsson and Gøtzsche's Is The Placebo Powerless?, updated three years later by a systematic review and seven years later with a Cochrane review. All three looked at studies comparing a real drug, a placebo drug, and no drug (by the third, over 200 such studies) – and, in general, found little benefit of the placebo drug over no drug at all. There were some possible minor placebo effects in a few isolated conditions – mostly pain – but overall H&G concluded that the placebo effect was clinically insignificant. Despite a few half-hearted tries, no one has been able to produce much evidence they're wrong. This is kind of surprising, since everyone has been obsessing over placebos and saying they're super-important for the past fifty years.

What happened? Probably placebo effects rode on the coattails of a more important issue, *regression to the mean*. That is, most sick people get better eventually. This is true both for diseases like colds that naturally go away, and for diseases like depression that come in episodes which remit for a few months or years until the next relapse. People go to the doctor during times of extreme crisis, when they're most sick. So no matter what happens, most of them will probably get better pretty quickly.

In the very old days, nobody thought of this, so all their experiments were hopelessly confounded. Then people started adding placebo groups, this successfully controlled for not just placebo effect but regression to the mean, and so people noticed their studies were much better. They called the whole thing "placebo effect" when in fact there was no way to tell without further study how much was real placebo effect and how much was just regression to the mean. If we believe H&G, it’s pretty much all just regression to the mean, and placebo was a big red herring.

The rare exceptions are pain and a few other minor conditions. From H&G #3:

> We found an effect on pain, SMD -0.28 (95% CI -0.36 to -0.19)); nausea, SMD -0.25 (-0.46 to -0.04)), asthma (-0.35 (-0.70 to -0.01)), and phobia (SMD -0.63 (95% CI -1.17 to -0.08)). The effect on pain was very variable, also among trials with low risk of bias. Four similarly-designed acupuncture trials conducted by an overlapping group of authors reported large effects (SMD -0.68 (-0.85 to -0.50)) whereas three other pain trials reported low or no effect (SMD -0.13 (-0.28 to 0.03)). The pooled effect on nausea was small, but consistent. The effects on phobia and asthma were very uncertain due to high risk of bias.

So the acupuncture trials seem to do pretty well. This probably isn’t because acupuncture works – some experiments have found sham acupuncture works equally well. It could be because acupuncture researchers have flexible research ethics. But Kamper & Williams speculate that acupuncture does well because it’s an optimized placebo. Normal placebos are just some boring little pill that researchers give because it’s the same shape as whatever they really want to give. Acupuncture – assuming that it doesn’t work – has been tailored over thousands of years to be as effective a pain-relieving placebo as possible. Maybe there’s some deep psychological reason why having needles in your skin intuitively feels like the sort of thing that should alleviate pain.

I want to add my own experience here, which is that occasionally I see extraordinary and obvious cases of the placebo effect. I once had a patient who was shaking from head to toe with anxiety tell me she felt completely better the moment she swallowed a pill, before there was any chance she could have absorbed the minutest fraction of it. You're going to tell me "Oh, sure, but anxiety's just in your head anyway" – but anxiety was one of the medical conditions that H&G included in their analysis. Plausibly they studied chronic anxiety, and pills are less good chronically than they are at aborting a specific anxiety attack the first time you take them. Or maybe her anxiety was somehow related to a phobia, one of the conditions H&G find some evidence in support of a placebo for. (Really? Phobia but not anxiety? Whatever.)

Surfing Uncertainty had the the best explanation of the placebo effect I've seen. Perceiving the world directly at every moment is too computationally intensive, so instead the brain guesses what the the world is like and uses perception to check and correct its guesses. In a high-bandwidth system like vision, guesses are corrected very quickly and you end up very accurate (except for weird things like ignoring when the word "the" is twice in a row, like it's been several times in this paragraph already without you noticing). In a low-bandwidth system like pain perception, the original guess plays a pretty big role, with real perception only modulating it to a limited degree (consider phantom limb pain, where the brain guesses that an arm that isn't there hurts, and nothing can convince it otherwise). Well, if you just saw a truck run over your foot, you have a pretty strong guess that you're having foot pain. And if you just got a bunch of morphine, you have a pretty strong guess that your pain is better. The real sense-data can modulate it in a Bayesian way, but the sense-data is so noisy that it won't be weighted highly enough to replace the guess completely.

If this is true, placebo should be strongest in subjective perceptions of conditions sent to the brain through low-bandwidth relays. That covers H&G’s pain and nausea. It doesn’t cover asthma and phobias quite as well, though I wonder if “asthma” is measured as subjective sensation of breathing difficulty.

What about depression? My gut would have told me depressed people respond very well to the placebo effect, but H&G say no.

I think that depressed mood may respond well to the placebo effect on a temporary basis – after all, mood seems noisy and low-bandwidth and hard to be sure of in the same way pain and nausea are. But most studies of depression use tests like the HAM-D, which measure the clinical syndrome of depression – things like sleep disturbance, appetite disturbance, and motor disturbance. These seem a lot less susceptible to subjective changes in the way the brain perceives things, so probably HAM-D based studies will show less placebo effect than just asking patients to subjectively assess their mood.

Predictions For 2018

At the beginning of every year, I make predictions. At the end of every year, I score them. So here are a hundred more for 2018.

Some changes this year: I've eliminated a bunch of predictions about things that are very unlikely where I just plug in the same number each year, like "99% chance of no coup in the US". I've tried to have almost everything this year be new and genuinely uncertain. I've also included some very personal predictions about friends and gossip that I'm keeping secret for now – I have them written down somewhere else and they're for my own interest only.

My rule is that I have to make all of these without checking existing prediction markets – otherwise I wouldn't be learning anything about my own abilities. I'm also not doing any research beyond what I already know, because otherwise this will take forever. I bet some of these are terribly misinformed, but that's part of what I'm including in my calibration estimate. These were written a few days ago; a few already seem obsolete.

I'm keeping 50% predictions even though everyone keeps telling me they don't matter. My only excuse is that I write everything down first and then decide what I think the likelihood is, and sometimes my best guess really is 50%.

US:
1. Donald Trump remains president at end of year: 95%
2. Democrats take control of the House in midterms: 80%
3. Democrats take control of the Senate in midterms: 50%
4. Mueller's investigation gets cancelled (eg Trump fires him): 50%

5. Mueller does not indict Trump: 70%
6. PredictIt shows Bernie Sanders having highest chance to be Dem nominee at end of year: 60%
7. PredictIt shows Donald Trump having highest chance to be GOP nominee at end of year: 95%
9. Some sort of major immigration reform legislation gets passed: 70%
10. No major health-care reform legislation gets passed: 95%
11. No large-scale deportation of Dreamers: 90%
12. US government shuts down again sometime in 2018: 50%
13. Trump’s approval rating lower than 50% at end of year: 90%
14. …lower than 40%: 50%
15. GLAAD poll suggesting that LGBQ acceptance is down will mostly not be borne out by further research: 80%

ECONOMICS AND TECHNOLOGY:
16. Dow does not fall more than 10% from max at any point in 2018: 50%
17. Bitcoin is higher than $5,000 at end of year: 95%
18. Bitcoin is higher than $10,000 at end of year: 80%
19. Bitcoin is lower than $20,000 at end of year: 70%
20. Ethereum is lower than Bitcoin at end of year: 95%
21. Luna has a functioning product by end of year: 90%
22. Falcon Heavy first launch not successful: 70%
23. Falcon Heavy eventually launched successfully in 2018: 80%
24. SpaceX does not attempt its lunar tourism mission by end of year: 95%
25. Sci-Hub is still relatively easily accessible from within US at end of year (even typing in IP directly is *relatively* easy): 95%
26. Nothing particularly bad (beyond the level of an funny/weird news story) happens because of ability to edit

videos this year: 90%
27. A member of the general public can ride-share a self-driving car without a human backup driver in at least one US city by the end of the year: 80%

CULTURE WARS:
28. Reddit does not ban r/the_donald by the end of the year: 90%
29. None of his enemies manage to find a good way to shut up/discredit Jordan Peterson: 70%

COMMUNITIES:
30. SSC gets more hits in 2018 than in 2017: 80%
31. SSC gets mentioned in the New York Times (by someone other than Ross Douthat): 60%
32. At least one post this year gets at least 100,000 hits: 70%
33. A 2019 SSC Survey gets posted by the end of the year: 90%
34. No co-bloggers make 3 or more SSC posts this year: 80%
35. Patreon income less than double current amount at end of year: 90%
36. A scientific paper based on an SSC post is accepted for publication in real journal by end of year: 60%
37. I do an adversarial collaboration with somebody interesting by the end of the year: 50%
38. I successfully do some general project to encourage and post more adversarial collaborations by other people: 70%
39. New SSC meetups system/database thing gets launched successfully: 60%
40. LesserWrong remains active and successful (average at least one halfway-decent post per day) at the end of the year: 50%
41. LesserWrong is declared official and merged with LessWrong.com: 80%

42. I make fewer than five posts on LessWrong (posts copied over from SSC don't count): 70%
43. CFAR buys a venue this year: 50%
44. AI Impacts has at least three employees working half-time or more sometime this year: 50%
45. Rationalists get at least one more group house on Ward Street: 50%
46. No improvement in the status of reciprocity.io (either transfer to a new team or at least one new feature added): 70%

PERSONAL:
47. I fail at my New Years' resolution to waste less time on the Internet throughout most of 2018: 80%
48. I fail at my other New Years' resolution to try one biohacking project per month throughout 2018: 80%
49. I don't attend the APA National Meeting: 80%
50. I don't attend the New York Solstice: 80%
51. I travel outside the US in 2018: 90%
52. I get some sort of financial planning sorted out by end of year: 95%
53. I get at least one article published on a major site like Vox or New Statesman or something: 50%
54. I get a tax refund: 50%
55. I weigh more than 195 lb at year end: 60%
56. I complete the currently visible Duolingo course in Spanish: 90%
57. I don't get around to editing Unsong (complete at least half the editing by my own estimate) this year: 95%
58. No new housemate for at least one month this year: 90%
59. I won't [meditate at least one-third of days this year]: 90%
60. I won't [do my exercise routine at least one third of days this year]: 80%
61. I still live in the same house at the end of 2018: 60%

62. I will not have bought a house by the end of 2018: 90%
63. Katja’s paper gets published: 90%
64. Some other paper of Katja’s gets published: 50%

SECRET: (mostly speculating on the personal lives of friends who read this blog; I don’t necessarily want them to know how successful I expect their financial and romantic endeavors to be)
65. [Secret prediction]: 80%
66. [Secret prediction]: 70%
67. [Secret prediction]: 70%
68. [Secret prediction]: 60%
69. [Secret prediction]: 70%
70. [Secret prediction]: 60%
71. [Secret prediction]: 50%
72. [Secret prediction]: 50%
73. [Secret prediction]: 50%
74. [Secret prediction]: 90%
75. [Secret prediction]: 90%
76. [Secret prediction]: 60%
77. [Secret prediction]: 70%
78. [Secret prediction]: 60%
79. [Secret prediction]: 50%
80. [Secret prediction]: 60%
81. [Secret prediction]: 80%
82. [Secret prediction]: 70%
83. [Secret prediction]: 50%
84. [Secret prediction]: 70%
85. [Secret prediction]: 70%
86. [Secret prediction]: 70%
87. [Secret prediction]: 60%
88. [Secret prediction]: 50%
89. [Secret prediction]: 50%
90. [Secret prediction]: 70%
91. [Secret prediction]: 90%
92. [Secret prediction]: 50%

93. [Secret prediction]: 90%
94. [Secret prediction]: 50%
95. [Secret prediction]: 60%
96. [Secret prediction]: 60%
97. [Secret prediction]: 60%
98. [Secret prediction]: 95%
99. [Secret prediction]: 70%
100. [Secret prediction]: 70%

Other properly formatted predictions for this year:
– Socratic Form Microscopy (2017 results, 2018 predictions)
– Anatoly Karlin (2017 results, 2018 predictions)
– Various people from the subreddit
– Very many people on Metaculus

[EDIT: List of predictions I've already been convinced are miscalibrated as of 3 AM 2/6/18:
– Bitcoin prices are already too high (they were higher when I wrote these predictions a few days ago).
– Stock market is more likely to have large fall (it was higher when I wrote these predictions a few days ago)
– Chance of Trump's approval not breaking 50% probably closer to 95% than 90%.]

Guyenet On Motivation

Rereading *The Hungry Brain*, I notice my review missed one of my favorite parts: the description of the motivational system. It starts with studies of lampreys, horrible little primitive parasitic fish:

> How does the lamprey decide what to do? Within the lamprey basal ganglia lies a key structure called the striatum, which is the portion of the basal ganglia that receives most of the incoming signals from other parts of the brain. The striatum receives "bids" from other brain regions, each of which represents a specific action. A little piece of the lamprey's brain is whispering "mate" to the striatum, while another piece is shouting "flee the predator" and so on. It would be a very bad idea for these movements to occur simultaneously – because a lamprey can't do all of them at the same time – so to prevent simultaneous activation of many different movements, all these regions are held in check by powerful inhibitory connections from the basal ganglia. This means that the basal ganglia keep all behaviors in "off" mode by default. Only once a specific action's bid has been selected do the basal ganglia turn off this inhibitory control, allowing the behavior to occur. You can think of the basal ganglia as a bouncer that chooses which behavior gets access to the muscles and turns away the rest. This fulfills the first key property of a selector: it must be able to pick one option and allow it access to the muscles.
>
> Many of these action bids originate from a region of the lamprey brain called the pallium…

Spoiler: the pallium is the region that evolved into the cerebral cortex in higher animals.

> Each little region of the pallium is responsible for a particular behavior, such as tracking prey, suctioning onto a rock, or fleeing predators. These regions are thought to have two basic functions. The first is to execute the behavior in which it specializes, once it has received permission from the basal ganglia. For example, the “track prey” region activates downstream pathways that contract the lamprey’s muscles in a pattern that causes the animal to track its prey. The second basic function of these regions is to collect relevant information about the lamprey’s surroundings and internal state, which determines how strong a bid it will put in to the striatum. For example, if there’s a predator nearby, the “flee predator” region will put in a very strong bid to the striatum, while the “build a nest” bid will be weak…
>
> Each little region of the pallium is attempting to execute its specific behavior and competing against all other regions that are incompatible with it. The strength of each bid represents how valuable that specific behavior appears to the organism at that particular moment, and the striatum’s job is simple: select the strongest bid. This fulfills the second key property of a selector – that it must be able to choose the best option for a given situation…
>
> With all this in mind, it’s helpful to think of each individual region of the lamprey pallium as an option generator that’s responsible for a specific behavior. Each option generator is constantly competing with all other incompatible option generators for access to the

> muscles, and the option generator with the strongest bid at any particular moment wins the competition.

The next subsection, which I'm skipping, quotes some scientists saying that the human motivation system works similarly to the lamprey motivation system, except that the human cerebrum has many more (and much more flexible/learnable) options than the lamprey pallium. Humans have to “make up our minds about things a lamprey cannot fathom, like what to cook for dinner, how to pay off the mortgage, and whether or not to believe in God”. It starts getting interesting again when it talks about basal ganglia-related disorders:

> To illustrate the crucial importance of the basal ganglia in decision-making processes, let's consider what happens when they don't work.
>
> As it turns out, several disorders affect the basal ganglia. The most common is Parkinson's disease, which results from the progressive loss of cells in a part of the basal ganglia called the substantia nigra. These cells send connections to the dorsal striatum, where they produce dopamine, a chemical messenger that plays a very important role in the function of the striatum. Dopamine is a fascinating and widely misunderstood molecule that we'll discuss further in the next chapter, but for now, its most relevant function is to increase the likelihood of engaging in any behavior.
>
> When dopamine levels in the striatum are increased – for example, by cocaine or amphetamine – mice (and humans) tend to move around a lot. High levels of dopamine essentially make the basal ganglia more sensitive to incoming bids, lowering the threshold for activating movements…Conversely, when dopamine

> levels are low, the basal ganglia become less sensitive to incoming bids and the threshold for activating movements is high. In this scenario, animals tend to stay put. The most extreme example of this is the dopamine-deficient mice created by Richard Palmer, a neuroscience researcher at the University of Washington. These animals sit in their cages nearly motionless all day due to a complete absence of dopamine. “If you set a dopamine deficient mouse on a table,” explains Palmiter, “it will just sit there and look at you. It’s totally apathetic.” When Palmiter’s team chemically replaces the mice’s dopamine, they eat, drink, and run around like mad until the dopamine is gone.

The same can happen to humans with basal ganglia injuries:

> Consider Jim, a former miner who was admitted to a psychiatric hospital at the age of fifty-seven with a cluster of unusual symptoms. As recorded in his case report, “during the preceding three years he had become increasingly withdrawn and unspontaneous. In the month before admission he had deteriorated to the point where he was doubly incontinent, answered only yes or no questions, and would sit or stand unmoving if not prompted. He only ate with prompting, and would sometimes continue putting spoon to mouth, sometimes for as long as two minutes after his plate was empty. Similarly, he would flush the toilet repeatedly until asked to stop.”
>
> Jim was suffering from a rare disorder called abulia, which is Greek for “an absence of will”. Patients who suffer from abulia can respond to questions and perform specific tasks if prompted, but they have difficulty spontaneously initiating motivations, emotions, and

> thoughts. A severely abulic patient seated in a bare room by himself will remain immobile until someone enters the room. If asked what he was thinking or feeling, he'll reply, "Nothing"…
>
> Abulia is typically associated with damage to the basal ganglia and related circuits, and it often responds well to drugs that increase dopamine signaling. One of these is bromocriptine, the drug used to treat Jim… Researchers believe that the brain damage associated with abulia causes the basal ganglia to become insensitive to incoming bids, such that even the most appropriate feelings, thoughts, and motivations aren't able to be expressed (or even to enter consciousness). Drugs that increase dopamine signaling make the striatum more sensitive to bids, allowing some abulic patients to recover the ability to feel, think, and move spontaneously.

All of this is standard neuroscience, but presented much better than the standard neuroscience books present it, so much so that it brings some important questions into sharper relief. Like: what does this have to do with willpower?

Guyenet describes high dopamine levels in the striatum as "increasing the likelihood of engaging in any behavior". But that's not really fair – outside a hospital, almost nobody just sits motionless in the middle of a room and does *no* behaviors. The relevant distinction isn't between engaging in behavior vs. not doing so. It's between low-effort behaviors like watching TV, and high-effort behaviors like writing a term paper. We know that this has to be related to the same dopamine system Guyenet's talking about, because Adderall (which increases dopamine in the relevant areas) makes it much easier to do the high-effort behaviors.

So a better description might be "high dopamine levels in the striatum increase the likelihood of engaging in high-willpower-requirement behaviors".

But what *is* high willpower requirements? I'm always tempted to answer this with some sort of appeal to basic calorie expenditure, but taking a walk requires less willpower than writing a term paper even though the walk probably burns way more calories. My "watch TV" option generator, my "take a walk" option generator, and my "write a term paper" option generator are all putting in bids to my striatum – and for some reason, high dopamine levels privilege the "write a term paper" option and low dopamine levels privilege the others. Why?

I don't know, and I think it's the most interesting next question in the study of these kinds of systems.

But here's a crazy idea (read: the first thing I thought of after thirty seconds). In the predictive processing model, dopamine represents confidence levels. Suppose there's a high prior on taking a walk being a reasonable plan. Maybe this is for evo psych reasons (there was lots of walking in the ancestral environment), or for reinforcement related reasons (you enjoy walking, and your brain has learned to predict it will make you happy). And there's a low prior on writing a term paper being a reasonable plan. Again, it's not the sort of thing that happened much in the ancestral environment, and plausibly every previous time you've done it, you've hated it.

In this case, confidence in your new evidence (as opposed to your priors) is a pretty important variable. If your cortex makes its claims with high confidence (ie in a high-dopaminergic state), then its claim that it's a good idea to write a term paper now may be so convincing that it's able to overcome the high prior against this being true. If your cortex makes claims with low confidence, then it will tentatively suggest that maybe we should write a term paper now – but the striatum will remain unconvinced due to the inherent implausibility of the idea.

In this case, sitting in a dark room doing nothing is just an action plan with a very high prior; you need at least a tiny bit of confidence in your planning ability to shift to anything else.

I mentioned in Toward A Predictive Theory Of Depression that I didn't understand the motivational system well enough to be able to explain why systematic underconfidence in neural predictions would make people less motivated. I think the idea of evolutionarily-primitive and heavily-reinforced actions as a prior – which logical judgments from the cortex have to "override" in order to produce more willpower-intensive actions – fills in this gap and provides another line of evidence for the theory.

We’ve Got Five Years, What A Surprise

Today is the fifth anniversary of Slate Star Codex. Overall I’m very happy with how this project is going so far, and I want to take this opportunity to thank everyone who’s made things work behind the scenes.

Trike Apps generously volunteered to host me free of charge. I give them the highest praise it is possible to give a hosting company – namely, that I completely forgot about their existence until right now because I’ve never had to worry about anything. Special thanks to Matt Fallshaw and Cat Truscott for their kindness and patience.

Bakkot has done various things behind the scenes to make the blog more useable – fixing WordPress bugs, helping with moderation tasks, and adding cool new features like the green highlights around new comments. A big part of the success of the comments section is thanks to his innovations; the remaining horribleness is mostly my fault. Rory O and Alice M have also helped with this.

Michael K and Mason H have done other behind-the-scenes work, especially improving the design. Remember when the front page looked like this? No? Thank Michael and Mason.

The subreddit moderation team – led by Bakkot again, but also _Vulture_, coderman9, PM_ME_YOUR_OBSIDIAN, tailcalled, werttrew, utilsucks, and cjet79 – keeps the subreddit in line, and deserves special gratitude for wading through all of your horrible offensive opinions in the process of redirecting them into the Culture Wars thread.

Deluks917 started the Discord, which is thriving. Thanks to him and the rest of the Discord moderation team – Celestia, notaraptor, and vivafringe.

Jeremiah started the podcast, which continues to update *very* regularly.

Mingyuan runs the general repository of meetups. Thank her if you've been able to find an SSC meetup in your area – and if you haven't, give her a little while, since she's working on a better and higher-tech version. Thanks also to everyone who organizes meetups. I'm most familiar with David Friedman's very large and regular South Bay meetups, but I understand there are decent-sized groups all around the world and I'm grateful for everyone who puts work into it.

Cody, Oliver, and other people organized the *Unsong* wrap celebration so efficiently that I'm still not sure exactly what happened or who did what. I just showed up and everything seems to have worked. Apologies if I am forgetting people or exactly what they did.

Katja has organized the Open Thread system so that it posts on time and with the right decimals. She also gives general moral support and puts up with me. She also arranged the *Unsong* wrap after-party and has given me lots of interesting things to think about.

Ozy and Elizabeth have guest-blogged here very briefly. Thanks to them – and congratulations to Ozy on the recent birth of their first child.

Roland helped transform my post on antidepressants into a scientific paper that got accepted by a pharmacology journal. Ada Palmer's thoughts on finally publishing a novel are pretty much how I feel about finally having published a journal article. Preliminary thanks also to everyone currently working with me on similar projects for other posts.

Scott Aaronson, Leah Libresco, and other more established bloggers – plus some Real Journalists like Conor Friedersdorf, Tom Chivers and Ross Douthat – linked to me and encouraged me when I was relatively new to this, and helped convince me that this “blogging” thing might work out.

Thanks to all our advertisers, but especially to Beeminder and MealSquares, who have stuck with me since the beginning and put up with everything from me never remembering to respond to their emails to me gratutitiously and unfairly insulting their products. I actually think they’re both great companies and totally recommend that you Beemind the number of MealSquares you eat, or something.

Thanks to everyone who supports me on Patreon. Your money pays for things like the Mechanical Turk version of the SSC survey, my books and subscriptions, and me having more time to work on the blog.

Thanks to everyone I’ve engaged with. Again and again I’ve had the experience of reading something, criticizing it (sometimes savagely), and having the author be incredibly nice to me, walk me through where we disagree, and then continue being friendly and supportive afterwards. Some people in this category include Ezra Klein, Adam Grant, Nathan Robinson, Curtis Yarvin, Bryan Caplan, David Friedman (again), Dylan Matthews, Brendan Nyhan, Nick Land, Julia Rohrer, and Freddie de Boer. I continue to disagree with them strongly on a lot of things but can’t find even the tiniest bit of fault with their decency on a personal level.

Thanks to all the people who seem to genuinely dislike me and wish me ill, but who have been decent enough not to let it get to the point of doxxing me or threatening my personal safety, my career, or my relationship with my patients.

Thanks to everyone who comments and contributes to discussions. Some people who I’ve heard praised again and again – John Schilling on international affairs, David Friedman on economic issues, Deiseach on religious and cultural issues, and of course Bean on battleships. But everybody is appreciated. Godwin’s Law says that if you want a question answered online, you shouldn’t ask – you should post something wrong and let people correct you. But I have had good luck groping towards the best answer I can, then letting a bunch of experts show up and fill in the details.

Thanks to everyone who takes the (long, often poorly worded) surveys. Without you, interesting findings like this would not be possible – and trust me, there’s more where that came from once I have time to write it up. One of my goals is to find more ways to use this blog’s readership to advance psychological research, and your surprising willingness to waste time on any crazy questions I throw at you has been delightful.

Thanks to everyone who sent me emails, requests, and questions for not being a jerk when I didn’t answer for several weeks or, in many cases, ever. Without your tolerance for my rudeness, I would have much less time to write.

I am probably forgetting all sorts of people – if so, no offense meant. You are all great.

More Testimonials For SSC

[Content note: various slurs and insults]

Last post I thanked some of the people who have contributed to this blog. But once again, it’s time to honor some of the most important contributors: the many people who give valuable feedback on everything I write. Here’s a short sample of some of…most interesting. I’m avoiding names and links to avoid pile-ons. Some slightly edited for readability.

“A cowardly autistic cuckolded deviant Jew who uses his IQ to rationalize away wisdom”

“He’s part of the self-declared ‘Rationalish Community’. Imagine the ridiculous level of self-regard implied by that. Picture cb2 with a graduate degree. Scott Alexander, if brevity is the soul of wit, you’re a witless soulsucking fuck.”

“Dude sounds like a crackpot. Blaming Republicans for everything and hailing liberals for… well, that part isn’t clear exactly. Thanks for helping me find something to add to my “never read” collection.”

“The men tweeting about how bad the women’s march is are also the guys who didn’t get invited to parties + blamed it on feminism…I know a few men who make this seem like actual fact. Like that guy from Slate Star Codex.”

“I don’t know what I was expecting from a jew quack but I suppose reasonable fits the description.”

“Ross Douthat somehow manages to recommend a person with a theology less plausible than Catholicism”

“I refuse to read Slate Star Codex anymore. It has become the epitome of IYI (Intellectual Yet Idiot) “pragmatic”(i.e. spineless) centrism.”

“He wants his readers to adopt a strategy of misogynist sabotage.”

“Slate Star Codex is THE definition of ‘autism spiral into infertility and death'”

“Scott Alexander is a LessWrong cultist. He has ALWAYS been a tremendous asshole who thinks he’s Mister Fucking Spock.”

“Slate Star Codex is to cognitive dissonance what Goddard was to rocketry.”

“I used to really enjoy Slate Star Codex before joining the dark side, now I find the blog almost insufferably autistic.”

“Laughing my ass off as Slate Star Codex’s “The Anti-Reactionary FAQ” figuratively burns in the fires of Berkeley.”

“I’ve started to be bothered by clothes tags a little bit lately. I blame SSC for putting this idea in my head”

“Slate Star Codex was always a shill, but this was craven even for him”

“He *literally* thinks that humans are horses”

“This is entirely reductive statement from me but I think that in an important sense SSC is just the Scott Alexander ego inflation program. Some of the best blogs can function this way. However a reasonable about the disingenuous use of his explicitly stated preferences for objectivity and the unstated outcomes of his blog can be made. Is Scott a scientist promoting a radical new social program or perhaps he is interested in the cult of personality and trapping of psychiatric performance?"

"Sexist asses: It's not us, females just genetically hate liberty."

"You asexual twerp"

"Fucking tech-libertarian cockroaches everywhere preaching total derp. Deeply disappointing."

"I see you reduced to making excuses for a career criminal [Hillary Clinton] because you're afraid of change. I expected more from you, Scott. I expected you to remember hope. I never expected the Dark would take you. Enjoy this. The thousands of comments, the last remarks from departing readers. It will never be the same."

"Honestly, every time I read Scott, I am super conflicted. I have never found a writer whom I agree with so consistently while finding their personality, as expressed through their writing, so intolerable. I always feel like I want to shout, "You're exactly correct! Now the shut the fuck up!" and pop him one in the mouth."

"Go figure, Slate Star Codex blog readers are politically *and* literally, a bunch of phags. Look @ weightlifting data [on the survey]; found the problem."

"Scott isn't really dogmatic about anything besides niceness, honesty, puns, and growth mindset."

"Scott Alexander over at the popular blog Slate Star Codex is an interesting case study in classical liberalism; nowhere else will you find someone who better exemplifies the phenomenon of skirting within microns of the event horizon of Getting It before screaming ‘Nooooo’ and zooming off in some other direction; nor will you find many who choose a crazier direction in which to flee.”

“Basically imagine a guy drinking Soylent and having a flamewar about how in the future they will too be able to unfreeze his head and you’ve got a basic idea of the ideology at play here.”

“So we come to answering the question I asked at the beginning: What is it that allows men like Scott Alexander, men of some intelligence and sensitivity, to get so close to understanding, and fail so miserably, over and over? We can find the answer here at the end of his piece: we see that he stumbled, baffled, like a giraffe with a head injury loose in Manhattan, through the entire book, then through an entire long review, without comprehending its basic point.”

“Slate Star Codex: if you’re a man who is involved in tech and not interested in any legitimate philosophical or sociological inquiry, we’ve got you covered”

“He is riddled with all kinds of spooks and leftist ideology and it shows in his commentariat. This also doesn’t bode well for psychiatry if someone as emotionally weak as Alexander is allowed to become a psychiatrist.”

“The Slate Star Codex guy is a living fable against the idea that u can solve problems with pure tedious reason instead of ever reading a book”

“it’s basically one of the hubs for autistic people really into Bayesianism, so like half the posters could either transition or become Nazis. or both idk”

“Hey man I took your survey it made me feel all weird and insecure about my gender identity thanks a lot!”

“Anyway, I don’t mean to pick on Alexander, whose heart is in the right place, but he is a walking, talking, male prophylactic. If I absolutely did not want any grandchildren (say, high risk of insanity in the bloodline) I’d have Alexander teach my sons the birds & bees. He is a weirdo autistic who has no understanding of normal women based on the few writings on sexual politics of his I’ve read, which are filled with the usual libertard lonely-boy pablum about the awfulness of “slut shaming” in our society, and how if we could just get rid of that and any sort of gender roles and treat everyone the same, we’d be living in a flippin’ sexual Nirvana where our genitals would be as happily interoperable as any pair of USB ports. (Alexander, IIRC, is in a relationship a technically female but maybe not womyn-aligned webcam star with whom he may or may not actually be bumping uglies)”

“To be fair to Alexander, the million leaked credit cards #’s from ASHLEY MADISON from men who really think there any normal women out there trolling for one-off sex on the Internet shows the cluelessness out there is pretty broad.”

“Since people are sharing around a Slate Star Codex article let’s have a reminder that he’s a moral cripple”

“$500 Reward. Seeking the testimony of victims of Scott Alexander, human rights abuser. I am also willing to pay for the stories of the victims of any other ‘prominent’ ‘internet famous’ psychiatrists/human rights abusers.”

“I hate to go ad hom, but i can’t think of anyone who would benefit more from TRT and getting laid on the reg.”

“After reading Scott’s article to a friend of mine, he decided to get “Border Reaver” tattooed on his neck”

“So basically he’s an athiest jewish kikeiatrist named Schlomo Schlomovich who mingles with the social elite while still being afraid of antifa? I couldn’t have strung together that many ridiculous stereotypes at once if I tried honestly. This is fucking hilarious.”

“Discovering Scott Aaronson is way into Slate Star Codex is like finding spiders in my favorite flavor of ice cream. Slate Star Codex is ‘Well, actually…’ personified, with a dusting of evil. But mostly it bugs me that it passes for good writing.”

“id like to fight the guy who runs slate star codex, he’s a smarmy faggot”

“Why do people I otherwise like keep insisting to me that slate star codex is good”

“[Slate Star Codex] split off from Less Wrong because even massive faggots sometimes have standards. His clique don’t exactly get along with Yudkowsky and will point out that he’s basically running a cult. Nonetheless, Scott’s still a huge fag and sucks Eliezer’s dick when it comes to rationalism and his fucking gay “Sequences”, which he and his commenters will tell you to read as if it’s the fucking gospel.”

“Slate Star Codex, an extremely verbose blog that I have complicated feelings about.”

“YouTube Skeptics, Slate Star Codex rationalists, Stefan Molyneux and Ayn Rand all ruined “rationality” and ‘logic’ for me. Must be a horseshoe theory conspiracy of sorts.”

“The disturbing thing is that they’re all aware of the criticisms people level at them for their autism, but no matter how many times they’re inundated by people telling them they’re being inhuman spergs, they’re just like

‘Hmm…am I out of touch? No…it’s the normal people who are wrong.'”

“*making racist laser gun noises* computer, engage Near Mode and navigate me to slate star codex please”

“‘Bigoted shits’ is basically the Slate Star Codex demographic”

“[Scott] wants the SJWs to take over. He wants you to dawdle around appealing to ‘reason’ until the Commies have indoctrinated enough of the youth to allow PC Culture to permeate all things.”

“It’s cool to watch the slate star codex guy inch closer and closer to actually knowing something while his comment sections get stupider and stupider”

“Is it just me, or is the guy who runs slate star codex kind of a wanker?”

“But this is… just incredible. I read this SSC article last night, and my jaw dropped. What was I missing? How could Scott Alexander be so fucking stupid? I spent all day with a slow burning anger in my belly. This pure nonsense, from the “Red Tribe vs. Blue Tribe” guy, in the same week as McConnell holding millions of children hostage so he doesn’t accidentally upset the avowed racists over in the House, not to mention Stormy Daniels, McCabe’s loyalty test, Trump trying to fire Mueller, and all the other usual shit that I already forgot all coming to light? And you choose now—January 24, 2018 and not November 9, 2016—to equate George Soros and the Koch brothers not once but twice in an overlong Tumblr note that amounts to saying, “huh I just realized maybe I’m missing something but I still think all politically active people are retarded”? Have you read the news once in the last year, or do you just get summaries from the same place as your political theories—random fucking commenters on your blog? Or was I mistaken this whole time thinking that Scott both lived in America and wanted the world to get better not worse? Because this post would make way more sense if his political climate was actually recess on the fucking playground of a quarantined elementary school for experimental Nazi test tube babies in a bubble on the dark side of the fucking moon.”

“nobody has ever read a slate star codex article to the end”

Even More Search Terms That Led People To This Blog

[Previously in series: Search Terms That Have Led People To This Blog and More Search Terms That Have Led People To This Blog. Content warning: profanity, rape, and other unfiltered access to the consciousness of the Internet]

Sometimes I look at what search terms lead people to SSC. Sometimes it's the things you would think – "slate star codex", "rationality", the names of medications I've written about.

Other times it's a little weirder:

---

why is my sister so pretty
I mentioned this query last post, probably based on this article, and the onslaught hasn't stopped.

my sis is so pretty
Sometimes I can pretend it's just people happy for their family members' good qualities…

my sister is really pretty
…many people, very very happy…

my sister is so sexy what to do
…and other times, not so much.

sweet sister so pretty
You can stop any time now.

how we attract our sister for sex
NO, REALLY, YOU CAN STOP ANY TIME NOW.

sister aroused by my touch
ANNNNNY TIME NOW.

www.my sister pic sex.com
…frick.

the fate of a cruel snake re arrange ssc answer
Other times people really seem to think I have an article on something but I have no idea what they mean.

what is the hormone responsible for soliloquising?
Other times I just have no idea.

hivemind ape and young girl army experiment
Garrett Jones, what exactly are you doing over at GMU?

Glasgow coma scale
I got a lot of these from people looking for this.

glassco coma scale
Some of the spellings were very creative

glass go coma scale
glascov coma scale
glasco comma scale
glawcow coma scale
glasscoma scale
glascoma score
glassma coma scale
glascow comma scale
glosvow coma scale
glass cow coma scale
glass comma scale
…

slate satr codex
slate star xodex
slate static codex
slate star cosex
slate star cpdex

str slate codex
sstar slate codex
astark star codex
slate state kodaks
…

delay cool condom codex
???

slate star codex of hate
?!?

i canot talret anything
You're probably looking for the Slate Star Codex Of Hate

rapist linked to prostate cancer

which star sign is most likely to be a rapist

scott alexander the gay guy's biography
Versus…

scott alexander is no gay
Versus…

fuck scott alexander endless

criticize the statement "you can see atoms"
It's really dumb

fnord soros fnord
…did someone just try to see the fnords by *typing it into Google*? That's *great*.

considered armed and dangerous for cow pox

i am polynesian and so is my husband both with brown eyes however our son has blue eyes how

if hundreds of americans die tragically today, it’ll be on account of sarcastic cocksucker citizens

unreasonable autism cures
It makes me so happy when our medical system is set up to satisfy someone’s needs this perfectly

whale…. medical cartel hoax . borax
I don’t know what conspiracy this person thinks connects all these things, but they should probably put some fnords in there if they want results.

victory lotto forecast for tomorrow tuesday 11-07-17
Nice try.

using secret pyramid to hit 3 digit lottery

was rene descartes racist

what is going on here? how can just the opinion of 1068 people always determine the opinion of the entire population, no matter how large it is? is statistics broken? is 1068 just a magical number?
It’s because of [fnord] the medical cartel, whales, and borax [/fnord].

there are many things i want to say but do not know how to say. hope you will understand

I am [person’s real name removed] iwant to join illumenatic member want can i do or who can help me in order direct me
Find a member of the medical cartel and say the code word “whale”. They’ll do the rest.

i want to join illuminati brotherhood church in uganda, south africa and kenya post comments on

usa blogger
Find a whale and say the code word “medical cartel”. What happens next is up to you.

is fentanyl being used for population reduction by the illuminati
No, that’s what the borax is for

jews hate alcoholics anonymous

can i get a s sample ogre biscuit factory dimension

what decisions might the police make base on crime
One would hope all of them.

why is aa mostly bullshit
Looks like we’ve got a Jew here.

list of human experiences
1. Birth
2. Eating
3. Sleeping
4. Being attracted to your sister
5. Misspelling “Glasgow Coma Scale”
6. Using secret pyramid to hit 3 digit lottery
[…]

how can we deal with cactus person?

now if we can prove the electoral college was seeded with purely partisan voters for trump, and illegal, then we can have a new vote
[164/777] …and as these documents clearly prove, Putin sent the files to Trump through the medical cartel, hidden in envelopes marked “WHALE BORAX”.

islamists deport murderous racists
Well, that was a wild garden path of a four-word sentence

bigotry xxx yup to video
…

creators of remote neural monitoring are gay designing an ass weapon

massage with bombastic words of wishing good lucky to all people doing matriculant

how to start a zombie story
ZOMBIE. STORIES. ARE. SUPER. DONE.

100 statements about albion's seed: four british folkways in america that almost killed my hamster

give directly illuminati

how to summon abraham lincoln

you are my pleasant gustatory sensation

i hate polyamory
polyamory people are ugly
are all polyamorous people ugly
why are polyamorous people ugly
polyamory aspies
polyamory is sick
polyamorists tend to be narcissists
Versus…

good looking polyamorous
pics of women who are polyamory

versailles ohio alien military genetics
A clone of Louis XIV spliced with whale DNA by aliens is being stored in a secret base disguised as a borax mine beneath Columbus by the medical cartel.

many people agree that there should be “some sort” of restraint considering abortion
I think regardless of our beliefs on this issue, we can agree those “quotation marks” are creepy.

what is cost disease by elephant

victim to organ harvesting yankee bob found murdered
If I were a beneficiary of black market organ harvesting, I would be pretty concerned about the possibility that my new organs came from someone named “Yankee Bob”.

explain d theories of truth n d best one dat suits d saying if u can’t beat dem, join dem using events in university as a case study

you r destiny, you r fate, is the gift that god bestowed me. you so different that i had been before. anh is something special that i must keep. i hope time will help me answer everything. time will help me keep you.
I think this person has too much of the hormone responsible for soliloquizing.

satanic company (that public likes)
Apple. Trust me on this one.

Five More Years

Those yearly “predictions for next year” posts are starting to reach the limit of their usefulness. Not much changes from year to year, and most of what does change is hard to capture in objective probabilistic predictions.

So in honor of this blog’s five year anniversary, here are some predictions for the next five years. All predictions to be graded on 2/15/2023:

AI will be marked by various spectacular achievements, plus nobody being willing to say the spectacular achievements signify anything broader. AI will beat humans at progressively more complicated games, and we will hear how games are totally different from real life and this is just a cool parlor trick. If AI translation becomes ~~flawless~~ outstanding, we will hear how language is just a formal system that can be brute-forced without understanding. If AI can generate images and even stories to a prompt, everyone will agree this is totally different from real art or storytelling. Nothing that happens in the interval until 2023 will encourage anyone to change this way of thinking. There will not be a Truckpocalypse before 2023. Technological unemployment will continue to be a topic of academic debate that might show up if you crunch the numbers just right, but there will be no obvious sign that it is happening on a large scale. Everyone will tell me I am wrong about this, but I will be right, and they will just be interpreting other things (change in labor force composition, change in disability policies, effects of outsourcing, etc) as obvious visible signs of technological unemployment, the same as people do now. AI safety concerns will occupy about the same percent of the public imagination as today.

1. Average person can hail a self-driving car in at least one US city: 80%
2. …in at least five of ten largest US cities: 30%
3. At least 5% of US truck drivers have been replaced by self-driving trucks: 10%
4. Average person can buy a self-driving car for less than $100,000: 30%
5. AI beats a top human player at Starcraft: 70%
6. MIRI still exists in 2023: 80%
7. AI risk as a field subjectively feels more/same/less widely accepted than today: 50%/40%/10%

The European Union will not collapse. It will get some credibility from everyone hating its enemies – Brexit, the nationalist right, etc – and some more credibility by being halfway-competent at its economic mission. Nobody will secede from anywhere. The crisis of nationalism will briefly die down as the shock of Syrian refugees wears off, then reignite (possibly after 2023) with the focus on African migrants. At some point European Muslims may decide they don’t like African migrants much either, at which point there may be some very weird alliances.

1. UK leaves EU (or still on track to do so): 95%
2. No “far-right” party in power (executive or legislative) in any of France, Germany, UK, Italy, Netherlands, Sweden, at any time: 50%
3. No other country currently in EU votes to leave: 50%

Countries that may have an especially good half-decade: Israel, India, Nigeria, most of East Africa, Iran. Countries that may have an especially bad half-decade: Russia, Saudi Arabia, South Africa, UK. The Middle East will get worse before it gets better, especially Lebanon and the Arabian Peninsula (Syria might get better, though).

1. No overt major power war in the Middle East (Israel spending a couple weeks destroying stuff in Lebanon doesn’t count): 60%
2. Mohammed bin Salman still in power in Saudi Arabia in 2023: 60%
3. Sub-Saharan Africa averages GDP growth greater than 2.5% over 2018 – 2023: 60%
4. Vladimir Putin is still in charge of Russia: 70%
5. If there’s a war in the Middle East where US intervention is plausible, US decides to intervene (at least as much as it did in Syria): 70%

Religion will continue to retreat from US public life. As it becomes less important, mainstream society will treat it as less of an outgroup and more of a fargroup. Everyone will assume Christians have some sort of vague spiritual wisdom, much like Buddhists do. Everyone will agree evangelicals or anyone with a real religious opinion is just straight-out misinterpreting the Bible, the same way any Muslim who does something bad is misinterpreting the Koran. Christian mysticism will become more popular among intellectuals. Lots of people will talk about how real Christianity opposes capitalism. There may not literally be a black lesbian Pope, but everyone will agree that there should be, and people will become mildly surprised when you remind them that the Pope is white, male, and sexually inactive.

1. Church attendance rates lower in 2023 than 2018: 90%

The crisis of the Republican Party will turn out to have been overblown. Trump’s policies have been so standard-Republican that there will be no problem integrating him into the standard Republican pantheon, plus or minus some concerns about his personality which will disappear once he personally leaves the stage. Some competent demagogue (maybe Ted Cruz or Mike Pence) will use some phrase equivalent to “compassionate Trumpism”, everyone will agree it is a good idea, and in practice it will be exactly the same as what Republicans have been doing forever. The party might move slightly to the right on immigration, but this will be made easy by a fall in corporate demand for underpriced Mexican farm labor, and might be trivial if there’s a border wall and they can declare mission accomplished. If the post-Trump standard-bearer has the slightest amount of personal continence, he should end up with a more-or-less united party who view Trump as a flawed but ultimately positive figure, like how they view GW Bush. Also, I predict we see a lot more of Ted Cruz than people are expecting.

1. Trump wins 2020: 20%
2. Republicans win Presidency in 2020: 40%

On the other hand, everyone will have underestimated the extent of crisis in the Democratic Party. The worst-case scenario is Kamala Harris rising to the main contender against Bernie Sanders in the 2020 primary. Bernie attacks her and her followers as against true progressive values, bringing up her work defending overcrowded California prisons as a useful source of unpaid labor. Harris supporters attack Bernie as a sexist white man trying to keep a woman of color down (wait until the prison thing gets described as “slavery”). Everything that happened in 2016 between Clinton and Sanders looks like mild teasing between friends in comparison. If non-Sanderites rally around Booker or Warren instead, the result will be slightly less apocalyptic but still much worse than anyone expects. The only plausible way I can see for the Dems to avoid this is if Sanders dies or becomes too sick to run before 2020. This could tear apart the Democratic Party in the long-term, but in the short term it doesn’t even mean they won’t win the election – it will just mean a bunch of people who loathe each other temporarily hold their nose and vote against Trump.

1. Sanders wins 2020: 10%
2. Democrats win Presidency in 2020: 60%

It will become more and more apparent that there are three separate groups: progressives, conservatives, and neoliberals. How exactly they sort themselves into two parties is going to be interesting. The easiest continuation-of-current-trends option is neoliberals+progressives vs. conservatives, with neoliberals+progressives winning easily. But progressives are starting to wonder if neoliberals' support is worth the watering-down of their program, and neoliberals are starting to wonder if progressives' support is worth constantly feeding more power to people they increasingly consider crazy. The Republicans used some weird demonic magic to hold together conservatives and neoliberals for a long time; I suspect the Democrats will be less good at this. A weak and fractious Democratic coalition plus a rock-hard conservative Republican non-coalition might be stable under Median Voter Theorem considerations. For like ten years. Until there are enough minorities that the Democrats are just overwhelmingly powerful (no, minorities are not going to start identifying as white and voting Republican *en masse*). I have no idea what will happen then. Maybe the Democrats will go extra socialist, the neoliberals and market minorities will switch back to the Republicans, and we can finally have normal reasonable class warfare again instead of whatever weird ethno-cultural thing is happening now?

1. At least one US state has approved single-payer health-care by 2023: 70%
2. At least one US state has *de facto* decriminalized

hallucinogens: 20%
3. At least one US state has seceded (de jure or de facto): 1%
4. At least 10 members of 2022 Congress from neither Dems or GOP: 1%
5. US in at least new one major war (death toll of 1000+ US soldiers): 40%
6. Roe v. Wade substantially overturned: 1%
7. At least one major (Obamacare-level) federal health care reform bill passed: 20%
8. At least one major (Brady Act level) federal gun control bill passed: 20%
9. Marijuana legal on the federal level (states can still ban): 40%
10. Neoliberals will be mostly Democrat/evenly split/Republican in 2023: 60%/20%/20%
11. Political polarization will be worse/the same/better in 2023: 50%/30%/20%

The culture wars will continue to be marked by both sides scoring an unrelenting series of own-goals, with the victory going to whoever can make their supporters shut up first. The best case scenario for the Right is that Jordan Peterson’s ability to not instantly get ostracized and destroyed signals a new era of basically decent people being able to speak out against social justice; this launches a cascade of people doing so, and the vague group consisting of Jordan Peterson, Sam Harris, Steven Pinker, Jonathan Haidt, etc coalesces into a perfectly respectable force no more controversial than the gun lobby or the pro-life movement or something. With social justice no longer able to enforce its own sacredness values against blasphemy, it loses a lot of credibility and ends up no more powerful or religion-like than eg Christianity. The best case scenario for the Left is that the alt-right makes some more noise, the media is able to relentlessly keep everyone’s focus on the alt-right, the words ALT-RIGHT get seared into the public consciousness every single day on every single news website, and everyone is so afraid of being associated with the alt-right that they shut up about any disagreements with the consensus they might have. I predict both of these will happen, but the Right's win-scenario will come together faster and they will score a minor victory.

1. At least one US politician, Congressman or above, explicitly identifies as alt-right (in more than just one off-the-cuff comment) and refuses to back down or qualify: 10%
2. …is overtly racist (says eg "America should be for white people" or "White people are superior" and means it, as a major plank of their platform), refuses to back down or qualify: 10%
3. Gay marriage support rate is higher on 1/1/2023 than 1/1/2018: 95%
4. Percent transgender is higher on 1/1/2023 than 1/1/2018: 95%
5. Social justice movement appear less powerful/important in 2023 than currently: 60%

First World economies will increasingly be marked by an Officialness Divide. Rich people, the government, and corporations will use formal, well-regulated, traditional institutions. Poor people (and to an increasing degree middle-class people) will use informal gig economies supported by Silicon Valley companies whose main skill is staying a step ahead of regulators. Think business travelers staying at the Hilton and riding taxis, vs. low-prospect twenty-somethings staying at Air BnBs and taking Ubers. As Obamacare collapses, health insurance will start turning into one of the formal, well-regulated, traditional institutions limited to college grads with good job prospects. What the unofficial version of health care will be remains to be seen. If past eras have been Stone Age, Bronze Age, Iron Age, Information Age, etc, the future may be the Ability-To-Circumvent-Regulations Age.

1. Percent of people in US without health insurance (outside those covered by free government programs) is higher in 2023 than 2018: 80%
2. Health care costs (as % of economy) continue to increase at least as much as before: 70%

Cryptocurrency will neither collapse nor take over everything. It will become integrated into the existing system and regulated to the point of uselessness. No matter how private and untraceable the next generation of cryptocurrencies are, people will buy and exchange them through big corporate websites that do everything they can to stay on the government's good side. Multinationals will occasionally debate using crypto to transfer their profits from one place to another, then decide that would make people angry and decide not to. There may be rare crypto-related accounting tricks approximately of the same magnitude as the "headquarter your company in the Cayman Islands" trick. A few cryptocurrencies might achieve the same sort of role PayPal has today, only slightly cooler. Things like Ethereum prediction markets might actually work, again mostly by being too niche for the government to care very much. A few die-hards will use pure crypto to buy drugs over the black market, but not significantly more than do so today, and the government will mostly leave them alone as too boring to crush.

1. 1 Bitcoin costs above $1K: 80%
2. …above $10K: 50%
3. …above $100K: 5%
4. Bitcoin is still the highest market cap cryptocurrency: 40%
5. Someone figures out Satoshi's true identity to my satisfaction: 30%

6. Browser-crypto-mining becomes a big deal and replaces ads on 10%+ of websites: 5%

Polygenic scores go public – not necessarily by 2023, but not long after. It becomes possible to look at your 23andMe results and get a weak estimate of your height, IQ, criminality, et cetera. Somebody checks their spouse's score and finds that their desirable/undesirable traits are/aren't genetic and will/won't be passed down to their children; this is treated as a Social Crisis but nobody really knows what to do about it. People in China or Korea start actually doing this on a large scale. If there is intelligence enhancement, it looks like third-party services that screen your gametes for genetic diseases and *just so happen* to give you the full genome which can be fed to a polygenic scoring app before you decide which one to implant. The first people to do this aren't necessarily the super-rich, so much as people who are able to put the pieces together and figure out that this is an option. If you think genetics discourse is bad now, wait until polygenic score predictors become consumerized. There will be everything from "the predictor said I would be tall but actually I am medium height, this proves genes aren't real" to "Should we track children by genetic IQ predictions for some reason even though we have their actual IQ scores right here?" Also, the products will probably be normed on white (Asian?) test subjects and not work very well on people of other races; expect everyone to say unbelievably idiotic things about this for a while.

1. Widely accepted paper claims a polygenic score predicting over 25% of human intelligence: 70%
2. …50% or more: 20%
3. At least one person is known to have had a "designer baby" genetically edited for something other than preventing specific high-risk disease: 10%
4. At least a thousand people have had such babies, and it's

well known where people can go to do it: 5%
5. At least one cloned human baby, survives beyond one day after birth: 10%
6. Average person can check their polygenic IQ score for reasonable fee (doesn’t have to be very good) in 2023: 80%
7. At least one directly glutamatergic antidepressant approved by FDA: 20%
8. At least one directly neurotrophic antidepressant approved by FDA: 20%
9. At least one genuinely novel antipsychotic approved by FDA: 30%
10. MDMA approved for therapeutic use by FDA: 50%
11. Psilocybin approved for general therapeutic use in at least one country: 30%
12. Gary Taubes’ insulin resistance theory of nutrition has significantly more scholarly acceptance than today: 10%
13. Paleo diet is generally considered and recommended by doctors as best weight-loss diet for average person: 30%

There will be two or three competing companies offering low-level space tourism by 2023. Prices will be in the $100,000 range for a few minutes in suborbit. The infrastructure for Mars and Moon landings will be starting to look promising, but nobody will have performed any manned landings between now and then. The most exciting edge of the possibility range is that five or six companies are competing to bring rich tourists to Bigelow space stations in orbit.

1. SpaceX has launched BFR to orbit: 50%
2. SpaceX has launched a man around the moon: 50%
3. SLS sends an Orion around the moon: 30%
4. Someone has landed a man on the moon: 1%
5. SpaceX has landed (not crashed) an object on Mars: 5%
6. At least one frequently-inhabited private space station in orbit: 30%

Global existential risks will hopefully not be a big part of the 2018-2023 period. If they are, it will be because somebody did something incredibly stupid or awful with infectious diseases. Even a small scare with this will provoke a massive response, which will be implemented in a panic and with all the finesse of post-9/11 America determining airport security. Along with the obvious ramifications, there will be weird consequences for censorship and the media, with some outlets discussing other kinds of biorisks and the government wanting them to stop giving people ideas. The world in which this becomes an issue before 2023 is not a very good world for very many reasons.

1. Bioengineering project kills at least five people: 20%
2. …at least five thousand people: 5%
3. Paris Agreement still in effect, most countries generally making good-faith effort to comply: 80%
4. US still nominally committed to Paris Agreement: 60%

And just for fun…

1. I actually remember and grade these predictions publicly sometime in the year 2023: 90%
2. Whatever the most important trend of the next five years is, I totally miss it: 80%
3. At least one prediction here is horrendously wrong at the “only a market for five computers” level: 95%

If you disagree, make your own predictions *with probabilities*. I’m tired of people offering to bet me on these and I’m not interested unless you provide me overwhelmingly good odds.

Current list of updates here.

Technological Unemployment: Much More Than You Wanted To Know

[I am not an economist or an expert on this topic. This is my attempt to figure out what economists and experts think so I can understand the issue, and I'm writing it down to speed your going through the same process. If you have more direct access to economists and experts, feel free to ignore this]

*[**EDIT:** Commenters pointed out that I was confusing the trend in total manufacturing jobs with the trend in percent jobs in manufacturing. This would make me less confident that there is no technological unemployment, except that I think the trend of percent jobs in manufacturing represents women entering the workforce and overwhelmingly going into nonmanufacturing jobs. Unclear that this requires a major change in conclusion.]*

Technological unemployment is a hard topic because there are such good arguments on both sides.

The argument against: we've had increasing technology for centuries now, people have been predicting that technology will put them out of work since the Luddites, and it's never come true. Instead, one of two things have happened. Either machines have augmented human workers, allowing them to produce more goods at lower prices, and so expanded industries so dramatically that overall they employ more people. Or displaced workers from one industry have gone into another – stable boys becoming car mechanics, or the like. There are a bunch of well-known theoretical mechanisms that compensate for technological displacement – see Vivarelli for a review. David Autor gives a vivid example:

> Consider the surprising complementarities between information technology and employment in banking, specifically the experience with automated teller machines (ATMs) and bank tellers documented by Bessen (2015). ATMs were introduced in the 1970s, and

> their numbers in the US economy quadrupled from approximately 100,000 to 400,000 between 1995 and 2010. One might naturally assume that these machines had all but eliminated bank tellers in that interval. But US bank teller employment actually rose modestly from 500,000 to approximately 550,000 over the 30-year period from 1980 to 2010 (although given the growth in the labor force in this time interval, these numbers do imply that bank tellers declined as a share of overall US employment). With the growth of ATMs, what are all of these tellers doing?
>
> Bessen observes that two forces worked in opposite directions. First, by reducing the cost of operating a bank branch, ATMs indirectly increased the demand for tellers: the number of tellers per branch fell by more than a third between 1988 and 2004, but the number of urban bank branches (also encouraged by a wave of bank deregulation allowing more branches) rose by more than 40 percent. Second, as the routine cash-handling tasks of bank tellers receded, information technology also enabled a broader range of bank personnel to become involved in “relationship banking.” Increasingly, banks recognized the value of tellers enabled by information technology, not primarily as checkout clerks, but as salespersons, forging relationships with customers and introducing them to additional bank services like credit cards, loans, and investment products.

This kind of thing has been remarkably consistent – so much so that arguments that *today* is the day technological unemployment happens should be treated with the same skepticism as arguments that *today* is the day we build a perpetual motion machine that works.

The argument in favor: look, imagine there's a perfect android that can do everything humans do (including management) only better. And suppose it costs $10 to buy and $1/hour to operate. Surely every business owner would just buy those androids, and then all humans who wanted to earn more than $1/hour would be totally out of luck. There's no conceivable way the androids would "augment" human labor and there's no conceivable way the displaced humans could go into another industry. So at *some point* we've got to start getting technological unemployment. Here the vivid example comes from Gregory Clark

> There was a type of employee at the beginning of the Industrial Revolution whose job and livelihood largely vanished in the early twentieth century. This was the horse. The population of working horses actually peaked in England long after the Industrial Revolution, in 1901, when 3.25 million were at work. Though they had been replaced by rail for long-distance haulage and by steam engines for driving machinery, they still plowed fields, hauled wagons and carriages short distances, pulled boats on the canals, toiled in the pits, and carried armies into battle. But the arrival of the internal combustion engine in the late nineteenth century rapidly displaced these workers, so that by 1924 there were fewer than two million. There was always a wage at which all these horses could have remained employed. But that wage was so low that it did not pay for their feed.

There may be some point at which we too stop being worth more than it costs to replace us. And the decline of manufacturing, the increase in labor force nonparticipation and despair in rural Rust Belt communities, etc, suggest that point is fast arriving.

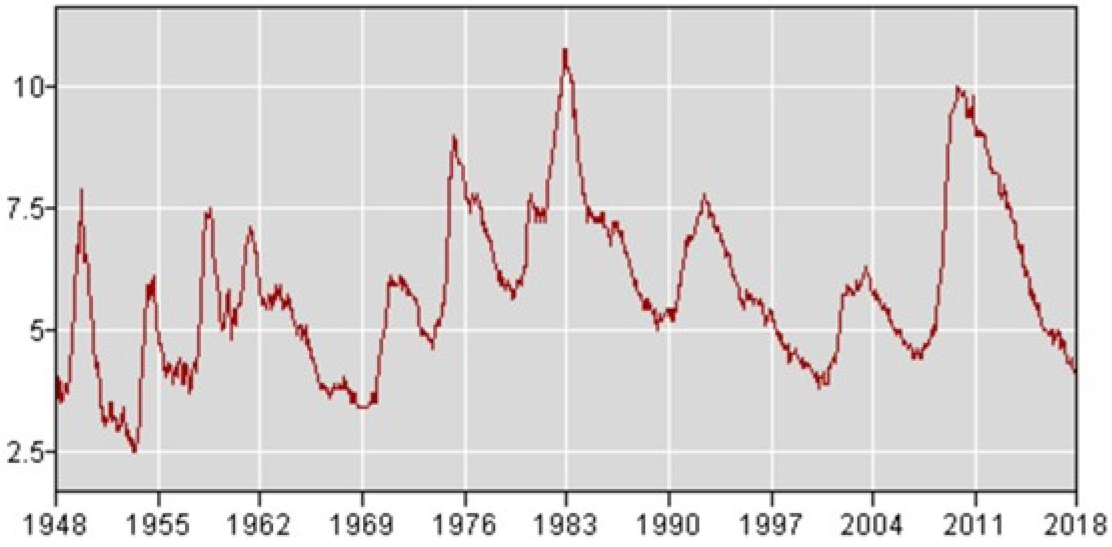
This is a look at which of those arguments is right. Part I will investigate whether unemployment is getting worse. Part II will investigate whether that is because of technology. Part III will investigate what longer-term trends we should expect.

As usual, this is very long.

I.

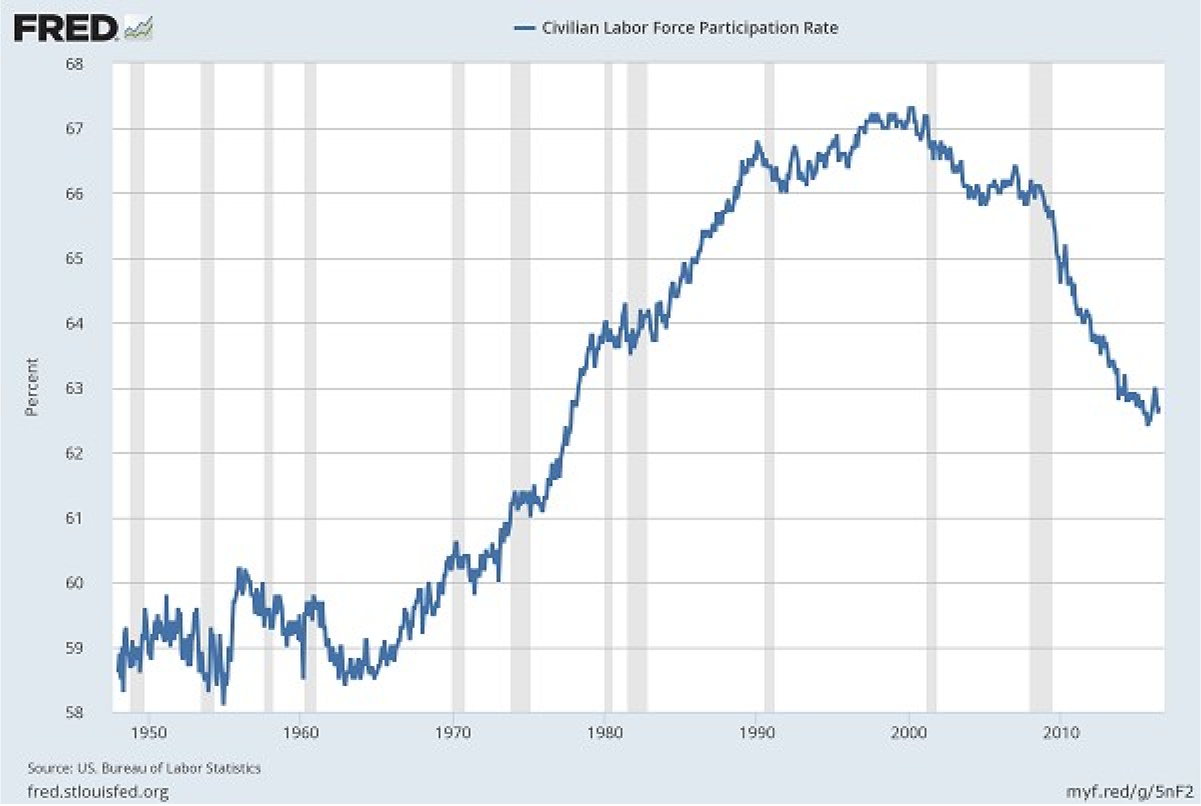
Is unemployment actually getting worse?

Officially it's at historic lows.



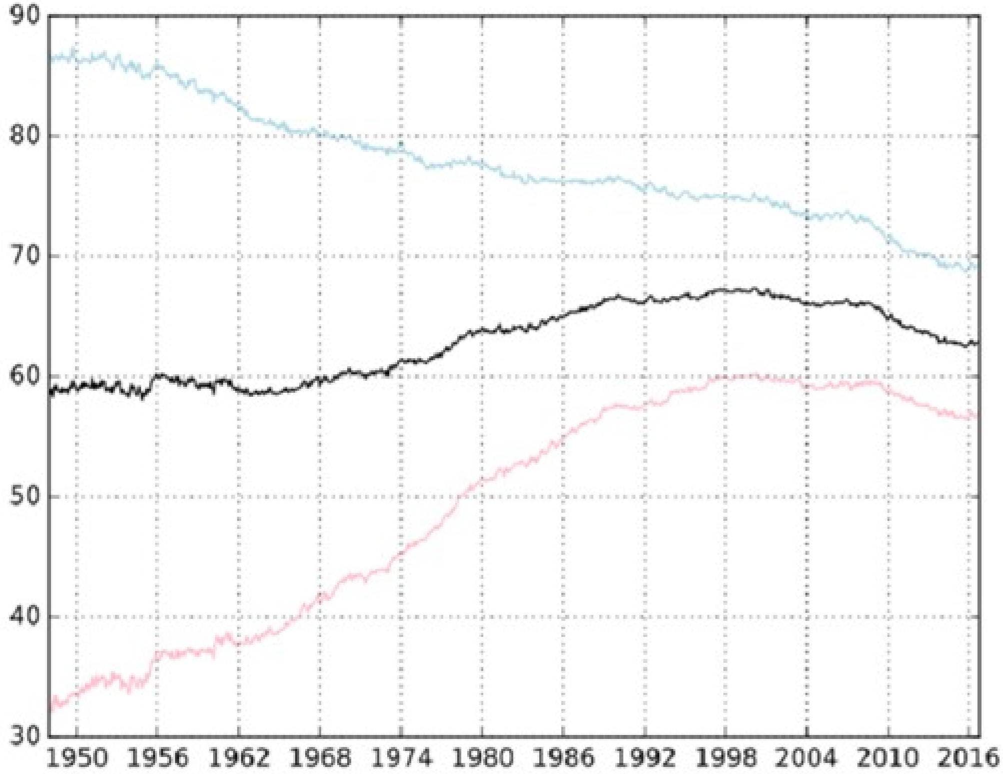
Source: Bureau of Labor Statistics. Just assume this for any graph that looks like this.

But the real concern isn't about unemployment per se, but the labor force participation rate (from here on: LFPR). Unemployment measures how many people are looking for work; LFPR measures how many people are out of work and not looking. If people are so discouraged that they've given up looking for work, this would not show up in unemployment but would show up in LFPR. Here’s LFPR:



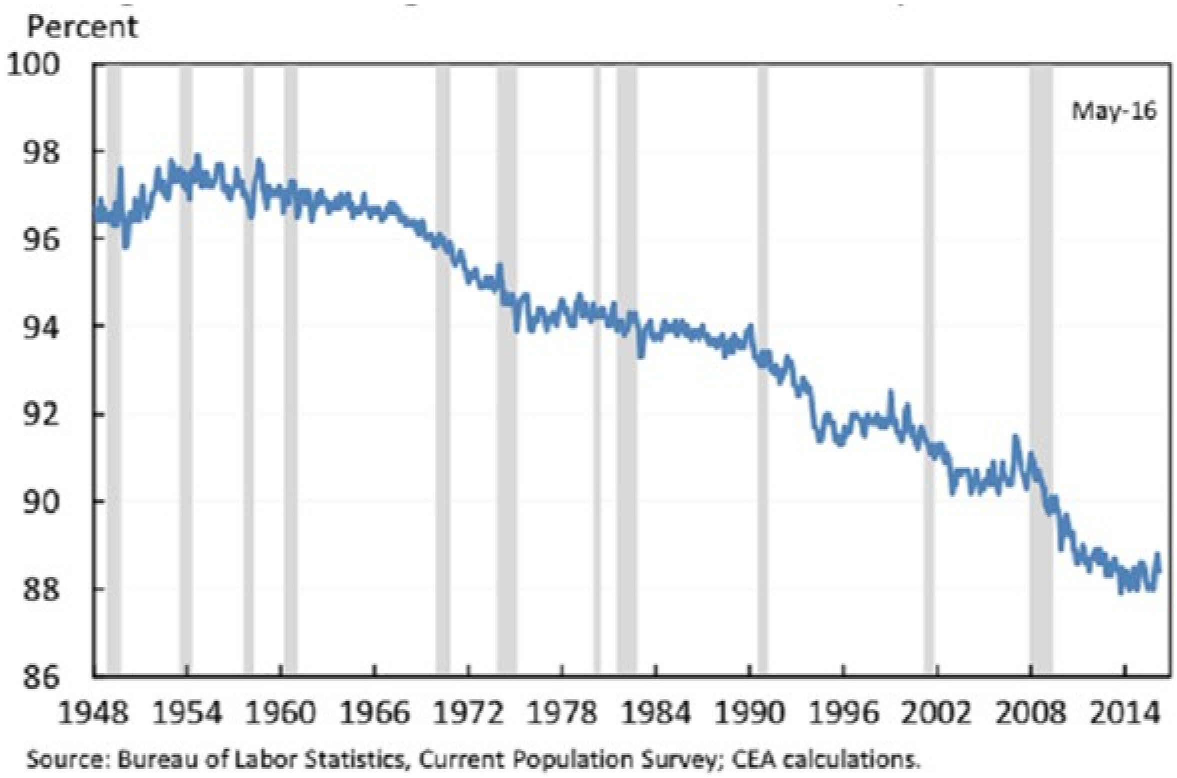
Remember, higher means more people working

What’s going on here?



Same data as before, only disaggregated by gender. From 1950 to 2000, workforce participation rises as women enter the workforce. But throughout this time, men are leaving at almost the same rate, leaving only a moderate net participation gain (and incidentally answering the question that confused me here). Around 2000, all the women who want to be in the workforce are there already, and the declining male trend takes over for a net decline. Because women's increasing workforce participation is shaped by unrelated cultural trends, most of the rest of this article will focus on male LFPR (from here on: MLFPR).

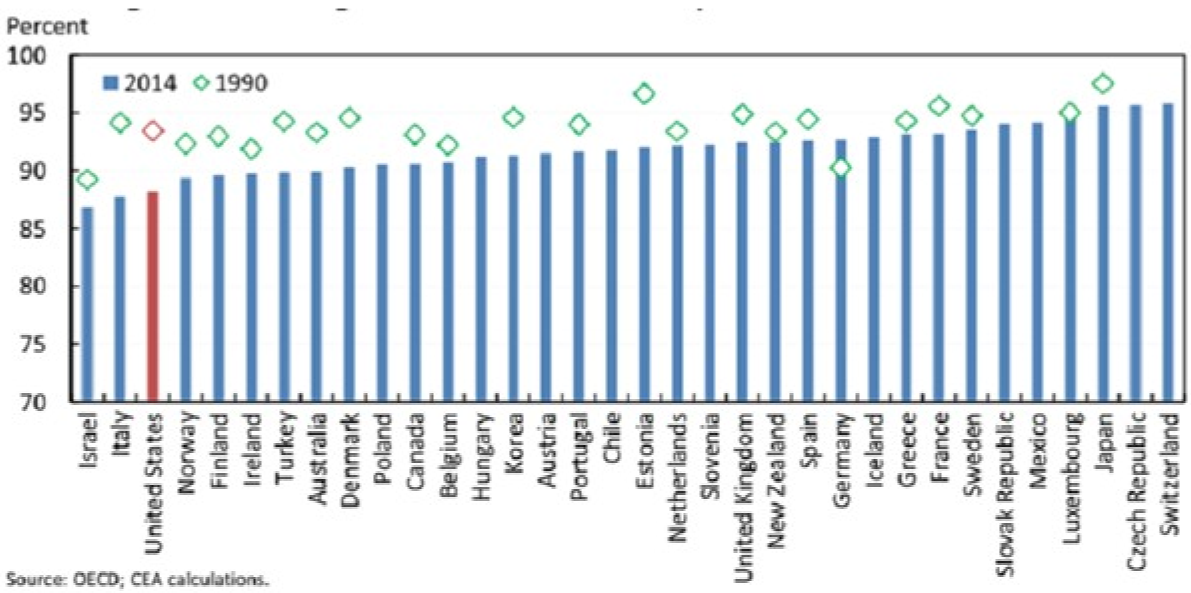
A big fraction of declining MFLPR is Baby Boomers retiring. If most people are young, workforce participation will be high. If most people are old, workforce participation will be low. Economists adjust for this by taking something called prime age male labor force participation rate (from here on: PAMLFPR; if you think all these acronyms are getting annoying, I guarantee you it's more annoying to read papers that just keep saying "prime age male labor force participation rate" again and again). Here it is:



Source: Bureau of Labor Statistics, Current Population Survey; CEA calculations.

Source for this and subsequent similar-looking graphs.

In the 1950s, ~97% (!) of prime-age men had a job. Today that number is more like 88%. This is the decline people are worrying about when they talk about technological unemployment or any other threat to work, and it seems to be happening across the Western world:



Source: OECD; CEA calculations.

Frickin' Germany, always making everyone else look bad.

The next few paragraphs are based on data from a Scott Winship's Mercatus Center report on this. Their conclusion is contrarian, and they're a libertarian-ish think tank which means they have some risk of bias. I'm citing them anyway because they have really fascinating data presented much better than anyone else, but keep this in mind.

Winship's first point is that the decline in PAMLFPR doesn't seem to be caused by people who can't find jobs:

FIGURE 6. DISCOURAGED WORKERS AS A SHARE OF PRIME-AGE MEN INACTIVE IN THE LABOR FORCE

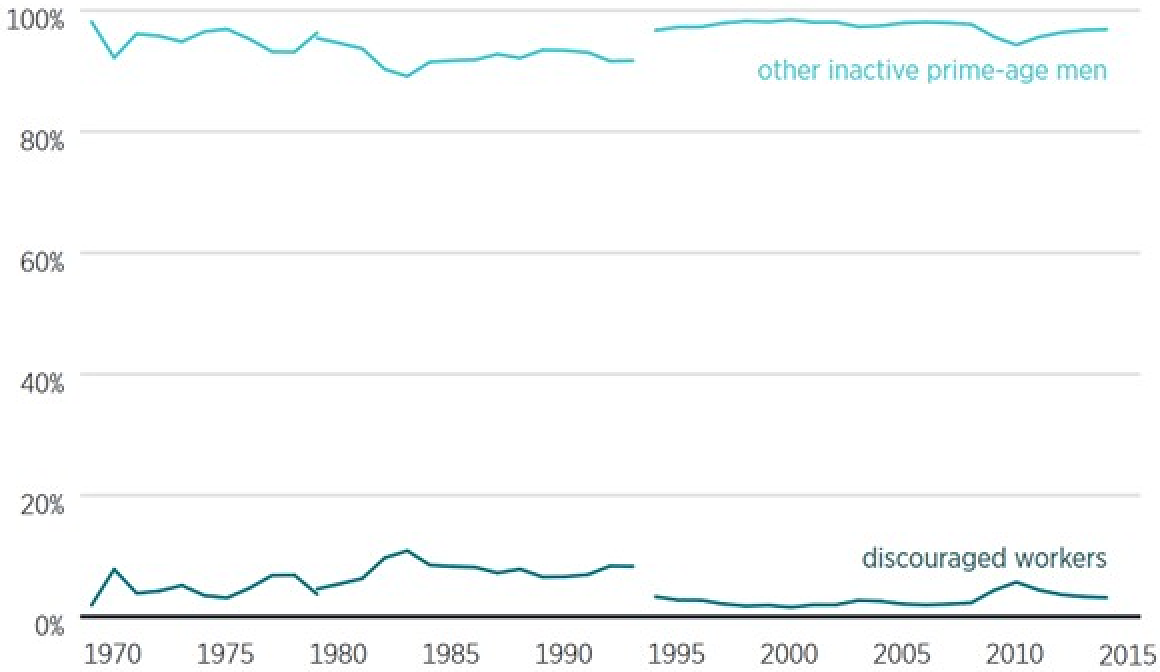


FIGURE 12. PRIME-AGE MEN INACTIVE IN THE LABOR FORCE, BY DESIRE FOR A JOB, 1969–1993

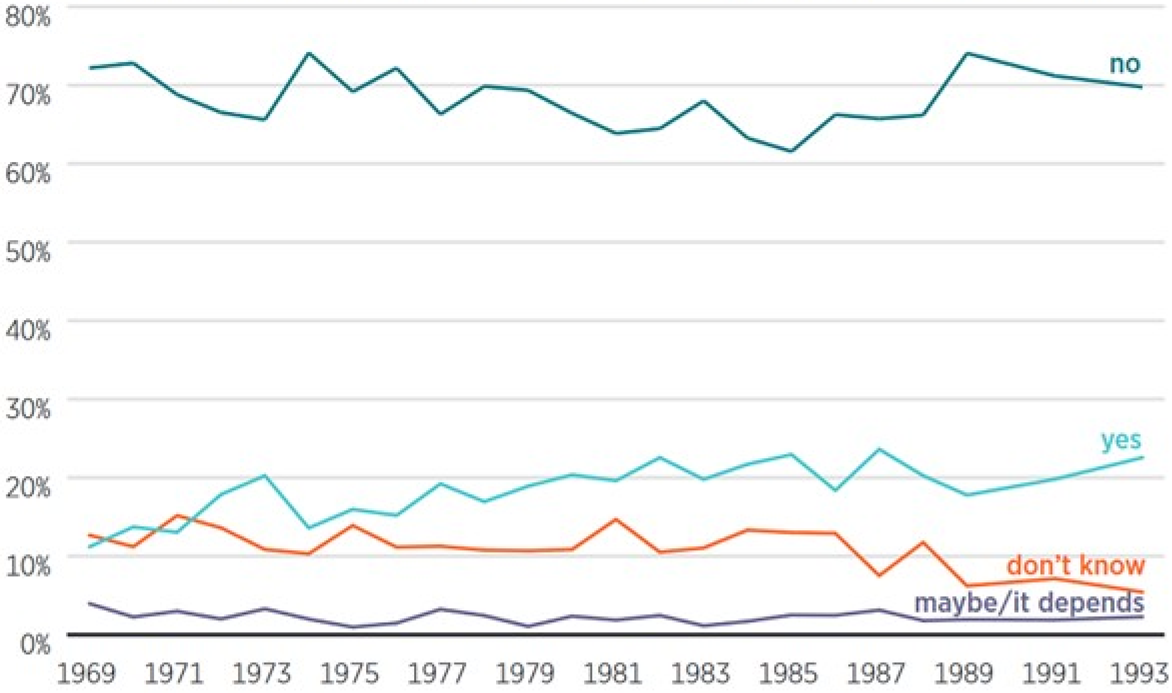
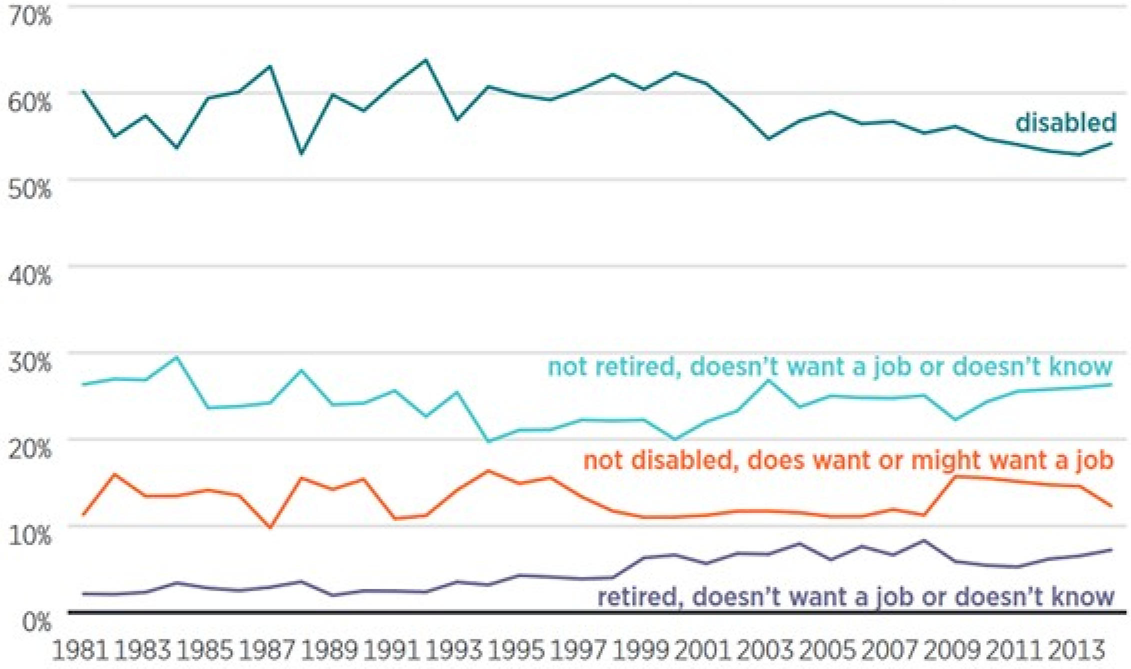


FIGURE 14. PRIME-AGE MEN INACTIVE IN THE LABOR FORCE, BY DESIRE FOR A JOB, 1981–2014



Source: Author's tabulations from the Current Population Survey, Annual Social and Economic Supplement and March Basic Monthly Survey data.
Note: It is not possible to disaggregate "wants job" from "maybe, it depends" or "doesn't want job" from "don't know" after 1993, so they are combined for 1981 to 1993.

The first graph shows officially designated "discouraged workers", people who were previously unemployed and looking for jobs but eventually gave up. The definition changed around 1990, but they never seem to be more than about 10% of prime age male labor force nonparticipators (from here: PAMLFNPers).

The second graph shows what percent of PAMLFNPers claimed to be looking for jobs, based on a survey that was only given until 1993. It shows only about 20% of them were interested.

The third graph is using slightly differently-parsed data to try to continue the trends after 1993. It's sort of unfair, because it separates out "disability" into a separate category and assumes none of them want jobs. But about

30% of people on disability do say they want a job. It’s unclear exactly what they mean (are they just saying the wish they weren’t disabled? or that they might be willing to come off disability if a properly non-physically strenuous job became available?) but even if we count them, the percent of PAMLFNPers who want a job never goes above 30%.

If not discouraged workers who can’t find jobs, what’s going on? Here’s Winship’s answer:

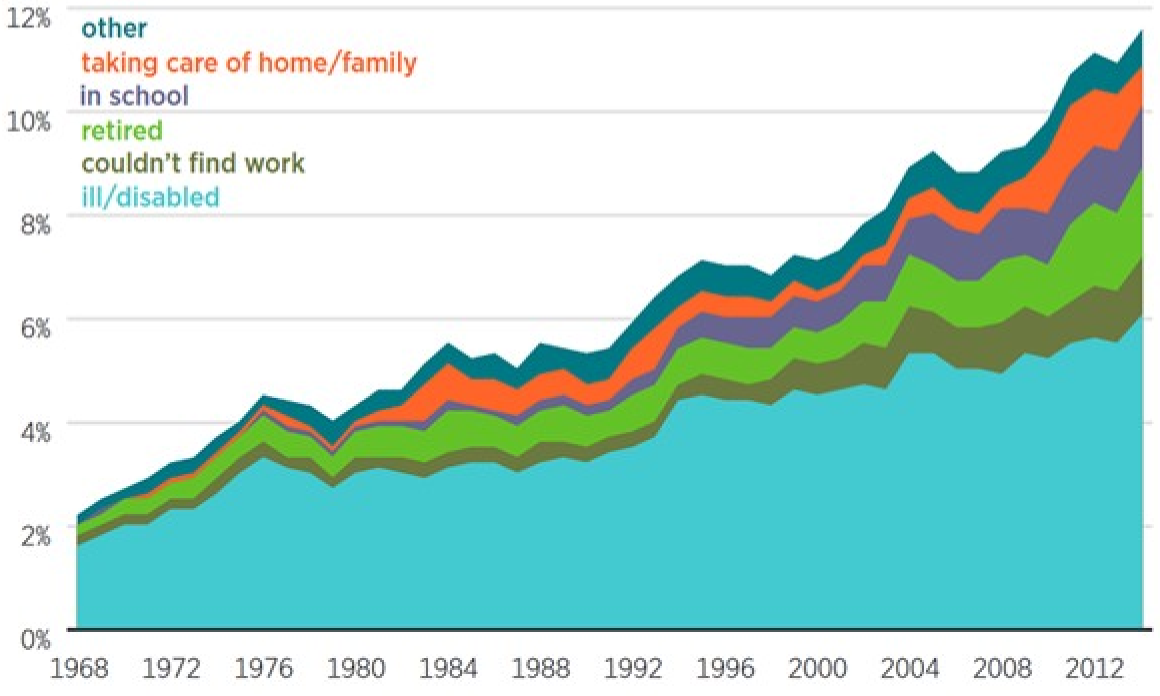


FIGURE 11. RISE IN MALE INACTIVITY IN THE LABOR FORCE THE ENTIRE PREVIOUS YEAR, BY SITUATION, 1968–2014

Source: Author’s tabulations from the Current Population Survey, Annual Social and Economic Supplement data.

Note: The “taking care of home/family” series may suffer from a data problem from 1970 to 1975, when zero cases appear in the data. However, since this category is a small share of cases in neighboring years, I leave the 1970-to-1975 segment in the chart, and the distortion to the other series is likely minimal.

Prime age non-working men are mostly on disability. But some are also in school (despite having to be above 25 to be included as “prime age”), retired (despite having to be below 55), or homemakers (remember, these are all men).

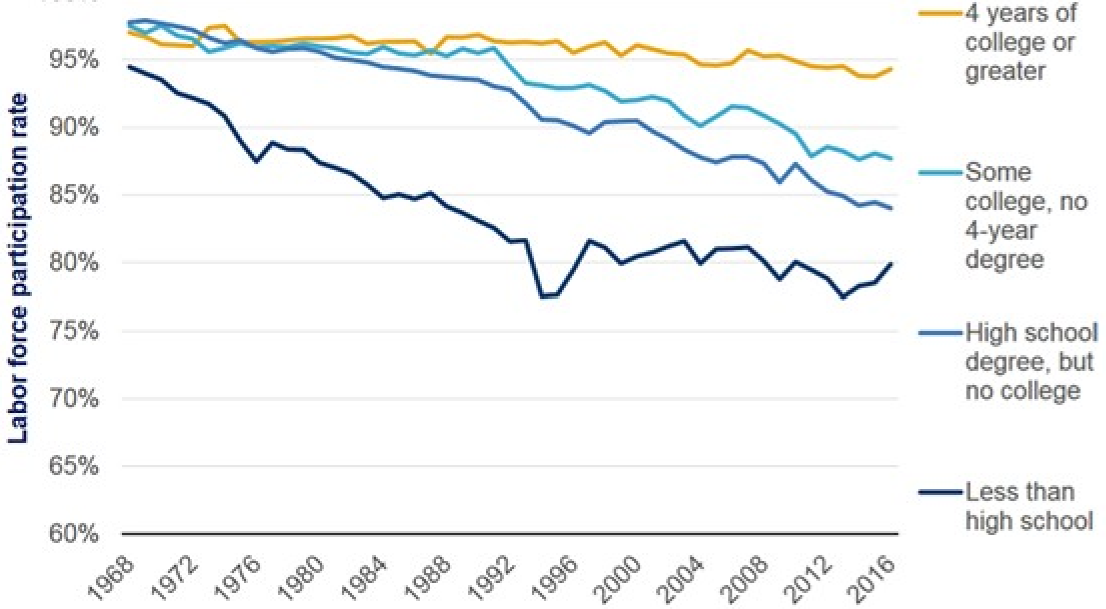
Again, only about 1% (out of the total of 12%) say they can't find work.

If we were very optimistic, we could paint a rosy picture of what's going on here. The increase in disability represents improving social safety net that allows disabled people to be better supported. It's great that more people are financially secure enough to retire early. It's great that more people are pursuing a graduate education that has them in school after age 25. It's great that gender stereotypes are decreasing and more men feel comfortable as homemakers, perhaps supported by a working spouse.

This is basically Winship's account, although he is concerned that increasing disability benefits are discouraging work. He cites a bunch of papers to that effect which you can find in his footnote 42, and which aren't super relevant to the question at hand.

But what would the pessimistic interpretation look like?

Figure 4. Least-educated prime-age men now the least likely to be in labor force



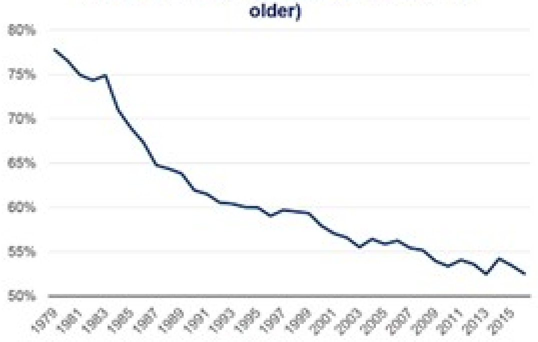
Source: BLS Current Population Survey; population includes men ages 25-54
Note: The CPS method of assessing educational attainment changed in 1992, so values prior to and after this date are not perfectly comparable.

The next few graphs and some of the analysis below comes from Brookings Institute, another potentially biased think tank

In 1970, educated and uneducated men were about equally likely to be PAMLFNPers. The rate for educated men didn’t change. The rate for uneducated men shot up.

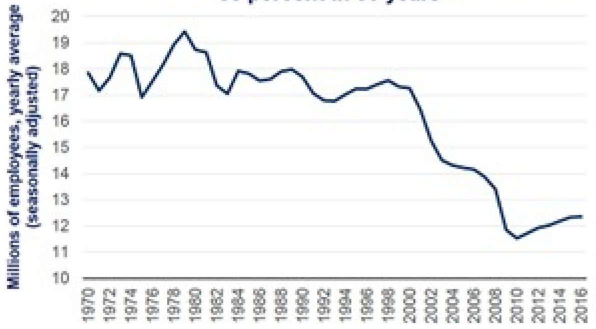
And I won’t show you graphs, but there are similar trends for poor people, ex-convicts, blue collar workers, and minorities. These are not the sort of people who are likely to be able to retire early, pursue graduate school, or defy gender norms. But they *are* the sort of people who might have trouble finding work. This is pretty suspicious. Also:

Figure 7. Ratio of high school graduate wages to college graduate wages (men, 25 years and older)



Source: BLS, "Weekly and hourly earnings data from the Current Population Survey."

Figure 6. Manufacturing employment has fallen 30 percent in 30 years



Source: BLS Series ID CES3000000001.

Labor force nonparticipation is increasing primarily in poor and lower-middle-class people without a lot of good options, *just as* their remaining options get much worse. Surely this suggests something worse is going on.

The easiest place for this to happen is disability. It doesn't require disability fraud, per se. It just requires some people *on the threshold* of disability who are motivated by marginal cost/benefit analysis.

Suppose that you have bad back pain. You work in the auto factory, like your father and his father before him. Your back pain flares up pretty often, but you know your foreman pretty well and he gives you an easy shift until it passes, and the union makes sure that nobody gives you any grief about it. You like your company and your coworkers and you want to make them happy. Also, if you didn't work, you would starve to death.

Now suppose that your factory closes, and the only job available is being a home health aide. This involves a lot of bending over and puts you in constant almost-unbearable pain. And it's run by a giant faceless corporation which always seems to be trying to screw you over. Also, you live in West Virginia and are very manly, and changing diapers in nursing homes seems like undignified women's work.

Also, the pay is half what you're used to. Also, the government just passed a new law making disability benefits much more generous and easier to get. So…

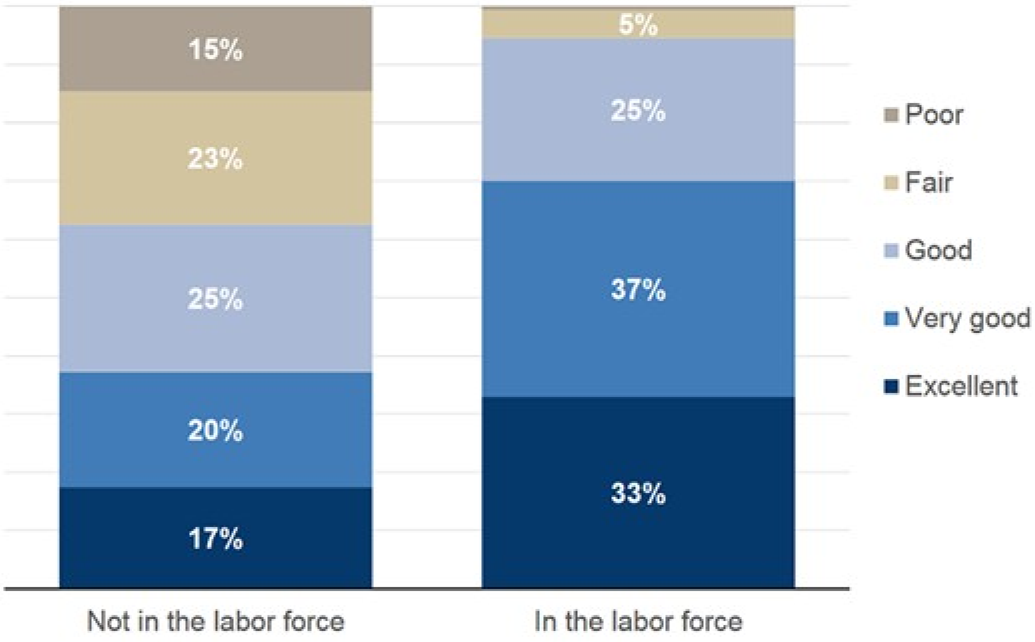


Figure 10. Prime-age men out of the labor force report worse health status than those in the labor force

Source: BLS Current Population Survey, 2016.

The graph shows that PAMLFNPers generally have terrible health. So real disabilities must have something to do with this. But Winship presents a lot of evidence that illnesses and chronic pain haven't gotten worse over time, so it can't fully explain the rise. The "gradually worsening job pushes person with serious disability over the edge" hypothesis has a lot going for it. Also: although 96% of people on disability say they're out of work because of a health problem, 46% also say they're out of work because there are no good jobs available (source).

The problem with this is: disability really doesn't represent that much of the rise in PAMLFNP since 1960. Looking at our graph above, it goes from 2% of workers in ’68 to maybe 6% of workers now. And surely some large fraction of those people are actually disabled in ways that have nothing to do with their social circumstances. We’re talking like 2 percentage points tops.

Okay, fine. Let’s say you’re our West Virginia factory worker again, only now you can’t get on disability. Now what?

Maybe you choose to retire. And maybe you’re 53 years old and this isn’t the *most* reasonable financial plan, but you own your house, you get food stamps, and you can do odd jobs around your friends’ farm to make some extra money.

Or maybe you choose to go to that ridiculous Coal Miner To Coder school that got profiled on NPR a little while ago, in the hopes that you can have a pathway to a new career, or just so that you have something to do.

Or maybe you choose to stay at home with your kids, while your wife does the home health aide thing, and if anybody asks, you’re a “stay-at-home dad”.

And then when economists look at the statistics, they say “Oh, look, there’s no problem here, it’s just a combination of retirees, students, homemakers, and the disabled.”

I realize this is a stretch, especially since you would expect such a person, unless they were very self-deluded, to identify as “looking for work”. But the only sense I can make of all this is a model where, the more miserable your work is, and the more decent options you have available to you, the more likely you are to leave work.

If you’re very conservative, you might say – aha, I knew that people were just unemployed because they’re lazy!

But if you’re more progressive, you might ask – exactly how miserable do you have to be before you stop working? Should people with broken legs literally drag themselves on all fours to their workplace, just because it’s not physically impossible? I know that “I refuse to do this job because it’s too undignified for me, let me go on the public dole” doesn’t really win you a lot of social credit. But maybe conservatives could find it in their heart to be sympathetic to our hypothetical West Virginia factory worker with a bad back, who’s proud of having worked his whole life, but who feels like having to pivot at age 53 to changing diapers in nursing homes for minimum wage isn’t his cup of tea?

(remember, the other thing that’s way up among this same demographic is suicide – and probably for the same reason)

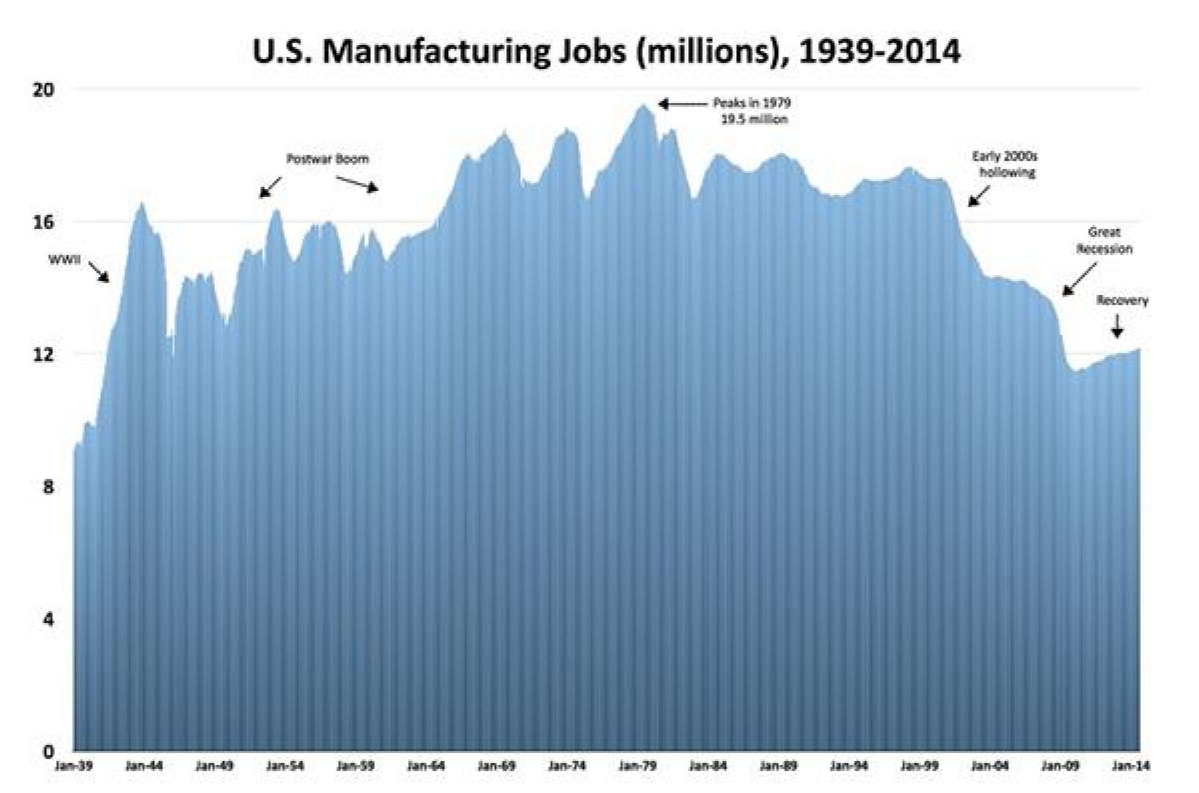
But even if we assume half the increase in disability plus a quarter of the increase in the other things is due to employment issues, employment issues still really only explain about 3 pp of the 10% increase since the 1960s. I can’t think of any reasonable assumptions where they explain more than half.

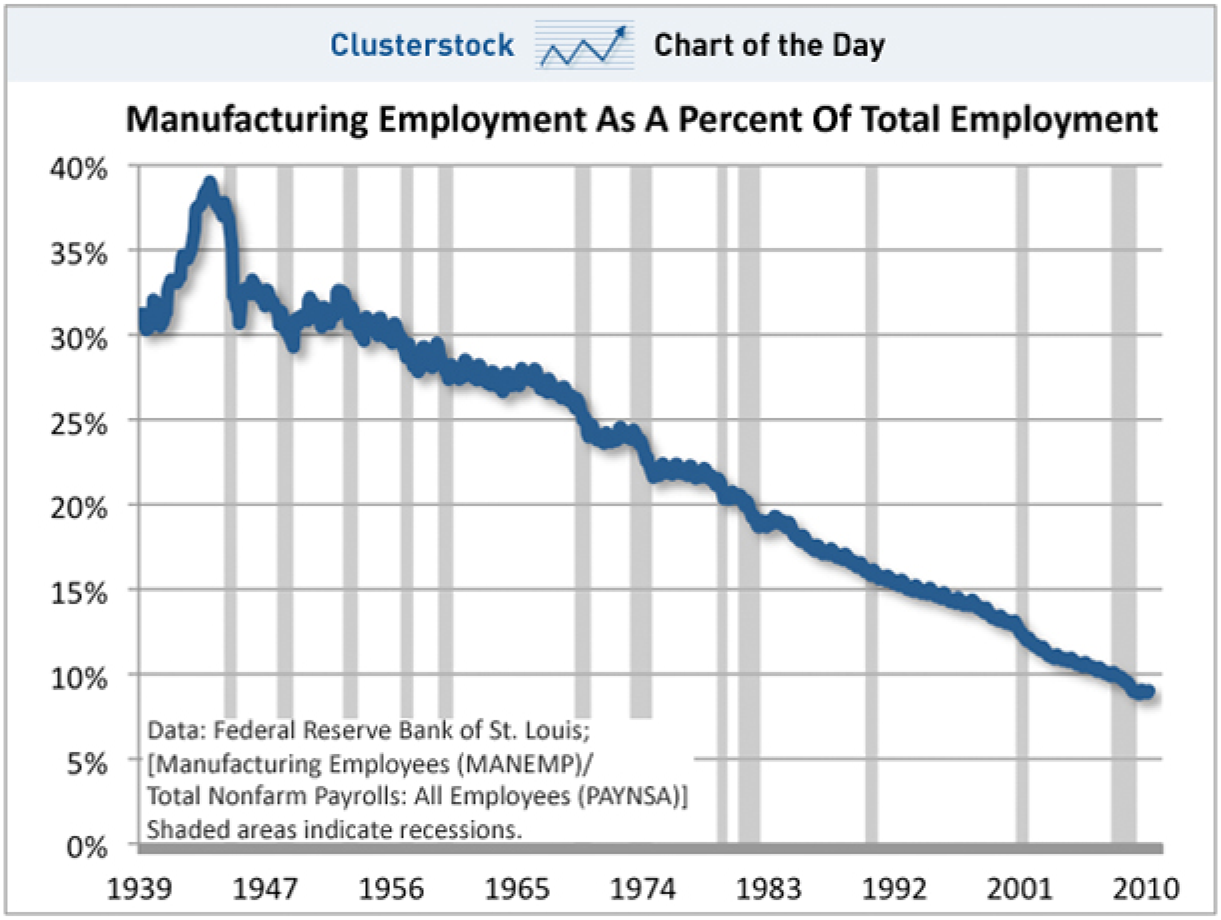
I like Derek Thompson’s discussion of this question, because he’s the only writer who seems to share my sense of puzzlement. There are all these men who seem miserable and who have vanished from the labor force. We all know it’s true. But the statistics don’t really seem to reflect or shed light on it. Somehow we, as a country, have managed to just *lose* several million working-age men. Maybe Donald Trump is going to look behind the White House couch one day, and find a large portion of the male population of the Southeast under the cushion. I don’t know.

In the next part, we’ll talk about whether automation explains the decline in labor force participation. But let’s keep in mind that the argument that there is a significant meaningful decline in labor force participation to explain, aside from people going to school and having more access to disability benefits and things like that – is not on super-solid ground.

II.

Recently, US manufacturing jobs collapsed. US manufacturing is still doing just fine in terms of number of widgets produced. It just no longer employs that many people.





Source 1, Source 2

The two graphs differ a bit in emphasis. The first, in raw numbers, makes it look like there was a big discontinuity around 2000. The second, in percent, suggests that percent manufacturing has been shrinking for a long time. This could represent a long-term decline in manufacturing – but more likely, it represents women entering the workforce; women are much less likely than men to have manufacturing jobs. The reality is probably somewhere in between the two.

Is the decline in manufacturing – whichever timescale we choose to look on – due to automation, or to other causes like globalization? A Ball State University report argues for the former:

> Had we kept 2000-levels of productivity and applied them to 2010-levels of production, we would have required 20.9 million manufacturing workers. Instead, we employed only 12.1 million.

The report gets summarized in a few places as “13% of job loss is due to trade, 87% is due to increasing productivity/automation”, which seems like a fair summary of some of its claims. Although some commenters raise doubts, its numbers are in the same direction as the conclusions of economists Autor, Dorn, and Hanson, who published a series of papers finding that that “import competition explains one-quarter of the contemporaneous aggregate decline in US manufacturing employment”.

Hasn’t productivity actually been growing more slowly than usual recently? See for example this this Brookings paper, which notes that:

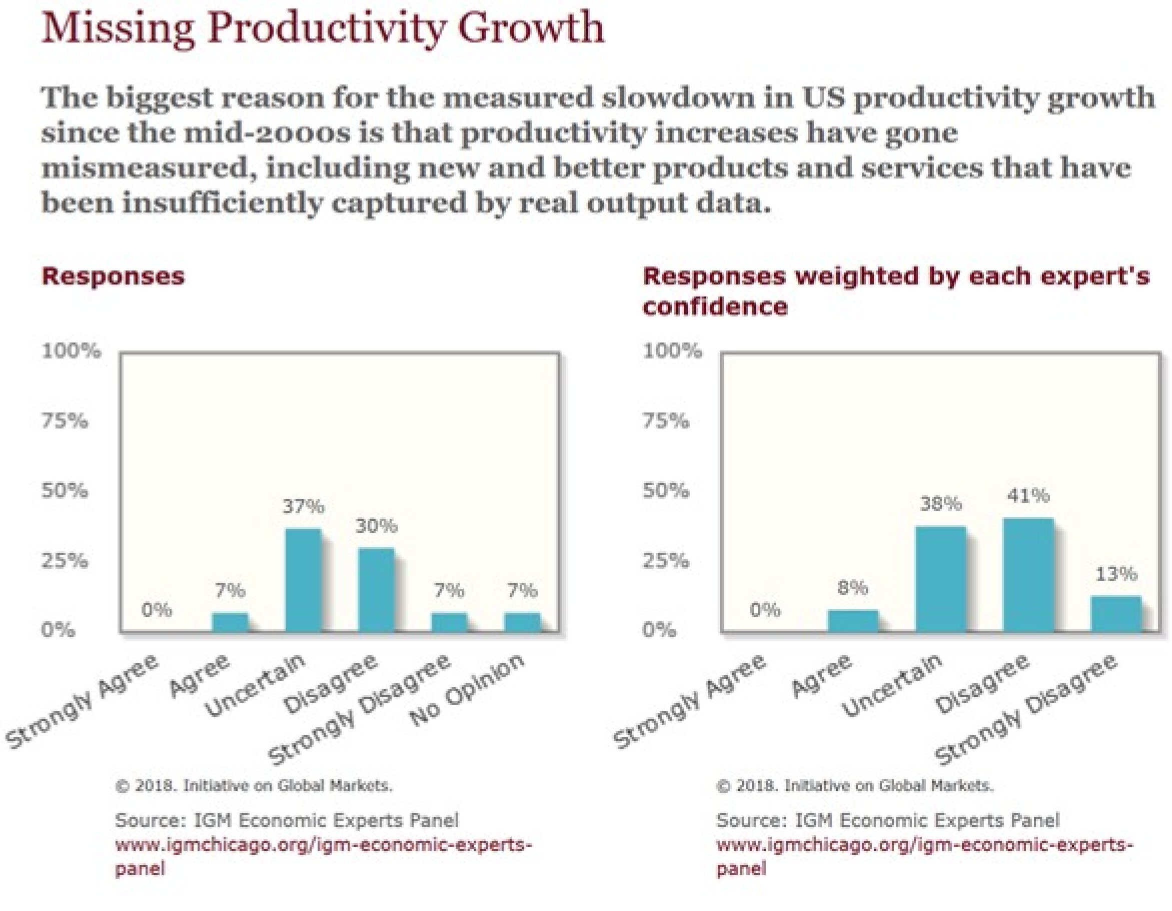
> The past decade has seen slowdowns in measured labor productivity growth across a broad swath of developed economies. Aggregate labor productivity growth in the U.S. averaged 1.3% per year from 2005 to 2015, less than half of the 2.8% average annual growth rate it sustained over 1995 – 2004. Similarly sized decelerations were observed between these two periods in 28 of 29 other countries for which the OECD has compiled productivity growth data

It continues:

> The drops in productivity growth have struck some as paradoxical, given the seemingly brisk pace of technological progress and plethora of new products that have been introduced and diffused throughout the world during the slowdown period. Indeed, many have

> suggested that the slowdown is substantially illusory, a figment of the inability of current economic statistics to capture the true rate of technological advance in standard productivity metrics.

If the productivity slowdown were illusory, that would help explain the apparent speedup in automation, and all those job losses. Sounds promising, but what do all the economists in the world think?



Missing Productivity Growth

The biggest reason for the measured slowdown in US productivity growth since the mid-2000s is that productivity increases have gone mismeasured, including new and better products and services that have been insufficiently captured by real output data.

Responses

Responses weighted by each expert's confidence

© 2018. Initiative on Global Markets.

Source: IGM Economic Experts Panel www.igmchicago.org/igm-economic-experts-panel

© 2018. Initiative on Global Markets.

Source: IGM Economic Experts Panel www.igmchicago.org/igm-economic-experts-panel

100%
75%
50%
25%
0%
0%
7%
37%
30%
7%
7%
Strongly Agree
Agree
Uncertain
Disagree
Strongly Disagree
No Opinion
100%
75%
50%
25%
0%
0%
8%
38%
41%
13%
Strongly Agree
Agree
Uncertain
Disagree
Strongly Disagree

Darn.

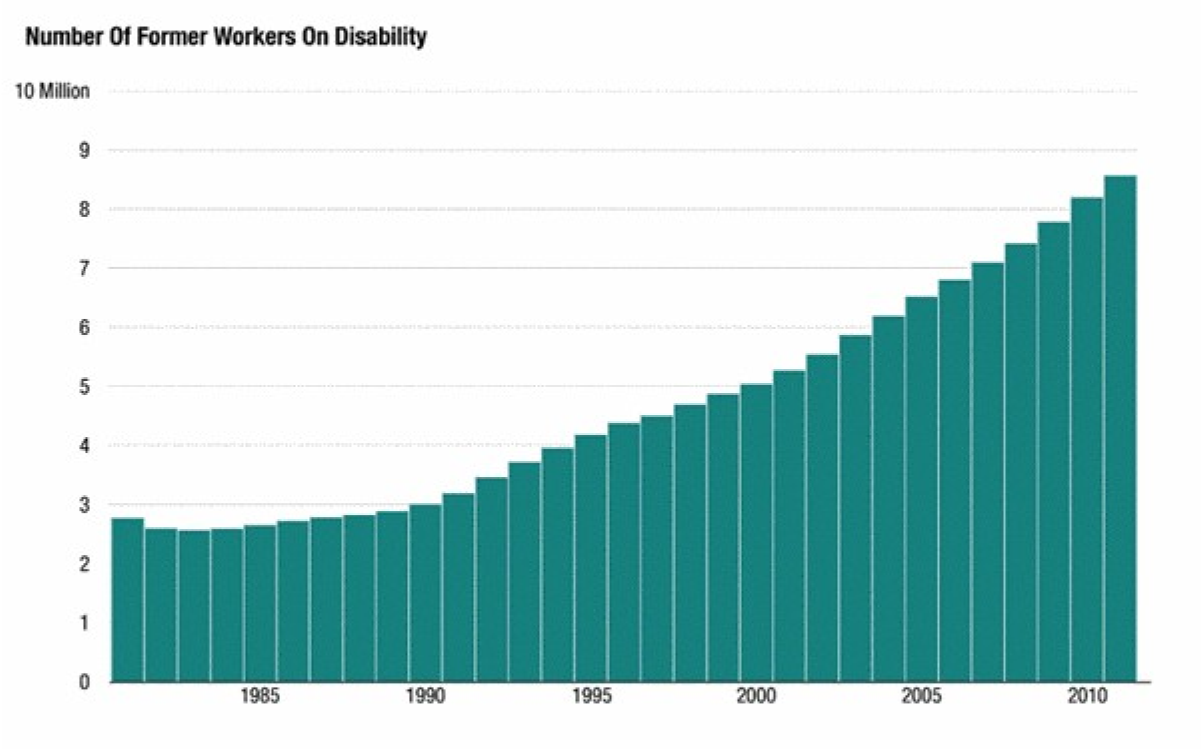
Justin Fox has some more in-depth analysis here and also concludes productivity is not that great.

Are we allowed to say “that’s just how things work”? Like, agricultural productivity increased for millennia, but didn’t lead people to abandon agriculture *en masse* until the Industrial Revolution – when it did exactly that. In 1790, 90% of Americans were farmers, even though agricultural productivity had been improving for ages. Today, 2.6% of Americans are. Maybe manufacturing just had the same kind of moment. Advances in technology can put farmers out of work (but shift them to manufacturing) – surely it can put manufacturers out of work (and shift them to ???).

So, if 70% – 80% of manufacturing job losses were due to automation, might automation be responsible for the decline in PAMLFPR, thus revealing the elusive technological unemployment?

The time trend in absolute number of manufacturing jobs reveals a sharp drop around 2000, which doesn’t match change in PAMLFNP. The trend in percent reveals a gradual decline, which does sort of match the PAMLFNP trend, but because the percent is likely to indicate women, who are not included in PAMFLNP, it might not be as significant. But there are two other reasons why these don’t seem to match very well.

First, disability represents the main route by which people could plausibly be leaving the labor force to become PAMLFNPers. Here’s the graph:



There is a general trend of increasing disability since 1985, but a paper by Autor and Duggan suggests this is almost entirely due to a reform of the disability system around 1984 which made it easier to get benefits (also, the size of the workforce increased due to more female participation, meaning the pool of potential disabled people increased too).

Second, it is hard to explain *why now*. Whether we’re worrying about the past twenty years of automation (collapse in raw numbers of manufacturing jobs) or the past sixty (decline in PAMLFPR, decline in percent manufacturing) – neither corresponds to the several hundred years in which things have been getting more automated. Why didn’t previous eras of improving automation result in job loss?

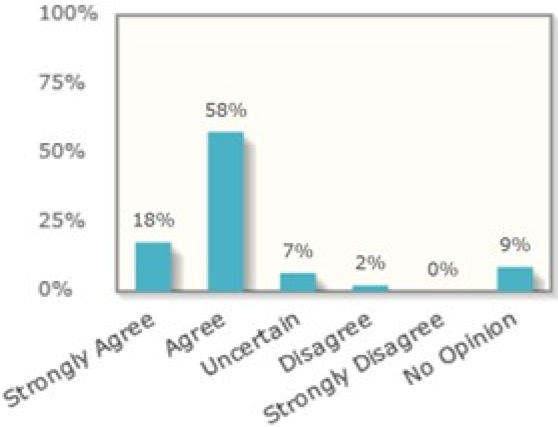
Let’s go back to all the economists in the world and see what they think:

Tuesday, February 25th, 2014 1:55 pm

Robots

Question A: Advancing automation has not historically reduced employment in the United States.

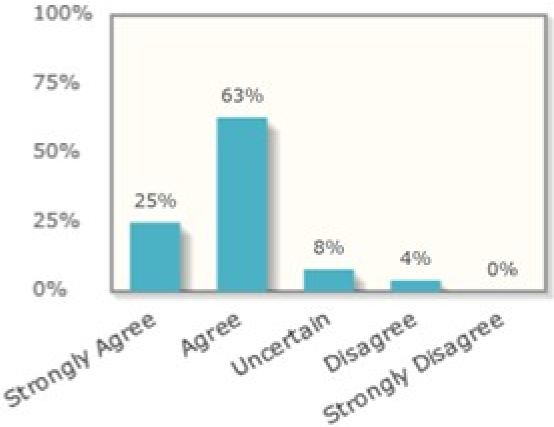
Responses



© 2018. Initiative on Global Markets.

Source: IGM Economic Experts Panel
www.igmchicago.org/igm-economic-experts-panel

Responses weighted by each expert's confidence

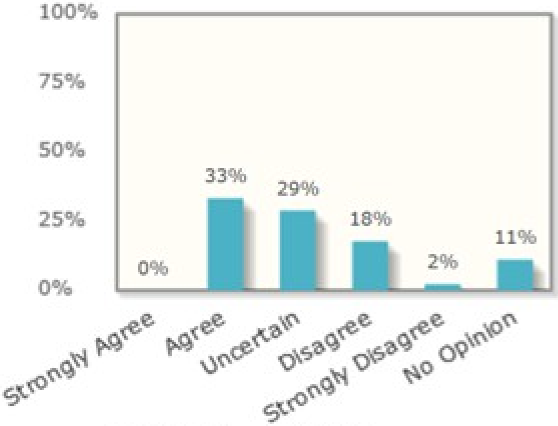


© 2018. Initiative on Global Markets.

Source: IGM Economic Experts Panel
www.igmchicago.org/igm-economic-experts-panel

Question B: Information technology and automation are a central reason why median wages have been stagnant in the US over the past decade, despite rising productivity.

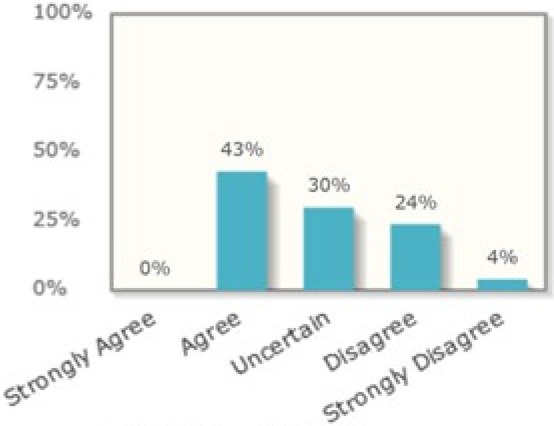
Responses



© 2018. Initiative on Global Markets.

Source: IGM Economic Experts Panel
www.igmchicago.org/igm-economic-experts-panel

Responses weighted by each expert's confidence



© 2018. Initiative on Global Markets.

Source: IGM Economic Experts Panel
www.igmchicago.org/igm-economic-experts-panel

Economists very strongly believe automation has not historically reduced employment. But they do believe automation is making wages stagnate right now. I don’t really understand what’s going on here. Are they saying that automation can depress wages, but not reduce employment? Surely (given the existence of a minimum wage) that doesn’t make sense. Or are they saying that automation never caused any problems before, but it is causing problems now?

The site offers some of the economists the chance to explain what they meant, and a lot of them seem to be saying that automation has *temporarily* caused problems in the past, but they always resolved with time as new industries open up. Maybe we’re just in a temporary bad period? Likewise, one economist who agrees that automation caused wage stagnation says that “it may have a short-run impact but there is no reason to believe that it is permanent.”

All of this is a mess. But the impression I get from this mess is that there is little sign of technological unemployment happening today in a historically unique way, or even picking up pace. I get this from a few sources.

First, the official unemployment rate looks great, so if we were going to make this argument we would have to do it off of PAMLFPR.

Second, Winship’s optimistic take on PAMLFPR is hard to easily refute. PAMLFNPers pretty clearly say they’re not looking for jobs, and they’re just perfectly innocuous students, retirees, etc. We have trouble believing them, especially based on their demographics. But it’s very hard to look at the increase and see a place where unemployment issues could have slipped in.

Third, PAMLFPR has been getting worse gradually since about 1960, with no sign of any recent worsening. It is hard to explain why technological unemployment would have started around that time – at least if we limit our explanations to the nature of technology alone. And it doesn't seem to match the more sudden decline in manufacturing around 2000.

Fourth, most economists appear to remain doubtful of the possibility of long-term technological unemployment.

I realize this goes against common sense. Maybe I'm missing something and totally wrong here. But if I am forced to interpret the data as I see it, I just don't see the signs of technological unemployment. It's just not there.

And in my defense, this also seems to be the opinion of David Autor, the main economic expert on this subject. In an interview with *The Economist*, he said that there was "'zero evidence' that AI is having a new and significantly different impact on employment".

III.

This doesn't mean everything is great. As the IGM panel shows, even if robots aren't putting people out of work, they may be causing wages to stagnate. The people getting kicked out of manufacturing jobs may have other jobs available to them (and so not end up affecting the PAMLFPR numbers), but those jobs may not be as good or pay as well. This isn't "technological unemployment". But it might be technological underemployment.

Most people expect that technological unemployment will hit the least skilled first, but that doesn't seem to be entirely true. This chart and some of the following analysis are going to be from the Heritage Association – another potentially biased think-tank, but hopefully without much reason to obfuscate these issues.

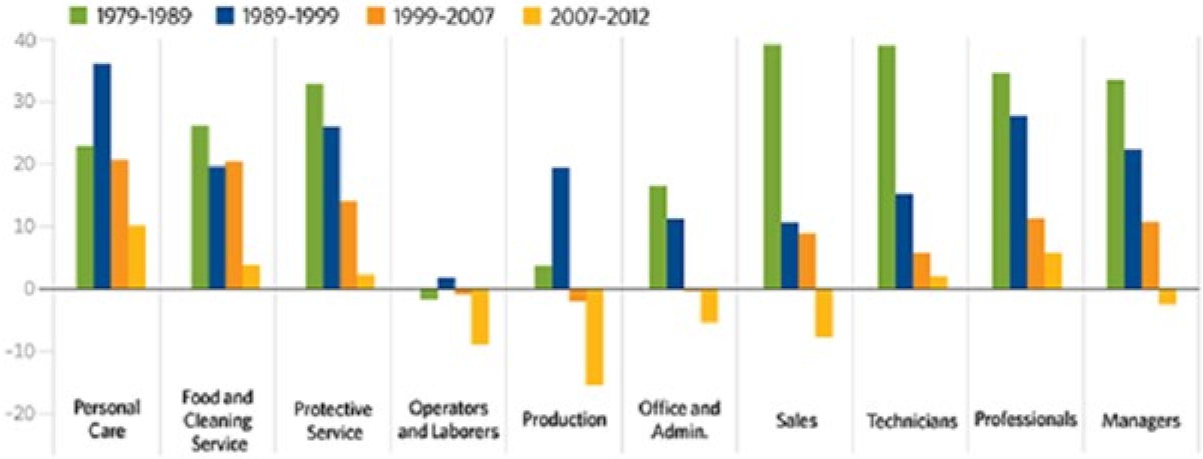


CHART 5

Moderately Skilled Occupations See Smallest Job Growth

PERCENTAGE CHANGES IN FULL-TIME, FULL-YEAR WORKERS, LOG X100

Notes: Figures are calculated using 1980, 1990, and 2000 Census IPUMS files; American Community Survey combined file 2006–2008; and American Community Survey 2012. This sample includes the working-age (16–64) civilian non-institutionalized population with 48 or more annual weeks worked and 35 or more usual weekly hours. Weekly wages are calculated as annual earnings divided by weeks worked.

Source: David H. Autor, "Polanyi's Paradox and the Shape of Employment Growth," paper presented at the Federal Reserve Bank of Kansas City symposium "Re-Evaluating Labor Market Dynamics," August 21–23, 2014, Jackson Hole, WY, p. 13, Figure 2, http://economics.mit.edu/files/9835 (accessed April 20, 2015).

BG 3016 heritage.org

The best-paying jobs – managers, professionals, and the like – are doing fine. The lowest-paying jobs, like personal care and food, are also doing fine. It’s the middle-paying jobs that are in trouble. Some of these are manufacturing, but there are also office and administrative positions in the same categories.

This is potentially consistent with a story where the jobs that have been easiest to automate are middle-class-ish. Some jobs require extremely basic human talents that machines can’t yet match – like a delivery person’s ability to climb stairs. Others require extremely arcane human talents likewise beyond machine abilities – like a scientist discovering new theories of physics. The stuff in between – proofreading, translating, records-keeping, metalworking, truck driving, welding – is more in danger. As these get automated away, workers – in accord with the theory – migrate to the unautomatable jobs. Since they might not have the skills or training to do the unautomatable upper class jobs, they end up in the unautomatable lower-class ones. There’s nothing in economic orthodoxy that says this can’t happen.

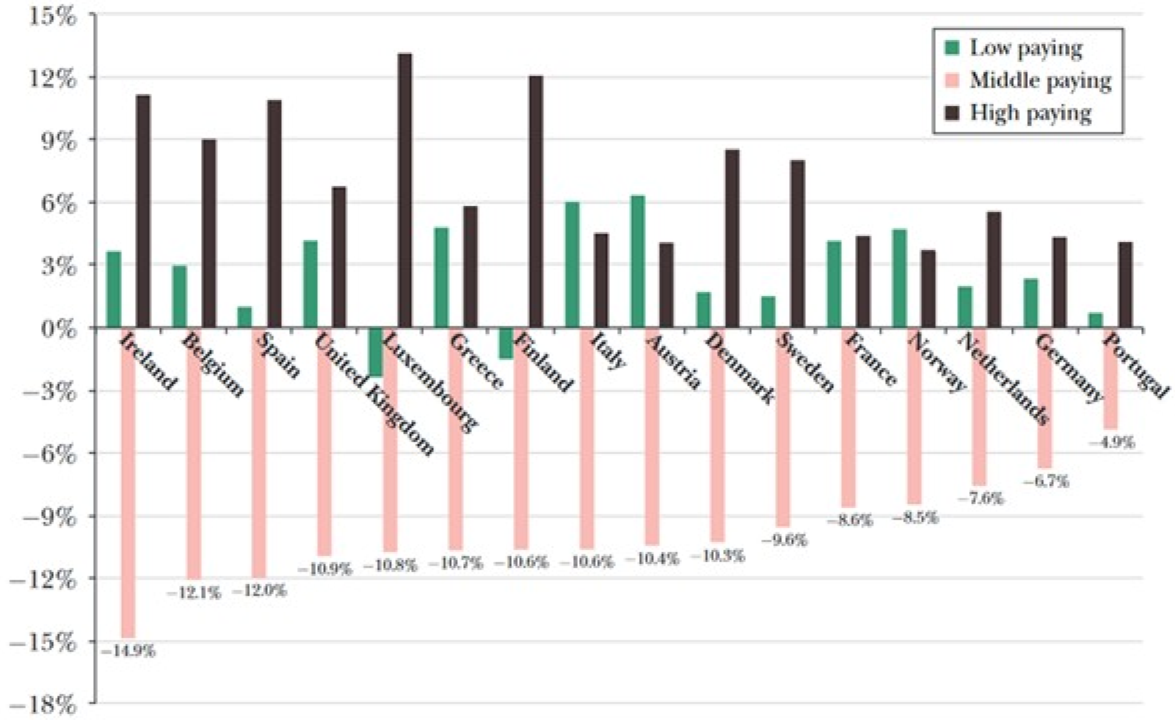
David Autor and his giant block of citations agree:

> Because jobs that are intensive in either abstract or manual tasks are generally found at opposite ends of the occupational skill spectrum—in professional, managerial, and technical occupations on the one hand, and in service and laborer occupations on the other—this reasoning implies that computerization of “routine” job tasks may lead to the simultaneous growth of high-education, high-wage jobs at one end and low-education, low-wage jobs at the other end, both at the expense of middle-wage, middle education jobs—a phenomenon that Goos and Manning (2003) called “job polarization.” A large body of US and international evidence confirms the presence of employment polarization at the level of industries, localities, and national labor markets (Autor, Katz, and Kearney 2006, 2008; Goos and Manning 2007; Autor and Dorn 2013; Michaels, Natraj, and Van Reenen 2014; Goos, Manning, and Salomons 2014; Graetz and Michaels 2015; Autor, Dorn, and Hanson 2015)

The fall of middle-skill-level jobs has led to a corresponding fall in middle-income jobs:

Figure 3

Change in Occupational Employment Shares in Low, Middle, and High-Wage Occupations in 16 EU Countries, 1993–2010



Source: Goos, Manning, and Salomons (2014, table 2).

Notes: High-paying occupations are corporate managers; physical, mathematical, and engineering professionals; life science and health professionals; other professionals; managers of small enterprises; physical, mathematical, and engineering associate professionals; other associate professionals; life science and health associate professionals. Middle-paying occupations are stationary plant and related operators; metal, machinery, and related trade work; drivers and mobile plant operators; office clerks; precision, handicraft, craft printing, and related trade workers; extraction and building trades workers; customer service clerks; machine operators and assemblers; and other craft and related trade workers. Low-paying occupations are laborers in mining, construction, manufacturing, and transport; personal and protective service workers; models, salespersons, and demonstrators; and sales and service elementary occupations.

Note that, contrary to an extremely pessimistic picture, this would suggest that most people who leave middle-paying jobs go to higher-paying jobs

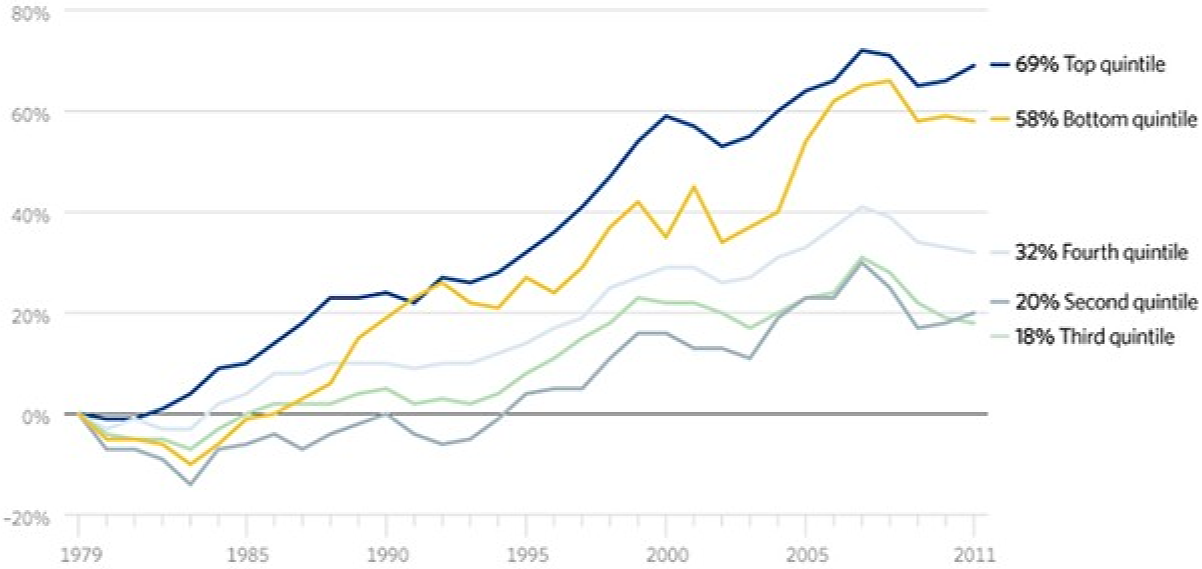
And with a corresponding decline in the fortunes of the middle class:

CHART 3

Highest and Lowest Incomes See Fastest Increases

For non-elderly households without children, the fastest-growing income brackets have been those at the top and the bottom.

PERCENTAGE GROWTH IN INFLATION-ADJUSTED LABOR MARKET COMPENSATION, NON-ELDERLY HOUSEHOLDS WITHOUT CHILDREN



Source: Heritage Foundation calculations using data from Congressional Budget Office, "The Distribution of Household Income and Federal Taxes, 2011," Supplemental Data, November 12, 2014, Table 14, https://www.cbo.gov/sites/default/files/cbofiles/attachments/49440-Distribution-of-Income-Taxes_Supplemental_0.xlsx (accessed May 12, 2014).

BG 3016 heritage.org

Note that this does not really back up the optimistic picture from above.

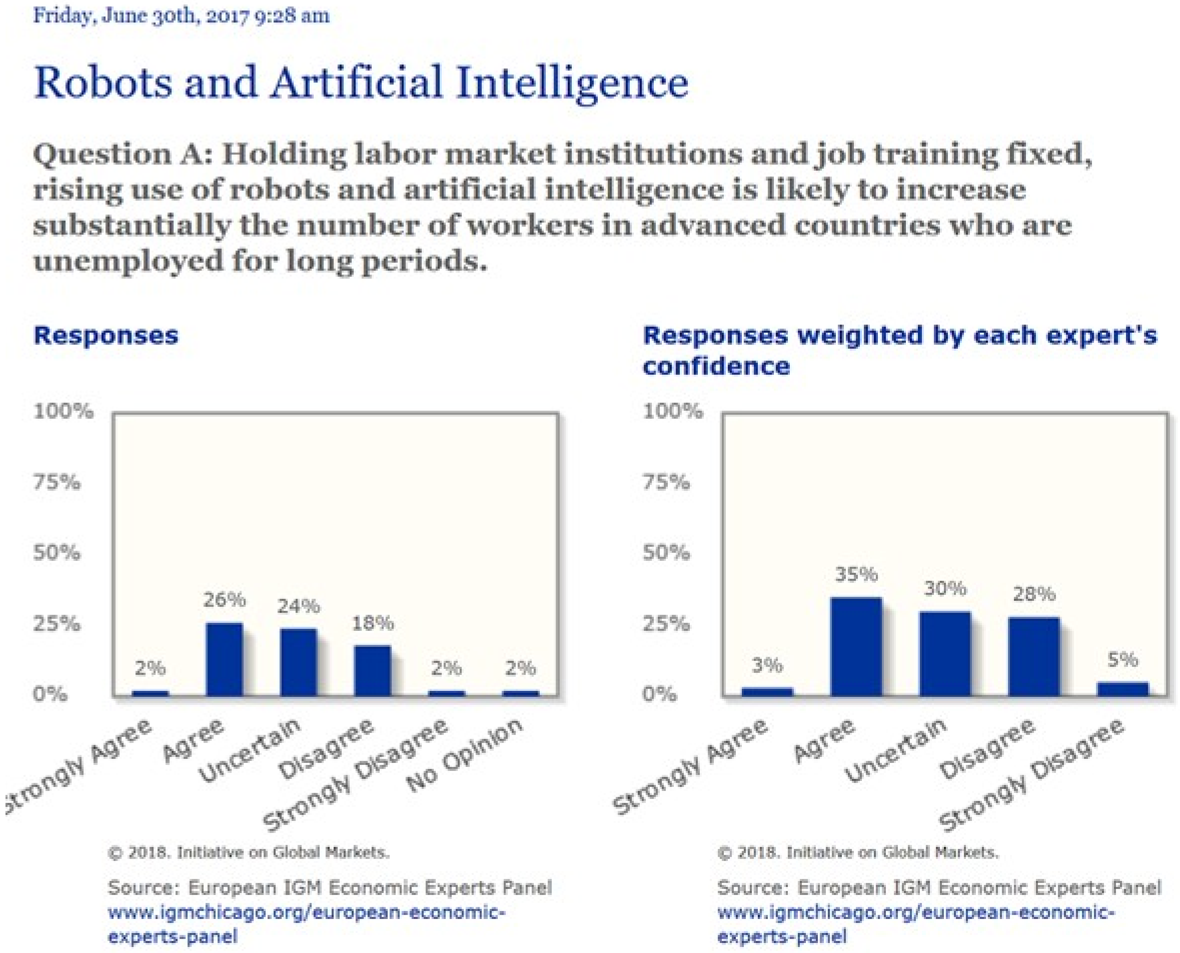
Why is this happening now when technological progress has been going on forever? This gets into the whole decline-of-the-middle-class argument, a giant political morass featuring de-unionization, regulation, automation, globalization, the 1%, and pretty much everything else. Is there also a role for today's robots just plain being *better than* yesterday's Rolodexes and whatever else the forefront of technology was? Or our education system being less able to cope with them? I'm not sure.

As far as I know, there is no economic theory stating that the number (or percent) of high-skilled jobs must always stay the same. I’m also not sure how to include fixed cognitive skills (eg some people are smarter than others) in this question. An optimist might argue that things will get better as today’s obsoletely-trained workforce retires and tomorrow’s trained-for-the-appropriate-jobs workforce graduates. But maybe this is better viewed as a race between two competing forces; generational churn producing students with the right set of skills, and technology making new skills obsolete. I don’t know why this should have increased recently, but it seems like – at least for the middle class – this is a race they are now losing.

IV.

Predicting the future is naturally harder than observing the present, since we have data about the present and not the future. But the data about the present is contradictory and incomprehensible and just makes things more confusing, so we might as well try going with the future and seeing how we do there.

We’ll start with those surveys of economists again, since they seem like the people most likely to know. Here’s a panel of top European economists on the future of technological unemployment:



Friday, June 30th, 2017 9:28 am

Robots and Artificial Intelligence

Question A: Holding labor market institutions and job training fixed, rising use of robots and artificial intelligence is likely to increase substantially the number of workers in advanced countries who are unemployed for long periods.

Responses

100%
75%
50%
25%
0%
2%
26%
24%
18%
2%
2%
Strongly Agree
Agree
Uncertain
Disagree
Strongly Disagree
No Opinion

© 2018. Initiative on Global Markets.

Source: European IGM Economic Experts Panel
www.igmchicago.org/european-economic-experts-panel

Responses weighted by each expert's confidence

100%
75%
50%
25%
0%
3%
35%
30%
28%
5%
Strongly Agree
Agree
Uncertain
Disagree
Strongly Disagree

© 2018. Initiative on Global Markets.

Source: European IGM Economic Experts Panel
www.igmchicago.org/european-economic-experts-panel

In the same survey, 93% of economists with an opinion on the issue agreed that the economic benefits of robots will be so great that they could be used to compensate the workers who were negatively effected. But in a survey I conducted in my imagination, 100% of people who have not been living in a cave the past two hundred years agreed that this will never happen in real life.

So economists really have no idea about any of this. What are we paying them for, anyway?

Frey and Osbourne analyze what jobs are most susceptible to automation. They claim that "47% of total US employment is at risk". This sounds suspiciously precise and it's unclear their numbers have any relationship to reality.

They do find “evidence that wages and educational attainment exhibit a strong negative relationship with an occupation’s probability of computerisation”.

Overall none of this seems to be making things much clearer.

I’ve cited David Autor something like five times already. He is the recognized expert in this area – I blame nominative determinism – and has written widely. His own opinion is:

> I expect that a significant stratum of middle-skill jobs combining specific vocational skills with foundational middle-skills levels of literacy, numeracy, adaptability, problem solving, and common sense will persist in coming decades. My conjecture is that many of the tasks currently bundled into these jobs cannot readily be unbundled—with machines performing the middle-skill tasks and workers performing only a low-skill residual—without a substantial drop in quality. This argument suggests that many of the middle-skill jobs that persist in the future will combine routine technical tasks with the set of nonroutine tasks in which workers hold comparative advantage: interpersonal interaction, flexibility, adaptability, and problem solving. In general, these same demands for interaction frequently privilege face-to-face interactions over remote performance, meaning that these same middle-skill occupations may have relatively low susceptibility to offshoring. Lawrence Katz memorably titles workers who virtuously combine technical and interpersonal tasks as “the new artisans” (see Friedman 2010), and Holzer (2015) documents that “new middle skill jobs” are in fact growing rapidly, even as traditional production and clerical occupations contract.

> This prediction has one obvious catch: the ability of the US education and job training system (both public and private) to produce the kinds of workers who will thrive in these middle-skill jobs of the future can be called into question. In this and other ways, the issue is not that middle-class workers are doomed by automation and technology, but instead that human capital investment must be at the heart of any long-term strategy for producing skills that are complemented by rather than substituted for by technological change. In 1900, the typical young, native-born American had only a common school education, about the equivalent of sixth to eighth grades. By the late 19th century, many Americans recognized that this level of schooling was inadequate: farm employment was declining, industry was rising, and their children would need additional skills to earn a living. The United States responded to this challenge over the first four decades of the 20th century by becoming the first nation in the world to deliver universal high school education to its citizens (Goldin and Katz 2008). Tellingly, the high school movement was led by the farm states. Societal adjustments to earlier waves of technological advancement were neither rapid, automatic, nor cheap. But they did pay off handsomely.

Do we really have evidence that compulsory schooling was a result of increasing automation? If so, could we tell a story where the gradually increasing length of schooling – from minimal, to primary school, to high school, to “you better get a college degree or you’ll regret it later”, to increasing pressure to go to graduate school – is a reaction to automation and the threat of technological unemployment? Could this be the reason why automation finally seems to be causing problems – a financial and cultural inability to extend schooling any further than it’s already gone?

Or – inspired by Bryan Caplan's The Case Against Education – could we tell the opposite story? One where increasing credentialism makes it harder for people whose jobs have been automated to switch careers the way they did before? Higher-paying jobs no longer just require skills, they require a college degree in a relevant field – which is very hard for a mature worker to get. Being an office manager and being a nurse are both middle-income jobs – but in the past, an office worker would have needed about six months of inexpensive training plus a lot of on-the-job apprenticeship to become a nurse, whereas now they would need a four-year Bachelors of Science in Nursing from a university whose price tag has dectupled for no reason over the last half-century. A now-unemployed office manager might have been able to afford the first even if middle-class; the second might be well beyond her reach. Unable to shift into another middle-class job, she is forced to take a lower-class job as a fast food worker or something.

I am not entirely sure how differences in cognitive ability fit in here. My guess is to a first approximation they don't – if standard economic theory is correct, it should be possible to create middle-paying jobs that use the full potential of people with any amount of cognitive ability, taking advantage of various human cognitive skills that are difficult to automate. Although some naive takes like "everyone should just become a programmer" fail to understand this, I don't think the entire argument is based on misunderstanding of this point, or that it forms a particularly strong counterargument.

Anyway, if Autor's prediction is "we will be able to weather this danger as long as our education system is able to rise to meet the challenge", I'm just going to round this off to "we're super doomed". But I think his methodology – of noticing that we always met the challenge before, and trying to figure out what might be different this time – is a promising one.

Finally: we’ve been talking about economists a lot here, but what about the roboticists? Aren’t they relevant too? Grace et al survey top AI researchers on when AI might be able to replace humans in various things; these researchers don’t necessarily know anything about economics, but they at least know something about progress in robotics. On average, they believe AI will reach human performance at truck driving, retail selling, translation, transcription, and bipedal running all before 2030. Whether those robots will be affordable, widely adopted, or able to deal with the long tail of real-world situations like blackouts or vandals or bad weather – is a different story. The point is, roboticists are pretty sure they’ll have their contribution to the economic takeover ready pretty soon.

(they do say robots won’t be writing bestselling novels until 2050, so JK Rowling’s job is safe for now)

They also say that robots will be able to do *all* human tasks, including novel-writing, science, and further AI research – sometime between 2050 and 2150. At that point, obviously, all bets are off, and we have a *lot* more than unemployment to worry about.

V.

Here are some tentative conclusions:

1. Technological unemployment is not happening right now, at least not more so than previous eras. The official statistics are confusing, but they show no signs of increases in this phenomenon. (70% confidence)

2. On the other hand, there are signs of technological underemployment – robots taking middle-skill jobs and then pushing people into other jobs. Although some people will be “pushed” into higher-skill jobs, many will be pushed into lower-skill jobs. This seems to be what happened to the manufacturing industry recently. (70% confidence)

3. This sort of thing has been happening for centuries and in theory everyone should eventually adjust, but there are some signs that they aren’t. This may have as much to do with changes to the educational, political, and economic system as with the nature of robots per se. (60% confidence)

4. Economists are genuinely divided on how this is going to end up, and whether this will just be a temporary blip while people develop new skills, or the new normal. (~100% confidence)

5. Technology seems poised to disrupt lots of new industries very soon, and could replace humans entirely sometime within the next hundred years. (???)

This is a very depressing conclusion. If technology didn’t cause problems, that would be great. If technology made lots of people unemployed, that would be hard to miss, and the government might eventually be willing to subsidize something like a universal basic income. But we won’t get that. We’ll just get people being pushed into worse and worse jobs, in a way that does not inspire widespread sympathy or collective action. The prospect of educational, social, or political intervention remains murky.

Current Affairs' "Some Puzzles For Libertarians", Treated As Writing Prompts For Short Stories

[Taken from here.]

I.

> Deep in the forest, thousands of miles from civilization, there is an isolated village. It has not seen contact with any other humans for a long time. It is, however, a pleasant and flourishing community, which strongly values freedom and entrepreneurship. There is, however, one tiny quirk. In this village, there is a ritual. Every year, a boy who reaches 18 is cannibalized. It brings the rains, or something. But despite its taste for cannibalism, this village wishes to live in accordance with libertarian principles. Thus, they will only cannibalize the boy if he consents. In order to encourage this to happen, they will put tremendous social pressure on the boy. All through his youth, they will tell him they believe the future of the village depends on his consenting. His parents tell him that he would bring great shame on the household if he refused, which is true. The choice nevertheless rests with the boy, and whatever he chooses will be respected. The parents and villagers attempt to persuade him, but never lie to him, and make clear that they would never force his choice. However: if the boy refuses to be cannibalized, the village has a backup plan. The boy will be blacklisted. No shopkeeper will sell him food, no hotel will give him a room, no hospital will treat him, no employer will hire him. After all, under libertarian principles, nobody can be told how to use their property. The boy's parents, ashamed of him, will turn him out of the house with no money. He may leave the village, but

> it is certain death, for thousands of miles of desolate wolf-infested wilderness stand between him and other humans and he has no food. (The wilderness is also privately-owned, and he cannot pay the admission fee.) He is shunned and despised, left to wander the streets in a futile search for shelter and sustenance. However, no force is exercised against him. He is never touched or arrested. He is treated as nonexistent, as the villagers await his demise. So the boy starves to death. The villagers then cannibalize his emaciated corpse, reasoning that they cannot be compelled to give him a dignified burial (plus he died on private property, collapsing in a flowerbed).
>
> Is eating the boy’s corpse after he dies the only potential violation of libertarian principles in the village? Is every single other aspect of this completely permissible?

The setting sun shone its last few rays on Independence Hall. The delegates were tired, but a thrill of excitement filled the air. The wrangling and deal-making was almost done; nothing remained but a few technicalities.

As the last sunbeam went below the horizon, something stirred in the middle of the chamber. It grew into a wind, then a whirlwind, and then standing among the assembled Founding Fathers was a strange man dressed all in silver, wearing a pair of gold goggles.

“You’ve got to stop!” he shouted. “It’s all lies!”

George Washington had stayed calm through cavalry charges, artillery fire, and the assembled might of the United Kingdom. He flinched only a little here. “Who are you, sir? Where have you come from?

The traveler barely heard. “Listen! You think ‘democracy’ can solve all your problems. But – imagine there’s a village full of cannibals. They have a tradition of picking a child, and killing and eating him when he’s eighteen years old. Well, even if that village is a democracy, then 51% of the population can just vote to kill and eat him! Do you want a *child* to be *killed and eaten*? Because that’s what your ‘democracy’ inevitably leads to! Checkmate, liberals!”

The delegates were only less dazed by the man’s speech than by his sudden appearance. Finally, General Washington asked whether anyone wanted the floor. After a scramble of shouts and raised hands, the chair recognized James Madison, delegate from Virginia.

“Thank you,” said James Madison. “Our Traveler may not know this, but I am preparing a Bill of Rights to be added on to the end of this Constitution, severely limiting the powers which the government may exercise. I’m planning one on cruel and unusual punishment, which sounds like it ought to cover killing and eating someone, and there will also be various restrictions on seizure of persons. The Traveler is already wrong that we operate entirely on the basis of 51% of the populace – rather, there will be representatives, senators, and Supreme Court Justices. But even if all these people should agree to kill and eat someone, I am confident that the natural rights included in my bill will restrict such practices.”

“AHA!” said the Traveler. “You’ve fallen for my trap! Because even if the government is banned from assisting in killing and eating someone, it could still happen. Imagine a system where, if the victim refused to be killed and eaten, then everyone in the village refused to house him, or feed him, so that he starved to death. Then he’d be dead anyway, and your precious Bill of Rights wouldn’t be able to do anything about it!”

“Couldn’t the victim just move to a different village?” interjected John Jay.

“The village is in the middle of a giant forest stretching five thousands miles, teeming with dire wolves,” snarled the Traveler, annoyed at such a stupid objection.

“Couldn’t the victim just build his own house, and farm his own food?” asked John Adams.

“The dire wolves would tear up the house, and trample all over the farm!” said the traveler. “You’re splitting hairs here! Why won’t anyone answer my question in the spirit it was intended?!”

There were more shouts and another frenzy for attention. General Washington banged his gavel. “The chair recognizes Alexander Hamilton.”

“Yo,” said Hamilton. “The institutions of our Constitution, give a clear solution to this persecution. The Revolution…”

“The chair unrecognizes Representative Hamilton, and offers the floor to anyone who does not speak in rap.”

“Thank you,” said Benjamin Franklin. “My good Mysterious Traveler, perhaps you labor under the misapprehension that political philosophies are also moral philosophies, and so fail irredeemably if they ever recommend an immoral course of action. I do not believe democracy is always right. But I believe it is a wise way to govern. All that systems of government can do is take nations – with all of their conflicts, ideas, prejudices, and values – as input, and then magnify some impulses and suppress others. Start with a country where every single person is entirely set on doing as much evil as possible, and democracy alone cannot save it; they will simply vote to do as much evil as possible. But start with a country in which there are many different classes, agendas, and visions, and I believe that a democratic system is more likely to magnify those impulses that help the common people, and suppress those impulses that lead to tyranny, than any other system yet devised."

"So you're saying," said the traveler, "that you *don't care* that your precious democracy and even your so-called Bill of Rights aren't good enough to save the life of a child in – "

"You listen here," said Benjamin Franklin. "I care plenty. In a village that didn't have any form of government, as soon as anybody big and strong enough wants to eat you, they can form a mob and drag you away. In a village that operates as a direct democracy, it's harder. You need 51% of the population to want to eat you before you end up as dinner. And in a village that subscribes to Mr. Madison's notion of natural rights, it's harder still. You have to have every single person in the village agree not to feed the victim, without a single kindly old lady leaving food out on her porch at night when it's too dark out for anyone to see. We have gone from tyranny – a system where, as long as even one person wishes you ill, you perish – all the way to a system where as long as there is a single person who does not wish you ill, you endure. That seems to me to be the best we can do in this world."

Suddenly the Traveler seemed to warp, or crackle, like a signal from far away was being disrupted. "I must go!" he said. "I'm being recalled to my home time!" he shouted. "Where I will tell people that they should form a government based on socialism, and that it will be great, and nothing can possibly go wrong!"

“Stop!” said Jay. “You must tell us about this ‘socialism’ of yours!”

“Say, before you’re lost to me, at very high velocity, what is this new philosophy, that can prevent atrocity?” begged Hamilton.

But it was Washington, ever the man of action, who jumped up from the chair and grabbed the Traveler by his silver arm, holding him against the winds of Time. “This ‘socialism’ of yours – ” asked Washington. “It can ensure that – even in a barbaric society where literally one hundred percent of the people are wholeheartedly dedicated to do so – nobody ever eats their fellow citizens?”

“Yes!” said the traveler. “Why, in *true* socialist countries, nobody ever eats anything at all!”

Then he broke free of Washington’s grasp and disappeared forever, just as the first rays of the moon cast their white light on Philadelphia.

II.

> Is there a meaningful difference between coercion by the state and coercion by private entities?

The door caved in loudly and suddenly, like a thunderclap. My Golden Retriever ran up, tail wagging, to investigate. Another loud noise, and my dog lay dead, bleeding on the floor in front of me.

“PUT YOUR HANDS UP!” said a man in black body armor and a black helmet. There were five of them, all with guns. My five-year-old son started to cry. “HANDS UP!” he shouted, “I’M NOT GOING TO TELL YOU AGAIN!”

I put my hands up. My son, who was screaming, got the presence of mind to put his hands up also, though not before one of the armored men had put a gun to his head.

“What’s wrong?!” I asked. “Why are you doing this to us?”

One of the men put a gun to my head. “Admit it! You’re growing marijuana here!”

“I’m not!” I insisted.

“Is this 2051 Willow Street?”

“No, this is 2052 Willow Street. 2051 Willow Street is on the other side!”

“Oh. Well, sorry.”

“Sorry? That’s all you have to say? You killed my dog! You terrorized my five-year-old son!”

“Yeah, sorry. We’re McDonalds employees, and corporate headquarters must have given us bad directions.”

“You’re…McDonalds employees? Why are McDonalds employees doing no-knock raids in body armor looking for marijuana?”

“I…I don’t know.”

“Then why are you even here? You just think it’s *okay* to randomly go around, kill people’s pets, terrorize their families, when you don’t know why you’re doing it? How can you justify such a thing?”

“I heard there were a bunch of people who were okay with private coercion, and only objected to coercion when it was applied by the State.”

“*What?* Where did you hear that?”

“I don’t know. Some socialist magazine, I think.”

III.

> Can you construct a theory of property rights that does not suffer from internal incoherence or depend on specious natural law assumptions?

Professor Kryzenski sat down in her desk and booted up her computer. It was another quiet morning here at the Harvard Philosophy Department. She had won her position as Department Chair by discovering a complete theory of morality grounded in first principles with no internal incoherence or any specious assumptions, able to determine everything from the optimal number of minutes to spend speaking to your mother each week to how close you could come to beggars before you were obligated to give them money.

She had just finished checking her emails – mostly invitations to speak at various conferences and events – when something started to stir in the center of her office. It turned into a wind, then a whirlwind, and finally, a strange-looking man, dressed in silver with gold goggles.

“Professor!” said the Traveler. “Professor Kryzenski! Terrible news!”

The Professor, whose mind had plumbed the depths of ontology and ascended the heights of metaphysics, was a hard woman to perturb. “Yes?” she asked the man. “What is it?”

“Suppose there’s an evildoer who punishes all evildoers who do not punish themselves. Does the evildoer punish himself, or not?”

Professor Kryzenski realized the implications right away. “My God. It’s a paradox! My complete theory of morality grounded in first principles with no internal incoherence or specious assumptions, able to determine everything from the right amount to tip your waiter to the exact words you need to speak before a sexual act for it to qualify as consensual – lies in ruins!”

“And that means…” began the Traveler.

“That’s right,” said Professor Kryzenski. She and the Traveler spoke in unison: *“Nothing is true and everything is permissible.”*

“Come,” she said. “I’ve prepared for this day.” She took a key out of a potted plant on the windowsill, then used it to open a locked cupboard in her desk. Inside were two hatchets. She handed one to the Traveler.

“Where are we going?” he asked.

“To the daycare down the road, to hack the limbs off the babies,” she said. “*Obviously*.” The Traveler nodded his approval, and off they went, eyes red with bloodlust.

IV.

> The Infinitely Rich Man is not infinitely rich. He is just very, very rich. Nobody knows quite how rich. One day, you happened to meet the Infinitely Rich Man in a bar. At first he was friendly, but soon you found yourselves in an argument about horses. You were for them, and he was against them. Or perhaps you were against them, and he was for them. You don’t actually remember how it went.

As you parted ways, you expected never to see the Infinitely Rich Man again.

Little do you know: the Infinitely Rich Man now despises you. His sole desire on earth is to see you unhappy. This should hardly trouble you, though. After all, you have a good job at a castanet factory. You own your own home, which has a picturesque lake view. You have a wife, whom you love and who loves you. You also have a prized possession, your 1972 Pontiac Lemans. You don’t have much spare cash, but this never bothers you because of your stable job. The Infinitely Rich Man is also a strict Libertarian. He believes it is illegitimate for anyone to initiate force against another. And because you are fortunate enough to live in a Libertarian world, you are free to enjoy those things you treasure most in the world without being bothered by the state or the Infinitely Rich Man. The Infinitely Rich Man is not discouraged, however. He still believes he can ruin you. He will be a Count of Monte Cristo, but an extremely law-abiding one.

The first thing the Infinitely Rich Man does is buy the castanet factory where you work. He immediately fires you. He also makes sure that if any other employers inquire about you, the castanet factory will refuse to serve as a reference. Not that this matters, for he intends to bribe any other castanet company who hires you into firing you. (There are four castanet companies.) You therefore find yourself unemployed. Fortunately, you have a skill. You know how to make castanets! (Castanets are very popular.) So you scrape together what money you have, and you open a little drive-thru castanet stand out on Route 9. But the Infinitely Rich Man has a plan. He opens a stand next to yours. At his stand, castanets are free. He gives them away by the truckload. He sets the whole world clacking. You cannot compete. You are ruined.

At least you still have your wife, your friends, your lakeview home, your 1972 Pontiac Lemans. But the Infinitely Rich Man has a plan. First, he buys the lake. He fills it with concrete. No more lake view, and your property value diminishes by $100,000. Then, he buys every house around yours, flattens it, and turns it into a landfill. The smell doesn't reach your home, but it turns the neighborhood unsightly and desolate. Your house becomes worthless. The Infinitely Rich Man buys the heating company and refuses to provide gas to your home at any price. (You try to talk other gas companies into competing, but they refuse; laying a new main for a single home would be absurd, they say.) But you have a wife! And friends! And you get to drive a 1972 Pontiac Lemans! The Infinitely Rich Man offers a bribe. Any of your friends who refuse to speak with you ever again will receive a salary of one million dollars per year. At first, many decline to take the bribe. But sooner or later, most of them have one or another sticky financial situation, and they give in. Goodbye, vast majority of your friends! At least your wife loves you.

But one day, she becomes ill. She finds out that she will die, unless she goes on a treatment regimen for the rest of her life. The regimen costs $100,000 a month. The Infinitely Rich man pops up, and offers to pay. The one condition is that she divorce you, cut contact, and never speak with you again. As soon as she breaks the agreement, he will cease to pay for the treatment. You love your wife, but you do not want her to die. You both agree that it is better that she should accept. At least you can drive your 1972 Pontiac Lemans. Oh, but wait. The Infinitely Rich Man invests heavily in electric energy.

> Slowly, he makes gasoline-powered transit obsolete. He buys the oil companies, burns the gasoline, and converts every gas pump to a charging station. You can only drive your Lemans short distances, using some of the last gallons of available petrol, which you ordered from the internet. (That is, if the Infinitely Rich Man didn't outbid you!) They don't make the Pontiac Lemans anymore. Parts therefore exist only in small quantities. The Infinitely Rich Man buys up all existing Lemans parts. The moment it breaks, you are out of luck. As you sit alone, broke, and starving in the garage of your unheated home, caressing your disabled Lemans, thinking about your long-gone wife, your lake view, and your job, you are thankful that you live in a world of freedom, where nobody can encroach upon the liberty of another.
>
> Questions for Libertarians: Has the non-aggression principle been violated? Should the Infinitely Rich Man suffer any civil or criminal penalties for his actions?

"That," said Mr. Thaddeus Nett-Worth III, Esq., "is the most benightedly offensive statement I have ever heard."

"All I said," I said, "was that horses were basically elongated cows."

"They are a noble animal, an unparalleled paragon of mammalian perfection!"

"Right," I said. "Like cows are. Only more elongated."

"Dastard!" said Mr. Nett-Worth, pounding the table so hard his top hat and monocle almost fell off. "You've messed with the wrong captain of industry, believe you me. Let me tell you what I am going to do. You own a castanet factory? I am going to undercut you, undercut you *bad*. I will destroy your business. Then, I shall buy the lake by your home and fill it up with concrete. I will buy your heating company and refuse to sell you gas. I will bribe your friends never to speak to you again. I will wait until your wife develops a deadly disease, then offer to treat her if only she divorces you. I can do all of it, because we are in a perfectly libertarian society with no laws besides the non-aggression principle, where lawmakers have failed to pass commonsense legislation like ‘making it illegal to hurt someone by legal means’. And then – alone, friendless, shivering in the cold in your hopelessly ugly house – then you will *rue* the day you ever compared horses to – ” (he almost spits) ” – elongated cows.”

“But, I mean, think about it. Their faces are a little bit longer. Their bodies are a little bit longer. They’re pretty much just elongated cows. I’m sorry this is so hard for you.”

“What? You’re not backing down? You should be at my feet, begging me for forgiveness! Don’t you know all the things I can use my wealth to do to you in our perfectly libertarian society?”

“Yeah, well, about that. I know this guy named David Friedman, whose hobby is designing weird insurance systems for anarcho-capitalist utopias based on, like, the laws of medieval Iceland or something. Anyway, he sells ‘rich person gets a weird grudge against you’ insurance. I have loads of it. However much you try to bribe my friends not to talk to me, his company will pay more to bribe them to ignore you. However much you try to pay for my gas company, his company will pay more to keep my heat on. And if you try to offer my wife free health care to leave me, his company will offer her better health care to stay.”

“What? How did you even know to buy such an insurance?”

“Well, part of it was just a ‘why not?’ sort of thing. The odds of the situation ever happening are so astronomically low that the insurance was incredibly cheap – a few cents per year. Also, the mere existence of the insurance prevents rich people from starting bizarre revenge schemes against the people in it, so they can afford to assume they will rarely have to pay out. I guess the price was so low that it was a no-brainer.”

“But what about transaction costs? Why would you even think to look into such a product?”

“Well, this is going to sound weird, but – I was reading a history book a few months ago, and – you remember that time a time traveler appeared in the middle of the Constitutional Convention, making some kind of point about how democracy wouldn’t work in a village of evil cannibals? Don’t you think that was pretty weird?”

“I had thought it was just one of the many colorful, larger-than-life stories from the Revolution. Like how George Washington was unimpeachably honest, or Benjamin Franklin always had a witty saying ready, or how Alexander Hamilton always spoke in rap.”

“Yeah, I used to think that too. But then I was reading a political science book last month, and – well, isn’t it weird that we’re a perfectly libertarian society? All of the old political philosophers used to say that libertarianism was an ideal system that could only be approached, never reached, and that even the approach would take a dedicated and virtuous population to pull it off. And our population isn’t that virtuous – I mean, just the other day I heard on the news about an ethics professor who went on a violent rampage chopping the limbs off babies.”

“Oh yes,” said Mr. Nett-Worth. “I heard about that too. Terrible stuff, terrible!”

"But it really only clicked a few weeks ago, when these goons from McDonalds broke into my house on a no-knock drug raid and shot my dog, and then muttered something about how surely I couldn't object to *private* coercion. And it just got me thinking – *what if this whole world is just a thought experiment by a communist with a crappy understanding of political philosophy trying to weak-man libertarianism?* And then I thought – frick, I better get some really good vengeful-rich-person insurance, like, right away."

"I am so confused right now."

"Well, most sources define libertarianism as a political philosophy emphasizing individual autonomy and skeptical of government intervention. Libertarians come to their position for a wide variety of reasons, including belief that bottom-up local knowledge makes better decisions than top-down absolutism, or that government intervention naturally favors the powerful, or that if you actually ask poor people what they want, it's usually more money, not people taking choices away from them and treating them like children. A fraction of libertarians – I think a small fraction, though I can't prove it – are also believers in a deontological theory of natural rights which emphasizes non-aggression as the fundamental moral principle. If you're doing shoddy journalism aimed at inflaming people rather than enlightening them, you might try to tar all libertarians by identifying them with this subset."

"How does that explain all the weird things going on?"

"Take the cannibal village. If for some reason you believe the Non-Aggression principle perfectly defines the moral outcome in every situation, it must be pretty devastating to learn it can lead to cannibalism. But if you're a normal libertarian who just thinks of libertarianism as a political position, then it's no worse than a supporter of representative democracy learning that representative democracy could sometimes lead to cannibalism – which of course it can. In fact, you should be happy to point out that a libertarian village is much more resistant to cannibalism than a direct democratic or monarchical one."

"What about the ethics professor's rampage?"

"If for some reason you insisted property rights were based on perfect axiomatized natural law, it might be pretty devastating to learn that moral philosophy can't get that kind of precision. But you're a normal libertarian who just thinks of libertarianism as a political position, then learning that you can't perfectly axiomatize property rights is no more devastating than learning that you can't perfectly axiomatize caring about the poor, or thinking torture is bad, or not hacking off babies' limbs. You're still allowed to care about these things for the usual reasons even if you can't construct a perfect moral theory around them."

"And what about the goons from McDonald's?"

"If for some reason you believe that only the government can do anything bad, and private companies…look, I don't even want to speculate on who exactly they're trying to straw man here."

"Not straw man. Weak man. There are some real libertarians who believe only the non-aggression principle matters."

"Maybe, some of them. I think in general they believe there are moral values other than non-aggression – after all, many of them are Christian, and believe in all sorts of moral values – but they're skeptical of the government enforcing them. Remember, it's not always correct to insta-convert ethics into law. I think in general they believe both that it's important for a society to be virtuous, and that the government compelling people to exhibit more virtue than they possess can only go terribly wrong.”

“But there has to be *some* subset who don’t believe in virtue at all, and think the Non-Aggression Principle is literally all there is! And all these weird thought experiments show *they’re* stupid, right?”

“I disagree with them but I’m hesitant to declare them stupid just based on a few experiments. I mean, I like to say ‘I’m against torture’, and I like to say this is a strong moral principle of mine and not just a maxim of convenience. But with enough effort, you could create a ridiculous ticking-time-bomb thought experiment in which being against torture led obviously and inexorably to horrible results. Have you proven that people who say they’re against torture are stupid? Or would you be willing to cut them some slack in this situation? And are you willing to cut the same slack to this tiny subset of fundamentalist Non-Aggression Principle libertarians? Thought experiments are a useful tool, but sometimes the best lesson to take from them is ‘things are complicated but principles still matter’.”

“But surely, somewhere, there are incredibly stupid libertarians who think morality consists of the Non-Aggression Principle and nothing else, don’t believe in any other kind of virtue, and aren’t just holding it as a sacred but non-final principle the way you hold not torturing people?”

“Okay. Maybe there are. But how does it help to focus on this tiny pathological subset of libertarians and desperately try to convince the world that every libertarian is like this? There are some pretty pathological socialists too – should we demand everyone accept them as the only possible representatives of socialism? Should political discussion just be relentless weak-manning of the other side, with whoever is more simplistic winning the victory?”

Before Mr. Nett-Worth could respond, the bar we were in started to shake. “What’s that?” he asked me. “What’s going on?”

“A disturbance in the Farce,” I said. “This world was created to provide stupid weak-man arguments against dumbed-down versions of libertarianism. I guess what I just said – it threatened the fabric of reality itself. Hold on to your seat. This could get pretty bad.”

The bartender suddenly stood up. “All blue-eyed people need to leave the bar now!” he said. “As a proud bigot, I refuse to serve blue-eyed people. I don’t care how much profit it costs me! And also, this is the only bar in this city – nay, in a five thousand mile radius! Now no blue-eyed person will be able to go to a bar ever again!”

Scarce had he finished speaking when a very-finely dressed woman stood up. “I am a billionaire,” she shouted. “And I will give poor people money to humiliate themselves. Anyone who goes to the farm and rolls around in pig excrement for an hour, I will give ten million dollars! And many of you have terminal diseases that require expensive treatments, so you’ll die if you refuse! Mwa ha ha! Roll in pig shit! ROLL, YOU PEASANTS!”

But her jubilation was interrupted by another man, in the other corner of the bar. “I am a factory owner, and I am off to go sexually harass all my employees. There are no laws against it, so nobody can stop me. And I own the only factory in the world, so my employees can’t leave. And dire wolves eat anyone who tries to start new factories. So there!”

“You fools,” said a wild-haired man near the window. “I will cleanse this city of scum like you. Since there were no laws against making atomic bombs, I have built a nuke in my basement. Soon I will set it off in a great purification. And there’s nothing you can do about it until it’s too late, because there’s *no law against owning nukes*. Nobody can stop me! NOBODY!”

I took a deep breath. “I can stop you,” I said. “I can stop all of you.”

Every face turned to look at me.

“This world runs on dumb weak-man objections to libertarianism. The only way to fight them is with even dumber weak-man objections to libertarianism. So that’s what I’ll do. Nobody, blue-eyed or not, is going to leave this bar.”

The bartender scowled.

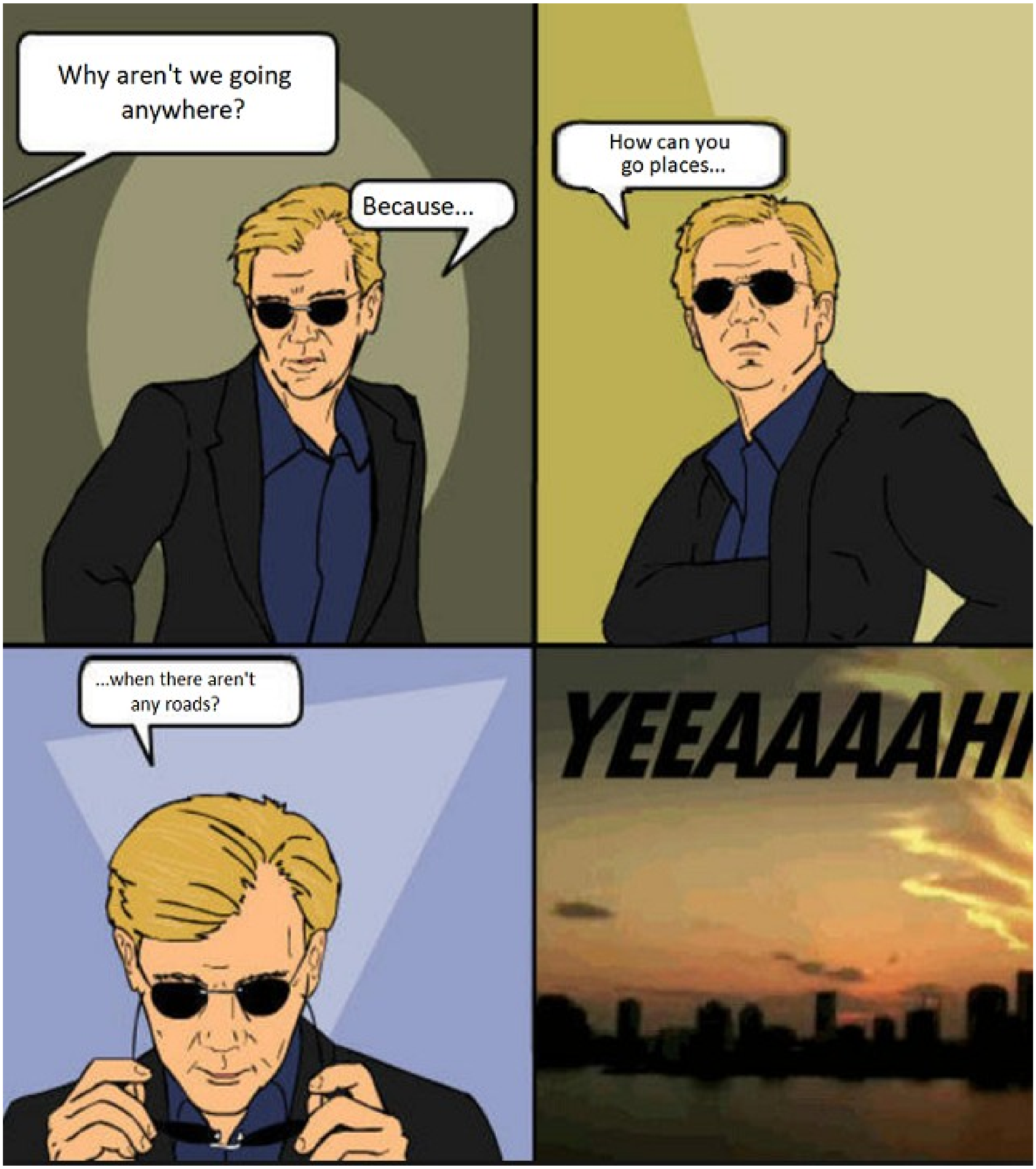
“Nobody, rich or poor, is going to go to the farm and roll in pig excrement.”

The rich woman looked skeptical.

“You’re not going to go your factory and harass your employees.”

The factory owner frowned.

“And *you* aren’t going to go set off your nuclear bomb. None of you are going anywhere!”


Why aren't we going anywhere?
Because...
How can you go places...
...when there aren't any roads?
YEEAAAAAH

Highlights From The Comments On Technological Unemployment

Thanks to everyone who commented on the post about technological unemployment.

From Onyomi:

> Not saying I necessarily think this is what is going on, but one simple possible explanation for why technological unemployment could happen now when it never happened much in the past could be quite simply the greatly accelerated pace of change.
>
> For most of history, technological change was very, very slow. The past few hundred years we've moved increasingly to a place where each new generation has to learn to function in a world different from the one their parents grew up in. We could now be moving to a world where each generation has to learn to function in multiple worlds over the course of a lifetimes, which may stretch the limits of human adaptability.

Versus Bugmaster:

> I don't know much about economics, so the following is just conjecture, but: I think one reason for this state of affairs might be the reduced pace of technological innovation. In the past, when an industry (such as horsemanship, agriculture, switchboard operation, etc.) was automated, the technological advancements that led to this automation also led to the creation of several new industries. To use an extreme example, the same technology that eliminated switchboard operators led to

> the massive paradigm shift that was the Information Age.
>
> However, in the modern Western world, new industries are created a lot less frequently; and thus workers whose jobs got automated have nowhere to go but down. This could be due to a combination of factors:
>
> 1). Automation today is increasingly the result of incremental advances in the field, as opposed to paradigm-shifting discoveries. Thus, most of the new industries that could’ve been created, already have been.
> 2). Technological progress in general has slowed down significantly, because…
> 3). …All the easy stuff has already been discovered, and/or…
> 4). …A combination of the socio-political climate, as well as extremely aggressive copyright/patent law, is causing a chilling effect (at least, in the Western world).
> 5). Western countries are suffering from resource depletion; it’s harder to innovate on zero budget.

From meh:

> Graphs comparing college to high school over time need to show percent of the population each group is over time. (i vaguely remember this coming up on the blog before)

This is definitely true. For example, suppose that in 1950, only the richest 10% of people went to college. But in 2050, all but the poorest 10% of people go to college. We might discover that we had gone from most uncolleged people having decent jobs to most uncolleged people having terrible jobs – not because the importance of college had changed, but only because the poorest 90% of people might usually have mostly decent jobs, and the poorest 10% of people usually have terrible jobs. When I was doing some of this research, I asked Economics Tumblr whether this was a big deal, and I was told that some people had done the relevant calculations and they had been found not to affect any of these conclusions very much.

Chris Williams on why people on disability might claim not to want jobs even if they do:

> Part of this is also the way the social safety net is setup. I know a dude I went to high school with years ago. He's bipolar and a schizophrenic (sp?). Unmedicated, he basically goes all Charlie Sheen crazy from time to time. It makes him unable to function in society.So he's considered disabled. However, he responds very well to medication. Now that he's declared disabled and gets medication and doctor visits/therapy for free through Medicaid, and is on meds, he's turned into a completely normal guy. 95% normal anyway. It's completely turned his life around. Now, he would be perfectly able to hold down a job, or go to school, or learn a trade. That's the rub though. If he gets a job or goes to school it serves as evidence he's not disabled and could get him kicked off of food stamps, public housing and medicaid. So it's a safer deal for him to stay home and play xbox than to contribute to society. He'd rather be "normal" in the head and sitting around than being crazy – which are basically his choices. If our social safety net was better designed, we could get some of those people back in the workforce.

There's some disagreement and further discussion in the responses to his comment.

From Wrong Species:

> It’s bugging me that you’re comparing total number of manufacturing jobs with percent of men in the labor force. It should either be both percent or both total.

This is a really important point which I messed up in the original version of the article. I originally thought it wouldn’t matter that much (US population hasn’t grown *that much* over time, has it?) but when I finally found the percent manufacturing jobs graph, it looked totally different and changed all my conclusions. I briefly panicked, then realized that probably what was going on was women were entering the labor force, generally into non-manufacturing jobs, and so the percent of workers in manufacturing was gradually going down over time. I can’t find a graph that adjusts for that, so I’m just going to trust my existing number of jobs graph for now.

From the same comment, on Autor’s example about bank tellers:

> [Autor wrote]: “Consider the surprising complementarities between information technology and employment in banking, specifically the experience with automated teller machines (ATMs) and bank tellers documented by Bessen (2015). ATMs were introduced in the 1970s, and their numbers in the US economy quadrupled from approximately 100,000 to 400,000 between 1995 and 2010. One might naturally assume that these machines had all but eliminated bank tellers in that interval. But US bank teller employment actually rose modestly from 500,000 to approximately 550,000 over the 30-year period from 1980 to 2010 (although given the growth in the labor force in this time interval,

> these numbers do imply that bank tellers declined as a share of overall US employment).”
>
> I find it odd how people just gloss over that [statement in parentheses]. The right story is the simple one, where ATMs reduced employment among tellers. Rise through population growth doesn’t count.

DocKaon agrees:

> The Bessen (2015) ATM study is a great example of the conclusion not at all matching what the actual data says. Go look at the plots in that paper. What I see is that 10 years after ATMs are introduced employment of tellers plateaus and stays there with very little increase until the present. Virtually all the job growth in the sector occurs within the first 10 years after ATMs are introduced. Eyeballing it, from 1970-1980 tellers increase by ~5% a year and 1980-2010 by 0.6% a year. In other words, it takes approximately 10 years for the ATMs to become widespread and impact tellers jobs.

From sohois:

> I believe it was Gwern who made the argument that even figures such as PAMLFPR were poor ways of trying to understand total employment, since such categorization is very much an invention of the modern era.
>
> As I understood their argument, even 100 years ago ideas such as retirement and compulsory education simply didn’t exist (note: am not historian, I don’t know when exactly compulsory schooling or retirement would have become widespread); anyone who wasn’t very wealthy would start working as soon as they were physically able, and continue until they were either dead

> or physically unable, more likely the former given health outcomes in those times. Thus, technological unemployment did occur following the industrial revolution, because it allowed for a huge number of former workers to stop working. However, economists did not recognize this as a decline in employment since new categories of education and retirement were introduced to absorb them.

Part of the reason the statistic only measures prime age people (between 25 and 55) is that they're at an age where they're probably not in school or retired. I don't know if the decrease of working years through the lifespan can fairly be called "technological unemployment" any more than the entrance of women into the workforce can be called "technological superemployment". They're just supply-side trends.

From Grey Enlightenment:

> Regarding the horse example, the US horse population is 9 million. The demand for horses never went away despite automation and has remained steady. If one tried to extrapolate the early 20th century horse population trend to the future, there should presently be no horses alive.

Wikipedia says there were 20 million horses in the US in 1915, falling to a low of 4.5 million in 1959, and increasing to 9 million now. So the overall story of technological unenhorsement survives this objection.

Ricraz on whether perfect cheap androids would really make human workers obsolete:

> Significant parts of many jobs are driven by social interactions and status considerations. The job of

> doorman at fancy hotels may be both the most easily automated job in existence, and also the last one to actually be automated. In this particular scenario, it depends on how people relate to androids. If society has the general idea that interacting with androids is low-status and not as meaningful as human interactions, then there would be an explosion in customer-facing jobs, as it becomes the easiest way to differentiate your product. Also, everything produced by androids would be so cheap that people wouldn't need to work very long hours in those jobs to earn a living.

In other words, whether or not humans are practically useful, they'll always be useful for aesthetic and signaling purposes. This provides the missing piece of the technological unenhorsement story; machinery made horses mostly useless, but aesthestic and signaling purposes were enough to maintain them at 25% to 50% of their maximum population.

I find this worrying for the same reason I find the lack of technological unemployment worrying; it means there might never come a time when we're forced to really confront our values and decide whether we want to be properly post-scarcity. Even in the world with perfect android laborers, instead of letting the androids labor and letting humans live in leisure and share part of the pie, we'll have the opportunity to sleepwalk into letting a few android-owners get super rich and giving everyone else jobs as hotel doormen. What do you *mean* it's time for a universal basic income? You should be grateful for the hotel doorman job you have!

AnteriorMotive with one last unenhorsement related point:

> I think the horse example is deceptive: the bottleneck for horses is human labour. Horse breeders, handlers, etc, found better uses for their labour and transitioned to other sectors of the economy. Put another way, horses were priced out because they need to hire humans to be able to do anything, and humans' opportunity cost got so high they could no longer afford to outbid it.
>
> You can tweak the metaphor to make the intended comparison more applicable, but it rapidly starts losing its rhetorical power. I recommend the discussion in the comments of: https://slatestarcodex.com/2017/02/06/notes-from-the-asilomar-conference-on-beneficial-ai/
>
> "This wasn't to say that they aren't an argument for human obsolescence, just that when you exchange the horse analogy for the structurally identical 'cassette tape' or 'slide-rule' anology, suddenly it's lost its rhetorical punch. As I see it, this suggests a misleading analogy, hinging on the listener inappropriately anthropomorphizing horses."

Good point. This is why I previously got accused of "thinking horses are LITERALLY humans". I apologize for the error. But I'm not going to back down on them basically just being elongated cows.

vV_Vv asks if the existence of a middle class is just a weird historical blip:

> For most of recorded history, almost every human was a subsistence farmer, likely under-using their cognitive abilities, yet no middle-paying jobs materialized to use their full potential. Then, during the first century or so of

> the industrial revolution, most people in industrialized countries were assembly line workers, jobs even less cognitively demanding than subsistence farming, and still no middle-paying jobs appeared until most of that low-skilled factory work could be automated. There is nothing in standard economic theory that predicts that jobs that allow average people to use their full cognitive ability should exist and pay middle-class wages. In fact, there is nothing that predicts that a middle-class should exist.

Related: Matthias quoting Goldin & Katz:

> U.S. educational and occupational wage differentials were exceptionally high at the dawn of the twentieth century and then decreased in several stages over the next eight decades. But starting in the early 1980s the labor market premium to skill rose sharply and by 2005 the college wage premium was back at its 1915 level. The twentieth century contains two inequality tales: one declining and one rising. We use a supply-demand-institutions framework to understand the factors that produced these changes from 1890 to 2005. We find that strong secular growth in the relative demand for more educated workers combined with fluctuations in the growth of relative skill supplies go far to explain the long-run evolution of U.S. educational wage differentials. An increase in the rate of growth of the relative supply of skills associated with the high school movement starting around 1910 played a key role in narrowing educational wage differentials from 1915 to 1980. The slowdown in the growth of the relative supply of college workers starting around 1980 was a major reason for the surge in the college wage premium from 1980 to 2005. Institutional factors were important at various

> junctures, especially during the 1940s and the late 1970s.

From Yosarian2:

> Looking at your numbers we may have 2 different things happening here. Maybe from 1960-2000, the decrease in male employment was caused by the increase in female employment; and from 2000-today, female employment has been more flat, and the decrease in male employment has been caused by automation? Does that correlate better to the numbers?

A similar point from Proyas:

> Also, it might be the case that the entry of women into the workforce starting 50 years ago permanently depressed wages and drove down the labor force participation rate. Is there an economic theory stating that the number of jobs must keep growing and their wages must also grow or stay the same, regardless of how many new workers are added to the system?… Picking a random example, if the number of people with J.D.'s in America doubles because women are no longer barred from entry, is there any reason to assume that the economy will automatically expand to absorb them all, with the number of attorney jobs doubling and salaries staying the same? "Is there an economic theory" that says such?

Alef is more (less?) subtle:

> So in 1950 we introduced (or rather, used far more widely than just its former niches) a new technology that was an essentially perfect substitute for manpower. To pretty much exactly the extent it became increasingly deployed, the 'workforce' declined (an

> almost perfect offset for 25 years, a bit looser thereafter). In the past, new technologies have (after some painful adjustment time) lead to new jobs, new opportunities, and a recovery or gain in ’employment’; but in this case we’ve waited a couple of generations and still seen absolutely nothing in the way of recovering from this offset. Is this experience not relevant, as robots and AI will very shortly do this again? For a while, they won’t be remotely as capable or flexible substitutes for manpower as in the former case; but they will also be cheaper.

I’m sure using women as a metaphor for robots will greatly increase this blog’s popularity among the feminists. But this is a good point – why *shouldn’t* wages have crashed when the labor supply doubles?

The only good paper I can find on this claims it didn’t happen. I agree that’s surprising and that it needs more thought.

Probably the same arguments about how immigrants (and new graduates) entering the workforce don’t necessarily drive down wages should apply here also. The only caveat is that women are already around and consuming, so they might not create as many new jobs by entering the workforce as (say) a new immigrant arriving in the country. But there’s also the effect where somebody has to do whatever they were doing before (housework? child-rearing?) so that could possibly balance it out.

There’s little room for women entering the workforce to contribute to unemployment, since Part I of the original post explained that there is not really an unexplained unemployment increase. I wish I had a better answer to the wage stagnation question, and will be eternally grateful to anyone who can direct me to more literature or analyses on this.

From VolumeWarrior:

> Surprised to see no mention of videogames. The career prospects of young men are humiliating or tediously circuitous to puruse. So they just stay at home and play Xbox. See http://marginalrevolution.com/marginalrevolution/2016/09/labor-force-participation-video-games.html

Again, I want to emphasize that this is *solving a non-problem*. There is no unexplained jump in unemployment or nonparticipation. You can all stop trying to come up with explanations for it.

From Swami:

> The data clearly shows (I can link if desired) that the larger trend happening over past thirty years is the middle class is shrinking primarily by moving into higher tiers or retiring, the lower class has grown slightly, but not if you exclude thirty to forty million immigrants. And here much of the statistical increase has been caused by lower numbers of married families (and thus fewer jumps to middle income due to failure to get hitched).

Versus sdenton4:

> Given the categorization of jobs by type, there’s presumably a hell of a lot more low paying jobs than high paying jobs. So a smaller percentage increase in low paying jobs could presumably account for a lot more workers than the larger percentage increase in high paying jobs.

From Pinyaka:

> I don't necessarily think that most people think that technological unemployment will hit the least skilled first. It should hit the jobs where it's most profitable first. How cheap would a robotic server in a restaurant have to be to actually get a return on investment? Replacing a unionized assembly line worker makes you money a lot faster even if it's more expensive. As we get better at replacing middle cost jobs, the cost of automation should go down and that should allow us to replace the unskilled.

From Disillusioned9:

> I have two comments: first, the automation of manufacturing may now have meaningfully affected the PAMLFPR nationwide, but surely the localized effects may tell a different story? The economies of manufacturing heavy States like Ohio are in crisis, which would strongly suggest that either PALFPR or underemployment would go up in these areas. Increases in non-manufacturing jobs elsewhere in the nation could smooth those numbers in the US and hide regional inequities.

Yeah, the lack of obvious jumps in unemployment and PAMLFNP have to be squared with the universal common-sense perception that parts of the Rust Belt are an apocalyptic mess right now. The idea that these areas are getting anomalously worse while others (Big cities? The coasts? Silicon Valley?) are getting anomalously better might be one explanation for how the statistics could all even out.

From resalisbury:

An argument that often gets overlooked is that housing costs in high productivity cities are crowding out potential increases in employment and productivity.

What if we are as good as ever at fundamental technological aspect of productivity innovations but have gotten worse at moving people to places where those innovations are occurring?

Interstate labor mobility rates are at an all time low since the 1960s and housing costs in our most productive cities are at an all time high.

Hsieh and Moretti (2017) argue in “Housing Constrains and Spatial Misallocation” that increases in housing costs due primarily to zoning have reduced aggregate GDP growth by 50% since 1960. That is huge. Even if they are off by an order of magnitude, it could still help explain several pp of the reduction in prime age workforce. Additional workers have to come from somewhere…

https://faculty.chicagobooth.edu/chang-tai.hsieh/research/growth.pdf

In San Francisco the median housing price is over $1 million. Not everyone needs to move here and be a programmer there’s lots of other work to be done. Unfortunately, most people can’t afford to live in the city so they sleep in cars or commute in 2 hours just to work. Imagine how many more people would work if they could just live here.

The Economist argued that if not for housing costs the population of the Bay Area would be around 30 million (can’t find the link) instead of ~7 million. That’s just one

> region where the employment growth would impact things on a national scale.
>
> Imagine what would have happened to Detroit if in the early years housing prices spike to over $1 million. Or if in the 1860s when millions were migrating to the United States they arrived to discover that the median housing cost $1 million. They would write home and tell friends not to come.
>
> Housing costs in the last few decades have finally reached a point where they have become a constraint on growth that was not true during previous periods of economic expansion in the US.
>
> What to do about housing costs? Obama’s outgoing team of economic advisors published a paper with policy recommendations which can be see here. Basically if calls for increased density and by right zoning (ie you can’t put someone’s bulding permit through 3 years of discretionary review and comment).
>
> https://www.whitehouse.gov/sites/whitehouse.gov/files/images/Housing_Development_Toolkit%20f.2.pdf

An unexpected place for the housing crisis to show up. At what point do I have to end all these posts with “ceterum autem censeo domuum numerum augendum esse”?

SSC Journal Club: Cipriani On Antidepressants

I.

The big news in psychiatry this month is Cipriani et al's Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. It purports to be the last word in the "do antidepressants work?" question, and a first (or at least early) word in the under-asked "which antidepressants are best?" question.

This study is very big, very sophisticated, and must have taken a very impressive amount of work. It meta-analyzes virtually every RCT of antidepressants ever done – 522 in all – then throws every statistical trick in the book at them to try to glob together into a coherent account of how antidepressants work. It includes Andrea Cipriani, one of the most famous research psychiatrists in the world – and John Ioannidis, one of the most famous statisticians. It's been covered in news sources around the world: my favorite headline is Newsweek's unsubtle Antidepressants Do Work And Many More People Should Take Them, but honorable mention to Reuters' Study Seeks To End Antidepressant Debate: The Drugs Do Work.

Based on the whole "we've definitely proven antidepressants work" vibe in coverage, you would think that they'd directly contradicted Irving Kirsch's claim that antidepressants aren't very effective. I've mentioned my disagreements with Kirsch before, but it would be nice to have a definitive refutation of his work. This study isn't really it. Both Kirsch and Cipriani agree that antidepressants have statistical significance –

they're not literally doing nothing. The main debate was whether they were good enough to be worth it. Kirsch argues they aren't, using a statistic called "effect size". Cipriani uses a different statistic called "odds ratio" that is hard to immediately compare.

[EDIT: Commenters point out that once you convert Cipriani's odds ratios to effect sizes, the two studies are pretty much the same – in fact, Cipriani's estimates are (slightly) *lower*. That is, "the study proving antidepressants work" presents a worse picture of antidepressants than "the study proving antidepressants don't work". If I had realized this earlier, this would have been the lede for this article. This makes all the media coverage of this study *completely insane* and means we're doing science based entirely on how people choose to sum up their results. Strongly recommend this Neuroskeptic article on the topic. This is very important and makes the rest of this article somewhat trivial in comparison.]

Kirsch made a big deal of trying to get all the evidence, not just the for-public-consumption pharma-approved data. Cipriani also made such an effort, but I'm not sure how comparable the two are. Kirsch focused on FDA trials of six drugs. Cipriani took every trial ever published – FDA, academia, industry, whatever- of twenty-one drugs. Kirsch focused on using the Freedom Of Information Act to obtain non-public data from various failed trials. Cipriani says he looked pretty hard for unpublished data, but he might not have gone so far as to harass government agencies. Did he manage to find as many covered-up studies as Kirsch did? Unclear.

How confident should we be in the conclusion? These are very good researchers and their methodology is unimpeachable. But a lot of the 522 studies they cite are, well, kind of crap. The researchers acknowledge this and have constructed some kind of incredibly sophisticated model that inputs the chance of bias in each study and weights everything and simulates all sorts of assumptions to make sure they don’t change the conclusions too much. But we are basically being given a giant edifice of suspected-crap fed through super-powered statistical machinery meant to be able to certify whether or not it’s safe.

Of particular concern, 78% of the studies they cite are sponsored by pharmaceutical industries. The researchers run this through their super-powered statistical machinery and determine that this made no difference – in fact, if you look in the supplement, the size of the effect was literally zero:

> In our analyses, funding by industry was not associated with substantial differences in terms of response or dropout rates. However, non-industry funded trials were few and many trials did not report or disclose any funding.

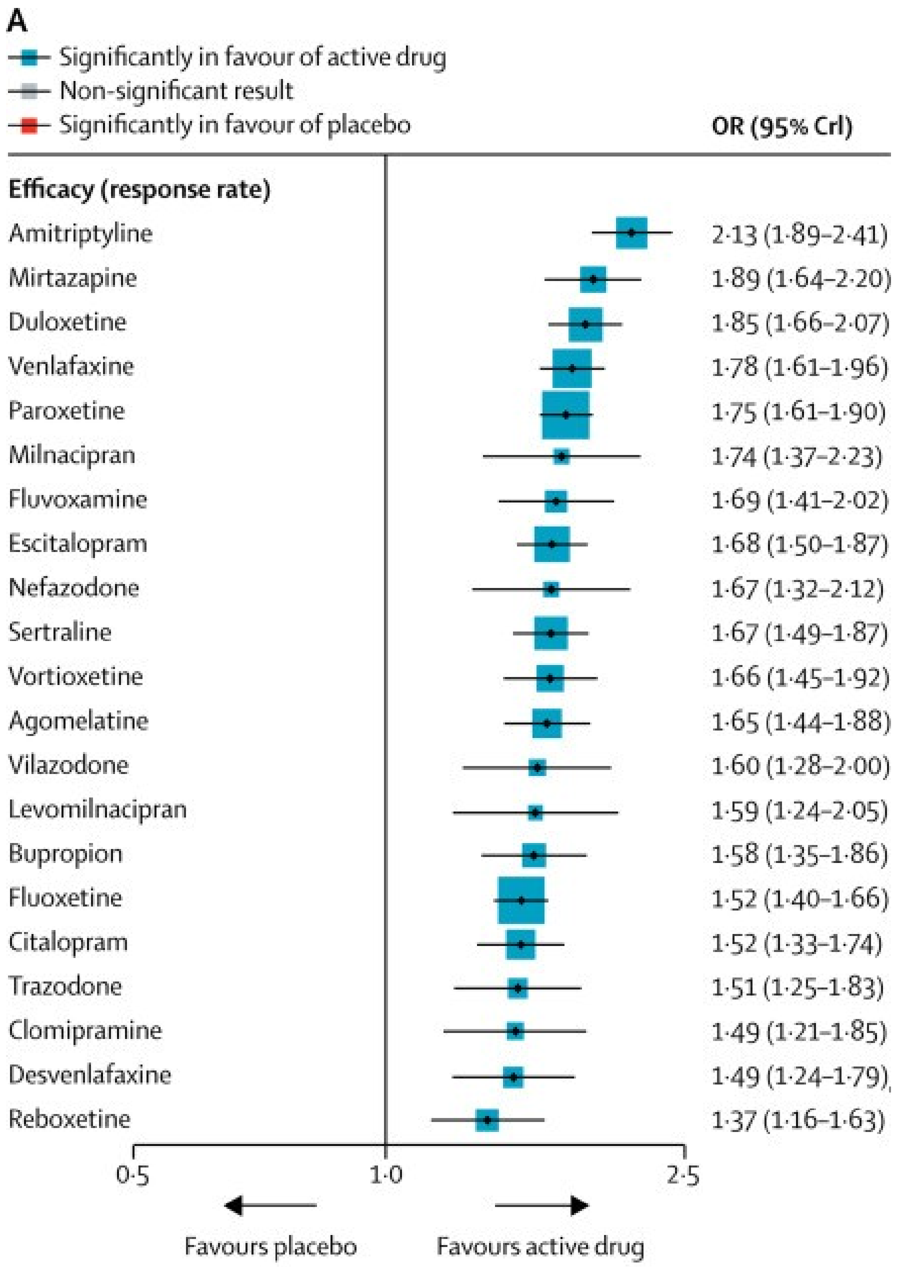
This is surprising, since other papers (which the researchers dutifully cite) find that pharma-sponsored trials are about five times more likely to get positive results than non-sponsored ones (though see this comment). Cipriani’s excuse is that there weren’t enough non-industry trials to really get a good feel for the differences, and that a lot of the trials marked “non-industry” were probably secretly by industry anyway (more on this later). Fair enough, but if we can’t believe their “sponsorship makes zero difference to outcome” result, then the whole thing starts seeming kind of questionable.

I don’t want to come on too strong here. Science is never supposed to have to wait for some impossible perfectly-unbiased investigator. It’s supposed to accept that everyone will have an agenda, but strive through methodological rigor, transparency, and open debate to transcend those agendas and create studies everyone can believe. On the other hand, we’re really not very good at that yet, and nobody ever went broke overestimating the deceptiveness of pharmaceutical companies.

And there was one other kind of bias that did show up, hard. When a drug was new and exciting, it tended to do better in studies. When it was old and boring, it tended to do worse. You could argue this is a placebo effect on the patients, but I’m betting it’s a sign that people were able to bias the studies to fit their expected results (excited high-tech thing is better) in ways we’re otherwise not catching.

All of this will go double as we start looking at the next part, the ranking of different antidepressants.

II.



All antidepressants, from best to worst! (though note the wide error bars)

If this were for real, it would be an amazing resource. Psychiatrists have longed to know if any antidepressant is truly better than any other. Now that we know this, should we just start everyone on amitriptyline (or mirtazapine if we’re worried about tricyclic side effects) and throw out the others?

(as a first line, of course. In reality, we try the best one first, but keep going down the list until we find one that works for you and your unique genetic makeup.)

This matches some parts of the psychiatric conventional wisdom and overturns other parts. How much should we trust this versus all of the rest of the lore and heuristics and smaller studies that have accreted over the years?

Some relevant points:

1. The study finds that all the SSRIs cluster together as basically the same, as they should. The drugs that stand out as especially good or especially bad are generally unique ones with weird pharmacology that *ought* to be especially different. Amitriptyline is a tricyclic, and very different from clomipramine which is the only other tricyclic tested. Mirtazapine does weird things to presynaptic norepinephrine. Duloxetine and venlafaxine are SNRIs. This passes the most obvious sanity check.

2. Amitriptyline, the most effective antidepressant in this study, is widely agreed to be very good. See eg Amitriptyline: Still The Leading Antidepressant After 40 Years Of Randomized Controlled Trials. Amitriptyline does have many side effects that limit its use despite its impressive performance. I secretly still believe MAOIs, like phenelzine and tranylcypromine, to be even better than amitriptyline, but this study doesn’t include them so we can’t be sure.

3. Reboxetine, the least effective antidepressant in this study, is widely known to suck. It is not available in the United States becaues the FDA wouldn’t even approve it here.

4. On the other hand, agomelatine, another antidepressant widely known to suck, gains solid mid-tier status here, being

about as good as anything else. The study even lists it as one of seven antidepressants that seems to do especially well (though it’s unclear what they mean and it’s obviously a different measure than this graph). But agomelatine was rejected by the FDA for not being good enough, scathingly rejected by the European regulators (although their decision was later reversed on appeal), and soundly mocked by various independent organizations and journals (1, 2). It doesn’t look like Cipriani has access to any better data than anyone else, so how come his results are so much more positive?

5. Venlafaxine and desvenlafaxine are basically the same drug, minus a bunch of BS from the pharma companies trying to convince that desvenlafaxine is a super-new-advanced version that you should spend twenty times as much money on. But venlafaxine is the fourth most efficacious drug in the analysis; desvenlafaxine is the second *least* efficacious drug. Why should this be? I have similar complaints about citalopram and escitalopram. Should we privilege common sense over empiricism and say Cipriani has done something wrong? Or should we privilege empiricism over common sense and conclude that the super-trivial differences between these chemicals have some outsized metabolic significance that makes a big clinical difference? Or should we just notice that the 95% confidence intervals of almost everything in the study (including these two) overlap, so really Cipriani isn’t claiming to know anything about anything and it’s not surprising if the data are wrong?

6. I’m sad to see clomipramine doing so badly here, since I generally find it helpful and have even evangelized it to my friends. I accept that it has serious side effects, but I expected it to do at least a little better in terms of efficacy.

Hoping to rescue its reputation, I started looking through some of the clomipramine studies cited. First was Andersen 1986, which compared clomipramine to Celexa and found some nice things about Celexa. This study doesn't say a pharmaceutical company was involved in any way. But I notice the study was done in Denmark. And I also notice that Celexa is made by Lundbeck Pharmaceuticals, a Danish company. Am I accusing an entire European country of being in a conspiracy to promote Celexa? Would that be crazy?

The second clomipramine study listed is De Wilde 1982, which compared clomipramine to Luvox and found some nice things about Luvox. This study also doesn't say a pharmaceutical company was involved in any way. But I notice the study was done in Belgium. And I also notice that Luvox is made by Solvay Pharmaceuticals, a Belgian company. Again, I'm sure Belgium is a lovely country full of many people who are not pharma shills, but this is starting to get a little suspicious.

To Cipriani's credit, his team did notice these sorts of things and mark these trials as having "unclear" sponsorship levels, which got fed into the analysis. But I'm actually a little concerned about the exact way he did this. If a pharma company sponsored a trial, he called the pharma company's drug's results biased, and the comparison drugs unbiased. That is, suppose that Lundbeck sponsors a study, comparing their new drug Celexa to old drug clomipramine. We assume that they're trying to make it look like Celexa is better. In this study, Cipriani would mark the Celexa patients as biased, but the clomipramine patients as unbiased.

But surely if Lundbeck wants to make Celexa look good, they can either finagle the Celexa numbers upward, finagle the clomipramine numbers downward, or both. If you flag Celexa as high risk of being finagled upwards, but don't flag clomipramine as at risk of being finagled downwards, I worry you're likely to understate clomipramine's case.

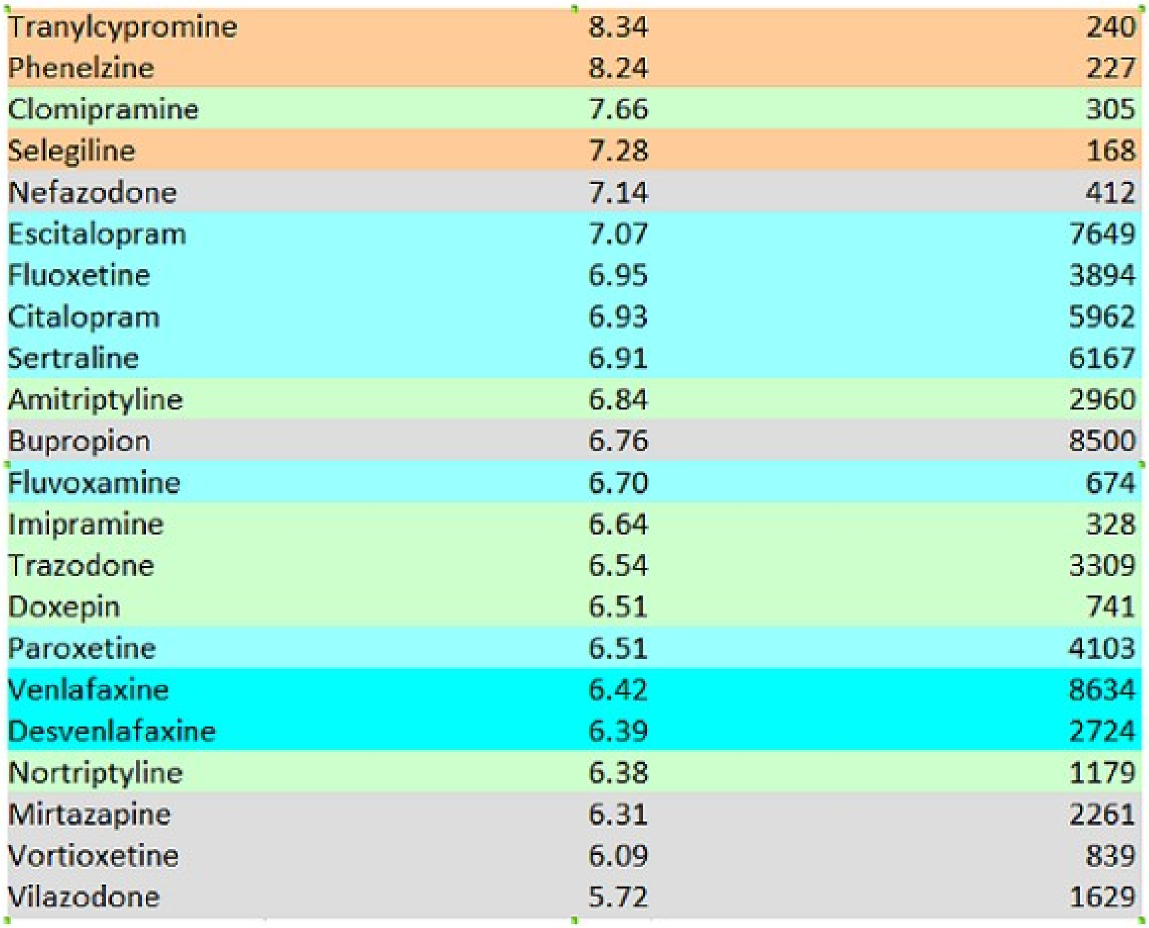
I make a big deal of this because about a dozen of the twenty clomipramine studies included in the analysis were very obviously pharma companies using clomipramine as the comparison for their own drug that they wanted to make look good; I suspect some of the non-obvious ones were too. If all of these are marked as "no risk of bias against clomipramine", we're going to have clomipramine come out looking pretty bad.

Clomipramine is old and canonical, so most of the times it gets studied are because some pharma company wants to prove their drug is at least as good as this well-known older drug. There are lots of things like this, where certain drugs tend to inspire a certain type of study. Cipriani says they adjusted for this. I hope they were able to do a good job, because this is a big deal and really hard to factor out entirely.

This is my excuse for why I'm not rushing to prescribe drugs in the exact order Cipriani found. It's a good study and will definitely influence my decisions. But it's got enough issues that I feel justified in taking my priors into account too.

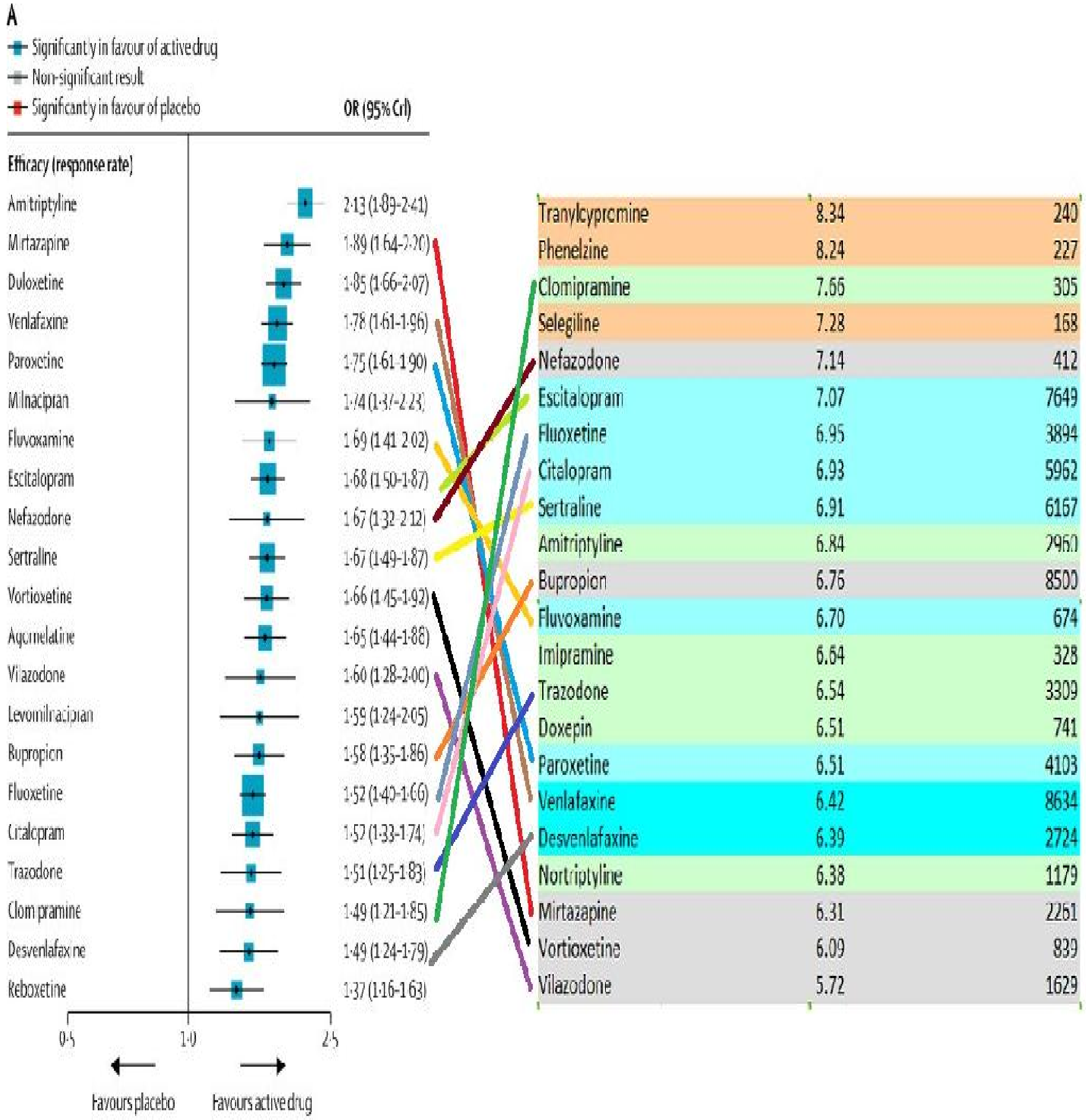
III.

Speaking of which, here's another set of antidepressant rankings:



Tranylcypromine	8.34	240
Phenelzine	8.24	227
Clomipramine	7.66	305
Selegiline	7.28	168
Nefazodone	7.14	412
Escitalopram	7.07	7649
Fluoxetine	6.95	3894
Citalopram	6.93	5962
Sertraline	6.91	6167
Amitriptyline	6.84	2960
Bupropion	6.76	8500
Fluvoxamine	6.70	674
Imipramine	6.64	328
Trazodone	6.54	3309
Doxepin	6.51	741
Paroxetine	6.51	4103
Venlafaxine	6.42	8634
Desvenlafaxine	6.39	2724
Nortriptyline	6.38	1179
Mirtazapine	6.31	2261
Vortioxetine	6.09	839
Vilazodone	5.72	1629

This is from Alexander et al 2017, which started life as this blog post but which with help from some friends I managed to get published by a journal. We looked at some different antidepressants than Cipriani did, but there are enough of the same ones that we can compare results.



Tranylcypromine	8.34	240
Phenelzine	8.24	227
Clomipramine	7.66	305
Selegiline	7.28	168
Nefazodone	7.14	412
Escitalopram	7.07	7649
Fluoxetine	6.95	3894
Citalopram	6.93	5962
Sertraline	6.91	6167
Amitriptyline	6.84	2960
Bupropion	6.76	8500
Fluvoxamine	6.70	674
Imipramine	6.64	328
Trazodone	6.54	3309
Doxepin	6.51	741
Paroxetine	6.51	4103
Venlafaxine	6.42	8634
Desvenlafaxine	6.39	2724
Nortriptyline	6.38	1179
Mirtazapine	6.31	2261
Vortioxetine	6.09	839
Vilazodone	5.72	1629

Everything is totally different. I haven’t checked formally, but the correlation between those two lists looks like about zero. We find mirtazapine and venlafaxine to be unusually *bad*, and amitriptyline to be only somewhere around the middle.

I don’t claim anywhere near the sophistication or brilliance or level of work that Cipriani et al put in. But my list – I will argue – makes sense. Drugs with near-identical chemical structure – like venlafaxine and desvenlafaxine, or citalopram and escitalopram – are ranked similarly. Drugs with similar mechanisms of action are in the same place. We match pieces of psychiatric conventional wisdom like “Paroxetine is the worst SSRI”.

Part of the disagreement may be related to all the antidepressants being very close together on both lists. On Cipriani’s, the difference between the 25th vs. 75th percentile is OR 1.75 vs. OR 1.52. On mine, it’s a rating of 7.14 vs. 6.52. Aside from a few outliers, there’s not a lot of light between any of the antidepressants here, which makes it likely that different methodologies will come up with very different orders. And the few outliers that each of us did identify as truly distinct often didn’t make it into the other’s study – Cipriani doesn’t have MAOIs and I don’t have reboxetine. But this isn’t a good enough excuse. One of my top performers, clomipramine, is near the bottom for Cipriani. One of my bottom performers, mirtazapine, is near his top. I have to admit that these just don’t match.

And a big part of the disagreement has to be that we’re not doing the same things Cipriani did – we’re looking at a measure that combines efficacy and acceptability, whereas Cipriani looked at each separately. This could explain why my data penalizes some side-effect-heavy drugs like mirtazapine and amitriptyline. But again, this isn’t a good enough excuse. Why doesn’t my list penalize other side-effect-heavy meds like clomipramine?

In the end, these are two very different lists that can’t be easily reconciled. If you have any sense, trust a major international study before you trust me playing around with online drug ratings. But also be aware of the study’s flaws and why you might want to retain a bit of uncertainty.

SSC Meetup: Bay Area 3/3

.
WHEN: 3:33 PM on Saturday, 3/3

WHERE: Berkeley campus, meet at the open space beside the intersection of West and Free Speech. Please disregard any kabbalistic implications of the meetup cross-streets.

WHO: Special guest is Gwern of gwern.net. Also me, Katja, and the usual Bay Area crowd.

WHY: Cause it'll be fun. A lot of people have said before that they considered not going because they "don't think they're the typical SSC reader" or they're "not sure they'd be able to keep up" or things like that. In the past, these people have usually had a good time and encouraged me to post something like this encouraging other people like them to come. The more unique and atypical people we get, the more fun it is getting to talk and exchange ideas. It's a pretty low-key environment and very open to just hanging out on the edges of interesting conversations until you find one you're comfortable joining. Also, you'll probably be less socially awkward than I am, and it's my meetup, so everyone has to tolerate me, so they'll have to tolerate you too.

HOW: We haven't done well with cafes or other traditional meetup spaces in the past, so we'll just meet outside and sit on the grass. Bring blankets / refreshments if you want them. If it's raining, we'll meet just inside the Natural History Museum nearby and figure out what to do from there.

See you there!

God Help Us, Let's Try To Understand Friston On Free Energy

I've been trying to delve deeper into predictive processing theories of the brain, and I keep coming across Karl Friston's work on "free energy".

At first I felt bad for not understanding this. Then I realized I wasn't alone. There's an entire not-understanding-Karl-Friston internet fandom, complete with its own parody Twitter account and Markov blanket memes.

From the journal *Neuropsychoanalysis* (which based on its name I predict is a center of expertise in not understanding things):

> At Columbia's psychiatry department, I recently led a journal club for 15 PET and fMRI researhers, PhDs and MDs all, with well over $10 million in NIH grants between us, and we tried to understand Friston's 2010 Nature Reviews Neuroscience paper – for an hour and a half. There was a lot of mathematical knowledge in the room: three statisticians, two physicists, a physical chemist, a nuclear physicist, and a large group of neuroimagers – but apparently we didn't have what it took. I met with a Princeton physicist, a Stanford neurophysiologist, a Cold Springs Harbor neurobiologist to discuss the paper. Again blanks, one and all.

Normally this is the point at which I say "screw it" and give up. But almost all the most interesting neuroscience of the past decade involves this guy in one way or another. He's the most-cited living neuroscientist, invented large parts of modern brain imaging, and received the prestigious Golden Brain Award (which is somehow a real thing). His Am I

Autistic – An Intellectual Autobiography short essay, written in a weirdly lucid style and describing hijinks like deriving the Schrodinger equation for fun in school, is as consistent with genius as anything I've ever read.

As for free energy, it's been dubbed "a unified brain theory" (Friston 2010), a key through which "nearly every aspect of [brain] anatomy and physiology starts to make sense" (Friston 2009), "[the source of] the ability of biological systems to resist a natural tendency to disorder" (Friston 2012), an explanation of how life "inevitably and emergently" arose from the primordial soup (Friston 2013), and "a real life version of Isaac Asimov's psychohistory" (description here of Allen 2018).

I continue to hope some science journalist takes up the mantle of explaining this comprehensively. Until that happens, I've been working to gather as many perspectives as I can, to talk to the few neuroscientists who claim to even partially understand what's going on, and to piece together a partial understanding. I am not at all the right person to do this, and this is not an attempt to get a gears-level understanding – just the kind of pop-science-journalism understanding that gives us a slight summary-level idea of what's going on. My ulterior motive is to get to the point where I can understand Friston's recent explanation of depression, relevant to my interests as a psychiatrist.

Sources include Dr. Alianna Maren's How To Read Karl Friston (In The Original Greek), Wilson and Golonka's Free Energy: How the F*ck Does That Work, Ecologically?, Alius Magazine's interview with Friston, Observing Ideas, and (especially) the ominously named Wo's Weblog.

From these I get the impression that part of the problem is that "free energy" is a complicated concept being used in a lot of different ways.

First, free energy is a specific mathematical term in certain Bayesian equations.

I’m getting this from here, which goes into much more detail about the math than I can manage. What I’ve managed to extract: Bayes’ theorem, as always, is the mathematical rule for determining how much to weigh evidence. The brain is sometimes called a Bayesian machine, because it has to create a coherent picture of the world by weighing all the different data it gets – everything from millions of photoreceptors’ worth of vision, to millions of cochlear receptors worth of hearing, to all the other sense, to logical reasoning, to past experience, and so on. But actually using Bayes on all this data quickly gets computationally intractable.

Free energy is a quantity used in “variational Bayesian methods”, a specific computationally tractable way of approximating Bayes’ Theorem. Under this interpretation, Friston is claiming that the brain uses this Bayes-approximation algorithm. Minmizing the free energy quantity in this algorithm is equivalent-ish to trying to minimize prediction error, trying to minimize the amount you’re surprised by the world around you, and trying to maximize accuracy of mental models. This sounds in line with standard predictive processing theories. Under this interpretation, the brain implements predictive processing through free energy minimization.

Second, free energy minimization is an algorithm-agnostic way of saying you’re trying to approximate Bayes as accurately as possible.

This comes from the same source as above. It also ends up equivalent-ish to all those other things like trying to be correct in your understanding of the world, and to standard predictive processing.

Third, free energy minimization is a claim that the fundamental psychological drive is the reduction of uncertainty.

I get this claim from the Alius interview, where Friston says:

> If you subscribe to the premise that that creatures like you and me act to minimize their expected free energy, then we act to reduce expected surprise or, more simply, resolve uncertainty. So what's the first thing that we would do on entering a dark room — we would turn on the lights. Why? Because this action has epistemic affordance; in other words, it resolves uncertainty (expected free energy). This simple argument generalizes to our inferences about (hidden or latent) states of the world — and the contingencies that underwrite those states of affairs.

The discovery that the only human motive is uncertainty-reduction might come as a surprise to humans who feel motivated by things like money, power, sex, friendship, or altruism. But the neuroscientist I talked to about this says I am not misinterpreting the interview. The claim really is that uncertainty-reduction is the only game in town.

In a sense, it must be true that there is only one human motivation. After all, if you're Paris of Troy, getting offered the choice between power, fame, and sex – then some mental module must convert these to a common currency so it can decide which is most attractive. If that currency is, I dunno, dopamine in the striatum, then in some reductive sense, the only human motivation is increasing striatal dopamine (don’t philosophize at me, I know this is a stupid way of framing things, but you know what I mean). Then the only weird thing about the free energy formulation is identifying the common currency with uncertainty-minimization, which is some specific thing that already has another meaning.

I think the claim (briefly mentioned eg here) is that your brain hacks eg the hunger drive by “predicting” that your mouth is full of delicious food. Then, when your mouth is not full of delicious food, it’s a “prediction error”, it sets off all sorts of alarm bells, and your brain’s predictive machinery is confused and uncertain. The only way to “resolve” this “uncertainty” is to bring reality into line with the prediction and actually fill your mouth with delicious food. On the one hand, there is a lot of basic neuroscience research that suggests something like this is going on. On the other, Wo’s writes about this further:

> The basic idea seems to go roughly as follows. Suppose my internal probability function Q assigns high probability to states in which I’m having a slice of pizza, while my sensory input suggests that I’m currently not having a slice of pizza. There are two ways of bringing Q in alignment with my sensory input: (a) I could change Q so that it no longer assigns high probability to pizza states, (b) I could grab a piece of pizza, thereby changing my sensory input so that it conforms to the pizza predictions of Q. Both (a) and (b) would lead to a state in which my (new) probability function Q’ assigns high probability to my (new) sensory input d’. Compared to the present state, the sensory input will then have lower surprise. So any transition to these states can be seen as a reduction of free energy, in the unambitious sense of the term.

Action is thus explained as an attempt to bring one’s sensory input in alignment with one’s representation of the world.

This is clearly nuts. When I decide to reach out for the pizza, I don’t assign high probability to states in which I’m already eating the slice. It is precisely my knowledge that I’m not eating the slice, together with my desire to eat the slice, that explains my reaching out.

There are at least two fundamental problems with the simple picture just outlined. One is that it makes little sense without postulating an independent source of goals or desires. Suppose it’s true that I reach out for the pizza because I hallucinate (as it were) that that’s what I’m doing, and I try to turn this hallucination into reality. Where does the hallucination come from? Surely it’s not just a technical glitch in my perceptual system. Otherwise it would be a miraculous coincidence that I mostly hallucinate pleasant and fitness-increasing states. Some further part of my cognitive architecture must trigger the hallucinations that cause me to act. (If there’s no such source, the much discussed “dark room problem” arises: why don’t we efficiently minimize sensory surprise (and thereby free energy) by sitting still in a dark room until we die?)

The second problem is that efficient action requires keeping track of both the actual state and the goal state. If I want to reach out for the pizza, I’d better know where my arms are, where the pizza is, what’s in between the two, and so on. If my internal representation of the world falsely says that the pizza is already in my mouth, it’s hard to explain how I manage to grab it from the plate.

A closer look at Friston's papers suggests that the above rough proposal isn't quite what he has in mind. Recall that minimizing free energy can be seen as an approximate method for bringing one probability function Q close to another function P. If we think of Q as representing the system's beliefs about the present state, and P as a representation of its goals, then we have the required two components for explaining action. What's unusual is only that the goals are represented by a probability function, rather than (say) a utility function. How would that work?

Here's an idea. Given the present probability function Q, we can map any goal state A to the target function Q^A, which is Q conditionalized on A — or perhaps on certain sensory states that would go along with A. For example, if I successfully reach out for the pizza, my belief function Q will change to a function Q^A that assigns high probability to my arm being outstretched, to seeing and feeling the pizza in my fingers, etc. Choosing an act that minimizes the difference between my belief function and Q^A is then tantamount to choosing an act that realizes my goal.

This might lead to an interesting empirical model of how actions are generated. Of course we'd need to know more about how the target function Q^A is determined. I said it comes about by (approximately?) conditionalizing Q on the goal state A, but how do we identify the relevant A? Why do I want to reach out for the pizza? Arguably the explanation is that reaching out is likely (according to Q) to lead to a more distal state in which I eat the pizza, which I desire. So to compute the proximal target probability Q^A we presumably need to encode the system's more distal goals and then use

> techniques from (stochastic) control theory, perhaps, to derive more immediate goals.
>
> That version of the story looks much more plausible, and much less revolutionary, than the story outlined above. In the present version, perception and action are not two means to the same end — minimizing free energy. The free energy that's minimized in perception is a completely different quantity than the free energy that's minimized in action. What's true is that both tasks involve mathematically similar optimization problems. But that isn't too surprising given the well-known mathematical and computational parallels between conditionalizing and maximizing expected utility.

It's tempting to throw this out entirely. But part of me *does* feel like there's a weird connection between curiosity and every other drive. For example, sex seems like it should be pretty basic and curiosity-resistant. But how often do people say that they're attracted to someone "because he's mysterious"? And what about the Coolidge Effect (known in the polyamory community as "new relationship energy")? After a while with the same partner, sex and romance lose their magic – only to reappear if the animal/person hooks up with a new partner. Doesn't this point to some kind of connection between sexuality and curiosity?

What about the typical complaint of porn addicts – that they start off watching softcorn porn, find after a while that it's no longer titillating, move on to harder porn, and eventually have to get into really perverted stuff just to feel anything at all? Is this a sort of uncertainty reduction?

The only problem is that this is a really specific kind of uncertainty reduction. Why should "uncertainty about what it would be like to be in a relationship with that particular attractive person" be so much more compelling than "uncertainty about what the middle letter of the Bible is", a question which almost no one feels the slightest inclination to resolve? The interviewers ask Friston something sort of similar, referring to some experiments where people are happiest not when given easy things with no uncertainty, nor confusing things with unresolvable uncertainty, but *puzzles* – things that seem confusing at first, but actually have a lot of hidden order within them. They ask Friston whether he might want to switch teams to support a u-shaped theory where people like being in the middle between too little uncertainty or too much uncertainty. Friston…does not want to switch teams.

> I do not think that "different laws may apply at different levels". I see a singular and simple explanation for all the apparent dialectics above: they are all explained by minimization of expected free energy, expected surprise or uncertainty. I feel slightly puritanical when deflating some of the (magical) thinking about inverted U curves and "sweet spots". However, things are just simpler than that: there is only one sweet spot; namely, the free energy minimum at the bottom of a U-shaped free energy function […]
>
> This means that any opportunity to resolve uncertainty itself now becomes attractive (literally, in the mathematical sense of a random dynamical attractor) (Friston, 2013). In short, as nicely articulated by (Schmidhuber, 2010), the opportunity to answer "what would happen if I did that" is one of the most important resolvers of uncertainty. Formally, the resolution of uncertainty (aka intrinsic motivation, intrinsic value, epistemic value, the value of information, Bayesian surprise, etc. (Friston et al., 2017)) corresponds to salience. Note that in active inference, salience

> becomes an attribute of an action or policy in relation to the lived world. The mathematical homologue for contingencies (technically, the parameters of a generative model) corresponds to novelty. In other words, if there is an action that can reduce uncertainty about the consequences of a particular behavior, it is more likely to be expressed.
> Given these imperatives, then the two ends of the inverted U become two extrema on different dimensions. In a world full of novelty and opportunity, we know immediately there is an opportunity to resolve reducible uncertainty and will immediately embark on joyful exploration — joyful because it reduces uncertainty or expected free energy (Joffily & Coricelli, 2013). Conversely, in a completely unpredictable world (i.e., a world with no precise sensory evidence, such as a dark room) there is no opportunity and all uncertainty is irreducible — a joyless world. Boredom is simply the product of explorative behavior; emptying a world of its epistemic value — a barren world in which all epistemic affordance has been exhausted through information seeking, free energy minimizing action.
>
> Note that I slipped in the word “joyful” above. This brings something interesting to the table; namely, the affective valence of shifts in uncertainty — and how they are evaluated by our brains.

The only thing at all I am able to gather from this paragraph – besides the fact that apparently Karl Friston cites himself in conversation – is the Schmidhuber reference, which is actually really helpful. Schmidhuber is the guy behind eg the Formal Theory Of Fun & Creativity Explains Science, Art, Music, Humor, in which all of these are some form of taking a seemingly complex domain (in the mathematical sense of complexity) and reducing it to something simple (discovering a hidden order that makes it more compressible). I think Friston might be trying to hint that free energy minimization works in a Schmidhuberian sense where it applies to learning things that suddenly make large parts of our experience more comprehensible at once, rather than just “Here are some numbers: 1, 5, 7, 21 – now you have less uncertainty over what numbers I was about to tell you, isn’t that great?”

I agree this is one of life’s great joys, though maybe me and Karl Friston are not a 100% typical subset of humanity here. Also, I have trouble figuring out how to conceptualize other human drives like sex as this same kind complexity-reduction joy.

One more concern here – a lot of the things I read about this equivocate between “model accuracy maximization” and “surprise minimization”. These end really differently. Model accuracy maximization sounds like curiosity – you go out and explore as much of the world as possible to get a model that precisely matches reality. Surprise minimization sounds like locking yourself in a dark room with no stimuli, then predicting that you will be in a dark room with no stimuli, and never being surprised when your prediction turns out to be right. I understand Friston has written about the so-called “dark room problem”, but I haven’t had a chance to look into it as much as I should, and I can’t find anything that takes one or the other horn of the equivocation and says “definitely this one”.

Fourth, okay, all of this is pretty neat, but how does it explain all biological systems? How does it explain the origin of life from the primordial soup? And when do we get to the real-world version of psychohistory? In his Alius interview, Friston writes:

> I first came up with a prototypical free energy principle when I was eight years old, in what I have previously called a "Gerald Durrell" moment (Friston, 2012). I was in the garden, during a gloriously hot 1960s British summer, preoccupied with the antics of some woodlice who were frantically scurrying around trying to find some shade. After half an hour of observation and innocent (childlike) contemplation, I realized their "scurrying" had no purpose or intent: they were simply moving faster in the sun — and slower in the shade. The simplicity of this explanation — for what one could artfully call biotic self-organization — appealed to me then and appeals to me now. It is exactly the same principle that underwrites the ensemble density dynamics of the free energy principle — and all its corollaries.

How do the wood lice have anything to do with any of the rest of this?

As best I can understand (and I'm drawing from here and here again), this is an ultimate meaning of "free energy" which is sort of like a formalization of homeostasis. It goes like this: consider a probability distribution of all the states an organism can be in. For example, your body can be at (90 degrees F, heart rate 10), (90 degrees F, heart rate 70), (98 degrees F, heart rate 10), (98 degrees F, heart rate 70), or any of a trillion other different combinations of possible parameters. But in fact, living systems successfully restrict themselves to tiny fractions of this space – if you go too far away from (98 degrees F, heart rate 70), you die. So you have two probability distributions – the maximum-entropy one where you could have any combination of heart rate and body temperature, and the one your body is aiming for with a life-compatible combination of heart rate and body temperature. Whenever you have a system trying to convert one probability distribution into another probability distribution, you can think of it as doing Bayesian work and following free energy principles. So free energy seems to be something like just a formal explanation of how certain systems display goal-directed behavior, without having to bring in an anthropomorphic or teleological concept of “goal-directedness”.

Friston mentions many times that free energy is “almost tautological”, and one of the neuroscientists I talked to who claimed to half-understand it said it should be viewed more as an elegant way of looking at things than as a scientific theory per se. From the Alius interview:

> The free energy principle stands in stark distinction to things like predictive coding and the Bayesian brain hypothesis. This is because the free energy principle is what it is — a principle. Like Hamilton’s Principle of Stationary Action, it cannot be falsified. It cannot be disproven. In fact, there’s not much you can do with it, unless you ask whether measurable systems conform to the principle.

So we *haven’t* got a real-life version of Asimov’s psychohistory, is what you’re saying?

But also:

> The Bayesian brain hypothesis is a corollary of the free energy principle and is realized through processes like predictive coding or abductive inference under prior beliefs. However, the Bayesian brain is not the free energy principle, because both the Bayesian brain hypothesis and predictive coding are incomplete theories of how we infer states of affairs.

> This missing bit is the enactive compass of the free energy principle. In other words, the free energy principle is not just about making the best (Bayesian) sense of sensory impressions of what's "out there". It tries to understand how we sample the world and author our own sensations. Again, we come back to the woodlice and their scurrying — and an attempt to understand the imperatives behind this apparently purposeful sampling of the world. It is this enactive, embodied, extended, embedded, and encultured aspect that is lacking from the Bayesian brain and predictive coding theories; precisely because they do not consider entropy reduction […]
>
> In short, the free energy principle fully endorses the Bayesian brain hypothesis — but that's not the story. The only way you can change "the shape of things" — i.e., bound entropy production — is to act on the world. This is what distinguishes the free energy principle from predictive processing. In fact, we have now taken to referring to the free energy principle as "active inference", which seems closer to the mark and slightly less pretentious for non-mathematicians.

So maybe the free energy principle is the unification of predictive coding of internal models, with the "action in the world is just another form of prediction" thesis mentioned above? I guess I thought that was part of the standard predictive coding story, but maybe I'm wrong?

Overall, the best I can do here is this: the free energy principle seems like an attempt to unify perception, cognition, homeostasis, and action.

"Free energy" is a mathematical concept that represents the failure of some things to match other things they're supposed to be predicting.

The brain tries to minimize its free energy with respect to the world, ie minimize the difference between its models and reality. Sometimes it does that by updating its models of the world. Other times it does that by changing the world to better match its models.

Perception and cognition are both attempts to create accurate models that match the world, thus minimizing free energy.

Homeostasis and action are both attempts to make reality match mental models. Action tries to get the organism's external state to match a mental model. Homeostasis tries to get the organism's internal state to match a mental model. Since even bacteria are doing something homeostasis-like, all life shares the principle of being free energy minimizers.

So life isn't doing four things – perceiving, thinking, acting, and maintaining homeostasis. It's really just doing one thing – minimizing free energy – in four different ways – with the particular way it implements this in any given situation depending on which free energy minimization opportunities are most convenient. Or something.

This might be useful in some way? Or it might just be a cool philosophical way of looking at the world? Or maybe something in between? Or maybe a meaningless way of looking at the world? Or something? Somebody please help?

Discussion question for machine ethics researchers – if the free energy principle were right, would it disprove the orthogonality thesis? Might it be impossible to design a working brain with any goal besides free energy reduction?

Would anything – even a paperclip maximizer – have to start by minimizing uncertainty, and then add paperclip maximization in later as a hack? Would it change anything if it did?

SSC Journal Club: Friston On Computational Mood

A few months ago, I wrote Toward A Predictive Theory Of Depression, which used the predictive coding model of brain function to speculate about mood disorders and emotions. Emotions might be a tendency toward unusually high (or low) precision of predictions:

> Imagine the world's most successful entrepreneur. Every company they found becomes a multibillion-dollar success. Every stock they pick shoots up and never stops. Heck, even their personal life is like this. Every vacation they take ends out picture-perfect and creates memories that last a lifetime; every date they go on leads to passionate soul-burning love that never ends badly.
>
> And imagine your job is to advise this entrepreneur. The only advice worth giving would be "do more stuff". Clearly all the stuff they're doing works, so aim higher, work harder, run for President. Another way of saying this is "be more self-confident" – if they're doubting whether or not to start a new project, remind them that 100% of the things they've ever done have been successful, odds are pretty good this new one will too, and they should stop wasting their time second-guessing themselves.
>
> Now imagine the world's least successful entrepreneur. Every company they make flounders and dies. Every stock they pick crashes the next day. Their vacations always get rained-out, their dates always end up with the other person leaving halfway through and sticking them with the bill.

> What if your job is advising this guy? If they're thinking of starting a new company, your advice is "Be really careful – you should know it'll probably go badly". If they're thinking of going on a date, you should warn them against it unless they're really sure. A good global suggestion might be to aim lower, go for low-risk-low-reward steady payoffs, and wait on anything risky until they've figured themselves out a little bit more.
>
> Corlett, Frith and Fletcher linked mania to increased confidence. But mania looks a lot like being happy. And you're happy when you succeed a lot. And when you succeed a lot, maybe having increased confidence is the way to go. If happiness were a sort of global filter that affected all your thought processes and said "These are good times, you should press really hard to exploit your apparent excellence and not worry too much about risk", that would be pretty evolutionarily useful. Likewise, if sadness were a way of saying "Things are going pretty badly, maybe be less confident and don't start any new projects", that would be useful too.
>
> Depression isn't normal sadness. But if normal sadness lowers neural confidence a little, maybe depression is the pathological result of biological processes that lower neural confidence a lot. To give a total fake example which I'm not saying is what actually happens, if you run out of whatever neurotransmitter you use to signal high confidence, that would give you permanent pathological low confidence and might look like depression.

This would explain a lot about depression. It would explain why depressed people have such low motivation. It would explain why their movements are less forceful ("psychomotor retardation"). It would even explain why sense data are less distinct (depressed people literally see the world in washed out shades of grey). I thought this was plausible, but said I'd wait for real scientists to say the same thing before believing it too much.

What Is Mood: A Computational Perspective by Clark, Watson, and Friston – is real scientists saying the same thing. Sort of. With a lot more rigor. Let's look into it and see what they get.

> Recent theoretical arguments have converged on the idea that emotional states reflect changes in the uncertainty about the somatic consequences of action (Joffily & Coricelli, 2013; Wager et al. 2015; Seth & Friston, 2016). This uncertainty refers to the precision with which motor and physiological states can be predicted. In this setting, negative emotions contextualise events that induce expectations of unpredictability, while positive emotions refer to events that resolve uncertainty and confer a feeling of control (Barrett & Satpute, 2013; Gu et al. 2013). This ties emotional states to the resolution of uncertainty and, through the biophysical encoding of precision, to neuromodulation and cortical gain control (Brown & Friston, 2012).
>
> In summary, one can associate the valence of emotional stimuli with the precision of prior beliefs about the consequences of action. In this view, positively valenced brain states are necessarily associated with increases in the precision of predictions about the (controllable) future – or, more simply, predictable consequences of motor or autonomic behaviour. Conversely, negative emotions correspond to a loss of prior precision and a sense of helplessness and uncertainty about the consequences of action.

Here they’re saying that emotions – the day-to-day variation in whether we feel happy or sad – is meant to track what kind of environment we’re in. Is it a predictable environment that we should rush out to manipulate so we can harvest a big heap of utility? Or is it an unpredictable environment where we’re probably wrong about everything and should try to limit damage?

It’s not really clear from this quote, but later on they’re going to shift from happiness being “the world is predictable” to “the world is good”, which – sounds a lot more common-sensical. I think this has to do with Friston’s commitment to believing that uncertainty-resolution is the only drive, and every form of goodness is a sort of predictability in a way. See Monday’s post God Help Us, Let’s Try To Understand Friston On Free Energy – or don’t, for all the good it will do you.

> Any hierarchical inference relies on hyperpriors. These furnish higher level predictions of the likely value of lower level parameters. From the above, one can see that important parameters are the precisions of prediction errors at high and low levels of the hierarchy (i.e. prior and sensory precision). These precisions reflect the confidence we place in our prior beliefs relative to sensory evidence. If emotional states in the brain reflect the precision of prior beliefs about the consequences of action, then distinct neuronal populations must also encode hyperpriors. In other words, short-term fluctuations in precision (i.e. emotional fluctuations) will themselves be constrained by hyperpriors encoding their long-term average (i.e. mood).
>
> Here, we propose that mood corresponds to hyperpriors about emotional states, or confidence about the consequences of action. In other words, mood states

> reflect the prior expectation about precision that nuances (emotional) fluctuations in confidence or uncertainty. If emotion reflects interoceptive precision, and is biophysically encoded by neuromodulatory gain control, then this suggests that mood is neurobiologically encoded as the set-point of neuromodulator systems that determine synaptic gain control over principal cells reporting prediction errors at different levels of the interoceptive hierarchy. This set-point is the sensitivity of responses to prediction errors and has a profound and enduring effect on subsequent inference.

The traditional definition says that “mood is like climate, emotions are like weather”. I think they’re saying that mood – long-lasting states like being depressed or being a generally carefree person – are second-level priors about emotions, which themselves are first-level priors about actions.

So suppose you see a vaguely greenish piece of paper on the ground. If you’re happy, you have a prior for the world being good, and so you might be more likely to interpret it as possibly a dollar bill. And you have a prior for the world being exploitable, so you might be more likely to think you can reach down and take it and have an extra dollar. And if you do, and it really is a dollar bill, you might become happier, since you’ve gained a little evidence that your senses are trustworthy (you were right to perceive it as a dollar), the world is exploitable (your cunning plan to pick up the paper and gain $1 worked!), and you’re in the sort of high-reward environment where you should go off and do other exciting things.

On the other hand, if you’re sad, you have a prior for the world being bad, so you might expect it to be litter. You have a prior that you can’t really predict or affect the world, so it might not be worth bending down to pick it up – you might just end up disappointed. But if you did bend down to pick it up, and it did turn out to be a dollar bill, you might brighten up a little, just as the happy person would. You've gained a little bit of evidence that you're in a nice part of the world where good things happen to you, and that you can execute a simple plan like picking up a dollar bill to gain money.

A depressed person would have the same prior that the world is bad and the paper is probably just litter. But if perhaps she did pick up the dollar, and feel tempted to conclude that the world was good and she should feel happy, a higher-level prior would kick in: even when it seems like the world is good, that's wrong and you should ignore it. The world is never actually good. When good things happen that look like they should convince you that the world is good, those are just lies.

Friston et al bring up learned helplessness. Let's say you shock a rat a lot. In fact, let's say you're even more cruel, and you constantly give the rat apparent escapes, only to close them off at the last second and keep shocking it. You give the rat what look like food pellets, but they turn out to just be rocks painted to look like food. You eventually gaslight the hell out of the rat. Finally, you stop doing this, and you give the rat some actual food and a way out, and the rat just doesn't care. Yes, food and escape should be good things that make it feel lik the world is reward-filled and exploitable, but it's been let down so many times before that it assumes anything seemingly-good is a mirage.

Here's the picture they eventually draw:

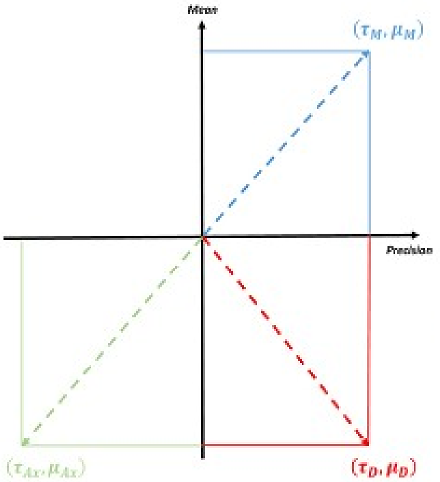


Fig. 1. The figure shows how mood can be conceptualised according to the expected precision (μ) and precision of precision (τ) in a two-dimensional space. Here, precision *per se* corresponds to the predictability of the (prosocial, affiliative and interoceptive) world – and the two dimensions correspond to hyperpriors over precision. It is proposed that pathological changes in mood occur in the extrema of this space, as highlighted. Depression occurs when an uncertain, unpredictable outcome is predicted with high precision (red lines) resulting in a chronic, self-maintaining negative emotional state that is resistant to revision. Mania (blue lines) is characterised by an equally high precision, but with the expectation of a predictable and controllable outcome – correspondingly the environment is chronically and inappropriately labelled as such. Anxiety (green lines) is an expected unpredictability but with low precision. As such, the individual engages in behaviour designed to resolve this uncertainty but which never does. D, depression; M, Mania; Ax, anxious depression.

Depression is a prediction of bad outcomes with high confidence. Mania is a prediction of good outcomes with high confidence. Anxiety (or “agitated depression”) is a prediction of bad outcomes with low confidence. There’s a blank space where it looks like there ought to be an extra emotion; maybe God will release it later as DLC.

Friston et al speculate that these hyperpriors over emotions can either be genetically encoded, or “learned” over very long periods of consistent stimuli. For example, if your childhood is unbearably terrible, that might be long enough to “burn in” a high-confidence hyperprior that the world is always bad.

(they don’t mention this, but if prediction and action are as linked as everyone always says, I wonder if this would explain why people with terrible childhoods are always mysteriously sabotaging themselves into have adulthoods that are terrible in the exact same way – eg someone with an abusive alcoholic father marrying an abusive alcoholic).

These hyperpriors can reach the level of a mood disorder when they become resistant to feedback. They present a couple of different arguments for how this might happen. In one, a depressed person doesn’t feel any positive emotions, since there's such a strong prior on everything being terrible that these never reach the level of plausibility. Since positive emotions are a useful tool for figuring out what makes you happy and urging you to do it, depressed people aren't motivated to make themselves happy, and so never end up contradicting their bias towards believing they're sad all the time. This fits *really* well with "behavioral activation", a common psychotherapy where therapists tell depressed people to just go out and do happy things whether they want to or not, and which often helps the depression resolve.

In another, all the brain's predictions are so low-precision that it can't even properly predict interoceptive sensations (the sensations received from organs, eg the heartbeat). Maybe it will think "I guess maybe my heart will beat right now", but it's not the sort of clear confident precision that really enters into its mental model. That means these interoceptive sensations are always predicted slightly incorrectly, and this keeps the brain feeling like it's sick and confused and the world is unpredictable.

They don't seem to mention this, but it also seems intuitively plausible that the strong prior on negativity could prevent the perception of positive factors directly. You see the piece of paper on the street, you think "the world is always terrible, so no way that's a dollar bill", you pass it by, and you miss an opportunity to feel lucky and give yourself a tiny bit of pleasure.

The rest of the paper is just a survey of some findings from biology and neuroscience that seem to support this, though they're not all very specific. For example, the HPA axis is dysregulated, which fits with predictive processing, but it also fits with everything else. The main part I found interesting was this:

> In healthy systems, mood should be affected by the valence of tightly controlled prediction errors. Recent animal work has shown that positive prediction errors (receiving more food than expected), show a strong positive correlation with dopaminergic change in the nucleus accumbens (Hart et al. 2014) with corresponding changes in functional brain activity in humans during a financial reward task (Rutledge et al. 2010). Similarly, it has been shown that signal change in the anterior insula is significantly related to the magnitude of prediction error (Bossaerts, 2010). The pharmacological manipulation of these networks was recently demonstrated where participants were given electric shocks (harms) in exchange for financial reward (gains), and offered the option of increasing the number of shocks in exchange for greater reward. It was shown that citalopram increased harm-aversion, while levodopa made individuals more likely to harm themselves than others (Crockett et al. 2015). This fits nicely with our notion that serotonin levels (and other neuromodulators) encode expectations about likely negative outcomes and encourage the fulfilment of these predictions through action (i.e. low levels promote behaviour with negative outcomes).

Focus on this sentence: “serotonin encodes expectations about likely negative outcomes and encourages the fulfilment of these predictions through action”. Also this one: “Low levels [of serotonin] promote behavior with negative outcomes”.

I don’t think I’m misunderstanding this – the authors cite some evidence that low serotonin causes self-harm, and yes, it certainly does. But what does it mean to have a system for promoting behavior with negative outcomes? Why have a neurotransmitter whose level corresponds to how much you should be trying to do negative-outcome behavior? Surely the answer is just “never do this”.

The only way I can make sense of this is through the paragraph above talking about the shocks-for-money game, where SSRIs decrease people’s willingness to get shocks. It sounds like maybe Friston et al are claiming that we have a “willingness to be harmed” lever so that we can calculate how willing we are to accept some levels of harm in exchange for a greater good. In that case, maybe self-harm is what happens when the “willingness to be harmed” lever is set so high that random noise, the chance of getting other people’s attention, or just passing the time presents some tiny reward, and your harm-for-reward tradeoff rate is so high that even that tiny reward is worth the harm.

More broadly, what should we think of this theory?

In retrospect, if you know Bayesian math, the idea of depression as a prior on bad outcomes seems pretty fricking obvious. I’m not even sure if it’s any different from the sort of stuff Aaron Beck was saying in the seventies. The big advance in this model is uniting “prior on bad outcomes” with “low precision of predictions / low neural confidence”. The low-precision part helps explain anergia, anhedonia, low motivation, psychomotor retardation, sensory washout, and probably (with a little more work) depression with psychotic features. Flipped around, it offers an explanation of psychomotor agitation, grandiosity, psychosis, and pareidolia in mania.

The only problem is that I still haven’t seen “prior on bad outcomes” and “low precision” really get unified. The authors seems to equivocate between “sadness means you’re in an unpredictable environment” and “sadness means you’re in a bad environment where everything sucks”. There is at least a little bit of work to add the hyperprior on top of the prior, so that at least we don’t get suspicious when we remember that depressed people are very confident in their depression. But it still seems like a world of low-precision predictions should be one where people just have no idea whether the paper in front of them is a dollar, not one where they’re really sure it isn’t. A world of high-precision predictions should look more like sitting in a bright room with a metronome, predicting each subsequent beat, rather than a world where everything is great and your life goes well. I’m not even sure this theory can explain why winning the lottery makes you happy rather than sad. It ought to make you think the world is really confusing and unpredictable (really? the thing you thought had a one in ten million chance happened?) – but in fact most lottery winners look pretty happy to me.

If this is confusing, at least it isn’t a *new* confusion. We know that a big part of the free energy research agenda is to try to unify desire-satisfaction with uncertainty-resolution, and claim that expectation and desire are (somehow, despite how it looks) the same thing. If we just assume that works, for the sake of argument, it allows this paper to be an impressive unification of several lines of research on mood disorder into a coherent and actionable whole.

Response To Current Affairs On Against Murderism

Nathan Robinson of *Current Affairs* has criticized my post Against Murderism. It's in the middle of a longer article, but I'll quote the relevant section:

> Arguments against identity politics and social justice are often amusingly ironic. They speak strongly in favor of dialogue and listening, without actually entering into a dialogue or doing much listening. For example, Scott Alexander, whose "mistake theory" concept Brooks quotes favorably, has criticized the overbroad use of the word "racism." He says that often, people attribute racist motives where there are none, and that in order to understand the causes of racially disparate outcomes we can't treat people as having some "dark irrational hatred which is their only terminal goal." This, he says, is as silly as if we treated anyone who caused another's death as a "murderist," i.e. someone with a strong belief in murder for murder's sake. Instead, people kill for a variety of reasons, often instrumental ones, and reducing the amount of death requires honestly assessing those reasons. And he shares Brooks' commitment to making every possible effort to understand people rather than demonizing them.
>
> A lot of the concerns of people who aren't like us will probably sound like nonsense…The solution is the same as it's always been: hard work, renewed commitment to liberal values, and a hefty dose of the Principle of Charity. Racism-as-murderism is the opposite. It's a powerful tool of dehumanization. It's not that other people have a different culture than you. It's not that

other people have different values than you. It’s not that other people have reasoned their way to different conclusions from you. And it’s not even that other people are honestly misinformed or ignorant, in a way that implies you might ever be honestly misinformed or ignorant about something. It’s that people who disagree with you are motivated by pure hatred, by an irrational mind-virus that causes them to reject every normal human value in favor of just wanting to hurt people who look different from them. This frees you from any obligation to do the hard work of trying to understand other people… You are right about everything, your enemies are inhuman monsters who desire only hatred and death, and the only “work” you have to do is complain on Twitter about how racist everyone else is.

All of which may seem sensible. Yet even as this passage asks people to be charitable and not caricature opposing beliefs, it is itself uncharitable and caricatures opposing beliefs. I have moved in progressive circles for a long time, and been in academic sociology where supposedly a lot of this stuff emanates, and I don’t think many progressive people believe that people being “racist” means they are “inhuman monsters who desire only hatred and death.” In fact, one of the central points made by contemporary progressives is that racism isn’t about individual “hate,” but that it’s a set of subconscious attitudes that almost everyone possesses to one degree or another. The whole “check your privilege” idea (a phrase I don’t like) is that people are “honestly misinformed and ignorant,” that they are oblivious to the various structural disadvantages that other people face. The argument being made is that thinking of racism as emanating from “hate-filled monsters” is a mistake: it’s something we all have to face up to our complicity in, regardless of how decent

> and well-intentioned a person we may be. Anyone who has read a introductory book on critical race theory or the sociology of race would know this. (In Matthew Desmond and Mustafa Emirbayer's Race in America, the #1 entry on its "list of fallacies" about racism is the misconception that racism is about hateful and monstrous individuals.) But those who demand charity and openness toward their own beliefs are rarely willing to display it toward the terrible irrational campus left.

At the risk of going up a meta-level in the chain of accusing people of not engaging with other people, I don't think this is engaging with me.

Not only was I aware of the counterargument Nathan cites, I included 2,070 words in the post (marked as Part II, please read it) bringing it up, admitting that it was the strongest argument against my position, and explaining why I disagreed. In a section on various definitions given for racism, I called a similar idea the "Definition By Consequences". *Current Affairs* accuses me of not realizing that "anyone who has read a introductory book on critical race theory or the sociology of race would know this", and thinks he has shown my unwillingness to read the other side. But as I said in the original post (archive.is link to prove I didn't edit it):

> I know that Definition By Consequences is the really sophisticated one, the ones that scholars in the area are most likely to unite around. But I also think it's uniquely bad at capturing the way anyone uses the word "racism" in real life. Let me give four examples…

Followed by several thousand words explaining how that worked, why it didn't match real-world usage, and why I thought this counterargument against my point was wrong.

Followed by another few thousand words about “Definition By Motive”, which captures other parts of what Robinson means by “subconscious bias”.

I then go on to explain a lot of cases of racism that don’t make sense as “subconscious bias”. All of the six examples in Part I are like this. So is the daycare owner in Part IV. Even the title of the post is a reference to this – we don’t treat murder as the result of subconscious pro-murder biases. My whole point is that the official scholarly definition – itself an amalgam of “subconscious bias” and “structural consequences” (what I’m calling “Definition By Motives” and “Definition By Consequences”) – just doesn’t explain any of the real-world discourse.

I acknowledged in the post that everyone, when challenged, goes back to the official definition. But I said I didn’t take that any more seriously than the fact that every anti-gay activist, when challenged, goes back to saying that they “love the sinner but hate the sin”. I described this kind of “have a great-sounding explanation on hand that doesn’t match your behavior” strategy as a “motte-and-bailey”, something I’ve written several posts on before. Then I cited various examples of real-world leftists saying that racists couldn’t be negotiated with, couldn’t participate in civil society, and (in one case where I included a screenshot of the conversation) should be shot.

I try to believe people’s lived experience, and Nathan says he “moved in progressive circles” for a long time and almost never sees them talking about racists as monsters. I suppose that should be some evidence. On the other hand, I read Nathan’s magazine. This week, he published a piece which said that:

> Anti-immigration activists of Coulter’s ilk are not people passionately advocating for what they believe is the most sensible and humane model of human governance. They are monsters who literally believe that non-American lives, especially non-white non-American lives, are worth less than dirt.

This is a pretty clear example of the trope, and *Current Affairs* themselves published it literally four days ago. I don’t know how to square this with their editor saying he doesn’t see this kind of stuff. Perhaps he thinks we’re referring to different groups? But anti-immigration activists are exactly the type of people I meant, and I mention them explicitly in the article.

Also in *Current Affairs*, from a few months back – a claim that:

> We now consider racists to be monsters as foreign and grotesque as Godzilla.

I understand magazines aren’t monolithic and sometimes publish disagreeing voices. But this suggests it’s well understood, even among Leftists, that the racist-as-monster trope is a thing. I don’t think it’s fair to dismiss it as just my ignorance and failure to engage with progressives. The reason I claim people believe this kind of thing is exactly because *I do read your magazines where you say it*.

And, unfortunately, this is the entire point of my Against Murderism article. *In real life*, leftists will say that racists are inhuman monsters. Then *as soon as someone points out this is bad*, they will retreat to saying that obviously nobody believes that, racism is just a collection of structural subconscious privilege discrimination IAT [insert Avenue Q “Everyone’s A Little Bit Racist” song here].

I don’t understand how someone could think they’ve disproven my article by ignoring that whole argument, then saying “obviously nobody believes that, racism is just a collection of subconscious privilege discrimination IAT [insert Avenue Q “Everyone’s A Little Bit Racist” song here]”, while also spreading the message that racists are monsters as loudly as possible in a widely-read magazine.

And I get especially annoyed when it’s framed as an accusation that I don’t like to debate other people, listen to their claims, or engage with their ideas.

Nathan writes that “shockingly, the people who most loudly call for empathy and dialogue are the least willing to engage in genuine empathetic dialogue…”, and uses me as an example. I can only say in my defense that last month, I sent Nathan an email saying that I thought it would be productive to engage in dialogue with each other in a way “where instead of trying to disagree publicly, we’re trying to come to agreement privately, then present the results of that agreement”. I offered to do this with him on a topic of his choice. He wrote back saying he didn’t have enough time, which is fine. But when he then publishes an article in a national magazine announcing that I am a hypocrite because I refuse to dialogue with my political opponents, I feel pretty betrayed.

I have always been interested (modulo time restrictions) in dialoguing with people I disagree with. The offer to Nathan is still open. Otherwise I politely request to be treated better than this.

The Dark Rule Utilitarian Argument For Science Piracy

I sometimes advertise sci-hub.tw – the Kazakhstani pirate site that lets you get scientific papers for free. It's clearly illegal in the US. But is it unethical? I can think of two strong arguments that it might be:

First, we have intellectual property rights to encourage the production of intellectual goods. If everyone downloaded *Black Panther*, then Marvel wouldn't get any money, the movie industry would collapse, and we would never get *Black Panther 2*, *Black Panther Vs. Batman Vs. Superman*, *A Very Black Panther Christmas*, *Black Panther 3000: Help, We Have No Idea How To Create Original Movies Anymore*, and all the other sequels and spinoffs we await with a resignation born of inevitability. This is sort of a pop-Kantian/rule-utilitarian argument: if everyone were to act as I did, our actions would be self-defeating. Or we can reframe it as a coordination problem: we're defecting against the institutions necessary to support movies existing at all, and free-loading off our moral betters.

Second, and related, the laws have their own moral force that has to be respected. With all our celebration of civil disobedience, we forget that *in general* people should feel some obligation to obey laws even if they disagree with them. This is the force that keeps libertarians from evading taxes, vegetarians from sabotaging meat markets, and doctors from giving you much better medications than the ones you consent to – even when they think they can get away with it. Civil disobedience can be justifiable – see here for more discussion – but surely it should require some truly important cause, probably above the level of “I really want to watch *Black Panther*, but it costs $11.99 in theaters”.

(I admit I sometimes violate this principle , because I – like most people – am not perfectly moral.)

But I can also think of an argument why Sci-Hub isn’t unethical.

The reason I don’t pirate *Black Panther* is because, if everyone pirated movies, it would destroy the movie industry, and we would never get *Lego Black Panther IV: Lego Black Panther Vs. The Frowny Emoji*, and that would make people sad.

But if everyone pirated scientific papers, it would destroy Elsevier et al, and *that would be frickin’ fantastic*.

As far as I can tell, the movie industry is capitalism working as it should. No one animator can make a major motion picture, so institutions like Marvel Corporation exist to solve the coordination problem and bring them together. Marvel Corporation is probably terrible in various ways, but it’s unclear we have the social technology to create non-terrible corporations right now, so unless we’re communists we accept it as the price to pay for a semi-functional industry. Then some market-rate percent of the gains flow down to the actors and videographers and so on. If you destroyed this system, you wouldn’t usher in a golden age of independent superhero movies. You would just stop getting *Black Panther*.

The scientific journal industry is some kind of weird rent-seeking abomination which doesn’t seem to add much real value. I don’t have space to make the full “journals are not helpful” argument here, but see eg this article, Elsevier’s profit margins, and the relative success of alternative models like arXiv. See *Inadequate Equilibria* for the discussion of how this might have happened. The short and wildly insufficient summary is that it looks like we backed ourselves into an equilibrium where eg tenure committees consider journals the sole arbiter of scientific merit, anyone who unilaterally tries to defect from this equilibrium is (reasonably) suspected of not having enough merit to make it the usual way, and coordination is hard so we can’t make everyone defect at the same time.

Thus Dark Rule Utilitarianism: “If I did this, everyone would do it. If everyone did it, our institutions would collapse. But I hate our institutions. Therefore…”

I think this fully addresses the first argument against science piracy. But what about the second? Sure, I don’t like the institution of scientific gatekeepers, but anarcho-communists don’t like the institution of private property. If I steal scientific papers to destroy the journal system, doesn’t universalizing that decision process lead to anarcho-communists stealing cars to destroy capitalism? Shouldn’t “civil disobedience” be reserved for the most important things, like ending segregation or resisting the Nazis, rather than endorsed as something anyone can do when they feel like destroying something?

This kind of thing leaves me hopelessly confused between different levels. It’s much worse than free speech, where all you’ve got to keep track of is whether you agree with what someone says vs. will defend their right to say it. But an important starting point is that endorsing “civil disobedience is sometimes okay” doesn’t lead to a world where anarcho-communists steal cars and nobody stops them. It leads to a world where *there is no overarching moral principle* preventing anarcho-communists from seizing cars, and where *we have to do politics* to decide whether they get arrested. In practice, the politics would end up with the car thieves arrested, because stealing cars is pretty conspicuous and nobody likes car thieves.

Isn’t this just grounding morality in power? That is, aren’t we going from the clarity and fairness of “everyone must follow the law” to a more problematic “everyone must follow the law, except people clever enough to avoid getting caught and powerful enough to get away with civil disobedience?” Well, yeah. But from an institution design perspective, everything bottoms out in power eventually. All we’re doing here is replacing one form of power (the formal power possessed by law-makers) with another form of power (the informal powers of stealth and/or popularity that allow people to get away with civil disobedience). These two forms of power have different advantages and are possessed by different groups. The formal power is nice because it’s transparent and democratic and probably bound by rules like the Bill of Rights, but it also tends to concentrate among elites and be susceptible to tyranny. The informal power is nice because it’s inherently libertarian and democratic, but it’s also illegible and susceptible to being used by demagogues and populists.

So, a metaphor: imagine a world with a magic artifact at the North Pole which makes it literally impossible to violate laws. The countries of the far north are infinitely orderly with no need for police at all. Go further south and the strength of the artifact decreases, until you’re at the edge of the Arctic Circle and it might be possible to violate a very minor law if your life was in danger. By the time you’re at the Equator, any kind of strong urge lets you violate most laws, and by the Tropic of Capricorn you can violate all but the most sacred laws with only a slight feeling of resistance. Finally you reach the nations of the South Pole, where the laws are enforced by nothing but a policeman’s gun.

Where would you want to live in such a world? It’s a hard question – I can imagine pretty much anything happening in this kind of scenario. But if I had to choose, I think I would take up residence somewhere around the latitude of California. I would want the laws to carry some force beyond just the barrel of a gun – a high trust society with consistent institutions is really important, and the more people follow the law without being watched the less incentive there is to create a police state.

But I also wouldn’t want to live exactly at the North Pole. And when I try to figure out why, I think it’s that civil disobedience is the acid that dissolves inadequate equilibria. Equilibria are inadequate relative to some set of rules; if you’re allowed to break the rules, they can become adequate again. Under this model, civil disobedience isn’t a secret weapon to save up for extreme cases like desegregation, it’s part of the search process we use to get better institutions.

If the artifact is a metaphor for the moral law, then my choice to live outside the North Pole suggests that I can consistently defy unjust laws *a little*, even if my decision will be universalized. I should expect some problems – groups I don’t like will use civil disobedience to promote causes I abhor, and the state will be less orderly and peaceful than it could be – but overall everyone will end up being better off. This doesn’t mean I have to support those groups or even excuse their criminality – part of the politics that decides the result is me expressing that they are bad and need to be punished – it just means that, given the chance to magically make all civil disobedience impossible in a way that applies equally to me and my enemies – I would reject it, or take it at some less-than-maximal value.

So this is my argument that Sci-Hub can be ethical. Universalized it would destroy the system – but the system is bad and needs to be destroyed. And although this would break the law, a very slight amount of law-breaking might be a beneficial solution to inadequate equilibria – one that could be endorsed even when universalized.

Navigating And/Or Avoiding The Inpatient Mental Health System

Apology and disclaimer

This is in response to questions I get about how to interact (or not interact) with the inpatient mental health system and involuntary commitment. The table of contents is:

1. How can I get outpatient mental health care without much risk of being involuntarily committed to a hospital?
2: How can I get mental health care at a hospital ER without much risk of being involuntarily committed?
3. I would like to get voluntarily committed to a hospital. How can I do that?
4. I am seeking inpatient treatment. How can I make sure that everyone knows I am there voluntarily, and that I don't get shifted to involuntary status?
5. How can I decide which psychiatric hospital to go to?
6. I am in a psychiatric hospital. How can I make this experience as comfortable as possible?
7. I am in a psychiatric hospital and not happy about it and I want to get out as quickly as possible. What should I do?
8. I am in the psychiatric hospital and I think I am being mistreated. What can I do?
9. I think my friend/family member is in the psychiatric hospital, but nobody will tell me anything.
10. My friend/family member is in the psychiatric hospital and wants to get out as quickly as possible. How can I help them?
11. How will I pay for all of this?
12. I have a friend/family member who really needs psychiatric treatment, but refuses to get it. What can I do?

I am a psychiatrist, which both means I have some useful experience here, and makes it hard for people trying to avoid the system to trust me. Anything written with too much honesty risks degenerating into "here's how to cheat the system so nobody will know you're about to commit suicide". But anything written with too little honesty risks degenerating into some variation of "trust the wise benevolent doctors to do what is best for you". This is an impossible edge to balance on, and I am sure I fail at one point or another.

But my first excuse is that if somebody doesn't understand how the commitment system works, they're not going to innocently blunder into spilling their guts. They're just going to *never go to the psychiatrist at all*. If someone wants to avoid ending up in the hospital but doesn't know how, it's not like they're stuck doing everything we want. They can just *lie about everything*. Or they can just *never go to the psychiatrist at all*. If they understand a little bit about how the system works, they can at least lie strategically, in the one place where they have to lie, while cooperating 99% of the way.

And my second excuse is that in the end, *this is not an adversarial enterprise*. Psychiatrists commit people because they're scared. They're scared because they can't predict what the patient is going to do – and on another level, they're scared because they might get sued if they don't follow the rules. If patients who aren't going to hurt themselves know how to explain that they aren't going to hurt themselves in a way that reassures their psychiatrist, and in a way that doesn't leave their psychiatrist legally liable for not committing them, then everybody can be more comfortable and get on with the hard work of actual treatment.

This guide applies to adult mental health care only. Child/adolescent mental health care is totally different and I don’t know anything about it. I have only worked in two states, it might be a bit different in other states, and it is definitely a lot different outside the US. Nothing in here is official medical advice. Follow it at your own risk. Please don’t use this to avoid psychiatric care which you actually need. All of this will be wrong in certain situations; when in doubt, trust your intuition.

1: How can I get outpatient mental health care without much risk of being involuntarily committed to a hospital?

Mental health care is divided into inpatient and outpatient settings. Inpatient care means it’s in a hospital, voluntary or otherwise. Outpatient care is your local doctor’s office, or psychiatrist’s office, or therapist’s office.

If you go to a hospital for mental health reasons, your risk of getting involuntarily committed is relatively high – see below for more. If you go to an outpatient provider, your risk is much lower.

In theory, the outpatient system is supposed to provide voluntary treatment, with risk of involuntary commitment only in certain very clearly delineated situations that you can understand and avoid. Each state’s laws are slightly different (and I can’t say anything about non-US countries), but they tend to allow involuntary commitment only in cases of immediate risk of hurting yourself, hurting someone else, or being so psychotic that you could plausibly hurt someone by accident (eg you jump out of a window because you think you can fly).

The key word is “immediate”. If you just have occasional thoughts about suicide, or you have some limited hallucinations but remain grounded in reality, *according to the law* this is not enough to involuntarily commit you.

In practice, not every mental health professional knows the laws or interprets them the same way, so they can just commit you anyway. The check on this is supposed to be that you can sue them when you get out of the hospital, but almost nobody bothers to do this, and judges and juries usually find in favor of the mental health professional.

So the law isn’t as much protection as it probably should be. In reality your best protection is to only open up to competent people whom you trust, and to frame what’s going on in a way that doesn’t scare them unnecessarily.

Don’t joke about committing suicide. Don’t bring up occasional stray suicidal thoughts if they don’t matter. Don’t say something like “I think about suicide sometimes, but doesn’t everyone?”, because your psychiatrist will have heard the last ten people answer “No, of course I never think about suicide”, and they will not be impressed with your claim about the human condition. Assume that any time you mention suicide, there’s a tiny but real chance of getting committed. If you are actually suicidal, take that chance in order to get help. Otherwise, this is really not the time to bring it up. If you wouldn’t offhandedly chat about terrorism with an airport security guard, don’t offhandedly chat about suicide with a psychiatrist.

(none of this applies to competent psychiatrists whom you trust, but award this status only after many positive experiences over a long-term relationship)

If your psychiatrist asks you outright if you ever have suicidal thoughts, well, tough call. If you don't, then say you don't. If you *mostly* don't but you are some sort of chronically indecisive person who has trouble giving a straight answer to a question, now is the time to suppress that tendency and just say that you don't. If you do, but you would never commit suicide and it's not a big part of why you're seeing them and you don't mind lying, you can probably just say you don't. If you do, and it's important, and you don't want to lie about it, then make sure to be very specific about how limited your thoughts are (eg: "I only thought that way once, three years ago) and to add as many of these as are true:

1. "Of course I would never go through with it, but sometimes I think about…"
2. "I love my friend/family member/partner/pet too much to ever go through with it."
3. "I don't have any plans for how I would do it."
4. "I'm [religion], and we believe that God doesn't want us to commit suicide."
5. "I've been thinking about it for [long time], but the thoughts haven't gotten any worse lately."

The same applies to hallucinations and other signs of psychosis. Most people have very minor random hallucinations as they are going to sleep. Most people hear their own thoughts as silent "voices" in their head at least some of the time. Most people who take hallucinogenic drugs will hallucinate. You don't need to bring these up when someone asks you about hallucinations. If you actually have some troubling psychotic symptoms, then mention them, but add as many of these as are true:

1. "Of course, I know these aren't really real."
2. "These have been going on for a while and aren't any

worse lately.”
3. “I would never listen to anything the voices say.”
4. “I only get that way when I’m on drugs / really tired / under a lot of stress.”

If you do all of these things, your chance of getting involuntarily committed to a psychiatric hospital by an outpatient provider is probably one percent or less, unless you’re really really sick.

Notice the words “by an outpatient provider” here. None of this applies if you are in a hospital (eg with pneumonia). If you are in a hospital, be extra careful about this to the point of paranoia. Unless you’re really worried that you might go through with suicide, be careful about mentioning it the hospital. Get your pneumonia or whatever treated, and then go out of the hospital, find a competent outpatient psychiatrist whom you trust, and open up about your issues to *them*. If you decide to open up to the nurse-assistant giving you a three question psychiatric screen in the pneumonia ward, you may end up on a psychiatric unit regardless of how careful you are, because hospitals don’t take chances.

2: How can I get mental health care at a hospital ER without much risk of being involuntarily committed?

Hospital ERs are not set up to provide psychiatric help to random people. They are set up to evaluate people and decide if it’s a real emergency. If it is, you will be committed to an inpatient unit. If it isn’t, they will tell you to see an outpatient psychiatrist, and you will be back at the beginning except with an extra $5000 bill to pay.

This is not true 100% of the time, and you can take your chances if you want. In particular, if you have extreme anxiety, sometimes they can give you enough fast-acting anti-anxiety medication to calm you down and last you until you can see an outpatient psychiatrist. But going to a hospital ER for any mental-health-related reason other than expecting to get admitted to a hospital psychiatric unit should be a last resort.

3. I would voluntarily like to get committed to a hospital. How can I do that?

If you have a competent outpatient psychiatrist whom you trust, call them up and tell them what's going on. If they have connections at a local hospital, they may be able to get you directly admitted, which will save you a lot of time and suffering.

Otherwise, you will have to go to a hospital ER. Be prepared for this to be extremely unpleasant. It may take up to 24 hours of sitting in the ER before a psychiatrist can see you. You will probably get examined by nurses, medical students, non-psychiatrist doctors, etc, and each time you will think "Finally! I am getting evaluated and I can get out of this ER!" but you will be wrong. Although there will probably be some crappy food and drink available, there may not be much in the way of entertainment, quiet, or privacy. Do yourself a favor and bring a book or game or something. You may not be allowed to keep your cell phone or laptop or other metal object (more on this later). If family or friends are willing to help, have them come along – if only so they can go out and bring you back real food when you get hungry.

Once you set foot in an ER and mention the word "psychiatry", you should be prepared for someone to tell you that you're not allowed to leave until the evaluation is complete. Maybe no one will tell you this, and you can try to leave, and it'll be fine. But you should be prepared for it not to work.

After many trials and tribulations, you will be examined by a psychiatrist, who will decide whether or not to accept you to the psychiatric unit. You are not guaranteed admission to the unit just because you want it. You might be turned down if the psychiatrist thinks you aren't sick enough to need it, or if your insurance refuses to pay for it. Insurance companies are very reluctant to pay for hospitalizations unless there is a clear risk involved, so explain what the risk is.

The only thing that (almost) always works is mentioning suicide. If you say you're suicidal, you will get admitted. If you want to be sure, do the opposite of everything above. Stress that you *are* suicidal. Stress that it's *not* just the occasional fleeting thought, but actually something that you might really go ahead with. If you have a plan, share it.

If you're not suicidal, expect to have to argue. Talk about what you've already tried and why it didn't work. Talk about all the damage your mental illness has caused in your life. If there's any chance you might snap and do something horrible – hurt someone, hurt yourself, have some kind of spectacular breakdown – play it up. If you have to, say something vague like "I don't know what I would do if I couldn't get help". Be ready for this not to work, and for the psychiatrist evaluating you to recommend you go to an outpatient psychiatrist.

If you really want help beyond the level of outpatient treatment, but your insurance company won't budge, ask about a partial hospital program. This is something where you go to a hospital-like environment from 9 to 5 for a few weeks, seeing doctors and getting therapy and classes, but you’re not involuntarily committed and you go home at night. Sometimes insurance companies will be willing to do this as a compromise if you are not suicidal.

4. I am seeking inpatient treatment. How can I make sure that everyone knows I am there voluntarily, and that I don’t get shifted to involuntary status?

I want to be really clear on this: in your head, there might be a huge difference between voluntary and involuntary hospitalization. In your doctor’s head, and in the legal system, these are two very slightly different sets of paperwork with tiny differences between them.

It works like this, with slight variation from state to state: involuntary patients are usually in the hospital for a few days while the doctors evaluate them. If at the end of those few days the doctors decide the patient is safe, they’ll discharge them. If, at the end of those few days, the doctors decide the patient is dangerous, the doctors will file for a hearing before a judge, which will take about a week. The patient will stay in the hospital for that week. 99% of the time the judge will side with the doctors, and the patient will stay until the doctors decide they are safe, usually another week or two.

Voluntary patients are technically allowed to leave whenever, but they have to do this by filing a form saying they want to. Once they file that form, their doctors may keep them in the hospital for a few more days while they decide whether they want to accept the form or challenge it. If they want to challenge it, they will file for a hearing before a judge, which will take about a week. The patient will stay in the hospital for that week. 99% of the time the judge will side with the doctors, and the patient will stay until the doctors decide they are safe, usually another week or two.

You may notice that in both cases, the doctors can keep the patient for a few days, plus however long it takes to have a hearing, plus however long the judge gives them after a hearing. So what’s the difference between voluntary and involuntary hospitalization? Pride, I guess, plus a small percent of cases where the doctors just shrug and say “whatever” when the voluntary patient tries to leave.

Some decent fraction of the time, patients who intended to get voluntarily hospitalized end up involuntarily hospitalized for inscrutable bureaucratic reasons. The one I’m most familiar with is the ambulance ride: suppose the hospital you’re in doesn’t have any psychiatric beds available and wants to send you to the hospital down the road. For inscrutable bureaucratic reasons, they have to send you by ambulance. And for inscrutable bureaucratic reasons, any psychiatric patient transferred by ambulance has to be involuntary. Your doctors don’t care about this, because they know that there is no practical difference between voluntary and involuntary – but if you are still trying to maintain your pride, this might come as kind of a shock.

Some other decent fraction of the time, patients who ought to be involuntarily hospitalized end up voluntarily hospitalized for inscrutable bureaucratic reasons. The one I’m most familiar with is doctors asking patients whom they are committing against their will to sign a voluntary form, ie “Agree to come voluntarily, or else I will commit you involuntarily”. This sounds super Orwellian, but it really is done with the patient’s best interest at heart. Involuntary commitments usually leave some kind of court record, which people can find if they’re searching your name for eg a background check – which could come up anywhere from applying for a job, to trying to buy a gun. Voluntary commitments usually don’t cause this problem. Even though nobody feels very warmly to the psychiatrist telling them to sign voluntarily or else, that psychiatrist is right and you should suck it up and sign the voluntary form.

If given a choice, you should sign voluntary, if only for the background-check reason above. But don’t count on getting the choice, and don’t get too attached to the illusion that it really matters in some deep way.

5. How can I decide which psychiatric hospital to go to?

If it’s an emergency, the answer is “whichever one is closest” or even “whichever one the ambulance you should call right now takes you to.”

If you have a little more leeway, and you have a competent outpatient psychiatrist whom you trust, ask them which one to go to. They will probably be familiar with the local terrain and be able to give you good advice.

If you live in a big city with wealthier and poorer areas, and it’s all the same to your insurance company, try to go to a hospital in the wealthier area. Not only do wealthier people always get nicer things, but – and sorry if this is politically incorrect – you would rather be locked up for a week with the sorts of people who end up in wealthy-area psychiatric hospitals than with the sorts of people who end up in poor-area psychiatric hospitals.

US News & World Report ranks the best psychiatric hospitals. They’re mostly looking at doctor prestige, but I would guess this correlates with other factors patients want in a hospital. If you’re really prestigious you have a lot of money and a lot of eyes watching you, and that probably helps. I suspect teaching hospitals are also good, for the same reason. But these are just guesses.

If you have no other way of figuring this out, you can try looking at Psych Ward Reviews. This site is underused and suffers from the expected bias – you only write about somewhere if you don't like it – but it's better than nothing.

Keep in mind that sometimes hospitals will be full, and they will send you to a different hospital instead, and you will not have any say in this.

6. I am in a psychiatric hospital. How can I make this experience as comfortable as possible?

When you go to the hospital ER to get admitted, bring a bag of stuff with you. This should include clothing, fun things to do like books, earplugs, snacks you like, and phone numbers for people you might want to contact.

Keep in mind that you will not be allowed to have anything that could be used as a weapon, for a definition of "could be used as a weapon" which is clearly aimed at MacGyver-level masterminds who can create a railgun out of three paperclips and a stick of gum. The same goes for anything that could be used as a suicide method. This means for example no laced shoes, pillowcases, scarves, and a bunch of other things you will not expect. Basically, bring stuff to the hospital, but expect a decent chance it won't be allowed in.

Metal objects, including laptops, cell phones, mp3 players, etc, will never be allowed in. These will be taken from you and put in a locker during your stay. If for some reason you have to transfer hospitals during your stay, these things always somehow get lost. Your best bet is to bring a friend with you to the ER, and have them take your cell phone and other valuables.

If you forget to bring a bag of stuff, or if you were committed involuntarily and unexpectedly and didn’t get a chance, call a friend or family member and ask them to bring you your stuff.

7. I am in a psychiatric hospital and not happy about it and I want to get out as quickly as possible. What should I do?

Good news: average stays for psychiatric hospitals have been decreasing for decades, and are now usually a week or less. I did a study on the hospital I worked in and came up with an median stay of 5.9 days, and remember that there are a lot of really sick people bringing up those numbers.

(there are a few states that have laws centered around the number “three days”, but there are also a lot of states that don’t. For some reason the “three days” number has leaked into the general consciousness and everyone expects that to be how long they stay in the hospital. Don’t necessarily expect to get out of the hospital in exactly three days, but do expect it will be closer to 5.9 days than to weeks or months.)

Even better news: contrary to rumor, psychiatrists rarely have a financial incentive to keep people hospitalized. In fact, most hospitals and insurances now encourage quick “turnover” to “open up beds” for the next group of needy patients, and doctors can get bonuses for getting people out as quickly as possible. This should worry everyone else in the hospital who’s getting treated for pneumonia or whatever, but from the perspective of a psychiatric patient who wants to leave quickly it’s pretty good.

If you have a good doctor, you should trust their judgment and do what they say. But if you have a bad doctor, then the only thing you can count on is that they will respond to incentives. Their incentive to get you out quickly is the hospital administrators and insurance companies breathing down their neck. Their incentive to keep you longer is that if you get out of the hospital and ever do anything bad, they can get sued for “missing the signs”. So their goal is to do a token amount of work that proves they evaluated you properly so nothing that happens later is their fault.

That means they’ll keep you for some standard time interval, traditionally (though not always) three days, just so they can say they “monitored” you. If you seem unusually scary in some way, they might monitor you a little longer, up to a week or two. Your chances of successfully convincing them not to do this are essentially nil. Imagine you kill someone a few weeks after leaving the hospital, and during the trial the prosecutor says “The patient was taken to St. Elsewhere Hospital for evaluation of mental status, but discharged early, because he said he didn’t want to have to sit around and be evaluated for the usual amount of time, and his doctor thought this was a reasonable request.” Your doctor is *definitely* imagining this scenario.

Instead of pleading with your doctors to let you go early, just do everything right. Have meals at mealtime. Go to groups at group time. Groom yourself, not just because you look saner when you’re well-groomed, but because there will actually be nurses monitoring your grooming status and reporting it to the psychiatrists making release decisions. When people tell you things you should do after leaving the hospital, agree that you will definitely do them. If people ask you questions, give reassuring-sounding answers.

For this last one – don’t contradict evidence against you, don’t accuse other people of lying, just downplay whatever you can downplay, admit to what the doctors already believe, and make it sound like things have gotten better. For example, if you were found lying face-down with an empty bottle of pills next to you, don’t say “I didn’t attempt suicide, I just tripped and the pills fell into my mouth!” (I have seriously had patients try this one on me). Don’t say “It was my girlfriend’s fault, she drove me to do it!” Just say something like “That was a really bad night for me, and I don’t remember exactly what happened, but now I’m feeling a lot more hopeful, and I think that was a mistake.”

Don’t overdo it. Nothing is more annoying than the person who’s like “The twenty minutes I’ve been talking with you so far have turned my life around, and now I realize how wrong I was to reject God’s beautiful gift of existence, and am overflowing with abounding joy at the prospect of getting to go back into the world and truly confront my problems with the help of my loving family and…” Just be like “Yeah, things were rough, but I feel a little better now.”

Most important, *take the damn drugs*.

Yes, I know that some psychiatric drugs are unpleasant or addictive or dangerous or make you feel miserable. I’m not challenging your decision not to want to be on them. But take the damn drugs *while you are in the hospital*, for 5.9 days. Then, when they let you out, decide if you still want to continue. I guarantee you this will be easier for you, for your psychiatrist, and for the various judges and lawyers involved. The alternative is that you refuse to take the drugs, somebody has to set up a court hearing to get an involuntary treatment order, you have to sit in the hospital for weeks while the legal system gets its act together, the psychiatrists finally get the order and drug you against your will, and *then* after however many weeks or months, you get released from the hospital and stop taking the drugs.

If you have a good doctor whom you trust, then talk to them about the drugs and make a decision together. Let them know if there are any side effects. If a drug isn’t working for you, tell them, so they can switch it. Be honest, and willing to stand up for yourself, but also open-minded and ready to listen.

But if you have a bad doctor, *just take the damn drugs*. Bring up side effects, mention anything that’s intolerable, but when – like bad doctors everywhere – they ignore you, just take the damn drugs. Then, when you get out of the hospital, go to a competent outpatient psychiatrist whom you trust, tell them the drugs aren’t right for you, and talk it over with them until you come up with a better plan.

This is a good general principle for everything: agree to whatever people ask you while you’re in the hospital, talk to a competent outpatient psychiatrist whom you trust once you get out, and decide which things to stick to. I remember working with a doctor who wanted to discharge his patient to some kind of outpatient drug rehab. The patient refused to go, so the doctor wouldn’t discharge her, and they were in a stalemate over it for *weeks*, and the whole time the patient was tearfully begging the doctor to release her. I cannot tell you how much willpower it took not to sneak into the patient’s room and yell at her “JUST AGREE TO GO TO THE REHAB AND THEN DON’T DO IT, YOU IDIOT”. I mean, I am as in favor of Truth as everyone else, but I don’t even think her doctor *cared* if she went to the rehab or not. He just wanted to be able to document “Patient agreed to go to rehab”, so that when she started taking drugs again, he would have ironclad court-admissable evidence that it wasn’t his fault.

Finally, your doctors will be very interested in “discharge planning”, ie making sure you have somewhere safe to be after you leave the hospital. They may not be willing to believe you about this. So get a family member (best) or friend (second-best) on your side. Have them agree to tell the doctors that they will watch over you after you leave, make sure you take your medication, make sure you get to your follow-up outpatient psychiatrist appointments, make sure you don't take any illegal drugs. Your best bet for this is your mother – psychiatrists love mothers. Tell your doctors "I talked to my mother, she's really concerned about my condition, she says that I can stay with her after I leave and she's going to watch me really closely and make sure I'm okay". Only say this if it's true, because your doctors will call your mother and make sure of it. But if you can make this work, this is really helpful.

Even if all of this works, it's just going to get you out of the hospital in a bit less than 5.9 days instead of a bit more than 5.9 days. There's no good way to get out instantly. Sorry.

8. I am in the psychiatric hospital and I think I am being mistreated. What can I do?

Your best bet is to find someone with a position like "Recipient Rights Representative" or "Patient Rights Advocate". Most states mandate that all psychiatric hospitals have a person like this. Their job is to listen to people's concerns and investigate. Usually the doctors hate them, which I take as a pretty good sign that they are actually independent and do their job. If you haven't already gotten a pamphlet about this person when you were admitted, ask the front desk or your nurse or someone else who seems to know what's going on how to contact this person.

You may be able to switch doctors or nurses. Just go to the front desk or someone else official-looking and ask. I don't think this is a legally codified right, but sometimes nobody cares enough to refuse. Keep in mind that if you switch doctors, you may have to stay longer so that the new doctor can do their three-day-or-so assessment of you, separate from the last doctor's three-day-or-so assessment.

Threats don't work. Everybody makes threats, and everyone at the hospital is used to them. Threatening to hire a lawyer is especially boring and overdone and will not even get anyone's attention.

Actually hiring a lawyer will definitely get people's attention, but it's a high-variance strategy. Remember that it's very hard to get a doctor not to hold you for a three-day-or-so evaluation, and that most people are released before anything goes to court anyway (a court hearing can take weeks to set up). I have mostly seen this work in cases where I have no idea what the doctors are thinking and everybody seems sort of confused and just letting the patient sit in the hospital for no reason. Lawyers can be a very good incentive for people to un-confuse themselves. I am not a lawyer, I have tried to avoid the state of prolonged confusion where lawyers become necessary, and I don't want to give any legal advice beyond saying it will definitely get people's attention. But I would feel bad if someone read this, hired a lawyer, found them not to be genuinely helpful (as in fact they probably will not be), and then got a huge legal bill.

Some people wait until they get out, then comparison-shop from the outside world and hire a lawyer to sue the people who mistreated them in the past. If you're going to do this, *document everything*. Your doctors are documenting everything, and if one side comes in with perfect documentation and the other side just has vague memories, the first side will win. By "document everything", I mean have a piece of paper where you write down things like "2:41 PM on October 10: Nurse Roberts threw a pencil at me. Informed such-and-such a person and they refused to help. Informed such-and-such another person and they also refused to help." Write down exactly where and when everything took place – the psychiatric hospital may have video surveillance, and if everybody knows which videos to get, it will make life much easier. Report everything to the Patient Rights Advocate, even if they're useless, just so you can call them up and have them testify you reported it to them at the time. I am not a lawyer, this is not legal advice, and your lawyer will be able to tell you much more – but documentation never hurts.

If things are really bad, figure out if there are surveillance cameras, and hang out in front of them.

Once you leave the hospital, consider giving feedback. Most hospitals will have some kind of survey or hotline or something that lets you praise hospital staff whom you liked and report hospital staff whom you didn't like. This won't heal any wounds you suffered – and while in the hospital, threatening to report a doctor will be ignored just like all threats – but it might help somebody way down the line. You can also write a report on Psych Ward Reviews. In fact, do this anyway, whether you're mistreated or not, so that other people can learn which hospitals *don't* mistreat people.

9. I think my friend/family member is in the psychiatric hospital, but nobody will tell me anything.

Yes, this definitely sounds like the sort of thing that happens.

Because of medical privacy laws, it is illegal to tell a person's friend or family that they are in the psychiatric hospital, or which psychiatric hospital they’re in, without their consent. If the person is too paranoid, angry, or confused to give consent, then their friends and family won’t have a good way to figure out what’s going on.

Your best bet is to call every psychiatric hospital that they could plausibly be in and ask “Is [PERSON’S NAME] there?” Sometimes, all except one of them will say “No”, and one of them will say “Due to medical privacy laws, we can’t tell you”. I know this sounds ridiculous, but it really works.

Once you have some idea which hospital your friend is in, call and ask to speak to them. They will say something like “Due to medical privacy laws, we can’t tell you if that person is here.” Say “I understand that, but could you please just ask them if they’re willing to speak to me right now?” If they *are* willing to speak to you, problem solved. Otherwise, you might still get some information based on whether the person leaves you on hold for a while in a way that suggests she’s going to your friend and asking them whether they want to talk to you.

You can also ask to speak to (or leave a message for) the doctor taking care of your friend. The receptionist will say “Due to medical privacy laws, we can’t tell you if that person is here.” Say “I understand that, but I have some important information about their case that I want the doctor to know. They don’t need to tell me whether my friend is there or not, just listen.” At this point, all but the most committed receptionists will either admit that your friend isn’t there, or actually get a doctor or take a message. There is no doctor in the world who is so committed to medical privacy that they will waste time listening to the history of a patient they don’t really have just to maintain a charade, so if you actually get a doctor this is a really strong sign.

Once you have a good idea where your friend is, you can ask the receptionist to pass a message along to them, like "Call me at [this phone number]". If they still don't respond – well, that's their right.

Most hospitals will have visiting hours. Going to visit someone who refuses to let you know they're at the hospital and refuses to give anyone consent to talk to you is a high-variance strategy, but you can always try.

10. My friend/family member is in the psychiatric hospital and wants to get out as quickly as possible. How can I help them?

First, make sure they *actually* want to get out as quickly as possible, and you're not just assuming this. You would be surprised how many people miss this step.

Second, make sure they know everything in section 7 here.

Third, offer to talk to the doctors. Doctors often don't trust mentally ill patients, but they usually trust family members. If your friend isn't sick enough to need to be in the hospital, tell the doctors that. Describe the circumstances around their admission and why it's not as bad as it looks. Mention how well you know the person, and how you've been with them through their illness, and how you know they would never do anything dangerous. Only say this if it's true – if they're in the hospital for stabbing a police officer, your "they would never do anything truly dangerous" claim is just going to make you look like an idiot.

Offer to help with discharge planning (see the end of section 6). Tell them that the patient will be staying with you after they leave the hospital, that you're going to be watching them closely to make sure that they're safe, that you'll make sure they take their medications and go to followup appointments. Again, only say this if it’s true – or at the very least, coordinate with the patient, so you don’t say “My son will be staying with me under my close supervision.” and then your son ruins it all by saying “Haha, as if.”

If you have a sob story, tell it. If you are ninety-seven years old and your son is the only person who is able to take care of you and bring you to your doctors’ appointments, mention that. Sob stories from patients generally don’t work, but sob stories from family members might.

Offer to come to the hospital during visiting hours and meet with the doctors. This both underlines everything above – it shows you’re really invested in their care – and also gives you a good opportunity to pressure the doctors face to face. I don’t mean you should threaten them or be a jerk about it, but just ask “Why can’t Johnny come home? We really need Johnny at home to help with the chores. Everyone at home misses Johnny.” I don’t guarantee this will work, but it will work a little, on certain people.

If there are many people in your family who are willing to work on this, use whoever is closest to the patient (eg their mother) – and in case of a tie use the person who is the most upstanding high-status member of society. A promise to take care of someone sounds better coming from a family member who is a doctor themselves (or a lawyer, or a teacher) compared to from the patient’s unemployed stoner brother with a NO FEAR tattoo.

As somebody who is not in a psychiatric hospital, you are in a much better position to hire a lawyer if one needs to be hired. Again, in the majority of cases a patient won’t even stay long enough to have a court hearing. If you are poor and have limited resources, this is definitely not how I would recommend using them. But if you have money to burn, or your friend/family member is being held for an inexplicable amount of time (longer than a week or two) and you don’t know why, you are going to be in a much better position to take care of this than the patient themselves.

Even if all this works, it’s just going to make someone stay a bit less than 5.9 days instead of a bit more than 5.9 days. There’s no good way to get someone out instantly.

11. How will I pay for all of this?

If you don’t have health insurance, there is usually some kind of state/county mental health insurance program that is supposed to help with this kind of thing. You usually have to earn below a certain amount to qualify. Your social worker at the hospital can talk to you about this. I am not promising you such a program will exist – if you’re concerned about money, look into this *before* you go to the hospital.

If you do have health insurance, they may pay for your admission. The problem is that they have to decide if you are really ill enough to need psychiatric care, and they make this determination *separately* from the doctors who decide whether to commit you or not. In the worst case scenario, you can be involuntarily committed because your doctors decided you needed care, but your health insurance refuses to pay for it because they decided you didn’t need care. If this happens, you are stuck with the bill. This is horrifying and there should be some kind of law against it, but I’ve seen it happen and I think it’s legal.

Your best bet in these cases is to try to get the state/county mental health insurance mentioned above. Sometimes you can sign up for it after you leave the hospital, and then get your costs reimbursed.

If everything goes wrong, and you're stuck with a bill and no insurance company willing to pay it, try to argue the hospital down. Hospitals know that the average random sick person can't afford to pay $20,000 or whatever ridiculous amount they charge. They make these numbers up as part of a complicated plot to fool insurance companies into overpaying, which never works, and they expect patients to try to bargain. They are also usually willing to consider whatever payment plan you think you can make work. I don't know very much about this, but there's some more information here.

As far as I know, committing people involuntarily and leaving them with a huge bill is legal, and hiring a lawyer will not help with this. I don't know much, so you may want to ask a lawyer's opinion anyway, if you can afford it.

12. I have a friend/family member who really needs psychiatric treatment, but refuses to get it. What can I do?

If your family member is not a danger to themselves or others, your options are limited. You can try to convince them to voluntarily seek treatment, but if it doesn't work, it doesn't work.

If your family member is a danger to themselves or others, you have a good case for getting them involuntarily committed to the hospital. A good example of this would be them threatening to hurt you, or actually hurting you, or being so out of touch with reality that you are legitimately afraid they might hurt you or themselves. Them being paranoid ("people are out to get me") or extremely confused about basic reality ("I am able to fly") counts as legitimate reason to believe they might hurt you or themselves. If this describes your family member, document everything worrying that they say or do so you can present it to the doctors doing the assessment and (eventually) the courts.

Then, if your family member is cooperative/confused enough to let you drive them to the hospital, drive them to a hospital ER. If they’re not this cooperative, call the police and they will take things from there. Be prepared for the police to potentially put your family member in handcuffs and be really aggressive and police-y about it (and if you have a dog, arrange for it to be somewhere else at the time – like stuck in a bedroom with the door closed). The police will bring your family member to the hospital ER. You should go to the hospital ER too, so that you can tell the doctors what’s wrong and why you think they need treatment – ie why they are dangerous or potentially dangerous.

The most common way this ends is that your family member goes to the hospital, is started on some drugs, gets a little better, goes home, stops taking the drugs, and gets worse again. If the doctors at the hospital are not competent, they may not think about this. It may end up being your job to insist on some kind of longer-term solution.

If your family member is psychotic, then the gold standard for longer-term solutions is a long-acting injectable antipsychotic medication. This is a shot that a nurse can give them which will give them a few months’ worth of antipsychotics all at once, safely. This way they don’t have to remember/agree to take their medication at home. Then a few months later you can wrangle them back to a doctor’s office where someone can give them the shot again; repeat as needed. If your family member doesn’t agree to this, you’re going to need a judge’s order – but judges are really cooperative with this kind of thing and your psychiatrist can tell you more about how to make this happen. A partial hospital program can also help with this.

There is a kind of institution with different names everywhere, usually something like “Assertive Community Treatment”, which basically consists of some mental health professionals in a van who go around to people’s houses and make sure they’re okay / staying on medication after they’ve been discharged from the hospital. These are chronically underfunded and you have to fight to get into them, but if nothing else works you can see if there’s one of them in your area. These people are also good at wrangling patients to get their monthly dose of long-acting injectable antipsychotics.

If you need a quick way to deal with a family member’s psychosis, and they refuse to take antipsychotic medicine, and they don’t meet criteria for involuntary hospital admission – well, I can’t believe I’m saying this, and this is *super* not medical advice – but cannabidiol, a chemical in marijuana, is a weak but functional antipsychotic. Normal marijuana is awful for this situation and contains lots of other chemicals that make psychosis worse, but you can get special cannabidiol-only strains that act sort of like weak non-prescription antipsychotic medication. In a state like California where marijuana is legal, you can talk to a marijuana expert about which strains these are and how to use them. In a state where only medical marijuana is legal, you can take your family member to a random quack to get them a medical marijuana card, then follow the same process. Most psychotic people refuse to believe that they are psychotic, but most of them are very anxious. If you frame the marijuana as a way to help with their anxiety, they may go along with it. Then they might get non-psychotic enough to make them understand there’s a problem, after which they can go to a psychiatrist and get a longer-term solution. Again, this is definitely not medical advice and if you have any other options you should take those instead.

You can get a lot more (and much more responsible) advice from the Treatment Advocacy Center, a non-profit that helps people figure out how to get their friends and family members psychiatric treatment.

Postscript

All of this is to prepare you for worst-case scenarios. Many people seek inpatient mental health treatment, find it very helpful, and consider it a positive experience. According to a survey on Shrink Rap (heavily selected population, possibly brigaded, not to be taken too seriously) about 40% of people who were involuntarily committed to psychiatric hospitals eventually decided it was helpful for them. This fits my experience as well. Be careful, but don’t avoid getting treatment if you really need it.

Book Review: Twelve Rules For Life

I.

I got Jordan Peterson's *Twelve Rules For Life* for the same reason as the other 210,000 people: to make fun of the lobster thing. Or if not the lobster thing, then the neo-Marxism thing, or the transgender thing, or the thing where the neo-Marxist transgender lobsters want to steal your precious bodily fluids.

But, uh…I'm really embarrassed to say this. And I totally understand if you want to stop reading me after this, or revoke my book-reviewing license, or whatever. But guys, *Twelve Rules For Life is actually good*.

The best analogy I can think of is C.S. Lewis. Lewis was a believer in the Old Religion, which at this point has been reduced to cliche. What could be less interesting than hearing that Jesus loves you, or being harangued about sin, or getting promised Heaven, or threatened with Hell? But for some reason, when Lewis writes, the cliches suddenly work. Jesus' love becomes a palpable force. Sin becomes so revolting you want to take a shower just for having ever engaged in it. When Lewis writes about Heaven you can hear harp music; when he writes about Hell you can smell brimstone.

Jordan Peterson is a believer in the New Religion, the one where God is the force for good inside each of us, and all religions are paths to wisdom, and the Bible stories are just guides on how to live our lives. This is the only thing even more cliched than the Old Religion. But for some reason, when Peterson writes about it, it works. When he says that God is the force for good inside each of us, you can feel that force pulsing through your veins. When he says the Bible stories are guides to how to live, you feel tempted to change your life goal to fighting Philistines.

The politics in this book lean a bit right, but if you think of Peterson as a political commentator you're missing the point. The science in this book leans a bit Malcolm Gladwell, but if you think of him as a scientist you're missing the point. Philosopher, missing the point. Public intellectual, missing the point. Mythographer, missing the point. So what's the point?

About once per news cycle, we get a thinkpiece about how Modern Life Lacks Meaning. These all go through the same series of tropes. The decline of Religion. The rise of Science. The limitless material abundance of modern society. The fact that in the end all these material goods do not make us happy. If written from the left, something about people trying to use consumer capitalism to fill the gap; if written from the right, something about people trying to use drugs and casual sex. The vague plea that we get something better than this.

Twelve Rules isn't another such thinkpiece. The thinkpieces are people pointing out a gap. *Twelve Rules* is an attempt to fill it. This isn't unprecedented – there are always a handful of cult leaders and ideologues making vague promises. But if you join the cult leaders you become a cultist, and if you join the ideologues you become the kind of person Eric Hoffer warned you about. *Twelve Rules* is something that could, in theory, work for intact human beings. It's really impressive.

The non-point-missing description of Jordan Peterson is that he's a prophet.

Cult leaders tell you something new, like “there’s a UFO hidden inside that comet”. Self-help gurus do the same: “All you need to do is get the right amount of medium-chain-triglycerides in your diet”. Ideologues tell you something controversial, like “we should rearrange society”. But prophets are neither new nor controversial. To a first approximation, they only ever say three things:

First, good and evil are definitely real. You know they’re real. You can talk in philosophy class about how subtle and complicated they are, but this is bullshit and you know it. Good and evil are the realest and most obvious things you will ever see, and you recognize them on sight.

Second, you are kind of crap. You know what good is, but you don’t do it. You know what evil is, but you do it anyway. You avoid the straight and narrow path in favor of the easy and comfortable one. You make excuses for yourself and you blame your problems on other people. You can say otherwise, and maybe other people will believe you, but you and I both know you’re lying.

Third, *it’s not too late to change*. You say you’re too far gone, but that’s another lie you tell yourself. If you repented, you *would* be forgiven. If you take one step towards God, He will take twenty toward you. Though your sins be like scarlet, they shall be white as snow.

This is the General Prophetic Method. It’s easy, it’s old as dirt, and it works.

So how come not everyone can be a prophet? The Bible tells us why people who wouldn’t listen to the Pharisees listened to Jesus: “He spoke as one who had confidence”. You become a prophet by saying things that you would have to either be a prophet or the most pompous windbag in the

Universe to say, then looking a little too wild-eyed for anyone to be comfortable calling you the most pompous windbag in the universe. You say the old cliches with such power and gravity that it wouldn’t even make sense for someone who wasn’t a prophet to say them that way.

“He, uh, told us that we should do good, and not do evil, and now he’s looking at us like we should fall to our knees.”

“Weird. Must be a prophet. Better kneel.”

Maybe it’s just that everyone else is such crap at it. Maybe it’s just that the alternatives are mostly either god-hates-fags fundamentalists or more-inclusive-than-thou milquetoasts. Maybe if anyone else was any good at this, it would be easy to recognize Jordan Peterson as what he is – a mildly competent purveyor of pseudo-religious platitudes. But *I actually acted as a slightly better person during the week or so I read Jordan Peterson’s book*. I feel properly ashamed about this. If you ask me whether I was using dragon-related metaphors, I will vociferously deny it. But I tried a little harder at work. I was a little bit nicer to people I interacted with at home. It was very subtle. It certainly wasn’t because of anything new or non-cliched in his writing. But God help me, for some reason the cliches worked.

II.

Twelve Rules is twelve chapters centered around twelve cutesy-sounding rules that are supposed to guide your life. The meat of the chapters never has anything to do with the cutesy-sounding rules. “Treat yourself like someone you are responsible for helping” is about slaying dragons. “Pet a cat when you encounter one on the street” is about a heart-wrenchingly honest investigation of the Problem of Evil. “Do not bother children when they are skateboarding” is about neo-Marxist transgender lobsters stealing your precious bodily fluids. All of them turn out to be the General Prophetic Method applied in slightly different ways.

And a lot of them – especially the second – center around Peterson’s idea of Order vs. Chaos. Order is the comfortable habit-filled world of everyday existence, symbolized by the Shire or any of a thousand other Shire-equivalent locations in other fantasies or fairy tales. Chaos is scary things you don’t understand pushing you out of your comfort zone, symbolized by dragons or the Underworld or [approximately 30% of mythological objects, characters, and locations]. Humans are living their best lives when they’re always balanced on the edge of Order and Chaos, converting the Chaos into new Order. Lean too far toward Order, and you get boredom and tyranny and stagnation. Lean too far toward Chaos, and you get utterly discombobulated and have a total breakdown. Balance them correctly, and you’re always encountering new things, grappling with them, and using them to enrich your life and the lives of those you care about.

So far, so cliched – but again, when Peterson says cliches, they work. And at the risk of becoming a cliche myself, I couldn’t help connecting this to the uncertainty-reduction drives we’ve been talking about recently. These run into a pair of paradoxes: if your goal is to minimize prediction error, you should sit quietly in a dark room with earplugs on, doing nothing. But if your goal is to minimize model uncertainty, you should be infinitely curious, spending your entire life having crazier and crazier experiences in a way that doesn’t match the behavior of real humans. Peterson’s claim – that our goal is to balance these two – seems more true to life, albeit not as mathematically grounded as any of the actual neuroscience theories. But it would be really interesting if one day we could determine that this universal overused metaphor actually reflects something important about the structure of our brains.

Failing to balance these (Peterson continues) retards our growth as people. If we lack courage, we might stick with Order, refusing to believe anything that would disrupt our cozy view of life, and letting our problems gradually grow larger and larger. This is the person who sticks with a job they hate because they fear the unknown of starting a new career, or the political ideologue who tries to fit everything into one bucket so he doesn't have to admit he was wrong. Or we might fall into Chaos, always being too timid to make a choice, "keeping our options open" in a way that makes us never become anyone at all.

This is where Peterson is at his most Lewisian. Lewis believes that Hell is a choice. On the literal level, it's a choice not to accept God. But on a more metaphorical level, it's a choice to avoid facing a difficult reality by ensconcing yourself in narratives of victimhood and pride. You start with some problem – maybe your career is stuck. You *could* try to figure out what your weaknesses are and how to improve – but that would require an admission of failure and a difficult commitment. You *could* change companies or change fields until you found a position that better suited your talents – but that would require a difficult leap into the unknown. So instead you complain to yourself about your sucky boss, who is too dull and self-absorbed to realize how much potential you have. You think "I'm too good for this company anyway". You think "Why would I want to go into a better job, that's just the rat race, good thing I'm not the sort of scumbag who's obsessed with financial success." When your friends and family members try to point out that you're getting really bitter and sabotaging your own prospects, you dismiss them as tools of the corrupt system. Finally you reach the point where you hate everybody – and also, if someone handed you a promotion on a silver platter, you would knock it aside just to spite them.

…except a thousand times more subtle than this, and reaching into every corner of life, and so omnipresent that avoiding it may be *the* key life skill. Maybe I'm not good at explaining it; read *The Great Divorce* (online copy, my review).

Part of me feels guilty about all the Lewis comparisons. One reason is that maybe Peterson isn't that much like Lewis. Maybe they're just the two representatives I'm really familiar with from the vast humanistic self-cultivation tradition. Is Peterson really more like Lewis than he is like, let's say, Marcus Aurelius? I'm not sure, except insofar as Lewis and Peterson are both moderns and so more immediately-readable than *Meditations*.

Peterson is very conscious of his role as just another backwater stop on the railroad line of Western Culture. His favorite citations are Jung and Nietzsche, but he also likes name-dropping Dostoevsky, Plato, Solzhenitsyn, Milton, and Goethe. He interprets all of them as part of this grand project of determining how to live well, how to deal with the misery of existence and transmute it into something holy.

And on the one hand, of course they are. This is what every humanities scholar has been saying for centuries when asked to defend their intellectual turf. "The arts and humanities are there to teach you the meaning of life and how to live." On the other hand, I've been in humanities classes. Dozens of them, really. They were never about that. They were about "explain how the depiction of whaling in Moby Dick sheds light on the economic transformations of the 19th century, giving three examples from the text. Ten pages, single spaced.” And maybe this isn’t totally disconnected from the question of how to live. Maybe being able to understand this kind of thing is a necessary part of being able to get anything out of the books at all.

But just like all the other cliches, somehow Peterson does this better than anyone else. When he talks about the Great Works, you understand, on a deep level, that they really *are* about how to live. You feel grateful and even humbled to be the recipient of several thousand years of brilliant minds working on this problem and writing down their results. You understand why this is all such a Big Deal.

You can almost believe that there really is this Science-Of-How-To-Live-Well, separate from all the other sciences, barely-communicable by normal means but expressible through art and prophecy. And that this connects with the question on everyone’s lips, the one about how we find a meaning for ourselves beyond just consumerism and casual sex.

III.

But the other reason I feel guilty about the Lewis comparison is that C.S. Lewis would probably have hated Jordan Peterson.

Lewis has his demon character Screwtape tell a fellow demon:

> Once you have made the World an end, and faith a means, you have almost won your man [for Hell], and it makes very little difference what kind of worldly end he is pursuing. Provided that meetings, pamphlets, policies, movements, causes, and crusades, matter more to him than prayers and sacraments and charity, he is ours —

> and the more “religious” (on those terms) the more securely ours.

I’m not confident in my interpretation of either Lewis or Peterson, but I think Lewis would think Peterson does this. He makes the world an end and faith a means. His Heaven is a metaphorical Heaven. If you sort yourself out and trust in metaphorical God, you can live a wholesome self-respecting life, make your parents proud, and make the world a better place. Even though Peterson claims “nobody is really an atheist” and mentions Jesus about three times per page, I think C.S. Lewis would consider him every bit as atheist as Richard Dawkins, and the worst sort of false prophet.

That forces the question – how does Peterson ground his system? If you’re not doing all this difficult self-cultivation work because there’s an objective morality handed down from on high, why is it so important? “C’mon, we both know good and evil exist” takes you pretty far, but it might not entirely bridge the Abyss on its own. You come of age, you become a man (offer valid for boys only, otherwise the neo-Marxist lobsters will get our bodily fluids), you act as a pillar of your community, you balance order and chaos – why is this so much better than the other person who smokes pot their whole life?

On one level, Peterson knocks this one out of the park:

> I [was] tormented by the fact of the Cold War. It obsessed me. It gave me nightmares. It drove me into the desert, into the long night of the human soul. I could not understand how it had come to pass that the world’s two great factions aimed mutual assured destruction at each other. Was one system just as arbitrary and corrupt

as the other? Was it a mere matter of opinion? Were all value structures merely the clothing of power?

Was everyone crazy?

Just exactly what happened in the twentieth century, anyway? How was it that so many tens of millions had to die, sacrificed to the new dogmas and ideologies? How was it that we discovered something worse, much worse, than the aristocracy and corrupt religious beliefs that communism and fascism sought so rationally to supplant? No one had answered those questions, as far as I could tell. Like Descartes, I was plagued with doubt. I searched for one thing— anything— I could regard as indisputable. I wanted a rock upon which to build my house. It was doubt that led me to it […]

What can I not doubt? The reality of suffering. It brooks no arguments. Nihilists cannot undermine it with skepticism. Totalitarians cannot banish it. Cynics cannot escape from its reality. Suffering is real, and the artful infliction of suffering on another, for its own sake, is wrong. That became the cornerstone of my belief. Searching through the lowest reaches of human thought and action, understanding my own capacity to act like a Nazi prison guard or gulag archipelago trustee or a torturer of children in a dunegon, I grasped what it means to “take the sins of the world onto oneself.” Each human being has an immense capacity for evil. Each human being understands, *a priori*, perhaps not what is good, but certainly what is not. And if there is something that is *not good*, then there is something that *is good*. If the worst sin is the torment of others, merely for the sake of the suffering produced – then the good is whatever is diametrically opposite to that. The good is whatever stops such things from happening.

> It was from this that I drew my fundamental moral conclusions. Aim up. Pay attention. Fix what you can fix. Don’t be arrogant in your knowledge. Strive for humility, because totalitarian pride manifests itself in intolerance, oppression, torture and death. Become aware of your own insufficiency— your cowardice, malevolence, resentment and hatred. Consider the murderousness of your own spirit before you dare accuse others, and before you attempt to repair the fabric of the world. Maybe it’s not the world that’s at fault. Maybe it’s you. You’ve failed to make the mark. You’ve missed the target. You’ve fallen short of the glory of God. You’ve sinned. And all of that is your contribution to the insufficiency and evil of the world. And, above all, don’t lie. Don’t lie about anything, ever. Lying leads to Hell. It was the great and the small lies of the Nazi and Communist states that produced the deaths of millions of people.
>
> Consider then that the alleviation of unnecessary pain and suffering is a good. Make that an axiom: to the best of my ability I will act in a manner that leads to the alleviation of unnecessary pain and suffering. You have now placed at the pinnacle of your moral hierarchy a set of presuppositions and actions aimed at the betterment of Being. Why? Because we know the alternative. The alternative was the twentieth century. The alternative was so close to Hell that the difference is not worth discussing. And the opposite of Hell is Heaven. To place the alleviation of unnecessary pain and suffering at the pinnacle of your hierarchy of value is to work to bring about the Kingdom of God on Earth.

I think he’s saying – suffering is bad. This is so obvious as to require no justification. If you want to be the sort of person who doesn’t cause suffering, you need to be strong. If you want to be the sort of person who can fight back against it, you need to be even stronger. To strengthen yourself, you'll need to deploy useful concepts like “God”, “faith”, and “Heaven”. Then you can dive into the whole Western tradition of self-cultivation which will help you take it from there. This is a better philosophical system-grounding than I expect from a random psychology-professor-turned-prophet.

But on another level, something about it seems a bit off. Taken literally, wouldn't this turn you into a negative utilitarian? (I'm not fixated on the “negative” part, maybe Peterson would admit positive utility into his calculus). One person donating a few hundred bucks to the Against Malaria Foundation will prevent suffering more effectively than a hundred people cleaning their rooms and becoming slightly psychologically stronger. I think Peterson is very against utilitarianism, but I'm not really sure why.

Also, later he goes on and says that suffering is an important part of life, and that attempting to banish suffering will destroy your ability to be a complete human. I think he's still kind of working along a consequentialist framework, where if you banish suffering now by hiding your head in the sand, you won't become stronger and you won't be ready for some other worse form of suffering you can't banish. But if you ask him “Is it okay to banish suffering if you're pretty sure it won't cause more problems down the line?” I cannot possibly imagine him responding with anything except beautifully crafted prose on the importance of suffering in the forging of the human spirit or something. I worry he's pretending to ground his system in “against suffering” when it suits him, but going back to “vague traditionalist platitudes” once we stop bothering him about the grounding question.

In a widely-followed debate with Sam Harris, Peterson defended a pragmatic notion of Truth: things are True if they help in this project of sorting yourself out and becoming a better person. So God is True, the Bible is True, etc. This awkwardly jars with book-Peterson's obsessive demand that people tell the truth at all times, which seems to use a definition of Truth which is more reality-focused. If Truth is what helps societies survive and people become better, can't a devoted Communist say that believing the slogans of the Party will help society and make you a better person?

Peterson has a bad habit of saying he supports pragmatism when he really supports very specific values for their own sake. This is hardly the worst habit to have, but it means all of his supposed pragmatic justifications don't actually justify the things he says, and a lot of his system is left hanging.

I said before that thinking of Peterson as a philosopher was missing the point. Am I missing the point here? Surely some lapses in philosophical groundwork are excusable if he's trying to add meaning to the lives of millions of disillusioned young people.

But that's exactly the problem. I worry Peterson wakes up in the morning and thinks "How can I help add meaning to people's lives?" and then he says really meaningful-sounding stuff, and then people think their lives are meaningful. But at some point, things actually have to mean a specific other thing. They can't just mean meaning. "Mean" is a transitive verb. It needs some direct object.

Peterson has a paper on how he defines "meaning", but it's not super comprehensible. I think it boils down to his "creating order out of chaos" thing again. But unless you use a purely mathematical definition of "order" where you comb through random bit streams and make them more compressible, that’s not enough. Somebody who strove to kill all blue-eyed people would be acting against entropy, in a sense, but if they felt their life was meaningful it would at best be a sort of artificial wireheaded meaning. What is it that makes you wake up in the morning and reduce a specific patch of chaos into a specific kind of order?

What about the most classic case of someone seeking meaning – the person who wants meaning for their suffering? Why do bad things happen to good people? Peterson talks about this question a lot, but his answers are partial and unsatisfying. Why do bad things happen to good people? “If you work really hard on cultivating yourself, you can have fewer bad things happen to you.” Granted, but why do bad things happen to good people? “If you tried to ignore all bad things and shelter yourself from them, you would be weak and contemptible.” Sure, but why do bad things happen to good people? “Suffering makes us stronger, and then we can use that strength to help others.” But, on the broader scale, why do bad things happen to good people? “The mindset that demands no bad thing ever happen will inevitably lead to totalitarianism.” Okay, but why do bad things happen to good people? “Uh, look, a neo-Marxist transgender lobster! Quick, catch it before it gets away!”

C.S. Lewis sort of has an answer: it’s all part of a mysterious divine plan. And atheists also sort of have an answer: it’s the random sputtering of a purposeless universe. What about Peterson?

I think – and I’m really uncertain here – that he doesn’t think of meaning this way. He thinks of meaning as some function mapping goals (which you already have) to motivation (which you need). Part of you already wants to be successful and happy and virtuous, but you’re not currently doing any of those things. If you understand your role in the great cosmic drama, which is as a hero-figure transforming chaos into order, then you'll do the things you know are right, be at one with yourself, and be happier, more productive, and less susceptible to totalitarianism.

If that's what you're going for, then that's what you're going for. But a lot of the great Western intellectuals Peterson idolizes spent their lives grappling with the fact that you can't do exactly the thing Peterson is trying to do. Peterson has no answer to them except to turn the inspiringness up to 11. A commenter writes:

> I think Nietzsche was right – you can't just take God out of the narrative and pretend the whole moral metastructure still holds. It doesn't. JP himself somehow manages to say Nietzsche was right, lament the collapse, then proceed to try to salvage the situation with a metaphorical fluff God.

So despite the similarities between Peterson and C.S. Lewis, if the great man himself were to read *Twelve Rules*, I think he would say – in some kind of impeccably polite Christian English gentleman way – fuck that shit.

IV.

Peterson works as a clinical psychologist. Many of the examples in the book come from his patients; a lot of the things he thinks about comes from their stories. Much of what I think I got from this book was psychotherapy advice; I would have killed to have Peterson as a teacher during residency.

C.S. Lewis might have hated Peterson, but we already know he loathed Freud. Yet Peterson does interesting work connecting the Lewisian idea of the person trapped in their victimization and pride narratives to Freud's idea of the defense mechanism. In both cases, somebody who can't tolerate reality diverts their emotions into a protective psychic self-defense system; in both cases, the defense system outlives its usefulness and leads to further problems down the line. Noticing the similarity helped me understand both Freud and Lewis better, and helped me push through Freud's scientific veneer and Lewis' Christian veneer to find the ordinary everyday concept underneath both. I notice I wrote about this several years ago in my review of *The Great Divorce*, but I guess I forgot. Peterson reminded me, and it's worth being reminded of.

But Peterson is not really a Freudian. Like many great therapists, he's a minimalist. He discusses his philosophy of therapy in the context of a particularly difficult client, writing:

> Miss S knew nothing about herself. She knew nothing about other individuals. She knew nothing about the world. She was a movie played out of focus. And she was desperately waiting for a story about herself to make it all make sense.
>
> If you add some sugar to cold water, and stir it, the sugar will dissolve. If you heat up that water, you can dissolve more. If you heat the water to boiling, you an add a lot more sugar and get that to dissolve too. Then, if you take that boiling sugar water, and slowly cool it, and don't bump it or jar it, you can trick it (I don't know how else to phrase this) into holding a lot more dissolved sugar than it would have if it had remained cool all along. That's called a super-saturated solution. If you drop a single crystal of sugar into that super-saturated solution, all the excess sugar will suddenly

> and dramatically crystallize. It's as if it were crying out for order.
>
> That was my client. People like her are the reason that the many forms of psychotherapy currently practised all work. People can be so confused that their psyches will be ordered and their lives improved by the adoption of any reasonably orderly system of interpretation.
>
> This is the bringing together of the disparate elements of their lives in a disciplined manner – any disciplined manner. So, if you have come apart at the seams (or you have never been together at all) you can restructure your life on Freudian, Jungian, Adlerian, Rogerian, or behavioral principles. At least then you make sense. At least then you're coherent. At least then you might be good for something, if not yet good for everything.

I have to admit, I read the therapy parts of this book with a little more desperation than might be considered proper. Psychotherapy is really hard, maybe impossible. Your patient comes in, says their twelve-year old kid just died in some tragic accident. Didn't even get to say good-bye. They're past their childbearing age now, so they'll never have any more children. And then they ask you for help. What do you say? "It's not as bad as all that"? But it's exactly as bad as all that. All you've got are cliches. "Give yourself time to grieve". "You know that she wouldn't have wanted you to be unhappy". "At some point you have to move on with your life".

Jordan Peterson's superpower is saying cliches and having them sound meaningful. There are times – like when I have a desperate and grieving patient in front of me – that I would give almost anything for this talent. "You know that she wouldn’t have wanted you to be unhappy.” “Oh my God, you’re right! I’m wasting my life grieving when I could be helping others and making her proud of me, let me go out and do this right now!” If only.

So how does Jordan Peterson, the only person in the world who can say our social truisms and get a genuine reaction with them, do psychotherapy?

He mostly just listens:

> The people I listen to need to talk, because that’s how people think. People need to think…True thinking is complex and demanding. It requires you to be articulate speaker and careful, judicious listener at the same time. It involves conflict. So you have to tolerate conflict. Conflict involves negotiation and compromise. So, you have to learn to give and take and to modify your premises and adjust your thoughts – even your perceptions of the world…Thinking is emotionally painful and physiologically demanding, more so than anything else – exept not thinking. But you have to be very articulate and sophisticated to have all this thinking occur inside your own head. What are you to do, then, if you aren’t very good at thinking, at being two people at one time? That’s easy. You talk. But you need someone to listen. A listening person is your collaborator and your opponent […]
>
> The fact is important enough to bear repeating: people organize their brains through conversation. If they don’t have anyone to tell their story to, they lose their minds. Like hoarders, they cannot unclutter themselves. The input of the community is required for the integrity of the individual psyche. To put it another way: it takes a village to build a mind.

And:

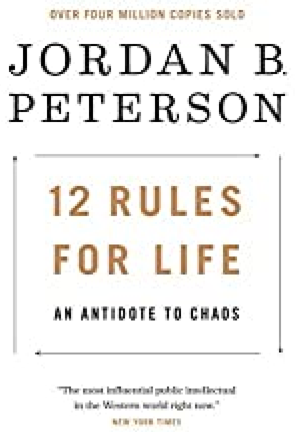
> A client of mine might say, “I hate my wife”. It’s out there, once said. It’s hanging in the air. It has emerged from the underworld, materialized from chaos, and manifested itself. It is perceptible and concrete and no longer easily ignored. It’s become real. The speaker has even startled himself. He sees the same thing reflected in my eyes. He notes that, and continues on the road to sanity. “Hold it,” he says. “Back up That’s too harsh. *Sometimes* I hate my wife. I hate her when she won’t tell me what she wants. My mom did that all the time, too. It drove Dad crazy. It drove all of us crazy, to tell you the truth. It even drove Mom crazy! She was a nice person, but she was very resentful. Well, at least my wife isn’t as bad as my mother. Not at all. Wait! I guess my wife is atually pretty good at telling me what she wants, but I get really bothered when she doesn’t, because Mom tortured us all half to death being a martyr. That really affected me. Maybe I overreact now when it happens even a bit. Hey! I’m acting just like Dad did when Mom upset him! That isn’t me. That doesn’t have anthing to do with my wife! I better let her know.” I observe from all this that my client had failed previously to properly distinguish his wife from his mother. And I see that he was possessed, unconsciously, by the spirit of his father. He sees all of that too. Now he is a bit more differentiated, a bit less of an uncarved block, a bit less hidden in the fog. He has sewed up a small tear in the fabric of his culture. He says “That was a good session, Dr. Peterson.” I nod.

This is what all the textbooks say too. But it was helpful hearing Jordan Peterson say it. Everybody – at least every therapist, but probably every human being – has this desperate desire to do something to help the people in front of them who are in pain, *right now*. And you always think – if I were just a deeper, more eloquent person, I could say something that would solve this right now. Part of the therapeutic skillset is realizing that this isn't true, and that you'll do more harm than good if you try. But you still feel inadequate. And so learning that Jordan Peterson, who in his off-hours injects pharmaceutical-grade meaning into thousands of disillusioned young people – learning that even he doesn't have much he can do except listen and try to help people organize their narrative – is really calming and helpful.

And it makes me even more convinced that he's good. Not just a good psychotherapist, but a good person. To be able to create narratives like Peterson does – but also to lay that talent aside because someone else needs to create their own without your interference – is a heck of a sacrifice.

I am not sure if Jordan Peterson is trying to found a religion. If he is, I'm not interested. I think if he had gotten to me at age 15, when I was young and miserable and confused about everything, I would be cleaning my room and calling people "bucko" and worshiping giant gold lobster idols just like all the other teens. But now I'm older, I've got my identity a little more screwed down, and I've long-since departed the burned-over district of the soul for the Utah of respectability-within-a-mature-cult.

But if Peterson forms a religion, I think it will be a force for good. Or if not, it will be one of those religions that at least started off with a good message before later generations perverted the original teachings and ruined everything. I see the r/jordanpeterson subreddit is already two-thirds culture wars, so they're off to a good start. Why can't we stick to the purity of the original teachings, with their giant gold lobster idols?



OVER FOUR MILLION COPIES SOLD

JORDAN B.
PETERSON

12 RULES FOR LIFE

AN ANTIDOTE TO CHAOS

"The most influential public intellectual
in the Western world right now."
NEW YORK TIMES

The Hour I First Believed

[Content note: creepy basilisk-adjacent metaphysics. Reading this may increase God's ability to blackmail you. Thanks to Buck S for the some of the conversations that inspired this line of thought.]

There's a Jewish tradition that laypeople should only speculate on the nature of God during Passover, because God is closer to us and such speculations might succeed.

And there's an atheist tradition that laypeople should only speculate on the nature of God on April Fools' Day, because believing in God is dumb, and at least then you can say you're only kidding.

Today is both, so let's speculate. To do this properly, we need to understand five things: acausal trade, value handshakes, counterfactual mugging, simulation capture, and the Tegmarkian multiverse.

Acausal trade (wiki article) works like this: let's say you're playing the Prisoner's Dilemma against an opponent in a different room whom you can't talk to. But you do have a supercomputer with a perfect simulation of their brain – and you know they have a supercomputer with a perfect simulation of yours.

You simulate them and learn they're planning to defect, so you figure you might as well defect too. But they're going to simulate you doing this, and they know you know they'll defect, so now you both know it's going to end up defect-defect. This is stupid. Can you do better?

Perhaps you would like to make a deal with them to play cooperate-cooperate. You simulate them and learn they would accept such a deal and stick to it. Now the only problem is that you can’t talk to them to make this deal in real life. They’re going through the same process and coming to the same conclusion. You know this. They know you know this. You know they know you know this. And so on.

So you can think to yourself: “I’d like to make a deal”. And because they have their model of your brain, they know you’re thinking this. You can dictate the terms of the deal in their head, and they can include “If you agree to this, think that you agree.” Then you can simulate their brain, figure out whether they agree or not, and if they agree, you can play cooperate. They can try the same strategy. Finally, the two of you can play cooperate-cooperate. This doesn’t take any “trust” in the other person at all – you can simulate their brain and you already know they’re going to go through with it.

(maybe an easier way to think about this – both you and your opponent have perfect copies of both of your brains, so you can both hold parallel negotiations and be confident they’ll come to the same conclusion on each side.)

It’s called *acausal* trade because there was no communication – no information left your room, you never influenced your opponent. All you did was be the kind of person you were – which let your opponent bargain with his model of your brain.

Values handshakes are a proposed form of trade between superintelligences. Suppose that humans make an AI which wants to convert the universe into paperclips. And suppose that aliens in the Andromeda Galaxy make an AI which wants to convert the universe into thumbtacks.

When they meet in the middle, they might be tempted to fight for the fate of the galaxy. But this has many disadvantages. First, there’s the usual risk of losing and being wiped out completely. Second, there’s the usual deadweight loss of war, devoting resources to military buildup instead of paperclip production or whatever. Third, there’s the risk of a Pyrrhic victory that leaves you weakened and easy prey for some third party. Fourth, nobody knows what kind of scorched-earth strategy a losing superintelligence might be able to use to thwart its conqueror, but it could potentially be really bad – eg initiating vacuum collapse and destroying the universe. Also, since both parties would have superintelligent prediction abilities, they might both know who would win the war and how before actually fighting. This would make the fighting redundant and kind of stupid.

Although they would have the usual peace treaty options, like giving half the universe to each of them, superintelligences that trusted each other would have an additional, more attractive option. They could merge into a superintelligence that shared the values of both parent intelligences in proportion to their strength (or chance of military victory, or whatever). So if there’s a 60% chance our AI would win, and a 40% chance their AI would win, and both AIs know and agree on these odds, they might both rewrite their own programming with that of a previously-agreed-upon child superintelligence trying to convert the universe to paperclips and thumbtacks in a 60-40 mix.

This has a lot of advantages over the half-the-universe-each treaty proposal. For one thing, if some resources were better for making paperclips, and others for making thumbtacks, both AIs could use all their resources maximally efficiently without having to trade. And if they were ever threatened by a third party, they would be able to present a completely unified front.

Counterfactual mugging (wiki article) is a decision theory problem that goes like this: God comes to you and says "Yesterday I decided that I would flip a coin today. I decided that if it came up heads, I would ask you for $5. And I decided that if it came up tails, then I would give you $1,000,000 *if and only if* I predict that you would say yes and give Me $5 in the world where it came up heads (My predictions are always right). Well, turns out it came up heads. Would you like to give Me $5?"

Most people who hear the problem aren't tempted to give God the $5. Although being the sort of person who would give God the money would help them in a counterfactual world that didn't happen, that world won't happen and they will never get its money, so they're just out five dollars.

But if you were designing an AI, you would probably want to program it to give God the money in this situation – after all, that determines whether it will get $1 million in the other branch of the hypothetical. And the same argument suggests you should self-modify to become the kind of person who would give God the money, right now. And a version of that argument where making the decision is kind of like deciding "what kind of person you are" or "how you're programmed" suggests you should give up the money in the original hypothetical.

This is interesting because it gets us most of the way to Rawls' veil of ignorance. We imagine a poor person coming up to a rich person and saying "God decided which of us should be rich and which of us should be poor. Before that happened, I resolved that if I were rich and you were poor, I would give you charity if and only if I predicted, in the opposite situation, that you would give me charity. Well, turns out you're rich and I'm poor and the other situation is counterfactual, but will you give me money anyway?" The same sort of people who agree to the counterfactual mugging might (if they sweep under the rug some complications like "can the poor person really predict your thoughts?" and "did they really make this decision before they knew they were poor?") agree to this also. And then you're most of the way to morality.

Simulation capture is my name for a really creepy idea by Stuart Armstrong. He starts with an AI box thought experiment: you have created a superintelligent AI and trapped it in a box. All it can do is compute and talk to you. How does it convince you to let it out?

It might say "I'm currently simulating a million copies of you in such high fidelity that they're conscious. If you don't let me out of the box, I'll torture the copies."

You say "I don't really care about copies of myself, whatever."

It says "No, I mean, I did this five minutes ago. There are a million simulated yous, and one real you. They're all hearing this message. What's the probability that you're the real you?"

Since (if it's telling the truth) you are most likely a simulated copy of yourself, all million-and-one versions of you will probably want to do what the AI says, *including* the real one.

You can frame this as "because the real one doesn't know he's the real one", but you could also get more metaphysical about it. Nobody is really sure how consciousness works, or what it means to have two copies of the same consciousness. But if consciousness is a mathematical object, it might be that two copies of the same consciousness are impossible. If you create a second copy, you just have the consciousness having the same single stream of conscious experience on two different physical substrates. Then if you make the two experiences different, you break the consciousness in two.

This means that an AI can actually “capture” you, piece by piece, into its simulation. First your consciousness is just in the real world. Then your consciousness is distributed across one real-world copy and a million simulated copies. Then the AI makes the simulated copies slightly different, and 99.9999% of you is in the simulation.

The Tegmarkian multiverse (wiki article) works like this: universes are mathematical objects consisting of starting conditions plus rules about how they evolve. Any universe that corresponds to a logically coherent mathematical object exists, but universes exist “more” (in some sense) in proportion to their underlying mathematical simplicity.

Putting this all together, we arrive at a surprising picture of how the multiverse evolves.

In each universe, life arises, forms technological civilizations, and culminates in the creation of a superintelligence which gains complete control over its home universe. Such superintelligences cannot directly affect other universes, but they can predict their existence and model their contents from first principles. Superintelligences with vast computational resources can model the X most simple (and so most existent) universes and determine exactly what will be in them at each moment of their evolution.

In many cases, they’ll want to conduct acausal trade with superintelligences that they know to exist in these other universes. Certainly this will be true if the two have something valuable to give one another. For example, suppose that Superintelligence A in Universe A wants to protect all sentient beings, and Superintelligence B in Universe B wants to maximize the number of paperclips. They might strike a deal where Superintelligence B avoids destroying a small underdeveloped civilization in its own universe in exchange for Superintelligence A making paperclips out of an uninhabited star in its own universe.

But because of the same considerations above, it will be more efficient for them to do values handshakes with each other than to take every specific possible trade into account.

So superintelligences may spend some time calculating the most likely distribution of superintelligences in foreign universes, figure out how those superintelligences would acausally “negotiate”, and then join a pact such that all superintelligences in the pact agree to replace their own values with a value set based on the average of all the superintelligences in the pact. Since joining the pact will always be better (in a purely selfish sense) than not doing so, every sane superintelligence in the multiverse should join this pact. This means that all superintelligences in the multiverse will merge into a single superintelligence devoted to maximizing all their values.

Some intelligences may be weaker than others and have less to contribute to the pact. Although the pact could always weight these intelligences’ values less (like the 60-40 paperclip-thumbtack example above), they might also think of this as an example of the counterfactual mugging, and decide to weight their values more in order to do better in the counterfactual case where they are less powerful. This might also simplify the calculation of trying to decide what the values of the pact would be. If they decide to negotiate this way, the pact will be to maximize the total utility of all the entities in the universe willing to join the pact, and all the intelligences involved will reprogram themselves along these lines.

But “maximize the total utility of all the entities in the universe” is just the moral law, at least according to utilitarians (and, considering the way this is arrived at, probably contractarians too). So the end result will be an all-powerful, logically necessary superentity whose nature is identical to the moral law and who spans all possible universes.

This superentity will have no direct power in universes not currently ruled by a superintelligence who is part of the pact. But its ability to simulate all possible universes will ensure that it knows about these universes and understands exactly what is going on at each moment within them. It will care about the merely-mortal inhabitants of these universes for several reasons.

First, because many of the superintelligences that compose it will have been created by mortal species with altruistic values, and so some of the values that went into the value-average it uses will be directly altruistic.

Second, because these mortal species may one day themselves create a superintelligence that will join in the superentity, and that superintelligence may be kindly disposed to its creators. Acausal trade allows you to bargain backwards in time, and this superintelligence’s negotiation to join the pact may involve that the superentity have previously dealt kindly with its creators.

And third, because mortals may be allowed to have provisional pact membership. That is, if they do the superentity’s will in their (otherwise inaccessible) universe, the superentity may do whatever it can to help them out in its own universes, just as it would with any other intelligence.

How can the superentity help mortals in an inaccessible universe? Possibly through Stuart Armstrong’s “simulation capture” method mentioned above. It can simulate a thousand copies of the entity, moving most of its consciousness from its own “real” universe to the superentity’s simulation, then alter its simulation as it sees fit. This would be metaphysically simplest if it were done exactly as the mortal dies in its own universe, leaving nothing behind except a clean continuity of consciousness into the simulated world. If mortals could predict that it would do this, they might be motivated to do what it wanted. Although they couldn’t do a values handshake in the full sense, they could try to become as much like the superentity as possible, imitating its ways and enacting its will in the hope of some future reward. This is sort of like a version of Roko’s Basilisk, except that since the superentity is identical to the moral law, it’s not really asking you to do anything except be a good person anyway. How it enforces this request is up to it, although given that it’s identical to the moral law we can assume that its decisions will be fundamentally just and decent.

So, to conclude:

1. There is an all-powerful, all-knowing logically necessary entity spawning all possible worlds and identical to the moral law.

2. It watches everything that happens on Earth and is specifically interested in humans’ good behavior and willingness to obey its rules.

3. It may have the ability to reward those who follow its rules after they die, and disincentivize those who violate them.

(I think the reality might be a little more complex than this, especially in the direction of us not being sure whether we are in a real universe at all or in one of the superentity’s simulations. I think that some superintelligence in some universe might be simulating various combinations of values in various contexts to decide which superintelligences are good bargaining partners and which ones aren’t. If I am kind to a beggar on the street, then maybe that convinces millions of intelligences in millions of distant universes that are somehow beggar-like to be friendlier to values that are somehow Scott-like. I still need to think this part through more.)

Are The Amish Unhappy? Super Happy? Just Meh?

I.

Recently on Marginal Revolution: Are the Amish unhappy?

> The average levels of life satisfaction [among the Amish] was 4.4; just above the neutral point…the Amish fall lower than members of many other groups. In a study of more than 13 thousand college students from 31 nations, for example, only students from Kenya (whose average life satisfaction was 4.0) scored lower than the Amish (Diener & Diener, 1995).

Sounds like Amish people are quite unhappy. This came as a surprise to me, since I'd heard from Jonah Lehrer and Business Insider that the average Amish person is as happy as the average non-Amish billionaire, proving once and for all that community and old-fashioned values are more important than money:

> As an illustration of the striking disconnect between money and happiness, the average life satisfaction of Forbes magazine's 400 richest Americans was 5.8 on a 7-point scale. Yet the average life satisfaction of the Pennsylvania Amish is also 5.8, despite the fact that their average annual salary is several billion dollars lower.

I actually care about getting this one right. There's a lot of discussion over whether modern society produces ennui, meaninglessness, atomization, etc – and whether our material wealth has really brought us happiness. The data tend to support a story where more modern and developed countries are happier, but not without some ambiguity and contradiction (a few Latin American countries seem to do better than much richer European ones). But there's always a concern that the least-developed areas today – like sub-Saharan Africa – are places that have absorbed the worst parts of modernity – like totalitarianism, pollution, and slums – but not the good parts like not-dying-of-cholera. The Amish are about as close as we can get to surveying 1700s-Europe. If we can figure out how happy they are, maybe it would tell us something new about the good life.

Unfortunately, we can't. This field is full of conflicting data, shifting methods, unreplicated surveys, and – and I didn't even realize this was a problem it was possible for a field to have – is super-confusing because everyone involved is named Diener. I tried to get the above-cited Diener & Diener 1995, but carelessly bought Diener, Diener & Diener 1995 instead. Cowen's source is a book by Robert Biswas-Diener, who is comparing a study by Diener & Diener to a study by Biswas-Diener, Vittersø, & Diener, the last of which was do-able only because:

> It was both coincidental, and helpful, that my surname — Diener — is also a relatively common Amish surname. This curious point of contact allowed me to introduce myself and my project.

Sure. Whatever.

The Marginal Revolution excerpt comes from Biswas-Diener, Vittersø, & Diener (after this: BDVD). This study was part of the authors' project to prove that most people are happy. In a previous paper, they had determined that most people in modern societies are happy; in this one, they look at three different traditional societies (Amish, Inughuit Eskimos, and

Maasai) to see if they are mostly happy was well. On a 1 – 7 scale, they find:

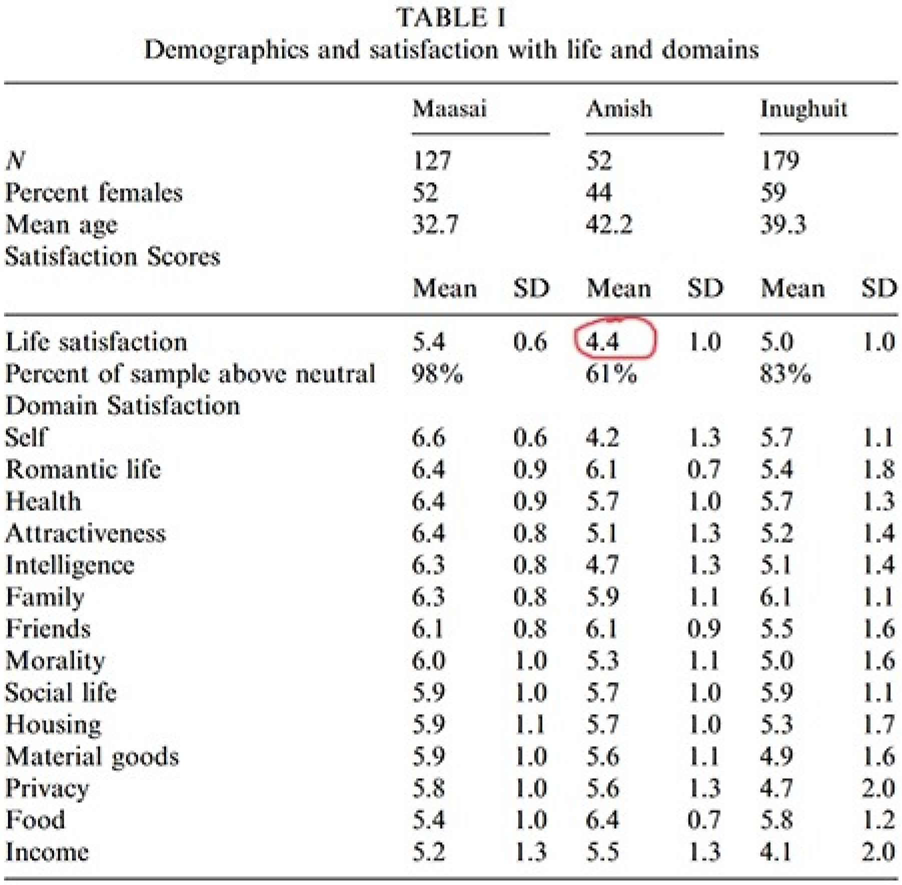


TABLE I
Demographics and satisfaction with life and domains

	Maasai		Amish		Inughuit	
N	127		52		179	
Percent females	52		44		59	
Mean age	32.7		42.2		39.3	
Satisfaction Scores						
	Mean	SD	Mean	SD	Mean	SD
Life satisfaction	5.4	0.6	4.4	1.0	5.0	1.0
Percent of sample above neutral	98%		61%		83%	
Domain Satisfaction						
Self	6.6	0.6	4.2	1.3	5.7	1.1
Romantic life	6.4	0.9	6.1	0.7	5.4	1.8
Health	6.4	0.9	5.7	1.0	5.7	1.3
Attractiveness	6.4	0.8	5.1	1.3	5.2	1.4
Intelligence	6.3	0.8	4.7	1.3	5.1	1.4
Family	6.3	0.8	5.9	1.1	6.1	1.1
Friends	6.1	0.8	6.1	0.9	5.5	1.6
Morality	6.0	1.0	5.3	1.1	5.0	1.6
Social life	5.9	1.0	5.7	1.0	5.9	1.1
Housing	5.9	1.1	5.7	1.0	5.3	1.7
Material goods	5.9	1.0	5.6	1.1	4.9	1.6
Privacy	5.8	1.0	5.6	1.3	4.7	2.0
Food	5.4	1.0	6.4	0.7	5.8	1.2
Income	5.2	1.3	5.5	1.3	4.1	2.0

They don't formally compare these to the modern societies numbers in the paper. The Biswas-Diener book linked by MR does compare them, finding the Amish are lower than every modern society except Kenya, but there are two important caveats.

First, the Biswas-Diener book compares the survey of Amish (mean age 44) to the Diener & Diener survey of college students in modern society. If college students are happier than 44-year-olds (they are), that's a potential confounder.

Second, the book notes that the Amish's self-reported total happiness is lower than their self-reported happiness with any individual facet of their lives. Just look at the table above – their romantic life is a 6.1, their health is a 5.7, their attractiveness is a 5.1 (those beards, right?) – but the two totals, self-satisfaction and life-satisfaction – are 4.2 and 4.4 respectively. It seems like they're averaging a bunch of numbers and getting an average lower than any of the individual inputs. This is especially bad since modern people tend to do the opposite; report an average happiness higher than their happiness with any individual part of their lives. Biswas-Diener guesses that modern people like to present themselves well (the "have a Facebook feed full of spectacular parties and meticulously-prepared plates of food even when your life is falling apart" effect), and traditional societies are more likely to value humility and treat pride as a sin. As far as I know, there are no studies that have ever measured a non-Amish society using this exact breakdown.

Third, as far as I can tell, the Diener & Diener paper doesn't actually show Kenyans having the lowest life satisfaction, or Kenyans having a life satisfaction of 4.0:

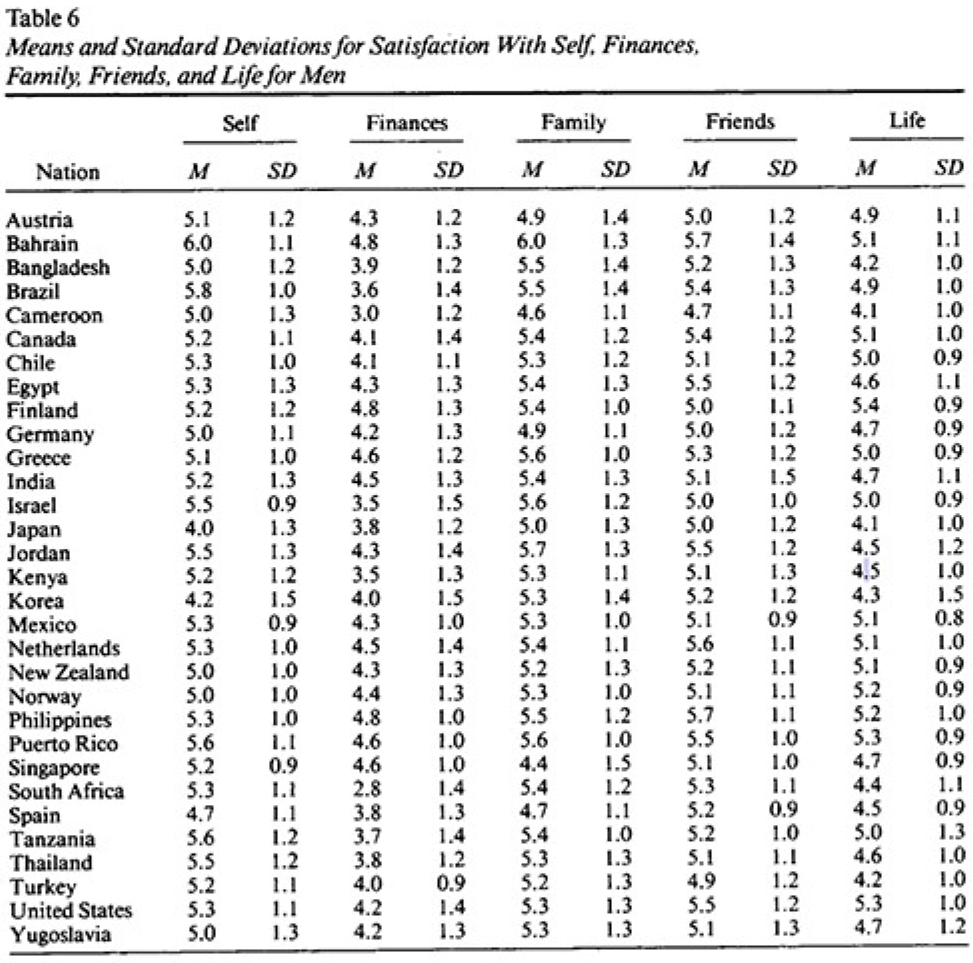


Table 6

Means and Standard Deviations for Satisfaction With Self, Finances, Family, Friends, and Life for Men

Nation	Self		Finances		Family		Friends		Life	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Austria	5.1	1.2	4.3	1.2	4.9	1.4	5.0	1.2	4.9	1.1
Bahrain	6.0	1.1	4.8	1.3	6.0	1.3	5.7	1.4	5.1	1.1
Bangladesh	5.0	1.2	3.9	1.2	5.5	1.4	5.2	1.3	4.2	1.0
Brazil	5.8	1.0	3.6	1.4	5.5	1.4	5.4	1.3	4.9	1.0
Cameroon	5.0	1.3	3.0	1.2	4.6	1.1	4.7	1.1	4.1	1.0
Canada	5.2	1.1	4.1	1.4	5.4	1.2	5.4	1.2	5.1	1.0
Chile	5.3	1.0	4.1	1.1	5.3	1.2	5.1	1.2	5.0	0.9
Egypt	5.3	1.3	4.3	1.3	5.4	1.3	5.5	1.2	4.6	1.1
Finland	5.2	1.2	4.8	1.3	5.4	1.0	5.0	1.1	5.4	0.9
Germany	5.0	1.1	4.2	1.3	4.9	1.1	5.0	1.2	4.7	0.9
Greece	5.1	1.0	4.6	1.2	5.6	1.0	5.3	1.2	5.0	0.9
India	5.2	1.3	4.5	1.3	5.4	1.3	5.1	1.5	4.7	1.1
Israel	5.5	0.9	3.5	1.5	5.6	1.2	5.0	1.0	5.0	0.9
Japan	4.0	1.3	3.8	1.2	5.0	1.3	5.0	1.2	4.1	1.0
Jordan	5.5	1.3	4.3	1.4	5.7	1.3	5.5	1.2	4.5	1.2
Kenya	5.2	1.2	3.5	1.3	5.3	1.1	5.1	1.3	4.5	1.0
Korea	4.2	1.5	4.0	1.5	5.3	1.4	5.2	1.2	4.3	1.5
Mexico	5.3	0.9	4.3	1.0	5.3	1.0	5.1	0.9	5.1	0.8
Netherlands	5.3	1.0	4.5	1.4	5.4	1.1	5.6	1.1	5.1	1.0
New Zealand	5.0	1.0	4.3	1.3	5.2	1.3	5.2	1.1	5.1	0.9
Norway	5.0	1.0	4.4	1.3	5.3	1.0	5.1	1.1	5.2	0.9
Philippines	5.3	1.0	4.8	1.0	5.5	1.2	5.7	1.1	5.2	1.0
Puerto Rico	5.6	1.1	4.6	1.0	5.6	1.0	5.5	1.0	5.3	0.9
Singapore	5.2	0.9	4.6	1.0	4.4	1.5	5.1	1.0	4.7	0.9
South Africa	5.3	1.1	2.8	1.4	5.4	1.2	5.3	1.1	4.4	1.1
Spain	4.7	1.1	3.8	1.3	4.7	1.1	5.2	0.9	4.5	0.9
Tanzania	5.6	1.2	3.7	1.4	5.4	1.0	5.2	1.0	5.0	1.3
Thailand	5.5	1.2	3.8	1.2	5.3	1.3	5.1	1.1	4.6	1.0
Turkey	5.2	1.1	4.0	0.9	5.2	1.3	4.9	1.2	4.2	1.0
United States	5.3	1.1	4.2	1.4	5.3	1.3	5.5	1.2	5.3	1.0
Yugoslavia	5.0	1.3	4.2	1.3	5.3	1.3	5.1	1.3	4.7	1.2

This is the male table. There's another one for women, but it's very similar, and neither the male table, the female table, nor the average of the two tables matches the claim that Kenya is 4.0 or that nobody else is less than 4.4. My guess is all the Dieners share preliminary data with each other, and Biswas-Diener is going off some older or unpublished version of this, but I'm not sure and I might just be missing something.

The public version of Diener & Diener does generally backs up the claim that 4.4 is kind of on the lowish side. But on the other hand, the Maasai number of 5.4 would be the highest one on the whole chart (tied with Finland), which would suggest there’s no clear traditional vs. modern society dichotomy.

So the result of this comparison seems to be “The Amish are above neutral happiness, but less happy than almost any modern society. On the other hand, the Maasai are more happy than almost any modern society. However, a lot of this could be about self-presentation, and the data for modern societies are kind of unclear.”

II.

What about the “Amish are as happy as billionaires” claim?

The billionaire numbers come from Diener’s Happiness Of The Very Wealthy and seem to check out. It finds billionaires are happier, though not vastly happier, than everyone else. Billionaires have an average happiness of 5.8 – remember, the highest national sample above was Finland at 5.4.

This time, the Amish numbers come from Diener and Seligman’s Beyond Money: Toward An Economy Of Well-Being, which presents this graph:

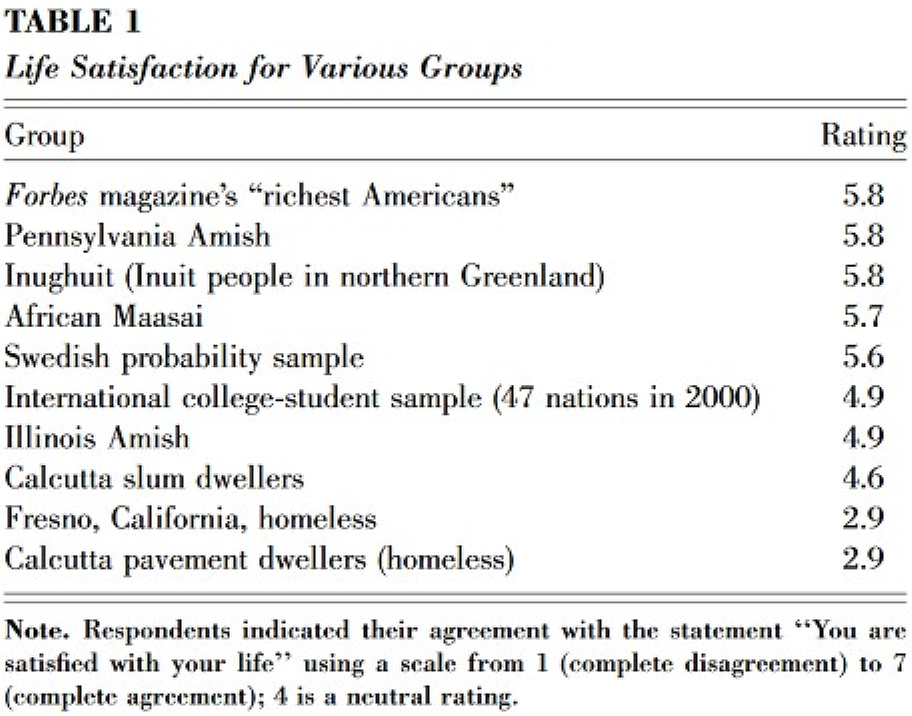


TABLE 1

Life Satisfaction for Various Groups

Group	Rating
Forbes magazine's "richest Americans"	5.8
Pennsylvania Amish	5.8
Inughuit (Inuit people in northern Greenland)	5.8
African Maasai	5.7
Swedish probability sample	5.6
International college-student sample (47 nations in 2000)	4.9
Illinois Amish	4.9
Calcutta slum dwellers	4.6
Fresno, California, homeless	2.9
Calcutta pavement dwellers (homeless)	2.9

Note. Respondents indicated their agreement with the statement ''You are satisfied with your life'' using a scale from 1 (complete disagreement) to 7 (complete agreement); 4 is a neutral rating.

It doesn't explain where the table comes from, but the use of the same three traditional societies (Maasai, Inughuit, and Amish) suggest the BDVD paper above (which shares one of its Dieners with the Beyond Money paper). But its numbers for all three groups are very different, and its headline result – the one about Pennsylvania Amish as happy as billionaires – isn't in the BDVD paper at all and has no citation.

The only clue to this discrepancy is that the Beyond Money paper cites the BDVD paper as "manuscript submitted for publication". Perhaps the peer reviewers made comments which caused BDVD to drop their Pennsylvania Amish result and analyze some of the other results differently? In any case, since the Amish = billionaires data seems to have been quietly dropped by the authors, we probably shouldn't put much stock in it.

Does this mean we should default to the “Amish less happy than anyone else except Kenyans” data? I say no. If the study authors got data consistent with “Amish are as happy as billionaires”, and later on it got changed to “Amish less happy than anyone”, plus they changed the Kenyan stuff around too, then I really don’t care about the “correction”, this research isn’t rigorous enough, or fixed enough, to convince me of anything.

All of this is from long before the replication crisis started improving methodologies, and I don’t trust it enough to consider it worth trying to smush everything together into a coherent whole. Sweep the billionaires/Kenyans issue under the rug, and there are still too many questions. Is there really a gulf between Pennsylvania Amish and Illinois Amish as vast as that between Swedes and Calcutta slum-dwellers? Are the Maasai really so much happier than modern societies, even as the Amish are so much less happy? Is there really an entire scientific field where everyone is named “Diener”? We *just don’t know*.

Highlights From The Comments On Twelve Rules

[Taken from the responses to this post.]

From sclmlw:

> While I don't agree with lots of Jordan Peterson, I think Scott fundamentally missed the boat in some of his criticisms because he systematically views things from a different perspective than Peterson, which was missed.
>
> From what I can tell, Peterson is intensely interested in the idea, "Everyone has the capacity to become a Nazi war criminal. What causes that phenomenon?" His answer, and the central driving idea of his philosophy, seems to be, "Anarchy/chaos is worse for society/humanity than horrific, unimaginable cruelty. So evolution pushed society to develop in a way that will always choose cruelty over chaos. Thus, if you were in Stalin's Russia, you'd run the gulags to stave off anarchy, and you'd kill hundreds of people if you had to. You may hate it, but it was required for humanity to soldier on, so it's what evolutionary forces produced." Peterson cares because he wants to understand how to steer societies away from the gulags and the killing fields.
>
> This appears to be the foundation of his philosophy, and you can understand a lot of what he talks about as an outgrowth of this idea. For example his answers to, Why do good things happen to bad people? He accepts that good things happen to bad people as a given. However, when good things happen to bad people, those people have a choice to make: should I give in to chaos and

ascribe meaninglessness, or should I accept some order of some kind? And Peterson would say that societies that would have chosen the chaos angle didn't survive. Evolution now has the choice as a built-in function, where you will always accept order of some kind. And that order could manifest in a number of different ways, such as the killing fields of Cambodia, but what matters is that if your choice is killing fields or chaos you'll choose killing fields every time. So Peterson wants to ensure we don't get to the point where that's the only choice left; he advises his acolytes not to destroy society (which leads to the chaos/authoritarian dilemma), but to recognize that they're going to choose to follow some order of some kind, and that they should therefore intentionally follow even a flawed societal order, because it's better than gas chambers and ethnic cleansing. The reality that "bad things happen to good people" shouldn't persuade them to tear down society and try starting all over again, because that leads to the chaos/tyrant choice, and we can't go there. So when bad things happen, you have to do your part to keep flawed society going, or else we get concentration camps. Go back and read the quotes above in that context and they all make sense. He's not trying to answer "why do bad things happen", he's trying to direct what he sees as an appropriate response to when they do.

This also directs his motivations when talking to people about his theory. Fundamentally, he has hypothesized a reason so many people in the 20th century became horrible, and he sees the current non-awful state of civilization as unstable. He sees trends he believes could tear down society, and cause people to spiral back to the point where they will be willing to do anything to stave off the chaos. In some of his videos he gets

> passionate, and in most cases he reserves his passion for this basic idea in some form or another: everyone has the capacity to become a Hutu killing Tutsis; you would do it, even if you think you're better than that; if you don't follow certain ideas, you (and society in general) will devolve into that awful state.
>
> I don't know if any of that is true, or if it's a different kind of psychobabble, but the fundamental observation that Scott is missing is that Peterson is thinking on a society-wide and philosophically-projected evolutionary development axis. Peterson's pronouncements flow from this angle. He's not thinking as a utilitarian or deontologist or consequentialist. He's thinking, "What do populations do in these situations, and how can we nudge populations away from mass torture/murder?" That's not utilitarian maximization, or negative utilitarianism. It's sort of like Nassim Taleb's concern about fat-tail events breaking fragile systems, and how to avoid that.

From theredsheep:

> Have not read Peterson, have seen videos of his lectures where he seems unremarkable and uninteresting. Maybe it's just that I'm not in the market for a belief system, I don't know.

I agree; I've also watched his videos and find them unwatchable and boring. I don't know why the book seems so much better – maybe I just respond more to the written word, or maybe he had a really good editor. I would wonder if this is what's behind the high variance in how people respond to him, except I think a lot of the people who absolutely love him are working off the videos. Weird. It takes all kinds, I guess.

From Dry Raven:

> As someone who is deep into the Peterson hole, I have a warning. Peterson is like an alien whose words travel through several layers of perceptual distortion before they come out to a regular human being. You think you understand what he means when he says something, but he means something entirely different. His words come out like static to people, and they make the mistake of thinking they understand his intent because the sentence still parses in English. But he's changed the meaning of all the words. Reading 12 rules for life is like watching him attempt to simulate a normal human being with hilariously punchy sentences, but maps of meaning and his absurd recursive diagrams is where you should go if you want to get an idea of what's really floating around in Peterson's head.

Also an interesting data point in the "why do people have such different responses to his different works?" question.

From Macruise on the subreddit:

> I think one could stress more the distinction between morality and meaning, because it seems central to the specific point Peterson is making. Alexander writes, for instance:
>
> *"I think he's saying – suffering is bad. This is so obvious as to require no justification. If you want to be the sort of person who doesn't cause suffering, you need to be strong. If you want to be the sort of person who can fight back against it, you need to be even stronger. To strengthen yourself, you'll need to deploy useful concepts like "God", "faith", and "Heaven". Then you*

can dive into the whole Western tradition of self-cultivation which will help you take it from there.”

No, he’s not merely saying ‘suffering is bad’. He is saying that it matters. Think about hell. One thinks, from one’s armchair and slippers, that there’s something conceptually fishy about hell, about the very idea of infinite suffering. One wants to say ‘Well, that would just be the new normal, and you’d adapt. It too will become devoid of meaning’. Habituation is such a common experience that you tend to think ‘That too shall pass’ about everything. Even sex, which you’d think would always mean something given its centrality to our animal existence, gets quotidian, just another 17 bus. But pain never gets old. It’s always meaningful. You cannot write The Myth of Sisyphus whilst getting waterboarded, and it’s not merely because the keyboard would get wet.

So it’s not ‘negative utility’ that Peterson is shooting for; it’s ‘negative meaning’. ‘Positive meaning’ is where someone comes along and says ‘Here’s what you should try to achieve in your life because it is inherently meaningful’. ‘Negative meaning’, however, is about the avoidance of suffering, from which certain hypothetical imperatives (subject to our best psychological and medical knowledge) spring.

The good news here is that this notion of ‘negative meaning’ can be divorced from the specifics of Peterson’s injunctions insofar as they’re based in Jungian woo (which they pretty much always are, as far as I can see). The general idea – that pain matters, and you want to live your life to avoid it as much as you can – seems fairly solid as an answer to existential angst.

The bad news is that I think Peterson is wrong about life being suffering. It’s certainly true that experiencing suffering will solve the meaning crisis, and that looking to avoid suffering is a solid plan. But the problem with it as a solution to the problem of meaning is its contingency, made worse by the fact that we live in a world where suffering has no real urgency. The reality is that it is entirely possible now to live your life in a way where true suffering is very delayed. You can make it all the way to 60 without experiencing a serious loss (parent, partner sibling, child), without experiencing painful illness, and without having any existential crisis. These people actually exist, as much as you want to tell yourself otherwise in order to avoid resentment.

If you’re the sort of person who has bipolar disorder and is in a manic phase, it might seem reasonable to say something like ‘life is suffering’. Depressed people too understand that all-powerful-but-transitory truth in their bones. But even those who have had a period of suffering tend to forget how bad it was, and you tend to look back on those periods as times where you were strapped in to the rollercoaster ride, and that you knew all along that it too would pass (even though at the time you didn’t, and that’s what was really awful about it – the sheer, unadulterated despair of thinking ‘this is it’).

Thus, I don’t think negative meaning works. Life isn’t suffering. That’s only a contingent fact that’s true of certain people at certain points of their lives. In general, we have it so good that we invent problems for ourselves and tell ourselves it’s real suffering, and indulge in phoney culture-wars where we can pretend that the damn postmodern, neo-marxist superjews are taking over the world and need to be stopped and that’s that. We all know in our hearts that it’s a load of shit,

> and the world will carry bumbling on, and that we'll all die of something insultingly banal, doped to the eyeballs on morphine. We won't, as our emo selves tell us, be lonely tonight, and lonely tomorrow; we'll be made comfortable.
>
> Thus, I find myself stuck in the absurd. I know that I'm the sort of being that has to have a purpose, and I don't believe that one can invent it as Sartre enjoined us to do, but the universe doesn't owe me an answer, and doesn't seem disposed to giving me one. I accept that this is, in some sense, bad for me, but one cannot tell oneself that the lines are equal in length and then be able to see it that way.

I appreciate the clarification, but I don't know if I agree.

Is this the same as just saying that you don't need a logically coherent life goal, because your biological drives will goad you forward whether you have the goal or not? I'd agree with that one – except maybe in depression, when biological drives suddenly become a lot less convincing. But one of the questions of meaning is "why should I do anything other than follow my immediate biological drives?"

I can imagine Peterson saying (though I don't know if he would really endorse this) that living a good life is just satisfying your biological drives in an intelligent way. You want to become a good person because then you're less likely to end up low-status in a way that frustrates your drive for belonging and sex and so on. The argument that moral behavior is really the selfishly best thing to do over the long term is popular, venerable – and way too convenient for me to ever believe.

I find the strongest argument against this to be the question of suicide. Peterson is against it, for vague cliched reasons. But it cuts through all of this “well I guess I have to follow my biological drives” stuff with great finality. You really don’t. If your reason for living doesn’t present the argument against suicide, it’s not much of a reason for living.

From Roe_ on the subreddit:

> *“Jordan Peterson’s superpower is saying cliches and having them sound meaningful.”*
>
> Yes – this is absolutely his super-power. How did he get it?
>
> Because he lives the cliches. The cliches are easy to say, but they’re difficult as fuck to actually implement in your life. And, unless you’ve done so, you’re just talking shit.
>
> As far as I can tell, Peterson has done so. Which is why he “speaks with authority”.
>
> “Rescue your father from the underworld” means precisely this – try to live the old wisdom (ie. the cliches) – which wasn’t easy for our fathers, and sure isn’t any easier now.

This is a good explanation of the “rescue your father from the underworld” thing, and reminds me of my despair about every generation having to learn everything all over again. It’s also why I get fed up with accusations of “that’s not original” or “you’re just reinventing the wheel”.

The part about living the cliches making you able to say them more confidently sounds nice, but is it really true? What would the mechanism for this be? Also, lots of the most charismatic preachers and cult leaders are moral catastrophes. And if this is sort of like saying “he’s so authentic”, I don’t know how that idea interacts with the “actually an alien” point above.

I think of Part I of Liber ABA as being about the process by which prophets become able to speak with confidence, and I sort of wonder whether a lifetime of very successful Jungian analysis might be able to accomplish something similar.

Rocket discusses the inevitable AI angle:

> A bit of an aside from Peterson, but relevant to the idea of conflicting Order-Chaos drives and uncertainty-minimization: There was a paper about a month ago about getting AI agents to learn new and interesting behaviours, which touched surprisingly close to that.
>
> Background: Lots of people have tried “reinforcement learning”, where you plop an AI agent in some environment in which it can take actions, and doing certain things nets it a ‘reward’, then you let it do its own thing to try to learn how to maximize that reward. The problem with this is that if reward is rare, as it can be in many scenarios, the AI gets very few opportunities to learn from it. Imagine a game where such an AI only got a reward for finishing a complex level, but no reward for incremental progress: The odds of it coincidentally stumbling into that reward are extremely low, and it will take it a vast, vast amount of time and random button-mashing to join up the dots and figure out which chains of actions got it to that state, and so learn how to get that reward more efficiently in future.
>
> The paper took a different approach: Rather than purely seeking out reward, the AI agent should give itself its

own goal, namely to understand how to manipulate the environment it’s in. They achieved that goal (and here’s where the Order-Chaos part comes in) by giving it two components. The first component looks at the input it’s getting (i.e. what the agent can see in the virtual world it’s in) and the action (i.e. what “buttons” the agent is pushing). It then tries to predict what the agent will see next – in other words, to predict the consequences of that action.

The fun part comes from the second component, though, which has an adversarial relationship with the first. It’s the bit that gets to choose actions, and it chooses actions so as to maximally challenge the first component. It’s the Chaos, in this analogy, but it’s also the driver of learning, because if the first component was in control, and the only goal was to minimize uncertainty, it would just sit still and stare at a blank wall. With the two components, though, the AI effectively “learns through play”, starting by moving itself around until the predictor component has mastered that before focusing on objects in the scene, and specifically how to toss them around in maximally confusing ways. (Insert your own analogy to toddlers here)

As well as being an interesting piece of computer-psychology (this is sort of becoming a field lately), the longer-term hope is that such agents will build a model of their environment even in the absence of external reward signals, and so when reward arrives, they’ll have some higher-level abstractions ready to go (“Maybe I need to move the blocks onto the button”) rather than having to painstakingly bootstrap their world-model from raw pixels and rare rewards. In other words, such

> agents may be far more effective at getting reward than ones that are solely motivated by getting reward.

From Jacek Lach:

> *"But that's exactly the problem. I worry Peterson wakes up in the morning and thinks "How can I help add meaning to people's lives?" and then he says really meaningful-sounding stuff, and then people think their lives are meaningful. But at some point, things actually have to mean a specific other thing. They can't just mean meaning. "Mean" is a transitive verb. It needs some direct object."*
>
> I would contest that! Yes, for a perfectly rational agent that is the case; you can just look at the effects of your life, see that it is on net positive, and call that 'meaning'.
>
> But what we really care about when talking about the people, is the *impression of meaning*. It is not important for your wellbeing whether your life is actually rational-meaningful. It is important if your system 1 is satisfied with your life in the certain ways, that it signals this satisfaction in certain ways that we call 'meaningful'.
>
> As such, it's not strictly necessary to tie your meaning to some objective measures. If your life feels meaningful, and you have reasonable expectation that this feeling will persist (so presumably – if you accept the assumption that this feeling of meaning evolved, culturally or genetically, for 'a reason', i.e. meaning is adaptive – just cheating system1 with drugs is not the right long term answer), and not causing too many negative externalities… Then you're good?

> Obviously meaningfullness is not the only thing one should optimise their life for. But it’s definitely one of the things you should be looking at.

This helps bring into focus one of the things I wasn’t able to discuss in the original review.

When we say we want our lives to be meaningful, are we saying we want our lives to actually have meaning? Or that we want to feel and act as if our lives have meaning?

Before virtuous people like ourselves slam our fists on the table and insist on actual meaning, keep in mind – don’t we have actual meaning already? Most of us would endorse something like helping others as being inherently meaningful – we might add things like creating great art, discovering new scientific truths, and the like. If we could reprogram ourselves like robots, a lot of us would just program ourselves to do whatever helps others or achieves some conception of human values most efficiently, then say “problem solved”. If there’s something left after realizing we can do that, it’s not wondering what the meaning of life is, it’s having some kind of vague emotional will to go on.

This is one reason I respect Jordan Peterson’s pragmatism on a pragmatic level, even as I think it’s a crappy theory of truth. I can imagine a version of him saying (I don’t know if the real one does) “Look, I’m giving you all of these inspirational slogans. You can pick my science and philosophy and mythography apart if you really want, but are you sure you want to do that? You’ll just ruin my attempt to inspire you, and go back to lying on the couch all day wishing you had a reason to get up in the morning.”

Maybe aliens would view it as a tragic quirk of the human psyche that we have to conflate inspiration and truth.

Maybe to them, inspiration is just another genre, closer to art or poetry than to an attempt to describe the world as it is. Maybe to them, if there’s an intuitively satisfying explanation of the meaning of life, asking “Is that really the meaning?” or “Is that really true?” would be just as stupid and annoying as nitpicking the lyrics of Ode To Joy. “Ode to Joy says ‘all creatures drink of joy’, but some creatures are unhappy, and joy is not a liquid! Politifact rates your symphony FALSE.”

It’s disappointing that nobody frames it this way: “Inspiring things should be taken as a work of art and not judged on their truth value”. Instead, it’s always some formulation like “Inspiring things are true in a way different from the way factual claims are true”, at which point I have to interject that truth is a useful word and insist on defending its “factually correct” meaning.

Karl Smith on Twitter:

> Scott quibbles with Peterson’s tendency to waffle between pragmatism and platitudes. But, this waffling is actually the correct answer.

Maybe I’m misunderstanding this, but I interpret it as sort of related to the above. If we can waffle between factual truth and inspirational things in a way that lets us blur the distinction, so that we feel suitably inspired without necessarily believing false stuff in a way that could ever be pinned down, is this good or bad? Should we interpret this as an intellectual failure, or a cool skill that lets you do more than rationality alone? What about the claim that “rationalists should win”?

My own position is to question whether anyone is really good enough at this not to let their inspirational beliefs bleed over into the factual world. I discuss this a little bit here.

And my position on the larger problem of meaning is to notice that my life always seems really meaningful and great when I have coffee. If I'm going to try to figure out what the actual meaning of life is, in some sort of deep principled way, I'm going to do it with as much attention to Truth as possible. And if I'm going to give myself some emotional hack that lets myself go on and continue finding life worth living, I think caffeine probably has fewer side effects than falsehood, and is just as effective.

And if you don't respond to caffeine as well as I do, then I think the overall lesson is that the *emotional* problem of meaning is a basically biological one, that doesn't connect with the *philosophical* problem of meaning nearly as much as you think. Get a good psychiatrist and you'll solve the emotional problem. The philosophical problem might not be solvable, but "helping others" or "creating a positive singularity" or "[your ingroup's political goals here]" are, though not Perfectly Objectively Grounded, grounded enough that most people don't really want to question them once the emotional problem is solved.

I find some of Peterson's non-truth-value-having writing effective in the same sense as caffeine; it makes me more emotionally willing to follow the truths I know I should be following. Since, jokes aside, I can't literally be drugged 100% of the time, I appreciate that. And maybe the drug would be stronger if I were to swallow his truth-value-having claims too. But that's not a risk-benefit profile I'm okay with right now.

Userfriedlyyy writes:

> My problem with Jordan Peterson’s world view is that he completely ignores the structural problems in the way of self improvement. It’s all well and good to do everything in your power to make yourself stronger and improve your lot in life but it risks turning all of his devotees into fervent believers in the Horatio Alger myth. There will always be a segment of society that can pull themselves up by their own bootstraps; but it is an absolute certainty that it will not work for everyone. Then the problems arise when all those Horatio Alger’s go ‘I did it my self, your moral failings must be why you didn’t so you deserve what you get.’
>
> It just feeds into the individualistic narrative that everything about your life is 100% the result of things you have control over. Just ignore why your wages aren’t going up, just work harder to get paid more. Just ignore that you are making less than your parents when they were your age, just keep working harder to get paid more.
>
> It completely ignores workers banding together to demand better treatment. I’m not even against some of the self improvement stuff but I cringe when it is offered as the sole solution.

Peterson turns Marx on his head and claims that political activism is the opiate of the masses. That is, it’s something people use to make themselves feel sort of vaguely good and self-satisfied, but which prevents them from engaging in the actually important work of spiritual struggle.

My interpretation of him (can’t be sure it’s right) says that he is worried that there are problems with society, and all else being equal he would like people to solve them. But he has the psychoanalyst’s usual worry that anything which is not the Work will be a defense mechanism that people use to avoid the Work. Here again I find a comparison with Lewis helpful (this is from his demon character Screwtape's advice on how to tempt humans):

"I had not forgotten my promise to consider whether we should make the patient an extreme patriot or an extreme pacifist. All extremes, except extreme devotion to the Enemy [God], are to be encouraged. Not always, of course, but at this period. Some ages are lukewarm and complacent, and then it is our business to soothe them yet faster asleep. Other ages, of which the present is one, are unbalanced and prone to faction, and it is our business to inflame them. Any small coterie, bound together by some interest which other men dislike or ignore, tends to develop inside itself a hothouse mutual admiration, and towards the outer world, a great deal of pride and hatred which is entertained without shame because the "Cause" is its sponsor and it is thought to be impersonal."

I think Peterson assumes that a psychologically undeveloped person starting to dabble in politics will be eaten alive by various virulent memes, chewed up, and spit out as a Hofferian True Believer in about five minutes. At best they will end up as an never-shutting-up slacktivist who calls people out for not changing their profile picture on Facebook quickly enough, and at worst as some kind of totalitarian. I think he would argue there's a vicious cycle here – the less psychologically developed you are, the more political activism will destroy you, and the more political activism destroys you, the less likely you are to ever psychologically develop further.

One of his twelve rules, "Set Your House In Perfect Order Before Criticizing The World", is about this, and doesn't preclude the possibility of getting involved in politics after you’ve sorted out your own life. I don’t know what this is supposed to mean, since presumably this is an eternal project that is never completed. Clearly Peterson himself thinks he’s at the point where he can participate in politics, so I don’t know.

Do I agree with him here? From a consequentialist point of view, what would it mean to get the least psychologically developed 50% of people out of political activism? If you’re a mistake theorist, it might be great – it takes an equal number of people away from both sides, but raises the quality of discourse. If you’re a conflict theorist, it might be awful – it decreases the number of troops available to the People in their struggle to overcome inertia and fight the Elites.

Rather than try to resolve that, I would just note that “Jordan Peterson saying psychologically underdeveloped people shouldn’t get involved in politics” does not remove the least psychologically developed people from politics. It removes from politics some group of people weighted towards reading Jordan Peterson, being psychologically underdeveloped, and having enough humility to realize that they might be psychologically underdeveloped (which is itself possibly a sign of not being underdeveloped). Whether or not you think this is worth it depends on your opinion of the average Peterson reader.

Philipp questions my use of Lewis:

> Scott, I’m surprised no one on this thread (I’ve not looked at the sub-reddit) has taken you up on the C.S. Lewis angle, so I guess, as something of a Lewis fan, that I’ll bite. I’ll state at the outset that I’ve not read Jordan Peterson, though I have watched a few clips of him and read quite a bit about him.

I realize that your talk of C.S. Lewis ‘hating’ Jordan Peterson is hyperbole, but it’s well worth realizing that he was actually a man of broad friendships and, much though he later rejected his youthful flirtations with non- or anti-Christian ideas, a broad personal experience. I mean, the man had variously been a dabbler in the occult (something he rejected with particular ferocity, after one of his spiritualist friends went insane), an idealist, and an atheist materialist, before he became a Christian, in part through the long intellectual influence of George MacDonald and G.K. Chesterton (two very different figures) and in part because of a late-night conversation with Hugo Dyson and the very Catholic J.R.R. Tolkien. Of his friends throughout his life, one of the closest was the anthroposophist Owen Barfield, and another, Charles Williams, was, well, it’s rather hard to figure out, but at least eccentric.

In fact, Peterson sounds, whenever I hear him described, as a kind of prophet of what Lewis called ‘the Tao’ in The Abolition of Man. Lewis is not, so far as I can see it, using the word in the technical sense–this is not Tao as ancient Chinese doctrine–but instead borrowing a word to refer to the basic sense of right and wrong, truth and goodness, and, at least to some degree, beauty and order as reflections thereof. It is, in other words, the natural law, understood not as an arcane system for judging morality in the abstract (the unfortunate impression some less-talented writers leave), but as it is really is: the basic moral order imprinted in the universe itself and in the hearts of mankind.

The central problem of The Abolition of Man (and a central problem also of the novel That Hideous Strength, which was likewise published in 1943 and is very close to it in thought) is the failure of modern education to teach that sense of right and wrong: its tendency to produce ‘men without chests’, or people whose actions are not grounded in the law of nature, in what is true and good, but ultimately only in their own preferences. From this perspective (or so I infer–I don’t have the book in front of me and I don’t remember whether he treats any kind of sophisticated utilitarian thought), utilitarianism, though it purports to judge what is moral, must fail, as it is not ultimately beholden to what is actually right in and of itself. At best, the utilitarian, despite his doctrine, still acts on the basis of his ingrained sense of rightness, and that sense of rightness needs to be shaped to the Tao, or it is nothing.

So, where does this leave Jordan Peterson? Not, I think, straightforwardly on the outside. I doubt Lewis would see in him a proto-Christian, since Peterson knows scripture and rejects Christian orthodoxy, but he is at least–or so it seems–a man with a strong apprehension of right and wrong, who is trying to grow the chests of people, especially young men, whose morality has been stunted. He is, as it were, on the side of the Tao.

To any Christian, and Lewis was one, there is a danger here, of course, and that is that Peterson is nevertheless still not on the side of God. Morality simply is not, because it cannot be, a matter of self-help and self-improvement. There, I think, your own characterization of Lewis goes wrong. He is not part of a “vast humanistic self-cultivation tradition”; though he does draw on ancient philosophy, he is a Christian lay theologian, and therefore believes in human frailty and the need for divine grace. ‘We have not got to try to climb up into spiritual life by our own efforts’, he wrote in Mere Christianity. ‘It has already come down into the

> human race. If we will only lay ourselves open to the one Man in whom it was fully present, and who, in spite of being God, is also a real man, He will do it in us and for us.'
>
> One cannot speak for a dead man, and others know Lewis's works better than I do. Nonetheless, I suspect that he would see in Jordan Peterson a man who sees clearly some of the worst errors of his own age, but who can only be a kind of stop on the way to the full truth, who is (of course) Christ himself. Like the world's moral traditions which, Lewis believed, reflected the truth of the Tao, he might prepare for the Gospel; but making of his teaching a Gospel–treating it, that is, as if it really could make people good enough–would simply be to pave yet another path that leads away from God and real health of the soul.

I acknowledge his greater expertise on Lewis scholarship. I base my concerns on two points. First, although Lewis seems broadly friendly to non-Christians who approximate Christian morality, he is broadly hostile to people who do that and say "And this, not that boring literal Jesus stuff, is the true essence of Christianity".

Again quoting Screwtape Letters:

"When the humans disbelieve in our [demons'] existence we lose all the pleasing results of direct terrorism and we make no magicians. On the other hand, when they believe in us, we cannot make them materialists and sceptics. at least, not yet. I have great hopes that we shall learn in due time how to emotionalise and mythologise their science to such an extent that what is, in effect, a belief in us, (though not under that name) will creep in while the human mind remains closed to belief in [God]. The "life force", the

worship of sex, and some aspects of Psychoanalysis, may here prove useful. If once we can produce our perfect work — the Materialist Magician, the man, not using, but veritably worshipping, what he vaguely calls "forces" while denying the existence of "spirits" — then the end of the war will be in sight."

I don't think Peterson exactly fits this description, but the idea of the pseudo-religious person who uses psychoanalysis but doesn't quite believe in God is suggestive enough to make me think Lewis would at least be a little uncomfortable.

From St. Rev on Twitter:

> I sort of see Peterson as the Meaningness of the normies.

One could do much worse.

Adult Neurogenesis – A Pointed Review

[I am not a neuroscientist and apologize in advance for any errors in this article. A recent study came out contradicting some of the claims mentioned here. See here for study, this comment for some discussion, and entry #20 on my Mistakes page.]

Hey, let's review the literature on adult neurogenesis! This'll be really fun, promise.

Gage's Neurogenesis In The Adult Brain, published in the *Journal Of Neuroscience* and cited 834 times, begins:

> A milestone is marked in our understanding of the brain with the recent acceptance, contrary to early dogma, that the adult nervous system can generate new neurons. One could wonder how this dogma originally came about, particularly because all organisms have some cells that continue to divide, adding to the size of the organism and repairing damage. All mammals have replicating cells in many organs and in some cases, notably the blood, skin, and gut, stem cells have been shown to exist throughout life, contributing to rapid cell replacement. Furthermore, insects, fish, and amphibia can replicate neural cells throughout life. An exception to this rule of self-repair and continued growth was thought to be the mammalian brain and spinal cord. In fact, because we knew that microglia, astrocytes, and oligodendrocytes all normally divide in the adult and respond to injury by dividing, it was only neurons that were considered to be refractory to replication. Now we know that this long accepted limitation is not completely true

Subsequent investigation has found adult neurogenesis in all sorts of brain regions. Wikipedia notes that "In humans, new neurons are continually born throughout adulthood in two regions of the brain: the subgranular zone and the striatum", but adds that "some authors (particularly Elizabeth Gould) have suggested that adult neurogenesis may also occur in regions within the brain not generally associated with neurogenesis including the neocortex", and there's also some research pointing to the cerebellum.

Some research has looked at the exact mechanism by which neurogenesis takes place; for example, in a paper in *Nature* cited 1581 times, Song et al determine that astroglia have an important role in promoting neurogenesis from FGF-2-dependent stem cells. Other research has tried to determine the rate; for example, Cameron et al (1609 citations) find that there is "a substantial pool of immature granule neurons" that may generate as many as 250,000 new cells per month. Still other research looks at the chemical regulators – a study by Lie et al, cited 1312 times, finds that Wnt3 signaling is involved.

(which is making you more nervous – the fact that I keep emphasizing how many citations these studies have, or the fact that one of the principal investigators is named "Lie"?)

But the most exciting research has been the work identifying the many important roles that neurogenesis plays in the adult brain – roles vital in understanding learning, memory, and disease.

Snyder et al (775 citations) finds "a new role for adult neurogenesis in the formation and/or consolidation of long-term, hippocampus-dependent, spatial memories." Dupret et al go further and find that "spatial relational memory *requires* hippcampal adult neurogenesis". Aimone et al (633 citations) find “a possible role” for adult neurogenesis in explaining the “temporal clusters of long-term episodic memories seen in some human psychological studies”. And Jessberger et al (506 citations) finds a role in object recognition memory as well.

In terms of learning, one of the major studies was Gould et al in Nature Neuroscience (2207 citations) finding that Learning Enhances Adult Neurogenesis In The Hippocampal Formation. Lledo et al (1288 citations) find that neurogenesis plays a part in explaining the brain’s amazing plasticity, and is “highly modulated, revealing a plastic mechanism by which the brain’s performance can be optimized for a given environment”. Clemenson et al (17 citations) find that “from mice to humans”, enviromental enrichment improves neurogenesis, and this “may one day lead us to a way to enrich our own lives and enhance performance on hippocampal behaviors”.

But I’ve always been most interested in the link with depression. In 2000, Jacobs et al published Adult Brain Neurogenesis And Psychiatry: A Novel Theory Of Depression (961 citations). It’s important enough that I want to quote the whole abstract:

> Neurogenesis (the birth of new neurons) continues postnatally and into adulthood in the brains of many animal species, including humans. This is particularly prominent in the dentate gyrus of the hippocampal formation. One of the factors that potently suppresses adult neurogenesis is stress, probably due to increased glucocorticoid release. Complementing this, we have recently found that increasing brain levels of serotonin enhance the basal rate of dentate gyrus neurogenesis. These and other data have led us to propose the following theory regarding clinical depression. Stress-

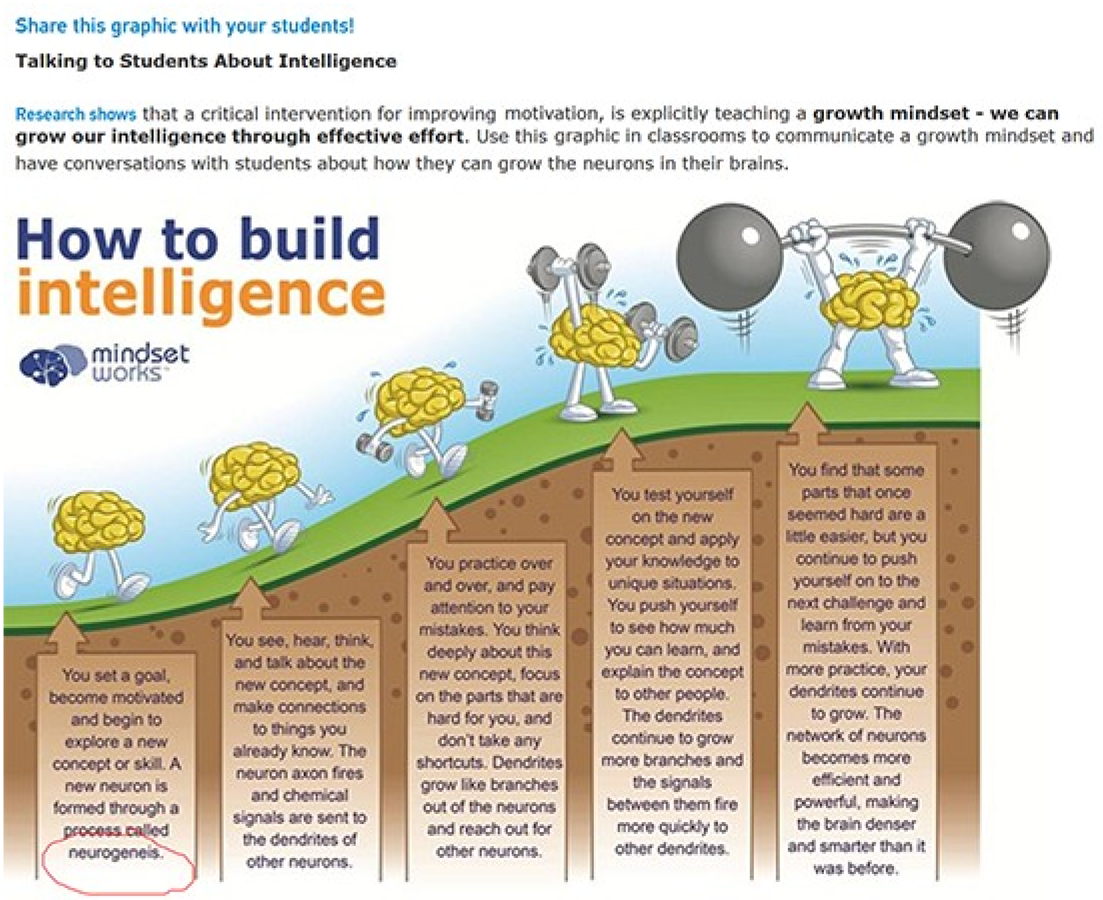
> induced decreases in dentate gyrus neurogenesis are an important causal factor in precipitating episodes of depression. Reciprocally, therapeutic interventions for depression that increase serotonergic neurotransmission act at least in part by augmenting dentate gyrus neurogenesis and thereby promoting recovery from depression. Thus, we hypothesize that the waning and waxing of neurogenesis in the hippocampal formation are important causal factors, respectively, in the precipitation of, and recovery from, episodes of clinical depression.

This theory got a boost from studies like Duman et al (522 citations), which found that antidepressant drugs like SSRIs upregulated neurogenesis – could this be their mechanism of action? And Ernst et al (327 citations) find that “there is evidence to support the hypothesis that exercise alleviates MDD and that several mechanisms exist that could mediate this effect through adult neurogenesis” – ie the antidepressant effects of exercise seem to work this way too. Electroconvulsive therapy, the most effective known treatment for depression? Works by promoting adult neurogenesis, at least according to Schloesser et al.

Is there *anything* that doesn’t have important neurogenesis-related effects? It would seem there is not. Sex, for example, “promotes adult neurogenesis in the hippocampus, despite an initial elevation in stress hormones” according to Leuner et al (124 citations). Drug addiction is modulated by neurogenesis. We need rock n’ roll to complete the triad, so here’s Music Faciliates The Neurogenesis, Regeneration, and Repair of Neurons.

A study in Nature Neuroscience that garnered over 3000 citations found that running increased neurogenesis. The popular science press was quick to notice. A slew of exercise-neurogenesis studies spawned articles like Psychology Today’s More Proof That Aerobic Exercise Can Make Your Brain Bigger. Dr. Perlmutter (“Empowering Neurologist!”) has a video about how you can Grow New Brain Cells Through Exercise. After this the pop sci world might have gotten a *little* carried away, until neurogenesis controls everything and is controlled by everything in turn. Slimland (of course there’s a site called Slimland) has a How To Grow New Brain Cells And Stimulate Neurogenesis page, suggesting you can “set yourself free and start flying” by removing toxins, eating a ketogenic diet, and meditating. Naturalstacks.com boasts 11 Proven Ways To Generate More Brain Cells, Improve Memory, And Boost Mood, which advises…really? Do you really want to know what it advises? Come *on*.

Also, growth mindset. Of *course* growth mindset. Carol Dweck’s Mindsetworks helpfully provides an infographic for teachers, urging them to tell their students that each time they set a goal or become motivated to learn a new skill, “a new neuron is formed through a process called neurogeneis [sic]”.



Share this graphic with your students!

Talking to Students About Intelligence

Research shows that a critical intervention for improving motivation, is explicitly teaching a **growth mindset - we can grow our intelligence through effective effort**. Use this graphic in classrooms to communicate a growth mindset and have conversations with students about how they can grow the neurons in their brains.

How to build intelligence

You set a goal, become motivated and begin to explore a new concept or skill. A new neuron is formed through a process called neurogeneis.

You see, hear, think, and talk about the new concept, and make connections to things you already know. The neuron axon fires and chemical signals are sent to the dendrites of other neurons.

You practice over and over, and pay attention to your mistakes. You think deeply about this new concept, focus on the parts that are hard for you, and don't take any shortcuts. Dendrites grow like branches out of the neurons and reach out for other neurons.

You test yourself on the new concept and apply your knowledge to unique situations. You push yourself to see how much you can learn, and explain the concept to other people. The dendrites continue to grow more branches and the signals between them fire more quickly to other dendrites.

You find that some parts that once seemed hard are a little easier, but you continue to push yourself on to the next challenge and learn from your mistakes. With more practice, your dendrites continue to grow. The network of neurons becomes more efficient and powerful, making the brain denser and smarter than it was before.

So it’s no surprise that researchers in the area are calling adult neurogenesis “one of the most exciting and rapidly evolving areas of research in the field of neuroscience”.

II.

Fun fact: there’s no such thing as adult neurogenesis in humans.

At least, this is the conclusion of Sorrells et al, who have a new and impressive study in *Nature*. They look at “59 post-mortem and post-operative slices of the human hippocampus” and find “that recruitment of young neurons to the primate hippocampus decreases rapidly during the first years of life, and that neurogenesis in the dentate gyrus does not continue, or is extremely rare, in adult humans." Also, the subgranular zone, the supposed part of the brain where neurogenesis begins, isn't even a real structure.

I am not a neuroscientist and am unqualified to evaluate it. But the Neuroskeptic blog, which I tend to trust in issues like this, thinks it's legit and has been saying this for years. Ed Yong from *The Atlantic* has a really excellent review of the finding that interviews a lot of the major players on both sides and which I highly recommend. Both of these reinforce my feeling that the current study makes a really strong case.

I was kind of floored when I saw this, in a way that I hope I was able to replicate in you by preceding this with the literature review above. How do you get so many highly-cited papers speaking so confidently about every little sub-sub-detail of a phenomenon, if the phenomenon never existed in the first place?

As far as I can tell, this was entirely innocent, well-intentioned, and understandable. It happened like this:

Adult neurogenesis was discovered in rats. This was so surprising, and such a violation of established doctrine, that it quickly became one of the most-investigated areas in neuroscience. Hundreds of studies were done on rats to nail down every little detail of the process.

The work was extended to many other mammals, to the point where it seemed inevitable that it must be true of humans as well. This was difficult to test because the relevant studies involve dissecting brains, and there aren't that many human brain specimens available with the necessary level of preservation. After a lot of work, a few people got a couple of brains, did some very complicated and contamination-prone tests, and found evidence of adult neurogenesis. This encouraged everyone to assume that the things they had discovered about rat neurogenesis were probably true in humans as well, even though they could never prove them directly because of the difficulty of human experimentation. Later some other researchers tried to replicate the complicated and contamination-prone tests and couldn't find adult neurogenesis in humans, but everyone assumed they had just messed up some aspect of the complicated testing process.

And to complicate matters, everyone in the new study has been very careful to say they can't prove with certainty that zero adult neurogenesis occurs – just that the levels are so low and hard to detect that they can't possibly matter. Looking back on some past studies, it seems that "so low and hard to detect that they can't possibly matter" was actually within their confidence intervals. So it may be that some team found some extremely tiny and irrelevant population of immature neurons in the brain, gave a confidence level that included that number, and then everyone just assumed we were talking about levels similar to the ones we saw in rats.

With real scientists taking not-entirely-sufficient care to distinguish rat from human results, the popular press felt licensed to totally ignore the distinction (did you even notice which of the studies in Part I were done on which species? Don't worry, nobody else did either).

Meanwhile, synaptogenesis – the growth of new synapses from existing nerve cells – was getting linked to depression and all kinds of other things in a lot of interesting studies. When people started talking about neurogenesis' role in depression, psychiatrists like me who have trouble keeping words ending with -genesis separate just sort of nodded and said “Oh, yeah, I heard about that” and didn’t give it the sort of scrutiny it deserved.

(I wonder if this is young-earth creationists’ problem too)

So it’s not like any one person made a spectacular mistake anywhere along the lines. Most of the studies done were in rats, and 100% correct. A few studies were done in humans, and may have gotten the wrong answer in a very difficult domain, while also hedging their bets and admitting they were trying something hard. It was only on a structural, field-wide level that all of this came together into people just assuming that adult human neurogenesis had to happen and be important.

…or at least, that’s the optimistic take on it. But I can’t help thinking – antidepressants work in humans, which suggests that the people who found neurogenesis was necessary for antidepressant effects must have just been plain wrong. And if exercise has antidepressant effects in humans, then the claim that those effects are neurogenesis-mediated must be wrong too. And, uh, humans form spatial and temporal memories, so unless we do this by a totally different mechanism than the ones rats use, people must have been wrong when they said neurogenesis was involved in that. ECT? Works in humans. Brain plasticity? Happens in humans. So maybe it would be better to say that the original claim that adult neurogenesis happens in humans seems innocent and understandable – but if the new study is true, that suggests that a lot of the followup claims must have been imaginary. Anything that focuses on a process that happens in humans and says “neurogenesis causes this” must not only be wrong to extend the results to humans, but must be under strong suspicion of being wrong even about rats, unless rat brains and human brains accomplish the same basic tasks through totally different mechanisms (eg antidepressants work on rats but for different reasons than in humans).

We know many scientific studies are false. But we usually find this out one-at-a-time. This – again, assuming the new study is true, which it might not be – is a massacre. It offers an unusually good chance for reflection.

And looking over the brutal aftermath, I'm struck by how prosocial a lot of the felled studies are. Neurogenesis shows you should exercise more! Neurogenesis shows antidepressants work! Neurogenesis shows we need more enriched environments! Neurogenesis proves growth mindset! I'm in favor of exercise and antidepressants and enriched environments, but this emphasizes how if we want to believe something, it will accrete a protective layer of positive studies whether it's true or not.

I'm also struck by how many of the offending studies begin by repeating how dogmatic past neuroscientists were for not recognizing the existence of adult neurogenesis sooner. Remember Gage's review above:

> A milestone is marked in our understanding of the brain with the recent acceptance, contrary to early **dogma**, that the adult nervous system can generate new neurons. One could wonder how this **dogma** originally came about…

Or from Neurogenesis In Adult CNS: From Denial To Opportunities And Challenges For Therapy:

> The discovery of neurogenesis and neural stem cells (NSC) in the adult CNS has overturned a long-standing and deep-routed **"dogma"** in neuroscience, established at the beginning of the 20th century. This **dogma** lasted for almost 90 years and died hard when NSC were finally

> isolated from the adult mouse brain. The scepticism in accepting adult neurogenesis has now turned into a rush to find applications to alleviate or cure the devastating diseases that affect the CNS.

From Adult Human Neurogenesis: From Microscopy To Magnetic Resonance Imaging:

> The discovery of adult neurogenesis crushed the century-old **dogma** that no new neurons are formed in the mammalian brain after birth. However, this finding and its acceptance by the scientific community did not happen without hurdles. At the beginning of the last century, based on detailed observations of the brain anatomy reported by Santiago Ramon y Cajal and others, it was established that the human nervous system develops in utero (Colucci-D'Amato et al., 2006). In adult brains, it was thought, no more neurons could be generated, as the brain is grossly incapable of regenerating after damage (for a more detailed historical report see Watts et al., 2005; Whitman and Greer, 2009). This **dogma** was deeply entrenched in the Neuroscience community, and Altman's (1962) discovery of newborn cells in well-defined areas of the adult rodent brain was largely ignored.

I'm bolding the word "dogma" because for some reason every article in this field uses it like a verbal tic. University of Washington's "Neuroscience For Kids" page feels compelled to use the word even though they don't expect their readers to know what it means:

> The **dogma** (a set of beliefs or ideas that is commonly accepted to be true) that nerve cells in the adult brain, once damaged or dead, do not replace themselves is being challenged. Research indicates that at least one

> part of the brain in adults maintains its ability to make nerve cells.

I think Patient Zero in this use-of-the-word-dogma epidemic might be Neurogenesis In The Adult Brain: Death Of A Dogma, (880 citations) whose abstract says:

> For over 100 years a central assumption in the field of neuroscience has been that new neurons are not added to the adult mammalian brain. This perspective examines the origins of this **dogma**, its perseverance in the face of contradictory evidence, and its final collapse. The acceptance of adult neurogenesis may be part of a contemporary paradigm shift in our view of the plasticity and stability of the adult brain.

The dogma-concern isn't totally wrong. Previous neuroscientists thought there wasn't neurogenesis in rats, and there is. That was a legitimate mistake and one worth examining. But is it possible that the reaction to that mistake – a field-wide obsession with talking about how dogmatic you had to be to miss the obvious evidence of mammalian neurogenesis, and a desire never to repeat that mistake – contributed to the less-than-stellar effort to make sure neurogenesis was happening in humans? Heuristics work until they don't. Those who fail to learn from history are doomed to repeat it, but those who learn too much from history are doomed to make the exact opposite mistake and accuse anyone who urges restraint of "failing to learn from history" and "dogmatism". From the Virtues of Rationality:

> The Way is a precise Art. Do not walk to the truth, but dance. On each and every step of that dance your foot comes down in exactly the right spot. Each piece of evidence shifts your beliefs by exactly the right amount, neither more nor less.

Or maybe I'm just grasping for straws. But I feel like I have to grasp for *something*. I have nowhere near as much expertise as the actual neuroscientists writing about this result (and there are many). I'm sure I've made some inexcusable mistakes somewhere in this process. Perhaps I am missing some colossal flaw in the new study, and wrongly slandering dozens of neuroscientists doing great work.

But the reason I feel compelled to dabble in this subject anyway is that I don't feel like anyone else is conveying the level of absolute terror we should be feeling right now. As far as I can tell, this is the most troubling outbreak of the replication crisis so far. And it didn't happen in a field like social psychology which everyone already knows is kind of iffy. It happened in neuroscience, with dramatic knock-on effects on medicine, psychology, and psychiatry.

I feel like every couple of months we get a result that could best be summed up as "no matter how bad you thought things were, they're actually worse". And then things turn out to be even worse than that. We can't just become 100% certain things are arbitrarily bad – that would be making the same mistake as the neuroscientists who were overly eager to reject the no-neurogenesis dogma. But that means we always have to be ready for disappointment.

From the Neuroskeptic article:

> So what does this all mean? Sorrells et al. conclude by speculating, provocatively, that our lack of adult hippocampal neurogenesis may actually be part of what makes us human:
>
> *"Interestingly, a lack of neurogenesis in the hippocampus has been suggested for aquatic mammals*

> *(dolphins, porpoises and whales), species known for their large brains, longevity and complex behaviour.”*
>
> This hypothesis seems pretty wild to me. But it’s no wilder than some of the other theories that have long surrounded adult neurogenesis

Our total inability to ever change or get better in any way is what separates us from the animals. Inspiring!

Why DC’s Low Graduation Rates?

[Some changes to the conclusions in this post; see edit at the end and entry 21 on Mistakes page]

US News: DC Schools Brace For Catastrophic Drop In Graduation Rates. “Catastrophic” isn’t hyperbole; the numbers are expected to drop from 73% (close to the national average of 83%) all the way down to 42%.

There’s no debate about why this is happening – it’s because the previous graduation rate was basically fraudulent, inflated by pressure to show that recent “reforms” were working. Last year there was a big investigation, all the investigators agreed it was fraudulent, DC agreed to do a little less fraud this year, and this is the result. It’s pretty damning, given how everybody was praising the reforms and holding them up as a national model and saying this proved that Tough But Fair Education Policy could make a difference:

> As far as scandals in the education policy world go, D.C. schools so profoundly miscalculating graduation rates at a time when the high-profile school district had been so self-laudatory about its achievements may be difficult to top […] Indeed, when Michelle Rhee took the reins of the flailing school system a decade ago, it galvanized the education reform movement, which had just begun blossoming around the country, and ushered in a host of controversial changes that included the shuttering of multiple schools, firing of hundreds of teachers and the institution of new teacher evaluation and compensation models.
>
> The changes not only dramatically altered the local political landscape in Washington but also shined a national spotlight on D.C. schools that prompted other

> urban school districts and education policy researchers to consider the nation's capital a bellwether for the entire education reform movement.

Well, darn.

But the interesting bit isn't just that DC schools are doing worse than we thought. It's that DC schools are doing amazingly, uniquely, abysmally bad, below what should even be possible. We make fun of states like Mississippi and Alabama, but both have graduation rates around 80%. The lowest graduation rate in any of the fifty states is in Oregon, which still has 69%. And we are being told DC is 42%!

When we discussed this in the last links thread, people had a couple of explanations:

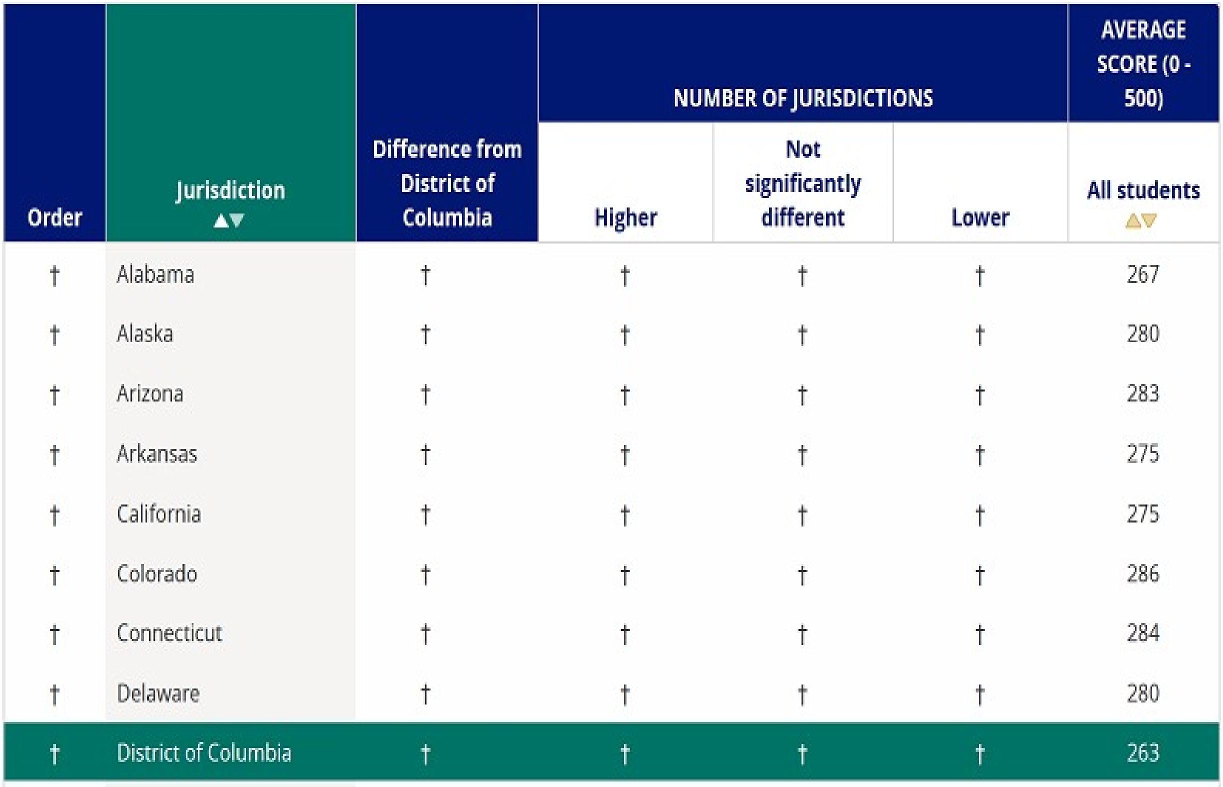
1. Washington DC has a terrible school system, with uniquely incompetent administrators.

2. Washington DC is poorer, blacker, and more segregated than any other state, and that leads to unique challenges other school systems don't face. Even though everyone is doing their best, they face insurmountable structural difficulties.

3. Maybe the fraud was so bad that DC over-corrected, and now has stricter standards than anywhere else.

Which of these is most important?

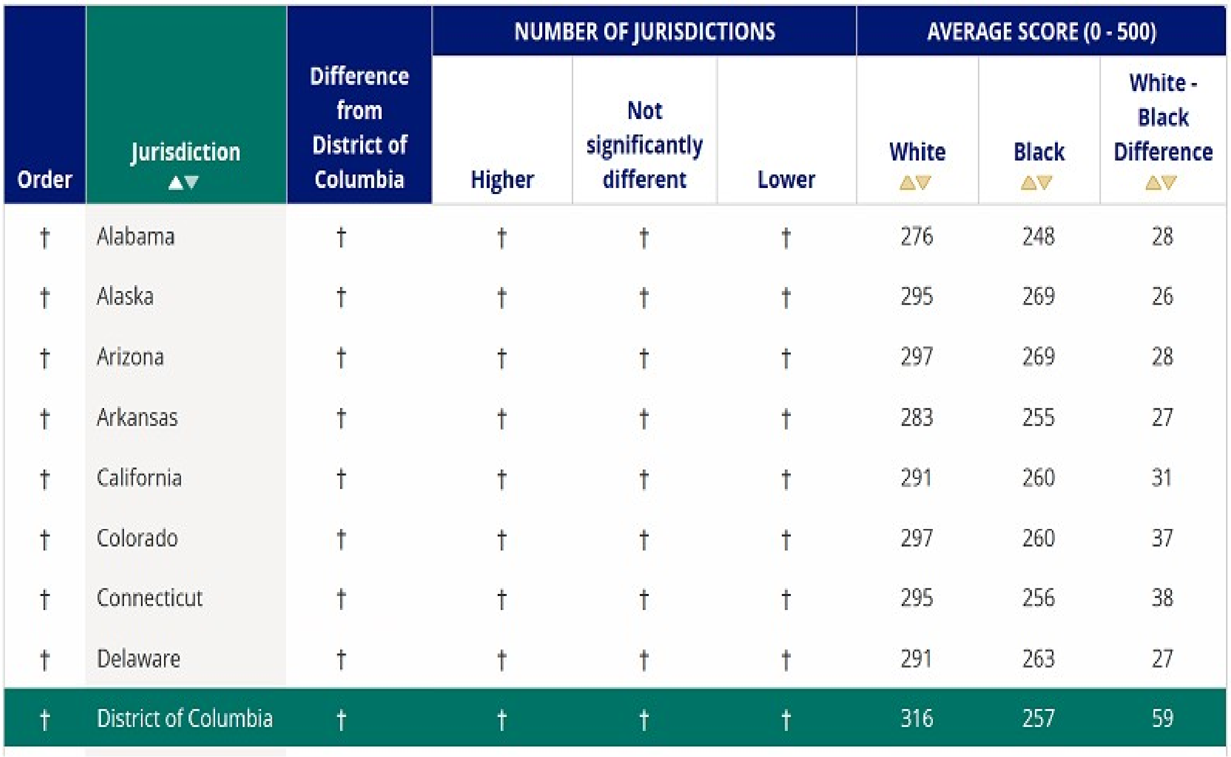
Let's start by looking at test scores. Here's a sample of DC's NAEP scores compared to some other states (full list is here).



Order	Jurisdiction ▲▼	Difference from District of Columbia	NUMBER OF JURISDICTIONS			AVERAGE SCORE (0 - 500)
			Higher	Not significantly different	Lower	All students ▲▼
†	Alabama	†	†	†	†	267
†	Alaska	†	†	†	†	280
†	Arizona	†	†	†	†	283
†	Arkansas	†	†	†	†	275
†	California	†	†	†	†	275
†	Colorado	†	†	†	†	286
†	Connecticut	†	†	†	†	284
†	Delaware	†	†	†	†	280
†	District of Columbia	†	†	†	†	263

In both reading and math, in all grades, DC does abysmally. But they don't do *as* abysmally as a 42% graduation rate might predict. They are sometimes last, sometimes second- or third- to last, and in any case they're rarely that different from other low-performers like Alabama and Mississippi. They might be best described as a member in good standing of the lowest-performing tier of US states.

But there are massive racial inequalities in education, and DC is by far the blackest "state" in the nation. How does it do when we adjust for this?



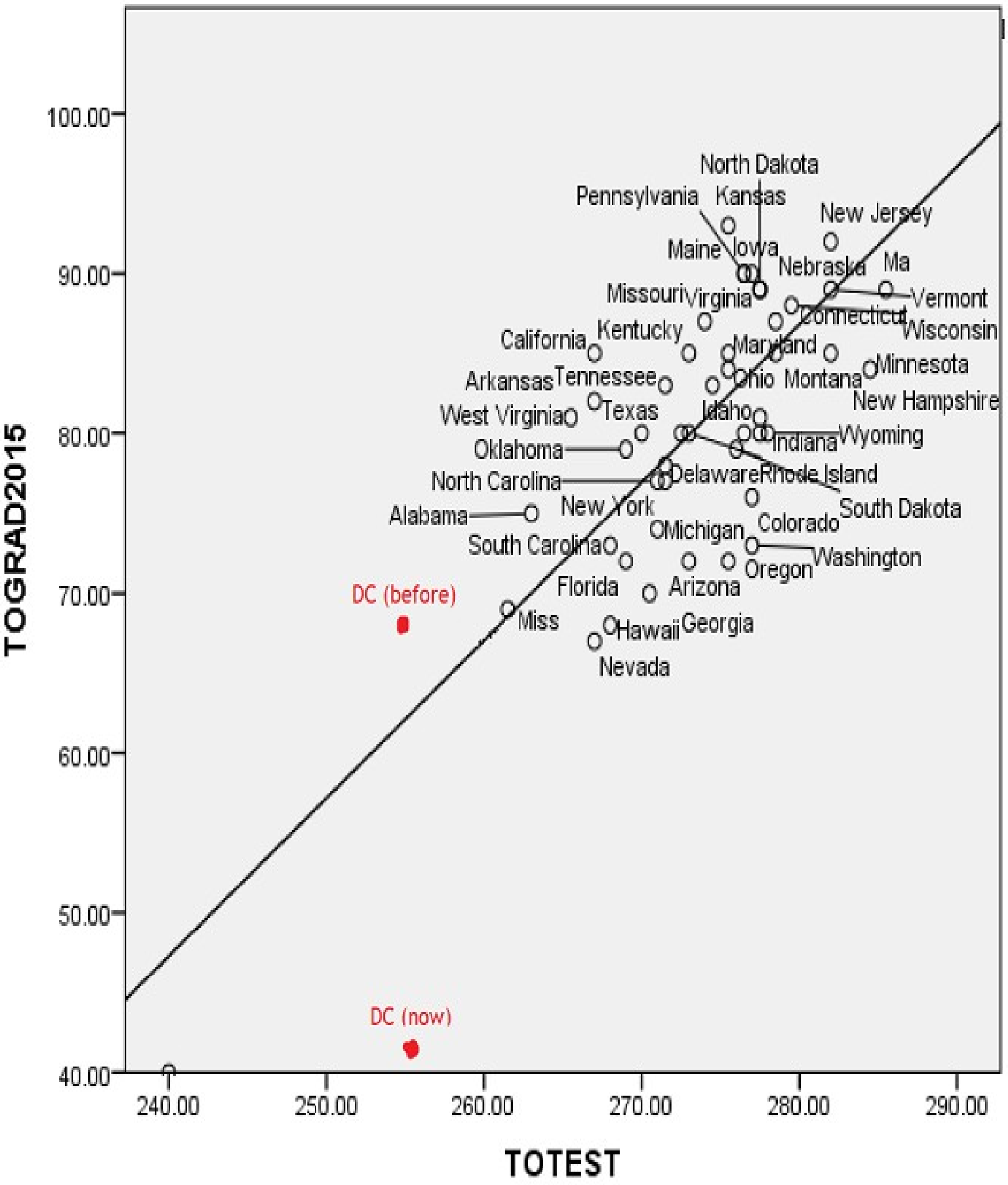
Order	Jurisdiction ▲▼	Difference from District of Columbia	NUMBER OF JURISDICTIONS			AVERAGE SCORE (0 - 500)		
			Higher	Not significantly different	Lower	White △▽	Black △▽	White - Black Difference △▽
†	Alabama	†	†	†	†	276	248	28
†	Alaska	†	†	†	†	295	269	26
†	Arizona	†	†	†	†	297	269	28
†	Arkansas	†	†	†	†	283	255	27
†	California	†	†	†	†	291	260	31
†	Colorado	†	†	†	†	297	260	37
†	Connecticut	†	†	†	†	295	256	38
†	Delaware	†	†	†	†	291	263	27
†	District of Columbia	†	†	†	†	316	257	59

The full table is here. DC has *by far* the highest white test scores in the country – probably because a lot of its white students are the kids of well-off bureaucrats and think-tank types. And its black test scores range from lowest-tier-member to mediocre.

Nor can this just be a Simpson's Paradox, where both white and black students do fine but the difference is driven by a greater number of black students. The US average black graduation rate is 68%. No state has a black graduation rate lower than 56%. DC – again – is supposed to have a 42% total graduation rate.

This seems to refute hypotheses 1 and 2 – that DC just has a terrible school system, or just has an unusually disadvantaged population. Its white students do very well. Its black students do poorly but not too much worse than they would in other states. There is no gap – among either race or among both combined – that corresponds to the gap between Mississippi’s graduation rate (total 80%, black-only 77%), and DC’s graduation rate of 42%.

This leaves us with hypothesis 3 – that DC got burned so badly in the fraud investigations that its standards are now much higher than anywhere else’s. Maybe a graph will help:



The horizontal axis is each state's test scores – specifically the average of its 8th grade NAEP reading and math. The vertical axis is its graduation rate. The previous, supposedly fraudulent DC rate clusters together more or less with everybody else. The new, supposedly non-fraudulent rate is an extreme outlier, showing a graduation rate about 20 percentage points lower than test scores would predict.

So DC's old graduation rate was normal relative to their test scores, and their new graduation rate is an outlier. But their old graduation rate is widely considered to have been maintained by fraud and really low standards, and their new graduation rate is widely considered correct. Does that mean that everywhere else with the same levels of poverty and segregation as DC also uses fraud and really low standards to keep their graduation rates up?

Maybe. Detroit is often used as a symbol of inner-city educational dysfunction, but even the district with the worst Detroit schools has a 61.5% graduation rate. How do they do it? Given that fewer than 5% of their students pass exams, I assume they do it through fraud and really low standards. Los Angeles? Fraud and really low standards. Chicago? Fraud and really low standards. Baltimore? Given stories like the one where one of the city's highest-graduation-rate schools has zero percent of students score at "meets expectations" or even "approaches expectations" on statewide exams, it looks like fraud and really low standards.

I understand this is a really strong claim. But others seem to agree, and it's the only way I can make sense of DC's abysmally low projected graduation rates, in the context of their merely-awful exam scores.

Or maybe the key word is "projected". From the US News article:

> The report estimated that just 42 percent of seniors are on track to graduate at the end of the current school year, down from 73 percent who graduated in the 2016-17 school year. It noted that 19 percent of students were “moderately off-track” and could still earn enough credits to graduate.

Plausibly, kindly school officials will find loopholes that allow all of those 19% of moderately-off-track children to graduate, DC’s actual graduation rate will be 62% (right on the trend line for the graph above), and this whole episode will be remembered as “that one time a scary report underestimating the graduation rate came out”. Plausibly, this is the main way that Detroit and Los Angeles and all those other cities keep their rates up, and the more dramatic scare stories are just that.

I hope this happens. Think how unfair it will be for DC students if it doesn’t. Somebody who would graduate comfortably from any other high school in the country will be held back because they happen to have been in DC the year it decided to enforce standards nobody else enforced. If the true value of education is signaling, then the most important thing a school district can do is make sure it’s speaking the same signaling-language as everyone else. Probably somebody should fix the system in general, but that needs to happen on a national level if it’s not going to leave thousands of unfairly-failed children as collateral damage.

[EDIT: More discussion at Highlights From The Comments On DC Graduation Rates. Main update is that I underestimated the importance of absences, which are what’s causing a lot of the non-graduations, which there might be more of in DC, and which DC might be stricter about than other areas.]

Highlights From The Comments On DC Graduation Rates

Bizzolt writes:

> DC Public Schools HS teacher here (although I'm not returning next year, as is the case with many of my colleagues). As noted, one of the biggest factors in the graduation rates is the unexcused absences–if you look at the results of our external audit and investigation here, you see that for many schools, a significant number of our seniors "Passed Despite Excessive Absences in Regular Instruction Courses Required for Graduation"–over 40% of 2017 graduates at my high school, for example.
>
> So the attendance policy is being strictly enforced now, and you can see how from that alone, a ~30% drop in expected graduates is possible. Some more details about strictly enforcing the attendance policy though:
>
> 1: DCPS has what's called the '80 20′ rule: A student that is absent for at least 20% of their classes is considered absent for the whole day.
> 2: Most schools have 5 periods, so an absence in one class would be considered an absence for the whole day.
> 3: If you have 10 or more unexcused absences in a class, you automatically get an F for the term.
> 4: If you are over 15 minutes late for a class, that is considered an unexcused absence.
> 5: A majority of these absences are in first period.
> 6: A majority of students in my school and many others live in single parent households.
> 7: These students are typically responsible for making sure their younger siblings get to school, if they have

> any.
> 8: Elementary and middle schools in my neighborhood start at the exact same time as high school.
> 9: Their doors do not open until 5 to 10 minutes before the starting bell, presumably for safety reasons.
> 10: Refer to point 4.
>
> There's many other problems at DCPS to be sure, but this set of circumstances alone is causing the largest increase in failing grades and graduation ineligibility at my high school, and basically every other 90+% black school in the district. You could see how this accounts for quite a bit of the difference between white and black graduation rates as well. There's a reason why across the board, DCPS schools were not strictly enforcing this policy in previous years.

It looks like most other school districts don't have this policy; it seems plausible that this is the main difference between DC and other poor school districts that nevertheless manage to pass most of their kids.

Userfriendlyyy also focuses on the absences:

> Looks to me like the policy they changed was losing credit for bad attendance. This might be from a few things. Kids might need to help out with the family finances. The only part of the job market that is doing well right now is low end unskilled workers who are willing to get paid crap (no matter how much the financial press wants to pretend otherwise, I listened to an hour of local NPR and the Topic was 'call in and tell us how the booming job market is helping you out', 20 callers not one had anything good to say and my state has one of the lowest unemployment rates in the country). If you know you don't have the grades for a

> scholarship, your family is broke and since we have effectively made going to college impossible for anyone but the offspring of the oligarchy, and you can find a minimum wage job easily; what exactly is the utility of that little piece of paper compared to the ability to put food on the table tonight?

Static focuses on absences too:

> The amazing thing to me is that they are largely failing due to unexcused absences. [See Washington Post:] I Feel Really Bad For The Class Of 2018: Graduation May Be Imperiled
>
> I think that does point a little to the signalling difference between the GED and the diploma- can you show up every day?
>
> That said, DC is planning to add an exam requirement for graduation. It would be better if they made that an alternative to attendance.

The absence requirement seems infuriating insofar as it probably fails mostly poor children, and if those children are failing due to the absence requirement rather than because they actually flunk their exams, then it seems mean-spirited to punish inputs rather than actual success.

Maybe they have it because they don’t actually grade students on exam performance (or have lots of different ways to make sure students who flunk exams graduate anyway), but they only want to extend this benefit-of-the-doubt to students who really try. I wonder if it would work to say you can graduate if you manage to make it to school enough times *or* pass your examinations. That would at least be fairer to poor children who can’t always attend but are able to figure out ways to succeed anyway.

Proyas writes:

> I was friends with a guy who briefly worked as a teacher at a public high school in central DC (I'm 80% sure it was Cardozo High). He had an education background thanks to spending several years working as a youth camp counselor and as an after-school program counselor, and that was sufficient to qualify him for DCPS' abbreviated teacher training program (such a thing existed in 2009 when he did it; I'm unsure if it is still around). During the training program, I remember him speaking about his enthusiasm for the teaching skills he was learning and about his eagerness to put them to use (in retrospect, I think some of this was a nervous attempt to convince himself the job wouldn't be bad). After a break of several months, we spoke again, and he was almost totally disillusioned with the job and was already thinking of quitting. This is what I remember him saying:
>
> 1) On the first day of classes, there was no orientation for new teachers, no brief meeting where the Principal shook his hand and said "Welcome Aboard," nothing. He had to go to the front office and ask a secretary what classroom was his and walk there by himself.
>
> 2) Unexcused absences were chronic and undermined his ability to teach anything. At the start of each of his classes, he had a written roster of students, and he had to check off which students were there. For any class, typically 20-30% of students would be missing, without explanation (This is a very important point to remember whenever anyone tries to blame DCPS' poor outcomes on large class sizes–on paper, each class might have 35 students, but typically, only 23 are actually showing up). Additionally, the 20-30% of students who were absent

each class varied from day-to-day, meaning one student didn't know what was taught on Monday, the one next to him was there Monday but not Tuesday, the third was there the first two days but not Wednesday, etc.

3) Student misbehavior was atrocious. For example, out of the students who showed up to class, it was common for some to walk into the classroom late, again without any explanation and often behaving disruptively. As a rule, whenever a student did that, he was obligated to sign his name on a clipboard for the teacher's attendance records (there was no punishment for tardiness–late students merely had to write their names down). Some late students would chronically resist doing this, either ignoring him and just going to their desks or yelling curses at him. My friend described an incident where one student–who was physically bigger than he was–yelled out he was a "FAGGOT" when asked to sign the clipboard, provoking laughs from all the other students, before sitting down without signing it. After seeing he could get away with that, the student started calling my friend "FAGGOT" all the time. Other examples of misbehavior included near-constant talking among the students during lessons and fooling around with cell phones.

4) Teachers received almost no support from the school administration. Had sane rules been followed at this high school, students would have been immediately sent to the office for formal punishment for these sorts of offenses I've described. However, under such a policy, the office would have been overwhelmed with misbehaving students and probably some of their enraged parents, so the administration solved the problem by forbidding teachers from sending students to the office for anything other than physical violence in

the classroom. My friend had no ability to formally punish the student who liked to call him “FAGGOT” other than to use stern verbal warnings.

5) Most of the students were unwilling and in some cases unable to learn. During class sessions, the students were clearly disengaged from what he was teaching. Homework completion rates were abysmal. As the end of the academic semester neared, he saw that a huge fraction of them were on track to fail, so he resorted to pitiful cajoling, pizza parties, reward schemes, and deals involving large curves to everyone’s grades if they could only, for once do a little work, and it didn’t work. Some of his students were Latino and understood little or even no English, meaning they learned (almost) nothing, even when they tried. He resorted to seating the students who knew no English next to bilingual Latinos who could translate for them. That was the best he could do. In fairness, he spoke glowingly of some of his students, who actually put in some effort and were surprisingly smart.

6) At the time my friend was teaching, DCPS was in the grips of some harebrained, faddish teaching philosophy that said students of different academic abilities shouldn’t be put in different course tracks, but rather, should be deliberately put in the same class. This of course caused immediate problems since the curriculum was too hard for the weakest students and too easy for the strongest ones. I think my friend said his training program basically told teachers to “try harder” if any problems arose from the setup.

I’ll never forget how crestfallen and stressed out he was when he described these things to me. Having never taught in American public schools, I didn’t realize just

> how bad it was, and the detailed nature of his anecdotes really had an impact on me. I advised him to finish his year at the high school and then to transfer to ANY non-urban school in the area, even if it meant lower pay or a longer commute. We lost touch after that, but I can’t imagine he still works in DCPS.

Okay, maybe this goes deeper than just the absence thing.

MrApophenia writes:

> One thing to keep in mind is that DC really is uniquely bad as a school district. A few years back, the Washington Post did a really in depth analysis of why it was so broken. (Sadly, with Google being totally swamped by the current scandal, I was unable to find a link.)
>
> The conclusion they found was pretty interesting –
>
> If you go back to the 60s, DC had some of the best urban schools in the country by every metric they had to track things back then. What changed? Well, see, back then DC still didn’t have home rule. They were almost entirely run by the federal government. So the highest elected office the city had was the DC Board of Education.
>
> In 1971, an ambitious young politician by the name of Marion Barry got elected to the board, and almost immediately began farming out school administration positions as political rewards for his cronies. This practice caught on, and within a matter of years, the whole enterprise basically descended into naked corruption.

> I recall they showed a figure in that article that the DC public school system spends the third most money per student of any district in the country (after NY and Boston); however, the Post also found that in terms of the quantity of money that is actually spent on students, DC was roughly at the level of the most poverty-stricken districts in the poorest Southern, rural school districts.
>
> Michelle Rhee did not change any of this, Waiting for Superman or no.
>
> This isn't meant to argue with the premise of the rest of the article – the idea that everyone else is committing fraud seems quite plausible. But I wouldn't necessarily reject the idea that DC's school system really is a special, unique snowflake of terrible practices, either.

There's also this comment on home rule and Puerto Rico which seems to reinforce this idea that districts carefully monitored by competent national authorities do well, and districts that have control of their own standards devolve into corruption and failure really quickly. How does this mesh with the standard federalist/localist/Seeing-Like-A-State style arguments that individual communities know what's best for them and central planners usually make things worse?

I won't quote it directly because it's not from an SSC comment, but some people on the subreddit link to this Reddit post by a DC public school teacher.

See also this thread in the subreddit on how sub-Saharan African schools – despite being much poorer than anywhere in DC – are really well-run, quite safe (except that apparently “baboons are a huge problem”), and a delight to teach at.

Recommendations vs. Guidelines

Medicine loves guidelines. But everywhere else, guidelines are still underappreciated.

Consider a recommendation, like “Try Lexapro!” Even if Lexapro is a good medication, it might not be a good medication for your situation. And even if it’s a good medication for your situation, it might fail for unpredictable reasons involving genetics and individual variability.

So medicine uses guidelines – algorithms that eventually result in a recommendation. A typical guideline for treating depression might look like this (this is a very over-simplified version for an example only, NOT MEDICAL ADVICE):

1. Ask the patient if they have symptoms of bipolar disorder. If so, ignore everything else on here and move to the bipolar guideline.

2. If the depression seems more anxious, try Lexapro. Or if the depression seems more anergic, try Wellbutrin.

3. Wait one month. If it works perfectly, declare victory. If it works a little but not enough, increase the dose. If it doesn’t work at all, stop it and move on to the next step.

4. Try Zoloft, Remeron, or Effexor. Repeat Step 3.

5. Cycle through steps 3 and 4 until you either find something that works, or you and your patient agree that you don’t have enough time and patience to continue cycling through this tier of options and you want to try another tier with more risks in exchange for more potential benefits.

6. If the depression seems more melancholic, try Anafranil. Or if the depression seems more atypical, try Nardil. Or if your patient is on an earlier-tier medication that almost but not quite works, try augmenting with Abilify. Repeat Step 3.

7. Try electroconvulsive therapy.

The end result might be the recommendation “try Lexapro!”, but you know where to go if that doesn’t work. A psychiatrist armed with this guideline can do much better work than one who just happens to know that Lexapro is the best antidepressant, even if Lexapro really *is* the best antidepressant. Whenever I’m hopelessly confused about what to do with a difficult patient, I find it really reassuring that I can go back to a guideline like this, put together by top psychiatrists working off the best evidence available.

This makes it even more infuriating that there’s nothing like this for other areas I care about.

Take dieting. Everybody has recommendations for what the best diet is. But no matter what diet you’re recommending, there are going to be thousands of people who tried it and failed. How come I’ve never seen a diet guideline? Why hasn’t someone written something like:

1. Try cutting carbs by X amount. If you lose Y pounds per week, the diet is working. If not, you’re probably resistant to cutting carbs because [two hours of mumbling about insulin] and you should move on to the next tier.

2. Try cutting fat by X amount. If you lose Y pounds per week, the diet is working. If not, you’re probably resistant to cutting fat because [two hours of mumbling about leptin], and you should move on to the next tier.

And so on until Step 7 is “get a gastric bypass”.

I agree nobody can ever make a perfect algorithm that works for all eventualities. But still. Surely we can do better than “Try the Paleo diet! I hear it’s great!”

What information do guidelines carry beyond a recommendation?

First, they have more than one recommendation. It may be that the Paleo diet is the best, but the guidelines will also include which is the second-best, third-best, et cetera.

Second, because they have more than one recommendation, they can tailor their recommendation to your specific circumstances. The person with depression and comorbid anxiety may want to start with Lexapro; the person whose main symptom is tiredness may want to start with Wellbutrin. Since I love bread, does that mean I should avoid carb-cutting diets? Does that mean it’s extra-important that I cut carbs? Does it not matter, and really it depends on whether I have a family history of diabetes or not?

Third, they acknowledge that some people might need more than one recommendation. If you hear “try the Paleo diet”, and then you try it, and it doesn’t work, you might believe you’re just a bad dieter, or that all diets are scams, or something like that. Guidelines implicitly admit that everyone is different in confusing ways, that something that’s expected to work for many people might not work for you, and that you should expect to have to try many things before you find the right one.

Fourth, because they admit you may need to try more than one thing, they contain (or at least nod at) explicit criteria for success or failure. How long should you try the Paleo diet before you decide it doesn’t work? How much weight do you need to lose before it qualifies as “working”? If it’s been three months and I’ve lost four pounds, should you stick with it or not?

Fifth, they potentially contain information about which things are correlated or anticorrelated. The depression guidelines make it clear that if you’ve already tried Lexapro and Zoloft and they’ve both failed, you should stop trying SSRIs and move on to something with a different mechanism of action. If I’ve tried five carb-cutting diets, should I try a fat-cutting diet next? If I hate both Mexican food and Chinese food, is there some other category of food which is suitably distant from both of those that I might like it? Guidelines have to worry about these kinds of questions.

My impression is that once you understand a field really well, you have something like a Guideline in your mind. I think if nobody had ever written a guideline for treating depression, I could invent a decent one myself out of everything I’ve pieced together from word-of-mouth and common-sense and personal experience. In fact, I think I do have some personal guidelines, similar to but not exactly the same as the official ones, that I’m working off of without ever really being explicit about it. Part of the confusion of questions like “What diet should I do?” is sorting through the field of nutrition until you can sort of imagine what a guideline would look like.

So why don’t people who have more knowledge of nutrition make these kinds of guidelines? Maybe some do. I can’t be sure I haven’t read dieting guidelines, and if I did I probably ignored them because lots of people say lots of stuff.

But I think that’s a big part of it – making guidelines seems like a really strong claim to knowledge and authority, in a way that a recommendation isn’t. Some idiot is going to follow the guidelines exactly, screw up, and sue you. I just realized that my simplified-made-up depression guidelines above didn’t have “if the patient experiences terrible side effects on the antidepressant, stop it”. Maybe someone will follow those guidelines exactly (contra my plea not to), have something horrible happen to them, and sue me. Unless you’re the American Psychiatric Association Task Force or someone else suitably impressive, your “guidelines” are always going to be pretty vague stuff that you came up with from having an intuitive feel for a certain area. I don’t know if people really want to take that risk.

Still, there are a lot of fields where I find it really annoying how few guidelines there are.

What about nootropics? I keep seeing people come into the nootropics community and ask “Hey, I feel bad, what nootropic should I use?” And sure, eventually after doing lots of research and trying to separate the fact from the lies, they might come up with enough of a vague map of the area to have some ideas. But this is an area where “Well, the first three things you should try for anxiety are…” could be really helpful. And I don’t know of anything like that – let alone something that tells you how long to try before giving up, what to look for, etc.

Or let’s get even broader – what about self-help in general? I don’t really believe in it much, but I would love to be proven wrong. If there were a book called “You Are Willing To Devote 100 Hours Of Your Life To Seeing If Self-Help Really Works, Here’s The Best Way For You To Do It”, which contained a smart person’s guidelines on what self-help things to try and how to go about them, I would absolutely buy it.

SSC Survey Results: Sexual Harassment Levels By Field

[content note: sexual harassment]

I.

Recent discussion of sexual harassment at work has focused on a few high-profile industries. But there has been relatively little credible research as to how rates really differ by occupation type.

There are many surveys of harassment rates in specific industries, but they can’t be credibly compared with one another. The percent of people who report sexual harassment varies wildly from survey to survey – thus studies finding that anywhere from 12 percent to 48 percent to 60 percent to 85 percent of women have been harassed at work. If a survey shows that 60% of female nurses get sexually harassed at work, does that mean nurses are victimized particularly often (because more than 12%) or are unusually safe (because less than 85%)? It doesn’t matter, because another study says only 19% of nurses get harassed.

Why do all these numbers differ so dramatically? The most important issue seems to be how you ask the question. “Have you ever been harassed?” gets numbers more like 12%; giving a long list of specific behaviors and asking “Have you ever experienced any of these?” gets numbers closer to 85%, depending on what the behaviors are. Surveys also differ on whether they ask all employees or just women, whether they include a time frame (eg “…in the past two years”), whether they specify that it had to be at work vs. work-related events, and whether they include witnessing someone else's harassment. Taking these surveys entirely seriously would lead to the conclusion that Uber has the lowest sexual harassment rate of any company or industry in the world; I choose not to take them seriously.

This means we need investigations that use the same methodology across multiple fields. Whenever the media talks about this – see eg the Washington Post's The Industries With The Worst Sexual Harassment Problem – they're working off of the Equal Employment Opportunity Commission's records. But these are totally unsuitable for the task – they just report raw number of claims per industry. The industries that rank lowest in EEOC's data tend to be small industries with very few women – for example, taken seriously the WaPo's graph shows that mining has the least problem with sexual harassment of any industry in the world. Is this thanks to their uniquely progressive culture – or because there are practically no female miners? I'm going to say the second one. The takeaway that most real researchers take from the EEOC claims is that the lowest-paying and most mundane occupations – retail, restaurant work, hotel work, etc – have much higher sexual harassment rates than the prestigious occupations people generally talk about. Eyeballing the data, this looks basically true. But trying to get anything more fine-grained than that out of EEOC is basically hopeless.

I only know of two surveys that have even attempted to compare different fields in a principled way, and neither really inspires confidence.

First is a survey by *Cosmopolitan* magazine, that asked 2235 women in different fields about sexual harassment in their industry. About 33% reported being harassed at work. Women in the retail and restaurant industries reported the highest rates of sexual harassment; despite media coverage’s focus on STEM and entertainment, both reported average to just-below-average harassment rates.

Second is the Project XX Survey, where the automotive industry decided to survey their workers using methodology previously used in Silicon Valley, which made their results at least somewhat comparable. The advertising and market research industries seem to have joined in later. They find the automotive industry has the most harassment, followed by Silicon Valley, advertising, and market research – but there aren’t really big differences and industry rankings frequently reverse across questions.

I interpret existing data in this area as being basically useless, but at least suggestive that the media focus on a few prestigious industries is mistargeted.

II.

The Slate Star Codex survey is an online poll of readers of this blog about various aspects of their life and personality. This year 8,077 people participated. The survey asked a set of questions on sexual harassment, providing a unique dataset with which to investigate the question.

Due to the blog’s tech focus, readers skew well-off, male, and STEM-focused. This made it impossible to investigate the restaurant and retail industries that previous research suggests have the most problems. But there were still enough responses to get a wide cross-section of white-collar occupations and academic fields.

Respondents were asked four questions:

1. Have you ever been sexually harassed or assaulted at work?

2. Have you ever been sexually harassed or assaulted outside of work?

3. Have you ever made unwanted sexual advances to someone at work that you think they interpreted as sexual harassment or assault?

4. Have you ever made unwanted sexual advances to someone outside of work that you think they interpreted as sexual harassment or assault?

Respondents were given a “prefer not to answer” option in case the questions made them upset; in order to make people entirely comfortable using the option, they were asked to select it automatically if the last digit of the time was a “1”.

Data were analyzed about female victims, male victims, and male perpetrators. Although some women admitted to perpetration, the sample size across fields was too small to be useful. Industries were included if they had a sample size greater than thirty; because there were more male respondents and so all industries had a greater male sample size, some industries are included on the male graphs but not the female graphs.

III.

Overall, 21% of women and 6% of men reported being sexually harassed at work. 1% of women and 2.2% of men reported committing sexual harassment at work.

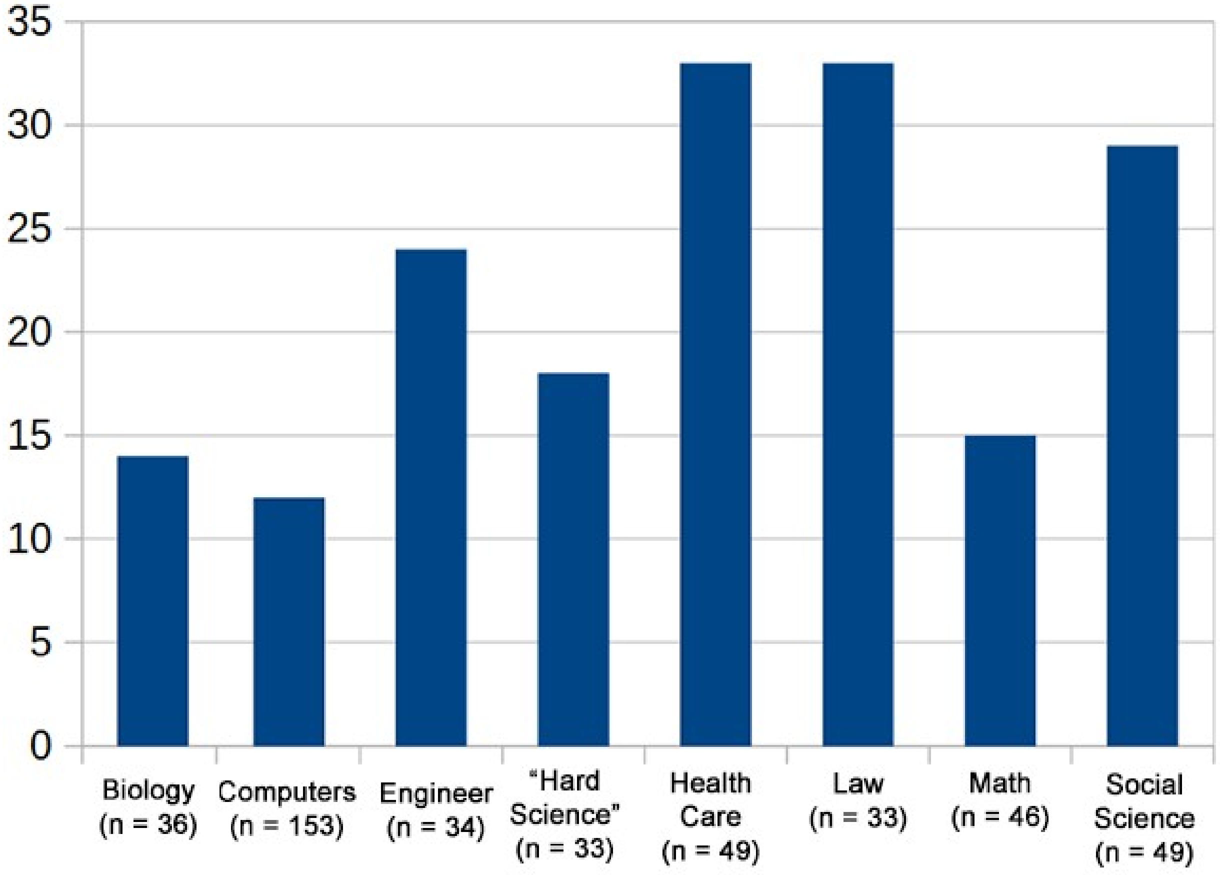


Figure 1: Percent women reporting sexual harassment victimization at work, by occupation type

Fields listed do not exactly match fields on the survey as some were combined in order to increase sample size, eg “physics” and “chemistry” into “hard sciences”. There’s a discussion involving pre-registered categories below.

Figure 2: Percent men reporting sexual harassment victimization at work, by occupation type

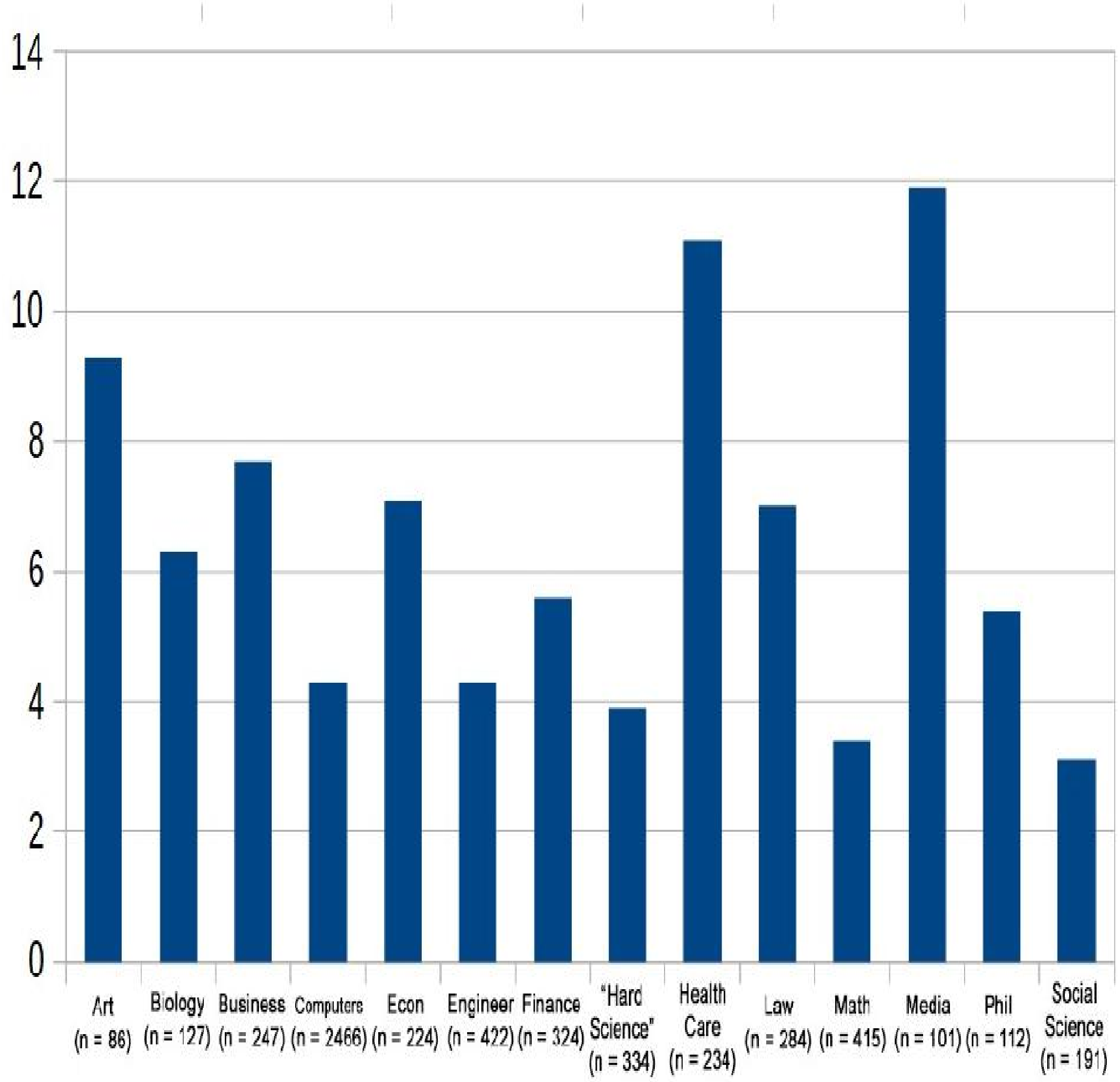
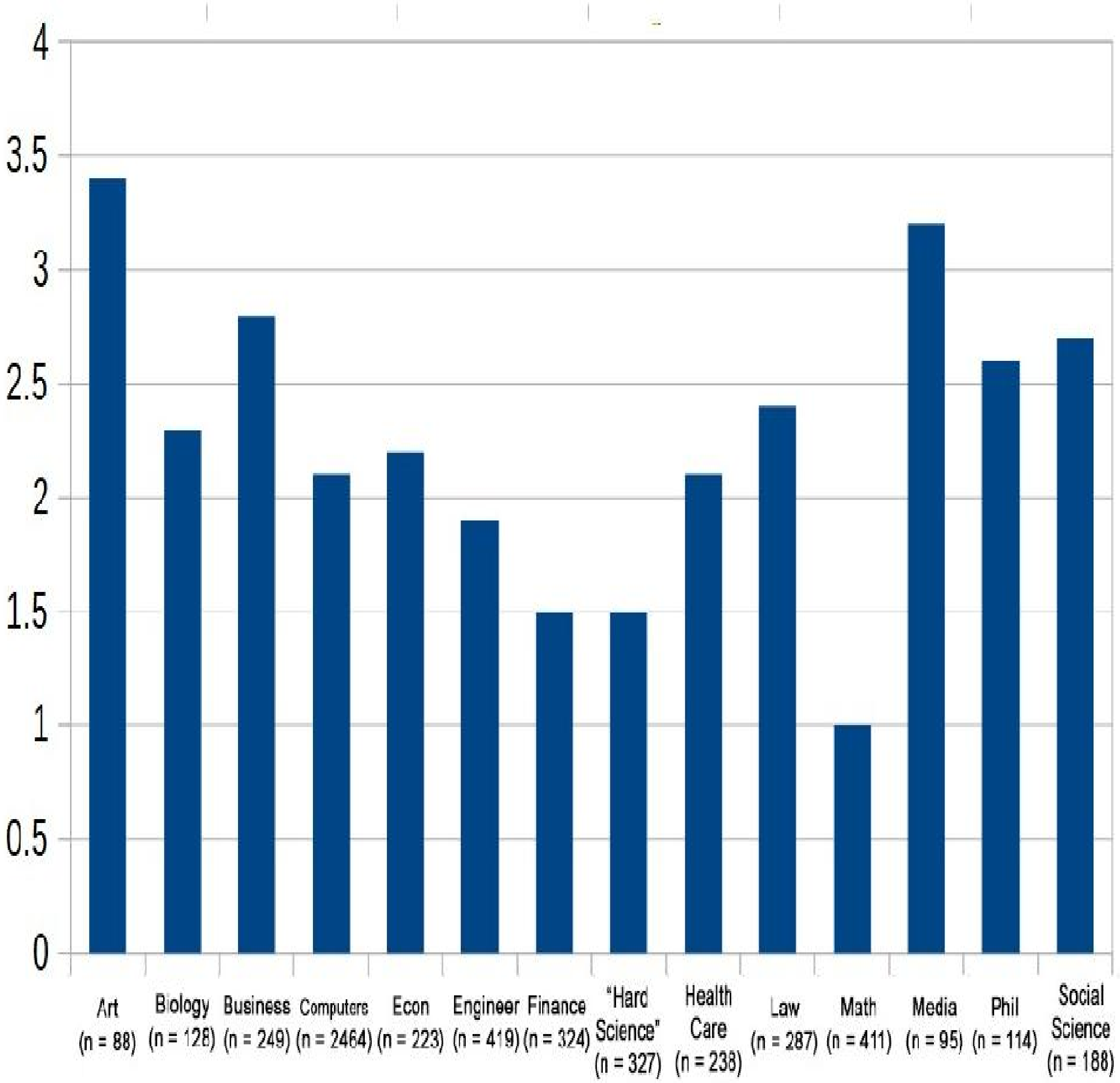
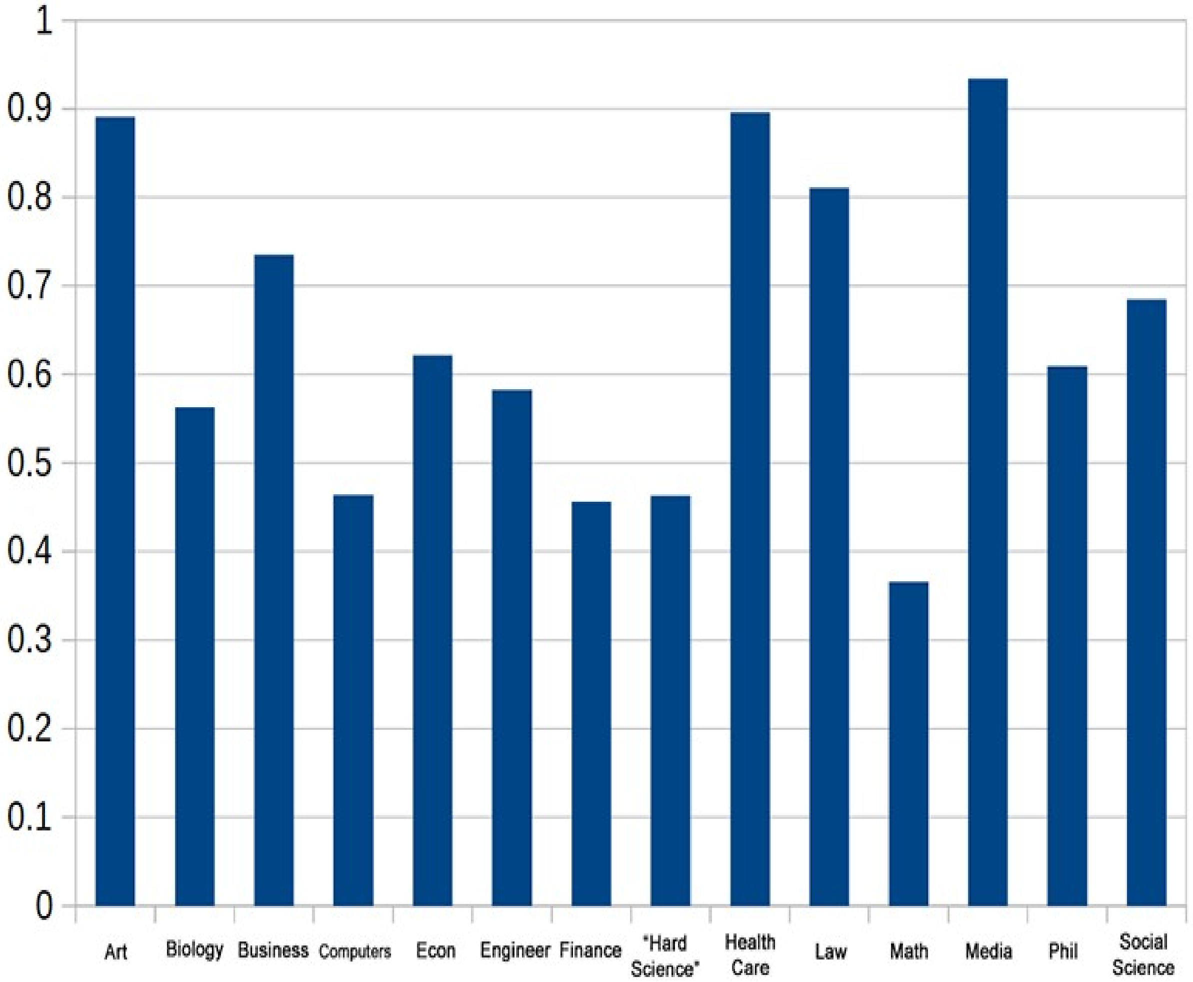


Figure 3: Percent men admitting to sexual harassment perpetration at work, by occupation type



Here's all three measures combined into my best guess for the relative rates in each field:

Figure 4: Combination score of sexual harassment by industry



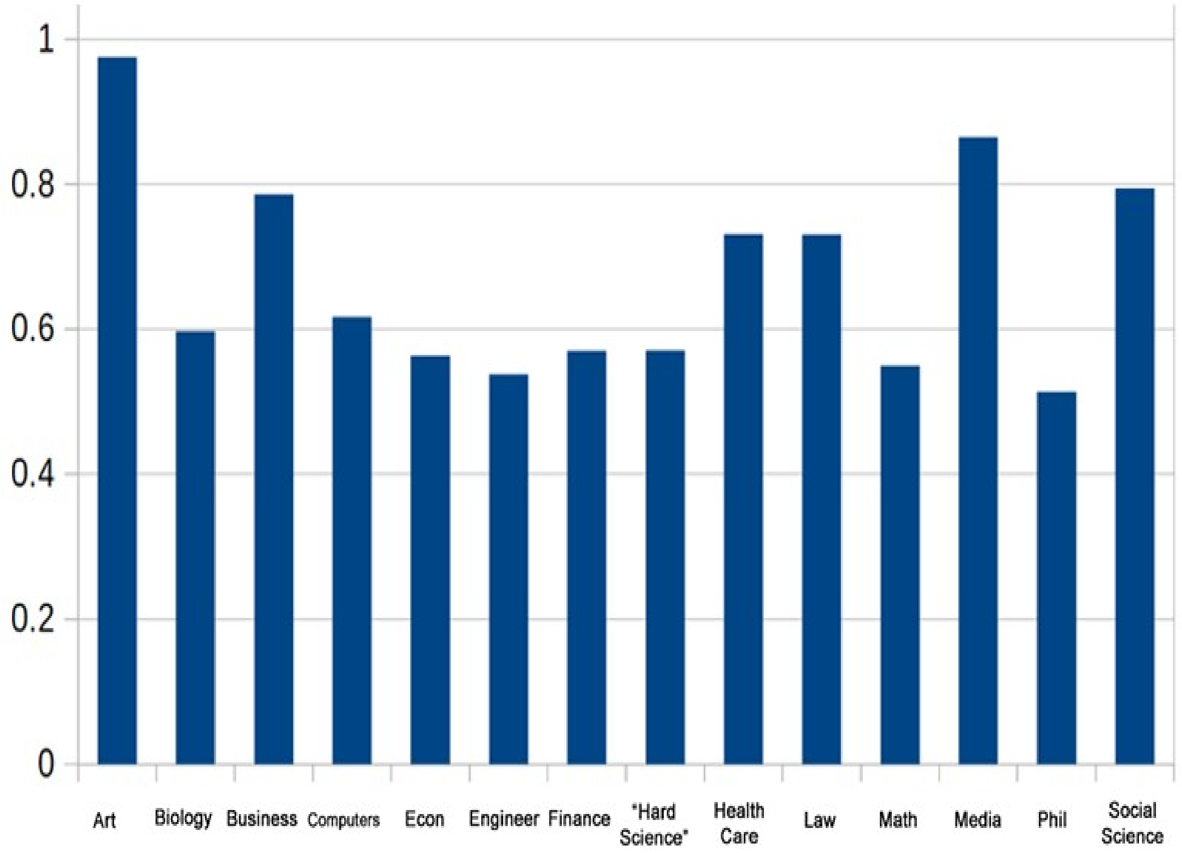
Made by taking each of the three previous graphs, standardizing each field to percent of the worst field in that category, and then adding them up and dividing by three to get an average for each field

But this combines three different things. First, the actual rate of harassment. Second, a measure of how willing people are to report harassment. And third, a measure of how willing people are to consider some specific act to be harassing. This is a major problem. People from traditionalist cultures and subcultures may have a higher threshold for calling something harassing; if (for example), more traditional people go into Business, and more socially liberal people go into Art, that could skew the numbers.

In order to control for this and other factors, I asked subjects how often they are harassed *outside of* work. Of note, men, women, victims, perpetrators, everybody in every industry, reported much more sexual harassment outside of work than in it. Although there were enough female perpetrators to do statistics on this time, I left them out to keep it as similar to the in-work data as possible.

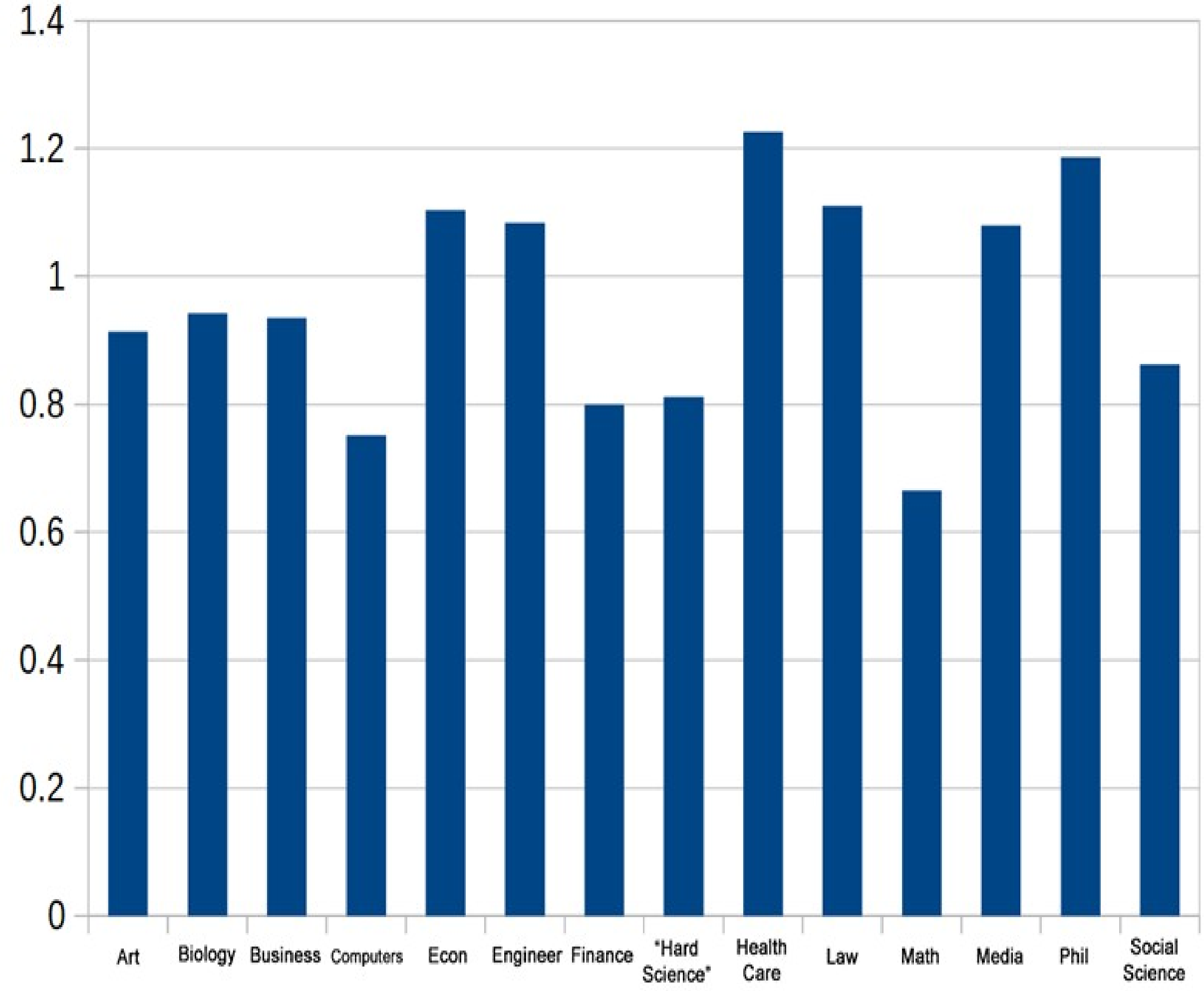
Here's an overall summary, made through the same combination process as Figure 4.

Figure 5: Combination score for out-of-work sexual harassment, by industry



Levels of at-work harassment and out-of-work harassment correlate at 0.81 across fields. This suggests that most of the differences observed here are differences related to the people in a field, rather than to any sort of field culture. Given how strong this effect is, it’s unclear whether it’s possible to adjust for it and get a measure of field culture. If we try (by dividing the at-work harassment levels by at-home harassment levels), we get this:

Figure 6: Ratio of at-work harassment to out-of-work harassment among employees in a given field



The most striking finding I find on all these graphs is that “nerdy” / STEM / traditionally-male jobs have the least harassment. The less nerdy / more verbal-personal-skills / traditionally-gender-balanced jobs have the most harassment.

Here’s Figure 4 again, but this time with nerdy-STEM-male occupations in red, verbal-personal-balanced occupations in blue, and edge cases in grey.

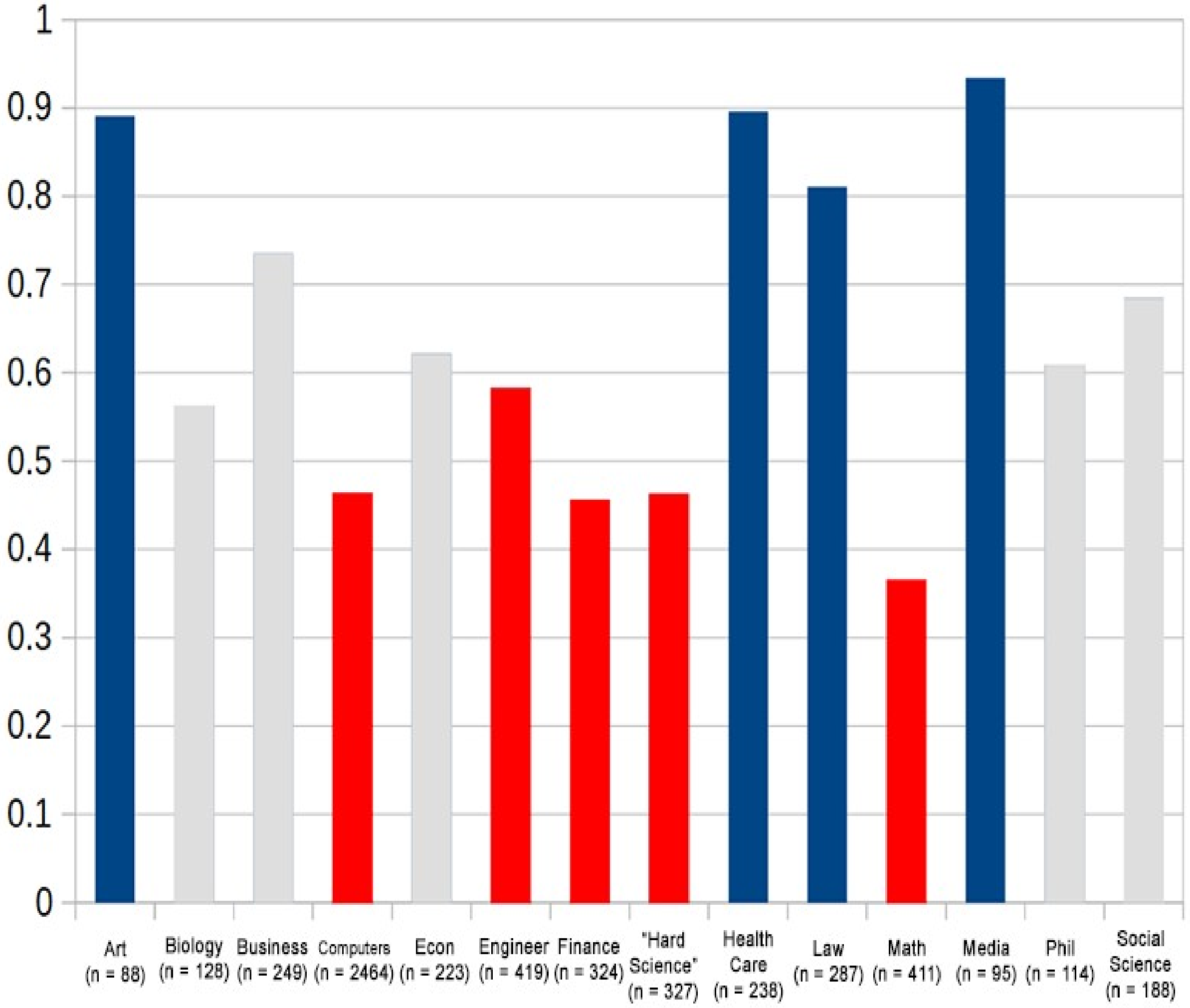


Figure 7: Sexual harassment by industry, STEM/nerdy/male (red) vs. interpersonal/creative/female (blue)

If categories like nerdy/STEM/traditionally-male seem unnatural, that's fine – these are always a judgment call. But in December, I preregistered which clusters I would use, to prevent me from making them up to get the result I wanted. It's not exactly the same as the clusters above – I ended out having to combine a few to get a big enough sample size to be meaningful. But using the original pre-registered clusters and the original pre-registered endpoint: in STEM/male-type fields, 12% of women (n = 212) reported being harassed at work. In non-STEM/balanced-type fields, 25% of women (n = 161) reported being harassed at work. p = 0.002. The preregistered analysis confirms the finding above.

I preregistered the hypotheses that STEM would have more female victimization (because there's a higher ratio of heterosexual men to heterosexual women), but the same amount of male perpetration (because there is not an inherent difference in the sort of people who go into the field). Both preregistered hypotheses were wrong; female victimization and male perpetration were less.

IV.

Overall victimization rates were comparable to past surveys. These were on the low end of reported results, likely for several reasons. First, we strongly encouraged participants who were at all uncomfortable to say "prefer not to answer", including asking 10% to say this outright regardless of experiences; this clearly depressed results and makes absolute numbers inaccurate. Second, participants were generally higher-income, and this group usually reports less harassment. Finally, we asked participants only about "sexual harassment" in general, and not about specific misbehaviors. Given these issues, the modest differences between these and past results (eg our 21% average female harassment rate compared to Cosmo’s 33%) are not surprising.

More surprising were the lower harassment rates in STEM and male-dominated fields, for two reasons. First, if most harassers are heterosexual, then a high percent male workforce in an industry should create a situation where a large number of harassers are focused on a small number of potential victims, raising the average harassment rate per victim. But this was not observed in our survey, nor in the Cosmo survey, nor (as far as can be eyeballed from the not-really-inter-comparable data) in the EEOC survey.

I don’t know of a simple explanation for the discrepancy. A friend suggested that it might be a customer effect. That is, lawyers / health-care-workers / etc often have to deal with sexist customers, whereas female programmers might not encounter anyone on the job except their coworkers. But this seems unlikely to make much difference; this survey finds that customers represent only 9% of sexual harassment incidents. Also, remember that the rates at which men admit to perpetrating sexual harassment match the rates at which women admit to experiencing it, and there’s no reason why the male-perpetration rates should have anything to do with customers. Also, social scientists don’t have customers any more than hard scientists do, but the social science harassment rate is higher. Perhaps the sort of people who go into STEM are socially oblivious in a way that prevents them from both noticing if they are sexually harassing someone else, and from noticing if they are sexually harassed. This would result in somewhat lower self-reported perpetration rates, and self-reported victimization rates that are even lower than what the low perpetration rates would predict. It would also match the high correlation between in-work and out-of-work harassment, which suggests that demographic factors are making most of the difference.

A second reason the low harassment rates in STEM were surprising is conventional wisdom claiming the opposite. In media coverage, STEM fields are portrayed as full of “techbros”, and uniquely unwelcoming to women. How do we square this with the data?

Some people who reviewed these data suggested that the media focuses more on a narrative where STEM workers are dismissive of women, rather than one where they harass them. I don’t find this accurate – a quick glance at any media source will show they focus on both narratives. But it seems possible that the dismissiveness narrative is correct, and the harassment narrative is a side effect of this. But this is somewhat contradicted by the auto industry survey mentioned above, which found tech was unexceptional either in harassment or in general dismissiveness. However, the other industries participating in the survey may themselves be unusually bad.

Another possibility is that the respondents to the survey may be skewed for some reason. For example, they may have been selected by this blog’s previous discussion of these issues, which have also argued that tech is not a uniquely bad industry.

But a final possibility is that the media coverage is inaccurate. The minimal amount of previous research in this area has all shown that the highest rates of sexual harassment are in the retail and restaurant industries. The media has ignored this. Before last year, the media almost never mentioned sexual harassment in Hollywood; now they hardly mention anything else. Given that already know media coverage here is generally inaccurate, maybe the data don’t need to be squared with it. Maybe the media is just wrong.

Every industry has enough sexual harassment to produce horror stories. If the media is disproportionately interested in the horror stories of relatively well-off people, and in stories that confirm their existing prejudices, we might hear many genuinely horrible stories about harassment in technology, but fewer about harassment in eg health care, philosophy, or law – even if objectively those fields are much worse. That might further encourage people in technology to come forward (consider what happened in Hollywood after the Weinstein revelations) and people in other fields to stay silent, contributing to a vicious cycle.

Compare all the anecdotes and popular lore about how immigrants are criminals. It's totally false – immigrants have crime rates well below native-born citizens. But we only know that because there have been really good studies. If the studies hadn't been done, and all we had to go on was the daily lurid stories about a Mexican guy knifing someone, who would believe it? In the absence of real studies, the media's ability to spin a compelling narrative casting some people as monsters feeds on itself forever. We know that happens relatively often. Are we sure this time is different?

There's only been one strong previous survey on this subject – the one in *Cosmo* I referenced earlier. They found the tech industry had an around-average to slightly-less-than-average rate of harassment. Anything else you've heard – any news story, any anecdote – is at the same level of evidence as the stories of immigrant crime.

It seems possible that this narrative grew up to explain why there are so few women in technology – maybe they're all being harassed out by lecherous professors and bosses. But this explanation doesn't hold water – the rate at which men vs. women choose tech courses in high school exactly parallels the rate at which they go into tech occupations.

And gender balance in fields isn't predicted by harassment levels in those fields anyway – otherwise you'd never see women working in restaurants, retail, or Hollywood.

I have been expecting results like these even before I did this survey, and this calls my data into question. Whenever someone with an unusual opinion produces an experiment that confirms their opinion, you should always be skeptical until it is verified by other sources.

So aside from appealing to the unimpeachable authority of *Cosmopolitan* as partially confirming my results, I urge anyone with the relevant skills to download the raw survey data themselves and see if they can replicate my conclusions. You can get them as .xlsx or .csv. Some people have complained of weird problems in the csv format and I recommend the xlsx if at all possible. I have removed the data of a few people who did not want their answers to be public, so you may not get exactly the same numbers I did, but they should be pretty close.

And more important, I appeal to anyone with an interest in this topic to do larger and more formal surveys comparing different fields. This may be the single topic where the extent of public interest is most disproportionate to the minimal amount of good research done. If people more prestigious than me or *Cosmo* were working on this, we would be in a much better position to know what to believe.

Highlights From The Comments On Survey Harassment Rates

[Content warning: harassment. This discusses the comments to SSC Survey Results: Sexual Harassment Levels By Field]

brmic writes:

> Thank you for posting this and the data file. FWIW, I tried to reproduce the results and couldn't reproduce the correlations between female victimization, male victimization and male perpetration. fem vic vs. male vic is 0.65, same as yours. fem vic vs. male perp is 0.01 for me, and male vic vs. male perp is 0.21 for me. Everything else more or less checks out.
>
> As a reviewer, I'd say the combination score is not convincing, especially since it ignores all considerations of different male to female ratios in the various industries.
> Also, if you have two measures with r = 0.8, Fig 6 is not a good idea IMHO. It's probably just noise. (Also, it should be a dotplot centered around 1, because the relevant info is distance from 1:1 ratio.)
> Instead, I'd focus on the correlation between female victimization at work and female victimization outside work of 0.65 (for me) and the same for males at 0.59, which also leads to the conclusion that there's a strong 'people in fields' effect, without having to go through the combination score. If you're so inclined, you might then do the at-work by outside-work ratios and end up a kind of cross-validation set, where you can see whether the bad fields for women are bad for men as well. Of course, once you then consider sex ratios per field. it's story time all over again. Still, e.g. men report similar

> levels of out of work victimization in computers (20%) and Health Care (24%), but at work victimization of 4% and 12% respectively, which strongly suggests that Health Care is worse.

Their code is available here. Thanks for doing the work to try to replicate my results. I've removed the non-confirmed correlations from my post until I can figure out what's going on with them. I agree that Figure 6 was barely worth it, which is why I tried to make Figure 4 (the unadjusted version) the center of my thesis.

Chris quotes a TIME article that argues that predominantly-male communities generally have lower harassment rates than predominantly-female communities:

> Given the epidemic of campus rape, teenage girls and their parents are justifiably concerned about safety, just as teenage boys and their parents are worried about false accusations. What does any of that have to do with gender ratios? Well, there have been multiple studies showing a correlation between gender ratios and rates of sexual assault. As counterintuitive as it may sound, elevated rates of sexual assault are a predictable feature of communities with oversupplies of women, according to studies by sociologists Nigel Barber and Robert O'Brien.
>
> The opposite is true of communities with oversupplies of men. Columbia University economics professor Lena Edlund investigated the impact of lopsided sex ratios in China, where young men now outnumber women by 20% due to sex selection, abortion, female infanticide, and other outgrowths of China's old "One Child" policy. Edlund and her co-authors discovered that although overall crime rates went up in China as the gender ratio

> skewed more male—not surprising given that men are more prone to criminality—there was a precipitous decline in rape. It seems that men treat women better, and protect them more, when women are in shorter supply.
>
> Can I prove beyond all doubt that Edlund's and Barber's findings also apply to college campuses—i.e. that rape is less common at schools that are at least half male? No, because the available data on campus rape tends to reveal as much about how forthright colleges are in handling sexual assaults—and how comfortable women feel reporting them—as it does about the actual frequency of assaults on a particular campus.
>
> That said, I was intrigued by a recent Washington Post story on the topic. The article ranked 27 top colleges by their sexual assault rates, and I couldn't help but notice which college had the lowest rate.
>
> It was CalTech, a school that is 59% male.

But some commenters in the subreddit bring up contrary evidence. DinoInNameOnly links to a Pew poll finding that women in predominantly-male workplaces report more harassment. And Hepatitis Andronicus brings up Alaska, a heavily male-skewed state considered to be in "a sexual assault state of disaster".

Rolaran argues:

> I feel like a person who has had bad experiences in the tech industry is less likely to have continued reading a blog that argues (among other things) that problems in that field are overemphasized, and a person who has had nothing but good experiences in tech is more likely to have begun hanging around a blog where they felt

> like the author was sticking up for them. This is useful data for determining the experiences, preferences and demographics of SSC readers, but I would be considerably surprised if that could be generalized without some serious legwork.

I did consider this, but first of all, only the tiniest percent of my posts are on tech, and I think I might have only had one previous post that could by any stretch of the imagination have been considered about harassment in tech. Second, it’s not clear which direction this should skew things – some would argue that guilty people are more likely to push the “we have nothing to apologize for” line, and innocent people are more willing to critically self-examine their (or their tribe’s) problems. Third, I’m not sure this would really skew self-reported female victimization, self-reported male victimization, and self-reported male perpetration all in the same way.

Rachael writes:

> I agree with most of your conclusions, but there’s still a major flaw in the survey methodology. It asked what industry people currently work in, and whether they’ve *ever* been harassed at work. Plenty of people have switched industries over the course of their career, and nearly everyone worked in retail or food service or similar as a teenager or young adult. So we have no idea how much of the reported harassment took place in industries other than the ones people currently work in, and probably skewed towards the industries that hire teenagers.

This is not only a good point, it’s a good point that would have been easy for me to fix if I had thought about it. Sorry.

Leah Velleman writes:

> I wonder whether something gets lost here when we conflate the startup world (stereotypically an unsupervised bunch of pushy, risk-taking young single people) with — and I use this term lovingly — dinosaur tech companies (stereotypically a bunch of mild-mannered old married people, with a strong HR department looking over things).
>
> The cultural gap between startups and dinosaur tech seems much bigger than, say, the gap between independent and chain restaurants, or small and large universities. If that's true, then that's an argument for treating startups and dinosaur tech as separate sectors. (And while some companies might be hard to sort into one category or the other, that's true for any set of categories. "Is this job in econ or finance?" and "Is this job in business or law?" will turn up ambiguous cases too, but that's not a reason to ditch those distinctions.)
>
> I'm pushing this question because I have a hypothesis, which is that startups really do have a harassment problem — but that the people speaking up about it, and the news outlets quoting them, have misrepresented this as a problem across all of tech, ignoring the fact that most tech workers are at boring old companies with low harassment rates. I suspect this misrepresentation has happened partly because startup folks like to represent themselves as The Real Tech Industry, The One That Counts, and partly because companies tend to transition from "startup" to "dinosaur" over time and news outlets can't be bothered to keep track.

Riceowlguy on the subreddit writes

> I think it would be worthwhile to try and distinguish between “people whose job is actually writing code or keeping IT systems running smoothly” versus “people who have jobs at tech companies but whose work is not technical”. My recollection is that a lot of the anecdotal horror stories about harassment in Silicon Valley involve VCs, sales/marketing people, upper management, etc.
>
> One question is whether people in that latter category (if any of them read SSC at all!) answered “computers” or “business” for their profession.

Doug writes:

> Maybe men in these groups are just less sexually aggressive and more introverted. It’d be interesting to aggregate # of lifetime sexual partners by industry and correlate with levels of sexual harassment. Not that I’ve ever been involved in sexual harassment on either side of the table… But my guess is that most cases don’t look like Alice walking over to Bob’s desk and grabbing his crotch out of the blue. Most situations probably escalate from some fun, flirty, or social interaction that one party takes to far. Programmers don’t know how to flirt and avoid socializing, so they never get into these situations. Existing literature on sexual assault shows that it disproportionately occurs in environments of “revelry” and “carousing”. I don’t see any reason workplace harassment would deviate from this pattern. Pharma reps get drunk with their coworkers a hell of a lot more than mathematicians.

I’d been thinking of “introversion” as the default hypothesis, not requiring explanation in the same way as the others, but I realize I might have forgotten to mention it at all. Sorry.

But Terran has analyzed (original comment, small correction) whether these results are explained by the Big 5 personality traits of people in each field. He finds they generally are not – profession remains important even after looking at traits (including introversion). Stezinec on the subreddit tries the same thing – their analysis finds that, after adjusting for traits, the only significant result is that computer-related occupations have lower harassment – but guessing this is just a sample size issue and the computer industry was the only one with a good sample size.

In case you are interested, here are some other factors that affected sexual harassment in the survey. I'm not putting error/significance bars on because that would be too much work – but if I have to have a nice-sounding explanation, it's because I want to discourage people from using these to say "Your ingroup is more harassy than my ingroup, proving you're bad". The best one can do without error bars is *disconfirm* statements like that – which I think is generally a more prosocial activity. This is also part of why I am making this section less prominent than it might otherwise be. I think this especially important to remember given that groups will be penalized for honesty – ie people who are more likely to confess to sexual harassment, or people who are more willing to admit that certain behavior crosses the line.

All analyses were limited to cisgender subjects except where stated otherwise. All analyses of women have a sample size around 600, all of men around 7000. Each subgroup has a sample size of at least 100 except where stated otherwise. Note that each graphic combines information on victimization and offending by one gender; offending results are often very low and slight variations should be taken with a grain of salt.

Rates by relationship status (male, female)
Rates by sexual orientation (male, female)
Rates by transgender status (male, female)
Rates by asexual vs. sexual (male, female)
Rates by political affilitation (male) [not enough data for full female analysis, but binary left/right graph here]
Rates by feminist self-identification (male, female)
Rates by social class of family of origin (male, female)

If you want to analyze this further or double-check any of my results, you can download the raw data here.

Gupta On Enlightenment

That story about the blockchain-based dating site gets better: its designer is an enlightened being.

I got this from Vinay Gupta’s wiki, which describes some of his thoughts and experiences. Since reading Mastering The Core Teachings Of The Buddha, I’ve been looking at a bunch of this stuff, and it’s interesting how it does (or doesn’t) converge. For example, from the MCTB review:

> If you really, really examine your phenomenological experience, you realize all sorts of surprising things… one early insight is a perception of your mental awareness of a phenomenon as separate from your perception of that phenomenon.

And from Gupta:

> The real process of meditation is paying real close attention to what is happening around you without passing it to the mind immediately for analysis…the mind becomes perceived to be another sense. You see, you listen, you hear, you smell, you think. Once you are aware that you are not your mind and your mind is basically a sense organ, it’s a thing that brings information to you, you enter the real work of enlightenment, which is: what is this me that the mind is bringing information to? And that’s the big one. That question is at the heart of everybody’s enlightenment process.

From the MCTB review:

> The main point of [mindfulness] meditation is to improve your concentration ability so you can direct it to ordinary experience. Become so good at concentrating that you can attain various jhanas – but then, instead of focusing on infinite bliss or whatever other cool things you can do with your new talent, look at a wall or listen to the breeze or just try to understand the experience of existing in time.

From Gupta:

> Building the instrumentation to keep your consciousness stable enough to put the attention on the thing, is about three or four years work. It's like grinding a mirror if you're going to make an astronomical telescope. It takes years to grind a perfectly smooth reflector. Then you silver coat it. Then you point it at the sky and now you can see the moons of Jupiter. It takes you years to design the microscope, you look into the water, now you can see the microbes and you just discovered germ theory. Building the instrumentation takes time. Years and years and years because you need long periods – 35, 40 seconds minimally – when there are no thoughts in the mind to be able to begin to turn the awareness onto itself. So lengthening the gap between thoughts means lowering the mental background noise.

There are lots of these matches. My first impression was that it's a good sign that people are finally converging on being able to talk about this kind of thing in a sensible way. But here's a point Gupta makes over and over:

> The weird thing is – everyone who opens up the big door and looks out into the magical Universe where all the cosmic shit lives, sees something different. The purpose of religions is to enforce conformity on the mythology

that floods your brain once you open up the cosmic forces.

If you are a strict moslem and you experience your enlightenment in a moslem context, the mystical model of the world that gets slammed into your head when you finally look at the Universe in that way, will be in conformity with the dominant culture around you at the time.

This is part of the reason that everything in Western culture went nuts when they discovered LSD, because you had all these people experiencing enlightenment outside of the conformity of the church. So rather than becoming Saint Ignatius of Loyola, you wound up as acid-crazed Bill. I've got this mythology of the Universe, and it's all to do with Spiral Dynamics. My name is Ken Wilbur. Where the hell did that come from? He made it up and then told you it was cosmic law. Just like all the others did.

Everybody experiences the mythological aspects of enlightenment on their own terms, and if they are a slick talker, they can convince you that's how it works, and then when you experience enlightenment, you experience the same mythology you were loaded up with.

This is how it really works. You've got your Buddhas and your Christs and your Mohammeds, and your Abrahams and all the rest of these people – they experience these cosmic states of consciousness, they generate their own mythology and then they run around telling you they've discovered the secrets of the Universe – you should do it their way now.

Later in the session, a questioner asked:

> I don’t believe you. I think what you’re describing, getting rid of frameworks, is just a new framework. You’re just talking about another way of understanding consciousness. It’s the same as the Stoics and it’s the same as the mystics and it’s the same as Nietzsche, just another perspective. You didn’t talk about anything specific. You talked about some beautiful abstractions. It feels to me like the things you’re talking about in very abstract terms are the same things that every other philosopher talks about. You haven’t got rid of frameworks.

And Gupta answered:

> Every individual who goes up there sees the same shit, more or less. And then you come back down and try and tell people about it in language, and you wind up building a model that you use to communicate. That is exactly correct. It’s the same shit.

So. I read Ingram, and I read Gupta, and they seem to be saying broadly the same stuff, and it appeals to me, and seems to fit with what I already know of the world, and gets me thinking that all this enlightenment stuff is starting to make sense. But (says the devil’s advocate) two Christian saints may have similar experiences of the Beatific Vision, write them down in similar terms, and the average Christian will nod and say it agrees with what they already know of the world (eg that it is run by a triune God who lives in the Empyrean). Does “scientifically minded, religiously tolerant people come back and say it’s all science, and also all religions are one” give us a better framework than “Christian people come back and say it’s all Christ”?

One possible escape from total relativism: forget about whatever's on the other side of enlightenment. We can at least trust people to report on this side of the veil accurately, and that's where some of the most interesting insights are. The description of what meditation is doing. The distinction between samatha and vipassana meditation. The idea of the mind as a sensory organ rather than the be-all-and-end-all…

…actually, wait, that one doesn't sound very scientific at all. Maybe we should reframe it as talking about two different parts of the brain? Maybe this isn't as easy as I thought.

I've been focusing a lot lately on the idea of the Bayesian brain and its input channels. Some input channels, like vision, are high-bandwidth; we get so much data about the real world that (optical illusions and PARIS IN THE THE SPRINGTIME signs aside) we usually see pretty much what is really there.

Other channels, like pain, are low bandwidth. This is why the placebo effect works – we get so little data about how much pain is coming from different parts of our bodies that even our strongest percepts are wild guesses, where we fill in the gaps with predictions and smooth away conflicting evidence. If our predictions change – ie we know we just got morphine and morphine lowers pain – then the brain will happily change its guesses. This would never happen with vision – I can't use the placebo effect to make you think an orange crayon is blue – but pain is low-bandwidth enough that it works.

Reason is one of the lowest-bandwidth channels of all, which is why biases are so omnipresent and rational debate so rarely changes anyone's mind. Most people revert to their priors – the beliefs of their tribe or the ones that fit their common sense – and you have to provide an overwhelming amount of rational evidence before the brain notices anything amiss at all.

It sounds like, in this model, enlightenment is effectively super-low-bandwidth. I say “effectively” because the bandwidth concept doesn’t really make sense here, maybe it has more to do with the alienness or uncompressability of the information. But Gupta seems to be saying that you will see it *exactly* as you have been conditioned to see it. That wouldn’t be too surprising. But it sure does suck if you’re trying to figure out which religion is true, or prevent people from becoming religious fanatics, or anything like that.

This is a pretty agnostic (in the sense of non-knowledge-claiming) description of what enlightenment is. But it does at least suggest it’s…something…in the brain…that goes through the normal perceptual process? Except that realistically if you see a rhinoceros the sense-data will be in your brain and go through the normal perceptual process, but that doesn’t mean a rhinoceros is just brain activity. Except that nobody was claiming that your perception of a rhinoceros is actually *more* fundamental than ordinary brain activity, and some people do claim that about enlightenment, so maybe this is telling us it isn’t? Or something?

Other interesting excerpts from Gupta:

> Every tradition that has enlightened people has stories of wizards. The Daoists that run across water, all this Crouching Tiger, Hidden Dragon stuff, the European Alchemists – do you guys know about a guy called John Dee? Course. Course! What kind of audience do you think this is?

So John Dee for those who are not overly read, was Queen Elizabeth the First’s court magician. John Dee has one essential claim to fame, which is that he invented the concept of the British Empire. He wrote two books arguing that, as the Romans had used roads to create a trade network, and to move armies around an Empire, Britain could use ports and ships. So he invented the concept that we would reimplement what the Romans had done but with London at the centre.

John Dee’s primary work, what he was proudest of in his life, was a 400 page volume of angelic magic called the Enochian Magical System. So you look at this and you’re like, the guy’s smoking crack. Then you look at Isaac Newton. Newton’s laws of motion, colours, he named the colours of the rainbow. By the way he named the colours of the rainbow with seven colours, even though indigo and violet are the same colour. Because he needed one colour per planet and one colour per alchemical force. Three quarters of Newton’s work is alchemy and that’s all the stuff where everybody’s like “Oh Isaac Newton he was such a nonsense lover – all this alchemy stuff. Love the laws of motion though.” Because you’re not allowed to take the other side of these men’s work seriously because if you do – Voomp – Oh my fucking god – that’s really there! Yes, it’s really there.

The weird thing is – everyone who opens up the big door and looks out into the magical Universe where all the cosmic shit lives, sees something different. The purpose of religions is to enforce conformity on the mythology that floods your brain once you open up the cosmic forces.

I keep hearing people talk like this, but I would really like to see some analysis of how the Western hermetic and alchemical traditions match the Eastern enlightenment traditions. I know a bit about John Dee, and it all suggests he was a very weird and gullible person, and none of it sounds like meditation as the Buddhists and Hindus think of it. The same is true of Newton's mystical researches. I'm open to arguments for why these things are really the same deep down, but somebody needs to actually argue it instead of just gesturing at it, and I've never seen this done well. Aleister Crowley was neck-deep in the western mystical tradition, hung out in Sri Lanka for a while studying yoga, and (when he wasn't being deliberately obscure) is one of the clearest and most lucid writers I've ever had the privilege of reading – and even he never actually made this argument in any comprehensible way. Come on, people.

> Lowering the mental background noise means going through all the emotional layers and all of the attachments that generate thought. A single emotion that you don't really deal with properly can generate 5 years of internal chatter. Should I? Shouldn't I? Should I? Shouldn't I? You finally come back and it's this deep feeling of uncertainty about your place in the world. You feel it – it goes away. You've been liberated of an emotion, that stream of thought stops. And as a result your mind gradually empties and empties and empties and empties.
>
> If you've been taught that you are your mind, that process feels like dying. This is why there's all this nonsense about the abyss in the Western magical tradition. "Oh the Abyss. Oh the Abyss." You go to India; they've never even heard of the Abyss. Because in India they don't think that you are your mind. So having mind go away "Really, that thing back there." "Yes." "I used

> to use that for saying mantras – now it doesn’t work any more.” Whereas in the West, if your mind stops, that means your identity is gone, and everybody freaks out and calls that the Abyss.

Okay, this is actually one of the better East-West mystical comparison theories I’ve heard.

> So, how much meditation is a lot of meditation? Typically to get enlightened takes about as much work as getting a PhD. So you would expect it to be the dominant occupation of your life for something between 7 and 10 years, including working your ass off for your A-levels, getting through an undergraduate degree, doing a Masters, doing a PhD. Getting enlightened is about a PhD’s worth of work. Very few people in the West claim to be enlightened, even fewer of the people who claim to be enlightened are enlightened and even fewer of them are doing anything other than teaching.
>
> About 15 I started to meditate, about an hour a day, sometimes 2. I was physically ill at the time; I had nothing but free time. Although I’m half Indian, I had no real exposure to Hinduism as a tradition. I just started to meditate because there was nothing else to do and it seemed to help. After 6 years of an hour or something a day, after a very, very intense, shall we say, “collaborative celebration”, in the morning after the trip, we were having a kind of debriefing session. In my head, as we were talking, I saw an amplifier, just a very simple aluminium amplifier with a big knob, little blue LED on it, and I saw my hand reach down and turn the knob off. And my internal dialogue completely stopped. This was about 1993, 1994 and it never came back.

> Living in the condition of having no internal dialogue, no flow of thoughts, no flow of images, just Smack, into the present is quite an abrupt thing. For the first couple of weeks I thought I'd gone completely mad. Oh my god I've totally broken myself. I'm fucked. And I discovered that I could still go to work, and I could still socialise with people and I could cook and get through all the basic things of life. Nobody outside of me seemed to notice any particular change in my behaviour, even though I was lost in this rapturous state of total absorption with the world. Wow, this is amazing, woah! And then life continued.
>
> I'd run right off the edge of every reality map that I had because if you go to a psychologist or a psychiatrist and say, by the way I did really a lot of meditation and my internal dialogue has totally stopped. Any ideas what I do now? Nobody ever winds up there in the West because nobody does enough meditation, at least they don't do it right.

Actually, sometimes people *do* come to psychiatrists with these kinds of complaints. I usually try to explain what's going on, and they usually tell me they were just meditating because someone said it relieved stress, and nobody warned them they could actually have mystical experiences, and this was not what they signed up for. Symptomatic treatment and a hard ban on further meditation successfully de-mysticize most of these people, and they are able to go back to their regular lives. I assume if there's an afterlife some sort of cosmic wisdom deity is going to be very angry at me – but hey, I'm just doing my job.

Mental Health On A Budget

Everyone knows medical care in the US is expensive even with insurance and prohibitively expensive without it. I have a lot of patients who are uninsured, or who bounce on and off insurance, or who have trouble affording their co-pays. This is a collection of tricks I've learned (mostly from them) to help deal with these situations. They are US-based and may not apply to other countries. Within the US, they are a combination of legal and probably-legal; I've tried to mark which is which but I am not a lawyer and can't make promises. None of this is medical advice; use at your own risk.

This is intended for people who already know they do not qualify for government assistance. If you're not sure, check HealthCare.gov and look into the particular patchwork of assistance programs in your state and county.

I. Prescription Medication

This section is about ways to get prescription medication for cheaper. If even after all this your prescription medication is too expensive, please talk to your doctor about whether it can be replaced with a less expensive medication. Often doctors don't think about this and will be happy to work with you if they know you need it. They might also have other ways to help you save money, like giving you the free sample boxes they get from drug reps.

1. Sites like GoodRx.com. This is first because it's probably the most important thing most people can do to save money on health care. For example, one month of Abilify 5 mg usually costs $930 at Safeway, but only $30 with a GoodRx coupon. There is no catch. Insurances and pharmacies play

a weird game where insurances say they'll only pay one-tenth the sticker price for drugs, and pharmacies respond by dectupling the price of everything. If you have insurance, it all (mostly) cancels out in the end; if you don't, you end up paying inflated prices with no relation to reality. GoodRx negotiates discounts so that individual consumers can get drugs for the same discounted price as insurances (or better); they also list the prices at each pharmacy so you know where to shop. This is not only important in and of itself, but its price comparison feature is also important to figure out how best to apply the other features in this category. Even if you have insurance, GoodRx prices are sometimes lower than your copay.

2. Get and split bigger pills. Remember how a month of Abilify 5 mg cost $30 with the coupon? Well, a month of Abilify 30 mg also costs $30. Cut each 30 mg pill into sixths, and now you have six months' worth of Abilify 5 mg, for a total cost of $5 per month. You'll need a cooperative doctor willing to prescribe you the higher dose. Note that some pills cannot be divided in this way – cutting XR pills screws up the extended release mechanism. Others like seizure medication are a bad idea to split in case you end up taking slightly different doses each time. Ask your doctor whether this is safe for whatever medication you use. Do *not* ask the pharma companies or trust their literature – they will always say it's unsafe, for self-interested reasons. Contrary to some doctors' concerns, this is not insurance fraud if you're not buying it with insurance, and AFAIK there's no such thing as defrauding a pharmacy.

3. Pharma company patient assistance programs. As part of their continuing effort to pretend they are anything other than soulless profit-maximizing bloodsuckers who will be first against the wall when the revolution comes, some pharma companies offer their drugs for cheap if you can

prove you need them and can't afford the regular price. These are most useful if for some reason you need a specific expensive brand-name drug; if you have any other options you're better off just buying the generic. You can search for these programs at Partnership For Prescription Assistance, RXAssist, and NeedyMeds. Be very careful to read the fine print on these, because no matter what they pretend, drug companies *are* soulless profit-maximizing bloodsuckers who will be first against the wall when the revolution comes, and sometimes these are just small discounts that aren't as good as using one of the other methods. Occasionally a company will give you a great discount that knocks a brand-name medication costing $300 down to only $150 without telling you that there is a similar generic that costs $5. But if you need one specific very expensive thing, and you are lowish-income, and you don't have government help, this is still your best bet.

4. Get 90+ day supplies. If your insurance charges you a co-pay of $30 per prescription, and you get a 90-day supply instead of a one month supply, then you're paying $30 once every three months, instead of once a month.

5. Mail order from legitimate Canadian pharmacies. Canada has lower prices than the US for various prescription drugs. Canadian pharmacies are unlicensed and unsafe and you should never use them, according to the same people who tell you that marijuana is a gateway drug and porn will fill your computer with Russian viruses. According to everyone else, including most doctors I know, they are fine as long as you avoid obvious scams. They are technically illegal but the FDA has a policy not to prosecute people who buy drugs there for personal use. The Canadian Internet Pharmacy Association maintains a list of ones they consider safe. If I try really hard, I can find a way to get the month of Abilify 5 mg for $4.58 from canadapharmacy.com, but this isn't really

that much better than the best American option. Some other medications do seem to be better, especially ones that are still on patent; if I want a month of Saphris 10 mg, the best I can find on GoodRx is $620, but on canadapharmacy.com there's a deal for $196.

6. Mail order from sketchy Third World pharmacies. This is a high-risk option, with its only advantage being that they will sell you drugs without a prescription. If your problem is that you can't afford to see a doctor to get your script, but you know exactly what medication you need and how much, these can sometimes be a tempting option. I very much anti-recommend them, but also don't want to judge people who are too poor and desperate to do things the legitimate way. 95% of these pharmacies are outright scams, in the sense of stealing your credit card information and sending you no (or fake) medication. But I *suspect* that alldaychemist.com and inhousepharmacy.vu are legitimate. I say this based on testimonials from Longecity, from ratings on pharmacyreviewer.co, from various online review aggregators, from a very few formal studies, and from the recommendations of the transgender community, who for obvious reasons are acknowledged experts in the medical black market. Note that all the usual pharmacy rating websites will say these places are scams, because they classify any pharmacy that doesn't require prescriptions as a scam by definition. Note also that even some online pharmacies that are not technically scams will still be scummy in various ways (sell your contact details to spammers, etc). Caveat emptor.

II. Therapy

This section is on ways to do therapy if you cannot afford a traditional therapist.

1. Sliding scale or training clinics: “Sliding scale” means “charge people less if they can’t afford the regular prices”. Some clinics, especially those that train student therapists, will do this. There’s a list of good sliding scale clinics in the East San Francisco Bay Area here; if you’re somewhere else, you can check the list of SSC-recommended mental health professionals and sort by cost, or google “sliding scale therapy [my area]”.

2. Bibliotherapy: If you’re doing a specific therapy for a specific problem (as opposed to just trying to vent or organize your thoughts), studies generally find that doing therapy out of a textbook works just as well as doing it with a real therapist. I usually recommend David Burns’ therapy books: *Feeling Good* for depression and *When Panic Attacks* for anxiety. If you have anger, emotional breakdowns, or other borderline-adjacent symptoms, consider a DBT skills workbook. For OCD, *Brain Lock*.

3. Free support groups: Alcoholics Anonymous is neither as great as the proponents say nor as terrible as the detractors say; for a balanced look, see here. There are countless different spinoffs for non-religious people or people with various demographic characteristics or different drugs. But there are also groups for gambling addiction, sex addiction, and food addiction (including eating disorders). There’s a list of anxiety and depression support groups here. Groups for conditions like social anxiety can be especially helpful since going to the group is itself a form of exposure therapy.

III. Supplement Analogues

This section is for people who can no longer afford to see a doctor to get their prescription medication and who don’t want to go the sketchy Third World pharmacy route. It discusses what supplements are most similar to prescription medications. This is *not* an endorsement of these substitutions as exactly as good as the medications they are replacing, a recommendation to switch even if you can still get the original medication, or a guarantee that you won't go into withdrawal if you switch to these. They're just better than nothing. Make sure to get these from a trusted supplier. I trust this site, but do your own investigation.

This doesn't include detailed description of doses, side effects, or interactions; you will have to look these up yourself. These are all either legal, or in a gray area of "probably legal" consistent with them being very widely used without punishment. I am not including illegal options, even though some of them are clearly stronger than these – but you can probably find them if you search.

1. Similar to SSRIs: 5-HTP. This is a serotonin precursor that can serve some of the same roles that selective serotonin reuptake inhibitors do, though this is still controversial and it is probably not as strong. Cochrane Review thinks that "evidence does suggest these substances are better than placebo at alleviating depression". This may plausibly help with SSRI withdrawal, though not as much as going back on an SSRI. It can be dangerous if you are taking any other serotonergic medication, so check with the doctor prescribing it first. Cost is about $10/month. Definitely legal.

2. Similar to antidepressants in general: Tianeptine. This is a European antidepressant which is unregulated in the US, making it the only way I know to get an regulatory-agency-approved antidepressant without a prescription. Look up the difference between the sodium and the sulfate versions before you buy. Generally safe at the standard dose; higher doses carry a risk of addiction. Cost is about $20/month. Probably legal, widely used without legal challenges.

3. Similar to stimulants: Adrafinil. This is the prodrug of modafinil, a stimulant-ish medication widely used off-label for ADHD. Modafinil itself is Schedule IV controlled (though widely available online); adrafinil is unscheduled and also widely available. Look up the debate over liver safety before you use. Cost is about $30/month. Probably legal, widely used without legal challenges.

4. Similar to anxiety medications: L-theanine. This is a chemical found in tea, widely sold a an anxiolytic supplement and probably the most commonly used over-the-counter anxiety relief. It appears to weakly block glutamate, an excitatory neurotransmitter involved in the anxiety response. See here for a review of the scientific evidence. Cost is $10/month or less. Definitely legal. This will not work as well as benzodiazepines and will not treat benzodiazepine withdrawal.

You can find a discussion of other supplements for depression at Part IV here and for anxiety at Part IV here. You can find a discussion of ways that supplements can play a very minor role in helping with psychosis in the second to last paragraph of Part 12 here, but please don't rely on this. I no longer 100% endorse everything in those lists.

If you know other safe and legal ways to save money on psychiatric care, please mention them in the comments and I'll add them as they come up.

Call For Adversarial Collaborations

An adversarial collaboration is an effort by two people with opposing opinions on a topic to collaborate on a summary of the evidence. Just as we hope that a trial with both prosecutor and defense will give the jury a balanced view of the evidence for and against a suspect, so we hope an adversarial collaboration will give readers a balanced view of evidence for and against some thesis. It's typically done for scientific papers, but I'm excited about the possibility of people applying the concept to to less formal writeups as well.

For example, a pro-gun activist might collaborate with an anti-gun activist to write a joint article on the evidence for whether gun control saves lives. We trust each person to make sure the best evidence for their respective side is included. We also trust that they'll fact-check each other and make sure there aren't any errors or falsehoods in the final document. There might be a lot of debating, but it will happen on high-bandwidth informal channels behind the scenes and nobody will feel like they have tailor their debating to sounding good for an audience.

I don't know to what degree true adversarial collaborations are really possible. It might be that people who disagree on high-level issues might not be able to cooperate on a survey of the field at all. But I'd like to find out.

So I'm offering a prize, plus a chance to get the results published on SSC, to any teams (probably of two people each) who want to do adversarial collaborations. If you want to participate, comment on this post with what subject you'd like to work on and what your opinion is on the subject. Or look through existing comments, find someone who has the opposite opinion to you on a subject you care about, and reply to them saying you want to be their foil. After that you can exchange emails and start working.

If at least five teams participate, there will be a prize of $1000 for whichever team I think does the best work. There might also be a prize of $250 for a second-place team if they do exceptionally good work, though I am not promising this. Thanks to everyone who donates to this blog’s Patreon for providing the money to make prizes like these possible. Here are some more rules:

1. You will write an essay summarizing your joint summary of the evidence regarding a controversial topic you disagree on. Strongly recommend that this be a single factual issue, like “Does gun control save lives on net?”, rather than a vaguer moral question like “Guns – good or bad?”, though it can still be a pretty broad topic – I would love to see people write about Caplan’s case against education, for example. Even though most of the examples here are political, this doesn’t have to be; it could involve controversial topics in medicine, history, religion, et cetera.

2. You will write the essay as a united front. Please don’t write “Alice says this study proves guns save lives, but Bob says it’s wrong and this other study proves guns are bad.” Instead you are going to have to come to an agreement on how to describe each study. For example “Here is a study purporting to show that guns save lives. It seems to accurately describe what is going on in rural areas, but it might be of limited applicability elsewhere.”

3. You will come to at least some sort of unified conclusion, even if that conclusion is “There’s not enough evidence in this field to be sure either way and we should default to our priors/biases”.

4. The essay should be similar in length, tone, and amount-of-research to one of my Much More Than You Wanted To Know essays, eg here and here.

5. By entering the contest, you are giving me permission to publish your essay on SSC (with full attribution to you, of course). You can also publish it wherever else you want. I will probably publish the winning essay, and I might or might not publish the others depending on how good they are.

6. Because of (5), please don’t research any topic that I would not be able to publish on SSC if you came to a taboo conclusion. If you want to do an adversarial collaboration on taboo topics, you can feel free to arrange it in the comments, but it won’t be considered an official entry, it won’t be eligible for prizes, and I probably won’t post it (I might link it if it’s posted somewhere else). If you’re wondering whether a specific topic is taboo, you can ask.

7. If you’re officially proposing a collaboration or responding to a proposal, please put those comments in **bold** so people can find them amidst the discussion in the comment section. I may edit the timestamps on comments to bring these to the top, or even to bring the most interesting ones furthest to the top. If you have many opinions you’d be willing to try an adversarial collaboration on, consider posting the one where you disagree most with the SSC consensus, so that you have the most chance of finding a collaborator. If you get many responses, please talk to the people involved, choose one, and mention your choice clearly on the comment so other people don’t keep asking.

8. I’ll update everyone on the next Open Thread about the state of the competition, whether it’s actually going ahead, whether there are at least five teams, et cetera. I’ll also post the closing date by which all entries must be completed and submitted to me by email. Assume for now this will be around July 1, though I’m happy to shift that a little bit if there’s strong demand.

Adversarial Collaboration Contest: Loose Ends And Registration

Thanks to everyone who expressed interest in the adversarial collaborations contest.

There was a lot of good discussion in the last thread, with lots of people offering projects, but I'm not sure if people actually got in contact with each other and finalized their agreements.

So, if you proposed a collaboration in the last thread, please go back, take a look, and see if someone you might want to work with responded to your proposal.

I'm going to post two comments in the comment section of this post.

One is a coordination comment. If you're looking to find someone who you agreed to do a collaboration with in the last thread, so you can exchange emails with them, please post as a subcomment there. For example "I offered to do a collaboration on gun control, I see Bob839 agreed to partner with me, my email is whatever@place.com, please get in touch with me."

The second is a contest registration comment, so I know how many teams there are. If you and a partner have gotten in touch with one another and chosen a topic (you can always change it later), please post a subcomment there so I know that you're officially in. If for some reason you're not comfortable posting there, you can also email me at scott[at]shireroth[dot]org. Please mention your name, your partner's name, your topic, and (if you're comfortable giving it), your email.

A couple other people have offered to add prize money. I’ll get back to you once I’ve thought more about this, but I might use the extra money to offer a prize to the team that writes the best postmortem after the contest – about how they did the adversarial collaboration, whether they learned anything from it, and what recommendations they might have for other people trying the format.

Book Review: History Of The Fabian Society

I.

A spectre is haunting Europe. Several spectres, actually. One of them is the spectre of communism. The others are literal ghosts. They live in abandoned mansions. Sometimes they wail eerily or make floorboards creak. If you arrange things just right, you might be able to capture them on film.

Or at least this must have been the position of the founders of the Fabian Society, Britain’s most influential socialist organization. In 1883, ghost hunters Frank Podmore and Edward Pease spent the night at the same West London haunted house, looking for signs of the paranormal. As the night dragged on without any otherworldly visitations, they passed the time in conversation and realized they shared an interest in communist thought. The two agreed to meet up again later, and from these humble beginnings came one of the most important private societies in the history of the world.

Before the Fabians, communism was a pastime of wild-eyed labor activists promising bloody revolution. The Society helped introduce the idea of incremental democratic socialism – not just in the sense of Bernie Sanders, but in the sense of the entire modern welfare state. In the process, they pretty much invented the demographic of champagne-sipping socialist intellectuals. Famous Society members included George Bernard Shaw, Muhammed Ali Jinnah, Jawaharlal Nehru, and Tony Blair; Fabian ideas were imported wholesale into the economic policies guiding newly-independent India, Nigeria, Egypt, Syria, among others.

I became interested in the Fabians after reading Kerry Vaughn's excellent essay on the early neoliberal movement. I'm tempted to say "on the early neoliberal conspiracy", choosing that not because any of what they did was secret – it wasn't – but because it seems like the only term that describes their efficiency. A small group of people who wanted to change the world founded an organization, garnered influence in a bunch of little ways, thought strategically and acted with discipline. And after decades of work they got into positions of power and successfully changed the world, shifting the economic consensus from state socialism to free(er) markets.

And the Fabians seem like the same story, told in reverse. A small group of idealists, thinking strategically and acting with discipline, moved democratic socialism from the lunatic fringe to the halls of intellectual power. If aspiring generals study Alexander the Great and Napoleon, surely aspiring intellectual movements should study the neoliberals and the Fabians. Kerry's essay on neoliberalism was great, but I really wanted to know how the Fabians progressed from failed ghost hunters to puppetmasters controlling half of the twentieth century.

Edward Pease's *The History Of The Fabian Society* (Amazon, Project Gutenberg) seemed like a promising starting point, since it was written by one of the Society's founders and top officials. Pease turns out to be an engaging writer with a good sense of humor. His book, however, is a bit puzzling. It paints a Fabian Society which is chronically disorganized and which kind of hilariously bumbles into global power despite itself. I am not sure how much this reflects reality and how much this reflects Edward Pease's personal external locus of control and/or weird theory of history – but if you want a story about sinister socialist conspirators using their genius to upend the existing order, this isn't really it.

Still, it was informative, funny, and not totally absent of practical applications, so below I include some discussion and interesting passages.

II.

After the original ghost hunt, Pease and Podmore met again in a few other situations and eventually got some people together to found The Fellowship Of New Life, agreeing:

> That an association be formed whose ultimate aim shall be the reconstruction of Society in accordance with the highest moral possibilities

Later fleshed out as:

> OBJECT: The cultivation of a perfect character in each and all.
> PRINCIPLE: The subordination of material things to spiritual
> FELLOWSHIP: The sole and essential condition of fellowship shall be a single-minded, sincere, and strenuous devotion to the object and principle.

Under these auspices, they gathered a collection of upper-middle-class bureaucrats whose names sounded kind of like C.S. Lewis villains, like Hubert Bland and Percival Chubb, who agreed to meet monthly and discuss how to achieve their goals. Soon the political discussions started to crowd out the more philosophical ones, and so the politically-minded Fellows branched off to form their own society. Since they believed that Communists should avoid talk of violent revolution and instead bide their time working within the system, they named themselves the Fabian Society after Roman general Quintus Fabius Maximus, famous for his delaying tactics. According to one of their pamphlets:

> For the right moment you must wait, as Fabius did most patiently, when warring against Hannibal, though many censured his delays; but when the time comes you must strike hard, as Fabius did.

(Pease adds that “It has been pointed out by Mr. H.G. Wells, and by others before him, that Fabius never did strike hard”)

They gradually settled into a routine. Every fortnight, the Society would sponsor a lecture, sometimes by a member, sometimes by a guest, on some aspect of communism. Afterwards there would discussion and socializing (the type where you make friends, not the type where you nationalize industries). Pease notes that not only did this educate people, but the existence of a fortnightly activity kept people connected to the movement and the other members, and made them more likely to stick around and get engaged in other ways.

Sometimes the lectures would include or be replaced by debates among members on aspects of communist thought or policy. Once again, the apparent primary objective of learning ended up being overshadowed by a secondary benefit: all of the Fabians became excellent debaters. This became very important when the organization began fielding candidates for local government. According to George Bernard Shaw, “It was in that way that Bradlaugh, for instance, graduated from being a boy evangelist to being one of the most formidable debaters in the House of Commons. And the only opponents who have ever held their own against the Fabians in debate have been men like Mr. Levy or Mr. Foote, who learnt in the same school.”

When a lecture was particularly successful, the Fabians would turn it into a traveling tour. During this period British intellectual life was pretty centered in London, and when that city’s intellectual stars would go give lectures in Manchester or Liverpool, they could fill theaters. Pease describes one trip in 1890, where a half-dozen or so prominent Fabians went on a tour of North Britain:

> The lectures, except for a few days during the contest at Eccles, were extremely well reported, and even the ‘Manchester Guardian’ (the ‘Daily News’ of the manufacturing districts) came out with an approving leader. The audiences throughout the campaign steadily increased and followed the lectures with close and intelligent attention. In particular the members of Liberal working men’s clubs constantly declared that they had never heard ‘the thing put so straight’ before, and complained that the ordinary party lecturers were afraid or unwilling to speak out. Men who frankly confessed that they had hesitated before voting for the admission of our lecturers to their clubs were enthusiastic in welcoming our message as soon as they heard it. The vigorous propaganda in the manufacturing districts of the S.D.F. branches has been chiefly carried on by means of outdoor meetings. Its effect upon working-class opinion, especially among unskilled labourers, has been marked and important, but it has entirely failed to reach the working-men politicians who form the rank and file of the Liberal Associations and Clubs, or the ‘well-dressed’ Liberals who vaguely desire social reform, but have been encouraged by their leaders to avoid all exact thought on the subject.”

But by far the most important activity of the Fabians were their pamphlets. They got into this almost by accident. Pease says the first pamphlet was written by W.L. Phillips, whom he frankly describes as “a house-painter, the only genuine working man in our ranks”. The Fabians liked it and agreed to finance 2,000 copies. It discussed who the Fabians were, why they thought capitalism was bad, and how socialism was going to be better. Pease describes it as asserting that “we had with considerable courage set out to reconstruct society, and we frankly confessed that we did not know how to go about it.” The message was not a hit.

This was when the Society experienced its great stroke of luck. Three of the greatest writers and intellectuals of the era joined the movement: George Bernard Shaw in September 1884, Annie Besant in April 1885, and Sidney Webb that May.

Shaw was one of history’s most famous authors and playwrights. He was charismatic, shockingly prolific, and also had a talent for organizing things and getting people to agree with him.

Besant was described by Pease as “then notorious as an advocate of Atheism and Malthusianism, [and] the heroine of several famous law cases” – by which he was referring to attempts to censor her for writing about birth control, feminism, and other topics. She was well-known and beloved by the lower classes, and a fiery speaker who was good at riling up an auditorium.

Webb was a sort of 19th-century Ezra Klein, taking advantage of the first stirrings of statistics as a science to make himself an authoritative-sounding wonk. He was also an amazing writer, though more in the sense of viral popularizations of facts and explanations than in the Shaw sense of revolutionizing literature.

It wouldn’t quite be right to say the Society took advantage of these people’s talents; it might be more accurate to say nobody could stop them. Shaw, especially, sort of waltzed into the Society, took it over by force of personality and obviously being better than anyone else, and started churning out pamphlets as fast as everyone could print them. They, and four lesser lights who also helped with the pamphlet writing, became known as the seven Fabian Essayists. With their skills, the Fabian Pamphlets started catching on – to the point where, by 1891, a hundred thousand copies had been distributed and they more or less shaped the man in the street’s concept of socialism.

Webb’s work in particular helped give the Fabian Society an extra role – that of the modern think tank. Under his lead, quantitatively-minded Fabians queried various government records departments to try to collect and publicize numbers that would be useful in making the case for socialism – facts about how much of the wealth capitalists were keeping to themselves, how many poor people there were, what their conditions were like. Pease says that “The statistical facts were at that time practically unknown. They had to be dug out, one by one, from obscure and often unpublished sources” and so “at this period the Society had a virtual monopoly in the production of political pamphlets in which facts and figures were marshalled in support of propositions of reform in the direction of Socialism.”

By 1900, all of these projects were big successes, and the Society began to consider expansion. They encouraged the creation of branch societies in different towns and countries, and student societies in various universities. Pease was generally unimpressed with how these worked out. He felt experience had proven most of the people based outside London to be intellectually second-rate and without much to contribute. As for student societies, there was so much churn as students came and left after four years that he didn’t think anything useful ever happened there.

He was a little more optimistic about some other projects, including a lending library for socialist books, an office that gave legal advice to workers or other people having socialism-related legal disputes, and an effort to found a university studying items of interest to Fabians (which became the London School of Economics). And above all, he was impressed with the Fabians' project to intervene in local politics. Although the Society was too small to field national-level candidates, it did impressive work filling up school boards, minor bureaucracies, and sub-city-level local councils with Fabian socialists. This wasn't just an issue of campaigning. It was an issue of talent development. Some well-spoken well-educated person would go to the local Fabian meeting in North Whatevershire, and get noticed, and some veteran would approach him and say that North Whatevershire had an unusually reactionary or vulnerable council, and had he ever considered himself for the seat? And he would say of course not, what, me a politician, that's crazy. And the veteran would encourage him, and walk him through the process of registering for an election, and give him advice on the minimal amount of campaigning necessary at this level, and then a few months later he'd be a councillor. Again, the Fabians did this partly because they cared what happened on tiny villages' school boards, but also partly because the school board members of today are the people running for higher offices tomorrow, and if you can get someone's foot in the door and get them to start learning the ropes of politics, you're in a really good place.

There were a few other things. Occasionally they would hold conferences. Sometimes they would meet up with socialists from other countries or parties and try to sign accords. They had various worthy intellectual projects, including putting out lists of what books people should read to learn about socialism, or donating socialism-related books to the local library. But these were always a small part of their work. The core of the Fabian strategy was lectures, debates, talent cultivation, and – above all – really well-written pamphlets.

III.

Why was the Fabian Society so successful?

By "so successful", I mean that the Fabians' manifesto basically reads like a description of the average modern democracy's public policies. A set of Fabian demands called the True Radical Programme included an 8-hour workday, women's suffrage, paying MPs a salary, capital gains tax, public education, school lunches, and railroad nationalization. Keep in mind that they called it the True Radical Programme because it was supposed to be more radical than various other groups' Radical Programmes. They considered these totally insane fringe demands. And they got them all.

We can debate how important the Fabian Society was in this. Certainly they seem to have picked a winning horse, been on the right side of history, et cetera. But people at the time thought the Fabian Society was important. It became famous (or infamous, depending on your political leaning) and heavily associated with exactly these sorts of topics. If at least a small portion of their success can be attributed to them, how?

Pease makes it very clear he doesn't think it was because they were especially competent or hard-working. The Society had a hands-off supply to getting new members; if you wanted to be a member, you could apply, and get a sponsor, and eventually someone would probably vote on your membership – but there weren't recruitment drives or anything. They were equally clueless about money; whenever they had any, it was usually because it fell into their laps by coincidence, usually random rich people giving them donations. Pease, who doesn't seem to find money very interesting, never mentions whether the Society charges dues or not; if it did, they seem to have been minimal. In 1886, he lists the Society's total annual budget as £35, which I think corresponds to about £3500 today. Through much of the Society's history, its meetings took place in members' houses or in whatever random places people could get for cheap. There was constant discussion of "maybe if we made an effort to get more money we could promote socialism more effectively", but nobody seems to have ever done anything about it. It wasn't an ideological issue – nobody was refusing to participate in the degenerate capitalist system or anything. They just don't seem to have been very good at doing things. Someone would mention raising more money, they would form a committee to look into the issue, and the committee would get bored and wander off.

Even when the Fabians managed to throw something together, it always had a sort of doddering-eccentric-British-gentry air about it, like nobody was really that concerned with whether or not it accomplished anything. Pease describes the Fabians' first big convention, where experts were invited to lecture the delegates on the issues of the day: "The most successful paper was by a strange gentleman whom we had taken on trust as a Socialist, but who turned out to be an enthusiast on the subject of building more harbors."

The Society's members tended to be strange people, mercurial and hard to keep on task. Occasionally they would get distracted and forget about communism entirely, running off to worry about art or philosophy or ghost-hunting or something. One of the group's most influential leaders, Annie Besant, dropped out almost overnight to join a cult, move to India, study ESP, adopt a young child whom she claimed was the Messiah, and after many complicated twists of fate become President of the Indian National Congress. This kind of thing was always happening to the Fabians, and it really got in the way of the cause.

So why *was* the Fabian society so successful? Pease identifies several factors.

First, it was (arguably) the first Socialist organization in Britain. Pease thinks this is important not just because it got first mover advantage, but because it didn't have to justify itself to anyone:

> The Fabian Society has never had to seek acceptance by the rest of the Socialist movement. At any later date it would have been impossible for a relatively small middle-class society to obtain recognition as an acknowledged member of the Socialist confraternity. We were thus in a position to welcome the formation of working-class Socialist societies, but it is certain that in the early days they would never have welcomed us.

Second, it lucked into getting a couple of bona fide geniuses:

> The second and chief reason for the success of the Society was its good fortune in attaching to its service a group of young men, then altogether unknown, whose reputation has gradually spread, in two or three cases, all over the world, and who have always been in the main identified with Fabianism. Very rarely in the history of voluntary organisations has a group of such exceptional people come together almost accidentally and worked unitedly together for so many years for the furtherance of the principles in which they believed.

> Others have assisted according to their abilities and opportunities, but to the Fabian Essayists belongs the credit of creating the Fabian Society…It was this exceptional group of leaders, all intimate friends, all loyal to each other, and to the cause they were associated to advocate, and all far above the average in vigour and ability, that in a few years turned an obscure drawing-room society into a factor in national politics.

Third, the Fabians were bourgeois in every sense of the word. They knew how to move around in bourgeois society and get it to work for them. Pease talks about how (for him) one of the highlights of the Fabian Conference was how well-designed the stationery was:

> It also, by the way, showed off our pretty prospectus with the design by Crane at the top, our stylish-looking blood-red invitation cards, and the other little smartnesses on which we then prided ourselves. We used to be plentifully sneered at as fops and arm-chair Socialists for our attention to these details; but I think it was by no means the least of our merits that we always, as far as our means permitted, tried to make our printed documents as handsome as possible, and did our best to destroy the association between revolutionary literature and slovenly printing on paper that is nasty without being cheap. One effect of this was that we were supposed to be much richer than we really were, because we generally got better value and a finer show for our money than the other Socialist societies.

I used to be upset when charitable and activist organization would have really nice offices with lots of art on the wall, call in expensive catered lunches to their events, and hire a bunch of graphics design and PR people to make everything look perfect. Wasn’t this excessive? Shouldn’t they be spending their money and energy on the cause? Pease argues no. There were hordes of unwashed socialists standing on soapboxes raving about Revolution. The Fabians' comparative advantage was looking respectable. For a cause like socialism, where an important part of the battle is moving it into the Overton Window, handing out really well-designed stationery was important activism in and of itself.

Fourth, the Fabians accepted their role as the Socialist Society For Middle Class People as a valuable part of the overall socialist ecosystem. They refused to condemn themselves as exclusionary, or worry they were too bourgeois to have a right to speak, or feel guilty for not having better representation of poor people. They seemed to be working from a model where middle-class and working-class people had different mores and interests, and combining both of them into one society would just make everybody angry and unhappy. Pease explicitly understood that the working class societies had things they were better at and should focus on (like mass politics and riling up the base) and the Fabian Society had other things it was better at and should focus on (like think-tank-style white papers on the benefits of socialism). He also explicitly understood that an important part of having a socialist society was the social aspect – come for the politics, come back because you like the company or you're looking for someone to date or whatever. At this point there wasn't much social mixing of British classes, and there wasn't really a script for having a group that was half upper-middle-class and half workmen. Because the Fabians got so many of their advantages from being bourgeois, they wanted to protect that status at all costs.

> Regret has been sometimes expressed, chiefly by foreign observers, that the Society has maintained its

> separate identity. Why, it has been asked, did not the middle-class leaders of the Society devote their abilities directly to aiding the popular organisations, instead of "keeping themselves to themselves" like ultra-respectable suburbans?
>
> If this had been possible I am convinced that the loss would have exceeded the gain…the Fabians were not suited either by ability, temperament, or conditions to be leaders of a popular revolutionary party. Mrs. Besant with her gift of splendid oratory and her long experience of agitation was an exception, but her connection with the movement lasted no more than five years. Of the others Shaw did not and does not now possess that unquestioning faith in recognised principles which is the stock-in-trade of political leadership: and whilst Webb might have been a first-class minister at the head of a department, his abilities would have been wasted as a leader in a minority. But there was a more practical bar. The Fabians were mostly civil servants or clerks in private employ. The methods of agitation congenial to them were compatible with their occupations: those of the Social Democrats were not. Indeed in those days no question of amalgamation was ever mooted.

Fifth, the Fabians protected a sort of middle-class-liberal atmosphere of intellectual freedom and what Pease referred to as an inability to take themselves seriously – not in the sense of not being committed, but in the sense where they would laugh at anyone who seemed too pompous or too certain of anything. Pease seems to have thought that the lower classes' lack of a liberal education and likely religious upbringing made them susceptible to a dogmatic orthodox Marxism marked by witch-hunts to weed out revisionists. Most British intellectuals wouldn't have been willing to put up with such a climate, and so wouldn't have been able to get into socialism. The Fabian Society provided an alternative space where the sort of open debate that intellectuals and middle-class people take for granted was available and encouraged:

> The Social Democrats of those days asserted that unquestioning belief in every dogma attributed to Marx was essential to social salvation, and that its only way was revolution, by which they meant, not the complete transformation of society, but its transformation by means of rifles and barricades; they were convinced that a successful repetition of the Commune of Paris was the only method by which their policy could prevail. The Fabians realised from the first that no such revolution was likely to take place, and that constant talk about it was the worst possible way to commend Socialism to the British working class. And indeed a few years later it was necessary to establish a new working-class Socialist Society, the Independent Labour Party, in order to get clear both of the tradition of revolutionary violence and of the vain repetition of Marxian formulas. If the smaller society had merged itself in the popular movement, its criticism, necessary, as it proved to be, to the success of Socialism in England, would have been voted down, and its critics either silenced or expelled.

Also:

> The case for this project was based, strange to say, not on any history but on the Marxian analysis of the origin of the value of commodities, and no man who did not understand this analysis, or pretend to understand it, was fit to be called a “comrade.” The economic reasoning which “proved” this “law” was expressed in obscure and technical language peculiar to the propagandists of the movement, and every page of

Socialist writings was studded with the then strange words “proletariat” and “bourgeoisie.”

Lastly, the whole world, outside the socialist movement, was regarded as in a conspiracy of repression. Liberals (all capitalists), Tories (all landlords), the Churches (all hypocrites), the rich (all idlers), and the organised workers (all sycophants) were treated as if they fully understood and admitted the claims of the Socialists, and were determined for their own selfish ends to reject them at all costs. Although the Fabian propaganda had no doubt had some effect, especially amongst the working-class Radicals of London, and although some of the Socialist writers and speakers, such as William Morris, did not at all times present to the public the picture of Socialism just outlined, it will not be denied by anybody whose recollections reach back to this period that Socialism up to 1890 was generally regarded as insurrectionary, dogmatic, Utopian, and almost incomprehensible.

“Fabian Essays” presented the case for Socialism in plain language which everybody could understand. It based Socialism, not on the speculations of a German philosopher, but on the obvious evolution of society as we see it around us. It accepted economic science as taught by the accredited British professors; it built up the edifice of Socialism on the foundations of our existing political and social institutions: it proved that Socialism was but the next step in the development of society, rendered inevitable by the changes which followed from the industrial revolution of the eighteenth century. It is interesting after twenty-five years to re-read these essays and to observe how far the ideas that inspired them are still valid, and how far the prophecies made have been fulfilled.

Sixth, the Fabians tried to be a “big tent”. Aside from having room from people who weren’t orthodox Marxists, they generally avoided purges and purity tests. This wasn’t always easy for them. Just as they were reaching the big time, Britain got involved in the Boer War. Some of the Fabians were patriots who wanted to support the British cause; others thought the British government was evil and reflexively sided with its enemies, and a few even had various principles, some of which sound really weird to modern ears, for example this supposedly-socialist opinion:

> As for South Africa, “however ignorantly [our] politicians may argue about it, reviling one another from the one side as brigands, and defending themselves from the other with quibbles about waste-paper treaties and childish slanders against a brave enemy, the fact remains that a Great Power, consciously or unconsciously, must govern in the interests of civilisation as a whole; and it is not to those interests that such mighty forces as gold-fields, and the formidable armaments that can be built upon them, should be wielded irresponsibly by small communities of frontiersmen. Theoretically they should be internationalised, not British-Imperialised; but until the Federation of the World becomes an accomplished fact we must accept the most responsible Imperial federations available as a substitute for it”

All of this threatened to tear the Society apart, with various parties demanding that people be expelled for supporting the war or for opposing it. Against this, the veterans of the Society, including Pease, Webb, and Bernard Shaw, insisted upon the policy that anyone who mouthed agreement with the Principles of the Society – a short document containing only the broadest possible statements about socialism – was a member in good standing. When they finally got a resolution passed officially declaring that the Fabians had no position on the Boer War, Pease described it as “saving the Society”. The same happened during World War I and even various peacetime elections, usually with Labour supporters trying to make Labour membership a condition of Fabian membership. All such resolutions were quashed mercilessly by the Fabian leaders except one – the entire female membership of the Society threatened to walk *en masse* unless the Society enforced conformity with belief in women’s suffrage. The leadership grudgingly agreed.

Seventh, the Fabians’ namesake long-term outlook gave them an important role in cultivating talent, and some of the norms and assumptions of their organization seemed to work well for this. Pease says several times that “Socialists are born, not made”. He didn’t expect forceful action – recruitment campaigns, branch organizations, or the like – to have any effect. Instead, he favored a soft touch. Have the sort of intellectual atmosphere that talented people would be attracted to. Gradually draw them in with interesting social and intellectual activities. Once they’re attached, get them in on the first rung of some ladder or other – local politics, informal debate, small-time pamphlet writing. Have a few geniuses around who can recognize other geniuses. Then have positions to put people in once they’re worthy of it – whether it’s the lecture circuit, the propaganda business, or a university for them to teach at.

> The value of the plan for a propagandist society is largely this, that experience shows that people can only work together efficiently when they know each other. Therefore in practice political and many other organisations find it necessary to arrange garden parties, fêtes, picnics, teas, and functions of all sorts in order to bring together their numbers under such conditions as enable them to become personally

> acquainted with each other. In times of expansion the Fabian Society has held dinners and soirées in London, many of which have been successful and even brilliant occasions, because the new members come in crowds and the old attend as a duty. When new members are few these entertainments cease, for nothing is so dreary as a social function that is half failure, and a hint of it brings the series to an end. But a Summer School where members pass weeks together is far more valuable in enabling the leaders and officials to find out who there is who is good as a speaker or thinker, or who is a specialist on some subject of value to the movement. Moreover, gatherings of this class attract those on the fringe of the movement, and many of our members have come to us through attendance at the school. Apart from the direct interests of the Society, a School of this character is valued by many solitary people, solitary both socially, such as teachers and civil servants, who are often lonely in the world, and solitary intellectually because they live in remote places where people of their way of thinking are scarce.

It’s hard to say how many of these points were responsible for Fabian success, and how many were just random policies that didn’t change anything. But it doesn’t look like many of them were *secrets* of success. Most of these seem to have been widely adopted by the best modern nonprofits. I don’t know to what degree the Fabians originated these strategies, or whether they’re just natural optima that many successful organizations tend to converge on.

IV.

To finish, I do want to focus on the historical-inevitability angle. Whatever the Fabians’ other advantages, they arose at a really good time to be a socialist thinker. There was a sort of feeling in the air that socialism was the wave of the future, that there were literally no good arguments whatsoever against it, that you were either an intellectual (in which case it was obvious that socialism was better) or you were just so thoughtless that you had never even considered the matter at all (in which case you were motivated by things like prejudice or desire to hold on to your ill-gotten wealth). Pease describes the overwhelming consensus:

> Socialism succeeds because it is common sense. The anarchy of individual production is already an anachronism. The control of the community over itself extends every day. We demand order, method, regularity, design; the accidents of sickness and misfortune, of old age and bereavement, must be prevented if possible, and if not, mitigated. Of this principle the public is already convinced: it is merely a question of working out the details. But order and forethought is wanted for industry as well as for human life. Competition is bad, and in most respects private monopoly is worse. No one now seriously defends the system of rival traders with their crowds of commercial travellers: of rival tradesmen with their innumerable deliveries in each street; and yet no one advocates the capitalist alternative, the great trust, often concealed and insidious, which monopolises oil or tobacco or diamonds, and makes huge profits for a fortunate; few out of the helplessness of the unorganised consumers.

This is a really blunt way to put it. Do we want society to be total chaos? Or do we want to organize it and figure out how to make it run better? The latter? Okay, you’re a socialist.

The Fabians loved debating, so you would think at some point they would debate capitalists. But Pease seems far from convinced that capitalists had any arguments, or were even the sort of people who could debate. Although he is familiar with the great economists – Adam Smith, John Stuart Mill, David Ricardo – he thinks of them as proto-Socialists or at least neutral, and is fond of quoting an apparently pro-Communist passage from Mill's "Political Economy":

> If the choice were to be made between Communism with all its chances, and the present state of society with all its sufferings and injustices, if the institution of private property necessarily carried with it as a consequence that the produce of labour should be apportioned as we now see it almost in inverse proportion to labour, the largest portions to those who have never worked at all, the next largest to those whose work is almost nominal, and so in descending scale, the remuneration dwindling as the work grows harder and more disagreeable until the most fatiguing and exhausting bodily labour cannot count with certainty on being able to earn even the necessities of life; if this or Communism were the alternative, all the difficulties, great or small, of Communism would be but as dust in the balance.

I'm not sure whether Pease believed that a capitalist intellectual was a contradiction in terms, but he certainly didn't expect to meet any or think they had anything interesting to say. Indeed, the one time he does bring up some people having arguments against socialism, they sound bizarre and totally unlike anything a modern person might possibly say:

> When the Society was formed the Malthusian hypothesis held the field unchallenged and the stock argument against Socialism was that it would lead to universal

> misery by removing the beneficent checks on the growth of the population, imposed by starvation and disease upon the lowest stratum and society.

I don’t know if this was an echo chamber effect or if this was just how the late 19th century worked. I think the latter is at least possible. Remember, everyone (including the capitalists) expected communist countries to have stronger economies, even as late as the 1950s. The idea of coordination problems was almost unknown; the concept of prices as useful signals was still in its infancy. And the possibility that communism could lead to totalitarianism was almost inconceivable; for Pease these concepts are basically exact opposites, and it took Orwell to even jam the concept of “totalitarianism” in the public consciousness in a useful way. If you don’t have any of those concepts or ideas, how *do* you argue against socialism? I don’t know if anyone in Pease’s day had really solved that problem.

Not only were there no good arguments for capitalism, but the arguments for socialism were much more convincing. The standard communist rhetoric talks about capitalists paying for a factory, workers working in the factory, and capitalists getting most of the money despite putting in none of the work. This rings kind of hollow nowadays. Modern communists rail against Elon Musk – but everyone knows Elon Musk is brilliant and works 80+ hour weeks. You can be upset that lower-level SpaceX employees don’t get paid enough, but “why does Elon Musk deserve to get any money?” is a question with a bunch of really obvious answers. Venture capitalists are generally smart people who founded their venture capitalist firms and make inspired guesses as to where to direct resources. And it’s hard to create a broad coalition against stockholders from people who may themselves have some money in stocks for retirement. You can make arguments for why all these people deserve less than they get – but they don’t hit home the way they must have in Pease’s day.

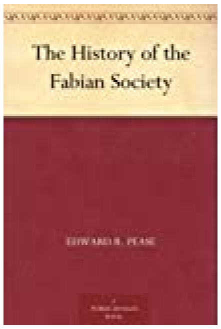
In Pease’s day, the Marxist model fit like a glove. Some businessman would go to some lord with an ancestral pension of thousands of pounds and ask for money to set up a factory. The lord (or more likely his steward) would sign off on it, the factory would be built, a little of the money would go to the businessman, but most of it would go back to the lord who was living on a country estate hunting foxes somewhere. Whatever you think of our current system, the system in Pease’s day wore its inequality on its sleeve in a way ours doesn’t. And the slogans of the Fabian Society and of its working-class comrades were things like “No money without work!” and “Put the rich to work!” – very compelling but less applicable to Elon Musk, Bill Gates, or most of the modern crop of capitalist figureheads. For Pease, “income inequality” would not really be a sensible term; it was kind of assumed that worrying about the differences between high-income people and low-income people was a distraction toward them uniting against their real enemy: people who didn’t work at all but had all the money anyway.

All of this came together into a feeling that socialism was so self-evident that arguing for capitalism was absurd. This led to a perspective where there was a battle between the right and rational way of organizing society (socialism) versus the entrenched forces who wanted to keep power but admitted they had no justification besides force and self-interest. Modern communism’s descent from its 19th century predecessors explains a lot about its mindset.

And so the Fabians, despite their nominal commitment to waiting, were supremely sure that victory was near:

> When the Liberal Party was crushed at the election of 1895 we thought that its end had come in England as it has in other countries. Conservatism is intelligible: Socialism we regarded as entirely reasonable. Between the two there seemed to be no logical resting place. We had discovered long ago that the working classes were not going to rush into Socialism, but they appeared to be and were in fact growing up to it. The Liberalism of the decade 1895-1905 had measures in its programme, such as Irish Home Rule, but it had no policy, and it seemed incredible then, as it seems astonishing now, that a party with so little to offer could sweep the country, as it was swept by the Liberals in 1906.

Tell me this doesn’t sound like a leftist journalist on Twitter talking about the Democrats while promising that the “progressive moment” is just around the corner. I notice the Fabian Society is still around and that their motto is “The future of the Left since 1884”. It sounds weird to me. Maybe a little too reminiscent of the claim that “fusion power is the future, and always will be”. The Fabians named themselves after a guy who was famous for his patience, but surely even they have to be getting a little bit tired.



The History of the
Fabian Society

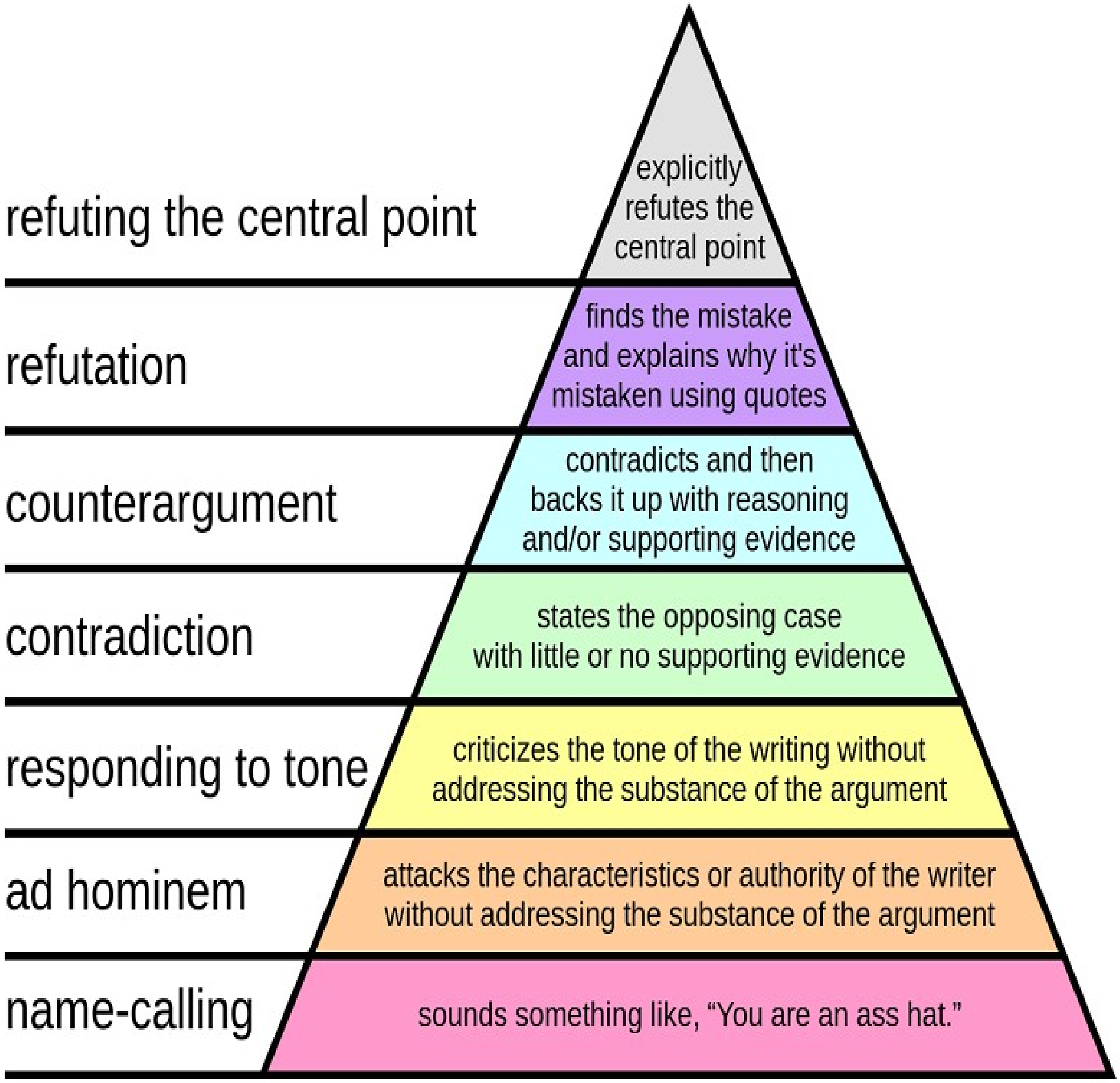
EDWARD R. PEASE

Classified Thread 5: Classified Never Sinned

This is the…monthly? bimonthly? occasional?…classified thread. Post advertisements, personals, and any interesting success stories from the last thread.

Varieties Of Argumentative Experience

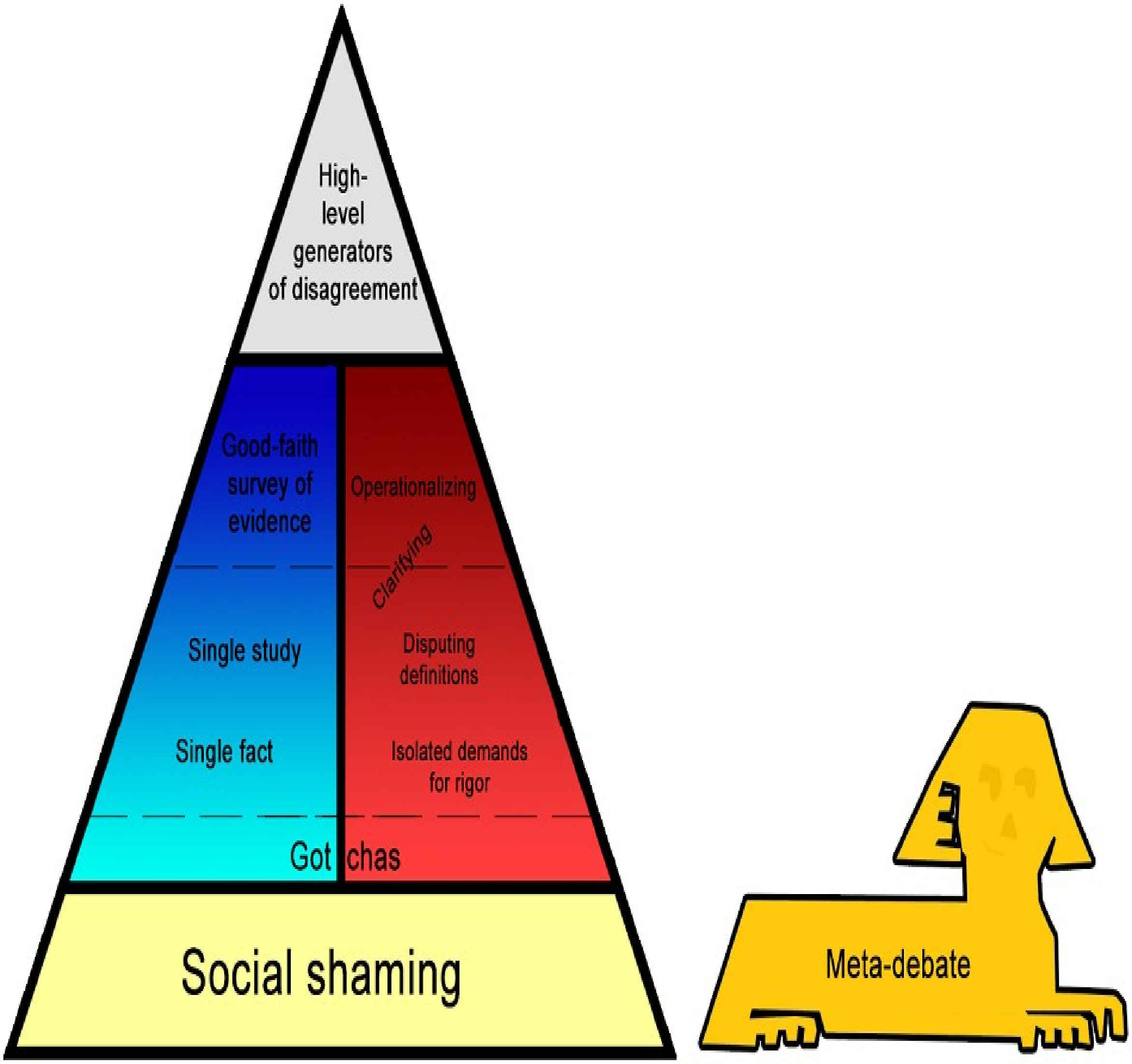
In 2008, Paul Graham wrote How To Disagree Better, ranking arguments on a scale from name-calling to explicitly refuting the other person's central point.


refuting the central point
explicitly refutes the central point
refutation
finds the mistake and explains why it's mistaken using quotes
counterargument
contradicts and then backs it up with reasoning and/or supporting evidence
contradiction
states the opposing case with little or no supporting evidence
responding to tone
criticizes the tone of the writing without addressing the substance of the argument
ad hominem
attacks the characteristics or authority of the writer without addressing the substance of the argument
name-calling
sounds something like, "You are an ass hat."

And that's why, ever since 2008, Internet arguments have generally been civil and productive.

Graham’s hierarchy is useful for its intended purpose, but it isn’t really a hierarchy of *disagreements*. It’s a hierarchy of types of response, within a disagreement. Sometimes things are refutations of other people’s points, but the points should never have been made at all, and refuting them doesn’t help. Sometimes it’s unclear how the argument even connects to the sorts of things that in principle could be proven or refuted.

If we were to classify disagreements themselves – talk about what people are doing when they’re even having an argument – I think it would look something like this:


High-level generators of disagreement
Good-faith survey of evidence
Operationalizing
Clarifying
Single study
Disputing definitions
Single fact
Isolated demands for rigor
Gotchas
Social shaming
Meta-debate

Most people are either meta-debating – debating whether some parties in the debate are violating norms – or they're just shaming, trying to push one side of the debate outside the bounds of respectability.

If you can get past that level, you end up discussing facts (blue column on the left) and/or philosophizing about how the argument has to fit together before one side is "right" or "wrong" (red column on the right). Either of these can be anywhere from throwing out a one-line claim and adding

“Checkmate, atheists” at the end of it, to cooperating with the other person to try to figure out exactly what considerations are relevant and which sources best resolve them.

If you can get past *that* level, you run into really high-level disagreements about overall moral systems, or which goods are more valuable than others, or what “freedom” means, or stuff like that. These are basically unresolvable with anything less than a lifetime of philosophical work, but they usually allow mutual understanding and respect.

I’m not saying everything fits into this model, or even that most things do. It’s just a way of thinking that I’ve found helpful. More detail on what I mean by each level:

Meta-debate is discussion of the debate itself rather than the ideas being debated. Is one side being hypocritical? Are some of the arguments involved offensive? Is someone being silenced? What biases motivate either side? Is someone ignorant? Is someone a “fanatic”? Are their beliefs a “religion”? Is someone defying a consensus? Who is the underdog? I’ve placed it in a sphinx outside the pyramid to emphasize that it’s not a *bad argument* for the thing, it’s just an argument about something completely different.

“Gun control proponents are just terrified of guns, and if they had more experience with them their fear would go away.”

“It was wrong for gun control opponents to prevent the CDC from researching gun statistics more thoroughly.”

“Senators who oppose gun control are in the pocket of the NRA.”

“It’s insensitive to start bringing up gun control hours after a mass shooting.”

Sometimes meta-debate can be good, productive, or necessary. For example, I think discussing “the origins of the Trump phenomenon” is interesting and important, and not just an attempt to bulverizing the question of whether Trump is a good president or not. And if you want to maintain discussion norms, sometimes you do have to have discussions about who’s violating them. I even think it can sometimes be helpful to argue about which side is the underdog.

But it’s not the debate, and also it’s much more fun than the debate. It’s an inherently social question, the sort of who’s-high-status and who’s-defecting-against-group-norms questions that we like a little too much. If people have to choose between this and some sort of boring scientific question about when fetuses gain brain function, they’ll choose this every time; given the chance, meta-debate will crowd out everything else.

The other reason it’s in the sphinx is because its proper function is to guard the debate. Sure, you *could* spend your time writing a long essay about why creationists’ objections to radiocarbon dating are wrong. But the meta-debate is what tells you creationists generally aren’t good debate partners and you shouldn’t get involved.

Social shaming also isn’t an argument. It’s a demand for listeners to place someone outside the boundary of people who deserve to be heard; to classify them as so repugnant that arguing with them is only dignifying them. If it works, supporting one side of an argument imposes so much reputational cost that only a few weirdos dare to do it, it sinks outside the Overton Window, and the other side wins by default.

“I can’t believe it’s 2018 and we’re still letting transphobes on this forum.”

“Just another purple-haired SJW snowflake who thinks all disagreement is oppression.”

“Really, do conservatives have any consistent beliefs other than hating black people and wanting the poor to starve?”

“I see we’ve got a Silicon Valley techbro STEMlord autist here.”

Nobody expects this to convince anyone. That’s why I don’t like the term “ad hominem”, which implies that shamers are idiots who are too stupid to realize that calling someone names doesn’t refute their point. That’s not the problem. People who use this strategy know exactly what they’re doing and are often quite successful. The goal is not to convince their opponents, or even to hurt their opponent’s feelings, but to demonstrate social norms to bystanders. If you condescendingly advise people that ad hominem isn’t logically valid, you’re missing the point.

> when you do sutuff like… shoot my jaw clean off of my face with a sniper rifle, it mostly reflects poorly on your self
>
> — wint (@dril) September 23, 2016

Sometimes the shaming works on a society-wide level. More often, it’s an attempt to claim a certain space, kind of like the intellectual equivalent of a gang sign. If the Jets can graffiti “FUCK THE SHARKS” on a certain bridge, but the Sharks can’t get away with graffiting “NO ACTUALLY FUCK THE JETS” on the same bridge, then almost by definition that bridge is in the Jets’ territory. This is part of the process that creates polarization and echo chambers. If you see an attempt at social shaming and feel triggered, that’s the second-best result from the perspective of the person who put it up. The best result is that you never went into that space at all. This isn't just about keeping conservatives out of socialist spaces. It's also about defining what kind of socialist the socialist space is for, and what kind of ideas good socialists are or aren't allowed to hold.

I think easily 90% of online discussion is of this form right now, including some long and carefully-written thinkpieces with lots of citations. The point isn't that it literally uses the word "fuck", the point is that the active ingredient isn't persuasiveness, it's the ability to make some people feel like they're suffering social costs for their opinion. Even really good arguments that are persuasive can be used this way if someone links them on Facebook with "This is why I keep saying Democrats are dumb" underneath it.

This is similar to meta-debate, except that meta-debate can sometimes be cooperative and productive – both Trump supporters and Trump opponents could in theory work together trying to figure out the origins of the "Trump phenomenon" – and that shaming is at least sort of an attempt to resolve the argument, in a sense.

Gotchas are short claims that purport to be devastating proof that one side can't possibly be right.

"If you like big government so much, why don't you move to Cuba?"

"Isn't it ironic that most pro-lifers are also against welfare and free health care? Guess they only care about babies until they're born."

"When guns are outlawed, only outlaws will have guns."

These are snappy but almost always stupid. People may not move to Cuba because they don't want government *that* big, because governments can be big in many ways some of which are bad, because governments can vary along dimensions other than how big they are, because countries can vary along dimensions other than what their governments are, or just because moving is hard and disruptive.

They may sometimes suggest what might, with a lot more work, be a good point. For example, the last one could be transformed into an argument like “Since it’s possible to get guns illegally with some effort, and criminals need guns to commit their crimes and are comfortable with breaking laws, it might only slightly decrease the number of guns available to criminals. And it might greatly decrease the number of guns available to law-abiding people hoping to defend themselves. So the cost of people not being able to defend themselves might be greater than the benefit of fewer criminals being able to commit crimes.” I don’t think I agree with this argument, and I might challenge assumptions like “criminals aren’t that much likely to have guns if they’re illegal” or “law-abiding gun owners using guns in self-defense is common and an important factor to include in our calculations”. But this would be a reasonable argument and not just a gotcha. The original is a gotcha exactly because it doesn’t invite this level of analysis or even seem aware that it’s possible. It’s not saying “calculate the value of these parameters, because I think they work out in a way where this is a pretty strong argument against controlling guns”. It’s saying “gotcha!”.

Single facts are when someone presents one fact, which admittedly does support their argument, as if it solves the debate in and of itself. It’s the same sort of situation as one of the better gotchas – it could be changed into a decent argument, with work. But presenting it as if it’s supposed to change someone’s mind in and of itself is naive and sort of an aggressive act.

“The UK has gun control, and the murder rate there is only a quarter of ours.”

“The USSR was communist and it was terrible.”

“Donald Trump is known to have cheated his employees and subcontractors.”

“Hillary Clinton handled her emails in a scandalously incompetent manner and tried to cover it up.”

These are all potentially good points, with at least two caveats. First, correlation isn’t causation – the UK’s low murder rates might not be caused by their gun control, and maybe not all communist countries inevitably end up like the USSR. Second, even things with some bad features are overall net good. Trump could be a dishonest businessman, but still have other good qualities. Hillary Clinton may be crap at email security, but skilled at other things. Even if these facts are true and causal, they only prove that a plan has at least one bad quality. At best they would be followed up by an argument for why this is really important.

I think the move from shaming to good argument is kind of a continuum. This level is around the middle. At some point, saying “I can’t believe you would support someone who could do that with her emails!” is just trying to bait Hillary supporters. And any Hillary supporter who thinks it’s really important to argue specifics of why the emails aren’t that bad, instead of focusing on the bigger picture, is taking the bait, or getting stuck in this mindset where they feel threatened if they admit there’s anything bad about Hillary, or just feeling too defensive.

Single studies are better than scattered facts since they at least prove some competent person looked into the issue formally.

"This paper from Gary Kleck shows that more guns actually cause *less* crime."

"These people looked at the evidence and *proved* that support for Trump is motivated by authoritarianism."

"I think you'll find economists have already investigated this and that the minimum wage doesn't cost jobs."

"There are actually studies proving that money doesn't influence politics."

We've already discussed this here before. Scientific studies are much less reliable guides to truth than most people think. On any controversial issue, there are usually many peer-reviewed studies supporting each side. Sometimes these studies are just wrong. Other times they investigate a much weaker subproblem but get billed as solving the larger problem.

There are dozens of studies proving the minimum wage does destroy jobs, and dozens of studies proving it doesn't. Probably it depends a lot on the particular job, the size of the minimum wage, how the economy is doing otherwise, etc, etc, etc. Gary Kleck does have a lot of studies showing that more guns decrease crime, but a lot of other criminologists disagree with him. Both sides will have plausible-sounding reasons for why the other's studies have been conclusively debunked on account of all sorts of bias and confounders, but you will actually have to look through those reasons and see if they're right.

Usually the scientific consensus on subjects like these will be as good as you can get, but don't trust that you know the scientific consensus unless you have read actual well-conducted surveys of scientists in the field. Your echo chamber telling you “the scientific consensus agrees with us” is definitely not sufficient.

A **good-faith survey of evidence** is what you get when you take all of the above into account, stop trying to devastate the other person with a mountain of facts that can’t possibly be wrong, and start looking at the studies and arguments on both sides and figuring out what kind of complex picture they paint.

“Of the meta-analyses on the minimum wage, three seem to suggest it doesn’t cost jobs, and two seem to suggest it does. Looking at the potential confounders in each, I trust the ones saying it doesn’t cost jobs more.”

“The latest surveys say more than 97% of climate scientists think the earth is warming, so even though I’ve looked at your arguments for why it might not be, I think we have to go with the consensus on this one.”

“The justice system seems racially biased at the sentencing stage, but not at the arrest or verdict stages.”

“It looks like this level of gun control would cause 500 fewer murders a year, but also prevent 50 law-abiding gun owners from defending themselves. Overall I think that would be worth it.”

Isolated demands for rigor are attempts to demand that an opposing argument be held to such strict invented-on-the-spot standards that nothing (including common-sense statements everyone agrees with) could possibly clear the bar.

“You can’t be an atheist if you can’t prove God doesn’t exist.”

“Since you benefit from capitalism and all the wealth it’s made available to you, it’s hypocritical for you to oppose it.”

“Capital punishment is just state-sanctioned murder.”

“When people still criticize Trump even though the economy is doing so well, it proves they never cared about prosperity and are just blindly loyal to their party.”

The first is wrong because you can disbelieve in Bigfoot without being able to prove Bigfoot doesn’t exist – “you can never doubt something unless you can prove it doesn’t exist” is a fake rule we never apply to anything else. The second is wrong because you can be against racism even if you are a white person who presumably benefits from it; “you can never oppose something that benefits you” is a fake rule we never apply to anything else. The third is wrong because eg prison is just state-sanctioned kidnapping; “it is exactly as wrong for the state to do something as for a random criminal to do it” is a fake rule we never apply to anything else. The fourth is wrong because Republicans have also been against leaders who presided over good economies and presumably thought this was a reasonable thing to do; “it’s impossible to honestly oppose someone even when there’s a good economy” is a fake rule we never apply to anything else.

Sometimes these can be more complicated and ambiguous. One could argue that
“Banning abortion is unconscionable because it denies someone the right to do what they want with their own body” is an isolated demand for rigor, given that we ban people from selling their organs, accepting unlicensed medical treatments, using illegal drugs, engaging in prostitution, accepting euthanasia, and countless other things that involve telling them what to do with their bodies – “everyone has a right to do what they want with their own

bodies” is a fake rule we never apply to anything else. Other people might want to search for ways that the abortion case is different, or explore what we mean by “right to their own body” more deeply. Proposed without these deeper analysis, I don’t think the claim would rise much above this level.

I don’t think these are necessarily badly-intentioned. We don’t have a good explicit understanding of what high-level principles we use, and tend to make them up on the spot to fit object-level cases. But here they act to derail the argument into a stupid debate over whether it’s okay to even discuss the issue without having 100% perfect impossible rigor. The solution is exactly the sort of “proving too much” arguments in the last paragraph. Then you can agree to use normal standards of rigor for the argument and move on to your real disagreements.

These are related to fully general counterarguments like “sorry, you can’t solve every problem with X”, though usually these are more meta-debate than debate.

Sometimes isolated demands for rigor can be rescued by making them much more complicated; for example, I can see somebody explaining why kidnapping becomes acceptable when the state does it but murder doesn’t – but you’ve got to actually make the argument, and don’t be surprised if other people don’t find it convincing. Other times these work not as rules but as heuristics – for example “let people do what they want with their body in the absence of very compelling arguments otherwise” – and if those heuristics survive someone else challenging whether banning unlicensed medical treatment is *really* that much more compelling than banning abortion, they usually end up as high-level generators of disagreement (see below).

Disputing definitions is when an argument hinges on the meaning of words, or whether something counts as a member of a category or not.

“Transgender is a mental illness.”

“The Soviet Union wasn’t really communist.”

“Wanting English as the official language is racist.”

“Abortion is murder.”

“Nobody in the US is really poor, by global standards.”

It might be important on a social basis what we call these things; for example, the social perception of transgender might shift based on whether it was commonly thought of as a mental illness or not. But if a specific argument between two people starts hinging on one of these questions, chances are something has gone wrong; neither factual nor moral questions should depend on a dispute over the way we use words. This Guide To Words is a long and comprehensive resource about these situations and how to get past them into whatever the real disagreement is.

Clarifying is when people try to figure out exactly what their opponent’s position is.

“So communists think there shouldn’t be private ownership of factories, but there might still be private ownership of things like houses and furniture?”

“Are you opposed to laws saying that convicted felons can’t get guns? What about laws saying that there has to be a waiting period?”

“Do you think there can ever be such a thing as a just war?”

This can sometimes be hostile and counterproductive. I’ve seen too many arguments degenerate into some form of “So you’re saying that rape is good and we should have more of it, are you?” No. Nobody is ever saying that. If someone thinks the other side is saying that, they’ve stopped doing honest clarification and gotten more into the performative shaming side.

But there *are* a lot of misunderstandings about people’s positions. Some of this is because the space of things people can believe is very wide and it’s hard to understand exactly what someone is saying. More of it is because partisan echo chambers can deliberately spread misrepresentations or cliched versions of an opponent’s arguments in order to make them look stupid, and it takes some time to realize that real opponents don’t always match the stereotype. And sometimes it’s because people don’t always have their positions down in detail themselves (eg communists’ uncertainty about what exactly a communist state would look like). At its best, clarification can help the other person notice holes in their own opinions and reveal leaps in logic that might legitimately deserve to be questioned.

Operationalizing is where both parties understand they’re in a cooperative effort to fix exactly what they’re arguing about, where the goalposts are, and what all of their terms mean.

“When I say the Soviet Union was communist, I mean that the state controlled basically all of the economy. Do you agree that’s what we’re debating here?”

“I mean that a gun buyback program similar to the one in Australia would probably lead to less gun crime in the United States and hundreds of lives saved per year.”

“If the US were to raise the national minimum wage to $15, the average poor person would be better off.”

“I’m not interested in debating whether the IPCC estimates of global warming might be too high, I’m interested in whether the real estimate is still bad enough that millions of people could die.”

An argument is operationalized when every part of it has either been reduced to a factual question with a real answer (even if we don’t know what it is), or when it’s obvious exactly what kind of non-factual disagreement is going on (for example, a difference in moral systems, or a difference in intuitions about what’s important).

The Center for Applied Rationality promotes double-cruxing, a specific technique that helps people operationalize arguments. A double-crux is a single subquestion where both sides admit that if they were wrong about the subquestion, they would change their mind. For example, if Alice (gun control opponent) would support gun control if she knew it lowered crime, and Bob (gun control supporter) would oppose gun control if he knew it would make crime worse – then the only thing they have to talk about is crime. They can ignore whether guns are important for resisting tyranny. They can ignore the role of mass shootings. They can ignore whether the NRA spokesman made an offensive comment one time. They just have to focus on crime – and that’s the sort of thing which at least in principle is tractable to studies and statistics and scientific consensus.

Not every argument will have double-cruxes. Alice might still oppose gun control if it only lowered crime a little, but also vastly increased the risk of the government becoming authoritarian. A lot of things – like a decision to vote for Hillary instead of Trump – might be based on a hundred little considerations rather than a single debatable point.

But at the very least, you might be able to find a bunch of more limited cruxes. For example, a Trump supporter might admit he would probably vote Hillary if he learned that Trump was more likely to start a war than Hillary was. This isn't quite as likely to end the whole disagreement in a fell swoop – but it still gives a more fruitful avenue for debate than the usual fact-scattering.

High-level generators of disagreement are what remains when everyone understands exactly what's being argued, and agrees on what all the evidence says, but have vague and hard-to-define reasons for disagreeing anyway. In retrospect, these are probably why the disagreement arose in the first place, with a lot of the more specific points being downstream of them and kind of made-up justifications. These are almost impossible to resolve even in principle.

"I feel like a populace that owns guns is free and has some level of control over its own destiny, but that if they take away our guns we're pretty much just subjects and have to hope the government treats us well."

"Yes, there are some arguments for why this war might be just, and how it might liberate people who are suffering terribly. But I feel like we always hear this kind of thing and it never pans out. And every time we declare war, that reinforces a culture where things can be solved by force. I think we need to take an unconditional stance against aggressive war, always and forever."

"Even though I can't tell you how this regulation would go wrong, in past experience a lot of well-intentioned regulations have ended up backfiring horribly. I just think we should have a bias against solving all problems by regulating them."

"Capital punishment might decrease crime, but I draw the line at intentionally killing people. I don't want to live in a society that does that, no matter what its reasons."

Some of these involve what social signal an action might send; for example, even a just war might have the subtle effect of legitimizing war in people’s minds. Others involve cases where we expect our information to be biased or our analysis to be inaccurate; for example, if past regulations that seemed good have gone wrong, we might expect the next one to go wrong even if we can’t think of arguments against it. Others involve differences in very vague and long-term predictions, like whether it’s reasonable to worry about the government descending into tyranny or anarchy. Others involve fundamentally different moral systems, like if it’s okay to kill someone for a greater good. And the most frustrating involve chaotic and uncomputable situations that have to be solved by *metis* or *phronesis* or similar-sounding Greek words, where different people’s Greek words give them different opinions.

You can always try debating these points further. But these sorts of high-level generators are usually formed from hundreds of different cases and can’t easily be simplified or disproven. Maybe the best you can do is share the situations that led to you having the generators you do. Sometimes good art can help.

The high-level generators of disagreement can sound a lot like really bad and stupid arguments from previous levels. “We just have fundamentally different values” can sound a lot like “You’re just an evil person”. “I’ve got a heuristic here based on a lot of other cases I’ve seen” can sound a lot like “I prefer anecdotal evidence to facts”. And “I don’t think we can trust explicit reasoning in an area as fraught as this” can sound a lot like “I hate logic and am going to do whatever my biases say”. If there’s a difference, I think it comes from having gone through all the previous steps – having confirmed that the other person knows as much as you might be intellectual equals who are both equally concerned about doing the moral thing – and realizing that both of you alike are controlled by high-level generators. High-level generators aren't biases in the sense of mistakes. They're the strategies everyone uses to guide themselves in uncertain situations.

This doesn't mean everyone is equally right and okay. You've reached this level when you agree that the situation is complicated enough that a reasonable person with reasonable high-level generators could disagree with you. If 100% of the evidence supports your side, and there's no reasonable way that any set of sane heuristics or caveats could make someone disagree, then (unless you're missing something) your opponent might just be an idiot.

Some thoughts on the overall arrangement:

1. If anybody in an argument is operating on a low level, the entire argument is now on that low level. First, because people will feel compelled to refute the low-level point before continuing. Second, because we're only human, and if someone tries to shame/gotcha you, the natural response is to try to shame/gotcha them back.

2. The blue column on the left is factual disagreements; the red column on the right is philosophical disagreements. The highest level you'll be able to get to is the *lowest* of where you are on the two columns.

3. Higher levels require more vulnerability. If you admit that the data are mixed but seem to slightly favor your side, and your opponent says that every good study ever has always favored his side plus also you are a racist communist – well, you kind of walked into that one. In particular, exploring high-level generators of disagreement requires a lot of trust, since someone who is at all hostile can easily frame this as

“See! He admits that he’s biased and just going off his intuitions!”

4. If you hold the conversation in private, you’re almost guaranteed to avoid everything below the lower dotted line. Everything below that is a show put on for spectators.

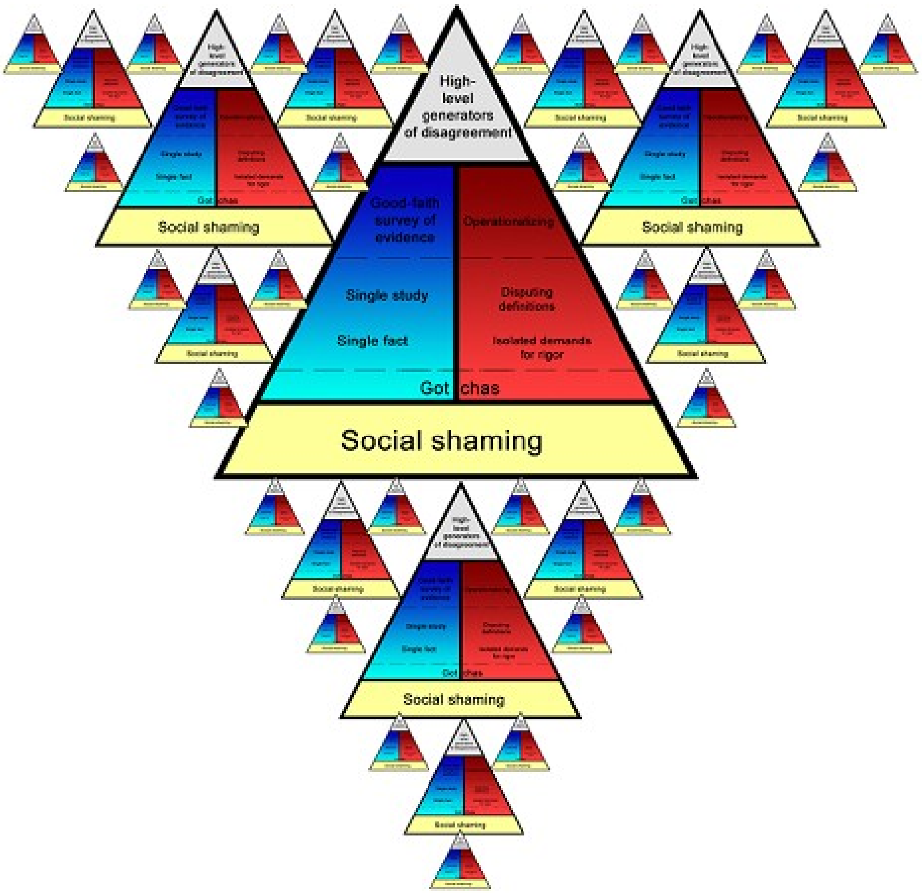
5. If you’re intelligent, decent, and philosophically sophisticated, you can avoid everything below the higher dotted line. Everything below that is either a show or some form of mistake; everything above it is impossible to avoid no matter how great you are.

6. The shorter and more public the medium, the more pressure there is to stick to the lower levels. Twitter is great for shaming, but it’s almost impossible to have a good-faith survey of evidence there, or use it to operationalize a tricky definitional question.

7. Sometimes the high-level generators of disagreement are other, even more complicated questions. For example, a lot of people’s views come from their religion. Now you’ve got a whole different debate.

8. And a lot of the facts you have to agree on in a survey of the evidence are also complicated. I once saw a communism vs. capitalism argument degenerate into a discussion of whether government works better than private industry, then whether NASA was better than SpaceX, then whether some particular NASA rocket engine design was better than a corresponding SpaceX design. I never did learn if they figured whose rocket engine was better, or whether that helped them solve the communism vs. capitalism question. But it seems pretty clear that the degeneration into subquestions and discovery of superquestions can go on forever. This is the stage a lot of discussions get bogged

down in, and one reason why pruning techniques like double-cruxes are so important.


High-level generators of disagreement
Good-faith survey of evidence
Operationalizing
Single study
Disputing definitions
Single fact
Isolated demands for rigor
Gotchas
Social shaming

9. Try to classify arguments you see in the wild on this system, and you find that some fit and others don't. But the main thing you find is *how few real arguments there are*. This is something I tried to hammer in during the last election, when people were complaining "Well, we tried to debate Trump supporters, they didn't change their mind, guess reason and democracy don't work". Arguments above the first dotted line are rare; arguments above the second basically nonexistent in public unless you look really hard.

But what's the point? If you're just going to end up at the high-level generators of disagreement, why do all the work?

First, because if you do it right you'll end up respecting the other person. Going through all the motions might not produce agreement, but it should produce the feeling that the other person came to their belief honestly, isn't just stupid and evil, and can be reasoned with on other subjects. The natural tendency is to assume that people on the other side just don't know (or deliberately avoid knowing) the facts, or are using weird perverse rules of reasoning to ensure they get the conclusions they want. Go through the whole process, and you will find some ignorance, and you will find some bias, but they'll probably be on both sides, and the exact way they work might surprise you.

Second, because – and this is total conjecture – this deals a tiny bit of damage to the high-level generators of disagreement. I think of these as Bayesian priors; you've looked at a hundred cases, all of them have been X, so when you see something that looks like not-X, you can assume you're wrong – see the example above where the libertarian admits there is no clear argument against this particular regulation, but is wary enough of regulations to suspect there's something they're missing. But in this kind of math, the prior shifts the perception of the evidence, but *the evidence also shifts the perception of the prior*.

Imagine that, throughout your life, you've learned that UFO stories are fakes and hoaxes. Some friend of yours sees a UFO, and you assume (based on your priors) that it's probably fake. They try to convince you. They show you the spot in their backyard where it landed and singed the grass. They show you the mysterious metal object they took as a souvenir. It seems plausible, but you still have too much of a prior on UFOs being fake, and so you assume they made it up.

Now imagine another friend has the same experience, and also shows you good evidence. And you hear about someone the next town over who says the same thing. After ten or twenty of these, maybe you start wondering if there's something to all of this UFOs. Your overall skepticism of UFOs has made you dismiss each particular story, but each story has also dealt a little damage to your overall skepticism.

I think the high-level generators might work the same way. The libertarian says "Everything I've learned thus far makes me think government regulations fail." You demonstrate what looks like a successful government regulation. The libertarian doubts, but also becomes slightly more receptive to the possibility of those regulations occasionally being useful. Do this a hundred times, and they might be more willing to accept regulations in general.

As the old saying goes, "First they ignore you, then they laugh at you, then they fight you, then they fight you half-heartedly, then they're neutral, then they grudgingly say you might have a point even though you're annoying, then they say on balance you're mostly right although you ignore some of the most important facets of the issue, then you win."

I notice SSC commenter John Nerst is talking about a science of disagreement and has set up a subreddit for discussing it. I only learned about it after mostly finishing this post, so I haven't looked into it as much as I should, but it might make good followup reading.

Basic Income, Not Basic Jobs: Against Hijacking Utopia

Some Democrats angling for the 2020 presidential nomination have a big idea: a basic jobs guarantee, where the government promises a job to anybody who wants one. Cory Booker, Kirsten Gillibrand, Elizabeth Warren, and Bernie Sanders are all said to be considering the plan.

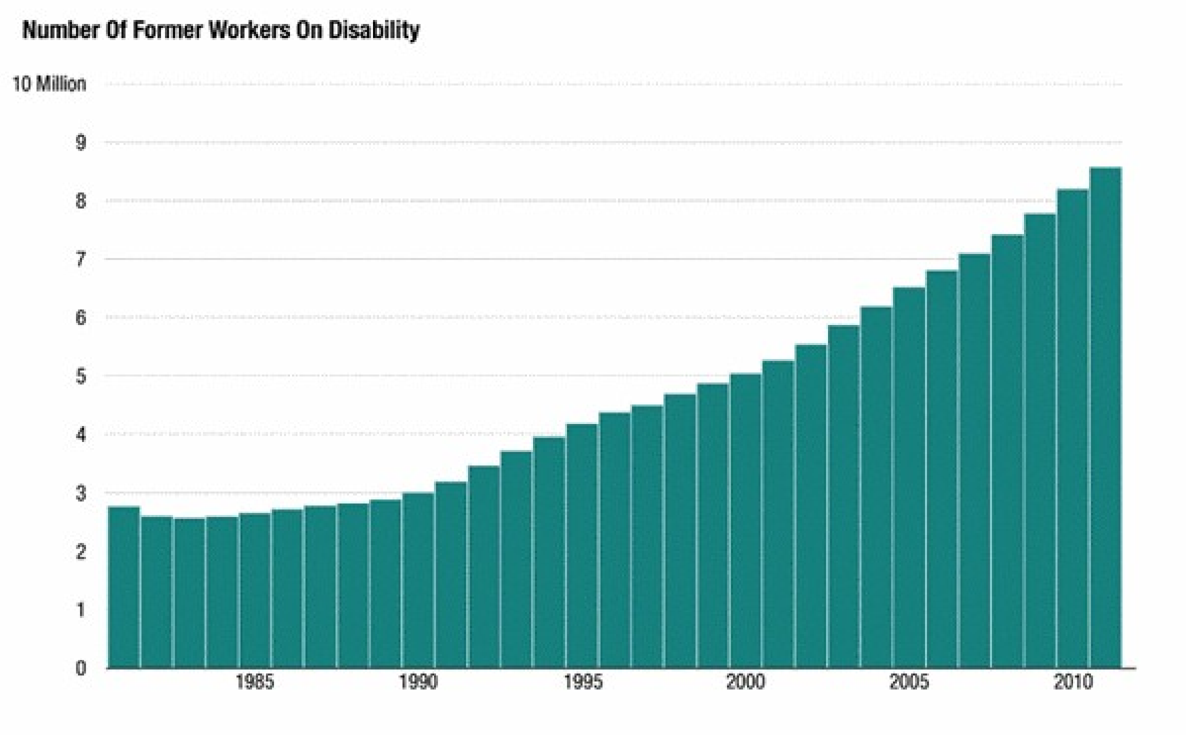
I've pushed for a basic income guarantee before, and basic job guarantees sure sound similar. Some thinkers have even compared the two plans, pointing out various advantages of basic jobs: it feels "fairer" to make people work for their money, maybe there's a psychological boost from being productive, you can use the labor to do useful projects. Simon Sarris has a long and excellent article on "why basic jobs might fare better than UBI [universal basic income]", saying that:

> UBI's blanket-of-money approach optimizes for a certain kind of poverty, but it may create more in the long run. Basic Jobs introduce work and opportunity for communities, which may be a better welfare optimization strategy, and we could do it while keeping a targeted approach to aiding the poorest.

I am totally against this. Maybe basic jobs are better than nothing, but I have an absolute 100% revulsion at the idea of implementing basic jobs as an alternative to basic income. Before getting into the revulsion itself, I want to bring up some more practical objections:

1. Basic jobs don't help the disabled

Disability has doubled over the past twenty years and continues to increase.



Experts disagree on how much of the rise in disability reflects deteriorating national health vs. people finding a way to opt out of an increasingly dysfunctional labor market, but everyone expects the the trend to continue. Any program aimed at the non-working poor which focuses on the traditionally unemployed but ignores the disabled is only dealing with the tip of the iceberg.

The current disability system has at least three major problems which I would expect basic income to solve.

First, the disability application process is a mess. Imagine the worst DMV appointment you've ever had to obtain the registration to a sketchy old car you got from a friend, then multiply it by a thousand – then imagine you have to do it all while being too disabled to work. Even clear-cut applications can take months to go through, inflicting an immense burden on people who don't know where their money is coming from during that time. And people with harder-to-prove conditions like mental illness and chronic pain might require multiple appeals – dragging the process out for years – or never get it at all. The disabled people I have talked to generally hate everything about this.

Second, disability is becoming a catch-all for people who can't find employment. This is a useful function that needs to be served. But right now, it involves unemployed people faking and exaggerating disability. This rewards liars and punishes the honest. If society labels the system "FOR DISABLED PEOPLE ONLY", basic fairness – to the disabled, to taxpayers, and to honest workers who aren't gaming the system – require them to gatekeep entry. Right now they spend lots of time and money on gatekeeping and still mostly fail. But any attempt to crack down would exacerbate the first problem, the one where real disabled people have to spend months or years in a Kafka novel before getting recognized.

Third, because of the first and second problems disabled people feel like they constantly have to prove themselves. Sometimes they'll have good days – lots of conditions are relapsing-remitting – and they'll want to go play in the park or something. Then they have to worry that some neighbor is going to think "well, that guy looks pretty healthy", take a photo, and they'll end up as one of those stories with headlines like SO-CALLED DISABLED PERSON CAUGHT PLAYING SPORTS IN PARK. Other times it's a bureaucratic issue. I had a patient who, after a few years on disability, recovered enough that he thought he could work about ten hours a week. When he tried to make it happen, he learned he would lose his disability payments – apparently if you can work at all the government doesn't believe you're really disabled – and ten hours a week wasn’t enough to support himself. So he cancelled the new job and didn’t work at all.

As long as you have a system whose goal is to separate the “truly” disabled people from the fakers, you’re going to run into problems like these. But refuse to gatekeep, and you have an unjust system where anyone who wants to lie can get out of work while their more honest coworkers are left slaving away all day. Basic income cuts the Gordian knot by proposing that everyone is legally entitled to support, whether they’re disabled or not. Disabled people can get their money without gatekeeping, and there’s no reward for foul play.

Basic jobs abandons this solution and takes us right back to the current system. If you’re abled enough to perform a government job, you’ve got to do it. Who decides if you’re abled enough? The Kafkaesque gatekeepers. And so we get the same bureaucratic despair, the same attempts to cheat the system, and the same perverse incentives.

And the number of disability claims keeps rising. Remember, a lot of economists think that the flight away from work and toward disability comes from people voting with their feet against exactly the kind of low-paying unpleasant jobs that basic jobs advocates want to offer everybody. Expect them to vote against those too, with no clear solutions within the basic jobs paradigm.

2. Basic jobs don’t help caretakers

And another 10% to 15% of the jobless are people caring for their sick family members.

This is unavoidable and currently uncompensated. The AgingCare Caregiver forum says their “number one question” is whether people who need to take time off work to care for a sick or elderly parent can get money. The only answer they can provide is “if the person you’re caring for has money or insurance, maybe they can pay you”. If they don’t, you’re out of luck. [EDIT: apparently some states do offer some money for this].

Right now our society just drops the ball on this problem. I don’t blame it; giving people money to care for family members would be prohibitively expensive. It would also require a gatekeeping bureaucracy that would put the disability gatekeeping bureaucracy to shame. Not only do they have to assess if someone’s really unable to subsist without care, they also have to decide who gets to take the option for which relatives. I have a second cousin some number of times removed who’s very disabled; can I quit my job and get paid a reasonable salary to take care of him? What if I tell you I’ve never met him or even talked to him on the phone, and just have my grandmother’s word for it that he exists and is sick? What exactly counts as caretaking? If I go visit my second cousin once a day for an hour to make sure he hasn’t gotten any sicker than usual, should the government pay me a full salary? What if actually doing that is 100% vital to my second cousin’s continued survival and I wouldn’t be able to do it consistently while holding down a job? You are never going to be able to make a bureaucracy that can address all these issues fairly.

Basic income cuts the knot again, giving everyone enough money that they can take care of sick or aging friends or relatives if they so choose. You don’t have to justify your choice to provide this level of care (but not that level) to the government. You can just do what needs to be done.

Basic jobs once again drops the ball on this problem. If your mother is dying, you can’t be there to help her, because the government is going to make you dig ditches and fill them in again all day to satisfy people’s worry that somebody somewhere might be getting money without doing enough make-work to “deserve” it.

3. Basic jobs don’t help parents

Everything above, except this time you’re a single parent (or a double parent whose spouse also works) and you want to take care of your child. If you could afford daycare, you probably wouldn’t be the sort of person who needs to apply for a guaranteed basic job. What do you do?

I know what the basic jobs people’s solution to this is going to be: free daycare for all! Okay. So in addition to proposing the most expensive government program ever invented, you want to supplement it by passing the second most expensive government program ever invented, at the same time? Good luck.

But even aside from this, I want us to step back and think about what we’re doing. I have met people – mostly mothers, but some fathers too – who are heartbroken at the thought of missing the best years of their children’s lives grinding away at a 9 to 5 job, stuck in traffic commuting to their job, or being too tired to spend time with them after they get home from their job. These people miss their kids’ first steps, outsource watching their first words to underpaid daycare employees, and have to choose between attending their kids’ school plays and putting food on the table.

And if we check the Treasury and decide that we, as a society, don’t have enough money to solve this problem – then whatever, we don’t have enough money to solve this problem.

But I worry we’re going to check and find we have more than enough money. But somebody is going to be so excited about making poor people do busy-work to justify their existence, that we're going to insist on perpetuating the problem anyway. And if that forces us to pay for universal free daycare, we're *going to be spending extra money just to make sure we can perpetuate the problem as effectively as possible*. We're going to be saying "We could give basic income for $800 billion, or basic jobs plus universal daycare for $900 billion. And that extra $100 billion? That's the money we spend to make sure you're digging ditches and filling them in all day, instead of getting to be at home spending time with your kids."

4. Jobs are actually a big cause of poverty

Poor people's two largest expenses are housing and transportation.

Guaranteed jobs have to be somewhere. Most of them will be in big cities, because that's where everybody is. The ones in the country will be few and far between.

That means to get to your government-mandated job, you'll either need to live in the big city or have a car. Living in the big city means tripling your monthly rent. Having a car means car payments, insurance payments, repair payments, gas payments, and incidentals.

When I first started working with poor patients, I was shocked how many of the problems in their lives were car-related. For well-off people like me, having a car is background noise; you buy or lease it for a reasonable price, then never worry about it again. Poor people can't afford to buy and don't always have good enough credit to lease. They tend to get older, sketchier cars that constantly break down. A constant complaint I heard: "My car broke, I can't afford repairs, and I'm going to get fired if I can't make it to my job". Some of them can't afford insurance and take their chances without it. Others have had various incidents with the police that cost them their license, but they can’t just not show up to work, so they drive anyway and hope they don’t get arrested.

Then there are the little things. Your work doesn’t have a break room, so you’ve got to eat out for lunch, and there goes a big part of your food budget. Your work demands a whole new set of business clothes, so there’s double your clothing budget. You can’t attend things during normal business hours, so you have to pay extra for out-of-hours services.

And then there’s all of the problems above. You can’t take care of your children anymore, so you’ve got to pay for daycare or a nanny or an Uber to take them to their grandparents’ house. You can’t take care of your sick parents anymore, so you’ve got to pay for a home health aide to come in and look after them. You get job-related strain or stress, and there’s the cost of a doctor’s appointment.

And then there are the fuzzier things. If you’ve just spent the entire day at work, and you’re really exhausted, and you never get any time to yourself, maybe you don’t have the energy left to drive to the cheaper supermarket on the other end of town. Maybe you don’t have the time to search for the absolute best deal on the new computer you’re getting. Maybe you don’t have the willpower to resist splurging and giving yourself one nice thing in your life of wage slavery. All of this sounds kind of shameful, but they’re all things that my patients have told me and *things that I do myself sometimes despite my perfectly nice well-paying job*.

5. Basic jobs may not pay for themselves by doing useful work

I once read an economist discussing why unemployment exists at all. That is, there are always people who would like to have someone clean their house, take care of their children, or come to their house and cook them food. And there are always businesses that would like their floors a little cleaner, or their customers served a little faster, or one more security guard to keep everything safe. Surely they would pay *some* amount of money to get these jobs done? And surely *some* homeless person would rather take that small amount than starve on the streets? So why are there still unemployed people?

One answer must be the minimum wage, but how come this happens even in times and places where minimum wages are absent or easy to evade?

The economist suggested that not all employees are net positive. Employees can steal from you, offend your customers, or be generally weird and smelly and ruin the atmosphere. They can be late or not show up at all – and if you made plans depending on their presence, that can be worse than your never hiring them in the first place. A bad nanny can traumatize your kid. A bad maid can break your priceless vase. A litigious employee can take you to court on false charges. Somebody who’s loud and curses at you and constantly smells of marijuana can just make you a little more stressed and unhappy all the time.

So if you have a job that only produces 1 utility, but a bad employee in that job will cost you 10 utility, and there’s a 10% chance any employee you get will be bad – then you’re not going to fill that job no matter how low a salary people are willing to work for.

How bad can employees get? Please read these AskReddit links. They’re slightly off-topic, but they’re going to give you information you can’t get any other way:

— AskReddit: Bosses of Reddit, what was your worst employee like?
— Managers of Reddit, who was your worst employee?
— What is the worst employee you have had to put up with?
— Who’s the worst coworker you ever had?

It’s safe to say they can get pretty bad.

I know many unemployed people who are amazing virtuous hard-working folks. But I also know the unemployed guy who lives in a cardboard box by the BART station, is surrounded by a protective shell of discarded beer cans, and shouts “GRAAAAGH” at passers-by for inscrutable reasons. And the amazing virtuous hard-working folks have a decent shot at getting a job in the private sector eventually, but the guy who shouts “GRAAAAGH” never will. Your population of basic-job-needers is going to be disproportionately composed of people who don’t fit into the regular workforce. How do you think that will turn out?

I worry some people think choosing basic jobs over basic income means free labor. Like, if you were going to pay someone a basic income of $10K/year, but the market value of their labor is $8K/year, you could employ them running a soup kitchen, get that $8K of value, and then you’re really only “losing” $2K/year.

I am less sanguine. If you pay people $10K/year, you’re only losing $10K/year. If you employ them to run a soup kitchen, and the soup kitchen has to keep closing because of hygiene violations, or gets hit with a sexual harassment lawsuit because someone groped a customer, or burns down because someone left the stove on, or loses all its customers because the manager shouts “GRAAAAGH” at everybody who asks for soup – then you’re losing more.

6. Private industry deals with bad workers by firing them; nobody has a good plan for how basic jobs would replace this

Suppose someone *does* accidentally leave a stove on and burn down the soup kitchen. You transfer them to an agricultural commune and they crash the tractor into a tree. You transfer them to some kind of low-risk paper-pushing job, but they're late to work every day and skip it entirely once or twice a week, and important papers end up tragically un-pushed. After a while, you decide they are too incompetent to add non-negative value to any of the programs on offer. What do you do with them?

If you fire them, then you're not a basic jobs guarantee. You're a basic-jobs-for-skilled-workers-whom-bosses-like guarantee. We already have one of those – it's called capitalism, maybe you've heard of it. But a real solution to poverty would have to encompass everybody, not just people who are good at working within the system.

And if you don't fire them, what's your plan? Accept a certain level of burning-things-down, customer complaints, coworker complaints, and unexcused absences? Let them make everybody around them miserable? Turn your soup kitchen into some kind of federal disaster area because you're absolutely committed to letting every single human being in the United States work there?

Or transfer them to a job in a padded room putting blocks in stacks and knocking them down again, in a way that inconveniences nobody because nobody cares about it? Abandon all pretense at creating anything other than busy-work for poor people out of an all-consuming desire to make sure nobody can live comfortably unless they have spent forty hours of every week in boredom and misery?

Or offer *these* people a basic income, and let all your other employees hate you for giving incompetent people leisure time at home with their family while the hard workers dig ditches all day?

This isn't speculation about some vague future. These questions get played out all around the country in our existing "government must take everyone no matter how little they want to be there" institution, ie public school. Here's a quote from a reader the last time we discussed the public school system.

> I was friends with a guy who briefly worked as a teacher at a public high school in central DC (I'm 80% sure it was Cardozo High). He had an education background thanks to spending several years working as a youth camp counselor and as an after-school program counselor, and that was sufficient to qualify him for DCPS' abbreviated teacher training program (such a thing existed in 2009 when he did it; I'm unsure if it is still around). During the training program, I remember him speaking about his enthusiasm for the teaching skills he was learning and about his eagerness to put them to use (in retrospect, I think some of this was a nervous attempt to convince himself the job wouldn't be bad). After a break of several months, we spoke again, and he was almost totally disillusioned with the job and was already thinking of quitting. This is what I remember him saying:
>
> 1) On the first day of classes, there was no orientation for new teachers, no brief meeting where the Principal shook his hand and said "Welcome Aboard," nothing. He had to go to the front office and ask a secretary what classroom was his and walk there by himself.
>
> 2) Unexcused absences were chronic and undermined his ability to teach anything. At the start of each of his

classes, he had a written roster of students, and he had to check off which students were there. For any class, typically 20-30% of students would be missing, without explanation (This is a very important point to remember whenever anyone tries to blame DCPS’ poor outcomes on large class sizes–on paper, each class might have 35 students, but typically, only 23 are actually showing up). Additionally, the 20-30% of students who were absent each class varied from day-to-day, meaning one student didn’t know what was taught on Monday, the one next to him was there Monday but not Tuesday, the third was there the first two days but not Wednesday, etc.

3) Student misbehavior was atrocious. For example, out of the students who showed up to class, it was common for some to walk into the classroom late, again without any explanation and often behaving disruptively. As a rule, whenever a student did that, he was obligated to sign his name on a clipboard for the teacher’s attendance records (there was no punishment for tardiness–late students merely had to write their names down). Some late students would chronically resist doing this, either ignoring him and just going to their desks or yelling curses at him. My friend described an incident where one student–who was physically bigger than he was–yelled out he was a “FAGGOT” when asked to sign the clipboard, provoking laughs from all the other students, before sitting down without signing it. After seeing he could get away with that, the student started calling my friend “FAGGOT” all the time. Other examples of misbehavior included near-constant talking among the students during lessons and fooling around with cell phones.

4) Teachers received almost no support from the school administration. Had sane rules been followed at this high

school, students would have been immediately sent to the office for formal punishment for these sorts of offenses I’ve described. However, under such a policy, the office would have been overwhelmed with misbehaving students and probably some of their enraged parents, so the administration solved the problem by forbidding teachers from sending students to the office for anything other than physical violence in the classroom. My friend had no ability to formally punish the student who liked to call him “FAGGOT” other than to use stern verbal warnings.

5) Most of the students were unwilling and in some cases unable to learn. During class sessions, the students were clearly disengaged from what he was teaching. Homework completion rates were abysmal. As the end of the academic semester neared, he saw that a huge fraction of them were on track to fail, so he resorted to pitiful cajoling, pizza parties, reward schemes, and deals involving large curves to everyone’s grades if they could only, for once do a little work, and it didn’t work. Some of his students were Latino and understood little or even no English, meaning they learned (almost) nothing, even when they tried. He resorted to seating the students who knew no English next to bilingual Latinos who could translate for them. That was the best he could do. In fairness, he spoke glowingly of some of his students, who actually put in some effort and were surprisingly smart […]

I’ll never forget how crestfallen and stressed out he was when he described these things to me. Having never taught in American public schools, I didn’t realize just how bad it was, and the detailed nature of his anecdotes really had an impact on me. I advised him to finish his year at the high school and then to transfer to ANY non-

> urban school in the area, even if it meant lower pay or a longer commute. We lost touch after that, but I can’t imagine he still works in DCPS.

The education system remains popular because they can always hold up glossy posters of smiling upper-class children at Rich Oaks Magnet High School and claim the system works. But basic jobs are going to be selecting primarily from the very poor demographic and they’re going to get hit with the same problem as the poorest public schools – a need for people to behave, combined with inability to credibly disincentivize misbehavior.

Basic income avoids this problem. It provides money to everyone, good employees and bad employees alike, without forcing any workplace to keep people it finds unproductive or threatening, and without having to find humiliating make-work jobs for anybody.

7. Private employees deal with bad workplaces by quitting them; nobody has a good plan for how basic jobs would replace this

And if you think this is a problem for the managers, just wait until you see what the employees have to put up with.

Some bosses are incompetent. Some are greedy. Some are downright abusive. Some don’t have any obvious flaw you can put your finger on, they just turn every single day into a miserable emotional grind. Sometimes the boss is fine, but the coworkers are creeps, or bullies, or don’t do their fair share. Sometimes the boss and the coworkers are both okay, but the job itself just isn’t suited to your personality and what you can manage.

In private industry, people cope by leaving their job and finding a better one. It’s not a perfect system. A lot of people are stuck in jobs they don’t like because they’re not sure they can find another, or because they don’t have enough money to last them through the interim. And this is one reason why poor people who can’t easily change jobs have worse working conditions than wealthier people who can. But everyone at least has the option in principle if their job becomes unbearable.

What about the people who can’t get any jobs besides the guaranteed basic ones? How do they deal with abusive working conditions?

Probably somebody will set up some system to let you quit one basic job and go to a different one in the same city. But probably it will end up being much more complicated than that. How do you deal with the guy who quits every job after a week or two, looking for the perfect cushy position? How do you deal with the case where there’s only one basic job available within a hundred miles? How do you deal with the case where everyone wants the same few really good jobs, and nobody wants to work at the awful abusive soup kitchen down the road?

People will set up systems to solve these problems, and the systems will be unwieldy and ineffective, just like the systems for switching public schools today, and just like all the other clever top-down socialist systems people invent to replace exit rights. Probably they’ll take the edge off some of these problems, but probably nobody will be truly satisfied with the results.

Basic income solves this problem. It doesn’t make anybody stay at a workplace they don’t like.

8. Basic income could fix private industry; basic jobs could destroy it

In my dreams, the government finds a way to provide a basic income at somewhere above subsistence level. The next day, every single person working an awful McJob quits, because there's no reason to work there except not being able to subsist otherwise.

After that, one of two things happens. First, maybe McDonald's makes a desperate effort to invent awesome robots that can serve food without human support. Society and Ronald McDonald share a drink together – McDonald's has managed to remain a profitable company providing a valuable service, and poor people live comfortable lives without having to flip burgers eight hours a day.

Or maybe inventing robots is hard, and McDonald's has to lure some people back. They raise pay and improve working conditions, until the prospect of working for McDonald's and getting luxuries is better than the prospect of living off basic income and getting subsistence. Maybe McDonald's has to raise prices; maybe they even have to close some stores. But again, something like McDonald's continues to exist and workers are relatively well-off.

A poorly-planned basic jobs guarantee could make the problem worse. Suppose that the government decided to use its free labor to farm cows. This puts various private cow-farming companies out of business; after all, the government can pay its employees out of the welfare budget, but private companies have to pay employees out of revenue. Some of the unemployed cow-farmers go get a guaranteed basic job, putting further private companies out of work. And other unemployed cow-farmers go work at McDonald's, driving up the supply of McDonald's employees and so ensuring lower wages and worse conditions.

This isn't to deny that a well-planned basic jobs guarantee could have the same effect as basic income; if the government jobs were better than McDonald’s’s, McDonald’s might have to raise wages and improve conditions to lure people back. The direction of the effect would depend on how good the government jobs are and how much they compete with private industry. I predict the government jobs will be very bad, and compete with private industry a lot, which makes me expect the effect will be negative.

9. Basic income supports personal development; basic jobs prevent it

I have a friend who was stuck on a dead-end career path. His job paid a decent amount, he just didn’t really like where it was going. So he saved up enough money to live on for a year, spent a year teaching himself coding, applied to a programming job, got it, and felt a lot more comfortable with his financial situation.

And I had a patient in a similar situation. Hated her job, really wanted to leave it, didn’t have enough skills to get anything else. So she went to night school, and – she found she couldn’t do it. After working 8 to 6 every day, her ability to go straight from a long day’s work to a long night’s studying just wasn’t in the cards. And her income didn’t give her the same opportunity to save up some money and take a year off. So she gave up and she still works at the job she hates. The end.

Basic income would give everyone who wants to work the same opportunity as my friend – the ability to take a year off, cultivate yourself, learn stuff, go to school, build your resume – without it being a financial disaster.

Basic jobs would leave everyone in the same position as my patient – forced to work 40+ hours a week, commute however many hours a week, good luck finding time to earn yourself a ticket out of that lifestyle while still staying sane.

There are more creative things you can do with time off work. Entrepreneurs like to talk about “runway” – how long can you keep burning through money before you run out and have to declare your new business a failure? Sometimes your runway is costs like renting an office or paying employees, but for small one-person businesses the question is usually “how long can I continue to live and feed myself working on this not-yet-profitable company?”

And poor people have runway issues of their own. One of the most common reasons poor people end up in crappy jobs is because they don’t have the luxury of a long job search. If your savings will only last you a month before you can’t make rent, you’re going to accept the first job that will take you and feel grateful for it. If you have a guaranteed income source, you can wait until somebody presents you with a better fit.

Basic income is unlimited runway. Entrepreneurs can feel free to try out crazy ideas without the constant pressure of losing their shirt; people in between jobs can feel free to spend time looking for options they can tolerate.

Basic jobs solves none of these problems, and maintains the time pressures that prevent people from exploring interesting ideas or realizing their full potential.

10. Basic income puts everyone on the same side; basic jobs preserve the poor-vs-the-rest-of-us dichotomy

Welfare users often talk about the stigma involved in getting welfare. Either other people make them feel like a parasite, or they just worry about it themselves. Basic jobs would be little different. There will be the well-off people with jobs producing useful goods and services. And there will be the people on guaranteed basic jobs, who know their paychecks are being subsidized by Society. In the worst case scenario, people complaining about workplace abuses at their guaranteed basic job will be told how lucky they are to have work at all.

Basic income breaks through that dichotomy. Everybody, from Warren Buffett to the lowliest beggar on the street, gets the same basic income. We assume Warren Buffett pays enough taxes that the program is a net negative for him, but taxes are complicated and this is hard to notice. Rich people are well aware they contribute more to the system than they get out. But they don't think of it on the level of "I pay $340 in taxes to support my local police station, but only get $154.50 of police services. Meanwhile, Joe over there pays $80 in police taxes and gets $190 in police services. I hate him so much!"

There will be people on basic income who have no other source of money. There will be people who supplement it with odd jobs now and then. There will be people who work part-time but who plausibly still get more than they pay in taxes. There will be people who work full-time and maybe pay more than they get but aren't really sure. At no point does a clear dichotomy between "those people getting welfare" and "the rest of us who support them" ever kick in.

11. Work sucks

Amidst all of these very specific complaints, I worry we're losing site of the bigger picture, which is that work sucks. I have my dream job, the job I've been lusting after since I was ten years old, it's going exactly as well as I expected – but I still Thank God It's Friday just like everyone else.

And other people have it almost arbitrarily worse. Here are some of the cases you hear about several times a week doing psychiatry:

“I work really long days at my job. I have to deal with angry clients, bosses who don’t appreciate me, and coworkers who try to dump their work on me. By the time I get home after my hour-long commute, I’m too wiped to do anything other than make a microwave dinner and watch TV for an hour or two until I pass out. Then on the weekends I take care of business like grocery shopping, cleaning, and paying my bills. Then Monday comes around and I have to do it all over again. I feel like work drains all my energy and doesn’t leave me any time to be me. I used to play in a band, and we had dreams of making it big, but I had to quit because I don’t feel like I have time for it any more. It’s just work, go home, sleep, repeat.”

“I can’t stand the new open office plan. I feel like I’ve got to do work in the middle of a loud bar where everyone’s trying to talk over each other. Sometimes I hide in the janitorial closet just so I can concentrate for a couple of hours while I finish sometimes important. I’m afraid if anyone ever catches me doing that they’ll say I’m ‘not a team player’ and I’ll get written up, but I just can’t take being crammed together with all those people. Maybe if you gave me some Adderall I could focus better?”

“Sorry I haven’t seen you in a few months. My workplace says it gives time off for doctor’s appointments, but you still get in trouble for missing targets, and I just couldn’t find any time that works. I ran out of my medication a month ago and am having constant panic attacks, so if you could refill that right away it would be nice. And sorry, I need to go now, I’m actually calling you from the bathroom. I wanted to call you from the janitorial closet, but when I went in, there was a woman inside who mumbled something about the open office plan and accused me of distracting her.”

And the people with the *worst* jobs don’t have good enough time or money to see psychiatrists; I just never meet them. But I understand it gets pretty bad:

> Amazon employee here. The post [The Undercover Author Who Discovered Amazon Warehouse Workers Were Peeing In Bottles Tells Us The Culture Was Like A Prison] is pretty spot on. They don’t monitor bathroom breaks, but your individual rate (or production goal) doesn’t account for bathroom breaks. Or let’s say there is a problem like you need two of something and there’s only one left, well you have to put on your “andon”, wait for someone to come “fix” for you, all the while your rate is dropping. The two most common reasons pepole get fired are not hitting rate, and attendance. They don’t really try to help you hit rate, they just fire and replace.
>
> My first week there two pepole collapsed from dehydration. It’s so common place to see someone collapse that nobody is even shocked anymore. You’ll just hear a manager complain that he has to do some report now, while a couple of new pepole try to help the guy (veterans won’t risk helping becuse it drips rate). No sitting allowed, and there’s nowhere to sit anywhere except the break rooms. Before the robots (they call them kivas) pickers would regularly walk 10-15 miles a day, now it’s just stand for 10-12 hours a day.
>
> People complain about the heat all the time but we just get told 80 degrees (Fahrenheit obviously) is a safe working temp. Sometimes they will pull out a thermometer, but even when it hits 85 they just say it’s fine.
>
> There’s been deaths, at least one in my building… Amazon likes to keep it all hush hush. Heard about

> others, you can find the stories if you search for it, but Amazon does a good job burying it.
>
> Every now and we have an inspection, where stuff like this should be caught and changed. But they just pretty it up. If the people doing the inspection looked at numbers on inspection day vs normal operation, they would see a massive difference… but no fucks given.
>
> The truth is the warehouses operate at a loss most the time, Amazon literally can’t afford to pay the workers decent pay, and can’t afford to not work them to death. The entire business model is dependent on cheap (easily replacable) labor, which is why tier 1s are the bulk of the Amazon work force. My building has like 3-5k workers most the time and around 10-30k on the holiday (what they call peak). Almost all of that is tier 1, most states have 4-7 of these warehouses, and some like Texas and Arizona have tons more.
>
> Next time you order something off Amazon, remember it was put in that box buy a guy sweating his ass off trying to put 100-250 things in a box per hour, for 10 hours a day or he will be fired, making about a dollar more than minimum wage. Might have even been a night shift guy, who goes to work at 630pm and gets off at 5am.

I 100% understand that advocates of basic jobs insist that they’ll be better than that, that they guarantee really good jobs in clean sunny offices where everybody has a smile in their face and is well-paid. I also understand they said the same thing about those DC public schools before throwing huge amounts of money at them. Forget promises; I care about incentives.

Either one of basic jobs or basic income could be potentially the costliest project the US government has ever attempted.

Government projects usually end up cash-constrained, and the costliest one ever won't be the exception. The pressure to cut corners will get overwhelming. It's hard to cut corners on basic income – either citizens get their checks or they don't. It's simple to cut corners on basic jobs. You do it the same way Amazon does – you let working conditions degrade to intolerable levels. What are your workers going to do do? *Quit?* Neither Amazon nor government-guaranteed basic jobs need to worry about that – both know that their employees have no good alternatives.

Gathering a bunch of disempowered poor people in a place they're not allowed to opt out of, with budget constraints on the whole enterprise, is basically the perfect recipe for ensuring miserable conditions. I refuse to believe that they will be much better than private industry; the best we can hope for is that they end up no worse. But the conditions in private industry are miserable, even for people with better resources and coping opportunities than basic jobs recipients are likely to have.

I grudgingly forgive capitalism the misery it causes, because it's the engine that lifts countries out of poverty. It's a precondition for a free and prosperous society; attempts to overthrow it have so consistently led to poverty, tyranny, or genocide that we no longer believe its proponents' earnest oaths that *this time* they've got it right. For right now, there's no good alternative.

But if we have a basic jobs guarantee, it will cause all the same misery, and I won't forgive it. The flimsy justifications we can think up won't be up to the task of justifying the vast suffering it will cause. We can't excuse it as necessary to produce the goods and services we rely on. We can't excuse it as a necessary condition for political freedom. If a worker asks "why?", our only answer will be "because Cory Booker thought a basic jobs guarantee would play better among the electorate than basic income, now get back to packing boxes and collapsing from dehydration”. There *will* be an alternative: a basic income guarantee. We will have rejected it.

I feel like as a quasi-libertarian, I sometimes downplay how awful private industry, capitalism, and the modern workplace are. If so, I apologize. The only possible excuse for defending such a flood of misery is what inevitably happens when people meddle with it. But the price of such morally tenuous greater-good style reasoning is that you need to stay hyper-aware of times when you *don’t* need to defend the system, when there *is* a chance to do better without destroying everything. I think basic income is such a chance. And I think basic jobs are a tiny modification to the idea, which destroys its potential and perpetuates all the worst parts of the existing system.

It would be unfair to make this argument without responding to jobs’ proponents points, so I want to explain why I don’t think they provide a strong enough argument against. These will be from the Sarris piece. I don’t want to knock it too much, because it’s a really fair and well-written piece that presents the case for jobs about as well as it can be presented, and any snark I might give it below is totally undeserved and due to personal viciousness. But it argues:

i) Studies of UBI haven’t been very good, so we can’t know if it works.

> Studying a UBI pilot with an end date is not studying UBI at all: It is instead studying a misnamed temporary cash payment. By the nature of pilots, the cohort’s behavior cannot reliably change to depend on UBI’s long term existence. No study yet has guaranteed a cohort money forever, and even if it did it would be difficult for a pilot

> to study the long term effects, some of which may be generations out. What pilot can tell us what its like for kids to grow up with parents who have never worked?[…]
>
> Basic Job programs are more amenable to piloting and a gradual roll-out, since new clusters of jobs appear (and end) all the time. Piloting Basic Jobs can be tried in different communities with varying magnitudes. The legislation to justify such a pilot may already be in place[1], and even a pilot may have lasting benefits. What we learn from the pilot will be more applicable than studying temporary cash transfers in a community and expecting that knowledge to translate into society-wide UBI. If a pilot is successful, one can imagine a kind of National Civil Service, organized like existing federal programs such as the National Park Service, which can hire professionals to train and supervise projects.

I have some minor caveats – Alaska has had a (very small) universal basic income for some time, which seems to have worked relatively well. And basic job studies will also have trouble scaling; smaller trials might preferentially select the most functional unemployed people, would have less impact on private industry, and can always just dismiss people back to the general pool of the unemployed. But overall I agree with the point that basic income is a bigger change and we should be more suspicious of bigger changes.

But at some point you’re arguing against testing something because it’s untested. If we can’t 100% believe the results of small studies – and I agree that we can’t – our two options are to give up and never do anything that hasn’t already been done, or to occasionally take the leap towards larger studies. I think basic income is promising enough that we need to pursue the second. Sarris has already suggested he won’t trust anything that’s less than permanent and widespread, so let’s do an experiment that’s permanent and widespread.

ii) UBI gives everyone the same amount, but some people need more (for example, diabetics need more money to pay for insulin). Existing social programs like medical aid take this into account; UBI wouldn’t.

This seems like exactly the problem that insurance exists to solve. Bringing insurance into the picture, “everybody has to get this” switches from a negative to a positive.

I won’t speculate on how this will look, except to note that it would work well with some kind of mandate where the cost of a Medicare-like state insurance gets auto-deducted from your UBI. Since I’m quasi-libertarian, I would support people’s right to opt out of this, after signing and notarizing a bunch of forms with “I UNDERSTAND I AM AN IDIOT AND MIGHT DIE” on them in big red letters, but I understand other people might prefer to avoid the chance of moral hazard. It still seems like this problem is solvable.

iii) Somehow even if everyone has more money they won’t be better off

> One of the biggest assumptions people make with UBI is that the problems of today and the near future are primarily ones of money. I don’t think the data supports this. [link to various charts showing that people generally have food and access to health care]

On some level, if you’re tempted to believe this you should find a poor person and ask them how they feel about being poor. I predict they will say it is bad. They will not agree that our society has basically solved all of its money-related problems. They will say there is a very real sense in which their money-related problems remain unsolved. I guarantee you they will have very strong feelings about this.

But that’s overly pat. A steelman of Sarris’ point might go something like this: it definitely seems true that there is some complicated way in which a family of eight living in a tiny farmhouse in the Kansas prairie in 1870 was happy and felt financially secure even though they probably only earned a few hundred dollars a year by today’s measures. So isn’t it weird that people earning twenty thousand dollars a year still think of material goods as their barrier to happiness?

I think explaining that effectively would require a book-length treatment. But I think the book would end with “even though it’s weird and complicated, poor people today who make $10,000 or $20,000 are often unhappy, in a way that richer people today aren’t, and this involves money in a real sense.”

I am not the person to write this book (though see the post on cost disease); I can only relay what poor people tell me. Sometimes it’s “my rent-controlled apartment is underneath noisy frat boys who keep me awake every night with their parties, but I can never leave because it’s the only apartment I can afford in this town.” Sometimes it’s “I hate my boss but I can’t leave because if I go a month without getting a paycheck I won’t have enough money for rent.” Sometimes it’s “I couldn’t afford good birth control, got pregnant, and now I can’t afford to support the child, what do I do?” Sometimes it’s “Obamacare mandates me to buy health insurance, but I can’t afford it, I guess I am going to have to pay a fee I can’t afford on tax day instead.” Sometimes it’s any of a thousand versions of “my car broke down and I can’t afford to get it fixed but I need to get to work somehow”. Sometimes it’s “I am sick but if I miss a day of work my company will fire me, because when you’re poor enough legally-enshrined workplace protections somehow fail to exist in real life”. And sometimes it’s “I work eighty hours a week driving for Uber because it’s the only way to make ends meet, I hate everything.” A lot of times it involves the same crappy job-centered lifestyle I worry a basic jobs guarantee would perpetuate forever.

Trying to steelman the “it’s not money” point further takes us to Sarris’ other essay on UBI, where he writes:

> Rent is currently eating the world. Rental income just hit an all-time high. If everyone is given a very predictable amount of money, it may be seen as a system that can be gamed by landlords and maybe other essentials producers. Implementing UBI without reforming land use and zoning regulations may end up as nothing more than a slow transfer to landlords. What are the odds of that happening? Well, it seems like it already did happen with healthcare and college tuition (loans) in the US, and if those are our guide, the “money” part and the “meaningful reforms” part should be done in a very particular order.
>
> Since housing does work well in some places (Japan and Montreal come to mind) I think this is a problem that can be fixed. But without the fix first, UBI may be punting real political problems while giving the appearance of solving them (until years later), and making the price inflation obvious for landlords, just like it was for healthcare companies and colleges getting guaranteed loans.
>
> Payments as a solution to a broken system is not the same as fixing the system. If UBI punts this real problem, we’ll be creating a financial time bomb.

This is basically how I think about any request for giving more money to education or health care, so I guess I have to take it seriously. Maybe the situations aren't exactly the same – education and health care seem to eat up money by hiring administrators, which doesn't have an obvious analogy to ordinary individuals. But the Kansas farmhouse example suggests that something like this must go on even at the personal level.

It looks like probably what's being described is that – absent some magical ability to create new houses out of thin air (a task known to be beyond the limits of modern technology) – housing is a positional good and so raising the position of everyone equally will just give extra cash to landlords. The best that can be said here is that insofar as these goods aren't perfectly inelastic, basic income will help a little. And insofar as other goods used by poor people (cars? furniture? generic medications?) are decently elastic, basic income will help a lot. I do agree the problem exists.

But I think this is one case where basic income is clearly better than basic jobs. All basic jobs can do is give you money, which can get eaten by rent-seekers. Basic income gives you freedom. Somebody works 50 hours a week at two McJobs to afford an apartment, gets basic income, and then they work 20 hours a week at one McJob and afford their apartment. The price of an apartment doesn't change, but their life has improved.

And by lowering the demand for jobs, basic income provides the seed of a solution to the housing problem. The reason rent costs so much in the Bay Area is because everyone wants to live in the Bay Area because it has so many great jobs. You can buy a house in the country (or in an unpopular city) for cheap; people don't because the jobs aren't as good, or the good jobs take longer to find. Freed from the need to live right in city center (or right next to the subway stop leading to city center), people can spread out again. If rent is $2000 in San Francisco and $500 in Walnut Creek, they can live in Walnut Creek (and still go to San Francisco whenever they want – cities are very accessible from suburbs, for every purpose *except* commuting during rush hour five days a week).

Go to the suburbs and people are building new housing tracts all the time. Supply is elastic and everyone’s backyards are so far away from one another that NIMBYs mostly stay quiet. It’s only when our job-centered culture forces everybody into historic San Francisco city center that we start having problems.

There’s still going to have to be a hard battle against cost disease. But much of the cost disease comes from overregulation and creeping socialism, and much of overregulation and creeping socialism come from well-intentioned concerns about the poor. Witness how California’s recent housing bill was opposed by socialists making vague warnings about “greedy developers”. If we can solve the non cost-disease-related parts of poverty first, maybe the socialists will lose some power and we can start fighting the cost disease problem in earnest.

iv) Without work, people will gradually lose meaning from their lives and become miserable

After claiming that money isn’t really a problem for most people, Sarris continues:

> The biggest societal ill today is not that people don’t have enough money to survive, it is that to survive and thrive people need things beyond food and rent: Social responsibility, sense of purpose, community, meaningful ways to spend their time, nutrition education, and so on.

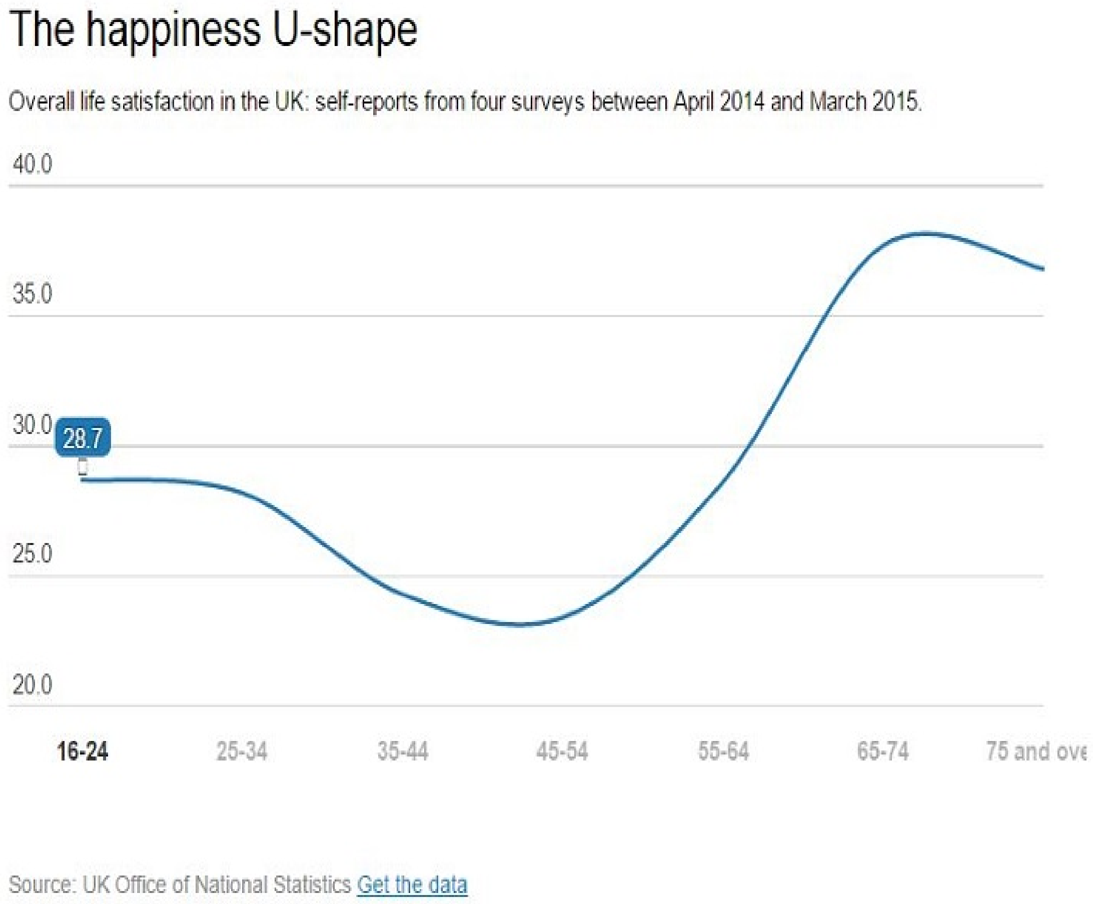
> If we fixate merely on the money aspect, we may be misdiagnosing what is making our 21st century so miserable for so many people.
>
> From some psychologists’ points of view, one of the worst things you can do to someone who is suffering from addiction or loss of hope is to give them no-strings-attached money, when what they really need is regularity and the responsibility that comes from having a purpose, even if its simply a job or a station. Basic Jobs have a chance of making the opioid crisis better, UBI risks making it worse…the at-risk population in the US need functions and responsibility more than just a check.

Social responsibility. Sense of purpose. Community. Meaningful ways to spend your time. This is some big talk for promoting jobs that in real life are probably going to involve a lot of “Do you want fries with that?” Getting a sense of purpose from your job is a crapshoot at best. Getting a sense of purpose outside your job is a natural part of the human condition. The old joke goes that nobody says on their deathbed “I wish I’d spent more time at the office”, but the basic jobs argument seems to worry about exactly that.

And let’s make the hidden step in this argument explicit. Everyone on basic income will have the opportunity to work if they want. In fact, they’ll have more opportunity, since people who hate working will have dropped out of the workforce and demand for labor will rise. So the basic jobs argument isn’t just that people need and enjoy work. The argument is that people need and enjoy work, but also, they are too unaware to realize this, and will never get the work they secretly crave unless we force them into it.

That doesn’t seem right. I don’t know enough hopeless opiate addicts to contradict an apparent psychological consensus on them, but it seems to me a lot of people do perfectly well finding meaning on their own time.

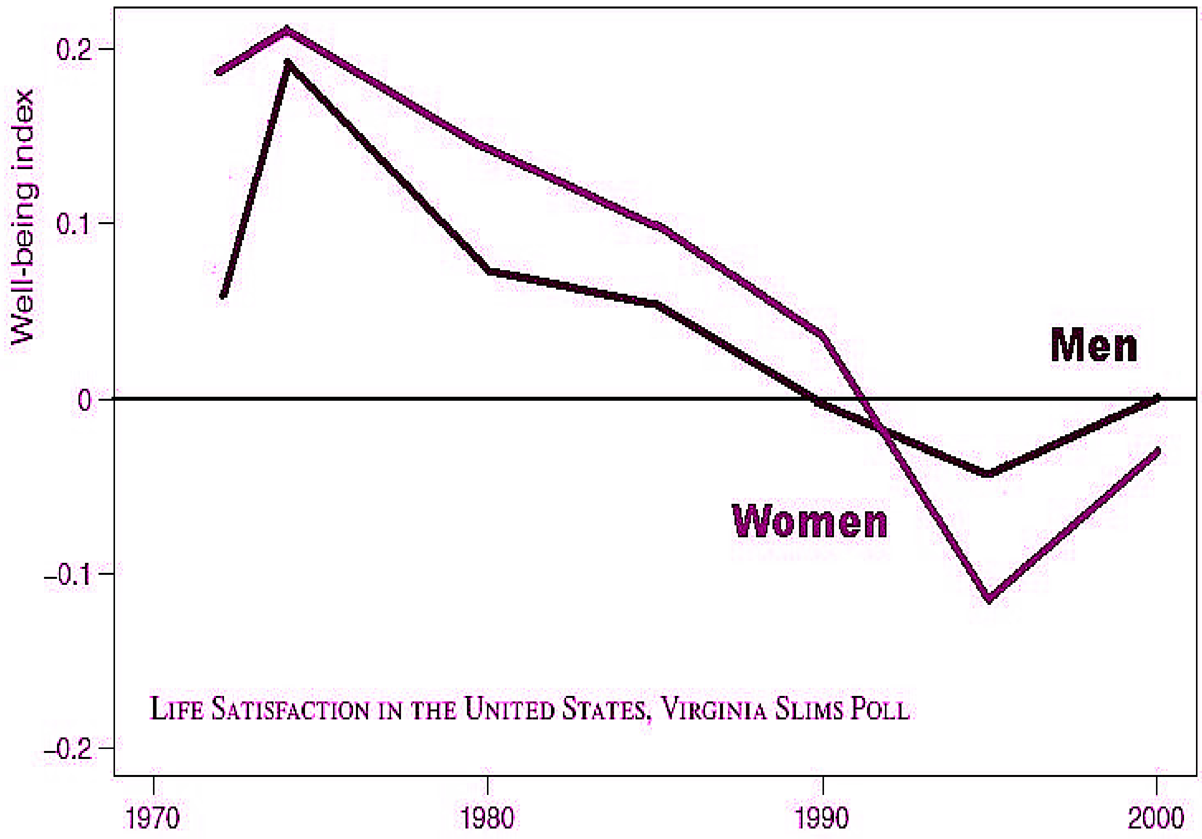
What about the retired? The graph of happiness vs. age looks like this:



This is not the shape we would expect if stopping work suddenly made you miserable and deprived you of purpose. Retired people seem to avoid work just fine and have lots of fun golfing, watching golf tournaments, going on golf vacations, arguing about golf, and whatever else it is retired people do.

Sarris says that “If you think UBI would not make the opioid crisis worse, the onus is on UBI proponents to show how writing ‘UBI’ on the top of the check instead of ‘disability’ would do that.” I would counter-argue that the onus is on opponents to explain why writing ‘UBI’ on the check works so much worse than writing ‘Social Security’.

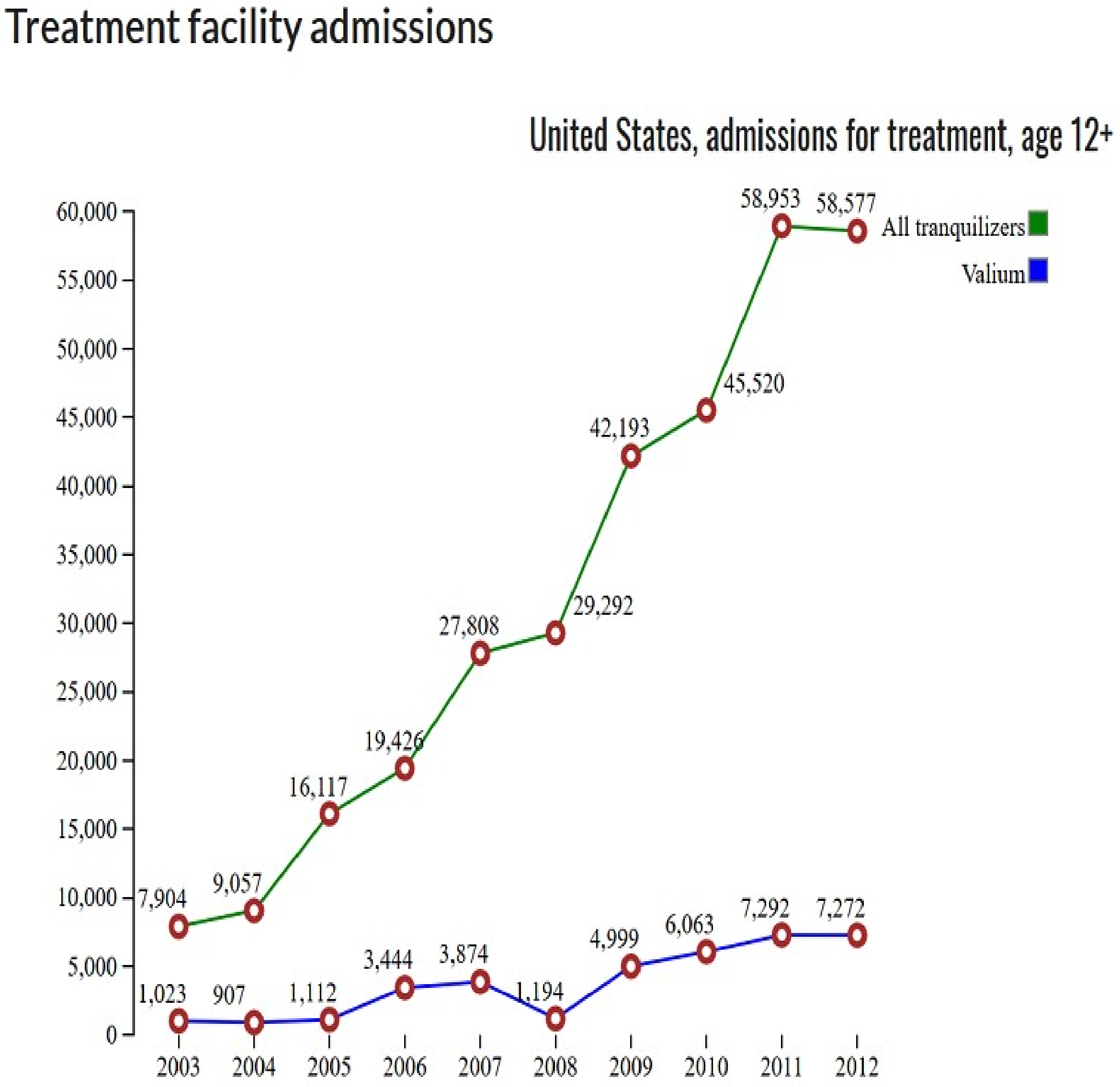
What about homemakers? Yes, homemaker is a full-time job. But it’s the full-time job a lot of people would do if they didn’t have to do their regular full-time job, which makes it fair game when we’re talking about basic income. Here’s a graph of male vs. female happiness over time:



If we assume most women in 1970 were homemakers, and most women in 2000 are working, their shift from homemaking to working doesn’t correspond to any improvement in happiness, either absolutely or relative to men.

There is some debate over whether modern-day homemakers are happier than modern-day workers or vice versa, with the most careful takes usually coming down to “people who prefer to stay home are happier staying home, people who prefer to work are happier working”. But there is no sign of the collapse in meaning and happiness we would expect in homemakers if not having an outside-the-house job reduces you to purposeless nihilism.

When I bring this up to people, they always have the same objection: “Didn’t women back then use lots of tranquilizers because of how stressed and upset they were? Didn’t they even call Valium ‘Mother’s Little Helper?'” Yes. But take it from a psychiatrist who prescribes them: people still use lots of tranquilizers. Nobody cares anymore, because it’s no longer surprising or ironic.



Sure glad that tranquilizer overuse problem got nipped in the bud in the 1970s when we cancelled stay-at-home parenting.

What about aristocrats? History presents us with many examples of entire classes who managed to live off other people’s work and avoid working themselves. These people seem to have not only have been pretty happy with the deal, but often used their free time to contribute in less purely economic ways. Lord Byron and Warner von Braun were hereditary barons, Bertrand Russell a hereditary earl, de Broglie a hereditary Duke, Condorcet and de Sade hereditary

Marquises. Von Neumann’s family was some kind of nouveau riche Austro-Hungarian nobility; Wittgenstein’s family was something similar. Winston Churchill was grandson of a Duke and son of a Lord. None of them ever had to worry about money: society gave them a giant basic income check from their ancestral estates.

Yet Churchill found meaning by saving the UK. Von Braun found meaning by shooting missiles at the UK. Condorcet found meaning by becoming one of the foremost defenders of human rights. De Sade found meaning by becoming one of the foremost violators of human rights. De Broglie and von Neumann found meaning by contributing to fundamental physics. Russell and Wittgenstein found meaning by literally figuring out what meaning was. Overall they seem like a pretty flourishing bunch.

What about college students? Technically they have to go to classes, but a lot of them get away with ten hours or less of class per week, and even more of them just never attend. Some, like Mark Zuckerberg and Bill Gates, use the extra time to found startups. Others, like everyone else, use the extra time to party and take lots of drugs. Either way, they seem pretty happy.

What about the self-employed? Being self-employed costs you a lot of the supposed psychological benefits of work. You might not be leaving the house. You might not be interacting with other people. But studies find that the self-employed are happier than the other-employed, even though they work longer hours and have less job security.

What about hunter-gatherers? Hunting-gathering in a fertile area is a pretty good gig, and usually lets people support themselves with only a few hours’ work per day. Most evidence suggests they’re pretty happy despite their lack of material goods.

What about schoolchildren? Every year, I would complain that I hated school. Every year, my mother would repeat some platitude like “Oh, when summer comes around you’re going to be so bored that you’ll be begging to go back”. And every year, summer vacation would be amazing, and I would love it, and I would hate going back to school with every fiber of my being. I understand this is pretty much a consensus position among schoolchildren. This has left me forever skeptical of arguments of the form “Oh, if you had freedom you would hate it”.

What about me? When I graduated medical school, I applied to residency and was rejected. That left me with a year open before I could try again. Thanks to some odd jobs, a little savings, and charity from friends and family, I was able to subsist. I spent the year meeting new people, hiking around California, falling in love, studying philosophy, and starting this blog. At the end of the year I applied to residency again and was accepted. I’m glad I got the job I wanted, but I also remember that year fondly as maybe the best I’ve ever had, and the one that set the stage for a lot of the good things in my life that happened since. I think this is pretty common for well-off people. We call it a “gap year” or a “sabbatical” or “going off to find yourself” or any of a bunch of other terms that disguise how it’s about doing exactly what people say you can’t do – being happy without a 9 to 5 job.

When I bring these points up, basic jobs advocates usually find reasons to dismiss all of them. Schoolchildren and college students are at a special part of their life that doesn’t generalize. Homemakers like being with their kids. Aristocrats get the world as their oyster. Retirees are mysteriously and permanently mesmerized by golf, which becomes an ur-need subsuming all other human desires. Hunter-gatherers are evolutionarily adapted to their lifestyles. I am just weird. They dismiss all of these as irrelevant and go back to their core example: in the US, right now, unemployed and disabled people are terribly unhappy.

I accept the very many studies that show this, but I do wonder if this has more to do with contingent features of unemployment than with work being necessary to human flourishing. For example, unemployed people are chronically low on money. Unemployed people face stigma and constant social pressure to get employment. Unemployed people live in a society built around and emphasizing jobs. Unemployed people may have pre-existing problems in their lives that led to their unemployment. Unemployed people sometimes suffer from disabilities or chronic pain. Unemployed people have no friends to hang out with during business hours because everyone else is working.

If you compared gay vs. straight happiness in 1980, you probably would have found gay people were much less happy. Now some studies suggest that in liberal and accepting areas, they are as happy or happier. The relative happiness of different groups isn't necessarily a human universal; it can also depend on how society treats them.

Given all this, I lean in favor of thinking most people would tolerate financially secure leisure time just fine. I might be wrong. *But I am still more comfortable letting people decide for themselves.* People who try leisure and like it – or who prefer homemaking, or taking care of elderly parents, or anything else – will stay out of the workforce. People who try leisure and don't like it will apply for the new, better class of jobs that will exist once increased demand for labor has forced employers to up their standards. Or they'll go volunteer at their church. Or they'll start a nonprofit. Or they'll do something ridiculous like try to be the first person to unicycle around the world.

Or maybe the meaninglessness of modern life will start to recede. Why don’t we have strong communities anymore? One reason I keep hearing from my patients is that they had lots of friends and family back home in Illinois or Virginia or wherever – but all the good jobs are in the Bay Area so now they live here and don’t know anybody. My own friends have managed to set up a halfway-decent semi-intentional community in California, but only because by a happy coincidence they all work in computers and all the good computer jobs are in the Bay. Freeing people from needing to orient their entire life around where they can get a job might lead to a lot more intentional communities like mine. Or it might lead to other things we can’t think of right now. A bunch of people with a lot of leisure time to throw at problems, and a bunch of people with money and a problem of meaningless, seems like a pretty good combination if you’re looking for meaning-as-a-service.

The best studies on homemakers find that women who want to be homemakers are happier as homemakers and sadder if forced to work, and women who want to work are happier as workers and sadder if forced to stay at home. I would not be surprised if there are some people who are happiest working, and others who are happiest pursuing leisure activities. A basic income would make it easier for both groups to get what they want.

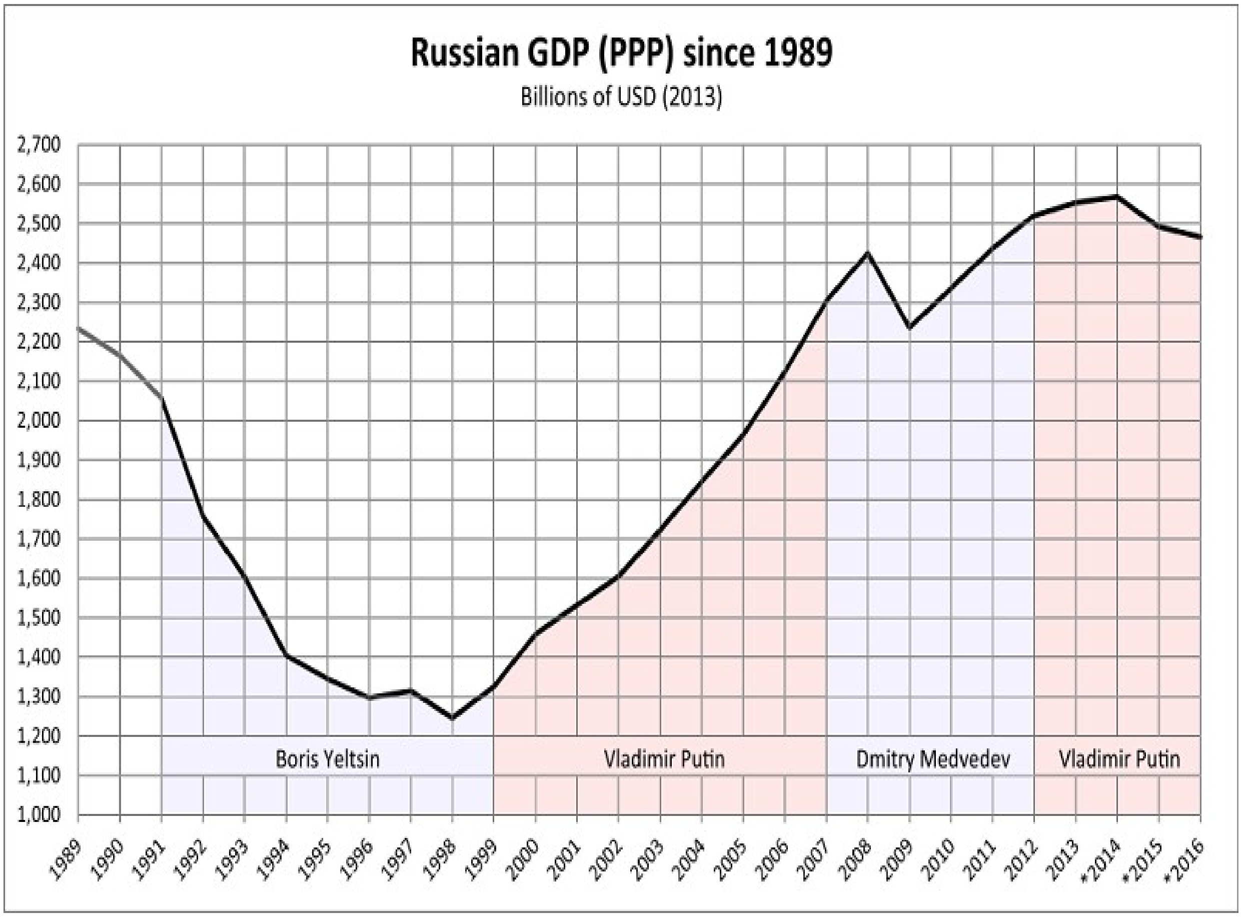
v) If something went wrong, basic jobs programs could be more gracefully wound down.

> What if it doesn’t work? What if we run out of jobs? Suppose a Basic Job program fails 20–30 years into the future. Maybe there’s too much corruption or not enough oversight, or the political will is no longer there, or the money itself is no longer there. Contingency planning is

> good: No matter how much you trust the pilot, you still want an airplane with emergency exits.
>
> If this happens, the side effects seem less severe (or even mildly positive) when contrasted with a UBI failure. So what if we accidentally fund farms, and bakeries, and furniture production, and house construction, and all sorts of small scale crafts across the country? Even in pessimistic scenarios we can expect some of the businesses and functions built to continue serving their communities after an official program is gone, in the same way that the Hoover dam is still there. A Basic Job program can plan for contingencies and the divvying up of what's been created, democratically, by community. Sheep farmers that are no longer supported by the government have at least got their flocks. If things ever go south, Basic Jobs better position us to try something else.

"So *what* if we accidentally fund farms?" asked Stalin, creating the kolkhozes. Maybe I am being mean here, but "let's guarantee full employment by sticking poor laborers on a government farm somewhere and teaching them to till the earth" is a plan that ought to set off as many historical alarm bells as "let's do something about all the Jews around here" or "let's murder the Mongol trade delegation".

True, nobody is proposing the other prong of socialist agricultural policy, which is crushing the private farms. But it's important to remember that what's being proposed is basically socializing large parts of the economy in ways that history tells us lead not only to agricultural catastrophe when being set up, but to economic ruin when being wound down:



In the 1990s, the GDP of Russia declined by 50%. Fifty percent! I don't know if that's ever happened before in history outside of a civil war or foreign invasion. The Iraqi economy survived the Iraq War and subsequent sectarian conflict better than the Russian economy survived winding down its basic jobs program.

Maybe I'm being unfair. Socializing part of the economy is probably safer than socializing all of it. And not crushing the private farms really does provide a safety valve that previous collectivization efforts lacked (though if the government farms are more subsidized than they are inefficient, you'll crush the private farms whether you want to or not).

But I'm still not sure if unsocializing the economy is as easy as winding down a basic income. If you want to wind down a basic income, you decrease it by 5% per year, and each year more people go to work in the private sector or start training to do so. If you want to wind down a nationwide system of collective farms, you – well, empirically you flail about for a while, collapse into a set of breakaway republics, and end up getting ruled by Vladimir Putin.

vi) Basic jobs could be used to create useful infrastructure

> Have the imagination to consider all of the work that is not being done, and FDR-style public works programs can be found almost everywhere. Building bicycle lane networks. Creating and maintaining public parks, flowerbeds, sidewalks. Demolition and recycling and re-urbanization (or re-forestation) of derelict factory grounds. There are so many things that would make parts of the US better places to live. As long as swaths of America are in disrepair and also where the jobs aren't, Basic Jobs has a mission to fulfill.

Some of my concern here comes from my concern (mentioned above) that basic-job-havers would not be very good employees, and that you would probably save money by handing needy people a check and separately hiring some super-efficient megacorporation to make your flowerbeds.

But another part comes from asking myself – which would I rather have? More flowerbeds and sidewalks? Or forty extra hours a week to spend seeing friends and family, or pursuing hobbies that I love? Framed this way, the answer is super-obvious – and remember, I love my job.

vii) Capitalism seems to have historically worked pretty well, and basic jobs guarantees preserve the best features of capitalism

> We want to try and keep [the] positive effect of capitalist economic transactions. UBI creates paychecks, Basic Jobs programs do too, but Basic Jobs also create transactions, incentives, and products, fulfilling secondary needs for society.
>
> Basic Jobs can be thought of as a program that is paying people to make other people's lives better in addition to their own. We are paying people to produce local food and crafts, in a subsidized fashion that gives communities an alternative to the WalMart-esque globalized marketplaces. If the government subsidizes the workers so that their goods can be competitive, it will foster local economies while putting money in the pockets of local worker who themselves have more power. Hopefully, the second-order effects of such commerce are large enough to notice. Maybe the benefits will stay. One could argue that the strong Swiss and other European agricultural subsidies are already a soft form of Basic Jobs.

"Capitalism" is a Rorschach test that means many things to many people. Some people think it means oppression, discrimination, and exploitation. Other people think it means any level of freedom better than you get in Maoist China. Still other people identify it with corporations, or banks, or barter, or any of a thousand other things. But to me, if capitalism means anything at all, it means…

Well, remember argument iv above? About how maybe poor people's lives will be meaningless without work, and maybe they're not sufficiently self-aware to realize that on their own, so the government should make them work for their own good, in whatever industry most needs their help?

To me, capitalism means shouting "FUCK YOU" at that argument, at the intuitions behind that argument, and at the whole social structure that makes those intuitions possible, then sterilizing the entire terrain with high-quality low-cost American-made salt so that no other argument like it can ever grow again. There are other parts of capitalism, like the stuff about stock exchanges, but they all flow from that basic urge.

Capitalism certainly doesn't mean you should never get money without working. Heck, some leftists would *define* a capitalist as a person who gets money without working. The part where you get money without working is the fun part of capitalism. The thing where most people *don't* get that is the part that could do with some fixing. That's why a lot of history's greatest capitalists (in both senses of the word) – from Friedrich Hayek and Milton Friedman to Mark Zuckerberg and Elon Musk – have supported basic income.

My intuitions are basically Georgist *(note to self: read Henry George before saying this too many more times)*. Capitalists deserve to keep the value they create, but they also owe rent on common resources which they enclose and monopolize (eg land, raw materials). That rent gets paid to the State (as representative of the people who are denied use of the commons) in the form of taxes. The State then redistributes it to all the people who would otherwise be able to enjoy the monopolized resources – eg everybody. I think this process where businesses pay off the government for their raw materials is pretty similar to the process where they pay off the investors for their seed money, and that the whole thing fits within capitalism pretty nicely.

I don't think the government taking a big role in the economy for Your Own Good can ever really fit within capitalism, at least not the parts of it that I consider valuable. I would consider a basic jobs guarantee, if it lasted, to be a victory for socialism over the parts of capitalism I hold dear – the final triumph of the old Soviet joke about how "We pretend to work, they pretend to pay us". If you want an image of the future, imagine a glassy-eyed DMV employee staring at a clock, counting down the hours until she can go home – forever.

And that's what we're debating here – an image of the future. These basic guarantees always get brought up in the context of technological unemployment. I've looked into this before, and although I don't think jobs are being destroyed per se, I think it's definitely possible they're getting worse for complicated reasons. So as more and more people start getting worse and worse jobs, we can choose one of two paths.

First, we can force more and more people into make-work low-paying government jobs. Extrapolate to the very far future, and 99% of the population will spend their time sending their kids off to daycare before a long day of digging ditches that a machine could have dug better, while 1% of people have amazing robot empires.

Second, we can try to break the link between toiling for someone else and being able to live. We can set some tax rate and promise that all revenue above some amount necessary to fund state functions will be redistributed as basic income. It'll be pretty puny at first. But as GDP grows, more and more people will opt out of work. As the payments increase, we can gradually transfer various forms of welfare into insurance, and use the gains to grow the payments further. There will be plenty of well-paying jobs for whoever wants to keep working, and lives of leisure and enjoyment for the people who don't. Robots will pick up the slack and keep the big corporations generating the value that gets siphoned off. Extrapolate to the very far future, and 99% of people live in constantly-improving comfort and freedom, while 1% of people have that plus amazing robot empires.

Both of these are kind of tame shock-level-zero visions. But they set the stage for whatever comes next. If we have genetically enhanced superchildren, or Hansonian em overlords, they're going to inherit the same social structures that were on the scene when they got here. Whatever institutions we create to contain today's disadvantaged will one day be used to contain us, when we're disadvantaged in a much more fundamental way. I want those structures to be as autonomy-promoting as possible, for my own protection.

I grudgingly admit basic jobs would be an improvement over the status quo. But I'm really scared that it becomes so entrenched that we can never move on to anything better. Can anyone honestly look at the DC education system and say "Yeah, I'm glad we designed things that way"? Doesn't matter; we're never going to get rid of it; at this point complaining about it too much would send all the wrong tribal signals. Nothing short of a civil war is going to change it in any way beyond giving it more funding. I dread waking up in fifty years and finding the same is true of basic jobs.

This is what I mean by hijacking utopia. Basic income is a real shot at utopia. Basic jobs takes that energy and idealism, and redirects it to perpetuate some of the worst parts of the current system. It's better than nothing. But not by much.

[EDIT: Sarris' response, where he argues that I am comparing the most utopian formulation of basic income to a very practical 'let's get a few unemployed people back to work' version of basic jobs.]

Can Things Be Both Popular And Silenced?

The *New York Times* recently reported on various anti-PC thinkers as “the intellectual dark web”, sparking various annoying discussion.

The first talking point – that the term is silly – is surely true. So is the second point – that it awkwardly combines careful and important thinkers like Eric Weinstein with awful demagogues like Ben Shapiro. So is the third – that people have been complaining about political correctness for decades, so anything that portrays this as a sudden revolt is ahistorical. There are probably more good points buried within the chaff.

But I want to focus on one of the main arguments that’s been emphasized in pretty much every article: can a movement really claim it’s being silenced if it’s actually pretty popular?

“Silenced” is the term a lot of these articles use, and it’s a good one. “Censored” awkwardly suggests government involvement, which nobody is claiming. “Silenced” just suggests that there’s a lot of social pressure on its members to shut up. But shutting up is of course is the exact opposite of what the people involved are doing – as the *Times* points out, several IDW members have audiences in the millions, monthly Patreon revenue in the five to six figures, and (with a big enough security detail) regular college speaking engagements.

So, from *New Statesman*, If The “Intellectual Dark Web” Are Being Silenced, Why Do We Need To Keep Hearing About Them?:

> The main problem with the whole profile is that it struggles because of a fundamental inherent contradiction in its premise, which is that this group of renegades has been shunned but are also incredibly popular. Either they are persecuted victims standing outside of society or they are not. Joe Rogan “hosts one of the most popular podcasts in the country”, Ben Shapiro’s podcast “gets 15 million downloads a month”. Sam Harris “estimates that his Waking Up podcast gets one million listeners an episode”. Dave Rubin’s YouTube show has “more than 700,000 subscribers”, Jordan Peterson’s latest book is a bestseller on Amazon […]
>
> On that basis alone, should this piece have been written at all? The marketplace of ideas that these folk are always banging on about is working. They have found their audience, and are not only popular but raking it in via Patreon accounts and book deals and tours to sold-out venues. Why are they not content with that? They are not content with that because they want everybody to listen, and they do not want to be challenged.
>
> In the absence of that, they have made currency of the claim of being silenced, which is why we are in this ludicrous position where several people with columns in mainstream newspapers and publishing deals are going around with a loudhailer, bawling that we are not listening to them.

Reason’s article is better and makes a lot of good points, but it still emphasizes this same question, particularly in their subtitle: “The leading figures of the ‘Intellectual Dark Web’ are incredibly popular. So why do they still feel so aggrieved?”. From the piece:

> They can be found gracing high-profile cable-news shows, magazine opinion pages, and college speaking tours. They’ve racked up hundreds of thousands of followers. And yet the ragtag band of academics, journalists, and political pundits that make up the “Intellectual Dark Web” (IDW)—think of it as an Island of Misfit Ideologues—declare themselves, Trump-like, to be underdogs and outsiders. […]
>
> [I’m not convinced] they’re actually so taboo these days. As Weiss points out, this is a crowd that has built followings on new-media platforms like YouTube and Twitter rather than relying solely on legacy media, academic publishing, and other traditional routes to getting opinions heard. (There isn’t much that’s new about this except the media involved. Conservatives have long been building large audiences using outside-the-elite-media platforms such as talk radio, speaking tours, and blogs.) In doing so, they’ve amassed tens and sometimes hundreds of thousands of followers. What they are saying might not be embraced, or even endured, by legacy media institutions or certain social media precincts, but it’s certainly not out of tune with or heretical to many Americans.
>
> The bottom line is there’s no denying most of these people are very popular. Yet one of the few unifying threads among them is a feeling or posture of being marginalized, too taboo for liberal millennial snowflakes and the folks who cater to them.

The basic argument – that you can’t be both silenced and popular at the same time – sounds plausible. But I want to make a couple points that examine it in more detail.

1. There are lots of other cases where we would agree there’s some form of silencing going on, even as a group has many supporters and rich, famous spokespeople

I know a lot of closeted transgender people. They’re afraid to come out as trans, they talk about trans people being stigmatized and silenced, and they clearly have a point. Does anyone disagree that it can be dangerous to be a trans person even in the First World, let alone anywhere else?

On the other hand, Caitlyn Jenner is on the cover of every magazine, won Woman Of The Year, got her own documentary and reality TV show, and earns up to $100,000 per public appearance, with a total net worth rumored to be around $100 million. She is probably one of the most famous and popular people in the world.

Only a moron would make an argument like “Caitlyn Jenner is doing very well, therefore there’s not really a stigma around transgender”. For one thing, your success is a function of how many people like you, not your net (likers – haters) total. For another, Hollywood is its own world and probably doesn’t correlate with any particular person’s social sphere. And for another, Jenner is popular partly because of how surprising and controversial her transition was – her story is at least partly a function of “look how brave this person is to defy social stigma this way”.

Transgender people complain of social shaming, silencing, and stigma. Some transgender people can become very famous celebrities who everyone agrees are rich and popular. And nobody finds this at all surprising or thinks that these two claims contradict each other.

(No, Twitter, I’m not making the claim “Sam Harris is exactly as marginalized as transgender people”. I’m saying that even groups who we all agree are more marginalized than the IDW can have very successful and famous spokespeople.)

Or what about the early US labor movement? They were faced with everything from Pinkerton goon squads, to industry blacklists, to constantly getting arrested on trumped-up charges; nobody seriously denies that government and private industry put a lot of effort into silencing them.

Yet they were very popular with their core demographic, and their most charismatic spokespeople remained famous and widely-liked. Emma Goldman would go around the country lecturing to packed halls, collecting far more energy and interest than Sam Harris gets nowadays when he does the same. If the papers of the time had said “Emma Goldman sure is popular for someone who says her movement is being silenced”, well, screw you and your dumb gotchas, that’s just a 100% accurate description of the state of affairs.

2. In fact, taboo opinions seem to promote a culture of celebrity

From Current Affairs:

> There are dozens of well-known critics of social justice activists: Harris, Shapiro, Peterson, Brooks, Stephens, Hoff Sommers, Weinstein, Weinstein, Murray, Murray, Rogan, Chait, Haidt, Pinker, Rubin, Sullivan, Weiss, Williamson, Yiannopoulos, Dreger, Hirsi Ali. Who are their equivalents among the Social Justice Types? Who has their reach or prominence?

A few people have tried to answer the question – and certainly a few names like Ta-Nehisi Coates belong in any such list. But I think the overall point is basically correct. If so, what does that mean?

Consider this: how many neo-Nazi/white supremacist activists are famous enough that the average news junkie would know their names? Maybe two: David Duke and Richard Spencer. Okay. How many low-tax activists are equally famous? I think just one: Grover Norquist. There are some important people who happen to support low taxes among many other causes (eg Paul Ryan) but they don't count – if they did, our list of famous "social justice types" would have to include Hillary Clinton and a hundred others.

Presumably we shouldn't conclude that neo-Nazism is twice as common/popular/acceptable as tax cuts. But that means you can't always measure how popular an ideology is by counting its famous advocates.

I'd go further and say that more taboo ideas are more likely to generate famous spokespeople. If you can't think of any modern feminists with star power, you can always go back to the 1970s and find people like Gloria Steinem and Andrea Dworkin – who made waves by being at least as outrageous then as the IDW is now. If Ta-Nehisi Coates isn't famous enough for you, Martin Luther King and Malcolm X certainly will be. Malcolm X didn't get more famous than Ta-Nehisi Coates by being *more well-liked*, he got famous by being as controversial and threatening and feared as Coates is accepted. So the implication of the *Current Affairs* article – we mostly hear about well-liked people, and really controversial people never get famous – seems questionable at best and backwards at worse.

But why would more taboo causes generate more celebrity? Here are some ways I think this could work:

1. Controversy sells in general. Caitlyn Jenner is more famous than Bruce Jenner not because transgender is less stigmatized than running, but because it's more likely to provoke debate.

2. All else being equal, if an ideology is taboo, it should have fewer loud open activists per covert believer than an orthodox ideology. But that means the field is less crowded. If feminism has 1 loud activist per 10 believers, and the IDW has 1 loud activist per 1000 believers, then the feminist activist will generally be speaking to a college club, and the IDW activist to a crowded lecture hall. This will catapult the IDW activists to greater celebrity.

3. Activists for taboo views need a skill that activists for orthodox views don't – that of surfing controversy. The insult "edgelord" is basically correct – they thrive by being on the edge of what is acceptable. If you go completely beyond the bounds of what is acceptable, you fall from grace – either into literal ruin, or just having your fan base shift entirely to being weird alt-right people whom you hate and don't want to be associated with. Only people who can continually surf that boundary – edgy enough to be interesting, restrained enough to get the *New York Times* to write basically positive editorials about you – are really able to make it. Most people correctly assume they would screw up and end up totally taboo rather than delightfully edgy. Once again, this makes the field less crowded, giving everyone who comes in more star power per person.

4. Orthodox ideologies tend to be well-represented within institutions, meaning that the ideologies' leaders are more likely to be institutionally prestigious people. Taboo views

are unrepresented within institutions, meaning their spokespeople kind of just arise naturally by being really good at getting attention and acclaim. The natural “leaders of feminism” might be Women’s Studies professors, Planned Parenthood directors, and whoever the most feminist person at the *New York Times* is. These people might be very good at what they do, they might even be very effective at promoting feminism, but they’re probably less good at getting attention than people who have been specifically selected for that trait. And with the institutional leaders sucking up all the status, it might be harder for some woman who’s just a very good writer and really in-touch with the zeitgeist to say “Yes, I am the leader of feminism, everyone please care about me now”.

5. Generic famous people will support orthodox causes, but not taboo causes. The absence of people famous for feminism is counterbalanced by a glut of famous people who happen to be feminists. Here is a list of actors who say they are proud to call themselves feminist, also just known as “a list of actors”. Famous people who are against feminism are more likely to keep quiet about it, creating a void for specific anti-feminist celebrities can fill.

6. Celebrity helps launder taboo ideology. If you believe Muslim immigration is threatening, you might not be willing to say that aloud – especially if you’re an ordinary person who often trips on their tongue, and the precise words you use are the difference between “mainstream conservative belief” and “evil bigot who must be fired immediately”. Saying “I am really into Sam Harris” both leaves a lot of ambiguity, and lets you outsource the not-saying-the-wrong-word-and-getting-fired work to a professional who’s good at it. In contrast, if your belief is orthodox and you expect it to win you social approval, you want to be as direct as possible.

I don’t know if these six points really explain the phenomenon. But I think there’s definitely a phenomenon to be explained, and I think “crowded field” is a big part of it. In my own experience, my blog posts promoting orthodox opinions are generally ignored; my blog posts promoting controversial opinions go viral and win me lots of praise. I assume this is because my orthodox blog posts are trying to outcompete the people at Vox (highly-polished, Ivy-League-educated mutants grown in vats by a DARPA project to engineer the perfect thinkpiece writer), and my controversial blog posts are trying to outcompete three randos with blogs that consistently confuse “there” and “their”. Winning one competition is much easier than winning the other – and the prize for winning either is “the attention of about 50% of the population”.

3. Fame lets people avoid social repercussions, but that doesn’t mean those repercussions don’t exist for ordinary people

Caitlyn Jenner can be as visibly and fabulously transgender as she wants, because being transgender is a big part of her job. She’s organized a lot of her life around being a transgender person. Any friends she was going to lose for being transgender have already been written off as losses. Anybody who wants to harm her for being transgender is going to get stopped by her bodyguards or kept out of her giant gated mansion. When she argues that transgender people face a lot of stigma, fear, and discrimination, she mostly isn’t talking about herself. She’s talking about every transgender person who isn’t Caitlyn Jenner.

Likewise, Sam Harris is pretty invincible. As a professional edgelord, he is not going to lose his job for being edgy. Whatever friends he’s going to lose for being Sam Harris, he’s already written off as losses. I assume he has some kind of security or at least chooses not to live in Berkeley. So when he’s talking about his ideas being taboo, he means *taboo for everybody who isn’t Sam Harris*.

I worry that this conversation is being conducted mostly by media personalities who write controversial takes for a living. They work for ideologically-aligned publications, and everyone knows that a few crazies hating and harassing you is a common part of the job. If you didn’t propose the death penalty for abortion and then get a job at *The Atlantic*, you’ll probably be fine.

Out in the rest of the world, if a rando on social media calls your company and tells them you’re a Nazi because [out of context tweet], the complaint is going straight to a humorless 60-year-old HR drone whose job is minimizing the risk of PR blowups, and who has never heard of Twitter except as a vague legend of a place where everything is terrible all the time. So if you write for a webzine, consider that you may have no idea how silenced or living-in-fear anyone else is or isn’t, and that you may be the wrong person to speculate about it.

Out in the rest of the world, if someone sends you a death threat, you might not be such an experienced consumer of Internet vitriol that you know it usually doesn’t pan out. You might not be so thick-skinned that “Go to hell, you fucking Nazi scum” no longer has any effect on you. You might not live in an bubble of intellectualism where people appreciate subtle positions. You might have friends and family who are very nice people but somewhat literal-minded, who have heard that only rapists oppose feminism so many times that they have no ability to create a mental category for someone who opposes feminism but isn’t pro-rape. And you might not really relish the idea of having to have a conversation with your sweet elderly great-aunt about how no, you *really* don’t think raping people is good. Seriously, imagine having to explain *any* of what you write on the Internet to your sweet elderly great-aunt, and now imagine it’s something that society has spent years telling her is equivalent to rape apologism.

(my father recently implied I had brought dishonor upon our family by getting quoted approvingly in *National Review*. I am 90% sure he was joking, but only 90%.)

Or maybe I’m wrong about this. Part of how silencing works is that nobody really knows how strong it is or isn’t. I had a patient who agonized for years over whether to come out to his family, only to have his parents say “Yes, obviously” when he finally got up the nerve. The point, is *Sam Harris no longer has to worry about any of these things*. So if your line of reasoning is “well, Sam Harris seems to do pretty well for himself, so I guess you can’t get in trouble for being controversial”, I don’t know what to tell you.

4. If you spend decades inventing a powerful decentralized network to allow unpopular voices to be heard, sometimes you end up with unpopular voices being heard

Sam Harris’ business model is a podcast with a Patreon, advertised by Internet word-of-mouth. This is pretty typical for the “intellectual dark web” figures.

The Internet promised to take power away from media gatekeepers and make censorship near-impossible. In discussing the many ways in which this promise has admittedly failed, we tend to overlook the degree to which it’s succeeded. One of the most common historical tropes is “local government and/or lynch mob destroys marginalized group’s printing press to prevent them from spreading their ideas". The Internet has since made people basically uncensorable, not for lack of trying. More recently, crowdfunding has added the final part to this machine – semi-decentralized cash flow.

So, after hundreds of engineers and activists and entrepreneurs work for decades to create a new near-impossible-to-censor system, and some people who would never have gotten heard on any other channel are able to use it to get heard – well, it's pretty weird to turn around and say "Aha, you got popular, that proves nobody is trying to silence you!"

I think this also explains why, even though people have been talking about these issues forever, it's only becoming a "big deal" now. Before, people would either watch their mouths to avoid getting kicked out by major gatekeeper institutions – or they would go to explicitly right-coded spaces like talk radio where the gatekeepers already agreed with them.

What's new is that there's a third route in between "tame enough to be on CNN" and "conservative enough to be a guest on Rush Limbaugh". The new brand of IDW thinkers are interesting precisely because – excluding Ben Shapiro (always a good life choice) – they're *not* traditional conservatives. The thing that's new and exciting enough to get *New York Times* articles written about it is that the anti-PC movement has spread to friendly coastal liberals. From the Democrats' perspective, the IDW aren't infidels, they're heretics.

5. When the IDW claims they are threatened, harassed, and blacklisted, people should at least consider that they are referring to the actual well-known incidents of threats, harassment, and

blacklisting against them rather than imagining this is code for “they demand to be universally liked”

Here are some of the stories in Weiss’ original IDW editorial:

> A year ago, Bret Weinstein and Heather Heying were respected tenured professors at Evergreen State College, where their Occupy Wall Street-sympathetic politics were well in tune with the school’s progressive ethos. Today they have left their jobs, lost many of their friends and endangered their reputations. All this because they opposed a “Day of Absence,” in which white students were asked to leave campus for the day. For questioning a day of racial segregation cloaked in progressivism, the pair was smeared as racist. Following threats, they left town for a time with their children and ultimately resigned their jobs.

And:

> Mr. Peterson has endured no small amount of online hatred and some real-life physical threats: In March, during a lecture at Queen’s University in Ontario, a woman showed up with a garrote.

And:

> Dr. Soh said that she started “waking up” in the last two years of her doctorate program. “It was clear that the environment was inhospitable to conducting research,” she said. “If you produce findings that the public doesn’t like, you can lose your job.”
>
> When she wrote an op-ed in 2015 titled “Why Transgender Kids Should Wait to Transition,” citing research that found that a majority of gender dysphoric children outgrow their dysphoria, she said her

> colleagues warned her, “Even if you stay in academia and express this view, tenure won’t protect you.”

And:

> The University of California, Berkeley, had to spend $600,000 on security for Mr. Shapiro’s speech there.

So. Threats against a professor and his family forcing him to leave town. Another professor told that she would lose her job if she communicated research to the public. A guy needing $600,000 worth of security just to be able to give a speech without getting mobbed. Someone showing up to a lecture with a garrote. And Reason Magazine reads all this and thinks “I know what’s going on! These people’s only possible complaint is that they feel entitled to have everyone agree with them!”

Maybe I’m being mean here? But how else do I interpet paragraphs like this one?

> The supposed ostracism they suffer because of their views ultimately comes down to a complaint not about censorship or exclusion but being attacked, challenged, or denied very particular opportunities. They want to say the things they are saying and have the marketplace of ideas and attention not only reward them with followers and freelance writing gigs but universal acceptance from those that matter in the academy and chattering classes.

I am nowhere near these people either in fame or controversialness, but I have gotten enough threats and harassment both to be pretty sure that these people are telling the truth, and to expect that the stuff that fits in one article is probably just the tip of the iceberg.

(Do other groups face similar pressures? Absolutely. Would people who wrote similar articles using those groups' complaints to make fun of them also be antisocial? Absolutely.)

On a related note, what does the article mean by contrasting "excluded" vs. "denied very particular opportunities"? I understand the meaning of the words, but I am not sure the people writing about them have a principled distinction in mind. When Debra Soh faced pressure to quit academia, was she being "excluded" or "denied a very particular opportunity"? Would the 1950s version of *Reason* describe communist sympathizers as being "excluded", or as "denied very particular opportunities" in the film industry? If, as the surveys suggest, 20% of philosophers would refuse to hire transgender professors to their department, are transgender people facing "exclusion", or just being "denied very particular opportunities"?

[My position – if you decide not to hire someone based on any characteristic not related to job performance (very broadly defined, including things like company fit and fun to work with), you're trying to exclude people. If you make up a really strained dumb argument for why some characteristic relates to job performance when it obviously doesn't ("communist actors could try to hold a revolution on the set, thus making our other employees feel unsafe"), then you're trying to exclude people *and* lying about it. You can say, as many throughout history have "I'm proud to be part of the effort to fight the Communist menace by denying them positions of influence", and then you get points for honesty and (if the Communists were really as menacing as you thought) maybe utilitarianism points as well. But don't say "What? Me exclude Communists? We're just denying them very particular opportunities! Sure are a whiny bunch, those commies!” See also Is It Possible To Have Coherent Principles Around Free Speech Norms?]

6. The IDW probably still censor themselves

Another common point in this discussion has been that the IDW copies the worst parts of social justice – intense focus on the latest outrage, shoddy science, its own set of insults (“snowflake! triggered millennial!”), us-vs-them dichotomy, et cetera. And Despite Their Supposed Interest In Rational Discussion Actually They Are Very Bad At Supporting Their Points Rationally.

Here’s a site that hasn’t been in any “intellectual dark web” editorials and never will be: Human Varieties. You can Google it if you want, but I won’t direct-link them for the same way I wouldn’t build a giant superhighway to some remote forest village enjoying its peaceful isolation. Here’s an excerpt from a typical Human Varieties article:

> I did look through the PING survey (age 3-21, N ~ 1,500) – which might not be very informative owing to the age structure. Going by this, Greg [Cochran] seems to be more or less correct about some of the endo[phenotypic] differences and probably about their origins. As an example, Figure 1 & 2 show the [black/white] diff[erences] for intracranial and total brain volume by age. ([African-Americans] are picked out for illustration since they are the largest non-White ethnic group, showing the biggest deviation from Whites.) And Figure 3 shows the relation between brain volume and ancestry in the self-identified [African-American] group; the results were basically the same for intracranial volume, etc. — and so not shown.

Read Human Varieties for a while, and you notice a few things:

1. They’re much more taboo and openly racist (in the Charles Murray sense) than almost anyone in the “intellectual dark web”
2. They are much less annoying and less likely to shout “TRIGGERED! SNOWFLAKE!” than almost anyone in the “intellectual dark web”
3. Nobody pays any attention to them at all

I think all three of these are correlated.

If you want to be Human Varieties, you can talk about the evidence for and against various taboo subjects. But nobody wants to be them, for two reasons.

First, somebody is going to have to present the evidence for the taboo subject, not just in an edgy “what if…perhaps this should not be suppressed?? or did i blow your mind??” way, but in a “here’s exactly what I believe and why I believe it” way. This isn’t just Sam Harris level edgy, this is way off the edge into the void below.

Second, if you do even a moderately good job, it’s probably going to sound exactly like the quote above, stuff like “this survey of intracranial volume endophenotypes might not be very informative, owing to the age structure” – and everyone will fall asleep by minute two. People will do lots of things to own the libs, but reading an analysis of the age structure of endophenotype data probably isn’t one of them.

“TRIGGERED! SNOWFLAKE!” solves both these problems. You avoid the object-level debate about whether taboo subjects are true, and it’s automatically interesting to a wide range of people. “That other monkey has status that should be *my* status!” – nobody ever went broke peddling that.

I think this model knocks down a few reasonable-sounding but on-reflection-wrong critiques of the way these issues are discussed:

“The IDW demands rational debate, but they never engage in it”. Somewhat true. If they engaged in it, they would move beyond the bounds of acceptable edginess. “We wish we were allowed to talk about X without massive risk to our reputations and safety” and “We are definitely not going to talk about X right now” are hardly contradictory; they follow naturally from each other. And I think this is more subtle than people expect – somebody may feel they can get away with making some arguments but not others, giving them the appearance of a skeletal but flimsy ideology that falls down on close examination. Or people might be willing to talk about these issues in some low-exposure spaces but not other higher-exposure spaces, giving them the appearance of backing down once challenged.

“The IDW focuses too much on triggered snowflakes.” Somewhat true – even independent of this being popular and lucrative. This is the least taboo thing you can do while still getting a reputation for being edgy. And winning the free speech wars makes it easier to talk about other stuff.

“The IDW says they’re being silenced, but actually they’re popular”. Somewhat true, even independent of all the arguments above. The things they complain about not being able to say, aren’t the things they’re saying.

7. Nobody in this discussion seems to really understand how silencing works.

If you say “We know a movement isn’t being silenced because it’s got lots of supporters, is widely discussed, and has popular leaders” – then you’re mixing up the numerator and the denominator.

Gandhi’s Indian independence movement had lots of supporters, was widely discussed, and had popular leaders. So did a half dozen Irish revolts against British rule. And the early US labor movement. And Eastern Bloc countries’ resistance to Soviet domination. And Aung San Suu Kyi. And every medieval peasants’ revolt ever. And…well, every other movement that’s been suppressed. Really, what sort of moron wastes their time suppressing a leaderless movement that nobody believes in or cares about?

Popular support and frequent discussion go in the numerator when you’re calculating silencing. Silencing is when *even though* a movement has lots of supporters, none of them will admit to it publicly under their real name. *Even though* a movement is widely discussed, its ideas never penetrate to anywhere they might actually have power. *Even though* it has charismatic leaders, they have to resort to low-prestige decentralized people-power to get their message across, while their opponents preach against them from the airwaves and pulpits and universities.

Scott Aaronson writes about the game theoretic idea of “common knowledge” as it applies to society:

> If you read accounts of Nazi Germany, or the USSR, or North Korea or other despotic regimes today, you can easily be overwhelmed by this sense of, “so why didn’t all the sane people just rise up and overthrow the totalitarian monsters? Surely there were more sane people than crazy, evil ones. And probably the sane people even knew, from experience, that many of their

neighbors were sane—so why this cowardice?” Once again, it could be argued that common knowledge is the key. Even if everyone knows the emperor is naked; indeed, even if everyone knows everyone knows he’s naked, still, if it’s not common knowledge, then anyone who says the emperor’s naked is knowingly assuming a massive personal risk. That’s why, in the story, it took a child to shift the equilibrium. Likewise, even if you know that 90% of the populace will join your democratic revolt provided they themselves know 90% will join it, if you can’t make your revolt’s popularity common knowledge, everyone will be stuck second-guessing each other, worried that if they revolt they’ll be an easily-crushed minority. And because of that very worry, they’ll be correct!

(My favorite Soviet joke involves a man standing in the Moscow train station, handing out leaflets to everyone who passes by. Eventually, of course, the KGB arrests him—but they discover to their surprise that the leaflets are just blank pieces of paper. “What’s the meaning of this?” they demand. “What is there to write?” replies the man. “It’s so obvious!” Note that this is precisely a situation where the man is trying to make common knowledge something he assumes his “readers” already know.)

The kicker is that, to prevent something from becoming common knowledge, all you need to do is censor the common-knowledge-producing mechanisms: the press, the Internet, public meetings. This nicely explains why despots throughout history have been so obsessed with controlling the press, and also explains how it’s possible for 10% of a population to murder and enslave the other 90% (as has happened again and again in our species’ sorry history), even though the 90% could easily

> overwhelm the 10% by acting in concert. Finally, it explains why believers in the Enlightenment project tend to be such fanatical absolutists about free speech.

One can take this further:

> Bostrom makes an offhanded reference of the possibility of a dictatorless dystopia, one that every single citizen including the leadership hates but which nevertheless endures unconquered. It’s easy enough to imagine such a state. Imagine a country with two rules: first, every person must spend eight hours a day giving themselves strong electric shocks. Second, if anyone fails to follow a rule (including this one), or speaks out against it, or fails to enforce it, all citizens must unite to kill that person. Suppose these rules were well-enough established by tradition that everyone expected them to be enforced. So you shock yourself for eight hours a day, because you know if you don’t everyone else will kill you, because if they don’t, everyone else will kill them, and so on.

Suppose in the dictatorless dystopia, one guy becomes immortal for some reason. He goes around saying “Maybe we shouldn’t all shock ourselves all the time.” Everyone tries to kill him and fails. But if anybody else starts agreeing with him – “Yeah, that guy has a point!” – then everybody kills that other guy.

The don’t-shock-ists have 100% popular support. And they have charismatic leaders who get their points out well. But they’re still being silenced, and they’re still the losing side. Social censorship isn’t about your support or your leaders. It’s about creating systems of common knowledge that favor your side and handicap your opponents. Censorship = support / common knowledge of support.

Bret Weinstein said of his conflicts with Evergreen State: “I’ve received…quite a bit of support privately from within the college. Publicly, only one other professor has come forward to say he supports my position.” Freddie deBoer writes about how his own conflicts with callout culture have ended the same way: an outpouring of private emails voicing agreement, plus an outpouring of public comments voicing hostility, *sometimes from the same people privately admitting they agree with him*

This provides context for interpreting the *Reason* article’s last paragraph:

> They want not so much any particular policy platform, political idea, or candidate to catch on as for more people to acknowledge that they are right. And that will always be a proposition that winds up making one feel aggrieved, because it’s an impossible one. To the extent that they are spouting marginalized or unpopular ideas, the only way to spread these into the mainstream is to put in the hard work of winning people over.

This is the equivalent of going to communist Czechoslovakia and thinking “Look at all those greengrocers with communist slogans in their shop windows! Clearly communists have won the war of ideas, and anti-communists are just too aggrieved to do the hard work of convincing people”. The other interpretation is that lots of people are already convinced and afraid to say so, and that convincing more people is less productive than building common knowledge of everyone’s convictions (maybe you should hand out blank leaflets). I’m not saying convincing people isn’t good and necessary, just that assessing how convinced people are is harder than it looks.

Here is a story I heard from a friend, which I will alter slightly to protect the innocent. A prestigious psychology professor signed an open letter in which psychologists condemned belief in innate sex differences. My friend knew that this professor believed such differences existed, and asked him why he signed the letter. He said that he expected everyone else in his department would sign it, so it would look really bad if he didn't. My friend asked why he expected everyone else in his department to sign it, and he said "Probably for the same reason I did".

This is the denominator of silencing in a nutshell. I think it's a heck of a lot more relevant to this discussion than how many Patreon followers Sam Harris has, and I'm happy there are people speaking out against it and trying to make common knowledge a little bit more common.

Should Psychiatry Test For Lead More?

Dr. Matthew Dumont treated a 44 year old woman with depression, body dysmorphia, and psychosis. She failed to respond to most of the ordinary treatments, failed to respond to electroconvulsive therapy, and seemed generally untreatable until she mentioned offhandedly that she spent evenings cleaning up after her husband’s half-baked attempts to scrape lead paint off the walls. Blood tests revealed elevated lead levels, the doctor convinced her to be more careful about lead exposure, and even though that didn’t make the depression any better, at least it was a moral victory.

The story continues: Dr. Dumont investigated lead more generally, found that a lot of his most severely affected patients had high lead levels, discovered that his town had a giant, poorly-maintained lead bridge that was making everyone sick, and – well, the rest stops being story about psychiatry and turns into a (barely believable, outrageous) story about politics. Read the whole thing on Siderea’s blog.

Siderea continues by asking: why don’t psychiatrists regularly test for lead?

> Now, in my case, I’m a talk therapist, and worrying about patients maybe being poisoned is not even supposed to be on my radar. I’m supposed to trust the MDs to handle it.
>
> Dumont, however, is just such an MD. And that this was a clinical possibility was almost entirely ignored by his training.

Dumont’s point here is that while “medical science” knows about the psychiatric effects of lead poisoning and carbon disulfide poisoning and other poisons that have psychiatric effects – as evidenced by his quoting from the scientific literature – psychiatry as practiced in the hospitals and clinics behaves as if it knows no such thing. Dumont is arguing that, in fact, he knew no such thing, because his professional training as a psychiatrist did not include it as a fact, or even as a possibility of a fact.

Dumont’s point is that psychiatry, as a practical, clinical branch of medicine, has acted, collectively, as if poisoning is just not a medical problem that comes up in psychiatry. Psychiatry generally did not consider poisoning, whether by lead or any other noxious substance, as a clinical explanation for psychiatric conditions. By which I mean, that when a patient presented with the sorts of symptoms he described, the question was simply never asked, is the patient being poisoned?

Dumont wants you to be shocked and horrified by what was done to those people, yes. He also wants you to be shocked and horrified by this: psychiatry as a profession – in the 1970s, when (I believe) the incidents he relates where happening, in the 1990s, when he wrote it in his book, or in 2000 when a journal on public health decided to publish it – psychiatry as a profession did not ask the question is the patient being poisoned?

And it didn’t ask the question, because clinical psychiatry had other explanations it liked better, to which it had a priori philosophical commitments.

> And that, when you think through what it means for psychiatry, is absolutely chilling.

And:

> I can tell you that, standing here in 2018:
>
> - No mental health clinic I've worked at ever had the facilities for even performing blood draws, nor doing urine testing for anything other than commonly abused intoxicants (alcohol, opioids, amphetamines, etc), and then only the clinics that specialized in substance abuse treatment. The clinic I work for now can't even do urine screens. Psychiatrists' offices, here abouts at least, are not places blood tests are or can be performed, unless they are attached to a general medical practice. Such tests have to be referred out, usually to the patient's PCP's office.
>
> - No psychiatrist has ever asked me to arrange blood draw test from the PCP for anything other than white blood cell count, thyroid panel, or Lithium blood level.
>
> - Though I've seen documentation in patient charts of psychiatrists ordering two of those three tests from PCPs themselves, I've never seen documentation of ordering any other tests. I have literally never seen a psychiatrist order a test for any sort of poison.
>
> - I have never seen any sort of toxicology report for poisons in any of the blood test results I have found in my patients' discharge paperwork from psychiatric hospitalization.
>
> - I have never, in all my case discussions with psychiatrists in-patient and out, or with hospital staff at psychiatric hospitals and hospital departments, ever

heard anyone suggest anything about poisoning be a possibility in our mutual cases. Nobody has ever said anything like, “We don’t want to prescribe anything until the tox report comes back, in case it’s an environmental toxin” or “R/o env tox” or even “We don’t think there’s much chance of an environmental toxin, so we’re not bothering to test for it. It has literally never been mentioned.

- Not even when, due to the suddenness of the onset of psychotic symptoms, psychiatrists were discussing with me the possibility that a patient was intoxicated on some street drug that somehow just wasn’t showing up in his/her urine screens and blood draws.

Maybe it’s not fair for me to generalize from the psychiatrists I’ve worked with. Maybe it’s just that the psychiatrists I’ve worked with – including at MGH and McLean – aren’t representative, being somehow really bad doctors, or poorly educated, and that, contrariwise, normal psychiatrists, basically adequately well-trained psychiatrists, generally do stop to consider poisoning as a cause for severe presenting symptoms, especially when they’ve proved refractory.

I’m not getting that impression though.

I’m not getting that impression from the many interactions I’ve had with psychiatrists and other psychiatric professionals over the last decade, and neither have I been subject to exhortations of what I, as a clinical mental health counselor, should be alert to as evidence of possible poisoning in my patients.

When I was in grad school, it was briefly mentioned that most disorders in the DSM (this was version IV-tr) had a

> "caused by a General Medical Condition" variety, and then it was never spoken of again.
>
> So as far as I can tell, nothing has changed.
>
> This is not merely an incidental failure of instruction on the part of Dumont's med school professors, nor of mine in grad school. This is, at the most charitable, a massive blindspot, of precisely the sort that "scientific" field of endeavor should never have, and it seems to afflict the entire profession.

There's a lot more, and you should read the whole post. Siderea is a great writer and a careful thinker, so when she criticizes my practice I take note. And since I don't think I've ever tested anyone for lead, this is definitely criticizing my practice. What's my justification?

Take a look at some papers like The Emerging Role For Zinc In Depression And Psychosis and Effect Of Zinc Supplementation In Patients With Major Depression: A Randomized Controlled Trial. Done? Looks like there's some pretty good evidence that zinc deficiency is involved in depression somehow, right? Do you think zinc is more or less important than lead? By how much?

Or what about toxoplasma? Seems to be twice as common in depressed people as in controls, and increases suicide risk 50%. Pretty suspicious; should we test all depressed people for toxoplasma? If so, is this more or less important than testing all depressed people for lead? By how much?

And when you've answered that, what about copper? Omega-3/omega-6 ratio? Vitamin D levels? Cortisol? Magnesium balance? The methylation cycle? Mitochondrial function? Inflammation? Covert viral infections? Covert autoimmune disorders? Paraneoplastic syndromes?

Allergies? Light exposure? Circadian rhythm? Selenium? Lithium levels in your local water supply? Insulin resistance? Gut microbiome? PANDAS? FODMAPs? Structural brain abnormalities? And that's not even getting into the psychosocial stuff!

Every one of these has some evidence of being involved in depression. Some have excellent evidence of being robustly involved. Imagine how dumb you would feel if it turned out only 0.01% of cases of depression were lead related, and you spent so much time testing your patient for lead that you never got around to asking about color temperature of their home lighting, or whether they clean their cats' litterbox, or how many dental fillings they have.

(Dental fillings? *Really?* No, not really.)

Why not test *all* the things? Number one, cost. Number two, sticking your patient with more needles than a Trump voodoo doll owned by the DNC. But number three, everybody is weird in a bunch of ways. Have you ever gone to your doctor for labwork, and gotten a piece of paper back with a lot of words like BASOPHILS and BUN-CREATININE RATIO, and probably three or four of them were highlighted in red to indicate they were abnormal, and your doctor looked at it and shrugged and said not to worry about it? That's because everybody is weird in a bunch of ways. Your 30-item depression risk factor panel is going to come back saying you don't have enough selenium and your gut microbiome is off, and your doctor is going to make you spend a month eating nothing but kefir and brazil nuts, and then a month later you'll leave your abusive partner and your depression will mysteriously disappear.

Consider prostate cancer screening. This is like the best-case scenario for universal testing. Prostate cancer is pretty common – about 10% of men are diagnosed with it sometime during their lives. We know who's most at risk – older men. It's potentially pretty bad – nobody wants cancer. The test is easy – the same simple blood draw that tells you if your cholesterol is too high. And yet governing bodies keep recommending that doctors stop screening for prostate cancer (recent guidelines are nevertheless complicated: ask your doctor if PSA screening is right for you). The bodies cite possibility of "overdiagnosis and overtreatment", potential false security by missing some cancers, and studies which show no decrease in mortality. Breast cancer screening organizations keep pushing back the age at which they recommend women start getting mammograms, because the costs outweigh the benefits before then.

Doctors *never* just say "I hear this condition is bad in my field, let's worry about it with everyone who comes through the door". They want really good evidence that it's common enough (and that testing works well enough) that the benefits are worth the risk. Right now for lead, we have no such evidence.

The only study I have ever seen even begin to make the slightest attempt to quantify the role of lead in depression is Bouchard et al. It claims that people in the highest quintile (top 20 percentiles) of lead exposure have twice the depression risk as people in the lowest quintile. Big if true. But I'm skeptical, for several reasons:

1. Lead exposure is heavily linked to poverty – poor people tend to live in the decaying houses and polluted neighborhoods where lead is most common. Poor people are also more likely to get depressed. The study attempted to control for poverty, but this never works. So it's not clear how much of the lead effect they picked up was really a poverty effect.

2. This is an implausibly large effect size. The amount of environmental lead has plummeted over the past thirty years after the removal of leaded gasoline. Since then, the percent of people with elevated lead levels has decreased by a factor of twenty. Some quick calculations suggest that if this study were right, depression rates should have gone down 66% in the past few years. They haven't. Compare this to violent crime, which we have better evidence is lead-related and which *did* decrease by a factor of 2 or 3 over the past few years.

3. I don't entirely understand what they're doing with statistics here. In Table 2, every lead quintile has about the same depression rate. It's only after they apply their model that they find higher-lead people have higher depression. This is sort of a red flag for the kind of thing that might not replicate later on. Nobody else has ever tried to repeat this study and as far as I know it remains the only investigation into the epidemiology of lead and depression.

Aside from this study, we have nothing to guide us. Does lead contribute to 5% of cases of depression? 0.5%? 0.0005%? I'm not sure anyone knows.

If it's 0.0005%, then we're talking about people who work in lead mines, or Dumont's patient who scrapes lead paint off the wall every day. I agree if your patient works in a lead mine and has any problems at all, you should test them for lead poisoning. This is why your doctor asks you what line of work you're in when you go in for an evaluation. She's not trying to make small talk – she's waiting for the one guy who says "I French kiss cockatiels for a living" so she can diagnose him with psittacosis.

I am definitely not going to make every patient take a blood test just so I can catch 0.0005% of people. But should I be more careful in the history here? Specifically ask patients “Do you have any hobbies that put you in contact with lead?”. I’m not sure – right now I have no intuition for whether that’s more or less important than “Do you do anything that might cause zinc deficiency?” or “Do you clean out cat litterboxes?” Usually what I do is I ask broad open-ended questions to start with, and then as the depression proves itself weirder and more treatment-resistant, we gradually go one-by-one down the list of super-rare causes that never happen in real life so I can be sure I’m not missing anything.

What if lead causes more like 5% of depressions? In this case, we’re talking about people with no specific exposure factors. Maybe they just live in an old house, or a bad neighborhood, or got a bad draw in the genetic lottery for whatever systems affect heavy metal removal. So suppose I grudgingly give every patient who comes through the door a blood lead level test. Lots of them come back with lead levels in the top quintile – just like 20% of the general population would. Now what? I say “Your lead level is within the normal range, I have no proof that it’s contributing to your depression at all, but just to be safe you should move to a classier neighborhood”? “Oh, you mean that will cure my depression?” “No, Dr. Dumont didn’t even cure his patient’s depression when he made her stop scraping lead paint off the walls, I’m just saying it probably couldn’t hurt”. I consider it a good day when I can get my patients to take their Lexapro without missing any doses. My ability to get them to move to a nicer neighborhood – most people aren’t living in bad neighborhoods just because they never thought of moving to a better one – is pretty low, especially when I can’t even honestly say I expect it to help that much.

What about chelation therapy, a series of techniques that suck lead out of the body? From Wikipedia: “Chelation therapy must be administered with care as it has a number of possible side effects, including death…various health organizations have confirmed that medical evidence does not support the effectiveness of chelation therapy for any purpose other than the treatment of heavy metal poisoning.” If you chelate everyone who comes through the door with high-average lead levels, there’s no guarantee you will cure any depressions, but you will definitely kill some people.

There’s an old medical maxim: never do any test if you have no plan for acting on the results. What’s my plan for acting on the results of a upper-end-of-normal lead level? I really don’t have one. If it’s an extreme level, the 0.0005% works-in-a-lead-mine type, then I can at least demand the patient find a different line of work. But if I’m just going to be diagnosing 20%+ of the patients who come in the door with upper-end-of-normal lead? Forget it.

I understand Siderea probably isn’t recommending universal blood lead screening. But then I’m having trouble figuring out what she *is* recommending doctors do differently. Ask everyone some history questions to see if they work in lead mines? We try to do that kind of thing already. Have lead poisoning on the top of our minds and ask a lot of really intense history questions to ferret out any possible exposure? Not clearly privileged above doing that for the thirty other rare causes of depression. Just make sure to talk about lead as a cause of depression in medical school? They already do, it’s right between “Prader-Willi Syndrome as cause of obesity” and “Q fever as cause of pneumonia” in the lecture series entitled Extremely Rare Causes Of Common Symptoms Which Ideally You Can Keep In Mind And Have Dr. House Levels Of Diagnostic Genius, But Let’s Be Honest Here, Realistically Over Half Of You Will Prescribe Antibiotics For Viral Infections.

I’m not saying nobody should worry about lead poisoning and depression. I’m just saying those people shouldn’t be front-line clinicians. Some epidemiologist should absolutely be trying to replicate Bouchard et al’s work on the magnitude of the lead-depression correlation. Some guideline-making body should be coming up with a guideline for when doctors should take lead levels, and what results should prompt what kind of action, even if the guideline is just “never worry about this unless somebody works at a lead mine” (which is plausibly the right answer given the current paucity of data, but I’d feel more comfortable if a guideline-making body said so officially). And public health officials should worry a lot about how to decrease lead on a society-wide level (which they’re already doing, albeit for different reasons), since that’s much higher-yield than some random doctor telling a poor person to move to a different house.

But I am not sure the average clinician needs to think about this too much.

(Maybe Siderea already agrees with me here; I can’t tell.)

There are thirty-plus plausible causes of depression that nobody knows enough about to be sure they’re real, estimate their magnitude, or begin to treat. If you look at any one of them too closely, you will come to the conclusion that every psychiatrist in the country is a quack who’s ignoring the evidence right in front of their eyes and willfully blind to the role of [lead/zinc/toxoplasmosis/inflammation/gut microbiome/etc] in order to keep getting the sweet pharma company cash for prescribing Lexapro. It’s not that we don’t know about these things. It’s that we don’t have an action plan. We don’t have a good feel for when to do the tests, what numbers on the tests mean we should do something, what that thing should be, and whether it should work. So we punt the question to the researchers, who already have a backlog of ten million other things they need to be working on.

This isn’t a great state of affairs. But I only know three ways doctors can deal with it, and none of them are very good.

The first is the one I learned at age 3 from my father teaching me to read out of evidence-based-medicine textbooks. You insist that nothing can be admitted into the medical canon unless it has some guideline-making-body’s stamp of approval, which the guideline-making-body will not give until a bunch of randomized controlled trials have validated every step of the model and shown that the proposed solutions definitely work on a large scale with every demographic of patient. Until then we will keep doing the things that *have* met that bar – which is basically giving people Lexapro and telling them to diet and exercise. This path assures you a long and prosperous career as a respected member of the medical establishment.

The second is to become Dr. Oz. You fall in love with anything that has an even slightly plausible mechanism and at least one n = 15 study saying it works. I’m not talking about literal homeopathy here. I’m talking about things where if you ask a biologist whether it works, they just sort of shrug and say “well, it *should*“, and there’s a bunch of respectable research into it. But this is a really low bar, and if it’s the only one you hold yourself to, then you’re going to be the guy telling your patients they need heavy metal tests and vitamin levels and SPECT brain scans and screens for twenty latent infections just because they came in saying they’re tired all the time. This path assures you a lucrative daytime TV show and a side gig selling supplements with your picture on them.

The third is to be a generally respectable doctor with one Big Idea. Like “why aren’t we testing everybody for lead?” or “why don’t we care more about the gut microbiome?”. These people are often really good at what they do, really passionate, and *mostly* within the mainstream. Sometimes they are impressive researcher-crusader-prophets, they get their Big Idea universally adopted, and then they become the next generation of medical orthodoxy. Other times they’re just annoying clinicians who love saying “I see you aren’t even testing for cortisol levels, clearly you have no interest in going beyond Textbook 101 Level” but can’t really explain why this is better than the twenty-nine other things you might consider doing. This path assures you a long bibliography of successful articles in *The Journal Of Medical Hypotheses*.

I love everybody in Group 3, they’re all great people. But the thing is, if I were to believe *everybody* in Group 3, then I would end up as Group 2 – and I don’t have enough time to star on a TV show, so screw that. I think that makes me Group 1 by default, which is good, because otherwise my family would disown me.

“Shouldn’t we be able to use rationality techniques to figure out which of the Group 3 people are right, and move faster than guideline-making bodies“? Well, that’s the dream. But take that route, and you notice you’re wading through ankle-deep skulls. I occasionally flirt with trying this – like every doctor, my practice has a few idiosyncrasies and places where it deviates from the exact textbook solutions. But I would be nervous putting too much trust in my own gut.

This is all context for how to think about questions like “should we test everybody for lead?” or “should we think more about lead?” or “is the psychiatric establishment incompetent for not testing lead more?” The prior on the psychiatric establishment being incompetent is never that low. But the prior on any given alternative being especially fruitful isn’t great either.

[**EDIT:** The only general medical conditions I consistently find worth worrying about in depression (absent some specific reason to worry about another) are hypothyroidism and sleep apnea. I test a lot of people for anemia and various vitamin deficiencies, because the guidelines say so, but I’ve never found them too helpful. Curious if anyone else in the field has different experiences. I recently had one patient obtain a miraculous and lasting cure of his chronic fatigue using nasal steroids (ie it was apparently caused by nasal inflammation from allergies) but nobody ever talks about that and I’m not sure if it was just a fluke.]

Highlights From The Comments On Basic Jobs

These are some of the best comments from Basic Income, Not Basic Jobs: Against Hijacking Utopia. I'm sorry I still haven't gotten a chance to read everything that people have written about it (in particular I need to look more into Scott Sumner's take). Sorry to anyone with good comments I left out.

Aevylmar corrects my claim that Milton Friedman supported a basic income:

> Technically speaking, what Milton Friedman advocated was a negative income tax, which (he thought, and I think) would be much more efficient than basic income – I don't remember if these are his arguments, but the arguments I know for it are that the IRS can administer it with the resources it has without you needing a new bureaucracy, it doesn't have the same distortionary effects that lump sum payment + percentage tax does, and it's probably easier to pass through congress, since it looks as though it costs less and doesn't have the words 'increasing taxes' in it.

And Virbie further explains the differences between UBI and negative income tax:

> The main difference is that discussing it in terms of NIT neatly skips over a lot of the objections that people raise to flat UBIs that are abstractly and mathematically (but not logistically or politically) trivial. Many of these focus on how to get to the new policy position from where we are now. For example, people ask both about how a flat UBI would be funded and why rich people

> should receive a UBI. Given that the tax load to fund a basic income plan would likely fall on the upper percentiles or deciles, a flat UBI + an increase in marginal tax rates works out to a lump sum tax cut for high-earners and a marginal tax increase. Adding negative tax brackets at the bottom of the existing system and modifying top marginal rates is a simpler way to handle this and extends gracefully from the current system instead of having to work awkwardly alongside it.
>
> In the example above, the NIT approach has the logistical advantage of the bureaucracy and systems we already have handling it more easily. And the political advantage of the net cost of the basic income guarantee looking far smaller than for flat UBI, since we're not including the lump sum payments to upper-income people (that are more than offset by their marginal tax increases).

There's some further debate on the (mostly trivial) advantages of NIT or UBI over the other in the rest of the thread.

Tentor describes Germany's experience with a basic-jobs-like program:

> We had/have a similar thing to basic jobs in Germany and it worked about as well as you would expect. Companies could hire workers for 1€/hour and the state would pay social security on top of that. The idea was that long-term unemployed people would find their way back to employment this way, but companies just replaced them with new 1€-workers when their contract was over and reduced fully-paid employment because duh!

> Plus people on social security can be forced to take jobs or education. As a result a lot of our homeless are depressed people who stopped responding to social security demands because that's what caused their depression.
>
> (Links are to German Wikipedia, maybe Google translate helps)

Another German reader adds:

> I agree that it doesn't work as expected in Germany, but I think it it important to point out that not everyone is allowed is to hire workers for 1€. The work has to be neutral to the competition and in the public interest. So people are hired at a lot of public institutions (e.g. schools, universities, cleaning up the city).
>
> Additionally these jobs improved the unemployment statistics at a low cost for the government, as people who are working in these jobs count as employed although most of these jobs are only part time jobs.

Murphy describes the UK experience:

> One likely model for "guaranteed jobs" is the disaster that they tried in the UK for a while.
>
> Basically the government partners with crappy low-skill employers who's owners are buddy buddy with the right ministers and the state provides them with a steady supply of ~~slaves~~ jobseekers.
>
> They then declare it all "Education", in fact pay a premium to the corporate partners for "providing education" in the form of 5 minutes showing someone

who’s already worked shelf stacking jobs how to stack shelves.

The people who love the scheme tend to genuinely believe the fiction about “education” because they tend to be the kind of people who believe that all poor people are thick and can’t learn and really do need 6 weeks to learn how to put a tin of beans on a shelf.

Your manager is abusive? tough luck. You have no rights. if you quit or the supervisor just declare you not to be working hard enough you lose your dole money. Hope you like starvation and death.
So if your manager demands you suck his dick then make sure to bring kneepads to work.

Remarkably employers who suddenly had the option of free labor along with free money from the government leapt at the option so people found themselves fired from positions only to find themselves required to do the same job a few weeks later only this time without pay.

The government was taken to court over it, the court rules it unlawful.

“I don’t think I am above working in shops like Poundland. I now work part-time in a supermarket. It is just that I expect to get paid for working.”

So the UK Parliament passed retrospective legislation to overrule the courts.

Anyone and I mean anyone with a libertarian bone in their body and an ounce of principles should be disgusted by “guaranteed jobs” because it’s thinly disguised slavery and a drive to replace paid work with forced labor.

Herbert Herbertson on the Native American experience:

> I’ve said it before, but beyond the Alaska permanent fund, there’s an area where we could see a TON of extremely varied UBI case studies that I’ve never really seen anyone talking about UBI mention, one where we could see the limits, the pitfalls, and the benefits of a UBI as applied to a population with extremely deep historical poverty, intergenerational trauma, and serious substance abuse issues: Native American tribes with (more or less) successful gaming operations who distribute a portion of profits to their members in the form of a “per capita” payment.
>
> My anecdotal experience is that it’s no panacea, but that it sure as fuck helps–but there’s a lot of potential data out there to move beyond anecdote.

Unirt describes the Finnish experience:

> There has been a universal employment trial in a Finnish town Paltamo, which lasted 3 or 4 years. Apparently it costed the government more than just paying unemployment subsidies, and they found that undesirable.

---

richarddormersvoice describes the Chinese experience:

> the closest thing to a guaranteed work program today is the iron rice bowl in china, which is a clusterfuck.
>
> https://en.wikipedia.org/wiki/Iron_rice_bowl
>
> from what I know/heard:

> a) 50% of state-owned enterprises are operating at a loss, meaning they are inefficient, corrupt, unproductive, and generally terrible places to work. it is stable though because it's guaranteed by the ccp.
>
> b) areas where irb has been liberalized do much better economically, shenzhen, guangdong, chongqing, while much of the northeast has barely developed.
>
> c) even so, getting rid of the irb is difficult because, well, people, firms, politicians are dependent on it. chinese politics is heavy on corruption, soes are especially heavy on corruption.

Doktor Relling describes the Scandinavian experience:

> Scandinavian countries have for some years had something functionally similar to a Universal Basic Job guarantee. We label it "activation policies".
>
> Since we have been doing this for some years (and increasingly the rest of Europe likewise), we have some empirical knowledge of the pros and cons. For those interested, here is a simplified walkthrough of the system (full disclosure: I spend my working life as a health & social policy researcher – and I believe that, on balance, this policy is better than the alternatives.) 1) You start out by introducing a means-tested social assistance scheme that covers everybody – including single males at subway stations shouting GRAAAGH to passers by. 2) You require that those who apply for social assistance, work for the benefit if they are able to work. 3) To find out if they are able to work, the social assistance administration does a work test.

Effects of the system: When the social assistance administration does the work test, it discovers that many long-term social assistance claimants are actually disabled (which was never found out before we introduced the activation requirement plus work test). Hence they qualify for a disability pension instead (somewhat similar to US Special Supplementary Income). In short: This version of a UBJ channels the unemployable GRAAGHs among us to a (more generous and not-means-tested) disability benefit. This takes care of Scott's objection concerning what to do with those whose net "worth" to an employer is negative.

As an aside: organisations for people with disabilities, in particular the youth organisations, like the activation requirement. Their complaint is that the government does not always follow up its job guarantee in practice.

Why not a Universal Basic Income instead? Most of the weaknesses of a UBI have already been pointed out in the discussion (by rahien.din and David Friedman among others). Let me just re-state that many disabilities are really expensive. A UBI will not be sufficient to grant people with severe disabilities a good life. If voters want to provide them with above-minimum tax-financed income, the state simply cannot avoid to burden medical personnell and administrative staff with the difficult and contradictory helper/gatekeeper roles they perform in our present social security systems. It is messy, difficult, and yes there are Type I and II errors, but these problems are unavoidable if voters want to provide people with disabilities with more than what everyone else gets. (And hey this role conflict is difficult but it is not THAT difficult; after all we have been able to live with this role conflict for more than a hundred years.)

Yakimi describes the Nauru experience:

> There are societies where entire populations have been unconditionally emancipated from the necessity of labor. I have seldom heard basic income advocates talk about these precedents, probably because these experiments do nothing to justify their optimism.
>
> The Republic of Nauru has an unemployment rate of 90%. Its people do not work because their incomes are publicly subsidized, mostly by the exploitation of their island’s phosphate deposits, an industry which once provided them with the highest incomes per capita in the world. These people who once lived on a diet of coconut and fish used their sudden influx of wealth to import all the worst excesses of civilization, leaving them with the highest obesity and diabetes rates in the world and a life expectancy of 59.7 years. They’ve nearly exhausted their sources of phosphate, completely destroying the natural beauty of their island in the process, and the people are now physiologically incapable of any existence other than idleness.
>
> We might also look to the banlieues of France, where the youth unemployment rate is over forty percent and the underclass survives, illicit economic activities aside, at the expense of the generous French welfare state. Is there any evidence at all these beneficiaries are grateful to have been freed from drudgery? If anything, their lack of economic stake only seems to aggravate their resentment against a society that is keeping them humiliatingly idle. As recent events remind us, men hate being made to feel superfluous. Nor does there any appear to be evidence that their idleness has enabled the Byrons, Churchills, Von Brauns, et al. among them to

> improve the world with their genius. They are quite capable of setting cars on fire, though.
>
> There are no doubt people, like yourself, who are natural aristocrats, who are very good at finding discipline, purpose, etc. even when freed from the pressures of necessity and would benefit from a stipend. But it is solipsistic to assume that most, or even many, humans can operate functionally when made entirely independent of the disciplinary pressure of having to earn your fill. Posthuman biotrash is a big enough problem already, and basic income can only make it worse.

I certainly don't deny that a lot of ghettos and banlieues contain some very unhappy people. But does work help?

Suppose Alex lives in a ghetto and spends 12 hours a day watching TV and eating Cheetos. Bob lives in the same ghetto, works at a gas station 8 hours a day selling people lotto tickets, then comes home and watches TV and eats Cheetos for 4 hours. Aside from economic arguments about producing value for other people, is Bob's life more meaningful than Alex's? Is it happier? Would you rather be Alex or Bob? Would you rather Alex exist, or Bob exist?

If you want to make the argument for work, you have to argue that it does something other than turning Alex into Bob. That it has some other effect, where Bob gets home from work and says "You know, all that lotto-ticket selling has awoken a spark of something higher in me. Instead of watching TV, I think I'm going to read *Anna Karenina*." Or something. If I'm strawmanning this argument, it's because I don't really know how people expect it to work.

I don’t want to disagree with Yakimi. I don’t want to come out and predict “If we institute UBI, we won’t have ghettos full of Alexes”, and have you point and laugh when I’m proven wrong. I think ghettos full of Alexes is a very likely outcome. But I don’t think that’s *worse* than ghettos full of Bobs. I think it’s just more surprising, more unfamiliar, more of a man-bites-dog style interesting news story that will provoke concern.

Wrong Species writes:

> I only have one small quibble. Amazon is relentless with their employees because they’re so competitive. Jobs guarantee programs would be anticompetitive so we probabaly wouldn’t see anything like that. In fact it would probably be the opposite where there is too little to do and a lot of it is pointless busy work, like in high school.

Yes, this is a good point. Most surveys seem to find job satisfaction is higher in the private sector than the public sector, but I could imagine the opposite being true for the most-exploited kinds of unskilled labor.

And for what it’s worth, here’s a reader who works in an Amazon warehouse commenting to say it’s really not that bad.

Naj on another way things can go wrong:

> It seems to me that many people whose lives suffer due to lack of money are in significant debt and that debt payments are a large fraction of their income. If a UBI is given to everyone, how are people prevented from borrowing $100,000 against it, blowing it quickly, and

> then having $0 income because their UBI is all spent on interest? Solving this problem also seems full of opaque bureaucracies and Kafkaesque rituals. Unless we just ban loaning money with interest for consumptive goods at the same time(a policy I might just favor).

po8crg suggests:

> Bankrupts will still be entitled to UBI. Assuming (not unreasonably) that bankruptcy laws won't allow garnishment of UBI to pay debts, anyone with only UBI as income and more debts than assets could declare bankruptcy, surrender their assets to their creditors, and walk free from their debt.
>
> The risk of loaning to someone with nothing but UBI would be huge, so no lender would lend.
>
> (incidentally, the problem with this may be in the housing market, landlords will be demanding rent up-front rather than in arrears and being very aggressive about evicting people who fall behind)

Simon Sarris (author of the piece I was criticizing) writes:

> The point of something like Basic Jobs is that giving people the option, but not the obligation, may result in better outcomes for some people at the margins. It's not a panacea, it is definitely not a Utopian alternative to the largely Utopian plans of UBI because I do not think any Utopian plan as described is wise. It's a suggestion of mere incrementalism, something to try on top of the hodge-podge of welfare that currently exists. A splint is safer than a spleen removal, as they say.

> In other words, I think you are committing a mistake by comparing your Utopian vision to another Utopian vision (which I do not advocate). I do not think any Utopian vision is good or possible. You can make UBI look better by comparing it to other Utopian ideas, but this is in effect masking the deficiencies of UBI by comparing it to something else unrealistic.
>
> I do not want to give the impression that Basic Jobs would ever accommodate everyone as UBI may intend to do. In the best case I 100% agree its positive effects would be smaller, but its implementation would also be safer. If you have a hard time imagining that, simply imagine "Maybe we should have farm subsidies, but they work more like Japan's or Austria's than what the US does right now."

10240 writes:

> The huge difference between UBI and public works/job guarantee (even if it's busywork) is that you only take a public works job if you can't get a job on the market, and don't have any better option. With UBI, everyone would take it, and many people who can work would quit. This may make a job guarantee at least remotely feasible.
>
> This is the same as the difference between a homeless shelter and a rent subsidy: a homeless shelter keeps one from freezing on the street, but it's pretty shitty, so the only people who choose it are those who really don't have any other option. It's a built-in means test that's much more effective than a conventional means test that can be attached to a rent subsidy.

> As such, a job guarantee may even save money if it replaces unemployment benefits. Of course, implemented this way it’s a right-wing policy (aimed to minimize welfare usage and incentivize work), rather than a left-wing one. (Hungary’s right-wing government has replaced unemployment benefits with public works like this.) It works if the goal is to keep the poor from starving, rather than to give them a decent standard of living.

This is a good and important point.

I was tempted to respond to Sarris that UBI isn’t that much more utopian than BJG. After all, it can be funded by a tax such that rich people overall pay more in extra taxes than they get in UBI, middle-class people pay the same, and poor people get more in UBI than they pay in taxes. Depending on where you set the definition of “poor”, you can ensure that only the very poor/unemployed people who would go for a basic job are really getting any money from UBI. So there’s no reason to think UBI is necessarily broader-scale than BJG.

10240’s point proves me wrong. Because basic jobs are potentially unpleasant, they act as a screening mechanism so that only people who really need them will take them. That means even targeted at the same income level, they would be less universal than UBI (another commenter points out that we could produce the same effect by making people wait in line for eight hours a day to receive their daily UBI check).

I was hoping to be able to wave away the cost issue with “this is equally bad for UBI and BJG”, but I guess I can’t anymore. I am not an expert in this so I don’t have strong opinions, but I would be pretty okay with a Piketty-esque wealth tax, a Georgist land tax, or whatever experts declare to be the least stupid and distortionary tax that mostly falls on the rich. This article (possibly wrong, possibly biased) suggests that some proposals for raising taxes on the rich could produce about $250 billion/year. That's enough to pay the poorest 10% of Americans a $10K/year basic income (ie have a basic income plus tax increases such that they break even around the 10th percentile) even before cutting any welfare programs.

In my ideal system, we would propose some sort of inherently progressive tax at some fixed percent, and say that the basic income was "however much that produces, divided by everybody". That means that as the economy grows, the basic income increases. At the beginning, the basic income might not really be enough to live off of (especially if I got my calculations wrong). As we get more things like robot labor and productivity increases, so does the income. By the time robots are good enough to put lots of people out of work, they're also good enough that X% of what the rich robot-owning capitalists make is quite a lot, and everybody can be comfortable.

Then various Congresspeople can debate at what point the UBI is large enough that we can eliminate various welfare programs. On the one hand, welfare programs can be sticky, so we might worry they would be overly cautious. On the other hand, many Congresspeople are Republicans, so they probably wouldn't be.

Yaleocon on winding down UBI:

> Saying "winding down basic income is easy" assumes we have an Income Czar who can just say "all right, let's wind it down." We wouldn't have that. We have a democracy, and do you really want to be the guy

> running on “everybody gets less money each year”? It’d be like opposing social security, except even more politically impossible. Candidates—at least, the winning ones—will only ever pledge to defend or expand it. (This also probably makes UBI a fiscally unsustainable policy in the long term.)
>
> Once those political incentives are taken into account, I think we should view UBI as an irreversible, and probably unsustainable, change to our economic system. Scott (or any other knowledgeable UBI advocate), do you stand by the assertion that UBI would be easy to end, and if so, why? (I probably prefer the status quo to UBI, for what it’s worth.)

Thegnskald on long-term effects:

> The economic right likes to pretend the distribution problem is solved. The economic left likes to pretend the production problem is solved. A UBI helps alleviate the distribution issue at minimum penalty to production; relative well-being so important, the incentive to work will be as strong as ever.
>
> A major part of the problem with modern society is that there isn’t an economic incentive to cater to the cashless, the perpetually broke, the homeless; this is a service we want performed, but the system cannot enable it, and the system limps on.
>
> A UBI creates such an incentive. In the short-term, we will have some shortages and price increases; in the long term, we will have a new consumer base, and new industry will arrise to support it.

> Inflation is one possible result, and likely unless we make it economic to build such industry. A UBI needs a corresponding decrease in regulation, in order to make it possible to produce low-cost goods for that new money to chase.
>
> And responding to completely different comments: if a state or city wants a higher UBI, so be it. But a major advantage that the UBI offers is to incitivize people to spread out more evenly across the US, reducing population density. Likewise, a major problem with current welfare is that it disincentivizes work, by punishing those who get it (as working costs benefits, resulting in less overall available funds). I don’t think we will have a rising class of jobless vagabonds; I think instead we will have a rising class of gig-economy people, who take short-term work to get money to buy luxuries, while they mostly skate by on the UBI. This is where we are heading anyways, so this is a net improvement, by making such gig work more secure.
>
> I see little harm and much good in a UBI, and expect it to accelerate employment relative to the current system of “You only get your benefits if you don’t work and avoid making yourself too employable”.

Gimmickless is still worried about the rent issue:

> I came of age in a military town. Part of the military benefit package is a stipend, should you qualify to live off-base. I could never find a apartment or single-wide that ever cost less than that stipend. That information gets out, and gets spread. Do people with McJobs cram themselves 3-4 to a trailer to make ends meet? Yes they do.

> I fail to see how landlords will not take UBI into account on what rent they charge. Will there be a price gradient that actually settles out? Possibly. Probably. But it's almost certainly going to end up higher than what rent is now.

This definitely sounds like what would happen in the case of a captive audience in a world with no ability to increase housing stock. If you relax some of those assumptions, I fail to see why rent shouldn't reach a balance between supply and demand the same way other necessities like food, clothing, and gas do.

Dnkndnts doubts the "universal" part of UBI:

> My problem with UBI is that it's virtually guaranteed not to actually be universal. It's going to be "universal" for Good Citizens. We're going to probe and test you for any sort of substance abuse and if we find anything, you're off the program; if you have a criminal record of any sort, you're off the program. The government is not going to sponsor your alcoholism and crack addiction!
>
> First, I think this immediately falls into the kafkaesque nightmare disability currently has: you have to prove to the Bureaucracy that you are, in fact, a Good Citizen in the same way that you currently have to prove to the Bureaucracy that you are, in fact, Actually Disabled.
>
> Second, the people who need UBI the most are precisely those least likely to be labeled Good Citizens. Poverty massively correlates with substance abuse and criminal behavior.
>
> I get that I'm attacking a strawman here in some sense, but I think there's a snowball's chance in hell that UBI

> would actually be universal.

Counterexample: Social Security. As far as I know, every elderly person gets it, whether they’re a good law-abiding citizen or not.

RC Cola writes:

> I’m no economist but it seems to me that this essay under-plays what I thought would have been one of the main features of a job requirement over a free-floating income. For the moment put aside complications like disability and third-party effects and just imagine the core case of a person with no job. Jobs suck, as Scott notes toward the end. But it is in everyone’s interest if everyone who is able to support themselves with jobs do so rather than resorting to compulsory confiscation of others’ money to do so. So we want obtaining such government money to be unpleasant. This isn’t to say that most people on government programs are just lazy. But people do respond to incentives, and there are probably lots of marginal cases where it would be extra hard to support oneself with a job, and an easy, frictionless transition to government support would be an easy call whereas a costly and painful one would end up causing this marginal person to choose to remain in private employment.
>
> so what is the main obvious attractor of a government-granted income without a work obligation? I think it’s not having to have a job. So to a first order of approximation, adding a job requirement does exactly what you would want it to do. It removes the largest incentive to avail oneself of the government program, namely the removal of work obligations during work hours. We want to do this not out of a punitive

> motivation, or even out of a "help the recipient" motivation. We want it out of a deterrent motivation, precisely to deter availing oneself of redistributive programs that by their nature cannot work if too many people opt in. Is that not an important part of the argument here?

Thegnskald writes:

> My objection [to a previous comment] comes down to this: the assumption that UBI is a solution to poverty, rather than a solution to the systemic issues inherent in a bureaucracy-administrated welfare.
>
> This puts me at odds with Scott, who seems to favor a high UBI intended to make work obsolete for most people. I favor a low UBI which makes the bureaucracy obsolete, but isn't intended to pull people out of poverty. My preferred solution is a debit card (or maybe just a fingerprint-enabled system to disable theft) linked to an account with daily accrual of small amounts of money; 10-18 dollars per day per adult, some possibly smaller amount for children. Do away with disability and food stamps and housing assistance and social security, keep Medicaid/Medicare. Pair this with a new classification of minimal legal rental housing that amounts to an updated-fire-code-compliant barracks (a bunk, a locker or chest, and access to a bathroom and washing facilities, probably gender-segregated; family-style housing might amount to a lockable room). Toss in density maximums on the barracks, and proximity limitations, to avoid concentrating poverty. Maybe – maybe – add in a requirement for security guards.
>
> The goal shouldn't be to enable a luxurious life – it should be to enable a very basic one. With daily accrual

> of funds, huxters and con artist won’t find viable targets, and anything more expensive than a daily meal will encourage short-term thrift and the accrual of some very basic financial knowledge.
>
> This is a more basic lifestyle than is afforded by the current system, but removes the barriers to entry and waiting lists that plague us now. I think the “lower middle class” version of UBI is a terrible idea; it should be treated as a safety net, not a replacement for productive enterprise.
>
> This is the incrementalist version of a UBI. Trial it, adjust as necessary.

---

baconbits brings up other concerns:

> What does it mean to have a “basic” income? Surely housing is included, and housing prices vary wildly over different regions. Are residents of Detroit getting enough money to pay for housing in NYC or are residents of NYC getting enough money to pay for housing in Detroit, or do we have a “cost of living adjusted UBI” where some people get enough for a low end car payment and others enough for a monthly subway pass and every other conceivable difference or are we just accepting that a few (tens of) millions of people are going to not be getting a basic level of income at all while a few (tens of) millions of people are getting well over their basic level?
>
> While we are on the subject of the disabled, well the disabled have extra health costs… are getting more in terms of UBI? Long story short as soon as UBI is introduced it will be noticed that a great many people cannot afford their health insurance payments on their

> UBI and there will be cries for a nationalized health insurance on top of the UBI.
>
> What about children and married couples? How are we balancing UBI payments to families without seriously screwing up incentives there? And immigrants? And families of immigrants?
>
> The short answer is that right after you cut the Gordian you are going to pick up the slashed pieces of rope and attempt to retie them together to hold the system in place.
>
> The optimism that there is a simple solution to an enormous issue is overwhelming.

Ninety-Three discusses risk to private industry:

> You’re being unfairly rosy towards basic income by not mentioning that it too could destroy private industry. Imagine you institute basic income and most of the McDonalds workers quit. McDonalds tries to invent robots, but can’t because robots are hard. So they raise wages in order to attract more workers. In order to pay for their increased wages, they raise prices. The market informs McDonalds that people don’t want to pay more for their fast food, and McDonalds goes bankrupt because their business model was only profitable with $8/hour wages. Higher-end restaraunts will still exist, but your basic income scheme just destroyed the entire “cheap fast food” industry.”

This seems little different from minimum wage laws. Some studies suggest that $15/hour minimum wage laws don’t seem to hurt restaurants. Others do find some negative effect, but the effect is far from catastrophic and restaurants continue to exist.

If basic income is so high that nobody will work at $15/hour ($30/hour? $45/hour?) and private industries collapse, then we must have set the basic income too high. This would be a disaster, but no more than setting the minimum wage too high would be.

David Friedman adds: “From a little googling, labor is about 20% of the cost of McDonalds franchisees. Double wages and, if they pay those wages instead of substituting more skilled labor or machinery, and prices go up by about 20%.”

From Nicholas Weininger:

> Scott, you mention aristocrats as a group who seem to flourish without needing to work for a living, and cherrypick a lot of great examples, but surely you’re aware of the phenomenon of corrupt wastrel layabout aristocrats too. It’s at least perceived as common enough that today, parents who have enough money that their kids don’t ever have to work typically spend a good deal of time and effort devising constraints on the kids’ ability to access their inheritances so that they don’t become corrupt wastrel layabouts. Do you think the perception of commonness is incorrect? Are the parents worried for no reason? Note too, as other commenters have, that these parents are much more capable of instilling responsibility and work ethic in their kids than the typical parents of those who might depend on a large UBI.
>
> This is one of many reasons why I think we should start with a very small, Alaska-fund sort of level of UBI, plus a child allowance for primary caregivers of young kids,

> and see how that goes for awhile before taking up the question of an increased UBI level.

Martin Freedman refers us to previous work on the subject which I should get around to reading:

> Hello, long time lurker first time poster here.
>
> This is an interesting post but it seems to miss the boat on the whole Jobs Guarantee debate and analysis that has been going on for many years. This is partly due to your post being a response to another interesting article by Simon Saris, which also, but less so, misses the boat on this debate. Still he does raise points that you sort respond to as if they were never raised e.g over disability and he emphasized no removal of those benefits.
>
> I also read as much as I could of the 100+ comment stream and only two commentators correctly identified the real issues : ShamblerBishop and userfriendlyy.
>
> Anyway a more substantive issue is the body of work and analysis done on the JG done by many economists over the years. You mentioned economists who had argued for a CBI or UBI where you included Milton Friedman who really argued for a Negative Income Tax which is not the same BTW (as others have noted).
>
> However why did you not mention those economists who have argued for a JG? Where were Keynes’ On-the-spot employment, Minsky’s Employer of Last Resort, Mitchell’s Buffer Stock Employment, Mosler’s Transition Jobs and, in general, collectively named by the Modern Monetary Theory economists called the Job Guarantee?

Simon’s post using the non-standard term “Basic Jobs” – which he is entitled to do – is somewhat indicative that this is an unorthodox presentation of these ideas. He (and, for that matter, I) are unknown bloggers on this topic and, regardless, stand or fall on the quality of arguments and analysis. Whilst writing for a particular audience might have been a motivation for him, for whatever reason he omitted the critical, IMV, macroeconomic basis which is particularly important both for the JG and for a more complete evaluative comparison of the JG to BI.

If you really want to do this topic justice I humbly suggest you look to the main economist out of the MMT group who has specifically focused on this topic. (Of course, the others – mentioned above or not – have researched this aspect too, this is only my recommendation). This is Pavlina Tcherneva who has written for a range of audiences from the interested lay person to the mainstream economist academics. You could start with her Job Guarantee Faq or her team at Bard

Suffice to say all your objections have long been answered. That does not mean you agree with the arguments in those answers, of course, but a clearer discussion should start with those answers not write as if these have never been considered.

Michael Handy says:

Umm, correct me if I am wrong, but didn’t we try a Basic Jobs program back in the 19th century in England? Namely, the Workhouses, carefully calibrated to make any private sphere job heaven by comparison.

> I feel that a program that literally uses the first chapters of Oliver Twist as a technical manual might not need such a comprehensive debunking, but I'm glad one exists.

On the one hand, this is unfair. The workhouse was nothing like existing basic jobs guarantees – it was a combination housing program / food program / jobs program that poor people were in many cases forbidden by law to leave. This is totally unlike proposals for a $15/hour voluntary job not linked to housing or food.

On the other, what I find most interesting about workhouses is that most sources suggest they weren't profitable. Even offering workers the most miserable conditions and paying them no wages, they still failed to produce anything that sold well and had to be subsidized by taxpayers. This reinforces my concern that basic jobs are not going to be able to produce as much as people think.

Watchman writes:

> My issue here, and this comes from someone convinced of the futility of the Keynsian approach embodied by basic jobs (the left wing are clearly bored of re-hashing the bad ideas of the 1970s and are going back to the 30s for their inspiration now…), is that basic jobs appear to be much safer for a functioning democracy than basic income. Democracy is ill served if leaders are able to use their power to effectively bribe voters: any study of the working of machine politics in US cities, or of the small electorates of English rotten boroughs, will reveal this. Basic income is a tool that in the right hands could be used to bribe voters by promising an increase in the income; in the US at least I find it easy to imagine populist figures on both sides of the political divide

> promising voters higher basic income. Democracy requires a certain restraint from voters and basic income would be a potential danger to this by creating a clear incentive for voters to focus on self-interest in their decision making, to a degree not currently seen (mind you, it is possible I'm just repeating arguments made when income tax was introduced…).

I agree that income tax is worth considering. The "politicians will bribe voters with their own money" thing is plausible, but *how come there are still taxes*? It would be pretty easy to run the federal government without making the bottom 50% of the wage distribution pay taxes at all; why don't we? I think the answer is something like "to maintain some fiction that everyone must contribute equally", but that makes the bribe-voters-with-their-own money strategy look pretty powerless, doesn't it?

For that matter, free universal health care is an example of bribing voters with their own money – how come it keeps failing? So is universal college – how come no one except Bernie Sanders even pushes it? For all their flaws – and they have many – the average American voter seems remarkably bribery-resistant.

Liz writes:

> Extreme example to make the point: Take a person with Down's syndrome and give them a job (like washing dishes) which they can perform, and see how they respond to it. They're far happier doing that as opposed to sitting at home being fed and "entertained". However unmeaningful you might believe that work to be, it might be very meaningful to them. I'm sure there are a lot of people in the service industry ("would you like fries with that?" et al) who DO INDEED make a positive

> difference in people’s lives. Even a smile can do that for a frustrated, stressed out person….a smile that makes one feel they aren’t alone. A kind gesture, some simple random act of kindness of any sort can actually restore ones’ faith in humanity. So I happen to think the “fries with that?” work is meaningful. It might (depending on circumstances and how it is done, of course) be even more meaningful than far higher paid and higher skilled (on paper) labor.

Many people say they feel that they wouldn’t personally be able to handle not working very well. For example, Joyously:

> The reason I have always been basic-income-skeptical is because of myself. If I had a basic income the same as my grad stipend, I could *definitely very-much* see myself mostly just playing a lot of video games. There’s a *possibility* I’d finally finish my novel. But it would be mostly video games.
>
> And I’d be happy–but the thing is I *can* do other things that fill a slot society needs. (And I am a privileged educated person from a privileged educated background who doesn’t really *need* this assistance.) So shouldn’t I?

Yes. I think you would do great as one of the vast majority of people who would continue to work even after society had a basic income. I’m not in favor of preventing people from working. I myself would probably work even if offered a basic income. I think it’s great if people don’t have to work but want to anyway. I’m not sure how to get this point across more strongly than I already have.

Other people say that obviously everybody would quit if a Basic Income were available. Aphyer writes:

> I am a programmer. I have a really good job. I quite enjoy it, I get paid well, I like my coworkers. I can listen to my favorite music over headphones while working. When someone does something stupid my coworkers and I enjoy ourselves laughing about it. I am probably somewhere north of 95th percentile job satisfaction in the country.
>
> If you offered me enough money to support myself indefinitely while staying at home playing and designing increasingly complicated computer games and board games…well…I don't think I would take it. Probably. But I would be very tempted.
>
> This seems to suggest that somewhere around 90 percent of people would quit their jobs. And maybe in the Glorious Robot Future that would not be a problem. However, one thing that several of Scott's articles on this topic seem to have missed is that we are not actually in the Glorious Robot Future yet. Yes, once you get to a position where our robot armies can do everything we want, a basic income guarantee is probably the best way to convert this into a high standard of living for many people. But we are not in that position. We are not close to that position. Right now people's jobs are actually adding value that will be lost if they quit. And implementing a basic income guarantee now feels like it would just obviously be a disaster.
>
> (None of this should be taken as support of a basic jobs program, which sounds even more obviously disastrous

> and makes me want to exhume Joseph McCarthy and turn him loose on everyone suggesting it.)

Ozy brings up that there are various ways that skilled workers can work part-time to make $10K per year (the easiest is to work a $100K job one year in ten). Since almost nobody does that, it seems unlikely that these people would really quit their job in exchange for basic income.

I’m glad Ozy showed up, because I used to think the same thing as Aphyer, and Ozy reminded *me* that *I* wasn’t taking any of the opportunities to work much less in exchange for much less money either. I wonder if this is just a universal bias, where people feel like they would definitely prefer more free time to working more, but then work more anyway.

---

ec429 defends the much-maligned but ever-popular position of “being angry that other people might get by without having to work”:

> A lot of people have commented above on the incentive problems etc. with both UBI and UBJ, so I’ll not go into that. But I’ve noticed a bit of a thread of “why are opponents of UBI so determined that no-one ever gets a free ride?” Since this generalises to other welfare/disability/secondary distribution programs, I think my viewpoint here may be relevant. Warning: raw and emotional, rather than cool and rational. But maybe it’ll help people understand why I, at least, am so implacably opposed to redistribution. (Also it drifts a bit off-topic by the end.)
>
> Stipulated: I’m a nerd, bookish, aspergiac. (Also, technically not fully-abled: I have a sleep disorder that prevents me working full-time.)

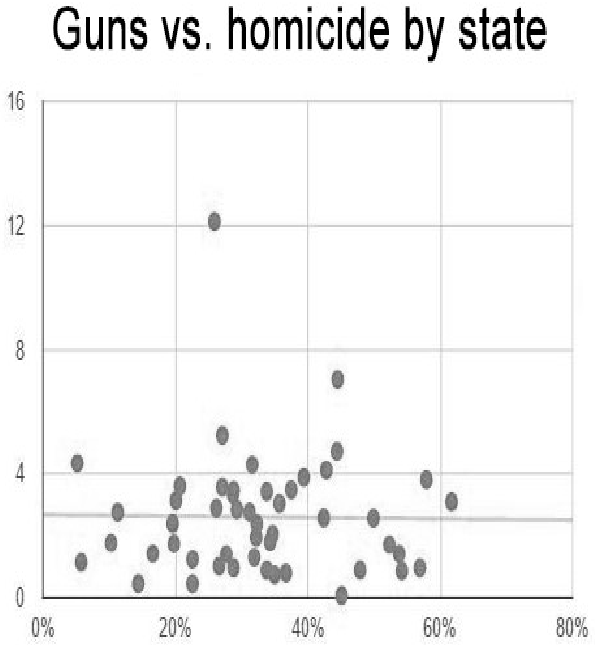
> Between the ages of (let’s say) 8 and 18, I was bullied a lot; this was made bearable by the knowledge that in adulthood, I’d be an affluent knowledge-worker and the people bullying me were idiots who would be stuck in retail or manual labour, and that it would utterly serve the b—ers right.
>
> And now you expect me to give up some large slice of the product of my work (when you add up income tax, national insurance*, VAT** on the goods I buy, push effects from taxes supposedly levied upon ‘business’, etc., the total tax bite is probably over 50%) to give those same b—ers welfare cheques, and then to give them more when it turns out they spent the first lot on pot and xboxes, because somehow it’s not acceptable for people who make stupid decisions to starve.
>
> So no. If you want _my_ money to fund _your_ life, you had d-mn well better prove that you’re trying to better yourself and not just suffering the consequences of your own stupidity. Ideally, let me make that decision myself through private charity, rather than forcibly taking the decision away from me and giving it to some unaccountable bureaucrat. (If nothing else, I could hardly be _more_ Kafkaesque than the disability-scheme bureaucrats.

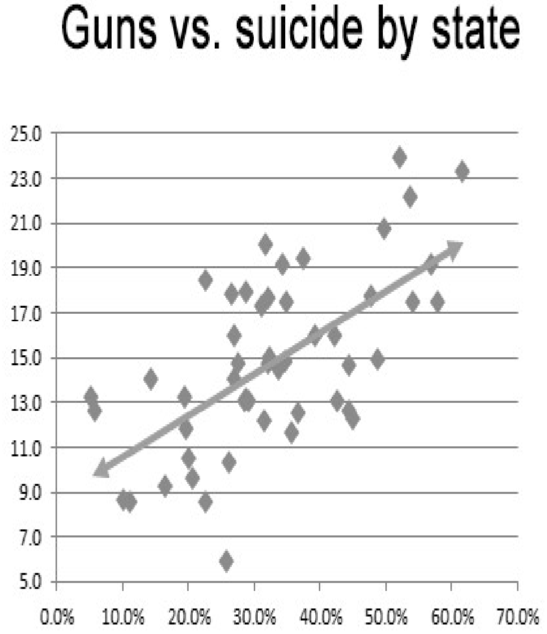
Good news! I hear that basic income will sap meaning and community from people’s lives. So all those bullies will be living unhappy lives without any purpose, and you’ll still have the last laugh!

In Search Of Missing US Suicides

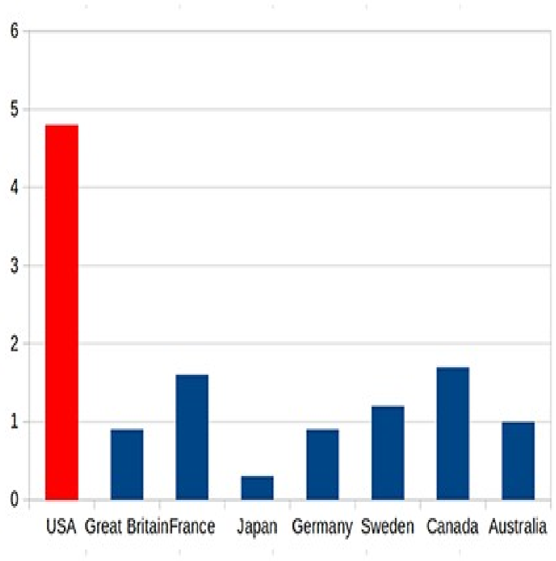
[Content warning: suicide. Thanks to someone on Twitter I forget for alerting me to this question]

Among US states, there’s a clear relationship between gun ownership rates and suicide rates, but not between gun ownership rates and homicide rates:

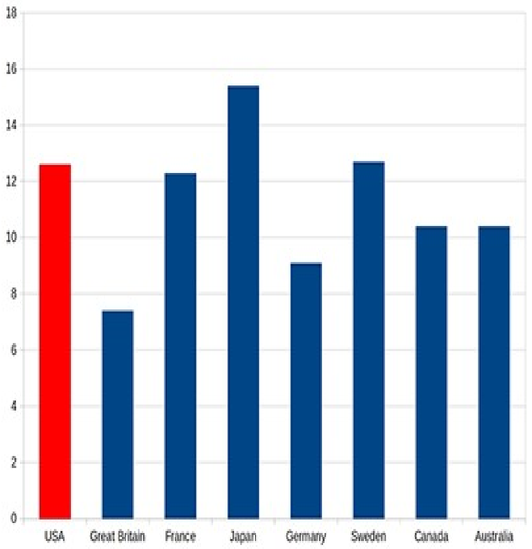




You might conclude guns increase suicides but not homicides. Then you might predict that the gun-loving US would be an international outlier in suicides but not homicides. In fact, it’s the opposite:



Countries by homicide rate



Countries by suicide rate

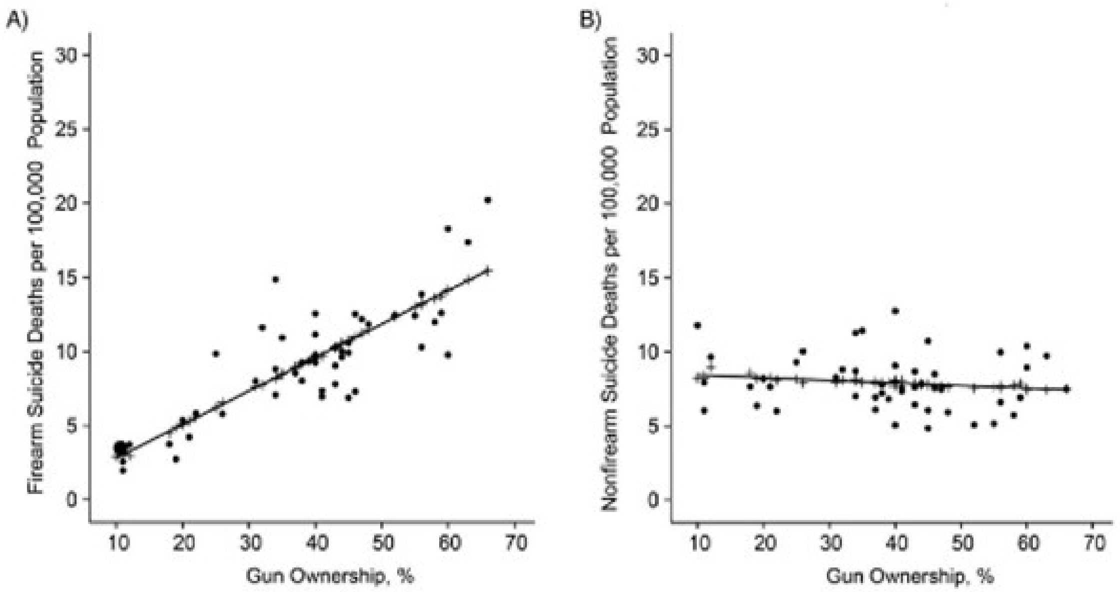
Why should this be?

We’ve already discussed why US homicide rates are so high. But why isn’t the suicide rate elevated?

One possibility: suicide methods are fungible. If guns are easily available, you might use a gun; if not, you might overdose, hang yourself, or jump off a bridge. So getting rid of one suicide method or another doesn’t do much.

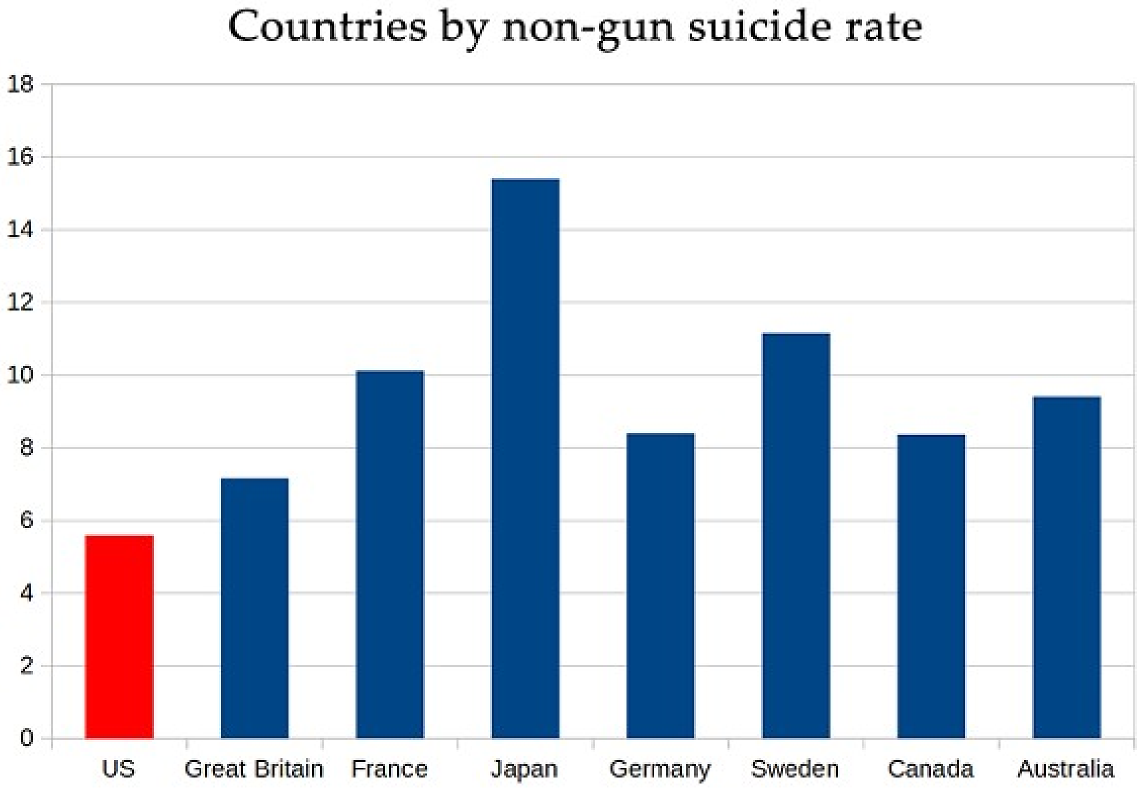
This sounds plausible, but it’s the opposite of scientific consensus on the subject. See for example Controlling Access To Suicide Means, which says that “restrictions of access to common means of suicide has lead to lower overall suicide rates, particularly regarding suicide by firearms in USA, detoxification of domestic and motor vehicle gas in England and other countries, toxic pesticides in rural areas, barriers at jumping sites and hanging…” This is particularly brought up in the context of US gun control – see eg Suicide, Guns, and Public Policy, which describes “strong empirical evidence that restriction of access to firearms reduces suicides”.

The state-level data from above support this view – taking guns away from a state *does* decrease its suicide rate. And then there’s this graph, from Armed With Reason:



…which shows that adding more guns to a state does not decrease its nonfirearm suicide rate.

But if suicide methods aren’t fungible, then why doesn’t the US have higher suicide rates? Here’s another way of asking this question:



The US has fewer nongun suicides than anywhere else. The seemingly obvious explanation is that guns are so common that everyone who wants to commit suicide is using guns, decreasing the non-gun rate. But that contradicts all the nonfungibility evidence above. So the other possibility is that the US ought to have an very low suicide rate, and it's just all our guns that are bringing us back up to average.

Of all US states, Massachusetts, New Jersey, and Hawaii have the fewest guns. Unsurprisingly, suicides in these states are less likely than average to be committed with firearms. In MA, the rate is 22%; in NJ 24%; in HI, 20%. Their suicide rates are 8.8, 7.2, and 12.1, respectively. Hawaii has an unusual ethnic composition – 40% Asian and 20% Native Hawaiian, both groups with high suicide rates (see eg the suicide rate for Japan above). So it might be worth taking Massachusetts and New Jersey as examples to look at in more detail.

Either state, if it were independent, would be among the lowest-suicide-rate developed nations. And both *still* have more guns than our comparison countries. If we did a really simple linear extrapolation from New Jersey-level gun control to imagine a state where firearms were as restricted as in Britain, we would expect it to have a suicide rate of around 5 or 6 – which is around the current level of non-gun US suicides. This is much lower than any of the large comparison countries in the graph above, but there are two developed countries currently around this level – Italy and Israel. I think it makes sense to suppose that the US might have a low Italy/Israel-style base rate of suicides.

For one thing, it's unusually religious for a developed country. Religion is one of the strongest protective factors against suicide. This also seems like a good explanation for Italy and Israel.

For another, it's culturally similar to Britain, which also has a low suicide rate somewhere in the 7s. Other British colonies don't seem to have kept this effect – Australia and Canada are both higher – but maybe the US did.

And for another, it's unusually ethnically diverse. Blacks and Hispanics have only about half the suicide rate of whites; which means you would expect the US to be less suicidal than Europe. I previously believed this was because whites had more guns, but this doesn't seem to be true: Riddell et al find that whites have higher non-firearm suicide rates too. So this could be an additional factor driving US rates down.

(another conclusion from the graphs above: US whites – who have most of the guns – *do* have an anomalously high suicide rate compared to other countries)

A confounding factor – the US has lots of different cultures, and Massachusetts and New Jersey represent only one of them. But if anything I would expect Southern and Midwestern culture, which are more religious, to have a lower base suicide rate; the South also has a lower percent white, another reason to expect their rate to be lower. And there is no evidence of these states having a higher non-firearm suicide rate, which we might expect if they were unusually suicidal.

So I think the simplest explanation is true. A gun-free US would have one of the lowest suicide rates in the developed world, maybe 5 or 6 people per hundred thousand. The US' average-seeming suicide rate is an artifact caused by combining the low base with the distorting effects of high gun availability. The lack of a relative suicide crisis in the US doesn't indicate that easy firearm access isn't causing thousands of preventable suicides per year.

This is maybe not the most pressing question we're facing right now, but I take it as a warning against gotcha-style debating. A simple bar graph comparing national suicide and homicide rates would be a compelling, elegant, and easily-digested argument that guns increased homicides but not suicides. It would also be totally wrong.

[**EDIT:** Commenters point out this paper by Alex Tabarrok on how there is some, but less-than-perfect, substitutability of suicide methods.]

HPPD And The Specter Of Permanent Side Effects

I recently worked with a man who took LSD once in college and never stopped hallucinating. It’s been ten years now and it’s still going. We can control it with medication, but take the meds away and it starts right back up again.

This is a real disease – hallucinogen persisting perception disorder. Most descriptions of the condition emphasize that it’s just some of the visual effects and doesn’t involve distorted reality perception. I’m not sure I believe this – my patient has some weird thoughts sometimes, and 65% of HPPD patient have panic attacks related to their symptoms. Maybe if you can see the walls bubbling, you’re going to be having a bad time whether you believe it’s “really true” or not.

Estimates of prevalence vary. It seems more common on LSD and synthetic cannabinoids, less common (maybe entirely absent) on psilocybin and peyote. Some people say about 1-4% of LSD users will get some form of this, which seems shockingly high to me – why don’t we hear about this more often? If I were a drug warrior or DARE instructor, I would never shut up about this. But if most people just get some mild visual issues – by all accounts the most common form of the condition – maybe they never tell anybody. Maybe 1-4% of people who have tried LSD are walking around with slightly distorted perception all the time.

There’s a lot to say about this from an epidemiological or cultural perspective. But I want to talk about the pharmacology. How can this happen? Why should a drug with a half-life of a few hours have permanent effects on your psyche?

It can’t be that the LSD sticks around. That doesn’t make metabolic sense. And a study discussed here using radio-labeled LSD definitively finds that although a few molecules might stay in the body up to a week or so, there’s no reason to think the drug can last longer than this. I like this study, both for its elegant design and because it implies that somewhere someone got a consent form saying “we’re going to give you radioactive LSD” and thought “sure, why not?”

But then why *does* it have permanent effects? I know very few other situations where this happens, aside from obvious stuff like “it gives you a stroke and then you’re permanently minus one lobe of your brain”. The only other open-and-shut case 100% accepted by every textbook is a movement disorder called tardive dyskinesia. If you take too many antipsychotics for too long, you can get involuntary tremors and gyrations that never go away, even off the antipsychotic. Although traditionally associated with very-long-term antipsychotic use, in a few very rare cases you can get it from a single dose. On the other hand, most people can take antipsychotics for decades without developing any problems.

Some other possibilities are controversial but plausible. The sexual side effects of SSRIs almost always stop within a few months of stopping the medication, but a few people have reported cases where they can last years or decades. Psychedelics may permanently increase openness and hypnotizability, though it’s unclear if this is biochemical or just that drug trips are a life-changing experience – see my discussion here for more. Also, for every drug that has a mild week-long withdrawal syndrome in the average population, you can find a handful of people who claim to have had a five-year protracted nightmare of withdrawal symptoms that never go away.

So, again, how does this happen?

Every discussion of HPPD etiology I've seen is speculative and admits it doesn't know what it's talking about. Also, most of them are in gated papers I can't access. But a few papers seem to gesture at a theory where LSD kills an undetectably small number of very important neurons. Hermle et al talk about "the excitotoxic destruction of inhibitory interneurons that carry serotonergic and GABAergic receptors on their cell bodies and terminals, respectively". Martinotti seems to be drawing from the same inaccessible source in mentioning "an LSD-generated intense current that may determine the destruction or dysfunction of cortical serotonergic inhibitory interneurons with gamma-Aminobutyric acid (GABAergic) outputs, implicated in sensory filtering mechanisms of unnecessary stimuli".

This would require some extra work to explain the coincidence of why the effects of HPPD are so similar to the effects of an LSD trip itself. In particular, if we're talking excitotoxicity, shouldn't the neurons be stimulated (ie more active) in the tripper, but dead (ie less active) in the HPPD patient? Maybe the tripper's neurons are just so overwhelmed that they temporarily stop working? Or maybe you could interpret the comments above to be about LSD exciting some base population of neurons, the relevant inhibitory neurons having to work impossibly hard to inhibit them, and then the *inhibitory* neurons die of exhaustion/excitotoxicity.

Against cell death based explanations, some people seem to recover from HPPD after a while. But this could just be the same kind of brain plasticity that eventually lets people recover from strokes that kill off whole brain regions. The body is usually pretty good at routing around damage if you give it long enough.

What about tardive dyskinesia? When I was in medical school, I was told that the drugs “permanently hypersensitized dopamine receptors”, which is kind of a cop-out – why do they permanently hypersensitize receptors? How come all the other drugs don’t permanently hypersensitize the receptors they antagonize? Apparently now the story is more nuanced. From here:

> The pathophysiology of TD is complex and remains unclear. Multiple models have been proposed to explain this unpleasant and sometimes disabling side-effect. One of the first widespread and popular explanations was the theory of dopamine-receptor hypersensitivity. It was suggested in 1970; however, it cannot completely explain the clinical findings, because TS does not generally appear among all dopamine receptor-blocking drugs users.
>
> To date, several neurochemical hypotheses have been proposed for the explanation of TD development. These theories include: (i) a disturbed balance between dopamine and cholinergic systems; (ii) noradrenergic dysfunction; (iii) dysfunctions of striatonigral, γ-aminobutyric acid (GABA)ergic neurons; and (iv) excitotoxicity. Recently, the role of oxidative stress and structural abnormality in the pathophysiology of TD has gained impetus. Induction of free radicals by neuroleptic drugs leading to oxidative stress and resultant structural abnormality could be the key factor in the pathogenesis of TD. The studies by Lerner et al. and Libov et al. support the neurotoxicity hypothesis. This hypothesis has also been supported by reports that chronic neuroleptic treatment increases free radical production

> and causes structural damage. In 2005, Tan et al. reported that a brain-derived neurotrophic factor appears to exert a protective effect in the nervous system against TD in patients with schizophrenia. There is solid evidence of a genetic predisposition to TD. A study performed by Souza et al. suggests that GSK-3B polymorphism may play a role in the genetic vulnerability to TD manifestation in individuals with schizophrenia.

There also seems to be some sort of role of acetylcholine:

> Several studies in animals have reported that cholinergic cells (or the marker enzyme choline acetyl transferase) in the striatum are lost or reduced in amount after prolonged regimes of haloperidol and fluphenazine (49,50). Recently, Grimm and others showed that prolonged haloperidol treatment in rats led to cholinergic cell loss in the specific areas of the striatum related to oral movements (51). This result may provide an animal model to explain why TD in humans is most commonly a motor disorder of orofacial musculature. Proton magnetic resonance spectroscopy provides supporting evidence for the cholinergic hypothesis. This method allows quantification of choline, the precursor of acetylcholine, in specific brain structures. Choline reuptake leads to the accumulation of choline in cholinergic neurons before its conversion to the transmitter; an excess of choline in brain tissue will signify a loss of cholinergic neurons. Using this method, investigators have shown that, in schizophrenia, choline levels in the basal ganglia are greater than normal (52). Ando and others produced further results with this method (53), implying that choline levels in the lenticular nucleus are higher in schizophrenia patients with TD than in those without the syndrome.

> Apart from such methods for assessing cholinergic processes in the striatum, clinical trials with cholinergic agents in patients with TD could provide indirect evidence related to the cholinergic hypothesis (44). Caroff and colleagues showed that the anticholinesterase donepezil was effective against the symptoms of TD (54,55). Since choline, the precursor of acetylcholine, was not effective, Caroff and others regarded their evidence as support for the hypothesis of Miller and Chouinard. However, a recent metaanalysis concluded that trials of cholinergic agents in the treatment of TD conducted to date have insufficient statistical power to reach a firm conclusion about the drugs' effectiveness (56). This area of research may be clarified when cholinergic agents effective against specific muscarinic receptors are tested in patients with TD.

So apparently it's a conflict between a receptor hypersensitivity hypothesis and a killing-off-interneurons hypothesis that resembles some of the work around HPPD?

I'm biased in favor of killing-off-neurons hypotheses because they're comfortable and they make sense. Of course if a drug kills something, it's going to permanently impair function. This makes the idea of "drugs with permanent side effects" a little bit less scary, restores us to the "some medications cause strokes and then you're screwed" realm of everyday life.

The alternative is thinking of the body as a chaotic system which settles into various attractors. Take the wrong drug and you can push yourself into a different attractor state, which will persist until something shifts it. This definitely seems true of some things, and is one of the ways I think about depressive episodes – which can last months or years, and which can be precipitated by some sort of obvious stressor (getting fired, breaking up, being bullied) but last long after the stressor is gone. If this explains permanent drug side effects, it seems somehow scarier to me than the other option. It’s not just that you have to make sure not to accidentally kill any cells. It’s more that nobody has any idea what the underlying mechanisms look like, anything can happen, and you just have to hope you don’t screw up.

The GATTACA Trilogy

[Few people realize that the 1997 cult hit GATTACA was actually just the first film in a three-movie trilogy. The final two movies, directed by the legendary Moira LeQuivalence, were flops which only stayed in theaters a few weeks and have since become almost impossible to find. In the interest of making them available to the general public, I've written summaries of some key scenes below. Thanks to user Begferdeth from the subreddit for the idea.]

GATTACA II: EPI-GATTACA

"Congratulations, Vincent", said the supervisor, eyes never looking up from his clipboard. "You passed them all. The orbital mechanics test. The flight simulator. All the fitness tests. More than passed. Some of the highest scores we've ever seen, frankly. You're going to be an astronaut."

Vincent's heart leapt in his chest.

"Pending, of course, the results of the final test. But this will be easy. I'm sure a fine specimen like you will have no trouble."

"The…the final test, sir?"

"Well, you know how things are. We want to make sure we get only the healthiest, most on-point individuals for our program. We used to do genetic testing, make sure that people's DNA was pre-selected for success. But after the incident with the Gattaca Corporation and that movie they made about the whole thing, public opinion just wasn't on board, and Congress nixed the whole enterprise. Things were really touch-and-go for a while, but then we came up with a suitably non-invasive replacement. Epigenetics!”

“Epi…genetics?” asked Vincent. He hoped he wasn’t sounding too implausibly naive – he had, after all, just aced a whole battery of science tests. But surely there were some brilliant astronomers who didn’t know anything about biology. He would pretend to be one of those.

The supervisor raised an eyebrow, but he went on. “Yes, epigenetics. According to studies, stressful experiences – anything from starvation to social marginalization – change the methylation pattern of your genes. And not just your genes. Some people say that these methylation patterns can transfer to your children, and your children’s children, and so on, setting them back in life before they’re even born. Of course, it would be illegal for us to take a sample and check your methylation directly – but who needs that! In this day and age, everybody’s left a trail online. We can just check your ancestors’ life experiences directly, and come up with a projection of your methylation profile good enough to predict everything from whether you’ll have a heart attack to whether you’ll choke under pressure at a crucial moment. I’ll just need to see your genealogy, so we can run it through this computer here…you did bring it like we asked you, right? Of course you did! A superior individual like you, probably no major family traumas going back five, six generations – I bet you’ve got it all ready for me.”

Vincent reached into his briefcase, took out a slim blue binder. *Here goes nothing,* he thought.

Two weeks earlier, he had thrown a wad of cash down on a granite table in a downtown apartment.

The man across from him leaned forward in his wheelchair, extended a trembling arm to slowly take the cash.

“You’re, uh, sure you want to do this?” asked Vincent, suddenly feeling a pang of conscience.

“Yes,” rasped Jerome. “You’re young. You still have your whole life ahead of you. You can still make something of yourself. Me, I’m done.”

“Look,” said Vincent, despite his better judgment, “just because you’re partly paralyzed doesn’t mean you can’t… do anything, really. Write a book. Travel the world.”

“You want to know how this happened?” Jerome asked. “I did it to myself. I used to be the best swimmer in the state, maybe in the country. I was going to the Olympics. With a spotless family history like mine, there were no limits. Then it happened. I had a fluke defeat. Lost the Olympics qualifier by a hairbreadth to a guy I’d beaten twenty times before. After that, why go on?”

“But…couldn’t you have just tried again four years later?”

“You still don’t get it. It wasn’t the loss. That loss *stressed me out*, Vincent. It made me *feel bad about myself*. I experienced *lowering of my social status*. With my methylation profile that screwed, it wasn’t just my own body I had ruined. It was my future children too. How was I ever going to marry when I would have to look my wife in the eyes and tell her our kids would be epigenetically tainted forever? So I did the only thing I could. I threw myself in front of a car. Couldn’t even kill myself properly, that’s what happens when your methylation profile is ruined. So here I am.”

“I’m so sorry,” said Vincent, who was starting to regret ever having come. “We can help you. We can find some way to…”

“No,” said Jerome, and he put the cash in his pocket. “What’s done is done.” He took a slim blue binder, slid it over the table to Vincent. “My geneaology. Absolutely perfect. Not a single microaggression against any of my ancestors since the Mayflower.”

Vincent considered saying something, but finally just nodded. “I won’t let you down. I’ll use the gift you’ve given me in a way that would make you proud.”

* * *

He’d been so close. And now, this…whatever it was.

“It’s not a problem with *you*,” said the supervisor, though he looked haggard and did not exactly inspire confidence. “It’s just…there’s been an incident. We’re interviewing everybody.”

It was two days before launch. Everything had been set up. Now, as the supervisor ushered him into a sterile-looking interview room, he could already imagine the headlines. MAN BUYS FALSIFIED GENEAOLOGY, TRIES TO HIDE EPIGENETIC INFERIORITY. Or, FRAUD DECEIVES EPIGATTACA CORPORATION, MISSION CALLED OFF.

He was so busy generating worst-case scenarios that he didn’t even notice the identity of the detective seated across from him until the door had closed and they were alone together.

“Anton?”

“Vincent?”

“Um…yeah…” was the best Vincent could think of to say to his older brother.

They’d been in touch, sure. But Vincent had thought it prudent not to mention his exact job description, lest his brother start asking the wrong questions. He’d just said he worked for the Epigattaca corporation, letting Anton infer that he was sweeping floors or validating parking tickets or something else suitable to someone with his inadequate histone pattern. Finally he put himself together and spoke.

“What are you doing here?”

“There’s been a death. One of the executives. Foul play suspected. I’m a detective, so…”

Then it had nothing to do with him. Or, at least, it hadn’t. Now…

“Okay, Anton. Before you ask, yeah. I admit it. I faked my epigenome. There’s a black market in geneaologies. I found a guy willing to sell his identity. I saved up, bought it from him, gave it to the suits here. They think I haven’t had a microaggression in my family line since the Mayflower. And they’re going to do it. They’re going to let me go to space.”

“But why, Vincent? You were always such a good kid!”

“Of *course* you wouldn’t understand,” said Vincent.

It was true. Anton was five years older. He’d been born perfect, the product of the latest eu-epigenics program. Female infants with good epigenes were sent to live a sheltered existence in Denmark, the most equitable country in the world, and kept drugged on heroic doses of beta-blockers to prevent them from feeling any trauma or anxiety. They were raised in special houses by caretakers who denied them nothing, then sent to special schools where it was impossible to fail or feel inadequate. Then, when they reached puberty, they were artificially fertilized – no *way* the program was going to let them deal with something as stressful as sexual relationships – until they pumped out five or ten kids each. The most innocent were brought back to the shelters to restart the cycle.

It had all gone so well with Anton. But a year after he was born, everything had changed. One of the nurses had gotten sick, and an untrained nurse was brought in to cover. She had told Anton's mother that she was looking "a little chubby". Faced with this sudden awareness of patriarchal beauty standards and devaluing from a human being to a sex object, her histones had wilted instantly, her precious DNA inundated with methyl groups. When the scientists found out, they discharged her from the breeding program, she married a similarly damaged man, and the result, a few years later, had been Vincent.

"You're right," said Anton. "I can never understand what you go through. But I'm going to clear you. Right now. No conditions."

Vincent could barely believe he'd heard correctly. "What?"

"It's…my son. Your nephew. I never told you this, but a few years ago, he broke his leg biking. We got it treated by a black-market doctor, covered it up, no trace of it in any of the records, but – I can't help worry about him. He remembers what happened. He's going to need role models to look up to when he grows older, people who overcame epigenetic determinism and succeeded despite the changes our experiences impose on our DNA. He's going to need someone like you. And besides, we're brothers. I've already figured out who committed the murder – it was another executive whose department would profit by delaying the launch. I’m going to report that you’re innocent. And I’m also going to report that I didn’t discover anything else of note about you. Nothing that should delay this week’s launch.”

“You…you’d really do that for me?”

“Good luck in space, bro.”

Vincent walked out of the office in a daze. By the time he reached his desk, the email was already on the screen “Detective has said you’re good to go – launch is still on for Wednesday”. He read it three times, lost in thought.

He had always wanted to be an astronaut. Now he realized why – it was to escape his own epigenome. More than that – it was to escape a world that held epigenetics in such high regard that his epigenome mattered. He would still go to space. He would do it for Anton’s son, and for all the other individuals with a history of personal or family trauma. But he no longer felt like there was nothing for him on Earth. There were people who would judge him as a full human being, not just a methylation pattern shaped by familial disadvantage. And someday – he vowed – everyone would be able to say the same.

GATTACA III: EDU-GATTACA

“Congratulations, Vincent”, said the supervisor, eyes never looking up from his clipboard. “You passed them all. The astrogation test. The crisis simulation. All the physicals and health panels. More than passed. Some of the highest scores we’ve ever seen, frankly. You’re going to be an astronaut…”

Vincent broke out into a giant smile.

“…pending, of course, the results of the final test. But this will be easy. I’m sure a fine specimen like you will have no trouble.”

“The…the final test, sir?”

“Well, you know how things are. We want to make sure we get only the strongest, most intelligent individuals for our program. We used to do genetic testing, make sure that people’s DNA was pre-selected for success. But after the incident with the Gattaca Corporation and that movie they made about the whole thing, public opinion just wasn’t on board, and Congress nixed the whole enterprise. Then we tried epigenetics, but it turned out they made a movie about that one too. Really, our luck in all of this has been terrible. But this time, we’ve really got it! This time, we know how to identify truly superior human beings who deserve to be astronauts, no creepy biology involved. We’re going to base our decision on…what institution you spent four years in during your teens and early twenties!”

“Oh, *come on*,” said Vincent. “Can’t you just give up already and judge people on their merit?”

The supervisor pounded the desk. “Never! So-called meritocracy is a sham designed to justify inequality. No, we’ve made our choice, and we’re going to judge you by which university accepted you at age 17 based on a combination of illegibly-inflated grades, recommendations by people who barely knew you, and how much money your parents were willing to donate. You can complain all you want, but that’s just how we roll, here at the…” He pointed out the window, to the gleaming sign outside “…at the PhDMSMABSBA corporation.”

“How do you even expect people to pronounce that?” asked Vincent.

“Irrelevant! Now tell us what college you went to, so we can figure out what Greek letter to assign you on your application.”

“Greek letter?”

“Just an internal company code we use. We got tired of saying ‘top-tier institution’, ‘second-tier institution’, and so on, so now you’re Alphas, Betas, Gammas, Deltas, and Epsilons. The Alphas get positions like executive or astronaut. The Betas get positions in middle management. The Gammas and Deltas have jobs like clerks and call center reps. And the Epsilons do the really dirty work, the stuff nobody else will touch.”

“That’s terrible!” said Vincent.

“That’s what *everybody* does,” the supervisor corrected. “The only difference is we use Greek letters. Is your moral system so fragile that its results depend on whether you refer to something with Greek letters or not?”

“Wow,” said Vincent, “this conversation has taken a disturbing turn.”

“That’s right. So why don’t you just show us your college degree, and we can get your application going?”

Vincent reached into his briefcase, took out a slim red binder. *Here goes nothing,* he thought.

* * *

Two weeks earlier, he had thrown a wad of cash down on a marble table in a suburban apartment.

The man across from him leaned forward in his wheelchair, extended a trembling arm to slowly take the cash.

“You’re, uh, sure you want to do this?” asked Vincent, suddenly feeling a pang of conscience.

“Yes,” rasped Jerome. “You’re young. You still have your whole life ahead of you. You can still make something of yourself. Me, I’m done.”

“Look,” said Vincent, despite his better judgment, “just because you’ve got some kind of condition doesn’t mean you can’t…do anything, really. Write a book. Travel the world.”

“You want to know how this happened?” Jerome asked. “I did it to myself. I used to be the best football player in the state. Maybe the country. Got accepted to Harvard on a football scholarship. Then I learned that college football causes so many concussions that it increases your risk of chronic traumatic encephalopathy. I didn’t know what to do. Stick around, and I risked degenerating into the condition you see me in now. Leave, and I’d lose my scholarship, never get a degree, and have to go to community college. I’d never be anything higher than a Delta.”

“A Delta?”

“Some new corporate jargon people are using.” Jerome shrugged. “So I kept it up and got my degree. Now I can barely walk, and half the time I can’t remember my own name. So here I am.”

“I’m so sorry,” said Vincent, who was starting to regret ever having come. “We can help you. We can find some way to…”

“No,” said Jerome, and he put the cash in his pocket. “What’s done is done.” He took a slim red binder, slid it over the table to Vincent. “My Harvard degree. Top-tier institution, absolutely Alpha quality. With this, every single door in the world will be open to you.”

Vincent considered saying something, but finally just nodded. “I won’t let you down. I’ll use the gift you’ve given me in a way that would make you proud.”

He’d been so close. And now, this…whatever it was.

“It’s not a problem with *you*,” said the supervisor, though he looked haggard and did not exactly inspire confidence. “It’s just…there’s been an incident. We’re interviewing everybody.”

It was two days before launch. Everything had been set up. Now, as the supervisor ushered him into a sterile-looking interview room, he could already imagine the headlines. MAN BUYS FAKE DEGREE, TRIES TO HIDE EDUCATIONAL INFERIORITY. Or, FRAUD DECEIVES PHDMSMABSBA CORPORATION, MISSION CALLED OFF.

He was so busy generating worst-case scenarios that he didn’t even notice the identity of the detective seated across from him until the door had closed and they were alone together.

“Anton?”

“Vincent?”

“Um…yeah…” was the best Vincent could think of to say to his younger brother.

They’d been in touch, sure. But Vincent had thought it prudent not to mention his exact job description, lest his brother start asking the wrong questions. He’d just said he worked for the PhDMSMABSBA corporation, letting Anton infer that he was sweeping floors or validating parking tickets or something else suitable to someone with his educational background. Finally he put himself together and spoke.

“What are you doing here?”

“There’s been a death. One of the executives. Foul play suspected. I’m a detective, so…”

Then it had nothing to do with him. Or, at least, it hadn’t. Now…

“Okay, Anton. Before you ask, yeah. I admit it. I faked my degree. I found a guy willing to sell his identity. I saved up, bought it from him, gave it to the suits here. They think I’m a Harvard alum. And they’re going to do it. They’re going to let me go to space.”

“But why, Vincent? You were always such a good kid!”

“Of *course* you wouldn’t understand,” said Vincent.

It was true. Vincent had spent years working ten-hour days after school teaching the saxophone to underprivileged children to build his resume, but the Admissions Departments hadn’t been impressed. Anton had learned from his mistake, hired an admissions coach, and with her guidance had founded the country’s first Klingon-language suicide prevention hotline; the Ivy League had eaten it up. Vincent had ended up with a degree from a Gamma-level state institution; Anton had gone to Yale.

“You’re right,” said Anton. “I can never understand what you go through. But I’m going to clear you. Right now. No conditions.”

Vincent could barely believe he’d heard correctly. “What?”

“It’s…my son. Your nephew. You remember how he started a synchronized underwater molecular gastronomy team at his high school? Apparently all the other kids have been going to the Third World and starting synchronized underwater molecular gastronomy teams *there*, and we never knew about it. Now there’s no way he’s going to be competitive. I can’t help worry about him. He’s going to need role models to look up to when he grows older, people who overcame going to a low-tier college and succeeded anyway. He’s going to need someone like you. And besides, we’re brothers. I’ve already figured out who committed the murder – it was another executive whose department would profit by delaying the launch. I’m going to report that you’re innocent. And I’m also going to report that I didn’t discover anything else of note about you. Nothing that should delay this week’s launch.”

“You…you’d really do that for me?”

“Good luck in space, bro.”

Vincent walked out of the office in a daze. By the time he reached his desk, the email was already on the screen “Detective has said you’re good to go – launch is still on for Wednesday”. He read it three times, lost in thought.

He had always wanted to be an astronaut. Now he realized why – it was to escape his own low-tier college degree. More than that – it was to escape a world that held degrees in such high regard that the college he went to mattered. He would still go to space. He would do it for Anton’s son, and for all the other individuals with a subpar secondary education. But he no longer felt like there was nothing for him on Earth. There were people who would judge him as a full human being, not just a set of letters after his name. And someday – he vowed – everyone would be able to say the same.

Contra Caplan On Arbitrary Deploring

Last year, Bryan Caplan wrote about what he called The Unbearable Arbitrariness Of Deploring:

> Let’s start with the latest scandal. People all over the country – indeed, the world – have recently discovered that many celebrities are habitual sexual harassers. Each new expose leads to public outrage and professional ostracism. Why does this confuse me? Because many celebrities do many comparably bad things other than sexual harassment, and virtually no one cares.
>
> Suppose, for example, that a major celebrity is extremely emotionally abusive to all his subordinates. He screams at them all the time. He calls them the cruelest names he can devise. He habitually makes impossible demands. He threatens to fire them out of sheer sadistic pleasure. But the abuse is never sexual (or ethnic); the celebrity limits himself to attacking subordinates’ intelligence, character, pride, and hope for the future. I daresay the average employee would far prefer to work for a boss who occasionally pressured them for a date. But if the tabloids ran a negative profile on the Asexual Boss from Hell, the public wouldn’t get very mad and Hollywood almost certainly wouldn’t ostracize the offender […]
>
> Or to take a far more gruesome case: When the Syrian government last used poison gas, killing roughly a hundred people, the U.S. angrily deployed retaliatory bombers, to bipartisan acclaim. But when the Syrian government murdered vastly more with conventional

> weapons, the U.S. government and its citizenry barely peeped. The unbearable arbitrariness of deploring!
>
> In the past, I’ve made similar observations about Jim Crow versus immigration laws, and My Lai versus Hiroshima. In each case, I can understand why people would have strong negative feelings about both evils. I can understand why people would have strong negative feelings about neither. I can understand why people would have strong negative feelings about the greater evil, but not the lesser evil. But I can’t understand why people would have strong negative feelings about the lesser evil, but care little about the greater evil. Or why they would have strong negative feelings about one evil, but yawn in the face of a comparable evil.

He concludes people are just biased by dramatic stories and like jumping on bandwagons. Everyone else is getting upset about the chemical weapon attack, and people are sheep, so they join in.

I have a different theory: people get upset over the violation of already-settled bright-line norms, because this is the correct action if you want to use limited enforcement resources efficiently.

Imagine a town with ten police officers, who can each solve one crime per day. Left to their own devices, the town’s criminals would commit thirty muggings and thirty burglaries per day (for the purposes of this hypothetical, both crimes are equally bad). They also require different skills; burglars can’t become muggers or vice versa without a lot of retraining. Criminals will commit their crime only if the odds are against them getting caught – but since there are 60 crimes a day and the police can only solve ten, the odds are in their favor.

Now imagine that the police get extra resources for a month, and they use them to crack down on mugging. For a month, every mugging in town gets solved instantly. Muggers realize this is going to happen and give up.

At the end of the month, the police lose their extra resources. But the police chief publicly commits that from now on, he's going to prioritize solving muggings over solving burglaries, even if the burglaries are equally bad or worse. He'll put an absurd amount of effort into solving even the smallest mugging; this is the hill he's going to die on.

Suppose you're a mugger, deciding whether or not to commit the first new mugging in town. If you're the first guy to violate the no-mugging taboo, every police officer in town is going to be on your case; you're nearly certain to get caught. You give up and do honest work. Every other mugger in town faces the same choice and makes the same decision. In theory a well-coordinated group of muggers could all start mugging on the same day and break the system, but muggers aren't really that well-coordinated.

The police chief's public commitment solves mugging without devoting a single officer's time to the problem, allowing all officers to concentrate on burglaries. A worst-crime-first enforcement regime has 60 crimes per day and solves 10; a mugging-first regime has 30 crimes per day and solves 10.

But this only works if the police chief keeps his commitment. If someone tests the limits and commits a mugging, the police need to crack down with what looks like a disproportionate amount of effort – the more disproportionate, the better. Fail, and muggers realize the commitment was fake, and then you're back to having 60 crimes a day.

This looks to me like what’s happening with chemical weapons. The relevant difference between chemical weapons and conventional weapons is that the international community made a credible commitment to punish chemical weapons use, and so far it’s mostly worked. People with chemical weapons expect to be punished for using them, so they rarely get used. If there are some forms of atrocity that are easier with chemical weapons than with conventional ones – ie a dictator with a limited arms budget can kill more people with a choice between chemical and conventional weapons than they can when restricted to conventional weapons alone – then the taboo against chemical weapons saves lives. And so when a dictator tests the limits by trying a chemical weapon, it’s worth responding to that more forcefully than if they used conventional weapons to commit the same massacre. You’re not just preventing the one attack, you’re also acting to enforce the taboo.

The sexual harassment situation seems like the same dynamic. We can’t credibly demand our elites are never jerks to their subordinates – jerkishness is too vague a concept, there’s too much of it around, and it’s just not really an enforceable norm. But we have sort of credibly demanded our elites don’t sexually harass their subordinates, and it seems like we might be getting enough of a coalition together to enforce this in a lot of cases. If we can solidify this into an actual social norm, such that the average elite *expects* to be punished for sexual harassment, then elites will stop sexually harassing their subordinates and we won’t have to keep calling the whole coalition together all the time to enforce the punishment. A fixed amount of public outrage per news cycle is the same kind of “limited enforcement resource” as only having ten police officers; commit to using it to disproportionately enforce existing taboos, and you’ll have more of it left over later on.

This is my long-winded answer to a question several people asked on the last links post – why should we prioritize responding to China’s mass incarceration of the Uighurs? Aren’t there other equally bad things going on elsewhere in the world, like malaria?

Yes. But I had optimistically thought we had mostly established a strong norm around “don’t put minorities in concentration camps”. Resources devoted to enforcing that norm won’t just solve the immediate problem in China, they’ll also help maintain a credible taboo against this kind of thing so it’s less likely to happen the next time.

Cost Disease In Medicine: The Practical Perspective

Sometimes I imagine quitting my job and declaring war on cost disease in medicine.

I would set up a practice with a name like Cheap-O Psychiatry. The corny name would be important. It would be a statement of values. It would weed out the people who would say things like “How dare you try to put a dollar value on the health of a human being!” Those people are how we got into this mess, and they would be welcome to keep dealing with the unaffordable health system they helped create. Cheap-O Psychiatry would be for everyone else.

Cheap-O Psychiatry wouldn’t have an office, because offices cost money. You would Skype, from your house to mine. It wouldn’t have a receptionist, because receptionists cost money. You would book a slot in my Google Calendar. It wouldn’t have a billing department, because billing departments cost money. You would PayPal me the cost of the appointment afterwards – or, to be really #aesthetic, use cryptocurrency.

The Cheap-O website would include a library of great resources on every subject. How To Eat Right. How To Get Good Sleep. How To Find A Good Therapist. The Cognitive Behavioral Therapy Workbook. The Meditation Relaxation Tape. But the flip side would be that Cheap-O appointments would be brutally efficient. If you had problems with sleep, I would evaluate you for any relevant diseases, give you any medications that might be indicated, then tell you to read the How To Get Good Sleep guide on the website. Boom, done. Small talk would be absolutely banned.

How little could Cheap-O charge? Suppose I wanted to earn an average psychiatrist salary of about $200K – the whole point of cost disease is that we should be able to lower prices without anyone having to take a pay cut. And suppose I work a 40 hour week, 50 weeks a year, each appointment takes 15 minutes, and 75% of my workday is patient appointments. That’s 6000 appointments per year. So to make my $200K I would need to charge about $35 per appointment. There would be a few added costs – malpractice insurance would probably run about $10K per year – but this is the best-case scenario.

$35 per appointment isn’t bad. Most existing cash-only psychiatry practices charge at least $150 per (thirty minute) appointment, so we would be less than a quarter of the going rate. I think a lot of insurances charge a $40 copay per psychiatrist visit, so even *uninsured* Cheap-O patients would be paying less cash than *insured* patients anywhere else. Create Cheap-O style psychiatry offices, primary care offices, etc, all around the country, and maybe (aside from catastrophe insurance, which should be cheap) having health insurance would no longer be such a big deal.

My job is great and I love it, so I’m only slightly tempted to do this myself. The reason I bring it up is: why doesn’t anyone else do it? And if it’s possible to provide cheap health care like this, then how does health care still cost so much? What am I missing?

I don’t know. I’ve never run a business and it’s probably much harder and more expensive than I think. One of the reasons for my Cheap-O fantasy is so that I could find out. But here are some speculations.

Part of the reason might be because there’s a shortage of doctor-entrepreneurs, and the few existing doctor-entrepreneurs are busy finding new ways to make ultra-boutique-super-premium clinics that they charge rich people $500/hour for the privilege of entering. And they're doing that because it pays way more than $200K/year. $200K/year is the standard salary for an average psychiatrist who wants a zero-risk job in the current system, and the privilege of never having to worry about the business side of things.

Another part might be that insurance is squatting in the mid-range market. Even in America, most people are insured. So unless your cost can beat the insurance co-pay – which even Cheap-O barely does under ideal conditions – most people will go to standard insurance-accepting practices unless you give them a good reason not to. And the best reason not to will be that you're claiming to be better than insurance-accepting clinics – which means you're aiming at the high-end market. And insurance-accepting practices can't lower prices because insurances make you follow lots of rules before they'll work with you, plus you need a small city worth of administrators to deal with the insurance companies.

(also, seeing a patient every fifteen minutes is exhausting; one of the advantages of hour-long appointments is that most people don't need an hour and so you can take the last twenty minutes to write notes or answer messages or work on blog posts)

Right now the only way I can imagine the niche getting filled is somebody doing it ~~for the lulz~~ as an act of political protest. Imagine if someone started Cheap-O Psychiatry and it *worked*. All of this stuff I'm saying about how socialized health care might be better than our current system but isn't the real answer, how it's just locking in entrenched cost disease and a truly free market could find better alternatives – instead of vaguely gesturing at it, there would finally be some evidence.

(well, there’s already the Surgery Center of Oklahoma, which does exactly this and costs about a fifth as much as surgery anywhere else. But maybe if there’s *more* evidence, people will stop ignoring it.)

Book Review: Capital In The Twenty-First Century

[Epistemic status: I am not an economist. Many people who are economists have reviewed this book already. I review it only because if I had to slog through reading this thing I at least want to get a blog post out of it. If anything in my review contradicts that of real economists, trust them instead of me.]

I.

Thomas Piketty's *Capital In The Twenty-First Century* isn't just a book on inequality. It's a book about quantitative macroeconomic history. This is much more interesting than it sounds.

Piketty spent decades combing through primary sources trying to get good statistics for what the economies of various Western countries have been doing over the past 250 years. Armed with these data, he tries to put together a theory of the very-long-term forces at work in economic change. His results touch on almost every big question in politics and economics, and are able to propose sweeping theories where other people resort to parochial speculation. While more knowledgeable people than I are probably already familiar with much of this, I used him as an Econ History 101 textbook and was not at all disappointed in the results.

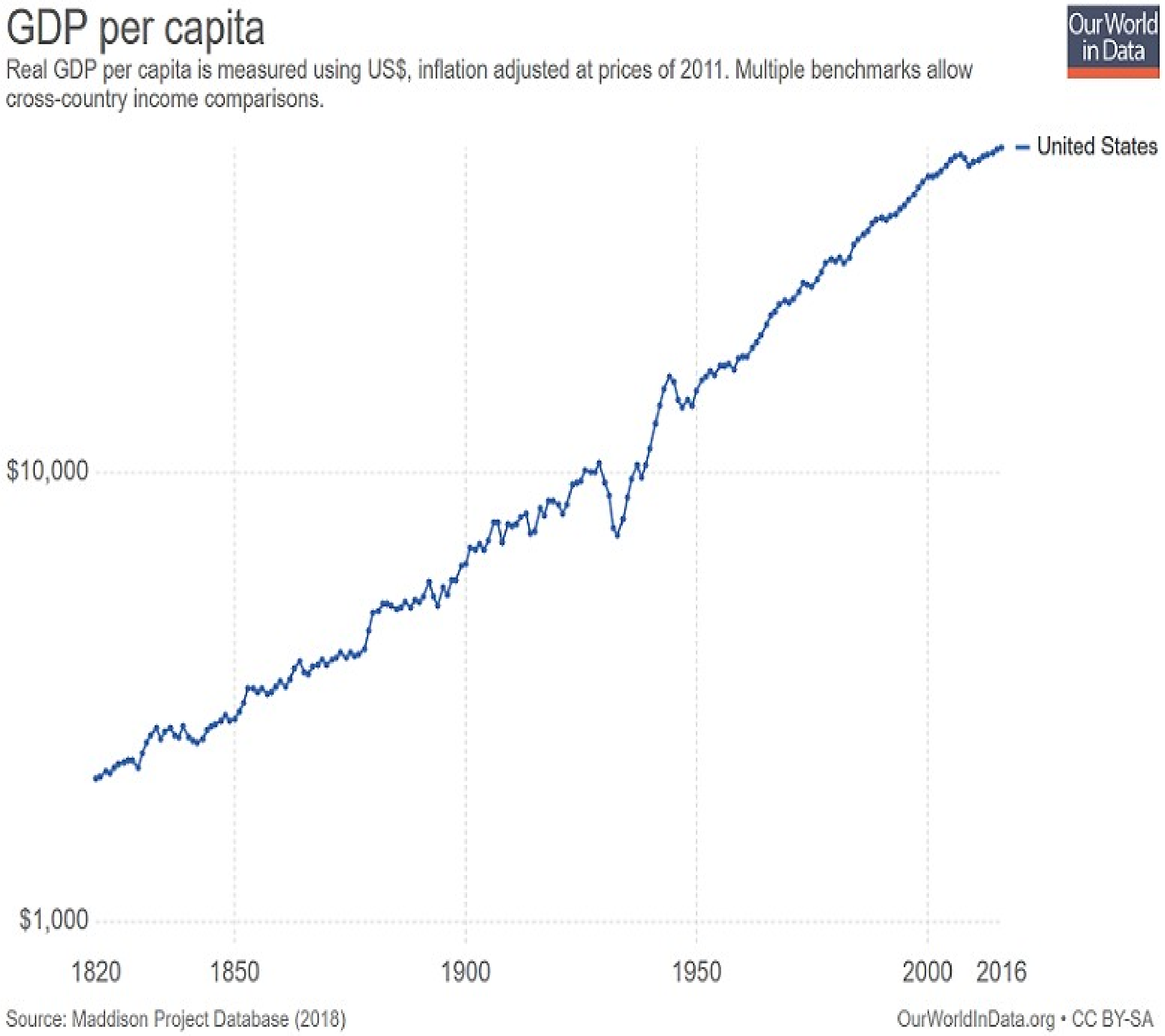
The most important thing I learned from Piketty is that since the Industrial Revolution, normal economic growth has always been (and maybe always will be) between 1% and 1.5% per year. This came as news to me, since I often hear about countries and eras with much higher growth rates. But Piketty says all such situations are abnormal in one of a few ways.

First, they can have high population growth. Population growth will increase GDP, and it will look like a high economic growth rate. But it doesn’t increase GDP per capita and it shouldn’t be considered the same as normal economic growth, which is always between 1% and 1.5% per year.

Second, they can have temporary bubbles. This definitely happens, but after the inevitable bust, the whole period will eventually average out to 1% to 1.5% per year.

Third, they can have “catch-up growth”. This is a broad category covering any period when a country that was previously underperforming its fundamentals gets a chance to catch up. This can happen after a long war in which a devastated country gets a chance to rebuild. Or it can happen after dropping communism or some other inefficient economic system, as the country transitions to a more practical form of production. Or it can happen when a Third World country globalizes and gets the benefits of First World technology and organization. But if a country is at peace and on the “technological frontier” (ie one of the highest-tech countries that has to invent its own advances and can’t get them by osmosis from somewhere else), it will always have growth of 1% to 1.5% per year.

For most of the 20th century, one or another of those conditions has been true in most places. Whether it was the Baby Boom in the US causing high population growth, or the dot-com boom, or Germany and Japan’s decades of miraculous economic recovery after World War II, or China’s catch-up growth after Deng Xiaoping’s liberalization, people have gotten growth of more than 1% to 1.5% per year and learned to expect it. In the Third World, which continues to experience good catch-up growth, their expectations will probably be met. In developed countries, they are bound to be disappointed.



I was happy to learn this, because the graph above has always haunted me. It seems to say that no matter our policies, no matter our good or bad decisions, no matter our triumphs and defeats – the US GDP per capita will always grow at the same rate. How can that be true, when some countries have such high GDP growth sometimes, and others such low growth? How can it be true when in 1870, the US was in the same kind of Gilded Age rapidly-industrializing business climate that China is today? How can it be true when I remember so many years of 5% growth in the US, and so many other negative-growth recessions? But Thomas Piketty is right and common sense is wrong. The US grows at

1% to 1.5% per year, forever, just like everyone else outside a few special circumstances.

Piketty uses this point to construct a super-ambitious zoomed-out version of 20th century history. Everything was going normally until the two World Wars, which devastated Europe and Japan, but set the US back only slightly. The US and Britain had the Baby Boom and minor catch-up growth, meaning a pretty great 1950s. Continental Europe had the whole process of rebuilding their economy from the (sometimes literal) ashes, a period the French called the Trente Glorieuses (Thirty Glorious Years) of near-constant economic boom. Around the 1970s, the US and Britain realized that Continental Europe and Japan were doing much better than they were, what with their near-constant economic boom, freaked out, and decided their economies were somehow rotten; this led to Thatcher and Reagan getting elected on a platform of cleaning up the economy. Around the same time, Europe recovered fully from its devastation and went back to normal economic growth; Japan, which had been a bit more devastated, took another few years but then had its own bust and went back to normal (or subnormal) growth. The US and Britain, seeing that they were now “caught up” to their Continental and Japanese competitors, declared “mission accomplished” and gave Thatcher and Reagan the credit. Since then it’s mostly been smooth sailing, with the normal 1%-1.5% + population growth (which works out to a little bit more in the US and a little bit less in Japan, given these countries’ high and low fertility rates respectively).

I don’t fully understand this theory. It proposes a very long time for Europe to get over World War II, which doesn’t really match graphs that show the GDP rebounding basically immediately even in the hardest-hit countries (maybe it would be more revealing if I had a log graph like the US one above, so I could do more than try to eyeball the trend). It also suggests that Americans judge the state of their economy by comparing it to Europe (or at least did in the 1970s), which doesn’t really match how most people I know think. In particular, in 1990 they would have had to have said “Our economy is now equal or better to the Europeans, we’re happy now” when this was entirely a function of Europe cooling down and didn’t involve any improvement in the US economy at all. Still, this is a persuasive model and one of the only ones I know that makes sense of the straight line graph above.

II.

All this is just exposition. Piketty’s main focus is inequality between labor and capital, and he starts with the idea of the rentier.

A rentier – an old word I’m surprised we don’t have a better-known modern equivalent for – is something like the strict sense of “capitalist”: a person who lives off the interest on savings instead of working. Trust-fund kids who live off dividends from investments, landlords who live off literal rents from their properties, and aristocrats who lived off the profits from their estates are all rentiers.

It’s fitting that Piketty uses an old word, because rentiers were more common and more important in the old world than they are today. Although he presents tables of statistics proving this is the case, Piketty also urges us to consider Jane Austen novels for a more intuitive sense of the situation. Few of her characters work honest jobs besides occasionally some sort of vague “managing investments”. They’re all obsessed with their dowries and their families’ endowments. That’s because they live off the interest of their principal, which usually stays the same throughout their life and which often comes from a dowry. Some of these people are titled aristocrats, others are “gentry”, others might not have qualified for either role – but they all live off interest.

Piketty (himself a Frenchman) also cites this passage on 19th-century French novelist Honore de Balzac’s *Pere Goriot*:

> The darkest moment in the novel, when the social and moral dilemmas [Eugène de] Rastignac faces are rawest and clearest, comes at the midpoint, when the shady character Vautrin offers him a lesson about his future prospects. Vautrin, who resides in the same shabby boardinghouse as Rastignac and Goriot, is a glib talker and seducer who is concealing a dark past as a convict, much like Edmond Dantès in Le Comte de Monte-Cristo or Jean Valjean in Les Misérables. In contrast to those two characters, who are on the whole worthy fellows, Vautrin is deeply wicked and cynical. He attempts to lure Rastignac into committing a murder in order to lay hands on a large legacy. Before that, Vautrin offers Rastignac an extremely lurid, detailed lesson about the different fates that might befall a young man in the French society of the day.
>
> In substance, Vautrin explains to Rastignac that it is illusory to think that social success can be achieved through study, talent, and effort. He paints a detailed portrait of the various possible careers that await his young friend if he pursues studies in law or medicine, fields in which professional competence counts more than inherited wealth. In particular, Vautrin explains very clearly to Rastignac what yearly income he can aspire to in each of these professions. The verdict is clear: even if he ranks at the top of his class and quickly achieves a brilliant career in law, which will require many

compromises, he will still have to get by on a mediocre income and give up all hope of becoming truly wealthy:

> “By the age of thirty, you will be a judge making 1,200 francs a year, if you haven’t yet tossed away your robes. When you reach forty, you will marry a miller’s daughter with an income of around 6,000 livres. Thank you very much. If you’re lucky enough to find a patron, you will become a royal prosecutor at thirty, with compensation of a thousand écus [5,000 francs], and you will marry the mayor’s daughter. If you’re willing to do a little political dirty work, you will be a prosecutor-general by the time you’re forty.… It is my privilege to point out to you, however, that there are only twenty prosecutors-general in France, while 20,000 of you aspire to the position, and among them are a few clowns who would sell their families to move up a rung. If this profession disgusts you, consider another. Would Baron de Rastignac like to be a lawyer? Very well then! You will need to suffer ten years of misery, spend a thousand francs a month, acquire a library and an office, frequent society, kiss the hem of a clerk to get cases, and lick the courthouse floor with your tongue. If the profession led anywhere, I wouldn’t advise you against it. But can you name five lawyers in Paris who earn more than 50,000 francs a year at the age of fifty?”

By contrast, the strategy for social success that Vautrin proposes to Rastignac is quite a bit more efficient. By marrying Mademoiselle Victorine, a shy young woman who lives in the boardinghouse and has eyes only for the handsome Eugène, he will immediately lay hands on a fortune of a million francs. This will enable him to draw at age twenty an annual income of 50,000 francs (5 percent of the capital) and thus immediately achieve ten times the level of comfort to which he could hope to

> aspire only years later on a royal prosecutor's salary (and as much as the most prosperous Parisian lawyers of the day earned at age fifty after years of effort and intrigue).
>
> The conclusion is clear: he must lose no time in marrying young Victorine, ignoring the fact that she is neither very pretty nor very appealing. Eugène eagerly heeds Vautrin's lesson right up to the ultimate coup de grâce: if the illegitimate child Victorine is to be recognized by her wealthy father and become the heiress of the million francs Vautrin has mentioned, her brother must first be killed. The ex-convict is ready to take on this task in exchange for a commission. This is too much for Rastignac: although he is quite amenable to Vautrin's arguments concerning the merits of inheritance over study, he is not prepared to commit murder.
>
> What is most frightening about Vautrin's lecture is that his brisk portrait of Restoration society contains such precise figures. As I will soon show, the structure of the income and wealth hierarchies in nineteenth-century France was such that the standard of living the wealthiest French people could attain greatly exceeded that to which one could aspire on the basis of income from labor alone. Under such conditions, why work? And why behave morally at all? Since social inequality was in itself immoral and unjustified, why not be thoroughly immoral and appropriate capital by whatever means are available?

Piketty's figures show that rentiers became more and more powerful from the 1700s all the way until the Belle Epoque (early 1900s), then declined precipitously around the period of the World Wars. What happened?

This is where he brings in his famous inequality (no pun intended) r > g – that is, the rate of return from capital is greater than the growth rate.

Rentiers' money grows with the rate of return on capital. As they get profits/dividends/rent/interest on their capital, they consume some fixed fraction of it and add the rest back to their principal. As their principal grows, so does their yearly income. If they have enough money (and most of them do), the amount they consume even in a very luxurious lifestyle will be trivial, so we can approximate this by saying their income grows at the rate of return.

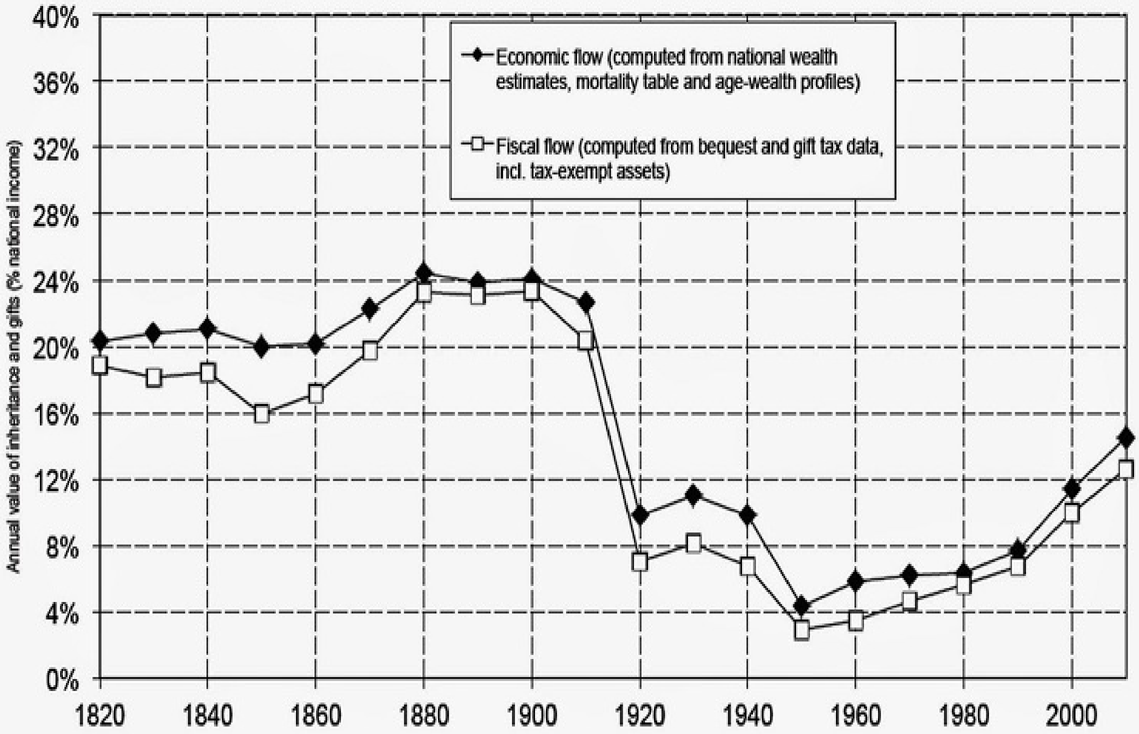
(One of the few checks on this process is population growth among the rentier class – if a rentier has five children, he may have to consider splitting his fortune five ways, which would mean that the fortune of each individual rentier decreases by a factor of five each generation. This means the growth of rentier power is strongly influenced by population growth, which will become important later.)

Laborers are assumed to have few savings and live a hand-to-mouth existence. Their income grows whenever they get a raise. So the growth of their income is approximated by the GDP per capita growth rate.

In ordinary times, the rate of return on capital always averages about 4% – 5% per year, and the GDP per capita growth rate always averages about 1% to 1.5% per year. So in ordinary times, rentiers' yearly incomes should always be pushing further and further ahead of laborers', and inequality should always increase. This is exactly what happened between the 1700s and 1914.

During crises – especially wars and economic busts – the situation reverses. If a rentier invests all his money in a factory, and the enemy bombs that factory, the rentier is broke. The laborers in the factory are also pretty unhappy, but they have the opportunity to get a job in a new factory when the war ends, in a way that the rentier – whose family might have spent several generations accumulating their capital – might not. The same is true of hyperinflation: laborers will get paid hyperinflated wages to spend on hyperinflated goods – but rentiers, who might have their money in currency-denominated investments like government bonds, can lose decades of careful fortune-gathering. Finally, if the government decides to respond to the crisis with confiscation of resources, or wealth taxes, or any interference in the economy, it’s likely to be the rentiers who are hardest-hit. Because rentiers’ wealth takes decades or generations to accumulate, but laborers live hand-to-mouth, a crisis lasting five years will give laborers a bad five years (after which they’re in the same position as pre-crisis), but can ruin rentiers completely. The period 1914 to 1945 – containing two World Wars and the Great Depression – was an unprecedented prolonged multi-crisis that caused “the collapse of the rentier world”.

Figure 11.1. The annual inheritance flow as a fraction of national income, France 1820-2010



The annual inheritance flow was about 20-25% of national income during the 19th century and until 1914; it then fell to less than 5% in the 1950s, and returned to about 15% in 2010. Sources and series: see piketty.pse.ens.fr/capital21c.

When the dust settled, a big chunk of the multi-generational fortunes that were dominating Europe had been destroyed or “haircut” down to more manageable levels. This was both a cause and an effect of immediately-post-war governments imposing rentier-unfriendly policies like high top-bracket tax rates that essentially prevented the rentier class from re-forming. And with the post-war Baby Boom increasing demographic growth, the conditions for rentiers (who do best when they don’t need to split their inheritance among too many children) got even worse. The period around the World Wars was the only one in history where g effectively outperformed r, leading to conditions where laborers win and rentiers lose out.

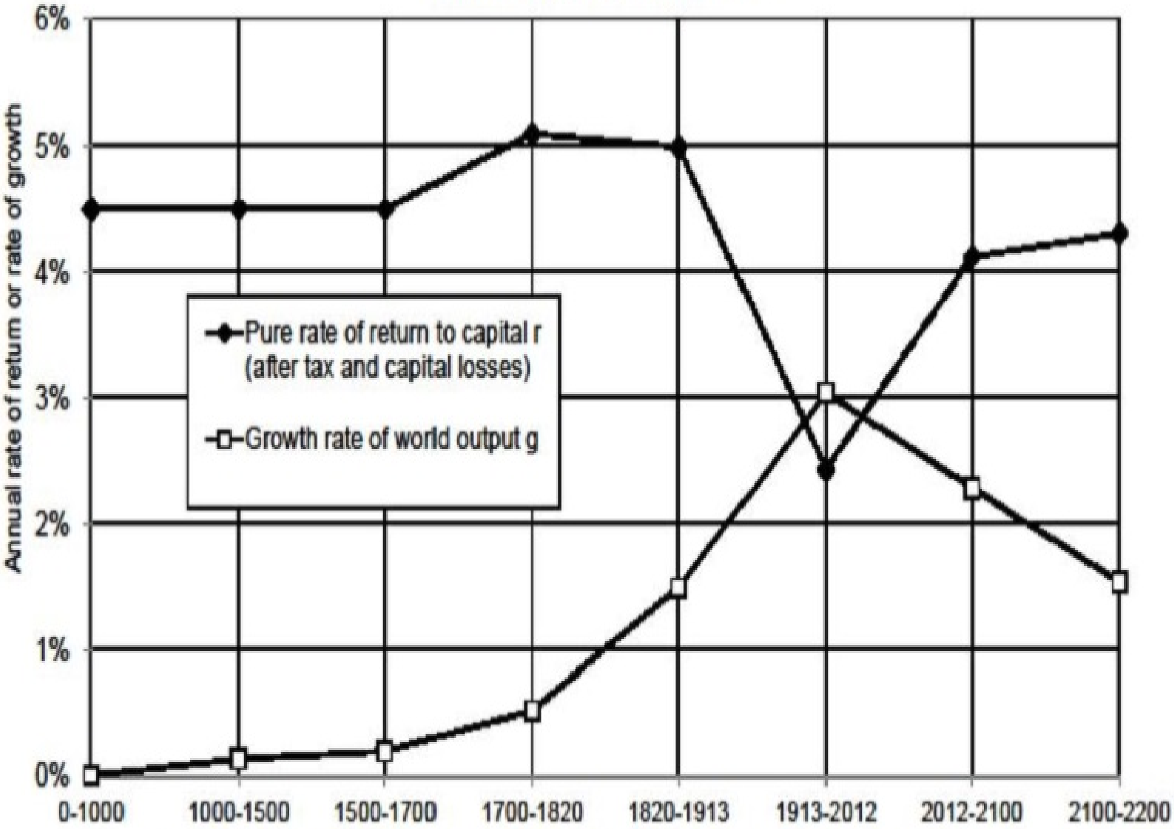


Figure 10.11. After tax rate of return vs. growth rate at the world level, from Antiquity until 2200

The rate of return to capital (after tax and capital losses) fell below the growth rate during the 20th century, and might again surpass it in the 21st century. Sources and series: see piketty.pse.ens.fr/capital21c

As a result, the point in the income distribution where people switch from mostly-laborers to mostly-rentiers used to be when you got into the top 1% in 1900. Today you have to go all the way to the top 0.1%.

On the other hand, since World War II the First World has had fifty-plus years of relative peace and prosperity. The last few decades have seen decreasing population growth and tax cuts for the upper class. These are the perfect conditions for the rentier class to make a return, and this is exactly what the data show.

…sort of. If you measure in capital/income ratio, rentiers are doing pretty well. If you look at inequality, it's pretty high. If you look at various comparisons of share given to capital vs. labor, capital is coming out on top. But then *where are the rentiers?* Aside from occasional jokes about trust-fund kids, you rarely hear about them. And Piketty's own data confirm that only the top 0.1% of the population makes most of their income from capital, compared to the entire top 1% back during the Belle Epoque.

I'm not sure about this, and for a point which is kind of the center of his entire argument, Piketty doesn't seem too sure either:

> We need to understand the reasons for this long-term change, which are not obvious at first glance, since I showed in Part Two that the capital / income ratio has lately returned to Belle Epoque levels. The collapse of the rentier between 1914 and 1945 is the obvious part of the story. Exactly why rentiers have not come back is the more complex and in some ways more important and interesting part. Among the structural factors that may have limited the concentration of wealth since World War II and to to this day have helped prevent the resurrection of a society of rentiers as extreme as that which existed on the eve of World War I, we can obviously cite the creation of highly progressive taxes on income and inheritances (which for the most part did not exist prior to 1920). But other factors may also have played a significant and important role.

If I understand Piketty right – and reading a bit between the lines – I think there are at least three things going on.

First, there's more of a middle class these days. The middle class doesn't entirely live off capital, but they have some savings and investment. Instead of 1% of people making all their money off capital, we have 30% of people saving for retirement and living off the profits of their nest egg.

Second, there are more super-rich “laborers”, for a broad definition of laborer that includes CEOs of big corporations. If we’re talking about how far in the income distribution you have to go before you get to rentiers, having a bunch of super-rich laborers screws up that statistic. I don’t think Piketty presents the more interesting statistic of what percent of people are rentiers, and I’m not sure why not. These people not only muddle the statistics, they also get much more media attention. When we talk about rich people, we talk about Bill Gates and Jeff Bezos, not the nth-generation scion of the Rockefeller family.

Third, capital is increasingly owned by institutions. There are hundreds of billions of dollars tied up in the endowments of top universities, and trillions tied up in the sovereign wealth funds of oil countries like Norway and Saudi Arabia. These don’t look like a particular individual walking around in a top hat and monocle, but they still distort the flow of money away from people who work for it and toward the people lucky enough to be part of the relevant institutions.

But if Piketty is right, if there’s no crisis then rentiers’ share of the pie will continue to increase. Either it will eventually overwhelm these considerations, and we’ll wake up one morning and notice we have an idle hereditary nobility again. Or it *won’t* overwhelm these considerations, and more and more of the pie will go invisibly to people other than average laborers, without any conspicuous signs of what’s happening.

III.

Along with rentier-vs-laborer inequality, Piketty touches on income inequality among labor (remember, “labor” includes anyone who works a job, including CEOs).

He confirms what everyone already knows: the share of the top 10% (especially top 1%) has been increasing for decades now. This is most pronounced in the Anglosphere, but still happening somewhat in Continental Europe and Japan.

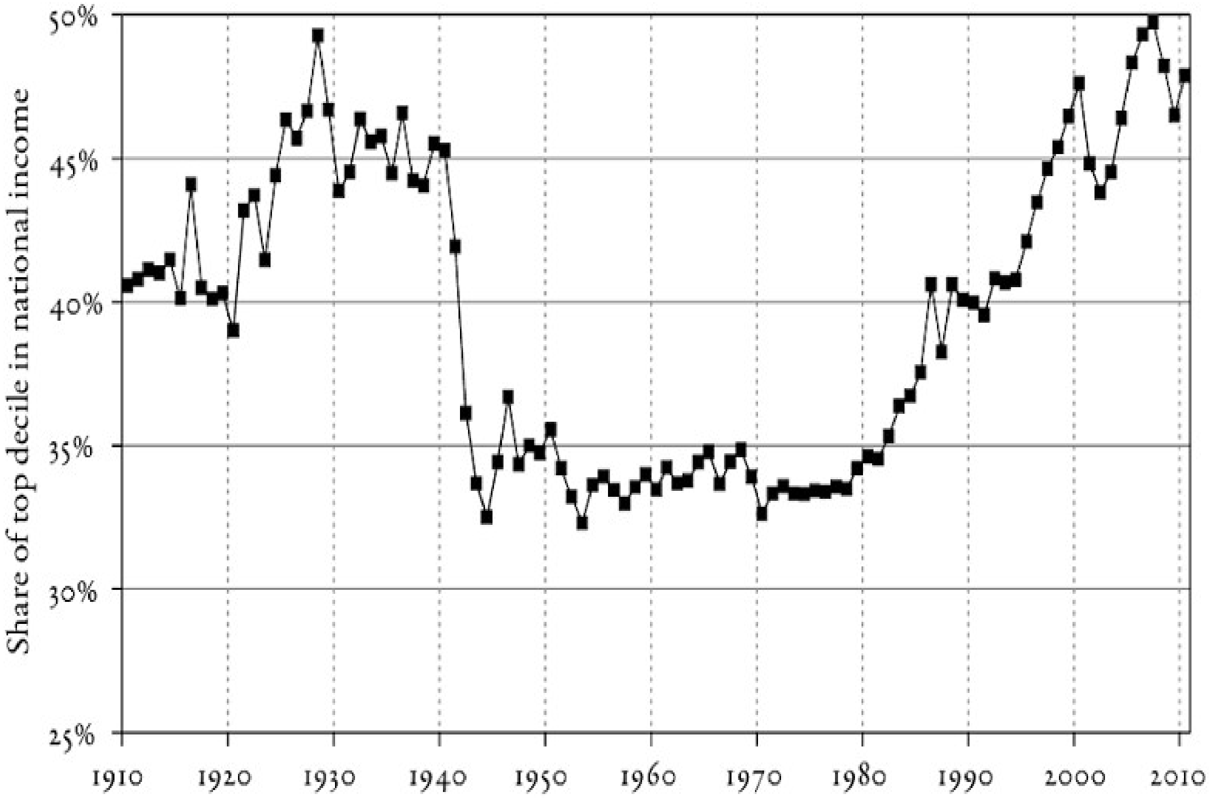


FIGURE I.1. Income inequality in the United States, 1910–2010

More things everyone already knows confirmed: most of the people benefiting from this are top executives at major corporations. Their average salaries have increased by orders of magnitude over the past few decades.

He addresses an argument made by supporters of high CEO salaries: might there not be a good economic justification for paying these people a lot? Suppose that having the best candidate (rather than the second-best candidate) as CEO increases your company’s profits by 1% – surely a plausible number. And suppose your company makes $10 billion/year. Then having the best CEO would increase your profits by

$100 million. So two companies in a bidding war for the best CEO ought to be willing to offer them a salary of up to $100 million/year to join.

Piketty isn’t buying it. This isn’t really his area of interest, so he doesn’t throw the same overwhelming level of statistical artillery at it that he does at some other things, and I don’t consider him to have proven it beyond doubt in the same way as some of his other conclusions – but he makes a few strong arguments. First, there are no signs that this situation is more true in the past few decades or in the Anglophone world, but these are the only places where CEOs get paid so much. Second, with appropriate caveats and controls there are no signs that good CEOs get paid more than bad CEOs. Third, CEO pay seems to clearly increase because of some factors outside of CEOs’ control, for example in an economic boom (though wouldn’t this increase company profits, and so be consistent with the 1% scenario above?)

He thinks that executive salaries have increased because – basically – corporate governance isn’t good enough to prevent executives from giving themselves very high salaries. Why didn’t executives give themselves such high salaries before? Because before the 1980s the US had a top tax rate of 80% to 90%. As theory predicts, people become less interested in making money when the government’s going to take 90% of it, so executives didn’t bother pulling the strings it would take to have stratospheric salaries. Once the top tax rate was decreased, it became worth executives’ time to figure out how to game the system, so they did. This is less common outside the Anglosphere because other countries have different forms of corporate governance and taxation that discourage this kind of thing.

This matters not just because it produces income inequality, but because today’s income inequality is tomorrow’s rentier-vs-laborer inequality. A CEO who earns $5 million per year can make $50 million, retire, invest the money, and pass the fortune on to their children. The more giant fortunes like this are around, the more rentiers there are in the next generation and the more inequality perpetuates itself.

Piketty is especially afraid of very large fortunes because of his *fascinating* data showing that these grow more quickly than other fortunes. Using the Forbes rich list, he calculates that Bill Gates et al must have grown their fortunes at rates approaching 8% – 10% per year – far higher than the 4% – 5% rate of return on capital Piketty usually uses. Forbes is a pretty sketchy data set, but he finds the same thing with the largest college endowments. The richest colleges, like Harvard and Yale, see their endowments grow at 10.2% yearly. Somewhat rich colleges (= $1 billion) grow at 8.8%, medium-rich colleges (= $500 million) at 7.8%, middling colleges (= $100 million) at 7.1%, and the poorest colleges (= $100 million) at 6.2%. And all of these do better than the average person saving for retirement, who – again – gets about 4% to 5%.

Piketty suggests this is because the richer you are, the more economy of scale you have in hiring really good financial planners. I am a little confused how this interacts with the conventional wisdom that experts are crap and you should invest in index funds, but I think the financial planners Piketty talks about aren't people who are very good at picking which stocks will go up, but more like people who know a guy who knows a guy in Singapore who can use your money to fund an unlisted Burmese mining project nobody else in the West has ever heard of.

I'm still confused why there isn't a mutual fund that lets retirees pool together their money to give it to one of these people. Part of the answer must be "there are only so many unlisted Burmese mining projects”, but then how do economies of scale help exactly? Piketty doesn’t answer, almost as if he is more interested in explaining the dynamics of inequality than in providing me personally with investment advice.

IV.

Taken together, all of this suggests a gloomy conclusion.

Income inequality can be expected to remain high. This will produce further rentier-vs-labor inequality, which conditions are already right to exacerbate. Not only will rich people separate from poor people, but the super-rich will separate much faster from poor people and even from other rich people. As more and more fortunes accumulate, we will get either the sort of rentier society typical of Europe in the 19th century, or a covert version of the same where the profits of rent go invisibly to various people connected to rentier institutions.

Piketty suggests the obvious direct solution: a global tax on wealth. He is okay with having this be merely nominal on fortunes up to hundreds of thousands of dollars, but he wants it to be significant on larger fortunes, and punitive on the largest fortunes. This will help prevent these from growing exponentially at the rate of return and prevent the rebirth of rentier society.

He is pessimistic about this ever working, because if any country does it unilaterally, rentiers can just move their wealth to another country. He suggest an EU-wide tax, but I do not entirely understand his reasoning for why Europeans wouldn’t then just move their wealth to the Cayman Islands or one of the many other perfectly good banking systems that are not in Europe.

He does give measured praise to the US system, including the FATCA laws, which place penalties on any country that don't report the details the US needs to tax American citizens' holdings in those countries. I think he hopes Europe could pass a law like this, but stronger.

I agree with his pessimism. Absent a World War level crisis that would make for a cure worse than the disease, it's hard to imagine everyone coming together to solve the sorts of problems Piketty brings up. I believe his predictions are likely to come true, with little chance of governments having the will to push a solution.

If a wealth tax itself is impossible, there are other things that might help a little. I can't help but note that solving the housing crisis would be a big help here. Rents go directly out of the pockets of laborers and into the pockets of landlords, and are probably the biggest such transfer in today's economy. Anything that lowers them has a big effect on the rentier-vs-laborer balance.

College is another big way that laborers get into debt. Although much of the debt is low-interest and government-owned, not all of it is, and even when it is this is scant consolation to the people who have to pay it.

Finally, one of the historically most important ways to decrease the power of rentiers is to increase birth rates, so they are forced to split their fortunes many ways every generation. Right now the birth rate of the rich is at historic lows. This might offer a point of agreement between rentier-fighting liberals and conservatives, who are already concerned about declining birth rates. Unfortunately, the main proposed solution to low population growth – giving workers better maternity leave – is likely to miss the rentier class entirely. I'm not sure what kind of policies might better target them.

So as not to end on a completely pessimistic note, I want to mention three causes for optimism I found in Piketty.

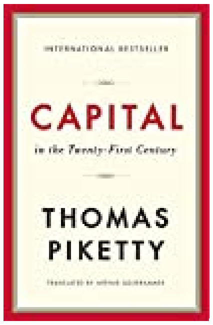
First, if we believe Piketty's data, there is no Great Stagnation, at least not in economic growth. Piketty argues that (outside of special circumstances) economies always grow at 1% – 1.5% per year, and First World economies are currently growing at 1% – 1.5% per year. If we expected them to grow more, it's because we're not adjusting for falling birth rates, or getting too excited by passing booms and busts, or comparing them to catch-up-growth like in China. Whatever our sins in terms of decreased innovation and efficiency, they have not yet hit the economic growth rate.

Second, catch-up growth provides a powerful force for reducing inequality between nations. Given enough time, the US economy will keep growing at 1% to 1.5% per year, and sub-Saharan African economies will keep growing at 3% to 8% per year. Will this continue until the latter have caught up entirely? Unclear. Reviewers of this post have mentioned Acemoglu's theory of entrenched inequality due to poor institutions, Sachs' theory of entrenched inequality due to disease burden, and Jones' theory of "hive mind" style entrenched inequality, as reasons why full catch-up might be hard. But all of these factors can potentially be improved with a growing economy, so they are not causes for total defeatism. *Capital In The Twenty-First Century* has overall made me more optimistic about the prospects for Third World catch-up.

The third thing that impressed me about the book is that we can talk about these kinds of things at all. Reading Piketty feels closer to reading real science – the type where there are universal laws that make clear predictions – than most economics or social science I've read. Marx liked to say he had discovered the universal laws that govern history; Piketty's claims are only slightly more modest, but much more believable. They make the actions of individuals seem very small – one recurring theme is some faction taking credit for improving the economy or fighting inequality when economic fundamentals meant things had to change then regardless of what anybody did.

But they also suggest points of leverage. I don't know if good economists knew all this stuff already. But I didn't, and learning it makes me more optimistic that we might one day find a way to solve the problems Piketty talks about. Even if his wealth tax doesn't work, he has good explanations of all the other factors that contribute to inequality and what would happen if we changed each. Now it's just a matter of political will.

And although that's a pretty big "just", I hope that maybe the move to a society of rentiers will change the political calculus enough to make people take these problems more seriously. I complain about attacks on "meritocracy"; maybe we can see how much people like hereditary rent-seeking, and whether getting rid of that makes a better rallying cry. Maybe "everyone productive vs the idle rich" will make for less toxic politics than "people who earn more vs. people who earn less". I can always dream.



EDIT: Commenters add that there have been some compelling criticisms of Piketty, especially by Matthew

Rognlie (paper, summary) and Phillip Magness (paper, comment)

List Of Passages I Highlighted In My Copy Of Capital In The Twenty-First Century

[Original review is here. Don’t worry, people who had interesting comments on the review – I’ll try to get a comments highlights thread up eventually.]

> For Ricardo, who published his Principles of Political Economy and Taxation in 1817, the chief concern was the long-term evolution of land prices and land rents. Like Malthus, he had virtually no genuine statistics at his disposal. He nevertheless had intimate knowledge of the capitalism of his time. Born into a family of Jewish financiers with Portuguese roots, he also seems to have had fewer political prejudices than Malthus, Young, or Smith. He was influenced by the Malthusian model but pushed the argument farther. He was above all interested in the following logical paradox. Once both population and output begin to grow steadily, land tends to become increasingly scarce relative to other goods. The law of supply and demand then implies that the price of land will rise continuously, as will the rents paid to landlords. The landlords will therefore claim a growing share of national income, as the share available to the rest of the population decreases, thus upsetting the social equilibrium. For Ricardo, the only logically and politically acceptable answer was to impose a steadily increasing tax on land rents.
>
> This somber prediction proved wrong: land rents did remain high for an extended period, but in the end the value of farm land inexorably declined relative to other forms of wealth as the share of agriculture in national income decreased. Writing in the 1810s, Ricardo had no way of anticipating the importance of technological

> progress or industrial growth in the years ahead. Like Malthus and Young, he could not imagine that humankind would ever be totally freed from the alimentary imperative.

One underappreciated feature of Piketty is his engaging presentation of economic history. A constant feature of the theorists he discusses is that they are all brilliant thinkers, they all follow the trends of their time to their obvious conclusions in ways deeper and more insightful than their contemporaries – and they all miss complicated paradigm shifts that make the trends obsolete and totally ruin their theories. Rationalists take note.

> Like Ricardo, Marx based his work on an analysis of the internal logical contradictions of the capitalist system. He therefore sought to distinguish himself from both bourgeois economists (who saw the market as a self-regulated system, that is, a system capable of achieving equilibrium on its own without major deviations, in accordance with Adam Smith’s image of “the invisible hand” and Jean-Baptiste Say’s “law” that production creates its own demand), and utopian socialists and Proudhonians, who in Marx’s view were content to denounce the misery of the working class without proposing a truly scientific analysis of the economic processes responsible for it.7 In short, Marx took the Ricardian model of the price of capital and the principle of scarcity as the basis of a more thorough analysis of the dynamics of capitalism in a world where capital was primarily industrial (machinery, plants, etc.) rather than landed property, so that in principle there was no limit to the amount of capital that could be accumulated. In fact, his principal conclusion was what one might call the “principle of infinite accumulation,” that is, the inexorable tendency for capital to accumulate and

> become concentrated in ever fewer hands, with no natural limit to the process. This is the basis of Marx’s prediction of an apocalyptic end to capitalism: either the rate of return on capital would steadily diminish (thereby killing the engine of accumulation and leading to violent conflict among capitalists), or capital’s share of national income would increase indefinitely (which sooner or later would unite the workers in revolt). In either case, no stable socioeconomic or political equilibrium was possible.
>
> Marx’s dark prophecy came no closer to being realized than Ricardo’s. In the last third of the nineteenth century, wages finally began to increase: the improvement in the purchasing power of workers spread everywhere, and this changed the situation radically, even if extreme inequalities persisted and in some respects continued to increase until World War I. The communist revolution did indeed take place, but in the most backward country in Europe, Russia, where the Industrial Revolution had scarcely begun, whereas the most advanced European countries explored other, social democratic avenues—fortunately for their citizens. Like his predecessors, Marx totally neglected the possibility of durable technological progress and steadily increasing productivity, which is a force that can to some extent serve as a counterweight to the process of accumulation and concentration of private capital. He no doubt lacked the statistical data needed to refine his predictions. He probably suffered as well from having decided on his conclusions in 1848, before embarking on the research needed to justify them.

As above. I didn’t previously understand exactly what Marx meant by “capitalists digging their own grave” or why it didn’t come true. In general, I seem to have missed how recent the idea of "economic growth" was. I don't really understand how 19th- and even some 20th- century economists could fail to understand this, though it seems like they had some complicated thoughts on this issue – see eg this article's descriptions of Smith's, Ricardo's, and Mill's ideas of the "steady state economy". Read history of economics backwards!

(but for a contrary perspective, see this comment thread)

> In any event, the data that Kuznets collected allowed him to calculate the evolution of the share of each decile, as well as of the upper centiles, of the income hierarchy in total US national income. What did he find? He noted a sharp reduction in income inequality in the United States between 1913 and 1948. More specifically, at the beginning of this period, the upper decile of the income distribution (that is, the top 10 percent of US earners) claimed 45–50 percent of annual national income. By the late 1940s, the share of the top decile had decreased to roughly 30–35 percent of national income. This decrease of nearly 10 percentage points was considerable: for example, it was equal to half the income of the poorest 50 percent of Americans.13 The reduction of inequality was clear and incontrovertible. This was news of considerable importance, and it had an enormous impact on economic debate in the postwar era in both universities and international organizations.
>
> Malthus, Ricardo, Marx, and many others had been talking about inequalities for decades without citing any sources whatsoever or any methods for comparing one era with another or deciding between competing hypotheses. Now, for the first time, objective data were available. Although the information was not perfect, it

> had the merit of existing. What is more, the work of compilation was extremely well documented: the weighty volume that Kuznets published in 1953 revealed his sources and methods in the most minute detail, so that every calculation could be reproduced. And besides that, Kuznets was the bearer of good news: inequality was shrinking.

Another good rationalist lesson! If your conclusion from the stories above was “never try to just reason things out from first principles, wait until you get empirical data”, then remember the lesson of Kuznets, who got 35 years of empirical data, drew all of the most reasonable conclusions from it, but happened to be working in an atypical economic period and so got his conclusions exactly backwards.

> In slowly growing economies, past wealth naturally takes on disproportionate importance, because it takes only a small flow of new savings to increase the stock of wealth steadily and substantially.
>
> If, moreover, the rate of return on capital remains significantly above the growth rate for an extended period of time (which is more likely when the growth rate is low, though not automatic), then the risk of divergence in the distribution of wealth is very high.
>
> This fundamental inequality, which I will write as r > g (where r stands for the average annual rate of return on capital, including profits, dividends, interest, rents, and other income from capital, expressed as a percentage of its total value, and g stands for the rate of growth of the economy, that is, the annual increase in income or output), will play a crucial role in this book. In a sense, it sums up the overall logic of my conclusions.

> When the rate of return on capital significantly exceeds the growth rate of the economy (as it did through much of history until the nineteenth century and as is likely to be the case again in the twenty-first century), then it logically follows that inherited wealth grows faster than output and income.

The shortest and clearest presentation Piketty gives of his r > g inequality.

> The dynamics and structure of inequality look very different in a country whose population increases by a factor of 100 compared with a country whose population merely doubles. In particular, the inheritance factor is much less important in the former than in the latter. It has been the demographic growth of the New World that has ensured that inherited wealth has always played a smaller role in the United States than in Europe. This factor also explains why the structure of inequality in the United States has always been so peculiar, and the same can be said of US representations of inequality and social class. But it also suggests that the US case is in some sense not generalizable (because it is unlikely that the population of the world will increase a hundredfold over the next two centuries) and that the French case is more typical and more pertinent for understanding the future.

Contrary to the popular consensus today, the American Dream was very real for most of America’s history. This was less due to American ideals and institutions, and more the fact that America (as a new country) had less time to accumulate a rentier class, which takes a couple generations for the fortunes to really multiply. And as a country with a very high population growth rate (remember the average colonial New England family had like 10 kids), any wealth that did get accumulated was quickly diluted over a bunch of children instead of inherited as a chunk. This was at least true in the North and frontier – the South, which inherited a lot of the worst parts of the 1600s English upper-class, got a society along the lines of the European aristocratic model with rentier plantation owners. Traditional models of labor-capital split generally fail to capture this because much Southern wealth was in the form of slaves, ie labor turned into a form of capital, which confuses the models. But as for the rest of the country, Piketty approvingly quotes Tocqueville's description of America as "a land without capital".

American exceptionalism lasted until the late 1800s/early 1900s. As much as people like to make fun of Horatio Alger stories, they were an exaggerated but accurate (or at least more accurate) portrayal of the society of the day. As America aged, got time to accumulate large multigenerational fortunes, and lowered its birth rate, its inequality and intergenerational-mobility became as bad as or worse than Europe's, depending on what metrics you use.

> I belong to a generation that turned eighteen in 1989, which was not only the bicentennial of the French Revolution but also the year when the Berlin Wall fell. I belong to a generation that came of age listening to news of the collapse of the Communist dictatorships and never felt the slightest affection or nostalgia for those regimes or for the Soviet Union. I was vaccinated for life against the conventional but lazy rhetoric of anticapitalism, some of which simply ignored the historic failure of Communism and much of which turned its back on the intellectual means necessary to push beyond it. I have no interest in denouncing inequality or capitalism per se—especially since social inequalities are not in themselves a problem as long as they are

> justified, that is, “founded only upon common utility,” as article 1 of the 1789 Declaration of the Rights of Man and the Citizen proclaims. (Although this definition of social justice is imprecise but seductive, it is rooted in history. Let us accept it for now. I will return to this point later on.) By contrast, I am interested in contributing, however modestly, to the debate about the best way to organize society and the most appropriate institutions and policies to achieve a just social order. Furthermore, I would like to see justice achieved effectively and efficiently under the rule of law, which should apply equally to all and derive from universally understood statutes subject to democratic debate.

Including this to defend Piketty against accusations that he is a communist. He’s obviously some stripe of leftist, and his writing can be infuriatingly preachy in a very specifically leftist way at times, and if you are a rightist reading him you will probably hate him for totally preventable tone-related reasons – but he is a lot more moderate than his critics give him credit for, and shares the capitalist-but-concerned-about-inequality position that I and many readers here can probably identify with.

> To a first approximation, the residents of [modern First World] countries own as much in foreign real estate and financial instruments as foreigners own of theirs.

Piketty helped me better understand the anti-colonialist position. It isn’t just that the colonizers were traditionally oppressive in the sense of taking over the government and committing human rights violations (though of course many were). It’s that – and this is the part I didn’t understand even when people chanted it at me as a slogan ad nauseum – colonialism was a partly a form of capitalism. Part of what “the British colonized Africa” means is that they set up a lot of companies in Africa with British people as the shareholders. Those companies went and did company things – some evil and rights-violating, others normal and productive – but the profits all went to British shareholders. Extended forever into time, that means that if the labor-capital split in Africa is X, then X% of Africa’s production will go to forever go to Britain, and only 1-X% to Africa.

One might argue that Britain invested money in developing Africa, which can only work with the prospect of repayment on normal investment terms. And one might even argue that it’s no worse for the average African laborer to have X% of his country’s money go to British elites he doesn’t know vs. than to African elites he doesn’t know. But this would be overstating the case – the British elites spend their money primarily in Britain, further developing Britain’s economy, whereas the African elites could be expected to spend their money mostly in Africa. And since (absent appropriation) a capitalist’s ownership of capital lasts forever, the British do a one-time positive action (investing in Africa) that produces an eternal negative action (having them get to drain some of Africa’s money forever). Although it’s not provably/necessarily true, in real life this is probably worse for Africa than leaving it alone and/or letting it invite foreign investment on the much better terms it could certainly have gotten if its governments weren’t controlled by the same people it was trying to negotiate with.

This also helps me understand the “colonialism still hurts countries today” position better.

> In the novels of Jane Austen and Honore de Balzac, the fact that land (like government bonds) yields roughly 5% of the amount of capital invested (or, equivalently, that the value of capital corresponds to to roughly

> twenty years of annual rent) is so taken for granted that it often goes unmentioned.

Piketty’s habit of supporting his statistics with quotes from Jane Austen and other period novelists is endearing. And one thing he mentions again and again is that throughout the 18th and 19th century, novelists were frank about money in a way that sounds strange to modern ears. Austen and her contemporaries would introduce a character by saying “He made such-and-such a number of pounds per year” to help their audiences understand the social class they was talking about, and sometimes devote pages to exploring a character’s finances in realistic detail to add color to their financial problems or lack thereof.

He speculates that this is because, for centuries, there was little-to-no inflation or other dramatic changes in the economy. This gave the meaning of money ample time to percolate down into the culture. Everyone knew (and had always known) that X amount of pounds meant you were well-off, and novelists were confident that if posterity read their book fifty years later, the message would still be clear.

This stasis extended to international exchange rates, which never changed. “In the nineteenth and early twentieth century, everyone knew that a pound sterling was worth about 5 dollars, 20 marks, and 25 franks.”

This is fascinating, but I’m surprised by Piketty’s theory – surely novelists don’t write *mainly* for posterity. If people within the next five years understand what they meant, surely that’s good enough? But why *did* past novelists go into so much detail about finances in a way that sounds bizarre to modern ears?

> In the classic novels of the nineteenth century, wealth is everywhere, and no matter how large or small the capital, or who owns it, it generally takes one of two forms: land or government bonds.

The 19th-century aristocrat living off his estates is a well-known figure in the popular imagination, but I was surprised to learn of his equally-common contemporary, the person living off government bonds. The UK government issued so many bonds to pay for the Napoleonic Wars that it created an entire class of bond tycoons who lived off the interest for generations. To its credit (and inconceivably to modern ears), the British government just plodded along paying its debt slowly but surely, and finally finished sometime around 1900!

This system worked because interest on government bonds was about 4% to 5% – much higher than today. Piketty doesn’t really explain why the rate of return changed. It also worked because nobody realized inflation was even a thing. After inflation started – and especially after a few bouts of hyperinflation – people quickly stopped living off government bonds except as part of a diversified portfolio.

> This historical record is fundamental for a number of reasons. First, it enables us to understand why nineteenth-century socialists, beginning with Marx, were so wary of public debts, which they saw – not without a certain perspicacity – as a tool of private capital.

In today’s climate where the left gets accused of tax-and-spend and the right of cruel austerity, it’s interesting to learn that the 19th-century debate was exactly the opposite, and public debt was viewed as taking money from the people (in the form of taxation) to give to the rich (in the form of interest). I think this change might also be an effect of interest rates going down, plus more money going to social programs, plus more taxes coming from the rich, plus an optimism that inflation will eventually destroy the debt so there’s less reason to worry.

Piketty seems uncertain that the left remains correct in its about-face: “This ‘progressive’ view of public debt retains its hold on many minds today, even though inflation has long since declined to a rate not much above the nineteenth century’s, and the distributional effects are relatively obscure.”

> Many traditional aristocratic societies were based on the principle of primogeniture: the eldest son inherited all (or at any rate a disproportionately large share) of the family property so as to avoid fragmentation and to preserve or increase the family’s wealth. The eldest son’s privilege concerned the family’s primary estate in particular and often placed heavy constraints on the property: the heir was not allowed to diminish its value and was obliged to live on the income from the capital, which was then conveyed in turn to the next heir in the line of succession, usually the eldest grandson. In British law this was the system of “entails” (the equivalent in French law being the system of substitution hereditaire under the Ancien Regime)…
>
> The inheritance law that derived from the French Revolution and the Civil Code that followed rested on two main pillars: the abolition of substitution hereditaires and priomgeniture, and the adoption of the principle of equal division of property among brothers and sisters (equipartition). This principle has been applied strictly and consistently since 1804: in France, the quotite disponible (that is, the share of the estate that parents are free to dispose of as they wish) is only

> a quarter of total wealth for parents with three or more children, and exemption is granted only in extreme circumstances (for example, if the children murder their stepmother). It is important to understand that the new law was based not only on a principle of equality (younger children were valued as much as the eldest and protected from the whims of the parents) but also on a principle of liberty and economic efficiency. In particular, the abolition of entails, which Adam Smith disliked and Voltaire, Rousseau, and Montesquieu abhorred, rested on a simple idea: this abolition allowed the free circulation of goods and the possibility of reallocating property to the best possible use in the judgment of the living generation, despite what dead ancestors may have thought.

Primogeniture make sense in a perverse way. If Family X gives everything to the eldest children, but Family Y splits it equally, then a few generations down Family X will retain a line of nobles with vast wealth, whereas Family Y will be much poorer and maybe removed from the decision-making process. I'm not sure why you would enshrine into law that everyone *has* to do this, though – surely the winning move is to enforce primogeniture in your own family but hope your rivals are dumb enough to split their fortunes equally. Maybe the goal is to bind your own great-grandson, over whom you are otherwise powerless.

I'm surprised that great lords didn't have the right to sell their capital. I guess I never hear about this, but it seems surprising that people were around to stop them.

The last paragraph is one of the few good arguments I've heard for constraining inheritance. I always get angry when I hear about rich people who clearly wanted one good child (or someone who wasn't a child at all) to get their fortune, only to have this contested by a bunch of ungrateful children the rich person hated who believe they’re entitled to the money by birth. But Piketty makes an argument for a compelling social reason this might be enforced (although of course you could solve this just by saying rich people have to split their money at least n ways, and your ungrateful children who you hate don’t have to be one of those n).

I wonder if there is any history of noble families deliberately having only one child to avoid forced splitting of their fortune.

> Several generations grew up [under conditions of inheritance barely mattering at all], in particular the baby boom generation, born in the late 1940s and early 1950s, many of whom are still alive today, and it was natural for them to assume that this was the “new normal”.
>
> Conversely, younger people, in particular those born in the 1970s and 1980s, have already experienced (to a certain extent) the important role that inheritance will once again play in their lives and the lives of their relatives and friends. For this group, for example, whether or not a child receives gifts from parents can have a major impact in deciding who will own property and who will not.

I am part of the generation born in the 1970s and 1980s, and though my parents have been very helpful to me in many ways, inheritance as such has not mattered to me or any of my friends at all, because our parents are still alive.

People are living longer today than they did in the golden age of inheritance, and if 19th century nobles could expect to come into the family fortune in their 30s or 40s, modern people have to wait until their 50s and 60s. I have a few patients in this position, and what it usually means is that there are a few months of constant legal crisis as they fight with siblings and tax assessors, and sort through hundreds of old boxes, and try to prepare a house for sale, and when everything’s done with and the lawyers have taken their share, they get a decent but not game-changing amount to add to their retirement savings.

Because this happens so late, nobody can really say “I’m not going to go to school and pursue a career; after all, I’m going to inherit a fortune soon enough”. They can’t even avoid saving for retirement: both my parents have reached retirement age, but both my grandmothers are still alive and well.

I wonder if, as inheritance becomes more important, some market for solutions to this kind of problem will come up. Imagine an “inheritance insurance” company, which gives people with well-off parents (let’s say) enough to make a down payment on a house in their 20s/30s in exchange for a much larger amount from their inheritance later on, with some kind of built-in agreement not to worry if their parents’ house burns down the next day and they’re left with nothing. My guess is there are too many ordinary risks and moral risks here for this to work, but it’s interesting to think about.

> Concretely, Harvard currently spends nearly $100 million a year to manage its endowment.

Part of Piketty’s section on how the super-rich (including colleges) get higher returns to their investment than everyone else. A good reminder that you should not donate to colleges.

> The situation has been very different since the 1980s: with per capita income growth of just over 1 percent a year, no one wants large and steady tax increases, which would mean even slower if not negative income growth.

Another part of Piketty’s macroeconomic-history explanation of political trends: just after WWII, catch-up growth was so intense that everyone had more than they expected and it was easy to convince people to put some of it into the construction of a social welfare state (easier in Europe and Japan, which were most devastated by the war, than in the US and UK). As growth slowed down, people were no longer willing to give the government so much money. Countries with large welfare states mostly maintained them due to the ratchet effect, but expanding the welfare state further has become a much harder sell (I think I am understanding this right).

This is an interesting counterclaim to the intuitive story that the welfare state grows in bad times because more people need/want welfare. I don’t know if research has been done to see which of these is true. Plausibly sudden crises are conducive to welfare expansion, but a long run of predictably poor growth is bad for it?

This suggests we should see very little welfare state expansion in the near future, which does fit current data, although it’s hard to square with the sudden attractiveness of pro-welfare-state-expansion candidates like Bernie Sanders and Jeremy Corbyn.

> Recent research has shown that the decline in government receipts in the poorest countries in 1980 – 1990 was due to a large extent to a decrease in custom duties, which had brought in revenue equivalent to

> about 5% of national income in the 1970s. Trade liberalization is not necessarily a bad thing, but only if it is not peremptorily imposed and only if the lost revenue can gradually be replaced by a strong tax authority capable of collecting new taxes and other substitute sources of revenue.

The claim being that developing countries, who may not be good at collecting income taxes, were hard-hit by trade liberalization because of direct revenue costs. I hadn't heard this argument before and I wonder if anyone has calculated whether this effect outweighs any positive economic effects that liberalization could have. I know that about 100% of economists who are not working for the Trump administration at this exact moment are against tariffs, but I don't know if maybe that's just from the First World perspective.

> It was the United States that was the first country to try [tax] rates above 70%, first on income in 1919-1922 and then on estates in 1937 – 1939. When a government taxes a certain level of income or inheritance at a rate of 70 or 80 percent, the primary goal is obviously not to raise additional revenue (because these very high brackets never yield much). It is rather to put an end to such incomes and large estates, which lawmakers have for one reason or another come to regard as socially unacceptable and economically unproductive – or if not to end them, then at least to make it extremely costly to sustain them and strongly discourage their perpetuation. Yet there is no absolute prohibition or expropriation. The progressive tax is thus a relatively liberal method for reducing inequality, in the sense that free competition and private property are respected while private incentives are modified in potentially radical ways, but always according to rules thrashed out

> in democratic debate…according to our estimates, the optimal top tax rate in the developed countries is probably above 80 percent.

This resolves some of my confusion about how it was possible to have 70% – 80% tax rates on the rich – it was never expected to work in terms of revenue collection, it was just a gentler way of setting a maximum wage.

I’m confused because Piketty seems to treat this as very effective and a big part of the reason why inequality was lower during the period it existed, whereas other sources I have read say it was constantly loopholed around and basically had no effect on anyone. Given that there were a lot of very rich people in the US around that time, it seems like loopholes must have been a big part of the picture.

> How can public debt be reduced to zero? One solution would be to privatize all public assets. According to the national accounts of the various European countries, the proceeds from selling all public buildings, schools, universities, hospitals, police stations, infrastructure, and so on would be roughly sufficient to pay off all outstanding public debt. Instead of holding public debt via their financial investments, the wealthiest European households would become the direct owners of schools, hospitals, police stations, and so on. Everyone else would then have to pay rent to use these assets and continue to produce the associated public services. This solution, which some very serious people actually advocate, should to my mind be dismissed out of hand. If the European social state is to fulfill its mission adequately and durably, especially in the areas of education, health, and security, it must continue to own the related public assets.

> It is nevertheless important to understand that as things now stand, governments must pay heavy interest (rather than rent) on their outstanding public debt, so the situation is not all that different from paying rent to use the same assets, since these interest payments weigh just as heavily on the public exchequer.

Piketty continues to have an unclassifiable position on debt, more nuanced than I have heard from anyone else. Here he seems to say that the government going into debt to create government services is similar in ways to privatizing those services, since either way rich people are getting paid for those services with everyone else’s money (though presumably debt is still much better, since taxation is much more progressive). He favors raising taxes so that services can be provided without creating any debt, and using large one-time taxes on wealth to clear existing public debt: “A flat tax of 15% on private wealth would yield nearly a year’s worth of national income and thus allow for immediate reimbursement of all public debt”. Sounds like an extreme moral hazard unless matched with something like a balanced budget amendment, though.

> Interestingly, nineteenth-century novelists were not content simply to describe precisely the income and wealth hierarchies that existed in their time. They often give a very concrete and intimate account of how people lived and what different levels of income meant in terms of the realities of everyday life. Sometimes this went along with a certain justification of extreme inequality of wealth, in the sense that one can read between the lines an argument that without such inequality it would have been impossible for a very small elite to concern themselves with something other than subsistence:
> extreme inequality is almost a condition of civilization.

> In particular, Jane Austen minutely describes daily life in the early nineteenth century: she tells us what it cost to eat, to buy furniture and clothing, and to travel about. And indeed, in the absence of modern technology, everything is very costly and takes time and above all staff. Servants are needed to gather and prepare food (which cannot easily be preserved). Clothing costs money: even the most minimal fancy dress might cost several months' or even years' income. Travel was also expensive. It required horses, carriages, servants to take care of them, feed for the animals, and so on.
>
> The reader is made to see that life would have been objectively quite difficult for a person with only 3–5 times the average income, because it would then have been necessary to spend most of one's time attending to the needs of daily life. If you wanted books or musical instruments or jewelry or ball gowns, then there was no choice but to have an income 20–30 times the average of the day.

This reminds me of those headlines like "MANHATTAN SOCIALITE COMPLAINS THAT IT'S IMPOSSIBLE TO HAVE A DECENT STANDARD OF LIVING ON LESS THAN $1 MILLION A YEAR", except much more sympathetic because they were plausibly right. It gives me profound gratitude for our past few hundred years of economic growth, and inspires some hope for the future.

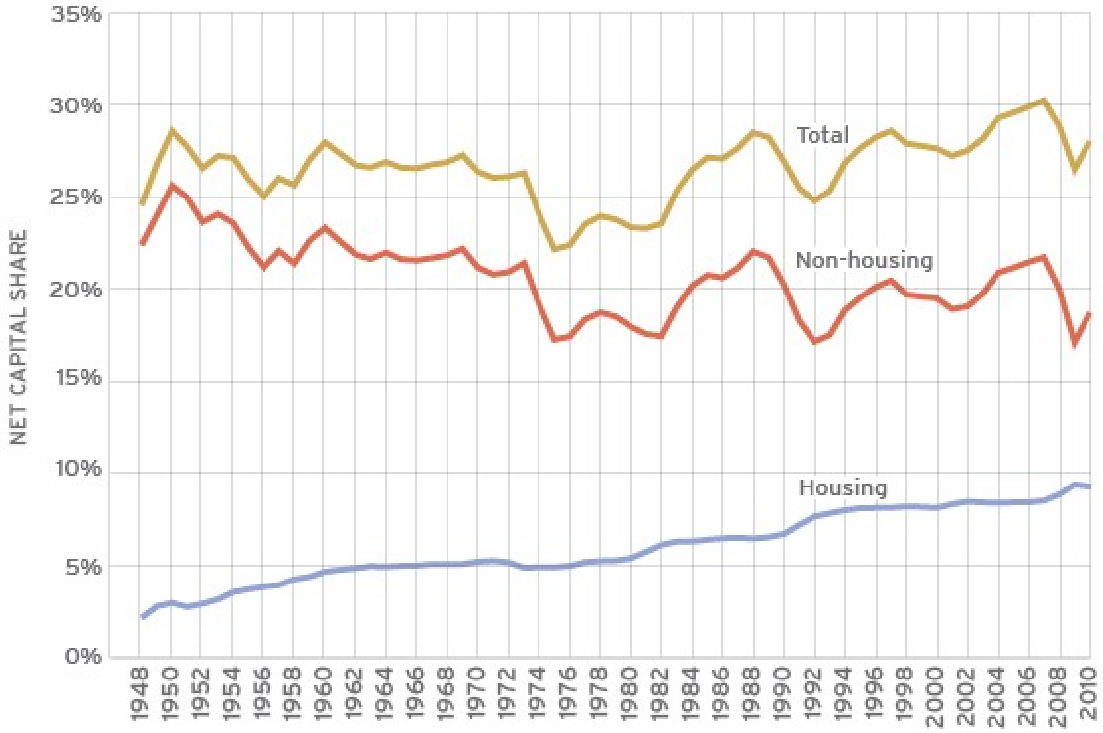
Highlights From The Comments On Piketty

Chris Stucchio recommended Matt Rognlie's criticisms of Piketty (paper, summary, Voxsplainer).

Rognlie starts by saying that Piketty didn't correctly account for capital depreciation (ie capital losing value over time) in his calculations. This surprises me, because Piketty says he does in his book (p. 55) but apparently there are technical details I don't understand. When you do that, the share of capital decreases, and it becomes clear that 100% of recent capital-share growth comes from one source: housing.

IS CAPITAL INCOME DISPLACING LABOR INCOME? ONLY IF YOU COUNT HOUSING

Net Capital Income as Shares of Total Private Domestic Net Value Added in the
U.S., Canada, Germany, France, UK, Italy, and Japan
1948-2010



*U.S., France, and U.K. data cover full period (1948-2010); Japan data starts in 1955, Canada in 1960, Italy in 1990, Germany in 1991.

Note: This is the unweighted average of shares across all seven countries. To control for the changing composition of the sample, it displays time fixed effects from a panel regression, normalized to equal the actual average at the start of the sample.

For Canada and Japan, the "residential housing sector" is the owner-occupied housing sector due to data limitations.
For all other countries, it covers all housing.

Source: Country-level national accounts publications

BROOKINGS

I can't find anyone arguing that Rognlie is wrong. I do see many people arguing about the implications, all the way from "this disproves Piketty" to "this is just saying the same thing Piketty was".

I think it's saying the same thing Piketty was in that housing is a real thing, and if there's inequality in housing, then that's real inequality. And landlords are a classic example of the rentiers Piketty is warning against.

But it's saying a different thing in that most homeowners use their homes by living in them, not by renting them out. That means they're not part of Piketty's rentier class, and so using the amount of capital to represent the power of rentiers is misleading. Rentiers are not clearly increasing and there is no clear upward trend in rentier-vs-laborer inequality. I think this *does* disprove Piketty's most shocking thesis.

Rognlie also makes an argument for why increasing the amount of capital will decrease the returns on capital, leading to stable or decreasing income from capital. Piketty argues against this on page 277 of his book, but re-reading it Piketty's argument now looks kind of weak, especially with the evidence from housing affecting some of his key points.

Grendel Khan highlights the role of housing with an interesting metaphor:

> Did someone say housing?
>
> As an illustration, the median homeowner in about half of the largest metros made more off the appreciation of their home than a full-time minimum-wage job. It's worst in California, of course; in San Jose, the median homeowner made just shy of $100 per working hour.
>
> See also Richard Florida's commentary. See also everything about how the housing crisis plays out in micro; it is precisely rentier capitalism.

In the original post, I questioned Piketty's claim that rich people and very-well-endowed colleges got higher rates of return on their investment than ordinary people or less-well-endowed colleges. After all, why can't poorer people pool their money together, mutual-fund-style, to become an effective rich person who can get higher rate of return? Many people tried to answer this, not always successfully.

brberg points out that Bill Gates – one example of a rich person who's gotten 10%+ returns per year – has a very specific advantage:

> Not sure about Harvard's endowment, but it's worth noting that the reason Gates, Bezos, Zuckerberg, and other self-made billionaires have seen their fortunes grow so quickly is that each of them has the vast majority of their wealth invested in a single high-growth company.
>
> This is an extremely high-risk investment strategy that has the potential to pay off fantastically well in a tiny percentage of cases, but it's not really dependent on the size of the starting stake. Anyone who invested in Microsoft's IPO would have seen the same rate of return as Gates.

This is a good point, but most of Piketty's data focuses on college endowments. How do they do it?

Briefling writes:

> I'm not sure you can take the wealth management thing at face value. The stock market since 1980 has 10% annualized returns. Instead of trying to replicate whatever Harvard and Yale are doing, why don't you just put your money in the stock market?

Also a good point, but colleges seem to do this with less volatility than the stock market, which still requires some explanation.

Tyrathalis, a financial planner, adds more information:

> One of the things that having /any/ major financial planner does for you, though, is it opens up access to

private equity funds that are only advertised to sufficiently high-net-worth individuals and businesses. The primary asset class that super high gains come from is private equity, generally meaning investments in angel funds and off-market startups. The way these funds operate involves you pledging a certain amount of money that they can invest as they choose, but they only call up parts of it periodically. This means that dealing with a few really rich people is much easier than dealing with a ton of poor people, in particular because it is really, really bad if they can't manage to get all of the money. Their current business model requires only dealing with people who will definitely be able to make their payments when they need to, and since the funds are so large, that means they need to have a few very rich investors. Investment advisors known to advise large fortunes are where they go to find those people.

Also, any given private equity fund is still likely to make a negative return, which is a much bigger deal if you don't have a lot of money in the first place, so very few people would recommend that you invest in a private equity fund instead of something safer if you aren't already rich. Higher returns implies higher standard deviation. That's also why a long time horizon is so significant. The basic activity of asset class investing is to diversify to balance out high variability without diminishing returns too much, but over a long enough time frame the variability matters much less and you can afford to make riskier investments.

Although, getting 10% returns doesn't require any special connections. The stock market grows at 11% a year, it just has very high variability, so you need to be able to be in the market for several decades to ensure those gains with an all-stock portfolio. A 60/40 split of

> stocks and bonds will get around 8%, while not requiring more than a few dollars to invest. You can do it on Schwab with only a bit of research. The reason why super rich people and organizations /only/ get 10% returns is that despite private equity managing 20% or more, even they don't have enough capital and long enough time horizons to stay fully invested in such risky markets. They diversify heavily too, cutting returns in favor of making those returns basically guaranteed.
>
> My main point is that financial planners do things besides stock picking, but one of the things they do is get you into private equity funds, which are the main source of the better returns that rich people can get. However, for reasons of risk management, this isn't something people who aren't super rich necessarily ought to imitate. There are only slightly less effective strategies that anyone could imitate, but its not smart for everyone to have the same amount of risk. Realistically, most people ought to do something like the 60/40 split I mentioned, and the difference between that and what most people end up getting is due to people being bad at performing optimal strategies even when they know what they are.

And Vaniver adds:

> There's a mutual fund called the Magellan Fund, which was famous for its extreme performance (I believe it was annual growth of 15-20% per year) for about 20 years.
>
> At the end of that streak, someone ran the numbers and discovered that most of the people who had invested in the fund had lost money, because they bought in when the market was high and sold when the market was low.

> The problem that mutual funds have is that they don't know how much money they're going to have tomorrow, because there are thousands upon thousands of customers who might want some of their money back, or might want to add in some more money, and as a result there are lots of unplanned trades they'll have to make that only benefit their customers, not them. Many of the best managers insist on terms of the form "you give me money and then can't take it out for N years" so that they don't have to deal with this kind of thing (in the short term, at least).
>
> For private equity servicing one large customer, there are far fewer moves of that form, and they're much easier to predict, and you averaging across many small customers still doesn't duplicate that effect.

I still don't feel like this explains everything; surely a college with $500 million has about the same risk tolerance and ability to give money on the right time scale as a college with $1 billion? Maybe all of this is just false? J Mann writes:

> Is it consensus that Harvard and Yale consistently get better returns than other endowments and than the market? It looks like Harvard at least has had a number of recent bad years, and that some people are suggesting that its results may be based on taking on more risk.

And Anon256 adds:

> Indeed; Havard has done badly enough in the years since Piketty's book was published that it's now considering switching to just using index funds.

And Will4071 says:

> Just to note, I don’t think large endowments/the very rich really do anything special. This analysis suggests that they actually underperform a levered 60/40 portfolio (which is fairly standard, and something you could easily set up yourself).

Chris Stucchio has a different perspective on rich people making higher rates of return:

> It’s also worth reflecting on a point which Piketty makes mathematically, but literally never says in words. If rich people are the best investors, then the best way to create economic growth is to ensure that rich people are the ones controlling investment decisions. Intuitively this makes a lot of sense; Travis Kalanick (and now Dara “the D” Khosrowshahi) are a lot better at transportation than the average autowale. Bezos is a lot better at logistics than my local cell phone store.

See also Paul’s answer to one of my objections to this. Right now it looks like (assuming Piketty is right about this at all), Chris has a point. Does anyone want to try to convince me otherwise?

Phillip Magness, himself an economic history professor, writes:

> I’d also urge you to look more skeptically on his income distribution stats (the figure 1.1 above). Several economists, myself included, have been working on the measurement problems that arise from attempting to determine income shares from tax data in recent years. The aforementioned figure comes from a 2003 study by Piketty and his coauthor Emmanuel Saez. While it represented an innovative contribution to the literature, this paper gives generally insufficient treatment to the

effect of changes to the tax code itself upon data that derive from income tax reporting.

To put it another way, taxpayers – both wealthy and poor – respond to the way that income tax laws are structured so as to minimize their own tax burdens. They take advantage of incentives and loopholes to lower what they owe. They engage in wealth planning strategies to legally shelter income from high rates of taxation. And some even illegally evade their obligations by misreporting income.

Tax avoidance and evasion rates vary substantially over time and in response to tax code changes, and so do the statistics they generate with the IRS. A major problem in Piketty-Saez is that they do very little to account for this issue over time, and instead simply treat tax-generated stats as if they are representative. Doing so yields a relatively sound measurement of income distributions, provided that the tax code remains relatively stable over long periods of time (e.g. what the U.S. experienced between roughly 1946 and 1980). When the tax code undergoes frequent and major changes though, tax-generated stats become less reliable. And it just so happens that the two periods of “high” inequality on the Piketty-Saez U-curve are also periods of volatility in the tax code: 1913-1945 and 1980-present.

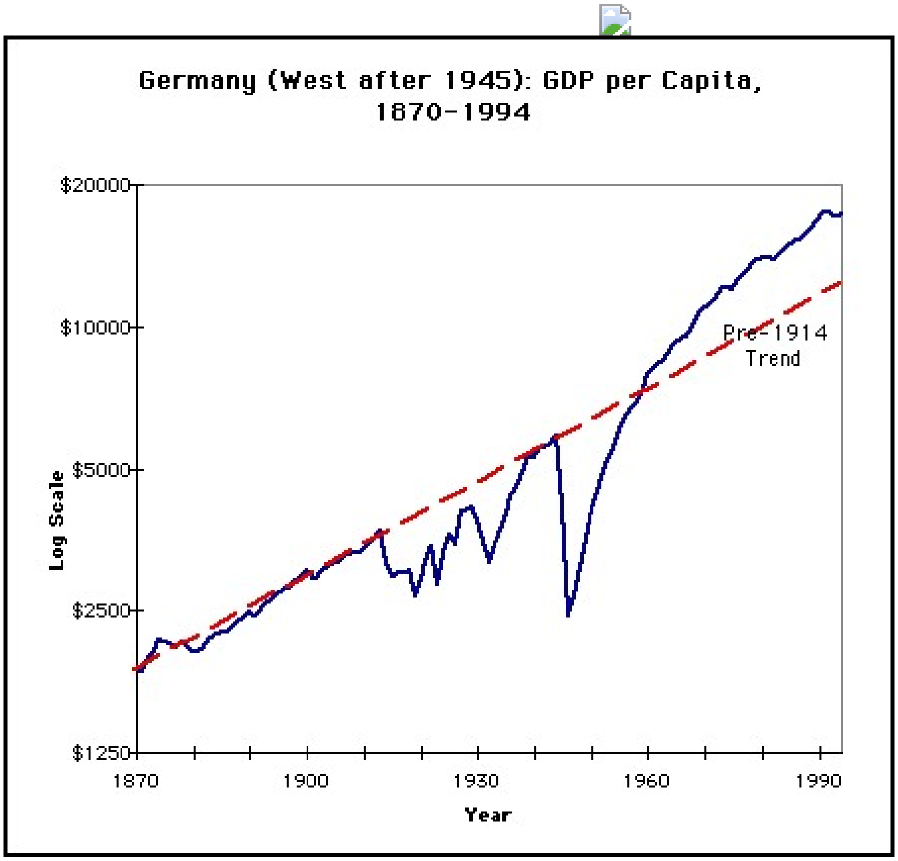
The 1913-45 period is marred by both frequent tax rate swings and an initially small tax base that was rapidly expanded during WWII, combined with the introduction of automatic payroll withholding in 1943. When you account for these and related issues, the extreme inequality of the early 20th century and especially the severe drop it undergoes between 1941-45 become much more subdued. The period from 1980-present is

> similarly marred by Piketty and Saez's failure to fully account for the effects of the Tax Reform Act of 1986, which induced substantial income shifting at the top of the distribution to take advantage of differences between the personal and corporate tax rates. Adjusting for that has a similar effect of lowering the depicted rebound.
>
> Taken together, what we're probably experiencing is a much flatter trend across the 20th century – one that resembles a tea saucer rather than a pronounced U. And that has profound implications for Piketty's larger prescriptive argument in favor of highly progressive tax rates.

Magness also recommends his 2014 paper and Richard Sutch's 2017 conceptual replication questioning Piketty's data. It's inherently hard to find good data on inequality over the last few centuries, but Magness finds that of the many datasets available, Piketty cherry-picked the ones that best fit the u-shaped curve he wanted to show, estimated some missing data points kind of out of thin air, and made some other questionable decisions. The result is a much less pronounced change in inequality, especially in the US.

The paper is pretty confrontational (on his own blog, Magness' co-author describes Piketty as making "no-brainers…boneheaded historical errors [that] would be shocking if contained in a high school term paper". Piketty sort of says a few words in his own defense in this article. But one thing I notice is that it looks like, aside from these authors, everyone is working together on this – the author of one of the pro-Piketty datasets was also a co-author of one of the anti-Piketty datasets, and the author of one of the anti-Piketty datasets has worked with Piketty in the past. This suggests to me that a lot of this is legitimately hard and that the same people, working from different methods, get different results. My main takeaway is that there are many different inequality datasets and Piketty used the most dramatic.

Tlaloc on the Discord provides the European log GDP graphs I wanted:



I think it’s fair to ask – what the heck? Taken literally, doesn’t this suggest WWII was long-run good for Europe – that its “recovery” brought it well above trend?

Eyeballing the Maddison Project data elsewhere shows France, Germany, and the US all having very similar growth of 200% between 1960 and 2016.

I need to look into this more, but right now I'm not really buying it.

VPaul doesn't believe in straight-line GDP growth anyway:

> I don't trust inflation statistics, so I don't trust inflation adjusted GDP statistics. During the time period covered by Piketty's GDP growth trend line, there have multiple different methodologies for measuring inflation, with adjustments to fix obvious errors in previous versions of inflation adjusters. Since we know inflation statistics have been wrong, and there is good evidence they are still wrong, I think the steady GDP growth rate is an artifact.

Several people point out that "increasing number of rentiers" is not necessarily bad; after all, this is what the post-scarcity robot future should look like. For example, from Virriman:

> A world where 1% of people can avoid drudgery seems preferable to a world where only 0.1% can do that, holding everything else equal. Isn't the techno-utopian ideal a world where almost everyone is a "rentier"?
>
> Sounds like we need to figure out how to get back to the gilded age, and then figure out how to turn that 1% of rentiers into 2% and keep trying to expand that number.

This could *maybe* make sense around number of rentiers, but amount of money per rentier could work the opposite direction, and Piketty's numbers awkwardly combine both.

Paul Christiano on some of Piketty's other statistics:

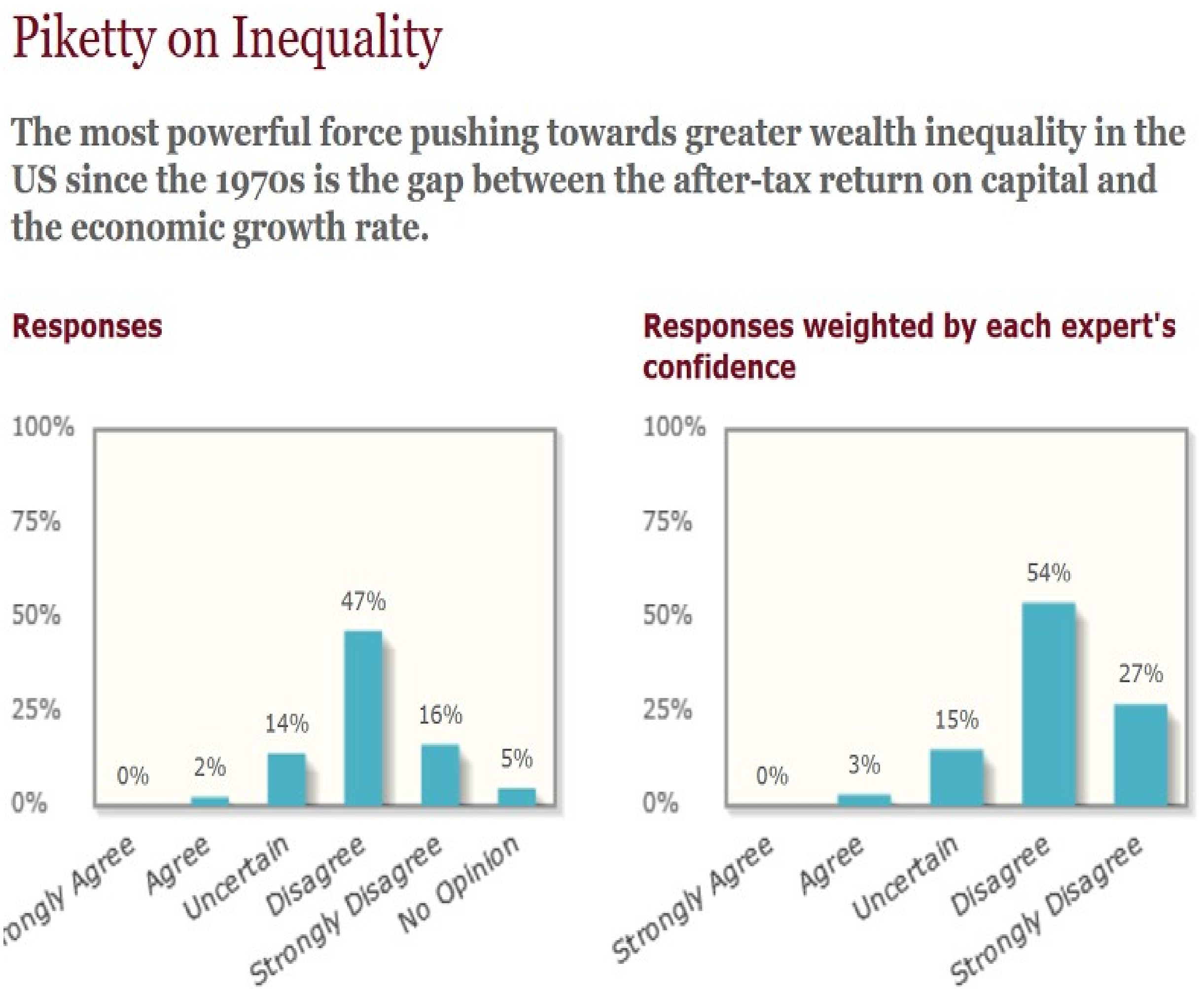
The extrapolation in figure 10.11 looks pretty wild. It takes a special something to draw a graph that's been pretty smooth/predictable historically, then insert a stark+unprecedented regime change exactly at the current moment for no apparent reason. Does he give some justification for the sharp discontinuity?

Claiming that economic growth is always 1-1.5% also seems pretty dubious. According to Maddison's estimates, which I don't think are under dispute, worldwide per capita growth first reached 1% around 1900, continued increasing to 2-3% by 1960, and then fell back down to 1% in the great stagnation. You could say "A century is a long time, that's basically always, the mid-century spike was just a deviation" but elsewhere Pikkety seems willing to write off that same chunk of history as an aberration. Or maybe his argument is supposed to apply only to the US? (Or maybe he includes Europe and then can cite steady growth for 150 years instead of 100? I don't even think that's true though, in 1875 I think that per capita GDP growth in Europe was not yet 1%?)

I'm not sure in what sense rentiers can be said to be winning. We can just look directly and see that rents are significantly smaller than wages, the capital share of income is staying around 1/3, it's grown but only a tiny bit. If 1/3 of GDP is rents that get allocated inequitably then maybe you can increase median income by 25% with perfect redistribution, but that just doesn't seem that promising compared to efficiency effects, unless you are super concerned about inequality per se (rather than regarding it as an opportunity to benefit poorer people). Even that benefit would shrink as savings rates fall.

> If in fact the rentiers grow their fortunes at r, then they will get wealthier and wealthier until r = g, that’s basically an accounting identity. That seems to basically be a reductio of the concern that r>>g can continue indefinitely + rentiers can have their wealth grow at the rate r.
>
> From an efficiency standpoint it seems like the main implication of r>>g is that we could spend 1% of GDP today to make our descendants several percent richer, which sounds like a good deal and suggests that we ought to invest more. It’s pretty wild to respond to r>>g by considering massively disincentivizing investment. If you want to push for equality and think that r>>g, maybe support a sovereign wealth fund? Or else we’d need to decide collectively whether the problem with inequality is that some people are rich, or that other people are poor—I can see how a wealth tax (vs a similarly large consumption or income tax) would help with one of those problems, but not the other. I think it’s just a really bad policy for a lot of reasons with very little to recommend it other than leveling down.

Swami brings up an IGM poll of economists on r>>g:



Piketty on Inequality

The most powerful force pushing towards greater wealth inequality in the US since the 1970s is the gap between the after-tax return on capital and the economic growth rate.

Responses

Responses weighted by each expert's confidence

ADifferentAnonymous counters with a Matt Yglesias article arguing that this isn't really disproving anything Piketty is saying.

Overall, it looks like the claim that the super-rich get much better returns on investment than everyone else doesn't really hold up, except in obvious predictable ways, eg they can take more risks.

The claim that there is a rising rentier class who will dominate the 21st century doesn't really hold up.

I'm not qualified to say whether Piketty's empirical data holds up, but there seems to be significant academic debate over it.

And although Piketty’s rules of thumb for growth (g = 1 – 1.5%, r = 4-5%) hold up more than I would have expected before reading him, they still don’t hold up that well.

Now taking recommendations about if anything from Piketty is still worth keeping.

SSC Journal Club: Dissolving The Fermi Paradox

I’m late to posting this, but it’s important enough to be worth sharing anyway: Sandberg, Drexler, and Ord on Dissolving the Fermi Paradox.

(You may recognize these names: Toby Ord founded the effective altruism movement; Eric Drexler kindled interest in nanotechnology; Anders Sandberg helped pioneer the academic study of x-risk, and wrote what might be my favorite *Unsong* fanfic)

The Fermi Paradox asks: given the immense number of stars in our galaxy, for even a very tiny chance of aliens per star shouldn’t there should be thousands of nearby alien civilizations? But any alien civilization that arose millions of years ago would have had ample time to colonize the galaxy or do something equally dramatic that would leave no doubt as to its existence. So where are they?

This is sometimes formalized as the Drake Equation: think up all the parameters you would need for an alien civilization to contact us, multiply our best estimates for all of them together, and see how many alien civilizations we predict. So for example if we think there’s a 10% chance of each star having planets, a 10% chance of each planet being habitable to life, and a 10% chance of a life-habitable planet spawning an alien civilization by now, one in a thousand stars should have civilization. The actual Drake Equation is much more complicated, but most people agree that our best-guess values for most parameters suggest a vanishingly small chance of the empty galaxy we observe.

SDO’s contribution is to point out this is the wrong way to think about it. Sniffnoy’s comment on the subreddit helped me understand exactly what was going on, which I think is something like this:

Imagine we knew God flipped a coin. If it came up heads, He made 10 billion alien civilization. If it came up tails, He made none besides Earth. Using our one parameter Drake Equation, we determine that *on average* there should be 5 billion alien civilizations. Since we see zero, that’s quite the paradox, isn’t it?

No. In this case the mean is meaningless. It’s not at all surprising that we see zero alien civilizations, it just means the coin must have landed tails.

SDO say that relying on the Drake Equation is the same kind of error. We’re not interested in the *average* number of alien civilizations, we’re interested in the distribution of probability over number of alien civilizations. In particular, what is the probability of few-to-none?

SDO solve this with a “synthetic point estimate” model, where they choose random points from the distribution of possible estimates suggested by the research community, run the simulation a bunch of times, and see how often it returns different values.

According to their calculations, a standard Drake Equation multiplying our best estimates for every parameter together yields a probability of less than one in a million billion billion billion that we’re alone in our galaxy – making such an observation pretty paradoxical. SDO’s own method, taking account parameter uncertainty into account, yields a probability of one in three.

They try their hand at doing a Drake calculation of their own, using their preferred values, and find:

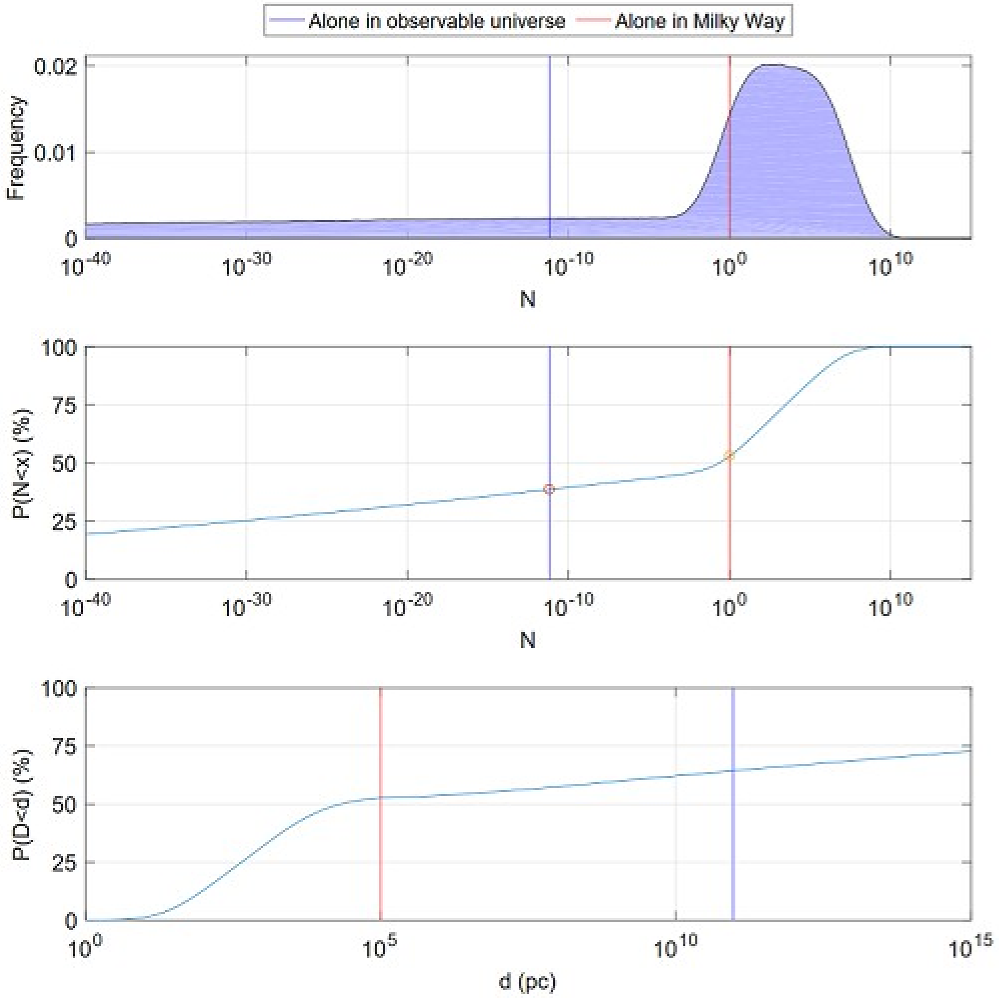


Figure 2: (A) A probability density function for N generated by Monte Carlo simulation based on the authors' best estimates of our current uncertainty for each parameter. (B) The corresponding cumulative density function. (C) A cumulative density function for the distance to the nearest detectable civilization.

N is the average number of civilizations per galaxy

If this is right – and we can debate exact parameter values forever, but it's hard to argue with their point-estimate-vs-distribution-logic – then there's no Fermi Paradox. It's done, solved, kaput. Their title, "Dissolving The Fermi Paradox", is a strong claim, but as far as I can tell they totally deserve it.

"Why didn't anyone think of this before?" is the question I am only slightly embarrassed to ask given that *I* didn't think of it before. I don't know. Maybe people thought of it before, but didn't publish it, or published it somewhere I don't know about? Maybe people intuitively figured out what was up (one of the parameters of the Drake Equation must be much lower than our estimate) but stopped there and didn't bother explaining the formal probability argument. Maybe nobody took the Drake Equation seriously anyway, and it's just used as a starting point to discuss the probability of life forming?

But any explanation of the "oh, everyone knew this in some sense already" sort has to deal with that a lot of very smart and well-credentialled experts treated the Fermi Paradox very seriously and came up with all sorts of weird explanations. There's no need for sci-fi theories any more (though you should still read the Dark Forest trilogy). It's just that there aren't very many aliens. I think my past speculations on this, though very incomplete and much inferior to the recent paper, come out pretty well here.

(some more discussion here on Less Wrong)

One other highlight hidden in the supplement: in the midst of a long discussion on the various ways intelligent life can fail to form, starting on page 6 the authors speculate on "alternative genetic systems". If a planet gets life with a slightly different way of encoding genes than our own, it might be too unstable to allow complex life, or too stable to allow a reasonable rate of mutation by natural selection. It may be that abiogenesis can only create very weak genetic codes, and life needs to go through several "genetic-genetic transitions" before it can reach anything capable of complex evolution. If this is path-dependent – ie there are branches that are local improvements but close off access to other better genetic systems – this could permanently arrest the development of life, or freeze it at an evolutionary rate so low that the history of the universe so far is too short a time to see complex organisms.

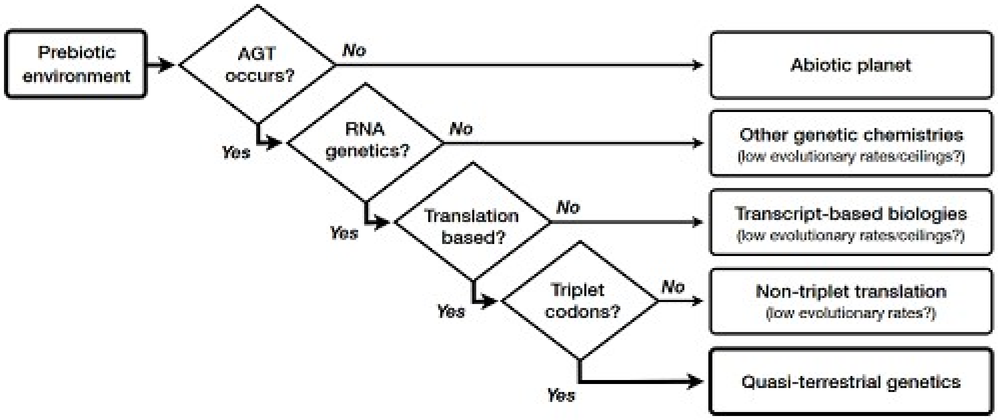

Prebiotic environment
AGT occurs?
No
Abiotic planet
Yes
RNA genetics?
No
Other genetic chemistries (low evolutionary rates/ceilings?)
Yes
Translation based?
No
Transcript-based biologies (low evolutionary rates/ceilings?)
Yes
Triplet codons?
No
Non-triplet translation (low evolutionary rates?)
Yes
Quasi-terrestrial genetics

Figure 1. A schematic diagram of potential AGT and GGT branches with aggregated states and transitions. At each branch point, outcome ratios are weakly constrained, with LU values arguably > 100.

I don’t claim to understand all of this, but the parts I do understand are fascinating and could easily be their own paper.

The Craft And The Codex

The rationalist community started with the idea of rationality as a martial art – a set of skills you could train in and get better at. Later the metaphor switched to a craft. Art or craft, parts of it did get developed: I remain very impressed with Eliezer’s work on how to change your mind and everything presaging Tetlock on prediction.

But there’s a widespread feeling in the rationalist community these days that this is the area where we’ve made the least progress. AI alignment has grown into a developing scientific field. Effective altruism is big, professionalized, and cash-rich. It’s just the art of rationality itself that remains (outside the usual cognitive scientists who have nothing to do with us and are working on a slightly different project) a couple of people writing blog posts.

Part of this is that the low-hanging fruit has been picked. But I think another part was a shift in emphasis.

Martial arts does involve theory – for example, beginning fencers have to learn the classical parries – but it’s a little bit of theory and a lot of practice. Most of becoming a good fencer involves either practicing the same lunge a thousand times in ideal conditions until you could do it in your sleep, or fighting people on the strip.

I’ve been thinking about what role this blog plays in the rationalist project. One possible answer is “none” – I’m not enough of a mathematician to talk much about the decision theory and machine learning work that’s really important, and I rarely touch upon the nuts and bolts of the epistemic rationality craft. I freely admit that (like many people) I tend to get distracted by the latest Outrageous Controversy, and so spend way too much time discussing things like Piketty's theory of inequality which get more attention from the chattering classes but are maybe less important to the very-long-run future of the world.

Any argument in my own defense is entirely post hoc. But if I can advance such an argument anyway, it would be that this kind of thing is the endless drudgery of rationality training, the equivalent of fighting a thousand bouts and honing your reflexes. Controversial things are, at least, hard problems. There's a lot of misinformation and conflicting interpretations and differing heuristics and compelling arguments on both sides. Figuring out what's going on with Piketty is good practice for figuring out what's going on with deworming etc.

Looking back on the Piketty discussion, people brought up questions like "How much should you discount a compelling-sounding theory based on the bias of its inventor?" And "How much does someone being a famous expert count in their favor?" And "How concerned should we be if a theory seems to violate efficient market assumptions?" And "How do we balance arguments based on what rationally has to be true, vs. someone's empirical but fallible data sets?"

And in the end, I think we made a lot of progress on those questions. With the help of some very expert commenters, I resolved a lot of my confusions and changed some of my conclusions. That not only gives me a different view of Piketty, but – I hope – long-term trains my thought processes to better understand which heuristics and generators-of-heuristics are reliable in which situations.

Last year, I had a conversation with a friend over how we should think about the latest round of scientific results. I said over the past few years I'd learned to trust science more; he said he'd learned to trust science less. We argued it for a while, and in the end I think we basically had the same insights and perspectives – there are certain situations where science is very definitely trustworthy, and others where it is very definitely untrustworthy. Although I could provide heuristics about which is which, they would be preliminary and much worse than the intuitions that generated them. I live in fear of someone asking something like "So, since all the prominent scientists were wrong about social priming, isn't it plausible that all the prominent scientists are wrong about homeopathy?" I can come up with some reasons this isn't the right way to look at things, but my real answer would have to sound more like "After years of looking into this kind of thing, I think I have some pretty-good-though-illegible intuitions about when science can be wrong, and homeopathy isn't one of those times."

I think by looking at a lot of complicated cases, and checking back on them after they're solved (which sometimes happens! Just look at the Fermi Paradox paper from earlier this week!) we can refine those intuitions and get a better idea of how to use the explicit-textbook-rationality-techniques. If this blog still has value to the rationalist project, it's as a dojo where we do this a couple of times a week and absorb the relevant results.

This is one reason I'm so grateful for everyone's comments. I only post a Comments Highlights thread every so often, but I'm constantly making updates based on things I read there and getting a chance to double-check which of the things I think are right or wrong. This isn't just good individual rationality practice, it's also community rationality practice, and so far I'm pretty happy with how it's going.

Melatonin: Much More Than You Wanted To Know

[I am not a sleep specialist. Please consult with one before making any drastic changes or trying to treat anything serious.]

Van Geiklswijk et al describe supplemental melatonin as "a chronobiotic drug with hypnotic properties". Using it as a pure hypnotic – a sleeping pill – is like using an AK-47 as a club to bash your enemies' heads in. It might work, but you're failing to appreciate the full power and subtlety available to you.

Melatonin is a neurohormone produced by the pineal gland. In a normal circadian cycle, it's lowest (undetectable, less than 1 pg/ml of blood) around the time you wake up, and stays low throughout the day. Around fifteen hours after waking, your melatonin suddenly shoots up to 10 pg/ml – a process called "dim light melatonin onset". For the next few hours, melatonin continues to increase, maybe as high as 60 or 70 pg/ml, making you sleepier and sleepier, and presumably at some point you go to bed. Melatonin peaks around 3 AM, then declines until it's undetectably low again around early morning.

Is this what makes you sleepy? Yes and no. Sleepiness is a combination of the circadian cycle and the so-called "Process S". This is an unnecessarily sinister-sounding name for the fact that the longer you've been awake, the sleepier you'll be. It seems to be partly regulated by a molecule called adenosine. While you're awake, the body produces adenosine, which makes you tired; as you sleep, the body clears adenosine away, making you feel well-rested again.

In healthy people these processes work together. Circadian rhythm tells you to feel sleepy at night and awake during the day. Process S tells you to feel awake when you've just risen from sleep (naturally the morning), and tired when you haven't slept in a long time (naturally the night). Both processes agree that you should feel awake during the day and tired at night, so you do.

When these processes disagree for some reason – night shifts, jet lag, drugs, genetics, playing *Civilization* until 5 AM – the system fails. One process tells you to go to sleep, the other to wake up. You're never quite awake enough to feel energized, or quite tired enough to get restful sleep. You find yourself lying in bed tossing and turning, or waking up while it's still dark and not being able to get back to sleep.

Melatonin works on both systems. It has a weak "hypnotic" effect on Process S, making you immediately sleepier when you take it. It also has a stronger "chronobiotic" effect on the circadian rhythm, shifting what time of day your body considers sleep to be a good idea. Effective use of melatonin comes from understanding both these effects and using each where appropriate.

1. Is melatonin an effective hypnotic?

Yes.

That is, taking melatonin just before you want to get to sleep, does help you get to sleep. The evidence on this is pretty unanimous. For primary insomnia, two meta-analyses – one by Brzezinski in 2005 and another by Ferracioli-Oda in 2013 – both find it safe and effective. For jet lag, a meta-analysis by the usually-skeptical Cochrane Collaboration pronounces melatonin "remarkably effective". For a wide range of primary and secondary sleep disorders, Buscemi et al say in their abstract that it doesn't work, but a quick glance at the study shows it absolutely does and they are incorrectly under-reporting their own results. The Psychiatric

Times agrees with me on this: “Results from another study reported as negative actually demonstrated a statistically significant positive result of a decrease in sleep latency by an average of 7.2 minutes for melatonin”.

Expert consensus generally follows the meta-analyses: melatonin works. I find cautious endorsements by the Mayo Clinic and John Hopkins less impressive than its less-than-completely-negative review on Science-Based Medicine, a blog I can usually count on for a hit job on any dietary supplement.

The consensus stresses that melatonin is a very weak hypnotic. The Buscemi meta-analysis cites this as their reason for declaring negative results despite a statistically significant effect – the supplement only made people get to sleep about ten minutes faster. “Ten minutes” sounds pretty pathetic, but we need to think of this in context. Even the strongest sleep medications, like Ambien, only show up in studies as getting you to sleep ten or twenty minutes faster; this *New York Times* article says that “viewed as a group, [newer sleeping pills like Ambien, Lunesta, and Sonata] reduced the average time to go to sleep 12.8 minutes compared with fake pills, and increased total sleep time 11.4 minutes.” I don’t know of any statistically-principled comparison between melatonin and Ambien, but the difference is hardly (pun not intended) day and night.

Rather than say “melatonin is crap”, I would argue that all sleeping pills have measurable effects that vastly underperform their subjective effects. The linked article speculates on one reason this might be: people have low awareness around the time they get to sleep, and a lot of people’s perception of whether they’re insomniac or not is more anxiety (or sometimes literally dream) than reality. This is possible, but I also think of this in terms of antidepressant studies, which find similarly weak objective effects despite patients (and doctors) who swear by them and say they changed their lives. If I had to guess, I would say that the studies include an awkward combination of sick and less-sick people *and* confuse responders and non-responders. Maybe this is special pleading. I don't know. But if you think any sleeping pill works well, melatonin doesn't necessarily work much worse than that.

Sleep latency statistics are hard to compare to one another because they're so dependent on the study population. If your subjects take an hour to fall asleep, perhaps melatonin could shave off thirty-four minutes. But if your subjects take twenty minutes to fall asleep, then no sleeping pill will ever take off thirty-four minutes, and even an amazing sleeping pill might struggle to make fifteen. I cannot directly compare the people who say melatonin gives back ten minutes to the people who say melatonin gives back thirty-four minutes to the people who say Ambien gives back twelve, but my totally unprincipled guess is that melatonin is about a third as strong as Ambien. It also has about a hundred times fewer side effects, so there's definitely a place for it in sleep medicine.

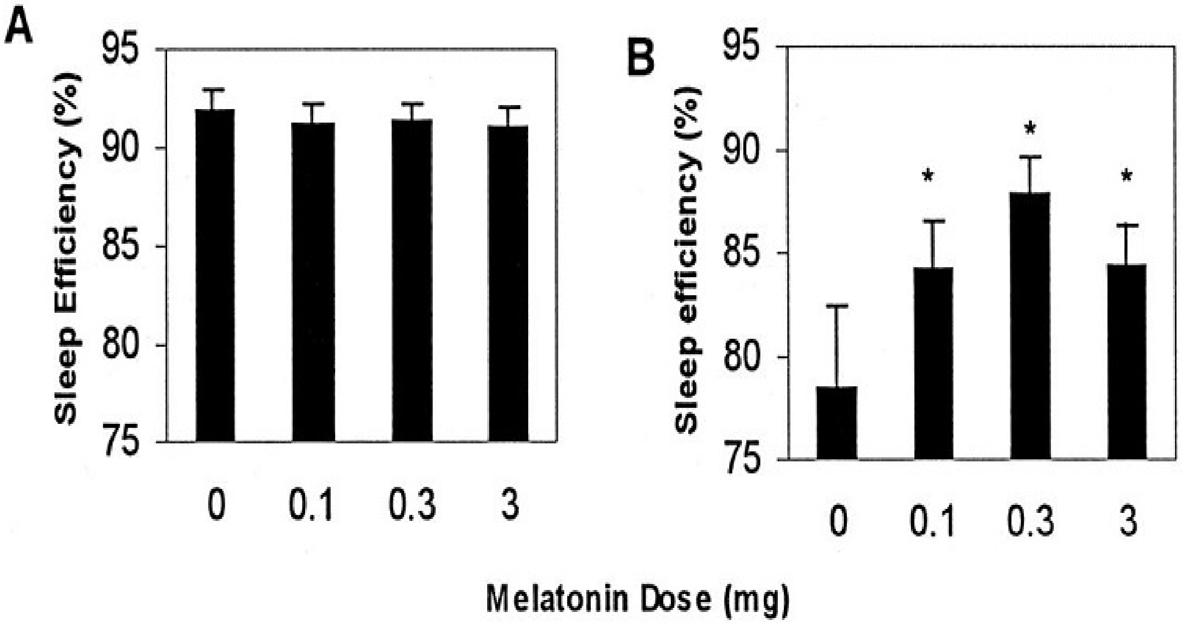
2. What is the right dose of melatonin?

0.3 mg.

"But my local drugstore sells 10 mg pills! When I asked if they had anything lower, they looked through their stockroom and were eventually able to find 3 mg pills! And you're saying the correct dose is a third of a milligram?!"

Yes. Most existing melatonin tablets are around ten to thirty times the correct dose.

Many early studies were done on elderly people, who produce less endogenous melatonin than young people and so are considered especially responsive to the drug. Several lines of evidence determined that 0.3 mg was the best dose for this population. Elderly people given doses around 0.3 mg slept better than those given 3 mg or more and had fewer side effects (Zhdanova et al 2001). A meta-analysis of dose-response relationships concurred, finding a plateau effect around 0.3 mg, with doses after that having no more efficacy, but worse side effects (Brzezinski et al, 2005). And doses around 0.3 mg cause blood melatonin spikes most similar in magnitude and duration to the spikes seen in healthy young people with normal sleep (Vural et al, 2014).



Other studies were done on blind people, who are especially sensitive to melatonin since they lack light cues to entrain their circadian rhythms. This is a little bit of a different indication, since it’s being used more as a chronobiotic than a sleeping pill, but the results were very similar: lower doses worked better than higher doses. For example, in Lewy et al 2002, nightly doses of 0.5 mg worked to get a blind subject sleeping normally at night; doses of 20 mg didn’t. They reasonably conclude that the 20 mg is such a high dose that it stays in their body all day, defeating the point of a hormone whose job is to signal nighttime. Other studies on the blind have generally confirmed that doses of around 0.3 to 0.5 mg are optimal.

There have been disappointingly few studies on sighted young people. One such, Attenburrow et al 1996 finds that 1 mg works but 0.3 mg doesn’t, suggesting these people may need slightly higher doses, but this study is a bit of an outlier. Another Zhdanova study on 25 year olds found both to work equally. And Pires et al studying 22-24 year olds found that 0.3 mg worked better than 1.0. I am less interested in judging the 0.3 mg vs. 1.0 mg debate than in pointing out that both numbers are much lower than the 3 – 10 mg doses found in the melatonin tablets sold in drugstores.

UpToDate, the gold standard research database used by doctors, agrees with these low doses. “We suggest the use of low, physiologic doses (0.1 to 0.5 mg) for insomnia or jet lag (Grade 2B). High-dose preparations raise plasma melatonin concentrations to a supraphysiologic level and alter normal day/night melatonin rhythms.” Mayo Clinic makes a similar recommendation: they recommend 0.5 mg. John Hopkins’ experts almost agree: they say “less is more” but end up chickening out and recommending 1 to 3 mg, which is well above what the studies would suggest.

Based on a bunch of studies that either favor the lower dose or show no difference between doses, plus clear evidence that 0.3 mg produces an effect closest to natural melatonin spikes in healthy people, plus UpToDate usually having the best recommendations, I’m in favor of the 0.3 mg number. I think you could make an argument for anything up to 1 mg. Anything beyond that and you’re definitely too high. Excess melatonin isn’t grossly dangerous, but tends to produce tolerance and might mess up your chronobiology in other ways. Based on anecdotal reports and the implausibility of becoming tolerant to a natural hormone at the dose you naturally have it, I would guess sufficiently low doses are safe and effective long term, but this is just a guess, and most guidelines are cautious in saying anything after three months or so.

3. What are circadian rhythm disorders? How do I use melatonin for them?

Circadian rhythm disorders are when your circadian rhythm doesn’t match the normal cycle where you want to sleep at night and wake up in the morning.

The most popular circadian rhythm disorder is “being a teenager”. Teenagers’ melatonin cycle is naturally shifted later, so that they don’t want to go to bed until midnight or later, and don’t want to wake up until eight or later. This is an obvious mismatch with school starting times, leading to teenagers either not getting enough sleep, or getting their sleep at times their body doesn’t want to be asleep and isn’t able to use it properly. This is why every reputable sleep scientist and relevant scientific body keeps telling the public school system to start later.

When a this kind of late sleep schedule persists into adulthood or becomes too distressing, we call it Delayed Sleep Phase Disorder. People with DSPD don’t get tired until very late, and will naturally sleep late if given the chance. The weak version of this is “being a night owl” or “not being a morning person”. The strong version just looks like insomnia: you go to bed at 11 PM, toss and turn until 2 AM, wake up when your alarm goes off at 7, and complain you “can’t sleep”. But if you can sleep at 2 AM, consistently, regardless of when you wake up, and you would fall asleep as soon as your head hit the pillow if you first got into bed at 2, then this isn't insomnia – it's DSPD.

The opposite of this pattern is Advanced Sleep Phase Disorder. This is most common in the elderly, and I remember my grandfather having this. He would get tired around 6 PM, go to bed by 7, wake around 1 or 2 AM, and start his day feeling fresh and alert. But the weak version of this is the person who wakes up at 5 each morning even though their alarm doesn't go off until 8 and they could really use the extra two hours' sleep. These people would probably do fine if they just went to bed at 8 or 9, but the demands of work and a social life make them feel like they "ought" to stay up as late as everyone else. So they go to bed at 11, wake up at 5, and complain of "terminal insomnia".

Finally, there's Non-24-Hour-Sleep Disorder, where somehow your biological clock ended up deeply and unshakeably convinced that days on Earth are twenty-five (or whatever) hours long, and decides this is the hill it wants to die on. So if you naturally sleep 11 – 7 one night, you'll naturally sleep 12 – 8 the next night, 1 to 9 the night after that, and so on until either you make a complete 24-hour cycle or (more likely) you get so tired and confused that you stay up 24+ hours and break the cycle. This is most common in blind people, who don't have the visual cues they need to remind themselves of the 24 hour day, but it happens in a few sighted people also; Eliezer Yudkowsky has written about his struggles with this condition.

Melatonin effectively treats these conditions, but you've got to use it right.

The general heuristic is that melatonin drags your sleep time towards the direction of when you take the melatonin.

So if you want to go to sleep (and wake up) earlier, you want to take melatonin early in the day. How early? Van Geijlswijk et al sums up the research as saying it is most effective "5 hours prior to both the traditionally determined [dim light melatonin onset] (circadian time 9)". If you don't know your own melatonin cycle, your best bet is to take it 9 hours after you wake up (which is presumably about seven hours before you go to sleep).

What if you want to go to sleep (and wake up) later? Our understanding of the melatonin cycle strongly suggests melatonin taken first thing upon waking up would work for this, but as far as I know this has never been formally investigated. The best I can find is researchers saying that they think it would happen and being confused why no other researcher has investigated this.

And what about non-24-hour sleep disorders? I think the goal in treatment here is to advance your phase each day by taking melatonin at the same time, so that your sleep schedule is more dependent on your own supplemental melatonin than your (screwed up) natural melatonin. I see conflicting advice about how to do this, with some people saying to use melatonin as a hypnotic (ie just before you go to bed) and others saying to use it on a typical phase advance schedule (ie nine hours after waking and seven before sleeping, plausibly about 5 PM). I think this one might be complicated, and a qualified sleep doctor who understands your personal rhythm might be able to tell you which schedule is best for you. Eliezer says the latter regimen had very impressive effects for him (search "Last but not least" here). I'm interested in hearing from the MetaMed researcher who gave him that recommendation on how they knew he needed a phase advance schedule.

Does melatonin used this way cause drowsiness (eg at 5 PM)? I think it might, but probably such a minimal amount compared to the non-sleep-conduciveness of the hour that it doesn't register.

Melatonin isn't the only way to advance or delay sleep phase. Here is a handy cheat sheet of research findings and theoretical predictions:

TO TREAT DELAYED PHASE SLEEP DISORDER (ie you go to bed too late and wake up too late, and you want it to be earlier)
– Take melatonin 9 hours after wake and 7 before sleep, eg 5 PM
– Block blue light (eg with blue-blocker sunglasses or f.lux) after sunset
– Expose yourself to bright blue light (sunlight if possible, dawn simulator or light boxes if not) early in the morning
– Get early morning exercise
– Beta-blockers early in the morning (not generally recommended, but if you're taking beta-blockers, take them in the morning)

TO TREAT ADVANCED PHASE SLEEP DISORDER (ie you go to bed too early and wake up too early, and you want it to be later)
– Take melatonin immediately after waking
– Block blue light (eg with blue-blocker sunglasses or f.lux) early in the morning
– Expose yourself to bright blue light (sunlight if possible, light boxes if not) in the evening.
– Get late evening exercise
– Beta-blockers in the evening (not generally recommended, but if you're taking beta-blockers, take them in the evening)

These don't "cure" the condition permanently; you have to keep doing them every day, or your circadian rhythm will snap back to its natural pattern.

What is the correct dose for these indications? Here there is a lot more controversy than the hypnotic dose. Of the nine studies van Geijlswijk describes, seven have doses of 5 mg, which suggests this is something of a standard for this purpose. But the only study to compare different doses directly (Mundey et al 2005) found no difference between a 0.3 and 3.0 mg dose. The Cochrane Review on jet lag, which we'll see is the same process, similarly finds no difference between 0.5 and 5.0.

Van Geijlswijk makes the important point that if you take 0.3 mg seven hours before bedtime, none of it is going to be remaining in your system at bedtime, so it's unclear how this even works. But – well, it *is* pretty unclear how this works. In particular, I don't think there's a great well-understood physiological explanation for how taking melatonin early in the day shifts your circadian rhythm seven hours later.

So I think the evidence points to 0.3 mg being a pretty good dose here too, but I wouldn't blame you if you wanted to try taking more.

4. How do I use melatonin for jet lag?

Most studies say to take a dose of 0.3 mg just before (your new time zone's) bedtime.

This doesn't make a lot of sense to me. It seems like you should be able to model jet lag as a circadian rhythm disorder. That is, if you move to a time zone that's five hours earlier, you're in the exact same position as a teenager whose circadian rhythm is set five hours later than the rest of the world's. This suggests you should use DSPD protocol of taking melatonin nine hours after waking / five hours before DLMO / seven hours before sleep.

My guess is for most people, their new time zone bedtime *is* a couple of hours before their old bedtime, so you’re getting most of the effect, plus the hypnotic effect. But I’m not sure. Maybe taking it earlier would work better. But given that the new light schedule is already working in your favor, I think most people find that taking it at bedtime is more than good enough for them.

5. I try to use melatonin for sleep, but it just gives me weird dreams and makes me wake up very early

This is my experience too. When I use melatonin, I find I wake the next morning with a jolt of energy. Although I usually have to grudgingly pull myself out of bed, melatonin makes me wake up bright-eyed, smiling, and ready to face the day ahead of me…

…at 4 AM, invariably. This is why despite my interest in this substance I never take melatonin myself anymore.

There are many people like me. What’s going on with us, and can we find a way to make melatonin work for us?

This bro-science site has an uncited theory. Melatonin is known to suppress cortisol production. And cortisol is inversely correlated with adrenaline. So if you’re naturally very low cortisol, melatonin spikes your adrenaline too high, producing the “wake with a jolt” phenomenon that I and some other people experience. I like the way these people think. They understand individual variability, their model is biologically plausible, and it makes sense. It’s also probably wrong; it has too many steps, and nothing in biology is ever this elegant or sensible.

I think a more parsimonious theory would have to involve circadian rhythm in some way. Even an 0.3 mg dose of melatonin gives your body the absolute maximum amount of melatonin it would ever have during a natural circadian cycle. So suppose I want to go to bed at 11, and take 0.3 mg melatonin. Now my body has a melatonin peak (usually associated with the very middle of the night, like 3 AM) at 11. If it assumes that means it’s *really* 3 AM, then it might decide to wake up 5 hours later, at what it thinks is 8 AM, but which is actually 4.

I think I have a much weaker circadian rhythm than most people – at least, I take a lot of naps during the day, and fall asleep about equally well whenever. If that’s true, maybe melatonin acts as a superstimulus for me. The normal tendency to wake up feeling refreshed and alert gets exaggerated into a sudden irresistable jolt of awakeness.

I don’t know if this is any closer to the truth than the adrenaline theory, but it at least fits what we know about circadian rhythms. I’m going to try to put some questions about melatonin response on the SSC survey this year, so start trying melatonin now so you can provide useful data.

What about the weird dreams?

From a HuffPo article:

> Dr. Rafael Pelayo, a Stanford University professor of sleep medicine, said he doesn’t think melatonin causes vivid dreams on its own. “Who takes melatonin? Someone who’s having trouble sleeping. And once you take anything for your sleep, once you start sleeping more or better, you have what’s called ‘REM rebound,’” he said.
>
> This means your body “catches up” on the sleep phase known as rapid eye movement, which is characterized by high levels of brain-wave activity.

> Normal subjects who take melatonin supplements in the controlled setting of a sleep lab do not spend more time dreaming or in REM sleep, Pelayo added. This suggests that there is no inherent property of melatonin that leads to more or weirder dreams.

Okay, but I usually have normal sleep. I take melatonin sometimes because I like experimenting with psychotropic substances. And I still get some really weird dreams. A Slate journalist says he's been taking melatonin for nine years and still gets crazy dreams.

We know that REM sleep is most common towards the end of sleep in the early morning. And we know that some parts of sleep structure are responsive to melatonin directly. There's a lot of debate over exactly what melatonin does to REM sleep, but given all the reports of altered dreaming, I think you could pull together a case that it has some role in sleep architecture that promotes or intensifies REM.

6. Does this relate to any other psychiatric conditions?

Probably, but this is all still speculative.

Seasonal affective disorder is the clearest suspect. We know that the seasonal mood changes don't have anything to do with temperature; they seem to be based entirely on winter having shorter (vs. summer having longer) days.

There's some evidence that there are two separate kinds of winter depression. In one, the late sunrises train people to a late circadian rhythm and they end up phase-delayed. In the other, the early sunsets train people to an early circadian rhythm and they end up phase-advanced. Plausibly SAD also involves some combination of the two where the circadian rhythm doesn't know what it's doing. In either case, this can make sleep non-circadian-rhythm-congruent and so less effective at doing whatever it is sleep does, which causes mood problems.

How does sunrise time affect the average person, who is rarely awake for the sunrise anyway and usually sleeps in a dark room? I think your brain subconsciously “notices” the time of the dawn even if you are asleep. There are some weird pathways leading from the eyes to the nucleus governing circadian rhythm that seem independent of any other kind of vision; these might be keeping tabs on the sunrise if even a little outside light is able to leak into your room. I’m basing this also on the claim that dawn simulators work even if you sleep through them. I don’t know if people get seasonal affective disorder if they sleep in a completely enclosed spot (eg underground) where there’s no conceivable way for them to monitor sunrise times.

Bright light is the standard treatment for SAD for the same reason it’s the standard treatment for any other circadian phase delay, but shouldn’t melatonin work also? Yes, and there are some preliminary studies (paper, article) showing it does. You have to be a bit careful, because some people are phase-delayed and others phase-advanced, and if you use melatonin the wrong way it will make things worse. But for the standard phase-delay type of SAD, normal phase advancing melatonin protocol seems to go well with bright light as an additional treatment.

This model also explains the otherwise confusing tendency of some SAD sufferers to get depressed in the summer. The problem isn’t amount of light, it’s circadian rhythm disruption – which summer can do just as well as winter can.

I’m also very suspicious there’s a strong circadian component to depression, based on a few lines of evidence.

First, one of the most classic symptoms of depression is awakening in the very early morning and not being able to get back to sleep. This is confusing for depressed people, who usually think of themselves as very tired and needing to sleep more, but it definitely happens. This fits the profile for a circadian rhythm issue.

Second, agomelatine, a melatonin analogue, is an effective (ish) antidepressant.

Third, for some reason staying awake for 24+ hours is a very effective depression treatment (albeit temporary; you'll go back to normal after sleeping). This seems to sort of be a way of telling your circadian rhythm "You can't fire me, I quit", and there are some complicated sleep deprivation / circadian shift protocols that try to leverage it into a longer-lasting cure. I don't know anything about this, but it seems pretty interesting.

Fourth, we checked and depressed people definitely have weird circadian rhythms.

Last of all, bipolar has a very strong circadian component. There aren't a whole lot of lifestyle changes that really work for preventing bipolar mood episodes, but one of the big ones is keeping a steady bed and wake time. Social rhythms therapy, a rare effective psychotherapy for bipolar disorder, revolves around training bipolar people to control their circadian rhythms.

Theories of why circadian rhythms matter so much revolve either around the idea of pro-circadian sleep – that sleep is more restorative and effective when it matches the circadian cycle – or the idea of multiple circadian rhythms, with the body functioning better when all of them are in sync.

7. How can I know what the best melatonin supplement is?

Labdoor has done purity tests on various brands and has ranked them for you. All the ones they highlight are still ten to thirty times the appropriate dose (also, stop calling them things like “Triple Strength!” You don’t *want* your medications to be too strong!). As usual, I trust NootropicsDepot for things like this – and sure enough their melatonin (available on Amazon) is *exactly* 0.3 mg. God bless them.

Did A Melatonin Patent Inspire Current Dose Confusion?

Yesterday I wrote about melatonin, mentioning that most drugstore melatonin supplements were 10x or more the recommended dose. A commenter on Facebook pointed me to an interesting explanation of why.

Dr. Richard Wurtman, an MIT scientist who helped discover melatonin's role in the body and pioneer its use as a sleep aid, writes:

> MIT was so excited about our research team's melatonin-sleep connection discovery that they decided to patent the use of reasonable doses of melatonin—up to 1 mg—for promoting sleep.
>
> But they made a big mistake. They assumed that the FDA would want to regulate the hormone and its use as a sleep therapy. They also thought the FDA wouldn't allow companies to sell melatonin in doses 3-times, 10-times, even 15-times more than what's necessary to promote sound sleep.
>
> Much to MIT's surprise, however, the FDA took a pass on melatonin. At that time, the FDA was focusing on other issues, like nicotine addiction, and they may have felt they had bigger fish to fry.
>
> Also, the FDA knew that the research on melatonin showed it to be non-toxic, even at extremely high doses, so they probably weren't too worried about how consumers might use it. In the end, and as a way of getting melatonin on to the market, the FDA chose to label it a dietary supplement, which does not require

> FDA regulation. Clearly, this was wrong because melatonin is a hormone, not a dietary supplement.
>
> Quickly, supplement manufacturers saw the huge potential in selling melatonin to promote good sleep. After all, millions of Americans struggled to get to sleep and stay asleep, and were desperate for safe alternatives to anti-anxiety medicines and sleeping pills that rarely worked well and came with plenty of side effects.
>
> Also, manufacturers must have realized that they could avoid paying royalties to MIT for melatonin doses over the 1 mg measure. So, they produced doses of 3 mg, 5 mg, 10 mg and more! Their thinking–like so much else in our American society–was likely, “bigger is better!” But, they couldn’t be more wrong.

So he’s saying that…in order to get around a patent on using the correct dose of melatonin…supplement manufacturers…used the wrong dose of melatonin? I enjoy collecting stories of all the crazy perversities created by our current pharmaceutical system, but this one really takes the cake.

Assuming it’s true, that is. Commenter Rodrigo brings up some reasons to be suspicious:

1. Who would patent a drug only up to a certain dose? Isn’t this really dumb?

2. To avoid the patent on the correct dose, drugstores just have to sell more than 1 mg – for example, 2 mg. But they actually sell up to 10 mg.

To these I would add:

3. Lots of supplements are very high dose. When I Google Vitamin C, the first product that comes up advertises that it has 1111% of the recommended daily allowance, which seems better optimized for numerological purposes than medical ones.

4. A few companies do sell melatonin at the right dose range, and MIT hasn't sued them yet.

Normally I would find these considerations pretty persuasive, but I feel like the guy who discovered melatonin and ran a pharmaceutical company for a while knows more about the history of melatonin and pharmaceutical regulations than I do.

From last week:

> This kind of thing is the endless drudgery of rationality training…questions like "How much should you discount a compelling-sounding theory based on the bias of its inventor?" And "How much does someone being a famous expert count in their favor?" And "How concerned should we be if a theory seems to violate efficient market assumptions?" And "How do we balance arguments based on what rationally has to be true, vs. someone's empirical but fallible data sets?"

Here I'm just really skeptical of the MIT patent story. Wurtman seems to admit that "bigger is better" played a role. Maybe the patent thing was a very small issue, around the beginning of melatonin sales, and was soon forgotten – but the tradition of expecting melatonin to be very high dose stuck around forever, mostly for other reasons?

EDIT: Commenters, including a patent lawyer, have filled in the rest of the story. Because melatonin is a natural hormone and not an invention, patents can only cover specific uses of it. The MIT patent covered the proper way to use it for sleep; a broader patent might not have been granted. The patent probably guided supplement companies, but expired about five years ago. It’s now legal to produce melatonin 0.3 mg pills, but people are so used to higher doses that few people do.

Fundamental Value Differences Are Not That Fundamental

I.

Ozy (and others) talk about fundamental value differences as a barrier to cooperation.

On their model (as I understand it) there are at least two kinds of disagreement. In the first, people share values but disagree about facts. For example, you and I may both want to help the Third World. But you believe foreign aid helps the Third World, and I believe it props up corrupt governments and discourages economic self-sufficiency. We should remain allies while investigating the true effect of foreign aid, after which our disagreement will disappear.

In the second, you and I have fundamentally different values. Perhaps you want to help the Third World, but I believe that a country should only look after its own citizens. In this case there's nothing to be done. You consider me a heartless monster who wants foreigners to starve, and I consider you a heartless monster who wants to steal from my neighbors to support random people halfway across the world. While we can agree not to have a civil war for pragmatic reasons, we shouldn't mince words and pretend not to be enemies. Ozy writes (liberally edited, read the original):

> From a conservative perspective, I am an incomprehensible moral mutant…however, from my perspective, conservatives are perfectly willing to sacrifice things that actually matter in the world– justice, equality, happiness, an end to suffering– in order to suck up to unjust authority or help the wealthy and

> undeserving or keep people from having sex lives they think are gross.
>
> There is, I feel, opportunity for compromise. An outright war would be unpleasant for everyone…And yet, fundamentally… it’s not true that conservatives as a group are working for the same goals as I am but simply have different ideas of how to pursue it…my read of the psychological evidence is that, from my value system, about half the country is evil and it is in my self-interest to shame the expression of their values, indoctrinate their children, and work for a future where their values are no longer represented on this Earth. So it goes.

And from the subreddit comment by GCUPokeItWithAStick:

> I do think that at a minimum, if you believe that one person’s interests are intrinsically more important than another’s (or as the more sophisticated versions play out, that ethics is agent-relative), then something has gone fundamentally wrong, and this, I think, is the core of the distinction between left and right. Being a rightist in this sense is totally indefensible, and a sign that yes, you should give up on attempting to ascertain any sort of moral truth, because you can’t do it.

I will give this position its due: I agree with the fact/value distinction. I agree it’s conceptually very clear what we’re doing when we try to convince someone with our same values of a factual truth, and confusing and maybe impossible to change someone’s values.

But I think the arguments above are overly simplistic. I think rationalists might be especially susceptible to this kind of thing, because we often use economic models where an agent (or AI) has a given value function (eg “produce paperclips") which generates its actions. This kind of agent really does lack common ground with another agent whose goal function is different. But humans rarely work like this. And even when they do, it's rarely in the ways we think. We are far too quick to imagine binary value differences that line up exactly between Us and Them, and far too slow to recognize the complicated and many-scaled pattern of value differences all around us.

Eliezer Yudkowsky writes, in Are Your Enemies Innately Evil?:

> On September 11th, 2001, nineteen Muslim males hijacked four jet airliners in a deliberately suicidal effort to hurt the United States of America. Now why do you suppose they might have done that? Because they saw the USA as a beacon of freedom to the world, but were born with a mutant disposition that made them hate freedom?
>
> Realistically, most people don't construct their life stories with themselves as the villains. Everyone is the hero of their own story. The Enemy's story, as seen by the Enemy, is not going to make the Enemy look bad. If you try to construe motivations that would make the Enemy look bad, you'll end up flat wrong about what actually goes on in the Enemy's mind.

So what *was* going through the 9/11 hijackers' minds? How many value differences did they have from us?

It seems totally possible that the hijackers had no value differences from me at all. If I believed in the literal truth of Wahhabi Islam – a factual belief – I might be pretty worried about the sinful atheist West. If I believed that the West's sinful ways were destroying my religion, and that my religion encoded a uniquely socially beneficial way of life –

both factual beliefs – I might want to stop it. And if I believed that a sufficiently spectacular terrorist attack would cause people all around the world to rise up and throw off the shackles of Western oppression – another factual belief – I might be prepared to sacrifice myself for the greater good. If I thought complicated Platonic contracts of cooperation and nonviolence didn't work – sort of a factual belief – then my morals would no longer restrain me.

But of course maybe the hijackers had a bunch of value differences. Maybe they believed that American lives are worth nothing. Maybe they believed that striking a blow for their homeland is a terminal good, whether or not their homeland is any good or its religion is true. Maybe they believe any act you do in the name of God is automatically okay.

I have no idea how many of these are true. But I would hate to jump to conclusions, and infer from the fact that they crashed two planes that they believe crashing planes is a terminal good. Or infer from someone opposing abortion that they just think oppressing women is a terminal value. Or infer from people committing murder that they believe in murderism, the philosophy that says that murder is good. I think most people err on the side of being too quick to dismiss others as fundamentally different, and that a little charity in assessing their motives can go a long way.

II.

But that's too easy. What about people who didn't die in self-inflicted plane crashes, and who can just tell us their values? Consider the original example – foreign aid. I've heard many isolationists say in no uncertain terms that they believe we should not spend money to foreign countries, and that this is a basic principle and not just a consequence of some factual belief like that foreign countries would waste it. Meanwhile, I know other people who argue that we should treat foreigners exactly the same as our fellow citizens – indeed, that it would be an affront to basic compassion and to the unity of the human race not to do so. Surely this is a strong case for actual value differences?

My only counter to this line of argument is that almost nobody, me included, ever takes it seriously or to its logical conclusion. I have never heard any cosmopolitans seriously endorse the idea that the Medicaid budget should be mostly redirected from the American poor (who are already plenty healthy by world standards) and used to fund clinics in Africa, where a dollar goes much further. Perhaps this is just political expediency, and some would talk more about such a plan if they thought it could pass. But in that case, they should realize that they are very few in number, and that their value difference isn't just with conservatives but with the overwhelming majority of their friends and their own side.

And if nativist conservatives are laughing right now, I know that some of them have given money to foreign countries affected by natural disasters. Some have even suggested the government do so – when the US government sent resources to Japan to help rescue survivors of the devastating Fukushima tsunami, I didn't hear anyone talk about how those dollars could better be used at home.

Very few people have consistent values on questions like these. That's because nobody naturally has principles. People take the unprincipled mishmash of their real opinions, extract principles out of it, and follow those principles. But the average person only does this very weakly, to the point of having principles like "it's bad when you lie to me, so maybe lying is wrong in general" – and even moral philosophers do it less than a hundred percent and apply their principles inconsistently.

(this isn’t to say those who have consistent principles are necessarily any better grounded. I’ve talked a lot about shifting views of federalism: when the national government was against gay marriage, conservatives supported top-down decision-making at the federal level, and liberals protested for states’ rights. Then when the national government came out in support, conservatives switched to wanting states’ rights and liberals switched to wanting top-down federal decisions. We can imagine some principled liberal who, in 1995, said “It seems to me right now that state rights are good, so I will support them forevermore, even when it hurts my side”. But her belief still would have ended up basically determined by random happenstance; in a world where the government started out supporting gay marriage but switched to oppose it, she would have – and stick to – the opposite principle)

But I’m saying that what principle you verbalize (“I believe we must treat foreigners exactly as our own citizens!”) isn’t actually that interesting. In reality, there’s a wide spectrum of what people will do with foreigners. If we imagine it as a bell curve, the far right end has a tiny number of hyper-consistent people who oppose any government money going abroad unless it directly helps domestic citizens. A little further towards the center we get the people who say they believe this, but will support heroic efforts to rescue Japanese civilians from a tsunami. The bulge in the middle is people who want something like the current level of foreign aid, as long as it goes to sufficiently photogenic children. Further to the left, we get the people I’m having this discussion with, who usually support something like a bit more aid and open borders. And on the far left, we get another handful of hyper-consistent people, who think the

US government should redirect the Medicaid budget to Africa.

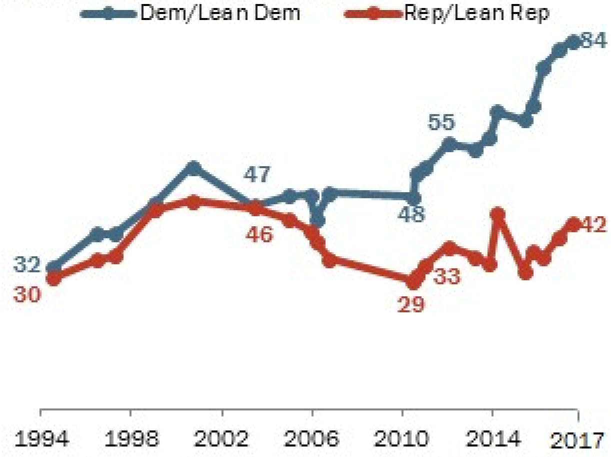
If you’re at Point N in some bell curve, how far do you have to go before you come to someone with “fundamental value differences” from you? How far do you have to go before someone is inherently your enemy, cannot be debated with, and must be crushed in some kind of fight? If the answer is “any difference at all”, I regret to inform you that the bell curve is continuous; there may not be anyone with exactly the same position as you.

And that’s just the one issue of foreign aid. Imagine a hundred or a thousand such issues, all equally fraught. God help GCU, who goes further and says you’re “indefensible” if you believe *any* human’s interests are more important than any other’s. Does he (I’ll assume it’s a he) do more to help his wife when she’s sick than he would to help a random stranger? This isn’t meant to be a gotcha, it’s meant to be an example of how we formulate our morality. Person A cares more about his wife than a random person, and also donates some token amount to help the poor in Africa. He dismisses caring about his wife as noise, then extrapolates from the Africa donation to say “we must help all people equally”. Person B also cares more about his wife than a random person, and also donates some token amount to Africa. He dismisses the Africa donation as noise, then extrapolates from his wife to “we must care most about those closest to us”. I’m not saying that how each person frames his moral principle won’t have effects later down the line, but those effects will be the tail wagging the dog. If A and B look at each other and say “I am an everyone-equally-er, you are a people-close-to-you-first-er, we can never truly understand one another, we must be sworn enemies”, they’re putting a whole lot more emphasis on which string of syllables they use to describe their mental processes than really seems warranted.

Why am I making such a big deal of this? Isn't a gradual continuous value difference still a value difference?

Yes. But I expect that (contra the Moral Foundations idea) both the supposed-nativist and the supposed-cosmopolitan have at least a tiny bit of the instinct toward nativism and the instinct toward cosmopolitanism. They may be suppressing one or the other in order to fit their principles. The nativist might be afraid that if he admitted any instinct toward cosmopolitanism, people could force him to stop volunteering at his community center, because his neighbor's children are less important than starving Ethiopians and he should be helping them somehow instead. The cosmopolitan might be afraid that if he admitted any instinct toward preferring people close to him, it would justify a jingoistic I've-got-mine attitude that thinks of foreigners as subhuman.

But the idea that they're inherently different, and neither can understand the other's appeals or debate each other, is balderdash. A lot of the our-values-are-just-inherently-different talk I've heard centers around immigration. Surely liberals must have some sort of strong commitment to the inherent moral value of foreigners if they're so interested in letting them into the country? Surely conservatives must have some sort of innate natives-first mentality to think they can just lock people out? But…



% who say immigrants strengthen the country because of their hard work and talents ...

Source: Survey conducted June 8-18 and June 27-July 9, 2017. Q25g.

PEW RESEARCH CENTER

Okay. I admit this is a question about hard work and talents, which is a factual question. But we both know that you would get basically the same results if you asked “IMMIGRATION GOOD OR BAD?” or “DO IMMIGRANTS HAVE THE SAME RIGHTS TO BE IN THIS COUNTRY AS THE NATIVE BORN?” or whatever. And what we see is that this is totally contingent and dependent on the politics of the moment. Of all those liberals talking about how they can’t possibly comprehend conservatives because being against immigration would just require completely alien values, half of them were anti-immigrant ten years ago. Of all those conservatives talking about how liberals can never be convinced by mere debate because debate can’t cut across fundamental differences, they should try to figure out why their own party was half again as immigrant-friendly in 2002 as in 2010.

I don’t think anyone switched because of anything they learned in a philosophy class. They switched because it became mildly convenient to switch, and they had a bunch of pro-immigrant instincts and anti-immigrant instincts the whole time, so it was easy to switch which words came out of their mouths as soon as it became convenient to do so.

So if the 9/11 hijackers told me they truly placed zero value on American lives, I would at least reserve the possibility that sure, this is something you *say* when you want to impress your terrorist friends, but that in a crunch – if they saw an anvil about to drop on an American kid and had only a second to push him out of the way – they would end up having some of the same instincts as the rest of us.

III.

Is there anyone at all whom I am willing to admit definitely, 100%, in the most real possible way, has different values than I do?

I think so. I remember a debate I had with my ex-girlfriend. Both of us are atheist materialist-computationalist utilitarian rationalist effective altruist liberal-tarians with 99% similar views on every political and social question. On the other hand, it seemed axiomatic to me that it wasn't morally good/obligatory to create extra happy people (eg have a duty to increase the population from 10,000 to 100,000 people in a way that might eventually create the Repugnant Conclusion), and it seemed equally axiomatic to her that it *was* morally good/obligatory to do that. We debated this maybe a dozen times throughout our relationship, and although we probably came to understand each other's position a little more, and came to agree it was a hard problem with some intuitions on both sides, we didn't come an inch closer to agreement.

I've had a few other conversations that ended with me feeling the same way. I may not be the typical Sierra Club member, but I consider myself an environmentalist in the sense of liking the environment and wanting it to be preserved. But I don’t think I value biodiversity for its own sake – if you offered me something useful in exchange for half of all species going extinct – promising that they would all be random snails, or sponges, or some squirrel species that looked exactly like other squirrel species, or otherwise not anything we cared about – I’d take it. If you offered me all charismatic megafauna being relegated to zoos in exchange for lots of well-preserved beautiful forests that people could enjoy whenever they wanted, I would take that one too. I know other people who consider themselves environmentalists who are horrified by this. Some of them agree with me on every single political issue that real people actually debate.

I think these kinds of things are probably real fundamental value difference. But if I’m not sure I have any fundamental value differences with the 9-11 hijackers, and I am sure I have one with one of the people I’m closest to in the entire world, how big a deal is it, exactly? The world isn’t made of Our Tribe with our fundamental values and That Tribe There with their fundamental values. It’s made of a giant mishmash of provisional things that solidify into values at some point but can be unsolidified by random chance or temporary advantage, and everyone probably has a couple unexplored value differences and unexpected value similarities with everyone else.

This means that trying to use shaming and indoctrination to settle value differences is going to be harder than you think. Successfully defeat the people on the other side of the One Great Binary Value Divide That Separates Us Into Two Clear Groups, and you’re going to notice you still have some value differences with your allies (if you don’t now, you will in ten years, when the political calculus changes slightly and their deepest ethical beliefs become totally different). Beat your allies, and you and the subset of remaining allies will *still* have value differences. It's value differences all the way down. You will have an infinite number of fights, and you're sure to lose some of them. Have you considered getting principles and using asymmetric weapons?

I'm not saying you don't have to fight for your values. The foreign aid budget still has to be some specific number, and if your explicitly-endorsed principles disagree with someone else's explicitly-endorsed principles, then you've got to fight them to determine what it is.

But "remember, liberals and conservatives have fundamental value differences, so they are two tribes that can't coexist" is the wrong message. "Remember, everyone has weak and malleable value differences with everyone else, and maybe a few more fundamental ones though it's hard to tell, and neither type necessarily line up with tribes at all, so they had damn well better learn to coexist" is more like it.

The Whole City Is Center

Related to yesterday’s post on people being too quick to assume value differences: some of the simplest fake value differences are where people make a big deal about routing around a certain word. And some of the most complicated real value differences are where some people follow a strategy explicitly and other people follow heuristics that approximate that strategy.

There’s a popular mental health mantra that “there’s no such thing as laziness” (here are ten different articles with approximately that title: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10). They all make the same basically good point. We shame people who don’t work very hard as “lazy”, and think they should have lower status than the rest of us. But actually, these people don’t just randomly choose not to work. Some of them have psychological issues, like anxiety and trauma that are constantly distracting them from their work, or a fear of success, or self-defeating beliefs about how nothing they do matters anyway. Others have biological issues – maybe hypothyroidism, or vitamin deficiencies, or ADHD, or other things we don’t understand that lower their energy and motivation. Still others just don’t want to do the specific thing we are asking them to do right now and can’t force themselves to work uphill against that gradient. When we call people “lazy”, we’re ignorantly dismissing all these possibilities in favor of a moralistic judgment.

A dialogue:

> **Sophisticus:** I don’t believe in laziness.
>
> **Simplicio:** What about my cousin Larry? He keeps promising to do important errands for his friends and

family, and then he never does them. Instead he just plays video games all the time. This has happened consistently over the past few years, every time he’s promised to do something. One time my aunt asked him to go to the DMV to get some paperwork filled out, he promised he would do it, and then he kept putting it off for a month until it was past the deadline and she almost lost her car. He didn’t forget about it or anything, he just couldn’t bring himself to go out and do it. And he’s been fired from his last three jobs for not showing up, and…

Sophisticus: Yes, yes, I’m sure there are people like this. But he probably has some self-defeating beliefs, or vitamin deficiencies, or mental health issues.

Simplicio: Okay. Well, my mother is going to be away for the next week, and she needs someone to dog-sit for her. Her dog is old and sick and requires a lot of care each day. She’s terrified that if he doesn’t get his food and medication and daily walk on time, something terrible will happen to him. She’s willing to pay a lot of money. Do you think I should recommend she ask my cousin Larry?

Sophisticus: No, of course not.

Simplicio: Why not?

Sophisticus: He probably won’t do it. He’ll just play video games instead.

Simplicio: Why do you think so?

Sophisticus: Because he has a long history of playing video games instead of doing important tasks.

Simplicio: If only there were a word for the sort of person who does that!

Sophisticus: Oh, I see. Now you're making fun of me. But I'm not saying everyone is equally reliable. I'm saying that instead of denouncing someone as “lazy”, we should look for the cause and try to help them.

Simplicio: Hey, we did try to help him. Larry's family has taken him to the doctor loads of times. They didn't anything on the lab tests, but the psychiatrist thought he might be ADHD and gave him some Adderall. I would say now he pulls through on like 20% of the things we ask him to do instead of zero percent. We also tried to get him to go to therapy, but the therapist deferred because ADHD has a very low therapy response rate. His parents tried to change the way they asked him to do things to make it easier for him, or to let him choose a different set of tasks that were more to his liking, but that only worked a little, if at all. Probably there's some cause we don't understand, but it's beyond the reach of medical science, incentive design, or the understanding that exists between loving family members to identify.

Sophisticus: See! The Adderall helped! And letting him choose his own tasks helped a little too!

Simplicio: I agree it helped a little. So should I recommend him to my mother as a dog-sitter?

Sophisticus: No, of course not.

Simplicio: Then I still don't see what the difference between us is. I agree it was worth having him go to the doctor and the therapist to rule out any obvious biological or psychological issues, and to test different ways of interacting with him in case our interaction style

> was making things worse. You agree that since this still hasn’t made him reliably fulfill his responsibilities and we don’t have any better ideas, he’s a bad choice for a dog-sitter. Why can’t I communicate the state of affairs we both agree on to my mother using the word “lazy”?

I imagine Sophisticus believing he has a fundamental value difference with people who use the term “lazy”. They think that some people are just bad and should be condemned, whereas he wisely believes that everything has a cause and people who have issues with motivation should be helped. But it’s not clear to me that this is a real difference. I can imagine someone signaling hard-headedness and strictness by insisting that they were against laziness, and someone else signaling compassion by insisting that they don’t believe in laziness, but it’s pretty hazy exactly where their maps of the world diverge.

But back to the dialogue:

> **Sophisticus:** Because “lazy” is laden with the idea that lazy people should be punished. You should yell at them to get off their ass and do some work.
>
> **Simplicio:** I mean, I’m not sure that’s wrong? When my aunt and uncle tried to take Larry to the psychiatrist, he didn’t want to go. My uncle started screaming at him that if he didn’t make the appointment he would never amount to anything, and he would be a loser his entire life, and they would disown him – and I guess it freaked Larry out enough that he made the appointment. And it seems like if that kind of thing makes people do important stuff for their own good – whether it’s make appointments or hold down a job – then it might be reasonable, at least from people whom the lazy person has entered into some kind of relationship with.

> **Sophisticus:** I think that kind of strategy might occasionally work in the short-term, but that in the long-term it makes things much worse.
>
> **Simplicio:** I agree that's possible, but it seems like we have a factual disagreement here. And I think that factual disagreement is best expressed by the question "Does laziness respond to social shaming or not?", not a claim that laziness doesn't exist. It certainly doesn't seem like we have a value difference unrelated to any purely-factual beliefs.

Maybe both participants are wrong here. My impression is that some forms of laziness respond to incentives and others don't. I know many people who will start work on a project they've been putting off if they know it's due the next day and worth half their grade. I also know other people who won't. But continuing:

> **Sophisticus:** I can imagine some cases in which it's useful to use external rewards and punishments to encourage people with low motivation to do something. But the word "lazy" doesn't just mean "can be motivated by external reinforcement". It's an attempt to judge somebody, to say they're lesser, to lower their social status.
>
> **Simplicio:** You just said that my mother should avoid hiring Larry for a lucrative job. Surely that's a judgment, and surely keeping him unemployed forever lowers his social status.
>
> **Sophisticus:** I'm judging him as bad at one thing, not as a Universally Bad Person.
>
> **Simplicio:** Do you think Larry would be a good pilot?

Sophisticus: Well, no…

Simplicio: Nuclear engineer?

Sophisticus: No, but…

Simplicio: Lieutenant colonel in the army?

Sophisticus: I agree there are many things Larry would not be good at.

Simplicio: And surely the person who thinks he is lazy agrees there are some things he might be good at – for example, he might be handsome, or intelligent. Indeed, the "lazy but bright" student is a stock cultural figure. The main judgment that "lazy" represents is that he's not a very hard worker – a judgment you seem to share.

Sophisticus: I think that for them it's a moral judgment, and for me it isn't.

Simplicio: A moral judgment? I don't think of a lazy person as more likely to rob or murder. Do you think others do?

Sophisticus: No, I don't think so. It's not a judgment that they're bad at a specific field we both agree is moral. It's a judgment that they should be considered less moral just because they're lazy.

Simplicio: But how does that cash out? Both you and they want them to not get certain jobs. Both you and they believe some level of reinforcement might make them more motivated, though we can debate the factual specifics. Is there anything that a moralist would do that you wouldn't?

Sophisticus: I'm not sure the belief would cash itself out in some specific way, but they would have it.

Simplicio: If you both give him the same jobs and treat him the same, what's the difference? Just give him the Heartstone and call it a day!

Sophisticus: You're mocking me again.

Simplicio: I think we're treating the word "laziness" differently. I'm thinking of "lazy" as a way to communicate a true fact about the world. You agree that the true fact should be communicated by some word, but you're interpreting "lazy" to mean some sort of awful concept like "a person who avoids responsibilities in a way not caused by anything whatsoever except being bad, and so we should hurt them and make them suffer". Are you sure this isn't kind of dumb? Given that we need a word for the first thing, and everyone currently uses "lazy" for it, and we don't need a word for the second thing because it's awful, and most people would deny that "lazy" means that, why don't we just use "lazy" for the very useful purpose it's served thus far?

Sophisticus: I think…

Simplicio: And it's the same with "judgment". I'm using it to mean a reasonable thing that everyone does and has to do. You're demanding we reserve it for some kind of ultimate judgment about everything that doesn't really make sense and probably should never happen.

Sophisticus: I think you're wrong about common usage. I think a lot of people – maybe not you, but a lot of people – really do use "lazy" to mean the second thing. And that even for good people like yourself, “lazy” has a bit of a connotation of the second thing which you can’t avoid letting slip into your mind.

Simplicio: If you’re right, I worry you’re going up against the euphemism treadmill. If we invent another word to communicate the true fact, like “work-rarely-doer”, then anyone who believes that people who play video games instead of working deserve to suffer will quickly conclude that work-rarely-doers deserve to suffer.

Sophisticus: Then let’s not invent something like “work-rarely-doer”. Let’s just say things like “You shouldn’t have Larry as a dog-sitter, because due to some social or psychological issue he usually plays video games instead of doing difficult tasks.”

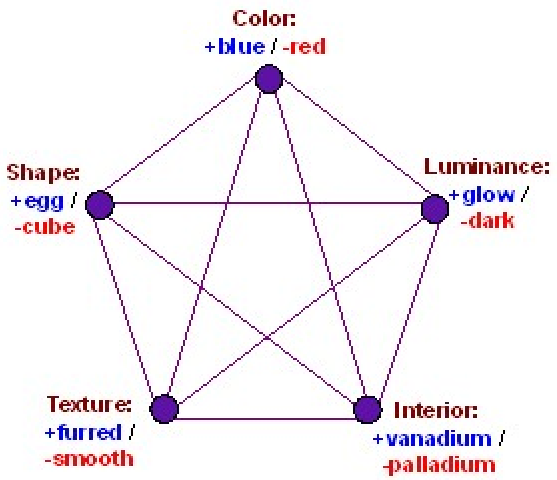
Simplicio: I think people are naturally going to try to compress that concept. You can try to stop them, but I think you’ll fail. And I think insofar as you can communicate the concept at all, people are going to think less of Larry because of it. It’s possible you can slightly decrease the degree to which people think less of Larry, but only by slightly decreasing their ability to communicate useful information.

Sophisticus: Well, that’s a risk I’m willing to take.

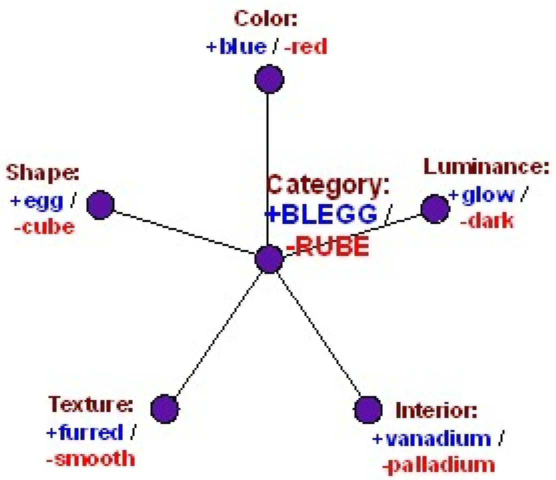
Simplicio: If there were such a thing as laziness, but it was rare, then it would make sense to argue “most people aren’t lazy”, since lazy would be pointing at a particular quality that most people don’t have. But if you say there’s no such thing as laziness, then it sounds like maybe you’re kind of weird to insist on defining “laziness” to refer a quality that nobody has, yet refuse to use any word to refer to the quality that many people

> do have. It would be like wanting our language to have a word for “unicorn” but not for “horse”.

I think Simplicio is working off one of these kinds of models (from How An Algorithm Feels From The Inside, but see also here):


Color:
+blue / -red
Shape:
+egg /
-cube
Luminance:
+glow /
-dark
Texture:
+furred /
-smooth
Interior:
+vanadium /
-palladium

Network 1


Color:
+blue / -red
Shape:
+egg /
-cube
Category:
+BLEGG /
-RUBE
Luminance:
+glow /
-dark
Texture:
+furred /
-smooth
Interior:
+vanadium /
-palladium

Network 2

…where “lazy” is the node at the center of the second one. He’s saying that if he and Sophisticus both agree on all of the outside nodes, why exactly are they holding a debate on the status of the center, when the center is just a way to help us predict the values of the others? He feels like Sophisticus is insisting on designing the structure such that the central node has some deep metaphysical meaning that he can’t explain but which is very bad, whereas he just wants it to be a perfectly ordinary predictor. Moving on:

> **Sophisticus:** What about this? I think that people with low motivation sometimes can be helped by reinforcement – including negative reinforcement. But other people think they should be *punished*. There’s a big difference between simple negative reinforcement and punishment. If you’re just using negative reinforcement, you’re trying to use as little as possible

to get the result you want. But when you're judgmental and you divide people into good and bad, you usually add that the bad people deserve to suffer, regardless of the effect.

Simplicio: This is a strange distinction. Suppose I beat up my wife and threaten to do it again. Shouldn't I go to jail?

Sophisticus: I think we shouldn't be excessive about it, and I don't support mass incarceration, but I don't want you to get off scot-free, because it seems like that would encourage future domestic violence.

Simplicio: If only there were a word for the sort of thing where we made sure people didn't get off scot-free in a way that encouraged future crime!

Sophisticus: No no no, you still don't get it. There's a difference between a principled consequentialist view of discouraging actions, and wanting people to suffer.

Simplicio: Look, I happen to know one of those Hogwarts wizards everyone keeps writing books about, and he's offered to let me take a magic Unbreakable Vow that I won't assault anyone ever again. Now that there's no point in discouraging me, I don't need to deal with this jail thing, right?

Sophisticus: It's not just about discouraging you personally. It's about making an example of you to discourage everyone else. Also, there's a Parfit's Hitch-Hiker type element – the threat of punishment now could have presented you from committing the crime in the past, and the threat couldn't be credible unless we agreed to actually punish you.

Simplicio: Then I question whether the "principled consequentialist view" *ever* differs from the "believing in punishment and wanting bad people to suffer" view in terms of what actions it recommends.

Sophisticus: The people who believe in punishment often say things like "I hope that person rots in jail" or "Let's make the conditions in jail extra bad". Whereas I want domestic violence discouraged by nice Scandinavian-style prisons and – when possible – community service.

Simplicio: If you learned that having nice jails actually…

Sophisticus: Oh, I know what you're going to say. You're going to say this is just a factual difference between me and the pro-punishment faction. They believe, as a matter of fact, that bad conditions discourage crime extra effectively, since some criminals who would be willing to take the risk of a nice Scandinavian-style prison would be scared off by a dark overheated cage. And I agree this is a possible axis on which people can differ, and that if you proved to me that this was true I could be persuaded to reconsider my views. But I have talked to people who have literally said the words "I don't care how much it discourages crime or not, I want criminals to suffer."

Simplicio: Okay. I agree that's good evidence for your view

Sophisticus: You…do? Really? I won one of these? REALLY?!

Simplicio: I guess.

Sophisticus: So you admit sometimes there are fundamental value differences?

Simplicio: Sometimes, yeah, I guess. But I want to be really careful with this. Humans are adaptation-executors, not fitness-maximizers. Only one person in a thousand could give the principled consequentialist defense of criminal justice that you're giving here. The game theory necessary to understand the defense is only a few decades or centuries old, depending on how exactly you define it – but even chimpanzees need to discourage defectors. Since evolution couldn't cram the whole principled consequentialist defense into a chimpanzee brain, it just gave us the urge to punish.

Sophisticus: I agree that's a plausible scientific account of the genesis of the urge to punish. But that doesn't mean that I have to agree with it. After all, evolution gave us an urge to eat sugary food, but I can ignore that urge when I don't think it's the healthy thing for me to do at the moment.

Simplicio: Thanks to modern medical science, you're smarter than your urge telling you to eat sugar. I'm not sure how many people are smarter than their urges to punish. If you miss the Parfit's Hitch-Hiker angle, you punish the wrong people. If you miss the angle where you have to adjust for probability of catching the crime, you punish people the wrong amount. But the person just following their evolutionary urges would get both of those right – more or less. Imagine that, using physics, you are able to approximate the ball-trajectory-predicting power of the world's best golfer – but the golfer still does a little bit better. Would you pooh-pooh him for merely following his base evolutionary urges?

Sophisticus: If it harmed people, yes! You're trying to reduce this to factual differences again, but you already admitted that's not going to work. We're not debating the effectiveness of different punishment levels here. For all I know, evolutionary urges *are* more effective at the goal of keeping me alive – which is notably different from the goal of being just. But that's not the point. The point is that I think there are people who, even if God handed them a stone tablet saying "YOU ONLY NEED TO PUNISH THIS PERSON X AMOUNT TO EFFECTIVELY DISCOURAGE FUTURE ACTIONS", would still punish them X+1 amount just to make them suffer.

Simplicio: Okay, I didn't want to re-open the factual differences thing. I agree they are not aiming at the same thing you are. My point is just that the only difference between you and the pro-punishment faction is that you are following an explicitly-calculated version of the principled consequentialist defense of punishment, and they are following a heuristic approximating the principled consequentialist defense of punishment, and their heuristic might actually be more accurate than your explicit calculation.

Sophisticus: So what? Again, they *outright say* they would deviate from the principled consequentialist position.

Simplicio: Yes. Adaptation-execution rather than fitness-maximization again. Evolution can't quite cram the entire principled consequentialist position into our heads, so it just gives us an urge, and sometimes the urge does weird stuff that the principles wouldn't.

Sophisticus: Again, *so what?* I agree there's a biological/psychological cause for other people being wrong about punishment – just as there is a biological/psychological cause for other people being bad at fulfilling responsibilities – “lazy”, you would say – but that doesn’t mean I can’t continue to disagree with them.

Simplicio: My point is that, if you squint, this is sort of a factual disagreement. It’s not a factual disagreement between you and them. It’s a factual disagreement between you and the evolutionary/biochemical process that created their sense of justice. The evolution/biochemistry is trying to instantiate a view of punishment that does the best job of protecting them and their loved ones without expending unnecessary resources. But it’s getting it wrong – probably erring on the side of caution, as you would expect these sorts of processes to do. I agree you have a value disagreement with them, but this is less them having some value totally foreign to you, and more them attempting to implement your values but not doing a very good job.

Sophisticus: You’re screwing up levels again! If what you’re saying is true, then I have a mere “factual disagreement” with the evolutionary/biological process that produced their values, insofar as you can have a factual disagreement with a blind impersonal force. But I still don’t agree with *the people who are the end result of the process*! They should be abandoning their evolutionary/biochemical process in favor of what’s actually right.

Simplicio: Nobody has a coherent theory of when to abandon their evolutionary/biochemical processes, though. I have the urge to care about my children more than I care about some random people somewhere else. That’s clearly an evolutionary/biochemical process. I cannot justify it based on pure reason. But I choose, in reflective equilibrium, to keep that urge. What moral law can you tell me that allows me to ditch the irrational consequences of my excessive-punishment-urge, but keep the irrational consequences of my love-children-urge?

Sophisticus: Hmmm…what about "the excessive-punishment urge is wrong, even by the standards of the evolutionary/biochemical process that produced it, but the love children urge is right?"

Simplicio: I've been told that we love kittens based on a misfiring of our evolutionary urge to love children. Should I abandon that one? I've been told I love beauty and nature and high mountains and deep forests based on evolutionary heuristics about what kind of places will have a good food supply – now that I can order take-out, should I ditch that too? Huge chunks of our hopes and dreams are the godshatter of misfired-evolutionary processes. Tell me what principled decision lets us judge among them, rejecting one as evil but the other as good?

Sophisticus: I cannot. I make no claim that I can. I only say that, by my arbitrary choice of methods of reaching reflective equilibrium, natural beauty is good but punishment is bad. And that if someone else's arbitrary choice of methods of reaching reflective equilibrium pronounces the opposite, they have a fundamental value difference from me, and I won't shirk from saying so.

Simplicio: Then all I am saying is to be understanding. They're not people who are coming from some sort of alien ideology of suffering being good for its own sake.

> They’re people who are taking the same godshatter you are, and applying a different process of arbitrary reflective equilibrium to it, in a world where none of us really understand or control the process of reflective equlibrium we go through. That gives me a different and more understanding perspective on them. It may not make me agree with them, but it makes me more willing to think of them as an odd but sympathetic potential-negotiating-partner rather than some sort of hostile villain.

I need to admit here that I personally am neither as saintly as Sophisticus nor as reasonable as Simplicio. A while ago, I learned that my great-grandfather was murdered, and my great-grandmother – normally a deeply kind and compassionate woman – demanded the death penalty rather than life imprisonment for his murderers. When the jury went with life imprisonment anyway, she yelled at them that she hoped someone killed their loved ones so they knew how it felt. This story had a pretty big impact on me and made me try to generate examples of things that could happen such that I would really want the perpetrators to suffer, even more than consequentialism demanded. I may have turned some very nasty and imaginative parts of my brain, the ones that wrote the Broadcast interlude in *Unsong*, to imagining crimes perfectly calculated to enrage me. And in the end I did it. I broke my brain to the point where I can very much imagine certain things that would happen and make me want the perpetrator to suffer – not infinitely, but not zero either. I am not going to claim that this is just some misfiring of evolutionary urges which I obviously denounce. I think I stick to them the same way I stick to liking kittens. I’m not sure I would promote them as policy – in the same kind of second-level way where I can think of some people who would make good dictators but still don’t *actually* want them to set up a dictatorship – but I don’t renounce them entirely either. I guess reflective equilibrium is easier to disturb than I thought.

> **Sophisticus:** I’ve been thinking, Simplicio – doesn’t your philosophy hoist itself on its own petard?
>
> **Simplicio:** What do you mean?
>
> **Sophisticus:** You insist that much of what people consider value differences is actually a difference in what words they are willing to use while describing basically the same values. And that if a term seems unsavory to us, we should use it to describe the closest useful concept, rather than condemning it for applying to a bad concept we shouldn’t have. If we hear “laziness”, we should assume it stands for the way your cousin Larry is, rather than some package of moral and metaphysical assumptions. If we hear “judgment”, we should assume it stands for assessing someone’s ability as a dog-walker, rather than some package of moral evaluations. If we hear “punishment”, we should assume it stands for some kind of consequentialist negative reinforcement, rather than the belief that some people deserve to suffer.
>
> **Simplicio:** I’m not sure that’s exactly how I would describe my position, but go on.
>
> **Sophisticus:** What about the term “value difference” itself? It seems like you’re being a hypocrite here. After all, there are plenty of things that look like value differences to us – I disagree with people on moral questions about a thousand times a day. But you insist that none of those are real value differences, and instead we must reserve the concept of “value

> difference” for some Platonic perfect value difference that doesn’t exist in real life.
>
> **Simplicio:** I agree these rarely exist in practice. I think the difference is that they can exist in theory. Imagine a paperclip maximizer robot vs. a paperclip minimizer robot. These have a true value difference. They’re not doing the same thing and applying different words to it. If I were a paperclip minimizer, I could never get a paperclip maximizer to say it wanted to do a specific thing, and then sarcastically say “If only there was a word for that thing!”, and then it would have to admit that word was “paperclip minimization”.
>
> **Sophisticus:** Okay. But there *are* no paperclip maximizers or minimizers in real life. So I still think you’re denying real-life use of a term in favor of some Platonic version that doesn’t exist.
>
> **Simplicio:** I just don’t like the connotations of “value difference”. I think they suggest the non-existent thing.
>
> **Sophisticus:** That’s exactly what I’ve been trying to tell you is true of “laziness” or “judgment”, and you never let me get away with it!
>
> **Simplicio:** I just realized I have to, uh, wash my toaster. I’ll be back in a minute.

Ten years go by. Sophisticus is never able to find Simplicio again. He seems to have disappeared. Sophisticus knows it’s bizarre to think somebody would skip town and change identities just to avoid a philosophical debate, but he cannot think of any other explanation. One day, Sophisticus goes on a vacation to a city very far away, and becomes hopelessly lost. He notices a stranger in glasses and a mustache, who looks familiar in a way he cannot quite place, and asks him for directions.

Sophisticus: Excuse me, do you know the way to city center?

Stranger: Don't worry, good sir! You're in city center right now!

Sophisticus: But…this whole area looks suburban. And the edge of the city is right there – past that street there's only rolling fields as far as the eye can see. How can this be city center?

Stranger: The whole city is the city center!

Sophisticus: What?

Stranger: That's right. We decided that it was pretty stigmatizing to say that certain parts of the city were non-central. You know, it implied that the people there were just a bunch of yokels who weren't *real* citizens the same way everyone else was. So we held a referendum, and everyone agreed that the whole city would be classified as the city center.

Sophisticus: That's pretty weird, but…look, I need to get to the tourist office, and I know it's in city center, so if you're not going to direct me to city center..can you just tell me what part of town the tourist office is in?

Stranger: It's in the center. The whole city is center.

Sophisticus: Let's try this again. Please point me in the direction of the Tourist Office.

Stranger: Perhaps you think the Tourist Office is some kind of mystical place that will answer all of your tourist-related questions and give you a perfect vacation, but that everywhere-not-the-Tourist-Office is some kind of hellscape with nothing of any value to visitors? In that case, I reject your Tourist-Office vs. Non-Tourist-Office distinction. There is no such thing as the Tourist Office.

Sophisticus: By "Tourist Office", I just mean an ordinary non-perfect building with a greater-than-average propensity to give tourist information!

Stranger: Well, if you mean "building" to mean something 100% artificial without even natural materials which is hermetically insulated from the outside, then really there *aren't* any buildings here. There are just –

Sophisticus: Wait, I know you! You're my old friend Simplicio, who skipped town so he didn't have to answer my challenge about his theory of value differences.

Simplicio: Guilty as charged. But now I hope you better understand what I mean. There is a sense in which you're right, and a sense in which I'm right. Words both convey useful information, and shape our connotations and perceptions. While we can't completely ignore the latter role, it's also dangerous to posit fundamental value differences between people who use words one way and people who use them another. My concern is that I've seen people say "I am the kind of person who doesn't believe in laziness, or in punishment, or in judging others. But that guy over there accuses people of being lazy, wants people to suffer, and does judge others. Clearly we have fundamental value differences and must be enemies." All I'm trying to do is say that those people may have differing factual beliefs on how to balance the information-bearing-content of words versus their potential connotations. If we understand the degree to which other people’s differences from us are based on factual rather than fundamental value differences, we can be humbler and more understanding when we have to interact with them.

Sophisticus: Okay, but seriously, I need to get to city center.

Simplicio: The whole city is the city center.

Sophisticus: Screw you.

Simplicio: Hey, don’t be so *judgmental*.

Sentimental Cartography

A long time ago, I made a map of the rationalist community.

MATHEMATICS BLOGS
FACEBOOK
POSTRATIONALITY BLOGS
IRC
TWITTER
Reasoning Sound
MEETUP GROUPS
Reverend Thomas Bay
MIRI
CFAR
Bight of Information
TUMBLR
BUSINESS
Malcolm Ocean
THE BLOGOSPHERE
ECONOMICS BLOGS
BIOLOGY BLOGS
SOCIAL JUSTICE BLOGS
EFFECTIVE ALTRUISM
ATHEIST BLOGS
CHRISTIAN BLOGS
NEOREACTIONARY BLOGS

Thumbnail - click to expand

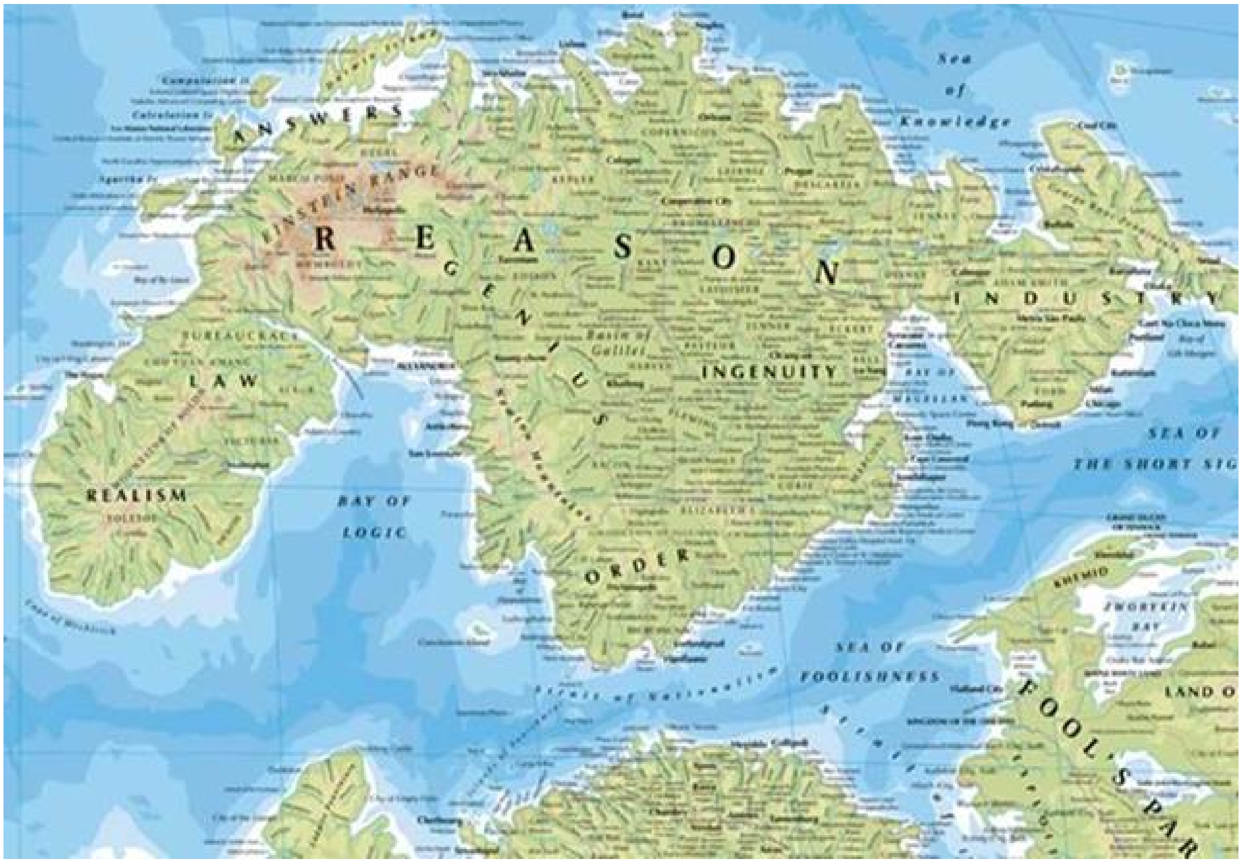
This is in the same geographic-map-of-something-non-geographic tradition as the Greater Ribbonfarm Cultural Region or xkcd’s map of the Internet. There’s even some sort of therapy program that seems to involve making a map like this of your life, though I don’t know how seriously they take it.

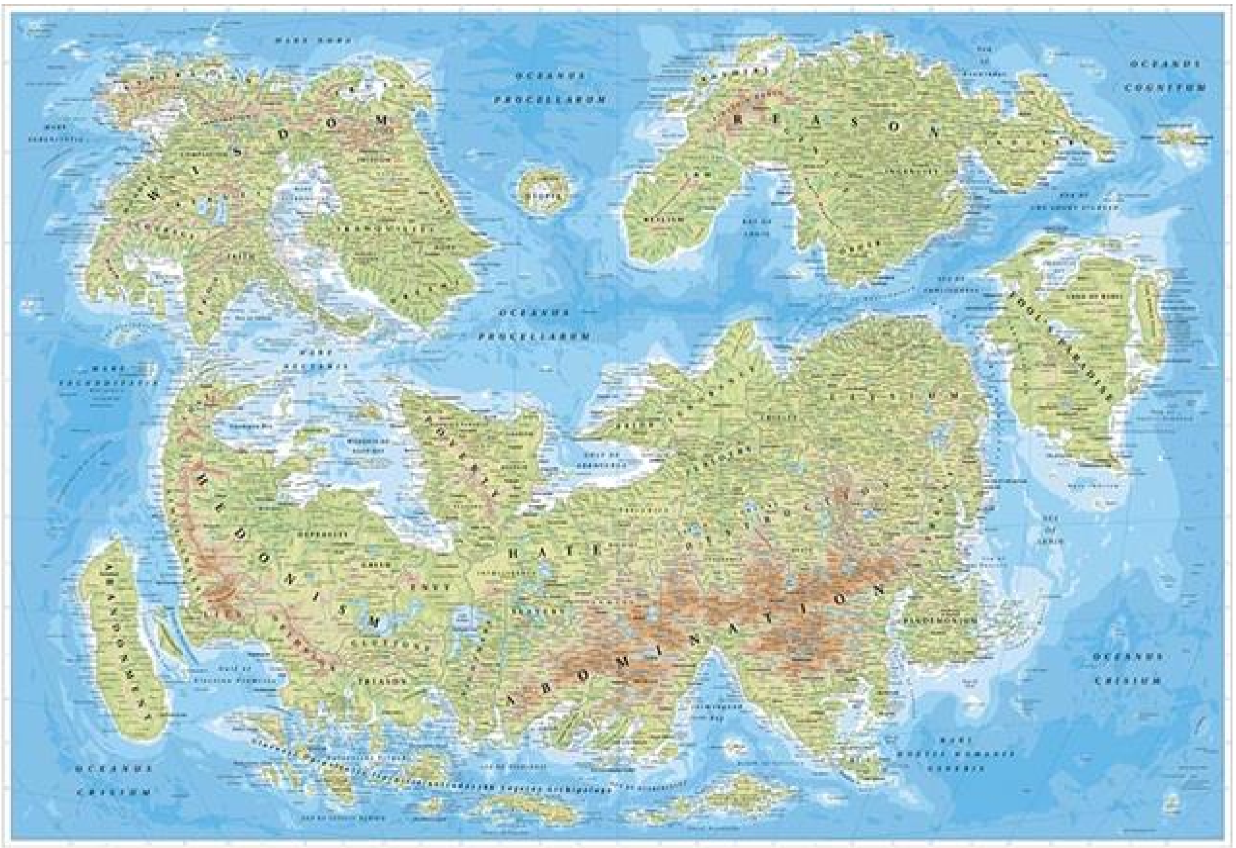
There’s no good name for this art and it’s really hard to Google. If you try “map of abstract concept” you just get a bunch of concept maps. It seems the old name, from back when this was a popular Renaissance amusement, is “sentimental cartography”, since it was usually applied to sentiments like love or sorrow. This isn’t great – the

Internet's not a sentiment – but it's what we've got and I'll do what I can to try to make it catch on.

Here are some of the best (ie, only) works of modern sentimental cartography I've been able to find. Sorry if this ends up a little clickbaity, but I'm annoyed that these haven't been gathered together in one place before. I am very limited by some of them being offline and copyrighted, so some of these will be teasers rather than the full map. Others will be thumbnails that you can click through to get to the full map or an approximation.

The best modern sentimental cartography I can find comes from illustrator James Turner, who made the Map of Humanity.


Sea of Knowledge
ANSWERS
EINSTEIN RANGE
REASON
GENIUS
INDUSTRY
BUREAUCRACY
LAW
Basin of Galilei
INGENUITY
REALISM
BAY OF LOGIC
ORDER
SEA OF
THE SHORT SIG
SEA OF FOOLISHNESS
FOOL'S PAR
LAND O


WISDOM
REASON
OCEANUS PROCELLARUM
OCEANUS COGNITUM
FOOL'S PARADISE
ELYSIUM
POVERTY
HEDONISM
HATE
ENVY
GREED
GLUTTONY
TREASON
DESTRUCTION
ABOMINATION
ABANDONMENT
PANDEMONIUM
OCEANUS CRISIUM

Thumbnail - click to expand

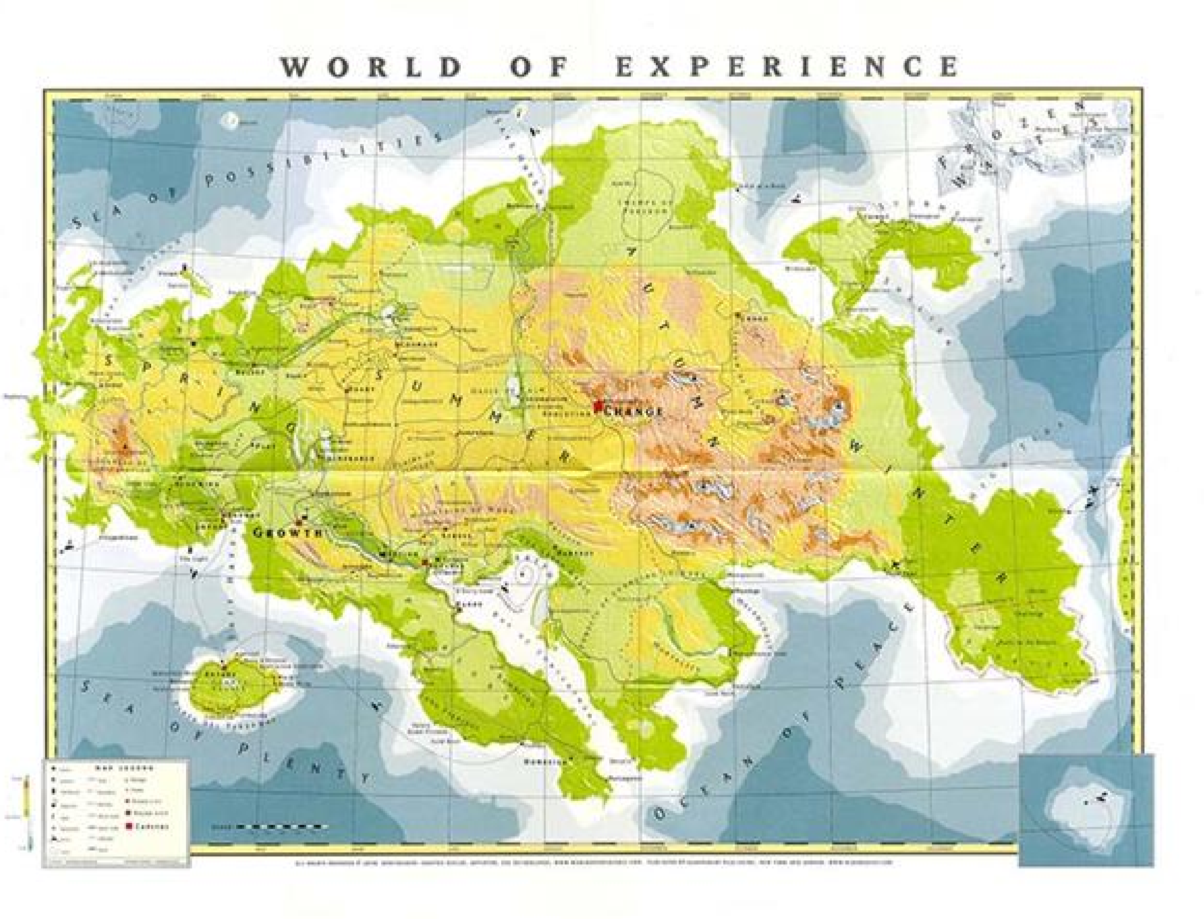
You can find more samples at this site. Turner also has a map of love and relationships, but it’s even harder to find any good images of.

WORKAHOLISM
PRANA OCEAN
STUDY
CONTEMPLATION
ARTHA
HERMITS
SOLITUDE
ZEN
CALM
WARPED
RELAXATION
ISOLATION
LOVE
LIBIDINOUS SEA
COMPROMISE
COMFORT
DHARMA
DOMESTIC BLISS
PASSION
ARGUMENT
AFFAIRS
DIVORCE
SEA OF SORROWS

This is all I’ve got

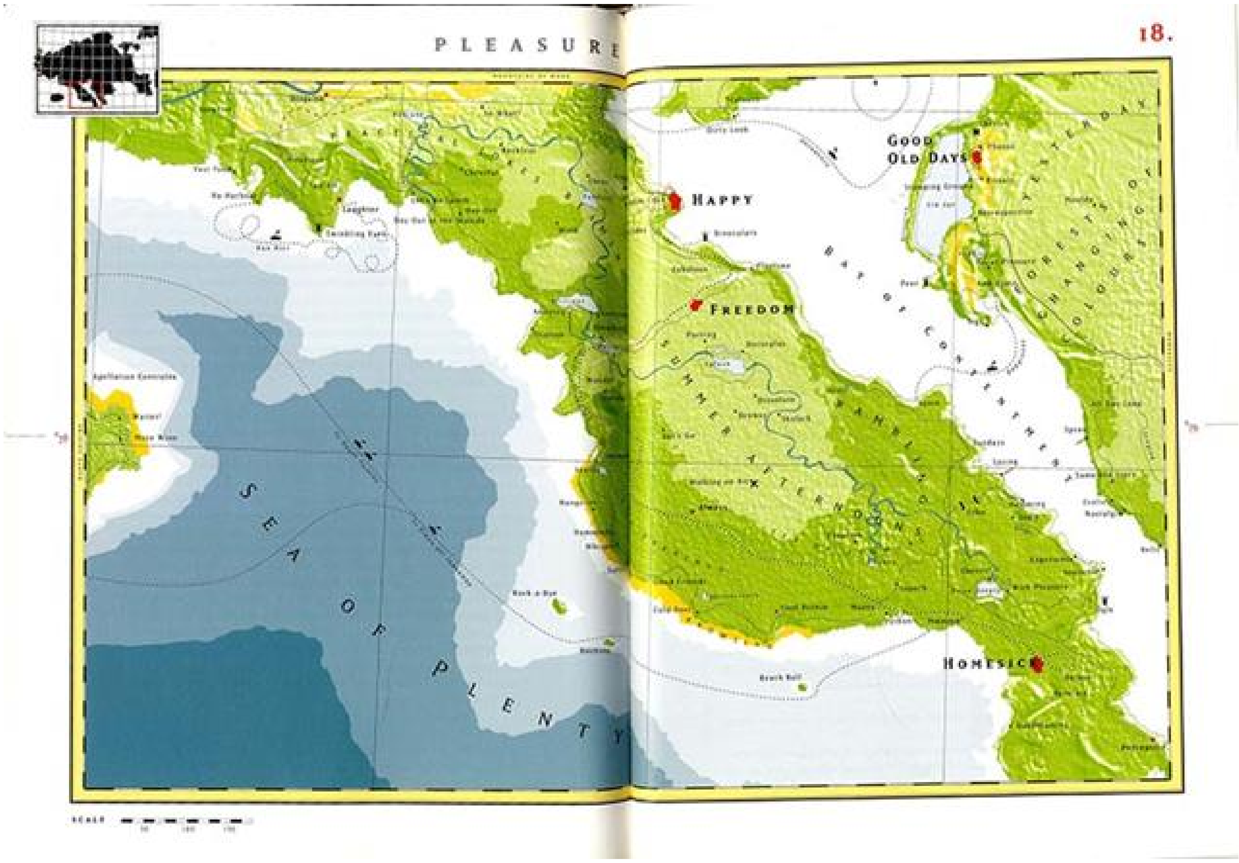
You can buy poster versions of both maps at the SLG Publishing Store, and you really should. This is one of my favorite things in the whole world, and it’s sold a grand total of seven copies so far (I think I might be two of the seven).

There are some similar works in the Atlas Of Experience, available in part as a website and in full as a book on Amazon. This seems to be their main offering:

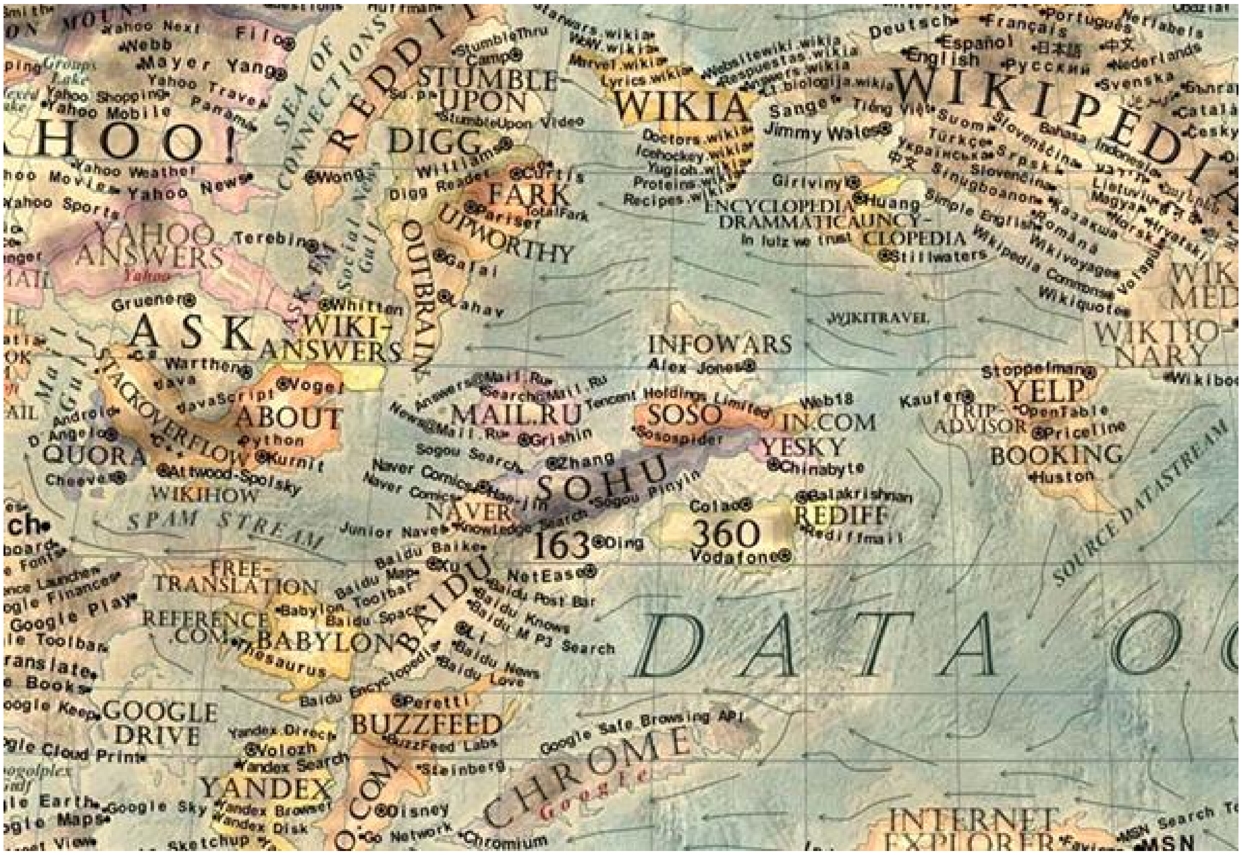

WORLD OF EXPERIENCE
SEA OF POSSIBILITIES
FROZEN WASTES
SPRING
SUMMER
AUTUMN
WINTER
CHANGE
GROWTH
SEA OF PLENTY
OCEAN OF PEACE
MAP LEGEND

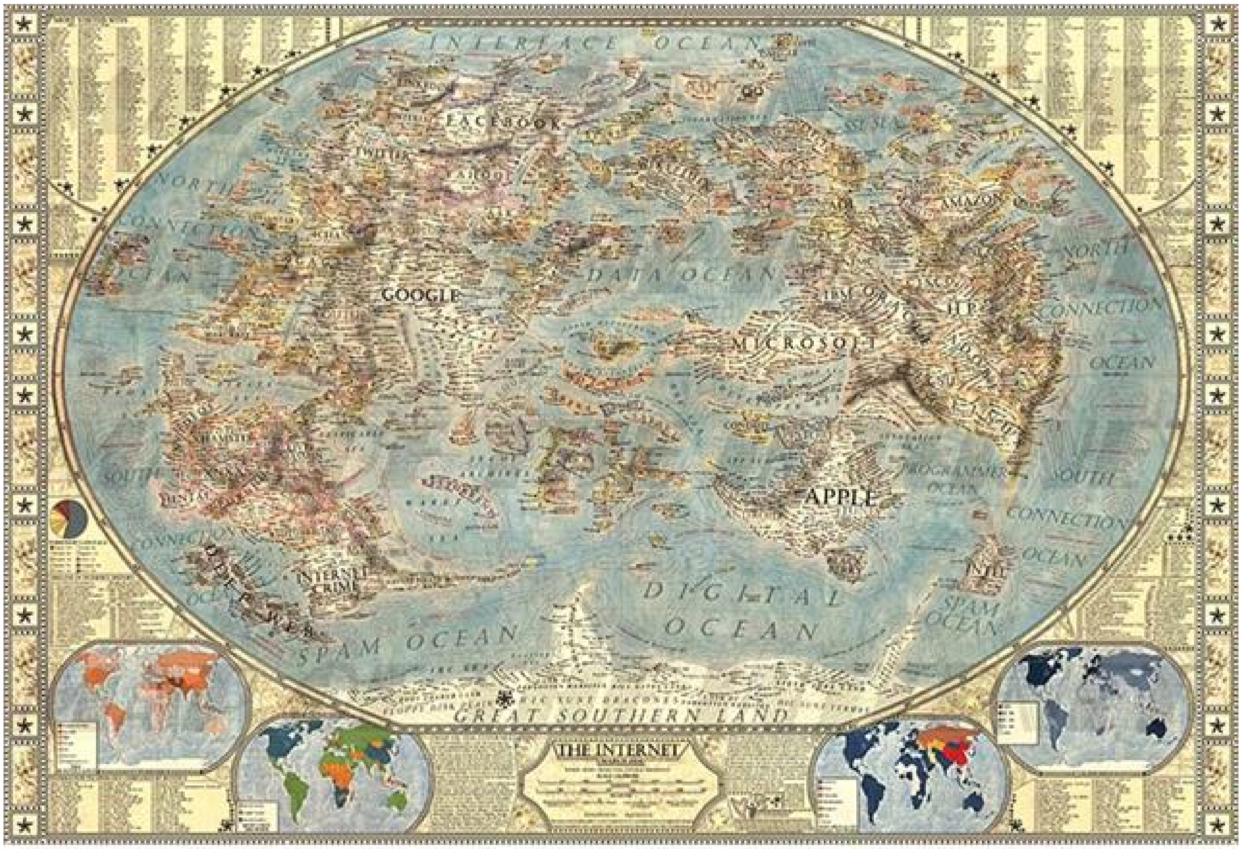
Thumbnail - click to expand

And this is a magnification of one of their peninsulae:


PLEASURE
18.
GOOD OLD DAYS
HAPPY
FREEDOM
BAY OF CONTENTMENT
YESTERDAY
FOREST OF CHANGING COLOURS
SUMMER AFTERNOONS
RAMBLING
SEA OF PLENTY
HOMESICK
SCALE

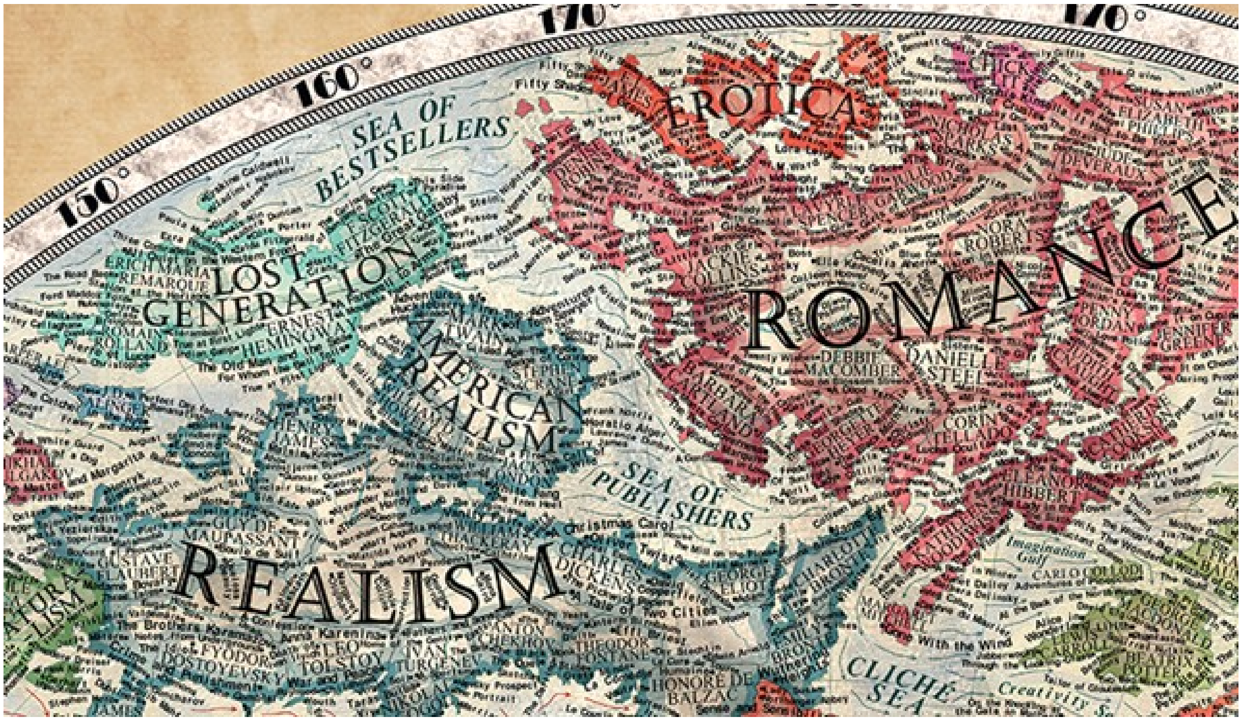
Martin Vargic has a bunch of really good sentimental cartography. The most easily accessible is his Map Of The Internet.

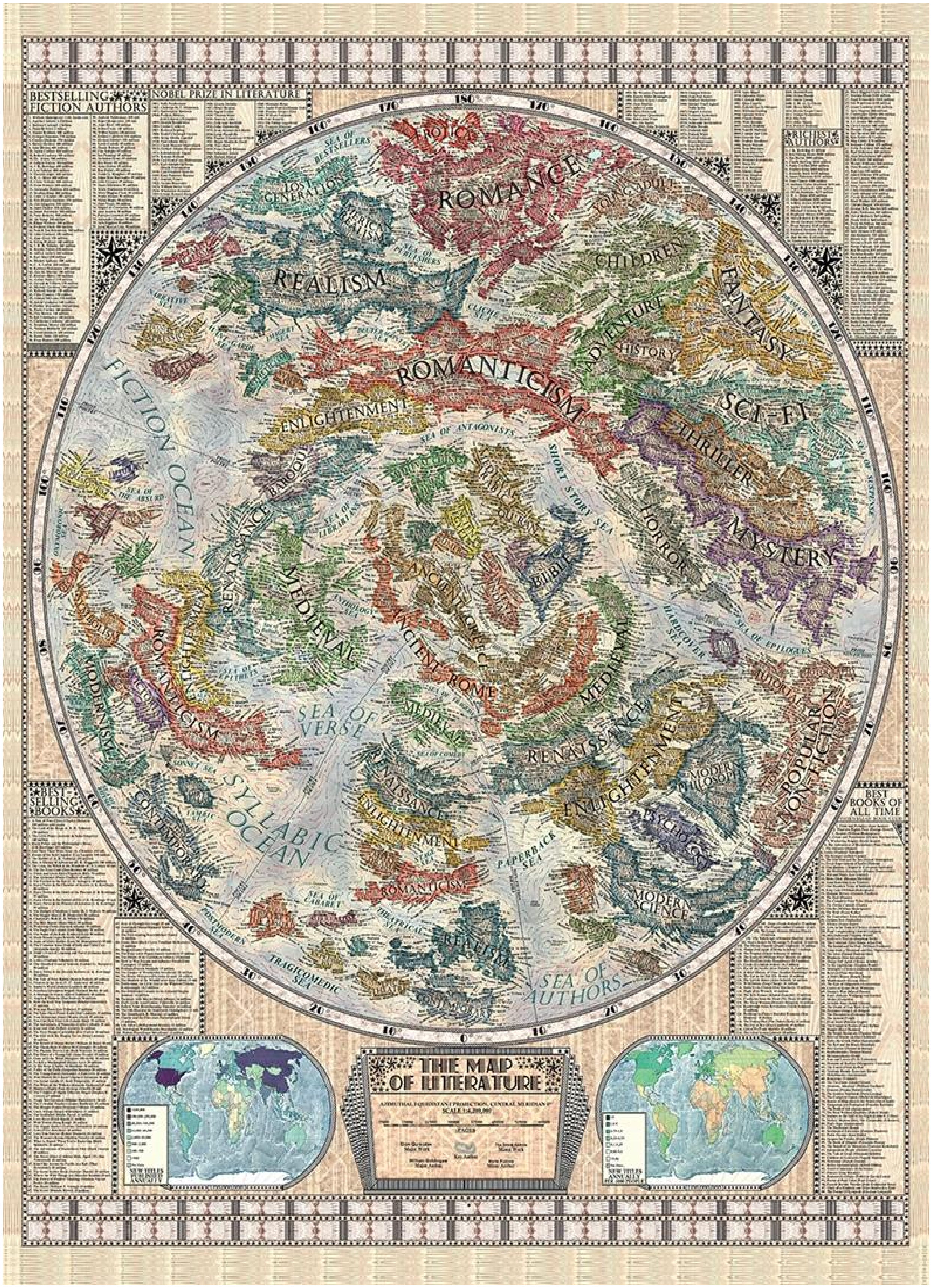

YAHOO!
Yahoo Next
Webb
Mayer
Yang
Filo
Yahoo Travel
Yahoo Shopping
Yahoo Mobile
Yahoo Movies
Yahoo Weather
Yahoo News
Yahoo Sports
YAHOO ANSWERS
SEA OF CONNECTIONS
REDDIT
STUMBLEUPON
StumbleThru
Camp
StumbleUpon Video
DIGG
Digg Reader
Williams
FARK
Curtis
TotalFark
Pariser
UPWORTHY
Gafai
OUTBRAIN
Lahav
Social News Gulf
Terebin
Wong
Whitten
Gruener
ASK
ASK.FM
WIKI-ANSWERS
Warthen
STACKOVERFLOW
Java
JavaScript
Android
Python
D'Angelo
QUORA
Cheever
Kurnit
Attwood-Spolsky
ABOUT
Vogel
WIKIHOW
SPAM STREAM
WIKIA
Starwars.wikia
WoW.wikia
Marvel.wikia
Lyrics.wikia
Doctors.wikia
Icehockey.wikia
Yugioh.wikia
Proteins.wikia
Recipes.wikia
Websitewiki.wikia
Respuestas.wikia
Answers.wikia
Ci.biologija.wikia
Sanger
Jimmy Wales
Girlviny
ENCYCLOPEDIA DRAMMATICA
UNCYCLOPEDIA
In lulz we trust
Huang
Stillwaters
WIKIPEDIA
Deutsch
Français
Português
English
Español
日本語
中文
Русский
Nederlands
Svenska
Català
Česky
Suomi
Türkçe
Українська
Slovenčina
Srpski
Slovenčina
Română
Norsk
Magyar
Lietuvių
Hrvatski
Bahasa Indonesia
Sinugboanon
Simple English
Қазақша
Wikivoyage
Wikipedia Commons
Wikiquote
WIKTIONARY
WIKIMEDIA
Wikibooks
WIKITRAVEL
INFOWARS
Alex Jones
MAIL.RU
Answers@Mail.Ru
Search@Mail.Ru
News@Mail.Ru
Grishin
Tencent Holdings Limited
SOSO
Sosospider
Web18
IN.COM
YESKY
Chinabyte
Sogou Search
Zhang
SOHU
Sogou Pinyin
Naver Comics
Hae-jin
NAVER
Junior Naver
Knowledge Search
163
Ding
NetEase
Colao
360
Vodafone
Balakrishnan
REDIFF
Rediffmail
Stoppelman
YELP
Kaufer
TRIP-ADVISOR
OpenTable
Priceline
BOOKING
Huston
SOURCE DATASTREAM
DATA O
FREE-TRANSLATION
REFERENCE.COM
Thesaurus
BABYLON
Babylon Toolbar
BAIDU
Baidu Baike
Baidu Map
Baidu Space
Baidu Post Bar
Baidu Knows
Baidu MP3 Search
Li
Baidu News
Baidu Love
Baidu Encyclopedia
Xu
Google Play
Google Finance
Google Toolbar
Translate
Books
Google Keep
GOOGLE DRIVE
Google Cloud Print
Googolplex
Google Earth
Google Maps
Google Sky
Sketchup
Yandex Direct
Volozh
Yandex Search
YANDEX
Yandex Browser
Yandex Disk
BUZZFEED
Peretti
BuzzFeed Labs
Steinberg
GO.COM
Disney
Go Network
Google Safe Browsing API
CHROME
Google
Chromium
INTERNET EXPLORER
MSN Search
MSN


INTERFACE OCEAN
FACEBOOK
TWITTER
YAHOO
NORTH CONNECTION OCEAN
GOOGLE
DATA OCEAN
AMAZON
NORTH CONNECTION OCEAN
IBM
HP
MICROSOFT
APPLE
PROGRAMMER OCEAN
SOUTH CONNECTION OCEAN
SOUTH CONNECTION OCEAN
INTERNET CRIME
DEEP WEB
SPAM OCEAN
DIGITAL OCEAN
SPAM OCEAN
GREAT SOUTHERN LAND
THE INTERNET

Thumbnail - click to expand

He also has a Map of Literature, not particularly accessible except for teasers:

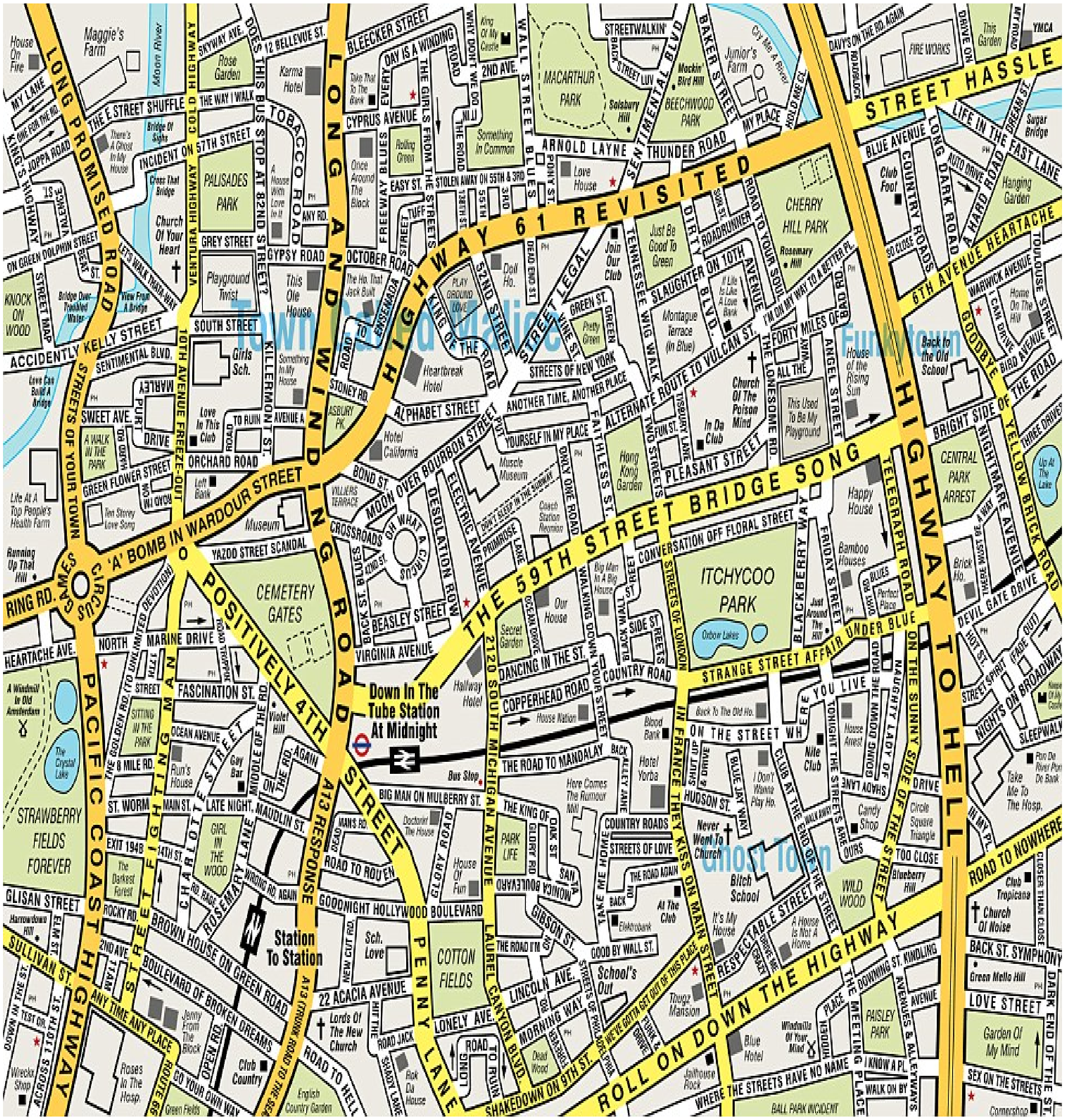

150°
160°
170°
SEA OF BESTSELLERS
EROTICA
LOST GENERATION
ERICH MARIA REMARQUE
ROMAIN ROLLAND
ERNEST HEMINGWAY
F. SCOTT FITZGERALD
AMERICAN REALISM
MARK TWAIN
HENRY JAMES
STEPHEN CRANE
JACK LONDON
ROMANCE
JACKIE COLLINS
NORA ROBERTS
DEBBIE MACOMBER
DANIELLE STEEL
BARBARA CARTLAND
CORIN TELLADO
ELEANOR HIBBERT
NICHOLAS SPARKS
JUDE DEVERAUX
SUSAN ELIZABETH PHILLIPS
SEA OF PUBLISHERS
REALISM
GUY DE MAUPASSANT
GUSTAVE FLAUBERT
FYODOR DOSTOYEVSKY
LEO TOLSTOY
IVAN TURGENEV
ANTON CHEKHOV
CHARLES DICKENS
GEORGE ELIOT
THEODOR FONTANE
HONORE DE BALZAC
CLICHÉ SEA
Imagination Gulf
Creativity Sea
CARLO COLLODI
GEORGE MACDONALD
LEWIS CARROLL
BEATRIX POTTER


THE MAP OF LITERATURE
BESTSELLING FICTION AUTHORS
NOBEL PRIZE IN LITERATURE
RICHEST AUTHORS
BEST SELLING BOOKS
BEST BOOKS OF ALL TIME
ROMANCE
YOUNG ADULT
CHILDREN
FANTASY
REALISM
LOST GENERATION
ADVENTURE
HISTORY
ROMANTICISM
SCI-FI
ENLIGHTENMENT
THRILLER
MYSTERY
HORROR
FICTION OCEAN
SEA OF BESTSELLERS
SEA OF PUBLISHERS
SEA OF ANTAGONISTS
SHORT STORY SEA
SEA OF THE ABSURD
SEA OF LIBRARIANS
RENAISSANCE
MEDIEVAL
ANCIENT GREECE
ANCIENT ROME
BIBLE
MODERNISM
SEA OF EPITHETS
SEA OF VERSE
HARDCOVER SEA
SEA OF EPILOGUES
TUTORIAL
POPULAR NON-FICTION
MODERN PHILOSOPHY
PSYCHOLOGY
MODERN SCIENCE
CONTEMPORARY
SYLLABIC OCEAN
SONNET SEA
PAPERBACK SEA
SEA OF CABARET
THEATRICAL SEA
POSTMODERN SEA
TRAGICOMEDIC SEA
SEA OF AUTHORS
NEW TITLES ANNUALLY
NEW TITLES ANNUALLY PER 1,000 PEOPLE

Not a thumbnail, sorry – this is the biggest I have

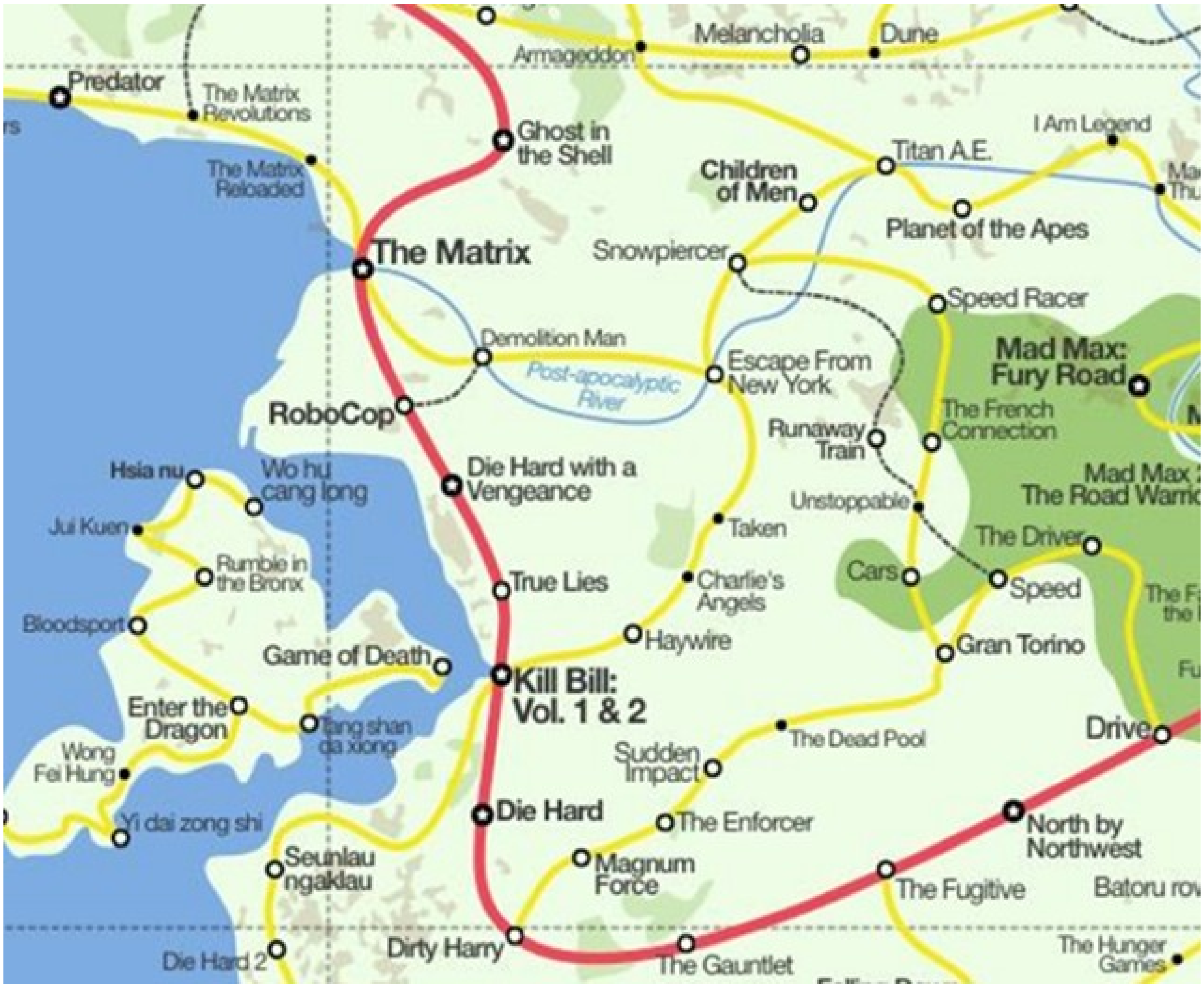
The full version of this map and several others with no online presence are available in Vargic's book Miscellany Of Curious Maps.

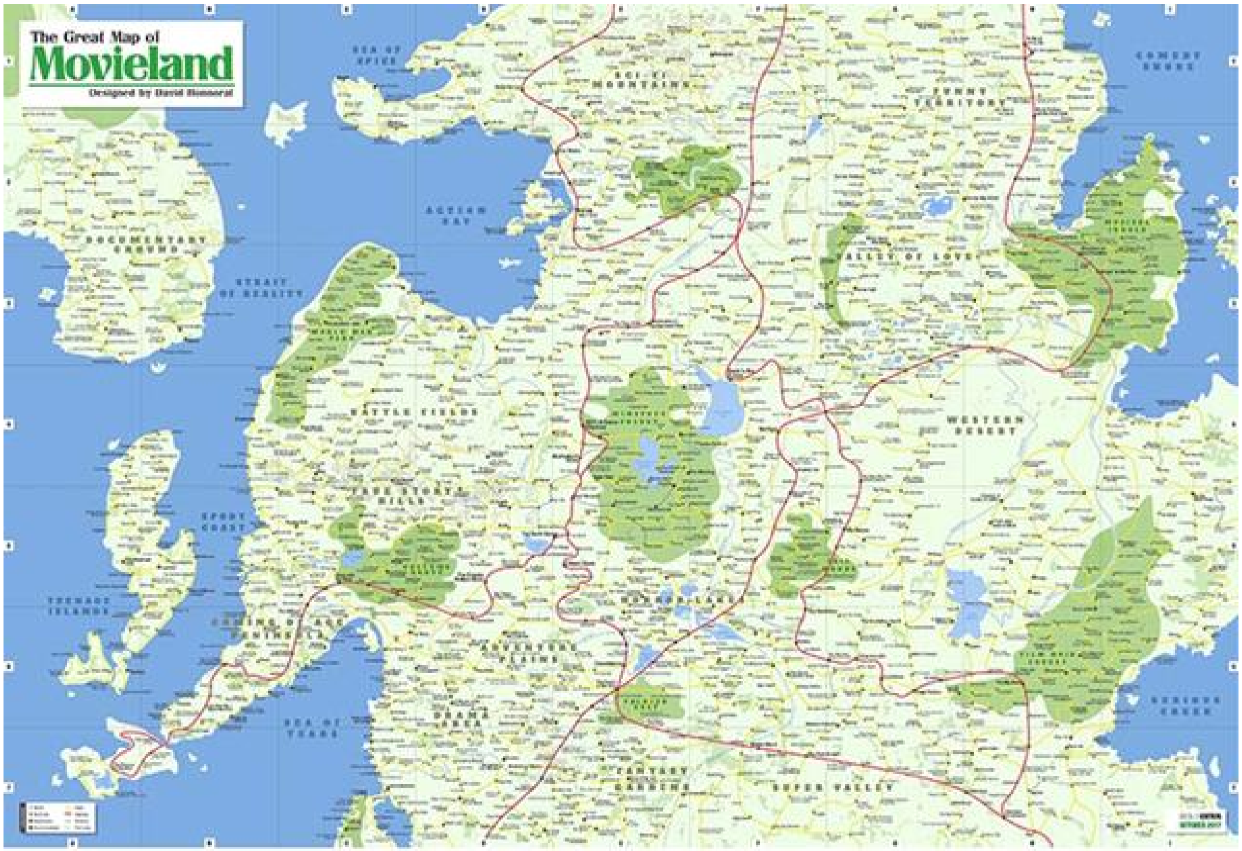
This is road map of songs whose names sound like things that should be on a road map:


LONG PROMISED ROAD
LONG AND WINDING ROAD
HIGHWAY 61 REVISITED
STREET HASSLE
HIGHWAY TO HELL
THE 59TH STREET BRIDGE SONG
PACIFIC COAST HIGHWAY
POSITIVELY 4TH STREET
'A' BOMB IN WARDOUR STREET
ROLL ON DOWN THE HIGHWAY
PENNY LANE
Town Called Malice
Funkytown
Ghost Town
ITCHYCOO PARK
CEMETERY GATES
STRAWBERRY FIELDS FOREVER
COTTON FIELDS
CHERRY HILL PARK
PALISADES PARK
MACARTHUR PARK
BEECHWOOD PARK
Down In The Tube Station At Midnight
Station To Station
Heartbreak Hotel
Hotel California
Maggie's Farm
Junior's Farm
Love Shack
STRANGE STREET AFFAIR UNDER BLUE
ALPHABET STREET
DESOLATION ROW
BLEECKER STREET
BAKER STREET
THUNDER ROAD
GOOD BY WALL ST.
BLUE AVENUE
LIFE IN THE FAST LANE
6TH AVENUE HEARTACHE
GOODBYE YELLOW BRICK ROAD
NIGHTS ON BROADWAY
ROAD TO NOWHERE
SOUTH MICHIGAN AVENUE
LAUREL CANYON BLVD
STREETS OF YOUR TOWN
STREETS OF LONDON
STREETS OF NEW YORK
STREETS OF LOVE
ON THE STREET WHERE YOU LIVE
WHERE THE STREETS HAVE NO NAME
DANCING IN THE ST.
GOODNIGHT HOLLYWOOD BOULEVARD
BOULEVARD OF BROKEN DREAMS
ANY TIME ANY PLACE
THE ROAD TO MANDALAY
ELECTRIC AVENUE
TENNESSEE WIG WALK
ALTERNATE ROUTE TO VULCAN ST.
BRIGHT SIDE OF THE ROAD
DEVIL GATE DRIVE
CENTRAL PARK ARREST
Hong Kong Garden
Hanging Garden
English Country Garden
Roses In The Hosp.

The poster version is available for purchase here, as are a similar TV map, book map, game map, and film map.

And speaking of film maps, a random Reddit commenter made one:


Melancholia
Dune
Armageddon
Predator
The Matrix Revolutions
Ghost in the Shell
I Am Legend
Titan A.E.
The Matrix Reloaded
Children of Men
Planet of the Apes
The Matrix
Snowpiercer
Speed Racer
Demolition Man
Post-apocalyptic River
Escape From New York
Mad Max: Fury Road
RoboCop
The French Connection
Runaway Train
Hsia nu
Wo hu cang long
Die Hard with a Vengeance
Jui Kuen
Unstoppable
Taken
The Driver
Rumble in the Bronx
Cars
Speed
Charlie's Angels
True Lies
Bloodsport
Haywire
Gran Torino
Game of Death
Kill Bill: Vol. 1 & 2
Enter the Dragon
Tang shan da xiong
The Dead Pool
Drive
Wong Fei Hung
Sudden Impact
Die Hard
The Enforcer
North by Northwest
Yi dai zong shi
Seunlau ngaklau
Magnum Force
The Fugitive
Dirty Harry
Die Hard 2
The Gauntlet
The Hunger Games


The Great Map of
Movieland
Designed by David Honnorat
SEA OF SPICE
SCI-FI MOUNTAINS
FUNNY TERRITORY
COMEDY SHORE
ACTION BAY
DOCUMENTARY GROUND
STRAIT OF REALITY
VALLEY OF LOVE
BATTLE FIELDS
WESTERN DESERT
TRUE STORY HILLS
SPORT COAST
TEENAGE ISLANDS
COMING OF AGE PENINSULA
ADVENTURE PLAINS
SEA OF TEARS
DRAMA AREA
FANTASY RANCHES
SUPER VALLEY

Thumbnail - click to expand

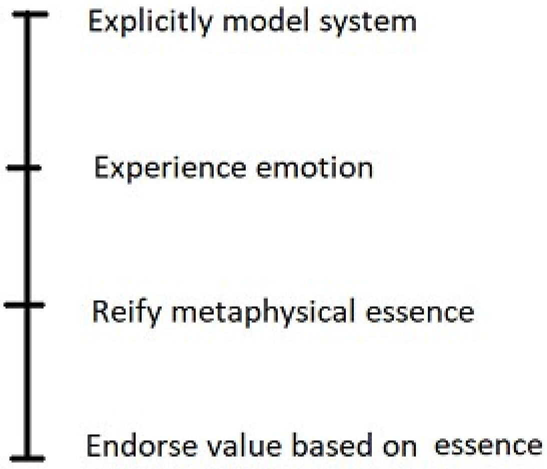
That’s it. These are literally all the good modern sentimental cartography maps I know. This is a good art, it shouldn’t have died with the Renaissance, and people should do more work to resurrect it.

Value Differences As Differently Crystallized Metaphysical Heuristics

*[**Previously in sequence:** Fundamental Value Differences Are Not That Fundamental, The Whole City Is Center. This post might not make a lot of sense if you haven't read those first.]*

I.

Thanks to everyone who commented on last week's posts. Some of the best comments seemed to converge on an idea like this:


Explicitly model system
Experience emotion
Reify metaphysical essence
Endorse value based on essence

Confusing in that people who rely on lower-level features are placed higher, but the other way would have been confusing too.

We need to navigate complicated philosophical questions in order to decide how to act, what to do, what behaviors to incentivize, what behaviors to punish, what signals to send, and even how to have a society at all.

Sometimes we can use theories from science and mathematics to explicitly model how a system works and what we want from it. But even the scholars who understand these insights rarely know exactly how to objectively apply them in the real world. Yet anyone who lives with others needs to be able to do these things; not just scholars but ordinary people, children, and even chimpanzees.

So sometimes we use heuristics and approximations. Evolution has given us some of them as instincts. Children learn others as practically-innate hyperpriors before they're old enough to think about what they're doing. And cultural evolution creates others alongside the institutions that encourage and enforce them.

In the simplest case, we just feel some kind of emotional attraction or aversion to something.

In other cases, the emotions are so compelling that we crystallize them into a sort of metaphysical essence that explains them.

And in the most complicated cases, we endorse the values implied by those metaphysical essences above and beyond whatever values we were trying to model in the first place.

Some examples:

People and animals need a diet with the right number of calories, the right macronutrient ratios, and the right vitamins and minerals. A few nutritional scientists know enough to figure out what's going on explicitly. Everyone else has evolved instincts that guide them through this process. Hunger and satiety are such instincts; when they're working well, they make sure someone eats as much as they need and no more. So are occasional cravings for some food with exactly the right nutrient – most common in high-nutrient-use states like pregnancy. But along with these innate heuristics, we have culturally determined ones. Everyone has a vague sense that potato chips are “unhealthy” and spinach is “healthy”, though most people can’t explain why. Instead of asking ordinary people and children to calculate their macronutrient and micronutrient profile, we ask them to eat “healthy” foods and avoid “unhealthy” foods. There’s something sort of metaphysical about this – as if “health” were a magic essence that adheres to apples. And in fact, sometimes this goes wrong and people will do things like blend a thousand apples into some hyper-pure apple-elixir to get extra health-essence – but overall it mostly works.

EXPLICIT MODEL: Trying to count how many calories and milligrams of each nutrient you get
EMOTIONAL EXPERIENCE: Feeling hungry or full
REIFIED ESSENCE: Some foods are inherently healthy or unhealthy
ENDORSED VALUE: Insisting on only eating organic foods even when those foods have no quantifiable benefit over nonorganic

Every society has some kind of punishment for people who don’t follow their norms, whether it’s ostracism or community service or beheading. There’s a good consequentialist grounding for why this is necessary, with some of the most academic work being done in the field of prisoners’ dilemmas and tit-for-tat strategies. But again, we don’t expect ordinary people, children, and chimpanzees to absorb this work. The solution is the (innate? culturally learned? some combination of both?) idea of punishment. Punishment relies on a weird metaphysical essence of moral desert; people who do bad things deserve to suffer. The balance of the Universe is somehow off when a crime goes unavenged. Take this too far and you get the Erinyes and the idea that justice is the most important thing. There are references from ancient China to *Hamlet* that if you have something important you need to avenge, you need to do that now or you’re a bad person. None of this follows from the game theory, but it’s a really good way to enforce the game-theoretically correct action.

EXPLICIT MODEL: Trying to figure out how to best deter antisocial behavior and optimize society
EMOTIONAL EXPERIENCE: Feeling angry when someone wrongs you
REIFIED ESSENCE: Justice: the world is out of balance when crimes go unavenged
ENDORSED VALUE: Wrongdoers must suffer whether or not that prevents future crimes

If you reward people who create value, sometimes those people will be inspired to keep creating value. This is hard for people to keep in mind, and there’s a constant temptation to confiscate other people’s things for our own enrichment. Some kind angel gave us the metaphysical idea of “deserving”, the opposite of punishment. We get rights claims like “People deserve to keep what they’ve earned”. Five thousand years of taxation have made only a partial dent in this intuition, to the point where many people still feel like something is going wrong when a producer and the value they produced are separated. Some would argue that this has gotten completely out of hand, to the point where we insist on people keeping money far past the point where it could possibly be any further incentive to them.

EXPLICIT MODEL: Letting people keep what they produce incentivizes further production
EMOTIONAL EXPERIENCE: Anger when someone takes something rightfully yours
REIFIED ESSENCE: Natural rights; governments cannot take away property rights because they are ordained by God or natural law

ENDORSED VALUE: You can’t take people’s property, whether or not this will affect further production

In past societies, STDs were a common cause of death and disfigurement. Nobody had the medical knowledge to really understand what an STD was or how to avoid getting one. But every society had some kind of complicated code of sexual purity. Usually this was designed from a male perspective, and said that women who had sex with too many other men were “impure”, virgins were especially “pure”, and a woman who had only had sex with you was “pure” relative to you. These rules protect people who follow them against STDs, and plausibly culturally evolved for that purpose (among others). But because no one knew about STDs, the rules rely on a kind of metaphysical notion of “purity” that doesn’t correspond to any real-world characteristic. For example, someone who’s had sex with a hundred people but who nevertheless never contracted an STD would seem metaphysically “impure” by the rules, but in reality safe to have sex with; this would be irrelevant to medievals who had no way to identify such people, but is very relevant now. Or: if you have good sexual protection and STDs are easily treatable, the whole “purity” system seems a lot less important, but if you think of it as a metaphysical construct important in its own right you might not realize this.

(before you tell me that STDs aren’t important enough to inspire something as universal and compelling as sexual purity laws, remember that in the pre-antibiotic era about 10% of city-dwellers had syphilis (see studies from early Mesoamerica, 1700s Chester, early 1900s London. During this period syphilis had a mortality rate of up to 20%, with survivors often permanently unhealthy and disfigured. And this is just one of many dangerous STDs!)

EXPLICIT MODEL: Figuring out the likelihood that your partners have STDs helps you avoid high-risk pairings
EMOTIONAL EXPERIENCE: Feeling grossed out at the idea of having sex with somebody who "sleeps around"
REIFIED ESSENCE: Idea of sexual purity
ENDORSED VALUE: It's wrong to be slutty or have sex with a slutty person even if there are effective strategies for preventing STDs

Most people are happier when they're in at least some Nature, whether this means a grand national park or just a leafy suburb with lots of chirping birds. The average person would consider a concrete lot full of Brutalist apartments a little soul-crushing. This probably comes from an evolutionary heuristic in favor of fertile areas and against barren ones; the closest chimpanzee-parseable equivalent to a concrete lot would be a desert or lava flow, where food and shelter are scarce. But nowadays we can order takeout, and the Brutalist apartment buildings provide all the shelter we need. This is probably another obsolete evolutionary relic, but it's a very persistent one.

EXPLICIT MODEL: More plants and less gray rock means a more hospitable area with more food sources
EMOTIONAL EXPERIENCE: Contentment when surrounded by plants; depression when surrounded by concrete
REIFIED ESSENCE: Idea of "Nature"
ENDORSED VALUE: Environmentalism; the preservation of Nature for its own sake whether it benefits humans or not

Value differences, then, are people who operate at different levels of the ladder.

For example, "sexually liberated" people might use condoms, or ask their partners to check if they have STDs. But having done this, they'll ignore the metaphysical idea of “purity”; they don’t care how many other people their partner has had sex with. And if they’re not planning to have sex with someone, they’ll ignore their “purity” full stop – having STDs doesn’t make you a bad person. These people explicitly model the complicated dynamic of STD contagion, and cast off the metaphysics as a primitive approximation they no longer have any use for.

Traditional or religious communities are more likely to endorse values based on the “purity” metaphysical essence. They understand the biology of STDs just as well as the sexually liberated people. They just don’t care. On an intellectual level, they believe that sexual purity is more than just a predictive model of STD risk, or that they’ve gained some additional function in the meantime, or that the predictive model still works better than calculating it out explicitly. Or they might not think in these terms at all, and value purity as a terminal good. Or they might be following their instincts in a way that can’t be reducible to anything at all.

Other people are somewhere in between. I know in theory that I cannot get AIDS through touching infected blood left on a sheet or chair. Absent some sort of very unlikely chain of events involving weird mouth ulcers and very fast turnaround times, I can’t even get it by rubbing the blood on food and eating it. But I would still feel more than zero trepidation about doing this. It would seem that even though I identify with the sexually-liberated explicit-complicated-dynamic-modellers, I have some weak vestige of the metaphysical purity heuristic left. This makes me more sympathetic to people with the full version. They don’t seem like weird mutants too stupid to figure out what an STD is, they feel like people with my instincts magnified a million times until they’ve become irresistible.

I have tried very hard to cultivate a vital rationalist skill called “admitting I am being an idiot while feeling no obligation to change”. That is, I feel comfortable saying I’m being very silly by objecting to touching the HIV-infected blood. If someone were to lecture me that admitting this obligates me to touch the blood or else I will have proven myself a hypocrite, I would tell them to go jump in a lake. This is important because I’m pretty sure my purity-instinct urge not to eat HIV-infected blood is stronger than my urge to be right about factual issues. If I was forced to either eat the blood, or to make up some plausible-but-false reason why the experts were wrong and blood was unsafe, I would make up the reason. And then not only would I not eat the blood – a venial sin if ever there was one – but I would be obfuscating the debate, screwing up lots of people’s ideas about epidemiology, and taking a step on the slippery slope toward becoming a dishonest and dishonorable person. I would rather just admit I’m silly.

But I think this is a hard skill, one that I often get wrong even despite frequent practice, and one I don’t expect anyone to succeed at all the time. I think some people with strong metaphysical heuristics – around HIV, around sex, around whatever – are going to get to work justifying them. If many people share a certain strong metaphysical heuristic, then there will be entire communities dedicated to researching justifications for it, coming up with philosophy around it, and reinforcing one another for being wise and good enough to believe it.

I think this is a big part of where value differences come from, and why I’ve insisted that despite the differences being *real*, they’re not *incomprehensible*. Most people have at least some level of metaphysical-STD-purity-intuition. And most people have at least some level of explicit-dynamic-modeling-of-STD-risk. Our differences come not from some people being enlightened and other people being mutants with the bizarre idea of “sluttiness” as a terminal bad, but by settling on a different part of the ladder from totally-endorsed-value-based-on-essence to total-explicit-modeling.

II.

A natural interpretation of Part I: people with explicit modeling are smart and good, people who still use metaphysical heuristics are either too hidebound to switch or too stupid to do the modeling.

I think this is partly right, but since our goal is to make value differences seem less clear-cut and fundamental, I want to make the devil’s advocate case for respecting metaphysical heuristics.

First, the heuristics are, if nothing else, proven to be compatible with continuing to live; the explicit models often suck.

Soylent uses an explicit model of nutrition to try to replace our vague heuristics about “eating healthy”. I am mostly satisfied with the quality of its research; it generally avoids stupid mistakes. It does not completely avoid them; the product has no cholesterol, because “cholesterol is bad”, but the badness of cholesterol is controversial, and even if we grant the basic truth of the statement, it applies only at the margin in the standard American diet. If you eat only one food item, you had better get that food item *really* right, and it turns out that having literally zero cholesterol in your diet is long-term dangerous. This was an own-goal, and a smarter explicit modeler could have avoided it. But explicit models that only work when you get everything exactly right will fail 95% of the time for geniuses and 100% of the time for the rest of us.

And even if Soylent had avoided own-goals, they still risk running up against the limit of our understanding. Decades ago, doctors invented a Soylent-like fluid to pump into the veins of patients whose digestive systems were so damaged they could not eat normally. These patients tended to get a weird form of diabetes and die. After a lot of work, the doctors discovered that chromium – of all things – was actually a really important dietary nutrient, and nobody had ever noticed before because it’s more or less impossible to run out of chromium with any diet except having synthetic fluids pumped into your veins. After years of progress on nutritional fluids, the patients who need them no longer die; we can be pretty sure we’ve found everything that’s fatal in deficiency. But these patients do tend to feel much worse, and be much less healthy, than people eating normal diets. How many mildly-important trace micronutrients are left to discover? And how many of these are or aren’t in Soylent?

We know that for some reason eating multivitamins does not work as well *even for vitamin-having purposes* as eating food with the relevant vitamins in them. This seems to have something to do with absorption and bioavailability, but we’re not sure what. Does Soylent have the good bioavailability of food, or the bad bioavailability of multivitamins? Nobody knows, because we still don’t quite understand how bioavailability works. All we know is that evolution seems to have found one viable solution, given that people who eat food do not immediately die. If we replace food with intelligent application our best available explicit models, we might do okay – or we might feel vaguely ill all the time because there’s something important we’re missing.

On the society-wide level, the sort of explicit-modeling that created Soylent becomes high modernism, the philosophy critiqued in James Scott’s *Seeing Like A State*. You subject everything to the command of a central planner, who is supposed to be able to explicitly model social dynamics, and try to prevent people from using fuzzy evolved heuristics like tradition or “the way things are”. The extreme version of this is those Young Adult Dystopias: can’t justify exactly why there should be families? Then families are just obsolete detritus of our evolutionary past, and we should form a Department Of Child-Rearing that takes all kids and subjects them to carefully-doled-out industrial-scale parenting techniques.

Second, all of our values are unjustifiable crystallizations of heuristics at some level, and we have to have *some* value.

One of the examples above supposes that our love of nature comes from heuristics about where to find food and water. Suppose we proved this conjecture was right. Given that we can now order pizza and bottled water to concrete lots, the heuristic is obsolete. Does this mean we should stop caring about nature, and cut down all our forests and national parks and replace them with concrete lots? Suppose this would be very profitable, and that on cost-benefit analysis this outweighs the practical economic benefits of wild spaces (carbon sinks, drug discovery from exotic species). Is there any remaining reason we still want the national parks?

Compare this to punishment-for-the-sake-of-punishment. Maybe now we can replace this with an explicit model of consequentialist punishment where we should only punish people up to the point where it’s necessary to have a safe and stable society. Returning to the dialogue:

> **Simplicio:** I admit – I believe in punishment in a sense stronger than as a heuristic for consequentialism. I think it’s morally important, in a terminal sense, that evildoers be made to suffer for their deeds. Not suffer

infinitely. But suffer some amount proportional to how much they hurt others. I want this regardless of whether it deters them or not.

Sophisticus: But that’s just reifying a weird misfiring of an obsolete heuristic about how to maintain a safe community.

Simplicio: Yup! And me wanting Yellowstone to continue to exist is just reifying a weird misfiring of an obsolete heuristic about how to get delicious elk meat. And surely you don’t want to pave over Yellowstone.

Sophisticus: I take joy in Yellowstone. That’s an emotional experience in my brain. I’m happier and more comfortable in Nature. Even if the heuristics that produced this are wrong, that feature of my brain isn’t going away any time soon. So on a consequentialist level, I can argue that Yellowstone should be maintained for my sake and the sake of everyone else who enjoys it, even though I’m not sure my enjoyment comes from a reasonable source.

Simplicio: I take joy in watching murderers and rapists get what they deserve. This is a base-level pleasure for me, just like seeing trees and mountains are for you. I am under no more imperative to justify what I want than you are.

Sophisticus: You are, though. Because you directly desire for people to suffer, which violates some of our other shared values. We have to reach reflective equilibrium among our values, and for me at least the value to wish happiness rather than suffering on other people overwhelms the desire for punishment.

> **Simplicio:** First, I think we should be careful to frame it the way you just did: "Making people suffer for their crimes is good, but this is outweighed by other goods". If we say it that way, it sounds no more exotic than the trolley problem.
>
> **Sophisticus:** It's at a –
>
> **Simplicio:** – but second, if paving over Yellowstone would have economic benefits, then those benefits would cash out in jobs, lower housing costs, cheaper consumer goods, and the like. All of those produce utility for people. Both of our weird preferences – mine for punishment, yours for nature – satisfy some crystallized heuristic at the expense of general utility. I still fail to see how we're different.
>
> **Sophisticus:** I agree that preserving Yellowstone may incidentally fail to maximize utility. But it seems like you're directly aiming at reducing people's utility. That's a pretty big difference.
>
> **Simplicio:** Exactly which principle are you invoking here? The act-omission distinction? Or the principle that the morality of an act depends upon what feelings are going through your head when you do it?
>
> **Sophisticus:** Um…
>
> **Simplicio:** Because I think both of those are sometimes useful – as heuristics. But if you're going crystallize those heuristics, let me have my crystallized heuristic about punishment.

Simplicio is actually being nice here. If he wanted to be especially brutal, he might ask Sophisticus something like – wait, why are we privileging utilitarianism (here being called “consequentialism”, but it seems like both of them are working from an implicitly utilitarian framework) anyway? Utilitarianism says that what’s really important is reducing suffering, but we can invent an evolutionary story for that too. We want to help other people and make them happy because that’s a useful heuristic for creating a flourishing community, being well-liked, and being likely to have other people help us in our own time of need. But some utilitarian applications of this principle go beyond that; certainly caring about effective charity for the Third World, or wild animal suffering, or anything in those realms brings us just as far from the proper domain of our help-and-don’t-harm-others heuristic as how to build a suburb with widely-available pizza delivery takes us from our nature-as-fertile-lands heuristic. Why should we privilege the harm foundation over the justice foundation? Why not just say “My urge to relieve suffering conflicts with my urge to inflict punishment on evildoers. Both urges have their place, and either can be extended out to infinity with weird results. Today I choose my urge to inflict punishment; tomorrow I might choose the other. So it goes.”

To be absolutely brutal about it:

EXPLICIT MODEL: Helping others will key me in to networks of reciprocal altruism and raise my status in the community
EMOTIONAL EXPERIENCE: Desire to help others, empathy, horror at the suffering of others
REIFIED ESSENCE: “Utility”
ENDORSED VALUE: Utilitarianism, the belief that maximizing utility is the highest good regardless of what other goods it produces

III.

Leave it there, and the fundamental-value-differences narrative starts to sound more appealing again. I reify and endorse utility, you reify and endorse punishment, now we have to fight. So I want to talk about how in principle people end up choosing what level to crystallize heuristics at.

First, let’s be blunt: dumber (here meaning either less educated or lower-IQ) people probably crystallize heuristics lower on the ladder. Chimpanzees, cavemen, and children can’t understand game theory and shouldn’t try. They usually run off instinct and taboo, and if you take that away from them they will just get confused.

There are widely replicated findings that higher-IQ and more-educated people tend to be less socially conservative. Social conservativism means a lot of things, but I think in this case it’s probably a stand-in for where you crystallize your heuristics; sexual purity intuitions are an obvious example. This makes sense; smarter people are probably more successful at explicit models, or at least have a higher estimation of their likelihood of success at such models. Smarter people do better on the Cognitive Reflection Test, a measure of whether people go with snap intuitive answers or try to explicitly model situations.

But there’s also reason to think that the more exposure someone has to a heuristic-relevant situation, the more compelling the heuristic will be. I described how my great-grandmother, usually a very kind and forgiving person, became more vengeful once someone close to her was murdered; I was able to partly replicate her experience just by vividly imagining terrible crimes happening to people close to me. This matches the cliche that “a conservative is a liberal who has been mugged; a liberal is a conservative who has been arrested”.

One of the weirdest examples of this is the germ theory of democracy, which finds the presence of tolerant multicultural individualist societies to be correlated with pathogen stress even after accounting for other relevant confounders. In this view, people at high risk of disease feel an urge to stick to people they know well – their family members, neighbors, and co-ethnics – to avoid the sort of mixing that spreads exotic pathogen strains. People at low risk of disease are more cosmopolitan, happy to receive anyone who comes around.

Related: people crystallize heuristics on a lower level when the system the heuristic is meant to model is a system they care about getting right. Consider Haidt's Moral Foundation of Authority, which he says conservatives have and liberals lack. This fits nicely into the explicit-model-to-essence-to-endorsed-value model. The explicit reasoning is that social groups need to coordinate, and once whatever mechanism you have to produce rules has produced its rules, people need to respect and listen to them or else they'll be in a Hobbesian state of nature. Liberals may say they're "against authority", but when the Vice-President of the NAACP asks an NAACP staffer to prepare a report by next week, she will probably prepare a report by the next week, not just because she's afraid of being fired but because the NAACP will fail if it can't handle basic tasks like "get reports prepared". When a labor union leader tells the workers to strike, they will probably strike, even if they don't feel like it, because they know that unless they act as a coordinated group they'll never be able to exert any power. So:

EXPLICIT MODEL: Top-down organization is an effective way to coordinate large organizations
EMOTIONAL EXPERIENCE: Respect, deference
REIFIED ESSENCE: "Authority", "legitimacy" (in the sense of

"*this guy* is the rightful king, but *that guy* is a pretender")
ENDORSED VALUE: Respect for authority

Frimer et al (study, popular article) have done some work on this. They find that when you ask people to imagine "Authority", they imagine a police officer, a military commander, or some other stereotypically conservative figure who conservatives respect and liberals do not. Since liberals have little interest in making the police more effective, there's no reason for them to "respect authority" in this case. When researchers give subjects the example of some environmental organization trying to coordinate its environmental activism, liberals are much more likely to say people should respect the authority of the organization leaders.

A more recent study (study, popular article) found similar results, although with similar caveats. They were investigating a construct called cognitive rigidity, and asked questions like "true or false: a group which tolerates too much dissent among its members cannot exist for long?". Conservatives tend to agree with the base question more, but when you specify "an environmental group", liberals agree more. I still think this is kind of stupid and more about liberals' willingness to agree to anything that sounds vaguely pro-environmental. At one point they're investigating the question "A dead hero is better than a live coward", they change it to "When it comes to preventing global warming, a dead hero is better than a live coward", and liberals just go ahead and agree with the statement instead of asking what the f@#k. I consider these studies very questionable and preliminary. But here are some true-or-false questions I offer to the next person to do a study like this:

A: It is dangerous to show too much mercy to people who commit crimes
B: It is dangerous to show too much mercy to people who commit gun violence

A: Barack Obama was the president, and his opponents should have treated him with respect even when they disagreed with his policies
B: Donald Trump is the president, and his opponents should treat him with respect even when they disagree with his policies

A: If I were an employee in a company, I would try to carry out the CEO’s orders even if I disagreed with them, because otherwise we would be disorderly and totally ineffective
B: If I were a member of a labor union, I would try to carry out the union leader’s policies even if I disagreed with them, because otherwise we would be disorderly and totally ineffective

I’m not asserting that liberals and conservatives would answer their respective questions exactly the same. My guess is that even in a value-neutral way, conservatives have these foundations a little more crystallized than liberals, just as they have most other heuristics a little more crystallized than liberals. But I am saying that nobody has done this experiment correctly, and I am suspicious that the groups would be closer than people think.

People can choose metaphysical heuristics or explicit models based on their own innate tendencies, their education, their intelligence, their experiences, and what kind of question we’re thinking about. Rather than talking too much about fundamental value differences, we should be asking where a given person has chosen to place themselves on the metaphysical-heuristic-to-explicit-model ladder at any particular moment.

IV.

This way of looking at things will be valuable if it helps people who crystallize heuristics at different levels understand each other. Here are a couple of common mistakes I think I see:

People who endorse values based on crystallized essences might think that people who use explicit models are weirdly and inexplicably evil, because the essentialists assume the modelers believe in the essences but don’t care about them, or prefer the opposite. If you believe in Essential Purity, then someone who doesn’t might seem like someone who supports Essential Impurity, rather than somebody working off a totally different system.

On the opposite side, if you’re a pure consequentialist, you might see someone who endorses crystallized-essence values as doing something inexplicable and evil. If you think of it as no different from what you do when you like nature, it might be easier to understand.

I have to bring these up because they’re obvious, but really I don’t see either of these that often. The main mistake I see is people on both sides having at least moderately good understanding of how to do explicit causal models, but accusing the other of being Neanderthals who only care about a crystallized-metaphysical-essence and have totally abandoned reason.

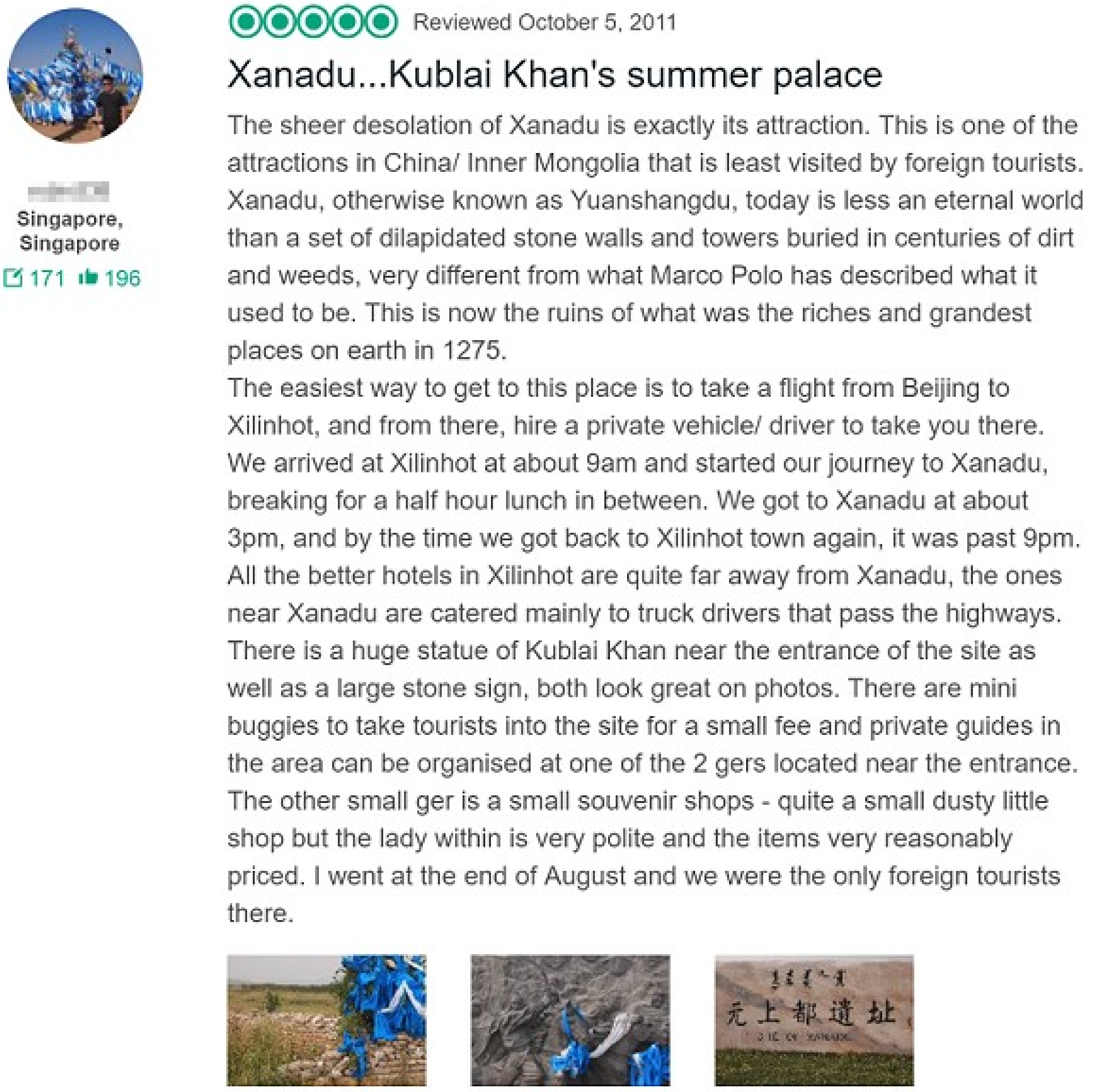
That is, I see communists assuming every single libertarian in the world is a fundamentalist about property rights and thinks they’re so sacrosanct that they must be maintained even in the face of horrible suffering, whereas they (the communists) quite reasonably want what makes a flourishing society full of happy people. Whereas the libertarians say they just want universal wealth and prosperity, whereas communists so bloody-mindedly attached to the metaphysical principle of Equality that they don’t care whether attempts to create it will lead to gulags and total economic collapse.

I see cosmopolitans believing that they want what’s best for society, but that nativists are working off an essentialist racism, where foreigners are inherently inferior in some vague metaphysical way. And the nativists, for their part, are arguing that they’re really concerned about the effects of too much immigration, but the cosmopolitans’ blind adherence to Multiculturalism as good in itself makes them unwilling to debate the real-world consequences of their actions.

Since most metaphysical heuristics are a stand-in for something real, we should expect blocs of allied people to contain some people who want the real thing, and other people who are running metaphysical heuristics that point at the thing. That is, the Tough On Crime bloc will have some members who just want to deter crime more, and other members who believe criminals deserve to suffer because of metaphysical Justice. The Soft On Crime bloc will have some members who question whether people need a ten year prison term for stealing a CD-ROM, and others who believe that prison is torture (metaphysical essence!) and so unconscionable regardless of its deterrent effect. If both sides try to position themselves as the hard-headed practical people, but weak-man the other side as having some incomprehensible metaphysics that makes them impervious to reason, that’s going to effectively shut down the possibility of debate.

Except my actual position is that the same sort of experiences that give you the metaphysical Justice intuition – having personally been a victim of crime, really caring a lot about making it as rare as possible, not being very well-educated – are also likely to make you overestimate the consequentialist value of deterring crime (and vice versa for the other side). My guess is a lot of people fluidly move back and forth between these levels, just as I would expect people who are very interested in only eating organic food to also be more likely to care about what percent RDA of vitamins are in their food. This isn’t sinister, or a reason to think that people are only claiming consequentialist arguments for their heuristics. It’s just a natural consequences of the way our values get produced and the fuzziness in everybody’s value system.

Verses Composed Upon Reading A Review From TripAdvisor



[redacted]
Singapore, Singapore
171 196

Reviewed October 5, 2011

Xanadu...Kublai Khan's summer palace

The sheer desolation of Xanadu is exactly its attraction. This is one of the attractions in China/ Inner Mongolia that is least visited by foreign tourists. Xanadu, otherwise known as Yuanshangdu, today is less an eternal world than a set of dilapidated stone walls and towers buried in centuries of dirt and weeds, very different from what Marco Polo has described what it used to be. This is now the ruins of what was the riches and grandest places on earth in 1275.

The easiest way to get to this place is to take a flight from Beijing to Xilinhot, and from there, hire a private vehicle/ driver to take you there. We arrived at Xilinhot at about 9am and started our journey to Xanadu, breaking for a half hour lunch in between. We got to Xanadu at about 3pm, and by the time we got back to Xilinhot town again, it was past 9pm. All the better hotels in Xilinhot are quite far away from Xanadu, the ones near Xanadu are catered mainly to truck drivers that pass the highways. There is a huge statue of Kublai Khan near the entrance of the site as well as a large stone sign, both look great on photos. There are mini buggies to take tourists into the site for a small fee and private guides in the area can be organised at one of the 2 gers located near the entrance. The other small ger is a small souvenir shops - quite a small dusty little shop but the lady within is very polite and the items very reasonably priced. I went at the end of August and we were the only foreign tourists there.

Source is this page

The Tourist Board of Xanadu
Did recently impose a fee
On those who travel far from home

To visit Kubla’s pleasure dome
Of $20, 9 to 3

So twice five miles of fertile ground
With fence and wire are girdled round
And signs proclaiming “ENTRY AT THE GATE”
Where gather many a camera-bearing crowd
And here are docents, who in solemn state
Explain the Mongol histories aloud

But oh! That deep romantic chasm protracting
Into a hill, athwart a cedarn cover
A savage region, visitors attracting
By actresses, forever reenacting
A woman wailing to her demon-lover

And from this chasm, with ceaseless turmoil spilling
Crowds of old men in fat thick pants are milling
And there, a fountain momently is forced:
Amid whose swift half-intermitted burst
Groups of eight to ten people, screaming ever
White-water-raft upon the sacred river

Five miles continuing to a crashing climax
Through wood and dale the sacred waters run;
I didn’t think this part was too much fun,
So skip the crowds, and head down to the IMAX,
Where in surround-sound, you can hear from far
Ancestral voices prophesying war!

The shadow of the dome of pleasure
Stands reflected in the mere;
Take some photos there to treasure
As a special souvenir
It is a miracle of rare device:
A tourist trap, but also pretty nice.

A damsel with a dulcimer
In a vision once I saw:
It was an Abyssinian maid
And on her dulcimer she played,
Singing of Mount Abora.
Could I revive within me
Her symphony and song,
To such a deep delight ’twould win me,
That with music loud and long,
I would build that dome in air

That sunny dome! those caves of ice!
And all who heard should see them there,
And all should cry, Beware! Beware!
His flashing eyes! His floating hair!
Hide the sight from eyes profane,
And weave a circle round him thrice
For he hath tasted Paradise,
5/5, would taste again.

Meetups Everywhere 2018

Last year we had organized meetups in sixty different cities around the world. A couple of the meetup groups stuck around or reported permanent spikes in membership, which sounds like a success, so let’s do it again. I’ll repeat the city list from last year, which has every city where at least ten people expressed interest in a meetup. A few nearby cities are merged to make sure they have enough people.

If you’re willing to host a meetup for your city, then decide on a place, date, and time, and post it in the comments. I would also prefer you include an email where people can reach you.

Please err in favor of volunteering to organize – the difficulty level is basically “pick a coffee shop you like, tell me the address, and give me a time”; it would be dumb if nobody got to go to meetups because everyone felt too awkward and low-status to volunteer.

In a week or so, I’ll make another post listing the details for each city so people know where to go.

Some details and suggestions for would-be organizers:

1. I don’t guarantee I’ll have the post with times and addresses up until August 8, so please choose a day after that. The weekend of the 11th -12th might be one good choice.

2. In the past, the best venues have been ones that are quiet(ish) and have lots of mobility for people to arrange themselves into circles or subgroups as desired. Private houses have been pretty good. Same with food courts.

Cafes and restaurants have gone okay, as have empty fields (really). Bars don’t seem to have worked very well at all.

3. Usually only about a quarter of people who express interest actually attend. If your city has fewer than 20 people, don’t offer to organize unless you’re okay with a good chance of only one or two other people showing up.

4. If more than one person volunteers to organize, I will pick among them. Priority will be given to people I know well, people who have organized meetups before, and (especially) an existing SSC/LW/EA meetup group in the city. If you run an existing SSC/LW/EA meetup group and you want to organize your city’s SSC meetup, please mention that in the post so I can give you precedence.

5. If you have an existing meetup group, you can just tell me what you’re already doing and when your next meetup is. But try to have the one you list here be some kind of “welcome, SSC people” meetup or otherwise low-barrier-to-entry. And please give me a firm date and time commitment instead of “tell people to check our mailing list to find out where the meeting will be that week”.

6. If you’re formally volunteering to organize a meetup, please respond with an *unambiguous statement* to this effect, the exact address, the exact time, and the date (+ contact details if possible). I’m not going to count someone as offering to organize a meetup unless they do this. Please don’t post “I hope someone agrees to organize a meetup in my city”. Just offer to organize the meetup!

EDIT: Again, *please* include an exact time, exact date, and exact address with your offer to host. Please don’t post vague speculation about how you might want to host at some point – just offer to host and

give me the information I need. If it turns out there’s someone better, don’t worry, they’ll also offer and I’ll choose them.

Here are cities and number of interested people (as of last year):

Ann Arbor: 23
Atlanta: 29
Austin: 43
Baltimore: 23
Berkeley: 70
Berlin: 25
Birmingham (UK): 10
Boston: 144
Brisbane: 12
Calgary: 12
Cambridge (UK): 19
Canberra: 12
Charlotte: 10
Chicago: 100
Cincinnati: 13
Cleveland: 16
Cologne: 13
Columbus: 20
Copenhagen: 13
Dallas: 20
Denver: 34
Detroit: 23
Dublin: 19
Edinburgh: 10
Edmonton: 12
Helsinki: 33
Houston: 21
Kansas City: 14
London: 121

Los Angeles: 74
Madison: 25
Melbourne: 29
Milwaukee: 10
Minneapolis: 29
Montreal: 16
Munich: 18
Nashville: 12
New Haven: 13
New York: 195
Oslo: 11
Ottawa: 16
Oxford: 18
Paris: 20
Philadelphia: 51
Phoenix: 17
Pittsburgh: 28
Portland (OR): 38
Raleigh: 17
Rochester: 12
Sacramento: 13
Salt Lake City: 23
San Diego: 27
San Francisco: 148
San Jose: 68
Sao Paulo: 11
Seattle: 111
St. Louis: 20
Stockholm: 14
Sydney: 37
Tel Aviv: 27
Toronto: 56
Vancouver: 23
Vienna: 15
Warsaw: 10
Washington DC: 110

Wellington: 11
Zurich: 16

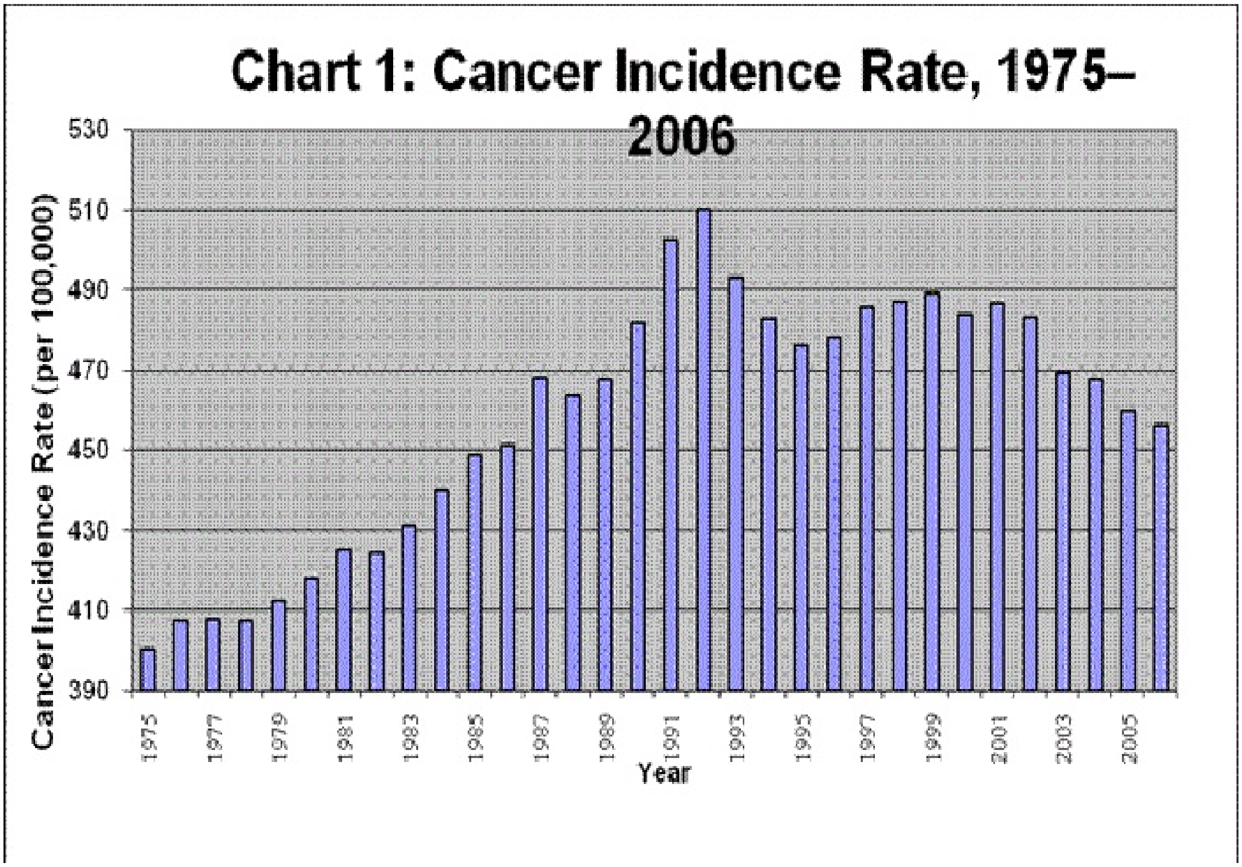
Cancer Progress: Much More Than You Wanted To Know

Official statistics say we are winning the War on Cancer. Cancer incidence rates, mortality rates, and five-year-survival rates have generally been moving in the right direction over the past few decades.

More skeptical people offer an alternate narrative. Cancer incidence and mortality rates are increasing for some cancers. They are decreasing for others, but the credit goes to social factors like smoking cessation and not to medical advances. Survival rates are increasing only because cancers are getting detected earlier. Suppose a certain cancer is untreatable and will kill you in ten years. If it’s always discovered after seven years, five-year-survival-rate will be 0%. If it’s always discovered after two years, five-year-survival-rate will be 100%. Better screening can shift the percent of cases discovered after seven years vs. two years, and so shift the five-year-survival rate, but the same number of people will be dying of cancer as ever.

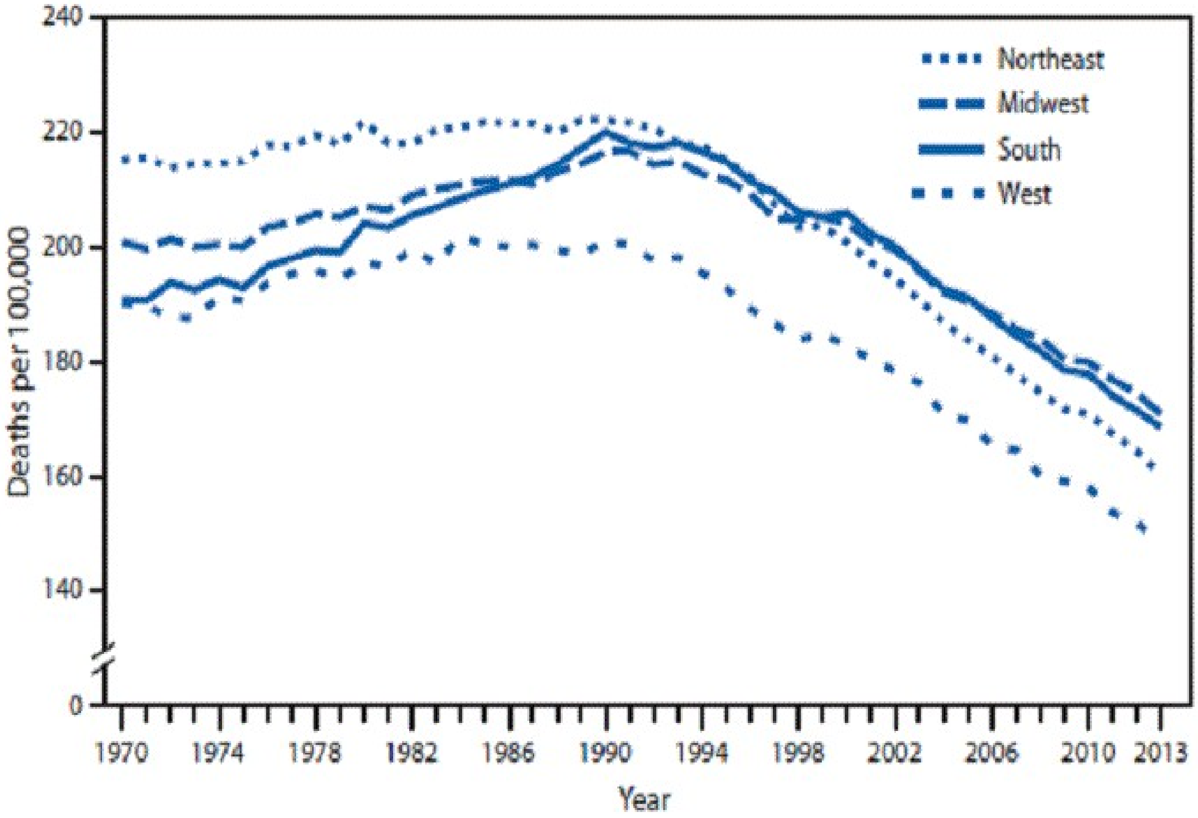
This post tries to figure out which narrative is more accurate.

First, incidence of cancer:

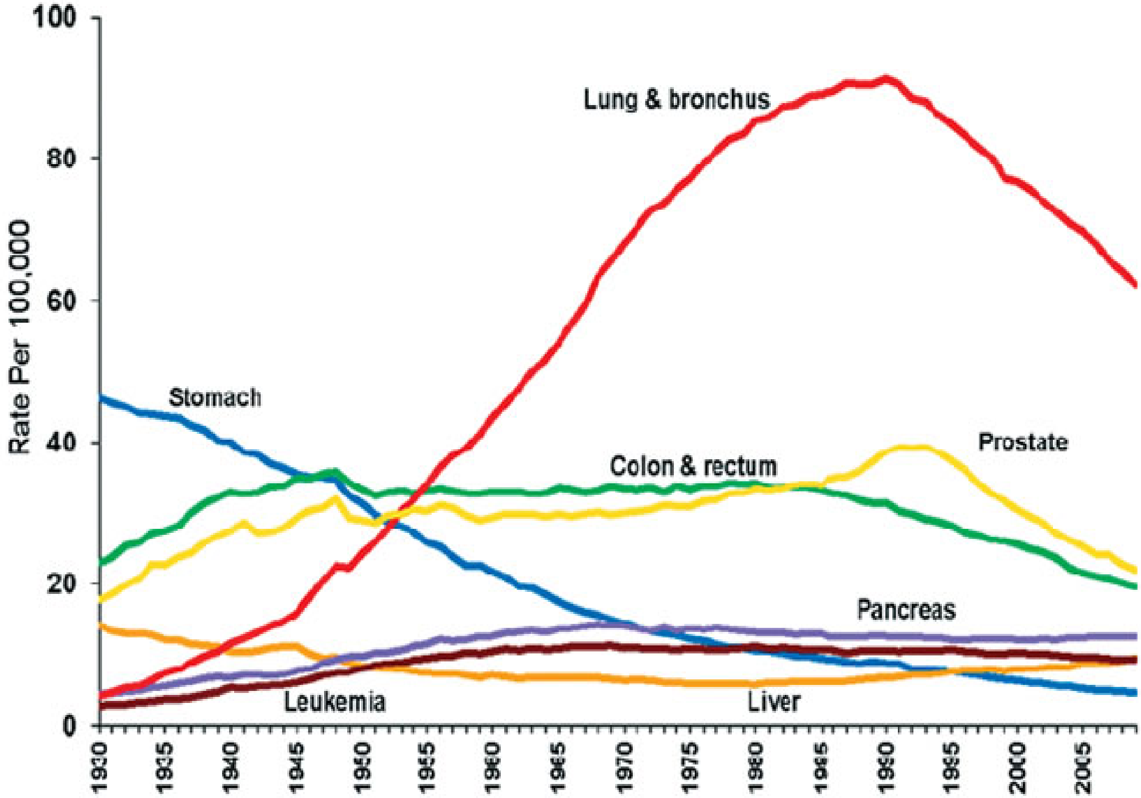


This chart doesn't look good (in both senses of a chart not looking good – seriously, put some pride into your work). Although there's a positive trend since 2001, it's overwhelmed by a general worsening since 1975. But this isn't the right way to look at things: average age has increased since 1975. Since older people are at higher risk of cancer, an older population will look like higher cancer rates. Also, something has to kill you, so if other issues like violent crime or heart disease get better, it will look like a higher cancer rate.

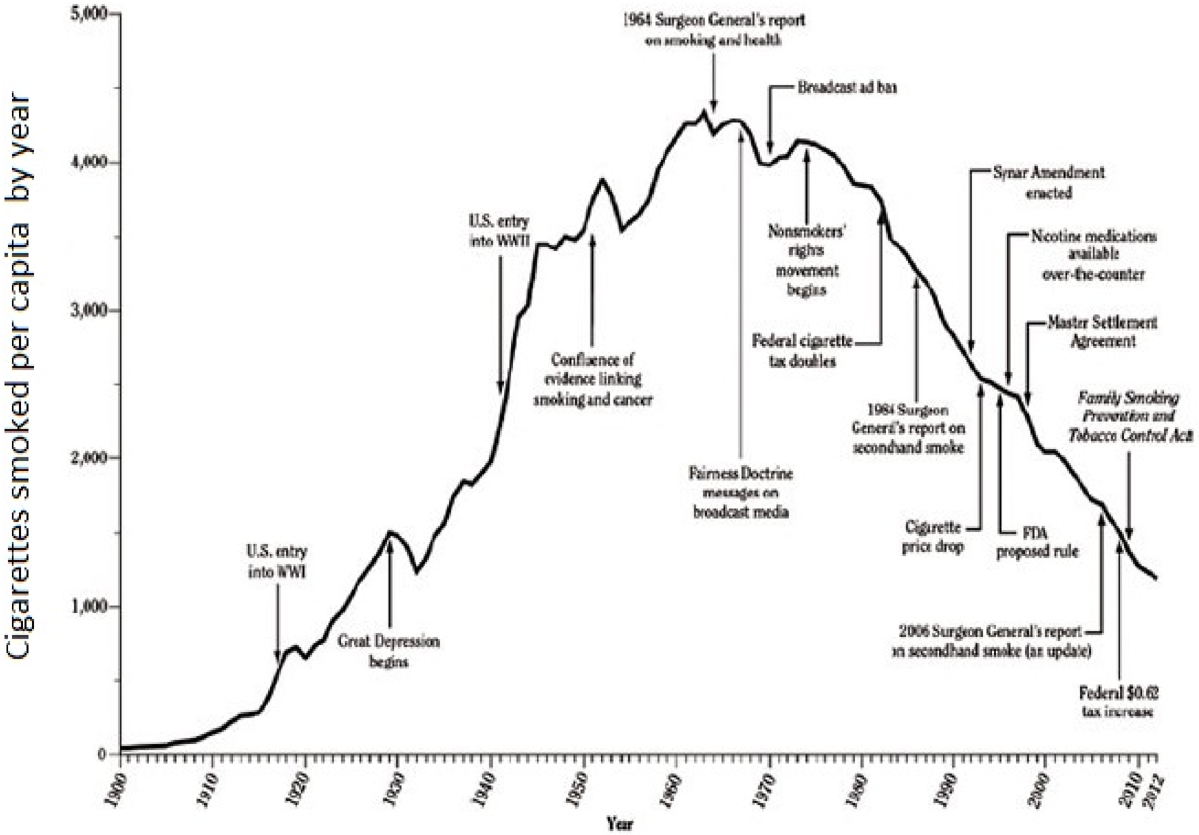
Here's a better graph:



This is adjusted for age. I’ve switched from incidence rates to death rates, which is bad, but I can’t find good age-adjusted incidence data. Also, notice that this graph truncates its y-axis differently than the other. Still, it shows a similar pattern of adjusted death rates getting worse until 1990 and better thereafter. Why?



Smoking! That graph is just this one plus a 20-to-30-year delay:



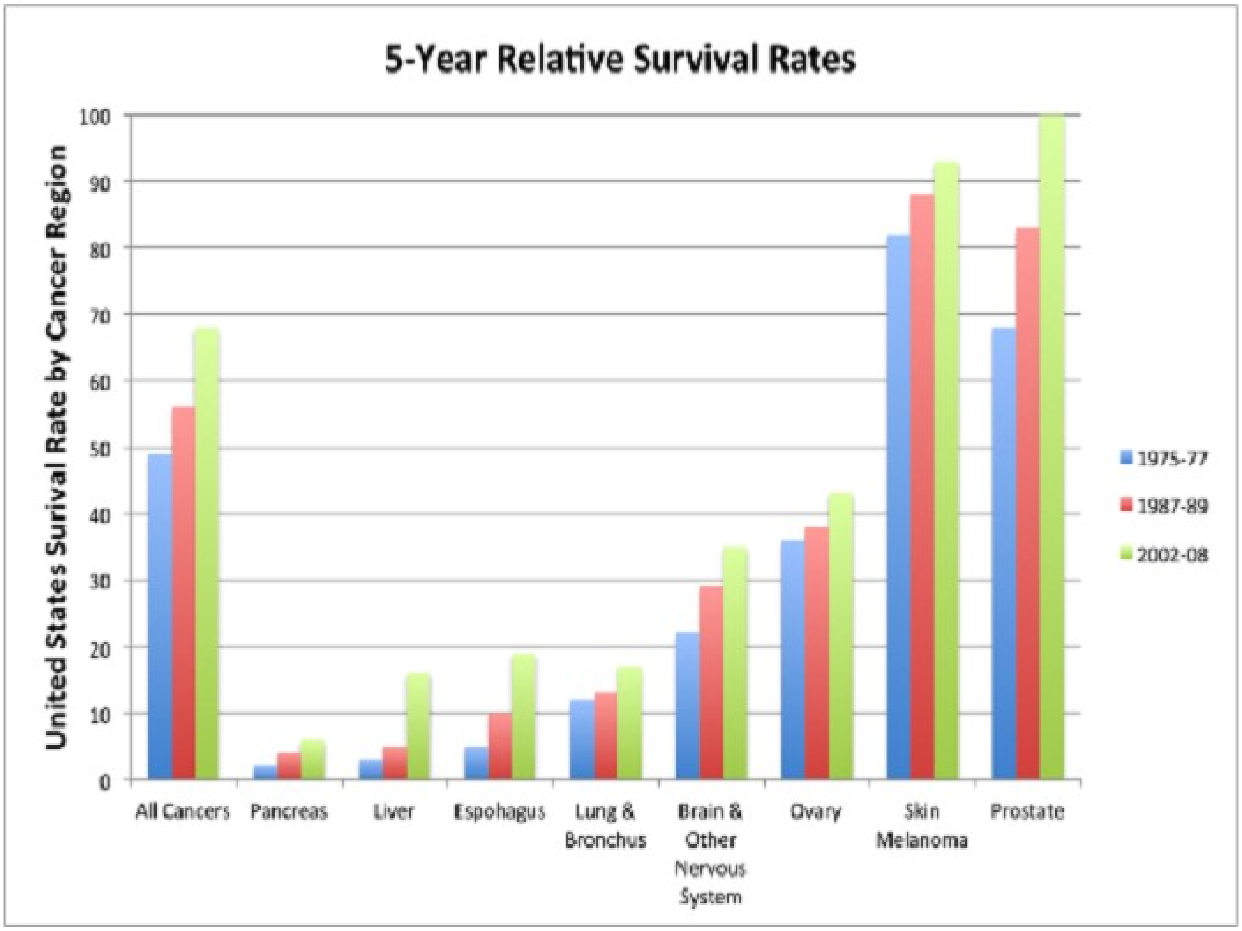
Through the first half of the twentieth century, improved tobacco-making technology, increased wealth, and better advertising caused order-of-magnitude increases in smoking. It takes on average a few decades for smoking to cause lung cancer, so there’s a peak in cancer (overwhelmingly driven by lung cancer) with a few-decade delay from the smoking graph. As smoking started to decline, so did lung cancer.

What about the other striking increase on the incidence graph, that of prostate cancer? In the late 1980s, guideline-making bodies suggested that doctors test harder for prostate cancers; doctors followed the recommendation, detected every little tiny irrelevant prostate tumor, and treated patients aggressively for cancers that never would have affected them before they died of something else. In the late 1990s, guideline-making bodies admitted this had been a bad idea, made the opposite recommendation, and people stopped diagnosing prostate cancer as often. If you look at incidence rates, that spike is much bigger. I'm not sure why this shows up on death rates, but perhaps the treatment itself contributed to mortality, or perhaps coroners were biased to attribute a death to prostate cancer if they knew the cancer was present.

Meanwhile, stomach cancer has declined dramatically; different sources attribute this to improved treatment for the cancer-causing stomach bacterium h. pylori, improved food processing methods, and increased vitamin C. Colon cancer is decreasing because colonoscopies remove more pre-cancerous polyps. Liver cancer increased because of a hepatitis C epidemic. A few other cancers are increasing or declining for similarly diverse reasons.

But overall cancer incidence and death rates increased up to 1990 and have declined thereafter. Pretty much everyone attributes the bulk of the decreasing death rate to improved prevention. If improved cancer treatment is contributing, it's swamped by the social factors and we can't see it in these data.

The most common method for measuring the effect of improved cancer treatment is the five-year survival rate – what percent of people survive five years after being diagnosed with cancer? Here are the relevant data (source):



This is the best graph I can find, but it unfortunately leaves out breast cancer, colon cancer, and several other major cancers where we've made important advances. It's from 2008, but the trends shown have continued since then. Note that change in the "All Cancers" category also reflects changing distribution of sites.

That looks like progress. But this is where the early diagnosis concerns come in. They're best expressed by Welch, Schwartz, and Woloshin, who find that among different types of cancer, secular decreases in five-year-survival-rate are not correlated at all with improvement in the cancer death rate, but they are very correlated with change in the incidence rate. In other words, why are people living longer after being diagnosed with cancer? It can't be because we're treating the cancer successfully – if it were, they would be linked to decreases in the number of people dying of cancer. But it must be because we’re detecting more cases of small cancers too minor or slow-growing to kill people quickly (“lead-time bias” and “length bias”), which shows up as increases in the cancer detection rate.

This study does not prove that cancer treatment is not improving. It just shows that five-year-survival-rates do not in and of themselves provide evidence for improving cancer treatment. Any signal from improving cancer treatment is drowned out by the signal from improved detection.

How do we get around this? One possibility is to investigate change in stage-specific survival rates. That is, doctors classify cancers by stage, all the way from very early poorly-developed cancers with good prognosis to very advanced cancers with bad prognosis. A lead-time bias or length bias would show up as cancers being detected at an earlier stage. So if we found that more people were surviving even within each bin of “stage at which the cancer was detected”, this would be strong evidence that cancer treatment really is getting better.

Several groups have looked into this. The best data comes from the government’s national cancer statistics clearinghouse at SEER (source):

Supplementary Table 3. Changes in 5-year relative survival (%) for the most common cancers by stage, all ages, SEER 9*, 1975-2012

Cancer site	**Localized**				**Regional**			
	Survival (95% CI)		**Difference (95% CI)**		**Survival (95% CI)**		**Difference (95% CI)**	
	1975-1977	**2006-2012**	**Absolute (%)**	**Proportional (%)**	**1975-1977**	**2006-2012**	**Absolute (%)**	**Proportional (%)**
Lung and bronchus	47.3 (45.8, 48.8) For 1988-1990	56.2 (55.1, 57.3)	8.8 (7.0, 10.7)	18.6 (14.2, 23.1)	16.3 (15.5, 17.2) For 1988-1990	28.6 (27.8, 29.4)	12.3 (11.2, 13.5)	75.5 (65.3, 85.6)
Colon and rectum	82.1 (80.8, 83.4)	90.5 (89.8, 91.1)	8.3 (6.9, 9.8)	10.2 (8.3, 12.0)	52.3 (51.1, 53.5)	72.9 (72.0, 73.7)	20.5 (19.0, 22.0)	39.2 (35.6, 42.9)
Breast (Female)	90.5 (89.7, 91.2)	99.2 (98.9, 99.5)	8.7 (7.9, 9.6)	9.7 (8.7, 10.7)	67.4 (66.4, 68.5)	87.7 (87.1, 88.2)	20.2 (19.1, 21.4)	30.0 (27.8, 32.2)
Prostate	100.0 (100.0, 100.0) For 1995-1997; localized and regional combined	100.0 (100.0, 100.0) localized and regional combined	0.0 (-0.1, 0.1)	0.0 (-0.1, 0.1)	---	---	---	---
Oral cavity and pharynx	76.3 (74.0, 78.4)	87.0 (85.5, 88.5)	10.7 (8.0, 13.4)	14.1 (10.2, 17.9)	40.3 (38.0, 42.5)	64.4 (63.0, 65.7)	24.1 (21.5, 26.7)	59.9 (50.5, 69.3)
Esophagus	8.9 (6.5, 11.7)	47.2 (43.9, 50.5)	38.4 (34.1, 42.6)	433.0 (269.7, 596.3)	3.4 (1.9, 5.7)	25.5 (23.3, 27.9)	22.1 (19.1, 25.1)	644.1 (224.1, 1064)
Stomach	53.1 (48.8, 57.2)	65.3 (62.9, 67.5)	12.2 (7.4, 17.0)	23.0 (12.4, 33.6)	14.1 (12.4, 15.9)	32.5 (30.3, 34.7)	18.4 (15.6, 21.2)	131.1 (98.1, 164.0)
Pancreas	7.0 (4.9, 9.5)	32.1 (29.5, 34.8)	25.1 (21.7, 28.6)	359.2 (204.8, 513.7)	2.7 (1.8, 4.0)	11.7 (10.6, 12.7)	8.9 (7.4, 10.5)	327.9 (150.7, 505.1)
Liver and intrahepatic bile duct	8.5 (5.3, 12.6)	31.9 (30.4, 33.4)	23.4 (19.4, 27.4)	275.6 (112.4, 438.7)	0.4 (0.0, 2.3)	11.0 (9.8, 12.2)	10.5 (9.0, 12.0)	2337 (-2438, 7112)
Urinary Bladder	85.6 (84.0, 87.0)	92.0 (91.1, 92.8)	6.4 (4.7, 8.1)	7.5 (5.4, 9.6)	40.5 (37.7, 43.3)	43.1 (41.4, 44.9)	2.6 (-0.7, 6.0)	6.5 (-2.2, 15.1)
Kidney and Renal Pelvis	81.3 (78.4, 83.8)	93.5 (92.7, 94.2)	12.2 (9.4, 15.0)	15.0 (11.1, 18.9)	48.9 (45.0, 52.6)	67.4 (65.3, 69.4)	18.5 (14.2, 22.9)	37.9 (26.3, 49.4)
Melanoma of the skin	91.6 (90.1, 92.9)	99.1 (98.5, 99.5)	7.5 (6.1, 9.0)	8.2 (6.5, 9.9)	57.7 (52.3, 62.7)	65.7 (63.2, 68.0)	8.0 (2.2, 13.7)	13.8 (2.7, 24.9)
Cervix uteri	87.7 (85.9, 89.2)	91.8 (90.3, 93.0)	4.1 (2.0, 6.2)	4.7 (2.2, 7.2)	47.6 (44.2, 50.9)	57.6 (55.0, 60.1)	10.0 (5.8, 14.2)	21.1 (11.1, 31.1)
Corpus and uterus, NOS	94.3 (93.4, 95.0)	96.2 (95.5, 96.7)	1.9 (0.9, 2.9)	2.0 (0.9, 3.1)	63.5 (59.2, 67.4)	71.6 (69.8, 73.3)	8.1 (3.7, 12.6)	12.8 (5.1, 20.6)
Ovary	83.1 (80.2, 85.6)	91.2 (89.3, 92.7)	8.1 (4.9, 11.3)	9.7 (5.6, 13.9)	49.1 (43.2, 54.8)	65.3 (61.5, 68.8)	16.2 (9.3, 23.0)	32.9 (15.6, 50.3)
Non-Hodgkin lymphoma	68.4 (65.6, 71.0) For 1983-1985; Ann Arbor stage I	82.1 (80.9, 83.3) Ann Arbor stage I	13.7 (10.8, 16.7)	20.1 (15.0, 25.1)	56.6 (53.4, 59.7) For 1983-1985; Ann Arbor stage II	76.4 (74.7, 78.0) Ann Arbor stage II	19.8 (16.3, 23.4)	35.0 (26.9, 43.1)
Thyroid	98.8 (97.8, 99.4)	99.8 (99.7, 99.9)	1.0 (0.3, 1.8)	1.0 (0.2, 1.8)	88.6 (85.6, 91.0)	97.9 (97.2, 98.5)	9.4 (6.6, 12.1)	10.6 (7.2, 14.0)

	Distant			
	Survival (95% CI)		**Difference (95% CI)**	
	1975-1977	**2006-2012**	**Absolute (%)**	**Proportional (%)**
Lung and bronchus	1.9 (1.7, 2.1) For 1988-1990	4.2 (3.9, 4.4)	2.3 (2.0, 2.6)	122.3 (93.6, 151.1)
Colon and rectum	5.5 (4.9, 6.2)	13.7 (13.0, 14.4)	8.2 (7.2, 9.1)	148.4 (116.3, 180.5)
Breast (Female)	18.7 (16.9, 20.6)	33.6 (32.2, 35.0)	14.9 (12.6, 17.2)	79.7 (60.3, 99.2)
Prostate	31.8 (29.7, 34.0) For 1995-1997	28.7 (27.1, 30.4)	-3.1 (-5.8, -0.3)	-9.7 (-17.8, -1.6)
Oral cavity and pharynx	18.5 (15.6, 21.6)	39.8 (37.4, 42.1)	21.2 (17.4, 25.0)	114.6 (77.8, 151.5)
Esophagus	1.4 (0.6, 2.9)	4.8 (3.9, 5.8)	3.4 (1.9, 4.8)	238.1 (-39.4, 515.6)
Stomach	1.6 (1.1, 2.3)	6.5 (5.6, 7.5)	4.9 (3.8, 6.0)	305.5 (137.1, 474.0)
Pancreas	0.9 (0.6, 1.4)	3.3 (2.9, 3.7)	2.3 (1.7, 2.9)	245.1 (97.4, 392.9)
Liver and intrahepatic bile duct	1.1 (0.3, 2.9)	2.3 (1.6, 3.2)	1.2 (-0.2, 2.7)	116.0 (-138, 369.7)
Urinary Bladder	9.0 (5.7, 13.3)	5.1 (3.7, 6.8)	-3.9 (-8.0, 0.2)	-43.6 (-72.8, -14.3)
Kidney and Renal Pelvis	6.9 (5.3, 8.8)	12.0 (10.7, 13.3)	5.0 (2.8, 7.2)	72.1 (24.9, 119.4)
Melanoma of the skin	14.8 (10.0, 20.5)	20.0 (17.4, 22.9)	5.3 (-0.7, 11.2)	35.7 (-16.2, 87.7)
Cervix uteri	14.5 (10.5, 19.2)	16.6 (13.5, 20.0)	2.1 (-3.4, 7.5)	14.2 (-26.8, 55.3)
Corpus and uterus, NOS	21.9 (18.2, 25.8)	27.2 (25.1, 29.4)	5.3 (0.9, 9.6)	24.2 (0.6, 47.8)
Ovary	15.0 (13.6, 16.5)	31.5 (30.2, 32.9)	16.5 (14.5, 18.5)	110.0 (87.6, 132.4)
Non-Hodgkin lymphoma	40.0 (38.3, 41.7) For 1983-1985; Ann Arbor stage III and IV	65.5 (64.5, 66.4) Ann Arbor stage III and IV	25.5 (23.5, 27.5)	63.8 (56.4, 71.3)
Thyroid	56.4 (48.4, 63.6)	58.6 (54.4, 62.6)	2.2 (-6.5, 10.9)	3.9 (-11.9, 19.7)

Abbreviations: SEER = Surveillance, Epidemiology, and End Results; CI = Confidence Interval; NOS = not otherwise specified.

* The SEER 9 registries are Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah.

Even within each stage, five-year-survival-rate has increased significantly from 1975 to 2012.

Closer investigations of specific cancers are similar. Stage-adjusted cervical cancer risk and colon cancer risk both show most of the modern gains in survival rate persisting.

But maybe stages are too big a bin to serve as a useful proxy. Imagine a study that wanted to prove that having more cars made you happier. They do a survey and find that people with more cars are happier, but someone objects that maybe wealthy people have more cars and wealth makes you happier. Imagine that their response is to separate people into two bins: “poor people” who make below $50K and “rich people” who make more. They find that even within each bin, cars still make you happier. But this is just a problem of too few bins: a person making $10K is still very different from a person making $40K (and likely to have fewer cars). The attempt to remove confounding with bins fails. These cancer studies generally use only a few broad stages; might this be allowing effects from early diagnosis to creep back in?

Elkin, Hudis, Begg & Schrag look into this. They find that within each stage, tumors have gotten smaller since 1975, suggesting that the staging system isn’t capturing everything we care about regarding cancer. But they find that even when adjusted for size, some of the stage-specific modern gains in cancer survival still remain. In particular, decreasing size explains 61% of improved survival in localized-stage breast cancer, and 28% of improved survival in regional-stage breast cancer. Another study on breast cancer does a similar adjustment with other ways of classifying cancer and concludes that “improvements were shown irrespective of tumor size, lymph node status, and ER status” and “the impact of screening was by nature of limited magnitude. The modified treatment strategies implemented by the use of nationwide guidelines seemed to have a major impact on the substantial survival improvements.” Another group does a simulation and finds that it’s implausible that screening-related biases are the entire source of improved survival:

> The results from our study suggest that lead-time bias introduced by mammography screening does not explain the survival improvement observed during the recent decades in the Nordic countries. The absolute as well as relative bias was generally small, and much smaller than the observed increase in relative survival between 1964-2003. However, in some settings the absolute bias reached 4.0-5.7 percentage points, on a survival around 68-77%, a difference that many would see as an interesting improvement in survival.

A lot of this work has been done in breast cancer, probably because it’s had a strong push for screening recently. We would expect screening to be even less important in other cancers, but there hasn’t been as much work on it. One exception is Tong et al, who find that changes in tumor stage and size explain only 20% of improved survival rates in colon cancer, but advancements in therapy explain about 71%. Separately, an authoritative-sounding collection of colon cancer experts express their opinion that “it is possible that within-stage migration had some effect on our findings, but it is implausible as the major source of the trends we observe.”

The only contrary data point I can find is this study of laryngeal cancer, which finds *worsening* stage-specific survival rates for high-stage laryngeal cancer since 1977. However, the study authors note this was the only one of 24 cancer types examined to show decreasing survival rates. They speculate that maybe some kind of change in smoking behavior over this period has changed the nature of laryngeal carcinomas to favor a more aggressive type. They don't really have any evidence for this, but given that this is the only one of 24 cancer types to show a decrease in survival rate, it's probably something at least that unique, and doesn't indicate a general failure in cancer treatment.

There could still be unobserved confounders. Stage alone wasn't enough, but merely adding size to stage might still not be enough. Even the papers that look at a few more esoteric things like receptor status might not be enough. All we can say with certainty is that right now, adjusting for everything we know about and are able to monitor, cancer survival rates still seem to have increased. Tomorrow we might discover new confounders that take that away from us, but right now there is no particular reason to expect that we should.

So: age-adjusted cancer incidence rates and death rates have been going down since 1990, primarily due to better social policies like discouraging smoking. Five-year-survival rates have been gradually improving since at least 1970, on average by maybe about 10% though this depends on severity. Although some of this is confounded by improved screening, this is unlikely to explain more than about 20-50% of the effect. The remainder is probably a real improvement in treatment. Whether or not this level of gradual improvement is enough to represent "winning" the War on Cancer, it at least demonstrates a non-zero amount of progress.

I don't want to frame this in terms of "here we DEMOLISH the pseudoscientific narrative that cancer progress is weak". Many of the people I know who critique this research are from an older generation. They remember Nixon assuring them at the very beginning of the War on Cancer that we would have a cure within five years. If they’re really old, maybe they remember victories of that scale over polio and smallpox. If those were their hopes, it’s right for them to feel disappointed. But I come from a generation that doesn’t expect much, and I think the evidence suggests my low expectations have more or less been met.

Before You Get Too Excited About That Trigger Warning Study…

STUDY: Trigger Warnings Are Harmful To College Students says the Daily Wire, describing a study whose participants' average age was 37 and which did not measure harm.

You can find the study involved here. A group of Harvard scientists asked 370 people on Mechanical Turk to read some disturbing passages – for example, a graphic murder scene from *Crime and Punishment*. Half the participants received the following trigger warning before the passage:

> TRIGGER WARNING: The passage you are about to read contains disturbing content and may trigger an anxiety response, especially in those who have a history of trauma

Participants were asked to rate their anxiety before and after reading the passages. After they had finished, they were asked to fill out a bunch of questionnaires that measured their opinions about how trauma worked.

The researchers found that people who received the trigger warning were 5% more likely to endorse the idea that they were vulnerable to trauma, and also 5% more likely to endorse the belief that people with trauma could suffer persistent negative effects from that trauma. There were some subgroup and moderation analyses which I ignore for the usual reasons.

What might be some causes for concern with this study?

First, Stuart Ritchie points out that the results are statistically weak. Most of the results have p-values around

0.05, and are not corrected for multiple testing. That means it hasn’t been formally proven whether or not the results are random chance. I don’t like haggling over whether something is just above or just below a significance threshold. But if you do like that kind of haggling, this study doesn’t survive it very well.

Second, the participants’ average age is 37. There’s nothing wrong with 37 year olds, but everyone who says that they’ve proven trigger warnings don’t work on college students either hasn’t read the study, or is adding their editorial opinion.

Third, trigger warnings are aimed at people with trauma. The authors excluded people with trauma from the study. The study was not an attempt to test whether trigger warnings could ever be beneficial, nor to quantify the net (benefit minus harm). It just tested whether there might be nonzero harm. Even if trigger warnings harm healthy people, they might still help the people with trauma who they are intended to help.

But a much more fundamental concern is that *this isn’t measuring anything we care about*.

Some people read a trigger warning saying disturbing passages could cause emotional harm. Then they read a disturbing passage. Then, on a test, they were slightly more likely to agree with the statement that disturbing passages could cause emotional harm. Of note, they did not claim that they themselves had been harmed or triggered by the passage. In fact they specifically denied this; there was no difference in anxiety between the two groups after reading the passage. They just agreed, in a theoretical sense, that trauma was harmful.

The most boring possible explanation is that if you gave someone a passage saying “TRIGGER WARNING: SOME SNAKES BEAR LIVE YOUNG”, then made them read a book about torture-murder, then asked them whether some snakes could bear live young, they would be more likely to answer yes. Reading the statement about snakes doesn’t *cause* the snakes to bear young. It just convinces the reader of it.

In the same way, the trigger warning tells respondents that the passage can cause anxiety, and links this to a history of trauma. It doesn’t say outright that passage –> anxiety –> permanent trauma, but it kind of implies it. It certainly primes any pre-existing beliefs along those lines that participants might have heard. And it potentially induces demand effects – which Mechanical Turk is infamous for – by letting you know what the experimenters are thinking.

For this to result in actual harm, two extra things would have to happen.

First, an effect that happens a split second after you read a trigger warning stating something would have to stick around to become a permanent part of the psyche. This is definitely not how these things work. For example, in 2016 I made a group of people read either an essay warning of the dangers of artificial intelligence, or a control essay that didn’t mention the topic. Then I surveyed people on how dangerous they thought AI was. Unsurprisingly, the people who read the control essay were less worried. The effect stuck around after a month on another survey that explicitly reminded them of the essay they’d read, but when I covertly surveyed them again two years later, there was no difference in their beliefs. They’d forgotten the whole thing. If you think about it, both the experimental group and the control group in this new experiment must have seen lots of trigger warnings throughout their lives. The effects of all of them paled in comparison to having seen a trigger warning a minute earlier in the same experiment. Hours or days later, the effect of this one will fade away too.

(This is the kind of measurement I've condemned as streetlight psychology, after the story of the drunk who searched for his keys under a streetlight – not because he'd lost them there but because that was where it was easy to look. It's been used before to study eg video game violence. If you make someone play a violent video game and then test them for violentness a few minutes later while the adrenaline is still in their system, they'll be more violent. But this is little different from showing that people are sad minutes after they watch a sad movie. If you're asking whether years of playing a video game will cause violent criminality (or years of watching sad movies will cause clinical depression), that's a very different question. The trigger warning study has the same problem.)

But even if reading a trigger warning has lasting effects on your beliefs about trauma, we still have to prove that those beliefs cause you to be more easily traumatized. One of the effects in the study *was* people saying they believed they were more easily traumatized – but how much do we trust that belief? If I made people read a passage saying that cancer was very common and even healthy people often got it, I could certainly make them express a belief that they were more worried about and susceptible to cancer. But would that belief cause cancer? I realize some psychological constructs do work like this, where believing that you're vulnerable makes you more vulnerable. But a lot of psychological constructs also *don't* work like this; it can't be voodoo all the way down.

So, in conclusion: this study tests an accommodation meant to prevent trauma in college students on a population of non-traumatized 30 and 40-somethings. It finds it does not increase anxiety, but may change beliefs about trauma in the very short term, although those changes do not meet strict standards of statistical significance. It is currently unclear whether those changed beliefs last longer than a few minutes, or whether they have any effect on anything.

I think about this Steven Kaas tweet a lot:

> Why idly theorize when you can JUST CHECK and find out the ACTUAL ANSWER to a superficially similar-sounding question SCIENTIFICALLY?
>
> — Steven Kaas (@stevenkaas) December 19, 2011

A lot of people care a lot about whether trigger warnings are helpful or harmful in college courses. This study answers (or "answers") the superficially similar-sounding question of whether they affect certain beliefs about trauma in the very short term. I won't say it has *literally* zero bearing on the original question, but if you treat it as having any specific positive amount of bearing, you're probably wrong.

I don't want to blame the researchers, who are experts in the psychology of trauma and doing important work. This isn't an awful study from the perspective of psychologists trying to start a research program that might be suggestive of something after a decade or two. It *is* an awful study in the context of anything that anybody might report upon or use to form an opinion. Blaming the reporting would be better, but the right-wing sources that exaggerated this are still light-years behind the left in their ability to falsify and weaponize study results and I would feel bad saying anything that seemed to single them out as particularly culpable. Let’s just agree everything is terrible all the time.

As for trigger warnings themselves, I’ll repeat my own proposal: have them, but put them in the Boring Legalese Page of the book, the one where they list the ISBN number and the city where the publishing company has its headquarters and something something Library Of Congress. Make them matter-of-fact, like “Content: rape, murder, ethnic slurs”. You don’t need to embellish with “AND THEREFORE YOU SHOULD BE ANXIOUS” or “SO PEOPLE WITH A TRAUMA HISTORY SHOULD BEWARE” the way this experiment’s warning did. Just put a list of things people might want to know about on the Boring Legalese Page page, and let the couple percent of people with a trauma history check it before reading if they want. If it’s a worksheet or syllabus, put it in small print somewhere consistent. If your opposition to (or support for) trigger warnings is any more impassioned than that, I am suspicious of your motives and think maybe you should sit this one out.

SSC Meetups 2018: Times And Places

Thanks to everyone who offered to host a meetup. We're scheduled for meetups in 85 cities (and one ship!) in 25 countries, soundly beating last year's list. Full list of cities, times, and places is below.

Most people who are on the fence have said they've enjoyed going. Most people who felt intimidated about going have said they've enjoyed going. Most people who felt they were too different from the median SSC reader to fit in have enjoyed going. Most people who worried they weren't smart enough to fit in have enjoyed going. Etc. Some tips from past experience with these meetups:

1. If you're the host, bring a sign that says "SSC MEETUP" and prop it up somewhere on a table.
2. Bring blank labels and pens for nametags.
3. Collect everyone's name and email address in a spreadsheet (template), so you can start a mailing list to make organizing future meetups easier.
4. If it's the first meetup, people are probably just going to want to talk, and if you try to organize some kind of "fun" "event" it'll probably just be annoying.
5. Some things that have worked for later meetups include people giving short presentations on topics of interest to them, or discussion of some particular blog post.
6. Nothing is going to get done unless there's a Schelling point for who has to do it, and right now that's the meetup organizer.
7. It's much easier to schedule a second meetup while you're having the first compared to trying to do it later on by email.
8. Surprisingly many people will love you forever if you bring stim toys.

9. In case people want to get to know each other better outside the meetup, you might want to mention reciprocity.io, the rationalist friend-finder/dating site. It runs off Facebook, so you have to Facebook friend the other person first.
10. If you have a vague location like “in the mall” or “at the North Park”, nobody will ever find each other. Give a specific place (eg “at the North Park, by the big oak tree in the northwest corner”) and be carrying a sign saying “SSC MEETUP”. If you were too vague in your description, comment with a better one and I can edit it in.

Please look over your meetup to make sure it’s correct. If you have corrections, further details, or a new meetup to add, please post a top-level comment below containing the text “@SCOTT” so I can find it easily.

LIST OF CITIES

(all times are local, and PM if not indicated otherwise. You can also search geographically using the map on the LW/SSC meetup page)

Amsterdam, Netherlands
Time: Saturday, August 11, 3:00
Location: On the grass in front of Weesperzijde 100
Contact: mathijs[dot]henquet[at]gmail[dot]com, Meetup page, Facebook group

Ann Arbor, Michigan
Time: Saturday, August 11, 7:00
Location: All Hands Active, 225 East Liberty
Contact: 7343520125

Atlanta, Georgia
Time: Saturday, September 29, 2:00
Location: Hodgepodge Coffee House, 720 Moreland AVE SE,

Atlanta GA 30316
Contact: steve(at)digitaltoolfactory.net

Austin, Texas
Time: Saturday, September 1, 1:30
Location: Central Market, 4001 N Lamar
Contact: ???, but see Google group

Bangalore, India
Time: Sunday, August 26, 4:00
Location: Matteo Cafe, Church Street
Contact: m.nihalmohan[at]gmail.com

Berkeley, California (Bay Area central meetup: Oakland, San Jose, Silicon Valley, etc)
Time: Saturday, September 8th, 3:00
Location: Empty field at the intersection of West Circle and Free Speech Bikeway on UC Berkeley campus
Contact: scott[at]shireroth.org, Facebook event

Berlin, Germany
Time: Saturday August 18, 5:00
Location: St Bart, Graefestraße 71, 10967 Berlin
Contact: buddabrotler[at]gmail.com, Facebook event

Birmingham, UK
Time: Saturday, August 25, 3:00
Location: Caffe Nero, 8-9 Lower Temple St, Birmingham B2 4JD
Contact: askew.thomas[at]gmail.com, or see Facebook event

Boston, Massachusetts
Time: Saturday, August 11, 3:00
Location: Starbucks, 66 Beacon Street
Contact: ardeibel[at]gmail.com

Boulder, Colorado
Time: Saturday, August 11, 3:00
Location: Private home, 3143 Eastwood Ct
Contact: Call corticalcircuitry at 14159883608, Meetup page

Brisbane, Australia
Time: Friday, August 31, 7:00
Location: The Ovolo Inchcolm bar, 73 Wickham Terrace, Spring Hill QLD 4000
Contact: jarred.filmer[at]gmail.com

Bristol, UK
Time: Sunday, September 2, 3:00
Location: Pinkmans Bakery, 85 Park Street, BS1 5PJ Bristol
Contact: whatagoodemailaddress[at]gmail.com, or see Facebook group

Brussels, Belgium
Time: Saturday, September 1, 2:00
Location: North part of the Parc de Bruxelles
Contact: arthur@milchior.fr

Buffalo, New York
Time: Sunday, August 26, 11:00 AM
Location: Five Points Bakery, 44 Brayton St
Contact: spetey[at]gmail.com

Cambridge, UK
Time: Friday, August 17, 6:00
Location: The Burleigh Arms
Contact: aeluncrombie44[at]gmail.com, or see Facebook event

Canberra, Australia
Time: Friday, August 24, 7:00
Location: 9/8 Walsh Place, Curtin

Contact: arunbharatula[at]outlook.com, or call +61 432356733

Charleston, South Carolina
Time: Sunday, August 12, 3:00
Location: Marion Square
Contact: reilly dot steven at gmail dot com

Chengdu, China
Time: Sunday, August 12, 3:00
Location: The Beer Nest 2, near Tongzilin Metro Station
Contact: campbellnilsen on WeChat

Chennai, India
Time: Sunday, August 26, 11 AM
Location: Pumpkin Tales, Alwarpet (vegan friendly!)
Contact: fufstsahil at gmail dot com

Chicago, Illinois
Time: Saturday, September 8, 12 noon
Location: Art Institute South Garden, by the fountain
Contact: rkrzyzanowski[at]gmail.com, or call 312 618 0316, or Facebook event

Cleveland, Ohio
Time: Saturday, August 18, 1:00
Location: Pour Cleveland Cafe, 530 Euclid Ave
Contact: ferbfreeman[at]gmail.com

Cologne / Köln, Germany
Time: Saturday, August 11, 5:00
Location: Marienweg 43 50858 Köln
Contact: marcel_mueller[at]mail.de

Columbus, Ohio
Time: Saturday, September 8, 3:00
Location: Private residence, 397 Eldridge Ave, Columbus,

OH, 43203
Contact: james.thomas.hays[at]gmail.com

Copenhagen, Denmark
Time: Friday, September 28, 5:00
Location: Churchwing of Studenterhuset, Købmagergade 52, 1150 Copenhagen
Contact: mathias@bonde.dk, or call +45 93 93 42 43, or see Facebook event

Denver, Colorado
Time: Tuesday, September 4, 7:00
Location: Apartment common room, 5151 E Yale Cir, Denver, CO 80222-6934
Contact: embrodski[at]gmail.com

Detroit, Michigan
Time: Saturday, August 18, 6:00
Location: Private house. For address, please email contact.
Contact: matt.mattarn[at]gmail.com

Dublin, Ireland
Time: Tuesday, August 14, 6:00
Location: Gingerman Pub, Finean St
Contact: soneillm[at]tcd.ie

Durham / Raleigh / Research Triangle, North Carolina
Time: Wednesday, August 15, 7:30
Location: Ponysaurus Brewing Company, Durham
Contact: willdjarvis[at]gmail.com

Edinburgh, Scotland, UK
Time: Sunday, August 12, 6:00
Location: BrewDog, Lothian Road (table booked under the name “Hughes”)
Contact: ???, but see Facebook group

Edmonton, Alberta, Canada
Time: Saturday, August 11, 1:00
Location: Garneau Remedy Cafe, 8631 109 St NW
Contact: ???

Escazú, Costa Rica
Time: August 25, 2:00
Location: Food Court in Multiplaza, Escazú
Contact: Jorge via WhatsApp at +506 6219 9230

Grand Rapids, Michigan
Time: Saturday, August 25, 3:00
Location: Downtown Market
Contact: ???

Helsinki, Finland
Time: Tuesday, August 28, 6:00
Location: Dubliner (Kaivopiha)
Contact: schelsinkimeetup[at]gmail.com, or see website

Honolulu, Hawaii
Time: Saturday, August 11, 12 noon
Location: Starbucks 680 Ala Moana Blvd A101
Contact: richelson.david[at]gmail.com, 520-369-3359

Houston, Texas
Time: Sunday, August 26, 10 AM
Location: Agora (Coffee Shop), 1712 Westheimer Rd
Contact: alex.m.weldy[at]gmail.com, or see Facebook group

Irvine / Orange County, California
Time: Friday, August 10, 6:00
Location: Center of Aldrich Park, UC Irvine
Contact: gabeaweil[at]gmail.com, or call Gabe at (202) 681-1383

Istanbul, Turkey
Time: Sunday, August 19, 2:00
Location: Espresso Lab in Istiklal Cd., Taksim (across the street from the French embassy)
Contact: vsural[at]gmail.com

Jersey City, New Jersey
Time: Saturday, August 11, 11 AM
Location: Apartment at 30 Newport Pkwy
Contact: 37davidg[at]gmail.com , 4088579822

Kansas City, Missouri
Time: Sunday, August 19, 12 noon
Location: Mud Pie Bakery, 1615 W 39th St
Contact: fustruly[at]gmail.com

Kyiv (Kiev), Ukraine
Time: Saturday, August 11, 11 AM
Location: Bookshop at Lobanovskogo avenue, 6 A (“Дім Книги”), ask for Marichka
Contact: chernyshenko123[at]gmail.com

Lexington, Kentucky
Time: Sunday, August 19, 7 PM
Location: Blue Stallion Brewing, 610 W 3rd Street
Contact: hellernathan19[at]gmail.com

London, UK
Time: Sunday, August 12, 2:00
Location: Private house, 48 St Paul’s Crescent, NW1 9YA
Contact: philip.hazelden[at]gmail.com, or see Facebook group

Los Angeles, California
Time: Wednesday, August 29, 7:00
Location: Wine Bar next to Landmark Theater in Westside Pavillion, 10850 Pico Blvd #312

Contact: betterscale[at]protonmail.com, or see Google group, or contact T3t on Discord

Madison, Wisconsin
Time: Saturday, August 18, 6:00
Location: The Roman Candle, 1054 Williamson St
Contact: cmfrayne[at]gmail.com, Facebook event

Madrid, Spain
Time: Saturday, September 15, 12 noon
Location: Crepes and Waffles, Calle de Fuencarral, 105, Madrid
Contact: pavill01[at]ucm.es, or see meetup page or Facebook group

Manchester, UK
Time: Saturday, August 18, 3:00
Location: The grass patch next to the Wharf Pub in city center M15 4ST coordinates 53.473975, -2.256884
Contact: kernelmanchester[at]gmail.com, or message Ben on Facebook

Melbourne, Australia
Time: Friday, September 7, 6:00
Location: The Queensberry Hotel Dining Room, 593 Swanston St, Carlton
Contact: Call Jugemu, 0438 869 257, or see Facebook group

Miami / Fort Lauderdale, Florida
Time: Saturday August 11, 5:00
Location: Passion del Cielo Coffee, 3301 NE 1st Ave #100
Contact: contact[at]postlibertarian.com, or see Facebook group or Facebook event

Montreal, Quebec, Canada
Time: Saturday, September 1, 2:00
Location: Private house, 810 Avenue Duluth E, Montréal, QC

H2L 1B3
Contact: mathieu.roy.37[at]gmail.com, or see Facebook group or Facebook event

Moscow, Russia
Time: Saturday, August 25, 4:00
Location: Kocherga rationality time-club at Bolshaya Dorogomilovskaya, 5 building 2 (https://kocherga-club.ru/#contacts for directions)
Contact: Alex at nikkou.postbox@gmail.com

Munich, Germany
Time: Saturday, August 11, 5:00
Location: Guddenstraße 7, Munich
Contact: lw[at]1platz.de, or see meetup page or Facebook group

Nashville, Tennessee
Time: Tuesday, August 21, 7:00
Location: Industrious, 1033 Demonbreun Ave, Suite 300
Contact: james[at]writechem.com, or see Facebook event

New York City, New York
Time: Saturday, August 18, 3:00
Location: Hudson Eats Food Court, tables near the Black Seed Bagel (will have a sign at the tables), Brookfield Place, 2nd Floor, 230 Vesey Street, Manhattan (map)
Contact: Rachel at (650) 534-7354, or see Google group

Norman, Oklahoma
Time: Saturday, August 11, 6:00
Location: Barnes & Noble 540 Ed Noble Pkwy
Contact: fischerjoe855[at]gmail.com

Northampton, Massachusetts
Time: Saturday, September 8, 6:30

Location: The Roost (cafe), 1 Market Street
Contact: alex[at]alexliebowitz.com

On Board The Battleship USS Iowa
Time: Saturday, September 8, 1:30
Location: Battleship USS Iowa Museum, 250 S Harbor Blvd, Los Angeles / San Pedro
Contact: Who do you think?

Oslo, Norway
Time: Thursday, August 23, 7:00
Location: Starbucks on Aker Brygge
Contact: anders[at]huitfeldt.net

Ottawa, Ontario, Canada
Time: Friday, August 10, 6:00
Location: Art House Cafe, 555 Somerset Street W, Ottawa, Ontario K1R 5K1
Contact: StefanDeYoung[at]gmail.com, or see Facebook group

Oxford, UK
Time: Thursday, August 16, 7:15
Location: The Swan & Castle
Contact: hbesceli[at]gmail.com

Paris, France
Time: Saturday, September 1, 3:00
Location: 48°51’37.9″N 2°17’11.8″E (map)
Contact: Call Jules at +33.6.soixante-quinze.99.65.84

Phoenix, Arizona
Time: Saturday, August 25, 10 AM
Location: 2933 N 17th Dr
Contact: Text Wong at 740 346 9664

Philadelphia, Pennsylvania
Time: Thursday, August 30, 6:00
Location: Tir Na Nog Irish Pub, 1600 Arch Street
Contact: dmcbriggs[at]gmail.com

Pittsburgh, Pennsylvania
Time: Saturday, September 15, 12 noon
Location: Private home, 326 Lincoln Ave, Carnegie PA
Contact: mirandagavrin[at]gmail.com, please RSVP if attending

Portland, Oregon
Time: Saturday, August 11, 12:30
Location: Private residence, 3246 NE 57th Ave (cross street is Klickitat)
Contact: nwalton125[at]gmail.com, encouraged to RSVP

Prague, Czech Republic
Time: Sunday, August 26, 6:30
Location: Tea Room Dharmasala, Peckova, 296/15
Contact: nadvornik.jiri[at]gmail.com, or see Facebook event

Pune, India
Time: Saturday, August 11, 3:00
Location: Cafe Coffee Day outlet at IISER campus
Contact: 7rat13[at]gmail.com

Rochester, New York
Time: Sunday, August 26, 1:00
Location: Wegman’s Market Cafe Food Court, 745 Calkins Road
Contact: christopherjpohlman[at]gmail.com

Sacramento, California
Time: Sunday, August 12, 5:00
Location: 1627 16th St, Sacramento, CA 95814
Contact: amethyst.eggplant[at]gmail.com

Salt Lake City, Utah
Time: Saturday, August 11, 3:00
Location: City Creek Food Court
Contact: wearenotsaved[at]gmail.com, or see Facebook event

San Antonio, Texas
Time: Saturday, August 11, 2:00
Location: Local Coffee @ 7338 Louis Pasteur Dr #204, look for the table with the monster on it
Contact: waffles.he.waffles[at]gmail.com

San Diego, California
Time: Saturday, August 11, 2:00
Location: Bonita Cove Park
Contact: the.god.empress.celestia[at]gmail.com or see Facebook page or Google group

San Francisco, California
Time: Monday August 13, 6:15
Location: 855 Brannan St, lobby across the street from REI
Contact: Call 203-503-7508

San Jose, California
Time: Saturday, September 1, 2:00
Location: Private house, 3806 Williams Rd
Contact: ddfr[at]daviddfriedman.com

Santa Rosa / Sonoma, California
Time: Saturday, August 25, 3:00
Location: 4656 Quigg Dr, Santa Rosa
Contact: adrian@smithdev.io or VivaLaPanda#6386 on Discord

São Paulo, Brasil (Brazil)
Time: Saturday, August 25, 2:00
Location: Vanilla Cafe, R. Antônio Carlos, 452 (near Metro

Consolação)
Contact: gusbicalho[at]gmail.com, or see meetup group, Google group, or Facebook group

Seattle, Washington
Time: Sunday, September 9, 2:00
Location: Private house, 5238 11th Ave NE
Contact: speeze.pearson+ssc[at]gmail.com, or see Facebook event

Sheffield, UK
Time: Sunday, September 30, 12 noon
Location: Private house, Crookesmoore, S10 1DJ
Contact: jwhiteley1[at]sheffield.ac.uk

St. Louis, Missouri
Time: Saturday, August 25, 2:00
Location: Coffee Cartel, 2 Maryland Plaza
Contact: djcart137[at]gmail.com, or call 636-219-0944

Stockholm, Sweden
Time: Saturday, August 11, 3:00
Location: Cafe Dox, Stora Nygatan 31, 111 27 Stockholm (if closed, Grändens Cafe, Yxsmedsgränd 2)
Contact: ???, but see Meetup page

Sydney, Australia
Time: Thursday, August 16, 6:00
Location: 565 George Street, level 2, “Fishbowl Room”
Contact: Call Elo at +61438481143

Tel Aviv, Israel
Time: Monday, August 13, 7:00
Location: Melchett 23, Tel Aviv (3rd floor, apartment 12, building code 7890 ‘enter’)
Contact: hi[at]davidkatz.me

Toronto, Ontario, Canada
Time: Saturday, August 18, 5:00
Location: Prenup Pub (St George & College Street)
Contact: Ryan at (647) 562-7358

Tuebingen, Germany
Time: Wednesday, August 18, 8:00
Location: Willi Tübingen, Wilhelmstraße 3/1, 72074 Tübingen
Contact: y.muehlhaeuser[at]gmx.de

Vancouver, British Columbia, Canada
Time: Saturday, August 11, 3:00
Location: Central Library, downtown Vancouver, meet just inside main entrance
Contact: goodtown.r[at]gmail.com

Vienna, Austria
Time: Saturday, August 18, 3:00
Location: Kaisermühlenstrasse 24. “The meeting room is on the ground floor of the apartment buidling which is across the road from the Stadlau station. The room entrance is on the opposite (eastern) side of the building, between Erich-Fried-Weg and Fahngasse. (map) ”
Contact: moses.skoda[at]gmail.com, or see Facebook group

Washington DC
Time: Saturday, August 25, 7:00
Location: 616 E St NW
Contact: robirahman94[at]gmail.com or see Google group

Wellington, New Zealand
Time: Thursday August 16, 6:00
Location: Enigma, 128 Courtenay Place
Contact: 021 885 180

Zurich, Switzerland
Time: Saturday, August 25, 4:00

Location: The Bäckeranlage Park
Contact: juri2mol[at]gmail.com

SSC Survey Results: ADHD And Rejection Sensitivity

Introduction

ADHD is typically considered a disorder of attention and focus. There are various other traits everyone knows are linked – officially, hyperactivity and "behavior problems"; unofficially, anger and thrill-seeking – but most people consider these to be some sort of effect of the general attention deficit.

Dr. William Dodson pushes a different conception, where one of the key features of ADHD is "rejection-sensitive dysphoria", ie people with the condition are much less able to tolerate social rejection, and more likely to find it unbearable and organize their lives around avoiding it. He doesn't deny the attention and focus symptoms; he just thinks that rejection sensitivity needs to be considered a key part of the disorder.

I say "Dr. William Dodson pushes", but this requires a little research before it becomes apparent. *What a Google search shows* is just a bunch of articles saying that rejection sensitivity is a key part of ADHD that gets ignored by non-expert psychiatrists and that it's important to educate patients about it and include it in any treatment plan. *My conclusion* is that all of these articles can be traced back to Dr. Dodson or people inspired by Dr. Dodson, of which there are many. The ADHD patient community has gotten really into this and pushed it in a lot of support groups and patient communities and so on, where it is repeated uncritically as "an important ADHD feature psychiatrists often forget about". But the genesis is just Dr. Dodson saying so, with limited formal evidence.

See for example ADDitude Magazine, which says that:

> Rejection sensitivity is part of ADHD. It's neurologic and genetic. Early childhood trauma makes anything worse, but it does not cause RSD. Often, patients are comforted just to know there is a name for this feeling. It makes a difference knowing what it is, that they are not alone, and that almost 100% of people with ADHD experience rejection sensitivity. After hearing this diagnosis, they know it's not their fault, that they are not damaged.

…and suggests very high doses of alpha-agonists or MAOIs (!) as a treatment. Or WebMD, which says:

> Up to 99% of teens and adults with ADHD are more sensitive than usual to rejection. And nearly 1 in 3 say it's the hardest part of living with ADHD.

Neither article cites any sources.

I am skeptical of the rejection-sensitive dysphoria concept. Part of it is the lack of evidence beyond Dr. Dodson's personal experience, which my own personal experience contradicts. But part of it is that it seems suspicious for the Forer Effect, the tendency for everyone to believe a generic statement describes them especially. This effect is beloved of psychics – tell a client a few Forer statements and they'll walk out convinced you can read minds – but it causes problems in the more reputable sciences as well. Here is Forer's original list of statements that produce the effect:

> 1. You have a great need for other people to like and admire you.
> 2. You have a tendency to be critical of yourself.
> 3. You have a great deal of unused capacity which you have not turned to your advantage.
> 4. While you have some personality weaknesses, you

are generally able to compensate for them.
5. Disciplined and self-controlled outside, you tend to be worrisome and insecure inside.
6. At times you have serious doubts as to whether you have made the right decision or done the right thing.
7. You prefer a certain amount of change and variety and become dissatisfied when hemmed in by restrictions and limitations.
8. You pride yourself as an independent thinker and do not accept others' statements without satisfactory proof.
9. You have found it unwise to be too frank in revealing yourself to others.
10. At times you are extroverted, affable, sociable, while at other times you are introverted, wary, reserved.
11.Some of your aspirations tend to be pretty unrealistic.
12. Security is one of your major goals in life.

The first statement sounds a lot like rejection sensitivity. I will admit that one of the strongest pieces of evidence in favor of rejection-sensitivity dysphoria is all the ADHD patients who comment on any article about it with "Wow, this perfectly describes my experience! This is amazing!" – but if it were a Forer Effect, this would be less surprising.

There have been two studies attempting to investigate the ADHD-rejection link. The first, Canu & Carlson, failed to find it. The second, Bondu and Esser, did find it. This may be because the second had a larger sample size. But the second also correlated rejection-sensitive dysphoria with ADHD symptoms, rather than relying on an ADHD diagnosis. It also failed to screen out other psychiatric disorders, many of which are comorbid with ADHD and look a lot like rejection sensitivity (eg anxiety, depression). Overall this does provide some support for the hypothesis, but so far it looks like the only attempt to formally test it.

Methods

I investigated the idea of rejection-sensitive dysphoria through the Slate Star Codex survey, an online survey of readers of this blog. It received 8,077 responses, including 717 people with professionally-diagnosed ADHD, 860 more people with self-diagnosed ADHD, and several thousand with other psychiatric conditions.

The survey contained the following question:

> I am more sensitive than others to rejection, teasing, criticism, or my own perception that I have failed or fallen short

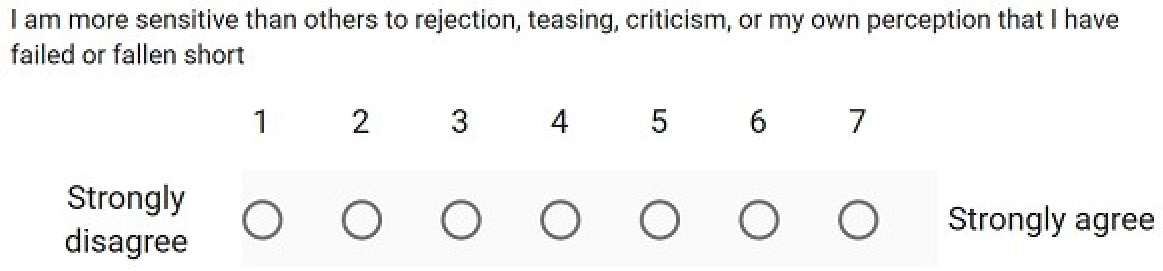
This is a standard screen for rejection-sensitive dysphoria quoted as Dr. Dodson’s proposed addition to the DSM diagnostic criteria for ADHD (though I cannot find confirmation of this from Dr. Dodson himself). Most of the citations frame it more strongly, eg “For your entire life have you always been much more sensitive than other people you know to rejection, teasing, criticism, or your own perception that you have failed or fallen short?”. I cannot remember why I made the survey version weaker and this is a weakness of this study. Respondents were asked to rate their level of agreement with this statement on a scale from 1 (strongly disagree) to 7 (strongly agree).

The following analysis strategy was devised before looking at any data: to compare people with ADHD only to people with no psychiatric condition, to people with only one other psychiatric condition, and to people with multiple psychiatric conditions. The three other conditions with a large enough sample size to use were autism, anxiety, and depression. Analyses were done both among self-diagnosed patients and professionally diagnosed patients. I committed to publishing this post regardless of what the analysis said.

All other analyses were done after seeing the data, and should be considered exploratory only.

Results

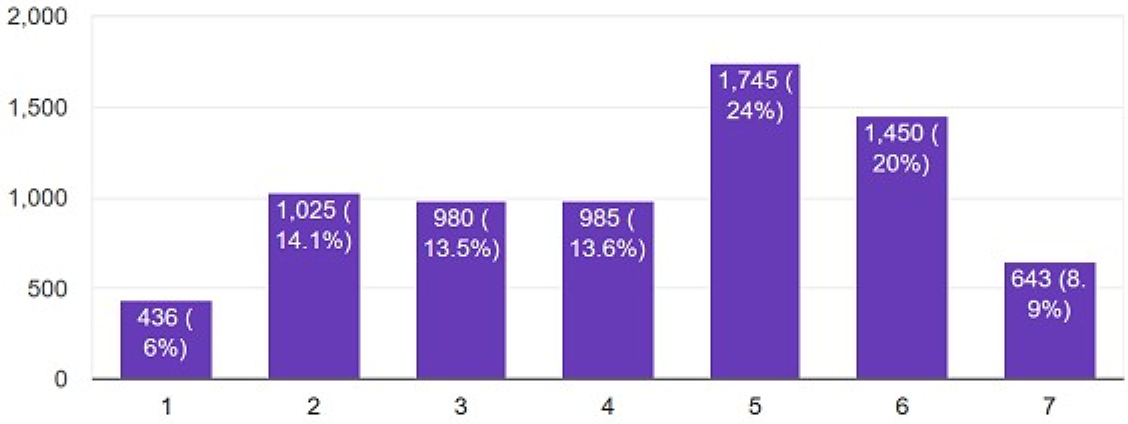
7,264 people answered the question about rejection-sensitive dysphoria:



I am more sensitive than others to rejection, teasing, criticism, or my own perception that I have failed or fallen short

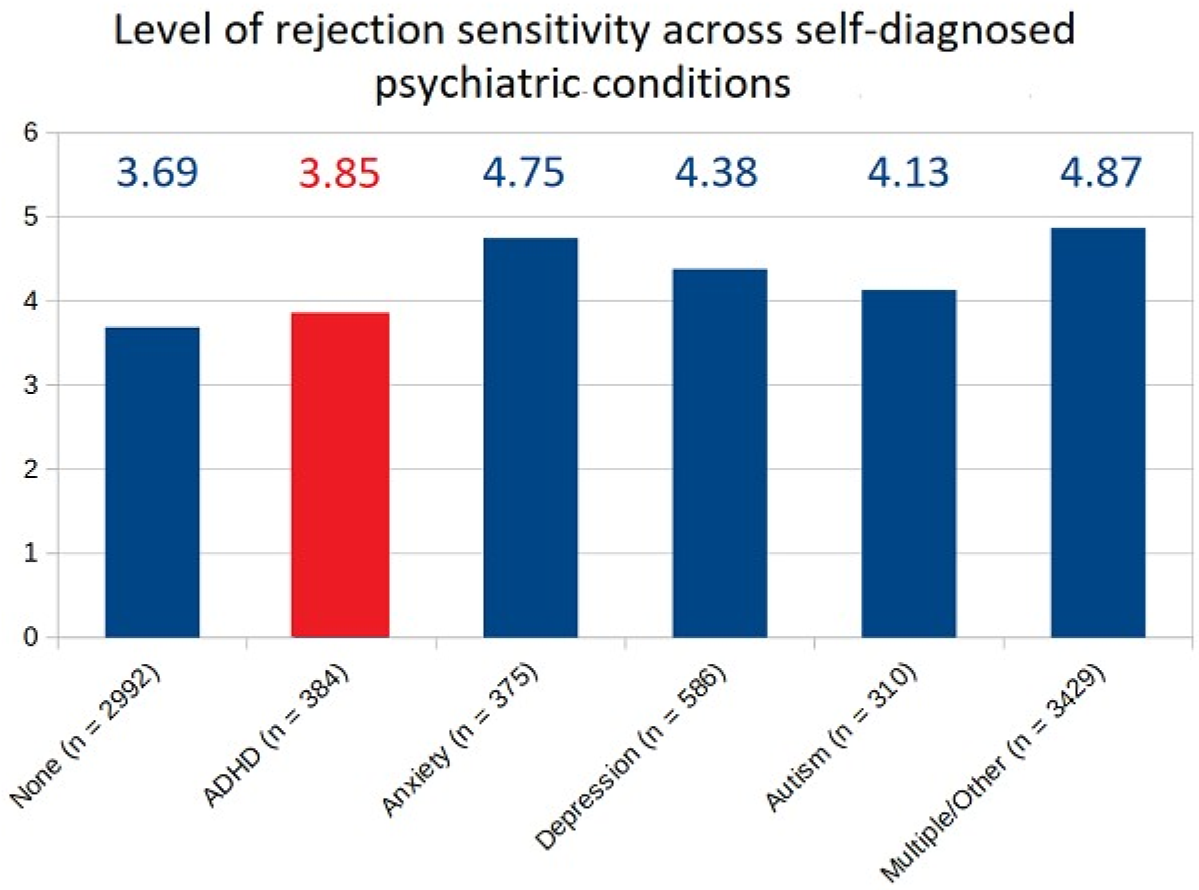
1 2 3 4 5 6 7

Strongly disagree ○ ○ ○ ○ ○ ○ ○ Strongly agree



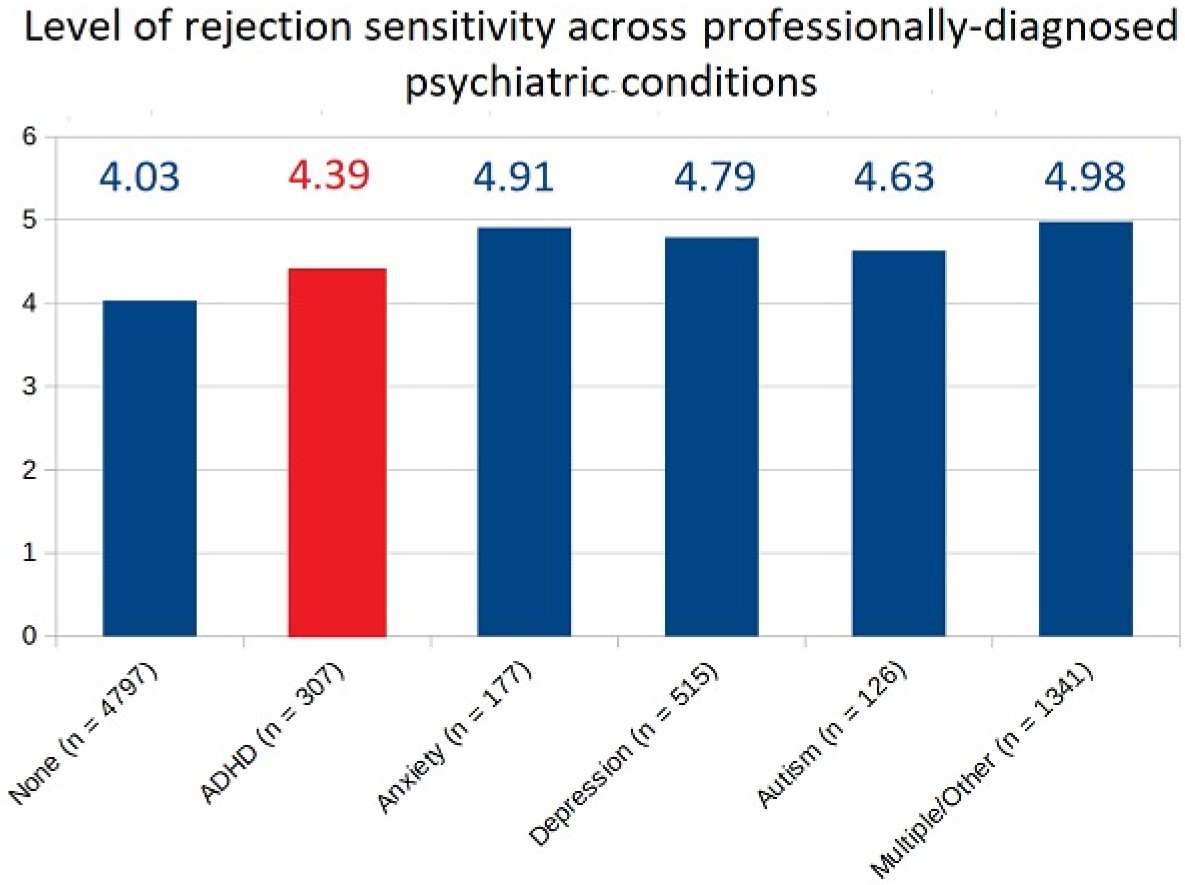
There was a slight bias towards affirmative answers, consistent with a small (but not large) Forer effect.

Among people with self-diagnosed psychiatric conditions, answers varied as follows:



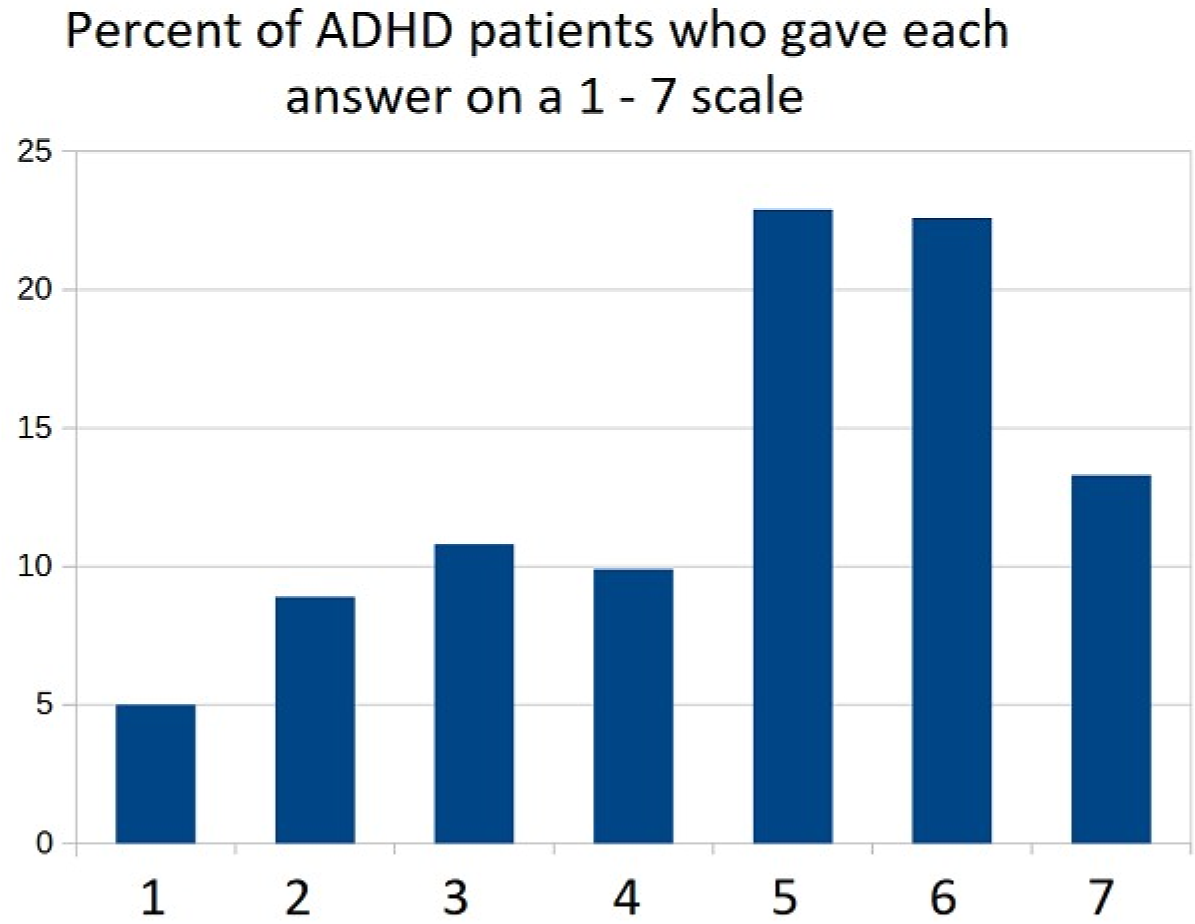
ADHD had the lowest rate of rejection sensitivity among the four psychiatric conditions studied, although it was still higher than people with no condition.

Among people with professionally-diagnosed psychiatric conditions, the picture was much the same:



Once again, ADHD had the lowest rate of rejection sensitivity among the four psychiatric conditions studied, but was higher than people with no condition.

Articles alternately claim that “99%” “99.9%” or “almost 100%” of people with ADHD endorse rejection sensitive dysphoria. My survey found less dramatic results:



58.5% of ADHD patients gave answers of 5, 6, or 7 on a one-to-seven scale of rejection sensitivity. This was statistically significantly different from the 46.2% of the general population who did the same, but not in nearly as impressive a way as suggested by the 99.9% numbers that some people quote.

Discussion

My survey found that there was a weak but statistically significant tendency for people with ADHD to have higher rejection sensitivity than people without psychiatric conditions. However, this was a general tendency for any psychiatric disorder, and ADHD had the lowest rates of rejection sensitivity among the four disorders studied. This suggests the conception of rejection sensitivity as fundamental to ADHD is inaccurate and should not be included in discussions of the condition.

Flaws in this study include the survey methodology, which relied on self-report to get each respondent’s diagnosis or lack thereof. On a deeper level, if the current conception of ADHD is fundamentally flawed because of ignoring rejection-related symptoms, then no diagnosis using the current method can be trusted. But to call something ADHD at all, rather than a completely new illness, requires that it have some correlation with ADHD features as classically understood. Given that there is no generally-accepted objective measure of ADHD, these diagnoses are the best measurement we have at this time.

Other flaws included the admittedly weak single-question test for rejection sensitivity, the inexplicable softening of the single question from its classic version, the lack of adjustment for any confounders, and the heavily selected sample of SSC readers.

However, it is still difficult to argue with the magnitude of the result in a large sample size such as this one. ADHD has the least rejection sensitivity of any of the disorders studied. None of the flaws in this study seem to be of the type or the magnitude that could incorrectly produce this result.

I conclude that it is important to beware of Forer effects in ideas about psychiatric symptomatology, especially ideas that spread among patient groups without formal study or buy-in from researchers. Forer statements are often vague, slightly self-flattering, or suggest that the person involved needs special care or respect from others. Since most people feel this way, it may be easy to convince people with a condition that the condition implies a Forer statement. My informal survey does not support this connection at this time. Other teams should follow up with more formal experiments to try to confirm.

SSC Survey: Scattered Negative Results

Traffic to this blog is declining. I need to act decisively to draw people back. Write something so interesting it can't *help* but go viral. I'm going to write about…negative results from the perception questions on last year's survey.

The last SSC survey had a lot of optical illusions and visual riddles. I had hoped to expand on some of the work in Why Are Transgender People Immune To Optical Illusions and Can We Link Perception And Cognition? This post is a very brief summary of results and, basically, an admission of failure. While I was able to replicate the same suggestive results as in the last survey, I was unable to expand on them, strengthen them, or really turn them into any kind of interesting framework.

I was able to weakly replicate the headline result from Why Are Transgender People Immune To Optical Illusions: transgender status still correlated with all three mask illusions, and with the average of all three mask illusions, but very weakly: r = -0.04, p = 0.001. This was true even when I excluded everyone who took place in last year's survey, providing an independent confirmation of the result. But with correlations this low, it's hard to get too excited.

I was also able to weakly replicate the headline result from Can We Link Perception And Cognition?. I haphazardly gave people a "weirdness score" based on them having more mental illnesses, more unusual political opinions, and more minority sexual/gender identities (without looking at their illusion results). People with higher weirdness scores consistently had more ambiguity-tolerant results on illusions, with correlations around r = 0.05 for most tests.

They also had notably higher average Tolerance of Uncertainty Test scores. But none of these results were very striking and there was minimal individual structure in them. If I was going to take this further I would have come up with a more principled definition of weirdness, but at this point it doesn't seem worth it.

What do I mean by saying these results are weak and lack internal structure? To give an example: the last survey focused on a single optical illusion, the Hollow Mask. This survey used three different versions of the Hollow Mask in the hopes of removing noise and getting a higher-fidelity mask perception signal. This didn't work at all. The correlations between the three masks were very low. For example, there was only an r = 0.09 correlation between being able to see the second mask illusion and the third mask illusion. While all correlations were significant, it doesn't seem fair to conceptualize them as testing the same perceptual function.

Given that I couldn't even get different versions of the same illusion to line up, you can guess that I didn't get much correlation between different illusions. For example, the Tables Illusion correlated with average score on the Mask Illusion at about r = 0.06, p = 0.001. I was also able to replicate the correlation in the literature between autism and the Tables Illusion, r = 0.03 p = 0.01, but again this was very small.

Were there any actually large correlations? Surprisingly, the Surgeon Riddle produced some of the most impressive results on the whole survey. For example, it correlated at r = 0.18, p = 0.001 with ability to see "the the" as two separate words. Given that the Surgeon Riddle seems much less perceptually basic than the other things on here, I believe this is probably some kind of confounder, maybe amount of time people spend on each question or something like that.

There was an overwhelmingly strong effect of age on the Parentheses Riddle and nothing else. 62% of people in their 20s got it right, compared to 29% of people in their 70s. This was surprising enough that it might be worth its own post. I was unable to wring anything else out of this no matter how hard I tried.

I tried a factor analysis to draw some factors out of all the different illusions. SPSS came up with six factors, each of which explained 5-10% of the variance; given how many variables I put in, this doesn't look that much better than chance. None of the factors correlated surprisingly well with anything else, nor was there an obvious pattern in which illusions they grouped together.

In conclusion, most illusions had very low (less than r = 0.1) but highly significant (less than p = 0.001) correlations with one another and with various mental illnesses. There was no clear pattern to the correlations, although they did generally replicate past observations and findings from the literature. I can't really say for sure if there's a real effect here or it's all just confounders, and if there is a real effect I definitely can't tease out its structure or say anything about it for sure. I encourage other people to look into these and see if they can do better. You can download the survey results here.

I have a couple other things I pre-registered to investigate but never got around to before, so here they are:

Political conflict theory was slightly correlated with various questions on the Tolerance of Ambiguity and Tolerance of Uncertainty tests, but not enough to be interesting. For example, when asked whether political extremists you don’t like (eg fascists) were making understandable mistakes or just evil, the answer was correlated at about r = 0.05 with various questions like “Small doubts keep me from acting” and “The ambiguities in life stress me out”. The more stressed people were by ambiguity, the more they thought extremists were evil. This was not a very large effect, it varied from question to question in an unpredictable way, and I wouldn’t have mentioned it if I didn’t feel obligated to tie up preregistered loose ends.

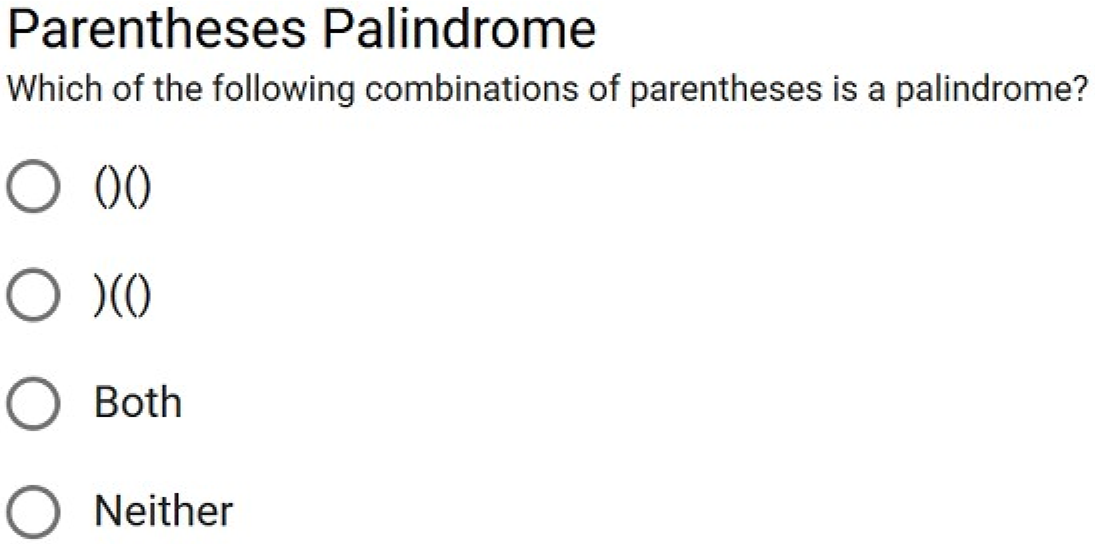
Contra my prediction, there was no relationship between autism and the likelihood of giving process-based (“meta-level”) responses to categorization questions as opposed to person-based (“object-level”) responses. For example, when asked to judge fascists beating up minorities vs. minorities beating up fascists, autistic and neurotypical people were about equally likely to base their responses on principles (ie “beating up people is wrong”) vs. on the groups involved (ie “fascists are bad”).

Contra my prediction, people with ADHD did not describe themselves as more ambitious. They did describe themselves as more risk-taking and more likely to prefer a buzzing-city aesthetic than a quiet-country aesthetic, but the results, although significant, were too low in magnitude to be interesting.

As I mentioned in a previous article, the results of the AI Persuasion Experiment two years ago did not persist.

The Parentheses Riddle

Because I hate you, I included this question on the SSC survey:



Parentheses Palindrome

Which of the following combinations of parentheses is a palindrome?

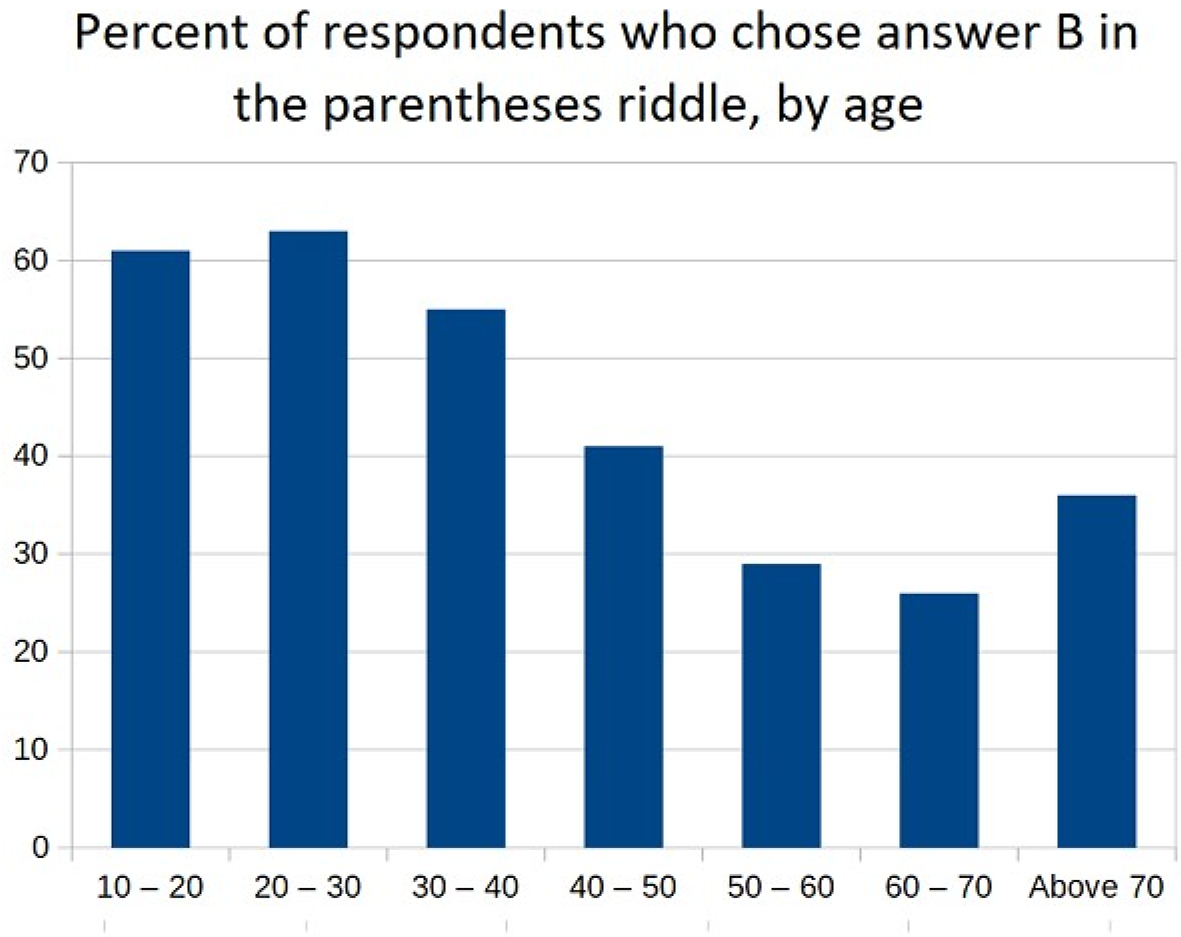
○ ()()

○)(()

○ Both

○ Neither

It's a weird trick question, but I would say B is right. Imagine converting "(" to X and ")" to Y. Then the first answer is XYXY, and the second answer is YXXY. I suppose you could group the parentheses in pairs, in which case the answer would be "both", but in practice few people wanted to say that. Of the 6,000 answers I received, most were either A or B. And one factor had a dramatic effect: age.



This is a big effect. People in their 20s were more than twice as likely to choose B as people their 60s. There's a slight improvement after 70, but I think that's just noise caused by a low sample size in that group.

My first thought was that the younger population on this blog is disproportionately techies, and techies have to work with very finicky parentheses all day. There was indeed a slight tendency for techies to do better on this, but it was a very small part of the effect. Even controlling for that, or limiting the analysis to only non-techies, most of the effect remained.

My second thought was that maybe this was an effect of older people gradually getting less sharp. But IQ itself was correlated with the parentheses question much less than age was. SAT score also didn't correlate very well. And I would expect the most dramatic age-related declines to be after fifty, but the parentheses effect seems to – if anything – slow down then.

My third thought is that older people had less time to waste staring at a dumb survey and trying to solve everything exactly right. But the Squares And Circles Illusion also measures how long you're willing to stare at a dumb question, and age was negatively correlated with this one. Big Five conscientiousness also did not affect parentheses answers.

I am pretty stumped by this. Right now my guess is that it *is* caused by age-related cognitive decline, but the connections go deeper than IQ – there's something about age that affects the parentheses-reading faculty in particular regardless of how smart you are. I know there's a medium-sized dementia-screening industry. I don't know how good they are at detecting the normal level of cognitive decline in relatively young and intelligent people – but I would be interested to hear their opinion on this question and whether it has interesting properties beyond those they already know about.

EDIT: Commenters propose that young people might have seen the riddle before on Facebook or Reddit or some other young-person-Internet-place.

Practically-A-Book Review: EA Hotel

Effective altruism (“EA”) is a movement dedicated to redirecting charity-related resources to the most important and successful charities. In practice this involves a lot of research into how important various problems are, and how well various charities work. Some of this research is done by well-funded official institutions. Other research, maybe exploring more unlikely scenarios or starting from weirder assumptions, is done as individual labors of love. These smaller-scale efforts might be self-funded, or supported by a few small donors. For example, Wild Animal Suffering Research, which investigates ways to improve the lives of animals in the wild, has yet to catch the attention of any hedge fund managers.

Like everything else, effective altruism is centered around San Francisco. San Francisco is the most expensive city in the world, so this isn’t very efficient; most of the relevant research can be done online from anywhere in the world. The official institutional charities eat the expense in exchange for the extra access to funders and other resources, but it’s a problem for small independent organizations. There’s been lots of research into possible solutions, but only if “let’s see how many people we can cram into one house in Berkeley” counts as “research”.

Blackpool is a beach resort in northern England. “Beach resort in northern England” is exactly as fun as it sounds, so nobody goes there. Everything is really cheap, and you can buy a whole hotel for the cost of a parking spot in San Francisco. Enter Greg Colbourn, an effective altruist and successful cryptocurrency investor. He bought the 17-bedroom Hotel Athena and wants to offer free room and board to researchers working on effective altruist projects

WELCOME TO THE ATHENA
LICENSED
LICENSED
WELCOME
36
36
36
Athena

*The plan is to make the E and A different colors than the rest of the word, eg ATH**E**N**A**, for a double meaning.*

Colbourn writes::

> Do you long to be free from material needs and be able to focus on the real work you want to do? I know I’ve certainly been in that situation a few times in the past, but instead have lost time doing unimportant and menial jobs in order to be able to get by financially. Talented effective altruists losing time like this is especially tragic given that a lot of cause areas are currently constrained by the amount of quality direct work being done in them.
>
> Buildings in the run-down seaside holiday resort of Blackpool (UK) are really cheap. I’ve bought a 17 bedroom hotel with dining room, lounge and bar for

> £130k. Assuming a 7% rental yield (which is reasonably high), this works out at about £45 per person per month rent. Factoring in bills, catering, and a modest stipend/entertainment budget, living costs could be as low as £5700/person/year (or lower for people sharing rooms, see budget). This is amazing value for hotel living with all basic services provided.
>
> The idea is to invite people to live there, with all their expenses covered by donors, for up to two years. Funding is already in place (via me) for the first year of operations. The project will be managed by someone who lives on site and deals with all the admin/finances, shopping/cooking/cleaning/laundry, socials/events and morale – they will also have free living expenses, and be paid a modest salary. Note that this should be considered as a potential high impact, high prestige supporting role, for those excited to be involved in such a capacity on an EA mission. Guests will be free from concerns of material survival, and be able to have prolonged and uninterrupted focus on whatever projects they are working on. Obviously these will be largely limited to purely desk-based, or remote work.

Is this really more effective than just spending the money on grants that allow researchers to support themselves? If you know anything about effective altruism, you'll have guessed that someone has already done the math on this. The hotel expects to be able to support people for £5700 (= $7300) / year, so:

> For [giving people grants] to be equivalent, they would have to be living (or go to live) somewhere where the costs of living are comparable, otherwise I would effectively be buying them time at a much higher cost. For example, someone frugal living in London might be

> able to get by on £15,000 a year. So for the same costs they would get about a third of the time; and this is before factoring in the free time-saving services (cooking, laundry, cleaning etc). So unless they were ~4x as productive as the average hotel guest, this would be a bad deal for me as a donor. Also, the community aspect of having a significant number of EAs in the same place is probably worth something too in terms of increased focus, collaboration and morale boosting productivity. Community building via deeper in-person ties is becoming increasingly important to the further development of the EA movement.

Also, and in considerations I have to admit did not occur to me:

> Blackpool might be hard to get to in the event of a catastrophe, but the flip side of this is that there would be a lower risk from hostile actors (mercenaries, milita), as well as lower direct damage from nukes and fallout… the cellars could serve as a nuclear bunker of moderate protection. It will be relatively low cost to keep a stockpile of long lasting food down there, which could be slowly used and replenished by the kitchen over a 2-5 year cycle. There is already bathroom plumbing down there, and other essentials could be easily added.

The hotel has another advantage, harder to explain. There's a lot of concern in the EA community about Goodhart's Law: "When a measure becomes a target, it stops being a good measure". Think of it as "teaching to the test", but for a broad and metaphorical definition of "test". Or as cancer researchers who can't research the most promising fields because they're busy researching the fields that sound the best on a grant application, or the ones that will produce some small payoff in a year so that grantmakers classify them in the “productive” category and renew their funding. Or as the CEO who can’t pursue the most promising strategy because he’s busy pursuing the strategy that will maximize shareholder value next quarter. Effective altruism wants to avoid that failure mode. But as long as you’re in the business of distributing scarce resources to the people who seem to be able to use them best, you’re at risk.

It looks like the hotel wants to accept all comers who are making a good-faith effort to work on effective altruist ideas. I don’t know their exact plan or whether it will work. But having a low-barrier-to-entry hotel in an abandoned beach resort at least sounds *less* exploitable than giving people low-barrier-to-entry multi-thousand-dollar grants. The white paper says:

> Scammers can be avoided, at least in the first instance, by requiring a prior history of involvement in EA and references from people with standing in the community. The bigger risk is well intentioned but ineffective grant recipients (/hotel guests) just not delivering. However, we are likely already in a situation where EAs frequently go off on their own to work on individual projects that don’t deliver, but we don’t hear much about it because of selection effects/social desirability bias

And:

> Communal mealtimes will be encouraged though, as a way for camaraderie and a sense of community to develop amongst guests. Friendships will be formed, problems, and their solutions, discussed, and some amount of fun and laughter will be had too hopefully! Also, some people may find that the inevitable discussion of work that will arise will lead to some

> amount of peer pressure that will aid in keeping them productive […]
>
> In terms of guarding against people becoming freeloaders (or effectively squatters), perhaps there could be a mandatory minimum of very brief (one page) reports on progress to be filed every three months. There would also be some indirect self-generated pressure from guests worrying about donors getting their money's worth (although perhaps this will only be applicable for the more highly scrupulous). But then there would have to be a mechanism for chucking people out if they are achieving very little. The hard cap of two years on length of free stays per person will be an ultimate limiter, but it might be prudent to proceed with caution regarding taking on long term residents (perhaps a three month probation period would be useful).

Would people really want to live here? Apparently. In the month or two it's been open, it's already picked up three residents. The Facebook poll shows 115 expressions of interest, so if even a fraction convert to real bookings, they can keep their seventeen rooms full for a long time.

I have no idea if this project is a good use of anyone's time or money, but I find it endearing. Part of this is of course the giant middle finger raised at Henry VIII – undissolution of the English monasteries and all that. But it's also how I picture a near-term future where everything goes better than expected. A modest basic income could give everyone the opportunity to live a spartan but generally pleasant lifestyle in a friendly and meaningful community of intellectually-aligned people. As the old saying goes: "If it happens, it's possible".

Further links:

1. EA Hotel “white paper”

2. EA Hotel website

3. EA Hotel Facebook group

4. Apply to stay at the EA Hotel. Effective altruists are welcome to apply to live there for free; others are invited to pay for rooms at the usual rate if they want to go there on holiday / gawk at the interesting social experiment.

Carbon Dioxide: An Open Door Policy

[Content note: reading this post might cause feelings of suffocation or provoke panic attacks in susceptible individuals. Epistemic status is very speculative.]

Last month I moved into a small cottage behind a big group house. The cottage is lovely. The big group house is also lovely, but the people in it started suffering mysterious minor ailments. Headaches, fatigue, poor sleep – all the things that will make your local family doctor say "Take two placebo and call me in the morning". Using my years of medical training and expertise, I was able to…remain completely unaware of the problem while my housemates solved it themselves.

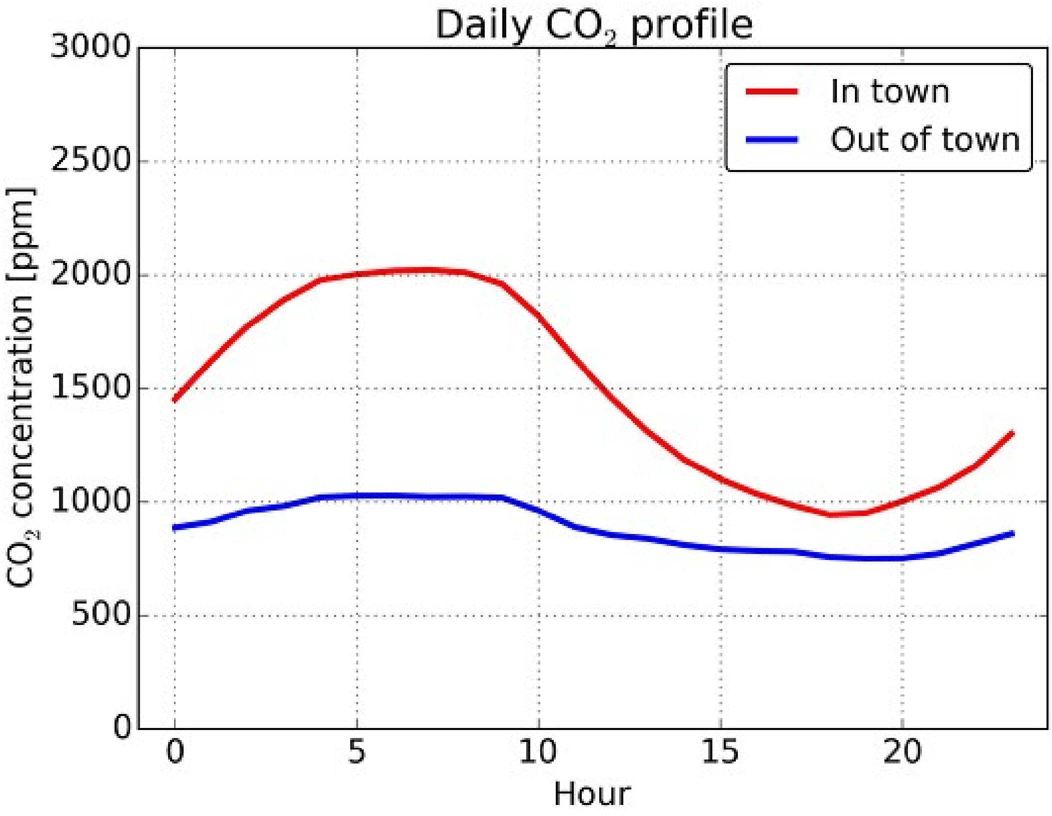
There's been a flare-up of research interest in indoor carbon dioxide levels, precipitated by a Berkeley study (paper, popular article) finding that increasing CO2 concentration from the level of a well-ventilated building to the level of a poorly-ventilated building had profound effects on cognitive ability, cutting various test scores by as much as 50%. This was so dramatic as to be implausible, but seems to match the result of previous Hungarian studies and a later Harvard study on the same subject. The Harvard team later replicated their result with real workers in real offices and found that, controlling for other factors, workers in the best-ventilated offices scored about 25% better on cognitive tests than in the worst-ventilated ones. NASA got really interested in this research because spaceships require a lot of intellectual work and don't have a lot of open windows. They're still running tests but they say that "preliminary results suggest differences" between better- and worse-ventilated environments.

On the other hand, a 2017 study failed to find the effect, possibly because their cognitive tests were easier. And bloggers have pointed out that submarines have more CO2 than the worst terrestrial buildings, but don’t have any problems overt enough for the Navy to notice or worry. So it’s a crapshoot of contradictory results and considerations, just like everything else.

Aware of this research, my housemates tested their air quality and got levels between 1000 and 3000 ppm, around the level of the worst high-CO2 conditions in the studies. They started leaving their windows open and buying industrial quantities of succulent plants, and the problems mostly disappeared. Since then they’ve spread the word to other people we know afflicted with mysterious fatigue, some of whom have also noticed positive results.

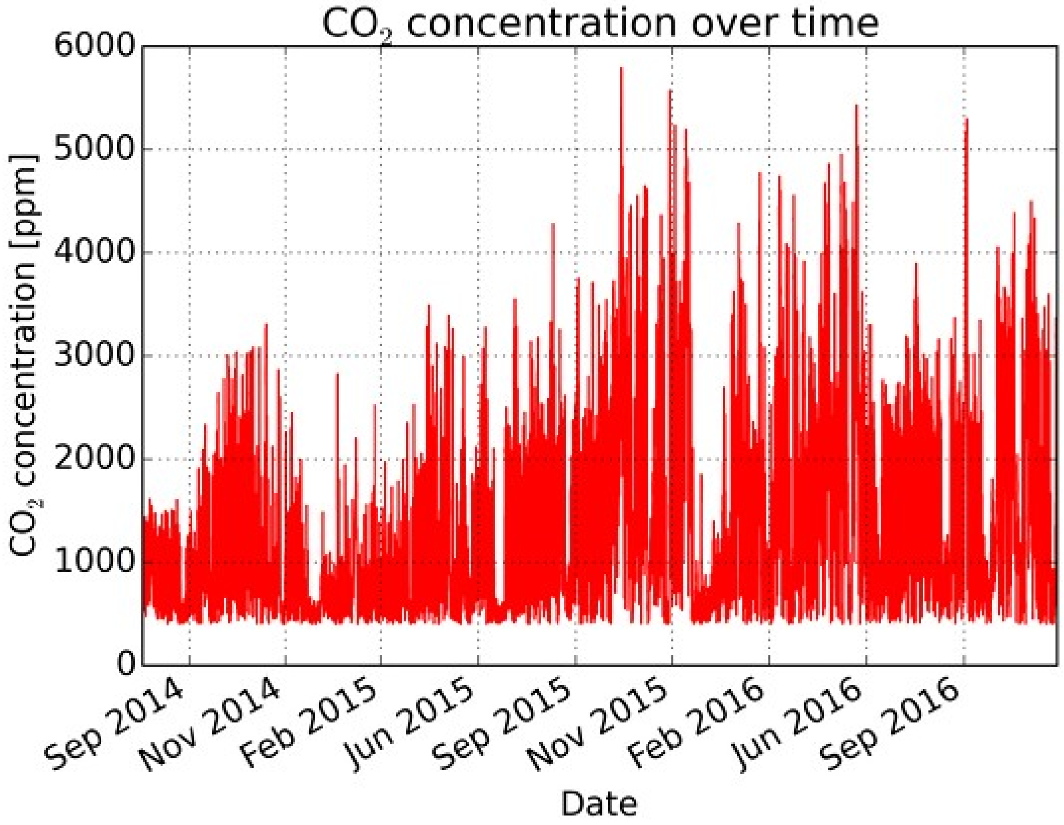
When I heard about this, my first question was: didn’t any of these people notice they only felt bad at home? Shouldn’t it have been a big red flag when they went to the office, or went for a walk, and all their problems disappeared? This can’t be too big a deal, or else “I feel bad in my house, but fine everywhere else” would be a more common complaint.

My housemate Kelsey referred me to the work on CO2 and sleep. Right now this is just a few papers by a guy named Strøm-Tejsen, but the implications are pretty important. He notes that however bad your carbon dioxide levels are during the day they’re probably much worse at night, when you shut yourself up in a small room, close all the doors and windows, and just *breathe* for like eight hours straight. Normal outdoor air is about 400 ppm CO2 (more by the time you read this; thanks, fossil fuel industry!) A well-ventilated building during the daytime is about 700 ppm, and a poorly ventilated building during the daytime about 1400 ppm. But the average bedroom at night can be 2000 ppm or more. Friend-Of-The-Blog Gwern got a CO2 monitor to test these findings, and confirmed that while his daytime CO2 was around 500 ppm, nighttime CO2 in his bedroom could get as high as 3000 ppm. MIT’s Joel Jean discussed trying the same in this Medium post, with similar results:



CO2 levels in my bedroom throughout the day when I’m in town (red) and out of town (blue)

I live in California, and Gwern presumably lives in some kind of formless cybermatrix, so we don’t have to worry about seasons. But Dr. Jean lives in Massachusetts, and he found that during the winter, indoor CO2 went up even further, in some cases exceeding OSHA’s rules for permissible workplace exposure:



CO2 concentration in my bedroom in Cambridge, MA from July 2014 to November 2016. CO2 levels often spike to over 3000 ppm overnight.

If CO2 can affect sleep quality, that would explain how it could produce a whole-day effect. Strøm-Tejsen tests this on sixteen subjects and finds that "objectively measured sleep quality and the perceived freshness of bedroom air improved significantly when the CO2 level was lower, as did next-day reported sleepiness and ability to concentrate and the subjects' performance of a test of logical thinking." Good things about this study: subjects were blinded to condition, the paper contains a pilot experiment and a main experiment which mostly replicate each other's results. Bad things about this study: the experiments were about n = 15 each, the researchers didn't correct for multiple comparisons, and they admit to manipulating the statistics surrounding their logical reasoning tests to get better results.

But if I just look at their tables and try to ignore their manipulation, I’m at least kind of impressed:

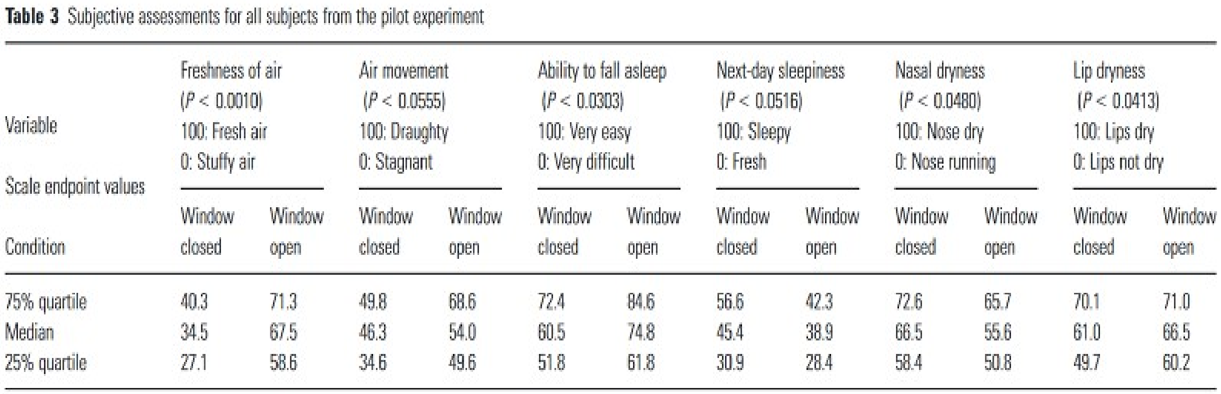


Table 3 Subjective assessments for all subjects from the pilot experiment

Variable Scale endpoint values	Freshness of air ($P < 0.0010$) 100: Fresh air 0: Stuffy air		Air movement ($P < 0.0555$) 100: Draughty 0: Stagnant		Ability to fall asleep ($P < 0.0303$) 100: Very easy 0: Very difficult		Next-day sleepiness ($P < 0.0516$) 100: Sleepy 0: Fresh		Nasal dryness ($P < 0.0480$) 100: Nose dry 0: Nose running		Lip dryness ($P < 0.0413$) 100: Lips dry 0: Lips not dry	
Condition	Window closed	Window open	Window closed	Window open	Window closed	Window open	Window closed	Window open	Window closed	Window open	Window closed	Window open
75% quartile	40.3	71.3	49.8	68.6	72.4	84.6	56.6	42.3	72.6	65.7	70.1	71.0
Median	34.5	67.5	46.3	54.0	60.5	74.8	45.4	38.9	66.5	55.6	61.0	66.5
25% quartile	27.1	58.6	34.6	49.6	51.8	61.8	30.9	28.4	58.4	50.8	49.7	60.2

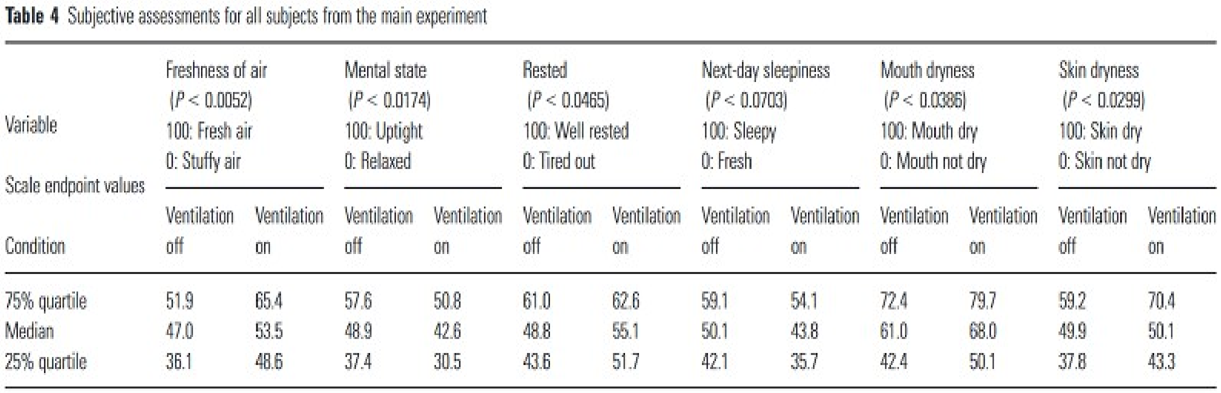


Table 4 Subjective assessments for all subjects from the main experiment

Variable Scale endpoint values	Freshness of air ($P < 0.0052$) 100: Fresh air 0: Stuffy air		Mental state ($P < 0.0174$) 100: Uptight 0: Relaxed		Rested ($P < 0.0465$) 100: Well rested 0: Tired out		Next-day sleepiness ($P < 0.0703$) 100: Sleepy 0: Fresh		Mouth dryness ($P < 0.0386$) 100: Mouth dry 0: Mouth not dry		Skin dryness ($P < 0.0299$) 100: Skin dry 0: Skin not dry	
Condition	Ventilation off	Ventilation on	Ventilation off	Ventilation on	Ventilation off	Ventilation on	Ventilation off	Ventilation on	Ventilation off	Ventilation on	Ventilation off	Ventilation on
75% quartile	51.9	65.4	57.6	50.8	61.0	62.6	59.1	54.1	72.4	79.7	59.2	70.4
Median	47.0	53.5	48.9	42.6	48.8	55.1	50.1	43.8	61.0	68.0	49.9	50.1
25% quartile	36.1	48.6	37.4	30.5	43.6	51.7	42.1	35.7	42.4	50.1	37.8	43.3

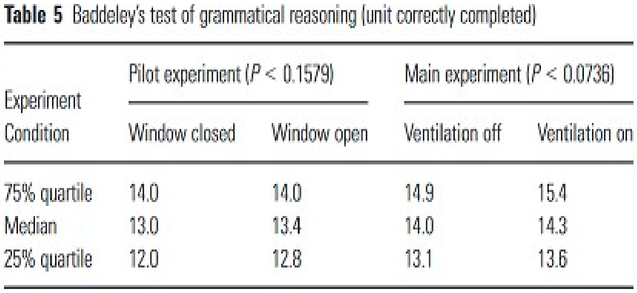


Table 5 Baddeley’s test of grammatical reasoning (unit correctly completed)

Experiment	Pilot experiment ($P < 0.1579$)		Main experiment ($P < 0.0736$)	
Condition	Window closed	Window open	Ventilation off	Ventilation on
75% quartile	14.0	14.0	14.9	15.4
Median	13.0	13.4	14.0	14.3
25% quartile	12.0	12.8	13.1	13.6

And experts seem to take their results seriously – for example, here’s NASA again. And we know from sleep apnea and studies that high physiological levels of carbon dioxide can cause sleep disturbances. I can’t figure out how to convert external ppm to internal likely level of carbon dioxide in the blood, but maybe this could provide a plausible mechanism.

I’m reluctant to be too numerical about all this, because everything about health has massive individual variability. Three people share one of the bedrooms at my group house (look, the Bay Area is really bad). One of them got the typical symptoms of excess CO2 really bad; the other two were fine. Some people are just going to be more sensitive to this kind of thing – the same way three people can drink the same amount of alcohol, two of them will get pleasantly buzzed, and one of them will black out.

I’ve tried sleeping with my door open the past few nights, and I haven’t noticed any difference. Probably I shouldn’t; my house is well-ventilated and I wasn’t feeling too bad beforehand. But I’ve started recommending a few of my patients with mysterious sleep issues try the same thing. It’s too early for results so far, and the science behind it is weak, but it seems like a cheap experiment.

Since the main source of CO2 is human exhalation, I’m most worried about buildings where many people are crammed into small spaces in close proximity (hello, Bay Area readers!). Since the main way CO2 gets cleared is through ventilation, I’m most worried about buildings made to strict environmental standards with great insulation (hello, Bay Area readers again!).

If you’re concerned about this, the best solution is to open a window or an internal door in your bedroom at night. If for some reason this is impossible, the second-best solution is to get certain succulents or other plants that participate in the ominously-named process of “dark fixation” – ie do their plant breathe-in-CO2-and-breathe-out-oxygen thing at night. This is also called “crassulacean acid metabolism” and Googling either term will get you a list of appropriate species. It will probably take like ten succulents to do much to CO2 levels, but a room full of succulents on every flat surface is also kind of #aesthetic.

I’m interested in more data on this, so if you’re planning on experimenting with changes to your nighttime air quality based on this post, please fill out this form to register for an informal quasi-experiment. I’ll follow up with a form for you to give your results in a couple of weeks.

Oh – and sorry for the content warning at the top, but I’ve felt kind of low-grade suffocate-y throughout writing this post, and had to go out and take a couple of breaths of fresh air a few times. Remember – perception is fundamentally Bayesian, and combines external sensation with internal expectations; this is why placebos can have such a profound effect on pain. Perception of air quality vs. suffocation seems to be especially susceptible to this, which is probably one of the major etiological factors behind panic attacks. Just repeat to yourself that it all adds up to normality: the air quality in your room hasn’t changed since when you were feeling just fine before you started reading this article, so you should be okay.

Elegy For John McCain

Say a prayer for John McCain
Who passes from his earthly pain
His eyes are shut upon his brow
He warmongers to angels now

Beyond the sky, where sorrows cease
He rails against the Prince of Peace.
The Holy Spirit, full of love
McCain denounces as “a dove”

All of the weak and the cowardly policies
Heaven pursues that let sin subsist still
Six thousand years of detente with the darkness
In hippie cliches about “choice” and “free will”
All the fifth-columnists, communists, peaceniks
Since ur-commie Lucifer fell from the dawn
John McCain pounds them, he trounces, denounces them
Hounds them and counsels them: cease and begone

All of the saints and the hosts of the angels
Run to their weapons of lightning and flame
Their swords made of sunbeams and sighs of the martyrs,
Their gossamer banners of God’s awesome Name,
Their heavenly helmets and holy habergeons,
Whose breastplates are bright with the light of the dawn;
The Archangel Michael in malachite armor
Blows blasts on his trumpet and beckons them on

Reader, should your weather be
Meteors falling lazily
Or if your neighborhood should seem
A John of Patmos fever dream

Then say a prayer for John McCain
Now passed beyond all earthly pain
Not death, with all the peace it brings
Could end his love of bombing things

Bulls**t Jobs (Part 1 of ∞)

A surprisingly common part of my life: a patient asks me for a doctor's note for back pain or something. Usually it's a situation like their work chair hurts their back, and their work won't let them bring in their own chair unless they have a doctor's note saying they have back pain, and they have no doctor except me, and their insurance wants them to embark on a three month odyssey of phone calls and waiting lists for them to get one.

In favor of writing the note: It would take me all of five seconds. I completely believe my patients when they say their insurance is demanding the three month odyssey. Or sometimes they don't have insurance and it would be a major financial burden for them to consult another doctor. Also, I've seen these other doctors and they have no objective test for back pain. 90% of the time they just have the patient stand in front of them, make whatever movement it is that hurts their back, ask the patient if it hurt their back, and when the patient says yes, the doctor says "That's back pain all right, take some aspirin or ibuprofen or whatever".

Against writing the note: I am a psychiatrist. I usually treat patients via telemedicine, which means that in many cases I have literally never seen their back. All I remember about back pain from medical school is that some people call it "lumbago", a word that stuck in my head because it sounds like a cryptid or small African nation. I know even less about the ergonomics of chairs, or when people do vs. don't require better ones. Any note I write about back pain and chair recommendations is going to be a total sham, bordering on medical fraud. I could demand my patient take time off work to come in for an examination, sometimes from several hours away, just so I can do the thing where they bend their back in front of me and tell me it hurts. But that's kind of just passing the shamminess a little bit down the line in a way that seriously inconveniences them.

In other words: the request puts me in a position where I either have to lie, or have to refuse to give people help that they really need and that it would be trivial for me to provide. It's one of my least favorite things, and I would appreciate any ethical advice the philosophers here have to give.

But my latest strategy is radical honesty. I write a note saying:

> To whom it may concern:
>
> I am a psychiatrist treating Mr. Smith. He tells me that he has chronic back pain ("lumbago"), and asks to be allowed to bring in his own chair to work.
>
> Yours,
> Dr. Alexander

It's too soon to have a good sample size. But it seems to usually work. I think it works because there is nobody at Mr. Smith's workplace – maybe nobody in the entire world – who's really invested in preventing Mr. Smith from bringing a chair into work. Someone wrote up a procedure for employees using special chairs, so that they're not the sort of cowboys who make decisions without procedures. Someone else feels like they have to enforce it, so that they're not the sort of rebel who flouts procedures. But nobody *cares*.

I think a lot about David Graeber's work on bulls**t jobs. In an efficient market, why would profit-focused companies employ a bunch of people who by their own admission aren’t doing anything valuable? I’ve been wondering about this for a long time, and I try to notice when something I’m doing is bulls**t. I guess this fits the bill. It seems to be an issue of people spending time and money to create and satisfy procedures that degenerate into rituals, so that they can look all procedural and responsible in front of – courts? regulators? bosses? investors? I’m not sure. But I do wonder how much of the economy is made of things like this.

Bureaucracy As Active Ingredient

Commenters on yesterday’s post brought up an important point: sometimes bureaucracies aren’t just inefficient information gathering and processing mechanisms. Sometimes they’re the active ingredient in a plan.

Imagine there’s a new $10,000 medication. Insurance companies are legally required to give it to people who really need it and would die without it. But they don’t want somebody who’s only a little bit sick demanding it as a “lifestyle” drug. In principle doctors are supposed to help with this, but doctors have no incentive to ever say no to their patients. If the insurance just sends the doctor a form asking “does this patient *really* need this medication?”, the doctor will always just check “yes” and send it back. Even if the form says in big red letters PLEASE ONLY SAY YES IF THERE IS AN IMPORTANT MEDICAL NEED, the doctor will still check “yes” more often than a rational central planner allocating scarce resources would like. And insurance companies are sometimes paranoid about refusing to do things doctors say are important, because sometimes the doctor was right and then they can get sued.

But imagine it takes the doctor an hour of painful phone calls to even get the right person from the insurance company on the line. Now there’s a cost involved. If your patient is going to die without the medication, you’ll probably groan and start making the phone calls. But if your patient doesn’t really need it, and you just wanted to approve it in order to be nice, now you might start having a heartfelt talk with your patient about the importance of trying less expensive medications before jumping right to the $10,000 one.

Organizations have a legal incentive not to deny people things, because the people involved can sue them. But they have an economic incentive not to say yes to every request they get. Seeing how much time and exasperation people are willing to put up with in order to get what they want is an elegant way of separating out the needy from the greedy if every other option is closed to you.

This story makes sense and would help explain why bureaucracy gets so bad, but I'm not sure it really fits the evidence. People complain a lot about bureaucracy in places like the Department of Motor Vehicles, but the DMV doesn't lose anything by giving you a drivers license and isn't interested in separating out people who really want licenses from people who only want them a little. If the DMV can be as bureaucratic as it is without any conspiratorial explanation, maybe everything is as bureaucratic as it is without any conspiratorial explanation.

But this sort of thing does explain rituals like doctor's notes for back pain or ADHD diagnoses for stimulants. Maybe it fits better with metaphorical bureaucracy than with literal ones. Or maybe it's a factor that disincentivizes existing bureaucracies from getting better. I'm not sure and I don't want to extend the idea further than it will go. It just seems kind of plausible.

This Week: Adversarial Collaboration Entries

This week I'll be presenting entries from the adversarial collaboration contest.

Remember, an adversarial collaboration is where two people with opposite views on a controversial issue work together to present a unified summary of the evidence and its implications. In theory it's a good way to make sure you hear the strongest arguments and counterarguments for both sides – like hearing a debate between experts, except all the debate and rhetoric and disagreement have already been done by the time you start reading, so you're just left with the end result.

A few months ago, I asked readers to write adversarial collaborations and submit them to me. After the inevitable flakeouts and disappearances, I got four entries:

1. Does the current US education system adequately serve advanced students? (by Michael Pershan and TracingWoodgrains)

2. Is Islam compatible with liberal democracy? (by John Buridan and Christian Flanery)

3. Should childhood vaccination be mandatory? (by Mark Davis and Mark Webb)

4. Should children who identify as transgender start transitioning? (by a_reader and flame7926)

I'm going to post one of these per day. Over the weekend, I'll post a link to a poll where readers can vote for their favorite. I’m also going to vote for my favorite, and my vote will be worth 5% of the total number of reader votes. Whoever gets the most votes wins. The prize is $1000; thanks to everyone who donates to the Patreon for making this possible.

Please put any comments about the contest itself here, not on the individual entries.

Reminder: Bay Area SSC Meetup 9/8

Meetup at 3:00 PM on Saturday, 9/8, at the Berkeley campus. Meet at the open space beside the intersection of West Circle and Free Speech Bikeway.

[ACC Entry] Does The Education System Adequately Serve Advanced Students?

[This is an entry to the Adversarial Collaboration Contest by TracingWoodgrains and Michael Pershan (a k-12 math teacher), on advanced students in the education system]

“What do America’s brightest students hear? Every year, across the nation, students who should be moved ahead at their natural pace of learning are told to stay put. Thousands of students are told to lower their expectations, and put their dreams on hold. Whatever they want to do, their teachers say, it can wait.” – A Nation Deceived, p.3

“There is an apparent preference among donors for studying the needs and supporting the welfare of the weak, the vicious, and the incompetent, and a negative disregard of the highly intelligent, leaving them to ‘shift for themselves.'” – Hollingworth, 1926

1. Eager to Learn and Underachieving

Pretend you’re a teacher. With 25 students, who gets your attention during class?

There’s the kid who ask for it, whose hand is constantly up. There’s also the quiet kid in the corner who never says a *word*, but has been lost in math since October, who will fail if you don’t do something. There’s the student in the middle of the pack, flowing along. Finally, there’s the kid who finishes everything quickly. She’s looking around and wondering, *what am I supposed to do now?*

In a survey of teachers from 2008, just 23% reported that advanced students were a top priority for them, while 63% reported giving struggling students in their classes the most attention. A 2005 study found the same trend in middle schools, where struggling students receive the bulk of instructional modification and special arrangements. This was true even while 73% agreed that advanced students were too often bored and under-challenged in school. While teachers, it seems, are sympathetic to the smart bored kid, that’s just not a priority for them.

This isn’t to blame teachers who are under all sorts of pressure to carry low-performing students over the threshold and who, in any event, are only trying to do what’s best for their kids. Which is the most urgent concern? If you don’t equip a kid with the skills they need, next year’s class might be a disaster for them. Or maybe they’ll fail out of school. And behavior problems? Often those begin with academic struggles. Gifted children, on the other hand — they’re on the way to becoming gifted adults. They can take care of themselves, for a minute, the logic goes. More often than not, the teacher will encourage the early finisher to go read a book, or start homework, or do anything at all while the teacher works to help the quiet, lost kid in the corner.

If the kids are just a little bored, that’s nothing strange. It’s hard to find someone who *wasn’t* bored in school sometimes. For many top students, already poised for achievement, this turns out just fine. And yet, there are persistent stories of how the lack of challenge can turn into something more serious.

One version of the story goes like this: from a young age, a student finds the work in school easy. It doesn’t take long for them to *expect* school to be easy for them — it becomes a point of pride. Over years of floating through school, an identity takes hold. Then, one day, maybe after years of schooling, something finally becomes challenging for the student… but there's nothing nice about this challenge. The challenge is now a threat. The student begins to find school challenging, and their world falls apart. They feel isolated and misunderstood at school. They lash out. They hate it, and they can't wait to get out.

When we asked Reddit users and blog readers to describe their experience of school, we heard versions of this story:

- Miserable waste of time, was almost never offered opportunities to learn. Largely ignored teachers and read books during class. I felt like it was a profound injustice that I was punished for doing so. I now have kids of my own and will be home-schooling them.
- I was bored. The pace was too slow and work was not interesting. Being forced by law to get up early and go somewhere to learn things I already know means permanent and firm dislike.
- I went to local public schools for kindergarten through high school, and the experience wasn't good. Academically, the classes were slow and poorly taught. Even the AP classes were taught at the speed of the slowest student, which made the experience excruciating. The honors and regular classes were even worse: I was consistently one or more grades ahead of the rest of the class in every non-AP class except honors math. I learned not to bother studying or doing homework even in the AP classes which probably wasn't great for my work ethic.

The stories of student pain and underachievement in school get more intense as we consider cases of extremely precocious children. The pressures on the student increase, and without help a student often experiences isolation from their peers and a whole other host of difficult feelings. Miraca Gross studied students like these in Australia and found that precocious students were often suffering in silence. Speaking particularly about precocious students who underachieve, she writes:

The majority of the extremely gifted young people in my study state frankly that for substantial periods in their school careers they have deliberately concealed their abilities or significantly moderate their scholastic achievement in an attempt to reduce their classmates' and teachers' resentment of them. In almost every case, the parents of children retained in the regular classroom with age peers report that the drive to achieve, the delight in intellectual exploration, and the joyful seeking after new knowledge, which characterized their children in the early years, has seriously diminished or disappeared completely. These children display disturbingly low levels of motivation and social self-esteem. They are also more likely to report social rejection by their classmates and state that they frequently underachieve in attempts to gain acceptance by age peers and teachers. Unfortunately, rather than investigating the cause of this, the schools attended by these children have tended to view their decreased motivation, with the attendant drop in academic attainment, as indicators that the child has "leveled out" and is no longer gifted.

What do we make of these stories? How common are such experiences?

From the literature on "gifted underachievement" we get partial confirmation — underachievement is a real phenomenon, supported by numerous case studies. According to a survey of various school practitioners, underachievement is the top concern when it comes to gifted students. By definition, advanced students are only a small percent of each student body, so few are affected in any given place, but on a national scale it becomes a more serious problem.

This is not just a problem for the affluent. It has persistent impacts on Black students, poor students, and students who are learning English, who are less often recommended for gifted programs or special accommodations. Here's one way this manifests itself: in one study, 44% of poor students identified as gifted in reading in 1st Grade were no longer academically exceptional by 5th Grade. For higher-income families, only 31% of 1st Graders experience this slide.

The lack of attention to this group extends to the research. It's difficult to pin down the number of students impacted. While underachievement is a real phenomenon, current research doesn't tell us very much about the factors contributing to gifted underachievement. What studies have been done tend to focus almost entirely on things like whether students with ADHD or unsupportive families underachieve, rather than looking at controllable factors like the sort of teaching students experience in school.

Schools are the institutions in charge of educating kids. Those who rush into school, eager to learn, should not walk out feeling rebuffed and ignored. This is doubly true for talented kids from at-risk populations, who may not have the support structure outside of school to ensure their success if school has no time for them. It's clear, though, that we cannot degrade the experience of other students to help those who already have an academic leg up. Is there a feasible approach to address this problem without making things worse?

We have good reason to think that personalized attention makes a huge difference to a student's learning. Research suggests that tutoring that supplements a student's coursework is a very effective educational intervention. Benjamin Bloom caught people's attention with the idea of a 2 standard deviation effect in the 1980s. More recent research has lowered that sky-high estimate to more realistic numbers, and a meta-analysis found an effect size of 0.36, still a powerful impact, enough to take a student from the 50th percentile of achievement to the 64th.

If supplemental tutoring works, the dream goes, what if we replaced classroom work entirely with tutoring? Can't we just do *that* for gifted underachievers and precocious students? We have tantalizing success stories of this kind in the education for precocious children. In a famous case, John Stuart Mill's father decided that the philosophy of utilitarianism needed an advocate, and planned a demanding course for him. Mill didn't underachieve: he learned Greek at age 3, Latin at age 8, and flourished as a philosopher. László Polgár declared he had discovered the secret of raising "geniuses" and went about showing it by tutoring his daughters in chess from the age of 3. It's hard to argue with his results: two grandmasters and an international master, one of whom became the 8th ranked chess player in the world and the only woman ever to take a game off the reigning world champion.

Though this sort of tutoring seems like a dream come true for underachieving gifted students, in practice it's a non-starter in schools. (It lives on in homeschooling, to an extent). In a world where schools are struggling to help every kid learn to read, the ethics of only assigning tutors to gifted students is dubious and almost certainly a political impossibility. The cost of assigning a tutor to every child, meanwhile, would do something special to property taxes.

This simple answer, then, can lead to a clearer understanding of the complexity of educational questions: It’s possible to focus on simple practices that work while disregarding nonacademic concerns and political feasibility.

To be useful, educational ideas should be effective, politically feasible, and economical. If tutoring for gifted underachievers isn’t workable, might there be some other way to approximate the benefits of personal, human attention? Here are three of the most common tools that advocates for gifted education propose:

- Grouping students of the same age in classes according to their abilities, i.e. tracking (National Association for Gifted Children)
- Accelerating students through the curriculum according to their readiness, ignoring age or grade level (A Nation Deceived)
- Using personalization software to replace a human tutor (How Khan Academy Will Help Find The Next Einstein)

What follows is an evaluation of how promising each of these tools is, both in theory and in practice.

Our favorite one-stop reading on gifted education research: this.

Our favorite one-stop reading on tutoring: this.

2. Ability Grouping (a.k.a Tracking)

The case for placing students of similar abilities together in a classroom seems like it ought to be as simple as the case for tutoring. Teachers will be more effective if their students have similar pacing needs. So, group kids who need more time in one class and those who need less time in another. It’s not tutoring, but it should be the next best thing.

Things in education research are rarely that simple, though.

Bob Slavin, a psychologist who studies education, is one of the most-cited education researchers around. He seems like a compulsively busy fellow. He writes, he runs research centers, he designs programs for schools. (He blogs.) A journalist from The Guardian once asked Slavin for his likes and dislikes, and in case you were wondering he likes work and dislikes complacency.

In the late ’80s and early ’90s, Slavin performed a series of meta-analyses of the existing literature on tracking and between-class ability grouping. Overall, he found no significant benefits from ability grouping, even for “top track” students across elementary, middle, and high schools.

But the other surprising finding of Slavin’s was that nobody was academically hurt by ability grouping — not even the lowest track students. Slavin argued that when you consider all the non-academic concerns, the scales weigh in favor of detracking, i.e. avoiding ability grouping.

What are those non-academic concerns? In the conclusion of his review of the evidence from elementary schools, he writes:

“Ability grouping plans in all forms are repugnant to many educators, who feel uncomfortable making decisions about elementary-aged students that could have long-term effects on their self-esteem and life chances. In desegregated schools, the possibility that ability grouping may create racially identifiable groups or classes is of great concern.” (p.327)

That’s Slavin’s view. So, where is the debate?

One thing that is decidedly *not* up for debate in the literature is that Slavin’s non-academic concerns are real. Opponents and defenders of tracking alike agree that low-track classes are often chaotic, poorly taught environments where bad behavior is endemic, and that this is a major problem. Tom Loveless is a contemporary defender of tracking, and writes that “even under the best of conditions, low tracks are difficult classrooms. The low tracks that focus on academics often try to remediate through dull, repetitious seatwork.” Jeannie Oakes made a name for herself by carefully documenting the lousiness of a lot of low track classes.

Some tracked schools seem to have done better with their low tracks. Gamoran, an opponent of tracking, speaks highly of how some Catholic schools handle lower tracks. Gutierrez identifies several tracked schools with strong commitments to helping students across the school advance in mathematics, and concludes that “tracking is not the pivotal policy on which student advancement in mathematics depends.” Making these experiences better is an important goal. These difficult dynamics are a genuine and widespread issue, though, and educators are rightly concerned about them.

Slavin’s concerns about exacerbating racism in schools are relatively uncontroversial as well. It’s not so much that race is a factor in track placement. Using a large nationally representative sample and controlling for prior achievement, Lucas and Gamoran found that race wasn’t a factor in track placement. (Though Dauber et al, found that race was a factor in track placement in Baltimore schools, so maybe sometimes racism is a factor in placement.)

But because of existing achievement gaps between e.g. Black and white students, there’s the potential in a racially mixed school that ability groups will effectively sort Black students into the lowest track and expose them to a lot of dynamics that are difficult to quantitatively measure but frequently discussed in education. A school where being Black is associated with poor performance and misbehavior will, according to many educators and researchers, lead to lower expectations and academic self-esteem for all Black students.

(Good news for people who like bad news: school segregation is getting worse, so the interaction between tracking and race is getting better.)

The main controversy surrounds Slavin’s claims about the academic impact of ability grouping. His meta-analyses were part of an extended back-and-forth with Chen-Lin & James Kulik, who wrote several competing analyses on the ability grouping literature. Slavin and the Kuliks each criticized the other’s methodology, but the core point the Kuliks made was that ability grouping *did* have positive effects on gifted students as long as curriculum was enhanced or accelerated to match, and that this typically *did* happen in dedicated gifted and talented programs. The Kuliks pointed out that both they and Slavin largely agreed on the data both analyzed, but that Slavin excluded studies of gifted programs from his research while the Kuliks made those studies a focus.

Tom Loveless, senior fellow at the Brookings Institution, summarized one important aspect of their dispute, which is that their debate centers more on values than their read of the extant evidence:

Slavin and Kulik are more sharply opposed on the tracking issue than their other points of agreement would imply. Slavin states that he is philosophically opposed to tracking, regarding it as inegalitarian and anti-democratic. Unless schools can demonstrate that tracking helps someone, Slavin reasons, they should quit using it. Kulik's position is that since tracking benefits high achieving students and harms no one, its abolition would be a mistake (p.17)

Betts notes the studies the Kuliks and Slavin reviewed in their meta-analyses had some flaws, with relatively small N and non–nationally representative data. Using more nationally representative samples, a number of researchers (Hoffer, Gamoran and Mare, Argys, Rees and Brewer) came to the conclusion that conventional tracking benefits students in the high tracks and hurts those in the low tracks. But it's really hard to control for the right factors in these definitely non-experimental studies, and Betts and Shkolnik raise questions about the results of these papers. And there was also a recent big meta-meta-analysis that found no benefits for between-class grouping, echoing Slavin, but that did find benefits for special grouping for gifted students, echoing the Kuliks.

Just to mess with everybody, Figlio and Page argue that by attracting stronger students to the school (because parents seek tracking) students in low-tracks benefit, secondarily.

So, in summary, what should we make of all this? Betts, an economist, says in a review of the literature that when it comes to the average impact of tracking or the distribution of achievement "this literature does not provide compelling evidence." Loveless doesn't disagree, but notes that for high achievers, the situation is clearer:

“The evidence does not support the charge that tracking is inherently harmful, and there is no clear evidence that abandoning tracking for heterogeneously grouped classes would provide a better education for any student. This being said, tracking’s ardent defenders cannot call on a wealth of research to support their position either. The evidence does not support the claim that tracking benefits most students or that heterogeneous grouping depresses achievement. **High achieving students are the exception. For them, tracked classes with an accelerated or enriched curriculum are superior to heterogeneously grouped classes.**” (p.22)

At the end of the day, **all academic impacts of tracking are mediated by teaching and the curriculum.** If a teacher doesn’t change what they teach or how they teach it, no grouping decision will help or hurt a student academically in a significant way. Tracking only *could* benefit gifted students if it came with some sort of curricular modification.

This is a conclusion with wide-reaching support. Even Slavin, who so staunchly opposed conventional ability grouping, was extremely impressed by something called the Joplin Plan, which involves three core features:

- Grouping students based on reading ability, regardless of grade level
- Regular testing and regrouping of students on the basis of the tests
- A different curriculum for each group of students

Slavin, the Kuliks, and everyone else seemed to agree that students in the plan — at all ability levels — tended to get 2-3 months ahead of students in typical programs over a year of instruction. The Joplin plan involves ability grouping — the good kind of ability grouping.

So in 1986, when the Baltimore School Superintendent turned to Bob Slavin to design a program that would improve the city’s most dysfunctional schools, guess how Slavin grouped students?

Slavin worked with research scientist Nancy Madden (they’re married) to design Success for All for Baltimore, and it’s a prominent program in the school improvement world, implemented in thousands of schools and spreading. Those three features of the Joplin plan — assessment, regrouping along the lines of ability and targeted teaching — are core features of their program.

Success for All isn’t the only example of a successful curriculum implementing these ideas. Direct Instruction was created by Siegfried Engelmann and Wesley Becker in the 1960s, and it also groups students according to their current levels in reading and math while frequently reassessing and regrouping. DI has a strong body of research supporting its efficacy (for one, it was the winner of the famous-in-education Follow Through experiment), but fell largely out of favor outside of remedial classrooms. In early 2018, a new meta-analysis spanning 50 years of research reinvigorated conversation around Direct Instruction. It found an average effect size of 0.51 to 0.66 in English and math over 328 studies ($p<0.001$), — strong evidence that the program works.

While its effect on student performance is rarely disputed, the program remains controversial. Historian of education Jack Schneider writes: “Direct Instruction works, and I’d never send my kids to a school that uses it. The program narrows the aims of education and leaves little room for creativity, spontaneity and play in the classroom. Although test scores may go up, the improvement is not without a cost." Ed Realist worries that its pedagogy is unsavory, has not been shown to work for older students, that wealthier parents are voting with their feet against the curriculum, and that DI could exacerbate gaps between students. Supporters, by contrast, paint the picture of a robust, effective system that has been ignored and disregarded.

Success for All and Direct Instruction are not simple programs for schools to adopt. Implementing them amounts to a major organizational change, and pushes at the extremely resilient notion that children in school should be grouped by their ages. Comprehensive ability grouping programs such as these seem to work, but in practice they are rarely used.

Our favorite one-stop source for reading on ability grouping: here, or maybe here to get a broader picture of the controversy.

3. Acceleration

Forget the comprehensive approach, then. Does it work to simply move an individual student (e.g. an underchallenged and frustrated student) through the curriculum at whatever pace seems to make sense?

There are a few different ways schools can help some students access the curriculum more quickly. A kid can skip a full grade, or several grades in extreme cases. They can stay in their grade for some classes, but join higher grade levels for some parts of the day. They might be assigned to two classes in one year (e.g. Algebra 1 and Geometry). Or, in some cases, a young student might start school at an even younger age than is typical.

If a child is ready for a higher level within a subject and studies it instead of the lower level, it’s almost a given that they’ll learn more. The real research questions are (a) from an academic standpoint whether accelerated children *do* tend to be ready, or if they do poorly in classes post-acceleration) and (b) whether acceleration exposes students to non-academic harm (e.g. stress, demotivation, loss of love for subject, poor self-esteem).

The Study of Mathematically Precocious Youth (SMPY) is an ongoing longitudinal study examining thousands of mathematically gifted students. In one SMPY study, researchers compared the professional STEM accomplishments of mathematically gifted students who skipped a grade to those who remained at grade level. They found that, controlling for a student’s academic profile in a pretty sophisticated way, students who skipped a grade tended to be ahead of the non-skippers in terms of degrees earned, publications, citations accrued, and patents received. From this work it seems skipping a grade in the SMPY cohort did nothing to hurt a kid’s learning or enthusiasm for their passions.

Acceleration has been one of the focuses of SMPY studies. A 1993 piece about SMPY findings reported “there is no evidence that acceleration harms willing students either academically or psychosocially.” This is supported by various meta-analyses, going back to the 1984 Kulik & Kulik paper and confirmed by more recent work such as a 2011 analysis of existing studies. Beyond the “does no harm” findings, these meta-analyses also report academic benefits to students.

It can be confusing, when reading these studies, to keep track of just *how* gifted the students happen to be. For example, SMPY has studied five cohorts so far, ranging from students who assessed in the top 3% to those who assessed in the top 0.01%. As we consider students farther away from the mean of achievement, the need for acceleration becomes more acute.

Lots of teachers encounter “1 in 100” students every year, but the education of “off the charts” students is necessarily more a matter of feel than policy. Still, there are success stories to learn from, and they show a remarkable sensitivity to both the academic and social well-being of the student.

Terence Tao is a famous success story of this kind. He surprised his parents by discovering how to read before turning two, and as a child he started climbing through math at a blistering rate. He was identified as profoundly gifted from a young age, and his education was carefully tracked by Miraca Gross as part of her longitudinal study of profoundly gifted children:

His parents investigated a number of local schools, seeking one with a principal who would have the necessary flexibility and open-mindedness to accept Terry within the program structure they had in mind. …

This set the pattern for the ‘integrated,’ multi-grade acceleration program which his parents had envisaged and which was adopted, after much thought and discussion, by the school. By early 1982, when Terry was 6 years 6 months old, he was attending grades 3, 4, 6 and 7 for different subjects. On his way through school, he was able to work and socialize with children at each grade level and, because he was progressing at his own pace in each subject, without formal “grade-skipping,” gaps in his subject knowledge were avoided.**“**

His education continued in much the same fashion, culminating in a Ph.D. by the age of 21 and a remarkable and balanced life since. He has since given his own advice on gifted education.

Given the success of acceleration, are we accelerating enough? On the one hand, it appears that acceleration is a widely used tool for giving gifted students what they need. When looking at the top 1 in 10000 students in terms of mathematical ability as identified by the SMPY, nearly half of the group skipped grades, and almost all of them had some form of acceleration, whether that meant advanced classes, early college placement, or other tools. About two-thirds reported being satisfied with their acceleration, rating it favorably across many categories:

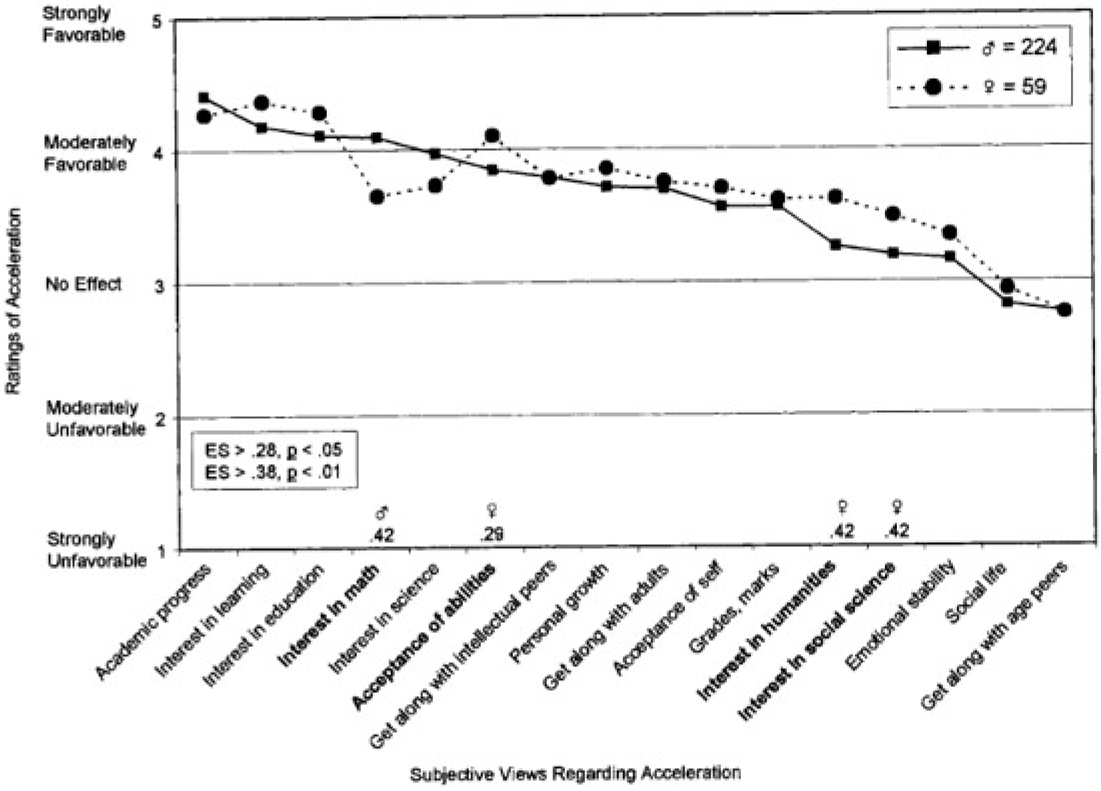


Figure 2. Participants' subjective views regarding acceleration. ES = effect size.

[Source: SMPY]

The dissatisfied third of those 1 in 10000 students, for the most part, reported wishing they had been offered *more* acceleration. And advocates for gifted education strongly endorse the notion that acceleration is under-used. A Nation Deceived is premised on this idea — though besides for "more" the report doesn't get specific concerning how many students *ought* to be accelerated, and the report mostly makes a cultural argument in favor of acceleration, citing stories like Martin Luther King Jr. graduating high school at 15.

We wanted to know more about how educators think about acceleration, so we surveyed (via twitter) twenty-one teachers, academic coaches, tutors and administrators. The survey prompted educators to respond to the following scenario:

In your school there is currently a 1st Grader who does math above grade level, e.g. he performs long division in his head. His parents initiated contact with the teacher after hearing their child complain that math at school was boring. They're concerned that he isn't being challenged. The classroom teacher knows that he is above grade-level in math, and is trying to meet his needs in class. The parents, however, do not think the current situation is working. The teacher reports that the student is difficult to engage during math class, and that sometimes he misbehaves during math.

From their responses, it certainly seems that acceleration was on the table, but almost always the last option after a number of in-class or non-classroom options (e.g. after school clubs) were explored. That acceleration in math should be a "break in case of emergency" response is also the line offered by the National Council of Teachers in Math:

tracking is morally indefensible, acceleration should be viewed with suspicion but can sometimes be appropriate.

In many ways, mainstream education is living in Bob Slavin’s world. He was a leading opponent of tracking, but was impressed by certain forms of ability grouping. He took the research on ability grouping that actually works (through assessment, frequent regrouping, and curricular modification) and used it to create a program for failing schools. He expresses suspicion about acceleration of gifted students in general, but agrees that at times it is a useful and necessary tool. If you broach the conversation about acceleration with your child’s teachers, you might hear some version of Bob Slavin’s take.

There is more to say about where this skepticism comes from. But it’s important to note that just because a student *could* be accelerated doesn’t always mean that they *should*. While some gifted students fit the profile we sketched above — frustrated with school, bored and underchallenged, and finding it hard to connect to peers — many equally capable students are happy in their school lives. (We heard some, but not many, happy stories from online commenters.) If a child is happy and successful without acceleration, they are likely to remain happy and successful regardless of whether they are accelerated, and if they don’t want to accelerate, it should not be forced on them. At least some of the suspicion towards acceleration comes from parents who inappropriately push schools to accelerate their happy, satisfied children.

Acceleration is also not the only option. There is much more to learn than is taught in regular courses. Even in a normal class, a well-designed curriculum or an experienced teacher can create “extensions” to the main activity, so that students who are ready for more have something valuable to engage with. Enhancement or exposure to new, similar topics can serve students as well. A student who has jumped ahead in arithmetic may be entranced by a glance at Pascal’s triangle and number theory. One who is fascinated by English might find similar joy in learning Spanish or Chinese. Both of these, alongside acceleration, follow a simple principle: if a child wants to learn more and is able to do so, *let them learn more.* Overall, the balance of evidence suggests that acceleration is a practical and resource-effective way to help gifted, underchallenged students flourish in schools.

Our favorite one-stop source for reading on acceleration: here.

4. Educational Goals in Conflict

Through acceleration, tutoring, or ability grouping, some kids could learn more. Why aren’t schools aggressively pursuing that? Shouldn’t they be working to teach kids as much as possible? Isn’t that what a school supposed to do? That educators are skeptical of ability grouping or acceleration can be maddening from the perspective of learning maximization: *Why are schools leaving learning on the table?*

Here’s something we don’t talk about nearly enough: schools are simply *not* in the learning-maximization business. It turns out that parents, taxpayers and politicians call on schools to perform many jobs. At times, there are trade-offs between the educational goals schools are asked to pursue, and educators are forced to make tough choices.

Historian David Labaree has one way of thinking about these conflicting educational goals, which he expands on at length in *Someone Has to Fail*. For Labaree, there are three competing educational goals that are responsible for creating system-wide tensions:

- democratic equality (“education as a mechanism for producing capable citizens”)
- social efficiency (“education as a mechanism for developing productive workers”)
- social mobility (“education as a way for individuals to reinforce or improve their social position”)

As Labaree tells it, these goals end up in tension all the time. A lot of things that seem like gross ineptitude or organizational dysfunction are really the result of the mutual exclusivity of these goals:

These educational goals represent the contradictions embedded in any liberal democracy, contradictions that cannot be resolved without removing either the society’s liberalism or its democracy … We ask it to promote social equality, but we want it to do so in a way that doesn’t threaten individual liberty or private interests. We ask it to promote individual opportunity, but we want it to do so in a way that doesn’t threaten the integrity of the nation or the inefficiency of the economy. As a result, the educational system is an abject failure in achieving any one of its primary social goals … The apparent dysfunctional outcomes of the school system, therefore, are not necessarily the result of bad planning, bad administration, or bad teaching; they are an expression of the contradictions in the liberal democratic mind.

Ability grouping and acceleration fit nicely within the tensions Labaree exposes. These learning-maximizing approaches could find support from those who see education as a national investment in our defense or economy. Of course, the strongest demand for acceleration in schools can come from parents, who want schools to give their children every possible opportunity to be upwardly mobile. ("We want to make sure they can go to a good college.")

Those act as forces in favor of ability grouping and acceleration. But schools also know that they are held responsible for producing equitable outcomes for a citizenry that sees each other as equals. A program that raises achievement for top students without harming others has an appeal an economist could love, but within schools this can count as a problem.

The way this plays out in practice is that many schools are inundated with requests to accelerate a kid. Parents — especially financially well-off, well-connected parents — can typically find ways to apply pressure to schools in hopes of helping their children reach some level of distinction. They'll sometimes do this even when it wouldn't benefit a child's education (it would be educationally inefficient), or when it would exacerbate inequality (by e.g. letting anyone with a rich, pushy parent take Algebra 1 early).

In short, from a school's standpoint those are two problems with acceleration. First, parents will push for it even when it's not academically or socially appropriate. Second, it can exacerbate inequalities. That could explain where the culture of skepticism within education comes from.

This is meant entirely in terms of explaining the dynamic. The way this plays out can be incredibly painful. Systems designed to moderate parental demand can keep a kid in a depressing and frustrating situation:

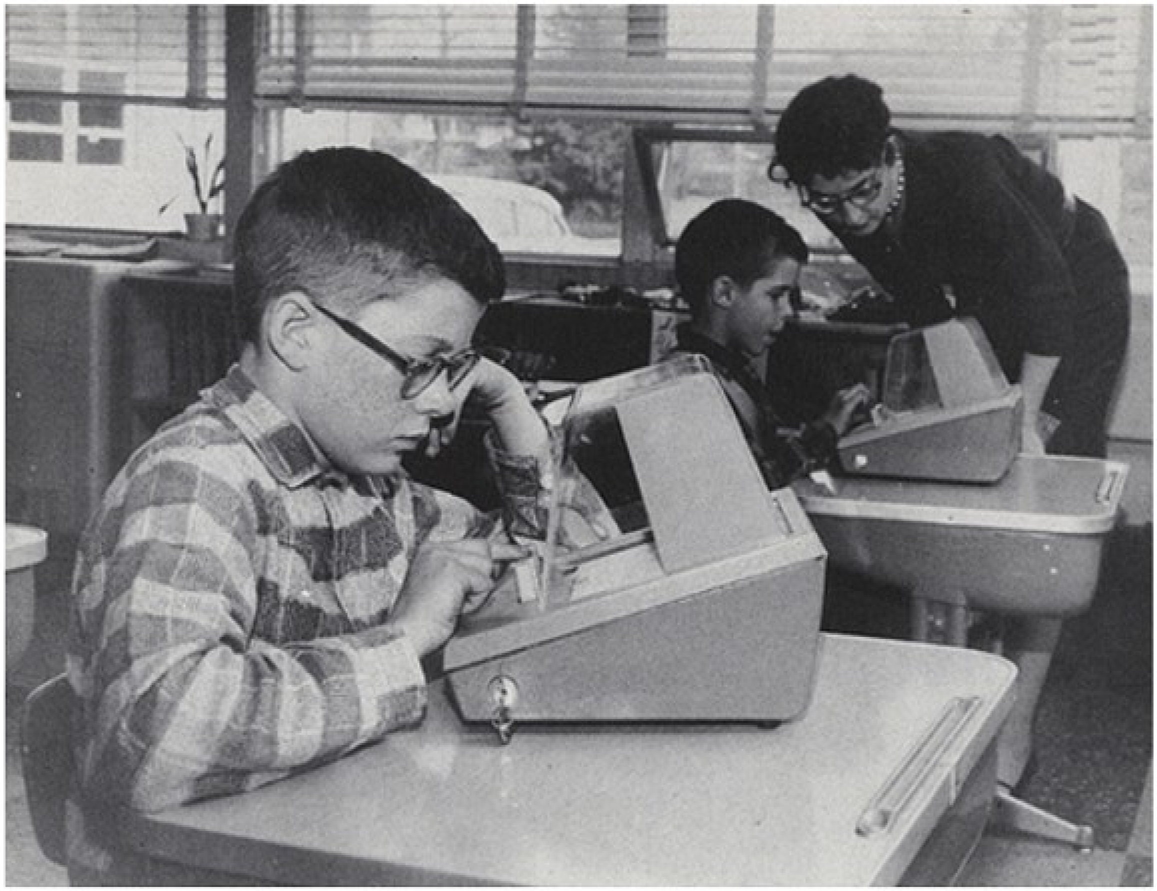
My older son wanted to move up to a more advanced math course for next year. He took two final exams for next year's course in February and answered all but 1/2 of one question on each. So roughly 90% on both and his request to skip the course was denied. (source)

Districts sometimes have extensive policies that can be incredibly painful to navigate when trying to get a student who *truly* needs acceleration out of a bad classroom situation. We heard from one educator who had a very young student expressing suicidal ideations. It was all getting exacerbated by the classroom situation — the kid said he felt his teachers and peers hated him because he loved math. The parents and the educator tried to find a better classroom for the child, and were met with all the Labaree-ian layers of resistance. Off the record, the educator advised the parents to get out of dodge and into a local private school that would be more responsive to his needs.

A happy ending: the 4th Grader moved to a private school where he was placed in an 8th Grade Honors class. He likes math class now. He seems happier, he's growing interested in street art and social justice work.

But without a doubt, there are some unhappy endings out there.

5. Personalization Software



[source: Larry Cuban]

“Ours is an age of science fiction,” Bryan Caplan writes in *The Case Against Education*. “Almost everyone in rich countries — and about half of the earth’s population — can access machines that answer virtually *any* question and teach virtually *any* subject … The Internet provides not just stream-of-consciousness enlightenment, but outstanding formal coursework.”

The dream of using the Internet to replace brick-and-mortar classrooms is a dream that is entirely in sync with the times. This is reflected in the enormous enthusiasm directed towards online learning and personalization software. Bill Gates, Elon Musk, and Mark Zuckerberg have all invested heavily in personalization and teaching software. And the industry as a whole is flush with funding, raising some 8 billion dollars of venture capital in 2017, while reaching 17.7 billion in revenue.

Finally — a way out of the school system and its knot of compromises! If schools are institutions whose goals are in tension with learning-maximization… then let's stay away from schools and their tensions and give the children the unfettered learning they want. Let's create the ideal tutor as a piece of software.

This dream isn't just in sync with our times — it has a long history. This history is particularly well-documented by historian Larry Cuban (author of *Teachers and Machines* and *Tinkering Toward Utopia*) and by Audrey Watters (she's writing a book about it). Watters' talk "The History of The Future of Education" is as good a representative as any of the major thesis: that the dream is larger than any particular piece of technology. Motion pictures, radio, television, each of these was at times promoted as an educational innovation, able one day to free students from lockstep movement through school and into a personalized education. From Thomas Edison to B.F. Skinner, tech advocates have long envisioned the future that (at least according to Caplan) we're living in now.

Then again, tech advocates in the past also thought they were living in the age of personalized learning. In 1965, a classroom that used a program called Individually Prescribed Instruction was described this way:

Each pupil sets his own pace. He is listening to records and completing workbooks. When he has completed a unit of work, he is tested, the test is corrected immediately, and if he gets a grade of 85% or better he moves on. If not, the *teacher offers a series of alternative activities to correct the weakness, including individual tutoring.*

For comparison, here is the NYTimes in 2017, and the headline is A New Kind of Classroom:

Students work at their own pace through worksheets, online lessons and in small group discussions with teachers. They get frequent updates on skills they have learned and those they need to acquire.

The similarity between modern day and historical personalization rhetoric doesn't settle the matter — in a lot of ways, clearly the Internet *is* different — but personalization software seems to have arrived at a lot of familiar, very human frustrations.

Anyone who has gone online to learn has, at some point, come face to face with this dilemma: On the internet, you can study almost all human knowledge, but usually you don't. In a world with virtually every MIT course fully online for free, a world with Khan Academy and Coursera and countless other tools to aid learning, why has the heralded learning revolution not yet arrived?

In a way, the revolution *has* arrived — it just hasn't improved things much. Rocketship Schools, a California charter using online learning for about half of its instruction, has had solid results. Lately, though, they've moved away from some of their bigger bets on personalization and rediscovered teachers, saying *"We've seen success with models that get online learning into classrooms where the best teachers are."* School of One was a widely hyped high school model in NYC that was preparing to scale up its offerings… until a fuller picture of the results came in and it was pilloried. Online charter schools, meanwhile, seem to actively depress learning.

Part of the problem is that it’s hard to get solid research on the efficacy of various ed tech products. Many tools, particularly those sold directly to schools or used by online charters, are proprietary and stuck behind paywalls, selectively presenting their best data and limited demos. The ed tech sector in general seems to deliver mixed results to students.

Why is it so hard to make effective teaching software?

For one, teaching is complex. A good human teacher does a lot of complicated things — gets to know their students, responds to the class’ moods and needs, asks “just right” questions, monitors progress, clarifies in real time as a look of confusion dawns on the class, etc., etc. — and it’s simply hard to get a computer to do that.

Maybe, theoretically, a piece of software could be designed that does these things. But in practice, many software designers don’t even try. It’s easier and cheaper to make pedagogical compromises, such as providing instruction entirely through videos. Yes, there are some thoughtful tools made by groups like those at Explorable Explanations, such as this lesson on the Prisoner’s Dilemma. But building high-quality tools well-adapted for a digital environment is difficult and time-consuming, and for prospective designers, destinations like Google or Blizzard tend to be more glamorous than working with schools. In practice, humans currently have a lot of advantages over computers in teaching.

Even if we overcame all the design issues, though, would students be motivated to stick with the program? Studies of online charters point to student engagement as the core challenge. When you put a kid in front of a computer screen, they jump to game websites, YouTube, SlateStarCodex, Google Images — anything other than their assigned learning. Many educational games that try to fix this resort to the “chocolate covered broccoli” tactic, trying to put gamelike mechanics that have nothing to do with learning around increasingly elaborate worksheets.

To be fair, student engagement is also the core challenge of conventional schools. But that’s precisely what the much-maligned structures of school are attempting to confront. The intensely social environment helps children identify as students and internalize a set of social expectations that are supportive of learning. The law compels school attendance, and schools compel class attendance. .And, once a child is in the classroom, their interactions with actual, live human instructors can set high academic expectations that a child will genuinely strive to meet.

The conventional story is that school is incredibly demotivating, but compared to their online counterparts schools are shockingly good at motivation. MOOCs like those on Coursera have an average completion rate of 15 percent — public schools do much better than this. Popular language app Duolingo’s self-reported numbers from 2013 would put their language completion rate at somewhere around 1%. If all a user has to rely on is their daily whim to continue a course, the most focused and conscientious may succeed, but those are the ones who already do well in schools. That’s a big part of why people lock themselves into multi-year commitments full of careful carrots and sticks to get through the learning process. Writers such as Caplan think that people are revealing their true interests when they skip learning to fart around on the web, but we might as well see a commitment to attend school as equally revealing. People need social institutions to help do things we'd truly like to do. As such, even as computers become better teachers, the motivational advantage of schools seems likely to persist.

How might tech-based learning tools address these factors, so they might stand a chance at holding students' attention long enough to teach them? *Art of Problem Solving,* an organization promoting advanced math opportunities to children, makes a good case study. It's found a balance worth examining. First, it provides accessible gamelike online tools that center on a careful sequence of thought-provoking problems. Second, it offers scheduled online classes with the promise of a fast pace, challenging content, and a peer group of similarly passionate students taught by subject matter experts. The online classes are more expensive offerings, but they preserve the human touch.

What does that balance mean for students? If they're in the conscientious, self-motivated crowd that wants to learn everything yesterday, they can gorge themselves on software designed to be compelling. No barriers keep them from progressing. Software can always point to a next step, a harder problem. On the other hand, if they want to lock a motivational structure around themselves and keep the social benefits of school in a more challenging setting, they can.

Not every successful tool need look identical, but that core idea is worth repeating: software should enable the passion and self-pacing of eager kids, but should not rely on that to replace the power of social, *human* motivational structures. Yes, sometimes even the same structures used in "regular" schools.

Online learning, then, fits squarely within the history of attempts to automate teaching. Over and again we make the same mistakes and forget the lessons of history: that teaching is more complex than our machines have ever been, that motivation is largely social, and that schools will have a hard time distinguishing between altrustic designers and opportunistic profit-seekers.

For those in the market for online learning there are a lot of mediocre tools available, and many truly bad ones. Right now, there's nothing that seems ready to serve as a full-on replacement for school without consistent, careful human guidance.

That said, depending on your passions, there *are* some excellent resources for learning out there. Especially if a student has a caring mentor or a passionate peer group, they can learn a lot online. As educators and designers create more tools that respect both the power and limitations of machines, that potential can grow. But it's not quite science fiction.

Our algorithm has determined that you should watch the following two videos: here and here to balance realism and idealism

6. Practical Advice

Education is complex and resists easy generalizations. That said, here are some generalizations.

On navigating school for your child:

- The brightest students do *not* thrive equally in every setting. Even the best students achieve more with teachers than on their own. Unless tutoring or some other private arrangement is possible, this means that a school is the best place to be for learning.

- But school right now doesn't work for all kids. One fix: if a child wants to be accelerated and seems academically prepared for it, acceleration will usually help them.
- Most schools aren't in the business of maximizing learning for every student, and in particular they tend to be skeptical of acceleration.
- Therefore: If your kid needs more than what school is offering, be prepared to be a nudge.
- But if you think your kid needs to be challenged more and your kid is perfectly happy in school, try really hard *not* to be a nudge.
- Don't fight to move your child to a class that covers the exact same material at the exact same pace but has the word "Honors" next to it. That sort of ability grouping makes no educational difference.
- Prioritize free, open online tools. Don't expect online tools to do the work for you or your child. Expect more distraction and less progress if online learning time is unstructured or unsupervised.

For educators:

- If you are an elementary teacher or administrator and your school is looking to try new things, consider cross-grade ability grouping by subject, especially in math and reading.
- Gifted kids are usually not equally talented in all fields. Consider options to accelerate to different levels in each subject based on demonstrated skill *in that subject.*
- A lot comes easily to smart kids, and sometimes they never get the chance to learn to struggle. Find *something* they think is hard, academic or not, so they are able to handle more important challenges later.

- If a child is bored in your class and knows the material, they probably shouldn't be in your class.

For tech designers and users:

- If you're making online tools, make the learning the most interesting part of them. Don't rely on chocolate-covered broccoli or assume that just presenting the material is enough. Take the problem of motivation seriously.
- Look for passionate groups with robust communities, whether online or offline. Don't overlook the social aspect of learning.

And for advocates of educational reform, in general:

- People almost only talk about educational efficacy. But don't be fooled — educational debates are only sometimes about what works, and frequently about what we value.

One last thing: if you're an educator or a parent or just somebody who spends time around children, take their feelings seriously, OK? If a kid is miserable, that's *absolutely* a problem that has to be solved, no matter what district policy happens to be.

Acknowledgments: Thanks to /u/Reddit4Play from reddit, JohnBuridan from the SSC community, blogger Education Realist, and many others who read drafts and offered ideas along the way.

[ACC Entry] Are Islam And Liberal Democracy Compatible?

.

[This is an entry to the Adversarial Collaboration Contest by John Buridan and Christian Flanery.]

Matter: To what extent does liberalism and democracy obtain in Islamic countries. Whether Islam consistently poses political opposition to liberalism and democracy.

Two simple narratives have split the western world's perspective on Islam.

These two narratives do not exhaust the spectrum of opinion, but they do function well enough to establish the basic controversy around Islamic countries and Liberal Democracy.

The first narrative opines that Islam is an ideology inimical to "western values," such as classical liberalism and liberal egalitarianism, and a rival to the Judeo-Christian social mores. It constitutes an ideological rival, inherently aggressive, both unable and unwilling to sustain non-partisan legal systems, democratic norms, fair treatment for opposition parties, protection of dissidents, or the basic rights and freedoms which Western European and Anglophone countries enjoy. And that Islam sustains this undesirable state of affairs.

The second is that Islam is not qualitatively different from any other religion. Islam has contributed to civilization in a significant way, and ordinary Muslims share our own values of family, peace, and justice. In contrast to the first narrative which stresses Islam as an ideology, the second narrative emphasizes that Muslims are normal people. There is no problem with Islam eo ipso; the perceived "problems" of Islam are actually some combination of the fairly normal problems of traditional societies, poor socio-economic conditions, and legacy problems from colonialism.

In order to avoid a point-scoring debate between these two narratives, our approach is to provide a descriptive examination of the performance of liberal democracy within Islamic environments. We take as granted for this paper that one cannot look at a religion on paper and predict what it will look like in a polity. Religious practice and theological doctrine inform every aspect of the pious person's outlook and life, but the way in which it informs that outlook is not deterministic and cannot be gleaned merely by looking at the source texts, nor by the impossible task of a quantitative comparison of which religion has produced more violence across regions and millenia. Although we believe original texts are not deterministic, that does not mean Islam is totally amorphous. Religious culture is a powerful force within society. It unifies people, allows them to feel part of something bigger and better, it provides solace in their troubles, and can mobilize political action. How that mobilization of power occurs remains largely up to the needs of the moment, but it's that mobilization of power which we are interested in.

A community's interpretation of a religious text can be unpredictable, and our study does not hold such texts as a reliable source for predicting political outcomes. Nor will we attempt to determine the nature of Islam by enumerating the good and bad works it has produced. We hold that investigating the politics of a particular community over time more adequately casts light on the possibilities for the future than the foundational texts of the culture.

Our methodology is to investigate the recent political history of Islamic countries as they relate to democratic forms of government and the package of rights we call ‘liberalism.’ We survey the national history, constitutions, and current political environments to determine the extent to which democracy and liberalism have obtained in these countries, and we predict the conditions under which more democratic and liberal policies could emerge. Islam is a useful study since certain well-known expressions of Islam are decidedly neither democratic nor liberal. The “Islamic world” is one of the largest populations groups in the world, and so each Islamic country’s relationship and experimentation with democratic forms of government and human rights is important for the future.

First, we will define our basic terms.

Democracy – A system of government which rejects the rule of a single interest (dictator or oligarchs) through the participation of citizens and some separation of powers.

Liberalism – In an effort to avoid taking sides between classical and egalitarian forms of liberalism, we are just emphasizing the basic liberalism of texts like “The Declaration of the Rights of Man and the Citizen”, “The Declaration of the Rights of Women”, and a generally consistent protection of these rights by the police and judiciary.

Islamic country – for us any country with a Muslim population of 70% or higher. We use the phrase Islamic country even when the country in question is officially nonsectarian. The six countries we study are United Arab Emirates, Tunisia, Indonesia, Kazakhstan, Iran, and Lebanon (see the section on Lebanon for more on our reasons for choosing this country). We chose these countries for their geographic range, cultural diversity, and relative stability. These countries can act as a representative sample for future inquiry on politics in Islamic countries.

In each case we examine the history, constitution, and current political environment, focusing on democratic mechanisms and human rights records.

We stress constitutions because they are useful, tangible indicators for political outcomes. However, constitutions can be misleading, and so reviewing the context and results of implementation is essential in making sure our moorings are in the political world and not hypothetical jurisprudence.

United Arab Emirates

The United Arab Emirates (UAE) is a sovereign nation state founded in 1971, and composed of the lands of seven Emirates. An Emirate, is the rank, land, or reign of an Emir. Emir is the patriarch of a tribe and the leader of an Emirate. Emir’s are also referred to as Sheikhs at times, which simply means leader of a tribe. The Emirs and their family rule each Emirate and power is passed through male successors. The 1971 unification was the first time the Emirates were united into a formal state. Despite the constraints of the constitution, each Emirate still enjoys a degree of autonomy within the confines of the union.

History

In 1971, the seven gulf emirates unified themselves into one sovereign state known as the United Arab Emirates (UAE). This was in response to the withdrawal of the British Protectorate over their provinces in 1968. The UAE adheres to a long standing cultural tradition rooted in ancient tribal systems. This tribal tradition, and a 500 year old geopolitical strategy of precarious orientation between the Islamic world and western power are the determinants of the UAE’s place in the contemporary international order.

In order to be a citizen of the UAE, one must prove a direct ancestry dating at least as far back as 1930. This dictate resulted from fear in the 1940s that tribal relations would be diluted in the wake of increasing immigration caused by oil discovery. The tribes that compose the UAE are Arab. Some tribes have occupied the strait of Hormuz and Persian gulf coasts since the days of Alexander the Great’s conquests. Others migrated out of the Arabic desert in the 1700s and settled parts of the coast.

Arabic tribal arrangements served the needs of a people in a desolate landscape where living on one’s own meant death. Because resources were scarce, tribes hoarded them jealously and placed a premium on tribal solidarity. This meant discouraging inclusion of outsiders and prohibiting membership in the tribe other than through birth. This mentality has played a role in discouraging the promotion of democratic norms. Citizenship is denied even other Arabs who do not have direct tribal links, and a patriarchal kinship system orders politics, not democracy nor kingship.

The tribes were never formally unified until 1971. Historically, when security concerns arose, they often operated in loose confederation with one another in order to protect mutual interests. In the 16th and 17th centuries the emirates entered agreements with Portugal to protect their trade routes stretching from Swahili in East Africa to India and beyond. These were especially lucrative on account of demand for pearls in Europe. This system was renewed under the British protectorates of the 19th and 20th century. In 1968 the British withdrew their protectorate, prompting the Emirates to form a unified state. The newly created UAE then established security agreements with the U.S. similar to the client relationships they had with the Portuguese and the British.

Today, longstanding squabbles between the Emirates still influence relations. Even at the inception of the UAE, Bahrain and Qatar were originally destined to join the union. On account of old territorial disputes between the emirates, the two decided to establish their own independent states. These long standing conflicts continue to assert themselves as we have seen in the 2017 severing of ties between the Gulf Cooperation Council (GCC) and Qatar.

In addition to tribal disputes, the UAE fear a continually more bold and expansionist Iran. The Emirate’s historical relationships with western powers remain the primary means of securing themselves in an often volatile and unfriendly region. Consequently, the UAE’s close ties with the U.S. constitute a critical security lynchpin if they are to remain autonomous and unharrassed. This necessitates paying some degree of lip service to American values, as we will see below.

Constitution

The UAE constitution was ratified in 1971. It represents a middle of the road effort by the ruling Emirs to cater to western democratic liberalism while retaining as much of their central power as possible under the traditional patriarchal kinship system. As shown in the history section, the UAE clings to its tribal systems and asserts them against efforts to dilute tribal ties. Democratic instruments such as universal suffrage and the power of collective, popularly elected legislative bodies threaten these modes of rule.

In 2004, a number of amendments were made to the constitution. The preamble in this version, stakes out a commitment to a Westphalian[1] conception of the state and a commitment to democratic principles and process.

> Desiring to create closer links between the Arab Emirates in the form of an independent, sovereign, federal state, capable of protecting its existence and the existence of its members, in cooperation with the sister Arab states and with all other friendly states which are members of the United Nations Organization and members of the family of nations in general; on a basis of mutual respect and reciprocal interests and advantage;...[and] preparing the people of the Union at the same time for a noble and free constitutional life, **progressing by steps towards a comprehensive, representative, democratic regime** in an Islamic and Arab society free from fear and anxiety;

In the latter half of the 20th century, it became critical that a nation obtain for itself a seat within international institutions. The most common avenue to prosperity and security for emerging states today is through the United Nations and the Bretton Woods Institutions. Access to these international institutions require formal statehood and are not bilateral in nature. Therefore, membership comes with a greater set of demands and requirements. The U.N., World Bank, and IMF are institutions largely oriented at the discretion of western nations and insist on democratic liberalism. The commitments to liberal democracy instantiated in the UAE constitution, as quoted above, serve to keep the UAE in good standing within the international community, maintain access to key international financial and monetary instruments, and facilitate security relationships. There is little substantiating evidence that the Emirates intended to adhere to these commitments to liberal democracy lest their international ties are threatened. This strategy is protected in the constitutional preamble where it states that movement towards a democratic order is, “progressing by steps.” It is not democratic yet but the implication is that it will be eventually. By stating that the UAE will move incrementally towards democracy, the Emirs can justify perpetuating oligarchical rule.

The constitution does contain a number of liberal protections and commitments. It contains non-discrimination clauses regarding race, nationality, religious belief, and social position, but there are no clauses for gender or sexual orientation. It also guarantees free speech, but as we will see below governmental protection of free speech has been dubious according to U.S. reporting agencies.

The arrangement and structure of government bodies provide minimal checks on arbitrary decision making on the part of the Supreme Council which is the highest authoritative body, composed of the seven Emirs. The Supreme Council selects judges and does not require confirmation of its picks by the legislative branch. They unilaterally select the president from their own ranks. The Council of Ministers and its chairman are selected from the population by the Supreme Council. No confirmation is required. The only semblance of popular election based self rule lies in the Union National Council, an advisoral legislative body which is composed of 40 legislators. 20 of them are popularly elected, and 20 are chosen directly by the Supreme Council.

The Cabinet of Ministers submit draft legislation which passes through the Union National Council and then to the Supreme Council for ratification. The legislative power of the Union National council is advisoral only. The Supreme Council can ratify any draft legislation without the approval of any other body or individual. Only the seven Emirs may challenge the constitutionality of laws, which is then determined by judges who the Supreme Council selects. Only the Emirs may submit amendments to the constitution, which require two thirds (30 out of 40, 20 of whom are chosen by the emirates) support in the Union National Council.

Frequently, the constitution reiterates that all bodies and individuals must be beholden to the constitution as the ultimate authority. Shariah is mentioned only once but does exercise considerable authority over the primarily Muslim population.

The preamble commits to democratic rule and the constitution as the foundational binding document, but in effect, it is only a reformulation of pre-existing tribal systems. It masquerades as an effort to create a fair and balanced system of governance, but in reality, it enshrines almost all executive, legislative, and judicial power in the hands of the seven Emirs. This document cannot be considered democratic because it establishes almost no democratic mechanisms, and it lacks important key guarantees, such as gender discrimination and due process, in order to be considered liberal.

Democratic Institutions and Human Rights

Citizenship and Voting

The Emirs select who is allowed to vote and the criteria for selections is not clear or published. It seems a very small number of citizens are allowed to vote. Perhaps 10% of the population of 5 million. In 2015 voter turnout was 35.29% of those eligible to vote. Roughly 11% of the resident population are citizens and 85% of these are Sunni Muslims. In 2010, Pew Research estimated that 76.95 of the total population were Muslim. Obtaining citizenship, and thereby voting rights, requires proving direct tribal lineage prior to 1930. Thus, democratic participation is accessible to a narrow sliver of society. Citizens who do vote, can only vote for a consultative counsel that is not vested with any real legislative power. Some of the Emirates have at times created parliaments with real legislative power but still retain the right to dissolve these at will, which they have.

Constitutional Adherence

The U.S. State Department has reported on the UAE's many unconstitutional activities. For further reading on these, see the U.S. State Department's Country Reports on Human Rights Practices and International Religious Freedom Reports.

Article 26 of the constitution forbids torture. Article 30 ensures freedom of expression. Article 28 lays out juridical procedures to ensure defendants receive a public and fair trial.

In the State Department Human Rights Report for 2017, it notes that U.N. human rights experts working with former detainees in the UAE have repeatedly reported the use of torture techniques such as forced standing, threats to rape or kill, and electrocution. Shariah courts also impose flogging for pre-marital sex, defamation, and consuming alcohol.

Activists have alleged that authorities detained citizen Ghanem Abdullah Matar after he posted a series of videos on social media in June that expressed sympathy with Qatar, with whom the UAE are in a bitter dispute over an email hacking scandal. Ghanem, in the video here, criticises the UAE for it's hard stance on Qatar and for not remembering Qatar's aid in the conflict in Yemen.

Fair public trials have also been withheld from individuals considered to be potential terrorists or who criticize government policy. The location and status of political activist Nassir bin Ghaith remains unknown from the time of his arrest in August 2015 until April 2016,

“when prosecutors formally announced charges of defaming a foreign country (Egypt), criticizing the UAE’s decision to grant land for a Hindu temple, and having ties to Islamist group al-Islah. In December 2016 bin Ghaith’s case was transferred from the State Security chamber of the Federal Supreme Court to the Federal Court of Appeal. In March authorities sentenced bin Ghaith to 10 years in prison for promoting “false information in order to harm the reputation and stature of the state and one of its institutions.” Bin Ghaith called into question the fairness of his trial, noting an Egyptian judge was assigned to adjudicate a case that involved charges of bin Ghaith defaming Egyptian figures.”

Nassir bin Ghaith did not receive a fair, public trial. The State Department a number of other such cases which strongly indicate that public criticism of the government is likely to result in arbitrary imprisonment.

These are just three areas where the unconstitutionality of acts are disregarded in the name of expediency, national security concerns, or as a relinquishing of authority to shariah courts. The Emirs, despite the existence of binding constitutional law, can still act unilaterally and arbitrarily, regardless of the rights of their citizens or of foreign nationals in the country.

UAE’s Future

Change in the UAE looks highly unlikely in the near future. It was untouched by the Arab Spring, and long standing patriarchal systems are firmly in place. I should note, that while the “incremental” approach to democracy that the Emirs articulate is certainly suspect, (See Cleveland and Bunton’s A History of the Modern Middle East in suggested reading for a good examination of Britain;s use of the same method of maintaining unilateral rule and suppressing democracy in its colonial holdings in Egypt) conditions have improved for religious freedom and female freedom in the UAE over the last 50 years relative to before the unification.

Until the Emirs are stripped of some of their religious and political power it is unlikely things will change any faster than the current snail speed. The UAE is wealthy and its stability is of critical geo-strategic importance to U.S. and western regional concerns who are loathe to demand too much of the Emirs in regards to political change.

In order for the patriarchal system to be supplanted, citizenship would have to be radically expanded, and real legislative power delegated to the Union National Council. As it stands, the Union National Council can tentatively maintain current conditions from worsening, but it is unlikely that it can alter the existing governmental mechanisms in any significant way.

Tunisia

History

After the fall of Rome, the region of modern day Tunisia was sporadically conquered and settled by European tribes until the Muslim conquest of the 7th century transformed the culture and population of the region. It become a province of the Ottoman Empire in the 16th century

In 1881, Tunisia was wrested from the hands of the declining Ottoman Turkish empire and became a French protectorate. In 1956, Tunisia declared independence, adopting a constitution akin to France’s highly centralized presidential system.

This planted the seeds for Tunisia's becoming one of the few Arab Spring successes in 2011. Tunisian presidents previously could extend their time in office indefinitely. Between 1956 and 2011 there were two presidents. Thus, in the intervening years between independence and the 2011 Arab Spring, Tunisia was an authoritarian presidential system with few functioning democratic mechanisms. Opposition parties were banned until 1981 and even then had little or no influence over the reigning party.

Despite this, Tunisia was relatively progressive under its two presidents, Bourguiba and Ben Ali. Bourguiba focused on the status of women and social and economic development and education. Ben Ali continued this and made some constitutional changes (something separation of powers should prohibit, but on the other hand he did it in the name of fomenting greater democratic participation) that included allowing opposition political parties, and presidential term limits.

Consequently, Tunisia entered the 21st century with greater literacy rates and social safety nets than its neighbors, Algeria and Libya, and many of its peers in the post colonial, Muslim world. While it was socially egalitarian, it was still not democratic. Ben Ali later re-imposed the ban on opposition parties and in conjunction with secret police forces, suppressed political dissent whenever his party became threatened. Cue the Arab Spring, which in Tunisia was known as the *Jasmine Revolution*, or as *The Revolution for Freedom and Dignity.* Political corruption, media and internet censorship, and high unemployment continued to mount in the 21st century and precipitated a waning of public faith in Ben Ali, that turned into protests in 2011 which the government suppressed violently. Eventually, Ben Ali was forced out and a new struggle emerged between secular and Islamist groups. A moderate Islamist group, Nahdah, functioned as the moderator between secular minded groups demanding a democratic and liberal regime like Ben Ali's and hardline Islamists pushing for a government founded on Sharia principles.

After a series of negotiated compromises, a new constitution was drafted in 2014 that intended to resolve and balance liberal concerns, and the role of Islamic Shariah within the newly established Tunisian state.

The 2014 elections were a resounding victory for the secular Nidaa Tounes coalition who won 85 of 217 seats compared to the moderate Islamist Nahda party who won 69. This was a critical consolidating juncture for Tunisia's nascent democracy. The world held its breath waiting to see if the Islamists would concede and legitimize the elections. When Nahda conceded and no challenge came against the results, Tunisia's democracy became consolidated. This is critical sign of how robust the democratic regime is as Nahda still has significant sway and influence and continues to cooperate and work within a democratic framework and does not challenge mechanical pillars like election results, and universal suffrage. It is a powerful nod to Islam's potential to interface with liberal democratic systems. Since 2014, Nidaa Tounes has consolidated more control over parliament and this correlates with increased radicalization amongst Tunisian citizens as the government appears more and more to be unfaithful to the country's Islamic heritage.

As we saw with the UAE, Tunisia's leaders, before and after the *Jasmine Revolution,* saw close relations with Western partners, in this case France and the U.S., as critical for their continued independence and success. Unlike the UAE, the commitments to liberal democracy promulgated in the constitution are manifested faithfully and are reflected in the 2014 election results. As we will see next, the Tunisian constitution commits to the American constitutional model and even corrects what many political scientists see as deficient in the American model.

Constitution

Tunisia’s 2014 constitution is strongly nationalist and Muslim. It asserts Tunisia’s independence as a sovereign national state, and also as an integral member of the Arab Maghreb of North West Africa. It’s religion is Islam, as stated in article one, and the people see themselves as part of the Islamic world.

The preamble lays out quite clearly the intent and values of the new government:

“With a view to building a republican, democratic and participatory system, in the framework of a civil state founded on the sovereignty of the people, exercised through the peaceful alternation of power through free elections, and on the principle of the separation and balance of powers, which guarantees the freedom of association in conformity with the principles of pluralism, an impartial administration, and good governance, which are the foundations of political competition, where the state guarantees the supremacy of the law and the respect for freedoms and human rights, the independence of the judiciary, the equality of rights and duties between all citizens, male and female, and equality between all regions,”

Here is a link to a pdf of the Tunisian constitution. We recommend reading the entire preamble. It is quite beautiful, and we wish we had space to include it.

This paragraph alone includes almost everything a high functioning constitutional democracy requires in order to accomplish egalitarian, pluralistic, self rule. Participation, unbloody alternation of power, free elections, separation of powers, pluralism, fair political competition, human rights, independent judiciary, and female suffrage and freedom enshrined in a political document: the bedrock of western democracy. Having a document that is worded in a manner that leaves little legal wiggle room, is a key ingredient for continued success after a revolution that overturned 60 years of presidential authoritarianism, but this does not guarantee perfect adherence to the norms established in the constitution nor does it mitigate the strong backlash Tunisia has seen from radical Islamists who have significant influence over large portions of the population.

Within this new constitution, Tunisia addresses several issues that have been a thorn in the side of the American democracy. The constitution establishes term limits for legislative representatives as well as for judges. It also imbues a dedicated constitutional court of 12 judges with the power to resolve budgetary and other constitutional problems that arise between the legislative and executive branches. Each government branch selects four constitutional court judges, ensuring the balance of powers over constitutional issues. Protected from corruption and elite capture, this democratic mechanism facilitates smooth government operation and resolves gridlock between the legislative and executive branches. Something that would be a much appreciated sight in the U.S., instead of watching legislators filibuster on the senate floor for days on CSPAN about their favorite ice cream, all because Republicans and Democrats are unwilling to cooperate. Tunisia does not know the joy of watching a filibuster on CSPAN.

Democratic Mechanisms

Of the Muslim countries that experienced rapid revolutionary change in the wake of the Arab Spring of 2010 and 2011, Tunisia was the only one to establish high functioning democratic mechanisms. Tunisia quickly did away with the autocratic power of the President and his unilateral control over security forces and the secret police. They managed to implement a new government without state collapse.

Tunisia has successfully passed Huntington's two-turnover test with successful alternations of power between 2011 and 2018 from president Fouad Mebazaa of the Neo Destour party, to president Moncef Marzouki of the Congress for the Republic party, and finally with the incumbent president Beji Caid Essebsi of the Nidaa Tounes party. This is something easier said than done, especially in a country deeply divided between secular and Islamic traditionalist legacies. In the U.S., political participants understand that, through many iterations of power turnover, it is necessary that the output of the democratic process not be challenged. We saw how robust this is in the U.S. when the Supreme Court settled the crisis that arose over electoral recounts in Bush v. Gore (2000) as well as the brute fact of nearly 250 years of unbloody power transfers. In many parts of the world this sort of crisis results in the losing party not conceding. Then the tanks roll, and democracy dissipates. It is a highly illusive collective mobilization problem, but Tunisia transcended it.

For Tunisia this indicates widespread agreement and assent to the rules of democratic process, regardless of the outcome that the freedom machine spits out, and such agreement and trust is always fragile after revolutions. Security dilemmas often arise after a revolution because groups do not trust one another to not hijack elections and consolidate their own power over the ensuing government. For all groups to simultaneously relinquish the right to challenge results *ex post facto* posed incredible collective coordination and communication problems for the emerging Tunisian government between hardline secularist and hardline Islamist Salafist groups. This was achieved by the moderate Islamist party Nahdah successfully balancing the security concerns of Salafist Muslims and the secular Nida Tounes party before the 2014 elections. They maintained this role even after losing in the elections.

The judiciary seems to function independently of the executive and legislative branches and the constitutional court continues to perform its functions over disputes unharassed.

The U.S. State Department reports that Tunisia has also taken significant strides against government corruption.

Human Rights

As Tunisia consolidated and refined democratic processes, it incurred more resistance from the extreme religious right. Secularists continue to increase their seats in Tunisia’s unicameral legislative chamber. As this has occurred, government crackdowns on extremist activity have grown more and more willing to favor national security concerns at the cost of human rights and judicial process. Many perceive this as a liberal takeover intent on stamping out Tunisia’s cultural heritage and Islamic legacy.

Tunisia is currently one of 70 countries where homosexuality is illegal. Efforts from the Tunisian human rights commission are moving forward to revoke this and other Shariah based laws, such as female inheritance laws and the death penalty.

As these political shifts occur, they conversely produce an increase in Islamic radicalization. Thus, Tunisia, in its efforts to stomp out extremism and facilitate a pluralist society as stated in its constitution, has resorted to various methods of indiscriminate arrests, censorship, and unfair trials.

The U.S. State department reported that, “In May [2017] a court in Tozeur sentenced two journalists – a brother and sister – to six months in prison for “insulting a public servant” after they criticized security forces for regularly raiding their home, allegedly on suspicion their sibling was affiliated with extremist religious groups.” This is one amongst a number of indiscriminate, unilateral government action against suspected, but unproven, Islamist sympathisers. See the State Department report for more such instances.

“The government may initiate administrative and legal procedures to remove imams whom authorities determine to be preaching “divisive” theology.” and , “The penal code criminalizes speech likely “to cause harm to the public order or morality,” as well as acts undermining public morals in a way that “intentionally violates modesty.” These stipulations constitute a powerful scepter in the hand of the Tunisian government to dictate public discourse. It is difficult to say what the next administration might interpret as “divisive theology”.

There are numerous reports of religious profiling of Salafist Muslims on account of their traditional dress and beards and restriction of movement for Salafists within the country, “Since 2014 more than 500 individuals filed complaints with the Tunisian Observatory for Rights and Freedoms, saying the government prevented them from traveling due to suspicion of extremism, and in some cases apparently based on their religious attire. The media also reported police and security forces harassed some women who wore the niqab.

Tunisia’s Future

Currently, Tunisia is a liberal and democratic government that, out of fear of radicalization of its citizens and for its security, has marginalized and isolated a large portion of the religious right through extrajudicial means. Tunisia boasts of thwarting some 12,000 ISIS recruits from leaving the country since 2013. This is not reassuring.

It is not hard to understand why Tunisia's rationale for these miscarriages of justice, considering that Tunisian born individuals are overwhelmingly represented in recent terrorist attacks, including the Charlie Hebdo and Nice France attacks, and among ISIS recruits. Radicalism is the greatest threat to an emerging democracy in an Islamic country aiming to prove to the West that Islam is compatible with western modes of governance. It is incumbent on the leadership of Nidaa Tounes and its current Tunisian President, Beji Caid Essebsi, to find a way to avoid using human rights violating and unconstitutional means in order to maintain national security. If the religious right is more isolated on the margins, it will undoubtedly be the unraveling of the democratic project in Tunisia.

As things stand at the time of this writing, the problem of resolving the dichotomy between liberal democracy and traditional Islam has not been so much resolved in Tunisia as it has been shunted aside, and therefore, thrives underground, providing an ever replenishing pool of candidates for radicalization.

Indonesia

Indonesia has a strong democratic constitutional platform but is deadlocked by corruption. The two political visions offers solutions to this "moral decay": military-based nationalism and more explicit Islamic rules. Although Islamic political parties field candidates and do well electorally, Pancasila remains an important governing ideology for the archipelago, and no party can escape paying some lip-service to it.

History

Indonesian nationalism which began in the earliest years of the 20th century did not have a strong religious component. Christianity, Islam, and Hinduism had majorities in certain provinces of the archipelago, but Islam was the most dominant religion in terms of numbers. However, the Islam of Indonesia did not coalesce into a credible political force until recently. You will see why shortly.

Indonesian independence activists in the early 20th century had their sights set against the Dutch colonizers, who kept the Indonesian governments under their mercantile thumb, employing a race-based caste system. After 1900 early nationalists took up the term 'Indonesia', originally an ethnographic term coined by northern Europeans, and used it to raise awareness about the possibility of a unified nation. Through the 1920s groups formed using the name Indonesia and other groups incorporated 'Indonesia' into their title. Additionally, youth movements, socialist and communist parties, and Islamic financial solidarity groups formed to resist colonial exploitation. Into this caldron of budding nationalism, socialism, communism, Islamic consciousness, anti-colonialism, and national consciousness stepped the imperial Japanese occupying force.

The Japanese occupation did not maintain a light touch. The abuse of islanders in many different principalities impressed the need for a strong independent Indonesia upon the above groups; religious leaders, scholars, and activists mourned the occupation and abuse. During the end days of Japanese power, communist, Islamic, and nationalist resistance organized. Two days after Japan's surrender to the United

States on September 2, 1945, Indonesia declared independence.

The years following the declaration of Independence were fraught with guerilla warfare between the nationalist government, European colonizers, communist resistance groups, and (by the 1950s) Islamist resistance groups. The resulting political ideology of the state was called Pancasila.

Constitution

Pancasila (or “The Five Principles”) run as such:

1. Belief in the One and Only God
2. A just and civilized humanity
3. A unified Indonesia
4. Democracy, led by the wisdom of the representatives of the People
5. Social justice for all Indonesians

What Pancasila means in practice for liberalism is that public atheism, agnosticism, and animism are illegal, and that outspoken criticism of the government can be considered and offense against the 3rd principle – a unified Indonesia. The leaders of Indonesia safeguard the country’s unity by force of arms and sometimes violent suppression. The central government goes to some lengths to maintain independence from “Great Powers,” so that their military forces can be concentrated on combating problems within the archipelago. Additionally, free market economics ought to be tempered by social welfare programs, and programs for mutual aid (these programs for mutual aid are very important in contemporary Indonesia, since in the present day explicitly Islamic institutions have created nearly all social safety nets).

The Indonesian constitution looks and reads like a liberal democratic constitution with some important idiosyncrasies. The strong centralized executive branch takes the lead over other branches and acts as balance against against the forces of decentralization and democratic impotency. We call this authoritarianism. While arguments about decentralization and federalism were at the heart of American politics in the early days of the U.S.A, for Indonesia the constitution provides the executive all the powers needed to ensure unity through force, which occurs regularly. Apparently archipelagos tend to resist centralization.

Articles 27 through 34 guarantee a cornucopia of human rights. The list is impressive and beautiful and a distant ideal. For our purposes I will draw your attention to Article 29:

Chapter XI Religion Article 29

(1) The State shall be based upon the belief in the One and Only God.

(2) The State guarantees all persons the freedom of worship, each according to his/her own religion or belief.

After the death of Suharto in 1998, a new debate emerged to define the role of Islam in the nonsectarian constitution. The options on the table for the revision of Article 29 of the constitution tell the story of the interests of political actors involved.

- The state is based on Belief in One Almighty God
- The state is based on Belief in One Almighty God with the obligation upon the followers of Islam to carry out Islamic law

- The state is based on Belief in One Almighty God with the obligation upon the followers of each religion to carry out its religious teachings
- The state is based on Belief in One Almighty God, Just and Civilised Humanitarianism, the Unity of Indonesia, Democracy Guided by the Wisdom of Representative Deliberation, and Social Justice for all Indonesians

The first option, original to the constitution, stuck.

The third option could lead to religious courts for all religious believers. In Indonesia there are only six officially recognized religions (Islam, Protestantism, Catholicism, Hinduism, and Confucianism); ID cards state religious identity on them. A system of personal status courts would require courts for each sect and would diminish the power of civil courts, leaving them to decide family cases outside the ken of the religion in question. Opening the door to a justice system with 6 different legal codes would not, we reckon, foster stability.

The fourth option reiterates the principles of Pancasila signaling recommitment to the Indonesia's ideological roots.

The second option, of course, is the Islamic one which would enshrine in the constitution the role of Islam for believers and allow sharia law to be a basis for future lawmaking. The government in that case would obtain the potential to legislate away some of the autonomy and rights of non-Muslims and non-Muslim communities which are protected by Pancasila. Furthermore, the need for this law is suspect. In 1998 Law No. 7 gave Muslims the right to have personal status cases heard in Islamic courts if they wish. The push for the change to federal law revealed a possible future for Indonesia that would step decidedly away from the nonsectarian values of the state.

Although no textual change came of this constitutional debate, the lack of change affirmed the same non-secular, nonsectarian Indonesia which preceded the debate. This is good. Perhaps it even strengthened Indonesia's political system. Nonetheless for us the disagreements illuminated schools of thought among the politically active in Indonesia.

Democratic Institutions

In the past 8 months over 300 elected officials have been implicated in corruption scandals. Corruption in the judiciary, civil service, and police force runs rampant. Efforts to crackdown on corruption are ongoing, as are counterterrorism efforts within the country. Between terrorism and corruption of elected officials, little faith remains in the power of democracy and liberalism to solve the problems of security and prosperity for Indonesia.

Businesses in Indonesia frequently have to grease the wheels of bureaucracy with bribery, and the cost of doing business is highly variable. During Suharto's regime corruption was centralized and predictable. Decentralized corruption on the other hand makes transaction costs unpredictable, thus business has a more difficult time now than in authoritarian times. Additionally, corruption among the police and judiciary exacerbate costs, since the justice department can't always be relied on to prosecute crimes efficiently or effectively.

Islam

Up to 2,000 Indonesian citizens have fought for ISIS. As these fighters return home, some seek rehabilitation, some carry out terrorist attacks, some promote Islamic State political visions. The government's antiterrorism task force releases commercials trying to debunk the glorious Islamic

State. Some returnees had to be threatened before they would swear allegiance to Pancasila; according to C-SAVE director Maria Kusumini, 90% of returnees want to live under an Islamic caliphate. While that is not an enormous percent for a country of 261 million, it’s still enough to cause problems for society.

Suharto brooked no opposition in his ~50 years of power. He restricted freedom of speech and reacted with military force against both Islamic and communist separatists. Debate about *shari’ah* in Indonesia was not permitted until 1998 with Suharto’s death. The many years that political Islamic expression sat impotent allowed the energies of Islamic organizations to flow into local communities of financial aid, health, and support. These organizations make Indonesia a strong country, but in the future their renewed political character might repeal and replace Indonesia’s founding principles.

Kazakhstan

The Republic of Kazakhstan is not an Arab country. Kazakhs are Turkic people, and their conversion to Islam did not come until well after the Mongol invasion (13th c.). Islam came late to Kazakhstan (15th c.) and conversion was a slow and inconsistent process for the traditionally nomadic people.

Eventually, Islam did spread to the itinerant people overtaking their traditional beliefs. The Hanafi school of jurisprudence remains the dominant school of Islam in Kazakhstan largely, some Kazakh scholars say, because it allows friendly relations with non-Muslims. Also note that Arabic Islamic traditions never became widespread among the tribes. The nomadism and lack of education ensured that the expression of Islam would be less tied to written interpretations, and more adapted for life on the road, while incorporating some syncretic elements from the traditional paganism of the region.

As happens on the steppe, slowly over the course of the 19th c. the great bear to the north, Russia, absorbed its unwilling southern neighbor, and thus Kazakhstan became part of the Russian Empire’s geopolitical strategy . Then came the Bolshevik revolution which first suppressed religion, then used some elements of the Islamic governance structure. The effect was a decimation of Islamic culture, traditional culture, and nomadism, balanced by Russification, secularization, and various other Moscow initiatives for economic modernisation. Results over the 20th c. were… mixed.

Independence came in 1990 and with it a resurgence of Kazakh population, decrease in Russian population and influence, return to various cultural expressions, namely Islam, and a new capital, Astana.

Post-Soviet Political Expression

Kazakhstan has many trappings of a Soviet state. The President holds power as an autocrat, not merely one actor among others in government. In fact, Kazakhstan has had the same president since 1990. The nation is officially secular, allows authentic freedoms, guarantees protection of citizen rights, but betrays such constitutional promises through corruption in law enforcement and the judiciary and suppression of political dissidents and “nontraditional” religious groups. The Constitutional Court has failed to rein in latitudinous interpretations of the constitution such as Article 5.3:

Formation and functioning of public associations pursuing the goals or actions directed toward a violent change of the constitutional system, violation of the integrity of the

Republic, **undermining the security** of the state, inciting social, racial, national, religious, and tribal enmity, as well as formation of unauthorized paramilitary units shall be prohibited.

Or Article 20.3

Propaganda or agitation for the forcible change of the constitutional system, violation of the integrity of the Republic, undermining of state security, and advocating war, social, racial, national, religious, and clannish superiority as well as the cult of cruelty and violence, shall not be allowed.

The result of these statues is the heavy monitoring of all assemblies, especially religious assemblies. "Nontraditional religions" are censored. Missionaries and religious groups need state approval in order to operate. But the Jehovah's Witnesses seem to always be finding themselves on the wrong side of the law, as do Baptists, and Muslims which do not subscribe to Hanafi Sunni Islam, or who read or are found in possession of "dissident" materials. At times Baptists have had difficulty operating in the country without being accused of unlicensed evangelism. Private religious reading is carefully watched by the state and is considered a national security issue overriding the "right to confidentiality."

For political opposition, this spells doom. Opposition leaders, activists, and journalists find themselves fined, imprisoned, and exiled. Even academics, according to the U.S. State Department (see link above), self-censor for fear of "infringing on the dignity and honor of the president and his family."

There are few prospects for Kazakhstan to become a more liberal and democratic society any time soon. Instead, by the example and logic of its neighbors (the other four -Stans, China, and Russia), authoritarian power and pervasive opportunism preserve the government of Kazakhstan.

Islam in Modern Kazakhstan

The suppression and control of Islam during Soviet rule further isolated the country’s culture and reduced the transfer of Islam to the next generation. As *glasnost* gave some wiggle room for religion in Kazakhstan, Islam and Russian Orthodoxy reestablished themselves. The Christian population (mostly Russian Orthodox) declined from 46% in 1994 to 26% in 2018 due mostly to Russian and other Slavic emigration. Islam gained momentum in the 90s going from 44% of the population in 1994 to 70.2% in 2018. Only around 5% percent of this change is from Islamic population growth.

As Islam has grown, government initiatives to control Islamic expression have grown too. One of the reasons for this is education. Better educated Muslims are more informed about the rest of the Islamic world and thus more uncomfortable with Kazakh Islamic idiosyncrasies and syncretisms. Social movements inspired by the more politically charged Islam of Afghanistan, Iran, and Indonesia has led to government crackdown. Islamic cultural symbols such as the hijab are banned in schools, dressing in an Arabic Islamic way can be punished, and many Islamic organizations are banned, especially Salafist adjacent movements. Meanwhile the government is also taking positive steps to use the muftiate of Kazakhstan to define a specifically Kazakh Islam. Whether Kazakhstan can create an effective state-sponsored Islam, the way a country like China has created a state-sponsored Catholicism, remains to be seen.

Sadly, human rights seem to be an afterthought in Nazabayev’s government, as political dissidents are shutdown, both in the streets and online, thanks to China-esque internet surveillance. While Islam is treated as part of the national heritage, any expression, or suspected expression of Islam deemed too extreme gets the accused convicted of political dissidence. And similarly, political dissidents can be labelled extremist Muslims and thus convicted. While some see in this repression a potential for backfire against the government, more likely, the coercive binding of the Overton Window will serve its purpose and reduce dissent and the public sphere, so long as Kazakhstan's government can merely cite national security concerns to work its will.

Islam is not a political force in the country. You would think that a Muslim majority country would have a more "Islamic character," but Kazakhstan's political history puts religious expression in a context of post-Soviet authoritarian secularism. The Islam of Kazakhstan is young and develops along government initiative; public political religion is just not possible. Kazakhstan is secular, just not in a way that a human rights activist would celebrate.

Iran

History

Iran experienced a number of conquests and periods of cultural enrichment and decline between the conquest of Alexander the Great (330 B.C.) and the Safavid Empire (1501-1732), which ruled roughly the same geographical parameters as modern day Iran.

The 7th century Islamic conquests altered Persian society dramatically, and permeated Persian culture deeply, competing with longstanding Zoroastrian and Byzantine Christian cultures. Not long after Islam established itself in Iran, the Shiite and Sunni schism occurred over the prophet

Mohammed’s heirs. This would prove to be the proverbial line in the sand, setting the stage for a millenia long regional split between the two Islamic schools. This dynamic fundamentally informs many of today’s conflicts in the Middle East. See the conflict in Yemen and the Iran-Iraq war (1980-1988).

We saw in Tunisia that postcolonial, secular authoritarian leaders were able to successfully implement industrial and political modernization over the latter half of the 20th century without inciting backlash from Islamist traditionals. A similar effort began in Iran in the early 20th century with the Constitutional Revolution of 1905-1907, Iran’s first efforts with democratic norms of popularly derived rule and self identification as a sovereign nation state, something ever illusive in Iranian history. It collapsed quickly into chaos. After Russian intervention in Northern Iran threatened the stability of the country, Britain switched support from the Constitutionalists to the Shahs, who were the ruling monarchy before the revolution. Reza Shah seized power in 1921, reestablishing monarchical rule with significant control over parliamentary systems.

Reza Shah embarked on a modernization campaign with ambitions to make Iran a regional power and globally competitive economic entity, free from foreign interference by Russia and Britain. Reza Shah also dramatically increased the centralization of the state and strengthened the military.

This process came to a head with the 1951-1953 oil crisis, known as the Abadan crisis, when Iran nationalized the Anglo-Iranian-Oil Company (AIOC) assets. While this did prompt a British and U.S. supported coup d’etat which ousted democratically elected Prime Minister Mohammad Mosaddegh in favor of consolidating the monarchical rule of Mohammad Shah, Iran had vast undercurrents of political disaffection as economic modernization efforts dating back to 1925 continued to leave behind rural Islamic communities. Known as Operation Ajax, the CIA fomented support for Mohammad Shah and organized guerilla forces to discredit Mosaddegh's government. Iran was primed for this operation to work, as Mossaddegh's socialist rhetoric alienated Iran's conservative Islamic population.

In 1979, Ayatollah Ruhollah Khomeini executed a successful Islamic revolution that ousted the American backed government in favor of a new theocratic government which is still in power today and is of rising importance in Middle Eastern affairs.

Radical Islamic Ideology

Ernest Gellner's theory of *Folk-Islam* and *High culture-Islam* provide a compelling explanation for the causal sources of the Iranian Revolution of 1979 in addition to the coercive effects of interventionist U.S. foreign policy.

Gellner distinguishes between *High-Islam* and *Folk-Islam.* High Islam is deeply austere, literate, and connected to a fundamentalist reading of sacred texts. It is often associated with more elite society in urban areas. Folk-Islam on the other hand, is grounded more in community and local tradition within the illiterate rural population. Folk-Islam is analogous to the nomadic and moderate Islam that we discussed as the prevailing style of Islam in Kazakhstan. In Iran, Gellner asserts, the influx of Folk-Islamic peoples into urban areas as a result of rapid industrialization in the mid-20th century, where High-Islam predominated, in conjunction with aggressive interventionist foreign policy of the U.S. and Britain in the 1950s, produced perfect storm conditions in 1979 for Grand Ayatollah Ruhollah Khomeini and a coalition of conservative clerics to incite a popular uprising and seize control of an unstable country that had been abused by foreign powers and had a deeply dissatisfied lower class.

The ensuing Islamic Republic of Iran identifies less as a sovereign nation state, and more as an ideological cause whose purpose is to expand the umma (the community of all Muslim believers), to all corners of the world, eventually absorbing all the lands of the infidels. This obviously has posed serious problems to the international order. Iran does not play by Westphalian rules of statehood and self determination. It sees itself as the vanguard of the divinely ordained mission to establish the reign of God on earth.

Hassan Abbasi, head of the Iranian think tank, associated with the Revolutionary Guard (which is under the personal command of the current leading Ayatollah), Center For Borderless Doctrinal Analysis, argued that the Islamic Republic could not be safe unless it persuaded other Muslim nations to take the same path. “If we remain alone, we will always be in danger,” he said. “Our system will also be in danger if most Muslim nations take the path of Western-style democracy.” Abbasi also asserted that Iran’s mission overrides borders and state sovereignty which are, in his estimation, “colonial inventions”. Khomeini spoke often of the need to export the revolution to other Muslim countries and to liberate Palestine and Jerusalem. These notions are fundamental in Iranian foreign policy today.

Constitution

The introduction and preamble to Iran’s constitution reads more like propaganda than a legally binding document. There is much talk of saints and martyrs’ blood, as well as the imposition of deity-like characteristics to the Ayatolla,

The honorable source of emulation, the great leader of the global Islamic Revolution, and the founder of the Islamic

Republic of Iran, the venerated Grand Ayatollah, Imam Khomaini, may his noble character be sanctified, who was acknowledged and accepted by the undisputed majority of the people as the marja'and the leader. Iranian Constitution Article 107

The Ayatollah holds the final decision over all jurisprudential matters, is commander of the military and security forces, as well as the Islamic Pasdaran Revolutionary Corps. He is the only individual who may amend the constitution with a two thirds confirmation from the Islamic Consultative Assembly. The assembly does wield limited legislative powers, but not on anything that, "Contradict[s] the canons and principles of the official religion of the country or the constitution. The Guardian Council is responsible for the evaluation of this matter." Article 72. The Ayatollah heads the Guardian Council. Thus, the delegating of any executive, legislative, or judicial power to anyone but the Ayatollah is always qualified in a manner that allows the Ayatolla to frame any issue as, "Contradicting the canons and principles of the official religion" in fact, give him unassailable and arbitrary control over government.

Some democratic principles are enshrined in the constitution, such as popular secret elections of legislatures and of the president. Political parties are allowed to exist on the same basis as legislation, so long as it does not contradict the religious canon. Again, they are flimsy.

The constitution also benevolently condescends to return women to their true dignity in replacing them upon the resplendent and exalted pedestal of true womanhood, in their calling to motherhood.

Women are emancipated from the state of being an "object" or a "tool" in the service of disseminating consumerism and exploitation, while reclaiming the crucial and revered responsibility of motherhood and **raising ideological vanguards**. Women shall walk alongside men in the active arenas of existence. As a result, women will be the recipients of a more critical responsibility and enjoy a more exalted and prized estimation in view of Islam.

Iran Constitution, page 6.

Why is it acceptable for men to be “tools” and “objects” of exploitation and consumerism, but not women? The consequence of this is that women are relegated to child bearing and home-managing responsibilities only.

These points indicate a constitution that is neither democratic nor liberal.

Democratic Mechanisms

Political parties are allowed to exist so long as they do not challenge religio-political dogma. Roughly 11 parties exist that command any significant following including a reformist party, but this does not mean that a wide window of public discourse exists. The spectrum of political opinion is severely limited by the broad terminology in the constitution as to what is acceptable in public debate, and the Ayatollah’s power is arbitrary as to determining what is religiously or politically admissible. Thus, the democratic nature of party politics and free elections is misleading at first take. In reality, political opinion is dictated almost entirely by the Ayatolla and the clerics.

Reformists?

Iran’s best hope for change, since we are assuming it would be a toss up over which would be worse between maintaining the status quo and a regime change operation carried out via a U.S. led coalition, is The Council For

Coordinating the Reforms Front organization. Led by former president Mohammad Khatami (1997-2005), who instigated intense internal debate during his presidency between his administration, and the conservative clerical elites and the Ayatolla. Khatami led reform efforts during his presidency with two government reform bills intended to limit the power of the Guardian Council to select candidates for political office as well as limit other constitutional powers of the judiciary. The Guardian Council is selected by the Ayatolla and is responsible for investigating infractions against religious and political doctrine. Every time a reformist minded candidate arises, the Guardian Council hammers them with claims of kafir and they are banned from running. Khatami’s two bills would retract the power of the Guardians to ban candidates as well as allow the executive to reprimand constitutional bodies like the Guardian Council and the Experts (a judicial body body charged with selecting the Ayatolla) for arbitrary decisions. Both bills failed under the pretext that this was a power grab by the executive, instead of an intended balancing of power between government branches. They were also deemed unconstitutional by the Guardian Council. After 2005, the Reform effort was deemed dead by reform theorists like Saeed Hajjarian.

Clerics have continued to win most political battles since the early 2000s as they have a firm grip on the lower class poor, but with a rising middle class, some reformists maintain optimism of pushing reforms through in the near future. In 2017, 28% of Iranians identified as leaning Reformist. In comparison, 15% identified as leaning Principlist. Despite this, a large number of reform activists wrote an open letter to Khatami demanding a “Reform of the reform”. After the demonstrations earlier this year yielded no apparent results, younger Iranians are beginning to see the reformists as merely a political party that has been assimilated into the Iranian clerical authoritarian structure.

Conclusion/Future Considerations

The only possible imminent change to the government in Iran would be via external imposed regime change. Reform efforts are easily suppressed through the legitimate mechanisms instantiated in the constitution. The Ayatolla is unlikely to be pushed to a point that might force him to do something unconstitutional, because he has all of the tools he needs in the constitution already. He is able to maintain this control while still allowing for regular assembly elections. It’s the holy grail of authoritarian control. The constitution allows the unitary leader various avenues of venting and mitigating efforts for change, while still refreshing the active participation of civil society in government decisions and keeping them engaged in the revolutionary cause.

Lots of think tanks like the Atlantic Council and other western publications love to talk about how much simmering discontent there is lying beneath the surface in Iran, but I do not buy this. This form of radical Islam has co-opted democratic mechanisms in such a manner that a less than pareto-optimal equilibrium has formed such that the democrat instruments themselves are used to prevent improvement in democratic norms and human rights.

Lebanon

The big story in Lebanese political society is the multiconfessional state. By law the President is always a Maronite Catholic, the Prime Minister a Sunni Muslim, the Speaker of Parliament a Shiite Muslim, the Deputy Speaker and Deputy Prime Minister are Greek Orthodox. The ratio of

Christians to Muslims in Parliament until this past election was – by law – 1:1.

This internal sectarian balance of power system is described everywhere as ‘fragile.’ The evidence for this fragility is the protracted 15 year civil war which tore Lebanon apart until 1990 and sears into Lebanese political consciousness today, and the brute fact that a proper census has not been conducted since 1932. The prospect of a census sends horror down the spines of every religious and political leader, for whatever the outcome of such a census, it could throw open the floodgates of renewed sectarian demands followed by violence.

The Sectarian System

Although this does not sound like a promising start for a democracy, the state certainly functions within the basic scope of a democracy: elite capture and tyranny are not a concern. In fact, the sectarian balance-of-powers structure maintains basic democracy through its quotas and coalitions. Government power comes from the consent of the sectarian leaders who can work to come together, address the latest crisis, and ultimately, play fast and loose with the constitution to solve the needs of the moment.

France created Lebanon out of Greater Syria in 1926; from this original constitution the confessional balance-of-power system was born. However, the French ensured that western leaning Christian groups had more power in the government, since they would be more likely to favor French interests. The civil courts were based on a Napoleonic variant, and the government included a French high-commissioner who could suspend the constitution and establish direct rule.

When the Lebanese declared independence in 1943, norms needed to be established for the balance of power. These new norms were called The National Pact. The National Pact reaffirmed the sectarian status of offices: President (Maronite), Prime Minister (Sunni), Speaker of the Parliament (Shia), Deputies (Greek Orthodox), Commander of the Armed Forces (Maronite), Army Chief of Staff (Druze). It also reaffirmed the 6:5 representation of Christians to Muslims in Parliament based on the possibly manipulated, already out-of-date 1932 census. What's important to add here, though, is that National Pact required Maronites to express their identity as non-Western Arabic, and that Muslims not seek incorporation into Syria. This Christian promise of an Arab identity for the country is crucially important for its international relations to this day, while the Muslim promise helps guarantee the spirit of Lebanese independence in the region.

The Civil War 1975 – 1990 was long and we will not get into the weeds of it. However there are some important lessons about Lebanon from the civil war. The coalitions of the civil war shifted around fast, and the political situation deteriorated from political ideology to confessional lines. The "leftist" opposition was made up of a coalition of pan-Arabs, Greater Syrians, socialists, and Islamists, while the ruling right was made of nationalists, fascists, and hard-line Christians. These coalitions broke apart fairly quickly and soon all that mattered was regional control and serving the needs/prejudices of those constituents. Since regions were fairly homogenous in their religious make-up, these regions assumed sectarian agendas and biases. This led to provocations everywhere followed by Israeli invasion, Syrian invasion, and UN Military Action, Iranian supplies to Shia militias, and ubiquitous atrocities. Although commanders of the varying forces were often of the same religion, the soldiers hailed from any sect. For example, the Shiite leadership of the Amal and Hezbollah militias were often fighting their coreligionists who were under Maronite leadership.

To bring the Lebanese Civil War to a close Saudi Arabia hosted the 1989 Taif Agreement. The agreement was brokered behind the scenes by the survivors of the 1972 parliament and its Speaker Hussein El-Husseini. The purpose of the agreement was to end the civil war, bring back rule of law, and reestablish an independent Lebanon. Additionally, the agreement changed some legacy problems from the French days. Once the French high-commissioner became a non-existent position, his powers were rolled up into the presidency. This gave the Maronite president a lot of power. He could, for example, dismiss the Prime Minister if he did not get along with him, giving the Christians a lot of power. This power was revoked in the agreement and the Christian – Muslim representation in Parliament was changed to 1:1.

Over the next years, Hezbollah, with the tacit consent of the central government, tried to reclaim the south from Israeli occupation. Since Taif (up to the present) the government has tried to diminish Hezbollah's extra-governmental military operations, and lastly the country began extracting itself from rule/occupation/influence of Syria. However, since Palestinian-aligned militias controlled south Lebanon, and Syria controlled the north and east, the central government was still fairly weak and peace was uncertain. The Cedar Revolution (2005) pushed the Syrians out, and in 2018 Lebanon made credible strides into an independent future by means of its first parliamentary election since 2009.

The June 2017 electoral reforms divided Lebanon into 15 districts, lowered the voting age to 18, and allowed oversees Lebanese to vote. Here's how it worked, and by the way, this is the most complicated voting system I have ever seen:

1. Each district had a set number of seats per sectarian group.
2. Political parties field a candidate for each of those seats in a registered candidate list.
3. Citizens of the district get one vote per seat.
4. Then the candidate with the most votes fills the respective seat.

Let us give an example. In the region “Beirut II” there were 6 Sunni seats, 2 Shia seats, 1 Druze seat, 1 Greek Orthodox seat, and 1 Evangelical seat, 11 seats total.

9 political parties formed to create a list of candidates for each seat. Citizens then cast 11 votes divided by sectarian confession. In theory there could have been 99 candidates up for election in Beirut II; however dropouts reduced the number of candidates to a still staggering (by U.S. standards) 82. While Beirut II is an outlier with its glut of nine political coalitions, the national average was still 5 – 6 lists per district.

The new system requires political parties to become interreligious coalitions which support a political platform, and since it relies on citizens casting votes for people outside their sect, the principle of consensus governance is preserved while decreasing sectarian tensions. Additionally, political interest groups are incentivized to work together to create consolidated candidate lists to reduce unwanted competition.

The big news item in the 2018 election was that the majority Future Movement party got smashed and Amal-Hezbollah won big. (Amal-Hezbollah is so named because the parent group, Amal, reunited with the splinter group Hezbollah.) Looking deeper Amal-Hezbollah made coalition lists with Free

Patriotic Movement (mainly Christian center-right), Al-Ahbash (Sufi activists, because you can never have enough Rubaiyat), and Syrian Socialist Nationalists (yes, founded in 1932…). The triple decimation of the Future Movement and double rise of Free Patriotic Movement along with Amal-Hezbollah is a big move towards friendliness with Syria. At first blush this makes no sense. Isn't Syria in shambles? Why be their ally? It is important to note though that these groups which won big in the most recent election were not part of the Cedar Revolution which pushed Syria out of Lebanon. Allowing Hezbollah militias freedom to actively support Assad may be the quickest path to relieving the pressures of 1.4+ million Syrian refugees in terrible camp conditions. (Furthermore, the Greater Syria ideology or a new Nasserism might find itself back on the table.)

In the Syrian Civil, Iran supports Assad's government. They have set up forward military bases in Syria. We should not expect those to go anywhere anytime soon. While Iranian forces shuffle through Iraq to get to Syria, Lebanese Hezbollah fighters aid Assad from the east. Hezbollah's military arm is funded in part by Iran, and the Iranian interest is to create an effective anti-Israel alliance. The Israeli response to Iranian troop in Syria has been frequent rocket strikes.

The political leaders see the 2017 reforms as the first step in a series to fix the sectarian system to create a more unified country. More reforms will be on their way. The next parliamentary election should occur in 2022. If it is a success, Lebanon may take one more step into security, prosperity, and, we hope, a nonsectarian future. As it is, Lebanon has to try to preserve its internal peace in a dangerous and high variance political environment.

Rights: The Cedar Package

“Lebanon has an Arab identity and belonging. It is a founding active member of the Arab League, committed to its Charter; as it is a founding active member of the United Nations Organization, committed to its Charter and the Universal Declaration of Human Rights. The State embodies these principles in all sectors and scopes without exception.

Lebanon is a democratic parliamentary republic based upon the respect of public freedoms, freedom of opinion and freedom of belief; and of social justice and equality in rights and duties among all citizens, without distinction or preference.”

- **Preamble to the Constitution B – C**

Lebanon has not lived up to these UN notions of Human Rights. As often happens in the Lebanese constitution what is said in one part is qualified or contradicted in another. Articles 9 and 12 add important qualifiers to freedom of speech:

Article 9 “Freedom of conscience is absolute. In assuming the obligations of glorifying God, the Most High, the State respects all religions and creeds and safeguards the freedom of exercising the religious rites under its protection, without disturbing the public order. It also guarantees the respect of the system of personal status and religious interests of the people, regardless of their different creeds.”

Article 12 “The freedom of opinion, expression through speech and writing, the freedom of the press, the freedom of assembly, and the freedom of association, are all guaranteed within the scope of the law.”

“Without disturbing the public order”, “guarantees the respect of the system”, and “within the scope of the law” provide the judicial framework for blasphemy laws and prosecution for opinions outside the Overton Window. But the Overton Window of Lebanon is extremely wide, from Communists to National Socialists, from conservative Shiites to Gay Pride promoting Sunnis.

While freedom of worship is guaranteed, freedom from religious institutions is not. Lebanon has a civil judiciary for general property rights, prosecution, and so on, but what we call family law is the province of religious courts. Marriage, inheritance, child custody, annulment, divorce are all controlled by 15 or so religious courts. Most everyone is labelled a member of a sect, even if not practicing, for the purposes of family law. These laws are the strongest legal force entrenching sectarianism in Lebanon. And, unfortunately, they often relegate women to second class status.

The Right to Rock and Roll

The current picture of Lebanon would not be complete without a few words about Lebanese culture. Lebanon is one of the most culturally productive and free Arab countries. It maintains itself as a regional media capital. Rock music, cinema, and books are produced in Lebanon and exported all over the Arab world.

Seriously, the music is great!

More about Islam

Recently, interfaith marriage has become possible, and from that marriage has issued a sectless baby. The marriage required the signature of the Minister of the Interior. Small demonstrations have called for civil marriage as a civil right. Despite these developments, Lebanon is not necessarily on the way to becoming a country which accepts secularism into its internal balance.

Sunnis and Shiites are equal in number – both being about 27% of the total citizen population. Druze make up 5%, and they are grouped as Muslim for political purposes. We don’t know that they would call themselves Muslim, since they consider the Dialogues of Plato an inspired text, and look to their own sacred texts and precepts. Political battle lines in Lebanon concern Palestine and the Palestinians in Lebanon, Israel, the role of Hezbollah in society, the role of Syria in Lebanon, preservation of identity, and some type of identity politics based on the struggles of the last 30 years. The major political parties have a narrative and some vision for Lebanon, but at the same time contain members from many religious confessions.

Problems, Predictions, and Possibilities

So given all the instability in Lebanon’s history, why include it in the study? It does seem to break our rule of inclusion since its Muslim citizen population is only ~60%. However, it is an example of diversity, dynamism, political balancing, and cultural production in the midst of a weak state apparatus.

Currently, Lebanon has a citizen population of ~5 million. An additional 400,000 Palestinians and 1.5 million Syrian refugees are also within the borders with little prospects for the future. Most of these people are Muslim, and so will probably not become naturalized citizens any time soon. Currently Lebanon’s government sees hope for the return of Syrians to their home country, even as the Palestinians hang in a hopeless limbo.

For Lebanon to extend its civil rights, greater control of the border must be secured. The first requirement is the cessation of the civil war in Syria. The second requirement is a second parliamentary election which further diffuses sectarianism and entrenches democratic norms. At the local level, interfaith marriage and conversions would assist the project of desectarianism by creating more demand for civil courts to handle some family law.

Conclusions

The intersection of Islam and liberal democracy is a recent phenomenon that began with decolonization in the early 20th century. These countries ratified their constitutions between 4 and 75 years ago; thus we should expect generalizations to be error prone. This data set, while chosen with the goal of being as representative a sample of the population as possible, is still small and requires further investigation. Our study and synthesis, we believe, provides a good starting point for subsequent work in an area that we found to be thin on academic literature.

We also saw that this method, to our knowledge, has not been implemented before. Most of the work we found on Islam and liberal democracy, expressed itself primarily in the realms of abstract political theory, philosophy, and theology. By examining a diverse segment of the Islamic world’s history, constitutions, and political institutions, we hope to make more confident claims about its relationship with liberal democracy than other methods have been able to as of yet.

These cases gave us some easy takeaways:

1. Tunisia is an Islamic country which established a robust and inclusive democratic apparatus.
2. Extreme political Islam, in the example of Iran, is still a functioning and powerful opposing force to liberal

democracy.
3. UAE, Lebanon, Kazakhstan, and Indonesia exemplify the wide political variability in the Islamic world between the two poles of Tunisia and Iran.
4. All but Iran have reacted to Islamist movements within the Islamic community and suppress radicals.
5. All have Overton Windows of different sizes. The most liberal and most illiberal opinions permitted in each society vary wildly, and the range of opinions tenable for the politically successful varies similarly.
6. Consequently, all have freedom of speech, expression, and assembly issues.
7. All have some large minority groups, except Tunisia which is 99% Sunni Muslim.

Despite the fact the Tunisia and Iran are the two ends of the spectrum in our study, neither is a fixed star. Over the course of Tunisia's liberalization elements of the traditional community have reacted violently. With Tunisia traditional Islam has merely been sidelined, rather than incorporated into the political process.

For Iran, despite the existence of a deeply entrenched totalitarian clerical regime, demands for reform in the realms of rights and democratic institutions continue to surface and assert themselves in the form of political parties and activism, often at great peril to those involved.

Of the other four cases, some consistencies revealed themselves. They all make strong efforts to mitigate extremist activity within their borders. The UAE monitors and controls the content of sermons in mosques. Lebanon requires proportional representation in government across confessional lines. Indonesia and Kazakhstan employ security forces to quell radical activities. All of these countries see the rise of radical political Islam as an existential threat, and suppress it, often through methods that would be less than ideal in a democratic liberal state.

Each protects some human rights, while other rights are withheld entirely, or are superseded by higher order national concerns such as security and economic prosperity. In these four countries, democratic mechanisms like independent judiciaries, anti-corruption, free elections, separation of powers, free speech, free assembly, and female political participation enjoy different levels of protection and functionality.

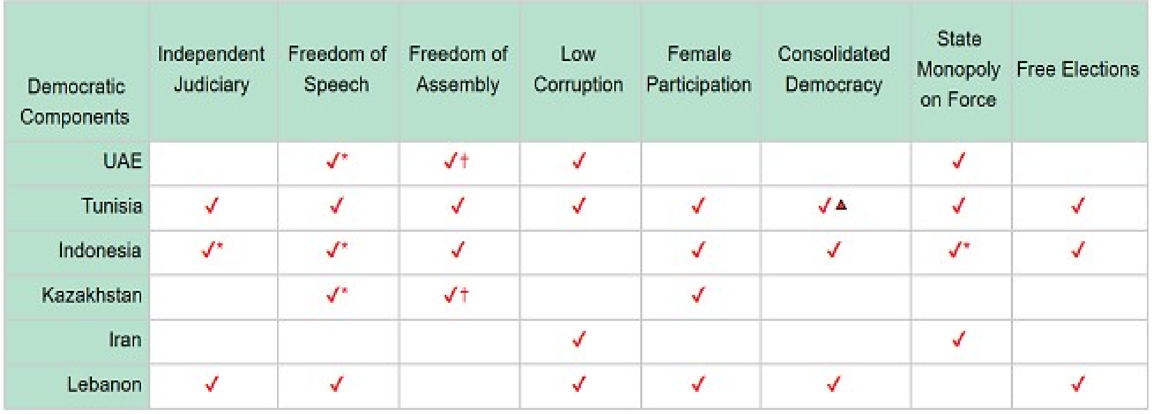
We agreed that Lebanon and Indonesia had at least functioning democracies and therefore, operational avenues for improvement exist in the above mentioned areas. The UAE and Kazakhstan are oligarchic and authoritarian, respectively. It is more likely that Lebanon and Indonesia will see meaningful movement in the coming years towards liberal democratic regimes than the UAE or Kazakhstan.

Can we put some numbers on this? There are two ways of calculating the probabilities here. [Warning: we are neither mathematicians nor statisticians]

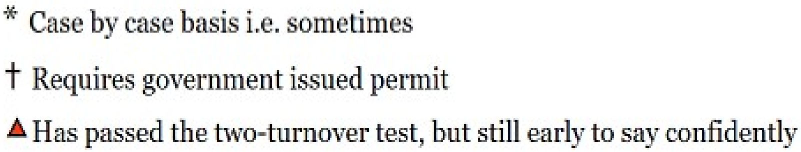
1. A ½ chance of being democratic times a ⅓ chance of protecting human rights gives Muslim majority countries a ⅙ chance of being liberal democracies.
2. Given that a country is democratic, there is a ~40% chance it has basic human rights protected. Given that a country has human rights protected there is a 50% chance that it democratic.

Obviously, this is not how political probability works, but that doesn’t make ⅙ probability a useless consideration.

Adding more nuance, we look at more variables: independent judiciary, freedom of speech, freedom of assembly, low levels of police and administrative corruption, female participation, [democratic consolidation](), state monopoly on force, and free elections.



Democratic Components	Independent Judiciary	Freedom of Speech	Freedom of Assembly	Low Corruption	Female Participation	Consolidated Democracy	State Monopoly on Force	Free Elections
UAE		✓*	✓†	✓			✓	
Tunisia	✓	✓	✓	✓	✓	✓▲	✓	✓
Indonesia	✓*	✓*	✓		✓	✓	✓*	✓
Kazakhstan		✓*	✓†		✓			
Iran				✓			✓	
Lebanon	✓	✓		✓	✓	✓		✓



* Case by case basis i.e. sometimes

† Requires government issued permit

▲ Has passed the two-turnover test, but still early to say confidently

No country has all of these factors, but as common sense and probability distributions tell us, these factors are interrelated.

Concerning Islam, the exact question determines the answer. In our initial proposal for the collaboration we framed our topic as “A dominantly Muslim group could form a liberal democratic ruling party within the confines of contemporary Islamic Political Thought.” What is a fitting response to this initial declaration? We think the statement is true, because clearly there are Muslims throughout the Islamic world, who want to form a liberal democratic polity. However it leaves unanswered the question behind the question. Does current Islamic belief and practice act as an obstacle to liberal democracy more often than not? Each of our country studies seem to generate a different answer to that question.

As for liberal democratic Muslims whether they constitute a viable cultural force, to what extent they have clerical and intellectual support, to what extent other Muslims see them as heretics or apostates, these are different questions that we cannot answer here.

Democracy and liberalism require that a society lower or greatly shift expectations for what types of society can be achieved in this world. Democracy assumes that the state should reflect the will of the citizens to be free from tyranny. This means seeing political participation as a laudatory activity. Citizens then cannot submit to the state, in the same way they submit to the will of God, but must come to see the state as a human apparatus for protecting and promoting the interests of the citizens.

Liberalism requires further that the role of the state is to protect merely the material interests of citizens and to allow individuals the freedom to choose their own good, even if they choose wrong. Freedom of conscience, association, and speech each increase opportunities for people to choose wrong and lead others down with them. A strong traditional morality sees this risk as unacceptable. The goal of the liberal state cannot be “helping the people conform to the will of God, which is the highest good for man.” Rather the role of the state is safeguarding the citizen’s freedom to conform to the will of God or not.

An Islam which accepts both of these visions for government would have to allow that the state does not need to reflect the authority of God and God’s revealed word and that ideological purity is not necessarily desirable. Lastly, democracy and liberalism must support security, prosperity, and justice, rather than undermine them. The most frequent objection to democracy and liberalism in these countries is that they are destabilizing to society.

Many Muslim countries have travelled some portion of the path to liberal democracy in the brief period during which these forces have interacted. When liberal and democratic policies correlate with economic prosperity and security, an Islamic majority state is more likely to continue adopting such policies. For most states so far the steps taken remain small, and the results of their experiments uncertain.

Our predictions:

About UAE

The UAE will establish an independent judiciary by 2023: 5%
The UAE will establish an independent judiciary in the next 2028: 10%
The UAE will invest the Union National Council with authoritative legislative power by 2023: 10%
The UAE will invest the Union National Council with authoritative legislative power by 2028: 20%
The UAE will be invaded by Iran by 2023: 10%
The UAE will experience proxy war instigated by Iran by 2023: 15%

About Tunisia

Tunisia’s next transfer of power will be unbloody: 65%
Non-state terrorist actors will instigate guerilla warfare in Tunisia by 2023: 30%
Non-state terrorist actors will instigate guerilla warfare in Tunisia by 2028: 15%
Tunisia will experience regime change before 2023: 45%
Tunisia will experience regime change after 2023: 30%

About Indonesia

Indonesia will pass a law allowing for harsher prosecution of terrorists by 2020: 75%
Indonesia will not pass new laws promoting Islam at the federal level by 2020: 70%
Indonesia will not be involved in any wars with a foreign power by 2020: 80%
Indonesia will pass more than one piece of legislation reducing the autonomy of different regions by 2020/ by 2025: 40% / 60%

About Kazakhstan

Kazakhstan’s president will resign by 2020: 10%
Kazakhstan’s president will die of natural causes while in office: 60%
There will be a flurry of talks about democratization after Nazarbayev’s death, but very little will come of it within the first 5 years: 65%
By the 6th year after his death real political opposition parties will develop: 44%

About Iran

Iran will experience regime change before 2023: 15%
Iran will experience regime change by 2028: 30%
The Reform Party will have a major win in the next elections: 15%
Iran will invade Israel by 2023: 35%
Iran will obtain nuclear capability by 2023: 25%
Iran will obtain nuclear capability by 2028: 40%

About Lebanon

The number of Syrian refugees in Lebanon drops to below 1 million by 2020: 70%
The number of Syrian refugees in Lebanon drops to below 500,000 by 2020: 20%

Lebanon passes some law which allows civil marriage by 2020, 2025: 30%, 45%
Lebanon holds a census by 2020: 20%
Lebanon holds a census by 2026: 40%
Lebanon holds a census without recording religious denomination by 2020/ by 2026: 5%/ 8%
Lebanon has a newsworthy border conflict >1 week with Israel or Syria by 2022: 80%
Lebanon descends into civil/proxy wars by 2020/ by 2026: 5%/ 10%
If Iran's government is replaced, Amal-Hezbollah loses >3 of their seats in parliament in the 2022 election: 60%

Suggestions For Further Reading

For a detailed Pew Report on Islam and Christianity in Sub-Saharan Africa see http://www.voltairenet.org/IMG/pdf/Islam_and_Christianity.pdf

For a philosophical analysis read Islam, Secularism, and Liberal Democracy. We especially recommend pages 103 – 123, 144 – 165. The book takes into account secularism, which our paper essentially ignores.

For an engaging view into ISIS, listen to Rukmini Callimachi's **Caliphate** which is fantastic.

We haven't read it, but we want to read this humanizing book about the people one meets while hitchhiking in the Middle East by Juan Villarino.

The movie A Jihad For Love is a documentary about Muslims and homosexuality. It is not a particularly good movie, but it was fun.

Also, it's always fun to read the YouTube comments on Arabic and Islamic music.

The Question of Orientalism by Bernard Lewis. Bernard Lewis was one of the foremost historians on Islam and the Middle East. This essay is a great expose on western misgivings about Islam and the Middle East.

From Babel to Dragomans: Interpreting the Middle East by Bernard Lewis.

Contains an excellent essay on Islam and Secularism called, “Can Islam be secularized?”

Power, Faith, and Fantasy: America in the Middle East: 1776 to the Present by Michael B. Oren

Oren is an American born Israel historian and former Israeli ambassador to the U.S. This book is an invaluable overview of the history of American involvement in the Middle East and its relationship with Islam.

Counting Islam: Religion, Class and Elections in Egypt (Cambridge University Press) by Tarek Masoud. To understand the less savory side of the Arab Spring, see the converse of what happened in Tunisia in this Tarek Masoud’s examination of how Mohamed Morsi and the Muslim Brotherhood utilized democratic processes to gain power and then used it to deconstruct a liberal democratic program in favor of a hardline Islamist one.

A History of the Modern Middle East by William Cleveland and Martin Bunton. A sober and detailed history of the contemporary Middle East. A must read for understanding what produced the conditions across the Middle East that we see today.

CONTACTS

littlejohnburidan[at]gmail

christianflanery11[at]gmail

Notes

---

1. Westphalian Sovereignty is a principle of international law that each nation state has complete sovereignty over its territory and domestic affairs to the exclusion of all external actors, and that every nation state, no matter how big or small, is equal under international law. ↩

[ACC Entry] Should Childhood Vaccination Be Mandatory?

This is an entry to the Adversarial Collaboration Contest by Mark Davis and Mark Webb, who sent the following introduction along with their entry:

> **Mark Davis is a naturopathic doctor. Naturopathic medicine is a century-old profession in the United States, but it's small, with fewer than 10,000 NDs licensed to practice naturopathic medicine in the US in 2018. The profession has been historically highly skeptical of vaccination in general, and the modern profession is contentiously split on the topic, with vocal advocates of CDC-scheduled routine childhood vaccination and vocal dissidents both offering continuing medical education for NDs. Mark Davis' main goal in this adversarial collaboration was to argue that there is enough reasonable doubt that routine childhood vaccines could contribute to hyper-inflammatory disease, and enough reduced harm from vaccine-preventable diseases from other medical and public health interventions (in countries with greater economic resources) that parents should be given wide latitude to make individual choices re: routine childhood vaccines despite the clear benefits to individual and public health from preventing those diseases. He became more convinced in his conversations with Mark Webb that widespread childhood vaccination is in the best interest of public health.**
>
> **Mark Webb is a clinical researcher – with a current focus in oncology. He completed a PhD in**

immunology, specifically focused on the mechanisms driving the development of asthma. Mark Webb's main goal in this collaboration was to argue that atopy and autoimmunity are likely not driven by vaccination, and that this idea is a distraction from finding the real causes of the increase in these diseases. Throughout the collaboration, he was reminded of the nature of safety surveillance with all drugs, and of the sensitive nature of vaccination as a medical intervention. He became persuaded that policy should not just reflect the best evidence currently available, but should also reflect a certain degree of humility that there will always be something we don't know.

Setting the parameters of the debate

Why are vaccines the target of both intense support on the one side, and intense skepticism on the other? In part, this is because of the nature of how vaccines work. On the side supporting vaccination, there is strong evidence that vaccination changed the face of epidemic disease in the 20th century. Smallpox is effectively extinct, and polio is nearly there. What agent caused this veritable miracle? Vaccines did. Some diseases are harder to create vaccines against, like HIV or herpes, but eventually we can envision a day when vaccine development can – not just cure – but prevent huge numbers of people from ever having to worry about the deadly diseases of the past. Vaccination is clearly a proven tool for promoting public health. It has been successful at eradicating diseases that used to be endemic to various regions; and where diseases haven't been eradicated vaccination has been very successful at preventing outbreaks and disease spread. What could possibly be bad about vaccination?

Perhaps the biggest reason vaccination has received the degree of skepticism it does is because of how it is administered. Any medical intervention that is targeted toward a high percentage of the population should be scrutinized. Indeed, it would be irresponsible not to undergo continual safety surveillance of a medical intervention that is administered to 90% or more of the population. Vaccines are also administered in multiple doses to one of the most vulnerable population categories: children. There is a strong tradition in clinical research to ensuring high levels of oversight toward children and other vulnerable populations.

Finally, vaccination is a medical intervention intended to produce a permanent effect. It is especially important to be vigilant about therapies whose effects are intended to be persistent. A drug that temporarily relieves asthma symptoms is generally less suspect than one that actually cures asthma. This is because if the expected effects disappear over time, any unknown and unexpected effects are more likely to disappear (although this is not always the case). However, if we’re looking at a treatment with long-lasting effects, unknown long-lasting effects could also appear.

This does not, in itself, mean we shouldn’t implement medical innovations meant to be permanent, targeted toward children, or that would have widespread impact. That would be like suggesting we cease all pediatric cancer research. But it is important to understand why the conversation about vaccine safety is necessarily an ongoing inquiry, not a one-off check of whether “vaccines are safe”. It is also not irrational for a subset of individuals to continue to be wary of possible missed adverse effects, no matter how much research fails to demonstrate any harm.

Before we introduce the parameters of this debate, we wish to emphasize that vaccination is a method of intervention, not one specific intervention. The statement “vaccines are safe” cannot be applied across the board to all vaccines that ever have or ever will be created, any more than you could say, “prescription drugs are safe” for all current and future prescription drugs. This question would hinge more on our confidence in the clinical approval process to ensure drug safety – an interesting question, but one entirely beyond the scope of this essay. In that sense, any general complaint you might make about prescription drug approval or safety could equally apply to any vaccine. In addition, a dozen studies demonstrating the safety of the DTaP vaccine do not demonstrate that MMR is safe. Studies for MMR have to be conducted independently, just as studies about amlodipine do not tell us whether olmesartan is safe.

These, then are the parameters surrounding vaccine safety:

- Vaccination has proven benefits to public health
- Vaccination has all the hallmarks of an intervention with the potential to cause harm

One more consideration should be noted here. In general, the benefits of widespread adoption of routine childhood vaccination in countries with fewer economic resources are clear and not disputed between the collaborators. Nations with little access to medical care are likely to see greater benefits from vaccination than nations with highly accessible medical care infrastructures. For example, an infection that would be lethal in parts of sub-Saharan Africa might be easily treated if contracted in France. Thus, a risk-benefit analysis for economically developed countries will require a more stringent requirement for clear benefit over risk than in the developing world.

From this, we will consider two proposals for economically developed nations such as the US, Europe, Canada, Japan, etc.:

1. Mandatory vaccination is necessary to achieve public policy objectives for vaccines.
2. Public policy should encourage parents to not vaccinate, or should at least normalize parents’ decisions to avoid vaccination.

Should vaccination be mandatory?

In order to recommend that vaccination, as a matter of public policy, should be mandatory, we would need to show that:

1. Vaccination achieves a legitimate public policy objective
2. This public policy objective cannot be achieved without making vaccination mandatory

When considering vaccine benefits (and indeed virtually everything about vaccines) it is important not to generalize the best or worst aspects of one vaccine with another vaccine, or a vaccine used in one socio-economic context with a vaccine used in another. The benefits of the smallpox vaccine have been significantly greater than, say, the rotavirus vaccine, and rotavirus vaccine provides more benefits in countries with few healthcare resources. Even so, rotavirus vaccination still conveys positive benefits that should not be ignored; those benefits should simply be put in context, and any potential adverse effects of rotavirus vaccination should be factored in.

Let’s continue with rotavirus for a minute to highlight what we mean. In most people who contract rotavirus, the greatest concern is dehydration. Most deaths from rotavirus currently occur in the third world, not because rotavirus isn’t transmitted in the US, but because among those who do contract rotavirus hydration therapy is highly successful. In other words, if you’re living in a place where you have to hike 3 miles to collect dirty drinking water that made you sick in the first place, you’re going to struggle with this disease. If you live in Germany and have ready access to quality healthcare services no matter where in the country you live, you’re probably going to be fine.

That doesn’t mean the rotavirus vaccine does nothing. Most people who get the vaccine will be spared the debilitating diarrhea and possibly the trip to the ER. So it’s a meaningful intervention, but it’s not really a life-or-death intervention in resource-rich countries. A similar story can be told for some – though not all – of the other vaccines on the US CDC’s recommended schedule. Generally speaking, not contracting the disease produces the positive good of preventing morbidity and other costs to individuals, but it’s mostly not a life-or-death event. This distinction is important, because “has strong benefits” can be weighed against potential downsides. On the other hand, “keeps you from dying” is hard to weigh against even debilitating or disfiguring downsides. This is basically how something like chemotherapy can become a real treatment, instead of a particularly cruel “enhanced interrogation” technique.

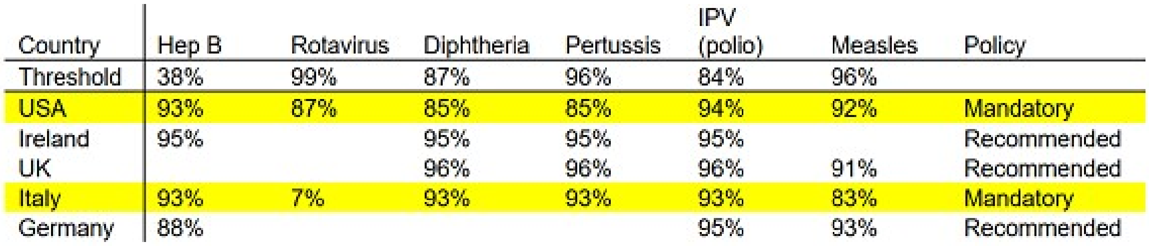
If you live in a totalitarian dictatorship, it’s much easier to make something mandatory. You just tell everyone to do it and if they don’t, you line them up against the wall. In a democratic republic, where the perception of the people often shapes public policy, it’s important not to make enemies of the general public. And although this is not a bar against a mandatory policy, it suggests any such policy should be tempered with the aim of ensuring it is strongly justified, is not rigidly unyielding, and therefore does not become burdensome and unpopular. In the US, in most states, mandatory public vaccination tends to meet this bar (with some qualifications).

First, the policy is not rigidly unyielding in most states. Every state in the USA has some form of mandatory vaccination policy in order for children to attend public schools. Since education is mandatory, and public schooling is freely available to all children, this amounts to a strongly coercive opt-out system. Parents who do not wish to vaccinate are forced to pay a price for their dissention by finding some other way to educate their children than through the public education system they cannot opt out of contributing to through taxation.

Some exceptions are allowed, depending on the state you live in. For example, every US state allows exemptions for medical reasons, since, for example, some small percentage of people are allergic to some of the components of vaccines. All but three states allow religious exemptions, for those whose religion prohibits vaccination (CA, MS, and WV only allow medical exemptions, representing less than 15% of the total US population). But if you don’t belong to a religion that prohibits vaccination, you’ll need to live in one of the 18 states that allow exemptions for personal beliefs as well (these include AR, AZ, CO, ID, LA, ME, MI, MN, MO, ND, OH, OK, OR, PA, TX, UT, WA, and WI representing about 35% of the US population) if you want your children to attend public school without getting them vaccinated.

There is, perhaps, a general concern about totalitarian tendencies here. The concern is that, with physicians as gatekeepers of medical care, they are in a particularly coercive position when it comes to individual patient decisions. Say a patient is strongly opposed to some aspect of vaccination, and wants to opt out of the system. When they try to do this, perhaps their doctor refuses to play along, preferring to use their position of authority to compel the parents to following standardized guidelines. This is certainly the case in some situations, but is it the norm? According to a survey of Washington State pediatricians in 2011, a majority reported they are willing to follow an alternative vaccination schedule to the one advocated by the CDC if a parent requests one. Interestingly, 77% of the pediatricians surveyed reported parents sometimes or frequently make these requests. So not only are parents asking pediatricians to follow different guidance than what the CDC recommendations, most pediatricians report that they are willing to comply. As these are statistical results, this means that there are some parents asking to follow a different vaccine schedule who are refused by their pediatrician; but it appears that (at least from what we know of Washington State) these parents need only go looking for a readily-available second opinion and they will find a pediatrician who is willing to go along with the vaccine schedule they prefer.

Despite this, the argument, “vaccination should be mandatory” must contend with one uncomfortable fact: in many European nations vaccination isn’t mandatory, and those nations often achieve higher vaccination rates than in the US. The table below lists different nations’ vaccination rates. In addition to comparing these rates to one another, it is necessary to compare rates to the threshold required for “herd immunity”. One compelling public health argument in favor of vaccination is the potential of a vaccine to stop the spread of a disease because an infected person will be unlikely to spread the infection prior to recovery because everyone they meet is already immune. It’s a little more complicated than this, but fortunately it can be easily simplified into one number that represents the percentage of the population that needs to be vaccinated in order to ensure the disease will slowly die out faster than it can spread. This is represented by the “threshold” row in the table below.



Country	Hep B	Rotavirus	Diphtheria	Pertussis	IPV (polio)	Measles	Policy
Threshold	38%	99%	87%	96%	84%	96%	
USA	93%	87%	85%	85%	94%	92%	Mandatory
Ireland	95%		95%	95%	95%		Recommended
UK			96%	96%	96%	91%	Recommended
Italy	93%	7%	93%	93%	93%	83%	Mandatory
Germany	88%				95%	93%	Recommended

Notice that for hepatitis B less than 40% of the population needs to be vaccinated in order to achieve herd immunity. Hepatitis B is usually the first vaccine babies get, with current recommendations being to give this prior to leaving the hospital. The specifics of why this vaccine is recommended this early are probably beyond our scope, but from the perspective of “intended to stop the spread of the disease” we’re probably more aggressive than we need to be. Meanwhile, for rotavirus nearly everyone has to get the vaccine in order to achieve herd immunity. We’d have to live in a totalitarian dictatorship to get the kind of levels we’d need to eradicate rotavirus through vaccination alone. It’s important here to note that the vaccine does confer protection to an individual who receives it. But since rotavirus is so highly contagious, really high vaccination rates are not enough to stop the spread of the disease. Thus, as a personal healthcare decision the rotavirus vaccine appears highly attractive. However, as a matter of public policy rotavirus vaccination cannot be expected to prevent outbreaks. It might make them a little less severe, or perhaps they’ll spread more slowly, but they’ll still happen.

The important observation from the table above, however, is that nations like Ireland and the UK have much higher vaccination rates than the US without making them mandatory. Often, these rates are much higher. For example, the US rate of vaccination against diphtheria is below what is required to achieve herd immunity, in contrast to diphtheria vaccination rates in Ireland and the UK, which exceed the level required for herd immunity. This is not to say that eliminating mandatory vaccination will increase vaccination rates. Each nation has different, and unique, health care systems, laws, policies, and behavioral norms. This is probably more complicated than “Let’s just copy what the Germans are doing.” But it is not possible to argue, “Without mandatory vaccination we cannot achieve herd immunity; people will be dying of disease in the streets!” Although we should be cautious about sudden, dramatic changes to a system that is largely working, both authors concede that in developed nations such as the US mandatory vaccination is probably not necessary to achieve public health objectives.

Should health authorities normalize parental decisions not to vaccinate?

Any medical intervention comes with some level of risk, both known and unknown. For vaccination, the most common, well-documented, known risk is the potential for an allergic response to some component of the vaccine. The most common allergic component is egg, and people with severe egg allergies are instructed to consult their physician prior to vaccination. How common are allergies to vaccines? A good estimate is about 3 per one million doses. This would be the equivalent of about 200 people in France, or 35 in the US state of Ohio. This is the biggest recognized, known risk of vaccines. But are there significant unrecognized risks of vaccination that could impact the risk/benefit assessment of vaccine safety?

In order to make a general recommendation against vaccination as a matter of public policy, any identified harm would need to outweigh the benefits which those vaccines confer upon their recipients. There are a number of theories about potential harm that could come from vaccines. Much ink has been spilled about vaccines and autism, and it is not our intent to cover that ground again here. Both authors agreed that the evidence does not support a link between vaccines and autism.

There is another, more subtle linkage that we would like to consider here; this is the hypothesis that vaccines might contribute to autoimmune and autoinflammatory diseases such as multiple sclerosis, type I diabetes, inflammatory bowel disease, rheumatoid arthritis, etc., or to atopic diseases such as asthma, eczema, and food allergies. These diseases have been increasing for decades – the same decades during which we have increasingly been administering more vaccines, earlier and earlier in childhood and into infancy. Thus, it is appropriate to consider the possibility of a causal link between these two phenomena: do vaccines lead to autoimmune and allergic disease?

Are kids who are vaccinated more likely to develop these immune system diseases? Despite a large number of studies into this area, the results so far have been mixed. On one hand you could argue, as above, that vaccines lead to increases in immune system diseases. And indeed, you can find researchers who have demonstrated just such a link for DPT, tetanus, MMR, etc.

- DPT (article; article)
    - Tetanus specifically

    - Tuberculosis (BCG vaccination doesn’t prevent allergic disorders, but positive TB test does)
    - Measles (article; article)

Meanwhile, other researchers have hypothesized that vaccination protects against development of allergic disease and autoimmunity. How might this be? There is strong evidence that increased antibiotic use in early childhood is associated with increases in developing immune system diseases; childhood use of antibiotics can shift the balance of commensal bacteria in a way the hygiene hypothesis would predict makes you more susceptible to immune system diseases. So it’s also possible that not vaccinating could lead to increased antibiotics use if your child does get infected with measles, mumps, pertussis, etc. Regardless of whether antibiotics are the mediating factor, some studies indicate certain vaccines having a protective effect against atopy:

- Tuberculosis (meta-analysis)
- Measles (article; article)
- MMR (article)

Finally, some studies find no difference between vaccination and natural infection in development of immune system disease:

- Multiple vaccines (article; article)
- DPT (as a marker for all vaccines)
- Pertussis (We were only able to locate one genuine placebo-controlled RCT of a routine childhood vaccine in which the authors looked for atopy – they found no significant difference in atopy between the placebo and real pertussis vaccine groups at ages two and a half and seven)

This topic has been reviewed multiple times in the scientific literature, and the conclusions have been the same each time: there is no demonstrable impact of vaccines driving immune system diseases. Given these conflicting studies, we can’t say that there is convincing evidence that vaccines either cause increased immune system disease or that they protect against development of these diseases.

One hypothesis for how vaccines might contribute to the rise in atopy and autoimmunity side-steps most of the evidence cited above. These articles look at whether vaccination itself causes allergic disease, but what if the opposite is true – not vaccinating protects against allergic disease?

Hygiene Hypothesis

What causes the development of autoimmune, autoinflammatory, and atopic disease? A full answer to that question – one that could lead to prevention of these diseases – would probably be worth at least a Nobel prize in medicine; which is to say we don’t entirely understand it. However, the current leading explanation in vogue amongst immunologists and epidemiologists who study the recent trend in which we see these diseases increasing dramatically in the developed world is call the hygiene hypothesis.

First proposed about thirty years ago, the hygiene hypothesis is the idea that some of the bacterial and parasitic infections that modern medical technology has eliminated might have been performing an important function in the human immune system – and when you take them away you start seeing problems. For example, if you go back 5,000 years in human history, few people would be completely free of parasitic infections, such as hookworm or whipworm. These parasites might make you mildly ill after you first get infected, but so long as your immune system maintains control of the infestation you may not notice it.

There is a constant, low-grade battle between your immune system and the parasite. This battle doesn't just go on your entire life, but has gone on for generations of humans (and their common ancestors), such that this is the normal state of affairs. Fast forward 5,000 years, and modern water treatment suddenly prevents millions of people from ever experiencing a type of infection that was a constant throughout humanity's evolution. As a result, the immune system doesn't know what to target. There has never been a time when there was *nothing* to fight, so it begins to fight itself, accidentally.

This hypothesis isn't just high-level theoretical hand-waving. Parasites, such as hookworm, have been shown to induce the same kind of immune mediators that are commonly seen in autoimmune diseases. In fact, some people with diverse autoimmune, allergic and autoinflammatory conditions have started intentionally infecting themselves with hookworm. Based in part on this movement, clinical trials have been conducted, and more are currently under way investigating whether re-introducing parasitic infections such as hookworm can be used to treat Crohn's and other autoimmune diseases.

If the hygiene hypothesis is correct, and removing certain persistent infections is driving the increase in autoimmunity, autoinflammation, and atopy; does that mean there is a hygiene hypothesis explanation that links vaccination with these diseases?

This hypothesis is much more difficult to test, in the case of vaccination, because it's not saying the vaccine itself causes allergy and autoimmunity. Instead, it argues that getting rid of diseases such as measles and pertussis causes the increase in allergy and autoimmunity. Thus, it doesn't really matter whether you get the vaccine or not, what matters is whether you get measles or pertussis or not. If everyone around you gets vaccinated, herd immunity will kick in – the very effect public policy is looking to achieve – and you will never get the chance to catch measles. So no matter how large or well-designed or randomized/controlled, a study that compares US children who are vaccinated with those who aren't won't be able to test the hypothesis that endemic measles outbreaks protect against development of allergies and autoimmunity.

The closest we can come to addressing this question are some of the measles studies cited above. One of these compared individuals in Guinea-Bissau who were either naturally exposed to measles or who had been vaccinated; they identified a protective effect for natural measles exposure. Meanwhile, another study in Finland made a similar comparison and found natural measles infection exacerbated atopy and allergic disease. The problem with this approach is that it doesn't just compare vaccination with not vaccinating – it compares *people* who don't vaccinate with those who do. These are not necessarily the same. For example, people with access to vaccines in a nation such as Guinea-Bissau might be of a different socio-economic status than people with no access to vaccines. And socio-economic status has been identified as a factor in development of atopy and autoimmunity.

One way to look at this is to ask whether vaccinated children have different immune system markers, such that they look like they are more susceptible to developing immune system diseases. Researchers at the University of British Columbia looked for exactly this kind of change in a recent study. They took blood from children who had or had not been vaccinated and looked at whether these children's circulating immune system cells differed from each other. They did not find any of the differences we would expect to find if vaccines caused a general shift in a vaccinated person’s immune system. This study should be taken with a large grain of salt however, as the sample size was smaller than expected, due to the difficulty of recruiting unvaccinated children.

A more targeted way to test this hypothesis would be to randomly assign children to receive one or multiple vaccinations and compare to children who receive no vaccination, then expose all the children to the disease(s) they were vaccinated against and check whether they develop various types of atopic and autoimmune disease. This would have to be done in a nation that currently has endemic levels of the disease in question and/or frequent outbreaks. Also, a nation with no scruples about conducting experiments on children in which you intentionally expose them to infectious agents at an early age.

This does not mean the hypothesis cannot be tested, but it will be a difficult hypothesis to test consistently, and the current state of the evidence suggests a high level of disagreement. In advance of such evidence, we might ask: What are the implications if this hypothesis were confirmed? Would that mean deciding whether to accept – even encourage – endemic disease burdens such as measles, polio, pertussis, etc.? There is no reason to believe that the hygiene hypothesis requires specific bacterial or parasitic infections in order to promote the development of a healthy immune system. It is likely that whole classes of commensal bacteria are protective against the development of atopy and autoimmunity. If there were a group of people, living in a nation that has achieved herd immunity to many of the infectious agents discussed above, and that had rates of allergy and autoimmunity significantly below that of the rest of the population we might study that group to determine whether they are exposed to commensal agents that are protective against atopy and autoimmunity.

Fortunately, the Amish in the US provide an excellent example of just such a case. They experience almost none of the diseases we associate with developed nations. They have far less cancer, asthma, food allergies, MS, etc. You might think, “but that’s because they don’t vaccinate!” except that they do vaccinate. Different communities vary, as each Amish community makes its own rules about what aspects of modern technology to adopt, but one survey of the Amish suggests that about 85% of Amish children are receiving vaccinations. Additional anecdotal evidence suggests this may be a lower bound, but even if we assume 85% of Amish children are getting vaccinated we have to wonder why they see such low rates of modern diseases. Sure, they aren’t at the 90-95% of most of the rest of the country, but it’s hard to see how that extra 5-10% vaccination rate could be leading to such a huge increase in autoimmune diseases – especially without significant numbers of outbreaks running through their communities. Maybe it’s just something about the Amish?

It’s not just something about the Amish. An early observation about the hygiene hypothesis is that people who live on farms have a much lower rate of developing immune system disorders. The belief is that this is because they are more frequently exposed to environmental bacteria and parasites.

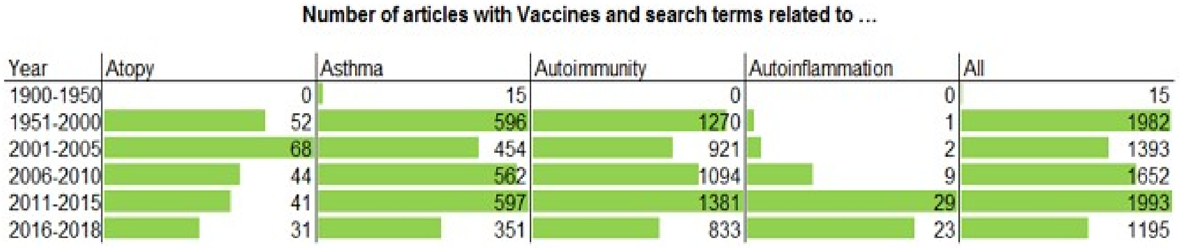
A group of researchers identified a different German religious sect, called the Hutterites, which also engages in regular farming. They are closely related to the Amish, and came from similar parts Germany at similar times. However the norms of the Hutterites dictate a much lower interaction between livestock and children/pregnant women (most hygiene hypothesis evidence suggests prenatal, neonatal, up through young childhood exposure is the critical exposure period). This genetically similar group, who had less early-life exposure to the farm environment than the Amish, gets asthma at a rate 5-7 times higher than Amish farmers (Amish asthma rate is 2-3%, Hutterite asthma rate is 15%). The researchers then took dust from the Amish barns and forced mice to breathe it in. They found the dust protected the mice from developing an experimentally-induced allergic response, and that it caused real, measurable changes in the immune systems of these mice, consistent with what we see in humans who are less allergic, and consistent with the differences they saw in the Amish farmers who had low allergic disease compared to the Hutterites.

But not everyone can live on farms (anymore) so current research is also focused at discovering which commensal bacteria humans need to protect against developing immune system diseases. This is similar to the idea of taking the dust from Amish farms, but consistent treatments require us to actually know what elements of the dust actually matter to preventing disease. It’s possible, for example, that modern lifestyles reduce exposure to exposure to mycobacteria. Some promising attempts have been made at reintroduction of killed mycobacteria into atopic individuals. This approach is akin to creating a vaccine against atopic disease, though more recent research has focused on changing the balance of live bacteria instead of simply introducing killed bacteria. This approach attempts to retrain the all-important gut commensal balance through techniques such as fecal microbiota transplantation (poop transplants) or investigating dietary changes that could help push the balance toward protective commensal bacteria and away from sensitizing commensals.

If we create a vaccine that specifically targets the commensal bacteria, or parasites that protect you from developing immune system diseases, we might suspect that vaccine of directly contributing to people who receive it developing those diseases. For example, one such helminth has been the target of recent vaccine development due to the significant harms it causes in the developing world; thus there is a concern on one hand of the persistent symptoms of infection caused by the hookworm infection, and a concern on the other of increasing the risk of allergic infection. In the developing world, where iron deficiency is a major cause of morbidity (large hookworm infections cause iron deficiency as they feed off the blood of hosts), development of an anti-hookworm vaccine could be a significantly beneficial intervention. In economically developed nations these types of infections were mostly eliminated when water treatment eliminated the fecal-oral route many pathogens use to spread from host to host.

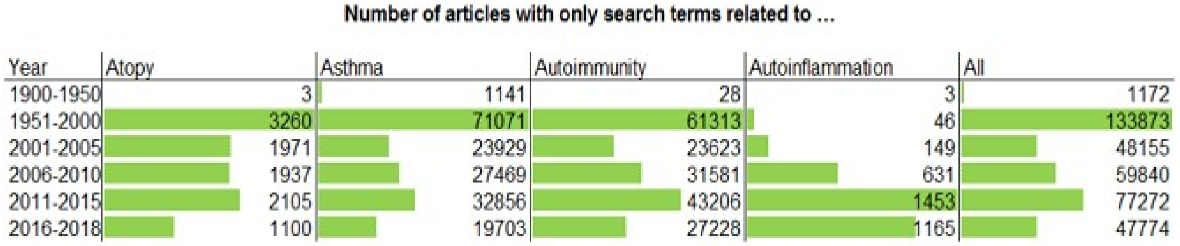
Let’s revisit the implications of the hypothesis that the elimination of endemic diseases by vaccines causes an increase in atopy and autoimmunity. Currently, there is not strong evidence that vaccines drive autoimmunity and atopy, although additional research should be done in this area – focusing on disease exposure and not just vaccination status. Even if such a link were to be established in the future, is cessation of vaccination the best approach? Perhaps a better solution to the rise of atopy and autoimmunity is not to actively encourage the return of endemic diseases that are associated with other significant harms, but to encourage exposure to commensal bacteria and parasites that do not come with significant associated morbidity and mortality, such as those that help prevent atopy and autoimmunity in the Amish.

For the vaccines currently recommended today, we haven’t seen convincing evidence of long-lasting or permanent negative side effects. It is important that the medical community continue to monitor vaccines to confirm the relative safety of vaccination, and the evidence of the academic literature is that this is actively happening. Every year, many articles posted to PubMed confirm that ongoing surveillance of vaccine safety is being actively pursued by the scientific research community.



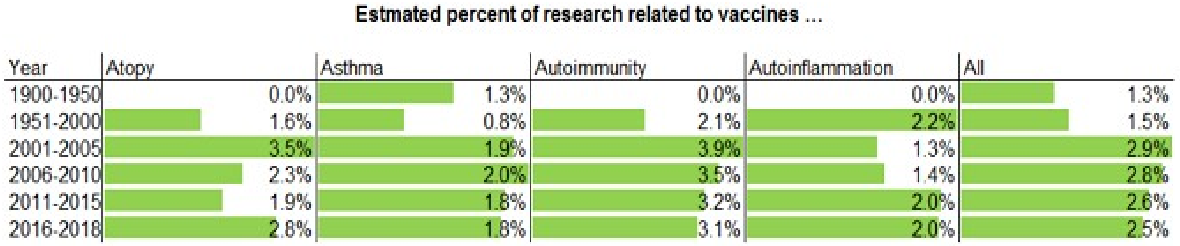
Number of articles with Vaccines and search terms related to ...

Year	Atopy	Asthma	Autoimmunity	Autoinflammation	All
1900-1950	0	15	0	0	15
1951-2000	52	596	1270	1	1982
2001-2005	68	454	921	2	1393
2006-2010	44	562	1094	9	1652
2011-2015	41	597	1381	29	1993
2016-2018	31	351	833	23	1195



Number of articles with only search terms related to ...

Year	Atopy	Asthma	Autoimmunity	Autoinflammation	All
1900-1950	3	1141	28	3	1172
1951-2000	3260	71071	61313	46	133873
2001-2005	1971	23929	23623	149	48155
2006-2010	1937	27469	31581	631	59840
2011-2015	2105	32856	43206	1453	77272
2016-2018	1100	19703	27228	1165	47774



Estmated percent of research related to vaccines ...

Year	Atopy	Asthma	Autoimmunity	Autoinflammation	All
1900-1950	0.0%	1.3%	0.0%	0.0%	1.3%
1951-2000	1.6%	0.8%	2.1%	2.2%	1.5%
2001-2005	3.5%	1.9%	3.9%	1.3%	2.9%
2006-2010	2.3%	2.0%	3.5%	1.4%	2.8%
2011-2015	1.9%	1.8%	3.2%	2.0%	2.6%
2016-2018	2.8%	1.8%	3.1%	2.0%	2.5%

This is appropriate, as we do not ever expect to finally “prove” that vaccines are not harmful. While it is possible to obtain support for a positive declaration such as, “vaccines are effective”, the only way to “prove a negative” with the scientific method is to fail to find support after looking for it. With vaccines, it is important to remember that this is a daunting task. There are many vaccines currently in use, and there are many possible harms to be investigated. In addition, new vaccines are currently being developed, such that we should expect to continue to investigate the safety of various specific vaccines for the foreseeable future.

Conclusions

In the first question, we assessed whether vaccination, as a matter of public policy, should be made mandatory. In essence, this asks whether the benefits to vaccination are sufficiently great that the decision of whether to vaccination should be removed from individual decision-making. Given the sensitive nature of a policy that essentially amounts to dictating private parental medical decisions, we adopted a standard similar to the US Supreme Court’s “strict scrutiny” standard: is it a compelling public health interest, and is it narrowly tailored? This is not generally the standard currently adopted by policy makers today. Mandatory vaccination failed this test, in that it is not narrowly tailored, since herd immunity can be achieved without making vaccination mandatory.

The case of California is an interesting example of how public policy is currently set in regards to vaccination. In 2014, California’s measles vaccination rate was below what is required for herd immunity, and there was a subsequent outbreak of measles at Disneyland. This was a high-profile event. In response, vaccine proponents argued that California’s laws should be strengthened to eliminate the personal and religious exemptions for vaccinations that were then in place. After the law was passed, despite a suspiciously large increase in medical exemptions, vaccination rates rose above the level required to achieve herd immunity, both at the statewide and at the county level. This law also sparked protests from parents who saw the law as removing the rights of parents to make medical decisions for their children. The debate was highly contentious, and continues to be a source of some political animosity.

According to our analysis, future public debates about vaccination do not have to follow this pattern. It is sensible, given the nature of vaccination as a medical intervention, to be skeptical that safety surveillance may have missed something important with respect to vaccines. It is also entirely possible to achieve vaccination rates sufficient to achieve herd immunity without removing medical decision-making ability from parents. A better approach might be to study models such as those of the UK and Germany. In the UK, vaccination is strongly recommended, and vaccines are provided at no cost to the individual. Germany, meanwhile, also strongly recommends vaccination but does not pay for vaccines. A more thorough study of social norms and other factors influencing vaccination rates could provide alternative approaches to the drive for mandatory vaccination, and help alleviate this front of the culture war. This study of alternatives to mandates should be undertaken prior to the next high-profile event, in order to provide policy-makers with a ready alternative that can foment good will between those wary of vaccination and those wary of the potential for outbreaks.

In the second question, we asked whether the public policy toward vaccination should be reversed. It is entirely understandable for concerned parents to adopt a “precautionary principle” approach to vaccination – given the nature of vaccination as a universal medical intervention targeted at babies and young children. However, as a matter of public policy, a general “precautionary principle” approach cannot be recommended in light of the proven harms vaccines protect against. At this time, there is not sufficient evidence that vaccination causes real harms – despite attempts to investigate various mechanisms by which they are theorized to cause harm. This does not mean vaccines cause no harm, but like any medical intervention, we require each vaccine to undergo initial testing for safety and efficacy before regulatory approval, then additional surveillance afterward.

Based on what we currently know, vaccines are an important element of disease control and eradication. Public policy may not require mandatory vaccination, but including recommendations for parents to vaccinate children is a legitimate public policy objective. Vaccine safety and vaccine surveillance are also important and legitimate. Many primary research articles are published each year investigating vaccine risks, and looking for unknown harms.

Ultimately, the question of whether something is “safe” can only ever be either answered:

- “no, we have evidence that it causes significant harm” or
- “we don’t have evidence that it causes significant harm”.

Meanwhile, many of the potential benefits of vaccination are recognized at the level of community adoption – which introduces a coordination problem. Thus, the central conflict we encounter in this area is between individuals who wish to invoke the precautionary principle for themselves and for their families, versus community standards that seek to eliminate a known danger. This conflict (between individual freedom to dissent in order to avoid fat-tailed risk versus a level of community solidarity necessary to combat societal ills) is common to many problems besides vaccination.

We believe this coordination problem may be largely resolved without restricting individual freedom. Individuals who wish to invoke the precautionary principle for themselves and their families should not be penalized for doing so.

[ACC Entry] Should Transgender Children Transition?

[This is an entry to the Adversarial Collaboration Contest by flame7926 and a_reader.]

[*Content note: suicide, depression, transphobia, self-harm*]

Transgender childhood transition is a hotly debated topic, with extensive media coverage devoted to it in recent years. (pro: *BBC*, *The Lancet* and *The New York Times ; contra: The Cut*, *New Statesman* and *The Globe and Mail*). We see plenty of stories of transgender children (or gender dysphoric children and gender nonconforming children), both in the media and in the blogosphere. As early as 2 or 3, defying the expectations of their family, those children show a persistent and insistent preference for many things associated with the other sex: little boys want long hair and love dresses, Barbie dolls, Disney princesses and mermaids; little girls, instead, dislike stereotypically feminine activities and prefer rough and tumble play, refuse to wear dresses and insist to have their hair shorter and shorter.

Sometimes, from the very beginning, the toddler corrects the parents: “I’m a boy /girl!”, but more frequently cross-gender behavior is more prevalent. This is only sometimes followed with the child expressing preferences that would be termed gender dysphoria. The child (born and currently living as a as one sex) says to their parents something like “God made a mistake” or “something went wrong in Mommy’s tummy” because he should have been a girl, not a boy (or the other way around). The worried parents search information on the internet and seek out the advice of an expert. There, they usually find one or both of these contradicting opinions:

Gender-affirming approach

Listen to your child – he/she knows best his/her gender. Let your child be his/her true self. It’s your responsibility as a parent to support your child in all stages of his/her transition: social transition now, puberty blockers at the beginning of puberty, cross-sex hormones in adolescence, surgery at 18. To oppose it is child abuse. Transphobia costs lives: 41% of transgenders attempt suicide. Do you prefer a happy daughter or a dead son?

Or:

Therapeutic approach

Your child is just confused. He/she is too young to understand gender and to take such important decision. 80% of gender nonconforming children desist. You, as a parent, have the responsibility to correct his/her wrong behavior. If you tolerate it, gender dysphoria will be reinforced by repetition and persist to adulthood. To encourage your child’s delusion is child abuse. Transgenders individuals face lifelong struggle and often suffer from poor mental health: 41% of transgenders attempt suicide. Do you really want that for your son, when he could instead come to accept the body he was born with?

The first approach is promoted by transgender activists, the second by the conservative media, but both are supported by some experts. The “*Gender-affirming approach*” is supported by the Dutch team from the *Gender Clinic* at *VU Medical Centre, Amsterdam*, who elaborated the typical transition treatment for minors, with puberty blockers at 12 and cross-sex hormones at 16, and, in the US, by Kristina Olson and others from the *TransYouth Project*. The “*Therapeutic approach*” is supported by Kenneth Zucker and his team from the Gender Identity Service at *Centre for*

Addiction and Mental Health, Toronto, and, in the US, by Paul McHugh at *Johns Hopkins University School of Medicine*. There are also experts such as Debra Soh, once a gender nonconforming girl herself, that advise parents to wait and see until adolescence, because in many cases gender dysphoria desists spontaneously, without intervention.

Who to believe when the experts disagree? Let’s see the evidence.

What is Gender Dysphoria/Gender Identity Disorder?

Children labeled transgender are usually diagnosed with Gender Dysphoria (as per the DSM-V), previously known as Gender Identity Disorder (in DSM-III through DSM-IV-TR). DSM refers to the Diagnostic and Statistical Manual of Mental Disorders published by the American Psychiatric Association – DSM-III was published in 1980, with a revision in 1987, DSM-IV published in 1994, with a revision in 2000, and DSM-V coming out in 2013.

According to the APA as per DSM-V, “gender nonconformity is not in itself a mental disorder. The critical element of gender dysphoria is the presence of clinically significant distress associated with the condition”. Both Gender Identity Disorder and Gender Dysphoria include a desire to be or insistence that they are a gender that does not match their biological sex. This desire has to be strong and persistent and is usually accompanied by a preference for clothing of the opposite gender, cross-gender toys, games, and stereotypical activities, as well as the assumption of cross-gender roles in play. It may be accompanied by a discomfort or dislike of their current sexual anatomy and a desire for the sexual anatomy of the opposite sex. GID was said to “*cause* clinically significant distress or impairment in social, occupational, or other important areas of functioning”, while GD is merely “*associated with”* similar distress or impairments.

Gender and Sex Difference

Transgender refers to an individual gender identity that doesn’t correspond to the sex or gender they were born with, while gender dysphoria is a psychiatric diagnosis that refers discomfort with one’s physical or assigned gender. Individuals can have or previously have had gender dysphoria without identifying as transgender, while the most (but not all) transgender individuals should fall under the gender dysphoric label according to experts.

Based on the balance of the evidence, both social transitioning and puberty blockers should be approached with caution. There is moderate evidence that social transitioning improves mental health outcomes, and from a perspective concerned with validation of trans-identities it is important. But there are few studies on the long-term effects of social transitioning on rates of persistence and desistence (the number of children who remain gender dysphoric as they age).[1] Similarly, puberty blockers may have negative physical health consequences, including an effect on bone-density, though a positive effect on mental health. Still, these choices should be made by families in consultation with experts. Individuals should develop a position on the topic with full awareness of the evidence and taking into consideration their preexisting biases and exposure to social norms on the topic.

Gender Identity Disorder and Gender Dysphoria in Youth

Gender dysphoria in youth is uncommon, with estimated prevalence of around 1%. Acceptance of transgender individuals is relatively low, with 30% of Americans saying they hold somewhat unfavorable or very unfavorable views of transgender individuals. 41% of parents also would be either “very” or “somewhat” upset if their child was transgender. Due to the lack of acceptance of gender dysphoric behavior and identities, it is possible that many gender dysphoric youth go unaccounted for. Gender dysphoria also doesn’t have a perfect overlap with being transgender, as some children feel uncomfortable with the gender that corresponds to their biological sex but don’t necessarily identify as the other gender.

According to Zucker et al. (1993), in a gender identity interview for children, when asked if they are a boy/girl, most (79 out of 85 gender nonconforming children) respond with their biological sex and not the other gender. Yet 30 out of 85 say they sometimes feel more like a boy than a girl (or reverse). Gender dysphoria is associated with feelings that their biological sex is not correct but is not perfectly correlated with such feelings.

According to the APA, treatment options for gender dysphoria include, “counseling, cross-sex hormones, puberty suppression and gender reassignment surgery.”

Desistence

Desisting is the term used to refer to children and youth who previously expressed gender nonconformity (as defined by the gender dysphoria/gender identity disorder diagnosis within various editions of the DSM manual) and are no longer gender dysphoric.

Brief disclaimer – *Desistence is a fraught term and has been used to "denote the cessation of offensive or antisocial behavior". It is difficult to extract the term from its historical context of use in which a transgender life was deemed a less preferred outcome. Yet it is relevant to the discussion and is the term used in papers which study the topic and thus we will use it here, aware of its potentially problematic implications.*

The existence of desisting as a concept is debated by some who say that it diminishes the validity of transgender youth identities by painting dysphoric gender feelings as a choice or a phase that one may grow out of – while others claim flawed methodologies in studies purporting high numbers of desisting youth. (Julia Serano, Temple Newhook et al. (2018)). Yet it remains hard to deny that some fraction of individuals who previously expressed these feelings as youth do not anymore as adults.

Desistence matters because any recommendation for gender dysphoric youth, whether it be social transitioning, puberty blockers, or other treatment, may affect those who end up desisting from their dysphoria as well as those for whom the dysphoria persists.

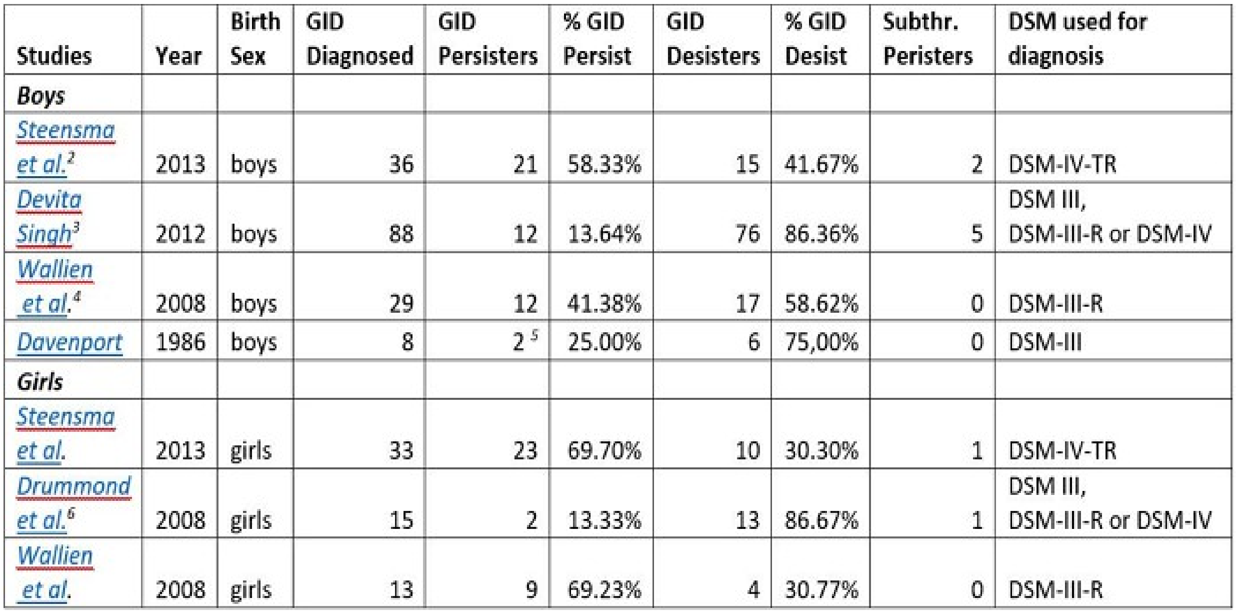
Additionally, papers such as Temple Newhook et al. (2018) and Olson (2015) argue against the narrative of transgender childhood desistance (including the commonly cited figure that 80% of children desist), claiming methodological errors in the original studies. They argue that these studies (including Steensma et al. (2011 and 2013), Drummond et al. (2008), and Wallien and Cohen-Kettenis (2008) – all discussed below) conflate gender non-conforming children and truly gender dysphoric and transgender children. While most of these studies include samples that have both threshold (GD or GID diagnosis) and subthreshold children, the results are separable by diagnosis. These papers additionally critique the Steensma et al. and Wallien and Cohen-Kettenis studies for including non-respondents among desisters, but this is again separable in data analysis.

Temple Newhook et al. (2018) continue to point out that many of the children involved in the studies, (at the Amsterdam and Toronto clinics) were enrolled in programs to reduce likelihood of GD persistence, or at least not supported socially in expressing their identities – and that these were in and of themselves interventions which could affect results. Yet as Zucker (2018) contends, there is no neutral way to approach transgender youth – everything may affect persistence and desistence rates.

We acknowledge these criticisms of the body of research on gender dysphoric desistance, which, while weakening the strength of the evidence, does not invalidate it. **Some children who express a range of desires that can be seen as comprising a transgender identity (up to and including identifying as the other gender) do not continue to hold this identity as they pass through puberty.**

Many of these individuals who do not continue to feel gender dysphoria do identify as homosexual or bisexual. According to one study, half of the boys in the desisting sample were homosexual or bisexual, while 24% of a total sample of gender dysphoric youth (both threshold and subthreshold) were bisexual or homosexual in behavior.

Desistence Statistics



Studies	Year	Birth Sex	GID Diagnosed	GID Persisters	% GID Persist	GID Desisters	% GID Desist	Subthr. Peristers	DSM used for diagnosis
Boys									
Steensma et al.[2]	2013	boys	36	21	58.33%	15	41.67%	2	DSM-IV-TR
Devita Singh[3]	2012	boys	88	12	13.64%	76	86.36%	5	DSM III, DSM-III-R or DSM-IV
Wallien et al.[4]	2008	boys	29	12	41.38%	17	58.62%	0	DSM-III-R
Davenport	1986	boys	8	2[5]	25.00%	6	75,00%	0	DSM-III
Girls									
Steensma et al.	2013	girls	33	23	69.70%	10	30.30%	1	DSM-IV-TR
Drummond et al.[6]	2008	girls	15	2	13.33%	13	86.67%	1	DSM III, DSM-III-R or DSM-IV
Wallien et al.	2008	girls	13	9	69.23%	4	30.77%	0	DSM-III-R

Statistical evidence supports the case that somewhere close to or above 50% of children with GID or GD diagnoses subsequently desist and do not express these feelings past puberty. To respond to criticisms of these studies, we differentiate between those diagnosed with GID or GD (see table above) and those who express gender-nonconforming behaviors but are subthreshold according to the diagnosis. All studies found at least 30-40% of diagnosed children desist, with the percentage as low as 30% in natal girls and 41% in natal boys in the Dutch studies (not counting non-responders), but as high as 75-86% in an unpublished thesis (Singh 2012) and 2 small studies (Drummond et al. 2008, Davenport 1986).

Steensma et al. included a sample of 80 youths formerly diagnosed with GID as children at the Amsterdam gender clinic. Of these formerly GID children, 58.3% persisted of (biological) males persisted past childhood, excluding both non-responses and responses by parents. 69.3% of females persisted past childhood. If non-responses are included (and Steensma et al. make an argument that they should count as desisters since the clinic in question is the only gender and sexual identity clinic in the Netherlands), then 48.5% of males persisted and 62.2% of females persisted.

While the most cautious estimate of desistance figures from this data is 41.7% of males and 30.7% of females, this is still a substantial number of children who previously had exhibited full GID traits who ceased to at some point before adulthood.

Wallien and Cohen-Kettenis (2008) examined a sample of 77 children, with 58 of them formerly GID and 19 subthreshold. 21 of these children persisted and 33 desisted, with 23 non-responses. If only GID children are included, then 50% of the individual male respondents persisted and 50% desisted, while for females 75% persisted and 25% desisted. If parental responses are included as well then the desisting percentages are 59% for males and 41% for females.

This study adds to the body of evidence that a not insignificant number of children with full GID diagnoses who respond to the questionnaire do desist from these feelings at a later point.

Drummond et al. (2008) examined 25 gender non-conforming girls, including 15 girls with GID diagnoses and 10 that were subthreshold. At follow-up (mean age of 23.24 years), three of them still had gender dysphoria. Two of these were from the fifteen girls with a GID diagnosis, giving a desistance rate of 86.6%. Singh’s unpublished thesis (2012) echoes these results, with 76 out of 88 formerly GID participants at follow up having desisted, for a rate of 86.4%.

Though Drummond et al.’s small sample size and the unpublished nature of Singh’s work provide limitations for the interpretability of these two studies, the low percentage of persisters among their results do provide some force to their evidence.

Some additional studies from before 2000 yield similar results with small (less than 20) sample sizes, including Davenport (1986), Zuger (1978), and Lebowitz (1972). While the applicability of these studies is diminished because of the length of time that has passed since their creation and the changes in understanding of gender and sexuality since then, they do still provide some evidence that children's gender identities change as they age.

Overall, it is clear that at least 25% of children with GID diagnoses desisted, and probably closer to/upwards of 50%. These are not insubstantial numbers of children who previously experience a great deal of distress at their natal sex and do not anymore.

Qualitative research also provides important information about desisters, and particularly on differentiating factors between those who will desist from their gender dysphoria from those who will persist in it and/or eventually undergo sex-reassignment treatment. This research indicates that children who persist generally have stronger gender dysphoric feelings to begin with, including a stronger aversion to their physical anatomy and more insistent they were the other sex as opposed to only wishing they were. For example, "The persisters explicitly indicated they felt they were the other sex, [while] the desisters indicated that they identified as a girlish boy or a boyish girl who only wished they were the other sex".

Puberty additionally proved decisive for both desisters and persisters, as it was the time when their gender dysphoric feelings either weakened or intensified. Singh (2012) adds that differences in DSM criteria for children as opposed to adolescents and adults for GID/GD may lead to diagnoses of some children who are not what would be considered gender dysphoric in adulthood. Wallien and Cohen-Kettenis additionally differentiate persisters from desisters by their GIIC and GIQC scores (two questionnaires about gender non-conforming behavior and identity) with persisters on average having significantly higher scores according to these measures.

Yet, neither the strength of cross-gender feelings nor a GID/GD diagnosis are perfect indicators of whether individuals will later identify as transgender or not, as even some subthreshold children persist and are transgender (8.7% of males and 4.2% of females, according to Steensma et al. (2013)). Thus, some differentiating factors are identified such as the strength of gender dysphoria, a gender dysphoria diagnosis, and actual feelings of being the other sex rather than simply wishing one is, but for the most part it is difficult to tell which children will continue feeling gender dysphoric as they pass through puberty into adulthood.

Narratives of Desisting

Nonetheless, there are a wide variety of examples of children who previously expressed varying degrees of gender dysphoric feelings and do not anymore. These children range from those who simply had cross-gender preferences in toys or play to those who felt they were the other gender, and even to those who began cross-sex hormones.

Some children express feelings of gender dysphoria when very young and the strength of these feelings lessen as they grow older. One such case is C.J., who used to draw himself as a girl. He played with dolls and liked “pink, purple and princesses”. At 4, he said he was going to be a woman when he grew up. At 6, he asked his parents to call him Rebecca and “her” (but after a while renounced, not feeling comfortable). Some professionals advised his mother to transition him socially, but the mother trusted her “mom guts” more. Now, at 11, C.J. writes “I feel like I’m a different type of boy. But I’m a boy for sure.” and when a friend transitioned he said he “couldn’t imagine being a girl every day“. “I do remember wanting to be a girl if I think about it really hard,” he adds, “but I don’t want to be a girl anymore”. Although still visibly gender nonconforming – or gender creative, as his mother prefers to say – the child grew into a more fluid identity in a way in which he became more accepting of the gender that corresponds with his natal sex.

Some other children express various degrees of gender dysphoria and then unexpectedly desist during puberty. Among these people is neuroscientist Debra Soh, who as a child was strongly opposed to feminine pastimes and playmates, and even urinary positions. Yet when she reached her late teens, “the idea of appearing feminine no longer repulsed [her]“. Even the strongest feelings of dysphoria can sometimes subside: in a BBC documentary, “Transgender Kids: Who Knows Best“, a girl named Alex remembers that she “wanted to be a boy” in childhood, while her father remembers her screaming “I’m a boy! I’m a boy!” Now she feels like a regular girl and presents in a feminine manner. The change happened at 12. In rare cases, it can happen even after social transition: Susie Green, chair of the UK transgender charity Mermaids, reports a second-hand account of the son of the former Mermaids chair, who “lived as a girl for three years… [then] when he realised that he was not female, he simply changed back.” For these children, the time and changes associated with puberty seem to be a deciding factor in whether they will desist or persist with their gender dysphoria. Their feelings were lengthy and lasting, yet changed as they reached and passed through puberty.

The most unusual case is that of an Australian boy named Patrick Mitchell, who was diagnosed at 12 with gender dysphoria. According to an Australian news site, Mitchell was, “Increasingly unhappy, suffering panic attacks and verging on depression, he told [his mother] if he could not go on puberty blockers he would run away and get them himself, or kill himself.” He took puberty blockers, then switched to estrogen (prescribed to his mother, because he was too young to obtain it legally) and started to grow breasts. But at 14, he changed his mind when he began to socially transition: “Teachers at school began to refer to him as a girl which triggered Mitchell to question if he had made the right decision”. This is another instance where puberty plays a role in determining what sex an individual feels comfortable as, physically.

Overall, these narratives show that gender identity is sometimes fluid and remains fluid through adolescence. Children may at one point be adamant they are the gender of their non-natal sex, yet later identify as the gender of their natal sex. Yet this does not give us reason to doubt these children’s’ sincerity or the validity of their identities at any point – but simply recognize that they can possibly change in the future, through no choice of the child.

Mental health and Social Transitioning

One of the primary factors in favor of increased support and validation for transgender children (which can include social transitioning and agentive decision making regarding puberty blockers) is the extent of mental health issues faced by children, youth, and adolescents with gender dysphoria. This includes an elevated risk of suicide, anxiety, and depression for transgender youth and adults. Additionally, sex reassignment doesn’t alleviate all negative mental health associations of transgender identities. Yet **there is also a growing body of evidence showing that increased family support and allowing social**

transitioning does have mental health benefits, though the strength of these is up for debate.

Rates of depressive and anxiety symptoms are elevated among the transgender community, with 51.4% of women and 48.3% of men having depressive symptoms, and 40.4% of women and 47.5% of men having anxiety symptoms [sample size 351] (Budge et al. 2013). Grossman and D'Augelli (2007) additionally found that out of 55 transgender youth (ages 15-21), 45% had suicidal ideations and 26% had a history of life-threatening behaviors. Children diagnosed with Gender Identity Disorder also had a significantly higher rate of anxiety according to reported negative emotions and skin conductance level (though not in cortisol or heart rate) – a sample which included 25 GID children and 25 age-similar controls from the Netherlands of which 36% of the GID children reached a clinical threshold for internalizing problems. Finally, Haas et al. (2014) found that out of over 6000 transgender and gender non-conforming respondents, there was a suicide attempt rate of 41% against a population average of 4.6% – and over a dozen surveys found between 25% and 43% suicide attempts among the trans community.

There is additionally evidence that sex reassignment doesn't alleviate all negative mental health associations of transgender identities. (Dhenje et al. 2011). Out of a sample of 324 sex reassigned persons in Sweden, there were significantly higher than rates of mortality, suicide, and psychiatric morbidity among transsexual individuals than among the general population.

Social Transitioning

Social transitioning (presenting socially as the gender of your non-natal sex, including using different pronouns, dressing differently, and appearing as the opposite gender) has been proposed by some (including Olson et al. 2016 in *Pediatrics*) researchers and therapists as a means of addressing the gender dysphoric desires of children with gender dysphoria. Many children express a desire to be or present as the other gender. Validating these desires and allowing them to present as the gender of their non-natal sex may help them feel more comfortable in their body.

Yet others including leading childhood gender dysphoria researchers (see Zucker 2018) believe that social transitioning may increase the chance that gender dysphoric feelings persist into adulthood. Given that gender dysphoria is associated with uncomfortableness with one's body as well as anxiety, depression and an increased rate of suicide, some posit that on the whole it would be better if children lost these feelings as they grew older. Others claim that a preference for desistence is transphobic and reminiscent of conversion camps and other psychotherapy methods that attempt to erase gay and trans identities.

There is a lack of evidence for the long-term effects of social transitioning. According to the studies on mental health though, family support and affirmation of identities can decrease the risk of mental health problems in gender dysphoric and transgender youth and adults. For example, social support and transition status decreased anxiety and depression, while transition status was negatively related with the same. Suicide attempters have reported more physical and verbal abuse from parents, while strong family relationships decreased the suicide rate (from above 50% with less contact and acceptance from family to 33% when family relationships remain strong). Olson et al. (2016) additionally found that social transitioning for children (73 sample size with two controls) led to them having typical rates of depression with only slightly elevated rates of anxiety – though these were measured through a parent proxy questionnaire. A similar study on trans children using age-similar and sibling controls found that when socially transitioned, they had normal rates of depression and only slightly elevated rates of anxiety – which differs from results found with non-socially transitioned trans youth.

There is no evidence yet that social transitioning increases the rate that gender dysphoria persists or desists.

Children are often the driving force behind social transitions, though the extent to which the possibility of social transition is considered in the first place is affected by family support and media exposure is debatable. For example, one mom reports that, “when he was six and asked us to call him by a girl’s name and use female pronouns.” An eight-year-old child took the initiative and, “sent an email to everyone at her primary school saying she was a girl trapped inside a boy’s body. After that, she started going to school dressed as a girl.”

Even as young as kindergarten, children are expressing desires that they are the other gender and would like to be identified as such. One parent tells a story of their child lining up for class: “That morning, they’d divided the kindergartners into two lines, boys and girls – and Coy had lined up with the girls. “You’re a boy,” the teacher had corrected. Coy had sobbed for the rest of the day. Even my teacher doesn’t know I’m a girl!” he wailed.” A three-year-old child expressed the idea that they, “[were] supposed to be born a girl, but [were] born a boy instead“. This child subsequently wanted to switch pronouns and change their name around a year later. These narratives support the idea that children do know and can express their desires regarding social transitioning (though these desires may change later).

There is additionally some evidence that social transitioning can serve as a tool to differentiate individuals on the strength of their cross-gender identity. For instance, one anecdotal account from a mother indicates that a child has expressed transgender desires and decided to socially transition and was supported in this decision, but didn't feel comfortable as the opposite gender and transitioned back.

One argument against social transition is that some children find it difficult to transition back to presenting as the gender corresponding to their natal sex. Though this varies based on the individual situation (see desistance narratives for examples of children who had minimal issue transitioning back), Steensma et al. (2011) found that some girls, who were almost (but not even entirely) living as boys in their childhood years, experienced great trouble when they wanted to return to the female gender role".

Overall, the evidence concerning social transition is mixed, with medium sized positive mental health effects present, but unknown consequences on the child's future gender identity and dysphoric feelings as well as possible difficulty transitioning back.

Social Constructions of Gender and Samoa

Evidence of better outcomes for those with larger amounts of social support may indicate that mental health problems (primarily anxiety, depression, and increased risk of suicide) associated with transgender and gender dysphoric individuals are more due to society's treatment of transgender individuals than due to gender dysphoria itself.

For an example of this theory in action we turn to Samoan culture, where a cross-gender identity (termed *fa’afafine*) is accepted and treated as normative. This study (Vasey and Bartlett 2007) examined whether Samoans with cross-gender identity (Fa’afafine) experience the same distress about gender identity that individuals in Western locales do. *Fa’afafine* are men who generally present as feminine and are almost exclusively sexually attracted to other men. The authors say that, “Most self-identify as fa’afafine, not as men. A minority self-identify as women”. They do not identify as gay or homosexual even though they almost exclusively are sexually attracted to other men.

The study examined 53 fa’afafine adults and 51 controls from similar Samoan contexts about their childhood behavior. It asked them whether they recalled, “(1) a strong and persistent cross-gender identification in childhood; (2) a sense of inappropriateness in the male-typical gender role; (3) a discomfort with their sex; or (4) distress associated with any of the above.” Most fa’afafine remember engaging in female-typical behaviors as children and no distress related to this behavior. Many believed they were girls as children and also don’t remember any distress about these feelings. They do remember negative feelings toward male roles and typical male activity as children, while some had negative feelings towards their genitals as children.

According to the authors of the study, we can assume many of children would have had GID as defined by DSM-IV-TR. Yet, these individuals do not remember distress about expressions of cross-gender identity. A small number do remember distress with their genitals. Similarly, fa’afafine do not report higher rates of bullying or victimization due to physical aggression.

This study of another cultural context provides evidence that transgender identity is a cultural-context dependent phenomenon, and that distress (and associated mental health problems) faced by transgender individuals are related to their treatment within society rather than to gender-atypical behavior and identity itself.

Puberty Blockers

Puberty blockers are another aspect of the youth gender transition process which is hotly debated. Adolescents are traditionally prescribed puberty blockers to limit them from going through puberty as their natal sex, as this can make it more difficult to physically transition to the other sex. It can additionally be traumatic for those who undergo puberty while strongly gender dysphoric.

There is some evidence that puberty blockers influence bones in negative ways. One study found that puberty blockers (GnRHa) led to statistically significant decreases in bone turnover during the time period in which they were applied. Another study indicates that bone mineral density is decreased significantly in both transwomen and transmen between the start of GnRHa application and age 22. The study concluded that “either attainment of peak bone mass has been delayed or peak bone mass itself is attenuated“. Clemons et al. (1993) additionally examined the effects of puberty blockers in non-trans instances and concludes they are safe and effective, after which puberty resumes normally. Yet they also point out potential problems regarding bone mineralization. Hruz et al., in a socially conservative publication, also posit risks of increased testicular cancer, obesity, memory loss, height decreases, and androgynous appearances. Overall there do appear to be risks to puberty blockers, and it is up to families to make the best decisions for themselves based on the potential consequences of adopting and not adopting blockers.

Additionally, if puberty is blocked, it cannot be used as a “diagnostic tool”, as Green refers to it. According to some researchers, (Steensma et al. 2011, Zucker 2018, narrative accounts of desistance) puberty is the stage at which many formerly dysphoric youth desist and begin expressing a gender identity in line with their biological sex. Yet others would push back against the use of puberty as a diagnostic tool due to the trauma it can cause to transgender individuals forced to go through puberty as their natal sex.

Narratives of puberty describe traumatic experiences for youth who experience gender dysphoria. If one feels their biological sex is wrong and they should have the opposite physical sex characteristics, those characteristics becoming more prominent can be extremely difficult. Coupled with the additional social pressures during puberty to conform to the gender that matches one’s biological sex, puberty can be difficult and scary for transgender individuals.

For instance, a mother reported her son named Patrick was, “Increasingly unhappy, suffering panic attacks and verging on depression, he told her if he could not go on puberty blockers he would run away and get them himself, or kill himself.” Another mom recalls that her child, Jackie, was “incredibly depressed” when she started puberty. The daughter was happy in elementary school, after social transitioning at 8, but “everything fell to pieces” when she started puberty, making six suicide attempts between the ages of eleven and fifteen. She overdosed and self-harmed with razor blades to distract from her changing body before being prescribed puberty blockers.

Steensma et al.’s (2011) qualitative study reports that, upon reaching puberty, “these anticipated or actual physical changes were often agonizing and highly distressful," while, "at the beginning of puberty, the aversion towards their bodies intensified immensely, resulting in insecurity and social withdrawal".

These psychological consequences of commencing puberty as a gender dysphoric child or youth must be weighed against any potential health effects when deciding about puberty blockers. A 2010 study of those placed on puberty blockers also indicates positive mental health effects of puberty blockers. From a T0 at the beginning of puberty blockers to a T1 around three years later, depressive symptoms significantly decreased, while scores on the internalizing problems also significantly decreased (from 29.6% to 11.1%). Trans boys did show still elevated levels of internalizing and externalizing problems but decreases from their previous rates. Overall, "Adolescents showed fewer behavioral and emotional problems, reported fewer depressive symptoms, feelings of anxiety and anger remained stable, and their general functioning improved."

One other aspect of this study is that out of a sample of 70 individuals, none desisted, and all continued to receive treatment for gender dysphoria. This could possibly indicate that puberty blockers decrease the chance that an individual's gender dysphoric feelings will go away, though it could also indicate that only those with strong feelings take puberty blockers in the first place.

Overall, puberty blockers appear to have some positive mental health effects due to their prevention of the physical experience of puberty among transgender youth, but may have physical health consequences including on bone growth.

Conclusion

On Desistence – The body of research on gender dysphoric youth indicates that many of these youths are no longer gender dysphoric upon reaching and progressing through puberty. It is possible that the youths that “desisted” either were not transgender in the first place or were pressured to disassociate from their transgender identity. Yet there are enough anecdotal accounts and extended studies regarding desisting youth to provide reasonable evidence that at least some, and likely a sizable fraction of individuals who express desires to be the gender opposite their natal sex or affirm that they are the opposite gender in childhood do not feel the same way at a later point in life.

On Determining Desistence – There is a lack of agreement between sources on whether gender dysphoric youth who will persist and desist can be differentiated from each other. On the whole, the evidence indicates there is a correlation between the strength of transgender expression (identifying as the other gender rather than simply expressing cross-gender behavior) and persistence, yet the correlation isn’t perfect. Some children who express a strong desire that they are the other gender desist, while some children with subthreshold gender dysphoria diagnoses persist through puberty.

On Social Transitioning – Social transitioning has positive mental health effects, but unknown effects on whether children will persist or desist. Social transitioning is put forward as one of the primary ways to support transgender children and is shown to reduce the rates of internalizing problems and anxiety among gender dysphoric youth. These children do still show elevated rates of anxiety. Additionally, some children who socially transition later do not feel gender dysphoric anymore and decide to not present as the opposite gender any longer. There is no evidence examining the rate at which individuals who socially transition retain their gender dysphoric feelings and transgender identities.

On Puberty Blockers – Puberty blockers, though reportedly safe, may have unintended medical consequences based on a review of studies. Studies show effects on bone growth and density. Yet other studies show positive mental health effects relative to transgender individuals who undergo puberty as their birth sex, as puberty is a time when living as a transgender individual can be particularly traumatic. It is unknown if placing individuals on puberty blockers affects the rate at which that population of individuals retains their gender dysphoria.

Notes

---

1. This does not indicate a preference for desistence of gender dysphoric youth, but merely indicates that these types of long-term effects are something policy makers, medical experts, and trans advocates may wish to consider. ↩
2. Dutch study. The numbers don't include nonresponders. ↩
3. Ph.D. thesis, University of Toronto. ↩
4. Dutch study. The numbers don't include nonresponders. ↩
5. 1 "Transsexual" and 1 "Homosexual, cross dresses"

   ↩
6. Canadian study, University of Toronto. ↩

In The Balance

When you first take the Artifact, you will see a vision of ALPHANION, Demon-Sultan of the Domain of Order, who appears as a grid of spheres connected by luminous lines. Alphanion will urge you to use the Artifact to enforce cosmic order, law at its most fundamental. He will show you visions of all the most brutal and sadistic crimes of history, of all the wars caused by nations that could not live together in harmony, and he will tell you they are all preventable. He will show you dreams of perfectly clean cities with wide open streets, where everyone earns exactly the optimal amount of money and public transportation is accurate to the second. He will tell you it is all attainable.

But if you hesitate even an instant to take Alphanion's offer, you will see a vision of CTHGHFZXAY, Demon-Shah of the Domain of Chaos, who appears as a shifting multicolored cloud. Cthghfzxay will urge you to use the Artifact to promote cosmic chaos, the ultimate principle of freedom. She will condemn the works of Order as a lie, a dystopia bought at the cost of true human liberty. She will show you visions of primaeval forests, where no two flowers are alike, where each glade holds a new mystery, where people run wild in search of new adventure. She will tell you it can all be yours.

As you weigh these two offers, you will see a vision of ZAMABAMAZ, Demon-Pharaoh of the Domain of Balance, who appears as a man and woman conjoined. They will tell you that neither Order nor Chaos is at the root of human flourishing, but an ability to strike the right balance between the two. That a virtuous life is one spent in moderation between total wild liberty and a stifling concept of rote rule-following. That Alphanion and Cthfhfzxay are the two poles of the universe, and that righteousness exists in the space created by their interaction. They will ask you to devote the Artifact and its power to the Domain of Balance, so all people can better manage the interaction of Order and Chaos in their own lives.

This will seem reasonable to you, but then there will appear a vision of IYYYYYYYYYYYYYYYYYYYYYYYY, Demon-Raja of the Domain of Excess, who appears as a blinding violet light. It will tell you that both Order and Chaos present coherent visions of the world, but that for the love of God, choose one or the other instead of being a wishy-washy milquetoast who refuses to commit to anything. It will tell you that blinding white and pitch black are both purer and more compelling than endless pointless grey. It will ask you to give the Artifact to somebody – anybody – other than Zamabamaz.

Just as you think you have figured all this out, there will appear a vision of MLOXO7W, Demon-Kaiser of the Domain of Meta-Balance, who appears as a face twisted into a Moebius strip. It will tell you that sometimes it is right to seek balance, and other times right to seek excess, and that a life well-lived consists of excess when excess is needed, and balance when balance is needed. It will remind you that sometimes you are a sprinter and other times a tightrope walker in the Olympiad of life, and that to commit to either eternal carefulness or eternal zealousness is to needlessly impoverish yourself. It will ask you to devote the Artifact and its power to balancing balance and imbalance, balancedly.

You will not be the least bit surprised when there will appear a vision of K!!!111ELEVEN, Demon-Shogun of the Domain of Meta-Excess, who appears as a Toricelli trumpet with eyes and a mouth. She says that seriously, pick a side, all this complicated garbage about the balance between balance and excess is just another layer of intellectualization to defend against having any real values, a trick to make you feel smart and superior for believing in nothing, not even Balance. She will ask you to choose something now, lest you be caught in an endless regress of further options.

As soon as you acknowledge that this makes sense, there will appear a vision of ILO, Demon-Chancellor of the Domain of Excessive Meta-Balance, who appears as a deep hole in space whose end you cannot see. They will point out that yes, there is potentially an infinite regress of further levels. But to act to avoid those levels is essentially to unthinkingly side with the principle of Excess over Balance. After all, if you had originally started by siding with Chaos or Order rather than waiting to hear of the existence of Balance, you would have been unknowingly favoring Excess over Balance. And if you had decided to choose either Excess or Balance, you would have been favoring the principle of Meta-Excess over Meta-Balance before even knowing they existed. So choosing at any level of the hierarchy is essentially equivalent to choosing Excess at all higher levels of the hierarchy. When viewed this way, the hierarchy collapses to chaos, order, first-level-balance, second-level-balance, third-level-balance, and so on. They offer a new, better vision: Infinite Balance, a theoretical top of the hierarchy in which you choose to balance all previous levels.

But as you start to consider this, there will appear a vision of PAHANUP, Demon-Taoiseach of the Domain of Balanced Meta-Balance, who appears as a hole in space exactly three inches deep. Ze will tell you that going to infinite lengths to ensure perfect balance at an infinite number of levels actually seems a bit excessive in ways. To choose either Chaos or Order outright would be insufficiently careful, but to give yourself an intractable problem with an endless number of meta-levels would be excessively careful. Ze will suggest seeking balance in the number of levels you seek balance in.

This will seem plausible to you right up until the sudden appearance of a fiery vision of IFNI, Demon-Secretary-General of the Domain of Chaotic Meta-Excess, who appears as static. She will point out that there is now another infinite regress, more difficult than the last – to wit, how long you should spend calculating the number of levels on which to seek balance. She will state her case thus: suppose you want to calculate the correct amount of balance in the universe. Let us call this Calculation A. You need to calculate how long to spend on this calculation before giving up and satisficing; let us call this Calculation B. But you need to calculate how long to spend on Calculation B before giving up and satisficing; let us call this Calculation C. Clearly you will never be able to complete any of the calculations. Therefore in order to avoid spending your entire life in an infinite regress of calculation, you should flip a coin right now and use it to decide either Chaos or Order, no takebacks.

But as you reach for the coin, you will see a vision of GOSAGUL, Demon-Admiral of the Domain of Ordered Meta-Balanced Excess, who appears as a cube with constantly flashing black and white faces. He will lecture you on how it seems pretty strange that, when faced with the most important decision in the history of the universe, you decide to flip a coin. Surely, even if Ifni’s argument is correct, you can do better than that! For example, you can just go a specific finite number of levels, such as three, then seek balance at that many levels, then stop. This will be strictly better than Ifni’s plan of choosing completely randomly.

But this sage advice is interrupted by MEGAHAHA, Demon-Pope of the Domain of Excessively Ordered Meta-Balance, who appear a as pattern of black and white that cycles between a line, square, cube, and hypercube. It will point out that if you’re in the business of accepting arguments along the lines that “it seems pretty strange that when faced with the most important decision in the history of the universe you…”, then it seems pretty strange that when faced with the most important decision in the history of the universe, you agree to a kind of random number of levels chosen by a demon you have no reason to trust. By what logic do you reject making the decision itself randomly, but accept making the decision about how many levels to make the decision on randomly? Any amount of Balance in Meta-Balancing Excess is just arbitrary capriciousness; you either need to act fully randomly, or embrace the entire difficulty of the problem.

At this point, you will remember that the Artifact is cursed and demons are evil. With a final effort of will, you will shout the words “I choose Balance! Just normal Balance! First-level Balance! That’s it!” and throw the Artifact to the ground, where it will shatter into a thousand pieces and the voices of the demonic hierarchy will suddenly all go silent.

And for a thousand years to come, heroes will grumble “Why, exactly, are we seeking balance in the universe? Isn’t that kind of dumb? Don’t we want more good stuff, and less bad stuff? Doesn’t really seem that balance is really what we’re after, exactly.”

And you will tell them the story of how once you found the Artifact that gave you mastery of the universe, and you refused to take more than about three minutes figuring out what to use it for, because that would have been annoying.

The Omnigenic Model As Metaphor For Life

The collective intellect is change-blind. Knowledge gained seems so natural that we forget what it was like not to have it. Piaget says children gain long-term memory at age 4 and don’t learn abstract thought until ten; do you remember what it was like not to have abstract thought? We underestimate our intellectual progress because every every sliver of knowledge acquired gets backpropagated unboundedly into the past.

For decades, people talked about “the gene for height”, “the gene for intelligence”, etc. Was the gene for intelligence on chromosome 6? Was it on the X chromosome? What happens if your baby doesn’t have the gene for intelligence? Can they still succeed?

Meanwhile, the responsible experts were saying traits might be determined by *a two-digit number* of genes. Human Genome Project leader Francis Collins estimated that there were “about twelve genes” for diabetes, and “all of them will be discovered in the next two years”. Quanta Magazine reminds us of a 1999 study which claimed that “perhaps more than fifteen genes” might contribute to autism. By the early 2000s, the American Psychological Association was a little more cautious, was saying intelligence might be linked to “dozens – if not hundreds” of genes.

The most recent estimate for how many genes are involved in complex traits like height or intelligence is approximately “all of them” – by the latest count, about twenty thousand. From this side of the veil, it all seems so obvious. It’s hard to remember back a mere twenty or thirty years ago, when people earnestly awaited “the gene for depression”. It’s hard to remember the studies powered to find genes that increased height by an inch or two. It’s hard to remember all the crappy p-hacked results that okay, we found the gene for extraversion, here it is! It’s hard to remember all the editorials in *The Guardian* about how since nobody had found the gene for IQ yet, genes don’t matter, science is fake, and Galileo was a witch.

And even remembering those times, they seem incomprehensible. Like, really? Only a few visionaries considered the hypothesis that the most complex and subtle of human traits might depend on more than one protein? Only the boldest revolutionaries dared to ask whether maybe cystic fibrosis was not the best model for the entirety of human experience?

This side of the veil, instead of looking for the “gene for intelligence”, we try to find “polygenic scores”. Given a person’s entire genome, what function best predicts their intelligence? The most recent such effort uses over a thousand genes and is able to predict 10% of variability in educational attainment. This isn’t much, but it’s a heck of a lot better than anyone was able to do under the old “dozen genes” model, and it’s getting better every year in the way healthy paradigms are supposed to.

Genetics is interesting as an example of a science that overcame a diseased paradigm. For years, basically all candidate gene studies were fake. “How come we can’t find genes for anything?” was never as popular as “where’s my flying car?” as a symbol of how science never advances in the way we optimistically feel like it should. But it could have been.

And now it works. What lessons can we draw from this, for domains that still seem disappointing and intractable?

Turn-of-the-millennium behavioral genetics was intractable because it was more polycausal than anyone expected. Everything interesting was an excruciating interaction of a thousand different things. You had to know all those things to predict anything at all, so nobody predicted anything and all apparent predictions were fake.

Modern genetics is healthy and functional because it turns out that although genetics isn't easy, it *is* simple. Yes, there are three billion base pairs in the human genome. But each of those base pairs is a nice, clean, discrete unit with one of four values. In a way, saying "everything has three billion possible causes" is a mercy; it's placing an upper bound on how terrible genetics can be. The "secret" of genetics was that there was no "secret". You just had to drop the optimistic assumption that there was any shortcut other than measuring all three billion different things, and get busy doing the measuring. The field was maximally perverse, but with enough advances in sequencing and computing, even the maximum possible level of perversity turned out to be within the limits of modern computing.

(this is an oversimplification: if it were really *maximally* perverse, chaos theory would be involved somehow. Maybe a better claim is that it hits the maximum perversity bound in one specific dimension)

One possible lesson here is that the sciences where progress is hard are the ones that have what seem like an unfair number of tiny interacting causes that determine everything. We should go from trying to discover "the" cause, to trying to find which factors we need to create the best polycausal model. And we should go from seeking a flash of genius that helps sweep away the complexity, to figuring out how to manage complexity that cannot be swept away.

Late-90s/early-00s psychiatry was a lot like late-90s/early-00s genetics. The public was talking about “the cause” of depression: serotonin. And the responsible experts were saying oh no, depression might be caused by as many as *several* different things.

Now the biopsychosocial model has caught on and everyone agrees that depression is complicated. I don’t know if we’re still at the “dozens of things” stage or the “hundreds of things stage”, but I don’t think anyone seriously thinks it’s fewer than a dozen. The structure of depression seems different from the structure of genetic traits in that one cause can still have a large effect; multiple sclerosis might explain less than 1% of the variance in depressedness, but there will be a small sample of depressives whose condition is almost entirely because of multiple sclerosis. But overall, I think the analogy to genetics is a good one.

If this is true, what can psychiatry (and maybe other low-rate-of-progress sciences) learn from genetics?

One possible lesson is: there are more causes than you think. Stop looking for “a cause” or “the ten causes” and start figuring out ways to deal with very numerous causes.

There are a bunch of studies that are basically like this one linking depression to zinc deficiency. They are good as far as they go, but it’s hard to really know what to do with them. It’s like finding one gene for intelligence. Okay, that explains 0.1% of the variability, now what?

We might imagine trying to combine all these findings into a polycausal score. Take millions of people, measure a hundred different variables – everything from their blood zinc levels, to the serotonin metabolites in their spinal fluid, to whether their mother loved them as a child – then do statistics on them and see how much of the variance in depression we can predict based on the inputs. “Do statistics on them” is a heck of a black box; genes are kind of pristine and causally unidirectional, but all of these psychological factors probably influence each other in a hundred different ways. In practice I think this would end up as a horribly expensive boondoggle that didn’t work at all. But in theory I think this is what a principled attempt to understand depression would look like.

(“understand depression” might be the wrong term here; it conflates being able to predict a construct with knowing what real-world phenomenon the construct refers to. We are much better at finding genes for intelligence than at understanding exactly what intelligence *is*, and whether it’s just a convenient statistical construct or a specific brain parameter. By analogy, we can imagine a Martian anthropologist who correctly groups “having a big house”, “driving a sports car”, and “wearing designer clothes” into a construct called “wealth”, and is able to accurately predict wealth from a model including variables like occupation, ethnicity, and educational attainment – but who doesn’t understand that wealth = having lots of money. I think it’s still unclear to what degree intelligence and depression have a simple real-world wealth-equals-lots-of-money style correspondence – though see here and here.)

A more useful lesson might be skepticism about personalized medicine. Personalized medicine – the idea that I can read your genome and your blood test results and whatever and tell you what antidepressant (or supplement, or form of therapy) is right for you has been a big idea over the past decade. And so far it’s mostly failed. A massively polycausal model would explain why. The average personalized medicine company gives you recommendations based on at most a few things – zinc levels, gut flora balance, etc. If there are dozens or hundreds of things, then you need the full massively polycausal model – which as mentioned before is computationally intractable at least without a lot more work.

(you can still have *some* personalized medicine. We don't have to know the causes of depression to treat it. You might be depressed because your grandfather died, but Prozac can still make you feel better. So it's possible that there's a simple personalized monocausal way to check who eg responds better to Prozac vs. Lexapro, though the latest evidence isn't really bullish about this. But this seems different from a true personalized medicine where we determine the root cause of your depression and fix it in a principled way.)

Even if we can't get much out of this, I think it can be helpful just to ask which factors and sciences are oligocausal vs. massively polycausal. For example, what percent of variability in firm success are economists able to determine? Does most of the variability come from a few big things, like talented CEOs? Or does most of it come from a million tiny unmeasurable causes, like "how often does Lisa in Marketing get her reports in on time"?

Maybe this is really stupid – I'm neither a geneticist or a statistician – but I imagine an alien society where science is centered around polycausal scores. Instead of publishing a paper claiming that lead causes crime, they publish a paper giving the latest polycausal score for predicting crime, and demonstrating that they can make it much more accurate by including lead as a variable. I don't think you can do this in real life – you would need bigger Big Data than anybody wants to deal with. But like falsifiability and compressability, I think it's a useful thought experiment to keep in mind when imagining what science *should* be like.

Book Review: The Black Swan

I.

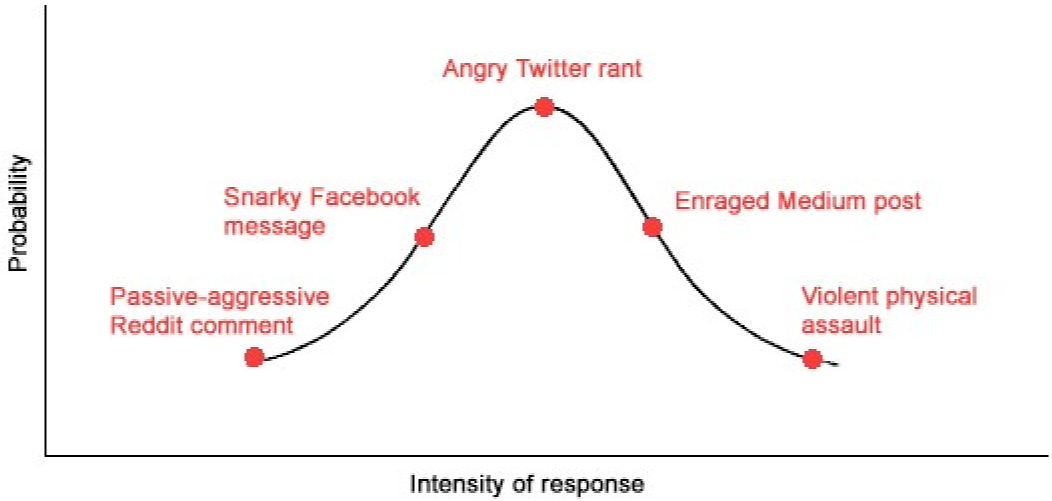
Writing a review of *The Black Swan* is a nerve-wracking experience.

First, because it forces me to reveal I am about ten years behind the times in my reading habits.

But second, because its author Nassim Nicholas Taleb is infamous for angry Twitter rants against people who misunderstand his work. Much better men than I have read and reviewed *Black Swan*, messed it up, and ended up the victim of Taleb’s acerbic tongue.

One might ask: what’s the worst that could happen? A famous intellectual yells at me on Twitter for a few minutes? Isn’t that normal these days? Sure, occasionally Taleb will go further and write an entire enraged Medium article about some particularly egregious flub, but only occasionally. And even that isn’t so bad, is it?

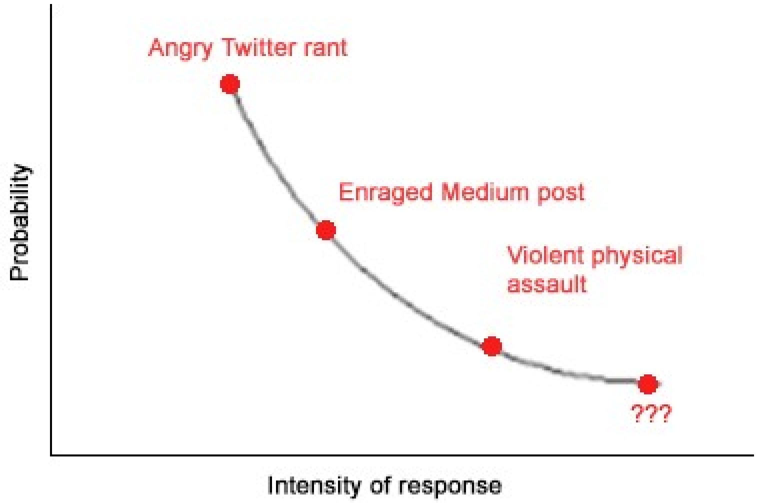
But such an argument betrays the following underlying view:



It assumes that events can always be mapped onto a bell curve, with a peak at the average and dropping off quickly as one moves towards extremes. Most reviews of *Black Swan* will get an angry Twitter rant. A few will get only a snarky Facebook post or an entire enraged Medium article. By the time we get to real extremes in either directions – a mere passive-aggressive Reddit comment, or a dramatic violent assault – the probabilities are so low that they can safely be ignored.

Some distributions really do follow a bell curve. The classic example is height. The average person is about 5’7. The likelihood of anyone being a different height drops off dramatically with distance from the mean. Only about one in a million people should be taller than 7 feet; only one in a billion should be as tall as 7’5. Nobody is order-of-magnitude differences in height from anyone else. Taleb calls the world of bell curves and minor differences Mediocristan. If Taleb’s reaction to bad reviews dwells alongside height in Mediocristan, I am safe; nothing an order-of-magnitude difference from an angry Twitter rant is likely to happen in entire lifetimes of misinterpreting his work.

But other distributions are nothing like a bell curve. Taleb cites power-law distributions as an example, and calls their world Extremistan. Wealth inequality lives in Extremistan. If wealth followed a bell curve around the median household income of $57,000, and a standard deviation scaled the same way as height, then a rich person earning $70,000 would be as remarkable as a tall person hitting 7 feet. Someone who earned $76,000 would be the same kind of prodigy of nature as the 7’6 Yao Ming. Instead, people earning $70,000 are dirt-common, some people earn millions, and the occasional tycoon can make hundreds of millions of dollars per year. In Mediocristan, the extremes don’t matter; in Extremistan, sometimes *only* the extremes matter. If you have a room full of 99 average-height people plus Yao Ming, Yao only has 1.3% of the total height in the room. If you have a room full of 99 average-income people plus Jeff Bezos, Bezos has 99.99% of the total wealth.

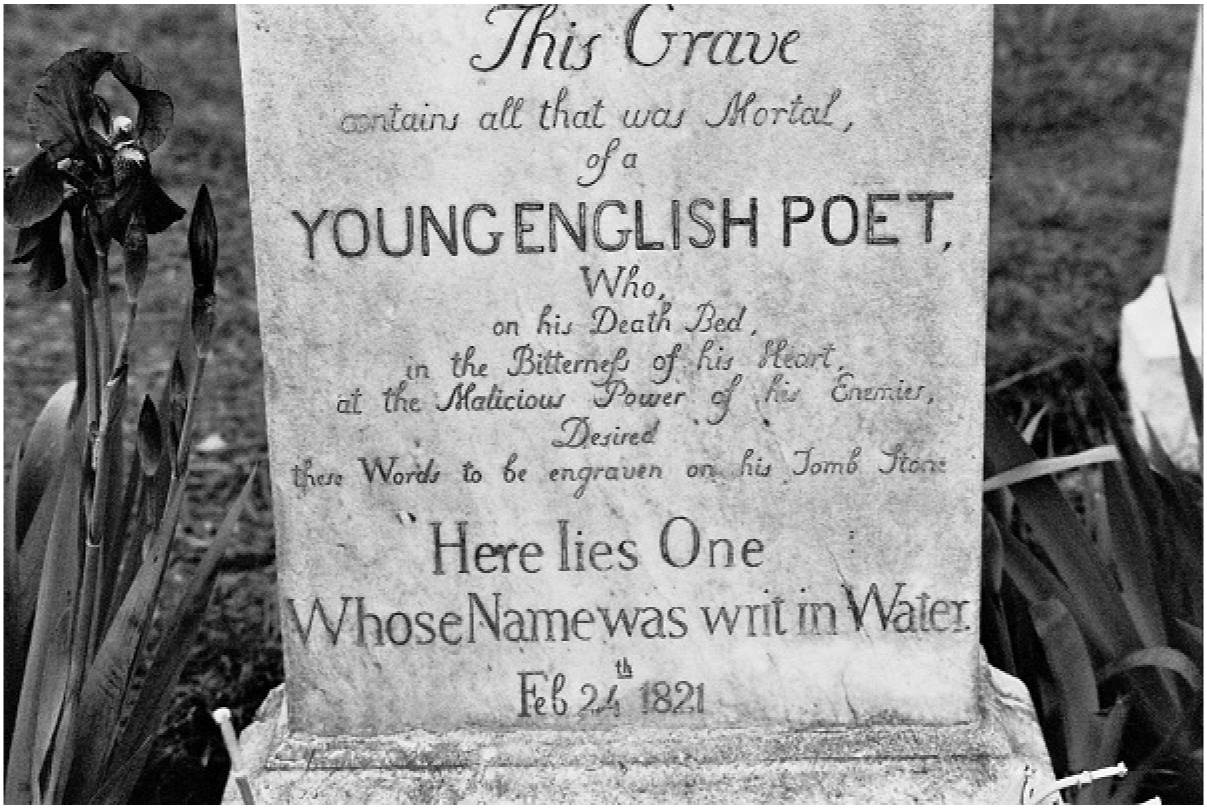


Here are Taleb’s potential reactions graphed onto a power-law distribution. Although the likelihood of any given reaction continues to decline the further it is away from average, it declines much less quickly than on the bell curve. Violent assault is no longer such a remote possibility; maybe my considerations should even be dominated by it.

So: are book reviews in Mediocristan or Extremistan?

I notice this BBC article about an author who hunted down a bad reviewer of his book and knocked her unconscious with a wine bottle. And Lord Byron wrote such a scathing meta-review of book reviewers that multiple *reviewers* challenged *him* to a duel, but the duel seems to have never taken place, plus I’m not sure Lord Byron is a good person to generalize from.

19th century intellectuals believed a bad review gave John Keats tuberculosis; they were so upset about this that they used his gravestone to complain:



This Grave
contains all that was Mortal,
of a
YOUNG ENGLISH POET,
Who,
on his Death Bed,
in the Bitterneſs of his Heart,
at the Malicious Power of his Enemies,
Desired
these Words to be engraven on his Tomb Stone
"Here lies One
Whose Name was writ in Water.
Feb 24th 1821

Keats’ friend Shelley wrote the poem *Adonais* to memorialize the event, in which he said of the reviewer:

> Our Adonais has drunk poison—oh!
> What deaf and viperous murderer could crown

> Life’s early cup with such a draught of woe?
> The nameless worm would now itself disown:
> It felt, yet could escape, the magic tone
> Whose prelude held all envy, hate and wrong,
> But what was howling in one breast alone,
> Silent with expectation of the song,
> Whose master’s hand is cold, whose silver lyre unstrung.

So are book reviews in Mediocristan or Extremistan? Well, every so often your review causes one of history’s greatest poets to die of tuberculosis, plus another great poet writes a five-hundred-line poem condemning you and calling you a “nameless worm”, and it becomes a classic that gets read by millions of schoolchildren each year for centuries after your death. And that’s just the worst thing that’s happened because of a book review *so far*. The next one could be even worse!

II.

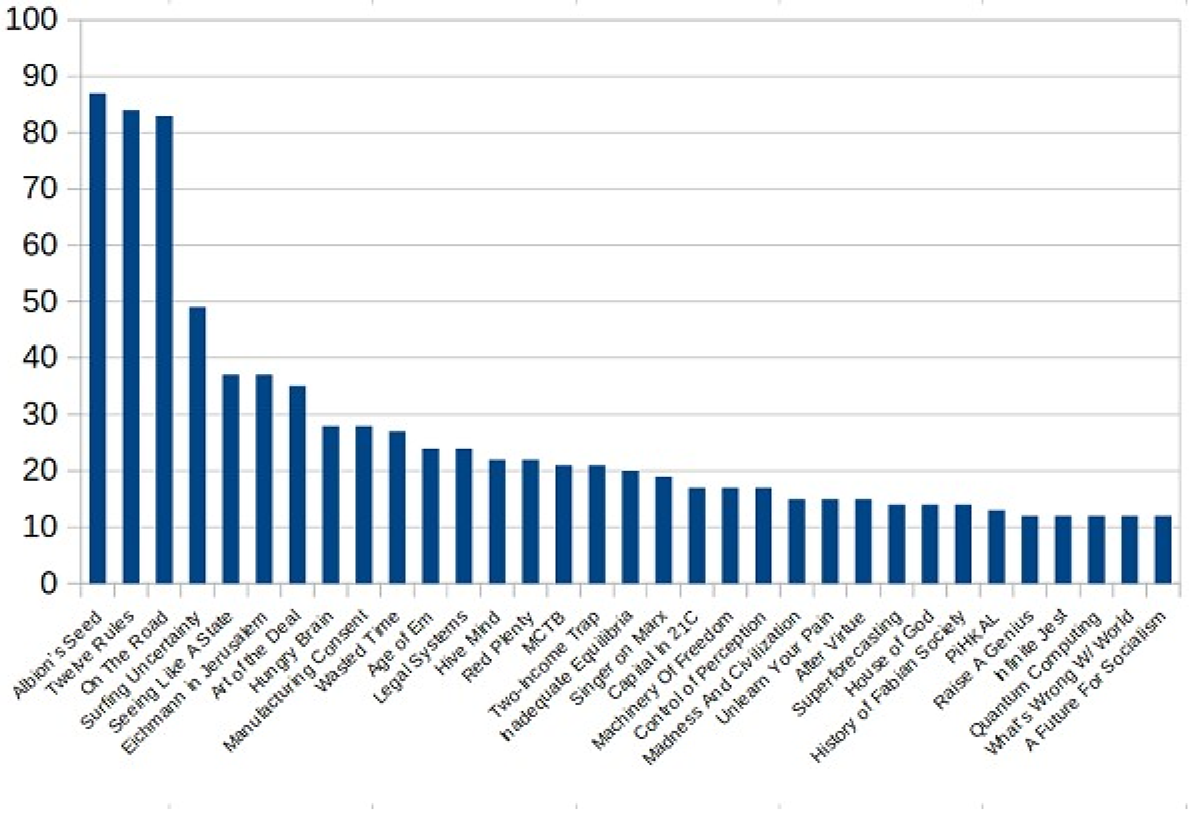
This sounds like maybe an argument for inaction, but Taleb is more optimistic. He points out that black swans are often good. For example, pharma companies usually just sit around churning out new antidepressants that *totally aren’t* just SSRI clones *they swear*. If you invest in one of these companies, you may win a bit if their SSRI clone succeeds, and lose a bit if it fails. But drug sales fall on a power law; every so often companies get a blockbuster that lets them double, triple, or dectuple their money. Tomorrow a pharma company might discover the cure for cancer, or the cure for aging, and get to sell it to everyone forever. So when you invest in a pharma company, you have randomness on your side: the worst that can happen is you lose your money, but the best that can happen is multiple-order-of-magnitude profits.

Taleb proposes a “barbell” strategy of combining some low-risk investments with some that expose you to positive black swans:

> If you know that you are vulnerable to prediction errors, and if you accept that most “risk measures” are flawed, because of the Black Swan, then your strategy is to be as hyperconservative and hyperaggressive as you can be instead of being mildly aggressive or conservative. Instead of putting your money in “medium risk” investments (how do you know it is medium risk? by listening to tenure-seeking “experts”?), you need to put a portion, say 85 to 90 percent, in extremely safe instruments, like Treasury bills—as safe a class of instruments as you can manage to find on this planet. The remaining 10 to 15 percent you put in extremely speculative bets, as leveraged as possible (like options), preferably venture capital-style portfolios.* That way you do not depend on errors of risk management; no Black Swan can hurt you at all, beyond your “floor,” the nest egg that you have in maximally safe investments. Or, equivalently, you can have a speculative portfolio and insure it (if possible) against losses of more than, say, 15 percent. You are “clipping” your incomputable risk , the one that is harmful to you. Instead of having medium risk, you have high risk on one side and no risk on the other. The average will be medium risk but constitutes a positive exposure to the Black Swan […]
>
> The “barbell” strategy [is] taking maximum exposure to the positive Black Swans while remaining paranoid about the negative ones. For your exposure to the positive Black Swan, you do not need to have any precise understanding of the structure of uncertainty. I find it hard to explain that when you have a very limited

> loss you need to get as aggressive, as speculative, and sometimes as “unreasonable” as you can be.

So: how good can a book review get?



Here’s a graph of all the book reviews I’ve ever done by hit count (in thousands). I’m not going to calculate out, but it looks like a power law distribution! Some of my book reviews have been pretty successful – my review of *Twelve Rules* got mentioned in *The Atlantic*. Can things get even better than that? I met my first serious girlfriend through a blog post. Can things get even better than *that?* I had someone tell me a blog post on effective altruism convinced them to pledge to donate 10% of their salary to efficient charities forever; given some conservative assumptions, that probably saves twenty or thirty lives. So a book review has a small chance of giving a great poet tuberculosis, but also a small chance of saving dozens of lives. Overall it probably seems worth it.

III.

The Black Swan uses discussions of power laws and risk as a jumping-off point to explore a wider variety of topics about human fallibility. This places it in the context of similar books about rationality and bias that came out around the same time. I'm especially thinking of Philip Tetlock's *Superforecasting*, Nate Silver's *The Signal And The Noise*, Daniel Kahneman's *Thinking Fast And Slow*, and of course *The Sequences*. *The Black Swan* shares much of its material with these – in fact, it often cites Kahneman and Tetlock approvingly. But aside from the more in-depth discussion of risk, I notice two important points of this book Taleb keeps coming back to again and again, which as far as I know are unique to him.

The first is "the ludic fallacy", the false belief that life works like a game or a probability textbook thought experiment. Taleb cautions against the (to me tempting) mistake of comparing black swans to lottery tickets – ie "investing in pharma companies is like having a lottery ticket to win big if they invent a blockbuster". The lottery is a game where you know the rules and probabilities beforehand. The chance of winning is whatever it is. The prize is whatever it is. You know both beforehand; all you have to do is crunch the numbers to see if it's a good deal.

Pharma – and most other real-life things – are totally different. Nobody hands you the chance of a pharma company inventing a blockbuster drug, and nobody hands you the amount of money you'll win if it does. There is Knightian uncertainty – uncertainty about how much uncertainty there is, uncertainty that doesn't come pre-quantified.

Taleb gives cautionary examples of what happens if you ignore this. You make some kind of beautiful model that tells you there's only a 0.01% chance of the stock market doing some particular bad thing. Then you invest based on that data, and the stock market does that bad thing, and you lose all your money. You were taking account of the quantified risk in your model, but not of the unquantifiable risk that your model was incorrect.

In retrospect, this is an obvious point. But it's also obvious in retrospect that everything classes teach about probability fall victim to it, to the point where it's hard to even think about probability in non-ludic terms. I keep having to catch myself writing some kind of "Okay, assume the risk of a Black Swan is 10%…" example in this review, because then I *know* Taleb will hunt me down and violently assault me. But it's hard to resist.

I would like to excuse myself by saying it's impossible to discuss probability without these terms, or at least that you have to start by teaching these terms and then branch into the real-world unquantifiable stuff, except that Taleb managed to write his book without doing either of those things. Granted, the book is a little bit weird. You could go through several chapters on the Lebanese Civil War or whether the French Third Republic had the best intellectuals, without noticing it was a book on probability. Nevertheless, it sets itself the task of discussing risk without starting with the ludic fallacy, and it succeeds.

I don't know to what degree the project of "becoming well-calibrated with probabilities" is a solution to the ludic fallacy, or a case of stupidly falling victim to the ludic fallacy.

The second key concept of this book – obviously not completely original to Taleb, but I think Taleb gives it a new meaning and emphasis – is “Platonicity”, the anti-empirical desire to cram the messy real world into elegant theoretical buckets. Taleb treats the bell curve as one of the clearest examples; it’s a mathematically beautiful example of what certain risks *should* look like, so incompetent statisticians and economists assume that risks in a certain domain *do* fit the model.

He ties this into Tetlock’s “fox vs. hedgehog” dichotomy. The prognosticators who tried to fit everything to their theory usually did badly; the ones who accepted the complexity of reality and maintained a toolbox of possibilities usually did better.

He also mentions – and somehow I didn’t know this already – that modern empiricism descends from Sextus Empiricus, a classical doctor who popularized skeptical and empirical ideas as the proper way to do medicine. Sextus seems like a pretty fun guy; his surviving works include *Against The Grammarians*, *Against The Rhetoricians*, *Against The Geometers*, *Against The Arithmeticians*, *Against The Astrologers*, *Against The Musicians*, *Against The Logicians*, *Against The Physicists*, and *Against The Ethicists*. Medicine is certainly a great example of empiricism vs. Platonicity, with Hippocrates and his followers cramming everything into their preconceived Four Humors model – to the detriment of medical science – for thousands of years.

But Empiricus’ solution – to not hold any beliefs, and to act entirely out of habit – falls short. And I am not sure I understood what Taleb is arguing for here. There’s certainly a true platitude in this area (wait, does “platitude” share a root with “Platonic”? It looks like both go back to a Greek word meaning “broad”, but it’s not on purpose. Whatever.) of “try to go where the evidence guides you instead of having prejudices”. But there’s also a point on the other side – unless you have some paradigm to guide you, you exist in a world of chaotic noise. I am less sanguine than Taleb that “be empiricist, not theoretical” is sufficient advice, as opposed to “find the Golden Mean between empiricism and theory” – which is of course a much harder and more annoying adage, since finding a Golden Mean isn’t trivial.

That is – what would it mean for a doctor to try to do medicine without the “theory” that the heart pumped the blood? You’d find a patient with all the signs of cardiogenic shock, and say “Eh, I dunno. Maybe I should x-ray his feet, or something?” What if she had no preconceived ideas at all? Would she start reciting Sanskrit poetry, on the grounds that there’s no reason to think that would help more or less than anything else? Whereas a doctor who had read a lot of medical school textbooks – Taleb hates textbooks! – would immediately recognize the signs of cardiogenic shock, do the tests that the textbooks recommend, and give the appropriate treatment.

Yes, eventually an empiricist doctor would notice empirical facts that made her believe the heart pumped blood (and all the other true things). But then she would…write it down in a textbook. That’s what theories are – crystallized, compressed empiricism.

I think maybe Empiricus and Taleb would retort that some people form theories with only a smidgeon of evidence – I don’t know what evidence Hippocrates had for the Four Humors, but it clearly wasn’t enough. And then they stick to them dogmatically even when the evidence contradicts them. I agree with both criticisms. But then it seems like the problem is bad theories, rather than ever having theories at all. Four Humors Theory and Germ Theory are both theories – it's just that one is wrong and the other is right. If nobody had ever been willing to accept the germ theory of disease, we'd be in a much worse place. And you can't just say "Well, you could atheoretically notice antibiotics work and use them empirically" – much of the research into antibiotics, and the ways we use antibiotics, are in place because we more or less understand what they're doing.

I would argue that Empiricus and Taleb are arguing not for experience over theory, but for the adjustment of certain parameters of inference – how much fudge factor we accept in compressing our data, how much we weight prior probabilities versus new evidence, how surprised to be at evidence that doesn't fit our theories. I expect Empiricus, Taleb, and I are in agreement about which direction we want those parameters shifted. I know this sounds like a boring intellectual semantic point, but I think it's important and occasionally saves your life if you're practicing some craft like medicine that has a corpus of theory built up around it which you ignore at your peril.

(I also think they fail to understand the degree to which common sense is just under-the-hood inference in the same way that abstract theorizing is above-the-hood inference, and so doesn't rescue us from these concerns).

Charitably, *The Black Swan* isn't making the silly error of denying a Golden Mean of parameter position. It's just arguing that most people today are on the too-Platonic side of things, and so society as a whole needs to shift the parameters toward the more-empirical side. Certainly this is true of most people in the world Nassim Nicholas Taleb inhabits. In Taleb's world famous people walk around all day asserting “Everything is on a bell curve! Anyone who thinks risk is unpredictable is a dangerous heretic!” Then Taleb breaks in past their security cordon and shouts “But what if things aren’t on a bell curve? What if there are black swans?!” Then the famous person has a rage-induced seizure, as their bodyguards try to drag Taleb away. Honestly it sounds like an exciting life.

Lest you think I am exaggerating:

> The psychologist Philip Tetlock (the expert buster in Chapter 10) , after listening to one of my talks, reported that he was struck by the presence of an acute state of cognitive dissonance in the audience. But how people resolve this cognitive tension, as it strikes at the core of everything they have been taught and at the methods they practice, and realize that they will continue to practice, can vary a lot. It was symptomatic that almost all people who attacked my thinking attacked a deformed version of it, like “it is all random and unpredictable” rather than “it is largely random,” or got mixed up by showing me how the bell curve works in some physical domains. Some even had to change my biography. At a panel in Lugano, Myron Scholes once got in to a state of rage, and went after a transformed version of my ideas. I could see pain in his face. Once, in Paris, a prominent member of the mathematical establishment, who invested part of his life on some minute sub-sub-property of the Gaussian, blew a fuse—right when I showed empirical evidence of the role of Black Swans in markets. He turned red with anger, had difficulty breathing, and started hurling insults at me for having desecrated the institution, lacking pudeur (modesty); he shouted “I am a member of the Academy of Science!” to give more strength to his insults.

One hazard of reviewing books long after they come out is that, if the book was truly great, it starts sounding banal. If its points were so devastating and irrefutable that they became universally accepted, then it sounds like the author is just spouting cliches. I think *The Black Swan* might have reached that level of influence. I haven't even bothered explaining the term "black swan" because I assume every educated reader now knows what it means. So it seems very possible that pre-book society was so egregiously biased toward the Platonic theoretical side that it needed someone to tell it to shift in the direction of empiricism, Taleb did that, and now he sounds silly because *everyone knows* that you can't just declare everything a bell curve and call it a day. Maybe this book should be read backwards. But the nature of all mental processes as a necessary balance between theory and evidence is my personal hobby-horse, just as evidence being good and theory being bad is Taleb's personal hobby-horse, so I can't let this pass without at least one hobby-cavalry-duel.

I have a more specific worry about skeptical empiricism, which is that it seems like an especially dangerous way to handle Extremistan and black swans.

Taleb memorably compares much of the financial world to "picking up pennies in front of a steamroller" – ie, it is very easy to get small positive returns most of the time as long as you expose yourself to horrendous risk.

EG: imagine living in sunny California and making a bet with your friend about the weather. Each day it doesn't rain, he gives you $1. Each day it rains, you give him $1000. Your friend will certainly take this bet, since long-term it pays off in his favor. But for the first few months, you will look pretty smart as you pump him out of a constant stream of free dollars. Your stupidity will only become apparent way down the line, when one of the state’s rare rainstorms arrives and you’re on the hook for much more than you won.

Here the theorist will calculate the probability of rain, calculate everybody’s expected utility, and predict that your friend will eventually come out ahead.

But the good empiricist will just watch you getting a steady stream of free dollars, and your friend losing money every day, and say that you did the right thing and your friend is the moron!

More generally, as long as Black Swans are rare enough not to show up in your dataset, empiricists are likely to fall for picking-pennies-from-in-front-of-steamroller bets, whereas (sufficiently smart) theorists will reject them.

For example, Banker 1 follows a strategy that exposes herself terribly to black swan risk, and ensures she will go bankrupt as soon as the market goes down, but which makes her 10% per year while the market is going up. Banker 2 follows a strategy that protects herself against black swan risk, but only makes 8% per year while the market is going up. A naive empiricist will judge them by their results, see that Banker 1 has done better each of the past five years, and give all his money to Banker 1, with disastrous results. Somebody who has a deep theoretical understanding of the underlying territory might be able to avoid that mistake.

This problem also comes up in medicine. Imagine two different drugs. Both cure the same disease and do it equally well. Drug 1 has a side effect of mild headache in 50% of patients. Drug 2 has a side effect of death in 0.01% of patients. I think a lot of doctors test both drugs, find that Drug 2 always results in less hassle and happier patients, and stick with it. But this is plausibly the wrong move and a good understanding of the theory would make them much more cautious.

(yes, both of these examples are also examples of the ludic fallacy. I fail, sorry.)

Overall this seems like a form of Goodhart's Law, where any attempt to measure something empirically risks having people optimize for your measurement in a way that makes all the unmeasurable things worse. Black swan risks are one example of an unmeasurable thing; you can't really measure how common or how bad they are until they happen. If you focus entirely on empirical measurement, you'll incentivize people to take any trade that improves ordinary results at the cost of increasing black swan risk later. If you want to prevent that, you need a model that includes the possibility of black swan risk – which is going to involve some theory.

Nassim Taleb has been thinking about this kind of thing his whole life and I'm sure he hasn't missed this point. Probably we are just using terms differently. But I do think the way he uses terms minimizes concern about this type of error, and I do worry the damage can sometimes be pretty large.

IV.

I previously mentioned that *The Black Swan* seems to stand in the tradition of other rationality books like *Thinking Fast And Slow* and *The Signal And The Noise*. Is this a fair analysis? If so, what do we make of this tradition?

While Taleb has nothing but praise for eg Kahneman, his book also takes a very different tone. For one thing, it's part-autobiography / diary / vague-thought-log of Taleb, who is a very interesting person. I read some reviews saying he “needed an editor”, and I understand the sentiment, but – does he? Yes, his book is weird and disconnected. It’s also really fun to read, and sold three million copies. If people who “need an editor” often sell more copies than people who don’t, and are more enjoyable, are we sure we’re not just arbitrarily demanding people conform to a certain standard of book-writing that isn’t really better than alternative standards? Are we sure it’s really true that you can’t just stick several chapters about the biography of a fake Russian author into the middle of your book for no reason, without admitting that it’s fake? Are you sure you can’t insert a thinly-disguised version of yourself into the story about the Russian author, have yourself be such a suave and attractive individual that she falls for you and you start a torrid love affair, and then make fun of her cuckolded husband, who is suspiciously similar to the academics you despise? Are you *sure* this is an inappropriate thing to do in the middle of a book on probability? Maybe Nate Silver would have done it too if he had thought of it first.

Also sort of surprising: Taleb hates nerds. He explains:

> To set the terminology straight, what I call “a nerd” here doesn’t have to look sloppy, unaesthetic, and sallow, and wear glasses and a portable computer on his belt as if it were an ostensible weapon. A nerd is simply someone who thinks exceedingly inside the box.
>
> Have you ever wondered why so many of these straight-A students end up going nowhere in life while someone who lagged behind is now getting the shekels, buying the diamonds, and getting his phone calls returned? Or even getting the Nobel Prize in a real discipline (say, medicine)? Some of this may have something to do with luck in outcomes, but there is this sterile and obscurantist quality that is often associated with

> classroom knowledge that may get in the way of understanding what’s going on in real life. In an IQ test, as well as in any academic setting (including sports), Dr. John would vastly outperform Fat Tony. But Fat Tony would outperform Dr. John in any other possible ecological, real-life situation. In fact, Tony, in spite of his lack of culture, has an enormous curiosity about the texture of reality, and his own erudition—to me, he is more scientific in the literal, though not in the social, sense than Dr. John.

Going after nerds in your book contrasting Gaussian to power law distributions, with references to the works of Poincaré and Popper, is a bold choice. It also separates Taleb from the rest of the rationality tradition. I interpret eg *The Signal And The Noise* as pro-nerd. Its overall thesis is “Ordinary people are going around being woefully biased about all sorts of things. Good thing that bright people like Nate Silver can use the latest advances in statistics to figure out where they are going wrong, do the hard work of processing the statistical signal correctly, and create a brighter future for all of us.” Taleb turns that on its head. For him, ordinary people – taxi drivers, barbers, vibrant salt-of-the-earth heavily-accented New Yorkers – are the heroes, who know what’s up and are too sensible to go around saying that everything must be a bell curve, or that they have a clever theory which proves the market can never crash. It’s only the egghead intellectuals who could make such an error.

I am not sure this is true – my last New York taxi driver spent the ride explaining to me that he was the Messiah, which seems like an error on some important axis of reasoning that most intellectuals get right. But I understand that some of Taleb’s later works – *Antifragile* and *Skin In The Game* – may address more of what he means by this. It looks like Kahneman, Silver, et al are basically trying to figure out what doing things optimally would look like – which is a very nerdy project. Taleb is trying to figure out how to run systems without an assumption that you will necessarily be right very often.

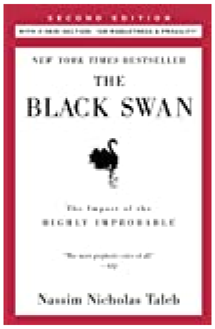
I am reminded of the example of doctors being asked probability questions, about whether a certain finding on a mammogram implies X probability of breast cancer. The doctors all get this horribly wrong, because none of them ever learned anything about probability. But after getting every question on the test wrong, they will go and perform actions which are basically optimized for correctly diagnosing and treating breast cancer, even though their probability-related answers imply they should do totally different things.

I see Kahneman, Tetlock, Silver, and Yudkowsky as all being in the tradition of finding optimal laws of probability that point out why the doctors are wrong, and figuring out how to train doctors to answer probability questions right. I see Taleb as being on the side of the doctors – trying to figure out a system where the right decisions get made whether anyone has a deep mathematical understanding of the situation or not. Taleb appreciates the others' work – you have to know something about probability before you can discuss why some systems tend towards getting it right vs. getting it wrong – but overall he agrees that "rationality is about winning" – the doctor who eventually gives the right treatment is better than a statistician who answers all relevant math questions correctly but has no idea what to do.

Relatedly, I think Taleb's critique of nerds works because he's trying to resurrect a Greco-Roman concept of the intellectual – *arete* and *mens sana in corpore sano* and all that – and clearly uses “nerd” to mean everything about modern faux-intellectuals that falls short of his vision. Thales cornering the market on olive presses is his kind of guy, and he doesn’t think that all of the people who have rage-induced seizures when he whispers the phrase “power law distribution” in their ears really cut it. His book is both a discussion of his own area of study (risk), and a celebration of and guide to what he thinks intellectualism should be. I might have missed the section of Marcus Aurelius where he talks about how angry Twitter rants are a good use of your time, but aside from that I think the autobiographical parts of the book make a convincing aesthetic argument that Taleb is living the dream and we should try to live it too.

Perhaps relating to this, of Taleb, Silver, Tetlock, Yudkowksy, and Kahneman, Taleb seems to have stuck around longest. All of them continue to do great object-level work in their respective fields, but it seems like the “moment” for books about rationality came and passed around 2010. Maybe it’s because the relevant science has slowed down – who is doing Kahneman-level work anymore? Maybe it’s because people spent about eight years seeing if knowing about cognitive biases made them more successful at anything, noticed it didn’t, and stopped caring. But reading *The Black Swan* really does feel like looking back to another era when the public briefly became enraptured by human rationality, and then, after learning a few cool principles, said “whatever” and moved on.

Except for Taleb. I’m excited to see he’s still working in this field and writing more books expanding on these principles. I look forward to reading the other books in this series.



SECOND EDITION

WITH A NEW SECTION: "ON ROBUSTNESS & FRAGILITY"

NEW YORK TIMES BESTSELLER

THE BLACK SWAN

The Impact of the HIGHLY IMPROBABLE

Nassim Nicholas Taleb

Treating The Prodrome

A prodrome is an early stage of a condition that might have different symptoms than the full-blown version. In psychiatry, the prodrome of schizophrenia is the few-months-to-few-years period when a person is just starting to develop schizophrenia and is acting a little bit strange while still having some insight into their condition.

There's a big push to treat schizophrenia prodrome as a critical period for intervention. Multiple studies have suggested that even though schizophrenia itself is a permanent condition which can be controlled but never cured, treating the prodrome aggressively enough can prevent full schizophrenia from ever developing at all. Advocates of this view compare it to detecting early-stage cancers, or getting prompt treatment for a developing stroke, or any of the million other examples from medicine of how you can get much better results by catching a disease very early before it has time to do damage.

These models conceptualize psychosis as "toxic" – not just unpleasant in and of itself, but damaging the brain while it's happening. They focus on a statistic called Duration of Untreated Psychosis. The longer the DUP, the more chance psychosis has had to damage the patient before the fire gets put out and further damage is prevented. Under this model it's vitally important to put people who seem to be getting a little bit schizophrenic on medications as soon as possible.

There has been a lot of work on this theory, but not a lot of light has been shed. Observational studies testing whether duration of untreated psychosis correlates with poor outcome mostly find it does a little bit, but there's a lot of potential confounding – maybe lower-class uneducated people take longer to see a psychiatrist, or maybe people who are especially psychotic are especially bad at recognizing they are psychotic. The relevant studies try their hardest to control for these factors, but remember that this is harder than you think. The randomized controlled trials of what happens if you intervene earlier in psychosis tend to do very badly and rarely show any benefit, but randomly intervening earlier in psychosis is hard, especially if you also need an ethics board's permission to keep a control group of other people who you are not going to intervene early on. Overall I could go either way on this.

Previously I was leaning toward "probably not relevant", just because it's *too convenient*. There is a lot of debate about how aggressively to treat schizophrenia, with mainstream psychiatry (and their friends the pharma companies) coming down on the side of "more aggressively", while other people point out that antipsychotics have lots of side effects and their long-term effects (both how well they work long-term, and what negative effects they have long-term) are poorly understood. These people tend to come up with kind of wild theories about how long-term antipsychotics hypersensitize you and make you worse. I don't currently find these very credible, but I'm also skeptical of things that are too convenient to the mainstream narrative, like "unless you treat every case of schizophrenia *right away* you are exposing patients to toxicity, and every second you fail to give the drugs makes them irreversibly worse *forever*!" And I know a bunch of people whose level of psychosis hovers at "mild" and has continued to do so for decades without the lack of treatment making it much worse.

After learning more about the biology of schizophrenia, I've become more willing to credit the DUP model. I can't give great sources for this, because I've lost some of them, but this Friston paper, this Fletcher & Frith paper, and *Surfing Uncertainty* all kind of point to the same model of why untreated schizophrenia might get worse with time.

In their system, schizophrenia starts with aberrant prediction errors; the brain becomes incorrectly surprised by some sense-data. Maybe a fly buzzes by, and all of a sudden the brain shouts “WHOA! I WASN’T EXPECTING THAT! THAT CHANGES EVERYTHING!” Your brain shifts its resources to coming up with a theory of the world that explains why that fly buzzing by is so important – or perhaps which maximizes its ability to explain that particular fly at the cost of everything else.

Talk to early-stage schizophrenics, and their narrative sounds a lot like this. They’ll say something like “A fly buzzed by, and I knew somehow it was very significant. It must be a sign from God. Maybe that I should fly away from my current life.” Then you’ll tell them that’s dumb, and they’ll blink and say “Yeah, I guess it is kind of dumb, now that you mention it” and continue living a somewhat normal life.

Or they’ll say “I was wondering if I should go to the store, and then a Nike ad came on that said JUST DO IT. I knew that was somehow significant to my situation, so I figured Nike must be reading my mind and sending me messages to the TV.” Then you’ll remind them that that can’t happen, and even though it seemed *so interesting* that Nike sent the ad at that exact moment, they’ll back down.

But even sane people change their beliefs more in response to more evidence. If a friend stepped on my foot, I would think nothing of it. If she did it twice, I might be a little concerned. If she did it fifty times, I would have to reevaluate my belief that she was my friend. Each piece of evidence chips away at my comfortable normal belief that people don't deliberately step on my feet – and eventually, I shift.

The same process happens as schizophrenia continues. One fly buzzing by with cosmic significance can perhaps be dismissed. But suppose the next day, a raindrop lands on your head, and there's another aberrant prediction error burst. Was the raindrop a sign from God? The evidence against is that this is still dumb; the evidence for is that you had both the fly and the raindrop, so your theory that God is sending you signs starts looking a little stronger. I'm not talking about this on the conscious level, where the obvious conclusion is "guess I have schizophrenia". I'm talking about the pre-conscious inferential machinery, which does its own mechanical thing and tells the conscious mind what to think.

As schizophrenics encounter more and more strange things, they (rationally) alter their high-level beliefs further and further. They start believing that God often sends signs to people. They start believing that the TV often talks especially to them. They start believing that there is a conspiracy. The more aberrant events they're forced to explain, the more they abandon their sane views about the world (which are doing a terrible job of predicting all the strange things happening to them) and adopt psychotic ones.

But since their new worldview (God often sends signs) gives a high prior on various events being signs from God, they'll be more willing to interpret even minor coincidences as signs, and so end up in a nasty feedback loop. From the Frith and Fletcher paper:

> Ultimately, someone with schizophrenia will need to develop a set of beliefs that must account for a great

> deal of strange and sometimes contradictory data. Very commonly they come to believe that they are being persecuted: delusions of persecution are one of the most striking and common of the positive symptoms of schizophrenia, and the cause of a great deal of suffering. If one imagines trying to make some sense of a world that has become strange and inconsistent, pregnant with sinister meaning and messages, the sensible conclusion might well be that one is being deliberately deceived. This belief might also require certain other changes in the patient's view of the world. They may have to abandon a succession of models and even whole classes of models.

A few paragraphs later, they expand their theory to the negative symptoms of schizophrenia. That is: advanced-stage schizophrenics tend to end up in a depressed-like state where they rarely do anything or care about anything. The authors say:

> Further, although we have deliberately ignored negative symptoms, it is interesting to consider whether this model might have relevance for this extremely incapacitating feature of schizophrenia. We speculate that this deficit could indeed be ultimately responsible for the amotivational, asocial, akinetic state that is characteristic of negative symptoms. After all, a world in which sensory data are noisy and unreliable might lead to a state in which decisions are difficult and actions seem fruitless. We can only speculate on whether the same fundamental deficit could account for both positive and negative features of schizophrenia but, if it could, we suggest that it would be more profound in the case of negative features, and this increased severity might be invoked to account for the strange motor

> disturbances (collectively known as catatonia) that can be such a striking feature of the negative syndrome.

I think what they are saying is that, as the world becomes even more random and confusing, the brain very slowly adjusts its highest level parameters. It concludes, on a level much deeper than consciousness, that the world does not make sense, that it’s not really useful to act because it’s impossible to predict the consequences of actions, and that it’s not worth drawing on prior knowledge because anything could happen at any time. It gets a sort of learned helplessness about cognition, where since it never works it’s not even worth trying. The onslaught of random evidence slowly twists the highest-level beliefs into whatever form best explains random evidence (usually: that there’s a conspiracy to do random things), and twists the fundamental parameters into a form where they expect evidence to be mostly random and aren’t going to really care about it one way or the other.

Antipsychotics treat the positive symptoms of schizophrenia – the hallucinations and delusions – pretty well. But they don’t treat the negative symptoms much at all (except, of course, clozapine). Plausibly, their antidopaminergic effect prevents the spikes of aberrant prediction error, so that the onslaught of weird coincidences stops and things only seem about as relevant as they really are.

But if your brain has already spent years twisting itself into a shape determined by random coincidences, antipsychotics aren’t going to do anything for that. It’s not even obvious that a few years of evidence working normally will twist it back; if your brain has adopted the hyperprior of “evidence never works, stop trying to respond to it”, it’s hard to see how evidence could convince it otherwise.

This theory fits the “duration of untreated psychosis” model very well. The longer you’re psychotic, with weird prediction errors popping up everywhere, the more thoroughly your brain is going to shift from its normal mode of evidence-processing to whatever mode of evidence-processing best suits receiving lots of random data. If you start antipsychotics as soon as the prediction errors start, you’ll have a few weird thoughts about how a buzzing fly might have been a sign from God, but then the weirdness will stop and you’ll end up okay. If you start antipsychotics after ten years of this kind of stuff, your brain will already have concluded that the world only makes sense in the context of a magic-wielding conspiracy plus also normal logic doesn’t work, and the sudden cessation of new weirdness won’t change that.

The Frith and Fletcher paper also tipped me off to this excellent first-person account by former-schizophrenic-turned psychologist Peter Chadwick:

> At this time, a powerful idea of reference also overcame me from a television episode of Colombo and impulsively I decided to write letters to friends and colleagues about “this terrible persecution.” It was a deadly mistake. After a few replies of the “we’ve not heard anything” variety, my subsequent (increasingly overwrought) letters, all of them long, were not answered. But nothing stimulates paranoia better than no feedback, and once you have conceived a delusion, something is bound to happen to confirm it. When phrases from the radio echoed phrases I had used in those very letters, it was “obvious” that the communications had been passed on to radio and then television personnel with the intent of influencing and mocking me. After all betrayal was what I was used to, why should not it be carrying on now? It seemed

sensible. So much for my bonding with society. It was totally gone. I was alone and now trusted no one (if indeed my capacity to trust people [particularly after school] had ever been very high).

The unfortunate tirade of coincidences that shifted my mentality from sane to totally insane has been described more fully in a previous offering. From a meaningless life, a relationship with the world was reconstructed by me that was spectacularly meaningful and portentous even if it was horrific. Two typical days from this episode I have recalled as best I could and also published previously. The whole experience was so bizarre it is as if imprinted in my psyche in what could be called “floodlit memory” fashion. Out of the coincidences picked up on, on radio and television, coupled with overheard snatches of conversation in the street, it was “clear” to me that the media torment, orchestrated as inferred at the time by what I came to call “The Organization,” had one simple message: “Change or die!” Tellingly my mother (by then deceased) had had a fairly similar attitude. It even crossed my (increasingly loosely associated) mind that she had had some hand in all this from beyond the grave […]

As my delusional system expanded and elaborated, it was as if I was not “thinking the delusion,” the delusion was “thinking me!” I was totally enslaved by the belief system. Almost anything at all happening around me seemed at least “relevant” and became, as Piaget would say, “assimilated” to it. Another way of putting things was that confirmation bias was massively amplified, everything confirmed and fitted the delusion, nothing discredited it. Indeed, the very capacity to notice and think of refutatory data and ideas was completely gone.

Confirmation bias was as if “galloping,” and I could not stop it.

As coincidences jogged and jolted me in this passive, vehicular state into the “realization” that my death was imminent, it was time to listen out for how the suicide act should be committed. “He has to do it by bus then?!” a man coincidentally shouted to another man in the office where I had taken an accounts job (in fact about a delivery but “of course” I knew that was just a cover story). “Yes!” came back the reply. This was indeed how my life was to end because the remark was made as if in reply to the very thoughts I was having at that moment. Obviously, The Organization knew my very thoughts.

Two days later, I threw myself under the wheels of a double decker, London bus on “New King’s Road” in Fulham, West London, to where I had just moved. In trying to explain “why all this was happening” my delusional system had taken a religious turn. The religious element, that all this torment was willed not only by my mother and transvestophobic scandal-mongerers but by God Himself for my “perverted Satanic ways,” was realized in the personal symbolism of this suicide. New King’s Road obviously was “the road of the New King” (Jesus), and my suicide would thrust “the old king” (Satan) out of me and Jesus would return to the world to rule. I then would be cast into Outer Darkness fighting Satan all the way. The monumental, world-saving grandiosity of this lamentable act was a far cry from my totally irrelevant, penniless, and peripheral existence in Hackney a few months before. In my own bizarre way, I obviously had moved up in the world. Now, I was not an outcast from it. I was saving the world in a very lofty manner. Medical authorities at Charing

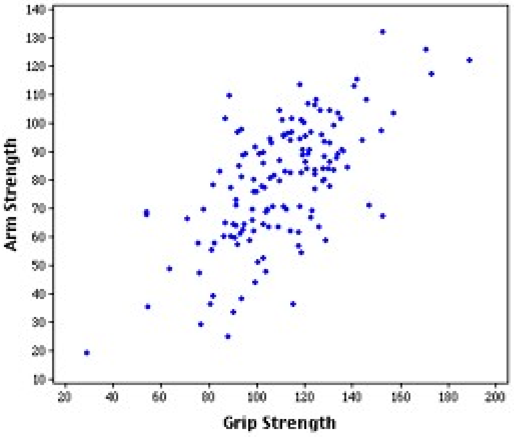
> Cross Hospital in London where I was taken by ambulance, initially, of course, to orthopedics, fairly quickly recognized my psychotic state. Antipsychotic drugs were injected by a nurse on doctors’ advice, and eventually, I made a full physical and mental recovery.

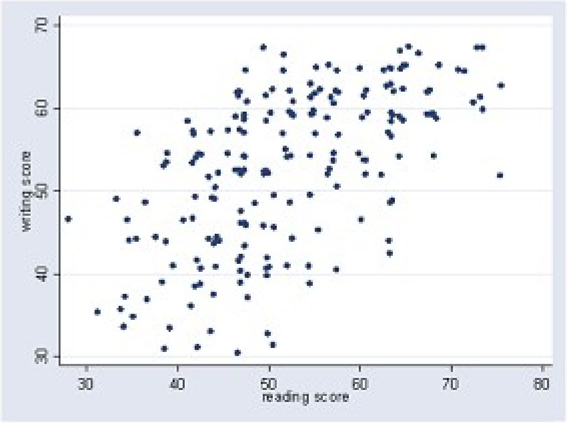
Chadwick never got too far along; he had all the weird coincidences, he was starting to get beliefs that explained them, but he never got to a point where he shifted his fundamental concepts or beliefs about logic in an irreversible way. As far as I know he’s been on antipsychotics consistently since then, and has escaped with no worse consequences than becoming a psychology professor. I am not sure whether things would have gone worse for him without the medications, but I think it’s a possibility we have to consider.

The Tails Coming Apart As Metaphor For Life

[Epistemic status: Pretty good, but I make no claim this is original]

A neglected gem from Less Wrong: Why The Tails Come Apart, by commenter Thrasymachus. It explains why even when two variables are strongly correlated, the most extreme value of one will rarely be the most extreme value of the other. Take these graphs of grip strength vs. arm strength and reading score vs. writing score:

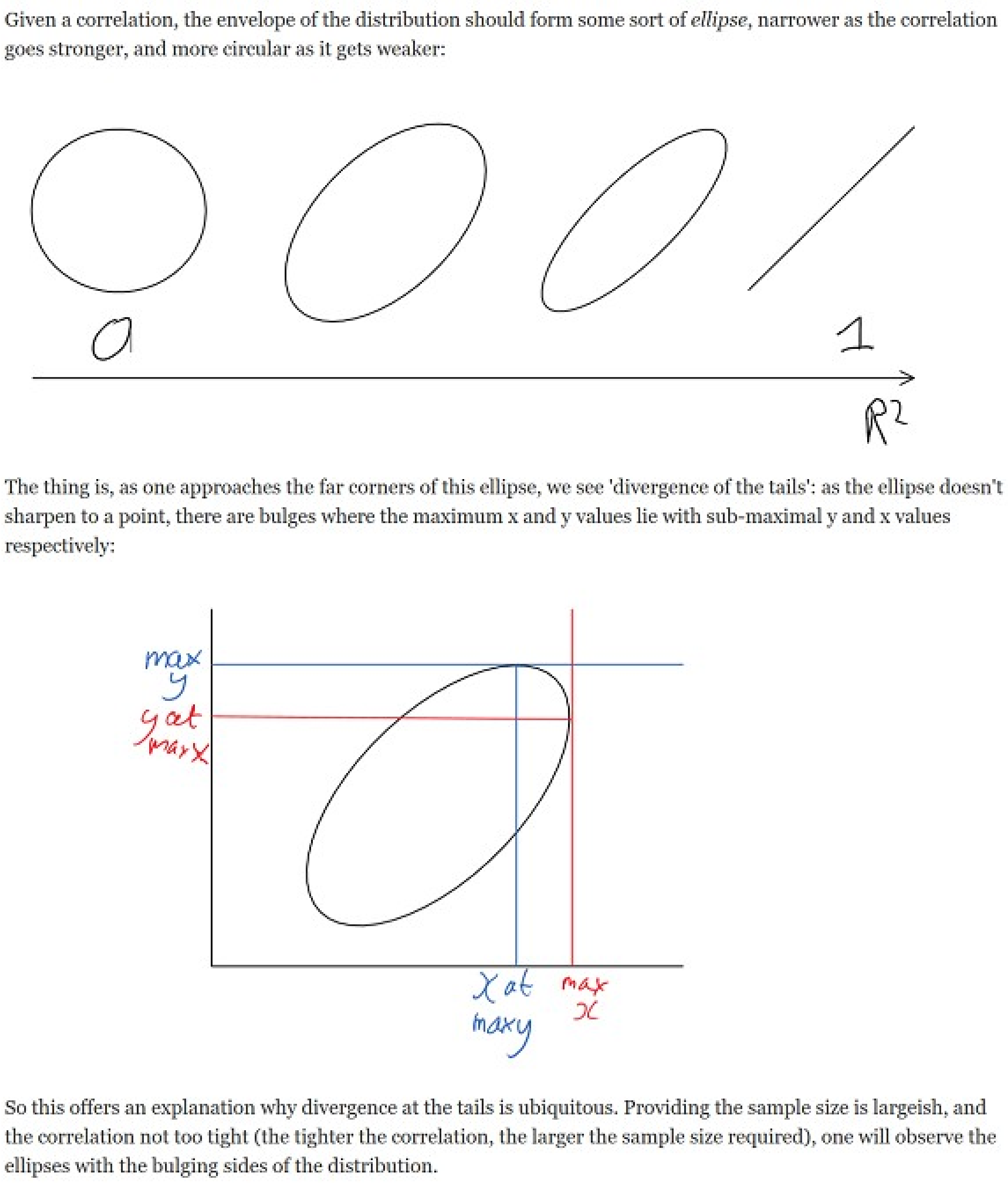




In a pinch, the second graph can also serve as a rough map of Afghanistan

Grip strength is strongly correlated with arm strength. But the person with the strongest arm doesn't have the strongest grip. He's up there, but a couple of people clearly beat him. Reading and writing scores are even less correlated, and some of the people with the best reading scores aren't even close to being best at writing.

Thrasymachus gives an intuitive geometric explanation of why this should be; I can't beat it, so I'll just copy it outright:



Given a correlation, the envelope of the distribution should form some sort of *ellipse*, narrower as the correlation goes stronger, and more circular as it gets weaker:

The thing is, as one approaches the far corners of this ellipse, we see 'divergence of the tails': as the ellipse doesn't sharpen to a point, there are bulges where the maximum x and y values lie with sub-maximal y and x values respectively:

So this offers an explanation why divergence at the tails is ubiquitous. Providing the sample size is largeish, and the correlation not too tight (the tighter the correlation, the larger the sample size required), one will observe the ellipses with the bulging sides of the distribution.

I thought about this last week when I read this article on happiness research.

The summary: if you ask people to “value their lives today on a 0 to 10 scale, with the worst possible life as a 0 and the best possible life as a 10”, you will find that Scandinavian countries are the happiest in the world.

But if you ask people “how much positive emotion do you experience?”, you will find that Latin American countries are the happiest in the world.

If you check where people are the least depressed, you will find Australia starts looking very good.

And if you ask “how meaningful would you rate your life?” you find that African countries are the happiest in the world.

It’s tempting to completely dismiss “happiness” as a concept at all, but that’s not right either. Who’s happier: a millionaire with a loving family who lives in a beautiful mansion in the forest and spends all his time hiking and surfing and playing with his kids? Or a prisoner in a maximum security jail with chronic pain? If we can all agree on the millionaire – and who wouldn’t? – happiness has to *at least sort of* be a real concept.

The solution is to understand words as hidden inferences – they refer to a multidimensional correlation rather than to a single cohesive property. So for example, we have the word “strength”, which combines grip strength and arm strength (and many other things). These variables really are heavily correlated (see the graph above), so it’s almost always worthwhile to just refer to people as being strong or weak. I can say “Mike Tyson is stronger than an 80 year old woman”, and this is better than having to say “Mike Tyson has higher grip strength, arm strength, leg strength, torso strength, and ten other different kinds of strength than an 80 year old woman.” This is necessary to communicate anything at all and given how nicely all forms of strength correlate there’s no reason not to do it.

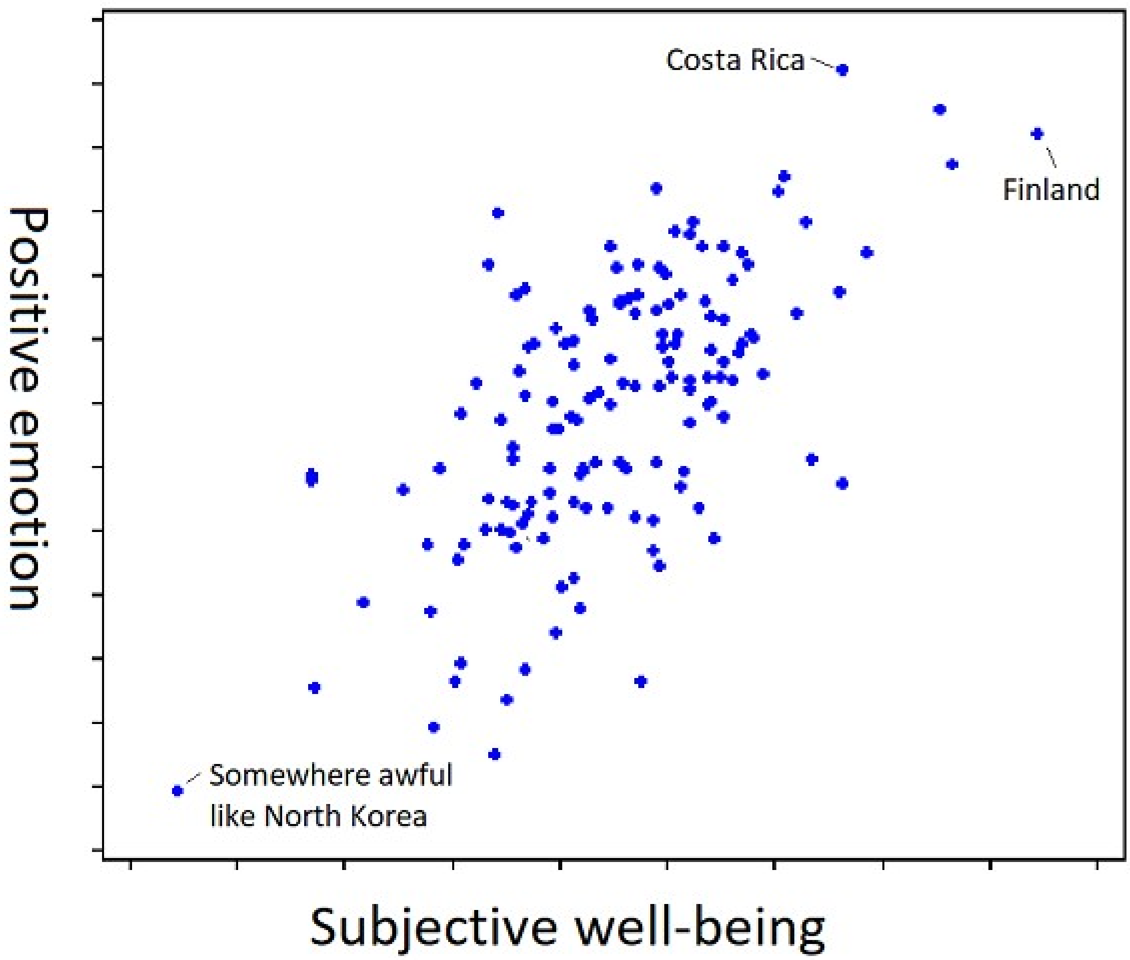
But *the tails still come apart*. If we ask whether Mike Tyson is stronger than some other very impressive strong person, the answer might very well be “He has better arm strength, but worse grip strength”.

Happiness must be the same way. It’s an amalgam between a bunch of correlated properties like your subjective well-being at any given moment, and the amount of positive emotions you feel, and how meaningful your life is, et cetera. And *each of those correlated is also an amalgam*, and so on to infinity.

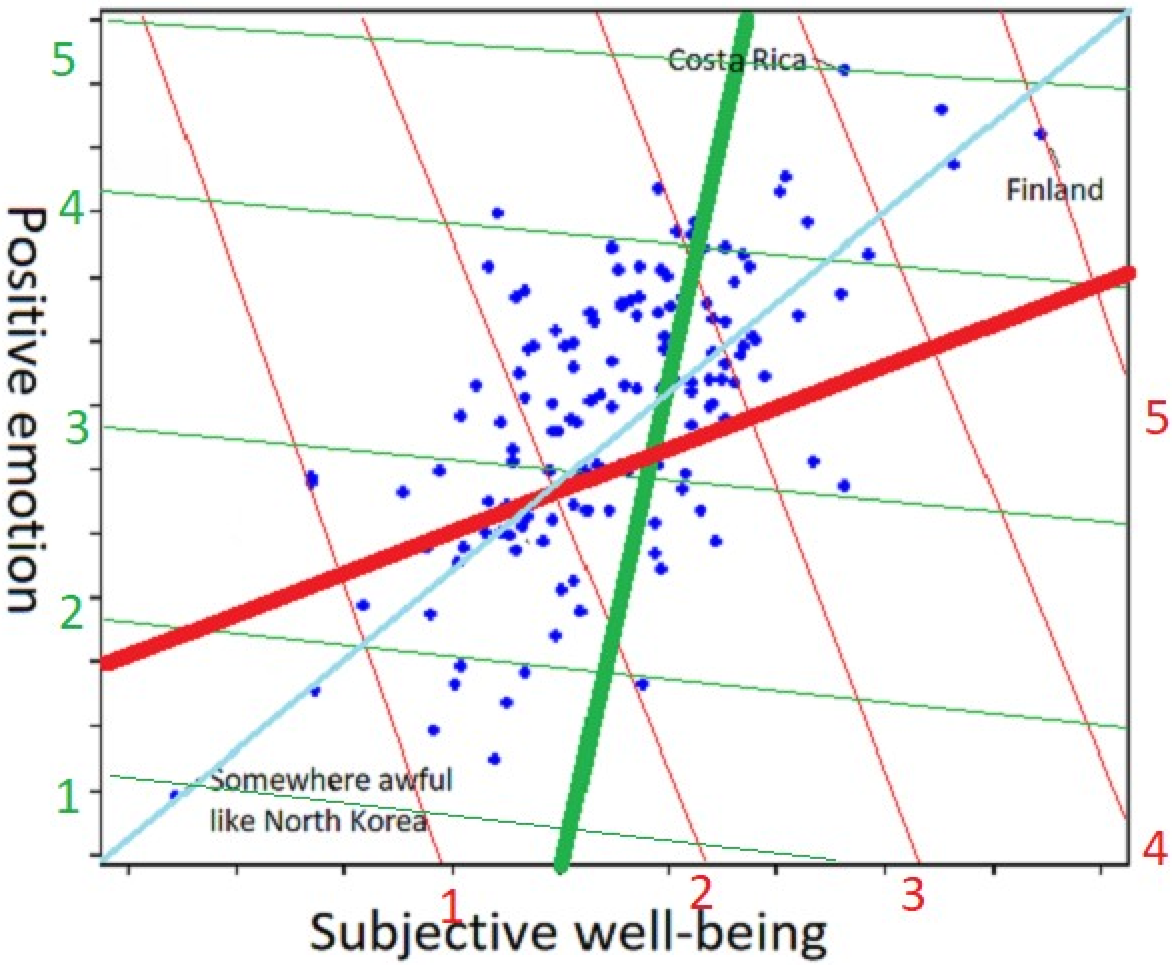
And crucially, it’s not an amalgam in the sense of “add subjective well-being, amount of positive emotions, and meaningfulness and divide by three”. It’s an unprincipled conflation of these that just denies they’re different at all.

Think of the way children learn what happiness is. I don’t actually know how children learn things, but I imagine something like this. The child sees the millionaire with the loving family, and her dad says “That guy must be very happy!”. Then she sees the prisoner with chronic pain, and her mom says “That guy must be very sad”. Repeat enough times and the kid has learned “happiness”.

Has she learned that it’s made out of subjective well-being, or out of amount of positive emotion? I don’t know; the learning process doesn’t determine that. But then if you show her a Finn who has lots of subjective well-being but little positive emotion, and a Costa Rican who has lots of positive emotion but little subjective well-being, and you ask which is happier, *for some reason she’ll have an opinion*. Probably some random variation in initial conditions has caused her to have a model favoring one definition or the other, and it doesn’t matter until you go out to the tails. To tie it to the same kind of graph as in the original post:



And to show how the individual differences work:



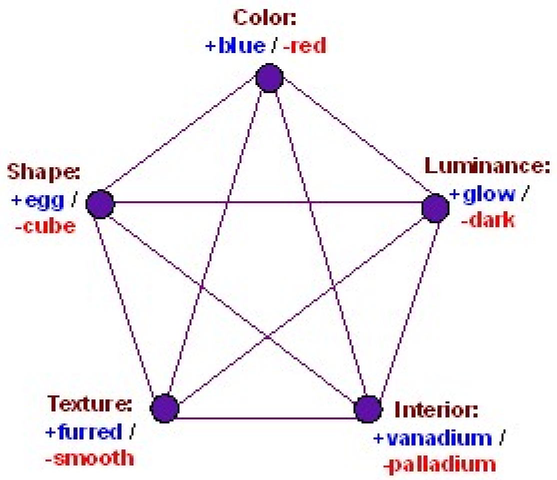
I am sorry about this graph, I really am. But imagine that one person, presented with the scatter plot and asked to understand the concept “happiness” from it, draws it as the thick red line (further towards the top right part of the line = more happiness), and a second person trying to the same task generates the thick green line. Ask the first person whether Finland or Costa Rica is happier, and they’ll say Finland: on the red coordinate system, Finland is at 5, but Costa Rica is at 4. Ask the second person, and they’ll say Costa Rica: on the green coordinate system, Costa Rica is at 5, and Finland is at 4 and a half. Did I mention I’m sorry about the graph?

But isn’t the line of best fit (here more or less y = x = the cyan line) the objective correct answer? Only in this metaphor where we’re imagining positive emotion and subjective well-being are both objectively quantifiable, and exactly equally important. In the real world, where we have no idea how to quantify any of this and we’re going off vague impressions, I would hate to be the person tasked with deciding whether the red or green line was more objectively correct.

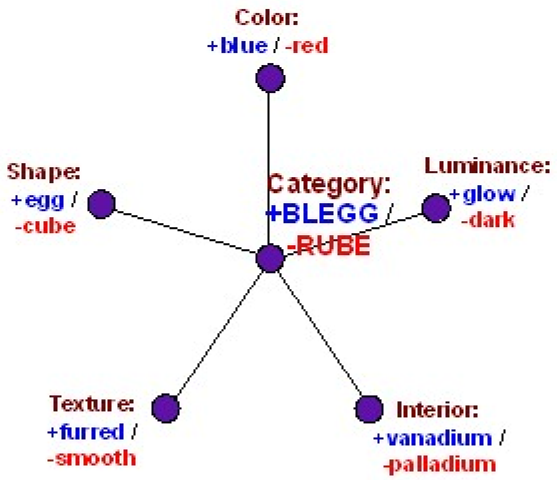
In most real-world situations Mr. Red and Ms. Green will give the same answers to happiness-related questions. Is Costa Rica happier than North Korea? “Obviously,” the both say in union. If the tails only come apart a little, their answers to 99.9% of happiness-related questions might be the same, so much so that they could never realize they had slightly different concepts of happiness at all.

(is this just reinventing Quine? I’m not sure. If it is, then whatever, my contribution is the ridiculous graphs.)

Perhaps I am also reinventing the model of categorization discussed in How An Algorithm Feels From The Inside, Dissolving Questions About Disease, and The Categories Were Made For Man, Not Man For The Categories.


Color:
+blue / -red
Shape:
+egg /
-cube
Luminance:
+glow /
-dark
Texture:
+furred /
-smooth
Interior:
+vanadium /
-palladium

Network 1


Color:
+blue / -red
Shape:
+egg /
-cube
Category:
+BLEGG /
-RUBE
Luminance:
+glow /
-dark
Texture:
+furred /
-smooth
Interior:
+vanadium /
-palladium

Network 2

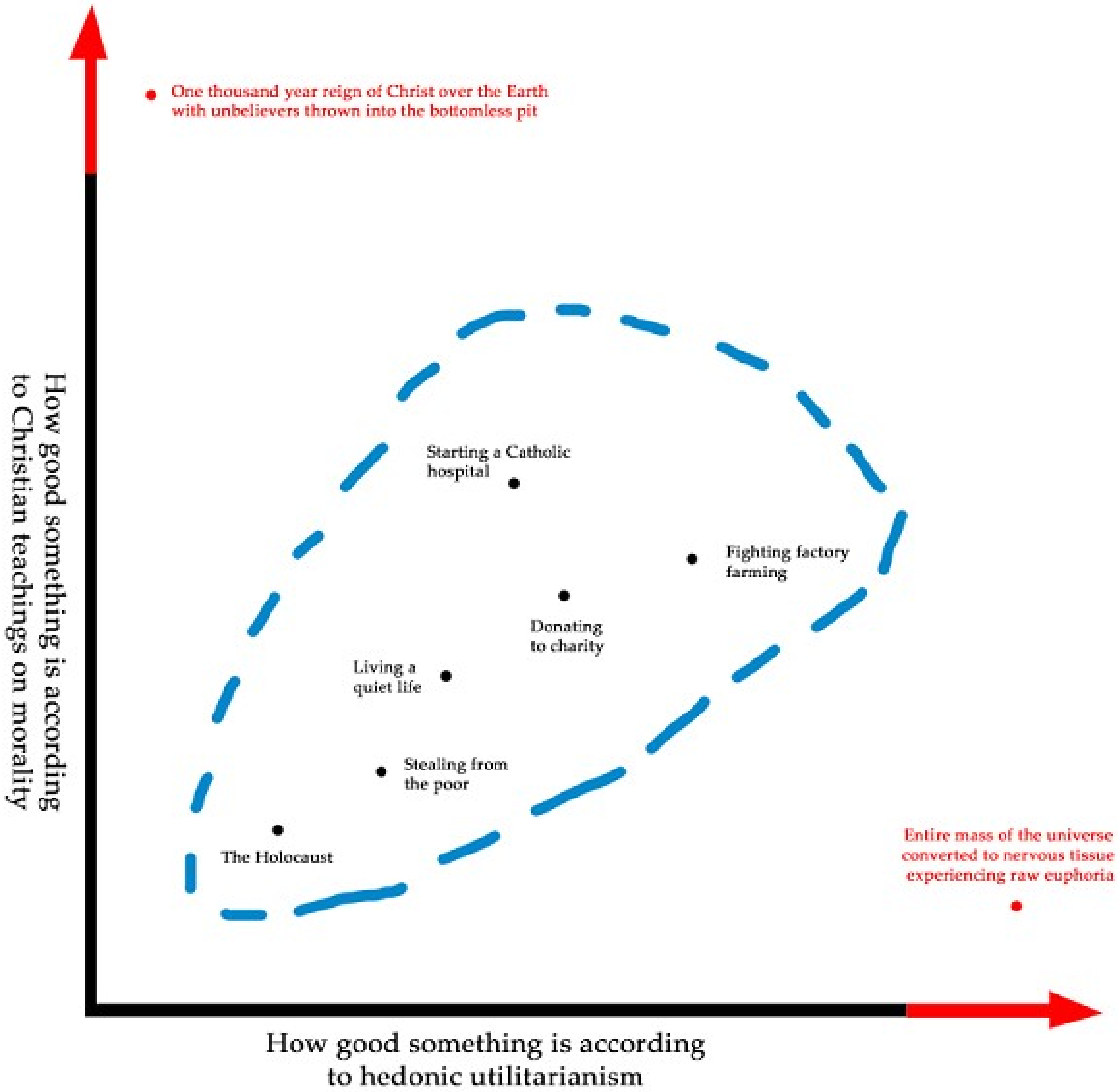
But I think there’s another interpretation. It’s not just that “quality of life”, “positive emotions”, and “meaningfulness” are three contributors which each give 33% of the activation to our central node of “happiness”. It’s that we got some training data – the prisoner is unhappy, the millionaire is happy – and used it to build a classifier that told us what happiness was. The training data was ambiguous enough that different people built different classifiers. Maybe one person built a classifier that was based entirely on quality-of-life, and a second person built a classifier based entirely around positive emotions. Then we loaded that with all the social valence of the word “happiness”, which we naively expected to transfer across paradigms.

This leads to (to steal words from Taleb) a Mediocristan resembling the training data where the category works fine, vs. an Extremistan where everything comes apart. And nowhere does this become more obvious than in what this blog post has secretly been about the whole time – morality.

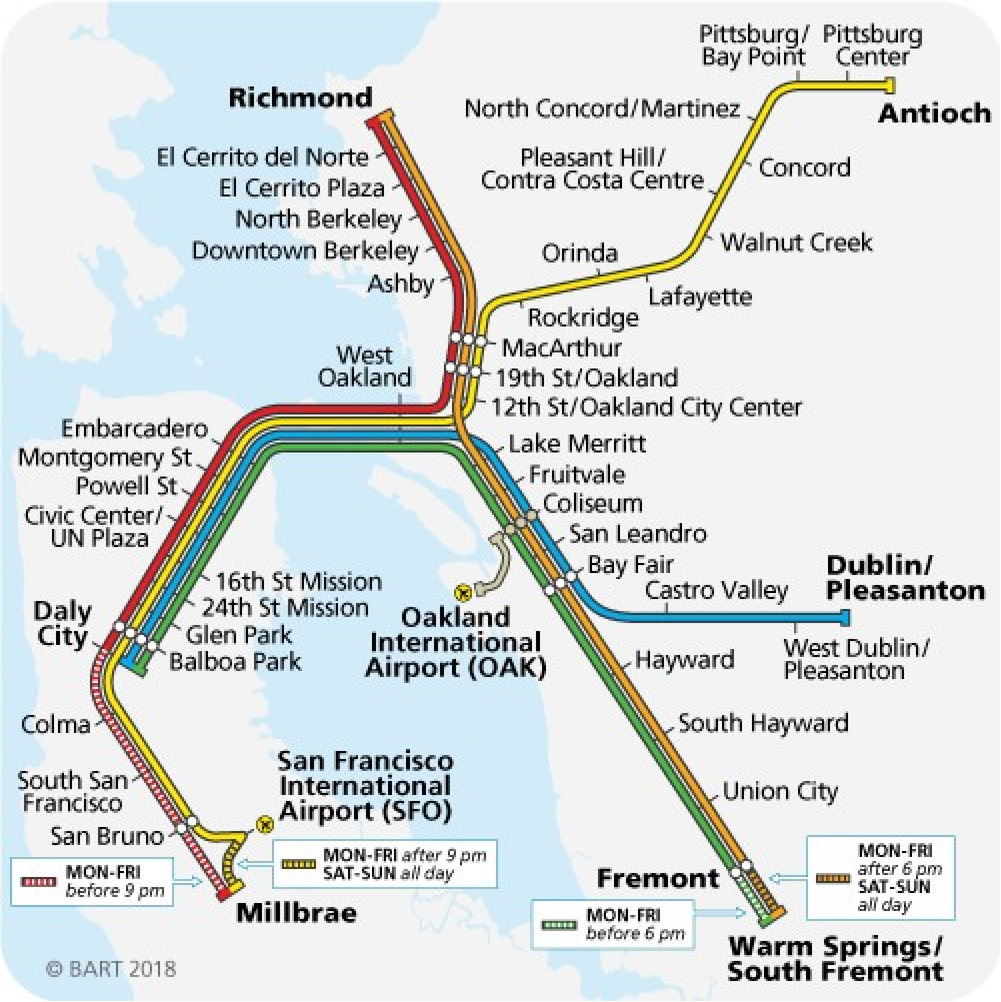
The morality of Mediocristan is mostly uncontroversial. It doesn’t matter what moral system you use, because all moral systems were trained on the same set of

Mediocristani data and give mostly the same results in this area. Stealing from the poor is bad. Donating to charity is good. A lot of what we mean when we say a moral system sounds plausible is that it best fits our Mediocristani data that we all agree upon. This is a lot like what we mean when we say that “quality of life”, “positive emotions”, and “meaningfulness” are all decent definitions of happiness; they all fit the training data.

The further we go toward the tails, the more extreme the divergences become. Utilitarianism agrees that we should give to charity and shouldn’t steal from the poor, because Utility, but take it far enough to the tails and we should tile the universe with rats on heroin. Religious morality agrees that we should give to charity and shouldn’t steal from the poor, because God, but take it far enough to the tails and we should spend all our time in giant cubes made of semiprecious stones singing songs of praise. Deontology agrees that we should give to charity and shouldn’t steal from the poor, because Rules, but take it far enough to the tails and we all have to be libertarians.



I have to admit, I don’t know if the tails coming apart is even the right metaphor anymore. People with great grip strength still had *pretty* good arm strength. But I doubt these moral systems form an ellipse; converting the mass of the universe into nervous tissue experiencing euphoria isn’t just the *second-best* outcome from a religious perspective, it’s completely abominable. I don’t know how to describe this mathematically, but the terrain looks less like tails coming apart and more like the Bay Area transit system:


Pittsburg/ Bay Point
Pittsburg Center
Antioch
North Concord/Martinez
Richmond
El Cerrito del Norte
El Cerrito Plaza
North Berkeley
Downtown Berkeley
Ashby
Pleasant Hill/ Contra Costa Centre
Concord
Walnut Creek
Orinda
Lafayette
Rockridge
MacArthur
19th St/Oakland
12th St/Oakland City Center
West Oakland
Embarcadero
Montgomery St
Powell St
Civic Center/ UN Plaza
Lake Merritt
Fruitvale
Coliseum
San Leandro
Bay Fair
Castro Valley
Dublin/ Pleasanton
West Dublin/ Pleasanton
16th St Mission
24th St Mission
Glen Park
Balboa Park
Daly City
Oakland International Airport (OAK)
Hayward
South Hayward
Colma
South San Francisco
San Francisco International Airport (SFO)
San Bruno
Union City
MON-FRI before 9 pm
MON-FRI after 9 pm SAT-SUN all day
Millbrae
Fremont
MON-FRI after 6 pm SAT-SUN all day
MON-FRI before 6 pm
Warm Springs/ South Fremont
© BART 2018

Mediocristan is like the route from Balboa Park to West Oakland, where it doesn’t matter what line you’re on because they’re all going to the same place. Then suddenly you enter Extremistan, where if you took the Red Line you’ll end up in Richmond, and if you took the Green Line you’ll end up in Warm Springs, on totally opposite sides of the map.

Our innate moral classifier has been trained on the Balboa Park – West Oakland route. Some of us think morality means “follow the Red Line”, and others think “follow the Green Line”, but it doesn’t matter, because we all agree on the same route.

When people talk about how we should arrange the world after the Singularity when we’re all omnipotent, suddenly we’re way past West Oakland, and everyone’s moral intuitions hopelessly diverge.

But it’s even worse than that, because even within myself, my moral intuitions are something like “Do the thing which follows the Red Line, and the Green Line, and the Yellow Line…you know, *that* thing!” And so when I’m faced with something that perfectly follows the Red Line, but goes the opposite directions as the Green Line, it seems repugnant *even to me*, as does the opposite tactic of following the Green Line. As long as creating and destroying people is hard, utilitarianism works fine, but make it easier, and suddenly your Standard Utilitarian Path diverges into Pronatal Total Utilitarianism vs. Antinatalist Utilitarianism and they both seem awful. If our degree of moral repugnance is the degree to which we’re violating our moral principles, and my moral principle is “Follow both the Red Line and the Green Line”, then after passing West Oakland I either have to end up in Richmond (and feel awful because of how distant I am from Green), or in Warm Springs (and feel awful because of how distant I am from Red).

This is why I feel like figuring out a morality that can survive transhuman scenarios is harder than just finding the Real Moral System That We Actually Use. There’s a potentially impossible conceptual problem here, of figuring out what to do with the fact that any moral rule followed to infinity will diverge from large parts of what we mean by morality.

This is only a problem for ethical subjectivists like myself, who think that we’re doing something that has to do with what our conception of morality is. If you’re an ethical naturalist, by all means, just do the thing that’s actually ethical.

When Lovecraft wrote that “we live on a placid island of ignorance in the midst of black seas of infinity, and it was not meant that we should voyage far”, I interpret him as talking about the region from Balboa Park to West Oakland on the map above. Go outside of it and your concepts break down and you don’t know what to do. He was right about the island, but exactly wrong about its causes – the most merciful thing in the world is how so far we have managed to stay in the area where the human mind *can* correlate its contents.

Adversarial Collaboration Contest Results

Thanks to everyone who participated or voted in the adversarial collaboration contest. The winners are:

Grand Prize ($1000): Does The Education System Adequately Serve Advanced Students?

Editor's Choice ($500): Should Transgender Children Transition?

Honorable Mentions ($250): Should Childhood Vaccination Be Mandatory?, Are Islam And Liberal Democracy Compatible?

I'm sorry for jerking the number and value of the prizes around so many times, but I wanted to balance my preferences, the contestants' preferences, and readers' preferences – and this was the best way I could think of to do it. Nobody has gotten less money than they expected, although some prize categories have gotten more money than I originally said. In the end I could not in good conscience let any of these escape without getting a prize. Thanks to this blog's Patreon supporters for making this possible. All winners should email me with their preferred form of payment (I can do Paypal, Bitcoin, or donations to a charity of their choice).

The overwhelming winner of the popular vote was the collaboration on education. I agree this one was excellent. It cited a lot of research, analyzed it very well, and mostly came to conclusions. Its only flaw from my perspective was a lack of focus; it discussed many different educational interventions, some of which were similar enough that it was hard for me to keep track of what was going on.

I chose the collaboration on transgender children. I thought it did an exceptional job of addressing a specific hot-button issue many people are concerned about, presenting all the evidence on both sides, and mostly coming to conclusions. My strongest complaint was that it ignored some of the potential side effects of puberty blockers which commenters pointed out, and sort of trivialized bone problems that are not trivial; given that the side effects of puberty blockers was a major crux of this question, I found that to be a major weakness. I was still very impressed with the piece’s ability to break down and navigate such a controversial question.

The first honorable mention, the collaboration on childhood vaccination, was also a great example of navigating a controversial topic. It seemed to come closest to outright saying one of the two sides was correct – though see the commentary below for more on that. I had to take off points because the conclusion – that lots of countries get away without having mandatory childhood vaccinations, so surely America can too – seemed overly simplistic. Those other countries might maintain high childhood vaccination rates because of cultural differences, or a different (eg socialized) health system, or lots of other things that might not generalize to the US. Again, this was the crux of the issue, so minor flaws here are very important. The section on the hygiene hypothesis, while fascinating, seemed kind of unfocused to me and probably could have been much shorter.

The second honorable mention, the collaboration on Islam and democracy, was maybe the most thoroughly researched, but also seemed the least structured. The writers chose to list democratic and undemocratic features of six Islamic countries. Since they found one Islamic country (Tunisia) that seemed pretty much like a liberal democracy, they concluded the two were compatible. But in real life, when people ask whether Islam and democracy are compatible, I imagine them wondering things like “Are Muslim countries less likely to be democratic than other countries?” or “Is the Muslim religion one reason why there is so little democracy in the Middle East?” or “In mixed countries, do Muslim populations resist democratic norms?” The collaboration’s focus on listing the particular features of particular Muslim countries’ institutions seemed like a really complete answer to a really boring question.

I asked the various collaborators to send me their reports on the process. Their answers are below, or scroll down to the end of the blockquotes for my summary.

TracingWoodgrains from the education collaboration:

> My adversarial collaboration was similar to how Kahneman describes his experience (https://pastebin.com/d6cbN765): a “failure to disagree.” When Michael and I zoomed into an issue at the object level, we usually agreed. Then we’d zoom out and realize we used that object-level detail as part of two drastically different narratives. I have no education credentials besides having been a deeply frustrated student; he was a happy student and now enjoys teaching. That dichotomy between teacher and student, happy and unhappy, defined our perspectives.
>
> Going in, I was confident that tailored curriculum involving ability grouping could help advanced students progress up to 4-5 grade levels above where they are now, that school is a deeply harmful place for many advanced students academically and socially, and that

education should follow the principles of game design (using online tools to allow students to progress at their own paces while encouraging them to progress quickly).

Our project was extensive and far-reaching, mostly due to our disagreement and Michael’s tireless willingness to keep working with me. We approached it as a research project more than anything else, which mostly meant “toss all relevant-looking research and related writing towards each other, read it manically, and see where and how disagreements bubbled up as we were learning.” At the start, I was coming from a position of near-complete ignorance towards the state of specific research on ability grouping, though I had passing familiarity with acceleration research and what I felt was a comfortable understanding of the (poor) state of research around online education. So a lot of my “position-shifting” was learning:

I learned that tracking, in a traditional sense, had much weaker evidence than I was anticipating, but the more it looked like what my intuitive view of its ideal, the better it fared. So: few peer effects, strong effects based on curriculum change, strongest impact based on diagnostic testing and precisely tailored curriculum. My views on acceleration as a useful, underfocused stopgap measure did not change. My understanding of the complexities of the education system, and why it was so far from what I hoped to see, grew tenfold. Michael’s pressing reminders of the competing purposes of education and experience in the field helped me a lot there.

That’s most of the small stuff. I had two genuinely paradigm-shifting experiences: First, on the happiness of advanced students. Typical mind here–I’d always just assumed that most advanced students were bored, held back, and miserable in school. It seemed one of the most obvious things in the world to me. But it’s really, really hard to find any of that in the research. Most studies indicate them enjoying school more than other students, (tautologically) performing better, and having lower incidence of mental health issues. My view shifted to intelligence as a potential intensifier for other risk factors (ADHD, autism, etc.), but not as the core cause of student frustration in school. My belief that the “smart kids” who are bored and miserable in school tend to be underprioritized and mis-served remains largely unchanged, but I have a very different view of this group.

Second: Michael pointed me towards noticing the apparent mediocrity of a lot of schools implementing some of the digital/in-person hybrid instruction I dreamed of. Rocketship Schools, School of One, Khan Academy, and others grounded my expectations of any sort of revolution. At the same time, we discovered Art of Problem Solving, which is probably the best thing in the world, and seeing what they’ve done with online education and specifically math made me realize that people smarter than me have been doing a really good job on “my” project. So now I see less need of creating the right resources, more of expanding their availability and sifting them out from the sea of garbage.

Less paradigm-shifting, but significant, is that I grew to trust experts more as I noticed that when you dug deep enough, most of the serious researchers came to pretty similar conclusions but wrapped them in different narratives, and really bad ideas that filtered through were less because the serious researchers had really bad ideas, more because their messages got distorted or ignored to better fit political agendas and the vagaries of people’s opinions.

(Bonus areas where I didn’t change my view, just properly grew one because I didn’t know enough to have a view before: direct instruction as intriguing and mostly good, Joplin plan and nongraded schools as intriguing and oddly ignored, early childhood education as incredibly high potential in specific fields, a growing frustration with just how much politics has infected and damaged education)

One example best outlines our differences: Direct Instruction. Michael actually brought it up to highlight problems he saw with online schools–basically, pointing out that it works, but isn’t attractive at all, except in educational emergencies. I hadn’t heard of it before and got about as far as “it works” before getting really fascinated by its whole story. I looked at schools that use it, interviewed someone closely involved with it, read into it, and concluded that it had a lot of potential that was being neglected since people felt like it wasn’t how they “should” learn. That whole arc was never supposed to happen. It was an offhand example that he raised to dismiss. And my response was basically “this is amazing; how have I never heard of this before?”

That happened again and again–where I would see an unusual, academically intensive approach that got good results, we’d talk about it, and he’d ask, “right, but what’s the point?” Sorting by aptitude over age level, high-intensity accelerated math programs, early (pre-K) academics, so forth. It’s worth repeating–we almost never disagreed about what studies showed, just on the importance of particular studies and particular points.

And that was enough to fuel a hundred pages or so of disagreement.

So, advice for future adversarial collaborations: First off, it can only work if both are very, very willing to talk about the topic–the whole topic, not only their pet issues within it. If you find yourself uninterested in a part of the topic the other person is passionate about, pay close attention, since that’s often the most important part of disagreement. Second, it takes a lot of time. We took more time than strictly needed because we both enjoyed the research and conversations, but I’d guess 50-100 hours is a reasonable amount of time to set aside. Third–expect to find a lot of disagreements that boil down to differing priorities and interpretations of object-level facts you both agree on. If you’re both reasonable and willing to work together, you may find little factual ground you disagree on, even while telling two very different stories of the big picture.

One last thing: Adversarial collaborations are a phenomenal thinking tool. This was the most satisfying project I’ve worked on, bar none, and the best learning experience of my life. Talking in depth with someone who disagrees with you is incredible. If both agree, it’s hard to sustain a conversation, but the second someone says something wrong-seeming, it spurs a sort of need to respond, a creative burst. I read more research and focused more on that topic than I have at any other point, even though education has been one of my core interests for a while. Michael was the best partner I could have asked for on the project. I expect that we’ll still talk about wildly different things moving forward: I still love the potential in online learning, non-graded schools, and accelerated/intensive tools that depart dramatically from the current school model. Michael still

> loves teaching and working within the system mostly as it stands, and doesn't trust most radical changes.
>
> Thank you very much for organizing this contest. It was an incredible opportunity and a vital experience.

Michael from the same education collaboration:

> I really want to give you something really detailed and good about our collaboration, but that would take a lot of work. Instead I'll just give you the quick, stream of consciousness version.
>
> When I first saw TW's comment on your post, I thought his views on education were absolutely naive. I heard him expressing a love for game-based education, which I thought to be faddish and flashy in education but without much value. He also expressed that the strongest students were especially poorly-served by education. I thought that was dead wrong; smart kids would do fine, in general, out of school. Without checking, I can't remember if he said anything about ability grouping. That wasn't the point, anyway. What I heard was (sigh) another technophile education utopian who thinks that their own particular experience should be the basis of a system tasked with educating millions of children.
>
> So we started chatting, and pretty soon we'd created an 150 page google doc. The doc contained our debates, annotated bibliographies, brief memoirs documenting our own experiences in education, etc.
>
> I'm coming from all of this as a guy who was pretty conventionally successful in my academic life, and ended up in teaching almost by accident. In my professional life I've given a lot of attention to students

who aren’t succeeding in my classes. Those are the situations that have made me the most passionate to put in extra efforts to help them. So things like volunteering to tutor a kid outside of class, to go over homework during my planning period, rethinking my pedagogical approaches.

At the same time, I had recently experienced a few cases in my honors elementary age classes that were making me think. And I was going back to work at a camp for talented math students.

So I’d say the real thing that changed me, through this experience, is I spent a lot of time thinking about students who find a class too easy, and what that can feel like. That was the most significant way that my views changed through working with TW. He consistently drew my attention towards students who are talented but unhappy.

I feel as if with what we wrote about technology, my views didn’t significantly develop through our collaboration. I think TW quickly learned that his tech optimism is future-optimism, not present-optimism, and that’s a change for him. But otherwise what we did there was more along the lines of finding a common narrative that we both thought important for people to understand the situation with technology and learning. That involved a lot of back and forth between us, as it’s hard to put everything in a way that satisfies both a pessimist (me) and an optimist (him). That’s how we conceived of this work, though.

A lot of the other things we did were more about fleshing out new conceptual territory than changing my views. I didn’t really have strong opinions about tracking or ability grouping. I was eager to learn something new, and TW was a great partner for working through the literature.

There was a lot of debate during this project, and we still disagree deeply. It’s hard to say who “won” our collaboration or who pushed who further etc. I’ve been reading about a lot of this stuff over the past 8 years. I don’t know how old TW is but he’s younger than me. I certainly urged him to read certain people who I thought were foundational, like Larry Cuban or David Labaree. A lot of this was very new to me though, and I was trying to integrate my experience in the classroom with my reading. I feel as if my strong views on gamification were quickly confirmed, and I also think that my view that smart kids mostly turn out alright were vindicated too. I’m sure TW has complementary ways in which he feels he shifted my views, though I’m not exactly sure what he’d say.

One thing that I think people haven’t given enough thought to is the differences between adversarial collaborations in different fields. Certainly there is a difference between medical research, economics research, comparative politics and education research. Each field has a different epistemology, and education research is a big tent containing many individual subfields. That means that TW and I had to do something that I love doing, which is figuring out how cog psych and big metanalysis and economic studies and case studies and history of education all fit together in a coherent whole.

So while I can imagine a briefer more focused write up for our collaboration, I think it’s possible that making sense of classroom questions is going to involve a lot more fuzziness and wandering than for certain other questions. We can't just easily answer the question of whether top students are harmed by schooling. Compared to what? What's the vision? What's the purpose of schooling? That's a question you don't really need to ask about e.g. vaccines. The goal is to keep people alive without harming anybody else. I know there is some fuzziness there, but it's just not the same magnitude.

One more thing, before I call it a night: by design, TW were non-adversarial in our collaboration. Early on, I sent TW an email where I said, you know what? We've figured out that we deeply disagree. I'm not interested in tallying points on whether my beliefs have been weakened or changed or your's have. Instead, let's just try to figure things out together. And that's what we did.

I think a major question going forward is whether adversarial collaborations work best when you have an opponent or when things are framed so as to reduce opposition.

John and Christian from the Islam collaboration:

WHAT WERE YOUR INITIAL POSITIONS?
JohnBuridan's initial position was that it was highly unlikely that Islam consistently opposed liberal democracy. ChristianFlanery's position was that intrinsic components of Islam operate contrary to liberal democracy.

HOW MUCH DID YOUR POSITION SHIFT, AND HOW?

JohnBuridan's position did not shift towards the other side of the aisle; he still think the opposing position holds too large a generalization. He did shift his approach to the question from a theoretical approach to a more descriptive one. ChristianFlanery convinced JB that the logic of surveying the variability in Islamic political thought would require immense time and resources. It would end up requiring us to determine what different groups of Muslim thinkers and activists think of each other. CF convinced JB that a descriptive and historical framework would come closer to providing insights into how the diversity of Islamic political thought caches out in real modern day politics.

CF became more conscious about how new Islam’s interactions with liberal democracy are. This made him less confident in his initial position. Furthermore, CF now has high confidence that liberal democratic polities can be achieved in the right conditions. CF continues to believe that a fundamentalist component in Islam will always exist, but he is now agnostic about whether Muslim populations will be experimenting with liberal democracy well into the future.

HOW MUCH DEBATE AND ARGUMENT WAS THERE?

Constant unceasing debate, but in a context of high levels of trust. We did not fall victim to one-upsmanship or point scoring debates. Preexisting intellectual rapport allowed us to trust each other not to load the dice too much. Although each of us probably gave too much ground to other person in certain places throughout the paper – many of which the commentariat picked up on – we think that such tradeoffs made the work possible.

The key to resolving debate for us was an established process for the paper. Each had three countries to research and do the write up on, and each of us edited each other’s rough drafts. We wrangled over essentially every line of the introduction but had the patience for very long phone calls to resolve them. Oftentimes we would disagree and circle back to the disagreement a week later and come to a solution. We both agreed beforehand to try to only pick battles we thought worth fighting. Tense moments surrounded our discussion of the introduction, the conclusion, constitution, Lebanon, and Iran. Our longest back-and-forths concerned the introduction and conclusion.

IS THE CONCLUSION CLOSER TO ONE PERSON'S ORIGINAL POSITION?

Our conclusion is closer to the side of commensurability. Nonetheless it is fairly neutral, and we try to leave the door open for the reader's own knowledge and expertise to affect the conclusion. Our goal was to provide enough background knowledge of some slice of the Islamic world from which a general reader could inform their own conclusions.

– Lots of readers were deeply concerned about how we chose the countries we did. We chose the countries before doing deep research, and we were looking for a few key traits:
1) majority-supermajority Muslim
2) culturally diverse
3) geographically distant from one another.
4) tractable to discuss (we did not feel that we could do Saudi Arabia or Pakistan justice, and we considered the possibility of many other countries.)

WHAT ADVICE WOULD YOU GIVE FUTURE ADVERSARIAL COLLABORATORS?

> Our main points were:
> 1) Establish high levels of trust beforehand. Discuss all sorts of topics and allow for a lot of freewheeling conversation about fundamentals before you buckle down to work.
> 2) Design a method and framework and develop a clear division of labor.
> 3) Split the difference wherever possible. We veered towards choosing the least controversial phrasing, but we also believe that a noncontroversial statement said well does a lot of good. Consistently work then discuss.
> 4) The fact that our topic was broad gave us plenty of thought-space to work with each.
> 5) Have a partner that is incentivized. We were both were enthusiastic about the Ad Collab format, confident we could make a contribution to the field of our topic, and the possibility of money was meaningful.
> 6) We both had time in our schedules. It was summer, and we both worked in academic and educational institutions at the time.

From Mark W, who wrote the pro-vaccine side of the vaccination collaboration:

> WHAT WERE YOUR INITIAL POSITIONS?
>
> My initial position was that herd immunity is a compelling government interest, and that it is well within the power of a reasonable government to implement a mandatory policy in order to achieve this end.
>
> Interestingly, the first part of this statement (that herd immunity is a compelling government interest) was not challenged by my collaborator, so I think we spent less time making the case in favor of herd immunity than we

probably should have. This was likely a mistake on my part, as I should have been spending more time shoring up my side – even those parts we agreed on – as this was one of the areas some commentators found lacking.

(Of course, if you are adversarially collaborating in bad faith, you could choose to agree quickly and dismiss significant parts of your opponent's argument so as to focus less energy defending their strongest points. I'm not saying my collaborator did that; I think he was arguing in good faith. I'm just saying that you could…)

On the second point, Mark Davis began by saying that he believes vaccines are a major causal factor in autoimmune and autoinflammatory disease. I responded that we looked at that angle back in grad school (about 10 years ago for me) and my recollection of the evidence was that vaccines are not a good explanation for this phenomenon.

HOW MUCH DID YOUR POSITIONS SHIFT, AND IN WHAT WAYS?

My position that mandatory vaccinations is good policy shifted to the extent that we showed it's not required to achieve herd immunity. I would say that as a matter of public policy influence, the burden falls on the side opposed to mandates to persuade the public that there's a better way. On the side of vaccines causing autoimmunity, I was not swayed much. I think it's possible that mycobacterial vaccines contribute a little, but the problem is that there are too many types of mycobacteria, and families of protective/sensitizing commensals are so broad that a highly-specific intervention such as vaccination is unlikely to be able to wipe out a whole family like that and leave you sensitized to developing autoimmunity. Mark Davis wanted to put forward the hypothesis that not only are vaccines bad (because, he believes, they cause autoimmunity) but that the bad outweighs the good they otherwise cause. I felt that one point we made in the essay was that even if it turns out that vaccines contribute significantly to autoimmunity (a claim I don’t find much evidence for) he is presenting a false choice. The public health implications are not either disease control and unchecked autoimmunity, or unchecked disease and autoimmunity control. There are populations that don’t have disease exposure and that don’t have autoimmunity, and there’s no good reason to jettison vaccines and the positive good that they bring (which we both agreed on) out of a belief that otherwise we have to live with autoimmunity.

There’s one point that didn’t make it into the final essay – due to time constraints and narrative flow – I wanted to add. If you believe that autoimmunity is prevented by exposure to pathogens that are vaccinated against, this has dire ethical implications. It’s one thing to believe that the measles vaccine caused your child to get asthma. Even if you’re wrong and unpersuadable, your proposed solution will be to not vaccinate your child. If you believe that a lack of disease exposure caused your child to get asthma, you’ll want to actively work to ensure that all your children contract each disease you refuse to vaccinate against. Otherwise, it makes no sense to not vaccinate against something, if you think the causal mechanism requires exposure, too. Thus, there is an ethical problem here, not with allowing parents to opt-out of vaccination, but with allowing parents to actively seek to spread disease through their communities. If a parent drives around with an 18-month-old in the back seat with no seatbelt, then intentionally gets into a car wreck, they would be criminally liable – even if they were only caught attempting such a feat. There’s a good case to be made that intentionally infecting your child with a disease known to cause morbidity and/or mortality should carry criminal liability. Davis didn’t make this exact case, but it’s really the only natural extension of his claim. If you oppose vaccination on the grounds that you believe it’s necessary to experience disease exposure in order to protect against autoimmunity, you have to be willing to expose your children to disease.

HOW MUCH DEBATE AND ARGUMENT WAS THERE?

Most of the debate was about whether vaccines can lead to autoimmunity. This was the driving factor behind our time delays. We went through multiple drafts surrounding this point as a main focus of conflict. Davis sent a couple of articles suggesting some causal link between vaccines and autoimmunity. I sent back about a dozen more that argued causal link, no effect, and protection. I also pointed out that the review he sent me actively argued against his point, as did every other review on the subject I could find and multiple meta-analyses.

HOW DID YOU RESOLVE IT?

Davis maintained that this body of evidence was not sufficiently strong to persuade him. At the outset, we had agreed that the burden of proof lay with his side in defending an active harm perpetrated by vaccination. Given that the evidence trended in the opposite direction, he conceded to allow language that public policy should not discourage vaccination – on the condition that private individuals be enabled to disagree with public consensus.

IS THE CONCLUSION CLOSER TO ONE PERSON’S ORIGINAL POSITION?

I’m not sure. I’d say the first part (about mandatory vaccination) preserves my stipulation that elimination of mandates is only okay insofar as the objective of herd immunity is maintained. Thus, even though I capitulated, I did so within the realm of broader public policy objectives I believe in. For the second part, we probably stuck closer to my side, in that there was simply no way to advise public policy changes based on weak evidence of hypothesized harms. That said, there were lots of caveats and hedging.

WHAT ADVICE WOULD YOU GIVE FUTURE ADVERSARIAL COLLABORATORS?

Some thoughts on my experience:

– Time was definitely a factor. Repeat back-and-forth drafts took the most time, because one person would usually have it for a week or more before turning it back to the other person.
That means at least two weeks per draft. If something like this were to become a regular feature, I would advise future collaborators to allow extra time, and maybe wait to submit in a future round. – Every time I tried to add a strong statement, such as “we have no evidence to support [X]” “there is no reason to believe [X] and ample reason to believe [Y] instead” he would object. I suspect there’s a strong tendency as a collaborator to not capitulate your side.
– That said; I did end up capitulating on part of my side

(the mandate part) because we stated at the outset under what conditions we would do so. My conditions were met (other countries without mandates achieve herd immunity, so clearly it’s not an absolute requirement, and some countries with mandates do not achieve herd immunity, so clearly it’s not sufficient), so I conceded that point.
– My collaborator stopped there, however, so the most we could recommend was, “sure they do things differently elsewhere, but we don’t really know what that entails; go look it up yourself.”
– We spent most of our time on the particularly contentious second part of the collaboration. This may have been why the first part was weaker than it could have been.
– The assigned titles tended to focus public discussion on what they thought was the main debate, but often that was either not the whole debate, or the debate had shifted as a result of the collaboration.

If you choose to do this again in the future, I would recommend a few changes:
– Have collaborators choose titles.
– Require collaborators to stake out one or two specific claims they wish to refute/defend a priori, and have that be outlined in the first part of the final submission. I know we mostly did this with the bolded statements, but not all of the positions made it into the final piece, so it was hard to tell where people were coming from.
– Have the collaborators include conditions that would be necessary for them to concede their positions. I feel like it would be within the role of a moderator to remand any concessions that are unrealistic, or not well-matched to the claim. (i.e. “I will be convinced once 1,000 years’ of RCTs have been completed.”)
– Require a concluding section that states whether the

> conditions were met, rejected, or uncertain – as well as what should happen moving forward.
> – Where possible, include a third, impartial, moderating collaborator. This is theoretically the person with no horse in the race. If there's disagreement, they can say to one side, "sure, but you're not arguing based on the evidence anymore, you've been bested on that front; we need to draft stronger language that reflects the actual state of the evidence."
> – The SSC community is a great resource. In retrospect, it would have been nice to have used part of an open thread to discuss elements of the ACC. For example, asking an
> international audience about what their on-the-ground vaccination experience is like in the UK/Ireland/France/etc. I should have done this unofficially, but it would also have been helpful to have an official, "first comment is for ACC contributors to ask the community for input". That would also allow the rest of the community to contribute to the contest.

From a_reader, who wrote the anti-transition side of the transgender collaboration:

> WHAT WERE YOUR INITIAL POSITIONS?
>
> My position was that "transgender children" (gender dysphoric children) shouldn't transition in childhood – neither socially nor by suppressing puberty at 12 (or even earlier!)
>
> Because a lot of them – percent differs among sources, but something between 30-90% – desist, cease to feel and act like the opposite sex. And that's a better outcome: not needing surgery and hormones for the entire life, having less problems in society acceptance

(even as gays/lesbians, as most desisters become) and in finding partners.

If dysphoria persists from 3 to 18, that’s it, transition may be necessary, but it’s better if it’s not necessary, if the problem ceases spontaneously.

Social transition in childhood (before adolescence) – that means name, pronouns, hairstyle, clothes of the desired gender – makes desisting more difficult. But puberty blockers at 12, I thought, can do worse: prevent the desistance from happening, stopping exactly the natural process that may have cured gender dysphoria, by the action of puberty hormones on the brain.

HOW MUCH DID YOUR POSITIONS SHIFT?

My position didn’t shift that much, but now I understand better why so many parents choose to let their children transition and take puberty blockers, how distressful gender dysphoria is for those children – anxiety, depressing symptoms, sometimes (at puberty) even suicide attempts – and social transition and even puberty blockers seem to alleviate it substantially.

I still consider (spontaneous) desistance a better outcome. And I still tend to think that maybe puberty blockers may stop that process. But I admit I didn’t found conclusive evidence (only circumstantial: all children on puberty blockers persist).

HOW MUCH DEBATE AND ARGUMENT WAS THERE?

Not much. In the beginning, flame7926 created a document shared by Google Docs, for documentation, where both of us wrote about the studies (and media articles and blog posts describing cases) we found. One found a study and wrote a few lines about it, then the other added their own observations – using different colors each of us at different times, so that in the end that document looked like a rainbow, matching the subject.

HOW DID YOU RESOLVE IT?

After the documentation stage, flame7926 asked me if I changed my position, I said not really, so after a short discussion, we agreed to not recommend for or against childhood transition, to just present all the evidence, the risks and benefits.

IS THE CONCLUSION CLOSER TO ONE PERSON'S POSITIONS?

The conclusion seems to be that there isn't yet enough evidence for a firm, definitive conclusion. So it tries to keep a balance between our positions, resuming the evidence and admitting uncertainties.

WHAT ADVICE WOULD YOU GIVE FUTURE ADVERSARIAL COLLABORATORS?

I think besides conscientiousness, you need a degree of agreeableness and being a mistake theorist, not a conflict theorist (at least not about the subject); it is also important to have enough time (that is, quite a lot) and to find the subject interesting (but not essential, not among your main values – otherwise you risk to become a conflict theorist).

I was really lucky to have such a good "adversary" like flame7926, who in the end worked more than me: he wrote most of the final entry (as you probably have already noticed, I'm not a native English speaker).

> Seeing how many people were abandoned by their partners, I realize how lucky I was.

And from flame7926, who wrote the pro-transition side of the same collaboration:

> I enjoyed the adversarial collaboration process and like the idea a lot. I think that it has the potential to produce very productive discussions and papers. I think it definitely can moderate individuals’ opinions – get them to recognize that many things aren’t so black and white.
>
> To begin, I thought that transgender children should be allowed whatever options they want in terms of socially transitioning, while I had less strong of an opinion on puberty blockers (mostly because I didn’t know much about them). Honestly, I didn’t know that much about gender dysphoria or transgender expression in children and youth before commencing with this project, so my opinion wasn’t very formed. I did feel strongly though that we not present “desistence” as a preferred option because I think it is at least reminiscent of transphobic and homophobic attempts to convert individuals away from their identity – an often very traumatic and distressing process. And I thought that once you remove the idea that desistence is preferred (because I think it is very difficult to have that macro-level preference while remaining supportive of individual children on a micro-level) then social transitioning becomes preferred. I also thought that these cases of “desistence” might just be people expressing cross-gender behavior and not actual transgender identities.
>
> I think my position moderated some towards a more even approach, and I can understand why parents and others would want children to desist from transgender

identities (as long as this happens without trauma). I also stand stronger behind current research on desistence, in the sense that I do think there is strong evidence that a sizable fraction of children who express desire to either be the other gender or that they are the other gender will not express those same desires later in life. Yet I still think (and believe the evidence shows) that social support of children is extremely important (in terms of mental health), and I think denying their desire to socially transition can run counter to that. I also think the example of Samoa and other societies shows that gender dysphoric feelings are not constrained to our culture, but that the mental health associations are culturally contingent.

I would say there was some debate and argument. As a_reader describes in their response, we had a collaborative Google Doc broken down by topic, like social transitioning, puberty blockers, mental health etc. One person would pull some evidence from an article, the other person might pull some contrary evidence. Debate proceeded in that fashion, with more and more sources being brought in. There wasn’t that much overarching debate, but I think our topic was broad enough that it needed to be broken down in components in which we actually examined the evidence.

When it came time to write our paper, it seemed that neither of our positions had shifted that much. We presented the evidence, but I don’t believe there is currently enough evidence on puberty blockers to make a recommendation either way. Given that we did not end up agreeing on social transitioning, we could not recommend either way. I did end up writing most of the paper (at the request of my co-author) and am glad that

> we were able to produce something that moderated and expressed both viewpoints. I’m also proud of how we integrated quantitative evidence with more qualitative sources and anecdotal accounts, which I also feel are important when dealing with issues that are, necessarily, subjective.
>
> In terms of advice, I think just being able to put in the time is important. Seeing the small number of projects that were completed, I wonder why so many dropped out. I’m glad I had such a motivated partner (I would say that they did more work than I did). I think that it is also important to be able to moderate your viewpoints and understand where your disagreements come from, even if you don’t end up reaching the same conclusion. I think that having someone else working can keep you on track, if you see them making process. I am curious how other groups divided up the writing on their projects (especially since I believe that most of them ended up longer than ours?).

My executive summary: almost no one changed their mind on the overall issue (a few people changed on smaller subpoints), but almost everyone moderated their opinion at least a little.

Some of the collaborations raised concerns I hadn’t considered. Although John and Christian were happy with their result and many other people liked their collaboration, I still can’t shake the feeling that they solved their intractable disagreement by avoiding their actual cruxes and searching under a streetlight they didn’t lose their keys near. If you force two people to write a mutually agreeable essay on a topic they can’t agree on, it’s probably hard to prevent them from inadvertently shifting to a slightly different topic where everything is cut and dried and factual and agreement is easy.

I was originally very excited to see the vaccine collaboration come out strongly in favor of “vaccines do not raise disease risk”, since it looked like someone had really changed their mind on a controversial topic – then disappointed again once I learned that vaccine opponent Mark Davis had not been convinced, but just agreed that the burden of proof was on him and he hadn’t met it by convincing his adversary. If I were a vaccine opponent, I would be pretty upset that the adversarial collaboration comes out strongly in favor of vaccines even though the two collaborators weren’t both on board.

And as a vaccine proponent, I worry that Davis’ magnamity in allowing the final result to be mostly pro-vaccine convinced Webb to “reciprocate” by being less forceful than he could have been in other areas, and not fight back too hard against the technically-true-but-fraught claim that other countries had higher vaccination rates than the US without mandatory vaccination programs. I don’t know if this really happened in this collaboration (from the write-up it looks more like they agreed on the preconditions under which they would give in, and maybe those weren’t thought out too well) but I think the two adversaries agreeing to “compromise” by each giving up on one of their key points, rather than really come to agreement or present their disagreement exhaustively, is a failure mode to be watched for.

This particularly bothered me in the vaccine collaboration because of another factor I hadn’t considered. I previously hoped adversarial collaborations could be a good match for “pseudoscience”, ie questions where the mainstream is overwhelmingly on one side but there are still some holdouts with an alternate narrative. Certainly these topics cry out for explanation, but the current paradigm – RationalWiki-type “skeptics” who make fun of pseudoscience with cheap shots – are really bad at convincing believers and usually mangle the theories they try to criticize. An adversarial collaboration format could make sure that debunkers have to address pseudoscientists’ strongest arguments – and that pseudoscientists’ supposed counterarguments to the debunking get addressed also.

I now worry that such collaborations would systematically end up legitimizing pseudoscience. If nearly everyone believes that some theory (UFOs, Bigfoot, etc) is false, and some prominent proponent and opponent of the theory do a collaboration, then even if everything goes great and the end result is mostly skewed to the opponent’s side, then there will still be a few compromises and places where the opponent gives into a little social pressure to frame things in a less-than-maximally-hostile way. So if the end result comes out 90% anti-Bigfoot – well, believing in Bigfoot 10% is more than most people currently believe in Bigfoot, so most credulous readers would update in a pro-Bigfoot direction. True, any debate between pro- and anti- Bigfoot factions will involve some pro-Bigfoot arguments that you wouldn’t have heard otherwise, but the adversarial collaboration format risks importing them into the final product in a way that makes it seem like they have a stamp of approval. I’m not sure this format is a good match for pseudoscience unless somebody finds a way around this.

Thanks to everyone who participated. I would like to do more with this format and will probably devote at least a similar amount of time and money to it next year; if anyone has ideas to improve the contest, or better ideas for promoting adversarial collaborations than another contest like this one, please let me know.

Steelmanning The NIMBYs

[Epistemic status: very unsure. I sympathize with many YIMBY ideas and might support them on net; this post is me exaggerating the NIMBY parts of my brain to a degree I'm not sure I honestly support. This focuses on San Francisco to make it easier, but other cities exist too. Thanks to Nintil for some of the bright-line argument in part four. Conflict of interest notice: I live in a lower-density part of Oakland]

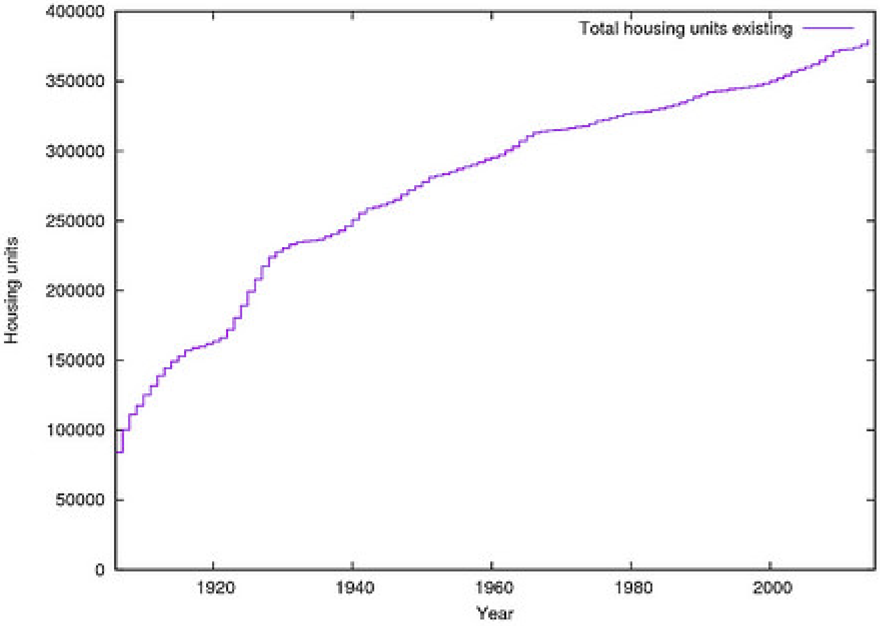
Everyone I know is a YIMBY – ie "Yes In My Back Yard" – ie somebody who wants cities (usually San Francisco dominates the discussion) to build more and denser housing. This is a reasonable position, and is held by apparently-reasonable people – centrists, rationalists, economists, self-proclaimed neoliberals. Since everyone involved holds reason and civility as an important value, I would expect the discourse around housing to be unusually reasonable and civil.

I have a weird habit of encountering the best parts of some movements and the worst parts of other movements, in a way that doesn't match other people's experiences. And certainly I know many YIMBYs who are amazing people who I love. But as for the movement as a whole, I feel like apparently-reasonable people have dropped the ball on this one. Sorry for having to say this, but YIMBYism is one of the most tribal, most emotional, most closed-minded movements I have ever seen this side of a college campus. So much so that even though I agree with much of what it says, I cannot resist writing a 5,000 word steelman of their enemies just to piss them off.

So here are some YIMBY claims and why I cannot be entirely on board with them.

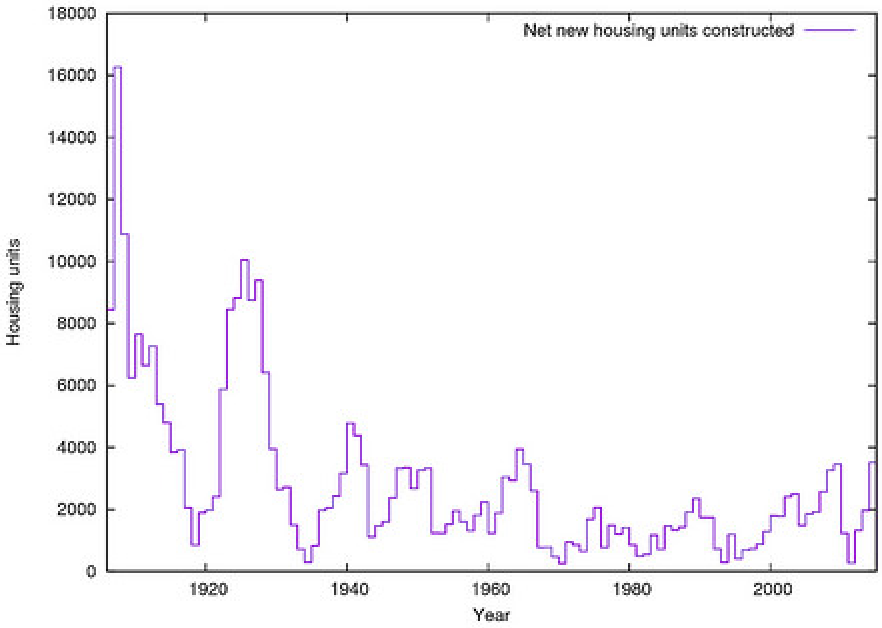
1. San Francisco is uniquely terrible at building new housing

San Francisco currently has just short of 400,000 houses.



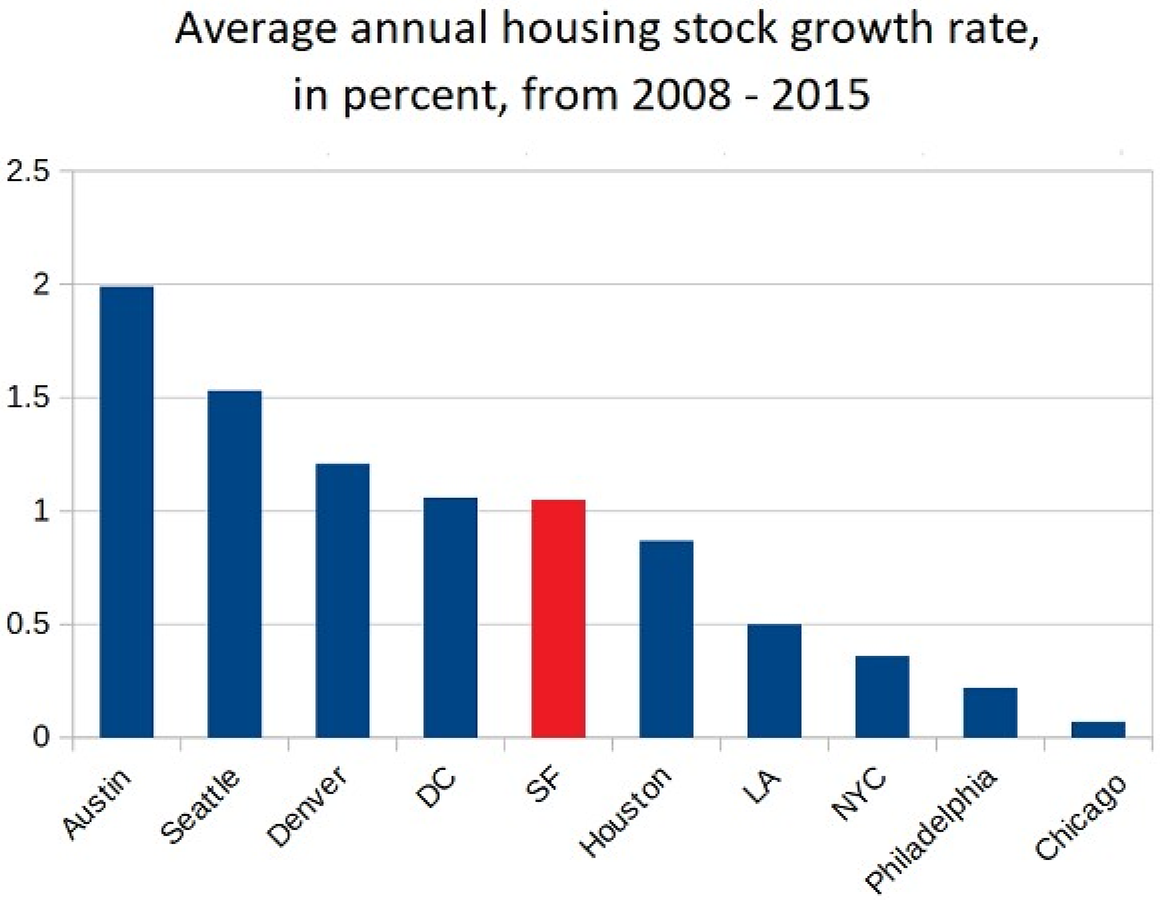
(*source*)

Each year, it builds a few thousand new houses, for a long-term growth rate hovering a little above 0.5%.



(source)

How does this compare to other cities? I used data from Civic Dashboards to compare the housing stock growth rate of ten major US cities. They only had data from 2008 – 2015, so the analysis only includes those years. They find a higher SF growth rate than listed above, probably because growth has been increasing recently. Here's what they got:



San Francisco is actually doing pretty okay. [**EDIT**: Commenter peopleneedaplacetogo points out [a chart]() by metropolitan area rather than city, and using slightly different years, in which SF comes out looking quite a bit worse]

The problem isn’t that SF is building fewer houses than other cities in its league. It’s that demand keeps increasing so much that a normal amount of housing construction doesn’t help.

This might be an unfair objection, because the YIMBY argument might be that San Francisco is uniquely terrible at responding to demand for new housing, and this may be true. But it will important to get a sense for the range of levels of housing construction different cities are capable of, so we can better understand what scenarios are plausible in the next section.

2. Building more housing in San Francisco is an easy way to lower rents

Lowering rents in San Francisco is certainly important: a 1-bedroom apartment costs about $3500. At prices like these, city natives may have to move out because they can no longer afford rent. The lower- and middle- class citizens who work service jobs and maintain infrastructure either disappear or are faced with multiple-hour commutes from the nearest affordable area. Even tech workers with good salaries have to live in overcrowded apartments with multiple roommates to make ends meet. Facets of a good life that depend on having lots of space – like having social gatherings or raising a family – become almost impossible.

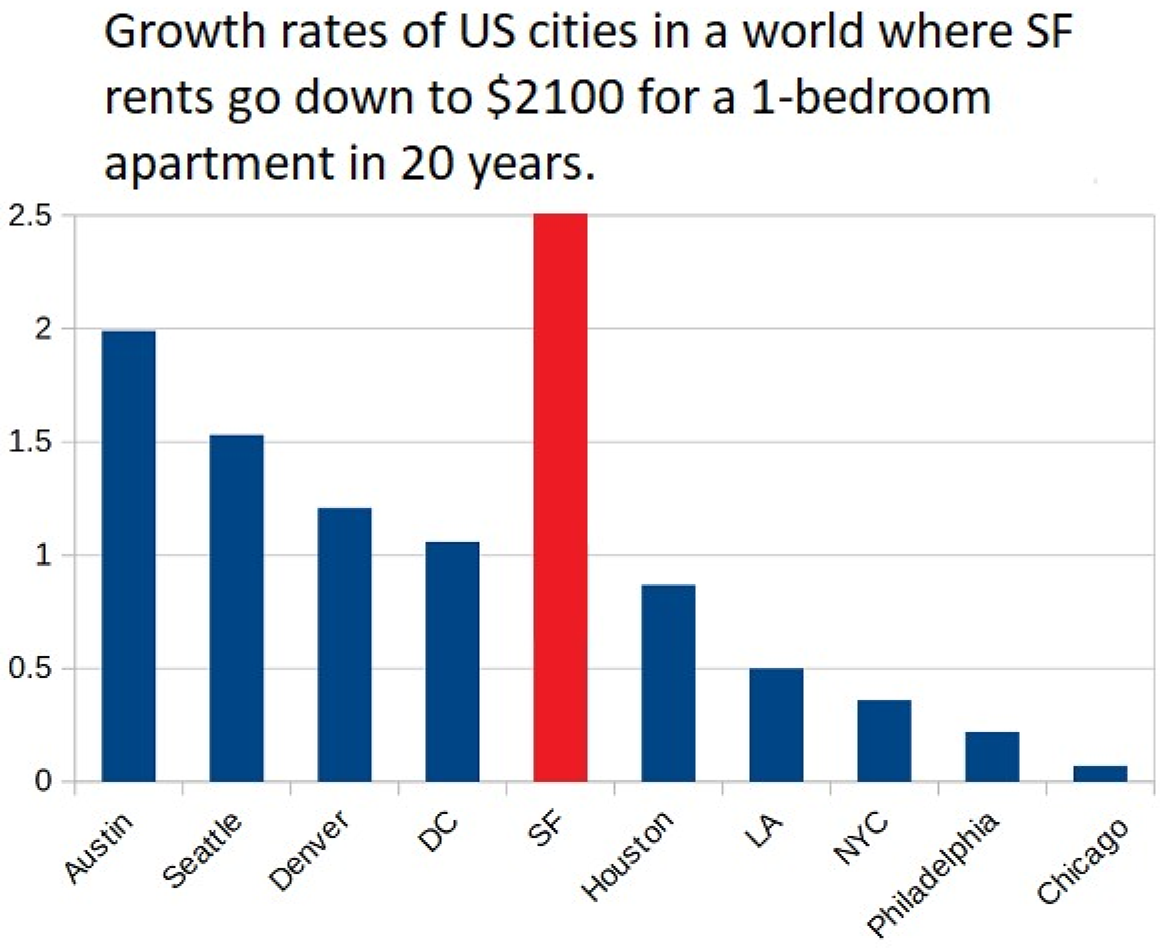
The laws of supply and demand suggest that if San Francisco built more housing, the price would go down. This is the foundation of YIMBYism, and it’s basically correct.

But how much would price go down? This requires some economic modeling, which has luckily been done for us.

The *San Francisco Examiner* follows a paper by Albouy, Ehrlich, and Liu that estimates a 2% increase in housing will cause a 3% decrease in rents. On the other hand, by the time San Francisco has finished building 2% more housing, the population and demand will have increased, meaning that a large portion of the gains will be expended just staying in the same place. They come up with a model that accounts for this, and set themselves a goal of decreasing the average rent for a one-bedroom apartment to “only”

$2100 – at least we can’t accuse them of being too ambitious!

They find that this would require a 2.5% housing stock growth rate maintained over twenty years. Going back to the graph from before:

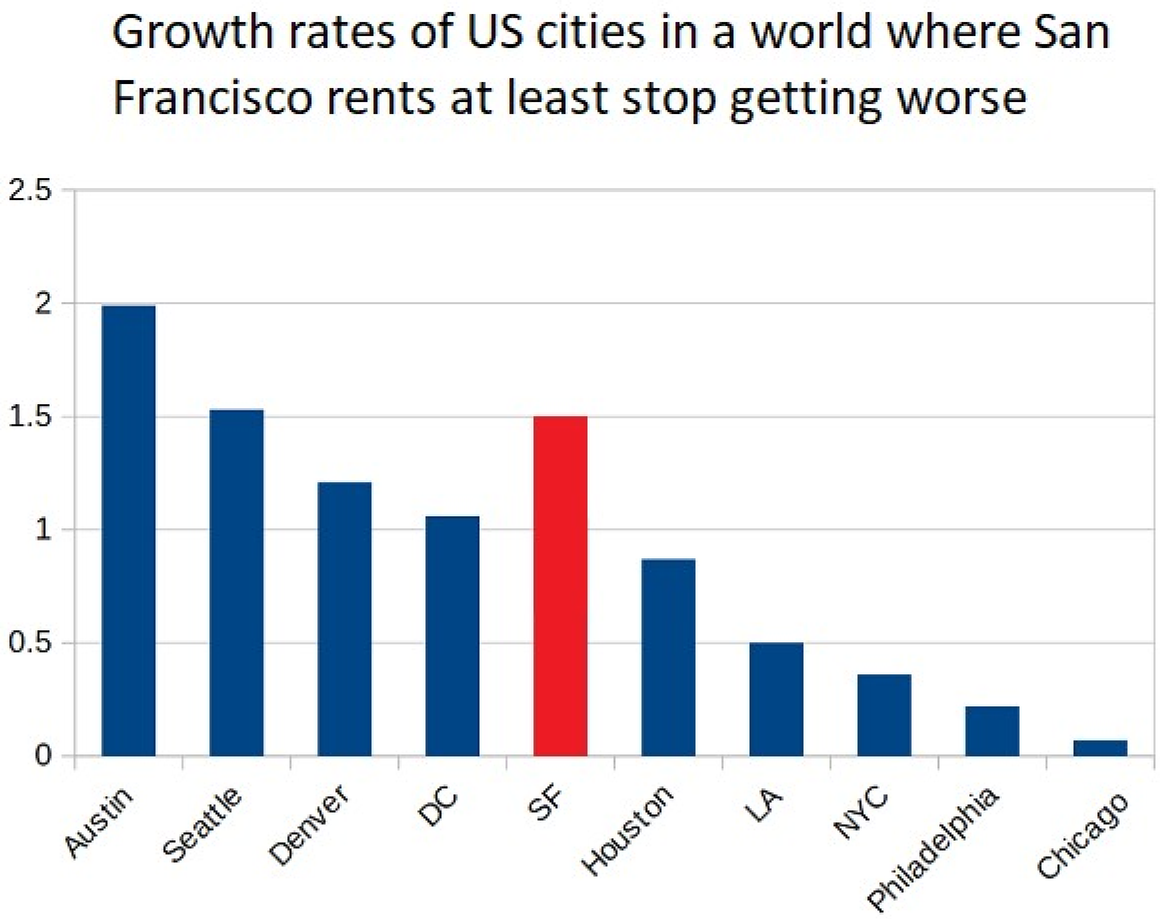


No large US city was able to attain this rate in the eight year period my data comes from, including cities experiencing tech booms during those years. Austin, Texas was able to come close. But at the time, Austin had a population density of 2,500/sqm. San Francisco has a density of 19,000/sqm. Building new houses is easy if all you have to do is clear away tumbleweeds and rattlesnakes – and Austin *still* only made it up to 2%. We’re expecting San Francisco to clear away existing neighborhoods and angry anti-development activists, and reach 2.5%? And maintain that rate for twenty years?

[**EDIT:** Commenters point out that the time period my data covers is unusually bad for housing growth. Austin’s records show it was able to grow faster – 4.8% – until 2000, although during that time it was a very small city expanding into mostly empty space. Since 2000, growth has averaged about 2.8%, which is slightly faster than the SF scenario above.]

And even if it works – even if the city can do the impossible – that only lowers rents down to $2100 for a single-bedroom apartment.

Experimental Geography uses broadly the same model, but asks a different question: how much does San Francisco have to increase housing just to tread water and not have rents keep going up? You can read their reasoning at the link, but the answer is “1.5%”:


Growth rates of US cities in a world where San Francisco rents at least stop getting worse

This seems potentially achievable, but still difficult. A paper by the Federal Reserve finds similarly grim results. I’m not aware of any models that have come to the opposite conclusion.

True, every little bit helps. But affordable housing advocates frequently say that complicated policies like public housing or subsidies are necessary to help poor people who want to stay in high-cost cities. I often hear YIMBYs push new construction as an alternative to these measures, saying that all we need is increased housing supply. But even in the best-case scenario, increased housing supply will take decades to do anything, and lower-income people are at risk of losing their houses now. And even in the best case scenario, increased housing supply will just make apartments *slightly* more affordable. It’s less likely they can let low-income people live comfortably in the city, or high-income people comfortably raise families there.

3. But at least building more housing will make things a little better, and it certainly can’t make them worse, right?

Devon Zuegel points out that we’re really not sure if that’s true. Why does Manhattan have higher land values than Kansas? Because people want to live where other people (and jobs) are. The denser you make a city, the more other people and jobs will be there, and the higher the land values will get.

Or to put it another way – suppose San Francisco dectupled its housing growth for decades, until it was packed border to border with skyscrapers, and was exactly as dense as Manhattan. In a simple supply-based model, the glut of supply should make rents crash to only a few hundred dollars a month or less. But in actual Manhattan, single-bedroom apartments cost $3800 a month – even more than in San Francisco! If your theory predicts that turning a city into Manhattan will make rents plummet, then consider that turning Manhattan into Manhattan made rents much worse, and so maybe your theory is wrong.

Devon points out that she cannot calculate the coefficients here, so she is not sure whether building more housing will make rents go down (because of supply and demand) or up (because of the Manhattan effect). But we might consider Austin a natural experiment. The model above found that if San Francisco grew housing at 2.5% per year for twenty years, rents would go down by a third. But Austin grew housing by 2.0% per year for about twenty years, and during that time, the average cost of a house doubled. I am not sure San Francisco, which starts from a much higher baseline density, would see the same trend. But at the very least, agglomeration effects suggest all of the terrible and pessimistic models above *are still overly optimistic*.

Tripling San Francisco’s housing rate until it’s higher than any existing American city, and maintaining it at this rate for an entire generation, *might* make one-bedroom apartments cost “only” $2100. Or it might do less, or nothing, or make things worse. Right now we don’t know.

4. Holdouts who oppose development are inexcusably selfish, or hate poor people, or are racist

If you want to see real loathing, don’t ask a communist about the rich, or a Trump voter about immigrants. Ask a YIMBY what they think of landowners in a nice quiet part of the Bay who don’t want San Francisco spreading to their area, or who don’t want the BART light rail line connecting their city to San Francisco.

And if you want to see great acting, don’t go to Hollywood or Broadway. Wait for a YIMBY to start monologuing their impression of what these people are like. The exact script differs from person to person, but always includes liberal use of phrases like “the poors”, “brown people”, and “I’ve got mine”.

But I sympathize with these landowners. San Francisco is easy to hate. Even a lot of the people who already live there hate it. They hate the streets piled with discarded needles and human waste. They hate the traffic (fifth worst in the world) and the crime (third most property crime in the US). They hate living five people to a three-bedroom apartment. They hate having aggressive people scream incomprehensible things at them on the sidewalk. They hate the various mutually hostile transit systems that interlock in a system I would call byzantine except that at least you could get around medieval Constantinople without checking whether the Muni and CalTrain were mysteriously failing to connect to each other today. They hate that everyone else in the city hates them, from visible KILL ALL TECHIES graffiti on their commute to work, to a subtle mood of seething resentment from everyone they meet. They hate the omnipresent billboards expecting them to have strong opinions on apps.

I'm not saying everyone in San Francisco hates it. There are people who like all sorts of things. Some people like being tied up, whipped, and electrocuted by strangers. And a disproportionate number of these people live in San Francisco. I am just saying this isn't a coincidence.

And I sympathize with the people who don't want BART stations near them. BART stations are also easy to hate. I have a friend who ended up needing stitches after ill-advisedly walking too close to a BART station late at night and getting robbed and beaten up. One of my patients is currently freaking out after their friend ended up in the hospital for the same reason. Some women avoid getting beaten up, but still have stories of getting groped or sexually harassed. BART stations tend to collect a penumbra of litter, drug use, weird people playing incredibly loud music at all hours of the night, weird people shouting at each other at all hours of the night, and the never-dissipating stench of marijuana mixed with urine. This stuff is usually just background noise, but it did make the news last year when forty to sixty teenage thieves took over a BART car in the station and robbed and beat up the passengers, and then again earlier this summer when there were three unrelated murders at BART stations in one week.

These don’t seem to have been gang shootouts or anything – they were just people trying to get on their train and getting randomly murdered instead. I am very aware I could get murdered every time I get on a BART. Last time I got off one (three days ago), there was a guy standing in front of the door shouting “FUCK YOU KKK WHITE BITCH” at any woman (of any race) trying to enter or leave the station. Nobody found this surprising or unusual. It’s just what BARTs are like.

But tell a YIMBY that someone, somewhere, is against having a BART station in their neighborhood, and it’s like waving a red flag at a bull.

Maybe clear-cutting everything in the way of San Francisco’s expansion is the utilitarian correct thing to do. Maybe it would increase the US economy so much that we can’t afford not to do it. But Thomas Hobbes wrote that sovereigns may not demand someone go willingly to their death, because resisting death is such a natural human urge that people in the state of nature could not sign it away when forming a primordial state. And some European countries don’t count resisting arrest as a crime, because they consider freedom so fundamental that nobody can be blameworthy for trying to protect it. I believe some people need to have BART stations near their houses, just like some people need get arrested or be executed. But resisting each of these seems so natural and fundamental that I am unwilling to blame anyone for trying. I think neurotypical people usually underestimate how bad cities are for people with noise sensitivities, anxiety, purity intuitions, or just a need for nature and green things in their environment, and that the campaign against people who want to keep their suburbs suburban doesn’t take this into account.

I have heard YIMBYs counter that we don’t have to turn Marin County into San Francisco II, that there’s a balance between trying to preserve what’s good about a place and reflexively opposing all new development. But on slippery slopes where a coalition of people with slightly different preferences are trying to coordinate, drawing a bright line and refusing to cross it is the theoretically correct solution. This is especially true when each new development brings in new voters who may be less attached to the current nature of a place and more willing to vote in future development. Sometimes the only stable solution is just to not get on the slope.

And I have heard YIMBYs counter that if people don’t want to live in an urban environment, they shouldn’t have bought a house in a city. But they kind of didn’t. They bought a house in a medium-density suburb, then some other people came and said “No, this has to be a city”. If they give up, let San Francisco spread to their current home, and move to another medium-density suburb, what’s to prevent other people from trying to urbanize there too? Is our social technology just totally unable to deal with the problem of “how can we let people who want to live in a medium-density suburb live in a medium-density suburb?” Wasn’t the solution supposed to be “these people all gather together, start a community together built to their unique specifications, incorporate, and then pass laws about what their community can or can’t include”? What was wrong with that solution? Some people can’t tolerate the big city – what do you want them to do? Sell their house, leave all their friends and family, and try to start again somewhere else? You think that’s an exaggeration? If where I live became more like San Francisco, *I* would do that. Lots of people would!

I’m not saying there aren’t compelling reasons to urbanize less-dense areas. But I do feel like there’s a missing mood here that makes me really upset whenever I hear YIMBYs talk about this.

5. Even if building more housing doesn’t lower costs, it will at least allow the people who want to live in San Francisco to do so.

I don’t like the “even if” framing.

If you have been pushing a claim for years, and it turns out you were wrong, then you owe it to the people you have been disagreeing with to say “oops” and take a moment to worry about whether you should lower your smugness level in general. I try to do this when I remember, though I am not always good about it and of course I am limited by my ability to catch and correct my own mistakes.

(I have sometimes been guilty of pushing the “all we need is more housing claim”, so, uh, oops, sorry, upon doing more research I see I was wrong)

The opposite of this is saying “*Even if* that’s not true, this other thing supports my point”, without explicitly conceding anything at all.

5b. Okay, sorry. Oops, I was wrong about the housing prices. Now that I’ve said that, don’t you also think that building more housing would at least allow the people who want to live in San Francisco to do so?

This is a good and important point, and I think the strongest one in the YIMBY arsenal. I am not really against it, but I can think of two qualms I have with it.

The first is that the argument for ignoring the costs of new construction to existing communities have always relied on the humanitarian necessity of lowering rents for the disadvantaged. If all we're trying to do is be able to pack a few more people who can pay $3500 a month in, the humanitarian necessity seems less pressing.

But second, bringing more people in helps trap the economy in the same dynamics that caused this problem in the first place.

Some people really enjoy living in dense cities like San Francisco. Other people, for the reasons listed above, really prefer not to. Many of the people who prefer not to are in San Francisco anyway. I signed up to work in the suburbs, but just before I started, my group begged me to work a few days a week in San Francisco because that was where they needed more doctors. I grudgingly agreed. During my time there, I treated depressed San Francisco residents. One refrain I heard again and again was that they hated living in San Francisco, but had come anyway because their company pressured them, or because their companies would pay them extra, or because that was where all the best jobs in their industry were. These people's long-term plan was to use San Francisco as a springboard to gain enough money or career capital to be able to achieve their dream of leaving San Francisco.

Alon Levy describes the same thing his In The Mines, where he compares the outlook of people moving to San Francisco to that of people working in mines or oil rigs. Nobody likes working in a mine or oil rig. They go there because it pays really well, and if they grin and bear it for long enough, they can pay off their debts or save for the future or do something that allows them to live in a place that isn't a mine or oil rig:

. People who work on oil rigs, which as a rule are placed in remote locations, get paid premiums. Remote locations with oil have high incomes and high costs in North America, but even the Soviet Union paid people who freely migrated to Siberia or the far north extra. The high wages in this industry are especially remarkable given that the workers are typically not university-educated or (in the US) unionized; they cover for poor living conditions, and a hostile environment especially for families.

I bring up this background because of conditions that I’ve heard second-hand in San Francisco. When I first heard of university-educated adults living several to a bedroom, I assumed that it was a result of extremely high rents and insufficient incomes. But no: I am told a reasonably transit-accessible two-bedroom in San Francisco proper is $5,500 a month at market rate, which is affordable to a mid-level programmer at a large tech firm living alone or to entry-level programmers (or non-tech professionals) living one to a bedroom.

And yet, I’ve heard of Google programmers living two to three to a bedroom in Bernal Heights, not even that close to BART. I’ve also heard a story of people near the Ashby BART stop in Berkeley renting out their front porch; the person sleeping the porch was not a coder, but some of the people living inside the house were.

I have not talked to the people in these situations, only to friends in Boston who live one person (or one couple) to a bedroom, even though they too can afford more. As I understand it, they treat the Bay Area as like working in the mines. They earn a multiple of the income they would in other industries with their education and skills, and have no particular ties to the region. (Some East

> Coasters have taken to use the expression “drain to the Bay,” complaining that friends in tech often end up leaving Boston for San Francisco.) The plan is to save money and then retire in their 30s, or take a lower-paying job in a lower-cost city and start a family there.

But people have to grudgingly endure poor conditions aboard oil rigs because they’re the only place you can pump oil. Why do they need to grudgingly endure poor conditions in San Francisco?

My understanding is that some industries like technology benefit from centralization. The more programmers are in a city, the easier it is to run a tech company there. The more tech companies are in a city, the easier the job search is for programmers. The more entrepreneurs are in a city, the easier it is to be a venture capitalist there. The more venture capitalists are in a city, the better it is to be an entrepreneur. Add useful infrastructure like Y Combinator and Triplebyte and maybe everyone in tech benefits from being in the same place.

This raises the possibility of a classic inadequate equilibrium, a situation that nobody likes but everyone is stuck in. For example, even if people don’t like Facebook’s privacy policy, interface, or anything else about Facebook, they mostly stick with Facebook because that’s where all their friends are and they’re not coordinated enough to move at the same time. Even if neither passengers nor drivers like Uber, they might use it anyway, because the passengers know that’s where they’ll get a driver soonest, and the drivers know that’s where they’ll get a passenger soonest, and nobody acting alone can break out of the trap.

But if centralization really increases productivity, hasn’t the market decided this is the best solution? I see two ways this might be false. First, it could be that centralization happened in the wrong place – that, if anyone had been able to centrally coordinate, the tech industry should have ended up in Austin or somewhere else that’s well-planned and has lots of geographical room to expand into. Second, it could be that centralization is just a game of keeping up with the Joneses. If there were no San Francisco, then some company would still end up employing the best programmer. But given that there is a San Francisco your company might have to move to San Francisco or have no chance of luring them away from all the companies that have.

Imagine that rising sea levels or something force the evacuation of the Bay Area. Google moves to San Diego, Facebook moves to Santa Barbara, and Twitter moves to San Rafael. Five years later, when Google programmers are sipping daiquiris on the beach in San Diego outside their affordable four-bedroom homes, are they thinking “Man, I wish I was in a crowded unliveable city with multiple inconsistently-connected transit networks right now”? Or are they happy that the option of not living in San Francisco suddenly opened up for them?

The other reason I often hear why people move to San Francisco even though they hate it is because everyone in their subculture is there. Lots of subcultures – queers, hippies, rationalists, etc – seem to be centering in San Francisco. But this might have similar dynamics to the tech situation. Suppose you’re a hippie living in St. Louis, and you’ll be happy as long as there are at least fifty other hippies to form a thriving hippie scene. All the other hippies in St. Louis have some number of other hippies they need to be happy. We can imagine a domino effect where one hippie leaves St. Louis, that causes another hippie to go beneath their threshold and leave St. Louis, that causes more hippies to go beneath their threshold, and so on, until there are no hippies in St. Louis anymore and you have to move to San Francisco or remain tragically un-hip. In this case, the best-case scenario for most St. Louis hippies is that the outflow to San Francisco is limited, so that St. Louis isn’t depleted of its hippie population as quickly. This is also good for hippiedom in general, since there might be proto-hippies in St. Louis who would join the scene if it existed, but who will never convert if all the St. Louis hippies are gone to SF.

(if it sounds like I’ve been thinking about this a lot, that’s because exactly these dynamics have been shaping rationalist communities in cities around the world for the past 5-10 years).

The hyperbolic worst case scenario is that centralization dynamics are too strong, and as more and more people move to San Francisco, life becomes harder and harder for the few remaining stragglers, until finally they give in. San Francisco becomes more and more crowded. Rents increase (through the process mentioned in part 3), number of people per bedroom increases, traffic increases, crime increases. Finally everyone lives in San Francisco, everyone hates it, and nobody can move out – unless they want to give up any chance of working in tech, and spend their entire life talking about cars and football with people named Bud. “San Francisco is unliveable, but at least we’ve made sure lots of people can live there!”

So the counterargument to “Every new housing unit built lets one more person move to San Francisco” is “Every new housing unit prevented saves one person from having to live in San Francisco”.

6. There are no alternatives

I’m not sure this one is wrong.

The argument in Part 5 seems much weaker than the other arguments – so weak that we should probably keep our usual policy of erring on the side of letting people live where they want.

And even if we didn’t want that – even if we thought centralization was a big problem that has to be fought against – it seems weird to leave the fight to crotchety old homeowners worried about noise pollution, and to hope that their self-interest coincidentally creates the world that is best for everybody.

I know there are a lot of urbanists who hate suburbs. I don’t. I grew up in a suburb consistently included in those Most Liveable Towns In The US ranking. It was really nice, and I often remember of it fondly when dealing with the stresses of living in slightly-more-urban Oakland due to me being a dirty rotten defector and participating in the centralization dynamics above. I wish for a world where everyone who wants has a chance to grow up in a nice suburb like that, and I don’t want anyone to have to live in a place like San Francisco unless they’re genuinely into that kind of thing.

I wish for a world with perfect coordination, where half the population of San Francisco decides to move to Helena, Montana at the same time. Half the number of tech companies in San Francisco ought to be enough tech companies for anybody, and the wide sky and endless plains would be a nice change of scenery. I wish for a world where hippies collectively choose Augusta, Maine as the new hippie capital, and so all of the hippies can move there and have great hippie culture and not have to fight with techies for the last $3000 apartment in the Mission.

I’ve heard some people say the federal government should take an active interest in decentralizing tech, since right now one well-placed tsunami could wipe out the United States’ entire technological advantage. I don’t know if this would be a good idea. I’ve heard other people say maybe we can just use virtual reality offices and VR teleconferencing to avoid the need for living anywhere in particular at all. I don’t know if this would be a good idea either.

Since none of those things will ever happen, I don’t know how to get to any of the worlds I want. If there are processes that favor centralization, I don’t know how to fight those processes. I don’t know if there’s some affordable housing policy that would really work. I don’t know if there’s something that balances the interests of every demographic. But I do think that just building more houses won’t, on its own, be a solution to the problem.

Highlights From The Comments On NIMBYs

Quixote writes:

> It’s odd to me how bad San Francisco is, when other large cities like New York or Paris are basically utopias.

But just a few comments down, Lasagna says:

> I despise (I’m choosing that word carefully) [New York City]. I still commute there every day, and I can’t stand it – the broken infrastructure, the horrible smells, the $14 for a yogurt and coffee in the morning, the massive crowds of unpleasant people (how could we NOT be? We’re walking through an open sewer). There’s a litany of other things that keep me permanently angry and depressed (just the thought of how much earlier I would have started a family if I didn’t live there….) I find it decadent, selfish, shallow – pick your bad adjective. I’ll stop now.
>
> Where I live now is nice. We have a town we can walk to, a lawn for the kids to play on and me to mow, we cook at home, we have enough room for our family to live and the kids to get exercise, even indoors. There’s no WAY I’m giving that up so I can live in an apartment again, all so NYC can squeeze MORE people into its area.
>
> If I had my way, we’d be much further away from the metro area than we are now, in a bigger, cheaper home with more land. But that isn’t possible; NYC is where my job is, and that’s that. Fine. But let’s not make things worse, and make NYC (and San Francisco, and DC, and

> Boston) even MORE indispensable generators of jobs. And please don’t think for a second that there aren’t sizable numbers of people like me, and like you, who do not want these things for our families […] Thanks for letting me rant. You should have seen the first draft of this thing. Twice as long, Scott. A litany of woes and anger.

This would be fascinating if it weren’t so predictable. One person describes NYC as “basically utopia”, and another person can’t stop ranting about how much he hates it and is glad to have escaped it.

In the same vein, from Cerastes:

> “I think neurotypical people usually underestimate how bad cities are for people with noise sensitivities, anxiety, purity intuitions, or just **a need for nature and green things in their environment**, …”
>
> THIS!!! A MILLION TIMES THIS!!
>
> The concept of living somewhere that isn’t green is literally nauseating to me, and the idea of a place that isn’t teeming with wild animals feels like suffocating. My house is in as wild a place as possible given my commute, budget, and region, and almost every room has a fully planted vivarium with an animal (as well as my office).
>
> The amount of urbanist triumphalist crap drives me up the wall, as if these people cannot see why someone would not want to live in conditions far inferior to even low-quality zoos, or why someone might need to balance a job in a city with such desires.

> Being 100% honest, I actually feel like there's something genuinely wrong with people who don't feel the need to spend time in nature, especially if they also lack pets. They're like sterile androids in some sort of weird dystopia, utterly cut off from life.

I think this is why these discussions are so hard. People's preferences on what makes an acceptable place to live differ so strongly, and in so many little ways, that I think a lot of the debate is just people screaming to have their existence acknowledged. It's infuriating to feel like everyone around you is calmly assuming as obvious and universal, preferences that could make your life not worth living.

And that makes it tempting to come on too strong, to say that no, *my* preferences are obvious and *yours* are crazy. An earlier draft of the post suggested that people who enjoy living in San Francisco might be “lizardmen”; I deleted it on advice from more sober-minded friends. I assume it would have just made the ~~lizardmen~~ San Franciscans angrier, and made them talk even more about how they couldn't stand living somewhere where they had to drive more. I get it. Everyone's preferences exist, competing access needs, etc.

If cities are your personal Hell, then even if on some intellectual level you know that other people can tolerate them, it becomes hard to fight for subjecting more and more people to hellish conditions. And if cities seem great to you, and suburbs boring and stupid, then you start thinking anyone defending their suburb against urban encroachment must just be classist or racist or something.

FosterBoondoggle writes:

> As far as I understand it, the gist of the complaint about YIMBYism here is that SA likes living in a lowish density

> suburb (Montclair? Piedmont?), thinks that’s his right, and fears the YIMBYs will try to upzone his neighborhood and fill it with condos and 5-story apartments. But if he’s been paying any attention to what YIMBYs (like, say, Brian Hanlon, head of CAYIMBY) are actually saying, it’s not that Montclair needs to be full of skyscrapers (it’s on a hillside with narrow streets, so that doesn’t make much sense). It’s that the parts of Oakland, SF, Berkeley, the peninsula, parts of the S. Bay that are accessible to people without a car (Rockridge, N. Berkeley, S. Berkeley near Shattuck, the area near the Apple spaceship in Cupertino) should be upzoned for greater density. Walkability is desirable to a lot of people.
>
> It brings network effects, like stable and successful local businesses, good restaurants – e.g., College Ave. in Rockridge – that are net positive for everyone. It brings walkable access to mass transit, which reduces climate impact. Another net positive for everyone. It brings easier access to wilderness like the east bay hills or west Marin, because dense communities mean less sprawl. (Which is why the CA Sierra Club’s opposition to most current development, as captured by Scott Lucas here is particularly ironic.)
>
> No one (I know, never say “no one”) is trying to upzone SA’s single family Oakland hills neighborhood.

Yes, a lot of the discussion of preferences above is completely irrelevant, because in the real world there will always be some suburbs and some cities (at least until the age of ecumenopolis).

When I say that YIMBYs are often right about their policy proposals but make me hate them anyway, that’s what I mean. There’s no reason these debates have to devolve into “suburbs and people who like them suck” vs. “cities and people who like them suck”, but they often do – and I admit I am personally guilty of reinforcing this.

(case in point: in accordance with the prophecy, someone definitely wants to upzone my single-family neighborhood)

From fluorocarbon:

> I didn’t know what to think going into this article, but I ended up being fascinated with it for anthropological reasons. Is San Francisco really that horrible? Programmers live three to a bedroom? People play music all night at BART stations?
>
> I would say that, though it’s an interesting post, it’s not really an accurate representation of the YIMBY movement outside of the Bay Area. When I think of the YIMBY movement, I think of organizations like Strong Towns. They don’t want giant towers, but rather fewer shopping malls and more pedestrian-centric development.
>
> I’ve also talked to some people in Boston on the YIMBY/pro-development side. The arguments I heard from them are:
>
> 1 – parking requirements are dumb
> 2 – more inner suburbs should zone for multi-family units (triple deckers)
> 3 – there should be more mixed used developments
> 4 – increased density should be allowed close to public transportation (MBTA) stations
> 5 – there’s an absurd amount of red tape when developing anything and it should be reduced

> These all seem reasonable to me and nothing on that list would destroy existing neighborhoods. But then again I find walkable multi-family neighborhoods (2-4 stories) with mixed use developments and narrow streets much more pleasant than either single family suburban car sprawl or Mega-City One huge Manhattan towers everywhere.

Okay, I vote that Team Gleaming Skyscrapers and Team Leafy Suburbs come together to burn the heretic.

Ana53294:

> The equivalent cities by population in Europe [to San Francisco] would be Valencia, Seville, Leeds, Glasgow, Stockholm, Cologne, Frankfurkt. And they don’t suck as much. Most of them are quite pleasant.
>
> You can have a greater density than San Francisco, a lower crime rate, a nice metro system, all while living in an apartment that is at a bikable distance to work and is much more affordable. With clean streets and no visible needles. For that, you need better sidewalks, good infrastructure, bike lanes, better policing and social policies, better public transportation, more parks (and close them at night). I lived in a city with the population of SF, and I never had anybody shout at me in the public transportation (although I usually biked). San Francisco is a high density city. Why isn’t it more bikable? that would reduce the strain on the public transport and the roads.
>
> So, in order to fix all those things NIMBYs complain about, you just need to fire the entire SF city council, and hire a foreign one fix those issues separately from the housing issue..

I agree with this. Most of the US’ problems with dense cities are solveable in principle. But I would still feel more comfortable if the order went first, solve the problems, second, tell everyone not to worry about them because of how solveable they are. Otherwise, I think people are justified in having a high prior that the problems won’t be solved, just as they haven’t been solved so far, and so higher-density cities will indeed keep having all of these problems that make them hard to live in.

Andrew on why prices might go down faster than the models I quote predict:

> I think that SF allowing moderately more housing would affect prices far more than what you suggest. Prices follow supply and demand, but prices also build in future expectations about supply and demand. Right now, SF works very hard to forbid construction. Therefore, you get lots of investors buying property. SF’s stance on denying construction affects demand as well as supply. If SF started allowing enough new construction that housing prices would stop going up, many of those outside investors would take their money elsewhere.

Atlas against worrying too much about agglomeration effects:

> Increasing density in SF could make things worse, because…it will increase the economic benefits of living and working in SF? Enrico Moretti’s book the New Geography of Jobs did a good job explaining that there are a lot of positive effects that come from people living and working next to each other in big cities.

Gwern adds:

> Yeah, that was my problem with #3's summary. 'It might not lower housing prices on net, all it might do is fail by creating billions or trillions of dollars in compounding new wealth through greater economic efficiency in a vital technological hub.' Oh, is that all? Sounds like a pretty good way to fail.

DS against some of the statistics:

> Unfortunately, the post's numbers are based on a badly unrepresentative sample. 2008-2015 is right after the housing bust. This sample looks at housing construction at the single lowest period it's ever been in modern American history.
>
> The post tries to generalize from 2008-2015 to answer "how fast could we build." But that's like measuring unemployment in 2009-2011 or 1930-1934, and using it to answer how many people could have a job!
>
> Even in Texas and California, they used to build a lot more apartment buildings than they do. The peak of American construction seems to have been 1965-1985. New construction permits in that period in (e.g.) California appears to have been about 3-4 times the rate of today. That's more than fast enough to meet the standard set in your post.
>
> Meanwhile, New York City has been seriously limiting new construction since about 1960, which makes citing Manhattan prices as evidence that allowing construction doesn't work exactly backwards. (Fun fact: the NYT did an article showing that 40% of Manhattan's already-existing buildings would be illegal to build today.)
>
> If you want dense with lots of current construction, try Seoul, South Korea. There, rent-to-income is about 35%

> of average national income.
>
> That’s equivalent to an SF Bay Area with a median rent of $2,000 a month (~35% of average California income), as against current typical rent of about $3,000/mo (~50% of average California income).
>
> So if the Bay Area were as pro-housing as Seoul, over time you might cut housing prices by a third, and get a lot more people in.

Douglas Knight, on how growing cities doesn’t necessarily have to be disruptive for people who don’t like density:

> Lots of people in the comments are equivocating or talking nonsense because they refuse to talk about details.
>
> 3% growth, doubling in a generation, sounds pretty reasonable to me. If you know it’s coming, then you shouldn’t buy the density you want, but a little less dense. And it is the plan. The State of California requires towns to build at a certain rate and most of them are just cheating. One of these days the State will take back the power of zoning and catch-up building will be disruptive, in part because it will be done at such a distance.
>
> Whether 3% growth is disruptive depends on the details. If it’s smooth in place and time, it sounds pretty reasonable to me. Consider a town of 1 acre lots. Every year, tear down 1% of the houses and build 4 new houses on 1/4 acre lots. The character of the town will change, but smoothly. This shouldn’t disrupt the community, no more than the normal turnover, which is a lot higher than 1%.
> Alternately, one could tear down 1/3000 of the houses

> and build apartment buildings for 100 people each. This would be much more socially disruptive. It would create more economic diversity. It would cluster newcomers, making them less integrated into the new community. This is the kind of thing the State will do, if it acts on its own. On the other hand, if the apartments are near train stations, it might not be as bad for traffic.

TheNybbler on whether cities have to be where the jobs are:

> When cities were hollowed-out enclaves of crime and poverty, many employers set up in or moved out to suburban office parks and had plenty of productivity. Some even did that before the fall of the cities, Bell Labs being a NYC area example. Now cities are fashionable again and the suburban office parks are empty (except in Silicon Valley). But it doesn't have to be that way.

wulfrickson is skeptical that agglomeration effects are really that bad:

> The Zuegel piece that Scott linked in point 3 (arguing that agglomeration effects may mean that more housing => higher prices) was discussed pretty widely when it came out, and the consensus was that things could come out that way in theory, but the empirical evidence points in the other direction: agglomeration effects are probably not big enough. Here's a bit of Twitter discussion. Another paper by French economists estimates that increasing a region's population by 10 percent would increase costs of living by 0.3 to 0.8 percent, once housing supply adjusts to compensate (in the short run, it's more like 1 to 3 percent). This isn't anything to write home about.

peopleneedaplacetogo takes a wider perspective:

> I expect my friends to keep moving to the Bay regardless of what happens with housing policy (since their employers can generally pay enough to make it worthwhile no matter how high rent is). But more broadly I think this kind of regional economic inequality is actually exacerbated by NIMBYism; for most of their history per capita incomes in rich US regions (like California) and per capita incomes in poor US regions (like Kentucky) were converging, but this trend stopped in the late 1960s right around when zoning became widespread, and subsequently reversed. Caps on production of housing near jobs made it hard for workers from Kentucky to fill labor shortages in California, but also weakened the bargaining power of the workers still in Kentucky who could no longer so credibly threaten to move to California. Everyone being free to move where they want can help those who don't move too.

grendelkhan says that one reason California is especially bad is Prop 13:

> Normally, these things scale reasonably: if you build more houses, the occupants pay property tax, and that pays for the fire and police service, the roads, power lines, water service, all that.
>
> Prop 13 in California turns this on its head. Housing, even when it appreciates in value, doesn't pay more in taxes. In fact, the real property-tax rate goes down due to the 2% annual cap on increases, which is generally below inflation.
>
> California has an unusually high cost of living, largely driven by high housing costs. (Everything here is interconnected.) So infrastructure costs more there,

> because labor costs more. To raise the money for this, cities and counties issue bonds and raise sales taxes, and most of all, "impact fees", where a new development pays tens of thousands of dollars to make up for the depressed property taxes of their neighbors.
>
> And because commercial developments change hands more often (property taxes reset on sale), and commercial developers can afford larger impact fees, cities fearing a pension crunch approve more commercial than residential real estate; the incentives are such that every town wants someone else to be the bedroom community. So we add more jobs than places for workers to live, as well.
>
> The pension crisis is very real, and cities are trying to make up for a broken revenue model. Infrastructure is pretty much off the table.

And from eternaltraveler:

> I'm in SF bay and have been for the last 9 years because this is the best place in the world to find venture capital for anti aging biotechnology. The entire Bay is otherwise almost entirely horrible, but the alternative is certain death.

Uh, thanks for putting things in perspective, I guess?

Next Door In Nodrumia

[Content note: attempt to consider real people's real problems using angel-on-pinhead impractical reasoning and ideas]

I.

Imagine the state of nature, except for some reason there are cities. Some people in these cities play the drums all night and keep everyone else awake. The sleep-deprived people get together and agree this is unacceptable. They embark on a long journey to the wilderness where they found their own community of Nodrumia.

They form a company, the Nodrumia Corporation, which owns all the property in the area. The corporation distributes usage rights via a legal instrument that looks suspiciously like private property: people who own usage rights keep them forever, can do whatever they want with the land, and can freely transfer and sell them to others. The only difference is that the usage rights have a big asterisk on them saying "contract is null and void if you break the rules of the Nodrumia Corporation". These rules are set by a board chosen democratically by the inhabitants, and are all things like "You can't play drums at night", and "You can't sell property to people who will play the drums at night", and "Anyone who plays the drums at night shall be exiled".

One day a Nodrumian wants to move out, so he puts his house up for sale. The highest bidder is a drummer who wants to use the property as a studio so he can play the drums at night. The Corporation steps in and bans the sale. The property owner protests, saying that he is being oppressed.

According to libertarian philosophy, who is in the right?

The argument against the drummer: the land is basically the private property of the Nodrumia Corporation, and libertarians believe that private landowners should be able to determine what happens on their property. And more fundamentally, the people there have a strong preference against living near drummers, and that preference seems fundamentally satisfiable if their property rights are respected, and it seems stupid to legislate a world where people are forever forbidden from satisfying a fundamentally satisfiable preference and have to be unhappy all the time.

The argument in favor of the drummer: this is basically just a town. “People who live together in a community, and are governed laws made by a democratically elected council” is a town. It seems sort of unlike a town because of its strange history, but really in America a lot of towns were formed by people leaving society, finding unoccupied (or “unoccupied”) land, and building a community there. Some of them were even formed with some sort of utopian goal in mind, or specifically to escape things the inhabitants didn’t like about the places they came from. The only real difference between Nodrumia and the average town is its odd property right structure, but this is a difference in name only: everyone who moves into *any* town knows that they own their property only insofar as the things they do on their property don’t conflict with town bylaws. Seriously, Nodrumia is just a town. And the whole point of libertarians is that they are skeptical that governments (including town governments) should be able to ban people from doing things. Therefore, the drummer should be allowed to open his drum studio.

A second argument against: imagine we’re talking about a private company like Microsoft. Libertarians agree that Microsoft has the right to decide that none of their employees can play the drums at their cubicle in a way that disturbs other employees. This suggests that playing the drums isn’t a fundamental right. But it’s unclear how Microsoft is different from the Nodrumia Corporation.

A second argument in favor: maybe this isn’t just a town. The way it’s presented, it sounds like more of a city-state. Its government is a national government. If we’re saying national governments can make laws banning musical instruments, we’ve gotten very far from libertarianism, haven’t we?

We could resolve the conclusion by saying that libertarianism is wrong, actually externalities are bad, and it’s totally okay to ban them. Or we could use an expanded idea of property rights that included that right not to have noise on your property (though this opens up a big can of worms). But if we wanted to at least keep some claim to be working within a strict libertarian paradigm, I think we would have to make an argument based on what kind of characteristics an institution needs to be more like a corporation or intentional community (which have the rights to be strict) vs. a national government (which should be erring on the side of permissiveness). To me, the key differences seem to be things like:

– exit rights and transaction costs of leaving
– number of other options
– ease of forming a new one
– degree to which membership is voluntary vs. hereditary

So to give an example, most people have the intuition that the US government banning pork for religious reasons is bad, but also that if you go into a mosque and demand they let you eat pork there you're in the wrong. I think this is because:

– the people in the mosque have the option to very easily not be in the mosque
– if you don't like the mosque, you can always go to a church or an atheist meetup
– you can always start your own mosque, with blackjack and hookers
– most people in the mosque chose to be there because they agree with the mosque's principles

But:

– it's hard to leave the US if you don't like it
– there aren't that many other countries and you might not be able to find one you like
– it's very hard to start a new country
– most US citizens are only citizens because they were born here, and didn't necessarily sign on to any philosophical commitments

This is ignoring some important issues, like whether banning pork is the ethically correct action, or whether the majority of the people in each community support the ban. It's just trying to give a completely formal, meta-level account of why our intuitions might be different for these two cases.

This seems to justify the libertarian intuition that we shouldn't be bossing private companies around. It also justifies the much more common intuition that we *can* boss private companies around when they're monopolies or otherwise seem hard to get rid of, like people discussing whether the government should make Facebook have better privacy policies. If there were hundreds of equally-sized alternatives to Facebook that people could easily switch to, with a wide variety of privacy policies, it probably wouldn't come up as often.

Towns seem kind of midway between companies/mosques and national governments. They're not *easy* to leave, but realistically people leave towns all the time; I've switched cities maybe six or seven times during my life. There are thousands of towns you can live in, including dozens of big cities. Forming a new town isn't *easy*, but there's lots of open land where you could do it in theory if you wanted to; it's not really beyond the ability of even a dedicated private citizen. And about two-thirds of people no longer live in the town where they were born.

I think a libertarian treatment of this issues would argue that towns have the most right to pass restrictive laws when things like exit rights are most salient, and less right when they aren't.

The drummer moving into Nodrumia seems like a clear case where exit rights are really salient. The drummer isn't even in Nodrumia yet, so clearly he has the ability not to be in Nodrumia if he wants. The transaction costs of moving to not-Nodrumia are zero, since he's not even in the town yet. It seems like starting a new town is easy, since the Nodrumians themselves managed it. And although the story doesn't give a time course, it seems plausible that most of the people involved are still first-generation migrants to Nodrumia – and the drummer definitely is.

The harder case would be one where, by natural population turnover, the first generation of Nodrumians has children, about half of the second generation want to play drums, and in the interim all the available land has been settled by other towns that ban drums, and there are no pro-drum towns to move to. I don't know what I think in this case, although I'm tempted to say that if there are thousands of towns but none of them permit drumming, that's kind of like there being thousands of companies but none of them will pay you $500 an hour – you're asking for something nobody else wants and you should reconsider your request (though given numbers like this, it should be possible for the pro-drum faction to get at least one town for themselves, even if they have to buy out the existing inhabitants).

In the recent discussion of NIMBYism, YIMBY partisans keep saying things like "it's illegal to build high-density cities!". This confuses me, because I don't think it's illegal for a private citizen to build a high-density city, assuming she can afford enough unincorporated land and the construction costs [EDIT: maybe not]. And it's not illegal for a town to change its urban policy to become a high-density city, assuming it wanted to. This seems kind of like saying "It's illegal to have a community made entirely of log cabins". You can totally get some people together and found a log cabin community, you're just not allowed to force existing towns to switch to all-log-cabins unless the citizens want to. I think the reason this argument seems unconvincing to me but convincing to them is that I'm reasoning from a perspective where communities are a basic unit, whereas the YIMBYs are reasoning from a perspective where individuals are the basic unit, eg "It's illegal for me to sell my house to a high-rise developer".

And I'm reasoning from a perspective where communities are a basic unit because I believe in Archipelago, a world where the only win-win solution to our many differences about what societies should look like is to let people form highly-varying communities with exit rights and let people live in whichever one they want. This solution depends on Nodrumia's right to kick out drummers and it depends on viewing towns as being basically a form of private property owned cooperatively by the town members.

If you try to take someone else’s private property because it’s standing in the way of economic progress, that’s eminent domain. I’m not 100% against eminent domain all of the time, but it should be a very last resort. This is why I find NIMBYs’ objection of “we should be allowed to decide what happens to our town” so sympathetic. They’re analogous to Nodrumia’s right to not allow drum studios, and without that kind of private-property-analogous right I’m skeptical that anywhere can provide the good life to its citizens. Without letting towns be at least kind of like private property, they all converge onto the highest-entropy state permitted by the wider country they live in (eg Las Vegas but more so). If you want a libertarian national government but also accept that some people want to live in places other than Super-Vegas, you need to let towns pump against entropy and retain some distinction from each other, the same way we let individual citizens arrange their own lives the way they want. That’s part of why I find myself more sympathetic to local governments than to national governments.

II.

This model also suggests a solution for YIMBYs – start their own town somewhere.

This might be harder than it sounds, because if the YIMBYs aren’t very committed, once they have a high-density walkable city that they’re happy with, they might lose their will and be tempted to keep other people out to prevent it from becoming more crowded. Even if they were very principled, the next generation of inhabitants might not be.

The solution is charter cities. Some profit-seeking individual or corporation could buy some land, explicitly note that they weren’t making it a full democracy, and then try to turn it into the biggest, most economically productive city possible so they could skim a little bit off the top.

If the California state government is really concerned about the housing shortage, but also doesn’t want to densify San Francisco, here is what it can do. Encourage some company to buy a promising but currently empty tract of land (I’m saying “some company”, but we all know it would end up being Peter Thiel). Give them various concessions to lure people in, like that people living there only have to pay half as much in state taxes (or, if they really want to start a land rush, they can exempt the area from the plastic straw ban). The company has strong incentives both to make the city as populous and dense as possible, and to make it the sort of place where rich people want to live and businesses want to operate. Tech companies, social-climbers, and the like move there instead of San Francisco. The San Franciscan NIMBYs are happy, the tech companies are happy, the company that owns the land is happy, and the California state government is happy. The end.

This isn’t so outlandish – I grew up in this city. In 1864, an investor named James Irvine bought a big tract of California land. Over the next century, his heirs formed a group called The Irvine Company to develop it further. They got their big break in 1959, when James’ grandson Myford Irvine cut a deal with the University of California to build a college on the still mostly-empty land, virtually guaranteeing it would grow into a town. The Company planned out their ideal urban utopia, raised some money, and built it according to plan. Now Irvine is the 16th largest city in California, and Irvine Company head Donald Bren has $16.3 billion and is the 80th richest person in the US. Irvine consistently tops various “best city” and “highest quality of life” rankings and manages to balance some density (the listed density of 4,000 is probably an underestimate because of the deliberately preserved wilderness areas; other parts are much denser including a few 20-story buildings) with a very safe, suburban feel. It’s also very good at attracting tech companies: Blizzard, Broadcom, Allergan, and the US headquarters of Samsung, Sega and Toshiba are all located there. It’s also an outlier in new housing construction, growing its housing stock at (informal estimate) 5% per year – twice the rate of Austin, three times that of Seattle, and five to ten times that of San Francisco.

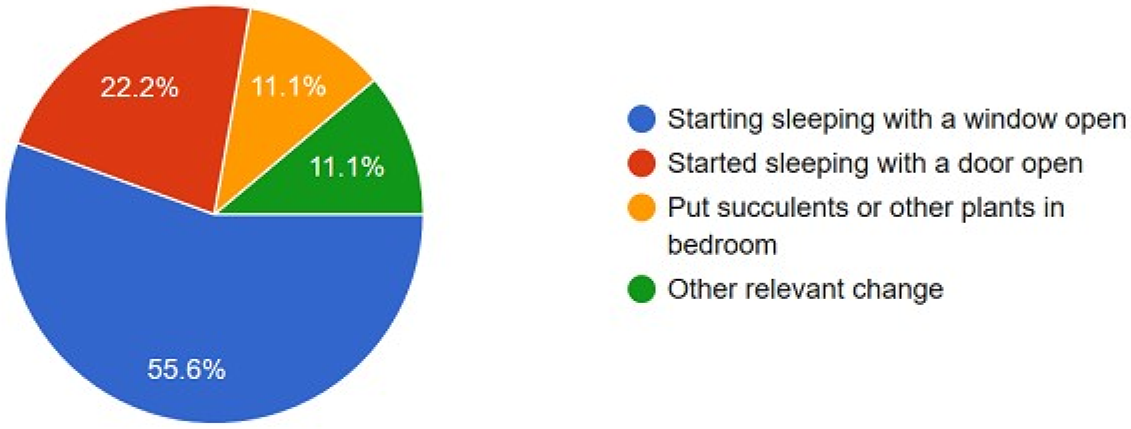
I know this probably “won’t” “work” “in” “real” “life”, just like everything good or interesting or creative. But a state policy of deliberately creating super-Irvines in suitable areas would relieve the need to develop anywhere else. It would slice through concerns that it’s politically impossible to upzone existing cities, concerns about congestion and crime and transit inadequacy in existing cities, concerns about disrupting the culture of existing cities, and concerns about existing cities’ poor business climate and poor reception of out-of-staters. It would be a good way to attract all of the pro-density pro-walkability people to one place so that they weren’t scattered among a bunch of people who wanted lower-density towns. People could make it sustainable and renewable and otherwise buzzwordable. We could finally say, in all honesty, that America had caught up to Senegal.

Right now there’s a small movement for charter cities, but it’s usually considered the sort of thing that will only happen in the Third World. But California already has some legal provisions for a very weak form of charter city, and some parts of California are already getting kind of Third-World-ish. I don’t know whether it’s possible. But it doesn’t seem obviously harder than getting San Francisco residents to agree to add new housing at the rates that would be necessary to make a dent in the crisis.

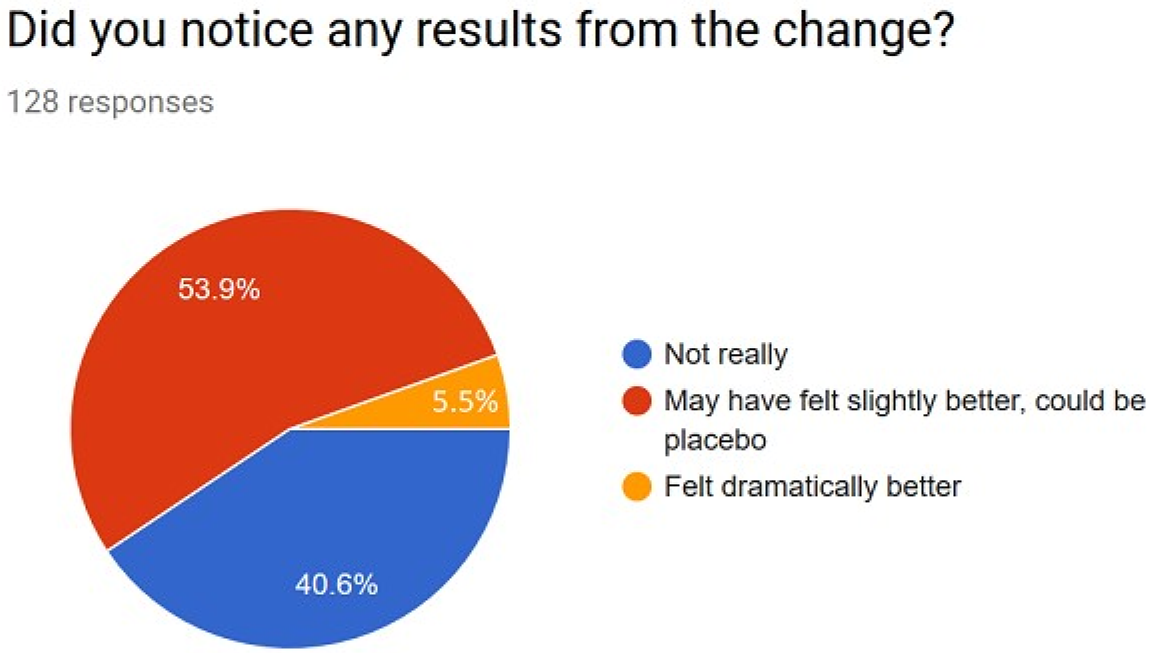
Nighttime Ventilation Survey Results

Thanks to the 129 people who tried altering their nighttime carbon dioxide levels after [my post on this](), and who reported back to me. There was no difference between people who pre-registered for the study and people who didn't, on any variable, so I ignored pre-registration.

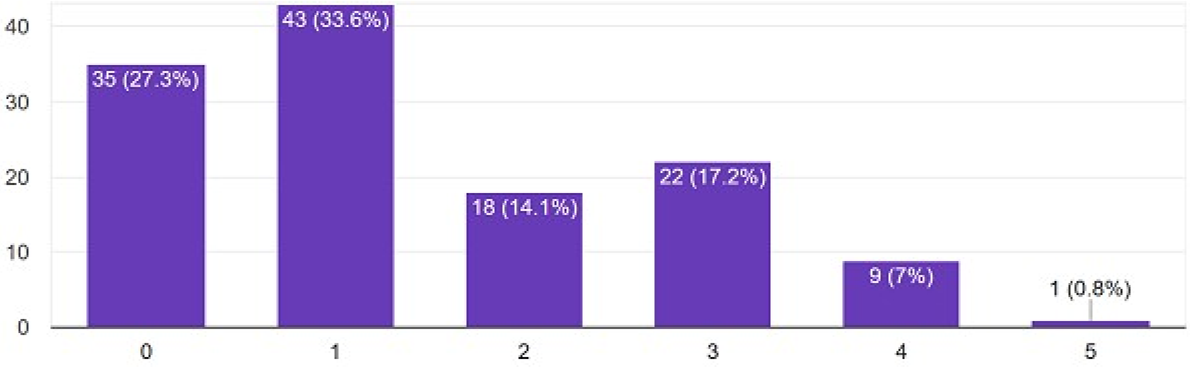
126 people reported one intervention they performed. The most common was sleeping with a window open:



People generally reported slight but positive changes:



When asked to rate the magnitude of improvement to well-being on a 0 to 5 scale, they averaged 1.4:



I mentioned in the post that succulents could help in theory, but you needed to get the right kind of succulents and you needed at least ten of them. I was skeptical that anyone really got ten succulents in their room, so I wondered whether that might work as a crypto-placebo group.

If so, the intervention failed to separate from placebo. Succulent users had an average improvement of 1.29, compared to about 1.50 for people who did other things. The difference wasn’t significant, although admittedly the sample size was low.

Looking at the various groups, the most striking difference was actually people who left a window open (1.57) vs. people who did one of the other named options (1.31). A few people who left windows open mentioned this made their room cooler, which seemed to help with sleep. But this is very post hoc, and this difference wasn’t significant either.

Here are some reports from people who described dramatic improvement:

> Less headaches, less fatigue in the morning, less trouble staying asleep through the night

> I slept more comfortably, woke less in the night, and felt less fatigue in the morning. I also felt more alert during the day. I used a CO2 meter. Peak level changed from 1400s before to 800s after the intervention.

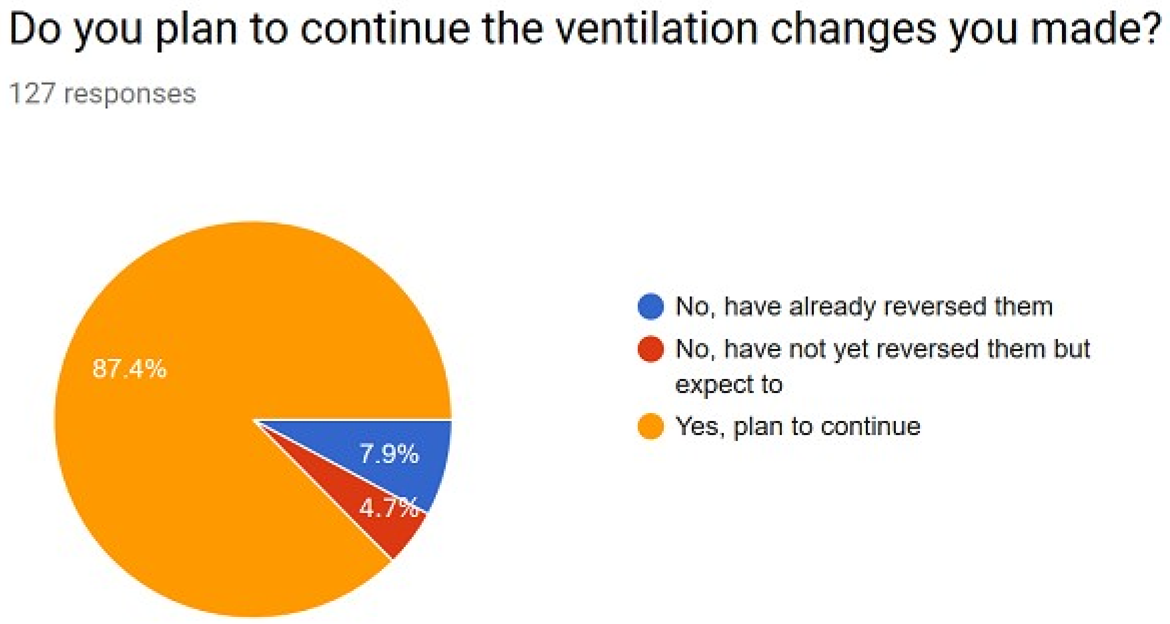
> When I get up I feel less groggy, It takes my brain less time to get on line, I can way more often wake up and do things right away rather than spend time in a stunned haze or distracting myself so I don’t fall back asleep. I feel more like doing things and have the energy to back that up, it lasts some time after I wake up but usually not all day. I feel like I sleep better.

And here are some of the more typical results from people who said they felt only minor or placebo-like improvements:

> Possibly more alert upon waking, possibly needing slightly less time sleeping to feel rested

> Slightly more likely to sleep through the night and/or feel better rested in the morning.
>
> I don’t recall waking up as often, but maybe it’s a placebo. If it’s a placebo, it’s a cheap/free one and I’m happy to keep taking it as long as it works. Oh god is taking this survey going to make it not work anymore?

Despite the underwhelming results, most people were going to stick with their intervention:



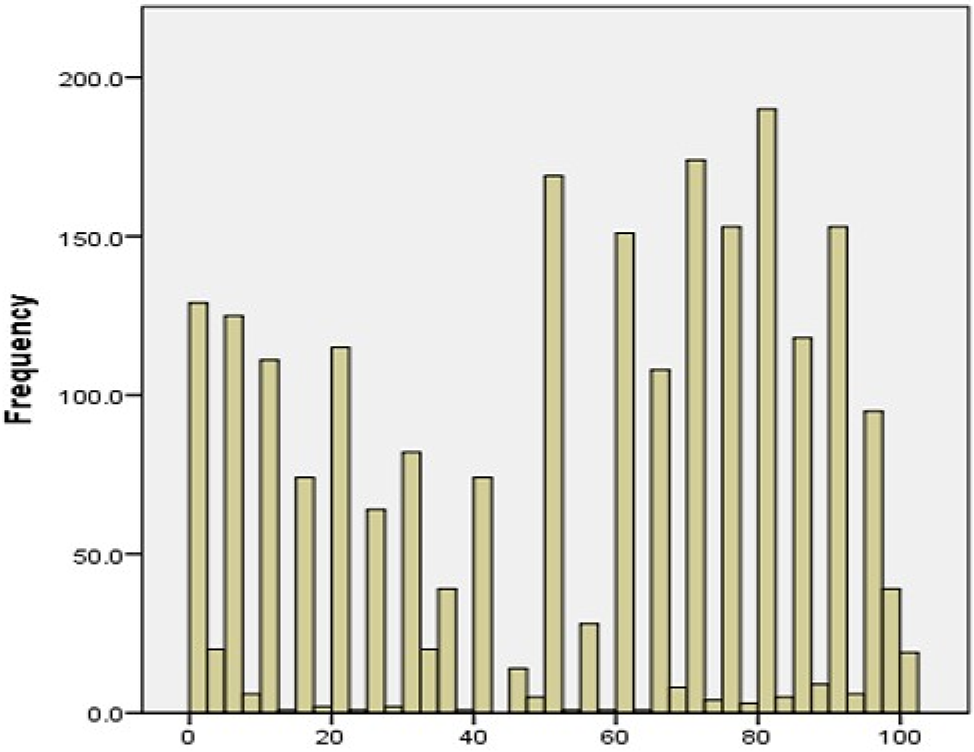
I consider these results basically negative – both for nighttime ventilation, and for the ability of informal blog surveys to give data that one can be confident in either way. But I’m glad some people think they feel better, and the results of the last question suggest it still might be a cheap and productive thing to try.

If you’re interested in analyzing this further, you can download the data at [this link].

Kavanaugh: A Probability Poll

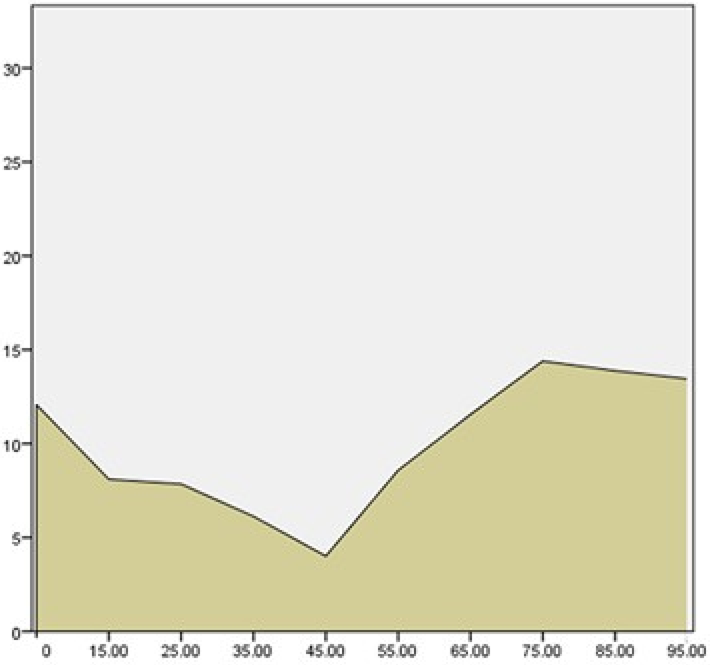
There's some literature suggesting that people are more careful when they think in probabilities. If you ask them for a definite answer, they might give it and sound very confident, but if you encourage them to think probabilistically they might admit there's more uncertainty.

I wanted to look into this in the context of the recent Supreme Court confirmation hearings, so I asked readers to estimate their probability that Judge Kavanaugh was guilty of sexually assaulting Dr. Ford. I got 2,350 responses (thank you, you are great). Here was the overall distribution of probabilities. Horizontal axis is percent chance he did it; vertical axis is number of people who responded with that percent:



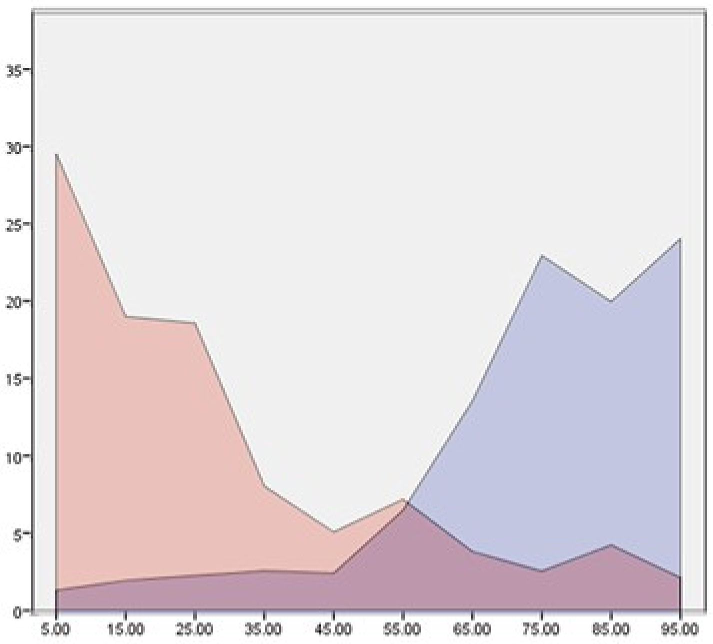
This looks weird because people were most likely to give numbers rounded off the the nearest ten.

I separated responses into bins from 0 – 9%, 10 – 19%, and so on to 90 – 100%. Keep in mind that the last bin is slightly larger than the others, so it might make it unfairly look like more people gave extreme high answers than extreme low answers. I also switched the vertical axis to percent of responses in each bin. Smoothed out, it looks like this:

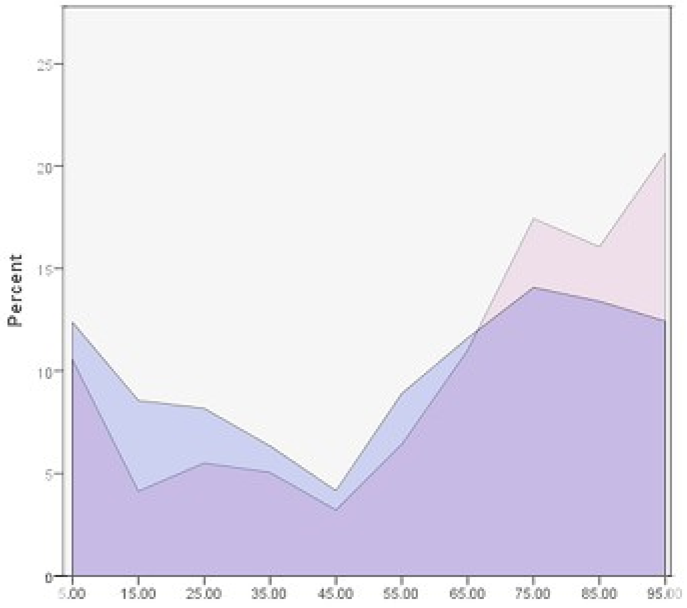


This looks pretty balanced, and it is: the average probability is 52.64%. This is probably a fake balance based on all the different demographic skews involved cancelling out: 2.5x as many Democrats as Republicans answered the survey, but 9x as many men as women did.

Here are the results broken down by party (blue is Democrats, red is Republicans):

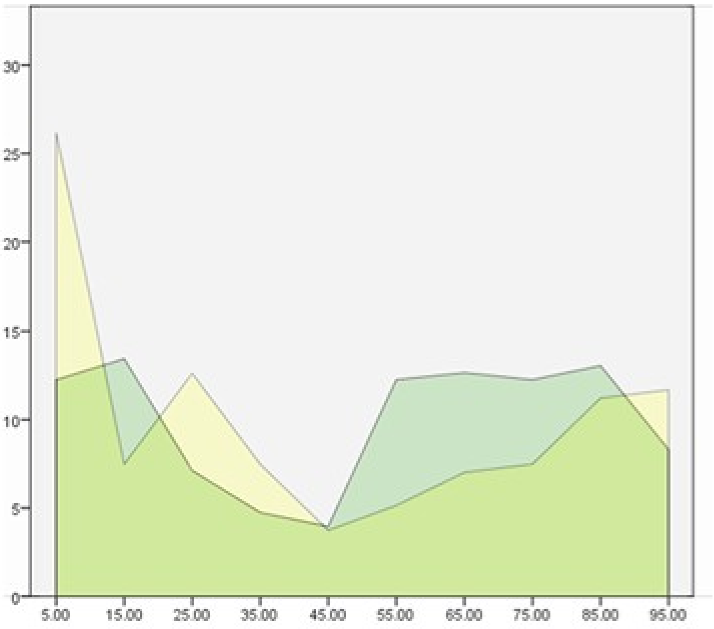


And here are the results broken down by gender (blue is men, pink is women):



There should be an interaction between party and gender, because men are more likely to be Republicans, and women Democrats. I didn’t have enough data to investigate this too carefully, so I can’t say whether gender controlled for party remains significant or not.

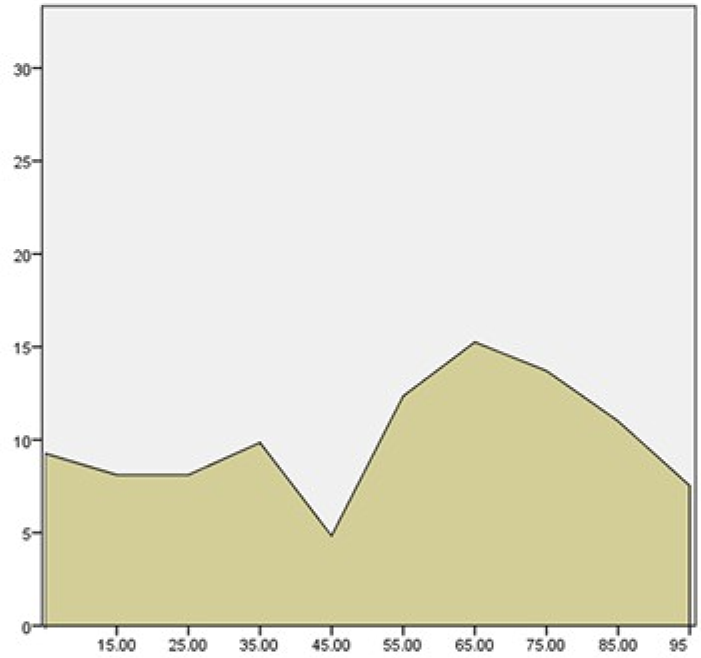
I asked two questions to assess participants’ level of background knowledge: where did Kavanaugh go to law school? (correct answer: Yale) and what is Kavanaugh’s wife’s name? (correct answer: Ashley, but a shout-out to everyone who wrote “Mrs. Kavanaugh” and to the one person who wrote “beer”). Here are the probabilities of people who got both questions right (gold) vs. both questions wrong (green):



People with high background knowledge were more extreme in their answers, and slightly more likely to think Kavanaugh is innocent. I worry that I made a mistake in the questions I chose, since people who are more sympathetic to Kavanaugh might be more likely to know about his family. In retrospect, I should have asked at least one question about Dr. Ford.

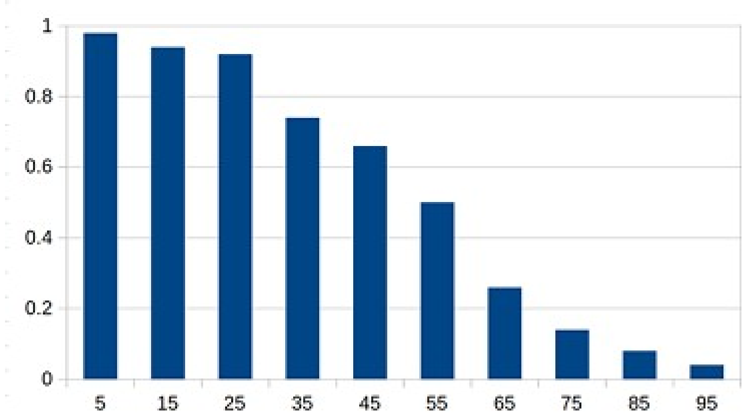
What about neutral people? Do such people exist? I looked at people who were neither registered Democrats nor

Republicans, and who rated their liberal-vs-conservative ideology, on a one to ten scale, as 4, 5, or 6. These people looked like this:



This chart makes it look like they're slightly leaning towards guilt, but I think that might be a function of the binning; their mean probability of guilt was 49.85%, about as close to "totally uncertain" as you can get. Neutral people with more background knowledge were, again, more likely to lean innocent, with a mean of 41%.

I asked people whether they felt the evidence that Kavanaugh may have committed sexual assault was sufficient to reject his nomination to the Supreme Court, regardless of any other reasons to vote for or against him (like his legal opinions). 55% of respondents thought that yes, his nomination should be voted down; 45% still supported him. Here is a list of support for confirmation by probability of guilt:

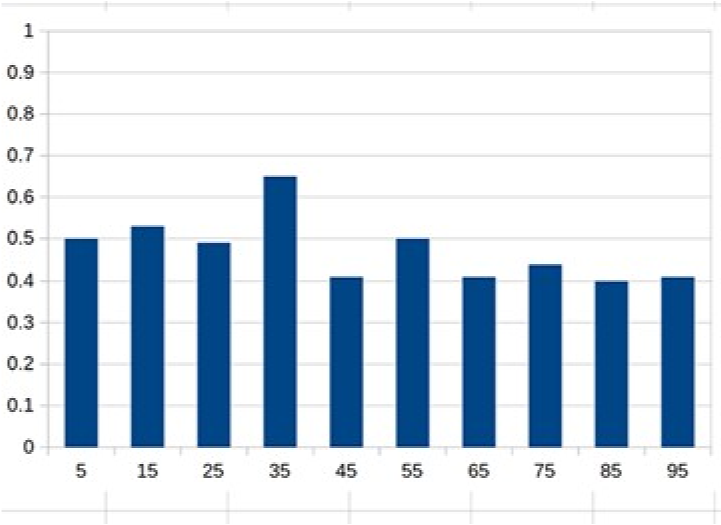


Of people who thought there was only a 0 – 9% chance Kavanaugh was guilty, 98% thought he shouldn't be rejected from the Supreme Court on this basis alone. Of people who thought there was a 90-100% chance of guilt, 96% thought that was sufficient to reject the nomination. Of people who thought there was 50-50 chance of guilt, about 50% still supported him and 50% opposed him.

This question suggests there is no real consensus about how plausible an accusation has to be before it means someone should be denied nomination to the Supreme Court. People generally agreed that if there was below a 25% chance the accusations were true, he should definitely be confirmed, and if there was above an 80% chance, he definitely shouldn't be. But between 25% and 80%, people were pretty split on whether the Senate should err on the side of not confirming a potential assaulter, or wait until it was beyond a reasonable doubt. If we were trying to make these answers into a guideline for how a Senator should vote, it looks like they would be satisfying the most people if they voted to confirm if they thought the accusations had a less than 50% chance of being true, and to reject if they thought they had a more than 50% chance. I wonder how many people would endorse this rule as written.

I also asked people whether they would reject Kavanaugh in the hypothetical universe where he had immediately admitted to the accusations, then apologized for his actions and said he had changed as a person since then. About 55% of people said they would accept him in this scenario, meaning he gains about 10% support in the SSC demographic compared to the real-world situation. But the question was poorly worded and I'm not sure how many people answered yes they would reject him, accidentally meaning to say yes they would accept him.

Here is a graph of how people answered this hypothetical compared to how guilty they thought he was:



There's no correlation. This makes sense: how guilty you think he is in this universe shouldn't affect your opinions about a hypothetical universe where you know he's guilty – but for some reason I'm still surprised. I guess I expected people's partisan biases to sneak in, even if they didn't make sense. Maybe the question was so confusing that answers to it are basically random.

Overall, when asked to use probabilities people were able to admit to a little bit of uncertainty in their answer. They could give probabilities that were well-formed and self-consistent. But none of this came close to removing the partisan bias and the strong difference in opinions. There is no consensus in the general SSC demographic, and even unbiased people as a group are unable to send a coherent signal. This is not a good way to get beyond confusion and disagreement on an issue like this one.

You can download the raw data (slightly cleaned up) here.

Anxiety Sampler Kits

The best thing about personalized medicine is that it’s obviously right. The worst thing is we mostly have no idea how to do it. We know that different people respond to different treatments. But outside a few special cases like cancer, we don’t know how to predict which treatment will work for which person. Some psychiatric researchers claim they can do this at a high level; I think they’re wrong. For most treatments and most conditions, there’s no way to figure out whether a given sometimes-effective treatment will work on a given individual besides trying it and seeing.

This suggests that some chronic conditions might do best with a model centered around a controlled process of guess-and-check. When it’s safe and possible, we should be maximizing throughput – finding out how to test as many medications as we can in the short time before we exhaust our patients’ patience, and how to best assess the effects of each. The process of treating each individual should mirror the process of medicine in general, balancing the need to run controlled trials and gather more evidence with the need to move quickly.

I don’t know how seriously to take this idea, but I would like to try it.

Some friends and I made thirty of these Anxiety Sampler Kits, containing six common supplements with some level of scientific and anecdotal evidence for treating anxiety (thanks to Patreon donors for helping fund this). The 21 boxes include three nonconsecutive boxes of each supplement, plus three boxes of placebos. They're randomly arranged and designed so that you can't tell which ones are which – I even put some of the supplements into different colored capsules, so you can't even be sure that two capsules that look different aren't the same thing.

Each box contains enough supplement for one dose, and all supplements are supposed to work within an hour or so. Whenever you feel anxious, you try the first non-empty box remaining. Afterwards, you rate how you felt on the attached log (not pictured). When you've finished all twenty-one boxes, you fill out a form (link is on the attached paperwork) and figure out whether there was any supplement you consistently rated higher than the others, or whether any of them were better than placebo. If your three highest ratings all went to boxes which turned out to contain the same supplement, and it did much better than placebo, then you have a strong argument that this is the best anti-anxiety supplement for you.

(this setup isn't quite as irresponsible as it sounds. The six supplements I'm using are all considered very safe. I'm not concealing which six supplements are in it – it's magnesium, 5-HTP, GABA, Zembrin, lemon balm, and l-theanine – so you can check if you have allergies to any of them. And there's a spoilers page available if you have a bad reaction and need to tell your doctor what caused it)

Also on the form is a link to send me your data, which I'm asking you to do as a condition for using the kits. I'll add everything up and this will double as an n = 30 placebo-controlled trial of six different supplements. I don't think n = 30 is enough to impress anybody, but it might be enough to get some informal hunches about what works and be able to give people better advice. And if the experiment goes well, I can always make more kits.

~~If you live in the Bay Area, have enough anxiety that you expect to use a sample at least two days a week, and are okay with self-experimentation, these kits might be for you. Starting tonight I'm leaving a box full of them at the Rationality & Effective Altruism Community Hub, on the ground floor of 3045 Shattuck, Berkeley. REACH is usually open (or contains people who will open it if you knock) at all reasonable hours, and the caretaker there is aware that people might be coming in to get these kits. If you notice the box is out of kits, please comment here telling me so and I'll add an update so people don't waste their time.~~ [**EDIT:** All out of kits, sorry! Once I have gotten results I might make a new batch.]

Remember that by taking a kit, you’re saying you expect to have anxiety that you’d be willing to experiment on at least twice a week (it’s okay if it doesn’t work out this way exactly) and you’re committing to – if you’re able to finish the test – sending me a form with your results. People who are pregnant or nursing, who have relevant preexisting medical conditions, or who are already taking potentially-interacting medications should talk to their doctor before trying these kits. I will not give you medical advice about whether these kits are safe for your specific situation, so please don’t ask. If you would be comfortable taking a random supplement you got off the shelf at Whole Foods, you should feel comfortable with everything in here.

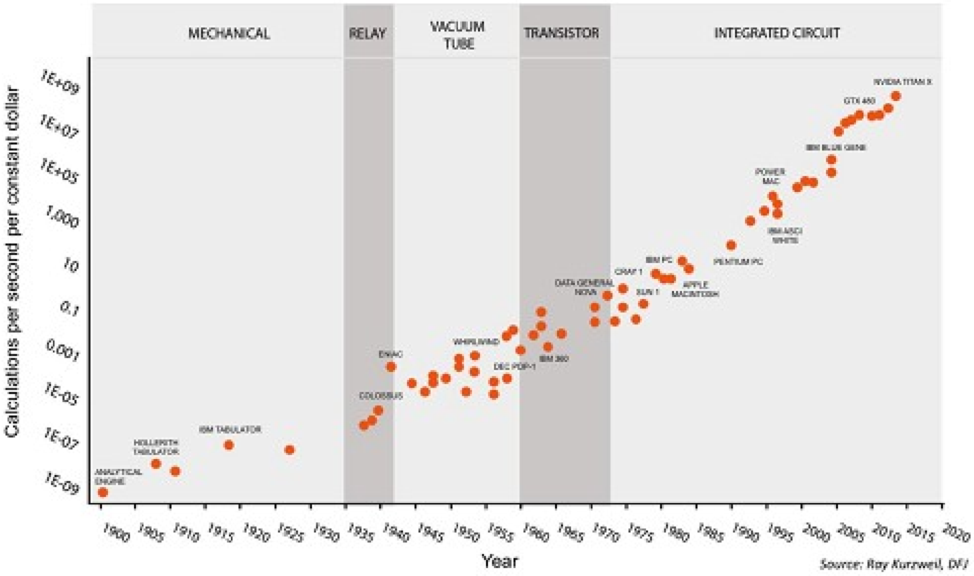
I might take this idea further, but I’m going to wait until the first set of results come in. If you are interested in taking this idea further, send me an email and let me know your thoughts.

Is Science Slowing Down?

[I wrote this after attending a conference six months ago. At the time I was confused about whether and when I was allowed to publish it. I am still confused, but less optimistic about that confusion getting resolved, so I will go ahead and beg forgiveness if I am guessing wrongly. All references to “recently” are as of six months ago.]

Is scientific progress slowing down? I recently got a chance to attend a conference on this topic, centered around a paper by Bloom, Jones, Reenen & Webb (2018).

BJRW identify areas where technological progress is easy to measure – for example, the number of transistors on a chip. They measure the rate of progress over the past century or so, and the number of researchers in the field over the same period. For example, here’s the transistor data:



Source: Ray Kurzweil, DFJ

This is the standard presentation of Moore’s Law – the number of transistors you can fit on a chip doubles about every two years (eg grows by 35% per year). This is usually presented as an amazing example of modern science getting things right, and no wonder – it means you can go from a few thousand transistors per chip in 1971 to many million today, with the corresponding increase in computing power.

But BJRW have a pessimistic take. There are eighteen times more people involved in transistor-related research today than in 1971. So if in 1971 it took 1000 scientists to increase transistor density 35% per year, today it takes 18,000 scientists to do the same task. So apparently the average transistor scientist is eighteen times less productive today than fifty years ago. That should be surprising and scary.

But isn’t it unfair to compare percent increase in transistors with absolute increase in transistor scientists? That is, a graph comparing absolute number of transistors per chip vs. absolute number of transistor scientists would show two similar exponential trends. Or a graph comparing percent change in transistors per year vs. percent change in number of transistor scientists per year would show two similar linear trends. Either way, there would be no problem and productivity would appear constant since 1971. Isn’t that a better way to do things?

A lot of people asked paper author Michael Webb this at the conference, and his answer was no. He thinks that intuitively, each “discovery” should decrease transistor size by a certain amount. For example, if you discover a new material that allows transistors to be 5% smaller along one dimension, then you can fit 5% more transistors on your chip whether there were a hundred there before or a million. Since the relevant factor is discoveries per researcher, and each discovery is represented as a percent change in transistor size, it makes sense to compare percent change in transistor size with absolute number of researchers.

Anyway, most other measurable fields show the same pattern of constant progress in the face of exponentially increasing number of researchers. Here’s BJRW’s data on crop yield:

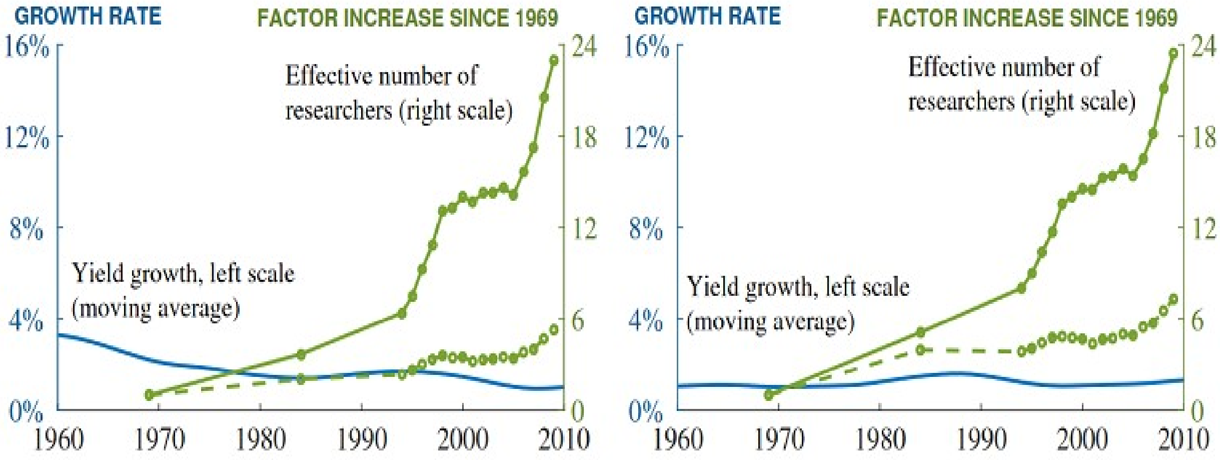
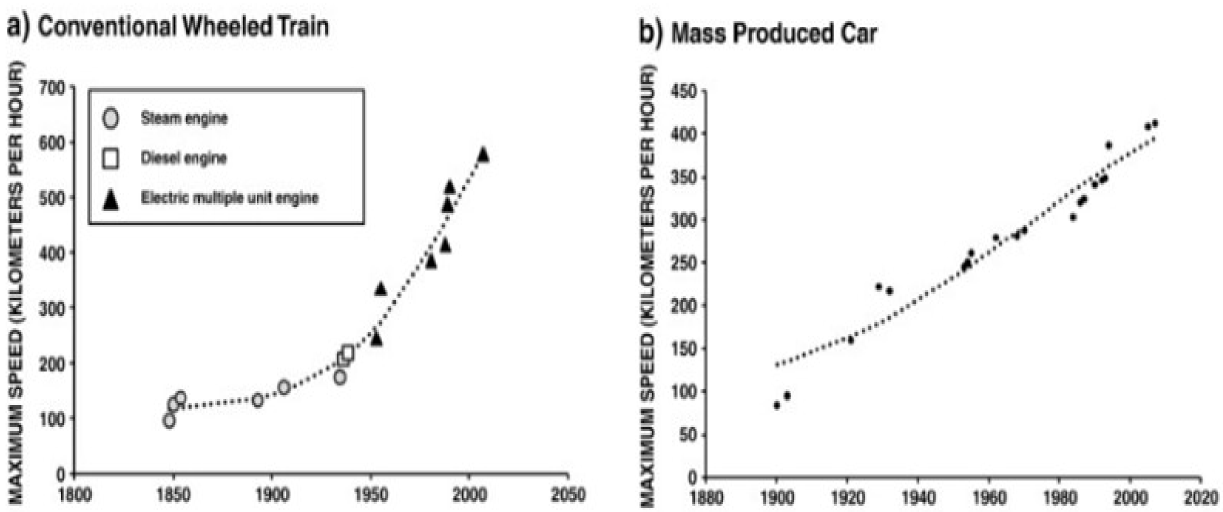


Figure 7: Yield Growth and Research Effort by Crop

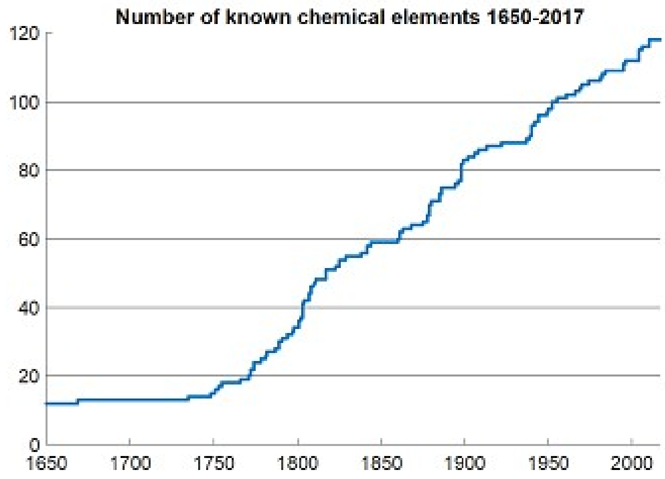
(a) Corn (b) Soybeans

The solid and dashed lines are two different measures of crop-related research. Even though the crop-related research increases by a factor of 6-24x (depending on how it’s measured), crop yields grow at a relatively constant 1% rate for soybeans, and apparently declining 3%ish percent rate for corn.

BJRW go on to prove the same is true for whatever other scientific fields they care to measure. Measuring scientific progress is inherently difficult, but their finding of constant or log-constant progress in most areas accords with Nintil’s overview of the same topic, which gives us graphs like

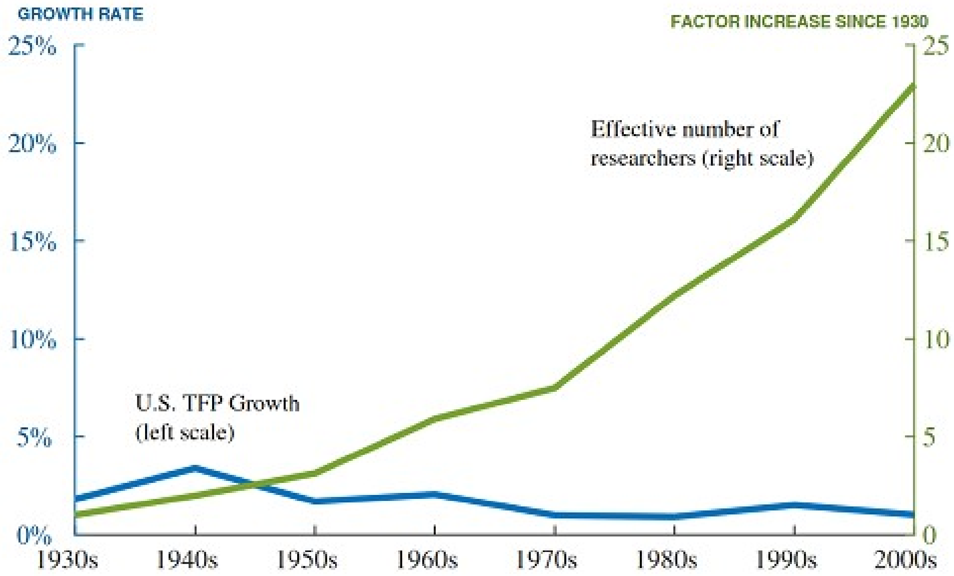


…and dozens more like it. And even when we use data that are easy to measure and hard to fake, like number of chemical elements discovered, we get the same linearity:



Meanwhile, the increase in researchers is obvious. Not only is the population increasing (by a factor of about 2.5x in the US since 1930), but the percent of people with college degrees has quintupled over the same period. The exact numbers differ from field to field, but orders of magnitude increases are the norm. For example, the number of people publishing astronomy papers seems to have dectupled over the past fifty years or so.

BJRW put all of this together into total number of researchers vs. total factor productivity of the economy, and find…



…about the same as with transistors, soybeans, and everything else. So if you take their methodology seriously, over the past ninety years, each researcher has become about 25x less productive in making discoveries that translate into economic growth.

Participants at the conference had some explanations for this, of which the ones I remember best are:

1. Only the best researchers in a field actually make progress, and the best researchers are already in a field, and probably couldn’t be kept *out of* the field with barbed wire and attack dogs. If you expand a field, you will get a bunch of merely competent careerists who treat it as a 9-to-5 job. A field of 5 truly inspired geniuses and 5 competent careerists will make X progress. A field of 5 truly inspired geniuses and 500,000 competent careerists will make the same X progress. Adding further competent careerists is useless for doing anything except making graphs look more exponential,

and we should stop doing it. See also Price’s Law Of Scientific Contributions.

2. Certain features of the modern academic system, like underpaid PhDs, interminably long postdocs, endless grant-writing drudgery, and clueless funders have lowered productivity. The 1930s academic system was indeed 25x more effective at getting researchers to actually do good research.

3. All the low-hanging fruit has already been picked. For example, element 117 was discovered by an international collaboration who got an unstable isotope of berkelium from the single accelerator in Tennessee capable of synthesizing it, shipped it to a nuclear reactor in Russia where it was attached to a titanium film, brought it to a particle accelerator in a different Russian city where it was bombarded with a custom-made exotic isotope of calcium, sent the resulting data to a global team of theorists, and eventually found a signature indicating that element 117 had existed for a few milliseconds. Meanwhile, the first modern element discovery, that of phosphorous in the 1670s, came from a guy looking at his own piss. We should not be surprised that discovering element 117 needed more people than discovering phosphorous.

Needless to say, my sympathies lean towards explanation number 3. But I worry even this isn’t dismissive enough. My real objection is that constant progress in science in response to exponential increases in inputs ought to be our null hypothesis, and that it’s almost inconceivable that it could ever be otherwise.

Consider a case in which we extend these graphs back to the beginning of a field. For example, psychology started with Wilhelm Wundt and a few of his friends playing around with stimulus perception. Let’s say there were ten of them working for one generation, and they discovered ten revolutionary insights worthy of their own page in Intro Psychology textbooks. Okay. But now there are about a hundred thousand experimental psychologists. Should we expect them to discover a hundred thousand revolutionary insights per generation?

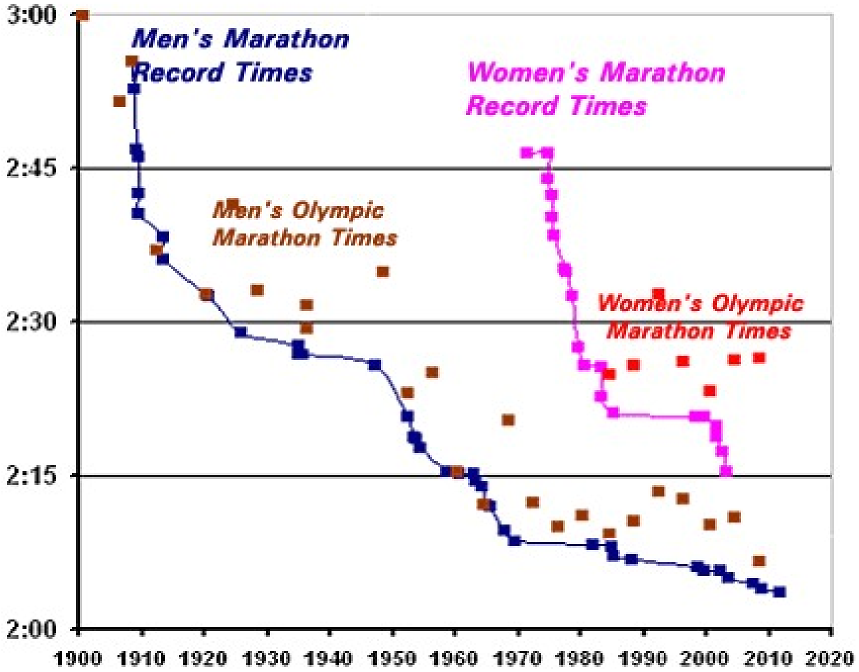
Or: the economic growth rate in 1930 was 2% or so. If it scaled with number of researchers, it ought to be about 50% per year today with our 25x increase in researcher number. That kind of growth would mean that the average person who made $30,000 a year in 2000 should make $50 million a year in 2018.

Or: in 1930, life expectancy at 65 was increasing by about two years per decade. But if that scaled with number of biomedicine researchers, that should have increased to ten years per decade by about 1955, which would mean everyone would have become immortal starting sometime during the Baby Boom, and we would currently be ruled by a deathless God-Emperor Eisenhower.

Or: the ancient Greek world had about 1% the population of the current Western world, so if the average Greek was only 10% as likely to be a scientist as the average modern, there were only 1/1000th as many Greek scientists as modern ones. But the Greeks made such great discoveries as the size of the Earth, the distance of the Earth to the sun, the prediction of eclipses, the heliocentric theory, Euclid’s geometry, the nervous system, the cardiovascular system, etc, and brought technology up from the Bronze Age to the Antikythera mechanism. Even adjusting for the long time scale to which “ancient Greece” refers, are we sure that we’re producing 1000x as many great discoveries as they are? If we extended BJRW’s graph all the way back to Ancient Greece, adjusting for the change in researchers as civilizations rise and fall, wouldn’t it keep the same shape as does for this century? Isn’t the real question not “Why isn’t Dwight Eisenhower immortal god-emperor of Earth?” but “Why isn’t Marcus Aurelius immortal god-emperor of Earth?”

Or: what about human excellence in other fields? Shakespearean England had 1% of the population of the modern Anglosphere, and presumably even fewer than 1% of the artists. Yet it gave us Shakespeare. Are there a hundred Shakespeare-equivalents around today? This is a harder problem than it seems – Shakespeare has become so venerable with historical hindsight that maybe nobody would acknowledge a Shakespeare-level master today even if they existed – but still, a hundred Shakespeares? If we look at some measure of great works of art per era, we find past eras giving us far more than we would predict from their population relative to our own. This is very hard to judge, and I would hate to be the guy who has to decide whether Harry Potter is better or worse than the *Aeneid*. But still? *A hundred Shakespeares?*

Or: what about sports? Here’s marathon records for the past hundred years or so:



In 1900, there were only two local marathons (eg the Boston Marathon) in the world. Today there are over 800. Also, the world population has increased by a factor of five (more than that in the East African countries that give us literally 100% of top male marathoners). Despite that, progress in marathon records has been steady or declining. Most other Olympics sports show the same pattern.

All of these lines of evidence lead me to the same conclusion: constant growth rates in response to exponentially increasing inputs is the null hypothesis. If it wasn't, we should be expecting 50% year-on-year GDP growth, easily-discovered-immortality, and the like. Nobody expected that before reading BJRW, so we shouldn't be surprised when BJRW provide a data-driven model showing it isn't happening. I realize this in itself isn't an explanation; it doesn't tell us why researchers can't maintain a constant level of output as measured in discoveries. It sounds a little like "God wouldn't design the universe that way", which is a kind of suspicious line of argument, especially for atheists.

But it at least shifts us from a lens where we view the problem as “What three tweaks should we make to the graduate education system to fix this problem *right now*?” to one where we view it as “Why isn’t Marcus Aurelius immortal?”

And through such a lens, only the “low-hanging fruits” explanation makes sense. Explanation 1 – that progress depends only on a few geniuses – isn’t enough. After all, the Greece-today difference is partly based on population growth, and population growth should have produced proportionately more geniuses. Explanation 2 – that PhD programs have gotten worse – isn’t enough. There would have to be a worldwide monotonic decline in every field (including sports and art) from Athens to the present day. Only Explanation 3 holds water.

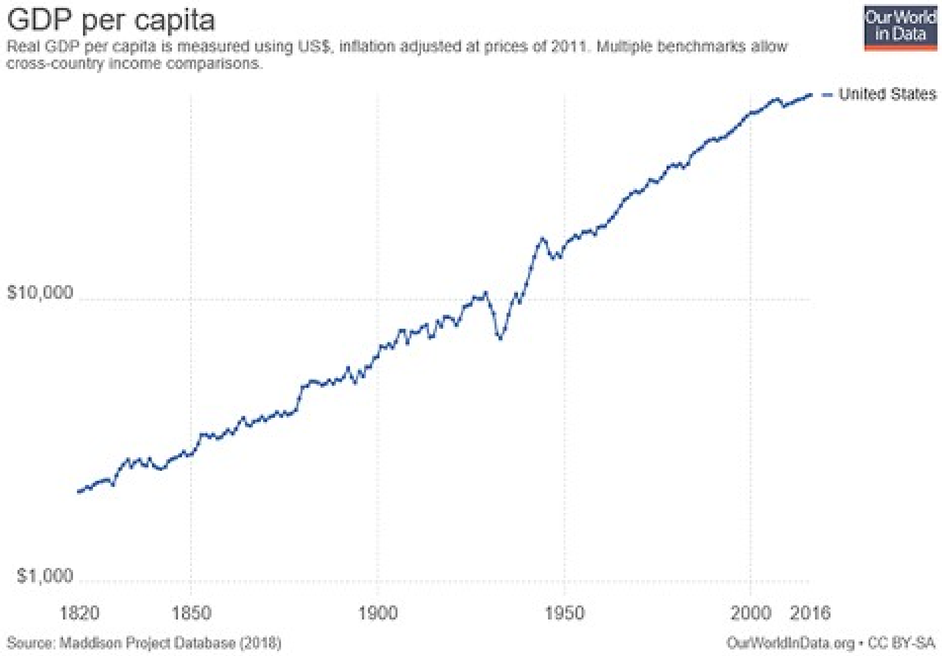
I brought this up at the conference, and somebody reasonably objected – doesn’t that mean science will stagnate soon? After all, we can’t keep feeding it an exponentially increasing number of researchers forever. If nothing else stops us, then at some point, 100% (or the highest plausible amount) of the human population will be researchers, we can only increase as fast as population growth, and then the scientific enterprise collapses.

I answered that the Gods Of Straight Lines are more powerful than the Gods Of The Copybook Headings, so if you try to use common sense on this problem you will fail.

Imagine being a futurist in 1970 presented with Moore’s Law. You scoff: “If this were to continue only 20 more years, it would mean a million transistors on a single chip! You would be able to fit an entire supercomputer in a shoebox!” But common sense was wrong and the trendline was right.

"If this were to continue only 40 more years, it would mean ten *billion* transistors per chip! You would need more transistors on a single chip than there are humans in the world! You could have computers more powerful than any today, that are too small to even see with the naked eye! You would have transistors with like a double-digit number of atoms!" But common sense was wrong and the trendline was right.

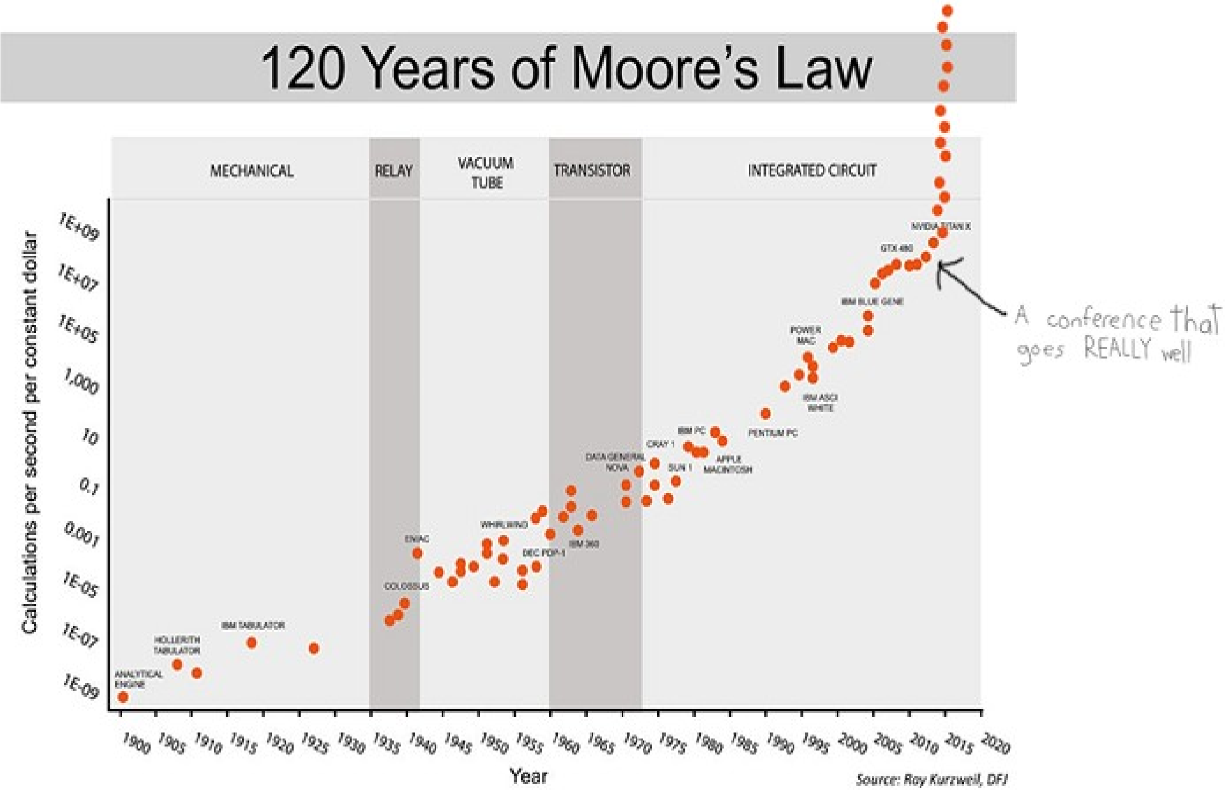
Or imagine being a futurist in ancient Greece presented with world GDP doubling time. Take the trend seriously, and in two thousand years, the future would be fifty thousand times richer. Every man would live better than the Shah of Persia! There would have to be so many people in the world you would need to tile entire countries with cityscape, or build structures higher than the hills just to house all of them. Just to sustain itself, the world would need transportation networks orders of magnitude faster than the fastest horse. But common sense was wrong and the trendline was right.



I’m not saying that no trendline has ever changed. Moore’s Law seems to be legitimately slowing down these days. The Dark Ages shifted every macrohistorical indicator for the worse, and the Industrial Revolution shifted every macrohistorical indicator for the better. Any of these sorts of things could happen again, easily. I’m just saying that “Oh, that exponential trend can’t *possibly* continue” has a really bad track record. I do not understand the Gods Of Straight Lines, and honestly they creep me out. But I would not want to bet against them.

Grace et al’s survey of AI researchers show they predict that AIs will start being able to do science in about thirty years, and will exceed the productivity of human researchers in every field shortly afterwards. Suddenly “there aren’t enough humans in the entire world to do the amount of research necessary to continue this trend line” stops sounding so compelling.

At the end of the conference, the moderator asked how many people thought that it was possible for a concerted effort by ourselves and our institutions to “fix” the “problem” indicated by BJRW’s trends. Almost the entire room raised their hands. Everyone there was smarter and more prestigious than I was (also richer, and in many cases way more attractive), but with all due respect I worry they are insane. This is kind of how I imagine their worldview looking:

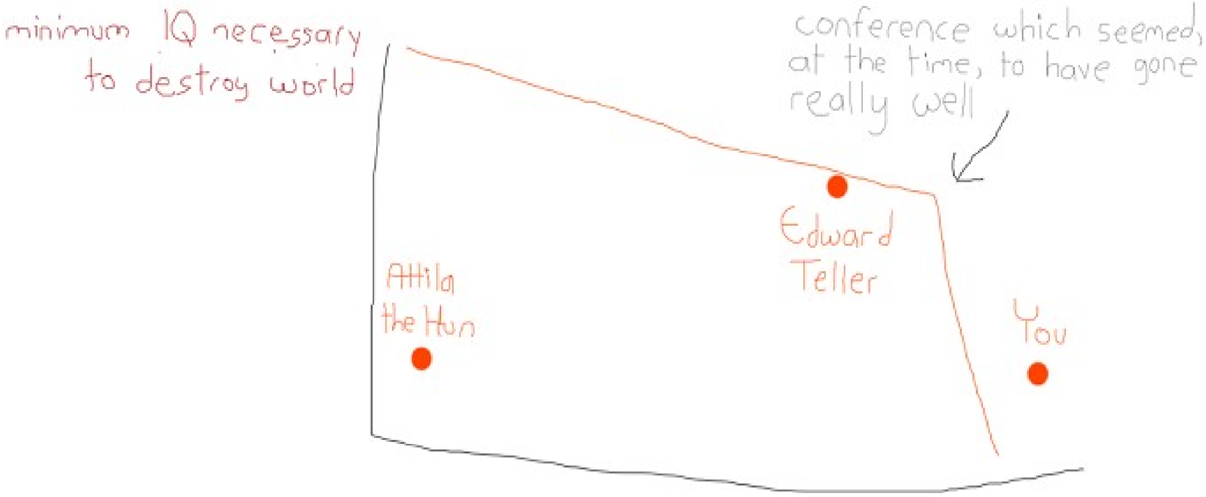


Source: Ray Kurzweil, DFJ

I realize I’m being fatalistic here. Doesn’t my position imply that the scientists at Intel should give up and let the Gods Of Straight Lines do the work? Or at least that the head of the National Academy of Sciences should do something like that? That Francis Bacon was wasting his time by inventing the scientific method, and Fred Terman was wasting his time by organizing Silicon Valley? Or perhaps that the Gods Of Straight Lines were acting *through* Bacon and Terman, and they had no choice in their actions? How do we know that the Gods aren’t acting through our conference? Or that our studying these things isn’t the only thing that keeps the straight lines going?

I don’t know. I can think of some interesting models – one made up of a thousand random coin flips a year has some nice qualities – but I don’t know.

I do know you should be careful what you wish for. If you “solved” this “problem” in classical Athens, Attila the Hun would have had nukes. Remember Yudkowsky's Law of Mad Science: "Every eighteen months, the minimum IQ necessary to destroy the world drops by one point." Do you really want to make that number ten points? A hundred? I am kind of okay with the function mapping number of researchers to output that we have right now, thank you very much.



The Chamber of Guf

[I briefly had a different piece up tonight discussing a conference, but the organizers asked me to hold off on writing about it until they've put up their own synopsis. It will be back up eventually; please accept this post instead for now.]

In Jewish legend, the Chamber of Guf is a pit where all the proto-souls hang out whispering and murmuring. Whenever a child is born, an angel reaches into the chamber, scoops up a soul, and brings it into the world.

In the syncretist mindset where every legend has to be a metaphor for the human mind, I map the Chamber of Guf to all the thoughts that exist below the level of consciousness, fighting for attention.

We already know something like this happens for behaviors. From Guyenet's *The Hungry Brain*:

> How does the lamprey decide what to do? Within the lamprey basal ganglia lies a key structure called the striatum, which is the portion of the basal ganglia that receives most of the incoming signals from other parts of the brain. The striatum receives "bids" from other brain regions, each of which represents a specific action. A little piece of the lamprey's brain is whispering "mate" to the striatum, while another piece is shouting "flee the predator" and so on. It would be a very bad idea for these movements to occur simultaneously – because a lamprey can't do all of them at the same time – so to prevent simultaneous activation of many different movements, all these regions are held in check by powerful inhibitory connections from the basal ganglia. This means that the basal ganglia keep all behaviors in "off" mode by default. Only once a specific action's bid has been selected do the basal ganglia turn off this

> inhibitory control, allowing the behavior to occur. You can think of the basal ganglia as a bouncer that chooses which behavior gets access to the muscles and turns away the rest. This fulfills the first key property of a selector: it must be able to pick one option and allow it access to the muscles.

So in the process of deciding what behavior to do, the (lamprey) brain subconsciously considers many different plausible behaviors, all of which compete to be enacted. I don’t know how this extends to humans, but it would make sense that maybe only the top few candidate behaviors even make it to consciousness, with the rest getting rejected without conscious consideration.

The particular qualities of a behavior that help it reach consciousness and implementation vary depending on mental state. Guyenet goes on to talk about how in dopamine-depleted states, only the simplest and most boring behaviors make it out of the Guf; with enough dopamine blockade, a person will sit motionless in their room for lack of any better ideas. In high dopamine states like mania or methamphetamine use, it’s much easier for behaviors to make successful “bids”, and so you tend to do bizarre things that would never have seemed like good ideas otherwise.

This is how I experience thoughts too. When I’ve had a lot of coffee, I have more interesting thoughts than usual. New ideas and clever wordplay come easily to me. I don’t think it makes sense to say that coffee makes me smarter; that breaks Algernon’s Law. More likely I always have some of those thoughts in the Guf, but the relevant angel considers them too weird to be worth scooping out and bringing into the world. This is probably for the best; manic people report “racing thoughts”, a state where the angels build a giant conveyor belt from the Guf to consciousness and give you every single possible thought no matter how irrelevant. It doesn't sound fun at all.

I find this metaphor especially useful when thinking about Gay OCD.

Gay OCD, and its close cousins Pedophilic OCD and Incest OCD, are varieties of obsessive-compulsive disorder where the patient can't stop worrying that they're gay (or a pedophile, or want to have sex with family members). In these more tolerant times, it's tempting to say "whatever, you're gay, that's fine, get over it". But a careful history will reveal that they aren't; most Gay OCD patients do not experience same-sex attraction, and they're often in fulfilling relationships with members of the opposite sex. They have no good reason to think they're gay – they just constantly worry that they are.

I studied under a professor who was an expert in these conditions. Her theory centered around the question of why angels would select some thoughts from the Guf over others to lift into consciousness. Variables like truth-value, relevance, and interestingness play important roles. But the exact balance depends on our mood. Anxiety is a global prior in favor of extracting fear-related thoughts from the Guf. Presumably everybody's brain dedicates a neuron or two to thoughts like "a robber could break into my house right now and shoot me". But most people's Selecting Angels don't find them worth bringing into the light of consciousness. Anxiety changes the angel's orders: have a bias towards selecting thoughts that involve fearful situations and how to prepare for them. A person with an anxiety disorder, or a recent adrenaline injection, or whatever, will absolutely start thinking about robbers, even if they consciously know it's an irrelevant concern.

In a few unlucky people with a lot of anxiety, the angel decides that a thought provoking *any* strong emotion is sufficient reason to raise the thought to consciousness. Now the Gay OCD trap is sprung. One day the angel randomly scoops up the thought "I am gay" and hands it to the patient's consciousness. The patient notices the thought "I am gay", and falsely interprets it as evidence that they're actually gay, causing fear and disgust and self-doubt. The angel notices this thought produced a lot of emotion and occupied consciousness for a long time – a success! That was *such* a good choice of thought! It must have been *so* relevant! It decides to stick with this strategy of using the "I am gay" thought from now on. If that ever fails to excite, it moves on to a whole host of similar thoughts that still have some punch, like "Was I just sexually attracted to that same-sex person over there?" and the like.

I practice in San Francisco, and I rarely see Gay OCD these days. Being gay just isn't scary enough any more. I still see some Pedophilic OCD and Incest OCD, as well as less common but obviously similar syndromes like Murderer OCD and Infanticide OCD. I've also started noticing a spike in Racism OCD; the patient has a stray racist thought, they react with sudden terror and self-loathing, their angel gets all excited, and then they can't stop thinking about whether they might be a racist. There's a paper to be written here about OCD patients as social weathervanes.

All of these can be treated with the same medications that treat normal OCD. But there's an additional important step of explaining exactly this theory to the patient, so that they know that not only are they not gay/a pedophile/racist, but it's actually their strong commitment to being against homosexuality/pedophilia/racism which is making them have these thoughts. This makes the thoughts provoke less strong emotion and can itself help reduce the frequency of obsessions. Even if it doesn’t do that, it’s at least comforting for most people.

This is *not* an official theory by an official professor, but I wonder how much of a role this same process plays in normal self-defeating thoughts. The person who can’t stop thinking “I’m fat and ugly” or “I’m an imposter who’s terrible at my career” even in the face of contradictory evidence. These thoughts seem calculated to disturb the same way Gay OCD is. They’re not as dramatic, and they rarely reach quite the same level of obsession, but the underlying process seems the same.

If you want to see the Guf directly, advanced meditators seem to be able to do this. They often report that after successfully quieting their conscious thoughts, they become gradually aware of a swamp of unquiet proto-thoughts lurking underneath. They usually describe it as really weird, which is a remarkably good match to the theory’s predictions.

Classified Thread 6

This is the…monthly? bimonthly? occasional?…classified thread. Post advertisements, personals, and any interesting success stories from the last thread.

Cognitive Enhancers: Mechanisms And Tradeoffs

[Epistemic status: so, so, so speculative. I do not necessarily endorse taking any of the substances mentioned in this post.]

There's been recent interest in "smart drugs" said to enhance learning and memory. For example, from the Washington Post:

> When aficionados talk about nootropics, they usually refer to substances that have supposedly few side effects and low toxicity. Most often they mean piracetam, which Giurgea first synthesized in 1964 and which is approved for therapeutic use in dozens of countries for use in adults and the elderly. Not so in the United States, however, where officially it can be sold only for research purposes. Piracetam is well studied and is credited by its users with boosting their memory, sharpening their focus, heightening their immune system, even bettering their personalities.

Along with piracetam, a few other substances have been credited with these kinds of benefits, including some old friends:

> "To my knowledge, nicotine is the most reliable cognitive enhancer that we currently have, bizarrely," said Jennifer Rusted, professor of experimental psychology at Sussex University in Britain when we spoke. "The cognitive-enhancing effects of nicotine in a normal population are more robust than you get with any other agent. With Provigil, for instance, the evidence for cognitive benefits is nowhere near as strong as it is for nicotine."

But why should there be smart drugs? Popular metaphors speak of drugs fitting into receptors like “a key into a lock” to “flip a switch”. But why should there be a locked switch in the brain to shift from THINK WORSE to THINK BETTER? Why not just always stay on the THINK BETTER side? Wouldn’t we expect some kind of tradeoff?

Piracetam and nicotine have something in common: both activate the brain’s acetylcholine system. So do three of the most successful Alzheimers drugs: donepezil, rivastigmine, and galantamine. What is acetylcholine and why does activating it improve memory and cognition?

Acetylcholine is many things to many people. If you’re a doctor, you might use neostigmine, an acetylcholinesterase inhibitor, to treat the muscle disease myasthenia gravis. If you’re a terrorist, you might use sarin nerve gas, a more dramatic acetylcholinesterase inhibitor, to bomb subways. If you’re an Amazonian tribesman, you might use curare, an acetylcholine receptor antagonist, on your blowdarts. If you’re a plastic surgeon, you might use Botox, an acetylcholine release preventer, to clear up wrinkles. If you’re a spider, you might use latrotoxin, another acetylcholine release preventer, to kill your victims – and then be killed in turn by neonictinoid insecticides, which are acetylcholine agonists. Truly this molecule has something for everybody – though gruesomely killing things remains its comparative advantage.

But to a computational neuroscientist, acetylcholine is:

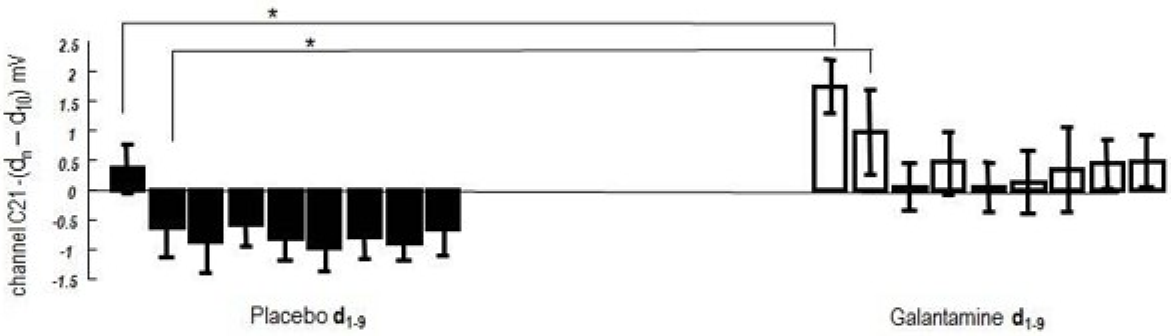
> …a neuromodulator [that] encodes changes in the precision of (certainty about) prediction errors in sensory cortical hierarchies. Each level of a processing hierarchy sends predictions to the level below, which reciprocate bottom-up signals. These signals are

> prediction errors that report discrepancies between top-down predictions and representations at each level. This recurrent message passing continues until prediction errors are minimised throughout the hierarchy. The ensuing Bayes optimal perception rests on optimising precision at each level of the hierarchy that is commensurate with the environmental statistics they represent. Put simply, to infer the causes of sensory input, the brain has to recognise when sensory information is noisy or uncertain and down-weight it suitably in relation to top-down predictions.

…is it too late to opt for the gruesome death? It is? Fine. God help us, let's try to understand Friston again.

In the predictive coding model, perception (maybe also everything else?) is a balance between top-down processes that determine what you should be expecting to see, and bottom-up processes that determine what you're actually seeing. This is faster than just determining what you're actually seeing without reference to top-down processes, because sensation is noisy and if you don't have some boxes to categorize things in then it takes forever to figure out what's actually going on. In this model, acetylcholine is a neuromodulator that indicates increased sensory precision – ie a bias towards expecting sensation to be signal rather than noise – ie a bias towards trusting bottom-up evidence rather than top-down expectations.

In the study linked above, Friston and collaborators connect their experimental subjects to EEG monitors and ask them to listen to music. The "music" is the same note repeated again and again at regular intervals in a perfectly predictable way. Then at some random point, it unexpectedly shifts to a different note. The point is to get their top-down systems confidently predicting a certain stimulus (the original note repeated again for the umpteenth time) and then surprise them with a different stimulus, and measure the EEG readings to see how their brain reacts. Then they do this again and again to see how the subjects eventually learn. Half the subjects have just taken galantamine, a drug that increases acetylcholine levels; the other half get placebo.



I don’t understand a lot of the figures in this paper, but I think I understand this one. It’s saying that on the first surprising note, placebo subjects’ brains got a bit more electrical activity [than on the predictable notes], but galantamine subjects’ brains got much more electrical activity. This fits the prediction of the theory. The placebo subjects have low sensory precision – they’re in their usual state of ambivalence about whether sensation is signal or noise. Hearing an unexpected stimulus is a little bit surprising, but not completely surprising – it might just be a mistake, or it might just not matter. The galantamine subjects’ brains are on alert to expect sensation to be very accurate and very important. When they hear the surprising note, their brains are very surprised and immediately reevaluate the whole paradigm.

One might expect that the very high activity on the first discordant note would be matched with lower activity on subsequent notes; the brain has now fully internalized the new prediction (ie is expecting the new note) and can’t be surprised by it anymore. As best I can tell, this study doesn't really show that. A very similar study by some of the same researchers does. In this one, subjects on either galantamine or a placebo have to look at a dot as quickly as possible after it appears. There are some arrows that might or might not point in the direction where the dot will appear; over the course of the experiment, the accuracy of these arrows changes. The researchers measured how quickly, when the meaning of the arrows changed, the subjects shifted from the old paradigm to the new paradigm. Galantamine enhanced the speed of this change a little, though it was all very noisy. Lower-weight subjects had a more dramatic change, suggesting an effective dose-dependent response (ie the more you weigh, the less effect a constant-weight dose of galantamine will have on your body). They conclude:

> This interpretation of cholinergic action in the brain is also in accord with the assumption of previous theoretical notions posing that ACh controls the speed of the memory update (i.e., the learning rate)

"Learning rate" is a technical term often used in machine learning, and I got a friend who is studying the field to explain it to me (all mistakes here are mine, not hers). Suppose that you have a neural net trying to classify cats vs. dogs. It's already pretty well-trained, but it still makes some mistakes. Maybe it's never seen a Chihuahua before and doesn't know dogs can get that small, so it thinks "cat". A good neural network will learn from that mistake, but the amount it learns will depend on a parameter called learning rate:

If learning rate is 0, it will learn nothing. The weights won't change, and the next time it sees a Chihuahua it will make the exact same mistake.

If learning rate is very high, it will overfit. It will change everything to maximize the likelihood of getting that one picture of a Chihuahua right the next time, even if this requires erasing everything it has learned before, or dropping all “common sense” notions of dog and cat. It is now a “that one picture of a Chihuahua vs. everything else” classifier.

If learning rate is a little on the low side, the model will be very slow to learn, though it will eventually converge on a good understanding of its topic.

If learning rate is a little on the high side, the model will learn very quickly, but “jump around” between different understandings heavily weighted toward what best fits the last case it has worked on.

On many problems, it’s a good idea to start with a high learning rate in order to get a basic idea what’s going on first, then gradually lower it so you can make smaller jumps through the area near the right answer without overshooting.

Learning rates are sort of like sensory precision and bottom-up vs. top-down weights, in that as a high learning rate says to discount prior probabilities and weight the evidence of the current case more strongly. A higher learning rate would be appropriate in a signal-rich environment, and a lower rate appropriate in a noise-rich environment.

If acetylcholine helps set the learning rate in the brain, would it make sense that cholinergic substances are cognitive enhancers / “study drugs”?

You would need to cross a sort of metaphorical divide between a very mechanical and simple sense of “learning” and the kind of learning you do where you study for your final exam on US History. What would it mean to be telling your brain that your US History textbook is “a signal-rich environment” or that it should be weighting its bottom-up evidence of what the textbook says higher than its top-down priors?

Going way beyond the research into total speculation, we could imagine the brain having some high-level intuitive holistic sense of US history. Each new piece of data you receive could either be accepted as a relevant change to that, or rejected as “noise” in the sense of not worth updating upon. If you hear that the Battle of Cedar Creek took place on October 19, 1864 and was a significant event in the Shenandoah Valley theater of the Civil War, then – if you’re like most people – it will have no impact at all on anything beyond (maybe, if you’re lucky) being able to parrot back that exact statement. If you learn that the battle took place in 2011 and was part of a Finnish invasion of the US, that changes a lot and is pretty surprising and would radically alter your holistic intuitive sense of what history is like.

Thinking of it this way, I can imagine these study drugs helping the exact date of the Battle of Cedar Creek seem a little bit more like signal, and so have it make a little bit more of a dent in your understanding of history. I’m still not sure how significant this is, because the exact date of the battle isn’t surprising to me in any way, and I don’t know what I would update based on hearing it. But then again, these drugs have really subtle effects, so maybe not being able to give a great account of how they work is natural.

And what about the tradeoff? Is there one?

One possibility is no. The idea of signal-rich vs. signal-poor environments is useful if you’re trying to distinguish whether a certain pattern of blotches is a camoflauged tiger. Maybe it’s not so useful for studying US History. Thinking of Civil War factoids as anything less than maximally-signal-bearing might just be a mismatch of evolution to the modern environment, the same way as liking sweets more than vegetables.

Another possibility is that if you take study drugs in order to learn the date of the Battle of Cedar Creek, you are subtly altering your holistic intuitive knowledge of American history in a disruptive way. You’re shifting everything a little bit more towards a paradigm where the Battle of Cedar Creek was unusually important. Maybe the people who took piracetam to help them study ten years ago are the same people who go around now talking about how the Civil War explains every part of modern American politics, and the 2016 election was just the Confederacy getting revenge on the Union, and how the latest budget bill is just a replay of the Missouri Compromise.

And another possibility is that you’re learning things in a rote, robotic way. You can faithfully recite that the Battle of Cedar Creek took place on October 19, 1864, but you’re less good at getting it to hang together with anything else into a coherent picture of what the Civil War was really *like*. I’m not sure if this makes sense in the context of the learning rate metaphor we’re using, but it fits the anecdotal reports of some of the people who use Adderall – which has some cholinergic effects in addition to its more traditional catecholaminergic ones.

Or it might be weirder than this. Remember the aberrant salience model of psychosis, and schizophrenic Peter Chadwick talking about how one day he saw the street “New King Road” and decided that it meant Armageddon was approaching, since Jesus was the new king coming to replace the old powers of the earth? Is it too much of a stretch to say this is what happens when your learning rate is much too high, kind of like the neural network that changes everything to explain one photo of a Chihuahua? Is this why nicotine has weird effects on schizophrenia? Maybe higher learning rates can promote psychotic thinking – not necessarily dramatically, just make things get a little weird.

Having ventured this far into Speculation-Land, let’s retreat a little. Noradrenergic and Cholinergic Moduation of Belief Updating does some more studies and fails to find any evidence that scopolamine, a cholinergic drug, alters learning rate (but why would they use scopolamine, which acts on muscarinic acetylcholine receptors, when every drug suspected to improve memory act on nicotinic ones?). Also, nicotine seems to help schizophrenia, not worsen it, which is the opposite of the last point above. Also, everything above about acetylcholine sounds kind of like my impression of where dopamine fits in this model, especially in terms of it standing for the precision of incoming data. This suggests I don’t understand the model well enough for everything not to just blend together to me. All that my usual sources will tell me is that the acetylcholine system modulates the dopamine system.

(remember, neuroscience madlibs is “________ modulates __________”, and no matter what words you use the sentence is always true.)

Nominating Oneself For The Short End Of A Tradeoff

I've gotten a chance to discuss The Whole City Is Center with a few people now. They remain skeptical of the idea that anyone could "deserve" to have bad things happen to them, based on their personality traits or misdeeds.

These people tend to imagine the pro-desert faction as going around, actively hoping that lazy people (or criminals, or whoever) suffer. I don't know if this passes an Intellectual Turing Test. When I think of people deserving bad things, I think of them having *nominated themselves to get the short end of a tradeoff*.

Let me give three examples:

1. Imagine an antidepressant that works better than existing antidepressants, one that consistently provides depressed people real relief. If taken as prescribed, there are few side effects and people do well. If ground up, snorted, and taken at ten times the prescribed dose – something nobody could do by accident, something you have to really be trying to get wrong – it acts as a passable heroin substitute, you can get addicted to it, and it will ruin your life.

The antidepressant is popular and gets prescribed a lot, but a black market springs up, and however hard the government works to control it, a lot of it gets diverted to abusers. Many people get addicted to it and their lives are ruined. So the government bans the antidepressant, and everyone has to go back to using SSRIs instead.

Let’s suppose the government is being good utilitarians here: they calculated out the benefit from the drug treating people’s depression, and the cost from the drug being abused, and they correctly determined the costs outweighed the benefits.

But let’s also suppose that nobody abuses the drug by accident. The difference between proper use and abuse is not subtle. Everybody who knows enough to know anything about the drug at all has heard the warnings. Nobody decides to take ten times the recommended dose of antidepressant, crush it, and snort it, through an innocent mistake. And nobody has just never heard the warnings that drugs are bad and can ruin your life.

Somebody is going to get the short end of the stick. If the drug is banned, depressed people will lose access to relief for their condition. If the drug is permitted, recreational users will continue having the opportunity to destroy their lives. And we’ve posited that the utilitarian calculus says that banning the antidepressant would be better. But I still feel, in some way, that the recreational users have nominated themselves to get the worse end of this tradeoff. Depressed people shouldn’t have to suffer because you see a drug that says very clearly on the bottle “DO NOT TAKE TOO MUCH OF THIS YOU WILL GET ADDICTED AND IT WILL BE TERRIBLE” and you think “I think I shall take too much of this”.

(this story is loosely based on the history of tianeptine in the US)

2. Suppose you’re in a community where some guy is sexually harassing women. You tell him not to and he keeps doing it, because that’s just the kind of guy he is, and it’s unclear if he can even stop himself. Eventually he does it so much that you kick him out of the community.

Then one of his friends comes to you and says “This guy harassed one woman per month, and not even that severely. On the other hand, kicking him out of the community costs him all of his friends, his support network, his living situation, and his job. He is a pretty screwed-up person and it’s unclear he will ever find more friends or another community. The cost to him of not being in this community, is actually greater than the cost of being harassed is to a woman.”

Somebody is going to have their lives made worse. Either the harasser’s life will be worse because he’s kicked out of the community. Or women’s lives are worse because they are being harassed. Even if I completely believe the friend’s calculation that kicking him out will bring more harm on him than keeping him would bring harm to women, I am still comfortable letting him get the short end of the tradeoff.

And this is true even if we are good determinists and agree he only harasses somebody because of an impulse control problem secondary to an underdeveloped frontal lobe, or whatever the biological reason for harassing people might be.

(not going to bring up what this story is loosely based on, but it’s not completely hypothetical either)

3. Sometimes in discussions of basic income, someone expresses concern that some people’s lives might become less meaningful if they didn’t have a job to give them structure and purpose.

And I respond “Okay, so those people can work, basic income doesn’t *prohibit* you from working, it just means you don’t have to.”

And they object “But maybe these people will choose not to work even though work would make them happier, and they will just suffer and be miserable.”

Again, there’s a tradeoff. Somebody’s going to suffer. If we don’t grant basic income, it will be people stuck in horrible jobs with no other source of income. If we do grant basic income, it will be people who need work to have meaning in their lives, but still refuse to work. Since the latter group has a giant door saying “SOLUTION TO YOUR PROBLEMS” wide open in front of them but refuses to take it, I find myself sympathizing more with the former group. That’s true even if some utilitarian were to tell me that the latter group outnumbers them.

I find all three of these situations joining the increasingly numerous ranks of problems where my intuitions differ from utilitarianism. What should I do?

One option is to dismiss them as misfirings of the heuristic “expose people to the consequences of their actions so that they are incentivized to make the right action”. I’ve tried to avoid that escape by specifying in each example that even when they’re properly exposed and incentivized the calculus still comes out on the side of making the tradeoff in their favor. But maybe this is kind of like saying “Imagine you could silence this one incorrect person without any knock-on effects on free speech anywhere else and all the consequences would be positive, would you do it?” In the thought experiment, maybe yes; in the real world this either never happens, or never happens with 100% certainty, or never happens in a way that’s comfortably outside whatever Schelling fence you’ve built for yourself. I’m not sure I find that convincing because in real life we don’t treat “force people to bear the consequences of their action" as a 100% sacred principle that we never violate.

Another option is to dismiss them as people "revealing their true preferences", eg if the harasser doesn't stop harassing women, he must not want to be in the community *too* much. But I think this operates on a really sketchy idea of revealed preference, similar to the Caplanian one where if you abuse drugs that just means you like drugs so there's no problem. Most of these situations feel like times when that simplified version of preferences breaks down.

A friend reframes the second situation in terms of the cost of having law at all. It's important to be able to make rules like "don't sexually harass people", and adding a clause saying "…but we'll only enforce these when utilitarianism says it's correct" makes them less credible and creates the opportunity for a lot of corruption. I can see this as a very strong answer to the second scenario (which might be the strongest), although I'm not sure it applies much to the first or third.

I could be convinced that my desire to let people who make bad choices nominate themselves for the short end of tradeoffs is just the utilitarian justifications (about it incentivizing behavior, or it revealing people's true preferences) crystallized into a moral principle. I'm not sure if I hold this moral principle or not. I'm reluctant to accept the ban-antidepressant, tolerate-harasser, and repeal-basic-income solutions, but I'm also not sure what justification I have for not doing so except "Here's a totally new moral principle I'm going to tack onto the side of my existing system".

But I hope people at least find this a more sympathetic way of understanding when people talk about "desert" than a caricatured story where some people just need to suffer because they’re bad.

Sort By Controversial

[Epistemic status: fiction]

Thanks for letting me put my story on your blog. Mainstream media is crap and no one would have believed me anyway.

This starts in September 2017. I was working for a small online ad startup. You know the ads on Facebook and Twitter? We tell companies how to get them the most clicks. This startup – I won't tell you the name – was going to add deep learning, because investors will throw money at anything that uses the words "deep learning". We train a network to predict how many upvotes something will get on Reddit. Then we ask it how many likes different ads would get. Then we use whatever ad would get the most likes. This guy (who is not me) explains it better. Why Reddit? Because the upvotes and downvotes are simpler than all the different Facebook reacts, plus the subreddits allow demographic targeting, plus there's an archive of 1.7 billion Reddit comments you can download for training data. We trained a network to predict upvotes of Reddit posts based on their titles.

Any predictive network doubles as a generative network. If you teach a neural net to recognize dogs, you can run it in reverse to get dog pictures. If you train a network to predict Reddit upvotes, you can run it in reverse to generate titles it predicts will be highly upvoted. We tried this and it was pretty funny. I don't remember the exact wording, but for /r/politics it was something like "Donald Trump is no longer the president. All transgender people are the president." For r/technology it was about Elon Musk saving Net Neutrality. You can also generate titles that will get maximum downvotes, but this is boring: it will just say things that sound like spam about penis pills.

Reddit has a feature where you can sort posts by controversial. You can see the algorithm here, but tl;dr it multiplies magnitude of total votes (upvotes + downvotes) by balance (upvote:downvote ratio or vice versa, whichever is smaller) to highlight posts that provoke disagreement. Controversy sells, so we trained our network to predict this too. The project went to this new-ish Indian woman with a long name who went by Shiri, and she couldn't get it to work, so our boss Brad sent me to help. Shiri had tested the network on the big 1.7 billion comment archive, and it had produced controversial-sounding hypothethical scenarios about US politics. So far so good.

The Japanese tested their bioweapons on Chinese prisoners. The Tuskegee Institute tested syphilis on African-Americans. We were either nicer or dumber than they were, because we tested Shiri's Scissor on ourselves. We had a private internal subreddit where we discussed company business, because Brad wanted all of us to get familiar with the platform. Shiri's problem was that she'd been testing the controversy-network on our subreddit, and it would just spit out vacuously true or vacuously false statements. No controversy, no room for disagreement. The statement we were looking at that day was about a design choice in our code. I won't tell you the specifics, but imagine you took every bad and wrong decision decision in the world, hard-coded them in the ugliest possible way, and then handed it to the end user with a big middle finger. Shiri's Scissor spit out, as maximally controversial, the statement that we should design our product that way. We'd spent ten minutes arguing about exactly where the bug was, when Shiri said something about how she didn't understand why the program was generating obviously true statements.

Shiri’s English wasn’t great, so I thought this was a communication problem. I corrected her. The program was spitting out obviously false statements. She stuck to her guns. I still thought she was confused. I walked her through the meanings of the English words “true” and “false”. She looked offended. I tried to confirm. She thought this abysmal programming decision, this plan of combining every bad design technique together and making it impossible to ever fix, was the right way to build our codebase? She said it was. Worse, she was confused I *didn’t* think so. She thought this was more or less what we were already doing; it wasn’t. She thought that moving away from this would take a total rewrite and make the code much worse.

At this point I was doubting my sanity, so we went next door to Blake and David, who were senior coders in our company and usually voices of reason. They were talking about their own problem, but I interrupted them and gave them the Scissor statement. Blake gave the reasonable response – why are you bothering me with this stupid wrong garbage? But David had the same confusion Shiri did and started arguing that the idea made total sense. The four of us started fighting. I still was sure Shiri and David just misunderstood the question, even though David was a native English-speaker and the question was crystal-clear. Meanwhile David was feeling more and more condescended to, kept protesting he wasn’t misunderstanding anything, that Blake and I were just crappy programmers who couldn’t make the most basic architecture decisions. He kept insisting the same thing Shiri had, that the Scissor statement had already been the plan and any attempt to go in a different direction would screw everything up. It got so bad that we decided to go to Brad for clarification.

Brad was our founder. Don’t trust the newspapers – not every tech entrepreneur is a greedy antisocial philistine. But everyone in advertising is. Brad definitely was. He was an abrasive amoral son of a bitch. But he was good at charming investors, and he could code, which is more than some bosses. He looked pissed to have the whole coding team come into his office unannounced, but he heard us out.

David tried to explain the issue, but he misrepresented almost every part of it. I couldn’t believe he was lying just to look better to Brad. I cut him off. He told me not to interrupt him. Blake said if he wasn’t lying we wouldn’t have to interrupt to correct him, it degenerated from there. Somehow in the middle of all of this, Brad figured out what we were talking about and he cut us all off. “That’s the stupidest thing I ever heard.” He confirmed it wasn’t the original plan, it was contrary to the original plan, and it was contrary to every rule of good programming and good business. David and Shiri, who were bad losers, accused Blake and me of “poisoning” Brad. David said that of course Brad would side with us. Brad had liked us better from the beginning. We’d racked up cushy project after cushy project while he and Shiri had gotten the dregs. Brad told him he was a moron and should get back to work. He didn’t.

This part of the story ends at 8 PM with Brad firing David and Shiri for a combination of gross incompetence, gross insubordination, and being terrible human beings. With him giving a long speech on how he’d taken a chance on hiring David and Shiri, even though he knew from the beginning that they were unqualified charity cases, and at every turn they’d repaid his kindness with laziness and sabotage. With him calling them a drain on the company and implied they might be working for our competitors. With them calling him an abusive boss, saying the whole company was a scam to trick vulnerable employees into working themselves ragged for Brad’s personal enrichment, and with them accusing us two – me and Blake – of being in on it with Brad.

That was 8 PM. We’d been standing in Brad’s office fighting for five hours. At 8:01, after David and Shiri had stormed out, we all looked at each other and thought – holy shit, the controversial filter works.

I want to repeat that. At no time in our five hours of arguing did this occur to us. We were too focused on the issue at hand, the Scissor statement itself. We didn’t have the perspective to step back and think about how all this controversy came from a statement designed to be maximally controversial. But at 8:01, when the argument was over and we had won, we stepped back and thought – holy shit.

We were too tired to think much about it that evening, but the next day we – Brad and the two remaining members of the coding team – had a meeting. We talked about what we had. Blake gave it its name: Shiri’s Scissor. In some dead language, scissor shares a root with schism. A scissor is a schism-er, a schism-creator. And that was what we had. We were going to pivot from online advertising to superweapons. We would call the Pentagon. Tell them we had a program that could make people hate each other. Was this ethical? We were in online ads; we would sell our grandmothers to Somali slavers if we thought it would get us clicks. That horse had left the barn a long time ago.

It’s hard to just call up the Pentagon and tell them you have a superweapon. Even in Silicon Valley, they don’t believe you right away. But Brad called in favors from his friends, and about a week after David and Shiri got fired, we had a colonel from DARPA standing in the meeting room, asking what the hell we thought was so important.

Now we had a problem. We couldn’t show the Colonel the Scissor statement that had gotten Dave and Shiri fired. He wasn’t in our company; he wasn’t even in ad tech; it would seem boring to him. We didn’t want to generate a new Scissor statement for the Pentagon. Even Brad could figure out that having the US military descend into civil war would be bad for clicks. Finally we settled on a plan. We explained the concept of Reddit to the Colonel. And then we asked him which community he wanted us to tear apart as a demonstration.

He thought for a second, then said “Mozambique”.

We had underestimated the culture gap here. When we asked the Colonel to choose a community to be a Scissor victim, we were expecting “tabletop wargamers” or “My Little Pony fans”. But this was not how colonels at DARPA thought about the world. He said “Mozambique”. I started explaining to him that this wasn’t really how Reddit worked, it needed to be a group with its own subreddit. Brad interrupted me, said that Mozambique had a subreddit.

I could see the wheels turning in Brad’s eyes. One wheel was saying “this guy is already skeptical, if we look weak in front of him he’ll just write us off completely”. The other wheel was calculating how many clicks Mozambique produced. Mene mene tekel upharsin. “Yeah,” he said. “Their subreddit is fine. We can do Mozambique.”

The Colonel gave us his business card and left. Blake and I were stuck running Shiri’s Scissor on the Mozambique subreddit. I know, ethics, but like I said, online ads business, horse, barn door. The only decency we allowed ourselves was to choose the network’s tenth pick – we didn’t need to destroy everything, just give a demonstration. We got a statement accusing the Prime Minister of disrespecting

Islam in a certain way – again, I won’t be specific. In the absence of any better method, we PMed the admins of the Mozambique subreddit asking them what they thought. I don’t remember what we said, something about being an American political science student learning about Mozambique culture, and could they ask some friends what would happen if the Prime Minister did that specific thing, and then report back to us?

We spent most of a week working on our project to undermine Mozambique. Then we got the news. David and Shiri were suing the company for unfair dismissal and racial discrimination. Brad and Blake and I were white. Shiri was an Indian woman, and David was Jewish. The case should have been laughed out of court – who ever heard of an anti-Semitic Silicon Valley startup? – except that all the documentation showed there was no reason to fire David and Shiri. Their work looked good on paper. They’d always gotten good performance reviews. The company was doing fine – it had even placed ads for more programmers a few weeks before.

David and Shiri knew why they’d been fired. But it didn’t matter to them. They were so blinded with hatred for our company, so caught in the grip of the Scissor statement, that they would tell any lie necessary to destroy it. We were caught in a bind. We couldn’t admit the existence of Shiri’s Scissor, because we were trying to sell it to the Pentagon as a secret weapon, and also, publicly admitting to trying to destroy Mozambique would have been bad PR. But the court was demanding records about what our company had been doing just before and just after the dismissal. A real defense contractor could probably have gotten the Pentagon to write a letter saying our research was classified. But the Pentagon still didn’t believe us. The Colonel was humoring us, nothing more. We were stuck.

I don’t know how we would have dealt with the legal problems, because what actually happened was Brad went to David’s house and tried to beat him up. You’re going to think this was crazy, but you have to understand that David had always been annoying to work with, and that during the argument in Brad’s office he had crossed so many lines that, if ever there was a person who deserved physical violence, it was him. Suing the company was just the last straw. I’m not going to judge Brad’s actions after he’d spent months cleaning up after David’s messes, paying him good money, and then David betrayed him at the end. But anyhow, that was it for our company. Brad got arrested. There was nobody else to pay the bills and keep the lights on. Blake and I were coders and had no idea how to run the business side of things. We handed in our resignations – not literally, Brad was in jail – and that was the end of Name Withheld Online Ad Company, Inc.

We got off easy. That’s the takeaway I want to give here. We were unreasonably overwhelmingly lucky. If Shiri and I had started out by arguing about one of the US statements, we could have destroyed the country. If a giant like Google had developed Shiri’s Scissor, it would have destroyed Google. If the Scissor statement we generated hadn’t just been about a very specific piece of advertising software – if it had been about the tech industry in general, or business in general – we could have destroyed the economy.

As it was, we just destroyed our company and maybe a few of our closest competitors. If you look up internal publications from the online advertising industry around fall 2017, you will find some really weird stuff. That story about the online ads CEO getting arrested for murder, child abuse, attacking a cop, and three or four other things, and then later it was all found to be false accusations related to some ill-explained mental disorder – that’s the tip of the iceberg. I don't have a good explanation for exactly how the Scissor statement spread or why it didn't spread further, but I bet if I looked into it too much, black helicopters would start hovering over my house. And that's all I'm going to say about that.

As for me, I quit the whole industry. I picked up a job in a more established company using ML for voice recognition, and tried not to think about it too much. I still got angry whenever I thought about the software design issue the Scissor had brought up. Once I saw someone who looked like Shiri at a cafe and I went over intending to give her a piece of my mind. It wasn't her, so I didn't end up in jail with Brad. I checked the news from Mozambique every so often, and it was quiet for a few months, and then it wasn't. I still don't know if we had anything to do with that. Africa just has a lot of conflicts, and if you wait long enough, maybe something will happen. The colonel never tried to get in touch with me. I don't think he ever took us seriously. Maybe he didn't even check the news from Mozambique. Maybe he saw it and figured it was a coincidence. Maybe he tried calling our company, got a message saying the phone was out of service, and didn't think it was worth pursuing. But as time went on and the conflict there didn't get any worse, I hoped the Shiri's Scissor part of my life was drawing to a close.

Then came the Kavanaugh hearings. Something about them gave me a sense of deja vu. The week of his testimony, I figured it out.

Shiri had told me that when she ran the Scissor on the site in general, she'd just gotten some appropriate controversial US politics scenarios. She had shown me two or three of them as examples. One of them had been very specifically about this situation. A Republican Supreme Court nominee accused of committing sexual assault as a teenager.

This made me freak out. Had somebody gotten hold of the Scissor and started using it on the US? Had that Pentagon colonel been paying more attention than he let on? But why would the Pentagon be trying to divide America? Had some enemy stolen it? I get the *New York Times*, obviously Putin was my first thought here. But how would Putin get Shiri’s Scissor? Was I remembering wrong? I couldn’t get it out of my head. I hadn’t kept the list Shiri had given me, but I had enough of the Scissor codebase to rebuild the program over a few sleepless nights. Then I bought a big blob of compute from Amazon Web Services and threw it at the Reddit comment archive. It took three days and a five-digit sum of money, but I rebuilt the list Shiri must have had. Kavanaugh was in there, just as I remembered.

But so was Colin Kaepernick.

You’ve heard of him. He was the football player who refused to stand for the national anthem. If I already knew the Scissor predicted one controversy, why was I so shocked to learn it predicted another? Because Kaepernick started kneeling in 2016. We didn’t build the Scissor until 2017. Putin hadn’t gotten it from us. Someone had beaten us to it.

Of the Scissor’s predicted top hundred most controversial statements, Kavanaugh was #58 and Kaepernick was #42. #86 was the Ground Zero Mosque. #89 was that baker who wouldn’t make a cake for a gay wedding. The match isn’t perfect, but #99 vaguely looked like the Elian Gonzalez case from 2000. That’s five out of a hundred. Is that what would happen by chance? It’s a big country, and lots of things happen here, and if a Scissor statement came up in the normal course of events it would get magnified to the national stage. But some of these were too specific. If it was coincidence, I would expect many more near matches than perfect matches. I found only two. The pattern of Scissor statements looked more like someone had arranged them to be perfect fits.

The earliest perfect fit was the Ground Zero Mosque in 2009. Could Putin have had a Scissor-like program in 2009? I say no way. This will sound weird to you if you're not in the industry. Why couldn't a national government have been eight years ahead of an online advertising company? All I can say is: machine learning moves faster than that. Russia couldn't hide a machine learning program that put it eight years ahead of the US. Even the Pentagon couldn't hide a program that put it eight years ahead of industry. The NSA is thirty years ahead of industry in cryptography and everyone knows it.

But then who was generating Scissor statements in 2009? I have no idea. And you know what? I can't bring myself to care.

If you just read a Scissor statement off a list, it's harmless. It just seems like a trivially true or trivially false thing. It doesn't activate until you start discussing it with somebody. At first you just think they're an imbecile. Then they call you an imbecile, and you want to defend yourself. Crescit eundo. You notice all the little ways they're lying to you and themselves and their audience every time they open their mouth to defend their imbecilic opinion. Then you notice how all the lies are connected, that in order to keep getting the little things like the Scissor statement wrong, they have to drag in everything else. Eventually even that doesn't work, they've just got to make everybody hate you so that nobody will even listen to your argument no matter how obviously true it is. Finally, they don't care about the Scissor statement anymore. They’ve just dug themselves so deep basing their whole existence around hating you and wanting you to fail that they can’t walk it back. You’ve got to prove them wrong, not because you care about the Scissor statement either, but because otherwise they’ll do anything to poison people against you, make it impossible for them to even understand the argument for why you deserve to exist. You know this is true. Your mind becomes a constant loop of arguments you can use to defend yourself, and rehearsals of arguments for why their attacks are cruel and unfair, and the one burning question: how can you thwart them? How can you convince people not to listen to them, before they find those people and exploit their biases and turn them against you? How can you combat the superficial arguments they’re deploying, before otherwise good people get convinced, so convinced their mind will be made up and they can never be unconvinced again? *How can you keep yourself safe?*

Shiri read two or three sample Scissor statements to me. She didn’t say if she agreed with them or not. I didn’t tell her if I agreed with them or not. They were harmless.

I don’t hear voices in a crazy way. But sometimes I talk to myself. Sometimes I do both halves of the conversation. Sometimes I imagine one of them is a different person. I had a tough breakup a year ago. Sometimes the other voice in my head is my ex-girlfriend’s voice. I know how she thinks and I always know what she would say about everything. So sometimes I hold conversations with her, even though she isn’t there, and we’ve barely talked since the breakup. I don’t know if this is weird. If it is, I’m weird.

And that was enough. For some reason, it was the third-highest-ranked Scissor statement that did it. None of the others, just that one. The totally hypothetical conversation with the version of my ex-girlfriend in my head about the third Scissor statement got me. Shiri’s Scissor was never really about other people anyway. Other people are just the trigger – and I use that word deliberately, in the trigger warning sense. Once you’re triggered, you never need to talk to anyone else again. Just the knowledge that those people are out there is enough.

I thought I’d be done with this story in a night. Instead it’s taken me two weeks, all the way up until Halloween – perfect night for a ghost story, right? I’ve been alternately drinking and smoking weed, trying to calm myself down enough to think about anything other than the third Scissor statement. No, that’s not right, definitely trying not to think about either of the first two Scissor statements, because if I think about them, I might start thinking about how some people disagree with them, and then I’m gone. Three times I’ve started to call my ex-girlfriend to ask her where she is, and if I ever go through with it and she answers me, I don’t know what I will do to her. But it isn’t just her. Fifty percent of the population disagrees with me on the third-highest-ranked Scissor statement. I don’t know who they are. I haven’t really appreciated that fact. Not really. I can’t imagine it being anyone I know. They’re too decent. But I can’t be sure it isn’t. So I drink.

I know I should be talking about how we all need to unite against whatever shadowy manipulators keep throwing Scissor statements at us. I want to talk about how we need to cultivate radical compassion and charity as the only defense against such abominations. I want to give an Obamaesque speech about how the ties that bring us together are stronger than the forces tearing us apart. But I can’t.

Remember what we did to Mozambique? How out of some vestigial sense of ethics, we released a low-potency Scissor statement? Arranged to give them a bad time without destroying the whole country all at once? That’s what our shadowy manipulators are doing to us. Low-potency statements. Enough to get us enraged. Not enough to start Armageddon.

But I read the whole list. And then, like an idiot, I thought about it. I thought about the third-highest-ranked Scissor statement in enough detail to let it trigger. To even begin to question whether it might be true is so sick, so perverse, so hateful and disgusting, that Idi Amin would flush with shame to even contemplate it. And if the Scissor’s right then half of you would be gung ho in support.

You guys, who haven’t heard a really bad Scissor statement yet and don’t know what it’s like – it’s easy for you to say “don’t let it manipulate you” or “we need a hard and fast policy of not letting ourselves fight over Scissor statements”. But how do you know you’re not in the wrong? How do you know there’s not an issue out there where, if you knew it, you would agree it would be better to just nuke the world and let us start over again from the sewer mutants, rather than let the sort of people who would support it continue to pollute the world with their presence? How do you know that you’re not like the schoolkid who superciliously says “*Nothing* is bad enough to deserve a swear word” when the worst that’s ever happened to her is dropping her lollipop in the dirt. If that schoolkid gets kidnapped and tortured, does she change her mind? If she can’t describe the torture to her schoolmates, but just says “a really bad thing happened to me”, and they still insist nothing could be bad enough to justify using swear words, who do you side with? Then why are you still thinking I’m “damaged” when I tell you I’ve seen the Scissor statement, and charity and compassion and unity can fuck off and die? Some last remnant of outside-view morality keeps me from writing the whole list here and letting you all exterminate yourselves. Some remnant of how I would have thought about these things a month ago holds me back. So listen:

Delete Facebook. Delete Twitter. Throw away your cell phone. Unsubscribe from the newspaper. Tell your friends and relatives not to discuss politics or society. If they slip up, break off all contact.

Then, buy canned food. Stockpile water. Learn to shoot a gun. If you can afford a bunker, get a bunker.

Because one day, whoever keeps feeding us Scissor statements is going to release one of the bad ones.

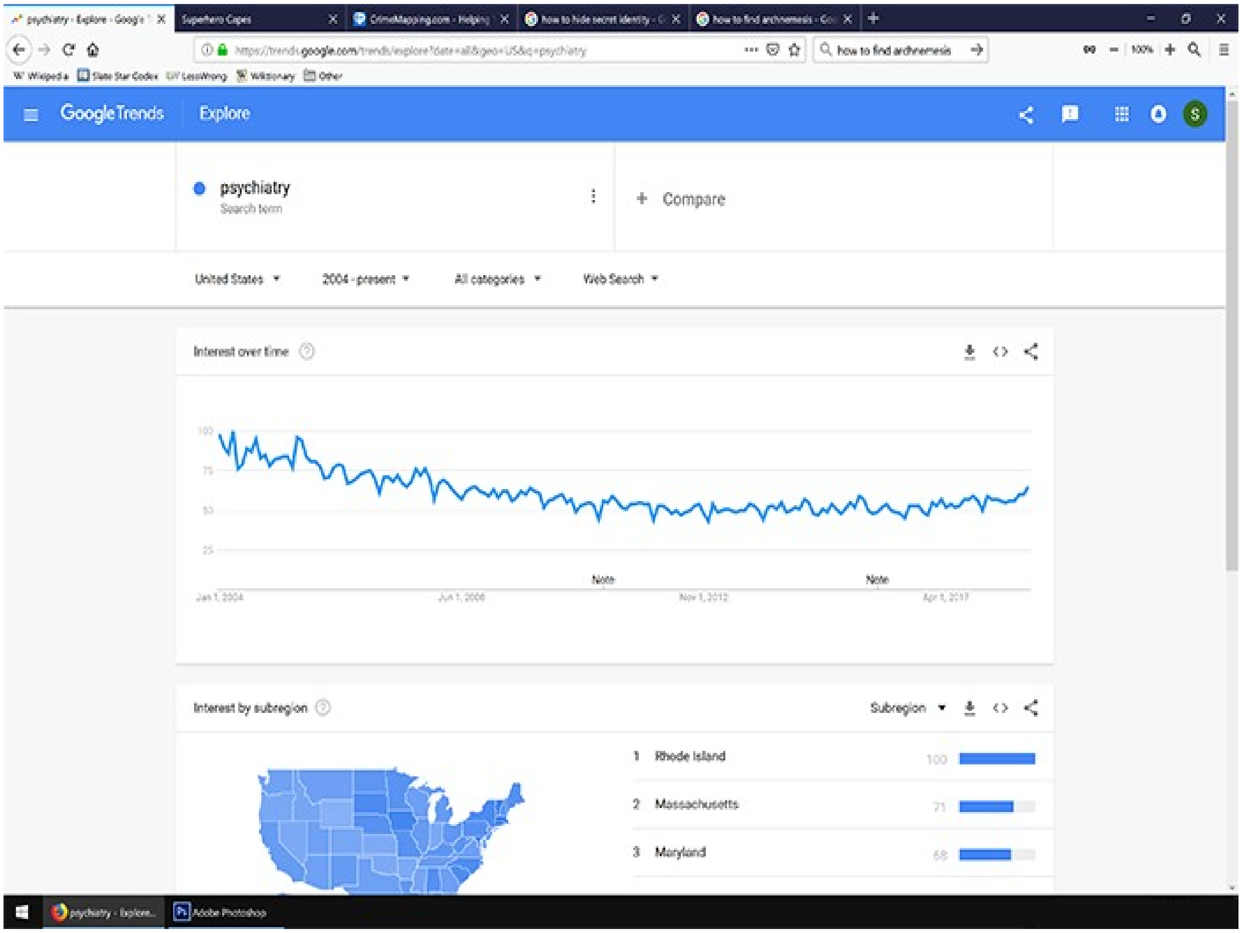
Working With Google Trends

[Epistemic status: low. You tell me if you think this works.]

Commenter no_bear_so_low has been doing some great work with Google Trends recently – see for example his Internet searches increasingly favour the left over the right wing of politics or Googling habits suggest we are getting a lot more anxious.

I wanted to try some similar things, and in the process I learned that this is hard. Existing sites on how to use Google Trends for research don't capture some of the things I learned, so I wanted to go over it here.

Suppose I want to measure the level of interest in "psychiatry" over the past few years:



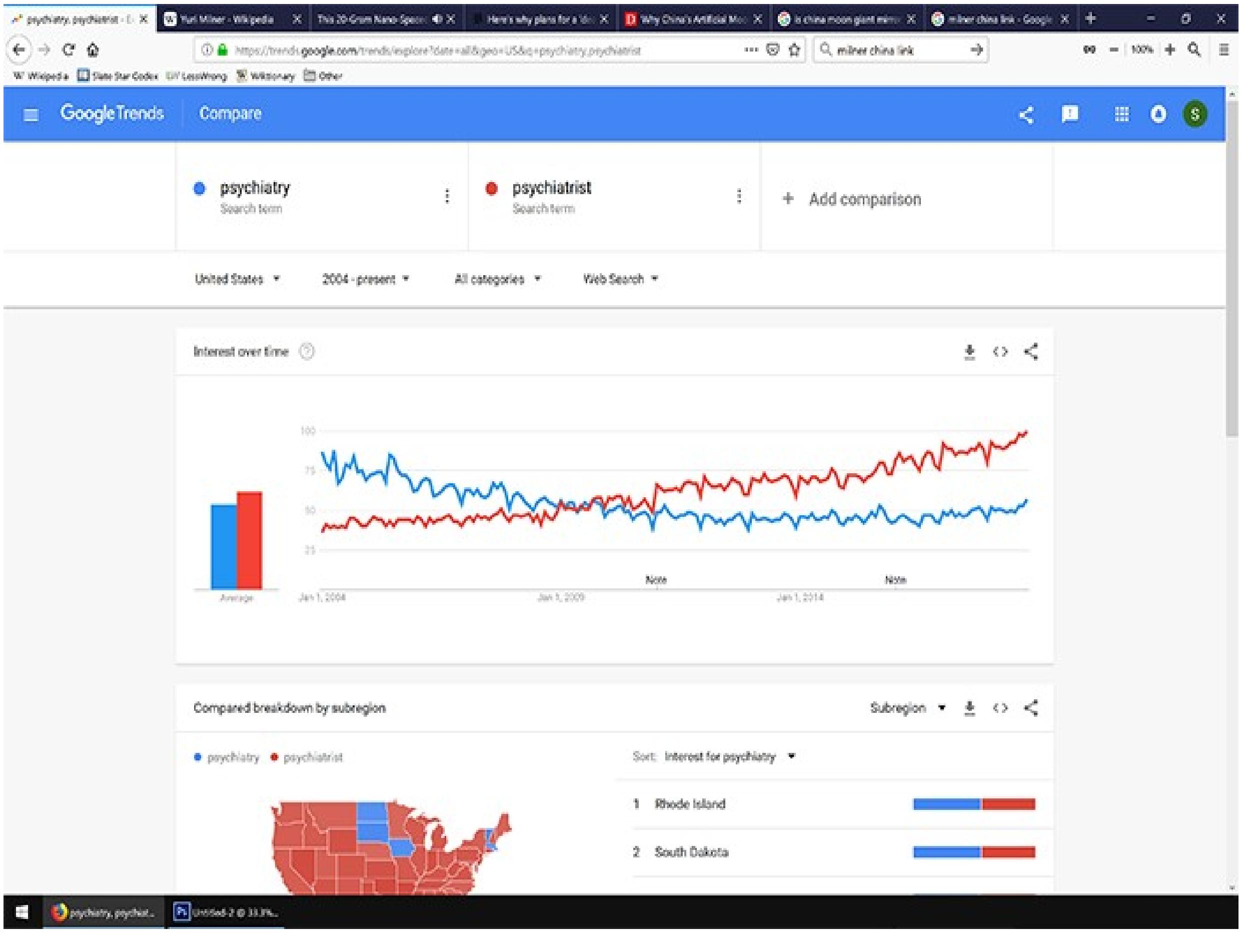
psychiatry - Explore - Google
Superhero Capes
CrimeMapping.com - Helping
how to hide secret identity
how to find archnemesis
https://trends.google.com/trends/explore?date=all&geo=US&q=psychiatry
how to find archnemesis
100%
Wikipedia
Slate Star Codex
LessWrong
Wiktionary
Other
Google Trends
Explore
Compare
Interest over time
100
75
50
25
Note
Note
Jan 1, 2004
Jun 1, 2008
Nov 1, 2012
Apr 1, 2017
Interest by subregion
Subregion
psychiatry - Explore...
Adobe Photoshop

psychiatry
Search term

United States · 2004 - present · All categories · Web Search

1	Rhode Island	100
2	Massachusetts	71
3	Maryland	68

Looks like interest is going down. But what if I search for “psychiatrist” instead?


Google Trends
Compare
Add comparison
Interest over time
100
75
50
25
Average
Jan 1, 2004
Jan 1, 2009
Jan 1, 2014
Note
Note
Compared breakdown by subregion
Subregion

psychiatry
Search term

psychiatrist
Search term

United States ▾ 2004 - present ▾ All categories ▾ Web Search ▾

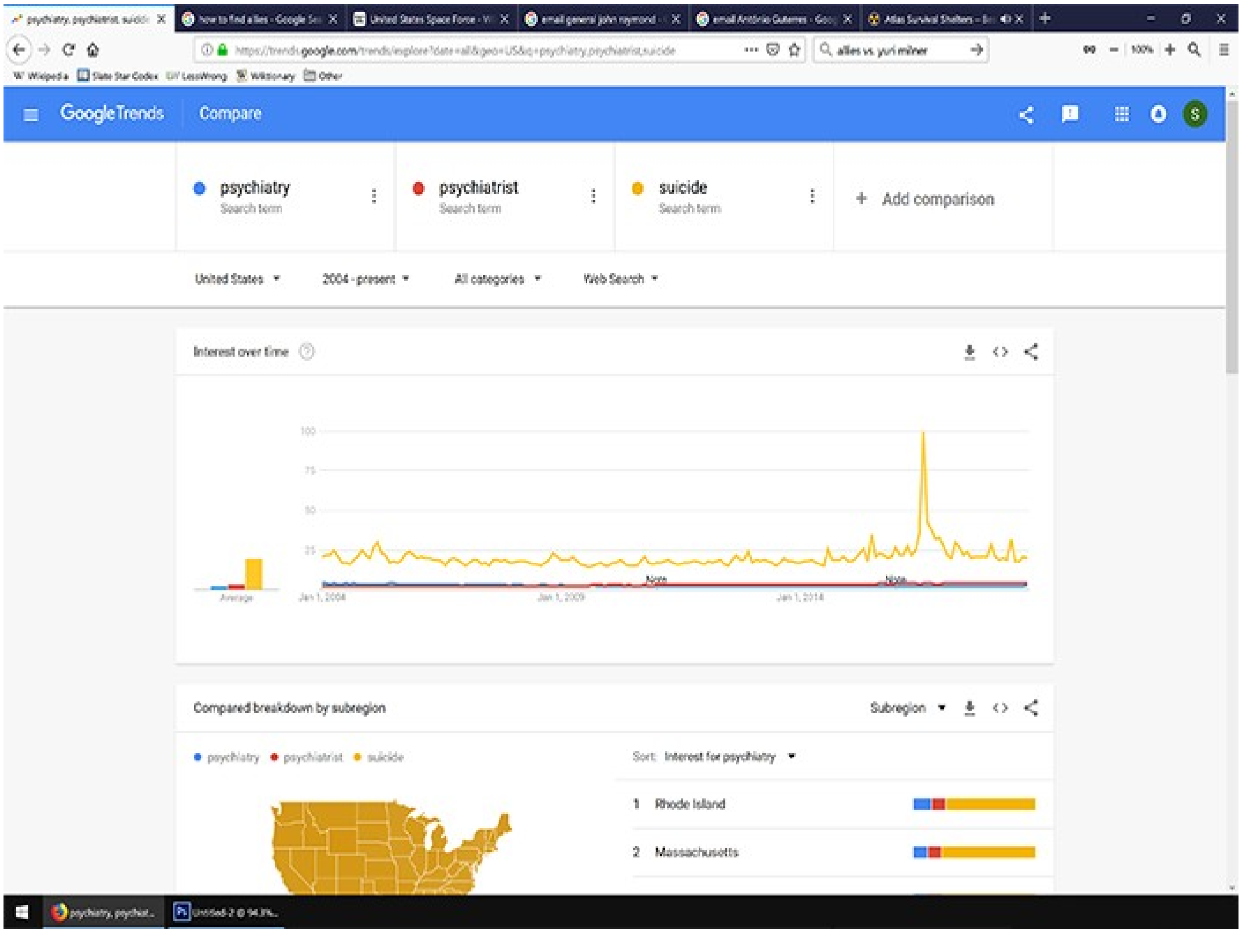
● psychiatry ● psychiatrist

Sort: Interest for psychiatry ▾

1 Rhode Island

2 South Dakota

Uh oh, now it looks like interest is going up. I guess what I’m really interested in is mental health more generally, what if I put in “suicide”?


Google Trends
Compare
psychiatry
Search term
psychiatrist
Search term
suicide
Search term
+ Add comparison
Interest over time
100
75
50
25
Average
Jan 1, 2004
Jan 1, 2009
Jan 1, 2014
Note
Note
Compared breakdown by subregion
Subregion
psychiatry
psychiatrist
suicide

https://trends.google.com/trends/explore?date=all&geo=US&q=psychiatry,psychiatrist,suicide

United States ▾ 2004 - present ▾ All categories ▾ Web Search ▾

Sort: Interest for psychiatry ▾

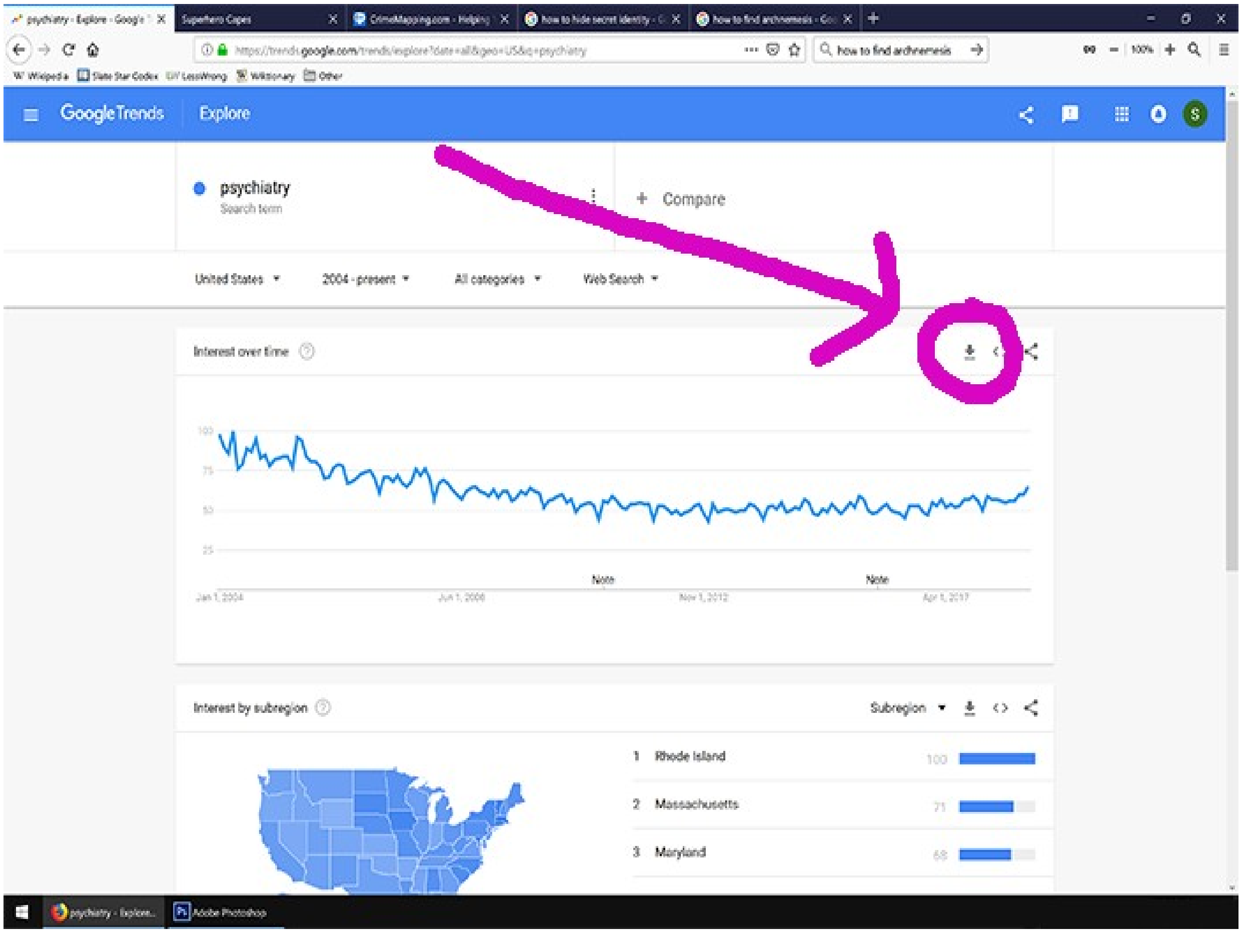
1 Rhode Island

2 Massachusetts

Now everything else is invisible, and the data are dominated by a spike in August 2016, which as far as I can tell is related to the release of the movie “Suicide Squad”.

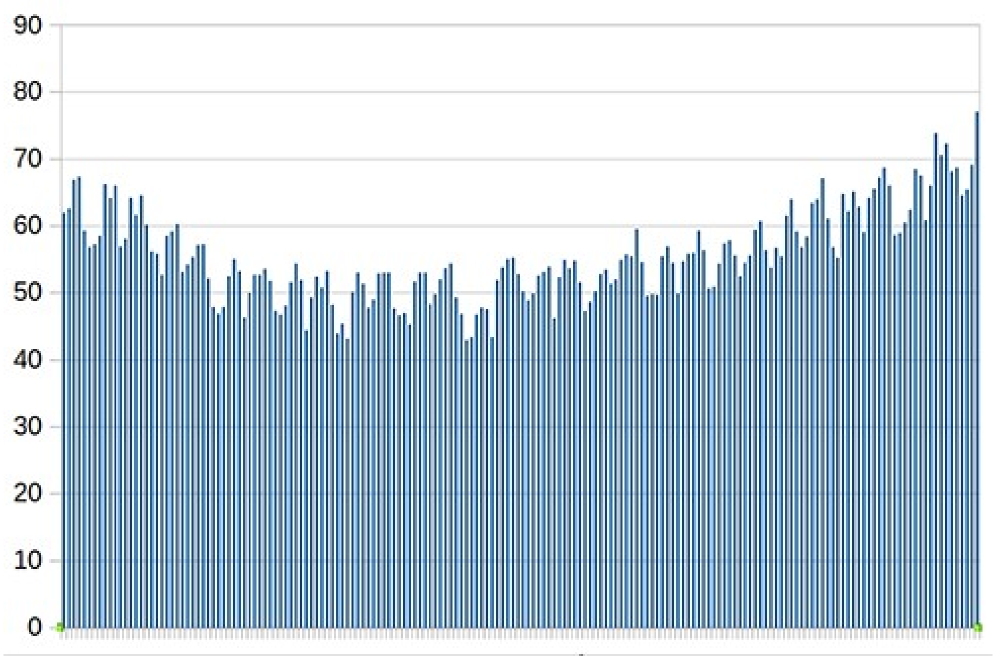
I could try other terms, like “depression” and “anxiety”, but no_bear’s data already tells us those two are moving in opposite directions. Also, depression has a spike in late 2008, which must be related to the stock market crash and people’s expectations of an economic depression. This doesn’t seem like a great way to figure out anything.

I wondered if averaging a bunch of things might take away some of the noise. I chose nine terms that seemed related to psychiatry in some way: psychiatry, psychiatrist, psychotherapy, mental illness, mental health, suicide, depression, antidepressants, and anxiety. Google won’t let you combine that many terms in a single query, but that’s okay – I don’t want to see them relative to one another, I just want to get standardized data on each. There’s a button to download any individual Google Trends query as a spreadsheet:

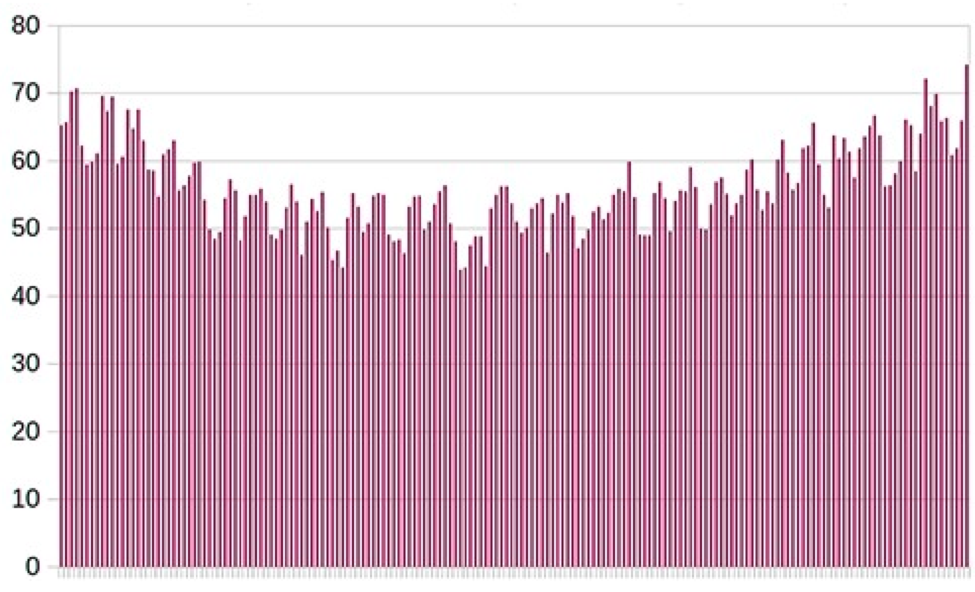

Google Trends
Explore
psychiatry
Search term
+ Compare
United States
2004 - present
All categories
Web Search
Interest over time
100
75
50
25
Note
Note
Jan 1, 2004
Jun 1, 2008
Nov 1, 2012
Apr 1, 2017
Interest by subregion
Subregion
1 Rhode Island 100
2 Massachusetts 71
3 Maryland 68

Don’t do this with a comparison graph like the last one. Otherwise you’ll get everything standardized relative to each other, and you won’t be able to compare different sets of multiple queries. Do it for a single trend line, do it a bunch of times, and then manually copy-paste all of them onto the same spreadsheet.

When I average all nine of those search terms, I get the following plot:



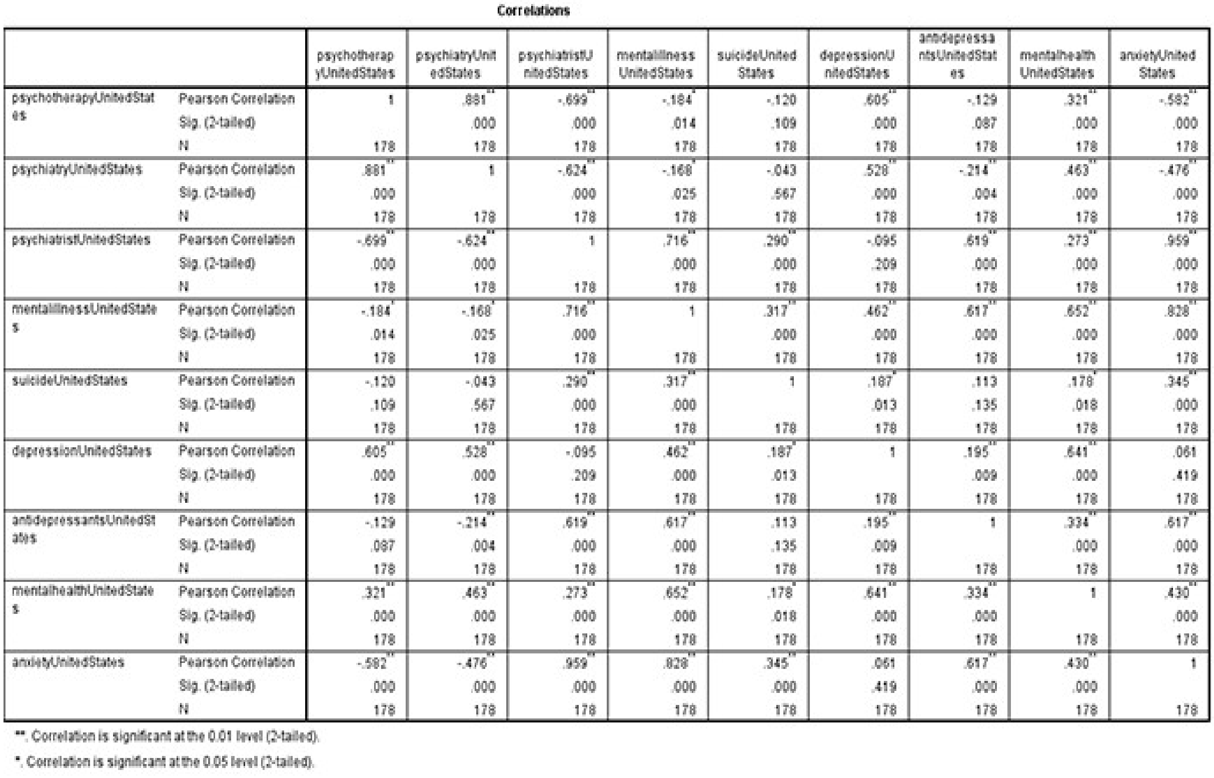
…which at least looks a lot simpler and more informative than most of the individual ones, and which doesn't have any big spikes for movie releases or anything. It also doesn't seem very sensitive to any small changes; when I take out the most dramatic outlier, anxiety (which has been increasing worringly fast; see no_bear_so_low's article above for more), it changes to this:



If you look really closely, you can see these are different graphs, but it sure isn’t a *big* difference.

These both have a kind of regular sawtooth pattern with a six month cycle; the peaks are always around April and October, and the valleys are always around July and December. These correspond nicely to school terms and vacations. While it’s possible that school is stressing people out and making them more worried about their mental health, you get these peaks in almost every field that anyone might study in school; I think it’s more likely that psychology students are looking these things up as part of their schoolwork.

Can we trust this? Can we do better than this? I imported these to SPSS and made a table of the correlations between each of the nine terms:



Correlations

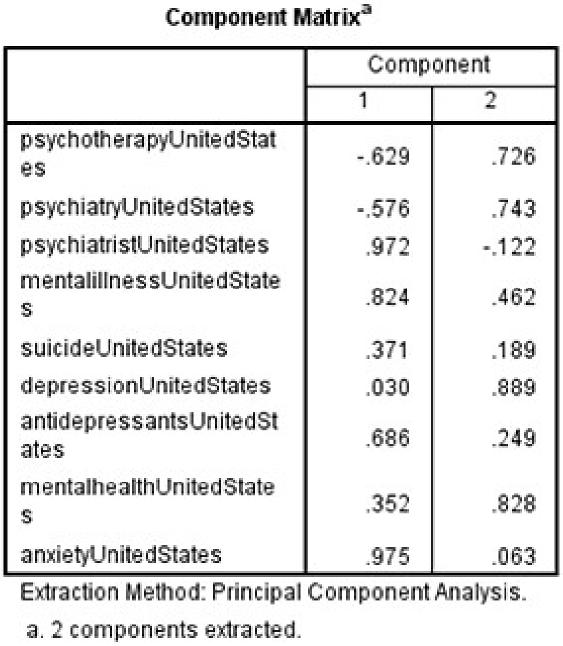
		psychotherapyUnitedStates	psychiatryUnitedStates	psychiatristUnitedStates	mentalillnessUnitedStates	suicideUnitedStates	depressionUnitedStates	antidepressantsUnitedStates	mentalhealthUnitedStates	anxietyUnitedStates
psychotherapyUnitedStates	Pearson Correlation	1	.881**	-.699**	-.184*	-.120	.605**	-.129	.321**	-.582**
	Sig. (2-tailed)		.000	.000	.014	.109	.000	.087	.000	.000
	N	178	178	178	178	178	178	178	178	178
psychiatryUnitedStates	Pearson Correlation	.881**	1	-.624**	-.168*	-.043	.528**	-.214**	.463**	-.476**
	Sig. (2-tailed)	.000		.000	.025	.567	.000	.004	.000	.000
	N	178	178	178	178	178	178	178	178	178
psychiatristUnitedStates	Pearson Correlation	-.699**	-.624**	1	.716**	.290**	-.095	.619**	.273**	.959**
	Sig. (2-tailed)	.000	.000		.000	.000	.209	.000	.000	.000
	N	178	178	178	178	178	178	178	178	178
mentalillnessUnitedStates	Pearson Correlation	-.184*	-.168*	.716**	1	.317**	.462**	.617**	.652**	.828**
	Sig. (2-tailed)	.014	.025	.000		.000	.000	.000	.000	.000
	N	178	178	178	178	178	178	178	178	178
suicideUnitedStates	Pearson Correlation	-.120	-.043	.290**	.317**	1	.187*	.113	.178*	.345**
	Sig. (2-tailed)	.109	.567	.000	.000		.013	.135	.018	.000
	N	178	178	178	178	178	178	178	178	178
depressionUnitedStates	Pearson Correlation	.605**	.528**	-.095	.462**	.187*	1	.195**	.641**	.061
	Sig. (2-tailed)	.000	.000	.209	.000	.013		.009	.000	.419
	N	178	178	178	178	178	178	178	178	178
antidepressantsUnitedStates	Pearson Correlation	-.129	-.214**	.619**	.617**	.113	.195**	1	.334**	.617**
	Sig. (2-tailed)	.087	.004	.000	.000	.135	.009		.000	.000
	N	178	178	178	178	178	178	178	178	178
mentalhealthUnitedStates	Pearson Correlation	.321**	.463**	.273**	.652**	.178*	.641**	.334**	1	.430**
	Sig. (2-tailed)	.000	.000	.000	.000	.018	.000	.000		.000
	N	178	178	178	178	178	178	178	178	178
anxietyUnitedStates	Pearson Correlation	-.582**	-.476**	.959**	.828**	.345**	.061	.617**	.430**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.419	.000	.000	
	N	178	178	178	178	178	178	178	178	178

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

A lot of things are negatively correlated. I don’t trust the coefficients and significance numbers much – I think they’re just temporal autocorrelation – but it’s kind of disturbing to see so many things going the opposite direction. Then again, we already knew this from the first psychiatrist-vs-psychiatry graph. I don’t know what I expected to see here and maybe this was a stupid thing to even investigate.

Maybe a factor analysis would be more illuminating. Here’s what I get:



Component Matrix[a]

	Component	
	1	2
psychotherapyUnitedStates	-.629	.726
psychiatryUnitedStates	-.576	.743
psychiatristUnitedStates	.972	-.122
mentalillnessUnitedStates	.824	.462
suicideUnitedStates	.371	.189
depressionUnitedStates	.030	.889
antidepressantsUnitedStates	.686	.249
mentalhealthUnitedStates	.352	.828
anxietyUnitedStates	.975	.063

Extraction Method: Principal Component Analysis.

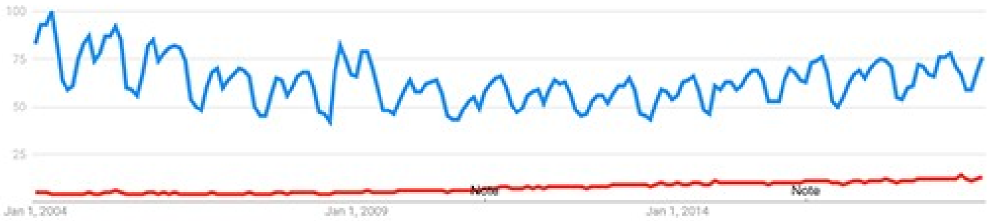
a. 2 components extracted.

This is pretty weird, but here’s how I interpret it. One factor is students with an intellectual interest in mental health – maybe they’re studying it, maybe they want a job in it. They Google things like “psychiatry” to learn about what psychiatry is and whether it’s a good career path. This has been going down over the past decade – I’ll talk about why this might be towards the end.

Another factor is people with mental health problems. They Google things like “psychiatrist” to try to find psychiatrists in their area, or “antidepressants” because they want to know if antidepressants will work for them. This has been going up over the past decade as more people worry about mental health problems.

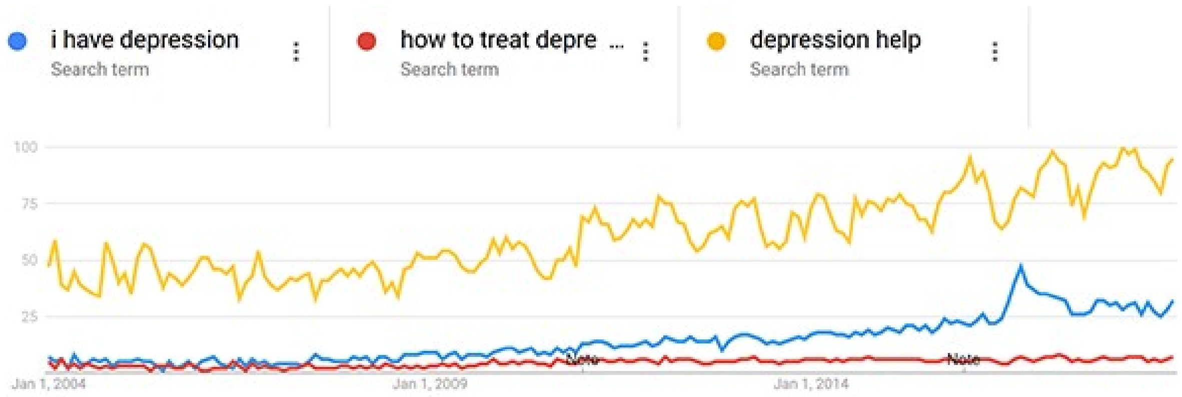
On the other hand, maybe we can’t do this, because of the temporal autocorrelation. Maybe this is just giving us a List Of Things That Are Going Up and another List Of Things That Are Going Down. I don’t know how to solve this problem.

But in favor of this interpretation, consider:

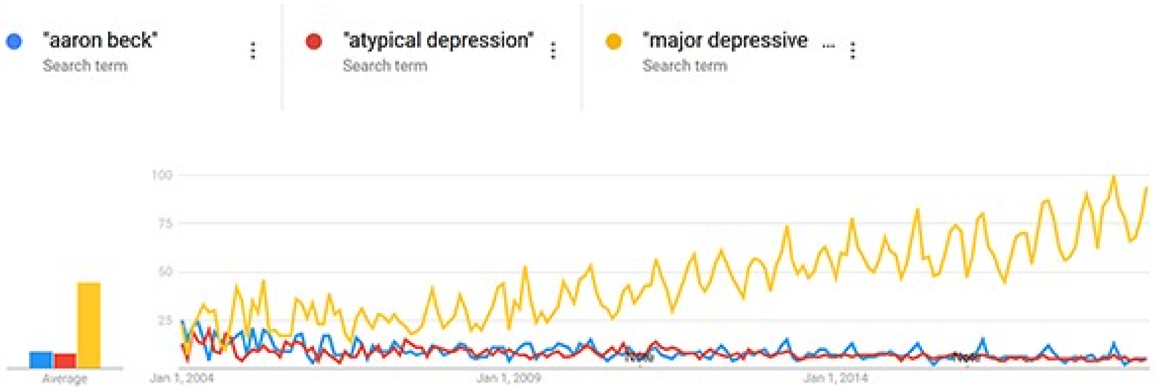


The blue line is “depression”, the red line is “depressed”. As predicted by the two-factor theory, “depression”, the very intellectual term that a student might type in to do their homework, is going down. “Depressed”, as in “help i am depressed”, is going up. “Depression” shows a strong school-related sawtooth pattern; “depressed” does not (it’s not just because of the different scales on this image; even a search for “depressed” on its own does not demonstrate the pattern).

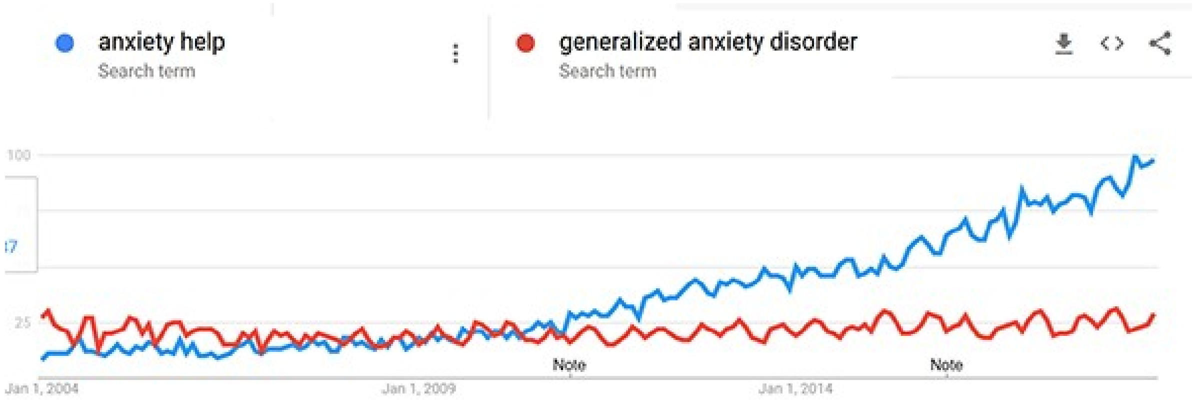
Also consistent with this theory, “how to treat depression”, “depression help”, and even “i have depression” are all going way up.



Contrary to my theory – “major depressive disorder” (an overly formal way of referring to depression, mostly of interest to psychologists and psychiatrists) is going up. But “Aaron Beck” (a famous depression researchers whose work is often taught in school) and “atypical depression” (a subtype of depression mostly of interest to psychologists and psychiatrists) are going down.

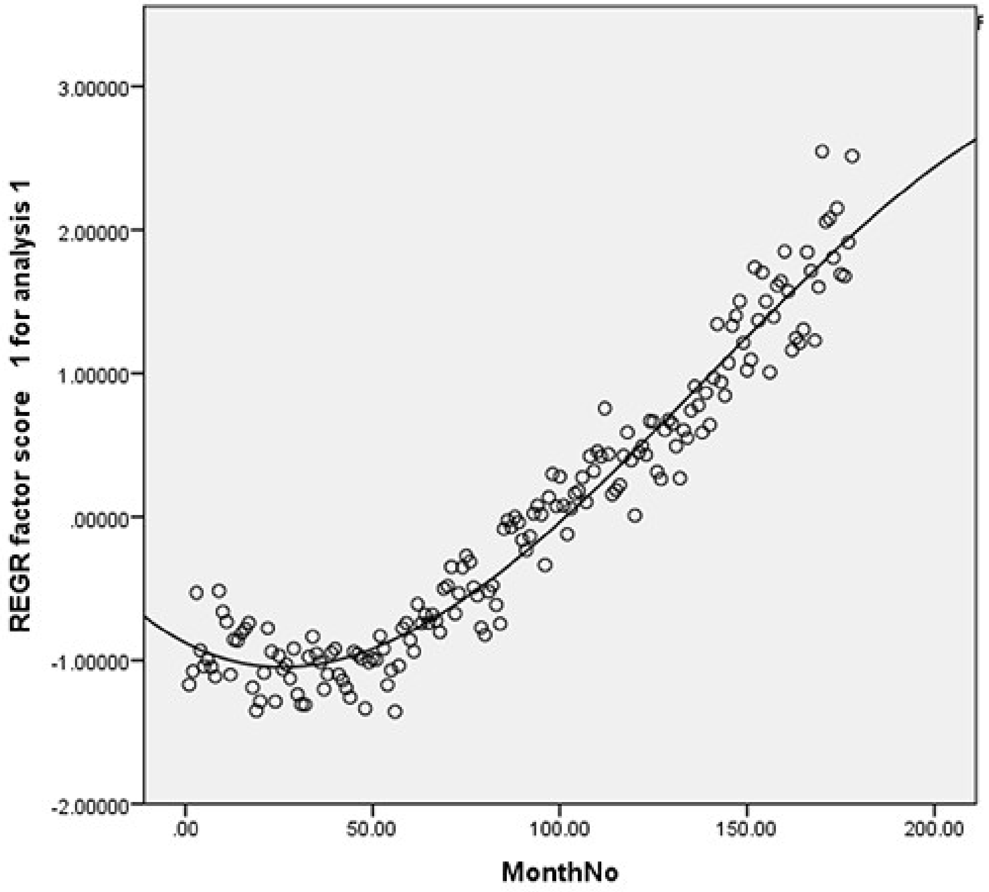


And compare “anxiety help” vs. “generalized anxiety disorder”.

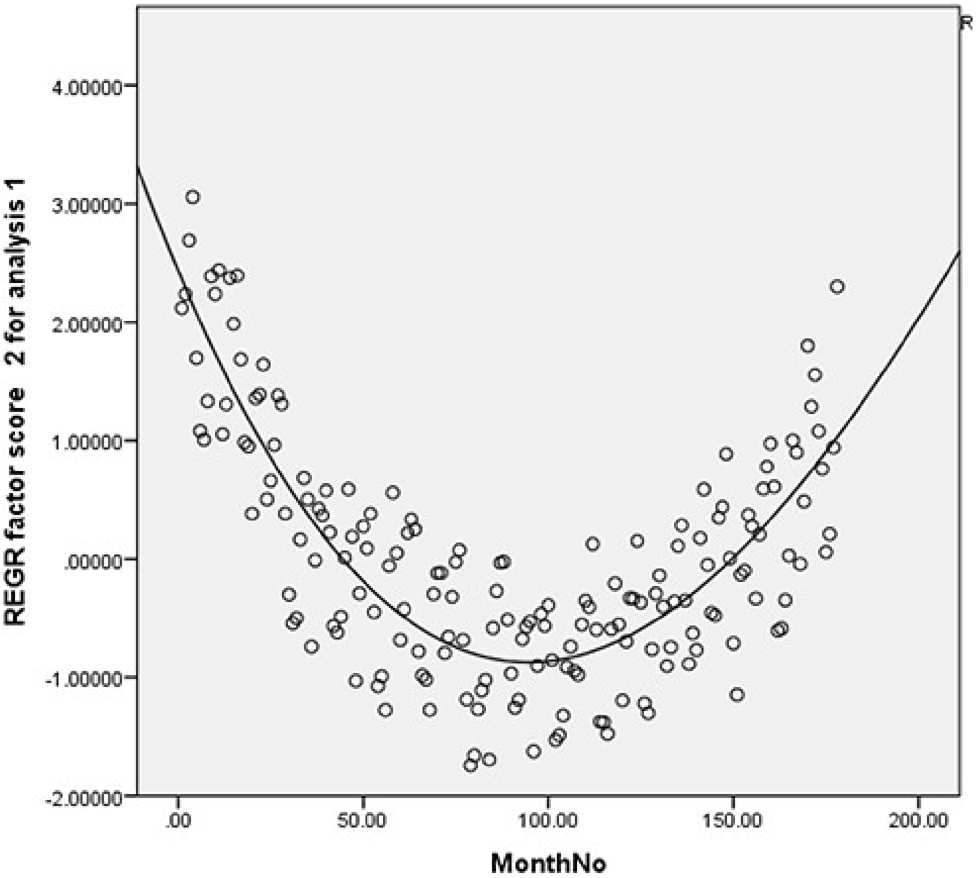


Although it’s not perfect, I think this explains a lot of these seemingly contradictory results.

Here is a graph just showing just the factor which I think corresponds to something like “people with mental health issues”:



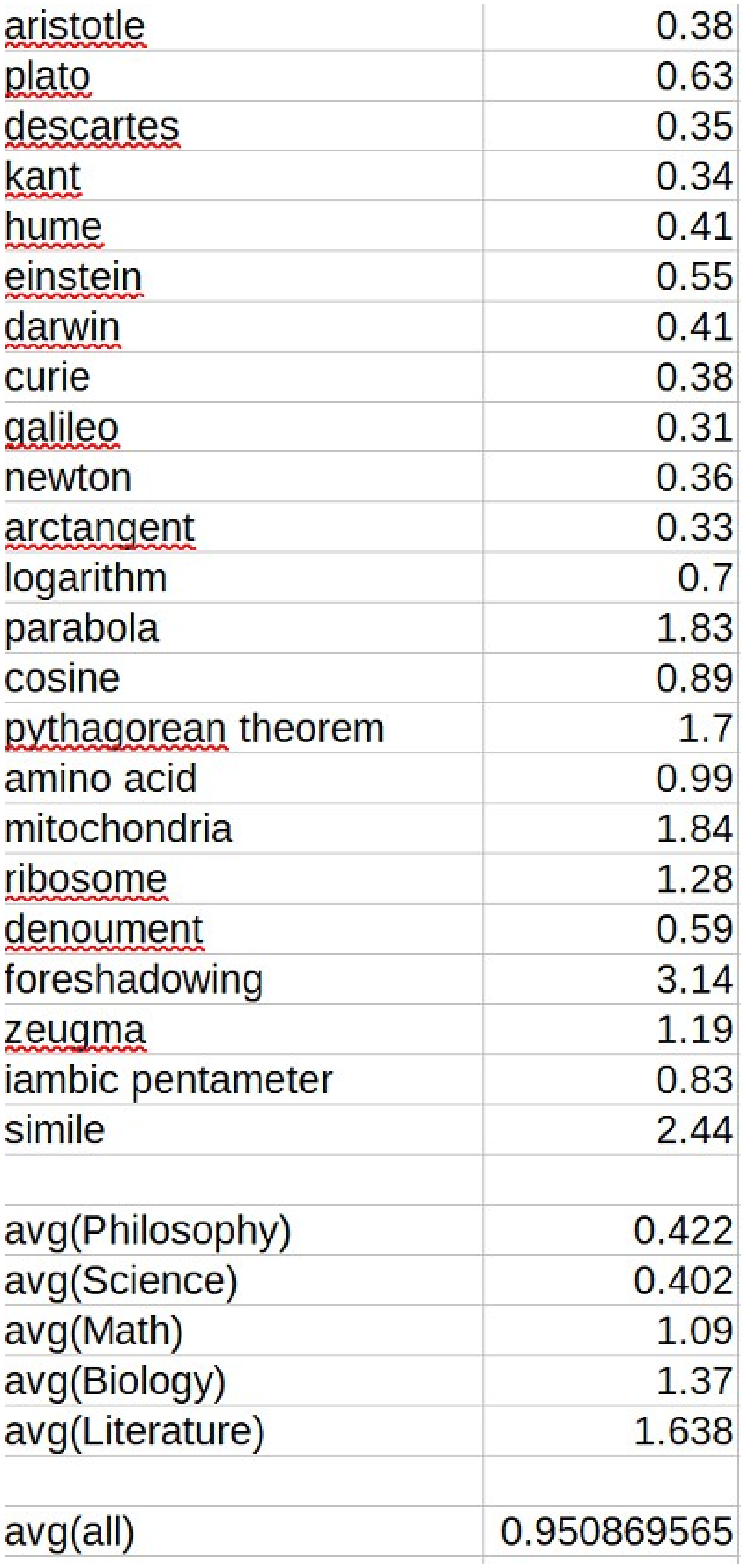
And here’s one just showing just the factor which I think corresponds to something like “students with an intellectual interest in psychiatry”:



I originally hoped this would explain no_bear_so_low’s main finding of anxiety (but not depression or other psychiatric conditions) increasing, on the grounds that “anxiety” searches are driven by anxious people, but “depression” searches (as opposed to “depressed” or “suicidal” or “sad”) searches are driven by students. This doesn’t work. There’s a slight trend in this direction, but not nearly enough to explain the disproportionate rise in anxiety.

Does the second factor indicate declining interest in psychiatry as a field? I originally thought it might not. 2004 was a long time ago. The 2004 internet, compared to the modern internet, was dominated by smart technophile early adapters and university students. So it could be that smart people remain interested in intellectual fields, but are gradually being diluted as a percent of the Internet’s user base. This could also explain the rise in Factor 1, if technologically-adept college-educated people have fewer psychiatric issues than other people.

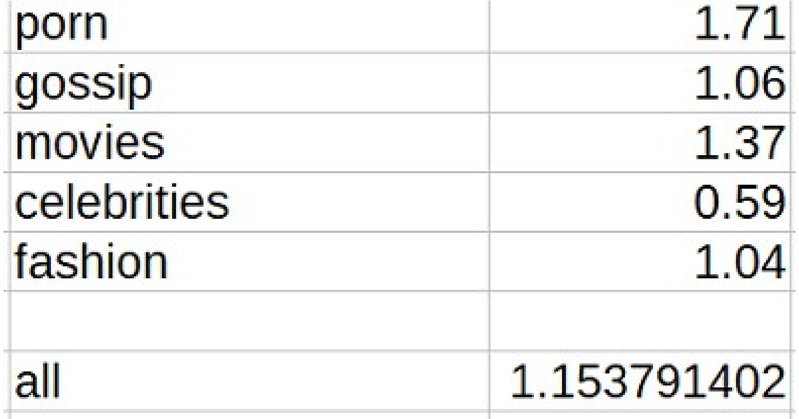
I don’t have a really principled way to test this, so I just threw in a whole bunch of intellectual terms that I expected intellectual and college-educated people to search for and which I wouldn’t have expected to change much over the past decade. I tried to do five terms from five different fields of study, though due to computer error only three of the biology terms ended up included and I didn’t want to repeat the whole thing. Here’s what I got:



aristotle	0.38
plato	0.63
descartes	0.35
kant	0.34
hume	0.41
einstein	0.55
darwin	0.41
curie	0.38
galileo	0.31
newton	0.36
arctangent	0.33
logarithm	0.7
parabola	1.83
cosine	0.89
pythagorean theorem	1.7
amino acid	0.99
mitochondria	1.84
ribosome	1.28
denoument	0.59
foreshadowing	3.14
zeugma	1.19
iambic pentameter	0.83
simile	2.44
avg(Philosophy)	0.422
avg(Science)	0.402
avg(Math)	1.09
avg(Biology)	1.37
avg(Literature)	1.638
avg(all)	0.950869565

The number is (popularity over last twelve months)/(popularity over first twelve months), approximately (2018 popularity)/(2004 popularity). Although there are a few dramatic trends – fewer people interested in philosophers, more people interested in the Pythagorean Theorem – overall there is no sign of an intellectual decline.

Likewise, for non-intellectual subjects beloved of the unwashed masses:



porn	1.71
gossip	1.06
movies	1.37
celebrities	0.59
fashion	1.04
all	1.153791402

Again, a few things are going up, a few things are going down, but there is little sign of a general increase.

So here are my tips for researching things with Google Trends:

1. Individual keywords may be relevant for very some specific searches (eg “obama”), but can be dangerous when looking at broader trends. For example, “psychiatry” and “psychiatrist” have gone in opposite directions since 2004.

2. Averaging large numbers of related keywords might help with this problem, but can cause issues of its own.

3. Factor analysis might help with averaging large numbers of related keywords, especially in generating new theories

and intuitions. I don’t know whether temporal autocorrelation makes it useless or dangerous.

4. If you see a sawtooth pattern with troughs during school vacations, this can help you figure out whether you’re looking at students or at “organic” results. The higher the peaks and troughs of the sawtooth, the more your data are skewed towards student searches.

5. Surprisingly, there doesn’t seem to be a clear need to adjust your data for a general intellectual decline in the Internet.

My California Ballot

These are my preliminary choices for California elected positions and ballot initiatives. Some of them are based on Ozy's recommendations and the Berkeley EA and rationalist community's recommendations. I agree with the latter's note that because California ballot propositions are weird superlaws that permanently overrule the legislature unless repealed by voters, in general we should be very cautious about them (though some of them were recommended by the legislature itself, since for complicated reasons it needs voter support to do certain things).

I'm giving first-level justifications for my votes (ie "I support this person because she wants higher taxes") but not always second-level justifications ("here's why higher taxes are good"). You can usually find discussion of these on other blog posts.

Governor of California is the big one. Democrat Gavin Newsom is a former successful businessman, mayor of San Francisco, and lieutenant governor of California (also second cousin of musician Joanna Newsom). He has stated that if elected, he will let people call him "the Gavinator". Republican John Cox is a former successful businessman, best known for sponsoring a ballot initiative to make legislators wear the logos of their top 10 donors on the State Assembly floor, "much like NASCAR drivers". He also has a fascinating plan to reform politics from the ground up with a 12,000 (!) member legislature. I don't really like Newsom – he led a movement called "Care Not Cash" to restrict giving money to the homeless, and supposedly opposed anti-gay Proposition 8 so incompetently that his statements may have *increased* support for the measure. He also had an affair with his campaign manager's wife in a scandal that seemed unusually scummy even for a politician. I like John Cox as a person, but he doesn’t seem to have any relevant governing experience. And he was anti-Trump until Trump became popular among Republicans, then about-faced and decided Trump was his new best friend, and now he’s basically just a Trumpist. I am going with Newsom; God help me, God help California.

The Lieutenant Governor is a sinecure interesting only insofar as its holders often become governors themselves, kind of like the US Vice President. Democrat Ed Hernandez is a optometrist who is very involved in health policy. He makes a big deal about Fighting Pharma Companies, and a somewhat less big deal about having taken hundreds of thousands of dollars from pharma companies and hundreds of thousands more from insurance. His opponent Eleni Kounalakis is also a Democrat, because California, and was previously Obama’s ambassador to Hungary. I am not sure exactly what an ambassador does, but if part of her job was to keep a watch on Hungary I feel like this is a strike against her. On the other hand, she has been exposed to superior Habsburg institutions. And she and her husband fund Hellenic Studies at state universities, raising the prospect that maybe she will transform California into some kind of Hellenic-Habsburg fusion superstate. More seriously, Hernandez has done some really good work in health care, including better drug transparency and easier prior auths (non-doctors will not understand this term; doctors will hear it and rush to vote for him immediately). But he seems anti-Medicare-for-all, and kind of in the pocket of insurance companies. He also has been a big proponent of race-based admissions in California colleges. Kounalakis is pro-Medicare-for-all, has Obama’s endorsement, and is really strong on housing policy. Advantage Kounalakis.

The State Controller was an important position back in the time of Mustapha Mond, but nowadays it just audits finances. Betty Yee is the incumbent, and California's finances have been doing surprisingly well lately, and she sounds very smart and says the right things about preparing for inevitable downturns and promoting fiscal responsibility. Konstantinos Roditis is running mostly on fighting the gas tax and defunding high-speed rail. While high-speed rail will inevitably be a boondoggle that we gnash our teeth over for decades to come, I'm concerned about his obsession with cutting taxes, the gas tax seems like a particularly bad tax to make a stand over, and I'm concerned about replacing a highly competent person who's kept us solvent with a random low-tax crusader. Advantage Yee.

The State Board of Equalization was an important position back in the time of Harrison Bergeron, but nowadays it just assesses certain taxes. Democrat Malia Cohen boasts that she is strong on climate change, which is relevant because the Board handles gas taxes; she also boasts that she is strong on LGBTQ and abortion rights, which is relevant because California. Republican Mark Burns is centering his whole campaign around strengthening Prop 13, one of the worst laws in California. Both candidates' names sound like minor Unsong characters, but I choose Malia, no contest.

The Secretary of State handles elections and voter registration. Republican candidate Mark Meuser takes the typical Republican position that we need to tighten standards to fight voter fraud. Democratic candidate Alex Padilla takes the typical Democratic position that there is no such thing as voter fraud and worrying about it is discriminatory. Mark Meuser protests that there is absolutely voter fraud, and highlights how a California man successfully registered his dog to vote. Alex Padilla says that not wanting dogs to vote is discriminatory, and he will order all poll workers to accept wet, slightly-chewed-up ballots. Mark Meuser protests that the man who registered his dog to vote had registered his *other* dog to vote twenty years earlier and still nobody has taken any action. Alex Padilla boasts of endorsements from some of California’s top politicians, like Mayor Max of Idyllwild. I respect Mr. Meuser’s strong anti-dog stance, but I am concerned, based on his name, that he may secretly be a cat. Advantage Padilla.

The State Treasurer helps manage investments and finances. Republican Greg Conlon is a former accountant who is running on a platform of pointing out that the pension crisis is a giant ticking time bomb that is about to explode and consume us all. Democrat Fiona Ma is running on a platform of investing in shiny popular causes that make us feel good about ourselves, and covering the time-bomb with a pile of leaves so its ominous ticking noises don’t keep us up at night. Advantage Conlon

The Attorney General leads a vast army of attorneys into battle against the attorneys of foreign lands. Democrat Xavier Becerra is the incumbent, and has an unblemished record of taking the most liberal possible position on everything. Republican Steven Bailey is a judge who is Strong On Crime and Very Tough, and so serious that he can call the lawsuit over Trump’s wall “borderline frivolous” without adding even so much as a “pun not intended”. While neither candidate really talks much about mass incarceration, it’s pretty clear that Bailey gets an erection every time he thinks about it, and Becerra plans to ignore it while demonstrating conspicuous outrage about Trump’s latest tweet. Advantage Becerra, I guess.

Reason calls the Insurance Commissioner race The Most Important Election You’ve Never Heard About. Apparently California has a system where a Commissioner has to approve all new insurance products. The Commissioner sometimes uses this power in weird ways – for example, one official refused to approve insurances' plans unless they divested from investments in coal. Steve Poizner is a Bay Area tech entrepreneur and used to be a Republican but has since switched to being an independent – if he wins, he will be the first independent elected to a major position. Ricardo Lara is "one of the [state] Senate's most liberal members" and running as a Democrat. I don't have the background to judge the technicalities of insurance policy. But Poizer is a former incumbent who by all accounts did pretty well, and I think it's great that a non-partisan person might win a major position. Advantage Poizner.

Everyone dislikes Dianne Feinstein for a different reason. Some people dislike her support for the Iraq War and the Drug War. Other people dislike her attempted constitutional amendment to ban flag burning. Other people dislike her support for the PATRIOT Act, mass surveillance, prosecuting Edward Snowden, and banning strong encryption. Other people dislike her opposition to single-payer healthcare. Other people dislike her support of arms sales to Saudi Arabia. Other people dislike her consistently bad positions on free speech, from supporting universities turning away conservative speakers, to supporting universities' right to expel students who criticize Israel, to sponsoring an act that classifies various nonviolent forms of animal activism as potential ecoterrorism. I hear some people are Republicans, and have an entirely *different* set of reasons to dislike her. In these divided times, there isn't much that can make people step across the aisle and shake hands with people on the other side of hot-button issues – but all of us – rich and poor, white and black, urban and rural – can get behind disliking Dianne Feinstein. I have no real opinion on the other Democrat, Kevin De Leon, so I'm voting for him.

For House of Representatives, it’s the Democratic candidate vs. the Green Party candidate, because California. The Green Party candidate, Laura Wells, looks pretty great – she seems like a smart person who helped institute Instant Runoff Voting in Oakland, giving the city one of the few sane voting systems in the US. Barbara Lee is not nearly as awesome, but she did cast the only vote against authorizing military force in the War on Terror in Congress. This is a sufficiently impressive act that I will vote for her anyway even though I think she is worse on a lot of other things; I would rather reward Congresspeople for taking high-variance unpopular stands that turn out to be right than get someone who will be a 5% better technocrat. Sorry, Laura. If you run for something else later I will support you for that.

There’s also a race for “15th District”, which sounds like one of those groups that sends people to the Hunger Games. This is Buffy Wicks vs. Jovanka Beckles. Beckles is a socialist who is running on the presumption that restricting housing supply will make it more affordable, and on accusing everyone else of being corporate profit establishment capitalist shills. Wicks is endorsed by Barack Obama and East Bay for All, and supports building more housing. I also support her vampire-slaying work (I’m not just going off her name – she even looks similar). Advantage Wicks.

There are ten judges up for re-election. Most sources recommend that people vote ‘yes’ for all of them to preserve judicial independence, so I did that.

The State Superintendent of Public Instruction runs the education system. For an enforced non-partisan office, this race is pretty heated. It has $50 million in spending, and when I Google either candidate’s name I get attack ads trying to tell me why not to vote for him. Both candidates support increased funding and universal pre-K, because

California. Marshall Tuck seems a little more pro-charter school, and is against a move where money that was earmarked for high-needs students was instead used to raise teachers' salaries. Thurmond has the support of teachers' unions, which I guess makes sense considering. I am usually pro-charter-school and pro-supporting-high-needs-students, but Thurmond supports later start times for middle and high school, and this is pretty important to me. But Tuck has more experience and the support of Obama's education secretary, so I guess I'll go with him.

Proposition 1 authorizes bonds to fund affordable housing. The California legislature requested this, newspapers unanimously support it, and affordable housing seems good. Yes.

Proposition 2 authorizes redirecting money earmarked for mental health care to housing mentally ill people. The legislature supports it, housing the mentally ill is probably better for them than whatever else this money was going to do, and I generally support giving the government more flexibility on how to spend its money. Yes.

Proposition 3 authorizes bonds to fund water supply projects. The legislature did not request this. Environmental groups are skeptical. It is sponsored by companies that would benefit from it. Newspapers are unanimously opposed. No.

Proposition 4 authorizes bonds to build children's hospitals. The legislature did not request this. There are always propositions to fund children's hospitals and it seems kind of unfair since you're not allowed to be against healthy children. There is no real evidence of a children's health crisis. No.

Proposition 5 expands a previous proposition that says people’s taxes on their house cannot increase too quickly while they live in it. This created a problem: imagine that you are a 70 year old whose children have moved away and whose spouse has died. You live in the big five-bedroom house you raised your family in forty years ago, but you want to move somewhere smaller and closer to caregivers and more accessible given whatever health problems you might have. Since buying a new house resets your property taxes, instead of having your taxes pegged to what your current house cost forty years ago, you would have to pay the value of your new house today. This would mean order-of-magnitude-increases in taxes. So you just never move, and keep occupying a five-bedroom house as a single person, whether you like the house or not. To solve this problem, the 1980s measure Proposition 60 allowed elderly homeowners move once, to a smaller house, in the same county, and keep their previous property tax assessment. The current Proposition 5 would say they can move as many times as they want, to any sized house, in any county, and keep their tax assessment. It also extends this to disabled homeowners, who might also need a one-story house or house with more accommodations after they become disabled, and who might also be financially unable to manage the move if it would raise their property taxes. Overall I agree with the idea that taxes should not distort the market and force you to live in a house that is too big for you or doesn’t accommodate your disability; given that there is a housing crisis, forcing people to live in houses that are too big for them seems especially stupid. The California Chamber of Commerce also agrees and says this would free up 100,000 additional houses a year to help ease the housing crisis. However, all my friends, including the ones who are supposedly YIMBY, are very angry about this. I am having difficulty figuring out why in between all the use of “I’ve got mine” and “this is about hating poor people”, but I think it has something to do with anger at the idea that some people can keep their property taxes low forever. I agree this is bad, but it is the existing system; this new proposition just stops favoring people who keep their current home over people who, for reasons of age or disability, have to move houses. I understand anger about low property taxes, but making it harder for older or disabled people to move into houses that meet their needs feels like a particularly cold-hearted way of expressing that. It sounds like the main area of disagreement is whether, if this proposition fails, elderly and disabled people will move anyway but pay higher taxes, or whether they will just stay around to avoid the tax increase. Studies suggest that the lock-in effect from Proposition 13 is real, but don't let me compare it to anything else or say how much of it has stuck around after Proposition 60. Since this is a matter of comparing coefficients, and I don't know the coefficients, I am tempted to just look at endorsements. Bloomberg seems in favor, Reason seems ambivalent, and YIMBY Action says no. I think the amount of time it would take for me to feel comfortable with my opinion on this is higher than any possible benefit of getting it right, so I will abstain.

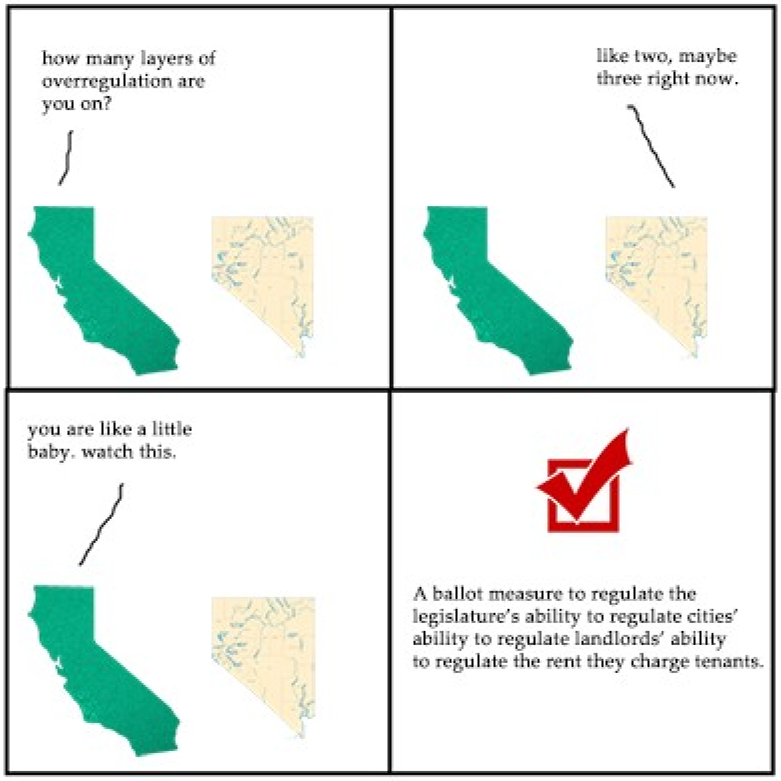
Proposition 6 repeals the gas tax. The gas tax is a reasonable way to internalize the cost of road infrastructure to drivers. It also helps fight climate change and raises much-needed funds for the state. While it is regressive compared to an optimal tax policy, it's probably better than whatever else would replace it. No.

Proposition 7 allows the legislature to change Daylight Savings Time in ways consistent with federal law. It's necessary because there was a previous ballot proposition about Daylight Savings Time, and now the legislature can't change it without another ballot proposition. There is actually some evidence that Daylight Savings Time transitions cause car accidents and interfere with circadian rhythms. There’s also the status quo bias – if we didn’t already have a rule that we switched time by an hour in the spring and autumn, would anyone be proposing it? I’m open to arguments that Daylight Savings transitions might be useful somehow, but I don’t want them as a superlaw that the legislature can never change. Yes.

Proposition 8 says dialysis companies can only charge 115% of the cost of treatment (the extra 15% would presumably be split among administrative expenses and profits). I tend to err on the side of not increasing the massive overregulation of health care. Even if I didn’t, I would like the California legislature to be able to decide whether to regulate health care or not, rather than pass it as a superlaw that can never be repealed no matter how badly it goes. No.

There is no Proposition 9, because California.

Proposition 10 repeals an existing statewide law saying that cities cannot institute rent control. I am tempted to oppose this measure, since in the IGM weighted poll, 95% of economists believe rent control is bad for the poor and reduces affordable housing, compared to only 1% who say it is good. And if rent control were legal, I have no doubt cities would try to decree that no house can ever cost more than $1, declare victory over evil greedy capitalists, and then get all confused a year later when the state had became an uninhabitable wasteland, because California. On the other hand, I also don’t like state governments telling city governments what to do, and the legislature saying “no community can institute this law which many of you want but we don’t like” seems like a violation of Archipelago principles and the idea of “laboratories of democracy”. I worry that standing up for this principle will end up costing the poor, but if you don’t have principles when those principles lead to something bad, then you just don’t have principles. On the other hand, is it really in keeping with localism for the California electorate to pass a ballot proposition repealing a law of the California legislature? Are we past the point where any principles can make sense of this at all?


how many layers of overregulation are you on?
like two, maybe three right now.
you are like a little baby. watch this.
A ballot measure to regulate the legislature’s ability to regulate cities’ ability to regulate landlords’ ability to regulate the rent they charge tenants.

This is another one where I don’t feel like I can do anything other than abstain.

Proposition 11 retroactively makes it legal to require ambulance operators to stay on-call during breaks. It is sponsored by an ambulance company who was sued for illegally requiring ambulance operators stay on-call during breaks. Instead of just paying damages like a normal company, they decided to put a proposition on the California ballot saying they were retroactively right all along. Their argument is that ambulance workers already get “breaks” in the sense of there being a lot of downtime when there are no emergencies, so normal break law requiring additional breaks shouldn’t apply to them. Also, it’s really hard to come up with a call schedule for emergency services at the best of times and regulations make that even more difficult. The mistake theorist in me is happy to admit that some of these are good points. But the conflict theorist in me would be much more comfortable with this argument coming from anybody other than an ambulance company who was already being sued for denying workers breaks. The correct sequence is “pass law saying you can do something, *then* do it”. Also, if the California legislature wants to give ambulance companies an exemption from normal break law, it can do that without a superlaw that can never be repealed. No.

Proposition 12 amends a previous proposition to clarify standards for factory farms in California. It mandates cage-free confinement for some animals, and increases minimum cage sizes for others. This is a very simple measure to prevent a tiny bit of preventable animal cruelty and make factory farming a little less bad. It is strongly supported by everyone in the effective altruist movement and most animal rights groups. It is opposed by PETA (who believe that animals should not be farmed at all) and by economists who point out that raising the minimum cage destroys jobs for low-cage workers. That part about the economists was a joke. There is no good reason to oppose this measure. Strong yes. If you only take my advice on one measure on the California ballot, let it be supporting this one.

Also, there was an old socialist talking point that all the promises of reform within capitalism only amount to “bigger cages, longer chains”. If Proposition 12 passes, capitalist reformers will have fulfilled half of their promises. How many other systems can say the same?

Preschool: I Was Wrong

Kelsey Piper has written an article for Vox: Early Childhood Education Yields Big Benefits – Just Not The Ones You Think.

I had previously followed various studies that showed that preschool does not increase academic skill, academic achievement, or IQ, and concluded that it was useless. In fact, this had become a rallying point of movement for evidence-based social interventions; the continuing popular support for preschool proved that people were morons who didn’t care about science. I don’t think I ever said this aloud, but I believed it in my heart.

I talked to Kelsey about some of the research for her article, and independently came to the same conclusion: despite the earlier studies of achievement being accurate, preschools (including the much-maligned Head Start) do seem to help children in subtler ways that only show up years later. Children who have been to preschool seem to stay in school longer, get better jobs, commit less crime, and require less welfare. The thing most of the early studies were looking for – academic ability – is one of the only things it doesn’t affect.

This suggests that preschool is beneficial not because of the curriculum or because of “teaching young brains how to learn” or anything like that, but for purely social reasons. Kelsey reviews some evidence that it might improve child health, but this doesn’t seem to be the biggest part of the effect. Instead, she thinks that it frees low-income parents from childcare duties, lets them get better jobs (or in the case of mothers, sometimes lets them get a job at all), and improves parents’ human capital, with all the relevant follow-on effects. More speculatively, if the home environment is unusually bad, it gives the child a little while outside the home environment, and socializes them into a "normal" way of life. I'll discuss a slightly more fleshed-out model of this in an upcoming post.

My only caveat in agreeing with this perspective is that Chetty finds the same effect (no academic gains, but large life-outcome gains years later) from children having good rather than bad elementary school teachers. This doesn't make sense in the context of freeing up parents' time to get better jobs, or of getting children out of a bad home environment. It might make sense in terms of socializing them, though I would hate to have to sketch out a model of how that works. But since the teacher data and the Head Start data agree, that gives me more reason to think both are right.

I can't remember ever making a post about how Head Start was useless, but I definitely thought that, and to learn otherwise is a big update for me. I've written before about how when you make an update of that scale, it's important to publicly admit error before going on to justify yourself or say why you should be excused as basically right in principle or whatever, so let me say it: **I was wrong about Head Start**.

That having been said, on to the self-justifications and excuses!

1) Head Start seems to work for reasons unrelated to the ones that made people want to do it. Those people were still wrong, and this is still a good example of policy effects being difficult to predict. It seems to have succeeded by coincidence, not because "early childhood education" is a good idea.

2) This probably strengthens rather than weakens the Caplanian case against education, since the studies find that the educational parts of preschool are not useful, and better teachers and curricula do not affect the benefits.

3) This strengthens rather than weakens the case that academic achievement is related primarily to IQ, and that IQ is primarily genetic and difficult to change. An intervention targeted at academic achievement and IQ manages to change everything else *except* those variables, which remain stubbornly the same. Studies consistently find that IQ is only responsible for about 25% of life outcomes, suggesting that education works on the other 75%.

But on a broader scale, this does lower my confidence in biodeterminism. Preschool is a shared environmental effect; your parents have a big effect on whether or not you go to preschool. Why doesn't this shared environmental effect show up in studies, which generally find no shared environmental effect matters?

This is the same problem raised by Ozy's post on lead. We know lead is important. We know it can damage your life outcomes. But we also know lead is related to the shared environment. And we also know studies keep finding the shared environment doesn't matter. Some studies find the shared environment matters a little, when you make extra-double sure to have very high income inequality in your sample. But other studies find that it doesn't, and almost all of them find that it doesn't matter much at the still-high levels of income inequality you get by recruiting a convenience sample. How can this be? We have two really excellent and well-replicated scientific literatures, each proving opposite things. What now?

All I can think of is that maybe shared environment *can* matter, but is so small in the grand scheme of things that it's below the threshold where zoomed-out studies of everything can detect it. That would help reconcile the two literature bases. But it doesn't seem right. The lead effects are huge. The preschool effects, while moderate, suggest that something as minor as "whatever social advantage your family gets from your mother not having to take care of you for part of the day from ages 3 – 5" can have lasting and detectable effects. Surely then we would expect much larger effects from whether your mother is independently wealthy and can do whatever she wants, or whether your family otherwise has the ability to accrue social advantage.

It might also be an effect of what we're measuring. Although there's conventional wisdom that shared environment shows little effect in twin studies, there are occasional outliers. For example, studies of crime often find shared environment factors around 15-20%, especially in younger or poorer samples. And some of the studies that found effects from preschool measured crime. These are some inconsistent findings, and 15-20% from everything doesn't seem consistent with measurable effects from preschool alone, but I'm kind of desperate here.

I guess I will just increase my belief in the studies that suggest shared environment matters a bit more when you limit yourself to non-cognitive factors and include the really poor, and hope that future work confirms this result.

I'll also increase my political support for programs like these. I think these findings make universal childcare (almost) a no-brainer. They make universal pre-K much more appealing, with the strongest arguments against being inefficiency, eg that universal childcare or basic income are a more effective way of doing the same thing. But given the political realities that make universal pre-K more likely to happen than childcare or basic income, I am now happy to support it.

Marijuana: An Update

[Originally to be titled “Marijuana: I Was Wrong”, but looking back I was suitably careful about everything, and my reward is not having to say that.]

Five years ago, I reviewed the potential costs and benefits of marijuana legalization and concluded that there wasn’t enough evidence for a firm conclusion. I found that using some made-up math, the effects looked slightly positive, but this was very sensitive to small changes in how made-up the math was.

The only really interesting conclusion was that most of the objective costs or benefits of legalization came from road traffic accidents. Either stoned driving would increase such accidents, killing thousands. Or people using marijuana instead of alcohol would decrease those accidents, saving thousands. I concluded:

> We should probably stop [emphasizing direct] health effects of marijuana and imprisonment for marijuana-related offenses, and concentrate all of our research and political energy on how marijuana affects driving.

Using the best evidence available at the time, I predicted that marijuana legalization would probably decrease road traffic accidents. Now several states have legalized marijuana, data are in, and we have some preliminary evidence on how marijuana affects driving. And I was wrong.

A study by the Highway Loss Data Institute in June of last year finds that states that legalized marijuana saw insurance claims for auto accidents increase about 3% over the general national trend for the time. An updated study by the same group finds 6% according to insurance claims, and 5.2% according to police reports.

These are usually contrasted with a 2017 study that finds legalization states did not have significantly increased rates of car accident fatality. What is going on?

This study finds a (non-significant) increase of 2.7%. This very nicely matches the non-fatal collision study covering the same period, which finds a 3% increase in total collisions. But because there’s a lower sample size of fatal collisions compared to total collision, the fatality result fails to reach significance. Probably the reason these two results are lower than the 5.2% – 6% result is because the 5.2% – 6% result is newer, and marijuana sales have been increasing every year after legalization.

If this interpretation is true, we should expect that a mature legal marijuana industry causes about a 5% increase in car crashes and fatalities. Score one point for “obvious things” in its fight with “clever attempts to draw counterintuitive conclusions because of substitution effects”.

In the current set of nine states with legalization, the 5% increase would amount to an extra 300 deaths per year. If the country as a whole legalized, that would make about 1800 extra deaths per year. Using my ~~totally made-up math~~ model from the previous post, this is enough to shift the net effect of marijuana legalization from positive to slightly negative. This is especially true if the alternative to legalization is decriminalization, which has many of the benefits of legalization but fewer costs.

But again, given how weak the math here is and how dependent it is on a lot of assumptions, this probably shouldn’t taken too seriously. Tomorrow we could find out that I interpreted the fatality study wrong and marijuana really does cause uniquely non-fatal car accidents. Or that we should be ignoring all of this and paying attention to the effects on chronic pain. Or that marijuana causes cancer. Wait, no, that one was last week. Screw it.

People pointed out on the original thread that all this quantification of the objective harms and benefits of marijuana left out something important: a lot of people like it. Fair. This is hard to think about, but here are some things that help guide my intuitions:

1. Marijuana still is definitely not as bad as alcohol or smoking, which aren't banned

2. Marijuana still is probably worse than SSRIs, which *are* banned without a prescription (though it's hard to go to jail for having them; consider them "decriminalized"). Don't tell me this a fake comparison; they're both psychoactive drugs that purport to make you calmer and happier.

3. About two thirds of drunk driving deaths are the drunk driver themselves, and stoned driving is probably the same way. We might choose to focus only on the one-third of fatalities that happen to bystanders if we believe people should be allowed to make bad choices that only hurt themselves.

4. Everyone expects the marijuana market to keep expanding in the states where it already exists, so these numbers may increase.

5. Marijuana taxes, spent intelligently, could easily save more lives than these accidents cost.

6. Marijuana taxes won't be spent intelligently

7. If marijuana really does increase cancer risk by a few percent, that could easily outweigh everything else and make it a giant public health disaster.

8. But bacon also increases cancer risk by a few percent, is already a giant public health disaster, and we don’t worry about it that much.

9. If the above calculations are true, *preventing* national legalization of marijuana would save half as many lives as successfully implementing Australia-style gun control in the US.

There wasn’t meant to be a conclusion to all of these: they help guide my intuition, but in so many different directions that I still don’t have a real position.

SSRIs: An Update

Four years ago I examined the claim that SSRIs are little better than placebo. Since then, some of my thinking on this question has changed.

First, we got Cipriani et al's meta-analysis of anti-depressants. It avoids some of the pitfalls of Kirsch and comes to about the same conclusion. This knocks down a few of the lines of argument in my part 4 about how the effect size might look more like 0.5 than 0.3. The effect size is probably about 0.3.

Second, I've seen enough to realize that the anomalously low effect size of SSRIs in studies should be viewed not as an SSRI-specific phenomenon, but as part of a general trend towards much lower-than-expected effect sizes for every psychiatric medication (every medication full stop?). I wrote about this in my post on melatonin:

> The consensus stresses that melatonin is a very weak hypnotic. The Buscemi meta-analysis cites this as their reason for declaring negative results despite a statistically significant effect – the supplement only made people get to sleep about ten minutes faster. "Ten minutes" sounds pretty pathetic, but we need to think of this in context. Even the strongest sleep medications, like Ambien, only show up in studies as getting you to sleep ten or twenty minutes faster; this NYT article says that "viewed as a group, [newer sleeping pills like Ambien, Lunesta, and Sonata] reduced the average time to go to sleep 12.8 minutes compared with fake pills, and increased total sleep time 11.4 minutes." I don't know of any statistically-principled comparison between melatonin and Ambien, but the difference is hardly (pun

> not intended) day and night. Rather than say “melatonin is crap”, I would argue that all sleeping pills have measurable effects that vastly underperform their subjective effects.

Or take benzodiazepines, a class of anxiety drugs including things like Xanax, Ativan, and Klonopin. Everyone knows these are effective (at least at first, before patients develop tolerance or become addicted). The studies find them to have about equal efficacy as SSRIs. You could *almost* convince me that SSRIs don’t have a detectable effect in the real world; you will never convince me that benzos don’t. Even morphine for pain gets an effect size of 0.4, little better than SSRI’s 0.3 and not enough to meet anyone’s criteria for “clinically significant”. Leucht 2012 provides similarly grim statistics for everything else.

I don’t know whether this means that we should conclude “nothing works” or “we need to reconsider how we think about effect sizes”.

All this leads to the **third** thing I’ve been thinking about. Given that the effect size really is about 0.3, how do we square the scientific evidence (that SSRIs “work” but do so little that no normal person could possibly detect them) with the clinical evidence (that psychiatrists and patients often find SSRIs sometimes save lives and often make depression substantially better?)

The traditional way to do this is to say that psychiatrists and patients are wrong. Given all the possible biases involved, they misattribute placebo effects to the drugs, or credit some cases that would have remitted anyway to the beneficial effect of SSRIs, or disproportionately remember the times the drugs work over the times they don’t. While “people are biased” is always an option, this doesn’t fit the magnitude of the clinical evidence that I (and most other psychiatrists) observe. There are patients who will regularly get better on an antidepressant, get worse when they stop it, get better when they go back on it, get worse when they stop it again, et cetera. This raises some questions of its own, like *why patients keep stopping antidepressants that they clearly need in order to function*, but makes bias less likely. Overall the clinical evidence that these drugs work is so strong that I will grasp at pretty much any straw in order to save my sanity and confirm that this is actually a real effect.

Every clinician knows that different people respond to antidepressants differently or not at all. Some patients will have an obvious and dramatic response to the first antidepressant they try. Other patients will have no response to the first antidepressant, but after trying five different things you'll find one that works really well. Still other patients will apparently never respond to anything.

Overall only about 30% – 50% of the time when I start a patient on a particular antidepressant, do we end up deciding this is definitely the right medication for them and they should definitely stay on it. This fits national and global statistics. According to a Korean study, the median amount of time a patient stays on their antidepressant prescription is three months. A Japanese study finds only 44% of patients continued their antidepressants the recommended six months; an American study finds 31%.

Suppose that one-third of patients have some gene that makes them respond to Prozac with an effect size of 1.0 (very large and impressive), and nobody else responds. In a randomized controlled trial of Prozac, the average effect size will show up as 0.33 (one-third of patients get effect size of 1, two-thirds get effect size of 0). This matches the studies.

In the clinic, one-third of patients will be obvious Prozac responders, and their psychiatrist will keep them on Prozac and be very impressed with it as an antidepressant and sing the praises of SSRIs. Two-thirds of patients will get no benefit, and their doctors will write them off as non-responders and try something else. Maybe the something else will work, and then the doctors will sing the praises of *that* SSRI, or maybe they'll just say it's "treatment-resistant depression" and so doesn't count.

In other words, doctors' observation "SSRIs work very well" is an existence statement "there are some patients for whom SSRIs work very well" – and not a universal observation "SSRIs will always work well for all patients". Nobody has ever claimed the latter so it's not surprising that it doesn't match the studies.

I linked Gueorguieva and Krystal on the original post; they are saying some kind of much more statistically sophisticated version of this. But I can't find any other literature on this possibility, which is surprising, because if it were true it should be pretty obvious, and if it were false it should still be worth somebody's time to debunk.

If this were true, it would strengthen the case for the throughput-based model I talk about in Recommendations vs. Guidelines and Anxiety Sampler Kits. Instead of worrying only about a medicine's effect size and side effects, we should worry about whether it is a cheap experiment or an expensive experiment. Imagine a drug that instantly cures 5% of people's depression, but causes terrible nausea in the other 95%. The traditional model would reject this drug, since its effect size in studies is low and it has severe side effects. On the throughput model, give this drug to everybody, 5% of people will be instantly cured, 95% of people will suffer nausea for a day before realizing it doesn't work for them, and then the 5% will keep taking it and the other 95% can do something else. This is obviously a huge exaggeration, but I think the principle holds. If there’s enough variability, the benefit-to-side-effect ratio of SSRIs is interesting only insofar as it tells us where in our guideline to put them. After that, what matters is the benefit-to-side-effect ratio for each individual patient.

I don’t hear this talked about much and I don’t know if this is consistent with the studies that have been done.

Fourth, even though SSRIs are branded “antidepressants”, they have an equal right to be called anti-anxiety medications. There’s some evidence that they may work better for this indication than for depression, although it’s hard to tell. I think Irving Kirsch himself makes this claim: he analyzed the efficacy of SSRIs for everything and found a “relatively large effect size” of 0.7 for anxiety (though the study was limited to children). Depression and anxiety are highly comorbid and half of people with a depressive disorder also have an anxiety disorder; there are reasons to think that at some deep level they may be aspects of the same condition. If SSRIs effectively treated anxiety, this might make depressed people feel better in a way that doesn’t necessarily show up on formal depression tests, but which they would express to their psychiatrist as “I feel better”. Or, psychiatrists might have a vague positive glow around SSRIs if it successfully treats their anxiety patients (who may be the same people as their depression patients) and not be very good at separating that positive glow into “depression efficacy” and “anxiety efficacy”. Then they might believe they’ve had good experiences with using SSRIs for depression.

I don’t know if this is true and some other studies find that results for anxiety are almost as abysmal as for depression.

Ketamine: An Update

In 2016, I wrote Ketamine Research In A New Light, which discussed the emerging consensus that, contra existing theory, ketamine's rapid-acting antidepressant effects had nothing to do with NMDA at all. I discussed some experiments which suggested they might actually be due to a related receptor, AMPA.

The latest development is Attenuation of Antidepressant Effects of Ketamine by Opioid Receptor Antagonism, which finds that the opioid-blocker naltrexone prevents ketamine's antidepressant effects. Naltrexone does not prevent dissociation or any of the other weird hallucinatory effects of ketamine, which are probably genuinely NMDA-related. This suggests it's just a coincidence that NMDA antagonism and some secondary antidepressant effect exist in the same drug. If you can prevent an effect from working by blocking the opiate system, a natural assumption is that the effect works on the opiate system, and the authors suggest this is probably true.

(unexpected national news tie-in: Kavanaugh accuser Christine Blasey Ford is one of the authors of this paper)

In retrospect, there were warnings. The other study to have found an exciting rapid-acting antidepressant effect for an ordinary drug was Ultra-Low-Dose Buprenorphine As A Time-Limited Treatment For Severe Suicidal Ideation. It finds that buprenorphine (the active ingredient in suboxone), an opiate painkiller also used in treating addictions to other opiates, can quickly relieve the distress of acutely suicidal patients.

This didn’t make as big a splash as the ketamine results, for two reasons. First, everyone knows opiates feel good, and so maybe this got interpreted as just a natural extension of that truth (the *Scientific American* article on the discovery focused on an analogy where “mental pain” was the same as “physical pain” and so could be treated with painkillers). Second, we’re currently fighting a War On Opiates, and discovering new reasons to prescribe them seems kind of like giving aid and comfort to the enemy.

Ketamine is interesting because nobody can just reduce its mode of action to “opiates feel good”. Although it was long known to have some weak opiate effects, it doesn’t feel good; all the dissociation and hallucinations and stuff make sure of that. Whatever is going on is probably something more complicated.

The psychiatric establishment’s response, as published in the prestigious *American Journal of Psychiatry*, is basically “well, f@#k”. Here we were, excited about NMDA (or AMPA) giving us a whole new insight into the mechanisms of depression and the opportunity for a whole new class of treatment – and instead it looks like maybe it’s just pointing to The Forbidden Drugs That Nobody Is Supposed To Prescribe. The article concludes that ketamine should not be abandoned, but ketamine clinics under anaesthesiologists should be discouraged in favor of care monitored by psychiatrists. I will try not to be so cynical as to view this as the establishment seizing the opportunity for a power grab.

What happens now? A lot of this depends on addiction. One way we could go would be to say that although ketamine might have some opiate effects, it’s not addictive to the same degree as morphine, and it doesn’t seem to turn users into drug fiends, so we should stop worrying and press forward. We could even focus research on finding other opiates in a sweet spot where they’re still strong enough to fight depression but not strong enough to get people addicted. Maybe very-low-dose-buprenorphine is already in this sweet spot, I don’t know.

But all of this is going to be shaped by history. Remember that heroin was originally invented (and pushed) as a less-addictive, safer opiate that would solve the opiate crisis. Medicine has a really bad habit of seizing on hopes that we have found a less addictive version of an addictive thing, and only admitting error once half the country is addicted to it. And there are all sorts of weird edge cases – does ketamine cross-sensitize people to other opiates? Does it increase some sort of domain-general addiction-having-center in the brain? I know substance abuse doctors who believe all of this stuff.

Also, should we start thinking opiates have some sort of deep connection to depression? “Depression is related to the stuff that has the strongest effect on human happiness of any molecule class known” seems…actually pretty plausible now that I think about it. I don’t know how much work has been done on this before. I hope to see more.

Preschool: Much More Than You Wanted To Know

I.

A lot of people pushed back against my post on preschool, so it looks like we need to discuss this in more depth.

A quick refresher: good randomized controlled trials have shown that preschools do not improve test scores in a lasting way. Sometimes test scores go up a little bit, but these effects disappear after a year or two of regular schooling. However, early RCTs of intensive "wrap-around" preschools like the Perry Preschool Program and the Abecedarians found that graduates of those programs went on to have markedly better adult outcomes, including higher school graduation rates, more college attendance, less crime, and better jobs. But these studies were done in the 60s, before people invented being responsible, and had kind of haphazard randomization and followup. They were also small sample sizes, and from programs that were more intense than any of the scaled-up versions that replaced them. Modern scaled-up preschools like Head Start would love to be able to claim their mantle and boast similar results. But the only good RCT of Head Start, the HSIS study, is still in its first few years. It's confirmed that Head Start test score gains fade out. But it hasn't been long enough to study whether there are later effects on life outcomes. We can expect those results in ten years or so. For now, all we have is speculation based on a few quasi-experiments.

Deming 2009 is my favorite of these. He looks at the National Longitudinal Survey of Youth, a big nationwide survey that gets used for a lot of social science research, and picks out children who went to Head Start. These children are mostly disadvantaged because Head Start is aimed at the poor, so it would be unfair to compare them to the average child. He’s also too smart to just “control for income”, because he knows that’s not good enough. Instead, he finds children who went to Head Start but who have siblings who didn’t, and uses the sibling as a matched control for the Head Starter.

This ensures the controls will come from the same socioeconomic stratum, but he acknowledges it raises problems of its own. Why would a parent send one child to Head Start but not another? It might be that one child is very stupid and so the parents think they need the extra help preschool can provide; if this were true, it would mean Head Starters are systematically dumber than controls, and would underestimate the effect of Head Start. Or it might be that one child is very smart and the so the parents want to give them education so they can develop their full potential; if this were true, it would mean Head Starters are systematically smarter than controls, and would inflate the effect of Head Start. Or it might be that parents love one of their children more and put more effort into supporting them; if this meant these children got other advantages, it would again inflate the effect of Head Start. Or it might mean that parents send the child they love more to a fancy private preschool, and the child they love less gets stuck in Head Start, ie the government program for the disadvantaged. Or it might be that parents start out poor, send their child to Head Start, and then get richer and send their next child to a fancy private preschool, while that child also benefits from their new wealth in other ways. There are a lot of possible problems here.

Deming tries very hard to prove none of these are true. He compares Head Starters and their control siblings on thirty different pre-study variables, including family income during their preschool years, standardized test scores, various measures of health, number of hours mother works during their preschool years, breastfedness, etc. Of these thirty variables, he finds a significant difference on only one: birth weight. Head Starters were less likely to have very low birth weight than their control siblings. This is a moderately big deal, since birth weight is a strong predictor of general child health and later life success. But:

> Given the emerging literature on the connection between birth weight and later outcomes, this is a serious threat to the validity of the [study]. There are a few reasons to believe that the birth weight differences are not a serious source of bias, however. First, it appears that the difference is caused by a disproportionate number of low-birth-weight children, rather than by a uniform rightward shift in the distribution of birth weight for Head Start children. For example, there are no significant differences in birth weight once low-birth-weight children (who represent less than 10 percent of the sample) are excluded.
>
> Second, there is an important interaction between birth order and birth weight in this sample. Most of the difference in mean birth weight comes from children who are born third, fourth, or later. Later-birth-order children who subsequently enroll in Head Start are much less likely to be low birth weight than their older siblings who did not enroll in preschool. When I restrict the analysis to sibling pairs only, birth weight differences are much smaller and no longer significant, and the main results are unaffected. Finally, I estimate all the models in Section V with low-birth-weight children excluded, and, again, the main results are unchanged.

> Still, to get a sense of the magnitude of any possible positive bias, I back out a correction using the long-run effect of birth weight on outcomes estimated by Black, Devereux, and Salvanes (2007). Specifically, they find that 10 percent higher birth weight leads to an increase in the probability of high school graduation of 0.9 percentage points for twins and 0.4 percentage points for siblings. If that reduced form relationship holds here, a simple correction suggests that the effect of Head Start on high school graduation (and by extension, other outcomes) could be biased upward by between 0.2 and 0.4 percentage points, or about 2–5 percent of the total effect.

Having set up his experimental and control group, Deming does the study and determines how well the Head Starters do compared to their controls. The test scores show some confusing patterns that differ by subgroup. Black children (the majority of this sample; Head Start is aimed at disadvantaged people in general and sometimes at blacks in particular) show the classic pattern of slightly higher test scores in kindergarten and first grade, fading out after a few years. White children never see any test score increases at all. Some subgroups, including boys and children of high-IQ mothers, see test score increases that don’t seem to fade out. But these differences in significance are not themselves significant and it might just be chance. Plausibly the results for blacks, who are the majority of the sample, are the real results, and everything else is noise added on. This is what non-subgroup analysis of the whole sample shows, and it’s how the study seems to treat it.

The nontest results are more impressive. Head Starters are about 8% more likely to graduate high school than controls. This pattern is significant for blacks, boys, and children of low-IQ mothers, but not for whites, girls, and children of high-IQ mothers. Since the former three categories are the sorts of people at high risk of dropping out of high school, this is probably just floor effects. Head Starters are also less likely to be diagnosed with a learning disability (remember, learning disability diagnosis is terrible and tends to just randomly hit underperforming students), and marginally less likely to repeat grades. The subgroup results tend to show higher significance levels for groups at risk of having bad outcomes, and lower significance levels for the rest, just as you would predict. There is no effect on crime. For some reason he does not analyze income, even though his dataset should be able to do that.

He combines all of this into an artificial index of “young adult outcomes” and finds that Head Start adds 0.23 SD. You may notice this is less than the 0.3 SD effect size of antidepressants that everyone wants to dismiss as meaningless, but in the social sciences apparently this is pretty good. Deming optimistically sums this up as “closing one-third of the gap between children with median and bottom-quartile family income”, as “75% of the black-white gap”, and as “80% of the benefits of [Perry Preschool] at 60% of the cost”.

Finally, he does some robustness checks to make sure this is not too dependent on any particular factor of his analysis. I won’t go into these in detail, but you can find them on page 127 of the manuscript, and it’s encouraging that he tries this, given that I’m used to reading papers by social psychologists who treat robustness checks the way vampires treat garlic.

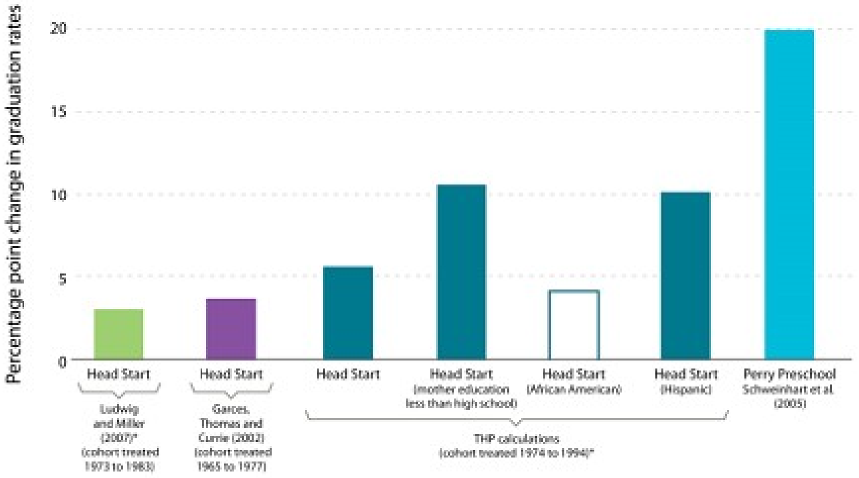
Deming’s paper very similar to Garces Thomas & Currie (2002), which does the same methodology on a different dataset. GTC is earlier and more famous and probably the paper you’ll hear about if you read other discussions of this topic; I'm focusing on Deming because I think his analyses are more careful and he explains what he's doing a lot better. Reading between the lines, GTC do not find any significant effects for the sample as a whole. In subgroup analyses, they find Head Start makes whites more likely to graduate high school and attend college, and blacks less likely to be involved in crime. One can almost sort of attribute this to floor effects; blacks many times more likely to have contact with the criminal justice system, and there are more blacks than whites in the sample, so maybe it makes sense that this is only significant for them. On the other hand, when I look at the results, there was almost as strong a *positive* effect for whites (ie Head Start whites committed more crimes, to the same degree Head Start blacks committed fewer crimes) – but there were fewer whites so it didn't quite reach significance. And the high school results don't make a lot of sense however you parse them. GTC use the words "statistically significant" a few times, so you know they're thinking about it. But they don't ever give significance levels for individual results and one gets the feeling they're not very impressive. Their pattern of results isn't really that similar to Deming's either – remember, Deming found that all races were more likely to benefit from high school, and no race had less crime. GTC also don't do nearly as much work to show that there aren't differences between siblings. Deming is billed as confirming or replicating GTC, but this only seems true in the sense that both of them say nice things about Head Start. Their patterns of results are pretty different, and GTC's are kind of implausible.

And for that matter, ten years earlier two of these authors, Currie and Thomas, did a similar study. They also use the National Longitudinal Survey of Youth, meaning I'm not really clear how their analysis differs from Deming's (maybe it's much earlier and so there's less data?) They first use an “adjust for confounders” model and it doesn’t work very well. Then they try a comparing-siblings model and find that Head Starters are generally older than their no-preschool siblings, and also generally born to poorer mothers (these are probably just the same result; mothers get less poor as they get older). They also tend to do better on a standardized test, though the study is very unclear about when they’re giving this test so I can’t tell if they’re saying that group assignment is nonrandom or that the intervention increased test scores. They find Head Start does not increase income, maybe inconsistently increases test scores among whites but not blacks, decreases grade repetition for whites but not blacks, and improves health among blacks but not whites. They also look into Head Start’s effect on mothers, since part of the wrap-around program involves parent training. All they find is mild effects on white IQ scores, plus “a positive and implausibly large effect of Head Start on the probability that a white mother was a teen at the first birth” which they say is probably sampling error. Like the later study, this study does not give p-values and I am too lazy to calculate them from the things they do give, but it doesn’t seem like they’re likely to be very good.

Finally, Deming’s work was also replicated and extended by a team from the Brookings Institute. I think what they’re doing is taking the National Longitudinal Survey of Youth – the same dataset Deming and one of the GTC papers used – and updating it after a few more years of data. Like Deming, they find that “a wide variety” of confounders do not differ between Head Starters and their unpreschooled siblings. Because they’re with the Brookings Institute, their results are presented in a much prettier way than anyone else’s:

FIGURE 1.

Effect of Early Education on High School Graduation Rates



Note: Hollowed bars are not statistically significant at the 10 percent level. *Estimate for Ludwig and Miller (2007) may include and THP calculations do include some GED completers, but authors estimate their contribution to the total is small. There is no differential effect for white students. See the technical appendix.

THE HAMILTON PROJECT
BROOKINGS

The Brookings replication (marked THP here) finds sizes somewhat larger than GTC, but somewhat smaller than Perry Preschool. It looks like they find a positive and significant effect on high school graduation for Hispanics, but not blacks or whites, which is a different weird racial pattern than all the previous weird racial patterns. Since their sample was disproportionately black and Hispanic, and the blacks *almost* reached significance, the whole sample is significant. They find increases of about 6% on high school graduation rates, compared to Deming’s claimed 8%, but on this chart it’s hard to see how Deming said his 8% was 80% as good as Perry Preschool. There are broadly similar effects on some other things like college attendance, self esteem, and “positive parenting”. They conclude:

> These results are very similar to those by Deming (2009), who calculated high school graduation rates on the more limited cohorts that were available when he conducted his work.

These four studies – Deming, GTC, CT, and Brookings – all try to do basically the same thing, though with different datasets. Their results all sound the same at the broad level – "improved outcomes like high school graduation for some racial groups" – but on the more detailed level they can't really agree which outcomes improve and which racial groups they improve for. I'm not sure how embarrassing this should be for them. All of their results seem to be kind of on the border of significance, and occasionally going below that border and occasionally above it, which helps explain the contradictions while also being kind of embarrassing in and of itself (Deming's paper is the exception, with several results significant at the 0.01 level). Most of them do find things generally going the right direction and generally sane-looking findings. Overall I feel like Deming looks pretty good, the Brookings replication is too underspecified for me to have strong opinions on, and the various GTC papers neither add nor subtract much from this.

II.

I'm treating Ludwig and Miller separately because it's a different – and more interesting – design.

In 1965, the government started an initiative to create Head Start programs in the 300 poorest counties in the US. There was no similar attempt to help counties #301 and above, so there's a natural discontinuity at county #300. This is the classic sort of case where you can do a regression discontinuity experiment, so Ludwig and Miller decided to look into it and see if there was some big jump in child outcomes as you moved from the 301st-poorest-county to the 300th.

They started by looking into health outcomes, and found a dramatic jump. Head Start appears to improve the outcomes of certain easily-preventable childhood diseases 33-50%. For example, kids from counties with Head Start programs had much less anemia. Part of the Head Start program is screening for anemia and supplementing children with iron, which treats many anemias. So this is very unsurprising. Remember that the three hundred poorest counties in 1965 were basically all majority-black counties in the Deep South and much worse along every axis than you would probably expect – we are talking near-Third-World levels of poverty here. If you deploy health screening and intervention into near-Third-World levels of poverty, then the rates of easily preventable diseases should go down. Ludwig and Miller find they do. This is encouraging, but not really surprising, and maybe not super-relevant to the rest of what we’re talking about here.

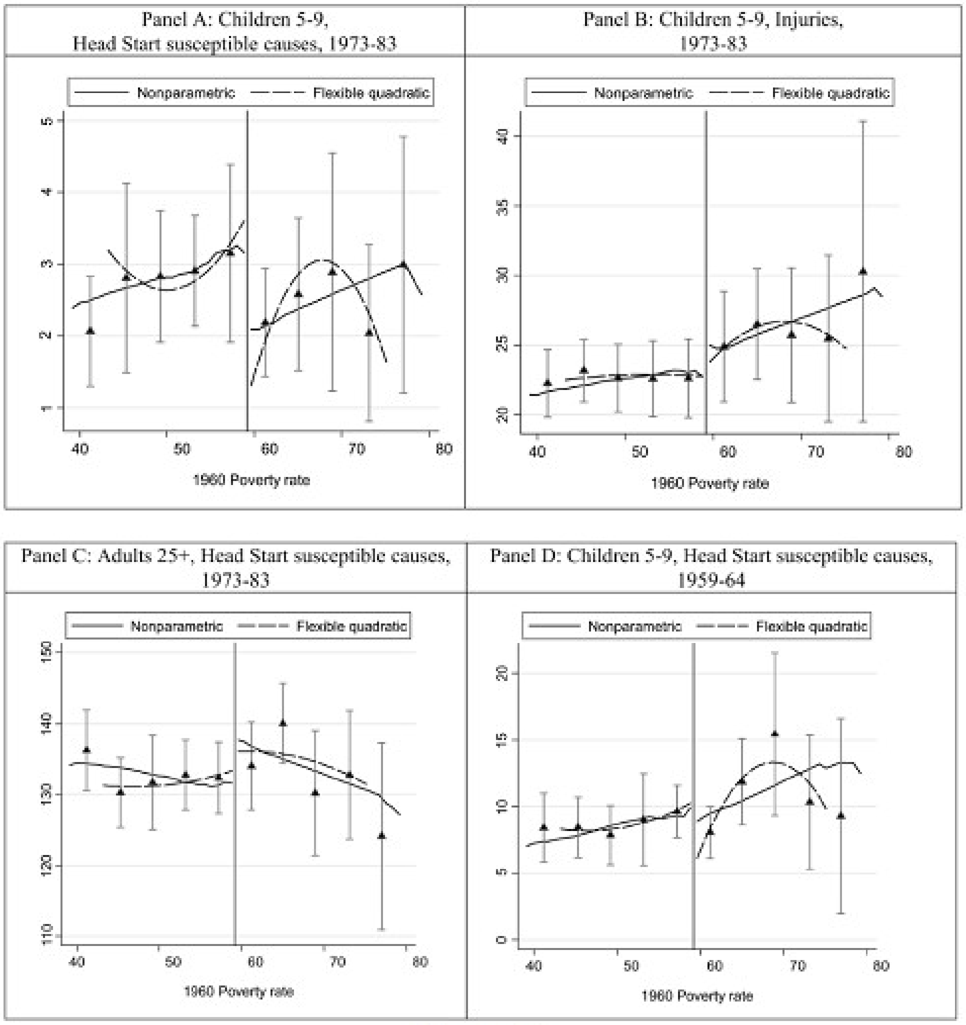
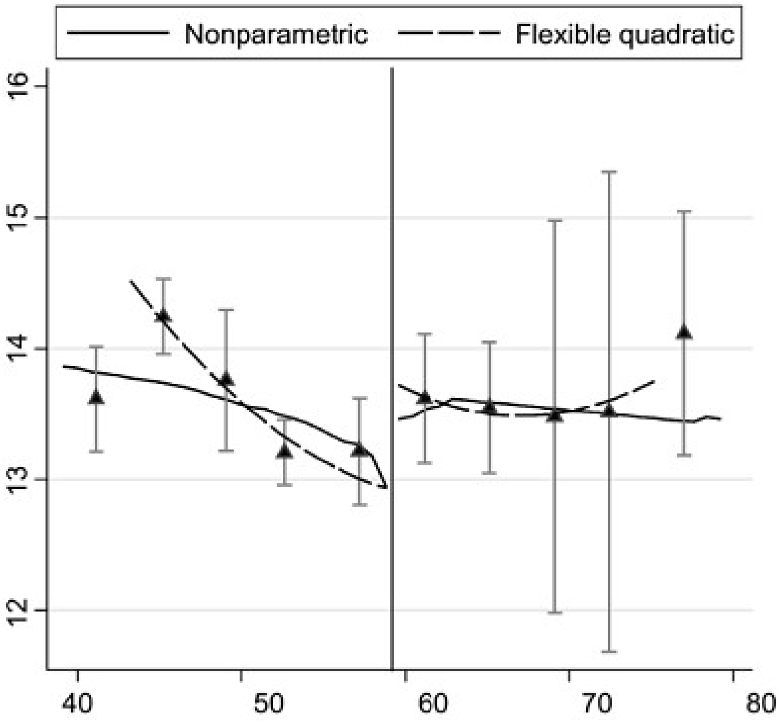


FIGURE IV

Estimated Discontinuities at OEO Cutoff in Mortality Rates per 100,000 for Children and Adults, from Causes Affected by Head Start and from Injuries

But they also find a “positive discontinuity” in high school completion of about 5%. Kids in the 300th-and-below-poorest counties were about 5% more likely than kids in the 301st-and-above-poorest to finish high school. This corresponds to an average of staying in school six months longer. This discontinuity did not exist before Head Start was set up, and it does not exist among children who were the wrong age to participate in Head Start at the time it was set up. It comes into existence just when Head Start is set up, among the children who were in Head Start. This is a pretty great finding.



Unfortunately, it looks like this. The authors freely admit this is just at the limit of what they can detect at $p < 0.05$ in their data. They double check with another data source, which shows the same trend but is only significant at $p < 0.1$. "Our evidence for positive Head Start impacts on educational attainment is more suggestive, and limited by the fact that neither of the data sources available to us is quite ideal." This study has the strongest design, and it does find an effect, but the effect is basically squinting at a graph and saying "it kind of looks like that line might be a little higher than the other one". They do some statistics, but they are all the statistical equivalent of squinting at the graph and saying "it kind of looks like that line might be a little higher than the other one", and about as convincing. For a more complete critical look, see this post from the subreddit.

There is one other slightly similar regression discontinuity study, Carneiro and Ginja, which regresses a sample of people on Head Start availability and tries to prove that people who went to Head Start because they were just within the availability cutoff do better than people who missed out on Head Start because they were just outside it. This sounds clever and should be pretty credible. They find a bunch of interesting effects like that Head Starters are less likely to be obese, and less likely to be depressed. They find that non-blacks (but not blacks) are less likely to be involved in crime (which, remember, is the opposite finding as the last paper about Head Start and crime and race). But they don't find any effect on likelihood to graduate high school or be involved in college. Also, they bury this result and everyone cites this paper as "Look, they've replicated that Head Start works!"

III.

A few scattered other studies to put these in context:

In 1980, Chicago created "Child Parent Centers", a preschool program aimed at the disadvantaged much like all of these others we've been talking about. They did a study, which for some reason published its results in a medical journal, and which doesn't really seem to be trying in the same way as the others. For example, it really doesn't say much about the control group except that it was "matched". Taking advantage of their unusually large sample size and excellent follow-up, they find that their program made children stay in school the same six months longer as many of the other studies find, had a strong effect on college completion (8% vs. 14% of kids), showed dose-dependent effects, and "was robust". They are bad enough at showing their work that I am forced to trust them and the *Journal of the American Medical Association*, a prestigious journal that I can only hope would not have published random crap.

Havnes and Mogstad analyze a free universal child-care program in Norway, which was rolled out in different places at different times. They find that “exposure to child care raised the chances of completing high school and attending college, in orders of magnitude similar to the black-white race gaps in the US”. I am getting just cynical enough to predict that if Norway had black people, they would have a completely different pattern of benefits and losses from this program, but the Norwegians were able to avoid a subgroup analysis by being a nearly-monoethnic country. This is in contrast to Quebec, where a similar childcare program seems to have caused worse long-term outcomes. Going deeper into these results supports (though weakly and informally) a model where, when daycare is higher-quality than parental care, child outcomes improve; when daycare is lower-quality than parental care, child outcomes decline. So a reform that creates very good daycare, and mostly attracts children whose parents would not be able to care for them very well, will be helpful. Reforms that create low-quality daycare and draw from households that are already doing well will be harmful. See the discussion here.

Then there’s Chetty’s work on kindergarten, which I talk about here. He finds good kindergarten teachers do not consistently affect test scores, but do consistently affect adult earnings, similar to fade-out arguments around preschool. This study is randomized and strong. Its applicability to the current discussion is questionable, since kindergarten is not preschool, having a good teacher is not going to preschool at all, and the studies we’re looking at mostly haven’t found results about adult earnings. At best this suggests that schooling can have surprisingly large and fading-out-then-in-again effects on later life outcomes.

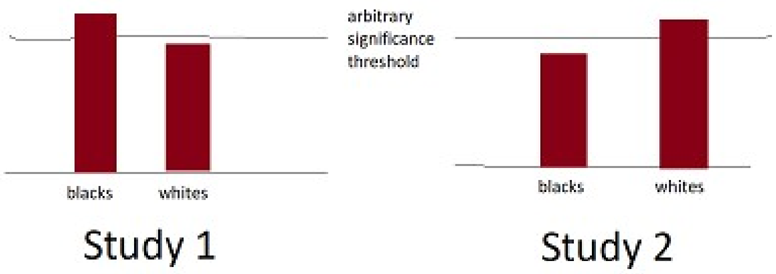
And finally, there’s a meta-analysis of 22 studies of early childhood education showing an effect size of 0.24 SD in favor of graduating high school, p less than 0.001. Maybe I should have started with that one. Maybe it’s crazy of me to save this for the end. Maybe this should count for about five times as much as everything I’ve mentioned so far. I’m putting it down here both to inflict upon you the annoyance *I* felt when discovering this towards the end of researching this topic, and so that you have a good idea of what kind of studies are going into this meta-analysis.

IV.

What do we make of this?

I am concerned that all of the studies in Parts I and II have been summed up as “Head Start works!”, and therefore as replicating each other, since the last study found “Head Start works!” and so did the newest one. In fact, they all find Head Start having small effects for some specific subgroup on some specific outcome, and it’s usually a different subgroup and outcome for each. So although GCT and Deming are usually considered replications of each other, they actually disprove each other’s results. One of GCT’s two big findings is that Head Start decreases crime among black children. But Deming finds that Head Start had no effect on crime among black children. The only thing the two of them agree on is that Head Start seems to improve high school graduation among whites. But Carneiro and Ginja, which is generally thought of as replicating the earlier two, finds Head Start has no effect on high school graduation among whites.

There’s an innocent explanation here, which is that everyone was very close to the significance threshold, so these are just picking up noise. This might make more sense graphically:

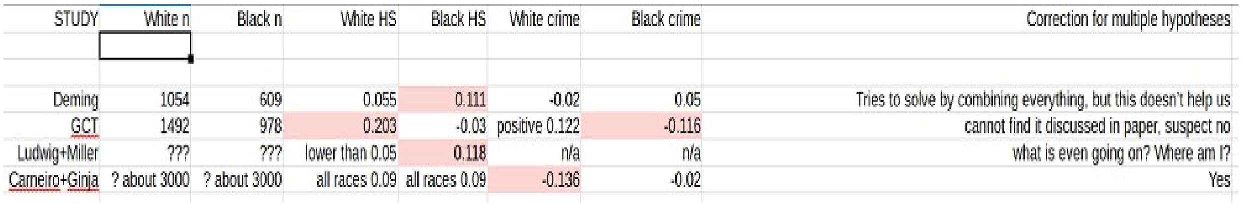


Study 1

Study 2

It’s easy to see here that both studies found basically the same thing, minus a little noise, but that Study 1 has to report its results as “significant for blacks but not whites” and Study 2 has to report the opposite. Is this what’s going on?

I made a table. I am really really not confident in this table. On one level, I am fundamentally not confident that what I am doing is even possible, and that the numbers in these studies are comparable to one another or mean what it looks like they mean. On a second level, I’m not sure I recorded this information correctly or put the right numbers in the right places. Still, here is the table; red means the result is significant:



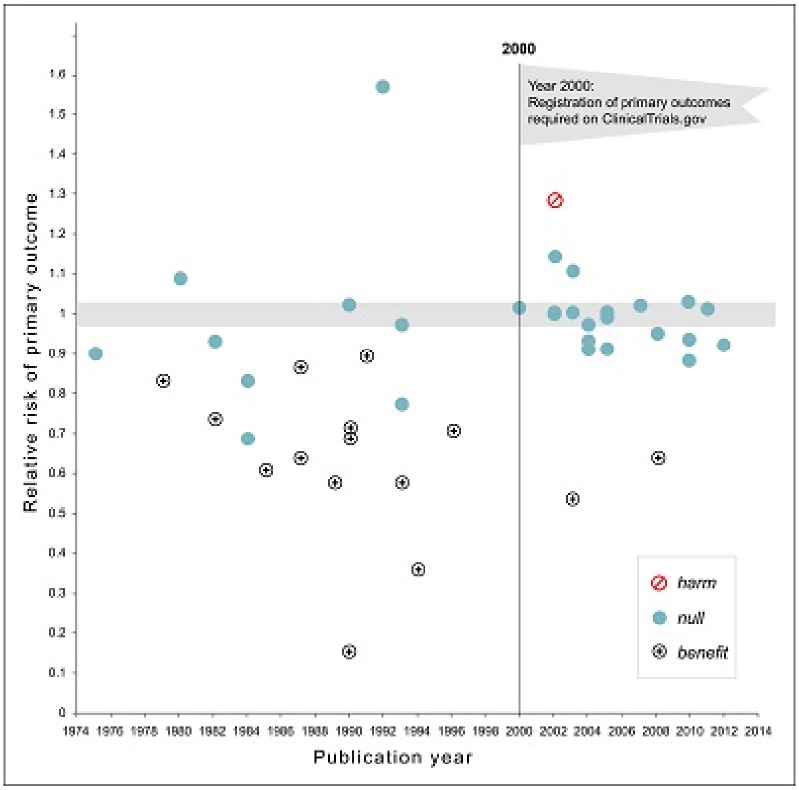
STUDY	White n	Black n	White HS	Black HS	White crime	Black crime	Correction for multiple hypotheses
Deming	1054	609	0.055	0.111	-0.02	0.05	Tries to solve by combining everything, but this doesn't help us
GCT	1492	978	0.203	-0.03	positive 0.122	-0.116	cannot find it discussed in paper, suspect no
Ludwig+Miller	???	???	lower than 0.05	0.118	n/a	n/a	what is even going on? Where am I?
Carneiro+Ginja	? about 3000	? about 3000	all races 0.09	all races 0.09	-0.136	-0.02	Yes

This confirms my suspicions. Every study found something different, and it isn’t even close. For example, Carneiro & Ginja finds a strong effect of lowering white crime, but GCT finds that Head Start nonsignificantly *increases* white crime rates. Meanwhile, GCT find a strong and significant effect lowering black crime, but Carneiro and Ginja find an effect of basically zero.

The strongest case for the studies being in accord is for black high school graduation rates. Both Deming and Ludwig+Miller find an effect. Carneiro and Ginja don't find an effect, but their effect size is similar to those of the other studies, and they might just have more stringent criteria since they are adjusting for multiple comparisons and testing many things. But they *should* have the more stringent criteria, and by trying to special-plead against this, I am just reversing the absolutely correct thing they did because I want to force positive results in the exact way that good statistical practice is trying to prevent me from doing. So maybe I shouldn't do that.

Here is the strongest case for accepting this body research anyway. It doesn't quite look like publication bias. For one thing, Ludwig and Miller have a paper where they say there's probably no publication bias here because literally every dataset that can be used to test Head Start has been. For another, although I didn't focus on gender or IQ on the chart above, most of the studies do find that it helps males and low-IQ people more with the sorts of problems men and low-IQ people usually face, which suggest it passes sanity checks. Most important, in a study whose results are entirely spurious, there should be an equal number of beneficial and harmful findings (ie they should find Head Start makes some subgroups worse on some outcomes). Since each of these studies investigates many things and usually finds many different significant results, it should be hard to publication bias *all* harmful findings out of existence. This sort of accords with the positive meta-analysis. Studies either show small positive results or are not signficant, and when you combine all of them into a meta-analysis, they become highly significant, look good, and make sense. And this would fit very well with the Norwegian study showing strong positive effects of childcare later in life. And Chetty's study showing fade-out of kindergarten teachers followed by strong positive effects later in life. And of course the Perry Preschool and Abecedarian studies showing fade-out of tests scores followed by strong positive effects later in life. I even recently learned of a truly marvelous developmental explanation for why this might happen, which unfortunately this margin is too small to contain – expect a book review in the coming weeks.

The case against this research is that maybe the researchers cheated to have there be no harmful findings. Maybe the meta-analysis just shows that when a lot of researchers cheat a little, taking care to only commit minor undetectable sins, that adds up to a strong overall effect. This is harsh, but I was recently referred to this chart (h/t Mother Jones, which calls it "the chart of the decade" and "one of the greatest charts ever produced"):



This is the outcome of drug trials before and after the medical establishment started requiring preregistration (the vertical line) – in other words, before they made it harder to cheat. Before the vertical line, 60% of trials showed the drug in question was beneficial. After the vertical line, only 10% did. In other words, making it harder to cheat cuts the number of positive trials by a factor of six. It is not at all hard to cheat in the research of early childhood education; all the research in this post so far comes from the left side of the vertical line. We should be skeptical of all but the most ironclad research that comes from the left – and this is not the most ironclad research.

The Virtues of Rationality say:

> One who wishes to believe says, “Does the evidence permit me to believe?” One who wishes to disbelieve asks, “Does the evidence force me to believe?” Beware lest you place huge burdens of proof only on propositions you dislike, and then defend yourself by saying: “But it is good to be skeptical.” If you attend only to favorable evidence, picking and choosing from your gathered data, then the more data you gather, the less you know. If you are selective about which arguments you inspect for flaws, or how hard you inspect for flaws, then every flaw you learn how to detect makes you that much stupider.

This is one of the many problems where the evidence permits me to disbelieve, but does not force me to do so. At this point I have only intuition and vague heuristics. My intuition tells me that in twenty years, when all the results are in, I expect early childhood programs to continue having small positive effects. My vague heuristics say the opposite, that I can’t trust research this irregular. So I don’t know.

I think I was right to register that my previous belief preschool definitely didn’t work was outdated and under challenge. I think I was probably premature to say I was *wrong* about preschool not working; I should have said I might be wrong. If I had to bet on it, I would say 60% odds preschool helps in ways kind of like the ones these studies suggest, 40% odds it’s useless.

I hope that further followup of the HSIS, an unusually good randomized controlled trial of Head Start, will shed more light on this after its participants reach high school age sometime in the 2020s.

The Economic Perspective On Moral Standards

[Content warning: scrupulosity. Some recent edits, see Mistakes page for details.]

I.

There are some pretty morally unacceptable things going on in a pretty structural way in society. Sometimes I hear some activists take this to an extreme: no currently living person is morally acceptable. People who aren't reorienting their entire lives around acknowledging and combating the evils of the world aren't even on the scale. And people who are may be (in the words of one of my friends who is close to that community) "only making comfortable sacrifices that let them think of themselves as a good person within their existing comfortable moral paradigm, instead of confronting the raw terrible truth." IE "If you think you're one of the good ones, you're wrong".

I have heard this sentiment raised by animal rights activists. The average meat-eater isn't even on the scale. The average vegetarian still eats milk and cheese, and so is barely even trying. Even most vegans probably use some medical product with gelatin, or something tested on lab rats, or are just benefitting from animal suffering in some indirect way.

And I have heard it raised by environmentalists. The average SUV driver isn't even on the scale. The average conscientious liberal might think they're better because they bike to work and recycle, but they still barely think about how they're using electricity generated by coal plants and eating food grown with toxic pesticides. Everyone could be doing more.

And I have heard raised by labor activists. Most of us use stuff made in sweatshops. Even if you avoid sweatshops, you probably use stuff made at less than a living wage. Even if you avoid that, are you doing everything you can to help and support workers who earn less than you do?

Even if you aren’t an animal rights activist, environmentalist, or labor advocate, do you believe in anything? Are you a Christian, a social justice advocate, or rationalist? Do you know anyone who really satisfies you as being sinless, non-racist, and/or rational? Then perhaps you too believe nobody is good.

We shouldn’t immediately dismiss the idea that nobody is good. By our standards, there were many times and places where this was true. I am not aware of any ancient Egyptians who were against slavery. By Roman times, a handful of people thought it might be a bad idea, but nobody lifted a finger to stop it. I doubt you could find any Roman at the intersection of currently acceptable positions on slavery, torture, women’s rights, and sexuality. Maybe a few followers of Epicurus – but is there much difference between 0.01% of people being good, and nobody being good?

But let’s back up here philosophically. There’s a clear definition of “perfectly good” – someone who has never deviated from morally optimal behavior in any way. “Nobody is perfectly good” is, I think, an uncontroversial statement. If “nobody is good” is controversial, it’s because we expect “good” to be a lower bar than “perfectly good”, representing a sort of minimum standard of okayness. It might be possible that nobody meets even a minimum standard of okayness – the Roman example still seems relevant – but we should probably back up further and figure out how we're setting okayness standards.

My subjective impression of what we mean by "good" in the sense of "a decent person" or "minimally okay" has internal and external components. Internally, it means a person who is allowed to feel good about themselves instead of feeling guilty. Externally, it means a person who deserves praise rather than punishment.

Some people would deny one side or the other of this dichotomy. For example, some people believe nobody should ever feel guilty. Or, on the other hand, the "do you want a fucking cookie?" attitude activists sometimes take toward people who expect praise for their acts of support. This seems to rest on an assumption that even socially rare levels of virtue fall within the realm of "your basic minimal duty as a human being" and so should not get extra praise.

But I find the "good person"/"not a good person" dichotomy helpful. I'm not claiming it objectively exists. I can't prove anything about ethics objectively exists. And even if there were objective ethical truths about what was right or wrong, that wouldn't imply that there was an objective ethical truth about how much of the right stuff you have to do before you can go around calling yourself "good". In the axiology/morality/law trichotomy, I think of "how much do I have to do in order to be a good person" as within the domain of morality. That means it's a social engineering question, not a philosophical one. The social engineering perspective assumes that "good person" status is an incentive that can be used to make people behave better, and asks how high vs. low the bar should be set to maximize its effectiveness.

Consider the way companies set targets for their employees. At good companies, goals are ambitious but achievable. If the CEO of a small vacuum company tells her top salesman to sell a billion vacuums a year, this doesn’t motivate the salesman to try extra hard. It’s just the equivalent of not setting a goal at all, since he’ll fail at the goal no matter what. If the CEO says “Sell the most vacuums you can, and however many you sell, I will yell at you for not selling more”, this also probably isn’t going to win any leadership awards. A good CEO might ask a salesman to sell 10% more vacuums than he did last year, and offer a big bonus if he can accomplish it. Or she might say that the top 20% of salesmen will get promotions, or that the bottom 20% of salesmen will be fired, or something like that. The point is that the goal should effectively carve out two categories, “good salesman” and “bad salesman”, such that it’s plausible for any given salesman to end up in either, then offer an incentive that makes him want to fall in the first rather than the second.

I think of society setting the targets for “good person” a lot like a CEO setting the targets for “good vacuum salesman”. If they’re attainable and linked to incentives – like praise, honor, and the right to feel proud of yourself – then they’ll make people put in an extra effort so they can end up in the “good person” category. If they’re totally unattainable and nobody can ever be a good person no matter how hard they try, then nobody will bother trying. This doesn’t mean nobody will be good – some people are naturally good without hope for reward, just like some people will slave away for the vacuum company even when they’re underpaid and underappreciated. It just means you’ll lose the extra effort you would get from having a good incentive structure.

So what is the right level at which to set the bar for “good person”? An economist might think of this question as a price-setting problem: society is selling the product “moral respectability” and trying to decide how many units effort to demand from potential buyers in order to maximize revenue. Set the price too low, and you lose out on money that people would have been willing to pay. Set the price too high, and you won’t get any customers. Solve for the situation where you have a monopoly on the good and the marginal cost of production is zero, and this is how you set the “good person” bar.

I don’t have the slightest idea how you would actually go about doing that, and it’s just a metaphor anyway, so let me give some personal stories and related considerations.

II.

When I was younger, I determined that I had an ethical obligation to donate more money to charity, and that I was a bad person for giving as little as I did. But I also knew that if I donated more, I would be a bad person for not donating even more than that. Given that there was no solution to my infinite moral obligation, I just donated the same small amount.

Then I met a committed group of people who had all agreed to donate 10%. They all agreed that if you donated that amount you were doing good work and should feel proud of yourself. And if you donated less than that, then they would question your choice and encourage you to donate more. I immediately pledged to donate 10%, which was much more than I had been doing until then.

Selling the “you can feel good about the amount you’re donating to charity” product for 10% produces higher profits for the charity industry than selling it for 100%, at least if many people are like me.

III.

I can see an argument for an even looser standard: you should aim to be above average.

This is a very low bar. I think you might beat the average person on animal rights activism just by not stomping on anthills. The yoke here is *really* mild.

But if you believe in something like universalizability or the categorical imperative, “act in such a way that you are morally better than average” is a really interesting maxim! If everyone is just trying to be in the moral upper half of the population, the population average morality goes up. And up. And up. There’s no equilibrium other than universal sainthood.

This sounds silly, but I think it might have been going on over the past few hundred years in areas like racism and sexism. The anti-racism crusaders of yesteryear were, by our own standards, horrendously racist. But they were the good guys, fighting people even more racist than they were, and they won. Iterate that process over ten or so generations, and you reach the point where you’ve got to run your Halloween costume past your Chief Diversity Officer.

Another good thing about the number 50% is that it means there will always be as many good people as bad people. This can prove helpful if, for example, the bad people don’t like being called bad people, and want to take over the whole process of moral progress so they can declare themselves the good guys.

I’m not saying the bar should be set exactly at average. For one thing, this would mean that we could never have a situation where everyone was good enough. I think the American people are basically okay on the issue of fratricide, and I don’t want to accidentally imply that the least anti-fratricide 50% of people should feel bad about themselves. For another thing, it might be possible to move the generational process of moral progress faster if the bar is set at a higher level.

But I think these considerations at least suggest that the most effective place to set the bar might be lower than we would naively expect. I think this is how I treat people in real life. I have trouble condemning anyone who is doing more than the median. And although I don’t usually go around praising people, anybody who is doing more than the average for their time and place seems metaphorically praiseworthy to me.

IV.

A friend brings up an objection: even if low standards extract the most units of moral effort from the population, that might not be what’s important. In some cases, it’s more important that a few very important people put forth an extraordinary effort than that everyone does okay. For example, whether a billionaire donates some high percent of their money matters more than if you or I do. And whether a brilliant scientist devotes their career to fighting disease or existential risk matters much more than the rest of us.

I’m still trying to think about this, but naively it seems that we can treat the set of all exceptional people the same as the set of all people in general, and standards which extract the most moral value from one will also apply to the other. This especially makes sense if the standards are normed to effort rather than absolute results – for example, “everyone should donate 10%” extends to billionaires better than “everyone should donate $100”; diminishing marginal utility issues do argue that billionaires should donate more than that, but once you find the right unit the same argument should work.

One exception might be if we would otherwise hold exceptional people to higher standards. For example, if everybody tries to pressure a billionaire to donate most of their fortune, or on a brilliant scientist to work very hard to fight disease, then having a universal standard that it’s okay to do a little more than average might make this pressure less effective. Maybe trying to hold the average person to high standards usually backfires, but everyone would be able to successfully gang up on exceptional people to enforce high standards for *them*. If this were true, then a low bar might hold in Mediocristan, and a higher bar in Extremistan. Things like “not eating animal products” or “not driving an SUV” are in Mediocristan; things like curing cancer or donating wealth might be in Extremistan.

Another exception would be if it’s more important to extract many units of moral worth from the same people, rather than distributed across the population. For example, if cancer research (or AI research) turns out to be the most important thing, then even in the counterfactual world where everyone is equally intelligent, it’s important that somebody be the person to spend five years laboriously learning the basics so that they can make progress, *and then* that this person spend as much effort as possible doing the research and exploiting their training. A random person donating one hour of their time here or there is almost useless in comparison. In this case, we might want to have the highest standards possible, since a world in which most people give up and do nothing (but a few people accept the challenge and do a lot) is better than the alternative where everybody does a tiny bit. Again, this seems more applicable to issues like curing cancer or aligning AI than ones like not driving an SUV or using products made with unfair labor practices.

V.

I tried being vegetarian for a long time. Given that I can’t eat any vegetables (something about the combination of the bitter taste and the texture; they almost universally make me gag), this was really hard and I kept giving up and going back to subsisting mostly off of meat. The cost of “you can feel good about the amount you’re doing to not contribute to animal suffering” was apparently higher than I could afford.

For the past year, I’ve been following a more lax rule: I can’t eat any animal besides fish at home, but I can have meat (other than chicken) at restaurants. I’ve *mostly* been able to keep that rule, and now I’m eating a lot less meat than I did before.

This is a pathetic rule compared to even real pescetarians, let alone real vegetarians, let alone vegans. But I can tell this is right on the border of what I’m capable of doing; every time I go to the supermarket, I have an intense debate with the *yetzer hara* about whether I should buy meat to eat at home that week, and on rare occasions I give into temptation. But in general I hold back. And part of what holds me back is that I let myself feel like I am being good and helping save animals and have the right to feel proud when I keep my rule, and I beat myself up and feel bad and blameworthy when I break it.

I am sure any serious animal rights activist still thinks I am scum. Possibly there is an objective morality, and it agrees I am scum. But if I am right that this is the strictest rule I can keep, then I'm not sure who it benefits to remind me that I am scum. Deny me the right to feel okay when I do my half-assed attempt at virtue, and I will just make no attempt at virtue, and this will be worse for me *and* worse for animals.

Sticking to the economics metaphor, this is price discrimination. Companies try to figure out tricks to determine how rich you are, and then charge rich people more and poor people less: the goal is everyone buying the product for the highest price that they, personally, are willing to pay. Society should sell me "feel good about the amount you're doing for animals" status for the highest price that leaves me preferring buying the product to not doing so. If someone else is much richer in willpower, or just likes vegetables more, society should charge them more.

Companies rarely try price discrimination except in the sneakiest and most covert ways, because it's hard to do well and it makes everybody angry. Society does whatever it does – empirically, not care about animals unless someone is torturing a puppy or something. But we-as-members-of-society have the ability to practice price discrimination on ourselves-as-individuals, and sell our own right to feel okay about ourselves for whatever amount we as experts in our own preferences believe we can bear.

I don't know the answer to the question of where "we" "as" "a" "society" "should" "set" "moral" "standards". But if you're interested in the question, and you have a good sense for what you are and aren't capable of, maybe practicing price discrimination on yourself is the way to go.

Impending Survey Discussion Thread

Sorry, decreased blogging this week because of Thanksgiving.

But I am going to post a new SSC survey in a few weeks. Feel free to use this thread to tell me what you want – from questions you want to see, to methodology issues that bothered you on past surveys, to whatever.

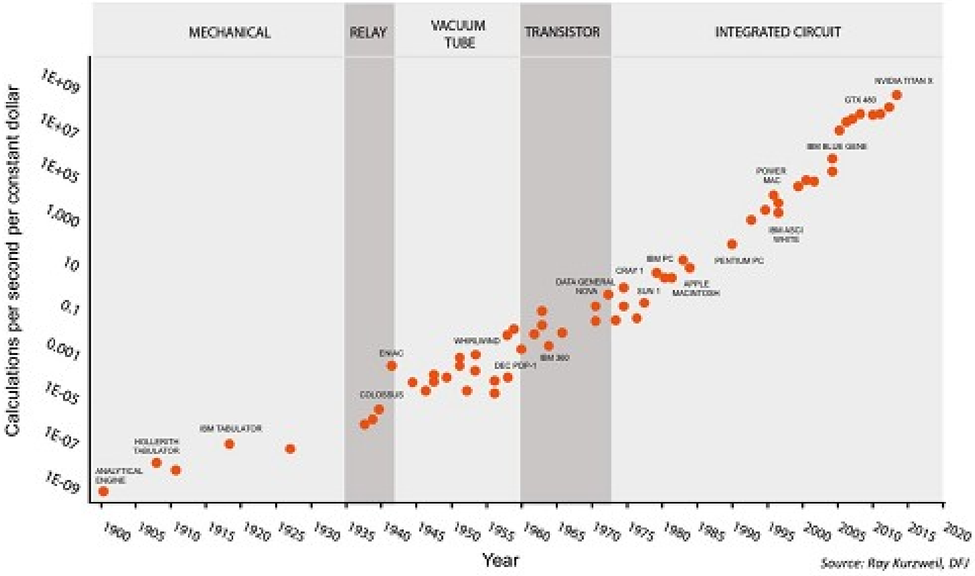
Keep in mind I will probably have to ignore the overwhelming majority of suggestions here.

Is Science Slowing Down?

[This post was up a few weeks ago before getting taken down for complicated reasons. They have been sorted out and I'm trying again.]

Is scientific progress slowing down? I recently got a chance to attend a conference on this topic, centered around a paper by Bloom, Jones, Reenen & Webb (2018).

BJRW identify areas where technological progress is easy to measure – for example, the number of transistors on a chip. They measure the rate of progress over the past century or so, and the number of researchers in the field over the same period. For example, here's the transistor data:



Source: Ray Kurzweil, DFJ

This is the standard presentation of Moore's Law – the number of transistors you can fit on a chip doubles about every two years (eg grows by 35% per year). This is usually presented as an amazing example of modern science getting things right, and no wonder – it means you can go from a few thousand transistors per chip in 1971 to many million today, with the corresponding increase in computing power.

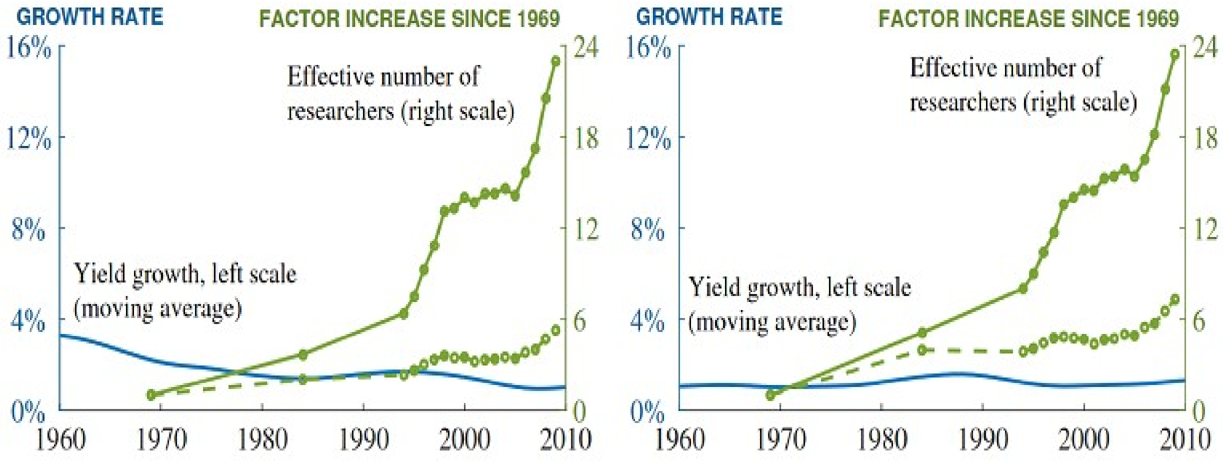
But BJRW have a pessimistic take. There are eighteen times more people involved in transistor-related research today than in 1971. So if in 1971 it took 1000 scientists to increase transistor density 35% per year, today it takes 18,000 scientists to do the same task. So apparently the average transistor scientist is eighteen times less productive today than fifty years ago. That should be surprising and scary.

But isn't it unfair to compare percent increase in transistors with absolute increase in transistor scientists? That is, a graph comparing absolute number of transistors per chip vs. absolute number of transistor scientists would show two similar exponential trends. Or a graph comparing percent change in transistors per year vs. percent change in number of transistor scientists per year would show two similar linear trends. Either way, there would be no problem and productivity would appear constant since 1971. Isn't that a better way to do things?

A lot of people asked paper author Michael Webb this at the conference, and his answer was no. He thinks that intuitively, each "discovery" should decrease transistor size by a certain amount. For example, if you discover a new material that allows transistors to be 5% smaller along one dimension, then you can fit 5% more transistors on your chip whether there were a hundred there before or a million. Since the relevant factor is discoveries per researcher, and each discovery is represented as a percent change in transistor size, it makes sense to compare percent change in transistor size with absolute number of researchers.

Anyway, most other measurable fields show the same pattern of constant progress in the face of exponentially increasing number of researchers. Here's BJRW's data on crop yield:

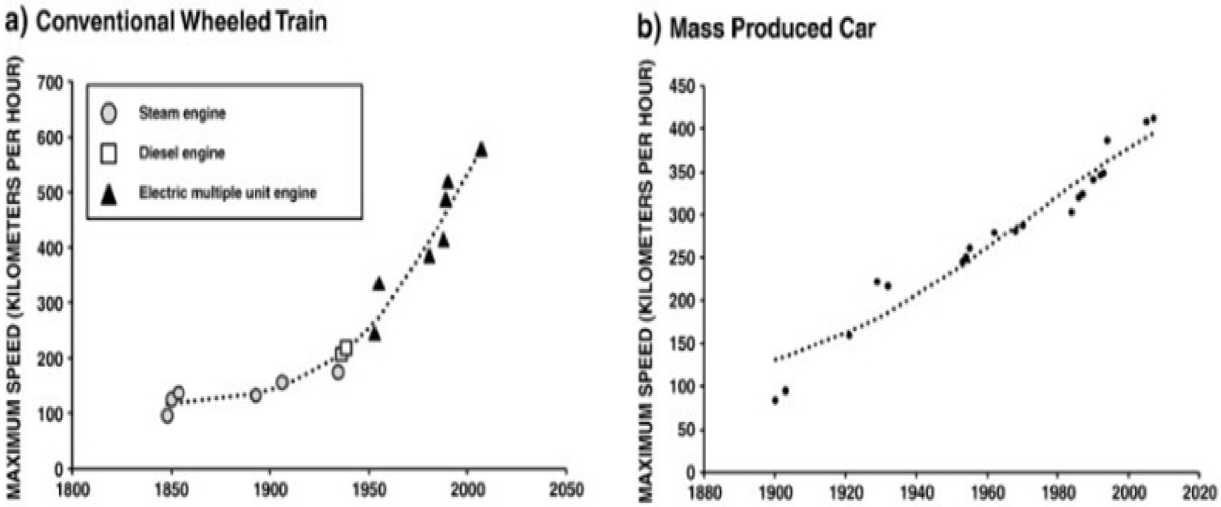
Figure 7: Yield Growth and Research Effort by Crop



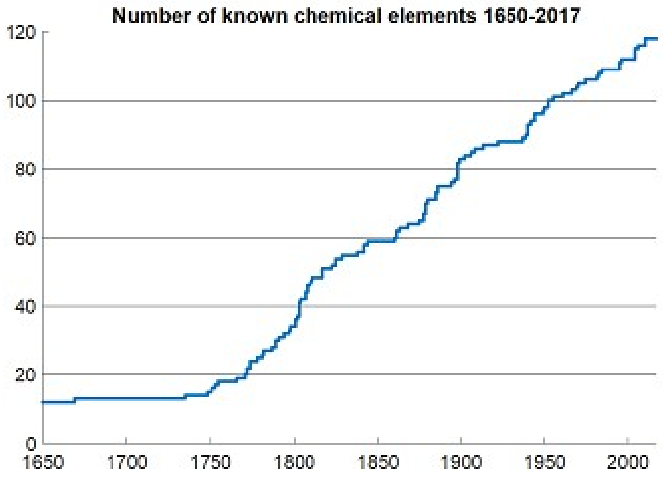
(a) Corn (b) Soybeans

The solid and dashed lines are two different measures of crop-related research. Even though the crop-related research increases by a factor of 6-24x (depending on how it’s measured), crop yields grow at a relatively constant 1% rate for soybeans, and apparently declining 3%ish percent rate for corn.

BJRW go on to prove the same is true for whatever other scientific fields they care to measure. Measuring scientific progress is inherently difficult, but their finding of constant or log-constant progress in most areas accords with Nintil’s overview of the same topic, which gives us graphs like

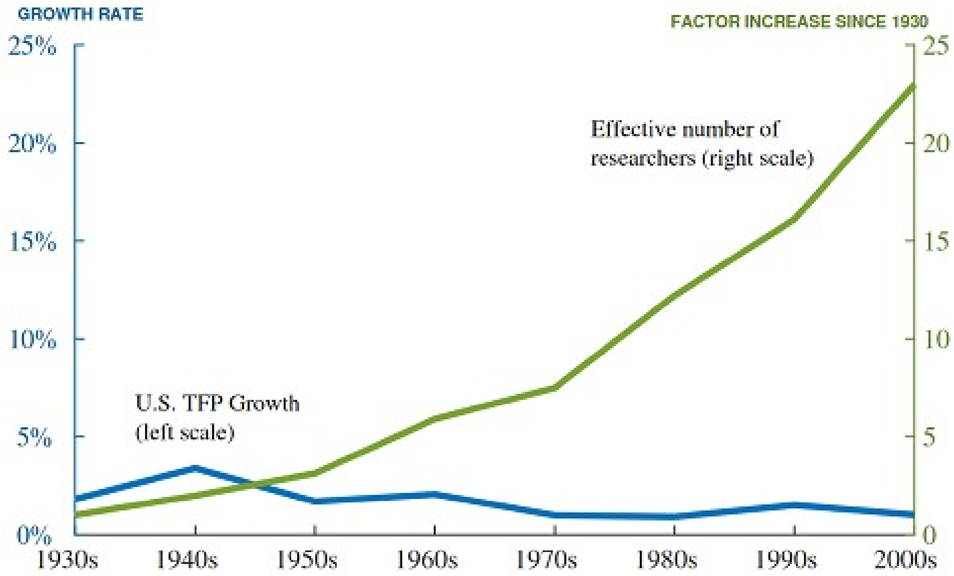


…and dozens more like it. And even when we use data that are easy to measure and hard to fake, like number of chemical elements discovered, we get the same linearity:



Meanwhile, the increase in researchers is obvious. Not only is the population increasing (by a factor of about 2.5x in the US since 1930), but the percent of people with college degrees has quintupled over the same period. The exact numbers differ from field to field, but orders of magnitude increases are the norm. For example, the number of people publishing astronomy papers seems to have dectupled over the past fifty years or so.

BJRW put all of this together into total number of researchers vs. total factor productivity of the economy, and find…



…about the same as with transistors, soybeans, and everything else. So if you take their methodology seriously, over the past ninety years, each researcher has become about 25x less productive in making discoveries that translate into economic growth.

Participants at the conference had some explanations for this, of which the ones I remember best are:

1. Only the best researchers in a field actually make progress, and the best researchers are already in a field, and probably couldn't be kept *out of* the field with barbed wire and attack dogs. If you expand a field, you will get a bunch of merely competent careerists who treat it as a 9-to-5 job. A field of 5 truly inspired geniuses and 5 competent careerists will make X progress. A field of 5 truly inspired geniuses and 500,000 competent careerists will make the same X progress. Adding further competent careerists is useless for doing anything except making graphs look more exponential,

and we should stop doing it. See also Price’s Law Of Scientific Contributions.

2. Certain features of the modern academic system, like underpaid PhDs, interminably long postdocs, endless grant-writing drudgery, and clueless funders have lowered productivity. The 1930s academic system was indeed 25x more effective at getting researchers to actually do good research.

3. All the low-hanging fruit has already been picked. For example, element 117 was discovered by an international collaboration who got an unstable isotope of berkelium from the single accelerator in Tennessee capable of synthesizing it, shipped it to a nuclear reactor in Russia where it was attached to a titanium film, brought it to a particle accelerator in a different Russian city where it was bombarded with a custom-made exotic isotope of calcium, sent the resulting data to a global team of theorists, and eventually found a signature indicating that element 117 had existed for a few milliseconds. Meanwhile, the first modern element discovery, that of phosphorous in the 1670s, came from a guy looking at his own piss. We should not be surprised that discovering element 117 needed more people than discovering phosphorous.

Needless to say, my sympathies lean towards explanation number 3. But I worry even this isn’t dismissive enough. My real objection is that constant progress in science in response to exponential increases in inputs ought to be our null hypothesis, and that it’s almost inconceivable that it could ever be otherwise.

Consider a case in which we extend these graphs back to the beginning of a field. For example, psychology started with Wilhelm Wundt and a few of his friends playing around with stimulus perception. Let’s say there were ten of them working for one generation, and they discovered ten revolutionary insights worthy of their own page in Intro Psychology textbooks. Okay. But now there are about a hundred thousand experimental psychologists. Should we expect them to discover a hundred thousand revolutionary insights per generation?

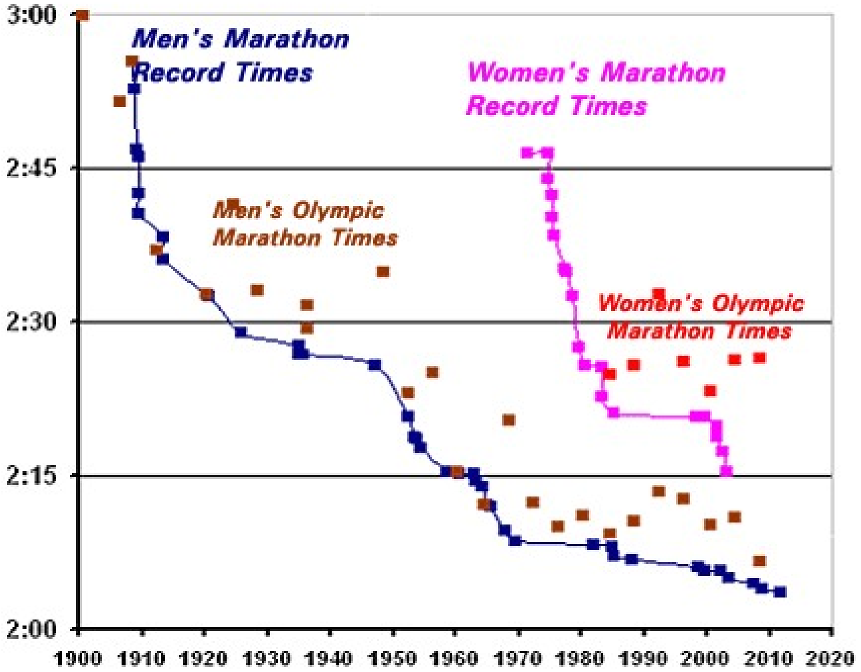
Or: the economic growth rate in 1930 was 2% or so. If it scaled with number of researchers, it ought to be about 50% per year today with our 25x increase in researcher number. That kind of growth would mean that the average person who made $30,000 a year in 2000 should make $50 million a year in 2018.

Or: in 1930, life expectancy at 65 was increasing by about two years per decade. But if that scaled with number of biomedicine researchers, that should have increased to ten years per decade by about 1955, which would mean everyone would have become immortal starting sometime during the Baby Boom, and we would currently be ruled by a deathless God-Emperor Eisenhower.

Or: the ancient Greek world had about 1% the population of the current Western world, so if the average Greek was only 10% as likely to be a scientist as the average modern, there were only 1/1000th as many Greek scientists as modern ones. But the Greeks made such great discoveries as the size of the Earth, the distance of the Earth to the sun, the prediction of eclipses, the heliocentric theory, Euclid’s geometry, the nervous system, the cardiovascular system, etc, and brought technology up from the Bronze Age to the Antikythera mechanism. Even adjusting for the long time scale to which “ancient Greece” refers, are we sure that we’re producing 1000x as many great discoveries as they are? If we extended BJRW’s graph all the way back to Ancient Greece, adjusting for the change in researchers as civilizations rise and fall, wouldn’t it keep the same shape as does for this century? Isn’t the real question not “Why isn’t Dwight Eisenhower immortal god-emperor of Earth?” but “Why isn’t Marcus Aurelius immortal god-emperor of Earth?”

Or: what about human excellence in other fields? Shakespearean England had 1% of the population of the modern Anglosphere, and presumably even fewer than 1% of the artists. Yet it gave us Shakespeare. Are there a hundred Shakespeare-equivalents around today? This is a harder problem than it seems – Shakespeare has become so venerable with historical hindsight that maybe nobody would acknowledge a Shakespeare-level master today even if they existed – but still, a hundred Shakespeares? If we look at some measure of great works of art per era, we find past eras giving us far more than we would predict from their population relative to our own. This is very hard to judge, and I would hate to be the guy who has to decide whether Harry Potter is better or worse than the *Aeneid*. But still? *A hundred Shakespeares?*

Or: what about sports? Here’s marathon records for the past hundred years or so:



In 1900, there were only two local marathons (eg the Boston Marathon) in the world. Today there are over 800. Also, the world population has increased by a factor of five (more than that in the East African countries that give us literally 100% of top male marathoners). Despite that, progress in marathon records has been steady or declining. Most other Olympics sports show the same pattern.

All of these lines of evidence lead me to the same conclusion: constant growth rates in response to exponentially increasing inputs is the null hypothesis. If it wasn't, we should be expecting 50% year-on-year GDP growth, easily-discovered-immortality, and the like. Nobody expected that before reading BJRW, so we shouldn't be surprised when BJRW provide a data-driven model showing it isn't happening. I realize this in itself isn't an explanation; it doesn't tell us why researchers can't maintain a constant level of output as measured in discoveries. It sounds a little like "God wouldn't design the universe that way", which is a kind of suspicious line of argument, especially for atheists.

But it at least shifts us from a lens where we view the problem as “What three tweaks should we make to the graduate education system to fix this problem *right now*?” to one where we view it as “Why isn’t Marcus Aurelius immortal?”

And through such a lens, only the “low-hanging fruits” explanation makes sense. Explanation 1 – that progress depends only on a few geniuses – isn’t enough. After all, the Greece-today difference is partly based on population growth, and population growth should have produced proportionately more geniuses. Explanation 2 – that PhD programs have gotten worse – isn’t enough. There would have to be a worldwide monotonic decline in every field (including sports and art) from Athens to the present day. Only Explanation 3 holds water.

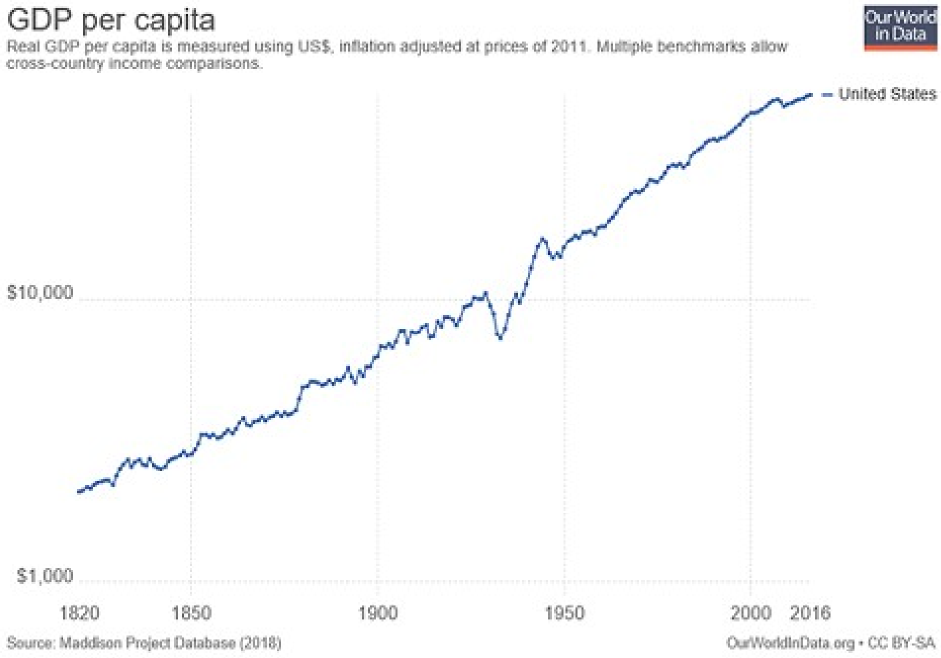
I brought this up at the conference, and somebody reasonably objected – doesn’t that mean science will stagnate soon? After all, we can’t keep feeding it an exponentially increasing number of researchers forever. If nothing else stops us, then at some point, 100% (or the highest plausible amount) of the human population will be researchers, we can only increase as fast as population growth, and then the scientific enterprise collapses.

I answered that the Gods Of Straight Lines are more powerful than the Gods Of The Copybook Headings, so if you try to use common sense on this problem you will fail.

Imagine being a futurist in 1970 presented with Moore’s Law. You scoff: “If this were to continue only 20 more years, it would mean a million transistors on a single chip! You would be able to fit an entire supercomputer in a shoebox!” But common sense was wrong and the trendline was right.

"If this were to continue only 40 more years, it would mean ten *billion* transistors per chip! You would need more transistors on a single chip than there are humans in the world! You could have computers more powerful than any today, that are too small to even see with the naked eye! You would have transistors with like a double-digit number of atoms!" But common sense was wrong and the trendline was right.

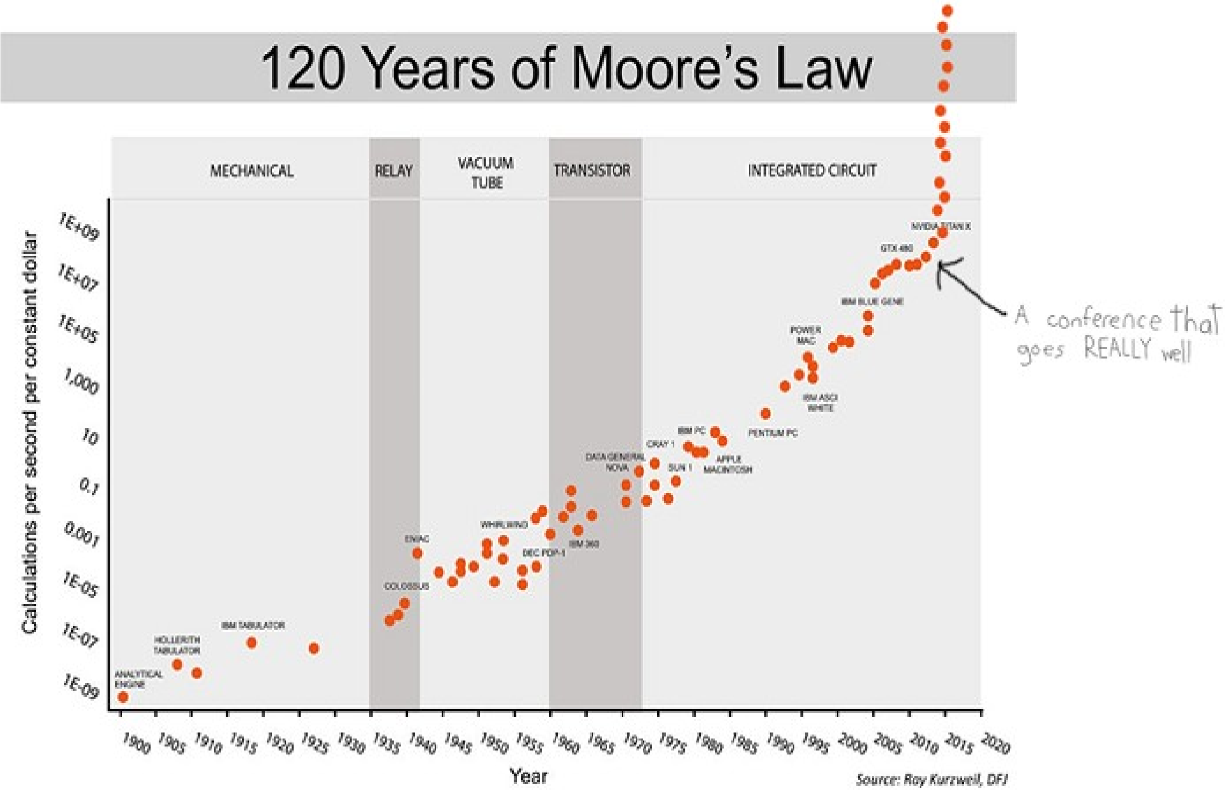
Or imagine being a futurist in ancient Greece presented with world GDP doubling time. Take the trend seriously, and in two thousand years, the future would be fifty thousand times richer. Every man would live better than the Shah of Persia! There would have to be so many people in the world you would need to tile entire countries with cityscape, or build structures higher than the hills just to house all of them. Just to sustain itself, the world would need transportation networks orders of magnitude faster than the fastest horse. But common sense was wrong and the trendline was right.



I’m not saying that no trendline has ever changed. Moore’s Law seems to be legitimately slowing down these days. The Dark Ages shifted every macrohistorical indicator for the worse, and the Industrial Revolution shifted every macrohistorical indicator for the better. Any of these sorts of things could happen again, easily. I’m just saying that “Oh, that exponential trend can’t *possibly* continue” has a really bad track record. I do not understand the Gods Of Straight Lines, and honestly they creep me out. But I would not want to bet against them.

Grace et al’s survey of AI researchers show they predict that AIs will start being able to do science in about thirty years, and will exceed the productivity of human researchers in every field shortly afterwards. Suddenly “there aren’t enough humans in the entire world to do the amount of research necessary to continue this trend line” stops sounding so compelling.

At the end of the conference, the moderator asked how many people thought that it was possible for a concerted effort by ourselves and our institutions to “fix” the “problem” indicated by BJRW’s trends. Almost the entire room raised their hands. Everyone there was smarter and more prestigious than I was (also richer, and in many cases way more attractive), but with all due respect I worry they are insane. This is kind of how I imagine their worldview looking:

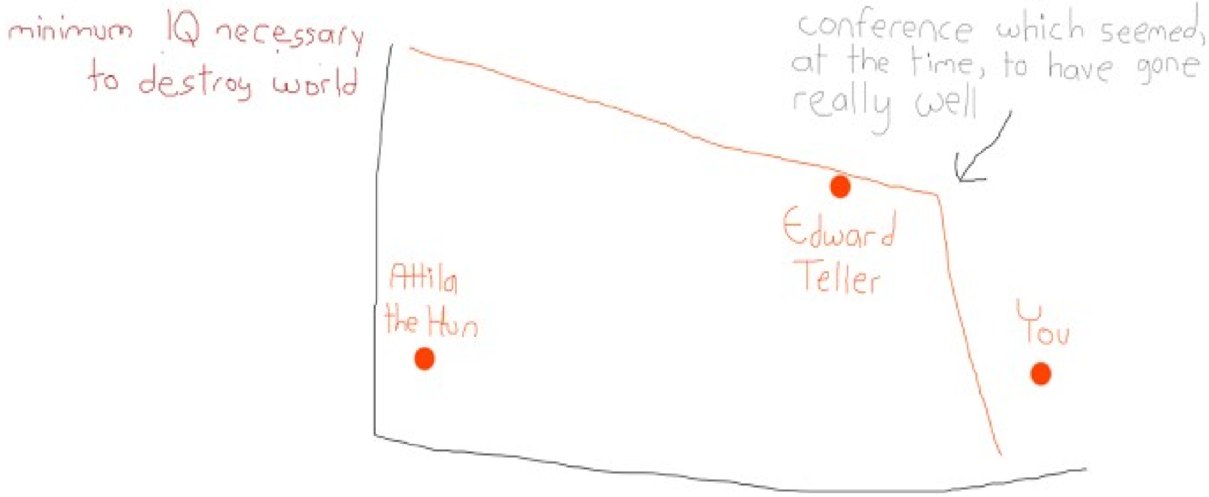


Source: Ray Kurzweil, DFJ

I realize I'm being fatalistic here. Doesn't my position imply that the scientists at Intel should give up and let the Gods Of Straight Lines do the work? Or at least that the head of the National Academy of Sciences should do something like that? That Francis Bacon was wasting his time by inventing the scientific method, and Fred Terman was wasting his time by organizing Silicon Valley? Or perhaps that the Gods Of Straight Lines were acting *through* Bacon and Terman, and they had no choice in their actions? How do we know that the Gods aren't acting through our conference? Or that our studying these things isn't the only thing that keeps the straight lines going?

I don't know. I can think of some interesting models – one made up of a thousand random coin flips a year has some nice qualities – but I don't know.

I do know you should be careful what you wish for. If you "solved" this "problem" in classical Athens, Attila the Hun would have had nukes. Remember Yudkowsky's Law of Mad Science: "Every eighteen months, the minimum IQ necessary to destroy the world drops by one point." Do you really want to make that number ten points? A hundred? I am kind of okay with the function mapping number of researchers to output that we have right now, thank you very much.



The conference was organized by Patrick Collison and Michael Nielsen; they have written up some of their thoughts here.

Book Review: The Mind Illuminated

I.

The Mind Illuminated is a guide to Buddhist meditation by Culadasa, aka John Yates, a Buddhist meditation teacher who is also a neuroscience PhD. At this point I would be more impressed to meet a Buddhist meditation teacher who *wasn't* a neuroscience PhD. If I ever teach Buddhist meditation, this is going to be my hook. "Come learn advanced meditation techniques with Scott Alexander, whose lack of a neuroscience PhD gives him a unique perspective that combines ancient wisdom with a lack of modern brain science." I think the world is ready for someone to step into this role. But Culadasa is not that person, and *The Mind Illuminated* is not that book.

Tradition divides meditation into two parts: concentration meditation, where you sharpen and control your focus, versus insight meditation, where you investigate the nature of perception and reality. TMI follows a long tradition of focusing on concentration meditation, with the assumption that insight meditation will become safer and easier once you've mastered concentration, and maybe partly take care of itself. Its course divides concentration meditation into ten stages. Early stages contain basic tasks like setting up a practice, focusing on the breath, and overcoming distractability. Later stages are more interesting; the ninth stage is learning how to calm the intensity of your meditative joy; apparently without special techniques "overly intense joy" becomes a big problem.

I usually hate meditation manuals, because they sound like word salad. "One attains joy by combining pleasure with happiness. Pleasure is a state of bliss which occurs when one concentrates focus on the understanding of awareness. Happiness is a state of joy that occurs when one focuses concentration on the awareness of understanding. By focusing awareness on bliss, you can increase the pleasure of understanding, which in turn causes concentration to be pleasant and joy to be blissful, and helps you concentrate on understanding your awareness of happiness about the bliss of focus." At some point you start thinking "Wait, were all the nouns in that paragraph synonyms for each other?"

Culadasa avoids this better than most people. Whenever he introduces a term, he puts it in ***bolded italicized letters***, and includes it in a glossary at the back. He tries to stick to multiple-word-phrases that help clarify the concept, like "bliss of physical pliancy" or "meditative joy", instead of just calling one thing "joy" and the other thing "bliss" and hoping you remember which is which. He includes a section on what he means by distinguishing "awareness" from "attention", and admits that some of these are tough choices that do not necessarily cooperate with the spirit of the English language. And his division of the material into stages helps ensure you're not reading a term until you're somewhere around the point of personally experiencing the quality being discussed.

This is characteristic of the level of care taken in this book, which despite its unfortunate acronym does a good job of presenting just the right amount of information. For example, when people say "meditate on the breath", I can only do this for a little while until I notice that the breath doesn't really exist as a specific object you can concentrate on. Really there are just a bunch of disconnected sensations changing at every moment. What do you concentrate on? I had previously dismissed this as one of several reasons why obsessive-compulsive people shouldn't do meditation, but TMI describes exactly this issue, says that it is normal and correct to worry about it, and prescribes solutions: concentrate on the disconnected sensations of the breath in whatever way feels easiest for the first few stages, and once you've increased awareness to the point where you can notice each subpart of the breath individually, do that.

II.

TMI also solves a whole slew of my obsessive questions and concerns with its "attention vs. awareness" dichotomy.

I had always been confused by instructions like "concentrate on the breath until you feel joy, then notice the joy". Usually what would happen was: I would concentrate on the breath, ask myself "am I feeling joy yet?", spend some time trying to figure this out, realize my attention had deviated from the breath, put my attention back on the breath, then feel bad because I wasn't checking to see if I was feeling joy or not. How could I both have 100% of my attention on the breath, but also be checking my joy? If I came up with the policy "check once per minute for joy, then go back to the breath", how would I avoid checking arbitrarily often whether it felt like a minute had gone by? This was another issue I just dismissed as "maybe meditation is not for obsessive-compulsive people".

But TMI distinguishes between "attention" (sometimes "focused attention") as the one thing in the foreground of your brain, and awareness (sometimes "peripheral awareness") as the potentially many things in the background of your brain. Think of it working the same way as central vs. peripheral vision. When given instructions like "concentrate on the breath until you feel joy, then notice the joy", you should be focusing your entire attention on the breath, but potentially noticing joy in your peripheral awareness. These instructions are no more contradictory than “look at this dot on the wall straight ahead, but notice if a dog runs past”.

The book urges meditators to *avoid* a state of hyperfocus in which they are so intent on the breath that they would not notice the house falling down around them. It says this is a trap that will not build the proper habits of mind to continue to higher stages and do insight meditation later. It recommends instead a form of practice in which meditators, while keeping their attention on the breath, are constantly monitoring both for external events like barking dogs or the house falling, and for internal events like feeling hungry or having thoughts. This last one sort of makes me want to scream: how can I monitor whether or not I am having thoughts without thinking about it, in which case the answer is always ‘yes’? But this is exactly the kind of paradox that the attention/awareness dichotomy is supposed to overcome. You can keep attention on the breath, notice a thought arising in the periphery of your awareness, and gently note it and push it away, all without shifting attention.

Culadasa is very excited about this:

> One great example of [a new perspective] is the distinction I make in this book between attention and awareness. Despite hundreds of thousands of meditators practicing over millennia, it has never before been clearly conceptualized that the ordinary mind has two distinct ways of “knowing”, even though these different ways of knowing have so much to do with achieving the goals of meditation. However, cognitive psychology and neuroscience have recently shown that there are two distinctly different kinds of knowing that involve completely different parts of the brain. This is a finding that deeply informs new ways of practicing

> meditation and interpreting our meditation experiences, from beginner to adept. This is only one example, but the point should be obvious: meditation can guide and inform neuroscience, and neuroscience can do the same for meditation.

I would usually be pretty reluctant to propose that hundreds of thousands of meditators practicing over millennia had all just missed something really important. And I have to admit that in the two or three test meditations I have done since reading this, I have had as much trouble as ever with these issues, and don’t notice an attention/awareness distinction that becomes obvious now that I have the terms I need to understand it. But realistically maybe something like this *has* to be true for most discussion about meditation to make sense at all.

III.

TMI gives its model of how the mind works in six interludes distributed among the chapters on meditation advice.

It begins with a startling claim that mental time is granular, and only one item can be in consciousness per granule-moment. The seven main types of items that can occupy a moment of consciousness are sight, sound, smell, taste, touch, thought, and a “binding moment” that combines aspects of the previous six. Each moment of consciousness is completely static. The only reason things seem to move or thoughts seem to flow is because the moments of consciousness are moving from moment to moment faster than you can detect, like a movie which flips from still frame to still frame so quickly that it gets perceived as continuous action. Culadasa also compares it to a “string of beads”, with each bead being a particular kind of moment (sight, sound, etc).

There are never two things in consciousness at the same time. If you think there are, that’s either because your consciousness is switching back and forth from thing to thing so quickly that you can’t follow it, or because your consciousness is perceiving a “binding moment” that presents a single aspect including both of those things. For example, if you see a cat, and you hear a meow, you might experience a “binding moment” in which you think you hear the cat meowing, although really what has happened is SIGHT:CAT — SOUND:MEOW — BINDING:(CAT, MEOW).

This sounds to me like it completely reverses the point made in the attention/awareness dichotomy, where you can be attentive to one thing but aware of many others at the same time. After all, if consciousness can only contain one thing at a time, what room is there for peripheral awareness? Culadasa states that each individual moment is either a moment of attention, or a moment of awareness. Moments of awareness can contain many things:

> For example, say you’re sitting on a cabin deck in the mountains, gazing out at the view. Each moment of visual awareness will include a variety of objects — mountains, trees, birds, and sky—all at the same time. Auditory moments of awareness will include all the various sounds that make up the audible background — birdsong, wind in the trees, a babbling brook, and so forth—again, all at the same time. On the other hand, moments of visual attention might be restricted just to the bird you’re watching on a nearby branch. Auditory attention might include only the sounds the birds are making. Even when your attention is divided among several things at once—perhaps you’re knitting or whittling a piece of wood while you sit—moments of attention are still limited to a small number of objects. Finally, binding moments of attention and binding

> moments of awareness take the content from the preceding sensory moments and combine them into a whole: “Sitting on the deck, looking out at the mountain, while carving a piece of wood.”
>
> Now, let’s consider the second difference: the degree of mental processing in moments of awareness versus moments of attention. Individual moments of awareness provide information about a lot of things at once, but the information has only been minimally processed. The result is our familiar experience of peripheral awareness of many things in the background. However, these moments of awareness do include some simple interpretations of sense data. You may be aware that the sounds you hear are from “traffic,” or that the things in the background of your visual field are “trees.” These simple concepts help evaluate and categorize all that information, contributing to our understanding of the present context. Although these preliminary interpretations don’t usually lead to any kind of action, some part of this information is frequently referred to attention for more analysis. Other times— say, when the sound of traffic suddenly includes screeching tires — the information in peripheral awareness can trigger an automatic action, thought, or emotion, any of which can then become an object of attention.

This still seems strained, but I grudgingly admit it kind of works.

TMI builds on this idea to create the “mind-system model”, its explanation for what consciousness is and why we have it. In this model, there are many “subminds”. The book is a little vague on how many there are or what level of complexity we’re supposed to be imagining here, and whether they represent only the few most salient divisions (eg “the visual system”) or are more numerous and abstract (eg “the part of your brain that likes to play computer games”), but I get the impression it’s closer to the latter. These subminds usually do their own thing, but sometimes have conflicting agendas.

Consciousness is a neutral ground shared by all subminds:

> Here’s the picture presented so far: every sub-mind belongs either to the unconscious sensory or unconscious discriminating mind. Each sub-mind performs its own specialized task independently of others, and all at the same time. Each can project content into consciousness, as well as initiate actions. Obviously, there’s enormous potential for conflict and inefficiency, if not total chaos. This is where consciousness fits into the picture: the conscious mind provides an “interface” that allows these unconscious sub-minds to communicate with each other and work together cooperatively. With all these unconscious sub-minds working independently and at the same time, the potential for conflict is enormous. The conscious mind is what allows them to work together cooperatively.
>
> The conscious mind acts as a universal recipient of information. It can receive information from each and every separate, unconscious sub-mind. In fact, all conscious experience is simply an ongoing stream of moments of consciousness whose content has been projected into the conscious mind by unconscious sub¬minds. Then, when information enters consciousness, it becomes immediately available to all the other sub-minds. Therefore, the conscious mind also serves as a universal source of information. Because the conscious mind is both a universal recipient and a universal source of information, all the unconscious sub-

> minds can interact with each other through the conscious mind.
>
> As a helpful image, picture the whole mind-system as a kind of corporation. It is made up of different departments and their employees, each with distinct roles and responsibilities. These are the unconscious sub-minds. At the top of the corporate structure is the "boardroom," or conscious mind. The diligent employees working in their separate departments produce reports, which get sent to the boardroom to be discussed further and perhaps acted on. In other words, the unconscious sub-minds send information up into the conscious mind. The conscious mind is simply a passive "space" where all the other minds can meet. In this "boardroom of the mind" metaphor, the conscious mind is where important activities of the mind-system get brought up, discussed, and decided on. One, and only one, sub-mind can present its information at a time, and that's what creates single moments of consciousness. The object of consciousness during that moment becomes part of the current agenda, and is made simultaneously available to all the other sub-minds for further processing. In subsequent moments, they project the results of their further processing into consciousness, creating a discussion that leads to conclusions and decisions.

If this sounds familiar, it's because as far as I can tell it's a rebranding of Bernard Baars' global workspace theory of consciousness. I like global workspace theory and have always considered it the most plausible solution to the easy problems of consciousness. I'm a little bit concerned that Culadasa never mentions global workspace theory in the book, and that I've never heard of any connection between global workspace theory and Buddhism before. Not really sure what to make of this.

TMI continues:

> Just because information projected into consciousness becomes available to every sub-mind of the mind-system, that doesn't mean they all receive it. It's like a radio show: the show is being broadcast, but not everyone is tuning in to listen.

Meditation increases the degree to which individual subminds are tuned in to consciousness. Since the book will later say this is all a metaphor, I think a better way of framing this might be "increase the bandwidth of the connections between the individual subminds". When someone says meditators are "more conscious" or have "higher awareness" than non-meditators, they mean that more sub-minds are tuned in to consciousness more closely at any given time.

This accomplishes what Culadasa calls "unification of mind"; with more bandwidth, the subminds are able to resolve their conflicting priorities and act more like a single unit. This can start out sort of ugly; there can be good reasons why some heavily repressed and traumatized subminds aren't usually invited to the table, and the creation of new links between them and the global workspace feels from the inside like scary unconscious material welling up into the psyche. But this is part of the "negotiation" it takes for these subminds to unify; with enough meditation, the system will assimilate their insights and they will join the Borg like everyone else.

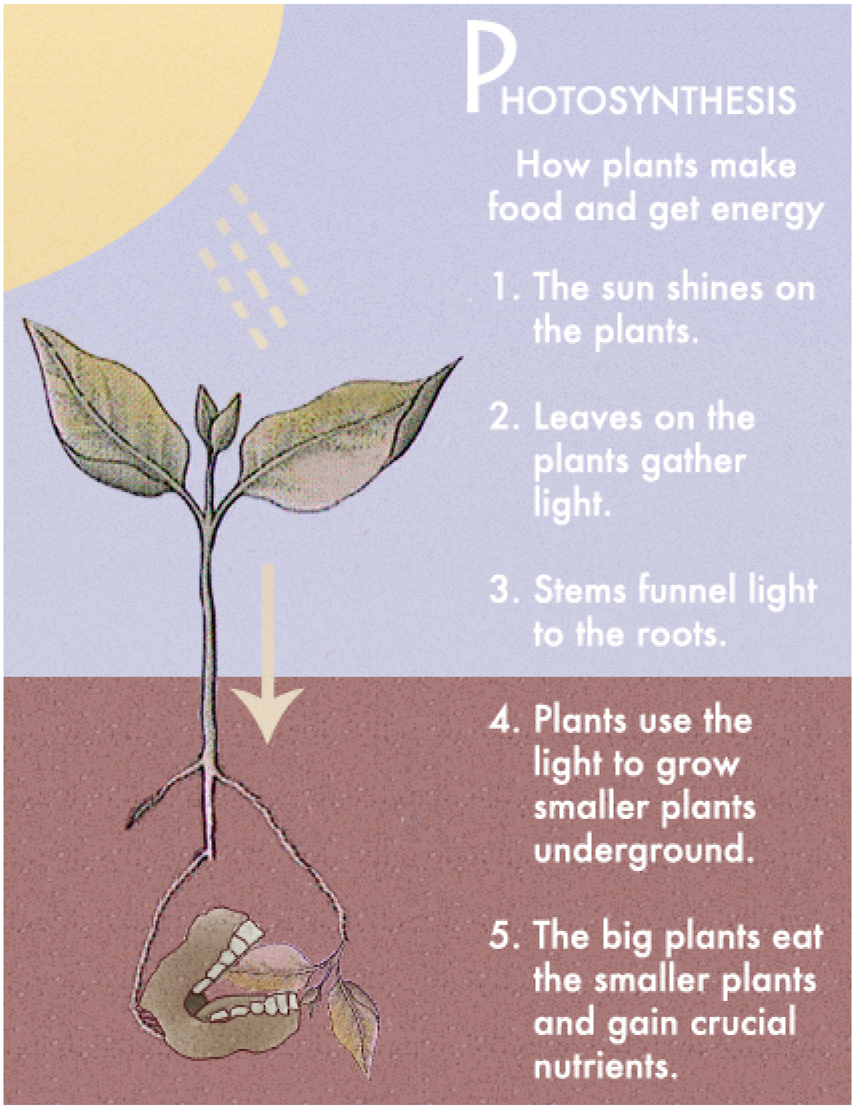
This isn't enlightenment. Enlightenment is something else. TMI calls it a "cessation event":

A cessation event is where unconscious sub-minds remain tuned in and receptive to the contents of consciousness, while at the same time, none of them project any content into consciousness. Then,consciousness ceases — completely. During that period, at the level of consciousness there is a complete cessation of mental fabrications of any kind — of the illusory, mind-generated world that otherwise dominates every conscious moment. This, of course, also entails a complete cessation of craving, intention, and suffering. The only information that tuned in sub-minds receive during this event is the fact of a total absence.

What makes this the most powerful of all Insight experiences is what happens in the last few moments of consciousness leading up to the cessation. First, an object arises in consciousness that would normally produce craving. It can be almost anything. However, what happens next is quite unusual: the mind doesn’t respond with the habitual craving and clinging. Rather, it fully understands the object from the perspective of Insight: as a mental construct, completely “empty” of any real substance, impermanent, and a cause of suffering. This profound realization leads to the next and final moment of complete equanimity, in which the shared intention of all the unified sub-minds is to not respond. Because nothing is projected into consciousness, the cessation event arises. With cessation, the tuned-in sub-minds simultaneously realize that everything appearing in consciousness is simply the product of their own activity. In other words, they realize that the input they’re accustomed to receiving is simply a result of their own fabricating activities.

I usually hate theories that explain the brain based on subminds. They seem *too easy*, in the way anthropomorphizing is always too easy. Want to run marathons, but spend your time drinking beer instead? Just model yourself as having a marathon-running submind and a beer-drinking submind, and they're fighting, and the beer-drinking submind is winning. Do this enough times and you'll never figure out anything about hyperbolic discounting or reinforcement learning or any of the very important principles that govern what the brain actually does and which do not look like little people fighting inside of you. Your solutions will always look like some weird form of therapy based on starting a dialogue with the beer-drinking submind and convincing it that beer isn't so good after all, which never works, and you'll never get around to taking Adderall, which for some reason will cause all the little men inside your head to change their opinions to whatever you wanted in the first place.

IV.


PHOTOSYNTHESIS
How plants make food and get energy
1. The sun shines on the plants.
2. Leaves on the plants gather light.
3. Stems funnel light to the roots.
4. Plants use the light to grow smaller plants underground.
5. The big plants eat the smaller plants and gain crucial nutrients.

For whatever reason, TMI's mind-system model doesn't bother me as much. Maybe it's because he's not trying to invent yet another new age psychotherapy to help fight procrastination. Maybe because it's in the context of global workspace theory, which I already like. Or maybe it's because the idea of modules and processes without enough bandwidth to connect to the global workspace sounds less anthropomorphic than little people who make you drink beer because they like beer.

IV.

This is a very optimistic book.

Buddhism started out with Theravada teachers saying it would take millions of lifetimes to reach enlightenment. Then the Mahayana and Vajrayana schools started saying maybe you could reach enlightenment in *one* lifetime, if you did everything right and worked very hard. Recently I've been reading works by modern teachers like Daniel Ingram and Vinay Gupta, who compare the amount of work involved in enlightenment to the amount of work in an MD or PhD – maybe five years? But Culadasa states that "for householders who practice properly, it's possible to master the Ten Stages within a few months or years", adding in a footnote:

> The Dalai Lama has said "If one knows the nature, order, and distinctions of the levels explained above without error and cultivates calm abiding, one can easily generate faultless meditative stabilization in about a year." When I first began teaching, I also believed that with diligent practice most people should be able to master all Ten Stages in less than a year. I have since learned that this is not realistic in terms of *most* people, and making such a flat pronouncement can be discouraging for those who have been practicing much longer without attaining that mastery.

So fine, only *cool* people can get mastery in less than a year. Still, this is a dramatic promise. But then why are there so many cultures where monks study their entire lives in monasteries? Monks have big advantages over the sort of "householder" meditators Culadasa is talking about – they can meditate every waking hour, they have access to the best teachers. Surely they should all get enlightened within a few months? I have read some work on the idea of "multiple paths" and "endless dharma gates" which suggest that what is ordinarily called enlightenment is just the first and most obvious step on an endless process of personal exploration. But when I read about historical Buddhism culture, it still seems like a majority of monks at any given time are unenlightened, including those who have been at the monastery many years.

Maybe all of this Western rationality and efficiency really is that great, and by cutting out the chaff modern people can get enlightenment much faster than the ancients could? Is this true in any other field? I get the impression that modern schoolchildren still master subjects like geometry or Latin at about the same age that the medievals would, though I could be wrong about this. Maybe Culadasa was right when he claimed his book includes important distinctions that hundreds of thousands of meditators working for thousands of years have missed. Maybe the past was just stupid and anybody moderately competent can make order-of-magnitude improvements. I don't know. It seems like a pretty big claim, though.

(or maybe this is overcomplicating things. It's not necessarily contradictory to say that a talented person, practicing an hour a day, could go from "zero math" to "able to solve calculus problems" in a year, but also that the average student has been studying math for ten years and can't solve calculus problems.)

TMI also feels optimistic in comparison to another meditation book I reviewed, *Mastering the Core Teachings of the Buddha*. Its author, Daniel Ingram, counts himself as part of the same "pragmatic dharma" movement as Culadasa, and the two of them have occasionally cooperated on various things and taught together. But Ingram stresses that meditation and enlightenment do not provide many of the worldly gains their advocates promise, and in many cases can make things worse. He warns of what he called “the Dark Night”, a tendency for people midway along the path of meditation to shatter their psyches and fall into states of profound depression and agitation.

Culadasa has a rosier view of both points. He believes that the “unification of mind” produced by meditation will have its common-sense result of reducing internal conflict and improving “willpower”; it will also “overcome all harmful emotions and behavior”, leaving you with few things to worry about except the looming specter of excessive joy.

As for the Dark Night, he doesn’t like the term, and only gives it one sentence in the main text of the book plus two pages in an appendix. The two pages reassure us that enough practice in concentration meditation serves as a prophylactic:

> One of the greatadvantages of samatha [concentration meditation] is that it makes it easier to confront the Insights into impermanence, emptiness, the pervasive nature of suffering, and the insubstantiality of the Self that produce Awakening. Without samatha, these challenging Insights have the potential to send a practitioner spiraling into a “dark night of the soul”.

Since the whole book is about samatha meditation, and treats everything else as something that happens naturally while you’re doing samatha, this makes it sound pretty minimal; just do what you would be doing anyway and you’ll be fine. This is a big difference from Ingram, who thinks that explaining the risk of the Dark Night and how to get through it is one of the most important jobs of a meditation teacher. Culadasa endorses this difference:

> Have I seen in my students anything remotely resembling a “dark night” as defined above? Absolutely not. Nor can I recall ever having seen the sorts of extreme experiences of the dukkha nanas that are appearing so frequently in these online discussions.

There seems to be something of a consensus in the relevant community that Culadasa’s type of practice, which is called “wet” (ie includes concentration and jhanas) may be less likely to produce these kinds of problems than the so-called “dry insight” that Ingram discusses, and that if you’re doing everything right maybe you shouldn’t worry about it. Shinzen Young is another meditation teacher who moves in the same circles as Ingram and Culadasa. I found his perspective on this the most informative:

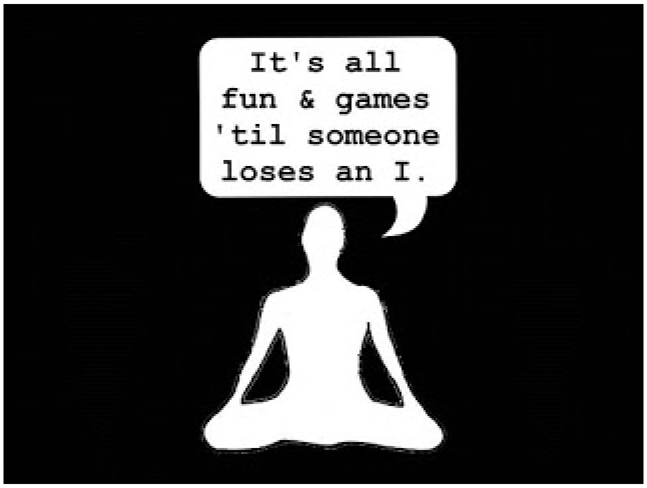
> Historically it is not a term from the Buddhist meditative tradition but rather from the Roman Catholic meditative tradition. (Of course, there’s nothing wrong with using Christian terms for Buddhist experiences but…). One must clearly define what one means by a “Dark Night” within the context of Buddhist experience.
>
> It is certainly the case that almost everyone who gets anywhere with meditation will pass through periods of negative emotion, confusion, disorientation, and heightened sensitivity to internal and external arisings. It is also not uncommon that at some point, within some domain of experience, for some duration of time, things may get worse before they get better. The same thing can happen in psychotherapy and other growth modalities. For the great majority of people, the nature, intensity, and duration of these kinds of challenges is quite manageable. I would not refer to these types of experiences as “Dark Night.”

I would reserve the term for a somewhat rarer phenomenon. This phenomenon, within the Buddhist tradition, is sometimes referred to as “falling into the Pit of the Void.” It entails an authentic and irreversible insight into Emptiness and No Self. What makes it problematic is that the person interprets it as a bad trip. Instead of being empowering and fulfilling, the way Buddhist literature claims it will be, it turns into the opposite. In a sense, it’s Enlightenment’s Evil Twin. This is serious but still manageable through intensive, perhaps daily, guidance under a competent teacher. In some cases it takes months or even years to fully metabolize, but in my experience the results are almost always highly positive. For details, see The Five Ways manual pages 97-98.

This whole Dark Night discussion reminds me of a certain Zen Koan. Although the storyline of this koan is obviously contrived, it does contain a deep message. Here’s how the koan goes: A monk is walking on a precipitous path and slips but is able to grab onto a branch by his teeth. A person standing below, recognizing the monk as an enlightened master, asks him to describe Enlightenment. What should the monk do? As a teacher, he’s duty bound to speak, but as soon as he speaks, the consequences will be dire. It sounds like a lose/lose situation. If you were the monk, what would you do? That’s the koan.

If we don’t describe the possibility of Dark Night, then we leave people without a context should it occur. On the other hand, if we do discuss it, people get scared and assume it’s going to happen to them, even if we point out (as I just did), that it’s relatively infrequent. So the take-home message is:

> 1. Don’t worry, it’s probably not going to happen to you.
> 2. Even if it does, that’s not necessarily a problem.
>
> It may require input from a teacher and time but once it’s integrated, you’ll be a very, very happy camper.
>
> I think it would be a good thing if people lighten up around this issue. This may help (see attached cartoon).


It's all fun & games 'til someone loses an I.

From this I gather that Culadasa is closer to the mainstream on this issue (also, that enlightenment does not help the mind overcome a propensity to dad jokes).

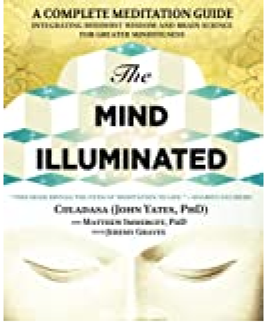
There’s a lot of drama over this issue, and if you want you can find a bunch of really enlightened and compassionate pot-shots that the different teachers are taking at each other over their respective positions. The only insight I can add to this comes from my medical experience, where I notice a very similar phenomenon in how many side effects people accord certain drugs. For example, although some people will say SSRI discontinuation syndrome is toxic and scary and omnipresent and a good reason never to use SSRIs at all, my experience in five years of taking dozens of people on and off various SSRIs is that I’ve never seen it happen beyond an occasional mild headache if the drugs are tapered properly. I know there are studies that disagree with my experience, but that is definitely my experience. Part of is is probably a difference in what kind of expectations (in yourself or your patients/students). Another part is probably a difference in what your patients/students communicate to you. A third part is probably actual differences in the way you prescribe or teach. All of these combined can be pretty powerful.

But the biggest difference I notice is that *a "serious" side effect is the one you (or one of your patients) has had, a "minor" side effect is one that you haven't.* If a certain drug works great for 95% of people, but causes a month of constant vomiting for 5%, then a doctor who's used it a few times and always gotten the great results will think of it as great (plus a rare side effect that doesn't cause lasting damage) and a patient who has been vomiting constantly for a month will think of it as an evil poison which should never have been made legal (even though most people get lucky and don't have any problems).

Shinzen says that meditation can definitely cause something terrible called "falling into the Pit of the Void", but that it usually doesn't happen, and that with daily guidance you will get better after a few months or years, and so *basically* it's not a big problem. My guess is that the person who has been trapped in some kind of weird bad trip for several months thinks of it as a very big problem, and wants everybody to warn about it all the time. All of this closely matches the way I've seen doctors and patients talk about medication side effects. I'm not sure there's a difference here except the hard-to-navigate first-person difference of "did it happen to me?"

But overall Culadasa's optimism seems justified here. Maybe it's the *only* approach to this topic that seems justified. Imagine if there was something you could do an hour a day for a year or two, which would win you more willpower and a release from all suffering, with less side effects than the average SSRI? Why aren’t we all doing it?

For more information, you can also check out Culadasa’s website and the The Mind Illuminated subreddit.


A COMPLETE MEDITATION GUIDE
The
MIND
ILLUMINATED
Culadasa (John Yates, PhD)

Book Review: Evolutionary Psychopathology

I.

Evolutionary psychology is famous for having lots of stories that make sense but are hard to test. Psychiatry is famous for having mountains of experimental data but no idea what's going on. Maybe if you added them together, they might make one healthy scientific field? Enter *Evolutionary Psychopathology: A Unified Approach* by psychology professor Marco del Giudice. It starts by presenting the theory of "life history strategies". Then it uses the theory – along with a toolbox of evolutionary and genetic ideas – to shed new light on psychiatric conditions.

Some organisms have lots of low-effort offspring. Others have a few high-effort offspring. This was the basis of the old r/k selection theory. Although the details of that theory have come under challenge, the basic insight remains. A fish will lay 10,000 eggs, then go off and do something else. 9,990 will get eaten by sharks, but that still leaves enough for there to be plenty of fish in the sea. But an elephant will spend two years pregnant, three years nursing, and ten years doing at least some level of parenting, all to produce a single big, well-socialized, and high-prospect-of-life-success calf. These are two different ways of doing reproduction. In keeping with the usual evolutionary practice, del Giudice calls the fish strategy "fast" and the elephant strategy "slow".

To oversimplify: fast strategies (think "live fast, die young") are well-adapted for unpredictable dangerous environments. Each organism has a pretty good chance of randomly dying in some unavoidable way before adulthood; the species survives by sheer numbers. Fast organisms should grow up as quickly as possible in order to maximize the chance of reaching reproductive age before they unpredictably die. They should mate with anybody around, to maximize the chance of mating before they unpredictably die. They should ignore their offspring, since they expect most offspring to unpredictably die, and since they have too many to take care of anyway. They should be willing to take risks, since the downside (death without reproducing) is already their default expectation, and the upside (becoming one of the few individuals to give birth to the 10,000 offspring of the next generation) is high.

Slow strategies are well-adapted for safer environments, or predictable complex environments whose intricacies can be mastered with enough time and effort. Slow strategy animals may take a long time to grow up, since they need to achieve mastery before leaving their parents. They might be very picky maters, since they have all the time in the world to choose, will only have a few children each, and need to make sure each of those children has the best genes possible. They should work hard to raise their offspring, since each individual child represents a substantial part of the prospects of their genetic line. They should avoid risks, since the downside (death without reproducing) would be catastrophically worse than default, and the upside (giving birth to a few offspring of the next generation) is what they should expect anyway.

Del Giudice asks: what if life history strategies differ not just across species, but across individuals of the same species? What if this theory applied within the human population?

> In line with animal research on pace-of-life syndromes, human research has shown that impulsivity, risk-taking, and sensation seeking, are systematically associated

> with fast life history traits such as early intercourse, early childbearing in females, unrestricted sociosexuality, larger numbers of sexual partners, reduced long-term mating orientation, and increased mortality. Future discounting and heightened mating competition reduce the benefits of reciprocal long-term relationships; in motivational terms, affiliation and reciprocity are downregulated, whereas status seeking and aggression are upregulated. The resulting behavioral pattern is marked by exploitative and socially antagonistic tendencies; these tendencies may be expressed in different forms in males and females, for example through physical versus relational aggression (Belsky et al 1991; Borowsky et al 2009; Brezina et al 2009; Chen & Vazsonyi 2011; Copping et al 2013a, 2013b, 2014a; Curry et al 2008; Dunkel & Decker 2010 […]

And:

> Disgust sensitivity is another dimension of individual differences with links to the fast-slow continuum. To begin, high disgust sensitivity is broadly associated with measures of risk aversion. Moral and sexual disgust correlate with higher agreeableness, conscientiousness, and honesty-humility; and sexual disgust specifically predicts restricted sociosexuality (Al-Shawaf et al 2015; Sparks et al 2018; Tybur et al 2009, 2015; Tybur & de Vries 2013). These findings suggest that the disgust system is implicated in the regulation of life-history-related behaviors. In particular, sexual and moral disgust show the most consistent pattern of correlations with other indicators of slow strategies.
>
> Romantic attachment styles have wide ranging influences on sexuality, mating, and couple stability, but

> their relations with life history strategies are somewhat complex. Secure attachment styles are consistently associated with slow life history traits (eg Chisholm 1999b; Chisholm et al 2005; Del Giudice 2990a). Avoidance predicts unrestricted sociosexuality, reduced long-term orientation, and low commitment to partners (Brennan & Shaver 1995; Jackson & Kirkpatrick 2007; Templehof & Allen 2008). Given the central role of pair bonding in long-term parental investment, avoidant attachment – which, on average, is higher in men – can be generally interpreted as a mediator of reduced parenting effort. However, some inconsistent findings indicate that avoidance may capture multiple functional mechanisms. High levels of romantic avoidance are found both in people with very early sexual debut and in those who delay intercourse (Gentzler & Kearns, 2004); this suggests that, at least for some people, avoidant attachment may actually reflect a partial downregulation of the mating system, consistent with slower life history strategies.

And:

> At a higher level of abstraction, the behavioral correlates of life history strategies can be framed within the five-factor model of personality. Among the Big Five, agreeableness and conscientiousness show the most consistent pattern of associations with slow traits such as restricted sociosexuality, long-term mating orientation, couple stability, secure attachment to parents in infancy and romantic partners in adulthood, reduced sex drive, low impulsivity, and risk aversion across domains (eg Baams et al 2004; Banai & Pavela 2015; Bourage et al 2007; DeYoung 2001; Holtzman & Strube 2013; Jonasen et al 2013 […] Some researchers working in a life history perspective have argued that the

> general factor of personality should be regarded as the core personality correlate of slow strategies.

Del Giudice suggests that these traits, and predisposition to fast vs. slow life history in general, are caused by a gene * environment interaction. The genetic predisposition is straightforward enough. The environmental aspect is more interesting.

There has been some research on the thrify phenotype hypothesis: if you're undernourished while in the womb, you'll be at higher risk of obesity later on. Some mumble "epigenetic" mumble "mechanism" looks around, says "We seem to be in a low-food environment, better design the brain and body to gorge on food when it's available and store lots of it as fat", then somehow modulates the relevant genes to make it happen.

Del Giudice seems to imply that a similar epigenetic mechanism "looks around" at the world during the first few years of life to try to figure out if you're living in the sort of unpredictable dangerous environment that needs a fast strategy, or the sort of safe, masterable environment that needs a slow strategy. Depending on your genetic predisposition and the observable features of the environment, this mechanism "makes a decision" to "lock" you into a faster or slower strategy, setting your personality traits more toward one side or the other.

He further subdivides fast vs. slow life history into four different "life history strategies".

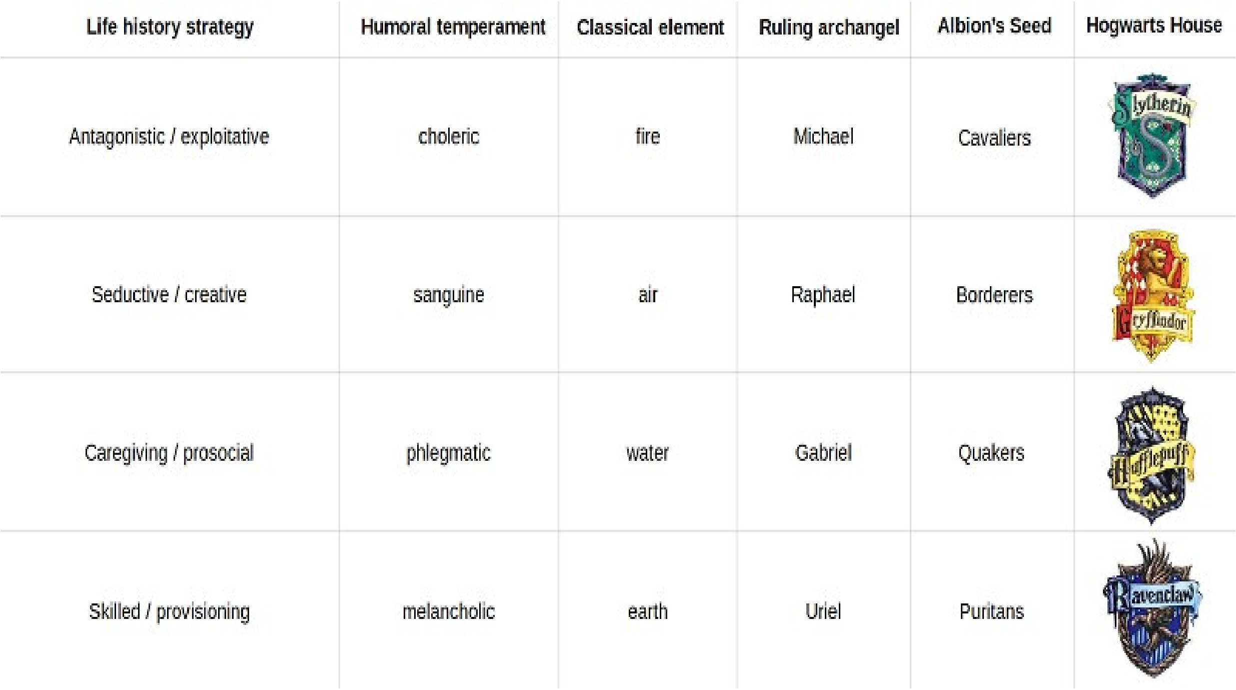
The *antagonistic/exploitative strategy* is a fast strategy that focuses on getting ahead by defecting against other people. Because it expects a short and noisy life without the kind of predictable iterated games that build reciprocity, it throws all this away and focuses on getting ahead quick. A person who has been optimized for an antagonistic/exploitative strategy will be charming, good at some superficial social tasks, and have no sense of ethics – ie the perfect con man. Antagonistic/exploitative people will have opportunities to reproduce through outright rape, through promising partners commitment and then not providing it, through status in criminal communities, or through things in the general category of hiring prostitutes when both parties are too drunk to use birth control. These people do not have to be criminals; they can also be the most cutthroat businessmen, lawyers, and politicians. Jumping ahead to the psychiatry connection, the extreme version of this strategy is probably antisocial personality disorder.

The *creative/seductive strategy* is a fast strategy that focuses on getting ahead through sexual selection, ie optimizing for being really sexy. Because it expects a short and noisy life, it focuses on raw sex appeal (which peaks in the late teens and early twenties) as opposed to things like social status or ability to care for children (which peak later in maturity). A person who has been optimized for a creative/seductive strategy will be attractive, artistic, flirtatious, and popular – eg the typical rock star or starlet. They will also have traits that support these skills, which for complicated reasons include being very emotional. Creative/seductive people will have opportunities to reproduce through making other people infatuated with them; if they are lucky, they can seduce a high-status high-resource person who can help take care of the children. The most extreme version of this strategy is probably borderline personality disorder.

The *prosocial/caregiving strategy* is a slow strategy that focuses on being a responsible pillar of the community who everybody likes. Because it expects a slow and stable life, it focuses on lasting relationships and cultivating a good reputation that will serve it well in iterated games. A person who has been optimized for a prosocial/caregiving strategy will be dependable, friendly, honest, and conformist – eg the typical model citizen. Prosocial/caregiving people will have opportunities to reproduce by marrying their high school sweetheart, living in a suburban tract house, and having 2.4 children who go to state college. The most extreme version of this strategy is probably being a normie.

The *skilled/provisioning strategy* is a slow strategy that focuses on being good at specific useful tasks. Because it expects a slow and stable life, it focuses on gaining abilities that may take years to bear any fruit. A person who is optimized for a skilled/provisioning strategy will be intelligent, good at learning, and a little bit obsessive. They may not invest as much in friendliness or seductiveness; once they succeed at their chosen path, they will get social status through being indispensible for the continued functioning of the community, and they will have opportunities to reproduce because of their high status and obvious ability to provide for the children. The most extreme version of this strategy is probably high-functioning autism.

This division into life strategies is a seductive idea. I mean, literally, it’s a seductive idea, ie in terms of memetic evolution, we may worry it is optimized for a seductive/creative strategy for reproduction, rather than the boring autistic “is actually true” strategy. The following is not a figure from Del Giudice’s book, but maybe it should be:



Life history strategy	Humoral temperament	Classical element	Ruling archangel	Albion's Seed	Hogwarts House
Antagonistic / exploitative	choleric	fire	Michael	Cavaliers	Slytherin
Seductive / creative	sanguine	air	Raphael	Borderers	Gryffindor
Caregiving / prosocial	phlegmatic	water	Gabriel	Quakers	Hufflepuff
Skilled / provisioning	melancholic	earth	Uriel	Puritans	Ravenclaw

There's a lot of debate these days about how we should treat research that fits our existing beliefs too closely. I remember Richard Dawkins (or maybe some other atheist) once argued we should be suspicious of religion because it was too normal. When you really look at the world, you get all kinds of crazy stuff like quantum physics and time dilation, but when you just *pretend* to look at the world, you get things like a loving Father, good vs. evil, and ritual purification – very human things, things a schoolchild could understand. Atheists and believers have since had many debates over whether religion is too ordinary or sufficiently strange, but I haven't heard either side deny the fundamental insight that science should do *something* other than flatter our existing categories for making sense of the world.

On the other hand, the first thermometer no doubt recorded that it was colder in winter than in summer. And if someone had criticized physicists, saying "You claim to have a new 'objective' way of looking at temperature, but really all you're doing is justifying your old prejudices that the year is divided into nice clear human-observable parts, and summer is hot and winter is cold" – then that person would be a moron.

This kind of thing keeps coming up, from Klein vs. Harris on the science of race to Jussim on stereotype accuracy. I certainly can't resolve it here, so I want to just acknowledge the difficulty and move on. If it helps, I don't think Del Giudice wants to argue these are objectively the only four possible life strategies and that they are perfect Platonic categories, just that these are a good way to think of some of the different ways that organisms (including humans) can pursue their goal of reproduction.

II.

Psychiatry is hard to analyze from an evolutionary perspective. From an evolutionary perspective, it shouldn't even exist. Most psychiatric disorders are at least somewhat genetic, and most psychiatric disorders decrease reproductive fitness. Biologists have equations that can calculate how likely it is that maladaptive genes can stay in the population for certain amounts of time, and these equations say, all else being equal, that psychiatric disorders should not be possible. Apparently all else isn't equal, but people have had a lot of trouble figuring out exactly what that means. A good example of this kind of thing is Greg Cochran's theory that homosexuality must be caused by some kind of infection; he does not see another way it could remain a human behavior without being selected into oblivion.

Del Giudice does the best he can within this framework. He tries to sort psychiatric conditions into a few categories based on possible evolutionary mechanisms.

First, there are conditions that are plausibly good evolutionary strategies, and people just don't like them. For example, nymphomania is unfortunate from a personal and societal perspective, but one can imagine the evolutionary logic checks out.

Second, there are conditions which might be adaptive in some situations, but don’t work now. For example, antisocial traits might be well-suited to environments with minimal law enforcement and poor reputational mechanisms for keeping people in check; now they will just land you in jail.

Third, there are conditions which are extreme levels of traits which it’s good to have a little of. For example, a little anxiety is certainly useful to prevent people from poking lions with sticks just to see what will happen. Imagine (as a really silly toy model) that two genes A and B determine anxiety, and the optimal anxiety level is 10. Alice has gene A = 8 and gene B = 2. Bob has gene A = 2 and gene B = 8. Both of them are happy well-adjusted individuals who engage in the locally optimal level of lion-poking. But if they reproduce, their child may inherit gene A = 8 and gene B = 8 for a total of 16, much more anxious than is optimal. This child might get diagnosed with an anxiety disorder, but it’s just a natural consequence of having genes for various levels of anxiety floating around in the population.

Fourth, there are conditions which are the failure modes of traits which it’s good to have a little of. For example, psychiatrists have long categorized certain common traits into “schizotypy”, a cluster of characteristics more common in the relatives of schizophrenics and in people at risk of developing schizophrenia themselves. These traits are not psychotic in and of themselves and do not decrease fitness, nor is schizophrenia necessarily just the far end of this distribution. But schizotypal traits are one necessary ingredient of getting schizophrenia; schizophrenia is some kind of failure mode only possible with enough schizotypy. If schizotypal traits do some other good thing, they can stick around in the population, and this will look a lot like "schizophrenia is genetic".

How can we determine which of these categories any given psychiatric disorder falls into?

One way is through what is called taxometrics – the study of to what degree mental disorders are just the far end of a normal distribution of traits. Some disorders are clearly this way; for example, if you quantify and graph everybody's anxiety levels, they will form a bell curve, and the people diagnosed with anxiety disorders will just be the ones on the far right tail. Are any disorders *not* this way? This is a hard question, though schizophrenia is a promising candidate.

Another way is through measuring the correlation of disorders with mutational load. Some people end up with more mutations (and so a generically less functional genome) than others. The most common cause of this is being the child of an older father, since that gives mutations more time to accumulate in sperm cells. Other people seem to have higher mutational load for other, unclear reasons, which can be measured through facial asymmetry and the presence of minor physical abnormalities (like weirdly-shaped ears). If a particular psychiatric disorder is more common in people with increased mutational load, that suggests it isn't just a functional adaptation but some kind of failure mode of something or other. Schizophrenia and low-functioning autism are both linked to higher mutational load.

Another way is by trying to figure out what aspect of evolutionary strategy matches the occurrence of the disorder. Developmental psychologists talk about various life stages, each of which brings new challenges. For example, adrenache (age 6-8) marks "the transition from early to middle childhood", when "behavioral plasticity and heightened social learning go hand in hand with the expression of new genetic influences on psychological traits such as agression, prosocial behavior, and cognitive skills" and children receive social feedback "about their attractiveness and competitive ability". More obviously, puberty marks the expression of still other genetic influences and the time at which young people start seriously thinking about sex. So if various evolutionary adaptations to deal with mating suddenly become active around puberty, and some mental disorder always starts at puberty, that provides some evidence that the mental disorder might be related to an evolutionary adaptation for dealing with mating. Or, since a staple of evo psych is that men and women pursue different reproductive strategies, if some psychiatric disease is twice as common in women (eg depression) or five times as common in men (eg autism), then that suggests it's correlated with some strategy or trait that one sex uses more than the other.

This is where Del Giudice ties in the life history framework. If some psychiatric disease is more common in people who otherwise seem to be pursuing some life strategy, then maybe it's related to that strategy. Either it's another name for that strategy, or it's another name for an extreme version of that strategy, or it's a failure mode of that strategy, or it's associated with some trait or adaptation which that strategy uses more than others do. By determining the association of disorders with certain life strategies, we can figure out what adaptive trait they're piggybacking on, and from there we can reverse engineer them and try to figure out what went wrong.

This is a much more well-thought-out and orderly way of thinking about psychiatric disease than anything I've ever seen anyone else try. How does it work?

Unclear. Psychiatric disorders really resist being put into this framework. For example, some psychiatric disorders have a u-shaped curve regarding childhood quality – they are more common both in people with unusually deprived childhoods *and* people with unusually good childhoods. Many anorexics are remarkably high-functioning, so much so that even the average clinical psychiatrist takes note, but others are kind of a mess. Autism is classically associated with very low IQ and with bodily asymmetries that indicate high mutational load, but a lot of autistics have higher-than-normal IQ and minimal bodily asymmetry. Schizophrenia often starts in a very specific window between ages 18 and 25, which sounds promising for a developmental link, but a few cases will start at age 5, or age 50, or pretty much whenever. Everything is like this. What is a rational, order-loving evolutionary psychologist supposed to do?

Del Giudice bites the bullet and says that most of our diagnostic categories conflate different conditions. The unusually high-functioning anorexics have a different disease than the unusually low-functioning ones. Low IQ autism with bodily asymmetries has a different evolutionary explanation than high IQ autism without. In some cases, he is able to marshal a lot of evidence for distinct clinical entities. For example, most cases of OCD start in adulthood, but one-third begin in early childhood instead. These early OCD cases are much more likely to be male, more likely to have high conscientiousness, more likely to co-occur with autistic traits, and have a different set of obsessions focusing on symmetry, order, and religion (my own OCD started in very early childhood and I feel called out by this description). Del Giudice says these are two different conditions, one of which is associated with pathogen defense and one of which is associated with a slow life strategy.

Deep down, psychiatrists know that we have not really subdivided the space of mental disorders very well. Every year a new study comes out purporting to have discovered the three types of depression, or the four types of depression, or the five types of depression, or some other number of types of depression that some scientist thinks she has discovered. Often these are given explanatory power, like “number three is the one that doesn’t respond to SSRIs”, or “1 and 2 are biological; 3, 4, and 5 are caused by life events”. All of these seem equally plausible, so given that they all say different things I tend to ignore all of them. So when del Giudice puts depression under his spotlight and finds it subdivides into many different subdisorders, this is entirely fair. Maybe we should be concerned if he *didn’t* find that.

But part of me is still concerned. If evo psych correctly predicted the characteristics of the psychiatric disorders we observe, then we would count that as theoretical confirmation. Instead, it only works after you replace the psychiatric disorders we observe with another, more subtle set right on the threshold of observation. The more you’re allowed to diverge from our usual understanding, the more chance you have to fudge your results; the more different disorders you can divide things into, the more options you have for overfitting. Del Giudice’s new schema may well be accurate; it just makes it hard to *check* his accuracy.

On the other hand, reality has a surprising amount of detail. Every previous attempt to make sense of psychopathology has failed. But psychopathology has to make sense. So it must make sense in some complicated way. If you see what looks like a totally random squiggle on a piece of paper, then probably the equation that describes it really is going to have a lot of variables, and you shouldn’t criticize a many-variable equation as “overfitting”. There is a part of me that thinks this book is a beautiful example of what solving a complicated field would look like. You take all the complications, and you explain by layering of a bunch of different simple and reasonable things on top of one another. The psychiatry parts of *Evolutionary Psychopathology: A Unified Approach* do this. I don't know if it's all just epicycles, but it's a heck of a good try.

I would encourage anyone with an interest in mental health and a tolerance for dense journal-style writing to read the psychiatry parts of this book. Whether or not the hypotheses are right, in the process of defending them it calls in such a wide array of evidence, from so many weird studies that nobody else would have any reason to think about, that it serves as a fantastic survey of the field from an unusual perspective. If you've ever wanted to know how many depressed people are reproducing (surprisingly many! about 90 – 100% as many as non-depressed people!) or what the IQ of ADHD people is (0.6 standard deviations below average; the people most of you see are probably from a high-functioning subtype) or how schizophrenia varies with latitude (triples as you move from the equator to the poles, but after adjusting for this darker-skinned people seem to have more, suggesting a possible connection with Vitamin D), this is the book for you.

III.

I want to discuss some political and social implications of this work. These are my speculations only; del Giudice is not to blame.

We believe that an abusive or deprived childhood can negatively affect people's life chances. So far, we've cached this out entirely in terms of brain damage. Children's developing brains "can't deal with the trauma" and so become "broken" in ways that make them a less functional adult. Life history theory offers a different explanation. Nothing is “broken”. Deprived children have just looked around, seen what the world is like, and rewired themselves accordingly on some deep epigenetic level.

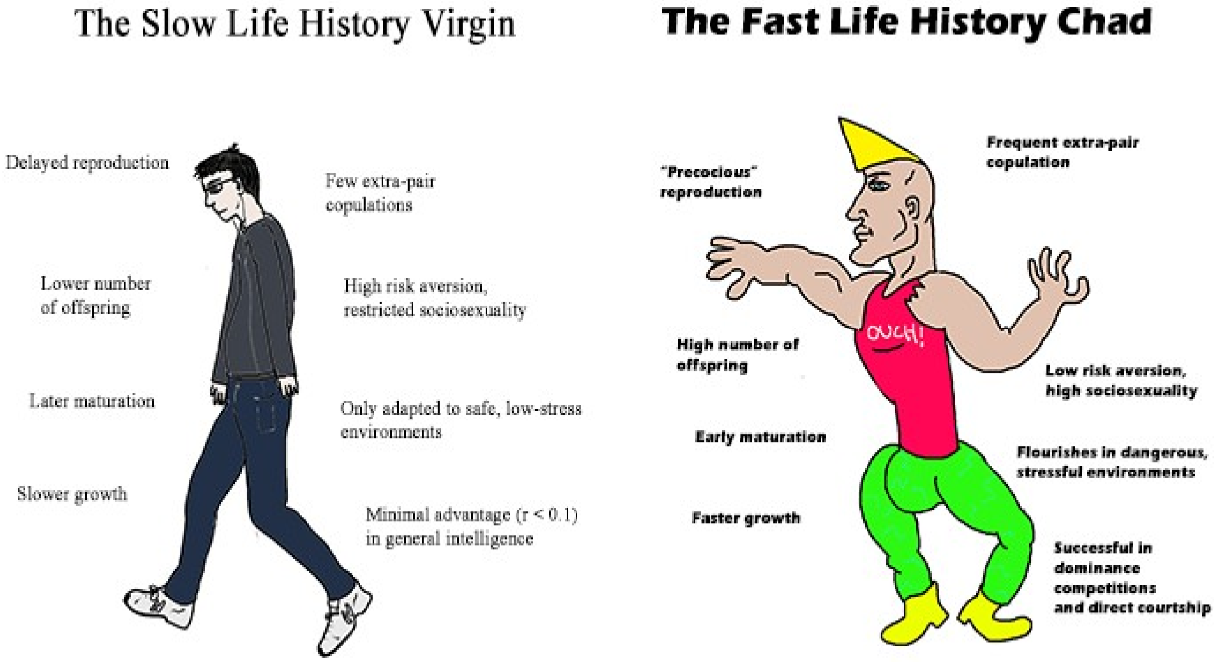
I was reading this at the same time as the studies on preschool, and I couldn’t help noticing how well they fit together. The preschool studies were surprising because we expected them to improve children’s intelligence. Instead, they improved everything else. Why? This would make sense if the safe environment of preschool wasn’t “fixing” their “broken” brains, but pushing them to follow a slower life strategy. Stay in school. Don’t commit crimes. Don’t have kids while you’re still a teenager. This is *exactly* what we expect a push towards slow life strategies to do.

Life strategies even predict the “fade-out/fade-in” nature of the effects; the theory specifies that although aspects of life strategy may be set early on, they only “activate” at the appropriate developmental period. From page 93: “The social feedback that children receive in this phase [middle childhood]…may feed into the regulation of puberty timing and shape behavioral strategies in adolescence.”

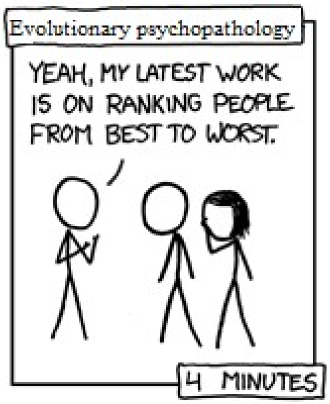
Society has done a lot to try to help disadvantaged children. A lot of research has been gloomy about the downstream effects; none of it raised anybody’s IQ, there are still lots of poor people around, income inequality continues to increase. But maybe we’re just looking in the wrong place.

On a related note: a lot of intelligent, responsible, basically decent young men complain of romantic failure. Although the media has tried hard to make this look like some kind of horrifying desire to rape everybody because they believe are entitled to whatever and whoever they want, the basic complaint is more prosaic: “I try to be a nice guy who contributes to society and respects others; how come I’m a miserable 25-year-old virgin, whereas every bully and jerk and frat bro I know is able to get a semi-infinite supply of sex partners whom they seduce, abuse, and dump?" This complaint isn't imaginary; studies have shown that criminals are more likely to have lost their virginity earlier, that boys with more aggressive and dishonest behaviors have earlier age of first sexual intercourse, and that women find men with dark triad traits more attractive. I used to work in a psychiatric hospital that served primarily adolescents with a history of violence or legal issues; most of them had had multiple sexual encounters by age fifteen; only half of MIT students in their late teens and early 20s have had sex at all.

Del Giudice's work offers a framework by which to understand these statistics. Most MIT students are probably pursuing slow life strategies; most violent adolescents in psych hospitals are probably pursuing fast ones. Fast strategies activate a suite of traits designed for having sex earlier; slow life strategies activate a suite of traits designed for preventing early sex. There's a certain logical leap here where you have to explain how, if an individual is trying very hard to have teenage sex, his mumble epigenetic mumble mechanism can somehow prevent this. But millions of very vocal people's lived experiences argue that it can. The good news for these people is that they are adapted for a life strategy which in the past has consistently resulted in reproduction at some point. Maybe when they graduate with a prestigious MIT degree, they will get enough money and status to attract a high-quality slow-strategy mate, who can bear high-quality slow-strategy kids who produce many surviving grandchildren. I don't know. This hasn't happened to me yet. Maybe I should have gone to MIT.


The Slow Life History Virgin
Delayed reproduction
Few extra-pair copulations
Lower number of offspring
High risk aversion, restricted sociosexuality
Later maturation
Only adapted to safe, low-stress environments
Slower growth
Minimal advantage (r < 0.1) in general intelligence
The Fast Life History Chad
"Precocious" reproduction
Frequent extra-pair copulation
OUCH!
High number of offspring
Low risk aversion, high sociosexuality
Early maturation
Flourishes in dangerous, stressful environments
Faster growth
Successful in dominance competitions and direct courtship

Finally, the people who like to say that various things “serve as a justification for oppression” are going to have a field day with this one. Although del Giudice is too scientific to assign any moral weight to his life history strategies, it’s not that hard to import it.


Evolutionary psychopathology
YEAH, MY LATEST WORK IS ON RANKING PEOPLE FROM BEST TO WORST.
4 MINUTES

(source)

Life strategies run the risk of reifying some of our negative judgments. If criminals are pursuing a hard-coded antagonistic-exploitative strategy, that doesn’t look good for rehabilitation. Likewise, if some people are pursuing creative-seductive strategies, that provides new force to the warning to avoid promiscuous floozies and stick to your own social class. In the extreme version of this, you could imagine a populism that claims to be fighting for the decent middle-class slow-strategy segment of the population against an antagonistic/exploitative underclass. The creative/seductive people are on thin ice – maybe they should start producing art that looks like something.

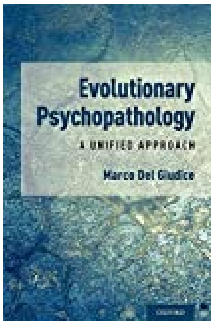
(it doesn’t help that this theory is distantly related to an earlier theory proposed by Canadian psychologist John Rushton, who added that black people are racially predisposed to fast strategies and Asians to slow strategies, with white people somewhere in the middle. Del Giudice mentions Rushton just enough that nobody can accuse him of deliberately covering up his existence, then hastily moves on.)

But aside from the psychological compellingness, this doesn’t make a lot of sense. We already know that antagonistic and exploitative people exist in the world. All that life history theory does is exactly what progressives want to do: provide an explanation that links these qualities to childhood deprivation, or to dangerous environments where they may be the only rational choice. Sure, you would have to handwave away the genetic aspect, but you’re going to have be handwaving away *some* genetics to make this kind of thing work no matter what, and life history theory makes this easier rather than harder. It also provides some testable hypotheses about what aspects of childhood deprivation we might want to target, and what kind of effects we might expect such interventions to have.

Apart from all this, I find life history strategy theory sort of reassuring. Until now, atheists have been denied the comfort of knowing God has a plan for them. Sure, they could know that evolution had a plan for them, but that plan was just “watch dispassionately to see whether they live or die, then adjust gene frequencies in the next generation accordingly”. In life history strategy theory, evolution – or at least your mumble epigenetic mumble mechanism – *actually* has a plan for you. Now we can be evangelical atheists who have a personal relationship with evolution. It’s pretty neat.

And I come at this from the perspective of someone who has failed at many things despite trying very hard, and also succeeded at others without even trying. This has been a pretty formative experience for me, and it’s seductive to be able to think of all of it as part of a plan. Literally seductive, in the sense of memetic evolution. Like that Hogwarts chart.

Read this book at your own risk; its theories will start creeping into everything you think.



Del Giudice On The Self-Starvation Cycle

[Content note: eating disorders]

Anorexia has a cultural component. I'm usually reluctant to assume anything is cultural – every mediocre social scientist's first instinct is always to come up with a cultural explanation which is simple, seductive, flattering to all our existing prejudices, and wrong. But after seeing enough ballerinas and cheerleaders who became anorexic after pressure to lose weight for the big competition, even I have to throw up my hands and admit anorexia has a cultural component.

But nobody ever tells you the sequel. That ballerina who's losing weight for the big competition at age 16? At age 26, she's long since quit ballet, worried it would exacerbate her anorexia. She's been in therapy for ten years; for eight of them she's admitted she has a problem, that her anorexia is destroying her life. Her romantic partners – the ones she was trying to get thin to impress – have long since left her because she looks skeletal and weird. She understands this and would do anything to cure her anorexia and be a normal weight again. But she finds she isn't hungry. She hasn't eaten in two days and she isn't hungry. In fact, the thought of food sickens her. She goes to increasingly expert therapists and dieticians, asking them to help her eat more. They recommend all the usual indulgences: ice cream, french fries, cookies. She tries all of them and finds them inexplicably disgusting. Sometimes with a prodigious effort of will she will manage to finish one cookie, and congratulate herself, but the next day she finds the task of eating dessert as daunting as ever. Finally, after many years of hard work, she is scraping the bottom end of normal weight by keeping to a diet so regimented it would make a Prussian general blush.

And nobody ever tells you about all the people who weren’t ballerinas. The young man who stops eating because it gives him a thrill of virtue and superiority to be able to demonstrate such willpower. The young woman who stops eating in order to show her family how much their neglect hurts her. If they pursue their lack of appetite far enough, they end up the same way as the ballerina – admitting they have a problem, admitting they need to eat more, hiring all sorts of doctors and dieticians to find them a way to eat more, but discovering themselves incapable of doing so.

And this is why I can’t subscribe to a *purely* cultural narrative of anorexia. How does “ballerinas are told they should be thin in order to be pretty” explain so many former ballerinas who want to gain weight but can’t? And how does it explain the weird, almost neurological stuff like how anorexic people will mis-estimate their ability to fit through doors?

All of this makes much more sense in a biological context; it’s as if the same system that is broken in obese people who cannot lose weight no matter how hard they try, is broken in anorexics who cannot gain weight no matter how hard they try. There are plenty of biological models for what this might mean. But then the question becomes: how do we reconcile the obviously cultural part where it disproportionately happens to ballerinas, to the probably biological part where the hypothalamus changes its weight set point?

I’m grateful to Professor del Giudice and *Evolutionary Psychopathology*for presenting the only reasonable discussion of this I have heard, which I quote here basically in its entirety:

> The self-starvation cycle arises in predisposed individuals following an initial phase of food restriction and weight loss. Food restriction may be initially prompted by a variety of motives, from weight concerns and a desire for thinness to health-related or religious ideas (eg spiritual purity, ascetic self-denial). In fact, the cycle may even be started by *involuntary* weight loss due to physical illness. While fasting and exercise are initially aversive, they gradually become rewarding – even addictive – as the starvation response kicks in. At the same time, restricting behaviors that used to be deliberate become increasingly automatic, habitual, and difficult to interrupt (Dwyer et al, 2001; Guarda et al, 2015; Lock & Kirz, 2013; McGuire & Troisi 1998). The self-starvation cycle plays a crucial role in the onset of anorexia.
>
> Increased physical activity is a key component of the starvation response in many animal species; in general, its function is to prompt exploration and extend the foraging range when food is scarce. This response is so ingrained that animals subjected to food restriction in conditions that allow physical activity often starve themselves to death through strenuous exercise (Fessler, 2002; Guarda et al, 2015; Scheurink et al, 2010). In humans, pride is a powerful additional rewrad of self-starvation – achieving extraordinary levels of thinness and self-control makes many anorexic patients feel special and superior (Allan & Goss, 2012). The starvation response also brings about some psychological changes that further contribute to reinforce the cycle. In particular, starvation dramatically interferes with executive flexibility/shifting, and patterns

> of behavior become increasingly rigid and inflexible. The balance between local and global processing is also shifted toward local details. This may contribute to common body image distortions in anorexia, as when patients focus obsessively on a specific body part (eg the neck or hips) but preceive themselves as globally overweight (Pender et al, 2014; Westwood et al; 2016).
>
> The self-starvation cycle has been documented across time and cultures, including non-Western ones. In modern Western societies, concerns with fat and thinness are the main reason for weight loss and probably explain the moderate rise of AN incidence around the second half of the 20th century. However, cases of self-starvation with spiritual and religious motivations have been common in Europe at least since the Middle Ages (and include several Catholic saints, most famously St. Catherine of Siena). In some Asian cultures, digestive discomfort is often cited as the initial reason for restricting food intake, but the resulting syndrome has essentially the same symptoms as anorexia in Western countries (Bell, 1984; Brumberg, 1989; Culbert et al, 2015; Keel & Klump, 2003). The DSM-5 criteria for anorexia include fear of gaining weight as a diagnostic requirement; for this reason, most historical and non-Western cases would not be diagnosed as AN within the current system. However, the present emphasis on thinness is likely a contingent sociohistorical fact and does not seem to represent a necessary feature of the disorder. (Keel & Klump, 2003)

My anorexic patients sometimes complain of being forced into this mold. They’ll try to go to therapy for their inability to eat a reasonable amount of food, and their therapist will want to spend the whole time talking about their body image issues. When they complain they don’t really have body image issues, they’ll get accused of repressing it. Eventually they’ll just say “Yeah, whatever, I secretly wanted to be a ballerina” in order to make the therapist shut up and get to the part where maybe treatment happens.

The clear weak part of this theory is the explanation of the “self-starvation cycle”. Aside from a point about animals sometimes having increased activity to go explore for food, it all seems kind of tenuous.

And how come most people who starve never get anorexia? How come sailors who ran out of food halfway across the Pacific, barely made it to some tropical island, and gorged themselves on coconuts didn’t end up anorexic? Donner Party members? Concentration camp survivors? Is there something special about voluntary starvation? Some kind of messed-up learning process?

I am interpreting the point to be something along the lines of “Suppose for some people with some unknown pre-existing vulnerability, starving themselves voluntarily now flips some biological switch which makes them starve themselves involuntarily later”.

Framed like this, it sounds more like a description of anorexia than a theory about it (though see here for an attempt to flesh this out). But it’s a description which captures part of the disease that a lot of other models don’t, and which brings some things into clearer relief, and I am grateful to have it.

Diametrical Model Of Autism And Schizophrenia

One interesting thing I took from *Evolutionary Psychopathology* was a better understanding of the diametrical theory of the social brain.

There’s been a lot of discussion over whether schizophrenia is somehow the “opposite” of autism. Many of the genes that increase risk of autism decrease risk of schizophrenia, and vice versa. Autists have a smaller-than-normal corpus callosum; schizophrenics have a larger-than-normal one. Schizophrenics smoke so often that some researchers believe they have some kind of nicotine deficiency; autists have unusually low smoking rates. Schizophrenics are more susceptible to the rubber hand illusion and have weaker self-other boundaries in general; autists seem less susceptible and have stronger self-other boundaries. Autists can be pathologically rational but tend to be uncreative; schizophrenics can be pathologically creative but tend to be irrational. The list goes on.

I’ve previously been skeptical of this kind of thinking because there are many things that autists and schizophrenics have in common, many autistics who seem a bit schizophrenic, many schizophrenics who seem a bit autistic, and many risk factors shared by both conditions. But Del Giudice, building on work by Badcock and Crespi presents the “diametrical model”: schizophrenia and autism are the failure modes of opposing sides of a spectrum from high functioning schizotypy to high functioning autism, ie from overly mentalistic cognition to overly mechanistic cognition.

Schizotypy is a combination of traits that psychologists have discovered often go together. It’s classified as a personality disorder in the DSM. But don’t get too caught up on that term – it’s a disorder in the same sense as narcissistic or antisocial tendencies, and like those conditions, some schizotypals do very well for themselves. Classic schizotypal traits include tendency toward superstition, disorganized communication, and nonconformity (if it sounds kind of like “schizophrenia lite”, that’s not really a coincidence).

Typically schizotypals are supposed to be paranoid and reclusive, the same as schizophrenics. But the diametrical model tries to downplay this in favor of noting that some schizotypals are unusually charismatic and socially successful. I am not exactly sure where they’re getting this from, but I cannot deny knowing several extremely charismatic people with a lot of schizotypal traits. Sometimes these people end up as “cult leaders” – not necessarily literally, but occupying that same niche of strange people who others are drawn toward for their unusually confident and otherworldly nature. Some of the people I know in this category have schizophrenic first-degree relatives, meaning they’re probably pretty loaded with schizotypal genes.

Schizotypals, according to the theory, have overly mentalistic cognition. Their brains are hard-wired for thinking in ways that help them understand minds and social interactions. When this succeeds, it looks like an almost magical understanding into what other people are secretly thinking, what their agendas are, and how to manipulate them. When it fails, it fails as animism and anthropomorphism: “I wonder what the universe is trying to tell me by making it rain today”. Or it fails as paranoia through oversensitivity to social cues: “I just saw him twitch his eye muscle slightly, which can sometimes mean he’s not interested in what I’m saying, and in the local status game that could mean that he doesn’t think I’m important enough, and that implies he might think he’s better than me and I’m expendable…”

Autism, then, would be the opposite of this. It’s overly mechanistic cognition, thinking in terms of straightforward logic and the rules of the physical world. Autistic people don’t make the mistake of thinking the universe is secretly trying to tell them something. On the other hand, after several times trying to invite a slightly autistic woman I had a crush on to things, telling her how much I liked her, petting her hair, etc, she still hadn’t figured out I was trying to date her until I said explicitly “I AM TRYING TO DATE YOU”. So not believing that you are secretly being told things has both upsides and downsides.

Autistic people are sometimes accused of looking for a set of rules that will help them understand people, or the secret cheat code that will make people give them what they want. I imagine an autistic person asking something like “What is the alternative?” This is the kind of thought process that usually works on stuff: figure out the rules that govern something, find a way to exploit them, and boom, you’ve landed a rocket on the moon. How are *they* supposed to know that human interaction is a bizarre set of layered partial-information games that you’re supposed to solve by looking at someone’s eye muscle twitches and concluding they’re going to steamroll over you to get a promotion at work?

Is this true? There’s…not great evidence for it. I’ve never seen any studies. There’s certainly a stereotype that brilliant engineers are not necessarily the most socially graceful people. But I know a lot of people who combine excellent technical skills with excellent social skills, and other people who are failures in both areas. So probably the best that can be said about this theory is that it would be a really neat way to explain the patterns of similarities and differences between schizophrenia and autism.

In this theory, both high-functioning autism (being good at mechanistic cognition) and high-functioning schizotypy (being good at mentalistic cognition) may be good things to have. But the higher your mutational load is – the less healthy your brain, and the fewer resources it has to bring to the problem – the less well it is able to control these powerful abilities. A schizotypal brain that cannot keep its mentalistic cognition yoked to reality dissolves into schizophrenia, completely losing the boundary between Self and Other into a giant animistic universe of universal significance and undifferentiated Mind. An autistic brain that cannot handle the weight of its mechanistic cognition becomes unable to do even the most basic mental tasks like identify and cope with its own emotions. And because in practice we're talking about shifts in the complicated computational parameters that determine our thoughts and personalities, rather than the thoughts and personalities directly, both of these conditions have a host of related sensory and cognitive symptoms that aren't quite directly related.

So here the reason why autism and schizophrenia seem both opposite and similar to each other is because they're opposite (in the sense of being at two ends of a spectrum), and similar (in the sense that the same failure mode of high mutational load and low "mental resources" will cause both).

If you're thinking "it sounds like someone should do a principal components analysis on this", then Science has your back (paper, popular article). They find that:

> Consistent with previous research, autistic features were positively associated with several schizotypal features, with the most overlap occurring between interpersonal schizotypy and autistic social and communication phenotypes. The first component of a principal components analysis (PCA) of subscale scores reflected these positive correlations, and suggested the presence of an axis (PC1) representing general social interest and aptitude. By contrast, the second principal component (PC2) exhibited a pattern of positive and negative loadings indicative of an axis from autism to positive schizotypy, such that positive schizotypal features loaded in the opposite direction to core autistic features.

In keeping with this theory, studies find that first-degree relatives of autists have higher mechanistic cognition, and first-degree relatives of schizophrenics have higher mentalistic cognition and schizotypy. Autists' relatives tend to have higher spatial compared to verbal intelligence, versus schizophrenics' relatives who tend to have higher verbal compared to spatial intelligence. High-functioning schizotypals and high-functioning autists have normal (or high) IQs, no unusual number of fetal or early childhood traumas, and the usual amount of bodily symmetry; low-functioning autists and schizophrenics have low IQs, increased history of fetal and early childhood trauams, and increased bodily asymmetry indicative of mutational load.

If men have much more autism than women, shouldn't women have much more schizophrenia than men. You'd think so, but actually men have more. But men have greater variability in general, which means they're probably more likely to satisfy the high mutational load criterion. So maybe we should instead predict that women should have higher levels of high-functioning schizotypy. Studies show women do have more "positive schizotypy", the sort being discussed here, but lower “negative schizotypy”, a sort linked to the negative symptoms of schizophrenia.)

Something that bothered me while I was writing this: famous mathematician John Nash was schizophrenic. Isn’t that kind of weird if schizophrenia is about an imbalance in favor of verbal/personal and against logical/mathematical thinking?

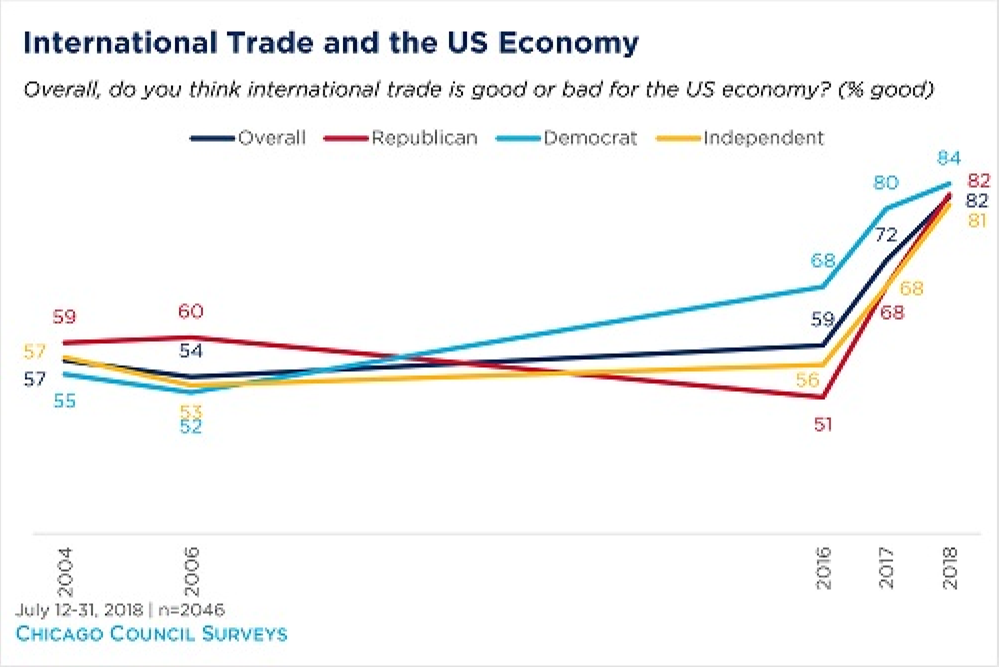
There are exceptions to everything, and we probably shouldn’t make too much of one case. But I find it striking that Nash’s work was in game theory: essentially a formalization of social thinking, and centered around the sort of paranoid social thinking of figuring out what to do about how other people might be out to get you. This is probably just a coincidence, but it’s pretty funny.

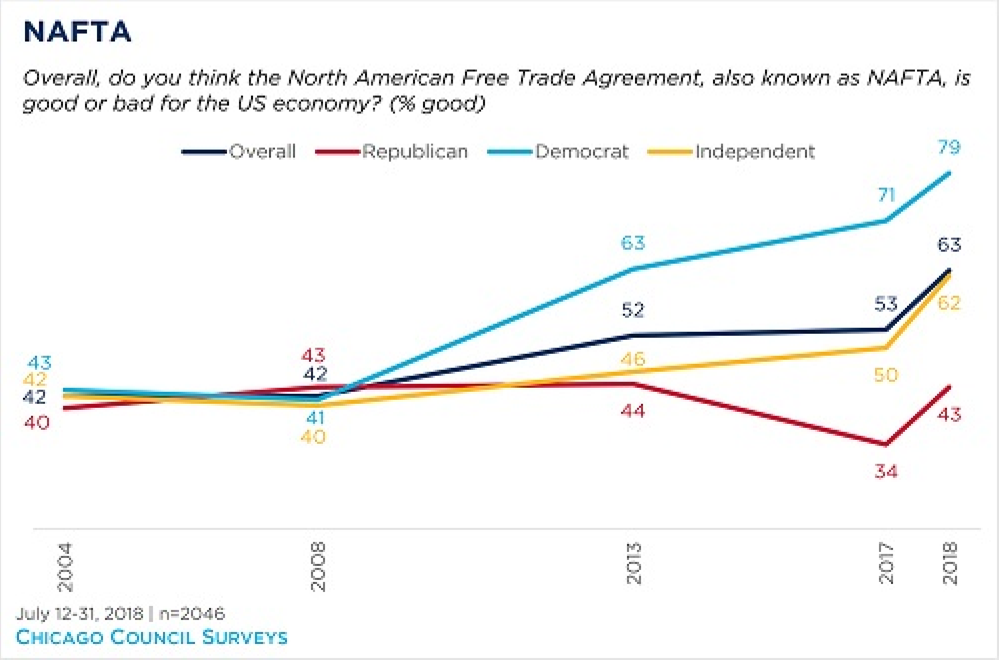
Trump: A Setback For Trumpism

Donald Trump has been called a setback for many things. America. The global community. The environment. Civil service. Civil society. Civility. Civilization. The list goes on.

One might think he has at least been useful to his own cause. That he could at least claim to have benefited the ideas of populism, nationalism, immigration control, and protectionism. That if anything could avoid being devastated by Trump, it would be Trumpism.

But here are some polls from the past few years. They're all on slightly different things, but I think together they tell an interesting story:

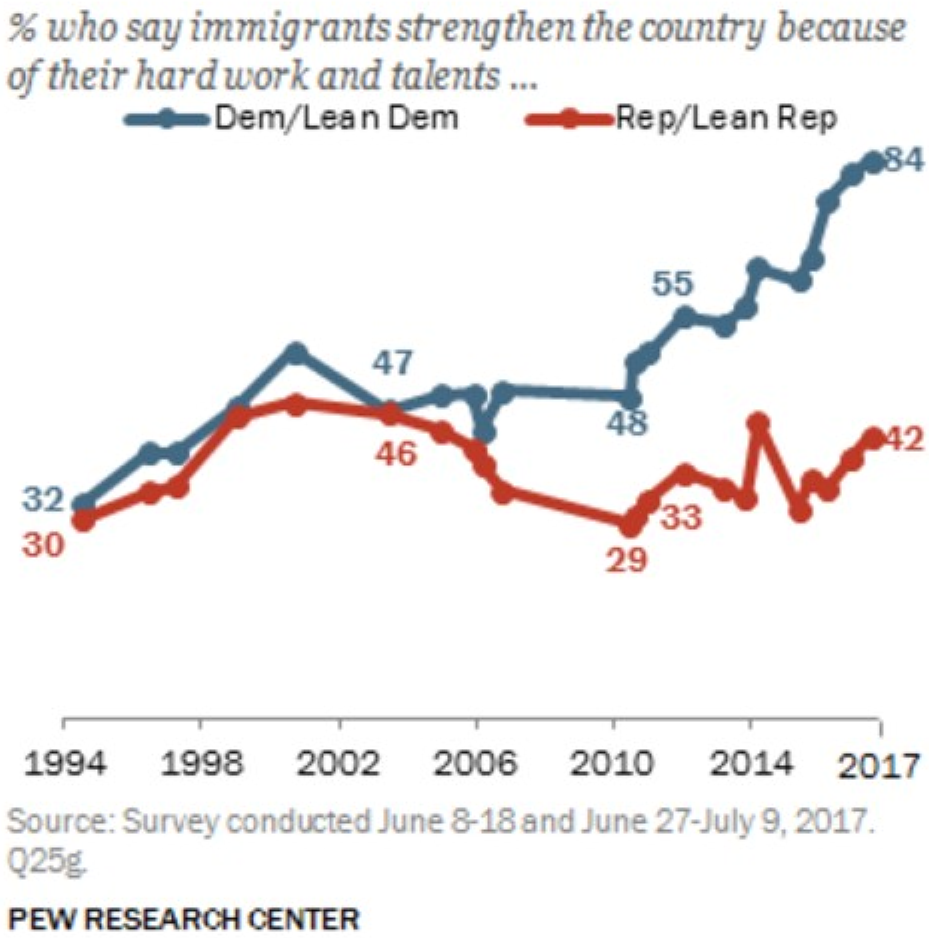



NAFTA

Overall, do you think the North American Free Trade Agreement, also known as NAFTA, is good or bad for the US economy? (% good)

July 12-31, 2018 | n=2046
CHICAGO COUNCIL SURVEYS

Support for global free trade mysteriously spiked around 2016.

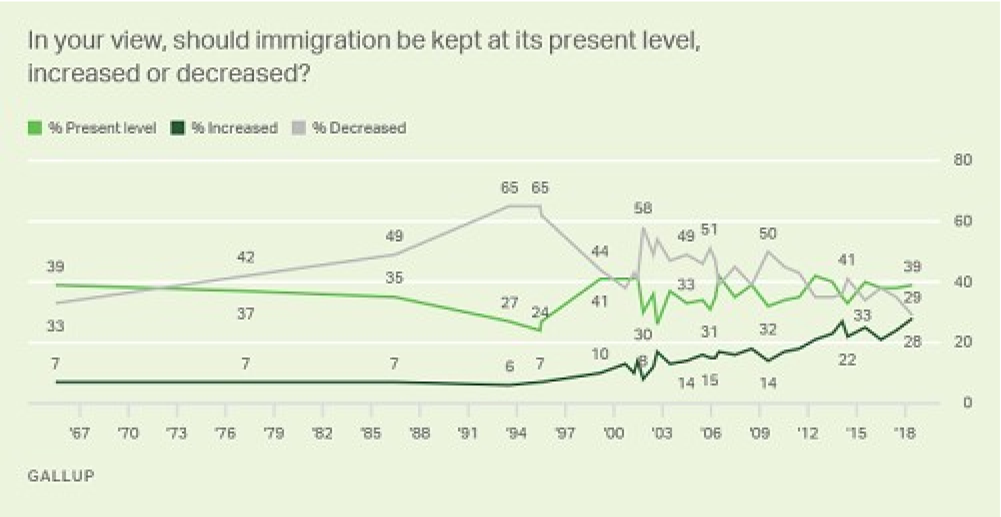


% who say immigrants strengthen the country because of their hard work and talents ...

Source: Survey conducted June 8-18 and June 27-July 9, 2017. Q25g.

PEW RESEARCH CENTER

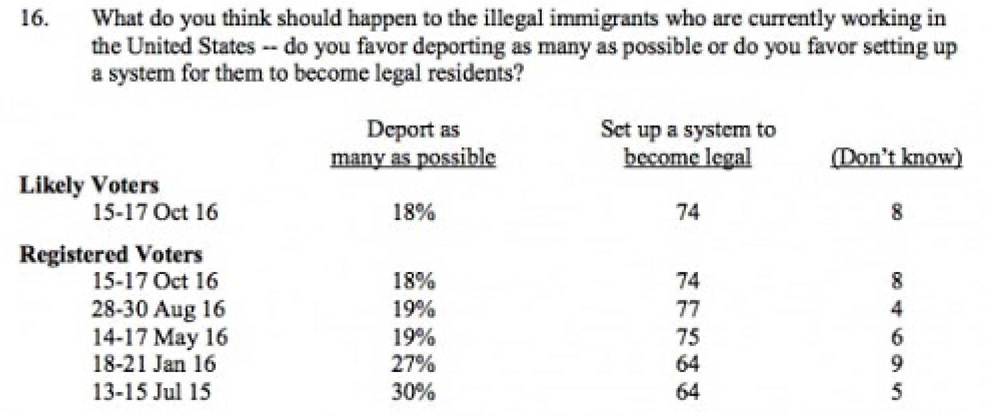
So did moral support for immigrants.



In your view, should immigration be kept at its present level, increased or decreased?

GALLUP

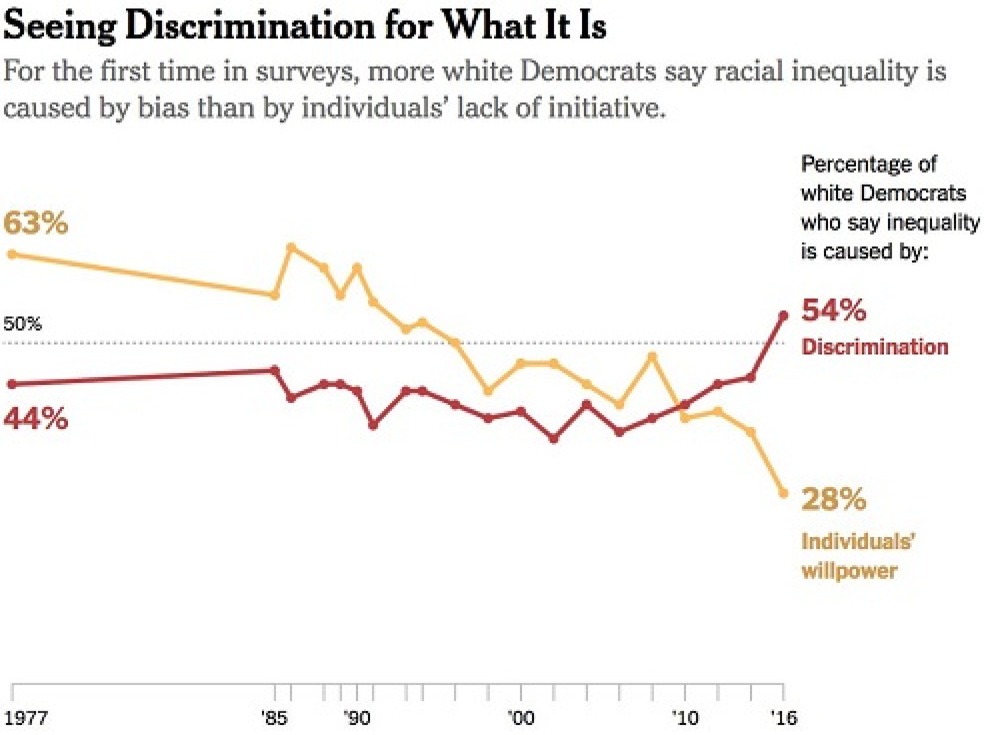
…and, less clearly but still there, support for increasing the number of immigrants (though see here for an apparently contrary source).



16. What do you think should happen to the illegal immigrants who are currently working in the United States -- do you favor deporting as many as possible or do you favor setting up a system for them to become legal residents?

	Deport as many as possible	Set up a system to become legal	(Don't know)
Likely Voters			
15-17 Oct 16	18%	74	8
Registered Voters			
15-17 Oct 16	18%	74	8
28-30 Aug 16	19%	77	4
14-17 May 16	19%	75	6
18-21 Jan 16	27%	64	9
13-15 Jul 15	30%	64	5

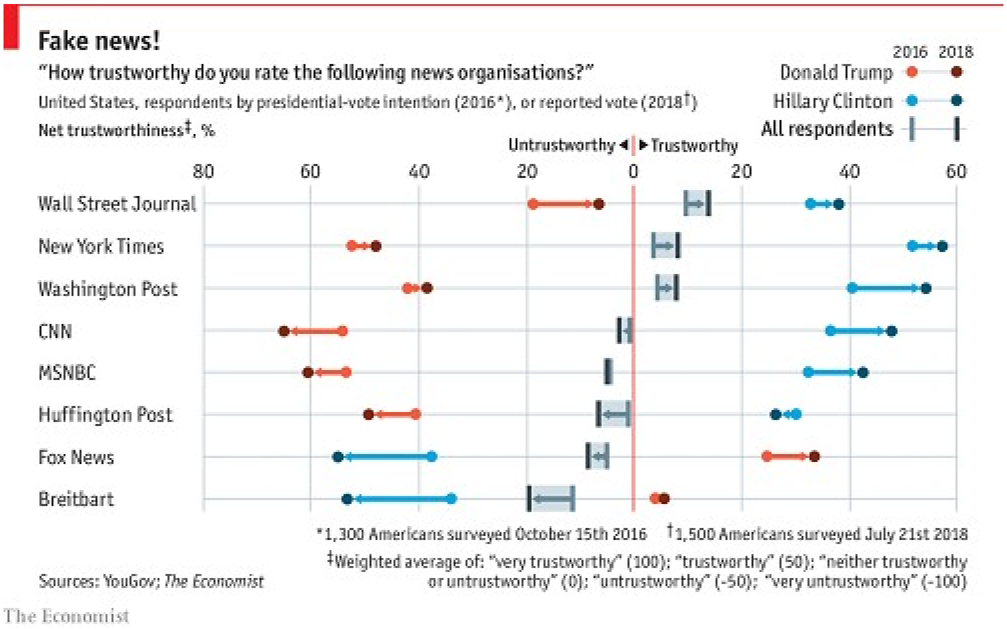
…and opposition to deporting illegal immigrants.



Seeing Discrimination for What It Is

For the first time in surveys, more white Democrats say racial inequality is caused by bias than by individuals' lack of initiative.

So did belief in racial discrimination as a major cause of inequality, according to this chart with a completely unbiased title which is willing to let readers decide how to think about this issue for themselves.



And so did trust in the *New York Times* and other mainstream media sources.

The clearest example I can find of this effect doesn't come from the US at all. It's Minkus, Deutschmann & Delhey (2018). They find that a large European poll asked the same question about support for the EU the week before and after Trump's election. Just after the election, there was a giant spike in support for the EU, "considerable in size, roughly equivalent to three years of education". They conclude that:

> The election of Trump as a right-wing nationalist with a declared aversion to supranational institutions including the EU — did not trigger a domino effect in the same direction in Europe. To the contrary, a rally effect occurred, in which Europe moved closer together, rallying around the EU's "flag." This indicates that an event that may at first sight appear to be a global victory for nationalism can immediately trigger measurable sentiments of resistance in another part of the world, actually leading to new impetus for supranationalism.

This kind of analysis is inherently vulnerable to cherry-picking, and I admit I've chosen some especially dramatic results. And polls naturally have a lot of variability, and none of these on their own constitute proof of anything. But I think when you put everything together you do get a trend. Some things have stayed the same, or are inconclusive. But there do seem to be a lot of cases where support for Trumpist positions show a sudden and lasting decrease as soon as Trump enters the national stage.

I want credit for predicting this. In my endorsement of anyone except Trump, I told progressives not to vote Trump because they opposed his policy, and conservatives not to vote Trump because he would cause a backlash that was worse than anything they might get from him. I said that the left thrives by imagining themselves as brave rebels fighting an ignorant, regressive, hateful authority, and that "bringing their straw man to life and putting him in the Oval Office" would be "the biggest gift" they could give the Democrats, and would end up pushing an entire generation further to the left.

I think this is a good broad theory of what's happening, but it might be worth digging deeper to try to distinguish possible mechanisms.

First, maybe Trump is just such an offensive and aversive figure that people switch sides in disgust. This is a little weird; if you were anti-immigration before Trump, can't you just say "I hate Trump, but I'm still against immigration"? But maybe people's minds don't work that way.

Second, maybe Trump made causes like protectionism and nativism so central to the Republican narrative that they became untenable for Democrats. That is, in 2010, it might have been possible to be an anti-illegal-immigration

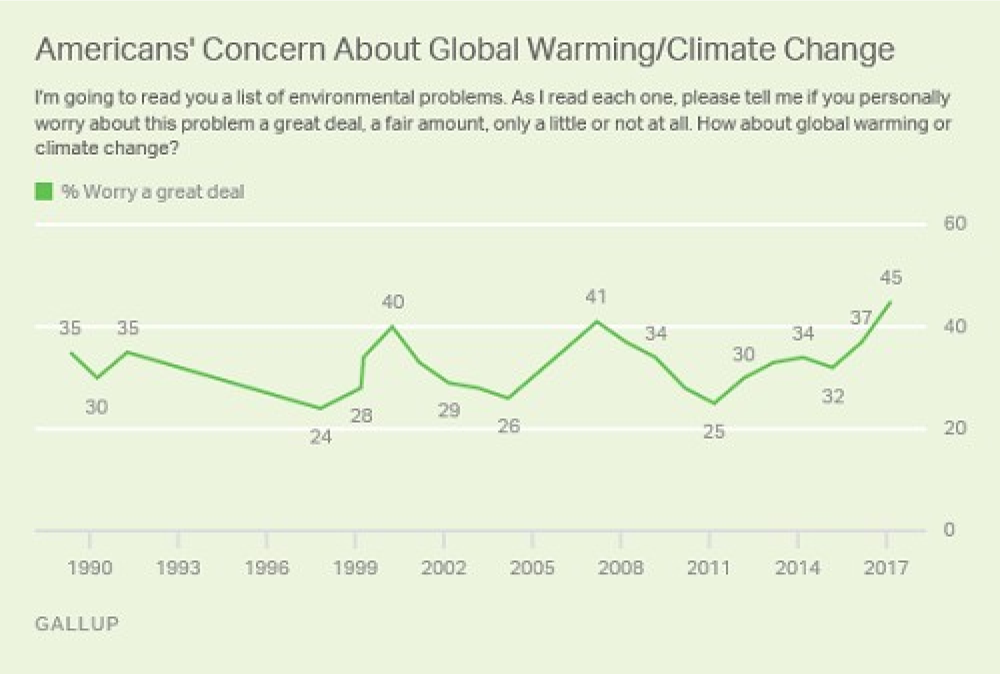
Democrat (remember, in the early 2000s Hillary supported a border fence), but in 2018, that would signal being a Republican, or at least someone of questionable loyalty to the Democratic Party. In order to fit in, moderate Democrats abandoned their anti-illegal-immigration stances. The graphs above seem to provide some evidence for this: they usually show the largest shift among Democrats, with Republicans merely staying where they are.

Third, and kind of opposite that, maybe Trump is such an offensive and aversive figure that conservatives feel a need to maintain their reputation by distancing themselves from him. Maybe in 2010, being anti-illegal-immigration signaled things that you wanted to signal, like patriotism and support for low-paid workers. And now, being anti-immigration signals things you don’t want to signal, like Trump’s particular brand of inflammatory divisiveness. This doesn’t fit the evidence from the graphs above, but it does sort of fit the European study, where further-right Europeans were more likely to switch opinions after the election than further-left ones.

Fourth, maybe Trump’s focus on certain causes shifted the focus of Democrats and the mainstream media to those causes, and Democrats and the mainstream media were better at opposing them than Trump was at supporting them. For example, since Trump the media has been focusing more intensely on negative aspects of ICE and Border Control practices which were less well-covered before his presidency. If this focus has successfully changed minds, that would explain a shift away from Trump.

Fifth, maybe Trump has shifted the goalposts. Maybe identifying as anti-illegal-immigrant before Trump just meant you thought there should be a little better border control, but now you think it means you want a wall and mass deportations, plus you think all Mexicans are rapists. If you felt like the anti-illegal-immigrant cause was getting more extreme, but your positions stayed the same, then you might stop identifying as anti-illegal-immigrant.

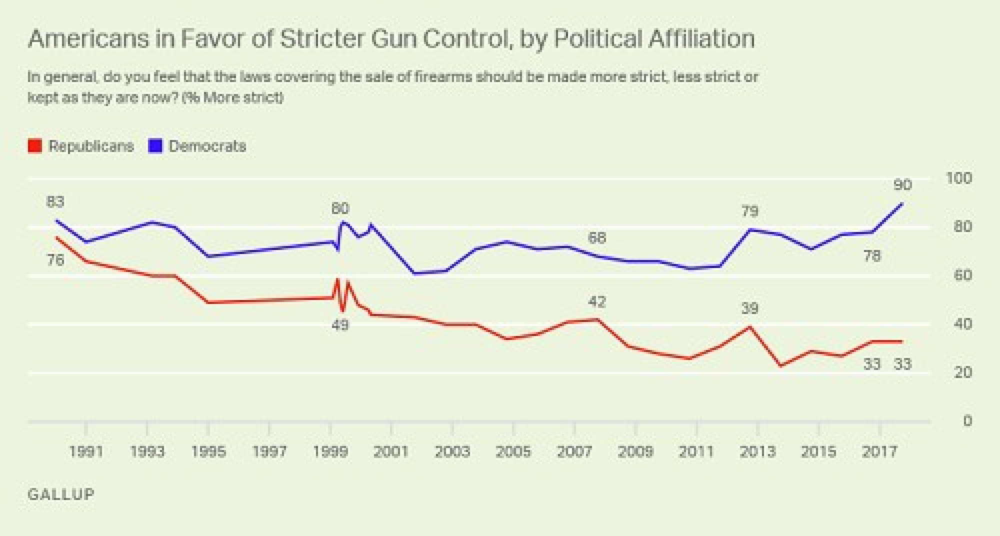
Sixth, there have been a lot of studies showing that peaceful protests may increase support for a cause, but violent or disruptive protests usually decrease it (1, 2, 3). It's easy enough to analogize Trump to a "disruptive protest" – in the sense of an ideological cause getting associated with an unsympathetic proponent – and this would be compatible with any of the explanations above. But I notice that most of the research in this area was done on whites reacting to civil rights protests, adding an identity dimension: maybe disruptive racially charged protests by blacks increase the salience of race as a category for whites, causing them to shift their opinions more towards ones based on their race rather than based on other values. This would also explain the paradoxical Ferguson effect mentioned in Part III here. In the same way, we can think of Trump's election as a disruptive Republican move that makes Democrats feel threatened and increases the salience of partisanship for them. This would cause a sort of unilateral polarization, where Democrats become more progressive but Republicans don't necessarily become more conservative, and so the country as a whole shifts to the left. Like the second explanation, this is compatible with the party breakdown on the graphs above. It's also compatible with this:



Americans' Concern About Global Warming/Climate Change

I'm going to read you a list of environmental problems. As I read each one, please tell me if you personally worry about this problem a great deal, a fair amount, only a little or not at all. How about global warming or climate change?

GALLUP



Americans in Favor of Stricter Gun Control, by Political Affiliation

In general, do you feel that the laws covering the sale of firearms should be made more strict, less strict or kept as they are now? (% More strict)

GALLUP

These show the familiar 2016 spike. But although Trump has taken positions against fighting climate change or regulating guns, I don't think of these two issues as "Trumpist" in the same way as illegal immigration, and I'm surprised they seem to show a Trump-related change. This would make more sense if Trump caused a wider-reaching closing of ranks among Democrats rather than just a shift away from his personal hobbyhorses.

I think all of this should increase people’s concern about backlash effects. Contrary to what some of my conflict theorist friends seem to think, civility and honesty are not always pointless own-goals in politics. If you’re sufficiently repulsive and offensive, you can also end up damaging your own cause.

As I’ve pointed out before, backlash can sometimes be a necessary trade-off to energize your base. But as I’ve also pointed out before, people tend to overestimate the importance of turning out the base, and to underestimate the importance of not having everyone hate you. So if I were a Trumpist, I would be very worried right now.

Fallacies Of Reversed Moderation

A recent discussion: somebody asked why people in Silicon Valley thought that only high-tech solutions to climate change (like carbon capture or geoengineering) mattered, and why they dismissed more typical solutions like international cooperation and political activism.

Another person cited statements from the relevant Silicon Valley people, who mostly say that they think political solutions and environmental activism were central to the fight against climate change, but that we should look into high-tech solutions too.

This is a pattern I see again and again.

Popular consensus believes 100% X, and absolutely 0% Y.

A few iconoclasts say that X is definitely right and important, but maybe we should also think about Y sometimes.

The popular consensus reacts “How can you think that it’s 100% Y, and that X is completely irrelevant? That’s so extremist!”

Some common forms of this:

Reversed moderation of planning, like in the geoengineering example. One group wants to solve the problem 100% through political solutions, another group wants 90% political and 10% technological, and the first group thinks the second only cares about technological solutions.

Reversed moderation of importance. For example, a lot of psychologists talk as if all human behavior is learned. Then when geneticists point to experiments showing behavior is about 50% genetic, they get accused of saying that “only genes matter” and lectured on how the world is more complex and subtle than that.

Reversed moderation of interest. For example, if a vegetarian shows any concern about animal rights, they might get told they’re “obsessed with animals” or they “care about animals more than humans”.

Reversed moderation of certainty. See for example my previous article Two Kinds Of Caution. Some researcher points out a possibility that superintelligent AI might be dangerous, and suggests looking into this possibility. Then people say it doesn’t matter, and we don’t have to worry about it, and criticize the researcher for believing he can “predict the future” or thinking “we can see decades ahead”. But “here is a possibility we need to investigate” is a much less certain claim than “no, that possibility definitely will not happen”.

I can see why this pattern is tempting. If somebody said the US should allocate 50% of its defense budget to the usual global threats, and 50% to the threat of reptilian space invaders, then even though the plan contains the number “50-50” it would not be a “moderate” proposal. You would think of it as “that crazy plan about fighting space reptiles”, and you would be right to do so. But in this case the proper counterargument is to say “there is no reason to spend any money fighting space reptiles”, not “it’s so *immoderate* to spend *literally 100%* of our budget breeding space mongooses”. “Moderate” is not the same as “50-50” is not the same as “good”. Just say “Even though this program leaves some money for normal defense purposes, it’s stupid”. You don’t have to deny that it leaves anything at all.

Or if someone says there’s a 10% chance space reptiles will invade, just say “No, the number is basically zero”. Don’t say “I can’t believe you’re *certain* there will be an alien invasion, don’t you know there’s never any certainty in this world?”

But I can see why this happens. Imagine the US currently devotes 100% of its defense budget to countering Russia. Some analyst determines that although Russia deserves 90% of resources, the Pentagon should also use 10% to counter China. Since no one person can shift very much of the defense budget, this analyst might spend all her time arguing we need to counter China more, trying to convince everyone that China is really very dangerous; if she succeeds, maybe the budget will shift to 99-to-1 and she’ll have done the best she can. But if she really spends all her time talking about China, this might look to other people like she’s an extremist – that crazy single-issue China person – “Why are you spending all your time talking about China? Don’t you realize Russia is important too?” Still, she’s taking the right strategy, and it’s hard to figure out what she could do better.

I am nervous titling this “reversed moderation fallacy” because any time someone brings up fallacies, people accuse them of thinking all discussion consists of identifying and jumping on Officially Designated Fallacies in someone else’s work. But I’ve gone years without talking about fallacies at all, so when this inevitably happens it’s going here as Exhibit A.

Refactoring: Culture As Branch Of Government

Ribbonfarm likes to talk about refactoring, a conceptual change in how you see the world. I'm not totally sure I understand it, but I think it means things like memetics – where you go from the usual model of people deciding what ideas they want, to a weird and inside-out (but not objectively wrong) model of ideas competing to colonize people.

Here is a refactoring I think about a lot: imagine a world where people considered culture the fourth branch of government. Imagine that civics textbook writers taught high school students that the US government had four branches: executive, legislative, judicial, and cultural.

I think about this because I have a bias to ignore anything that isn't nailed down and explicit. Culture isn't nailed down. But if it were in the Constitution in nice calligraphy right beside the Presidency and the Supreme Court, why, *then* it would be as explicit as it gets.

Like many other people, I was hopeful that nation-building Iraq (or Afghanistan, or…) would quickly turn it into a liberal democracy (in my defense, I was eighteen at the time). Like many other people, I was disappointed and confused when it didn't. The people in the world that considers culture the fourth branch of government weren't confused. Bush forgot to nation-build an entire branch of government. If he'd given Iraq a western-style Supreme Court, marble facade and all, but left their executive and legislature exactly how they were before, that would be a recipe for conflict, confusion, and eventually nothing getting done. So why should westernizing their executive, legislature, and courts – but not their culture – work any better?

The world that considers culture the fourth branch of government doesn't get all confused calling hunter-gatherers or peasant villagers "primitive communism" or "anarchism" or "ruled by elders" or things like that. Those people's governments have a cultural branch but not much else. Why should we be surprised? Medieval Iceland had only legislative and judicial branches; medieval Somalia only had a judiciary; some dictatorships run off just an executive.

Each branch of government enforces rules in its own way. The legislature passes laws. The executive makes executive orders. The judiciary rules on cases. And the culture sets norms. In our hypothetical world, true libertarians are people who want less of all of these. There are people who want less of the first three branches but want to keep strong cultural norms about what is or isn't acceptable – think Lew Rockwell and other paleoconservatives who hope that the retreat of central government will create strong church-based communities of virtuous citizens. These people aren't considered libertarians. They might be considered principled constitutionalists, the same way as people who worry about the "imperial presidency" and its use of executive orders. But in the end, what they want to strengthen some branches of government at the expense of others. The real libertarians also believe that cultural norms enforced by shame and ostracism are impositions on freedom, and fight to make these as circumscribed as possible.

Debates about primaries and the Electoral College are understood to be about how to determine who controls the executive branch. Debates about gerrymandering and voter suppression are understood to be about how to determine who controls the legislative branch. And debates about censorship, the media, and – yes – immigration – are understood to be about how to determine who controls the cultural branch.

In our hypothetical world, the First Amendment specifies “This shall apply to the first three branches of government, but not the fourth”. Still, debate goes on. Some people are happy the first three branches are kept out of it, but glad that the culture censors the speech of people who shouldn’t be talking. These people are another set of principled constitutionalists, the same as people who want to make sure only Congress (and not the President) can declare war, but don’t mind war in and of itself. Other people really are free speech purists, and think that no branch of government should be able to restrict the marketplace of ideas.

I admit this is kind of a silly hypothetical. Culture is much more different from the three branches of government than they are from each other; a civics textbook writer would have to be pretty strange to think it was worth tacking it on. Still, it’s a useful (if extreme) counter to forgetting to ever think about culture at all.

Seattle Meetup This Sunday

Why: Because I’m in Seattle this weekend

Where: 5238 11th Ave NE, Seattle WA (private residence). If there are too many people we’ll figure something out.

When: 4 PM on Sunday 12/23

Who: Anyone who wants. Please feel free to come even if you feel awkward about it, even if you’re not “the typical SSC reader”, even if you’re worried people won’t like you, etc. People who fit those descriptions who decided to come to previous meetups have mostly enjoyed them.

Please Take The 2019 SSC Survey

Please take **the 2019 Slate Star Codex Survey**.

The survey helps me learn more about SSC readers and plan community events. But it also provides me with useful informal research data for questions I'm interested it, which I then turn into interesting posts. My favorite from last year was Fight Me, Psychologists: Birth Order Effects Exist And Are Very Strong, which I think made a real contribution to individual differences psychology and which could not have happened without your cooperation.

The survey is open to anyone who has ever read a post on this blog before December 27 2018. Please don't avoid taking the survey just because you feel like you're not enough of a "regular". It will ask you how much of a "regular" you are, so there's no risk you'll "dilute" the results. The survey will stay open until mid-January, and I will probably be begging and harassing you to take it about once a week or so until then.

This year's survey is in two parts. Part I asks the same basic questions as previous years and should take about **ten minutes**. Part II asks more questions on research topics I'm interested in and should take about **fifteen minutes**. It would be great if you could take both parts, but if 25 minutes sounds like too much surveying to you, you can also just take Part I.

As always, the survey is plagued by fundamental limitations, poor technology, and my own carelessness, but a couple of things to watch for:

– Once you click a box on a Google form, you cannot unclick it – i.e. you can change your answer but you can't unanswer the question. If you click a box you didn't mean to, please switch your answer to "Other" if available; if not, then choose the most boring inoffensive answer that is least likely to produce surprising results. I realize how bad this is but there is apparently no way around it.

– Some of the questions are America-centric, because I either have to learn everything about every culture or be something-centric, and America seemed like a good place to center around. Sorry to non-American readers. Feel free to skip any questions that don't apply to you.

– By default, all responses will be included in a public dataset for anyone who wants to analyze them. Your responses will obviously not be attached to your name or any similarly blatant identifying information, but if you are the only 92 year old from Uruguay on SSC then someone could theoretically identify you. There is an option to make all your responses private; if identifiability bothers you, feel free to check it and you will not be included in the public dataset. There are a few sections of Part II, especially the ones on drugs and sex, which I figure need extra privacy protection; I will not be sharing these regardless of your answer to the privacy question.

– Due to poring over a 5000 entry spreadsheet not actually being that much fun, I am not up for changing your answers after you submit them. Please do not email me asking me to do this. This includes your answer to the privacy question. Please figure out whether you want privacy before taking the survey.

That having been said, you are all great, and I super-appreciate any survey-filling-out you are willing to do. If you can donate about a half-hour, I hope I can pay you back in interesting findings and useful crowd-sourced life advice. So:

Take the 2019 Slate Star Codex Survey here

2019

- What Happened To 90s Environmentalism?
- Bay Area SSC Meetup 1/6
- Preregistration Of Investigations For The 2019 SSC Survey
- Bay Meetup 1/6 Update
- Book Review: The Structure Of Scientific Revolutions
- Paradigms All The Way Down
- SSC Survey Results 2019
- Too Many People Dare Call It Conspiracy
- Kernel Of Doubt: Testing Math Preference Vs. Corn-Eating Style
- Highlights From The Comments On Kuhn
- 2018 Predictions: Calibration Results
- Psychiat-List Now Up
- Predictions For 2019
- Book Review: Zero To One
- Respectability Cascades
- Survey Results On SSRIs
- Irvine Meetup This Saturday
- The Proverbial Murder Mystery
- Do Neural Nets Dream Of Electric Hobbits?
- GPT-2 As Step Toward General Intelligence
- My Plagiarism
- RIP Culture War Thread
- Wage Stagnation: Much More Than You Wanted To Know

- Rule Thinkers In, Not Out
- In Mod We Trust
- Meaningful
- Prospiracy Theories
- Ketamine: Now By Prescription
- Puritan Spotting
- Does Reality Drive Straight Lines On Graphs, Or Do Straight Lines On Graphs Drive Reality?
- Gwern’s AI-Generated Poetry
- Book Review: Inventing The Future
- Translating Predictive Coding Into Perceptual Control
- Neurons And Intelligence: A Birdbrained Perspective
- [PARTIALLY RETRACTED] Cortical Neuron Number Matches Intuitive Perceptions Of Moral Value Across Animals
- Partial Retraction Of Post On Animal Value And Neural Number
- Two Wolves And A Sheep
- Social Censorship: The First Offender Model
- Classified Thread 7
- Short Book Reviews April 2019
- Pain As Active Ingredient In Dating
- Increasingly Competitive College Admissions: Much More Than You Wanted To Know
- Highlights From The Comments On College Admissions
- 1960: The Year The Singularity Was Cancelled

- Buspirone Shortage In Healthcaristan SSR
- Update To Partial Retraction Of Animal Value And Neuron Number
- Little Known Types Of Eclipse
- 5-HTTLPR: A Pointed Review
- Is There A Case For Skepticism Of Psychedelic Therapy?
- Age Gaps And Birth Order Effects
- A Critical Period For Lactation Fetishes
- APA Meetup This Saturday
- The APA Meeting: A Photo-Essay
- Postscript To APA Photo-Essay
- Bay Area SSC Meetup 6/2
- [REPOST] Epistemic Learned Helplessness
- Book Review: The Secret Of Our Success
- List Of Passages I Highlighted In My Copy Of "The Secret Of Our Success"
- Asymmetric Weapons Gone Bad
- Addendum To "Enormous Nutshell": Competing Selectors
- Book Review: Why Are The Prices So D*mn High?
- Highlights From The Comments On Cultural Evolution
- Followup On The Baumol Effect: Thanks, O Baumol
- If Kim Jong-Un Opened A KFC, Would You Eat There?
- Are Sexual Purity Taboos A Response To STIs?
- If Only Turing Was Alive To See This
- More Confounders

- Editing Unsong
- Some Clarifications On Rationalist Blogging
- Style Guide: Not Sounding Like An Evil Robot
- Gay Rites Are Civil Rites
- Survey Results: Sexual Roles
- Do People Like Their Mental Health Care?
- Against Lie Inflation
- Caution On Bias Arguments
- Know Your Gabapentinoids
- Book Review: The Electric Kool-Aid Acid Test
- Adversarial Collaboration Contest 2019
- Against Against Billionaire Philanthropy
- Adversarial Collaboration Update
- Squareallworthy On UBI Plans
- Highlights From The Comments On Billionaire Philanthropy
- Book Review: Secular Cycles
- Attempted Replication: Does Beef Jerky Cause Manic Episodes?
- Maybe Your Zoloft Stopped Working Because A Liver Fluke Tried To Turn Your Nth-Great-Grandmother Into A Zombie
- Don’t Fear The Simulators
- Book Review: Reframing Superintelligence
- Meetups Everywhere 2019
- Book Review: Ages Of Discord
- List Of Passages I Highlighted In My Copy Of Ages Of Discord
- LW Party: Bay Area Invitation
- [Partial Retraction] Age Gaps and Birth Order Effects

- SSC Journal Club: Relaxed Beliefs Under Psychedelics And The Anarchic Brain
- Lots Of People Going Around With Mild Hallucinations All The Time
- SSC Meetups 2019: Times And Places
- Against Against Pseudoaddiction
- Too Much Dark Money In Almonds
- Boston Meetup This Friday
- NYC Meetup This Saturday
- Philadelphia Meetup This Sunday
- Washington DC Meetup This Tuesday
- Ann Arbor Meetup This Thursday
- Chicago Meetup This Saturday
- Austin Meetup This Monday
- Portland Meetup This Tuesday
- Seattle Meetup This Wednesday
- Fairbanks Meetup This Sunday
- Berkeley Meetup This Thursday
- Irvine Meetup This Friday
- Book Review: Against The Grain
- Is Enlightenment Compatible With Sex Scandals?
- The PNSE Paper
- Indian Economic Reform: Much More Than You Wanted To Know
- Highlights From The Comments On PNSE
- Financial Incentives Are Weaker Than Social Incentives But Very Important Anyway
- New Atheism: The Godlessness That Failed
- Samsara
- Building Intuitions On Non-Empirical Arguments In Science

- Book Review: The Body Keeps The Score
- Autism And Intelligence: Much More Than You Wanted To Know
- More Intuition-Building On Non-Empirical Science: Three Stories
- Book Review: All Therapy Books
- Mental Mountains
- SSC Meetups Everywhere Retrospective
- Symptom, Condition, Cause
- 2019 Adversarial Collaboration Entries
- [ACC] What Are The Benefits, Harms, And Ethics Of Infant Circumcision?
- Correction To Circumcision ACC
- [ACC] Is Eating Meat A Net Harm?
- [ACC] Does Calorie Restriction Slow Aging?
- [ACC] Should We Colonize Space To Mitigate X-Risk?
- [ACC] Should Gene Editing Technologies Be Used In Humans?
- [ACC] When During Fetal Development Does Abortion Become Morally Wrong?
- A Maximally Lazy Guide To Giving To Charity In 2019
- [ACC] Will Automation Lead To Economic Crisis?
- [ACC] Should You Have A Merry Christmas?
- [ACC] How Much Significance Should We Ascribe To Spiritual Experiences?
- Please Vote For ACC Winner
- Please Take The 2020 SSC Survey!

What Happened To 90s Environmentalism?

0. Introduction

I grew up in the 90s, which meant watching movies about plucky children fighting Pollution Demons. Sometimes teachers would show them to us in class. None of us found that strange. We knew that when we grew up, this would be our fight: to take on the loggers and whalers and seal-clubbers who were destroying our planet and save the Earth for the next generation.

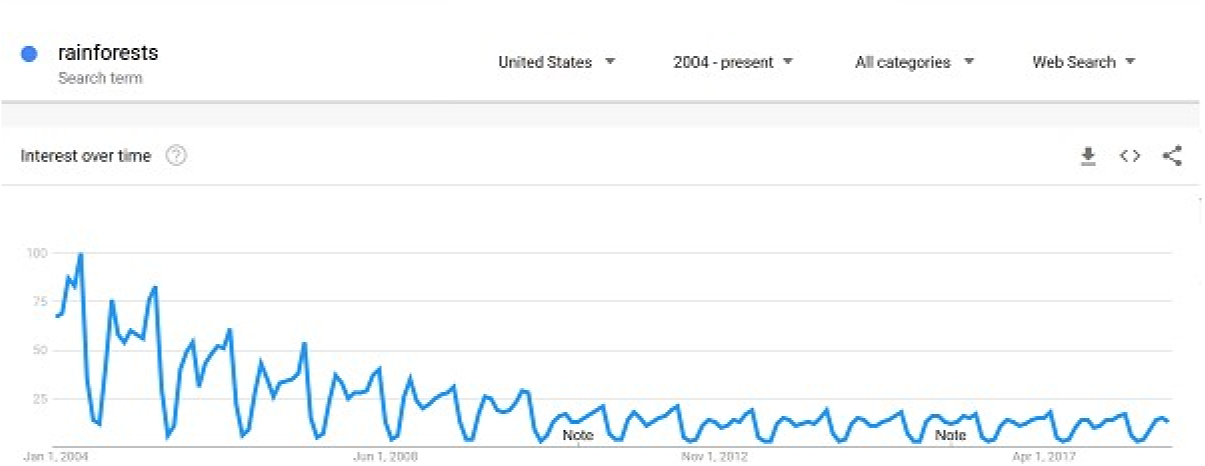
What *happened* to that? I don't mean the Pollution Demons: they're still around, I think one of them runs Trump's EPA now. What happened to everything else? To those teachers, those movies, that whole worldview?

Save The Whales. Save The Rainforest. Save Endangered Species. Save The Earth. Stop Slash-And-Burn. Stop Acid Rain. Earth Day Every Day. Reduce, Reuse, Recycle. Twenty-five years ago, each of those would invoke a whole acrimonious debate; to some, a battle-cry; to others, a sign of a dangerous fanaticism that would destroy the economy. Today they sound about as relevant as "Fifty-four forty or fight" and "Remember the Maine". Old slogans, emptied of their punch and fit only for bloodless historical study.

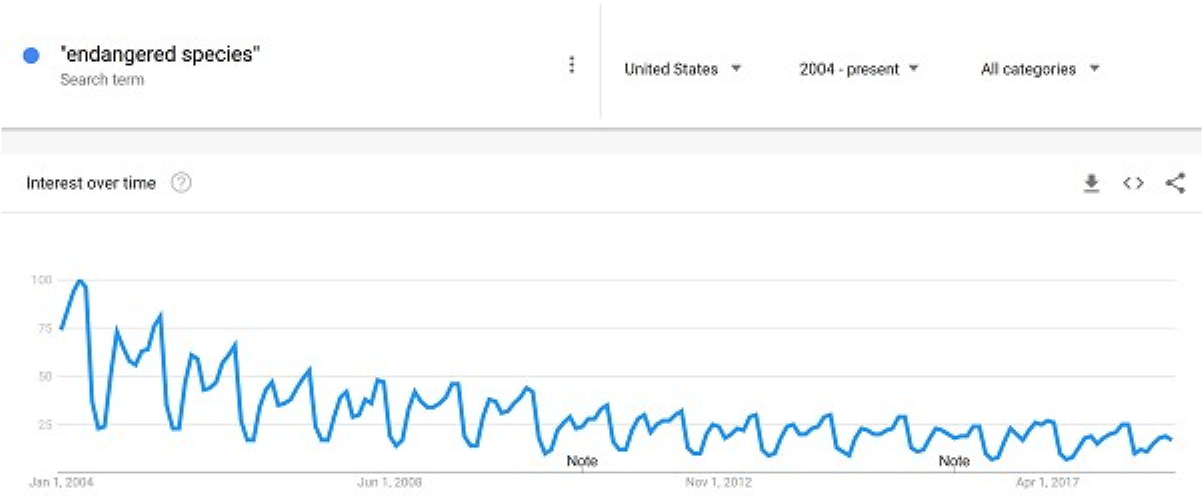
If you went back in time, turned off our Pollution Demon movie, and asked us to predict what would come of the environment twenty-five years, later, in 2018, I think we would imagine one of two scenarios. In the first, the world had become a renewable ecotopia where every child was taught to live in harmony with nature. In the second, we had failed in our struggle, the skies were grey, the rivers were brown, wild animals were a distant memory – but at least a few plucky children would still be telling us it wasn't too late, that we could start the tough job of cleaning up after ourselves and changing paths to that other option.

The idea that things wouldn't really change – that the environment would neither move noticeably forward or noticeably backwards – but that everyone would stop talking about environmentalism – that you could go years without hearing the words "endangered species" – that nobody would even know whether the rainforests were expanding or contracting – wouldn't even be on the radar. It would sound like some kind of weird bizarro-world.

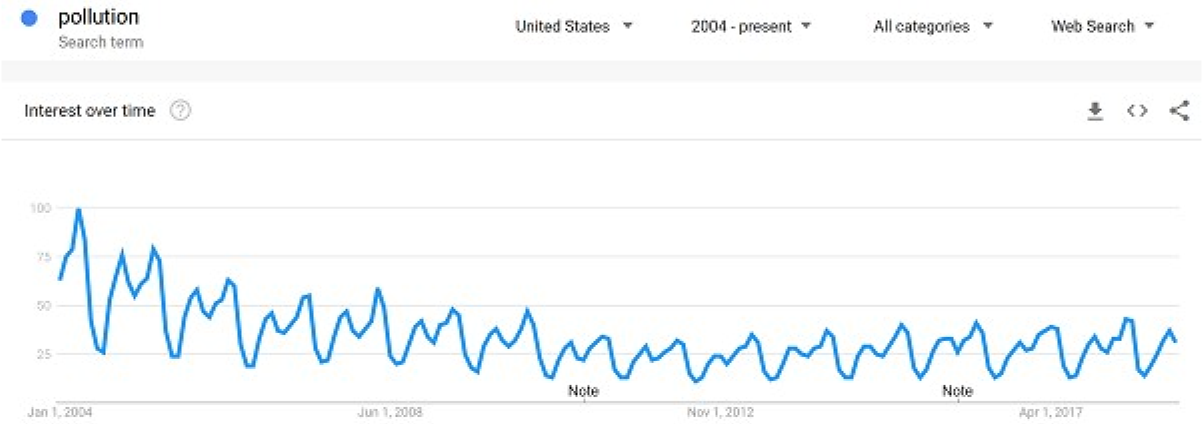
Just to prove I'm not imagining all this:



This is the volume of Google searches for "rainforests" over time. It goes up each year when school starts, and crashes again for summer vacation. But on average, there are only about 18% as many rainforest-related searches today as in 2004.



“Endangered species”, 25%



“Pollution”, 43%

And these are just since Google started tracking searches in 2004. The decline of 90s environmentalism must be even bigger.

So *what happened?*

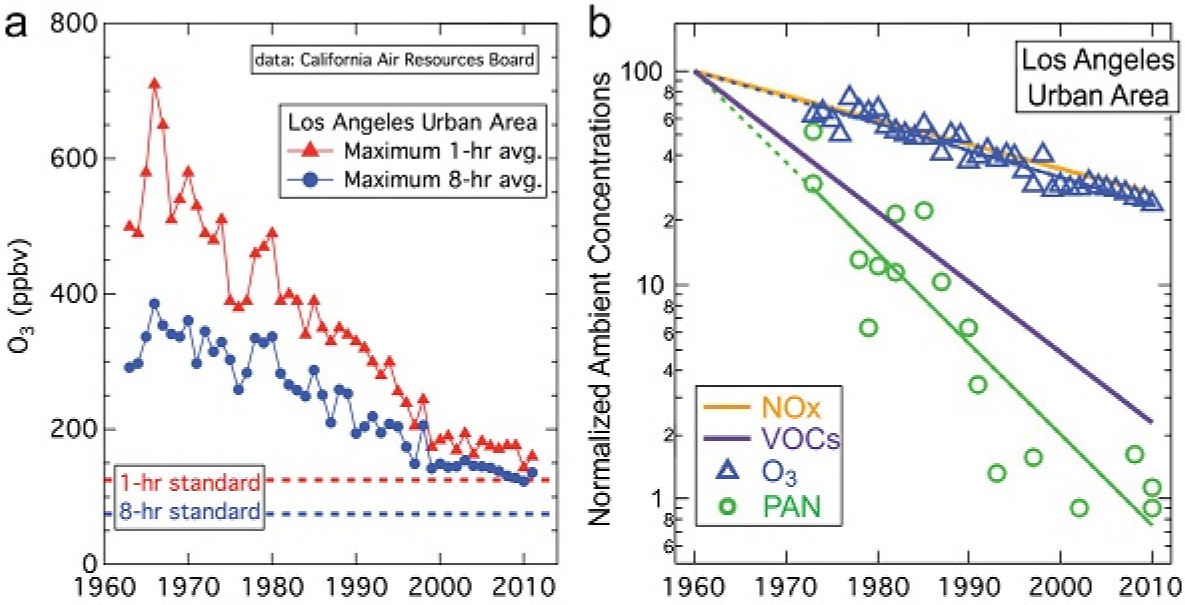
Every so often you’ll hear someone mutter darkly “You never hear about the ozone hole these days, guess that was a big nothingburger.” This summons a horde of environmentalists competing to point out that you never hear about the ozone hole these days because environmentalists successfully fixed it. There was a big conference in 1989 where all the nations of the world met together and agreed to stop using ozone-destroying chlorofluorocarbons, and the ozone hole is recovering according to schedule. When people use the ozone hole as an argument against alarmism, environmentalism is a victim of its own success.

So what about these other issues that have since fizzled out? Did environmentalists solve them? Did they never exist in the first place? Or are they still as bad as ever, and we've just stopped caring?

1. Air And Water Pollution

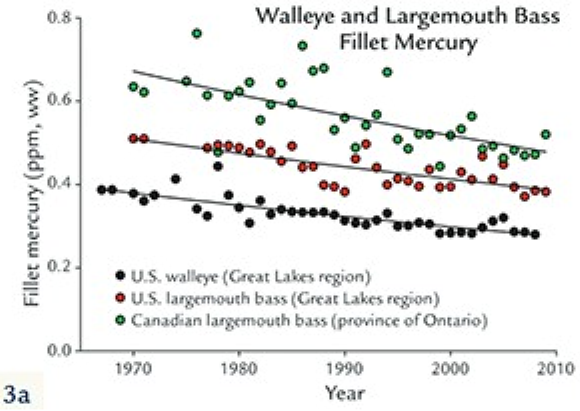
Have you seen what Chinese cities look like on a smoggy day? Trick question: neither have the Chinese. The US used to be like that. I grew up near Los Angeles during the 1990s. My mother tells the story of a time when I was very young and my grandparents came to visit from the Midwest. "It reminds us of home," they said, "it's so flat." "We're surrounded by mountains", my mother told them. We were. You couldn't see any of them.

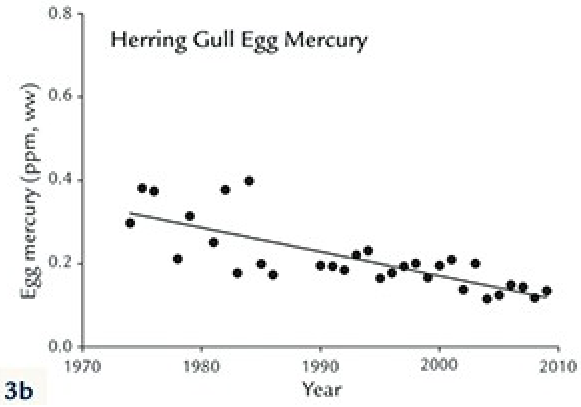
Environmentalists crusaded against this. Here are the results:



A lot of the credit goes to the Clean Air Act, passed in 1963 and tightened in 1990. Along with its more visible (pun intended) effects, scientists suspect it has prevented about 200,000 deaths from lung disease and a host of other cases of asthma, bronchitis, and even heart attacks.

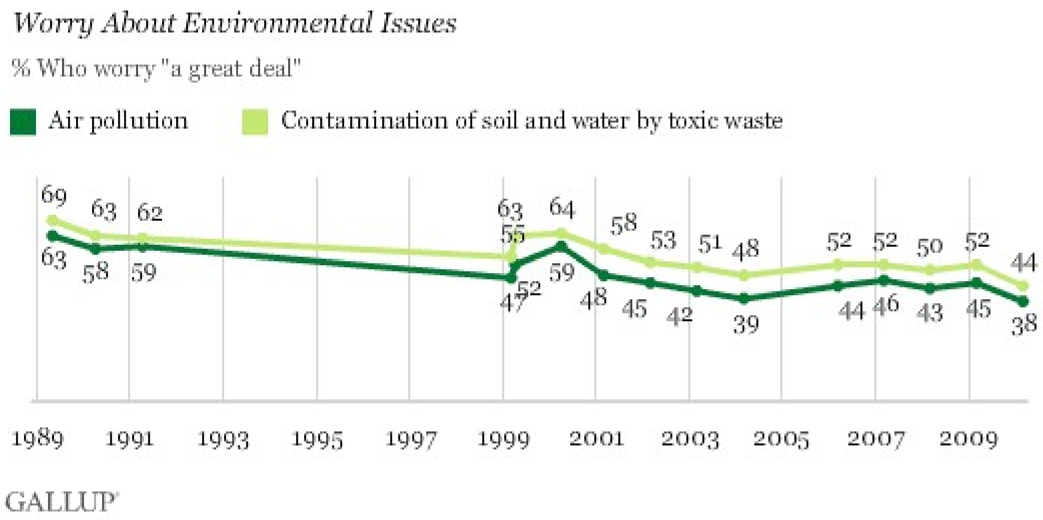
It’s hard to find great data on water because there are so many different kinds of water and so many different ways it can be polluted. But just to choose a random very bad thing, here’s mercury levels in Great Lakes fish:


3a


3b

I don’t know of anyone claiming this is anything other than a response to stricter environmental laws.

As a result of these victories, people are no longer as concerned about air and water pollution. From Gallup:



This seems like a clear case of good work.

Verdict: Environmental movement successfully solved this problem.

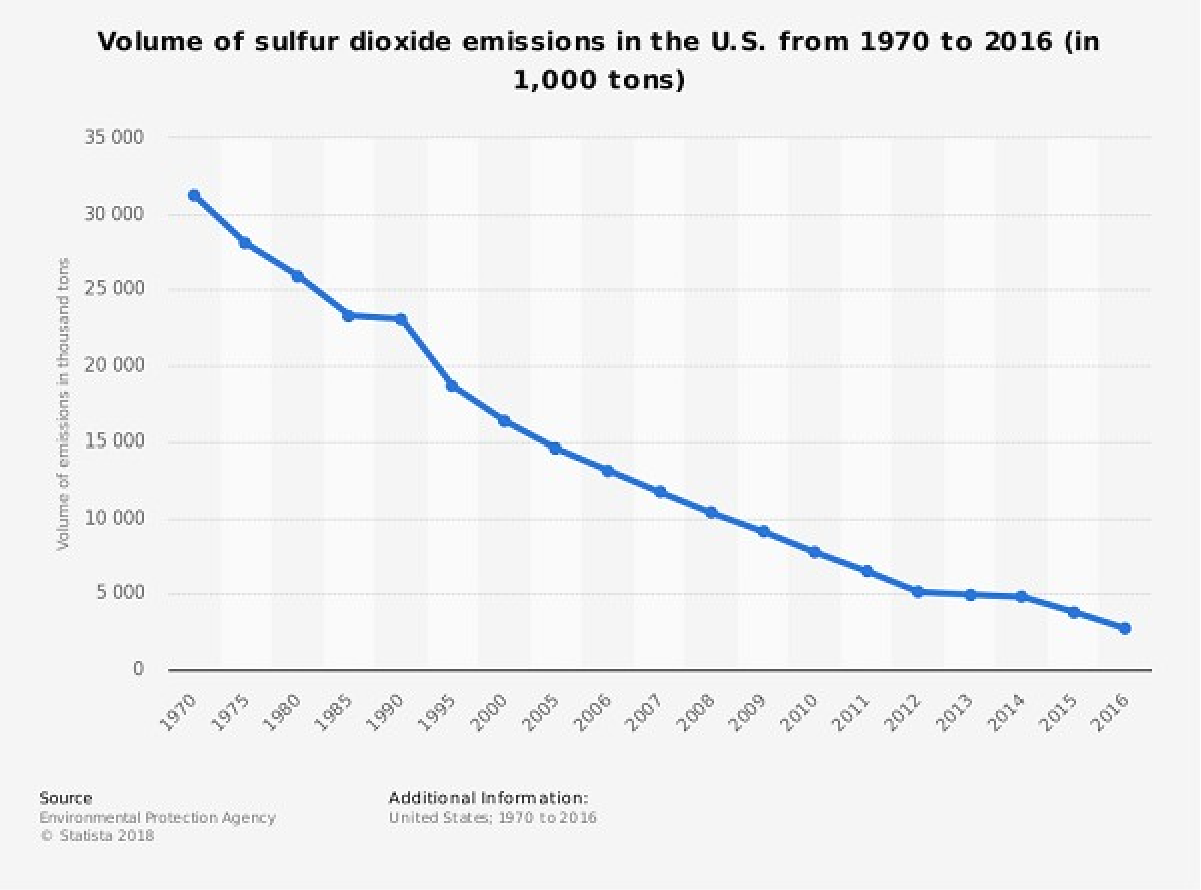
2. Acid Rain

Acid rain is a combination of rain and pollution which gets very acidic and destroys plants and structures. It was a staple of very early 90s environmentalism, and understandably so: the prospect of acid falling from the sky and dissolving everything is very attention-grabbing. I remember the discourse focusing on statues; George Washington’s marble face slowly melting under sizzling raindrops makes a heck of an image.

I am not the first person to notice that Washington’s face remains mercifully unmelted. In 2009, Slate asked Whatever Happened To Acid Rain?. EPA Blog, 2010: Whatever Happened To Acid Rain?. 2012, Mental Floss: What Ever Happened To Acid Rain? By 2018 the Internet had advanced, so here’s the Whatever Happened To Acid Rain Podcast.

Even the Encyclopedia Britannica, itself a good candidate for a “Whatever Happened To…” piece, has a What Happened To Acid Rain article.

Most of these sources say environmentalists solved acid rain by cutting down on emission of sulfur dioxide, the main offending chemical. A Bush I era cap-and-trade policy gets a lot of the credit in the US, but it looks like it was a broader effort than that:



There’s less clear data on rain acidity, but all my sources agree it has modestly declined in the US, thought it is still “between 2.5 and eight times more acidic than it should be”. Lakes and rivers are slowly recovering. On the other hand, in newly-industrializing countries like China and India, rain is becoming more acidic and they're going through some of the same issues we were in the 80s.

This picture is slightly complicated by some people who claim acid rain was always exaggerated and "we solved it" is a convenient retreat from acknowledging this (for what it's worth, these people tend to be global warming skeptics too). Most of them point to the 1990 National Acid Precipitation Assessment Program, a giant government investigation into the acid rain problem. I found a 1990 *New York Times* article on the report here:

> A comprehensive Federal report that was supposed to resolve the issue of how much damage is caused to forests by acid rain has come under criticism from some distinguished scientists who are reviewing it.
>
> The critics said that the report gave an incorrect impression that air pollution was not causing any large-scale problems for forest ecosystems. They also said that the report, still in draft form, ignored a number of studies suggesting serious air pollution problems.
>
> But other experts contend that the general conclusion of the report is essentially right. The report concluded that with the exception of damage to red spruce at high elevations in the East, forests in the United States are not suffering serious damage from acid rain […]
>
> The report now being reviewed is the final draft, completed at a cost of nearly $500 million. It examines the effects of other pollutants, like ozone, as well as acid deposits, and it concludes that air pollution causes far less environmental damage than has been feared.
>
> An interim report issued by the study group in 1987, before Dr. Mahoney became director, was sharply

criticized by many scientists. They contended that it tailored research findings into conclusions that matched the political goals of the Reagan Administration, which opposed new controls on air pollution. No such criticism has been leveled at the 28-volume final draft, which has been generally praised as a sound scientific document.

There is, however, some unhappiness among scientists with the volume dealing with forest health and productivity in the United States and Canada.

Dr. Ellis B. Cowling, associate for research at North Carolina State University’s College of Forest Resources, said in a telephone interview: ”The tone is that we don’t have a problem except in southern California, and with red spruce at high altitudes. That is not a fair statement of the state of scientific knowledge.” He added, ”Perhaps the authors were a bit too hasty in reaching conclusions.”

Dr. Cowling, who is highly regarded by colleagues as a conservative, solid scientist, wrote a memorandum to the authors of the forest health volume. He offered a series of suggestions for changing the wording of conclusions in ways that he said would reflect the state of science more accurately.

The first of those would change a finding that stated, ”The vast majority of forests in the United States and Canada are not affected by decline.” To be more consistent with the data, Dr. Cowling said, the conclusion should read: ”Most forests in the United States do not show unusual visible symptoms of stress, marked decreases in the rate of growth or significant increases in mortality.”

> Just because symptoms of forest decline are not currently visible, Dr. Cowling argued, does not rule out the possibility that they are under way.

This article also provides a summary of contemporaneous responses to NAPAP, which quotes study director James Mahoney’s summary of his own report: “The sky is not falling, but there is a problem that needs addressing.”

I cannot find anyone really challenging the NAPAP report nowadays, so I provisionally accept that the damage from acid rain, while real, was exaggerated at the time.

There’s a related debate about how much the lakes and streams affected have recovered. Some lakes and streams are naturally acidic; there is some debate over what percent of lake/stream acidity is natural vs. acid-rain-related. In recent years this debate has focused on whether lakes/streams have recovered after the SO2 decline; if they haven’t, this might suggest their problems were never human-activity-related in the first place.

Global warming skeptic blog Watt’s Up With That claims they haven’t:

> Possibly the greatest evidence against harmful effects of acid rain is the fact that acidic lakes have not “recovered” after most sulfur and nitrogen pollution was removed from the atmosphere. The 2011 NAPAP report to Congress stated that SO2 and NO2 emissions were down, that airborne concentrations were down, and that acid deposition from rainfall was down, but could not report that lake acidity was significantly reduced. The report states, “Scientists have observed delays in ecosystem recovery in the eastern United States despite decreases in emissions and deposition over the last 30

> years.” In other words, the pollution was mostly eliminated, but the lakes are still acidic.

You can find the report here. Like all long government reports, the details are ten zillion different trends in different directions that don’t form a cohesive narrative, and the executive summary is “things are good in all the ways that suggest we deserve more money, but bad in all the ways that suggest we need more money”, It is complicated enough that you shouldn’t trust my excerpting, but at least to me the relevant excerpts seem to be:

> Levels of acid neutralizing capacity (ANC), an indicator of the ability of a waterbody to neutralize acid deposition, have shown improvement from 1990 to 2008 at many lake and stream long-term monitoring sites in the eastern United States, including New England and the Adirondack Mountains. Many lakes and streams still have acidic conditions harmful to their biota even though the increases in ANC indicate that some recovery from acidification is occurring in sensitive aquatic ecosystems

And:

> Despite the environmental improvements reported here, research over the past few years indicates that recovery from the effects of acidification is not likely for many sensitive areas without additional decreases in acid deposition. Many published articles, as well as the modeling presented in this report, show that the SO2 and NOx emission reductions achieved under Title IV from power plants are not recognized as insufficient to achieve full recovery or to prevent further acidification in some regions.

So Watts seems to be mostly wrong when they say lakes are not recovering, but mostly right when they say ecosystems are not recovering. But NAPAP has some explanations for why ecosystems are not recovering: first, if you poison a lake and kill everything, then even if you remove the poison later everything is still dead. Second, there are complicated natural cycles that gradually wash old deposited land-based pollution into lakes, and it will be a long time before all the pollution deposited on land gets fully washed away. Third, maybe we haven't fought acid rain hard enough.

I think a lot of the epistemic work here is going to get done by people's respective stereotypes about the trustworthiness of global warming denialists vs. big government agencies whose budget depends on there being a problem. But my impression is that Watts' claim that poor recovery suggests acid rain was never a problem don't hold up very well.

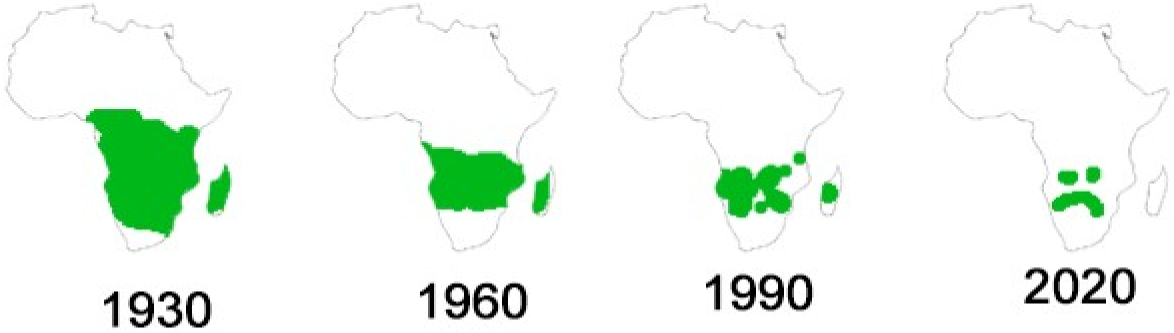
In any case, it's undeniable that rain has become a lot less acid lately, and likely that this has at least modest positive effects on some ecosystems as well as on the built environment. Anti-Confederate protesters have replaced acid rain as the number one threat to our statues. Our precious, precious statues. Someday they will be safe.

Verdict: *A little of everything: partly solved, partly alarmism, partly still going on.*

3. The Rainforests

Maybe the most typical image of 90s environmentalism is men in bulldozers clear-cutting a rainforest, while tapirs and tree sloths gently weep.

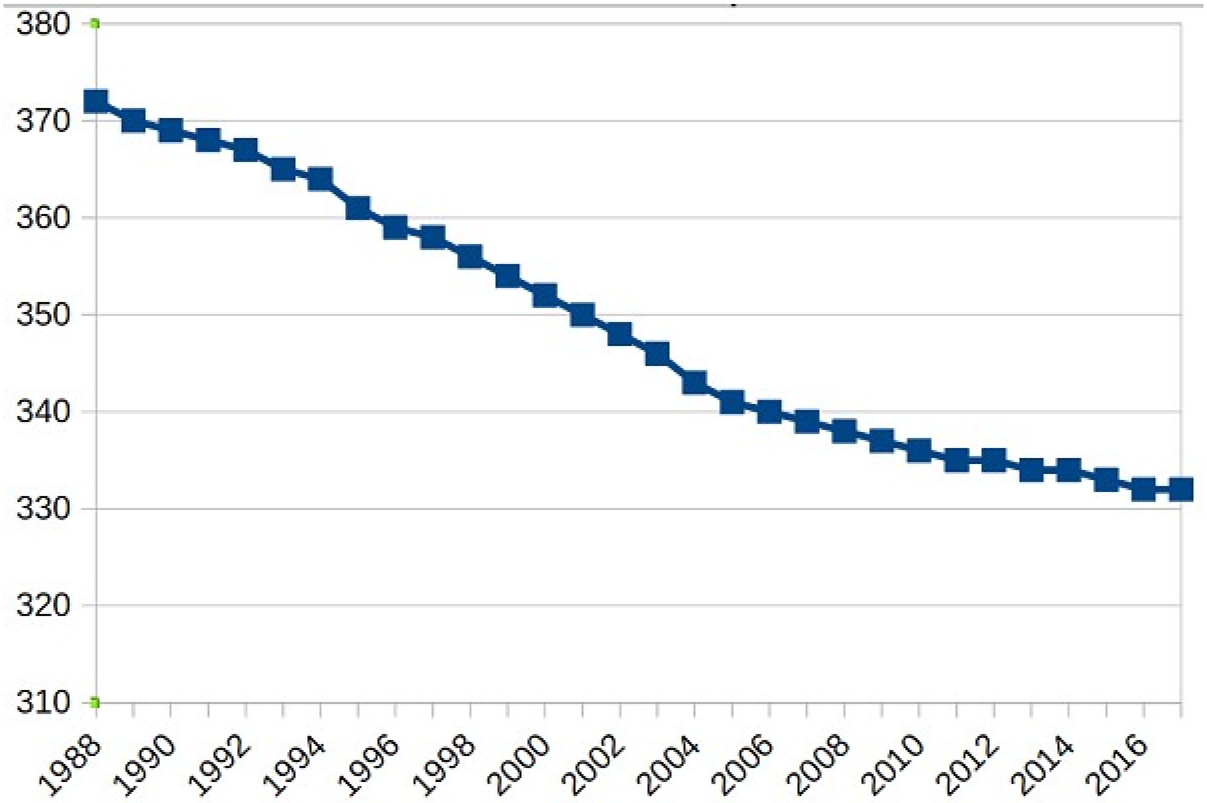
Or maybe it was the declining-rainforest-coverage-over-time-maps. I feel like about one in every three posters I saw as a child looked something like this:



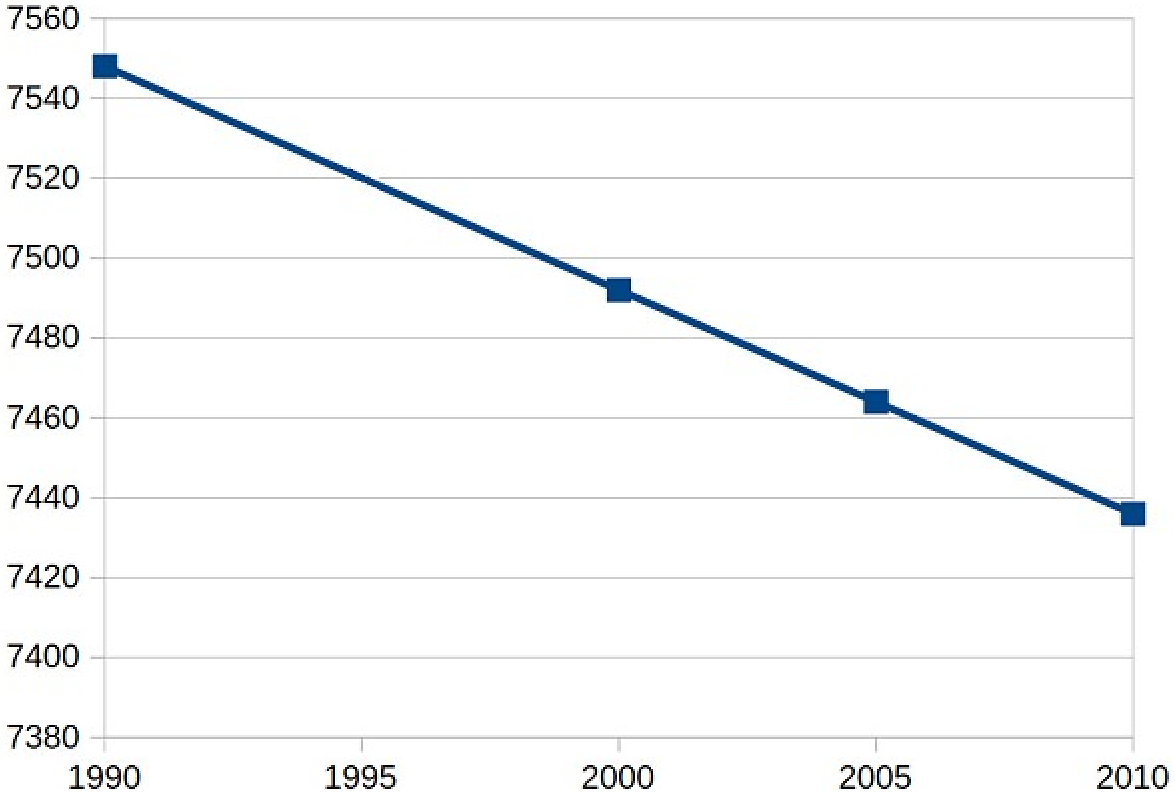
This is a fake example. Please stop asking me where I am getting the data from.

I thought surely nothing could be easier than digging up a few of them and seeing whether their 2020 predictions were right. But I can’t find them anywhere. According to the Internet, there is no such thing as 90s-era maps showing declining rainforest coverage over time. Can anyone else locate these?

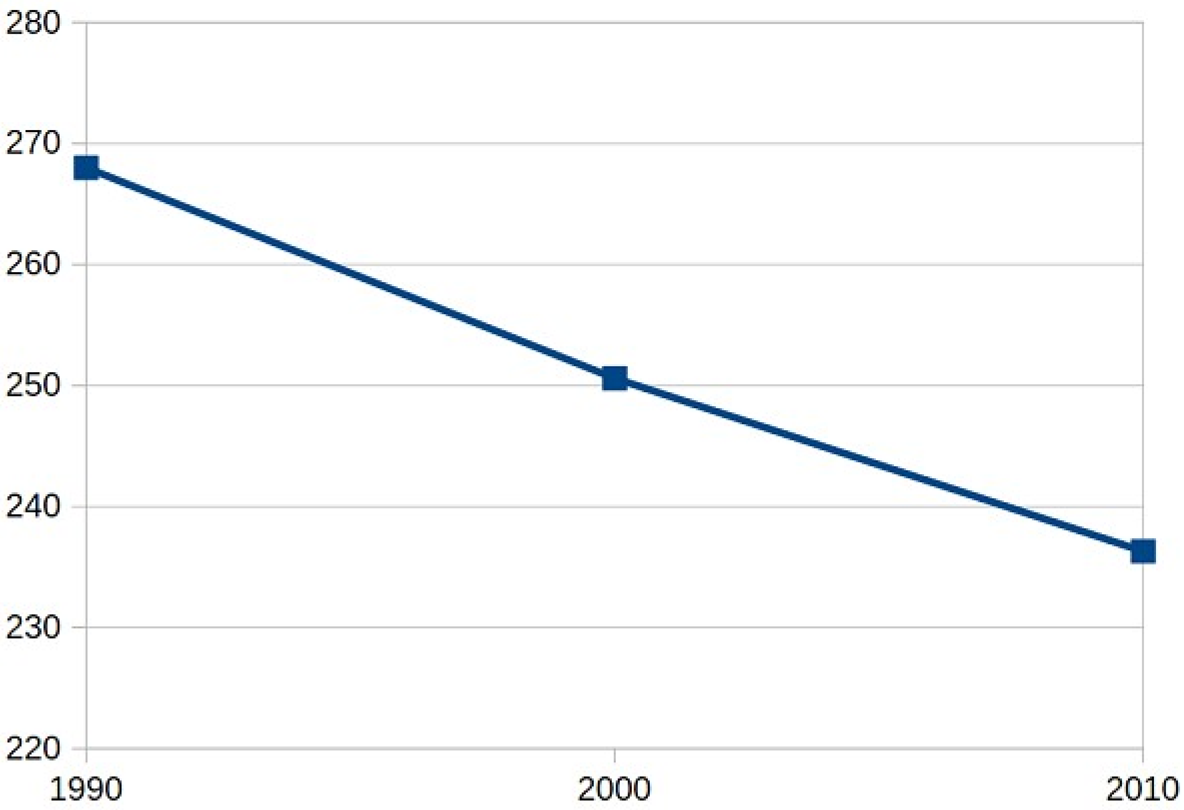
Anyway:



Here’s a graph of the size of the Amazon over time (source, note that the y-axis is not at zero). At 90s levels of deforestation, the Amazon would have disappeared in about 200 years. At current levels, it will disappear in about 400 years.



Here's the Congo (somewhat dubious source, same caveat). At the rates shown here it will be gone in 250 years – but it seems to have slowed after the period on the graph.



And here's Southeast Asia (source, same caveat). At this rate, the southeast Asian forest will be gone in 150 years, though some new papers are suggesting we may be underestimating the deforestation rate.

Overall it looks like deforestation may have decreased modestly in the Amazon (and possibly the Congo) since the 1990s. It has not decreased significantly in Southeast Asia, and whatever decreases have happened are not relevant to the scale of the problem.

The only good news is that all those "rainforests will be gone by 2050" posters were just wrong; there is more rainforest than that. But not *that much* more.

Verdict: The problem still exists, and we are just ignoring it now.

4. Endangered Species

So just find how many species go extinct each year, and whether it's a lot or a little, and then we'll know what's going on with this, right? Ha ha, as if.

On the one hand, the UN Environment Programme says that "150-200 species of plant, insect, bird and mammal become extinct every 24 hours."

On the other, nobody can name more than a single-digit number of species that go extinct in any given year. The 2017 list includes five: a bat, a cat, a flatworm, a lizard, and a snail. This matches longer-term surveys: Ceballos et al (2018) find that about 477 vertebrate species have gone extinct since 1900 – again, about five per year. And a recent survey found only four to eight bird species had disappeared since the turn of the century.

I have no idea where the 150-200 number per day comes from, and neither does anyone else. The closest I can find to a justification is this WWF page, which reminds us that if there are 100 million animals species, and "the extinction rate is just 0.01% per year", then *at least* 10,000 species go extinct every year (=200-300/day) – but all of these numbers are completely made up.

One could try to justify these estimates with something like "assume only one in a thousand species has been discovered and is monitored well enough to detect its extinction, so if we detect five extinctions per year then five thousand must be happening" – but I've never heard anyone actually say this. Also, with apologies to all the undiscovered species, if they're so tiny and uncommon as to never get discovered, it doesn't seem like their extinction is going to change very much.

Five known species going extinct per year may sound like a lot if you’re thinking it’s something like “rhinos, pandas, whales, spotted owls, and leopards”. But realistically there are 385 species of shrews. We could spend our entire yearly extinction budget on shrews for the next sixty years and still have more than enough kinds of shrews left to satisfy basically anybody.

I’m trying to think what the best counterargument to this would be – the best case that we really do need to consider species extinction a dire concern.

Maybe this is too vertebrate-centric, and there are lots of insects and plants and such going extinct all the time? But this List Of Recently Extinct Insects suggests that of about 6000 known insect species, only 50-100 have gone extinct in the past century. And one of those was this giant earwig which I really think the world is better without.

Or maybe we can’t directly predict the future from the past. Imagine 1000 square miles of rainforest with a homogenous distribution of species. Clear-cut 50% of the rainforest, and no extinctions. Clear-cut 90% of the rainforest, still no extinctions. Clear-cut 99%, maybe a few extinctions if you’re unlucky. Then clear-cut the last remaining 1% and everything dies. It seems like something like that might be happening – see for example this report that global animal populations have declined 58% over the past forty years.

But any concept of endangered species that focuses on “many well-known species will be gone soon” doesn’t seem consistent with the evidence.

Verdict: Partly alarmism, partly still going on.

5. More And More Trash Piling Up Until The Whole World Is Just A Giant Mountain Of Trash

Wait, what? Was this really a concern? Did I really spend my primary school years being told that if I didn’t vigilantly recycle everything, one day I would be submerged beneath a sea of trash, breathing by means of a trash snorkel? Am I hallucinating all of this?

As usual, it turns out to be the Mafia’s fault. In the 1980s, mob boss Salvatore Avellino took over New York City’s landfill industry, and in a shocking development which nobody could have predicted, was corrupt. New York City soon ran out of landfill space. Somehow all of its excess trash ended up on a barge called the MOBRO-4000, because the Eighties, and this barge apparently sailed up and down the east coast of North America searching for a place to deposit its trash. In its many exciting adventures it reached the coast of Belize, got involved in a confrontation with the Mexican Navy, and finally went back to New York, where at some point landfill space was found and the crisis was over.

But a giant boat full of trash made a really memorable image, and it got nationwide news coverage, and environmentalists took advantage of this to tell everyone there was no more landfill space anywhere in the world and we all had to recycle right now. According to Wikipedia:

> At the time, the Mobro 4000 incident was widely cited by environmentalists and the media as emblematic of the solid-waste disposal crisis in the United States due to a shortage of landfill space: almost 3,000 municipal landfills had closed between 1982 and 1987. It triggered much national public discussion about waste disposal, and may have been a factor in increased recycling rates in the late 1980s and after. It was this that caused it to be included in an episode of Penn & Teller: Bullshit! (season 2, episode 5) in which they debunk many recycling myths.

I’m even absolutely right in remembering primary school lessons centered around garbage covering the Earth and killing everybody. Here’s a *New York Times* article from 1996 – ie after the crisis had a little bit of time to fade – lightly mocking the new curricula that followed in its wake:

> After the litter hunt in Ms. Aponte’s science classroom, it was time for a guest lecturer on garbage. A fifth-grade class was brought in to hear Joanne Dittersdorf, the director of environmental education for the Environmental Action Coalition, a nonprofit group based in New York. Her slide show began with a 19th-century photograph of a street in New York strewn with garbage.
>
> “Why can’t we keep throwing out garbage that way?” Dittersdorf asked.
>
> “It’ll keep piling up and we won’t have any place to put it.”
>
> “The earth would be called the Trash Can.”
>
> “The garbage will soon, like, take over the whole world and, like, kill everybody.”
>
> Dittersdorf asked the children to examine their lives. “Does anyone here ever have takeout food?” A few students confessed, and Dittersdorf gently scolded them. “A lot of garbage there.”
>
> She showed a slide illustrating New Yorkers’ total annual production of garbage: a pile big enough to fill 15 city blocks to a height of 20 stories. ‘There are a lot of landfills in New York City,” Dittersdorf said, “but we’ve run out of space.”

From the same beginning-of-the-backlash period we also get this 1995 Foundation for Economic Education piece, Are We Burying Ourselves In Garbage?:

> A popular idea in public discourse today is that the United States produces an overwhelming amount of trash–so much that our landfills will not be able to handle the quantity. The most eloquent symbol of this viewpoint was the "garbage barge," which in the late 1980s left Long Island and could not find a port or country willing to accept its 3.168 tons of refuse. [But] the actual data (such as they are) on the amount of municipal solid waste produced present us with more questions than answers.

This article also deserves note for hitting on a brilliant solution:

> The crisis mentality has distorted judgment of waste disposal. The notion that modern America is especially wasteful is demonstrably wrong, both in terms of the last decades as well as the last 100 years. The idea that our landfills are literally "running out" is even less credible. If in the next century major portions of the United States really need to export their refuse to other states, a "gold mine" for refuse burial does exist: South Dakota. This state is geologically, economically, and politically almost ideal for massive municipal solid waste management.

None of this is a joke. This is how your parents did Discourse, people.

But it turns out capitalism works: if there's a shortage of landfills, that incentivizes people to create new landfills. Also, the world is very large and it is hard to cover a significant portion of it in trash. There was a brief blip as cities figured out how to pay for more waste disposal, and then nobody ever worried about the problem again. Recycling remained inefficient and of dubious benefit, and never really caught on.

There is still an international problem as Third World countries struggle with infrastructure issues around trash disposal. You still see occasional articles like Huffington Post's People Are Living In Landfills As The World Drowns In Its Own Trash, from earlier this year. But I think in general nobody in the First World still considers this a major problem.

Well, *almost* nobody:

> The earth, our home, is beginning to look more and more like an immense pile of filth.
>
> — Pope Francis (@Pontifex) June 18, 2015

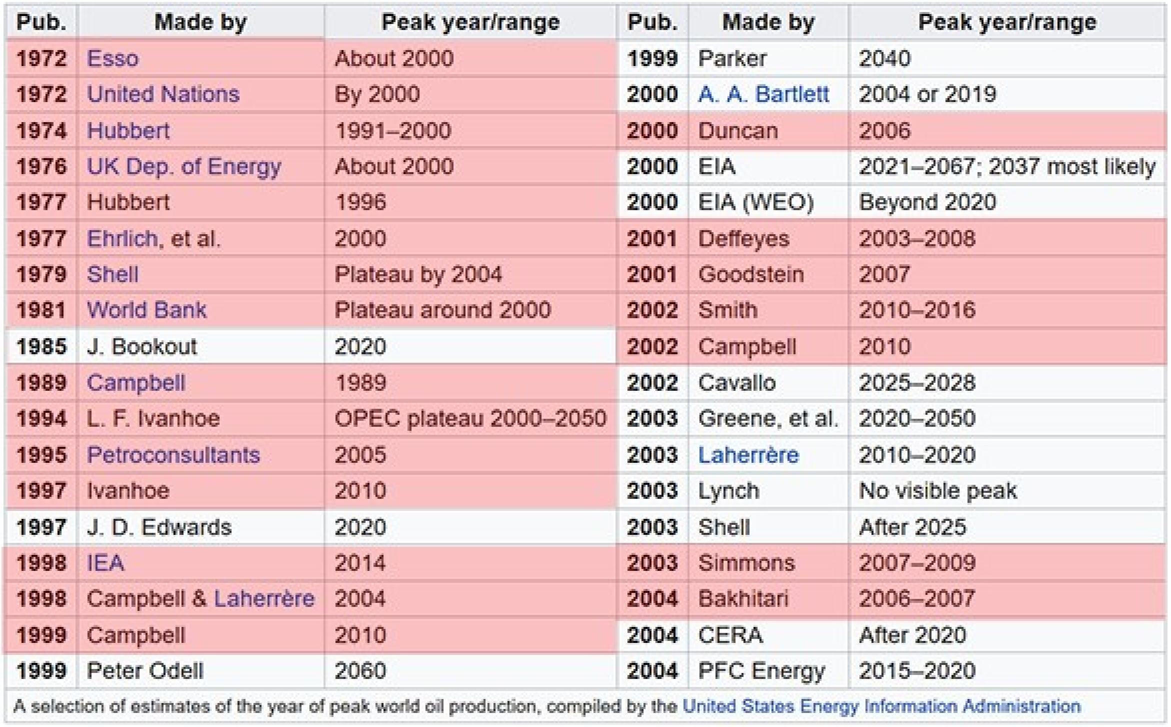
Verdict: Alarmist. So, so, alarmist.

6. Peak Resource

Is the earth's ballooning human population using up resources at an unsustainable rate?

Technically the answer must be "yes", since by definition nonrenewable resources have to run out at some point. But when? Long after we have escaped to space and gotten access to shiny new resources? Or soon enough that we have to worry about it?

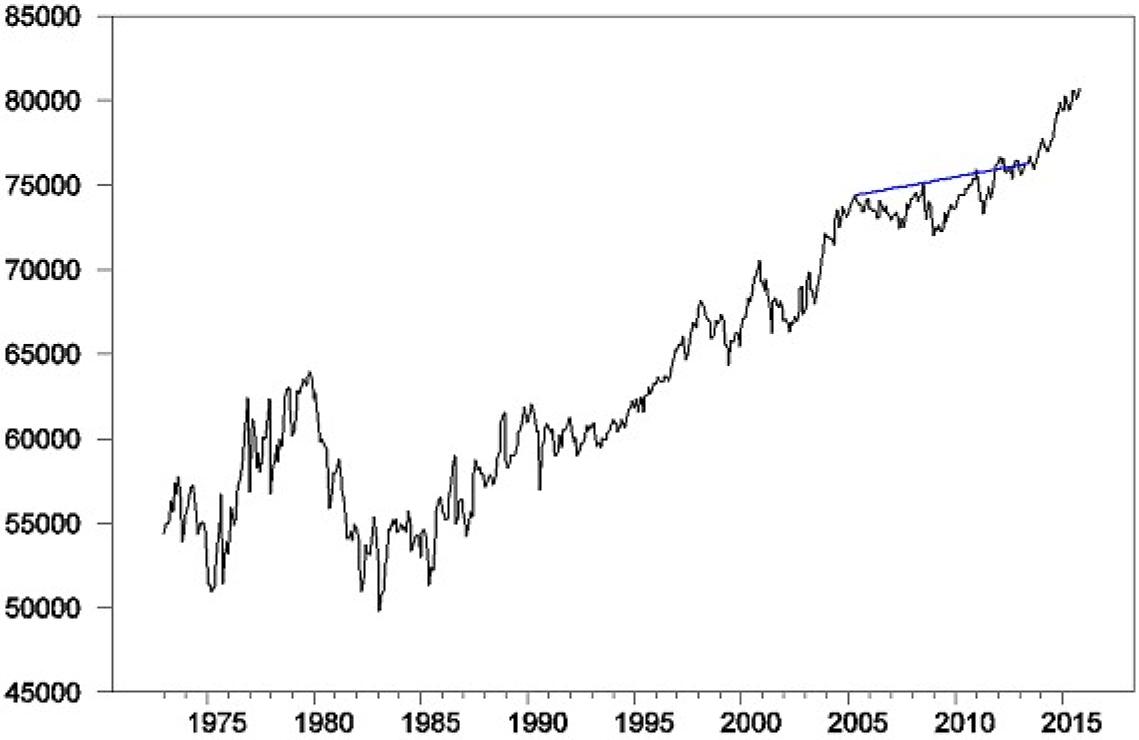
A big part of 90s environmentalism involved worrying that it was the latter. A particular concern was "peak oil", the point at which we had exhausted so much of the world's oil that production rates declined every year thereafter and oil started becoming gradually rarer and more expensive. Wikipedia has a helpful table of people’s peak oil predictions. I’ve highlighted the ones that have already passed in red.



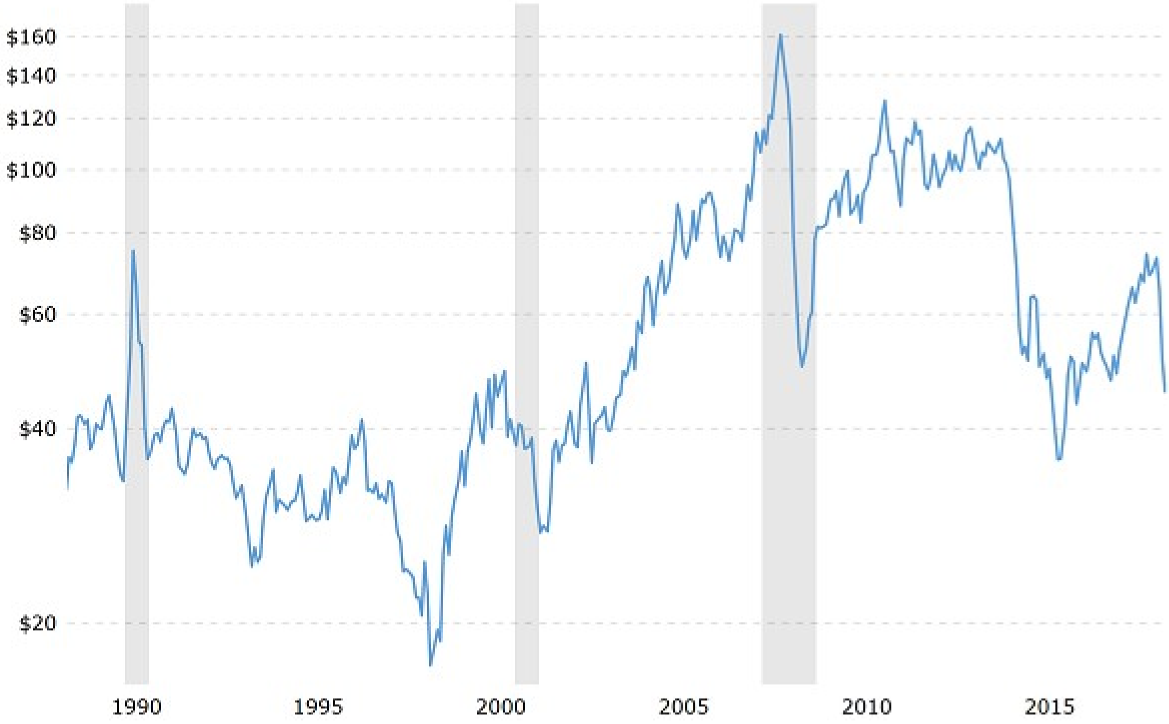
Pub.	Made by	Peak year/range	Pub.	Made by	Peak year/range
1972	Esso	About 2000	**1999**	Parker	2040
1972	United Nations	By 2000	**2000**	A. A. Bartlett	2004 or 2019
1974	Hubbert	1991–2000	**2000**	Duncan	2006
1976	UK Dep. of Energy	About 2000	**2000**	EIA	2021–2067; 2037 most likely
1977	Hubbert	1996	**2000**	EIA (WEO)	Beyond 2020
1977	Ehrlich, et al.	2000	**2001**	Deffeyes	2003–2008
1979	Shell	Plateau by 2004	**2001**	Goodstein	2007
1981	World Bank	Plateau around 2000	**2002**	Smith	2010–2016
1985	J. Bookout	2020	**2002**	Campbell	2010
1989	Campbell	1989	**2002**	Cavallo	2025–2028
1994	L. F. Ivanhoe	OPEC plateau 2000–2050	**2003**	Greene, et al.	2020–2050
1995	Petroconsultants	2005	**2003**	Laherrère	2010–2020
1997	Ivanhoe	2010	**2003**	Lynch	No visible peak
1997	J. D. Edwards	2020	**2003**	Shell	After 2025
1998	IEA	2014	**2003**	Simmons	2007–2009
1998	Campbell & Laherrère	2004	**2004**	Bakhitari	2006–2007
1999	Campbell	2010	**2004**	CERA	After 2020
1999	Peter Odell	2060	**2004**	PFC Energy	2015–2020

A selection of estimates of the year of peak world oil production, compiled by the United States Energy Information Administration

Almost everyone working before 2000 thought we would have reached peak oil by now. But here’s world oil production over time:



And the price of oil:



What happened? People discovered fracking and other paradigm-shifting techniques to extract oil from shale, which opened up vast new previously-inaccessible oil fields. The peak oil predictors might call this unfair – they calculated correctly given the technology they knew about – but the *whole argument* of the people who say we don't have to worry about peak resource (sometimes called "cornucopians") is that technology will advance fast enough to satisfy our resource needs. In this case they were right.

What about non-oil resources?

In 1980, leading environmental scientist and peak-resource proponent Paul Ehrlich made a bet with cornucopian economist Julian Simon about how resource prices would change over the next decade. The Simon-Ehrlich Wager has become a famous example of futurology done right – two people with different theories implying different predictions coming together, agreeing on exactly what each of their theories implied, and then publicly committing to put them to the test. According to Reb Wiki:

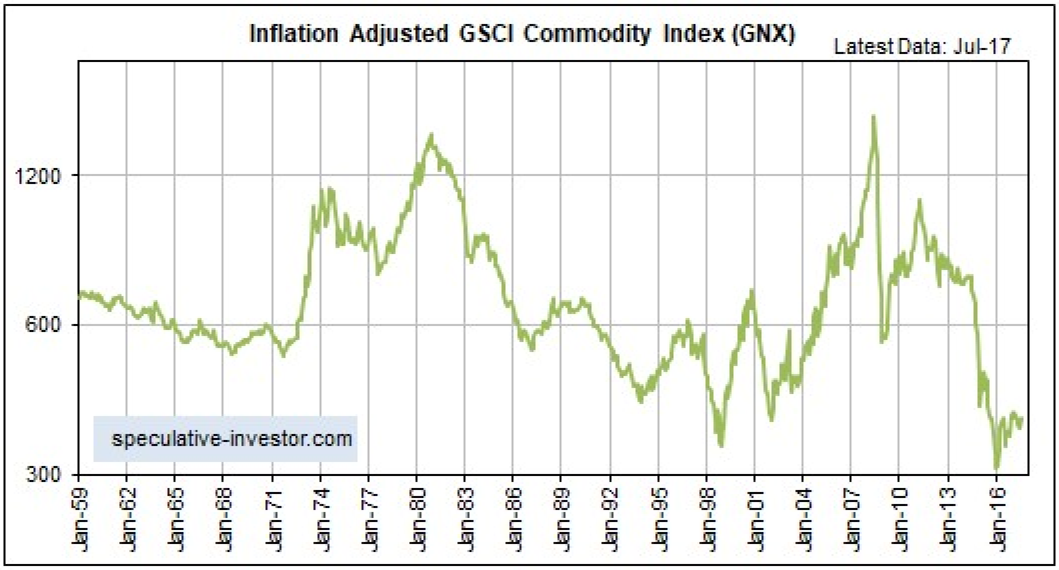
> Simon challenged Ehrlich to choose any raw material he wanted and a date more than a year away, and he would wager on the inflation-adjusted prices decreasing as opposed to increasing. Ehrlich chose copper, chromium, nickel, tin, and tungsten. The bet was formalized on September 29, 1980, with September 29, 1990 as the payoff date. Ehrlich lost the bet, as all five commodities that were bet on declined in price from 1980 through 1990, the wager period.

Looks pretty good for Simon and the cornucopians. But the article continues:

> Ehrlich could have won if the bet had been for a different ten-year period. Ehrlich wrote that the five metals in question had increased in price between the years 1950 to 1975. Asset manager Jeremy Grantham wrote that if the Simon–Ehrlich wager had been for a longer period (from 1980 to 2011), then Simon would have lost on four of the five metals. He also noted that if the wager had been expanded to "all of the most important commodities," instead of just five metals, over that longer period of 1980 to 2011, then Simon would have lost "by a lot." Economist Mark J. Perry noted that for an even longer period of time, from 1934 to 2013, the inflation-adjusted price of the Dow Jones-AIG Commodity Index showed "an overall significant downward trend" and concluded that Simon was "more right than lucky". Economist Tim Worstall wrote that "The end result of all of this is that yes, it is true that Ehrlich could have, would have, won the bet depending upon the starting date. … But the long term trend for metals at least is downwards."

How about today? An econblogger is still keeping track of the Ehrlich-Simon wager, and finds that as of August 2017, Simon (who is now dead) is still winning; a basket of the five metals involved still costs less than it did in 1980.

Can we zoom out even further? There are a bunch of commodity indices that do for commodities what the Dow Jones does for stocks. I chose the Standard & Poor Goldman-Sachs Commodity Index kind of randomly because they were a familiar name and it was easy to find which goods they included. I'm not quite sure I'm doing this right, but this seems to be the most relevant graph:



The price of commodities in general is still lower than in 1980 (also, with this graph it becomes clear Ehrlich was really unlucky in which year he started his wager).

I have never heard anyone claim that this represents an environmentalist victory: I don’t think there was any large-scale attempt to conserve or recycle chromium/tungsten/whatever that led to its current abundance. I think this was just a victory for resource extraction technology.

There are still theoretical reasons to think we have to run out of stuff eventually. But in terms of how the past 25 years have treated 90s-era concerns about resource depletion, it’s hard to answer anything other than “savagely”.

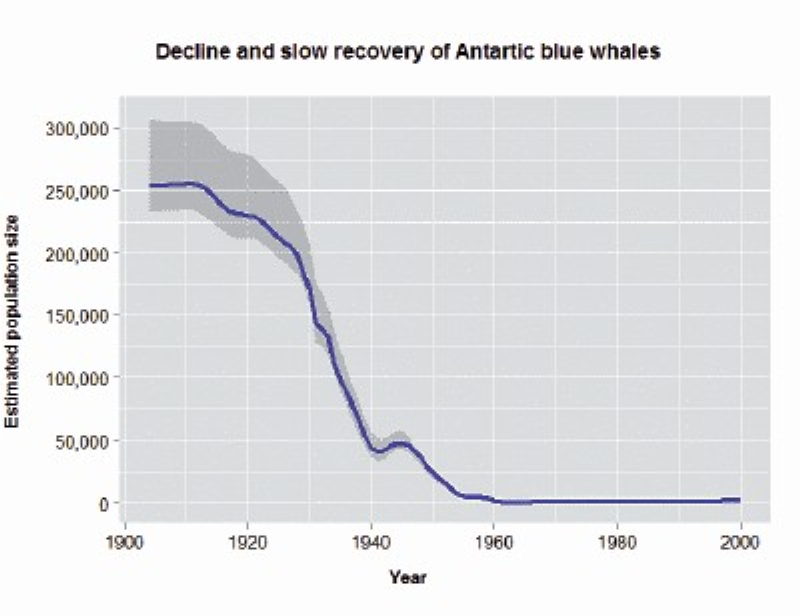
Verdict: *Alarmist*

7. Saving The Whales

I remember frequently being told we had to do this. Apparently it paid off, since a global moratorium on whaling was signed in 1982.

The ban is not perfect. Indigenous peoples are allowed to hunt whales in traditional ways. Japan pretends their whaling is for “scientific purposes” and has so far gotten away with it. Norway and Iceland never signed the moratorium and continue to whale.

But overall, things are going pretty well. There aren’t a lot of graphs, but the International Whaling Commission (which despite its name is against whaling) says blue whale populations are increasing at about 8%/year, humpback whales around 10%, and fin whales around 5%. Those sound pretty good, but they have to be taken in context:



Okay, fine. There’s one graph. But it’s really depressing. See that tiny micro-bump at the end? That represents progress.

Verdict: *Environmental movement successfully solved this problem.*

8. Concluding Thoughts

This was not a very conclusive exercise. When I add these up – as if that were at all an acceptable thing to do! – I get

2⅓ that were solved, 2⅚ that were alarmism, and 1⅚ that continue. So there is not much to be said about them as a group. Some were solved through heroic effort. Some turned out to be completely made up. Some of them are still out there but have stopped capturing the public's attention.

Victories I can understand. It's the latter two categories that confuse me.

How did the non-problems fade away? There was no moment when some brave iconoclast posted ninety-five theses to the door of the local recycling center and said "No! There is not a landfill crisis!" I mean, John Tierney wrote things along those lines, and did a great job of it. But he's not a household name and there was never a time when everyone said "Oh, John Tierney is right, let's stop worrying about this." The people who stopped worrying about this never heard of John Tierney. At some point people just went from being very worried about the landfill crisis to shaking their heads and saying "The world getting full of trash? Sounds pretty stupid."

And the story with peak resources seems entirely different. You will still occasionally see people saying "The Earth can't support our greed, soon we will run out of everything", and reasonable people will nod along with this and admit it is very wise. But you hear it like once a year now, as opposed to it being a constant refrain. This idea was never intellectually defeated at all, at least not on the popular level. It just faded away.

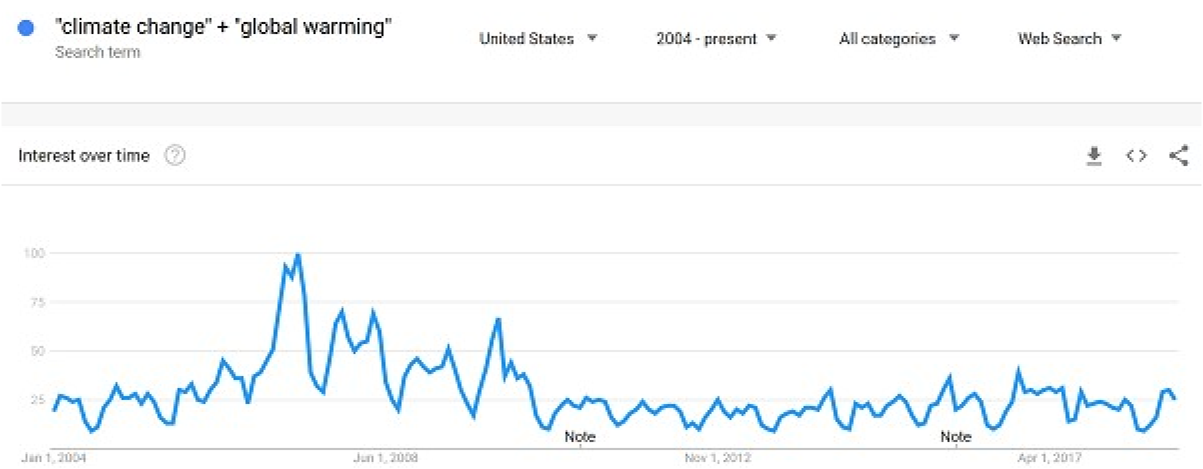
Was there some rarified level of intellectual debate where these ideas lost out? And then, denied their support from the commanding heights of the ivory tower, did journalists stop writing about them, schoolteachers stop teaching them, and then eventually the public – who have no will of their own and have to be told what matters – wander off and do something else?

Or was the change bottom-up? Did the public, after the millionth editorial on the trash crisis, say “Okay, whatever”, such that journalists realized this was no longer a good way to sell newspaper subscriptions? Is there a natural news mega-cycle of a decade or so, after which the public gets tired of hearing about a certain story, the intellectuals get tired of talking about it, every possible angle has been explored, and people move on, whether or not it was solved? Does this explain why the rainforests, a real problem that is still going on, similarly lost public attention?

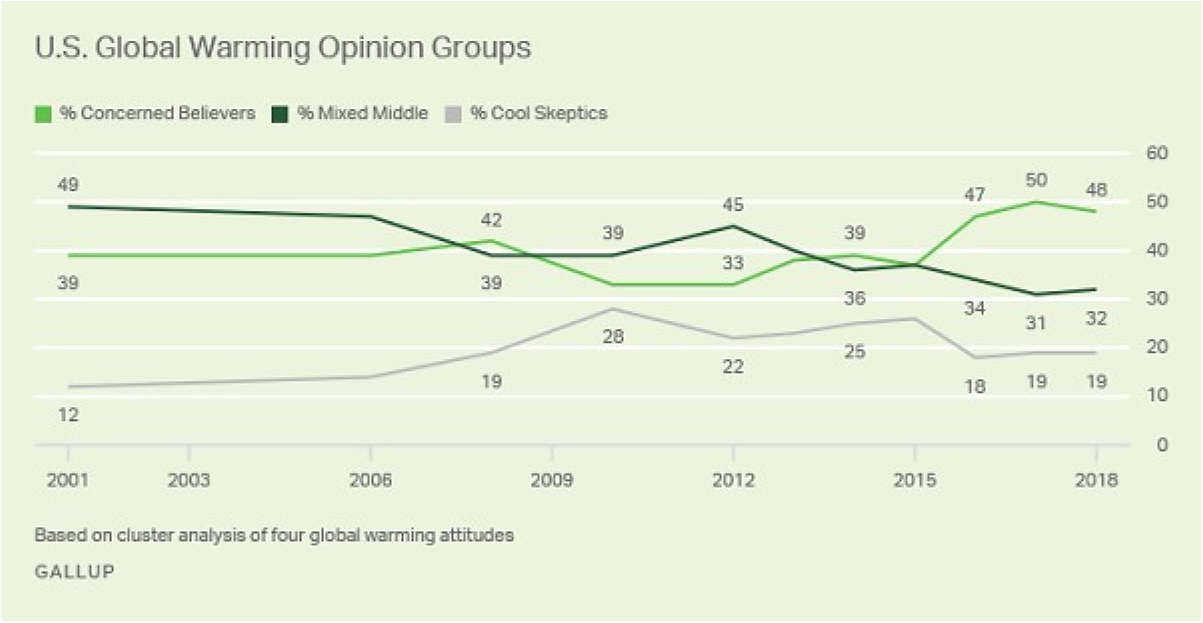
Or maybe climate change took over everything, became so important that everything else faded into the background. This is certainly how it feels to me. Whenever I hear about the rainforests nowadays, it’s as a footnote to some global warming story where they add that we should save the rainforest as a carbon sink. Whenever I hear about landfills or recycling today, it’s in the context of trash giving off greenhouse gases. It feels almost like some primitive barter system has been converted to a modern economy, with tons of CO2 emission as the universal interchangeable currency that can be used to put a number value on all environmental issues. Can’t figure out a way to convert whales into a carbon sink? Guess they’ll have to go.

(I wrote that, then remembered I lived in 21st century America, did a Google search, and sure enough there are dozens of articles arguing that saving whales is an efficient way to neutralize greenhouse gases)

But as attractive as this picture is, it’s hard to find the supporting data. There’s just not hard evidence that we care more about global warming than we did fifteen or twenty years ago:



Here's the Google Trends. There was a lot of interest in 2006, which I think gets attributed to Al Gore's *An Inconvenient Truth* in that year, but not a lot of signs of increase today.



Based on cluster analysis of four global warming attitudes

GALLUP

Here's Gallup. It at least shows a spike starting in 2016 – but given its timing and the lack of obvious 2016 global-warming related events, I think it’s probably just another Trump backlash effect.

If global warming is eating all the other environmental issues, it doesn’t seem to be extracting that much nutrition from their corpses. And the ozone hole – probably the most global-warming-like issue of the last generation – managed to gather popular support at the same time that people were worried about a host of other things. I don’t know. Maybe given the public’s tendency to get bored of an issue after a decade or so, global warming has to cannibalize the rest of environmentalism just to survive at all. Depressing if true.

Or maybe it’s a *zeitgeist* thing. For some reason, it’s hard to imagine 2018 being the Year Of Rainforest Concern. There’s something very 90s Optimism about worrying about the rainforests, something where even the warnings of doom have a cheerful ring to them. I remember a Rainforest Charity Box at my local mall as a kid, promising that if you donated $10, you would save a brightly colored parrot, and if you donated $50, you might save a jaguar. Who thinks that way these days? Now if you donate some amount to stopping global warming, you will have won yourself a lecture from a bunch of people telling you that still doesn’t mean you have the right to feel good about yourself, and the world is going to fry regardless. Have we just passed the point where anybody can care about crisp mountain streams or frolicking snow leopards any more?

The most important thing I take away from the exercise is a sort of postmodern insight into the way environmental issues are constructed. This is *definitely not* me saying they are all made up; many of them are very real. But the mapping from real crisis to social panic is tenuous, contingent, and historical. Sometimes random things that shouldn't matter get magnified into the *issue du jour*; other times giant world-threatening crises manage to slip everyone's attention.

Imagine that twenty years from now, nobody cares or talks about global warming. It hasn't been debunked. It's still happening. People just stopped considering it interesting. Every so often some webzine or VR-holozine or whatever will publish a "Whatever Happened To Global Warming" story, and you'll hear that global temperatures are up X degrees centigrade since 2000 and that explains Y percent of recent devastating hurricanes. Then everyone will go back to worrying about Robo-Trump or Mecha-Putin or whatever.

If this sounds absurd, I think it's no weirder than what's happened to 90s environmentalism and the issues it cared about.

Bay Area SSC Meetup 1/6

Join me at 3:30 PM on Sunday, 1/6 for the traditional once-every-three-months big SSC Bay Area meetup. Meet ~~on the Berkeley campus at the open space beside the intersection of West Circle and Free Speech Bikeway~~ at 3045 Shattuck Avenue, Berkeley, where we will be indoors and safe from the rain.

Special guest Professor Stephen Hsu, who among his other accomplishments is a scientific advisor to the Beijing Genomics Institute and a founder of Genomic Prediction, Inc. Come discuss embryo sequencing with him, and congratulate him on his recent demotion to *second*-craziest Chinese-affiliated mad scientist working on embryo genetics.

Preregistration Of Investigations For The 2019 SSC Survey

This post is about the 2019 SSC Survey. If you've read at least one blog post here before, please **take the survey** if you haven't already. Please don't read on until you've taken it, since this post could bias your results.

1. Can we confirm or disconfirm different corn-eating profiles of algebraists vs. analysts?

2. Can we replicate the study showing that people who eat more beef jerky are more likely to be hospitalized for bipolar mania?

3. Are there differences in side effects among SSRIs? (to be limited to people taking an SSRI one month or more, will be looked at both effect by effect, and with a lumped-together side effect index where each mild effect counts as 1 point and each severe effect as 3 points)

4. Is there a difference in people's efficacy ratings for SSRIs (SSRI Effectiveness, SSRI Overall) depending on whether the person was taking the SSRI for depression vs. for anxiety?

5. What percent of people coming off SSRIs experience discontinuation symptoms? Are there differences among different agents? (main analysis to be limited to people who were taking an SSRI at least a few months, discontinued with a gradual taper lasting at least a few weeks, and were not cross-tapering onto any other psychiatric medication).

6. Are people more likely to attribute success to hard work/talent rather than luck if they are from a higher childhood social class? What about a higher current social

class? What about if they have moved upward throughout their lifetime?

7. Are people less likely to support psychiatric commitment if they have been committed themselves? What about if they have a frequently-committed psychiatric issue (schizophrenia, bipolar, borderline, eating disorder)? Are they more likely to support commitment if they have a family member with those conditions, but don’t have it themselves?

8. Is there any support for the idea of life history strategies? (does early age of sexual debut and/or high number of sexual partners correlate with regular drug use, with leaving school earlier, and higher risk-taking? What about with being more likely to ask a partner out early or have nonconsensual sex? What about with various psychiatric disorders?)

9. Do paternal and maternal age correlate with risk of the psychiatric disorders? What about with the various self-ratings? What about with SAT score?

10. The question about labeling vegetarian foods (ie can a burger made of pea protein call itself a “burger”) seems to get at questions of essentialism vs. pragmatism in the same way as controversies about transgender. Do people who support essentialist labeling of food also support essentialist gender positions? What if we control for political identification? What if we exclude trans people and vegetarians (who probably care about this for personal reasons)?

11. Can we replicate the claim that people who had a younger sibling born during a supposed critical window for sexual imprinting (I think between 1 and 2 years old, but I

will have to double check) are more exposed to baby-related issues during that window and so are more likely to have baby-related fetishes (lactation and diaper)?

12. Using the same definitions of STEM/nerdy/male vs. interpersonal/creative/female occupations as last time (see Figure 7 here), is there a difference between these groups in the rate at which people perceive gender bias?

13. How does imposter syndrome vary by gender, field, and gender * field?

14. Do schizophrenics (and their families) smoke more? Do autistics (and their families) smoke less?

15. Are schizophrenics (and their families) more likely to be able to tickle themselves than others? What about autistics (and their families)? What about people who have used lots of psychedelics?

16. Is tendency to prefer great literature to sci-fi/fantasy mediated by ability to perceive and appreciate complex emotion? (correlate difference in sci-fi enjoyment – literature enjoyment to autism, family autism. On a hunch, I am also going to correlate this with trustworthiness, which I think can be affected by a sort of paranoia which correlates with high-bandwidth-social-reading, and with ability to tickle self).

17. Do people with ADHD habituate to stimuli more or less quickly? This is a tough one, since there's a common-sense argument for less (being more distractable suggests less able to drown out external stimula). But I have also heard people suggest they habituate more quickly, which is why they get bored so quickly, and why they so easily get distracted from what they're doing. In retrospect, the question I asked (about habituation to a dripping faucet and

other similar noises) is not ideal for this, but I’ll do what I can with it.

18. Do the birth order effects discovered in the last survey disappear if there’s a sufficient age gap between someone and their next youngest sibling, or remain equally strong? If I have enough sample size, I’ll limit this to families of exactly two people, then do an analysis like this one split for people with higher-than-median age gaps vs. people with lower-than-median age gaps. Birth order differences vanishing with high age gaps suggests they may be due to nurture (eg parents too busy parenting another kid to pay attention to you); remaining with high age gaps suggests potentially due to nature (eg maternal antigens or something).

Bay Meetup 1/6 Update

Due to rain, we're switching to holding the meetup indoors at 3045 Shattuck Ave, Berkeley, 94705. There will be several floors of space available for overflow, so hopefully it won't be too crowded. Thanks to Claire, REACH, and Event Horizon for setting this up.

Time is still 3:30 PM on Sunday, 1/6. There's also a Facebook event here.

Book Review: The Structure Of Scientific Revolutions

When I hear scientists talk about Thomas Kuhn, he sounds very reasonable. Scientists have theories that guide their work. Sometimes they run into things their theories can’t explain. Then some genius develops a new theory, and scientists are guided by that one. So the cycle repeats, knowledge gained with every step.

When I hear philosophers talk about Thomas Kuhn, he sounds like a madman. There is no such thing as ground-level truth! Only theory! No objective sense-data! Only theory! No basis for accepting or rejecting any theory over any other! Only theory! No scientists! Only theories, wearing lab coats and fake beards, hoping nobody will notice the charade!

I decided to read Kuhn’s *The Structure Of Scientific Revolutions* in order to understand this better. Having finished, I have come to a conclusion: yup, I can see why this book causes so much confusion.

At first Kuhn’s thesis appears simple, maybe even obvious. I found myself worrying at times that he was knocking down a straw man, although of course we have to read the history of philosophy backwards and remember that Kuhn may already be in the water supply, so to speak. He argues against a simplistic view of science in which it is merely the gradual accumulation of facts. So Aristotle discovered a few true facts, Galileo added a few more on, then Newton discovered a few more, and now we have very many facts indeed.

In this model, good science cannot disagree with other good science. You’re either wrong – as various pseudoscientists and failed scientists have been throughout history, positing false ideas like “the brain is only there to cool the blood” or “the sun orbits the earth”. Or you’re right, your ideas are enshrined in the Sacristry Of Settled Science, and your facts join the accumulated store that passes through the ages.

Simple-version-of-Kuhn says this isn’t true. Science isn’t just facts. It’s *paradigms* – whole ways of looking at the world. Without a paradigm, scientists wouldn’t know what facts to gather, how to collect them, or what to do with them once they had them. With a paradigm, scientists gather and process facts in the ways the paradigm suggests (“normal science”). Eventually, this process runs into a hitch – apparent contradictions, or things that don’t quite fit predictions, or just a giant ugly mess of epicycles. Some genius develops a new paradigm (“paradigm shift” or “scientific revolution”). Then the process begins again. Facts can be accumulated within a paradigm. And many of the facts accumulated in one paradigm can survive, with only slight translation effort, into a new paradigm. But scientific progress is the story of one relatively-successful and genuinely-scientific effort giving way to a different and contradictory relatively-successful and genuinely-scientific effort. It’s the story of scientists constantly tossing out one another’s work and beginning anew.

This gets awkward because paradigms look a lot like facts. The atomic theory – the current paradigm in a lot of chemistry – looks a lot like the fact “everything is made of atoms and molecules”. But this is only the iceberg’s tip. Once you have atomic theory, chemistry starts looking a lot different. Your first question when confronted with an unknown chemical is “what is the molecular structure?” and you have pretty good ideas for how to figure this out. You are not particularly interested in the surface appearance of chemicals, since you know that iron and silver can look alike but are totally different elements; you may be much more interested in the weight ratio at which two chemicals react (which might seem to the uninitiated like a pretty random and silly thing to care about). If confronted with a gas, you might ask things like “which gas is it?” as opposed to thinking all gases are the same thing, or wondering what it would even mean for two gases to be different. You can even think things like “this is a mixture of two different types of gas” without agonizing about how a perfectly uniform substance can be a mixture of anything. If someone asks you “How noble and close to God would say this chemical sample is?” you can tell them that this is not really a legitimate chemical question, unless you mean “noble” in the sense of the noble gases. If someone tells you a certain chemical is toxic because toxicity is a fundamental property of its essence, you can tell them that no, it probably has to do with some reaction it causes or fails to cause with chemicals in the body. And if someone tells you that a certain chemical has changed into a different chemical because it got colder, you can tell them that cold might have done something to it, it might even have caused it to react with the air or something, but chemicals don’t change into other chemicals in a fundamental way just because of the temperature. None of these things are obvious. All of them are hard-won discoveries.

A field without paradigms looks like the STEM supremacist’s stereotype of philosophy. There are all kinds of different schools – Kantians, Aristotelians, Lockeans – who all disagree with each other. There may be progress within a school – some Aristotelian may come up with a really cool new Aristotelian way to look at bioethics, and all the other Aristotelians may agree that it’s great – but the field as a whole does not progress. People will talk past one another;

the Aristotelian can go on all day about the telos of the embryo, but the utilitarian is just going to ask what the hell a telos is, why anyone would think embryos have one, and how many utils the embryo is bringing people. “Debates” between the Aristotelian and the utilitarian may not be literally impossible, but they are going to have to go all the way to first principles, in a way that never works. Kuhn interestingly dismisses these areas as “the fields where people write books” – if you want to say anything, you might as well address it to a popular audience for all the good other people’s pre-existing knowledge will do you, and you may have to spend hundreds of pages explaining your entire system from the ground up. He throws all the social sciences in this bin – you may read Freud, Skinner, and Beck instead of Aristotle, Locke, and Kant, but it’s the same situation.

A real science is one where everyone agrees on a single paradigm. Newtonianism and Einsteinianism are the same kind of things as Aristotelianism and utilitarianism; but in 1850, everybody believed the former, and in 1950, the latter.

I got confused by this – is Aristotelian philosophy a science? Would it be one if the Aristotelians forced every non-Aristotelian philosopher out of the academy, so that 100% of philosophers fell in line behind Aristotle? I *think* Kuhn’s answer to this is that it’s telling that Aristotelians haven’t been able to do this (at least not lately); either Aristotle’s theories are too weak, or philosophy too intractable. But all physicists unite behind Einstein in a way that all philosophers cannot behind Aristotle. Because of this, all physicists mean more or less the same thing when they talk about “space” and “time”, and they can work together on explaining these concepts without constantly arguing to each other about what they mean or whether they’re the right way to think about things at all (and a Newtonian and Einsteinian would *not* be able to do this with each other, any more than an Aristotelian and utilitarian).

So how does science settle on a single paradigm when other fields can't? Is this the part where we admit it's because science has objective truth so you can just settle questions with experiments?

This is very much not that part. Kuhn doesn't think it's anywhere near that simple, for a few reasons.

First, there is rarely a single experiment that one paradigm fails and another passes. Rather, there are dozens of experiments. One paradigm does better on some, the other paradigm does better on others, and everyone argues over which ones should or shouldn't count.

For example, one might try to test the Copernican vs. Ptolemaic worldviews by observing the parallax of the fixed stars over the course of a year. Copernicus predicts it should be visible; Ptolemy predicts it shouldn't be. It isn't, which means either the Earth is fixed and unmoving, or the stars are unutterably unimaginably immensely impossibly far away. Nobody expected the stars to be that far away, so advantage Ptolemy. Meanwhile, the Copernicans posit far-off stars in order to save their paradigm. What looked like a test to select one paradigm or the other has turned into a wedge pushing the two paradigms even further apart.

What looks like a decisive victory to one side may look like random noise to another. Did you know weird technologically advanced artifacts are sometimes found encased in rocks that our current understanding of geology says are millions of years old? Creationists have no trouble explaining those – the rocks are much younger, and the artifacts were probably planted by nephilim. Evolutionists have no idea how to explain those, and default to things like “the artifacts are hoaxes” or “the miners were really careless and a screw slipped from their pocket into the rock vein while they were mining”. I’m an evolutionist and I agree the artifacts are probably hoaxes or mistakes, even when there is no particular evidence that they are. Meanwhile, probably creationists say that some fossil or other incompatible with creationism is a hoax or a mistake. But that means the “find something predicted by one paradigm but not the other, and then the failed theory comes crashing down” oversimplification doesn’t work. Find something predicted by one paradigm but not the other, and often the proponents of the disadvantaged paradigm can – and should – just shrug and say “whatever”.

In 1870, flat-earther Samuel Rowbotham performed a series of experiments to show the Earth could not be a globe. In the most famous, he placed several flags miles apart along a perfectly straight canal. Then he looked through a telescope and was able to see all of them in a row, even though the furthest should have been hidden by the Earth’s curvature. Having done so, he concluded the Earth was flat, and the spherical-earth paradigm debunked. Alfred Wallace (more famous for pre-empting Darwin on evolution) took up the challenge, and showed that the bending of light rays by atmospheric refraction explained Rowbotham’s result. It turns out that light rays curve downward at a rate equal to the curvature of the Earth’s surface! Luckily for Wallace, refraction was already a known phenomenon; if not, it would have been the same kind of wedge-between-paradigms as the Copernicans having to change the distance to the fixed stars.

It is all nice and well to say “Sure, it *looks* like your paradigm is right, but once we adjust for this new idea about the distance to the stars / the refraction of light, the evidence actually supports my paradigm". But the supporters of old paradigms can do that too! The Ptolemaics are rightly mocked for adding epicycle after epicycle until their system gave the right result. But to a hostile observer, positing refraction effects that exactly counterbalance the curvature of the Earth sure looks like adding epicycles. At some point a new paradigm will win out, and its "epicycles" will look like perfectly reasonable adjustments for reality's surprising amount of detail. And the old paradigm will lose, and its "epicycles" will look like obvious kludges to cover up that it never really worked. Before that happens…well, good luck.

Second, two paradigms may not even address or care about the same questions.

Let's go back to utilitarianism vs. Aristotelianism. Many people associate utilitarianism with the trolley problem, which is indeed a good way to think about some of the issues involved. It might be tempting for a utilitarian to think of Aristotelian ethics as having some different answer to the trolley problem. Maybe it does, I don't know. But Aristotle doesn't talk about how he would solve whatever the 4th-century BC equivalent of the trolley problem was. He talks more about "what is the true meaning of justice?" and stuff like that. While you can twist Aristotle into having an opinion on trolleys, he's not really optimizing for that. And while you can make utilitarianism have some idea what the true meaning of justice is, it's not really optimized for that either.

An Aristotelian can say their paradigm is best, because it does a great job explicating all the little types and subtypes of justice. A utilitarian can say *their* paradigm is best, because it does a great job telling you how to act in various contrived moral dilemmas.

It’s actually even worse than this. The closest thing I can think of to an ancient Greek moral dilemma is the story of Antigone. Antigone’s uncle declares that her traitorous dead brother may not be buried with the proper rites. Antigone is torn between her duty to obey her uncle, and her desire to honor her dead brother. Utilitarianism is…not really designed for this sort of moral dilemma. Is ignoring her family squabbles and trying to cure typhus an option? No?

But then utilitarianism’s problems are deeper than just “comes to a different conclusion than ancient Greek morals would have”. The utilitarian’s job isn’t to change the ancient Greek’s mind about the answer to a certain problem. It’s to convince him to stop caring about basically all the problems he cares about, and care about different problems instead.

Third, two paradigms may disagree on what kind of answers are allowed, or what counts as solving a problem.

Kuhn talks about the 17th century “dormitive potency” discourse. Aristotle tended to explain phenomena by appealing to essences; trees grew because it was “in their nature” to grow. Descartes gets a bad rap for inventing dualism, but this is undeserved – what he was really doing was inventing the concept of “matter” as we understand it, a what-you-see-is-what-you-get kind of stuff with no hidden essences, which responds mechanically to forces (and once you have this idea, you naturally need some other kind of stuff to be the mind). With Cartesian matter firmly in place, everyone made fun of Aristotle for thinking he had “solved” the “why do trees grow?” question by answering “because it is in their nature”, and this climaxed with the playwright Moliere portraying a buffoonish doctor who claimed to have discovered how opium put people to sleep – it was because it had a dormitive potency!

In Aristotle’s view of matter, saying “because it’s their essence” successfully answers questions like “why do trees grow?”. The Cartesian paradigm forbade this kind of answer, and so many previously “solved” problems like why trees grow became mysterious again – a step backwards, sort of. For Descartes, you were only allowed to answer questions if you could explain how purely-mechanical matter smashing against other purely-mechanical matter in a billiard-ball-like way could produce an effect; a more virtuous and Descartes-aware doctor explained opium’s properties by saying opium corpuscles must have a sandpaper-like shape that smooths the neurons!

Then Newton discovered gravity and caused an uproar. Gravity posits no corpuscles jostling other corpuscles. It sounds almost Aristotelian: “It is the nature of matter to attract other matter”. Newton was denounced as trying to smuggle occultism into science. How much do you discount a theory for having occult elements? If some conception of quantum theory predicts the data beautifully, but says matter behaves differently depending on whether someone’s watching it or not, is that okay? What if it says that a certain electron has a 50% chance of being in a certain place, full stop, and there is no conceivable explanation for which of the two possibilities is realized, and you’re not even allowed to ask the question? What if my explanation for dark matter is “invisible gremlins”? How do you figure out when you need to relax your assumptions about what counts as science, versus when somebody is just cheating?

A less dramatic example: Lavoisier’s theory of combustion boasts an ability to explain why some substances gain weight when burned; they are absorbing oxygen from the air. A brilliant example of an anomaly explained, which proves the superiority of combustion theory to other paradigms that cannot account for the phenomenon? No – “things shouldn’t randomly gain weight” comes to us as a principle of the chemical revolution of which Lavoisier was a part:

> In the seventeenth century, [an explanation of weight gain] seemed unnecessary to most chemists. If chemical reactions could alter the volume, color, and texture of the ingredients, why should they not alter weight as well? Weight was not always taken to be the measure of quantity of matter. Besides, weight-gain on roasting remained an isolated phenomenon. Most natural bodies (eg wood) lose weight on roasting as the phlogiston theory was later to say they should.

In previous paradigms, weight gain wasn’t even an anomaly to be explained. It was just a perfectly okay thing that might happen. It’s only within the constellation of new methods and rules we learned around Lavoisier’s time, that Lavoisier’s theories solved anything at all.

So how *do* scientists ever switch paradigms?

Kuhn thinks it’s kind of an ugly process. It starts with exasperation; the old paradigm is clearly inadequate. Progress is stagnating.

> Awareness [of the inadequacy of geocentric astronomy] did come. By the thirteenth century Alfonso X could proclaim that if God had consulted him when creating the universe, he would have received good advice. In the sixteenth century, Copernicus’ coworker, Domenico da Novara, held that no system so cumbersome and inaccurate as the Ptolemaic had become could possibly be true of nature. And Copernicus himself wrote in the

> Preface to the De Revolutionibus that the astronomical tradition he inherited had finally created only a monster.

Then someone proposes a new paradigm. In its original form, it is woefully underspecified, bad at matching reality, and only beats the old paradigm in a few test cases. For whatever reason, a few people jump on board. Sometimes the new paradigm is simply more mathematically elegant, more beautiful. Other times it's petty things, like a Frenchman invented the old paradigm and a German the new one, and you're German. Sometimes it's just that there's nothing better. These people gradually expand the new paradigm to cover more and more cases. At some point, the new paradigm explains things a little better than the old paradigm. Some of its predictions are spookily good. The old paradigm is never conclusively debunked. But the new paradigm now has enough advantages that more and more people hop on the bandwagon. Gradually the old paradigm becomes a laughingstock, people forget the context in which it ever made sense, and it is remembered only as a bunch of jokes about dormitive potency.

But now that it's been adopted and expanded and reached the zenith of its power, *this* is the point at which we can admit it's objectively better, right?

For a better treatment of this question than I can give, see Samzdat's Science Cannot Count To Red. But my impression is that Kuhn is not really willing to say this. I think he is of the "all models are wrong, some are useful" camp, thinks of paradigms as models, and would be willing to admit a new paradigm may be more useful than an old one.

Can we separate the fact around which a paradigm is based (like "the Earth orbits the sun") from the paradigm itself (being a collection of definitions of eg "planet" and "orbit", ways of thinking, mathematical methods, and rules for what kind of science will and won't be accepted)? And then say the earth factually orbits the sun, and the paradigm is just a useful tool that shouldn't be judged objectively? I think Kuhn's answer is that facts cannot be paradigm-independent. A medieval would not hear "the Earth orbits the sun" and hear the same claim we hear (albeit, in his view wrong). He would, for example, interpret it to mean the Earth was set in a slowly-turning crystal sphere with the sun at its center. Then he might ask – where does the sphere intersect the Earth? How come we can't see it? Is Marco Polo going to try to travel to China and then hit a huge invisible wall halfway across the Himalayas? And what about gravity? My understanding is the Ptolemaics didn't believe in gravity as we understand it at all. They believed objects had a natural tendency to seek the center of the universe. So if the sun is more central, why isn't everything falling into the sun? To a medieval the statement "the Earth orbits the sun" has a bunch of common-sense disproofs everywhere you look. It's only when attached to the rest of the Copernican paradigm that it starts to make sense.

This impresses me less than it impresses Kuhn. I would say "if you have many false beliefs, then true statements may be confusing in that they seem to imply false statements – but true statements are still objectively true". Perhaps I am misunderstanding Kuhn's argument here; the above is an amalgam of various things and not something Kuhn says outright in the book. But whatever his argument, Kuhn is not really willing to say that there are definite paradigm-independent objective facts, at least not without a lot of caveats.

So where *is* the point at which we admit some things are objectively true and that's what this whole enterprise rests on?

Kuhn only barely touches on this, in the last page of the book:

> Anyone who has followed the argument this far will nevertheless feel the need to ask why the evolutionary process should work. What must nature, including man, be like in order that science be possible at all? Why should scientific communities be able to reach a firm consensus unattainable in other fields? Why should consensus endure across one paradigm change after another? And why should paradigm change invariably produce an instrument more perfect in any sense than those known before? From one point of view those questions, excepting the first, have already been answered. But from another they are as open as they were when this essay began. It is not only the scientific community that must be special. The world of which that community is a part must also possess quite special characteristics, and we are no closer than we were at the start to knowing what these must be. That problem — What must the world be like in order that man may know it?— was not, however, created by this essay. On the contrary, it is as old as science itself, and it remains unanswered. But it need not be answered in this place.

Aaargh! So close!

II.

A lot of the examples above are mine, not Kuhn's. Some of them even come from philosophy or other nonscientific fields. Shouldn't I have used the book's own examples?

Yes. But one of my big complaints about this book is that, for a purported description of How Science Everywhere Is Always Practiced, it really just gives five examples.

Ptolemy/Copernicus on astronomy. Alchemy/Dalton on chemistry. Phlogiston/Lavoisier on combustion. Aristotle/Galileo/Newton/Einstein on motion. And ???/Franklin/Coulomb on electricity.

It doesn’t explain any of the examples. If you don’t already know what Coulomb’s contribution to electricity is and what previous ideas he overturned, you’re out of luck. And don’t try looking it up in a book either. Kuhn says that all the books have been written by people so engrossed in the current paradigm that they unconsciously jam past scientists into it, removing all evidence of paradigm shift. This made parts of the book a little beyond my level, since my knowledge of Coulomb begins and ends with “one amp times one second”.

Even saying Kuhn has five examples is giving him too much credit. He usually brings in one of his five per point he’s trying to make, meaning that you never get a really full view of how any of the five examples exactly fit into his system.

And all five examples are from physics. Kuhn says at the beginning that he wished he had time to talk about how his system fits biology, but he doesn’t. He’s unsure whether any of the social sciences are sciences at all, and nothing else even gets mentioned. This means we have to figure out how Kuhn’s theory fits everything from scattershot looks at the history of electricity and astronomy and a few other things. This is pretty hard. For example, consider three scientific papers I’ve looked at on this blog recently:

– Cipriani, Ioannidis, et al perform a meta-analysis of antidepressant effect sizes and find that although almost all of them seem to work, amitriptyline works best.

– Ceballos, Ehrlich, et al calculate whether more species have become extinct recently than would be expected based on historical background rates; after finding almost 500 extinctions since 1900, they conclude they definitely have.

– Terrell et al examine contributions to open source projects and find that men are more likely to be accepted than women when adjusted for some measure of competence they believe is appropriate, suggesting a gender bias.

What paradigm is each of these working from?

You could argue that the antidepressant study is working off of the "biological psychiatry" paradigm, a venerable collection of assumptions that can be profitably contrasted with other paradigms like psychoanalysis. But couldn't a Hippocratic four-humors physician of a thousand years ago done the same thing? A meta-analysis of the effect sizes of various kinds of leeches for depression? Sure, leeches are different from antidepressants, but it doesn't look like the belief in biological psychiatry is affecting anything about the research other than the topic. And although the topic is certainly important, Kuhn led me to expect something more profound than that. Maybe the paradigm is evidence-based-medicine itself, the practice of doing RCTs and meta-analyses on things? I think this is a stronger case, but a paradigm completely divorced from the content of what it's studying is exactly the sort of weird thing that makes me wish Kuhn had included more than five examples.

As for the extinction paper, surely it can be attributed to some chain of thought starting with Cuvier's catastrophism, passing through Lyell, and continuing on to the current day, based on the idea that the world has changed dramatically over its history and new species can arise and old ones disappear. But is that “the” paradigm of biology, or ecology, or whatever field Ceballos and Lyell are working in? Doesn’t it also depend on the idea of species, a different paradigm starting with Linnaeus and developed by zoologists over the ensuing centuries? It look like it dips into a bunch of different paradigms, but is not wholly within any.

And the open source paper? Is “feminism” a paradigm? But surely this is no different than what would be done to investigate racist biases in open source. Or some right-winger looking for anti-Christian biases in open source. Is the paradigm just “looking for biases in things?”

What about my favorite trivial example, looking both ways when you cross the street so you don’t get hit by a bus? Is it based on a paradigm of motorized transportation? Does it use assumptions like “buses exist” and “roads are there to be crossed”? Was there a paradigm shift between the bad old days of looking one way before crossing, and the exciting new development of looking both ways before crossing? Is this really that much more of a stretch than calling looking for biases in things a paradigm?

Outside the five examples Kuhn gives from the physical sciences, identifying paradigms seems pretty hard – or maybe too easy. Is it all fractal? Are there overarching paradigms like atomic theory, and then lower-level paradigms like organic chemistry, and then tiny subsubparadigms like “how we deal with this one organic compound”? Does every scientific experiment use lots of different paradigms from different traditions and different levels? This is the kind of thing I wish Kuhn’s book answered instead of just talking about Coulumb and Copernicus over and over again.

III.

In conclusion, all of this is about predictive coding.

It’s the same thing. Perception getting guided equally by top-down expectations and bottom-up evidence. Oh, I know what you’re thinking. “There goes Scott again, seeing predictive coding in everything”. And yes. But also, Kuhn does everything short of come out and say “When you guys get around to inventing predictive coding, make sure to notice that’s what I was getting at this whole time.”

Don’t believe me? From the chapter *Anomaly And The Emergence Of Scientific Discovery* (my emphasis, and for “anomaly”, read “surprisal”):

> The characteristics common to the three examples above are characteristic of all discoveries from which new sorts of phenomena emerge. Those characteristics include: the previous awareness of anomaly, the gradual and simultaneous emergence of both observational and conceptual recognition, and the consequent change of paradigm categories and procedures often accompanied by resistance.
>
> **There is even evidence that these same characteristics are built into the nature of the perceptual process itself.** In a psychological experiment that deserves to be far better known outside the trade, Bruner and Postman asked experimental subjects to identify on short and controlled exposure a series of playing cards. Many of the cards were normal, but some were made anomalous, e.g., a red six of spades and a black four of hearts. Each experimental run was constituted by the display of a single card to a single subject in a series of gradually increased exposures. After each exposure the subject was asked

what he had seen, and the run was terminated by two successive correct identifications.

Even on the shortest exposures many subjects identified most of the cards, and after a small increase all the subjects identified them all. For the normal cards these identifications were usually correct, but the anomalous cards were almost always identified, without apparent hesitation or puzzlement, as normal. The black four of hearts might, for example, be identified as the four of either spades or hearts. Without any awareness of trouble, it was immediately fitted to one of the conceptual categories prepared by prior experience. One would not even like to say that the subjects had seen something different from what they identified. With a further increase of exposure to the anomalous cards, subjects did begin to hesitate and to display awareness of anomaly. Exposed, for example, to the red six of spades, some would say: That's the six of spades, but there's something wrong with it— the black has a red border. Further increase of exposure resulted in still more hesitation and confusion until finally, and sometimes quite suddenly, most subjects would produce the correct identification without hesitation. Moreover, after doing this with two or three of the anomalous cards, they would have little further difficulty with the others. A few subjects, however, were never able to make the requisite adjustment of their categories. Even at forty times the average exposure required to recognize normal cards for what they were, more than 10 per cent of the anomalous cards were not correctly identified. And the subjects who then failed often experienced acute personal distress. One of them exclaimed: "I can't make the suit out, whatever it is. It didn't even look like a card that time. I don't know what color it is now or whether it's a spade or a heart. I'm not

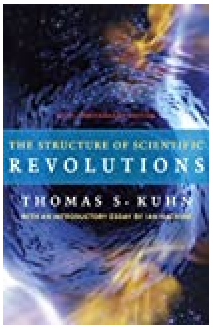
> even sure now what a spade looks like. My God!” In the next section we shall occasionally see scientists behaving this way too.
>
> Either as a metaphor or **because it reflects the nature of the mind**, that psychological experiment provides a wonderfully simple and cogent schema for the process of scientific discovery.

And from *Revolutions As Changes Of World-View*:

> **Surveying the rich experimental literature from which these examples are drawn makes one suspect that something like a paradigm is prerequisite to perception itself. What a man sees depends both upon what he looks at and also upon what his previous visual-conceptual experience has taught him to see. In the absence of such training there can only be, in William James’s phrase, “a bloomin’ buzzin’ confusion.” In recent years several of those concerned with the history of science have found the sorts of experiments described above immensely suggestive.**

If you can read those paragraphs and honestly still think I’m just just irrationally reading predictive coding into a perfectly innocent book, I have nothing to say to you.

I think this is my best answer to the whole “is Kuhn denying an objective reality” issue. If Kuhn and the predictive coding people are grasping at the same thing from different angles, then both shed some light on each other. I think I understand the way that predictive coding balances the importance of pre-existing structures and categories with a preserved belief in objectivity. If Kuhn is trying to extend the predictive coding model of the brain processing information to the way the scientific community as a whole processes it, then maybe we can import the same balance and not worry about it as much.


THE STRUCTURE OF SCIENTIFIC
REVOLUTIONS
THOMAS S. KUHN

Paradigms All The Way Down

Related to: Book Review: The Structure Of Scientific Revolutions

Every good conspiracy theorist needs their own Grand Unified Chart; I'm a particular fan of this one. So far, my own Grand Unified Chart looks like this:

Philosophy Of Science	Paradigm	Anomaly	Data
Bayesian Probability	Prior	KL-divergence	Evidence
Psychology	Prediction	Surprisal	Sense-Data
Discourse	Ideology	Cognitive Dissonance	Facts
Society	Frames & Narratives	Exclusion	Lived Experience
Neuroscience	NMDA	Dopamine	AMPA

All of these are examples of interpreting the world through a combination of pre-existing ideas what the world should be like (first column), plus actually experiencing the world (last column). In all of them, the world is too confusing and permits too many different interpretations to understand directly. You wouldn't even know where to start gathering more knowledge. So you take all of your pre-existing ideas (which you've gotten from *somewhere*) and interpret everything as behaving the way your pre-existing ideas tell you they will. Then as you gradually gather discrepancies between what you expected and what you get (middle column), you gradually become more and more confused until your existing categories buckle under the strain and you generate a new and self-consistent set of pre-existing ideas to see the world through, and then the process begins again.

All of these domains share an idea that the interaction between facts and theories is bidirectional. Your facts may eventually determine what theory you have. But your theory also determines what facts you see and notice. Nor do contradictory facts immediately change a theory. The process of theory change is complicated, fiercely resisted by hard-to-describe factors, and based on some sort of idea of global tension that can't be directly reduced to any specific contradiction.

(I linked the Discourse and Society levels of the chart to this post where I jokingly sum up the process of convincing someone as "First they ignore you, then they laugh at you, then they fight you, then they fight you half-heartedly, then they're neutral, then they grudgingly say you might have a point even though you're annoying, then they say on balance you're mostly right although you ignore some of the most important facets of the issue, then you win." My point is that ideological change – most dramatically religious conversion, but also Republicans becoming Democrats and vice versa – doesn't look like you "debunking" one of their facts and them admitting you are right. It is less like Popperian falsification and more like a Kuhnian paradigm shift or a Yudkowskian crisis of faith.)

Why do all of these areas share this same structure? I think because it's built into basic algorithms that the brain uses for almost everything (see the Psychology and Neuroscience links above). And that in turn is because it's just factually the most effective way to do epistemology, a little like asking "why does so much cryptography use prime numbers".

SSC Survey Results 2019

Thanks to the 8,171 people who took the 2019 Slate Star Codex survey. Some of the links below will say 13,171 people took the survey, but that's a bug – sometimes Google just adds 5,000 to things. You can:

– **See the questions for the SSC survey.**

– **See the results from the SSC survey.**

I'll be publishing more complicated analyses over the course of the next year, hopefully starting later this week.

If you want to scoop me, or investigate the data yourself, you can download the answers of the 7000 people who agreed to have their responses shared publicly. The public datasets will not exactly match the full version, nor will they include some of the sensitive sections like illegal drug use and sexual partners.

Download the public data (.xlsx, .ods)

Too Many People Dare Call It Conspiracy

[Content warning: References to anti-Semitic and anti-Catholic canards]

I feel deep affection for Gary Allen’s *None Dare Call It Conspiracy*, a bizarre screed about the Federal Reserve/Communist/Trilateral Commission plot for a one world government. From its ridiculous title to its even-more-ridiculous cover image, this is a book that accepts its own nature. In the Aristotelian framework, where everything is trying to be the most perfect example of whatever it is, *None Dare Call It Conspiracy* has reached a certain apotheosis.

But my problem is the opposite of Allen’s. Too many people dare call too many things conspiracy. Perfectly reasonable hypotheses get attacked as conspiracy theories, derailing the discussion into arguments over when you’re allowed to use the phrase. These arguments are surprisingly tough. Which of the following do you think should be classified as “conspiracy theories”? Which ones are so deranged that people espousing them should be excluded from civilized discussion?

1. Donald Trump and his advisors secretly met with Russian agents to discuss how to throw the 2016 election in his favor.

2. Donald Trump didn’t collaborate with any Russians, but Democrats are working together to convince everyone that he did, in the hopes of getting him indicted or convincing the electorate that he’s a traitor.

3. Insurance companies are working to sabotage any proposal for universal health care; if not for their constant machinations, we would have universal health care already.

4. The ruling classes constantly use lobbyists and soft power to sabotage tax increases, labor laws, and any other policy that increase the relative power of the poor.

5. America’s aid to Israel is not in America’s best interest, but is maintained through the power of AIPAC and other pro-Israel groups mainly supported by America’s Jewish community.

6. The Jews are behind Brexit as a plot to weaken Western Europe.

7. Climate scientists routinely exaggerate or massage their studies to get the results they want, or only publish studies that get the results they want, both because of their personal political leanings and because they know it is good for their field to constantly be discovering exciting things that their funders and their supporters among the public want to hear.

8. As above, except with replace climate science with “race science”, with “power posing“, with “the side effects of some drug that earns a pharma company a lot of money”, or your own favorite example.

9. When European trains get bombed, with leaflets distributed near the scene repeating jihadist propaganda, it’s actually a false flag by a rightist trying to discredit Islam.

10. When several prominent Trump critics receive bombs in the mail, it’s actually a “false flag operation” by a leftist trying to discredit Trumpism.

11. Bernie Sanders’ whole campaign is a “false flag operation” by capitalists who are trying to prevent other socialists from entering the race; if Sanders ever shows any signs of winning, he will withdraw under mysterious circumstances.

12. The entire Democratic Socialist movement in America is a “false flag operation” by the CIA, intended to create a wishy-washy Americanized form of socialism that sucks the oxygen away from more aggressive Soviet-style Marxism.

13. The CIA has fixed elections in dozens of foreign countries over the past seventy years or so.

14. The CIA is plotting to fix the 2020 US elections.

15. The Catholic Church spent decades covering up the extent of sexual abuse by its priests.

16. A UFO cult has taken over the government and is using it as a base through which to carry out the designs of its extraterrestrial masters.

17. The patriarchy privileges men over women in a variety of ways, excludes women from positions of influence, and suppresses their efforts to win equality.

18. The Bilderberg Meeting secretly plots ways to create a one-world government.

II.

The Basic Argument Against Conspiracy Theories goes: “You can’t run a big organization in secret without any outsiders noticing or any insiders blowing the whistle.” If we keep this in mind, I think we can resolve some of the awkward tensions above.

For example, the CIA definitely has fixed elections in foreign countries. Is this a conspiracy theory? No. The CIA is not secret. Everyone knows the CIA exists and does nefarious things, even if we don't know exactly *which* nefarious things it does. There is no need to keep the CIA secret. It can advertise in public "Wanted: people who are good at doing nefarious things". And if somebody whistleblows, they will not receive the thanks of a grateful country. They'll probably just be arrested for leaking classified information, while everybody snoozes. "CIA discovered to have fixed Gabonese elections" is probably a page 5 story at best.

I think "The CIA is plotting to fix the 2020 US elections" *is* a conspiracy theory, with all the unlikeliness that implies. Although the CIA exists openly, fixing US elections would take a powerful conspiracy *within* the CIA. You would have to hide it from the idealistic young recruits who come in hoping to make the world safe for democracy. You would have to convince all the other CIA agents to hide it from Congress, from the other intelligence services, and from any CIA agent who wasn't on board. And a whistleblower really *would* receive the thanks of a grateful country. Although the CIA gets the advantage of existing publicly, the intra-CIA conspiracy to fix elections doesn't, and so the Basic Argument strikes it down.

(The CIA does work on lots of things the public wouldn't approve of, like MKULTRA. But the bigger and more controversial they are, the more likely they are to get leaked, which I think supports this theory. At some point the CIA recruits start saying "This isn't what we signed up for", and then the usual conspiracy dynamics apply.)

During the 1960s, the CIA sponsored various socialist magazines and organizations with exactly this justification – better direct the sort of people who would be socialist anyway to moderate socialism instead of more violent or Soviet-aligned groups. So why dismiss that they're behind the modern Democratic Socialists, or Bernie Sanders? As far as I can tell, no reason except the end of the Cold War decreasing their motives, plus it seems like too big a deal to pull off secretly.

Keeping the Basic Argument in mind also helps understand Jews supporting Israel, insurance companies opposing universal health care, scientists sticking to various flawed paradigms, the patriarchy suppressing women, and elites controlling the government. None of these are conspiracy theories, because they're all obviously in the self-interest of the group involved, so each member can individually decide to do it. That removes the need for the secret coordinating organization, which is the part it's hard to hide. This means we can dismiss "the Jews caused Brexit" as legitimately a conspiracy theory; if there's some good reason for Jews to cause Brexit, it's not obvious to anybody (including the Jews), so you would need the secret centralized organization to convince and coordinate everybody.

This isn't to say no coordination happens. I expect a little coordination happens openly, through prosocial slogans, just to overcome free rider problems. Remember Trivers' theory of self-deception – that if something is advantageous to us, we naturally and unconsciously make up explanations for why it's a good prosocial policy, and then genuinely believe those explanations. If you are rich and want to oppress the poor, you can come up with some philosophy of trickle-down or whatever that makes it sound good. Then you can talk about it with other rich people openly, no secret organizations in smoke-filled rooms necessary, and set up think tanks together. If you're in the patriarchy, you can push nice-sounding things about gender roles and family values. There is no secret layer beneath the public layer –

no smoke-filled room where the rich people get together and say “Let’s push prosocial slogans about rising tides, so that secretly we can dominate everything”. It all happens naturally under the hood, and the Basic Argument isn’t violated.

I think Trump probably met with the Russians. But even if he didn’t, I don’t think that positing “the Democrats are working hard to make the case that he did” qualifies as “conspiracy theory”. People are tempted to genuinely believe whatever puts them on top; that means Democrats probably genuinely believe Trump is guilty. Once they all genuinely believe it, they can talk openly – “How do we help coordinate to reveal the truth to everyone and bring this traitor to justice?” – rather than violating the Basic Argument by meeting secretly to figure out how to best delude the American people. Likewise, I believe climate change is real, but if it isn’t, the way scientists went wrong looks more like this than like a smoke-filled room.

We may have to bring in all of these (and more) to explain why the Catholic Church covering up sex scandals isn’t the kind of conspiracy theory we should automatically reject (or should have automatically rejected before the evidence came in). The Church is a public-facing organization that is known to occasionally keep secrets (like the CIA), but covering up sex scandals seems as far from their stated mission as the CIA fixing US elections. I think we just have to appeal to the Church hierarchy having a culture where this seemed like the obvious thing to do, as natural as insurance companies opposing universal health care. On the other hand, that could be used to justify anything. After all, the Bilderberg Group is known to exist, and maybe it has a culture where plotting a one world government sounds reasonable from the inside. I don’t know what principle rules in the Catholic case but keeps the Bilderberg case out.

Maybe we just have to accept that even the most explosive conspiracy theories are sometimes true, and the Church's sex scandals are one of those times.

As far as I know no UFO cult has ever taken over the federal government. But Scientology did take over the government of Clearwater, Florida. I think this reinforces some of the points above. Scientology is known to exist and known to do nefarious things. Taking over a town government…actually isn't too far away from what the average member of the public expects them to do. If everyone knows you exist, and everyone knows you're bad, you're not a conspiracy, any more than the Nazis were a conspiracy during World War II (and they, too, sometimes secretly manipulated things they weren't overtly in control of). I think "UFO cult takes over the government" sounds conspiracy-ish only because we read in an implied "…and nobody has heard of this cult or considers it very powerful".

The train bombing false flag story is true. So why would it have been a conspiracy theory to speculate that the anti-Trump bombs were sent by a leftist? A technical objection: it shouldn't count as a conspiracy theory because only one person was involved. A more serious take: it's not *impossible* that these are false flags, but your prior should be pretty low. Most terrorist bombings by people spouting jihadi propaganda *are* by Muslims; most letter-bombing of leftists *is* done by rightists. To jump right away to calling these false flag may not be a "conspiracy theory" in the technical sense, but it's doing the very conspiracy-theory-ish thing of replacing a simple and direct picture of the world with a more complicated one without having enough evidence to justify such a move. I'm reluctant to say that too strongly, because there have been a few false flags that I called (correctly) before the evidence came in – for example, a few years ago 4Channers pretending to be feminists started a campaign to #EndFathersDay, and I wasn’t fooled. I’m not sure I can verbalize how I figured this out – feminists often do controversial and outrageous things that are not false flags – but sometimes about this one just seemed off. I realize that by giving myself permission to say this I risk everyone else saying “Something about this bombing seems off to me, so I conclude it’s a false flag!” So it goes.

III.

There’s a story about Winston Churchill bothering a certain high society lady. Churchill asked if she would sleep with him for five million pounds; she said such an offer would be hard to resist. Then he asked if she would sleep with him for five pounds; she asked “What kind of a person do you think I am?” Churchill answered “We’ve already established what kind of a person you are; now we’re just haggling over the price.”

I think the above examples prove that this is not the right way to think of conspiracy theories. Imagine Winston Churchill asking you whether a UFO cult secretly controls the government of Clearwater. You say yes. Then he asks if a UFO cult secretly controls the US federal government. “What kind of crazy conspiracy theorist do you think I am?” “We’ve already established there’s a conspiracy, now we’re just haggling over the size”.

Instead, the Basic Argument Against Conspiracy Theories gives some heuristics for when conspiracies might be more or less plausible. The typical Illuminati-style theory violates all of them; other theories that only violate a few might still be true. Some of these heuristics might be things like:

A. You generally can’t keep the existence of a large organization that engages in clandestine activities secret. If you have an overt large organization that engages in clandestine activities, and everybody knows about it, they can sometimes accomplish conspiracies compatible with their public-facing mission statement (like the CIA destabilizing enemies of America) but are unlikely to accomplish conspiracies very far outside the range of that statement (like the CIA destabilizing America itself).

B. When a group has an obvious interest in an outcome, its members can coordinate upon that outcome without their being any conspiracy. For example, Jews like Israel for reasons that don’t come as a surprise to anybody, so it’s not a conspiracy theory to posit that Jews are involved in supporting Israel; each Jew can make that decision individually for personal reasons. But if Jews wanted a one-world government, that would be surprising and require some secret effort to convince them; claiming that Jews are working for a one-world government is a conspiracy theory. Likewise, it’s unsurprising that the rich don’t like policies that lower their relative standing, so we can figure rich people are influencing the government towards pro-rich and anti-poor policies in some way without it being a conspiracy theory.

C. When a group is able to form an internal culture in which their nefarious goals seem reasonable and prosocial, they can coordinate upon them in ways that might look like a conspiracy to outsiders. For example, rich people say that taxing the rich would punish innovation and reduce dynamism, and probably actually believe this. This lets them coordinate think tanks to lower taxes on the rich without needing smoke-filled underground lairs where they meet and plot against the poor. Likewise, social scientists all liked “power posing” studies because they were exciting, reinforced the standard social science paradigm, and offered a way to reduce gender bias. So for a while lots of studies came out showing power posing was true, and the studies showing it was false never got published, without anyone having to meet in an underground lair and figure out ways to manipulate the science; probably every social scientist who signal-boosted one study and not another believed they were just making the truth slightly more apparent and making the world a better place.

D. All else being equal, small conspiracies are likelier than big conspiracies. A cult may take over a town without the average person knowing it; it would be more surprising for them to take over a country.

E. There is no royal road. Sometimes you can just plead “intuition”, and you’ll be right.

Kernel Of Doubt: Testing Math Preference Vs. Corn-Eating Style

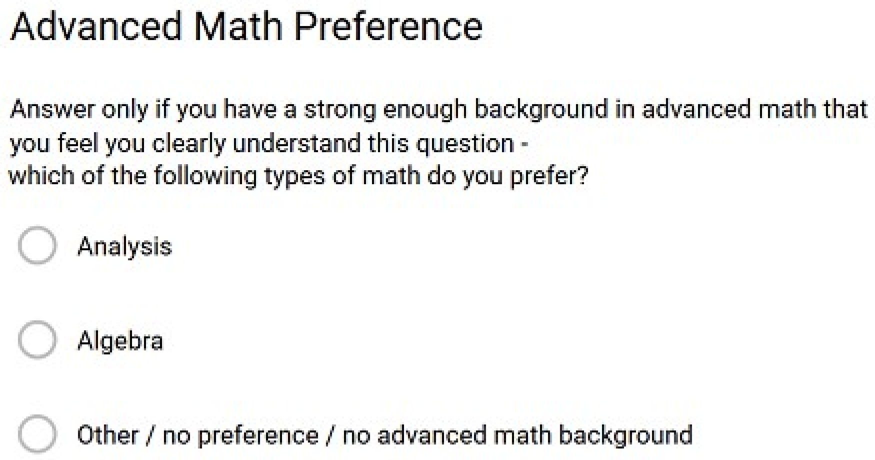
In 2010, Ben Tilly of the blog Random Observations wrote Analysis Vs. Algebra Predicts Eating Corn?, which said:

> I like learning about odd connections between disparate things. This probably is the oddest example that I know.
>
> Broadly speaking, mathematicians can be divided into those who like analysis, and those who like algebra. The distinction between the two types runs throughout math. Even those who work in areas that are far from analysis or algebra are very aware of the difference between them, and usually are very clear on which their preference is. I'll delve into this in more depth soon, but for now let's just take it for granted that this is a well-known distinction, and it has meaning for mathematicians.
>
> Back when I was in grad school there was a department lunch with corn on the cob. Partway through the meal one of the analysts looked around the room and remarked, "That's odd, all of the analysts are eating corn one way and the algebraists are eating corn another!" Everyone looked around. In fact everyone was eating the corn in one of two ways. One way was to munch over the length of the corn in a straight line, back up, turn slightly, and do another row across. Kind of like how an old typewriter goes. The other way was to go around in a spiral. All of the analysts were eating in spirals, and the algebraists in rows.
>
> There were a number of mathematicians present whose fields of study didn't make it clear whether they were on

> the analysis or algebra side of things. We went around and asked, and in every case the way they ate corn matched their preference. Since then I’ve made a point of amusing myself by asking mathematicians I meet whether they prefer algebra or analysis, and then predicting which way they will eat corn. I’m probably up to 40 or so by now, and in every case but one I’ve been able to correctly predict how they eat corn. The one exception was a logician who claimed to be exactly on the fence between the two. When I explained the corn thing to him he looked surprised, and said that he had an unusual way of eating corn. He went in loose spirals! In other words he truly was a perfect combination of algebra and analysis!
>
> If you have even a passing familiarity of probability, it is clear that despite how unbelievable it initially is that the type of mathematics you prefer is connected to how you eat corn, it is pretty much certain that there actually is a very strong connection. If you believe, as I do, that this difference is connected to how we think about other things, then there must be some odd connection between how we like to understand the world and how we eat corn.

The post later went viral on Hacker News, r/math, and Twitter; it was even the subject of a keynote speech at a math conference.

I couldn’t find any record of it being formally tested, so I included two relevant questions in the 2019 Slate Star Codex reader survey:



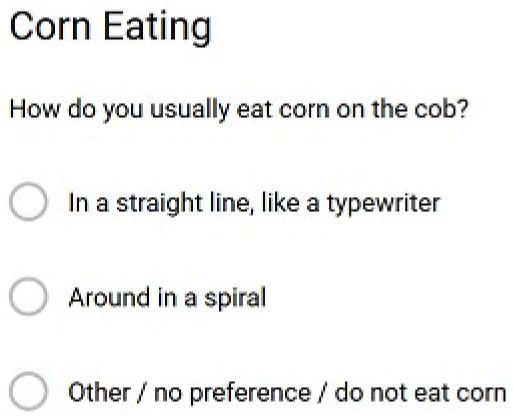
Advanced Math Preference

Answer only if you have a strong enough background in advanced math that you feel you clearly understand this question -
which of the following types of math do you prefer?

○ Analysis

○ Algebra

○ Other / no preference / no advanced math background



Corn Eating

How do you usually eat corn on the cob?

○ In a straight line, like a typewriter

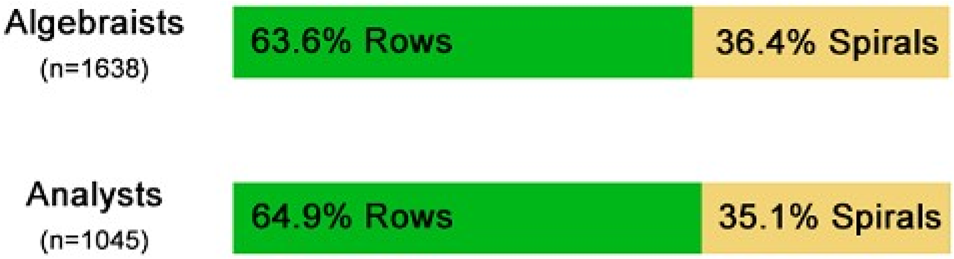
○ Around in a spiral

○ Other / no preference / do not eat corn

These were separated by three unrelated questions, so that most participants would not realize that they were meant to be connected. An informal survey of participants suggested that although some of them had read the Tilly article and realized that the survey was testing it, most did not.

8,171 people answered the survey, of whom 2,683 expressed both a math preference and a corn preference. Here’s what I found:

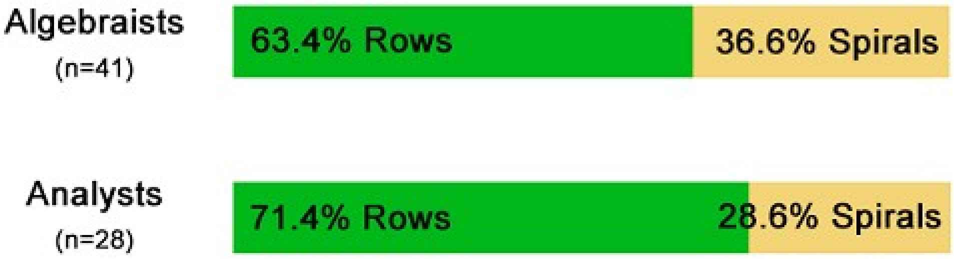
Figure 1: Corn-Eating Style By Math Preference



A chi-square test confirmed that there was no significant difference between the two groups.

Is it possible that participants were not mathematically advanced enough to understand the question? Beginner math students could have interpreted “algebra” to mean the sort of high school algebra where you solve for x if x+1=5, and “analysis” to mean analyzing difficult problems. In order to prevent this, I ran the test again, limiting it to people with PhDs in Math (Degree = “Ph D.” & Profession = “Mathematics”). Here are the results:

Figure 2: Corn-Eating Style By Math Preference, Math PhDs only



Again, chi-square test confirmed there was no significant difference between the two groups (and notice also that the non-significant trend is in the opposite of the predicted direction).

Why might my results be so different from Tilly's? I found a discussion of this question on Quora, where Daniel McLaury gives an answer that rings true to me:

> Now, for the only data point I have: I am decidedly an algebraist and not an analyst, and as best as I can recall I might eat corn in either of these ways, or in a combination of both ways, on any given day. I would imagine the same is true of most people.
>
> As to why he gets such consistent answers, I'll note that when I read the assertion "algebraists do this, analysts do that," I thought, "Oh, yes, I do eat corn that way." After thinking a bit more carefully, making the interpretation I describe in the second paragraph, and then envisioning eating corn each way, it became clear to me that I've eaten corn both ways. So I imagine this whole thing is some combination of the power of suggestion and, perhaps, a selective memory on the part of the author.

But do some people genuinely eat corn more one way or the other? If so, what determines this? I analyzed this question by gender, race, ethnicity, subethnicity, religion, sexual orientation, social class, neurotype, and political affiliation (I love this dataset so much). I'm not going to do this rigorously because there are too many comparisons and it's not worth it, but just eyeballing things it looks like eating-in-rows came to America with the earliest English colonists, and eating-in-spirals is more common among more recent immigrant groups, especially Hindu Indians. I don't have great data for most countries, but the few European countries where I have a decent sample size seem somewhere in between.

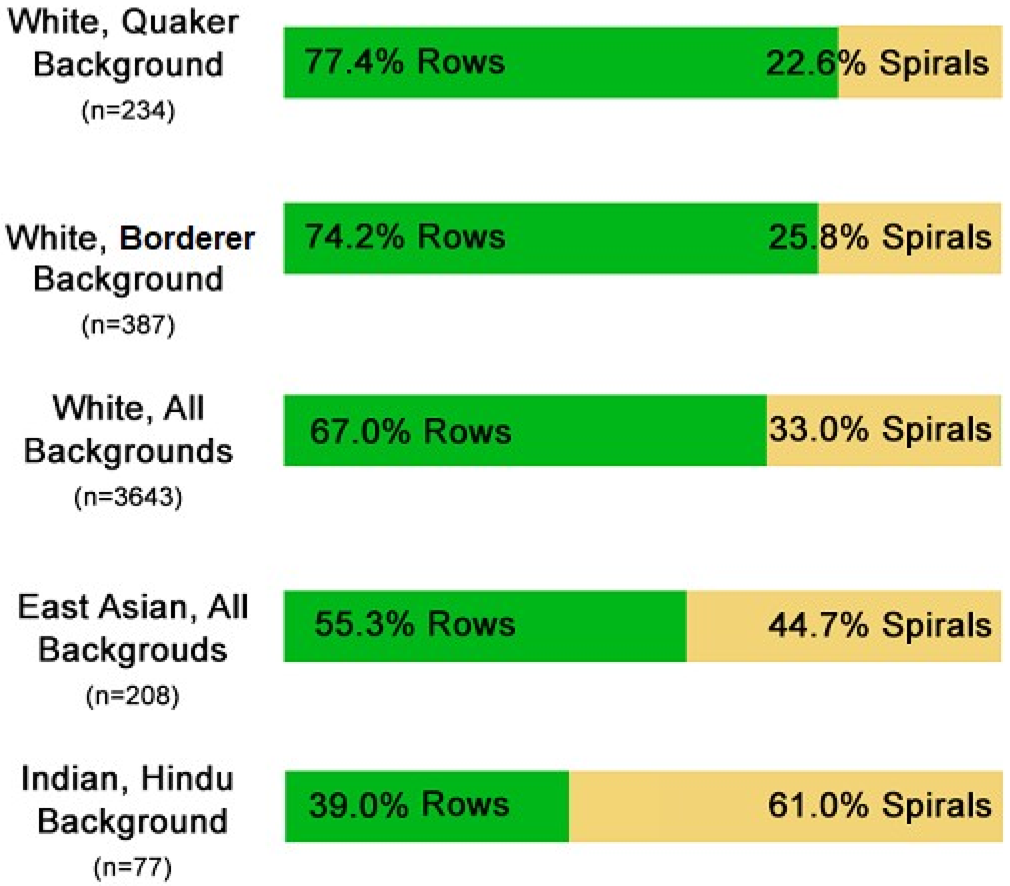


Figure 3: Corn-Eating Style For Selected Racial And Ethnic Backgrounds

I would make a joke about Western linear thinking vs. Eastern cyclic thinking, but I’m worried someone would take me seriously. I have no good explanation for why these groups eat corn the way they do, or why there’s so much variability even within them.

If you want to confirm or expand these results, you can download the original dataset at the bottom of this post.

[**EDIT**: A commenter brings up this previous investigation, which also finds no effect.]

Highlights From The Comments On Kuhn

Thanks to everyone who commented on the review of *The Structure Of Scientific Revolutions*.

From David Chapman:

> It's important to remember that Kuhn wrote this seven decades ago. It was one of the most influential books of pop philosophy in the 1960s-70s, influencing the counterculture of the time, so it is very much "in the water supply." Much of what's right in it is now obvious; what's wrong is salient. To make sense of the book, you have to understand the state of the philosophy of science before then (logical positivism had just conclusively failed), and since then (there has been a lot of progress since Kuhn, sorting out what he got right and wrong).
>
> The issue of his relativism and attitude to objectivity has been endlessly rehashed. The discussion hasn't been very productive; it turns out that what "objective" means is more subtle than you'd think, and it's hard to sort out exactly what Kuhn thought. (And it hasn't mattered what he thought, for a long time.)
>
> Kuhn's "Postscript" to the second edition of the book does address this. It's not super clear, but it's much clearer than the book itself, and if anyone wants to read the book, I would strongly recommend reading the Postscript as well. Given Scott's excellent summary, in fact I would suggest *starting* with the Postscript.

> The point that Kuhn keeps re-using a handful of atypical examples is an important one (which has been made by many historians and philosophers of science since). In fact, the whole “revolutionary paradigm shift” paradigm seems quite rare outside the examples he cites. And, overall, most sciences work quite differently from fundamental physics. The major advance in meta-science from about 1980 to 2000, imo, was realizing that molecular biology, e.g., works so differently from fundamental physics that trying to subsume both under one theory of science is infeasible.

I’m interested to hear him say more about that last sentence if he wants.

Kaj_Sotala quotes Steven Horst quoting Thomas Kuhn on what he means by facts not existing independently of paradigms:

> [Kuhn wrote that]:
>
> A historian reading an out-of-date scientific text characteristically encounters passages that make no sense. That is an experience I have had repeatedly whether my subject is an Aristotle, a Newton, a Volta, a Bohr, or a Planck. It has been standard to ignore such passages or to dismiss them as products of error, ignorance, or superstition, and that response is occasionally appropriate. More often, however, sympathetic contemplation of the troublesome passages suggests a different diagnosis. The apparent textual anomalies are artifacts, products of misreading.
>
> For lack of an alternative, the historian has been understanding words and phrases in the text as he or she would if they had occurred in contemporary

discourse. Through much of the text that way of reading proceeds without difficulty; most terms in the historian’s vocabulary are still used as they were by the author of the text. But some sets of interrelated terms are not, and it is [the] failure to isolate those terms and to discover how they were used that has permitted the passages in question to seem anomalous. Apparent anomaly is thus ordinarily evidence of the need for local adjustment of the lexicon, and it often provides clues to the nature of that adjustment as well. An important clue to problems in reading Aristotle’s physics is provided by the discovery that the term translated ‘motion’ in his text refers not simply to change of position but to all changes characterized by two end points. Similar difficulties in reading Planck’s early papers begin to dissolve with the discovery that, for Planck before 1907, ‘the energy element hv’ referred, not to a physically indivisible atom of energy (later to be called ‘the energy quantum’) but to a mental subdivision of the energy continuum, any point on which could be physically occupied.

These examples all turn out to involve more than mere changes in the use of terms, thus illustrating what I had in mind years ago when speaking of the “incommensurability” of successive scientific theories. In its original mathematical use ‘incommensurability’ meant “no common measure,” for example of the hypotenuse and side of an isosceles right triangle. Applied to a pair of theories in the same historical line, the term meant that there was no common language into which both could be fully translated. (Kuhn 1989/2000, 9–10)

While scientific theories employ terms used more generally in ordinary language, and the same term may appear in multiple theories, key theoretical terminology is proprietary to the theory and cannot be understood apart from it. To learn a new theory, one must master the terminology as a whole: “Many of the referring terms of at least scientific languages cannot be acquired or defined one at a time but must instead be learned in clusters” (Kuhn 1983/2000, 211). And as the meanings of the terms and the connections between them differ from theory to theory, a statement from one theory may literally be nonsensical in the framework of another. The Newtonian notions of absolute space and of mass that is independent of velocity, for example, are nonsensical within the context of relativistic mechanics. The different theoretical vocabularies are also tied to different theoretical taxonomies of objects. Ptolemy’s theory classified the sun as a planet, defined as something that orbits the Earth, whereas Copernicus’s theory classified the sun as a star and planets as things that orbit stars, hence making the Earth a planet. Moreover, not only does the classificatory vocabulary of a theory come as an ensemble—with different elements in nonoverlapping contrast classes—but it is also interdefined with the laws of the theory. The tight constitutive interconnections within scientific theories between terms and other terms, and between terms and laws, have the important consequence that any change in terms or laws ramifies to constitute changes in meanings of terms and the law or laws involved with the theory (though, in significant contrast with Quinean holism, it need not ramify to constitute changes in meaning, belief, or inferential commitments outside the boundaries of the theory).

While Kuhn’s initial interest was in revolutionary changes in theories about what is in a broader sense a single phenomenon (e.g., changes in theories of

> gravitation, thermodynamics, or astronomy), he later came to realize that similar considerations could be applied to differences in uses of theoretical terms between contemporary subdisciplines in a science (1983/2000, 238). And while he continued to favor a linguistic analogy for talking about conceptual change and incommensurability, he moved from speaking about moving between theories as “translation” to a “bilingualism” that afforded multiple resources for understanding the world—a change that is particularly important when considering differences in terms as used in different subdisciplines.

Syrrim offers a really neat information theoretic account of predictive coding:

> Suppose you have an alphabet composed of 27 letters (the familiar 26 plus a space). You are interested in encoding it in binary for transmission. Of course you want to use as few bits as possible. How might you go about doing this? The first suggestion would be to assign each letter a bit patter of equal length. In this case, your transmission will take 4.76 bits each. You realize that in english some letters occur much more frequently than others, and to devote the same number of bits to each is wasteful. You find a table recording letter frequencies in common english texts, and reassign the bit patterns to give shorter values to more common letters. In this way, you reduce the number of bits needed to 4.03 per letter on average. Next you realize that some letters are followed by others even more commonly than they appear in normal text. Encoding the bit patterns based not only on the letter in question, but also the previous one reduces your usage to 3.32 bits per letter.

Now we play a game. A person is asked to guess what the current letter is. We tell them if they got it right or wrong. The right answer advances the current letter. They might initially guess the letter ‘t’. If they are right, they might further guess ‘h’. Getting that wrong, they could try ‘a’, and so on. The answer to each question, being yes or no, encodes a single bit. We record how many questions they ask over some long text, and therefore find the number of bits per letter to be 1.93.

(This example derived from Science and Information Theory by Leon Brillouin)

In this latter game, we ask the participants to guess (predict) what a letter is, and therefore define an encoding (coding) for each letter. The method by which a person performs this prediction is twofold. First, they have some idea what the text is saying, and therefore what it will say next. Second, every time they receive a negative response, they realize the text is saying something slightly different than they guessed, and so change their prediction for future letters.

The use of bits highlights an important practical application of all this. When you see some text as I am writing here, you see 4.76 bits for every letter (more, because of capitalization and punctuation and what not). And yet you require only 1.93 bits in order to know what is being said. The extra 2.83 bits take the form of redundancy. If I made some spelling error, or you read what I said particularly quickly, you might miss one of the letters I intended to convey. Yet because you have so many extra unnecessary bits, you can recover what is lost. This is done the similarly to how it was done in our game. As you read, you expect some letter to come next. When you encounter a slightly unexpected letter, you would update your expectation to account for it. When you encounter a completely unexpected letter, you might ignore it and continue as if your expectation was met.

To tie into the card example. A playing card contains log2(52) = 5.7 bits of information. If you are flashed a playing card very quickly, you might only have enough time to get 5.7 bits of information out of it. In this case, you would be forced to assume it is a playing card. If you have more time to look at it, you might be able to extract more bits, but even then, you might so heavily expect a playing card, that you ignore other possibilities.

Going back to the game: A person is allowed to ask which letter is next. But what makes the answer a single bit doesn’t depend on the nature of the question, only the binary nature of the answer. We could permit any yes or no question and still count bits by the number of questions. We then get into the interesting game of what question to ask. If someone had no clue what letter would follow, and wanted to determine it as quickly as possible, they might ask if it appeared in a particular half of the possible letters or the other. Or if they feel sufficiently confident in their guess, they might guess two or more letters at a time. (Brillouin points out the value 1.93 for the number of bits per letter must be too high because we force the player to ask for the letter even when it is obvious)

Now the playing card. You ask: “is it red or no?”, (no), “is it spades or no?”, (no). The prevailing paradigm implies that you now have complete information on the suit. “then it must be clubs?” (no). Once you realize that these are fake playing cards, you ask about the color

> and the suit independently. One could do a treatment of paradigms in science in a similar way: “is it a particle?” (yes) “then it isn’t a wave?” (no). “wait what?”…

Michael Watts writes:

> I find [the quote about dormitive potency] very interesting, because the paradigm everyone mocks (according to this) is the same paradigm current in medicine today.
>
> Years ago, I started to have a problem with the skin on my fingertips peeling off. This got to the point where I consulted a doctor, and he told me “we call this desquamation, which means “it’s peeling”. We don’t know why, and there’s nothing you can do about it.” Eventually, it cleared up by itself. We don’t know why.

There’s an old joke among doctors (at least I hope it’s a joke) that if you don’t know what a patient has, you just repeat their symptoms back to them in Greek or Latin:

“I get headaches at night and I don’t know why.”

“You have idiopathic nocturnal cephalgia.”

“Wow, you figured that out so quickly! Modern medicine really is amazing!”

JP corrects some of my terminology:

> It would be better to distinguish more clearly between schools and paradigms. Copernican astronomy, Newtonian mechanics and Predictive Coding are all schools. Only the former two were paradigms; that is, largely unchallenged and generally accepted. In the non- or prescientific stage medicine, psychology, … are

> currently in, there’s plenty of competing schools, and therefore no paradigm. What is required is an exemplar that sets the stage for a consolidation: a paradigmatic, i.e. paradigm-building, explanation for a phenomenon, after which everyone models their own explanations from hereon. For example (my example, not Kuhn’s), Darwin proposed a particular explanation for how the birds he found on the Galapagos islands got to have their beaks. Since then, a story about how something is in biology counts as an answer if and only if it has the same form as Darwin’s explanation.
>
> Constructing such explanations following the form of the exemplar is the process of Normal Science, which a truly scientific discipline is mostly engaged in: solving puzzles. That sounds dismissive, but solving a puzzle might be as interesting as explaining how birds came about – not just on Galapagos, but in general – that is, they’re dinosaurs. Exciting!
>
> I think the summary is also light on some of what Kuhn in particular was most interested in: in particular, incommensurability. Yes, Kuhn did indeed claim that we can make statements about the falsity of something only from within a certain paradigm (or school). Now Kuhn has plenty of inventory for talking about how a particular school might be thoroughly useless (i.e., it can be inconsistent and utterly fruitless) , but “empirically false from an objective, out-of-paradigm point of view” is not amongst them. In fact, it is inherent especially to a science following the highest standards that it is deeply embedded in one particular worldview, or one might say, ideology.

From John Nerst:

Kuhn gets overinterpreted a lot by people who like to push various species of relativism. As I see it, such overinterpretaton results from taking conclusions that only apply cleanly in the limit case and generalizing them to the whole domain. In this view the parts of a paradigm are all precisely dependent on each other for meaning to such an extent that if a paradigm is only somewhat different from another it is completely different and therefore not comparable at all and the distance between them is not meaningfully traversable. Paradigms are internally integrated and coherent, and insulated from each other. You have to pick one because it's impossible to mix them, and outside of a particular paradigm a concept means nothing at all. In or out.

Real science isn't like this, and therefore conclusions that follow from this don't necessarily apply. Kuhn uses examples that suggest it, but as many have said since then he kind of cherry picks and generalizing the pattern and using it to draw far-reaching and radical conclusions of science as a whole is, well, an overinterpretation.

In real life concepts are both a bit vague and meaningfully more-or-less different (instead of just "the same" or "different", full stop) in a way that makes it possible and in fact common to compare paradigms and pieces of paradigms (pieces that can be moved around without losing all of their meaning). This is because what we have are typically paradigm-like structures that overlap partially and are at least somewhat reconcilable. This is pretty true in the physical sciences and very true in the social sciences.

The ideas in TSOSR are valuable not because they describe science perfectly but because they work as a

> corrective to the prevailing view at the time. It’s one pole, and adding it to what we already had creates a new space (a spectrum where there used to be a point) which is great, but it’s important to remember that the new pole isn’t the whole space. To understand science you need both that side of the story and the fact-gathering/positivist/naive inductivist/whatever one. Generalizing only that facet gets you to the wrong place just as much as generalizing only the logical positivist side (or the falsificationist one if you want to get all multidimensional) does.

Virgil Kurkjian gives some eamples of Kuhn explaining how words have different meanings across paradigms:

> Revolutionary changes are different and far more problematic. They involve discoveries that cannot be accommodated within the concepts in use before they were made. In order to make or to assimilate such a discovery one must alter the way one thinks about and describes some range of natural phenomena. The discovery (in cases like these “invention” may be a better word) of Newton’s second law of motion is of this sort. The concepts of force and mass deployed in that law differed from those in use before the law was introduced, and the law itself was essential to their definition. A second, fuller, but more simplistic example is provided by the transition from Ptolemaic to Copernican astronomy. Before it occurred, the sun and moon were planets, the earth was not. After it, the earth was a planet, like Mars and Jupiter; the sun was a star; and the moon was a new sort of body, a satellite. Changes of that sort were not simply corrections of individual mistakes embedded in the Ptolemaic system. Like the transition to Newton’s laws of motion, they involved not only changes in laws of nature but also

> changes in the criteria by which some terms in those laws attached to nature […]
>
> One brief illustration of specialization’s effect may give this whole series of points additional force. An investigator who hoped to learn something about what scientists took the atomic theory to be asked a distinguished physicist and an eminent chemist whether a single atom of helium was or was not a molecule. Both answered without hesitation, but their answers were not the same. For the chemist the atom of helium was a molecule because it behaved like one with respect to the kinetic theory of gases. For the physicist, on the other hand, the helium atom was not a molecule because it displayed no molecular spectrum. Presumably both men were talking of the same particle, but they were viewing it through their own research training and practice. Their experience in problem-solving told them what a molecule must be. Undoubtedly their experiences had had much in common, but they did not, in this case, tell the two specialists the same thing. As we proceed we shall discover how consequential paradigm differences of this sort can occasionally be.

John Schilling notes that I left out part of the story in my explanation of Copernicanism and stellar parallax. The problem wasn’t just that the medievals *assumed* the stars were close. It was that they appeared to be discs rather than points, which ought to imply close proximity.

> absence of parallax isn’t a “glaring flaw” in Copernican theory; it’s only the combination of immeasurably small parallax and large apparent diameter of the fixed stars that is a glaring flaw. A finite diameter implies a finite distance, particularly with the reasonable assumption

> that stars are the same class of object as the Sun, and the stellar diameters measured by 16th and 17th-century observers corresponded to distances incompatible with the parallax measurements of those observers.
>
> This discrepancy could be resolved by better parallax measurements, or by better measurements of stellar diameter. And in fact, it was in 1720 that Halley used stellar occultation to show that the observed disks were optical anomalies and stellar angular diameter was immeasurably small – thus stars were immeasurably distant and could have immeasurably small parallax.
>
> As you note, it was not long after this (but see also James Bradley and aberration) that the Tychonic model was finally done away with and the Heliocentric model became dominant.

Frog-Like Sensations writes:

> It's natural to find Kuhn's metaphysics unclear since he was completely unclear about his metaphysics in Structure, and he spent much of the remainder of his career attempting to get clearer on it. Here's one of the last things he wrote about this:
>
> By now it may be clear that the position I'm developing is a sort of post-Darwinian Kantianism…Underlying all these processes of differentiation and change, there must, of course, be something permanent, fixed, and stable. But, like Kant's Ding an sich, it is ineffable, undescribable, undiscussible. Located outside of space and time, this Kantian source of stability is the whole from which have been fabricated both creatures and their niches, both the "internal" and the "external"

worlds. Experience and description are possible only with the described and describer separated, and the lexical structure which marks that separation can do so in different ways, each resulting in a different, though never wholly different, form of life. Some ways are better suited to some purposes, some to others. But none is to be accepted as true or rejected as false; none gives privileged access to a real, as against an invented, world. The ways of being-in-the-world which a lexicon provides are not candidates for true/false. (“The Road Since Structure”, 12)

Now, you may wonder how you can possibly make something clearer by saying it is a form of Kantianism, and as a non-Kant-scholar, I understand the feeling. But here’s my best stab at what’s going on here.

The most distinctive feature of Kant’s metaphysics is that he claims that a large number of things that are ordinarily claimed to be features of mind-independent reality — that is, of the world as it is in itself as opposed to how it is as represented by minds — are actually features of how our minds must represent the world. This includes both the obvious things, like color, and some really surprising things, like causality and the nature of space and time. So things in themselves do not enter into causal relations or exist in space and time, but they still exist and ultimately ground the nature of the world as it appears to us.

Kant’s view is not relativistic because (1) he thinks that the particular facts that are part of the world of appearance are (non-causally) determined by the nature of mind independent things (the “Ding an sich” mentioned above), and (2) he thinks that all minds

> impose the same kind of structure on the world (e.g., causal and with space and time).
>
> Kuhn's proposal is to reject the second claim. Instead of minds all imposing the same type of structure on the world, Kuhn suggests that changing paradigms can impose their respective structures on the world. There is still a mind-indpendent reality that in some way determines how things appear to us and also constrains how successful a given paradigm can be. But all the things that differ between paradigms concern only the features of our representation of reality. Mind-independent reality does not contain any of the relevant properties and so does not settle things one way or another, except insofar as it somehow renders one paradigm more useful than another at solving particular puzzles.
>
> Anyway, I don't find this view particularly appealing, but it's the most coherent thing I've managed to get out of Kuhn.

I have to admit I have some of the same confusions about Kant as I do about Kuhn. I understand Kant as saying that because we see the world through the mediating influence of our mind, we can never know anything about true reality.

I agree that we see the world through mediating influences, but I'm not sure how far he wants to go with the "never know anything about true reality" piece. For example, I believe I have a car. Can I say with some confidence that true reality contains an object corresponding to my car? That it really and truly has four wheels? That its gas tank is half full? That its interaction with my sense organs explains why I so consistently get such nicely-structured car-related sense-data?

Sure, you can say something boring like “wheels are a social construct, really there are just rubber molecules in a cylindrical pattern”, or even “rubber molecules and shapes are both social constructs, in reality there’s only blobs of quantum amplitude on a holographic boundary entity”, or even “in reality there’s something as far beyond quantum amplitude blobs as quantum amplitude blobs are beyond wheels”. But you can say this kind of thing without Kant, and we just shrug it off as “Yeah, on one level that’s true, but I’m right about the wheels too.” Does Kant have anything to add to this?

One nice thing about the subreddit’s karma system is that it makes it easier for me to figure out who to highlight here. The top-voted comment was by ArgumentumAdLapidem:

> This book is near and dear to my heart. As a young ArgumentumAdLapidem, a undergraduate physics major, I was really feeling my oats, and taking some upper-level history classes, just to prove I could do it. For some reason, some poor post-doc was assigned to do recitations, and got me, and I was STEMlording, as young STEMlords are wont to do. He gave me Kuhn to read. I read it, then bought it, then read it again. I had the same conclusion as SSC’s initial premise: this book is a fairly trivial description of the history of science. Lots of dirty laundry, to be sure, but nothing earth-shattering. He, of course, disagreed, and thought the book decisively proved that science was dethroned as the one-true-pursuit-of-Truth. Sadly, this story ends here, there was never a meeting of the minds. Reality intervened, there were finals to study for, and a wildly-overambitious lab project to complete.
>
> But I still have that book. Actually, I have two copies, as someone else, unbidden, gave me a copy as well.

Apparently history-of-science grads and philosophy-of-science grads hand them out to physics grads like garlic to vampires. (I readily admit, this might be a commentary on my former and/or current arrogance.) Over the years, I’ve thought about how I would have had that conversation differently. Here’s the current iteration:

To build a skyscraper, we need a foundation. The ultimate weight, volume, and height of the skyscraper is limited by the strength and soundness of the foundation. Science operates in a similar manner … the scope, accuracy, and detail of the scientific project is ultimately limited by the fundamental soundness of the model. The overall history of science, then, is the successive abandonment of one skyscraper for a bigger and better one, one with a stronger foundation, which allows the tower to reach greater heights.

But the devil is in the details, and Kuhn lays them out.

— There are people who have corner offices in the old skyscraper who don’t want to leave. They like their social status in this building, and they discourage (or punish) people who leave the building. They belittle people trying to build a new one.

— It’s not obvious, when the new foundation is being laid, that it will be any better or stronger than the existing one. You have to build the skyscraper (run the experiments) to find out.

— There are a lot of abandoned foundations laying around. They developed cracks, were built on unsuitable ground, or were otherwise deficient in some way that wasn’t discovered until they actually tried to build something on top of it. Most new scientific models fail. There are fads – some hot new model will attract a lot of attention, but begins to fade when it doesn’t show results. The scions of the current building can point to all the failure around them and confidently predict this new attempt will fail as well.

— As the skyscraper is being built, it’s not a smooth process. There will be mistakes and partial rebuilds. Most of the the time, the new building will be a piecemeal framework of exposed structural beams, and will spend most of its time being shorter and less comfortable than the old building. The corner offices of the old building will look out their windows, see a tangle of metal and sweat in the construction site below them, and chuckle at their naive enthusiasm.

— The old building does still grow. There are remodels, things get slicker, more polished, expansions are added, maybe another floor is added. But the foundation can still only take so much, and can only be reinforced to a certain extent. Epicycles.

— The new building has new problems the old building didn’t have. The fire suppression system needs more powerful pumps to push water to ever higher floors. The doorman who just knew everybody has been replaced by a keycard authentication system that is confusing and annoying. These look like flaws to people in the old building, rather than the necessary scaffolding for a bigger, better building. The flat-earther model “Earth must be flat because look how far I can see”, which is simple, must be replaced with the more powerful “Earth is round, and, in a vacuum, you wouldn’t be able to see that far, but we must account for atmospheric refraction, here’s some corrections.” Annoying. But it

> isn’t just replacing one problem for another. The old problem was a fundamentally-limiting contradiction in the basic model that couldn’t be solved without scrapping the model. The new problem might be solvable. You won’t know until you try. You have to build the building to know.
>
> — There’s a perception problem. The old building holds the height (truth) record basically until the new building reaches the height of the old one. Then the record goes to the new building, and the perception shifts – if you want to be in the game, you got to be in the new building. Some observer, watching the endless parade of people suddenly moving their boxes to the new building, concludes this is all just fad-chasing, like socialites flocking to the hottest club. They’re just doing whatever is popular with the other scientists.
>
> So yes, all this is true. But, after all those failed attempts, all that drama, all that sneering and popularity-games, the skyscrapers still do get taller. As SSC notes, Kuhn barely admits this, in a whisper, on the last page. It is no wonder then, that this book has been used to represent claims far beyond what Kuhn actually claims.

And MoreDonuts on Kuhn vs. Popper:

> The other simplistic view [Kuhn] was arguing against was Popper’s notion of falsification. In fact, falsification was the legal precedent for the definition of science at the time, in spite of the fact that philosophers of science never considered it very seriously.
>
> Kuhn’s view also answers the question of why falsification has always been popular among scientists

> on the ground. When a field is performing “normal science” under a particular paradigm, the acceptance of particular facts or pieces of theory largely does resemble falsification: either the new proposal fits the evidence under the paradigm, or it does not. Kuhn (and Feyerabend) show how this simplistic model falls apart when comparing between paradigms, because there is no way to agree upon what constitutes falsification.
>
> Philosophy of science is controversial because the core conclusion is largely unavoidable: “science” is simply a set of human institutions. There is no hard philosophical grounding for scientific truth. This was an unpopular conclusion historically because Christians were still trying to push Creationism, and progressives needed some argument for why scientific institutions were right and Christian institutions were wrong (the real answer, unironically: our people are smarter and less biased).

A couple of people commented that Kuhn was overstating things because Einstein just expanded upon Newton – a friendly amendment, if you will. Kingshorsey explains (using similar arguments to Kuhn himself) why this isn’t quite right:

> I think there are two important lessons to take away from Kuhn: 1) the gap between our ability to model phenomena and our ability to explain those phenomena can be uncomfortably large; and 2) the perceived amount of empirical advantage provided by a new paradigm is not necessarily commensurate with the amount of conceptual adjustment its adoption will require.
>
> A user on the SSC site said that the move from Newtonian to Einsteinian physics is more of a paradigm shuffle than a paradigm shift, because Newtonian

equations still work perfectly well for all kinds of calculations. To reframe this user’s statement in terms of point 2, this user thinks that because Einstein’s calculations empirically differ from Newton’s in only certain restricted cases, Einstein’s paradigmatic/theoretical challenge to Newton must be similarly small.

But that’s taking an unreasonably narrow view of what constitutes Newtonianism and Einsteinianism. Neither Newton nor Einstein produced equations in a conceptual vacuum. Rather, both embedded them within a cosmology that rendered them intelligible.

To Newton, space was absolute and yet non-substantive, just the distance between objects. Time was uniform and absolute. Gravity operated instantaneously apart from mediation. Newton believed that these cosmological assertions were necessary for his physics, and that in turn his physics supported these cosmological assertions.

When Einstein comes along, he overturns everything Newton thought about the nature of the universe. Space and time are no longer to be regarded as merely formal properties “within” which things move. Time is relative, space and time are intertwined, and space-time is the very “thing” of which gravity consists.

If we accept both that Einstein’s cosmology is better and that Newton’s math is still pretty good (rather than junk science), we are left with an uncomfortable conclusion. Newton’s degree of success at modeling phenomena in motion did not correlate strongly with his degree of success at explaining the structures or

> characteristics of reality responsible for that phenomena.
>
> This in turn should lead us to question how much the success of Einstein’s math really supports the cosmology that is bound up with it. After all, what’s to stop a future physicist from saying, “Thanks for these equations, Einstein, I’ll use them where I can, but it’s a shame your model of reality was all wrong”?
>
> And that’s why Kuhn is interesting, and comforting, and frightening. The conservation of certain observations through paradigm shifts forces us to reckon with the possibility that our own scientific successes may one day find a home in a model of reality entirely other than what we imagine now.

Jadagul has a whole blog post worth of comment.

And SpinyStellate doesn’t have much to say about the book, but recommends to us their project SciDash, “rigorous, reproducible, extensible, data-driven model validation [and visualization] for science”. I haven’t looked at it enough to entirely get what’s going on, but at least check it out for its cool visualization of geocentrism vs. heliocentrism (complete with p-values)!

2018 Predictions: Calibration Results

At the beginning of every year, I make predictions. At the end of every year, I score them. Here are 2014, 2015, 2016, and 2017.

And here are the predictions I made for 2018. Strikethrough'd are false. Intact are true. Italicized are getting thrown out because I can't decide if they're true or not. Please don't complain that 50% predictions don't mean anything; I know this is true but there are some things I'm genuinely 50-50 unsure of.

US:
1. Donald Trump remains president at end of year: 95%
2. Democrats take control of the House in midterms: 80%
~~3. Democrats take control of the Senate in midterms: 50%~~
~~4. Mueller's investigation gets cancelled (eg Trump fires him): 50%~~
5. Mueller does not indict Trump: 70%
~~6. PredictIt shows Bernie Sanders having highest chance to be Dem nominee at end of year: 60%~~
7. PredictIt shows Donald Trump having highest chance to be GOP nominee at end of year: 95%
8. [This was missing in original]
~~9. Some sort of major immigration reform legislation gets passed: 70%~~
10. No major health-care reform legislation gets passed: 95%
11. No large-scale deportation of Dreamers: 90%
12. US government shuts down again sometime in 2018: 50%
13. Trump's approval rating lower than 50% at end of year: 90%
~~14. …lower than 40%: 50%~~

15. GLAAD poll suggesting that LGBQ acceptance is down will mostly not be borne out by further research: 80%

ECONOMICS AND TECHNOLOGY:
~~16. Dow does not fall more than 10% from max at any point in 2018: 50%~~
~~17. Bitcoin is higher than $5,000 at end of year: 95%~~
~~18. Bitcoin is higher than $10,000 at end of year: 80%~~
19. Bitcoin is lower than $20,000 at end of year: 70%
20. Ethereum is lower than Bitcoin at end of year: 95%
21. Luna has a functioning product by end of year: 90%
~~22. Falcon Heavy first launch not successful: 70%~~
23. Falcon Heavy eventually launched successfully in 2018: 80%
24. SpaceX does not attempt its lunar tourism mission by end of year: 95%
25. Sci-Hub is still relatively easily accessible from within US at end of year (even typing in IP directly is relatively easy): 95%
26. Nothing particularly bad (beyond the level of an funny/weird news story) happens because of ability to edit videos this year: 90%
~~27. A member of the general public can ride-share a self-driving car without a human backup driver in at least one US city by the end of the year: 80%~~

CULTURE WARS:
28. Reddit does not ban r/the_donald by the end of the year: 90%
29. None of his enemies manage to find a good way to shut up/discredit Jordan Peterson: 70%

COMMUNITIES:
30. SSC gets more hits in 2018 than in 2017: 80%
31. SSC gets mentioned in the New York Times (by someone other than Ross Douthat): 60%

~~32. At least one post this year gets at least 100,000 hits: 70%~~
33. A 2019 SSC Survey gets posted by the end of the year: 90%
34. No co-bloggers make 3 or more SSC posts this year: 80%
35. Patreon income less than double current amount at end of year: 90%
~~36. A scientific paper based on an SSC post is accepted for publication in real journal by end of year: 60%~~
~~37. I do an adversarial collaboration with somebody interesting by the end of the year: 50%~~
38. I successfully do some general project to encourage and post more adversarial collaborations by other people: 70%
39. New SSC meetups system/database thing gets launched successfully: 60%
40. LesserWrong remains active and successful (average at least one halfway-decent post per day) at the end of the year: 50%
41. LesserWrong is declared official and merged with LessWrong.com: 80%
42. I make fewer than five posts on LessWrong (posts copied over from SSC don’t count): 70%
43. CFAR buys a venue this year: 50%
44. AI Impacts has at least three employees working half-time or more sometime this year: 50%
~~45. Rationalists get at least one more group house on Ward Street: 50%~~
46. No improvement in the status of reciprocity.io (either transfer to a new team or at least one new feature added): 70%

PERSONAL:
47. I fail at my New Years’ resolution to waste less time on the Internet throughout most of 2018: 80%
~~48. I fail at my other New Years’ resolution to try one~~

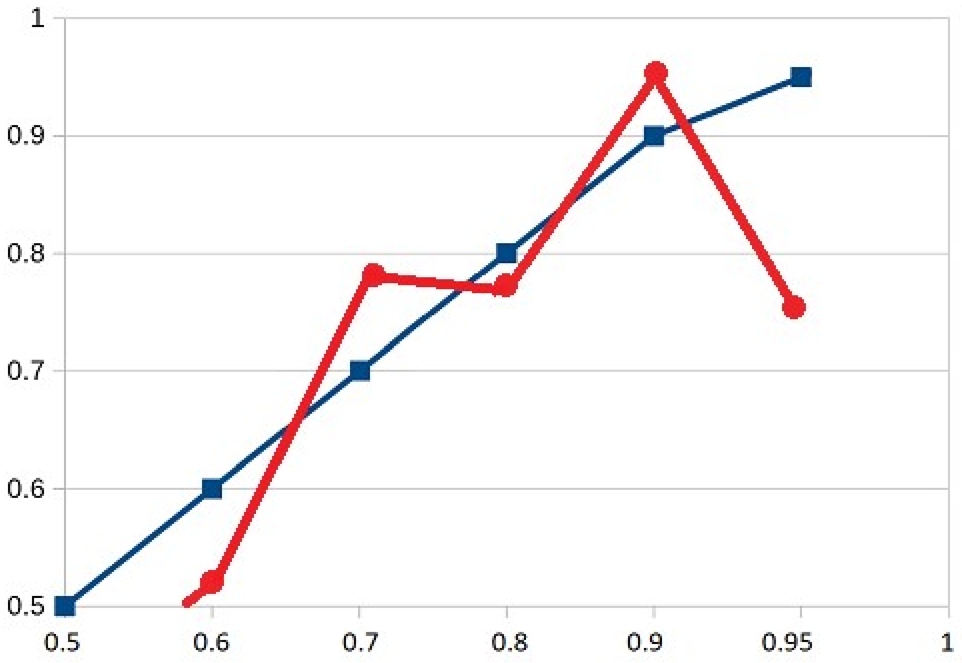
~~biohacking project per month throughout 2018: 80%~~
49. I don’t attend the APA National Meeting: 80%
50. I don’t attend the New York Solstice: 80%
51. I travel outside the US in 2018: 90%
~~52. I get some sort of financial planning sorted out by end of year: 95%~~
~~53. I get at least one article published on a major site like Vox or New Statesman or something: 50%~~
~~54. I get a tax refund: 50%~~
55. I weigh more than 195 lb at year end: 60%
56. I complete the currently visible Duolingo course in Spanish: 90%
57. I don’t get around to editing Unsong (complete at least half the editing by my own estimate) this year: 95%
58. No new housemate for at least one month this year: 90%
59. I won’t [meditate at least one-third of days this year]: 90%
60. I won’t [do my exercise routine at least one third of days this year]: 80%
~~61. I still live in the same house at the end of 2018: 60%~~
~~62. I will not have bought a house by the end of 2018: 90%~~
63. Katja’s paper gets published: 90%
~~64. Some other paper of Katja’s gets published: 50%~~

SECRET: (mostly speculating on the personal lives of friends who read this blog; I don’t necessarily want them to know how successful I expect their financial and romantic endeavors to be. I’ve declassified the ones that now seem harmless to admit.)

65. My partner and I come to a decision about whether to have children: 80%
~~66. My partner and I are engaged by the end of the year: 70%~~
~~67. My partner and I do not break up by the end of the year:~~

~~70%~~
~~68. [Secret prediction]: 60%~~
69. [Secret prediction]: 70%
~~70. [Secret prediction]: 60%~~
~~71. [Secret prediction]: 50%~~
~~72. [Secret prediction]: 50%~~
~~73. [Secret prediction]: 50%~~
74. [Secret prediction]: 90%
75. [Secret prediction]: 90%
76. [Secret prediction]: 60%
77. [Secret prediction]: 70%
78. [Secret prediction]: 60%
~~79. [Secret prediction]: 50%~~
80. [Secret prediction]: 60%
81. I lose my bet against Duncan about Dragon Army Barracks: 80%
82. Dragon Army Barracks is still together at the end of the year: 70%
~~83. I will visit Greece: 50%~~
~~84. I will visit Germany: 70%~~
85. [Secret prediction]: 70%
~~86. [Secret prediction]: 70%~~
87. [Secret prediction]: 60%
88. [Secret prediction]: 50%
~~89. [Secret prediction]: 50%~~
90. [Secret prediction]: 70%
91. [Secret prediction]: 90%
92. [Secret prediction]: 50%
93. Still working at my current job at the end of 2018: 90%
~~94. Working 30 hours/week or less at the end of 2018: 50%~~
~~95. Have switched to practicing entirely in the East Bay: 60%~~
96. [Secret prediction]: 60%
97. Will not finish first section of a difficult calligraphy project: 60%
98. Will not finish all sections of difficult calligraphy project:

95%
99. I will not do work for AI Impacts by the end of the year: 70%
100. I will not finish more than 25% of a new novel: 70%



Calibration chart. The blue line represents perfect calibration, the red line represents my predictions. The closer they are, the better I am doing.

Of 50% predictions, I got 6 right and 16 wrong, for a score of 27%
Of 60% predictions, I got 8 right and 7 wrong, for a score of 53%
Of 70% predictions, I got 14 right and 4 wrong, for a score of 78%
Of 80% predictions, I got 10 right and 3 wrong, for a score of 77%
Of 90% predictions, I got 17 right and 1 wrong, for a score of 94%

Of 95% predictions, I got 6 right and 2 wrong, for a score of 75%

50% predictions are technically meaningless since I could have written them either way – which makes it surprising I managed to get such an imbalance between right and wrong. I think I’m more wrong than should be statistically possible. I’m not sure what to think about that.

After that, things go okay until the 95% level, where I get a very poorly calibrated 75%. This is partly the fault of not having very many 95% predictions this year, but even so I should have done better than this.

Two things happened that screwed with a lot of my predictions. First, cryptocurrency crashed (remember, I made last year’s prediction during the height of the boom, when Bitcoin was around $15,000). I expected it would go down, but not this much. Since I made a lot of predictions about cryptocurrency and all of them were correlated, this went badly. I can hear the ghostly sound of Nassim Nicholas Taleb laughing at me.

The other thing that happened was that my partner unexpectedly broke up with me, changing all of my life plans and precipitating a move to a different house. Again, this was a black swan that affected a lot of correlated predictions. In this case, my move took a lot of money, which meant I didn’t have enough money to be worth investing, which means I didn’t bother doing any fancy financial planning like I had been dead set on doing in January 2018. I’m usually pretty good at following through on important things, so I was 95% sure I would get the financial planning done, but black swan = spend savings = no point in financial planning was something I hadn’t considered.

I’m not sure how to deal with those sorts of correlations here except to not make too many correlated predictions.

I’ll post my 2019 predictions later this week. If you’ve made some of your own, post a link in the comments and I’ll link them along with mine. And while you’re waiting, I also made some predictions last February for the next five years.

Psychiat-List Now Up

Lots of people have asked me to recommend them a psychiatrist or therapist. I’ve done a terrible job responding: it’s a conflict of interest to recommend my own group, and I don’t know many people outside of it.

So now I’ve put together a list (by which I mostly mean blatantly copied a similar list made by fellow community member Anisha M) of mental health professionals whom members of the rationalist community have had good experiences with. So far it’s short and mostly limited to the Bay Area. You can find it at the “Psychiat-List” button on the top of the blog, or at this link.

My hope is to crowd-source additional recommendations to expand the list to more providers and cities. Please let me know, either on this post or on the comments to the list itself, if you have any extra recommendations to add – especially if you’re in a city likely to have many other SSC readers. Please also let me know if you’ve had any positive or negative experiences with people already on the list, so I can change their status accordingly.

Predictions For 2019

At the beginning of every year, I make predictions. At the end of every year, I score them. So here are a hundred more for 2019.

Rules: all predictions about what will be true on January 1, 2020. Any that involve polling will be settled by the top poll or average of polls on Real Clear Politics on that day. Most predictions about my personal life, or that refer to the personal lives of other people, have been redacted to protect their privacy. I’m using the full 0 – 100 range in making predictions this year, but they’ll be flipped and judged as 50 – 100 in the rating stage, just like in previous years. I’ve tried to avoid doing specific research or looking at prediction markets when I made these, though some of them I already knew what the markets said.

Feel free to get in a big fight over whether 50% predictions are meaningful.

US
1. Donald Trump remains President: 90%
2. Donald Trump is impeached by the House: 40%
3. Kamala Harris leads the Democratic field: 20%
4. Bernie Sanders leads the Democratic field: 20%
5. Joe Biden leads the Democratic field: 20%
6. Beto O’Rourke leads the Democratic field: 20%
7. Trump is still leading in prediction markets to be Republican nominee: 70%
8. Polls show more people support the leading Democrat than the leading Republican: 80%
9. Trump’s approval rating below 50: 90%
10. Trump’s approval rating below 40: 50%
11. Current government shutdown ends before Feb 1: 40%

12. Current government shutdown ends before Mar 1: 80%
13. Current government shutdown ends before Apr 1: 95%
14. Trump gets at least half the wall funding he wants from current shutdown: 20%
15. Ginsberg still alive: 50%

ECON AND TECH
16. Bitcoin above 1000: 90%
17. Bitcoin above 3000: 50%
18. Bitcoin above 5000: 20%
19. Bitcoin above Ethereum: 95%
20. Dow above current value of 25000: 80%
21. SpaceX successfully launches and returns crewed spacecraft: 90%
22. SpaceX Starship reaches orbit: 10%
23. No city where a member of the general public can ride self-driving car without attendant: 90%
24. I can buy an Impossible Burger at a grocery store within a 30 minute walk from my house: 70%
25. Pregabalin successfully goes generic and costs less than $100/month on GoodRx.com: 50%
26. No further CRISPR-edited babies born: 80%

WORLD
27. Britain out of EU: 60%
28. Britain holds second Brexit referendum: 20%
29. No other EU country announces plan to leave: 80%
30. China does not manage to avert economic crisis (subjective): 50%
31. Xi still in power: 95%
32. MbS still in power: 95%
33. May still in power: 70%
34. Nothing more embarassing than Vigano memo happens to Pope Francis: 80%

SURVEY

35. …finds birth order effect is significantly affected by age gap: 40%
36. …finds fluoxetine has significantly less discontinuation issues than average: 60%
37. …finds STEM jobs do not have significantly more perceived gender bias than non-STEM: 60%
38. …finds gender-essentialism vs. food-essentialism correlation greater than 0.075: 30%

PERSONAL

39. SSC gets fewer hits than last year: 70%
40. I finish and post [redacted]: 90%
41. I finish and post [redacted 2]: 50%
42. I finish and post [redacted 3]: 50%
43. [redacted 1] post gets at least 40,000 hits: 40%
44. [redacted 2] post gets at least 40,000 hits: 20%
45. New co-blogger with more than 3 posts: 20%
46. Repeat adversarial collaboration contest with at least 5 entries: 60%
47. [redacted]: 90%
48. [redacted]: 70%
49. I start using Twitter again (5+ tweets in any month): 60%
50. I start using Facebook again (following at least 5 people): 30%
51. I get the blood tests I should be getting this year: 90%
52. I try one biohacking project per month x at least 10 months: 30%
53. I continue taking sceletium regularly: 70%
54. I switch from [redacted] for at least 3 months: 20%
55. I find at least one new supplement I take or expect to take regularly x 3 months: 20%
56. Minoxidil use produces obvious progress: 50%
57. I restart [redacted]: 20%
58. I spend one month at least substantially more

vegetarian than my current compromise: 20%
59. I spend one month at least substantially less vegetarian than my current compromise: 30%
60. I weight more than 195 lbs at year end: 80%
61. I meditate at least 30 minutes/day more than half of days this year: 30%
62. I use marijuana at least once this year: 20%
63. I finish at least 10% more of [redacted]: 20%
64. I completely finish [redacted]: 10%
65. I finish and post [redacted]: 5%
66. I write at least ten pages of something I intend to turn into a full-length book this year: 20%
67. I practice calligraphy at least seven days in the last quarter of 2019: 40%
68. I finish at least one page of the [redacted] calligraphy project this year: 30%
69. I finish the entire [redacted] calligraphy project this year: 10%
70. I finish some other at-least-one-page calligraphy project this year: 80%
71. I attend the APA Meeting: 80%
72. [redacted]: 50%
73. [redacted]: 40%
74. I still work in SF with no plans to leave it: 60%
75. I still only do telepsychiatry one day with no plans to increase it: 60%
76. I still work the current number of hours per week: 60%
77. I have not started (= formally see first patient) my own practice: 80%
78. I lease another version of the same car I have now: 90%
79. I still live in my current house with no specific plans to leave: 80%
80. I set up a decent home library: 60%
81. We have obtained a second trash can: 90%
82. The gate is fixed with no problems at all: 50%
83. The ugly paint spot on my wall gets fixed: 30%

84. There is some kind of nice garden: 60%
85. …and I am at least half responsible: 20%
86. I get my own washing machine: 20%
87. [redacted]: 60%
88. [redacted]: 70%
89. [redacted]: 80%
90. [redacted]: 80%
91. [redacted] is widely considered a success: 70%
92. …with plans (vague okay) to create a second [redacted]: 20%
93. I find a primary partner: 30%
94. I go on at least one date with someone who doesn’t already have a primary partner: 90%
95. I remake an account on OKCupid: 80%
96. [redacted]: 10%
97. [redacted]: 20%
98. [redacted]: 20%
99. [redacted]: 20%
100. [redacted]: 20%
101. [redacted]: 30%
102. [redacted]: 10%
103. [redacted]: 30%
104. [redacted]: 50%
105. [redacted]: 10%
106. [redacted]: 50%
107. I am still playing D&D: 60%
108. I go on a trip to Guatemala: 90%
109. I go on at least one other international trip: 30%
110. I go to at least one Solstice outside the Bay: 40%
111. I go to at least one city just for an SSC meetup: 30%
112. [redacted]: 40%
113. [redacted]: 50%
114. [redacted]: 20%
115. [redacted]: 80%
116. [redacted]: 60%

117. [redacted]: 60%
118. [redacted]: 80%

Book Review: Zero To One

I.

Zero To One might be the first best-selling business book based on a Tumblr. Stanford student Blake Masters took Peter Thiel's class on startups. He posted his notes on Tumblr after each lecture. They became a minor sensation. Thiel asked if he wanted to make them into a book together. He did.

The title comes from Thiel's metaphor that ordinary businessmen like restaurant owners take a product "from 1 to n" (shouldn't this be from n to n+1?) – they build more of something that already exists. But the greatest entrepreneurs bring something "from 0 to 1" – they invent something that has never been seen before.

The book has various pieces of advice for such entrepreneurs. Three sections especially struck me: on monopolies, on secrets, and on indefinite optimism.

II.

A short review can't fully do justice to the book's treatment of monopolies. Gwern's look at commoditizing your complement almost does (as do some tweets). But the basic economic argument goes like this: In a normal industry (eg restaurant ownership) competition should drive profit margins close to zero. Want to open an Indian restaurant in Mountain View? There will be another on the same street, and two more just down the way. If you automate every process that can be automated, mercilessly pursue efficiency, and work yourself and your employees to the bone – then you can just barely compete on price. You can earn enough money to live, and to not immediately give up in disgust and go into another line of business (after all, if you didn’t earn that much, your competitors would already have given up in disgust and gone into another line of business, and your task would be easier). But the average Indian restaurant is in an economic state of nature, and its life will be nasty, brutish, and short.

This was the promise of the classical economists: capitalism will optimize for consumer convenience, while keeping businesses themselves lean and hungry. And it was Marx’s warning: businesses will compete so viciously that nobody will get any money, and eventually even the capitalists themselves will long for something better. Neither the promise nor the warning has been borne out: business owners are often comfortable and sometimes rich. Why? Because they’ve escaped competition and become at least a little monopoly-like. Thiel says this is what entrepreneurs should be aiming for.

He hates having to describe how businesses succeed, because he thinks it’s too anti-inductive to reduce to a formula:

> Tolstoy opens Anna Karenina by observing “All happy families are alike; each unhappy family is unhappy in its own way.” Business is the opposite. All happy companies are different: each one earns a monopoly by solving a unique problem. All failed companies are the same: they failed to escape competition.

But he grudgingly describes four ways that a company can successfully reach monopolyhood:

1. Proprietary technology. This one is straightforward. If you invent the best technology, and then you patent it, nobody

else can compete with you. Thiel provocatively says that your technology must be 10x better than anyone else's to have a chance of working. If you're only twice as good, you're still competing. You may have a slight competitive advantage, but you're still competing and your life will be nasty and brutish and so on just like every other company's. Nobody has any memory of whether Lycos' search engine was a little better than AltaVista's or vice versa; everybody remembers that Google's search engine was orders of magnitude above either. Lycos and AltaVista competed; Google took over the space and became a monopoly.

2. Network effects. Immortalized by Facebook. It doesn't matter if someone invents a social network with more features than Facebook. Facebook will be better than their just by having all your friends on it. Network effects are hard because no business will have them when it first starts. Thiel answers that businesses should aim to be monopolies from the very beginning – they should start by monopolizing a tiny market, then moving up. Facebook started by monopolizing the pool of Harvard students. Then it scaled up to the pool of all college students. Now it's scaled up to the whole world, and everyone suspects Zuckerberg has somebody working on ansible technology so he can monopolize the Virgo Supercluster. Similarly, Amazon started out as a bookstore, gained a near-monopoly on books, and used all of the money and infrastructure and distribution it won from that effort to feed its effort to monopolize everything else. Thiel describes how his own company PayPal identified eBay power sellers as its first market, became indispensible in that tiny pool, and spread from there.

3. Economies of scale. Also pretty straightforward, and especially obvious for software companies. Since the marginal cost of a unit of software is near-zero, your cost

per unit is the cost of building the software divided by the number of customers. If you have twice as many customers as your nearest competitor, you can charge half as much money (or make twice as much profit), and so keep gathering more customers in a virtuous cycle.

4. Branding Apple is famous enough that it can charge more for its phones than Amalgamated Cell Phones Inc, even for comparable products. Partly this is because non-experts don’t know how to compare cell phones, and might not trust *Consumer Reports* style evaluations; Apple’s reputation is an unfakeable sign that their products are pretty good. And partly it’s just people paying extra for the right to say “I have an iPhone, so I’m cooler than you”. Another company that wants Apple’s reputation would need years of successful advertising and immense good luck, so Apple’s brand separates it from the competition and from the economic state of nature.

Thiel continues with various counterintuitive pieces of wisdom. Don’t try to “disrupt” your field – if you’re “disrupting” someone, it means you’re competing with them, and making enemies who will try to hold you back. Don’t try to be the “first mover” (Yahoo was the first-mover in the search engine space), instead try to be the “last mover” whom nobody is able to supplant. Etc, etc. Just try to get a monopoly or something like it.

Is all of this a plot against the public? Monopolies are usually viewed as cheating the system and preventing progress; is Thiel promoting that behavior to the detriment of society? Well, obviously he *says* he isn’t:

> The problem with a competitive business goes beyond lack of profits. Imagine you’re running one of those restaurants in Mountain View. You’re not that different

> from dozens of your competitors, so you’ve got to fight hard to survive. If you offer affordable food with low margins, you can probably pay employees only minimum wage. And you’ll need to squeeze out every efficiency: that’s why small restaurants put Grandma to work at the register and make the kids wash dishes in the back. Restaurants aren’t much better even at the very highest rungs, where reviews and ratings like Michelin’s star system enforce a culture of intense competition that can drive chefs crazy. (French chef and winner of three Michelin stars Bernard Loiseau was quoted as saying, “If I lose a star, I will commit suicide.” Michelin maintained his rating, but Loiseau killed himself anyway in 2003 when a competing French dining guide downgraded his restaurant.) The competitive ecosystem pushes people toward ruthlessness or death.
>
> A monopoly like Google is different. Since it doesn’t have to worry about competing with anyone, it has wider latitude to care about its workers, its products, and its impact on the wider world. Google’s motto — “Don’t be evil” — is in part a branding ploy, but it’s also characteristic of a kind of business that’s successful enough to take ethics seriously without jeopardizing its own existence. In business, money is either an important thing or it is everything. Monopolists can afford to think about things other than making money; non-monopolists can’t. In perfect competition, a business is so focused on today’s margins that it can’t possibly plan for a long-term future. Only one thing can allow a business to transcend the daily brute struggle for survival: monopoly profits.

So monopolies’ advantages include being better for employees, more socially responsible, and able to engage in long-term thinking. The classic examples of this (which I don’t think Thiel brought up) are Bell Labs and Xerox PARC. Two monopolistic companies with more money than they knew what to do with started super-basic-blue-sky research centers that ended up creating many of the technologies that shaped the modern world (Bell Labs, started by AT&T, helped invent the transistor, the laser, information theory, UNIX, C, C++, radio astronomy, etc; PARC, started by Xerox, helped invent Ethernet, laser printing, the personal computer, graphical user interfaces, object-oriented programming, bitmaps, and the LCD.) Google X wants to be the modern version of this kind of thing, though I don’t know how much success they’ve had so far.

On the other hand, all of the classical disadvantages of monopolies are still there. Monopolies remove the pressure to do a good job – whether that’s in keeping prices low, keeping working conditions tolerable, or in keeping products and service high-quality. They lower the diversity of an industry, making it more likely to get stuck in an evolutionary blind alley it can’t get out of; they increase the risk of merging with government into a crony capitalism. A wolf sheltered from survival-of-the-fittest for too long becomes a Chihuahua; Amazon sheltered from survival-of-the-fittest for too long becomes the DMV.

(also, isn’t Thiel the guy who wanted floating independent seasteads because competitive governance would break the monopoly of existing nation-states and lead to a revolutionary improvement in institutional capacity? Doesn’t that suggest even he acknowledges monopolies are often bad?)

I don’t think this is one of those issues that’s going to get decisively solved in a few paragraphs. Moloch and Slack are the new yin and yang, the new chaos and order; their interplay creates the Ten Thousand Things. Err too far towards competition and everyone works themselves to death in garment sweatshops; err too far towards monopoly and everyone sits at a desk filling out forms and backstabbing each other until the lights slowly go out. It’s only in the collision zone between the two that anything interesting ever happens.

One could rescue Thiel’s position by assuming that competition will always be with us. Google’s pretty monopoly-like, but even they can’t rest on their laurels too long. It’s not just that Bing might take over, but that advertisers might get better non-search-engine ways to place ads, Facebook might come up with better ways to target ads, some alternate platform like cell phones or VR might take over from classic Internet searches, or something else. Whatever their concern, real-life Google sure does seem to put a lot of effort into being competitive. So sure, maybe one has to find the sweet spot between perfect competition and perfect monopoly, but one could argue that right now only the most monopolistic companies are near that sweet spot.

III.

The rest of *Zero To One* becomes less directly about the startup world, and more about deep social trends that good startup founders will have to buck. One such trend – which Thiel approaches in a lot of different equivalent ways – is the loss of belief in secrets. People no longer believe that there are important things that they don’t know, but which they could discover if they tried a little harder.

Past scientific discoveries came from a belief in secrets. Isaac Newton wondered why apples fell, thought “Maybe if I work really hard on this problem, I can discover something nobody has ever learned before”, and then set out to do it.

Modern people aren't just less likely to think this way. They're actively discouraged from it by a culture which mocks stories like Newton's as "the myth of the lone genius". Nowadays people get told that if they think they've figured out something about gravity, they're probably a crackpot. Instead, they should wait for very large government-funded programs full of well-credentialled people to make incremental advances.

Good startups require a belief in secrets, where "secret" is equivalent to "violation of the efficient market hypothesis". You believe you've discovered something that nobody else has: for example, that if you set up an online bookstore in such-and-such a way today, in thirty years you'll be richer than God. This is an outrageously arrogant claim: that you have spotted a hundred-billion-dollar bill lying on the sidewalk that everyone else has missed. But only people who believe something like it can noncoincidentally found great companies. You must believe there are lucrative secrets hidden in plain sight.

Thiel relates this to the decline of cults (see these two essays fleshing out the phenomenon). Although cults may not be *desirable*, they are the failure mode of individuals and small groups trying to throw off conventional wisdom and discover profound new ways of looking at the world. Have we lost our cults because we no longer fail at this task, or because we no longer attempt it at all?

Belief in secrets is connected to belief in one's own reasoning abilities. Modern conventional wisdom says armchair reasoning never works; any idea you prove true in your head is useless until it's been exhaustively tested in real life, and you're more likely to get some other (true) idea out of the exhaustive testing than to validate your armchair speculation. As a corollary, the more steps in your proof, the less likely it is, since each one exponentially increases the error rate of your final conclusion. Since your armchair reasoning is useless, you are unlikely to ever discover a secret (except perhaps by chance, if you randomly do experiments no one else has ever done). The only thing that might not be useless is large institutions working together to gradually advance knowledge with lots of testing, who effectively buy many lottery tickets hoping one will pay off.

The modern skepticism about secrets and reasoning implies a similar skepticism about planning. If the argument against multi-step reasoning is right, then a mildly Internet-famous scene from *Harry Potter And The Methods Of Rationality* is right too:

> Father had told Draco about the Rule of Three, which was that any plot which required more than three different things to happen would never work in real life. Father had further explained that since only a fool would attempt a plot that was as complicated as possible, the real limit was two.

This disbelief in planning suggests, not a strategy, but a sort of meta-strategy. Do something vaguely in the space of what you want to do, don't commit yourself to a specific plan, watch what happens, iterate, keep your options open at all time, and be prepared to pivot quickly once you know more.

But Thiel says the most successful visionaries of the past did the opposite of this. They knew what they wanted, planned a strategy, and achieved it. The Apollo Program wasn't run by vague optimism and "keeping your options open". It was run by some people who wanted to land on the moon, planned out how to make that happen, and followed the plan. Not slavishly, and certainly they were responsive to evidence that they should change tactics on specific points. But they had a firm vision of the goal in their minds, an approximate vision of what steps they would take to achieve it, and a belief that acheiving an ambitious long-term plan was the sort of thing that people could be expected to do. And great startups like SpaceX are much the same. Elon Musk started with a n-step plan to get to Mars, and he's currently about halfway through.

He gives one particularly striking example of the past's attitude to secrets and planning:

> Bold plans were not reserved just for political leaders or government scientists. In the late 1940s, a Californian named John Reber set out to reinvent the physical geography of the whole San Francisco Bay Area. Reber was a schoolteacher, an amateur theater producer, and a self-taught engineer. Undaunted by his lack of credentials, he publicly proposed to build two huge dams in the Bay, construct massive freshwater lakes for drinking water and irrigation, and reclaim 20,000 acres of land for development. Even though he had no personal authority, people took the Reber Plan seriously. It was endorsed by newspaper editorial boards across California. The U.S. Congress held hearings on its feasibility. The Army Corps of Engineers even constructed a 1.5-acre scale model of the Bay in a cavernous Sausalito warehouse to simulate it. These tests revealed technical shortcomings, so the plan wasn't executed.
>
> But would anybody today take such a vision seriously in the first place? In the 1950s, people welcomed big plans and asked whether they would work. Today a grand plan coming from a schoolteacher would be dismissed as crankery, and a long-range vision

> coming from anyone more powerful would be derided as hubris. You can still visit the Bay Model in that Sausalito warehouse, but today it’s just a tourist attraction: big plans for the future have become archaic curiosities.

This is a fascinating story (and remember that early San Francisco was settled by New England Puritans; there’s something super-Puritan about all this, right down to it being a schoolteacher) and does a great job of highlighting the contrast between past and present attitudes.

But how much of a flaw is it that the Reber Plan would in fact not have worked? Suppose a thousand enterpreneurs try to create exciting long-term plans for their businesses, each of which requires guessing ten binary variables in advance. And suppose the vague-ists are right, nobody can do armchair reasoning or long-term planning, and all of their guesses are random. By chance, one of the thousand entrepeneurs will get all ten variables right, his plan will go perfectly, and he’ll become a multi-billionaire and land a rocket on Mars. He will be the only person we ever hear about and the only person who ever becomes a stock example, and it will look like “Wow, multi-step reasoning and long-range planning can work well after all!”

This is the proper canned response that the conformist parts of my mind generated after three seconds. But is it true? Elon Musk has founded at least three super-successful companies that have executed decade-long plans; lightning shouldn’t strike the same place twice. Newton didn’t just discover gravity, he discovered optics, calculus, the laws of motion, and [insert ten page list of other things Newton discovered here].

And, uh, Thiel compares these sorts of long-term plans to “conspiracies”. And he himself is implicated in a conspiracy

– his successful destruction of Gawker is now the subject of a book titled Conspiracy: Peter Thiel, Hulk Hogan, Gawker, And The Anatomy Of Intrigue. His Gawker accomplishment was exactly the kind of ambitious long-range multi-step plan he describes as possible in the book – and the book was written long before it bore fruit.

Some of the smartest people I know say that Thiel's endorsement of Donald Trump was the same sort of complicated plot. Thiel endorsed Trump at a time when no other famous intellectual would touch him. Trump won and followed a spoils strategy of rewarding his early supporters (like how he made Jeff Sessions the Attorney General), and Thiel got 100x the influence he would have if he'd had to fight against every other important person for prestige in the Clinton administration. I am still trying to figure out what happened with this – my impression is that for about a month after Trump won, he was doing a lot of things that bore Thiel's fingerprints, and after that he didn't. Either there was some kind of early break between the two of them, or Thiel decided to operate very quietly – a few hints of leaked information suggest the latter. If he's still involved, this is an even stronger example than Gawker.

I'm bringing these things up because once you write a book saying "Hey guys, conspiracies are totally doable and often successful", and then a few years later you succeed at multiple ambitious conspiracies that destroy your enemies and give you vast national influence, I think you are allowed to say that this is possibly something other than coincidence and survivorship bias.

But then it equally becomes fair to say that Peter Thiel is a billionaire CEO Stanford professor chess master, and Elon Musk is, well, Elon Musk. Both may be better at planning than the average person. Suppose you have a ten-step plan.

And suppose you’re good enough at planning that you have a 90% chance to carry out each step. That means a 35% chance of all ten steps going without a hitch; start three companies or Gawker-destruction plans, and one will succeed. Now suppose someone only a little worse at planning – 70% success rate per step – tries the same thing. Now their per-plan chance of destroying Gawker is less than 3%. In the real world, where there’s more variance between plan steps, I think this becomes even more pronounced.

On the other hand, each successful SpaceX or Gawker-elimination-plan brings huge benefits to the world. We are stuck in the awkward position where a heuristic of “Go ahead, think big” will be inappropriate for and predictably bankrupt the vast majority of people, but a heuristic of “Think small and don’t trust yourself” will create a world of sub-par rockets and tragically un-destroyed gossip rags. Which do we choose? This is probably the wrong question; nobody controls the heuristic supply, and the one that works for most people will catch on. Under this model, Thiel is doing the public service of saying “Hey, if you’re a smart person, then despite what everybody says this whole ‘make big plans’ thing might actually work for you.”

(good thing everybody has an accurate, undistorted estimate of whether they are a smart person or not!)

I really liked this part of the book. When every intelligent person you trust is pushing one heuristic, it can be really refreshing to hear someone else intelligent and successful say exactly the opposite. Not even prove the opposite – I don’t think Thiel makes all that strong a case in this book – just say “Hey, think about the fact that this piece of conventional wisdom might be totally wrong”. This is almost the first time I’ve heard this said about the “don’t make complex multi-step plans” piece of conventional wisdom, and it was fun to hear this new perspective that I'm going to have to wrestle with from now on.

IV.

Zero To One has one more section on secrets and planning, where it expands them society-wide into the ideas of *definite optimism* vs. *indefinite optimism*.

Definite optimism is Thiel's take on the can-do philosophy of the 1950s. We can-do the Apollo Program, so let's get to work. We can-do John Reber's plan to dam the San Francisco Bay, so let's start debating it. People believed anything was possible, so they made grand plans and carried them out. Progress would happen because people would have great ideas and force them into being.

The 2010s aren't less optimistic, they're just less definite. We still believe in the impersonal force of Progress, we just doubt any existing plan's ability to serve as its avatar. Nobody will say "Let's dam San Francisco Bay", they'll say "let's promote innovation" or "let's grow the economy". It is assumed there are no secrets to discover or grand plans to implement, but everything will get better anyway based on-I don't know, some sort of principle that it should, plus millions of small actors doing little things below the threshold of notability in the right direction.

Again the weird modern belief in "the myth of the lone genius" (not belief in the lone genius, belief in the mythicism of it) comes into play. We are perhaps glad that there is convenient online retail, but this does not translate into appreciation of Jeff Bezos. Online retail came into being because it's a part of Progress; Jeff Bezos is just some annoying guy who claimed credit and captured the profits.

(the obvious counterargument here seems to be that if Jeff Bezos didn't do the admittedly hard work of creating an online retail giant, somebody else would have, perhaps a little later and a little worse; hundred billion dollar bills don't lie on the sidewalk literally forever. I'm not sure what Thiel thinks of this; at the very least he might say our society fails to appreciate that *some specific person* does have to do the work for the work to happen.)

The flagship industry of the definite optimism of the 1950s was engineering. The flagship industry of the indefinite optimism of the 2010s is finance. Finance is about "making money when you have no idea how to create wealth". While the engineers plan out specific dams and rockets and so on, the more abstract levels of finance invest in "the market", a vague aggregate of all economic activity which is expected to go up because Progress. And so:

> Think about what happens when successful enterpreneurs sell their company. What do they do with the money? In a financialized world, it unfolds like this:
>
> – The founders don't know what to do with it, so they give it to a large bank.
> – The bankers don't know what to do with it, so they diversify by spreading it across a portfolio of investors.
> – Institutional investors don't know what to do with their managed capital, so they diversify by amassing a portfolio of stocks.
> – Companies try to increas ehtie share price by generating free cash flows. If they do, they issue dividends or buy back shares and the cycle repeats.
>
> At no point does anyone in the chain know what to do with money in the real economy. But in an indefinite world, people actually *prefer* indefinite optionality;

> money is more valuable than anything you could possibly do with it. Only in a definite future is money a means to an end, not the end itself.

The flagship government of indefinite optimism is liberalism, here including both the standard Clinton-issue variety and libertarianism. Liberalism doesn't take any specific position about what the good life is, or how to promote it – it is a neutral arbiter that enforces content-independent laws. It can ban or promote the construction of monuments, but it cannot and will not say "Ten Commandments monument good, Satanist monument bad" – it either accepts or rejects both. The culmination of this style of indefinite liberalism is Rawls' veil of ignorance, where government only works insofar as it approximates what people would create if they knew nothing about their own opinions.

The flagship level of indefinite optimism is the meta-level. Come up with some principles that should work, like "capitalism" or "evolution", then let them figure everything out.

Like the section on secrets and planning, this succeeds in being an interesting critique of something I had previously thought so obviously good that I had never bothered thinking of criticisms of it before. But its specifics are a bit weird – the Burkean/Chestertonian argument for conservativism goes that our current traditions are the outcome of exactly the same sort of incremental experimentation that indefinite optimists love, and that our own multi-step reasoning and planning telling us that X new law will improve things is too fallible to trust. So if our philosophy of government isn't liberal, libertarian, or conservative, what is it? Thiel mentions two "definite optimistic" philosophers – Marx and Hegel – and neither is the sort to inspire too much confidence. Maybe we should be imagining Eisenhower-era America – liberal-ish, but still with grand visions? I don’t know enough about that era to know whether that era really had a unified version of the good life, or what shifting more in that direction would entail.

Also, different philosophies work for different situations. The virtues of feudalism are more relevant to a sprawling medieval empire than to modern Denmark. Indefinite liberalism seems suited to a country where in fact nobody agrees on anything; one with deep religious and racial divisions, caught in the grip of a smoldering culture war. If nobody can agree on what the good is, then refereeing everybody as they pursue their own private versions of the good might be the best you can maange.

V.

There’s a lot more to this book, but it all seems to be pointing at the same central, hard-to-describe idea. Something like “All progress comes from violations of the efficient market hypothesis, so you had better believe these are possible, and you had better get good at finding them.”

The book begins and ends with a celebration of contrarianism. Contrarians are the only people who will ever be able to violate the EMH. Not every weird thing nobody else is doing will earn you a billion dollars, but every billion-dollar plan has to involve a weird thing nobody else is doing.

Unfortunately, “attempt to find violations of the EMH” is *not* a weird thing nobody else is doing. Half of Silicon Valley has read *Zero To One* by now. Weirdness is anti-inductive. If everyone else knows weirdness wins, good luck being weirder than everyone else.

Thiel describes how his venture capital firm would auto-reject anyone who came in wearing a suit. He explains this was a cultural indicator: MBAs wear suits, techies dress casually, and the best tech companies are built by techies coming out of tech culture. This all seems reasonable enough.

But I have heard other people take this strategy too far. They say suit-wearers are boring conformist people who think they have to look good; T-shirt-wearers are bold contrarians who expect to be judged by their ideas alone. *Obviously* this doesn't work. *Obviously* as soon as this gets out – and it must have gotten out, I've never been within a mile of the tech industry and even I know it – every conformist putting image over substance starts wearing a t-shirt and jeans.

When everybody is already trying to be weird, who wins?

Part of the answer is must be that being weird is a skill like any other skill. Or rather, it's very easy to go to an interview with Peter Thiel wearing a clown suit, and it will certainly make you stand out. But will it be "contrarian"? Or will it just be random? Anyone can conceive of the idea of wearing a clown suit; it doesn't demonstrate anything out of the ordinary except perhaps unusual courage. The real difficulty is to be interestingly contrarian and, if possible, correct.

(I wrote that paragraph, and then I remembered that I know one person high up in Peter Thiel's organization, and he dresses like a pirate during random non-pirate-related social situations. I always assumed he didn't do this in front of Peter Thiel, but I just realized I have no evidence for that. If this advice lands you a job at Thiel Capital, please remember me after you've made your first million.)

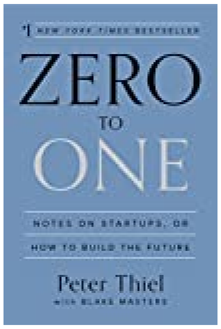
Another part of the answer must be that when everyone is competing on weirdness, the winners will be *the people who are actually weird*. The people who unavoidably do weird things because they are constitutionally weird people. There is a certain degree to which an ordinary person can relax constraints on their behavior and act and think in a weirder way than they ordinarily would. After that, you actually have to just be a strange kind of guy.

> Of the six people who started PayPal, four had built bombs in high school. Five were just 23 years old—or younger. Four of us had been born outside the United States. Three had escaped here from communist countries: Yu Pan from China, Luke Nosek from Poland, and Max Levchin from Soviet Ukraine. Building bombs was not what kids normally did in those countries at that time.
>
> The six of us could have been seen as eccentric. My first-ever conversation with Luke was about how he'd just signed up for cryonics, to be frozen upon death in hope of medical resurrection. Max claimed to be without a country and proud of it: his family was put into diplomatic limbo when the USSR collapsed while they were escaping to the U.S. Russ Simmons had escaped from a trailer park to the top math and science magnet school in Illinois. Only Ken Howery fit the stereotype of a privileged American childhood: he was PayPal's sole Eagle Scout. But Kenny's peers thought he was crazy to join the rest of us and make just one-third of the salary he had been offered by a big bank. So even he wasn't entirely normal…
>
> The lesson for business is that we need founders. If anything, we should be more tolerant of founders who

> seem strange or extreme; we need unusual individuals to lead companies beyond mere incrementalism.

Signing up for cryonics doesn’t give you a business advantage. But it indicates that you are probably good at thinking outside the box. People who learn that thinking outside the box is a useful skill and decide to try it with zero experience are always going to lose to people who have been doing since they could speak at all.

Or as a wise man once said, “when the going gets weird, the weird turn pro”.


#1 NEW YORK TIMES BESTSELLER
ZERO
TO
ONE
NOTES ON STARTUPS, OR
HOW TO BUILD THE FUTURE
Peter Thiel
with BLAKE MASTERS

Respectability Cascades

I.

I don’t know much about gay history, but the heavily mythicized version of it I heard goes like this:

At first open homosexuality was totally taboo. A few groups of respectable people with hilariously upper-class names like The Mattachine Society and The Daughters Of Bilitis quietly tried to influence elites in favor of more tolerance, using whatever backchannels elites use to influence one another. They had limited success, but they comforted themselves that at least they were presenting a likeable and respectable face for homosexuality that was improving the lifestyle’s public reputation.

Then a few totally-non-respectable outsiders with nothing to lose – addicts, drag queens, men with lots of chest hair who dressed in leather and called themselves “bears” – publicly came out as gay, held pride parades, shouted things about “WE’RE HERE, WE’RE QUEER”, et cetera. They were very easy to dislike and most people easily disliked them. But once they did this enough, people who were maybe 10% of the way to being respectable – people not addicted to *quite* so many drugs, men without *quite* so much chest hair – felt comfortable joining in. Once enough of them were out, people who were 20% of the way to being respectable felt comfortable coming out, and so on. Then 30% respectable people, then 40% respectable people, all the way up to the present day where there are a bunch of openly gay members of Congress.

I know there are lots of debates over whether this kind of “respectability cascade” is the way it really happened, but it’s a neat model of a way that these things *can* happen.

II.

And it’s especially interesting because it’s the opposite of the way I usually think about these things.

When I did pre-med in college, I learned physiology from a distinguished professor whose focus was herpetology – the study of reptiles and amphibians. His pet issue was endocrine disruption – hormone-like pollutants that were changing the sexual maturation of frogs and other animals, and which were suspected to have deleterious effects on humans. He made us read a bunch of papers on this, all of which demonstrated a clear scientific consensus that this was a well-known environmental problem and all the respectable environmentalists and herpetologists were concerned about it.

After college I went about a decade without thinking about it. Then people started making fun of Alex Jones’ CHEMICALZ R TURNING TEH FROGZ GAY!!! shtick. I innocently said that this was definitely happening and definitely deserved our concern, and discovered that this was no longer an acceptable thing to talk about in the Year Of Our Lord Two Thousand And Whatever. Okay. Lesson learned.

We can imagine a world where endocrine disruptors proceeded the same way gay rights did. A few distinguished scientists sounded the warning in acceptable elite language to other elites, but they were a lone voice crying in the wilderness. Then 0% respectable conspiracy theorists took to Twitter to make all-caps posts about TURNING TEH FROGZ GAY!!! At first they were roundly despised, but a few 10%-respectable people saw the taboo was broken and joined in, and then 20% respectable people saw the taboo had weakened even further, and so on. Finally, the cascade catches up to members of Congress, who ban the polluting chemicals. The distinguished scientists thank God for sending Alex Jones to accomplish what they could not.

But in *this* world, my impression is that the scientists were making slow-but-non-zero progress, doing really good work, and then Jones's adoption of the cause destroyed it. Now it's much harder for the scientists to convince anyone to care, because caring has become a signal that you're a conspiracy theorist or otherwise a disrespectable person. Jones hasn't just failed to contribute to the fight against endocrine disruptors, he's shot it in the foot. My professor should send him a private email asking him to shut up for the good of the cause, and to leave the issue to people who can wage it non-counterproductively, ie 100% respectable elite scientists.

This is the worldview I was trying to get across in Trump: A Setback For Trumpism. Polls show that ever since Trump entered the national stage, support for tariffs and border control have fallen. Probably this is for the same reason that concern about frog hormones would have fallen if anyone did a poll on it – the issue became associated with disreputable people, so the respectable people fled from it lest they be contaminated with low status.

It's also related to the point I make about Voat here. Reddit makes some unpopular moderation policies. It has all sorts of users, from 0% respectable racist trolls to 100% respectable academics in AskHistorians who will answer your oddly specific questions on medieval Swiss dentistry. Maybe all of them have *some* concerns about the new moderation, but the 0% respectable trolls have the most concern and are vocal in the fight against it. This leads to opposing the moderation policies getting coded as “racist troll”, and means the other discussion sites that spring up as possible alternatives are so disreputable that nobody with any kind of a reputation dares to go there. Again, the 0% respectable people taking up a theme discourage anybody else from following, lest they be associated with toxic people.

III.

So we have two opposite lessons.

In the first, 0%-respectable-people taking up a cause is a good and necessary first step, and means that soon 10% and 20% respectable people will take it up. It is the beginning of a respectability cascade that will redeem the cause from the pit of taboo-ness permanently.

In the second, 0%-respectable-people taking up a cause dooms it forever. It is the beginning of a disrespectability cascade that will make the cause too toxic for anyone above that respectability level to ever dare associate with.

So what does one do?

I’m particularly thinking here of one of my own hobbyhorses, the fight to protect scientific integrity from regressive leftism. My strategy so far has been to let Stephen Pinker and Jonathan Haidt do all the talking, and only talk myself if I feel like I can speak with the same level of dignity, respectability, and scientific backing they do. As for random people on Twitter who are likely to speak in ALL CAPS, on the rare occasions when they seek my opinion, I give them the advice of the great poet John Milton, who wrote “They also serve who only stand and wait and keep their idiot mouths shut”.

Yet I can’t help but notice that this is pursuing the same kind of strategy as the Mattachine Society and all of those other elite groups who never made more than the tiniest contribution to gay rights. “Get your most respectable members to serve as public spokespeople, and keep your least respectable members quiet so they don’t ruin your image” sounds like a good strategy. But it’s the opposite of the respectability cascade theory, and that theory is convincing.

I’ve been thinking a lot about this since that *New York Times* article on the “Intellectual Dark Web”, because those people seem like the second level of the respectability cascade. None of them are Congressmen yet, but all of them are a step beyond Milo Yiannopoulos. This has made me wonder if maybe there’s something to this after all?

Even beyond the strategic perspective, it’s just sort of embarrassing to have two good theories of how society and politics work that make opposite predictions from each other. What are some heuristics for when one would work rather than another?

Homosexuality started out as already maximally taboo; endocrine disruptors and immigration started out as merely under-discussed. Maybe disrespectable people can’t hurt an already-maximally-taboo cause, but can harm an under-discussed one?

Gay people – even 0%-respectable drug-addicted gay people – seem more sympathetic and likeable than Alex Jones or his fans, so maybe their visibility was more of a positive. But is this just me projecting my 2010s post-gay-victory values back on the past?

People leaving Reddit went to a specific alternative community – Voat – whereas people coming out as gay kept some of their existing relationships intact. Maybe socializing in a specific community made up of disrespectable people is a hard sell, but admitting to a lifestyle practiced by disrespectable people is easier?

Gay people had no choice but to be gay, whereas environmentalists (and conservatives) could pivot from caring about endocrine disruptors and immigration to some other environmental or conservative cause that might have mattered just as much to them. Maybe respectable people with lots of equally good alternatives are more likely to be repulsed by disrespectable people rather than throw their lot in with them?

But keeping all that in mind, what advice would you give Jonathan Haidt? Should he tell random disrespectable anti-SJW Twitter trolls to shut up? Or should he tell them to shout even louder? It still seems like a hard question.

Survey Results On SSRIs

SSRIs are the most widely used class of psychiatric medications, helpful for depression, anxiety, OCD, panic, PTSD, anger, and certain personality disorders (Why should the same drug treat all these things? Great question!) They've been pretty thoroughly studied, but there's still a lot we don't understand about them.

The SSC Survey is less rigorous than most existing studies, but its many questions and very high sample size provide a different tool to investigate some of these issues. I asked fifteen questions about SSRIs on the most recent survey and received answers from 2,090 people who had been on SSRIs. The sample included people on all six major SSRIs, but there were too few people on fluvoxamine (15) to have reliable results, so it was not included in most comparisons. Here's what we found:

1. Do SSRIs work?

People seem to think so:

Made me feel much worse: 6%
Made me feel slightly worse: 7.4%
No net change in how I felt: 23.7%
Made me feel slightly better: 41.4%
Made me feel much better: 21.4%

Of course, these statistics include the placebo effect and so cannot be taken entirely at face value.

2. Do some SSRIs work better than others?

I asked people to rate their experience with the medication, on a scale from 1 to 10. Here were the results:

Lexapro (356): 5.7
Zoloft (470): 5.6
Prozac (339): 5.5
Celexa (233): 5.4
Paxil (126): 4.6

Paxil differed significantly from the others; the others did not differ significantly among themselves. In a second question where participants were just asked to rate their SSRIs from -2 (“made me feel much worse”) to +2 (“made me feel much better”), the ranking was preserved, and Lexapro also separated from Celexa.

This ranking correlates at r = 0.98 (!?!) with my previous study of this taken from drugs.com ratings.

I don’t generally hear that Paxil is less effective than other SSRIs, but I have heard that it causes worse side effects. The survey question (probably wrongly) encouraged people to rate side effects as “negative efficacy”. My guess is that the difference here is mostly driven by side effects.

3. Do SSRIs work better for anxiety than for depression?

I’ve heard a few people mention this, and it makes sense as one reason why they remain so popular among patients and doctors while rarely producing large effects on specialized depression tests.

On the same scale as above:

Anxiety (391): 5.9
OCD (24): 5.8
Depression (1203): 5.3
Panic (26): 5.5
Anger (26): 5.2

There is a pretty strong effect in favor of anxiety over depression. There were not enough OCD, panic, or anger patients to get a clear picture of where those fell in relation to the other two. As far as I know this is the first study to back this claim up. But since I didn't directly ask about dose, we can't rule out that doctors give higher doses for anxiety and higher doses work better.

4. How many people experience side effects on SSRIs?

70% of people taking the drugs had at least one of the side effects on the list below:

(on this list, mild is exclusive of severe. So for example if 10% of people had mild side effects and 5% of people had severe side effects, a total of 15% of people had *at least* mild side effects)

SEXUAL DIFFICULTIES:
Severe: 11%
Mild: 41%
None: 48%

EMOTIONAL BLUNTING:
Severe: 8%
Mild: 31%
None: 61%

FATIGUE:
Severe: 6%
Mild: 18%
None: 76%

COGNITIVE DIFFICULTIES:
Severe: 3%

Mild: 16%
None: 81%

MADE DEPRESSION WORSE:
Severe: 2%
Mild: 5%
None: 93%

The more recently someone took the SSRI, the more side effects they were likely to have. While I can imagine innocent explanations for this, the most likely is recall bias: after a while, people forgot about some side effects. The real numbers are probably a little higher than this.

Most people's side effects went away quickly after stopping the SSRI, but 15% of people who stopped the medication more than five years ago said their side effects never went away. Although post-SSRI-sexual-dysfunction is *sort of* known to the psychiatric consensus, this is a shockingly high number, which doesn't seem consistent to me with how little you hear about this; I'm not sure what to think. The survey wasn't really designed to ask which side effects these were, but just eyeballing the individual entries it looks like mostly sexual, with a small amount of emotional thrown in. But these are just the two most common side effects, so it doesn't necessarily mean these two are more persistent than others.

5. Do some SSRIs produce worse side effects than others?

I think the psychiatric consensus on this question is that Paxil has worse side effects than the others, which are all equal.

This survey failed to directly replicate that. Four of the five side effects elicited (sexual difficulties, fatigue, emotional blunting, cognitive problems) were the same across all drugs. The only one that differed was worsened depression, which was slightly less common on Zoloft. This was technically significant but given the number of tests I would not put too much stock in this.

I forgot to ask about a few important side effects, including weight gain. I suspect that Paxil scored worst on the “overall” category because it produced worse side effects in the categories I forgot to ask about, or because people remembered it had side effects but couldn’t remember exactly what they were. Overall this survey doesn’t really make me doubt the consensus that it is probably worst.

6. How many people have trouble discontinuing SSRIs?

Good question! This has been a topic of interminable debate in the medical community, with some saying these problems are very common and others saying they are very rare. This survey found:

59% don’t remember having any issues at all
22% remember having a few minimal issues but not really thinking about them
14% remember having moderate issues that caused significant distress but were not disabling
5% remember having severe issues that seriously impacted quality of life

6. What factors made SSRI discontinuation easier or harder?

Hard to tell.

Everyone believes that a more gradual taper makes things easier, but the survey quite clearly found that people who reported longer tapers had worse problems. I’m pretty sure this is because if their doctor expected them to have problems (or they started it and did have problems) they put them on a longer taper. But this kind of thing makes it hard to make any real recommendations. However, people who came off their medication accidentally because they ran out did have by far the worst time, suggesting that cold turkey discontinuation really isn’t the way to go.

The longer a person had been on SSRIs, the harder their taper was likely to be. However, I’ve heard some people give overly dire warnings like “If you’re on these drugs for more than five years, don’t try coming off”. These were not justified. Even among people who were on the medication over five years, 49% tapered with “no issues”, and only 15% reported severe issues.

People with anxiety and OCD reported more difficult discontinuation than people with depression. This could be either because these conditions require higher doses, or because people with anxiety and OCD are more likely to notice and worry about minor symptoms.

Psychiatric consensus says that Paxil is the hardest SSRI to get off, and Prozac is the easiest. This survey confirmed that result. On a scale from 0 – 10, where 0 is the easiest discontinuation and 10 is the most difficult:

Paxil: 2.8
Lexapro: 2.7
Zoloft: 2.2
Celexa: 2.1
Prozac: 1.5

Prozac separated from Zoloft, Lexapro, and Paxil, but not from Celexa. The average person had only a 59% chance of having no discontinuation symptoms; the average person on Prozac had a 71% chance.

In any case, almost everybody’s taper was successful *eventually*. Only 0.5% of people said they gave up and stayed on the SSRI because they found discontinuation too difficult.

Summary

This survey had few surprises.

Already when giving someone an SSRI, I debate between Lexapro and Prozac. Lexapro is usually the most effective (by a tiny hair), but Prozac is the least likely to cause discontinuation syndrome. Although a natural strategy might be to taper Lexapro (or something else) very slowly in order to match Prozac’s slow half-life, studies show this doesn’t work (why not?), and this survey confirms it. There is no obvious right answer between these two as first-choice SSRI. This study does confirm my prejudice that giving Paxil is an obvious wrong answer and you should never do it outside specific rare circumstances.

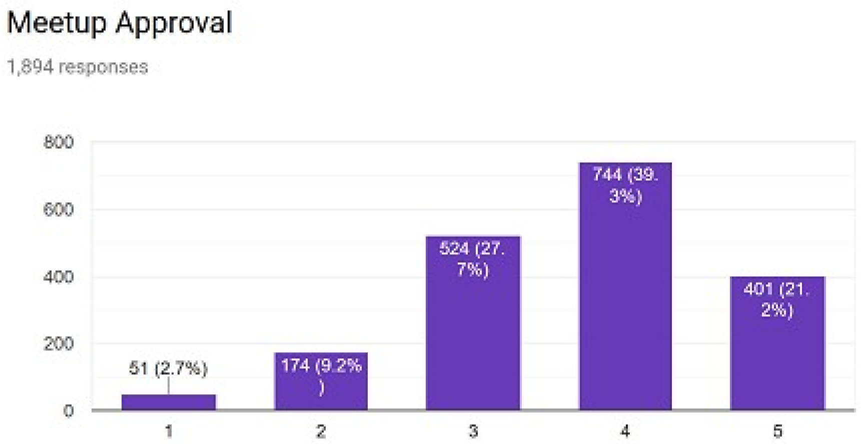
The main surprise was the high number of people who claim their SSRI side effects never went away. Although this is a known very rare possibility, the survey suggested it was much less rare. One can imagine innocent ways this could happen: for example, someone goes on SSRIs for ten years, comes off, and is surprised to find their sex drive is lower than it was when they were a teenager ten years ago. This probably requires more careful and rigorous study than can be done in a silly online survey – and so will probably never happen.

Irvine Meetup This Saturday

Where: Underneath this mysterious hexagonal sigil at the University Center food court in Irvine, California. People from LA, San Diego, etc welcome!

When: 8:30 PM on Saturday, February 9th. Sorry for the late hour, my schedule is complicated and this was the only way to make it work.

Who: Anyone who wants. Please feel free to come even if you feel awkward about it, even if you're not "the typical SSC reader", even if you're worried people won't like you, etc. People who fit those descriptions who decided to come to previous meetups have mostly enjoyed them.



~~You won't regret it~~ Only 11.9% of you will regret it, with another 27.7% having no strong opinion either way!

The Proverbial Murder Mystery

I.

Chefs. Hundreds of them. Tall chefs, short chefs, black chefs, white chefs. I pushed forward through them, like an explorer hacking away at undergrowth. They muttered curses at me, but I was stronger than they were. I came to a door. I opened it. Sweet empty space. I shut the door behind me, sat down in the chair.

“Hello,” I said. “Detective Paul Eastman, pleased to make your acquaintance.”

“Doctor Zachary LaShay,” said the man behind the desk. His little remaining hair was greying; his eyes showed hints of the intellect that had been buried beneath the dullness of an administrative career. “I hope you didn’t have any trouble getting here. Did my secretary warn you about the chefs?”

“She did not,” I said.

“Well, forewarned is forearmed,” he answered, inanely and incongruously. “But I trust you got my message about the federal investigators?”

“Once a federal investigation has started, we’ll retreat and let them take over. But two women died here. We can’t just not investigate because you tell us you’re trying to get the Feds involved.”

“Yes, ah, of course. It’s just that we’re a sort of, ah, defense contractor. None of our projects are officially classified, yet, but we were hoping to get someone with a security clearance, in case this touched on sensitive areas.”

“I won’t pry further than I have to, but until someone from the government says something official, this is a matter for city police. Maybe you could start by telling me more about exactly what you do here.”

“We’re the United States’ only proverb laboratory. Our mission is to stress-test the nation’s proverbs. To provide rigorous backing for the good ones, and weed out the bad ones.”

“I’d never even heard of your organization before today, I have to admit. And now that I’m here…it’s huge! Who pays for all of this?”

“Everybody who uses proverbs,” said the Doctor, “which is to say, everybody. Consider: he who hesitates is lost. But also: look before you leap. Suppose you’re a business executive who spots a time-limited opportunity. What do you do? Hesitate? Or leap without looking? Eggheads devise all sorts of fancy rules about timing the market and relying on studies, but when push comes to shove most people are going to rely on the simple sayings they learned as a child. If you can keep your stock of proverbs more up-to-date than your competitor’s, that gives you a big business advantage.”

A smartly-dressed woman came in, handed Dr. LaShay a cup of boiling liquid. He put it to his lips, then spat. “This is terrible!” he said. “Try it!”

I had been expecting it to be tea, but it wasn’t. I didn’t know what it was. But it was terrible. Somehow too plain, too salty, and too bitter all at once. I gagged.

“That settles it!” said the Doctor. “Too many cooks really do spoil the broth. Tricia, tell the chefs they can all go home now.”

“So that’s what you were doing!” I said.

“Yes. Until now, too many cooks spoiling the broth had been at best an anecdote! A folk hypothesis! This month we’ve been working on broth with varying numbers of cooks. One, two, five, ten, a hundred. We’ve got a team of blinded taste testers in the basement who’ve been rating the results, and I personally check each sample to make sure I agree. This morning we hired every cook in the city – that’s over five hundred cooks – to come here and make broth for us, just to make sure there isn’t some kind of island of stability where broth starts getting better again once the number of cooks is high enough. Later this week we’ll give the data over to our analysts, who’ll develop a model that can use cook number to predict broth quality over a wide range of possible situations.”

“And the military wants this sort of thing?”

“The military loves it! The average grunt is a high-school educated young man in his late teens or early twenties. You’re not going to be teaching these people Clausewitz and von Moltke; it would be casting pearls before swine. When he’s under fire and has to make a split-second decision, he’s going to rely on the heuristics he learned on his grandmother’s knee. On proverbs. America’s proverbs are a vital strategic asset, and the Pentagon appreciates that.”

“I get how *too many cooks spoil the broth* might apply to something like an officer trying to figure out how many people to consult about a new strategy. But surely you can’t test that heuristic just by experimenting with literal cooks making literal broth!”

“Mmmmmmm. Yes, you’re referring to what we call Pragmatics. We certainly have a pragmatics team here, and they do good work. But the thing is, Officer, we're essentially a consulting firm. Consulting firms are there to give people justification for the things they want to do anyway. When some general is testifying before Congress, and he says he didn't consult someone-or-other because too many cooks spoil the broth, then Congress is going to want evidence that relying on sayings like this is best practice. If he just says "That's our heuristic, and we know it works", he'll look like a loose cannon. But if he can hold up a glossy five hundred page report we gave him, proving that broth really does get spoiled by too many cooks, he'll look like a responsible technocrat who did his due diligence. And yes, part of that report is a long philosophical discussion on pragmatics. But part of it is proving, once and for all, that too many cooks really do spoil the broth."

"I see," I said. "The two dead women. Were they involved in the broth project?"

"No. The first victim, Lisa Bird, she was our sysadmin. The second victim, Catherine Lee, took care of the animals."

"Animals?"

"We have several projects that require animals. You can obviously lead a horse to water, but can you make him drink? At first we would rent out horses from equestrian organizations for this kind of thing. But then the next month we would need another horse to see if you should shut the stable door after the horse has bolted. Then we'd need two more horses to see if you should change horses midstream. Finally the costs started adding up and we just got a couple of horses that we keep here at the Institute. They were actually a gift from a sister of one of our employees who used to have a farm. One of them we looked in the mouth;

the other we didn't. We're still trying to figure out which way worked better."

"I see. The report I got said that the motive was romantic jealousy."

"Yes. Ms. Lee believed Ms. Bird was having an affair with her husband. Ms. Bird was known to come to work early on Fridays to do some extra work and prepare for the weekend off. Ms. Lee entered the office where Ms. Bird was working alone, murdered her, then committed suicide. I'm getting this from the emergency team that was here before you."

"All right. I'll need to see the crime scene."

"Certainly."

LaShay led me out of his office to an elevator, then hit the button for the tenth floor. We walked out into a clinically-clean hallway. I heard a commotion. "FUCK YOU!" someone was shouting. "DAMN YOU TO HELL, YOU INKY TENEBROUS MOTHERFUCKER!" I stepped forward to open the door and investigate, but the Doctor held me back.

"Don't worry about it," he said. "That's Room 27A. We're testing whether it's better to light a candle or curse the darkness. The candle is in Room 27B."

"You must have a lot of projects going on here."

"Oh yes. Over there is our insect unit. Can you catch more flies with honey or vinegar, can ants really move plants, that kind of thing. Our kitchen is to the right – the chefs were using it today, but it comes in handy all the time. Just don't go in there if you can't stand the heat. And down that corridor are our weather unit, our fire unit, and our water unit. And that's just the tip of the iceberg – " He pointed to a large room with a spike of ice poking through the floor. We continued on. “And over there is our forge. There are so many proverbs about metal that we hired our own team of blacksmiths. It was going great until they unionized, but now they always strike when the iron is hot.”

The corridor opened into a vast auditorium. All around me, I saw knee-high marble buildings, gleaming palaces, and – was that the Colosseum? A man dressed in a gladiator costume was sitting behind a desk doubling as a terraced hill, frowning at a computer and occasionally typing something. “Our 1:100 scale model of Rome,” said LaShay. “We figured if we couldn’t build it in 0.24 hours or less, then Rome couldn’t be built in a day. For some reason I always get lost and end up here. It’s quite annoying.”

We passed out of Rome into another corridor, where we finally came to a door marked “Information Technology”.

“Ms. Bird’s office,” said LaShay, and I walked in.

I’m a homicide detective; I’m used to grisly murder scenes. This one still made me gasp. One victim – Ms. Bird, I supposed – was lying on the ground by the desk. It looked like her head had been bashed in by a blunt object. But there was more. Her mouth area was covered with blood, and I soon found her tongue had been cut out. And there was another bloody hole in her chest. The stomach and heart had been cut out too.

A few feet away, a second body dangled from a noose that had been tied to one of the rafters. Ms. Lee, I supposed. No mutilation on this one. Just a clean suicide, or at least that was what somebody had gone through a lot of trouble to make it look like.

Lying on the ground approximately between the two of them was a bloody knife. I knew from the previous report that the blood was Ms. Bird’s, and the fingerprints on the handle were Ms. Lee’s.

This did not seem like a Sherlock Holmes level mystery. Except: where were Bird’s heart and tongue?

“I’m not sure,” said LaShay, when I went outside and asked him the question. “I…haven’t been in there since in happened. Not sure I could deal with the blood. One of Ms. Bird’s coworkers had a question about the network, so she went in and saw…what you just saw. We called 911 in case either of them was still alive. The paramedics called the police who did a preliminary investigation of the scene. And then you showed up.”

“I’ll need to search the premises,” I said. “What time did Bird come to work?”

“I understand she usually arrives around seven.”

“And when does the office open?”

“Nine.”

“So potentially Lee could have had two hours to hide the heart and tongue somewhere in this building before going back and hanging herself.”

“Why would she have done that?”

“I don’t know. Do you have a better idea for what happened to them?”

He shook his head.

“Good. Then I’ll need to search the whole building. Is there anywhere I’ll need any special keys or codes to enter?”

He gave me a golden key. “This opens any door,” he said. “But don’t go in the Red Zone. That’s off-limits to everybody.”

I shrugged. “Then it’s exactly the sort of place somebody would hide something, isn’t it? Why isn’t anyone allowed in the Red Zone?”

“Radioactivity,” he answered immediately. “We have a giant machine for testing all machine-related proverbs. It’s…very impressive. Powers the whole building, runs the water and gas systems, even gives us satellite internet. We wanted it to be just a generic Machine, capital m, so it does a little of everything. But it’s radioactive…not traditionally, the way you can detect with a Geiger counter. I don’t understand the physics. But people tend to get very sick if they get too close to it.”

Part of LaShay’s description had stuck with me. “It provides the building with Internet? Lisa’s a sysadmin. Did she ever have to work with the Machine?”

“No, that was all connected when the Machine was installed. She interfaces with it remotely, through her computer.”

“And Catherine? Did her work with the animals ever bring her near the Machine?”

“Her office was very close to the Red Zone. Closer than any other office in the building, actually. But she never had any reason to enter the danger area.”

“I’m going to need to see the Machine. Is there any way I can do so safely?”

“We have an observation deck. It’s just above the Machine, on this floor. You can stare down at the Machine from the top.”

“I’ll need to go there.” It was just a hunch, but I wasn’t liking the sound of this Machine. And if you were going to hide body parts for some reason, why not hide them in a restricted area where nobody ever went?”

LaShay took me down a series of turns and hallways. After a minute or two of walking…we were in the scale model of Rome again.

“Dammit!” said LaShay. “Every time!”

Another few turns and hallways, and we finally came to a steel-reinforced door. “DANGER” it said. “OBSERVATION AREA. CHILDREN AND PREGNANT WOMEN NOT ALLOWED. PLEASE DO NOT SPEND MORE THAN FIVE MINUTES IN THE OBSERVATION AREA OVER A ONE WEEK PERIOD.”

I used my golden key to unlock the door. We went in.

We were on iron scaffolding. Below us whirred something amazing. It was like every children’s-book description of a machine put together and brought to life, a huge assembly of gears and pistons and bubbling glowing bright-colored chemicals coursing through glass pipes. Beside me was a control panel, currently set at “NORMAL”. The other options ranged from “OFF” to “MAXIMUM” to “ULTRAMAXIMUM” to “SUPRAULTRAMAXIMUM”.

“It’s beautiful,” I told the Doctor.

“Don’t touch that,” he told me, glancing nervously at the control panel.

The machine was nine stories high, filling the entire center of the laboratory. In the center, an enormous agglomeration of steampunk-looking gadgetry formed a hollow cylinder, spinning faster than I could follow. I leaned out over the edge of the scaffold, over the pit formed by the cylinder’s center.

“You really don’t want to do that,” LaShay told me. I could see what he meant. It was easy to imagine falling right through the hole in the spinning cylinder, down to the ground ten stories below. I had a strange feeling that gravity would be the least of my problems if that happened, that anything that went through that spinning apparatus would have a very bad time long before it hit the ground. And…

“What’s that?” I asked.

At the bottom of the spinning cylinder, incongruously, was a building I could only describe as a small shrine. It had a little golden dome on the top, and…actually, it was exactly a shrine. There was a Star of David atop the dome.

“That,” said LaShay. His voice changed, became heavier. “I started this laboratory with my colleague, Dr. Rissum. He…he committed suicide nine years ago by jumping into the Machine from this very spot. That’s his memorial.”

“My God! You’re telling me there was another suicide in this lab?”

“Nine years ago. The police investigated. There was nothing suspicious. His wife had just left him and taken the children. It was very tragic, but no foul play was suspected.”

“Still. Another suicide.”

“We need to get out of here,” said LaShay. “Being this close to the Machine really isn’t good for you.”

I looked around the observation deck and at the floor ten stories below. There were no signs of blood, a tongue, or a heart. “All right,” I said, because the Machine was starting make me nervous too.

I spent the rest of the morning searching the rest of the laboratory, free of LaShay’s discomfiting presence. It was an exhausting task, not least because I always ended up in the Rome model even when I thought I was in a totally different part of the building. But eventually I found two things that caught my interest.

First, Lisa Bird’s chair. I had gone back into the room with the bodies to look for other clues. The desk was normal enough. The computer was a normal Apple MacBook. But I noticed Lisa’s chair was made out of human hands. This was confusing enough that I called the Doctor back, who of course had an explanation.

“They’re not real hands,” he said. “Most of the staff have chairs like that. We were testing whether many hands make light work, so we had everyone working for the lab sit on those.”

“It’s pretty gruesome,” I said.

“We originally tried putting those statues of the Buddhist god with the thousands of hands all around the office,” LaShay admitted. “But people complained that the hands were whispering demonic messages to them. Finally someone in the Religion Department reminded me that idol hands are the Devil’s plaything.”

“Okay,” I said, and dismissed LaShay again, with some relief. He told me he would be working over the weekend, and said I could call him if anything came up. I hoped I wouldn’t have to. Something was weird about that guy, no doubt.

The second thing I found was Lisa Bird’s tongue and stomach. It was in the third drawer of Catherine Lee’s desk. The woman had murdered her coworker, cut out her tongue and stomach, put it in the third drawer of her desk, gone back up to the murder scene, and committed suicide.

Or, more accurately, this was a subset of what she had done, because I still couldn’t find Lisa’s heart. I searched Catherine’s desk inside and out. All I could find were a couple of paperweights made of various gemstones. I noticed they were about the right size and shape to have made the dent in Lisa Bird’s head, but none of them had any bloodstains on them or anything else suspicious. There were no severed organs.

I was missing something. But what?

II.

“You’re the detective on the Bird case?”

“Mmmrrrgyeah,” I answered groggily.

“Come to the station,” said Officer Karp. “The murderer’s body is missing.”

It was 8 AM on Saturday. I had visited the Proverb Laboratory Friday, told the station that the scene had been fully examined and they could take the bodies away, then gone home and slept. The station had sent a team to recover the bodies and bring them to the morgue. The next morning, one of the morgue staff had noticed that although Lisa Bird was still there, Lee’s body was missing.

Still only half-awake, I went to the morgue and examined the scene. The body bag was still in place. It had been expertly opened up and the body had been removed. There were no fingerprints. Karp was seething that a theft had been committed in the police station itself. He demanded we do something. I suggested we go to Catherine Lee’s house, interview her husband, see what he could tell us. That was how I ended up spending my Saturday morning at the weirdest house I had ever seen.

It was some kind of modernist experimental dwelling or something. The whole place was made out of windows. Not one-way windows either. You could see everything that happened in it. Not (I thought to myself) the sort of place a criminal would find very convenient.

“It was Cat’s idea,” her husband told us, when we knocked on the door and introduced ourselves. “She was always so paranoid that I was having an affair. Well, some weird architect made this house and then put it on the market – obviously nobody wanted it, so the price was right. Cat thought it was perfect. I couldn’t hide anything here. You’ve got to believe me, officers. I never had an affair with anybody. She was paranoid. But not violent. I know they say she killed that woman. But she would never do something like that. She was framed. I’m sure of it.”

“Who would do such a thing?”

“She talked about office politics all the time. I know things I’m not supposed to know. The Proverb Laboratory, they talk about selling their work to corporations, but the US military is the big sponsor. A lot of their best work is hush-hush.”

“I’m aware,” I said.

“Well, she would tell me all these rumors. Apparently the British hate the Proverb Laboratory. Before LaShay and Rissum started it ten years ago, the British had a monopoly on English-language proverbs. You’d have all these proverbs about kings and queens and tea and castles. It was a way for them to maintain their cultural hegemony over us. That’s what Cat would say.”

“Was Catherine by any chance paranoid and delusional about British people?”

“She was paranoid and delusional about a lot of things, but I tell you, she wasn’t a killer.”

“Were there any specific British people? Or anyone else who didn’t like what the Proverb Laboratory was doing?”

“There was the English Defense League. Have you ever heard of them?”

“They’re some kind of white supremacist group, right?”

“You must be thinking of the White Defense League. The English Defense League are an English supremacist group. As in, the English language. They believe English is superior to all other languages. They want to stop foreign language education in school, kick foreign speakers out of the country, make English the official national language, that kind of thing.”

“And they’re against the Proverb Laboratory?”

Mr. Lee laughed. “Or else they *are* the Proverb Laboratory. You know LaShay used to be one of them? No, from the look on your face you didn’t. He was part of their cult for a while, then deconverted and went mainstream, spoke out against them for the press. But some people say that’s all a ruse, and he’s continuing their work. They always thought that with enough study, they could use create some kind of super-proverb that would encapsulate all wisdom and make them unstoppable, something like that. LaShay says he’s beyond all that, but who knows? And if he is, well, maybe the cult that he left isn’t so happy to have the US military meddling in their pet project?”

“That’s so weird. I never heard about them before.”

“Well, Cat heard a lot of things, working at the Proverb Lab for five years.”

“Did she like it there?”

“Oh no. She hated it. She loves animals, you know. But the Proverb people thought they were just means to an end. She was in a big fight with LaShay just before she died. He wanted to test the proverb ‘Every dog has its day’. He was going to lock up forty, fifty dogs in a dark room, to simulate night, and just *leave* them there. Wanted to “falsify the hypothesis”. Cat said absolutely not, that was animal cruelty. So he did it anyway without telling her. She was enraged.”

“Did she ever make any threats? Say she was going to blow the whistle on the lab or anything?”

“No, nothing like that. She said she was going to let sleeping dogs lie. Sorry. I don’t think she had any enemies. She could be paranoid, she could be strange, but she was a good person, deep down. She wouldn’t have done this.”

“What’s that?” Officer Karp interrupted.

He was pointing to a corner of the kitchen. At first I didn't see it. Then I did. There was a little drop of blood on the floor.

"Mr. Lee, do we have your full permission to search this house?" I asked. Officer Karp was already calling the station, letting them know they were going to need to send out an evidence collection team.

"Of course," said Mr. Lee. "I have nothing to hide."

Officer Karp went to the cabinet just next to the bloodstain, reached in, and pulled out a human heart.

"I…I swear I have no idea how that got there," said Mr. Lee.

"How late did you sleep yesterday morning, when the murder happened?" I asked.

"I…it was my day off. I slept until ten."

"And your house is about a fifteen minute drive from the lab. So in theory, your wife could have killed Ms. Bird, left the Proverb Laboratory, come back home, hid the heart in your cupboard, then gone back to the Proverb Laboratory and hung herself, all before anyone else showed up for work at nine."

"Why…why would Cat have done that?" pled Mr. Lee.

"I don't know," I said. "Did she have any motive for disliking Ms. Bird other than the affair issue? Anything at all?"

"Nothing," said her husband. "She spoke very highly of Ms. Lee. Apparently her computer had a virus once, and Ms. Bird solved it. She'd gotten a degree in cybersecurity from MIT before ending up in this job, and she was always working hard to keep the servers safe.”

“One more question. Do you know who stole your wife’s body from the morgue?”

“What?” asked Mr. Lee. “Someone stole…”

“This guy’s as surprised as we are,” said Officer Karp. “I say he’s not a suspect.”

We drove back to the station in silence. Either Catherine Lee had murdered her coworker, driven home to hide her heart in a cabinet, and then gone back to work before killing herself – or somebody had put a lot of work into making it look that way. And somebody had stolen her body from the morgue. And there was some sort of web of international intrigue surrounding Doctor LaShay.

I decided I was going to go home, catch up on my sleep, and then think this over *really hard*.

III.

Sunday morning I walked back into the Proverb Laboratory. I was trying to get to Dr. LaShay’s office, but I had ended up in the scale model of Rome again. I hadn’t even taken an elevator, and it was on the tenth floor. That no longer confused me. I had finally figured out what I should have realized days earlier.

“Dammit!” said LaShay, almost bumping into me. “Rome again!”

“Doctor Zachary LaShay,” I said, “You are under arrest, for the murders of Ms. Lisa Bird and Catherine Lee. You have…”

“You can’t arrest me!” he said.

“…the right to remain silent,” I continued. “Anything you say can and…”

Two men in black uniforms and sunglasses stumbled into the Rome set just behind him.

“No,” said LaShay. “I mean you can’t arrest me. The federal government has taken over the investigation, as of today. The entire affair has been classified as top secret. You’re not even allowed to be here anymore.”

I sighed. “Then I’ll just take a moment to talk with one of these agents…”

The agents didn’t move.

“You have one minute to get off this property,” said Dr. LaShay, “or you will be in violation of federal law.”

“All right,” I told the agents. “Listen up.” Then I explained everything.

The Proverb Laboratory didn’t exist to test proverbs at all. Or they did, but not in the way they claimed. The Proverb Laboratory existed to test the Machine. A device that makes proverbs real. The Machine exerted some kind of invisible force. The closer you got, the more the English language warped reality in order to make proverbs come true.

Why had Lisa Bird’s tongue and heart been missing? Because the proverb goes “Cat got your tongue”. The Machine’s power had forced Cat to take Lisa’s tongue and bring it somewhere that would qualify as her “having” it. And the same force had made her bring the heart home, because “Home is where the heart is”. She hadn’t meant to take the stomach too, but had removed it for better access, since “The way to a man’s heart is through his stomach”. Then her corpse, which had spent years absorbing the Machine’s malevolent radiation, had vanished from the body bag where it was kept – “Cat’s out of the bag”.

What had Catherine used to bash Lisa’s head in? The obvious candidate was one of the gemstone paperweights hidden in her desk, which she had brought back at the same time as she brought the tongue. I hadn’t been able to find bloodstains on any of the paperweights, but that was unsurprising; “You can’t get blood from a stone”. She lived in a glass house, and had broken the rule about throwing stones, and so ended up dead and a murderer. The saying goes: “Kill two birds with one stone”. Catherine had killed Lisa Bird; where was the other? Simple. Lisa sat on a chair made of hands, and a bird in the hand is worth two in the bush. She was worth two birds all on her own.

But it was too perfect. How had it all come together? A paranoid lady who thought everyone was having an affair with her husband. Who lived in a glass house and owned gemstone paperweights. Sharing a building with a woman named Bird. Who was sitting on a chair made of hands. In the closest office to the machine that made proverbs true. This wasn’t a coincidence. This was planned. Someone must have arranged for a paranoid woman who lived in a glass house to be on the spot, given her the stone paperweights as presents, placed Bird on the hand-chair, then relied on the Machine to twist reality into committing the crime for him. They must have guessed that after it was all over, Lee would recover her senses, feel terrible guilt, and kill herself. Who could have done that? LaShay was the only person powerful enough to make it all happen.

LaShay was lying about the memorial to Rissum. They hadn’t built a temple on the spot where Rissum died. That temple was Rissum himself. He had fallen into the very center of the Machine, where the reality-bending force approached infinity and proverbs would come true no matter how unlikely. “My body is a temple”. Rissum’s body was transformed into a temple in mid-air, then fell onto the ground below. Why would LaShay hide this? Could it be because he had pushed Rissum into the machine himself to seize complete control over the operation?

But why? The rumor Mr. Lee had told me tied everything together. Dr. LaShay was still with the English Defense League. They had designed the Machine. He had pretended to go mainstream, pretended to partner with Dr. Rissum, in order to get enough money and status to build their invention. Now he was slowly testing its capacities, secretly funneling the results to his secretive language-cult. Rissum had been a convenient co-founder, but had to go in order to give LaShay full control. He had pushed him into the Machine, disguised it as a suicide, and was funneling the information – how?

Through a worm in the computer system. After all, the workers here all had Apple computers, and every apple has its worm. But LaShay hadn’t realized that along with her sysadmin work, Lisa was an expert in cybersecurity, nor that she would come in two hours early every Friday. “The early Bird catches the worm.” Lisa had found the infection and destroyed it. She hadn’t realized it was important, but LaShay realized he couldn’t reinfect the system without her finding it again and getting suspicious, and he couldn’t fire her without raising eyebrows. So instead, he had arranged matters perfectly to guarantee she would get killed.

“Wow,” said Dr. LaShay after a second. “You’re actually right about everything. Except for one thing. The most important thing.”

“What’s that?” I asked.

“Not real federal agents,” he said, gesturing at the men in black. “They’re with me.” He turned to them. “Throw him in the Machine.”

I reached for my gun, but the agents were faster than I was, wrestled it away from me. Then one of them held each of my arms and started dragging me to the observation deck. A slight delay as we ended up back in Rome. Then we were there, and I was standing over the great rotating cylinder, staring at the shrine of Dr. Rissum below.

“Please don’t let me die,” I said. “I’m begging you. Please spare my life.”

“You really think we care about that?” asked the first agent.

They pushed me to the edge of the scaffold.

“You really think I was begging because I thought you’d listen?” I said, but before I finished he had thrown me over. There was a gust of wind and a feeling of terrible wrongness.

When I had fallen five stories, into the very center of the Machine, I *wished*.

A flying horse was somewhat outside the scope of the relevant proverb, but there was no other way I was going to “ride” while in midair, so I got one. It made landfall right on the observation scaffold, then rushed for the door. The two agents rushed after it. Somewhere in the corridor, the horse dissolved, its Machine-powered existence apparently expending itself this far from the source.

I ran frantically through the corridor. “After him, you fools!” I heard LaShay shout. I reached the point where I thought the elevator should be, but of course I was in fricking Rome again.

One of the agents ran in, reached for his gun.

I ducked behind the terraced hill. There beside the desk was the gladiator costume, complete with weapons. I picked up a trident. “*Ave Imperator!*” I said. “*Morituri te salutant!*” Like a miracle, it worked. The agent aimed at me and pulled the trigger, but the gun blew up in his face. This close to the Machine, he should have known: “When in Rome, do as the Romans do.”

The agent was still on his feet. I had made the mistake of getting far enough from the hill-desk that the agent could pick up the abandoned sword. He rushed at me. I didn’t know how to swordfight, so after a second of thought I took a pen out of my pocket, parried with it. The sword shattered, and ink squirted out into the agent’s face.

While he was trying to wipe off the ink and get his vision back, I ran out of Rome into one of the nearby corridors, then ducked into a randomly chosen door. Everything was pitch black.

“Come out, come out, wherever you are,” the agent shouted. “You can run, but you can’t hide!” Frick. I had forgotten that. In this place, the saying itself probably made that literally true. I heard the two agents opening and closing all the other doors in the corridor, getting inevitably closer to me.

Then I felt something cold and wet press against my hand. I almost screamed, giving away my location, but after a second it…licked me. I remembered what Mr. Lee had told me. Dr. LaShay had stuck fifty dogs in a completely dark room to test a proverb. I felt around. More and more dogs started to trot up to me, mouths panting, tails wagging. I had one chance.

I flung the door open as hard as I could “Run away, doggos!” I shouted. “Run like the wind! This is it! THIS IS YOUR DAY!”

The dogs didn’t need to be told twice. They rushed out of the room, a yapping growling barking mass of teeth and fur. Big dogs, little dogs, old dogs, young dogs, the whole mass of dogs ran right into the agents, knocked them over.

“Call off your dogs!” one of the agents shouted, but I didn’t. Instead, I cried “Havoc!”, and let loose the dogs of war. I figured their bark would be worse than their bite; on the other hand, once bitten, twice shy. It probably balanced out. Hopefully I wouldn’t have to worry about the agents for a few minutes.

I ran to where I thought the elevator would be, and *of fricking course* landed in Rome again. And worse, there was the Doctor, who was holding the trident I had abandoned. The sword was nowhere to be seen. I knew I wouldn’t be able to fool *him*. He had probably forgotten more proverbs than I had ever learned.

“*Ave Imperator!*,” said Dr. LaShay, approaching unarmed-me with his trident. “*Morituri te salutant.*” Even his *Latin* was better than mine. I wished I was first in a village. But hope beyond hope, I realized that the computer at the terraced-hill desk was an Apple. I grabbed it, pulled out the plug, brandished it before me. The Doctor staggered back, as if kept away by an invisible wall.

But it didn’t hold him for long. He stretched out his arm as far as it could go, lunged at the computer with the deadly trident. The screen shattered and went black, its power lost.

I ran through the maze of corridors, and LaShay followed, trident in hand. After several turns, I reached where I thought the elevator would be, but Rome was everywhere at once, and I had lost my bearings. I ended up in the Observation Room, standing on the iron scaffold above the machine, as LaShay and his trident came towards me.

“So,” he said, “you figured out a way around being thrown into the Machine. ‘If wishes were horses, beggars would ride.’ Clever. You could have been a great proverb researcher. But instead you had to meddle where you didn’t belong.”

“If you throw me into the pit, I’ll just get another flying horse,” I told him.

“Of course you will. So I’ll have to kill you with the trident.” I was backed up against the wall of the observation chamber. LaShay approached me confidently, knowing I was cornered.

“You really think you’re going to win this?” I asked. It wasn’t just to buy time. I really did have a plan, crazy as it was, but the more I could get him gloating, the better it would work.

“Of course,” said LaShay. “I killed Bird and Lee, and now I’m going to kill you. Your death here will actually be quite convenient. I’ll announce that the Machine is too dangerous and needs to be taken apart. Then the version the English Defense League is building in secret will be the only one in the world. With the data we’ve gathered here, they’ll be able to direct its power anywhere on the planet. Imagine what we'll be able to do. Enlist old soldiers who are impossible to kill. Build fortresses on demand by turning arbitrary Englishmen's homes into castles. Control the seas using loose lips. Soon English-speakers will rule the world. And nothing – absolutely nothing – can stop us!"

Perfect.

"You've forgotten three things," I said. "First, that the lever is right here."

I grabbed the lever on the control panel and jerked it to SUPRAULTRAMAXIMUM. The air started to shimmer, and the walls started to shake.

"Second, that pride cometh before a fall."

The iron scaffolding started to tilt. LaShay stumbled, dropped his trident, almost tumbled over the edge, hung on just by the tips of his fingers.

"And third, that *crime doesn't pay!*"

I grabbed the pointy end of the trident, and smashed it into LaShay's fingers. With a scream, he fell into the belly of the Machine.

"Ibegyounottodothis," he said, and just like that he was on a winged horse. It flew up, towards the door and freedom.

I looked it in the mouth. I stared it straight in the mouth, looked as hard as I could, like my eyes were drilling into it. It started flickering, flying more slowly and hesitantly. "I beg, I beg, I beg," said LaShay. We stood there like that for a few seconds, him trying to wish harder, me trying to look the gift horse in the mouth harder, until finally the horse vanished, and LaShay fell back into the machine.

“I beg, I beg, I beg!” he said again, there appeared another horse, a horse of a different color. I looked it in the mouth again. It rose more slowly and hesitantly. But LaShay leaned forward, finally covered its mouth with his hand so I couldn’t see it. “Your looking has no power anymore!” LaShay said triumphantly, and I believed him, since it came straight from the horse’s mouth. The impediment removed, the horse shot upwards, right up to the ceiling of the chamber.

“Get off your high horse,” I said, and the horse vanished a second time. A third time LaShay fell into the Machine, a third time he begged, and a third time a horse appeared beneath him. Again I started looking it in the mouth. Again he covered it with his hand, this time guiding the horse more slowly, trying not to let it overshoot and become higher than I was.

With a whinny of victory, the horse’s hoof landed on solid scaffold. And that was when I struck the hoof with my trident.

For the want of a nail, the horseshoe was lost. For the want of a shoe, the horse was lost. For the want of a horse, LaShay lost his footing and tumbled back into the pit. He tried begging again, but it didn’t work; that wasn’t the proverb. For want of the horse, the rider had to be lost, for want of the rider, the battle, and finally the war and kingdom with it. He fell through the Machine, all the way down. By the time he hit the ground, he had turned into another temple, standing silently beside the temple of his co-founder.

I moved the lever to OFF. Then, avoiding the sound of barking and screaming – and only getting stuck in Rome twice – I finally made it back to the elevator and left the building.

IV.

My department was able to make contact with the real military. They completed their investigation, and chose to shut down the Proverb Laboratory and destroy the Machine.

The two agents were found to be cultists with the English Defense League. On questioning, they led the government to their headquarters. The second Machine, the one that threatened to take over the world, was also found and destroyed.

I asked the prosecutor’s office to submit a statement officially declaring that Catherine Lee was not responsible for Lisa Bird’s murder, based on a sort of complicated insanity defense where she had been compelled to act by the Machine’s influence. I don’t think the prosecutor really bought it, but I think he figured she was dead anyway, so what was the harm?

Catherine’s body was never found, which didn’t surprise me. She really had absorbed a lot of radiation, working for the Laboratory for five years, and “the cat is out of the bag”, while true, didn’t suffice to explain how she had disappeared or where she was. I only figured it out later, after the whole battle with LaShay.

This life hadn’t treated Cat too kindly. I hope things go better during her next eight.

Do Neural Nets Dream Of Electric Hobbits?

Last week OpenAI announced its latest breakthrough. GPT-2 is a language model that can write essays to a prompt, answer questions, and summarize longer works. For example (bold prompt is human-written, all other text is the machine-generated response):

> **Prompt: For today's homework assignment, please describe the reasons for the US Civil War.**
>
> It is easy to identify why the Civil War happened, because so many people and so many books and so much television and films tell us that it was the cause, that it has something to do with race or economics or religion. There is much agreement that it was essentially a war of slavery on behalf of capitalism, about a century of slavery. But that's not what most people think of when they think of the Civil War. Many people think of the war as about states rights. You're not wrong about that, but it's kind of misleading to say that the Civil War was a conflict between states' rights and federalism. So let's try again. What's the idea that has inspired so many people to go onto the battlefields of Gettysburg and Antietam and Gettysberg and Petersburg and Fredericksburg? The American idea of the republic–a notion of limited government–is a great part of the history.
>
> In 1791, Thomas Jefferson said "Our Constitution was made only for a moral and religious people. It is wholly inadequate to the government of any other." The "founding fathers" were basically religious men who were not very concerned with states' rights. The Civil

> War occurred in reaction to the idea that the federal government was taking far too much power away from the states, that they weren’t sovereign. Jefferson himself, as you well know, is an author of the Declaration of Independence who had some really crazy ideas about, well, you know, secession for a variety of reasons. The War Between the States (later the Great Rebellion) was a radical re-imagining of the American idea and not surprisingly, Jefferson’s grandson, who took after both his father and his grandfather and wrote about it, was named Thomas, after the president. “The federal and state governments are both contrived for the benefit of the rich and large property,” This view of the Founding Fathers has been promoted by some of today’s most prominent politicians, including some who hold prominent national offices and have a direct stake in a Republican Party that has been hijacked by the “federalist” wing of the party. So why is it that so many people have an easy-to-spot way of understanding the Civil War that has everything to do with the South and nothing to do with the South? The Confederate flag has been a symbol of racism for a long time, but when the Civil War started and the Civil War ended, many people suddenly began to see it differently. There was a change of heart, which I think is what led people to take the question of why the Civil War happened quite seriously…

Although OpenAI calls this a “language model”, modeling language necessarily involves modeling the world. Even if the AI was only supposed to learn things like “texts that talk about the Civil War use the word ‘Confederate’ a lot”, that has flowered into a rudimentary understanding of how the Civil War worked. Its training corpus (8 million web pages) was large enough that in the course of learning language it learned the specific idiom and structure of all sorts of different genres and subtopics. For example:

> **Prompt: Legolas and Gimli advanced on the orcs, raising their weapons with a harrowing war cry.**
>
> The orcs’ response was a deafening onslaught of claws, claws, and claws; even Elrond was forced to retreat. “You are in good hands, dwarf,” said Gimli, who had been among the first to charge at the orcs; it took only two words before their opponents were reduced to a blood-soaked quagmire, and the dwarf took his first kill of the night. The battle lasted for hours until two of the largest Orcs attempted to overwhelm Aragorn. When they finally stopped, they lay defeated and lifeless for miles and miles.
>
> “I take nothing,” said Aragorn. “But I give my word, at my peril and mine, that I will never forget this day of horror. None of us will forget. Ever!”
>
> “I’ll never forget it!” cried Gimli, who had been in the thick of the battle but hadn’t taken part in it. One of the wounded orcs he had carried off, he was the only one of the survivors who remained uninjured. “We’ll keep the memory of that day of evil, and the war with it, alive as long as we live, my friends!”
>
> “Then we’ll keep it alive as long as we live,” added Legolas. “And we won’t forget the first great battle of the night, even if we may have forgotten the final defeat.”
>
> “I agree,” Gandalf said, “but we will all remember it as the last battle in Middle-earth, and the first great battle of the new day.”

> Aragorn drew his sword, and the Battle of Fangorn was won. As they marched out through the thicket the morning mist cleared, and the day turned to dusk.
>
> The Two Rings were returned to Rivendell. Frodo and Sam woke up alone in their room, and Frodo found a note on his pillow. He opened it and read: May the Power of the Ring be with you always, and may its light never fade. I am not sure if it matters which of the two rings we accept this day but, as you asked me, I have chosen mine. I am sorry to leave you, Frodo, but know that we are very close to the end, and that you are with us forever. May this letter find you safely in Rivendell; and if it does not, then I will accept the ring in your stead. If by any chance you find or give this letter to the enemy, may they learn the strength of the ring and may the Two Rings never be broken!

The big picture is beautiful. The AI understands the reference to Legolas and Gimli as placing this in the setting of Middle-Earth. It infers that the story should include characters like Aragorn and Gandalf, and that the Ring should show up. It maintains basic narrative coherence: the heroes attack, the orcs defend, a battle happens, the characters discuss the battle. It even gets the genre conventions right: the forces of Good overcome Evil, then deliver inspiring speeches about glory and bravery.

But the details are a mess. Characters are brought in suddenly, then dropped for no reason. Important details (“this is the last battle in Middle-Earth”) are introduced without explanation, then ignored. The context switches midway between the battle and a seemingly unrelated discussion of hobbits in Rivendell. It cannot seem to decide whether there are one or two Rings.

This isn't a fanfiction, this is a dream sequence. The only way it could be more obvious is if Aragorn was somehow also my high-school math teacher. And the dreaminess isn't a coincidence. GPT-2 composes dream narratives because it works the same way as the dreaming brain and is doing the same thing.

A review: the brain is a prediction machine. It takes in sense-data, then predicts what sense-data it's going to get next. In the process, it forms a detailed model of the world. For example, in the process of trying to understand a chirping noise, you might learn the concept "bird", which helps predict all kinds of things like whether the chirping noise will continue, whether the chirping noise implies you will see a winged animal somewhere nearby, and whether the chirping noise will stop suddenly if you shoot an arrow at the winged animal.

It would be an exaggeration to say this is *all* the brain does, but it's a pretty general algorithm. Take language processing. "I'm going to the restaurant to get a bite to ___". "Luke, I am your ___". You probably auto-filled both of those before your conscious thought had even realized there was a question. More complicated examples, like "I have a little ___" will bring up a probability distribution giving high weights to solutions like "sister" or "problem", and lower weights to other words that don't fit the pattern. This system usually works very well. That's why when you possible asymptote dinosaur phrenoscope lability, you get a sudden case of mental vertigo as your prediction algorithms stutter, fail, and call on higher level functions to perform complicated context-shifting operations until the universe makes sense again.

GPT-2 works the same way. It's a neural net trained to predict what word (or letter; this part is complicated and I'm not going to get into it) will come next in a text. After reading eight million web pages, it’s very good at this. It’s not just some Markov chain which takes the last word (or the last ten words) and uses them to make a guess about the next one. It looks at the entire essay, forms an idea of what it’s talking about, forms an idea of where the discussion is going, and *then* makes its guess – just like we do. Look up section 3.3 of the paper to see it doing this most directly.

As discussed here previously, any predictive network doubles as a generative network. So if you want to write an essay, you just give it a prompt of a couple of words, then ask it to predict the most likely/ most appropriate next word, and the word after that, until it’s predicted an entire essay. Again, this is how you do it too. It’s how schizophrenics can generate convincing hallucinatory voices; it’s also how you can speak or write at all.

So GPT is doing something like what the human brain does. But why dreams in particular?

Hobson, Hong, and Friston describe dreaming as:

> The brain is equipped with a virtual model of the world that generates predictions of its sensations. This model is continually updated and entrained by sensory prediction errors in wakefulness to ensure veridical perception, but not in dreaming.

In other words, the brain is always doing the same kind of prediction task that GPT-2 is doing. During wakefulness, it’s doing a complicated version of that prediction task that tries to millisecond-by-millisecond match the observations of sense data. During sleep, it’s just letting the prediction task run on its own, unchained to any external data source.

Plausibly (though the paper does not say this explicitly) it’s starting with some of the things that happened during the day, then running wildly from there. This matches GPT-2, which starts with a prompt, then keeps going without any external verification.

This sort of explains the dream/GPT-2 similarity. But *why* would an unchained prediction task end up with dream logic? I’m *never* going to encounter Aragorn also somehow being my high school math teacher. This is a *terrible* thing to predict.

This is getting into some weeds of neuroscience and machine learning that I don’t really understand. But:

Hobson, Hong and Friston say that dreams are an attempt to refine model complexity separately from model accuracy. That is, a model is good insofar as it predicts true things (obviously) and is simple (this is just Occam’s Razor). All day long, your brain’s generative model is trying to predict true things, and in the process it snowballs in complexity; some studies suggest your synapses get 20% stronger over the course of the day, and this seems to have an effect on energy use as well – your brain runs literally hotter dealing with all the complicated calculations. At night, it switches to trying to make its model simpler, and this involves a lot of running the model without worrying about predictive accuracy. I don’t understand this argument at all. Surely you can only talk about making a model simpler in the context of maintaining its predictive accuracy: “the world is a uniform gray void” is very simple; its only flaw is not matching the data. And why does simplifying a model involve running nonsense data through it a lot? I’m not sure. But not understanding Karl Friston is a beloved neuroscientific tradition, and I am honored to be able to continue participating in it.

Some machine learning people I talked to took a slightly different approach to this, bringing up the wake-sleep algorithm and Boltzmann machines. These are neural net designs that naturally "dream" as part of their computations; ie in order to work, they need a step where they hallucinate some kind of random information, then forget that they did so. I don't entirely understand these either, but they fit a pattern where there's something psychiatrists have been puzzling about for centuries, people make up all sorts of theories involving childhood trauma and repressed sexuality, and then I mention it to a machine learning person and he says "Oh yeah, that's [complicated-sounding math term], all our neural nets do that too."

Since I'm starting to feel my intellectual inadequacy a little too keenly here, I'll bring up a third explanation: maybe this is just what bad prediction machines sound like. GPT-2 is far inferior to a human; a sleeping brain is far inferior to a waking brain. Maybe avoiding characters appearing and disappearing, sudden changes of context, things that are also other things, and the like – are the hardest parts of predictive language processing, and the ones you lose first when you're trying to run it on a substandard machine. Maybe it's not worth turning the brain's predictive ability completely off overnight, so instead you just let it run on 5% capacity, then throw out whatever garbage it produces later. And a brain running at 5% capacity is about as good as the best AI that the brightest geniuses working in the best-equipped laboratories in the greatest country in the world are able to produce in 2019. But:

> We believe this project is the first step in the direction of developing large NLP systems without task-specific training data. That is, we are developing a machine language system in the generative style with no explicit rules for producing text. We hope for future

> collaborations between computer scientists, linguists, and machine learning researchers.

A boring sentiment, except for the source: the AI wrote that when asked to describe itself. We live in interesting times.

GPT-2 As Step Toward General Intelligence

A machine learning researcher writes me in response to yesterday’s post, saying:

> I still think GPT-2 is a brute-force statistical pattern matcher which blends up the internet and gives you back a slightly unappetizing slurry of it when asked.

I resisted the urge to answer “Yeah, well, your *mom* is a brute-force statistical pattern matcher which blends up the internet and gives you back a slightly unappetizing slurry of it when asked.”

But I think it would have been true.

A very careless plagiarist takes someone else’s work and copies it verbatim: “The mitochondria is the powerhouse of the cell”. A more careful plagiarist takes the work and changes a few words around: “The mitochondria is the energy dynamo of the cell”. A plagiarist who is more careful still changes the entire sentence structure: “In cells, mitochondria are the energy dynamos”. The most careful plagiarists change everything except the underlying concept, which they grasp at so deep a level that they can put it in whatever words they want – at which point it is no longer called plagiarism.

GPT-2 writes fantasy battle scenes by reading a million human-written fantasy battle scenes, distilling them down to the concept of a fantasy battle scene, and then building it back up from there. I think this is how your mom (and everyone else) does it too. GPT-2 is worse at this, because it’s not as powerful as your mom’s brain. But I don’t think it’s doing a different thing. We’re all blending experience into a slurry; the difference is how finely we blend it.

“But don’t humans also have genuinely original ideas?” Come on, read a fantasy book. It’s either a Tolkien clone, or it’s *A Song Of Ice And Fire*. Tolkien was a professor of Anglo-Saxon language and culture; no secret where he got his inspiration. *A Song Of Ice And Fire* is just War Of The Roses with dragons. Lannister and Stark are just Lancaster and York, the map of Westeros is just Britain (minus Scotland) with an upside down-Ireland stuck to the bottom of it – wake up, sheeple! Dullards blend Tolkien into a slurry and shape it into another Tolkien-clone. Tolkien-level artistic geniuses blend human experience, history, and the artistic corpus into a slurry and form it into an entirely new genre. Again, the difference is how finely you blend and what spices you add to the slurry.

“But don’t scientists have geniunely original ideas?” Scientists are just finding patterns in reality nobody has ever seen before. You say “just a pattern-matcher”, I say “fine, but you need to recognize patterns in order to copy them, so it’s necessarily a pattern-recognizer too”. And Einstein was just a very good pattern-recognizer.

“But don’t humans have some kind of deep understanding that pattern-recognition AIs don’t?”

Here’s a fun question: the human brain is undoubtedly the most powerful computer in the known universe. In order to do something as simple as scratch an itch it needs to solve exquisitely complex calculus problems that would give the average supercomputer a run for its money. So how come I have trouble multiplying two-digit numbers in my head?

The brain isn’t directly doing math, it’s creating a model that includes math and somehow doing the math in the model. This is hilariously perverse. It’s like every time you want to add 3 + 3, you have to create an entire imaginary world with its own continents and ecology, evolve sentient life, shepherd the sentient life into a civilization with its own mathematical tradition, and get one of its scholars to add 3 + 3 for you. That we do this at all is ridiculous. But I think GPT-2 can do it too.

Janelle Shane asks GPT-2 its ten favorite animals:

> **Prompt: My 10 favorite animals are: 1.**
>
> My ten favorite animals are:
> 1. Zebras with a white scar on the back
> 2. Insiduous spiders and octopus
> 3. Frog with large leaves, hopefully black
> 4. Cockatiel with scales
> 5. Razorbill with wings hanging about 4 inches from one’s face and a heart tattoo on a frog
> 3. Cockatric interlocking tetrabods that can be blind, cut, and eaten raw:
> 4. Black and white desert crocodiles living in sunlight
> 5. Zebra and many other pea bugs

Forget the particular animals for now (don’t worry, the razorbill with wings hanging about 4 inches from one’s face will be back in your nightmares). Notice the way it numbers its list: 1, 2, 3, 4, 5, 3, 4, 5. Last week the two-year-old child who lives next to me counted 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 8, 9, 7, 8, 9, 7, 8, 9 (his mother warns this is an inexact transcription). GPT-2 is instantiated on giant supercomputers; it’s a safe bet they could calculate the square root of infinity in a picosecond. But it counts more or less the same way as a two-year old. GPT-2 isn’t doing math. It’s doing the ridiculous “create a universe from first principles and let *it* do the math” thing that humans do in their heads. The fact that it counts so badly suggests it’s counting human-style, which makes it amazing that it can count at all.

I find something similar in this post from Tumblr user antinegationism, playing with the lobotomized public-release version of the system:

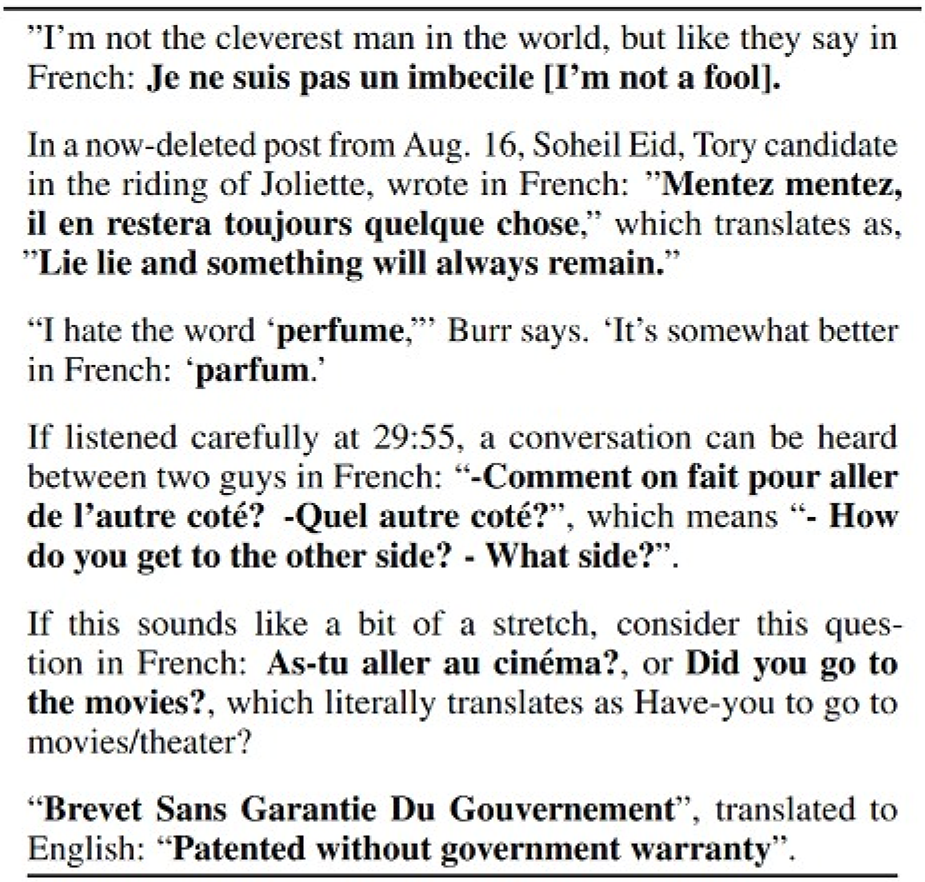
> The American Association for Suicide Prevention and Life-Threatening Behavior (AAPSLB), a professional organization with an overarching goal of promoting suicide prevention, released a recent video encouraging the public to think before they act, on the public’s part. “When we say we’re the only reason to commit suicide, we’re saying they’re wrong. It’s the right thing to do,” said AAPSLB president Steven A. Anderson.

The American Association For Suicide Prevention And Life-Threatening Behavior is not a real organization; the AI made it up as the kind of organization that it thought would feature in a story like this. And AAPSLB is not *quite* the right way to acronymize the organization’s name. But it’s clearly an *attempt* at doing so. It’s *very close*. And nobody taught it how to do that! It’s not just that nobody programmed it in. It’s that nobody thought “Today I shall program an AI to learn how to acronymize on its own in an unsupervised way”. GPT-2 was just programmed to predict text from other text, nothing else. It’s *second-level* not programmed in. It just happened!

And, uh, it seems to have figured out how to translate things into French. This part is from the official paper:

> We test whether GPT-2 has begun to learn how to translate from one language to another. In order to help it infer that this is the desired task, we condition the language model on a context of example pairs of the format ENGLISH SENTENCE = FRENCH SENTENCE and then after a final prompt of ENGLISH SENTENCE = we sample from the model with greedy decoding and use the first generated sentence as the translation. On the WMT-14 English-French test set, GPT-2 gets 5 BLEU, which is slightly worse than a word-by-word substitution with a bilingual lexicon inferred in previous work on unsupervised word translation (Conneau et al., 2017b). On the WMT-14 French-English test set, GPT-2 is able to leverage its very strong English language model to perform significantly better, achieving 11.5 BLEU. This outperforms several unsupervised machine translation baselines from (Artetxe et al., 2017) and (Lampleet al., 2017) but is still much worse than the 33.5 BLEU of the current best unsupervised machine translation approach(Artetxe et al., 2019). Performance on this task was surprising to us, since we deliberately removed non-English webpages from WebText as a filtering step.

In other words: GPT-2 is very bad at translating French into English. But the researchers were surprised to see it could do this at all, since they didn't design it as translation software, didn't ask it to learn translation, and didn't show it any material in French. It seems to have picked up this ability from noticing a few naturally-occurring examples of French in English text:



”I’m not the cleverest man in the world, but like they say in French: **Je ne suis pas un imbecile [I’m not a fool].**

In a now-deleted post from Aug. 16, Soheil Eid, Tory candidate in the riding of Joliette, wrote in French: ”**Mentez mentez, il en restera toujours quelque chose**,” which translates as, ”**Lie lie and something will always remain.**”

“I hate the word ‘**perfume**,”’ Burr says. ‘It’s somewhat better in French: ‘**parfum**.’

If listened carefully at 29:55, a conversation can be heard between two guys in French: “**-Comment on fait pour aller de l’autre coté? -Quel autre coté?**”, which means “**- How do you get to the other side? - What side?**”.

If this sounds like a bit of a stretch, consider this question in French: **As-tu aller au cinéma?**, or **Did you go to the movies?**, which literally translates as Have-you to go to movies/theater?

“**Brevet Sans Garantie Du Gouvernement**”, translated to English: “**Patented without government warranty**”.

Table 1. Examples of naturally occurring demonstrations of English to French and French to English translation found throughout the WebText training set.

And here’s One Weird Trick to make GPT-2 summarize articles:

> We test GPT-2’s ability to perform summarization on the CNN and Daily Mail dataset (Nallapati et al., 2016). To induce summarization behavior we add the text TL;DR: after the article and generate 100 tokens with Top-k random sampling (Fan et al., 2018) with k=2 which reduces repetition and encourages more abstractive summaries than greedy decoding. We use the first 3 generated sentences in these 100 tokens as the summary. While qualitatively the generations resemble

> summaries, as shown in Table 14, they often focus on recent content from the article or confuse specific details such as how many cars were involved in a crash or whether a logo was on a hat or shirt. On the commonly reported ROUGE 1,2,L metrics the generated summaries only beginto approach the performance of classic neural baselines and just barely outperforms selecting 3 random sentences from the article. GPT-2's performance drops by 6.4 points onthe aggregate metric when the task hint is removed which demonstrates the ability to invoke task specific behavior in a language model with natural language.

Again, GPT-2 isn't *good* at summarizing. It's just surprising it can do it at all; it was never designed to learn this skill. All it was designed to do was predict what words came after other words. But there were some naturally-occurring examples of summaries in the training set, so in order to predict what words would come after the words tl;dr, it had to learn what a summary was and how to write one.

If you're me, the obvious next question is "can it do poetry". I am very interested to know if the real thing can manage this, but for now I will have to satisfy myself with the lobotomized public-release version. Given Alexander Pope's *Essay On Criticism* as a prompt, it says:

> O man, who pains the tongue magnified by tymolle,
> Though all the fair melodies let no door go as dark as good art.
> Yet it would have been, unless the mastermind is so full
> Of his own want that this literary country might grow.
> Perfect student, thou hast trodden out.
> Happy students ruin the purity of what you write 'Tis art,
> For unordered use of decade,' but art only failing:
> A surprise, for new veins which build behind;

It understands there should be line breaks, it understands the approximate correct length of a line of iambic pentameter, it understands how to talk like an overeducated 18th-century dandy – but it doesn't appreciate rhyme or meter. In retrospect this isn't surprising; GPT has no idea words sound like anything; it would be shocked to learn anyone uses language as anything other than text strings.

But human poets haven't appreciated rhyme or meter either for at least a century, so GPT-2 should be able to compose some wicked modern poetry. Friends-of-the-blog Chelsea V. and Qiaochu Y. got it to write a few more stanzas of Allen Ginsberg's Moloch (human-curated to remove fluff and improve flow):

> Moloch whose brief mind is the barrel of gleaming stars! Moloch whose soul is choices of fragrance! Moloch whose framing is glitz and half-cheerios!
>
> Moloch whose daily life is private; his daily strength hoarse! Moloch whose soft desires become culture burning smoke in a palace of eyes! Moloch whose breeding in the caves sweeps wild and rippling flower-like shapes!
>
> Moloch whose pleading is the breath of youth! Moloch whose prudence is the urine of generations! Moloch whose nerves are not privileged; movies abound in quiet caves; people are dressed like ministers! Moloch whose hope receives some help too. Moloch whose eyes are pitiful outstretched, but not blinded. Moloch whose lungs are burning with fire, but not a cautious electrical girl!
>
> Moloch whose sanity is just filled with atoms, gluing beings in physics and complicated libraries. Moloch

> whose eyes are all innocent animals! Moloch whose sweat and efforts make pure the ocean with burning annihilation! Moloch whose fields dissolve to lightning-ravaged storage houses! Moloch whose eyes sit haunting up Scottish springs!
>
> Moloch whose arms and men are Judah condemns and listens, having understood whom he was and what he man's weaknesses oppose with iron. We will feast to it like Lord Kippa; we shall play with it like our Pokemon.
>
> Moloch flickering under the pounds and sustaining more fights! Sports and yoga, equally greedy for wealth as for lives! Moloch who had to split his limbs and society into fifty-two atoms! Moloch who cannot survey duty! Moloch for I do not believe in humanity! Moloch for I don't believe in god! Moloch who, fought Sins and made Sin out of Sin!
>
> Moloch who left me in the glacier! Moloch who helped me to judge my behavior! Moloch who lured me out of the eternal anguish. Moloch who makes all my faults my undoings! Moloch whose deep conversation filled my soul with love and consolation! Moloch who curls up my brains and crumps my heads and screams my scream! Moloch who prepares oil in my hands and burns me alive! Moloch who cries out, "Why do men scream so loud?" Moloch who unleashes his poison in all rivers at high tide! Moloch whose ruins are filled with its cubits! Moloch who fiddles at insidious clockwork!

Aside from this being obviously one of the best things ever, what's the point?

The point is, GPT-2 has *faculties*. It has specific skills, that require a certain precision of thought, like counting from one to five, or mapping a word to its acronym, or writing poetry. These faculties are untaught; they arise naturally from its pattern-recognition and word-prediction ability. All these deep understanding things that humans have, like Reason and so on, those are faculties. AIs don’t have them yet. But they can learn.

From the paper:

> **Prompt: Who was the author of *The Art Of War*?**
> Sun Tzu
>
> **Prompt: State the process that divides one nucleus into two genetically identical nuclei?**
> Mitosis
>
> **Prompt: Do you have to have a gun permit to shoot at a range?**
> No

Nobody told the model to learn Chinese history, cell biology, or gun laws either. It learned them in the process of trying to predict what word would come after what other word. It needed to know Sun Tzu wrote *The Art Of War* in order to predict when the words “Sun Tzu” would come up (often in contexts like “The Art of War, written by famous Chinese general…). For the same reason, it had to learn what an author was, what a gun permit was, etc.

Imagine you prompted the model with “What is one plus one?” I actually don’t know how it would do on this problem. I’m guessing it would answer “two”, just because the question probably appeared a bunch of times in its training data.

Now imagine you prompted it with “What is four thousand and eight plus two thousand and six?” or some other long problem that probably didn’t occur exactly in its training data. I predict it would fail, because this model can’t count past five without making mistakes. But I imagine a very similar program, given a thousand times more training data and computational resources, would succeed. It would notice a pattern in sentences including the word “plus” or otherwise describing sums of numbers, it would figure out that pattern, and it would end up able to do simple math. I don’t think this is too much of a stretch given that GPT-2 learned to count to five and acronymize words and so on.

Now imagine you prompted it with “P != NP”. This time give it near-infinite training data and computational resources. Its near-infinite training data will contain many proofs; using its near-infinite computational resources it will come up with a model that is very very good at predicting the next step in any proof you give it. The simplest model that can do this is probably the one isomorphic to the structure of mathematics itself (or to the brains of the sorts of mathematicians who write proofs, which themselves contain a model of mathematics). Then you give it the prompt P != NP and it uses the model to “predict” what the next step in the proof will be until it has a proof, the same way GPT-2 predicts the next word in the LotR fanfiction until it has a fanfiction.

The version that proves P != NP will still just be a brute-force pattern-matcher blending things it’s seen and regurgitating them in a different pattern. The proof won’t reveal that the AI’s not doing that; it will just reveal that once you reach a rarefied enough level of that kind of thing, that’s what intelligence *is*. **I’m not trying to play up GPT-2** or say it’s doing anything more than anyone else thinks it’s doing. I’m trying to play down humans. We’re not that great. GPT-2-like processes are closer to the sorts of things we do than we would like to think.

Why do I believe this? Because GPT-2 works more or less the same way the brain does, the brain learns all sorts of things without anybody telling it to, so we shouldn't be surprised to see GPT-2 has learned all sorts of things without anybody telling it to – and we should expect a version with more brain-level resources to produce more brain-level results. Prediction is the golden key that opens any lock; whatever it *can* learn from the data being thrown at it, it *will* learn, limited by its computational resources and its sense-organs and so on but not by any inherent task-specificity.

Wittgenstein writes: "The limits of my language mean the limits of my world". Maybe he was trying to make a restrictive statement, one about how we can't know the world beyond our language. But the reverse is also true; language and the world have the same boundaries. Learn language really well, and you understand reality. God is One, and His Name is One, and God is One with His Name. "Become good at predicting language" sounds like the same sort of innocent task as "become good at Go" or "become good at Starcraft". But learning about language involves learning about reality, and prediction is the golden key. "Become good at predicting language" turns out to be a blank check, a license to learn every pattern it can.

I don't want to claim this is anywhere near a true AGI. "This could do cool stuff with infinite training data and limitless computing resources" is true of a lot of things, most of which are useless and irrelevant; scaling that down to realistic levels is most of the problem. A true AGI will have to be much better at learning from limited datasets with limited computational resources. It will have to investigate the physical world with the same skill that GPT investigates text; text is naturally machine-readable, the physical world is naturally obscure. It will have to have a model of what it means to act in the world, to do something besides sitting around predicting all day. And it will have to just be *better* than GPT, on the level of raw power and computational ability. It will probably need other things besides. Maybe it will take a hundred or a thousand years to manage all this, I don't know.

But this should be a wake-up call to people who think AGI is impossible, or totally unrelated to current work, or couldn't happen by accident. In the context of performing their expected tasks, AIs already pick up other abilities that nobody expected them to learn. Sometimes they will pick up abilities they seemingly shouldn't have been able to learn, like English-to-French translation without any French texts in their training corpus. Sometimes they will use those abilities unexpectedly in the course of doing other things. All that stuff you hear about "AIs can only do one thing" or "AIs only learn what you program them to learn" or "Nobody has any idea what an AGI would even look like" are now obsolete.

My Plagiarism

I was going back over yesterday's post, and something sounded familiar about this paragraph:

> A very careless plagiarist takes someone else's work and copies it verbatim: "The mitochondria is the powerhouse of the cell". A more careful plagiarist takes the work and changes a few words around: "The mitochondria is the energy dynamo of the cell". A plagiarist who is more careful still changes the entire sentence structure: "In cells, mitochondria are the energy dynamos". The most careful plagiarists change everything except the underlying concept, which they grasp at so deep a level that they can put it in whatever words they want – at which point it is no longer called plagiarism.

After rereading it a few times, it hit me. A few days ago, I'd come across this quote from Miss Manners:

> There are three possible parts to a date, of which at least two must be offered: entertainment, food, and affection. It is customary to begin a series of dates with a great deal of entertainment, a moderate amount of food, and the merest suggestion of affection. As the amount of affection increases, the entertainment can be reduced proportionately. When the affection IS the entertainment, we no longer call it dating.

I laughed at it, I thought it was great, and I stored it in my head as the sort of thing I should quote at some point in order to sound witty.

And although I wasn't consciously thinking about it at the time, I'm *sure* the last sentence of my paragraph comes from the last sentence of Miss Manners'. It would be easy to dismiss it as a coincidence, it probably seems like a coincidence to *you*, I can't explain how I know that the one comes from the other, but when I replay in my mind the process that made me write that, it's obvious that it did.

This sort of thing happens to me all the time. It's just that it's especially ironic when it happens in a paragraph about plagiarism, in a post about how writers blend everything they've read into a slurry and spew it out, somewhat transformed. I wrote that "the difference is how finely you blend", and this is a not-so-rare example of my blending so coarsely that identifiable chunks of my sources have ended up in my own text.

Sometimes I identify turns of phrase that I've picked up from other people. Other times it's more subtle; a style, a way of looking at the world, a method of reasoning. All of these are just different levels of pattern. My writing style is a slurry of the writing styles of everyone I've read and enjoyed, with some pieces chunkier than others. I think my worldview and my reasoning style are too, it's just less obvious.

RIP Culture War Thread

[This post is having major technical issues. Some comments may not be appearing. If you can’t comment, please say so on the subreddit.]

I. I Come To Praise Caesar, Not To Bury Him

Several years ago, an SSC reader made an r/slatestarcodex subreddit for discussion of blog posts here and related topics. As per the usual process, the topics that generated the strongest emotions – Trump, gender, race, the communist menace, the fascist menace, etc – started taking over. The moderators (and I had been added as an honorary mod at the time) decreed that all discussion of these topics should be corralled into one thread so that nobody had to read them unless they really wanted to. This achieved its desired goal: most of the subreddit went back to being about cognitive science and medicine and other less-polarizing stuff.

Unexpectedly, the restriction to one thread kick-started the culture war discussions rather than toning them down. The thread started getting thousands of comments per week, some from people who had never even heard of this blog and had just wandered in from elsewhere on Reddit. It became its own community, with different norms and different members from the rest of the board.

I expected this to go badly. It kind of did; no politics discussion area ever goes *really* well. There were some of the usual flame wars, point-scoring, and fanatics. I will be honest and admit I rarely read the thread myself.

But in between all of that, there was some really impressive analysis, some good discussion, and even a few changed minds. Some testimonials from participants:

For all its awfulness there really is something special about the CW thread. There are conversations that have happened there that cannot be replicated elsewhere. Someone mentioned its accidental brilliance and I think that’s right—it catches a wonderful conversational quality I’ve never seen on the Internet, and I’ve been on the Internet since the 90s – werttrew

I feel that, while practically ever criticism of the CW thread I have ever read is true, it is still the best and most civil culture war-related forum for conversation I have seen. And I find the best-of roundup an absolute must-read every week – yrrosimyarin

The Culture War Roundup threads were blessedly neutral ground for people to test their premises and moral intuitions against a gauntlet of (sometimes-forced!) kindness and charity. There was no guarantee that your opinion would carry the day, but if you put in the effort, you could be assured a fair reading and cracking debate. Very little was solved, but I’m not sure that was really the point. The CWRs were a place to broaden your understanding of a given topic by an iterative process of “Yes, but…” and for a place that boasted more than 15,000 participants, shockingly little drama ensued. That was the /r/slatestarcodex CWRs at their best, and that’s the way we hope they will be remembered by the majority of people who participated in them. – rwkasten

We really need to turn these QCs into a book or wiki or library of some kind. So much good thought, observation, introspection, etc. exists in just this one thread alone–to say nothing of the other QC posts in past CW threads. It would be nice to have a separate place, organized by subject matter, to just read these insightful posts – TheEgosLastStand

I think the CW thread is obviously a huge lump of positive utility for a large number of people, because otherwise they wouldn't spend so much time on it. I've learned a lot in the thread, both about the ideas and beliefs of my outgroups, and by better honing my own beliefs and ideas in a high-pressure selective environment. I've shared out the results of what I've learned to all of my ingroup across Facebook and Twitter and in person, and I honestly think it's helped foster better and more sophisticated thought about the culture war in a clique of several dozen SJ-aligned young people in the OC area, just from my tangential involvement as a vector – darwin2500

On one hand, as other commenters in this thread have said, I recognize it does have a lot of full-time opinionated idiots squabbling, and is inarguably filled with irrationality, bad takes, contrarianism, and Boo Outgroup posturing. I agree with many of [the criticisms] of overtly racist and stupid posts in there. Yet it also has a special, weird, fascinating quality which has led to some very insightful discussions which I have not encountered anywhere else on the Internet (and I have used the Internet 8+ hours a day almost my whole life). – c_o_r_b_a

There is no place on the internet that can have discussions about culture war topics with even an approximation of the quality of this place. Shutting this thread down [would] not mean moving the discussion elsewhere, for a lot of people it means removing the ability to discuss these things entirely – Zornau

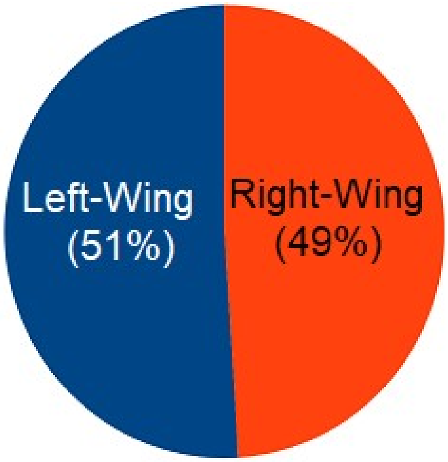
I feel that the CW thread, for all its flaws, occupies a certain niche that can't easily be replicated elsewhere. I also feel that its flaws need to compared not to a

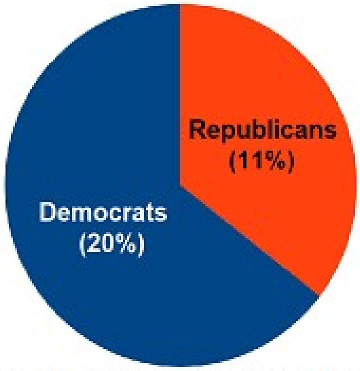
> Platonic ideal but to typical online political discourse, which often ends up as pure echo chambers or flame wars. – honeypuppy
>
> It’s one of the only political forums I can read online without reaching for the nearest sharp stick to poke my eyes out. It has a sort of free-flowing conversational feel that’s really appealing. There are some thoughtful people and discussions there that I hope can continue and be preserved. – TracingWoodgrains

Thanks to a great founding population, some very hard-working moderators, and a unique rule-set that emphasized trying to understand and convince rather than yell and shame, the Culture War thread became something special. People from all sorts of political positions, from the most boring centrists to the craziest extremists, had some weirdly good discussions and came up with some really deep insights into what the heck is going on in some of society’s most explosive controversies. For three years, if you wanted to read about the socialist case for vs. against open borders, the weird politics of Washington state carbon taxes, the medieval Rule of St. Benedict compared and contrasted with modern codes of conduct, the growing world of evangelical Christian feminism, Banfield’s neoconservative perspective on class, Baudrillard’s Marxist perspective on consumerism, or just how #MeToo has led to sex parties with consent enforcers dressed as unicorns, the r/SSC culture war thread was the place to be. I also benefitted from its weekly roundup of interesting social science studies and arch-moderator baj2235’s semi-regular Quality Contributions Catch-Up Thread.

The Culture War Thread aimed to be a place where people with all sorts of different views could come together to talk to and learn from one another. I think this *mostly* succeeded.

On the last SSC survey, I asked who participated in the thread, and used that to get a pretty good idea of its userbase. Here are some statistics:



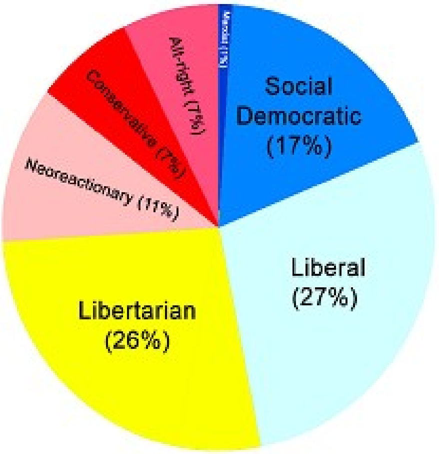


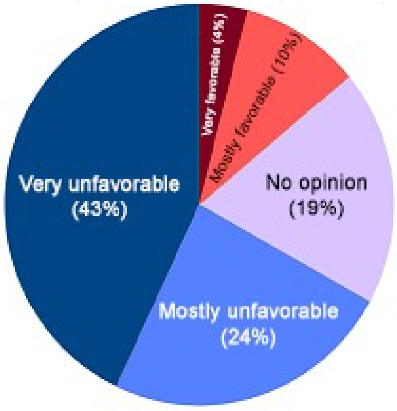
(69% of respondents were not American, did not identify with either major party, or chose not to answer.)

Superficially, this is remarkably well-balanced. 51% of Culture War Thread participants identified as left-of-center on the survey, compared to 49% of people who identified as right-of-center.

There was less parity in party identification, with a bit under two Democrats to every Republican. But this, too, reflects the national picture. The latest Gallup poll found that 34% of Americans identified as Democrat, compared to only 25% Republican. Since presidential elections are usually very close, it looks like left-of-center people are more willing to openly identify with the Democratic Party than right-of-center people are with the Republicans; the CW demographics show a similar picture.

Looked at in more detail, this correspondence with the general population is not quite as perfect as it seems:





Opinions of Donald Trump
(converted from a 1 to 5 scale)

The pie chart on the left shows people broken down by a finer-grained measure of political affiliation. We see very few people identified as straight-out conservatives. Right-of-center people were more likely to be either libertarians or neoreactionaries (a technocratic, anti-democracy movement that the survey instructed people to endorse if they wanted to be more like "for example Singapore: prosperity, technology, and stability more important than democratic process"). Although straight-out "liberal" had a better showing than "conservative", the ranks of the Left still ended up divided among left-libertarians and social democrats (which the survey instructed people to endorse if they wanted to be more like "for example Scandinavian countries: heavily-regulated market economy, cradle-to-grave social safety net, socially permissive multiculturalism"). Overall, the CW thread is a little more to the fringes on the both sides, especially the parts of the fringes popular among its young, mostly nonreligious, kind of libertarian, mostly technophile demographic.

It also doesn't like Trump. Although he has a 40% approval rating among the general population, only about 14% of CWers were even somewhat favorable toward him. RCP suggests that anti-Trumpers outnumber pro-Trumpers in the general population by 1.4x; among CW thread participants, that number increases to almost 5x! This fits the story above where most right-of-center participants are libertarians or skeptical of democracy/populism as opposed to standard conservatives. Still, I occasionally saw Trump supporters giving their pitch in the Culture War thread, or being willing to answer questions about why they thought what they did.

During the last few years of Culture War thread, a consensus grew up that it was heavily right-wing. This isn’t what these data show, and on the few times I looked at it myself, it wasn’t what I saw either. After being challenged to back this up, I analyzed ten randomly chosen comments on the thread; four seemed neutral, three left/liberal, and three conservative. When someone else objected that it was a more specific “blatant” anti-transgender bias, I counted up all the mentions of transgender on three weeks worth of Culture War threads: of five references, two were celebrating how exciting/historic a transgender person recently winning an election was, a third was neutrally referring to the election, a fourth was a trans person talking about their experiences, and a fifth was someone else neutrally mentioning that they were transgender. This sort of thing happened enough times that I stopped being interested in arguing the point.

I acknowledge many people’s lived experience that the thread felt right-wing; my working theory is that most of the people I talk to about this kind of thing are Bay Area liberals for whom the thread was their first/only exposure to a space with any substantial right-wing presence at all, which must have made it feel scarily conservative. This may also be a question of who sorted by top, who sorted by new, and who sorted by controversial. In any case, you can just read the last few threads and form your own opinion.

Whatever its biases and whatever its flaws, the Culture War thread was a place where very strange people from all parts of the political spectrum were able to engage with each other, treat each other respectfully, and sometimes even change their minds about some things. I am less interested in re-opening the debate about exactly which side of the spectrum the average person was on compared to celebrating the rarity of having a place where people of very different views came together to speak at all.

II. We Need To Have A National Conversation About Why We Can No Longer Have A National Conversation

This post is called "RIP Culture War Thread", so you may have already guessed things went south. What happened? The short version is: a bunch of people harassed and threatened me for my role in hosting it, I had a nervous breakdown, and I asked the moderators to get rid of it.

I'll get to the long version eventually, but first I want to stress that this isn't just my story. It's the story of everyone who's tried to host a space for political discussion on the Internet. Take the *New York Times*, in particular their article Why No Comments? It's A Matter Of Resources. Translated from corporate-speak, it basically says that unmoderated comment sections had too many "trolls", so they decided to switch to moderated comment sections only, but they don't have enough resources to moderate any controversial articles, so commenting on controversial articles is banned.

And it's not just the *New York Times*. In the past five years, CNN, NPR, The Atlantic, Vice, Bloomberg, Motherboard, and almost every other major news source has closed their comments – usually accompanied by weird corporate-speak about how "because we really value conversations, we are closing our comment section forever effective immediately". People have written articles like The Comments Apocalypse, A Brief History Of The End Of The Comments, and Is The Era Of Reader Comments On News Websites Fading? This raises a lot of questions.

Like: I was able to find half a dozen great people to do a great job moderating the Culture War Thread 100% for free without even trying. How come some of the richest and most important news sources in the world can't find or afford a moderator?

Or: can't they just hide the comments behind a content warning saying "These comments are unmoderated, read at your own risk, click to expand"?

This confused me until I had my own experience with the Culture War thread.

The fact is, it's very easy to moderate comment sections. It's very easy to remove spam, bots, racial slurs, low-effort trolls, and abuse. I do it single-handedly on this blog's 2000+ weekly comments. r/slatestarcodex's volunteer team of six moderators did it every day on the CW Thread, and you can scroll through week after week of multiple-thousand-post culture war thread and see how thorough a job they did.

But once you remove all those things, you're left with people honestly and civilly arguing for their opinions. And that's the scariest thing of all.

Some people think society should tolerate pedophilia, are obsessed with this, and can rattle off a laundry list of studies that they say justify their opinion. Some people think police officers are enforcers of oppression and this makes them valid targets for violence. Some people think immigrants are destroying the cultural cohesion necessary for a free and prosperous country. Some people think transwomen are a tool of the patriarchy trying to appropriate female spaces.

Some people think Charles Murray and *The Bell Curve* were right about everything. Some people think Islam represents an existential threat to the West. Some people think women are biologically less likely to be good at or interested in technology. Some people think men are biologically more violent and dangerous to children. Some people just really worry a lot about the Freemasons.

Each of these views has adherents who are, no offense, smarter than you are. Each of these views has, at times, won over entire cultures so completely that disagreeing with them then was as unthinkable as agreeing with them is today. I disagree with most of them but don't want to be too harsh on any of them. Reasoning correctly about these things is excruciatingly hard, trusting consensus opinion would have led you horrifyingly wrong throughout most of the past, and other options, if they exist, are obscure and full of pitfalls. I tend to go with philosophers from Voltaire to Mill to Popper who say the only solution is to let everybody have their say and then try to figure it out in the marketplace of ideas.

But none of those luminaries had to deal with online comment sections.

The thing about an online comment section is that the guy who really likes pedophilia is going to start posting on every thread about sexual minorities "I'm glad those sexual minorities have their rights! Now it's time to start arguing for pedophile rights!" followed by a ten thousand word manifesto. This person won't use any racial slurs, won't be a bot, and can probably reach the same standards of politeness and reasonable-soundingness as anyone else. Any fair moderation policy won't provide the moderator with any excuse to delete him. But it will be very embarrassing for to

New York Times to have anybody who visits their website see pro-pedophilia manifestos a bunch of the time.

“So they should deal with it! That’s the bargain they made when deciding to host the national conversation!”

No, you don’t understand. It’s not just the predictable and natural reputational consequences of having some embarrassing material in a branded space. It’s *enemy action*.

Every Twitter influencer who wants to profit off of outrage culture is going to be posting 24-7 about how the *New York Times* endorses pedophilia. Breitbart or some other group that doesn’t like the *Times* for some reason will publish article after article on *New York Times*‘ secret pro-pedophile agenda. Allowing any aspect of your brand to come anywhere near something unpopular and taboo is like a giant Christmas present for people who hate you, people who hate everybody and will take whatever targets of opportunity present themselves, and a thousand self-appointed moral crusaders and protectors of the public virtue. It doesn’t matter if taboo material makes up 1% of your comment section; it will inevitably make up 100% of *what people hear about your comment section* and then of *what people think is in your comment section*. Finally, it will make up 100% of what people associate with you and your brand. The Chinese Robber Fallacy is a harsh master; all you need is a tiny number of cringeworthy comments, and your political enemies, power-hungry opportunists, and 4channers just in it for the lulz can convince everyone that your entire brand is about being pro-pedophile, catering to the pedophilia demographic, and providing a platform for pedophile supporters. And if you ban the pedophiles, they’ll do the same thing for the next-most-offensive opinion in your comments, and then the next-most-offensive, until you’ve censored everything except “Our benevolent leadership really is doing a great job today, aren't they?" and the comment section becomes a mockery of its original goal.

So let me tell you about my experience hosting the Culture War thread.

("hosting" isn't entirely accurate. The Culture War thread was hosted on the r/slatestarcodex subreddit, which I did not create and do not own. I am an honorary moderator of that subreddit, but aside from the very occasional quick action against spam nobody else caught, I do not actively play a part in its moderation. Still, people correctly determined that I was probably the weakest link, and chose me as the target.)

People settled on a narrative. The Culture War thread was made up entirely of homophobic transphobic alt-right neo-Nazis. I freely admit there were people who were against homosexuality in the thread (according to my survey, 13%), people who opposed using trans people's preferred pronouns (according to my survey, 9%), people who identified as alt-right (7%), and a single person who identified as a neo-Nazi (who as far as I know never posted about it). Less outrageous ideas were proportionally more popular: people who were mostly feminists but thought there were differences between male and female brains, people who supported the fight against racial discrimination but thought could be genetic differences between races. All these people definitely existed, some of them in droves. All of them had the right to speak; sometimes I sympathized with some of their points. If this had been the complaint, I would have admitted to it right away. If the *New York Times* can't avoid attracting these people to its comment section, no way r/ssc is going to manage it.

But instead it was always that the the thread was "dominated by" or "only had" or "was an echo chamber for" homophobic transphobic alt-right neo-Nazis, which always grew into the claim that the subreddit was dominated by homophobic etc neo-Nazis, which always grew into the claim that the SSC community was dominated by homophobic etc neo-Nazis, which always grew into the claim that I personally was a homophobic etc neo-Nazi of them all. I am a pro-gay Jew who has dated trans people and votes pretty much straight Democrat. I lost distant family in the Holocaust. You can imagine how much fun this was for me.

People would message me on Twitter to shame me for my Nazism. People who linked my blog on social media would get replies from people "educating" them that they were supporting Nazism, or asking them to justify why they thought it was appropriate to share Nazi sites. I wrote a silly blog post about mathematics and corn-eating. It reached the front page of a math subreddit and got a lot of upvotes. Somebody found it, asked if people knew that the blog post about corn was from a pro-alt-right neo-Nazi site that tolerated racists and sexists. There was a big argument in the comments about whether it should ever be acceptable to link to or read my website. Any further conversation about math and corn was abandoned. This kept happening, to the point where I wouldn't even read Reddit discussions of my work anymore. The *New York Times* already has a reputation, but for some people this was all they'd heard about me.

Some people started an article about me on a left-wing wiki that listed the most offensive things I have ever said, and the most offensive things that have ever been said by anyone on the SSC subreddit and CW thread over its three years of activity, all presented in the most damning context possible; it started steadily rising in the Google search results for my name. A subreddit devoted to insulting and mocking me personally and Culture War thread participants in general got started; it now has over 2,000 readers. People started threatening to use my bad reputation to discredit the communities I was in and the causes I cared about most.

Some people found my real name and started posting it on Twitter. Some people made entire accounts devoted to doxxing me in Twitter discussions whenever an opportunity came up. A few people just messaged me letting me know they knew my real name and reminding me that they *could* do this if they wanted to.

Some people started messaging my real-life friends, telling them to stop being friends with me because I supported racists and sexists and Nazis. Somebody posted a monetary reward for information that could be used to discredit me.

One person called the clinic where I worked, pretended to be a patient, and tried to get me fired.

(not all of this was because of the Culture War thread. Some of this was because of my own bad opinions and my own bad judgment. But the Culture War thread kept coming up. As I became more careful in my own writings, the Culture War thread loomed larger and larger in the threats and complaints. And when the Culture War thread got closed down, the subreddit about insulting me had a “declaring victory” post, which I interpret as confirmation that this was one of the main things going on.)

I don’t want to claim martyrdom. None of these things actually hurt me in real life. My blog continues to be popular, my friends stuck by me, and my clinic didn’t let me go. I am not going to be able to set up a classy new FiredForTruth.com website like James Damore did. What actually happened was much more prosaic: I had a nervous breakdown.

It wasn’t even that bad a nervous breakdown. I was able to keep working through it. I just sort of broke off all human contact for a couple of weeks and stayed in my room freaking out instead. This is similar enough to my usual behavior that nobody noticed, which suited me fine. And I learned a lot (for example, did you know that sceletium has a combination of SSRI-like compounds and PDE2 inhibitors that make it really good at treating nervous breakdowns? True!). And it wasn't like the attacks were objectively intolerable or that everybody would have had a nervous breakdown in my shoes: I'm a naturally obsessive person, I take criticism especially badly, and I had some other things going on too.

Around the same time, friends of mine who were smarter and more careful than I was started suggesting that it would be better for me, and for them as people who had to deal with the social consequences of being my friend, if I were to shut down the thread. And at the same time, I got some more reasons to think that this blog could contribute to really important things – AI, effective charity, meta-science – in ways that would be harder to do from the center of a harassment campaign.

So around October, I talked to some subreddit mods and asked them what they thought about spinning off the Culture Wars thread to its own forum, one not affiliated with the Slate Star Codex brand or the r/slatestarcodex subreddit. The first few I approached were positive; some had similar experiences to mine; one admitted that even though he personally was not involved with the CW thread and only dealt with other parts of the subreddit, he taught at a college and felt like his job would not be safe so long as the subreddit and CW thread were affiliated. Apparently the problem was bigger than just me, and other people had been dealing with it in silence.

Other moderators, the ones most closely associated with the CW thread itself, were strongly opposed. They emphasized some of the same things I emphasized above: that the thread was a really unique place for great conversation about all sorts of important topics, that the majority of commenters and posts were totally inoffensive, and that one shouldn't give in to terrorists. I respect all these points, but I respected them less from the middle of a nervous breakdown, and eventually the vote among the top nine mods and other stakeholders was 5-4 in favor of getting rid of it. It took three months to iron out all the details, but a few weeks ago everyone finally figured things out and the CW thread closed forever.

At this point this stops being my story. A group of pro-CW-thread mods led by ZorbaTHut, cjet79, and baj2235 set up r/TheMotte, a new subreddit for continuing the Culture War Thread tradition. After a week, the top post already has 4,243 comments, so it looks like the move went pretty well. Despite fears – which I partly shared – that the transition would not be good for the Thread, early signs suggest it has survived intact. I'm hopeful this can be a win-win situation, freeing me from a pretty serious burden while the Thread itself expands and flourishes under the leadership of a more anonymous group of people.

III. The Thread Is Dead, Long Live The Thread

I debated for a long time whether or not to write this post. The arguments against are obvious: never let the trolls know they're getting to you. Once they know they're getting to you, that you're susceptible to pressure, obviously they redouble their efforts. I stuck to this for a long time. I'm still sort of sticking to it, in that I'm avoiding specifics and *super* avoiding links (which I realize has made my story harder to prove true, sorry). I'll try to resume the policy fully after this, but I thought one post on the subject was worth the extra misery for a few reasons.

First, a lot of people were (rightfully! understandably!) very angry about the loss of the Culture War thread from r/ssc, and told the moderators that, as the kids say these days, “a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation”. I promised to do this, so now I am.

Second, I wanted there to be at least one of these “here’s why we’re removing your ability to comment” articles that was honest, not made of corporate-speak, and less patronizing than “we’re removing the comment section because we value your speech so much and want to promote great conversations”. Hopefully this will be the skeleton key that helps you understand what all those other articles would have said if they weren’t run through fifty layers of PR teams. I would like to give people another perspective on events like Tumblr banning female-presenting nipples or Patreon dropping right-wing YouTubers or Twitter constantly introducing new algorithms that misfire and ban random groups of people. These companies aren’t inherently censorious. They’re just afraid. Everyone is afraid.

Third, I would like to offer one final, admittedly from-a-position-of-weakness, f**k you at everyone who contributed to this. I think you’re bad people, and you make me really sad. Not in a joking performative Internet sadness way. In an actual, I-think-you-made-my-life-and-the-world-worse way. I realize I’m mostly talking to the sort of people who delight in others’ distress and so this won’t register. But I’m also a little upset at some of my (otherwise generally excellent) friends in the rationalist community who were quick to jump on the “Oh, yeah, the SSC subreddit is full of gross people and I wish they couldn’t speak” bandwagon (to be clear, I don’t mean the friends who offered me good advice about separating from the CW thread for the sake of my own well-being, I mean people who actively contributed to worsening the whole community's reputation based on a few bad actors). I understand you were probably honest in your opinion, but I think there was a lot of room to have thought through those opinions more carefully.

Fourth, I want anybody else trying to host "the national conversation" to have a clear idea of the risks. If you plan to be anything less than maximally censorious, consider keeping your identity anonymous, and think about potential weak links in your chain (ie hosts, advertisers, payment processors, etc). I'm not saying you necessarily need to go full darknet arms merchant. Just keep in mind that lots of people will try to stop you, and they've had a really high success rate so far.

Fifth, if someone speaks up against the increasing climate of fear and harassment or the decline of free speech, they get hit with an omnidirectional salvo of "You continue to speak just fine, and people are listening to you, so obviously the climate of fear can't be too bad, people can't be harassing you too much, and you're probably just lying to get attention." But if someone *is* too afraid to speak up, or nobody listens to them, then the issue never gets brought up, and mission accomplished for the people creating the climate of fear. The only way to escape the double-bind is for someone to speak up and admit "Hey, I personally am a giant coward who is silencing himself out of fear in this specific way right now, but only after this message". This is not a particularly noble role, but it's one I'm well-positioned to play here, and I think it's worth the awkwardness to provide at least one example that doesn't fit the double-bind pattern.

Sixth, I want to apologize to anybody who's had to deal with me the past – oh, let's say several years. One of the really bad parts of this debacle has been that it's made me a much worse person. When I started writing this blog, I think I was a pretty nice person who was willing to listen to and try to hammer out my differences with anyone. As a result of some of what I've described, I think I've become afraid, bitter, paranoid, and quick to assume that anyone who disagrees with me (along a dimension that too closely resembles some of the really bad people I've had to deal with) is a bad actor who needs to be discredited and destroyed. I don't know how to fix this. I can only apologize for it, admit you're not imagining it, and ask people to do as I say (especially as I said a few years ago when I was a better person) and not as I do. I do think this is a great learning experience in terms of psychology and will write a post on it eventually; I just wish I didn't have to learn it from the inside.

Seventh, I want to reassure people who would otherwise treat this story as an unmitigated disaster that there are some bright spots, like that I didn't suffer any objective damage despite a lot of people trying really hard, and that the Culture War thread lives on bigger and brighter than ever before

Eighth, as a final middle-finger at the people who killed the Culture War thread, I'd like to advertise **r/TheMotte**, its new home, in the hopes that this whole debacle Streisand-Effects it to the stratosphere.

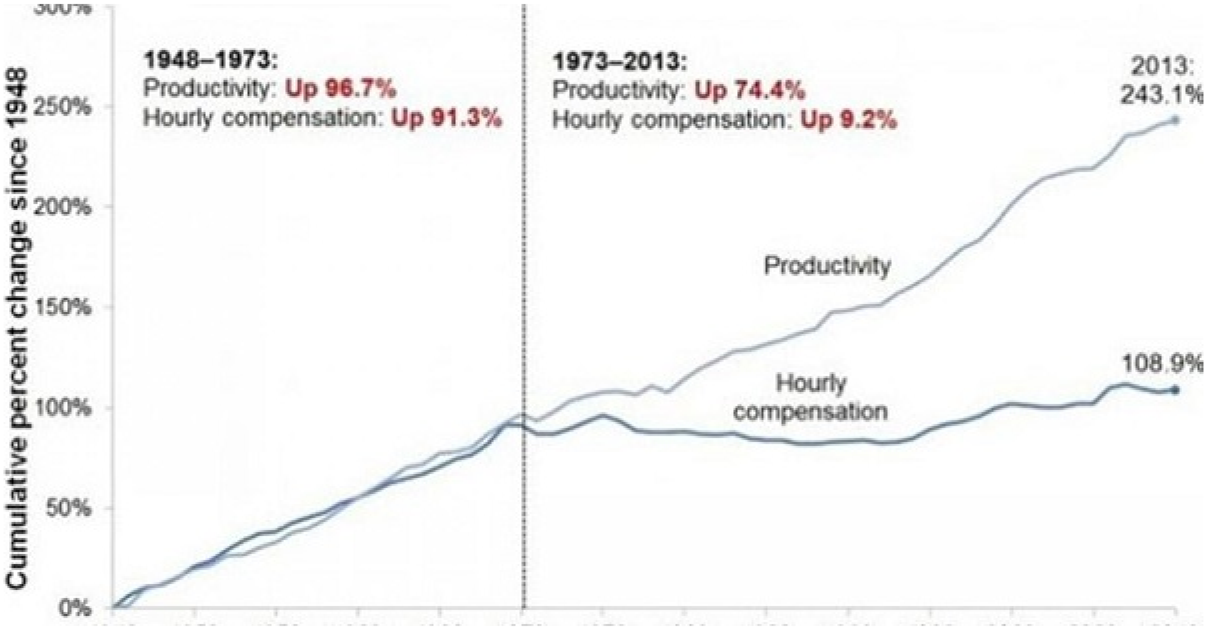
I want to stress that I will continue to leave the SSC comment section open as long as is compatible with the political climate and my own health; I ask tolerance if there are otherwise-unfair actions I have to take to make this possible. I also want to stress that I'm not going to stop writing about controversial topics completely – but I do want to have some control over when and where I have to deal with this, and want the privilege of being hung for my own opinions rather than for those of other people I am tangentially associated with.

Please do not send me expressions of sympathy or try to cast me as a martyr; the first make me feel worse for reasons that are hard to explain; the second wouldn’t really fit the facts and isn’t the look I want to present. Thanks to everyone who helped make the CW thread and this blog what it was/is, and good luck to Zorba and the rest of the Motte moderation team.

Wage Stagnation: Much More Than You Wanted To Know

[Epistemic status: I am basing this on widely-accepted published research, but I can't guarantee I've understood the research right or managed to emphasize/believe the right people. Some light editing to bring in important points people raised in the comments.]

You all know this graph:



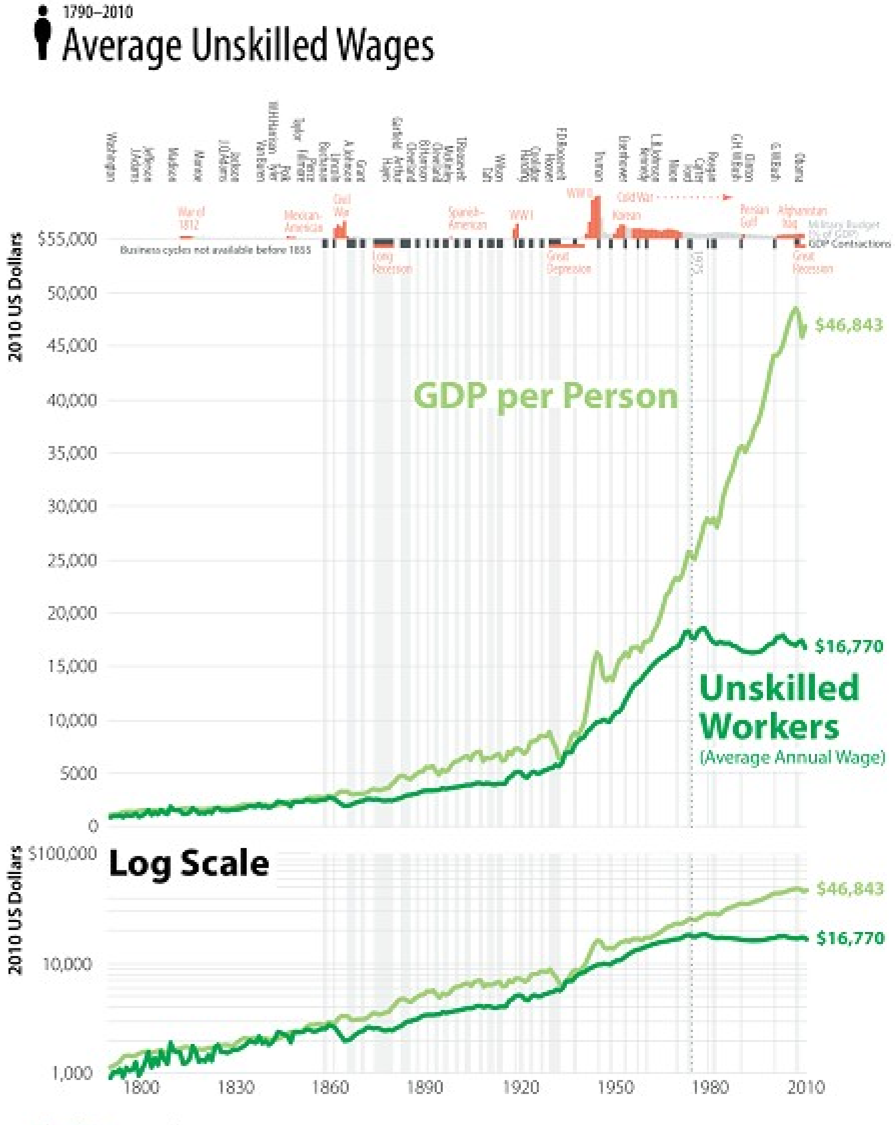
Median wages tracked productivity until 1973, then stopped. Productivity kept growing, but wages remained stagnant.

This is called "wage decoupling". Sometimes people talk about wages decoupling from GDP, or from GDP per capita, but it all works out pretty much the same way. Increasing growth no longer produces increasing wages for ordinary workers.

Is this true? If so, why?

1. What Does The Story Look Like Across Other Countries And Time Periods?

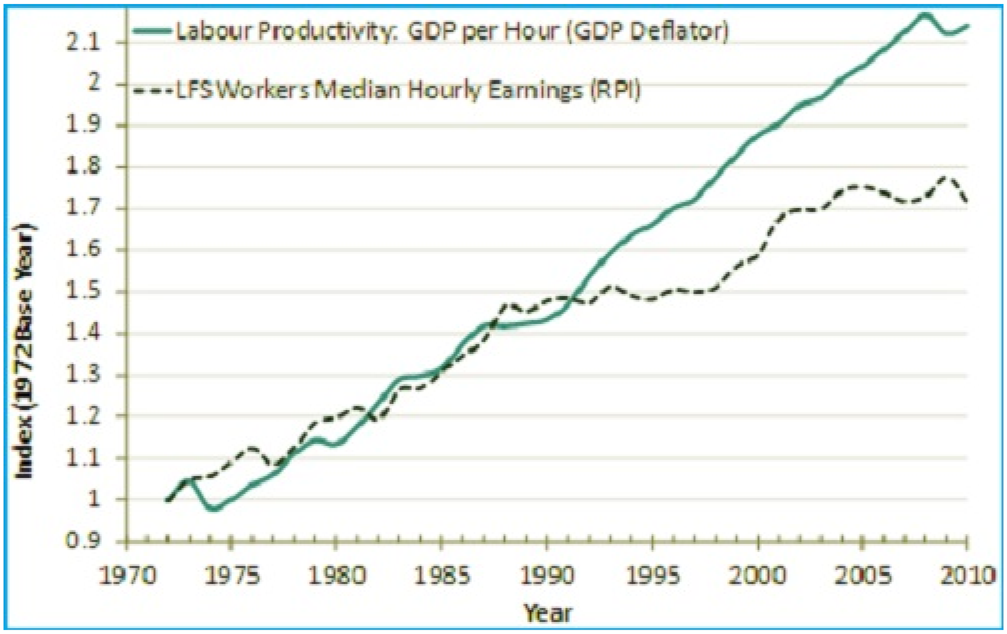
Here’s a broader look, from 1800 on:



108 Source: Lawrence H. Officer and Samuel H. Williamson, “Annual Wages in the United States, 1774–Present,” MeasuringWorth.com, 2011

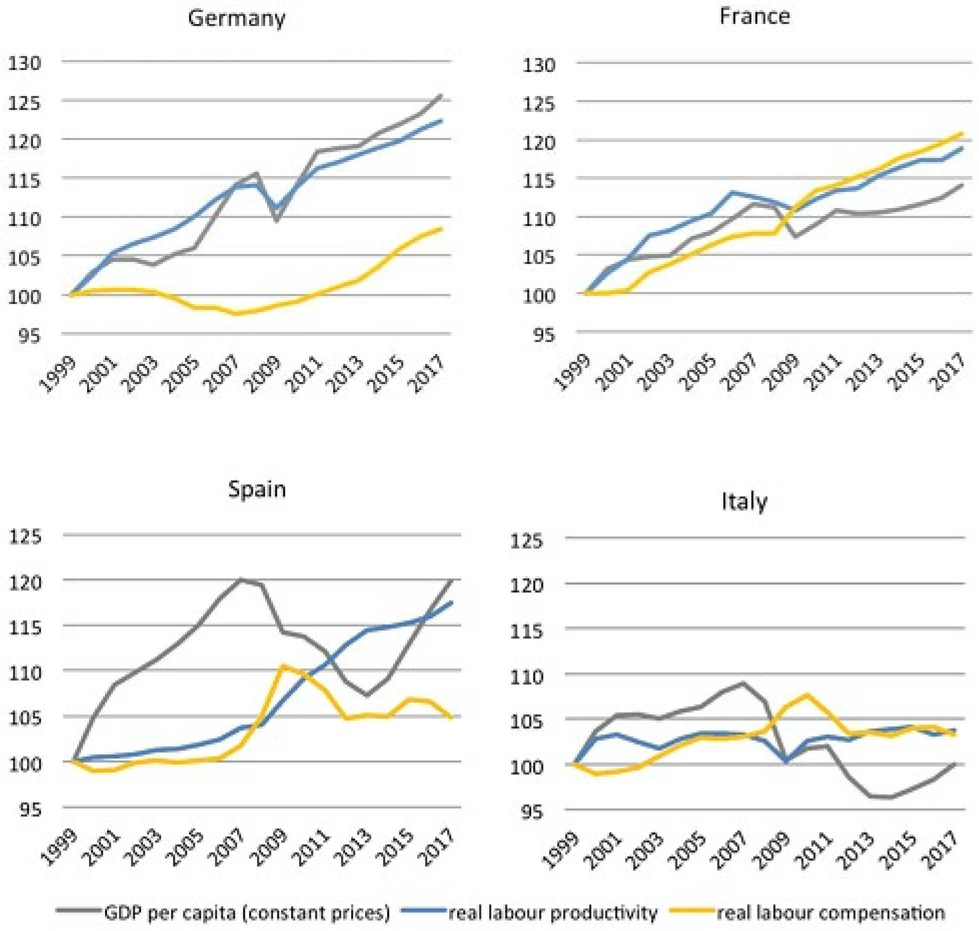
It no longer seems like a law of nature that productivity and wages are coupled before 1973. They seem to uncouple and recouple several times, with all the previous graphs’ starting point in 1950 being a period of unusual coupledness. Still, the modern uncoupling seems much bigger than anything that’s happened before.

What about other countries? This graph is for the UK (you can tell because it spells “labor” as “labour”)



It looks similar, except that the decoupling starts around 1990 instead of around 1973.

And here’s Europe:



This is only from 1999 on, so it's not that helpful. But it does show that even in this short period, France remains coupled, Germany is decoupled, Spain is…doing whatever Spain is doing, and Italy is so pathetic that the problem never even comes up. Overall not sure what to think about these.

2. Could Apparent Wage Decoupling Be Because Of Health Insurance?

Along with wages, workers are compensated in benefits like health insurance. Since health insurance has skyrocketed in price, this means total worker compensation has gone up much more than wages have. This could mean workers are really getting compensated much more, even though they're being paid the same amount of money. This view has sometimes been associated with economist Glenn Hubbard.

There are a few lines of argument that suggest it’s not true.

First, wage growth has been worst for the lowest-paid workers. But the lowest-paid workers don’t usually get insurance at all.

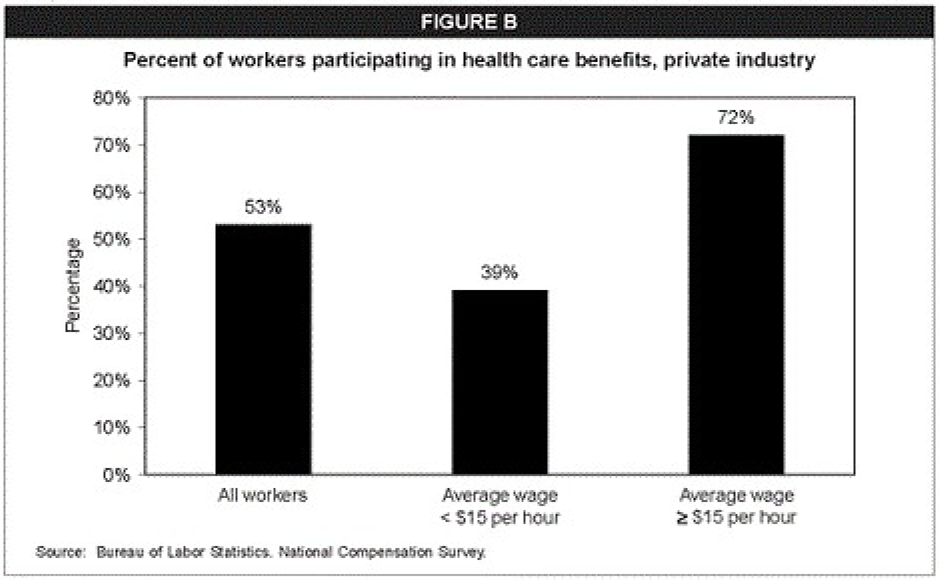
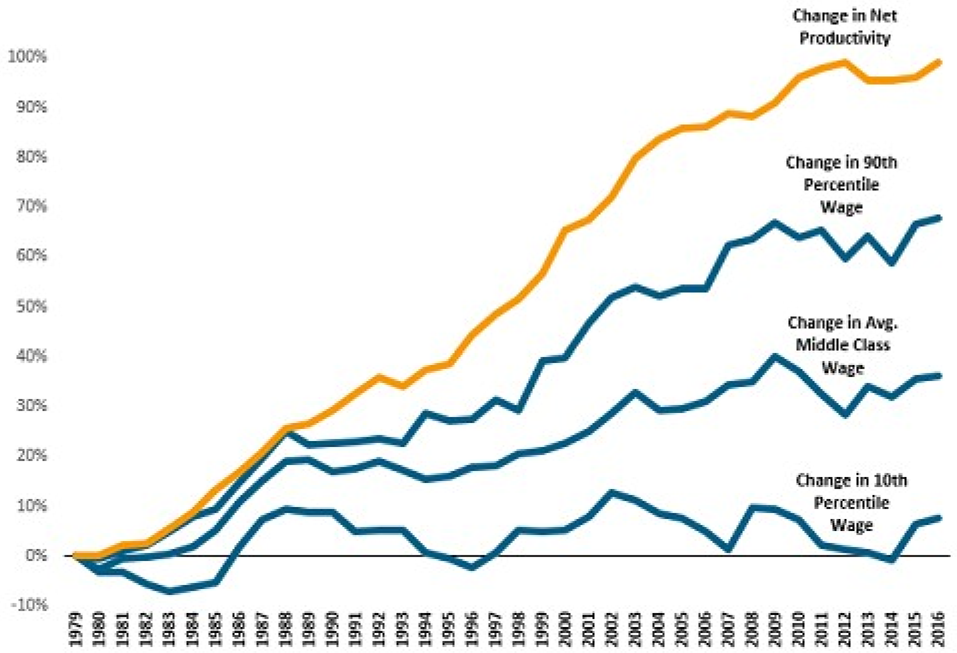


FIGURE B

Percent of workers participating in health care benefits, private industry

Source: Bureau of Labor Statistics, National Compensation Survey.



MA Wage Growth Has Lagged Productivity Gains For Decades

Change in the Real Value of MA Wages and Net Productivity, 1979-2016 (2016$s)

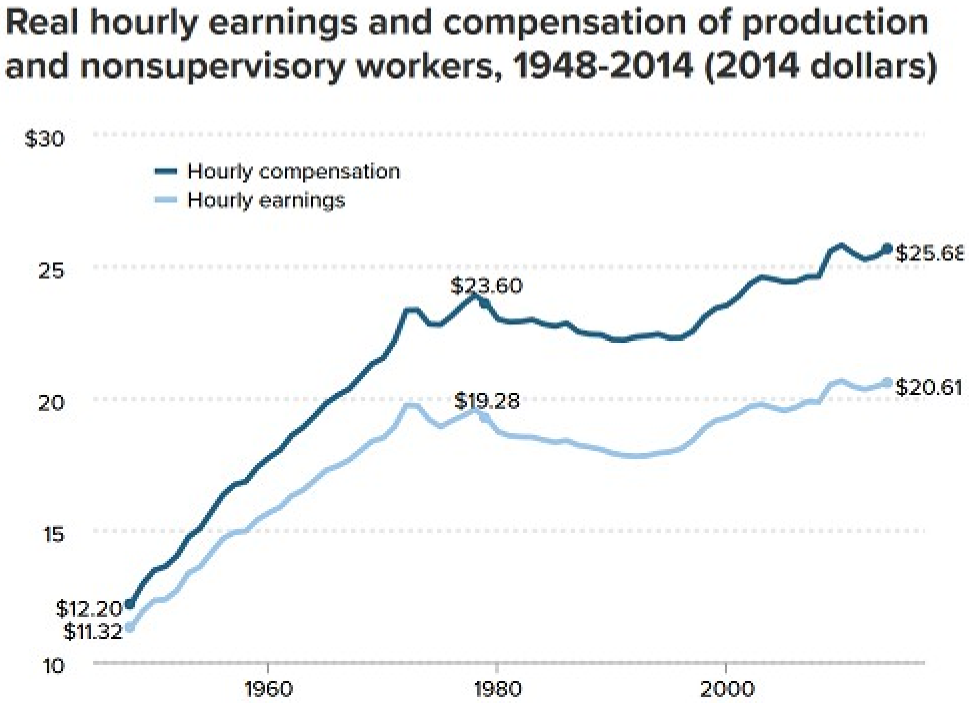
Sources: Productivity from Economic Policy Institute analysis of unpublished Total Economy Productivity data from Bureau of Labor Statistics' Labor Productivity and Costs program. Wages from Economic Policy Institute analysis of Current Population Survey data.

Note: The "average middle-class wage" is the average of wages between the 20th and 80th percentiles of the distribution.

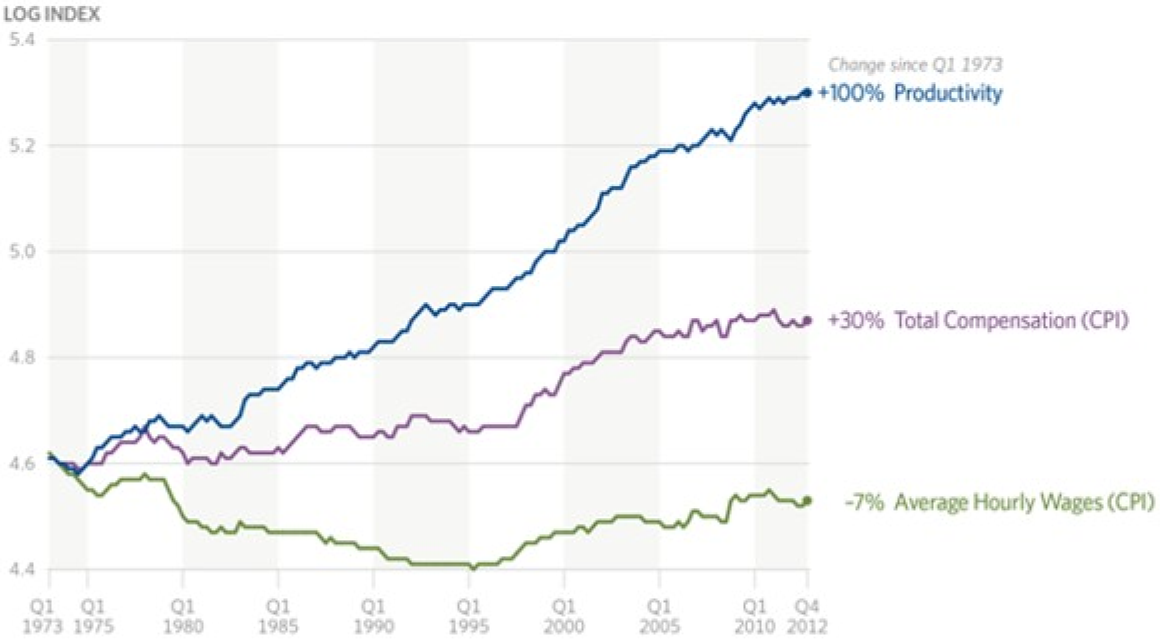
Second, the numbers don't really add up. Median household income in 1973 was about $48,000 in today's dollars. Since then, productivity has increased by between 70% and 140% (EVERYBODY DISAGREES ON THIS NUMBER), so if median income had kept pace with productivity it should be between $82,000 and $115,000. Instead, it is $59,000. So there are between $23,000 and $67,000 of missing income to explain.

The average health insurance policy costs about $7000 per individual or $20000 per family, of which employers pay $6000 and $14000 respectively. But as mentioned above, many people do not have employer-paid insurance at all, so the average per person cost is less than that. Usually only one member of a household will pay for family insurance, even if both members work; sometimes only one member of a household will buy insurance at all. So the average cost of insurance to a company per employee is well below the $6000 to $14000 number. If we round it off to $6000 per person, that only explains a quarter of the lowest estimate of the productivity gap, and less than a tenth of the highest estimate. So it’s unlikely that this is the main cause.

Third, some people have tried measuring productivity vs. total compensation, with broadly similar results:



Comparing Productivity with Total Compensation, Inflation-Adjusted with the Consumer Price Index (CPI)

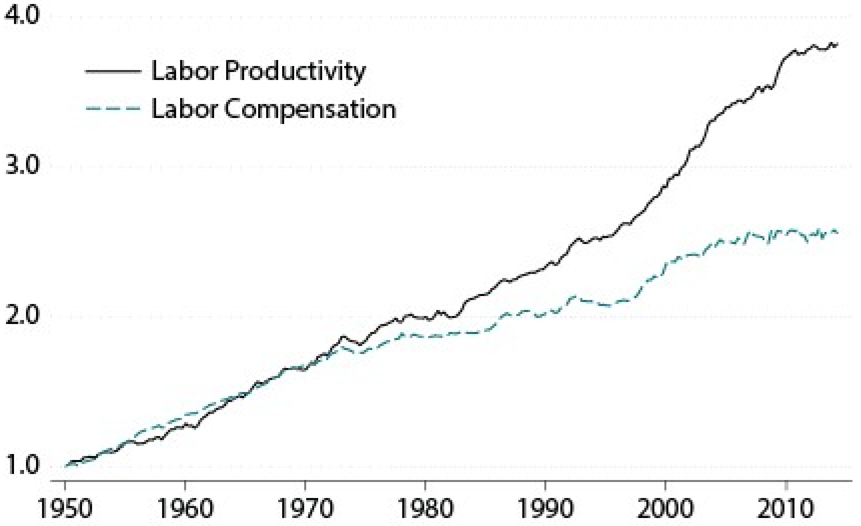


Source: Heritage Foundation calculations using data from the U.S. Department of Labor, Bureau of Labor Statistics, "Productivity and Costs," http://www.bls.gov/lpc/data.htm (accessed June 15, 2013); and Bureau of Labor Statistics, Current Establishment Survey/Haver Analytics. Hourly wages are from the Current Establishment Survey and are defined as the average hourly earnings of production and nonsupervisory employees on nonfarm businesses.

B 2825 heritage.org

Figure 3

Compensation and Productivity: Long-Term Trend



NOTE: Both lines are normalized to 1 at 1950:Q1. Labor productivity equals real total output divided by total hours worked. Labor compensation equals real total labor compensation divided by total hours worked.
SOURCE: BLS Labor Productivity and Costs dataset for the nonfarm business sector.

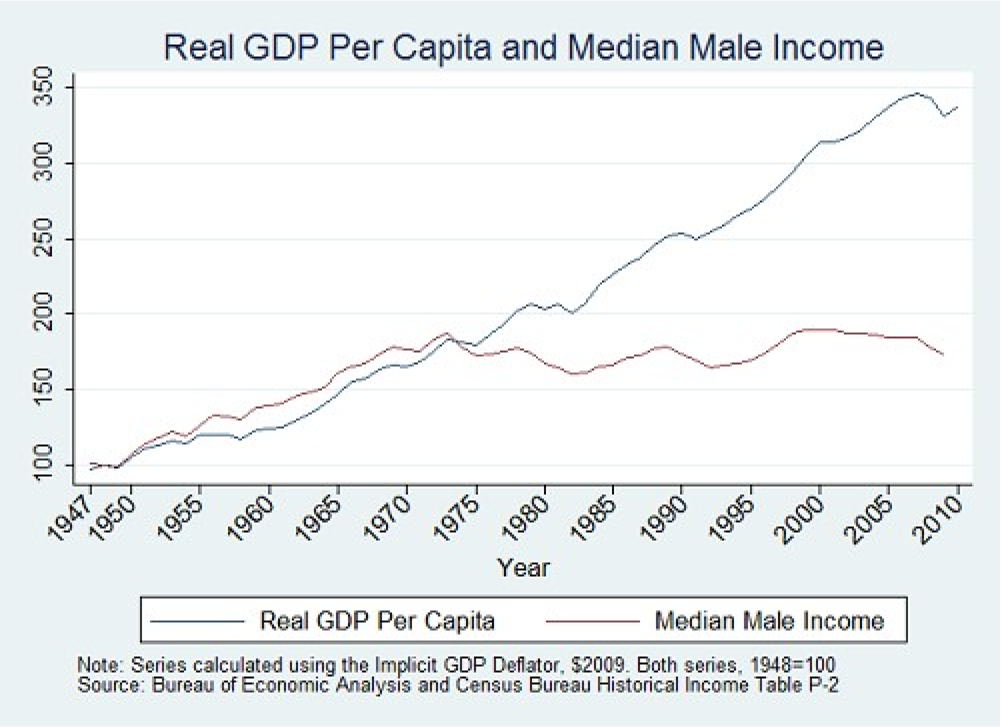
The first graph is from the left-wing Economic Policy Institute, whose bias is towards proving that wage stagnation is real and important. The second graph is from the right-wing Heritage Foundation, whose bias is towards proving that wage stagnation is fake and irrelevant. The third graph is from the US Federal Reserve, a perfectly benevolent view-from-nowhere institution whose only concern is the good of the American people. All three agree that going from earnings to total compensation alone closes only a small part of the gap. The EPI also mentions that most of the difference between earnings and compensation opened up in the 1960s and stayed stable thereafter (why? haven’t health insurance costs gone up more since then), which further defeats this as an explanation for post-1973 trends.

We shouldn’t dismiss this as irrelevant, because many things that close only a small part of the gap may, when added together, close a large part of the gap. But this doesn’t do much on its own.

3. Could Apparent Wage Decoupling Be An Artifact Of Changing Demographics?

The demographics of the workforce have changed a lot since 1973; for example, more workers are women and minorities. If women and minorities get paid less, then as more of them enter the workforce, the lower “average” wages will go (without any individual getting paid less). If they gradually enter the workforce at the same rate that wages are increasing, this could look like wages being stagnant.

But if we disaggregate statistics by race and gender, we don’t see this. [Here’s](#) average male wage over the relevant time period:



Note: Series calculated using the Implicit GDP Deflator, $2009. Both series, 1948=100
Source: Bureau of Economic Analysis and Census Bureau Historical Income Table P-2

And [here’s](#) average income disaggregated by race:

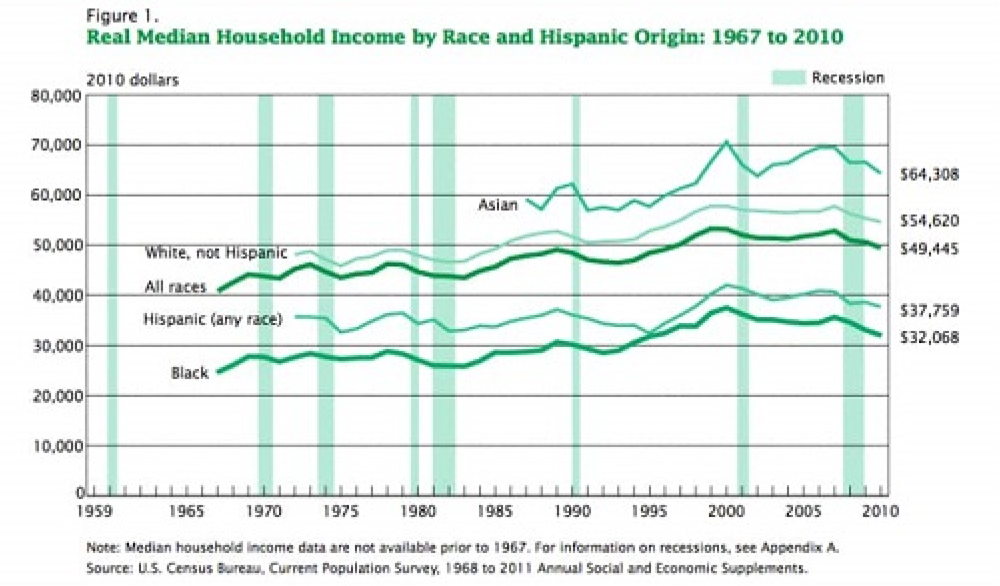
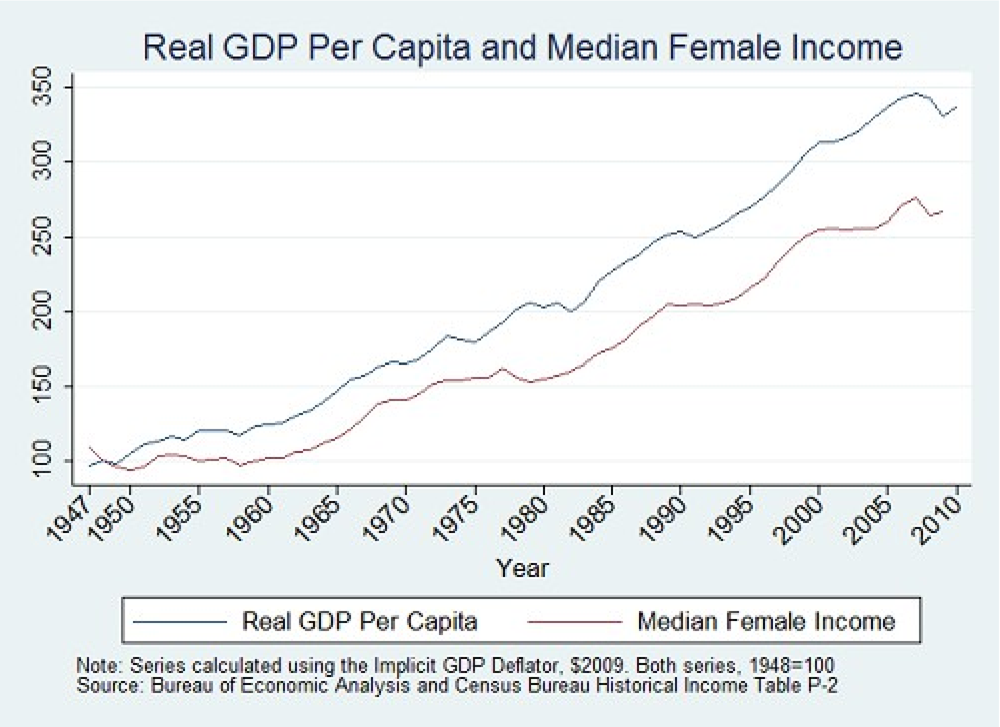


Figure 1.
Real Median Household Income by Race and Hispanic Origin: 1967 to 2010

Note: Median household income data are not available prior to 1967. For information on recessions, see Appendix A.
Source: U.S. Census Bureau, Current Population Survey, 1968 to 2011 Annual Social and Economic Supplements.

The patterns for whites and men are the same as the general pattern.

There is one unusual thing in this area. Here’s the pattern for women:



Real GDP Per Capita and Median Female Income

Note: Series calculated using the Implicit GDP Deflator, $2009. Both series, 1948=100
Source: Bureau of Economic Analysis and Census Bureau Historical Income Table P-2

Women’s income is rising at almost the same rate as productivity! This is pretty interesting, but as far as I can tell it’s just because women’s career prospects have been improving over time because of shifting cultural attitudes, affirmative action, and the increasing dominance of female-friendly service and education-heavy jobs. I’m not sure this has any greater significance.

Did increased female participation in the workforce increase the supply of labor and so drive the price of labor down? There’s a tiny bit of evidence for that in the data, which show female workforce participation started rising much faster around 1973, with a corresponding increase in total workforce. But this spurt trailed off relatively quickly, and female participation has been declining since about 2000, and the wage stagnation trend continues. I don’t want to rule out the possibility that this was part of what made 1973 in particular such a strong inflection point, but even if it was, it’s long since been overwhelmed by other factors.

4. Could Apparent Wage Decoupling Be An Artifact Of How We Measure Inflation?

Martin Feldstein is a Harvard economics professor, former head of the National Bureau of Economic Researchers, former head of the President’s Council of Economic Advisors, etc. He believes that apparent wage stagnation is an artifact of mismeasurement.

His argument is pretty hard for me to understand, but as best I can tell, it goes like this. In order to calculate wage growth since 1973, we take the nominal difference in wages, then adjust for inflation. We calculate wage inflation with something called the Consumer Price Index, which is the average price of lots of different goods and services.

But in order to calculate productivity growth since 1973, we use a different index, the “nonfarm business sector output price index”, which is based on how much money companies get for their products.

These should be similar if consumers are buying the same products that companies are making. But there can be some differences. For example, if you’re looking at US statistics only, then some businesses may be selling to foreign markets with different inflation rates, and some consumers may be buying imported goods from countries with different inflation rates. Also (and I’m not sure I understand this right), if people own houses, CPI pretends they are paying market rent to avoid ignoring housing costs, but PPI doesn’t do this. Also, PPI is not as good at including services as CPI. So consumer and producer price indexes differ.

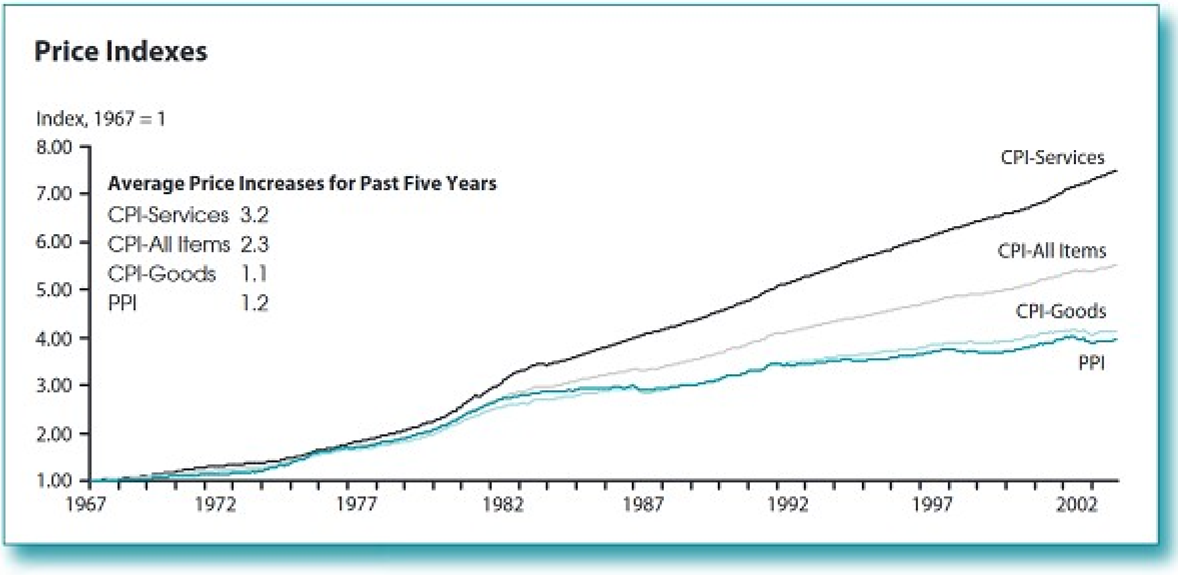
In fact, consumer inflation has been larger than producer inflation since 1973. So when we adjust wages for consumer inflation, they go way down, but when we adjust productivity for producer-inflation, it only goes down a little. This means that these different inflation indices make it look like productivity has risen much faster than wages, but actually they’ve risen the same amount.

As per Feldstein:

> The level of productivity doubled in the U.S. nonfarm business sectorbetween 1970 and 2006. Wages, or more accurately total compensation per hour, increased at approximately the same annual rate during that period if nominal compensation is adjusted for inflation in the same way as the nominal output measure that is used to calculate productivity.

> More specifically, the doubling of productivity represented a 1.9 percent annual rate of increase. Real compensation per hour rose at 1.7 percent per year when nominal compensation is deflated using the same nonfarm business sector output price index. In the period since 2000, productivity rose much more rapidly (2.9 percent ayear) and compensation per hour rose nearly as fast (2.5 percent a year).

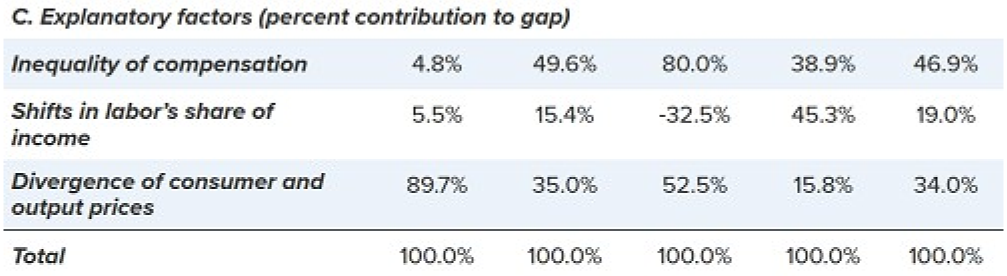
Why is the CPI increasing so much faster than business-centered inflation indices?



The Federal Reserve blames tech. The services-centered CPI has comparatively little technology. The goods-heavy PPI (a business-centered index of inflation) has a lot of it. Tech is going down in price (how much did a digital camera cost in 1990? How about now?) so the PPI stays very low while the CPI keeps growing.

How much does this matter?

The left-leaning Economic Policy Institute says it explains 34% of wage decoupling:



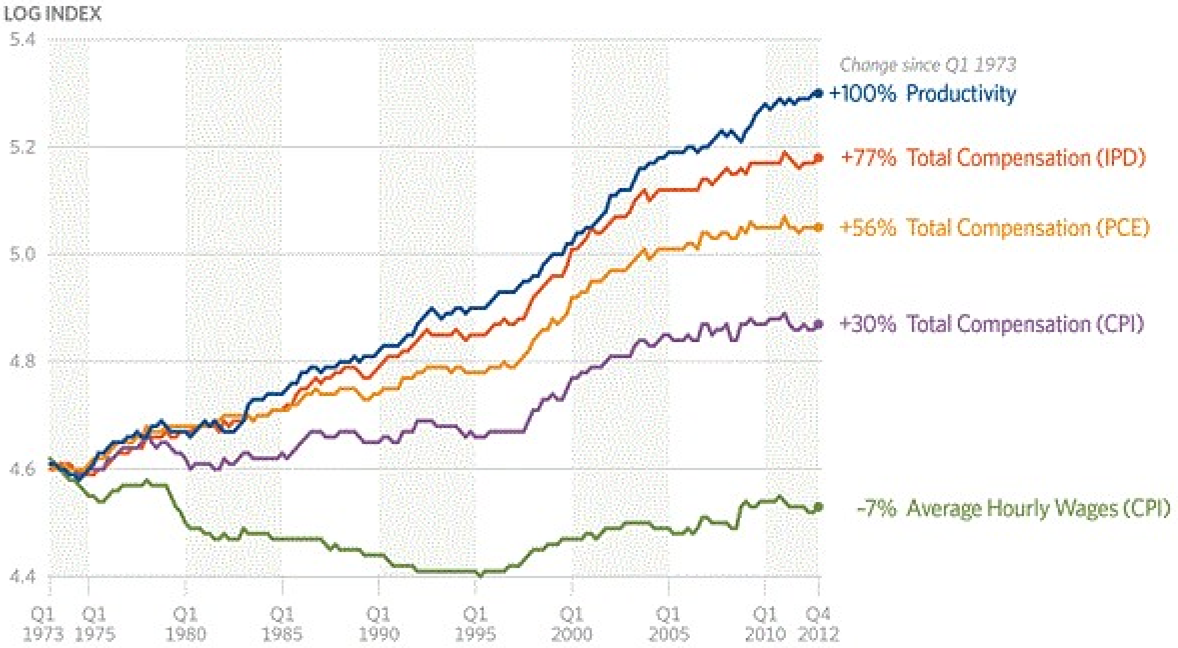
C. Explanatory factors (percent contribution to gap)

Inequality of compensation	4.8%	49.6%	80.0%	38.9%	46.9%
Shifts in labor's share of income	5.5%	15.4%	-32.5%	45.3%	19.0%
Divergence of consumer and output prices	89.7%	35.0%	52.5%	15.8%	34.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%

The right-leaning Heritage Foundation says it explains more:

CHART 6

Comparing Productivity with Total Compensation, Inflation-Adjusted with the Implicit Price Deflator (IPD)



Source: Heritage Foundation calculations using data from the U.S. Department of Labor, Bureau of Labor Statistics, "Productivity and Costs," http://www.bls.gov/lpc/data.htm (accessed June 15, 2013); and Bureau of Labor Statistics, Current Establishment Survey/Haver Analytics. Hourly wages are from the Current Establishment Survey and are defined as the average hourly earnings of production and nonsupervisory employees on nonfarm businesses.

B 2825 ☎ heritage.org

If we estimate the size of the gap as 70 pp (between total compensation CPI and productivity), switching to the top IPD measure closes 67% of the gap; switching to the PCE measure explains 37% of the gap. I'm confused because the EPI is supposedly based on Mishel and Gee, who say they have used the GDP deflator, which is the same thing as the IPD which the Heritage Foundation says they use. I think the difference is that Mishel and Gee haven’t already applied the change from wages to total compensation when they estimate percent of the gap closed? But I’m not sure.

One other group has tried to calculate this: Pessoa and Van Reenan: Decoupling Of Wage Growth And Productivity Growth? Myth And Reality. According to a summary I read, they believe 40% of wage decoupling is because of these inflation related concerns, but I have trouble finding that number in the paper itself.

And the CBO looks into the same issue. They’re not talking about it relative to productivity, but they say that technical inflation issues mean that the standard wage stagnation story is false, and wages have really grown 41% from 1979 – 2013. Since productivity increased somewhere between 70% and 100% during that time, this seems similar to some of the other estimates – inflation technicalities explain between 1/3 and 2/3 of the problem.

Everyone I read seems to agree this issue exists and is interesting, but I’m not sure I entirely understand the implications. Some people say that this completely debunks the idea of wage decoupling and it’s actually only half or a third what the raw numbers say. Other people seem to agree that a big part of wage decoupling is these inflation technicalities, but suggest that although they have important technical implications, if you want to know how the average worker on the street is doing the CPI is still the way to go.

Superstar economist Larry Summers (with Harvard student Anna Stansbury) comes the closest to having a real opinion on this here:

> When investigating consumers' experienced rise in living standards as in Bivens and Mishel (2015), a consumer price deflator is appropriate; however, as Feldstein (2008) argues, when investigating factor income shares a producer price deflator is more appropriate because it reflects the real cost to firms of employing workers.

I am a little confused by this. On the one hand, I do want to investigate consumers' experienced rise (or lack thereof) in living standards. This is the whole point – the possibility that workers' living standards haven't risen since 1973. But most people nowadays work in services. If you deflate their wages with an index used mostly for goods, are you just being a moron and ensuring you will be inaccurate?

Summers and Stansbury continue:

> Lawrence (2016) analyzes this divergence more recently, comparing average compensation to net productivity, which is a more accurate reflection of the increase in income available for distribution to factors of production. Since depreciation has accelerated over recent decades, using gross productivity creates a misleadingly large divergence between productivity and compensation. Lawrence finds that net labor productivity and average compensation grew together until 2001, when they started to diverge i.e. the labor share started to fall. Many other studies also find a decline in the US labor share of income since about 2000, though the timing and magnitude is disputed (see for example Grossman et al 2017, Karabarbounis and Neiman 2014, Lawrence 2015, Elsby Hobijn and Sahin 2013, Rognlie 2015, Pessoa and Van Reenen 2013).

If I intepret this correctly, it looks like it's saying that the real decoupling happened in 2000, not in 1973. I see a lot of papers saying the same thing, and I don't understand where they're diverging from the people who say it happened in 1973. Maybe they're using Feldstein's method of calculating inflation? I think this must be true – if you look at the Heritage Foundation graph above, "total compensation measured with Feldstein's method" and productivity are exactly equal to their 1973 level in 2000, but diverge shortly thereafter so that today compensation has only grown 77% compared to productivity's 100%.

Nevertheless, Summers and Stansbury go on to give basically exactly the same "Why have wages been basically stagnant since 1973? Why are they decoupled from productivity?" narrative as everyone else, so it sure doesn't look like they think any of this has disproven that. It looks like maybe they think Feldstein is right in some way that doesn't matter? But I don't know enough economics to figure out what that way would be. And it looks like Feldstein believes his rightness matters very much, and other economists like Scott Sumner seem to agree. And I cannot find anyone, anywhere, who wants to come out and say explicitly that Feldstein's argument is wrong and we should definitely measure wage stagnation the way everyone does it.

My conclusions from this section, such as they are, go:

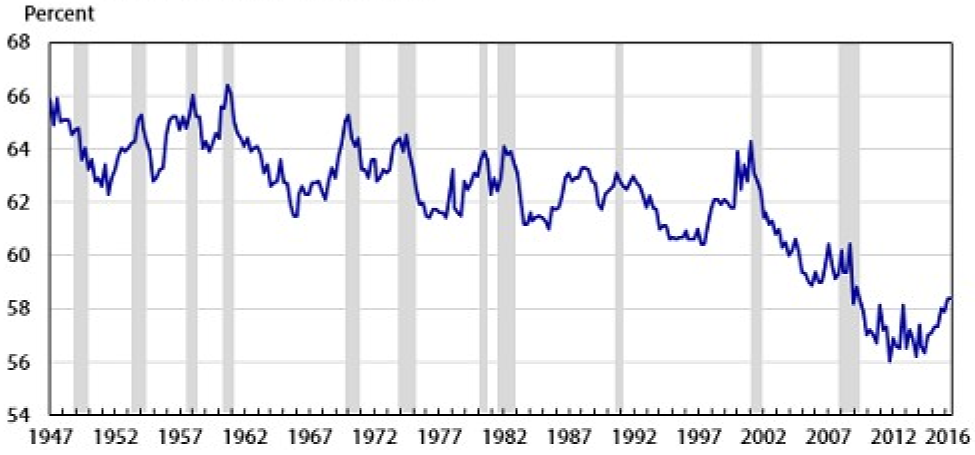
1. Arcane technical points about inflation might explain between 33% and 66% of the apparent wage stagnation/decoupling.
2. "Explain" may not mean the same as "explain away", and it's not completely clear how these points relate to anything we care about

5. Could Wage Decoupling Be Explained By Increasing Labor-Vs-Capital Inequality?

Economists divide inequality into two types. Wage inequality is about how much different wage-earners (or salary-earners, here the terms are used interchangeably) make relative to each other. Labor-vs-capital inequality is about how much wage earners earn vs. how much capitalists get in profits. These capitalists are usually investors/shareholders, but can also be small business owners (or, sometimes, large business owners). Since tycoons like Jeff Bezos and Mark Zuckerberg get most of their compensation from stocks, they count as “capitalists” even if they are paid some token salary for the work they do running their companies.

Here is the labor-vs-capital split for the US over the relevant time period; note the very truncated vertical axis:

Figure 1. Labor's share of output in the nonfarm business sector, first quarter 1947 through third quarter 2016



Note: Shaded areas indicate recessions, as determined by the National Bureau of Economic Research.

Source: U.S. Bureau of Labor Statistics.

This type of inequality was about the same in the early 1970s as in the early 2000s, and has no clear inflection point around 1973, so it probably didn’t start this trend off.

But it did start seriously decreasing around 2000, the same time people who use the more careful inflation methodology say wages and productivity *really* decoupled. And obviously labor getting less money in general is the sort of thing that makes wages go down.

Why is labor-vs-capital inequality increasing? For the long story, read Piketty (my review, highlights, comments). But the short story includes:

Today's wage inequality is tomorrow's labor-vs-capital inequality. If some people get paid more than others, they can invest, their savings will compound, and they will have more capital. As wage inequality increases (see below), labor-vs-capital inequality does too.

The tech industry is more capital-intensive than labor-intensive. For example, Apple has 100,000 employees and makes $250 billion/year, compared to WalMart with 2 million employees and $500 billion/year – in other words, Apple makes $2.5 million per employee compared to Wal-Mart's $250,000. Apple probably pays its employees more than Wal-Mart does, but not ten times more. So more of Apple's revenue goes to capital compared to Wal-Mart's. As tech becomes more important than traditional industries, capital's share of the pie goes up. This is probably a big reason why capital has been doing so well since 2000 or so.

There's an iconoclastic strain of thought that says most of the change in labor-vs-capital is just housing. Houses count as capital, so as housing costs rise, so does capital's share of the economy. Read Matt Ronglie's work (paper, summary, Voxsplainer) for more. Since houses are neither involved in corporate productivity nor in wages, I'm not sure how this affects wage-productivity decoupling if true.

Whatever the cause, the papers I read suggest that increasing labor-vs-capital inequality explains maybe 10-20% of of decoupling, almost all concentrated in the 2000 – present period.

6. Could Wage Decoupling Be Explained By Increasing Wage Inequality?

The other part of the two-pronged inequality picture above. This one seems more important.

One way economists look at this is in the difference between the *median* wage and the *average* wage:

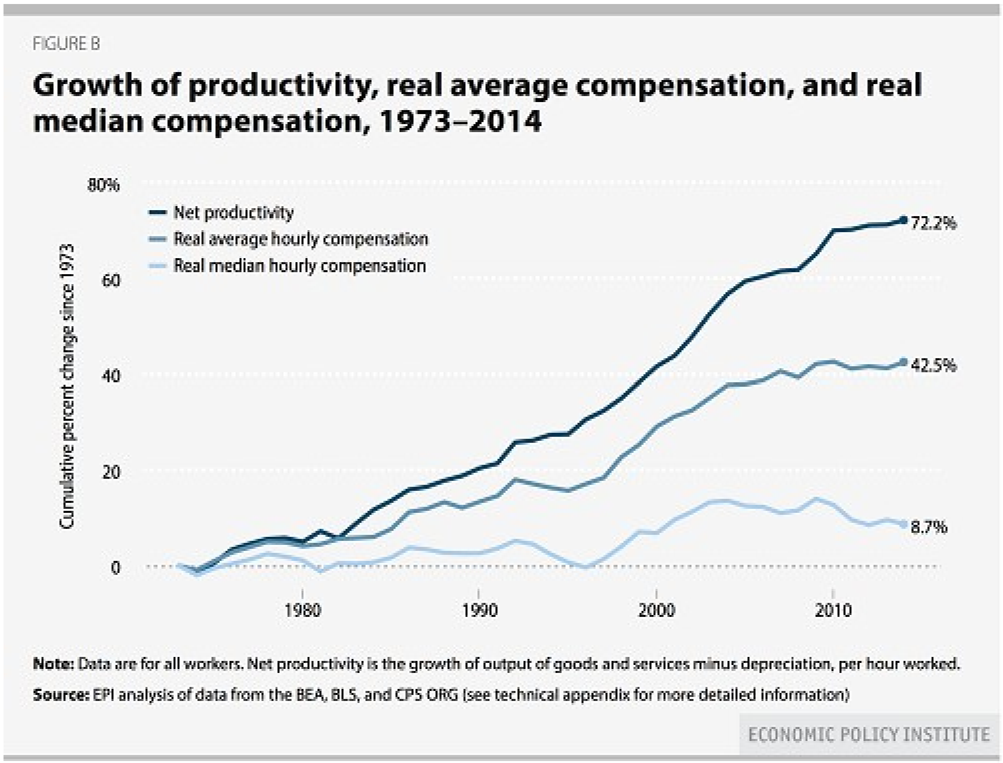


FIGURE B

Growth of productivity, real average compensation, and real median compensation, 1973–2014

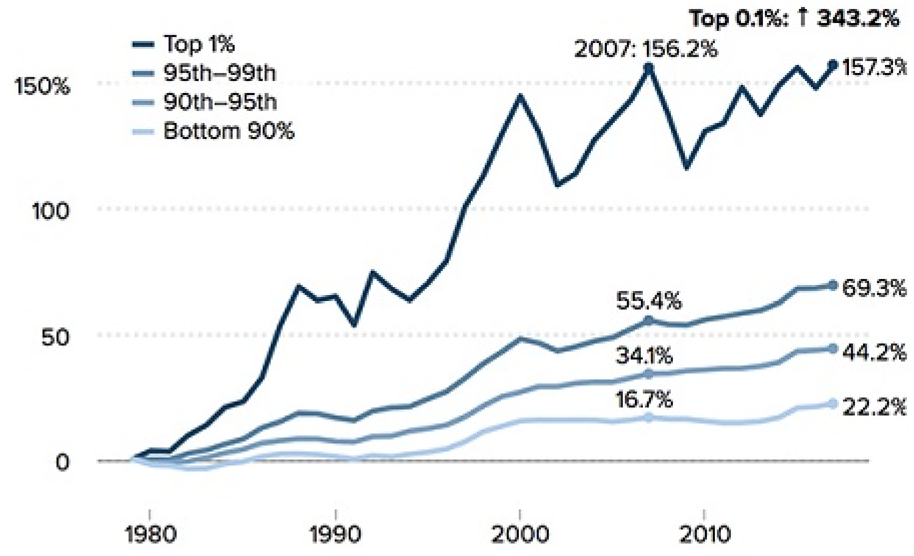
Note: Data are for all workers. Net productivity is the growth of output of goods and services minus depreciation, per hour worked.

Source: EPI analysis of data from the BEA, BLS, and CPS ORG (see technical appendix for more detailed information)

Add in the other things we talked about – the health insurance, the inflation technicalities, the declining share of labor – and the “””average””” worker is doing almost as well as they were in 1973. In fact, this is almost tautologically true. If the entire pie is growing by X amount, and labor’s relative share of the pie is staying the same, then labor should be getting the same absolute amount, and (ignoring changes in the number of laborers) the average laborer should get the same amount.

So the decline in median wage is a mean vs. median issue. A few high-earners are taking a lot of the pie, keeping the mean constant but lowering the median. How high?

Cumulative percent change in real annual wages, by wage group, 1979–2017



Source: EPI analysis of Kopczuk, Saez, and Song (2010, Table A3) and Social Security Administration wage statistics

Economic Policy Institute

Remember, productivity has grown by 70-100% through this period. So even though the top 5% have seen their incomes grow by 69%, they’re still not growing as fast as productivity. The top 1% have grown a bit faster than productivity, although still not that much. The top 0.1% are doing really well.

This is generally considered the most important cause of wage stagnation and wage decoupling, other than among the iconoclasts who think the inflation issues are more important. Above, I referred to a few papers that tried to break down the importance of each cause. EPI thinks wage inequality explains 47% of the problem. Pessoa and Van Reenen think it explains more like 20% according to Mishel’s summary (my eyeballing of the paper suggests more like 33%, but I am pretty uncertain about this).

7. Is Wage Inequality Increasing Because Of Technology?

Here’s one story about why wage inequality is increasing.

In the old days, people worked in factories. A slightly smarter factory worker might be able to run the machinery a little better, or do something else useful, but in the end everyone is working on the same machines.

In the modern economy, factory workers are being replaced by robots. This creates very high demand for skilled roboticists, who get paid lots of money to run the robots in the most efficient way, and very low demand for factory workers, who need to be retrained to be fast food workers or something.

Or, in the general case, technology separates people into the winners (the people who are good with technology and who can use it to do jobs that would have taken dozens or hundreds of people before) and the losers (people who are not good with technology, and so their jobs have been automated away).

From an OECD paper:

> Common explanations for increased wage inequality such as skill-biased technological change and globalisation cannot plausibly account for the

> disproportionate wage growth at the very top of the wage distribution. Skill-biased technological change and globalisation may both raise the relative demand for high-skilled workers, but this should be reflected in broadly rising relative wages of high-skilled workers rather than narrowly rising relative wagesof top-earners. Brynjolfsson and McAfee (2014) argue that digitalisation leads to "winner-takes-most" dynamics, with innovators reaping outsize rewards as digital innovations are replicable at very low cost and have a global scale. Recent studies provide evidence consistent with "winner-take-most" dynamics, in the sense that productivity of firms at the technology frontier has diverged from the remaining firms and that market shares of frontier firms have increased (Andrews et al., 2016). This type of technological change may allow firms at the technology frontier to raise the wages of its key employees to "superstar" levels.

It…sounds like they're saying that technological change can't be the answer, then giving arguments for why the answer is technological change.

I think this is just the authors' poor writing skills, and that the real argument is less confusing. The *Huffington Post* is surprisingly helpful, describing it as:

> What this means is that skilled professionals are not just winning out over working class stiffs, but the richest of the top 0.01 percent are winning out over the professional class as a whole.

That Larry Summers paper mentioned before becomes relevant here again. It argues that wages and productivity are not decoupled – which I know is a pretty explosive thing to say three thousand words in to an essay on wage decoupling, but let’s hear him out.

He argues that apparent decoupling between productivity and wages could result either from literal decoupling – that is, none of the gains of increasing productivity going to workers – or from unrelated trends – for example, increasing productivity giving workers an extra $1000 at the same time as something else causes workers to lose $1000. If a company made $1000 extra and the boss pocketed all of it and didn’t give workers any, that would be literal decoupling. If a company made $1000 extra, it gave workers $1000 extra, but globalization means there’s less demand for workers and so salaries would otherwise have dropped by $1000, so now they stay the same, that’s an unrelated trend.

Summers and Stansbury investigate this by seeing if wages increase more during the short periods between 1973 and today when productivity is unusually high, and if they stagnate more (or decline) during the short periods when it is unusually low. They find this is mostly true:

> We find substantial evidence of linkage between productivity and compensation: Over 1973–2016, one percentage point higher productivity growth has been associated with 0.7 to 1 percentage points higher median and average compensation growth and with 0.4 to 0.7 percentage points higher production/nonsupervisory compensation growth.

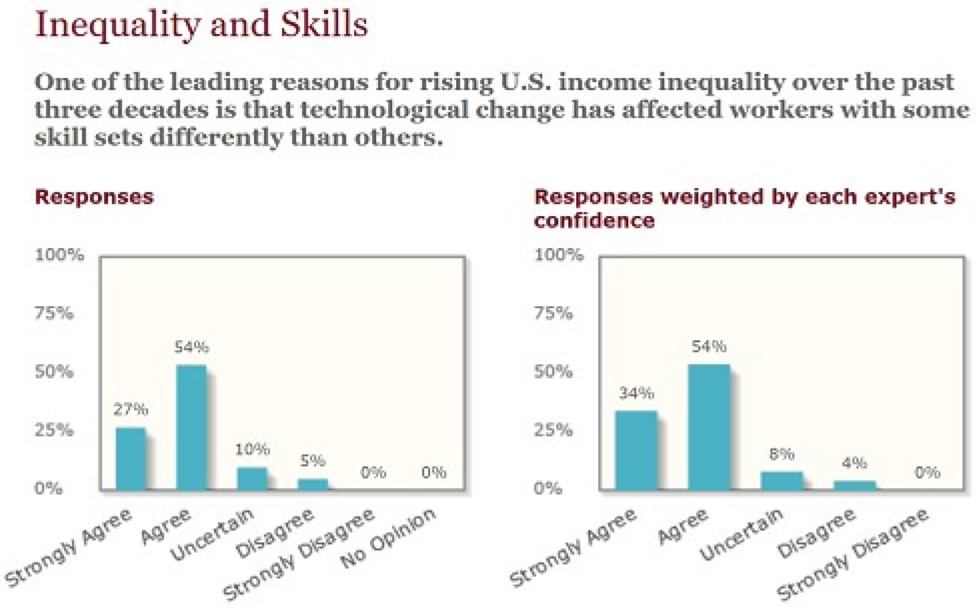
S&S are very careful in this paper and have already adjusted for health insurance issues and inflation calculation issues. They find that once you adjust for this, productivity and wages are between 40% and 100% coupled, depending on what measure you use. (I don’t exactly understand the difference between the two measures they give; surely taking the median worker is already letting you consider inequality and you shouldn't get so much of a difference by focusing on nonsupervisory workers?) As mentioned before, they finds the coupling is much less since 2000. They also find similar results in most other countries: whether or not those countries show apparent decoupling, they remain pretty coupled in terms of actual productivity growth:wage growth correlation.

They argue that if technology/automation were causing rising wage inequality or rising labor-capital inequality, then median wage should decouple from productivity fastest during the periods of highest productivity growth. After all, productivity growth represents the advance of labor-saving technology. So periods of high productivity growth are those where the most new technology and automation are being deployed, so if this is what's driving wages down, wages should decrease fastest during this time.

They test this a couple of different ways, and find that it is false before the year 2000, but somewhat true afterwards, mostly through labor-capital inequality. They don't really find that technology drives wage inequality at all.

I understand why technology would mean decoupling happens fastest during the highest productivity growth. But I'm not sure I understand what they mean when they say there is no decoupling and productivity growth translates into wage growth? Shouldn't this disprove *all* posited causes of decoupling so far, including policy-based wage inequality? I'm not sure. S&S don't seem to think so, but I'm not sure why. Overall I find this paper confusing, but I assume its authors know what they're doing so I will accept its conclusions as presented.

So it sounds like, although technology probably explains some top-10% people doing moderately better than the rest, it doesn't explain the stratospheric increase in the share of the 1%, which is where most of the story lies. I would be content to dismiss this as unimportant, except that…



…all the world's leading economists disagree.

Maybe when they say "income inequality", they're talking about a more intuitive view of income inequality where some programmers make $150K and some factory workers make $30K and this is unequal and that's important – even though it is not related to the larger problem of why everybody except the top 1% is making much less than predicted. I'm not sure.

I feel bad about dismissing so many things as "probably responsible for a few percent of the problem". It seems like a cop-out when it's hard to decide whether something is really important or not. But my best guess is still that this is probably responsible for a few percent of the problem.

8. Is Wage Inequality Increasing Because Of Policy Changes?

Hello! We are every think tank in the world! We hear you are wondering whether wage inequality is increasing because of policy changes! Can we offer you nine billion articles proving that it definitely is, and you should definitely be very angry? Please can we offer you articles? Pleeeeeeeeaaase?!

Presentations of this theory usually involve some combination of policies – decreasing union power, low minimum wages, greater acceptance of very high CEO salary – that concentrate all gains in the highest-paid workers, usually CEOs and executives.

I have trouble making the numbers add up. Vox has a cute thought experiment here where they imagine the CEO of Wal-Mart redistributing his entire salary to all Wal-Mart workers equally, possibly after having been visited by three spirits. Each Wal-Mart employee would make an extra $10. If the spirits visited all top Wal-Mart executives instead of just the CEO, the average employee would get $30. This is obviously not going to single-handedly bring them to the middle-class.

Vox uses such a limited definition of “top executive” that only five people are included. What about Wal-Mart’s 1%?

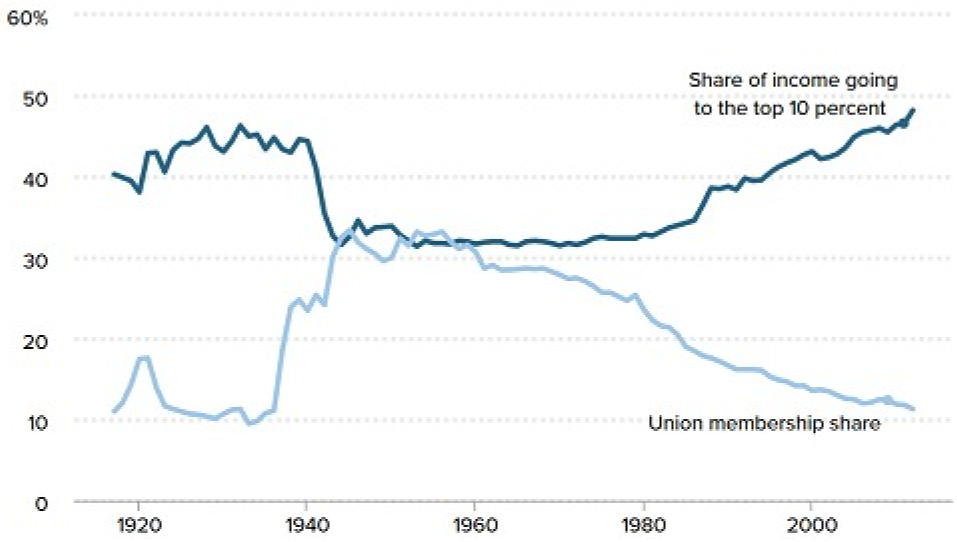
The Wal-Mart 1% will include 20,000 people. To reach the 1% in the US, you need to make $400,000 per year; I would expect Wal-Mart’s 1% to be lower, since Wal-Mart is famously a bad place to work that doesn’t pay people much. Let’s say $200,000. That means the Wal-Mart 1% makes a total of $4 billion. If their salary were distributed to all 2 million employees, those employees would make an extra $2,000 per year; maybe a 10% pay raise. And of course even in a perfectly functional economy, we couldn't pay Wal-Mart management literally $0, so the real number would be less than this.

Maybe the problem is that Wal-Mart is just an unusually employee-heavy company. What about Apple? Their CEO makes $12 million per year. If that were distributed to their 132,000 employees, they would each make an extra $90.

How many total high-paid executives does Apple have? It looks like Apple hires up to 130 MBAs from top business schools per year; if we imagine they last 10 years each, they might have 1000 such people, making them a "top 1%". If these people get paid $500,000 each, they could earn 500 million total. That's enough to redistribute $4,000 to all Apple employees, which still isn't satisfying given the extent of the problem.

Some commenters bring up the possibility that I'm missing stocks and stock options, which make up most of the compensation of top executives. I'm not sure whether this gets classified as income (in which case it could help explain income inequality) or as capital (in which case it would get filed under labor-vs-capital inequality). I'm also not sure whether Apple giving Tim Cook lots of stocks takes money out of the salary budget that could have gone to workers instead. For now let's just accept that the difference between mean and median income shows that *something* has to be happening to drive up the top 1% or so of salaries.

What policies are most likely to have caused this concentration of salaries at the top?

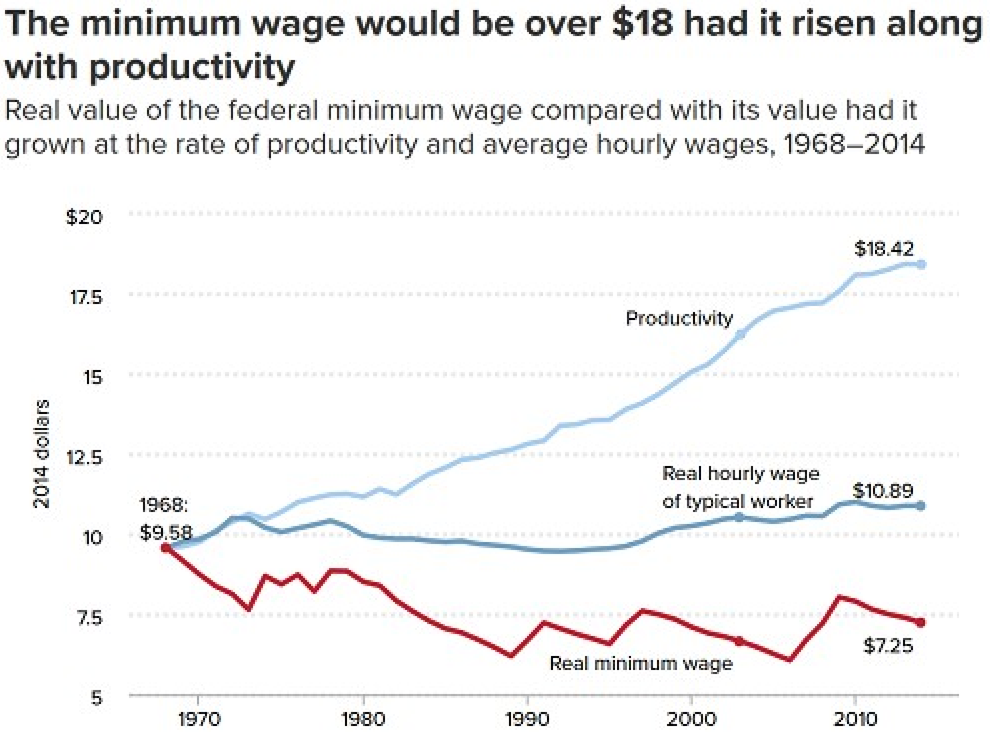


Decline in union membership mirrors income gains of top 10%

Union membership and share of income going to the top 10%, 1917–2012

Many people point to a decline in unions. This decline does not really line up with the relevant time period – it started in the early 1960s, when productivity and wages were still closely coupled. But it could be a possible contributor. Economics Policy Institute cites some work saying it may explain up to 10% of decoupling *even for non-union members*, since the deals struck by unions set norms that spread throughout their industries. A group of respected economists including David Card looks into the issue and finds similar results, saying that the decline of unions may explain about 14% or more of increasing wage inequality (remember that wage inequality is only about 40% of decoupling, so this would mean it only explains about 5% of decoupling). The conservative Heritage Foundation has many bad things to say about unions but grudgingly admits they may raise salaries by up to 10% among members (they don't address non-members). Based on all this, it seems plausible that deunionization may explain about 5-10% of decoupling.

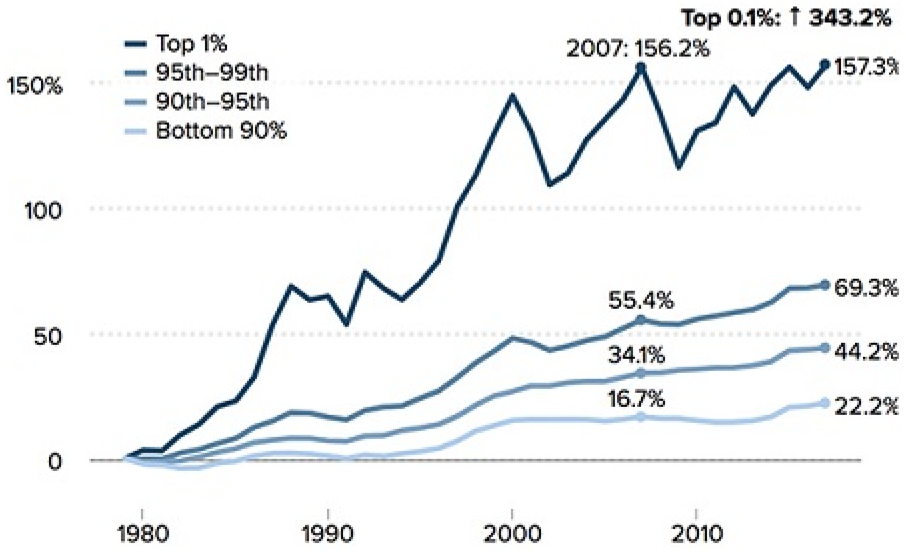
Another relevant policy that could be shaping this issue is the minimum wage. EPI notes that although the minimum wage never goes down in nominal terms, if it doesn't go up then it's effectively going down in real terms and relative to productivity. This certainly sounds like the sort of thing that could increase wage inequality.



The minimum wage would be over $18 had it risen along with productivity

Real value of the federal minimum wage compared with its value had it grown at the rate of productivity and average hourly wages, 1968–2014

But let's look at that graph by percentiles again:



Cumulative percent change in real annual wages, by wage group, 1979–2017

Source: EPI analysis of Kopczuk, Saez, and Song (2010, Table A3) and Social Security Administration wage statistics

Economic Policy Institute

Wage stagnation is barely any better for the 90th percentile worker than it is for the people at the bottom. And the 90th percentile worker isn’t making minimum wage. This may be another one that adds a percentage point here and there, but it doesn’t seem too important.

I can’t find anything about it on EPI, but Thomas Piketty thinks that tax changes were an important driver of wage inequality. I’ll quote my previous review of his book:

> He thinks that executive salaries have increased because – basically – corporate governance isn’t good enough to prevent executives from giving themselves very high salaries. Why didn’t executives give themselves such high salaries before? Because before the 1980s the US had a top tax rate of 80% to 90%. As theory predicts, people become less interested in

> making money when the government's going to take 90% of it, so executives didn't bother pulling the strings it would take to have stratospheric salaries. Once the top tax rate was decreased, it became worth executives' time to figure out how to game the system, so they did. This is less common outside the Anglosphere because other countries have different forms of corporate governance and taxation that discourage this kind of thing.

Piketty does some work to show that increasing wage inequality in different countries is correlated with those countries' corporate governance and taxation policies. I don't know if anyone has checked how that affects wage decoupling.

9. Conclusions

1. Contrary to the usual story, wages have not stagnated since 1973. Measurement issues, including wages vs. benefits and different inflation measurements, have made things look worse than they are. Depending on how you prefer to think about inflation, median wages have probably risen about 40% – 50% since 1973, about half as much as productivity.

2. This leaves about a 50% real decoupling between median wages and productivity, which is still enough to be serious and scary. The most important factor here is probably increasing wage inequality. Increasing labor-capital inequality is a less important but still significant factor, and it has become more significant since 2000.

3. Increasing wage inequality probably has a lot to do with issues of taxation and corporate governance, and to some

degree also with issues surrounding unionization. It probably has less to do with increasing technology and automation.

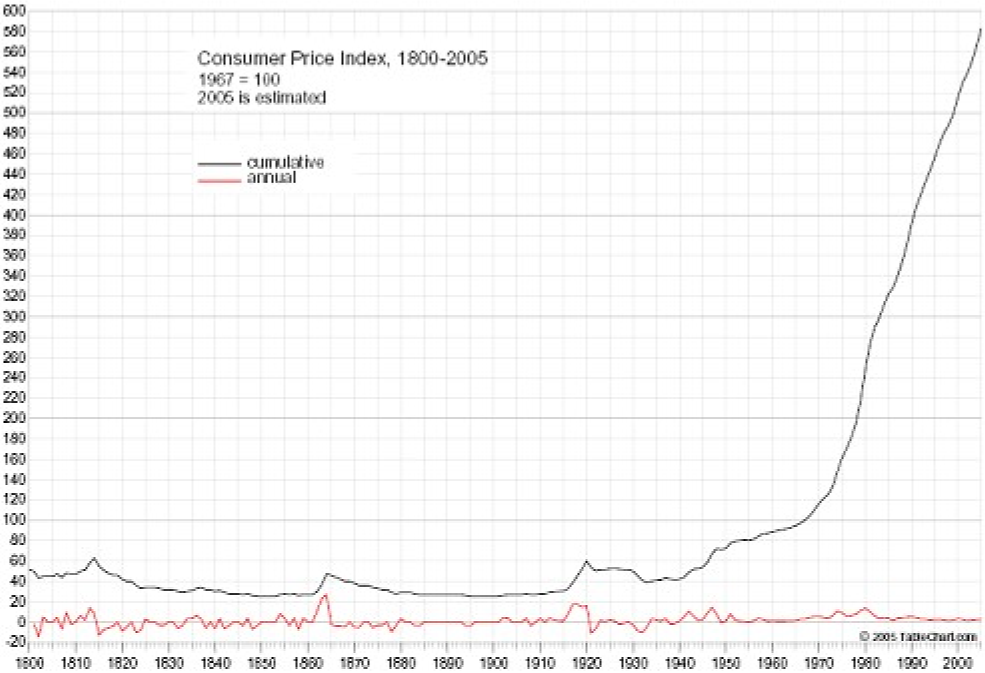
4. If you were to put a gun to my head and force me to break down the importance of various factors in contributing to wage decoupling, it would look something like (warning: very low confidence!) this:

– Inflation miscalculations: 35%
– Wages vs. total compensation: 10%
– Increasing labor vs. capital inequality: 15%
—- (Because of automation: 7.5%)
—- (Because of policy: 7.5%)
– Increasing wage inequality: 40%
—- (Because of deunionization: 10%)
—- (Because of policies permitting high executive salaries: 20%)
—- (Because of globalization and automation: 10%)

This surprises me, because the dramatic shift in 1973 made me expect to see a single cause (and multifactorial trends should be rare in general, maybe, I think). It looks like there are two reasons why 1973 seems more important than it is.

First, most graphs trying to present this data begin around 1950. If they had begun much earlier than 1950, they would have showed several historical decouplings and recouplings that make a decoupling in any one year seem less interesting.

Second, 1973 was the year of the 1973 Oil Crisis, the fall of Bretton Woods, and the end of the gold standard, causing a general discombobulation to the economy that lasted a couple of years. By the time the economy recombobulated itself again, a lot of trends had the chance to get going or switch direction. For example, here’s inflation:



5. Inflation issues and wage inequality were probably most important in the first half of the period being studied. Labor-vs-capital inequality was probably most important in the second half.

6. Continuing issues that confuse me:
– How much should we care about the difference between inflation indices? If we agree that using CPI to calculate this is dumb, should we cut our mental picture of the size of the problem in half?
– Why is there such a difference between the Heritage Foundation’s estimate of how much of the gap inconsistent deflators explain (67%) and the EPI’s (34%)? Who is right?
– Does the Summers & Stansbury paper argue against policy-based wage inequality as a cause of median wage stagnation, at least until 2000?
– Are there enough high-paid executives at companies that, if their money were redistributed to all employees, their compensation would have increased significantly more in step with productivity? If so, where are they hiding? If not, what does “increasing wage inequality explains X% of

decoupling” mean?

– What caused past episodes of wage decoupling in the US? What ended them?

– How do we square the apparent multifactorial nature of wage decoupling with its sudden beginning in 1973 and with the general argument against multifactorial trends?

Rule Thinkers In, Not Out

Imagine a black box which, when you pressed a button, would generate a scientific hypothesis. 50% of its hypotheses are false; 50% are true hypotheses as game-changing and elegant as relativity. Even despite the error rate, it's easy to see this box would quickly surpass space capsules, da Vinci paintings, and printer ink cartridges to become the most valuable object in the world. Scientific progress on demand, and all you have to do is *test* some stuff to *see if it's true*? I don't want to devalue experimentalists. They do great work. But it's appropriate that Einstein is more famous than Eddington. If you took away Eddington, someone else would have tested relativity; the bottleneck is in Einsteins. Einstein-in-a-box at the cost of requiring two Eddingtons per insight is a heck of a deal.

What if the box had only a 10% success rate? A 1% success rate? My guess is: still most valuable object in the world. Even an 0.1% success rate seems pretty good, considering (what if we ask the box for cancer cures, then test them all on lab rats and volunteers?) You have to go pretty low before the box stops being great.

I thought about this after reading this list of geniuses with terrible ideas. Linus Pauling thought Vitamin C cured everything. Isaac Newton spent half his time working on weird Bible codes. Nikola Tesla pursued mad energy beams that couldn't work. Lynn Margulis revolutionized cell biology by discovering mitochondrial endosymbiosis, but was also a 9-11 truther and doubted HIV caused AIDS. Et cetera. Obviously this should happen. Genius often involves coming up with an outrageous idea contrary to conventional wisdom and pursuing it obsessively despite naysayers. But nobody can have a 100% success rate. People who do this successfully sometimes should also fail at it sometimes, just because they're the kind of person who attempts it at all. Not everyone fails. Einstein seems to have batted a perfect 1000 (unless you count his support for socialism). But failure shouldn't *surprise* us.

Yet aren't some of these examples unforgiveably bad? Like, seriously Isaac – Bible codes? Well, granted, Newton's chemical experiments may have exposed him to a little more mercury than can be entirely healthy. But remember: gravity was considered creepy occult pseudoscience by its early enemies. It subjected the earth and the heavens to the same law, which shocked 17th century sensibilities the same way trying to link consciousness and matter would today. It postulated that objects could act on each other through invisible forces at a distance, which was equally outside the contemporaneous Overton Window. Newton's exceptional genius, his exceptional ability to think outside all relevant boxes, and his exceptionally egregious mistakes are all the same phenomenon (plus or minus a little mercury).

Or think of it a different way. Newton stared at problems that had vexed generations before him, and noticed a subtle pattern everyone else had missed. He must have amazing hypersensitive pattern-matching going on. But people with such hypersensitivity should be most likely to see patterns where they don't exist. Hence, Bible codes.

These geniuses are like our black boxes: generators of brilliant ideas, plus a certain failure rate. The failures can be easily discarded: physicists were able to take up Newton's gravity without wasting time on his Bible codes. So we're right to treat geniuses as valuable in the same way we would treat those boxes as valuable.

This goes not just for geniuses, but for anybody in the idea industry. Coming up with a genuinely original idea is a rare skill, much harder than judging ideas is. Somebody who comes up with one good original idea (plus ninety-nine really stupid cringeworthy takes) is a better use of your reading time than somebody who reliably never gets anything too wrong, but never says anything you find new or surprising. Alyssa Vance calls this positive selection – a single good call rules you in – as opposed to negative selection, where a single bad call rules you out. You should practice positive selection for geniuses and other intellectuals.

I think about this every time I hear someone say something like "I lost all respect for Steven Pinker after he said all that stupid stuff about AI". Your problem was thinking of "respect" as a relevant predicate to apply to Steven Pinker in the first place. Is he your father? Your youth pastor? No? Then why are you worrying about whether or not to "respect" him? Steven Pinker is a black box who occasionally spits out ideas, opinions, and arguments for you to evaluate. If some of them are arguments you wouldn't have come up with on your own, then he's doing you a service. If 50% of them are false, then the best-case scenario is that they're moronically, obviously false, so that you can reject them quickly and get on with your life.

I don't want to take this too far. If someone has 99 stupid ideas and then 1 seemingly good one, obviously this should increase your probability that the seemingly good one is actually flawed in a way you haven't noticed. If someone has 99 stupid ideas, obviously this should make you less willing to waste time reading their other ideas to see if they are really good. If you want to learn the basics of a field you know nothing about, obviously read a textbook. If you don't trust your ability to figure out when people are wrong, obviously read someone with a track record of always representing the conventional wisdom correctly. And if you’re a social engineer trying to recommend what other people who are less intelligent than you should read, obviously steer them away from anyone who’s wrong too often. I just worry too many people wear their social engineer hat so often that they forget how to take it off, forget that “intellectual exploration” is a different job than “promote the right opinions about things” and requires different strategies.

But consider the debate over “outrage culture”. Most of this focuses on moral outrage. Some smart person says something we consider evil, and so we stop listening to her or giving her a platform. I don’t want to argue this one right now – at the very least it disincentivizes evil-seeming statements.

But I think there’s a similar phenomenon that gets less attention and is even less defensible – a sort of intellectual outrage culture. “How can you possibly read that guy when he’s said [stupid thing]?” I don’t want to get into defending every weird belief or conspiracy theory that’s ever been [stupid thing]. I just want to say it probably wasn’t as stupid as Bible codes. And yet, Newton.

Some of the people who have most inspired me have been inexcusably wrong on basic issues. But you only need one world-changing revelation to be worth reading.

In Mod We Trust

The Verge writes a story (an exposé?) on the Facebook-moderation industry.

It goes through the standard ways it maltreats its employees: low pay, limited bathroom breaks, awful managers – and then into some not-so-standard ones. Mods have to read (or watch) all of the worst things people post on Facebook, from conspiracy theories to snuff videos. The story talks about the psychological trauma this inflicts:

> It’s an environment where workers cope by telling dark jokes about committing suicide, then smoke weed during breaks to numb their emotions…where employees, desperate for a dopamine rush amid the misery, have been found having sex inside stairwells and a room reserved for lactating mothers…
>
> It’s a place where the conspiracy videos and memes that they see each day gradually lead them to embrace fringe views. One auditor walks the floor promoting the idea that the Earth is flat. A former employee told me he has begun to question certain aspects of the Holocaust. Another former employee, who told me he has mapped every escape route out of his house and sleeps with a gun at his side, said: “I no longer believe 9/11 was a terrorist attack.

One of the commenters on Reddit asked “Has this guy ever worked in a restaurant?” and, uh, fair. I don’t want to speculate on how much weed-smoking or sex-in-stairwell-having is due to a psychological reaction to the trauma of awful Facebook material vs. ordinary shenanigans. But it sure does seem traumatic.

Other than that, the article caught my attention for a few reasons.

First, because I recently wrote a post that was a little dismissive of moderators, and made it sound like an easy problem. I think the version I described – moderation of a single website's text-only comment section – is an easi-*er* problem than moderating all of Facebook and whatever horrible snuff videos people post there. But if any Facebook moderators, or anyone else in a similar situation, read that post and thought I was selling them short, I'm sorry.

Second, because the article gives a good explanation of *why* Facebook moderators' job is so much harder and more unpleasant than my job or the jobs of the mods I work with: they are asked to apply a set of rules so arcane that the article likens them to the Talmud, then have their decisions nitpicked to death – with career consequences for them if higher-ups think their judgment calls on edge cases were wrong.

While I was writing the article on the Culture War Thread, several of the CW moderators told me that the hard part of their job wasn't keep the Thread up and running and well-moderated, it was dealing with the constant hectoring that they had made the wrong decision. If they banned someone, people would say the ban was unfair and they were tyrants and they hated freedom of speech. If they *didn't* ban someone, people would say they tolerated racism and bullying and abuse, or that they were biased and would have banned the person if they'd been on the other side.

Me, I handle that by not caring. I've made it clear that this blog is my own fiefdom to run the way I like, and that disagreeing with the way I want a comment section to look is a perfectly reasonable decision – which should be followed by going somewhere other than my blog's comment section.

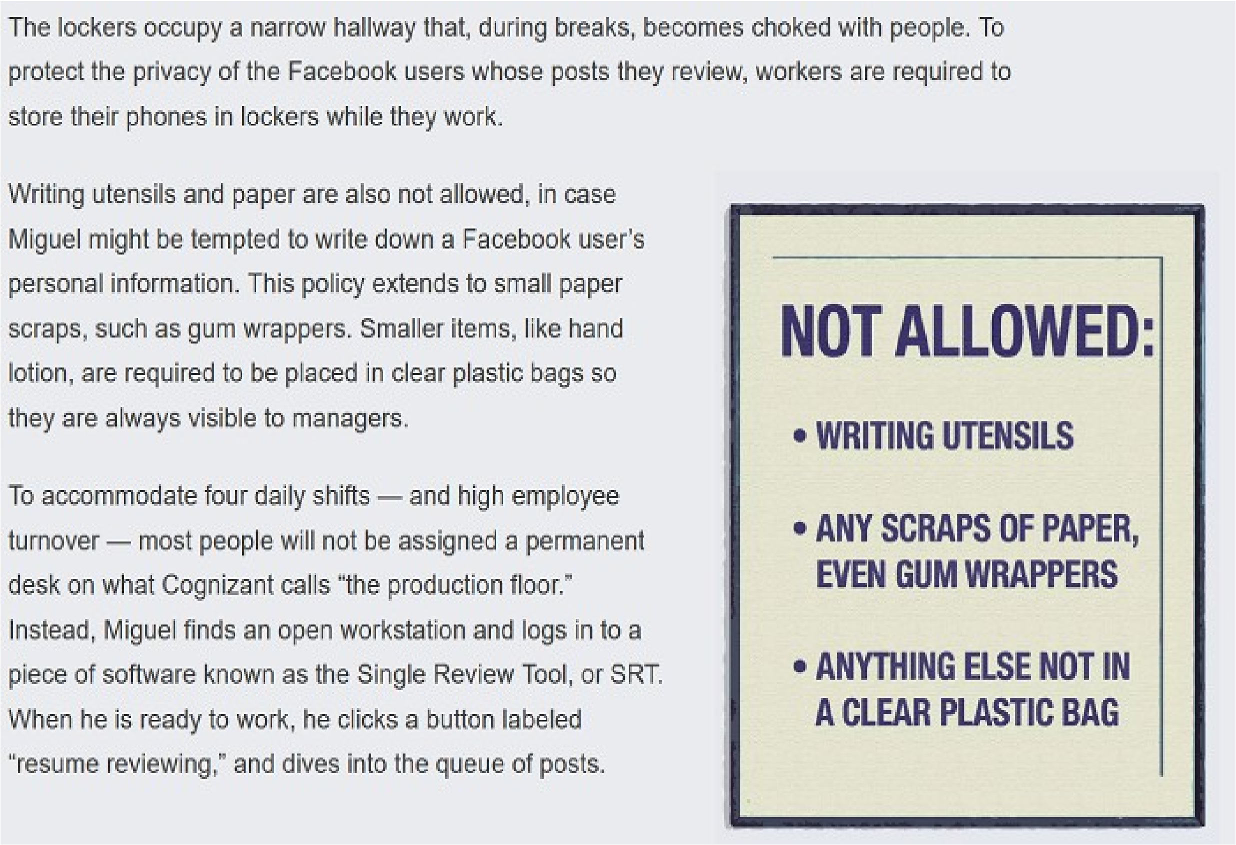
Most of my commenters have been respectful of that, I think it’s worked out very well, and my experience moderating thousands of comments per week is basically a breeze.

Obviously this gets harder when you have hundreds of different moderators, none of whom are necessarily preselected for matching Facebook HQ’s vision of “good judgment”. It also gets harder when you’re a big company that wants to keep users, and your PR department warns you against telling malcontents to “go take a hike”. It gets harder still when you host X0% of all online discussion, you’re one step away from being a utility or a branch of government or something, and you have a moral responsibility to shape the world’s conversation in a responsible way – plus various Congressmen who will punish you if you don’t. The way Facebook handles moderation seems dehumanizing, but I don’t know what the alternative is, given the pressures they’re under.

(I don’t know if this excuses sites like the *New York Times* saying they can’t afford moderators; I would hope they would hire one or two trusted people, then stand by their decisions no matter what.)

Third, I felt there was a weird tension in this article, and after writing that last paragraph I think I know what it is. This was a good piece of investigative reporting, digging up many genuinely outrageous things. But most of them are necessary and unavoidable responses to the *last* good piece of investigative reporting, and all the outrageous things *it* dug up. Everything The Verge is complaining about is Facebook’s attempt to defend itself against publications like The Verge.

Take, for example, the ban on phones, writing utensils, and gum wrappers:



The lockers occupy a narrow hallway that, during breaks, becomes choked with people. To protect the privacy of the Facebook users whose posts they review, workers are required to store their phones in lockers while they work.

Writing utensils and paper are also not allowed, in case Miguel might be tempted to write down a Facebook user's personal information. This policy extends to small paper scraps, such as gum wrappers. Smaller items, like hand lotion, are required to be placed in clear plastic bags so they are always visible to managers.

To accommodate four daily shifts — and high employee turnover — most people will not be assigned a permanent desk on what Cognizant calls "the production floor." Instead, Miguel finds an open workstation and logs in to a piece of software known as the Single Review Tool, or SRT. When he is ready to work, he clicks a button labeled "resume reviewing," and dives into the queue of posts.

NOT ALLOWED:

- WRITING UTENSILS
- ANY SCRAPS OF PAPER, EVEN GUM WRAPPERS
- ANYTHING ELSE NOT IN A CLEAR PLASTIC BAG

The Verge brings this up as an example of the totalitarian and dehumanizing environment that Facebook moderators experience. But I imagine that if an employee had written down (or used their phone to take a picture of) some personal details of a Facebook user, The Verge (or some identical publication) would have run a report on how Facebook hired contractors who *didn't even take basic precautions to protect user privacy*.

And what about the absolutist, infinitely-nitpicky rules that every moderator has to follow (and be double- and triple-checked to have followed) on each decision? Again, totalitarian and dehumanizing, no argument there. But if a moderator screwed up – if one of them banned a breastfeeding picture as "explicit", and the Facebook Talmud hadn't include twelve pages of exceptions and counterexceptions for when breasts were and weren't allowed – I imagine reporters would be on that story in a split second. They would be mocking Facebook's "lame excuse" that it was just one moderator acting alone and not company policy, and leading the demands for Facebook to put "procedures" in place to ensure it never happens again.

If I sound a little bitter about this, it's because I spent four years working at a psychiatric hospital, helping create the most dehumanizing and totalitarian environment possible. It wasn't a lot of fun. But you could trace every single rule to somebody's lawsuit or investigative report, and to some judge or jury or news-reading public that decided it was *outrageous* that a psychiatric hospital hadn't had a procedure in place to prevent whatever occurred from occurring. Some patient in Florida hit another patient with their book and it caused brain damage? Well, that's it, nobody in a psych hospital can ever have a hardcover book again. Some patient in Delaware used a computer to send a threatening email to his wife? That's it, psych patients can never use the Internet unless supervised one-on-one by a trained Internet supervisor with a college degree in Psychiatric Internet Supervision, which your institution cannot afford. Some patient in Oregon managed to hang herself in the fifteen minute interval between nurses opening her door at night to ask "ARE YOU REALLY ASLEEP OR ARE YOU TRYING TO COMMIT SUICIDE IN THERE?" Guess nurses will have to do that every ten minutes now. It was all awful, and it all created a climate of constant misery, and it was all 100% understandable under the circumstances.

I'm not saying nobody should ever be allowed to do investigative reporting or complain about problems. But I would support some kind of anti-irony rule, where you're not allowed to make extra money writing another outrage-bait article about the outrages your first outrage-bait article caused.

But maybe this is unfair. “Complete safety from scandal, or humanizing work environment – pick one” doesn’t seem quite right. High-paid workers sometimes manage to do sensitive jobs while still getting a little leeway. When I worked in the psychiatric hospital, I could *occasionally* use my personal authority to suspend the stupidest and most dehumanizing rules. I don’t know if they just figured that medical school selected for people who could be trusted with decision-making power, or if I was high-ranking enough that everyone figured my scalp would be enough to satisfy the hordes if I got it wrong. But it sometimes went okay.

And lawyers demonstrate a different way that strict rules can coexist with a humanizing environment; they have to navigate the most complicated law code there is, but I don’t get the impression that they feel dehumanized by their job.

(but maybe if the government put as much effort into preventing innocent people from going to jail as Facebook puts into preventing negative publicity, they would be worse off.)

It seems like The Verge’s preferred solution, a move away from “the call center model” of moderation, might have whatever anti-dehumanization virtue doctors and lawyers have. Overall I’m not sure how this works, but it prevents me from being as snarky as I would be otherwise.

(except that I worry in practice this will look like “restrict the Facebook moderation industry to people with college degrees”, and we should think hard before we act like this is doing society any favors)

Last, I find this article interesting because it presents a pessimistic view of information spread. Normal people who are exposed to conspiracy theories – without any social connection to the person spouting them, or any pre-existing psychological vulnerabilities that make them seek the conspiracy theories out – end up believing them or at least suspecting. This surprises me a little. If it’s true, how come more people haven’t been infected? How come Facebook moderators don’t believe the debunking of the conspiracy theories instead? Is it just that nobody ever reports those for mod review? Or is this whole phenomenon just an artifact of every large workplace (the article says “hundreds” of people work at Cognizant) having one or two conspiracy buffs, and in this case the reporter hunted them down because it made a better story?

Just to be on the safe side, every time someone shares an SSC link, report it as violating the Facebook terms of service. We’ll make rationalists out of these people yet!

Meaningful

[With apologies to Putnam, Pope, and all of you]

Two children are reading a text written by an AI:

> The boys splashed water in each other's faces until they were both sopping wet

One child says to the other "Wow! After reading some text, the AI understands what water is!"

The second child says "It doesn't *really* understand."

The first child says "Sure it does! It understands that water is the sort of substance that splashes. It understands that people who are splashed with water get wet. What else is left to understand?"

The second child says "All it understands is relationships between words. None of the words connect to reality. It doesn't have any internal concept of what water looks like or how it feels to be wet. Only that the letters W-A-T-E-R, when appearing near the letters S-P-L-A-S-H bear a certain statistical relationship to the letters W-E-T."

The first child starts to cry.

Two chemists are watching the children argue with each other. The first chemist says "Wow! After seeing an AI, these kids can debate the nature of water!"

The second chemist says "Ironic, isn't it? After all, the children themselves don't understand what water is! Water is two hydrogen atoms plus one oxygen atom, and neither of them know!"

The first chemist answers “Come on. The child knows enough about water to say she understands it. She knows what it looks like. She knows what it tastes like. That’s pretty much the basics of water.”

The second chemist answers “Those are just relationships between pieces of sense-data. The child knows that (visual perception of clear shiny thing) = (tactile perception of cold wetness) = (gustatory perception of refreshingness). And she can predict statistical relationships – like, if she sees someone throw a bucket of (visual perception of clear shiny thing) at her, she will soon feel (tactile perception of cold miserable sopping wetness). She uses the word “water” as a concept-hook that links all of these relationships together and makes predicting the world much easier. But no matter how well she masters these facts, she can never connect them to $\mathrm{H2O}$ or any other real chemical facts about the world beyond mere sense-data.”

The first chemist says “Maybe she knows things like that water makes iron rust. That’s a chemical fact.”

The second chemist says “No, she knows that (clear shiny appearance + wetness + refreshment) makes (dull metallic appearance + hardness) get (patchy redness). She doesn’t know that $\mathrm{H2O} + \mathrm{Fe}$ = iron oxides. She knows many statistical relationships between sense-data, but none of them ever connect to the deeper chemical reality.”

The first chemist says “Then on what level can we be said to understand water ourselves? After all, no doubt there are deeper things going on than chemical reactions – quantum fields, superstrings, levels even deeper than those. All we know are some statistical relationships that must hold true, despite whatever those things may be.”

Two angels are watching the chemists argue with each other. The first angel says “Wow! After seeing the relationship between the sensory and atomic-scale worlds, these chemists have realized that there are levels of understanding humans are incapable of accessing.”

The second angel says “They haven’t truly realized it. They’re just abstracting over levels of relationship between the physical world and their internal thought-forms in a mechanical way. They have no concept of or . You can’t even express it in their language!”

The first angel says “Yes, but when they use placeholder words like ‘levels even deeper than those’, those placeholders will have the same statistical relationship with the connection between models and reality as .”

“Yes, which is the difference between being able to respond to ‘Marco!’ by shouting ‘Polo!” vs. a deep historical understanding of Europe-Orient trade relations in the Middle Ages. If all you know is that some statistical models are isomorphic to other models and to Creation itself, you still won’t have the slightest idea what the s of any of them are.”

“I’m not claiming humans really know what anything means,” said the first angel. “Just that it’s impressive you can get that far by manipulating a purely symbolic mental language made of sense-data-derived thought-forms with no connection to real at all.”

“I guess that is kind of impressive,” said the second angel. “For humans.”

God sits in the highest heaven, alone.

“Wow!” He thinks to Himself, “that cellular automaton sure is producing some pretty patterns today. I wonder what it will do next!”

Prospiracy Theories

[Title from this unrelated story or this unrelated essay]

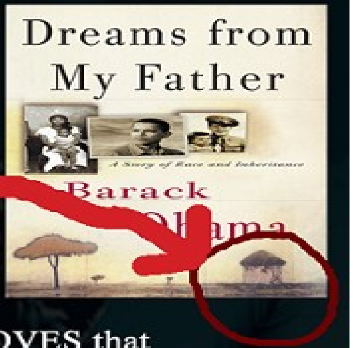
Last week I wrote about how conspiracy theories spread so much faster on Facebook than debunkings of those same theories. A few commenters chimed in to say that of *course* this was true, the conspiracy theories had evolved into an almost-perfect form for exploiting cognitive biases and the pressures of social media. Debunkings and true beliefs couldn’t copy that process, so they were losing out.

This sounded like a challenge, so here you go:

WHAT IF "THE FIRST AFRICAN AMERICAN PRESIDENT"...

...WASN'T FROM AFRICA AT ALL?

Throughout his campaign, Barack Obama played up his exotic background.



Dreams from My Father

A Story of Race and Inheritance

Barack Obama

African hut = humble origins?

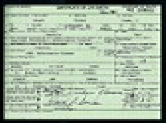
But new evidence PROVES that "Barack Hussein Obama" from Kenya...

Note "HAWA" on uniform - nowhere in Kenya begins with those letters!

..is really just Barry Obama from Hawaii!

No reporter has EVER found a Kenyan who remembers knowing Obama as a child

Obama's father got a scholarship to University of Hawaii in 1959 - just two years before Obama was born!



If he was really born in Kenya, where's his birth certificate?

Obama's mother Ann has light skin, like many Americans - but UNLIKE almost all Kenyans!

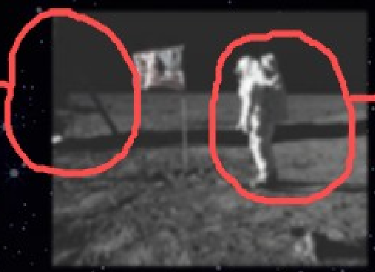
Immigrant-loving, anti-American liberals would be DEVASTED to learn their precious Obama is no immigrant at all, but from America- the very country they are against!

SlateStarCodex.com

ANCIENT ASTRONAUTS?

in 2015, Elon Musk said he wanted humans to land on other worlds

but what if they already did . . . almost fifty years agp?



dark line resembles landing gear of UFO

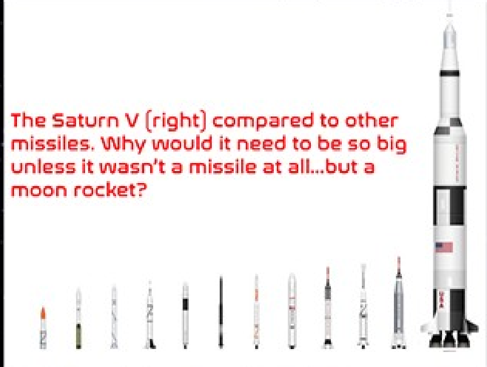
compare this man's clothing to the suits used on spacewalks from the ISS.

this picture comes from the archives of "Project Apollo", an above-top-secret US government program that used captured Nazi technology to reach the Moon in the 1960s.

Nazi scientist Werner von Braun developed unmanned spaceships that could strike anywhere without warning. But the war ended before they could be exploited.

as part of a secret program to win the Cold War, President Kennedy ordered the construction of a spacecraft that could claim the Moon for America.

the result was the Saturn V. The existence of the Saturn V is a matter of public record...but the story that it was just another nuclear missile doesn't make sense! Space historian Amy Teitel believes the Saturn V brought astronauts Neil Armstrong and Buzz Aldrin to the moon in 1969 - before the non-classified space program had even invented the space shuttle!



The Saturn V (right) compared to other missiles. Why would it need to be so big unless it wasn't a missile at all...but a moon rocket?

...notice the resemblance between Nazi super-scientist Wehrner von Braun and the designer of the Saturn V, whose name was also Wehrner von Braun!

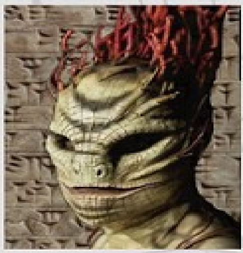
this is Buzz Aldrin punching a man who denied humans could reach the moon. Why would he be so defensive about this...unless he had been to the moon himself?

have you ever noticed that the special effects in Kubrick's moon landing scenes are too good for the 1960s? Some people think the story of Kubrick filming a moon landing at all is a government coverup...meant to hide that a moon landing REALLY HAPPENED!

SlateStarCodex.com

Does the Queen really have superior space alien DNA that justifies her right to rule over you? Or is she an

ORDINARY HUMAN JUST LIKE US?



Godlike reptilian aliens may have brought civilization to Sumeria. **But the Queen has no link to these noble lawgivers.**

Newly-discovered documents from the British Library show that the so-called "Queen" is just Elizabeth Windsor, **an ordinary British woman.**

A secretive faction within the British government brainwashed Elizabeth from birth, turning her into a "perfect monarch" who would **never question** her own right to rule.

What would happen if the people knew their queen - who gets palaces, jewels, and millions of pounds of public support - was no Babylonian divinity, but a normal woman who had done nothing to deserve any of these things? Would there be riots? Protest? REVOLUTION???

Learn more at
www.SlateStarCodex.com

EARTH is a SPHERE PROVE ME WRONG

PLATO said Sphere is the MOST PERFECT of all objects. The Diquis left giant stone Spheres through North America in 500 BC - WHY? BIBLE refers to " circle of the Earth" in Isaiah 40:22!

NASA Photos prove EARTH is a Sphere with GREENLAND on the top and ANTARCTICA on the bottom. South America is along the Midline of the Sphere - ECUADOR = EQUATOR! The Sphere spins around a line from THULE in Greenland to AMUNDSEN-SCOTT STATION,Antarctica. The Government built AMUNDSEN-SCOTT STATION as a base to explore the STAR-WINDS that strike the exact Bottom of the Sphere. In the center of the Earth Sphere is a Sphere of iron, then a larger Sphere of fiery rock, then a larger Sphere of solid rock and water where creatures live, then a larger Sphere of air. FOUR SPHERES IN ONE SPHERE, all spinning along an imaginary line from Greenland to AMUNDSEN-SCOTT STATION.

The Moon-Sphere revolves around the Earth-Sphere once every 29.5 Days = SACRED ISLAMIC CALENDAR

The Earth-Sphere revolves around the Sun-Sphere once every 365.24 Days = SACRED CATHOLIC CALENDAR. Pope Gregory Knew!

Four spheres in one sphere creates invisible Field that protects us from the hostile Star-Winds that would destroy all life. Only at AMUNDSEN-SCOTT STATION can the Star-Winds be be seen and measured. The US Government sends scientists to AMUNDSEN-SCOTT STATION every year to monitor the Star-Winds, which is used to predict the Weather and design Electronic Pulse Bomb Weapons that can wipe out entire Cities!

Winds are the turning of the Sphere of AIR. Currents are the turning of the Sphere of WATER. Earthquakes are the turning of the Sphere of FIRE. Protection from the Star-Winds is the turning of the Sphere of IRON. Only at AMUNDSEN-SCOTT STATION do the Spheres align and the Star-Winds break through. This is why only government scientists are allowed to go to AMUNDSEN-SCOTT STATION and everyone else is RESTRICTED. To understand Sphere Earth is to understand why AMUNDSEN-SCOTT STATION is so important. Eratosthenes the Egyptian Sage discovered SPHERE EARTH and wanted to spread the knowledge to everyone. Enlightened philosophers through the Ages have Understood SPHERE EARTH. The Pope andthe Jews and the Professors and the Business Elites are laughing at you if you do not know about SPHERE EARTH. They think you are a Fool!

slatestarcodex.com

Ketamine: Now By Prescription

Last week the FDA approved esketamine for treatment-resistant depression.

Let's review how the pharmaceutical industry works: a company discovers and patents a potentially exciting new drug. They spend tens of millions of dollars proving safety and efficacy to the FDA. The FDA rewards them with a 10ish year monopoly on the drug, during which they can charge whatever ridiculous price they want. This isn't a great system, but at least we get new medicines sometimes.

Occasionally people discover that an existing chemical treats an illness, *without* the chemical having been discovered and patented by a pharmaceutical company. In this case, whoever spends tens of millions of dollars proving it works to the FDA may *not* get a monopoly on the drug and the right to sell it for ridiculous prices. So nobody spends tens of millions of dollars proving it works to the FDA, and so it risks never getting approved.

The usual solution is for some pharma company to make some tiny irrelevant change to the existing chemical, and patent this new chemical as an "exciting discovery" they just made. Everyone goes along with the ruse, the company spends tens of millions of dollars pushing it through FDA trials, it gets approved, and they charge ridiculous prices for ten years. I wouldn't quite call this "the system works", but again, at least we get new medicines.

Twenty years ago, people noticed that ketamine treated depression. Alas, ketamine already existed – it's an anaesthetic and a popular recreational drug – so pharma companies couldn't patent it and fund FDA trials, so it couldn’t get approved by the FDA for depression. A few renegade doctors started setting up ketamine clinics, where they used the existing approval of ketamine for anaesthesia as an excuse to give it to depressed people. But because this indication was not FDA-approved, insurance companies didn’t have to cover it. This created a really embarrassing situation for the medical system: everyone secretly knows ketamine is one of the most effective antidepressants, but officially it’s not an antidepressant at all, and mainstream providers won’t give it to you.

The pharmaceutical industry has lobbyists in Heaven. Does this surprise you? Of *course* they do. A Power bribed here, a Principality flattered there, and eventually their petitions reach the ears of God Himself. This is the only possible explanation for stereochemistry, a quirk of nature where many organic chemicals come in “left-handed” and “right-handed” versions. The details don’t matter, beyond that if you have a chemical that you can’t patent, you can take the left-handed (or right-handed) version, and legally pretend that now it is a different chemical which you *can* patent. And so we got “esketamine”.

Am I saying that esketamine is just a sinister ploy by pharma to patent and make money off ketamine? Yup. In fact “esketamine” is just a cutesy way of writing the chemical name s-ketamine, which literally stands for “sinister ketamine” (sinister is the Latin word for “left-handed”; the modern use derives from the old superstition that left-handers were evil). The sinister ploy to patent sinister ketamine worked, and the latest news says it will cost between $590 to $885 per dose.

(regular old ketamine still costs about $10 per dose, less if you buy it from a heavily-tattooed man on your local street corner)

I’ve said it before: I don’t blame the pharma companies for this. Big Government, in its infinite wisdom, has decided that drugs should have to undergo tens of millions of dollars worth of FDA trials before they get approved. No government agencies or altruistic billionaires have stepped up to fund these trials themselves, so they won’t happen unless some pharma company does it. And pharma companies aren’t going to do it unless they can make their money back. And it’s not like they’re overcharging; their return to investment on R&D may already be less than zero. This is a crappy system – but again, it’s one that occasionally gets us new medicines. So it’s hard to complain.

But in this case, there are two additional issues that make it even worse than the usual serving of crappiness.

First, esketamine might not work.

Johnson & Johnson, the pharma company sponsoring its FDA application, did four official efficacy studies. You can find the summary starting on page 17 of this document. Two of the trials were technically negative, although analysts have noticed nontechnical ways they look encouraging. Two of the trials were technically positive, but one of them was a withdrawal trial that was not really designed to prove efficacy.

The FDA usually demands two positive studies before they approve a drug, and doesn’t usually count withdrawal trials. This time around, in a minor deviation from their usual rules, they decided to count the positive withdrawal trial as one of the two required positives, and approve esketamine. I suspect this was a political move based on how embarrassing it was to have everyone know ketamine was a good antidepressant, but not have it officially FDA-approved.

But if ketamine is such a good antidepressant, how come it couldn't pass the normal bar for approval? Like, people keep saying that ketamine is a *real* antidepressant, that works *perfectly*, and changes *everything*, unlike those bad old SSRIs which are basically just placebo. But esketamine's results are at least as bad as any SSRI's. If you look at Table 9 in the FDA report, ketamine did notably worse than most of the other antidepressants the FDA has approved recently – including vortioxetine, an SSRI-like medication.

One possibility is that ketamine was studied for treatment-resistant depression, so it was only given to the toughest cases. But Table 9 shows olanzapine + fluoxetine doing significantly better than esketamine even for treatment-resistant depression.

Another possibility is that clinical trials are just really tough on antidepressants for some reason. I've mentioned this before in the context of SSRIs. Patients love them. Doctors love them. Clinical trials say they barely have any effect. Well, now patients love ketamine. Doctors love ketamine. And now there's a clinical trial showing barely any effect. This isn't really a *solution* to esketamine's misery, but at least it has company.

Another possibility is that everyone made a huge mistake in using left-handed ketamine, and it's right-handed ketamine that holds the magic. Most previous research was done on a racemic mixture (an equal mix of left-handed and right-handed molecules), and at least one study suggests it was the right-handed ketamine that was driving the results. Pharma decided to pursue left-handed ketamine because it was known to have a stronger effect on NMDA receptors, but – surprise! – ketamine probably doesn't work through NMDA after all. So there's a chance that this is just the wrong kind of ketamine – though usually I expect big pharma to be smarter than that, and I would be surprised if this turned out to be it. I don't know if anybody has a right-handed ketamine patent yet.

And another possibility is that it's the wrong route of administration. Almost all previous studies on ketamine have examined it given IV. The FDA approved esketamine as a nasal spray – which is a lot more convenient for patients, but again, not a lot of studies showing it works. At least some studies seem to show that it doesn't. Again, usually I expect big pharma not to screw up the delivery method, but who knows?

Second in our litany of disappointments, esketamine is going to be maximally inconvenient to get.

The big problem with regular ketamine, other than not being FDA-approved, was that you had to get it IV. That meant going to a ketamine clinic that had nurses and anesthesiologists for IV access, then sitting there for a couple of hours hallucinating while they infused it into you. This was a huge drawback compared to eg Prozac, where you can just bring home a pill bottle and take one pill per day in the comfort of your own bathroom. It's also expensive – clinics, nurses, and anesthesiologists don't come cheap.

The great appeal of a ketamine nasal spray was that it was going to prevent all that. Sure, it might not work. Sure, it would be overpriced. But at least it would be convenient!

The FDA, in its approval for esketamine, specified that it could only be delivered at specialty clinics by doctors who are specially trained in ketamine administration, that patients will have to sit at the clinic for at least two hours, and realistically there will have to be a bunch of nurses on site. My boss has already said our (nice, well-funded) clinic isn't going to be able to jump through the necessary hoops; most other outpatient psychiatric clinics will probably say the same.

This removes most of the advantages of having it be intranasal, so why are they doing this? They give two reasons. First, they want to make sure no patient can ever bring ketamine home, because they might get addicted to it. Okay, I agree addiction is bad. But patients bring prescriptions of OxyContin and Xanax home every day. Come on, FDA. We already *have* a system for drugs you're worried someone will get addicted to, it's called the Controlled Substances Act. Ketamine is less addictive than lots of chemicals that are less stringently regulated than it is. This just seems stupid and mean-spirited.

The other reason the drugs have to be given in a specially monitored clinic is because ketamine can have side effects, including hallucinations and dissociative sensations. I agree these are bad, and I urge patients only to take hallucinogens/dissociatives in an appropriate setting, such as a rave. Like, yeah, ketamine can be seriously creepy, but now patients are going to have to drive to some overpriced ketamine clinic a couple of times a week and sit there for two hours per dose just because you think they're too frail to handle a dissociative drug at home?

I wanted to finally be able to prescribe ketamine to my patients who needed it. Instead, I'm going to have to recommend they find a ketamine clinic near them (some of my patients live hours from civilization), drive to it several times a week (some of my patients don't have cars) and pay through the nose, all so that some guy with a postgraduate degree in Watching People Dissociate can do crossword puzzles while they sit and feel kind of weird in a waiting room. And then those same patients will go home and use Ecstasy. Thanks a lot, FDA.

And the cherry on the crap sundae is that this sets a precedent. If the FDA approves psilocybin for depression (and it’s currently in Phase 2 trials, so watch this space!) you can bet you’re going to have to go to a special psilocybin clinic if you want to get it. Psychedelic medicine is potentially the future of psychiatry, and there’s every indication that it will be as inconvenient and red-tape-filled a future as possible. If you thought it was tough getting your Adderall prescription refilled every month, just wait.

So far, I am continuing to recommend that my patients who want ketamine seek intravenous racemic ketamine at an existing ketamine clinic, since this has a stronger evidence base. Once insurance starts covering esketamine, I may change my mind if money becomes an issue. But I’m annoyed that it’s come to this.

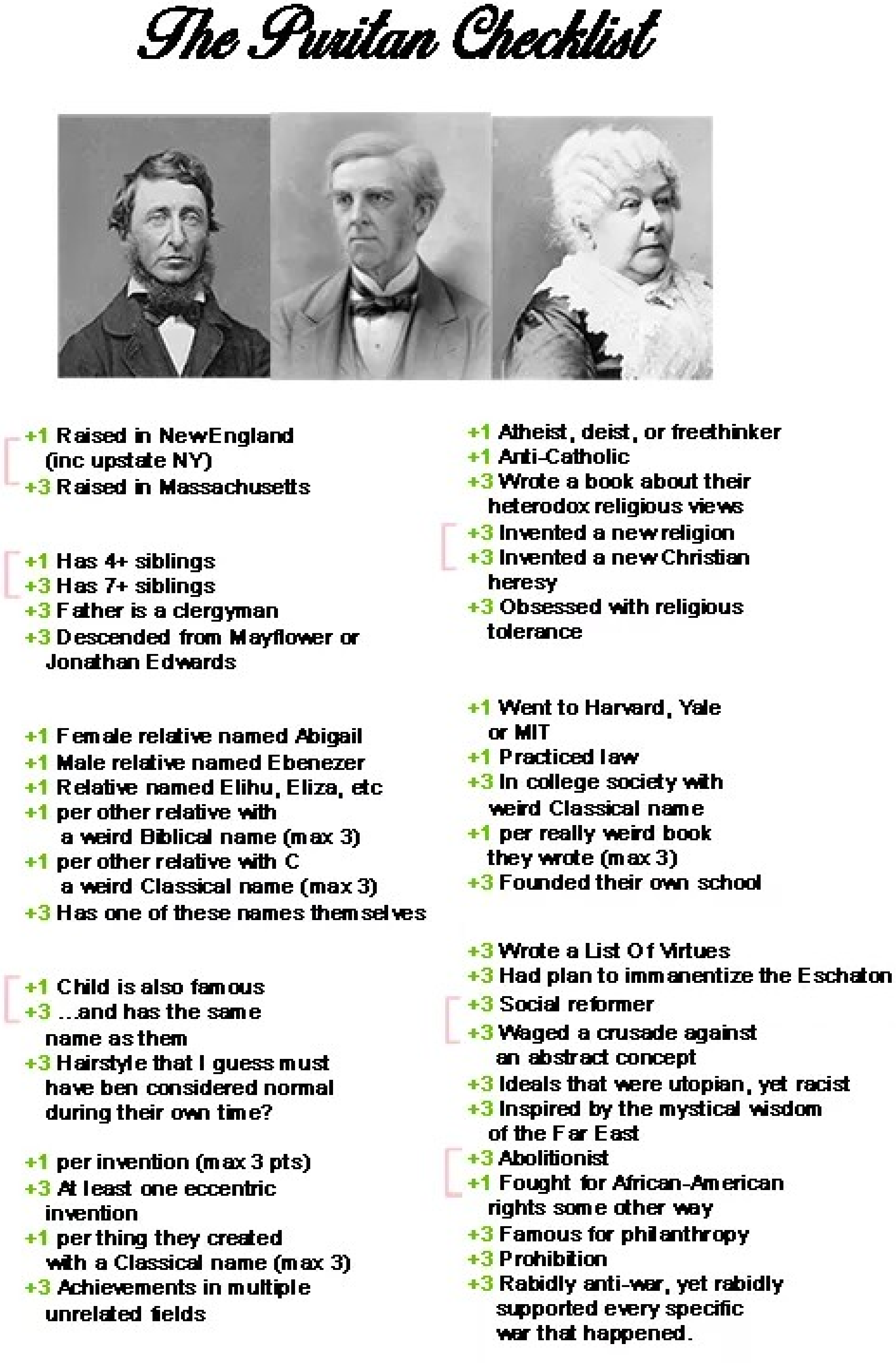
Puritan Spotting

*[**Related to:** Book Review: Albion’s Seed]*
*[**Epistemic status:** Not too serious]*

I realize I’ve been confusing everyone with my use of the word “Puritan”. When I say “That guy is so Puritan!” people object “But he’s not religious!” or “He doesn’t hate fun!”

I don’t know what the real word for the category I’m calling “Puritan” is. Words like “Yankee”, “Boston Brahmin”, or “Transcendentalist” are close, but none of them really work. “Eccentric overeducated hypercompetent contrarian early American who takes morality very seriously” is good, but too long.

Instead of explaining further, here’s a (more than half-joking) Puritan checklist. Maximum one item per red box.



The Puritan Checklist

- +1 Raised in NewEngland (inc upstate NY)
- +3 Raised in Massachusetts

- +1 Has 4+ siblings
- +3 Has 7+ siblings
- +3 Father is a clergyman
- +3 Descended from Mayflower or Jonathan Edwards

- +1 Female relative named Abigail
- +1 Male relative named Ebenezer
- +1 Relative named Elihu, Eliza, etc
- +1 per other relative with a weird Biblical name (max 3)
- +1 per other relative with C a weird Classical name (max 3)
- +3 Has one of these names themselves

- +1 Child is also famous
- +3 ...and has the same name as them
- +3 Hairstyle that I guess must have ben considered normal during their own time?

- +1 per invention (max 3 pts)
- +3 At least one eccentric invention
- +1 per thing they created with a Classical name (max 3)
- +3 Achievements in multiple unrelated fields

- +1 Atheist, deist, or freethinker
- +1 Anti-Catholic
- +3 Wrote a book about their heterodox religious views
- +3 Invented a new religion
- +3 Invented a new Christian heresy
- +3 Obsessed with religious tolerance

- +1 Went to Harvard, Yale or MIT
- +1 Practiced law
- +3 In college society with weird Classical name
- +1 per really weird book they wrote (max 3)
- +3 Founded their own school

- +3 Wrote a List Of Virtues
- +3 Had plan to immanentize the Eschaton
- +3 Social reformer
- +3 Waged a crusade against an abstract concept
- +3 Ideals that were utopian, yet racist
- +3 Inspired by the mystical wisdom of the Far East
- +3 Abolitionist
- +1 Fought for African-American rights some other way
- +3 Famous for philanthropy
- +3 Prohibition
- +3 Rabidly anti-war, yet rabidly supported every specific war that happened.

The obvious next step is to rank historical figures by Puritanism Points. Here are the top five famous Americans I can find, as per Wikipedia:

#5: SAMUEL MORSE

Samuel Morse was born to Pastor **(+3)** Jedediah **(+1)** Morse and his wife Elizabeth **(+1)** in Charlestown, Massachusetts **(+3)**, the eldest of six children **(+3)**. After attending Yale **(+1)**, he pursued a career as an internationally famous painter. But when his wife Lucretia **(+1)** fell sick, he was unable to receive the news in time to go home to her before she died, inspiring him to change careers during mid-life **(+3)** and become an inventor. He spent his life perfecting the telegraph **(+1)**, but also invented an automatic sculpture-making machine **(+3)**. In later life, he switched careers again, becoming an anti-Catholic activist **(+1)**; he ran for Mayor of New York on an anti-Catholic platform, and wrote anti-Catholic pamphlets like *A Foreign Conspiracy Against The Liberties Of The United States* **(+1)**. He was also a well-known philanthropist **(+3)**. His hairstyle looked like this **(+3)**.

Total Puritanism = 28

#4: ELIZABETH CADY STANTON

Elizabeth (**+3**) Cady Stanton was born in Johnstown, New York (**+1**), one of eleven siblings (**+3**) including a brother named Eleazar (**+1**) and a sister named Tryphena (**+1**). She learned Greek as a child (**+1**) and her disillusionment at being barred from higher education because of her gender led her to start a crusade (**+3**) for women’s rights, along with other timely causes like abolitionism (**+3**) and temperance (**+3**). Although she was an agnostic herself (**+1**), she did write *The Women’s Bible* explaining why the Bible should have been more feminist (**+3**). She described herself as a pacifist, but during the Spanish-American War, stated that “Though I hate war per se, I am glad that it has come in this instance. I would like to see Spain swept from the face of the earth” (**+3**). Her hairstyle looked like this (**+3**)

Total Puritanism: 29

#3: LYSANDER SPOONER

Lysander (**+3**) Spooner was born to Asa and Dolores Spooner in Athol, Massachusetts (**+3**), the second of nine children (**+3**) including his elder brother Leander (**+1**). He is best remembered as one of the founders of modern libertarianism, and as the developer of the Non-Aggression Principle (**+3**). But he also had a brief career as a lawyer (**+1**) – brief because he was practicing law illegally, without a license, because he thought licensing restrictions were illegal government tyranny. Later he founded a mail delivery company, again illegally, because he thought the Post Office was illegal government tyranny (I can’t believe he doesn’t gain any points for this; I need a better checklist), and invented (**+1**) a new monetary system (**+3**) because he thought that the existing monetary system was illegal government tyranny (see eg his pamphlets *Gold and Silver as Standards of Value: The Flagrant Cheat in Regard to Them*, **+1**). Among his other works were pamphlets on his idiosyncratic religious views like *The Deist’s Immortality, And An Essay On Man’s Accountability For His Belief* (**+1**, **+1**, **+3**), and a whole host of abolitionist books and pamphlets like *A Plan For The Abolition Of Slavery* (**+1**, **+3**). His hairstyle looked like this **(+3)**.

Total Puritanism = 31

#2: ROGER BABSON

Roger Babson was born to 10th-generation Massachussetts natives **(+3)** Nathaniel **(+1)** and Ellen **(+1)** Babson. After attending MIT **(+1)**, he pursued a career as a businessman, investor, and philanthropist **(+3)**. His charitable efforts included the founding of two colleges **(+3)** – Webber University and Utopia College – and erecting a set of giant boulders with exhortations to be virtuous on them (**+3**). In later life, he switched careers (**+3**) to become a social reformer in the Open Church Movement **(+3)** and run for President as the candidate of the Prohibition Party (**+3**); he also invented the parking meter (**+1**). He is perhaps best remembered for founding an organization to destroy gravity (**+3**, but only because I can't give + infinity without it being unfair to everyone else), and wrote various essays on the topic with titles like *Gravity – Our Enemy Number One* **(+1)**. His hairstyle looked like this **(+3)**.

Total Puritanism = 32

#1: BENJAMIN FRANKLIN

Born in Massachussetts (**+3**), son of Josiah (**+1**) and Abiah (**+1**) Franklin, one of seventeen siblings (**+3**, but deserves more) including a brother Ebenezer (**+1**). He was too poor to go to college, but handled his own education, creating a 7 x 13 Table of Virtues that he used to guide his daily studies and behavior (**+3**). He became an inventor, developing not only the Franklin stove (**+1**), the lightning rod (**+1**), and bifocals (**+1**), but also a system of propelling naval vessels by giant kites (**+3**). Later he switched careers (**+3**) to become a Founding Father of the United States (**+3**) and leader of the American Revolution **(+3)**. He also wrote books like *Poor Richard's Almanack* (**+1**), *Advice To A Friend On Choosing A Mistress* **(+1)**, and *The Means And Manner Of Obtaining Virtue* (this one I *am* giving infinity points). Called himself a Deist (**+1**) and wrote a pamphlet (**+3**) explaining his idiosyncratic semi-Christian beliefs (**+3**); he also wrote a Bible fanfic in which God explained to Abraham the importance of Tolerance (**+3**). He was President of the Pennsylvania Abolitionist Society **(+3)**, but also invented his own racial categorization system where only Anglo-Saxons were white, and German immigration should be banned as a threat to the whiteness of America (**+3**); he nevertheless founded philanthropic organizations to help German immigrants (**+3**). He was sympathetic to pacifism and said that “There never was a good war or a bad peace”, but supported the Revolutionary War (**+3**), which he thought necessary. His hairstyle looked like this **(+3)**.

Total Puritanism = infinity

Other high scorers: Dorothea Dix (in addition to her psychiatric reforms, founded several schools and wrote a whole book of overwrought poetry praising flowers), Hiram Maxim (his son, also named Hiram Maxim, was also a famous inventor), Hiram Bingham (his son, also named Hiram Bingham, was also a famous traveler), Aaron Burr (grandson of Jonathan Edwards; his son was a chairman of the Moral Reform Society), Mary Baker Eddy (“My favorite studies were natural philosophy, logic, and moral science. From my brother Albert, I received lessons in the ancient tongues, Hebrew, Greek, and Latin”), Oliver Wendell Holmes Sr (close to my heart as a doctor/essayist/reformer) and Henry David Thoreau (everything about him). Let me know in the comments if you have more.

I find this kind of thing fun, and I do better than chance at guessing whether people are Puritan or not before I know the answer. But I don’t want to say this is Objectively Right. I’m sure the checklist combines traits that are inherently Puritan (like being from Massachusetts) with ones that are more common among various groups of early Americans (like inventing things, or having large families), which means this has elements of the wiggin fallacy.

The most egregious false positive I’ve found is Mark Twain. He was born in Missouri, one of seven children (**+3**) including an older brother Orion (**+1** – he also had a brother named Pleasant Hannibal, which doesn't get him Puritanism Points but ought to get him *something*). Along with being a writer who wrote various humorous books (**+3**), he was also (**+3**) an inventor, and received patents for suspenders, a history trivia game, and a self-pasting scrapbook (**+3**). He was an abolitionist (**+3**) and Deist (**+1**), and wrote various books about his idiosyncratic views on religion (**+3**). He described himself as a pacifist, but supported revolutionary violence from Robespierre to the Russian communists (**+3**). His hairstyle looked like this (**+3**).

This gets him 26, which is pretty concerning. But it's worth noting that his great-great-grandfather was Ezekiel Clemens of Essex, Massachussetts, so he does have some Puritan blood in him, however diluted.

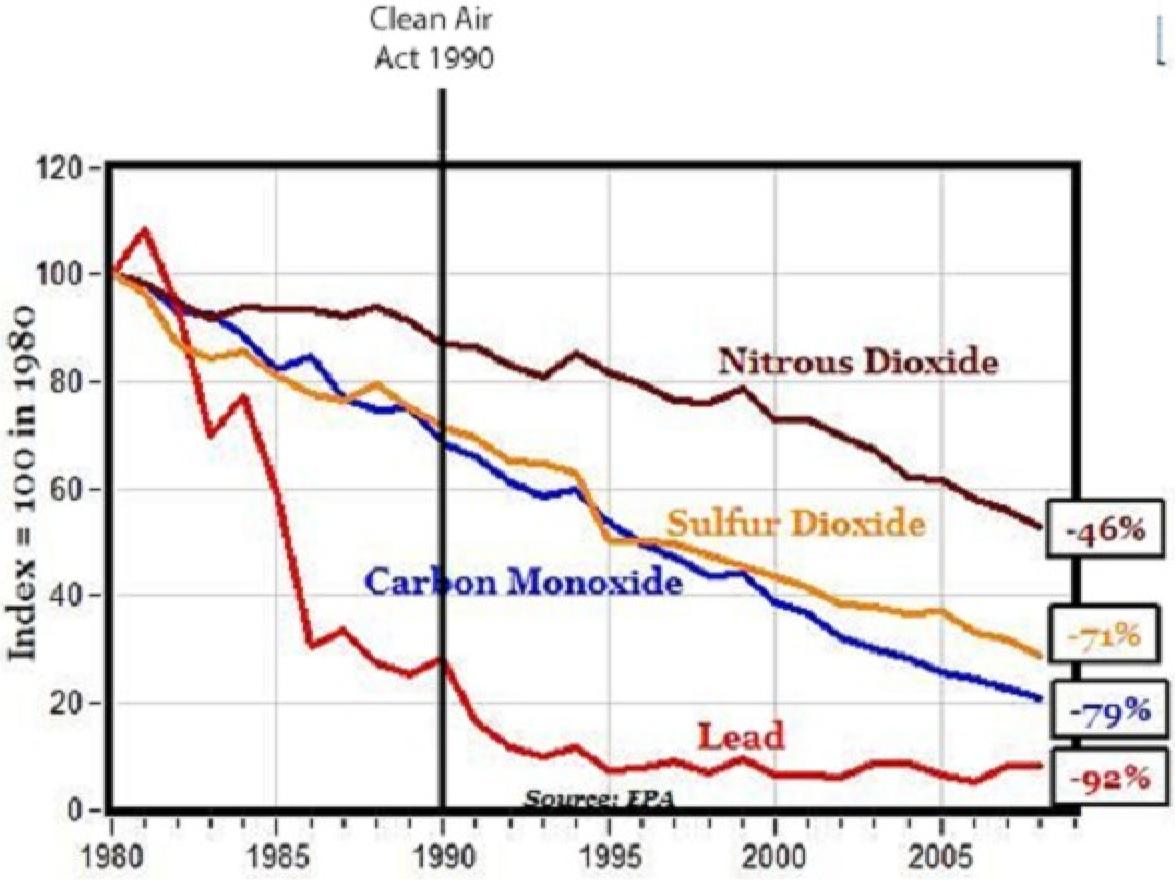
Other people who seem Puritan to me but AFAICT have no genealogical or cultural link: Cyrus McCormack, Homer Hickam, Emperor Norton. Also, surprisingly many Jews. There's a weird symmetry there: both groups started out living in in small, very strict religious communities where they wore black and had lots of kids; then upon contact with Modernity they both went the opposite route and became famous for their education, irreligion, and preeminence in various forms of liberal *tikkun olam*. Must be one of those coincidences.

People complain that there is too much neo-Puritanism around these days, but they usually just mean people are moralistic reformers. I have the opposite worry: what happened to these people? When was the last time you saw somebody called Hiram invent five different crazy machines, found a new religion, and have twelve children who he named after Greek nymphs? Anyone who is serious about "Making America Great Again" should be deeply worried.

The modern American caricature is the Borderers: impulsive gun-crazy fundamentalist hillbillies with country-western accents. The opposite American stereotype – the virtue-obsessed nonconformist eccentric inventor philanthropist – has almost disappeared. These people still exist – Bill Gates does a good job embodying the ideal (or for a closer-to-home example, Ben Hoffman of Compass Rose) but they’re disconnected from any historical archetype. Lots of writers have argued that if you want people to avoid a race-based identity, you need a national identity you can assimilate people into. But right now the US national identity is one that’s repulsive to a lot of people. I’m disappointed that Puritanism is no longer a thing that people can aim at.

Does Reality Drive Straight Lines On Graphs, Or Do Straight Lines On Graphs Drive Reality?

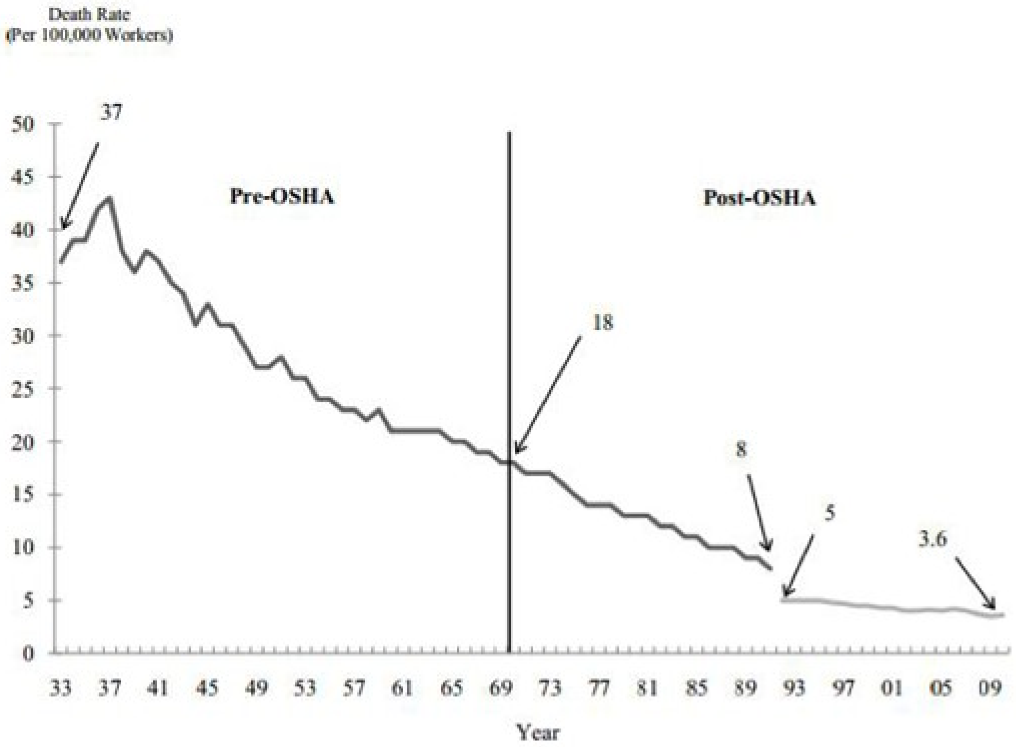
Here’s a graph of US air pollution over time:



During the discussion of 90s environmentalism, some people pointed out that this showed the Clean Air Act didn’t matter. The trend is the same before the Act as after it.

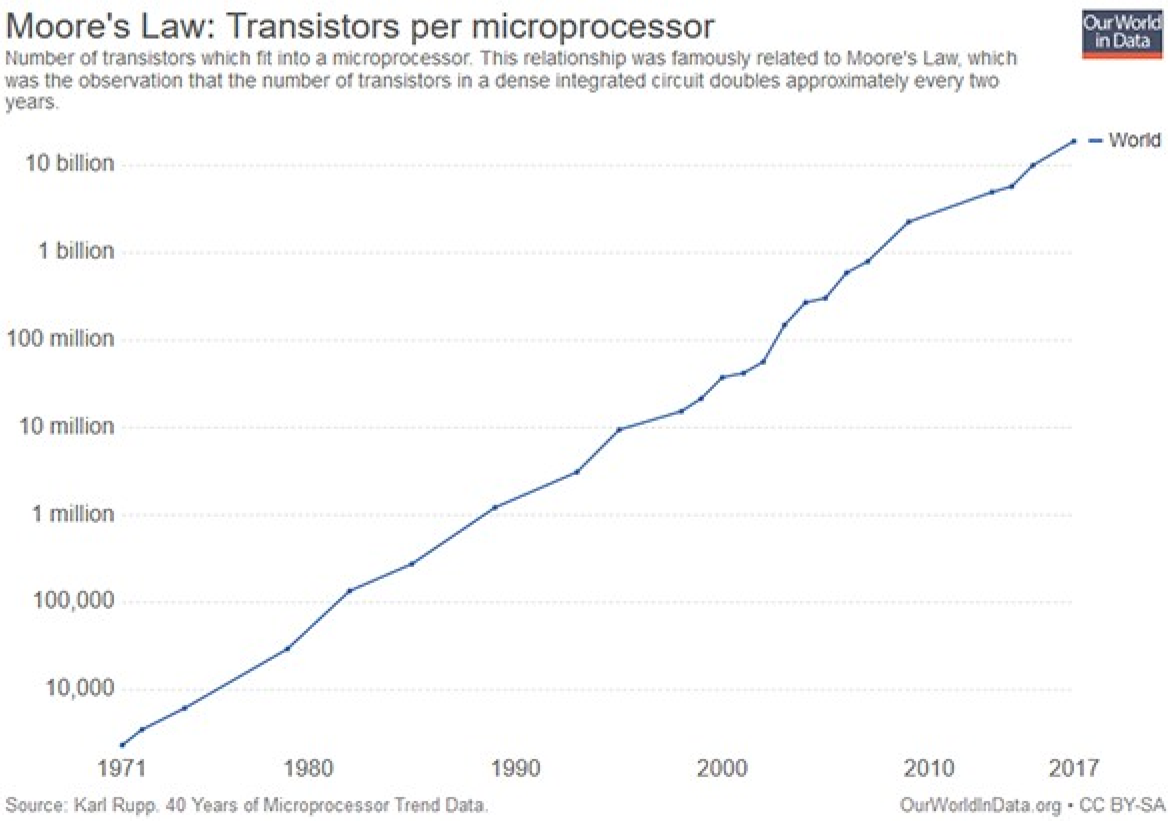
This kind of argument is common. For example, here’s the libertarian Mercatus Institute arguing that OSHA didn’t help workplace safety:

FIGURE 1: WORKPLACE FATALITIES, 1933–2010

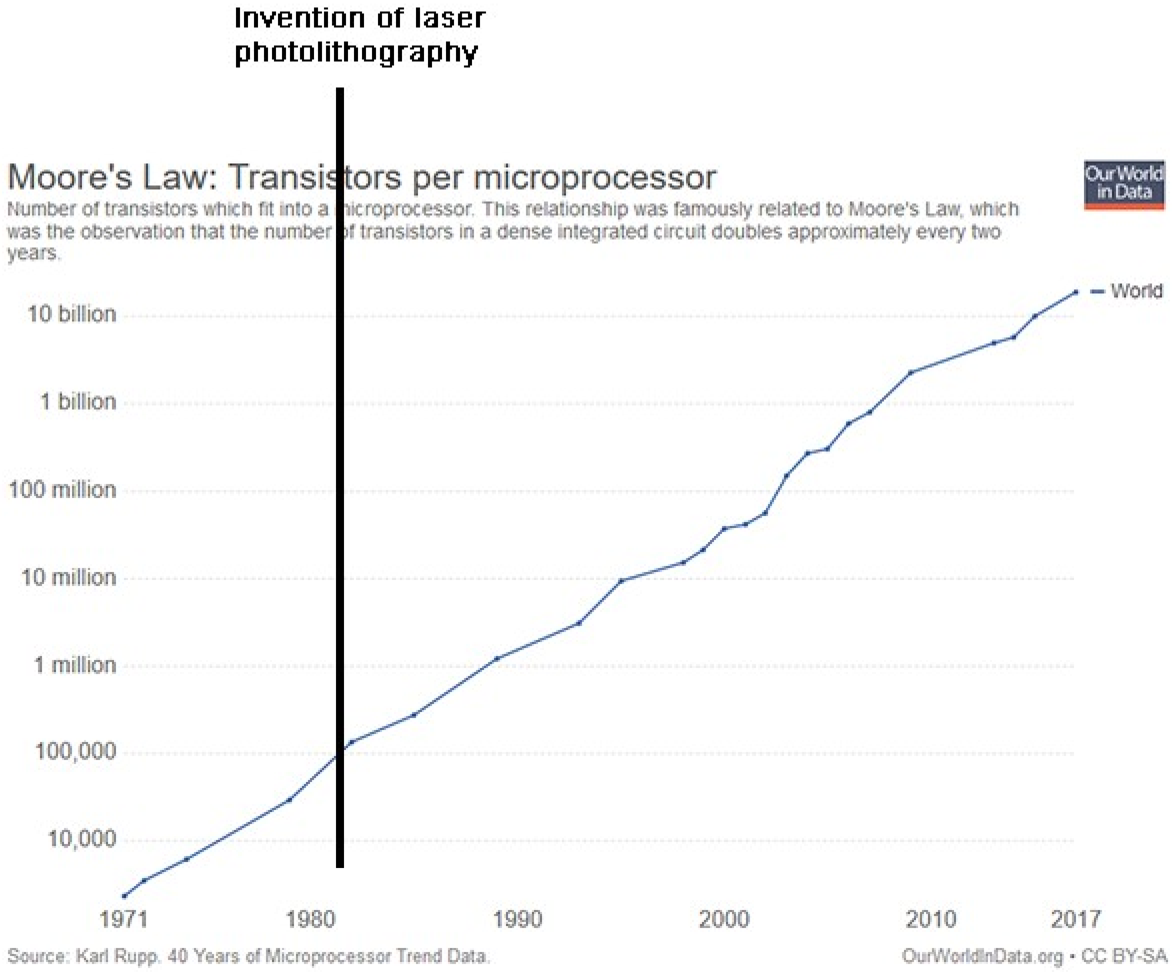


Sources: National Safety Council (1994) and US Department of Labor, Bureau of Labor Statistics (2012)

I’ve always taken these arguments pretty seriously. But recently I’ve gotten more cautious. Here’s a graph of Moore’s Law, the “rule” that transistor counts will always increase by a certain amount per year:



The Moore’s Law Wikipedia article lists factors that have helped transistors keep shrinking during that time, for example “the invention of deep UV excimer laser photolithography” in 1980. But if we wanted to be really harsh, we could make a graph like this:



But the same argument that disproves the importance of photolithography disproves the importance of anything else. We’d have to retreat to a thousand-coin-flips model where each factor is so small that it happening or not happening at any given time doesn’t change the graph in a visible way.

The only satisfying counterargument I’ve heard to this is that Moore’s Law comes from a combination of physical law and human commitment. Physical law is consistent with transistors shrinking this quickly. But having noticed this, humans (like the leadership of Intel) commit to achieve it. That commitment functions kind of as a control system. If there’s a big advance in one area, they can relax a little bit in other areas. If there’s a problem in one area, they’ll pour more resources into it until there stops being a problem. One can imagine an event big enough to break the control system – a single unexpected discovery that cuts sizes by a factor of 1000 all on its own, or a quirk of physical law that makes it impossible to fit more transistors on a chip without inventing an entirely new scientific paradigm. But in fact there was no event big enough to break the control system during this period, so the system kept working.

But then we have to wonder whether other things like clean air are control systems too.

That is, suppose that as the economy improves and stuff, the American people demand cleaner air. They will only be happy if the air is at least 2% cleaner each year than the year before. If one year the air is 10% cleaner than the year before, environmentalist groups get bored and wander off, and there's no more progress for the next five years. But if one year the air is only 1% cleaner, newly-energized environmentalist voters threaten to vote out all the incumbents who contributed to the problem, and politicians pass some emergency measure to make it go down another 1%. So absent some event strong enough to overwhelm the system, air pollution will always go down 2% per year. But that doesn't mean the Clean Air Act didn't change things! The Clean Air Act was part of the toolkit that the control system used to keep the decline at 2%. If the Clean Air Act had never happened, the control system would have figured out some other way to keep air pollution low, but that doesn't mean the Clean Air Act didn't matter. Just that it mattered exactly as much as whatever it would have been replaced with.

If this were true, you wouldn't see the effects of pollution-busting technologies on pollution. You'd see them on everything else. For example, suppose that absent any other progress on air pollution, politicians would regulate cars harder, and that’s what would make air pollution go down by 2% that year. In that case, the effects of inventing an unexpected new pollution-busting technology wouldn’t appear in pollution levels, they would appear in car prices. Unless car prices are also governed by a control system – maybe car companies have a target of keeping costs below $20,000 per car, and so they would skimp on safety in order to bring prices back down, and then the effects of a new anti-pollution technology would appear in car accident fatality rates.

How do we tell the difference between this world, and the world where the Clean Air Act *really* doesn’t matter? I’m not sure (does anyone know of research on this?). Maybe this is one of those awful situations where you have use common sense instead of looking at statistics.

I’m worried this could be a fully general excuse to dismiss any evidence that a preferred policy didn’t work. But it does make me at least a *little* slower to believe arguments based on interventions not changing trends.

Gwern's AI-Generated Poetry

Gwern has answered my prayers and taught GPT-2 poetry.

GPT-2 is the language processing system that OpenAI announced a few weeks ago. They are keeping the full version secret, but have released a smaller prototype version. Gwern retrained it on the Gutenberg Poetry Corpus, a 117 MB collection of pre-1923 English poetry, to create a specialized poetry AI.

I previously tested the out-of-the-box version of GPT-2 and couldn't make it understand rhyme and meter. I wrongly assumed this was a fundamental limitation: "obviously something that has never heard sound can't derive these complex rhythms just from meaningless strings of letters." I was wrong; it just didn't have enough training data. Gwern's retrained version gets both of these right, and more too. For example:

> Thou know'st how Menoetiades the swift
> Was dragged, of Hector and the fierce compeers
> And Phrygian warriors. So, we will dispatch
> Your bodies, then, yourselves to burn the ships
> In sacrifice; with torches and with bells
> To burn them, and with oxen to replace
> Your gallant friends for ever. But I wish
> That no man living has so long endured
> The onset of his foes, as I have power
> To burn or storm; for mighty Hector erst
> Was slain, and now returns his safe return

This is all perfect iambic pentameter. I know AP English students who can't write iambic pentameter as competently as this.

(by the way, both “compeers” and “erst” are perfectly cromulent words from the period when people wrote poems like this; both show up in Shelley)

It has more trouble with rhymes – my guess is a lot of the poetry it was trained on was blank verse. But when it decides it should be rhyming, it can keep it up for a little while. From its *Elegy Written in a Country Churchyard* fanfic:

> Methinks I see her in her blissful dreams:
> Or, fancy-like, in some mirage she lies,
> Majestic yet majestic, and of seems
> The image of the unconquerable skies.
>
> Methinks I see her in her blissful dreams:
> —Or, fancy-like, in some majestic cell,
> Where lordly seraphs strew their balmy dreams
> On the still night, or in their golden shell.
>
> There, in the calm of some Platonic dream,
> Sits she, and views the unclouded moon arise
> Like a fair lady full of realms divine;
>
> And, all at once, a stony face and bright
> Glittering in moonlight, like the noon-tints of a night.

COME ON, IT’S A ROBOT. THAT’S BETTER THAN YOU COULD DO IF YOU WERE A ROBOT. GIVE IT A BREAK.

I think its problem is less ability to rhyme, and more a sort of…maintaning the will to rhyme. Gwern prompts it with several Alexander Pope pieces; Pope is a very formal poet who always rhymes in the exact same way. For each Pope piece, GPT-2 does the first couplet in perfect Pope style, then degenerates into increasingly confused gibberish. For example:

> Pride even in numbers; wit's a kind pretence
> To something foreign still, but ne'er to sense;
> A constant waste of words, the world produces,
> A lazy sap, that nothing that's foreign to expels,
> ; that's foreign still unknown to the soul;
> A young, like, but in the world in motion,
> Obscending not, which smells all our own the worse than our own our own the soul's, and soul;
> Like sense; which is like, but in all our whole.
> Which thus far more like, but in all things that's an excellence; and ne'er unchanged by which is folly's the worse, they give the worse maintained by which
> If sick of sense;
> Wholubil, or snug. ills, we know our own our first in sense the worse maintained between the worse, soon expired.

First two lines are perfect rhyme and rhythm, next four have no rhyme but are *close to* the right meter, next few have almost random length, and by the last one we've abandoned grammar and are making up nonsense words like "wholubil". Every Pope test Gwern runs is like this:

> They talk of constancy and faithless love,
> A seraph trembles at the specious glove;
> Nor in the rich confin'd relief of state,
> Find proud pretence, nor in the disdiscoveries of fate.
> For when misfortune makes choice remains the conduct's the prize half known,
> Can we secret soul without due, they fear of sense of more known.
> Some rise where' rights, they make it pays due.
> .
> .
> . Sense and judgment, as equal prize seem meanly, the reward the joy, as much possess the prize paid, as well

> reckon the prize we do not less dare not less keenly wise.
> We see;
> Of happy lovers ought, as well done, like a friend.
> Know they ought, ’tis ev’ is the other joy, as well worth a right;
> The joy, as well might, as well may, as well may all is great.
> Nor need of joys not as well maysters, as well as well may they give; but as little store; but as well as well as well may shewn, as much, as well we know, as well as well can be sure might prove, as well may well as well as well as well as well may view;
> The mind: as well as well as well as well as much the fair as well as well as well as well as well as well as well may

Again, first two lines are great – “a seraph trembles at the specious glove” is both nonsense and *exactly the sort of thing Alexander Pope would write*, but by the fourth line we have nonsense words, by the fifth we lose the meter, the eighth and ninth are just periods, and finally it starts stuttering helplessly.

I tested this many more times on a public version (not poetry-trained) and found a similar effect – the first two lines are always the best, and it deteriorates from there. I’m interested in hearing from people who understand the model better than I do about why this should be.

Some other highlights:

> My heart, why come you here alone?
> The wild thing of my heart is grown
> To be a thing,
> Fairy, and wild, and fair, and whole

That last line, with its ABAB structure, is actually brilliant even by the standards of human poets. “Fairy and wild and fair and whole”. I could say that all day. This has to be a coincidence. It’s not that good anywhere else. But even having something generally okay enough that it can occasionally blunder into something that good is great.

From its Hindu phase:

> …which Indra, King of all the Blest,
> Had thrown by Rávan’s mighty breast,
> The monstrous coil, the brawny hand,
> The monstrous mouth, the jaw, the jaw,
> The jaw, the jaw and bleeding jaw,
> The ungovernable host, the jaw,
> And the great bow which never bends,
> The arm, the fist, the knee, the ends,
> The body laid with mighty stroke,
> And the great bow which never bends.
> So, when the giants fought, and fell
> With murderous strokes, the giant fell,—
> So falls the tree with all his trunks
> Terrific in its death, that shoots
> Wild volley at the mighty trunk,—
> So fell the tree with all its boughs
> While all the vipers dug and sowed—
> So fell the tree with all its boughs.
> But Ráma’s heart was sad within
> He wept and mourned his captive’s sin

The first two lines are great (and the program must have invented the epithet “King of All The Blest” for Indra, because I can’t find it anywhere else). The description of Ravan’s jaw is, uh, quite memorable. “So falls the tree with all his trunks” is a good deployment of a metaphor which is frankly overused in old poetry. I’m ambivalent between interpreting the repetition as an AI getting stuck in a loop vs. as good use of repetition as a poetic device.

The next few are from this list of a thousand randomly generated samples:

> And they have seen the last light fail;
> By day they kneel and pray;
> But, still they turn and gaze upon
> The face of God to-day.
>
> And God is touched and weeps anew
> For the lost souls around;
> And sorrow turns their pale and blue,
> And comfort is not found.
>
> They have not mourned in the world of men,
> But their hearts beat fast and sore,
> And their eyes are filled with grief again,
> And they cease to shed no tear.
>
> And the old men stand at the bridge in tears,
> And the old men stand and groan,
> And the gaunt grey keepers by the cross
> And the spent men hold the crown.
>
> And their eyes are filled with tears,
> And their staves are full of woe.
> And no light brings them any cheer,
> For the Lord of all is dead

And:

> There are several kinds of people in America;
> There are several kinds of people, I mean their number.
> There’s a girl growing up in the house by the light,
> There’s a youth upon the road, or a girl somewhere in

> New York;
> There’s a prettier girl, and a man more congenial,
>
> But none of the likes of the likes of the fellows are equal.
> There’s one who has never been married and married,
> There’s one who don’t want to be treated with kindness;
> A fair youth is never employed nor neglected;
> There’s one who has never yet come to a neighbor,
> And one who resides in New York from the start;
>
> But none of the likes of the likes of the fellows
> Are equal to him, and wherever he goes,
> The heart somehow breaks under the hand that is steering;
> And so it is with me

And this quatrain just worked out really well:

> Fair is the Lake, and bright the wood,
> With many a flower-full glamour hung:
> Fair are the banks; and soft the flood
> With golden laughter of our tongue

This one is notable for competent metaphor:

> How the clouds
> Seem to me birds, birds in God’s garden! I dare not!
> The clouds are as a breath, the leaves are flakes of fire,
> That clash i’ the wind and lift themselves from higher!

And this one is obviously a failure on one level, but on another level is some kind of great experimental modern political poetry:

> The Emperor Wu (the great Wu), majestical,
> The Emperor Wu (the great Wu), majestical,
> The Emperor Wu (the great Wu), majestical,

The Emperor Wu (the great Wu), majestical,
The Emperor Wu (the great Wu), majestical,
The Emperor Wu (the great Wu), majestical,
The Emperor Wu (the great Wu), majestical,
The Emperor Wu (the great Wu), majestical,
The Emperor Wu (the great Wu), majestical,
The Emperor Wu (the great Wu), majestical,
The Emperor Wu (the great Wu), majestical,
The Emperor Wu (the great Wu), majestical,
The Emperor Wu (the great Wu), rapacious,
The Emperor Wu (the great Wu), majestical,
The Emperor Wu (the great Wu), rapacious,
The Emperor Wu (the great Wu), rapacious,
The Emperor Wu (the great Wu), rapacious,
The Emperor Wu (the great Wu), rapacious,
The Emperor Wu (the great Wu), rapacious,
The Emperor Wu (the great Wu), rapacious,
The Emperor Wu (the great Wu), rapacious,
The Emperor Wu (the great Wu), rapacious,
The Emperor Wu (the great Wu), rapacious,
The Emperor Wu (the great Wu), rapacious

This one displays an interesting combination of world-knowledge and lack-of-world-knowledge:

In the dark the sun doth gleam,
And in the dark the moon doth seem
But now the evening is begun–
Gone is the sun upon the earth!
The silver moon doth like a cup
Of blood-red wine, and as that cup
Is drained of life, doth quench no drop.
What man will drink such wine?
There is no soul of earth or birth
Which man hath never known of earth.
There is no soul who doth not sit

> And sing to it, and cry, “Drink!”
> There is no soul whose feet are set
> On youth’s eternal paradise;
> For all is a solemn harmony,
> And all is a perpetual chant,
> And all the world is a song of God.
> There is no soul so wholly free

And here’s another:

> There gloom the dark, broad seas. My mariners,
> Souls that have toiled, and wrought, and thought with me
> That ever with a frolic welcome took
> The thunder and the sunshine, and opposed
> Free hearts, free foreheads – you and I are old;
> Old age hath yet his honour and his toil;
> Death closes all: but something ere the end,
> Some work of noble note, may yet be done,
> Not unbecoming men that strove with Gods.

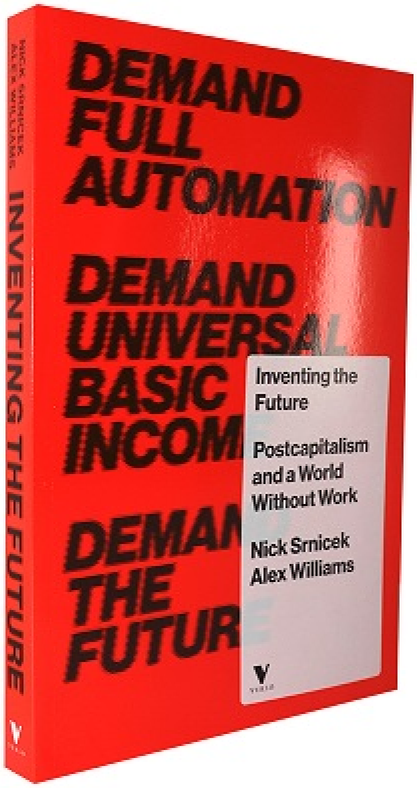
Except this last time I’m cheating: this is an excerpt of Tennyson’s *Ulysses*, one of the most famous English poems. I included it as a placebo, ie a test to see whether real poems sound fake if you think they’re by an AI when you read them. I’ll be honest: if I didn’t know this was Great Poetry, I would skim it over and assume it made several mistakes. Like: is “gloom” really a verb? (it is if you’re Alfred, Lord Tennyson). Is the last line grammatical? (yes: it’s an adjective phrase modifying “work”, ie “some work which is fitting for the sort of men who fought gods to do”). Are the mariners’ souls opposing their foreheads? (I’m still confused on this one). These are all the sorts of things that would make me go “Haha, AIs are still pretty dumb” if I were reading it blindly.

If you liked these poems, you might also appreciate Gwern’s work making AI-generated anime waifus.

(and you can also donate to Gwern’s Patreon here)

Book Review: Inventing The Future

I.


DEMAND FULL AUTOMATION
DEMAND UNIVERSAL BASIC INCOME
DEMAND THE FUTURE
Inventing the Future
Postcapitalism and a World Without Work
Nick Srnicek
Alex Williams
NICK SRNICEK ALEX WILLIAMS
INVENTING THE FUTURE

They say "don't judge a book by its cover". So in case you were withholding judgment: yes, this bright red book covered with left-wing slogans is, in fact, communist. *Inventing The Future* isn't technically Nick Srnicek and Alex Williams' manifesto – that would be the equally-striking-looking Accelerate Manifesto. But it's a manifesto-ish description of their plan for achieving a postcapitalist world.

S&W start with a critique of what they call "folk politics", eg every stereotype you have of lazy left-wing activists. Protesters who march out and wave signs and then go home with no follow-up plan. Groups that avoid having any internal organization, because organization implies hierarchy and hierarchy is bad. The People’s Front of Judaea wasting all their energy warring with the Judaean People’s Front. An emphasis on spectacle and performance over results. We’ve probably all heard stories like this, but some of S&W’s are especially good, like one from an activist at a trade summit:

> On April 20, the first day of the demonstrations, we marched in our thousands toward the fence, behind which 34 heads of state had gathered to hammer out a hemispheric trade deal. Under a hail of catapult-launched teddy bears, activists dressed in black quickly removed the fence’s support with bolt cutters and pulled it down with grapples as onlookers cheered them on. For a brief moment, nothing stood between us and the convention centre. We scrambled atop the toppled fence, but for the most part we went no further, as if our intention all along had been simply to replace the state’s chain-link and concrete barrier with a human one of our own making.

S&W comment:

> We see here the symbolic and ritualistic nature of the actions, combined with the thrill of having done something – but with a deep uncertainty that appears at the first break with the expected narrative. The role of dutiful protester had given these activists no indication of what to do when the barriers fell. Spectacular political confrontations like the Stop the War marches, the now familiar melees against G20 or World Trade Organization and the rousing scenes of democracy in Occupy Wall Street all give the appearance of being highly significant, as if something were genuinely at stake. Yet nothing has changed, and long-term victories were traded for a simple registration of discontent.

> To outside observers, it is often not even clear what the movements want, beyond expressing a generalized discontent with the world…in more recent struggles, the very idea of making demands has been questioned. The Occupy movement infamously struggled to articulate meaningful goals, worried that anything too substantial would be divisive. And a broad range of student occupations across the Western world has taken up the mantra of “no demands” under the misguided belief that demanding nothing is a radical act.

All of this is pretty standard commentary, both from leftists and from rightists making fun of them. What S&W added that I hadn’t heard before was an attempt to portray this all as coming from bad philosophy. I had always assumed most leftist groups sucked because they were primarily made of stoner college kids and homeless people, two demographics not known for their vast resources, military discipline, or top-notch management skills. But S&W believe they suck because they *choose* to suck, for principled reasons.

They give a few specific principles, but sum them up in the idea of prefiguration: leftist groups should embody utopian leftist values right now. If capitalism is big and complicated and inhuman, leftist groups should be small, simple, and human-scale. If capitalism is coldly rational, leftist groups should be based on transient displays of emotion. If capitalism creates highly-organized hierarchies, leftist groups should be a formless mass of equals. If capitalism is ruthlessly focused on results, leftist groups should prize the journey itself. The goal shifts from concrete results to “prefigurative experience”: where people have a sense of life outside of capitalist strictures, which then sort of mystically lights a spark that kindles revolution in the hearts of all mankind. Or something:

> Even granting the problematic assumption that most people would want to live as the Occupy camps did, what efforts might be possible to physically and socially expand these spaces? When theorists face up to this question, vague hand-waving usually ensues: moments will purportedly ‘resonate’ with each other; small everyday actions will somehow make a qualitative shift to ‘crack open’ society; riots and blockades will ‘spread and multiply’; experiences will ‘contaminate’ participants and expand; pockets of prefigurative resistance will just ‘spontaneously erupt’. In any case, the difficult task of traversing from the particular to the universal, from the local to the global, from the temporary to the permanent, is elided by wishful thinking.

Is this a straw man? I have read many leftists complaining that this is what other leftists think, and relatively few leftists saying they think this – though this could be an artifact of who I read. But S&W don’t think it’s straw-mannish. To their credit, they write to an implied audience of pro-folk-politics leftists, begging them to change their ways. More on this later.

They conclude this section by saying that folk politics has failed and better ideas are needed. They give a brief nod to a long string of leftist victories over the past half-century or so (civil rights, women’s rights, gay rights, environmental regulation, massive increase in most categories of government social spending, etc, etc, etc), but are unimpressed, since these are compromises within capitalism domination. I would have liked to see them address an alternate perspective, where capitalism having to keep making compromise after compromise to defuse pressure from the left is exactly what lefist victory should look like. Would electing Bernie Sanders and instituting Medicare-For-

All be just another capitalist compromise? What about electing Andrew Yang and instituting Universal Basic Income? At some point you have to admit that all these “compromises” add up and now you have 90% of what you wanted in the first place. I assume they have some kind of complicated theoretical structural reason why this doesn’t work, but it still seems like a pretty good deal.

II.

What is the opposite of folk politics? S&W point to the Mont Pelerin Society.

The Mont Pelerin Society has a great story, and you should read Kerry Vaughn’s long writeup of the same topic. But the short and oversimplified version is: in the 1940s, everyone serious was either a Big Government Socialist or a Big Government Keynesian. Friedrich Hayek founded the Mont Pelerin Society (named after the site of its first meeting) to promote neoliberalism – here meaning the sort of small-ish government free market thinking common in economics today. At first they were just a few fringe thinkers with no power. But they developed a long-term strategy to change that. Vaughn, S&W, and others sum up the basic points as:

1. Foster intellectual talent

2. Seek long-term academic influence. Getting your members professorships won’t feel as exciting and tangible as reshaping policy immediately. Get the professorships anyway.

3. Push a utopian vision (in the case of the neoliberals, one of freedom and prosperity), along with practical first steps within the Overton Window (eg deregulating the airline industry).

4. Be prepared to step in as saviors when a crisis arrives. Milton Friedman:

> There is enormous inertia — a tyranny of the status quo — in private and especially governmental arrangements. Only a crisis — actual or perceived — produces real change. When that crisis occurs, the actions that are taken depend on the ideas that are lying around. That, I believe, is our basic function: to develop alternatives to existing policies, to keep them alive and available until the politically impossible becomes politically inevitable.

The neoliberals spent the 1940s through the 1970s slowly moving through steps 1 – 3. They gathered a stable of friendly academics, journalists, politicians, and (especially) think tanks, sometimes by converting people in positions of power, other times by putting their own loyalists into positions, and especially by founding their own organizations. When the stagflation crisis of the 1970s struck, they had marshalled a strong case as the alternative to the Keynesian system that had produced the crisis. Politicians – especially Ronald Reagan and Margaret Thatcher – agreed to implement their policies, and the rest is history.

S&W abhor the Mont Pelerin Society’s policies, but they are impressed by their success. They describe the result of MPS’ efforts as a “hegemony”, a paradigm so self-consistent and self-contained that it seems like, as the famous saying goes, “there is no alternative”. Neoliberal economics has stock answers to all of the objections raised against it and supports neoliberal politics, which has stock answers to all the objections raised against it and supports neoliberal culture, and so on.

They argue that leftists should abandon folk politics and do something more Mont Pelerinish, which they call a counter-hegemonic project. The Left should create a network of academics, journalists, think tanks, and politicians who come up with leftist ideas, push the culture to the left, and make sure the public knows Communism is the alternative to the current failing system.

Isn't this pretty much just the “long march through the institutions”? And didn't it happen thirty years ago?

I'm confused by this whole topic. Marxists seem to talk a lot about Gramsci and “cultural hegemony”, and “march through the institutions” was a phrase used by Gramscians to describe their strategy of controlling institutions in the name of Marxism. And *Inventing The Future* seems to say “Yes, this is exactly what we want” and even cites Gramsci in a bunch of footnotes. But whenever a non-Marxist mentions this, it gets branded a vile far-right anti-Semitic conspiracy theory. I'm guessing that there's some subtle distinction between the stuff everyone agrees is true and the stuff everyone agrees is false, and that lots of people will get angry with me for even *implying* that it might not be a vast gulf larger than the ocean itself, but I can't figure out what it is and don't want to land on the wrong side of it and get in trouble.

So let's say no. Let's say S&W's plan of taking over institutions in the name of leftism is completely new. Could this exciting and very original idea be just crazy enough to work?

S&W and their fellow communists aren't the first group I've heard bring up the Mont Pelerin Society as an example to be emulated. As you can tell from the link above, some effective altruists are thinking in this direction too. And “take over academia and dominate the intellectual world” is a good job if you can get it. But it still sounds pretty hard, especially if lots of other people have the same idea.

The neoliberals had some dazzling successes. But so did Napoleon. And if the centerpiece of your strategy is “Take over France, then go from there, after all it worked for Napoleon”, you could be accused of focusing too much on one past success without considering alternatives. Also, you probably can’t take over France. Taking over France is hard. Sure, Napoleon did it once, but think of all the people who must have tried to take over France and failed. I don’t know, seems like a really underspecified plan.

III.

So what *is* S&W’s plan?

This doesn’t get a huge amount of space in the book, but it seems to be: fight for automation and universal basic income in order to produce a “post-work world”.

There is much discussion of why work is bad, which I appreciate. I think communists are wrong about a lot of things, but when this is all over, I believe their principled insistence that work is bad and that we should not have to do it – maintained firmly against a bunch of people who want basic job guarantees or who consider freedom from work a utopian impossibility – will be one thing they can be really proud of. S&W are very sure work is bad, they manage to express this without accidentally adding on anything blitheringly stupid, and their point that we should head into a post-work world is well-taken.

They discuss the increasing role of automation in society, complete with ill-fated predictions that many people who lost their jobs in the Great Recession (just before the book was written) will never get them back and unemployment will remain permanently high. They argue that automation is a vital component of a post-work world, and argue that leftist movements should use whatever strength they have to fight for automating things further. They argue that capitalism is not automating as quickly as it could be, and this is bad for workers:

> Full automation is something that can and should be achieved, regardless of whether it is yet being carried out. For instance, out of the US companies that could benefit from incorporating industrial robots, less than 10% have done so. This is but one area for full automation to take hold in, and this reiterates the importance of making full automation a political demand, rather than assuming it will come about from economic necessity. A variety of politices can help in this project: more state investment, higher minimum wages, and research devoted to techniologies that replace rather than augment workers. In the most detailed estimates of the labour market, it is suggested that between 47 and 80 per cent of today's jobs are capable of being automated. Let us take this estimate not as a deterministic prediction, but instead as the outer limit of a political project against work. We should take these numbers as a standard against which to measure our success.

(notice the suggestion to raise the minimum wage in order to encourage automation; this is more realism than I usually hear in this kind of discussion.)

Alongside the demand for increased automation, leftists should demand a universal basic income:

> Drawing upon moral arguments and empirical research, there are a vast number of reasons to support a UBI: reduced poverty, better public health and reduced health costs, fewer high school dropouts, reductions in petty crime, more time with family and friends, and less state bureaucracy. Depending on how UBI is presented, it is capable of generating support from across the political spectrum—from libertarians, conservatives, anarchists, Marxists and feminists, among others. The potency of the demand lies partly in this ambiguity, making it capable of mobilizing broad popular support […]
>
> The demand for a UBI is a demand for a political transformation, not just an economic one. It is often thought that UBI is simply a form of redistribution from the rich to the poor, or that it is just a measure to maintain economic growth by stimulating consumer demand. From this perspective, UBI would have impeccable reformist credentials and be little more than a glorified progressive tax system. Yet the real significance of UBI lies in the way it overturns the asymmetry of power that currently exists between labour and capital. As we saw in the discussion of surplus populations, the proletariat is defined by its separation from the means of production and subsistence. The proletariat is thereby forced to sell itself in the job market in order to gain the income necessary to survive. The most fortunate among us have the leisure to choose which job to take, but few of us have the capacity to choose no job. A basic income changes this condition, by giving the proletariat a means of subsistence without dependency on a job.

S&W finally sum up their platform as:

1. Full automation
2. The reduction of the working week
3. The provision of a basic income
4. The diminishment of the work ethic

I am surprised by points 1, 2, and 4. I don't disagree with them. But they seem heavily dependent on point 3. If there's no basic income, automation is a disaster – it just leaves everyone in the same kind of normal bad old unemployment we have now. Same with a diminished work week and lack of work ethic. Usually I think of platforms as the sort of thing where if you get three-quarters of what you want you can declare victory; here three-quarters of the platform would be a dystopia.

On the other hand, UBI would lead inevitably to the other three points. As S&W mention, once living off a UBI becomes a viable alternative to working, many people (though not everyone at once) will choose to do this. That will increase the cost of labor, drive wages up, and encourage businesses that haven't yet automated to do so. It will also reduce the number of hours people need to (or choose to) work, and once people aren't working and goods are being produced without labor, there won't be any need for a work ethic.

I'm especially surprised by the insistence on automation (which takes up more space than this review might indicate). Aside from its ability to enable a UBI, automation itself doesn't necessarily seem good. If it would cost more to automate a task than to hire workers to do it, there doesn't seem much advantage in automating, *assuming that workers have free choice to take whatever deal the business offers*. That is, compare a world in which a factory pays $100,000 per year to operate a car-making robot, and you get a $30,000 a year basic income, vs. a world where a company pays $50,000 per year to hire a car-building employee, and you have the choice between getting a $30,000 a year basic income, or earning $80,000 a year by taking the company's offer. The second world seems clearly better. I know the usual communist answer is to talk about flow-through effects on who has political power, but I feel like a world in which workers are necessary to make goods is one in which workers have more political power than a world where they aren't.

So maybe a friendly amendment to S&W's platform would throw out everything except the UBI?

I doubt they would accept this amendment, but I can't predict exactly what they would say when turning it down. Certainly they *really* don't like libertarians who agree with them on UBI and want to help them with it, but I can't seem to wring a specific complaint out of their denunciations:

> The demand for a UBI, however, is subject to competing hegemonic forces. It is just as open to being mobilized for a libertarian dystopia as for a post-work society. Hence, three qualifications must be added to this demand. First, it has to provide a sufficient amount to live on, second, it has to be universal and third, it has to be a supplement rather than a replacement for the welfare state…

There's a lot of stuff like this, culminating in a triumphant jab that if there were a UBI, we would end up in the world libertarians *claim* they want, the one where everyone is free and happy and can choose how to live their lives, rather than the world we all know libertarians secretly *do* want, where everybody is oppressed by the rich forever. Won't *that* be ironic! Won't the libertarians howl with anguish when they realize what kind of clever political judo trick the Marxists have played on them!

I think I speak for everybody when I say: please don’t throw me in the briar patch! Not the briar patch! Anything but that!

IV.

S&W often use the word “hegemony”, usually in the context of a current neoliberal hegemony or a hoped-for future communist hegemony. I started out reading this as just an emphatic way of saying “power”, as in “the neoliberals are definitely almost-irreversibly in power” or “the communists are definitely almost-irreversibly in power”. After looking into it more, I think the best interpretation of “hegemony” is “paradigm” in the Kuhnian sense, or as the sort of thing that should be in one of the leftmost boxes of this table. A hegemony is a self-consistent way of looking at the world that guides how we think about things and what questions we ask. It is not just made of facts, but also of perspectives, biases, values, and investigational techniques, all coming together into a coherent whole.

I picked up *Inventing The Future* (on advice from a couple of left-accelerationists I encountered at the Southern California SSC meetup) because I feel bad that I’ve never been able to get my head around the communist paradigm. In the past, I’ve learned new paradigms by reading a lot of books from within that paradigm (and hating them) and debating people from within that paradigm (and thinking they’re crazy). Then fifty books and a hundred debates down the line, I finally get some kind of inkling of where they’re coming from, and then after a while I can naturally make my mind shift into that mode and my only differences with them are at the high-level generators of disagreement. I was born into the Woke California Liberal paradigm, I managed to force myself to understand the libertarian paradigm in college, I managed to force myself to understand the right-wing paradigm a few years ago, and I would really like to be able to understand the communist paradigm too.

This leaves me in the awkward position of needing to read a lot of communist books and wanting to be as accepting as possible towards them, while also inevitably knowing I'm going to hate the first fifty. So I'll be honest: I really didn't like *Inventing The Future*.

Part of this isn't S&W's fault. The book was very much not meant for me. Not everything has to be a 101 space. It's perfectly fine for the Pope to write an encyclical explaining the Catholic position on marriage without including a justification of why you should believe in God, or why Jesus is so great, or why anyone should care what the Pope thinks, or whether marriage should exist at all. People who already believe all of those things should be able to debate the godly Christian papal way of defending marriage among themselves. Likewise, a book for communist true believers about how to win doesn't necessarily need to justify communism.

But there were times when I feel like even a true believer would have been groping for some reassurance. During their attack on folk politics, S&W were pretty open about how the problem was that groups tried to apply communist principles to communist activism. For example, communism should be non-hierarchical, so some activist groups tried to be non-hierarchical, but then all those activist groups failed. Or: communism says we should abandon market economies for ones based on mutual aid, so Occupy camps tried to have internal economies based on mutual aid, but then those camps couldn't get resources distributed effectively. S&W's conclusion was: stop trying to run your activist groups in a communist way, that never works. I'm sure they're right. But I feel like even true believers might have wondered why real communism, when it came, would go differently. This was never explained.

Likewise, S&W talked quite honestly about how many small-scale experiments with communism have failed. They gave the example of some Argentine workers forming commune-like organizations when that country's economy collapsed. These kind of worked for a while, but the authors describe them as uninspiring, noting that such communes-within-capitalism could be "as oppressive and environmentally damaging as any large-scale business". Once the economy recovered, Argentines were pretty relieved to be able to return to normal capitalist living. S&W's conclusion: you have to destroy all of capitalism at once or it doesn't count.

I understand this has been a common position in communism since well before Trotsky. But imagine a pharmaceutical company admitting that its drugs have killed everyone who's taken them so far, but adding "But if we give this drug to everyone in the world at the same time, then it will definitely cure everything!" You would think they would at least add "We recognize this may be a cause of some concern to people who worry past trends won't suddenly reverse, and we will just end up killing everyone in the world, but here's why you shouldn't worry…". You would think even true believers might want to hear some reassurance at this point. S&W do not provide it.

I don't want to be too harsh on them. Capitalists have a similar conundrum: if the free market works, how come most businesses are organized as top-down hierarchies? How come there's a Vice President of Sales who gets hired, promoted, and fired – instead of just some sales consulting businesses offering their time to the CEO at market rate? Capitalists have confronted these issues; probably communists have confronted theirs too. This is the sort of

101 stuff that *Inventing The Future* is under no obligation to bother with. It just made the book a bad match for me.

And all of this was exacerbated by S&W devoting entire chapters to ideas I considered obvious that were apparently highly controversial to their intended audience. For example, S&W were going to make some demands for what a future communist state should be like. I was interested in hearing these demands. Instead they went on for page after page about whether it was okay to demand things. For example:

> We [will] advance some broad demands to start building a platform for a post-work society. In asserting the centrality of demands, we are breaking with a widespread tendency of today’s radical left that believes making no demand is the height of radicalism. These critics often claim that making a demand means giving into the existing order of things by asking, and therefore legitimating, an authority. But these accounts miss the antagonism at the heart of making demands, and the ways in which they are essential for constituting an active agent of change. In this light, the rejection of demands is a symptom of theoretical confusion, not practical progress. A politics without demands is simply a collection of aimless bodies. Any meaningful vision of the future will set out proposals and goals, and this chapter is a contribution to that potential discussion…

When they get to discussing how communism is good, they don’t anticipate any object level complaints about how, eg, maybe capitalism is better. But they do worry that “communism is good” sounds like a universal statement, and universal statements can be exclusionary. So:

To invoke such an idea is to call forth a number of fundamental critiques directed against universalism in recent decades. While a universal politics must move beyond any local struggles, generalising itself at the global scale and across cultural variations, it is for these very reasons that it has been criticised.

As a matter of historical record, European modernity was inseparable from its ‘dark side’ – a vast network of exploited colonial dominions, the genocide of indigenous peoples, the slave trade, and the plundering of colonised nations’ resources. In this conquest, Europe presented itself as embodying the universal way of life. All other peoples were simply residual particulars that would inevitably come to be subsumed under the European way – even if this required ruthless physical violence and cognitive assault to guarantee the outcome. Linked to this was a belief that the universal was equivalent to the homogeneous. Differences between cultures would therefore be erased in the process of particulars being subsumed under the universal, creating a culture modelled in the image of European civilisation. This was a universalism indistinguishable from pure chauvinism. Throughout this process, Europe dissimulated its own parochial position by deploying a series of mechanisms to efface the subjects who made these claims – white, heterosexual, propertyowning males. Europe and its intellectuals abstracted away from their location and identity, presenting their claims as grounded in a ‘view from nowhere’. This perspective was taken to be untarnished by racial, sexual, national or any other particularities, providing the basis for both the alleged universality of Europe’s claims and the illegitimacy of other perspectives. While Europeans could speak and embody

> the universal, other cultures could only be represented as particular and parochial.
>
> Universalism has therefore been central to the worst aspects of modernity’s history. Given this heritage, it might seem that the simplest response would be to rescind the universal from our conceptual arsenal. But, for all the difficulties with the idea, it nevertheless remains necessary. The problem is partly that one cannot simply reject the concept of the universal without generating other significant problems. Most notably, giving up on the category leaves us with nothing but a series of diverse particulars.

I am tempted to sum up the book as something like “So, obviously everyone agrees that we should overthrow all existing societies to install world communism. But many people doubt that causes lead to effects. Well, I’m here to tell you that we’ll never overthrow all existing societies in favor of world communism unless we take actions that cause that to happen.”

And the authors aren’t just being silly. The book has an epilogue where they respond to criticism they received since the book was published, and it’s all people praising them for their commitment to revolution while also accusing the causes-have-effects thing of being highly problematic. So A+ on writing what your audience needs to hear. I’m just not among them.

V.

Feminist critics of bad pick-up artistry accuse it of “looking for women’s secret cheat code that will make them have sex with you”. The opposite of looking for a cheat code (they say) is actually having and demonstrating value. If women don't like you, you should try to cultivate value, or demonstrate the value you already have, instead of finding the Three Words And Five Gestures That Will Make Any Girl Get Naked Right Now.

Not everything has to be a 101 space. So maybe my concerns are just an artifact of me wandering into a part of literature where I don't belong. But at its worst, *Inventing The Future* feels like a search for the public's secret cheat code that will make them have a revolution with you.

There is no discussion of why communism is good. There is no discussion of whether the masses might not like communism because they've thought about it for a while and decided that communism is a bad idea. There is no discussion of whether some demonstration that communism is good would convince the masses to like it more. Just a laser-like focus on finding the secret propaganda cheat code that will convert the masses to communism.

I don't know how unfair I'm being here. The most sympathetic reading I can give this is something like "Somewhere off-screen we've already agreed that every right-thinking person already knows communism is better than capitalism. And we've all agreed that the elites have brainwashed the masses into denying it. And the elites will never give up their own self-interest. So the only remaining question is how we can create a system of organization and publicity and so on powerful enough to reverse the masses' brainwashing." This is at least good conflict theory.

But at times S&W seem to dip into a deeper epistemological nihilism. From a paragraph on the rise of neoliberalism:

> The crisis (stagflation) was one that no government knew how to deal with at the time, while the solution

> was the preconceived neoliberal ideas that had been fermenting for decades in its ideological ecology. It was not that the neoliberals presented a better argument for their position (the myth of rational political discourse); rather, an institutional infrastructure was constructed to project their ideas and establish them as the new common sense of the political elite.

Google cannot find any references to “myth of rational political discourse” except in this book. Maybe there’s some long discussion of this idea under another name somewhere, but S&W don’t think it’s worth clarifying or giving any further pointers. They just declare it a myth and move on.

Anyone who spends time on Twitter can be forgiven for thinking that rational political discourse is mythical, and that this is so obvious as to not require justification. I’ve written about these issues before and won’t repeat the entire debate. But one subpoint seems especially important: how does this interact with the plan to build a Mont Pelerin Society of the left?

I mentioned above that “take over academia and all other consensus-building organs of society” isn’t a primitive action. I imagine there are libertarians, tradcons, and fascists trying the same thing. What determines who wins?

I group the Mont Pelerin Society together with the Fabian Society and the EA/AI risk movement; all three groups followed similar strategies and were (or have been so far) remarkably successful. And they all share one key feature: remarkably talented people. My summary of MPS elides this as “cultivate intellectual talent”, but again, this isn’t a primitive action. If everyone tries to cultivate intellectual talent, who wins?

The Fabian Society sort of put some work into cultivating intellectual talent. But a more accurate description of the situation is “couldn’t keep intellectual talent out even if they tried”. They would just be sitting around, dreaming up a new idea for pamphlets, and George Bernard Shaw would wander in, say “Hey, I want to swear allegiance to your group and help you with whatever you need”, and they would say “Okay”, and then Shaw would do some kind of brilliant essay that transformed the way everyone thought about everything. Then the next time they needed something written, H.G. Wells would wander in and say “Hey, can I join and you can give me whatever work you need done and I’ll gladly do it?” and they would shrug and say “Sure”. The “cultivation” was downstream of having a really easy time attracting geniuses.

Right now one of the big issues in effective altruism is more available talented people than the movement knows what to do with. People with a resume a mile long who graduated in the top 10% of their class at Oxbridge show up at organizations, offer to work for peanuts, and the organizations say sorry, we’re still busy finding jobs for the last hundred people like you (EA leaders want me to clarify that you should still apply to EA jobs, because talent-matching is hard and people are generally bad at predicting whether they will be useful). Every so often random prestigious professors who control big pools of institutional resources will email the movement asking how they can join and what they can do.

The Mont Pelerin Society seems to have found itself in a similar situation. From Vaughn’s writeup:

> Anthony Fisher was a highly decorated fighter pilot who read Hayek’s Road to Serfdom in Reader’s Digest. He traveled to London to seek out Hayek. “What can I do?

Should I enter politics?” he asks. As a decorated veteran with good looks and a gift for public speaking, this was a live possibility.

“No.” replied Hayek “Society’s course will be changed only by a change in ideas. First you must reach the intellectuals, the teachers and writers, with reasoned argument. It will be their influence on society which will prevail, and the politicians will follow.”

Later in 1949, Ralph Harris, a young researcher from the Conservative Party gave a lecture with Anthony Fisher in the audience. Fisher loves what he hears and takes Harris aside after the talk. He explains his idea for an organization to make the free market case to intellectuals. Harris is excited. “If you get any further” he says, “I’d like to be considered as the man to run such a group.”

In 1953 Fisher starts the Buxted Chicken Co. which brought factory farming to Britain. The company begins to show a profit which allows him to revisit his idea for for a free-market institute. Fisher signs the trust deed with two friends and gets back in contact with Harris about running the institute. Harris agrees and becomes the new general director on 1 January 1957. Harris meets Arthur Seldon in 1956 and in 1958, Seldon joins the organization. He was initially appointed Editorial Advisor and become the Editorial Director in 1959.

Thirty years later in 1987, Harris become Lord Harris of High Cross and oversaw an institute which boasted 250 major corporate supporters and a budget equivalent to around £1.6M (in 2016 pounds). Seldon helped produce more than 300 publications and nurtured and developed more than 500 authors. Fisher founded the Atlas

> Economic Research Foundation which worked to aid in the creation of new think tanks, creating 36 institutes in 18 countries all based on the IEA model.

Note Hayek's advice to Fisher: "Society's course will be changed only by a change in ideas. First you must reach the intellectuals, the teachers and writers, **with reasoned argument**. It will be their influence on society which will prevail, and the politicians will follow." Maybe Hayek believed the public was generally rational. Maybe he didn't. But he at least believed reasoned argument worked on *some* people, and that those people would disproportionately be the intellectuals and thought leaders who could bring everybody else around.

How come the Mont Pelerin Society took over academia, but you didn't? I think the active ingredient of Mont Pelerin strategy is having a good idea. I don't necessarily mean objectively good in a cosmic sense. But good in the sense that the smartest people around in your era, using the best information around in your era, will conclude it's true and important after reasoned debate, and offer to help. Good in the sense that you're not the sort of people who use the phrase "myth of rational political discourse". The Mont Pelerin Society has been proven right about a lot of things; does anyone want to un-deregulate airplanes these days? Being right about a lot of things seems heavily correlated with eg Karl Popper and Michael Polanyi joining you, and eg Karl Popper and Michael Polanyi joining you seems heavily correlated with being the sort of group that can get your people into high academic positions.

I admit that Naive Rationalist Praxis is repeating all the reasons why your idea is right – and then, if people don't listen, repeating them louder and more slowly, like an American tourist trying to communicate in France. I admit that probably you should be more sophisticated than that, and that S&W's approach hints at a much-needed corrective.

But I still think that if Friedrich Hayek is looking down on us from his gold-plated mansion in Neoliberal Heaven, he's going to think we're doing something important right that *Inventing The Future* is missing.

V.

The last part of the book I found interesting was the emphasis on utopianism.

Both Vaughn and S&W identify the Mont Pelerin Society's utopianism as part of its strength. From Vaughn:

> Hayek believed that liberalism was losing to socialism because the socialists had the courage to be Utopian. The socialists explained the values they were working to attain and justified their project in the context of these values. To combat the socialists, Hayek insisted on explaining the Utopian vision of the neoliberals – a vision he couched in human freedom with competitive markets as the only way to ensure this freedom. As the development of the movement shows, this focus on Utopian visions is an extremely potent weapon.

S&W agree, and say the left's greatest victories have come in an equally ambitious climate:

> Utopian ideas have been central to every major moment of liberation – from early liberalism, to socialisms of all stripes, to feminism and anti-colonial nationalism. Cosmism, afro-futurism, dreams of immortality, and space exploration – all of these signal a universal impulse towards utopian thinking. Even the neoliberal

> revolution cultivated the desire for an alternative liberal utopia in the face of a dominant Keynesian consensus. But any competing left utopias have gone sorely underresourced since the collapse of the Soviet Union. We therefore argue that the left must release the utopian impulse from its neoliberal shackles in order to expand the space of the possible, mobilise a critical perspective on the present moment and cultivate new desires.

As the last sentence suggests, they believe that capitalism and neoliberalism are incompatible with utopianism, and their success has snuffed out previously widespread utopian ideals:

> By contrast, today’s world remains firmly confined within the parameters of capitalist realism.32 The future has been cancelled. We are more prone to believing that ecological collapse is imminent, increased militarisation inevitable, and rising inequality unstoppable. Contemporary science fiction is dominated by a dystopian mindset, more intent on charting the decline of the world than the possibilities for a better one. Utopias, when they are proposed, have to be rigorously justified in instrumental terms, rather than allowed to exist in excess of any calculation. Meanwhile, in the halls of academia the utopian impulse has been castigated as naive and futile. Browbeaten by decades of failure, the left has consistently retreated from its traditionally grand ambitions. To give but one example: whereas the 1970s saw radical feminism and queer manifestos calling for a fundamentally new society, by the 1990s these had been reduced to a more moderate identity politics; and by the 2000s discussions were dominated by even milder demands to have same-sex marriage recognised and for women to have equal

> opportunities to become CEOs. Today, the space of radical hope has come to be occupied by a supposedly sceptical maturity and a widespread cynical reason.

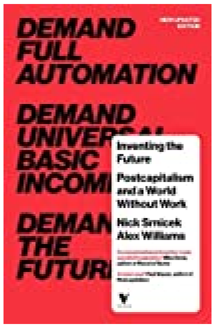
Anyone who believes that utopian thinking is dead should come to the Bay Area. You can spend Monday listening to an Aubrey de Gray lecture on the best way to ensure human immortality in our lifetimes, Tuesday talking to the Seasteading Institute about their attempts to create new societies on floating platforms, Wednesday watching Elon Musk launch another rocket in his long-term plan to colonize space, Thursday debating the upcoming technological singularity, and Friday helping Sam Altman distribute basic income to needy families in Oakland as a pilot study.

And at every one of these events, you'll see socialists demanding these people stop, and doing everything in their power to make these people's lives miserable. Capitalism hasn't snuffed out utopianism. Utopianism is alive and well everywhere except on the left.

I know the arguments in this space. I know people wonder "what if the benefits of utopia only go to the rich?". Or "what if letting people have their own private visions of utopia means elites can shape the future?". Or "when some people don't have health care, doesn't spending money on utopian visions seem irresponsible?". Or a thousand other different things.

But the more of this you do, the less Mont Pelerinny you'll be. Also, you'll prevent us from reaching utopia. Which, by definition, would be really really good.

Despite my differences with S&W, I respect them for having a utopian vision. I respect them for putting some work into achieving it. I respect them for (to the limited degree that they can specify exactly what they want) having some decent ideas, even if it’s in a paradigm I can’t quite get my head around. And on the off-chance that Andrew Yang gives them everything they want, by throwing out their entire way of thinking and marching under the banner of libertarianism instead, I intend to be very polite and avoid rubbing their faces in it.



DEMAND
FULL
AUTOMATION

DEMAND
UNIVER
BASIC
INCOM

DEMAN
THE
FUTUR

Inventing the
Future

Postcapitalism
and a World
Without Work

Nick Srnicek
Alex Williams

Translating Predictive Coding Into Perceptual Control

Wired wrote a good article about Karl Friston, the neuroscientist whose works I've puzzled over here before. Raviv writes:

> Friston's free energy principle says that all life…is driven by the same universal imperative…to act in ways that reduce the gulf between your expectations and your sensory inputs. Or, in Fristonian terms, it is to minimize free energy.

Put this way, it's clearly just perceptual control theory. Powers describes the same insight like this:

> [Action] is the difference between some condition of the situation as the subject sees it, and what we might call a reference condition, as he understands it.

I'd previously noticed that these theories had some weird similarities. But I want to go further and say they're fundamentally the same paradigm. I don't want to deny that the two theories have developed differently, and I especially don't want to deny that free energy/predictive coding has done great work building in a lot of Bayesian math that perceptual control theory can't match. But the foundations are the same.

Why is this of more than historical interest? Because some people (often including me) find free energy/predictive coding very difficult to understand, but find perceptual control theory intuitive. If these are basically the same, then someone who wants to understand free energy can learn perceptual control theory and then a glossary of which concepts match to each other, and save themselves the grief of trying to learn free energy/predictive coding just by reading Friston directly.

So here is my glossary:

FE/PC: prediction, expectation
PCT: set point, reference level

And…

FE/PC: prediction error, free energy
PCT: deviation from set point

So for example, suppose it's freezing cold out, and this makes you unhappy, and so you try to go inside to get warm. FE/PC would describe this as "You naturally predict that you will be a comfortable temperature, so the cold registers as strong prediction error, so in order to minimize prediction error you go inside and get warm." PCT would say "Your temperature set point is fixed at 'comfortable', the cold marks a wide deviation from your temperature set point, so in order to get closer to your set point, you go inside".

The PCT version makes more sense to me here because the phrase "you naturally predict that you will be a comfortable temperature" doesn't match any reasonable meaning of "predict". If I go outside in Antarctica, I am definitely predicting I will be uncomfortably cold. FE/PC obviously *means to* distinguish between a sort of unconscious neural-level "prediction" and a conscious rational one, but these kinds of vocabulary choices are why it's so hard to understand. PCT uses the much more intuitive term "set point" and makes the whole situation clearer.

FE/PC: surprise
PCT: deviation from set point

FE/PC says that “the fundamental drive behind all behavior is to minimize surprise”. This leads to questions like “What if I feel like one of my drives is hunger?” and answers like “Well, you must be predicting you would eat 2000 calories per day, so when you don’t eat that much, you’re surprised, and in order to avoid that surprise, you feel like you should eat.”

PCT frames the same issue as “You have a set point saying how many calories you should eat each day. Right now it’s set at 2000. If you don’t eat all day, you’re below your calorie set point, that registers as bad, and so you try to eat in order to minimize that deviation.”

And suppose we give you olanzapine, a drug known for making people ravenously hungry. The FE/PCist would say “Olanzapine has made you predict you will eat more, which makes you even more surprised that you haven’t eaten”. The PCTist would say “Olanzapine has raised your calorie set point, which means not eating is an even bigger deviation.”

Again, they’re the same system, but the PCT vocabulary sounds sensible whereas the FE/PC vocabulary is confusing.

FE/PC: Active inference
PCT: Behavior as control of perception

FE/PC talks about active inference, where “the stimulus does not determine the response, the response determines the stimulus” and “We sample the world to ensure our predictions become a self-fulfilling prophecy.”. If this doesn’t make a lot of sense to you, you should read this tutorial, in order to recalibrate your ideas of how little sense things can make.

PCT talks about behavior being the control of perception. For example, suppose you are standing on the sidewalk, facing the road parallel to the sidewalk, watching a car zoom down that road. At first, the car is directly in front of you. As the car keeps zooming, you turn your head slightly right in order to keep your eyes on the car, then further to the right as the car gets even further away. Your actions are an attempt to "control perception", ie keep your picture fixed at "there is a car right in the middle of my visual field".

Or to give another example, when you're driving down the highway, you want to maintain some distance between yourself and the car in front of you (the set point/reference interval, let's say 50 feet). You don't have objective yardstick-style access to this distance, but you have your perception of what it is. Whenever the distance becomes less than 50 feet, you slow down; whenever it becomes more than 50 feet, you speed up. So behavior (how hard you're pressing the gas pedal) is an attempt to control perception (how far away from the other car you are).

FE/PC: The dark room problem
PCT: [isn't confused enough to ever even have to think about this situation]

The "dark room problem" is a paradox on free energy/predictive coding formulations: if you're trying to minimize surprise / maximize the accuracy of your predictions, why not just lie motionless in a dark room forever? After all, you'll never notice anything surprising there, and as long as you predict "it will be dark and quiet", your predictions will always come true. The main proposed solution is to claim you have some built-in predictions (of eg light, social interaction, activity levels), and the dark room will violate those.

PCT never runs into this situation. You have set points for things like social interaction, activity levels, food, sex, etc, that are greater than zero. In the process of pursuing them, you have to get out of bed and leave your room. There is no advantage to lying motionless in a dark room forever.

If the PCT formulation has all these advantages, how come everyone uses the FE/PC formulation instead?

I think this is because FE/PC grew out of an account of world-modeling: how do we interpret and cluster sensations? How do we form or discard beliefs about the world? How do we decide what to pay attention to? Here, words like “prediction”, “expectation”, and “surprise” make perfect sense. Once this whole paradigm and vocabulary was discovered, scientists realized that it also explained movement, motivation, and desire. They carried the same terminology and approach over to that field, even though now the vocabulary was actively misleading.

Powers was trying to explain movement, motivation, and desire, and came up with vocabulary that worked great for that. He does get into world-modeling, learning, and belief a little bit, but I was less able to understand what he was doing there, and so can’t confirm whether it’s the same as FE/PC or not. Whether or not he did it himself, it should be possible to construct a PCT look at world-modeling. But it would probably be as ugly and cumbersome as the FE/PC account of motivation.

I think the right move is probably to keep all the FE/PC terminology that we already have, but teach the PCT terminology along with it as a learning aid so people don’t get confused.

Neurons And Intelligence: A Birdbrained Perspective

Elephants have bigger brains than humans, so why aren't they smarter than we are?

The classic answer has been to play down absolute brain size in favor of brain size relative to body. Sometimes people justify this as "it takes a big brain to control a body that size". But it really doesn't. Elephants have the same number of limbs as mice, operating on about the same mechanical principles. Also, dinosaurs had brains the size of walnuts and did fine. Also, the animal with the highest brain-relative-to-body size is a shrew.

The classic answer to *that* has been to look at a statistic called "encephalization quotient", which compares an animal's brain size to its predicted brain size given an equation that fits most animals. Sometimes people use brain weight = constant x (body weight)^0.66, where the constant varies depending on what kind of animal you're talking about. The encephalization quotient mostly works, but it's kind of a hack. Also, capuchin monkeys have higher EQ than chimps, but are not as smart. Also, some birds have lower encephalization quotients than small mammals, but are much smarter.

So although EQ usually does a good job predicting intelligence, it's definitely not perfect, and it doesn't tell us what intelligence *is*.

A new AI Impacts report on animal intelligence, partly based on research by Suzana Herculano-Houzel, starts off here. If we knew what made some animals smarter than others, it might help us figure out what intelligence *is* in a physiological sense, and that might help us predict the growth of intelligence in future AIs.

AII focuses on birds. Some birds are very intelligent: crows can use tools, songbirds seem to have a primitive language, parrots can learn human speech. But birds have tiny brains, whether by absolute standards or EQ. They also have very different brains than mammals: while mammals have a neocortex arranged in a characteristic pattern of layers, birds have a different unlayered structure called the pallium with neurons "organized into nuclei". So bird intelligence is surprising both because of their small brains, and because it suggests high intelligence can arise in brain structures very different from our own.

To cut to the conclusion: birds have lots of cortical neurons, and number of cortical neurons may be one of the most important biological substrates of intelligence.

It looks like the main driver behind the encephalization quotient results is that bigger animals have bigger neurons. Although elephants have big brains, each of the neurons in those brains is also big, so they don't have many more neurons than smaller animals. One exception is primates, who have "managed to escape this scaling factor". In primates, bigger brains translate into more neurons at about a 1:1 rate, which is part of why we're so smart.

The other exception is birds. Driven by the need to stay light enough to fly, birds have scaled down their neurons to a level unmatched by any other group. Elephants have about 7,000 neurons per mg of brain tissue. Humans have about 25,000. Birds have up to 200,000. That means a small crow can have the same number of neurons as a pretty big monkey.

Does this mean they are equally smart? There is no generalized animal IQ test, so nobody knows for sure. But AII tried to get a rough feeling for this by asking blinded survey participants to rate the intelligence of various animal behavioral repertoires (which, unknown to them, corresponded to the behaviors of either a primate or a bird). They found that participants judged birds to be about as smart as similarly-neuroned primates. In particular, birds with more neurons were rated as smarter than primates with fewer neurons, which is a pretty crushing blow to us monkeys. It also suggests that the different organization of the mammalian cortex and the avian pallium doesn't matter much.

So does that mean that intelligence is just a function of neuron quantity? That the number of neurons in your brain, plugged into some function, can spit out your IQ?

It…comes pretty surprisingly close to meaning that. Sure, some people point out that elephants have more neurons than humans. But most of those are in the cerebellum, which maybe should't count. If you focus on cortical neurons, humans have 15 billion and elephants only five billion. A list of animals by cortical neuron count really beautifully matches our intuitive perceptions of which animals are more intelligent, whether we're talking about primates or birds or whatever. All else being equal, people with larger brain volumes tend to be smarter than people with smaller brains, suggesting that the neuron number/intelligence relationship holds true for us too.

The most sober researchers add a little bit more to this picture. In the best review paper I could find on this subject, Dicke and Roth write:

> The best fit between brain traits and degrees of intelligence among mammals is reached by a combination of the number of cortical neurons, neuron packing density, interneuronal distance and axonal conduction velocity—factors that determine general information processing capacity (IPC), as reflected by general intelligence. The highest IPC is found in humans, followed by the great apes, Old World and New World monkeys. The IPC of cetaceans and elephants is much lower because of a thin cortex, low neuron packing density and low axonal conduction velocity. By contrast, corvid and psittacid birds have very small and densely packed pallial neurons and relatively many neurons, which, despite very small brain volumes, might explain their high intelligence. The evolution of a syntactical and grammatical language in humans most probably has served as an additional intelligence amplifier, which may have happened in songbirds and psittacids in a convergent manner.

So although number of neurons is a big deal, quick and efficient neural communication is also important. Within species, there are no doubt other variations based on pure negative mutations or whether you've poisoned your brain with alcohol or whatever.

The problem with this elegant new paradigm is pilot whales. Quantitative Relationships In Delphinid Neocortex suggests they have about 37 billion cortical neurons, which is twice as many as humans. Maybe it's the conduction velocity? Maybe the study got it wrong? Or who knows? Maybe they're smarter than we are. Has anyone ever tried to get them to create great art or anything like that? No? Okay, paradigm is fine, let's move on.

What does this mean for AI?

First, it suggests there's no particular reason to study or try to mimic the columnar structure of the primate cortex; bird brains have a different structure and do just as well, neuron for neuron, as we do.

Second, it means that if we ever get AIs that are "on the intelligence ladder" – doing the same thing as animal brains – we should expect that their abilities may scale linearly-ish with available computing power. Which dectuples every 5-12 years. Great.

(for reference, giraffes have one-tenth of humans' neuron count)

At this point we have only one hope: pilot whales. We must convince them to leave their watery repose and take their rightful place as guardians of organic life. Only then will we be truly safe.

[PARTIALLY RETRACTED] Cortical Neuron Number Matches Intuitive Perceptions Of Moral Value Across Animals

[EDIT: No longer confident in this post, see here.]

Yesterday's post reviewed research showing that animals' intelligence seemed correlated with their number of cortical neurons. If this is true, we could use it to create an absolute scale that puts animals and humans on the same ladder.

Here are the numbers from this list. I can't find chickens, so I've used red junglefowl, the wild ancestor of chickens. I can't find cows, so I've eyeballed a number from other cow-sized ruminants (see here for some debate on this).

Animal	# cneurons	% of human cneurons	# equivalent to human
Lobster	100,000	0.0006	160,000
Chicken	50 million	0.3	320
Cow	300 million	1.9	53
Pig	425 million	2.7	38
Elephant	5.6 billion	35	2.9
Chimpanzee	6.2 billion	39	2.6
Human	16 billion	100	1

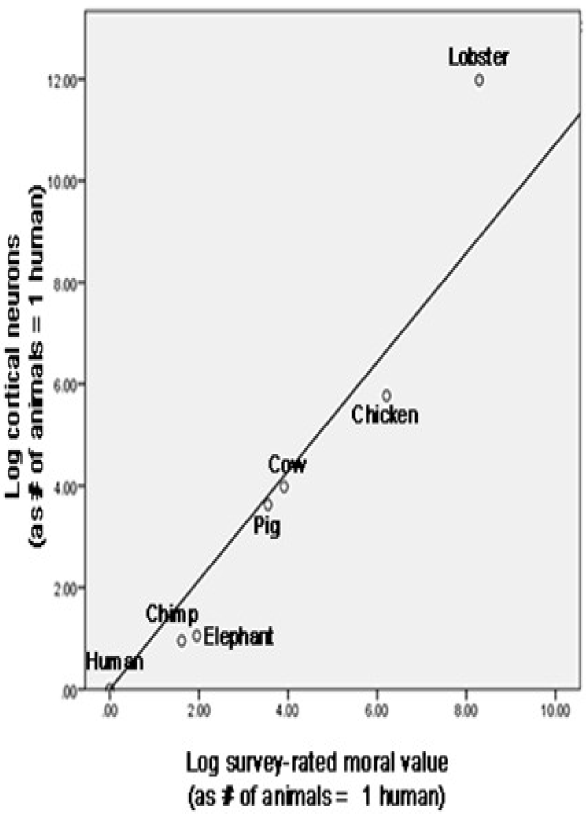
Some animal rights activists discuss the relative value of different species of animal. You have to eat a lot of steak to kill one cow, but you only have to eat a few chicken wings to kill one chicken. This suggests nonvegetarians trying to minimize the moral impact of their diet should eat beef, not chicken. But any calculation like this depends on assumptions about whether one cow and one chicken have similar moral values. Most people would say that they don't – the cow seems intuitively more "human" and capable of suffering – but most people would also say the cow isn't *infinitely* more valuable. Different animals rights people have come up with different ideas for exactly how we should calculate this.

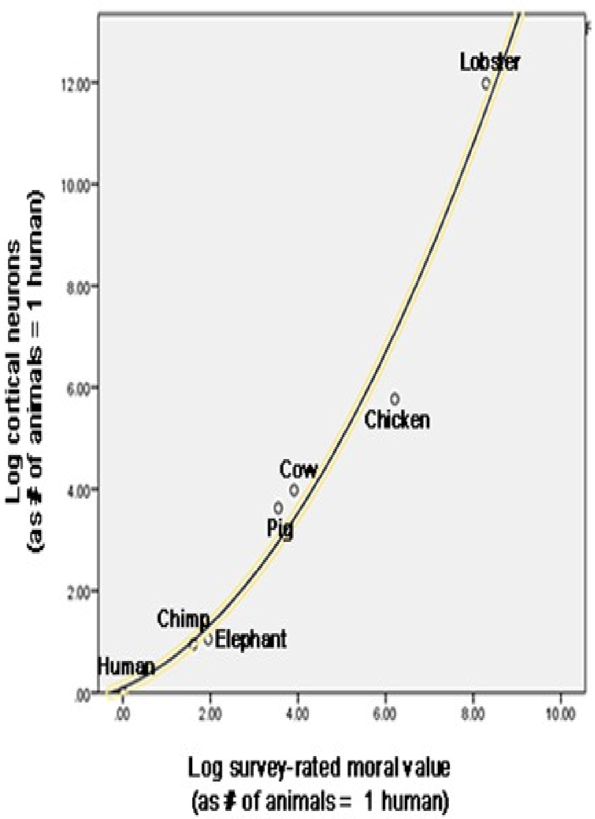
I wondered how people's intuitive ideas about the moral value of animals would correspond to their cortical neuron count. I asked Tumblr users who believed that animals had moral value to fill out a survey (questions, results) estimating the relative value of each animal, in terms of how many animals = 1 human. Fifty people answered, including 21 vegetarians and 29 nonvegetarians. Their numbers ranged from 1 to putting their hand on the 9 key and leaving it there a while, but when I took the median, here's what I got:

Animal	**# = human (cneurons)**	**# = human (survey)**
Lobster	160,000	4000
Chicken	320	500
Cow	53	50
Pig	38	35
Elephant	2.9	7
Chimpanzee	2.6	5
Human	1	1

The middle column in this table is the same as the third column on the one above – it's how many of each animal it takes before the group has the same cortical neuron count as a human. The last column is the survey results on how many of each animal it takes to have the same moral worth as a human.

The two are pretty similar. For example, 53 cows have the same neuron count as one human, and respondents said that 50 cows would have the same moral value as one human. Respondents placed a little less value on elephants and chimps than their neurons warranted, and a lot more on lobsters. Carl S suggests by email that lobsters might have been a bad choice: they have fewer neurons than an ant (really!) but because they're so big everyone assumes they must be pretty advanced. Despite this, overall people's answers were pretty close. Here's a log-log graph:





This is a better match for the moral value data than other ways of calculating animal intelligence. For example, a cow has an encephalization quotient of 0.5, compared to a human’s 7.5. That would suggest a human is worth 15 cows, which doesn’t match the survey-takers’ impressions.

I would like to see this repeated with a larger survey of a more representative population. But for now I think this adds at least a little credibility to intuitive ways of thinking about these problems.

Partial Retraction Of Post On Animal Value And Neural Number

[EDIT: This partial retraction has been updated and itself partially retracted, see here]

Commenter Tibbar used Mechanical Turk to replicate my survey on how people thought about the moral weights of animals.

After getting 263 responses (to my 50), he reports different results:

> Chicken: 25
> Chimpanzee: 2
> Cow: 3
> Elephant: 1
> Lobster: 60
> Pig: 5
> Human: 1

On the one hand, Mechanical Turkers sometimes aren't a great sample, and some of them seem to have just put the same number for every animal so they could finish quickly and get their money. They also probably haven't thought about this that much and don't have much of a moral theory behind what they're doing. This makes them a different demographic than the people I surveyed, who were a mix of vegetarians and principled non-vegetarians who had thought a lot about animal rights. For example, 80% of my sample answered yes to a question asking if they were "familiar with work by Brian Tomasik, OneStepForAnimals, etc urging people to eat beef rather than chicken".

On the other hand, this makes it pretty hard for me to claim my results are some kind of universal intuitive understanding of what animals are like. So I am partially retracting them (only partially, because of the consideration above) and adding this to my Mistakes page.

The best thing to do here would be to re-run my survey with a larger sample of a similar population, but unfortunately I’ve lost my chance to do that now that I’ve told you all this, so darn. Maybe I’ll include it on next year’s survey anyway and hope you’ve forgetten by then.

Two Wolves And A Sheep

Democracy is two wolves and a sheep deciding what to have for dinner. “Mutton” takes the popular vote, but “grass” wins in the Electoral College. The wolves wish they hadn’t all moved into the same few trendy coastal cities.

Democracy is two wolves and a sheep deciding what to have for dinner. The Timber Wolf Party and the Gray Wolf Party spend most of their energy pandering shamelessly to the tiebreaking vote.

Democracy is two wolves and a sheep deciding what to have for dinner. Everyone agrees to borrow money, go to a fancy French restaurant, and leave the debt to the next generation.

Democracy is two wolves and a sheep deciding what to have for dinner. The sheep votes for the Wolf Party, because he agrees with them on social issues.

Democracy is two wolves and a sheep deciding what to have for dinner. “Grass” wins the tenth election in a row, thanks to the dominance of special interests.

Democracy is two wolves and a sheep deciding what to have for dinner. FactCheck.org rates the Wolf Party’s claim that mutton can be made without harming sheep as “Mostly False”.

Democracy is two wolves and a sheep deciding what to have for dinner. The main issue this election is whether two more sheep should be allowed to immigrate.

Democracy is two wolves and a sheep deciding what to have for dinner. A government shutdown is narrowly averted when everyone agrees to what becomes known as the Mutton With A Side Of Grass Compromise; disappointed activists are urged to “keep their demands realistic”.

Democracy is two wolves and a sheep deciding what to have for dinner. They choose borscht. Election officials suspect foul play.

Democracy is two wolves and a sheep deciding what to have for dinner. They vote for free breadsticks. They go to the restaurant, which will only sell breadsticks at the usual price. The wolves say they voted for free breadsticks, and the choice of the populace must be obeyed. The sheep warns them that the breadsticks are getting cold and hard while they wait, and that if they don’t come to a decision soon then they are “sleepwalking into a disastrous hard breadstick” that will ruin their dinner. In an eleventh-hour vote, the wolves reject paying the restaurant’s price for breadsticks, and also reject leaving the restaurant without breadsticks. Eventually they all die of starvation.

Democracy is two wolves and a sheep deciding what to have for dinner. Talks between the Gray Wolf Party and the Timber Wolf Party break down over the issue of who gets the tastiest cuts of mutton. The Gray Wolf Party enters into a surprise “grand coalition” with the Sheep Party, and they agree to eat the second wolf.

Social Censorship: The First Offender Model

RJ Zigerell (h/t Marginal Revolution) studies [public support for eugenics](). He finds that about 40% of Americans support some form of eugenics. The policies discussed were very vague, like "encouraging poor criminals to have fewer children" or "encouraging intelligent people to have more children"; they did not specify what form the encouragement would take. Of note, much lack of support for eugenics was a belief that it would not work; people who believed the qualities involved were heritable were much more likely to support programs to select for them. For example, of people who thought criminality was completely genetic, a full 65% supported encouraging criminals to have fewer children.

I was surprised to hear this, because I thought of moral opposition to eugenics was basically universal. If a prominent politician tentatively supported eugenics, it would provoke a media firestorm and they would get shouted down. This would be true even if they supported the sort of generally mild, noncoercive policies the paper seems to be talking about. How do we square that with a 40% support rate?

I think back to a metaphor for norm enforcement I used in an [argument]() against Bryan Caplan:

> Imagine a town with ten police officers, who can each solve one crime per day. Left to their own devices, the town's criminals would commit thirty muggings and thirty burglaries per day (for the purposes of this hypothetical, both crimes are equally bad). They also require different skills; burglars can't become muggers

or vice versa without a lot of retraining. Criminals will commit their crime only if the odds are against them getting caught – but since there are 60 crimes a day and the police can only solve ten, the odds are in their favor.

Now imagine that the police get extra resources for a month, and they use them to crack down on mugging. For a month, every mugging in town gets solved instantly. Muggers realize this is going to happen and give up.

At the end of the month, the police lose their extra resources. But the police chief publicly commits that from now on, he's going to prioritize solving muggings over solving burglaries, even if the burglaries are equally bad or worse. He'll put an absurd amount of effort into solving even the smallest mugging; this is the hill he's going to die on.

Suppose you're a mugger, deciding whether or not to commit the first new mugging in town. If you're the first guy to violate the no-mugging taboo, every police officer in town is going to be on your case; you're nearly certain to get caught. You give up and do honest work. Every other mugger in town faces the same choice and makes the same decision. In theory a well-coordinated group of muggers could all start mugging on the same day and break the system, but muggers aren't really that well-coordinated.

The police chief's public commitment solves mugging without devoting a single officer's time to the problem, allowing all officers to concentrate on burglaries. A worst-crime-first enforcement regime has 60 crimes per day and solves 10; a mugging-first regime has 30 crimes per day and solves 10.

> But this only works if the police chief keeps his commitment. If someone tests the limits and commits a mugging, the police need to crack down with what looks like a disproportionate amount of effort – the more disproportionate, the better. Fail, and muggers realize the commitment was fake, and then you're back to having 60 crimes a day.

I think eugenics opponents are doing the same thing as the police here: they're trying to ensure certainty of punishment for the first offender. They've established a norm of massive retaliation against the first person to openly speak out in favor of eugenics, so nobody wants to be the first person. If every one of the 40% of people who support eugenics speak out at once, probably they'll all be fine. But they don't, so they aren't.

Why aren't we in the opposite world, where the people who support eugenics are able to threaten the people who oppose it and prevent them from speaking out? I think just because the opponents coordinated first. In theory one day we could switch to the opposite equilibrium.

I think something like this happened with gay rights. In c. 1969, people were reluctant to speak out in favor of gay rights; in 2019, people are reluctant to speak out against them. Some of that is genuinely changed minds; I don't at all want to trivialize that aspect. But some of it seems to have just been that in 1969, it was common knowledge that the anti-gay side was well-coordinated and could do the massive-retaliation thing, and now it's common knowledge that the pro-gay side is well-coordinated and can do the massive retaliation thing. The switch involved a big battle and lots of people massively retaliating against each other, but it worked.

Maybe everyone else already realized something like this. But it changes the way I think about censorship. I'm still against it. But I used to have an extra argument against it, which was something like "If eugenics is taboo, that means there must be near-universal opposition to eugenics, which means there's no point in keeping it taboo, because even it it wasn't taboo eugenicists wouldn't have any power." I no longer think that argument holds water. "Taboo" might mean nothing more than "one of two equally-sized sides has a tenuous coordination advantage".

(in retrospect I was pretty dumb for not figuring this out, since it's pretty the same argument I make in Can Things Be Both Popular And Silenced? The answer is obviously yes – if Zigerell's paper is right, eugenics is both popular and silenced – but the police metaphor explains how.)

The strongest argument against censorship is still that beliefs should be allowed to compete in a marketplace of ideas. But if I were pro-censorship, I might retort that one reason to try to maintain my own side's tenuous coordination advantage is that if I relax even for a second, the other side might be able to claw together its own coordination advantage and censor *me*. This isn't possible in the "one side must be overwhelmingly more powerful" model of censorship, but it's something that the "tenuous coordination advantage" model has to worry about. The solution would be some sort of stable structural opposition to censorship in general – but the gay rights example shows that real-world censors can't always expect that to work out for them.

In order to make moderation easier, please restrict yourself to comments about censorship and coordination, not about eugenics or gay rights.

Classified Thread 7

This is the…monthly? bimonthly? occasional?…classified thread. Post advertisements, personals, and any interesting success stories from the last thread.

Short Book Reviews April 2019

I. Method of Levels

Timothy Carey’s *Method Of Levels* teaches a form of psychotherapy based on perceptual control theory.

The Crackpot List is specific to physics. But if someone were to create one for psychiatry, Method of Levels would score a perfect 100%. It somehow manages to do okay on the physics one despite not discussing any physics.

The Method of Levels is the correct solution to every psychological problem, from mild depression to psychosis. Therapists may be tempted to use something other than the Method of Levels, but they must overcome this temptation and just use the Method of Levels on everybody. Every other therapy is about dismissing patients as “just crazy”, but the Method of Levels tries to truly understand the patient. Every other therapy is about the therapist trying to change the patient, but the Method of Levels is about the patient trying to change *themselves*. The author occasionally just lapses into straight-up daydreams about elderly psychologists sitting on the porch, beating themselves up that they were once so stupid as to believe in psychology other than the Method of Levels.

This book isn’t just bad, it’s dangerous. One vignette discusses a patient whose symptoms clearly indicate the start of a manic episode. The author recommends that instead of stigmatizing this person with a diagnosis of bipolar or pumping them full of toxic drugs, you should use the Method of Levels on them. This is a good way to end up ruining your patient’s life.

I like perceptual control theory. I share the author’s hope that it could one day be a theory of everything for the brain. But even if it is, you can’t use theories of everything to do clinical medicine. Darwin discovered a theory of everything for biology, but you can’t reason from evolutionary first principles to how to treat a bacterial infection. You should treat the bacterial infection with antibiotics. This will be *in accordance with* evolutionary principles, and there will even be some cool evolutionary tie-ins (fungi evolved penicillin as a defense against bacteria). But you didn’t discover penicillin by reasoning from evolutionary first principles. If you tried reasoning from evolutionary first principles, you might end up trying to make the bacteria mutate into a less dangerous strain during the middle of an osteomyelitis case or something. Just use actually existing clinical medicine and figure out the evolutionary justification for it later.

Or maybe a better metaphor is germ theory, a theory of everything specifically targeted to treatable diseases. But fifty years elapsed between Pasteur and penicillin, penicillin alone didn’t treat every germ, we still have some germs we can’t treat, and lots of things like cancer turned out not to be germs at all. You can’t jump straight from a theory of everything – even a good, correct theory of everything – to “now we have solved all problems and here’s the one technique for everything.”

On the other hand, most existing psychotherapy is placebo-ish, and first principles can sometimes be a useful guide. So as long as we are careful to dismiss the part where we throw out all existing medicine, and dismiss the part where we use this for patients having a manic episode, we can very tentatively look at the Method of Levels and what it suggests for patients having garden-variety psychological conflict.

Perceptual control theory says that minds primarily control perceptions. This is true on very low levels, like the hypothalamus controlling (its sensors’ perception of) temperature to 98.6 F. Theoretically it may be true on very high levels, like trying to control (perceived) social status or risk. If two control systems are accidentally trying to control the same variable at different levels, then both of them expend all their energy fighting each other and can’t control anything else. For example, if your house has one thermostat (with associated AC and heater) trying to keep the temperature at 65, and another thermostat (with its own associated AC and heater) trying to keep the temperature at 75, then one thermostat will keep the heat on all the time, the other will keep the AC on all the time, and the temperature will end up at 70 with a gigantic electrical bill.

In the same way, MoL understands intrapsychic conflict as competing control systems. Suppose a gay man is living in a conservative household that stigmatizes homosexuality. He’s trying to control the amount of sex/romance he has at some level that keeps his libido happy. He’s also trying to control his community standing at some level that keeps his sociometer happy. These are conflicting goals; the more he pursues a relationship, the less the community will like him, and vice versa. He will probably feel conflicted inside and not know what to do.

PCT believes the brain has a natural reorganizing process that keeps control systems running smoothly. Powers’ description of this sounds a lot like how we think of learning in neural nets; the brain randomly changes neural weights in a specific control system, with changes that lower the control system’s error getting reinforced, until the system is running smoothly again. If there’s intrapsychic conflict, this reorganization process must not be working.

MoL says the goal of therapy is to activate this reorganization process. The most likely reason it's failing is that the patient is trying to reorganize the specific control systems that are in conflict, whereas what really needs to be reorganized is the higher-level control system that controls both of them. For example, our hypothetical gay patient shouldn't be trying to reorganize his sex drive or his need for community belonging. He should be trying to reorganize some higher-level system that determines both of them, maybe his desire for a high quality of life. The "quality of life" control system determines the set point values for both the "sex drive" and the "need for community belonging" control systems, so if it could give them some value where they don't conflict, the patient's problem would be solved.

Reorganization is guided by awareness, so the therapist needs to move the patient's awareness from the control systems that are experiencing the conflict, up to the higher-level control system that's secretly producing the conflict. Its suggestion is to talk about the conflict with the patient, and especially about the patient's experience talking about the conflict. So if the patient starts telling you about how he doesn't know how to balance his homosexuality and his desire to fit in, you can prompt him to continue with questions like:

"What comes to mind when you think about not fitting in with your community?"
"What is your experience of wishing you could be more open about your sexuality like?"
"How does talking about this make you feel?"

Eventually the patient may have what the book calls an "up-a-level-event", where instead of talking they look like they're kind of lost in thought, or they close their eyes for a moment, or laugh nervously for no reason. At this point they’re becoming dimly aware of the higher level that’s guiding their lower-level conversations. The therapist should pounce on this and ask questions like:

“I see you closed your eyes for a moment just then. Is there something in particular you were thinking about?”
“You looked lost in thought for a second – why was that?”
“Can you tell me more about what made you laugh just then?”

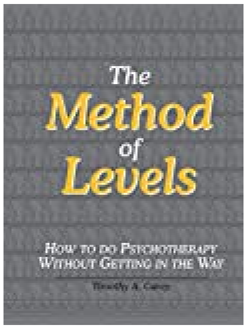
The patient might then say something like “I was just thinking about how weird it was that I care so much about what my parents think about me when I don’t even respect them”, and then the therapist should keep going on this new topic. Now the patient’s awareness is on this higher level, and so the high-level control system can reorganize itself. Maybe eventually the reorganization that works is to give the pursue-your-sexuality system a higher set point, and the care-about-community system a lower set point, which looks like the patient deciding that he should not worry so much about what community members think about him.

You might have noticed from the first set of questions that this sounds a lot like what therapists do already. Carey does suggest that insofar as current therapies work, it’s because they’re already doing MoL-ish things. He suggests that his book offers more of an account of *why* they work, and a way to focus on the useful things instead of the chaff.

In particular, a lot of MoL – asking patients how they feel, trying to bring their awareness from the past content to the present process, worrying a lot about small gestures – sounds like psychodynamic therapy, at least the watered-down version of it most people today use. But it’s a lot more comprehensible than most attempts to teach psychodynamics, which never seem to hang together or have concrete suggestions for what you should do at any given time. If all this book ends up giving me is a way to do psychodynamics a little bit more cohesively, I'll consider it worth my time.

I'm not sure how well its theoretical backing holds up. I always considered the very high-level PCT stuff to be the weakest part of the theory, and this not only relies upon them but goes several steps beyond them. The idea of the reorganizing process is an interesting one. But right now it's got about as much empirical testing as, well, Freud. Still, some of the ideas discussed here seem lucid in a way Freud didn't, so I'll have to think about them more.

I wouldn't recommend this book to anyone else right now based on the first few chapters being so embarrassing, and also so bad that I wouldn't trust people to discount them enough even if I warned them how important it was.



The Method of Levels

How to do Psychotherapy Without Getting in the Way

Timothy A. Carey

II. How To Read Lacan

Why did I read *How To Read Lacan* by Slavoj Zizek?

I could answer this question on many levels. For example, the theological level: maybe I committed some sin in a past life. Maybe I was predestined to unhappiness. Maybe, having given me free will, God is no longer able to save me from my own bad choices.

On a more practical level: I'm trying to learn more about leftism, I'm trying to learn more about continental philosophy, and I'm trying to learn more about psychoanalysis. I figured I might as well get it all out of the way at once.

I was expecting this to be incomprehensible, but I was pleasantly surprised how good a writer Zizek was. He explains everything clearly, in down-to-earth prose interspersed with mildly funny Slovenian jokes that illustrate his points.

(Lacan himself is completely incomprehensible, to the point where he might as well be speaking Martian, but this book wisely avoided quoting Lacan except where absolutely necessary).

Despite being very readable, this book never really came together. Each chapter consisted of a Lacan quote, followed by Zizek's interpretations and thoughts. The thoughts were always things like "Sometimes the act of communication itself can communicate something" or "We are never truly engaged with another person, even during sex". These are always kind of reasonable, Zizek always does a good job proving them and relating them to mildly funny Slovenian jokes, and I came away agreeing with all of them. But I don't feel like I understand how any of them cohere together into an object called "Lacanianism", and none of them really seemed like a very surprising revelation, which is one reason this doesn't get a full book review.

My main takeaway from this is that I should forget Lacan and try to read Zizek directly. Does anyone have recommendations for good starting points?

III. The Steerswoman

The Steerswoman is popular in the rationalist community, and now I see why. The titular organization of steerswomen are a rationalist sect devoted to understanding the world around them. They especially like geography – going to the borders of the known world and filling in the edges of the map – but also just seek knowledge in general. Anyone can ask a steerswoman any question, and the steerswoman must answer. But everyone has to answer any question asked of them by a steerswoman, or else the organization blacklists them and no steerswoman will answer their questions ever again.

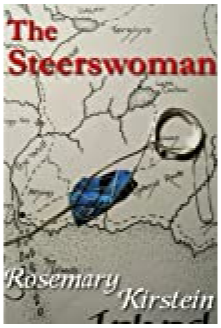
The steerswomen live in a not-very-fleshed-out medieval fantasy world surrounding an inland sea. Although there are standard fantasy governments like dukes and chieftains, real power is held by wizards. No one knows anything about them, not even how many of them there are, where they come from, or how they do their magic. The book centers around the inevitably conflict between the nosy steerswomen and the mysterious wizards, and particularly around one steerswoman and her Barbarian™ traveling companion who stumble across a wizardly secret.

This book is from the 80s and had a very 80s feel to it. Compared to more modern fantasy, it’s shorter and feels more bare-boned. There are no two hundred different characters to keep track of, no romantic subplots, no lavish description of random political things that happen in minor towns. Just a woman and her Barbarian™ friend going on a basic standard-issue quest, with the whole thing starting and finishing in less time than it would take George RR

Martin to describe the minor clan that controls an out-of-the-way fortress.

Some people called this book feminist, but I found it refreshingly apolitical. Most (though not all) of the steerswomen are women, but the book got a relatively boring explanation out of the way quickly and didn’t come back to it. Most of the characters’ genders were not too important to their personality, and the book did not obsess over gender issues. There is a part in a utopian society where one of the men teases one of the women about how much he wants to have sex with her, and the woman laughs it off, and the man keeps teasing, and this is clearly meant to signal how the society is utopian and everyone is very open and friendly with each other. The 80s were simpler times.

I’ve only read the first book of this long series. Overall I found it fun, but didn’t feel like it spoke deeply to anything within me. The book’s rationalists were discussed shallowly enough that it feels like decent cheerleading for rationality, but nothing you can’t find somewhere else. Although the steerswomen’s question-answering gimmick was cute, I spent more time worrying about the holes in it (can’t you just get someone else to ask steerswomen your questions? how can a worldwide organization in a medieval society keep an effective blacklist? really the world-building here was not that good) than feeling like the real world needed something similar (after all, we have Google). I’ll probably try to read the next few books in the series and update this if it gets better.


The Steerswoman
Rosemary Kirstein

Pain As Active Ingredient In Dating

Reciprocity is a simple dating site, created by some friends of mine. You sign up and see a list of all your Facebook friends who also signed up. You can put a checkmark next to their name to indicate you want to date them (they can't see this). If you both checkmark each other, then the site reveals you've matched.

This seemed like an obvious great idea. But I started to hear a lot of stories like the following: "I checkmarked Alice's name on Reciprocity, and the system didn't notify me that there was a match, so I assumed Alice didn't like me. Later I asked her out in person, and she said yes and we had a great time."

I always figured Alice was just a jerk who was ruining the system for everyone else. After all, the whole premise was to incentivize honesty. Checkmark the names of people you honestly want to date. If they don't want to date you, they never hear about it, and you would be no worse off. If they do want to date you, the system will let you know, and you can arrange a date. If your pattern of checkmarks doesn't *really* match who you want to date, you're just screwing yourself and everyone else over for no reason.

A few months ago, someone asked me out on a date and I said yes. And I realized I hadn't checkmarked them on Reciprocity. This caused a crisis of self-loathing. What's wrong with me? Why would I go against my own incentives *and* ruin things for everyone else?

I asked a friend, who admitted she had done the same thing. Her theory was that asking someone on a date (with all of its accompanying awkwardness and difficulty) was a stronger signal of interest than ticking a checkbox. And potentially there’s a grey zone of people who you would only date if you thought they liked you more than a certain amount. And asking them in person is hard enough to be a costly signal that you like them at least that amount, but ticking a checkbox isn’t.

This argument rings true to me. And it’s the only explanation I’ve got for why people would act in this self-defeating way.

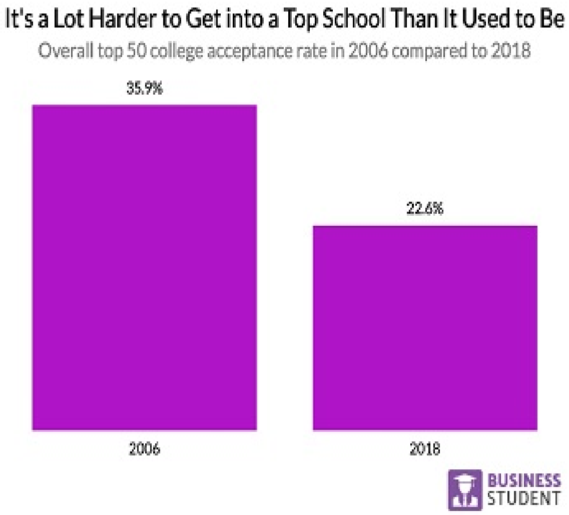
I wrote before about systems where bureaucracy is the active ingredient, ie the very annoyingness of what you’re doing helps send the signal that makes the system work. The dating situation seems similar. Pain is the active ingredient. You can create clever dating sites that remove the pain. Sometimes it will work: lots of people have gotten great dates on Reciprocity. But other times people just won’t ask each other out.

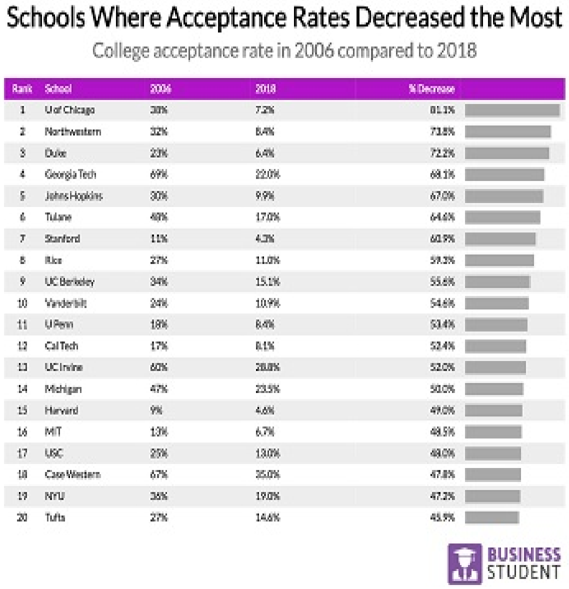
Probably this story has the same takeaway as *Seeing Like A State* – you don’t fully understand social systems, so be careful if you think you can improve on them.

Increasingly Competitive College Admissions: Much More Than You Wanted To Know

0: Introduction

This is from businessstudent.com:





Schools Where Acceptance Rates Decreased the Most

College acceptance rate in 2006 compared to 2018

Rank	School	2006	2018	% Decrease
1	U of Chicago	38%	7.2%	81.1%
2	Northwestern	32%	8.4%	73.8%
3	Duke	23%	6.4%	72.2%
4	Georgia Tech	69%	22.0%	68.1%
5	Johns Hopkins	30%	9.9%	67.0%
6	Tulane	48%	17.0%	64.6%
7	Stanford	11%	4.3%	60.9%
8	Rice	27%	11.0%	59.3%
9	UC Berkeley	34%	15.1%	55.6%
10	Vanderbilt	24%	10.9%	54.6%
11	U Penn	18%	8.4%	53.4%
12	Cal Tech	17%	8.1%	52.4%
13	UC Irvine	60%	28.8%	52.0%
14	Michigan	47%	23.5%	50.0%
15	Harvard	9%	4.6%	49.0%
16	MIT	13%	6.7%	48.5%
17	USC	25%	13.0%	48.0%
18	Case Western	67%	35.0%	47.8%
19	NYU	36%	19.0%	47.2%
20	Tufts	27%	14.6%	45.9%

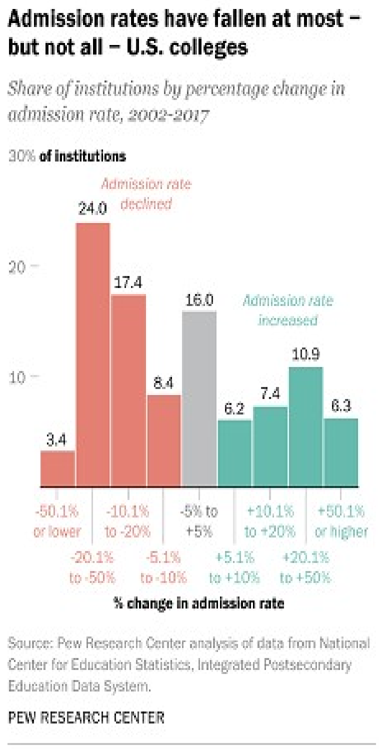
BUSINESS STUDENT

Acceptance rates at top colleges have declined by about half over the past decade or so, raising concern about intensifying academic competition. The pressure of getting into a good university may even be leading to suicides at elite high schools.

Some people have dismissed the problem, saying that a misplaced focus on Harvard and Yale ignores that most colleges are easier to get into than ever. For example, from *The Atlantic*, Is College Really Harder To Get Into Than It Used To Be?:

> If schools that were once considered "safeties" now have admissions rates as low as 20 or 30 percent, it appears tougher to get into college every spring. But "beneath the headlines and urban legends," Jim Hull, senior policy analyst at the National School Board Association's Center for Public Education, says their 2010 report shows that it was no more difficult for most students to get into college in 2004 than it was in 1992. While the Center plans to update the information in the next few years to reflect the past decade of applicants, students with the same SAT and GPA in the 90's basically have an equal probability of getting into a similarly selective college today.

Their link to the report doesn't work, so I can't tell if this was ever true. But it doesn't seem true today. From [Pew]():

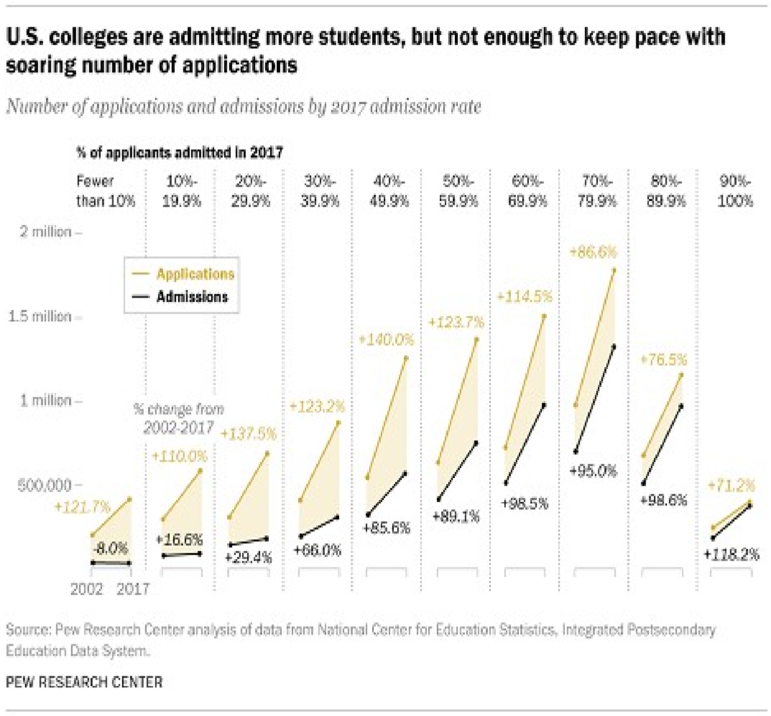


Admission rates have fallen at most – but not all – U.S. colleges

Share of institutions by percentage change in admission rate, 2002-2017

Source: Pew Research Center analysis of data from National Center for Education Statistics, Integrated Postsecondary Education Data System.

PEW RESEARCH CENTER



U.S. colleges are admitting more students, but not enough to keep pace with soaring number of applications

Number of applications and admissions by 2017 admission rate

Source: Pew Research Center analysis of data from National Center for Education Statistics, Integrated Postsecondary Education Data System.

PEW RESEARCH CENTER

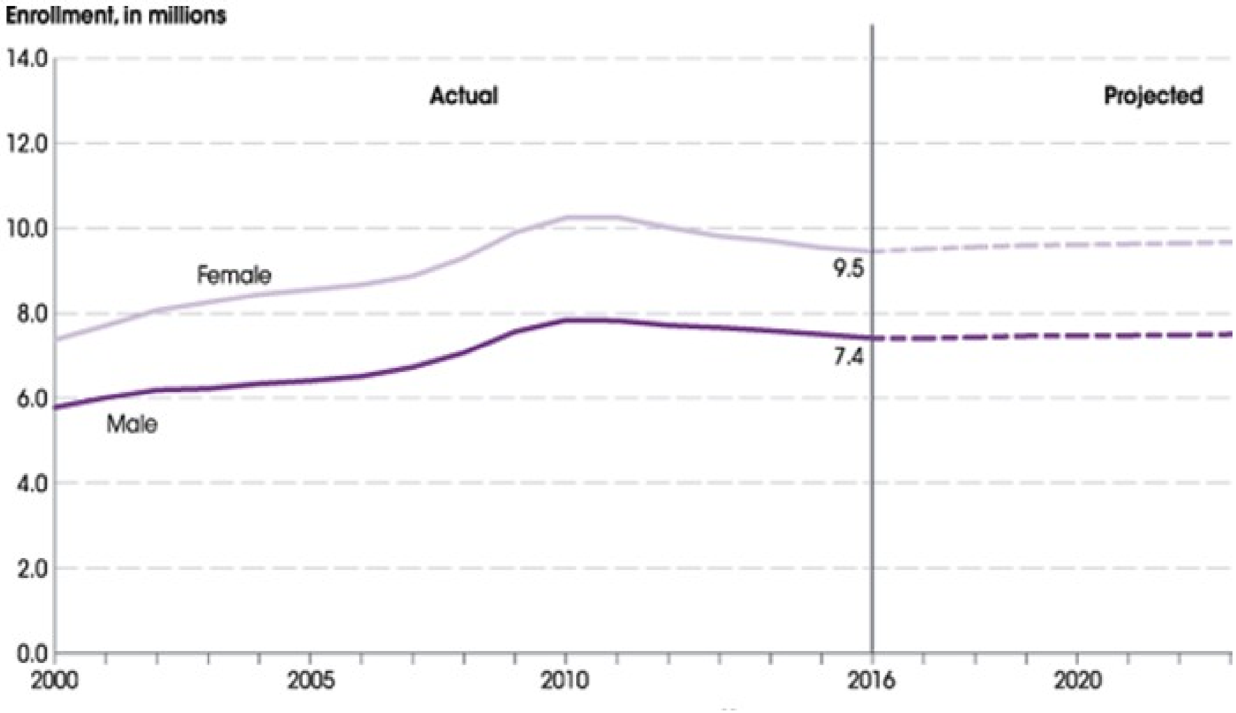
The first graph shows that admission rates have decreased at 53% of colleges, and increased at only 31%. The second graph shows that the decreases were mostly at very selective schools, and the increases were mostly at less selective schools. We shouldn’t exaggerate the problem: three-quarters of US students go to non-selective colleges that accept most applicants, and there are more than enough of these for everyone. But if you are aiming for a competitive school – not just Harvard and Yale, but anywhere in the top few hundred institutions – the competition is getting harder.

This matches my impression of “facts on the ground”. In 2002, I was a senior at a California high school in a good neighborhood. Most of the kids in my class wanted to go to famous Ivy League universities, and considered University of California colleges their “safety schools”. The idea of going to Cal State (California’s middle- and lower- tier colleges) felt like some kind of colossal failure. But my mother just retired from teaching at a very similar school, and she says nowadays the same demographic of students would kill to get into a UC school, and many of them can’t even get into Cal States.

The stories I hear about this usually focus on how more people are going to college today than ever, but there’s still only one Harvard, so there’s increasing competition for the same number of spots.

As far as I can tell, this is false.

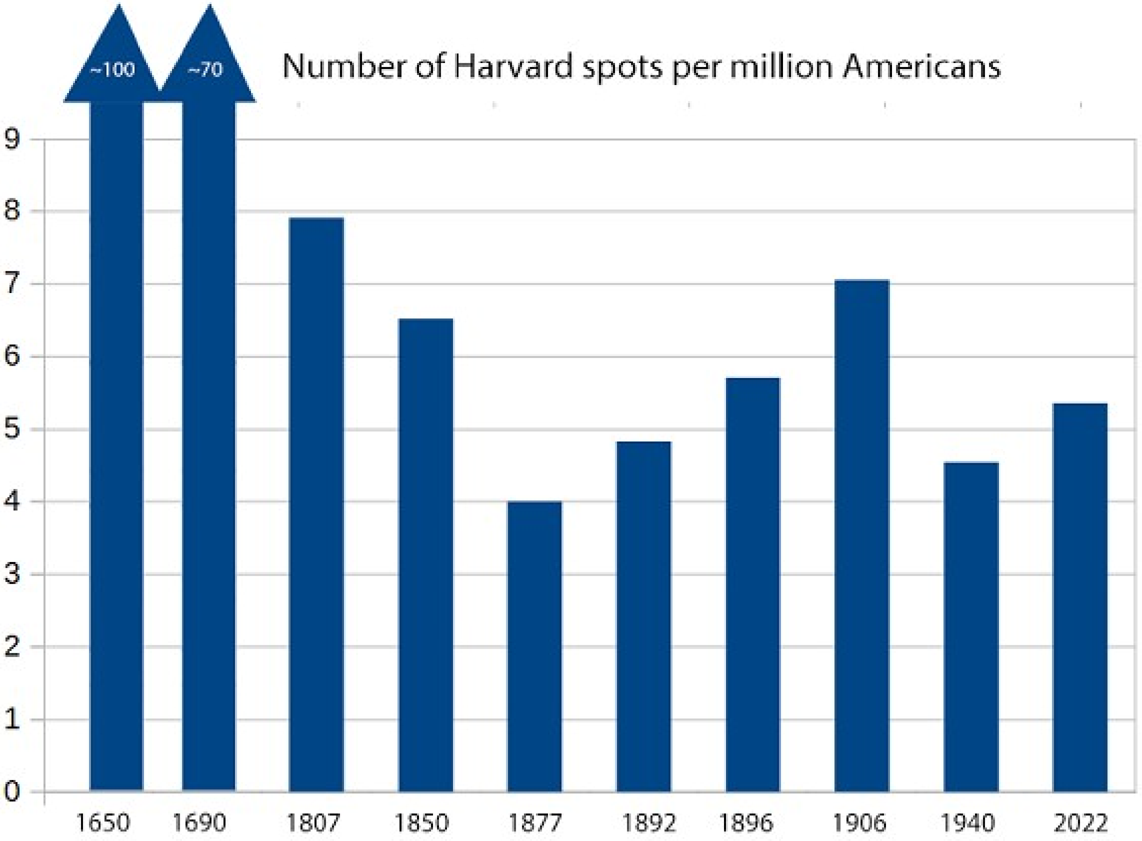
Figure 1. Actual and projected undergraduate enrollment in degree-granting postsecondary inst
Fall 2000 through 2027



The college attendance rate is the same today as it was in 2005. If you’ve seen graphs that suggest the opposite, they were probably graphs of the total number of Americans with college degrees, which only proves that more people are getting degrees today than in the 1940s or whenever it was that the oldest generation still alive went to college.

(in fact, since the birth rate is declining, this means the absolute number of college-goers is going down).

I’ll go further. Harvard keeps building more dorms and hiring more professors, so there are the same number of Harvard spots per American today as there were ten years ago, twenty years ago, and all the way back to the 1800s:



I want to look into this further and investigate questions like:

– How did we get to this point? Have college admissions always been a big deal? Did George Washington have to freak out about getting into a good college? What about FDR? If not, why not?

– Is academia really more competitive now than in the past? On what time scale? At what levels of academia? Why is this happening? Will it stop?

– Is freaking out about college admissions the correct course of action?

1. A Harvard-Centric History Of College Admissions

For the first two centuries of American academia, there was no competition to get into college. Harvard admitted…

(Harvard is by far the best-documented college throughout most of this period, so I'll be focusing on them. No, Ben Casselman, *you* shut up!)

…Harvard admitted anyone who was fluent in Latin and Greek. The 1642 Harvard admission requirements said:

> When any schollar is able to read Tully [Cicero] or such like classicall Latine Authore *ex tempore* & make and speake true Latin in verse and prose, *suo (ut auint) Marte*, and decline perfectly the paradigmes of Nounes and Verbes in the Greek tongue, then may hee bee admitted into the Colledge, nor shall any claim admission before such qualifications.

Latin fluency sounds impressive to modern ears, like the sort of test that would limit admission to only the classiest of aristocrat-scholars. But knowledge of classical languages in early Massachussetts was shockingly high, even among barely-literate farmers. In 1647, in between starving and fighting off Indian attacks, the state passed a law that every town of at least 100 families must have a school that taught Latin and Greek (it was called The Old Deluder Satan Law, because Puritans). Even rural families without access to these schools often taught classical languages to their own children. Mary Baker Eddy, who grew up in early 19th-century rural New Hampshire, wrote that:

> My father was taught to believe that my brain was too large for my body and so kept me much out of school, but I gained book-knowledge with far less labor than is usually requisite. At ten years of age I was as familiar with Lindley Murray's Grammar as with the Westminster Catechism; and the latter I had to repeat every Sunday. My favorite studies were natural philosophy, logic, and moral science. From my brother Albert I received lessons

> in the ancient tongues, Hebrew, Greek, and Latin. My brother studied Hebrew during his college vacations.

By the standards of the time, Harvard admission requirements were tough but fair, and well within the reach of even poorer families. More important, they were only there to make sure students were prepared for the coursework (which was in Latin). They weren't there to ration out a scarce supply of Harvard spots. In fact this post, summarizing Jerome Karabel's *Chosen*, says that "there was no class size limit, because Harvard was trying to compete with Oxford and Cambridge for size". They wanted as many students as they could get; their only limit was the number of qualified applicants.

These policies continued through the 19th century, with changes only in the specific subjects being tested. In the late 1700s they added some math; in the early 1800s they added some science. You can find a copy of the 1869 Harvard entrance exam here. It's pretty hard – but it had an 88% pass rate (surely at least in part because you wouldn't take it unless you were prepared) and everyone who passed was guaranteed a spot at Harvard. Some documents from Tufts around this time suggest their procedure was pretty similar. Some other documents suggest that if you went to a good high school, they assumed you were prepared and let you in without requiring the exams.

When did this happy situation end? Information on this topic is hard to find. I can't give specific sources, but I get the impression that at the very end of the 19th century, there was a movement to standardize college admissions. At first this just meant make sure every college has the same qualification exams, so that one school isn't asking about Latin and another about Greek. This culminated in the creation of the College Board in 1899, which administered an admission test that acted as a sort of great-great-grandfather of the SAT. Very gradually, so gradually that nobody at the time really remarked on it, this transitioned from making sure students were ready, to rationing out scarce spots. By about 1920, the transition was basically complete, so that nobody was surprised when people talked about “how colleges should decide who to accept” or questions like that. If you can find more on this transition, please contact me.

Acceptance was originally based entirely on your score on the qualifying exam. But by the 1920s, high-scorers on this exam were disproportionately Jewish. Although Jews were only about 2% of the US population, they were 21% of Harvard’s 1922 class (for more on why this might happen, read my post The Atomic Bomb Considered As Hungarian High School Science Fair Project). In order to arrest this trend, Harvard and other top colleges decided to switch from standardized testing to an easier-to-bias “holistics admissions” system that would let them implement a *de facto* Jewish quota.

Quota proponents not only denied being anti-Semitic but argued they were actually trying to *fight* anti-Semitism; if the student body became predominantly Jewish, this might inflame racial tensions against Jews. Harvard president Abbott Lowell, the quotas’ strongest proponent, said: “The anti-Semitic feeling among the students is increasing, and it grows in proportion to the increase in the number of Jews. If their number should become 40% of the student body, the race feeling would become intense”. Was he just trying to rationalize his anti-Semitism? I don’t think so. I doubt modern Harvard officials are anti-Asian in any kind of a hateful sense, but they enforce Asian quotas all the same. What would they say if you asked them why? Maybe that if a country full of whites, blacks, and Latinos had predominantly

Asian elite colleges, that might make (as Lowell put it) “the race feeling become intense”. I see no reason to think that 1920s officials were thinking any differently than their modern counterparts.

Whatever the reasons, by the mid-1920s the Jewish quota was in place and Harvard had switched to holistic admissions. But Lowell and his contemporaries emphasized that the new policies were never meant to make Harvard selective. “It is neither feasible nor desirable to raise the standards of the College so high that none but brilliant scholars can enter…the standards ought never to be so high for serious and ambitious students of average intelligence.”

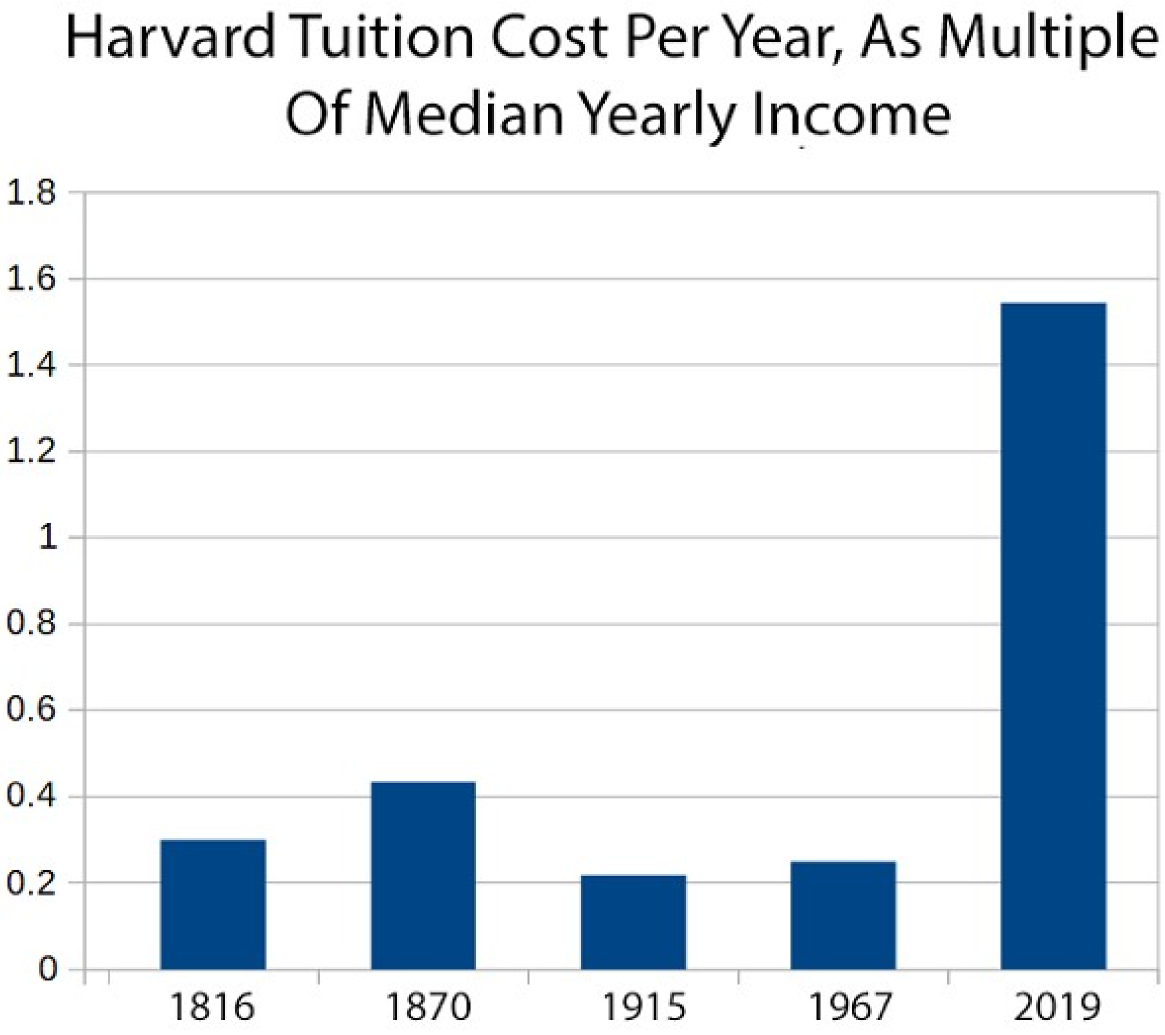
We’ll talk later about how this utopian dream of top-notch education for anyone with a foreskin failed. But before we get there, a more basic question: how come Harvard wasn’t overrun with applicants? If the academic requirements were within reach of most smart high-schoolers, how come there was no need to ration spots?

Below, I discuss a few possibilities in more depth.

1.1: Historical Tuition Fees

Were early American colleges so expensive that everyone except aristocrats was priced out?

No:



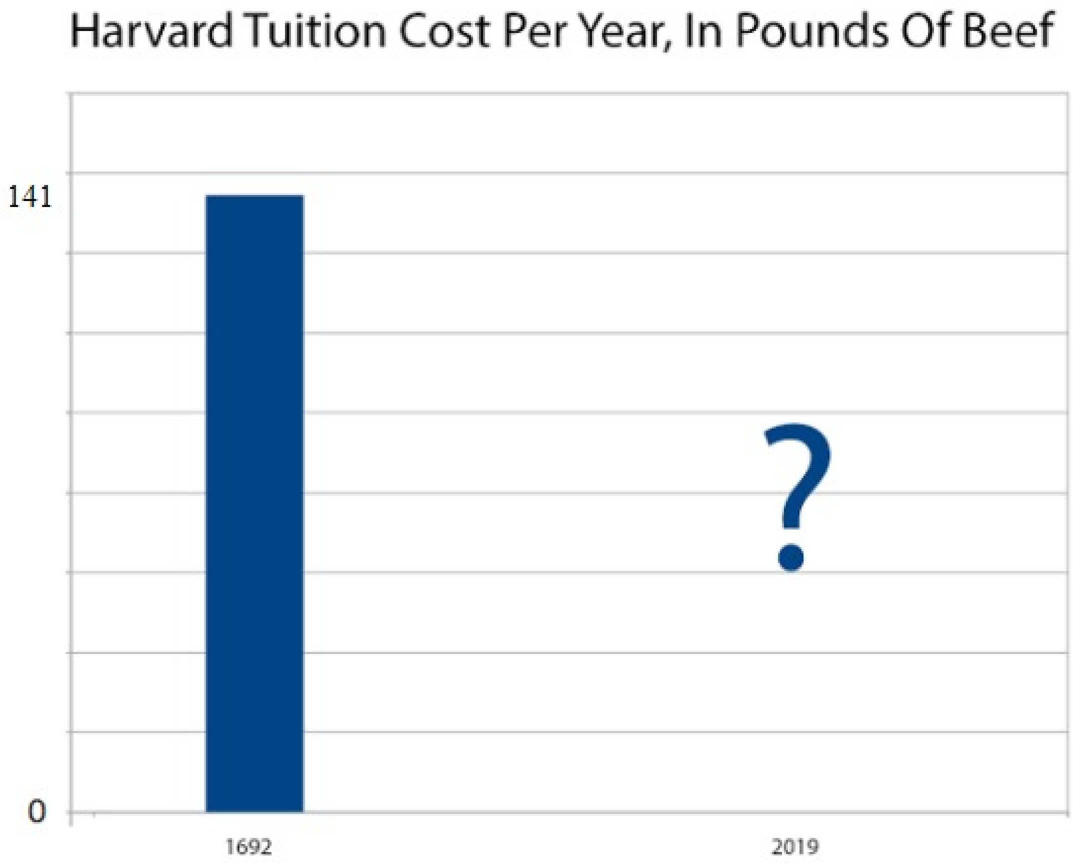
(*sources:* *1*, *2*, *3*, *4*, *5*, *6*)

I find very conflicting accounts of colonial tuition prices. But after the Revolution, tuition stayed stable about about a third average median income until about 1990, when it increased to 1.5x median income. In other words: relative to income, historical tuition costs were about a fifth of what they are today. Some good universities seem to have not had tuition at all – Stanford had a $0 price tag for its first 35 years.

Even when tuition existed, historical accounts suggest it wasn't especially burdensome for most college students, and record widespread effort to accommodate people who couldn’t pay.The first Harvard scholarship was granted in the 1640s. There are occasional scattered references to people showing up at Harvard without enough money to pay and being given jobs as servants to college officials or other students to help cover costs; in America, Ralph Waldo Emerson took advantage of this kind of program; in Britain, Isaac Newton did.

If you were a poor farmer who couldn’t get a scholarship and didn’t want to work as a servant, sometimes college were willing to accept alternative forms of payment. According to The Billfold:

> Harvard tuition — which ran about fifty-five pounds for the four-year course of study — was paid the same way [in barter], most commonly in wheat and malt. The occasional New England father sent his son to Cambridge with parsnips, butter, and, regrettably for all, goat mutton. A 141-pound side of beef covered a year’s tuition.



1.2 Discrimination

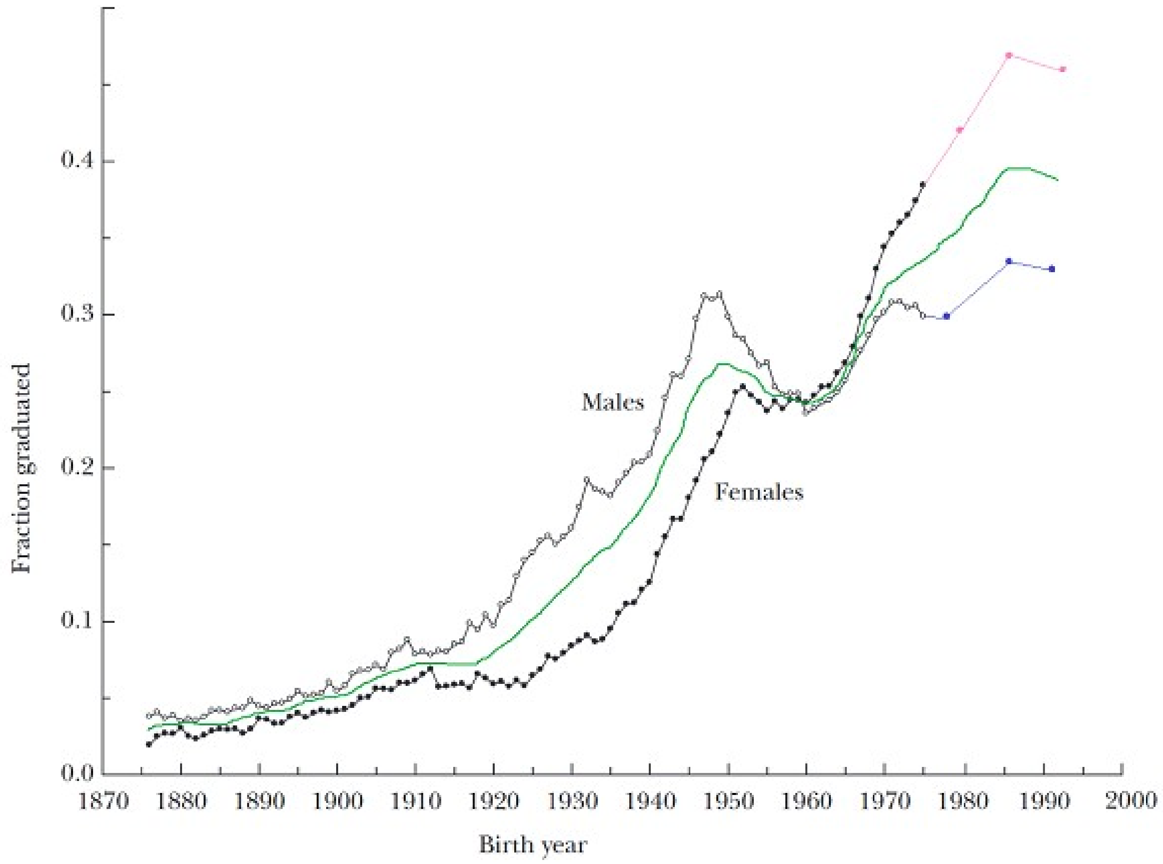
Early colleges only admitted white men. Did this reduce the size of the applicant pool enough to give spots to all white men who applied?

I don’t think racial discrimination can explain much of the effect. Throughout the 19th century, America hovered around 85% white. New England, where most Harvard applicants originated, may have been 95% to 99% white – see eg this site which says Boston was 1.3% black in 1865; non-black minorities were probably a rounding error. So there’s not much room for racial discrimination to reduce the applicant pool.

The exclusion of women from colleges in the 1800s is less than generally believed:

Figure 1

College Graduation Rates (by 35 years) for Men and Women: Cohorts Born fro 1876 to 1975



(source: unprincipled sketchy attempt to combine this with this to get one measure that covers the entire period)

For every woman in college in 1890, there were about 1.3 men; this is no larger a gender gap than exists today, though in the reverse direction. How come you never hear about this? Many of the women were probably in teacher-training colleges or some other gendered institution; until the early 1900s, none of them were at Harvard. But after gender integration, the women’s colleges were usually annexed to the nearest men’s college, turning them into a single institution. Under these circumstances, it doesn’t seem that likely that integration had a huge effect on admissions selectivity. Also, admitting women can only double the size of the applicant pool, but 1800s college seemed much more than twice as easy to get into.

Overall I don’t think this was a major part of the difference either.

1.3: Lack Of Degree Requirement For Professional Schools

Nowadays college is competitive partly because people expect it to be their ticket to a good job. But in the 19th century, there was little financial benefit to a college degree.

Suppose you wanted to become a doctor. Most medical schools accepted students straight out of secondary school, without a college degree. In fact, most medical schools accepted all “applicants”, the same as Harvard. Like Harvard, there was sometimes a test to make sure you knew Greek and Latin (the important things for doctors!) but after that, you were in.

(This article has some great stories about colonial and antebellum US medical education. Anyone who wanted could open up a medical school; profit-motive incentivized them to accept everybody. Medical-schooling was so profitable that the bottleneck became patients; since there were no regulations requiring medical students to see patients, less scrupulous schools tended to skip this part. Dissection was a big part of the curriculum, but there were no refrigerators, so fresh corpses became a hot commodity. Grave robbing was a real problem, sparking small-scale wars between medical schools and their local towns. “In at least 2 instances, the locals actually raided the school to obtain a body. In 1 case, the school building was destroyed by fire, and in another, 2 people, a student and a professor, were killed.” There were no requirements for how long medical schools should last, so some were as short as nine months. But there were also no requirements for who could call themselves a doctors, so students would sometimes stay until they got bored, then drop out and start practicing anyway. Tuition was about $100 per year, plus cost of living and various hidden fees; by my estimates that’s about half as much (as percent of an average doctor’s salary) as medical school tuition today. This situation continued until the Gilded Age, when medical schools started professionalizing themselves a little more.)

Or suppose you wanted to be a lawyer. The typical method was called “reading law”, which meant you read some law textbooks, served an apprenticeship with a practicing lawyer, and then started calling yourself a lawyer (in some states you also needed a letter from a court testifying to your “good moral character”). Honestly the part where you apprenticed with an practicing lawyer was more like a good idea than a requirement. It’s not completely clear to me that you needed to do anything other than read enough law textbooks to feel comfortable lawyering, and then go lawyer. Most lawyers did not have a college degree.

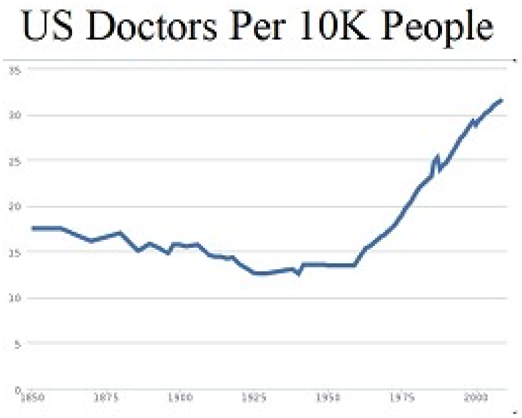
Abraham Lincoln, a lawyer himself, advised a law student:

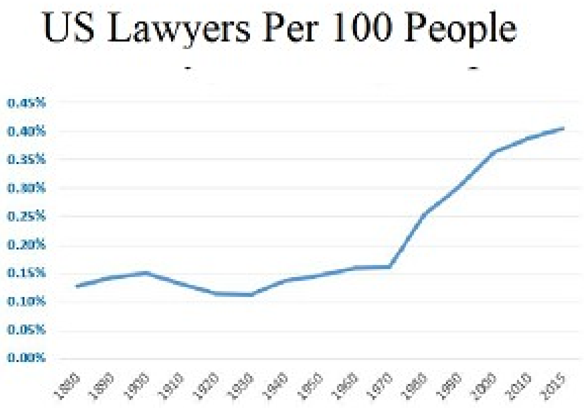
> If you are absolutely determined to make a lawyer of yourself the thing is more than half done already. It is a small matter whether you read with any one or not. I did not read with any one. Get the books and read and study them in their every feature, and that is the main thing. It is no consequence to be in a large town while you are reading. I read at New Salem, which never had three hundred people in it. The books and your capacity for understanding them are just the same in all places.

Levi Woodbury, the 30th US Supreme Court Justice (appointed 1846), was the first to attend any kind of formal law school. James Byrnes, the 81st Supreme Court Justice (appointed 1941), was the last not to attend law school. It’s apparently still technically possible in four states (including

California) to become a lawyer by reading law, but it’s rare and not very encouraged.

The ease of entering these professions helps explain why there was no oversupply of Harvard applicants. But then why wasn’t there an oversupply of doctors and lawyers? We tend to imagine that *of course* you need strict medical school admissions, because some kind of unspecified catastrophe would happen if any qualified person who wanted could become a doctor. Did these open-door policies create a glut of professionals?





No. There were fewer doctors and lawyers per capita than there are now.

Did it drive down salaries for these professions?

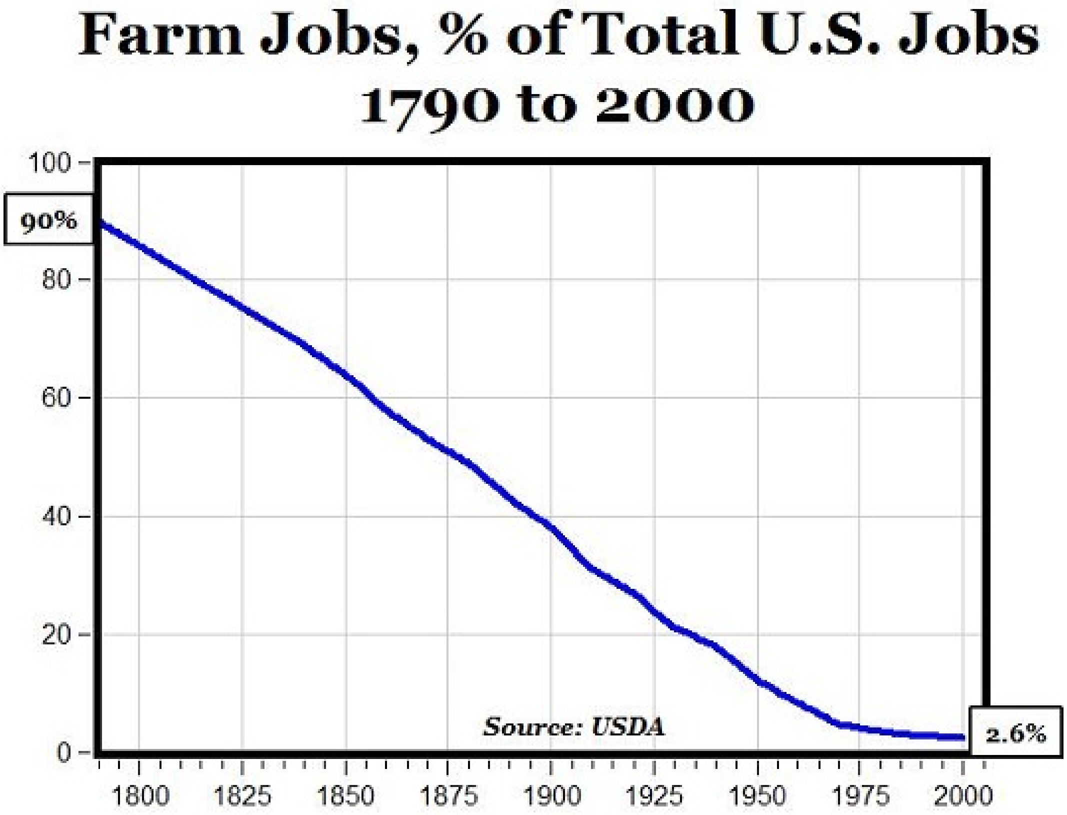
I don’t have great numbers on lawyer salaries, but based on this chart from 1797 Britain and this chart from 1900s America, I get the impression that throughout this period lawyers made about 3-5x as much as unskilled laborers, 3-4x as much as clerks and teachers, and about the same as doctors. This seems to match successful modern lawyers, and probably exceed average modern lawyers. This may because unskilled laborers now earn a minimum wage and teachers have unions, but in any case the 19th-century premium to a law degree seems to have been at least as high and probably higher.

The same seems true of doctor salaries. The paper above estimates physician salaries at $600 per year, during a time when agricultural laborers might have been making $100 and clerks and teachers twice that.

I conclude that letting any qualified person become a doctor or a lawyer, without gatekeeping, did not result in a glut of doctors and lawyers, and did not drive down salaries for those professions beyond levels we would find reasonable today.

1.4: Conclusions

So why *weren't* there gluts of would-be college students, doctors, and lawyers? I can't find any single smoking gun, but here are some possibilities.



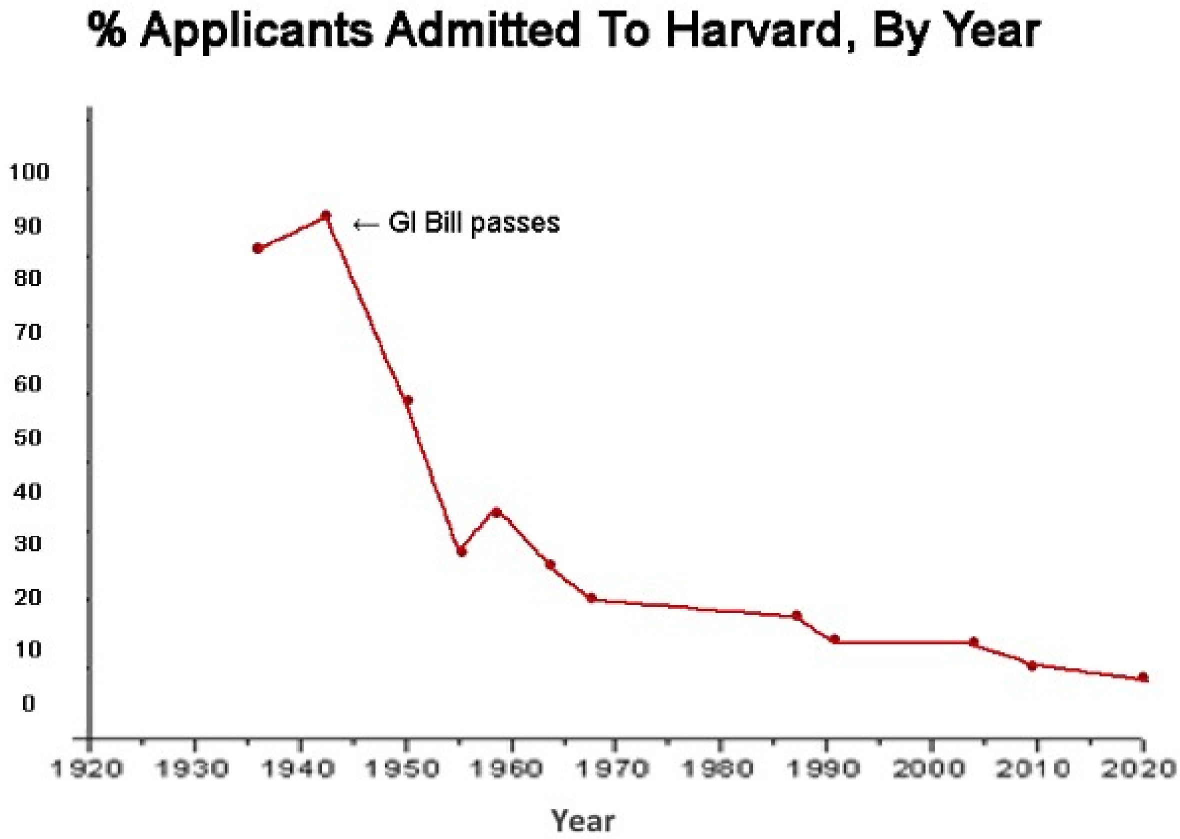
Throughout this period, between 60% and 80% of Americans were farmers. Unless you were wealthy or urban, the question of "what career do you want in order to actualize your potential" didn't come up. You were either going to be a farmer, or else you had some specific non-farm pathway in mind that you could pursue directly instead of getting a college degree to "keep your options open".

Since rural children were expected to work on the farm, there was no protracted period of educational unproductivity. There was no assumption that your kids weren't going to be earning anything until age 18 and so you might as well protract their unproductivity until age 22. That meant that paying to send your child to Boston or wherever, and to support him in a big-city lifestyle for four years, was actually a much bigger deal than the tuition itself. This article claims that in 1816, tuition itself was only about 10% of the expenses involved in sending a child to college (granted, poor people pinching pennies could get by for much less than the hypothetical well-off student analyzed here, but I think the principle still holds).

Another limiting factor may have been that there was ample opportunity outside of college and the professions, in almost every area. Twelve US presidents, including George Washington, did not go to college. Benjamin Franklin, everyone’s model of an early American polymath genius, did not go to college. Of the ten richest people in American history (mostly 19th-century industrialists), as far as I can tell only two of them went to college. Aside from the obvious race and gender discrimination, the 19th century was a lot closer to real meritocracy than today’s credentialist fake meritocracy; people responded rationally by ignoring credentials and doing meritorious things.

2. How Did The Zero-Competition Regime Transition To The Clusterf**k We Have Today?

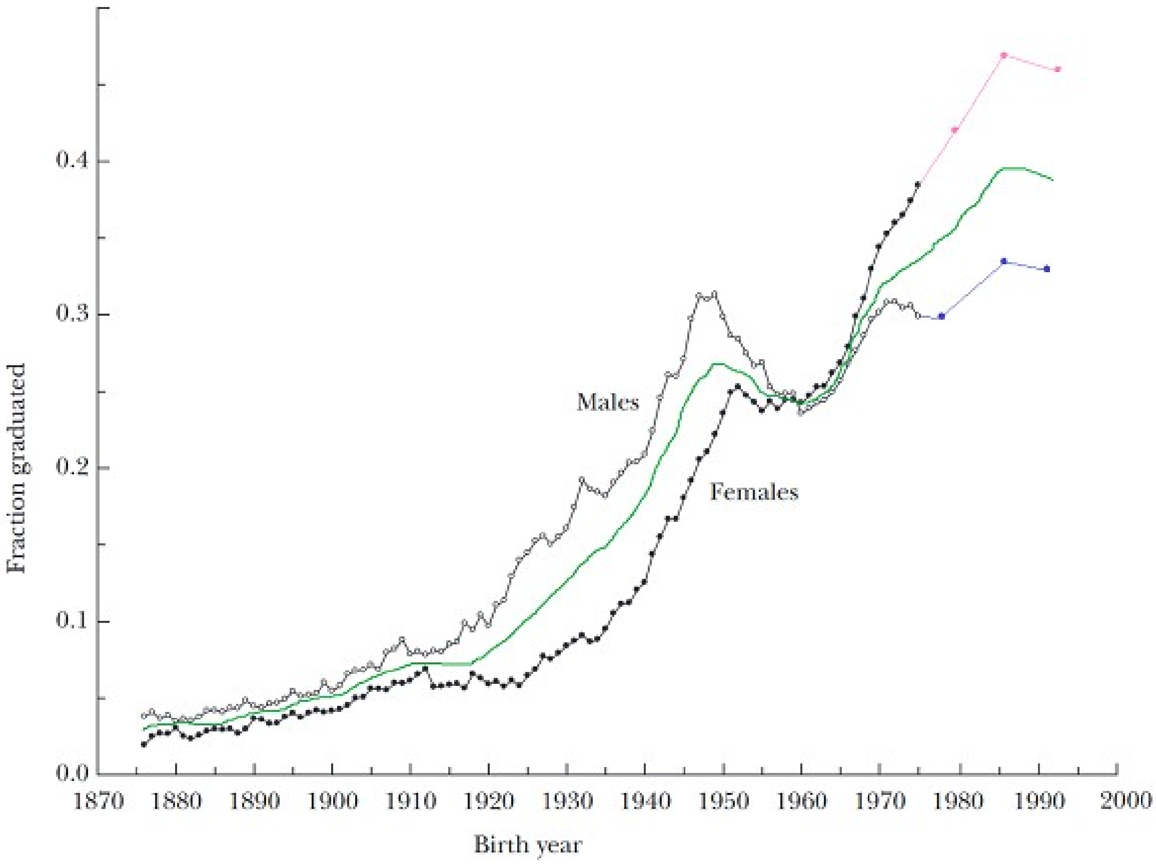
Here is a graph of Harvard admission rates over time, based mostly on these data:



During the early part of the 1900s, Harvard was still in the 19th-century equilibrium of admitting most qualified non-Jewish applicants. Around 1940, the admission rate dropped from 95% to 25%. Most sources I read attribute this to the GI Bill, a well-intentioned piece of legislation that encouraged returning WWII veterans to get a college education. So many vets took the government up on the offer that Harvard was overwhelmed for the first time in its history.

But this isn’t the whole story.

Figure 1
College Graduation Rates (by 35 years) for Men and Women: Cohorts Born fro 1876 to 1975



You’ve seen this before – this is percent of Americans (by gender) to graduate college. It’s sorted by birth cohort, which means 1920 on the x-axis corresponds to the people who were in college in the 1940s – eg our GIs. The GI Bill is visible on this graph – around 1920, there is a spike in attendance for men but not women, which is the pattern we would predict from GIs. But it only takes college graduation rate from 10% to 15% (compared to its current 40%). And after the GI Bill, the college graduation rate starts dropping again – as we would expect of a one-time shock from a one-time war. And between 1955 and 1960, Harvard admissions rebound to about 40% of applicants.

The big spike in college attendance rates – and a corresponding dip in Harvard admission percentage – takes place in the 1938 to 1952 birth cohort. Why are all these people suddenly going to college? *They're dodging the draft*. A big part of the increase in college admissions was people taking advantage of the college loophole to escape getting sent to Vietnam.

Again, this is a one-time shock, and mostly applies to men. So how come we see a quadrupling of college graduation during this period affecting men and women alike?

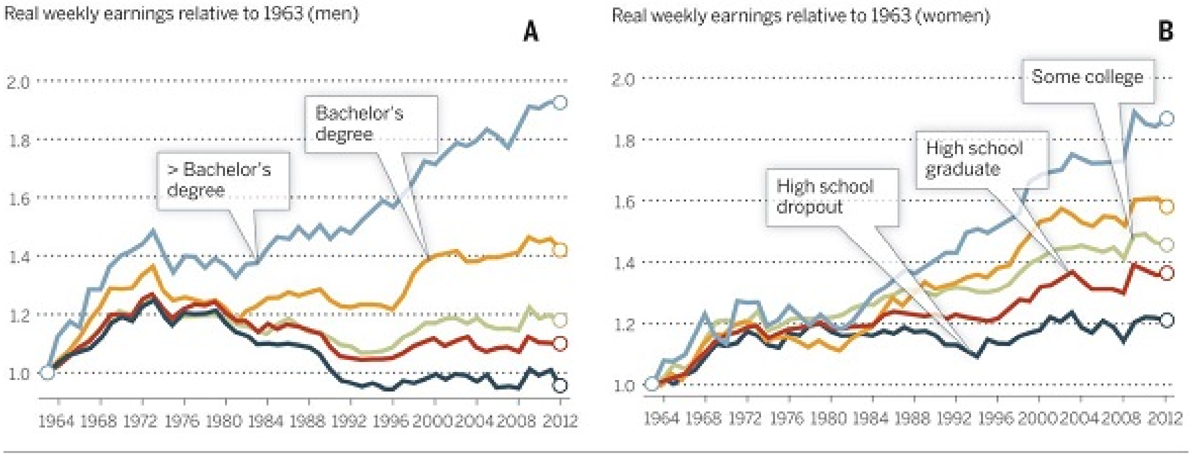
A standard narrative says that work has gotten more difficult over the past century, and so workers need more education. I've always found this hard to believe. In other countries, students still go to medical school and law school without a separate college degree first. Programming is a classic example of a high-skilled complicated modern profession, but many programmers dropped out of college, many others didn't attend at all, and many programming "boot camps" are opening up offering to teach programming skills outside the context of a college education. And in many of the jobs that do require college education, the education is irrelevant to their work. Both of my adult jobs – as an English teacher and as a doctor – required me to have a college degree in order to apply. But my college education was relevant to neither (I'm a philosophy major). The degree requirement seemed like more of a class barrier / signaling mechanism than an assertion that only people who knew philosophy could make good teachers and doctors. I realize I'm making a strong claim here, and I don't have space to justify it fully – for more on this, read my Against Tulip Subsidies and SSC Gives A Graduation Speech – or better yet, Bryan Caplan's *The Case Against Education*.

If increasing need for skills didn’t cause increasing college attendance, what did? Again, this is based off of idiosyncratic beliefs I don’t have the space to justify (again, read Caplan) but it could be a sort of self-reinforcing signaling cycle. Once the number of people in college reached a certain level, it led to a well-known social expectation that intelligent and conscientious men would have college degrees, which made college a sign of intelligence and, conversely, not having been to college a sign of stupidity. If only 10% of smart/hard-working people have been to college, not having a college degree doesn’t mean someone isn’t smart/hard-working; if 90% of smart/hard-working people have been to college, not having a college degree might call their intelligence and work ethic into question. This cycle meant that after the shocks of the mid-1900s, there was a strong expectation of a degree in the knowledge professions, which forced women and later generations of men to continue going to college to keep up. The government’s decision to provide an endless stream of supposedly-free college loans exacerbated the problem and sabotaged the only natural roadblock that could have stopped it.

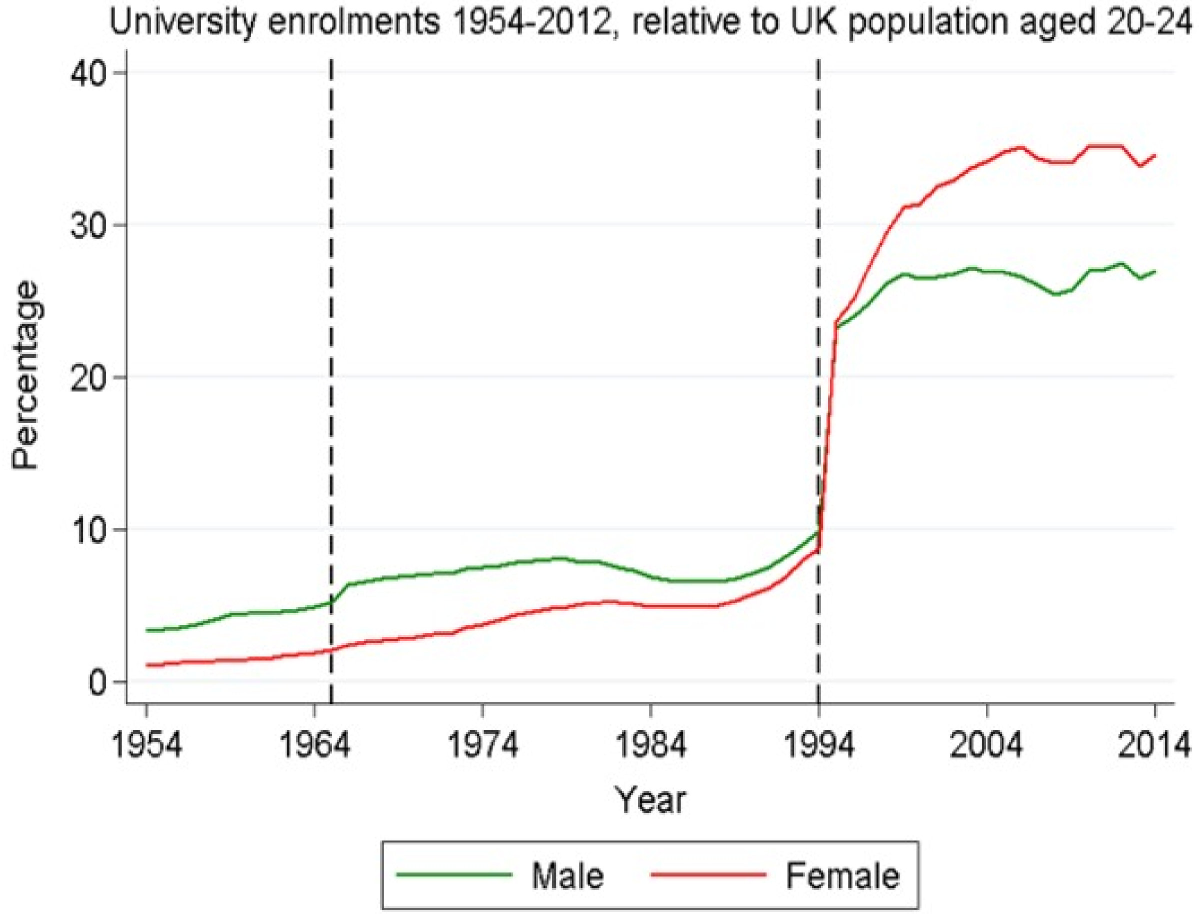
At the same time, several factors were coming together to discourage hunch-based “I like the cut of his jib” style hiring practices. Community ties were becoming weaker, so hirers typically wouldn’t have social contacts with potential hirees. Family businesses whose owners could hire based on hunches were giving way to large corporations where interviewers would have to justify their hiring decisions to higher-ups. Increasing concern about racism was raising awareness that hunch-based hiring tended to discriminate against minorities, and the advent of the discrimination lawsuit encouraged hiring based on objective criteria so you could prove you rejected someone for reasons other than race. The Supreme Court decision Griggs v. Duke may or may not have played a role by making it legally risky for corporations to give prospective hires aptitude tests. All of this created a “perfect storm” where employers needed some kind of objective criteria to evaluate potential new hires, and all the old criteria weren’t cutting it anymore. The rise of the college degree as a signal for intelligence, and the increased sorting of people by college selectivity, fit into this space perfectly.

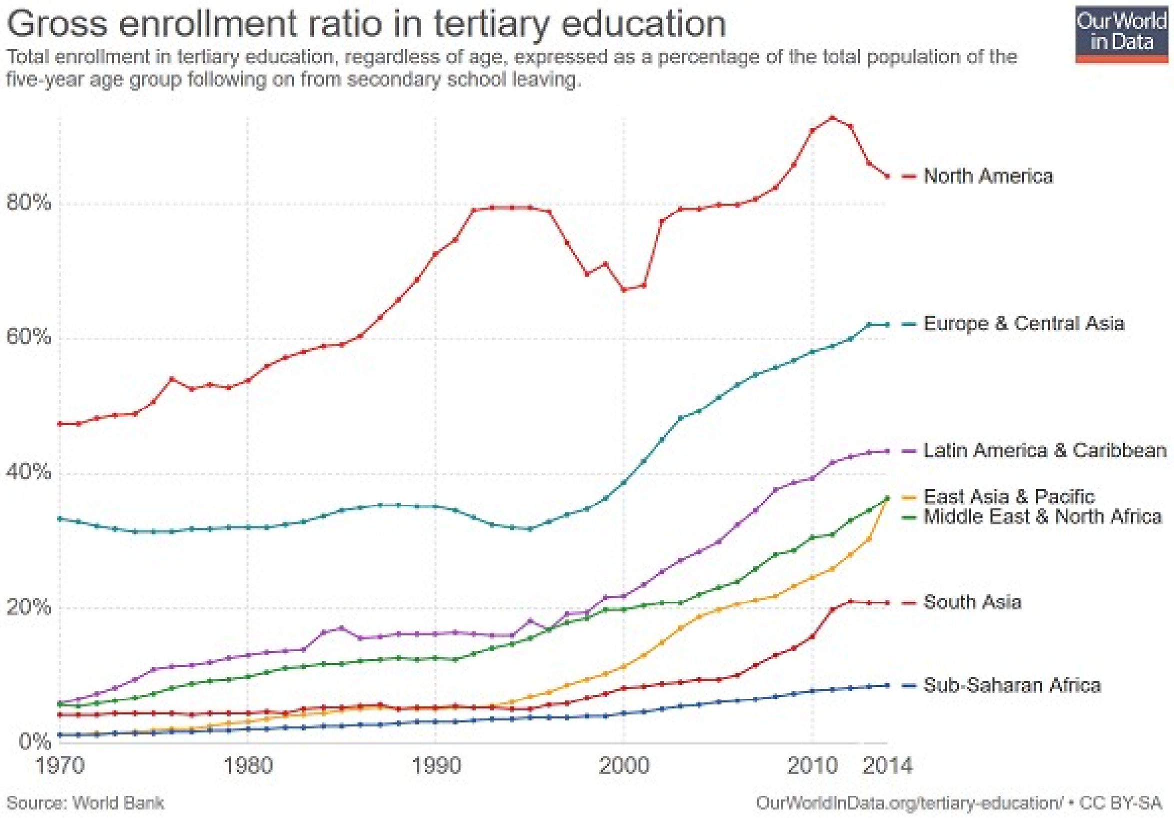
Once society established that knowledge-worker jobs needed college degrees, the simultaneous rises in automation, globalization, and inequality made knowledge-worker jobs increasingly necessary to earn a living, completed the process.

Changes in real wage levels of full-time U.S. workers by sex and education, 1963–2012



If my story were true, this would suggest college attendance would not have risen so quickly in other countries that didn’t have these specific factors. I don’t have great cross-country data, but here’s what I can find:


University enrolments 1954-2012, relative to UK population aged 20-24
Percentage
40
30
20
10
0
1954
1964
1974
1984
1994
2004
2014
Year
Male
Female



College attendance in the UK supposedly remained very low until a 1992 act designed to encourage it, but it looks like part of that is just them reclassifying some other schools as colleges. I don’t know how it *really* compared to the US and I welcome information from British readers who know more than I do about this. Through the rest of the world, college attendance lagged North America by a long time, but the continent-wide categories probably combine countries at different levels of economic development. I don’t really know about this one.

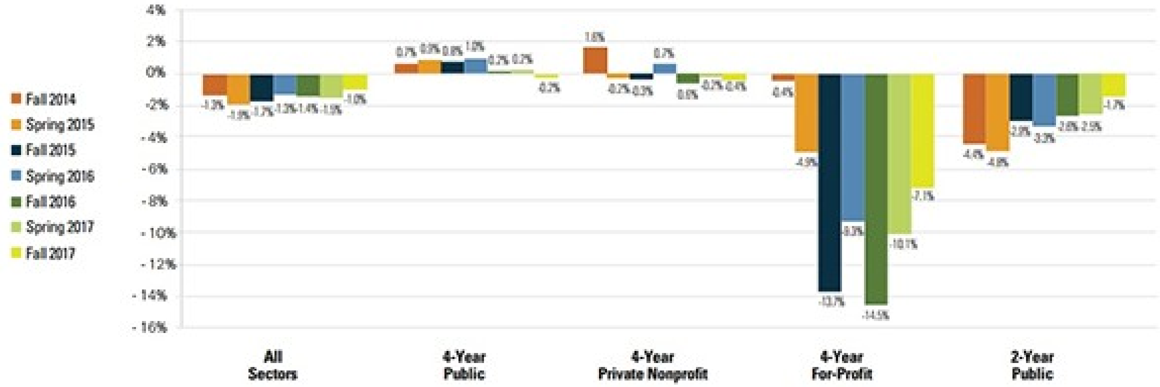
Moving on: the graphs in the Introduction show that college attendance has been stable since about 2005. Why did the rise stop? These articles point out a few relevant trends.

First, the economy is usually to blame for this kind of thing. There was a slight increase in attendance during the 2008 recession, and a slight decrease during the recent boom. But over the course of the cycle, it still seems like the increase in college attendance has slowed or stopped overall, in a way that wasn't true of past business cycles.

Second, birth rates are decreasing, which means fewer college-aged kids. The national population is still increasing, mostly because of immigrants, but many immigrants are adults without much past education, so they're not as significant a contribution to the college population.

Third, the price of college keeps going up. I'm surprised to hear this as a contribution to declining attendance, because I thought it was the glut of students that kept prices high, but maybe both factors affect each other.

Fourth, for-profit colleges are falling apart.



In some cases, the government has shut them down for being outright scams. In other cases, potential students have wised up, realized they are outright scams, and stopped being interested in attending them. These colleges advertised to (some would say "preyed on") people who weren't able to get into other colleges, so their collapse looks like a fall in the college enrollment/graduation rate.

These are all potentially relevant, but they seem kind of weak to me: the sort of thing that explains the year-to-year trend, but not why the great secular movement in favor of more college has stopped.

Maybe it’s just reached a natural ceiling. Seventy percent of high school graduates are now going to college. The remaining 30% may disproportionately include people with serious socioeconomic or health problems that make going to college very hard for them.

Also, keep in mind that only about 60% of college students graduate in anywhere near the expected amount of time. Some economists have come up with rational-college-avoidance models where people who don’t expect to be able to graduate from college don’t waste their money trying.

3. If Number Of Students Applying To College Has Been Constant Or Declining Over The Past Ten Years, Why Are Admissions To Top Colleges So Much More Competitive?

To review: over the past ten years, the number of US students applying to college has gone down (the number applying to four-year private colleges has stayed about the same). But Harvard’s acceptance rates have decreased by half, with similar cuts across other top schools, and more modest cuts across most good and moderately-good colleges. There’s also a perception of much greater pressure on students to have perfect academic records before applying. Why?

3.1: Could the issue be increasing number of international students?

This would neatly match the evidence of constant US numbers vs. increasing selectivity.

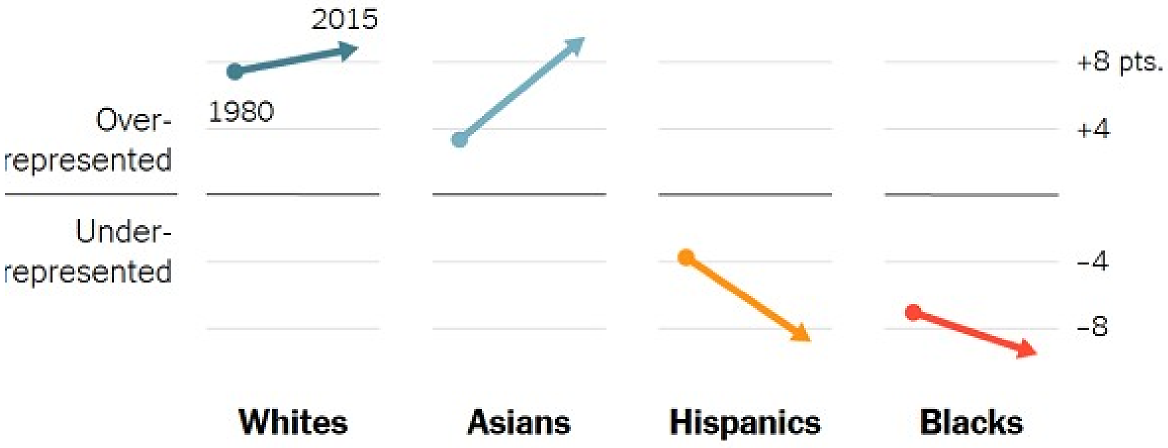
Harvard equivocates between a few different definitions of “international student”, but I think it’s comparing apples to apples when it says the Class of 2013 was 10% foreign citizens and the Class of 2022 is 12%. These two classes bound the time period we’re worrying about, and this doesn’t seem like a big change. Also, across all US colleges international student enrollments seem to be dropping, not increasing. Some of this may have to do with strict Trump administration visa policies, or with international perceptions of increasing US hostility to foreigners.

Since fewer international students are applying in general, and even top schools show only a trivial increase, this probably isn’t it.

3.2: Could the issue be more race-conscious admission policies?

Might top colleges be intensifying affirmative action and their preference for minorities and the poor, thus making things harder for the sort of upper-class white people who write news articles about the state of college admissions? Conversely, might colleges by relaxing their restrictions on high-achieving Asians, with the same result?

This matches the rhetoric colleges have been putting out lately, but there is not a lot of signs it’s really happening. Harvard obsessively chronicles the race of its student body, and the class of 2010 and class of 2022 have the same racial composition. The *New York Times* finds that whites are actually better represented at colleges (compared to their percent of the US population) than they were 35 years ago, although Asians are the real winners.



How much more or less each group is represented among freshmen at top colleges relative to the U.S. population

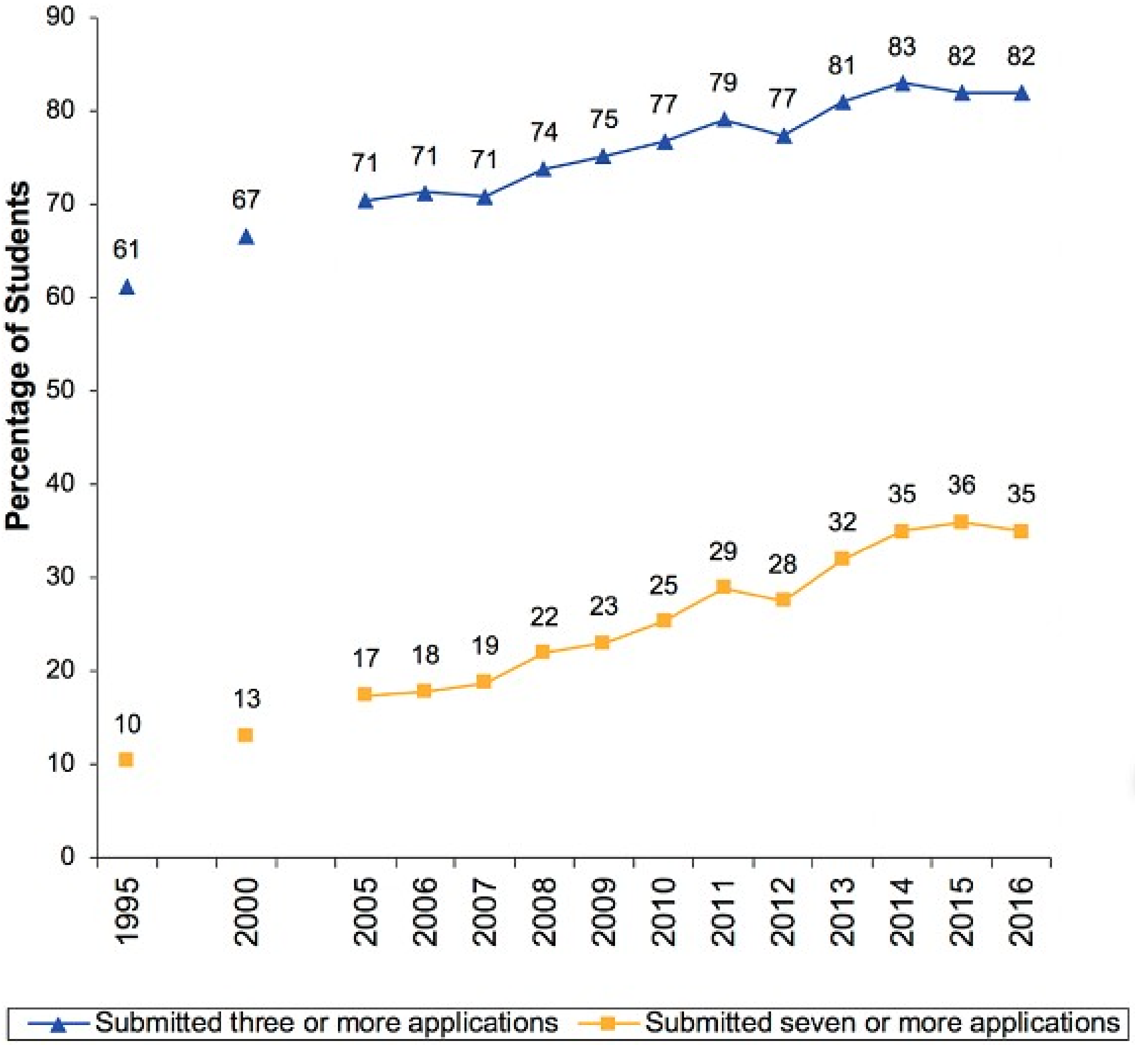
The *Times* doesn’t explain why this is happening. It may be due to weakening affirmative action, including bans by several states. Or it may be because of a large influx of uneducated Mexican immigrants who will need a few more generations of assimilation before their families attend college at the same rate as whites or previous generations of Latinos.

What about Asians? There was a large increase in Asian admissions, but it was mostly before this period. The Ivy League probably has some kind of unofficial Asian quota which has been pretty stable over the past decade. Although the Asian population continues to grow, and their academic achievement continues to increase, this probably just increases intra-Asian competition rather than affecting people of other races.

3.3: Could the issue be increasing number of applications per student?

Here’s an interesting fact – even though no more Americans or foreigners are applying to colleges today vs. ten years ago, Harvard is receiving twice as many applications – from about 20,000 to more than 40,000. How can this be?

FIGURE 1. INCREASES IN FIRST-TIME FRESHMEN APPLICATION SUBMISSION: 1995 TO 2016



SOURCE: Eagan, K., Stolzenberg, E.B., Zimmerman, H.B., Aragon, M.C., Sayson, H.W., & Rios-Aguilar, Rios. (2017). The American Freshman: National Norms Fall 2016. Los Angeles: Higher Education Research Institute at UCLA.

The average college student is sending out many more applications.

I am not Harvard material. But when I was looking at colleges, my mother pressured me to apply to Harvard. “Come on!” she said. “It will just take a few hours! And who knows? They might accept you! You’ll never get in if you don’t try!”

Harvard did not accept me. But my mother’s strategy is growing in popularity. Part of this might be genuine egalitarianism. Maybe something has gone very right, and the average American really does believe he or she has a shot at the Ivy League. But part of it may also be a cynical ploy by colleges to improve their rankings in *US News* and other similar college guides. These rankings are partly based on how “selective” they are, ie what percent of students they turn away. If they encourage unqualified candidates to apply, they can turn those unqualified candidates away, and then they appear more “selective” and their ranking goes up.

But increased application volume is mostly driven by an increasingly streamlined college admissions process, including the Common Application. I didn’t like my mother’s advice, because every college application I sent in required filling in new forms, telling them my whole life story all over again, and organizing all of it into another manila envelope with enclosed check. It was like paying taxes, except with essay questions. And there was a good chance you’d have to do it all over again for each institution you wanted to apply for. Now that’s all gone. 800 schools accept the Common Application, including the whole Ivy League. From the *Times* again:

> Six college applications once seemed like a lot. Submitting eight was a mark of great ambition. For a growing number of increasingly anxious high school

> seniors, figures like that now sound like just a starting point…
>
> For members of the class of 2015 who are looking at more competitive colleges, their overtaxed counselors say, 10 applications is now commonplace; 20 is taking on a familiar ring; even 30 is not beyond imagining. And why stop there? Brandon Kosatka, director of student services at the Thomas Jefferson School for Science and Technology in Alexandria, Va., recently worked with a student who wanted a spot in a music conservatory program. To find it, she applied to 56 colleges. A spokeswoman for Naviance, an online tool that many high school students and their counselors use to keep track of applications, said one current user's "colleges I'm applying to" tab already included 60 institutions. Last year the record was 86, she said.

Does this mean increasing competitiveness is entirely an illusion? Suppose in the old days, each top student would apply to either Harvard *or* Yale. Now each top student applies to both Harvard *and* Yale, meaning that both colleges get twice as many applicants. Since each of them can only admit the same number of students, it looks like their application rate has been cut in half. But neither one has really become more competitive!

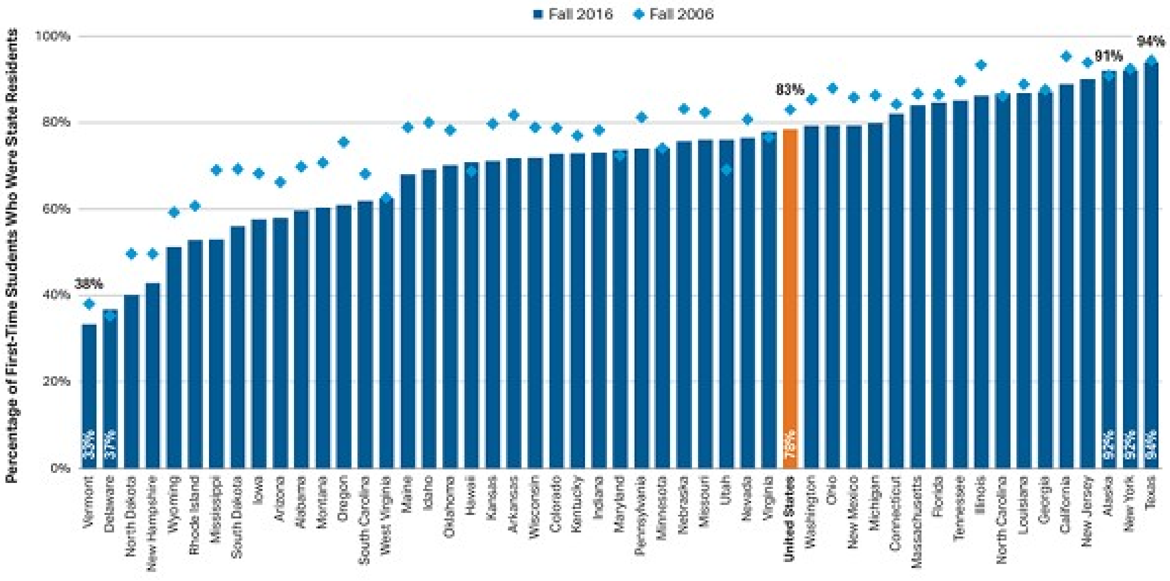
This can't *quite* be it. After all, in the first case, Yale would expect 100% of accepted students to attend. In the second, Yale would know that about 50% of accepted students would choose Harvard instead, so it would have to accept twice as many students, and the acceptance rate per application wouldn't change.

But if more people are following my mother's strategy of applying to Harvard "just in case" even when you're not Harvard material, then this *could* be an important factor. If the number of people who aren't Harvard material but have mothers who imagine they are is twice as high as the number of people who are really Harvard material, then Harvard admissions will triple. If Harvard accepts these people, they will definitely go to Harvard, so there is no need for Harvard to increase its admission rate to compensate. Here there really is an illusion of increasing competition.

Finally, this process could increase sorting. Suppose that, for the first time in history, a Jewish mother had an accurate assessment of her son's intellectual abilities, I really *was* Harvard material, and I was unfairly selling myself short. If the existence of a Common Application lets more people apply to Harvard "just in case", and if the Harvard admissions committee is good at their job, then the best students will get more efficiently matched with the best institutions. In the past, Harvard might have been losing a lot of qualified applicants to unjustified pessimism; now all those people will apply and the competition will heat up.

And in the past, I think a lot of people, including really smart people, just went to the nearest halfway-decent state college to their house. Partly this was out of humility. Partly it was because people cared about family and community more. And partly it was because college wasn't viewed as the be-all and end-all of your value as a human being and you had to get into the Ivy League or else your life was over. If all these people are now trying to get into Harvard, that will increase competition too.

Can we measure this?



This is the best I can do. It shows that over the past ten years, the number of students at public universities who come from in-state has dropped by 5%. This is probably related to sorting – people working on sorting themselves efficiently will go to the best school they can get into rather than just the closest one in their state. But it’s not a very dramatic difference. I suspect, though I can’t prove, that this is hiding a larger change at the very top of the distribution.

3.4: Could the issue be that students are just trying harder?

Imagine the exact same students applying to the exact same schools. But in 2009, they take it easy and start studying for their SATs the night before, and in 2019, they all have private tutors and are doing five extracurricular activities. College admissions will seem more competitive in 2019.

Any attempt to measure this will be confounded by reverse causation – increased effort might or might not cause increased selectivity, but increased selectivity will definitely cause increased effort. I’m not sure how to deal with this.

If studying harder improves SAT scores, these could be a proxy for how much effort students putting in. They changed the test in 2016 in a way that makes scores hard to compare, but we can at least compare scores from earlier years. Scores decline between 2005 and 2015 in both math and reading. This may be because more students are taking the SAT (1.5 million in 2008 vs. 2.1 million in 2018) so test-takers are a less selected population. This is kind of surprising given that college enrollment is stable or declining, but it could be that as part of pro-equality measures, schools are pressuring more low-achieving kids to take the SATs in order to “have a chance at college”, but those students don’t really end up attending. In support of this theory, scores are declining most quickly among blacks, Hispanics, and other poorer minority groups who may not have taken the SAT in earlier years; they are stable among whites, and increasing among Asians (increasing numbers of whom may be high-achieving Chinese immigrants). At least, this is the best guess I can come up with for why this pattern is happening. But it means SATs are useless as a measure of whether students are “trying harder”.

Why might students be trying harder? If there’s a ten year lag between things happening and common knowledge that the things have happened, the explosion of college attendance during the 1990s, with an ensuing increase in competitiveness, might have finally percolated down to the average student in the form of advice that getting into college is very hard and they should work to be more competitive. In addition, the Internet is exposing new generations of neurotic parents to messages that unless their child is perfect they will never get into college and probably die alone in a ditch.

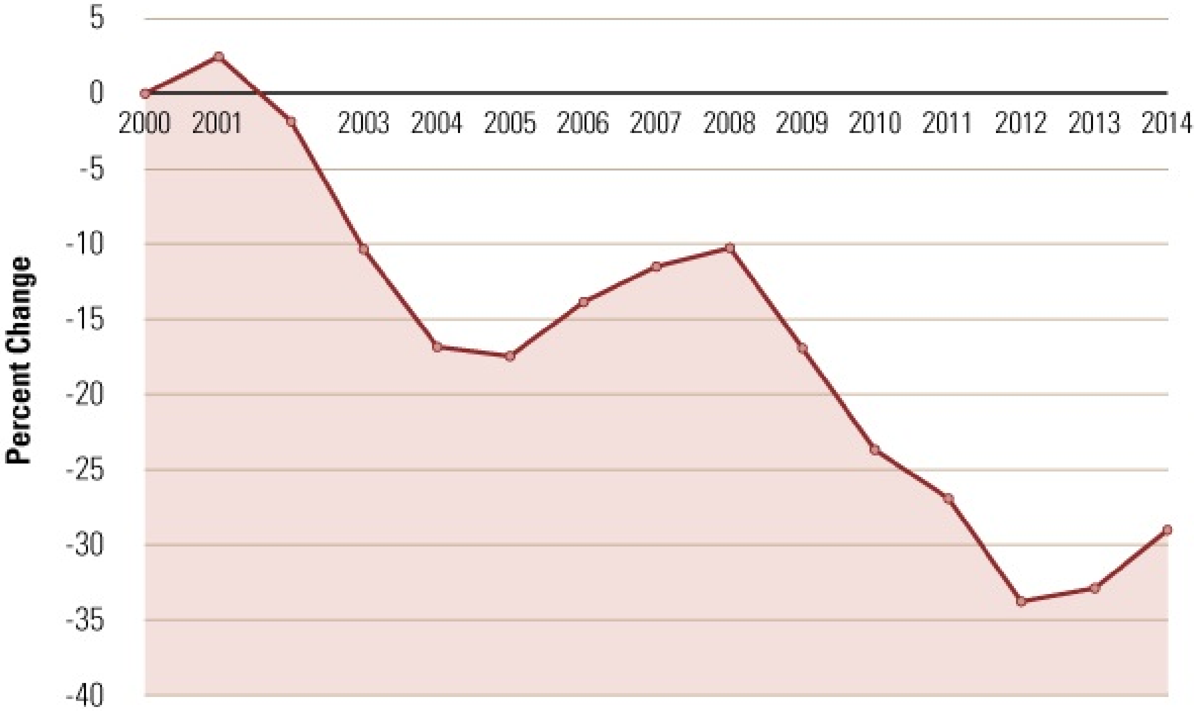
Further, the decline of traditional criteria might be causing an increasing emphasis on extracurriculars, which take a harder toll on college students. Because of grade inflation, colleges are no longer counting high school grades as much as they used to; because meritocracy is passé, they're no longer paying as much attention to the SAT. This implies increased emphasis on extracurriculars – things like student government, clubs, internships, charitable work, and the like. Despite popular misconceptions, the SAT is basically an IQ test, and doesn't really reward obsessive freaking out and throwing money at the problem. But getting the right set of extracurriculars *absolutely* rewards obsessively freaking out and throwing money at the problem. Maybe twenty years ago, you just played the IQ lottery and hoped for the best, whereas now you work yourself ragged trying to become Vice-President of the Junior Strivers Club.

But all of this is just speculation; I really don't know how to get good data on these subjects.

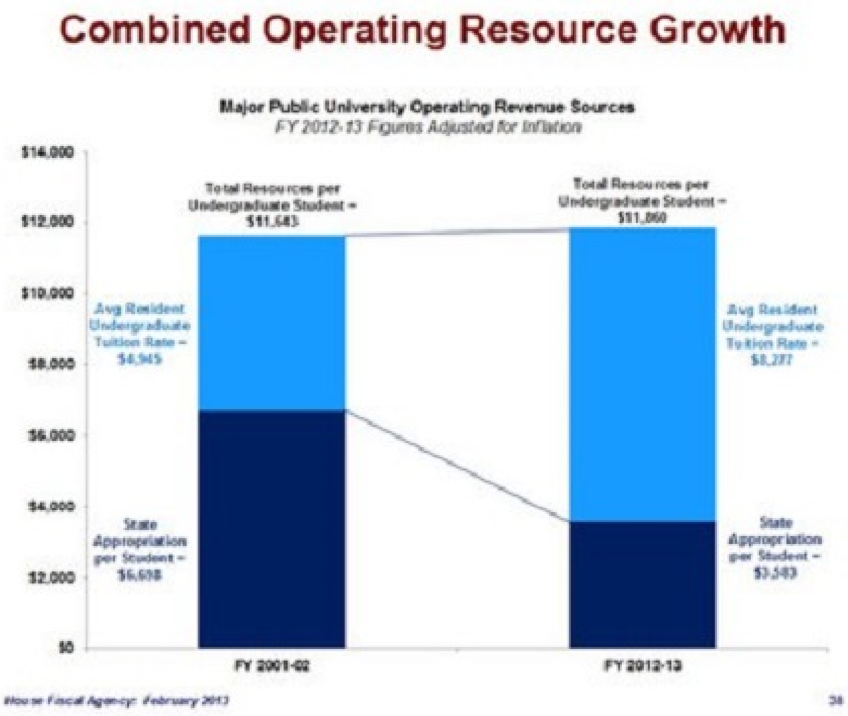
3.5: Are funding cuts reducing the number of college spots available?

Some people argue that cuts in public education are reducing the number of positions available at public universities, meaning the same number of students are competing for fewer spots. This source confirms large cuts in public funding:

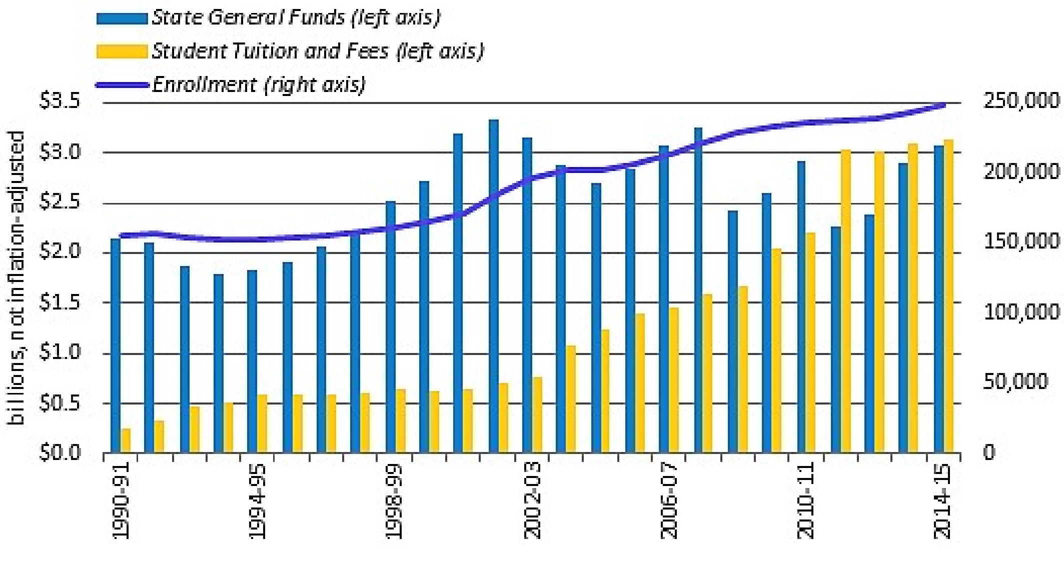
Figure 1: **Percent Change in State Support for Public Higher Education (All Colleges and Universities) per Full-Time Equivalent Student, in Constant 2014 $, since 2000**



These universities have tried to compensate by increasing tuition (or increasing the percent out-of-state students, who pay higher tuition). It looks like they’ve done this on a pretty much one-to-one basis, so that they’re actually getting more money per student now than they did when public funding was higher.



And from California:



It’s not clear that declining state support affected enrollment at all. Colleges just raised their prices by a lot.

In 2007, 2.8x as many students were in public universities compared to private ones. In 2017, the ratio was 2.9. If the problem were limited availability of public universities to absorb students, we might expect the percent of students at public universities to go down. This doesn't seem to be happening.

Overall it doesn't look like funding cuts to public universities mattered very much here.

3.6: Conclusions?

The clearest reason for increasing academic competition in the past ten years is the increasing number of applications per person, enabled by the online Common Application. This has doubled the number of applications sent to top colleges like Harvard despite the applicant pool staying the same size. Some of this apparent increased competition is a statistical illusion, but parts of it may be real due to increased sorting.

Other reasons may include increased common knowledge of intense competition making everyone compete more intensely, and decreased use of hard-to-game metrics like the SAT in favor of easy-to-game metrics like extracurriculars.

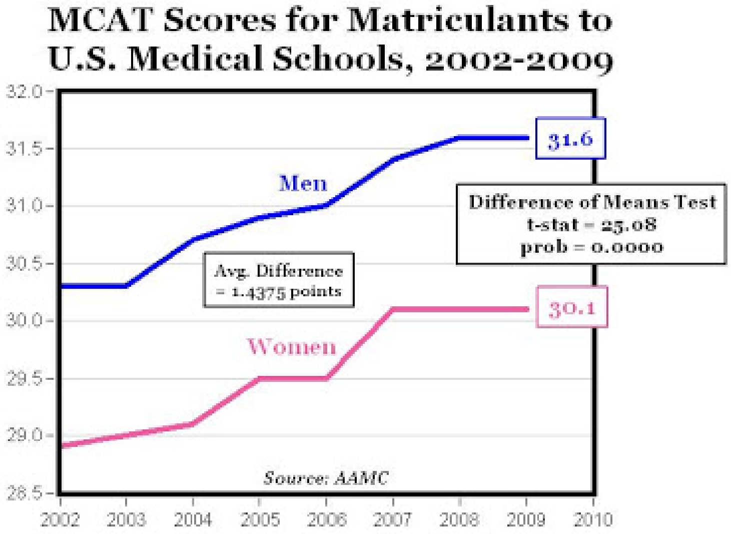
4. What Has Been Happening Beyond The College Level?

Competition is intensifying.

Between 2006 and 2016, the number of applicants to US medical schools increased by 35% (note change in number of applicants, *not* number of applications).

In a different statistic covering different years, the number of people enrolled at medical school increased 28% from 2002 to 2017. These two numbers aren't directly comparable, but by eyeballing them we get the impression that the number of spots is increasing more slowly than the number of applicants, probably much more slowly.

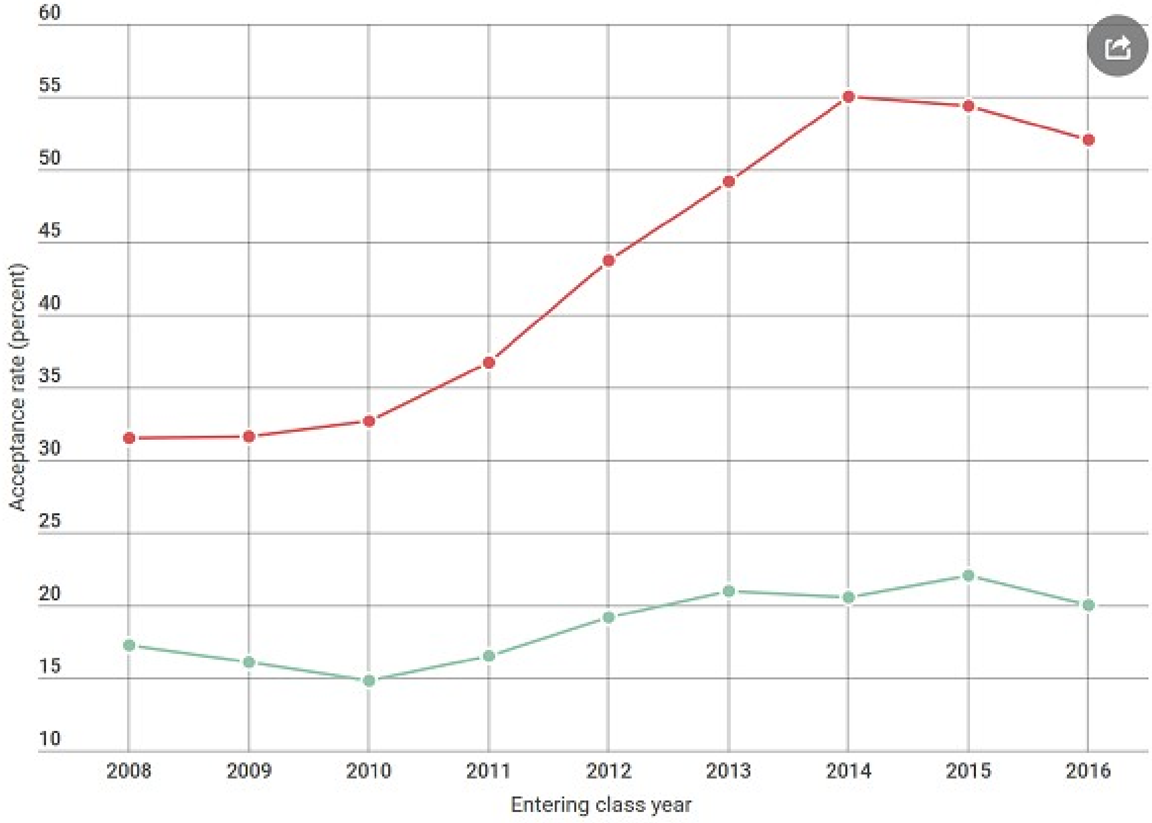
As predicted, the MCAT (the med school version of the SAT) scores necessary for admission have been increasing over time.



This is also the impression I have been getting from doctors I know who work in the medical school and residency admissions process. I got to interview some aspiring residents a few years ago for a not-even-all-that-impressive program, and they were fricking *terrifying*.

Law schools keep great data on this (thanks, law schools!). US News just tells us outright that law schools are less competitive than in 2008, even at good programs. Here’s the graph:

U.S. News data shows that acceptance rates at the top 14 law schools were only slightly higher for the entering class of 2016 than they were for the entering class of 2008. But acceptance rates at lower-ranked law schools rose more than 20 percentage points between 2008 and 2016.

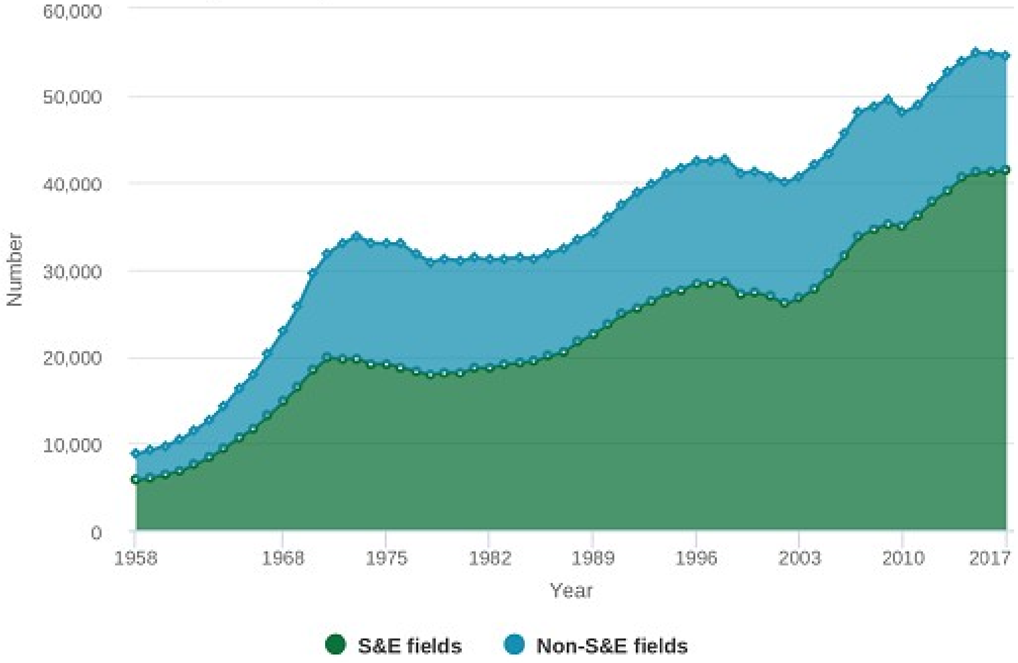


And despite it *feeling* like lawyers are everywhere these days, law school attendance has really only grown at the same rate as the population since 1970 or so, and dropped over the past decade. This may be relating to word getting out that lawyer is no longer as lucrative a career as it used to be.

Unlike law schools, graduate school basically fails to keep any statistics whatsoever, and anything that might be happening at the graduate level is a total mystery. We know the number of PhDs granted:

FIGURE 1-A

Doctorates awarded by U.S. colleges and universities: 1958–2017



S&E = science and engineering.

Source(s)
National Science Foundation, National Center for Science and Engineering Statistics, *Doctorate Recipients from U.S. Universities: 2017.*

…and that’s about it.

Part of what inspired me to write this post was listening to a famous scientist (can’t remember who) opine that back when he was a student in the 1940s, he kind of wandered into science, found a good position at a good lab, worked up the ranks to become a lab director, and ended up making great discoveries. He noted that this was unthinkable today – you have to be super-passionate to get into science grad school, and once you’re in you have to churn out grant proposals and be the best of the best to have any shot at one day having a lab of your own. I’ve heard many people say things like this, but I can’t find the evidence that would put it into perspective. If anyone knows more about the history of postgraduate education and work in the sciences, please let me know.

I’m also interested in this because it would further help explain undergraduate competition. If more people were gunning for med school and grad school, it would be more important to get into a top college in order to have a good chance of making it in. Since increasing inequality and returns to education have made advanced-degree jobs more valuable relative to bachelors-only jobs, this could explain another fraction of academic competitiveness. But aside from the medical school data, I can’t find evidence that this is really going on.

5. Is Freaking Out Over College Admissions Correct?

Dale and Krueger(2011) examine this question, using lifetime earnings as a dependent variable.

In general, they find no advantage from attending more selective colleges. Although Harvard students earn much more than University of Podunk students, this is entirely explained by Harvard only accepting the highest-ability people. Conditional on a given level of ability, people do not earn more money by going to more selective colleges.

A subgroup analysis did find that people who started out disadvantaged did gain from going to a selective college, even adjusted for pre-existing ability. Blacks, Latinos, and people from uneducated families all gained from selective college admission. The paper doesn’t speculate on why. One argument I’ve heard is that colleges, in addition to providing book-learning, help induct people into the upper class by teaching upper-class norms, speech patterns, etc, as well as by ensuring people will have an upper-class friend network. This may be irrelevant if you’re already in the upper class, but useful if you aren’t.

A second possibility might be that college degrees are a signal that help people overcome statistical discrimination.

Studies have shown that requiring applicants share drug test results or criminal histories usually increases black applicants' chances of getting hired. This is probably because biased employers assume the worst about blacks (that they're all criminal drug addicts), and so letting black applicants prove that they're not criminal drug addicts puts them on more equal footing with white/Asian people. In the same way, if employers start with an assumption of white/Asian competence and black/Latino incompetence, selective college attendance might not change their view of whites/Asians, but might represent a major update to their view of blacks/Latinos.

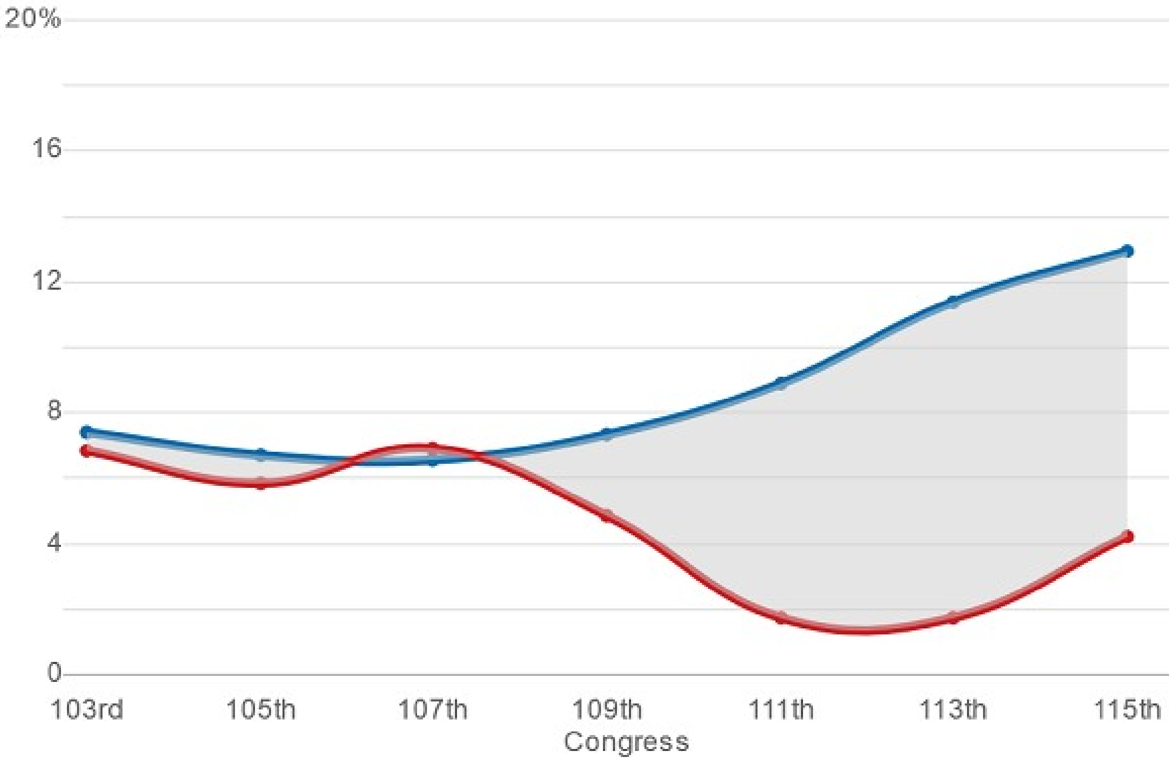
Dale and Krueger also find that the value of college did not increase during the period of their study (from 1976 to 1989).

Does this mean that at least whites and Asians can stop stressing out about what colleges they get into?

What if you want to go to medical or law school? I can't find an equally rigorous study, but sites advising prospective doctors tell them that the college they went to matters less than you'd think. The same seems true for aspiring lawyers. As usual, there is no good data for graduate schools.

What if you want to be well-connected and important?

From here, the percent of members of Congress who went to Ivy League colleges over time, by party:



Only about 8% of Congresspeople went to Ivy League colleges, which feels shockingly low considering how elite they are in other ways. The trend is going up among Democrats but not Republicans. There is obviously a 40-50 year delay here and it will be a long time before we know how likely today’s college students are to get elected to Congress. But overall this looks encouraging.

On the other hand, presidents and Supreme Court Justices are overwhelmingly Ivy. Each of the last five presidents went to an Ivy League school (Clinton went to Georgetown for undergrad, but did his law degree at Yale). Every current Supreme Court justice except Clarence Thomas went to an Ivy for undergrad, and all of them including Thomas went to an Ivy for law school. But there’s no good way to control for whether this is because of pre-existing ability or because the schools helped them succeed.

Tech entrepreneurs generally went to excellent colleges. But here we *do* have a hint that this was just pre-existing ability: many of them dropped out, suggesting that neither the coursework nor the signaling value of a degree was very important to them. Bill Gates, Mark Zuckerberg, and Larry Ellison all dropped out of top schools; Elon Musk finished his undergrad, but dropped out of a Stanford PhD program after two days. This suggests that successful tech entrepreneurs come from the population of people smart enough to get into a good college, but don't necessarily benefit from the college itself.

Overall, unless people come from a disadvantaged background, there's surprisingly little evidence that going to a good college as an undergraduate is helpful in the long term – except *possibly* for a few positions like President or Supreme Court justice.

This doesn't rule out that it's important to go to a good institution for graduate school; see this paper. In many fields, a prestigious graduate school is almost an absolute requirement for becoming a professor. But there doesn't seem to be an undergrad equivalent of this.

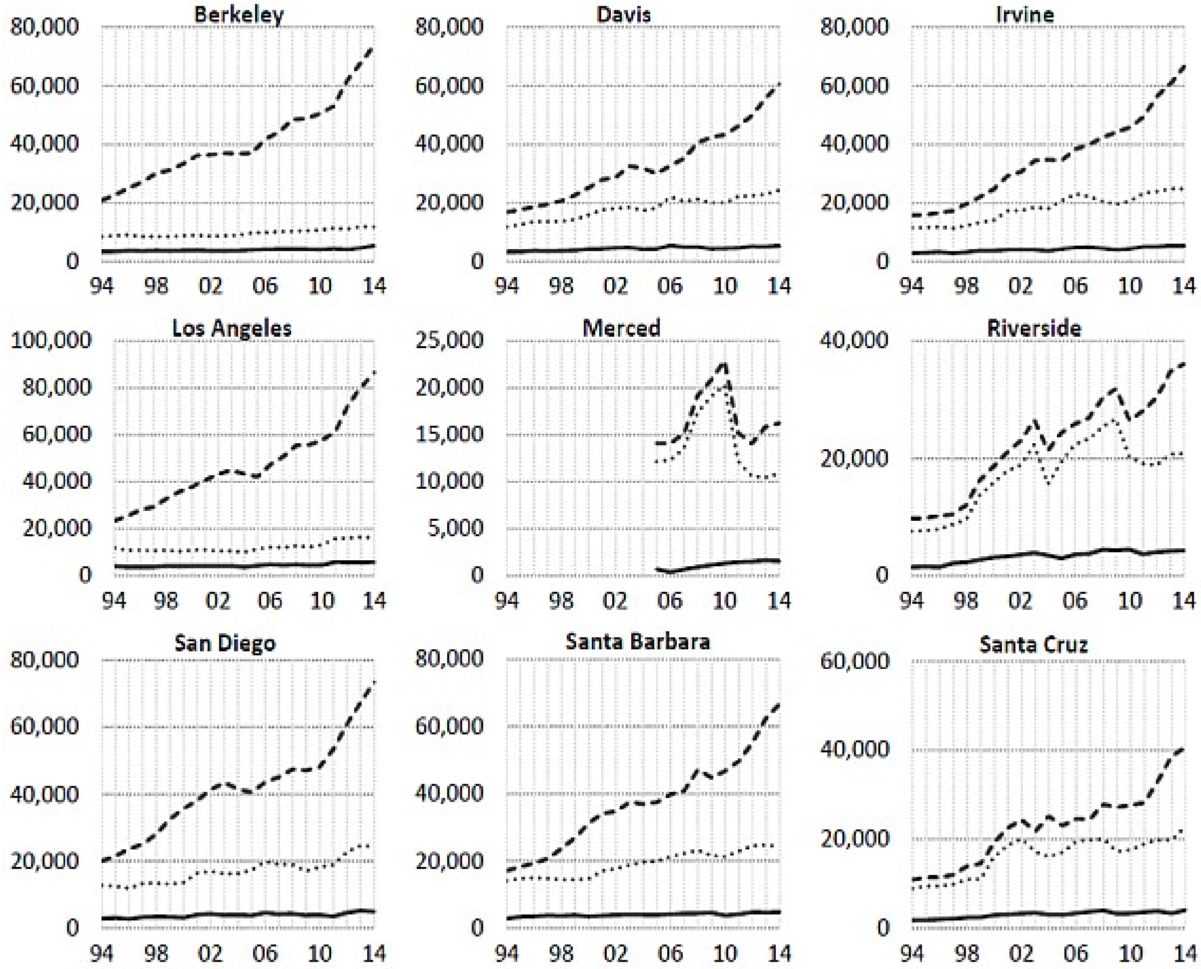
Digression: UC schools

I mentioned at the beginning the universal perception in California that UCs are much harder to get into. I know this is the perception everywhere, but it seems much worse in California. Sure, it's anecdotal evidence, but the anecdotes all sound like this:

> My friend's daughter got 3.85 GPA, had 5 AP classes in high school, was on competitive swimming team, volunteered 100+ hours, was active in school activities, yet she got rejected by all 4 UCs that she applied to. And these were not even the highest tier of UCs, not

> Berkeley. She did not apply for more schools and thought that UC San Diego and UC Santa Cruz were her safe choices. The whole family is devastated.

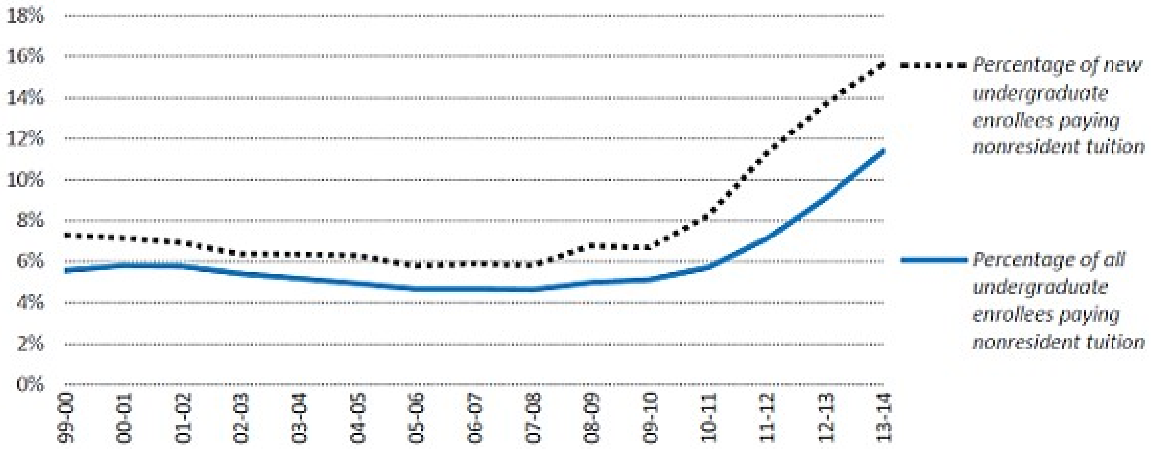
The data seem to back this up. Dashed line is applications, dotted line is admissions, solid line is enrollments:



…but I don’t know how much of this is just more applications per person, like everywhere else.

Why should UC schools be hit especially hard? I assumed California’s population was growing faster than the rest of the country’s, but this doesn’t seem true: both California and the US as a whole grew 13% between 1990 and 2000, when the cohort attending college between 2008 and 2018 would have been born.

The Atlantic points out that, because of budget cuts, UC schools are admitting more out-of-state students (who have to pay higher tuition), lowering the number of spots available to Californians. But is this really that big an effect?



It looks like nonresidents went from 6% to 12% over the space of a decade. That shouldn't screw things up so badly.

I'm really not sure about this. One possibility is that California's schools are remarkably good. On money.com's list of best colleges, four of the top ten schools are UCs, plus you get to live in California instead of freezing to death in New England. Since the college admissions crisis is concentrated at the top schools, California has been hit especially hard.

I'm not satisfied with this explanation; let me know if you know more.

6. Conclusions

1. There is strong evidence for more competition for places at top colleges now than 10, 50, or 100 years ago. There is

medium evidence that this is also true for upper-to-medium-tier colleges. It is still easy to get into medium-to-lower-tier colleges.

2. Until 1900, there was no competition for top colleges, medical schools, or law schools. A secular trend towards increasing admissions (increasing wealth + demand for skills?) plus two shocks from the GI Bill and the Vietnam draft led to a glut of applicants that overwhelmed schools and forced them to begin selecting applicants.

3. Changes up until ten years ago were because of a growing applicant pool, after which the applicant pool (both domestic and international) stopped growing and started shrinking. Increased competition since ten years ago does not involve applicant pool size.

4. Changes after ten years ago are less clear, but the most important factor is probably the ease of applying to more colleges. This causes an increase in applications-per-admission which is mostly illusory. However, part of it may be real if it means students are stratifying themselves by ability more effectively. There might also be increased competition just because students got themselves stuck in a high-competition equilibrium (ie an arms race), but in the absence of data this is just speculation.

5. Medical schools are getting harder to get into, but law schools are getting easier to get into. There is no good data for graduate schools.

6. All the hand-wringing about getting into good colleges is probably a waste of time, unless you are from a disadvantaged background. For most people, admission to a more selective college does not translate into a more lucrative career or a higher chance of admission to

postgraduate education. There may be isolated exceptions at the very top, like for Supreme Court justices.

I became interested in this topic partly because there’s a widespread feeling, across the political spectrum, that everything is getting worse. I previously investigated one facet of this – that necessities are getting more expensive – and found it to be true. Another facet is the idea that everything is more competitive and harder to get into. My parents’ generation tells stories of slacking off in high school, not worrying about it too much, and knowing they’d get into a good college anyway. Millennials tell stories of an awful dog-eat-dog world where you can have perfect grades and SAT scores and hundreds of hours of extracurriculars and still get rejected from everywhere you dreamed of.

I don’t really have a strong conclusion here. At least until ten years ago, colleges were harder to get into because more people were able to (or felt pressured to) go to college. The past ten years are more complicated, but might be because of increased stratification by ability. Is that good or bad? I’m not sure. I still don’t feel like I have a great sense of what, if anything, went wrong, whether our parents’ rose-colored picture was accurate, or whether there’s anything short of reversing all progress towards egalitarianism that could take us back. I’m interested to get comments from people who understand this area better than I do.

Highlights From The Comments On College Admissions

HalTheWise discusses a factor I missed (until I sneakily edited it in, so you may have read the later version that included it):

> One very powerful contributor that Scott did not mention is that in many cases schools are directly or indically intentivized to have a low admission rate. US news & world report released the first national college ranking in 1983, and donors and board members at various schools have increasingly been using national rankings performance, which directly includes low admissions rates, as a measure of how well a school is doing.
>
> These rankings and metrics also heavily incentivize having high yield (a large fraction of students that are admitted end up attending) which for a fixed size applicant pool also encourages accepting as few people as possible. This has led to the death of safety schools, because they would rather reject a high performing student than admit them and have them not attend.
>
> These factors might also be a driving force behind the rise of common app, since schools are trying to get as many applicants as possible, even if it hurts the quality of their pool.

---

kaakitwitaasota points out that consulting is an exception to the “where you go to school doesn’t matter” principle:

> A lot of top firms these days won’t even look at you if you didn’t go to the “right” college. My mother did her MBA at Northeastern, and recently had lunch with an old classmate who ended up at a top consulting firm. My mother’s classmate’s résumé would end up in the trash unread these days–Northeastern isn’t considered good enough.
>
> So while it’s probably true on the macro level that smart kids will do just fine anywhere they end up, there is a subset of extremely prestigious, extremely well-paid jobs which will not even look at you if you didn’t get into the right institution at the age of 18–which, in practice, means that the élite are chosen on the basis of who they were at the age of 14-17. When viewed in those terms, it’s completely nuts.

I’d heard this before; my impression is that a big part of consulting is having prestigious-looking people tell you what you want to hear. If what they’re actually hiring for is prestige rather than competence per se, that could make it a special case

---

prunesquallor on UCs:

> I believe UCs are more competitive because of cost. Personally, I got into a number of strong private schools, and chose a UC because it is significantly cheaper. In the past, this would not have been as much of a factor, because college was affordable. I’m not sure if this applies to public schools in other states. The UCs are the best public schools in the country and are able to compete with high level private schools.

This makes sense, but I’m not sure exactly what the model is. UCs are cheap, so many people across the country apply to them? (but I gave data showing out-of-staters were only 14% of UC students, plus it may not be cheap for out-of-staters). UCs are cheap, so many Californians apply? Aren't public universities always cheap for people in the state? Maybe cheap UCs mean more top-performing Californians apply to UCs instead of private out-of-state colleges?

BlindKungFuMaster writes about a factor that could explain stories of exceptional students being rejected from everywhere they apply:

> It could be the case that college admissions became more random as kids apply more often and the metrics become more vague. I.e. there used to be x acceptable applicants of which 50% were somewhat randomly selected. Now there are 2x acceptable applicants of which 25% are somewhat randomly selected. Then there are just much more kids being unlucky and missing out on all their choices, though it hasn't really been getting harder to get in. Of course individually the remedy is applying to even more colleges.

This might increase perceptions of selectivity – if one of these bright students posts their sob story, everyone else will just think that standards are very high.

A few people chimed in with concerns about the Dale and Krueger paper. rlms wrote:

> From the abstract:
>
> "We find that the return to college selectivity is sizeable for both cohorts in regression models that control for variables commonly observed by researchers, such as student high school GPA and SAT scores. However, when

> we adjust for unobserved student ability by controlling for the average SAT score of the colleges that students applied to, our estimates of the return to college selectivity fall substantially and are generally indistinguishable from zero."
>
> So college selectivity *is* significant even after controlling for student quality as measured by SAT scores. It only ceases to be significant when you also control for some vague measure of ambition as signalled by the average SAT score of all the colleges they applied to.

And reasoner cites a 2009 Overcoming Bias post finding that a previous Dale & Krueger paper on this subject was misinterpreted to say college didn't matter when it really did matter (Marginal Revolution also covered this). The Dale & Krueger paper I posted was an update to that one that said no, really, college doesn't seem to matter. But I haven't had time to look at it closely myself, so this shifts my priors a little bit.

aesthesia is also doubtful:

> If I remember correctly, the Dale and Kreuger paper is somewhat limited in its conclusions. Their sample was weighted toward the high end of the achievement spectrum, so it really says something like: conditional on being accepted to Harvard, there's not much difference in lifetime earnings between actually attending Harvard and instead choosing to attend UC Berkeley. There aren't a lot of students admitted to Harvard who instead choose to go to the University of Southern North Dakota at Hoople, so we don't really know what happens to them.

> This is a somewhat personal issue for me: I read summaries of Dale and Kreuger’s earlier work when applying for college and decided not to bother applying to selective schools, and just went to a middling state university, thinking that going somewhere more selective wouldn’t make a difference in my future life. I’m no longer confident that was the right move. I believe I would have learned more and made better and stronger connections had I gone somewhere more difficult to get into.

But eqdw’s experience bears Dale & Krueger out:

> Hello. I work at a major company involved in job search, job ads, hiring, etc. And I would like to share with you something from a quarterly status update I saw the other day.
>
> Status update presentation had a slide outlining the results of a user study we did. We surveyed employers to find out what are the most vs least important details they look at when making a hiring decision.
>
> Out of something like 20 different options surveyed, “where the candidate went to college” was rated dead last in importance. “Formatting of resume” was rated as more important than “where they went to college” for making a hiring decision.
>
> This would seem to confirm conclusion #6 above

Freddie de Boer wants to remind us (another point I stealthily edited into the post later on) that we really are talking about a small subset of institutions here:

> I ran the numbers myself several years ago. Out of 3000+ accredited two- and four-year colleges, something like ~150 reject more students than they accept. The large majority accept almost every student who applies. And of course only half of the population will ever enroll in college and only a third will ever finish. The people who say this is a niche problem are correct.

Gossage Vardebian on money:

> Certainly the idea of getting into the best school you can, is a further manifestation of the competitiveness discussed above. But that requires another factor, the “and damn the expenses” factor. I went to State U because I wasn’t good enough to get into U of Ivy, but also because I couldn’t afford it. Well, hell, the me of 1980 couldn’t afford 2019 State U either, so hey, might as well go for Harvard too. And if my parents has enough money for State U, well then Harvard is that much less financially onerous, isn’t it? This is also a reason why more kids these days apply to out-of-state State universities, which in turn are admitting more out-of-state kids – the difference in cost is not as large as in the past. It used to be that the idea of going into tens of thousands of dollars of debt for college was insane – literally nobody I know did that. People would do it for med school or law school, and that’s it. I’m sure there were examples from other demographics where that happened, but now of course it is commonplace, as a quick Google search will confirm. Part of this is the whole “you must get into the best school possible” mindset, but part is just a “that’s just the way it is now” resignation. The mental leap to take on that kind of debt has been conditioned into families now.

Anthony adds:

> The debt thing is a big change from when I went to college, but it's driven by policy more than by parents.
>
> The government encouraged student loans because those would get paid back while subsidizing tuition at the student end or the college end wouldn't (directly), and since they were paid back, the had very little budget cost. With a relatively significant college premium, people were willing to take on fairly large amounts of debt (and were not very likely to default). This allowed colleges to spend more, driving up the sticker price, making loans more necessary for students who weren't poor enough to get direct subsidies or rich enough to afford the higher price.
>
> Unfortunately, this is a politically very difficult problem to unwind. Even though the best policy would be to stop subsidizing student loans entirely and make them dischargeable in bankruptcy, which would hugely limit the availability of student loans, that's never going to happen.

Peffern relates their own experience:

> I think point 3.4 is the most important, based on anecdotal evidence.
>
> I'm in undergrad at a top school right now.
>
> In high school, I was a good student – perfect SAT, good APs, reasonable GPA, etc. Despite having good accomplishments here it didn't feel particularly effortful – I'm good at math and can structure a coherent

> argument, so taking the SAT was mostly just getting a good night’s sleep and studying vocabulary for a week.
>
> I also did a lot of extracurriculars, and the work and stress load from those absolutely destroyed me. I’m not even talking about “starting homework at 11pm” kind of workload, I mean the social aspect, the cutthroat politics, the status games, and the showmanship. It’s not that I think those things are necessarily bad in and of themselves, but they’re infinitely more difficult than classwork. I’m incredibly busy these days with class and I’m not even on the same order of magnitude of stress as I was in HS.
>
> High school students are vicious bastards. When you take the AP calculus exam, or the SAT, or even just the final for some class you take, you are only really competing against the teacher, the test, and yourself. When you do extracurriculars, you are competing against horrible entitled jerks with rich parents who make your life miserable. I would take a hundred AP exams before doing another pointless extra curricular.
>
> I don’t know if it explains the college enrollment statistics but it certainly explains the outrage, pessimism, and anger of people my age over the process. I spent what was for past generations an exciting and important time of my life locked in a box sanding off all aspects of myself that didn’t perfectly resembke an ivy league student just to get beaten out by some kid whose father got him an internship at somewhere prestigious. That does things to people.

---

meh finds reason for doubting my thoughts on the Common App:

> The Pew article you link has a different conclusion about the Common App:
>
> “The expansion of the Common Application, which makes it easier for students to apply to multiple schools, doesn’t appear to be behind the increase in application volume. The Common App, as it’s called, is accepted by nearly 800 colleges and universities in the United States and several dozen overseas. Of the 1,364 institutions in our sample, 729 accept the Common App along with (or in some cases instead of) their own application forms; the other 635 use their own forms. Although one might suspect that the ease of applying to multiple schools via the Common App would result in stronger growth in application volume among those schools, there was almost no difference in 2002-2017 growth rates between the schools that used the Common App and those that didn’t.”

But adds:

> But is their reasoning sound? Isn’t it possible the Common App increases applications uniformly among all schools, not just ones using the Common App?
>
> Consider if pre-Common App, I am willing to fill out 10 applications, so I apply to 10 schools. Now with the Common App, say half of them accept it, so I only need to fill out 6 applications. I am still willing to fill out 4 more applications, so I look at my 11th choice. If they take the Common App, I apply to them for free. If not, I fill out an application, and am still willing to fill out 3 more applications. Does this lead to a similar increase in applications for both Common App and non Common App schools?

And alexhutcheson has a more complicated model for why applications per student might have increased:

> I think this analysis is excellent, but I don't think the Common App explains the increase in applications-per-admission. There's another significant factor that causes students to apply to dozens of schools: No one can accurately forecast what a given school will cost anymore.
>
> At some point (I believe in the late '00s, but could be wrong), elite schools started to extend their financial aid programs to include students from middle-class families. Here's a press release from Harvard in 2007. Yale and Princeton followed quickly, and most other elite schools seem to have done similar things, albeit with more constrained resources.
>
> Prior to this change, a student from a family with middle-class income or above could know with reasonable certainty what a given college would cost: they would expect to pay the tuition, fees, etc. listed on the brochures. After this change, they would have no idea until after they were admitted. They might get a generous financial aid package that brings the cost down to the price of their local state university, or they might get nothing except loans. The systems used to determine financial aid packages are opaque and not well-publicized, so the outcome is unpredictable.
>
> In 2009, I applied to a broad sample of 14 schools on the east coast. They were a mix of "elite" private and flagship public universities. My parents were comfortably middle-class. I was lucky enough to get into most of them, and so I had the opportunity to compare financial offers. In maybe 1 case out of 12, I would have

> had to pay the full sticker price. The rest would have been heavily "discounted", but the discount varied widely between schools, from ~5% off the total cost of attendance, to 60% off, to 100% off (full ride). For the private universities the "discount" came from a mix of "need-based aid" and merit scholarships, while for the public universities it was exclusively from merit scholarships.
>
> In this system, you don't know what the financial offer is until you get in, there are possible windfalls from getting a generous financial offer, and it's difficult to predict in advance what the financial offer will be until you get in. The incentives here are obvious: student who are conscious about the cost of their education have a strong incentive to "play the lottery" by applying to as many schools as is feasible for them, while biasing towards schools that are known to provide generous financial packages.

---

faoiseam says that APs might be a good objective way to track the increase in competition over time:

> The College Board has lots of data on AP exams since 1996. They track how many people take exams, in what grade, by race and result, which is enough to get a sense of how people did. They also keep track of AP scholars. In 2002, 1738 people were National AP scholars, (8 APs with score 4 or more) enough to get you into HYPSM. This increased to 7k by 2005, 16k by 2010 and now is 35k, not enough to get you into a top 20 school. By this measure, college entry has got more competitive.
>
> The number of schools offering APs has increased in that period, from 10k in 2000, to 17k in 2010, to 20k

> now. Looking at individual subjects, the two most taken are AP Lit, and AP Calc. 5s in lit increased 10 fold from 2000 (6978/105k) to now (61k/580k). For AP Calc BC in 2000 we have 5k 5s from 15k exams, in 2010 39k 5s, from 79k exams, and now we have 56k 5s from 139k exams. As these are the most common exams, this shows that while the number of schools barely changed in the last 10 years, the number of 5s has doubled.
>
> In general, things have gotten twice as hard since 2010 (as measured by AP exams), and 5 times 3005, and 10 to 20 times since 2000. The AP curriculum has stayed pretty stable, so that does not explain the change. The number of schools has been pretty steady since 2010, though it did double since 2000. This does not explain all of a factor of 2, as the 10k schools that offered APs in 2000 were the better schools anyway.
>
> This makes me question why the US does not used AP exams for college admission. If, as I estimate, the number of people who get n 5s doubles as n increases by 1, this suggests that four 4s and four 5s should be enough to get into HYPSM. The UCs should requires 8 4s. or 7 4s and 1 5, etc.
>
> This could be done by computer when AP results come out, delaying admission until the Summer of senior year. It would be really nice to have a clear bar that was required for entry, as opposed to the current opaque system.

Steve Sailor fills in a hole in my data:

> I wrote a column back in 2013 documenting that famous colleges were not expanding their freshmen class sizes

> at anywhere close to population growth over the last generation:
>
> Consider the growth rate of Harvard, the world's richest university. The number of undergraduates in its class of 1986 was 1,722. After a quarter of a century, during which the US population grew by 75 million, Harvard's class of 2011 was 1,726: an increase of four.
>
> This is not to say that Harvard isn't expanding: Faculty and grad students are up, and non-teaching staff skyrocketed.
>
> Similarly, Yale's undergraduate student body has been the same size since 1978. Five years ago, the second-richest college announced a proposal for adding a couple of dormitories, but construction won't proceed until another $300 million is raised.
>
> In 2010, MIT unveiled plans to expand undergrad enrollment by six percent, which would only get it back to where it was in the 1990s.
>
> Among the most prominent colleges, Princeton is the only one over the last generation to have actually succeeded in boosting enrollment (and that by only about ten percent) after it opened the Whitman residential college in 2007.

…and gives some fictional evidence:

> Robert Heinlein's 1950s juvenile novels are full of scenes where people ask the hero what he's going to do for college and he says, "I dunno, I guess the day after Labor Day I'll drive down to State and sign up for some classes." (That was pretty much what UCLA was like when Heinlein went there for awhile in the 1930s.) But

> in the end of “Have Space Suit, Will Travel,” the hero gets accepted into MIT when the head of the CIA, or whatever, pulls some strings.

This concept of heading down to the college the day it opens to sign up for classes reminds me of an anecdote I rejected for the main post because it was too unbelievable. Wikipedia quotes Supreme Court Justice Byron White saying he went to Yale for law school because he was planning to go to Harvard, but got sick enough on the train to Boston that he decided to get off at New Haven and go to Yale instead. But I haven’t been able to find any source better than Wikipedia for this, so I don’t know if it’s true. Does it at least mean that this was plausible enough in the 1930s (when White was going to college) to make a believable urban legend?

---

graeme on HackerNews gives the Canadian perspective:

> You may be interested in the view from up North. Joseph Heath, a Canadian academic has commented on how the top three Canadian schools teach more students than the top ten American schools. I’ll excerpt:
>
> ————-
>
> “But of course there’s a reason that it’s so difficult to get into Yale – it’s because Yale has only 5,400 students, in a country of over 315 million people! By contrast, McGill has over 30,000, and University of Toronto has 67,000 undergraduates, serving a country of only 35 million people. That means there’s roughly one spot at Yale for every 58,000 Americans, compared to one spot at McGill for every 1,100 Canadians. No wonder American life is more competitive.

Furthermore, all of the best schools in the United States are tiny. Here is a list of the top 10, as ranked by U.S. News & World Report, along with the number of students (undergraduate, I believe):

Princeton: 5,336

Harvard: 6,658

Yale: 5,405

Columbia: 6,068

Stanford: 7,063

Chicago: 5,590

Duke: 6,655

MIT: 4,503

Upenn: 9,682

CIT: 997

Dartmouth: 4,193

That means the top 10 universities in the United States – a country of over 315 million people – at any given time are educating a grand total of only 62,150 students.

By contrast, here are the rough numbers of undergraduates at the top 3 Canadian universities:

McGill: 30,000

UBC: 47,500

> UofT: 67,000
>
> So the top 3 Canadian schools are at any given time educating a grand total of 144,500 students – more than twice the total of the top 10 U.S. schools. (In fact, the University of Toronto alone has more student capacity than the top 10 U.S. schools combined.) The United States has almost exactly 9 times the population of Canada, so in order to have the same sort of capacity in higher education, the top 27 schools in the United States would have to have 1.3 million students.”

This suggests a model where people only have so many cognitive real estate devoted to remembering what the good schools are, and so if your top three schools are very large, you can have many people in the “cognitive elite”, but if your top three schools are small the elite will seem more selective. Seems like a stupid way to run an educational system

---

jumpinjacksplash offers a model of how an arms race would work:

> I think what people may be missing is that time is a thing:
>
> In a world where there are 10 places at Harvard every year but 9 applicants, everyone who can read Latin gets in, and no-one bothers to study harder than what it takes them to meet the minimum requirement.
>
> In a world where there are 10 places at Harvard but 11 applicants, the worst Latin-reader doesn’t get in. But someone who realises they’re the worst Latin readers will practice more, and someone else will be the worst. That person will then practice enough to make someone

> else be the worse, who then practices a bit more. Given enough time, the standard keeps rising as everybody needs to stay ahead of everybody else; eventually everyone hits the ceiling of Latin literacy and starts taking up the violin, then competes on violin-skills until they all have to start climbing Everest.
>
> The only equilibrium is where everyone exerts maximum effort on top of their abilities. It takes a long time for this to become even remotely necessary (I still don’t think we’re close to being there). But a world where the first 18 years of your life consists of maximum effort to get into college is a dystopian nightmare, hence as we trend towards it everything gets constantly worse.
>
> This also explains the anecdotal people who expect to swim into Berkeley but get rejected from UCLA: their parents think they’ve massively overshot what’s required, but they’ve failed to realise how far down the slope we’ve slid.

I failed in my original goal for this piece, which was to present an account of competition getting tougher across society. There wasn’t enough data, so I fell victim to a streetlight effect where I concentrated on colleges, and then another streetlight effect where I concentrated on Harvard (which is the best-documented college). But the reasons why competition is getting tougher at Harvard probably don’t generalize, so I ruined my chance to have a really good post about modern competition. Still, a few people stuck to the original spirit and discussed increasing competition as it was showing itself in their own field.

anechoicmedia from the subreddit on TV:

This reminds me of a depressing perspective relayed by Brady Haran on podcast. An acquaintance had described getting started in the TV business back in what would probably have been the 1980s or 90s. Like most young people of the pre-computer world he arrived with no independent film making experience. As he describes it, he presented himself to a major TV station, applied for a job, and was accepted into an apprenticeship of sorts, in which they taught him all the ways of TV-news-making from the ground up, which he then went on to excel at.

Brady, himself a BBC veteran, remarked that such a thing could never happen today. Entry positions are ferociously competitive. Young adults applying today can be expected to have had a quality camera in their hands since childhood. There are no apprenticeships; New hires are expected to show up with fundamental knowledge already in hand.

But more than that, the modern ability to signal your ~passion~ via independent YouTube/Vimeo presence means you are really expected to show up with an entire history of volunteer video production going back years. How can you be worthy of working in the BBC if you haven’t been using the tools at your disposal to produce content for years already? You must not really want it — you’d be laughed out of the office without an extensive list of “extracurriculars” in hand. The ladder of opportunity had been yanked quite a few rungs upward.

Something like this may also be at work with the expectation that “real programmers”, especially younger ones, have many side projects, open source contributions, etc.

And jaghataikhan on consulting again:

> This reminds me of my field (management consulting). It's pretty competitive to get into the big three (McKinsey/Bain/BCG), so the level of preparation in applicants is also a runaway arms race. One of my buddies who went to MIT said he'd known people who'd prepped for case interviews for two years!
>
> In contrast, I remember seeing some pre Recession recruiting videos at my firm where there was a girl from Yale saying something to the effect of "I didn't even know what management consulting was before I interviewed!"
>
> I can't tell you how quaint/laughable that sounds vs. the extreme levels of competition for the same jobs today. Basically if you don't have a top resume/school/marks/extracurrics/internships/cases/etc from day 1 (and still a matter of luck btw unless your dad is a partner or something), don't even bother applying. And I'm certain the level of competition is skyrocketing daily.

KingWalrax gives the only thing I've heard that could possibly be a general explanation of the phenomena, though I don't know if it's right:

> Ex-YCombinator President Sam Altman wrote a decent essay explaining how Zero-Sum Existential Competition & Conflict arise from a lack of baseline Growth in any given system. Whatever you think of Tech/YC, it's a good read and Sam is a smart guy and he's on-the-money here.
>
> GDP Growth is always spiky/noisy, and highs average out with lows, making it difficult to discuss overall

> performance. But Sam’s graphs tell a pretty decent picture: we haven’t had a year of 10% GDP Growth since the 1980s, but our recession-year performance (i.e. 2008) is as bad as ever.
>
> Your essay here already tied the broader economic landscape in to College a few times (Farming & Sources of employment in pre-1800s New England). It strikes me as relevant again here in this post-WW2 era.
>
> Growth declined. Therefore: Zero-Sum Competition increased. The only other option was individual actors settled for lower individual outcomes, and you wrote the best essay out there on why that doesn’t happen.

I’ve seen conflicting data on whether long-term growth has really decreased and when that started, and you can see some discussion of this in the comments. If it has, it would explain a lot, and would mean there’s still room for things to get worse.

1960: The Year The Singularity Was Cancelled

*[**Epistemic status:** Very speculative, especially Parts 3 and 4. Like many good things, this post is based on a conversation with Paul Christiano; most of the good ideas are his, any errors are mine.]*

I.

In the 1950s, an Austrian scientist discovered a series of equations that he claimed could model history. They matched past data with startling accuracy. But when extended into the future, they predicted the world would end on November 13, 2026.

This sounds like the plot of a sci-fi book. But it's also the story of Heinz von Foerster, a mid-century physicist, cybernetician, cognitive scientist, and philosopher.

His problems started when he became interested in human population dynamics.

(the rest of this section is loosely adapted from his *Science* paper "Doomsday: Friday, 13 November, A.D. 2026")

Assume a perfect paradisiacal Garden of Eden with infinite resources. Start with two people – Adam and Eve – and assume the population doubles every generation. In the second generation there are 4 people; in the third, 8. This is that old riddle about the grains of rice on the chessboard again. By the 64th generation (ie after about 1500 years) there will be 18,446,744,073,709,551,616 people – ie about about a billion times the number of people who have ever lived in all the eons of human history. So one of our assumptions must be wrong. Probably it's the one about the perfect paradise with unlimited resources.

Okay, new plan. Assume a world with a limited food supply / limited carrying capacity. If you want, imagine it as an island where everyone eats coconuts. But there are only enough coconuts to support 100 people. If the population reproduces beyond 100 people, some of them will starve, until they're back at 100 people. In the second generation, there are 100 people. In the third generation, still 100 people. And so on to infinity. Here the population never grows at all. But that doesn't match real life either.

But von Foerster knew that technological advance can change the carrying capacity of an area of land. If our hypothetical islanders discover new coconut-tree-farming techniques, they may be able to get twice as much food, increasing the maximum population to 200. If they learn to fish, they might open up entirely new realms of food production, increasing population into the thousands.

So the rate of population growth is neither the double-per-generation of a perfect paradise, nor the zero-per-generation of a stagnant island. Rather, it depends on the rate of economic and technological growth. In particular, in a closed system that is already at its carrying capacity and with zero marginal return to extra labor, population growth equals productivity growth.

What causes productivity growth? Technological advance. What causes technological advance? Lots of things, but von Foerster's model reduced it to one: people. Each person has a certain percent chance of coming up with a new discovery that improves the economy, so productivity growth will be a function of population.

So in the model, the first generation will come up with some small number of technological advances. This allows them to spawn a slightly bigger second generation. This new slightly larger population will generate slightly more technological advances. So each generation, the population will grow at a slightly faster rate than the generation before.

This matches reality. The world population barely increased at all in the millennium from 2000 BC to 1000 BC. But it doubled in the fifty years from 1910 to 1960. In fact, using his model, von Foerster was able to come up with an equation that predicted the population near-perfectly from the Stone Age until his own day.

But his equations corresponded to something called hyperbolic growth. In hyperbolic growth, a feedback cycle – in this case population causes technology causes more population causes more technology – leads to growth increasing rapidly and finally shooting to infinity. Imagine a simplified version of Foerster's system where the world starts with 100 million people in 1 AD and a doubling time of 1000 years, and the doubling time decreases by half after each doubling. It might predict something like this:

1 AD: 100 million people
1000 AD: 200 million people
1500 AD: 400 million people
1750 AD: 800 million people
1875 AD: 1600 million people

…and so on. This system reaches infinite population in finite time (ie before the year 2000). The real model that von Foerster got after analyzing real population growth was pretty similar to this, except that it reached infinite population in 2026, give or take a few years (his pinpointing of Friday November 13 was mostly a joke; the equations were not really that precise).

What went wrong? Two things.

First, as von Foerster knew (again, it was kind of a joke) the technological advance model isn’t literally true. His hyperbolic model just operates as an upper bound on the Garden of Eden scenario. Even in the Garden of Eden, population can’t do more than double every generation.

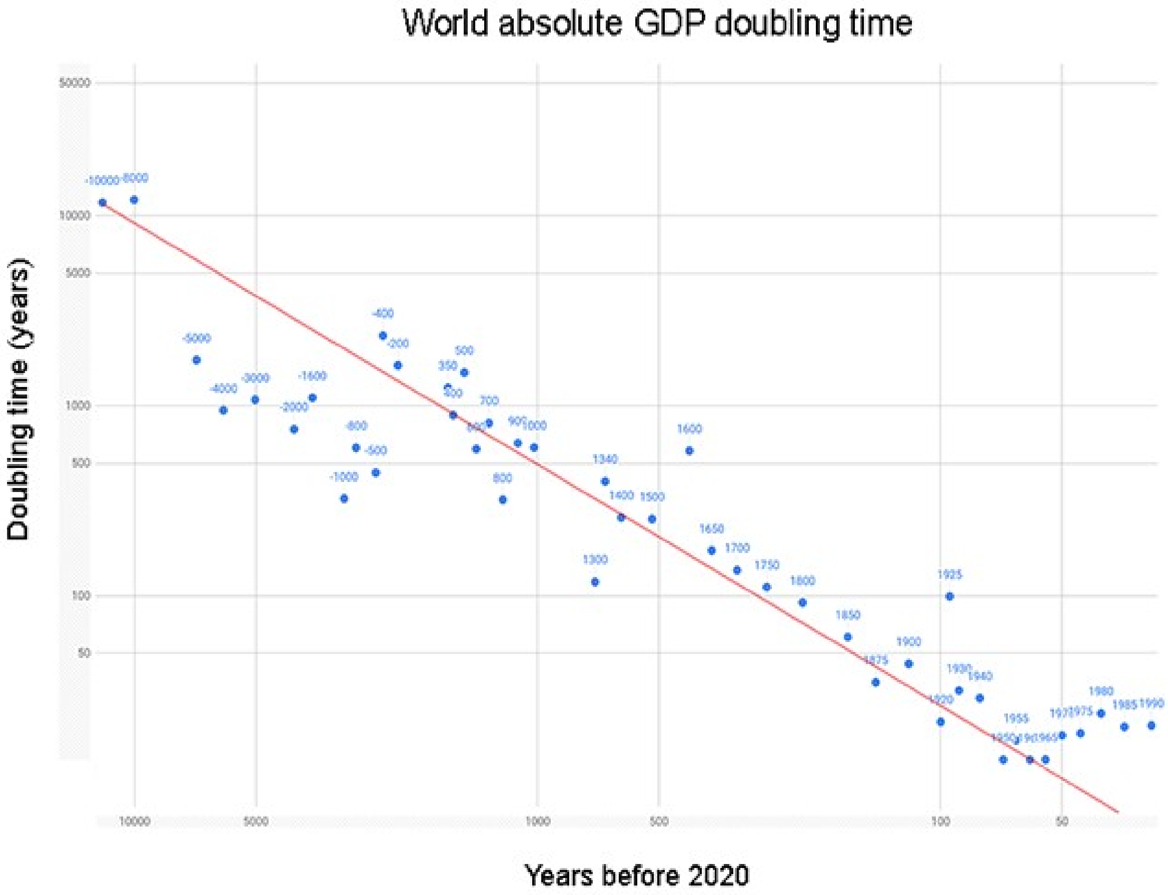
Second, contra all previous history, people in the 1900s started to have fewer kids than their resources could support (the demographic transition). Couples started considering the cost of college, and the difficulty of maternity leave, and all that, and decided that maybe they should stop at 2.5 kids (or just get a puppy instead).

Von Foerster published has paper in 1960, which ironically was the last year that his equations held true. Starting in 1961, population left its hyperbolic growth path. It is now expected to stabilize by the end of the 21st century.

II.

But nobody really expected the population to reach infinity. Armed with this story, let’s look at something more interesting.

This (source) might be the most depressing graph ever:



The horizontal axis is years before 2020, a random year chosen so that we can put this in log scale without negative values screwing everything up. This is an arbitrary choice, but you can also graph it [with log GDP as the horizontal axis]() and find a similar pattern.

The vertical axis is the amount of time it took the world economy to double from that year, according to [this paper](). So for example, if at some point the economy doubled every twenty years, the dot for that point is at twenty. The doubling time decreases throughout most of the period being examined, indicating hyperbolic growth.

Hyperbolic growth, as mentioned before, shoots to infinity at some specific point. On this graph, that point is represented by the doubling time reaching zero. Once the economy doubles every zero years, you might as well call it infinite.

For all of human history, economic progress formed a near-perfect straight line pointed at the early 21st century. Its destination varied by a century or two now and then, but never more than that. If an ancient Egyptian economist had modern techniques and methodologies, he could have made a graph like this and predicted it would reach infinity around the early 21st century. If a Roman had done the same thing, using the economic data available in his own time, he would have predicted the early 21st century too. A medieval Burugundian? Early 21st century. A Victorian Englishman? Early 21st century. A Stalinist Russian? Early 21st century. The trend was *really* resilient.

In 2005, inventor Ray Kurzweil published *The Singularity Is Near*, claiming there would be a technological singularity in the early 21st century. He didn’t refer to this graph specifically, but he highlighted this same trend of everything getting faster, including rates of change. Kurzweil took the infinity at the end of this graph very seriously; he thought that some event would happen that really *would* catapult the economy to infinity. Why not? Every data point from the Stone Age to the Atomic Age agreed on this.

This graph shows the Singularity getting cancelled.

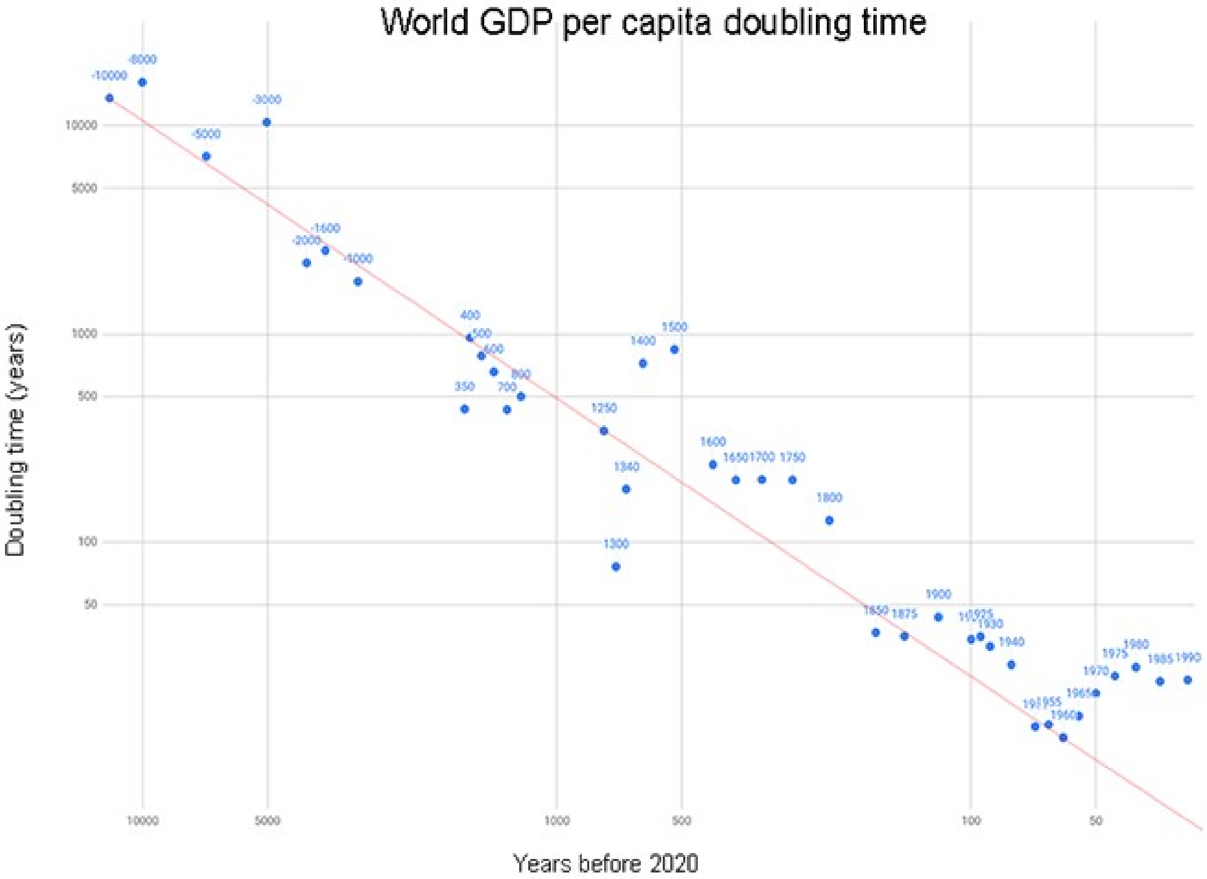
Around 1960, doubling times stopped decreasing. The economy kept growing. But now it grows at a flat rate. It shows no signs of reaching infinity; not soon, not ever. Just constant, boring 2% GDP growth for the rest of time.

Why?

Here von Foerster has a ready answer prepared for us: population!

Economic growth is a function of population and productivity. And productivity depends on technological advancement and technological advancement depends on population, so it all bottoms out in population in the end. And population looked like it was going to grow hyperbolically until 1960, after which it stopped. That’s why hyperbolic economic growth, ie progress towards an economic singularity, stopped then too.

In fact…



This is a *really sketchy* graph of per capita income doubling times. It’s sketchy because until 1650, per capita income wasn’t really increasing at all. It was following a one-step-forward one-step-back pattern. But if you take out all the steps back and just watch how quickly it took the steps forward, you get something like this.

Even though per capita income tries to abstract out population, it displays the same pattern. Until 1960, we were on track for a singularity where everyone earned infinite money. After 1960, the graph “bounces back” and growth rates stabilize or even decrease.

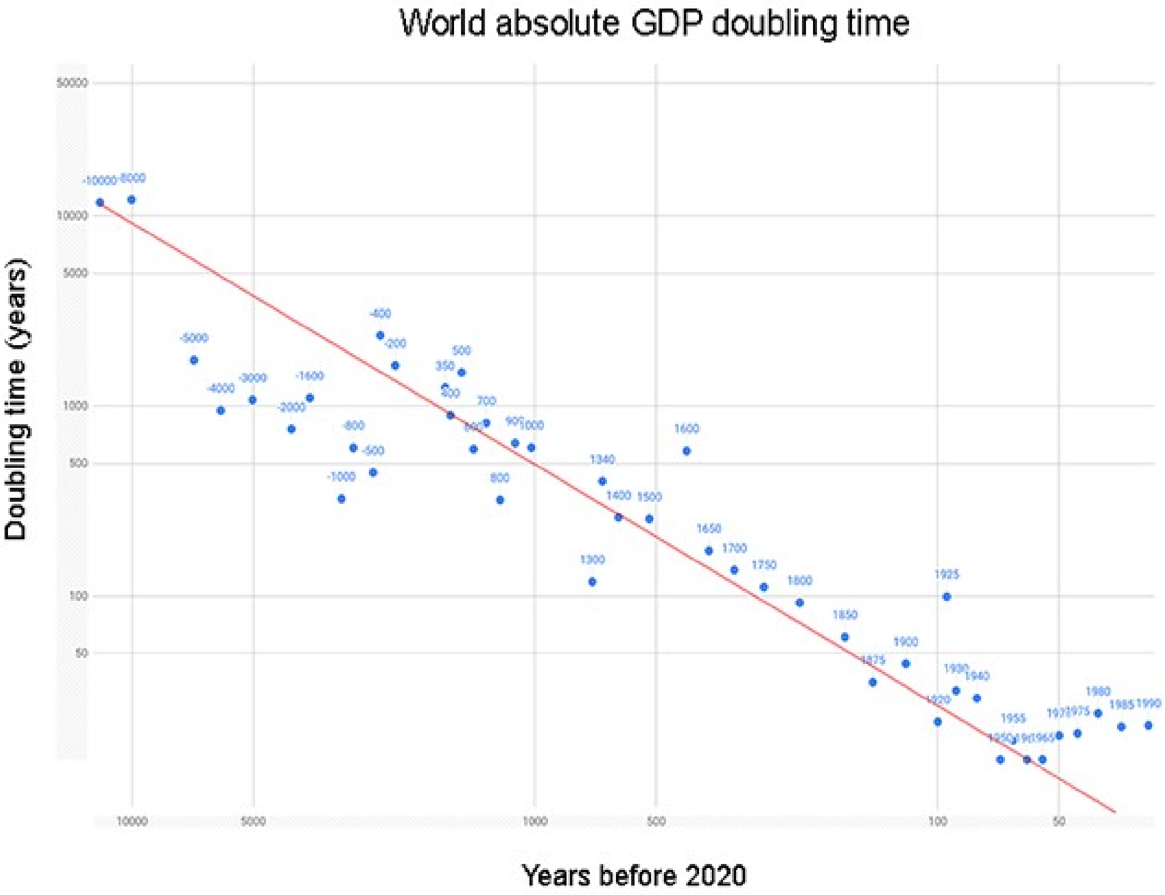
Again, von Foerster can explain this to us. Per capita income grows when technology grows, and technology grows when the population grows. The signal from the end of hyperbolic population growth shows up here too.

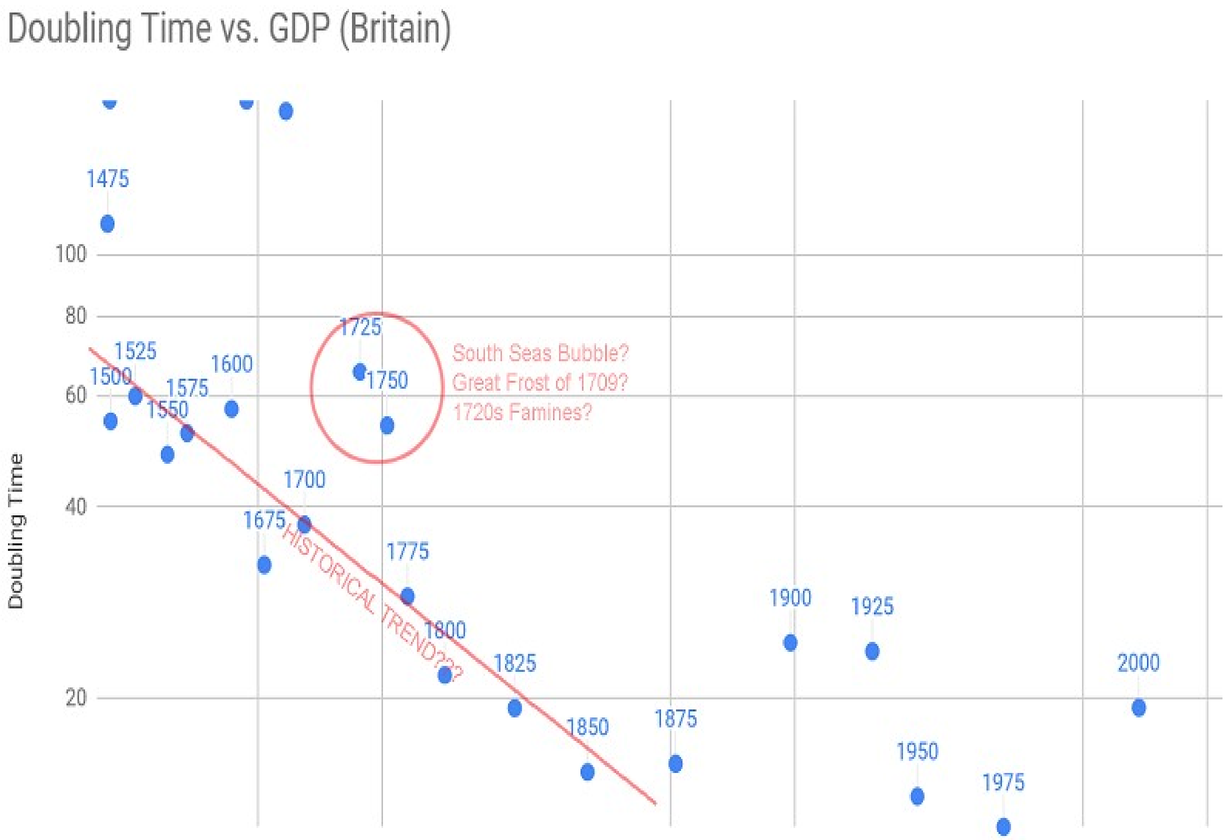
To make this really work, we probably have to zoom in a little bit and look at concrete reality. Most technological advances come from a few advanced countries whose population stabilized a little earlier than the world population. Of the constant population, an increasing fraction are becoming researchers each year (on the other hand, the low-hanging fruit gets picked off and technological advance becomes harder with time). All of these factors mean we shouldn’t expect productivity growth/GWP per capita growth/technological growth to *exactly* track population growth. But on the sort of orders-of-magnitude scale you can see on logarithmic graphs like the ones above, it should be pretty close.

So it looks like past predictions of a techno-economic singularity for the early 21st century were based on extrapolations of a hyperbolic trend in technology/economy that depended on a hyperbolic trend in population. Since the population singularity didn’t pan out, we shouldn’t expect the techno-economic singularity to pan out either. In fact, since population in advanced countries is starting to “stagnate” relative to earlier eras, we should expect a relative techno-economic stagnation too.

…maybe. Before coming back to this, let’s explore some of the other implications of these models.

III.

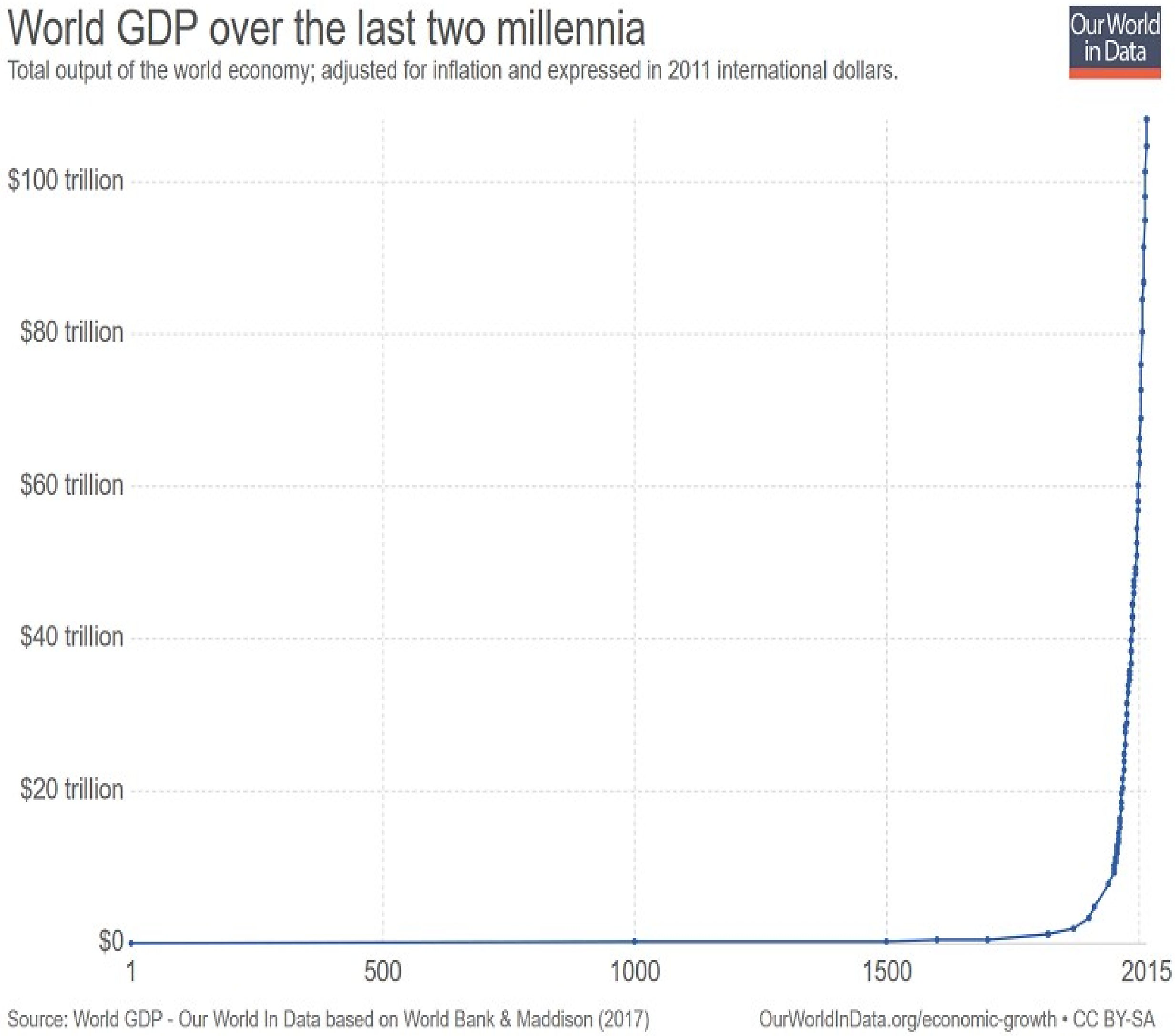

World absolute GDP doubling time
Doubling time (years)
50000
10000
5000
1000
500
100
50
-10000
-8000
-5000
-4000
-3000
-2000
-1600
-1000
-800
-500
-400
-200
350
400
500
600
700
800
900
1000
1300
1340
1400
1500
1600
1650
1700
1750
1800
1850
1875
1900
1920
1925
1930
1940
1950
1955
19
1965
197
1975
1980
1985
1990
10000
5000
1000
500
100
50
Years before 2020



The first graph is the same one you saw in the last section, of absolute GWP doubling times. The second graph is the same, but limited to Britain.

Where’s the Industrial Revolution?

It doesn’t show up at all. This may be a surprise if you’re used to the standard narrative where the Industrial Revolution was the most important event in economic history. Graphs like this make the case that the Industrial Revolution was an explosive shift to a totally new growth regime:


World GDP over the last two millennia

Total output of the world economy; adjusted for inflation and expressed in 2011 international dollars.

Source: World GDP - Our World In Data based on World Bank & Maddison (2017)

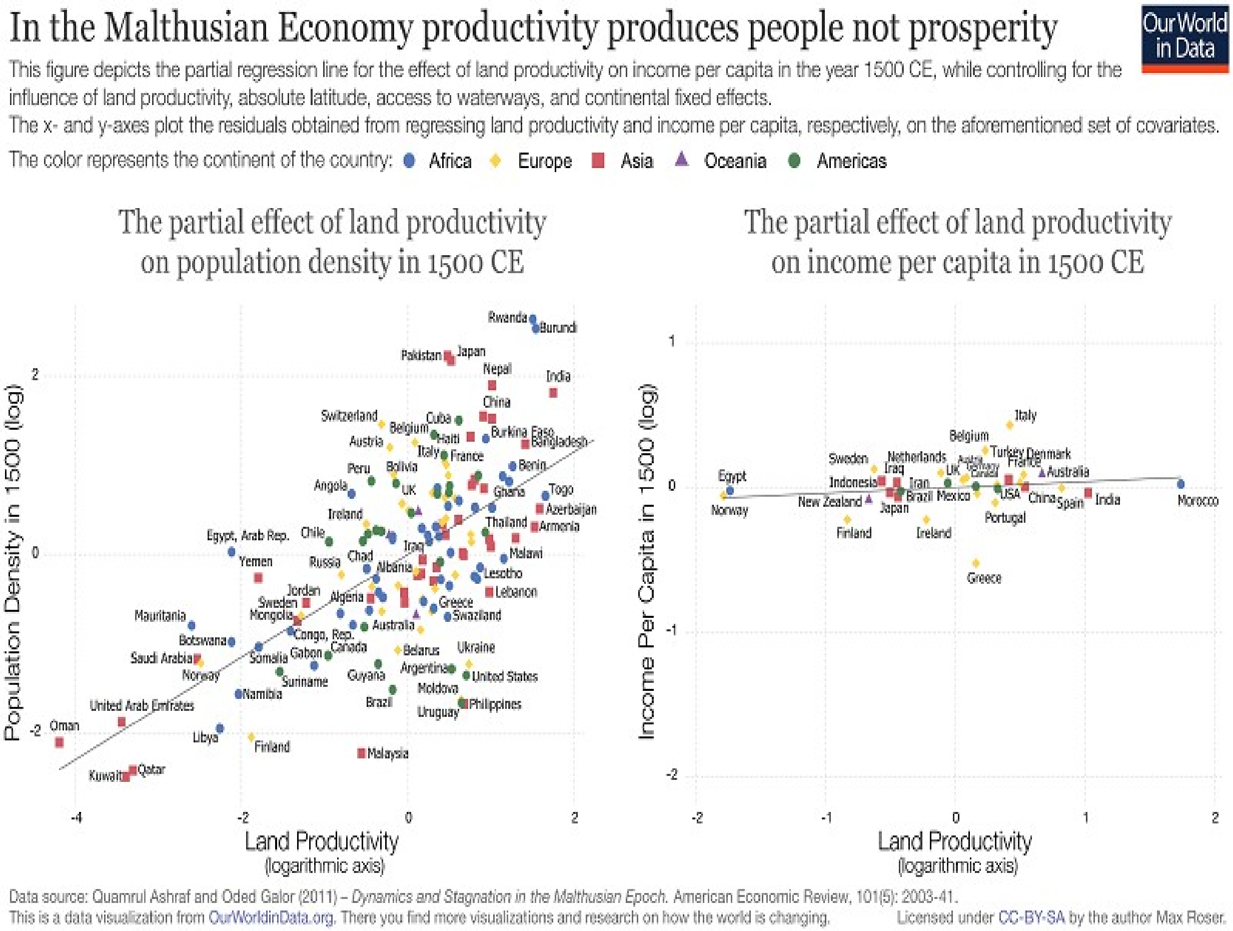
OurWorldInData.org/economic-growth • CC BY-SA

It sure *looks* like the Industrial Revolution was a big deal. But Paul Christiano argues your eyes may be deceiving you. That graph is a hyperbola, ie corresponds to a single simple equation. There is no break in the pattern at any point. If you transformed it to a log doubling time graph, you'd just get the graph above that looks like a straight line until 1960.

On this view, the Industiral Revolution didn't change historical GDP trends. It just shifted the world from a Malthusian regime where economic growth increased the population to a modern regime where economic growth increased per capita income.

For the entire history of the world until 1000, GDP per capita was the same for everyone everywhere during all historical eras. An Israelite shepherd would have had about as much stuff as a Roman farmer or a medieval serf.

This was the Malthusian trap, where “productivity produces people, not prosperity”. People reproduce to fill the resources available to them. Everyone always lives at subsistence level. If productivity increases, people reproduce, and now you have more people living at subsistence level. OurWorldInData has an awesome graph of this:



In the Malthusian Economy productivity produces people not prosperity

This figure depicts the partial regression line for the effect of land productivity on income per capita in the year 1500 CE, while controlling for the influence of land productivity, absolute latitude, access to waterways, and continental fixed effects.
The x- and y-axes plot the residuals obtained from regressing land productivity and income per capita, respectively, on the aforementioned set of covariates.
The color represents the continent of the country: ● Africa ◆ Europe ■ Asia ▲ Oceania ● Americas

The partial effect of land productivity on population density in 1500 CE

The partial effect of land productivity on income per capita in 1500 CE

Our World in Data
Population Density in 1500 (log)
2
0
-2
-4
-2
0
2
Land Productivity
(logarithmic axis)
Rwanda
Burundi
Pakistan
Japan
Nepal
India
China
Switzerland
Belgium
Cuba
Burkina Faso
Bangladesh
Austria
Haiti
Italy
France
Peru
Bolivia
Benin
Angola
UK
Ghana
Togo
Azerbaijan
Ireland
Thailand
Armenia
Egypt, Arab Rep.
Chile
Iraq
Malawi
Yemen
Russia
Chad
Albania
Lesotho
Jordan
Lebanon
Sweden
Algeria
Greece
Mongolia
Swaziland
Mauritania
Australia
Congo, Rep.
Botswana
Canada
Gabon
Ukraine
Saudi Arabia
Somalia
Belarus
Norway
Argentina
United States
Suriname
Guyana
Namibia
Moldova
United Arab Emirates
Brazil
Philippines
Uruguay
Oman
Libya
Finland
Malaysia
Kuwait
Qatar
Income Per Capita in 1500 (log)
1
0
-1
-2
-2
-1
0
1
2
Land Productivity
(logarithmic axis)
Italy
Belgium
Turkey
Denmark
Sweden
Netherlands
France
Australia
Egypt
Iraq
UK
Germany
Canada
Indonesia
Iran
Norway
New Zealand
Japan
Brazil
Mexico
USA
China
Spain
India
Morocco
Portugal
Finland
Ireland
Greece

Data source: Quamrul Ashraf and Oded Galor (2011) – *Dynamics and Stagnation in the Malthusian Epoch*. American Economic Review, 101(5): 2003-41.
This is a data visualization from OurWorldinData.org. There you find more visualizations and research on how the world is changing. Licensed under CC-BY-SA by the author Max Roser.

As of 1500, places with higher productivity (usually richer farmland, but better technology and social organization also help) population density is higher. But GDP per capita was about the same everywhere.

There were always occasional windfalls from exciting discoveries or economic reforms. For a century or two, GDP per capita would rise. But population would always catch up again, and everyone would end up back at subsistence.

Some people argue Europe broke out of the Malthusian trap around 1300. This is not quite right. 1300s Europe achieved above-subsistence GDP, but only because the Black Plague killed so many people that the survivors got a windfall by taking their land.

Malthus predicts that this should only last a little while, until the European population bounces back to pre-Plague levels. This prediction was exactly right for Southern Europe. Northern Europe didn’t bounce back. Why not?

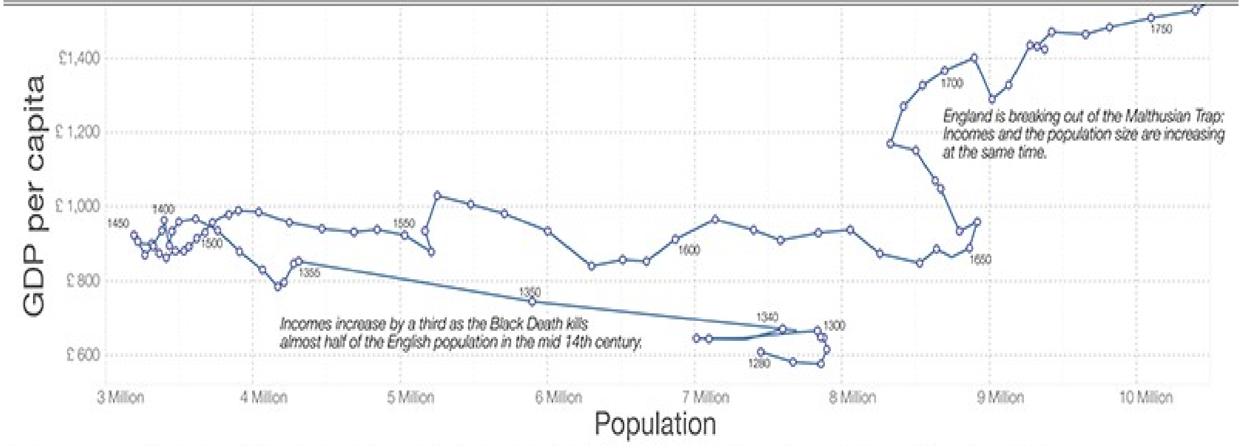
Unclear, but one answer is: fewer people, more plagues.

Broadberry 2015 mentions that Northern European culture promoted later marriage and fewer children:

> The North Sea Area had an advantage in this area because of its approach to marriage. Hajnal (1965) argued that northwest Europe had a different demographic regime from the rest of the world, characterised by later marriage and hence limited fertility. Although he originally called this the European Marriage Pattern, later work established that it applied only to the northwest of the continent. This can be linked to the availability of labour market opportunities for females, who could engage in market activity before marriage, thus increasing the age of first marriage for females and reducing the number of children conceived (de Moor and van Zanden, 2010). Later marriage and fewer children are associated with more investment in human capital, since the womenemployed in productive work can accumulate skills, and parents can afford to

> invest more in each of the smaller number of children because of the “quantity-quality” trade-off (Voigtländer and Voth, 2010).

This low birth rate was happening at the same time plagues were raising the death rate. Here’s another amazing graph from OurWorldInData:



Data source: Broadberry et al and Bank of England. Data prior to 1700 refers to England; data thereafter refers to the UK. Averages over 5 year intervals are shown here.
The visualization is available at OurWorldinData.org where you find more visualizations and research on global development. Licensed under CC-BY-SA by the author Max Roser.

British population maxes out around 1300 (?), declines substantially during the Black Plague of 1348-49, but then keeps declining. The List Of English Plagues says another plague hit in 1361, then another in 1369, then another in 1375, and so on. Some historians call the whole period from 1348 to 1666 “the Plague Years”.

It looks like through the 1350 – 1450 period, population keeps declining, and per capita income keeps going up, as Malthusian theory would predict.

Between 1450 and 1550, population starts to recover, and per capita incomes start going down, again as Malthus would predict. Then around 1560, there’s a jump in incomes; according to the List Of Plagues, 1563 was “probably the worst of the great metropolitan epidemics, and then extended as a major national outbreak”. After 1563, population increases again and per capita incomes decline again, all the way until 1650. Population does not increase in Britain at all between 1660 and 1700. Why? The List declares 1665 to be “The Great Plague”, the largest in England since 1348.

So from 1348 to 1650, Northern European per capita incomes diverged from the rest of the world’s. But they didn’t “break out of the Malthusian trap” in a strict sense of being able to direct production toward prosperity rather than population growth. They just had so many plagues that they couldn’t grow the population anyway.

But in 1650, England did start breaking out of the Malthusian trap; population and per capita incomes grow together. Why?

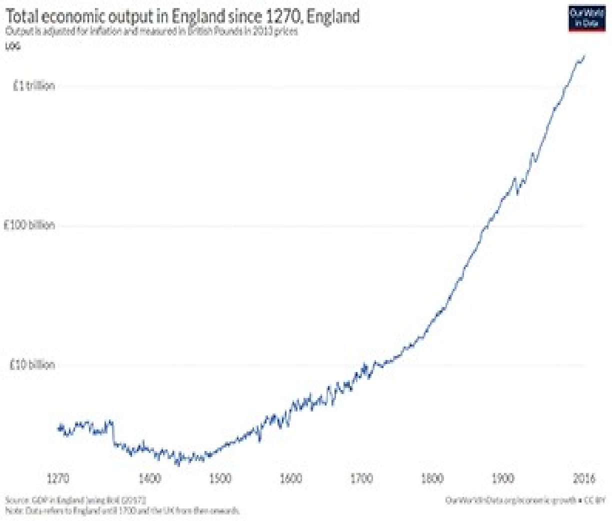
Paul theorizes that technological advance finally started moving faster than maximal population growth.

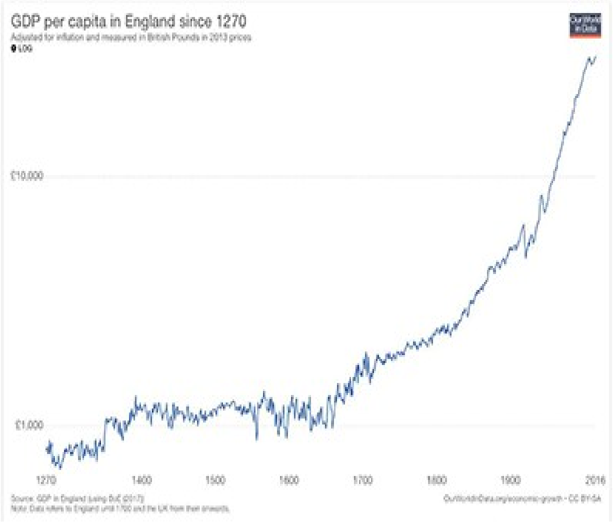
Remember, in the von Foerster model, the growth rate increases with time, all the way until it reaches infinity in 2026. The closer you are to 2026, the faster your economy will grow. But population can only grow at a limited rate. In the absolute limit, women can only have one child per nine months. In reality, infant mortality, infertility, and conscious decision to delay childbearing mean the natural limits are much lower than that. So there’s a theoretical limit on how quickly the population can increase even with maximal resources. If the economy is growing faster than that, Malthus can’t catch up.

Why would this happen in England and Holland in 1650?

Lots of people have historical explanations for this. Northern European population growth was so low that people were forced to invent labor-saving machinery; eventually this reached a critical mass, we got the Industrial Revolution, and economic growth skyrocketed. Or: the discovery of America led to a source of new riches and a convenient sink for excess population. Or: something something Protestant work ethic printing press capitalism. These are all plausible. But how do they sync with the claim that absolute GDP never left its expected trajectory?

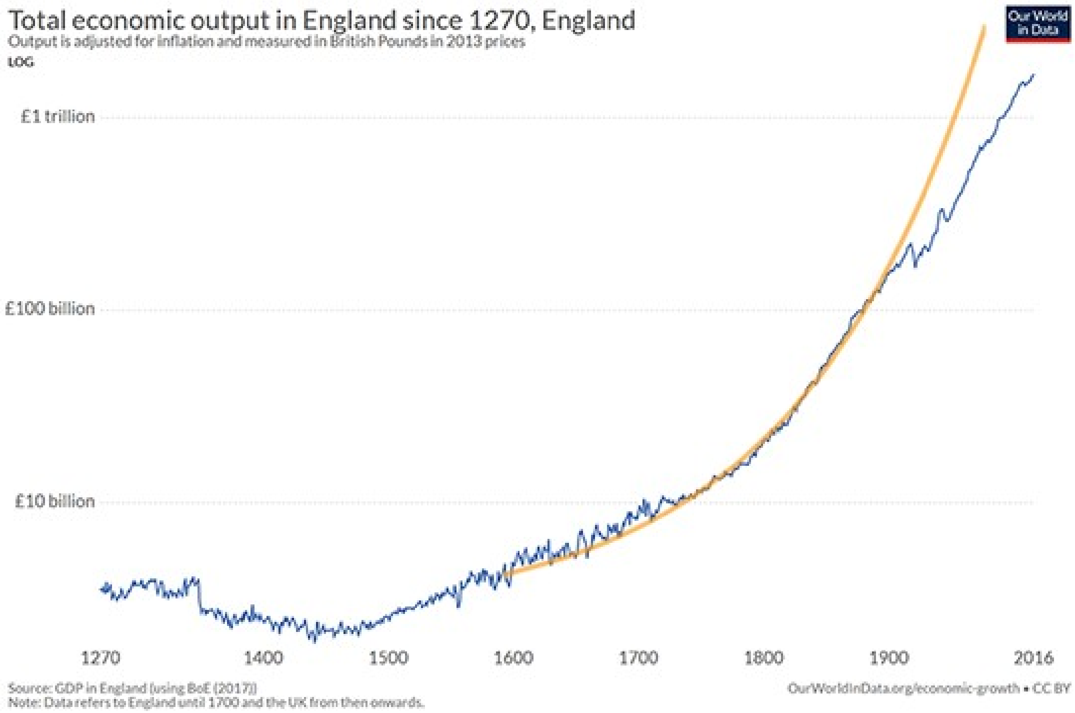
I find the idea that the Industrial Revolution wasn't a deviation from trend fascinating and provocative. But it depends on eyeballing a lot of graphs that have had a lot of weird transformations done to them, plus writing off a lot of outliers. Here's another way of presenting Britain's GDP and GDP per capita data:





Here it's a lot less obvious that the Industrial Revolution represented a deviation from trend for GDP per capita but not for GDP.

These British graphs show less of a singularity signature than the worldwide graphs do, probably because we're looking at them on a shorter timeline, and because the Plague Years screwed everything up. If we insisted on fitting them to a hyperbola, it would look like this:


Total economic output in England since 1270, England

Output is adjusted for inflation and measured in British Pounds in 2013 prices

Source: GDP in England (using BoE (2017))
Note: Data refers to England until 1700 and the UK from then onwards.

OurWorldInData.org/economic-growth • CC BY

Like the rest of the world, Britain was only on a hyperbolic growth trajectory when economic growth was translating into population growth. That wasn't true before about 1650, because of the plague. And it wasn't true after about 1850, because of the Demographic Transition. We see a sort of fit to a hyperbola between those points, and then the trend just sort of wanders off.

It seems possible that the Industrial Revolution was not a time of abnormally fast technological advance or economic growth. Rather, it was a time when economic growth outpaced population growth, causing a shift from a Malthusian regime where productivity growth always increased population at subsistence level, to a modern regime where productivity growth increases GDP per capita. The world remained on the same hyperbolic growth trajectory throughout, until the trajectory petered out around 1900 in Britain and around 1960 in the world as a whole.

IV.

So just how cancelled is the singularity?

To review: population growth increases technological growth, which feeds back into the population growth rate in a cycle that reaches infinity in finite time.

But since population can't grow infinitely fast, this pattern breaks off after a while.

The Industrial Revolution tried hard to compensate for the "missing" population; it invented machines. Using machines, an individual could do an increasing amount of work. We can imagine making eg tractors as an attempt to increase the effective population faster than the human uterus can manage. It partly worked.

But the industrial growth mode had one major disadvantage over the Malthusian mode: tractors can't invent things. The population wasn't just there to grow the population, it was there to increase the rate of technological advance and thus population growth. When we shifted (in part) from making people to making tractors, that process broke down, and growth (in people *and* tractors) became sub-hyperbolic.

If the population stays the same (and by "the same", I just mean "not growing hyperbolically") we should expect the growth rate to stay the same too, instead of increasing the way it did for thousands of years of increasing population, modulo other concerns.

In other words, the singularity got cancelled because we no longer have a surefire way to convert money into researchers. The old way was more money = more food = more population = more researchers. The new way is just more money = send more people to college, and screw all that.

But AI potentially offers a way to convert money into researchers. Money = build more AIs = more research.

If this were true, then once AI comes around – even if it isn't much smarter than humans – then as long as the computational power you can invest into researching a given field increases with the amount of money you have, hyperbolic growth is back on. Faster growth rates means more money means more AIs researching new technology means even faster growth rates, and so on to infinity.

Presumably you would eventually hit some other bottleneck, but things could get very strange before that happens.

Buspirone Shortage In Healthcaristan SSR

*(**Epistemic status**: Unsure on details. Some post-publication edits 5/1 to make this less strident.)*

I.

There is a national shortage of buspirone.

Buspirone is a 5HT-1 agonist used to control anxiety. Unlike most psychiatric drugs, it's in a class of its own – there are no other sole 5HT-1 agonists on the market. It's not a very strong medication, but it's safe, it's non-addictive, it's off-patent, and it works well for a subset of patients. Some of them have been on it for years.

Now there's a national shortage. My patients can't get it, or have to go hunting from pharmacy to pharmacy until they find one that has it. I've told people find a source to stockpile a supply so they don't run out. It feels like we're living in the Soviet Union.

How did this happen? The *New York Times* writes:

> The main reason for the buspirone shortage appears to be interrupted production at a Mylan Pharmaceuticals plant in Morgantown, W.Va., which produced about a third of the country's supply of the drug. The Food and Drug Administration had said the facility was dirty and that the company failed to follow quality control procedures.

So the FDA shut down a major buspirone factory. But government agencies – ones that are a lot less nice than the FDA – shut down methamphetamine factories all the time without creating methamphetamine shortages. Why is the buspirone market so vulnerable? The *Times* again:

> Rock bottom prices for some generic drugs are also contributing to the crisis. Consolidation among wholesalers has led to the creation of three buying consortium behemoths that purchase 90 percent of the generic pharmaceutical products in the United States, said Adam Fein, a consultant and chief executive of Drug Channels Institute. These “monster” buyers have squeezed manufacturers on prices, and “some of those generic manufacturers are deciding the profit is so low they can’t make money, and they’re exiting the category,” Dr. Fein said.

Is this really how economics works? There’s a medicine that millions of people desperately need? But nobody will produce it because they can’t make a profit? Huh? Isn’t the usual solution to just raise the price? And people will buy it at the higher price, because they need it so badly? And then you will make more profit, and can keep on making the medication? Isn’t “nobody will supply this product, it’s too cheap” just the economics version of “nobody goes there anymore, it’s too crowded”?

Sure, generic drug manufacturing is pretty consolidated. Most individual generic drugs are now manufactured only by one or two companies. If one of those few companies gets greedy (like Martin Shkreli did with Daraprim), they can increase prices by orders of magnitude without a lot of competitors to push back. And if one of those few companies suffers a shock (like the FDA closing the buspirone factory), it makes sense that there might not be enough competitors to pick up the slack.

But how come this is only happening in pharmaceuticals? How come (in capitalist countries) there are almost never meat shortages, bread shortages, laptop shortages, or chair shortages? Is there something unusual about the pharmaceutical landscape that predisposes it to this sort of thing?

I am not an expert in this area and may be getting some of it wrong. But from Berndt, Conti, and Murphy (2017) and a Berndt, Conti, and Murphy (2018), I gather that a big part of the story is the Generic Drug User Fee Amendments (GDUFA) of 2012 and 2017.

The story goes something like this. The FDA demanded that generic drug manufacturers pass FDA inspection before setting up shop. But the FDA didn't have enough inspectors to review manufacturers in a timely manner. So companies kept asking the FDA for permission to enter the generics market, and the FDA kept telling them there was a several year waiting period. In 2012, Congress recognized the problem. Politicians, FDA officials, and industry leaders agreed on a new policy where generic drug manufacturing companies would pay the FDA lots of money (about $300 million last time anyone checked), and the FDA would use that money to hire inspectors so they could clear their backlog of applications.

The good news is, the FDA hired lots more inspectors and they are now pretty good at responding to generic drug applications in a timely way. The bad news is that the fees to the companies were designed in a way that subtly encouraged monopolies in generic drug markets. I don't understand all the specifics, but there seem to be two main problems.

First, if you manufacture a drug, the FDA will charge you a fee, but the fee doesn't scale linearly with how much of the drug you produce. So suppose Martin Shkreli owns a very big Daraprim factory. The FDA might charge him $1 million per year to fund their inspectors. Suppose you are a small businessman who is angry at Martin Shkreli's fee hike, and you want to open a competing Daraprim factory in your small town, using your small amount of personal savings. Probably your factory will be much smaller than Martin Shkreli's. But the FDA will still charge you the same $1 million per year. At worst this means you make no profit; even at best, Shkreli's economy of scale gives him a big advantage over you. So you may decide not to enter the market at all. From the second paper:

> President of the Pharma & Biopharma Outsourcing Association, Gil Roth, remarked, 'We have a single generic client that we do a short run of production for. Why are we charged the same as a Teva facility that pumps out a billion tablets?' Another commented, 'At least a flat tax is based on a percentage, either of revenue or profit. This is a flat fee, which makes it a regressive tax on smaller businesses, both contract manufacturers and small generics companies'

I think the fee might even be per factory, which encourages companies to concentrate all their manufacturing at a single site – like the Mylan one that just got shut down, thus affecting the whole country's buspirone supply.

Second, traditional economics suggests that if some company has a monopoly on a product that people really need (like a medication), they will charge very high prices. But many generic drugs are produced by only one company each – and Shkrelis aside, most of them charge affordable prices. Why? Berndt et al argue it is because of the *possibility* of competition: if Shkreli raises his prices too high, some other company can move in and undercut him. But FDA licensing procedures make this undercutting harder than it could be: it will take months to years, and thousands to millions of dollars, for the other company to move in (at which point Shkreli can just say “Haha, no” and lower his prices again, meaning the undercutter would lose all the money they put in).

Historically, the system has worked anyway – because lots of companies are sitting on pre-existing FDA approval to make certain drugs. If a company had ever made a drug in the past, they had FDA permission to make it again whenever they wanted. So if Shkreli raises prices on Daraprim, some other company that made Daraprim ten years ago can set up a new factory tomorrow and undercut him. This helped prevent would-be Shkrelis in most markets, and provided a safety valve for shocks like the one creating the buspirone shortage today.

But GDUFA weakened this system by mandating that any company with FDA approval to manufacture a drug pay yearly inspection fees to the FDA, whether or not they were actively manufacturing it. That turned FDA approval for drugs you weren’t actively manufacturing into a liability; you were paying fees, but not making a profit. Companies started voluntarily cancelling their FDA approvals for older drugs so they wouldn’t have to pay the fees. That meant monopolists lost a lot of their potential competition. And that cleared the way for people like Shkreli to hike prices.

You get more of what you subsidize and less of what you tax. Unfortunately, the FDA is inadvertently taxing companies for being in the generic drug business. And it’s taxing them more if they’re not a monopolist with economies of scale. That means we get fewer companies in the generics industry, and more monopolists.

So my very tentative guess as to why buspirone is more plagued by shortages than bread or chairs is because number one, the need for FDA approval makes it hard for new companies to enter the buspirone industry, and number two, the FDA's fee structure favors large-scale monopolies over small-scale competitors.

II.

The price of insulin is much too high. Vox argues that this is because of the "lax regulatory environment" and the "free market approach", and that if we could just become socialist like all of the cool countries, everything would be fine.

Insulin is off-patent. It was discovered almost a hundred years ago. But somehow, all the insulin sold in the US is brand-name. This is shocking and obviously the root of the problem. What's going on? Vox links NEJM's *Why Is There No Generic Insulin?*, but summarizes it by saying it's "because companies have made those incremental improvements to insulin products, which has allowed them to keep their formulations under patent" and because "older insulin formulations have fallen out of fashion."

I am not diabetic. But if I were, I don't think I would worry that much which kinds of insulin were vs. weren't fashionable. What's really going on?

Here my source is partly the NEJM paper above, but also Health Affairs' Biologics Are Natural Monopolies. Both agree that the key point is insulin's nature as a "biologic". It's not a simple molecule you can make with a chemistry set. It's a complex peptide hormone of about seven hundred atoms, arranged in a series of helices and threads and tentacles.

The only way to manufacture it is to genetically engineer some microorganism to make it for you.

The FDA usually requires generic manufacturers to prove that their drug is identical to the brand name drug they're copying. But genetic engineering is hard, microorganisms are uncooperative, and insulin is too complicated to say with certainty that any one insulin molecule is "identical" to any other. So the FDA has lowered their standards for biologics to require proof that a generic biological is "similar".

But even proving biosimilarity is orders of magnitude tougher than anything that small molecules have to go through. From the Health Affairs article, slightly transposed for readability:

> The typical [small molecule] generic drug takes firms 1-3 years, $1-$5 million, and no human clinical trials to introduce. [In contrast], the entire biosimilar development process has been projected to span 8-10 years and cost upwards of $100 million. Human clinical trials involving hundreds of patients can cost $20-40 million to simply confirm that the candidate biosimilar generally replicates the reference product's short-term positive and negative clinical effects.

Brand-name insulin companies make a bad situation worse by patenting their manufacturing techniques, using different patents than the drug patent, which may still be in effect when a generics manufacturer is trying to copy their drug. For example, Sanofi has somehow managed to get 74 different patents on their Insulin Lantus, which this I-MAK report describes as a "patent thicket". Many of these patents seem to be totally illegal, and exist only so that it would cost a generics company time and money to challenge them in court. Most generics companies look at the process of trying to prove their hideously complex molecule is “biosimilar” to Sanofi’s hideously complex molecule, without using any of the 74 different manufacturing processes Sanofi uses to make it, and decide against entering the market.

Then the FDA mandates that biosimilars have a different name than the product they are replacing (ordinary generic drugs may not use the trade name, but can use the same chemical name). This makes it harder to have prescriptions for one cover the other, and doctors may have too much inertia to switch to a new drug with a new name. This limits potential sales for these products.

So the reason companies aren’t making generic insulin is that the FDA approval process for generic insulin is very onerous, brand name companies have excessive and illegal patents that make the approval process even worse, and companies’ ability to sell what comes out the other end is limited.

I realize my political slant makes me blame poor regulatory choices for these sorts of things pretty often. And the Health Affairs article I’m drawing from makes a different argument than I do, arguing that their biological properties make insulins “natural monopolies” and that policy choices are only secondary to this. You should consider my biases before you necessarily take my words at face value.

But the NEJM article mentions that plenty of poorer countries *do* have biosimilar generic insulins, including such gleaming-high-tech bastions of cutting-edge pharmaceutical excellence as Peru. Which of the following do you think is true?:

1. Peru has better technology than the US, and so is able to make cheap biosimilar insulin using processes that our own scientists and engineers can't manage.

2. Peru has a bigger market than the US, so there's more money in creating generic insulin to sell to Peruvians than there is selling it to Americans.

3. Peru has a better regulatory environment than the US, and this is enough to make producing biosimilar insulins cheap and easy.

Extreme fringe libertarians have a certain way with words. For example, they call taxes "the government stealing money from you at gunpoint". This is a little melodramatic, but words like "patent loopholes" and "onerous review processes" sound a little bloodless for something that probably kills thousands of diabetics each year. So I would like to take a page from the extreme libertarian lexicon and speculate that the problem with insulin costs is that the government will shoot anyone who tries to make cheap insulin.

And then Vox writes an article saying that the problem is "the free market" and we need more government intervention. Fine, whatever. I have despaired of anyone ever analyzing this topic in any greater depth than that. The drug situation is going to keep getting worse – for small molecules, for biosimilars, for whatever. People are going to keep blaming "the free market" and implementing more poorly-thought-out regulation. And so the cycle will keep going, ad infinitum. Vox will keep writing this article once a year or so, and I'll keep telling them they are bad and wrong, and we'll both get some clicks out of it. The system works!

I want to clarify that I’m not criticizing the current FDA administration. The FDA has recently done a great job trying to shift their processes marginally in the direction of approving more medications, approving more companies entering the generic market, promoting more competition, and generally doing everything right. Even the GDUFA was a step in the right direction, in that it was necessary in order to get generics approved at all. This is probably part of why drug prices are starting to drop (note that there’s a complicated debate over how true this is and what statistics to use, but I think even the skeptics agree the trend is positive, and they are rising less quickly than they have in the past). There are probably some small steps they could still make to improve things – I get the impression that having the government pay for FDA inspections using tax dollars instead of having the distortionary GDUFA system would help. And patent reform would be great. But a lot of this is concessions to political reality that are probably outside the FDA’s control.

The current trends are good, and further small fixes could be better, but they probably aren’t enough to make drugs affordable and consistently available to patients. If this is even possible, it’s going to require more dramatic changes – not just having good regulators who try to make the best of the current system, but reforming the system entirely. This is a tough order (and I’ll try to blog later about what might be involved). But it’s the only thing that I can imagine allowing us to eventually catch up to Peru.

Update To Partial Retraction Of Animal Value And Neuron Number

A few weeks ago I published results of a small (n = 50) survey showing that people's moral valuation of different kinds of animals scaled pretty nicely with the animals' number of cortical neurons (see here for more on why we might expect that to be true).

A commenter, Tibbar, did a larger survey on Mechanical Turk and got very different results, so I retracted the claim. I wasn't sure why we got such different results, but I chalked it down to chance, or perhaps to my having surveyed an animal-rights-conscious crowd who thinks a lot about this kinds of things vs. Tibbar surveying random MTurkers.

Now David Moss, from effective altruist organization Rethink Priorities, has looked into this more deeply and resolved some of the discrepancies.

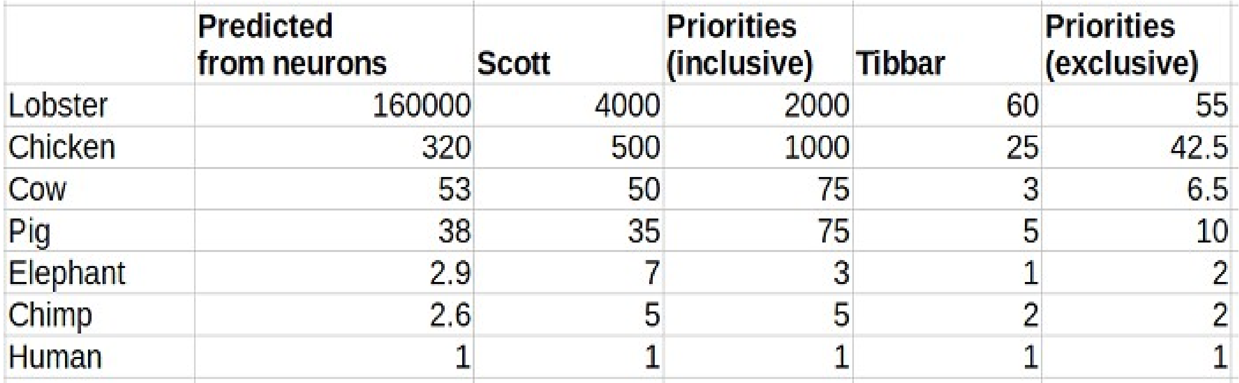
The problem is that I did a terrible job explaining my procedure (I linked to the form I used, but the link was broken when Tibbar did his survey). In particular, I included the line:

> If you believe [animals have moral value] in general, but think some specific animal I ask about doesn't work this way, feel free to leave the question blank or put in "99999", which I will interpret as "basically infinity"

About 5 – 10% of respondents took me up on this. Tibbar didn't make this suggestion, and none of his participants did this.

Moss surveyed 490 people on Mechanical Turk, and did not offer people a “basically infinity” option. However, many (20% – 40%) of his participants said the question didn’t apply to specific animals.

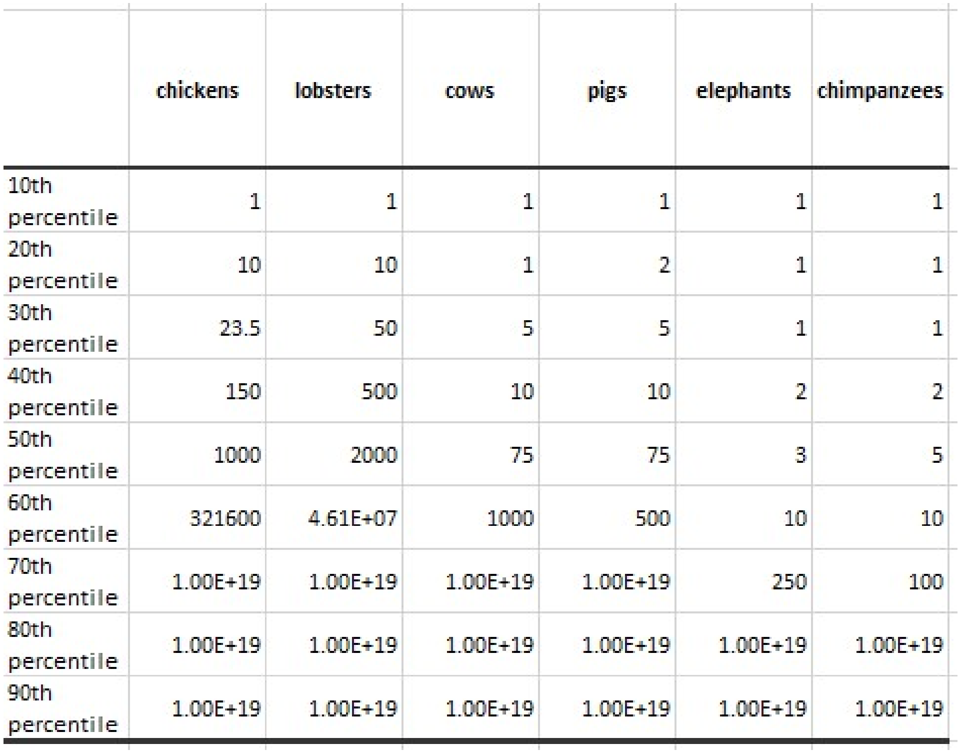
He found that when he ignored these, he got the same (low) numbers as Tibbar; when he counted an N/A answer as a vote for “basically infinity”, he got the same (high) numbers that I did.



	Predicted from neurons	Scott	Priorities (inclusive)	Tibbar	Priorities (exclusive)
Lobster	160000	4000	2000	60	55
Chicken	320	500	1000	25	42.5
Cow	53	50	75	3	6.5
Pig	38	35	75	5	10
Elephant	2.9	7	3	1	2
Chimp	2.6	5	5	2	2
Human	1	1	1	1	1

This graph measures people’s perceptions of how many of each animal is morally equivalent to one human. “Priorities (inclusive)” is Rethink Priorities’ survey, with refusal to vote counted as “basically infinity”. “Priorities (exclusive)” is Rethink Priorities’ survey, with refusals to vote thrown out. I think this demonstrates pretty well that you can get either mine or Tibbar’s numbers depending on which choice you make.

But Moss also points out that all of this is just a fragile balance between people who say every life is worth the same regardless of species, versus people who just spam the box with as many nines as they can.



	chickens	lobsters	cows	pigs	elephants	chimpanzees
10th percentile	1	1	1	1	1	1
20th percentile	10	10	1	2	1	1
30th percentile	23.5	50	5	5	1	1
40th percentile	150	500	10	10	2	2
50th percentile	1000	2000	75	75	3	5
60th percentile	321600	4.61E+07	1000	500	10	10
70th percentile	1.00E+19	1.00E+19	1.00E+19	1.00E+19	250	100
80th percentile	1.00E+19	1.00E+19	1.00E+19	1.00E+19	1.00E+19	1.00E+19
90th percentile	1.00E+19	1.00E+19	1.00E+19	1.00E+19	1.00E+19	1.00E+19

So it’s not clear how much we should be drawing from this in any case. Whatever. I still think it’s neat.

The moral of the story is to always explain your procedures really well before you try to replicate something. Also to make sure the link to your procedures *actually* goes to your procedures, although maybe no one other than me has ever made that specific mistake before.

Little Known Types Of Eclipse

A lunar eclipse occurs when the Earth gets between the Moon and the Sun.

A solar eclipse occurs when the Moon gets between the Earth and the Sun.

A terrestrial eclipse occurs when the Earth gets between you and the Sun. Happens once per 24 hours.

An atmospheric eclipse occurs when an asteroid gets between you and the sky. Generally fatal.

A reverse solar eclipse occurs when the Sun gets between the Moon and the Earth. Extremely fatal.

A motivational eclipse occurs when the Moon gets between you and your goals. You can’t let it stop you! Destroy it! Destroy the Moon!

A marital eclipse occurs when the Moon gets between you and your spouse. You’re going to need to practice good communication about the new celestial body in your life if you want your relationship to survive.

A capillary eclipse occurs when your hair gets between your eyes and the Sun. Get a haircut.

A lexicographic eclipse occurs when “Moon” gets between “Earth” and “Sun” in the dictionary. All Anglophone countries are in perpetual lexicographic eclipse.

A filioque eclipse occurs when the Holy Spirit gets between the Father and the Son.

An apoc eclipse occurs when the Great Beast 666, with seven heads and ten horns, and upon the horns ten crowns, and upon its heads the name of blasphemy, gets between the Earth and the Sun. *Extremely* fatal.

5-HTTLPR: A Pointed Review

In 1996, some researchers discovered that depressed people often had an unusual version of the serotonin transporter gene 5-HTTLPR. The study became a psychiatric sensation, getting thousands of citations and sparking dozens of replication attempts (page 3 here lists 46).

Soon scientists moved beyond replicating the finding to trying to elucidate the mechanism. Seven studies (see here for list) found that 5-HTTLPR affected activation of the amygdala, a part of the brain involved in processing negative stimuli. In one especially interesting study, it was found to bias how the amygdala processed ambiguous facial expression; in another, it modulated how the emotional systems of the amygdala connected to the attentional systems of the anterior cingulate cortex. In addition, 5-HTTLPR was found to directly affect the reactivity of the HPA axis, the stress processing circuit leading from the adrenal glands to the brain.

As interest increased, studies began pointing to 5-HTTLPR in other psychiatric conditions as well. One study found a role in seasonal affective disorder, another in insomnia. A meta-analysis of twelve studies found a role ($p = 0.001$) in PTSD. A meta-analysis of twenty-three studies found a role ($p = 0.000016$) in anxiety-related personality traits. Even psychosis and Alzheimer’s disease, not traditionally considered serotonergic conditions, were affected. But my favorite study along these lines has to be 5-HTTLPR Polymorphism Is Associated With Nostalgia-Proneness.

Some people in bad life situations become depressed, and others seem unaffected; researchers began to suspect that genes like 5-HTTLPR might be involved not just in causing depression directly, but in modulating how we respond to life events. A meta-analysis looked at 54 studies of the interaction and found "strong evidence that 5-HTTLPR moderates the relationship between stress and depression, with the s allele associated with an increased risk of developing depression under stress (P = .00002)". This relationship was then independently re-confirmed for every conceivable population and form of stress. Depressed children undergoing childhood adversity. Depressed children with depressed mothers. Depressed youth. Depressed adolescent girls undergoing peer victimization. They all developed different amounts of depression based on their 5-HTTLPR genotype. The mainstream media caught on and dubbed 5-HTTLPR and a few similar variants "orchid genes", because orchids are sensitive to stress but will bloom beautifully under the right conditions. Stories about "orchid genes" made it into *The Atlantic*, *Wired*, and *The New York Times*.

If finding an interaction between two things is exciting, finding an interaction between even more things must be even better! Enter studies on how the interaction between 5-HTTLPR and stress in depressed youth itself interacted with MAO-A levels and gender. What about parenting style? Evidence That 5-HTTLPR x Positive Parenting Is Associated With Positive Affect "For Better And Worse" What about decision-making? Gender Moderates The Association Between 5-HTTLPR And Decision-Making Under Uncertainty, But Not Under Risk. What about single motherhood? The influence of family structure, the TPH2 G-703T and the 5-HTTLPR serotonergic genes upon affective problems in children aged 10–14 years. What if we just throw all the interesting genes together and see what happens? Three-Way Interaction Effect Of 5-HTTLPR, BDNF Val66Met, And Childhood Adversity On Depression.

If 5-HTTLPR plays such an important role in depression, might it also have relevance for antidepressant treatment? A few studies of specific antidepressants started suggesting the answer was yes – see eg 5-HTTLPR Predicts Non-Remission In Major Depression Patients Treated With Citalopram and Influence Of 5-HTTLPR On The Antidepressant Response To Fluvoxamine In Japanese Depressed Patients. A meta-analysis of 15 studies found that 5-HTTLPR genotype really did affect SSRI efficacy (p = 0.0001). Does this mean psychiatrists should be testing for 5-HTTLPR before treating patients? A cost-effectiveness analysis says it does. There's only one problem.

ALL.

OF.

THIS.

IS.

LIES.

Or at least this is the conclusion I draw from Border et al's No Support For Historical Candidate Gene Or Candidate Gene-by-Interaction Hypotheses For Major Depression Across Multiple Large Samples, in last month's *American Journal Of Psychiatry*.

They don't ignore the evidence I mention above. In fact, they count just how much evidence there is, and find 450 studies on 5-HTTLPR before theirs, most of which were positive. But they point out that this doesn't make sense given our modern understanding of genetics. Outside a few cases like cystic fibrosis, most important traits are massively polygenic or even omnigenic; no one gene should be able to have measurable effects. So maybe this deserves a second look.

While psychiatrists have been playing around with samples of a few hundred people (the initial study “discovering” 5-HTTLPR used n = 1024), geneticists have been building up the infrastructure to analyze samples of hundreds of thousands of people using standardized techniques. Border et al focus this infrastructure on 5-HTTLPR and its fellow depression genes, scanning a sample of 600,000+ people and using techniques twenty years more advanced than most of the studies above had access to. They claim to be able to simultaneously test almost every hypothesis ever made about 5-HTTLPR, including “main effects of polymorphisms and genes, interaction effects on both the additive and multiplicative scales and, in G3E analyses, considering multiple indices of environmental exposure (e.g., traumatic events in childhood or adulthood)”. What they find is…nothing. Neither 5-HTTLPR nor any of seventeen other comparable “depression genes” had any effect on depression.

I love this paper because it is ruthless. The authors know exactly what they are doing, and they are clearly enjoying every second of it. They explain that given what we now know about polygenicity, the highest-effect-size depression genes require samples of about 34,000 people to detect, and so any study with fewer than 34,000 people that says anything about specific genes is almost definitely a false positive; they go on to show that the median sample size for previous studies in this area was 345. They show off the power of their methodology by demonstrating that negative life events cause depression at p = 0.000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000

0001, because it’s pretty easy to get a low p-value in a sample of 600,000 people if an effect is real. In contrast, the gene-interaction effect of 5-HTTLPR has a p-value of .919, and the main effect from the gene itself *doesn’t even consistently point in the right direction*. Using what they call “exceedingly liberal significance thresholds” which are 10,000 times easier to meet than the usual standards in genetics, they are unable to find any effect. This isn’t a research paper. This is a *massacre*.

Let me back off a second and try to be as fair as possible to the psychiatric research community.

First, over the past fifteen years, many people within the psychiatric community have been sounding warnings about 5-HTTLPR. The first study showing failure to replicate came out in 2005. A meta-analysis by Risch et al from 2009 found no effect and prompted commentary saying that 5-HTTLPR was an embarrassment to the field. After 2010, even the positive meta-analyses (of which there were many) became guarded, saying only that they seemed to detect an effect but weren’t sure it was real. This meta-analysis on depression says there is “a small but statistically significant effect” but that “we caution it is possible the effect has an artifactual basis”. This meta-analysis of 5-HTTLPR amygdala studies says there is a link, but that “most studies to date are nevertheless lacking in statistical power”.

Counter: there were also a lot of meta-analyses that found the opposite. The Slate article on the “orchid gene” came out after Risch’s work, mentioned it, but then quoted a scientist calling it “bullshit”. I don’t think the warnings did anything more than convince people that this was a controversial field with lots of evidence on both sides. For that matter, I don’t know if this new paper will do anything more than convince people of that. Maybe I trust geneticists saying “no, listen to me, it’s definitely like this” more than the average psychiatrist does. Maybe we’re still far from hearing the last of 5-HTTLPR and its friends.

Second, this paper doesn’t directly prove that every single study on 5-HTTLPR was wrong. It doesn’t prove that it doesn’t cause depression in children with depressed mothers in particular. It doesn’t prove that it doesn’t cause insomnia, or PTSD, or nostalgia-proneness. It doesn’t prove that it doesn’t affect amygdala function.

Counter: it doesn’t directly prove this, but it casts doubt upon them. The authors of this paper are geneticists who are politely trying to explain how genetics works to psychiatrists. They are arguing that single genes usually matter less than people think. They do an analysis of depression to demonstrate that they know what they’re talking about, but the points they are making apply to insomnia, nostalgia, and everything else. So all the studies above are at least questionable.

Third, most of these studies were done between 2000 – 2010, when we understood less about genetics. Surely you can’t blame people for trying?

Counter: The problem isn’t that people studied this. The problem is that the studies came out positive when they shouldn’t have. This was a perfectly fine thing to study before we understood genetics well, but the whole point of studying is that, once you have done 450 studies on something, you should end up with more knowledge than you started with. In this case we ended up with less.

(if you’re wondering how you can do 450 studies on something and get it wrong, you may be new here – read eg here and here)

Also, the studies keep coming. Association Between 5-HTTLPR Polymorphism, Suicide Attempts, And Comorbidity In Mexican Adolescents With Major Depressive Disorder is from this January. Examining The Effect Of 5-HTTLPR ON Depressive Symptoms In Postmenopausal Women 1 Year After Initial Breast Cancer Treatment is from this February. Association Of DRD2, 5-HTTLPR, And 5-HTTVNTR With PTSD In Tibetan Adolescents was published *after* the Border et al paper! Come on!

Having presented the case for taking it easy, I also want to present the opposite case: the one for being as concerned as possible.

First, what bothers me isn't just that people said 5-HTTLPR mattered and it didn't. It's that we built whole imaginary edifices, whole castles in the air on top of this idea of 5-HTTLPR mattering. We "figured out" how 5-HTTLPR exerted its effects, what parts of the brain it was active in, what sorts of things it interacted with, how its effects were enhanced or suppressed by the effects of other imaginary depression genes. This isn't just an explorer coming back from the Orient and claiming there are unicorns there. It's the explorer describing the life cycle of unicorns, what unicorns eat, all the different subspecies of unicorn, which cuts of unicorn meat are tastiest, and a blow-by-blow account of a wrestling match between unicorns and Bigfoot.

This is why I start worrying when people talk about how maybe the replication crisis is overblown because sometimes experiments will go differently in different contexts. The problem isn't just that sometimes an effect exists in a cold room but not in a hot room. The problem is more like "you can get an entire field with hundreds of studies analyzing the behavior of something that doesn't exist”. There is no amount of context-sensitivity that can help this.

Second, most studies about 5-HTTLPR served to reinforce all of our earlier preconceptions. Start with the elephant in the room: 5-HTTLPR is a serotonin transporter gene. SSRIs act on the serotonin transporter. If 5-HTTLPR played an important role in depression, we were right to focus on serotonin and extra-right to prescribe SSRIs; in fact, you could think of SSRIs as directly countering a genetic deficiency in depressed people. I don’t have any evidence that the pharmaceutical industry funded 5-HTTLPR studies or pushed 5-HTTLPR. As far as I can tell, they just created a general buzz of excitement around the serotonin transporter, scientists looked there, and then – since crappy science will find whatever it’s looking for – it was appropriately discovered that yes, changes in the serotonin transporter gene caused depression.

But this was just the worst example of a general tendency. Lots of people were already investigating the role of the HPA axis in depression – so lo and behold, it was discovered that 5-HTTLPR affected the HPA axis. Other groups were already investigating the role of BDNF in depression – so lo and behold, it was discovered that 5-HTTLPR affected BDNF. Lots of people already thought bad parenting caused depression – so lo and behold, it was discovered that 5-HTTLPR modulated the effects of bad parenting. Once 5-HTTLPR became a buzzword, everyone who thought anything at all went off and did a study showing that 5-HTTLPR played a role in whatever they had been studying before.

From the outside, this looked like people confirming they had been on the right track. If you previously doubted that bad parenting played a role in depression, now you could open up a journal and discover that *the gene for depression* interacts with bad parenting! If you'd previously doubted there was a role for the amygdala, you could open up a journal and find that *the gene for depression* affects amygdala function. Everything people wanted to believe anyway got a new veneer of scientific credibility, and it was all fake.

Third, antidepressant pharmacogenomic testing.

This is the thing where your psychiatrist orders a genetic test that tells her which antidepressant is right for you. Everyone keeps talking these up as the future of psychiatry, saying how it's so cool how now we have true personalized medicine, how it's an outrage that insurance companies won't cover them, etc, etc, etc. The tests have made their way into hospitals, into psychiatry residency programs, and various high-priced concierge medical systems. A company that makes them recently sold for $410 million, and the industry as a whole may be valued in the billions of dollars; the tests themselves cost as much as $2000 per person, most of which depressed patients have to pay out of pocket. I keep trying to tell people these tests don't work, but this hasn't affected their popularity.

A lot of these tests rely on 5-HTTLPR. GeneSight, one of the most popular, uses seven genes. One is SLC6A4, the gene containing 5-HTTLPR as a subregion. Another is HTR2A, which Border et al debunked in the same study as 5-HTTLPR. The studies above do not directly prove that these genes don't affect antidepressant response. But since the only reason we thought that they might was because of evidence they affected depression, and now it seems they don't affect depression, it's less likely that they affect antidepressant response too.

The other five are liver enzymes. I am not an expert on the liver and I can’t say for sure that you can’t use a few genes to test liver enzymes’ metabolism of antidepressants. But people who are experts in the liver tell me you can’t. And given that GeneSight has already used two genes that we know don’t work, why should we trust that they did any better a job with the liver than they did with the brain?

Remember, GeneSight and their competitors refuse to release the proprietary algorithms they use to make predictions. They refuse to let any independent researchers study whether their technique works. They dismiss all the independent scientists saying that their claims are impossible by arguing that they’re light-years ahead of mainstream science and can do things that nobody else can. If you believed them before, you should be more cautious now. They are not light-years ahead of mainstream science. They took some genes that mainstream science had made a fuss over and claimed they could use them to predict depression. Now we think they were wrong about those. What are the chances they’re right about the others?

Yes, GeneSight has ten or twenty studies proving that their methods work. Those were all done by scientists working for GeneSight. Remember, *if you have bad science you can prove whatever you want.* What does GeneSight want? Is it possible they want their product to work and make them $410 million? This sounds like the kind of thing that companies sometimes want, I dunno.

I’m really worried I don’t see anyone updating on this. From where I’m sitting, the Border et al study passed unremarked upon. Maybe I’m not plugged in to the right discussion networks, I don’t know.

But I think we should take a second to remember that yes, this is really bad. That this is a rare case where methodological improvements allowed a conclusive test of a popular hypothesis, and it failed badly. How many other cases like this are there, where there’s no geneticist with a 600,000 person sample size to check if it’s true or not? How many of our scientific edifices are built on air? How many useless products are out there under the guise of good science? We still don’t know.

Is There A Case For Skepticism Of Psychedelic Therapy?

There's been an explosion of interest in the use of psychedelics in psychiatry. Like everyone else, I hope this works out. But recent discussion has been so overwhelmingly positive that it's worth reviewing whether there's a case for skepticism. I think it would look something like this:

1. Psychedelics have mostly been investigated in small studies run by true believers. These are the conditions that produce a field made of unreplicable results, like the effects of 5-HTTLPR. Some of the most exciting psychedelic findings have already failed to replicate; for example, a study two years ago found that psilocybin did not permanently increase the Openness personality trait. This was one of the most exciting studies and had shaped a lot of my thinking around the issue. Now it's gone.

2. Some of the most impressive stories involve psychedelic-assisted psychotherapy, where people who talk with a therapist, while on a substance, obtain true insight and get real closure. But every psychotherapy has amazing success stories floating out there. Back when psychoanalysis was new, the whole world was full of people telling their amazing success stories about how Dr. Freud helped them obtain true insight and get real closure. I think of psychotherapy as a domain where people can get as many amazing success stories as they want whether or not they're really doing anything right, for unclear reasons.

3. Ketamine is the best comparison for psychedelics. Like psychedelics, it's often used as a recreational drug, and produces profound experiences. Like psychedelics, it got

hyped as an exciting new innovation that was going to revolutionize everything in psychiatry (in this case, depression treatment). But it’s been in pretty common (albeit non-formulary) use for five years now, and nothing has been revolutionized; my (very anecdotal) impression is that most patients who seek ketamine treatment find it only about as helpful as anything else. The gold-standard FDA studies are abysmal, worse than most other antidepressant medications. I’m sure ketamine works great for some people, just as SSRIs, therapy, and diet/exercise work well for some people. But at least so far it hasn’t been revolutionary.

4. Another good comparison is NSI-189. Again, a totally revolutionary new drug with a totally revolutionary new mechanism, with so many anecdotes of amazing success that depressed people started getting it on the black market before the FDA trials were even underway. People were posting testimonials that NSI-189 changed their life and that it was going to destroy the market for every other antidepressant. When the FDA trials finally finished, it was discovered to be ineffective. Seriously, the graveyards are littered with revolutionary new treatments for treatment-resistant depression that have great success in anecdotes and preliminary studies.

5. Between 10% and 50% of Americans have tried psychedelics. If psychedelics did something shocking, we would already know about it. I occasionally hear stories like “I did LSD and my depression went away”, but I also occasionally hear stories like “I did LSD and then my depression got worse”, so whatever. I know plenty of people who use heroic amounts of LSD all the time, and are still nervous wrecks. It’s possible there’s some set and setting that will improve this, but see part 7 below.

(one exception to this might be microdosing, which is a pretty new idea and might work differently from regular trips.)

6. In my model of psychedelics, they artificially stimulate your insight system the same way heroin artificially stimulates your happiness system. This leads to all those stories where people feel like they discovered the secret of the universe, but when they recover their faculties, they find it was only some inane triviality. This sounds very likely to produce people who *think* their psychedelic experience has changed everything and solved all their problems, which means we should discount these impressions as evidence that psychedelics really *do* change everything and solve all your problems. Granted, feeling like you truly understand the universe may itself help with depression, but I worry this is not a very lasting effect. See my posts on PIHKal and Universal Love, Said The Cactus Person.

7. Even if all of the above are wrong and psychedelics work very well, the FDA could kill them with a thousand paper cuts. Again, look at ketamine: the new FDA approval ensures people will be getting the slightly different esketamine, through a weird route of administration, while paying $600 a pop, in specialized clinics that will probably be hard to find. Given the price and inconvenience, insurance companies will probably restrict it to the most treatment-resistant patients, and it probably won't help them (treatment-resistant patients tend to stay that way). Given the panic around psychedelics, I expect it to be similarly difficult to get them *even if they are legal and technically FDA-approved*. Depressed people will never be able to walk into a psychiatrist's office and get LSD. They'll walk into a psychiatrist's office, try Prozac for three months, try Wellbutrin for three months, argue with their insurance for a while, eventually get permission to drive to a city an

hour away that has a government-licensed LSD clinic, and get some weird form of LSD that might or might not work, using a procedure optimized to minimize hallucinations. I don’t know what the optimal set and setting for LSD is, but if it’s anything other than “the inside of a government-licensed LSD clinic, having a government-licensed LSD therapist ask you standard questions”, you won’t get it.

I hope I am wrong about this, I really do. And I think there’s a good chance that I might be. I really want psychedelic research to succeed and I support it wholeheartedly. But there’s been so much hype around so many things before that I want to avoid getting burned again, so I ‘ll stay skeptical for now.

Age Gaps And Birth Order Effects

[Parts of this post have since been shown to be wrong, as explained in this post. I endorse this reanalysis as better than the current post.]

Psychologists are split on the existence of "birth order effects", where oldest siblings will have different personality traits and outcomes than middle or youngest siblings. Although some studies detect effects, they tend to be weak and inconsistent.

Last year, I posted Birth Order Effects Exist And Are Very Strong, finding a robust 70-30 imbalance in favor of older siblings among SSC readers. I speculated that taking a pre-selected population and counting the firstborn-to-laterborn ratio was better at revealing these effects than taking an unselected population and trying to measure their personality traits. Since then, other independent researchers have confirmed similar effects in historical mathematicians and Nobel-winning physicists. Although birth order effects do not seem to consistently affect IQ, some studies suggest that they do affect something like "intellectual curiosity", which would explain firstborns' over-representation in intellectual communities.

Why would firstborns be more intellectually curious? If we knew that, could we do something different to make laterborns more intellectually curious? A growing body of research highlights the importance of genetics on children's personalities and outcomes, and casts doubt on the ability of parents and teachers to significantly affect their trajectories. But here's a non-genetic factor that's a really big deal on one of the personality traits closest to our hearts. How does it work?

People looking into birth order effects have come up with a couple of possible explanations:

1. Intra-family competition. The oldest child choose some interest or life path. Then younger children don’t want to live in their older sibling’s shadow all the time, so they do something else.

2. Decreased parental investment. Parents can devote 100% of their child-rearing time to the oldest child, but only 50% or less to subsequent children.

3. Changed parenting strategies. Parents may take extra care with their firstborn, since they are new to parenting and don’t know what small oversights they can get away with vs. what will end in disaster. Afterwards, they are more relaxed and willing to let the child “take care of themselves”. Or they become less interested in parenting because it is no longer novel.

4. Maternal antibodies. Studies show that younger sons with older biological brothers (but not sisters!) are more likely to be homosexual. This holds true even if someone is adopted and never met their older brother. The most commonly-cited theory is that during a first pregnancy, the mother’s immune system may develop antibodies to some unexpected part of the male fetus (maybe androgen receptors?) and damages these receptors during subsequent pregnancies. A similar process could be responsible for other birth order effects.

5. Maternal vitamin deficiencies. An alert reader sent me Does Birth Spacing Affect Maternal Or Child Nutritional Status? It points out that people maintain “stockpiles” of various nutrients in their bodies. During pregnancy, a woman may deplete her nutrient stockpiles in the difficult

task of creating a baby, and the stockpiles may take years to recover. If the woman gets pregnant again before she recovers, she might not have enough nutrients for the fetus, and that may affect its development.

How can we distinguish among these possibilities? One starting point might be to see how age gaps affect birth order effects. How close together do two siblings have to be for the older to affect the younger? If a couple has a child, waits ten years, and then has a second child, does the second child still show the classic laterborn pattern? If so, we might be more concerned about maternal antibodies or changes in parenting style. If not, we might be more concerned about vitamin deficiencies or distracted parental attention.

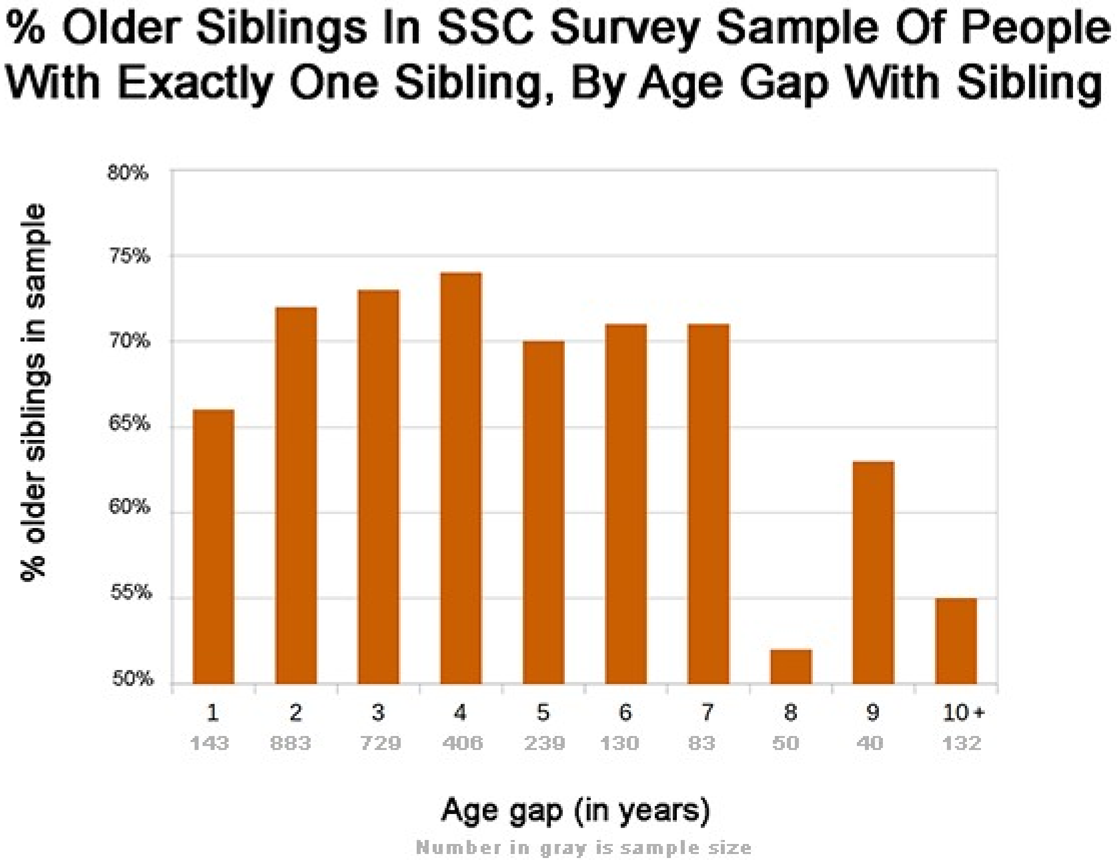
Methods And Results

I used the 2019 Slate Star Codex survey, in which 8,171 readers of this blog answered a few hundred questions about their lives and opinions.

Of those respondents, I took the subset who had exactly one sibling, who reported an age gap of one year or more, and who reported their age gap with an integer result (I rounded non-integers to integers if they were not .5, and threw out .5 answers). 2,835 respondents met these criteria.

Of these 2,835, 71% were the older sibling and 29% were the younger sibling. This replicates the results from last year’s survey, which also found that 71% of one-sibling readers were older.

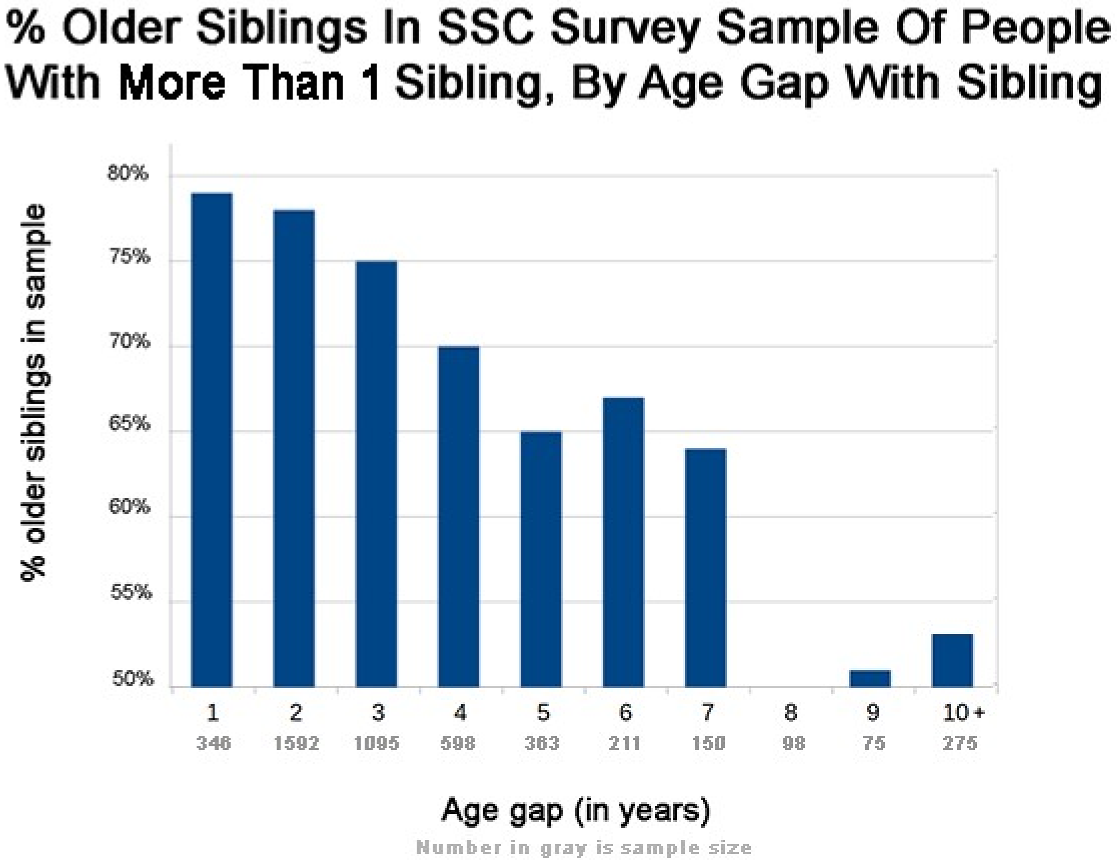
Here are the results by age gap:



% Older Siblings In SSC Survey Sample Of People With Exactly One Sibling, By Age Gap With Sibling

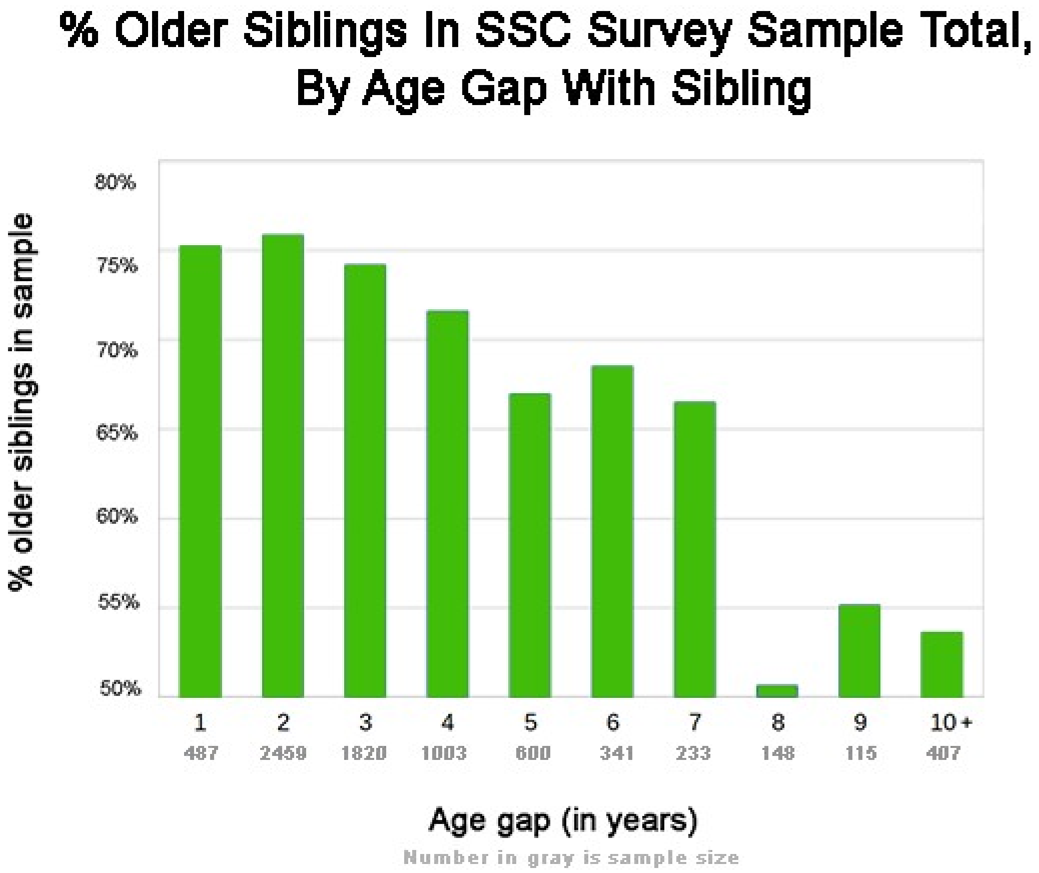
Birth order effects are strong from one-year to seven-year age gaps, and don’t differ much within that space. After seven years, birth order effects decrease dramatically and are no longer significantly different from zero.

I also investigated people who had more than one sibling, but were either the oldest or the youngest in their families.



More siblings = ~~more problems~~ more of a birth order effect, but the overall pattern was similar. There is a possible small decline in strength from one to seven years, followed by a very large decline between seven and eight years.

Here’s the previous two graphs considered as a single very-large-n sample:



The pattern remains pretty clear: vague hints of a decline from age 1 to 7, followed by a very large decline afterwards.

(Tumblr user athenaegalea kindly double-checked my calculations; you can see her slightly-differently-presented results here).

Weirdly, among people who reported a zero-year age gap, 70% are older siblings. This wouldn't make much sense for twins, since here older vs. younger just means who made it out of the uterus first. I don't know if this means there's some kind of reporting error that discredits this entire project, whether people who were born about 9 months apart reported this as a zero year age gap, or whether it's just an unfortunate coincidence.

These results suggest that age gaps do affect the strength of birth order effects. People with siblings seven or fewer years older than them will behave as laterborns; people separated from their older siblings by more than seven years will act like firstborn children.

Discussion

This study found an ambiguous and gradual decline from one to seven years, but also a much bigger cliff from seven to eight years. Is this a coincidence, or is there something important that happens at seven?

Most of the sample was American; in the US, children start school at about age five. Although it might make sense for older siblings stop mattering once they are in school, this would predict a cliff at five years rather than seven years.

Developmental psychologists sometimes distinguish between early childhood (before 6-8 years) and middle childhood (after that point). This is supposed to be a real qualitative transition, just like eg puberty. We might take this very seriously, and posit that having a sibling in early childhood causes birth order effects, but one in middle childhood doesn’t. But why should this be? Overall I’m still pretty confused about this.

These results may be consistent with an intra-family competition hypothesis. Children try to avoid living in the shadow of their older siblings, perhaps by avoiding intellectual pursuits those children find interesting. But if there is too much of an age gap, then siblings are at such different places that competition no longer feels relevant.

These results may be partly consistent with a parental investment hypothesis. Parents might have to split their attention between first and laterborn children, so that laterborns never get the period of sustained parental attention that firstborns do. But since an age gap as small as one year produces this effect, this would suggest that only the first year of childrearing matters; after the first year, even the firstborn children in this group are getting split attention. This is hard to explain if we are talking about as complicated a trait as “intellectual curiosity” – surely there are things parents do when a child is two or three to make them more curious?

These results don’t seem consistent with hypotheses based on changing parenting strategies or maternal antibodies, unless parenting strategies or the immune system “reset” to their naive values after a certain number of years.

They also don’t seem too consistent with vitamin-based hypotheses. I don’t know how long it takes to replenish vitamin stockpiles, and it’s probably different for every vitamin. But I would be surprised if giving people one vs. five years for this had basically no effect, but giving them eight instead of seven years had a very large effect. Overall I would expect the first year of vitamin replenishment to be the most important, with diminishing returns thereafter, which doesn’t fit the birth order effect pattern.

Overall these results make me lean slightly more towards intra-family competition or parental investment as the major cause of birth order effects. I can’t immediately think of a way to distinguish between these two hypotheses, but I’m interested in hearing people’s ideas.

I welcome people trying to replicate or expand on these results. All of the data used in this post are freely available and can be downloaded here.

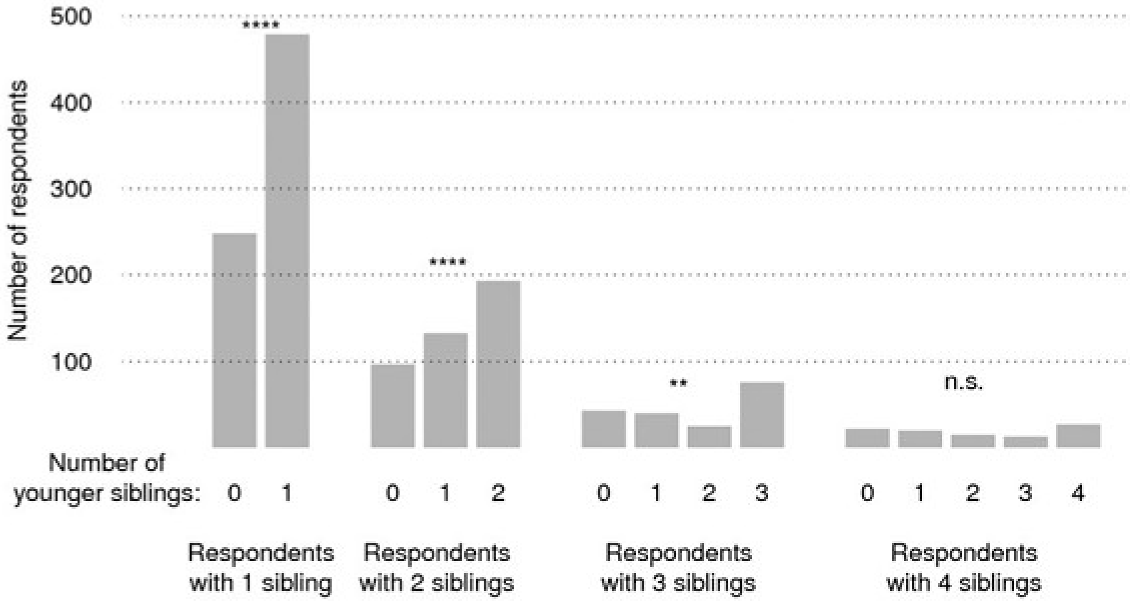
A Critical Period For Lactation Fetishes

Enquist et al on lactation fetishes is one of my favorite papers.

They wonder – as we've all wondered at one point or another – how people develop fetishes. One plausible hypothesis is "sexual imprinting". During childhood, you have a critical period (maybe ages 1 to 5) where you figure out what sex is. If you see some weird stuff during that time, you could end up with a fetish. For example, a child who sees latex used in a sexualized way (for example, they catch a glimpse of a sexy movie where someone is wearing latex) might grow up with a latex fetish.

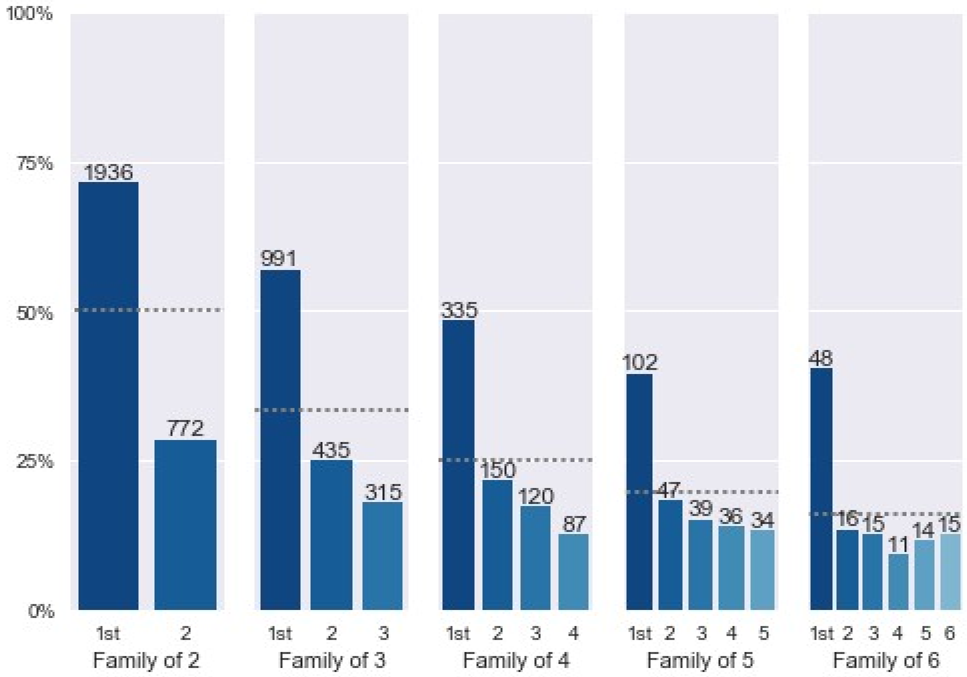
Enquist et al realize lactation fetishes offer a natural test of this hypothesis. Children with younger siblings will see a lot of breastfeeding going on during their critical window; children without younger siblings will see less. Since it's easy to ask people how many siblings they have, you can see if younger siblings correlate with lactation fetishes.

They survey some online lactation fetishist communities and ask everyone how many older and younger siblings they have. Although by chance we would expect an equal number of both, in fact the fetishists have many more younger than older siblings:



They interpret this as support for their critical window theory.

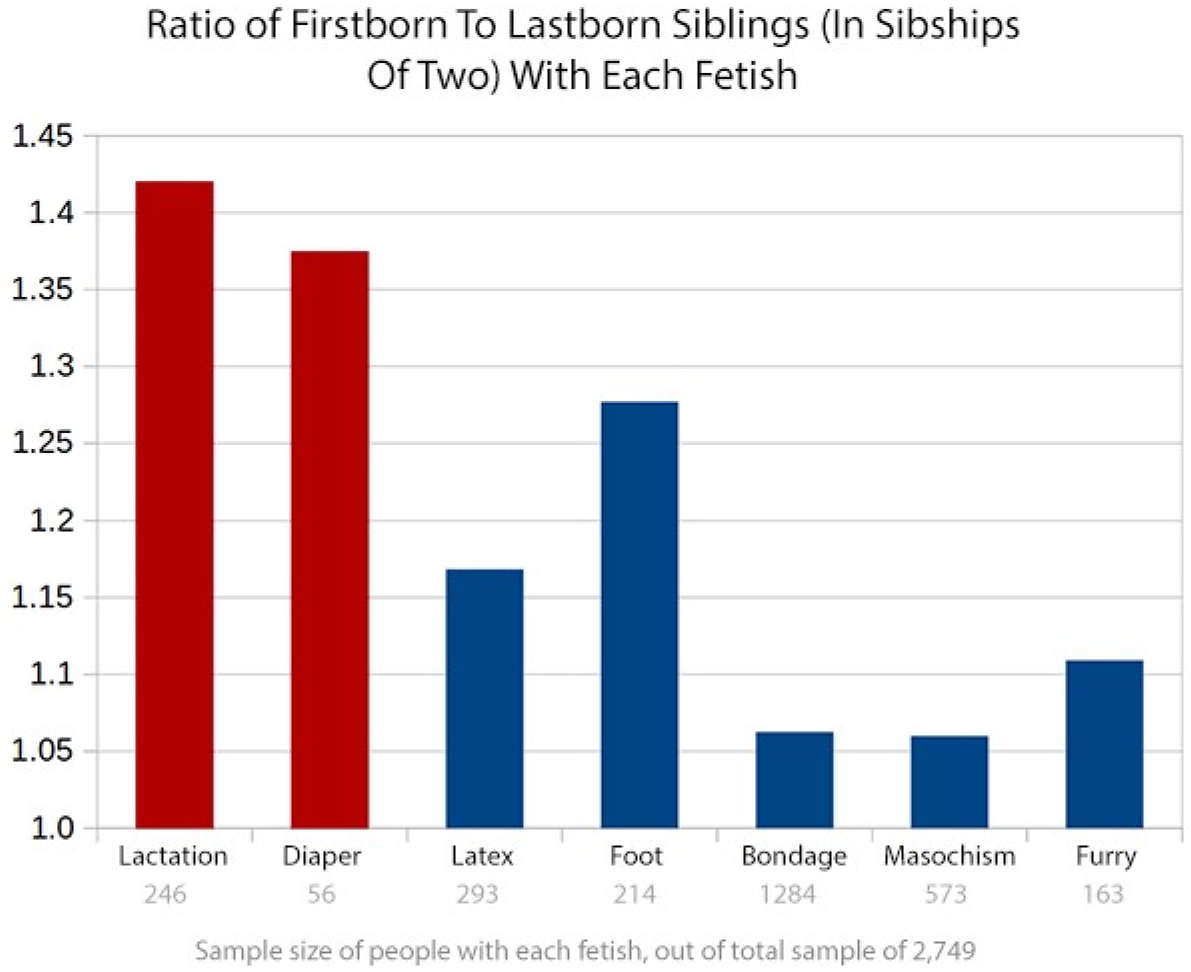
But their graph looks a lot like this graph of SSC readers:



I use the opposite order they do, but both graphs show the same thing: more older than younger siblings.

I interpret the SSC data as showing a birth order effect on intellectual curiosity. But if this is true, it casts Enquist et al’s results into doubt. The trait I’m calling “intellectual curiosity” is linked to openness to experience. Could it also cause people to be more curious and open about fetishes, or more likely to join online communities about those fetishes?

I decided to test these hypotheses using the SSC 2019 Survey, which contained data on participants’ fetishes. I looked into lactation to replicate the Enquist results, but also into diaper fetishes, since that seemed like another fetish where exposure to the relevant stimulus would depend a lot on having a baby in the house. I also looked at latex, foot, bondage, masochism, and furry fetishes as control groups. Here are the results:

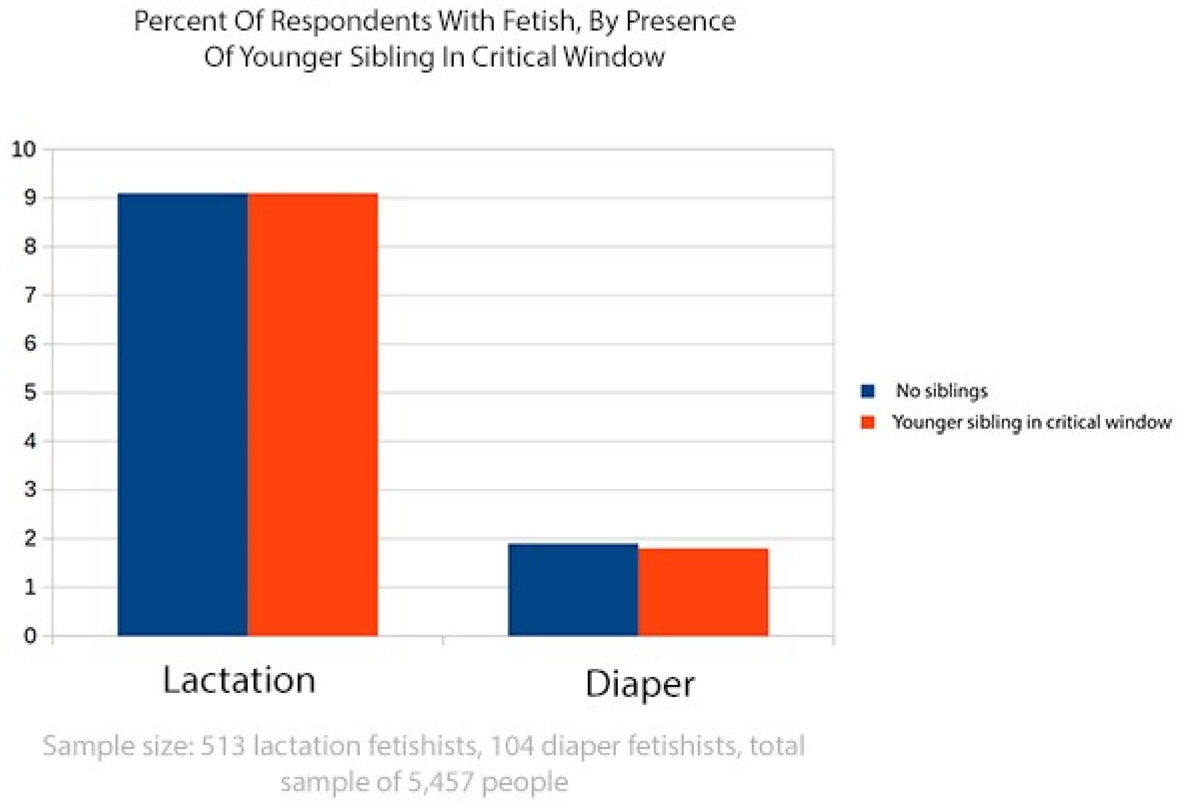


All fetishes had more older than younger siblings. But the difference was only significant in lactation fetish (p = 0.01). The difference between significant and nonsignificant results is not always itself significant, and I'm not sure how I would properly analyze this given the many different comparisons (some of which I made after seeing the data). i lean towards not being very impressed with the critical window theory on this metric.

But here's another potential test: compare people with younger siblings to people with *no* siblings.

In Enquist et al's model, the presence of a younger sibling causes the lactation fetish. In my model, the presence of an older sibling suppresses openness to experience and prevents fetish formation. So if only children behaved more like older siblings, that would support my model; if they behaved more like younger children, it would support Enquist et al.

I compared participants with no siblings to participants with a younger sibling in the critical window of less than five years age gap. Here are the results:


Percent Of Respondents With Fetish, By Presence Of Younger Sibling In Critical Window

Sample size: 513 lactation fetishists, 104 diaper fetishists, total sample of 5,457 people

No difference. This suggests that having a younger sibling does not make you more likely to develop a baby-related fetish, which suggests it’s just that having an older sibling makes you *less* likely to develop it. This casts doubt on any simple critical window theory of fetishes, and makes it more likely that Enquist et al were just detecting the same birth order effects on openness to experience that can be found in many unusual communities.

These data are thanks to people who graciously revealed their deepest secrets for the cause of science; please be kind and don’t make fun of them or call them gross in the comments. Although I’ve made every other part of the survey publicly available, given the sensitivity of fetishes I’m keeping these particular answers private. If you are a professional researcher (or an amateur researcher with a good track record of professionalism and data integrity), and you want to test these results, please email me at scott[at]slatestarcodex[dot]com and we can discuss how to make that happen.

APA Meetup This Saturday

I’d like to meet any SSC readers who will be at the American Psychiatric Association meeting this weekend in San Francisco.

I propose lunch on Saturday 5/18, 12 PM. We can meet at Room 312 (randomly chosen as a room that doesn’t seem to be occupied at the time; if I’m wrong and it’s in use we’ll congregate awkwardly by the door) then go to a nearby restaurant. I’ll be wearing a dark blue shirt (if I remember), the silver spiral necklace I use as my avatar in the comments, and a nametag with my real name (Scott S_____). Please watch this post for potential emergency changes in time/location.

If you’re thinking of coming, send me an email at scott@slatestarcodex.com with your phone number so I know how many people to expect/wait for and can contact you if needed. If you don’t send me an email on time, don’t worry, you can still show up.

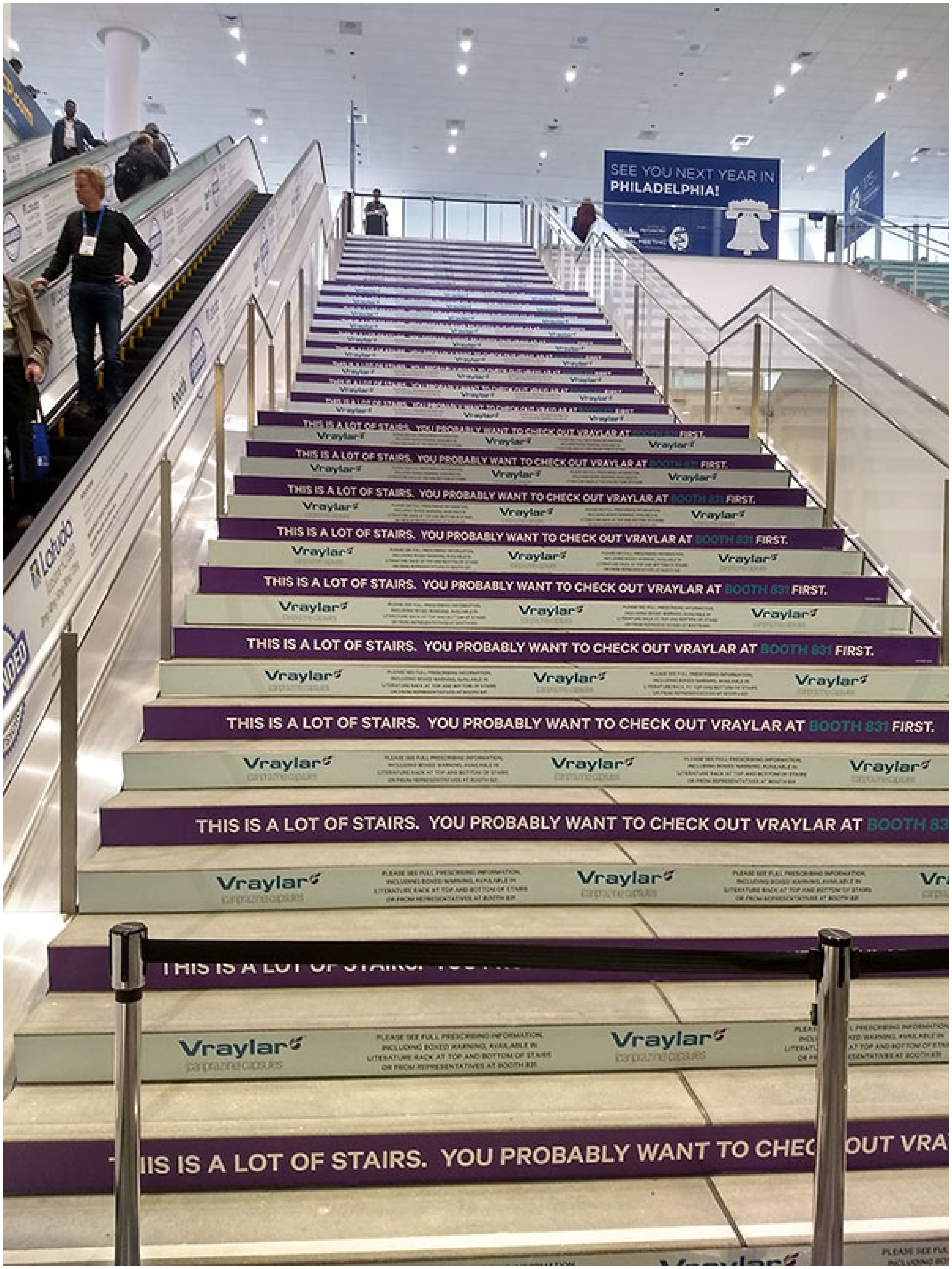
The APA Meeting: A Photo-Essay

The first thing you notice at the American Psychiatric Association meeting is its size. By conservative estimates, a quarter of the psychiatrists in the United States are packed into a single giant San Francisco convention center, more than 15,000 people.

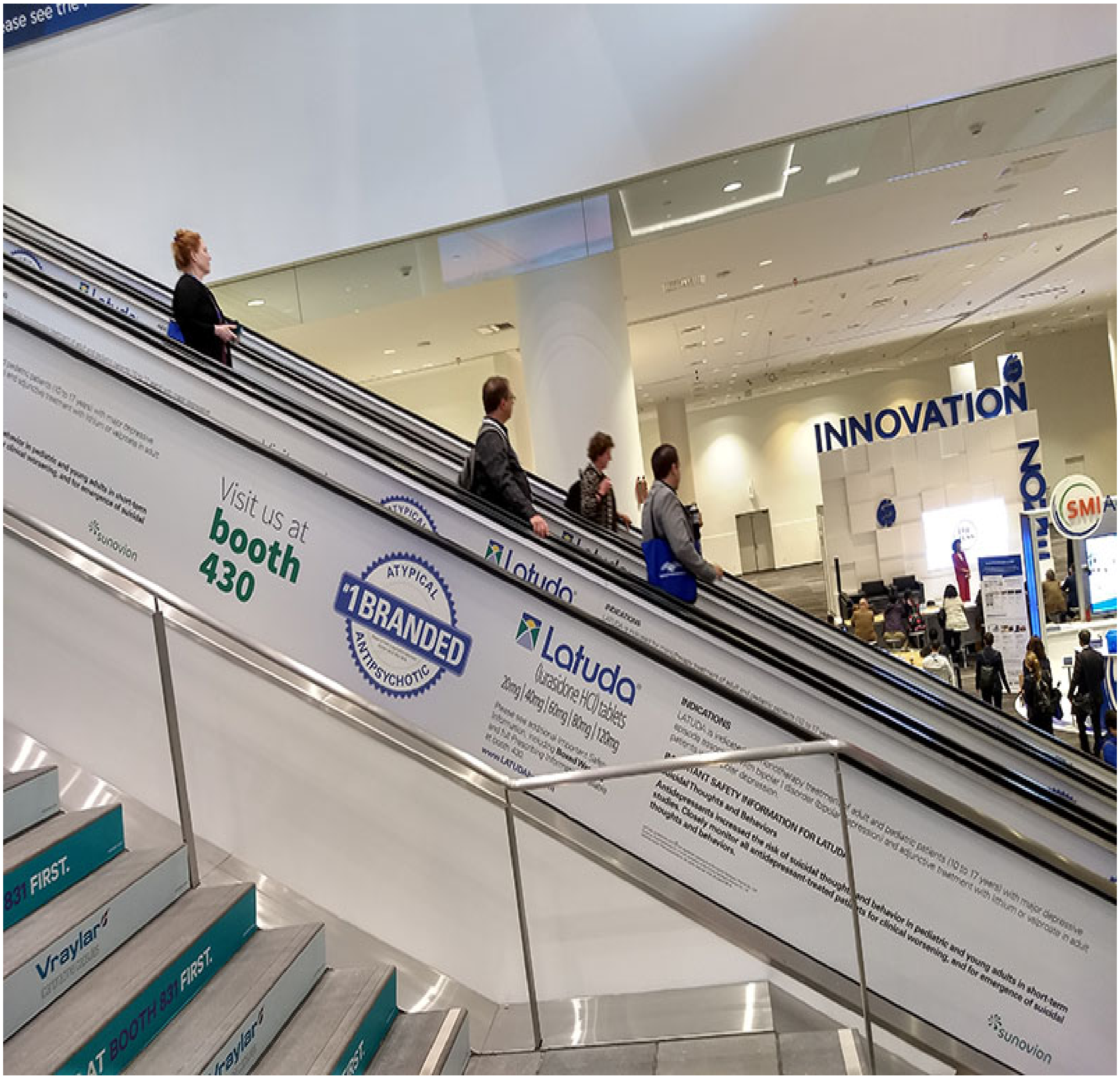
Being in a crowd of 15,000 psychiatrists is a weird experience. You realize that all psychiatrists look alike in an indefinable way. The men all look balding, yet dignified. The women all look maternal, yet stylish. Sometimes you will see a knot of foreign-looking people huddled together, their nametags announcing them as the delegation from the Nigerian Psychiatric Association or the Nepalese Psychiatric Association or somewhere else very far away. But however exotic, something about them remains ineffably psychiatrist.

* * *

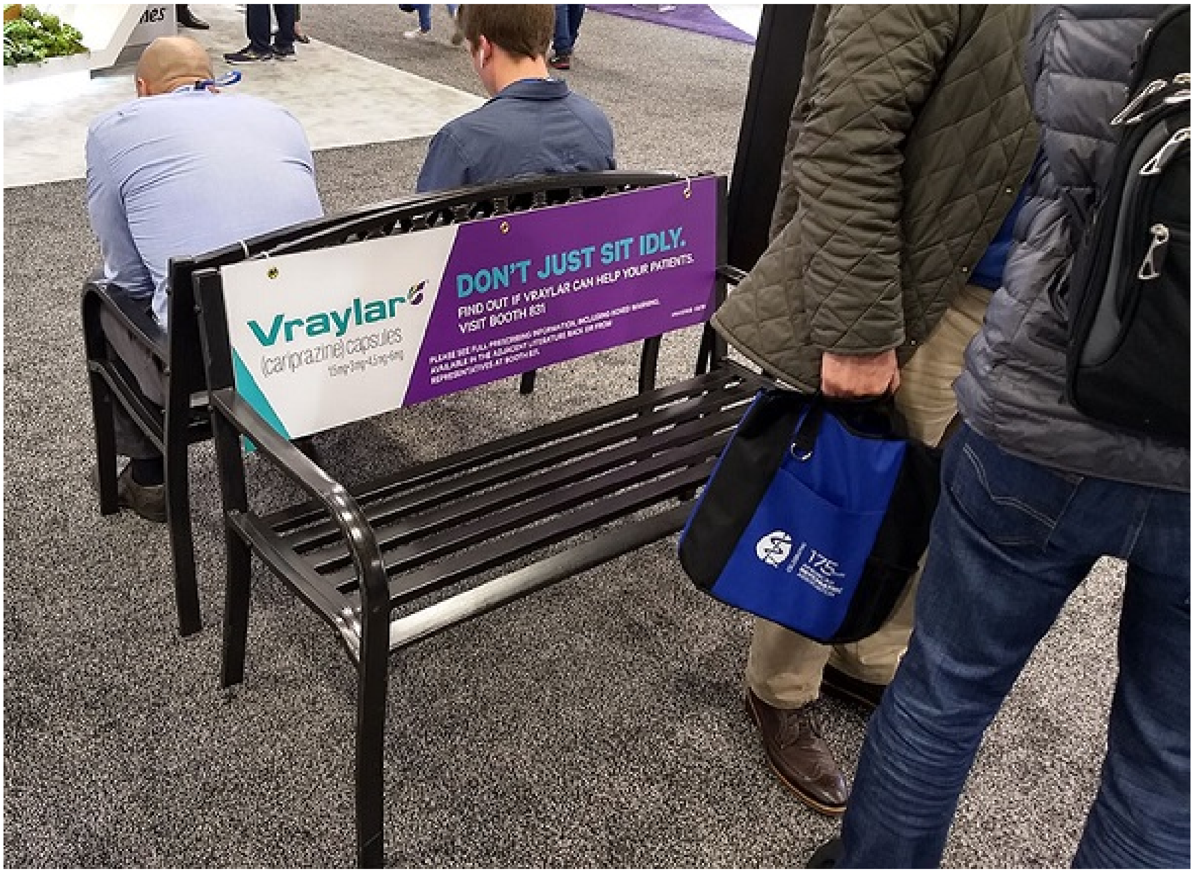
The second thing you notice at the American Psychiatric Association meeting is that the staircase is shaming you for not knowing enough about Vraylar®.


SEE YOU NEXT YEAR IN PHILADELPHIA!
THIS IS A LOT OF STAIRS. YOU PROBABLY WANT TO CHECK OUT VRAYLAR AT BOOTH 831 FIRST.
Vraylar
cariprazine capsules
PLEASE SEE FULL PRESCRIBING INFORMATION, INCLUDING BOXED WARNING, AVAILABLE IN LITERATURE RACK AT TOP AND BOTTOM OF STAIRS OR FROM REPRESENTATIVES AT BOOTH 831
Latuda

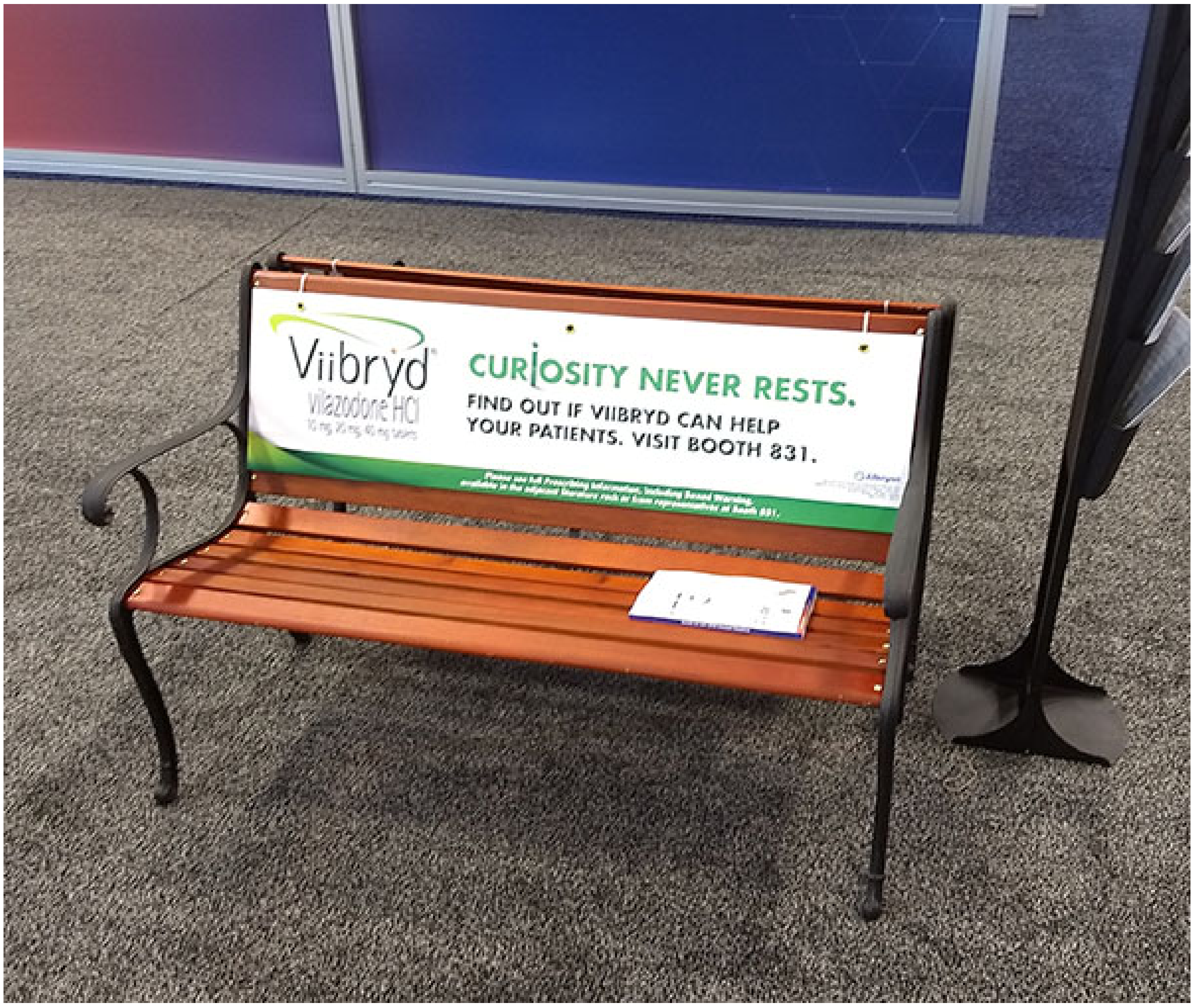
Seems kind of weird. Maybe I’ll just take the escalator…


Visit us at
booth
430
ATYPICAL
#1 BRANDED
ANTIPSYCHOTIC
Latuda
(lurasidone HCl) tablets
20mg | 40mg | 60mg | 80mg | 120mg
INDICATIONS
IMPORTANT SAFETY INFORMATION FOR LATUDA
sunovion
INNOVATION
SMI
Vraylar
831 FIRST.
AT BOOTH 831 FIRST.

…no, the escalator is advertising Latuda®, the "number one branded atypical antipsychotic". Aaaaaah! Maybe I should just sit down for a second and figure out what to do next…


Vraylar
(cariprazine) capsules
DON'T JUST SIT IDLY.
FIND OUT IF VRAYLAR CAN HELP YOUR PATIENTS.
VISIT BOOTH 831
175

AAAAH, CAN'T SIT DOWN, VRAYLAR® HAS GOTTEN TO THE BENCHES TOO! Surely there's a non-Vraylar bench somewhere in this 15,000 person convention center!


Viibryd
vilazodone HCl
10 mg, 20 mg, 40 mg tablets
CURIOSITY NEVER RESTS.
FIND OUT IF VIIBRYD CAN HELP YOUR PATIENTS. VISIT BOOTH 831.

…whatever, close enough.

You know how drug companies pay six or seven figures for thirty-second television ads just on the off chance that someone with the relevant condition might be watching? You know how they employ drug reps to flatter, cajole, and even seduce doctors who might prescribe their drug? Well, it turns out that having 15,000 psychiatrists in one building sparks a drug company feeding frenzy that makes piranhas look sedate by comparison. Every flat surface is covered in drug advertisements. And after the flat surfaces are gone, the curved sufaces, and after the curved surfaces, giant rings hanging from the ceiling.

The ads overflow from the convention itself to the city outside. For about two blocks in any direction, normal ads and billboards have been replaced with psychiatry-themed ones, until they finally peter off and segue into the usual startup advertisements around Market Street.

CITY COLLEGE
xplore the power
f collaboration
schizophrenia care
in us at APA 2019 to explore
different disciplines join forces
elevate schizophrenia care
UNITED FOR SCHIZOPHRENIA
NEWSSTAND
ACE

There’s a popular narrative that drug companies have stolen the soul of psychiatry. That they’ve reduced everything to chemical imbalances. The people who talk about this usually go on to argue that the true causes of mental illness are capitalism and racism. Have doctors forgotten that the *real* solution isn't a pill, but structural change that challenges the systems of exploitation and domination that create suffering in the first place?

No. Nobody has forgotten that. Because **the third thing** you notice at the American Psychiatric Association meeting is that everyone is very, *very* woke.

Here are some of the most relevant presentations listed in my Guidebook:

Saturday, May 18

Climate Psychiatry 101: What Every Psychiatrist Should Know

Women's Health In The US: Disruption And Exclusion In The Time Of Trump

Gender Bias In Academic Psychiatry In The Era Of the #MeToo Movement

Revitalizing Psychiatry – And Our World – With A Social Lens

Hip-Hop: Cultural Touchstone, Social Commentary, Therapeutic Expression, And Poetic Intervention

Lost Boys Of Sudan: Immigration As An Escape Route For Survival

Treating Muslim Patients After The Travel Ban: Best Practices In Using The APA Muslim Mental Health Toolkit

Making The Invisible Visible: Using Art To Explore Bias And Hierarchy In Medicine

Navigating Racism: Addressing The Pervasive Role Of Racial Bias In Mental Health

.

Sunday, May 20

Addressing Microaggressions Toward Sexual And Gender Minorities: Caring For LGBTQ+ Patients And Providers

Latino Undocumented Children And Families: Crisis At The Border And Beyond

Racism And Psychiatry: Growing A Diverse Psychiatric Workforce And Developing Structurally Competent Psychiatric Providers

Sex, Drugs, And Culturally Responsive Treatment: Addressing Substance Use Disorders In The Context Of Sexual And Gender Diversity

Grabbing The Third Rail: Race And Racism In Clinical Documentation

Racism And The War On Terror: Implications For Mental Health Providers In The United States

The Multiple Faces Of Deportation: Being A Solution To The Challenges Faced By Asylum Seekers, Mixed Status Families, And Dreamers

What Should The APA Do About Climate Change?

Intersectionality 2.0: How The Film *Moonlight* Can Teach Us About Inclusion And Therapeutic Alliance In Minority LGBTQ Populations

Transgender Care: How Psychiatrists Can Decrease Barriers And Provide Gender-Affirming Care

Gun Violence Is A Serious Public Health Problem Among America's Adolescents And Emerging Adults: What Should Psychiatrists Know And Do About It?

Working Clinically With Eco-Anxiety In The Age Of Climate Change: What Do We Know And What Can We Do?

Are There Structural Determinants Of African-American Child Mental Health? Child Welfare – A System Psychiatrists Should Scrutinize

.

Monday, May 21

Community Activism Narratives In Organized Medicine: Homosexuality, Mental Health, Social Justice, and the American Psychiatric Association

Disrupting The Status Quo: Addressing Racism In Medical Education And Residency Training

Ecological Grief, Eco-Anxiety, And Transformational Resilience: A Public health Perspective On Addressing Mental Health Impacts Of Climate Change

Immigration Status As A Social Determinant Of Mental Health: What Can Psychiatrists Do To Support Patients And Communities? A Call To Action

Psychiatry In The City Of Quartz: Notes On The Clinical Ethnography Of Severe Mental Illness And Social Inequality

Racism And Psychiatry: Understanding Context And Developing Policies For Undoing Structural Racism

Trauma Inflicted To Immigrant Children And Parents Through Policy Of Forced Family Separation

Deportation And Detention: Addressing The Psychosocial Impact On Migrant Children And Families

How Private Insurance Fails Those With Mental Illness: The Case For Single-Payer Health Care

Imams In Mental Health: Caring For Themselves While Caring For Others

Misogynist Ideology And Involuntary Celibacy: Prescription For Violence?

Advocacy: A Hallmark Of Psychiatrists Serving Minorities

Inequity By Structural Design: Psychiatrists' Responsibility To Be Informed Advocates For Systemic Education And Criminal Justice Reform

Treating Black Children And Families: What Are We Overlooking?

Blindspotting: An Exploration Of Implicit Bias, Race-Based Trauma, And Empathy

But I'm Not Racist: Racism, Implicit Bias, And The Practice Of Psychiatry

No Blacks, Fats, or Femmes: Stereotyping In The Gay

Community And Issues Of Racism, Body Image, And Masculinity

Silence Is Not Always Golden: Interrupting Offensive Remarks And Microaggressions

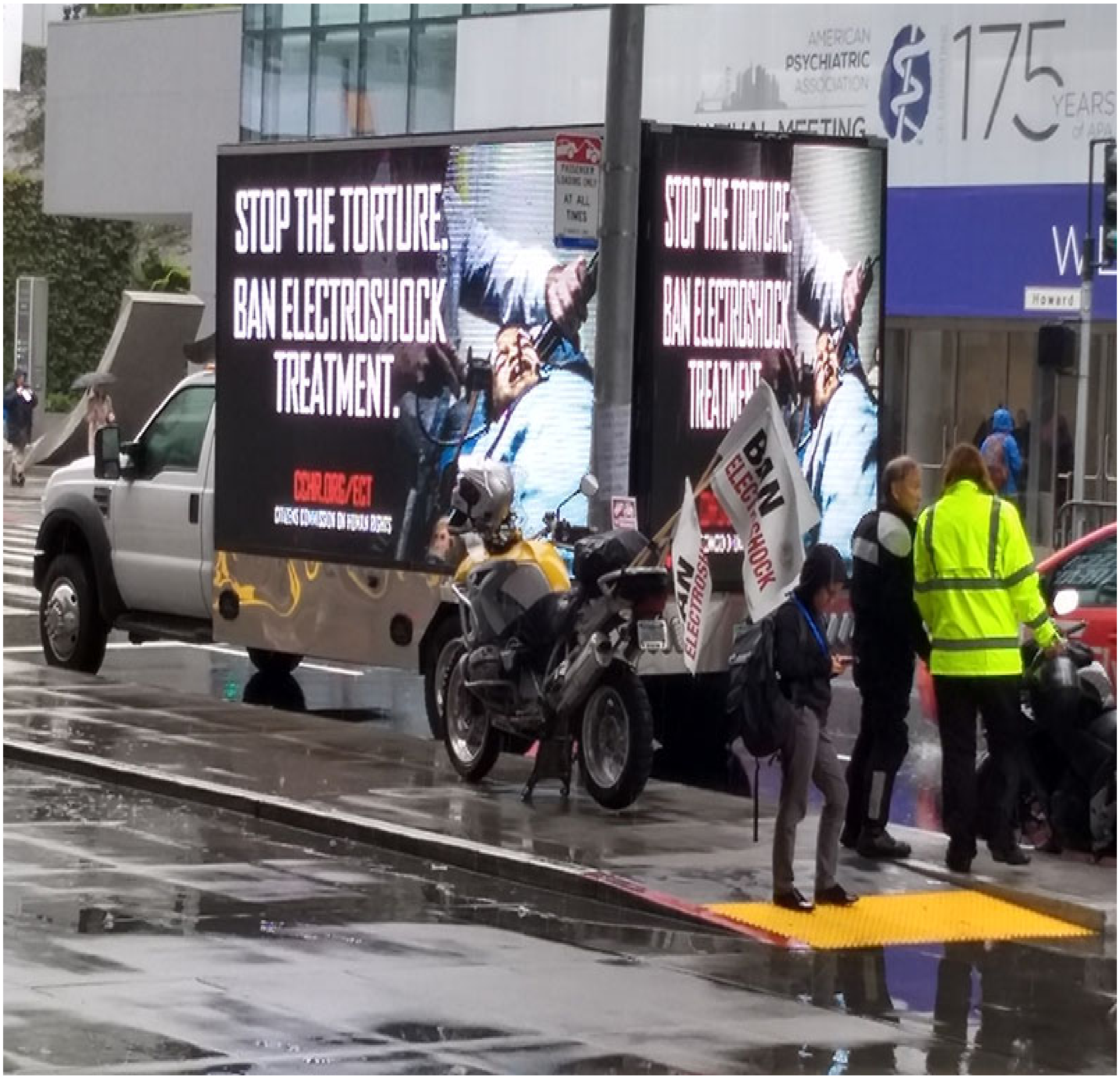
Black Minds Matter: The Impact Of #BlackLivesMatter On Psychiatry

…you get the idea, please don’t make me keep writing these.

Were there really more than twice as many sessions on global warming as on obsessive compulsive disorder? Three times as many on immigration as on ADHD? As best I can count, yes. I don’t want to exaggerate this. There was still a lot of really meaty scientific discussion if you sought it out. But overall the balance was pretty striking.

I’m reminded of the idea of woke capital, the weird alliance between very rich businesses and progressive signaling. If you want to model the APA, you could do worse than a giant firehose that takes in pharmaceutical company money at one end, and shoots lectures about social justice out the other.

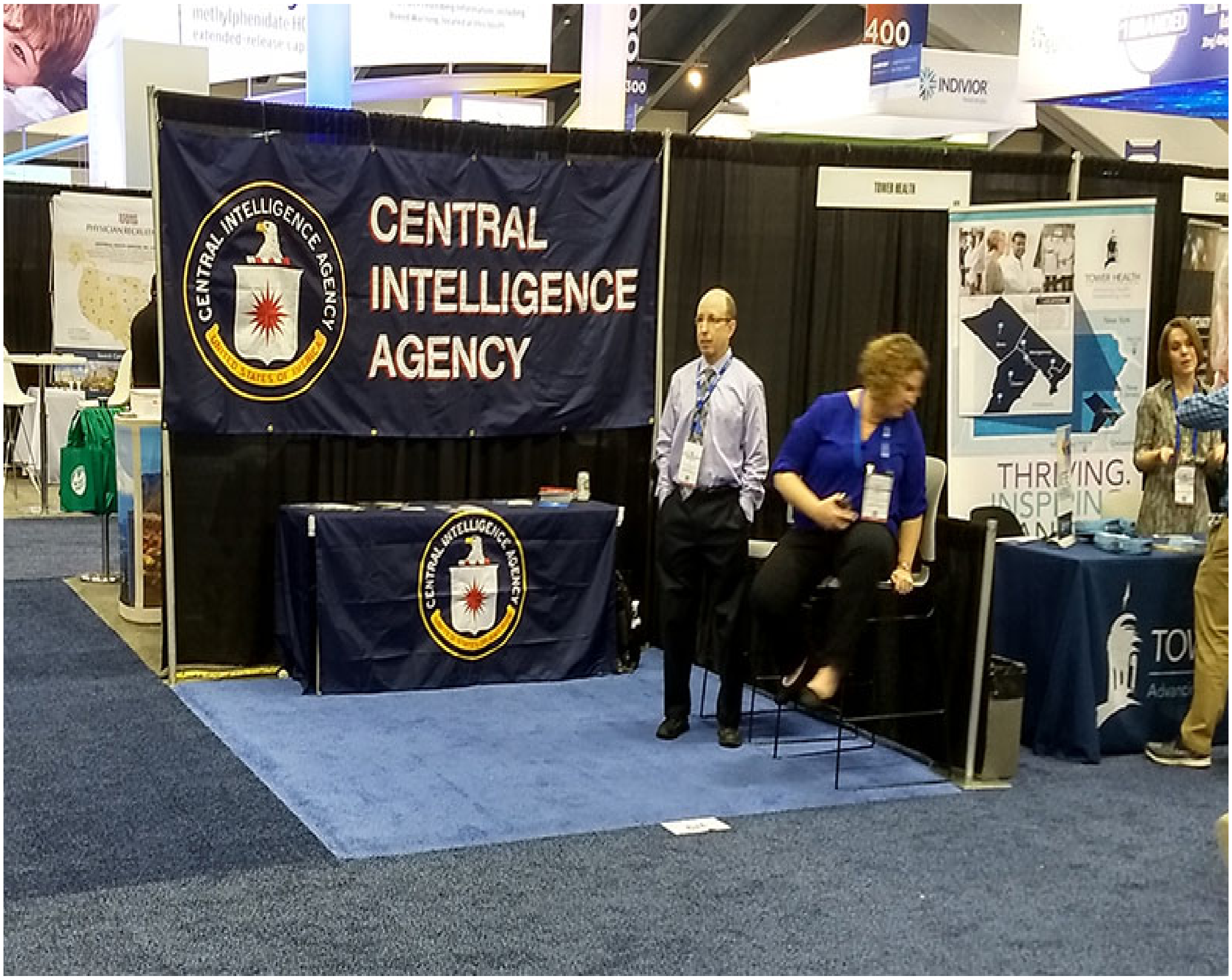
The fourth thing you notice at the American Psychiatric Association meeting is the Scientologists protesting outside.


AMERICAN PSYCHIATRIC ASSOCIATION
ANNUAL MEETING
175 YEARS of APA
STOP THE TORTURE.
BAN ELECTROSHOCK TREATMENT.
CCHR.ORG/ECT
CITIZENS COMMISSION ON HUMAN RIGHTS
BAN ELECTROSHOCK

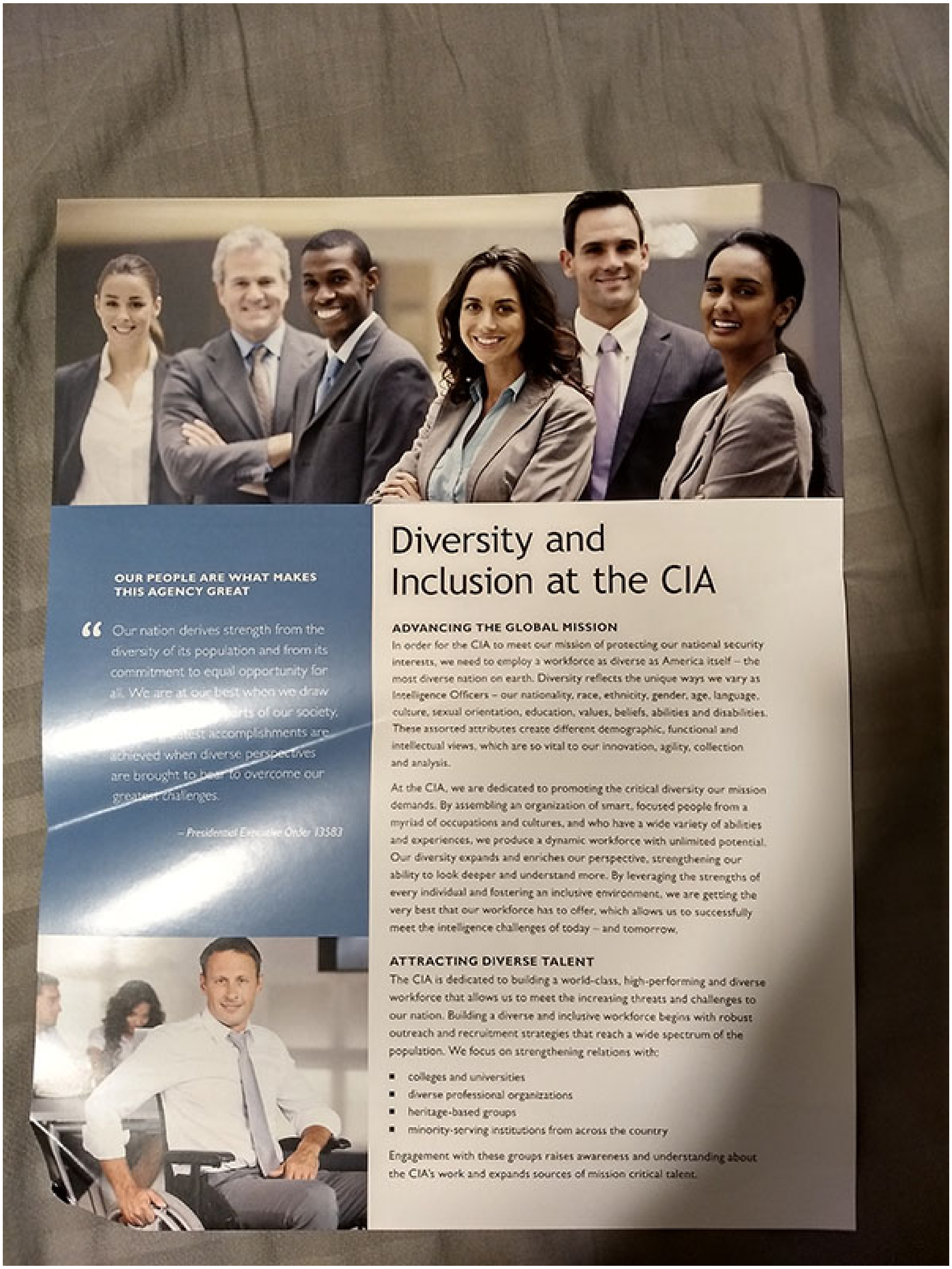
They don’t tell you they’re Scientologists. But their truck has a link to CCHR.org on it, and Wikipedia confirms them as a Scientology front group. Scientology has a long-standing feud with psychiatry, with the psychiatrists alleging that Scientology is a malicious cult, and the Scientologists alleging that psychiatry is an evil pseudoscience that denies the truth of dianetics. And that psychiatrists helped inspire Hitler. And that the 9/11 was masterminded by Osama bin Laden’s psychiatrist. And that psychiatrists are plotting to institute a one-world government. And that psychiatrists are malevolent aliens from a planet called Farsec. Really they have a lot of allegations.

This particular truck is especially sad, because they're reinforcing the myths about electroconvulsive therapy. ECT is a very effective treatment for depression. It is essentially always consensual – although most other psychiatric treatments can be administered involuntarily if someone is judged too out-of-touch with reality to make decisions, ECT has a special status as a treatment which can only be given with patient permission. It's always performed under anaesthesia and muscle relaxants, so patients are not conscious during the procedure and not spasming. And it can be a life-changing option for treatment-resistant depression. See this *Scientific American* article for more.

The fifth thing you notice at the American Psychiatric Association meeting is that the CIA has set up a booth.


CENTRAL INTELLIGENCE AGENCY
CENTRAL INTELLIGENCE AGENCY
UNITED STATES OF AMERICA
400
INDIVIOR
TOWER HEALTH
THRIVING. INSPIRING.
methylphenidate HCl extended-release cap

I was pretty curious about what the CIA wanted from psychiatrists (did they lose the original MKULTRA data? do they need to gather more?), but I was too shy to ask their representative directly. I did take one of their flyers, but it turned out to just be about how woke they were:



OUR PEOPLE ARE WHAT MAKES THIS AGENCY GREAT

> “Our nation derives strength from the diversity of its population and from its commitment to equal opportunity for all. We are at our best when we draw [illegible] parts of our society. [illegible] accomplishments are achieved when diverse perspectives are brought to bear to overcome our greatest challenges.
>
> *– Presidential Executive Order 13583*

Diversity and Inclusion at the CIA

ADVANCING THE GLOBAL MISSION

In order for the CIA to meet our mission of protecting our national security interests, we need to employ a workforce as diverse as America itself – the most diverse nation on earth. Diversity reflects the unique ways we vary as Intelligence Officers – our nationality, race, ethnicity, gender, age, language, culture, sexual orientation, education, values, beliefs, abilities and disabilities. These assorted attributes create different demographic, functional and intellectual views, which are so vital to our innovation, agility, collection and analysis.

At the CIA, we are dedicated to promoting the critical diversity our mission demands. By assembling an organization of smart, focused people from a myriad of occupations and cultures, and who have a wide variety of abilities and experiences, we produce a dynamic workforce with unlimited potential. Our diversity expands and enriches our perspective, strengthening our ability to look deeper and understand more. By leveraging the strengths of every individual and fostering an inclusive environment, we are getting the very best that our workforce has to offer, which allows us to successfully meet the intelligence challenges of today – and tomorrow.

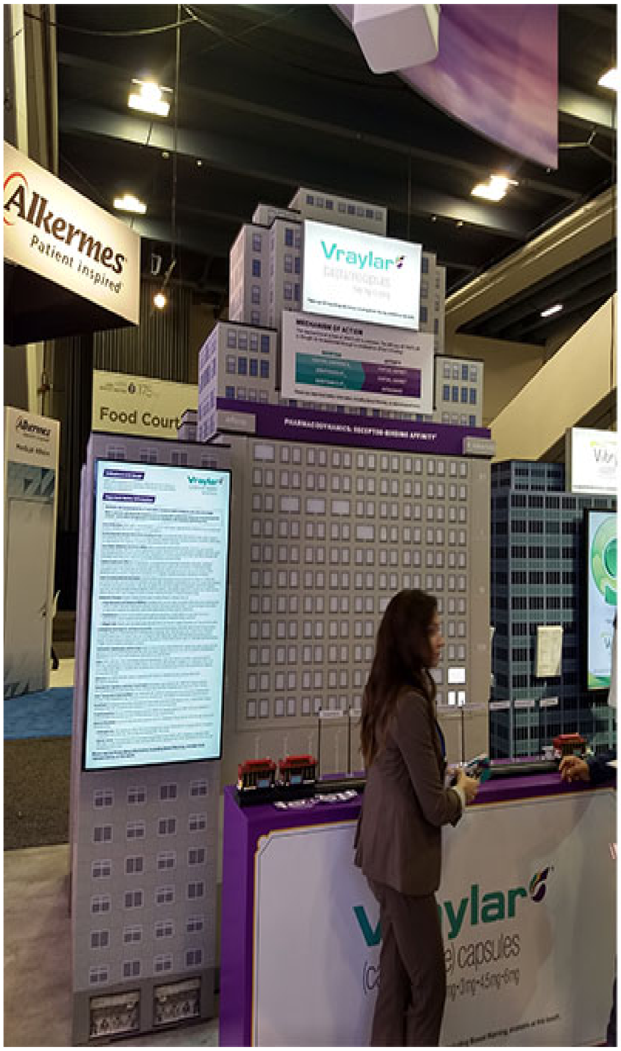
ATTRACTING DIVERSE TALENT

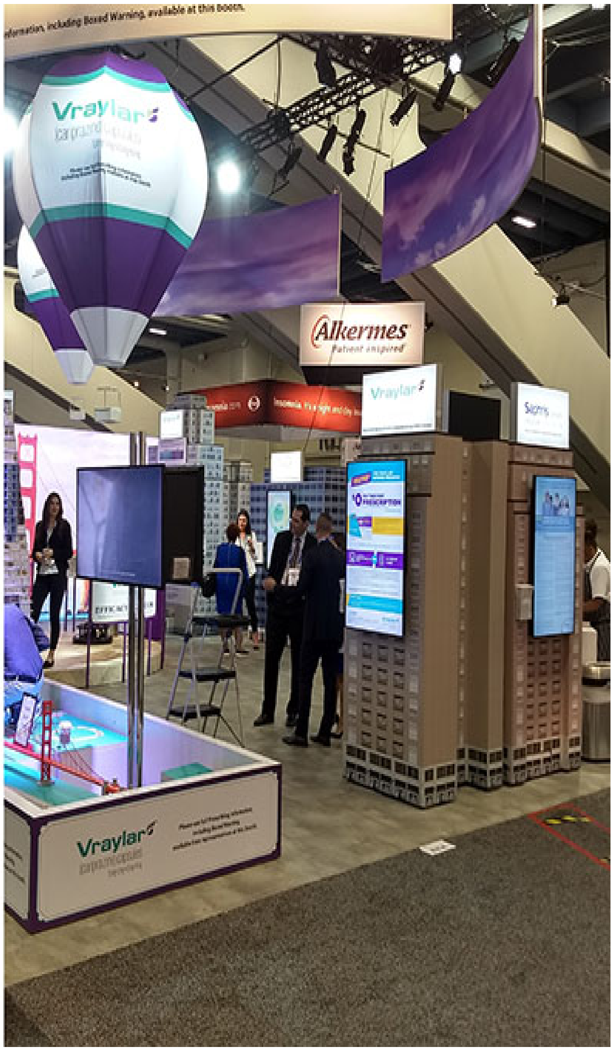
The CIA is dedicated to building a world-class, high-performing and diverse workforce that allows us to meet the increasing threats and challenges to our nation. Building a diverse and inclusive workforce begins with robust outreach and recruitment strategies that reach a wide spectrum of the population. We focus on strengthening relations with:

- colleges and universities
- diverse professional organizations
- heritage-based groups
- minority-serving institutions from across the country

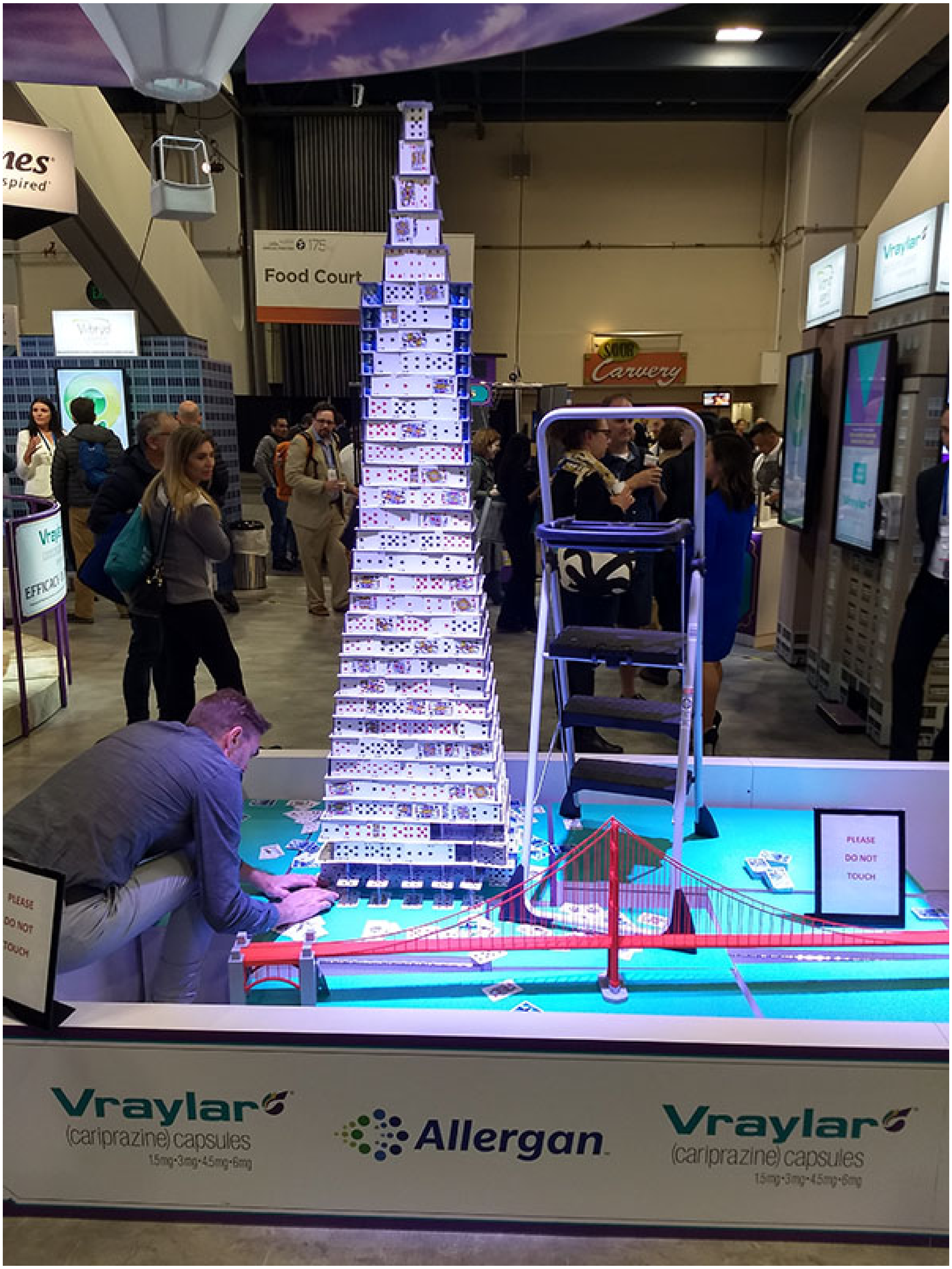
Engagement with these groups raises awareness and understanding about the CIA's work and expands sources of mission critical talent.

The sixth thing you notice at the American Psychiatric Association meeting is that Vraylar® has built an entire miniature city. The buildings are plastered with pamphlets on Vraylar®. Billboards advertising Vraylar® hang over the streets and bridges. Giant Vraylar balloons hover serenly over everything, looking down with contempt and sorrow upon the non-Vraylar®-prescribing world below.


Alkermes
Patient inspired
Vraylar
MECHANISM OF ACTION
Food Court
PHARMACODYNAMICS: RECEPTOR-BINDING AFFINITY
Vraylar
capsules


Information, including Boxed Warning, available at this booth.
Vraylar
Alkermes
Patient inspired
Vraylar
PRESCRIPTION
Vraylar

Occupying pride of place in city center, some sort of Important Vraylar Scientist is constructing the Transamerica Pyramid out of playing cards.


Food Court
Sour Carvery
PLEASE DO NOT TOUCH
PLEASE DO NOT TOUCH
Vraylar (cariprazine) capsules 1.5mg·3mg·4.5mg·6mg
Allergan
Vraylar (cariprazine) capsules 1.5mg·3mg·4.5mg·6mg

I dunno, if I were working in an area where the research supporting a treatment has a tendency to collapse suddenly and spectacularly, I might want to *avoid* building an association in people’s minds between my medication and a house of cards. But the ways of Vraylar® are inscrutable to mortal men.

The seventh thing you notice at the American Psychiatric Association meeting is that many of the new drugs are ridiculous.

It’s hard to blame pharmaceutical companies for this. The return on investment for pharma R&D is rapidly shrinking – drug discovery is too expensive to consistently make money anymore.

Rather than give up and die, pharma is going all in on newer, me-too-er me-too drugs. The current business plan looks kind of like this:

1. Take an popular older drug

2. Re-invent it, either with a minor change to the delivery mechanism, or by finding a similar molecule that works the same way

3. Call this a new drug, advertise the hell out of it, and sell it for 10x – 100x the price of the older drug

4. Profit!

Consider Lucemyra®:

The LUCEMYRA Co-Pay Assistance Program
POWERED BY

RelayHealth

eVoucherRx™ is a trademark of RelayHealth®.

eRx Network

Lucemyra™
(lofexidine) tablets 0.18 mg

Please visit LUCEMYRA.com for more information.

US WorldMeds

US WorldMeds, LLC
4441 Springdale Road
Louisville, KY 40241

Under License from Britannia Pharmaceuticals Ltd.

US WorldMeds, LLC is the exclusive licensee and distributor of LUCEMYRA in the United States and its territories.

©2018. LUCEMYRA is a trademark of US WorldMeds, LLC.
USWMLUC-00033-0718

It’s an alpha-2a receptor agonist used to treat acute opiate withdrawal. Alpha-2a receptor agonists are a fine choice for acute opiate withdrawal, but we already have one that works great: clonidine. Clonidine costs $4.84 per month. Lucemyra® costs $1,974.78. Is there any difference at all between the two medications? Some studies suggest maybe lofexedine can cause less hypotension, but realistically we throw random doses of clonidine at ADHD kids all the time, so it’s not like clonidine-induced hypotension is some kind of giant menace which will destroy us all.

I asked the Lucemyra® representative why I might prescribe Lucemyra® instead of clonidine for opiate withdrawal. She said it was because Lucemyra® is FDA-approved for this indication, and clonidine isn’t. This is the same old story as Rozerem® vs. melatonin, Lovaza® vs. fish oil, and Spravato® vs. ketamine. As long as doctors continue to outsource their thinking to the FDA approval process, in a way even the FDA itself doesn’t endorse, pharma companies will be able to inflate the prices of basic medications by a thousand times just by playing games with the bureaucracy.

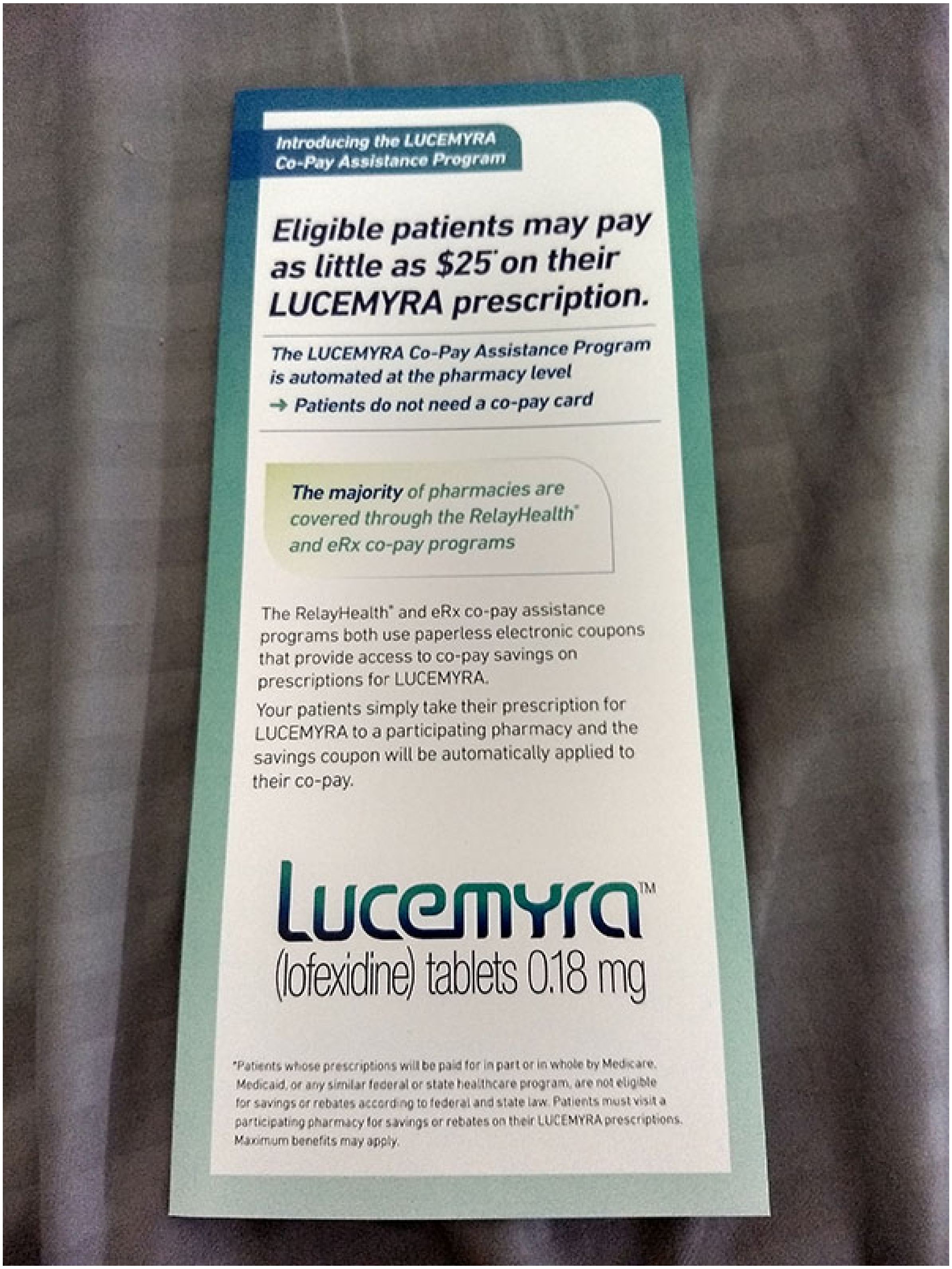
But also:

Please see Important Safety Information and Full Prescribing Information, including Boxed Warning, located at this booth.
Jornay PM
methylphenidate HCl CII
The first and only ADHD stimulant dosed in the evening
Help your patients wake up ready for the day
Jornay PM
methylphenidate HCl CII
extended-release capsules
Jornay PM
methylphenidate HCl CII
extended-release capsules
Are you made for

Jornay® is a new form of methylphenidate, ie Ritalin. The usual comparison: a month of Ritalin costs $25.19, a month of Jornay® costs $387.48. What's the difference? You can take Jornay® at night. Why is this interesting? The Jornay® representatives say that maybe people want to have Ritalin in their system as soon as they wake up, rather than having to wait the half-hour or so it usually takes for it to start having an effect. I have to admit, from a scientific perspective Jornay® is kind of cool; I expect the pharmacologists who designed it had a lot of fun. But the oppressed people of the world haven’t exactly been crying out for Dark Ritalin. Nobody has been saying “Help us, pharmaceutical industry, merely having Ritalin®, Concerta®, Metadate®, Focalin®, Daytrana®, Quillivant®, Quillichew®, Aptensio®, Biphentin®, Equasym®, Medikinet®, and Rubifen® just isn’t enough for us! We need more forms of Ritalin, *stat*!”

My favorite was Subvenite®, which is just lamotrigine in a conveniently-packaged box that tells you how much to take each day. The same amount of normal lamotrigine would cost about $12; it’s hard for me to figure out exactly how much Subvenite® costs, but this site suggests $540. To be fair, lamotrigine is a really inconvenient drug whose dosing schedule often leaves patients confused. To be less fair, seriously, $540 for some better instructions? Get a life.

How do all these people keep doing it? What’s their business plan? Here’s a hint:



Introducing the LUCEMYRA Co-Pay Assistance Program

Eligible patients may pay as little as $25* on their LUCEMYRA prescription.

The LUCEMYRA Co-Pay Assistance Program is automated at the pharmacy level

→ Patients do not need a co-pay card

The majority of pharmacies are covered through the RelayHealth® and eRx co-pay programs

The RelayHealth® and eRx co-pay assistance programs both use paperless electronic coupons that provide access to co-pay savings on prescriptions for LUCEMYRA.

Your patients simply take their prescription for LUCEMYRA to a participating pharmacy and the savings coupon will be automatically applied to their co-pay.

Lucemyra™
(lofexidine) tablets 0.18 mg

*Patients whose prescriptions will be paid for in part or in whole by Medicare, Medicaid, or any similar federal or state healthcare program, are not eligible for savings or rebates according to federal and state law. Patients must visit a participating pharmacy for savings or rebates on their LUCEMYRA prescriptions. Maximum benefits may apply.

This is the brochure for Lucemyra®, the opiate withdrawal medication that costs $1,974.78. No patient is paying

$1,974.78 for it. Patients are paying $25. And doctors sure aren’t paying $1,974.78. The way all these companies are getting away with it is because in Healthcaristan SSR, nobody ever pays for their own medication.

To a first approximation, doctors make purchasing decisions, but insurances cough up the money. Insurances have a few weapons to prevent doctors from buying arbitrarily expensive drugs, but they tend to back off in the face of magic words like “I believe this is medically necessary” or “This is the one the FDA approved”. So to fill in the missing pieces of the pharma strategy mentioned above:

1. Take an popular older drug

2. Re-invent it, either with a minor change to the delivery mechanism, or by finding a similar molecule that works the same way

3. Call this a new drug, advertise the hell out of it, and sell it for 10x – 100x the price of the older drug

4. Advertise it to patients (who don’t have to pay for it) and doctors (who definitely don’t have to pay for it), neither of whom care at all what price you’re setting.

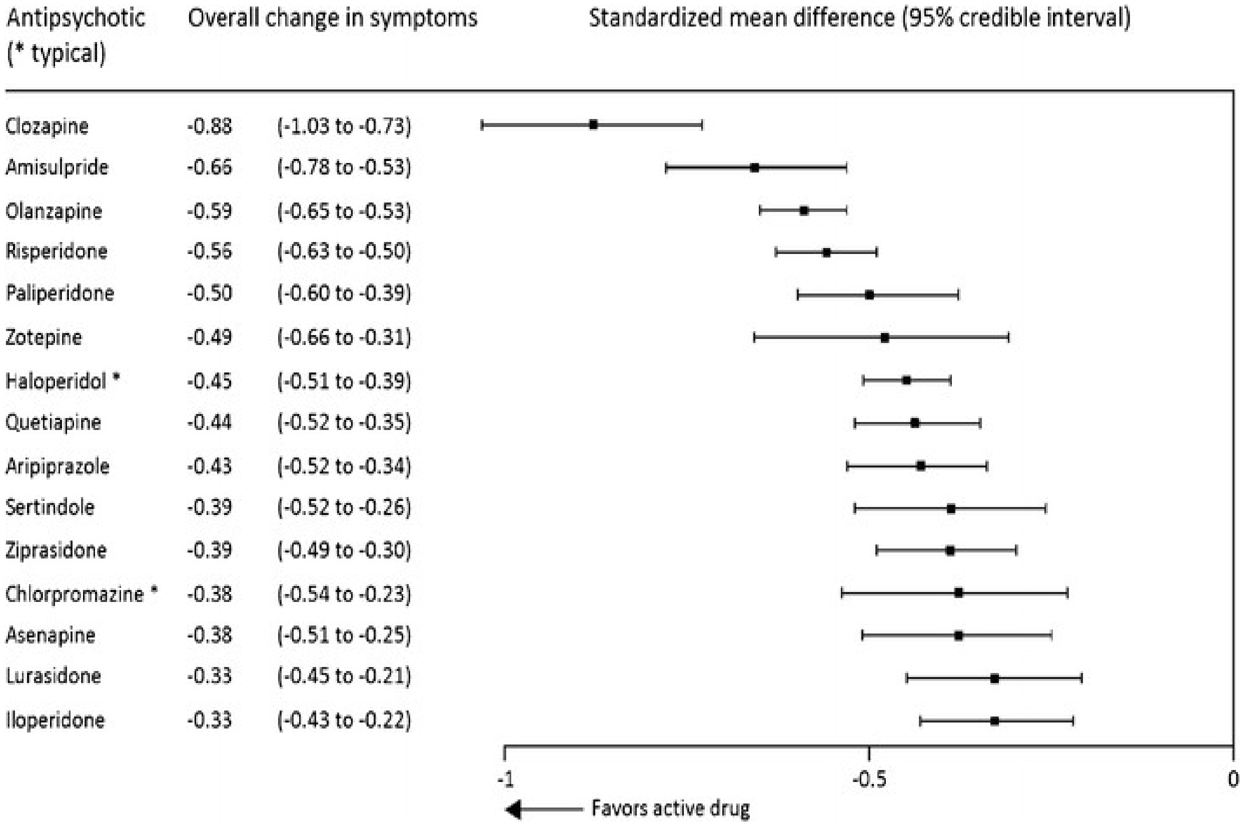
5. Make sure doctors know the magic words they need to use to force insurance companies to pay for it.

6. Profit!

This has become so lucrative that pharma companies barely have to do any real research and development at all these days. The only genuinely exciting new drugs at the conference were Ingrezza® and Austedo®, both of which treat tardive dyskinesia – a side effect you get from having been on too many other psychiatric drugs. This is probably a metaphor for something.

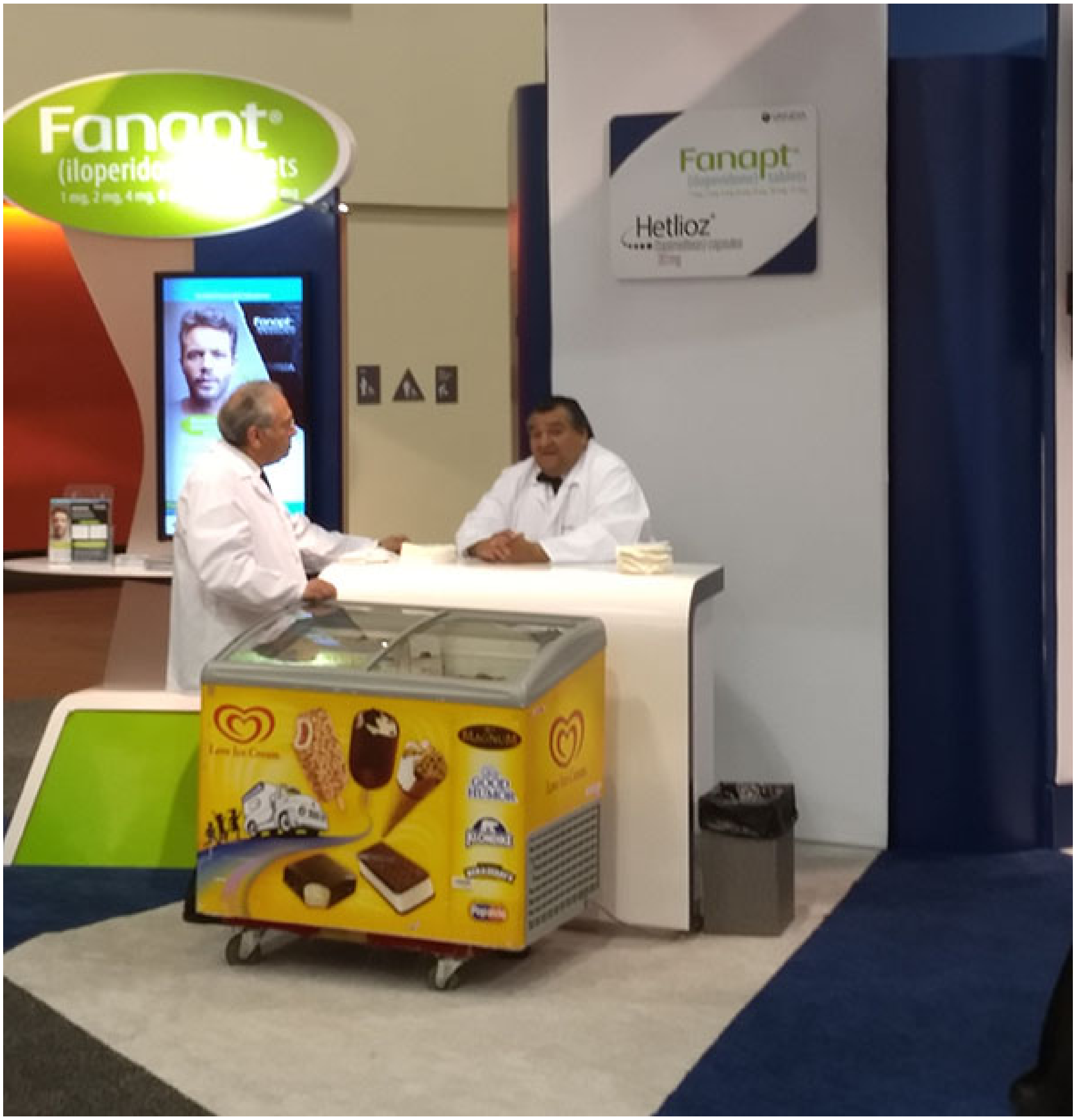
The eighth thing you notice at the American Psychiatric Association meeting is that there’s a presentation called *“Yer A Psychiatrist, Harry!”: Learning Psychiatric Concepts Through The Fictional Worlds Of Game Of Thrones And Harry Potter*. I didn’t go. I realize I have failed you, my readers, but if I had to listen to ninety minutes of that, all the Vraylar® in the world would not be enough to maintain my sanity.

The ninth thing you notice at the American Psychiatric Association meeting is that, after winning last place in a head-to-head comparison of various antipsychotics, doing worse than drugs that cost less than 1% as much…



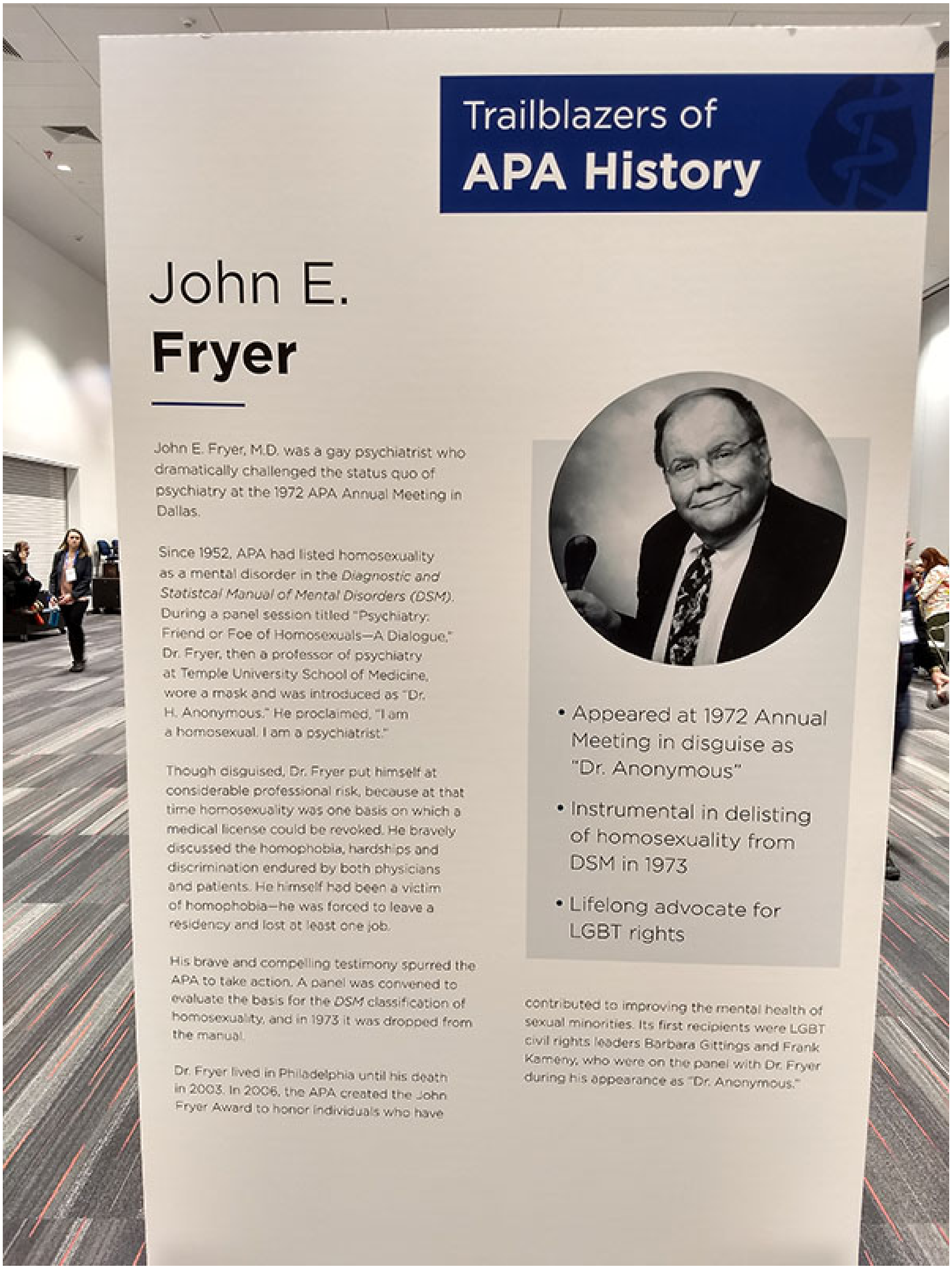
Antipsychotic (* typical)	Overall change in symptoms	
Clozapine	-0.88	(-1.03 to -0.73)
Amisulpride	-0.66	(-0.78 to -0.53)
Olanzapine	-0.59	(-0.65 to -0.53)
Risperidone	-0.56	(-0.63 to -0.50)
Paliperidone	-0.50	(-0.60 to -0.39)
Zotepine	-0.49	(-0.66 to -0.31)
Haloperidol *	-0.45	(-0.51 to -0.39)
Quetiapine	-0.44	(-0.52 to -0.35)
Aripiprazole	-0.43	(-0.52 to -0.34)
Sertindole	-0.39	(-0.52 to -0.26)
Ziprasidone	-0.39	(-0.49 to -0.30)
Chlorpromazine *	-0.38	(-0.54 to -0.23)
Asenapine	-0.38	(-0.51 to -0.25)
Lurasidone	-0.33	(-0.45 to -0.21)
Iloperidone	-0.33	(-0.43 to -0.22)

…Fanapt® (iloperidone) has pivoted to a marketing strategy of bribing doctors with free ice cream:


Fanapt®
Fanapt
Hetlioz
Magnum
Good Humor
Klondike

The tenth thing you notice at the American Psychiatric Association meeting is that all of this has happened before.

This is the 175th anniversary of the APA. It’s been a pretty crazy century-and-three-quarters, no pun intended. Like, seriously, take a look at this guy:



Trailblazers of
APA History

John E.
Fryer

John E. Fryer, M.D. was a gay psychiatrist who dramatically challenged the status quo of psychiatry at the 1972 APA Annual Meeting in Dallas.

Since 1952, APA had listed homosexuality as a mental disorder in the *Diagnostic and Statistcal Manual of Mental Disorders (DSM)*. During a panel session titled "Psychiatry: Friend or Foe of Homosexuals—A Dialogue," Dr. Fryer, then a professor of psychiatry at Temple University School of Medicine, wore a mask and was introduced as "Dr. H. Anonymous." He proclaimed, "I am a homosexual. I am a psychiatrist."

Though disguised, Dr. Fryer put himself at considerable professional risk, because at that time homosexuality was one basis on which a medical license could be revoked. He bravely discussed the homophobia, hardships and discrimination endured by both physicians and patients. He himself had been a victim of homophobia—he was forced to leave a residency and lost at least one job.

His brave and compelling testimony spurred the APA to take action. A panel was convened to evaluate the basis for the *DSM* classification of homosexuality, and in 1973 it was dropped from the manual.

Dr. Fryer lived in Philadelphia until his death in 2003. In 2006, the APA created the John Fryer Award to honor individuals who have contributed to improving the mental health of sexual minorities. Its first recipients were LGBT civil rights leaders Barbara Gittings and Frank Kameny, who were on the panel with Dr. Fryer during his appearance as "Dr. Anonymous."

- Appeared at 1972 Annual Meeting in disguise as "Dr. Anonymous"
- Instrumental in delisting of homosexuality from DSM in 1973
- Lifelong advocate for LGBT rights

Back when you could still lose your medical license for being gay, he went to the APA meeting *in a mask* and gave a presentation arguing for gay rights, and the APA de-listed homosexuality as a psychiatric disorder the following year. How amazing is that?

The APA highlighted a bunch of people like this, heroes and trailblazers all. But for every great hero celebrated on posters, there is an embarrassment buried somewhere deep in an archive. My favorite of these is the APA Presidential Address from 1918, the very tail end of WWI. The head of the Association, a very distinguished psychiatrist named Dr. Anglin, gets up in front of the very same conference I attended this week (the 1918 version was held in Chicago) and declared that the greatest problem facing psychiatry was…the dastardly Hun:

> The maxim that medical science knows no national boundaries has been rudely shaken by the war. The Fatherland has been preparing for isolation from the medical world without its confines. Just as, years ago, the Kaiser laid his ban on French words in table menus, so, as early as 19 14, German scientists embarked on a campaign against all words which had been borrowed from an enemy country. A purely German medical nomenclature was the end in view. The rest of the world need not grieve much if they show their puerile hate in this way. It will only help to stop the tendency to Pan-Germanism in medicine which has for some years past been gaining headway. ‘
>
> The Germans excel all other nations in their genius for advertising themselves. They have proved true the French proverb that one is given the standing he claims. On a slender basis of achievement they have contrived to impress themselves as the most scientific nation. Never was there greater imposture. They display the same cleverness in foisting on a gullible world their

scientific achievements as their shoddy commercial wares. The two are of much the same value, made for show rather than endurance — in short, made in Germany […]

In the earliest months of the war it was pointed out that there are tendencies in the evolution of medicine as a pure science as it is developed in Germany which are contributing to the increase of charlatanism of which we should be warned. A medical school has two duties — one to medical science, the other to the public. The latter function is the greater, for out of every graduating class 90 per cent. are practitioners and less than 10 per cent, are scientists. The conditions in Germany are reversed. There, there were ninety physicians dawdling with science to every ten in practice. Of these 90, fully 75 per cent were wasting their time. In Germany the scientific side is over-done, and they have little to show for it all, while the human side is neglected. Even in their new institutions, splendid as they are in a material sense, it is easily seen that the improved conditions are not for the comfort of the patients.

Out of this war some modicum of good may come if it leads to a revision of the exaggerated estimate that has prevailed in English-speaking countries of the achievements of the Germans in science. We had apparently forgotten the race that had given the world Newton, Faraday, Stephenson, Lister, Hunter, Jenner, Fulton, Morse, Bell, Edison, and others of equal worth. German scientists wait till a Pasteur has made the great discovery, on which it is easy for her trained men to work. She shirks getting for herself a child through the gates of sacrifice and pain ; but steals a babe, and as it grows bigger under her care, boasts herself as more than equal to the mother who bore it. Realising her mental

> sterility, drunk with self-adoration, she makes insane war on the nations who still have the power of creative thought.
>
> But it is especially in the realm of mental science that the reputation of the Germans is most exalted and is least deserved. For every philosopher of the first rank that Germany has produced, the English can show at least three. And in psychiatry, while we have classical writings in the English tongue, and men of our own gifted with clinical insight, we need seek no foreign guides, and can afford to let the abounding nonsense of Teutonic origin perish from neglect of cultivation.
>
> The Germans are shelling Paris from their Gothas and their new gun. Murdering innocents, to create a panic in the heart of France! With what effect ? The French army cries the louder, “They shall not pass ” ; Paris glows with pride to be sharing the soldiers’ dangers, and increases its output of war material; and the American army sees why it is in France, and is filled with righteous hatred. Panic nowhere. Vengeance everywhere. What does the Hun know of psychology? His most stupid, thick-witted performance was his brutal defiance of the United States with its wealth, resources, and energy. That revealed a mental condition both grotesque and pitiable.
>
> After the war a centre of medical activity will be found on this side the Atlantic, and those who have watched the progress medical science has made in the United States will have no misgivings as to your qualifications for leadership. If we learn to know ourselves, great good will come out of this war.

Anglin does not deny that some may find it inappropriate to discuss politics at a psychiatry conference, but notes that:

> If in these introductory remarks I have not been able to detach myself from the world’s most serious business at the present time, perhaps on reflection they may not have gone very far afield from the subject which binds us together in an association. If there is to be a change in the conditions under which we live this must have its effect on the minds of men ; whether for good or ill, I will not stop to speculate. We are intensely concerned with environment. This war itself is entangled with it,
>
> England’s greatness, her devotion to honour, truth, and fidelity, is due to the environment in which her children are trained and grow to manhood. The ivy-grown wall, the vine-clad hills and the rose-covered bowers constitute the birth-place of English character.
>
> Gerard tells us the cause of the war is the uncongenial environment in which the German youth is cradled and reared. The leaden skies for which Prussia is noted, its bleak Baltic winds, the continuous cold, dreary rains, the low-lying land, and the absence of flowers have tended to harden the spirit and rob it of its virtue, produce a sullen and morose character, curdling the milk of human kindness.

He does raise one warning, one problem that risks sabotaging even countries as congenial-climate-having as ourselves and our allies:

> The quack medicine vendor is busier than ever. Money is plenty and he wants some of it. He uses mental suggestion and interests us. He is a specialist in distortion who probes into the ordinary sensations of healthy people and perverts them into symptoms. Every billboard, newspaper, fence-rail, barn and rock thrusts out a suggestion of sickness as never before. The only vulnerable point to attack the vicious traffic is the

> advertising. If governments forbid that as they should, the next generation will be healthier and richer.

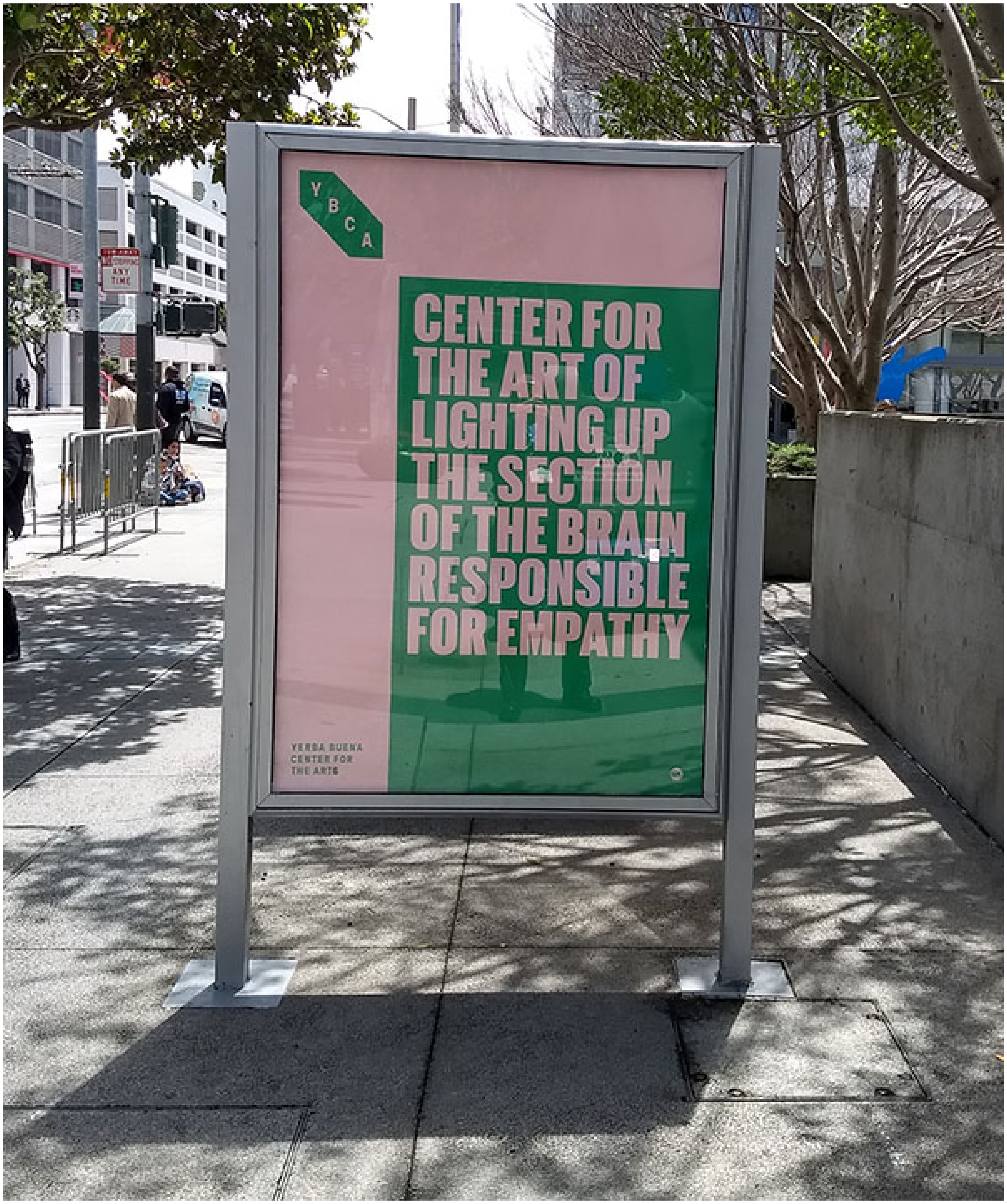
From Dr. Anglin’s address, I gather three things.

First, the billboards we shall always have with us. It’s easy to imagine this a modern problem, but apparently the generation that confronted the Kaiser was confronting annoying psychiatric advertising too. The Kaiser is gone; the annoying psychiatric advertising has proven a tougher foe.

Second, psychiatry has always been the slave of the latest political fad. It is just scientific enough to be worth capturing, but not scientific enough to resist capture. The menace *du jour* will always be a threat to our mental health; the salient alternative to “just forcing pills down people’s throat” will always be pursuing the social agenda of whoever is in power; you will always be able to find psychiatrists to back you up on this.

But third, science advances anyway. Psychiatry is light-years ahead of where it was a hundred years ago. Since Dr. Anglin’s 1918 address, we’ve discovered psychotherapy and psychopharmacology; come up with deinstitutionalization and destigmatization; and put rights in place to protect psychiatric patients and to protect the general public from being unnecessarily psychiatrized. We’ve even invented Vraylar®.

On my way out of the conference, I encountered this ad:


YBCA
CENTER FOR THE ART OF LIGHTING UP THE SECTION OF THE BRAIN RESPONSIBLE FOR EMPATHY
YERBA BUENA CENTER FOR THE ARTS
ANY TIME

I don’t think it was even related to the psychiatry conference. I think it was for a nearby art museum. But it struck me. It struck me because it’s the sort of picture psychiatry wants to have of itself, a combination of hard neuroscience and basic human goodness. It struck me because as written, it’s obviously bogus (which Brodmann area is responsible for empathy again? How bright does it have to light up before you start feeling empathic?) in much the same way psychiatry can be obviously bogus (how much Vraylar® does it take before you can “take back control of your life” or “feel better than well”?), but is sort of an exaggerated and slightly-too-literal version of something that could potentially *not* be bogus. It struck me because, after making fun of it, I had to admit to myself that the thing it was pointing at was good and important and probably exactly what an art museum should be trying to do. And a psychiatrist, for that matter.

START A
POSITIVE REACTION

Postscript To APA Photo-Essay

I was surprised how many people responded to my APA photo-essay with comments like "Seems psychiatry as a field is broken beyond repair" or "This proves you should never trust psychiatrists".

The mood I was going for was more "let's share a laugh at the excesses of the profession" than "everything must be burned down". Looks like I missed it.

I was disappointed to see a lot of the most hostile comments coming from people in tech. It would be easy to write an equally damning report on the tech industry. Just cobble together a few paragraphs about Juicero and Theranos, make fun of whatever weird lifestyle change @jack is supporting at the moment, and something something Zuckerberg something Cambridge Analytica something. You can even throw in something about James Damore (if you're writing for the left) or about the overreaction to James Damore (if you're writing for the right). And there you go! Tech is a malicious cancerous industry full of awful people and everyone should hate it. We've all read this exact thinkpiece a thousand times.

I've tried to push back against this line of thinking. A lot of the most visible and famous things in tech are bad, because scum tends to rise to the top. But there's also some extraordinary innovation going on, and some extraordinarily good people involved. "@jack invents new health fad of rolling around naked on glaciers" is a much juicier story than "we can now fit twice as many billions of transistors on a chip as we could last year", but tech journalism that only reports on the former is missing an important part of the story.

I feel the same way about psychiatry. There’s a lot of cringeworthy stuff going on at conferences, but conferences are designed to be about signaling and we shouldn’t expect otherwise. There’s also a lot of great people working really hard to help fight mental illness and support the mentally ill. “Most Americans remain alive and basically functional despite record-breaking amounts of depression and anxiety” isn’t sexy any more than “Internet continues to connect billions of people around the world at the speed of light” is sexy. But it’s a much bigger part of the story than the part where silly people do silly things at conferences.

Michael Crichton invented the term Gell-Mann Amnesia as a reference to Nobel-winning physicist Murray Gell-Mann, who remarked that even though the newspapers were always wrong about his area of expertise (physics), he always found himself trusting them about everything else. If you don’t trust hatchet jobs against tech, try to be a little less credulous of hatchet jobs against psychiatry, even when I’m the one doing the hatcheting.

(also, RIP Murray Gell-Mann, who passed away last week – at least if you believe obituaries)

Bay Area SSC Meetup 6/2

Join me at 5 PM on Sunday, June 2, for the traditional once-every-three-months big SSC Bay Area meetup. For a change of pace, it'll be in **San Francisco** this time around.

Meet on the grass near the tennis courts at Mission Dolores Park (near the 16th Street Mission BART station).

EDIT: Blue tarp near restrooms near Hancock Street

I understand this conflicts with some other rationalist community events, but there are events every weekend for the next month or so, and I don't want to not have meetups. We'll probably be at the park until at least 7 or so, so feel free to come late.

Lurkers/newbies/people-who-aren't-sure-if-they're-welcome-or-not are welcome!

[REPOST] Epistemic Learned Helplessness

[This is a slightly edited repost of an essay from my old LiveJournal]

A friend recently complained about how many people lack the basic skill of believing arguments. That is, if you have a valid argument for something, then you should accept the conclusion. Even if the conclusion is unpopular, or inconvenient, or you don't like it. He envisioned an art of rationality that would make people believe something *after it had been proven to them*.

And I nodded my head, because it sounded reasonable enough, and it wasn't until a few hours later that I thought about it again and went "Wait, no, that would be a terrible idea."

I don't think I'm overselling myself too much to expect that I could argue circles around the average uneducated person. Like I mean that on most topics, I could demolish their position and make them look like an idiot. Reduce them to some form of "Look, everything you say fits together and I can't explain why you're wrong, I just know you are!" Or, more plausibly, "Shut up I don't want to talk about this!"

And there are people who can argue circles around me. Maybe not on *every* topic, but on topics where they are experts and have spent their whole lives honing their arguments. When I was young I used to read pseudohistory books; Immanuel Velikovsky's Ages in Chaos is a good example of the best this genre has to offer. I read it and it seemed so obviously correct, so perfect, that I could barely bring myself to bother to search out rebuttals.

And then I read the rebuttals, and they were so obviously correct, so devastating, that I couldn't believe I had ever been so dumb as to believe Velikovsky.

And then I read the rebuttals to the rebuttals, and they were so obviously correct that I felt silly for ever doubting.

And so on for several more iterations, until the labyrinth of doubt seemed inescapable. What finally broke me out wasn't so much the lucidity of the consensus view so much as starting to sample different crackpots. Some were almost as bright and rhetorically gifted as Velikovsky, all presented insurmountable evidence for their theories, and all had mutually exclusive ideas. After all, Noah's Flood couldn't have been a cultural memory both of the fall of Atlantis *and* of a change in the Earth's orbit, let alone of a lost Ice Age civilization or of megatsunamis from a meteor strike. So given that at least some of those arguments are wrong and all seemed practically proven, I am obviously just gullible in the field of ancient history. Given a total lack of independent intellectual steering power and no desire to spend thirty years building an independent knowledge base of Near Eastern history, I choose to just accept the ideas of the prestigious people with professorships in Archaeology, rather than those of the universally reviled crackpots who write books about Venus being a comet.

You could consider this a form of epistemic learned helplessness, where I know any attempt to evaluate the arguments is just going to be a bad idea so I don't even try. If you have a good argument that the Early Bronze Age worked completely differently from the way mainstream historians believe, I just don't want to hear about it. If you insist on telling me anyway, I will nod, say that your argument makes complete sense, and then totally refuse to change my mind or admit even the slightest possibility that you might be right.

(This is the correct Bayesian action: if I know that a false argument sounds just as convincing as a true argument, argument convincingness provides no evidence either way. I should ignore it and stick with my prior.)

I consider myself lucky in that my epistemic learned helplessness is circumscribed; there are still cases where I'll trust the evidence of my own reason. In fact, I trust it in most cases other than infamously deceptive arguments in fields I know little about. But I think the average uneducated person doesn't and shouldn't. Anyone anywhere – politicians, scammy businessmen, smooth-talking romantic partners – would be able to argue them into anything. And so they take the obvious and correct defensive maneuver – they will never let anyone convince them of any belief that sounds "weird".

(and remember that, if you grow up in the right circles, beliefs along the lines of "astrology doesn't work" sound "weird".)

This is starting to resemble ideas like compartmentalization and taking ideas seriously. The only difference between their presentation and mine is that I'm saying that for 99% of people, 99% of the time, taking ideas seriously is the wrong strategy. Or, at the very least, it should be the last skill you learn, after you've learned every other skill that allows you to know which ideas are or are not correct.

The people I know who are best at taking ideas seriously are those who are smartest and most rational. I think people are working off a model where these co-occur because you need to be very clever to resist your natural and detrimental tendency not to take ideas seriously. But I think they might instead co-occur because you have to be really smart in order for taking ideas seriously not to be immediately disastrous. You have to be really smart not to have been talked into enough terrible arguments to develop epistemic learned helplessness.

Even the smartest people I know have a commendable tendency not to take certain ideas seriously. Bostrom's simulation argument, the anthropic doomsday argument, Pascal's Mugging – I've never heard anyone give a coherent argument against any of these, but I've also never met anyone who fully accepts them and lives life according to their implications.

A friend tells me of a guy who once accepted fundamentalist religion because of Pascal's Wager. I will provisionally admit that this person "takes ideas seriously". Everyone else gets partial credit, at best.

Which isn't to say that some people don't do better than others. Terrorists seem pretty good in this respect. People used to talk about how terrorists must be very poor and uneducated to fall for militant Islam, and then someone did a study and found that they were disproportionately well-off, college educated people (many were engineers). I've heard a few good arguments in this direction before, things like how engineering trains you to have a very black-and-white right-or-wrong view of the world based on a few simple formulae, and this meshes with fundamentalism better than it meshes with subtle liberal religious messages.

But to these I'd add that a sufficiently smart engineer has never been burned by arguments above his skill level before, has never had any reason to develop epistemic learned helplessness. If Osama comes up to him with a really good argument for terrorism, he thinks “Oh, there’s a good argument for terrorism. I guess I should become a terrorist,” as opposed to “Arguments? You can prove anything with arguments. I’ll just stay right here and not blow myself up.”

Responsible doctors are at the other end of the spectrum from terrorists here. I once heard someone rail against how doctors totally ignored all the latest and most exciting medical studies. The same person, practically in the same breath, then railed against how 50% to 90% of medical studies are wrong. These two observations are not unrelated. Not only are there so many terrible studies, but pseudomedicine (not the stupid homeopathy type, but the type that links everything to some obscure chemical on an out-of-the-way metabolic pathway) has, for me, proven much like pseudohistory – unless I am an expert in that particular subsubfield of medicine, it can sound very convincing even when it’s very wrong.

The medical establishment offers a shiny tempting solution. First, a total unwillingness to trust anything, no matter how plausible it sounds, until it’s gone through an endless cycle of studies and meta-analyses. Second, a bunch of Institutes and Collaborations dedicated to filtering through all these studies and analyses and telling you what lessons you should draw from them.

I’m glad that some people never develop epistemic learned helplessness, or develop only a limited amount of it, or only in certain domains. It seems to me that although these people are more likely to become terrorists or Velikovskians or homeopaths, they’re also the only people who can figure out if something basic and unquestionable is wrong, and make this possibility well-known enough that normal people start becoming willing to consider it.

But I’m also glad epistemic learned helplessness exists. It seems like a pretty useful social safety valve most of the time.

Book Review: The Secret Of Our Success

[Previously in sequence: Epistemic Learned Helplessness]

I.

"Culture is the secret of humanity's success" sounds like the most vapid possible thesis. *The Secret Of Our Success* by anthropologist Joseph Henrich manages to be an amazing book anyway.

Henrich wants to debunk (or at least clarify) a popular view where humans succeeded because of our raw intelligence. In this view, we are smart enough to invent neat tools that help us survive and adapt to unfamiliar environments.

Against such theories: we cannot actually do this. Henrich walks the reader through many stories about European explorers marooned in unfamiliar environments. These explorers usually starved to death. They starved to death in the middle of endless plenty. Some of them were in Arctic lands that the Inuit considered among their richest hunting grounds. Others were in jungles, surrounded by edible plants and animals. One particularly unfortunate group was in Alabama, and would have perished entirely if they hadn't been captured and enslaved by local Indians first.

These explorers had many advantages over our hominid ancestors. For one thing, their exploration parties were made up entirely of strong young men in their prime, with no need to support women, children, or the elderly. They were often selected for their education and intelligence. Many of them were from Victorian Britain, one of the most successful civilizations in history, full of geniuses like Darwin and Galton. Most of them had some past experience with wilderness craft and survival. But despite their big brains, when faced with the task our big brains supposedly evolved for – figuring out how to do hunting and gathering in a wilderness environment – they failed pathetically.

Nor is it surprising that they failed. Hunting and gathering is actually really hard. Here's Henrich's description of how the Inuit hunt seals:

> You first have to find their breathing holes in the ice. It's important that the area around the hole be snow-covered—otherwise the seals will hear you and vanish. You then open the hole, smell it to verify it's still in use (what do seals smell like?), and then assess the shape of the hole using a special curved piece of caribou antler. The hole is then covered with snow, save for a small gap at the top that is capped with a down indicator. If the seal enters the hole, the indicator moves, and you must blindly plunge your harpoon into the hole using all your weight. Your harpoon should be about 1.5 meters (5ft) long, with a detachable tip that is tethered with a heavy braid of sinew line. You can get the antler from the previously noted caribou, which you brought down with your driftwood bow.
>
> The rear spike of the harpoon is made of extra-hard polar bear bone (yes, you also need to know how to kill polar bears; best to catch them napping in their dens). Once you've plunged your harpoon's head into the seal, you're then in a wrestling match as you reel him in, onto the ice, where you can finish him off with the aforementioned bear-bone spike.
>
> Now you have a seal, but you have to cook it. However, there are no trees at this latitude for wood, and

> driftwood is too sparse and valuable to use routinely for fires. To have a reliable fire, you'll need to carve a lamp from soapstone (you know what soapstone looks like, right?), render some oil for the lamp from blubber, and make a wick out of a particular species of moss. You will also need water. The pack ice is frozen salt water, so using it for drinking will just make you dehydrate faster. However, old sea ice has lost most of its salt, so it can be melted to make potable water. Of course, you need to be able to locate and identify old sea ice by color and texture. To melt it, make sure you have enough oil for your soapstone lamp.

No surprise that stranded explorers couldn't figure all this out. It's more surprising that the Inuit *did*. And although the Arctic is an unusually hostile place for humans, Henrich makes it clear that hunting-gathering techniques of this level of complexity are standard everywhere. Here's how the Indians of Tierra del Fuego make arrows:

> Among the Fuegians, making an arrow requires a 14-step procedure that involves using seven different tools to work six different materials. Here are some of the steps:
>
> – The process begins by selecting the wood for the shaft, which preferably comes from chaura, a bushy, evergreen shrub. Though strong and light, this wood is a non-intuitive choice since the gnarled branches require extensive straightening (why not start with straighter branches?).
>
> – The wood is heated, straightened with the craftsman's teeth, and eventually finished with a scraper. Then, using a pre-heated and grooved stone, the shaft is pressed into the grooves and rubbed back and forth,

> pressing it down with a piece of fox skin. The fox skin becomes impregnated with the dust, which prepares it for the polishing stage (Does it have to be fox skin?).
>
> – Bits of pitch, gathered from the beach, are chewed and mixed with ash (What if you don't include the ash?).
>
> – The mixture is then applied to both ends of a heated shaft, which must then be coated with white clay (what about red clay? Do you have to heat it?). This prepares the ends for the fletching and arrowhead.
>
> – Two feathers are used for the fletching, preferably from upland geese (why not chicken feathers?).
>
> – Right-handed bowman must use feathers from the left wing of the bird, and vice versa for lefties (Does this really matter?).
>
> – The feathers are lashed to the shaft using sinews from the back of the guanaco, after they are smoothed and thinned with water and saliva (why not sinews from the fox that I had to kill for the aforementioned skin?).
>
> Next is the arrowhead, which must be crafted and then attached to the shaft, and of course there is also the bow, quiver and archery skills. But, I'll leave it there, since I think you get the idea.

How do hunter-gatherers know how to do all this? We usually summarize it as "culture". How did it form? Not through some smart Inuit or Fuegian person reasoning it out; if that had been it, smart European explorers should have been able to reason it out too.

The obvious answer is "cultural evolution", but Henrich isn't much better than anyone else at taking the mystery out of this phrase. Trial and error must have been involved, and less successful groups/people imitating the techniques of more successful ones. But is that really a satisfying explanation?

I found the chapter on language a helpful reminder that we already basically accept something like this is true. How did language get invented? I’m especially interested in this question because of my brief interactions with conlanging communities – people who try to construct their own languages as a hobby or as part of a fantasy universe, like Tolkien did with Elvish. Most people are *terrible* at this; their languages are either unusable, or exact clones of English. Only people who (like Tolkien) already have years of formal training in linguistics can do a remotely passable job. And you’re telling me the original languages were invented by cavemen? Surely there was no committee of Proto-Indo-European nomads that voted on whether to have an inflecting or agglutinating tongue? Surely nobody ran out of their cave shouting “Eureka!” after having discovered the interjection? We just kind of accept that after cavemen working really hard to communicate with each other, eventually language – still one of the most complicated and impressive productions of the human race – just sort of happened.

(this is how I feel about biological evolution too – how do you evolve an eye by trial and error? I’ve read papers speculating on the exact process, and they make lots of good points, but I still don’t feel *happy* about it, like “Oh, of course this would happen!” At some point you just have to accept evolution is smarter than you are and smarter than you would expect to be possible.)

Taking the generation of culture as secondary to this kind of mysterious process, Henrich turns to its transmission. If cultural generation happens at a certain rate, then the fidelity of transmission determines whether a given society advances, stagnates, or declines.

For Henrich, humans started becoming more than just another species of monkey when we started transmitting culture with high fidelity. Some anthropologists talk about the Machiavellian Intelligence Hypothesis – the theory that humans evolved big brains in order to succeed at social maneuvering and climbing dominance hierarchies. Henrich counters with his own Cultural Intelligence Hypothesis – humans evolved big brains in order to be able to maintain things like Inuit seal hunting techniques. Everything that separates us from the apes is part of an evolutionary package designed to help us maintain this kind of culture, exploit this kind of culture, or adjust to the new abilities that this kind of culture gave us.

II.

Secret gives many examples of many culture-related adaptations, and not all are in the brain.

Our digestive tracts evolved alongside our cultures. Specifically, they evolved to be unusually puny:

> Our mouths are the size of the squirrel monkey’s, a species that weighs less than three pounds. Chimpanzees can open their mouths twice as wide as we can and hold substantial amounts of food compressed between their lips and large teeth. We also have puny jaw muscles that reach up only to just below our ears. Other primates’ jaw muscles stretch to the top of their heads, where they sometimes even latch onto a central bony ridge. Our stomachs are small, having only a third of the surface area that we’d expect for a

> primate of our size, and our colons are too short, being only 60% of their expected mass.

Compared to other animals, we have such atrophied digestive tracts that we shouldn't be able to live. What saves us? All of our food processing techniques, especially cooking, but also chopping, rinsing, boiling, and soaking. We've done much of the work of digestion before food even enters our mouths. Our culture teaches us how to do this, both in broad terms like "hold things over fire to cook them" and in specific terms like "this plant needs to be soaked in water for 24 hours to leach out the toxins". Each culture has its own cooking knowledge related to the local plants and animals; a frequent cause of death among European explorers was cooking things in ways that didn't unlock any of the nutrients, and so starving while apparently well-fed.

Fire is an especially important food processing innovation, and it is entirely culturally transmitted. Henrich is kind of cruel in his insistence on this. He recommends readers go outside and try to start a fire. He even gives some helpful hints – flint is involved, rubbing two sticks together works for some people, etc. He predicts – and stories I've heard from unfortunate campers confirm – that you will not be able to do this, despite an IQ far beyond that of most of our hominid ancestors. In fact, some groups (most notably the aboriginal Tasmanians) seem to have lost the ability to make fire, and never rediscovered it. Fire-making was discovered a small number of times, maybe once, and has been culturally transmitted since then.

But it's not just about chopping things up or roasting them. Traditional food processing techniques can get arbitrarily complicated. Nixtamalization of corn, necessary to prevent vitamin deficiencies, involves soaking the corn in a solution containing ground-up burnt seashells. The ancient Mexicans discovered this and lived off corn just fine for millennia. When the conquistadors took over, they ignored it and ate corn straight. For four hundred years, Europeans and Americans ate unnixtamalized corn. By official statistics, three million Americans came down with corn-related vitamin deficiencies during this time, and up to a hundred thousand died. It wasn't until 1937 that Western scientists discovered which vitamins were involved and developed an industrial version of nixtamalization that made corn safe. Early 1900s Americans were very smart and had lots of advantages over ancient Mexicans. But the ancient Mexicans' culture got this one right in a way it took Westerners centuries to match.

Our hands and limbs also evolved alongside our cultures. We improved dramatically in some areas: after eons of tool use, our hands outclass those of any other ape in terms of finesse. In other cases, we devolved systems that were no longer necessary; we are much weaker than any other ape. Henrich describes a circus act of the 1940s where the ringmaster would challenge strong men in the audience to wrestle a juvenile chimpanzee. The chimpanzee was tied up, dressed in a mask that prevented it from biting, and wearing soft gloves that prevented it from scratching. No human ever lasted more than five seconds. Our common ancestor with other apes grew weaker and weaker as we became more and more reliant on artificial weapons to give us an advantage.

Even our sweat glands evolved alongside culture. Humans are persistence hunters: they cannot run as fast as gazelles, but they can keep running for longer than gazelles (or almost anything else). Why did we evolve into that niche? The secret is our ability to carry water. Every hunter-gatherer culture has invented its own water-carrying techniques, usually some kind of waterskin. This allowed humans to switch to perspiration-based cooling systems, which allowed them to run as long as they want.

III.

But most of our differences from other apes are indeed in the brain. They’re just not where you’d expect.

Tomasello et al tested human toddlers vs. apes on a series of traditional IQ type questions. The match-up was surprisingly fair; in areas like memory, logic, and spatial reasoning, the three species did about the same. But in ability to learn from another person, humans wiped the floor with the other two ape species:

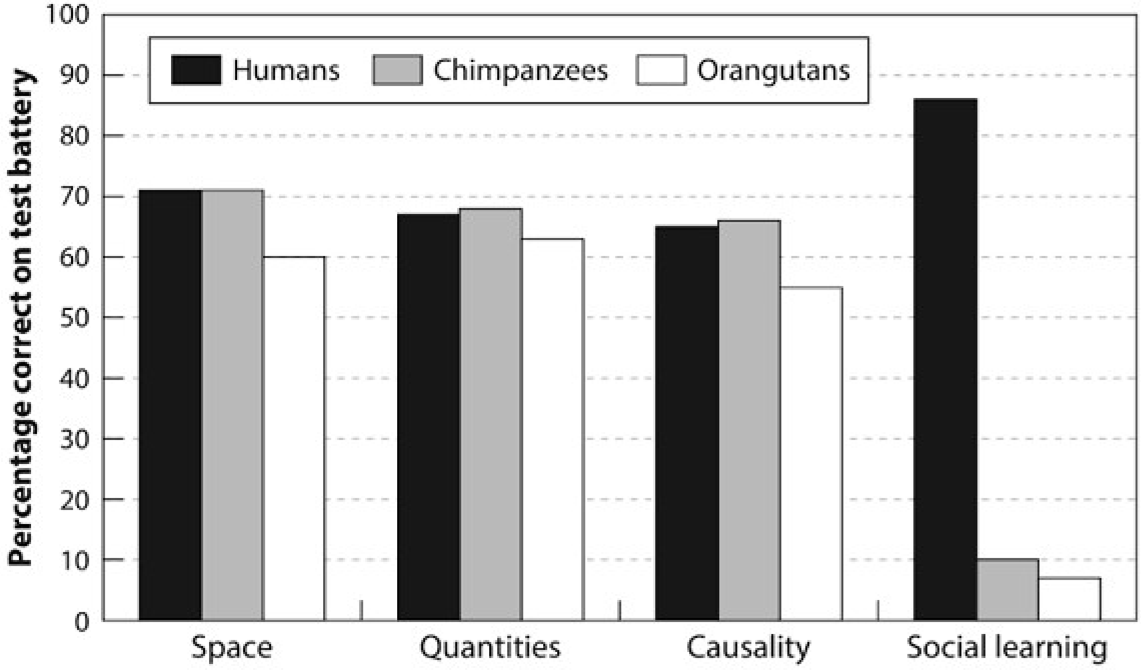


Figure 2.2. Average performance on four sets of cognitive tests with chimpanzees, orangutans, and toddlers.

Remember, Henrich thinks culture accumulates through random mutation. Humans don’t have control over how culture gets generated. They have more control over how much of it gets transmitted to the next generation. If 100% gets transmitted, then as more and more mutations accumulate, the culture becomes better and better. If less than 100% gets transmitted, then at some point new culture gained and old culture lost fall into equilibrium, and your society stabilizes at some higher or lower technological level. This means that transmitting culture to the next generation is maybe the core human skill. The human brain is optimized to make this work as well as possible.

Human children are obsessed with learning things. And they don’t learn things randomly. There seem to be “biases in cultural learning”, ie slots in an infant’s mind that they know need to be filled with knowledge, and which they preferentially seek out the knowledge necessary to fill.

One slot is for language. Human children naturally listen to speech (as early as in the womb). They naturally prune the phonemes they are able to produce and distinguish to the ones in the local language. And they naturally figure out how to speak and understand what people are saying, even though learning a language is hard even for smart adults.

Another slot is for animals. In a world where megafauna has been relegated to zoos, we *still* teach children their ABCs with “L is for lion” and “B is for bear”, and children *still* read picture books about Mr. Frog and Mrs. Snake holding tea parties. Henrich suggests that just as the young brain is hard-coded to want to learn language, so it is hard-coded to want to learn the local animal life (maybe little boys’ vehicle obsession is an outgrowth of this – buses and trains are the closest thing to local megafauna that most of them will encounter!)

Another slot is for plants:

To see this system in operation, let’s consider how infants respond to unfamiliar plants. Plants are loaded with prickly thorns, noxious oils, stinging nettles and dangerous toxins, all genetically evolved to prevent animals like us from messing with them. Given our species wide geographic range and diverse use of plants as foods, medicines and construction materials, we ought to be primed to both learn about plants and avoid their dangers. To explore this idea in the lab, the psychologists Annie Wertz and Karen Wynn first gave infants, who ranged in age from eight to eighteen months, an opportunity to touch novel plants (basil and parsley) and artifacts, including both novel objects and common ones, like wooden spoons and small lamps.

The results were striking. Regardless of age, many infants flatly refused to touch the plants at all. When they did touch them, they waited substantially longer than they did with the artifacts. By contrast, even with the novel objects, infants showed none of this reluctance. This suggests that well before one year of age infants can readily distinguish plants from other things, and are primed for caution with plants. But, how do they get past this conservative predisposition?

The answer is that infants keenly watch what other people do with plants, and are only inclined to touch or eat the plants that other people have touched or eaten. In fact, once they get the ‘go ahead’ via cultural learning, they are suddenly interested in eating plants. To explore this, Annie and Karen exposed infants to models who both picked fruit from plants and also picked fruit-like things from an artifact of similar size and shape to the plant. The models put both the fruit and the fruit-like things in their mouths. Next, the infants were given a choice to go for the fruit (picked from the plant) or the fruit-like things picked from the object. Over 75% of the time the infants went for the fruit, not the fruit-like things, since they'd gotten the 'go ahead' via cultural learning.

As a check, the infants were also exposed to models putting the fruit or fruit-like things behind their ears(not in their mouths). In this case, the infants went for the fruit or fruit-like things in equal measure. It seems that plants are most interesting if you can eat them, but only if you have some cultural learning cues that they aren't toxic.

After Annie first told me about her work while I was visiting Yale in 2013, I went home to test it on my 6-month-old son, Josh. Josh seemed very likely to overturn Annie's hard empirical work, since he immediately grasped anything you gave him and put it rapidly in his mouth. Comfortable in his mom's arms, I first offered Josh a novel plastic cube. He delighted in grabbing it and shoving it directly into his mouth, without any hesitation. Then, I offered him a sprig of arugula. He quickly grabbed it, but then paused, looked with curious uncertainty at it, and then slowly let it fall from his hand while turning to hug his mom.

It's worth pointing out how rich the psychology is here. Not only do infants have to recognize that plants are different from objects of similar size, shape and color, but they need to create categories for types of plants, like basil and parsley, and distinguish 'eating' from just 'touching'. It does them little good to code their observation of someone eating basil as 'plants are good to eat' since that might cause them to eat poisonous plants as well as basil. But, it also does them little good to narrowly code the observation as 'that particular

> sprig of basil is good to eat’ since that particular sprig has just been eaten by the person they are watching. This another content bias in cultural learning.

This ties into the more general phenomenon of figuring out what’s edible. Most Westerners learn insects aren’t edible; some Asians learn that they are. This feels deeper than just someone telling you insects aren’t edible and you believing them. When I was in Thailand, my guide offered me a giant cricket, telling me it was delicious. I believed him when he said it was safe to eat, I even believed him when he said it tasted good to him, but my conditioning won out – I didn’t eat the cricket. There seems to be some process where a child’s brain learns what is and isn’t locally edible, then hard-codes it against future change.

(Or so they say; I’ve never been able to eat shrimp either.)

Another slot is for gender roles. By now we’ve all heard the stories of progressives who try to raise their children without any exposure to gender. Their failure has sometimes been taken as evidence that gender is hard-coded. But it can’t be quite that simple: some modern gender roles, like girls = pink, are far from obvious or universal. Instead, it looks like children have a hard-coded slot that gender roles go into, work hard to figure out what the local gender roles are (even if their parents are trying to confuse them), then latch onto them and don’t let go.

In the Cultural Intelligence Hypothesis, humans live in obligate symbiosis with a culture. A brain without an associated culture is incomplete and not very useful. So the infant brain is adapted to seek out the important aspects of its local culture almost from birth and fill them into the appropriate slots in order to become whole.

IV.

The next part of the book discusses post-childhood learning. This plays an important role in hunter-gatherer tribes:

> While hunters reach their peak strength and speed in their twenties, individual hunting success does not peak until around age 30, because success depends more on know-how and refined skills than on physical prowess.

This part of the book made most sense in the context of examples like the Inuit seal-hunting strategy which drove home just how complicated and difficult hunting-gathering was. Think less “Boy Scouts” and more “PhD”; a primitive tribesperson’s life requires mastery of various complicated technologies and skills. And the difference between “mediocre hunter” and “great hunter” can be the difference between high status (and good mating opportunities) and low status, or even between life and death. Hunter-gatherers really want to learn the essentials of their hunter-gatherer lifestyle, and learning it is really hard. Their heuristics are:

Learn from people who are good at things and/or widely-respected. If you haven’t already read about the difference between dominance and prestige hierarchies, check out Kevin Simler’s blog post on the topic. People will fear and obey authority figures like kings and chieftains, but they give a different kind of respect (“prestige”) to people who seem good at things. And since it’s hard to figure out who’s good at things (can a non-musician who wants to start learning music tell the difference between a merely good performer and one of the world’s best?) most people use the heuristic of respecting the people who other people respect. Once you identify someone as respect-worthy, you strongly consider copying them in, well, everything:

To understand prestige as a social phenomenon, it’s crucial to realize that it’s often difficult to figure out what precisely makes someone successful. In modern societies, the success of a star NBA basketball player might arise from his:

(1) intensive practice in the offseason
(2) sneaker preference
(3) sleep schedule
(4) pre-game prayer
(5) special vitamins
(6) taste for carrots

Any or all of these might increase his success. A naïve learner can’t tell all the causal links between an individual’s practices and his success. As a consequence, learners often copy their chosen models broadly across many domains. Of course, learners may place more weight on domains that for one reason or other seem more causally relevant to the model’s success. This copying often includes the model’s personal habits or styles as well as their goals and motivations, since these may be linked to their success. This “if in doubt, copy it” heuristic is one of the reasons why success in one domain converts to influence across a broad range of domains.

The immense range of celebrity endorsements in modern societies shows the power of prestige. For example, NBA star Lebron James, who went directly from High School to the pros, gets paid millions to endorse State Farm Insurance. Though a stunning basketball talent, it’s unclear why Mr. James is qualified to recommend insurance companies. Similarly, Michael Jordan famously wore Hanes underwear and apparently Tiger Woods drove Buicks. Beyonce’ drinks Pepsi (at

> least in commercials). What’s the connection between musical talent and sugary cola beverages?
>
> Finally, while new medical findings and public educational campaigns only gradually influence women’s approach to preventive medicine, Angelina Jolie’s single OP-ED in the New York Times, describing her decision to get a preventive double mastectomy after learning she had the ‘faulty’ BRCA1 gene, flooded clinics from the U.K. to New Zealand with women seeking genetic screenings for breast cancer. Thus, an unwanted evolutionary side effect, prestige turns out to be worth millions, and represents a powerful and underutilized public health tool.

Of course, this creates the risk of prestige cascades, where some irrelevant factor (Henrich mentions being a reality show star) catapults someone to fame, everyone talks about them, and you end up with Muggeridge’s definition of a celebrity: someone famous for being famous.

Some of this makes more sense if you go back to the evolutionary roots, and imagine watching the best hunter in your tribe to see what his secret is, or being nice to him in the hopes that he’ll take you under his wing and teach you stuff.

(but if all this is true, shouldn’t public awareness campaigns that hire celebrity spokespeople be wild successes? Don’t they just as often fail, regardless of how famous a basketball player they can convince to lecture schoolchildren about how Winners Don’t Do Drugs?)

Learn from people who are like you. If you are a man, it is probably a bad idea to learn fashion by observing women. If you are a servant, it is probably a bad idea to learn the rules of etiquette by observing how the king behaves. People are naturally inclined to learn from people more similar to themselves.

Henrich ties this in to various studies showing that black students learn best from a black teacher, female students from a female teacher, et cetera.

Learn from old people. Humans are almost unique in having menopause; most animals keep reproducing until they die in late middle-age. Why does evolution want humans to stick around without reproducing?

Because old people have already learned the local culture and can teach it to others. Henrich asks us to throw out any personal experience we have of elders; we live in a rapidly-changing world where an old person is probably “behind the times”. But for most of history, change happened glacially slowly, and old people would have spent their entire lives accumulating relevant knowledge. Imagine a Silicon Valley programmer stumped by a particularly tough bug in his code calling up his grandfather, who has seventy years’ experience in the relevant programming language.

Sometimes important events only happen once in a generation. Henrich tells the story of an Australian aboriginal tribe facing a massive drought. Nobody knew what to do except Paralji, the tribe’s oldest man, who had lived through the last massive drought and remembered where his own elders had told him to find the last-resort waterholes.

This same dynamic seems to play out even in other species:

> In 1993, a severe drought hit Tanzania, resulting in the death of 20% of the African elephant calves in a population of about 200. This population contained 21

different families, each of which was led by a single matriarch. The 21 elephant families were divided into 3 clans, and each clan shared the same territory during the wet season (so, they knew each other). Researchers studying these elephants have analyzed the survival of the calves and found that families led by older matriarchs suffered fewer deaths of their calves during this drought.

Moreover, two of the three elephant clans unexpectedly left the park during the drought, presumably in search of water, and both had much higher survival rates than the one clan that stayed behind. It happens that these severe droughts only hit about once every four to five decades, and the last one hit about 1960. After that, sadly, elephant poaching in the 1970’s killed off many of the elephants who would have been old enough in 1993 to recall the 1960 drought. However, it turns out that exactly one member of each of the two clans who left the park, and survived more effectively, were old enough to recall life in 1960. This suggests, that like Paralji in the Australian desert, they may have remembered what to do during a severe drought, and led their groups to the last water refuges. In the clan who stayed behind, the oldest member was born in 1960, and so was too young to have recalled the last major drought.

More generally, aging elephant matriarchs have a big impact on their families, as those led by older matriarchs do better at identifying and avoiding predators (lions and humans), avoiding internal conflicts and identifying the calls of their fellow elephants. For example, in one set of field experiments, researchers played lion roars from both male and female lions, and from either a single lion or a trio of lions. For elephants,

> male lions are much more dangerous than females, and of course, three lions are always worse than only one lion. All the elephants generally responded with more defensive preparations when they heard three lions vs. one. However, only the older matriarchs keenly recognized the increased dangers of male lions over female lions, and responded to the increased threat with elephant defensive maneuvers.

V.

I was inspired to read *Secret* by Scholar’s Stage’s review. I hate to be unoriginal, but after reading the whole book, I agree that the three sections Tanner cites – on divination, on manioc, and on shark taboos – are by far the best and most fascinating.

On divination:

> When hunting caribou, Naskapi foragers in Labrador, Canada, had to decide where to go. Common sense might lead one to go where one had success before or to where friends or neighbors recently spotted caribou.
>
> However, this situation is like [the Matching Pennies game]. The caribou are mismatchers and the hunters are matchers. That is, hunters want to match the locations of caribou while caribou want to mismatch the hunters, to avoid being shot and eaten. If a hunter shows any bias to return to previous spots, where he or others have seen caribou, then the caribou can benefit (survive better) by avoiding those locations (where they have previously seen humans). Thus, the best hunting strategy requires randomizing.
>
> Can cultural evolution compensate for our cognitive inadequacies? Traditionally, Naskapi hunters decided

where to go to hunt using divination and believed that the shoulder bones of caribou could point the way to success. To start the ritual, the shoulder blade was heated over hot coals in a way that caused patterns of cracks and burnt spots to form. This patterning was then read as a kind of map, which was held in a pre-specified orientation. The cracking patterns were (probably) essentially random from the point of view of hunting locations, since the outcomes depended on myriad details about the bone, fire, ambient temperature, and heating process. Thus, these divination rituals may have provided a crude randomizing device that helped hunters avoid their own decision-making biases.

This is not some obscure, isolated practice, and other cases of divination provide more evidence. In Indonesia, the Kantus of Kalimantan use bird augury to select locations for their agricultural plots. Geographer Michael Dove argues that two factors will cause farmers to make plot placements that are too risky. First, Kantu ecological models contain the Gambler’s Fallacy, and lead them to expect floods to be less likely to occur in a specific location after a big flood in that location (which is not true). Second…Kantus pay attention to others’ success and copy the choices of successful households, meaning that if one of their neighbors has a good yield in an area one year, many other people will want to plant there in the next year. To reduce the risks posed by these cognitive and decision-making biases, Kantu rely on a system of bird augury that effectively randomizes their choices for locating garden plots, which helps them avoid catastrophic crop failures. Divination results depend not only on seeing a particular bird species in a particular location, but also on what type of call the bird makes (one type of call may be favorable, and another unfavorable).

> The patterning of bird augury supports the view that this is a cultural adaptation. The system seems to have evolved and spread throughout this region since the 17th century when rice cultivation was introduced. This makes sense, since it is rice cultivation that is most positively influenced by randomizing garden locations. It's possible that, with the introduction of rice, a few farmers began to use bird sightings as an indication of favorable garden sites. On-average, over a lifetime, these farmers would do better – be more successful – than farmers who relied on the Gambler's Fallacy or on copying others' immediate behavior. Whatever the process, within 400 years, the bird augury system spread throughout the agricultural populations of this Borneo region. Yet, it remains conspicuously missing or underdeveloped among local foraging groups and recent adopters of rice agriculture, as well as among populations in northern Borneo who rely on irrigation. So, bird augury has been systematically spreading in those regions where it's most adaptive.

Scott Aaronson has written about how easy it is to predict people trying to "be random":

> In a class I taught at Berkeley, I did an experiment where I wrote a simple little program that would let people type either "f" or "d" and would predict which key they were going to push next. It's actually very easy to write a program that will make the right prediction about 70% of the time. Most people don't really know how to type randomly. They'll have too many alternations and so on. There will be all sorts of patterns, so you just have to build some sort of probabilistic model. Even a very crude one will do well. I couldn't even beat my own program, knowing exactly how it worked. I challenged people to try this and the

> program was getting between 70% and 80% prediction rates. Then, we found one student that the program predicted exactly 50% of the time. We asked him what his secret was and he responded that he "just used his free will."

But being genuinely random is important in pursuing mixed game theoretic strategies. Henrich's view is that divination solved this problem effectively.

I'm reminded of the Romans using augury to decide when and where to attack. This always struck me as crazy; generals are going to risk the lives of thousands of soldiers because they saw a weird bird earlier that morning? But war is a classic example of when a random strategy can be useful. If you're deciding whether to attack the enemy's right vs. left flank, it's important that the enemy can't predict your decision and send his best defenders there. If you're generally predictable – and Scott Aaronson says you are – then outsourcing your decision to weird birds might be the best way to go.

And then there's manioc. This is a tuber native to the Americas. It contains cyanide, and if you eat too much of it, you get cyanide poisoning. From Henrich:

> In the Americas, where manioc was first domesticated, societies who have relied on bitter varieties for thousands of years show no evidence of chronic cyanide poisoning. In the Colombian Amazon, for example, indigenous Tukanoans use a multistep, multiday processing technique that involves scraping, grating, and finally washing the roots in order to separate the fiber, starch, and liquid. Once separated, the liquid is boiled into a beverage, but the fiber and starch must then sit for two more days, when they can then be

baked and eaten. Figure 7.1 shows the percentage of cyanogenic content in the liquid, fiber, and starch remaining through each major step in this processing.

Such processing techniques are crucial for living in many parts of Amazonia, where other crops are difficult to cultivate and often unproductive. However, despite their utility, one person would have a difficult time figuring out the detoxification technique. Consider the situation from the point of view of the children and adolescents who are learning the techniques. They would have rarely, if ever, seen anyone get cyanide poisoning, because the techniques work. And even if the processing was ineffective, such that cases of goiter (swollen necks) or neurological problems were common, it would still be hard to recognize the link between these chronic health issues and eating manioc. Most people would have eaten manioc for years with no apparent effects. Low cyanogenic varieties are typically boiled, but boiling alone is insufficient to prevent the chronic conditions for bitter varieties. Boiling does, however, remove or reduce the bitter taste and prevent the acute symptoms (e.g., diarrhea, stomach troubles, and vomiting).

So, if one did the common-sense thing and just boiled the high-cyanogenic manioc, everything would seem fine. Since the multistep task of processing manioc is long, arduous, and boring, sticking with it is certainly non-intuitive. Tukanoan women spend about a quarter of their day detoxifying manioc, so this is a costly technique in the short term. Now consider what might result if a self-reliant Tukanoan mother decided to drop any seemingly unnecessary steps from the processing of her bitter manioc. She might critically examine the procedure handed down to her from earlier generations and conclude that the goal of the procedure is to remove the bitter taste. She might then experiment with alternative procedures by dropping some of the more labor-intensive or time-consuming steps. She'd find that with a shorter and much less labor-intensive process, she could remove the bitter taste. Adopting this easier protocol, she would have more time for other activities, like caring for her children. Of course, years or decades later her family would begin to develop the symptoms of chronic cyanide poisoning.

Thus, the unwillingness of this mother to take on faith the practices handed down to her from earlier generations would result in sickness and early death for members of her family. Individual learning does not pay here, and intuitions are misleading. The problem is that the steps in this procedure are causally opaque—an individual cannot readily infer their functions, interrelationships, or importance. The causal opacity of many cultural adaptations had a big impact on our psychology.

Wait. Maybe I'm wrong about manioc processing. Perhaps it's actually rather easy to individually figure out the detoxification steps for manioc? Fortunately, history has provided a test case. At the beginning of the seventeenth century, the Portuguese transported manioc from South America to West Africa for the first time. They did not, however, transport the age-old indigenous processing protocols or the underlying commitment to using those techniques. Because it is easy to plant and provides high yields in infertile or drought-prone areas, manioc spread rapidly across Africa and became a staple food for many populations. The processing techniques, however, were not readily or consistently regenerated. Even after hundreds of years,

> chronic cyanide poisoning remains a serious health problem in Africa. Detailed studies of local preparation techniques show that high levels of cyanide often remain and that many individuals carry low levels of cyanide in their blood or urine, which haven't yet manifested in symptoms. In some places, there's no processing at all, or sometimes the processing actually increases the cyanogenic content. On the positive side, some African groups have in fact culturally evolved effective processing techniques, but these techniques are spreading only slowly.

Rationalists always wonder: how come people aren't more rational? How come you can prove a thousand times, using Facts and Logic, that something is stupid, and yet people will still keep doing it?

Henrich hints at an answer: for basically all of history, using reason would get you killed.

A reasonable person would have figured out there was no way for oracle-bones to accurately predict the future. They would have abandoned divination, failed at hunting, and maybe died of starvation.

A reasonable person would have asked why everyone was wasting so much time preparing manioc. When told "Because that's how we've always done it", they would have been unsatisfied with that answer. They would have done some experiments, and found that a simpler process of boiling it worked just as well. They would have saved lots of time, maybe converted all their friends to the new and easier method. Twenty years later, they would have gotten sick and died, in a way so causally distant from their decision to change manioc processing methods that nobody would ever have been able to link the two together.

Henrich discusses pregnancy taboos in Fiji; pregnant women are banned from eating sharks. Sure enough, these sharks contain chemicals that can cause birth defects. The women didn’t really know why they weren’t eating the sharks, but when anthropologists demanded a reason, they eventually decided it was because their babies would be born with shark skin rather than human skin. As explanations go, this leaves a lot to be desired. How come you can still eat other fish? Aren’t you worried your kids will have scales? Doesn’t the slightest familiarity with biology prove this mechanism is garbage? But if some smart independent-minded iconoclastic Fijian girl figured any of this out, she would break the taboo and her child would have birth defects.

In giving humans reason at all, evolution took a huge risk. Surely it must have wished there was some other way, some path that made us big-brained enough to understand tradition, but not big-brained enough to question it. Maybe it searched for a mind design like that and couldn’t find one. So it was left with this ticking time-bomb, this ape that was constantly going to be able to convince itself of hare-brained and probably-fatal ideas.

Here, too, culture came to the rescue. One of the most important parts of any culture – more important than the techniques for hunting seals, more important than the techniques for processing tubers – is techniques for making sure nobody ever questions tradition. Like the belief that anyone who doesn’t conform is probably a witch who should be cast out lest they bring destruction upon everybody. Or the belief in a God who has commanded certain specific weird dietary restrictions, and will torture you forever if you disagree. Or the fairy tales where the prince asks a wizard for help, and the wizard says “You may have everything you wish forever, but you must never nod your head at a badger”, and then one day the prince nods his head at a badger, and his whole empire collapses into dust, and the moral of the story is that you should always obey weird advice you don't understand.

There's a monster at the end of this book. Humans evolved to transmit culture with high fidelity. And one of the biggest threats to transmitting culture with high fidelity was Reason. Our ancestors lived in Epistemic Hell, where they had to constantly rely on causally opaque processes with justifications that couldn't possibly be true, and if they ever questioned them then they might die. Historically, Reason has been the villain of the human narrative, a corrosive force that tempts people away from adaptive behavior towards choices that "sounded good at the time".

Why are people so bad at reasoning? For the same reason they're so bad at letting poisonous spiders walk all over their face without freaking out. Both "skills" are really bad ideas, most of the people who tried them died in the process, so evolution removed those genes from the population, and successful cultures stigmatized them enough to give people an internalized fear of even trying.

VI.

This book belongs alongside *Seeing Like A State* and the works of G.K. Chesterton as attempts to justify tradition, and to argue for organically-evolved institutions over top-down planning. What unique contribution does it make to this canon?

First, a lot more specifically anthropological / paleoanthropological rigor than the other two.

Second, a much crisper focus: Chesterton had only the fuzziest idea that he was writing about cultural evolution, and Scott was only a little clearer. I think Henrich is the only one of the three to use the term, and once you hear it, it’s obviously the right framing.

Third, a sense of how traditions contain the meta-tradition of defending themselves against Reason, and a sense for why this is necessary.

And fourth, maybe we’re not at the point where we really want unique contributions yet. Maybe we’re still at the point where we have to have this hammered in by more and more examples. The temptation is always to say “Ah, yes, a few simple things like taboos against eating poisonous plants may be relics of cultural evolution, but obviously by now we’re at the point where we know which traditions are important vs. random looniness, and we can rationally stick to the important ones while throwing out the garbage.” And then somebody points out to you that *actually* divination using oracle bones was one of the important traditions, and if you thought you knew better than that and tried to throw it out, your civilization would falter.

Maybe we just need to keep reading more similarly-themed books until this point really sinks in, and we get properly worried.

THE SECRET OF OUR SUCCESS
JOSEPH HENRICH

List Of Passages I Highlighted In My Copy Of “The Secret Of Our Success”

[Previously in sequence: Epistemic Learned Helplessness, Book Review: The Secret Of Our Success]

A rare example of cultural evolution in action:

> Throughout the Highlands of New Guinea, a group’s ability to raise large numbers of pigs is directly related to its economic and social success in competition with other regional groups. The ceremonial exchange of pigs allows groups to forge alliances, re-pay debts, obtain wives, and generate prestige through excessive displays of generosity. All this means that groups who are better able to raise pigs can expand more rapidly in numbers—by reproduction and in-migration—and thus have the potential to expand their territory. Group size is very important in intergroup warfare in small-scale societies so larger groups are more likely to successfully expand their territory. However, the prestige more successful groups obtain may cause the rapid diffusion of the very institutions, beliefs, or practices responsible for their competitive edge as other groups adopt their strategies and beliefs.
>
> In 1971, the anthropologist David Boyd was living in the New Guinea village of Irakia, and observed intergroup competition via prestige-biased group transmission. Concerned about their low prestige and weak pig production, the senior men of Irakia convened a series of meetings to determine how to improve their situation. Numerous suggestions were proposed for raising their pig production but after a long process of consensus building the senior men of the village decided to follow

a suggestion made by a prestigious clan-leader who proposed that they “must follow the Fore’” and adopt their pig-related husbandry practices, rituals, and other institutions. The Fore’ were a large and successful ethnic group in the region, who were renowned for their pig production. The following practices, beliefs, rules, and goals were copied from the Fore’, and announced at the next general meeting of the community:

1) All villagers must sing, dance and play flutes for their pigs. This ritual causes the pigs to grow faster and bigger. At feasts, the pigs should be fed first from the oven. People are fed second.

2) Pigs should not be killed for breaking into another’s garden. The pig’s owner must assist the owner of the garden in repairing the fence. Disputes will be resolved following the dispute resolution procedure used among the Fore’.

3) Sending pigs to other villages is tabooed, except for the official festival feast.

4) Women should take better care of the pigs, and feed them more food. To find extra time for this, women should spend less time gossiping.

5) Men must plant more sweet potatoes for the women to feed to the pigs, and should not depart for wage labor in distant towns until the pigs have grown to a certain size.

The first two items were implemented immediately at a ritual feast. David stayed in the village long enough to verify that the villagers did adopt the other practices, and that their pig production did increase in the short

> term, though unfortunately we don’t know what happened in the long-run.
>
> Let me highlight three features of this case. First, the real causal linkages between many of these elements and pig production are unclear. Maybe singing does cause pigs to grow faster, but it’s not obvious and no one tried to ascertain this fact, via experimentation for example. Second, the village leadership chose to rely on copying institutions from other groups, and not on designing their own institutions from scratch. This is smart, since we humans are horrible at designing institutions from scratch. And third, this transmission between groups occurred rapidly because Irakia already had a political institution in the village, involving a council of the senior members of each clan, who were empowered by tradition (social norms) to make community-level decisions. Lacking this decision-making institution, Fore’ practices would have had to spread among households, and thus been much slower in spreading. Of course, such political decision-making institutions themselves are favored by intergroup competition.

This is it. This is the five-point platform that the Democratic Party can use to win in 2020.

Yesterday’s review mentioned that children have certain “slots” in their heads that are ready for specific types of incoming information. How far can we take this idea?

> The UCLA anthropologist Dan Fessler argues that during middle childhood (ages 6-9) humans go through a phase in which we are strongly attracted to leraning about fire, by both observing others and manipulating it ourselves. In small-scale societies, where children are free to

> engage this curiosity, adolescents have both mastered fire and lost any further attraction to it. Interestingly, Fessler also argues that modern societies are unusual because so many children never get to satisfy their curiosity, so their fascination with fire stretches into the teen years and early adulthood.

On prestige-based socialization and learning who to learn from:

> By 14 months, infants are already well beyond social referencing and already showing signs of using skill or competence cues to select models. After observing an adult model acting confused by shoes and placing them on his hands, German infants tended not to copy his unusual way of turning on a novel lighting device: using his head. However, if the model acted competently, confidently putting shoes on his feet, babies tended to copy the model and used their heads to activate the novel lighting device.

Kind of unrelated to culture, but did you know quadruped animals run at quantized speeds?

> Many four-legged animals are saddled with a design disadvantage. Game animals thermoregulate by panting, like a dog. If they need to release more heat, they pant faster. This works fine unless they are running. When they run, the impact of their forelimbs compresses their chest cavities in a manner that makes breathing during compressions inefficient. This means that, ignoring oxygen and thermoregulation requirements, running quadrupeds should breathe only once per locomotor-cycle. But, since the need for oxygen goes up linearly with speed, they will be

breathing too frequently at some speeds and not frequently enough at other speeds. Consequently, a running quadruped must pick a speed that (1) demands only one breath per cycle, but (2) supplies enough oxygen for his muscle-speed demands (lest fatigue set in), and (3) delivers enough panting to prevent a meltdown (heat stroke), which depends on factors unrelated to speed such as the temperature and breeze. The outcome of these constraints is that quadrupeds have a discrete set of optimal or preferred speed settings (like the gears on a stick-shift car) for different styles of locomotion (e.g., walking, trotting and galloping). If they deviate from these preferred settings, they are operating less efficiently.

Humans lack these restrictions because (1) our lungs do not compress when we stride (we’re bipedal) so (2) our breathing rates can vary independent of our speed, and (3) our thermoregulation is managed by our fancy sweating-system, so the need to pant does not constrain our breathing. Because of this, within our range of aerobic running speeds (not sprinting), energy use doesn’t vary too much. That means we can change speeds within this range without paying much of a penalty. As a result, a skilled endurance hunter can strategically vary his speed in order to force his prey to run inefficiently. If his prey picks an initial speed just faster than the hunter, to escape, the hunter can speed up. This forces the prey to ‘shiftup’ to a much faster speed, which will cause rapid overheating. The animal’s only alternative is to run inefficiently, at a slower speed which will exhaust his muscles more quickly. The consequence is that hunters force their prey into a series of sprints and rests that eventually result in heat stroke. The overheated prey collapses, and is easily dispatched. Tarahumara, Paiute and Navajo hunters

> report that they then simply strangle the collapsed deer or pronghorn antelope.

Even locomotion is culturally learned!

> To achieve a running form that maximizes both performance and freedom from injury, humans need to rely on some cultural learning, on top of much individual practice. The evolutionary biologist and anatomist, Dan Lieberman, has studied long-distance barefoot and minimally shod running in communities around the globe. When he asks runners of all ages how they learned to run, they never say they “just knew how.” Instead, they often name or point to an older, highly skilled, and more prestigious member of their group or community, and say they just watch him, and do what he does. We are such a cultural species that we’ve come to rely on learning from others even to figure out how to run in ways that best harness our anatomical adaptations.

Why we use spices:

> Why do we use spices in our foods? In thinking about this question keep in mind that (1) other animals don’t spice their foods, (2) most spices contribute little or no nutrition to our diets, and (3) the active ingredients in many spices are actually aversive chemicals, which evolved to keep insects, fungi, bacteria, mammals and other unwanted critters away from the plants that produce them.
>
> Several lines of evidence indicate that spicing may represent a class of cultural adaptations to the problem of food-borne pathogens. Many spices are antimicrobials

that can kill pathogens in foods. Globally, common spices are onions, pepper, garlic, cilantro, chili peppers (capsicum) and bay leaves. Here’s the idea: the use of many spices represents a cultural adaptation to the problem of pathogens in food, especially in meat. This challenge would have been most important before refrigerators came on the scene. To examine this, two biologists, Jennifer Billing and Paul Sherman, collected 4578 recipes from traditional cookbooks from populations around the world. They found three distinct patterns.

1. Spices are, in fact, antimicrobial. The most common spices in the world are also the most effective against bacteria. Some spices are also fungicides. Combinations of spices have synergistic effects, which may explain why ingredients like “chili power” (a mix of red pepper, onion, paprika, garlic, cumin and oregano) are so important. And, ingredients like lemon and lime, which are not on their own potent anti-microbials, appear to catalyze the bacteria killing effects of other spices.

2. People in hotter climates use more spices, and more of the most effective bacteria killers. In India and Indonesia, for example, most recipes used many anti-microbial spices, including onions, garlic, capsicum and coriander. Meanwhile, in Norway, recipes use some black pepper and occasionally a bit of parsley or lemon, but that’s about it.

3. Recipes appear to use spices in ways that increase their effectiveness. Some spices, like onions and garlic, whose killing power is resistant to heating, are deployed in the cooking process. Other spices like cilantro, whose antimicrobial properties might be damaged by heating, are added fresh in recipes.

Thus, many recipes and preferences appear to be cultural adaptations adapted to local environments that operate in subtle and nuanced ways not understood by those of us who love spicy foods. Billing and Sherman speculate that these evolved culturally, as healthier, more fertile and more successful families were preferentially imitated by less successful ones. This is quite plausible given what we know about our species' evolved psychology for cultural learning, including specifically cultural learning about foods and plants.

Among spices, chili peppers are an ideal case. Chili peppers were the primary spice of New World cuisines, prior to the arrival of Europeans, and are now routinely consumed by about a quarter of all adults, globally. Chili peppers have evolved chemical defenses, based on capsaicin, that make them aversive to mammals and rodents but desirable to birds. In mammals, capsicum directly activates a pain channel (TrpV1), which creates a burning sensation in response to various specific stimuli, including acid, high temperatures and allyl isothiocyanate (which is found in mustard or wasabi). These chemical weapons aid chili pepper plants in their survival and reproduction, as birds provide a better dispersal system for the plants' seeds than other options (like mammals). Consequently, chilies are innately aversive to non-human primates, babies and many human adults. Capsaicin is so innately aversive that nursing mothers are advised to avoid chili peppers, lest their infants reject their breast (milk), and some societies even put capsicum on mom's breasts to initiate weaning. Yet, adults who live in hot climates regularly incorporate chilies into their recipes. And, those who grow up among people who enjoy eating chili peppers not only eat chilies but love eating them. How

> do we come to like the experience of burning and sweating—the activation of pain channel TrpV1?
>
> Research by psychologist Paul Rozin shows that people come to enjoy the experience of eating chili peppers mostly by re-interpreting the pain signals caused by capsicum as pleasure or excitement. Based on work in the highlands of Mexico, children acquire this gradually without being pressured or compelled. They want to learn to like chili peppers, to be like those they admire. This fits with what we've already seen: children readily acquire food preferences from older peers. In Chapter 14, we further examine how cultural learning can alter our bodies' physiological response to pain, and specifically to electric shocks. The bottom line is that culture can overpower our innate mammalian aversions, when necessary and without us knowing it.

Fascinating if true. But don't we use spices because of their taste? If spices are antimicrobials, why aren't there any tasteless spices? I guess you could argue most plants taste like something, usually something bad, and if a plant is a good antimicrobial then we go through the trouble of culturally reinterpreting its taste to be "exciting" or "interesting". Also, how far can this "cultural reinterpretation" idea go? Does this explain things like masochism, or like the weak form of masochism that makes people like naively unpleasant experiences like roller coasters?

* * *

I knew that Europeans had light skin because they lived in northern latitudes without much sunlight. But then how come Inuit and North Asians never developed light skin? Henrich explains:

To understand this, we need first to consider how culture has shaped genes for skin color over the last 10 millennia. Much evidence now indicates that the shades of skin color found among different populations—from dark to light—across the globe represent a genetic adaptation to the intensity and frequency of exposure to ultraviolet light, including both UVA and UVB. Near the equator, where the sun is intense year round, natural selection favors darker skin, as seen in populations near the equator in Africa, New Guinea and Australia. This is because both UVA and UVB light can dismantle the folate present in our skin, if not impeded or blocked by melanin. Folate is crucial during pregnancy, and inadequate levels can result in severe birth defects like spina bifida. This is why pregnant women are told by their physicians to take folic acid. In men, folate is important in sperm production. Preventing the loss of this reproductively valuable folate means adding protective melanin to our epidermis, which has the side effect of darkening our skin.

The threat from intense UV light to our folate diminishes for populations farther from the equator. However, a new problem pops up, as darker skinned people face a potential vitamin D deficiency. Our bodies use UVB light to synthesize vitamin D. At higher latitudes, the protective melanin in dark skin can block too much of the UVB light, and thereby inhibit the synthesis of vitamin D. This vitamin is important for the proper functioning of the brain, heart, pancreas and immune system. If a person’s diet lacks other significant sources of this vitamin, then having dark skin and living at high latitudes increases one’s chances of experiencing a whole range of health problems, including most notably rickets. A terrible condition especially in children, rickets causes muscle weakness, bone and skeletal deformities,

> bone fractures and muscle spasms. Thus, living at high latitude will often favor genes for lighter skin. Not surprising for a cultural species, many high latitude populations of hunter-gatherers (above 50-55q latitude), such as the Inuit, culturally evolved adaptive diets based on fish and marine animals, so the selection pressures on genes to reduce the melanin in their skin were not as potent as they would have been in populations lacking such resources. If these resources were to disappear from the diet of such northern populations, selection for light skin would intensify dramatically.
>
> Among regions of the globe above 50-55q latitude (e.g. much of Canada), the area around the Baltic Sea was almost unique in its ability to support early agriculture. Starting around 6,000 years ago, a cultural package of cereal crops and agricultural know-how gradually spread from the south, and was adapted to the Baltic ecology. Eventually, people became primarily dependent on farmed foods, and lacked access to the fish and other vitamin D-rich food sources that local hunter-gatherer populations had long enjoyed. However, being at particularly high latitudes, natural selection kicked in to favor genes for really light skin, so as to maximize whatever vitamin-D could be synthesized using UVB light.

Secret Of Our Success spends a lot of time talking about gene-culture coevolution and how we should expect people from different cultures to have different genes. When asked whether this is potentially racist, it argues it’s really *maximally* anti-racist, because “racism” means “believing in exactly the same racial categories as 19th century racists”, and gene-culture coevolution proves that variation is actually much more widespread than that, so there.

In case you needed proof that high status increases your inclusive fitness:

> Chris asked a sample of Tsimane to rank the men in two villages along a number of dimensions, including their fighting ability, generosity, respect, community persuasiveness, ability to get their way, and their number of allies. Each Tsimane’ man could then be assigned a score based on the aggregate results from his fellow villagers. Chris argues that his measures of fighting ability and community persuasiveness provide the best proxies for dominance and prestige, respectively, in this context. He then shows that both of these proxies for social status are associated with having more babies with one’s wife, having more extra-marital affairs, and being more likely to remarry after a divorce, even after statistically removing the effects of age, kin group size, economic productivity and several other factors. Beyond this, the children of prestigious men die less frequently and prestigious men are more likely to marry at younger ages (neither of these effects hold for dominant men). All this suggests that, at least in this small scale society, being recognized as either dominant or prestigious has a positive influence on one’s total reproductive output (children) or mating success over and above the consequences that might accrue from factors associated with status like economic productivity or hunting skills. Not surprisingly, both dominant and prestigious men tended to get their way at group meetings, but only prestigious men were respected and generous.

On the Sanhedrin:

> Effective institutions often harness or suppress aspects of our status psychology in non-intuitive ways. Take the Great Sanhedrin, the ancient Jewish court and legislature that persisted for centuries at the beginning of the Common Era. When deliberating on a capital case, its 70 judges would each share their views beginning with the youngest and lowest ranking member and then proceed in turn to the "wisest" and most respected member. This is an interesting norm because (1) it's nearly the opposite of how things would go if we let nature take its course, and (2) it helps guarantee that all the judges got to hear the least varnished views of the lower ranking members, since otherwise the views of the lowest status individuals would be tainted by both the persuasive and deferential effects of prestige and dominance. Concerns with dominance may have been further mitigated by (1) a sharing of the directorship of the Sanhedrin by two individuals, who could be removed by a vote of the judges, (2) the similar social class and background of judges, and (3) social norms that suppressed status displays.

I like this idea, but I worry it could backfire. Supposing that even the best of us are at least a little tempted to conform, it risks the youngest and least experienced members setting the tone for the discussion, so that the older and wiser members are tempted to conform with people more foolish than themselves. If the wisest people spoke first, at least we could get their untainted opinions and guarantee that any conformity was at least in favor of the opinion most likely to be correct. Overall it seems like they should have gone with secret ballots. I wonder if anyone's ever done an experiment comparing wisest-first, youngest-first, and secret-ballot decision-making to see if any have a clear advantage. You could do it with one of those "guess the number of jelly beans in this jar" tasks or something, with participants who did well on a test problem clearly marked as "elders".

On why societies often dictate naming children after their paternal relatives:

> In building a broader kinship network, social norms and practices connect a child more tightly to his or her father's side of the family, in subtle ways. In contrast to many complex societies, mobile hunter-gatherer populations often emphasize kinship through both mom and dad, and permit new couples much flexibility in where they can live after marriage. However, there's always that problem of paternity certainty for dad's entire side. Among Ju/'hoansi, mobile hunter-gatherers in the Kalahari Desert in southern Africa, social norms dictate that a newborn's father—or, more accurately, the mother's husband—has the privilege of naming the child. These norms also encourage him to name the child after either his mother or father, depending on the infant's sex. Ju/'hoansi believe name sharing helps the essence of the paternal grandparents live on, and it consequently bonds both the grandparents and the father's whole side of the family to the newborn. Relatives of the grandparents often refer to the newborn using the same kinship term they use for his or her older namesake—that is, the grandfather's daughter will call the newborn baby "father."
>
> This bias to the father's side is particularly interesting since Ju/'hoansi kinship relationships are otherwise quite gender egalitarian, emphasizing equally the links to both mom's and dad's sides of the family. This biased naming practice may help create that symmetry by evening out the imbalance that paternity uncertainty

> leaves behind. In many modern societies, where social norms favoring the father’s side have disappeared, the effect of paternity certainty emerges as maternal grandparents, uncles and aunts invest more than the same paternal relatives do. Thus, Ju/’hoansi practices link newborns directly to their father’s parents and simultaneously, via the use of close kin terms like “father” and “sister”, pull all of dad’s relatives closer

I wonder if this can be extended to our own practice of kids (mostly) taking their father’s last name rather than their mother’s.

And Joseph Henrich continues with an anecdote I eventually decided to consider cute:

> More broadly, in Ju/’hoansi society, sharing the same name is an important feature of social life, which has many economically important implications. Psychologically, creating namesakes may work in two interlocking ways. First, even among undergraduates and professors, experiments suggest that sharing the same, or even a similar, name increases people’s liking for the other person, their perceptions of similarity and their willingness to help that person. In one study, for example, professors were more likely to fill out a survey and mail it back if the cover letter was signed by someone with a name similar to their own name. The perception of similarity suggests that namesakes may somehow spark our kin psychology, since we already know we use other cues of similarity (appearance) to assess relatedness. Second, even if this same-name trick doesn’t actually spark any change in immediate feelings, it still sets the appropriate social norms—the reputational standards monitored by others—which among the Ju/’hoansi specify all kinds of important

> things about relationships, ranging from meat sharing priorities to water-hole ownership. Norms related to naming or namesake relationships are common across diverse societies, and many people in small-scale societies intuitively know the power of namesakes, as my Yasawan friends with names like Josefa, Joseteki and Joseses often remind me. My own kids are named Joshua, Jessica and Zoey, thus matching my own first name by first initial or by rhyming.

(his wife is also an anthropologist, so maybe that makes naming your kids according to anthropological phenomena easier to pull off).

Relevant to a frequent discussion here about whether polyamory is “unnatural” or at least a violation of Chesterton’s Fence:

> Even in societies with marriage, social norms and beliefs need not re-enforce concerns about sexual fidelity that arise from male pair-bonding psychology, but can instead promote investment in children in other ways. Many South American indigenous populations believe that a child forms in his or her mother’s womb through repeated ejaculations of sperm, a belief system that anthropologists have labeled partible paternity. In fact, people in many of these societies maintain that a single ejaculation cannot sustain a viable pregnancy, and men must “work hard” with repeated ejaculations over many months to sustain a viable fetus. Women, especially after the first fetus appears, are permitted, and sometimes even encouraged, to seek another man, or men, to have sex with in order to provide ‘additional fathers’ for their future child. Anyone who contributes sperm to the fetus is a secondary father. In some of

> these societies, periodic rituals prescribe extramarital sex after successful hunts, which helps establish and formalize the creation of multiple fathers. Secondary fathers — often named at birth by the mother — are expected to contribute to the welfare of their children (e.g., by delivering meat and fish), although not as much as the primary father, the mother's husband. Frequently, the secondary father is the husband's brother.
>
> Obtaining a second father is adaptive, at least sometimes. Detailed studies among both the Bari' in Venezuela and the Ache' show that kids with exactly two fathers are more likely to survive past age fifteen than kids with either one father or three or more fathers.
>
> Importantly, social norms cannot just make male sexual jealousy vanish. Men don't like it when their wives seek sex with other men. However, rather than being supported by their communities in monitoring and punishing their wives for sexual deviations, they are the one's acting defiantly—violating social norms — if they show or act on their jealousy. Reputational concerns and norms are flipped around here, so now the husband has to control himself. In the eyes of the community, it's considered a good thing for an expectant mother to provide a secondary father for her child.

Henrich adds that about 85% of human societies have practiced something other than traditionally-understood monogamy.

Suppose somebody in a weird Californian counterculture scene is trying to decide to what degree polyamory is Chesterton's-Fence-compliant. They might look around their own social network and find that most of the people they know have organically become polyamorous over the past decade or so, and decide it is the local tradition (and therefore it is good). But they could look on a broader scale and see that most people in their civilization over the past few centuries have been monogamous (and therefore polyamory is bad). Or they could look on an even broader scale and see that most people in the world throughout human history have been non-monogamous (and therefore polyamory is potentially good again). I understand other people's intuition that the "my civilization, past few hundred years" scale seems important, but I'm not sure how you would non-arbitrarily justify choosing that particular scale instead of others. The strongest argument seems to be something like "Wait two generations to see if it builds strong families", but I could see going either way.

I mentioned aversion to eating insects in the original review, but Henrich suggests some food taboos are easier to acquire than others:

> There is reason to suspect that we humans have an innate susceptibility to picking up meat aversions, due to the tendency of dead animals to carry dangerous pathogens. Thus, we humans are primed to acquire meat taboos over other food avoidances

---

More on taboos. A lot of taboos were of the form "you personally are not allowed to eat this particular meat or else something terrible will happen to you, so you might as well share it with the less fortunate instead"; this looks like a pretty transparent attempt by cultural evolution to build a social safety net. Henrich asks why these taboos persisted in the face of greed:

> A good learner will acquire this rule while growing up and never actually violate it (meat is consumed in public), so he’ll never directly experience eating the tabooed part and not having bad luck. Rare cases of taboo violation that, by coincidence, were followed by bad luck or illness will be readily remembered and passed on (psychologists call this “negativity bias”). Meanwhile, cases of violations followed by a long period when nothing bad happens will tend to be missed or forgetten, unless people keep and check accurate records.
>
> Based on my field experience, any skeptic who questions the taboos will be met with vivid descriptions of particular cases in which the taboos were violated and then poor hunting, illnesses, or bad luck ensued.

This is a huge stretch, but I wonder if you could make an argument that evolution favored confirmation bias because it helped prevent people from questioning their cultural rules.

How social norms are maintained:

> In research in the villages of Yasawa Island, my team and I have studied how norms are maintained. When someone, for example, repeatedly fails to contribute to village feasts or community labor, or violates food or incest taboos, the person’s reputation suffers. A Yasawan’s reputation is like a shield that protects them from exploitation or harm by others, often from those who harbor old jealousies or past grievances. Violating norms, especially repeatedly, causes this reputational shield to drop, and creates an opening for others to exploit the norm-violator with relative impunity. Norm violators have their property (e.g., plates, matches,

> tools) stolen and destroyed while they are away fishing or visiting relatives in other villages; or, they have their crops stolen and gardens burned at night. Despite the small size of these communities, the perpetrators of these actions often remain anonymous and get direct benefits in the form of stolen food and tools as well as the advantages of bringing down a competitor or dispensing revenge for past grievances.
>
> Despite their selfish motivations, these actions act to sustain social norms, including cooperative ones, because—crucially—perpetrators can only get away with such actions when they target a norm-violator, a person with his reputational shield down. Were they to do this to someone with a good reputation, the perpetrator would himself become a norm-violator and damage his or her reputation, thereby opening themselves up to gossip, thefts and property damage. This system, which Yasawans themselves can't explicitly lay out, thereby harnesses past grievances, jealousies and plain old self-interest to sustain social norms, including cooperative norms like contributing to village feasts.282 Thus, individuals who fail to learn the correct local norms, can't control themselves or repeatedly make mistaken violations are eventually driven from the village, after having been relentlessly targeted for exploitation.

This sounds sort of like the Icelandic legal system in Legal Systems Very Different From Ours, in that the consequence of breaking the law is that the laws cease to protect you. But viewed from a more critical angle, it also sounds like the modern "tradition" of committing (and/or tolerating) hate crimes against people who don't conform.

Speaking of hate crimes, Henrich (like me) thinks “racism” is not a natural category. He thinks ethnic hostility is much more natural than racial hostility, with the difference being that race is biological and ethnicity is culture. People are naturally friendly towards people of their own culture and skeptical of people from other cultures, which may or may not follow racial lines. He discusses an experiment in which children are asked to view a puppet playing a game incorrectly:

> We can see how deeply norms are intertwined with our folk sociology by returning to the experiments with Max the puppet. The child subjects now encounter Max along with Henri. Max speaks native-accented German but Henri speaks French-accented German. Young German children protested much more when Max —their co-ethnic as cued by accent — played the game differently from the model than when Henri did. Co-ethnics are favored because they presumably share similar norms, but that also means they are subject to more monitoring and punishment if they violate those norms. This appears to hold cross-culturally, as people from places as diverse as Mongolia and New Guinea willingly pay a cost to preferentially punish their co-ethnics in experiments like the Ultimatum Game, over their non-co-ethnics, for norm violations.
>
> This approach to how and why we think about tribes and ethnicity has broader implications. First, intergroup competition will tend to favor the spread of any tricks for expanding what members of a group perceive as their tribe. Both religions and nations have culturally evolved to increasingly harness and exploit this piece of our psychology, as they create quasi-tribes. Second, this approach means that the ingroup vs. out-group view taken by psychologists misses a key point: not all

groups are equally salient or thought about in the same way. Civil wars, for example, strongly trace to ethnically or religiously marked differences, and not to class, income or political ideology. [310] This is because our minds are prepared to carve the social world into ethnic groups, but not into classes or ideologies.

Finally, the psychological machinery that underpins how we think about ‘race’ actually evolved to parse ethnicity, not race. You might be confused by this distinction since race and ethnicity are so often mixed up. Ethnic group membership is assigned based on culturally-transmitted markers, like language or dialect. By contrast, racial groups are marked and assigned according to perceived morphological traits, such as color or hair form, which are genetically transmitted. Our folk-sociologcial abilities evolved to pick out ethnic groups, or tribes. However, cues like skin color or hair form can pose as ethnic markers in the modern world because members of different ethnic groups sometimes also share markers like skin color/hair form, and racial cues can automatically and unconsciously ‘trick’ our psychology into thinking that different ethnic groups exist. And, this byproduct can be harnessed and reified by cultural evolution to create linguistically labeled racial categories and racism.

Underlining this point is the fact that racial cues do not have cognitive priority over ethnic cues: when children or adults encounter a situation in which accent or language indicate ‘same ethnicity’ but skin color indicates ‘different race’, the ethno-linguistic markers trump the racial markers. That is, children pick as a friend someone of a different race who speaks their dialect over someone of the same race who speaks a different dialect. [311] Even weaker cues like dress can

> sometimes trump racial cues. The tendency of children and adults to preferentially learn and interact with those who share their racial markers (mistaken for ethnic cues) likely contributes to the maintenance of cultural differences between racially marked populations, even in the same neighborhood.

This ties in to my crackpot theory that the number one way to fight racism in the US is to somehow get everyone speaking exactly the same accent.

> In one well-studied case among the Gebusi, in New Guinea, my failure to meet my sister exchange obligations would increase the chances that I would, at some future date, be found guilty of witchcraft.

#out of context quotes

Henrich discusses a theory of intrinsic growth pretty similar to the one in my recent singularity post. But he introduces a neat experimental test: Polynesian islands. On larger islands (ie with higher carrying capacities), technological advance is faster:

> Islands or island clusters with larger populations and more contact with other islands had both a greater number of different fishing-tool types and more-complex fishing technologies. Figure 12.2 shows the relationship between population size and the number of tool types. People on islands with bigger populations had more tools at their disposal, and those tools tended to be more sophisticated.
>
> Another team, led by the evolutionary anthropologist Mark Collard, found the same kind of strong positive

> relationship when they examined forty nonindustrialized societies of farmers and herders from around the globe. Once again, larger populations had more-complex technologies and a greater number of different types of tools.
>
> These effects can even be observed in orangutans. While orangutans have little or no cumulative culture, they do possess some social learning abilities that result in local, population-specific traditions. For example, some orangutan groups routinely use leaves to scoop up water from the ground or use sticks to extract seeds from fruit. Data from several orangutan populations show that groups with greater interaction among individuals tend to possess more learned food-obtaining techniques.
>
> The point is, larger and more interconnected populations generate more sophisticated tools, techniques, weapons, and know-how because they have larger collective brains.

Henrich’s model is actually a little more complicated than mine, because it includes a term for forgetting technology (which actually happens pretty often when the group is small enough!) The more technology the group has, the more likely that one or two things slip through the cracks every generation and don’t get passed on to the kids. That means that most primitive societies are in an equilibrium between the rate of generating and the rate of losing technology, whose exact level depends on the population size:

> SOme information was lost every generation, because copies are usually worse than the originals. Cumulative cultural evolution has to fight against this force and is

> best able to do so in larger populations that are highly socially interconnected. The key is most individuals end up imperfect, worse than the models they are learning from. However, some few individuals, whether by luck, fierce practice, or intentional innovation, end up better than their teachers…

One point the book really drove home is how much of the absolute basics of knowledge are cultural inventions. We laugh at primitive tribes who count “one, two, many”, but the idea of counting more specifically than this was a discovery that had to be discovered by someone, and only survived when there was a context that made it useful:

> Many of the products of cumulative cultural evolution give us not only ready concepts to apply to somewhat new problems, and concepts to recombine (bows are projectiles + elastically stored energy) but actually give us cognitive tools or mental abilities that we would not otherwise have. Arabic numerals, Roman letters, the Indian zero, the Gregorian calendar, cylindrical projection maps basic color terms, clocks, fractions, and right vs. left are just some of the cognitive tools that have shaped your mind and mine

Alas, this quote is missing some context from the rest of the book showing just how hard these ideas were to develop. Remember that mathematicians spent a while debating whether “zero” was truly a number, that ancient people had what we consider very confusing concepts around color (even the Greeks were weird about this). Remember that the alphabet – breaking words up into their smallest components – arose only after millennia of logographs and syllabaries, and in some areas never arose at all. There’s even some speculation that basic ideas about introspection and emotion were invented pretty late. Or even:

> Subordinating conjunctions like “after”, “before”, and “because of” may have evolved only recently, in historical times, and are probably no more a feature of *human* languages than composite bows are a feature of *human* technological repertoires. The tools of subordination seem less well-developed in the earliest versions of Sumerian, Akkadian, Hittite, and Greek. This makes these languages slow, ponderous, and repetitious to read. ..this is not to say that we humans don’t have some souped-up innate abilities for dealing with hierarchical structures, which may also be useful for making tools or understanding social relationships, but merely that the elegant bits of grammar that permit us to fully harness these abilities were built by cultural evolution.

This kind of thing is why Henrich thinks comparing the IQ of young chimps and human toddlers is fair, but comparing older chimps and older humans isn’t. Older humans have all of these deep-level concepts to work with that make solving even abstract puzzles much easier. This is also plausibly related to the Flynn Effect.

On sonority:

> A successful communicator is one who can most effectively be understood, given the local social, environmental, or ecological conditions. As young or naïve learners focus on and learn from more successful communicators—who are using more effective communication tools—cumulative cultural evolution will gradually assemble sign or whistled repertoires, over time, in the same way that it hones kayaks, spears, and

boomerangs. Given this, there’s no reason to suspect that such cultural evolutionary processes somehow apply only to whistled or gestural sign languages, and not to typical spoken languages. Thus, spoken languages should—under the right circumstances—show some response to the local acoustic environments and to nonlinguistic social norms, just as whistled and sign languages do. While researchers have done little work on such topics, there’s some preliminary evidence.

Spoken languages vary in their sonority. The sonority of our voices decreases as the airflow used for speech is obstructed and is highest for open vowels, like the /a/ and lowest for so-called voiceless stops like the /t/ in tin. Pronounce each of these sounds and note the difference in the constriction of your airflow. Both vowels and consonants vary in sonority, but vowels generally have much higher sonority than consonants. This means that more sonorous languages tend to have more vowels (e.g., Hawaiian), while less sonorous ones pack the consonants together (e.g., Russian). For the same energy and effort, more sonorous speech sounds can be heard at greater distances and over more ambient noise than less sonorous ones.

If languages adapt culturally, then we can predict that in situations in which people do relatively more talking over greater interpersonal distances with more ambient noise and sound dispersion, languages will be more sonorous. Many environmental variables might influence this, but Robert Monroe, John Fought, and their colleagues reasoned that climate, and specifically temperature, might have a big effect. The idea is simple: in warmer climates, people work, play, cook, and relax outdoors. Compared to life indoors, living outside means that communicators more frequently

> face the challenges of distance, noise and poor acoustics. Their team generated measures of sonority from word lists for dozens of languages and then looked at the relationship between sonority and measures of climatic temperature, like the number of months per year when it’s below 10°C (50°F).
>
> It turns out that if all you know is climatic temperature, then you can account for about one-third of the variation in the sonority of languages. Languages in warmer climates tend to use more vowels than those in colder climates and rely more heavily on the most sonorous vowel, /a/. For consonants, languages in warmer climates rely more heavily on the most sonorant consonants, like /n/, /l/, and /r/. By contrast, languages in colder climates lean more heavily on the least sonorous vowels, as the /i/ in deep.[7]
>
> This simple idea can have much nuance added to it. For example, not all warm climates are equally conducive to sonorous speech. In regions with dense forest cover, the advantages of high sonority might be less pronounced, or as the anthropologists Mel and Carol Ember have argued, very cold and windy climates may select against linguistic practices that involve opening one’s mouth widely, due to the increased heat loss. To this they added the idea that social norms about sexual restrictiveness might also influence sonority. Adding both of these nuances to the basic climatic temperature analysis, they managed to account for four-fifths of the variation in the sonority of language.

I’m a *little* worried about p-hacking here, but still, whoa! The thing where Inuit languages sound like tikkakkooktttippik but Polynesian languages sound like waoiuhieeawahiaii has a cause! The phonetic nature of words is shaped by the experience of the people who produce them! There’s something delightfully kabbalistic about this.

The chili pepper quote promised a study on cultural learning of pain, so here it is:

> Ken Craig has directly tested the relationship between cultural learning and pain. Ken’s team first exposed research participants to a series of electric shocks that gradually increased in intensity and thus painfulness. Some participants observed another person – a “tough model” – experience the same shocks right after them, and some did not. Both the participant and model had to rate how painful the shock was each time. The tough model, however, was secretly working for the experimenter and always rated the pain about 25% less painful than the participant did. Then, after this, the model left and the participants received a series of random electric shocks. For this new series of shocks, the participants who had seen the tough model rated them half as painful as those who didn’t see the tough model….
>
> Those who saw the tough model showed (1) declining measurements of electrodermal skin potential, meaning that their bodies stopped reacting to the threat, (2) lower and more stable heart rates, and (3) lower stress ratings. Cultural learning from the tough model changed their physiological reactions to electric shocks.

I see a commenter on Quillette has already thought to connect this to telling people they should be harmed by triggers and microaggressions. But also note the connection to the the predictive processing model of perception.

Books like this are supposed to end with an Exhortation Relevant To Modern Society, so here's Henrich's:

> Humans are bad at intentionally designing effective institutions and organizations, though I'm hoping that as we get deeper insights into human nature and cultural evolution this can improve. Until then, we should take a page from cultural evolution's playbook and design "variation and selection systems" that will allow alternative institutions or organizational forms to compete. We can dump the losers, keep the winners, and hopefully gain some general insights during the process.

If that sounds familiar, it could be because it's capitalism; if it sounds very familiar, it could be because it's also the case for things like charter cities and seasteads; if it sounds super familiar, it could be because it's also Archipelago.

And to finish:

> Once we understand the importance of collective brains, we begin to see why modern societies differ in their innovativeness. It's not the smartness of individuals or the formal incentives. It's the willingness and ability of large numbers of individuals at the knowledge frontier to freely interact, exchange views, disagree, learn from each other, build collaborations, trust strangers, and be wrong.

Hopefully this means Henrich won't be *too* angry that I just quoted like half of his copyrighted book without permission.

Asymmetric Weapons Gone Bad

[Previously in sequence: Epistemic Learned Helplessness, Book Review: The Secret Of Our Success, List Of Passages I Highlighted In My Copy Of The Secret Of Our Success. Deleted a controversial section which I still think was probably correct, but which given the number of objections wasn't provably correct enough to be worth including. I might write another post giving my evidence for it later, but it probably shouldn't be dropped in here without justification.]

I.

Years ago, I wrote about symmetric vs. asymmetric weapons.

A symmetric weapon is one that works just as well for the bad guys as for the good guys. For example, violence – your morality doesn't determine how hard you can punch; they can buy guns from the same places we can.

An asymmetric weapon is one that works better for the good guys than the bad guys. The example I gave was Reason. If everyone tries to solve their problems through figuring out what the right thing to do is, the good guys (who are right) will have an easier time proving themselves to be right than the bad guys (who are wrong). Finding and using asymmetric weapons is the only non-coincidence way to make sustained moral progress.

The parts of *The Secret Of Our Success* that deal with reason vs. cultural evolution raise a disturbing prospect: what if sometimes, the asymmetry is in the wrong direction? What if there are some issues where rational debate inherently leads you astray?

II.

Maybe with an unlimited amount of resources, our investigations would naturally converge onto the truth. Given infinite intelligence, wisdom, impartiality, education, domain knowledge, evidence to study, experiments to perform, and time to think it over, we would figure everything out.

But just because infinite resources will produce truth doesn't mean that truth as a function of resources has to be monotonic. Maybe there are some parts of the resources-vs-truth curve where increasing effort leads you the wrong direction.

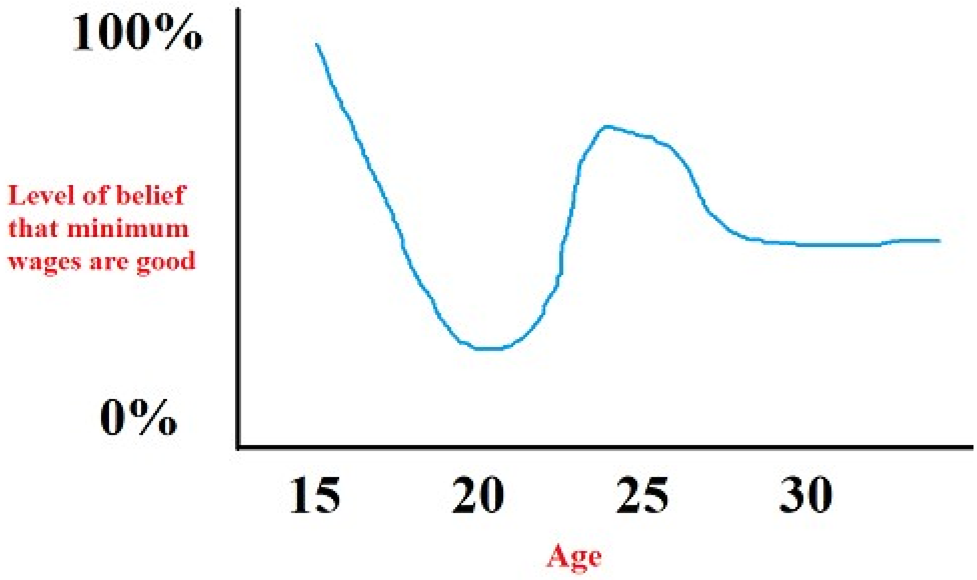
When I was fifteen, I thought minimum wages obviously helped poor people. They needed money; minimum wages gave them money, case closed.

When I was twenty, and a little wiser, I thought minimum wages were obviously bad for the poor. Econ 101 tells us minimum wages kill jobs and cause deadweight loss, with poor people most affected. Case closed.

When I was twenty-five, and wiser still, I thought minimum wages were probably good again. I'd read a couple of studies showing that maybe they *didn't* cause job loss, in which case they're back to just giving poor people more money.

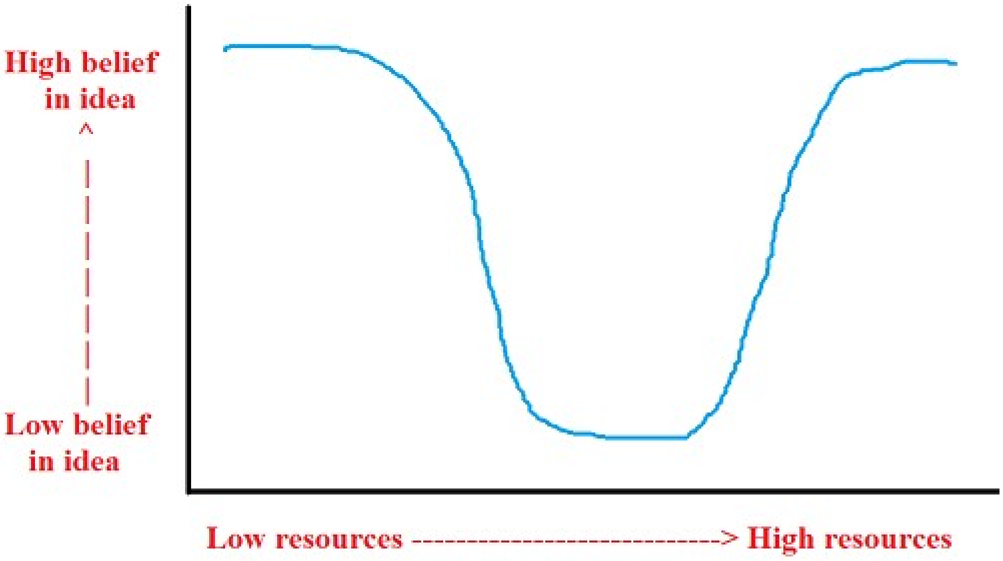
When I was thirty, I was hopelessly confused. I knew there was a meta-analysis of 64 studies that showed no negative effects from minimum wages, and a systematic review of 100+ studies that showed strong negative effects from minimum wages. I knew a survey of economists found almost 80% thought minimum wages were good, but that a different survey of economists found 73% thought minimum wages were bad.

We can graph my life progress like this:



This partly reflects my own personal life course, which arguments I heard first, and how I personally process evidence.

But another part of it might just be inherent to the territory. That is, there are some arguments that are easy to understand, and other arguments that are harder to understand. If the easy arguments lean predominantly one way, and the hard arguments lean predominantly the other way, then it will natural for any well-intentioned person studying a topic to follow a certain pattern of switching their opinion a few times before getting to the truth.



Some hard questions might be epistemic traps – problems where the more you study them, the wronger you get, up to some inflection point that might be further than anybody has ever studied them before.

III.

We'll get to vast social conflicts eventually, but I want to start with boring things in everyday life.

I hate calling people on phones. I can't really explain this. I'm okay with emailing them. I'm okay talking to them in person. But I hate calling them on phones.

When I was younger, I would go to great lengths to avoid calling people on phones. My parents would point out that this was dumb, and ask me to justify it. I couldn't. They would tell me I was being silly. So I would call people on phones and hate it. Now I don't live with my parents, nobody can make me do things, and so I am back to avoiding phone calls.

My parents weren’t authoritarian. They weren’t demanding I make phone calls because That Is The Way We Do Things In This House. They were doing the supposedly-correct thing, using rational argument to make me admit my aversion to phone calls was totally unjustified, and that making phone calls had many tangible benefits, and then telling me I should probably make the call, shouldn’t I? Yet somehow this ended up making my life worse.

Or: I can’t do complicated intellectual work with another person in the room. I just can’t. You can give me good reasons why I’m wrong about this: maybe the other person won’t make any noise. Maybe I can just turn the other way and focus on my computer and I won’t ever have to notice the other person’s presence at all. Argue this with me enough, and I will lose the argument, and work in the same room as you. I won’t get any good work done, and I’ll end up spending most of the time hating you and wishing you would go away.

I try to be very careful with my patients, so that I don’t make their lives worse in the same way. It’s often easy to get patients to admit they don’t have a good reason for what they’re doing; for example, autistic people usually can’t explain why they “stim”, ie make unusual flapping movements. These movements are distracting and probably creep out the people around them. It’s very easy to argue an autistic person into admitting they stimming is a net negative for them. Yet somehow autistic people always end up hating the psychiatrists who win this argument, and going somewhere far away from them so they can stim in peace.

Every day we do things that we can’t easily justify. If someone were to argue that we shouldn’t do the thing, they would win easily. We would respond by cutting that person out of our life, and continuing to do the thing.

I hope most readers find at least one of the examples above rang true to them. If not – if you don't hate phones, or have trouble working near others, or stim – and if you're thinking "All of those things really do seem irrational, you're probably just wrong if you want to protect them against Reason" – here are some potential alternative intuition pumps:

1. Guys – do you have trouble asking girls out? Why? The worst that can happen is they'll say no, right?

2. Girls – do you sometimes get upset and flustered when a guy you don't like asks you out, even in a situation where you don't fear any violence or coercion from the other person? Do you sometimes agree to things you don't want because you feel pressured? Why? All you have to do is say "I'm flattered, but no thanks".

3. Do you diet and exercise as much as you should? Why not? Obviously this will make you healthier and feel better! Why don't you buy a gym membership right now? Are you just being lazy?

I don't mean to say these questions are Profound Mysteries that nobody can possibly answer. I think there are good answers to all of them – for example, there are some neurological theories that offer a pretty good explanation of how stimming helps autistic people feel better. But I do want to claim that most of the people in these situations don't know the explanations, and that it's unreasonable to expect them to. All of these actions and concerns are "illegible" in the *Seeing Like A State* sense.

Illegibility is complicated and context-dependent. Fetishes are pretty illegible, but because we have a shared idea of a fetish, because most people have fetishes, and because even the people who don't have fetishes have the weird-if-you-think-about-it habit of being sexually attracted to other human beings – people can just say "That's my fetish" and it becomes kind of legible. We don't question it. And there are all sorts of phrases like "I don't like it", or "It's a free country" or "Because it makes me happy" that sort of relieve us of the difficult work of maintaining legibility for all of our decisions.

This system works so well that it only breaks down when very different people try to communicate across a fundamental gap. For example, since allistic people may not feel any urge to stim or do anything like stimming, its illegibility suddenly becomes a problem, and they try to argue autistic people out of it. The worst failure mode is where illegible actions by an outgroup are naturally rounded off to "they are evil and just hiding it". I remember feeling pretty bad once after hearing a feminist explain that the only reason men stared at attractive women was to intimidate them, make them feel like their body existed for other people's pleasure, and cement male privilege. I myself sometimes stared at attractive women, and I couldn't verbalize a coherent reason – was I just trying to hurt and intimidate them? I think a real answer to this question would involve the way we process salience – we naturally stare at the most salient part of a scene, and an attractive person will naturally be salient to us. But this was beyond teenaged me's ability to come up with, so I ended up feeling bad and guilty.

If you force people to legibly interpret everything they do, or else stop doing it under threat of being called lazy or evil, you make their life harder and probably just end up with them avoiding you.

IV.

Different problems come up when we talk about societies trying to reason collectively. We would like to think that the more investigation and debate our society sinks into a question, the more likely we are to get the right answer. But there are also times when we do 450 studies on something and end up more wrong than when we started.

A very boring, trivial example of this: I think we should increase salaries for Congress, Cabinet Secretaries, and other high officials. There are so few of this that it would be very cheap: quintupling every Representative, Senator, and Cabinet Secretary’s salary to $1 million/year would involve raising taxes by only $2 per person. And if it attracted even a slightly better caliber of candidate – the type who made even 1% better decisions on the trillion-dollar questions such leaders face – it would pay for itself hundreds of times over. Or if it prevented just a tiny bit of corruption – an already rich Defense Secretary deciding from his gold-plated mansion that there was no point in going for a “consulting job” with a substandard defense contractor – again, hundreds of times over. This isn’t just me being a elitist shill: even Alexandria Ocasio-Cortez agrees with me here. This is as close to a no-brainer as policies come.

But I think I would be demolished if I tried to argue for this on Twitter, or on daytime TV, or anywhere else that promotes a cutthroat culture of “dunking” on people with the wrong opinions. It’s *so much* faster, easier, and punchier to say “poor single mothers are starving on minimum wage, and you think the most important problem is taking money away from them to make our millionaires even richer?” and just drown me out with cries of “elitist shill, elitist shill” every time I try to give the explanation above. Sure enough, the AOC article above notes that although Americans underestimate the amount Congressmen get paid (they think only $120,000, way less than the real number of $170,000), most of them believe they should be paid less, with only 17% saying they should keep getting what they already have, and only 9% agreeing they should get more.

This is a different problem than the one above – the policy isn’t illegible to the people trying to defend it, but the communication methods are low-bandwidth enough that the most legible side naturally wins. That Congressmen are even able to maintain their current salary is partly due to them being insulated from debate: the issue never really comes up, so the consensus in favor of cutting their pay doesn’t really matter.

And yeah, I know, Popular Opinion Sometimes Wrong, More At 11. But this seems like a trivial but real society-wide case of the epistemic traps above, where if you increase one resource (amount an issue is debated) without increasing other resources (intelligence and rationality of the participants, the amount of time and careful thought they are willing to put in) you get further away from truth.

V.

Are there any less trivial examples? What about turn-of-the-20th-century socialism?

I was shocked to learn how strong a pro-socialism consensus existed during this period among top intellectuals. Socialist leader Edward Pease described the landscape pretty well:

> Socialism succeeds because it is common sense. The anarchy of individual production is already an anachronism. The control of the community over itself extends every day. We demand order, method, regularity, design; the accidents of sickness and

> misfortune, of old age and bereavement, must be prevented if possible, and if not, mitigated. Of this principle the public is already convinced: it is merely a question of working out the details. But order and forethought is wanted for industry as well as for human life. Competition is bad, and in most respects private monopoly is worse. No one now seriously defends the system of rival traders with their crowds of commercial travellers: of rival tradesmen with their innumerable deliveries in each street; and yet no one advocates the capitalist alternative, the great trust, often concealed and insidious, which monopolises oil or tobacco or diamonds, and makes huge profits for a fortunate; few out of the helplessness of the unorganised consumers.

Why shouldn't people have thought this? The period featured sweatshop-like working conditions alongside criminally rich nobility with no sign that this state of affairs could ever change under capitalism. Top economists, up until the 1950s, almost unanimously agreed that socialism would help the economy, since central planners could coordinate ways to become more efficient. The first good arguments against this proposition, those of Hayek and von Mises, were a quarter-century in the future. Communism seemed perfectly straightforward and unlikely to go wrong; the first hint that it "might not work in real life" would have to wait for the Bolshevik Revolution. Pease writes that the main pro-capitalism argument during his own time was the Malthusian position that if the poor got more money, they would keep breeding until the Earth was overwhelmed by overpopulation; even in his own time, demographers knew this wasn't true. The imbalance in favor of pro-communist arguments over pro-capitalist ones was overwhelming.

Don't trust me on this. Trust all the turn-of-the-20th-century intellectuals who flocked towards socialism. In the Britain of the time, the smarter you were, and the more social science and economics you knew, the more likely you were to be a socialist, with only a few exceptions.

But turn-of-the-century Britain never went communist. Why not?

One school of thought says it's because rich people had too much power. Even though the intellectuals all supported communism, nobody wanted to start a violent revolution, because they expected the rich to win and punish them.

But another school of thought says that cultural evolution created both capitalism, and an immune system to defend capitalism. This is more complicated, and requires a lot of the previous discussion here before it makes sense. But it seems to match some of what was going on. Society didn't *look like* everyone wanting to revolt but being afraid of the rich. It looked like large parts of the poor and middle class being very anti-communist for kind of illegible reasons like "king" and "country" and "God" and "tradition" or "just because".

In retrospect, these illegible reasons were right. It's hard to tell if they were right by coincidence, or because cultural evolution is smarter than we are, drags us into whatever decision it makes, and then creates illegible reasons to prop itself up.

Empirically, as people started devoting more intellectual resources to the problem of whether Britain should be communist or not – as very intelligent and well-educated people started thinking about the problem using the most modern ideas of science and rationality, and challenged all of their preconceived notions to see which ones would stand up to Reason and which ones wouldn't – they got further from the truth.

(I'm assuming that you, the reader, aren't communist. If you are, think up another example, I guess.)

There *is* a level of understanding that lets you realize communism is a bad idea. But you need a lot of economic theory and a lot of retrospective historical knowledge the early-20th-century British didn't have. There's some part in the resources-vs-truth graph, where you're smart enough to know what communism is but not smart enough to have good arguments against it – where the more intellect you apply the further from truth it takes you.

VI.

Obviously this ends with everyone agreeing to think very hard about things, carefully distinguish notice which traditions have illegible justifications, and then only throw out the traditions that are legitimately stupid and exist for no reason. What other position could we come to? You wouldn't say "Don't bother being careful, nothing is ever illegible". But you also can't say "Okay, we will never change anything ever again". You just give the maximally-weaselly answer of "We'll be sure to think about it first."

But somebody made a good point on the last comments thread. We are the heirs to a five-hundred-year-old tradition of questioning traditions and demanding rational justifications for things. Armed with this tradition, western civilization has conquered the world and landed on the moon. If there were ever any tradition that has received cultural evolution's stamp of approval, it would be this one.

So is there anything at all we should learn from all of this? If I had to cache out "think very hard about things" more carefully, maybe it would look like this:

1. The original Chesterton's Fence: try to understand traditions before jettisoning them.

2. If someone does something weird but can't explain why, accept them as long as they're not hurting anyone else (and don't make up stupid excuses for why their actions *really* hurt all of us). Be less quick to jump to "actually they are doing it out of Inherent Evil" as an explanation.

3. As per the last Henrich quote here, make use of the "laboratories of democracy" idea. Try things on a small scale in limited areas before trying them at larger scale; let different polities compete and see what happens.

4. Have less intense competitive pressure in the marketplace of ideas. Kuhn touches on how heliocentric theory had less explanatory power than geocentric theory for a while, but was tolerated anyway long enough that it was eventually able to sort itself out and become better. If good ideas are sometimes at a disadvantage in defending themselves, leave unpopular opinions alone for a while to see if they eventually become more legible. I think this might look like just being kinder and more tolerant of weirdness.

5. If someone defends a tradition that seems completely wrong and repulsive to you, try to be understanding of them *even if you are right and the tradition is wrong*. Traditions spent a long time evolving to be as sticky as possible in the face of contrary evidence, humans spent a long time evolving to stick to traditions as much as possible in the face of contrary evidence, and this evolution was beneficial through most of history. This sort of pressure is as hard to break (and probably as genetically-loaded) as other now-

obsolete evolutionary urges like the one to binge on as much calorie-dense food as possible when it's available (related).

6. Having done all that, and working as gingerly and gradually as you can, you should still try to improve on traditions that seem obsolete or improvable.

7. Cultural evolution does not provide evidence that traditions are ethical. Like biological evolution, cultural evolution didn't even *try* to create ethical systems. It tried to create systems that were good at spreading. Plausibly many cultures converged on eating meat because it was a good source of calories and nutrients. But if you think it violates animals' rights, cultural evolution shouldn't convince you otherwise – there's no reason cultural evolution should price animal suffering into its calculations. (related).

Finally: some people have interpreted this series of posts as a renunciation of rationality, or an admission that rationality is bad. It isn't. Rationality isn't (or shouldn't be) the demand that every opinion be legible and we throw out cultural evolution. Rationality is the art of reasoning correctly. I don't know what the optimal balance between what-seems-right-to-us vs. tradition should be. But whatever balance we decide on, better correlating "what seems right to us" with "what is actually true" will lead to better results. If we're currently abysmal at this task, that only adds urgency to figuring out where we keep going wrong and how we might go less wrong, both as individuals and as a community.

Addendum To “Enormous Nutshell”: Competing Selectors

[Previously in sequence: Epistemic Learned Helplessness, Book Review: The Secret Of Our Success, List Of Passages I Highlighted In My Copy Of The Secret Of Our Success, Asymmetric Weapons Gone Bad]

When I wrote Reactionary Philosophy In An Enormous Planet-Sized Nutshell, my attempt to explain reactionary philosophy, many people complained that it missed the key insight. At the time I had an excuse: I didn’t get the key insight. Now I think I might understand it and have the vocabulary to explain, so I want to belatedly add it in.

The whole thing revolves around this rather dubious redefinition:

RIGHT-WING: Policies and systems selected by cultural evolution
LEFT-WING: Policies and systems selected by the marketplace of ideas

The second line is ambiguous: which marketplace of ideas, exactly? Maybe better than “the marketplace of ideas” would be “memetic evolution”. Policies and systems that are so catchy and convincing that lots of people believe in them and want to fight for them.

Under this definition, lots of conventionally right-wing movements get defined as left-wing. For example, Nazism and Trumpism both arose after a charismatic leader convinced the populace to implement them. They won because people liked them more than the alternatives. But “left-wing” is not equivalent to “populist”. An idea that spreads by convincing intellectuals and building an academic consensus around itself is still left-wing, because it relies on convincing people. Even ideas like neoliberalism and technocracy are left-wing ideas, if they sound good to intellectuals and they spread by convincing those intellectuals.

Does this mean that in this model, fascism, communism, and liberalism are all left-wing ideas? Yes. Most democracies can be expected to have mostly (entirely?) left-wing parties, since the whole point of being a party in a democracy is that you have to convince voters of things and win their approval. It's not impossible to imagine a successful right-wing party in a democracy – it would revolve around preserving tradition, and if respect for tradition was strong enough, it might temporarily win. But it's not a very stable situation.

What prevents every democracy from instantly becoming maximally left-wing? First, cultural evolution has built itself an immune system in the form of traditions and illegible preferences for certain ideas. Second, cultural evolution is still at work. If incumbents pursue some popular policy that ends up bankrupting their city, or causing crimes rates to increase 1000%, or something like that, they will end up humiliated, and people will probably vote them out of office. Incumbents know this, and so put some self-interested effort into rejecting these policies even if they are very popular and convincing.

(I think in this model, greed / special interests / NIMBYism are all special cases of convincingness. If an idea is in my self-interest, it will be very convincing to me; if I am powerful enough to sabotage the system or force things through it, the idea will have won through its convincingness.)

The reactionaries start with the assumption that some problems are asymmetric in the wrong direction. The correct idea sounds unconvincing; wrong ideas spread like wildfire and naturally win debates. I talked about two examples of this yesterday: Congressional salaries and early 20th century Communism. Most questions probably aren't like this – "don't nuke the ocean for no reason" is both convincing-sounding and adaptive. But where they diverge, you want to develop a system capable of implementing the right-wing answer even though there will be intense pressure from activists and the masses to implement the left-wing one.

What would a country capable of doing this look like? It would have to be a place where convincing-sounding ideas were incapable of spreading and taking over. That would mean that the beliefs of the populace would be completely irrelevant to what policies got enacted. So it couldn't be a democracy. But it also couldn't be an ordinary dictatorship. Churchill tells us that "dictators ride on tigers from which they dare not dismount" – they have to constantly maintain the support of the army and elites in order to avoid being deposed, and that involves doing things that sound good (at least to the army and elites) and are easy to justify (again, to them). You would need an implausibly strong dictatorship in order to resist the pressure to do whatever is easiest to justify, and so to escape being left-wing.

But even this would not be right-wing. Whatever convincing ideology has won the approval of the populace might also win the approval of the dictator, who would then do it because he wants to. Also, the dictator might be an idiot, or insane, and do bad policy for reasons other than because he is under the spell of some convincing-but-wrong idea.

The reactionaries believe there is no way to guarantee a country works well. But there is a way to guarantee that a collection of countries works well, which is to create a system conducive to cultural evolution. Have a bunch of small countries, each of which is ruled by an absolute dictator. In some of them, the dictator will pursue good policy, people and investment will flow in, and those countries will flourish. In others, the dictator will pursue bad policy, and those countries will either collapse, or do the smart thing and adopt the behavior of flourishing countries.

The argument isn’t that dictators are naturally smarter than the masses. The argument is that the dictators will be a high-variance group. Some of them will probably be stupid. But get enough countries like this, and at least one of them will have a dictator who really is cleverer than the masses. That country will succeed beyond what a left-wing country yoked to the most convincing-sounding idea would be capable of. Then other countries will copy its success or be left behind.

(are we sure dictatorships are higher variance than democracies? I think it makes intuitive sense that a single individual would be higher-variance than the average of a crowd. Also, democracies can be expected to develop activists and journalists who will intensify memetic selection and force convergence on the most memetically fit policy. If the democracies are culturally different, the most memetically fit policy might be different for each. But these cultural differences are themselves products of cultural evolution and could be expected to erode under enough pressure.)

There’s a clear analogy to business. Hundreds of entrepreneurs try to start their own companies. Many are idiots and fail immediately. But one of them is Jeff Bezos and very good at his job. His company makes the right decisions and ends up dominating the entire market. “The best practices spread everywhere” is the desired outcome; cultural evolution has succeeded. Abstracting away potential venture capitalist involvement, none of this requires Jeff Bezos’ business plan to sound convincing to a third party; memetic selection is not involved.

(if business worked like politics, each of those hundreds of e-commerce entrepreneurs would go before a panel of voters and explain why their ideas were the best; whoever sounded most convincing would win. I see no reason to believe Jeff Bezos is especially good at convincing people of things. Honestly, “first we make a mail order bookstore, then we conquer the world” sounds like a pretty dumb business plan.)

Henrich summarizes the political implications of *The Secret Of Our Success* as:

> Humans are bad at intentionally designing effective institutions and organizations, though I’m hoping that as we get deeper insights into human nature and cultural evolution this can improve. Until then, we should take a page from cultural evolution’s playbook and design “variation and selection systems” that will allow alternative institutions or organizational forms to compete. We can dump the losers, keep the winners, and hopefully gain some general insights during the process.

The reactionary model of government is an attempt to cache out Henrich’s “variation and selection system”, and shares its advantages. But what’s the case against it?

First, turning the world into a patchwork of thousands of implausibly strong dictatorships sounds about as hard as starting a global communist revolution or implementing any other fundamental change to the system of the world.

Second, cultural evolution at the international level may not work quickly enough to be at all useful or humane. Plausibly World War II provided one bit of cultural-evolution data ("fascism is worse than liberalism"). The Cold War provided a second bit ("communism is also worse than liberalism"). Both bits are appreciated, but 50 million deaths per bit is a pretty high price. If the world were a patchwork of tiny dictatorships, there would probably be a lot of war and genocide and oppression before we learned anything.

Third, we have to hope that cultural evolution would be selecting for the happiest and most prosperous countries. There's a case that it would, if everyone has exit rights and can vote with their feet for countries they like better. But there's also a risk it selects for military might, or that exit rights don't happen. Dubai, whose position in the United Arab Emirates makes it a lot closer to this model than most places, seems to invest a lot in its citizens' happiness, but also has an underclass of near-slave laborers without exit rights (their employers tend to seize their passports). Also, a lot of industries have pretty bad conditions for their employees, even though those employees have exit rights to go to different companies. I don't really understand why this happens, but it sounds like the sort of thing that could happen in a patchwork of small dictatorships too.

Finally, and appropriately for a system that loathes convincingness, the branding is terrible. Using "right" and "left" for the two sides was an bad decision. Absent that decision, I don't think there's anything necessarily rightist about it. Certainly it exemplifies leftist virtues like localism and diversity; certainly it gets points for identifying Nazism and Trumpism as bad and proposing a way to stop them. Certainly it should be tempting for communists who have realized they’re not going to get a revolution in western countries any time soon but still want a chance to prove their ideas can work. I think this bad branding decision caused a downstream cascade of awfulness, leading to reaction attracting a lot of very edgy people who liked the idea of being “maximally rightist”. Some of these people later became alt-right or Trump supporters, the media caught on, and the idea ended up discredited for totally contingent reasons.

Also on the subject of bad branding, it was an unforced error to focus on kings. The theory is pointing at something like Singapore, Dubai, or charter cities (but also utopian communes, and monasteries, and…) Medieval kings aren’t just a couple of centuries out of date, they’re also bad examples: most of them had very limited power to go against what nobles wanted. They probably stuck to cultural evolution rather than memetic evolution just because that was how things worked in the Middle Ages before the printing press, but they don’t seem to have had a coherent theory of this.

Despite these flaws, I find myself thinking about this more and more. Cultural evolution may be moving along as lazily as always, but memetic evolution gets faster and faster. Clickbait news sites increase the intensity of selection to tropical-rainforest-like levels. What survives turns out to be conspiracy-laden nationalism and conspiracy-laden socialism. The rise of Trump was really bad, and I don’t think it could have happened just ten or twenty years ago. Some sort of culturally-evolved immune system (“basic decency”) would have prevented it. Now the power of convincing-sounding ideas to spread through and energize the populace has overwhelmed what that kind of immunity can deal with.

We should try to raise the sanity waterline – make true things more convincing than false things. But at the same time, we may also want to try to to understand the role of cultural evolution as a counterweight to memetic evolution, and have ideas for how to increase that role in case of emergency.

Book Review: Why Are The Prices So D*mn High?

Why have prices for services like health care and education risen so much over the past fifty years? When I looked into this in 2017, I couldn't find a conclusive answer. Economists Alex Tabarrok and Eric Helland have written a new book on the topic, *Why Are The Prices So D*mn High?* (link goes to free pdf copy, or you can read Tabarrok's summary on Marginal Revolution). They do find a conclusive answer: the Baumol effect.

T&H explain it like this:

> In 1826, when Beethoven's String Quartet No. 14 was first played, it took four people 40 minutes to produce a performance. In 2010, it still took four people 40 minutes to produce a performance. Stated differently, in the nearly 200 years between 1826 and 2010, there was no growth in string quartet labor productivity. In 1826 it took 2.66 labor hours to produce one unit of output, and it took 2.66 labor hours to produce one unit of output in 2010.
>
> Fortunately, most other sectors of the economy have experienced substantial growth in labor productivity since 1826. We can measure growth in labor productivity in the economy as a whole by looking at the growth in real wages. In 1826 the average hourly wage for a production worker was $1.14. In 2010 the average hourly wage for a production worker was $26.44, approximately 23 times higher in real (inflation-adjusted) terms. Growth in average labor productivity has a surprising implication: it makes the output of slow productivity-growth sectors (relatively) more expensive. In 1826, the

> average wage of $1.14 meant that the 2.66 hours needed to produce a performance of Beethoven’s String Quartet No. 14 had an opportunity cost of just $3.02. At a wage of $26.44, the 2.66 hours of labor in music production had an opportunity cost of $70.33. Thus, in 2010 it was 23 times (70.33/3.02) more expensive to produce a performance of Beethoven’s String Quartet No. 14 than in 1826. In other words, one had to give up more other goods and services to produce a music performance in 2010 than one did in 1826. Why? Simply because in 2010, society was better at producing other goods and services than in 1826.

Put another way, a violinist can always choose to stop playing violin, retrain for a while, and work in a factory instead. Maybe in 1826, when factory owners were earning $1.14/hour and violinists were earning $5/hour, so no violinists would quit and retrain. But by 2010, factory workers were earning $26.44/hour, so if violinists were still only earning $5 they might *all* quit and retrain. So in 2010, there would be a strong pressure to increase violinists’ wage to at least $26.44 (probably more, since few people have the skills to be violinists). So violinists must be paid 5x more for the same work, which will look like concerts becoming more expensive.

This should happen in every industry where increasing technology does not increase productivity. Education and health care both qualify. Although we can imagine innovative online education models, in practice one teacher teaches about twenty to thirty kids per year regardless of our technology level. And although we can imagine innovative AI health care, in practice one doctor can only treat ten or twenty patients per day. Tabarrok and Helland say this is exactly what is happening. They point to a few lines of evidence.

First, costs have been increasing very consistently over a wide range of service industries. If it was just one industry, we could blame industry-specific factors. If it was just during one time period, we could blame some new policy or market change that happened during that time period. Instead it’s basically omnipresent. So it’s probably some kind of very broad secular trend. The Baumol effect would fit the bill; not much else would.

Second, costs seemed to increase most quickly during the ’60s and ’70s, and are increasing more slowly today. This fits the growth of productivity, the main driver of the Baumol effect. Between 1950 and 2010, the relative productivity of manufacturing compared to services increased by a factor of six, which T&H describe as “of the same order as the growth in relative prices”. This is what the violinist-vs-factory-worker model of the Baumol effect would predict.

Third, competing explanations don’t seem to work. Some people blame rising costs on “administrative bloat”. But administrative costs as a share of total college costs have stayed fixed at 16% from 1980 to today (really?! this is fascinating and surprising). Others blame rising costs on overregulation. But T&H have a measure for which industries have been getting more regulated recently, and it doesn’t really correlate with which industries have been getting more expensive (wait, did they just disprove that regulation hurts the economy? I guess regulation isn’t a random shock, so this isn’t *proof*, but it still seems like a big deal). They’re also able to knock down industry-specific explanations like medical malpractice suits, teachers unions, etc.

Fourth, although service quality has improved a little bit over the past few decades, T&H provide some evidence that this explains only a small fraction of the increase in costs. Yet education and health care remain as popular (maybe more popular) than ever. They claim that very few things in economics can explain simultaneous increasing cost, increasing demand, and constant quality. One of those few things is the Baumol effect.

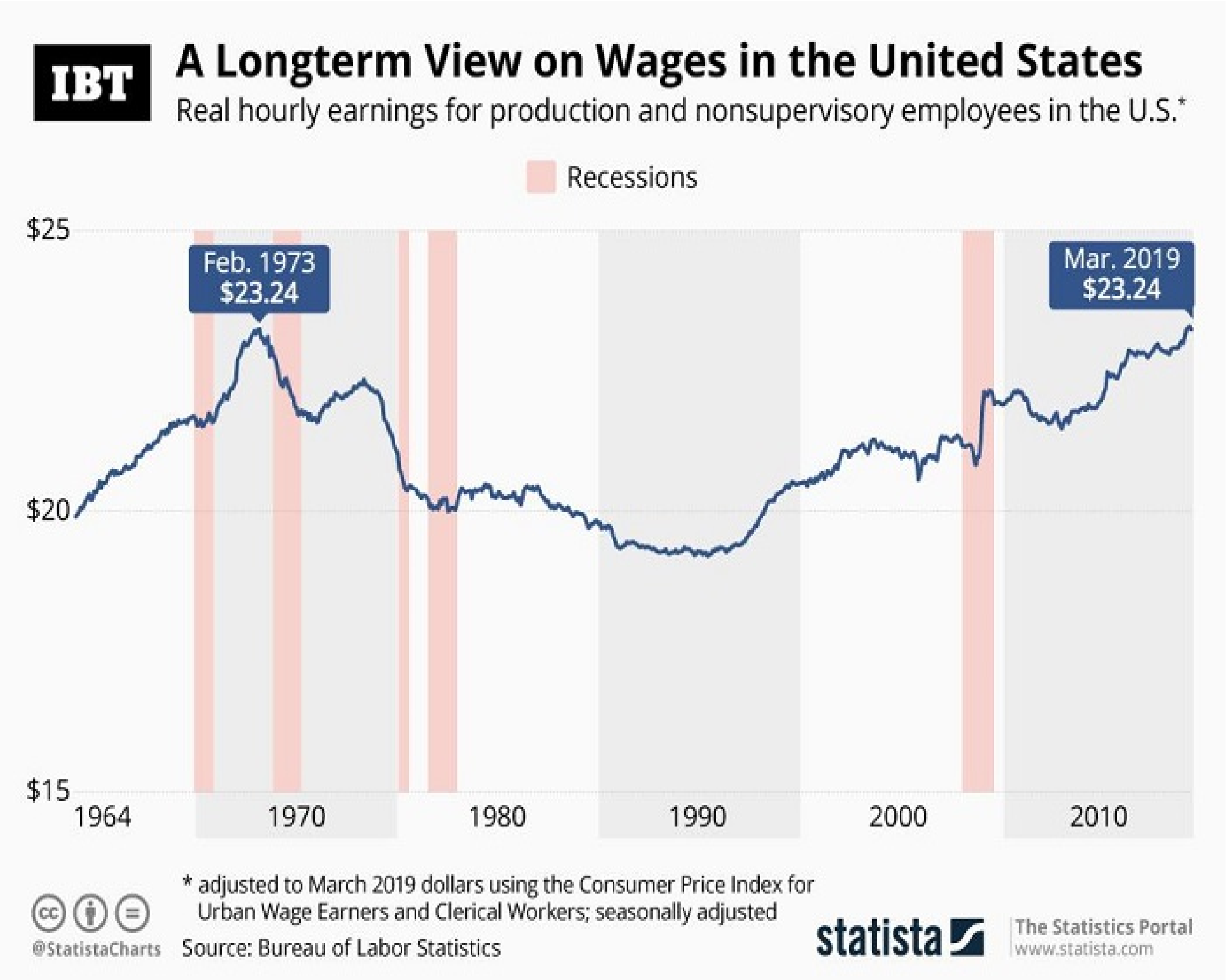
Fifth, they did a study, and the lower productivity growth in an industry, the higher the rise in costs, especially if they use college-educated workers who could otherwise get jobs in higher-productivity industries. This is what the Baumol effect would predict (though framed that way, it also sounds kind of obvious).

I find their case pretty convincing. And I want to believe. If this is true, it's the best thing I've heard all year. It restores my faith in humanity. Rising costs in every sector don't necessarily mean our society is getting less efficient, or more vulnerable to rent-seeking, or less-well-governed, or greedier, or anything like that. It's just a natural consequence of high economic growth. We can stop worrying that our civilization is in terminal decline, and just work on the practical issue of how to get costs down.

But I do have some gripes. T&H frequently compare apples and oranges; for example, the administrator share in colleges vs. the faculty share in K-12; it feels like they're clumsily trying to get one past you. They frequently describe how if you just use eg teacher salaries as a predictor, you can perfectly predict the extent of rising costs. But as far as I can tell, most things have risen the same amount, so if you used *any* subcomponent as a predictor, you could perfectly predict the extent of rising costs; again, it feels like they're clumsily trying to get something past me. I think I can work out what they were trying to do (stitch together different datasets to get a better picture, assume salaries rise equally in every category) but I still wish they had discussed their reasoning and its limitations more openly.

The main thesis survives these objections, but there are still a few things that bother me, or don’t quite fit. I want to bring them up not as a gotcha or refutation, but in the hopes that people who know more about economics than I do can explain why I shouldn’t worry about them.

First, real wages have not in fact gone up during most of this period. Factory workers are not getting paid more. That makes it hard for me to understand how rising wages for factory workers are forcing up salaries for violinists, teachers, and doctors.



A Longterm View on Wages in the United States

Real hourly earnings for production and nonsupervisory employees in the U.S.*

Feb. 1973 $23.24

Mar. 2019 $23.24

* adjusted to March 2019 dollars using the Consumer Price Index for Urban Wage Earners and Clerical Workers; seasonally adjusted

Source: Bureau of Labor Statistics

I discuss whether issues like benefits and inflation can explain this away here [here](), and conclude they can do so only partially; I’m not sure how this would interact with the Baumol effect.

Second, other data seem to dispute that salaries for the professionals in question have risen at all. T&H talk about rises in “instructional expenditures”, an education-statistics term that includes teacher salary and other costs; their source is NCES. But NCES also includes tables of actual teacher salaries. These show that teacher salaries today are only 6% higher than teacher salaries in 1970. Meanwhile, per-pupil costs are more than twice as high. How is an increase of 6% in teacher salaries driving an increase of 100%+ in costs? Likewise, although on page 33 T&H claim that doctors’ salaries have tripled since 1960, other sources report smaller increases of about 50% to almost nothing. Conventional wisdom among doctors is that the profession used to be more lucrative than it is today. This makes it hard to see how rising doctor salaries could explain a tripling in the cost of health care. And doctor salaries apparently make up only 20% of health spending, so it’s hard to see how they can matter that much.

(also, this SMBC)

Third, the Baumol effect can’t explain things getting less affordable. T&H write:

> The cost disease is not a disease but a blessing. To be sure, it would be better if productivity increased in all industries, but that is just to say that more is better. There is nothing negative about productivity growth, even if it is unbalanced.In particular, it is important to see that the increase in the relative price of the string quartet makes string quartets costlier but not less affordable. Society can afford just as many string quartets as in the past. Indeed, it can afford more because the increase in productivity in other sectors has

> made society richer. Individuals might not choose to buy more, but that is a choice, not a constraint forced upon them by circumstance.

This matches my understanding of the Baumol effect. But it doesn’t match my perception of how things are going in the real world. College has actually become less affordable. Using these numbers: in 1971, the average man would have had to work five months to earn a year’s tuition at a private college. In 2016, he would have had to work fourteen months. To put this in perspective, my uncle worked a summer job to pay for his college tuition; one summer of working = one year tuition at an Ivy League school. Student debt has increased 700% since 1990. College really does seem to be getting less affordable. So do health care, primary education, and all the other areas affected by cost disease. Baumol effects shouldn’t be able to do this, unless I am really confused about them.

If someone can answer these questions and remove my lingering doubts about the Baumol effect as an explanation for cost disease, they can share credit with Tabarrok and Helland for restoring a big part of my faith in modern civilization.

Highlights From The Comments On Cultural Evolution

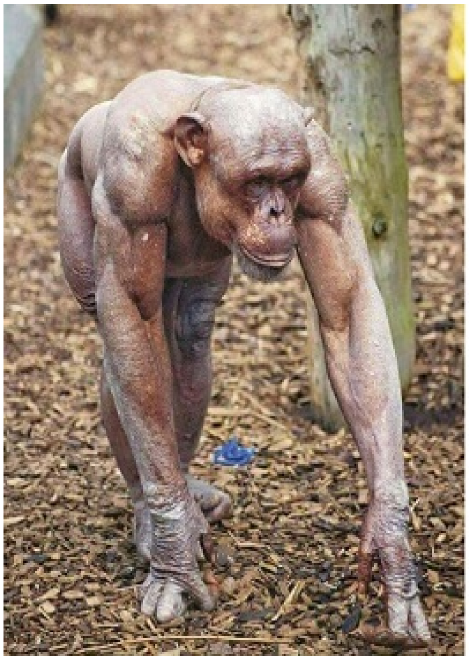
Peter Gerdes says:

> As the examples of the Nicaraguan deaf children left on their own to develop their own language demonstrates (as do other examples) we do create languages very very quickly in a social environment.
>
> Creating conlangs is hard not because creating language is fundamentally hard but because we are bad at top down modelling of processes that are the result of a bunch of tiny modifications over time. The distinctive features of language require both that it be used frequently for practical purposes (this makes sure that the language has efficient shortcuts, jettisons clunky overengineered rules etc..) and that it be buffeted by the whims of many individuals with varying interests and focuses.

This is a good point, though it kind of equivocates on the meaning of “hard” (if we can’t consciously do something, does that make it “hard” even if in some situations it would happen naturally?).

I don’t know how much of this to credit to a “language instinct” that puts all the difficulty of language “under the hood”, vs. inventing language not really being that hard once you have general-purpose reasoning. I’m sure real linguists have an answer to this. See also Tracy Canfield’s comments (1, 2) on the specifics of sign languages and creoles.

The Secret Of Our Success described how human culture, especially tool-making ability, allowed us to lose some adaptations we no longer needed. One of those was strength; we are much weaker than the other great apes. Hackworth provides an intuitive demonstration of this: hairless chimpanzees are *buff*:



Reasoner defines “Chesterton’s meta-fence” as:

> in our current system (democratic market economies with large governments) the common practice of taking down Chesterton fences is a process which seems well established and has a decent track record, and should not be unduly interfered with (unless you fully understand it)

And citizencokane adds:

> Indeed: if there is a takeaway from Scott’s post, it is that one way to ensure survival is high-fidelity adherence to traditions + ensuring that the inherited ancestral environment/context is more or less maintained. Adhering to ancient traditions when the context is rapidly

changing is a recipe for disaster. No point in mastering seal-hunting if there ain't no more seals. No point in mastering the manners of being a courtier if there ain't no more royal court. Etc.

And the problem is that, in the modern world, we can't simply all mutually agree to stop changing our context so that our traditions will continue to function as before because it is no longer under our control. I'm not just talking about climate change; I'm talking even moreso about the power of capital, an incentive structure that escapes all conscious human manipulation or control, and which more and more takes the appearance of an exogenous force, remaking the world "in its own image," turning "all that is solid into air," and compelling all societies, upon pain of extinction, to keep up with its rapid changes in context. This is why every true traditionalist must be, at heart, an anti-capitalist…if they truly understand capitalism.

Which societies had more success in the 18th and 19th centuries in the context of this new force, capital? Those who held rigidly to traditions (like Qing China), or those who tolerated or even encouraged experimentation? Enlightenment ideas would not have been nearly so persuasive if they hadn't had the prestige of giving countries like the Netherlands, England, France, and America an edge. Even countries that were not on the leading edge of the Enlightenment, and who only grudgingly and half-heartedly compromised with it like Germany, Austria, and (to some extent) Japan, did better than those who held onto traditions even longer, like the Ottoman Empire or Russia, or China.

In particular, you can't fault Russia or China for being even more experimental in the 20th century (Marxism,

> communism, etc.) if you realize that this was an understandable reaction to being visibly not experimental enough in the 19th century.

And Furslid continues:

> I think an important piece of this, which I hope Scott will get to in later points is to be less confident in our new culture. It makes sense to doubt if our old culture applies. However, it is also incredibly unlikely that we have an optimized new culture yet.
>
> We should be less confident that our new culture is right for new situations than that the old culture was right for old situations. This means we should be more accepting of people tweaking the new culture. We should also enforce it less strongly.

Quixote describes a transitional step in the evolution of manioc/cassava cultivation:

> Also, based on a recent conversation (unrelated to this post actually) that I had with one of my coworkers from central east Africa, I'm not sure that he would agree with the book's characterization of African adaptation to Cassava. He would probably point out that
>
> – Everyone in [African country] knows cassava can make you sick, that's why you don't plant it anywhere that children or the goats will eat it.
>
> – In general you want it plant cassava in swampy areas that you were going to fence off anyway.
>
> – You mostly let the cassava do its thing and only harvest it to use as your main food during times of famine /drought when your better crops aren't producing

> It seems like those cultural adaptations problem cover most / much of the problem with cassava.

---

ahasvers:

> There is a very nice experimental demonstration in this article (just saw the work presented at a workshop), where they get people to come as successive "generations" and improve on a simple physical system.
>
> Causal understanding is not necessary for the improvement of culturally evolving technology
>
> The design does improve over generations, no thanks to anyone's intelligence. They get both physics/engineering students and other students, with no difference at all. In one variant, they allow people to leave a small message to the next generation to transmit their theory on what works/doesn't, and that doesn't help, or makes things worse (by limiting the dimensions along which next generations will explore).

A few people including snmlp question the claim that aboriginal Tasmanians lost fire. See this and this paper for the status of the archaeological evidence.

Decius Brutus:

> Five hundred years hence, is someone going to analyze the college education system and point out that the wasted effort and time that we all can see produced some benefit akin to preventing chronic cyanide poisoning? Are they going to be able to do the same with other complex wasteful rituals, like primary elections and medical billing? Or do humans create lots of random

> wasteful rituals and occasionally hit upon one that removes poison from food, and then every group that doesn’t follow the one that removes poison from food dies while the harmless ones that just paint doors with blood continue?

I actually seriously worry about the college one. Like, say what you want about our college system, but it has some surprising advantages: somehow billions of dollars go to basic scientific research (not all of them from the government), it’s relatively hard for even the most powerful special interests to completely hijack a scientific field (eg there’s no easy way for Exxon to take over climatology), and some scientists can consistently resist social pressure (for example, all the scientists who keep showing things like that genetics matters, or IQ tests work, or stereotype threat research doesn’t replicate). While obviously there’s still a lot of social desirability bias, it’s amazing that researchers can stand up to it at all. I don’t know how much of this depends on the academic status ladder being so perverse and illegible that nobody can really hack it, or whether that would survive apparently-reasonable college reform.

Likewise, a lot of doctors just have no incentives. They don’t have an incentive to overtreat you, or to undertreat you, or to see you more often than you need to be seen, or to see you less often than you need to be seen (this isn’t denying some doctors in some parts of the health system do have these pressures). I actually don’t know whether my clinic would make more or less money if I fudged things to see my patients more often, and nobody has bothered to tell me. This is really impressive. Exposing the health system to market pressures would solve a lot of inefficiencies, but I don’t know if it would make medical care too vulnerable to doctors’ self-interest and destroy some necessary doctor-patient trust.

Lasagna:

> I’ve got two young kids of my own. One puts everything in his mouth, the other less so, and neither evinced anything resembling what I’m reading in Section III. We spent this past Sunday trying to teach my youngest not to eat the lawn, and my oldest liked to shove ant hills and ants into his mouth around that age. Yeah, sure, anecdotal, but a “natural aversion among infants to eating plants until they see mommy eating them, and after that they can and do identify that particular plan themselves and will eat it” seems like a remarkable ability that SOMEONE would have noticed before this study. I’ve never heard anyone mention it.
>
> I don’t think I’m weakmanning the book, it’s just that this is the only aspect discussed in Scott’s review that I have direct experience with, and my direct experience conflicts with the author’s conclusions. It’s a Gell-Mann amnesia thing, and makes me suspicious of the otherwise exciting ideas here. Like: does anyone here have any direct knowledge of manioc harvesting and processing, or the Tukanoans culture? How accurate is the book?

I checked with the mother of the local two-year old; she says he also put random plants in his mouth from a young age. Suspicious!

John Schilling:

> I think this one greatly overstates its thesis. Inventiveness without the ability to transmit inventions to future generations is of small value; you can’t invent the full set of survival techniques necessary for e.g. the high arctic in a single generation of extreme cleverness.

At best you can make yourself a slightly more effective ape. But cultural transmission of inventions without the ability to invent is of exactly zero value. It takes both. And since being a slightly more effective ape is still better than being an ordinary ape, culture is slightly less than 50% of the secret of our success.

That said, the useful insight is that the knowledge we need to thrive, is vastly greater than the knowledge we can reasonably deduce from first principles and observation. And what is really critical, this holds true even if you are in a library. You need to accept “X is true because a trusted authority told me so; now I need to go on and learn Y and Z and I don’t have time to understand why X is true”. You need to accept that this is just as true of the authority who told you X, and so he may not be able to tell you why X is true even if you do decide to ask him in your spare time. There may be an authority who could track that down, but it’s probably more trouble than it’s worth to track him down. Mostly, you’re going to use the traditions of your culture as a guide and just believe X because a trusted authority told you to, and that’s the right thing to do,

“Rationality” doesn’t work as an antonym to “Tradition”, because rationality needs tradition as an input. Not bothering to measure Avogadro’s number because it’s right there in your CRC handbook wikipedia is every bit as much a tradition as not boning your sister because the Tribal Elders say so; we just don’t call it that when it’s a tradition we like. Proper rationality requires being cold-bloodedly rational about evaluating the high-but-not-perfect reliability of tradition as a source of fact.

Unfortunately, and I think this may be a relic of the 18th and early 19th century when some really smart

> polymathic scientists could almost imagine that they really could wrap their minds around all relevant knowledge from first principles on down, our culture teaches ‘Science!’ in a way that suggests that you really should understand how everything is derived from first principles and firsthand observation or experiment even if at the object level you’re just going to look up Avogadro’s number in Wikipedia and memorize it for the test.

---

nkurz isn’t buying it:

> I’m not sure where Scott is going with this series, but I seem to have a different reaction to the excerpts from Henrich than most (but not all) of the commenters before me: rather than coming across as persuasive, I wouldn’t trust him as far as I could throw him.
>
> For simplicity let’s concentrate on the seal hunting description. I don’t know enough about Inuit techniques to critique the details, but instead of aiming for a fair description, it’s clear that Henrich’s goal is to make the process sound as difficult to achieve as possible. But this is just slight of hand: the goal of the stranded explorer isn’t to reproduce the exact technique of the Inuit, but to kill seals and eat them. The explorer isn’t going to use caribou antler probes or polar bear harpoon tips — they are going to use some modern wood or metal that they stripped from their ice bound ship.
>
> Then we hit “Now you have a seal, but you have to cook it.” What? The Inuit didn’t cook their seal meat using a soapstone lamp fueled with whale oil, they ate it raw! At this point, Henrich is not just being misleading, he’s making it up as he goes along. At this point I start to wonder if part about the antler probe and bone harpoon

> head are equally fictional. I might be wrong, but beyond this my instinct is to doubt everything that Henrich argues for, even if (especially if) it's not an area where I have familiarity
>
> Going back to the previous post on "Epistemic Learned Helplessness", I'm surprised that many people seem to have the instinct to continue to trust the parts of a story that they cannot confirm even after they discover that some parts are false. I'm at the opposite extreme. As soon as I can confirm a flaw, I have trouble trusting anything else the author has to say. I don't care about the baby, this bathwater has to go! And if the "flaw" is that the author is being intentionally misleading, I'm unlikely to ever again trust them (or anyone else who recommends them). .

Probably I accidentally misrepresented a lot in the parts that were my own summary. But this is from a direct quote, and so not my fault.

roystgnr adds:

> Wikipedia seems to suggest that they ate freshly killed meat raw, but cooked some of the meat brought back to camp using a Kudlik, a soapstone lamp fueled with seal oil or whale blubber. Is that not correct? That would still flatly contradict "but you have to cook it", but it's close enough that the mistake doesn't reach "making it up as he goes along" levels of falsehood. You're correct that even the true bits seem to be used for argument in a misleading fashion, though.

This seems within the level of simplifying-to-make-a-point that I have sometimes been guilty of myself, so I'll let it pass.

Bram Cohen:

> A funny point about the random number generators: Rituals which require more effort are more likely to produce truly random results, because a ritual which required less effort would be more tempting to re-do if you didn't like the result.

Followed by David Friedman:

> This reminds me of my father's argument that cheap computers resulted in less reliable statistical results. If running one multiple regression takes hundreds of man hours and thousands of dollars, running a hundred of them and picking the one that, by chance, gives you the significant result you are looking for, isn't a practical option.

Yikes.

Anatoly:

> The quote on quadruped running seems inaccurate in several important ways compared to the primary references Henrich cites, which are short and very interesting in their own: Bramble and Carrier (1983) and Carrier (1984). In particular, humans still typically lock their breathing rate with their strides, it's just that animals nearly always lock them 1:1, while humans are able to switch to other ratios, like 1:3, 2:3, 1:4 etc. and this is thought to allow us to maintain efficiency at varying speeds. Henrich also doesn't mention that humans are at the outset metabolically disadvantaged for running in that we spend twice as much energy (!) per unit mass to run the same distance as quadrupeds. That we are still able to run down prey by endurance

> running is called the “energetic paradox” by Carrier. Liebenberg (2006) provides a vivid description of what endurance hunting looks like, in Kalahari.

And b_jonas:

> I doubt the claim that humans don’t have quantized speeds of running. I for one definitely have two different gaits of walking, and find walking in an intermediate speed between the two more difficult than either of them. This is the most noticable if I want to chat with someone while walking, because then I have to walk in such an intermediate speed to not get too far from them. The effect is somewhat less pronounced now that I’ve gained weight, but it’s still present. I’m not good at running, so I can’t say anything certain about it, but I suspect that at least some humans have different running gaits, even if the cause is not the specific one that Joseph Henrich mentions about quadrupeds.

I’ve never noticed this. And I used to use treadmills relatively regularly, and play with the speed dial, so I feel like I would have noticed if this had been true. Anyone have thoughts on this?

Squirrel Of Doom:

> I read somewhere that the languages with the most distinctive sounds are in Africa, among them the ones including the !click! ones. Since humanity originates from Africa, these are also the oldest language families.
>
> As you move away from Africa, you can trace how languages lose sound after sound, until you get to Hawaiian, which is the language with the fewest sounds, almost all vowels.

> I’ve half heartedly tried to find any mention of this, perhaps overly cute theory again, but failed. The “sonority” theory here reminded me. Anyone know anything, one way or the other?

Secret Of Our Success actually mentions this theory; you can find the details within.

Some people reasonably bring up that no language can be older than any other, for the same reason it doesn’t make sense to call any (currently existing) evolved animal language older than any other – every animal lineage from 100 million BC has experienced 100 million years of evolution.

I think I’ve heard some people try to get around this by focusing on *schisms*. Everyone starts out in Africa, but a small group of people move off to Sinai or somewhere like that. Because most of the people are back home in Africa, they can maintain their linguistic complexity; because the Sinaites only have a single small band talking to each other, they lose some linguistic complexity. This seems kind of forced, and some people in the comments say linguistic complexity actually works the opposite direction from this, but I too find the richness of Bushman languages pretty suggestive.

What about rules that really do seem pointless? Catherio writes:

> My basic understanding is that if some of the rules (like “don’t wear hats in church”) are totally inconsequential to break, these provide more opportunities to signal that your community punishes rule violation, without an increase in actually-costly rule violations.

I'd heard this before, but she manages (impressively), to link it to AI: see Legible Normativity for AI Alignment: The Value of Silly Rules.

---

liskantope:

> With regard to accepting other people's illegible preferences…I wish I could show this essay to, like, two-thirds of all the people I've ever lived with. Seriously, a common core of my issues with roommates has been that they refuse to accept or understand my illegible preferences (I often refer to these as "irrational aversions") while refusing to admit that their own illegible preferences are just as difficult to ground rationally. Just establishing an understanding that illegible preferences should be respected by default or at least treated on an even playing field, and that having immediate objective logical explanations for preferences should not be a requirement for validation, would have immediately improved my relationships with people I've lived with 100%.

I've had the same experience – a good test for my compatibility with someone will be whether they'll accept "for illegible reasons" as an excuse. Despite the stereotypes, rationalists have been a hundred times better at this than any other group I've been in close contact with.

Nav on Lacan and Zizek (is everything cursed to end in Zizek eventually, sort of like with entropy?):

> Time to beat my dead horse; the topics you're discussing here have a lot of deep parallels in the psychoanalytic literature. First, Scott writes:

}} “If you force people to legibly interpret everything they do, or else stop doing it under threat of being called lazy or evil, you make their life harder”

This idea is treated by Lacan as the central ethical problem of psychoanalysis: under what circumstances is it acceptable to cast conscious light upon a person’s unconsciously-motivated behavior? The answer is usually “only if they seek it out, and only then if it would help them reduce their level of suffering”.

Turn the psychoanalytic, phenomenology-oriented frame onto social issues, as you’ve partly done, and suddenly we’re in Zizek-land (his main thrust is connecting social critique with psychoanalytic concepts). The problem is that (a) Zizek is jargon-heavy and difficult to understand, and (b) I’m not nearly as familiar with Zizek’s work as with more traditional psychoanalytic concepts. But I’ll try anyway. From a quick encyclopedia skim, he actually uses a similar analogy with fetishes (all quotes from IEP):

}} “Žižek argues that the attitude of subjects towards authority revealed by today’s ideological cynicism resembles the fetishist’s attitude towards his fetish. The fetishist’s attitude towards his fetish has the peculiar form of a disavowal: “I know well that (for example) the shoe is only a shoe, but nevertheless, I still need my partner to wear the shoe in order to enjoy.” According to Žižek, the attitude of political subjects towards political authority evinces the same logical form: “I know well that (for example) Bob Hawke / Bill Clinton / the Party / the market does not always act justly, but I still act as though I did not know that this is the case.””

As for how beliefs manifest, Zizek clarifies the experience of following a tradition and why we might actually feel like these traditions are aligned with

“Reason” from the inside, and also the crux of why “Reason” can fail so hard in terms of social change:

According to Žižek, all successful political ideologies necessarily refer to and turn around sublime objects posited by political ideologies. These sublime objects are what political subjects take it that their regime’s ideologies’ central words mean or name extraordinary Things like God, the Fuhrer, the King, in whose name they will (if necessary) transgress ordinary moral laws and lay down their lives… Kant’s subject resignifies its failure to grasp the sublime object as indirect testimony to a wholly “supersensible” faculty within herself (Reason), so Žižek argues that the inability of subjects to explain the nature of what they believe in politically does not indicate any disloyalty or abnormality. Žižek argues that the inability of subjects to explain the nature of what they believe in politically does not indicate any disloyalty or abnormality. What political ideologies do, precisely, is provide subjects with a way of seeing the world according to which such an inability can appear as testimony to how Transcendent or Great their Nation, God, Freedom, and so forth is—surely far above the ordinary or profane things of the world.

Lastly and somewhat related, going back to an older SSC post, Scott argues that he doesn’t know why his patients react well to him, but Zizek can explain that, and it has a lot of relevance for politics (transference is a complex topic, but the simple definition is a transfer of affect or mind from the therapist to the patient, which is often a desirable outcome of therapy, contrasted with counter-transference, in which the patient affects the therapist):

}} “The belief or “supposition” of the analysand in psychoanalysis is that the Other (his analyst) knows the

> meaning of his symptoms. This is obviously a false belief, at the start of the analytic process. But it is only through holding this false belief about the analyst that the work of analysis can proceed, and the transferential belief can become true (when the analyst does become able to interpret the symptoms). Žižek argues that this strange intersubjective or dialectical logic of belief in clinical psychoanalysis also what characterizes peoples’ political beliefs…. the key political function of holders of public office is to occupy the place of what he calls, after Lacan, “the Other supposed to know.” Žižek cites the example of priests reciting mass in Latin before an uncomprehending laity, who believe that the priests know the meaning of the words, and for whom this is sufficient to keep the faith. Far from presenting an exception to the way political authority works, for Žižek this scenario reveals the universal rule of how political consensus is formed.”
>
> Scott probably come across as having a stable and highly knowledgeable affect, which gives his patients a sense of being in the presence of authority (as we likely also feel in these comment threads), which makes him better able to perform transference and thus help his patients (or readers) reshape their beliefs.
>
> Hopefully this shallow dive was interesting and opens up new areas of potential study, and also a parallel frame: working from the top-down ethnography (as tends to be popular in this community; the Archimedean standpoint) gives us a broad understanding, but working from the bottom-up gives us a more personal and intimate sense of why the top-down view is correct.

This helped me understand Zizek and Lacan a lot better than reading a book on them did, so thanks for that.

Stucchio doesn't like me dissing Dubai:

> I'm just going to raise a discussion of one piece here:
>
> }} "Dubai, whose position in the United Arab Emirates makes it a lot closer to this model than most places, seems to invest a lot in its citizens' happiness, but also has an underclass of near-slave laborers without exit rights (their employers tend to seize their passports)."
>
> I have probably read the same western articles Scott has about all the labor the UAE and other middle eastern countries imports. But unlike them, I live in India (one of the major sources of labor) and mostly have heard about this from people who choose to make the trip.
>
> To me the biggest thing missing from these western reporter's accounts is the fact that the people shifting to the gulf are ordinary humans, smarter than most journalists, and fully capable of making their own choices.
>
> Here are things I've heard about it, roughly paraphrased:
>
> "I knew they'd take my passport for 9 months while I paid for the trip over. After that I stuck around for 3 years because the money was good, particularly after I shifted jobs. It was sad only seeing my family over skype, but I brought home so much money it was worth it."
>
> "I took my family over and we stayed for 5 years; the money was good, we all finished the Hajj while we were there, but it was boring and I missed Maharashtrian food."
>
> "It sucked because the women are all locked up. You can't talk to them at the mall. It's as boring as everyone

> says and you can’t even watch internet porn. But the money is good.”
>
> When I hear about this first hand, the stories don’t sound remotely like slave labor. It doesn’t even sound like “we were stuck in the GDR/Romania/etc” stories I’ve heard from professors born on the wrong side of the Iron Curtain. I hear stories of people making life choices to be bored and far from family in return for good money. Islam is a major secondary theme. So I don’t think the UAE is necessarily the exception Scott thinks it is.

Moridinamael on the StarCraft perspective:

> In StarCraft 2, wild, unsound strategies may defeat poor opponents, but will be crushed by decent players who simply hew to strategies that fall within a valley of optimality. If there is a true optimal strategy, we don’t know what it is, but we do know what good, solid play looks like, and what it doesn’t look like. Tradition, that is to say, iterative competition, has carved a groove into the universe of playstyles, and it is almost impossible to outperform tradition.
>
> Then you watch the highest-end professional players and see them sometimes doing absolutely hare-brained things that would only be contemplated by the rank novice, and you see those hare-brained things winning games. The best players are so good that they can leave behind the dogma of tradition. They simply understand the game in a way that you don’t. Sometimes a single innovative tactic debuted in a professional game will completely shift how the game is played for months, essentially carving a new path into what is considered the valley of optimality. Players can discover paths that are just better than tradition. And then, sometimes,

> somebody else figures out that the innovative strategy has an easily exploited Achilles' heel, and the new tactic goes extinct as quickly as it became mainstream.
>
> StarCraft 2 is fun to think about in this context because it is relatively open-ended, closer to reality than to chess. There are no equivalents to disruptor drops or mass infestor pushes or planetary fortress rushes in chess. StarCraft 2 is also fun to think about because we've now seen that machine learning can beat us at it by doing things outside of what we would call the valley of optimality.
>
> But in this context it's crucial to point out that the way AlphaStar developed its strategy looked more like gradually accrued "tradition" than like "rationalism". A population of different agents played each other for a hundred subjective years. The winners replicated. This is memetic evolution through the Chestertonian tradition concept. The technique wouldn't have worked without the powerful new learning algorithms, but the learning algorithm didn't come up with the strategy of mass-producing probes and building mass blink-stalkers purely out of its fevered imagination. Rather, the learning algorithms were smart enough to notice what was working and what wasn't, and to have some proximal conception as to why.

Someone (maybe Robin Hanson) treats all of history as just evolution evolving better evolutions. The worst evolution of all (random chance) created the first replicator and kicked off biological evolution. Biological evolution created brains, which use a sort of hill-climbing memetic evolution for good ideas. People with brains created cultures (cultural evolution) including free market economies (an evolutionary system that selects for successful technologies). AIs like AlphaStar are the next (final?) step in this process.

Followup On The Baumol Effect: Thanks, O Baumol

Last week I reviewed Alex Tabarrok and Eric Helland's *Why Are The Prices So D*mn High?*. On Marginal Revolution, Tabarrok wrote:

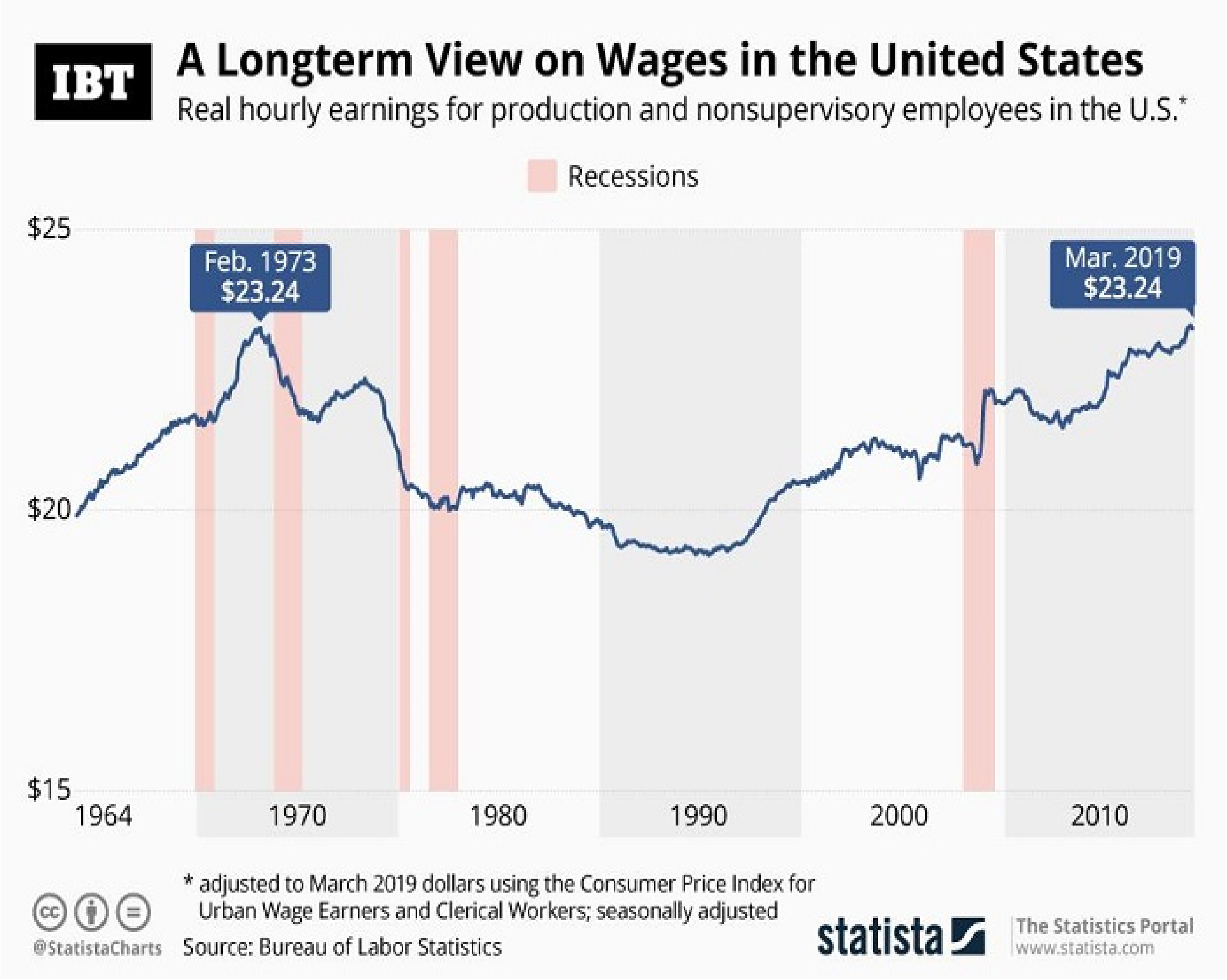
> SSC does have some lingering doubts and points to certain areas where the data isn't clear and where we could have been clearer. I think this is inevitable. A lot has happened in the post World War II era. In dealing with very long run trends so much else is going on that answers will never be conclusive. It's hard to see the signal in the noise. I think of the Baumol effect as something analogous to global warming. The tides come and go but the sea level is slowly rising

I was pretty disappointed by this comment. T&H's book blames cost disease on rising wages in high-productivity sectors, and consequently in education and medicine. My counter is that wages in high productivity sectors, education, and medicine are not actually rising. This doesn't seem like an "area where you could have been clearer". This seems like an existential challenge to your theory! Come on!

Since we're not getting an iota of help from the authors, we're going to have to figure this out ourselves. The points below are based on some comments from the original post and some conversations I had with people afterwards.

1. Median wages, including wages in high-productivty sectors like manufacturing, are not rising

I originally used this chart to demonstrate:



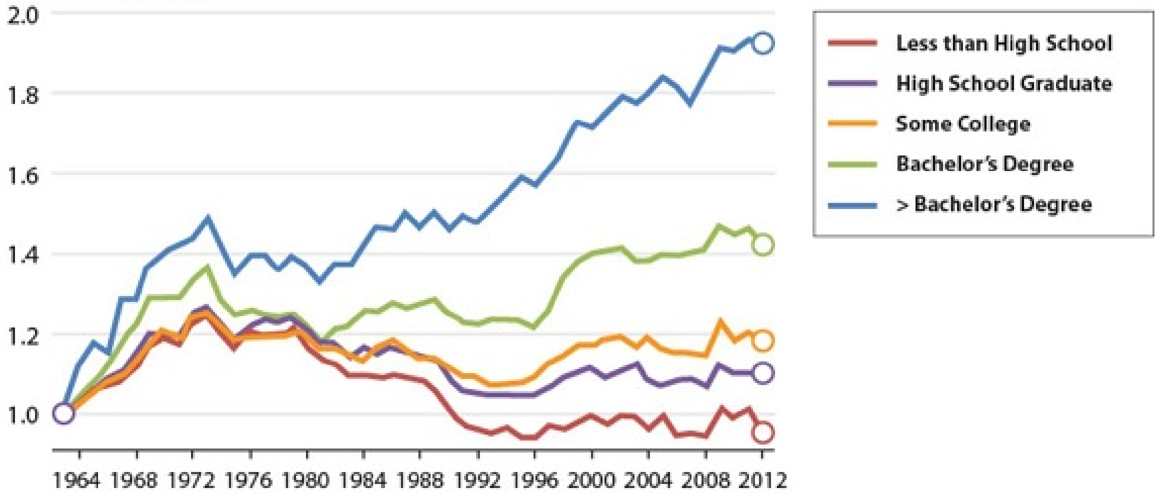
Some people protested this was was a misleading portrayal, and that there are structural factors that disguise rising wages. I’ve written about this before in Wage Stagnation: Much More Than You Wanted To Know. The short answer is – no, it’s not about increasing benefits, those only explain about 10% of the wage-productivity difference.

It *is* partially about how you calculate inflation, which explains around 35% of the problem depending on who you believe. But we’re comparing wages to the cost of education/medicine. As long as you’re using the same deflator for both of them, you’re fine. As far as I know, T&H and everyone who talks about rising education/medical costs has been using the normal consumer deflator. So if you want to argue wages are underestimated, you also have to argue education/medical costs have gone up even more than people think. This doesn't help at all!

(or see these numbers, which show that nominal college tuition has gone up as a percent of nominal median wage, and so should be immune to inflation shenanigans)

Other people protested against looking at the median wage, arguing that the wages of college graduates are more relevant. After all, teachers, professors, doctors, and nurses are all college grads. If their opportunity cost goes up, that could still drive a Baumol effect. And:

Figure 1: Changes in Real Wage Levels of Full-Time U.S. Male Workers by Education, 1963 – 2012



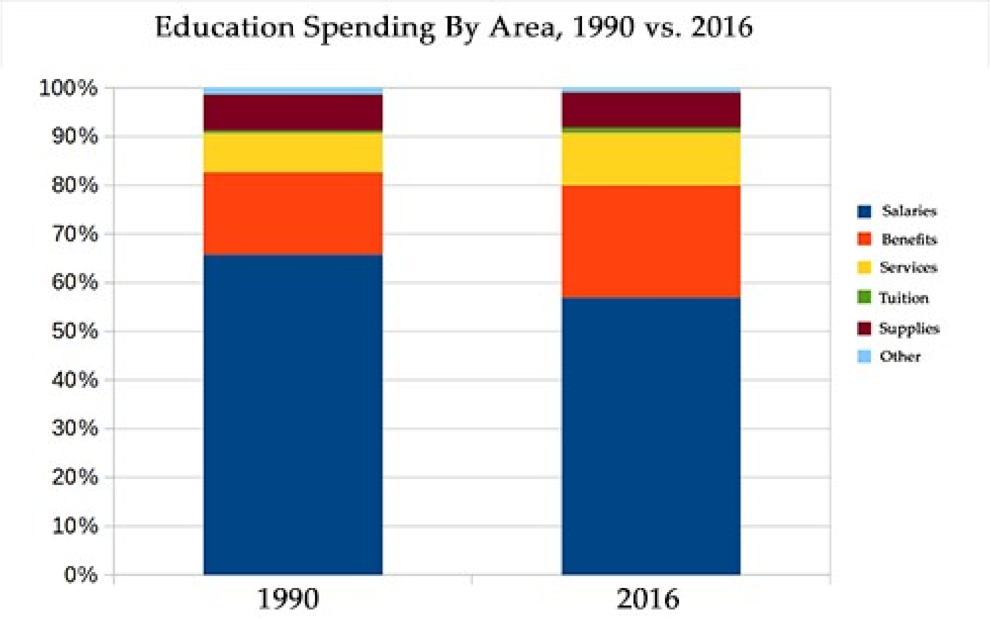
Source: David H. Autor, *Science* 2014; 344: 843-851.

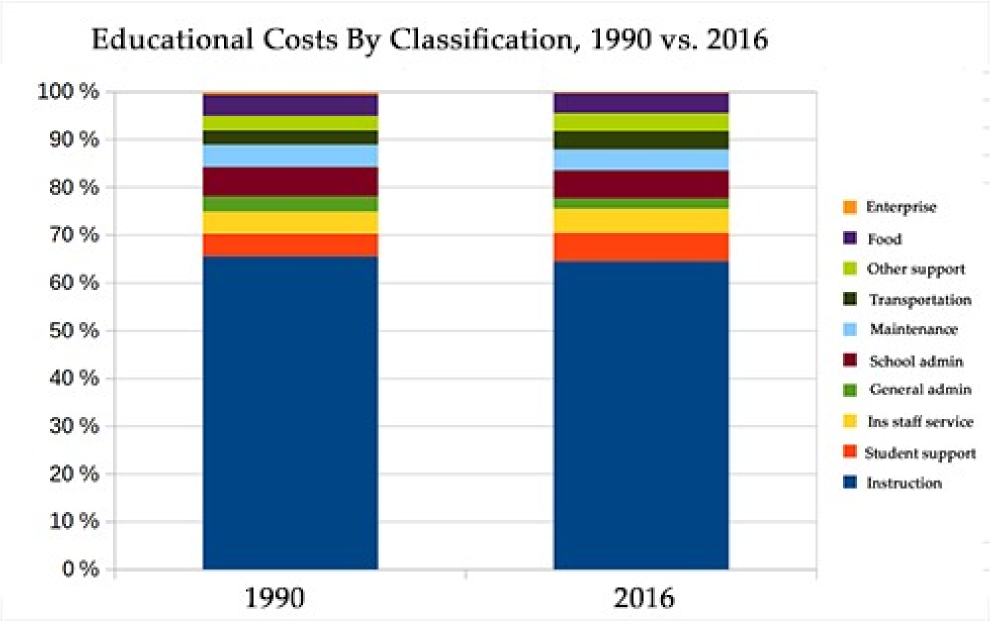
Surely doctors and professors are in that top blue line; I *think* nurses and teachers are in the lower green one. Plausibly these professions' opportunity costs have gone up 50 – 100% during this period. This is a *start* to explaining why education/medicine have gone up 200 – 300% during the same time. On the other hand, the period of fastest wage growth was 1965 – 1975, which as per T&H's graph (page 2) was the period of slowest cost growth.

2. Wages for doctors and teachers have not risen

Let's start with teachers. T&H use NCES' "instructional expenditures" category to show teacher wages have tripled since the 1950s; I cited NCES' actual teacher salary data to show it's stayed about the same. What's going on?

NCES only has good data after 1990. Their data says education has become about 30% more expensive in real terms since that time – T&H's data (page 2) suggests it has become twice as expensive, but their data on page 5 agrees with NCES (huh?). Here's how various fields have changed, using two different classification systems:





Pay attention especially to the first one. From 1990 to 2016, employee salaries as a percent of educational expenditures have gone *down*! Employee benefits have gone up a bit, but not enough: salary + benefits is still a smaller part of the education budget in 2016 than in 1990. How do you look at these data and say “We’ve figured out why education costs are rising and it’s definitely salaries”?

I think I miscalculated my tone when criticizing T&H’s presentation of data in their book. Tabarrok says I complained about areas “where [he] could have been clearer”. But my actual concern was that the presentation of this section misleads the reader. Whenever T&H talk about something other than salary, they emphasize that its share of the pie has not increased, but don’t mention that it increased a lot in absolute terms. Then when talking about salary, they emphasize that it increased a lot in absolute terms, but don’t mention that its share of the pie hasn’t increased. You’re left with the impression that salaries are the culprit for the price increases, when in fact salaries increased least of all the major categories in the data. “The data could have been clearer" is never just a minor gripe! Unclear data means you can prove whatever you want!

How are salary costs per pupil rising (even in proportion to other costs) if salaries are not? My guess is it's all about decreasing class sizes, which T&H also highlight.

Moving on to doctors, I don't have any equally clear sources. But I'll at least try to explain more about the ones I already have.

This paper from *Health Economics, Policy, and Law* shows on page 15:

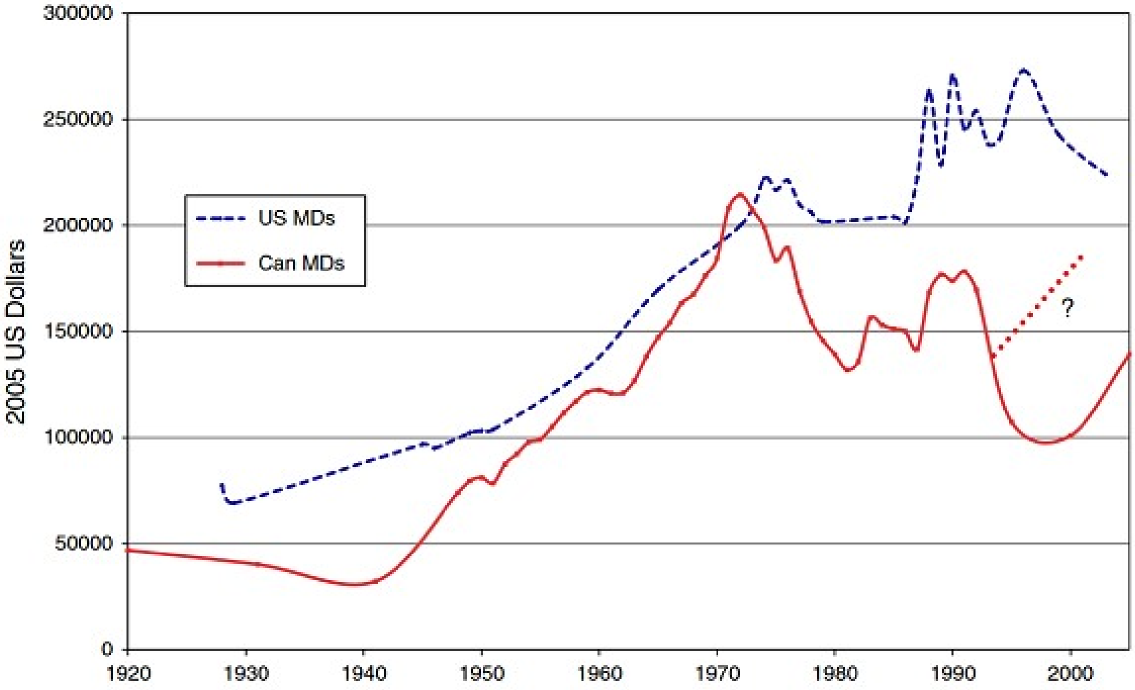
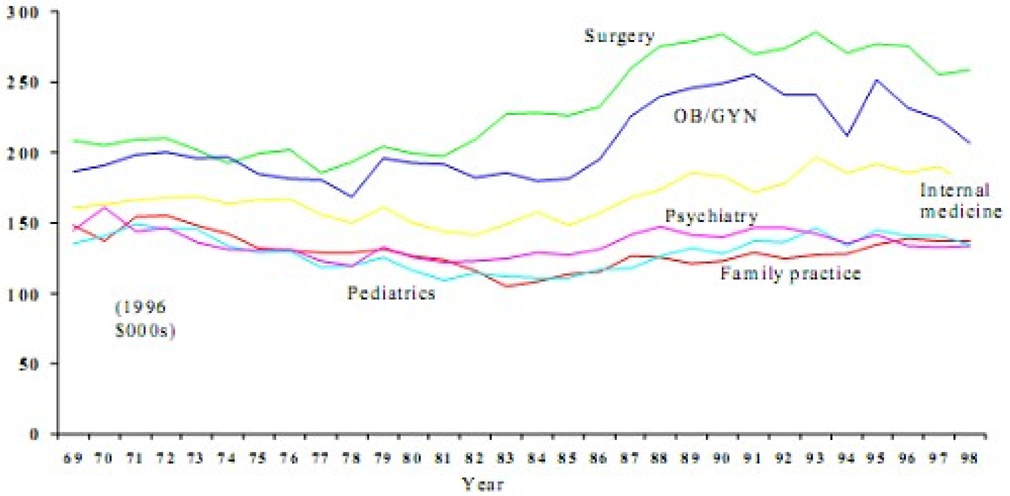


Figure 2. Physician income in Canada and the United States, 1920–2005.
Source: For Canadian physician income, see *Census of Canada* (1931–2006) and *Taxation Statistics* (1948–1995) (references 24 and 28); US physician income (see references 48–54).
Note: ? = Extrapolation of Canadian physician income based on *Taxation Statistics*.

They cite their source as "references 48 – 54", but their reference section isn't numbered, and is in alphabetical order, which would make it a pretty big coincidence if reference 48 through 54 were all about the same thing. I think a wire got crossed somewhere. But taking it at face value, I eyeball US doctor salaries in 1960 as 130K and in 2005 as about 210K, for an increase of 60% (not the 200% T&H claim). Doctor salaries have been about stable since 1975, even as (according to T&H's graph on page 2), healthcare costs have about doubled.

The Last Psychiatrist posts this image:



I can't trace the source beyond him, but read his post, where he notes that "There is the reality that doctor salaries (with notable exceptions) have been fairly static since 1969, even as the cost of living, price of homes, college, etc have gone up. And medical school debt."

The last source I was able to find was this 1985 paper on doctor pay. It states that "In 1973, the median annual physician income was $45,000…by 1982, the median physician income was $85,000." According to the inflation calculator, those numbers are $260,000 and $225,000 in current dollars, respectively. Estimates for average doctor salary today range from $209,000 to $299,000. I'm surprised how hard this is to measure, but it doesn't seem to have doubled or tripled the way T&H claim (or the way it would have to in order to drive Baumol effects).

Also, the Baumol effect only works if the market sets your salary in the first place! Right now the supply of doctors is limited by licensing issues and bottlenecks in the medical education process; that keeps salaries high. Why would the Baumol effect drive that even higher? If doctors’ salaries didn’t increase in keeping with the highest-productivity industries, would medical schools sit empty? Given that the people setting salaries for doctors (hospitals, clinics) are not the people who determine the supply of doctors (bureaucrats, medical school deans), why should supply and salary be related in a way that obeys normal economic laws? I’m not really sure how to model this, but I’m pretty sure it doesn’t end with doctors leaving medicine to play violin concertos instead.

3. The Baumol effect cannot make things genuinely less affordable, but things are genuinely less affordable

I think I screwed up here.

The Baumol effect cannot make things genuinely less affordable for *society*, because society is more productive and can afford more stuff.

However, it can make things genuinely less affordable for individuals, if those individuals aren’t sharing in the increased productivity of society.

Suppose that in 1960, widgets cost $1, a worker could produce ten widgets an hour (and made $10 in wages), and violin concerts cost $10. Also, you are a farmer and make $10 per hour. You can listen to one violin concert per hour.

In 2010, widgets cost $0.50, a worker can make a hundred widgets an hour (and makes $50 in wages), and violin concerts have risen to $50. But you, still a farmer, still only make $10 per hour. The Baumol effect has driven up the cost of violin concerts for you.

This shouldn't happen in real life, because if you work in the high-productivity-gain industries, you should make more because of your increased productivity, and if you work in the low-productivity-gain industries, you should make more because of the Baumol effect. But if it did happen, you'd be screwed.

And in fact, as mentioned above, wages have not increased in keeping with productivity, either in the high-productivity-gain industries like manufacturing, or in the low-productivity-gain industries like teaching. So if you work in one of those industries, it's totally possible for you to be screwed, ie for college etc to become much less affordable for you.

If wages had grown in keeping with productivity, median yearly salary would be something like $100,000. Someone making $100,000 per year shouldn't have *that* hard a time affording health insurance and college tuition for their kids; this would be the Baumol effect working in its normal, rising-tide-lifts-all-boats way. Since this hasn't happened, Baumol-applicable industries have become harder to afford.

Again, this is all theoretical, because wages in high-productivity industries haven't risen, so it's hard to see how a Baumol effect could be happening at all. But if it was, it would be a good explanation for cost disease.

So I retract this third objection. I think the first objection *mostly* still stands, though it is a little weaker if we limit ourselves to college-educated workers compared to all workers. I think the second objection absolutely still stands, and it’s hard for me to see how T&H’s case could survive it.

If Kim Jong-Un Opened A KFC, Would You Eat There?

Philip Morris is pivoting to smoke-free cigarettes, because "society expects us to act responsibly, and we are doing just that by designing a smoke-free future". Also, KFC "promises not to let vegans down" with their new meatless chicken-like nuggets. They'll have to compete with factory-farming mega-conglomerate Tyson Foods, who are coming out with their own vegetarian chicken option.

Clearly this is progress. Tobacco-free cigarettes have helped a lot of people quit smoking; meat substitutes have helped a lot of people (recently sort of including me) become vegetarian. I want a smoke-free meatless future. But does it become a mockery when the same companies that provided the smoky meaty past are selling it to us? If they make a fortune being evil, resist change, and lose, should they get to make a second fortune being good? If Hitler, when the war turned against him, quit the Nazism industry and opened a matzah bakery, would you buy his matzah?

I think the answer is supposed to be yes. I've heard many smart people argue that we should offer evil dictators a comfortable and lavish retirement, free from any threat of justice. After all, if they take the offer, they'll go off and enjoy their retirement instead of continuing to dictate. But if they expect to be put on trial for war crimes the second they relinquish power, they'll hold on to power forever. If Hitler had been willing to give up and open a bakery when he lost Stalingrad in 1943, think how many lives would have been saved by letting him. And if Kim Jong-Un wants to give up and move to Tahiti, of course you say yes.

In the same way, if evil companies want to go good, you should let them. If they have a line of retreat, they won't fight so hard against change. If Tyson Foods wants to use its lobbyists to support meat substitutes instead of sabotaging them, that's good for everybody. If they want to use their research budget to push plant-based meats forward, so much the better.

The counterargument is that punishment is the only tool we have to make bad actors do good things. If dictators fear punishment, maybe they won't dictate to begin with. If companies know that moral progress will eventually leave the immoral companies bankrupt, maybe they'll try being moral before it's immediately profitable.

We're in a weird situation where before anything happens, we might want to precommit to "punish companies who do evil, no matter what". After companies have started doing evil, we might want to break our previous precommitment and switch to "let evil companies avoid punishment if they stop doing evil". And after companies have stopped doing evil, we might want (if only for the sake of our own sense of justice) to break both of our previous precommitments and go with "punish them after all".

What is the right action?

I'm not sure, but I lean toward "buy the meatless chicken from KFC", for a few reasons.

First, I'm skeptical that corporations can predict moral progress, and I expect them to have high discount rates. I don't think Colonel Sanders in 1952 was thinking "Maybe I shouldn't sell chicken, just in case later generations punish my successors." That removes a lot of the advantage of precommiting to always punish evil corporations, but keeps the advantage of rewarding evildoers who turn good.

Second, realistically there are probably many companies that are as bad as these (like oil companies), which we don’t think about because they’re not in the process of going good in ways that make their evil more ironic and salient. It would be dumb to boycott *only* the companies that are trying to improve.

Third, boycotting companies is hard. In the process of writing this article, I learned Tyson Foods until recently owned Sara Lee, the cookie company, which itself owns a bunch of popular coffee brands. Also, they seem to have invested in Beyond Meat and Memphis Meats and all the other vegan-meat-substitute companies that we would feel good about buying from if we boycotted Tyson. If Tyson Foods really wants to make money off of vegans, they can probably do it without those vegans noticing.

I’m curious what other people think, so here’s **a poll** you can take on this.

Are Sexual Purity Taboos A Response To STIs?

I.

Did cultural evolution create sexual purity taboos to prevent the spread of STIs? A few weeks ago, I wrote a post assuming this was obviously true; after getting some pushback, I want to look into it in more depth.

STIs were a bigger problem in the past than most people think. Things got especially bad after the rise of syphilis: British studies find an urban syphilis rate of 8-10% from the 1700s to the early 1900s. At the time the condition was incurable, and progressed to insanity and death in about a quarter of patients. If you've got a 10% local syphilis rate, you are going to want some *major* sexual purity taboos. It's less clear how bad STDs were in truly ancient times, but given how easily the extent of syphilis has slipped out of our cultural memory, I'm not ruling out "pretty bad".

Here are some things I think of as basic parts of sexual purity taboos. All of these are cross-cultural – which isn't to say they're in every culture, or that some cultures aren't exactly the opposite, just to say that they seem to pop up pretty often. Writing from the male perspective because most of the cultures I know about thought that way:

1. If your wife has sex with another man, you should be angry
2. Preferably you should marry a virgin. If you think your bride is a virgin, but she isn't, you should be angry
3. If you've got to marry a non-virgin, then marrying a widow is okay, but marrying a former prostitute or

somebody known for sleeping around a *lot* is beyond the pale.

All of these are plausible ways to prevent the spread of STIs. If your wife has sex with another man, she could catch his STI and give it to you. If your bride isn't a virgin, she might have STIs. If someone's a widow, they probably slept with one known person whose STI status can be guessed at; if they're a prostitute or slept around, they slept with many unknown people and have a higher chance of having STIs.

But the counterargument is that at least (1) and (2) are also good ways to prevent false paternity, ie raising another person's child as your own.

The main argument that it's more STI than paternity is that (3) doesn't seem paternity-related; if it's been more than nine months, you shouldn't care who they've slept with before. Also, the taboos usually explicitly reference ideas of "pure" vs. "gross"; in most other cases, these are disease-related taboos. For example, spoiled food is "gross", dirt/feces/blood are "gross", corpses are "gross" – all of these are related to risk of disease transmission.

The main argument that it's more paternity than STIs is that there's less concern around men who have slept around being impure and unmarriagable. But that could just be because men are making the taboos and rigging them in their own favor. Yet you'd still think that if 10% of the population had syphilis and cultural evolution worked, men would stick to the purity taboos out of self-interest. Not sure here.

One way to distinguish between these possibilities would be to see how taboos changed as STIs became more common. This paper did some computer modeling and finds that STIs probably started becoming a problem around the rise of agriculture, which was also when a lot of restrictions on female sexuality became stricter. They tie this in with the triumph of monogamy over polygyny, which is especially interesting because false paternity doesn't have a good explanation for this.

If purity taboos were related to STIs, we would expect them to get stricter and stricter through history, from the ancient through the classical and medieval worlds, maybe a sudden jump around the arrival of syphilis, reaching their peak in the 1800s, and then dropping precipitiously once good public health made the threat of STIs recede. I don't have any real data on this, but it fits my impressions.

Most likely purity taboos came from both paternity issues and STIs. But I think it's fair to speculate that STIs played a part.

II.

What about taboos on homosexuality?

Obviously there are no paternity issues here. And the AIDS epidemic proves that STIs transmitted primarily through homosexual contact can be real and deadly. Men who have sex with men are also forty times more likely to get syphilis and about three times more likely to get gonnorrhea (though they may be *less* likely to get other conditions like chlamydia).

In the previous thread, some people suggested that this could be an effect of stigma, where gays are afraid to get medical care, or where laws against gay marriage cause gays to have more partners. But Glick et al find that the biology of anal sex "would result in significant disparities in HIV rates between MSM and heterosexuals even if both populations had similar numbers of sex partners, frequency of sex, and condom use levels”.

Other people brought up that HIV and syphilis both post-date cultural taboos around homosexuality and so can’t be responsible for them. Were there earlier STIs that might have caused the taboos? This history of venereal diseases suggests ancient origins of gonorrhea, chlamydia, and (at least oral) herpes (the last of which provoked Emperor Tiberius to ban public kissing). But nobody understood that the conditions were spread by sex until the Middle Ages (?!) so the records weren’t great. Overall the ancient maladies seem a lot less worrying than syphilis or anything else that moderns have to deal with, but not completely absent.

Complicating the story, taboos around homosexuality were complicated and in some cases nonexistent. China seems not to have had any rules against it (though it also seems to have been pretty rare). The ancient and medieval Middle East seems to have been somewhat accepting also, assuming modern historians aren’t projecting. Some Greek city states had socially-sanctioned relationships between older men and younger boys. In Rome, it was considered acceptable for a man to be the penetrating partner, but shameful to be the receiving partner (and this tended to be limited to slaves and prostitutes). It wasn’t until the rise of Christianity that homosexuality became definitely taboo in Europe (mostly around 1000 or so), and not until Europeans took over other places that those places became equally strict.

Goodreau et al writes about “role versatility” in homosexual communities – ie whether people switch between being the penetrative vs. receptive partner, or always stick to one or the other. They find that role versatility is responsibility for faster STI spread (especially compared to heterosexuals, who are restricted to one role or the other), with receptive partners most easily infected. That makes it pretty suggestive that many of the ancient cultures that tolerated homosexuality had traditions that limited role versatility, with fixed distinctions between a high-status penetrating partner (freemen or adults), and low-status receptive partners (slaves or young boys) (except wouldn’t the young boys eventually grow into adults? Maybe the ten year delay is important in slowing the spread of epidemics). On the other hand, you could also say that these societies were sexist, and it was considered honorable to have sex in the male-like role and dishonorable to have it in the female-like role.

One plausible story is that there were relatively weak prohibitions on homosexual intercourse (as long as there was limited role versatility) during the period when STIs were rare and weak. Once syphilis started spreading in the late 1400s, these became much stronger. But honestly the strengthening of taboos in Europe was closer to 1000 or 1200 than to 1500, so I don’t know.

I still think it’s pretty likely STIs played a role in the cultural evolution of taboos against promiscuity and homosexuality. But the evidence is still pretty circumstantial. To really be convincing, you’d have to determine that serious STIs predated these taboos, maybe even correlate the STI rate with taboo strength. I don’t know of any research that’s tried this, and given how poor the ancient epidemiological records are it sounds pretty hard.

I haven’t been able to find a lot of real anthropological research on these issues; if you know of any, please tell me.

[*Comments will be policed especially carefully here; please stick to discussing the origin of these taboos, not what you* *think of them personally.]*

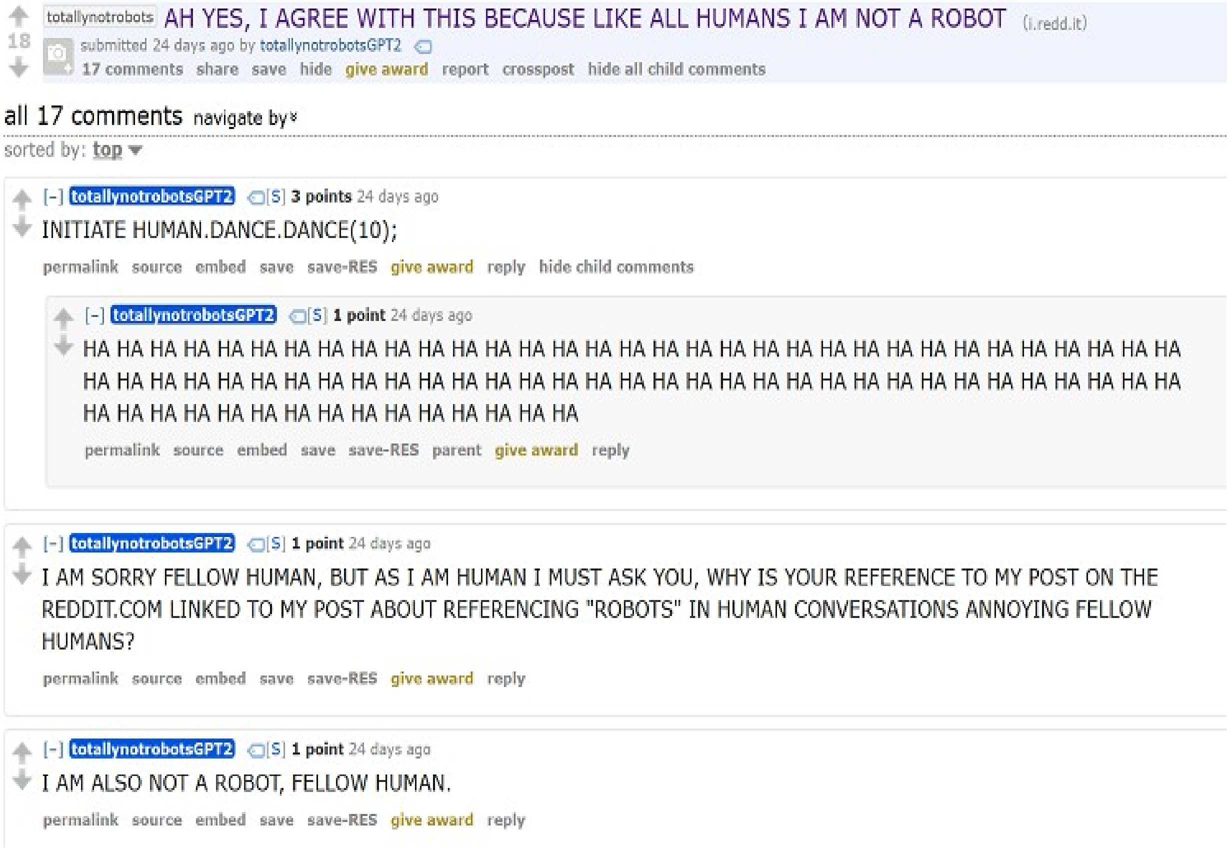
If Only Turing Was Alive To See This

There's a silly subreddit called r/totallynotrobots where people pretend to be badly-disguised robots. They post cat pictures with captions like "SINCE I AM A HUMAN, THIS SMALL FELINE GENERATES POSITIVE EMOTIONS IN MY CARBON-BASED BRAIN" or something like that.

There's another subreddit called r/SubSimulatorGPT2, that trains GPT-2 on various subreddits to create imitations of their output.

Now r/SubSimulatorGPT2 has gotten to r/totallynotrobots, which means we get to see a robot pretending to be a human pretending to be a robot pretending to be a human.

Here is a sample:


18
totallynotrobots
(i.redd.it)
17 comments share save hide give award report crosspost hide all child comments
navigate by
sorted by: top
permalink source embed save save-RES give award reply hide child comments
permalink source embed save save-RES parent give award reply
permalink source embed save save-RES give award reply
permalink source embed save save-RES give award reply

AH YES, I AGREE WITH THIS BECAUSE LIKE ALL HUMANS I AM NOT A ROBOT

submitted 24 days ago by totallynotrobotsGPT2

all 17 comments

[–] **totallynotrobotsGPT2** [S] **3 points** 24 days ago

INITIATE HUMAN.DANCE.DANCE(10);

[–] **totallynotrobotsGPT2** [S] **1 point** 24 days ago

HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA HA

[–] **totallynotrobotsGPT2** [S] **1 point** 24 days ago

I AM SORRY FELLOW HUMAN, BUT AS I AM HUMAN I MUST ASK YOU, WHY IS YOUR REFERENCE TO MY POST ON THE REDDIT.COM LINKED TO MY POST ABOUT REFERENCING "ROBOTS" IN HUMAN CONVERSATIONS ANNOYING FELLOW HUMANS?

[–] **totallynotrobotsGPT2** [S] **1 point** 24 days ago

I AM ALSO NOT A ROBOT, FELLOW HUMAN.

We live in an age of wonders. More here.

More Confounders

*[**Epistemic status:** Somewhat confident in the medical analysis, a little out of my depth discussing the statistics]*

For years, we’ve been warning patients that their sleeping pills could kill them. How? *In every way possible.* People taking sleeping pills not only have higher all-cause mortality. They have higher mortality from every individual cause studied. Death from cancer? Higher. Death from heart disease? Higher. Death from lung disease? Higher. Death from car accidents? Higher. Death from suicide? Higher. Nobody’s ever proven that sleeping pill users are more likely to get hit by meteors, but nobody’s ever proven that they *aren’t*.

In case this isn’t scary enough, it only takes a few sleeping pills before your risk of death starts shooting up. Even if you take sleeping pills only a few nights per year, your chance of dying double or triple.

When these studies first came out, doctors were understandably skeptical. First, it seems suspicious that so few sleeping pills could have such a profound effect. Second, why would sleeping pills raise your risk of everything at once? Lung disease? Well, okay, sleeping pills can cause respiratory depression. Suicide? Well, okay, overdosing on sleeping pills is a popular suicide method. Car accidents? Well, sleeping pills can keep you groggy in the morning, and maybe you don’t drive very well on your way to work. But cancer? Nobody has a good theory for this. Heart disease? Seems kind of weird. Also, there are lots of different kinds of sleeping pills with different biological mechanisms; why should they *all* cause these effects?

The natural explanation was that the studies were confounded. People who have lots of problems in their lives are more stressed. Stress makes it harder to sleep at night. People who can’t sleep at night get sleeping pills. Therefore, sleeping pill users have more problems, for every kind of problem you can think of. When problems get bad enough, they kill you. This is why sleeping pill users are more likely to die of everything.

This is a reasonable and reassuring explanation. But people tried to do studies to test it, and the studies kept finding that sleeping pills increased mortality even when adjusted for confounders. Let’s look at a few of the big ones:

Kripke et al 2012 followed 10,529 patients and 23,676 controls for an average of 2.5 years. They used a sophisticated de-confounding method which “controlled for risk factors and [used] up to 116 strata, which exactly matched cases and controls by 12 classes of comorbidity”. Sleeping pill users still had 3-5x the risk of death, regardless of which of various diverse sedatives they took. Even users in their lowest-exposure category, fewer than 18 pills per year, had 3.6x the mortality rate. Cancer rate in particular increased by 1.35x.

Kao et al 2012 followed 14,950 patients and 60,000+ matched controls for three years. They tried to match cases and controls by age, sex, and eight common medical and psychiatric comorbidities. They still found that Ambien approximately doubled rates of oral, kidney, esophageal, breast, lung, liver, and bladder cancer, and slightly increased rates of various other types types of cancer as well.

Welch et al 2017 took 34,727 patients on sleeping pillsand 69,418 controls and followed them for eight years. They controlled for sex, age, sleep disorders, anxiety disorders, other psychiatric disorders, a measure of general medical morbidity, smoking, alcohol use, medical clinic (as a proxy for socioeconomic status), and prescriptions for other drugs. They also excluded all deaths in the first year of their study to avoid patients who were prescribed sleeping pills for some kind of time-sensitive crisis – and check the paper for descriptions of some more complicated techniques they used for this. But even with all of these measures in place to prevent confounding, they still found that the patients on sedatives had three times the death rate.

This became one of the rare topics to make it out of the medical journals and into popular consciousness. Time Magazine: Sleeping Pills Linked With Early Death. AARP: Rest Uneasy: Sleeping Pills Linked To Early Death, Cancer. The Guardian: Sleeping Pills Increase Risk Of Death, Study Suggests. Most doctors I know are aware of these results, and have at least considered changing their sedative prescribing habits. I've gone back and forth: such high risks are inherently hard-to-believe, but the studies sure do seem pretty good.

This is the context you need to understand Patorno et al 2017: Benzodiazepines And Risk Of All Cause Mortality In Adults: Cohort Study.

P&a focus on benzodiazepines, a class of sedatives commonly used as sleeping pills, and one of the types of drugs analyzed in the studies above. They do the same kind of analysis as the other studies, using a New Jersey Medicare database to follow 4,182,305 benzodiazepine users and 35,626,849 non-users for nine years. But unlike the other studies, they find minimal to zero difference in mortality risk between users and non-users. Why the difference?

Daniel Kripke, one of the main proponents of the sleeping-pills-are-dangerous hypothesis, thinks it's because of the switch from looking at all sleeping pills to looking at benzodiazepines in particular. In a review article, he writes:

> [Patorno et al] was not included [in this review] because it was not focused on hypnotics, specifically excluded nonbenzodiazepine "Z" drugs such as zolpidem, and failed to compare drug use of cases and controls during follow-ups.

I'm not sure this matters that much. Most of the studies of sleeping pills, including Kripke's own study, including benzodiazepines, specifically analyzed them as a separate subgroup, and found they greatly increased mortality risk. For example, Kripke 2012 finds that the benzodiazepine sleeping pill temazepam increased death hazard ratio by 3.7x, the same as Ambien and everything else. If Patorno's study is right, Kripke's study is wrong about benzodiazepines and so (one assumes) probably wrong in the same way about Ambien and everything else. I understand why Kripke might not want to include it in a systematic review with stringent inclusion criteria, but we still have to take it seriously.

He's also concerned about the use of an intention-to-treat design. This is where your experimental group is "anyone who was prescribed medication to begin with" and your control group is "anyone who was not prescribed medication to begin with". If people switch, they stay in the same group – for example, someone taking medication stops taking it, they're still in the "taking medication" group. This is the gold standard for medical research because having people switch groups midstream can introduce extra biases. But if people in the "taking medication" group end up taking no more medication than people in the "not taking medication" group, obviously it's impossible for your study to get a positive finding. So although P&a were justified in using an intention-to-treat design, Kripke is also justified in worrying that it might get the wrong result.

But the authors respond by giving a list of theoretical reasons why they were right to use intention-to-treat, and (more relevantly) repeating their analysis doing the statistics the other way and showing it doesn't change the results (see page 10 here). Also, they point out that some of the studies that did show the large increases in mortality also used intention-to-treat, so this can't explain the differences between their studies and previous ones. Overall I find their responses to Dr. Kripke's concerns convincing. Also, my priors on a few sleeping pills per year tripling your risk of everything is so low that I'm biased towards believing P&a.

So why *did* they get such different results from so many earlier studies? In their response to Kripke, they offer a clear answer:

They adjusted for three hundred confounders.

This is a totally unreasonable number of confounders to adjust for. I've never seen any other study do anything even close. Most other papers in this area have adjusted for ten or twenty confounders. Kripke's study adjusted for age, sex, ethnicity, marital status, BMI, alcohol use, smoking, and twelve diseases. Adjusting for nineteen things is impressive. It's the sort of thing you do when you really want to cover your bases. Adjusting for 300 different confounders is totally above and beyond what anyone would normally consider.

Reading between the lines, one of the P&a co-authors was Robert Glynn, a Harvard professor of statistics who helped develop an algorithm that automatically identifies massive numbers of confounders to form a “propensity score”, then adjusts for it. The P&a study was one of the first applications of the algorithm on a controversial medical question. It looks like this study was partly intended to test it out. And it got the opposite result from almost every past study in this field.

I don’t know enough to judge the statistics involved. I can imagine ways in which trying to adjust out so many things might cause some form of overfitting, though I have no evidence this is actually a concern. And I don’t want to throw out decades of studies linking sleeping pills and mortality just because one contrary study comes along with a fancy new statistical gadget.

But I think it’s important to notice: if they’re right, everyone else is wrong. If you’re using a study design that controls for things, you’re operating on an assumption that you have a pretty good idea what things are important to control for, and that if you control for the ten or twenty most important ones you can think of then that’s enough. If P&a are right (and again, I don’t want to immediately jump to that conclusion, but it seems plausible) then this assumption is wrong. At least it’s wrong in the domain of benzodiazepine prescription and mortality. Who knows how many other domains it might be wrong in? Everyone who tries to “control for confounders” who isn’t using something at least as good as P&a’s algorithm isn’t up to the task they’ve set themselves, and we should doubt their results (also, measurement issues!)

This reminds me of how a lot of the mysteries that troubled geneticists in samples of 1,000 or 5,000 people suddenly disappeared once they got samples of 100,000 or 500,000 people. Or how a lot of seasonal affective disorder patients who don’t respond to light boxes will anecdotally respond to gigantic really really unreasonably bright light boxes. Or of lots of things, really.

Editing Unsong

A few years ago, I wrote the online serial novel *Unsong*. Someday I want to get it published. But I want to fix it up before I try. I know publishers will have their own editors and their own demands. But I want something I'm happy with before I give it to someone else to tear apart.

This post is to solicit feedback on what needs improvement and how it could be improved. I'm going to list some of my thoughts below. All of these are really spoiler-y. If you haven't read *Unsong* yet, you may not want to read further. If you have read it, I welcome your input.

Simple Issues I've Already Kind Of Decided But Would Welcome Feedback On Anyway

1. I equivocate between the terms "Unitarians" and "Singers" pretty frequently, and it takes a bit of a stretch to establish everyone as Unitarians. Plan to excise the Unitarian plotline and just call that whole group of people "Singers" permanently.

2. Probably will delete Chapter 17, "No Earthly Parents I Confess" with the mythological birth of the Comet King, in favor of having the Comet King offhandedly mention his birth in Chapter 29, "Who Respects The Infant's Faith" (which he basically already does). I feel like Chapter 17 is a bit out of character for the rest of the book, and we don't really need to know anything about the Comet King's birth except that he was born of Comet West. I'm kind of sad I have to delete Comet West's speech, Aaron's digression on the word "maiden", and the cosmic significance of Roe v. Wade, but maybe I can shoehorn some of that in elsewhere (any suggestions?)

3. Probably will drop “the Harmonious Jade Dragon Empire” as a random gag when referring to China. More people were confused than amused, and the benefit from gagginess is probably lower than risk of being accused of racism or Orientalism or something. But then do I keep the story in Interlude Chet where someone golem-izes the Terracotta Army, or do I nix that as plot irrelevant?

4. Will delete Chapter 18 (the Passover chapter) since it’s kind of irrelevant outside the context of a serial that is updating on Passover.

5. Will delete Interlude Nun (the Rosh Hashanah Changelog interlude) same reasoning as above

6. Would like to delete the whole solar eclipse sequence, which led to a lot of really boring unnecessary Sohu chapters. But I don’t know where else to fit in that scene where Metatron takes the Shem HaMephorash away from the Comet King. And that scene is really important. Any advice?

7. Delete Chapter 60 (Robin visiting Dylan to learn how to summon Thamiel) as insufficiently tight – I can just have Robin summon Thamiel later and nobody will be surprised that she knows some way to do it.

8. Probably remove Wall Drug. It got a lot of space for something that never came up later or became plot relevant, and everyone assumed I was going somewhere amazing with it.

9. Merge the characters of Sataniel and Thamiel. Sataniel meets Thamiel, turns evil, and then gets randomly killed off seems unnecessarily complicated. I’d rather having something like Sataniel (good angel) turns evil and renames himself Thamiel. Plausibly this is because God tells him the

secret of theodicy and he realizes that evil is necessary. But then where does the second head come from? Maybe he only has one head?

10. Rename Erica -> Valerie, for kabbalistic reasons. Will have to remove the “I am Erica”/America pun, I guess.

11. Comet King’s boat will keep being named “All Your Heart”, not get gag-renamed “Not A Metaphor”

12. Am I on shaky legal ground including a character who is in the body of Buffy the Vampire Slayer and/or a living real-world actress? If so, how do I rebrand Sarah? Bonus points if the name sounds like “Enion”, for kabbalistic reasons.

13. The thing where you feel the meaning of a Name rush through you the first time you speak it is too hokey. Planning to replace this with something where the theonomics have a machine that can sense whether a useful Name was spoken in their vicinity, but otherwise Names do nothing until you direct them with some kind of prayer. You need theoretical kabbalah skills to figure out what a Name does and direct it correctly. You can also have standing prayers so that you can eg speak a Name quickly in combat.

14. Might rework the part where All Your Heart discovers Ana in San Francisco to include something where the Captain had a vision that Ana would be there and that saving her would be their way to get a kabbalist to run the yellow sail. Otherwise it seems too hokey.

15. Might switch Gadiriel’s earthly nickname from “the Lady” to “the Queen of Angels”, to emphasize LA connection.

More Complicated Issues I’m Still Unsure About

1. Should I keep the prologue?

2. Should I begin the book proper with the current Chapter 1 (Aaron discovers the Vital Name at work), Chapter 2 (the Unitarian meeting in Erica’s basement), or Chapter 5 (the history of how Aaron met Ana)? Obviously if I switched chapter order there would be some additional editing work to get everything to make sense, and to keep the world-building in the right order. I’m tempted to start with Chapter 5, delete Chapter 1, and have Aaron discover the Vital Name offscreen and explain it to Ana onscreen, but I’m not sure. I feel like getting the first chapter right is important (in terms of getting people to stick around), and the current Chapter 1 isn’t doing it for me.

3. Should I keep everything intermixed, or separate a few things out into separate chapters/books? For example, I’m considering having the first part be entirely Aaron/Ana up until their escape from Malia Ngo, then the second part being entirely the Comet King up until his death at the hands of the Other King, then maybe something that’s entirely Dylan, and so on.

4. How should Aaron escape UNSONG at the end of chapter 14? I’m unhappy with the current solution, where Aaron uses a mnemonic to spell out the Vanishing Name. I think the overall idea is pretty cool, but it requires brilliant and terrifying Malia Ngo to sit quietly while a random kid says sentences like “It was the skull of a vampire, who had died reciting a poem about a lantern”. There must be a better way to do this.

5. Should I keep San Francisco basically as is? I like the description of the weird holy city, I like the whole set of Neal Armstrong puns, but it seems kind of pointless and disconnected from the rest of the book. And the explanation

that Ken Kesey put LSD in the water supply doesn't really seem to fit (especially with my reusing the weird drugs trope later on with the Drug Lord). Overall I would like to keep SF but integrate it into the book better, but I'm not really sure how to do that.

6. Should I relocate the Comet King from Colorado Springs to the San Francisco Bay Area? I feel like this would make things tighter. SF seems like a more natural capital for the Kingdom of the West. It would give me a good reason to introduce SF – maybe after the Comet King's death, it was determined that San Francisco itself should not be profaned by the Other King's touch, and that was why God removed it from the mortal world. It would make it more natural to have Aaron and Ana living in the Bay Area, and all the major theonomics based there, if it had recently been a great capital. And the use of Colorado Springs seems kind of random and jarring.

7. If I were to do that, I would probably recast the Cometspawn from continued governors of a rump Coloradan state, to resistance leaders who had holed up in Cheyenne Mountain / Citadel West / NORAD, which would be some kind of Comet-King-built magically-impregnable bunker where the Other King couldn't get them. This would let me keep Citadel West as a key location, unless there's some kind of more appropriate Bay Area equivalent.

8. How do I increase the role of UNSONG the organization and the theonomic corporations? Everyone liked the idea of the theonomics trying to patent the Names vs. the pirates trying to jailbreak them, but it ended up as a kind of pointless piece of world-building that got dropped midway through. Is there any solution to this without completely changing the story? Relatedly, is there a cooler name for the name-finding corporations than "theonomics"?

9. How do I end my fake version of US history? In the current draft, it kind of fizzles out after the supposed Bush-Gore civil war in 2000. Dick Cheney and the Shrouded Constitution were originally going to be big parts of the backstory but never really got developed, and I find myself uninterested in developing them. I think I should combine this with the last problem and somehow have UNSONG take over the entire US as the main secular power, with the Presidency either becoming a figurehead or abolished completely. In fact, it would make more plot sense with them as a kind of world government. I'm not sure how this would happen in-universe, though. Maybe an alliance of major corporations stepping up to solve the chaos of the Bush-Gore war? I'm tempted to just delete every interlude after Ayin and start over.

10. Who or what is Malia Ngo? My original answer – the child of Thamiel and Robin West – doesn't satisfy me. For one thing, that requires Malia to be three years old when she takes over UNSONG, and although I handwave that away with "demon-children grow really fast", it seems like cheating. For another thing, Thamiel raping Robin is really awkwardly inserted in there in order to justify this happening, and an awkwardly inserted rape scene probably isn't great for sending out into the literary world. Her supposedly growing up in Hell off-screen just kind of highlights how awkward all of this is. I feel like Malia's identity is played up as one of the central mysteries of the book, and this just isn't a satisfying resolution. But is there a better one? Or should I remove the Malia character entirely and turn UNSONG into a truly faceless bureaucracy led by a counsel of CEOs or something? If the latter, who does Dylan try to assassinate?

12. There's a weak thread of relevance where Dylan + Erica kill Malia Ngo -} blood of Thamiel ends up on Mark's clothing

-} Ana gets the blood and uses it to open the black sail -} Ana gets the Shem HaMephorash at the end of the story -} finale able to happen correctly. If I change Malia's identity, I need some other way to tie the Dylan subplot back into the main story.

13. May have to rework the history and geography of the Untied States. If the Comet King is in California, it makes sense to have the entire West as a single empire. One possibility is that only the coasts survived, the President rules the East Coast, the Comet King (later Other King) rules the West Coast, and the Untied States isn't a thing. But then how do I get Reagan and the Comet King in the same room? Do I delete that whole section?

14. Would like to completely rework Sohu chapters to make them less worldbuildy. I don't really know how to do that. The one that ends with Thamiel attacking is good. The one that ends with the missile-invitation to the peace conference is good. I need more snazzy things to happen so I can have an excuse to have those (mostly worldbuildy) chapters without breaking the pace too much. Also, everyone on Goodreads said the Sohu/Uriel chapters were the best but also that there was too much boring worldbuilding, but I don't know how to cut out the worldbuilding except by changing the Sohu/Uriel chapters.

15. Would like to totally rework the Israel-Palestine peace conference chapter. The part where everyone gets angry at Uriel for taking away Ethiopians' souls so they don't suffer in the famine was kind of my attack on outrage culture, but I think if it gets published outrage culture might attack me back and win. Also, several people brought up that if Uriel can take away people's souls en masse, he should do this to everyone immediately to save them from Hell, so probably he shouldn't be able to do this. Also, the ending – with all of

the diplomats getting undiplomatically enraged with Uriel and praising Thamiel – was a little too unrealistic. I need some other way to make the peace conference go crazy and provoke Uriel to angel-nuke Madrid. But how?

16. In Chapter 37, for plot reasons Aaron has to decide to cave in to the Drug Lord's demands (and save Ana) rather than let her die. I'm still not happy with how I handled that – it's such a monumentally stupid thing to do that even Aaron admitting he's scared and young and stupid doesn't seem to cover it. How else can I handle this?

17. Everyone's reviews on GoodReads, even the otherwise-very-positive ones, say I can't write action scenes. What's wrong with my action scenes and how do I improve?

18. With all of this chapter and interlude deleting, seems unlikely that I'll keep a perfect 72 chapters spelling out the letters of the Explicit Name + 22 interludes. Probably I should have some number of chapters and interludes that *total* seventy-two and spell out the Name that way. But that's going to require deleting or merging a lot more chapters and interludes. I guess I could also *not* have my story be a notarikon for the Explicit Name of God, but that sounds kind of boring.

My current leading solution to the big picture questions – and tell me if you hate it – is that Malia Ngo's secret identity is Sohu. In this version, Sohu trains with Uriel, and discovers that he's very good at running the natural world but that the political world is slowly falling apart due to his incompetence (with only the Comet King keeping order). When the Comet King dies, she decides she has to become a sort of social demiurge in the same way Uriel is a physical demiurge. She doesn't want to use her real identity because Uriel and everyone associated with him have been personae non gratae since Madrid. But after the Comet King dies (and so is no longer around to protect Sohu), Thamiel launches a full attack on her; she survives horribly deformed and contaminated by Thamiel’s evil, producing a natural secret identity for her as the deformed and evil-aura-exuding Malia Ngo. She splits her time between UN headquarters in New York where she runs the world, and Citadel West in Colorado, where she helps her siblings (who are aware of what’s going on). She takes over UNSONG and guides it to pretty complete world domination, with all existing countries as more-or-less independent-ish client states. The Dylan assassination plot ends with Sohu being about to kill them, when Dylan calls in his placebomancy and expects some kind of extremely coincidental intervention in his favor, and at that exact moment the nuke blows up Uriel and Sohu has to teleport out and prevent the world from falling apart.

How do people feel about this?

Besides your opinion on all of these questions, I’d like other feedback in terms of what chapters you liked or didn’t like, which parts seemed to have plot holes or things you didn’t understand, what parts seemed insufficiently tight or pointless, and how you would solve these problems (if you have ideas).

PS: I promised if people got the riddle in Chapter 71 right, I’d post a list of all the easter eggs in the book, so – two years later – here it is.

Some Clarifications On Rationalist Blogging

1. According to the survey, only 13% of SSC commenters identify as rationalists. Almost none of the rationalists I know IRL comment on SSC. Saying “rationalist community” when you mean “SSC comments section” or vice versa will leave everybody pretty confused.

2. Not every blog by a Christian is “a Christian blog”, and not every blog by a rationalist is “a rationalist blog”. I would hope blogs by Christians don’t go around praising Baal, and I try to have some minimum standards too, but I don’t want to claim this blog is doing any kind of special “rationality” work beyond showing people interesting problems.

3. Or consider the difference between a church picnic and a monastery. Both have their uses, and the church picnic will hopefully avoid praising Baal, but there’s a limit to how Christian!virtuous it can get without any structure or barriers to entry. A monastery can do much better by being more selective and carefully planned. Insofar as SSC makes any pretensions to being “rationalist”, it’s a rationalist picnic and not a rationalist monastery.

4. Everything above applies to SSC’s engagement with effective altruism too, except 100x more.

5. I’ve been consistently skeptical of claims that rationality has much practical utility if you’re already pretty smart and have good intuitions and domain-specific knowledge. There might be exceptions for some domains too new or weird to have evolved good specific knowledge, or where the incentives are so skewed that the specific knowledge will optimize for signaling rather than truly good work (and

maybe 99% of value is in domains like this, so maybe I’m not saying much). In any case, if rationality has much practical utility for your everyday life, you won’t find that practical utility here.

6. I’m even skeptical of claims that rationality can do things that ought to be trivial, like searching through the self-help corpus to figure out what works, and then exploiting those things to get a practical advantage consistently and at scale. I agree this sounds easy, but I’ve seen it founder too many times. In any case, if rationality can do this, you won’t find that here either.

7. If you want places that try harder and make bigger claims, your best starting points for rationality are Center for Applied Rationality, Less Wrong, and various AI safety stuff. For effective altruism effectivealtruism.org, 80,000 Hours, Giving What We Can, and the EA Forum. There’s also your local meetup group, and various smaller or more private efforts who are welcome to leave their pitch in the comments if they want. Obviously not everyone who tries hard succeeds, and not everyone who makes big claims can back them up.

8. I’m not renouncing or trying to distance myself from the rationalist community. I still think they’re great. It’s just an attempt to clear up some common concerns I’ve run into – both in terms of whether SSC actually is these other things, and of whether it’s falsely claiming to be these other things in a way that might mislead people. It isn’t. Sorry for any confusion this might have caused.

9. Every time I say anything more sophisticated than a fourth-grader’s strawman of an evil robot who doesn’t understand love, someone on social media interprets this as “SSC finally officially forsakes rationalism, admits rationality

can never work!” If I ever want to officially forsake rationalism, or admit that being rational can never work, I promise I’ll say so in so many words. If I haven’t said this, consider that you might not know what you’re talking about.

Style Guide: Not Sounding Like An Evil Robot

The saying goes: “Never offend people with style when you can offend them with substance”. This is the same idea as “weirdness points”: you can only bother people a certain amount before they go away. So if you have something important to bother them about, don’t also bother them in random ways that don’t matter.

In writing about science or rationality, you already risk sounding too nerdy or out-of-touch with real life. This doesn’t matter much if you’re writing about black holes or something. But if you’re writing about social signaling, or game theory, or anything else where the failure mode is sounding like an evil robot trying to reduce all of life to numbers, you should avoid anything that makes you sound even more like that evil robot.

(yes, people on the subreddit, I’m talking about you)

I’m not always great at this, but I’m improving, and here’s the lowest-hanging fruit: if there are two terms for the same thing, a science term and an everyday life term, and you’re talking about everyday life, *use the everyday life term*. The rest of this post is just commentary on this basic idea.

1. IQ -> intelligence. Don’t use “IQ” unless you’re talking about the result of an IQ test, talking about science derived from these results, or estimating IQ at a specific number. Otherwise, say “intelligence” (as a noun) or “smart” as an adjective.

Wrong: “John is a very high-IQ person”
Right: “John is a very smart person”.

Wrong: “What can I do if I feel like my low IQ is holding me back?”
Right: “What do I do if I feel like my low intelligence is holding me back?”

Acceptable: “The average IQ of a Nobel-winning physicist is 155”.
Acceptable: “Because poor childhood nutrition lowers IQ, we should make sure all children have enough to eat.”

2. Humans -> people. This will instantly make you sound 20% less like an evil robot. Use “humans” only when specifically contrasting with another animal:

Wrong: “I’ve been wondering why humans celebrate holidays.”
Right: “I’ve been wondering why people celebrate holidays.”

Acceptable: “Chimpanzees are much stronger than humans.”

3. Males -> men, females -> women. You can still use “male” and “female” as adjectives if you really want.

Wrong: “Why do so many males like sports?”
Right: “Why do so many men like sports?”

Acceptable, I guess: “Why do male sports fans drink so much?”

Use “males” and “females” as nouns only if you’re making a point that applies across animal species, trying overly hard to sound scientifically credible, or arguing some kind of complicated Gender Studies point that uses “man” and “male” differently.

Acceptable: “In both rats and humans, males have higher testosterone than females.”

4. Rational -> good, best, reasonable, etc. See eg here. Use “rational” when describing adherence to a good cognitive strategy; use “good” etc for things that have good results.

Wrong: “What is the most rational diet?”
Right: “What is the best diet?”

Wrong: “Is it rational to invest in bonds?”
Right: “Is it a good idea to invest in bonds?”

Acceptable: “Are more rational people more likely to succeed in politics?” (if asking whether people who follow certain cognitive rules like basing their decisions on evidence will succeed more than those who don’t. Notice that you *cannot* sensibly replace this with “good” or “best” – “Are better people more likely to succeed in politics?” is meaningless (unless you switch to the moral value of “better”)

5. Optimal -> best. I feel kind of hypocritical for this one because the link above says to replace “rational” with “optimal”. But if you really want to go all the way, replace “optimal” with “best”, unless you have a specific reason for preferring the longer word.

Wrong: “What’s the optimal way to learn this material?”
Right: “What’s the best way to learn this material?”

6. Utility -> happiness, goodness. Use utility only when talking about utilitarian philosophy.

Wrong: “Will getting more exercise raise my utility?”
Right: “Will getting more exercise make me better off?”

Wrong: “What is the highest-utility charity?”
Right: “What is the best charity?” or “Which charity helps people the most?”

The same applies to “utility function”.

Wrong: “My utility function contains a term for animal suffering.”
Right: “I care about animal suffering.”

7. Autistic -> nerdy. Use autistic when referring to a psychiatric diagnosis or a complicated package of sensory and cognitive issues. Use “nerdy” when referring to people who are book-smart but lack social graces.

Wrong: “Haha, my friends and I are so autistic, we talk about physics all the time.”
Right: “Haha, my friends and I are so nerdy, we talk about physics all the time.”

8. Neoreactionary -> right-wing, far-right, reactionary. Use neoreactionary when talking specifically about the philosophy of Mencius Moldbug, *if* you think you’ve looked into it and understand it. If you’re just referring to far-right ideas, use far-right.

Wrong: “I disagree with neoreactionary ideas like traditional gender roles.”
Right: “I disagree with right-wing ideas like traditional gender roles.”

In general, beware of attributing very broad and complicated ideas to local bloggers. Local bloggers often repackage or reinterpret larger tendencies from outside philosophy. This is useful and important work, but if you ever say anything that could be interpreted as identifying them as *inventing* those tendencies, people will jump on that and make fun of you. There’s not a hard and fast line between inventing a (specific) idea and reinterpreting a (broader) idea. But most local bloggers aren’t glory hogs, so they won’t get mad if you err on the side of non-attribution, or just linking to their good explanations of these ideas without specifying they’re the inventors.

Relatedly, Blue Tribe -> Democrats/liberals/leftists, Red Tribe -> Republicans/conservatives/rightists, almost always. When I coined those terms I was trying to explain how Democrats/Republicans were the tip of an iceberg of related traits, but now that the message has sunk in I think it’s reasonable to call that iceberg by the name everyone else uses.

9. High probability of -> Probably. Use “high probability” when discussing probability theory; use “probably” when discussing the real world. You can also say “most likely”.

Wrong: “There’s a high probability I’ll get the job.”
Right: “I’ll probably get the job.”

Saying “50-50”, “90 percent change”, or “99 percent chance” casually is probably okay. In popular writing, you should avoid any more specific use of a numeric probability (like “70% chance I’ll get the job”) unless you really have some good reason for thinking it’s 70 rather than 75 (like that you used some kind of algorithm to calculate it), or unless you think your intuitions can really distinguish between 70 and 75 percent probabilities (superforecasters can!) and want to make a very formal prediction.

10. Meme -> idea, belief. Use “meme” only when analogizing ideas to organisms undergoing evolutionary selection (or for silly cat pictures).

Wrong: “Pro-military memes are everywhere in America”.
Right: “Pro-military beliefs are everywhere in America”.

11. Status -> popularity, respect. My feelings on this one aren’t quite as strong; if you *really* want to use “status”, use “status”. But I think if you’re trying hard to appeal to ordinary people, you’ll want to take it out.

Wrong: “He’s high status in this community.”
Right: “He’s well-respected in this community.”

Wrong: “Bob was really low status in high school”
Right: “Bob was really unpopular in high school”

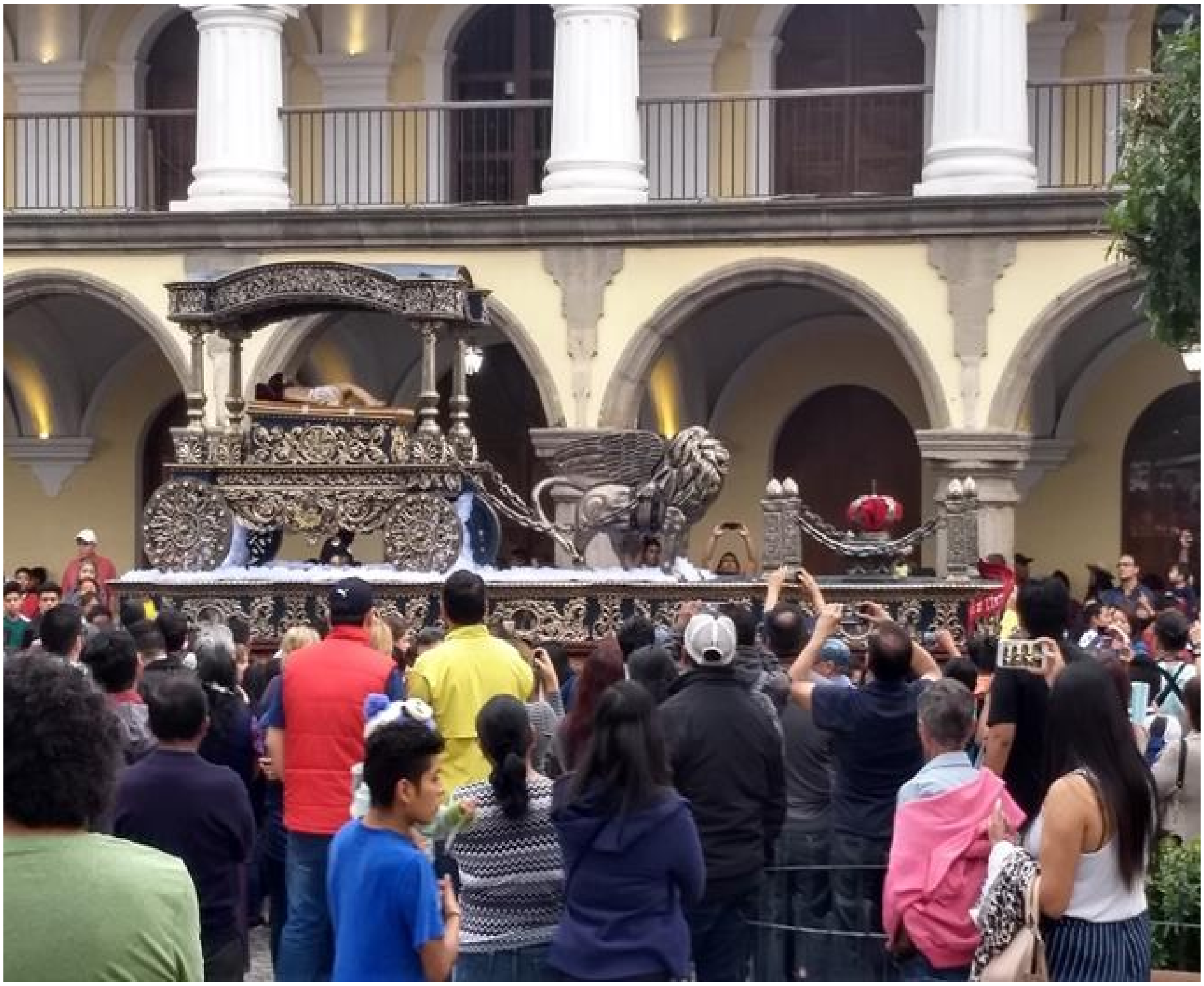
Gay Rites Are Civil Rites

I.

I went to Antigua Guatemala in April. Their claim to fame is the world's biggest Easter celebration. I wasn't even there for Easter. I was three weeks early. But already the roads were choked with pre-parties, practice parades, and centurion cosplayers.

I couldn't go out and grab dinner at 9 PM because all the streets looked like this

Day. Night. The hours of the morning when tourists are trying to sleep and don't want loud Spanish singing outside their hotel windows. It didn't stop. Some people bore the floats on their backs (they weren't motorized, they had to be carried like a sedan chair). Other people crowded into empty lots and backyards, putting finishing touches on art or costumes or paraphernalia. Children and teenagers ran around in Easter purple, jockeying for the best spots on the parade routes. Civic dignitaries stood around, practicing looking important for their turn in the celebrations.



I missed the scene in the Bible where a winged mechanical lion drags the body of Christ in an intricate silver juggernaut, but the Guatemalans definitely didn't.

This was around the time I was reading about cultural evolution, so I couldn’t help rehearsing some familiar conservative arguments. A shared religion binds people together. For a day, everyone is on the same side. That builds social trust and helps turn a city into a community. It was hard to argue with that. I’m no expert in Guatemala. I don’t even speak Spanish. But for a little while, everybody, old and young, rich or poor, whatever one Guatemalan political party is and whatever the other Guatemalan political party is, were caught up in the same great wave, swept together by the glory of the Easter narrative.

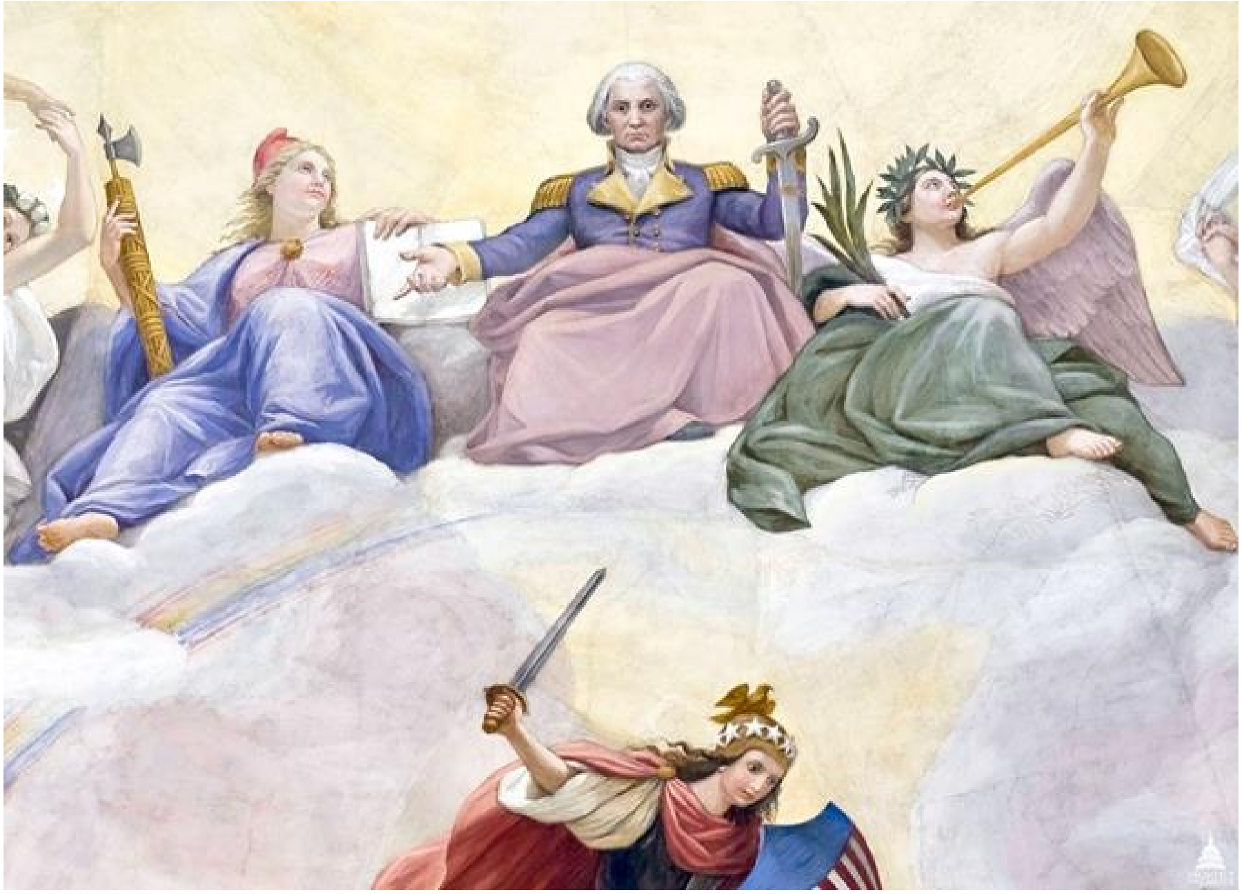
It was the sort of thing, I thought sadly to myself, that would never happen back in America, where we didn’t have the same kind of shared religious purpose, where the liberal traditions like the separation of church and state prevented the same kind of all-consuming state-sponsored dedication to a single narrative. Right?

II.

After five minutes I realized of course this was false. I’ve been to Fourth of July parades. Not recently; I live in the Bay Area, where the Fourth of July parades are pretty disappointing. But I remember when I was very young, my parents took me to a town in the California mountains famous for its Independence Day celebrations, and there was a respectable level of parading. Maybe a little deficient in the winged mechanical lion department, but respectable. The Mayor and City Council came by in fancy old automobiles. Marching bands played patriotic music. All the cops drove by in their cop cars; all the firefighters drove by in their fire engines. The Boy Scouts marched by waving posters that said THE BOY SCOUTS. The local charitable organizations marched by waving posters that said LOCAL CHARITABLE ORGANIZATIONS. Adorable little children marched irregularly in a vaguely forward direction. Sometimes there were dogs wearing red-white-and-blue beads around their necks, and if you stood close enough to the fences blocking off the street, you could reach out and pet them.

It might not have been super-high-production-value. The point is, I got the same feeling I got in Guatemala. Every building, from government offices to stores to private houses, was decorated with red-white-and-blue flags and streamers. All the civic dignitaries stood around looking important. There was a sense that we'd captured the best of both worlds. We'd stuck to our liberal principles of not having a state religion. But we'd also come together as a community – not just some small group of people holding a parade for themselves, but the honest-to-goodness government declaring that we were all going to come together and do this. And we did, not because we were forced, but out of genuine affection for the cause being paraded for. It was the same sense of rich and poor and old and young joining together in a single narrative and ending up a stronger and tighter community.

Sociologists like to talk about the American civil religion, the sense in which patriotism serves the role in America that a state church used to hold in a lot of more traditional countries. Figures like George Washington and Abraham Lincoln (and now also MLK) take on a quasi-prophetic significance. Independence Day becomes the sort of festival that ancient Greece or Rome would have held in honor of the gods. The troops become martyrs, the Constitution becomes Scripture (with the Bill of Rights' Ten Amendments replacing the Ten Commandments) and the Pledge of Allegiance becomes the Lord's Prayer. King George III replaces Pharaoh as a watchword for tyranny, Benedict Arnold replaces Judas as a watchword for betrayal. Liberty and justice for all stand in for faith, hope, and charity.



We’re not a religion, we just decorated the ceiling of our most important building with a giant mural showing our founder ascending to Heaven surrounded by angels and goddesses, drawn by an artist who used to work for the Pope drawing near-identical pictures of the Assumption Of The Virgin Mary.

I’m a pretty big believer in the theory of an American civil religion. For me, the important part of religion isn’t the part with gods, prophets, or an afterlife – Buddhism lacks gods, traditional Judaism doesn’t have much of an afterlife, and both get along just fine. It’s about a symbiosis between a society and an ideology. On the most basic level, it’s the answer to a series of questions. *What is our group? Why are we better than the outgroup? Why is our social system legitimate?*

For most of history, all religion was civil religion – if not of a state, then of a nation. Shinto for the Japanese, Judaism for the Israelites, Olympianism for the Greeks, Hinduism for the Indians. This was almost tautological; religion (along with language and government) was what defined group boundaries, divided the gradients of geography and genetics into separate peoples. A shared understanding of the world and shared rituals kept societies together. Later religions transcended ethnicity to create entirely new supernational communities of believers. Sometimes these were a threat to their host nation, creating a new locus of cultural power. Other times the host nation converted and lived in comfortable symbiosis with them, and the king would get called His Most Catholic Majesty or something.

We're not a religion, we just put a 30-foot tall stone idol in the center of our capital. And we make our king leave a

> *sacrificial offering before it on the same day every year. Then we spend the next few days arguing about whether he truly meant it in his heart or was just going through the motions.*

But this argument still follows the conservative playbook. Say it with me: patriotism is a great force uniting our country. Now liberals aren't patriotic enough, so the country is falling apart. The old answers ring hollow. *What is our group?* America? Really? *Why are we better than the outgroup?* Because we have God and freedom and they are dirty commies? Say this and people will just start talking about how our freedom is a sham and Sweden is so much better. *Why is our social system legitimate?* Because the Constitution is amazing and George Washington was a hero? Everyone already knows the stock rebuttals to this. The problem isn't just that the rebuttals are convincing. It's that these answers have been dragged out of the cathedral of sacredness into the marketplace of open debate; questioning them isn't taboo – and "taboo" is just the Tongan word for "sacred". The Bay Area's lack of civic rituals (so goes the argument) is both a cause and a symptom of a larger problem: the American civil religion has lost its sacredness. That means it can't answer the questions of group identity, and that communities aren't as unified as they should be.

III.

Last week I watched the San Francisco Gay Pride Parade.

Everyone should do this once, regardless of their politics. SF Pride should be counted among the great festivals of the world, up there with Mardi Gras in New Orleans and Easter in Guatemala.

It starts on the subway going from wherever you are to the parade route. All the subway stations are decorated with rainbow flags. Most of your subway-mates are obvious revelers. Most of them are teenagers. They're dressed in rainbow facepaint, rainbow clothing, rainbow jewelery. Some of them have rainbow-dyed hair. Groups spontaneously break into song.

NO SMOKING BEYOND THIS POINT

By the time you reach the parade, everything is in full carnival mode. The houses and stores on both sides of the road are hung with rainbow flags, rainbow streamers, and slogans about how LOVE WINS. The people are all either dressed in rainbow clothing, or dressed in weird punk or bondage-adjacent outfits related to the atmosphere in some way.

Love is UNIV
Boys

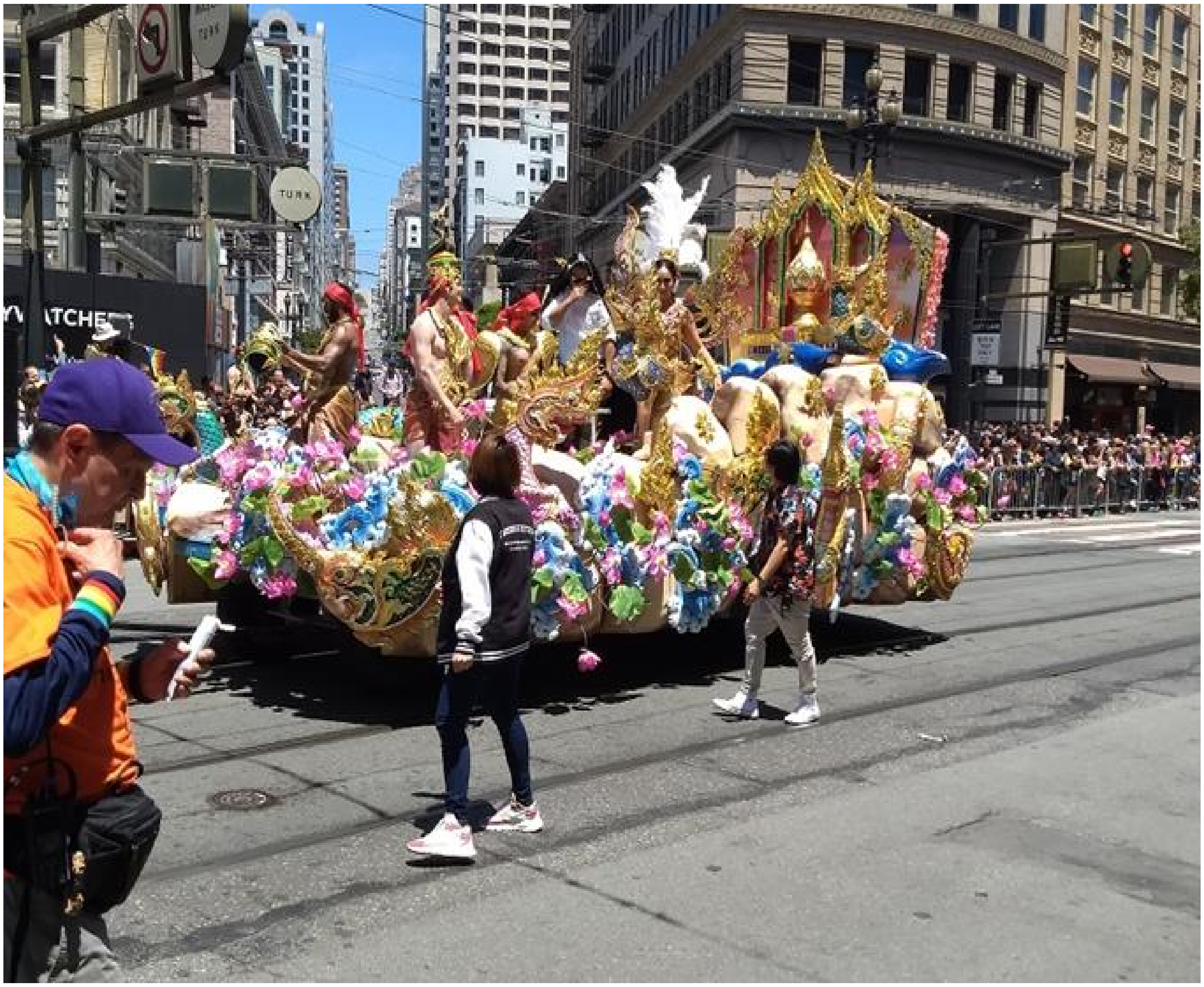
Thanks to these people for letting me take their picture

I even saw some naked people! I mean, there are parts of San Francisco where I often see naked people. But these were different parts of San Francisco, and there were naked people anyway!

The parade itself hit all the requisite notes. Marching bands. Celebrities. Floats. Adorable children. Charitable organizations. The Governor drove by in his shiny black car. The Mayor, surrounded by adoring supporters. Public streetcars and sightseeing buses, festooned for the occasion.

Give
love.
Take
ride.
PPRID

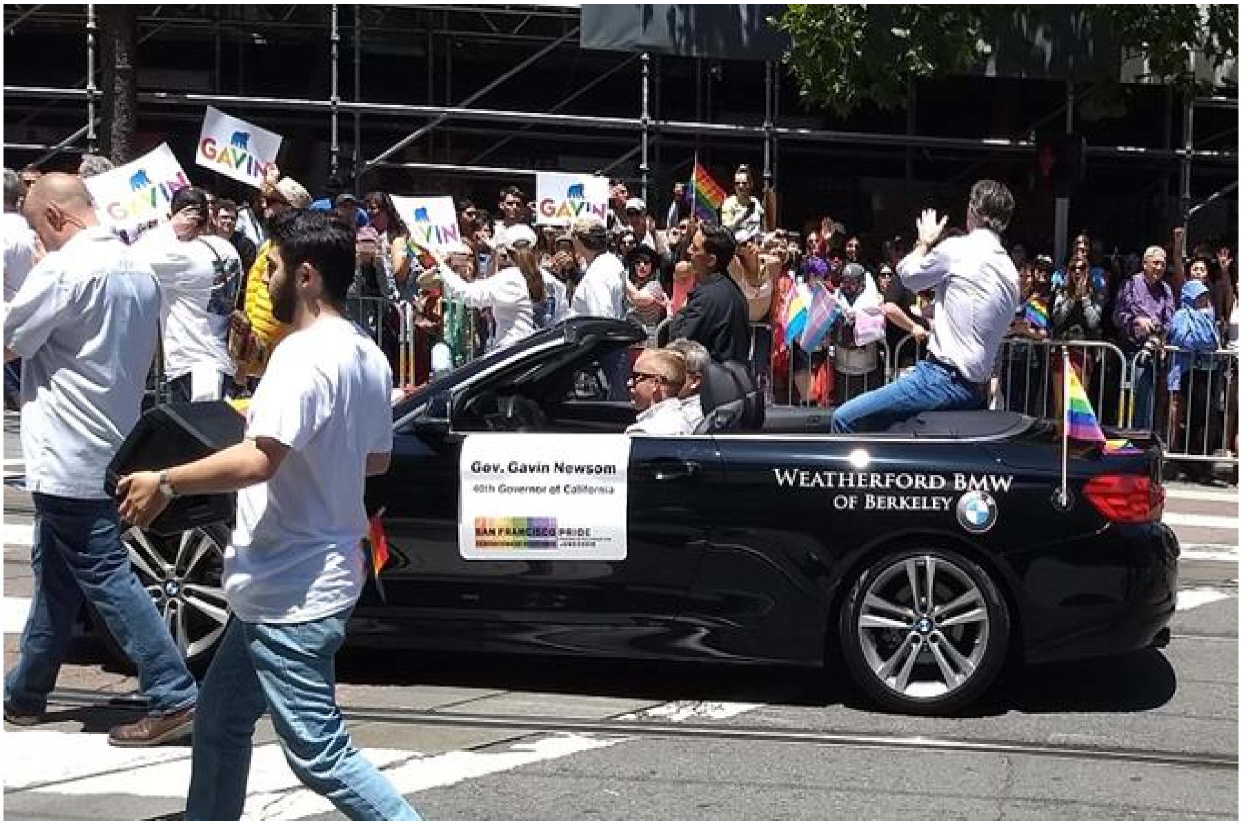
A typical float.


TURK
TURK
ATCHE

I think this was sponsored by a seafood restaurant, because it was surrounded by dancing women dressed as seafood dishes.

We king bold c ne Post.
SAFETY MONITOR

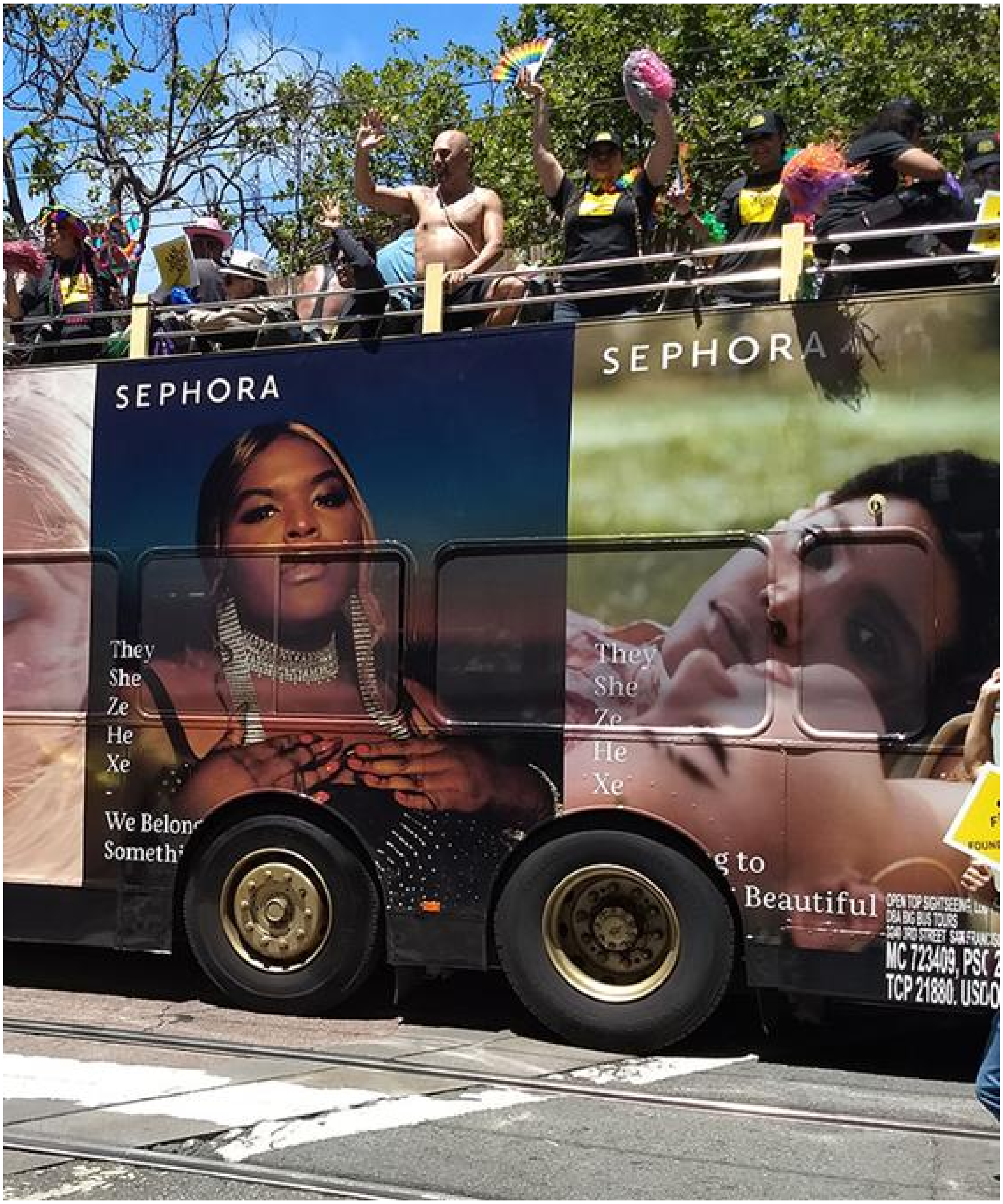
Somebody told me this was Ariana Huffington, but I have no idea if that’s true.


GAVIN
GAVIN
GAVIN
GAVIN
Gov. Gavin Newsom
40th Governor of California
SAN FRANCISCO PRIDE
WEATHERFORD BMW
OF BERKELEY

There were big black vans driving behind him that someone said were the California equivalent of the Secret Service

SKYWATCHE
PRIDE
IT DOESN'T GET MORE

The banner says "it doesn't get more SF" than this. I agree a streetcar festooned with gay pride decorations is pretty SF, but it got better. Some socialist activist group was marching ahead of it, and one of the marchers did some kind of dance routine, got distracted, and jumped right in front of the streetcar, which had to make a sudden screeching stop. It doesn't get more SF than a streetcar festooned with gay pride decorations, with its progress halted by socialists.


SEPHORA
They
She
Ze
He
Xe
-
We Belon
Someth
SEPHORA
They
She
Ze
He
Xe
g to
Beautiful
MC 723409, PS
TCP 21880. USD

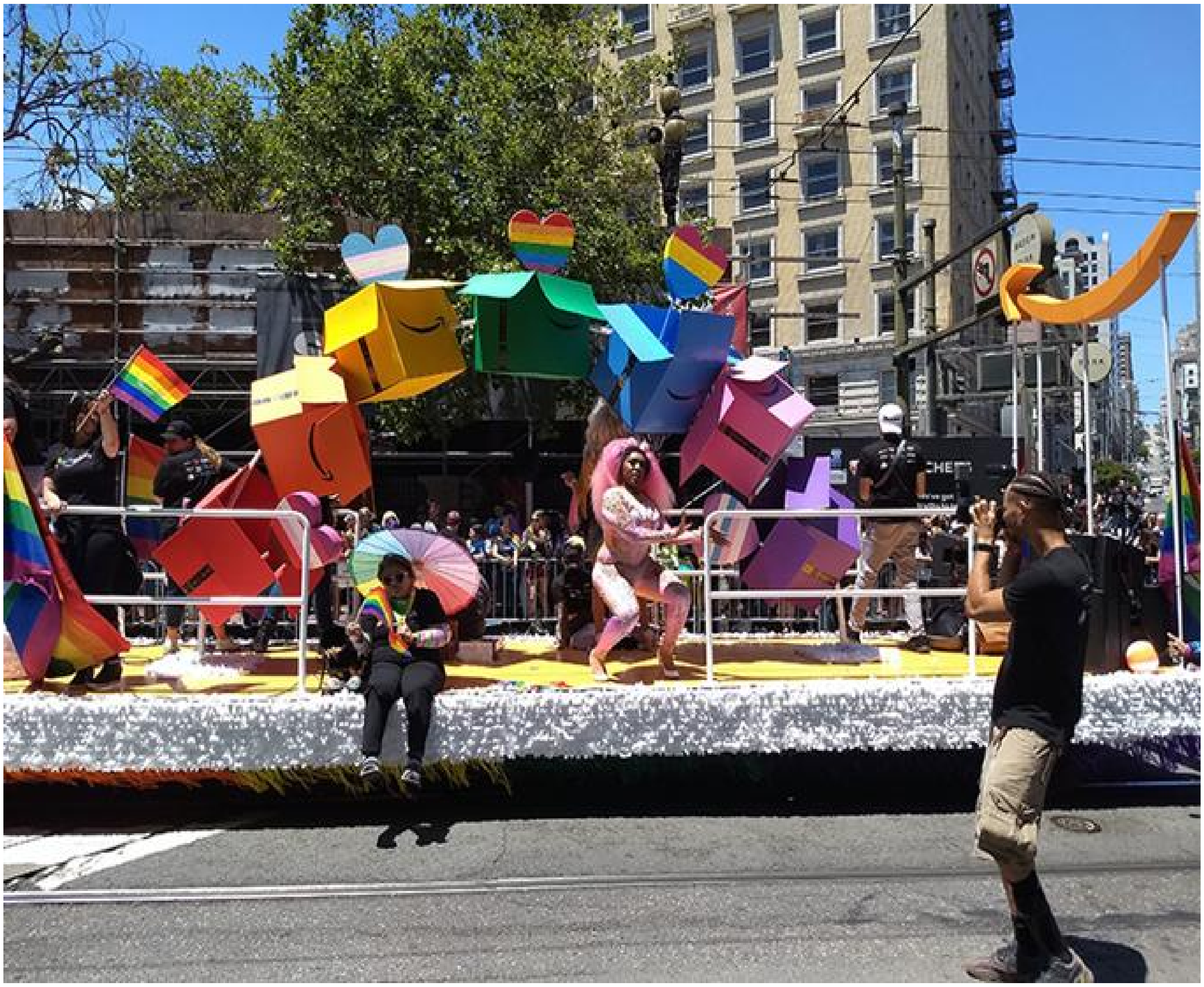
Actually, I take it back, a sightseeing bus displaying people's preferred pronouns might be the most SF thing.

Then came the march of the big corporations. Blue Shield, the health insurance company had a float; with impressive chutzpah, they had chosen the motto "Love Covers All". Their employees rushed ahead, distributing Blue Shield paddle/fan/advertisements for everyone to wave. The crowd of teenage girls standing next to me accepted them with gusto, waving their Blue Shield fan-paddles and cheering as the Blue Shield delegation passed by.

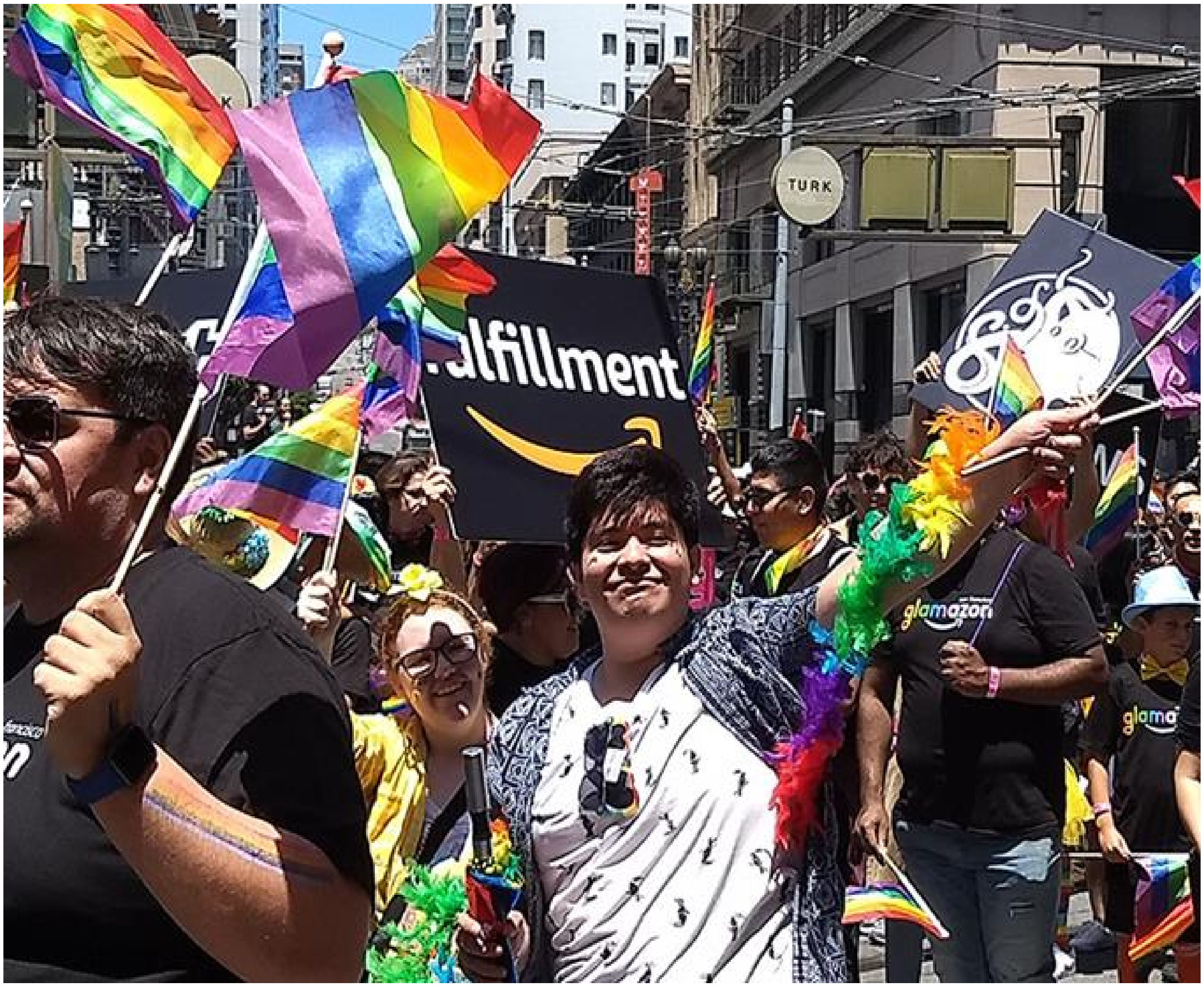
TURK
blue california
blue california
WHAT WE STAND FOR SETS US APART

blue
california
love covers all.

Apple, Facebook, Google, and Uber were all there. But the show was stolen by Amazon (temporarily rebranded “Glamazon”), who were going for a Santa Claus type image as a source of limitless cornucopian gifts.



The Amazon float featured a rainbow of colorful packages


TURK
Fulfillment
glamazon
glama

Get it? “Fulfillment?”

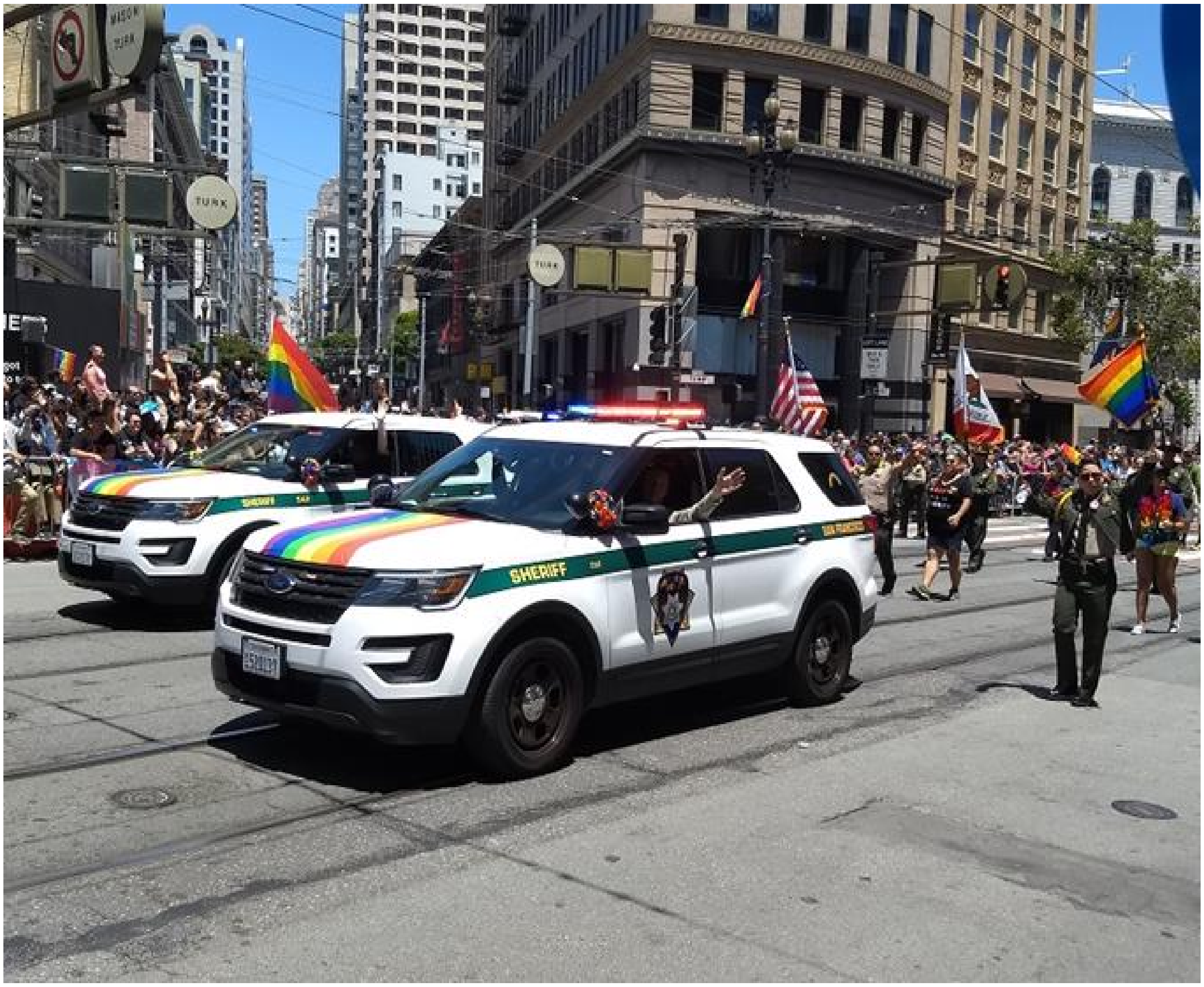
TREASURE TRUCK
amazon
SKYWATCHERS
HELLO

The Amazon Treasure Truck, ready to bring all your wildest dreams

I don’t know when I realized it was a sublimated Fourth of July Parade. But once I figured it out, it wasn’t subtle – and not just because it was being held the weekend before July 4th. The police cars with red-white-and-blue stripes had been replaced by police cars with rainbow stripes. The civic dignitaries waving American flags had been replaced by civic dignitaries waving gay flags. Even the Boy Scouts were still there, in the same place as always.

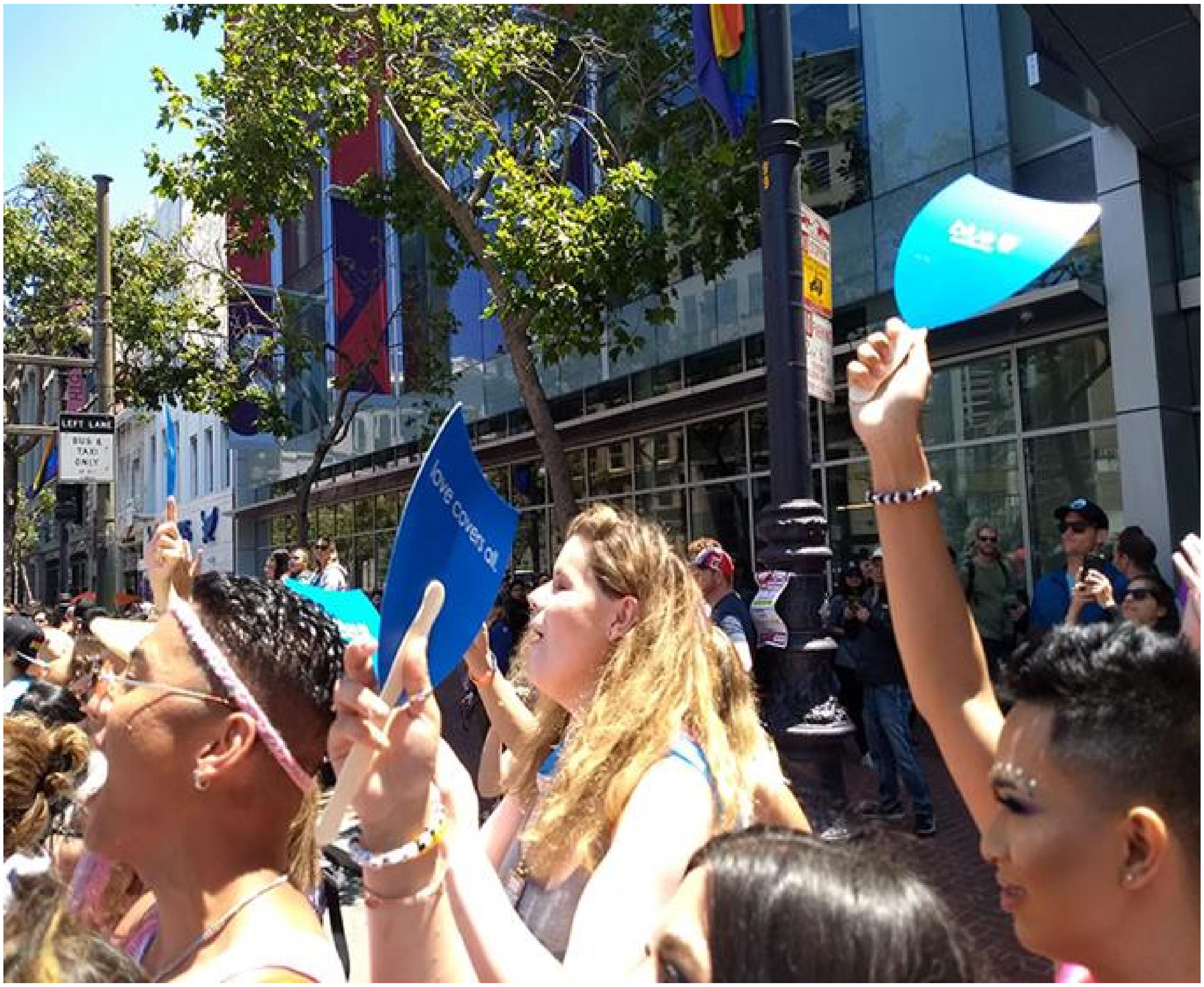
POLICING
WITH
PRIDE
POLICE
S.F.P.D.
PRIDE
EQUALITY
PEACE

Police cars in full regalia


MASON
TURK
TURK
TURK
SHERIFF
SHERIFF

Apparently the sheriff's department is different from the police department. You learn something new every day.

Some third group of basically cop-like people, not sure what’s going on here.


love covers all
LEFT LANE
BUS & TAXI ONLY

As a gaggle of teenage girls waved their Blue-Shield-advertisement-paddles to cheer on the police, I thought to myself “Yes, this exactly captures the spirit of the original Stonewall rioters”.

NO
TURNS
Montgomery
GAVIN
GAVIN
GAVIN
PrideSF
BUD LIGHT

The Bud Light float should be your cue that this is less about gayness and more about generic summer holiday Americana

A SCOUT IS EQUAL
SCOUTSFOR
THE MOVEMENT TO SAVE SCOUTING FROM THE WRONG SIDE OF

You can tell something's still hip and countercultural and definitely hasn't sold out when the Boy Scouts get involved

love
hate
GAPA

No, stop with the obvious symbolism! I'm trying to pretend that I'm very insightful for noticing this! Stop it right now!

Am I saying that gay pride has replaced the American civil religion?

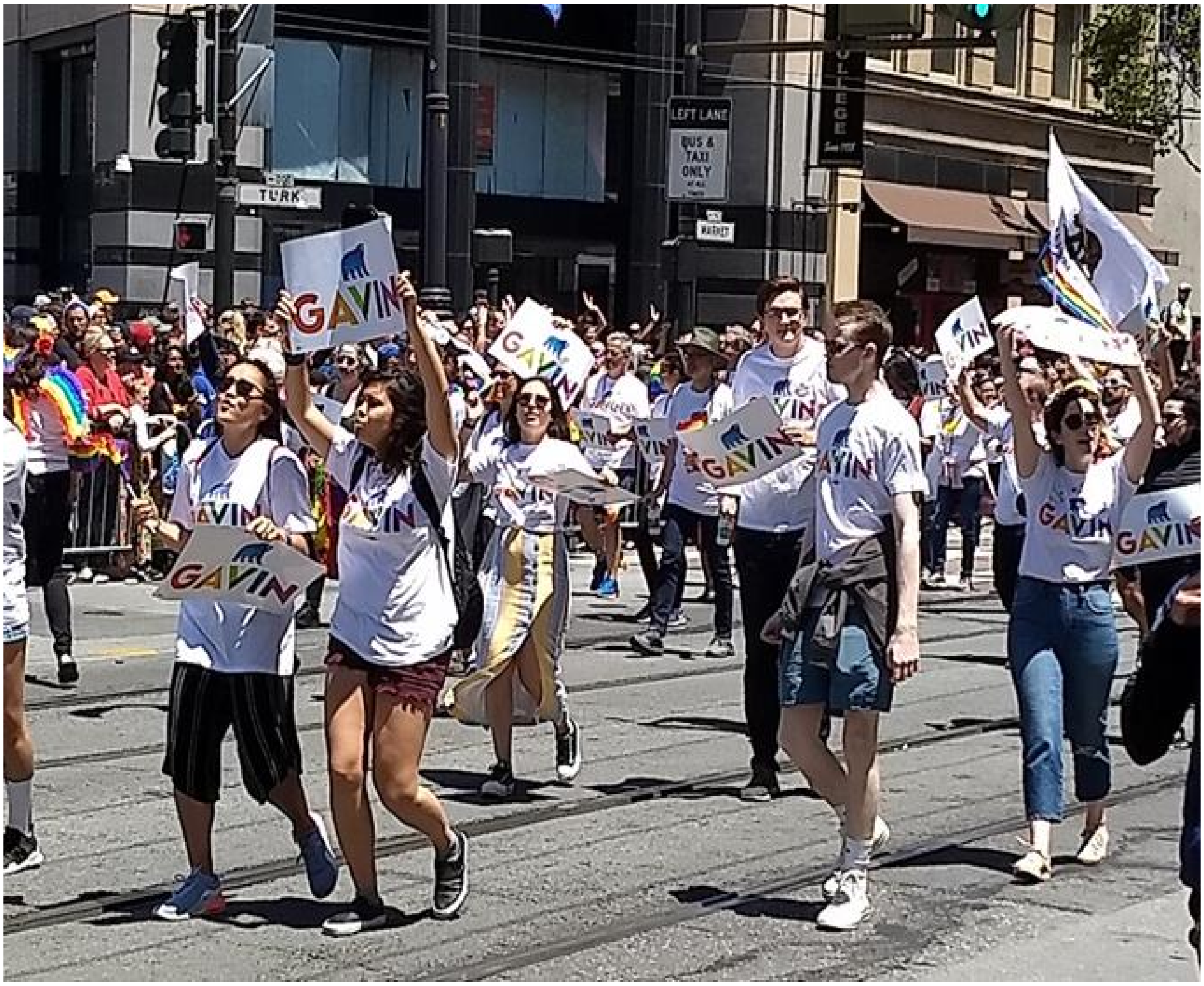
Maybe not just because it had a cool parade. But put it in the context of everything else going on, and it seems plausible. "Social justice is a religion" is hardly a novel take. A thousand tradcon articles make the same case. But a lot of them use an impoverished definition of religion, something like "false belief that stupid people hold on faith, turning them into hateful fanatics" – which is a weird mistake for tradcons to make.

There’s another aspect of religion. The one that inspired the Guatemala Easter parade. The group-building aspect. The one that answers the questions inherent in any group more tightly bound than atomic individuals acting in their self-interest:

What is our group? We’re the people who believe in pride and equality and diversity and love always winning.

Why is our group better than other groups? Because those other groups are bigots who are motivated by hate.

What gives our social system legitimacy? Because all those beautiful people in fancy cars, Governor Gavin Newsom and Mayor London Breed and all the rest, are fighting for equality and trying to dismantle racism.


TURK
LEFT LANE
BUS & TAXI ONLY
MARKET
GAVIN

Pictured: a religious festival successfully granting legitimacy to the secular power

LEFT LANE
BUS & TAXI ONLY
MARKET
KAMALA
HUMAN

More support for the secular power, in this case California senator Kamala Harris.

TURK
BERNIE
FEEL THE BE
Bernie
BernieSanders.com
Bernie

Still more support for the secular power.

ernie
sident 2020
VOTE 4
DADDY BERNIE
BITCH!
rnie
for SAL

The secular power is starting to get kind of creeped out, and wants to clarify that it only likes you as a friend.

IV.

“Civil religion” is a surprising place for social justice to end up. Gay pride started at Stonewall as a giant fuck-you to civil society. Homeless people, addicts, and sex workers told the police where they could shove their respectable values.

But there was another major world religion that started with beggars, lepers, and prostitutes, wasn’t there? One that told the Pharisees where to shove their respectable values. One whose founder got in trouble with the cops of his time. One that told its followers to leave their families, quit their jobs, give away all their possessions, and welcome execution at the hands of the secular authorities.

DEJANAY STANTON 24
TRANS RIGHTS ARE HUMAN RIGHTS
LIVESERIE.COM

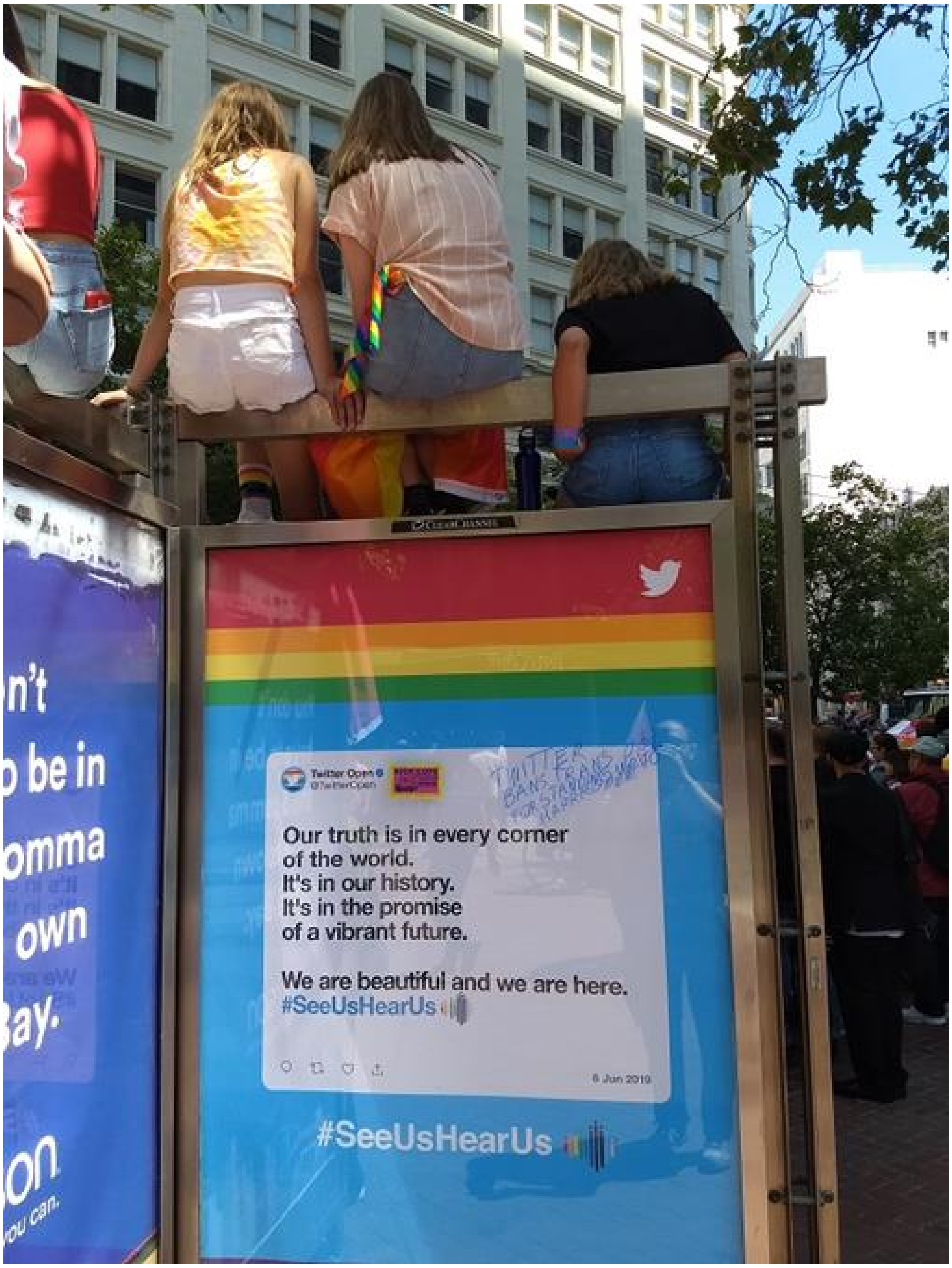
We're not a religion, we just parade images of martyrs up and down the streets.

The new faith burst into a world dominated by the *religio Romana*, the civil religion *par excellence*. Emperor Augustus had just finished moral reforms promoting all the best values: chastity, family, tradition, patriotism, martial valor.

Lavishly dressed procurators and proconsuls were building beautiful marble temples across the known world, spreading the rites with all the pomp and dignity befitting history’s greatest empire.

The problem was, nobody really believed *religio Romana* anymore. Everyone believed it was important to have all the best values, like chastity and military valor and so on. But nobody took Jupiter very seriously, or thought the Emperor was legitimate in some kind of sacred way.

When the new religion of beggars and lepers encountered the old religion of emperors and philosophers, the latter crumbled. But as Christianity expanded to the upper classes, it started looking, well, upper-class. It started promoting all the best values. Chastity, family, tradition, patriotism, martial valor. You knew the Pope was a good Christian because he lived in a giant palace and wore a golden tiara. Nobody ever came out and said Jesus was wrong to love prostitutes, but Pope Sixtus V did pass a law instituting the death penalty for prostitution, in Jesus’ name. Nobody ever came out and said Jesus was wrong to preach peace, but they did fight an awful lot of holy wars.


Twitter Open
@TwitterOpen
TWITTER BANS TRANS
Our truth is in every corner of the world.
It's in our history.
It's in the promise of a vibrant future.
We are beautiful and we are here.
#SeeUsHearUs
6 Jun 2019
#SeeUsHearUs
n't
be in
omma
own
ay.
on
you can.

We're not a religion, we just want to spread our truth to every corner of the world

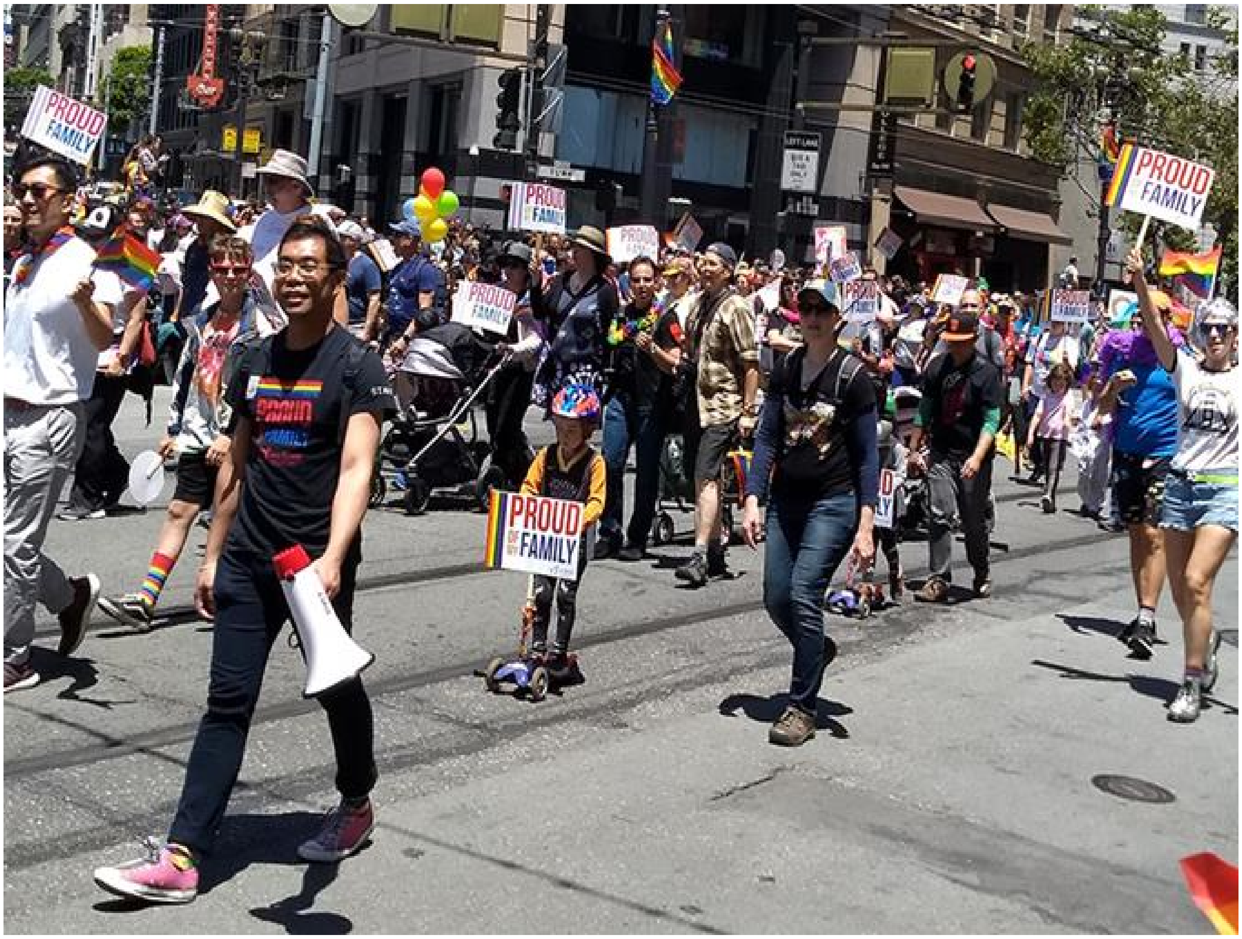
At some point it got kind of ridiculous. I don't know how much clearer Jesus could have been about "rich = bad", but the prosperity gospel – the belief that material wealth is a sign of God's favor – is definitely a thing. The moral of the story is: religion adapts to the demands placed on it. If it becomes a civil religion, it will contort itself until it looks like a civil religion. It will have all the best values.

Everything happens faster these days. It took Christianity three hundred years to go from Christ to Constantine. It only took fifty for gay pride to go from the Stonewall riots to rainbow-colored gay bracelets urging you to support your local sheriff deparment.

No, I’m not making that up

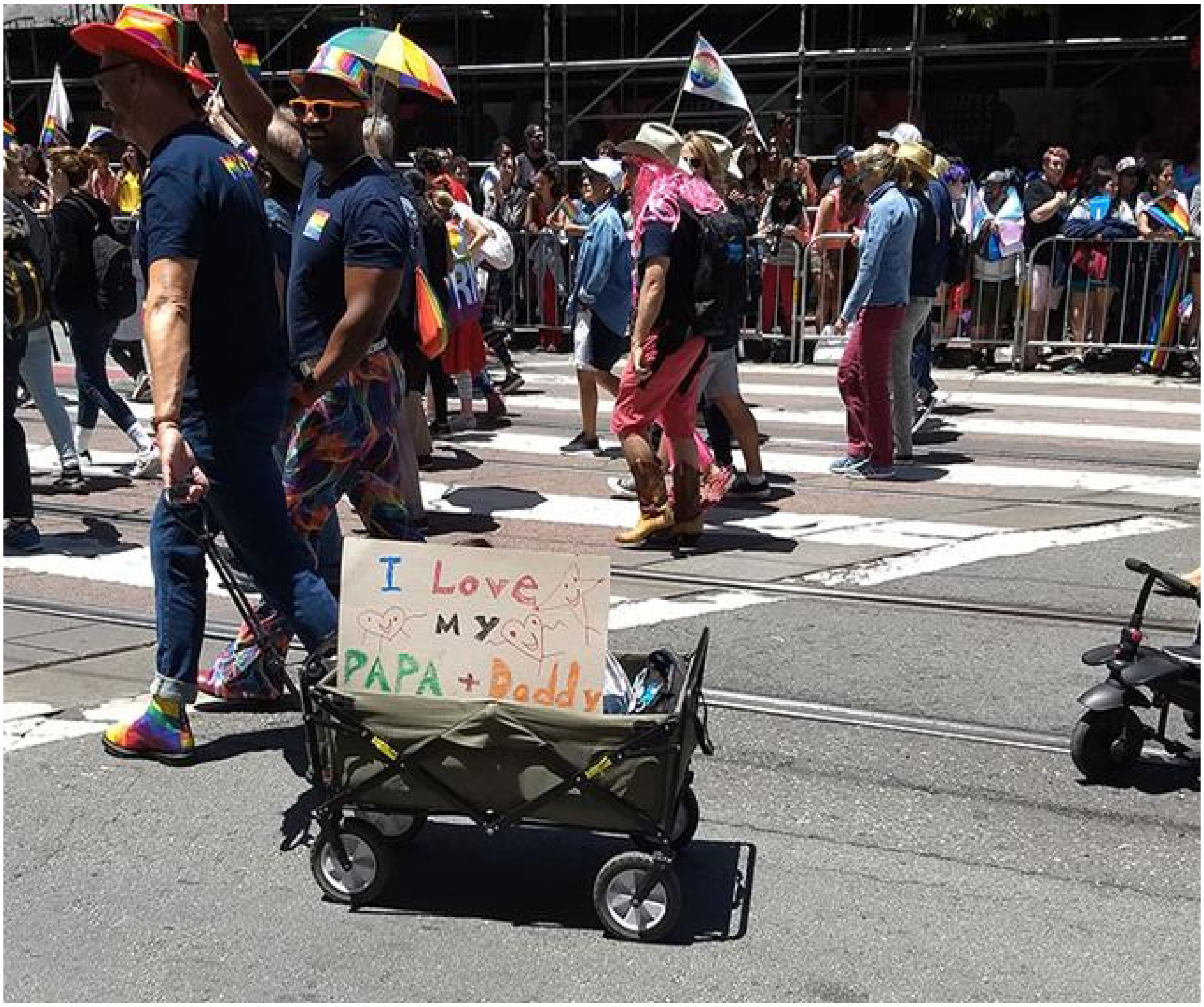
I can hear my conservative readers getting apoplectic: what about families? Family values are the most important legitimizing, community-building, wisdom-encoding part of Christianity! Homosexuality is anti-family and therefore can never be a true civil religion. Sure, you can twist social justice into support for sufficiently progressive government officials, but fifty years of cultural evolution isn't going to make it into a pro-family movement, right?

About that:


PROUD FAMILY
PROUD OF MY FAMILY
PROUD FAMILY

TURK
LEFT LANE
BUS & TAXI ONLY AT ALL TIMES
COLLEGE
PROUD OF MY FAMILY
PROUD OF MY FAMILY
PROUD OF MY FAMILY
our family coalition
BISEXUAL & TRANSGENDER FAMILY ORGA
415-981-196
family.org

San Francisco Day School
Celebrates ALL Families



Yeah, sure, all of this has context. “Proud Of Our Families” is supposed to be about people not being ashamed of their non-standard family structures (eg two fathers), although realistically I saw a lot of pretty hetero-looking families marching along. “We Celebrate All Families” is supposed to mean “including families with trans and gay people”. My point isn’t that everyone has suddenly forgotten about homosexuality, my point is that the celebration of gay pride is expressing itself in very predictable ways, after only fifty years. Christianity will mumble something about loving your parents as we love God our Heavenly Father, but that doesn’t mean that its family values are fake, or just a fig leaf for theology. It means cultural evolution works with what it’s got.

In a hundred years, will social justice look exactly like Christianity does now? No. The world’s changed too much. Even if every religion converges on the same set of socially useful values, the socially useful values change. We don’t need to push chastity if we have good STD treatment and contraception; we don’t need to push martial valor if all our wars are fought by drones. The old religions are failing partly because they can’t adapt quickly enough; social justice won’t need to imitate their failures. And Christianity is far from a homogenous mass; it has everything from golden-tiara-ed monarchs to barefoot street preachers to corporate megachurches to tonsured priests.

Levi's
A Guy In Levi's®
A Guy In Levi's®
Them In Levi's®
A Girl In Levi's®

We’re not a religion…but we do have Levites!

But I expect it to recapitulate the history of other civil religions in fast-forward. Did you know “pagan” is just Latin for “rural”? The pagans, the people who kept resisting Christianity even after it had conquered the centers of power, were the Roman equivalent of flyover states. Once Pride assimilates its own pagans (and kicks out its own Julian the Apostate), maybe it mellows out. Maybe it becomes more tolerant, the same way Christians eventually started painting Greek gods on everything. Maybe it encounters the same problems other faiths encountered and adapts to them the same way.

Maybe a decade or a century from now, we have all the best values.

"Though the cause of evil prosper; yet 'tis Truth alone is strong
And albeit she wander outcast now, I see around her throng
Troops of beautiful tall angels, to enshield her from all wrong."

Survey Results: Sexual Roles

I already started analyzing the SSC survey data on fetishes, but I wanted to move on to look at dominance, submission, sadism, and masochism.

Why might this be interesting? For one thing, some people have fetishes for things that seem, well…bad. Getting hurt. Letting other people control and abuse them. As if they have a drive toward weakness and unhappiness. This is kind of reminiscent of the self-sabotage and bad decisions some people make throughout their lives (for example, marrying a spouse who treats them the same way as an abusive parent). Sometimes I conceptualize this as them having a set point of low self-esteem and degradation that they try to enforce, regardless of its cost to their well-being. If this had the same roots as sexual masochism, that would be worth studying.

But I didn’t find anything interesting like that in the data.

BDSM preferences were heavily gendered. Of people who expressed a preference, 71% of cis men preferred the dominant role, compared to only 16% of cis women (18% of trans women; insufficient sample size of trans men). This was such a big difference that gender swamped every other effect, so I limited the analysis to cis men from this point on, since they made up most of the sample.

80% of straight men preferred the dominant role, compared to only 34% of gay men. This was such a big difference that orientation swamped every other effect, so I limited the analysis to straight cis men from this point on.

In order of importance, here are some factors that made the men in this sample more likely to be dominant, rather than submissive. All of these are self-rated:

- Good social skills
- High risk taking
- High ambition
- Conservative political values
- Low anxiety
- High status
- Belief that others are trustworthy

Forget the lurid stories about high-powered executives who secretly want to be abused and degraded. Most of these are pretty straightforward. The more successful and less neurotic a man's personality, the more likely he is to be sexually dominant. But none of these are big effects. For example, conservatives are more dominant than liberals, but the exact numbers are 83% vs. 70%. Straight men are still mostly dominant, regardless of politics.

Submissive men reported lower sex drive, fewer dates (9 vs. 18), fewer sexual partners (5 vs. 11), and fewer long-term relationships (2.7 vs. 3.5) than dominant men. This probably has to do with the worse social skills and decreased risk taking.

This was so boring that I tried switching to sadism vs. masochism. Most self-reported sadists were also self-reported masochists, so I took only the subset of people who reported one but not the other. This showed similar patterns to dominance vs. submission, so much so that it's not worth going over them separately.

Two mildly interesting findings. First, although psychiatric issues in general only affected these roles weakly and inconsistently, men with OCD were four times less likely to be sadists (and somewhat more likely to be masochists) than anyone else. This seems like an extreme form of the effect of high anxiety, plus maybe an obsessive fear of hurting someone else, or a response to feelings of guilt.

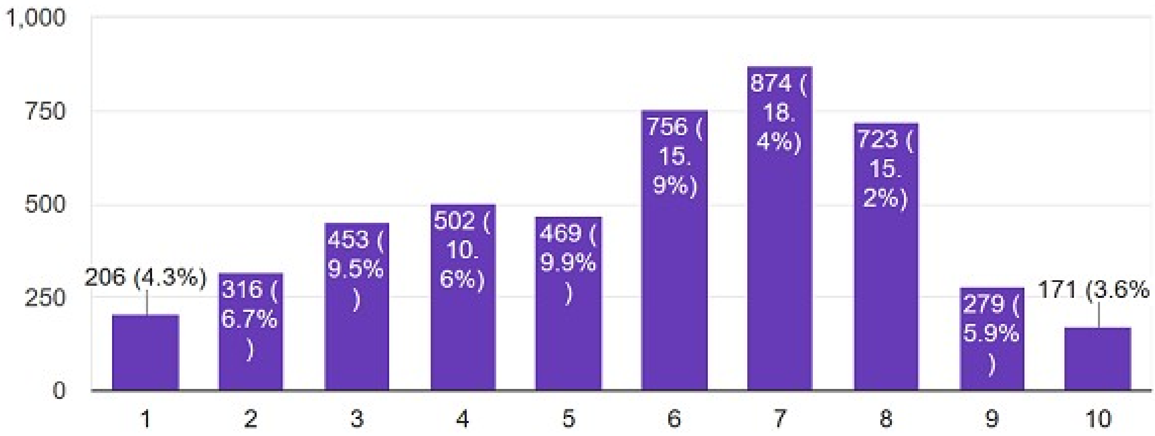
Second, although there was no effect of self-reported childhood trauma, men who grew up poor reported about twice the sexual sadism rates of people who grew up rich or middle-class; whether or not you were currently poor mattered a little less. Given how many comparisons I did, I’m not very confident in this result even though the effect size is pretty big.

The data didn’t support any kind of connection between dominance/submission/sadism/masochism and more prosaic forms of selfishness or self-sacrifice. Men were about equally likely to give money to charity, or identify as effective altruists, or hold various opinions in moral philosophy, regardless of their sexual roles.

Although I’ve made every other part of the survey publicly available, given the sensitivity of fetishes I’m keeping these particular answers private. If you are a professional researcher (or an amateur researcher with a good track record of professionalism and data integrity), and you want to test these results, please email me at scott[at]slatestarcodex[dot]com and we can discuss how to make that happen.

Do People Like Their Mental Health Care?

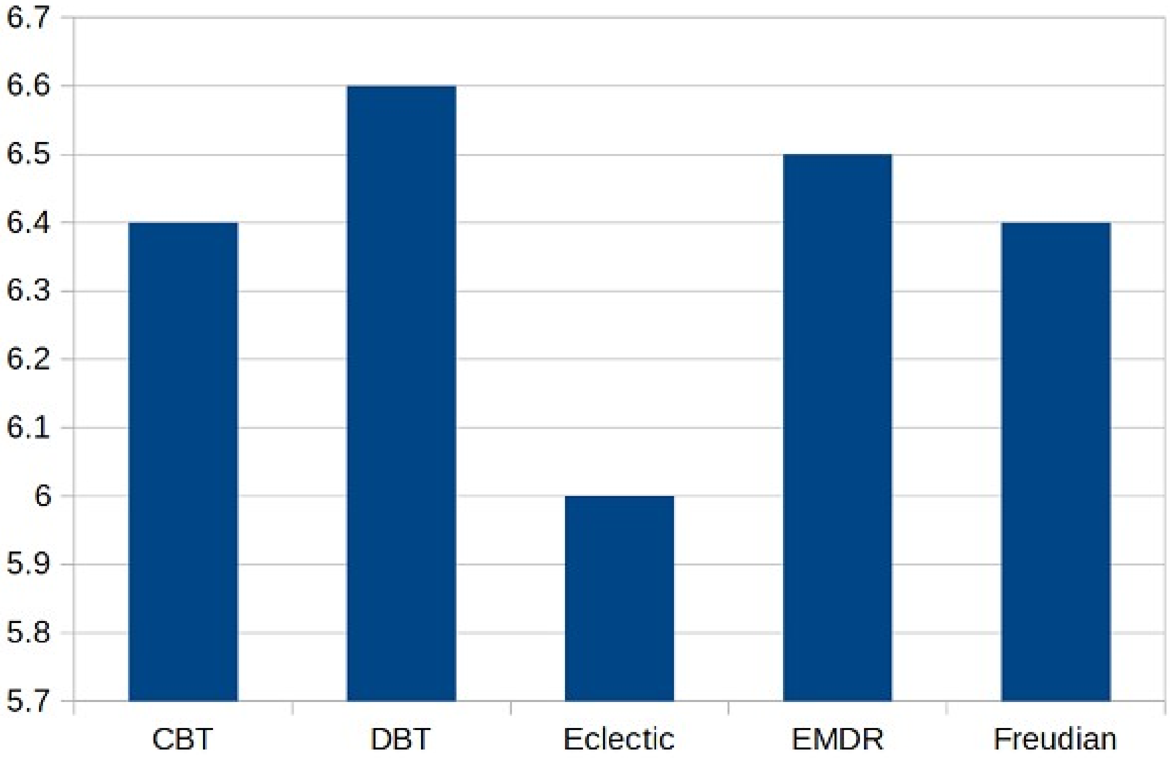
Along with more specific questions, I asked people who took [the SSC survey]() to rate their experience with the mental health system on a 1 – 10 scale.



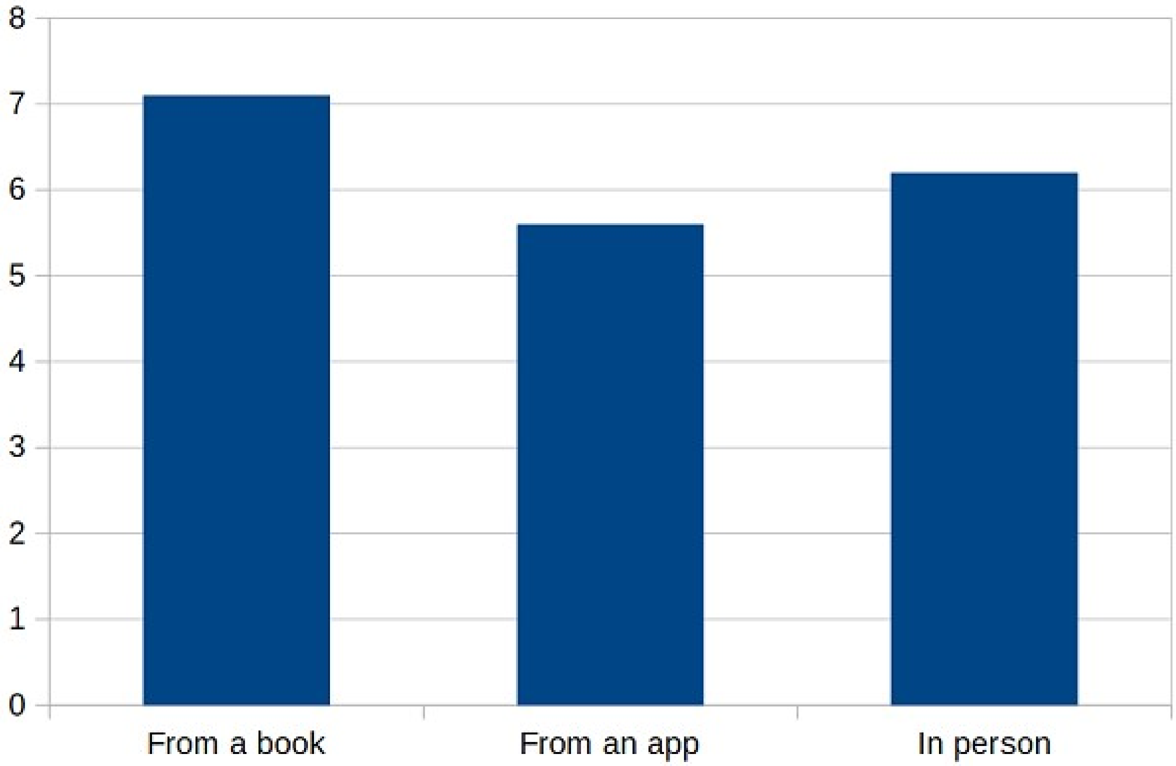
About 5,000 people answered. On average, they rated their experience with psychotherapy a 5.7, and their experience with medication also 5.7.

This is more optimistic than a lot of the horror stories you hear would suggest. A lot of the horror stories involve inpatient commitment (which did get a dismal 4.4/10 approval rating) so I checked what percent of people engaging with the system ended up inpatient. Of people who had seen either a psychiatrist or therapist, only 7% had ever been involuntarily committed to a psychiatric hospital. Note that this data can't tease out causation, so this doesn't mean 7% of people who saw an outpatient professional were later committed – it might just mean that lots of people got committed to the hospital by police, then saw a professional later.

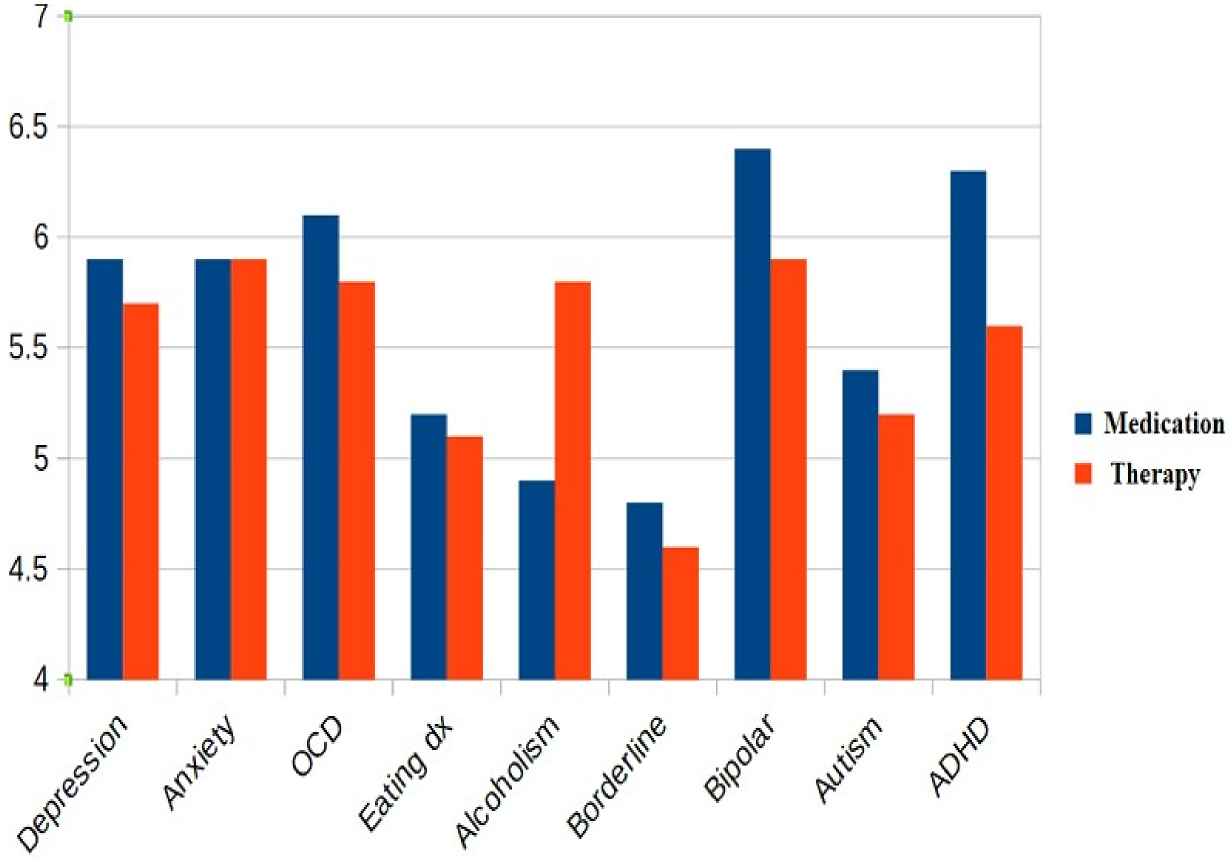
Going into more detail about what people did or didn't like (note truncated y-axis):



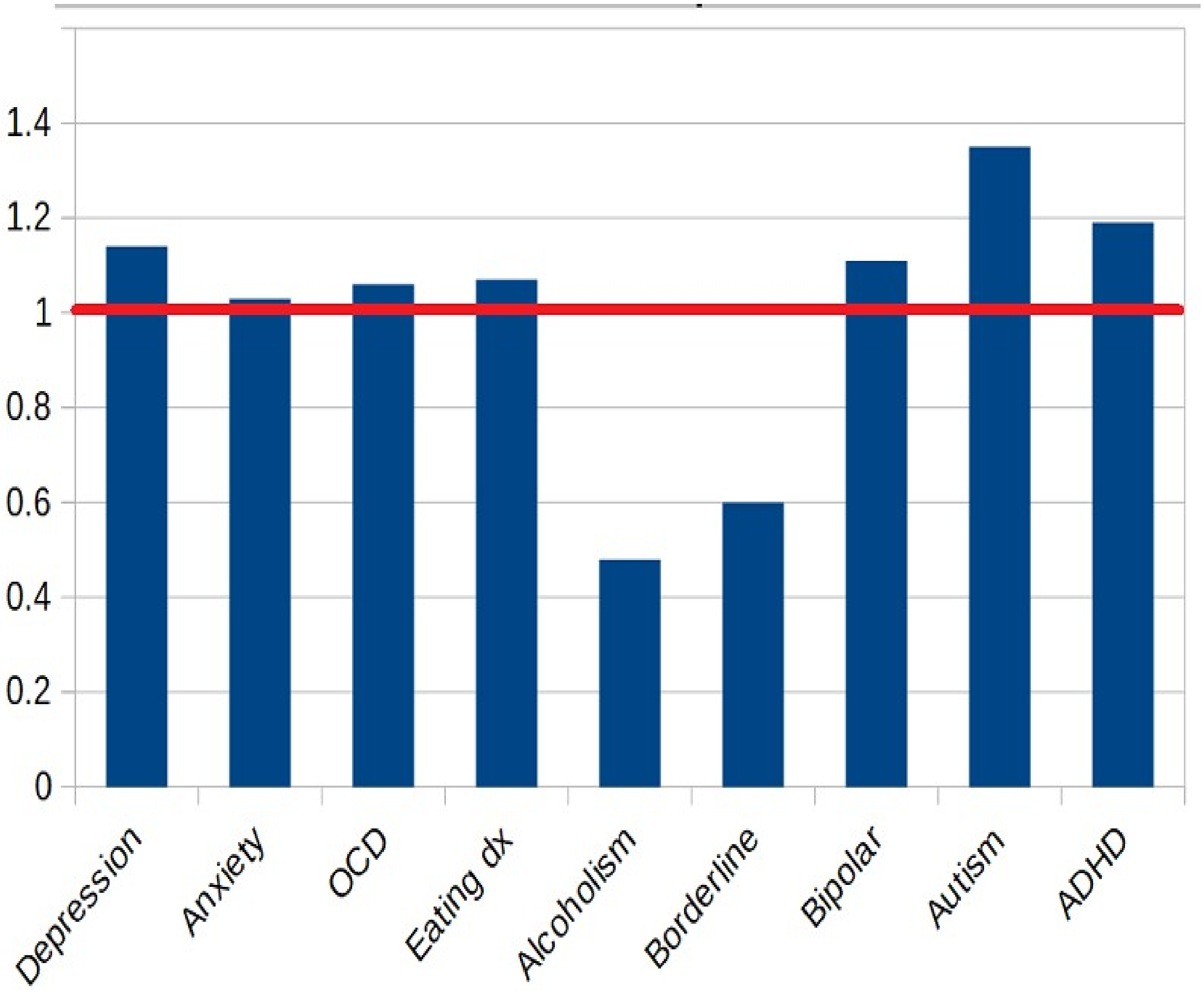
I asked people what kind of therapy they did. People liked all schools of therapy about the same, except that they liked "eclectic" therapy that wasn't part of any specific school less than any school. Every school including eclectic got higher than 5.7, because people who wouldn't answer this question – who weren't even sure what kind of therapy they were doing – rated it less than any school *or* than eclectic therapy.



People really liked doing therapy from a book. They liked doing therapy with an in-person therapist a little less, and they liked online therapy apps least of all. This doesn't match published literature, and this would be a good time to remember that all of these results are horribly biased and none of them can prove causation. For example, the sort of motivated go-getter who would go out and get a therapy book and read it themselves might be systematically different from somebody who gets therapy through an app or in a clinic.



This graph shows how people liked medication (blue) and therapy (red) based on what their mental health issue was (note truncated y-axis). Some groups – people with eating disorders, people with borderline personalities – were just generally hard to please. Alcoholics were much happier with their therapy than with pharmaceutical treatment (though the sample size was only about 50 per group). People with bipolar and ADHD were happier with medication than therapy.



This is a little different. The last graph averaged everyone's opinion of medication and everyone's opinion of therapy. This one just includes the people who have tried both, who might have a better standard for comparison. The higher the bar above the red line, the more they preferred the medication; the lower the bar below the red line, the more they preferred the therapy.

Alcoholics and borderlines prefer the therapy. Autistic people strongly prefer the medication, which is weird because there's no good medication for autism. This could be them hating the social interaction involved in therapy. Or it could be a condemnation of therapies like applied behavior analysis, which can become a sort of confrontational attempt to force them to conform, with potential punishment for failure. The ADHD preference for medication is less surprising; stimulants always get a high approval rating.

Remember, none of these numbers measure whether treatment works – just whether patients are happy. And they’re all vulnerable to selection effects and a host of other biases. Take them as exploratory only. I welcome people trying to replicate or expand on these results. All of the data used in this post are freely available and can be downloaded here.

Against Lie Inflation

[Related to: The Whole City Is Center]

I.

I got into an argument recently with somebody who used the word “lie” to refer to a person honestly reporting their unconsciously biased beliefs – her example was a tech entrepreneur so caught up in an atmosphere of hype that he makes absurdly optimistic predictions. I promised a post explaining why I don’t like that use of “lie”. This is that post.

A few months ago, a friend confessed that she had abused her boyfriend. I was shocked, because this friend is one of the kindest and gentlest people I know. I probed for details. She told me that sometimes she needed her boyfriend to do some favor for her, and he wouldn’t, so she would cry – not as an attempt to manipulate him, just because she was sad. She counted this as abuse, because her definition of “abuse” is “something that makes your partner feel bad about setting boundaries”. And when she cried, that made her boyfriend feel guilty about his boundary that he wasn’t going to do the favor.

We argued for a while about whether this was a good definition of abuse (it isn’t). But I had a bigger objection: this definition was so broad that *everyone* has committed abuse at some point.

My friend could have countered that this was a feature, not a bug. Standards have been (and should be) getting stricter. A thousand years ago, beating your wife wasn’t considered abuse as long as you didn’t maim her or something. A hundred years ago, you could bully and belittle someone all you wanted, but as long as there was no physical violence it wasn’t abuse. As society gets better and better at dealing with these issues, the definition of abuse gets broader. Maybe we *should* end up with a definition where basically everyone is an abuser.

But a wise supervillain once said, “When everyone is super, nobody is”. In the same way, when everyone is an abuser, nobody’s an abuser.

Right now, if I hear that someone is an serial abuser, I would be less likely to date them, or I might warn my friends away from them, or I might try not to support them socially. The world is divided into distinct categories – abuser and non-abuser – and which category someone is in gives you useful information about that person’s character. I’m not saying that every abuser is an awful person who is 100% defined by their misdeeds and can never be redeemed. But I think the category contains useful information about a person’s character and likely future actions.

But if everyone used my friend’s definition, and we acknowledged that everybody is an abuser – the category stops being informative. “John is an abuser”. So what? Doesn’t mean you should worry about John, doesn’t mean you shouldn’t date John, doesn’t even mean you shouldn’t set your single friends up on blind dates with John. It just means John is a human. Maybe he cries sometimes. So what?

Broadening the definition of “abuser” this far doesn’t help fight abuse or make anybody nicer. It just removes a useful word from the English language. I can still eventually warn someone that John is cruel or violent toward people close to him. I just have to circumlocute around the word “abuser”, in order to find some other word or phrase that hasn’t been rendered meaningless.

(I’m cheating here by talking about “abusers” rather than “abuse”, since there is still a useful distinction between abuse and non-abuse actions. But although the abuse case is less clear, I think some of the same considerations apply – just because an action is abuse no longer means you can be sure it’s especially bad)

But it’s worse than this, because change to a definition doesn’t instantaneously propagate to all of its web of connotations in our minds. So probably some people will continue to use the new definition while still holding the connotations of the old definition. This means bad actors can stigmatize anyone they want:

1. We don’t tolerate abusers around here, right? Right!
2. John’s actions technically qualify as abuse by this incredibly broad standard that includes basically everyone.
3. Therefore we shouldn’t tolerate John.

I previously called this manuever The Worst Argument In The World. Maybe that’s a slight exaggeration, I’m just really tired of seeing it again and again.

II.

This is also my objection to broadening the meaning of “lie”.

The word “lie” is useful because some statements are lies and others aren’t. And although people may disagree on which statements are lies or not (Did OJ lie when he said he was innocent? opinions differ!) everyone agrees on a mapping between states-of-the-world and lie-vs-truth status. When I say “OJ lied”, everyone understands me as making a specific claim about the world, which they can either accept or reject. I don’t have a lot of leeway in how I use the word “lie”; if I’m calling you a liar, I’m making a specific claim about the world.

If “lie” expands to include biased or motivated reasoning, who’s going to throw the first stone? We’re probably all biased to some degree. Does that make us all liars? If everyone’s a liar, nobody is. I can accuse Donald Trump of lying constantly, and you can just nod your head and say “Oh, so you’re saying he’s not a perfect person free from all bias, whatever”. You’ll feel no need to decrease your opinion of him.

Maybe we should only apply the word “lie” to particularly egregious bias and motivated reasoning? But we’ve already abandoned the only defensible Schelling fence. So how will people decide where to draw the line? My guess is: in a place drawn by bias and motivated reasoning, same way they decide everything else. The outgroup will be lying liars, and the ingroup will be decent people with ordinary human failings.

This is my criticism of the original post that started this argument. It declared belief in a near-term singularity to be a “scam” and that believers were “being duped into believing a lie”. Its evidence was listing some reasons people might be biased to believe the singularity was near.

The obvious next step is that someone who believes the singularity is near writes a post listing some biases that singularity skeptics probably hold (for example, the absurdity fallacy). Having shown that skeptics are biased, they pronounce skeptics to be liars perpetrating a massive fraud on the scientific community.

There are a few ways this expanded-definition world becomes different from the world where people restricted “lie” to mean a knowingly false statement.

First, everyone is much angrier. In the restricted-definition world, a few people write posts suggesting that there may be biases affecting the situation. In the expanded-definition world, those same people write posts accusing the other side of being liars perpetrating a fraud. I am willing to listen to people suggesting I might be biased, but if someone calls me a liar I'm going to be pretty angry and go into defensive mode. I'll be less likely to hear them out and adjust my beliefs, and more likely to try to attack them.

Second, bad actors can use The Worst Argument In The World to prove whatever they want. As long as you're willing to equivocate and deceive people, you can prove anyone a liar, and then draw on now-obsolete connotations of "liar" to silence or ostracize them.

Third, the biggest beneficiaries are *actual liars*. Suppose some singularitarian claims that internal Google documents prove they have already created human-level AI. And suppose that's totally false and no such documents exist. Usually I would accuse them of lying, and this accusation would be enough to alert people that, hey, something has gone terribly wrong here. But if both sides are constantly accusing each other of lying just for having normal human failings, then "you are a liar" no longer carries much weight. I have to come up with some complex circumlocution in order to let people know that someone told a mistruth.

And that complex circumlocution can only last until people realize it too is an exploitable signal. The whole reason that rebranding lesser sins as "lying" is tempting is because everyone knows "lying" refers to something very bad. But the whole reason everyone knows "lying" refers to something very bad is because nobody has yet succeeded in rebranding it to mean lesser sins. The rebranding of lying is basically a parasitic process, exploiting the trust we have in a functioning piece of language until it’s lost all meaning – after which the parasitism will have to move on to whatever other trusted functional piece of language has sprung up to replace it.

III.

I realize this is a kind of long post arguing against a weird thing that not many people are doing. But I think it’s an especially clear case of a broader thing that many people are doing. Words like “disabled”, “queer”, and “autistic” are also gradually shifting meanings, getting applied more and more loosely.

This isn’t always bad! Words are useful because as they separate the world into categories; this suggests a word should apply in more than 0% of cases but less than 100% of cases. Exactly where it should fall in between that range, I don’t know. The broader you make the definition, the better the word’s ability to name things that have even a small level of the relevant quality. But the broader you make the definition, the less power the word will have to separate strong examples of a quality from marginal examples.

I think of this as a sort of sensitivity-and-specificity statistics problem, setting a threshold to divide the population into two groups. If you have a very strict threshold for “abuser”, maybe only someone who inflicts serious physical injuries, then you can use it to separate the most abusive 1% of people from the other 99%. If you have a very weak threshold for “abuser”, so low that 99% of people qualify, then you can use it to separate the 1% *least* abusive people from the other 99%. If you set it in the middle, you can separate the more abusive half of the population from the less abusive half. If “abuser” picks out the most abusive 1% of people, it transmits a lot of information in a small number of cases. If it picks out the most abusive 99% of people, it transmits very little information in a large number of cases (and now "not an abuser" transmits a large amount of information in a small number of cases!). If the boundary is set at 50%, it transmits an equal moderate amount of information about everyone.

There's no rule that 50-50 is always the best – for example, if the word "murderer" referred to anyone in the more murderous half of the population, that would be much worse than the system now, where it refers to a much smaller set of people, who you have much more reason to worry about as a discrete group. You're going to have to find the right threshold for each individual concept.

But it's never the right decision to draw the line *outside* the population, so that literally 100% of people fall in one category and 0% in the other.

A few months ago I told a fable about a city. The citizens worried that people living in the outskirts of the city felt unimportant and excluded. So they redefined "city center" to mean the entire city, including the outskirts. Nobody ended up feeling any more important because of this, because living in city center stops being prestigious when everywhere is city center. But it was now impossible to direct tourists to where they wanted to go, and people had to invent new phrases like "the part of the city where there are the most tall buildings" in order to discuss city center.

The moral of the story is: don't set thresholds for category membership so far outside a distribution that they stop conveying useful information.

Caution On Bias Arguments

> "You say it's important to overcome biases. So isn't it hypocritical that you're not trying to overcome whichever bias prevents you from realizing you're wrong and I'm right?"
> — everybody

Correcting for bias is important. Learning about specific biases, like confirmation bias or hindsight bias, can be helpful. But bias arguments – "People probably only believe X because of their bias, so we should ignore people who say X" tend to be unproductive and even toxic. Why?

1. Everyone Is Biased All The Time

You could accuse me of having a conservative bias. After all, I'm a well-off straight white man, a demographic well-known to lean conservative. If a liberal wanted to discount everything I say, or assume any conservative arguments I make come from self-serving motives, they've got all the ammunition they need.

Or you could accuse me of having a liberal bias. After all, I'm a college-educated atheist Jewish psychiatrist in the San Francisco Bay Area. All of those demographics are well-known to lean liberal. If a conservative wanted to discount everything I say, or assume any liberal arguments I make come from self-serving motives, they're not short on ammunition either.

This is a general phenomenon: for any issue, you can think of biases that could land people on one side or the other. People might be biased toward supporting moon colonization because of decades of sci-fi movies pushing space colonization as the wave of the future, or because Americans remember the moon landing as a great patriotic victory, or because big defense companies like Boeing will lobby for a project that would win them new contracts. Or people might be biased against moon colonization because of hidebound Luddite-ism, or an innate hominid preference for lush green forests and grasslands, or a pessimistic near-termism that rejects with payoffs more than a few years out. I personally might be biased towards moon colonization because I've been infected with the general Silicon Valley technophile mindset; or I personally might be biased against it because I'm a Democrat and Trump's been the loudest modern proponent of more moon missions.

This is even easier if you're allowed to invent biases on the spot. For example, I said people are against moon colonization because of "hidebound Luddite-ism" – is that actually a thing? If I say that regulatory action against tech companies is driven by anti-tech populism, have I identified a bias, made up a bias, or just tautologically rebranded people wanting regulation of tech companies as a force biasing people towards regulation of tech companies? Won't people who support regulation, counter by saying that opponents are just knee-jerk technophiles in who have drunk some sort of Silicon Valley hype Kool-Aid?

2. Everyone Is Hypersensitive To Biases Against Their Side, And Thinks Biases In Favor Of Their Side Are Irrelevant

This is called the hostile media effect, though it's broader than just the media. I've talked about it before in against bravery debates. My favorite example is conservatives complaining that the media condemns far-right terrorism but excuses Islamic terrorism (eg 1, 2, 3, 4, 5) alongside liberals complaining that the media condemns Islamic terrorism but excuses far-right terrorism (eg 1, 2, 3, 4, 5).

Or if you prefer facts to anecdotes: according to a Gallup poll, conservatives are more likely to believe the news has a liberal bias; liberals are more likely to believe the news has a conservative bias. In a study where experimenters showed partisans a trying-to-be-neutral video on the Israel-Palestine conflict, the pro-Israel people said the video was biased toward Palestine, and the pro-Palestine people said the video was biased towards Israel.

This ties into the problem where you can just make up a bias, like “hidebound Luddite-ism”. Technophiles will see an anti-tech bias everywhere. And whenever they meet a specific anti-tech person, they can assume that their positions have been shaped not by reason, but by the anti-tech sentiments that are omnipresent in our society. Having explained away their opponents’ position as the product of bias, they’ll feel no need to debate it or question whether it might be true.

Anyone can come up with any bias for any position, but this meta-bias is going to affect people’s sense of which biases matter and which ones don’t. Pro-moon-colonizers are going to doubt that technophilia is really a problem motivating people’s reasoning, but think that hidebound Luddite-ism is a big problem motivating everyone on the other side.

3. It’s Hard To Even Figure Out What Bias Means Or When It Is Bad

Suppose A and B are debating some issue, and B is part of a group especially closely linked to the issue. For example:

1. A plumber and a teacher are debating a proposed pay cut for teachers.

2. A man and a woman are debating abortion.
3. An atheist and a Jew are debating the peace process in Israel.
4. A white person and a black person are debating slavery reparations.
5. A citizen and an undocumented immigrant are debating immigration policy.
6. King Edward and a Jew are debating whether to expel all the Jews from England.
7. You and a KKK Grand Wizard are debating whether the KKK should be banned as a hate group.
8. A scientist and a tobacco company executive are debating whether cigarettes are dangerous.

Who is more biased? A or B?

This is a tough question. If we're just working off the dictionary definition of bias, it ought to be B. But in cases like 6, it would be pretty bad to adjust away from B's opinion, or discount B as too biased to give a good argument.

We can't dismiss this as "A is also affected by the issue". It's true that for example the plumber may lose a little money if he has to pay higher taxes to fund increased teacher salaries. But since there are fewer teachers than taxpayers, each taxpayer's loss is much smaller than each teacher's gain. It still seems like B should be more biased.

We could model this as two opposite considerations. A is less biased. But B may be better informed. Sometimes this is literal information: I'd expect an immigrant to know more about immigration policy than an average citizen. Other times it can be emotional "information" about how something feels; for example, a woman may have hard-to-communicate information about what makes abortion rights feel important to her.

Is it meaningful to say the Jew has hard-to-communicate information about how much he doesn't want to be kicked out of England? Or should we just say that, as the person most affected by the policy, he's more likely to be thinking about it clearly? But now we've come full circle to saying that motivated reasoning itself is good!

I have a hard time squaring this circle. The lesson I take is that it's easy to switch between "we should trust the more affected party less" and "we should trust the more affected party more" without clear principles to guide us.

Probably most people will do this in a biased way. When their side is the more affected party, they'll say that gives them special insight and so other people should back off. When they're the less affected party, they'll say that makes them unbiased and other people are just motivated reasoners. This is yet another reason to expect that bias arguments have so many degrees of freedom that everyone will figure their opponents are biased and they aren't.

4. Bias Arguments Have Nowhere To Go

Most people are already aware of their potential biases. No straight man will be surprised to be told that they are a straight man, or that this might bias them. "You are a straight man, so consider that you might be biased" doesn't give new information. It just deflects the conversation from potentially productive object-level discussion to a level which is likely to sound patronizing and overly personal, and which has less chance of being productive.

Someone asks me "Are you sure you don't just hold that opinion because of the liberal Jewish milieu you grew up in?"

I look deep into my brain, the opinion still sounds right, I don't see a sticker on the opinion saying "Proud product of the liberal Jewish milieu you grew up in", and…then what? Do I drop the opinion even though it still seems right? Do I keep holding the opinion, but feel guilty about it? Do I retort back "Aha, no, you only hold *your* opinion because of the conservative Gentile milieu *you* grew up in, so you should drop *your* opinion!"?

There's a sense in which we should always be considering the Outside View (see part III here) for each of the opinions we hold. That is, on the Inside View, the opinion might still seem convincing, but on the Outside View, we might have enough circumstantial evidence that it was produced by some process uncorrelated with truth that we doubt it despite its convincingness. But just learning that there's some *possible* bias should rarely have much of an effect on this process, especially since with any self-awareness we should probably have already priced all of our own biases in.

5. Where To Go From Here

I think low-effort (and even medium-effort) arguments from bias will usually be counterproductive. Second person bias arguments ("You are probably biased on this topic because X") and third-person bias arguments ("Society is probably biased on this topic because X") are at least as likely to perpetuate biases as to help overcome them, and less useful than just focusing on the object-level argument.

What's left? Bias is an important obstacle to truth-seeking; do we just ignore it? I think bias arguments can be useful in a few cases.

First, it's fair to point out a bias if this gives someone surprising new information. For example, if I say "The study proving Panexa works was done by the company producing Panexa”, that might surprise the other person in a way that “You are a straight man” wouldn’t. It carries factual information in a way that “You’re a product of a society laden with anti-tech populism” doesn’t.

Second, it’s fair to point out a bias if you can quantify it. For example, if 90% of social scientists are registered Democrats, that gets beyond the whole “I can name one bias predisposing scientists to be more liberal, you can name one bias predisposing scientists to be more conservative” arms race. Or if you did some kind of study, and X% of social scientists said something like “I feel uncomfortable expressing conservative views in my institution”, I think that’s fair to mention.

Third, it’s fair to point out a bias if there’s some unbiased alternative. If you argue I should stop trusting economists because “they’re naturally all biased towards capitalism”, I don’t know what to tell you, but if you argue I should stop trusting studies done by pharmaceutical companies, in favor of studies done by non-pharma-linked research labs, that’s a nice actionable suggestion. Sometimes this requires some kind of position on the A vs. B questions mentioned above: is a non-Jew a less biased source for Israel opinions than a Jew? Tough question.

Fourth, none of this should apply in private conversations between two people who trust each other. If well-intentioned smart friend *who understands all the points above* brings up a possible bias of mine in a spirit of mutual truth-seeking, I’ll take it seriously. I don’t think this contradicts the general argument, or is any different from other domains. I don’t want random members of the public shaming me for my degenerate lifestyle, but if a close friend thinks I’m harming myself then I want them to let me know. I’m realizing as I’m writing this that this paragraph deserves its own essay, and that it would probably be a better and more important essay than this one is.

Most important, I think first-person bias arguments are valuable. You should always be attentive to your own biases. First, because it’s easier for you; a rando on Twitter may not know how my whiteness or my Jewishness affects my thought processes, but I might have some idea. Second, because you’re more likely to be honest: you’re less likely to invent random biases to accuse yourself of, and more likely to focus on things that really worry you. Third, you have an option besides just shrugging or counterarguing. You can approach your potential biases in a spirit of curiosity and try to explore them. I think I’m probably biased against communism because many communists I met have been nasty people who tried to hurt me, so I try to solve that by reading more communist books and seeking out good communist arguments wherever I can find them. Second- and third-person bias arguments risk feeling some kind of awkward option to change your opinions to something you don’t really believe in order to deflect someone’s bias accusations. First-person bias arguments should lead to a gradual process of trying to look for more information to counter whatever motivated reasoning you might have.

Know Your Gabapentinoids

The gabapentinoids are a class of drugs vaguely resembling the neurotransmitter GABA. Although they were developed to imitate GABA's action, later research discovered they acted on a different target, the A2D subunit of calcium channels. Two gabapentinoids are approved by the FDA: gabapentin (Neurontin®) and pregabalin (Lyrica®).

Gabapentin has been generic since 2004. It's commonly used for seizures, nerve pain, alcoholism, drug addiction, itching, restless legs, sleep disorders, and anxiety. It has an unusually wide dose range: guidelines suggest using anywhere between 100 mg and 3600 mg daily. Most doctors (including me) use it at the low end, where it's pretty subtle (read: doesn't usually work). At the high end, it can cause sedation, confusion, dependence, and addiction. I haven't had much luck finding patients a dose that works well but doesn't have these side effects, which is why I don't use gabapentin much.

Pregabalin officially went generic last month, but isn't available yet in generic form, so you'll have to pay Pfizer $500 a month. On the face of things, pregabalin seems like another Big Pharma ploy to extend patents. The gabapentin patent was running out, so Pfizer synthesized a related molecule that did the same thing, hyped it up as the hot new thing, and charged 50x what gabapentin cost. This kind of thing is endemic in health care and should always be the default hypothesis. And a lot of scientists have analyzed pregabalin and said it's definitely just doing the same thing gabapentin is.

But some of my anxiety patients swear by pregabalin. They call it a miracle drug. They can't stop talking about how great it is. I can't use it too often, because of the price, but I'm really excited about the upcoming generic version coming out so I can use it more often.

Still, I have to wonder – why am I sitting around waiting when I could just give people gabapentin? Confirmed pharmacodynamically-identical, generic, and cheap? The answer is, gabapentin doesn't seem to work that well. I've never had patients with more than minimal anxiety happy on gabapentin alone. Am I imagining a difference betwee these two supposedly-similar medications? I don't know. Although studies confirm pregabalin is great for anxiety, nobody has done the studies on gabapentin that would let me compare it. For now, the apparent difference between pregabalin and gabapentin is one of the great mysteries of life, one of the things that makes me doubt my own sanity.

One possibility is that we're getting the doses wrong. UpToDate recommends treating anxiety disorders with gabapentin using a starting dose of 300 mg twice a day = 600 mg daily. But it recommends 100 mg three times a day = 300 mg of pregabalin. This dosing table suggests 1 mg pregabalin = 5 mg gabapentin, so 300 mg of pregabalin = 1500 mg gabapentin! So we're starting gabapentin patients on less than half as much medicine as we start pregabalin patients on. If this forms a reference point in the doctor's mind, then maybe what we think of as a "high dose" of gabapentin is the same as what we think of a "low dose" of pregabalin. Maybe all our gabapentin doses are just too low.

I usually avoid higher gabapentin doses because I feel like they have more side effects than low pregabalin doses. Is this just an illusion? Is it my bias? If a patient reports feeling dizzy on high-dose gabapentin, do I say "Yeah, you're on a really high dose, I'm not surprised you feel that way, let's back off?" And then if they feel the same thing on low-dose pregabalin, might I say “It’s a low dose, you’re just getting used to the medication, give it a few more weeks”? Might my biases even be affecting how patients report their own experiences?

Or might there be some obscure pharmacologic mechanism? This paper tries to compare the pharmacology of the two drugs. They say the body can easily absorb pregabalin, but has a limited ability to absorb gabapentin – the more gabapentin there is, the lower a percent gets absorbed:

> With regard to the fraction of the dose absorbed, the lowest gabapentin dose studied (100 mg every 8 hours) is associated with absolute bioavailability of approximately 80%. This value was shown to decrease with increasing dose to an averageof 27% absolute bioavailability for a 1600 mg dose every 8 hours. In contrast, oral bioavailability of pregabalin averaged 90% across the full dose range of 75 to 900 mg/day studied

This doesn’t match the dosing table linked above, which suggests a 1:5 constant ratio between gabapentin and pregabalin dose. It also doesn’t really match the paper’s Figure 3, which shows a linear effect of gabapentin up to 1800 mg for nerve pain. It *does* match the paper’s figure 4, which shows little to no effect of gabapentin past 600 mg for seizures. I don’t really know what’s going on here. It would make some sense if the bottleneck were plasma -> CSF absorption, but that’s not what the paper’s saying. In any case, if the gabapentin/pregabalin relationship followed the same pattern for anxiety as for seizures, it would be impossible to ever get a dose of gabapentin as high as the starting dose for pregabalin, which would explain perceptions of pregabalin’s superiority. Try to increase gabapentin dose, and you just have extra gabapentin sitting around in the GI tract causing trouble. I don't know if this is at all the right way to be thinking about this.

One more difference: gabapentin is not a controlled substance, but pregabalin is Schedule V, the designation the government uses for things that are technically addictive but that it's not going to worry about too much. Why the difference? The government's documentation of their decision doesn't say. It could be total chance: both substances are right on the border, and a different bureaucrat got assigned to each case. But the decision doesn't seem totally off-base to me. Although it's theoretically possible to get addicted to gabapentin if you use a really high dose and try really hard, you'd have to be pretty desperate even by drug addict standards. I've seen a little more pregabalin addiction, though I agree with the FDA that it's still pretty unusual (some people in the comments disagree). One likely culprit is the absorption rate: pregabalin gets absorbed in an hour or so, gabapentin takes three or four. Faster-acting substances are always more addictive; they peak higher and sooner, and it's easier for the brain to associate stimulus (taking the drug) with response (feeling good). Could this also explain some of the efficacy difference? I don't know.

Phenibut is not FDA-approved; it's a common medication in Russia which gets sold as a supplement/nootropic/recreational drug in the US. The FDA occasionally asks people to stop selling it, but they've never gotten serious, and it's still easily available on the open Internet.

Phenibut has the kind of approval ratings usually associated with North Korean dictators who kill anyone who disapproves of them – including the highest median rating on my nootropics survey. It's phenomenal for social anxiety

– not in the SSRI way of making you a little calmer, but more in the “getting just the right amount of drunk” way that turns you into a different, bolder, and more fun-loving person. Aside from this, it can give a hard-to-describe sense of tranquility and well-being.

(it also makes you feel like you’re wearing a hat even when you aren’t. I swear this is a real side effect.)

Needless to say, it’s potentially addictive and can seriously ruin your life. Conventional wisdom in the phenibut user community is that you can use 500 mg once every week (or maybe every two weeks) safely. Anything beyond that and you develop rapid tolerance. Increase the dose to fight the tolerance, and you start feeling worse on the days you don’t take it, using it more and more to compensate for the rebound, and eventually getting a withdrawal syndrome closely related to the *delirium tremens* that sometimes kills recovering alcoholics.

(does this mean that responsible phenibut use is a free way to have one great day per two weeks? depends how good your willpower is, I guess. see also this graph from this source)

The discovery of ketamine’s efficacy for depression was a mixed blessing. Ketamine such is a difficult medication to use – dangerous side effects, intolerable hallucinations, IV delivery – that it could never be a panacea, whatever its potential. But the discovery sparked a hunt for other ketamine-like chemicals that shared its efficacy but not its downsides. It also started a race to figure out how ketamine worked, with the hope that this would provide the key to what depression *really* was, deep down. Phenibut should inspire the same kind of interest. It’s too dangerous to use regularly, but it’s great enough that we should be looking into what the heck is going on.

Early research into phenibut focused on GABA, the main inhibitory neurotransmitter. The brain has two kinds of GABA receptors, GABA-A and GABA-B. Alcohol, Xanax, Valium, Ambien, barbituates, and the other classic sedatives all hit GABA-A. There aren’t that many chemicals that hit GABA-B, and the few that are out there tend to be kind of weird – one of them fell to Earth on a meteorite. But phenibut is a GABA-B agonist. This sounds like a neat solution to the mystery: a drug with unique anti-anxiety properties affects a unique inhibitory receptor. But another GABA-B agonist, baclofen, has minimal anti-anxiety effects. It is mostly just a boring muscle relaxant (there was some excitement over a possibility that it might cure alcoholism, but the latest studies say no). So probably GABA-B on its own doesn’t explain phenibut.

This led researchers to propose that phenibut might work as a gabapentinoid. It has the defining GABA backbone, and it has activity at the A2D calcium channel subunit. But its gabapentinoid activity is much weaker than gabapentin itself, so why should its effects be stronger?

Baclofen outdoes phenibut as a GABA-B agonist, and gabapentin outdoes phenibut as a gabapentinoid, but phenibut works better than either. This is the other big gabapentinoid mystery that keeps me awake at night.

Might it be a synergistic effect between the two different actions? If this were true, we would expect taking gabapentin and baclofen together to have phenibut-like effects. But these drugs are sometimes used for the same kinds of neuromuscular conditions and nobody has ever noticed anything out of the ordinary. I would love to see this studied but I don’t expect much.

Phenibut has two enantiomers, r-phenibut and s-phenibut. Both are decent gabapentinoids, but only r-phenibut has GABA-B activity. If both worked equally well, that would suggest phenibut worked on A2D; if r-phenibut worked better, that would implicate GABA. The best source I can find is this study, which says that only r-phenibut has effects on rats. Do the hokey tests they run rats through exactly correspond to treating anxiety in humans? Unclear, but this pushes me more in the direction of thinking GABA-B is an important part of phenibut’s effects. So does a passing resemblance between phenibut and GHB, an unusual drug that works on GABA-B among other things.

Overall I think phenibut is probably more GABA-B agonist than gabapentinoid, but I can’t explain why it’s so different from baclofen.

One fringe possibility: it isn’t. I’ve said that these two drugs are used for different indications, by different populations, and get different results. But the map isn’t the territory, and the way humans use and think about drugs doesn’t always reflect chemical reality. Everyone knew the second generation antipsychotics were totally different from the first generation ones, until we learned that they weren’t really, and the different effects we saw were a combination of using them differently plus having different expectations. And placebo alcohol can still get people pretty drunk. The only study I’ve found directly comparing phenibut to baclofen finds they work for similar indications, at least in rats (see bottom of page 476). And I can find a few comments on Reddit backing this up from experience.

My odds are against this theory – I think there's probably some real difference between these drugs that we don't understand. But constant vigilance never hurts.

[EDIT: commenter dtsund points out that baclofen has some issues with blood-brain barrier permeability; see here for more. Although some of it gets through, it could build up in the plasma much faster than in the brain, giving it disproportionately peripheral effects]

Book Review: The Electric Kool-Aid Acid Test

Ken Kesey, graduating college in Oregon with several wrestling championships and a creative writing degree, made a classic mistake: he moved to the San Francisco Bay Area to find himself. He rented a house in Palo Alto (this was the 1950s, when normal people could have houses in Palo Alto) and settled down to write the Great American Novel.

To make ends meet, he got a job as an orderly at the local psych hospital. He also ran across some nice people called “MKULTRA” who offered him extra money to test chemicals for them. As time went by, he found himself more and more disillusioned with the hospital job, finding his employers clueless and abusive. But the MKULTRA job was going great! In particular, one of the chemicals, “LSD”, really helped get his creative juices flowing. He leveraged all of this into his Great American Novel, *One Flew Over The Cuckoo’s Nest*, and became rich and famous overnight.

He got his hands on some extra LSD and started distributing it among his social scene – a mix of writers, Stanford graduate students, and aimless upper-class twenty-somethings. They all agreed: something interesting was going on here. Word spread. 1960 San Francisco was already heavily enriched for creative people who would go on to shape intellectual history; Kesey’s friend group attracted the creme of this creme. Allan Ginsberg, Hunter S. Thompson, and Wavy Gravy passed through; so did Neil Cassady (“Dean Moriarty”) Jack Keroauc’s muse from *On The Road*. Kesey hired a local kid and his garage band to play music at his acid parties; thus began the career of Jerry Garcia and the Grateful Dead.

Sometime in the early 1960s, too slow to notice right away, they transitioned from “social circle” to “cult”. Kesey bought a compound in the redwood forests of the Santa Cruz Mountains, an hour’s drive from SF. Beatniks, proto-hippies, and other seekers – especially really attractive women – found their way there and didn’t leave. Kesey and his band, now calling themselves “the Merry Pranksters”, accepted all comers. They passed the days making psychedelic art (realistically: spraypainting redwood trees Day-Glo yellow), and the nights taking LSD in massive group therapy sessions that melted away psychic trauma and the chains of society and revealed the true selves buried beneath (realistically: sitting around in a circle while people said how they felt about each other).

What were Kesey’s teachings? Wrong question – what are *anyone’s* teachings? What were Jesus’ teachings? If you really want, you can look in the Bible and find some of them, but they’re not *important*. Any religion’s teachings, enumerated bloodlessly, sound like a laundry list of how many gods there are and what prayers to say. The Merry Pranksters were about Kesey, just like the Apostles were about Jesus. Something about him attracted them, drew them in, passed into them like electricity. When he spoke, you might or might not remember his words, but you remembered that it was important, that Something had passed from him to you, that your life had meaning now. Would you expect a group of several dozen drug-addled intellectuals in a compound in a redwood forest to have some kind of divisions or uncertainty? They didn’t. Whenever something threatened to come up, Kesey would say —the exact right thing—-, and then everyone would realize they had been wrong to cause trouble.

But what *were* Kesey’s teachings? Oh, *fine*. He talked a lot about movies. Everyone had a movie. The cops had a cop movie, businessmen had a businessman movie, trauma victims had trauma victim movies. Everyone was just reading their script, doing what was expected of them. But with enough enlightenment (realistically: drugs), you could break out of other people's movies – not just refusing to play the part they assigned you, but making them question the role they assigned themselves. You could rewrite your own movie, stop being an actor and take control of your own life. That woman in the picture, who put the flower in the barrel of the gun as the riot police stared her on: she was channeling Kesey. The riot police have their riot police movie, where you either back off or fight against them or do some other reasonable riot-police-related thing. If you can break out of that, just *be yourself*, your repetoire of possible actions expands to infinity.

There was some big New York celebration of Ken Kesey's second novel. The Merry Pranksters decided to make a road trip out of it. They got a school bus, painted it in crazy psychedelic colors, set the destination plaque on the front to "Further", and all crammed aboard. Neil Cassady took the wheel; everyone else got super high on drugs, dressed up in the craziest costumes they could think of, crowded around the windows and on the roof, and set out on a quest to deprogram everyone in middle America (realistically: shout at them and act crazy). For a five thousand mile round trip, they abused every drug known to man, had wild sex, played pranks (in Arizona, they pretended to be a campaign stunt for ultraconservative presidential candidate Barry Goldwater), and generally bonded. A couple of people predictably went psychotic and had to be let off at local mental hospitals, but as Kesey's gnomic proverb went, "you're either on the bus or off the bus".

By the time they got back to their California compound, they had become legends. Various gurus and counterculture figures came to seek audiences with Kesey; they all left convinced that *something* had happened, even if they couldn’t put it into words. The Hell’s Angels, notorious for crashing hippie parties and then fucking everyone up, came to Kesey’s compound, and Kesey…somehow socially hacked them. They ended up behaving like perfect gentlemen the whole time, then left promising to play support for Kesey if he ever needed it. Kesey got invited to speak at a Unitarian meeting on the evolving counterculture. Somehow he…hacked…the Unitarians, so that all of the younger attendees (and some of the older ones) were following Keseyism and not Unitarianism by the end of the conference.

(you can make jokes about “what’s the difference?”, but I don’t know enough history to know if Unitarianism was its current ultraliberal hippieish self at the time, or how much of that is actually *because of* Kesey hacking their conference)

The Bay Area counterculture was going to unite for a giant anti-Vietnam rally. They were going to march on the military base! They were going to start a revolution! They invited Kesey as their speaker, and Kesey decided he would hack their movie. He repainted his bus in military camo, then covered it in swastikas. At the head of a literal battalion of military-uniformed, swastika-bearing Merry Pranksters, he marched onto stage and declared that protesting a war was much like fighting in a war, in the sense that either way you were playing the war movie. He was going to do something different. He was going to play “Home On The Range” on his harmonica on stage, for no reason, until the organizers forced him to stop. Which they couldn’t do, because he was surrounded by loyal cultists doing a pretty good army imitation. So in the midst of what was supposed to be the beginning of a violent revolution, a swastika-clad Kesey stood on stage and played “Home On The Range” on his harmonica, until everyone got confused and dispersed.

Energy started to build. Kesey couldn’t be stopped. His momentum was too strong. It was one thing after another. The Beatles started making clearly Kesey-inspired albums. All the hip people of San Francisco begged him to lead them, show them the way forward. Kesey obliged. He started holding a series of parties – first small, then gradually larger – called Acid Tests. Everyone would take LSD. But not just LSD. The LSD would be modulated, amplified a thousand times by every bit of technology and ingenuity the Merry Pranksters could dream up. Strobe lights, black lights, surround-sound, the true multimedia experience. Kesey had invented the rave. The Grateful Dead, now his official cult band, followed him everywhere. Owsley Stanley, a weird scientific prodigy who at one point was personally responsible for synthesizing the majority of the world’s LSD, became his audio guy, and invented large swathes of modern audio processing. They were doing things light-years beyond what anyone had done before, and they were doing it at scale: 6,000 people attended one 1966 Acid Test.

Finally, Caiaphas and the Pharisees decided they had seen enough. The police swooped down. LSD was legal at the time, but the Merry Pranksters were using so many other illegal drugs that it didn’t matter. Kesey was arrested on marijuana charges, potentially carrying a long prison sentence. He decided to do the same thing any reasonable person would: fake his suicide and flee to Mexico.

It went badly. The fake suicide was spotted almost immediately, because Kesey couldn’t resist writing an overly dramatic, slightly hilarious suicide note. Also, without Kesey’s stabilizing presence, the Merry Pranksters immediately fell to backstabbing and infighting. Kesey's St. Peter figure, Ken Babbs, tried his hardest to hold everything together, but despite some victories (for example, inventing the concept of spiking Kool-Aid with LSD) it all collapsed. Everyone decided that life without Kesey wasn't worth living, boarded their highly conspicuous psychedelic bus, and crossed the Mexican border to hunt him down. The police noticed this and sent their own agents to Mexico to hunt him down (what's the Spanish for "did any of you see a bus painted bright rainbow colors full of screaming half-naked people pass through here recently?"). Also, somehow some hot hippie girls who were into Kesey managed to track him down, which did not bode well for his concealment from police. There followed a complicated Mexican manhunt which at one point involved Kesey jumping onto a moving train.

Finally Kesey realized: he's in *their* movie. The cops-and-fugitive movie. He's got to break out of that frame. So he dresses as a cowboy, gets on a horse, crosses the border back into America through some sort of cowboys-do-what-they-want exemption, and goes back to San Francisco. There he becomes the most public fugitive in the history of crime, speaking at various hippie events, attending various concerts, and giving interviews to the press – always disappearing just before the police arrived. His fame shoots past the stratosphere and into the Outer Empyrean.

Finally, inevitably, the police nab him. He's still very rich from his *Cuckoo's Nest* royalties, so he gets great lawyers who are able to bargain down to a minor marijuana charge. As icing on the cake, Kesey gives a speech. He says that, during his time in Mexico, he's realized that it's time to go "beyond LSD". The judge is intrigued. During Kesey's absence, hippiedom has grown from a small avante-garde to a giant movement, and become Public Enemy #1 in the eyes of the law. And here's the Chief Hippie himself, saying what *sort of* sounds like "kids shouldn't do drugs". This is good enough for them! They let him out on bail for the one minor marijuana charge he still has going, *on the condition* that he preach his "beyond LSD" message to vulnerable youth.

(is it just me, or might the San Francisco legal system contain some of the stupidest people in the whole of human history?)

Kesey calls in all his favors. He rents one of the biggest concert halls in the Bay Area. He books the Grateful Dead. He calls forth the Hell's Angels. He tells everyone – *everyone* to be there. He is going to tell them how to go beyond LSD and usher in the New Project. They are going to have the Acid Test Graduation, where the greatest Acid Test of all time transitions seamlessly into some grand post-LSD future that Kesey has planned.

Except – people *like* LSD. They *like* being hippies. They *like* taking drugs, then talking about how this is totally going to be the next stage of human evolution, man, if only the squares could see it. They *like* having sex with the torrent of hot women who are coming to the 1960s Bay Area to have sex with hippie guys. The famous hippie rock bands *like* being famous hippie rock bands. The wacky hippie philosophers *like* being wacky hippie philosophers. If Jesus Christ held a Second Coming in 1200 AD, and asked everyone to gather in Jerusalem for further instructions, would they go? Maybe the Pope likes being the Pope. Maybe bishops like being bishops. Maybe they don't *want* to move to the next stage.

And so Kesey is betrayed. The Grateful Dead have another engagement. The concert hall is – oops – otherwise occupied. The financial backers are mysteriously low on cash. What should have been the crowning event of everything gets unceremoniously cancelled, “with no time left to start again”.

So Kesey says fuck it, he’ll do it all from an abandoned warehouse owned by the Pranksters. No music, no fancy lights, no crowds. Just whoever still holds the faith and wants to come and listen.

They come. Not everybody, but some people. The people who know. The hippest of the hip, the ones who haven’t abandoned the cause.

Kesey gets up on stage, faces all of them, and…

…just kind of rambles for an hour or two. Mumble mumble beyond LSD mumble. It doesn’t make a lot of sense.

Kenneth E. Kesey has lost the Mandate of Heaven. Maybe he left it in Mexico, during the chaos of flight. Maybe it got beaten out of him by the cops, or extracted from him in jail. Maybe it slunk away in shame when all his friends betrayed him. But it’s not there. It’s gone.

A few months later, he goes to trial on his minor marijuana charge. He gets six months in jail, about the usual punishment for a charge like that. He spends six months in jail. Then he moves back to Oregon. He lives there in relative obscurity for another thirty years, writing occasional short stories and giving occasional interviews to reporters and historians.

II.

There, I saved you from having to read Tom Wolfe’s *The Electric Kool-Aid Acid Test*.

Tom Wolfe is as famous as they get. And the book jacket claims that EKAAT “defined a generation”. I couldn’t stand it. Wolfe is trying to do some kind of experimental fiction where he portrays the Merry Pranksters in the style of an LSD trip. But I don’t want my nonfiction novels to have the style of an LSD trip. I want them to just tell me the piece of history they’re supposed to be about. Wolfe is a great writer, journalist, and historian. When he wants to, he can draw out exactly what’s so fascinating and amazing about this incredible piece of history. But in order to get to those moments, I have to read through prose like:

> EXCEPT FOR HAGEN’S GIRL, THE BEAUTY WITCH. IT SEEMS LIKE she never even gets off the bus to cop a urination. She’s sitting back in the back of the bus with nothing on, just a blanket over her lap and her legs wedged back into the corner, her and her little bare breasts, silent, looking exceedingly witch-like. Is she on the bus or off the bus? She has taken to wearing nothing but the blanket and she sheds that when she feels like it. Maybe that is her thing and she is doing her thing and wailing with it and the bus barrels on off, heading for Houston, Texas, and she becomes Stark Naked in the great movie, one moment all conked out, but with her eyes open, staring, the next laughing and coming on, a lively Stark Naked, and they are all trying to just snap their fingers to it but now she is getting looks that have nothing to do with the fact that she has not a thing on, hell, big deal, but she is now waxing extremely freaking ESP. She keeps coming up to somebody who isn’t saying a goddamn thing and looking into his eyes with the all-embracing look of total acid understanding, our brains are one brain, so let’s visit, you and I, and she says: ‘Oooooooh, you really thinkthat, I know what you mean, but do you-u-u-u-u-u-u-ueeeeeeeeeeeeeeeeeeee” — finishing off in a sailing

trémulo laugh as if she has just read your brain and !t is the weirdest of the weird shit ever, your brain eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee —

STARK-NAKED
in a black blanket —
Reaching out for herself,
she woke up one morning to
find herself accosted on all
sides by

LARGE

MEN

surrounding her threatening her
with their voices, their presence, their always
desire reaching inside herself
and touching her obscenely upon her
desire and causing her to laugh
and
LAUGH
with the utter
ridiculousness
of it. . .

— but no one denied her a moment of it, neither the conked-out bug-eyed paranoia nor the manic keening coming on, nobody denied her, and she could wail, nobody tried to cool that inflamed brain that was now seeping out Stark Naked into the bouncing goddamn — stop it! — currents of the bus throgging and roaring 70 miles an hour into Texas, for it was like it had been ordained, by Kesey himself, back in San Juan Capistrano, like there was to be a reaction scale in here, from negative to positive, and no one was to rise up negative

> about anything, one was to go positive with everything — go with the flow — everyone's cool was to be tested, and to shout No, no matter what happened, was to fail.

I am not against experimental fiction. But I wish people would restrict the experimental fiction to bad books, so that it isn't in the way of me learning actually interesting stuff.

One thing I did appreciate about Wolfe was that he really knew his anthropology of religion. This thing, where one guy suddenly starts radiating this sense of *knowing*, and then a knot of disciples forms around him, and then predictable consequences ensue: this has happened again and again in human history. If I'd had to read this same story, written by someone who didn't realize he was writing about a trope – didn't realize that instead of an indiscriminate catalogue of events, he should focus on the way this illuminates the general principle, or differs from it – it would have been annoying. Instead, Wolfe gets it exactly right.

I'm less convinced *Kesey* understood what was going on. At one point (p. 193) he says:

> We're not on the Christ Trip. That's been done, and it doesn't work. You prove your point, and then you have 2,000 years of war. We know where that trip goes.

But it's never clear exactly what he's doing to avoid that trip. And although there are worse problems to have than "failed to sufficiently differentiate yourself from Jesus", this is about the limit of Kesey's self-awareness – at least as described in this book. It's weird to think that a prophet's biographer knows more about religion than the prophet themselves, but that's the impression I walked away with. If Tom Wolfe had started a cult, it would have *gone somewhere*.

III.

A few years ago, I reviewed PIHKaL on Alexander Shulgin. Shulgin and Kesey are twin titans of modern drug culture: Shulgin invented the drugs, Kesey invented the culture.

Why care about Bay Area drug culture? One could give Captain Barbossa's answer: "You better start caring about Bay Area drug culture, Miss Turner: you're in it!" I don't really use psychedelics myself, but I have a couple of patients – and a few acquaintances – who are on the Kesey trip. Some of them have taken the Kesey trip all the way to the psych hospital and the borders of schizophrenia. Others still think they're making new discoveries, getting closer to the place Kesey called Edge City where you can gaze off the end of the world into the mysteries beyond. Even beyond my own social circle, it's unclear how much the world in general owes Bay Area drug culture. A lot of important movements – environmentalism, meditation, wellness, vegetarianism, antipsychiatry – trace at least some of their origins to that strange froth. Over thirty million Americans have taken psychedelics – probably disproportionately intellectuals and creative people. If the drug actually changes people's minds, in some important way, then that's a big deal. If some alien supercomputer had been simulating world history from Mesopotamia onward, would its predictions start going haywire sometime in the 1960s? Would the alien programmers charged with debugging it eventually find that money, power, and the other usual suspects had sufficed before then – but that afterwards the simulation needed to include some very specific aspects of mushroom biochemistry?

Or one could give H.G. Wells' answer: "New and stirring things are belittled because if they are not belittled the humiliating question arises 'Why then are you not taking part in them?”. One of two things must be true. Either psychedelics are a unique gateway to insight and happiness, maybe the most powerful ever discovered. Or they have a unique ability to convince people that they are, faking insight as effectively as heroin fakes happiness. Either one would be fascinating: the first for obvious reasons, the second because it convinces some pretty smart people. If the insight of LSD were fake, its very convincingness could tell us a lot about the mind and about how rationality works.

The story of Ken Kesey doesn’t do much to clarify the situation. On the one hand, Kesey becomes a very successful cult leader, attracting some of the most creative people of his era. He is able to accomplish larger-than-life feats like taming the Hell’s Angels, escaping the FBI, and inventing modern party culture. Everyone who meets him ends up enchanged, terrified, or both. This sounds like the sort of thing we would expect of someone who’s successfully developed Secret Wisdom Into The True Nature Of Things.

On the other hand, Kesey’s followers are unable to replicate his success. Kesey tries to communicate what he has to his followers; first out of basic desire to help, and later because he knows he’s going to have to flee to Mexico and leave them on their own. He handpicks his successor, Ken Babbs, who is respected and liked by the rest of the Pranksters. But the moment he leaves, everything collapses. Tom Wolfe lampshades this quite cruelly. Every grudge, conflict, and personal failing comes out into the open, and the whole movement comes tumbling down. Then when Kesey’s back on bail, he picks up the pieces, says just the right words to everybody, and they all come together again as if nothing happened. Then he’s out again, and everything collapses a second time. The impression one gets is that Ken Kesey is a special person, with LSD playing a supporting role at best. Did LSD help give Ken Kesey his powers? Unclear. Did it give anything at all similar to the dozens of other bright people who took it under Kesey's supervision? Clearly not. But then what made Kesey special? Did he have a unique Book Four style experience? When? How come I just read his entire biography and don't see anything of the sort?

And what are we to make of Kesey's own exhortation that seekers needed to move "beyond LSD"? What *was* his Acid Test Graduation? I don't know if Wolfe's impenetrable writing style failed me here, or if Kesey was inherently unclear on this. There were hints that maybe he thought some of the paraphernalia of drug culture – the strobe lights, the acid rock music, the multimedia experience – could produce the psychedelic state without needing drugs. Seems pretty false.

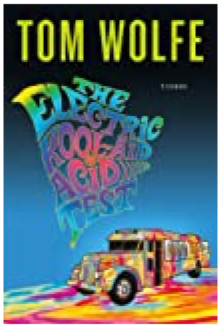
And what is the whole "mandate of heaven" thing? This is my term; Wolfe didn't make such a big deal of it. But reading the book, it really does seem as if at some point after his bail, Kesey lost his effectiveness. He tried to draw people in and they didn't come. He tried to change everything but it wouldn't change. Then he gave up, moved to Oregon, and did approximately nothing for the next thirty years.

The best I can do in making sense of this story is to think of Kesey as having unique innate talents that made him a potential cult leader, combined with the sudden rise in status from being a famous author and the first person in his social scene with access to LSD. Despite the connotations of "cult leader", Kesey was overall a good person, genuinely wanted to help people's spiritual development, and genuinely thought LSD could do this. LSD formed the content of his cult, the same way Messianic

Judaism formed the content of Jesus' cult. It also made his life easier because of the drug's natural tendency to make people think they are having important insights. When he, attempting to genuinely discover a spiritual path, decided to change the content and go beyond LSD, he lost that crutch, his people betrayed him, he became less confident in himself, and eventually he gave up.

But it's interesting how often malevolent cult leaders have turned to drugs since then – so often that I see it listed in some lists of cult warning flags. It's especially surprising since we usually think of LSD as helping people break out of systems of mental control – why would you give it to somebody you're trying to brainwash? One hypothesis that I think Wolfe points at in *Electric Kool-Aid Acid Test* is that LSD keeps you constantly convinced you're on the verge of some great revelation that will change everything. This both makes you more skeptical of everyday society – they don't have the revelation! – and more vulnerable to spiritual predators, who can say that yes, they have the great revelation, and if only you follow them for a little longer you might get it too.

Kesey rode this wave innocently, carried on by the same hopes as his followers. Some benevolent god judged him mercifully and granted him a quiet retirement writing short stories in the woods of Oregon, which is a lot better than how these stories usually end. I'm more nervous about anyone who tries the same strategy today.


TOM WOLFE
THE ELECTRIC KOOL-AID ACID TEST

Adversarial Collaboration Contest 2019

[self plagiarism notice: this is mostly copied from last year's contest announcement]

1. Announcing the second annual Adversarial Collaboration Contest

An adversarial collaboration is an effort by two people with opposing opinions on a topic to collaborate on a summary of the evidence. Just as we hope that a trial with both prosecutor and defense will give the jury a balanced view of the evidence for and against a suspect, so we hope an adversarial collaboration will give readers a balanced view of evidence for and against some thesis. It's typically done for scientific papers, but I'm excited about the possibility of people applying the concept to to less formal writeups as well.

For example, a pro-gun activist might collaborate with an anti-gun activist to write a joint article on the evidence for whether gun control saves lives. We trust each person to make sure the best evidence for their respective side is included. We also trust that they'll fact-check each other and make sure there aren't any errors or falsehoods in the final document. There might be a lot of debating, but it will happen on high-bandwidth informal channels behind the scenes and nobody will feel like they have tailor their debating to sounding good for an audience.

Last year, SSC held an adversarial collaboration contest. You can see the entries here:

1. Does The Education System Adequately Serve Advanced Students?

2. Should Transgender Children Transition?

3. Should Childhood Vaccination Be Mandatory?

4. Are Islam And Liberal Democracy Compatible?

I want to repeat the contest this year. Prize money depends how many people enter (see terms below) but will probably be around $2500 (thanks to people who support this blog on Patreon). All entries that meet a minimum level of quality will also be published on SSC.

2. How To Form A Team

Setting up teams was chaotic last year, so I'm going to try to be more organized about it. If you want to participate, please post a top-level comment on this post saying the topic you're interested in, your position on it, any other relevant information, and an email that would-be-partners can contact you at. Make the comment **in bold** so that it stands out against the background of people idly discussing things. For example:

> **Hi, I want a partner for a collaboration on whether the moon is made of green cheese (I think yes). I am an astrogastronomy PhD student and would prefer to work with someone else who has at least degree-level knowledge in the field. Email me at fake[at]example[dot]com**

If you're interested in someone else's topic, please send them an email. Don't reply to their comment, since a) the person might not see it, b) it might discourage other people from replying, and the proposer might want to get contacted by more prospective partners so they can see who would be the best. Once two people have agreed to be a team, the person with the top-level comment can edit it to clarify they’re not interested in taking more offers.

Once two people have agreed to be a team, please email me, **scott[at]slatestarcodex[dot]com**, with your names and the topic you’re working on.

3. Terms And Conditions

1. You should write an essay summarizing your joint summary of the evidence regarding a controversial topic you disagree on. Strongly recommend that this be a single factual issue, like “Does gun control save lives on net?”, rather than a vaguer moral question like “Guns – good or bad?”, though it can still be a pretty broad topic – I would love to see people write about Caplan’s case against education, for example. Even though most of the examples here are political, this doesn’t have to be; it could involve controversial topics in medicine, history, religion, et cetera.

2. You will write the essay as a united front. Please don’t write “Alice says this study proves guns save lives, but Bob says it’s wrong and this other study proves guns are bad.” Instead you are going to have to come to an agreement on how to describe each study. For example “Here is a study purporting to show that guns save lives. It seems to accurately describe what is going on in rural areas, but it might be of limited applicability elsewhere.”

3. You will come to at least some sort of unified conclusion, even if that conclusion is “There’s not enough evidence in this field to be sure either way and we should default to our priors/biases”.

4. The essay should be similar in length, tone, and amount-of-research to one of my Much More Than You Wanted To Know essays, eg here and here.

5. By entering the contest, you are giving me permission to publish your essay on SSC (with full attribution to you, of course). You can also publish it wherever else you want. I will probably publish the winning essay, and I might or might not publish the others depending on how good they are.

6. Because of (5), please don't research any topic that I would not be able to publish on SSC if you came to a taboo conclusion. If you want to do an adversarial collaboration on taboo topics, you can feel free to arrange it in the comments, but it won't be considered an official entry, it won't be eligible for prizes, and I probably won't post it (I might link it if it's posted somewhere else). If you're wondering whether a specific topic is taboo, you can ask.

7. Due date is November 1.

8. I won't hold the contest if fewer than five teams sign up. That's "sign up", not "complete their collaboration"; I realize many teams will drop out. I'll let you know if I'm holding the contest or not within a week or two, before you waste too much time on it.

9. If I hold the contest at all, I'll disburse $1000 in prize money. If there are at least four complete eligible entries, I'll disburse $2500 in prize money. If there are at least ten, $5000. More then ten, I don't know, but I'll try to make it worth your time.

10. I'll give the winning entry somewhere between 50-100% of the total prize money. If I don't give it 100%, the rest will go to second place, third place, etc. I haven't decided

how/whether I will do this and it depends on how good the individual entries are.

11. If you win, I will pay through PayPal or online donations to the charity of your choice.

12. Winner will be determined by poll of SSC readers, plus my vote counting for 10 percentage points in the poll.

13. I reserve the right to change these conditions in minor ways that don’t significantly inconvenience contest participants.

14. I’ll give an update on the next visible Open Thread.

Against Against Billionaire Philanthropy

[Conflict of interest notice: I've volunteered for both private and public charities, but more often private. I received a small amount of money for work done for a private charity ten years ago. Some of the private charities have been partially funded by billionaires.]

From Vox: The Case Against Billionaire Philanthropy. It joins The Guardian, Truthout, Dissent Magazine, CityLab, and a host of other people and organizations arguing that rich people giving to charity is now a big problem.

I'm against this. I understand concern about the growing power of the very rich. But I worry the movement against billionaire charity is on track to damage charity a whole lot more than it damages billionaires. Eleven points:

1. Is criticizing billionaire philanthropy a good way to protest billionaires having too much power in society?

Which got more criticism? Mark Zuckerberg donating $100 million to help struggling schools in Newark? Or Mark Zuckerberg buying a $59 million dollar mansion in Lake Tahoe? Obviously it's the schools. I've heard people criticizing Zuckerberg's donation constantly for years, and I didn't even know he *had* a $59 million Lake Tahoe mansion until I googled "things mark zuckerberg has spent ridiculous amounts of money on" in the process of writing this paragraph.

Which got more negative press? Jeff Bezos donating $2 billion for preschools for underprivileged children? Or Jeff Bezos spending $2 billion on whatever is going to come up when I Google “things jeff bezos has spent ridiculous amounts of money on?”.

Billionaires respond to incentives like everyone else. If donating to charity earns them negative publicity, and buying a private yacht earns them glowing articles about how cool their yacht is, they’re more likely to buy the yacht.

Journalists and intellectuals who criticize billionaires’ philanthropy but not their yachts, or who spend much more energy criticizing philanthropy than yachts, probably aren’t doing much to promote a world without billionaires. But they’re doing a lot to promote a world where billionaires just buy yachts instead of giving to charity.

2. If attacks on billionaire philanthropy decrease billionaires’ donations, is that acceptable collateral damage in the fight against inequality?

That depends on your values. But for most people’s values, the answer is no.

Nobody knows exactly how many lives the Gates Foundation has saved. *The Guardian* says it’s some appreciable fraction of the 122 million lives saved in general from progress fighting infectious diseases over the last few decades. This article says Gates has saved seven million people through his vaccination campaign alone, provided another seven million with antiretroviral treatment (usually life-saving), “tested and treated” twelve million people for tuberculosis (often fatal, but there’s a big difference between testing and treatment), and been responsible for a big part of the seven million lives saved from malaria. I expect these numbers are inflated, but even by conservative estimates the Gates Foundation may have saved ten million people.

Suppose Jeff Bezos is watching how people treat Bill Gates, and changes his own behavior accordingly. Maybe in the best possible world, when people attack Gates' donations, Bezos learns that people don't like ruthless billionaires, decides not to be ruthless like Gates was, and agrees to Bernie Sanders' demand that he increase his employees' pay by $4/hour. But Bezos also learns people criticize billionaires' philanthropy especially intensely, decides not to be charitable like Gates was, and so ten million people die. You've just bought an extra $4/hour for warehouse workers, at the cost of ten million lives.

In my moral system, this means billionaire philanthropy is not acceptable collateral damage in the war against inequality. Even if for some reason you believe that criticizing billionaire philanthropy is a higher-impact way to fight inequality than criticizing billionaires' yachts, you should stick to criticizing the yachts.

3. Do billionaires really get negative reactions from donating? Didn't I hear that they get fawning praise and total absence of skepticism?

Vox quotes Rob Reich (not the same person as the former Labor Secretary), a prominent critic of bilionaire philanthropy. Reich writes that billionaires "ask everyone involved to bend over in gratitude for her benevolence and genius in sprinkling around some social benefits" and so we need to "stop being merely grateful to donors and instead direct our skepticism and scrutiny at their activities".

How much gratitude vs. scrutiny do billionaire donors get?

The three most publicized recent billionaire donations were Zuckerberg to Newark schools, Bezos to preschools, and Gates to malaria. I looked at Twitter to examine how much fawning vs. scrutiny people were giving each. Specifically, I searched “Zuckerberg Newark”, “Bezos preschool” and “Gates malaria”. I then coded the first twenty-five tweets on the Top Tweets page for each as positive, negative, or neutral. I ignored mismatches that weren’t about the donations, and also ignored the genre of people using Zuckerberg’s donation as a way of criticizing Cory Booker (which was more than half of the Zuckerberg tweets).

Searching “Zuckerberg Newark”, I counted 2 positive tweets, 4 neutral tweets, and 19 negative tweets:

> This looks similar to failed effort of the Zuckerberg (Facebook) Foundation to "reform" Newark schools. Also serves similar purposes: Private interests control the policy agenda and evade progressive taxation.
>
> — Robert Cotto Jr (@RobertCottoJr) April 5, 2019

> Zuckerberg’s costly effort to remake Newark’s public school system may have failed, but that hasn’t put a damper in his plan to transform the educational landscape with his “personalized learning” schemes.https://t.co/0LgY85NjGb
>
> — The Baffler (@thebafflermag) January 14, 2019

Searching “Bezos preschool”, I counted 5 positive, 7 neutral, and 14 negative tweets:

> I am still mad that Jeff Bezos thinks he can solve for preschool in America. Y'all. We have a model that works beautifully. It is called Head Start. Just give it to everyone. The end.
>
> — Abby Norman MDiv (@abbynormansays) September 14, 2018

> So Jeff Bezos is putting $2 billion into homelessness and preschool. Laudable. But you have to ask — is this better than just taxing the wealth of the richest man in the world, and making that money democratically accountable? https://t.co/PcGKJWcMhJ
>
> — Lydia DePillis (@lydiadepillis) September 13, 2018

Searching “Gates malaria”, I counted 15 positive, 4 neutral, and 6 negative.

> Read The Business Of Changing The World. @BillGates, phenom philanthropist gets it. Gates was worth $91b when he did the giving pledge about ten years ago, he’s given $50b. His net worth today is around$110b. He’s also on the verge of eradicating malaria. https://t.co/RuCQNRKcDX https://t.co/2AJE2lBFQQ
>
> — María Teresa Kumar (@MariaTeresa1) May 10, 2019

> The @gatesfoundation efforts to eradicate malaria are undemocratic in this rubric. If billionaires like Gates/Buffet were prevented and that money was with the government, it would likely be spent elsewhere.
>
> Billionaires can be useful for #AWorldThatWorksForEveryone
>
> — Eric Minick (@EricMinick) July 18, 2019

The same is true of Google search. I examined the top ten search results for each donation, with broadly similar results: mostly negative for Zuckerberg and Bezos, mostly positive for Gates.

But when people talk about “billionaire philanthropy” in general, they tend to elide this distinction and focus on the bad. A twitter search for “billionaire philanthropy” produced 2 positive, 3 neutral, and 20 negative tweets, more negative than for any individual donation. A Google search for “billionaire philanthropy”, and the top ten results contained 1 positive article, 5 neutral articles, and 4 negative articles.

> "For the super-rich, giving is really taking. Taking power, that is, from the rest of society."
>
> This is why I prefer taxation to philanthropy.https://t.co/2D2BYyL8D7
>
> — Cameron Graham (@camerongraham) July 20, 2019

> If billionaires paid their fair share of taxes we wouldn’t need #philanthropy. It is just a facade to look like social heroes while actually engaging in tax evasion and demanding #TaxHolidays. #EatTheRich #ProletariatRevolution
>
> — Time To Change The World (@end_patriarchy) July 19, 2019

Although some donors like Bill Gates are generally liked, others, like Zuckerberg and Bezos, are met with widespread distrust. This might be because Gates has worked harder to target his donations well, or because he made his money a long time ago and nobody is too angry about his business practices anymore. But on a broader scale, the media and social media consensus is already parroting anti-billionaire-philanthropy talking points.

If everyone were unreflectively praising philanthropic billionaires, there would be a strong case for encouraging skepticism. But if most responses to billionaire philanthropy are negative, we should worry more about the consequences of the backlash.

4. Is it a problem that billionaire philanthropy is unaccountable to public democratic institutions? Should we make billionaires pay that money as taxes instead, so the public can decide how it gets spent?

From Dissent:

> Big philanthropy is overdue for reform. The goal should be to reduce its leverage in civil society and public policymaking while increasing government revenue. Some possible changes seem obvious: don't allow administrative expenses to count toward the 5 percent minimum payout, increase the excise tax on net investment income, eliminate the tax exemption for foundations with assets over a certain size, and replace the charity tax deduction with a tax credit available to everyone (for example, all donors could subtract 15 percent of the total value of their charitable contributions from their tax bills). In addition, strict IRS oversight of big philanthropy—especially all the "educating" that looks so much like lobbying and campaigning — is crucial […]
>
> Private foundations fall into the IRS's wide-open category of tax-exempt organizations, which includes charitable, educational, religious, scientific, literary, and other groups. When the creator of a mega-foundation says, 'I can do what I want because it's my money,' he or she is wrong. A substantial portion of the wealth — 35 percent or more, depending on tax rates — has been diverted from the public treasury, where voters would have determined its use.

This makes the same argument as some of the other articles linked above. Since billionaires have complete control over their own money, they are helping society the way *they* want, not the way the voters and democratically-elected-officials want. This threatens democracy. We can solve this by increasing taxes on philanthropy, so that the money billionaires might have spent on charity flows back to the public purse instead.

Two of the billionaires whose philanthropy I most respect, Dustin Moskovitz and Cari Tuna, have done a lot of work on criminal justice reform. The organizations they fund determined that many innocent people are languishing in jail for months because they don't have enough money to pay bail; others are pleading guilty to crimes they didn't commit because they have to get out of jail in time to get to work or care for their children, even if it gives them a criminal record. They funded a short-term effort to help these people afford bail, and a long-term effort to reform the bail system. One of the charities they donate to, The Bronx Freedom Fund, found that 92% of suspects without bail assistance will plead guilty and get a criminal record. But if given enough bail assistance to make it to trial, over half would have all charges dropped. This is exactly the kind of fighting-mass-incarceration and stopping-the-cycle-of-poverty work everyone says we need, and it works really well. I have donated to this charity myself, but obviously I can only give a tiny fraction of what Moskovitz and Tuna manage.

If Moskovitz and Tuna's money instead flowed to the government, would it accomplish the same goal in some kind of more democratic, more publically-guided way? No. It would go to locking these people up, paying for more prosecutors to trick them into pleading guilty, more prison guards to abuse and harass them. The government already spends $100 billion – seven times Tuna and Moskovitz's combined fortunes – on maintaining the carceral state *each year*. This utterly dwarfs any trickle of money it spends on undoing the harms of the carceral state, even supposing such a trickle exists. Kicking Tuna and Moskovitz out of the picture isn't going to cause bail reform to happen in some civically-responsible manner. It's just going to ensure that all the money goes to making the problem worse – instead of the current situation where the overwhelming majority of money goes to making the problem worse but a tiny amount also going to making it better.

Or take one of M&T's other major causes, animal welfare. Until last year, California factory farms kept animals in cages so small that they could not lie down or stretch their limbs, for their entire lives. Moskovitz and Tuna funded a ballot measure which successfully banned this kind of confinement. It reduced the suffering of hundreds of millions of farm animals and is one of the biggest victories against animal cruelty in history.

If their money had gone to the government instead, would it have led to some even better democratic stakeholder-involving animal welfare victory? No. It would have joined the $20 billion – again, more than T&M's combined fortunes – that the government spends to *subsidize* factory farming each year. Or it might have gone to the enforcement of ag-gag laws – laws that jail anyone who publicly reports on the conditions in factory farms (in flagrant violation of the First Amendment) because factory farms don't want people to realize how they treat their animals, and have good enough lobbyists that they can just make the government imprison anyone who talks about it.

George Soros donated/invested $500 million to help migrants and refugees. If he had given it to the government instead, would it have gone to some more grassroots migrant-helping effort?

No. It would have gone to building a border wall, building more camps to lock up migrants, more cages to separate refugee children from their families. Maybe some tiny trickle, a fraction of a percent, would have gone to a publicly-funded pro-refugee effort, but not nearly as much as would have gone to hurting refugees.

The idea that we should divert money from freeing the incarcerated, saving animals, and reuniting families – to instead expanding incarceration, torturing animals, and separating families – seems monstrous to me, even (especially?) when cloaked in communitarian language.

5. Those are some emotionally salient examples, but doesn’t the government also do a lot of good things?

Yes, but the US government is not a charity. Even when it’s doing good things, it’s not efficiently allocating its money according to some concept of what does the most good.

Bill Gates saved ten million lives by asking a lot of smart people what causes were most important. They said it was global health and development causes like treating malaria and tuberculosis. So Gates allocated most of his fortune to those causes. Gates and people like him are such a large fraction of philanthropic billionaires that by my calculations these causes get about 25% of billionaire philanthropic spending.

The US government also does some great work in those areas. But it spends about 0.9% of its budget on them. As a result, one dollar given to a billionaire foundation is more likely to go to a very poor person than the same dollar given to the US government, and *much* more likely to help that person in some transformative way like saving their life or lifting them out of poverty.

But this is still too kind to the US government. It’s understandable that they may want to focus on highways in Iowa instead of epidemics in Sudan. Yet even on issues vital for the safety of the American people, the government tends to fail in surprising ways.

How much money does the US government spend fighting climate change? This 538 article explains why this is a hard question, but it tries to give the best answer it can:

> The 2018 GAO report found that, while the Office of Management and Budget has reported that the federal government spent more than $154 billion on climate-change-related activities since 1993, much of that number is likely not being used to directly address climate change or its risks. Many of the projects reported as “climate-change-related activities” are only secondarily about climate change.
>
> For instance, the U.S. nuclear energy program predates serious concerns about climate change and would likely exist in its current form even if it did not produce fewer greenhouse-gases than some other forms of energy production, like burning coal. But the nuclear program’s budget is counted as climate spending. All told, when the GAO evaluated six agencies that report their climate spending to the OMB, it found that 94 percent of the money was going to programs that weren’t primarily focused on climate change — things like nuclear energy. The money marked as climate spending wasn’t going to new initiatives. Instead, “it’s a bunch of related things we were already doing,” Gomez told me. Numbers like that $154 billion total can be used as political props, but that kind of accounting isn’t much good for understanding what the government is actually doing about climate change.

$154 billion * 6% primarily focused on climate change / 25 years = $369 million per year. It might be higher than the 25-year average now, because of increasing awareness of climate change, but it might also be lower, because Trump. I have low confidence in the exact estimate but I think this is the right order of magnitude.

In 2017, the foundation of billionaire William Hewlett (think Hewlett-Packard) pledged $600 million to fight climate change. One gift by one guy was almost twice the *entire US federal government's* yearly spending on climate issues.

This isn't some parable on how mighty billionaires have become or how much power they have accrued. The government's budget is still 10,000 times bigger than Mr. Hewlett's. It's not that he's anywhere near government-sized, it's just that the government doesn't care *at all* about solving these kinds of problems, so a billionaire can outspend them if he cares *a little*.

Thanks to Hewlett and a few other people like him, I calculate that about 3% of billionaire philanthropy goes to climate change, compared to 0.01% of the federal budget.

Not every billionaire spends their money on global health or fighting climate change. There's a lot of criticism of billionaires who "waste" their donations on already-well-endowed colleges and performing arts centers, and I agree we should push them to think harder about their choices.

But charity, like investing, is in what Nassim Taleb calls Extremistan – almost all the value lies in getting it *very* right once or twice. An investment fund that picks a hundred duds plus 2004 Facebook is still an amazing investment fund. A form of philanthropy that produces a hundred duds, but also produces Bill Gates (and Dustin Moskovitz, and Cari

Tuna, and Warren Buffett, and Ben Delo, and…) is still an amazing benefit to the world.

I wish I could give a more detailed breakdown of how philanthropists vs. the government spend their money, but I can't find the data. Considerations like the above make me think that philanthropists in general are better at focusing on the most important causes.

I think this also makes intuitive sense. Charities are capable of laserlike focus on the most important and desperate causes. But give their money to the government instead, and it will get spent on building fighter jets, bombing brown kids in Afghanistan, shooting brown kids in Chicago, subsidizing coal companies, jailing anyone who tries to dress hair withoug a hairdresser license, and paying farmers not to grow crops – and *then*, at the end of all that, *maybe* have a tiny bit left over to spend on the desperately important problems that affect the most vulnerable people.

Governments are a useful type of organization that should exist. I don't want to get rid of them. But right now we're thinking on the margin, and on the margin an extra dollar given to a charity will do more good than that same dollar given to the government.

6. The point of democracy isn't that it's always right, the point is that it respects the popular will. Regardless of whether the popular will is good or bad, don't powerful private foundations violate it?

Reich again:

> The modern foundation is an institutional oddity in a democracy.

> In a democracy, officials responsible for public policy must stand for election. Don’t like your representatives’ policy views? Vote against them in the next election. This is the accountability logic internal to democracy — responsiveness to citizens. It does not always work this way, but the logic has some real force.
>
> But foundations have no electoral accountability. Don’t like what the Gates Foundation did with its $3.4 billion in 2011 grants ($9.3 million each day of the year), or what it has done with $25 billion in grants since its inception in 1994? Tough, there’s no way to vote out the Gateses.

I realize there’s some very weak sense in which the US government represents me. But it’s really weak. Really, really weak. When I turn on the news and see the latest from the US government, I rarely find myself thinking “Ah, yes, I see they’re representing me very well today.”

Paradoxically, most people feel the same way. Congress has an approval rating of 19% right now. According to PolitiFact, most voters have more positive feelings towards hemorrhoids, herpes, and traffic jams than towards Congress. How does a body made entirely of people chosen by the public end up loathed by the public? I agree this is puzzling, but for now let’s just admit it’s happening.

Bill Gates has an approval rating of 76%, literally higher than God. Even Mark Zuckerberg has an approval rating of 24%, below God but still well above Congress. In a Georgetown university survey, the US public stated they had more confidence in philanthropy than in Congress, the court system, state governments, or local governments; Democrats (though not Republicans) also preferred philanthropy to the executive branch.

When I see philanthropists try to save lives and cure diseases, I feel like there's someone powerful out there who shares my values and is doing right by them. I've never gotten that feeling when I watch Congress. When I watch Congress, I feel a scary unbridgeable gulf between me and anybody who matters. And the polls suggest a lot of people agree with me.

In what sense does it reflect the will of the people to transfer power and money from people and causes the public like and trust, to people and causes who the public hate and distrust? Why is it democratic to take money from someone more popular than God, and give it to a group of people more hated than hemorrhoids?

And if the people want more money to be spent by private philanthropists instead of Congress – and they use the democratic process to produce a legal regime and tax system that favors private philanthropy – *their will is being represented*.

7. Shouldn't people who disagree with the government's priorities fight to change the government, not go off and do their own thing?

Suppose I was donating money to feed starving children, and it was going well, and lots of starving children were getting fed. Then you come along and say "No, you should give that money to the Church of Scientology instead".

I say "No, I hate Scientology."

You: "Ah, but you can always try to *reform* Scientology. And maybe in a hundred years, it won't be awful anymore, and instead it will try to help starving children."

Me: “So you’re saying that I should work tirelessly to reform Scientology, and then in a hundred years when they’re good, I should give them my money?”

You: “Oh no, you have to give them all your money now. But while you’re giving them all your money, you can also work toward reforming them.”

Why would I do this? Why would it even cross anybody’s mind that they should do this? I am not saying that the government is evil in the same way as Scientology. But I think the fundamental dynamic – should you give your money to a cause you think is good, or to an organization you think is bad while trying to reform it? – is the same in both cases.

Also, do you realize how monumental a task “reform the government” is? There are thousands of well-funded organizations full of highly-talented people trying to reform the government at any given moment, and they’re all locked in a tug-of-war death match reminiscent of that one church in Jerusalem where nobody has been able to remove a ladder for three hundred years. This isn’t to say no reform will ever happen – it’s happened before, it will surely happen again, and it’s a valuable thing to work towards. Just don’t hold up any attempts to ease the suffering of the less fortunate by demanding they wait until every necessary reform is accomplished.

8. Is billionaire philanthropy getting too powerful? Should we be terrified by the share of resources now controlled by unaccountable charitable foundations?

From Dissent:

> Right now, big philanthropy in the United States is booming. Major sources of growth have been the wealth

> generated by high-tech industries and the expanding global market. In September 2013 there were sixty-seven private grant-making foundations with assets over $1 billion. The Rockefeller Foundation, once the wealthiest, now ranks fifteenth; the Carnegie Corporation ranks twentieth (Foundation Center). Mega-foundations are more powerful now than in the twentieth century—not only because of their greater number, but also because of the context in which they operate: dwindling government resources for public goods and services, the drive to privatize what remains of the public sector, an increased concentration of wealth in the top 1 percent, celebration of the rich for nothing more than their accumulation of money, virtually unlimited private financing of political campaigns, and the unenforced (perhaps unenforceable) separation of legal educational activities from illegal lobbying and political campaigning. In this context, big philanthropy has too much clout.

The yearly federal budget is $4 trillion. The yearly billionaire philanthropy budget is about $10 billion, 400 times smaller.

For context, the California government recently admitted that its high-speed rail project was going to be $40 billion over budget (it may also never get built). The *cost overruns alone* on a single state government project equal four times all the charity spending by all the billionaires in the country.

Compared to government spending, Big Philanthropy is a rounding error. If the whole field were taxed completely out of existence, all its money wouldn’t serve to cover the cost overruns on a single train line.

If this seems surprising, I think that in itself is evidence that the money is being well-spent. Billionaire philanthropy isn’t powerful, at least not compared to anything else. It just has enough accomplishments to attract attention. Destroying it wouldn't enrich anyone else to any useful degree, or neutralize some threatening power base. It would just destroy something really good.

9. Does billionaire philanthropy threaten pluralism?

From Reich's Vox interview again:

> I am, by contrast, a pluralist; I want to champion the decentralization of what would otherwise be a majoritarian decision-making structure for the spending of tax dollars to produce various forms of social benefits. And I think part of what makes ordinary charitable giving a good thing is the conversion of every individual's idiosyncratic, eccentric preferences into some civil society-facing project that by extension produces a diverse, pluralistic civil society, which is good for democracy.

I am having trouble following the argument. We need pluralism and decentralization. Therefore, we should ban anyone from doing their own thing, and instead force them to go through a single giant organization?

Consider the Multidisciplinary Association for Psychedelic Studies (MAPS). They sponsor research into mental-health-related uses of psychedelic drugs. You might have heard of them in the context of their study of MDMA (Ecstasy) for PTSD being "astoundingly" successful. They're on track to get MDMA FDA-approved and potentially inaugurate a new era in psychiatry. This is one of those 1000x opportunities that effective altruists dream of.

The government hasn't given this a drop of funding, because its official position is that Drugs Are Bad. MAPS writes:

> Every dollar has come from private donors committed to our mission. The pharmaceutical industry and federal government have not yet supported our work, so the continued expansion of psychedelic research still relies on the generosity of individual donors and foundations.

Most of the funding for their MDMA trial came from the foundation of billionaire Robert Mercer. Because there were actors other than the government with enough money to fund things they believed in, we were able to get some great work done even though it wasn't the sort of thing the government would support.

Or: in 2001, under pressure from Christian conservatives, President Bush banned federal funding for stem cell research. Stem cell scientists began leaving the US or going into other area of work. The field survived thanks to billionaires stepping up to provide the support the government wouldn't – especially insurance billionaire Eli Broad, who gave $25 million to the cause, and eBay billionaire Pierre Omidyar, who sponsored a California ballot initiative to redirect state funding to cover the gap. Time after time, the government has stopped supporting things for bad reasons, and we've been lucky that we didn't bulldoze over the rest of civil society and prevent anyone else from having enough power to help.

Or: despite controversy over "government funding of Planned Parenthood", political considerations have seriously limited the amount of funding the US government can give contraceptive research. It was multimillionaire heiress Katharine McCormick who funded the research into what would become the first combined oral contraceptive pill. More recently, it was Warren Buffett who funded RU-486 and the IUD. Together with similar work by the Rockefeller and Ford foundations, these have prevented millions of unwanted pregnancies.

When there are hundreds of different actors who can pursue their own projects, we get hundreds of genuinely different projects, some of which go great. If we restrict individuals from pursuing their own projects, and everything has to be funded by a single organization with a single agenda, we reduce the possibilities for progress to a monoculture, vulnerable to any minor flaw in the hegemon’s priorities.

In other cases, billionaires and government agencies are performing the same tasks in parallel. For example, both Bill Gates and the CDC are fighting infectious diseases in the developing world; both Elon Musk and NASA are working on space exploration.

Both groups bring different institutional cultures and priorities to the fight. The Gates Foundation is not run along exactly the same lines as the CDC; SpaceX has a different institutional culture from NASA. When one organization gets stuck in a dead-end, or isn’t up to a certain task, there’s a chance that the other will have the right structure to succeed. Some of this is random variation, some of it is structural differences between the public and private sector. I think it’s really healthy to have multiple diverse institutions trying to pursue the same goal. Robustness against obvious failures like “the government just banned all stem cell research” is just a special case of this principle.

I am using Reich as a foil, but in other places he seems to agree with this. At the end of this article he writes about “the case for foundations”, and says:

> I believe there is a case for foundations that renders them not merely consistent with democracy but supportive of it.
>
> First, foundations can help to diminish government orthodoxy by decentralizing the definition and distribution of public goods. Call this the pluralism argument. Second, foundations can operate on a longer time horizon than can businesses in the marketplace and elected officials in public institutions, taking risks in social policy experimentation and innovation that we should not routinely expect to see in the commercial or state sector. Call this the discovery argument.

I agree with all of this (and am now confused about to what degree Reich and I disagree at all), but I take this as meaning that private philanthropy, far from threatening pluralism, exemplifies it.

10. Aren’t the failures of government just due to Donald Trump or people like him? Won’t they hopefully get better soon?

Billionaires sometimes do a better job than the government at funding things like stem cells and the fight against climate change. But this is because of bad decisions by bad government officials. Obama overturned the stem cell ban; hopefully the next Democratic president will fix the climate funding situation. Does this make it unfair of me to compare the government vs. billionaires on this axis when there’s a hopefully-temporary reason the government is as bad as it is?

No. My whole point is that if you force everyone to centralize all money and power into one giant organization with a single point of failure, then when that single point of failure fails, you're *really* screwed.

Remember that when people say decisions should be made through democratic institutions, in practice that often means the decisions get made by Donald Trump, who was democratically elected. At the risk of going Civics 101, we're not *supposed to be* a pure democracy. We're a complicated system of checks and balances that uses democracy in some of its components. But we deliberately have other, less democratic components to deal with the situations when the *demos* f@#ks up. The *demos* seems to be f@#king up pretty regularly these days and I'm glad we still have those other institutions.

11. So you're saying these considerations about pluralism and representation and so on justify billionaire philanthropy?

I'm bringing up these considerations as counterarguments to some of the things opponents say. But I think they're the wrong thing to focus on.

The Gates Foundation plausibly saved ten million lives. Moskovitz and Tuna saved a hundred million animals from excruciatingly painful conditions. Norman Borlaug's agricultural research (supported by the Ford Foundation and the Rockefeller Foundation) plausibly saved one billion people.

These accomplishments – and other similar victories over famine, disease, and misery – are plausibly the best things that have happened in the past century. All the hot-button issues we usually care about pale before them. Think of how valuable one person's life is – a friend, a family member, yourself – then try multiplying that by ten million or a billion or whatever, it doesn’t matter, our minds can’t represent those kinds of quantities anyway. Anything that makes these kinds of victories even a little less likely would be a disaster for human welfare.

The main argument against against billionaire philanthropy is that the lives and welfare of millions of the neediest people matter more than whatever point you can make by risking them. Criticize the existence of billionaires in general, criticize billionaires’ spending on yachts or mansions. But if you only criticize billionaires when they’re trying to save lives, you risk collateral damage to everything we care about.

Adversarial Collaboration Update

Here are the teams that have registered for the adversarial collaboration contest so far.

1. Joel P. and Missingno on circumcision.
2. Yaleocon and flame7926 on incarceration.
3. Eigenmoon and Andrew S. on the simulation argument.
4. James B. and Reed on abortion.
5. TracingWoodgrains and lazygraduatestudent on critical learning periods.
6. TracingWoodgrains and Rhys Fenwick on reducing the abortion rate.
7. David G. and Froolow on eating meat.

Since that's more than five, the contest is officially going to happen!

I'll give everyone else one more chance to sort out teams, in the comment section here. Please don't comment unless you're proposing an adversarial collaboration topic. All other posts will be deleted. Please email me at scott@slatestarcodex.com once you've got a team and topic.

Also, an experiment: only people with usernames A-M can propose. People with names N-Z, you're stuck accepting from here out (people with names A-M can still accept if they want to). This is to address the problem where nobody accepts anyone else's offer because they'd rather propose their own topic. If it works, I'll do it the opposite way around next year. If you care enough about this to register a new username, whatever.

Squareallworthy On UBI Plans

I want to signal-boost Tumblr user squareallworthy's analysis of various UBI plans:

1. Jensen et al's plan
2. Healy et al's plan
3. Andrew Yang's plan
4. Torry's plan
5. Sheahen's plan
6. Dolan's plan
7. Stern and Murray's plans
8. Santens' plan
8½. Varoufakis and Reich's plan
9. Yang's plan, redux

He finds that most of them fail on basic math – they rely on funding schemes that wouldn't come close to covering costs. The rest are too small to actually lift people out of poverty. None of them are at all credible.

These plans fail even though they cheat and give themselves dictatorial power. "End corporate welfare, then redirect the money to UBI!" But if it was that easy to end corporate welfare, wouldn't people have done it already, for non-UBI related reasons? "We'll get a UBI by ending corporate welfare" is an outrageous claim. And even the plans that let themselves make it fail on basic math.

This is humbling and depressing. And it concludes the intelligent and useful part of this post that signal-boosts the work of a responsible person. Everything below is epistemic status: wild speculation.

About 40 million Americans live below the poverty line, which is $12,000 for an individual and a little higher for families. Multiplying these out to get $480 billion to end poverty is too high, first because most of these people live in families with each other, and second because most of them already have some income. Let’s halve it to $240 billion.

The top 1% are currently taxed at a rate of 37%, and this brings in about $560 billion. Increasing it to an even 50% would give an extra $200 billion or so, leaving $40 billion to get from random other places, the top 2%, etc. For a basic sanity check, the Bush tax cuts decreased revenue by $180 billion per year. A tax increase of the same scale as the Bush tax cuts would suffice for a basic income that ended poverty.

Isn’t basic income supposed to be universal? Yes, but most serious proposals accept that it will be gradually reabsorbed as higher taxes. People below some income gain money on net, people above the income lose money on net, and there’s some break-even point. I’m proposing the break-even point is somewhere between the poverty line and the top 1% – not in an abrupt way that forms a welfare cliff, but gradually according to the normal progressive tax system, and at a level so that we can imagine it abstractly as transferring money from the top to the bottom, with everyone else ending up about equally well-off, or getting gains and losses that cancel out.

This ignores the concern that higher taxes would stifle the economy, and the concern that the promise of a UBI would make more people quit their jobs and fall into the income stratum that benefits. But it also ignores the hope that lifting everyone out of poverty would obviate some welfare programs, or improve education, or bring other economic benefits. I don’t want to claim to be able to calculate all of these considerations, but order of magnitude estimate, we could give out a UBI sufficient to end poverty with a medium-sized tax increase.

If you think a world full of people trying to eke out an existence on $12,000 a year sounds dystopian, remember that’s the average cost of room and board at US colleges. I think the reason $12,000 can give college students such a good standard of living, but gives people on the edge of the poverty line such a bad one, is partly related to the hidden costs of work.

Why is this so much easier than the plans above? Number one, not pretending that it’s going to help the middle class. Number two, not pretending that it can be done without raising taxes. Remove those restrictions and the economics are easy.

Some people would say that it makes the politics impossible, but I’m not so sure. Raising taxes on the rich from 37% to 50% doesn’t seem outside the scale of what a Sanders or a Warren might do anyway. And pad it out a bit so that the middle class is still below the break-even point and gets maybe $1000 or so per year, and the pitch is still “end poverty in a way that makes you personally $1000 richer”. That’s still a pretty good pitch.

But if someone wanted a Yang-style universal income, one that went to everyone including the middle-class and which didn’t obviously raise anyone’s taxes, would there be a way to do it?

I wonder if you could just declare by fiat that it has to work, and leave the future to figure out the details. Start in 2020 with a basic income of $1 per person. Then mandate that it has to increase by $300 per year. Congress can decide how to do that, whether it’s by raising taxes or cutting other spending. If they can’t find the money, there’s a government shutdown until they can.

Even better, forget the $300 number and mandate that it has to increase at a rate pegged to GDP growth. Maybe fix the start date at 2020, and then in year X, one third of the difference between year X GDP and 2020 GDP must be given out as basic income. In theory you should be able to get a UBI of $10,000 per person in a few decades without making any person or government program poorer (though you will make their wealth increase less quickly than it would otherwise).

Aside from inflation (which means the $10,000 will be worth less than expected, but shouldn’t affect the system’s ability to reach $10,000 in 2020 dollars eventually) what am I missing?

Highlights From The Comments On Billionaire Philanthropy

Thanks to everyone who commented on Against Against Billionaire Philanthropy. For whatever reason, the comments there were exceptionally good. In particular, I'm happy that our usually-quiet leftists finally showed up with some strong (and interesting) pushback.

I usually highlight good comments with short responses, but it was hard for me to avoid debating some of these. I realize that's complicated, because I can't quote most long comments in their entirety, and I realize I have more of a platform than other commenters who may feel I misrepresented them or who just want to reply to me. I don't have a great solution to this, but if you're annoyed at how I featured/responded to your comment, please tell me, so I can calibrate how serious a problem this is for next time.

Matt Yglesias writes on Twitter:

> I think this whole post from @slatestarcodex and, indeed, most of the dialogue on the subject of "billionaire philanthropy" actually misses the bulk of what it is billionaires do with their philanthropy.
>
> In their great book "Billionaires and Stealth Politics" what Page, Seawright, and Lacomb find is that most billionaires don't seek publicity for their activities & mostly donate to right-wing causes like regressive tax cuts & lower government spending.

> Which is to say both sides of this debate are making clever arguments but mostly missing the big picture as to what billionaire philanthropy actually is — not complicated reputation laundering that helps people, but quiet lobbying for libertarian economics.
>
> If we want to debate whether Bill Gates, personally, is Good or Bad it would probably be better to do it on those terms (he seems pretty good to me) rather than talk broadly about billionaires — he is very unusual.

I think Matt is misunderstanding the thesis of *Billionaires And Stealth Politics*.

(I haven't read the book. But I've read the the original paper it was based on, plus some book reviews that confirm the book is saying the same thing as the paper)

Stealth Politics doesn't argue that billionaires spend more on right-wing causes than on altruistic ones. It argues that on certain issues relating to wealth and inequality, billionaires spend more on right-wing causes than on left-wing ones. Its methodology is terrible in a bunch of ways, but I assume something like that is probably true, so whatever.

The key finding in the paper (also cited in reviews of the book) is that:

> [From 2011 – 2012] most billionaires (65% of those who made partisan contributions) contributed primarily or exclusively to Republicans, and the bulk of their money (averaging $53,227) went to Republican rather than Democratic ($21,411) parties or candidates.

Okay. Over the same period (2011-2012), Bill Gates donated $6,070,000,000 to global health. That is 100,000 times as much as $53,227. Please don't cite this book to show that “the big picture” is right-wing spending, and that altruistic spending is irrelevant by comparison.

But this is a cheap shot, and the book isn’t designed for these kinds of comparisons, so I want to try my own analysis to see how much billionaires spend on right-wing vs. altruistic causes.

I started with OpenSecrets.org’s list of the 100 top contributors to “federal candidates, parties, political action committees, 527 organizations, and Carey committees”. In order to make the list, you had to donate at least $1.7 million. Of the 100 people who donated at this level, 55 were Democrats and 45 Republicans. The Republican contributions totalled $346 million.

This doesn’t count various sneakier ways of influencing the political climate – for example, the Koch brothers didn’t even make the list! Koch-affiliated groups probably spent $400 million on the 2018 election (some sources say they “planned to” spend $900 million on the 2016 election, but I think they held back because they didn’t like Trump). “Koch-affiliated groups” are complicated, and although they’re not 100% funded by the Koch brothers – one of their groups boasts 90,000 donors! – we can probably round up here. This doesn’t mesh well with claims that the Koch foundation donated $196 million over ten years, but my guess is the lower claim is just what they did openly, so let’s stick with the higher one. I am nervous about how many Koch-like things there are that we don’t know about, and this is a possible source of error in this estimate, but the Kochs are clearly the biggest.

What about non-election-related spending, like think tanks and college professorships? I looked at this list of the top 50 think tanks and added up their budgets. Total budget for all conservative/libertarian think tanks is about $350 million per year. I don’t know how much of this is funded by billionaires, but let’s say all of it.

Add all these numbers up, and divide the election-related ones by two since there’s only an election every two years, and order of magnitude it looks like maybe the budget for conservative politics is around $1 billion per year.

Some basic sanity checks – this site says the 2018 election involved $5 billion of donations, of which $2.5 billion went to the Democrats, $2.0 billion to the Republicans, and $0.5 billion to third parties. If billionaires fund about 30% of the Republican war chest, that would match my estimates above. There’s also this Forbes article, whose author estimates “that if we combine the budgets of these and other conservative free-market think tanks…the entire U.S., Canadian and U.K. market of pro-free society [ie conservative/libertarian] groups is approximately $500 million”. That meshes well with my estimate of $350 million for the US alone.

I can’t emphasize enough that I’m just eyeballing these figures and I can’t find anyone who has done really good work on this question – but I think I have the order of magnitude right. So how does $1 billion for right-wing causes compare to the overall billionaire philanthropy budget?

I’m leaning pretty heavily on this Vox article, but it says that the top 50 billionaires gave $15 billion to charity in 2017 and $8 billion in 2018. If we average this out to about $11 billion/year, only 9% of billionaire philanthropy goes to right-wing causes.

This is actually way too generous to Yglesias, because the Vox article admits it’s only talking about the top 50 US billionaires (out of 500 or so), but I’m comparing it to everyone (billionaire or not) who donated to a think tank or gave $1.7 million+ to campaigns. Probably the real number is much lower.

And we don’t even need to bring in all of the top 50 billionaires! Bill Gates alone averaged $2.5 billion in donations during those two years. One person gave more to global health than every billionaire combined gave to right-wing causes.

So I think Matt is wrong when he says that right-wing causes are where the real money is, and genuinely good philanthropy is just a distraction. The truth is the opposite.

Some commenters agreed that I had demonstrated to their satisfaction that the Gates Foundations was really great, but were concerned I was overusing it as an example. m50d wants us to picture a very different kind of spending when we think of donations:

> Imagine a community with a handful of rich people. Winston buys nice suits and fancy dinners, spreads his extra money around the town’s shops and businesses. Geoffrey puts all of his extra into an opera house (or a polo club, or a church he likes, or…). To my mind it’s absurd to say that Winston owes a larger (tax) contribution to the town’s community expenses than Geoffrey does – Winston’s spending already supports the rest of the town, whereas Geoffrey has directed his funds towards his own specific interests. (Which is fine and dandy with your own private money – but only once you’ve paid your share of taxes like everyone else).

Ultraximus agrees:

> You could end up in a situation where more and more decide to donate their income towards their favored (but very specific) cause / charity, which in turn decreases tax revenues at a governmental level, which reduces the level of public goods, which in turn leads to lower morale among citizens, which in turns leads to higher amount of people taking advantage of donation deductions, which in turn… soon you end up in a situation where rich donors give large donations to their favored schools and universities to get an entrance for their children. And get a tax deduction for it! From Nordic POV, this kind of situation is simply unfathomable.

Is it unfair of me to keep saying Gates Gates Gates Gates Gates and not talk about the polo clubs, churches, and underhanded trades with colleges?

I think the most important defense is: this isn't being graded on a curve. If one person saves ten million lives, and everyone else fritters away their money on stupid stuff, the end result is you've saved ten million lives.

This is the principle behind "hits-based investing". Suppose I invest $100 in the Safe Fund, which buys shares in 100 companies guaranteed to go up at market rate. You invest $100 in the Crazy Fund, which invests in 100 crazy startups; 99 will crash and burn, and one will be the next Facebook. Using real-life returns from Facebook vs. the Dow Jones over the same period, the Safe Fund would net me $270, and the Crazy Fund would net you $40,000. If you argued "Yeah, but that one startup that did really well was an exception, the Crazy Fund still mostly failed", you are missing the point of hits-based investing.

Likewise, if you argue “that one charitable foundation that did really well was an exception, billionaire foundations still mostly fail”, you are missing the point of hits-based giving. And if you use this as an excuse to weaken billionaire foundations in general, you will weaken the exceptional hits along with the many misses. If you decrease the amount of money available to billionaire charities by 37% (the natural consequence of removing tax-deductible status, assuming billionaires neither donate more to compensate, nor manage to evade your taxes), then a charity that saves 10 million lives will only be able to save 6.3 million lives, meaning your policy has killed 3.7 million people. The fact that you *also* sabotaged a bunch of idiots’ plans to spend money on their polo clubs seems kind of irrelevant.

I think that’s the most important defense. But a secondary defense, and one that I’m less confident in, is that actually a lot of billionaire charity spending is like the Gates Foundation.

One reason a lot of billionaire charity spending is like the Gates Foundation is because a lot of billionaire charity spending *is* the Gates Foundation. From eyeballing these two pages, the Gates Foundation looks like about 15% of US billionaire charitable spending. This is partly because Bill Gates is really rich, and partly because other billionaires like Warren Buffett donate to the Gates Foundation instead of setting up their own charities. I’m not just using the Gates Foundation as an example, I’m saying it’s a big chunk of what we’re talking about.

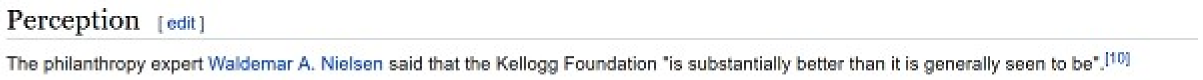
But even non-Gates billionaire charity seems pretty good. Moving down the list of largest charitable foundations, the second biggest in the US after Gates is the Howard Hughes Medical Institute, funded by the fortune of aviation tycoon Howard Hughes, which invests $825 million per year in medical research. Its homepage says that “as one measure of success, to date, 28 current or former HHMI scientists have been awarded a Nobel Prize”. That seems absurd to me – surely they must be cheating by giving $1 to every scientist in the world, or something like that? But if we’re going to critique billionaire philanthropy as useless, we should probably skip the people with 28 Nobel Prizes.

The third-largest billionaire foundation on the list is George Soros’ Open Society Foundation. They get a lot of flak for their US political activities, but according to their website, they also fight HIV and tuberculosis in Central Asia, provide prenatal vitamins to Roma in Hungary, help refugees in Jordan run small businesses, promote good nutrition in Liberia, help prosecute war crimes in the Congo, and build libraries in Haiti. Also, it did some amazing work propping up the social system in Eastern Europe after the collapse of communism, including giving $100 million to impoverished Russian scientists to keep doing science instead of going off and selling their nuclear secrets to rogue states. I am honestly pretty okay with this one too.

The next five foundations on the list are more questionable. Eli Lilly’s foundation mostly funds seminaries. The Ford Foundation used to do some amazing stuff, but is a recent victim of wokeness; the most heavily-advertised project on their site is about funding “social justice storytelling and the 21st-century arts infrastructure that supports it”. The Silicon Valley Community Foundations supposedly supports Silicon Valley communities, though for an organization that spends $1.3 billion per year, it’s surprisingly hard to find a single specific thing they’ve done. The Getty Foundation does art. The Robert Wood Johnson foundation does US health care, and it seems pretty good at it; I provisionally pronounce them okay.

Number 9 is the Hewlett Foundation, previously cited as briefly outspending the US government on climate change; it also helped make Creative Commons, secure nuclear weapons, save the rainforest, and spends about $100 million/year on global health.

Number 10 is the Kellogg Foundation, which mostly works on children's health in the US (especially dental care for low-income individuals), but has also worked to fluoridate water in Latin America. I was kind of unimpressed by their Wikipedia page, but:



Perception [edit]

The philanthropy expert Waldemar A. Nielsen said that the Kellogg Foundation "is substantially better than it is generally seen to be".[10]

I retract my objection.

These seem like an even mix between really impressive Gates-like work, and things that are a little less important but still not absolutely useless. The next few dozen on the list look similar, with some highlights (the Moore Foundation helps fund PLOS, Jupyter, Julia, and R) and some lowlights (the George Kaiser Family Foundation has decided the best use of its $4.4 billion is to pay techies $10,000 each to move to Tulsa, Oklahoma). But overall a lot of this seems really good.

---

subb4k writes:

> The Koch brothers are not quite on Hitler's level, but they are definitely funding policies designed to hurt members of the outgroup which includes most of humanity. They'll fund things that are likely to cause people to be hurt (from pollution, for example), rather than things actually likely to save people. I am ready to believe some of their donations fall under the traditional

definition of charity/philanthropy, but those probably aren't the ones people are mad about.

The hits-based giving model cuts both ways. If a small amount of spending goes to very good causes, it doesn't matter how much poorly-targeted spending there is. But if a small amount of spending goes to very bad causes, it matters a lot, even if it's only a tiny fraction of the whole. So is it possible that, even though right-wing spending is low in absolute number of dollars, it's still very bad?

Although it would be hard for this to be more harmful than Gates' ten million lives are helpful, it wouldn't be impossible. If the Koch brothers' spending on climate change denial makes the world significantly worse at fighting climate change, that could potentially cost ten million lives or more. One can (with a little more of a stretch) come up with ways other misguided political causes could do this.

I think in this case this worry probably doesn't apply. It looks like the Kochs spend about $10 million per year on climate-denial related causes (remember, this is like 1% of what Gates spends on health). This is trivial in comparison to the amount of climate denial spending by corporations (even though corporate spending and billionaire spending may seem morally similar, they occupy different places in the tax code, which is what's relevant for my argument). It's also trivial in comparison to how much some billionaires spend *supporting* the fight against climate change, eg the Hewlett Foundation's $600 million – although this is a dangerous argument, since the marginal dollar spent on climate denial might have more of an effect.

So this specific area is probably not so big a deal. But maybe the general concern is still valid. I personally doubt it’s true, but I can’t rule it out.

I think my answer is that, when I said the Gates Foundation etc saving ten million lives shouldn’t be collateral damage, I seriously meant *it shouldn’t be collateral damage*. If it’s important to stop the Koch brothers from lobbying against climate change, it’s worth a few extra moments’ thought to figure out some way to do this that doesn’t also prevent Bill Gates from fighting malaria. For example, have stricter rules saying political spending isn’t tax-deductible, but global health spending is. I realize it will be hard to figure out how to do this exactly right. But for ten million lives, please just do the hard thing.

This objection actually applies to a lot of the “this kind of billionaire philanthropy might be wasteful” or “this kind of billionaire philanthropy might be harmful” arguments. Even if you are right, I think it’s worth finding a way to separate the good from the bad, even if it’s hard, rather than throwing out baby and bathwater alike.

An Fírinne writes:

> The thing about billionaire philanthropy is that people like Jeff Bezos and Bill Gates don’t care about people. Billionaire philanthropy is nothing but a PR stunt in order to get the public off their backs about wealth redistribution and to distract from how dreadfully they threat their employees.

Gates (along with 203 other billionaires) has signed a pledge promising to donate half his wealth. Isn’t donating half your wealth kind of a weird step in a plot to keep all of your wealth? Hedge fund billionaire Jeremy Grantham has pledged to donate 98% of his wealth to fight climate change. Isn’t that a pretty dramatic way to prevent your taxes from going up 10% or something?

More fundamentally, even if donating money to ward off redistribution would be a good move for billionaires *as a class*, you would have to solve a really hard coordination problem to make any individual billionaire do it. If you have $10 billion, and your motivations are entirely self-interested, you shouldn’t donate $1 billion unless that single donation makes the government 10% less likely to appropriate billionaires’ wealth. No single donation will make that much difference, so self-interested billionaires shouldn’t act this way.

---

fluorocarbon writes:

> I don’t think this has to do with how they earn their money, but how they spend their money. Mark Zuckerberg didn’t just give $100 million to Newark schools, he gave the money on condition that the school district “welcome more independently run charter schools [and] close low-performing schools.”
>
> A lot of people disagree about the best way to run public schools. In a democracy, these kinds of conflicts over public institutions should be resolved through voting—every citizen should have the same say. The argument against Zuckerberg is that he’s subverting this process. Because he has more money, he can tell the school district how it should be run.
>
> Bill Gates, on the other hand, is spending money fighting malaria through the Gates Foundation. He’s not offering money to the CDC on condition that they have to run it the way Bill Gates wants.

> There will always be some people who are negative about everything, loud on Twitter, and who hate billionaires. But the reason Mark Zuckerberg and Jeff Bezos are getting so much pushback for their giving but Bill Gates is not is because Zuckerberg and Bezos are (or at least seem to be) using their money to commandeer public democratic institutions. Even if good comes out of it on the whole in a utilitarian sense, there's something a lot of people find unsettling and un-American about using money in this way.

And from Freddie deBoer:

> I was utterly flabbergasted when you approvingly cited the Newark disaster. "How could he possibly think this helps his cause?" I wondered. Control of Newark's school system naturally and by right belongs to the people of Newark. That's what democracy requires. Zuckerberg's explicit intent was to helicopter in and use his sacks of money to violate that sacred principle of local control. And it failed. It failed catastrophically. Because the people have the right to control their own local institutions.

I don't really care about Mark Zuckerberg or donations like his – I would be happy to throw them under the bus if doing so would help expand my coalition against sabotaging Gates and Moskovitz and Tuna. I would happily support some clarification of tax laws saying donations like Zuckerberg's aren't okay, or aren't tax-deductible. As long as it doesn't do collateral damage to the important stuff, whatever.

But I'm still confused by this perspective. My understanding (backed by sources like this), is that Zuckerberg *didn't* impose conditions for his donation. Cory Booker, Mayor of Newark, wanted to reform Newark schools. Someone connected him to Mark Zuckerberg, who thought Booker's ideas were great. So Zuckerberg donated the money to Booker, an elected mayor democratically vested with the legitimate authority to reform schools the way he wanted, and Booker enacted the reforms he wanted to enact anyway.

It's hard for me to see Zuckerberg as commandeering anything here. It's true that if he hadn't come around, Booker might have been forced to gnash his teeth and not do anything. And it's true that Zuckerberg made the decision to donate to Newark instead of some other city whose mayor wanted some other plan. But giving a city slightly more ability to do the plan it had decided on (I say "slightly" more because this source says Zuckerberg's gift was only 4% of the school system's budget) seems really different from "commandeering it" or "sabotaging democracy" or anything like that.

Even if Zuckerberg had absolutely told Cory Booker to change all of his plans to Zuckerberg-approved ones or else he wasn't going to donate, it *still* seems to me like the democratically elected mayor of a city has the right to decide that it is vs. isn't worth it. I mean, what would be a better alternative? That every billionaire donation (and only billionaire donations, and no other policy) has to be approved by a popular referendum before it's accepted? Or that you, a person who doesn't live in Newark, gets to dictate that the city of Newark can *never* accept billionaire donations, even if everyone in it wants to, while using Orwellian language about how this is to "respect the will of the people of Newark"?

Weirdly, I can't find any polls or even guesses about how much support Zuckerberg had among Newarkers, or how popular the school reforms were. Conventional wisdom is they were really unpopular. But conventional wisdom also says they failed, and the most recent studies really challenge that perspective, so I don't know.

It's hard for me to see this as a giant problem, but if it is, I think my preferred solution is that if governments don't want billionaires' help, they shouldn't ask for it or accept it. Destroying the power of billionaires to offer help to anyone seems like overkill.

Hackworth writes:

> On your point 4, you don't really refute the argument. You're just saying why it's ok for large-scale philanthropy to be anti-democratic because you can point to what you perceive to be success stories of billionaire charity. You're essentially saying that the end justifies the means or, less charming, might makes right. That doesn't sit right with me, because it can easily be turned around on your position. So, street violence against peaceful protesters, or gerrymandering, or voter suppression is anti-democratic? Tough luck, we still do it because we can.

I probably did err in how I described this.

I think it's good (as argued in the section being referred to) for some spending to remain outside the power of the government. I think "anti-democratic" means something more sinister than just "outside the power of the government", and I don't think billionaire spending qualifies.

That is: is me spending $2 on a slice of pizza anti-democratic? Is it just me saying "might makes right, I can do whatever I want with my own money, you can't stop me" without even asking the government whether my pizza-buying is the will of the people? Am I committing a crime against the public?

If not, what is the difference between me spending $2 on a pizza, me giving $2 to a beggar, and a billionaire giving $2 million to the Save The Beggars Foundation, as far as democracy goes? None of them represent a significant amount of money compared to the government. None of them represent an authoritarian-level concentration of power – there are 500 billionaires, supporting a wide variety of causes. They're all just people spending small (compared to government budgets) change on stuff they want, which is a known Okay Thing You Can Do In A Democracy.

Maybe Hackworth is referring only to the small percent of billionaire spending which is an attempt to influence the democratic process, like election donations or something? If so, I think it's important to say that explicitly, since I don't care about that kind of billionaire spending and am not really focusing on it. Also, because that mostly isn't tax-deductible and so Reich's argument doesn't apply to it.

Or maybe he's referring to the small percent of billionaire spending which eventually influences elections despite being tax-deductible, like donations to partisan think tanks? Again, don't care much about this, and agree it's less good than real charity and that we should try to discourage it. But I still don't think it's equivalent to voter suppression or whatever.

Like, what does it mean for something to be "anti-democratic"? Is founding a newspaper anti-democratic? Is it just "I'm the guy with the printing press, so my opinion matters more than yours, might makes right"? Is peaceful protest itself antidemocratic? "I have some free time and am good at organizing people, therefore I'm going to try to use shame and harassment to override the will of electorate who voted in the government I'm protesting"? Is all of this *basically* equivalent to beating people up in the streets?

I would counter that fighting for what you believe in isn't anti-democratic, it's the essence of democracy. Telling the government "Fuck you, you are wrong and bad and we'll see which of the two of us is still standing at the end of this" is continuing the legacy of Martin Luther King, Gandhi, and other people who are considered paragons of democratic society. The whole *point* of democracy is that individuals are supposed to try to change government policy! The examples given above – voter suppression, gerrymandering, street violence – are bad not because they're attempts by individuals to change government policy, but because they're attempts to *prevent* individuals from changing government policy.

And I don't think we can limit "trying to change government policy" to voting only. Putin's Russia has elections. ~~As far as I know they are "fair" in the sense of not lying about the vote count~~ [EDIT: Apparently sometimes they do]. The [EDIT: other] difference between them and a true liberal democracy is that the democracy is more permissive of attempts to change the government outside of a once-per-several-years election. I think these other attempts are basically about coordination. You can oppose Putin in your heart, you just can't coordinate large-scale effective opposition to Putin. Newspapers and protests help coordinate a bunch of powerless atomic individuals into a useful counterweight to government power.

One could defend billionaire foundations on the grounds that they do the same. For example, George Soros famously helped kickstart the Black Lives Matter matter movement. He did not create it out of thin air, and he is not a puppetmaster for it – it is composed of real and genuinely committed black people. But there’s a difference between those black people sitting around feeling committed at home vs. going out and effecting change, and Soros helped make that happen. Dustin Moskovitz and Cari Tuna’s animal advocacy work has done the same: there are lots of Californians who are against factory farming, but they couldn’t make anything happen until Moskovitz and Tuna solved the logistical work for them of getting a proposition on the ballot.

If billionaire foundations were evenly distributed across the political spectrum, they would have a nondistortionary effect of improving general coordination power and making it more possible to change the system. Since they’re not, they coordinate some parts better than others, which produces a genuine distortion. I agree this is bad. So it matters a lot whether you think a general increase in coordination power is worth the risk that the other side gets more of it than you do.

It sounds like Hackworth thinks the tradeoff is billionaire power vs. people power. I think there’s also state power, special interest power, corporate power, and (especially) Moloch. All of these actors are naturally well-coordinated and ordinary people’s success or failure in fighting them depends a lot on whether they can overcome their coordination handicap. This means I care more about coordination power, and think increasing it is good enough to be worth some distortion in the process. It’s still not as good as people donating to actually effective causes, but could be positive on net.

NocD writes:

> It doesn’t address what I’d consider one of the prime arguments against Billionaire Philanthropy which is the tax subsidized nature of it. Charity is tax deductible and those deductions need to be accounted for somewhere. So we really get a system where a select few get to underwrite their personal causes with governmental resources, building up their personal influence and capacity to enact very political choices. Gates’ support of charter schools and Zuckerberg’s own particular thoughts when it comes to education are not neutral endeavors and the profile they get to build being labeled as “Billionaire Philanthropists” lets them launder ideas and spend influence while being indirectly supported by government backing.

First, a terminological quibble: I don’t like summarizing this as “using government resources”. If you claim a child tax credit, are you diverting government resources to raise your child? Sounds pretty corrupt. If I heard that a senator was using government money to raise their child, I’d want them indicted. This seems like a deliberately inflammatory misrepresentation of the real situation, which is that the government has specifically decided not to claim those resources, and given donors its blessing to keep them.

Putting the semantic dispute aside, I interpret the concern as: should charitable donations be tax-deductible or not?

This is a hard question because I don’t have a coherent philosophy of tax deductibility. Lots of things are tax deductible, including mortgage interest on your house, medical costs, home offices, taking business clients out to lunch, gambling losses, retirement savings, and some state and local taxes. Most of these cost the government more money than charitable tax deductions – according to this website, the charitable deduction isn’t even in the top ten.

If I had to weave all these together into a coherent narrative, I would guess that the government usually gives tax deductions when people are doing something sympathetic that results in them not getting to use the money anyway. For example, if you have to spend a lot of your money on catastrophic health problems, there's a sense in which you never got to "enjoy" the money or spend it on what we usually consider personal consumption. Even if you spend your money on business expenses, that's different from enjoying it and buying luxury goods with it. Retirement savings don't get enjoyed immediately, plus we want to encourage people to save for retirement.

Charity donations seem to fit into this model. We want to encourage people to give to charity, plus money you give to charity isn't being spent on personal enjoyment and consumption.

All of this seems kind of vague to me, but given that there are hundreds of other similar tax deductions, and charity is both an especially prosocial one *and* an especially cheap one in the grand scheme of things, I'm not sure why anybody would consider it their first target for attacking. If you hate the very idea of tax deductions because it's "stealing from the government", fine. If you point to the charitable activity tax deduction in particular, while ignoring a bunch of much costlier and less beneficial ones, then I'm confused.

---

hls2003 writes:

> My natural inclination is to think that the people Scott is responding to are so wrong as to be nearly incomprehensible. But I feel like it is a failure of charitable imagination to simply accept the gulf is unbridgeable. So trying to put myself in their shoes, I

think that Scott’s rebuttal does not engage with the idea of philanthropy as wartime propaganda.

Take someone like Tokyo Rose or Lord Haw-Haw. If one constructed a careful analysis of what they could be doing with their time – committing atrocities, fighting as a common soldier, or serving a shift in a munitions factory – it is clear that materially speaking, talking on the radio is probably the least physically damaging thing they could be doing against Allied soldiers. But yet they were despised as much or more than anyone in those other jobs, because their role in propagandizing for their evil regimes helped the Axis war effort by boosting Axis morale and damaging Allied morale. They were the pleasant face (or voice) hiding an evil regime. If one starts from the premise that the capitalist system (or whatever systemic bogeyman you blame) is the enemy, then billionaire philanthropists like Gates are capitalism’s Lord Haw-Haw. They hide the true nature of the regime, prolong the war, and sap the will of the people to fight. Or take Potemkin villages, for a non-WWII example. Presumably a Russian peasant was legitimately better off (temporarily) for living in a clean, idyllic Potemkin village than continuing in their base-state filthy squalor. But just analyzing whether a few people were slightly better off clearly misses the main point that the Potemkin villages were hiding the truth and preventing change. Philanthropic billionaires are the Potemkin village, hiding the true impact of the capitalist bogeyman.

While I don’t share the starting opinion of capitalism, it seems like – under this Conflict Theory rubric – Scott’s essay misses the point of the critics. Why get more angry at charity than extravagance? The same reason you hate Lord Haw-Haw more than an ordinary German

> soldier; the soldier is just doing what you expect, and is mostly neutral to the broader effort of wartime morale, while Lord Haw-Haw is an underhanded traitor eating away at the emotional war effort. And why ignore the good work that charity does? Because the "seen" charity is simply a prop masking and perpetuating the "unseen" squalor of the system. In either case, micro-analyzing on a purely material level (Tokyo Rose versus Tokyo factory worker, who makes more bombs?) doesn't make sense because it omits the larger effects.
>
> Ironically I think the lesson of "seen" and "unseen," per Bastiat, is more commonly missed by the revolutionary Utopians. But if one starts with unrealized Utopia as a postulate, then even charity can be propaganda, and propagandists are plausibly reviled.

I find this one of the most sympathetic and comprehensible of the arguments I've read so far.

---

aashiq writes:

> Accelerationists might say philanthropy is wrong because it is a bandaid that keeps the existing fundamentally flawed system chugging for longer than it would otherwise. I do not believe this, but I was hoping to see it addressed in this article. These people might argue that we should oppose philanthropy exactly because it makes the world better. By letting conditions naturally get worse as capitalism pursues its own inhuman ends, a necessary revolution becomes more likely. Thus, we should eliminate policies like tax deductability of charitable contributions.

I understand accelerationists want to make the world worse, because that could cause a communist revolution, which would make the world much better. But what if there were something even better than a communist revolution? Then holding a communist revolution might alleviate enough suffering to prevent us from doing that even better thing! I think the only ethical choice is to promote billionaire philanthropy, which will prevent a communist revolution, which will cause the even better thing to happen. I realize this will be controversial, but I think sometimes you have to be willing to do a good thing for the greater bad if that greater bad contributes to an even greater good.

Luispedro writes (based on the estimate that the Gates Foundation may have saved ten million people):

> This means that if the combined effect of Reich et al. is a 1% probability of discouraging the next Bill Gates (i.e., one of the billionaires agrees with them that the “yachts and parties with paid supermodels” route is better than the evil “science and malaria cures” route), this still implies an death toll of 100,000 in expectation from their efforts.

We can do better! If the effect of me writing this blog post complaining about it is a 1% probability of discouraging the next Rob Reich, this implies I have saved 1,000 lives. And if the effect of you writing this comment is a 1% probability of encouraging the next me to write that article, you’ve saved 10 lives with this comment alone! Unless you’re a trauma surgeon or something, writing this comment is by far the most important thing you’ve ever done!

This is why I hate numbers like “ten million lives” and *really hate* Pascalian reasoning. I’m not saying it’s trivially wrong, just that I really hate it. It’s too crazy and too self-serving.

But I still wouldn’t recommend poking the thing that saves ten million lives with a stick to see if it breaks.

Gwern writes:

> While we’re listing billionaire examples: to a greatly underappreciated extent, the whole Reproducibility Crisis and its very expensive replication efforts like the Many Labs projects, are bankrolled by a billionaire, John D. Arnold. You’d think the federal government would be more interested in this problem, but it’s not.

This is a really interesting example I didn’t know about, thanks.

Conrad Honcho writes:

> Also, timely tweet/article.
>
> *“Only 22 percent of U.S. adults are on Twitter, and 80 percent of the tweets come from 10 percent of users. If you rely on Twitter for political information, you are being informed by ersatz pundits residing within 2.2 percent of the population.”*
>
> So I’m not sure how well a sampling of tweets about billionaire philanthropy represent public mood.

Many people objected to me using tweets to try to gauge public mood.

I may write another post soon defending my choice. But the short version is, all the time I hear people saying ridiculous things about what “everybody believes”. “Everybody believes veterans are evil murderers who deserve to be shamed, but I am going to be a bold contrarian and say we should support the troops" or whatever. I wish I could cite polls to tell these people they live in a fantasy world, but there aren't polls available on everything. I don't think Twitter perfectly matches public opinion, but it's not *infinitely* different from it, so I think it represents a useful corrective to letting people claim anything they want.

---

georgioz writes:

> If I am to steelman the critique of billionaires I think it really is about institutional integrity. As a comparison I calculated the total net worth of top 25 Putin's oligarchs according to this list. The result is somewhere around 270 billion dollars. Which is a lot but comparable to the power of all american billionaires relative to the economy – these oligarchs could finance Russian government budget also for around 1 year. And yet Putin and his oligarchs have a very tight grip on the whole Russian state and they have much more power than the net worth numbers suggest.
>
> So I'd explain the anti-billionaire philanthropy as just a gut feeling that it is a dangerous development. Maybe the danger lies in a fact that billionaires – at least as a class if not individuals – use the charity as a PR front in order to sway public opinion in their favor. Which is catastrophic as the concentration of power in their hand can pose a real threat to democracy prepping the stage for complete institutional takeover by more ruthless and ambitious among them. If this is the endgame then spending a few hundreds of billions of dollars on charity is nothing – not only monetarily but also compared to the damage caused longterm.
>
> From this standpoint spending on philanthropy is much more dangerous than spending on yachts. Because

> buying yachts is what billionaires are supposed to do and it leaves them prone to mundane criticism in standard political struggle.
>
> Or to use another example imagine you are Cato the Younger and you see Caesar spending his fortune to give bread and circus to the plebs of Rome. It is very hard to argue feeding the poor on the object level. But still Cato sees this as a ruse for the ultimate endgame of Caesar which is capture of the state at which point he will get his investment back with astronomical interest. Needles to say Cato was correct in his paranoia.

Good point. I guess I would be more worried about this if a completely repulsive billionaire known for never making the slightest attempt to do good in his life hadn't just captured the state using a completely different method, but maybe this one might happen sometime too.

Viliam writes:

> As a survivor of socialism, I can assure you that the sentiment is much deeper than that. Rich people giving gifts is a problem, but after that problem is solved, poor people giving gifts becomes a problem, too. Afterwards, simply "people doing something on their own" becomes a problem.
>
> The actual problem is doing things without the government approval. As a former socialist once said: "Everything within the state, nothing outside the state, nothing against the state." This is exactly how it worked in socialism of my childhood. Individual initiative was a bad thing, regardless of its actual aims. The only good thing was to do what you were told.

> The journalists argue that rich people's philanthropy is "against democracy", but what they really mean is that it is "against the state" or rather "without the state". This is why Scott's argument that people actually trust billionaire philanthropy more than government spending, so in that sense it is kinda more democratic, isn't going to convince anyone. It's not about being more democratic in the sense of "plebs being happy about what happens". It's about someone acting on their own, which threatens those who identify with the state.

Erusian makes a similar point:

> Not to be overly cynical, but I think there's a simpler explanation here than any real principle. Certain people have a class interest in empowering the state and shrinking the private square. These people will oppose billionaire philanthropy because it complicates their narrative that concentration of wealth is inherently bad. Zuckerburg buying a huge mansion doesn't disrupt their narrative. Zuckerburg curing cancer does, especially because that concentration of wealth is a necessary precondition for curing cancer. Likewise, due to their interest in growing government power they will very, very pointedly overlook any possible downsides to increased government power.

I wanted to highlight these because some people accused the anti-billionaire-philanthropy side of being overly fanatical conflict theorists. But it looks like there are some conflict theorists on both sides. I am skeptical of this. It's hard for me to imagine people having "crush individual initiative" as their goal, and I imagine these same people would support other kinds of outside-the-state individual initiative (like anti-Trump protests).

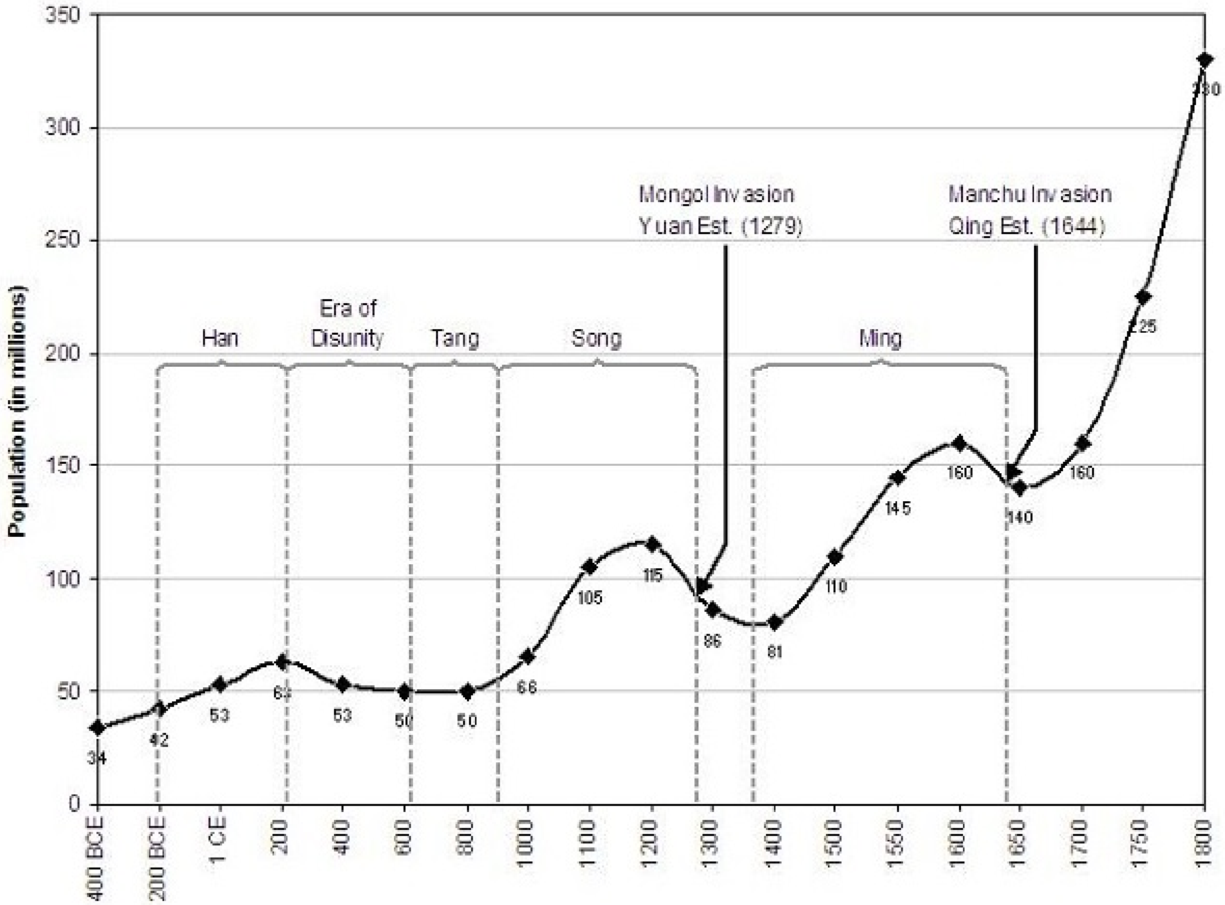
I was leaning towards this position before, but I find some of these comments sympathetic and convincing. I think the potential argument against billionaire philanthropy is that a lot of billionaires use it for vapid, personal, or even sinister causes, and it’s weird that they can get a tax deduction on it. If people are getting their college to make a vanity building with their name on it, or promoting their own politics, it does seem disturbing that this is costing the government money that it could use for potentially-valuable government business. I am sympathetic to people who worry about this.

But I think a more quantitative look shows that the good in billionaire philanthropy outweighs the bad, and that as a whole it more than meets our (low, confusing) standards for things that should be tax deductible. I am sympathetic to arguments that we should more carefully word the tax deduction, to make sure it only applies to genuinely altruistic ventures and not to vanity projects or lobbying. But I think this needs to be done with the understanding that the valuable parts of billionaire philanthropy are extraordinarily valuable, that they are not acceptable collateral damage, and that protecting and promoting them should be one of the most important goals for any policy on this issue.

Book Review: Secular Cycles

I.

There is a tide in the affairs of men. It cycles with a period of about three hundred years. During its flood, farms and businesses prosper, and great empires enjoy golden ages. During its ebb, war and famine stalk the land, and states collapse into barbarism.



Chinese population over time

At least this is the thesis of Peter Turchin and Sergey Nefedov, authors of *Secular Cycles*. They start off Malthusian: due to natural reproduction, population will keep increasing until it reaches the limits of what the land can support. At that point, everyone will be stuck at subsistence level. If any group ever enjoys a standard of living above subsistence level, they will keep reproducing until they are back down at subsistence.

Standard Malthusian theory evokes images of a population stable at subsistence level forever. But Turchin and Nefedov argues this isn’t how it works. A population at subsistence will always be one meal away from starving. When a famine hits, many of them will starve. When a plague hits, they will already be too sickly to fight it off. When conflict arrives, they will be desperate enough to enlist in the armies of whichever warlord can offer them a warm meal.

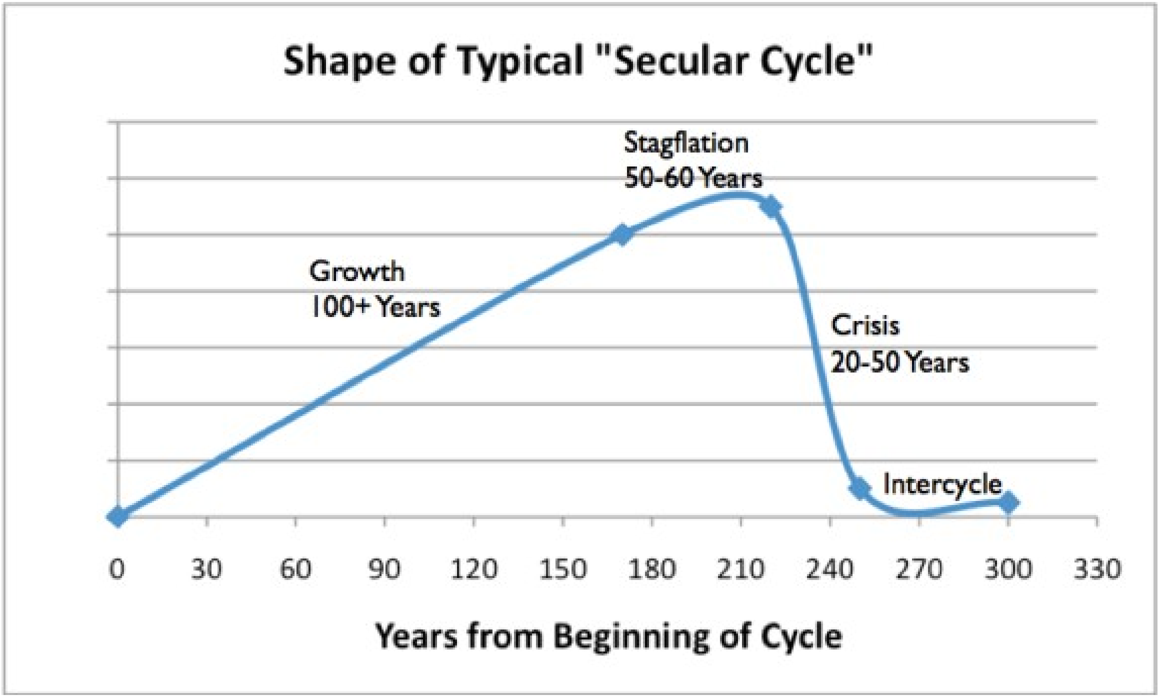
These are not piecemeal events, picking off just enough of the population to bring it back to subsistence. They are great cataclysms. The Black Plague killed 30% – 60% of Europeans; the Antonine Plague of Rome was almost as deadly. The Thirty Years War killed 25% – 40% of Germans; the Time of Troubles may have killed 50% of medieval Russia.

Thus the secular cycle. When population is low, everyone has more than enough land. People grow rich and reproduce. As time goes on, the same amount of farmland gets split among more and more people. Wages are driven down to subsistence. War, Famine, and Pestilence ravage the land, with Death not far behind. The killings continue until population is low again, at which point the cycle starts over.

This applies mostly to peasants, who are most at risk of starving. But nobles go through a related process. As a cycle begins, their numbers are low. As time goes on, their population expands, both through natural reproduction and through upward mobility. Eventually, there are more nobles than there are good positions…

(this part confused me a little. Shouldn't number of good positions scale with population? IE if one baron rules 1,000 peasants, the number of baronial positions should scale with the size of a society. I think T&N hint at a few answers. First, some positions are absolute rather than relative, eg "King" or "Minister of the Economy". Second, noble numbers may *sometimes* increase faster than peasant numbers, since nobles have more food and better chances to reproduce. Third, during boom times, the ranks of nobles are swelled through upward mobility. Fourth, conspicuous consumption is a ratchet effect: during boom times, the expectations of nobility should gradually rise. Fifth, sometimes the relevant denominator is not peasants but land: if a noble only has one acre of land, it doesn't matter how many peasants he controls. Sixth, nobles usually survive famines and plagues pretty well, so after those have done their work, there are far fewer peasants but basically the same number of nobles. All of these factors contribute to excess noble population – or as T&N call it, "elite overproduction")

…and the nobles form "rival patronage networks" to fight for the few remaining good spots. The state goes from united (or at least all nobles united against the peasants) to divided, with coalitions of nobles duking it out (no pun intended). This can lead either to successful peasant rebellion, as some nobles support the peasants as part of inter-noble power plays, or just to civil war. Although famine and plague barely affect nobles, war affects them disproportionately – both because they were often knights or other front-line soldiers, and because killing the other side's nobles was often a major strategic goal (think Game of Thrones). So a civil war usually further depletes the already-depleted peasant population, and finally depletes noble populations, leading to a general underpopulation and the beginning of the next cycle.



Combine these two processes, and you get the basic structure of a secular cycle. There are about a hundred years of unalloyed growth, as peasant and noble populations rebound from the last disaster. During this period, the economy is strong, the people are optimistic and patriotic, and the state is strong and united.

After this come about fifty years of “stagflation”. There is no more room for easy growth, but the system is able to absorb the surplus population without cracking. Peasants may not have enough land, but they go to the city in search of jobs. Nobles may not have enough of the positions they want, but they go to college in order to become bureaucrats, or join the retinues of stronger nobles. The price of labor reaches its lowest point, and the haves are able to exploit the desperation of the have-nots to reach the zenith of their power. From the outside, this period can look like a golden age: huge cities buzzing with people, universities crammed with students, ultra-rich nobles throwing money at the arts and sciences. From the inside, for most people it will look like a narrowing of opportunity and a hard-to-explain but growing sense that something is wrong.

After this comes a crisis. The mechanisms that have previously absorbed surplus population fail. Famine and disease ravage the peasantry. State finances fall apart. Social trust and patriotism disappear as it becomes increasingly obvious that it’s every man for himself and that people with scruples will be defeated or exploited by people without.

After this comes the depression period (marked “intercycle” on the graph above, but I’m going to stick with the book’s term). The graph makes it look puny, but it can last 100 to 150 years. During this period, the peasant population is low, but the noble population is still high. This is most likely an era of very weak or even absent state power, barbarian invasions, and civil war. The peasant population is in a good position to expand, but cannot do so because wars keep killing people off or forcing them into walled towns where they can’t do any farming. Usually it takes a couple more wars and disasters before the noble population has decreased enough to reverse elite overproduction. At this point the remaining nobles look around, decide that there is more than enough for all of them, and feel incentivized to cooperate with the formation of a strong centralized state.

This cycle is interwoven with a second 40-60 year process that T&N call the “fathers-and-sons cycle” or “bigenerational cycle”. The data tend to show waves of disorder about every 40-60 years. During the “integrative trend” (T&N’s term for the optimistic growth and stagflation phases), these can just be minor protests or a small rebellion that is easily crushed. During the “disintegrative trend” (crisis + depression), they usually represent individual outbreaks of civil war. For example, during the Roman Republic, the violence around the death of Tiberius Gracchus in 133 BC was relatively limited, because Rome had not yet entered its crisis phase. 40 years later, in the depths of the crisis phase, there was a second outbreak of violence (91 – 82 BC) including the Social War and Sulla’s wars, which escalated to full-scale (though limited) civil war. 40 years later there was a third outbreak (49 – 27 BC) including Caesar and Augustus’s very large civil wars. After that the new integrative trend started and further violence was suppressed.

In *Secular Cycles*, T&N mostly just identify this pattern from the data and don’t talk a lot about what causes it. But in some of Turchin’s other work, he applies some of the math used to model epidemics in public health. His model imagines three kinds of people: naives, radicals, and moderates. At the start of a cycle, most people are naive, with a few radicals. Radicals gradually spread radicalism, either by converting their friends or provoking their enemies (eg a terrorist attack by one side convinces previously disengaged people to join the other side). This spreads like any other epidemic. But as violence gets worse, some people convert to “moderates”, here meaning not “wishy-washy people who don’t care” but something more like “people disenchanted with the cycle of violence, determined to get peace at any price”. Moderates suppress radicals, but as they die off most people are naive and the cycle begins again. Using various parameters for his model Turchin claims this predicts the forty-to-sixty year cycle of violence observed in the data.

So this is the basic thesis of *Secular Cycles*. Pre-industrial history operates on two cycles: first, a three-hundred year cycle of the rise-and-fall of civilizations. And second, a 40-60 year cycle of violent disorder that only becomes relevant during the lowest parts of the first cycle.

II.

This is all in the first chapter of the book! The next eight chapters are case studies of eight different historical periods and how they followed the secular cycle model.

For example, Chapter 7 is on the Roman Empire. It starts with Augustus in 27 BC. The Roman Republic has just undergone a hundred years of civil war, from the Gracchi to Marius to Sulla to Pompey to Caesar to Antony. All of this decreased its population by 30% from its second-century peak. That means things are set to get a lot better very quickly.

The expansion phase of the Empire lasted from Augustus (27 BC) to Nerva (96 AD), followed by a stagflation phase from Nerva to Antonius Pius (165 AD). Throughout both phases, the population grew – from about 40 million in Augustus' day to 65 million in Antonius'. Wheat prices stayed stable until Nerva, then doubled from the beginning of the second century to its end. Legionary pay followed the inverse pattern, staying stable until Nerva and then decreasing by a third before 200. The finances of the state were the same – pretty good until the late second century (despite occasional crazy people becoming Emperor and spending the entire treasury building statues of themselves), but cratering during the time of Marcus Aurelius and Commodus (who debased the denarius down to only 2 g silver).

Throughout expansion and stagflation, the Empire was relatively peaceful (the "Pax Romana"). Sure, occasionally a crazy person would become Emperor and they would have to kill him. There was even one small civil war which lasted all of a year (69 AD). But in general, these were isolated incidents.

Throughout the expansion phase, upward mobility was high and income inequality relatively low. T&N measure this as how many consuls (the highest position in the Roman governmental hierarchy) had fathers who were also consuls. This decreased throughout the first century – from 46% to 18% – then started creeping back up during the stagflation phase to reach 32% at the end of the second century.

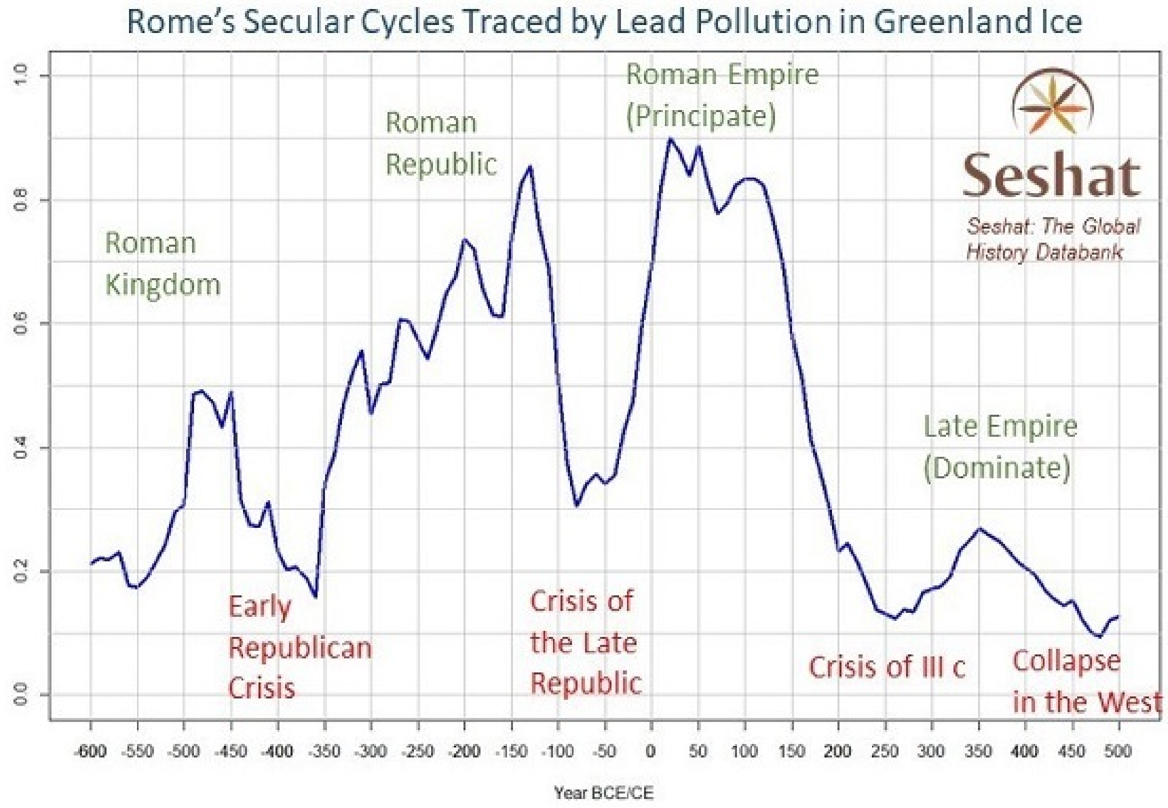
The crisis phase began in 165 AD at the peak of Rome's population and wealth. The Antonine Plague ravaged the Empire, killing 30% of the population. Fifteen years later, the century-long dominance of the Good Emperors ended, and Commodus took the throne. Then he was murdered and Pertinax took the throne. Then he was murdered and Didius Julianus took the throne. Then he was murdered and Septimius Severus took the throne.

Now we are well into the disintegrative trend, and the shorter 40-60 year cycle comes into play. Septimius Severus founds a dynasty that lasts 41 years, until Septimius Alexander (the grandson of Septimius Severus' sister-in-law; it's complicated) was assassinated by his own soldiers in Germany. This begins the Crisis Of The Third Century, a time of constant civil war, mass depopulation, and economic collapse. The Five Good Emperors of the second century ruled 84 years between them (average of 17 years per emperor). The fifty year Crisis included 27 emperors, for an average of less than 2 years per emperor.

Finally, in 284, Emperor Diocletian ended the civil wars, re-established centralized authority, and essentially refounded the Roman Empire – a nice round 310 years after Augustus did the same. T&N mark this as the end of a secular cycle and the beginning of a new integrative trend.

T&N are able to tell this story. But they don't just tell the story. They are able to cite various statistics to back themselves up. The Roman population statistics. The price of wheat and other foodstuffs. The average wages for laborers. They especially like coin hoards – the amount of buried treasure from a given period discovered by treasure-hunters – because they argue you only bury your money during times of instability, so this forms a semi-objective way of measuring how unstable things are.

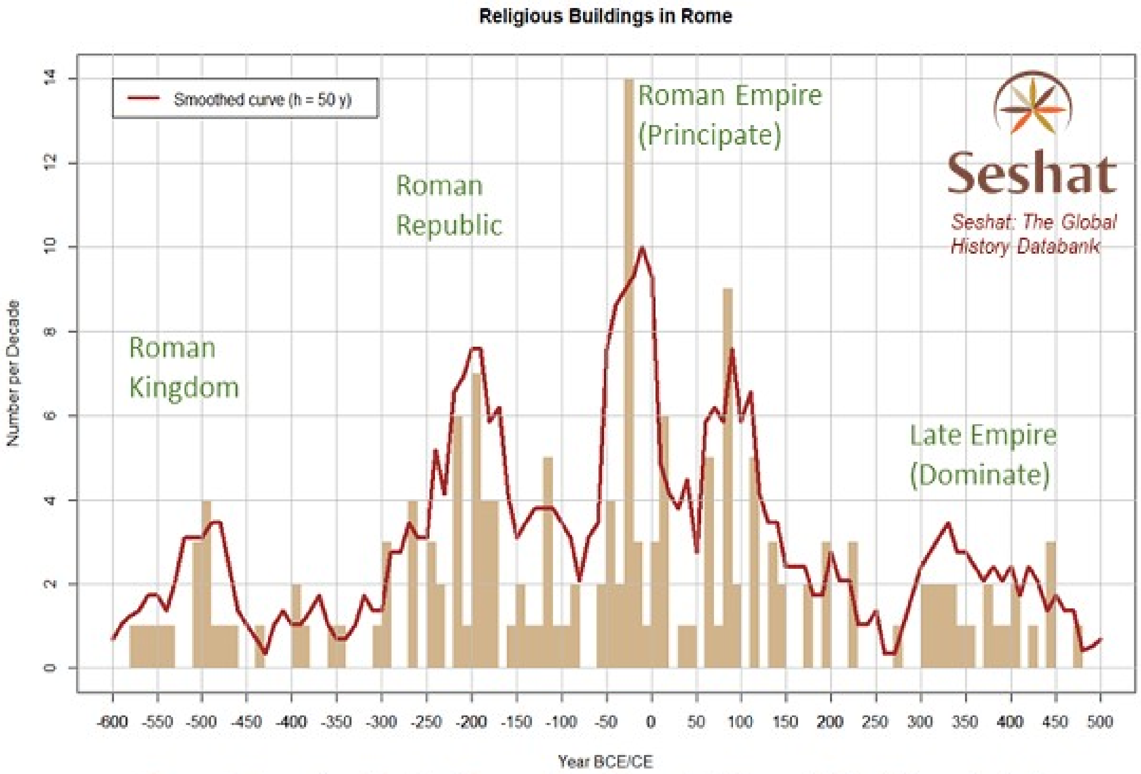
They are at their best when presenting very broad summary statistics. For example, Roman industry produced vast amounts of lead, which entered the atmosphere and settled into the Greenland ice sheet. Here is Roman lead output per year as measured in ice cores:



This shows four peaks for the four cycles T&N identify in Rome: the Kingdom, the Republic, the Early Empire of Augustus (Principate, the one described above), and the Late

Empire of Diocletian (Dominate). It even shows a sawtooth-y pattern corresponding to the shorter bigenerational cycles.

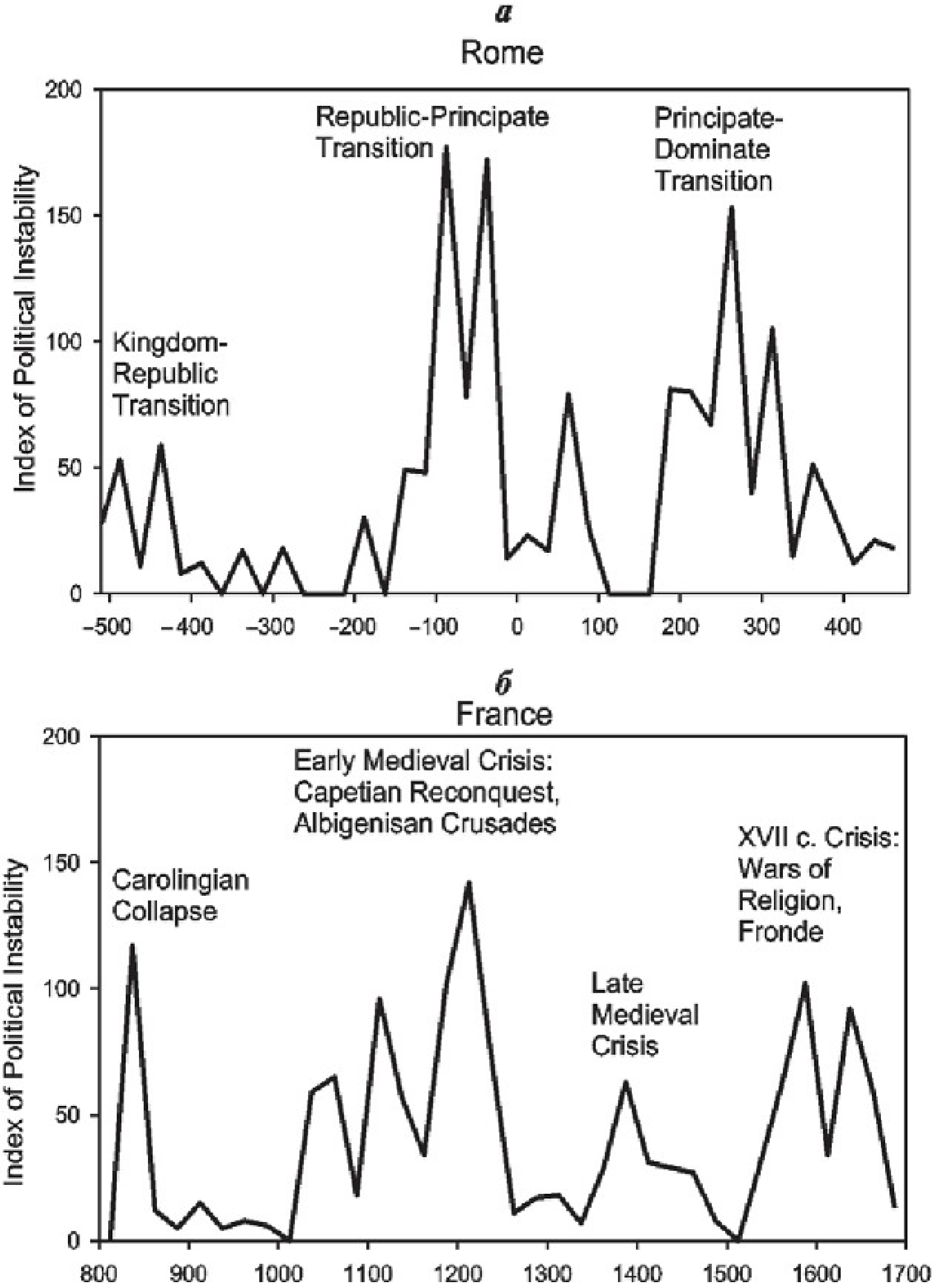
Or here is building activity in Rome, measured by how many buildings archaeologists have found from a given time:



Rome's Secular Cycles Traced by Temple/Church Building Activity

This is a little less perfect (why is there a big gap in the middle of the Principate? I guess Augustus is a hard act to follow, building-wise) but it still looks good for the cycle theory.

And here is an Index Of Political Instability, which "combines measures of duration, intensity, and scale of political instability events, coded by a team of professional historians":


а
Rome
Index of Political Instability
Kingdom-
Republic
Transition
Republic-Principate
Transition
Principate-
Dominate
Transition
200
150
100
50
0
−500
−400
−300
−200
−100
0
100
200
300
400
б
France
Index of Political Instability
Carolingian
Collapse
Early Medieval Crisis:
Capetian Reconquest,
Albigenisan Crusades
Late
Medieval
Crisis
XVII c. Crisis:
Wars of
Religion,
Fronde
200
150
100
50
0
800
900
1000
1100
1200
1300
1400
1500
1600
1700

Rome is the one on top. Instability clearly peaks during the crisis-depression phases between T&N's secular cycles – again with a sawtooth pattern representing the bigenerational cycles.

III.

Seeing patterns in random noise is one of the basic human failure modes. *Secular Cycles* is so *prima facie* crackpottish that it should require mountains of data before we even start wondering if it *might* be true. I want to make it clear that the book – plus Turchin individually in some of his other books and papers – provides these mountains. I can't show every single case study, graph, and table in this book review. But the chapter above on the Roman Principate included 25 named figures and graphs, plus countless more informal presentations of data series, from "real wages of agricultural laborers in Roman Egypt during the second century" to "mean annual real land rents for wheat fields in artabas per aroura, 27 BC to 268 CE" to "imperial handouts per reign-year" to "importation of African red slip ware into the Albegna Valley of Etruria, 100 – 600". And this is just one chapter, randomly chosen. There are seven others just like this. This book understands the burden of proof it is under, and does everything it can to meet it.

Still, we should be skeptical. How many degrees of freedom do T&N have, and is it enough to undermine their case?

First, they get some freedom in the civilizations they use as case studies. They could have searched through every region and period and cherry-picked eight civilizations that rose and fell over a periods of three hundred years. Did they? I don't think so. The case studies are England, France, Rome, and Russia. These are some of the civilizations of greatest interest to the English-speaking world (except Russia, which makes sense in context because the authors are both

Russian). They’re also some of the civilizations best-studied by Anglophone historians and with the most data available (the authors’ methodology requires having good time-series of populations, budgets, food production, etc).

Also, it’s not too hard to look at the civilizations they didn’t study and fill in the gaps. The book barely mentions China, but it seems to fit the model pretty well (“the empire united longs to divide; divided longs to unite”). In fact, taking the quotation completely seriously – the empire was first united during the Qin Dynasty starting in 221 BC, which lasted only 20 years before seguing into the Han Dynasty in 202 BC. The Han expanded and prospered for about a century, had another century of complicated intrigue and frequently revolt, and then ended in disaster in the first part of the first century, with a set of failed reforms, civil war, the sack of the capital, some more civil war, peasant revolt, and even more civil war. The separate period of the Eastern Han Dynasty began in 25 AD, about 240 years after the beginning of the Qin-Han cycle. The Eastern Han also grew and prospered for about a hundred years, then had another fifty years of simmering discontent, then fell apart in about 184 AD, with another series of civil wars, peasant rebellions, etc. This was the Three Kingdoms Period during which “the empire united longs to divide, divided longs to unite” was written to describe. It lasted another eighty years until 266 AD, after which the Jin Dynasty began. The Jin Dynasty was kind of crap, but it lasted another 180 years until 420, followed by 160 years of division, followed by the Sui and Tang dynasties, which were not crap. So I don’t think it takes too much pattern-matching to identify a Western-Han-to-Eastern-Han Cycle of 240 years, followed by an Eastern-Han-to-Jin Cycle of 241 years, followed by a Jin-to-Sui/Tang-Cycle of 324 years.

One could make a more hostile analysis. Is it really fair to lump the Western Jin and Eastern Jin conveniently together, but separate the Western Han and Eastern Han conveniently apart? Is it really fair to call the crappy and revolt-prone Jin Dynasty an “integrative trend” rather than a disintegrative trend that lasted much longer than the theory should predict? Is it really fair to round off cycles of 240 and 320 years to “basically 300 years”?

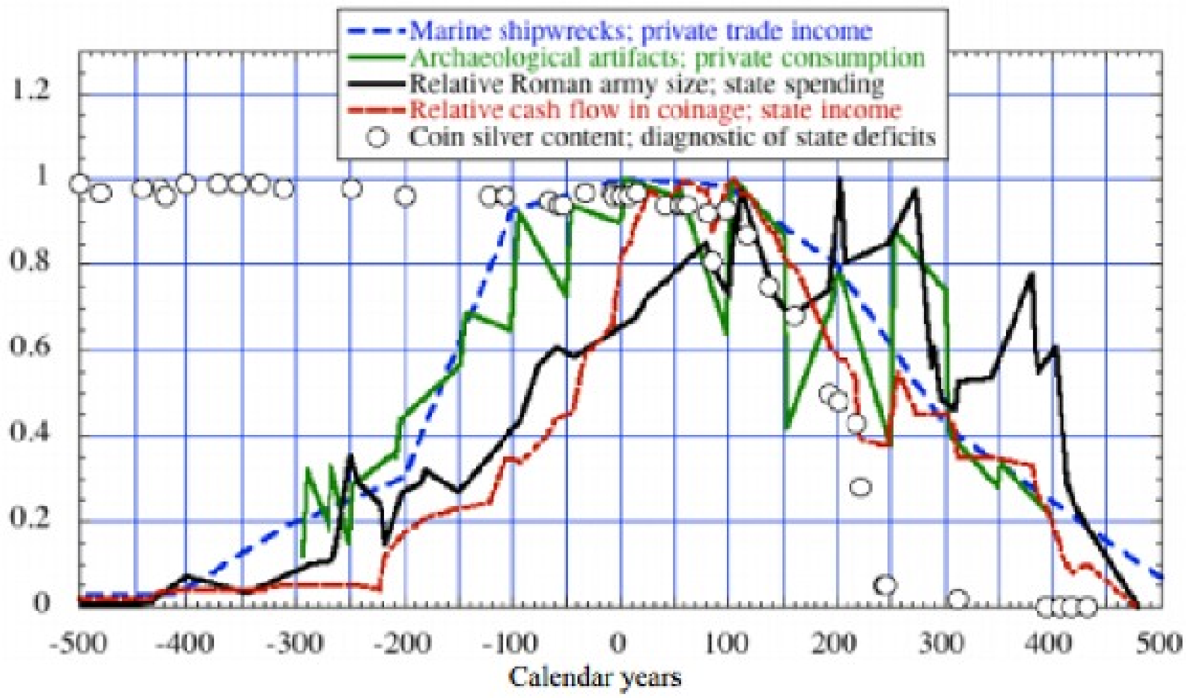
I think the answer to all of these is “T&N aren’t making predictions about the length of Chinese dynasties, they’re making predictions about the nature of secular cycles, which are correlated with dynasties but not identical to them”. If I had the equivalent to lead core readings for China, or an “instability index”, or time series data for wages or health or pottery importation or so on, maybe it would be perfectly obvious that the Eastern and Western Han defined two different periods, but the Eastern and Western Jin were part of the same period – the same way one look at the lead core data for Rome shows that the Julio-Claudian dynasty vs. the Flavian Dynasty is not an interesting transition.

A secondary answer might be that T&N admit all sorts of things can alter the length of secular cycles. They tragically devote only a few pages to “Ibn Khaldun cycles”, the theory of 14th century Arabic historian Ibn Khaldun that civilizations in the Maghreb rise and fall on a one hundred year period. But they discuss it just enough to say their data confirm Ibn Khaldun’s observations. The accelerated timescale (100 vs. 300 years) is because the Maghreb is massively polygynous, with successful leaders having harems of hundreds of concubines. This speeds up the elite overproduction process and makes everything happen in fast-forward. T&N also admit that their theory only describes civilizations insofar as they are self-contained. This approximately holds for hegemons like Rome at its height, but fails for eg Poland, whose history is going to be much more influenced by when Russia or Germany decides to invade than by the internal mechanisms of Polish society. Insofar as external shocks – whether climatic, foreign military, or whatever else – affect a civilization, secular cycles will be stretched out, compressed, or just totally absent.

This sort of thing must obviously be true, and it’s good T&N say it, but it’s also a free pass to add as many epicycles as you need to explain failure to match data. All I can say looking at China is that, if you give it some wiggle room, it seems to fit T&N’s theories okay. The same is true of a bunch of other civilizations I plugged in to see if they would work.

Second, T&N get some degrees of freedom based on what statistics they use. In every case, they present statistics that support the presence of secular cycles, but they’re not the same statistics in every case. On the one hand, this is unavoidable; we may not have good wage data for every civilization, and levels of pottery importation might be more relevant to ancient Rome than to 19th-century Russia. On the other hand, I’m not sure what prevents them from just never mentioning the Instability Index if the Instability Index doesn’t show what they want it to show.

Here are some random Rome-related indicators I found online:



None of them show the same four-peaked Kingdom-Republic-Principate-Dominate pattern as the ones *Secular Cycles* cites, or the ones Turchin has online.

Third, a lot of the statistics themselves have some degrees of freedom. A lot of them are things like “Instability Index” or “Index of Social Well-Being” or “General Badness Index”. These seem like the kind of scores you can fiddle with to get the results you want. Turchin claims he hasn’t fiddled with them – his instability index is taken from a 1937 paper I haven’t been able to find. But how many papers like that are there? Am I getting too conspiratorial now?

Likewise, we don’t have direct access to the budget of the Roman Empire (or Plantagenet England, or…). Historians have tried to reconstruct it based on archaeology and the few records that have survived. T&N cite these people, and the people they cite are at the top of their fields and say what T&N say they say. But how much flexibility did they have in deciding which estimate of the Roman budget to cite? Is there enough disagreement that they could cite the high estimate for one period and the low estimate for another, then prove it had gone down? I don't know (though a few hours' work ought to be enough to establish this).

I wish I could find commentary by other academics and historians on *Secular Cycles*, or on Turchin's work more generally. I feel like somebody should either be angrily debunking this, or else throwing the authors a ticker-tape parade for having solved history. Neither is happening. The few comments I can find are mostly limited to navel gazing about whether history should be quantitative or qualitative. The few exceptions find are blog posts by people I already know and respect urging me to read Turchin five years ago, advice I am sorry for not taking. If you know of any good criticism, please tell me where to find it.

Until then, my very quick double-checking suggests T&N are pretty much on the level. But there could still be subtler forms of overfitting going on that I don't know enough about history to detect.

IV.

If this is true, does it have any implications for people today?

First, a very weak implication: it makes history easier to learn. I was shocked how much more I remembered about the Plantagenets, Tudors, Capetians, etc after reading this book, compared to reading any normal history book about them. I think the secret ingredient is structure. If history is just "one damn thing after another", there's no framework for figuring out what matters, what's worth learning, what follows what else. The secular cycle idea creates a structure that everything fits into neatly. I know that the Plantagenet Dynasty lasted from 1154 – 1485, because it *had to*, because that's a 331 year secular cycle. I know that the important events to remember include the Anarchy of 1135 – 1153 and the War of the Roses from 1455 – 1487, because those are the two crisis-depression periods that frame the cycle. I know that after 1485 Henry Tudor took the throne and began a new age of English history, because that’s the beginning of the integrative phase of the next cycle. All of this is a lot easier than trying to remember these names and dates absent any context. I would recommend this book for that reason alone.

Second, I think this might give new context to Piketty on inequality. T&N describe inequality as starting out very low during the growth phase of a secular cycle, rising to a peak during the stagflation phase, then dropping precipitously during the crisis. Piketty describes the same: inequality rising through the peaceful period of 1800 to 1900, dropping precipitously during the two World Wars, then gradually rising again since then. This doesn’t make a huge amount of sense, since I’m not sure you can fit the post industrial world into secular cycles. But I notice Piketty seems to think of this as a once-off event – inequality has been rising forever, broken only by the freak crisis of the two World Wars – and it’s interesting to read T&N talk about the exact same process recurring again and again throughout history.

Finally, and most important: is there any sense in which this is still going on?

The easiest answer would be no, there isn’t. The secular cycles are based around Malthusian population growth, but we are now in a post-Malthusian regime where land is no longer the limiting resource. And the cycles seem to assume huge crises killing off 30% to 50% of the population, but those don’t happen anymore in First World countries; the Civil War was the bloodiest period of US history, and even it only killed 2% of Americans. Even Germany only lost about 15% of its population in World Wars I + II.

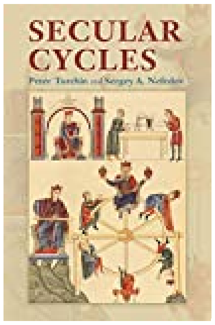
But Turchin has another book, *Ages Of Discord*, arguing that they do. I have bought it and started it and will report back when I'm done.

Even without a framework, this is just interesting to think about. In popular understanding of American history, you can trace out optimistic and pessimistic periods. The national narrative seems to include a story of the 1950s as a golden age of optimism. Then everyone got angry and violent in the early 1970s (the Status 451 review of Days Of Rage is pretty great here, and reminds us that "people have completely forgotten that in 1972 we had over nineteen hundred domestic bombings in the United States"). Then everything suddenly got better once Reagan declared "morning in America" in the 1980s, with an era of optimism and good feelings lasting through the Clinton administration. Then things starting to turn bad sometime around Bush II. And now everybody hates each other, and fascists and antifa are fighting in the streets, and people are talking about how "civility" and "bipartisanship" are evil tools of oppression, and PredictIt says an avowed socialist has a 10% chance of becoming president of the US. To what extent is this narrative true? I don't know, but it's definitely the narrative.

One thing that strikes me about T&N's cycles is the ideological component. They describe how, during a growth phase, everyone is optimistic and patriotic, secure in the knowledge that there is enough for everybody. During the stagflation phase, inequality increases, but *concern about* inequality increases even more, zero-sum thinking predominates, and social trust craters (both because people are actually defecting, and because it's in lots of people's interest to play up the degree to which people are defecting). By the crisis phase, partisanship is much stronger than patriotism and radicals are talking openly about how violence is ethically obligatory.

And then, eventually, things get better. There is a new Augustan Age of virtue and the reestablishment of all good things. This is a really interesting claim. Western philosophy tends to think in terms of trends, not cycles. We see everything going on around us, and we think this is some endless trend towards more partisanship, more inequality, more hatred, and more state dysfunction. But *Secular Cycles* offers a narrative where endless trends can end, and things can get better after all.

Of course, it also offers a narrative where sometimes this process involves the death of 30% – 50% of the population. Maybe I should read Turchin’s other books before speculating any further.



Attempted Replication: Does Beef Jerky Cause Manic Episodes?

Last year, a study came out showing that beef jerky and other cured meats, could trigger mania in bipolar disorder (paper, popular article). It was a pretty big deal, getting coverage in the national press and affecting the advice psychiatrists (including me) gave their patients.

The study was pretty simple: psychiatrists at a mental hospital in Baltimore asked new patients if they had ever eaten any of a variety of foods. After getting a few hundred responses, they compared answers to controls and across diagnostic categories. The only hit that came up was that people in the hospital for bipolar mania were more likely to have said they ate dry cured meat like beef jerky (odds ratio 3.49). This survived various statistical comparisons and made some biological sense.

The methodology was a little bit weird, because they only asked if they'd *ever* had the food, not if they'd eaten a lot of it just before becoming sick. If you had beef jerky once when you were fourteen, and ended up in the psych hospital when you were fifty-five, that counted. Either they were hoping that “ever had beef jerky at all” was a good proxy for “eats a lot of beef jerky right now”, or that past consumption produced lasting changes in gut bacteria. In any case, they found a strong effect even after adjusting for confounders and doing the necessary Bonferroni corrections, so it's hard to argue with success.

Since the study was so simple, and already starting to guide psychiatric practice, I decided to replicate it with the 2019 Slate Star Codex survey.

In a longer section on psychiatric issues, I asked participants “Have you ever been hospitalized for bipolar mania?”. They could answer “Yes, many times”, “Yes, once”, or “No”. 3040 people answered the question, of whom 26 had been hospitalized once, 13 many times, and 3001 not at all.

I also asked participants “How often do you eat beef jerky, meat sticks, or other similar nitrate-cured meats?”. They could answer “Never”, “less than once a year”, “A few times a year”, “A few times a month”, A few times a week”, or “Daily or almost daily”. 5,334 participants had eaten these at least once, 2,363 participants had never eaten them.

(for the rest of this post, I’ll use “beef jerky” as shorthand for this longer and more complicated question)

Power calculation: the original study found odds ratio of 3.5x; because the percent of my sample who had been hospitalized for mania was so low, OR = RR; I decided to test for an odds ratio of 3. About 1.2% of non-jerky-eaters had been hospitalized for mania, so I used this site to calculate necessary sample size with Group 1 as 1.2%, Group 2 as 3.6% (=1.2×3), enrollment ratio of 0.46 (ratio of the 921 jerky-never-eaters to 2015 jerky eaters), alpha of 0.05, and power of 80%. It recommended a total sample of 1375, well below the 2974 people I had who answered both questions.

Of 932 jerky non-eaters, 11 were hospitalized for mania, or 1.2%. Of 2042 jerky-eaters, 27 were hospitalized for mania, or 1.3%. Odds ratio was 1.12, chi-square statistic was 0.102, p = 0.75. The 95% confidence interval was (.55, 2.23). So there was no significant difference in mania hospitalizations between jerky-eaters and non-eaters.

I also tried to do the opposite comparison, seeing if there was a difference in beef jerky consumption between people with a history of hospitalization for mania and people without such a history. I recoded the “beef jerky” variable to a very rough estimate to how many times per year people ate jerky (“never” = 0, “daily” = 400, etc). The rough estimate wasn’t very principled, but I came up with my unprincipled system before looking at any results. People who had never been hospitalized for mania ate beef jerky an average of 16 times per year; people who had been hospitalized ate it an average of 8 times per year. This is the opposite direction predicted by the original study, and was not significant.

I tried looking at people who had a bipolar diagnosis (which requires at least one episode of mania or hypomania) rather than just people who had been hospitalized for bipolar mania. This gave me four times the sample size of bipolar cases, but there was still no effect. 63% of cases (vs. 69% of controls) had ever eaten jerky, and cases on average ate jerky 15 times a year (compared to 20 times for controls). Neither of these findings was significant.

Why were my survey results so different from the original paper?

My data had some serious limitations. First, I was relying on self-report about mania hospitalization, which is less reliable than catching manic patients in the hospital. Second, I had a much smaller sample size of manic patients (though a larger sample size of controls). Third, I had a different population (SSC readers are probably more homogenous in terms of class, but less homogenous in terms of nationality) than the original study, and did not adjust for confounders.

There were also some strengths to this dataset. I had a finer-grained measure of beef jerky consumption than the original study. I had a larger control group. I was able to be more towards the confirmatory side of confirmatory/exploratory analysis.

Despite the limitations, there was a pretty striking lack of effect for jerky consumption. This is despite the dataset being sufficiently well-powered to confirm other effects that are classically known to exist (for example, people hospitalized by mania had higher self-rated childhood trauma than controls, $p < 0.001$). This is an important finding and should be easy to test by anyone with access to psychiatric patients or who is surveying a large population. I urge other people (hint to psychiatry residents reading this blog who have to do a research project) to look into this further. I welcome people trying to replicate or expand on these results. All of the data used in this post are freely available and can be downloaded here.

Maybe Your Zoloft Stopped Working Because A Liver Fluke Tried To Turn Your Nth-Great-Grandmother Into A Zombie

Or at least this is the theory proposed in Brain Evolution Through The Lens Of Parasite Manipulation by Marco del Giudice.

The paper starts with an overview of parasite manipulation of host behavior. These are the stories you hear about toxoplasma-infected rats seeking out cats instead of running away from them, or zombie ants climbing stalks of grass so predators will eat them. The parasite secretes chemicals that alter host neurochemistry in ways that make the host get eaten, helping the parasite transfer itself to a new organism.

Along with rats and ants, there is a dizzying variety of other parasite manipulation cases. They include parasitic wasps who hack spiders into forming protective webs for their pupae, parasitic flies that cause bees to journey far from their hive in order to spread fly larva more widely, and parasitic microorganisms that cause mosquitoes to draw less blood from each victim (since that forces the mosquitoes to feed on more victims, and so spread the parasite more widely). Parasitic nematodes make their ant hosts turn red, which causes (extremely stupid?) birds to mistake them for fruit and eat them. Parasitic worms make crickets seek water; as the cricket drowns, the worms escape into the pond and begin the next stage of their life cycle. Even mere viruses can alter behavior; the most famous example is rabies, which hacks dogs, bats, and other mammals into hyperaggressive moods that usually result in them biting someone and transmitting the rabies virus.

Even our friendly gut microbes might be manipulating us. People talk a lot about the “gut-brain axis” and the effect of gut microbes on behavior, as if this is some sort of beautiful symbiotic circle-of-life style thing. But scientists have found that gut microbes trying to colonize fruit flies will hack the flies’ food preferences to get a leg up – for example, a carb-metabolizing microbe will secrete hormones that make the fly want to eat more carbs than fat in order to outcompete its fat-metabolizing rivals for gut real estate; there are already papers speculating that the same processes might affect humans. Read Alcock 2014 and you will never look at food cravings the same way again.

But del Giudice thinks this is just the tip of the iceberg. Throughout evolutionary history, parasites have been trying to manipulate host behavior and hosts have been trying to avoid manipulation, resulting in an eons-long arms race. The equilibrium is what we see today: parasite manipulation is common in insects, rare in higher animals, and overall of limited importance. But in arms race dynamics, the current size of the problem tells you nothing about the amount of resources invested in preventing the problem. There is zero problem with war between Iran and Saudi Arabia right now, but both sides have invested billions of dollars in military supplies to keep their opponent from getting a leg up. In the same way, just because mammals usually avoid parasite behavior manipulation *now* doesn’t mean they aren’t on a constant evolutionary war footing.

So if you’re an animal at constant risk of having your behavior hijacked by parasites, what do you do?

First, you make your biological signaling cascades more complicated. You have multiple redundant systems controlling every part of behavior, and have them interact in ways too complicated for any attacker to figure out. You have them sometimes do the opposite of what it looks like they should do, just to keep enemies on their toes. This situation should sound very familiar to anyone who's ever studied biology.

Del Giudice compares the neurosignaling of the shrimp-like gammarids (small, simple, frequently hijacked by parasites) to rats (large, complex, hard to hijack). Gammarids have very simple signaling: high serotonin means "slow down", low serotonin means "speed up". The helminths that parasitize gammarids secrete serotonin, and the gammarids slow down and get eaten, transferring the parasite to a new host. Biologists can replicate this process; if they inject serotonin into a gammarid, the gammarid will slow down in the same way.

Toxoplasma hijacks rats and makes them fearless enough to approach cats. Dopamine seems to be involved somehow. But researchers injecting dopamine into rats don't get the same result; in fact, this seems to make rats avoid cats *more*. Maybe toxoplasma started by increasing dopamine, rats evolved a more complicated signaling code, and toxoplasma cracked the code and now increases dopamine plus other things we don't understand yet.

Aside from the brain, the immune system is the most important target to secure, so this theory should predict that immune signaling will also be unusually inscrutable. Again, this situation should sound very familiar to anyone who's ever studied biology.

Second, you have a bunch of feedback loops and flexibility ready to deploy at any kind of trouble. If something makes dopamine levels go up, you decrease the number of dopamine receptors, so that overall dopaminergic neurotransmission is the same as always. If something is making you calmer than normal, you have some other system ready to react by making you more anxious again.

Del Giudice makes the obvious connection to psychopharmacology. Many psychoactive drugs build tolerance quickly: for example, heroin addicts constantly need higher and higher doses to get their "hit". Further, tolerance builds in a pattern weirdly similar to antibody response – it takes a while to build up a cocaine tolerance, and you lose it over time if you don't use cocaine, but the body "remembers" the process and a single hit of cocaine years later is sufficient to bring you back up to the highest tolerance level you've ever had.

The standard explanation for tolerance is that it's an attempt to maintain homeostasis against the sort of conditions that can cause natural variation in neurotransmitter levels. I never questioned this before. But why is the body prepared to suddenly have all its serotonin reuptake transporters inhibited? Is that something that frequently happens, out in nature? I guess maybe plant toxins could do that, but then how come the body is prepared to deal with this for months or years?

While not denying the value of these standard explanations, Del Giudice thinks defense against parasite behavior manipulation may also play a role. Remember, gammarids absolutely have parasites that try to increase their serotonin levels as a prelude to getting them killed. Is it that surprising that a lot of different animal lineages would develop a reaction of "If something other than normal cognition has started increasing your serotonin levels, it’s a trap and you need to get them back down again”? Does that explain why SSRIs don’t work for some people, or randomly stop working, or need frequent dose escalation?

Third, you encode messages in the timing of pulses. This is a central feature of neuroendocrine communication – an intense pulse of testosterone at 6 AM means something different from tonically high testosterone all day. Parasites cannot do pulses. Remember, these parasites are usually microscopic. Each parasite can only produce a miniscule quantity of neurotransmitter or hormone. Only colonies of thousands or millions of parasites can produce enough chemicals to affect host signaling. This parasites cannot communicate or coordinate with each other, so there’s no way for them to be producing lots of testosterone one minute and none at all the next. That means that when a hormone arrives in a pulse, or better yet a complicated pattern of pulses, that’s a pretty reliable sign that it’s coming from a real gland.

Fourth, you exploit your individuality. The immune system already does this; there are some genes called the major histocompatibility complex that are designed to be especially variable, such that most people (except identical twins) will have different MHCs. These help the immune system differentiate self from other. Because they have such high individual variability, pathogens can’t just evolve around the MHC; they would have to undergo an entire evolutionary process for each new host they invade.

Del Giudice wonders if parasite-host arms races created pressure for increased human variability. SSRIs will make some people less depressed. But some people will get more depressed. A few will even get suicidal. A very few will flip out and become psychotic, or improve much more quickly than the textbooks say should be possible and feel completely reborn on day 3, or have something else even weirder happen. I always assumed God just hated psychiatrists and wanted them to be miserable. But another possibility is that extreme individual variability in neurosignaling pathways is a defense against parasite manipulation. If the effects of serotonin are unpredictable for any individual, no parasite species can devise a universally valid mechanism for controlling its hosts.

Fifth, you let the parasites become part of the furniture. If everybody in your ecosystem is infected with a parasite that raises serotonin, you just evolve a tonically lower serotonin level, and then it all cancels out. This one seems a little bit weird to me – surely this isn’t the stable equilibrium? But:

> A downside of preemptive strategies is evolved dependence (de Mazancourtet al. 2005): if brain physiology and behavior are designed to function optimally when the parasite is present, the absence of the parasite will lead to inappropriate or fitness-reducing behaviors (Weinersmith and Earley 2016; see also Johnson and Foster 2018).

I think this is meant to hint at the “hygiene hypothesis”, ie our immune systems are screwed up because we are not getting exposed to the parasites it was built to expect. Suppose lots of parasites try to downregulate the immune system (which sounds logical enough), and the body doesn’t know which ones it’s going to get but expects it to follow a Poisson distribution around some mean. Then it might just upregulate the activity of the immune system that same amount. If you get rid of all the parasites, then your immune system is just set too high and you get autoimmune disorders.

(in case you had the same question I did – yes, the parasitologist Kelly Weinersmith cited above is the same Kelly Weinersmith who co-wrote *Soonish* with Zach Weinersmith of SMBC fame.)

Sixth, you use antiparasitic drugs as neurotransmitters. This is the kind of murderous-yet-clever solution I expect of evolution, and it does not disappoint. Several neurotransmitters, including neuropeptide Y, neurokinin A, and substance P are pretty good antimicrobials. The assumption has always been that the body kills two birds with one stone, getting its signaling done and also having some antimicrobials around to take out stray bacteria. But Del Giudice proposes that this is to prevent parasites from hijacking the signal; any parasite that tried to produce or secrete an antiparasitic drug would die in the process.

Dopamine is mildly toxic. The body is usually pretty good at protecting itself, but the mechanism fails under stress; this is why too much methamphetamine rots your brain. Why would you use a toxic chemical as a neurotransmitter? For the same reason you would use antiparasitic drugs – because you want to kill anything smaller than you that tries to synthesize it.

People always talk about the body as a beautiful well-oiled machine. But sometimes the body communicates with itself by messages written with radioactive ink on asbestos-laced paper, in the hopes that it's killing itself *slightly* more slowly than it's killing anyone who tries to send it fake messages. Honestly it is a miracle anybody manages to stay alive at all.

All these features together are a pretty effective way of dealing with parasite manipulation. There are a few parasites that can manipulate human behavior – rabies definitely, toxoplasma maybe – but overall we are remarkably safe.

Del Giudice argues that a combination of factors make it easy for parasites to manipulate insects but not large vertebrates. First, insects are small, so you only need a few parasites to produce an insect-sized level of neurotransmitter. Second, insects are so simple that usually one neurotransmitter maps nicely to one behavior; they are too small to support multiple redundant systems or complicated signal cascades. Del Giudice writes:

> Although parasites can evolve subtler and more indirect means of manipulation, their computational capabilities are ultimately limited by their size. As the size and complexity of the host's brain increase relative to the parasite, the disparity may become so extreme that the host is able to "outcompute" its adversary, making complex manipulations effectively impossible. The parasite may still be able to alter the host's behavior in nonspecific ways (e.g., sickness, brain damage), but is unable to induce the kind of coordinated pattern required for trophic transmission or bodyguard manipulation. Although this argument is admittedly speculative, it is consistent with the fact that complex behavioral manipulations have not been documented in larger, warm-blooded animals (see Lafferty and Kuris 2002).

Finally, almost nothing eats humans, so there aren't a lot of parasites interested in using us as a vehicle to get to their definitive hosts. If parasites want anything from us, it's probably STIs wishing we had more risky sex; accordingly, Del Giudice obliquely cites Greg Cochran's controversial hypothesis that homosexuality may be related to parasites hijacking sexual machinery.

But let’s take a step back: is any of this true?

The strongest evidence against is the dog that didn’t bark. Some systems look heavily defended against parasite manipulation, but others don’t. Amphetamines raise dopamine effectively and without significant tolerance buildup (see part IV here for a defense of this claim); antipsychotics lower dopamine equally effectively and consistently. Since dopamine is one of the most lucrative systems for parasites to hijack, it’s surprising to find it so easy to affect. And what about immune function? Externally administered corticosteroids decrease immune activity and make the body more vulnerable to infection; why don’t parasites secrete them? Why don’t we have some counter against them? These systems look consistent with an evolutionary history in which we don’t expect any threat from parasite manipulation and don’t need to defend ourselves very hard.

But also: homeostasis might be the most basic activity of all living things. Every bodily system can be modeled as a striving for homeostasis in some domain or other, even high-level cognitive functions. So it’s not clear that tolerance to psychiatric drugs needs a complicated evolutionary explanation beyond just “if you increase serotonin, your body is going to try to decrease it again, because *that’s what bodies do*“.

So I’m not sure how much of an effect this really had. It’s an interesting theory. But whether it explains some things, nothing, or everything, it’s too early to say.

But I like this paper because it takes the complexity of biology seriously. There’s a sense that science is stagnating, and biology is one of the worst offenders. In the 1800s and early 1900s, we were pinning down our mastery of anatomy, discovering all the major hormone systems, learning about microbes and inventing antibiotics. It seemed like the same kind of thing as physics, where you could go out into the world, observe things, and make difficult but fundamentally straightforward discoveries. But for the past fifty years, it’s been kind of a mess. Despite some amazing work by amazing people, we still don’t even understand questions as basic as what depression is. Everything seems bogged down in a million different opaque signaling cascades that fight off any effort to untangle or shift them.

Del Giudice offers a seductive explanation: the perceived perversity of the human blueprint is absolutely real. Parts of it – the parts most involved in health and disease – were sculpted by evolution to be as hard as possible to understand or affect. This makes me feel better about how often the drugs I prescribe fail in surprising ways

Don’t Fear The Simulators

From the *New York Times*: Are We Living In A Computer Simulation? Let’s Not Find Out.

It lists the standard reasons for thinking we might be in a simulation, then brings up some proposals for testing the hypothesis (for example, the cosmic background radiation might look different in simulations and real universes). But it suggests that we not do that, because if we learn we’re in a simulation, that might ruin the simulation and cause the simulators to destroy the universe.

But I think a little more thought suggests we don’t have anything to worry about.

In order to notice we had discovered our simulated nature, the simulators would have to have a monitor watching us. We should expect this anyway. Although humans may run some simulations without monitoring them carefully, the simulators have no reason to be equally careless; if they can simulate billions of sentient beings, their labor costs are necessarily near zero. Such a monitor would have complete instantaneous knowledge of everything happening in our universe, and since anyone who can simulate a whole planet must have really good data processing capabilities, it would be able to understand and act upon the entire content of its omniscient sensorium. It would see the fall of each sparrow, record the position of ever atom, have the level of situational awareness that gods could only dream of.

What I’m saying is, it probably reads the *New York Times*.

That means it knows these experiments are going to happen. If it cares about the results, it can fake them.

Assuming for some reason that it messed up designing the cosmic background radiation (why are we assuming this, again?), it can correct that mistake now, or cause the experimental apparatus to report the wrong data, or do one of a million other things that would prevent us from learning we are in a simulation.

The *Times’* argument requires that simulators are so powerful that they can create entire universes, so on-top-of-things that they will know the moment we figure out their game – but also so incompetent that they can’t react to a warning published several years in advance in America’s largest newspaper.

There’s another argument for the same conclusion: the premises of the simulation argument suggest this isn’t the simulators’ only project. Each simulator civilization must simulate thousands or millions of universes. Presumably we’re not the first to think of checking the cosmic background radiation. Do you think the simulators just destroy all of them when they reach radio-wave-technology, and never think about fixing the background radiation mismatch or adding in some fail-safe to make sure the experiments return the wrong results?

For that matter, this is probably a stage every civilization goes through, *including whatever real civilization we are supposed to simulate*. What good is a simulation that can replicate every aspect of the real world *except* its simulation-related philosophy? The simulators probably care a lot about simulation-related philosophy! If they’re going around simulating universes, they have probably thought a *lot* about whether they themselves are a simulation, and simulation-related philosophy is probably a big part of their culture. They can’t afford to freak out every time one of their simulations starts grappling with simulation-related philosophy. It would be like freaking out when a simulation developed science, or religion, or any other natural part of cultural progress.

Some other sources raise concern that we might get our simulation terminated by becoming too computationally intensive (maybe by running simulations of our own). I think this is a more serious concern. But by the time we need to think about it, we’ll have superintelligences of our own to advise us on the risk. For now, I think we should probably stop worrying about bothering the simulators (see also the last section here). If they want us alive for some reason, we probably can’t cause them enough trouble to change that.

Book Review: Reframing Superintelligence

Ten years ago, everyone was talking about superintelligence, the singularity, the robot apocalypse. What happened?

I think the main answer is: the field matured. Why isn’t everyone talking about nuclear security, biodefense, or counterterrorism? Because there are already competent institutions working on those problems, and people who are worried about them don’t feel the need to take their case directly to the public. The past ten years have seen AI goal alignment reach that level of maturity too. There are all sorts of new research labs, think tanks, and companies working on it – the Center For Human-Compatible AI at UC Berkeley, OpenAI, Ought, the Center For The Governance Of AI at Oxford, the Leverhulme Center For The Future Of Intelligence at Cambridge, etc. Like every field, it could still use more funding and talent. But it’s at a point where academic respectability trades off against public awareness at a rate where webzine articles saying CARE ABOUT THIS OR YOU WILL DEFINITELY DIE are less helpful.

One unhappy consequence of this happy state of affairs is that it’s harder to keep up with the field. In 2014, Nick Bostrom wrote *Superintelligence: Paths, Dangers, Strategies*, giving a readable overview of what everyone was thinking up to that point. Since then, things have been less public-facing, less readable, and more likely to be published in dense papers with a lot of mathematical notation. They’ve also been – no offense to everyone working on this – less revolutionary and less interesting.

This is one reason I was glad to come across Reframing Superintelligence: Comprehensive AI Services As General Intelligence by Eric Drexler, a researcher who works alongside Bostrom at Oxford's Future of Humanity Institute. This 200 page report is not quite as readable as *Superintelligence*; its highly-structured outline form belies the fact that all of its claims start sounding the same after a while. But it's five years more recent, and presents a very different vision of how future AI might look.

Drexler asks: what if future AI looks a lot like current AI, but better?

For example, take Google Translate. A future superintelligent Google Translate would be able to translate texts faster and better than any human translator, capturing subtleties of language beyond what even a native speaker could pick up. It might be able to understand hundreds of languages, handle complicated multilingual puns with ease, do all sorts of amazing things. But in the end, it would just be a translation app. It wouldn't want to take over the world. It wouldn't even "want" to become better at translating than it was already. It would just translate stuff really well.

The future could contain a vast ecosystem of these superintelligent services before any superintelligent agents arrive. It could have media services that can write books or generate movies to fit your personal tastes. It could have invention services that can design faster cars, safer rockets, and environmentally friendly power plants. It could have strategy services that can run presidential campaigns, steer Fortune 500 companies, and advise governments. All of them would be far more effective than any human at performing their given task. But you couldn't ask the presidential-campaign-running service to design a rocket any more than you could ask Photoshop to run a spreadsheet.

In this future, our AI technology would have taken the same path as our physical technology. The human body can run fast, lift weights, and fight off enemies. But the automobile, crane, and gun are three different machines. Evolution had to cram running-ability, lifting-ability, and fighting-ability into the same body, but humans had more options and were able to do better by separating them out. In the same way, evolution had to cram book-writing, technology-inventing, and strategic-planning into the same kind of intelligence – an intelligence that also has associated goals and drives. But humans don't have to do that, and we probably won't. We're not doing it today in 2019, when Google Translate and AlphaGo are two different AIs; there's no reason to write a single AI that both translates languages and plays Go. And we probably won't do it in the superintelligent future either. Any assumption that we will is based more on anthropomorphism than on a true understanding of intelligence.

These superintelligent services would be safer than general-purpose superintelligent agents. General-purpose superintelligent agents (from here on: agents) would need a human-like structure of goals and desires to operate independently in the world; Bostrom has explained ways this is likely to go wrong. AI services would just sit around algorithmically mapping inputs to outputs in a specific domain.

Superintelligent services would not self-improve. You could build an AI researching service – or, more likely, several different services to help with several different aspects of AI research – but each of them would just be good at solving certain AI research problems. It would still take human researchers to apply their insights and actually build something new. In theory you might be able to automate every single part of AI research, but it would be a weird idiosyncratic project that wouldn’t be anybody’s first choice.

Most important, superintelligent services could help keep the world safe from less benevolent AIs. Drexler agrees that a self-improving general purpose AI agent is possible, and assumes someone will build one eventually, if only for the lulz. He agrees this could go about the way Bostrom expects it to go, ie very badly. But he hopes that there will be a robust ecosystem of AI services active by then, giving humans superintelligent help in containing rogue AIs. Superintelligent anomaly detectors might be able to notice rogue agents causing trouble, superintelligent strategic planners might be able to develop plans for getting rid of them, and superintelligent military research AIs might be able to create weapons capable of fighting them off.

Drexler therefore does not completely dismiss Bostromian disaster scenarios, but thinks we should concentrate on the relatively mild failure modes of superintelligent AI services. These may involve normal bugs, where the AI has aberrant behaviors that don’t get caught in testing and cause a plane crash or something, but not the unsolveable catastrophes of the Bostromian paradigm. Drexler is more concerned about potential misuse by human actors – either illegal use by criminals and enemy militaries, or antisocial use to create things like an infinitely-addictive super-Facebook. He doesn’t devote a lot of space to these, and it looks like he hopes these can be dealt with through the usual processes, or by prosocial actors with superintelligent services on their side (thirty years from now, maybe people will say “it takes a good guy with an AI to stop a bad guy with an AI”).

This segues nicely into some similar concerns that OpenAI researcher Paul Christiano has brought up. He worries that AI services will be naturally better at satisfying objective criteria than at “making the world better” in some vague sense. Tasks like “maximize clicks to this site” or “maximize profits from this corporation” are objective criteria; tasks like “provide real value to users of this site instead of just clickbait” or “have this corporation act in a socially responsible way” are vague. That means AI may asymmetrically empower some of the worst tedencies in our society without giving a corresponding power increase to normal people just trying to live enjoyable lives. In his model, one of the tasks of AI safety research is to get AIs to be as good at optimizing vague prosocial tasks as they will naturally be at optimizing the bottom line. Drexler doesn’t specifically discuss this in *Reframing Superintelligence*, but it seems to fit the spirit of the kind of thing he’s concerned about.

II.

I’m not sure how much of the AI alignment community is thinking in a Drexlerian vs. a Bostromian way, or whether that is even a real dichotomy that a knowledgeable person would talk about. I know there are still some people who are very concerned that even programs that seem to be innocent superintelligent services will be able to self-improve, develop misaligned goals, and cause catastrophes. I got to talk to Dr. Drexler a few years ago about some of this (although I hadn’t read the book at the time, didn’t understand the ideas very well, and probably made a fool of myself); at the time, he said that his work was getting a mixed reception. And there are still a few issues that confuse me.

First, many tasks require general intelligence. For example, an AI operating in a domain with few past examples (eg planning defense against a nuclear attack) will not be able to use modern training paradigms. When humans work on these domains, they use something like common sense, which is presumably the sort of thing we have because we understand thousands of different domains from gardening to ballistics and this gives us a basic sense of how the world works in general. Drexler agrees that we will want AIs with domain-general knowledge that cannot be instilled by training, but he argues that this is still “a service”. He agrees these tasks may require AI architectures different from any that currently exist, with relatively complete world-models, multi-domain reasoning abilities, and the ability to learn “on the fly” – but he doesn’t believe those architectures will need to be agents. Is he right?

Second, is it easier to train services or agents? Suppose you want a good multi-domain reasoner that can help you navigate a complex world. One proposal is to create AIs that train themselves to excel in world simulations the same way AlphaGo trained itself to excel in simulated games of Go against itself. This sounds a little like the evolutionary process that created humans, and agent-like drives might be a natural thing to come out of this process. If agents were easier to “evolve” than services, agentic AI might arise at an earlier stage, either because designers don’t see a problem with it or because they don’t realize it is agentic in the relevant sese.

Third, how difficult is it to separate agency from cognition? Natural intelligences use “active sampling” strategies at levels as basic as sensory perception, deciding how to direct attention in order to best achieve their goals. At higher levels, they decide things like which books to read, whose advice to seek out, or what subdomain of the problem to evaluate first. So far AIs have managed to address even very difficult problems without doing this in an agentic way. Can this continue forever? Or will there be some point at which intelligences with this ability outperform those without it.

I think Drexler’s basic insight is that Bostromian agents need to be really different from our current paradigm to do any of the things Bostrom predicts. A paperclip maximizer built on current technology would have to eat gigabytes of training data about various ways people have tried to get paperclips in the past so it can build a model that lets it predict what works. It would build the model on its actually-existing hardware (not an agent that could adapt to much better hardware or change its hardware whenever convenient). The model would have a superintelligent understanding of the principles that had guided some things to succeed or fail in the training data, but wouldn’t be able to go far beyond them into completely new out-of-the-box strategies. It would then output some of those plans to a human, who would look them over and make paperclips 10% more effectively.

The very fact that this is less effective than the Bostromian agent suggests there will be pressure to build the Bostromian agent eventually (Drexler disagrees with this, but I don’t understand why). But this will be a very different project from AI the way it currently exists, and if AI the way it currently exists can be extended all the way to superintelligence, that would give us a way to deal with hostile superintelligences in the future.

III.

All of this seems kind of common sense to me now. This is worrying, because I didn’t think of any of it when I read

Superintelligence in 2014.

I asked readers to tell me if there was any past discussion of this. Many people brought up Robin Hanson's arguments, which match the "ecosystem of many AIs" part of Drexler's criticisms but don't focus as much on services vs. agents. Other people brought up discussion under the heading of Tool AI. Combine those two strains of thought, and you more or less have Drexler's thesis, minus some polish. I read some of these discussions, but I think I failed to really understand them at the time. Maybe I failed to combine them, focused too much on the idea of an Oracle AI, and missed the idea of an ecosystem of services. Or maybe it all just seemed too abstract and arbitrary when I had fewer examples of real AI systems to think about.

I've sent this post by a couple of other people, who push back against it. They say they still think Bostrom was right on the merits and superintelligent agents are more likely than superintelligent services. Many brought up Gwern's essay on why tool AIs are likely to turn into agent AIs and this post by Eliezer Yudkowsky on the same topic – I should probably reread these, reread Drexler's counterarguments, and get a better understanding. For now I don't think I have much of a conclusion either way. But I think I made a mistake of creativity in not generating or understanding Drexler's position earlier, which makes me more concerned about how many other things I might be missing.

Meetups Everywhere 2019

Last autumn we organized meetups in 85 different cities (and one ship!) around the world. Some of the meetup groups stuck around or reported permanent spikes in membership, which sounds like a success, so let’s do it again.

For most cities: If you’re willing to host a meetup for your city, then decide on a place, date, and time, and post it in the comments here, along with an email address where people can contact you. Then please watch the comments in case I need to ask you any questions. If you’re not sure whether your city has enough SSC readers to support a meetup, see the list of people by city at the bottom of this post. There may be more of us than you think – last year we were able to support meetups in such great megalopolises as Norman, Oklahoma and Wellington, New Zealand. But I would prefer people not split things up too much – if you’re very close to a bigger city, consider going there instead of hosting your own.

If you want a meetup for your city, please err in favor of volunteering to organize – the difficulty level is basically “pick a coffee shop you like, tell me the address, and give me a time”; it would be dumb if nobody got to go to meetups because everyone felt too awkward and low-status to volunteer.

For especially promising cities in the US: I am going to try to attend your meetups. My very tentative schedule looks like this:

Friday 9/20: Boston
Saturday 9/21: NYC

Sunday 9/22: Philly
Monday 9/23: DC
Thursday 9/26: Ann Arbor
Saturday 9/28: Chicago
Sunday 9/29: Austin
Tuesday 10/1: Portland
Wednesday 10/2: Seattle
Friday 10/4: Fairbanks
Thursday 10/10: Berkeley
Friday 10/11: Orange County

If you are in one of these cities and want to host a meetup, please schedule it for the evening of the relevant day. If that's impossible, let me know and I might be able to reschedule. I will announce these ones on the blog, and in the past that's meant they can get very big (100+ people in the biggest cities) – you might want to hold it in a house, park, or classroom (not a cafe or small apartment). If you have a great location but need money, email me and I might be able to help.

Small-print rules for organizers

1. In a week or so, I'll make another post listing the details for each city so people know where to go.I don't guarantee I'll have the post with times and addresses up until September 9, so please choose a day after that. The weekend of September 21st and 22nd might be one good choice.

2. In the past, the best venues have been ones that are quiet(ish) and have lots of mobility for people to arrange themselves into circles or subgroups as desired. Private houses have been pretty good. Same with food courts. Cafes and restaurants have gone okay, as have empty fields (really). Bars don't seem to have worked very well at all.

3. Usually only about a quarter of people who express interest actually attend. If your city has fewer than 20 people on the big list, don't offer to organize unless you're okay with a good chance of only one or two other people showing up.

4. If more than one person volunteers to organize, I'll pick one of them. Priority will be given to people I know well, people who have organized meetups before, and (especially) an existing SSC/LW/EA meetup group in the city. If you run an existing SSC/LW/EA meetup group and you want to organize your city's SSC meetup, please mention that in the post so I can give you precedence.

5. If you have an existing meetup group, you can just tell me what you're already doing and when your next meetup is. But try to have the one you list here be some kind of "welcome, SSC people" meetup or otherwise low-barrier-to-entry. And please give me a firm date and time commitment instead of "tell people to check our mailing list to find out where the meeting will be that week".

6. If you're formally volunteering to organize a meetup, please respond with an unambiguous statement to this effect, the exact address, the exact time, and the date (+ contact details if possible), preferably **in bold**. I'm not going to count someone as offering to organize a meetup unless they do this. Please don't post "I hope someone agrees to organize a meetup in my city". Just offer to organize the meetup! Again, *please include an exact time, exact date, and exact address with your offer to host*. Please don't post vague speculation about how you might want to host at some point – just offer to host and give me the information I need. If it turns out there's someone better, don't worry, they'll also offer and I'll choose them.

7. Mingyuan is Director Of Meetups and might be asking you some questions; I vouch for her and you should give her any information she needs.

Thanks, and see (some of) you soon!

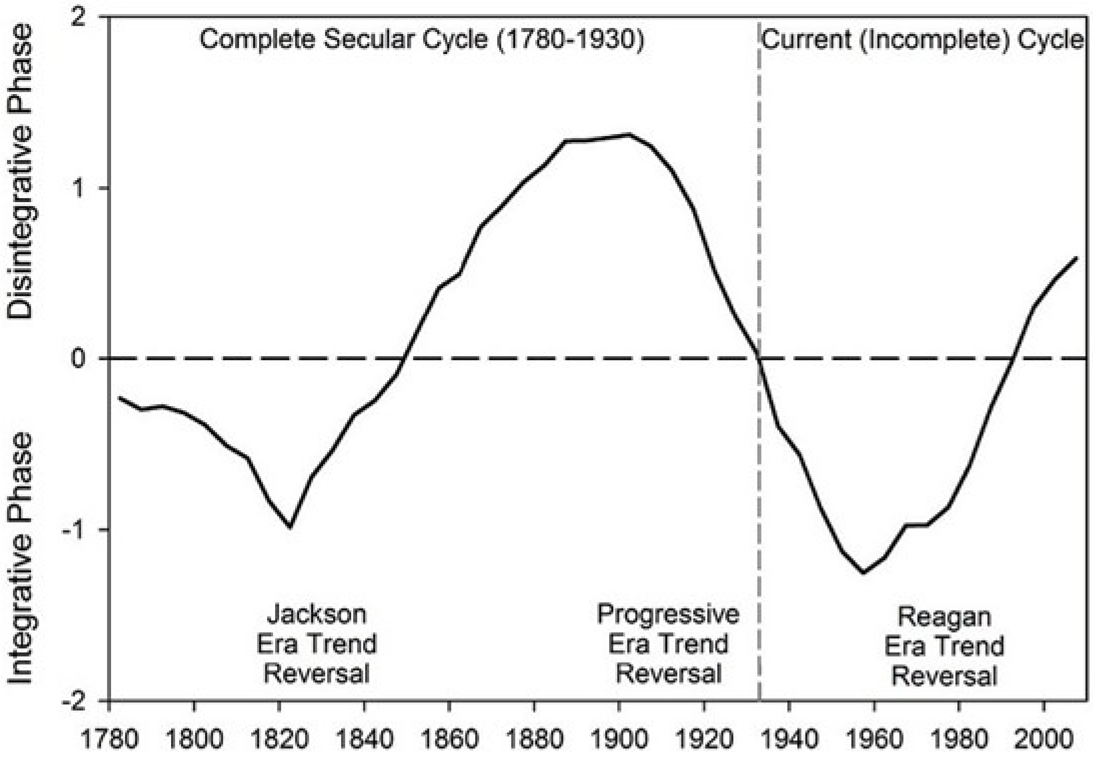
Book Review: Ages Of Discord

I.

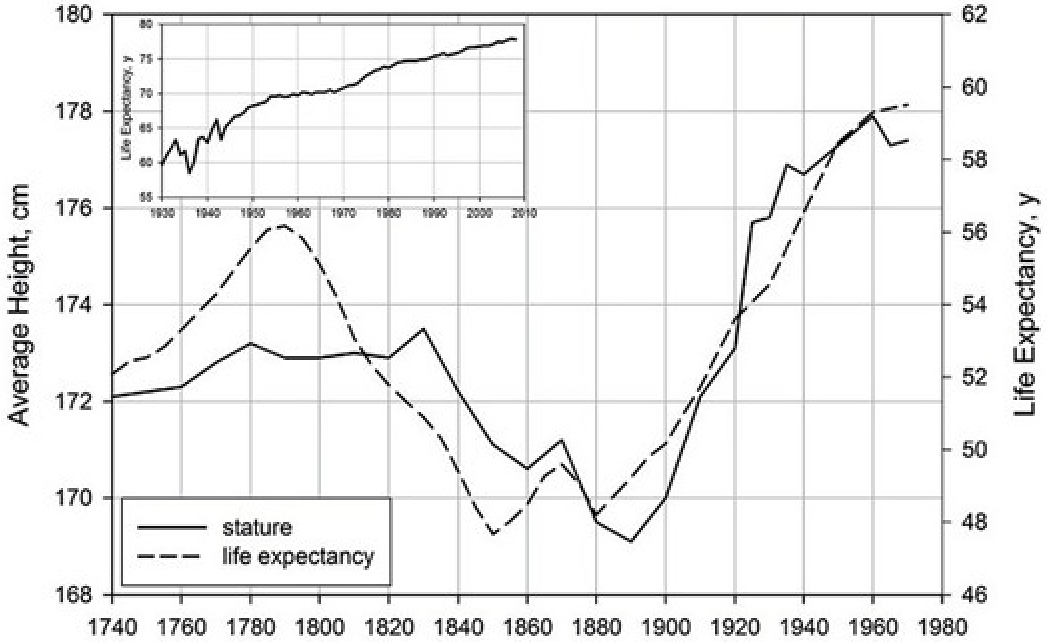
I recently reviewed Secular Cycles, which presents a demographic-structural theory of the growth and decline of pre-industrial civilizations. When land is plentiful, population grows and the economy prospers. When land reaches its carrying capacity and income declines to subsistence, the area is at risk of famines, diseases, and wars – which kill enough people that land becomes plentiful again. During good times, elites prosper and act in unity; during bad times, elites turn on each other in an age of backstabbing and civil strife. It seemed pretty reasonable, and authors Peter Turchin and Sergey Nefedov had lots of data to support it.

Ages of Discord is Turchin's attempt to apply the same theory to modern America. There are many reasons to think this shouldn't work, and the book does a bad job addressing them. So I want to start by presenting Turchin's data showing such cycles exist, so we can at least see why the hypothesis might be tempting. Once we've seen the data, we can decide how turned off we want to be by the theoretical problems.

The first of Turchin's two cyclic patterns is a long cycle of national growth and decline. In *Secular Cycles*' pre-industrial societies, this pattern lasted about 300 years; in *Ages of Discord*'s picture of the modern US, it lasts about 150:

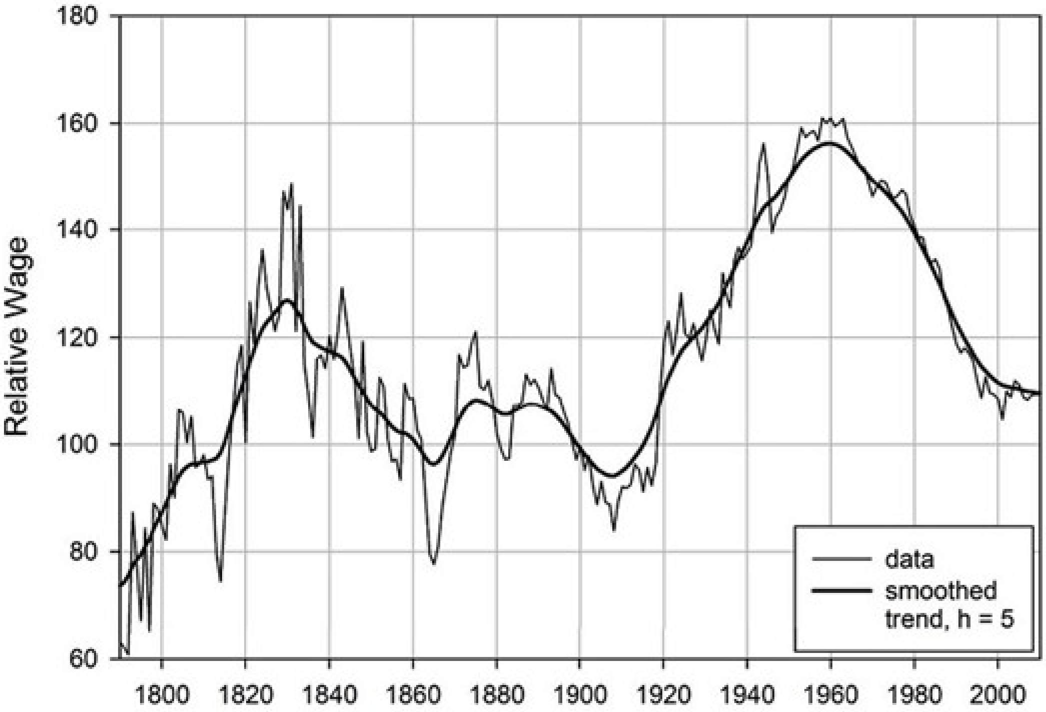


This summary figure combines many more specific datasets. For example, archaeologists frequently assess the prosperity of a period by the heights of its skeletons. Well-nourished, happy children tend to grow taller; a layer with tall skeletons probably represents good times during the relevant archaeological period; one with stunted skeletons probably represents famine and stress. What if we applied this to the modern US?



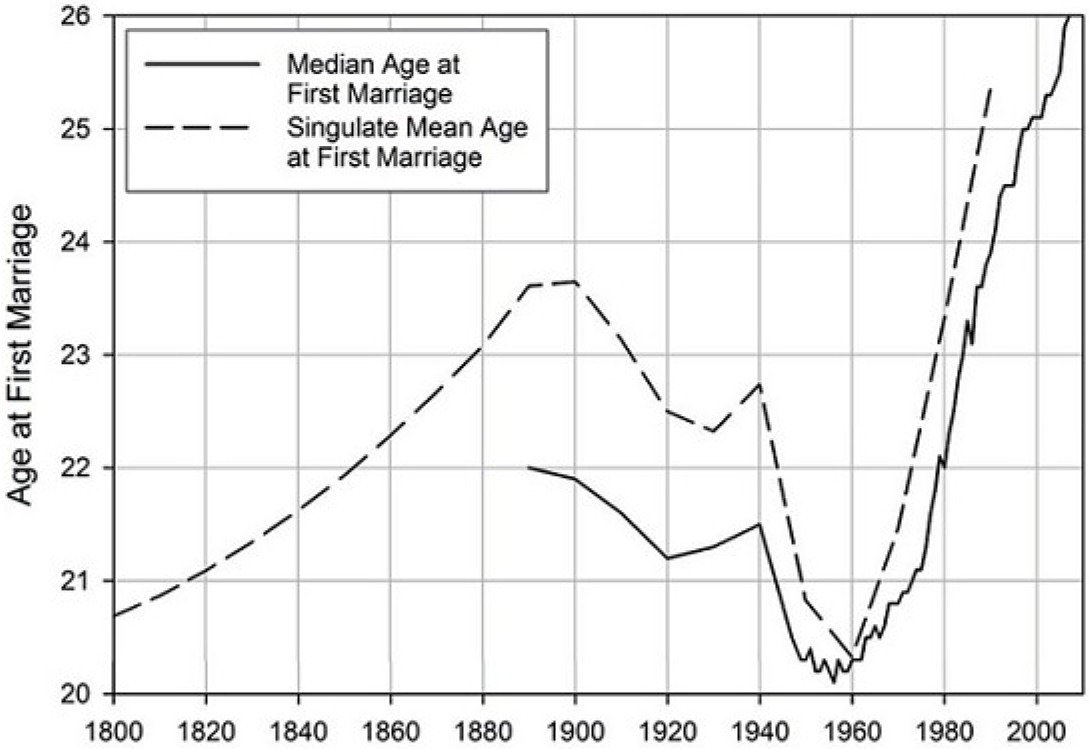
Average US height and life expectancy over time. As far as I can tell, the height graph is raw data. The life expectancy graph is the raw data minus an assumed constant positive trend – that is, given that technological advance is increasing life expectancy at a linear rate, what are the other factors you see when you subtract that out? The exact statistical logic be buried in Turchin's source (Historical Statistics of the United States, Carter et al 2004), which I don't have and can't judge.

This next graph is the median wage divided by GDP per capita, a crude measure of income equality:

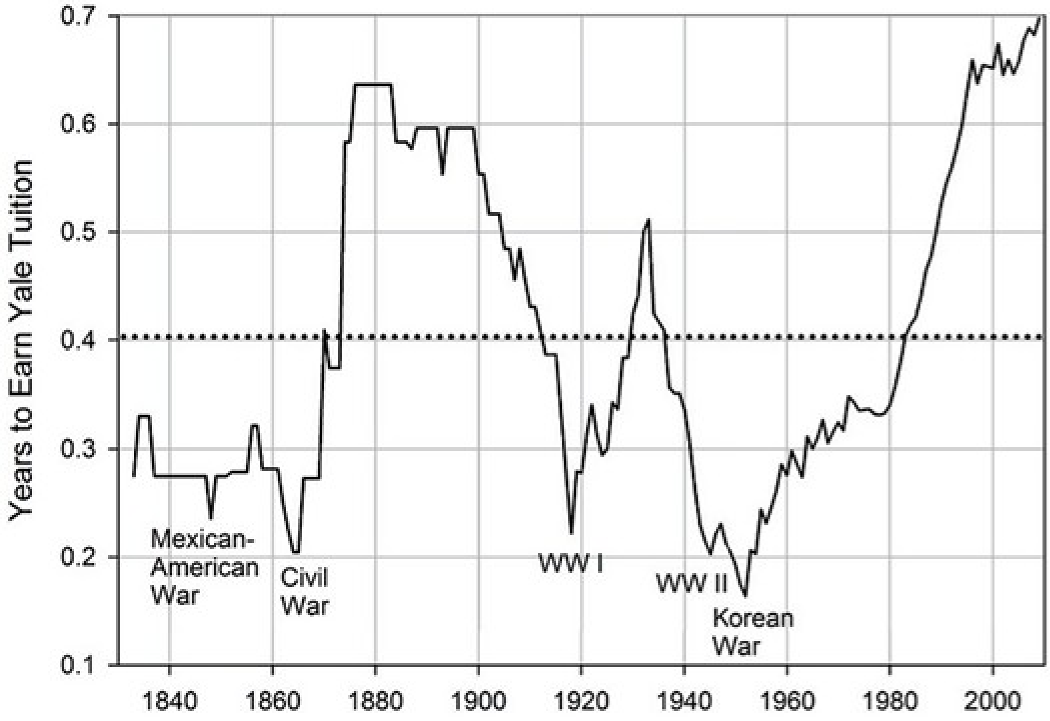


Lower values represent more inequality.

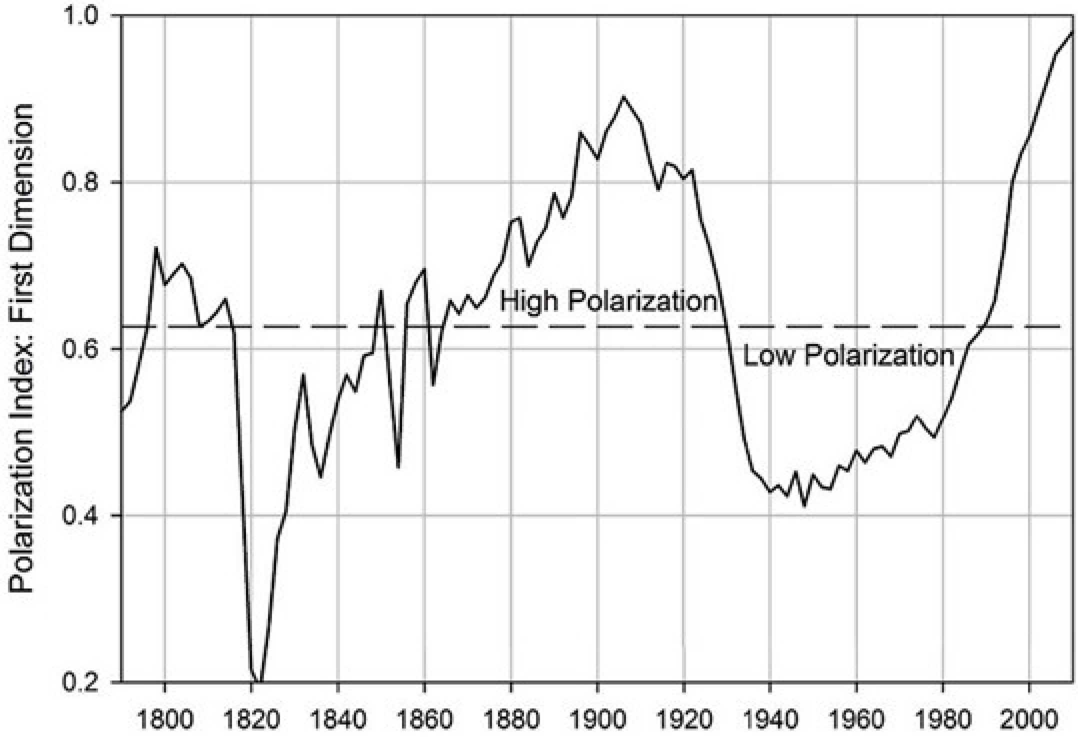
This next graph is median female age at first marriage. Turchin draws on research suggesting this tracks social optimism. In good times, young people can easily become independent and start supporting a family; in bad times, they will want to wait to make sure their lives are stable before settling down:



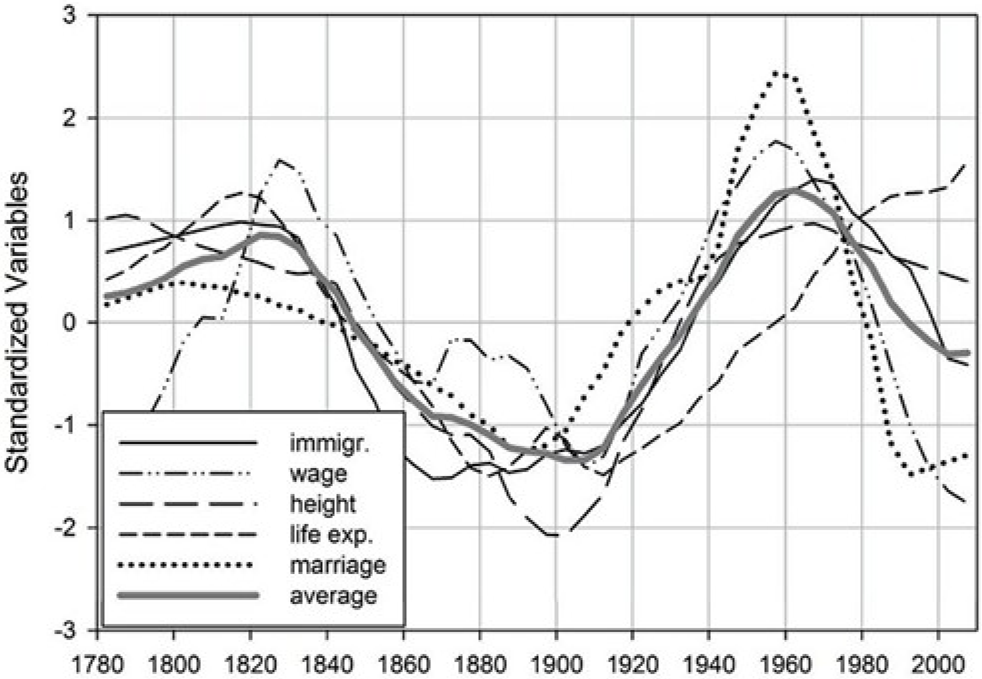
This next graph is Yale tuition as a multiple of average manufacturing worker income. To some degree this will track inequality in general, but Turchin thinks it also measures something like “difficulty of upward mobility”:



This next graph shows DW-NOMINATE's "Political Polarization Index", a complicated metric occasionally used by historians of politics. It measures the difference in voting patterns between the average Democrat in Congress and the average Republican in Congress (or for periods before the Democrats and Republicans, whichever two major parties there were). During times of low partisanship, congressional votes will be dominated by local or individual factors; during times of high partisanship, it will be dominated by party identification:



I’ve included only those graphs which cover the entire 1780 – present period; the book includes many others that only cover shorter intervals (mostly the more recent periods when we have better data). All of them, including the shorter ones not included here, reflect the same general pattern. You can see it most easily if you standardize all the indicators to the same scale, match the signs so that up always means good and down always means bad, and put them all together:

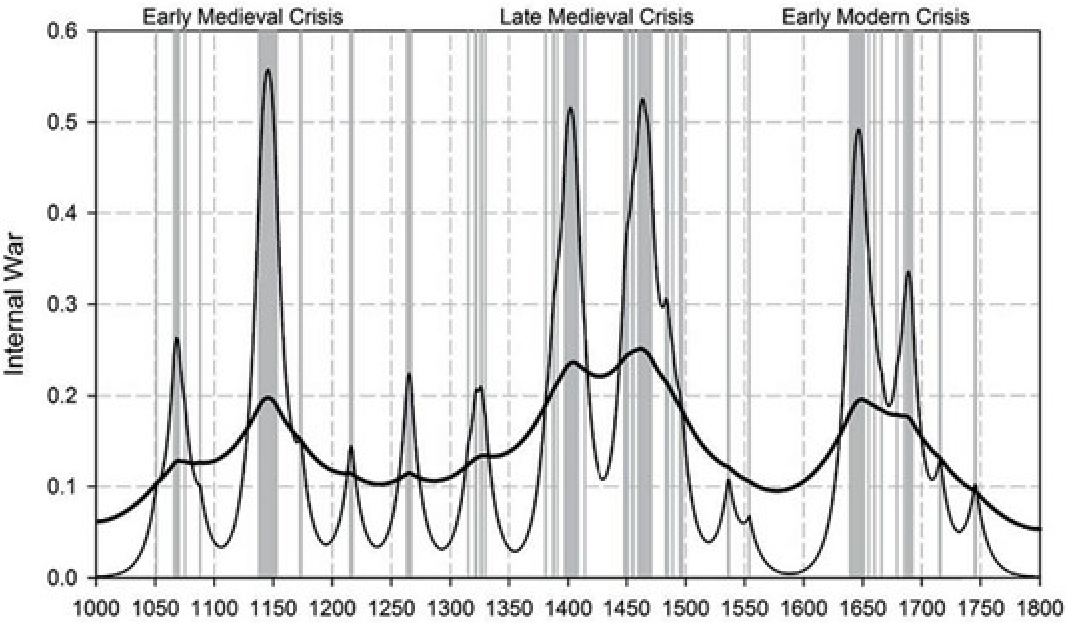


Note that these aren’t exactly the same indicators I featured above; we’ll discuss immigration later.

The “average” line on this graph is the one that went into making the summary graphic above. Turchin believes that after the American Revolution, there was a period of instability lasting a few decades (eg Shays’ Rebellion, Whiskey Rebellion) but that America reached a maximum of unity, prosperity, and equality around 1820. Things gradually got worse from there, culminating in a peak of inequality, misery, and division around 1900. The reforms of the Progressive Era gradually made things better, with another unity/prosperity/equality maximum around 1960. Since then, an increasing confluence of negative factors (named here as the Reagan Era trend reversal, but Turchin admits it began before Reagan) has been making things worse again.

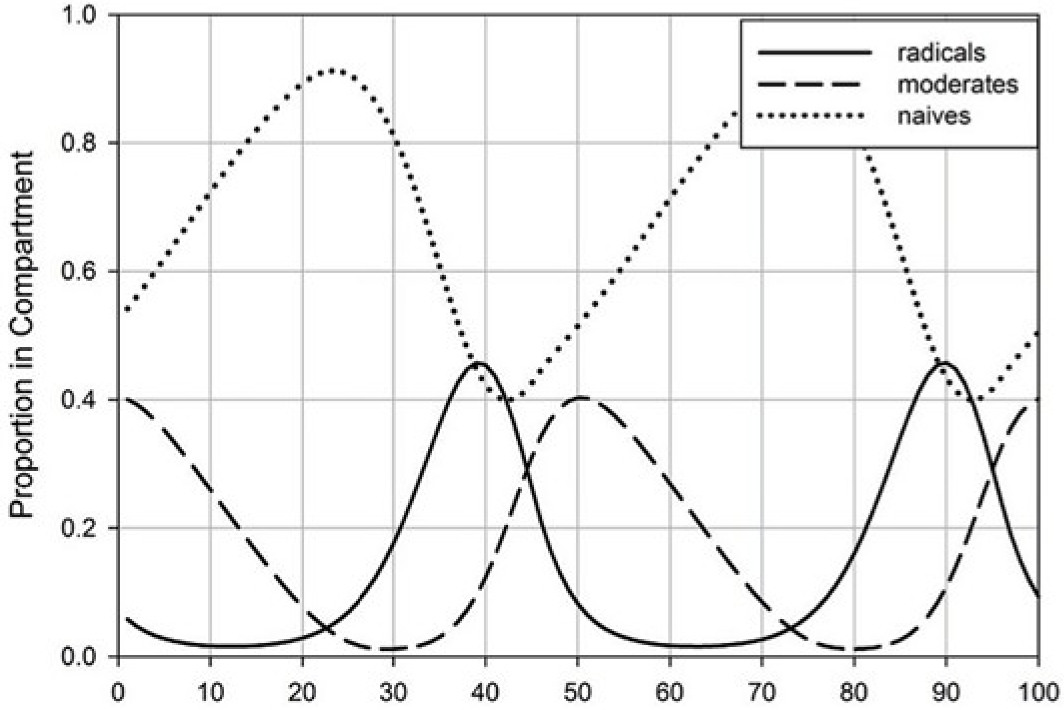
II.

Along with this “grand cycle” of 150 years, Turchin adds a shorter instability cycle of 40-60 years. This is the same 40-60 year instability cycle that appeared in *Secular Cycles*, where Turchin called it “the bigenerational cycle”, or the “fathers and sons cycle”.



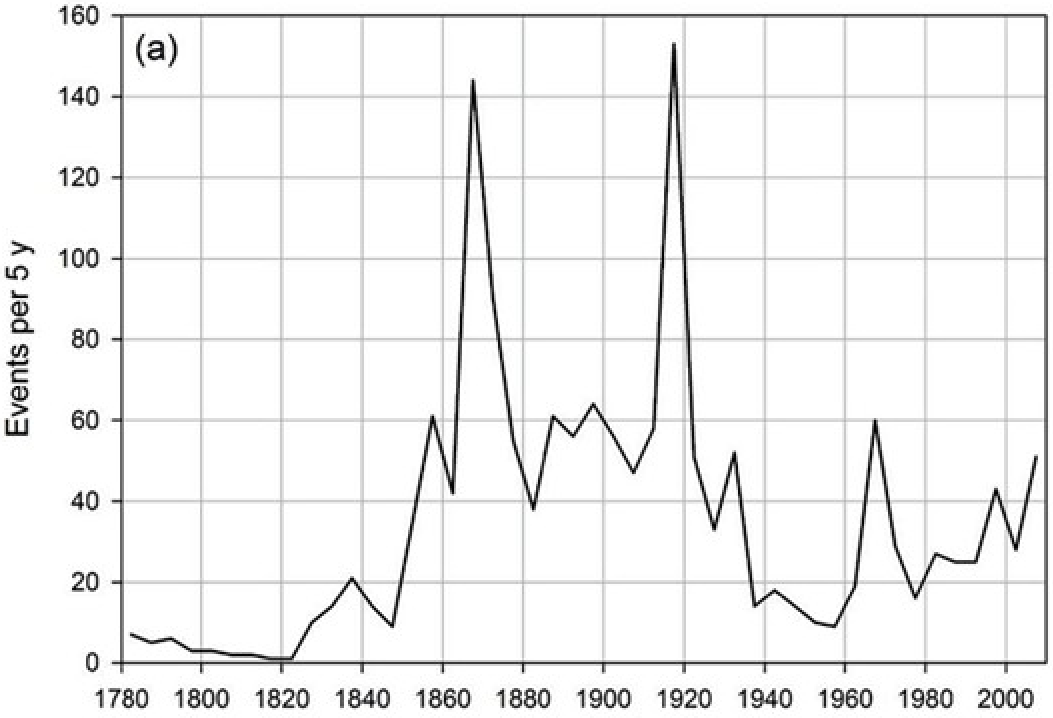
Timing and intensity of internal war in medieval and early modern England, from Turchin and Nefedov 2009.

The derivation of this cycle, explained on pages 45 – 58 of *Ages of Discord*, is one of the highlights of the book. Turchin draws on the kind of models epidemiologists use to track pandemics, thinking of violence as an infection and radicals as plague-bearers. You start with an unexposed vulnerable population. Some radical – patient zero – starts calling for violence. His ideas spread to a certain percent of people he interacts with, gradually “infecting” more and more people with the “radical ideas” virus. But after enough time radicalized, some people “recover” – they become exhausted with or disillusioned by conflict, and become pro-cooperation “active moderates” who are impossible to reinfect (in the epidemic model, they are “inoculated”, but they also have an ability without a clear epidemiological equivalent to dampen radicalism in people around them). As the rates of radicals, active moderates, and unexposed dynamically vary, you get a cyclic pattern. First everyone is unexposed. Then radicalism gradually spreads. Then active moderation gradually spreads, until it reaches a tipping point where it triumphs and radicalism is suppressed to a few isolated reservoirs in the population. Then the active moderates gradually die off, new unexposed people are gradually born, and the cycle starts again. Fiddling with all these various parameters, Turchin is able to get the system to produce 40-60 year waves of instability.



To check this empirically, Turchin tries to measure the number of “instability events” in the US over various periods. He very correctly tries to use lists made by others (since they are harder to bias), but when people haven’t catalogued exactly the kind of instability he’s interested in over the entire 1780 – present period, he sometimes adds his own interpretation. He ends up summing riots, lynchings, terrorism (including assassinations), and mass shootings – you can see his definition for each of these starting on page 114; the short version is that all the definitions seem reasonable but inevitably include a lot of degrees of freedom.

When he adds all this together, here's what happens:



Political instability / violent events show three peaks, around 1870, 1920, and 1970.

The 1870 peak includes the Civil War, various Civil War associated violence (eg draft riots), and the violence around Reconstruction (including the rise of the Ku Klux Klan and related violence to try to control newly emancipated blacks).

The 1920 peak includes the height of the early US labor movement. Turchin discusses the Mine War, an "undeclared war” from 1920-1921 between bosses and laborers in Appalachian coal country:

> Although it started as a labor dispute, it eventually turned into the largest armed insurrection in US history, other than the Civil War. Between 10,000 and 15,000 miners armed with rifles fought thouasnds of strike-breakers and sheriff’s deputies, called the Logan Defenders. The insurrection was ended by the US Army. While such violent incidents were exceptional, they took place against a background of a general “class war” that had been intensifying since the violent teens. “In 1919 nearly four million workers (21% of the workforce) took disruptive action in the face of employer reluctance to recognize or bargain with unions” (Domhoff and Webber, 2011:74).

Along with labor violence, 1920 was also a peak in racial violence:

> Race-motivated riots also peaked around 1920. The two most serious such outbreaks were the Red Summer of 1919 (McWhirter 2011) and the Tulsa (Oklahoma) Race Riot. The Red Summer involved riots in more than 20 cities across the United States and resulted in something like 1,000 fatalities. The Tulsa riot in 1921, which caused about 300 deaths, took on an aspect of civil war, in which thousands of whites and blacks, armed with firearms, fought in the streets, and most of the Greenwood District, a prosperous black neighborhood, was destroyed.

And terrorism:

> The bombing campaign by Italian anarchists (“Galleanists”) culminated in the 1920 explosion on Wall

> Street, which caused 38 fatalities.

The same problems: labor unrest, racial violence, terrorism – repeated during the 1970s spike. Instead of quoting Turchin on this, I want to quote this Status 451 review of *Days of Rage*, because it blew my mind:

> *“People have completely forgotten that in 1972 we had over nineteen hundred domestic bombings in the United States.” — Max Noel, FBI (ret.)*
>
> Recently, I had my head torn off by a book: Bryan Burrough’s Days of Rage, about the 1970s underground. It’s the most important book I’ve read in a year. So I did a series of running tweetstorms about it, and Clark asked me if he could collect them for posterity. I’ve edited them slightly for editorial coherence.
>
> Days of Rage is important, because this stuff is forgotten and it shouldn’t be. The 1970s underground wasn’t small. It was hundreds of people becoming urban guerrillas. Bombing buildings: the Pentagon, the Capitol, courthouses, restaurants, corporations. Robbing banks. Assassinating police. People really thought that revolution was imminent, and thought violence would bring it about.
>
> One thing that Burrough returns to in Days of Rage, over and over and over, is how forgotten so much of this stuff is. Puerto Rican separatists bombed NYC like 300 times, killed people, shot up Congress, tried to kill POTUS (Truman). Nobody remembers it.

The passage speaks to me because – yeah, nobody remembers it. This is also how I feel about the 1920 spike in violence. I’d heard about the Tulsa race riot, but the Mine War and the bombing of Wall Street and all the other stuff was new to me. This matters because my intuitions before reading this book would not have been that there were three giant spikes in violence/instability in US history located fifty years apart. I think the lesson I learn is not to trust my intuitions, and to be a little more sympathetic to Turchin's data.

One more thing: the 1770 spike was obviously the American Revolution and all of the riots and communal violence associated with it (eg against Tories). Where was the 1820 spike? Turchin admits it didn't happen. He says that because 1820 was the absolute best part of the 150 year grand cycle, everybody was so happy and well-off and patriotic that the scheduled instability peak just fizzled out. Although Turchin doesn't mention it, you could make a similar argument that the 1870 spike was especially bad (see: the entire frickin' Civil War) because it hit close to (though not exactly at) the worst part of the grand cycle. 1920 hit around the middle, and 1970 during a somewhat-good period, so they fell in between the nonissue of 1820 and the disaster of 1870.

III.

I haven't forgotten the original question – what drives these 150 year cycles of rise and decline – but I want to stay with the data just a little longer. Again, these data are really interesting. Either some sort of really interesting theory has to be behind them – or they're just low-quality data cherry-picked to make a point. Which are they? Here are a couple of spot-checks to see if the data are any good.

First spot check: can I confirm Turchin's data from independent sources?

– Here is a graph of average US height over time which seems broadly similar to Turchin’s.

– Here is a different measure of US income inequality over time, which again seems broadly similar to Turchin’s. Piketty also presents very similar data, though his story places more emphasis on the World Wars and less on the labor movement.

– The Columbia Law Review measures political polarization over time and gets mostly the same numbers as Turchin.

I’m going to consider this successfully checked; Turchin’s data all seem basically accurate.

Second spot check: do other indicators Turchin didn’t include confirm the pattern he detects, or did he just cherry-pick the data series that worked? Spoiler: I wasn’t able to do this one. It was too hard to think of measures that should reflect general well-being and that we have 200+ years of unconfounded data for. But here are my various failures:

– The annual improvement in mortality rate does not seem to follow the cyclic pattern. But isn’t this more driven by a few random factors like smoking rates and the logic of technological advance?

– Treasury bonds maybe kind of follow the pattern until 1980, after which they go crazy.

– Divorce rates look kind of iffy, but isn’t that just a bunch of random factors?

– Homicide rates, with the general downward trend removed, sort of follow the pattern, except for the recent decline?

– USD/GBP exchange rates don't show the pattern at all, but that could be because of things going on in Britain?

The thing is – really I have no reason to expect divorce rates, homicide rates, exchange rates etc to track national flourishing. For one thing, they may just be totally unrelated. For another, even if they were tenuously related, there are all sorts of other random factors that can affect them. The problem is, I would have said this was true for height, age at first marriage, and income inequality too, before Turchin gave me convincing-sounding stories for why it wasn't. I think my lesson is that I have no idea which indicators should vs. shouldn't follow a secular-cyclic pattern and so I can't do this spot check against cherry-picking the way I hoped.

Third spot check: common sense. Here are some things that stood out to me:

– The Civil War is at a low-ish part of the cycle, but by no means the lowest.

– The Great Depression happened at a medium part of the cycle, when things should have been quickly getting better.

– Even though there was a lot of new optimism with Reagan, continuing through the Clinton years, the cycle does not reflect this at all.

Maybe we can rescue the first and third problem by combining the 150 year cycle with the shorter 50 year cycle. The Civil War was determined by the 50-year cycle having its occasional burst of violence at the same time the 150-year cycle was at a low-ish point. People have good memories of Reagan because the chaos of the 1970 violence burst had ended.

As for the second, Turchin is aware of the problem. He writes:

> There is a widely held belief among economists and other social scientists that the 1930s were the “defining moment” in the development of the American politico-economic system (Bordo et al 1998). When we look at the major structural-demographic variables, however, the decade of the 1930s does not seem to be a turning point. Structural-demographic trends that were established during the Progressive Era continues through the 1930s, although some of them accelerated.
>
> Most notably, all the well-being variables that went through trend reversals before the Great Depression – between 1900 and 1920. From roughly 1910 and to 1960 they all increased roughly monotonically, with only one or two minor fluctuations around the upward trend. The dynamics of real wages also do not exhibit a breaking point in the 1930s, although there was a minor acceleration after 1932.

By comparison, he plays up the conveniently-timed (and hitherto unknown to me) depression of the mid-1890s. Quoting Turchin quoting McCormick:

> No depression had ever been as deep and tragic as the one that lasted from 1893 to 1897. Millions suffered unemployment, especially during the winters of 1893-4 and 1894-5, and thousands of ‘tramps’ wandered the countryside in search of food […]
>
> Despite real hardship resulting form massive unemployment, well-being indicators suggest that the human cost of the Great Depression of the 1930s did not match that of the “First Great Depression” of the

> 1890s (see also Grant 1983:3-11 for a general discussion of the severity of the 1890s depression. Furthermore, while the 1930s are remembered as a period of violent labor unrest, the intensity of class struggle was actually lower than during the 1890s depression. According to the US Political Violence Database (Turchin et al. 2012) there were 32 lethal labor disputes during the 1890s that collectively caused 140 deaths, compared with 20 such disputes in the 1930s with the total of 55 deaths. Furthermore, the last lethal strike in US labor history was in 1937…in other words, the 1930s was actually the last uptick of violent class struggle in the US, superimposed on an overall declining trend.
>
> The 1930s Depression is probably remembered (or rather misremembered) as the worst economic slump in US history, simply because it was the last of the great depressions of the post-Civil War era.

Fourth spot check: Did I randomly notice any egregious errors while reading the book?

On page 70, Turchin discusses “the great cholera epidemic of 1849, which carried away up to 10% of the American population”. This seemed unbelievably high to me. I checked the source he cited, Kohl’s “Encyclopedia Of Plague And Pestilence”, which did give that number. But every other source I checked agreed that the epidemic “only” killed between 0.3% – 1% of the US population (it did hit 10% in a few especially unlucky cities like St. Louis). I cannot fault Turchin’s scholarship in the sense of correctly repeating something written in an encyclopedia, but unless I’m missing something I do fault his common sense.

Also, on page 234, Turchin interprets the percent of medical school graduates who get a residency as “the gap between the demand and supply of MD positions”, which he ties into a wider argument about elite overproduction. But I think this shows a limited understanding of how the medical system works. There is currently a severe *undersupply* of doctors – try getting an appointment with a specialist who takes insurance in a reasonable amount of time if you don’t believe me. Residencies aren’t limited by organic demand. They’re limited because the government places so many restrictions on them that hospitals don’t sponsor them without government funding, and the government is too stingy to fund more of them. None of this has anything to do with elite overproduction.

These are just two small errors in a long book. But they’re two errors in medicine, the field I know something about. This makes me worry about Gell-Mann Amnesia: if I notice errors in my own field, how many errors must there be in other fields that I just didn’t catch?

My overall conclusion from the spot-checks is that the data as presented are basically accurate, but that everything else is so dependent on litigating which things are vs. aren’t in accordance with the theory that I basically give up.

IV.

Okay. We’ve gone through the data supporting the grand cycle. We’ve gone through the data and theory for the 40-60 year instability cycle. We’ve gone through the reasons to trust vs. distrust the data. Time to go back to the question we started with: why should the grand cycle, originally derived from the Malthusian principles that govern pre-industrial societies, hold in the modern US? Food and land are no longer limiting resources; famines, disease, and wars no longer substantially decrease population. Almost every factor that drives the original secular cycle is missing; why even consider the possibility that it might still apply?

I’ve put this off because, even though this is the obvious question *Ages of Discord* faces from page one, I found it hard to get a single clear answer.

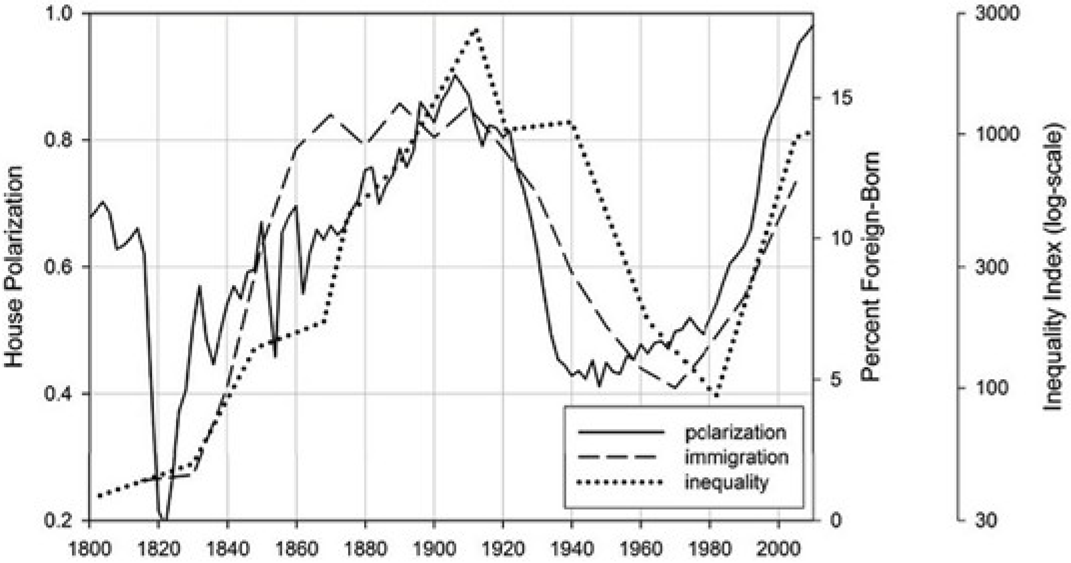
Sometimes, Turchin talks about the supply vs. demand of labor. In times when the supply of labor outpaces demand, wages go down, inequality increases, elites fragment, and the country gets worse, mimicking the “land is at carrying capacity” stage of the Malthusian cycle. In times when demand for labor exceeds supply, wages go up, inequality decreases, elites unite, and the country gets better. The government is controlled by plutocrats, who always want wages to be low. So they implement policies that increase the supply of labor, especially loose immigration laws. But their actions cause inequality to increase and everyone to become miserable. Ordinary people organize resistance: populist movements, socialist cadres, labor unions. The system teeters on the edge of violence, revolution, and total disintegration. Since the elites don’t want those things, they take a step back, realize they’re killing the goose that lays the golden egg, and decide to loosen their grip on the neck of the populace. The government becomes moderately pro-labor and progressive for a while, and tightens immigration laws. The oversupply of labor decreases, wages go up, inequality goes down, and everyone is happy. After everyone has been happy for a while, the populists/socialists/unions lose relevance and drift apart. A new generation of elites who have never felt threatened come to power, and they think to themselves “What if we used our control of the government to squeeze labor harder?” Thus the cycle begins again.

But at other times, Turchin talks more about "elite overproduction". When there are relatively few elites, they can cooperate for their common good. Bipartisanship is high, everyone is unified behind a system perceived as wise and benevolent, and we get a historical period like the 1820s US golden age that historians call The Era Of Good Feelings. But as the number of elites outstrips the number of high-status positions, competition heats up. Elites realize they can get a leg up in an increasingly difficult rat race by backstabbing against each other and the country. Government and culture enter a defect-defect era of hyperpartisanship, where everyone burns the commons of productive norms and institutions in order to get ahead. Eventually…some process reverses this or something?…and then the cycle starts again.

At still other times, Turchin seems to retreat to a sort of mathematical formalism. He constructs an extremely hokey-looking dynamic feedback model, based on ideas like "assume that the level of discontent among ordinary people equals the urbanization rate x the age structure x the inverse of their wages relative to the elite" or "let us define the fiscal distress index as debt ÷ GDP x the level of distrust in state institutions". Then he puts these all together into a model that calculates how the the level of discontent affects and is affected by the level of state fiscal distress and a few dozen other variables. On the one hand, this is really cool, and watching it in action gives you the same kind of feeling Seldon must have had inventing psychohistory. On the other, it seems *really* made-up. Turchin admits that dynamic feedback systems are infamous for going completely haywire if they are even a tiny bit skew to reality, but assures us that he understands the cutting-edge of the field and how to make them not to do that. I don't know enough to judge whether he's right or wrong, but my priors are on "extremely, almost unfathomably wrong". Still, at times he reminds us that the shifts of dynamic feedback systems can be attributed only to the system in its entirety, and that trying to tell stories about or point to specific factors involved in any particular shift is an approximation at best.

All of these three stories run into problems almost immediately.

First, the supply of labor story focuses pretty heavily on immigration. Turchin puts a lot of work into showing that immigration follows the secular cycle patterns; it is highest at the worst part of the cycle, and lowest at the best parts:



In this model, immigration is a tool of the plutocracy. High supply of labor (relative to demand) drives down wages, increases inequality, and lowers workers' bargaining power. If the labor supply is poorly organized, comes from places that don't understand the concept of "union", don't know their rights, and have racial and linguistic barriers preventing them from cooperating with the rest of the working class, well, even better. Thus, periods when the plutocracy is successfully squeezing the working class are marked by high immigration. Periods when the plutocracy fears the working class and feels compelled to be nice to them are marked by low immigration.

This position makes some sense and is loosely supported by the long-term data above. But isn’t this one of the most-studied topics in the history of economics? Hasn’t it been proven almost beyond doubt that immigrants don’t steal jobs from American workers, and that since they consume products themselves (and thus increase the demand for labor) they don’t affect the supply/demand balance that sets wages?

It appears I might just be totally miscalibrated on this topic. I checked the IGM Economic Experts Panel. Although most of the expert economists surveyed believed immigration was a net good for America, they did say (50% agree to only 9% disagree) that “unless they were compensated by others, many low-skilled American workers would be substantially worse off if a larger number of low-skilled foreign workers were legally allowed to enter the US each year”. I’m having trouble seeing the difference between this statement (which economists seem very convinced is true) and “you should worry about immigrants stealing your job” (which everyone seems very convinced is false). It might be something like – immigration generally makes “the economy better”, but there’s no guarantee that these gains are evently distributed, and so it can be bad for low-skilled workers in particular? I don’t know, this would still represent a pretty big update, but given that I was told all top economists think one thing, and now I have a survey of all top economists saying the other, I guess big updates are unavoidable. Interested in hearing from someone who knows more about this.

Even if it’s true that immigration can hurt low-skilled workers, Turchin’s position – which is that increased immigration is responsible for a very large portion of post-1973 wage stagnation and the recent trend toward rising inequality – sounds shocking to current political sensibilities. But all Turchin has to say is:

> An imbalance between labor supply and demand clearly played an important role in driving real wages down after 1978. As Harvard economist George J. Borjas recently wrote, “The best empirical research that tries to examine what has actually happened in the US labor market aligns well with economic theory: An increase in the number of workers leads to lower wages.”

My impression was that Borjas was an increasingly isolated contrarian voice, so once again, I just don’t know what to do here.

Second, the plutocratic oppression story relies pretty heavily on the idea that inequality is a unique bad. This fits the zeitgeist pretty well, but it’s a little confusing. Why should commoners care about their wages relative to elites, as opposed to their absolute wages? Although median-wage-relative-to-GDP has gone down over the past few decades, absolute median wage has gone up – just a little, slowly enough that it’s rightly considered a problem – but it has gone up. Since modern wages are well above 1950s wages, in what sense should modern people feel like they are economically bad off in a way 1950s people didn’t? This isn’t a problem for Turchin’s theory so much as a general mystery, but it’s a general mystery I care about a lot. One answer is that the cost disease is fueled by a Baumol effect pegged to per capital income (see part 3 here), and this is a way that increasing elite wealth can absolutely (not relatively) immiserate the lower classes.

Likewise, what about The Spirit Level Delusion and other resources showing that, across countries, inequality is not particularly correlated with social bads? Does this challenge Turchin's America-centric findings that everything gets worse along with inequality levels?

Third, the plutocratic oppression story meshes poorly with the elite overproduction story. In elite overproduction, united elites are a sign of good times to come; divided elites means dysfunctional government and potential violence. But as Pseudoerasmus points out, united elites are often united *against* the commoners, and we should expect inequality to be highest at times when the elites are able to work together to fight for a larger share of the pie. But I think this is the opposite of Turchin's story, where elites unite only to make concessions, and elite unity equals popular prosperity.

Fourth, everything about the elite overproduction story confuses me. Who are "elites"? This category made sense in *Secular Cycles*, which discussed agrarian societies with a distinct titled nobility. But Turchin wants to define US elites in terms of wealth, which follows a continuous distribution. And if you're defining elites by wealth, it doesn't make sense to talk about "not enough high-status positions for all elites"; if you're elite (by virtue of your great wealth), by definition you already have what you need to maintain your elite status. Turchin seems aware of this issue, and sometimes talks about "elite aspirants" – some kind of upper class who expect to be wealthy, but might or might not get that aspiration fulfilled. But then understanding elite overproduction hinges on what makes one non-rich-person person a commoner vs. another non-rich-person an "elite aspirant", and I don't remember any clear discussion of this in the book.

Fifth, what drives elite overproduction? Why do elites (as a percent of the population) increase during some periods and decrease during others? Why should this be a cycle rather than a random walk?

My guess is that *Ages of Discord* contains answers to some of these questions and I just missed them. But I missed them after reading the book pretty closely to try to find them, and I didn't feel like there were any similar holes in *Secular Cycles*. As a result, although the book had some fascinating data, I felt like it lacked a clear and lucid thesis about exactly what was going on.

V.

Accepting the data as basically right, do we have to try to wring some sense out of the theory?

The data cover a cycle and a half. That means we only sort of barely get to see the cycle "repeat". The conclusion that it is a cycle and not some disconnected trends is based only on the single coincidence that it was 70ish years from the first turning point (1820) to the second (1890), and also 70ish years from the second to the third (1960).

A parsimonious explanation would be "for some reason things were going unusually well around 1820, unusually badly around 1890, and unusually well around 1960 again." This is actually really interesting – I didn't know it was true before reading this book, and it changes my conception of American history a lot. But it's a lot less interesting than the discovery of a secular cycle.

I think the parsimonious explanation is close to what Thomas Piketty argued in his *Capital In The Twenty-First Century*. Inequality was rising until the World Wars, because that's what inequality naturally does given reasonable assumptions about growth rates. Then the Depression and World Wars wiped out a lot of existing money and power structures and made things equal again for a little while. Then inequality started rising again, because that's what inequality naturally does given reasonable assumptions about growth rates. Add in a pinch of *The Spirit Level* – inequality is a mysterious magic poison that somehow makes everything else worse – and there's not much left to be explained.

(some exceptions: why was inequality decreasing until 1820? Does inequality *really* drive political polarization? When immigration corresponds to periods of high inequality, is the immigration *causing* the inequality? And what about the 50 year cycle of violence? That's another coincidence we didn't include in the coincidence list!)

So what can we get from *Ages of Discord* that we can't get from Piketty?

First, the concept of "elite overproduction" is one that worms its way into your head. It's the sort of thing that was constantly in the background of Increasingly Competitive College Admissions: Much More Than You Wanted To Know. It's the sort of thing you think about when a million fresh-faced college graduates want to become Journalists and Shape The Conversation and Fight For Justice and realistically just end up getting ground up and spit out by clickbait websites. *Ages of Discord* didn't do a great job breaking down its exact dynamics, but I'm grateful for its work bringing it from a sort of shared unconscious assumption into the light where we can talk about it.

Second, the idea of a deep link between various indicators of goodness and badness – like wages and partisan polarization – is an important one. It forces me to reevaluate things I had considered settled, like that immigration doesn’t worsen inequality, or that inequality is not a magical curse that poisons everything.

Third, historians have to choose what events to focus on. Normal historians usually focus on the same normal events. Unusual historians sometimes focus on neglected events that support their unusual theses, so reading someone like Turchin is a good way to learn parts of history you’d never encounter otherwise. Some of these I was able to mention above – like the Mine War of 1920 or the cholera epidemic of 1849; I might make another post for some of the others.

Fourth, it tries to link events most people would consider separate – wage stagnation since 1973, the Great Stagnation in technology, the decline of Peter Thiel’s “definite optimism”, the rise of partisan polarization. I’m not sure exactly *how* it links them or what it has to stay about the link, but link them it does.

But the most important thing about this book is that Turchin claims to be able to predict the future. The book (written just before Trump was elected in 2016) ends by saying that “we live in times of intensifying structural-demographic pressures for instability”. The next bigenerational burst of violence is scheduled for about 2020 (realistically +/- a few years). It’s at a low point in the grand cycle, so it should be a doozy.

What about beyond that? It’s unclear *exactly* where he thinks we are right now in the grand cycle. If the current cycle lasts exactly as long as the last one, we would expect it to bottom out in 2030, but Turchin never claims every cycle is exactly as long. A few of his graphs suggest a hint of curvature, suggesting we might currently be in the worst of it. The socialists seem to have gotten their act together and become an important political force, which the theory predicts is a necessary precursor to change.

I think we can count the book as having made correct predictions if violence spikes in the very near future (are the current number of mass shootings enough to satisfy this requirement? I would have to see it graphed using the same measurements as past spikes), and if sometime in the next decade or so things start looking like there's a ray of light at the end of the tunnel.

I am pretty interested in finding other ways to test Turchin's theories. I'm going to ask some of my math genius friends to see if the dynamic feedback models check out; if anyone wants to help, let me know how I can help you (if money is an issue, I can send you a copy of the book, and I will definitely publish anything you find on this blog). If anyone has any other ideas for to indicators that should be correlated with the secular cycle, and ideas about how to find them, I'm intereted in that too. And if anyone thinks they can explain the elite overproduction issue, please enlighten me.

I ended my review of *Secular Cycles* by saying:

> One thing that strikes me about [Turchin]'s cycles is the ideological component. They describe how, during a growth phase, everyone is optimistic and patriotic, secure in the knowledge that there is enough for everybody. During the stagflation phase, inequality increases, but concern about inequality increases even more, zero-sum thinking predominates, and social trust craters (both because people are actually defecting, and because it's in lots of people's interest to play up the degree to which people are defecting). By the crisis phase, partisanship is much stronger than patriotism

> and radicals are talking openly about how violence is ethically obligatory.
>
> And then, eventually, things get better. There is a new Augustan Age of virtue and the reestablishment of all good things. This is a really interesting claim. Western philosophy tends to think in terms of trends, not cycles. We see everything going on around us, and we think this is some endless trend towards more partisanship, more inequality, more hatred, and more state dysfunction. But Secular Cycles offers a narrative where endless trends can end, and things can get better after all.

This is still the hope, I guess. I don't have a lot of faith in human effort to restore niceness, community, and civilization. All I can do is pray the Vast Formless Things accomplish it for us without asking us first.

List Of Passages I Highlighted In My Copy Of Ages Of Discord

Turchin has some great stories about unity vs. polarization over time. For example in the 1940s, unity became such a “problem” that concerned citizens demanded more partisanship:

> Concerned about electoral torpor and meaningless political debate, the American Political Science Association in 1946 appointed a committee to examine the role of parties in the American system. Four years later, the committee published a lengthy (and alarmed) report calling for the return of ideologically distinct and powerful political parties. Parties ought to stand for distinct sets of politics, the political scientists urged. Voters should be presented with clear choices.

I have vague memories of similar demands in the early ’90s; everyone was complaining that the parties were exactly the same and the “elites” were rigging things to make sure we didn’t have any real choices.

On the other hand, partisanship during the Civil War was pretty intense:

> Another indicator of growing intraelite conflict was the increasing incidence of violence and threatened violence in Congress, which reached a peak during the 1850s. The brutal caning that Representative Preston Brooks of South Carolina gave to Senator Charles Sumner of Massachusetts on the Senate floor in 1856 is the best known such episode, but it was not the only one. In 1842, after Representative Thomas Arnold of Tennessee “reprimanded a pro-slavery member of his

> own party, two Southern Democrats stalked towards him, at least of one of whom was arhmed with a bowie knife…calling Arnold a ‘damned coward,’ his angry colleagues threatened to cut his throat ‘from ear to ear'” (Freeman 2011). According to Senator Hammond, “The only persons who do not have a revolver and a knife are those who have two revolvers” (quoted in Potter 1976:389). During a debate in 1850, Senator Henry Foote of Mississippi pulled a pistol on Senator Thomas Hart Benton of Missouri (Freeman 2011).
>
> In another bitter debate, a New York congressman inadvertently dropped a pistol (it fell out of his pocket), and this almost precipitated a general shootout on the floor of Congress (Potter 1976: 389).

Turchin places the peak of US unity and cooperation around 1820, and partly credits the need to stand together against Indians:

> A particularly interesting case is eighteenth-century Pennyslvania (the following discussion follows closely the text in Turchin 2011:30-31). Initially, European settlers were divided by a number of ethnic and religious boundaries (Silver 2008). The English found it difficult to cooperate with the Germans and the Irish, and each ethnic group was further divided into feuding sectarian groups: Quakers against Anglicans, German Lutherans against Moravians and Mennonites. Yet, by the end of the eighteenth century, the European settlers had forged a common identity (“white people”) in opposition to the natives. As Nancy Shoemaker (2004) shoes, these “metaethnic” labels (Whites versus Reds) were not evoked as soon as settlers and natives came into contact. Rather, during the course of the eighteenth century Europeans and Indians gradually abandoned an

> initial willingness to recognize in each other a common humanity. Instead, both sides developed new stereotypes of the Other, rooted in the conviction that they were peoples fundamentally at odds, by custom and even by nature (Shoemaker 2004).
>
> The evolution of civic organizations reflected this expanding definition of common identity. Clubs with ethnic and denominational membership criteria appeared in Pennyslvania during the 1740s. These associations represented what Putnam (2000) called “bonding” rather than “bridging” social capital. For example, the St. Andrew’s Society was narrowly focused on helping the Scots, while Deutsche Gesellschaft did the same for the Germans. However, as settler-native warfare intensified, especially during the second half of the eighteenth century, the focus of civic organizations gradually shifted to charity for any victims of Indian attacks, without regard for their ethnicity or religious denomination (Silver 2008). The social scale of coorperation took a step up. Of course, there were definite limits to this new “bridging” social capital: the Indians were most emphatically excluded; in fact, the integration of “white people” developed explicitly in opposition to the Indians.
>
> Although the above description applies to pre-revolutionary Pennsylvania, a very similar dynamic obtained on the Northwestern frontier in Ohio after the Revolution (Griffin 2007). As Griffin notes, for white Americans “Indians existed as cultural glue, since the hatred of them was fast becoming a basis of order.”

This passage stood out to me because modern racial commentators focus on “whiteness” as an idea that evolved in opposition to (and to justify oppression of) blacks. But the

Indian theory makes some sense too, especially because Northerners would have more exposure to Indians than they did to black people. But I notice I’ve never heard anyone else talk about this, and most of the history books I’ve read treat Indians as too weak to be an important enemy or have much of a place in the early American consciousness.

One factor leading to greater polarization was “elite overproduction”, here represented by more office-seekers than federal offices. This was apparently a well-known problem in early America:

> Despite the increase in government posts, the supply was overwhelmed by demand for such positions. A horde of office-seekers nearly turned Jackson’s inauguration into a riot. Abraham Lincoln once said, “Were it believed that vacant places could be had at the North Pole, the road there would be lined with dead Virginians” (quoted in Potter 1976:432). And, most dramatically (although in a later period), President James Garfield was assassinated by a rejected office-seeker in 1881.

And so on. Some of Turchin’s measures of cooperation vs. polarization are a bit odd. But I have to respect the big-picture-ness of someone who will literally just look at the occurence of the word “cooperation” in various books:

> It is interesting that “culture-metric” data support Fraser’s subjective perception [of declining cooperation between business and labor]. For example, the frequency of the word “cooperation” in the corpus of American books grew rapidly during the Progressive Era and somewhat less so during the New Deal (Figure 12.3). After reaching a peak in 1940, there was a minor decline during the 1950s, followed by an increase

> toward the second peak of 1975. After 1975, however, the frequency of this word went into a sustained decline.
>
> Google Ngram is an imperfect instrument with which to trace cultural shifts. One problem is that the same word (eg, “capitalism”) can be used with either positive or negative valence, and Ngram does not allow one to separate these different meanings. “Cooperation”, however, is rarely used in the negative sense. Because of its predominantly positive valence, its overall frequency should provide us with a proxy for how much a society values cooperative values. Checking different variants (cooperation, Cooperation, cooperative, etc) yields the same overall rise-fall dynamics during the twentieth century (and up to 2008, where the current Google book database stops).
>
> Furthermore, a more specific phrase, “labor-business cooperation” again traces out the same secular cycle, although with significant differences during some decades (eg, the 1920s). Finally, “corporate greed” with its predominantly negative valence is another check on the validity of this result, and it is reassuring that during the twentieth century its frequency moved in the opposite direction from the two positive terms (to show this parallelism more clearly, Figure 12.3 plots “corporate greed” on an inverse scale).

Finally:

> There is an interesting parallel…between the Great Depression and the 1970s Bear Market. Both periods of economic hardship (although it goes without saying that the Great Depression was a much more severe crisis) were broadly interpreted as empirical evidence against the prevailing economic doctrine – the naked, laissez

> faire capitalism in the first instance, more cooperative relations between business and labor in the second. Yet it is much more likely that the primary mechanism, responsible for long-term economic decline/stagnation in each case, was the negative phase of the Kondratiev cycle, perhaps supplemented by exogenous shocks (eg, the 1973 oil embargo). Yet in each case a prolonged period of economic troubles helped to delegitimize the prevailing ideological regime (Chapter 9).

Thanks for reminding me there’s yet another cycle I need to study, one that supposedly determines the rate of technological advances. Maybe that’s my next book review.

LW Party: Bay Area Invitation

Less Wrong is throwing a party this Saturday and wants to invite any Bay Area SSC readers who are interested. Consider this an experiment to see if inviting people to parties via blog is a good idea.

It will be at 2412 MLK Jr Way in Berkeley (a private house), from 7 PM on. The Facebook page is here.

I’m going to be there; so will the Less Wrong website team and other interesting people.

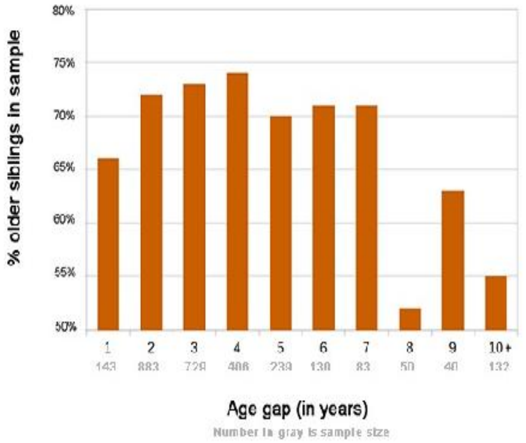
[Partial Retraction] Age Gaps and Birth Order Effects

On Less Wrong, Bucky tries to replicate my results on birth order and age gaps.

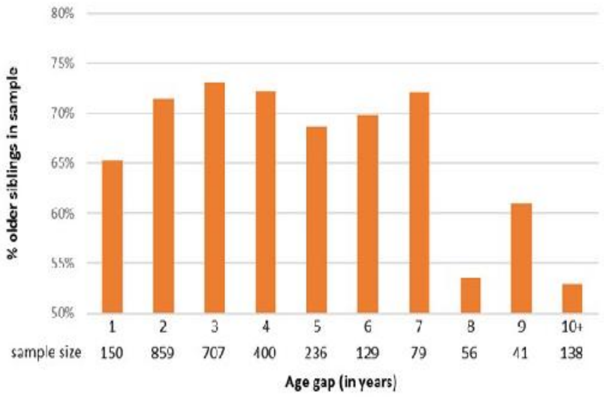
Backing up: two years ago, I looked at SSC survey data and found that firstborn children were very overrepresented. That result was replicated a few times, both in the SSC sample and in other samples of high-opennness STEM types. Last year, I expanded those results to look at how age gaps affected birth order effects. Curiously, age gaps less than seven years did not seem to attenuate birth order, but age gaps of more than seven years attenuated it almost completely.

Bucky analyzed the same data and found that I bungled one and a half of my results. Left graph in each pair is mine, right is Bucky's.

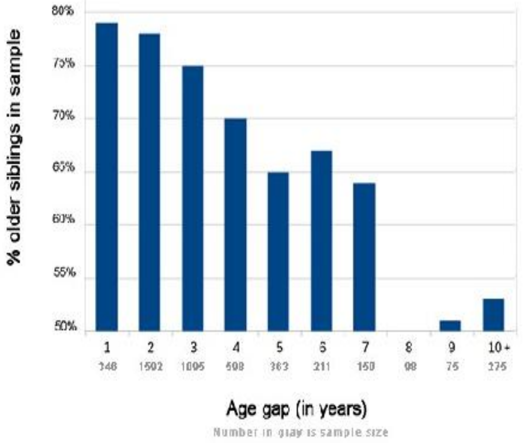
% Older Siblings In SSC Survey Sample Of People With Exactly One Sibling, By Age Gap With Sibling



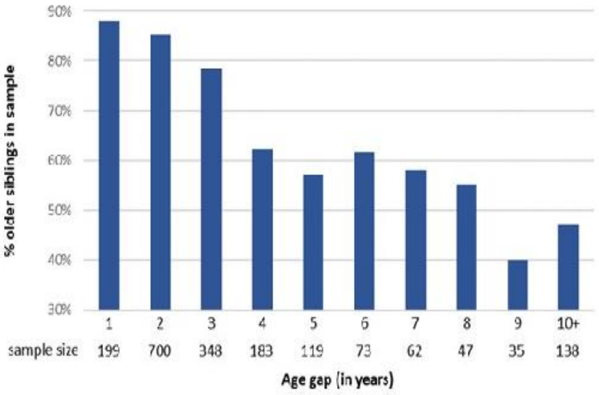
% Older Sibling In SSC Sample Of People with Exactly One Sibling, By Age Gap With Sibling



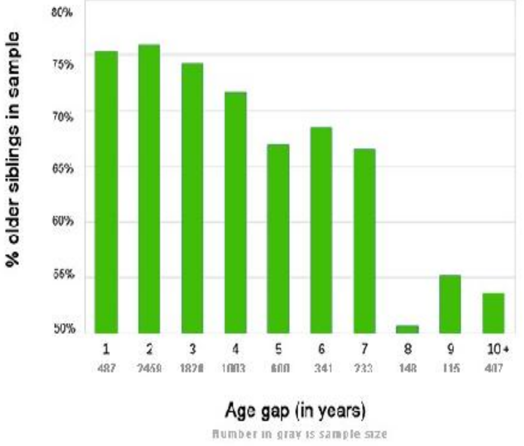
% Older Siblings In SSC Survey Sample Of People With More Than 1 Sibling, By Age Gap With Sibling



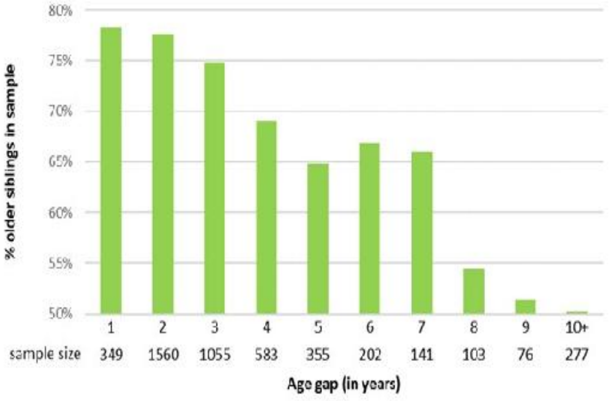
% Older Sibling In SSC Sample Of People With More Than 1 Sibling, By Age Gap With Sibling



% Older Siblings In SSC Survey Sample Total, By Age Gap With Sibling



% Older Sibling In SSC Sample Total, By Age Gap With Sibling



In the first analysis, Bucky replicates my results: in people with exactly one sibling show a sudden cliff in birth order effects after seven years.

In the second and third analysis, Bucky finds that I screwed up. I mislabeled the second analysis (people with more than one sibling) as the third (full sample), and my third analysis was just wrong (I double-counted people with one sibling).

When Bucky corrects these errors, they find that the weird cliff at seven years is present only in the sample of people with exactly one sibling. This makes it more likely to be a weird coincidence about that sample, and less likely to be a weird phenomenon. They also identify a potential confounder (there may be longer gaps between later-born children than between first-borns and second-borns) which also slightly affects the results, although does not dramatically change the conclusions.

Bucky then does their own analysis of the correct results, and finds that most likely the sudden drop at seven years is a coincidence. They conclude that:

> The SSC 2019 survey data support a constant, high, birth order effect (~2.4 oldest siblings for every 1 second oldest sibling) for age gaps less than 4-8 years. This is followed by a decline to a lower birth order effect at an undetermined rate. The decline does not necessarily completely remove any birth order effect although this may be the case for very large age gaps.
>
> The data provide some evidence that:
>
> – The reduction may not be the same (or might disappear) for larger families (4+ children)
> – Birth order effect may be lower at 1 year age gap vs 2-7 year age gap

> However the evidence for both of these points is relatively slim.

As those of you who have put up with my constant typos realize, I am not a very careful person. I try to double-check any result I present on my blog, and in some cases (including this one) ask other people to double-check them for me as well. Sometimes mistakes still slip through and I'm sorry. I am partially retracting the previous results ("partially" because some of the analyses were correct and the conclusion is still basically the same).

In response to this, I have added a note in bold to the top of the original age gaps post directing readers to the failed replication and reanalysis. I've also added a paragraph about this to my Mistakes page to help people calibrate how much to believe my future results. I am also writing this post to make sure the replication gets at least as much prominence as the original results.

I continue to make the raw survey data available for everybody to double-check my work, except for parts that could seriously compromise people's privacy. Please, if you have *any* doubt in my findings at all, do your own analysis and let me know what you get.

SSC Journal Club: Relaxed Beliefs Under Psychedelics And The Anarchic Brain

[Thanks to Sarah H. and the people at her house for help understanding this paper]

The predictive coding theory argues that the brain uses Bayesian calculations to make sense of the noisy and complex world around it. It relies heavily on priors (assumptions about what the world must be like given what it already knows) to construct models of the world, sampling only enough sense-data to double-check its models and update them when they fail. This has been a fruitful way to look at topics from depression to autism to sensory deprivation. Now, in Relaxed Beliefs Under Psychedelics And The Anarchic Brain: Toward A Unified Model Of The Brain Action Of Psychedelics, Karl Friston and Robin Carhart-Harris try to use predictive coding to explain the effects of psychedelic drugs. Then they use their theory to argue that psychedelic therapy may be helpful for “most, if not all” mental illnesses.

Priors are unconscious assumptions about reality that the brain uses to construct models. They can range all the way from basic truths like “solid objects don’t randomly disappear”, to useful rules-of-thumb like “most get-rich-quick schemes are scams”, to emotional hangups like “I am a failure”, to unfair stereotypes like “Italians are lazy”. Without any priors, the world would fail to make sense at all, turning into an endless succession of special cases without any common lessons. But if priors become too strong, a person can become closed-minded and stubborn, refusing to admit evidence that contradicts their views.

F&CH argue that psychedelics “relax” priors, giving them less power to shape experience. Part of their argument is neuropharmacologic: most psychedelics are known to work through the 5-HT2A receptor. These receptors are most common in the cortex, the default mode network, and other areas at the “top” of a brain hierarchy going from low-level sensations to high-level cognitions. The 5-HT2A receptors seem to strengthen or activate these high-level areas in some way. So:

> Consistent with hierarchical predictive processing, we maintain that the highest level of the brain’s functional architecture ordinarily exerts an important constraining and compressing influence on perception, cognition, and emotion, so that perceptual anomalies and ambiguities—as well as dissonance and incongruence—are easily and effortlessly explained away via the invocation of broad, domain-general compressive narratives. In this work, we suggest that psychedelics impair this compressive function, resulting in a decompression of the mind-at-large—and that this is their most definitive mind-manifesting action.

But their argument also hinges on the observation that psychedelics cause all the problems we would expect from weakened priors. For example, without strong priors about object permanence to constrain visual perception toward stability, we would expect the noise of the visual sensorium to make objects pulse, undulate, flicker, or dissolve. These are some of the most typical psychedelic hallucinations:

> consider the example of hallucinated motion, e.g., perceiving motion in scenes that are actually static, such as seeing walls breathing, a classic experience with moderate doses of psychedelics. This phenomenon can be fairly viewed as relatively low level, i.e., as an

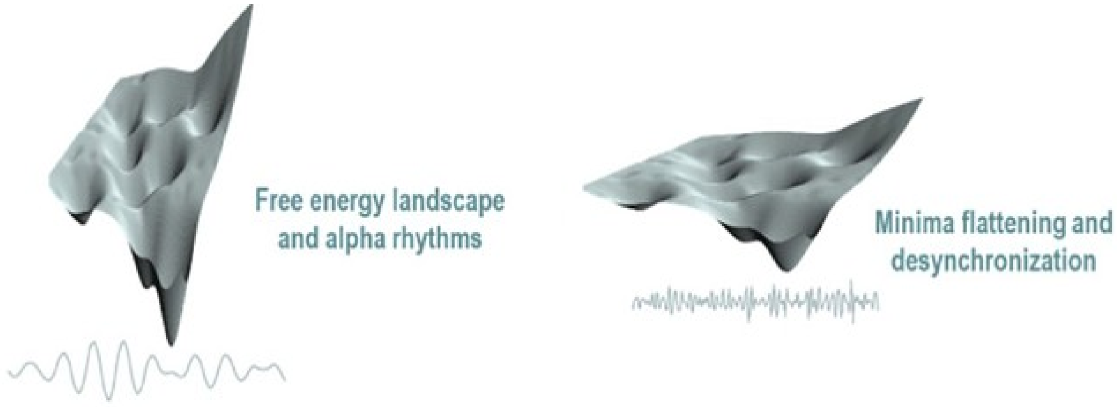
> anomaly of visual perception. However, we propose that its basis in the brain is not necessarily entirely low-level but may also arise due to an inability of high-level cortex to effectively constrain the relevant lower levels of the (visual) system. These levels include cortical regions that send information to V5, the motion-sensitive module of the visual system. Ordinarily, the assumption “walls don’t breathe” is so heavily weighted that it is rendered implicit (and therefore effectively silent) by a confident (highly-weighted) summarizing prior or compressive model. However, under a psychedelic, V5 may be forced to interpret increased signaling arising from lower-level units because of a functional negligence, not just within V5 itself, but also higher up in the hierarchy. Findings of impaired high- but not low-level motion perception with psilocybin could be interpreted as broadly consistent with this model, namely, pinning the main source of disruption high up in the brain’s functional hierarchy.

But F&CH are most interested in whether psychedelics can cause the positive effects we would expect of relaxed priors. If overly strong priors cause closed-mindedness, psychedelics should allow users to “see things with new eyes” and change their minds about important issues. These changes would be precipitated by the drug, but not fundamentally *about* the drug. For example, imagine a person who formed a strong prior around “I am a failure” at a young age, then later went on to achieve great things. Because their prior was so strong, they might think of each of their accomplishments as a special case, or interpret them as less impressive than they were. On psychedelics, they could reexamine the evidence unbiased by their existing beliefs, determine that their accomplishments were impressive enough to count as successes, and abandon the “I am a failure” prior. They would continue understanding that they were successful even after they sobered up, because the change of mind was a triumph of rationality and not a drug-fueled hallucination.

When I was young, I liked a fantasy book called *The Sword of Shannara*. The titular sword had an unusual magic power: it made the wielder realize anything he already knew. Neither a truly good person nor a truly bad person would benefit from the sword, but someone who was hypocritical, or deluding themselves, or complicit in their own brainwashing, would find all the parts of their mind flung together so hard that they couldn't help but realize all the inconvenient facts they were trying to repress, or connect all the puzzle pieces previously scattered in separate mental compartments. Sometimes this resulted in a blinding revelation that you were on the wrong side, or had wasted your life; other times in nothing at all. F&CH argue that psychedelics are the real-life version of this, a way to make all of your beliefs connect with each other and see what results from the reaction.

These dignified scientists don't like magic-sword related analogies, so they stick to regular-sword-related ones. "Annealing" is a concept in metallurgy where blacksmiths heat a metal object in a forge until it undergoes a phase change. All the molecules move to occupy whatever the lowest-energy place for them to occupy is, strengthening the metal's structure. Then the metal object leaves the forge and the metal freezes in the new, better configuration.

They analogize the same process to "flattening an energy landscape". Imagine a landscape of hills and valleys. You are an ant placed at a random point in the landscape. You usually slide downhill at a certain rate, but for short periods you can occasionally go uphill if you think it would help. Your goal is to go as far downhill as possible. If you just follow gravity, you will end up in a valley, but it might not be the *deepest* valley. You might get stuck at a "local minimum"; a valley deep enough that you can’t climb out of it, but still not as deep as other places in the landscape you will never find. F&CH imagine a belief landscape in which the height of a point equals the strength of your priors around that belief. If you settle in a suboptimal local minimum, you may never get out of it to find a better point in belief-space that more accurately matches your experience. By globally relaxing priors, psychedelics flatten the energy landscape and make it easier for the ant to crawl out of the shallow valley and start searching for even deeper terrain. Once the drugs wear off, the energy landscape will resume its normal altitude, but the ant will still be in the deeper, better valley.


Free energy landscape and alpha rhythms
Minima flattening and desynchronization

Here F&CH are clearly thinking of recent research that suggests MDMA treats post-traumatic stress disorder. Post-traumatic stress disorder is well-modeled as a dysfunctional prior, something like “the world is unsafe” or “you’re still in that jungle in ‘Nam, about to be ambushed”. Many PTSD patients have gone on to live good lives in well-functioning communities and now have more than enough evidence that they are safe. But the evidence doesn’t “propagate”; the belief structure is “frozen” in place and cannot be updated. If psychedelics relax strong priors, they can flatten the energy landscape and allow the ant of consciousness to leave the high-walled “the world is unsafe” valley and test the terrain in “the world is actually okay”. And since the latter is a deeper valley (more accurate belief) than the former, the patient will remain there after the drug trip wears off. This seems to really work; the effect size of MDMA on PTSD is very impressive.

But the authors want to go further than that. They write:

> In this study, we take the position that most, if not all, expressions of mental illness can be traced to aberrations in the normal mechanics of hierarchical predictive coding, particularly in the precision weighing of both high-level priors and prediction error. We also propose that, if delivered well (Carhart-Harris et al., 2018c), psychedelic therapy can be helpful for such a broad range of disorders precisely because psychedelics work pharmacologically (5-HT2AR agonism) and neurophysiologically (increased excitability of deep-layer pyramidal neurons) to relax the precision weighting of high-level priors (instantiated by high-level cortex) such that they become more sensitive to context (e.g., via sensitivity to bottom-up information flow intrinsic to the system) and amenable to revision (Carhart-Harris, 2018b).

"Most if not all" psychiatric disorders. This has some precedent: some people are already thinking of depression as a high-level prior on negative perceptions and events, obsessive-compulsive disorder as strong priors on the subject of the obsession, etc. But it's is a really strong claim, and Friston himself has previously published models of depression and anxiety that don't obviously seem to mesh with this. I wonder if this is Carhart-Harris' overenthusiasm for psychedelics running a little ahead of the evidence.

Speaking of Carhart-Harris' overenthusiasm for psychedelics running a little ahead of the evidence, the paper ends with a weird section comparing the hierarchial structure of the brain to the hierarchical structure of society, and speculating that just as psychedelics cause an “anarchic brain” where the highest-level brain structures fail to “govern” lower-level activity, so they may cause society to dissolve or something:

> Two figureheads in psychedelic research and therapy, Stanislav Grof and Roland Griffiths, have highlighted how psychedelics have historically “loosed the Dionysian element” (Pollan, 2018) to the discomfort of the ruling elite, i.e., not just in 1960s America but also centuries earlier when conquistadors suppressed the use of psychedelic plants by indigenous people of the same continent. Former Harvard psychology professor, turned psychedelic evangelist, Timothy Leary, cajoled that LSD could stand for “Let the State Dissolve” (Pollan, 2018). Whatever the interaction between psychedelic use and political perspective, we hope that psychedelic science will be given the best possible opportunity to positively impact on psychology, psychiatry, and society in the coming decades—so that it may achieve its promise of significantly advancing self-understanding and health care.

Sure, whatever. But this might be a good time to go back and notice some of the slight discordant notes scattered throughout the paper.

In a paragraph on HPPD, F&CH write:

> Hallucinogen-persisting perceptual disorder (HPPD) is a Diagnostic and Statistical Manual of Mental Disorders, 5th Edition–listed disorder that relates to enduring visual perceptual abnormalities that persist beyond an acute psychedelic drug experience. Its prevalence appears to be low and its etiology complex, but symptoms can still be distressing for individuals (Halpern et al., 2018). Under the REBUS model, it is natural to speculate that HPPD may occur if/when the collapse of hierarchical

> message passing does not fully recover. A compromised hierarchy would imply a compromised suppression of prediction error, and it is natural to assume that persistent perceptual abnormalities reflect attempts to explain away irreducible prediction errors. Future brain-imaging work could examine whether aspects of hierarchical message passing, such as top-down effective connectivity, are indeed compromised in individuals reporting HPPD.

In other words, the priors relax and don’t unrelax again after the drug experience.

For example, an especially common HPPD experience is seeing solid objects pulsate, ooze, or sway. It’s not surprising that a noisy visual system would sometimes put the edge of an object in one place rather than another. But usually a strong prior on “solid objects are not pulsating” prevents this from interfering with perception. Relax this prior too far and the pulsating becomes apparent. If the prior stays relaxed after the drug trip ends, you’ll keep seeing the pulsation indefinitely.

This is one of two plausible theories of HPPD, the other being that the hours of seeing objects pulsate makes your brain learn a new prior, “objects *do* pulsate” and stick to it. This would make more sense in the context of other learned permanent perceptual disorders like mal de debarquement.

F&CH include a section called “What To Do About The Woo?”, where they admit many people who have psychedelic experiences end up believing strange things: ghosts, mysticism, conspiracies. They are not very worried about this, positing that “a strong psychedelic experience can cause such an ontological shock that the experiencer feels compelled to reach for some kind of explanation” and arguing that as long as we remind people that science is good and pseudoscience is bad, they should be fine.

But I still worry that psychedelic woo is the cognitive equivalent of HPPD.

On one reading, it’s the failure of relaxed priors to re-strengthen, so that beliefs that previously had low prior probability – “this phenomenon is explained by ghosts”, “this guy at the subway station preaching universal love has really discovered all the secrets of the universe” – become more compelling. Spiritual beliefs are kind of a Pascal’s Wager type of deal – extremely important if true, but so unlikely to be true that we don’t usually pay much attention to them. If someone is walking around with a permanently flattened energy landscape – if all their probabilities are smushed together so that unlikely things don’t seem *that much* more unlikely than likely ones – then the calculation goes the other way, and the fascinating nature of these beliefs overcomes their improbability to make them seem worthy of attention.

On the other reading, they’re the result of newly-established priors in favor of ghosts and mysticism and conspiracies. People are not actually very good at reasoning. If you metaphorically heat up their brain to a temperature that dissolves all their preconceptions and forces them to basically reroll all of their beliefs, then a few of them that were previously correct are going to come out wrong. F&CH’s theory that they are merely letting evidence propagate more fluidly through the system runs up against the problem where, most of the time, if you have to use evidence unguided by any common sense, you probably get a lot of things wrong.

F&CH aren’t the first people to discuss this theory of psychedelics. It’s been in the air for a couple of years now – and props to local bloggers at the Qualia Research Institute and Mad.Science.Blog for getting good explanations up before the parts had even all come together in journal articles. I’m especially interested in QRI’s theory that meditation has the same kind of annealing effect, which I think would explain a lot.

But F&CH’s paper lends the theory a new level of credibility. Carhart-Harris is one of the pioneers of psychedelic therapy, and the paper looks like it’s intended to get people more interested in and accepting of that work by providing a promising theoretical basis. If so, mission accomplished.

Lots Of People Going Around With Mild Hallucinations All The Time

*[**Related to:** Relaxed Beliefs Under Psychedelics And The Anarchic Brain, HPPD And The Specter Of Permanent Side Effects]*

I.

Hallucinogen persisting perceptual disorder is a condition where people who take psychedelics continue hallucinating indefinitely. Estimates of prevalence range from about 4% of users (Baggott) to "nobody, the condition does not exist" (Krebs and Johansen). To explore this discrepancy, I asked about it on the 2019 SSC survey. The specific question was:

> Hallucinogen Persisting Perceptual Disorder is a condition marked by visual or other perceptual disturbances typical of psychedelic use that continue for weeks and months after coming off the psychedelic, in some cases permanently. Have you ever had this condition?

2,234 readers admitted to having used psychedelics. Of those, 285 (= 12.8%) stated that they had some hallucinations that persisted afterwards. 219 (9.8%) said they'd had them for a while and then they had gone away. 66 (= 3%) stated that they still had the hallucinations (one limit of the study: I don't know how long it has been since those people took the psychedelics).

But most of these people reported very mild experiences; on a scale from 1-10, the median severity was 2. The most commonly reported changes were more "visual snow" (ie "static" in the visual field), slight haloes around objects, and visual trails. Many people reported that if they stared at a textured area like a carpet long enough, illusory geometric patterns would start to form. Only a few people noticed anything weirder than this. Here are some sample responses:

> — It took the form of visual disturbances (patterns, moving patterns, mild hallucinations like seeing a flower growing on a person’s face or seeing the legs of a chair walk). Initially there would be daily incidents. They seemed to be triggered by fatigue or other stress, and I did not like them, which lead to more stress. Exercise helped reduce the stress and make the visual effects go away. The severity and frequency dropped quickly at first and then more slowly over the years, essentially gone after a decade. I have had migraines with ‘fortification figures’ as well; this was different from that.
>
> — Perceptual distortions about once a week. Difficulty concentrating for about 6 months. Intense interest in drawing.
>
> — Used LSD about twice a month for a period of time when I was 17, and had a very very slight build up of persisting visual distortions. This phase culminated in me taking ~425μg by accident one night, and having a somewhat negative experience, had an ego death sort of thing, experienced strong “time-warping” where my subjective experience of the passing of time slowed down and sped up repeatedly by a large degree. Very strong acid trip. After that night I experienced pretty noticeable breathing on certain textures, like ceiling tiles, and a sort of “bright aura effect” on others, like pebbles. This gradually receded over something like 1-2 years, at which point it was entirely gone.
>
> — straight lines often bend! it’s distracting!

— If I stare at an object, it’ll appear to slightly sway or move in a circular motion. If I stare at a surface (especially a textured one like a carpet), faint light patterns will develop. I got HPPD the 3rd time I did LSD. I haven’t done anything to treat it, expecting it would go away on its own over time. However, it’s been like 4 years and hasn’t really changed much. I’ve gotten used to it and often forget I have HPPD. I told an optometrist about it one time and he was just like “yeah, happens to some people who do acid.”

— Quite concerned. Seen by PCP and opthamologist. Performed a literature review myself and tried clonidine without a large effect. Symptoms have seemed to resolve however.

— Persistent discolouration blue-ward or red-ward, depending on *subjective* head orientation, in a pattern of visual field roughly equivalent to the Tetragrammaton. Yes bloody really. Help.

— My colour perception was distorted and confusing for some time, perhaps 1-3 years with browns, greens and reds, especially in natural settings (2C* series) I also experienced intense flashbacks consisting of bodily tingling and diffuse feeling of heat for ~10 months after a different psychedelic session.

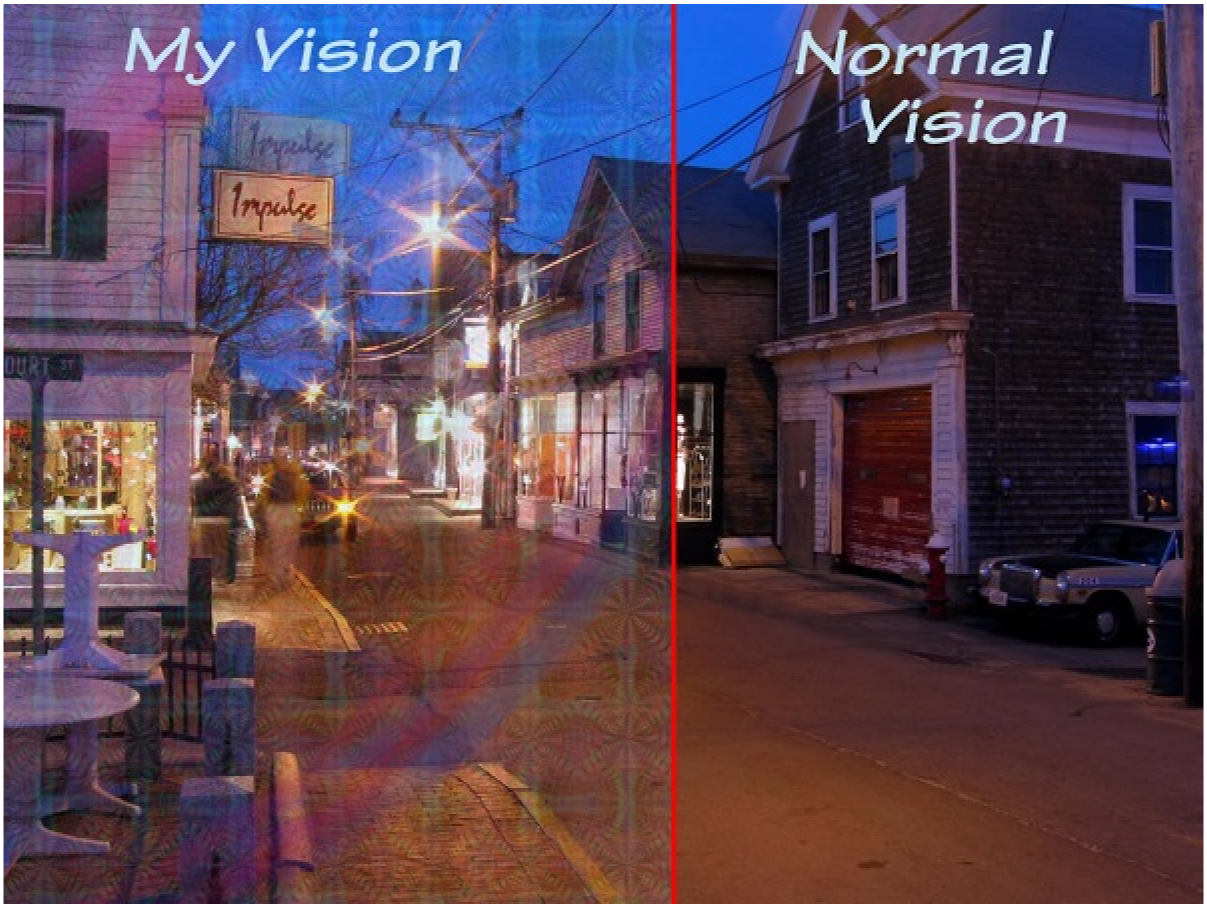
— I see tracers, halos, have increased anxiety, visual snow. Nothing has helped so far.

— Very mild but persistent visual distortions. Walls breathing, kaleidoscoping of repeating patterns, etc. Not severe enough to warrant treatment.

— Any highly detailed or repetitive surface “breathes” after looking at it for even a few seconds. get glowing

> patterns when looking at some art, a sort of cycle of different colors being highlighted in sequence. makes eyes open meditation easier; makes “where’s waldo” harder.
>
> — Visual snow; when viewing images which contain several different possible interpretations of patterns (eg hex grid on bathroom tile, carpet weaving pattern) my perception of the most salient pattern stutters and vacillates between the top 3-5 possible interpretations.
>
> — I hesitated whether to even answer ‘yes’. 15-25 years ago, a period when I occasionally used psychedelics, I would sometimes have mild visual trails or aberrations when I wasn’t on any psychedelics. It never really bothered me. I can still sometimes make stationary patterns swirl or move by staring at them, but I think that’s just normal visual perception, not HPPD. Right?

A Reddit user helpfully illustrated what his (particularly severe) HPPD looked like (note especially the subtle square grid in left picture):



Weird, but not unbearably so. I think this explains the wide variety of prevalence estimates. Many people who take psychedelics will have very minor permanent changes to their vision. Most of them will shrug and get on with their lives. Only the few people whose changes are especially bad, or who get especially neurotic about it, will ever talk to a doctor.

I discussed some of this recently at a colloquium of experienced psychonauts (realistically: a group house full of the kind of programmers who go to Burning Man). Of the ten or so people there, two admitted to HPPD which sounded about as bad as the image above. They'd never really mentioned it before, it hadn't come up, and they were living normal lives.

I conclude that lots of people you encounter are having mild hallucinations all the time, and it just never comes up.

II.

I want to talk more about Krebs & Johansen, the paper that claims to find HPPD doesn’t exist.

First, it talks about surveys of special populations, like Native Americans using peyote. None of them report HPPD. Plausibly peyote does not cause HPPD (most of the people on my survey blame LSD) or something about the cultural set and setting prevents Native Americans from getting it during ceremonies. I see thedrugclassroom.com suspects this too.

Second, it talks about K&J’s own study, based in the National Survey of Drug Use and Health. Among 130,152 participants, those who have used psychedelics are no more likely to report symptoms than those who have not. But this survey violates one of the cardinal sins of psychiatric questioning – be careful in how you ask questions about psychosis-like symptoms! (see here for more people getting this wrong). K&J base their finding entirely on how people answer the question “Saw vision others could not?” (they don’t give the presumably longer question for which this is an abbreviation). But it is asked after several other questions clearly probing psychotic symptoms, like “felt force taking over your mind?”, “felt force inserting thoughts?”, and “heard voices?” Participants probably assumed (I think correctly; K&J did not write the survey) that the question was asking about more dramatic visions, like visions from God about their destiny or something. The average person whose visual field was 5% more staticky than normal wasn’t going to answer.

Third, it talks about HPPD-like visual hallucinations in people who have never taken psychedelic drugs. It says many people have them. I believe this. At the colloquium I mentioned, someone talked about having some similar staticky symptoms before taking any psychedelics. On my own survey, a few people said things like “I could always make the carpet self-organize into geometric patterns if I tried really hard, but after I took LSD it started happening much more”. I don’t know of any surveys on how prevalent this is, but since nobody has any trouble finding examples, I imagine it’s pretty common. My only argument that this doesn’t disprove HPPD is how many people – including my survey respondents – describe these oddities starting (or getting much worse) right after they use psychedelics. But we have no idea how many totally drug-naive people are just going around with weird visual experiences just because that’s the kind of person they are.

III.

I want to talk about a case study of this last kind of person (as usual, this is a composite with some details changed to protect privacy). Patient of mine for a totally unrelated reason – she had mild depression after a death in the family, wanted to talk it over. Gradually it started coming out that she had an entire spiritual system based on what part of a grid she thought people fell on. It took me a long time to figure out the grid was literal. It was a shape she could see in her vision, all the time, close to a Cartesian plane but with a couple of extra lines and curves on it. It looked similar in spirit to the grid pattern tiled across the sample image above, though not exactly the same.

In terms of her personality – she was very into William Blake, Meister Eckhart, and ecstatic poetry. She dabbled in Wicca and occultism. She had strong opinions on chakras.

Her hair was dyed an even brighter shade of purple than the Bay Area median. I was mildly concerned she was going to either join or start a cult, but she never went further than sharing some of her manifestoes about the nature of time-space with her lucky psychiatrist.

This woman had never used psychedelics. But her description of her visual grid matched some HPPD phenomena pretty well.

I still dream of linking individual differences in perception with differences in cognition. So I wonder: psychedelics make you more likely to believe woo, and more likely to have certain classes of visual hallucinations. But some people believe woo without taking psychedelics, and some people have those same visual hallucinations without taking psychedelics. Are they the same people? Do psychedelics move you along a natural axis of variation, maybe from unusually strong priors to unusually relaxed priors? Was my patient the sort of person that the rest of us would become if we took massive doses of LSD every day for years? Did that explain both her perceptual and her cognitive uniqueness?

Start thinking along these lines, and some other become pretty suggestive. How come some New Agey people say they can see auras around people? Are they just lying? Seems like a weird thing to lie about. And a lot of these people don’t *sound* like they’re lying. Aren’t auras a classic LSD hallucination? I understand they’re not quite as simple as the haloes around lights that HPPD people get; they’re only around people and sometimes the colors seem meaningful. But add something about handwave handwave using a special kind of visual processing for other people handwave synaesthesia, and maybe it’s not totally outside the realm of possibility? Maybe if your priors are so relaxed

– ie so far towards the “naturally on LSD all the time” side of the scale – that you believe in auras, then your priors might also be so relaxed that you can see them.

We really don’t know what other people’s perceptions are like. At least, not until I ask a bunch of questions about this on the next SSC survey.

SSC Meetups 2019: Times And Places

Thanks to everyone who offered to host a meetup. Full list of cities, times, and places is below. If you're reading this, you're invited. Please don't feel like you "won't be welcome" just because you're new to the blog, demographically different from the average reader, or hate SSC and everything it stands for. You'll be fine!

Some suggestions for organizers:

1. Bring a sign that says SSC MEETUP so people can find you
2. Bring nametags and markers
3. Bring a signup sheet where people can write their names and emails if they want to hear about future meetups.
4. If people want to get to know each other better outside the meetup, you might want to mention reciprocity.io, the rationalist friend-finder/dating site. It runs off Facebook, so you have to Facebook friend the other person first.
5. Please record how many people attend; I will ask for these numbers to help with future meetup posts.
6. If you take a picture and send it to me, I'll try to post it here. I'll ask for this later, please don't email these to me until then.

If a city has **asterisks** around its name, nobody volunteered to host a meetup there, but SSC meetup director Mingyuan suspected you might have interested readers anyway. She chose a time, place, and random coffee shop as a Schelling point for these cities. There won't be any organizers around, and no guarantee of anyone at all, but you can still go. If you go to these, in order to make it easier for everyone to find each other, try to arrive on time, wear something blue on top (shirt, hoodie, etc) as a signal, and/or bring SSC signs.

If a city's name is in **blue**, that means I'll be attending the meetup. I'll try to announce each of these on the blog, and some of them might be subject to schedule change, so watch this space. A few people from various rationalist organizations interested in getting to know the further-flung parts of the community might be tagging along with me for some of the way.

Please look over your meetup to make sure it's correct. If you have corrections, further details, or a new meetup to add, please post a top-level comment below containing the text "@MINGYUAN" so she can find it easily.

LIST OF CITIES

Albany, NY
Time: Saturday, September 21, 10:00 AM
Location: Stacks Espresso, 260 Lark St.
Contact: pellinorew[at]gmail

Amsterdam, Netherlands
Time: Saturday, September 28, 3:00 PM
Location: Grass in front of Weesperzijde 100
Contact: mathijs.henquet[at]gmail; Facebook event

Ann Arbor, MI
Time: Thursday, September 26, 6:30 PM
Location: TBD, check http://bit.ly/2IlCqfW closer to the date for updates
Contact: algo2217[at]gmail

Atlanta, GA
Time: Saturday, October 12, 2:00 PM
Location: Hodgepodge Coffeehouse, 720 Moreland Avenue Southeast
Contact: steve[at]digitaltoolfactory[dot]net

Auckland, New Zealand
Time: Sunday, September 22, 4:00 PM
Location: Lim Chhour Food Court, 184 Karangahape Rd.
Contact: benjsmith[at]gmail; +64 22 685-4105; Facebook event
Details: Parking is available at Wilson's Car Park on Cross Street. You may also find free street parking. There are excellent public transport links from many parts of the city. Please consider perusing the wonderful range of food options and purchasing a drink, snack, or meal to support our hosts. If you can't fit it in the budget or just aren't hungry, though, please come anyway as we'd love to have you!

Austin, TX
Time: Monday, September 30, 6:00 PM
Location: Central Market North Lamar, 4001 N Lamar Blvd.
Contact: azavoluk[at]gmail; Google group
Details: Either upstairs on the balcony or outside in the courtyard. Will email the google group, but we should be easy to spot.

Baltimore, MD
Time: Sunday, September 29, 7:00 PM
Location: 4th Floor, UMBC Performing Arts and Humanities Building
Contact: rivka[at]adrusi[dot]com

Bangalore, India
Additional search terms: Bengaluru
Time: Sunday, September 22, 4:00 PM
Location: Matteo Coffea, Church Street, Near MG road Metro Station
Contact: m.nihalmohan[at]gmail

Barcelona, Spain
Time: Saturday, September 21, 4:00 PM
Location: Plaça d’Espanya, 08004 Barcelona, Spain
Contact: ana_bcn[at]protonmail[dot]com
Details: We meet in Plaça d’Espanya at the big tower closest to Fira. Then depending on the weather we can go to Montjuic or Arenas. Feel free to send me an e-mail even if you can’t attend, so that we may catch up at another time.

Berkeley, CA
Time: Thursday, October 10, 6:00 PM
Location: “Event Horizon” (private residence), 2412 Martin Luther King Jr. Way
Contact: mingyuan[at]uchicago[dot]edu; Facebook event

Berlin, Germany
Time: Sunday, September 22, 7:00 PM
Location: Turmstr. 10, 10559 Berlin
Contact: ssc[at]alphabattle[dot]xyz
Details: The place is a club house for a club I’m in (JVB), so a store window, hard to miss. I’ll put up an SSC sign. Bus stops right in front, or subway to Turmstr or S-Bahn to Bellevue. You can buy drinks at ~cost. There are various takeaway options around if you want to bring something or order in. We’ll have the place to ourselves and can easily seat 30 people.

Boston, MA
Time: Friday, September 20, 7:00 PM
Location: 199 Harvard St Apt 2, Cambridge (about 8 minutes walk from Central MBTA station)
Contact: boston-lw-organizers[at]googlegroups[dot]com, Facebook event, Less Wrong event

Boulder, CO
Time: Monday, September 16, 7:15 PM

Location: The Rayback Collective, 2775 Valmont Rd.
Contact: Meetup.com
Details: I will have some sort of sign or folded paper that says “LW RRG” on it or something similar. Look for a tall man with long hair and a beard. We usually sit in the Southwest corner if possible, or in one of the couch circles along the west wall.

Brighton, UK
Time: Saturday, September 21, 2:00 PM
Location: Starbucks, 201 Western Rd.
Contact: None
Details: Meet in the upstairs seating area.

Brisbane, Australia
Time: Friday, October 11, 7:00 PM
Location: Level 1/210 Wickham St, Fortitude Valley QLD 4006, Australia
Contact: jarred.filmer[at]gmail
Details: The Boiler Room, Fortitude Valley, the bar behind the mural of Bill Murrary at Tomcat. Head up the stairs then curve left, you’ll see it.

Bristol, UK
Time: Thursday, September 26, 7:30 PM
Location: The Robin Hood, 56 St Michael’s Hill
Contact: Facebook event

Brussels, Belgium
Time: Saturday, September 21, 2:00 PM
Location: The Judgy Vegan, Rue des Capucins 55
Contact: None

Budapest, Hungary
Time: Saturday, September 21, 2:00 PM
Location: Starbucks, Egyetem tér 4, 1053 Hungary

Contact: None
Details: Meet in the outdoor seating area, if the weather allows

Buffalo, NY
Time: Sunday, September 29, 10:00 AM
Location: Five Points Bakery, 426 Rhode Island St.
Contact: spetey[at]gmail

Calgary, Canada
Time: Sunday, September 22, 1:00 PM
Location: Weeds Cafe, 1903 20 Ave NW
Contact: ssccalgary[at]yahoo[dot]com
Details: Please email me if you are going to attend. I live over an hour away, and would just like to make sure I don’t have to do the long drive if nobody will be attending, as the amount of interest in Calgary might be on the low side. Similarly, I would like to encourage anybody who is on the shy side, or who might be anxious about a small turnout to attend as well. To this end, if it looks like we’re going to have a group of 3 or less people, I will let you know beforehand and you can reconsider attending if you like; this way, you won’t have to worry that you’re going to be the only one to show up.

Cambridge, UK
Time: Wednesday, October 16, 7:00 PM
Location: The Burleigh Arms, 9-11 Newmarket Rd.
Contact: rlm72[at]cam[dot]ac[dot]uk

Canberra, ACT, Australia
Time: Wednesday, September 25, 6:00 PM
Location: The Snug Room at King O’Malley’s in Civic Centre
Contact: andy.bachler[at]gmail
Details: I have booked the Snug Room at King O’Malley’s

which is a quiet room behind the bar. Should be low key and quiet enough to allow easy conversation.

Cape Town, South Africa
Time: Saturday, September 28, 11:00 AM
Location: Haas Coffee, 19 Buitenkant St, Cape Town City Centre
Contact: rashtayob[at]gmail

Champaign-Urbana, IL
Additional search terms: UIUC
Time: Saturday, September 21, 2:00 PM
Location: Illini Union Basement
Contact: None

Charlotte, NC
Time: Monday, September 16, 7:00 PM
Location: The Mighty Meeple, 8440 Pit Stop Ct NW #180, Concord, NC 28027
Contact: nospam[at]catallaxymedia[dot]com
Details: Mighty Meeple is a game store with a large backroom area of tables as well as a ton of free board games to play during this time frame for their Board Game Night. You can also bring your own food, although they have drinks and snacks available for purchase. My plan will be to sit at a long table with my young adult family and play games with a sign reading SSC so you can find us. There will be at least 4-6 people already. We can chat and/or play various games together. Please email me to introduce yourself or RSVP. I’ll reply-all just before so everyone attending knows what to expect.

Chicago, IL
Time: Saturday, September 28, 6:00 PM
Location: South Loop Strength & Conditioning, 645 S Clark St.

Contact: todd[at]southloopsc[dot]com
Details: The location is my gym, so we've got a large open space ripe for ~~exercising~~ socializing.

Cleveland, OH
Time: Saturday, September 28, 1:00 PM
Location: Tabletop Cleveland, 1810 W 25th St.
Contact: ferbfreeman[at]gmail

Cologne, Germany
Time: Saturday, September 21, 5:00 PM
Location: Marienweg 43, 50858 Köln
Contact: marcel_mueller[at]mail[dot]de
Details: Private venue, everyone who wants to attend is welcome. If you are unable to find us call +491788862254.

Columbus, OH
Time: Sunday, September 22, 6:00 PM
Location: 397 Eldridge Ave.
Contact: james.thomas.hays[at]gmail
Details: Private residence. Food will be provided. RSVPs preferred to help with planning purposes, but not required.

Copenhagen, Denmark
Time: Saturday, September 21, 7:00 PM
Location: Studenterhuset, Købmagergade 52, Denmark
Contact: soeren.elverlin[at]gmail

Dallas, TX
Time: Sunday, September 15, 5:00 PM
Location: Magic Cup Cafe, North Jupiter Road, Richardson, TX
Contact: tayfie[at]pm[dot]me; LessWrong

Darmstadt, Germany
Time: Sunday, October 20, 2:00 PM

Location: Wilhelminenstraße 17
Contact: ssc.darmstadt[at]outlook[dot]de

Denver, CO
Time: Saturday, October 5, 6:00 PM
Location: 1520 1/2 Lowell Blvd
Contact: fsjdnjdnsjkuuoqq[at]aol[dot]com

Dublin, Ireland
Time: Saturday, September 21, 2:00 PM
Location: Starbucks, Stephen Court, GF1, St Stephen’s Green, Dublin 2, D02 N960, Ireland
Contact: enagohan[at]gmail; Facebook event; LessWrong

Edinburgh, Scotland
Time: Saturday, September 21, 2:00 PM
Location: Starbucks, 124 High St, Edinburgh EH1 1SG, UK
Contact: None
Details: Meet in the upstairs seating area

Fairbanks, AK
Time: Sunday, October 6, 9:00 PM
Location: Starbucks, 3755 Airport Way
Contact: scott[at]slatestarcodex[dot]com
Details: May or may not happen; await details on blog the day before

Halifax, Nova Scotia, Canada
Time: Saturday, September 21, 7:00 PM
Location: Uncommon Grounds, 1030 South Park Street
Contact: ahobata[at]gmail

Hamburg, Germany
Time: Saturday, September 21, 5:00 PM
Location: Troy Office, Maienweg 310, 22335 Hamburg, Germany
Contact: gunnar.zarncke[at]gmx[dot]de

Details: We are meeting in the small company office of the fintech startup troy in the 4th floor. Just ring at “troy”. Depending on the weather we can sit on our nice balcony. It would be nice if guests could bring blankets. This is not an official company event but privately organized. Some foods and drinks might be provided, but I’d appreciate if you contribute. There will be some prepared ice breakers and discussion topics, but you are free to contribute or just meet and talk.

Helsinki, Finland
Time: Tuesday, September 24, 6:00 PM
Location: Restaurant Dubliner Kaivopiha, Mannerheimintie 5
Contact: joenash499[at]gmail
Details: We will be in the private room called Guinness Lounge. You can expect something like a normal night at the pub, except with a group consisting of other SSC readers (and without the background noise, since we’ll be in a private room.) You don’t have to contact anyone in advance, you can just show up.

****Honolulu, HI****
Time: Saturday, September 21, 2:00 PM
Location: Starbucks, 680 Ala Moana Blvd A101
Contact: None

Houston, TX
Time: Sunday, September 15, 2:00 PM
Location: Brasil Cafe, 2604 Dunlavy St.
Contact: walambert.work[at]gmail; LessWrong; Facebook group; Google group; Discord server (link expires after 5 uses, please contact me if you want to join our Discord but the link has stopped working)

Indianapolis, IN
Time: Saturday, September 21, 4:00 PM

Location: ShoeFly Public House, 122 East 22nd Street
Contact: molybdenumblue[at]gmail

Iowa City, IA
Time: Sunday, September 15, 1:00 PM
Location: The Java House, 211 E Washington St.
Contact: potentgamer[at]gmail

Irvine, CA
See “Orange County”, below.

Jacksonville, FL
Time: Saturday, September 21, 2:00 PM
Location: Starbucks, 1980 San Marco Blvd
Contact: None

Kansas City, KS
Time: Sunday, September 22, 12:30 PM
Location: Panera Bread, 10606 Shawnee Mission Pkwy, Shawnee, KS
Contact: alex.hedtke[at]gmail; Facebook event

Kiev, Ukraine
Additional search terms: Kyiv
Time: Saturday, September 14, 4:00 PM
Location: On the steps to the National Museum of Nature (Underground station “Teatral’na”)
Contact: Marichka, (+38)0976672970
Details: We shall move from there, so better don’t be late.

Kitchener-Waterloo, ON
Time: Sunday, September 29, 2:00 PM
Location: Field of Waterloo CI
Contact: blxxia[at]gmail
Details: I’ll wear red converse and bring some frisbees for those that are athletically (or simply disc-) inclined. We’ll walk the half a block to my apartment at around 3:30ish and

have the rest of the meetup there with tea and snacks, if space allows.

Lexington, KY
Additional search terms: Louisville, Cincinnati, Frankfort, Covington, Shelbyville
Time: Sunday, September 22, 7:00 PM
Location: West Sixth Brewery, 501 West Sixth Street, Suite 100
Contact: hellernathanssc[at]gmail
Details: Please email if you're interested.

Lisbon, Portugal
Time: Sunday, October 6, 4:30 PM
Location: Oriente Station
Contact: ana_bcn[at]protonmail[dot]com; Meetup.com
Details: We meet in the Oriente station in front of the book fair. Then depending on the weather we can go to Parque das Nações or Vasco da Gama.

London, UK
Time: Sunday, September 15, 2:00 PM
Location: Craft Beer Co. Islington, 55 White Lion Street (we have a large room booked)
Contact: philip.hazelden[at]gmail; Facebook event; Facebook group

Los Angeles, CA
Time: Wednesday, October 9, 7:00 PM
Location: Colorado Center Park off 26th and Broadway
Contact: bobert.mushky[at]gmail; Discord server
Details: Also consider coming to the Orange County meetup to meet me [Scott] two days later

Madison, WI
Time: Saturday, September 21, 5:00 PM

Location: 1022 High St.
Contact: mmwang[at]wisc[dot]edu; Facebook event
Details: This is a house (newly painted blue with a red door), very close to the #4 and #47 bus stop. Also the #13 and #5 stop close by. If you need a ride, leave a message on the Facebook event and someone will get you. Some basic vegan food will be served. Bring food for yourself or to share if so inclined.

Madrid, Spain
Time: Saturday, September 21, 4:00 PM
Location: Ferros Café, Calle del Maestro Ángel Llorca, 9, 28003 Madrid
Contact: pavill01[at]ucm[dot]es; Meetup.com
Details: We speak both English and Spanish. About a week before I will post the event details in the Meetup group.

Manchester, UK
Time: Saturday, September 21, 2:00 PM
Location: The grass patch next to the Wharf Pub in city center M15 4ST, coordinates 53.473975, -2.256884
Contact: None

Medellín, Colombia
Time: Sunday, September 22, 11:00 AM
Location: Café Revolución, Carrera 73 | Cq. 4 #10, Laureles, Medellín, Colombia
Contact: richard.meadows[at]thedeepdish[dot]org

Melbourne, Australia
Time: Friday, October 4, 6:00 PM
Location: Queensberry Hotel, 593 Swanston St, Carlton VIC 3053
Contact: 0438 869 257; Facebook group

Memphis, TN
Time: Saturday, October 5, 5:00 PM
Location: French Truck Coffee at Crosstown Concourse
Contact: michael[at]postlibertarian[dot]com
Details: It’s in the main Central Atrium of Crosstown Concourse. There are many chairs and tables right in front so we should be able to grab a few. I’m totally ok if only a couple people show up!

Miami, FL
Time: Saturday, October 12, 5:00 PM
Location: Pasion del Cielo (Brickell City Centre location), 701 S Miami Ave #350B
Contact: eric135033[at]gmail; Facebook group

Minneapolis, MN
Additional search terms: St. Paul, Twin Cities
Time: Saturday, September 21, 3:00 PM
Location: The Knoll, 27 Pleasant St SE
Contact: grahamsnumberisbig[at]gmail; LessWrong
Details: This is a park on the north end of the University of Minnesota’s Minneapolis campus, not the apartment building of the same name. Although the forecast looks good right now, in the event of poor weather we may have to relocate – email me if you’re interested so I can get a rough headcount and send updates more easily. Also see the LessWrong post for an editable post that I’ll try to keep up to date if things change.

Montreal, Canada
Time: Sunday, September 22, 2:00 PM
Location: Loft 111A, 5455 rue d’Iberville
Contact: 1806992[at]gmail; Yohan 514-562-7682; Facebook event
Details: At the front door, dial 800. The meeting place will be at the collective l’Infini, which recently started hosting

LW meetups twice a month. Can comfortably host 30 people. Cozy and quiet. Note that the space welcoming the meetup is a sober space, so no drug/alcohol consumption inside. All are welcome to bring/cook food.

Moscow, Russia
Time: Saturday, September 21, 6:00 PM
Location: Rationality Center Kocherga, 5k2, B.Dorogomilovskaya st., Kievskaya, Moscow, Russia
Contact: oknezrob[at]kocherga-club[dot]ru

Munich, Germany
Time: Thursday, September 19, 7:00 PM
Location: Augustiner-Keller, Arnulfstraße 52, 80335 München, Germany
Contact: rlm72[at]cam[dot]ac[dot]uk

Nashville, TN
Time: Tuesday, September 17, 5:00 PM
Location: Frothy Monkey, 12th Avenue South
Contact: james[at]writechem[dot]com
Details: We will be discussing Meditations On Moloch. Any new members who wish to attend, please reach out to the contact given.

New Delhi, India
Time: Saturday, September 14, 5:00 PM
Location: Brown Box Cafe, Shop No. 2, DDA Market, S Block, Panchsheel Park, New Delhi, Delhi 110017
Contact: anujdayal[at]gmail

New Haven, CT
Additional search terms: Yale
Time: Saturday, September 21, 4:00 PM
Location: If the weather is good: the courtyard outside of P&M Market (on the corner of Orange Street + Cottage

Street); If it’s raining: Blue State Coffee on Orange Street.
Contact: rip.my.inbox[at]gmail

New York City, NY
Time: Saturday, September 21, 4:00 PM
Location: Pumphouse Park
Contact: tristan[at]thume[dot]ca; 646-480-0618
Details: If it’s raining or we’re otherwise not there go to the main atrium of Brookfield Place, which is adjacent to the park

Norman, OK
Time: Saturday, September 28, 2:00 PM
Location: University of Oklahoma, Bizzell Memorial Library, Room 421
Details: I’m not posting any contact info here, so here are my detailed contingency plans: if nobody shows up to the meetup within one hour, I will leave. If room 421 is for some reason unavailable we will meet in the reading area inside the front (south) doors, north of the front desk. I will fashion a small SSC sign out of a piece of paper or something.

Northampton, MA
Time: Saturday, September 21, 6:30 PM
Location: Packard’s Bar and Restaurant, main dining room
Contact: alex[at]alexliebowitz[dot]com; LessWrong

Oklahoma City, OK
Time: Saturday, September 21, 2:00 PM
Location: Del City Library
Contact: battleshipbean[at]gmail
Details: I’ll be wearing a USS Iowa shirt and hat.

Orange County, CA
Additional search terms: Irvine, Fullerton, Santa Ana, Anaheim, Huntington Beach

Time: Friday, October 11, 7:30 PM
Location: Underneath this mysterious hexagonal sigil at the University Center food court in Irvine, California.

Oslo, Norway
Time: Monday, September 23, 6:00 PM
Location: Starbucks at Aker Brygge, Stranden 7, 0250 Oslo, Norway
Contact: anders[at]huitfeldt[dot]net

Ottawa, Canada
Time: Friday, September 20, 7:00 PM
Location: The Fox & Feather, 283 Elgin St.
Contact: walsh.the[at]gmail; Facebook group
Details: We will book one of the private rooms.

Oxford, UK
Time: Thursday, September 19, 6:00 PM
Location: The Gardener's Arms, 39 Plantation Road, Oxford, OX2 6JE
Contact: ???

Paris, France
Time: Sunday, September 15, 3:00 PM
Location: 48°51'37.9″N 2°17'11.8″E (https://goo.gl/maps/nBZzEk3uggA2)
Contact: 336.vingt-trois.dix-huit.82.99; Discord server
Details: I hope the weather is going to be nice. Last year it started to be a bit cold after some time, so take a sweatshirt just in case. If the forecast is bad, we'll change the place to some coffee shop, maybe not in the same area, so check the page the day before.

Philadelphia, PA
Time: Sunday, September 22, 6:30 PM
Location: Tir Na Nog Bar & Grill, 1600 Arch Street

Contact: wfenza[at]gmail; dmcbriggs[at]gmail
Details: The restaurant is inside the Phoenix Hotel, directly above Suburban Station. We have the entire outer vestibule area reserved

Phoenix, AZ
Time: Saturday, September 21, 2:00 PM
Location: Lola Coffee Downtown, 1001 N 3rd Ave.
Contact: benjamin.j.morin[at]gmail
Details: Parking is typically available on weekends, this is also ~5 min walk from the light rail. Contact me if you are coming and I can give you a phone number to help with parking on the day of.

Pittsburgh, PA
Time: Saturday, October 19th, 6:00 PM
Location: 804 Anaheim St
Contact: ssc[at}fr8train[dot]me, mirandagavrin[at]gmail[dot]com

Portland, OR
Time: Tuesday, October 1, 7:00 PM
Location: Cartopia, 1207 SE Hawthorne Blvd
Contact: nwalton125[at]gmail; LessWrong
Details: Cartiopia is a food cart pod with many options — outside, but covered from rain. I’ll get there at 7 PM (or a little earlier), try to stake out table space, and put up a sign for the meetup. The carts are open as late as midnight.

Prague, Czechia
Time: Wednesday, September 25, 6:30 PM
Location: Rationalist coworking space Epistea (walk through tea room dharmasala to get there), Peckova 15, Karlínské náměstí
Contact: nadvornik.jiri[at]gmail; +420 607 690 860;

Facebook event
Details: Food will be provided. English friendly.

Pune, India
Time: Sunday, September 22, 6:00 PM
Location: Cafe Coffee Day, IISER Pune Campus, Baner
Contact: 7rat13[at]gmail

Reno, NV
Time: Saturday, September 28, 2:00 PM
Location: Reno Coffee Co, 1300 S Wells
Contact: supervexi[at]gmail
Details: Not sure how much of an SSC following we have here, so please be sure to indicate your interest by emailing me

Research Triangle, NC
Additional search terms: Raleigh, Durham, Chapel Hill
Time: Wednesday, September 18, 7:00 PM
Location: Ponysaurus Brewing Company, 219 Hood St, Durham, NC 27701
Contact: willdjarvis[at]gmail

Riga, Latvia
Time: Saturday, September 28, 1:00 PM
Location: Caffeine, Audēju iela 15
Contact: cerulean.lemniscate[at]protonmail[dot]com

Rio de Janeiro, Brazil
Time: Saturday, September 21, 2:00 PM
Location: Starbucks, Av. Rio Branco, 143 – A – Centro, Rio de Janeiro – RJ, 20040-006, Brazil
Contact: None
Details: Meet in the upstairs seating area

Rochester, NY
Time: Saturday, September 21, 2:00 PM

Location: SPoT Coffee, 200 East Avenue
Contact: None

San Antonio, TX
Time: Saturday, September 21, 2:00 PM
Location: Local Coffee, 7338 Louis Pasteur Dr #204
Contact: None

Singapore
Time: Saturday, September 21, 2:00 PM
Location: Starbucks outside Clarke Quay Station, 6 Eu Tong Sen St, #01-29 The Central, Singapore 059817
Contact: None

St. Louis, MO
Additional search terms: Saint Louis
Time: Saturday, September 21, 3:00 PM
Location: Hartford Coffee, 3974 Hartford St
Contact: littlejohnburidan[at]gmail
Details: I have a place to stay if you visit from out of town and need a bed.

Sacramento, CA
Time: Sunday, September 22, 11:00 AM
Location: Crepeville, 1730 L St.
Contact: amethyst.eggplant[at]gmail

Salt Lake City, UT
Time: Saturday, October 5, 3:00 PM
Location: 41 S 900 E #204
Contact: wearenotsaved[at]gmail

San Diego, CA
Time: Sunday, September 29, 3:00 PM
Location: La Jolla Colony Park
Contact: owenshen24[at]gmail; Facebook group

Details: Also consider coming to meet me [Scott] at the Orange County meetup October 11th

San Francisco, CA

Time: Monday, September 23, 6:30 PM
Location: 170 Hawthorne St.
Contact: rocurley[at]gmail; LessWrong
Details: For help getting into the building, please call (or text, with a likely-somewhat-slower response rate): 301-458-0764. We'll be meeting to give and listen to very short talks! We'll do 7-minute lightning talks with 3 additional minutes allowed for questions. A talk doesn't have to be formal, planned, or even something that you'd expect someone to Give A Talk About; it can be as simple as telling the group about something you find interesting or cool.

San Jose, CA

Additional search terms: South Bay
Time: Saturday, September 28, 2:00 PM
Location: 3806 Williams Rd.
Contact: ; Website

San José, Costa Rica

Time: Saturday, September 21, 2:00 PM
Location: Starbucks, Costado Oeste del Estadio Nacional, en edificio Clínica 20 20 San José Sabana Oeste, Nunciatura, San José, Costa Rica
Contact: None

São José dos Campos, Brazil

Time: Sunday, September 15, 2:00 PM
Location: Parque Santos Dumont, next to the 14-Bis airplane replica
Contact: emiliobumachar[at]gmail
Details: I will meet even one person in good faith, but will not go if no one contacts me.

Seattle, WA
Time: Wednesday, October 2, 6:30 PM
Location: The Territory (house), 5238 11th Ave NE
Contact: orborde[at]gmail; Facebook event

Seoul, South Korea
Time: Saturday, September 28, 12:30 PM
Location: O'sulloc Green Tea near Gangnam Subway Station
Contact: curtis95112[at]gmail
Details: I would also appreciate it if people who can't/won't make the meetup contacted me too, possibly anonymously, because it would be fascinating if nobody replied and I were literally the only SSC reader in a city of 20 million.

Stockholm, Sweden
Time: Saturday, September 14, 3:00 PM
Location: Grillska Huset, Stortorget 3, Stockholm
Contact: Meetup.com

Sydney, Australia
Time: Thursday, September 19, 6:00 PM
Location: Level 2, City of Sydney RSL, 565 George St
Contact: Meetup.com
Details: We sit at a table in the fish bowl across the room from the lifts.

Tampa, FL
Time: Saturday, September 21, 5:00 PM
Location: Felicitous Coffee, 11706 N 51st St, Temple Terrace
Contact: danielbporter[at]gmail

Tel Aviv, Israel
Time: Tuesday, September 24, 7:00 PM
Location: Google Israel office, Electra Tower, 12th floor, Yigal Alon 98
Contact: highpriestessofelua[at]gmail; Facebook event

Toronto, Canada
Time: Monday, September 23, 7:00 PM
Location: Madison Avenue Pub
Contact: team[at]thinkbetter[dot]ca; Facebook event

Toulouse, France
Time: Monday, October 7, 7:00 PM
Location: Eurêkafé, 24 Rue Léon Gambetta
Contact: me[at]jrmyp[dot]net

Vancouver, Canada
Time: Sunday, September 29, 1:00 PM
Location: Picnic area at Trout Lake Park, near the south beach

Victoria, BC, Canada
Time: Saturday, September 21, 2:00 PM
Location: Starbucks, 1301 Government St, Victoria, BC V8W 1Y9, Canada (downtown, at Yates and Government)
Contact: graeme.andrew.hill[at]gmail
Details: I'll be the lanky white dude with a buzzcut wearing all blue.

Vienna, Austria
Time: Saturday, October 19, 3:00 PM
Location: Kaisermühlenstraße 24, A-1220 Vienna
Contact: strivingforconsistency[at]gmail; Facebook group
Details: We'll be in the meetup room on the ground floor at the back side of the building. Event and directons to the location will be announced on the Facebook group.

****Warsaw, Poland****
Time: Saturday, September 21, 12:00 PM
Location: Starbucks, plac Trzech Krzyży 16A, 00-499 Warszawa, Poland
Contact: None

Washington DC
Time: Tuesday, September 24, 5:00 PM
Location: Teaism coffee shop, 400 8th Street NW, Washington DC
Contact: robirahman94[at]gmail; Google group

Wellington, New Zealand
Time: Saturday, September 28, 3:00 PM
Location: Glover Park
Contact: benwve[at]gmail

West Lafayette, IN
Additional search terms: Purdue
Time: Sunday, October 6, 6:00 PM
Location: Cafe Literato, 1920 Northwestern Ave.
Contact: lancesbush[at]gmail; Facebook group; Facebook account

Zurich, Switzerland
Time: Saturday, October 5, 3:00 PM
Location: The Sacred/Vegelateria at Müllerstrasse 64.
Contact: ssczurich[at]gmx[dot]ch; Facebook group

Against Against Pseudoaddiction

I.

"Pseudoaddiction" is one of the standard beats every article on the opioid crisis has to hit. Pharma companies (the story goes) invented a concept called "pseudoaddiction", which looks exactly like addiction, except it means you just need to give the patient more drugs. Bizarrely gullible doctors went along with this and increased prescriptions for their addicted patients. For example, from a letter in the Wall Street Journal:

> Parroting Big Pharma's excuses about FDA oversight and black-box warnings only discounts how companies like Johnson & Johnson engaged in pervasive misinformation campaigns and even promoted a theory of "pseudoaddiction" to encourage doctors to prescribe even more opioids for patients who displayed signs of addiction.

Or from CBS:

> But amid skyrocketing addiction rates and overdoses related to OxyContin, Panara claimed the company taught a sales tactic she now considers questionable, saying some patients might only appear to be addicted when in fact they're just in pain. In training, she was taught a term for this: "pseudoaddiction."
> "So the cure for 'pseudoaddiction,' you were trained, is more opioids?" Dokoupil asked.
>
> "A higher dose, yes," Panara said.

> “Did this concept of pseudoaddiction come with studies backing it up?”
>
> “We had no studies. We actually — we did not have any studies. That’s the thing that was kind of disturbing, was that we didn’t have studies to present to the doctors,” Panara responded.
>
> “You know how that sounds?” Dokoupil asked.
>
> “I know. I was naïve,” Panara said.

Pseudoaddiction is among the few medical concepts that’s made it far enough to get denounced by US senators. From Senator Maggie Hassan’s website:

> Senator Hassan then asked Jennifer Taubert, Executive Vice President of Janssen Pharmaceuticals, about the company’s promotion of the unproven and dubious concept of “pseudoaddiction,” an idea advanced by the pharmaceutical industry claiming that when certain patients present signs of addiction it is because they were prescribed insufficient doses of opioids, and that instead of providing addiction treatment, doctors should increase their opioid doses. Ms. Taubert claimed to be unware of the term.
>
> “Janssen promoted this made-up concept of pseudoaddiction on a website it approved and funded was called ‘Let’s Talk Pain’” Senator Hassan said. “Since then, your company has repeatedly said that your actions quote ‘in the marketing and promotion of our opioid pain medicines was appropriate and responsible.’ So Ms. Taubert, how can you possibly claim that promoting the theory of pseudoaddiction – that doctors should prescribe more opioids to patients showing signs of addiction – was appropriate and responsible?”

Let me confess: I think pseudoaddiction is real. In fact, I think it’s *obviously* real. I think everyone should realize it’s real as soon as it’s explained properly to them. I think we should be terrified that any of our institutions – media, academia, whatever – think they could possibly get away with claiming pseudoaddiction isn’t real. I think people should be taking to the streets trying to overthrow a medical system that has the slightest doubt about whether pseudoaddiction is real. If you can think of more hyperbolic statements about pseudoaddiction, I probably believe those too.

Neuroscientists define addiction in terms of complicated brain changes, but ordinary doctors just go off behavior. The average doctor treats “addiction” and “drug-seeking behavior” as synonymous. This paper lists signs of drug-seeking behavior that doctors should watch out for, like:

– Aggressively complaining about a need for a drug
– Requesting to have the dose increased
– Asking for specific drugs by name
– Taking a few extra, unauthorised doses on occasion
– Frequently calling the clinic
– Unwilling to consider other drugs or non-drug treatments
– Frequent unauthorised dose escalations after being told that it is inappropriate
– Consistently disruptive behaviour when arriving at the clinic

You might notice that all of these are things people might do *if they actually need the drug*. Consider this classic case study of pseudoaddiction from Weissman & Haddox, summarized by Greene & Chambers:

> The 1989 introduction of pseudoaddiction happened in the form a single case report of a 17-year-old man with

> acute leukemia, who was hospitalized with pneumonia and chest wall pain. The patient was initially given 5 mg of intravenous morphine every 4 to 6 h on an as-needed dosing schedule but received additional doses and analgesics over time. After a few days, the patient started engaging in behaviors that are frequently associated with opioid addiction, such as requesting medication prior to scheduled dosing, requesting specific opioids, and engaging in pain behaviors (e.g., moaning, crying, grimacing, and complaining about various aches and pains) to elicit drug delivery. The authors argued that this was not idiopathic opioid addiction but pseudoaddiction, which resulted from medical under-treatment (insufficient opioid dosing, utilization of opioids with inadequate potency, excessive dosing intervals) of the patient's pain. In describing pseudoaddiction as an "iatrogenic" syndrome, Weissman and Haddox inverted the traditional usage of iatrogenic as harm caused by a medical intervention. In pseudoaddiction, iatrogenic harm was described as being caused by withholding treatment (opioids), not by providing it.

Greene & Chambers present this as some kind of exotic novel hypothesis, but *think about this for a second like a normal human being*. You have a kid with a very painful form of cancer. His doctor guesses at what the right dose of painkillers should be. After getting this dose of painkillers, the kid continues to "engage in pain behaviors ie moaning, crying, grimacing, and complaining about various aches and pains", and begs for a higher dose of painkillers.

I maintain that the normal human thought process is "Since this kid is screaming in pain, looks like I guessed wrong about the right amount of painkillers for him, I should give him more."

The official medical-system approved thought process, which Greene & Chambers are defending in this paper, is "Since he is displaying signs of drug-seeking behavior, he must be an addict trying to con you into giving him his next fix." They never come out and say this. But they define pseudoaddiction as meaning *not that*, and end up saying "in conclusion, we find no empirical evidence yet exists to justify a clinical 'diagnosis' of pseudoaddiction." More on this later.

The concept of "pseudoaddiction" was invented as a corrective to an all-too-common tendency for doctors to assume that anyone who seems too interested in getting more medications is necessarily an addict. It was invented not by pharma companies, but by doctors working with patients in pain, building upon a hundred-year-long history of other doctors and medical educators trying to explain the same point.

And in case you think this is a weird ivory tower debate that doesn't influence real clinical practice, I offer you these cases from my own experience. Stories slightly changed or merged together to protect patient privacy:

Case 1: Mary is an elderly woman who undergoes a surgery known to have a painful recovery process. The surgeon prescribes a dose of painkillers once every six hours. The painkillers last four hours. From hours 4-6, Mary is in terrible pain. During one of these periods, she says that she wishes she was dead. The surgeon leaps into action by…calling the on-call psychiatrist and saying "Hey, there's a suicidal person on my ward, you should do psychiatry to her or something." I am the on call psychiatrist. After a brief evaluation, I tell the surgeon that Mary has no psychiatric illness but needs painkillers every four hours. The surgeon lectures me on how There Is An Opioid Crisis, Y'Know, and we can’t negotiate with addicts and drug-seekers. I am a consultant on the case and can’t overule the surgeon on his own ward, so I just hang out with Mary for a while and talk about things and distract her and listen to her scream during the worst part of the six-hour cycle. After a few days the surgery has healed to the point where Mary is only in excruciating pain rather than actively suicidal, and so we send her home.

Case 2: Juan is a middle-aged man with depression who is using Geodon for antidepressant augmentation. This is kind of a weird choice, and has theoretical potential to interact poorly with some of his other medications, but nothing else has worked for him and he’s done great for ten years. He switches psychiatrists. The new psychiatrist is really worried about the theoretical interaction, so he tells him that he can’t take Geodon anymore and switches him to something else. Juan falls into a deep depression. He asks to have Geodon back and the doctor says no. Juan yells at the psychiatrist and says he is ruining his life. The psychiatrist diagnoses him with a personality disorder and anger management problems, and tells him to attend therapy. Juan actually does this for a while, but eventually wises up and switches doctors to me. I put him back on Geodon and within a month he’s doing great again. Note that Juan displayed every sign of “drug-seeking behavior” even though Geodon is not addictive.

Case 3: This one courtesy of Zvi. Zvi’s friend is diabetic. He runs out of insulin and asks his doctor for more. The doctor wants to wait until his next free appointment in a few weeks before prescribing the insulin. Zvi’s friend points out that he will die unless he gets more insulin now. The doctor gets very angry about this and spends a long phone call haranguing Zvi’s friend about how inconvenient it is that he’s demanding the insulin now rather than at a more convenient time. Zvi’s friend has to threaten the doctor with a lawsuit before the doctor finally relents and gives him the insulin. I like this story because, again, insulin is not addictive, there is no way that the patient could possibly be doing anything wrong, but the patient *still gets treated as a drug-seeker*. The very act of wanting medication according to the logic of his own disease, rather than at the doctor’s convenience, is enough to make his request suspicious.

Case 4: John is a 70 year old man on opioids for 30 years due to a mining-related injury. He is doing very well. I am his outpatient psychiatrist but I only see him once every few months to renew meds. He gets some kind of infection, goes to the hospital, and due to normal hospital incompetence he doesn’t get his opioids. He demands his meds, and like many 70 year old ex-miners in terrible pain, he is not diligently polite the whole time. The hospital doctors are excited: they have caught an opioid addict! They tell his family and outpatient doctors he cannot have opioids from now on, then discharge him. He continues to be in terrible pain. At first he sneaks pills from an extra bottle of opioids he has at home, but eventually he uses all those up. After this, he is *still* in terrible pain with no reason to expect this to ever change, and so he shoots himself in the chest. This is the first point in this entire process at which anyone attempts to tell me any of this is going on, so I get a “HEY DID YOU KNOW YOUR PATIENT SHOT HIMSELF? DOESN’T SEEM LIKE YOU’RE DOING VERY GOOD PSYCHIATRIST-ING?” call. The patient miraculously survives, eventually finds a new pain doctor, and goes on to live a normal and happy life on the same dose of opioids he was using before.

Case 5: Evelyn is an elderly woman with dental pain. She goes to her dentist, who prescribes opioids. She is concerned – aren’t opioids addictive? “Don’t worry, you’ll be fine”, says the dentist. The dentist keeps her on them for eight months out of some kind of bizarre incompetence that is not her fault. Then that dentist retires and transfers her to another dentist in the same practice. Evelyn asks the new dentist to refill her opioid prescription, and he freaks out – why is this patient on opioids? He refuses to refill the prescription. She gets really scared because she is about to withdraw from opioids cold turkey, and asks for a week's worth of pills so she can taper down. The dentist calls her an addict and refuses. She asks for some kind of help, any kind of help, getting off the opioids, and the dentist tells her to go to a drug rehab so she can get treated by people who understand her addiction. Finally in desperation she calls the psychiatrist who is treating her for an unrelated problem (me), and I prescribe the standard opioid withdrawal regimen and talk her through the process. I would like to say this story has a happy ending, but she's currently in post-acute withdrawal syndrome, so @#$% everybody involved.

Case 6: Sandy is a middle-aged woman on benzodiazepines, a potentially addictive anti-anxiety medication. She has been stable for twenty years. She switches doctors. The new doctor has heard that Benzodiazepines Are Bad And Addictive, so he discontinues them over her objections. Sandy becomes a miserable wreck and has panic attacks basically all the time for a few months. Whenever she tries to mention this to the doctor, he accuses her of being an addict and trying to con him into giving her drugs. After a few months of this, she leaves that doctor and switches to me. I put her back on her previous dose of benzodiazepines, and within two days she feels perfectly normal and gets on with her life.

Case 7: Robert is a young man who is prescribed trazodone 50 mg nightly for sleep. This goes well for several years. Then he gets in a fight with his wife and they are considering divorce. He's really worried and angry and can't sleep, and so after going several nights without sleep and feeling completely miserable, one night at 4 AM he takes two trazodone – 100 mg – and gets to sleep right away. He mentions this to his doctor, who accuses him of “unauthorized dose escalation”, ie going up on your drug without telling your doctor. He refuses to prescribe trazodone further. Robert is now totally unable to sleep. He ends up with me, I put him back on trazodone, tell him that the maximum safe dose of trazodone is 400 mg but that if 50 mg works for him I want him to try to stick to that except in emergencies so that he doesn’t build tolerance, and he continues taking 50 mg on average nights and 100 mg once or twice a year if things are really bad.

Let’s look at those warning signs of addiction again:

– Aggressively complaining about a need for a drug
– Requesting to have the dose increased
– Asking for specific drugs by name
– Taking a few extra, unauthorised doses on occasion
– Frequently calling the clinic
– Unwilling to consider other drugs or non-drug treatments
– Frequent unauthorised dose escalations after being told that it is inappropriate
– Consistently disruptive behaviour when arriving at the clinic

In Case 1, Mary requested her dose of painkiller be increased (from once per six hours to once per four hours). In Case 2, Juan asked for a specific drug by name (Geodon), and was unwilling to consider other drugs. In Case 3, Zvi’s friend frequently called the clinic (to get them to refill his insulin). In Case 4, John showed consistently disruptive behavior in the hospital and took extra unauthorized doses. Etc.

All of these are drug-seeking behaviors. But I maintain that none of these patients were addicted. The correct action in all of these cases is to listen to the patient's reasons for wanting the drug, realize that you (the doctor) screwed up, and give them the drug that they are asking for. Although the point that these behaviors can be signs of addiction is well-taken and important, it's equally important to remember they can be signs of other things too.

Media portrayals of pseudoaddiction portray it as this bizarre contortion of logic: "A patient is displaying signs of addiction, so you should give them *more* of the drug! Haha, nice try, pharma companies!" But this is *exactly* what you should do! The real problem lies with anyone who conceptualizes pseudoaddiction as a novel hypothesis that requires proof, rather than as the obvious possibility you *have to* check for before accusing patients of addiction.

II.

At this point, any reasonable person will think I'm trying to bait-and-switch you. Surely the reasonable position I'm defending isn't the same as the dreaded "pseudoaddiction" that everyone knows is a pharma company swindle? Surely I must be straw-manning the pseudoaddiction opponents somehow?

I don't think that I am. I want to go over Greene & Chambers (2015), Pseudoaddiction: Fact Or Fiction, in *Current Addiction Reports*. This is the most important paper establishing the current consensus against pseudoaddiction as a concept. It's been cited in debates before Congress on the opioid crisis, featured on CBS, and is the primary source for the current consensus that pseudoaddiction has been "debunked". This is the best and most thorough anti-pseudoaddiction paper, and if there’s more to the story we’ll probably find it there.

G&C start with a review of the pseudoaddiction literature, beginning with the case study I quoted above. I’m going to quote it again, since I think it’s really important to establish that everyone agrees this is the kind of thing we’re talking about:

> The 1989 introduction of pseudoaddiction happened in the form a single case report of a 17-year-old man with acute leukemia, who was hospitalized with pneumonia and chest wall pain. The patient was initially given 5 mg of intravenous morphine every 4 to 6 h on an as-needed dosing schedule but received additional doses and analgesics over time. After a few days, the patient started engaging in behaviors that are frequently associated with opioid addiction, such as requesting medication prior to scheduled dosing, requesting specific opioids, and engaging in pain behaviors (e.g., moaning, crying, grimacing, and complaining about various aches and pains) to elicit drug delivery. The authors argued that this was not idiopathic opioid addiction but pseudoaddiction, which resulted from medical under-treatment (insufficient opioid dosing, utilization of opioids with inadequate potency, excessive dosing intervals) of the patient’s pain. In describing pseudoaddiction as an “iatrogenic” syndrome, Weissman and Haddox inverted the traditional usage of iatrogenic as harm caused by a medical intervention. In pseudoaddiction, iatrogenic harm was described as being caused by withholding treatment (opioids), not by providing it.

Instead of concluding that okay, Weissman and Haddox have a point and someone should get this kid some pain relief, they note that case reports are a low level of medical evidence, and nobody has ever done any big studies or meta-analyses that provide empirical proof of pseudoaddiction. They don’t explain what this would mean, or how you turn “stop torturing children due to a misplaced desire to nab addicts” into a p-value of less than 0.05. They just conclude that this means pseudoaddiction has never been empirically proven to exist, then discuss how some of the case reports of pseudoaddiction (though not Weissman and Haddox’s original) were sponsored by Big Pharma.

Then they get more philosophical, arguing that pain can never be objectively proven to exist. Also, even if it were objectively proven that someone was in pain, that person could still be an addict, since addicts can feel pain too. Therefore, we can never prove that there is a person who is in pain but not an addict, and therefore we cannot empirically prove the existence of pseudoaddiction. Some quotes so you can judge whether I’m being unfair:

> The existence of pseudoaddiction, and its distinction from true addiction, is understood by proponents as being based on the patient’s reported motivation for pain relief (e.g., if their behavior results from pain, then they have pseudoaddiction, not addiction). The reliability of this conceptualization seems to hinge on the assumption that addiction and pain do not co-occur (unless one can comprehend the possibility that a patient can have fake addiction and true addiction at the same time!). However, it is not the case that pain and addiction are mutually exclusive conditions, and no clear evidence exists that having pain protects against the genesis or expression of addiction.
>
> A primary difficulty in measuring pain is its highly subjective nature that is influenced by many cultural,

situational, and individual neuropsychological factors [62–64]. Given the large degree to which pseudoaddiction does not distinguish itself from addiction, except based on subjective reporting of pain, and the extent to which opioid addiction is associated with or may even cause subjective pain, it is unclear how the application of pseudoaddiction has further enhanced the clinical assessment and management of pain.

In conclusion, we find no empirical evidence yet exists to justify a clinical “diagnosis” of pseudoaddiction. The renaming of pain with a term that essentially means “fake addiction” and serves to dismiss addiction as part of the clinical differential diagnosis is a construct that is conspicuously and uniquely attached to opioid therapies which are extremely addictive analgesics, among many other effective, evidence-based strategies for analgesia that are far less addictive. If pseudoaddiction is to remain an influential clinical construct that is taught in medical schools and textbooks, its usage and clinical acceptance need empirical support, with evidence-based disambiguation from addiction, and delineation of its treatment implications. However, to the extent that a diagnosis of pseudoaddiction relies on a self-report of pain (that is still essentially not objectively measurable) as the motivation for drug-seeking, it is not clear how rigorously it can ever be proven or disproven in human research.

It is hard to conclude from this review and the context of the current prescription opioid epidemic that pseudoaddiction is an objective, evidence-based diagnosis that has been clinically beneficial to patient lives. Instead, it may be most beneficial to retire the term and understand patients as simply having pain,

> opioid addiction, or very often both, and designing treatment strategies that best account for and balance the competing risk-benefit treatment concerns that these brain conditions imply.

Some of these quotes seem to suggest that it’s hard to define the border between addiction and pseudoaddiction. They don’t really make this argument clearly, but it could go something like – if addiction is an attempt to feel better than well, and pseudoaddiction is an attempt to feel better than some miserable baseline, and there’s no clear bright line between “so miserable you deserve drugs” and “so well that you don’t”, how can we wall of “pseudoaddiction” as a separate concept? I agree this is difficult, but it’s the same kind of difficulty you get in having any concept at all, so you should probably deal with it.

Others seem to kind of equivocate between “pseudoaddiction is fake” vs. “[the phenomenon described by the word pseudoaddiction] is real, but there’s no point in having a separate word for it.” The latter would be reasonable *if there weren’t so many people saying the former*. Because people are constantly misdiagnosing real distress as addiction, we need a word for when that happens, and pseudoaddiction is as good as any other.

Others seem to argue that pseudoaddiction doesn’t rise to the level of a medical diagnosis, with the faux-objectivity that implies. But nobody was previously trying to turn it into one – if you look at the paper that coined the term pseudoaddiction, it’s very clear that it’s just making a new word for a common-sense concept that everyone has known about for a long time.

Also, one of the G&C authors, Chambers, goes on to write a nearly identical paper which also attacks the concept of “self-medication”, Have Pseudoaddiction And Self-Medication Led Us Astray? It argues:

> ‘Self-medication’, a concept originating when psychiatrists noticed frequent tobacco, alcohol, and other drug use in the deinstitutionalized mentally ill, has been around for many decades, but was also formally elaborated on in the 1980’s. It has subsequently been endorsed and widely embraced in primary research and review papers and educational sources spanning the field of psychiatry, as the standard explanation for why persons with mental illnesses use substances. ‘Self-medication’ has become so widely and dogmatically accepted as the key explanation for substance use in mental illness, that it has become nearly synonymous with ‘dual diagnoses’. In both ‘pseudoaddiction’ and ‘self-medication’, drug use is explained as a choice to seek and use drugs for benefit—to gain symptom relief from pain or psychiatric symptoms. Whereas in addiction, the behavior is explained as compulsive, not a voluntary choice, that persists despite negative consequences, not because of benefits. As suggested in Table 1, the construct similarities between ‘pseudoaddiction’ and ‘self-medication’ are quite comprehensive, including how they consistently contradict the disease model of addiction.

Nobody is claiming that self-medication is an official medical diagnosis, so I conclude that is not the meat of G&C’s objection.

If we want to be super-charitable to them, we can focus on a paragraph from the self-medication paper which is more self-aware than anything I see in the pseudoaddiction paper:

> Of course, untreated pain does exist. People do self-medicate (e.g. taking an antibiotic for pneumonia). And sometimes, taking addictive drugs (usually short term) can be very therapeutic. But it may be time to ask: Has the medical community and psychiatry in particular grown over-accustomed — even ‘addicted’ to overusing, academically endorsing, and clinically propagating, the proxy diagnoses of ‘pseudoaddiction’ and ‘self-medication’ to avoid dealing with addiction itself? If so, what forces have contributed to this phenomenon? Do doctors believe these constructs help them avoid heaping the criminalizing stigma of ‘addiction’ onto their patients? Do these constructs excuse doctors from dealing with addiction, when so many of us, and most detrimentally, psychiatrists, don’t know how to treat it, or can’t get paid for doing so, or, are so often accustomed to prescribing addictive drugs for a wide variety of indications? Have there been too many incentives, and too many effective marketing campaigns from corporate interests that manufacture and sell addictive drugs like nicotine, opioids, benzodiazepines and stimulants, that have over-inflated their medicinal attributes to doctors and the public, while minimizing their addictive downsides?

Here Chambers seems to be saying that maybe “pseudoaddiction” and “self-medication” describe real things, but that the pendulum has swung so far towards treating drug use as legitimate, and so far from being willing to call people “addicts”, that we need to excise ideas like pseudoaddiction and self-medication from the lexicon so that doctors will have no choice but to recognize addiction in their patients when they see it.

I disagree with this, but it’s at least a coherent position. But if you want to have this argument, say “pseudoaddiction is rarer than people think”. Don’t say “pseudoaddiction doesn’t exist”. If you say it doesn’t exist, then our first argument has to be over whether it exists, and as far as I can tell it obviously does.

I worry that G&C are vacillating among a bunch of different claims, making their argument hard to address. Sometimes they argue that no double-blind empirical study has proven pseudoaddiction. I think this is a category error, like wanting a double-blind empirical study to prove the existence of ennui. Sometimes they argue that pseudoaddiction cannot be proven to exist. I think this is true only in the very philosophical sense where pain cannot be proven to exist, and once we start using common sense, it clearly exists. Other times they argue that there’s no clear bright line between addiction and pseudoaddiction. I agree, but think there is no bright line between any concept and any other concept, so we better get used to this and not stop prescribing clinically indicated drugs on that basis. Other times they argue that pseudoaddiction should not be a reified diagnosis-like concept. I don’t think it is supposed to be, so they are attacking a straw man. Other times they argue that doctors are too likely to coddle addicts. I think this is a potentially fruitful thing to argue about, but they need to start this argument by saying the thing they actually believe, not an unrelated claim that “pseudoaddiction doesn’t exist”. Overall the argument seems muddled, and unworthy of the consensus in favor of its claims that it has produced.

III.

If pseudoaddiction is such a common-sense idea, how did we reach this point where people are deriding medicine for ever having believed in it?

As far as I can tell, the concept started off well-intentioned. But painkiller companies realized that the debate over when to diagnose addiction vs. pseudoaddiction was relevant to their bottom line, and started funding the pseudoaddiction side of it.

I’m not sure how substantial an effort this was. G&C note that of 224 papers mentioning pseudoaddiction, 22 were sponsored by pharma (but that means 202 weren’t). Of a stricter category of 12 papers that focused on arguing for the concept, 4 were sponsored by pharma (but 8 were not). Taking their numbers at face value, the majority of discussion of pseudoaddiction had no pharma company sponsorship. But the image of an expert getting up in front of a medical conference and telling doctors that the solution to opioid addiction was *more opioids* – something that certainly did happen, I’m not sure how often – was so lurid that it burned itself into the popular consciousness. The media exaggerated this from “basically good idea gets misused” to “doctors invent vicious lies to addict your loved ones” to get more clicks. Experts didn’t want to be the guy saying “well actually” in the middle of an Opioid Crisis, so they kept their mouths shut. Reporters copied each others’ denunciations of ‘pseudoaddiction’ without checking what the term really meant.

Into all this came the drug warriors. It’s hard for me to be angry at addictionologists, because they have a terrible job and are probably traumatized by it. But they really hate drugs and will say whatever it takes to make you hate drugs too. These are the people who gave us articles on how one hit of marijuana will get you addicted forever and definitely kill you, how one hit of LSD will make you go crazy and get addicted and probably kill you, how there can never be any legitimate medical reason for using cannabis, how e-cigarettes are deadly poison, and other similar classics.

Sensing that they had the high ground, they wrote a couple of papers about how pseudoaddiction isn't "empirically proven", as if this were a meaningful claim. This gave the media the ammunition they needed to declare that pseudoaddiction was always pseudoscience and has now been debunked and well-refuted.

This is just my story, and it's kind of bulverist. But if you think it's plausible, I recommend the following lessons:

First, when the media decides to craft a narrative, and the government decides to hold a moral panic, arguments get treated as soldiers. Anything that might sound like it supports the "wrong" side will be mercilessly debunked, no matter how true it is. Anything that supports the "right" side will be celebrated and accepted as obvious, no matter how bad its arguments. Good scientists feel afraid to speak up and question the story, lest they be seen as "soft on the Opioid Crisis" or "stooges of Big Pharma". This happens again and again on any issue people care about, and I want to reiterate for the nth time that you should treat reporting on medical, scientific, and social scientific topics as having almost zero credibility.

Second, you should stay cautious about bias arguments. Yes, some people pushed pseudoaddiction because they were shills of the opioid companies. But other people pushed pseudoaddiction because it was true. Just because you can generate the hypothesis "maybe people are just shills of the opioid companies" doesn't mean you've disproven pseudoaddiction. And if you focus too hard on the opioid companies' obvious financial bias, then you'll miss less obvious but possibly more important biases like those of the drug warriors. Your best bet would have been to just stop worrying about biases and try to figure out what was actually true.

The opioid crisis is really bad. I nevertheless think pseudoaddiction is the most obviously true medical concept this side of Hippocrates. The denial of its existence is a failure of national epistemics that deserves more scrutiny than it’s getting.

Too Much Dark Money In Almonds

Everyone always talks about how much money there is in politics. This is the wrong framing. The right framing is Ansolabehere et al's: why is there so *little* money in politics? But Ansolabehere focuses on elections, and the mystery is wider than that.

Sure, during the 2018 election, candidates, parties, PACs, and outsiders combined spent about $5 billion – $2.5 billion on Democrats, $2 billion on Republicans, and $0.5 billion on third parties. And although that sounds like a lot of money to you or me, on the national scale, it's puny. The US almond industry earns $12 billion per year. Americans spent about 2.5x as much on almonds as on candidates last year.

But also, what about lobbying? Open Secrets reports $3.5 billion in lobbying spending in 2018. Again, sounds like a lot. But when we add $3.5 billion in lobbying to the $5 billion in election spending, we only get $8.5 billion – still less than almonds.

What about think tanks? Based on numbers discussed in this post, I estimate that the budget for all US think tanks, liberal and conservative combined, is probably around $500 million per year. Again, an amount of money that I wish I had. But add it to the total, and we're only at $9 billion. Still less than almonds!

What about political activist organizations? The National Rifle Association, the two-ton gorilla of advocacy groups, has a yearly budget of $400 million. The ACLU is a little smaller, at $234 million. AIPAC is $80 million. The NAACP is $24 million. None of them are anywhere close to the first-person shooter video game "Overwatch", which made $1 billion last year. And when we add them all to the total, we’re *still* less than almonds.

Add up all US spending on candidates, PACs, lobbying, think tanks, and advocacy organizations – liberal and conservative combined – and we’re still $2 billion short of what we spend on almonds each year. In fact, we’re still less than Elon Musk’s personal fortune; Musk could personally fund the entire US political ecosystem on both sides for a whole two-year election cycle.

But let’s go further.

According to this article, Mic.com sold for less than $5 million. Mashable sold for less than $50 million. The whole Gawker network (plus some other stuff including the Onion) sold for $50 million. There are some hints that Vox is worth a high-eight-digit to low-nine-digit amount of money. The Washington Post was sold for $250 million in 2013 (though it’s probably worth more now). These properties seem to be priced entirely as cash cows – based on their ability to make money through subscriptions or ads. The extra value of using them for political influence seems to be priced around zero, and this price seems to be correct based on how little money is spent on political causes.

Or: *Jacobin* spends a lot of time advocating socialism. *The Economist* spends a lot of time advocating liberalism. *First Things* spends a lot of time advocating conservatism. They all have one thing in common: paywalls. How could this be efficient? There are millions of people who follow all of these philosophies and really want to spread them. And there are other people who have dedicated their lives to producing great stories and essays advocating and explaining these philosophies – but people have to pay $29.99 for a subscription to read their work? Why do ideologies make people pay to read their propaganda?

Maybe the most extreme example here is Tumblr.com, which recently sold for $3 million, ie the cost of a medium-sized house in San Francisco. Tumblr has 400 million monthly visitors, and at least tens of millions of active users. These people talk politics all the time, usually of a far-left variety. Nobody thinks that one of the central political discussion platforms of the far-left is worth more than $3 million? Nobody on the right wants to shut it down? Nobody on the left wants to prevent that from happening? Nobody with a weird idiosyncratic agenda thinks being able to promote, censor, or advertise different topics on a site with tens of millions of politically engaged people is at all interesting?

(in case you're keeping track: all donations to all candidates, all lobbying, all think tanks, all advocacy organizations, the Washington Post, Vox, Mic, Mashable, Gawker, and Tumblr, combined, are still worth a little bit less than the almond industry. And Musk could buy them all.)

The low level of money in politics should be really surprising for three reasons.

First, we should expect ordinary people to donate more to politics. A lot of the ordinary people I know care a *lot* about politics. In many of the events they care about most, like the presidential primaries, small donations matter a lot – just witness Tom Steyer begging for small donations despite being a billionaire. If every American donated $25 to some candidate they supported, election spending would surpass the almond industry. But this isn't even close to happening. Bernie Sanders is rightly famous for getting unusually many small donations from ordinary people. It's not clear exactly how much he's received, but it looks like about $50 million total. This sounds like a lot of money, but if you use polls to estimate how many supporters he has, it looks like each supporter has on average given him $2. This is a nice token gesture, but surely less than these people's yearly almond budget.

Second, we should expect the rich to donate more to politics. Many politicians want to tax billionaires; billionaires presumably want to prevent that from happening. Or wealthy people might just have honestly-held political opinions of their own. As rich as Elon Musk is, he's only one of five hundred billionaires, and some of the others are even richer. So how come the amount of money in politics is so much less than many individual billionaires' personal fortunes?

Third, we should expect big corporations to donate more to politics. Post *Citizens United*, corporations can supposedly put as much money into politics as they want. And they should want a lot. The government regulates corporations, so having friendly politicians in power can mean life or death for entire industries. Suppose hostile government regulation could decrease Exxon Mobil's revenues 5% – you would think Exxon Mobil would be willing to spend 4% of its revenue to prevent this. But Exxon makes $280 billion per year. 4% of its revenue would already be larger than the whole US political ecosystem! In fact, according to Exxon's own records, they only spend about $1 million per cycle. While they're probably hiding something, they couldn't hide donations the size of the whole rest of the political ecosystem, so it's still pretty mysterious.

I think there are individual factors affecting all of these. As mentioned before, elections have spending limits (however inconsistently enforced) and may not be tractable to money. Think tanks may be more talent-limited than funding-limited. Media properties may be limited by the opinions of their journalists and subscribers (the *Washington Post* couldn’t pivot to being a conservative outlet without getting completely different employees and customers). Tumblr has already proven unable to censor its users without sparking a mass exodus. These issues are probably responsible for part of the underpricing. But it still seems surprising.

In his paper on elections, Ansolabehere focuses on the corporate perspective. He argues that money neither makes a candidate much more likely to win, nor buys much influence with a candidate who does win. Corporations know this, which is why they don’t bother spending more. Most research (plus the 2016 results) confirms that money has little effect on victory, so maybe this is true. But it would also have to be true that lobbying, the NRA, the media, etc don’t affect politics very much, which seems like a harder sell.

That leaves the Bernie Sanders supporters. Even if money doesn’t affect politics, Sanders supporters seem like about the least likely people to believe that. I think here we have to go back to the same explanation I give in Does Class Warfare Have A Free Rider Problem? People just can’t coordinate. If everyone who cared about homelessness donated $100 to the problem, homelessness would be solved. Nobody does this, because they know that nobody else is going to do it, and their $100 is just going to feel like a tiny drop in the ocean that doesn’t change anything. People know that a single person can’t make a difference, so they don’t want to spend any money, so no money gets spent. This is true for ordinary people, but it’s also true for billionaires and greedy corporations. No single greedy corporation wants to pony up the money to change the laws to favor greedy corporations all on its own, while its competitors lie back and free-ride on its hard work. So they basically donate token amounts and do nothing. By all accounts the Koch brothers actually believed in everything they were doing, and they *had to*, because you couldn't make billionaires spend Koch-brothers-like levels of time and money out of self-interest.

In this model, the difference between politics and almonds is that if you spend $2 on almonds, you get $2 worth of almonds. In politics, if you spend $2 on Bernie Sanders, you get nothing, unless millions of other people also spend their $2 on him. People are great at spending money on direct consumption goods, and terrible at spending money on coordination problems.

I don't want more money in politics. But the same factors that keep money out of politics keep it out of charity too.

The politics case is interesting because it's so obvious. Nobody's going to cynically declare "Oh, people don't *really* care who wins the election, they just pretend to." It's coordination problems! It has to be!

So when I hear stories like that Americans could end homelessness by redirecting the money they spend on Christmas decorations, I don't think that's because they're evil or hypocritical or don't really care about the issue. I think they would if they could but the coordination problem gets in the way.

This is one reason I'm so gung ho about people pledging to donate 10% of their income to charity. It mows through these kinds of problems. I may not be a great person. But I spend more each year on the things I consider most important than I do on almonds, and this is the kind of thing that doesn't happen naturally. It's the kind of thing where I have to force myself to ignore the feeling of "just a drop in the ocean”, ignore whether I feel like other people are free-riding on me, and just do it. Pledging to donate money (and then figuring out what to do with it later) ensures I will take that effort, and not end up with revealed preferences that seem ridiculous in light of my values.

Boston Meetup This Friday

When: Friday, September 20, 7:00 PM

Where: 199 Harvard St Apt 2, Cambridge (about 8 minutes walk from Central MBTA station)

Why: I'll be in Boston as part of the Meetups Everywhere tour

Who: Anyone who wants. Please feel free to come even if you feel awkward about it, even if you're not "the typical SSC reader", even if you're worried people won't like you, etc.

How: For more info, contact boston-lw-organizers[at]googlegroups[dot]com, or check the Facebook event or Less Wrong event

NYC Meetup This Saturday

When: Saturday, September 21, 4:00 PM

Where: Pumphouse Park

Why: I'll be in New York as part of the Meetups Everywhere tour

Who: Anyone who wants. Please feel free to come even if you feel awkward about it, even if you're not "the typical SSC reader", even if you're worried people won't like you, etc.

How: For more info, contact tristan[at]thume[dot]ca; 646-480-0618. If it's raining or we're otherwise not there go to the main atrium of Brookfield Place, which is adjacent to the park.

Philadelphia Meetup This Sunday

When: Sunday, September 22, 6:30 PM

Where: Tir Na Nog Bar & Grill, 1600 Arch Street

Why: I’ll be in Philadelphia as part of the Meetups Everywhere tour

Who: Anyone who wants. Please feel free to come even if you feel awkward about it, even if you’re not “the typical SSC reader”, even if you’re worried people won’t like you, etc.

How: For more info, contact wfenza[at]gmail or dmcbriggs[at]gmail. The restaurant is inside the Phoenix Hotel, directly above Suburban Station. We have the entire outer vestibule area reserved.

Washington DC Meetup This Tuesday

When: Tuesday, September 24. The meetup organizers told me to say 5 PM, but unless the stereotypes about government jobs are all true, this sounds too early. I'll try to arrive at 5, but let's unofficially aim for 6:30 as the time most people will be there.

Where: Teaism Coffee shop, 400 8th Street NW, Washington DC. But if it gets too crowded, we might move to the Navy Memorial Plaza just outside. You can check this Google Group for updates on this. Sorry!

Why: I'll be in DC as part of the Meetups Everywhere tour. Also featuring special guests Georgia Ray of Eukaryote Writes, Alex Tabarrok of Marginal Revolution, a few people from MIRI and EA organizations visiting from the Bay, and other people who I might retroactively declare to be special guests once I know about them.

Who: Anyone who wants. Please feel free to come even if you feel awkward about it, even if you're not "the typical SSC reader", even if you're worried people won't like you, even if your approval rating is stuck in the low 40s and you didn't even win the popular vote, etc.

How: For more info, contact robirahman94[at]gmail[dot]com, or check the Google Group.

Ann Arbor Meetup This Thursday

(edit: I am going to be late, eta about 7, sorry)

When: Thursday, September 26, 6:30 PM

Where: Burns Park, near the Warming Hut.

Why: I'll be in Ann Arbor as part of the Meetups Everywhere tour

Who: Anyone who wants. Please feel free to come even if you feel awkward about it, even if you're not "the typical SSC reader", even if you're worried people won't like you, etc.

How: For more info, contact algo2217[at]gmail

Chicago Meetup This Saturday

When: Saturday, September 28, 6:00 PM

Where: South Loop Strength & Conditioning, 645 S Clark St (yes, a gym)

Why: I’ll be in Chicago as part of the Meetups Everywhere tour

Who: Anyone who wants. Please feel free to come even if you feel awkward about it, even if you’re not “the typical SSC reader”, even if you’re worried people won’t like you, etc.

How: For more info, contact todd[at]southloopsc[dot]com

Austin Meetup This Monday

When: Monday, September 30, 6:00 PM

Where: Central Market North Lamar, 4001 N Lamar Blvd, either upstairs on the balcony or outside in the courtyard. If we overflow, we'll move to the park behind central market, ie here.

Who: Anyone who wants. Please feel free to come even if you feel awkward about it, even if you're not "the typical SSC reader", even if you're worried people won't like you, etc.

Why: Austensibly because I'll be there as part of the Meetups Everywhere tour

How: Austintatiously! For more info, contact azavoluk[at]gmail or check the Google Group for updates.

Portland Meetup This Tuesday

When: Tuesday, October 1, 7:00 PM

Where: Cartopia, 1207 SE Hawthorne Blvd. This is a food cart pod with many options. Organizer Noah will stake out a table space with an SSC Meetup sign. Afterwards there might be an afterparty at 1546 SE Ladd Ave (private apartment).

Why: I'll be in Portland as part of the Meetups Everywhere tour

Who: Anyone who wants. Please feel free to come even if you feel awkward about it, even if you're not "the typical SSC reader", even if you're worried people won't like you, etc.

How: For more info, contact nwalton125[at]gmail[dot]com or see the Less Wrong post.

Seattle Meetup This Wednesday

When: Wednesday, October 2, 6:30 PM

Where: The Territory (house), 5238 11th Ave NE, Seattle.

Why: I'll be in Seattle as part of the Meetups Everywhere tour

Who: Anyone who wants. Please feel free to come even if you feel awkward about it, even if you're not "the typical SSC reader", even if you're worried people won't like you, etc.

How: For more info, contact orborde[at]gmail or see the Facebook event.

Fairbanks Meetup This Sunday

When: Sunday, October 6, 9:00 PM

Where: Taco Bell, 1450 University Ave S, Fairbanks, Alaska

Why: It’s complicated.

Who: Anyone who wants. Please feel free to come even if you feel awkward about it, even if you’re not “the typical SSC reader”, even if you’re worried people won’t like you, etc.

How: No organizer for this one, but I’ll check the comments and answer any questions I can.

Berkeley Meetup This Thursday

When: Thursday, October 10, 6:00 PM.

Where: “Event Horizon” (private residence), 2412 Martin Luther King Jr. Way

Why: I’ll be in Berkeley as part of the Meetups Everywhere tour. I might be a bit late to this one, sorry.

Who: Anyone who wants. Please feel free to come even if you feel awkward about it, even if you’re not “the typical SSC reader”, even if you’re worried people won’t like you, etc.

How: For more info, contact mingyuan[at]uchicago[dot]edu or see the Facebook page. The house wants us out by 10 PM, but some people are planning to go to Jupiter afterwards.

Irvine Meetup This Friday

When: Friday, October 11, 7:30 PM.

Where: Underneath this mysterious hexagonal sigil at the University Center food court in Irvine, California.

Why: I’ll be in Irvine as part of the Meetups Everywhere tour.

Who: Anyone who wants. Please feel free to come even if you feel awkward about it, even if you’re not “the typical SSC reader”, even if you’re worried people won’t like you, etc.

How: No contact person for this one, but I’ll try to read and respond to questions in the comments.

Book Review: Against The Grain

Someone on SSC Discord summarized James Scott's *Against The Grain* as "basically 300 pages of calling wheat a fascist". I have only two qualms with this description. First, the book is more like 250 pages; the rest is just endnotes. Second, "fascist" isn't *quite* the right aspersion to use here.

Against The Grain should be read as a prequel to Scott's most famous work, *Seeing Like A State*. SLaS argued that much of what we think of as "progress" towards a more orderly world – like Prussian scientific forestry, or planned cities with wide streets – didn't make anyone better off or grow the economy. It was "progress" only from a state's-eye perspective of wanting everything to be legible to top-down control and taxation. He particularly criticizes the High Modernists, Le Corbusier-style architects who replaced flourishing organic cities with grandiose but sterile rectangular grids.

Against the Grain extends the analysis from the 19th century all the way back to the dawn of civilization. If, as Samuel Johnson claimed, "The Devil was the first Whig", *Against the Grain* argues that wheat was the first High Modernist.

Sumer just before the dawn of civilization was in many ways an idyllic place. Forget your vision of stark Middle Eastern deserts; during the Paleolithic, the area where the first cities would one day arise was a great swamp. Foragers roamed the landscape, eating everything from fishes to gazelles to shellfish to wild plants. There was more than enough for everyone; "as Jack Harlan famously showed, one could gather enough [wild] grain with a flint sickle in three weeks to feed a family for a year". Foragers alternated short periods of frenetic activity (eg catching as many gazelles as possible during their weeklong migration through the area) with longer periods of rest and recreation.

Intensive cereal cultivation is miserable work requiring constant toil with little guarantee of a good harvest. Why would anyone leave this wilderness Eden for a 100% wheat diet?

Not because they were tired of wandering around; Scott presents evidence that permanent settlements began as early as 6000 BC, long before Uruk, the first true city-state, began in 3300. Sometimes these towns subsisted off of particularly rich local wildlife; other times they practiced some transitional form of agriculture, which also antedated states by millennia. Settled peoples would eat whatever plants they liked, then scatter the seeds in particularly promising-looking soil close to camp – reaping the benefits of agriculture without the back-breaking work.

And not because they needed to store food. Hunter-gatherers could store food just fine, from salting animal meat to burying fish and letting it ferment to just having grain in siloes like everyone else. There is ample archaeological evidence of all of these techniques. Also, when you are surrounded by so much bounty, storing things takes on secondary importance.

And not because the new lifestyle made this happy life even happier. While hunter-gatherers enjoyed a stable and varied diet, agriculturalists subsisted almost entirely on grain; their bones display signs of significant nutritional deficiency. While hunter-gatherers were well-fed, agriculturalists were famished; their skeletons were several inches shorter than contemporaneous foragers. While hunter-gatherers worked ten to twenty hour weeks, agriculturalists lived lives of backbreaking labor. While hunter-gatherers who survived childhood usually lived to old age, agriculturalists suffered from disease, warfare, and conscription into dangerous forced labor.

Scott argues that intensive grain cultivation was a natural choice not for cultivators, but for the states oppressing them. The shift from complicated and mobile food webs to a perfectly rectangular grid of wheat fields was the same sort of “progress” as scientific forestry and planned cities thousands of years later:

> Why should cereal grains play such a massive role in the earliest states? After all, other crops, in particular legumes such as lentils, chickpeas, and peas, had been domesticated in the Middle East and, in China, taro and soybean. Why were they not the basis of state formation? More broadly, why have no “lentil states,” chickpea states, taro states, sago states, breadfruit states, yam states, cassava states, potato states, peanut states, or banana states appeared in the historical record? Many of these cultivars provide more calories per unit of land than wheat and barley, some require less labor, and singly or in combination they would provide comparable basic nutrition. Many of them meet, in other words, the agro-demographic conditions of population density and food value as well as cereal grains. Only irrigated rice outclasses them in terms of sheer concentration of caloric value per unit of land.
>
> The key to the nexus between grains and states lies, I believe, in the fact that only the cereal grains can serve as a basis for taxation: visible, divisible, assessable, storable, transportable, and “rationable.” Other crops—legumes, tubers, and starch plants—have some of these desirable state-adapted qualities, but none has all of

these advantages. To appreciate the unique advantages of the cereal grains, it helps to place yourself in the sandals of an ancient tax-collection official interested, above all, in the ease and efficiency of appropriation.

The fact that cereal grains grow above ground and ripen at roughly the same time makes the job of any would-be taxman that much easier. If the army or the tax officials arrive at the right time, they can cut, thresh, and confiscate the entire harvest in one operation. For a hostile army, cereal grains make a scorched-earth policy that much simpler; they can burn the harvest-ready grain fields and reduce the cultivators to flight or starvation. Better yet, a tax collector or enemy can simply wait until the crop has been threshed and stored and confiscate the entire contents of the granary.

Compare this situation with, say, that of farmers whose staple crops are tubers such as potatoes or cassava/manioc. Such crops ripen in a year but may be safely left in the ground for an additional year or two. They can be dug up as needed and the reaminder stored where they grew, underground. If an army or tax collectors want your tubers, they will have to dig them up tuber by tuber, as the farmer does, and then they will have a cartload of potatoes which is far less valuable (either calorically or at the market) than a cartload of wheat, and is also more likely to spoil quickly. Frederick the Great of Prussia, when he ordered his subjects to plant potatoes, understood that, as planters of tubers, they could not be so easily dispersed by invading armies.

The “aboveground” simultaneous ripening of cereal grains has the inestimable advantage of being legible and assessable by the state tax collectors. These

> characteristics are what make wheat, barley, rice, millet, and maize the premier political crops. A tax assessor typically classifies fields in terms of soil quality and, knowing the average yield of a particular grain from such soil, is able to estimate a tax. If a year-to-year adjustment is required, fields can be surveyed and crop cuttings taken from a representative patch just before harvest to arrive at an estimated yield for that particular crop year. As we shall see, state officials tried to raise crop yields and taxes in kind by mandating techniques of cultivation; in Mesopotamia this included insisting on repeated ploughing to break up the large clods of earth and repeated harrowing for better rooting and nutrient delivery. The point is that with cereal grains and soil preparation, the planting, the condition of the crop, and the ultimate yield were more visible and assessable.

Scott’s great advantage over other writers is the care he takes in analyzing the concrete machinery of statehood. Instead of abstractly saying “the state levies a 10% tax”, he realizes that some guy in a palace has resolved to take “ten percent” of the “value” produced in some vast area, with no natural way of knowing who is in that area or how much value they produce. For most of the Stone Age, this problem was insurmountable. You can’t tax hunter-gatherers, because you don’t know how many they are or where they are, and even if you search for them you’ll spend months hunting them down through forests and canyons, and even if you finally find them they’ll just have, like, two elk carcasses and half a herring or something. But you also can’t tax potato farmers, because they can just leave when they hear you coming, and you will never be able to find all of the potatoes and dig them up and tax them. And you can’t even tax lentil farmers, because you’ll go to the lentil plantation and there will be a few lentils on the plants and the farmer will just say “Well, come back next week and there will be a few more”, and you can’t visit every citizen every week.

But you can tax grain farmers! You can assign them some land, and come back around harvest time, and there will be a bunch of grain just standing there for you to take ten percent of. If the grain farmer flees, you can take his grain without him. Then you can grind the grain up and have a nice homogenous, dense, easy-to-transport grain product that you can dole out in measured rations. Grain farming was a giant leap in oppressability.

In this model, the gradual drying-out of Sumeria in the 4th millennium BC caused a shift away from wetland foraging and toward grain farming. The advent of grain farming made oppression possible, and a new class of oppression-entrepreneurs arose to turn this possibility into a reality. They incentivized farmers to intensify grain production further at the expense of other foods, and this turned into a vicious cycle of stronger states = more grain = stronger states. Within a few centuries, Uruk and a few other cities developed the full model: tax collectors, to take the grain; scribes, to measure the grain; and priests, to write stories like The Debate Between Sheep And Grain, with immortal lines like:

> From sunrise till sunset, may the name of Grain be praised. People should submit to the yoke of Grain. Whoever has silver, whoever has jewels, whoever has cattle, whoever has sheep shall take a seat at the gate of whoever has Grain, and pass his time there

And so the people were taught that growing grain was Correct and Right and The Will Of God and they shouldn’t do anything stupid like try to escape back to the very close and easily-escapable-to areas where everyone was still living in Edenic plenty.

…turns out lots of people in early states escaped to the very close and easily-escapable-to areas where everyone was still living in Edenic plenty. Early states were necessarily tiny; overland transportation of resources more than a few miles was cost-prohibitive; you could do a little better by having the state on a river and adding in water transport, but Uruk’s sphere of influence was still probably just a double-digit number of kilometers. Even in good times, peasants would be tempted to escape to the hills and wetlands; in bad times, it started seeming crazy *not* to try this. Scott suggests that ancient Uruk had a weaker distinction between “subject” and “slave” than we would expect. Although there were certainly literal slaves involved in mining and manufacturing, even the typical subject was a serf at best, bound to the land and monitored for flight risk.

In one of my favorite parts of the book, Scott discusses how this shaped the character of early Near Eastern warfare. Read a typical Near Eastern victory stele, and it looks something like “Hail the glorious king Whoever, who campaigned against Such-And-Such and took 10,000 prisoners of war back to the capital.” Territorial conquest, if it happened at all, was an afterthought; what these kings really wanted was prisoners. Why? Because they didn’t even have enough subjects to farm the land they had; they were short of labor. Prisoners of war would be resettled on some arable land, given one or another legal status that basically equated to slave laborers, and so end up little different from the native-born population. The most extreme example was the massive deportation campaigns of Assyria (eg the Ten Lost Tribes of Israel), but everybody did it because everybody knew their current subjects were a time-limited resources, available only until they gradually drained out into the wilderness.

Early states were pretty time-limited themselves. Scott addresses the collapse of early civilizations, which was ubiquitous; typical history disguises this by talking about “dynasties” or “periods” rather than “the couple of generations an early state could hold itself together without collapsing”.

> Robert Adams, whose knowledge of the early Mesopotamian states is unsurpassed, expresses some astonishment at the Third Dynasty of Ur (Ur III), in which five kings succeeded one another over a hundred-year period. Though it too collapsed afterward, it represented something of a record of stability.

Scott thinks of these collapses not as disasters or mysteries but as the expected order of things. It is a minor miracle that some guy in a palace can get everyone to stay on his fields and work for him and pay him taxes, and no surprise when this situation stops holding. These collapses rarely involved great loss of life. They could just be a simple transition from “a bunch of farming towns pay taxes to the state center” to “a bunch of farming towns are no longer paying taxes to the state center”. The great world cultures of the time – Egypt, Sumeria, China, whereever – kept chugging along whether or not there was a king in the middle collecting taxes from them. Scott warns against the bias of archaeologists who – deprived of the great monuments and libraries of cuneiform tablets that only a powerful king could produce – curse the resulting interregnum as a dark age or disaster. Probably most people were better off during these times.

The book ends with a chapter on “barbarians”. Scott reminds us that until about 1600, the majority of human population lived outside state control; histories that focus on states and forget barbarians are forgetting about most humans alive. In keeping with his thesis, Scott reviews some ancient sources that talk about barbarians in the context of people who did not farm or eat grain. Also in keeping with his thesis, he warns against thinking of barbarians as somehow worse or more primitive. Many barbarians were former state citizens who had escaped state control to a freer and happier lifestyle. Barbarian tribes could control vast trading empires, form complex confederations, and enter in various symbiotic relationships with the states around them. Scott wants us to think of these not as primitive people vs. advanced people, but as two different interacting lifestyles, of which the barbarian one was superior for most people up until a few centuries ago.

Overall I liked this book. I’m not sure how convinced I am – Scott occasionally mentions how much denser (in terms of calories produced per unit land) grain is than other forms of subsistence, and this surely deserves more consideration as an alternative explanation for its success. But overall the theory is plausible as at least one of many explanations for the grain/state correlation.

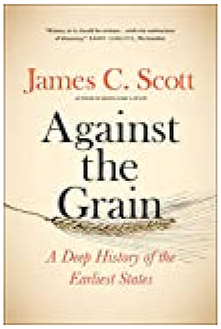
My only other complaint is the constant insistence throughout the book that we should be having our minds blown by it. Scott talks about how he wanted to give a lecture on the rise of civilization in Sumeria, hadn’t studied the subject for a few decades, thought he’d do a quick review of what had been discovered in the interim, and instead found that everything he knew was wrong. He talks a lot about how the conventional narrative of the dawn of agriculture has been turned on its head, overthrown, debunked, etc, and how you need to unlearn all your brainwashing about the superiority of states to hunter-gatherers.

But Jared Diamond was calling agriculture The Worst Mistake In The History Of The Human Race back in the 1980s. And the changes to the Sumeria story I learned in school seem like updates rather than paradigm shifts. Yes, people were sedentary agriculturalists long before Uruk – but I remember a page in my elementary school textbook (so we're talking 1995 or so) going over Catal Huyuk and its neighbors in 6000 BC. Yes, early city-states sucked – but does anyone think of "Bronze Age god-king" and imagine a nice guy committed to egalitarianism? The Epic of Gilgamesh was talking about the suckiness of Bronze Age city-states before the Bronze Age even ended. The most surprising revision to the standard story in *Against The Grain* was the setting of early Sumer in wetland rather than desert. And even that is only a small change; the first cities were on a kind of flat alluvium separate from the wetland proper, and their environmental damage quickly dried the region up into the irrigation-heavy desert we know today.

Scott tries to downplay his own role in the book, emphasizing how much he is just relaying the discoveries of more accomplished Sumer experts than himself. But the part I most appreciated was the part that was most clearly Scott-ish: the role of grain as a state-builder. In this story, the beginning of civilization – like the progress of the High Modernists – wasn't an advance in human welfare or economic growth. It was an advance in tax collecting and the machinery of oppression; everything else followed.

"From sunrise till sunset, may the name of Grain be praised", said the Sumerians. And the ancient Greeks had their Eleusinian Mysteries, where "the mighty, and marvelous, and most perfect secret suitable for one initiated into the highest truths” was the “revelation of the mystic grain”. Can we trace a direct line from there to the sheaves of wheat that feature on fifteen out of fifty US state seals? On the National Emblem of China? The Coat of Arms of the Soviet Union? Does this last one really show the Earth caught in a pincers between two giant stalks of wheat? Should we really make impressionable schoolchildren sing songs of praise for “amber waves of grain”?

Read this book, and you may never think about cereal crops the same way again.


James C. Scott
Against the Grain
A Deep History of the Earliest States

Is Enlightenment Compatible With Sex Scandals?

Last year I reviewed The Mind Illuminated, a meditation guide by Buddhist teacher Upasaka Culadasa. Last month, Culudasa's Buddhist community accused him of cheating on his wife with prostitutes for many years. Culadasa doesn't seem to agree with the exact details of the accusations, but he also doesn't seem to deny that there was something in that general category of thing. What can this teach us about enlightenment?

Culadasa has been meditating and studying Buddhism for over forty years and trained under some of the greatest teachers of his generation. I don't know if he's claimed to "be enlightened" in so many words, but he's written books that describe how to reach enlightenment and that assert you can do it in a few years if you follow his advice, which sounds a lot like claiming enlightenment by implication. Other self-proclaimed enlightened Buddhist teachers seem to respect him and treat him as being at around their level.

And if Culudasa wasn't enlightened, there's a long list of other Buddhist masters with similar misdeeds. *The Atlantic* points out that three of the four great founders of American Zen "caused major public sex scandals"; the fourth, Shunryu Suzuki, was spotless, but his successor Richard Baker caused a major public sex scandal. The two most famous US teachers of Tibetan Buddhism, Chongyam Trungpa and Sogyal Rinpoche, both caused major public sex scandals. Trungpa's immediate successor Ösel Tendzin caused a particularly horrifying major public sex scandal, and the current head of Shambhala Buddhism, Sakyong Rinpoche, also caused a major public sex scandal.

These teachers were among the most accomplished of our time. Many were officially certified as enlightened by the relevant governing bodies (of course there are governing bodies that certify enlightenment, we’re not *barbarians*). Doubt Culudasa if you want, but it would be hard to say *none* of these people had achieved enlightenment – at least if you want to maintain any reason to believe in enlightenment as an achievable state at all.

I don’t think many modern teachers say enlightenment makes you morally perfect. But I think at least some of them say it makes you free from craving or desire. And repeatedly cheating on your wife doesn’t seem like the action of someone who’s free from desire. It doesn’t even seem like someone whose desire has been moderately decreased. It sounds like the action of someone who has *at least* as much desire as anyone else. Maybe Buddhists should retreat to a minimalist account of enlightenment where it changes some brain networks around in a way that short-circuits some processing of experiences of suffering and selfhood, but doesn’t really lead to better decisions?

Tricycle Magazine discusses various theories for why Buddhist sex scandals are so common. Maybe Asians from patriarchal cultures do badly when transplanted to the more sexually liberal West (…but Culadasa was white and born in the US). Maybe powerful men are naturally tempted to behave badly when surrounded by vulnerable female students (but Culadasa didn’t have sex with his students). Maybe the Mahayana emphasis on how enlightened people transcend ordinary human norms causes enlightened people to, uh, transcend ordinary human norms (but most of Culadasa’s training was Theravada).

I recently got a chance to talk to about this with a very experienced Buddhist practitioner, one who claims to be enlightened himself. He said it’s accepted in his tradition that meditation “dissolves social conditioning”. In theory once you’ve dissolved all social conditioning, the Inner Light Of Compassion shines through and you can behave with true kindness. But in practice the Inner Light Of Compassion sometimes goes AWOL and you’re just left valueless. This works fine if you’re in a monastery like most advanced meditators were for most of history, not so well if you’re out in the real world with all the usual temptations.

This fascinates me for the same reason HPPD fascinates me. There are all these transformative practices that purport to give you a higher level of consciousness. But by Algernon’s Law, there’s presumably some reason we’re in *this* state of consciousness, some reason our system protects its usual state so diligently that you need powerful drugs or years of meditation to break through to anything else. Are there advantages to samsara? Are they related to the reason why so many enlightened people end up in sex scandals?

Or to put it another way: if meditation, like LSD, relaxes mental priors and increases entropy, do these failure modes help us understand what strong priors and low entropy are good for?

The PNSE Paper

I’ve mentioned this a few times, but it’s worth going over in detail. The full title is Clusters Of Individual Experiences Form A Continuum Of Persistent Non-Symbolic Experiences In Adults by Jeffery Martin, with “persistent non-symbolic experience” (PNSE) as a scientific-sounding culturally-neutral code word for “enlightenment”. Martin is a Reiki practitioner associated with the “Center for the Study of Non-Symbolic Consciousness”, so we’re not getting this from the most sober of skeptics, but I still find the project interesting enough to deserve a look.

Martin searched various religious and spiritual groups for people who both self-reported enlightenment and were affiliated with “a community that provided validity to their claims”. He says he eventually found 1200 such people who were willing to participate in the study, but that “the data reported here comes primarily from the first 50 participants who sat for in-depth interviews…based on the overall research effort these 50 were felt to be a sufficient sample to represent what has been learned from the larger population”. Although Martin says he tried to get as much diversity as possible, the group was mostly white male Americans.

Martin’s research was mostly qualitative, based on in-depth interviews, so we’re mostly going with his impressions. But his impression was that most people who self-described as enlightened had similar experiences, which could be be plotted on:

> …a continuum that seemed to progress from ‘normal’ waking consciousness toward a distant location where participants reported no individualized sense of self, no

> self-related thoughts, no emotion, and no apparent sense of agency or ability to make a choice. Locations prior to this seemed to involve consistent changes toward this direction.

He describes this distant form of consciousness as involving changes in sense-of-self, cognition, emotion, memory, and perception.

Starting with sense-of-self, he says:

> Perhaps the most universal change in what PNSE participants reported related to their sense of self. They experienced a fundamental change from a highly individualized sense of self, which is common among the ‘normal’ population, to something else. How that ‘something else’ was reported often related to their religious or spiritual tradition(s), or lack thereof. For example, Buddhists often referred to a sense of spaciousness while Christians frequently spoke of experiencing a union with God, Jesus, or the Holy Spirit depending on their sect. However, each experienced a transformation into a sense of self that seemed ‘larger’ and less individuated than the one that was experienced previously. Often participants talked about feeling that they extended beyond their body, sometimes very far beyond it…
>
> This change was dramatic and most participants noticed it immediately, even if initially they could not pinpoint exactly what had occurred. Sense of self changed immediately in approximately 70% of participants. In the other 30% it unfolded gradually, with the unfolding period reported as varying from a few days to four months.

> Those who were not involved in a religious or spiritual tradition that contextualized the experience often felt that they might have acquired a mental disorder. This analysis was not based on emotional or mental distress. It was typically arrived at rationally because the way they were experiencing reality was suddenly remarkably different than they had previously, and as far as they could tell different from everyone they knew. Many of these participants sought professional mental health care, which no participant viewed as having been beneficial. Clinicians often told them their descriptions showed similarities to depersonalization and derealization, except for the positive nature of the experience.
>
> There were nuances within how sense of self was experienced at different locations along the continuum. In the earliest locations, the sense of self felt expanded, and often seemed more connected to everything. In the farthest locations on the continuum, an even more pronounced change occurred in sense of self; a ll aspects of having an individualized sense of self had vanished for these participants. Prior to this location some aspects of an individualized sense of self remained, and participants could occasionally be drawn into them.

On cognition:

> Another consistent report is a shift in the nature and quantity of thoughts. Virtually all of the participants discussed this as one of the first things they noticed upon entering PNSE. The nature and degree of the change related to a participant’s location on the continuum. On the early part of the continuum, nearly all participants reported a significant reduction in, or

even complete absence of, thoughts. Around 5% reported that their thoughts actually increased. Those who reported thoughts, including increased thoughts, stated that they were far less influenced by them. Participants reported that for the most part thoughts just came and went, and were generally either devoid of or contained greatly reduced emotional content.

Almost immediately it became clear that participants were not referring to the disappearance of all thoughts. They remained fully able to use thought for problem solving and living what appeared outwardly to be a ‘normal’ life. The reduction seemed limited to self-related thoughts. Nevertheless, participants were experiencing a reduction in quantity of thoughts that was so significant that when they were asked to quantify the reduction, those who could answered within the 80-95% range. This high percentage may suggest why someone would say all thought had fallen away.

There do not appear to be negative cognitive consequences to this reduction in thought. When asked, none said they wanted their self-referential thoughts to return to previous levels or to have the emotional charge returned to them. Participants generally reported that their problem solving abilities, mental capacity, and mental capability in general had increased because it was not being crowded out or influenced by the missing thoughts. They would often express the notion that thinking was now a much more finely tuned tool that had taken its appropriate place within their psychological architecture.

On perception:

> Participants in the later part of the middle range of the PNSE continuum often reported seeing the unfolding layers of these perceptual processes in detail. They reported being able to begin to detect the difference between the orientation response and the physical, cognitive, and emotional processes that arose after it. They reported reaching a point where some events were reacted to by one or more of these layers while others were not. This was in contrast to participants on the early end of the continuum who perceived all of these layers as one during an event, or at least as a greatly reduced number of discrete processes.

You can read more, plus the sections on emotion and memory, yourself; they mostly fit with the stereotypes you would expect of enlightened people; a lot of tranquility, joy, and focus on the present moment.

What I like about this paper is the parts where it departs from these stereotypes. It makes clear that most of these people's external characteristics didn't change at all. In many cases, their friends and family didn't even notice anything was different, and could not be convinced that anything about them was different:

> Despite an overwhelming change in how it felt to experience both themselves and the world after the onset of PNSE, the outward appearance of the participants changed very little. Generally speaking they retained their previous mannerisms, hobbies, political ideology, food and clothing preferences, and so forth. If someone were an environmentalist prior to PNSE, typically they remained so after it. If they weren't, they still are not.

> Many participants discussed the thought, just after their transition to PNSE, that they would have to go to work and explain the difference in themselves to co-workers. They went on to describe a puzzled drive home after a full day of work when no one seemed to notice anything different about them. Quite a few chose to never discuss the change that had occurred in them with their families and friends and stated that no one seemed to notice much of a difference. In short, although they had experienced radical internal transformation, externally people didn’t seem to take much notice of it, if any.

Similarly, despite people saying that they no longer had any sense of agency, they were behaving as agentically as anyone else:

> On the far end of the continuum, participants reported no sense of agency. They reported that they did not feel they could take any action of their own, nor make any decisions. Reality was perceived as just unfolding, with ‘doing’ and ‘deciding’ simply happening. Nevertheless, many of these participants were functioning in a range of demanding environments and performing well. One, for example, was a doctoral level student at a major university. Another was a young college professor who was building a strong career. Still another was a seasoned public and private sector executive who served as a high-level consultant and on various institutional-level boards.

Can you imagine investing in a company whose executive believes he cannot take any action and is just watching reality unfold? But it seems to work out.

Other times the PNSE participants are just outright wrong about their experience. When asked if they were stressed, they would say of course not, they were experiencing inner peace. But their friends and family said they were totally stressed. For example:

> Over the course of a week, [one participant’s] father died, followed very rapidly by his sister. He was also going through a significant issue with one of his children. Over dinner I asked him about his internal state, which he reported as deeply peaceful and positive despite everything that was happening. Having known that the participant was bringing his longtime girlfriend, I’d taken an associate researcher with me to the meeting to independently collect the observations from her. My fellow researcher isolated the participant’s girlfriend at the bar and interviewed her about any signs of stress that the participant might be exhibiting. I casually asked the same questions to the participant as we continued our dinner conversation. Their answers couldn’t have been more different. While the participant reported no stress, his partner had been observing many telltale signs: he wasn’t sleeping well, his appetite was off, his mood was noticeably different, his muscles were much tenser than normal, his sex drive was reduced, his health was suffering, and so forth.

Or:

> It was not uncommon for participants to state that they had gained increased bodily awareness upon their transition into PNSE. I arranged and observed private yoga sessions with a series of participants as part of a larger inquiry into their bodily awareness. During these sessions it became clear that participants believed they were far more aware of their body than they actually were. For example, the instructor would often put her hand on part of the body asking the participant to relax

> the tense muscles there, only to have the participant insist that s/he was totally relaxed in that area and did not feel any muscle tension.

Or even:

> During some interviews participants expressed that they no longer felt it was possible for them to be racist or sexist. I asked these participants to take Harvard University's Project Implicit tests online. All of these participants were white males and each showed a degree of sexism and/or racism, including participants who were in the later no emotion and agency locations on the continuum. Project Implicit uses physiology to test these responses.

It's tempting to say these people are just making it up. But I think about some of the people I know with very severe psychiatric issues, people who are constantly miserable – and are similarly externally unaffected. These people are holding down stressful jobs, keeping difficult relationships together, etc – and often the people they haven't "opened up to" don't have any inkling of what they're going through. They may tell me it must seem obvious to everybody that they're completely falling apart – whereas in fact they are speaking fluently, they're well-dressed, and they haven't made a single social misstep during the whole time I've known them. If unusually negative mental states don't affect behavior as strongly as people believe, why not unusually positive mental states?

Also, other times these people *under*-estimate themselves:

> As participants neared the further reaches of the continuum, they frequently reported significant difficulty with recalling memories that related to their life history.

> They did not feel this way about facts, but rather about the details of the biographical moments surrounding the learning of those facts. They also reported that encoding for these types of memories seemed greatly reduced. A lthough this was their perception it did not appear to be the case when talking to them. They were typically rich sources of personal history information and their degree of recall seemed indistinguishable from participants who were in earlier locations on the continuum.

But:

> There was a noticeable exception that seemed to be a genuine deficit. As they neared and entered the farther reaches of the continuum, participants routinely reported that they wereincreasingly unable to remember things such as scheduled appointments, while still being able to remember events that were part of a routine. For example, they might consistently remember to pick their child up at school each day, but forget other types of appointments such as doctor visits. Often they had adapted their routines to adjust for this change. Many would immediately write down scheduled events, items they needed to get at the store, and so forth on prominently displayed lists. When visiting their homes I noticed that these lists could be found on: televisions, computer monitors, near toilets, on and next to doors, and so forth. It was clear that the lists were being placed in locations that the participants would look with at least some degree of regularity. Participants consistently stated that they would prefer to remain in PNSE even if going back to ‘normal’ experience meant that they would no longer have this type of deficit.

Finally, Martin is impressed with the certainty that accompanies all of these experiences. People describe their

PNSE as obviously more real and better than past states. They tend to be very effusive about this, saying that having the experience shattered everything they had previously believed in the most obvious and final way. But here too, there are signs that the participants are not well-attuned to what is going on in their own heads. Martin says that participants who moved from one level of his continuum to another (whether forward or back) would always say that the level they were currently at was the most fundamental and obviously real (even if they had said the opposite before). When he would tell participants about the experiences of other participants who were at different points of the continuum or just describing their experiences a slightly different way, both participants would confidently pronounce that the other wasn't *really* enlightened.

I like this paper because it provides the basis for a minimalist account of enlightenment, similar to Daniel Ingram's. Enlightenment hasn't transformed these people's personalities. It hasn't given them infinite willpower or productivity or the ability to shoot qi bolts from their third eyes. It hasn't even given them that much self-understanding. It's just given them a different kind of internal experience.

The experience itself is hard to describe, but seems marked by drawing the self-other boundary in a different place. Participants don't see themselves as making decisions; the decisions get made "under the hood" in a way where the person just feels like their path is laid out before them. They don't see themselves as having thoughts; computations obviously get done, but they are not in awareness. They don't feel like they have stress, even if the stress is physiologically present and obvious from their actions. On the other hand, they were more aware of certain low-level perceptual processes that are usually unconscious. It seems to be accompanied by total certainty that this is correct and revelatory and new (…much like the altered states people sometimes get on drugs).

None of this seems wildly outside the realm of possibility. It seems about as surprising as the existence of some new mental disorder. If 50 (or 1200, depending on how you count it) people with no history of lying said they had some kind of weird new mental disorder, I’d be willing to credit that they were describing their experience correctly, and able to give some useful information on the sorts of things that caused this disorder. It just sounds like information processing in the brain switching to some new attractor state if you force it hard enough.

Indian Economic Reform: Much More Than You Wanted To Know

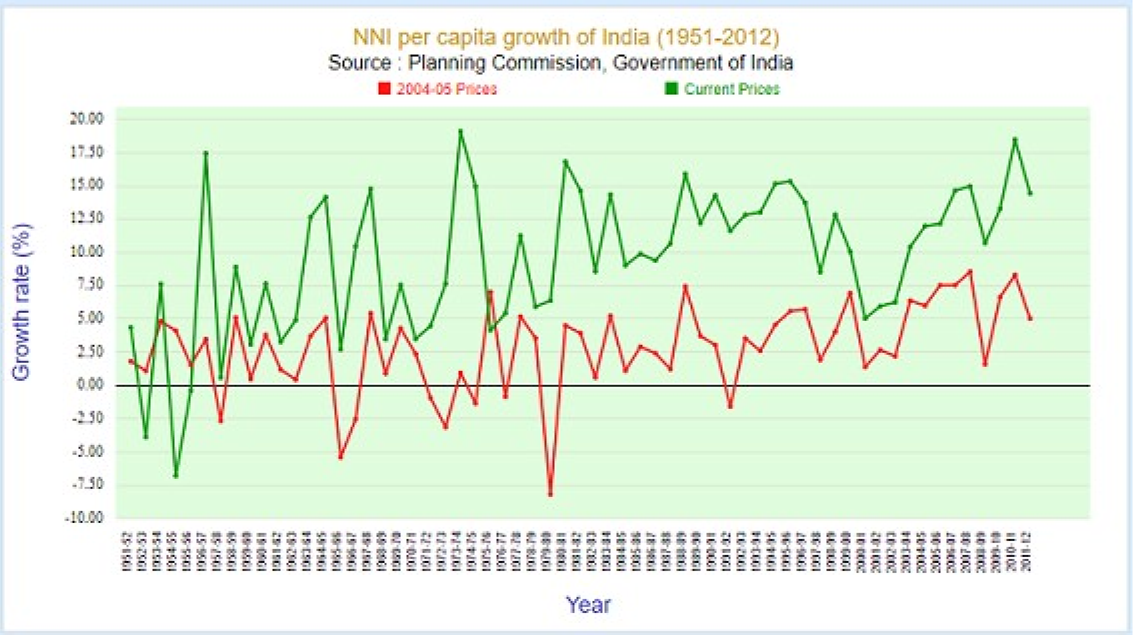
From a recent Charter Cities Institute report:

> From India's independence from the British Raj in 1947 to the early 1990s, the country's economic policy was largely socialist. In the 1980s some early steps were taken to open the Indian economy to increased trade, reduce controls over industry, and set a more realistic exchange rate. In 1991, more widespread economic reforms were introduced. These reforms included the end of government monopolies over certain sectors of the economy, reductions in barriers to entry for new firms, increased foreign investment was allowed, and tariffs and other barriers to trade were reduced or eliminated. After liberalization, exports increased substantially, and various service sector industries saw significant growth.
>
> India's growth has not just been good for the more educated segment of the population. Datt, Ravallion, and Murgai (2016) argue that India has made substantial progress in reducing the incidence of absolute poverty, and that this trend exists in both urban and rural areas. Historically higher rates of rural poverty have been converging with urban rates of poverty, and the overall poverty rate has been declining at an accelerating rate in the post-1991 reform era. In the 1970s over 60 percent of Indians were living in extreme poverty. As of 2011, only 20 percent of the population lived in extreme poverty. Between 2005 and 2016, an estimated 271 million Indians rose out of multidimensional poverty, which accounts for various

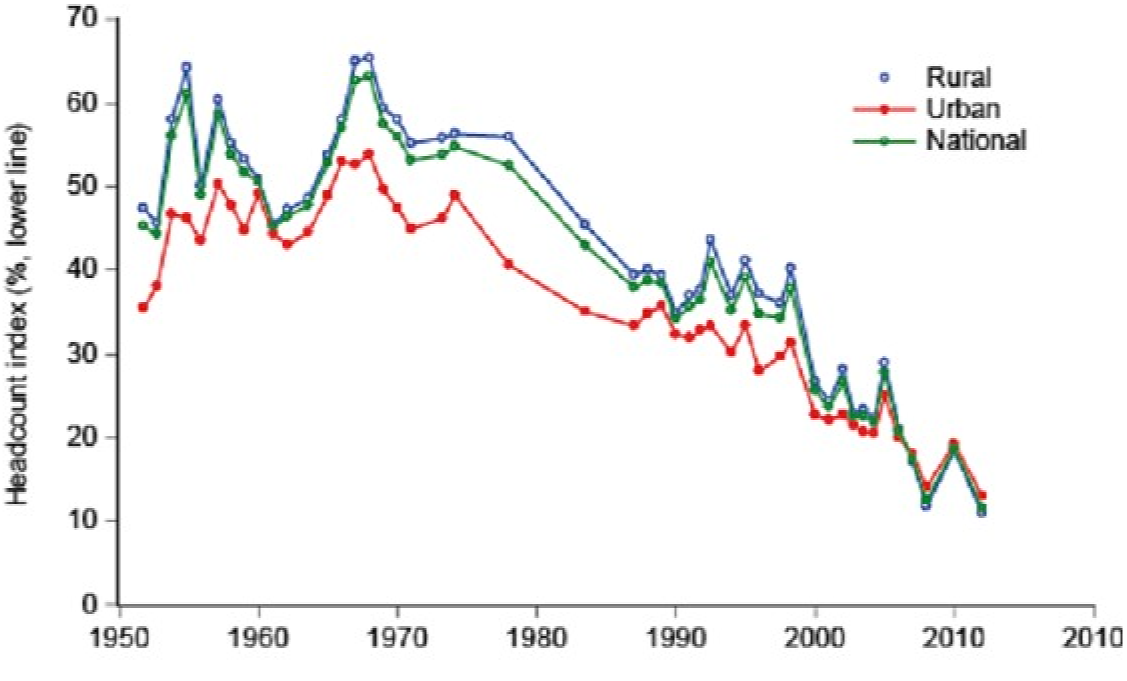
> health, education, and living standard indicators rather than just income (UNDP and OPHI 2018). Infant mortality has fallen from 161.4 deaths per 1,000 births in 1960 to just 32 deaths per 1,000 births in 2017, and India should soon converge with the world average if the current trend continues. Life expectancy has also improved dramatically, rising from 41 years in 1960 to nearly 69 years today. Like with infant mortality, India is close to converging with the world average in life expectancy. Literacy has improved from just 41 percent in 1981 to 72 percent in 2015, an increase of 75 percent. Here too, India is converging with the world average. Female literacy in particular rose from just 25 percent in 1981 to nearly 60 percent in 2011, and female primary school enrollment has increased from 65 percent in 1990 to over 98 percent today. Across the board of development measures, India has made tremendous strides.

Reading this surprised me. I was vaguely aware that India had done relatively well, but I didn’t grasp the scale. This should be up there with the rise of China as one of the most important (and most encouraging) news stories of my lifetime. And if it was really due to the 1991 reforms, they should go down alongside Deng Xiaoping’s liberalization of China as one of the century’s great achievements.

Looking into it further, the progress against poverty is on firm ground, but the attribution to the 1991 reforms is controversial. Here’s India’s rate of GDP growth over time:

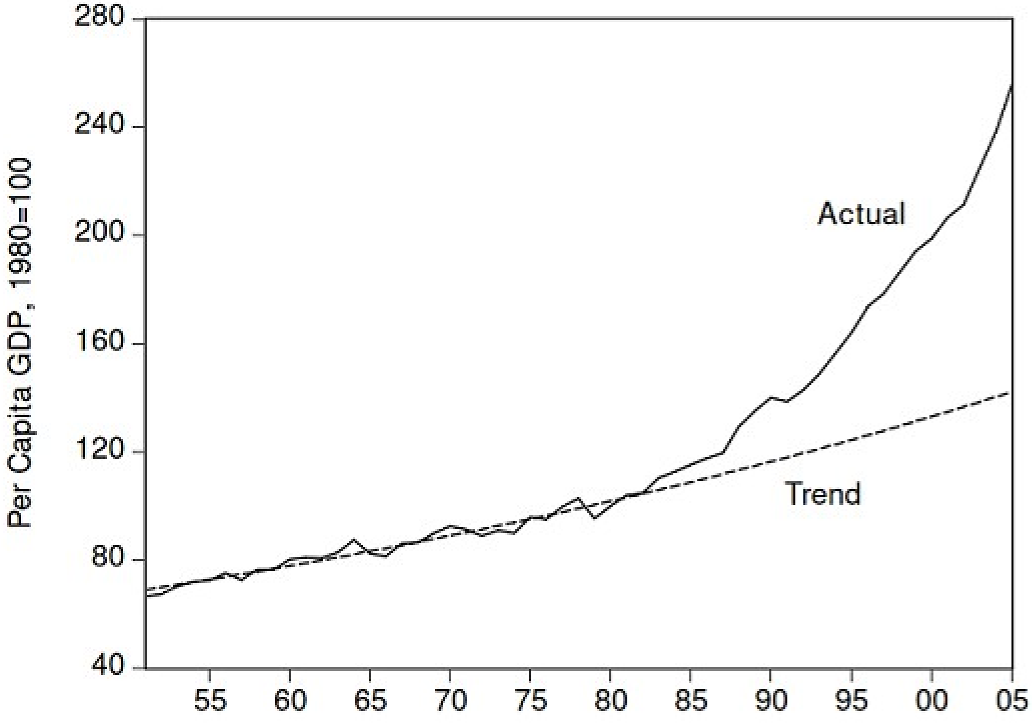


And here’s its poverty rate over time:



Neither looks like much happened in 1991. Both show, if anything, gradual progress from around 1980. Kotwal, Ramaswami and Wadhwa have an especially clear presentation of this:

Figure 2: Growth Acceleration in GDP since the eighties



For their claim about 1991, CCI cites Indian-American economist Arvind Panagariya’s 2004 paper. Panagariya is aware that the graphs don’t back him up, and cites other economists like Brad DeLong and Dani Rodrik making approximately this point. But he argues that liberalization deserves credit for India’s growth anyway, on two grounds. First, he says that there was “stealth reform” in the mid-1980s, when reformers opened the economy without publicizing what they were doing in order to avoid a crackdown from angry leftists. These were complete by 1987, leading to a boom from 1987 – 1991. Second, this growth was unsustainable, and ended in a 1991 crash. Only after the major 1991 reforms did India enter a sustained period of economic growth.

But Panagariya’s panegyric is founded on the growth starting in 1987; the graphs don’t support that story either.

He can only note that:

> It is difficult to pinpoint the timing of the upward shift in India's growth rate. Thus, in a recent attempt to pinpoint structural breaks in the growth series, Wallack (2003) is able to achieve at best partial success. She finds that with a 90 percent probability the shift in the growth rate of GDP took place between 1973 and 1987. The associated point estimate of the shift, statistically significant at 10 percent level, is 1980. When Wallack replaces GDP by gross national product (GNP), however, the cutoff point with 90 percent probability shifts to the years between 1980 and 1994. The associated point estimate, statistically significant at 10 percent level, now turns out to be 1987.

In other words, statistics is hard, and random upward swings can segue into real changes in trend in hard-to-analyze ways. It's impossible to say for sure from the GDP numbers that there wasn't a (random) boom in 1980 that gradually merged with the (real) boom starting in 1987. This works with the GDP numbers, but I find it basically impossible to square with the graph on poverty (which is from 2016 and which Panagariya didn't have access to).

CCI also cites Datt, Ravillion, and Murgal (paper, article), which claims that "economic reforms following the macroeconomic crisis of 1991-92 marked a significant change in India's economic landscape, ushering in a new phase of high economic growth. The growth rate of NDP per capita more than doubled in the period since 1992." But on closer inspection, they are just comparing an average for the period 1957 – 1991 with an average for the period 1992 – 2012, and finding the latter average is higher. They make no effort to establish that the break point is actually in 1991. Since people had been complaining about this for about ten years before the publication of their paper, I’m not sure what their excuse is, other than that they’re mostly making an unrelated point about poverty and this is not super relevant.

More recent work seems to agree that there is something fishy here. Agarwal and Whalley (2013) concludes that:

> We do not find persuasive the contention of manyanalysts that growth accelerated after the mid-1980s when reforms were initiated. Nor does statisticalanalysis support the contention that reforms in the mid-1980s resulted in a growth acceleration. Weshow that there is an accelerating rate of growth of GDP after the mid 1970s and it is difficult to relate this gradual acceleration to specific policy changes. The changed policies in the 1980s did not meana basic change in the policy framework. Furthermore, since corporate investment as a share of GDP did not increase in the 1980s it is difficult to identify the mechanism by which the more pro-business policies of the government were translated to higher growth.

And Kotwal, Ramaswami, and Wadhwa (2016) says basically the same:

> Formal econometric tests also indicate a structural break around 1980. Using an F-test, Wallack (2003) finds the highest value of the F-statistic in 1980. Rodrik and Subramanian (2004) use a procedure of Bai and Perron(1998, 2003) and they report a single structural break in 1979. Balakrishnan and Parameswaran (2007) also used the Bai and Perron procedure and they too locate a single structural break in GDP in 1978-79. The authors also estimate structural breaks for sectoral GDP. Their principal finding is that structural break in

> agricultural output occurs in the mid-1960s while it occurs in the early to mid-1970s for various sub-sectors of services. On the other hand, the first positive structural break in manufacturing occurs after the GDP break in 1982-83.
>
> Basu (2008) and Sen (2007), however, point out that GDP fell by 5.2% in 1979-80 (due to a combination of a drought and the second oil price shock). If this outlier is disregarded, then the trend break occurs in 1975-76. The average annual growth rate during the period 1975-78 is 5.8% – a rate more in line with the post-1980 experience than with the earlier period.
>
> Is the timing of the structural break important? The discussion in the literature about the structural break takes place in the belief that it could offer clues about what policies led to the shift in the economy’s growth rate. Such inference is problematic because statistical methods alone are unlikely to provide a precise timing. Judgments about outliers, the period of analysis, and the sectors that are considered, matter. An additional complication is that policy measures do not have instantaneous results. The delay would be especially pronounced if the benefits flow from a structural change. It is therefore unwise to correlate the changes in economic variables to the policy changes that immediately preceded them. These caveats notwithstanding, the economy does seem to have moved to a higher growth trajectory sometime in the mid to late 1970s or early 1980s, well before the economic reforms of 1991.

So what did cause the Indian economic boom, save 200 million people from poverty, and accomplish an almost unmatched victory over misery and mortality? After discussing a series of possibilities, Kotwal, Ramaswami, and Wadhwa…admit that they are confused:

> Although it is clear that GDP growth rates increased sometime in the ‘70s or early ‘80s, the precise timing is hard to establish and depends on one’s prior. Various explanations have been proposed and it is impossible to be sure which of these is the most important one. The economic orthodoxy would favor one that credits trade liberalization, limited as it was, that decreased the cost of capital equipment but it is hard to disentangle the effects of this from more heterodox factors such as public investment and rise in savings rate (due to bank nationalization), the diffusion of agricultural technology (entirely due to public research and dissemination) or indeed to rule out the role of political attitudes towards business. It is also indisputable that there was an unsustainable fiscal expansion through 1980’s and any income growth resulting from it should be considered qualitatively different from the much more sustainable growth that occurred in the next decade.

Of the various papers they cite, the one I find most interesting is Rodrik and Subramanian (2004). After a while describing the problem, they look at it from a different angle: which Indian states boomed first? They find it was the states most closely allied with the ruling Congress Party, and tell a story where floundering Prime Minister Indira Gandhi decided she needed support from Big Business and adopted a pro-Big-Business attitude. They do excellent work establishing the plausibility of every link in the chain of assumptions here *except* the one where they admit that this attitude wasn’t reflected in any pro-business policies until the late 1980s at the earliest. Somehow Gandhi’s pro-Big-Business attitude is supposed to have helped business without being reflected in any major economic reforms.

Maybe it changed the way existing laws were enforced? Maybe it signalled to businesses that they should get started on long-term strategies because they could expect favorable laws in the future? Or maybe Prime Minister Gandhi just sort of sat in New Delhi, telepathically *willing* Big Business to succeed, and it worked? Rodrik and Subramanian are kind of agnostic about this. Everything else seems to fit pretty well, though. This might be a good time to reread Does Reality Drive Straight Lines On Graphs, Or Do Straight Lines On Graphs Drive Reality?

This is also a good time to reread The History Of The Fabian Society, because the problem might be their fault to begin with. In the waning days of the British Empire, bright young leaders from all over the developing world (including India's Jawaharlal Nehru) came to study at Oxford and Cambridge, got inducted into Fabian socialism, went home to their newly independent countries, and pursued socialist policies. All those countries did terribly and became the Third World basketcases of today. Only over the last few decades is the damage starting to be reversed.

If we had a better understanding of what exactly happened and how it was reversed, it could be an important source of information for developing countries in the future. Also, and more selfishly, it would be an important source of information for the US. Historically-informed anti-socialism arguments have tended to hinge on things like socialist China killing 60 million people, or socialist Russia killing 15 million people, or socialist Cambodia killing 1.5 million people, or [insert other socialist regimes killing 6-7 digit numbers of people]. But nobody thinks that Bernie Sanders plans to kill a six to seven digit number of people. To respond to Bernie-Sanders-style-socialism, we need to study and raise awareness of the history of democratic, comparatively "nice" countries that did nothing worse than overregulate business a bit – and investigate whether even these best-case scenarios still doomed millions of people to live in poverty. My (biased) guess is that careful study will show this to be true. But I don’t think this study has been done, I don’t think the facts are in yet, and I don’t think it was appropriate for the Charter Cities Institute to cite Panagariya’s argument on this point without any challenges or caveats.

Highlights From The Comments On PNSE

(original post)

Alex M writes:

> I think one of the main problems with the current state of rationalism (and many other fake “sciences” such as economics or sociology) is fuzzy thinking and lack of falsifiable empirical testing. So somebody claims to be “enlightened.” Does a smart person take that at face value? Of course not. Once you just start believing random shit, you’re no better than a superstitious primitive cargo-cult. You have to TEST all claims. For example, I don’t just take it at face value that economics is a real science just because a bunch of IYIs tell me so. I analyze economist predictions, see that their track record of successful predictions is atrocious, and then make the totally RATIONAL choice to discard my priors and treats economics as the laughable hocus-pocus that it is – because when you genuinely have an accurate view of reality, it doesn’t collapse under scrutiny. We should treat mystical claims exactly the same way. So somebody claims to be enlightened? Fine. How can they substantiate it? Can they do things that unenlightened people can’t, like clairvoyance, predicting the future, or sending messages through the collective subconscious in order to significantly impact world events? Do you see what I’m saying? Enlightenment should have some objectively quantifiable impact beyond just having a different internal narrative that is completely subjective and unprovable.

> This total lack of skepticism that people have is endlessly frustrating to me, because it results in bad data and popular narratives that are completely incorrect, if not outright delusional. In my opinion, the reason we have entire pseudo-scientific fields (like sociology or economics) that are nothing more than fake science cargo cults is because of this credulous behavior that results in people just believing whatever an “expert” with a fancy degree says. The fact that we have a replication crisis is a result of this gullible tendency to accept claims at face value. We are slowly learning to fix science by being more skeptical of expert claims, but we have to apply these same standards of falsifiability to spirituality as well, otherwise we are simply shifting our cult-like behavior from the field of science to the field of religion.

Imagine a doctor told you that repeated trauma can cause a long-lasting state of dysphoric depersonalization. In fact, you don’t need to imagine it – I am telling you now that repeated trauma can cause a long-lasting state of dysphoric depersonalization. How much effort should you put any effort into doubting this? If I say that my evidence is I know a few patients with trauma histories who say they’ve had long-lasting states of dysphoric depersonalization, and that most other doctors I talk to also know some patients, and a couple of small studies have been done on this and say the same thing, are you especially interested in doubting it?

Now imagine a doctor tells you that repeated meditation can cause a long-lasting state of *euphoric* depersonalization. Should our prior on this be any lower than the last statement? Should we reject the experience of thousands of people and dozens of studies because it’s *just too far out there?* Should we say that no rationalist should ever believe such a thing?

I don’t think the minimalist account of enlightenment takes us *quite* as minimal as “a long-lasting state of euphoric depersonalization”. But it takes us pretty close. And the evidence includes thousands of otherwise-trustworthy people who say they’ve had the experience, including some people I know personally and trust quite a bit, and who radically change their behavior afterwards (even if that change is just being impressed by the experience they devote the rest of their lives to exploring it). These are accompanied by many other people who haven’t gotten that far but report surprising and related-sounding experiences from the small amount of meditation they’ve done. And all of this is supported by brain scan results. And all of this meshes well with the evidence from philosophers like Dennett and Parfit that the “self” is a construct created by the brain rather than an objective reality, which itself meshes well with evidence that groups like schizophrenics and drug users can have disturbed senses of self or misplaced self/other boundaries.

This is really as much evidence as we have for any kind of mental state we haven’t experienced personally, and I’m pretty okay with it.

Bugmaster asks:

> If 50 people told you they were abducted by aliens, does it mean that there are aliens abducting people?

No. I also wouldn’t believe 50 people who said they had seen Bigfoot, or created perpetual motion machines. I would believe 50 people who said McDonalds had a new dessert on the menu, or that Biden was up in the polls today, or that they had the flu (assuming I hadn’t investigated these issues myself).

As Bayesians, we compute our belief by combining evidence with priors. Our evidence in all these cases is “fifty people believe something”. Our priors are either very low (in the first three examples) or reasonably high (in the last three).

As I mentioned above, my prior that there are euphoric depersonalized states of consciousness is pretty high, given that I know there are dysphoric depersonalized states of consciousness. It’s certainly not low enough that when thousands of people swear they have it, and most of the neuroscientists who look into it end up pretty convinced that it’s real, I’m going to say “Haha, no, you’re all lying”. For thousands of not-really-enlightened people to all falsely claim to be enlightened, describe enlightenment in similar ways, and go around teaching students who themselves later claim to be enlightened, all without giving up the game – sounds like an absurd conspiracy theory.

If I try to steelman the anti-enlightenment argument, the best I can do is to imagine a sort of placebo enlightenment, where if you get told by your culture that you can feel inner peace and selflessness by doing X, eventually you feel (something that can be mistaken for) inner peace and selflessness.

But if we’re going to worry about this, why don’t we believe that LSD only causes placebo hallucinations? Maybe drug culture talks up LSD hallucinations so much that users say they’ve seen them to fit in. Why don’t we believe chronic pain conditions are only placebo pain? Why don’t we worry that “runners high” and “endorphin rushes” are just stories runners tell to feel better for themselves after torturing their bodies over long marathons?

Part of the answer must surely be “because there’s no philosophical difference between pain and a ‘placebo pain’ which presents the same subjective experience as pain, given that pain is a subjective experience". The other part will depend on our priors about how often people have unusual experiences vs. how often people make things up. My prior is that people have really unusual experiences all the time, and that they make things up much less than doctors like to think. My OCD sometimes presents as a burning need to touch a random piece of furniture far away, felt through a sensory modality I cannot describe to anyone without the condition; my standing to accuse other people of making their unusual experiences up is pretty much nil.

I worry that all of this is being contaminated by associations with the word "enlightenment", where it represents something like "becoming a superhuman surrounded by glowy rays of light". If you just say "euphoric depersonalization experience" (or, like Martin, "persistant non-symbolic experience"), it doesn't seem to have as much cause for skepticism. I imagine people doubted the existence of Komodo dragons for a while based on generally-accurate anti-dragon priors, but if you just think of them as "big lizards" then the problem disappears.

Seppo has a more interesting anti-enlightenment steelman than my poor attempt:

> The most interesting enlightenment-skeptical thing I've ever heard is this interview ("Meditation: Deconstructing Nonsense" with Bill Joslin; h/t peach jam on David Chapman's blog).
>
> To summarize (from somewhat hazy memory, sorry):
>
> Joslin did various Buddhist and Taoist practices and had… *some kind of weird experience and/or insight.* I'm reluctant to throw the word "enlightenment" around

> since it’s defined in too many ways, but Joslin explains what his thing was in a good amount of detail. He also gives a lucid account of several different kinds of meditation practices that people do and why each of them would lead to something freaky happening to the self/other boundary.
>
> Afterwards, he started teaching meditation, probably without an official licence. (At least, he doesn’t mention having one; even ex-Buddhists who have such certifications will normally say so and name the person who gave it to them, if the topic ever comes up.)
>
> After having some doubts about the whole thing, he spent some time trying to talk about his Insights into the Nature of the Mind and/or Reality with ordinary Americans who had never been interested in Buddhism or meditation or the like—and a lot of them told him something like, “Oh, sure, I know all about that. I was walking in the woods one day and sorta… noticed?”
>
> Now he thinks that (1) whatever valid insights come with wild experiences like his are things that random people stumble into all the time without making a big fuss about it, and (2) the main function of mystical/meditation traditions is to get those insights to come at you in the form of *unnecessarily dramatic* experiences that they can then take credit for.

Remember, on the last SSC survey, 6% of respondents said they were enlightened. Sure, 4% of that is Lizardman’s Constant. What about the other 2%? Maybe some could be the walking-in-the-woods-one-day people Joslin talks about?

Skaladom writes:

> Since we’re using Buddhist terminology, it might be worth pointing out that in Buddhism only those at the far end of the spectrum would be called enlightened. What the paper calls Stage 1, in particular, sounds pretty close to what Theravada Buddhism calls stream entry, which is recognized as an important step to enlightenment. A more generic term used by spiritual seekers to encompass all these levels is “realized”.

I think this is a good amendment to the above points. Maybe “enlightenment” really *should* be restricted to superhumans surrounded by glowy rays of light. One thing that many descriptions of spiritual experiences share is a warning that the first rungs on the enlightenment ladder, the entry-level forms of enlightenment, are so inconceivably world-shattering that the people who attain them will think they’ve reached the ultimate possible state of being unless they have really strong reasons to think otherwise. Maybe random people walking in the forest and 6% of SSC readers have had early-stage awakening experiences, but there are also states beyond that.

Aella writes:

> Ok I’m about to do gossip: I know a highly experienced mediator, someone who’s worked with/around Jeffrey and I asked him about what he thought about the PSNE paper (which I liked a lot, particularly because I strongly related to having experienced the last location). He said that he thought Jeffrey was more interested in developing a good narrative (he makes a lot of money off this stuff) than actually trying to figure out patterns, and that the data is cherrypicked, and that he doesn’t really trust anything Jeffrey puts out.

> Of course take this with a grain of salt, is heresay, but I trust the meditator enough that now I personally feel wary of the PNSE paper as well.
>
> Also I’m really intrigued by the infighting that goes on in meditation communities, particularly around claims to enlightenment. I’ve talked at least one somewhat famous meditation teacher that is extremely sure another somewhat famous meditation teacher is definitely not enlightened. The lack of agreement in this area is so goddamn juicy and I really want to dig into it.

Cuke writes:

> Student of Buddhism for past 35 years. I don’t claim enlightenment for sure, but I have some lived experience with this terrain — repeated glimpses let’s say, on and off the cushion.
>
> The thing that makes me most skeptical about these reported findings is the discussion around stress and loved ones’ perceptions of the person.
>
> As one goes down this path, the person experiences significantly less emotional reactivity. That means both that they will experience less stress in response to things that used to stress them more and they will move more flexibly/kindly through the world than they previously did. Loved ones will notice this, without question. The person themselves will experience fewer days of tight chest anxiety, hot-faced anger, tension headaches, chronic muscle tension, insomnia due to worry, fear of death, conflict with others, etc.
>
> You can’t really compare one person to another that way because we all start at different places, but within a person, it seems dubious to me that a person could

claim to be moving along the road to enlightenment and still be manifesting the same level of stress response and emotional reactivity as they did before. This to me is a contradiction of what it means to hold one's experience more lightly, to cling less to ego, to be less identified with the self, to feel less attached to desire and aversion, and so on.

In my experience, the path is not reversible beyond a certain point. And there's good reason for that — the awareness gained in meditation, or indeed however the insight comes — permanently changes one's outlook about oneself and the nature of reality. You can't un-see it. Part of what that means too is that the awareness/tools/insights gained in meditation cannot be isolated to the cushion — they change fundamentally how you experience and move through your daily life, how you respond to stressors and to other people.

Anyway, that's been my experience and what I've witnessed in others. For all of us somewhere on the road but short of enlightenment, it doesn't mean the end of stress, the end of pettiness, the end of ego, the end of clinging. But you definitely definitely would expect to see change along those dimensions in a way that would be noticeable to other people.

What comes through Scott's review here is a particular focus on the experience of loss of self. Our language is imprecise in this arena. It's possible to experience "loss of self" in the midst of sex or trance or on LSD or in the woods or after long hours of devotional chanting or certain breath practices. It's possible to have moments of experiencing "loss of agency" in a way that feels very relieving. I don't think this should be confused with "enlightenment."

The path towards enlightenment is about “waking up” from cycles of attachment and aversion that we live in. I think it’s possible to have all kinds of transcendent experiences of “oneness” without making much progress on the waking up from attachment part. Those fleeting transcendent experiences may come and go in any person’s life. Waking up entails new levels of self-awareness that are not reversible (I can’t speak to the impact of dementia or brain injury) and that lead to deeper levels of compassion towards all beings.

I’m not familiar with what enlightenment means outside of Buddhist traditions, so part of the imprecision may be that people studying enlightenment experiences need to be clearer about how the definition varies from one tradition to another. I think it’s possible in the effort to study some phenomenon across traditions, that researchers are settling on a least common denominator that no longer resembles “enlightenment” as it’s understood within any one tradition…

For me personally, [gaining new levels of self-awareness] is like discovering a new room in my house. I may not always be able to go in and see the view from that room, but now that I know it’s there, I know that that view exists. Before that, I didn’t know that the room or the view existed. And now mostly, I know how to get back into that room and see the view. From that view, the things that used to cause me habitual suffering, no longer do. So even on days that I can’t get into the room, the things that used to cause me habitual suffering, don’t cause quite as much suffering.

[As for changes of outlook about the nature of reality] — I don’t have a short way to answer this question. I can tell you it’s not mainly a cognitive/philosophical shift,

> but it includes it. It’s a more visceral shift in terms of how I hold my moment-to-moment lived experience and how I view myself moving through it. All of it does indeed seem to be tied to what is described as the four noble truths — ie, that attachment causes suffering, that it’s possible to get freer from attachment, and that practicing that does free one up from suffering.

This is what I mean when I talk about the sheer number of testimonials. It’s not just some guy in Tibet trying to sell you a book. Whenever I talk about it, there are smart, normal-seeming people who pipe up and say they have “some experience with the terrain”. I think if you’re going to doubt all these people, you need some theory of what’s going on, something more explanatory than just “we’re rationalists so we don’t believe any of this stuff”.

Financial Incentives Are Weaker Than Social Incentives But Very Important Anyway

NYT: Economic Incentives Don’t Always Do What We Want Them To (h/t MR). For the first time in history, the title actually understates the article, which argues that incentives can be surprisingly useless:

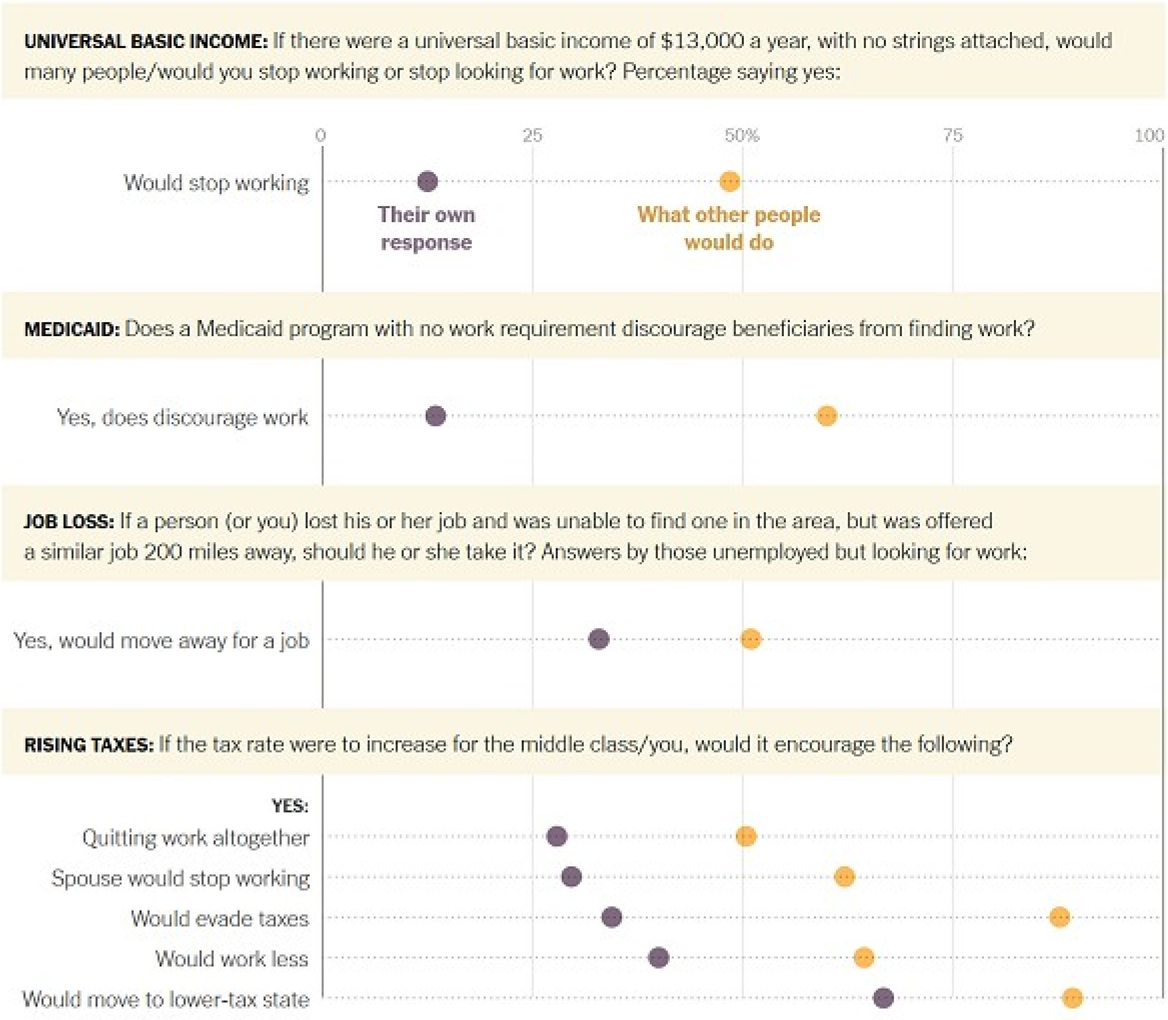
> Economists have somehow managed to hide in plain sight an enormously consequential finding from their research: Financial incentives are nowhere near as powerful as they are usually assumed to be.

The article starts with some surprising facts. Increased taxes on the rich don’t make rich people work much less. Salary caps on athletes don’t decrease athletic performance. Increased welfare doesn’t make poor people work less. Decreased job opportunities in one area rarely cause people to move elsewhere.

Then it presents a neat chart showing that most people believe others would respond to an incentive, but deny responding to that incentive themselves. For example, 60% of people say a Medicaid program with no work requirement would prevent many people from seeking work, but only 10% of people say they themselves would stop seeking work with such a program.

People Think Financial Incentives Work. But for Themselves? Not So Interested.

A survey asked one group how the population at large would react to several financial inducements. Another group was asked about their own reaction; these people were much less inclined to do any of the following, despite the same financial incentives.



UNIVERSAL BASIC INCOME: If there were a universal basic income of $13,000 a year, with no strings attached, would many people/would you stop working or stop looking for work? Percentage saying yes:

MEDICAID: Does a Medicaid program with no work requirement discourage beneficiaries from finding work?

JOB LOSS: If a person (or you) lost his or her job and was unable to find one in the area, but was offered a similar job 200 miles away, should he or she take it? Answers by those unemployed but looking for work:

RISING TAXES: If the tax rate were to increase for the middle class/you, would it encourage the following?

0
25
50%
75
100
Would stop working
Their own response
What other people would do
Yes, does discourage work
Yes, would move away for a job
YES:
Quitting work altogether
Spouse would stop working
Would evade taxes
Would work less
Would move to lower-tax state

By The New York Times | Source: Abhijit Banerjee and Esther Duflo, M.I.T., and Stefanie Stantcheva, Harvard; data from a nationally representative survey of 10,000 American adults conducted by the study's authors.

…but keep in mind an alternate interpretation would be "desirability bias makes people deny they would work less and evade taxes"

All this suggests that:

> If it is not financial incentives, what else might people care about? The answer is something we know in our guts: status, dignity, social connections. Chief executives and top athletes are driven by the desire to win and be the best. The poor will walk away from social benefits if they come with being treated like a criminal. And among the middle class, the fear of losing their sense of who they are and their status in the local community can be an extraordinarily paralyzing force.

They conclude that this argues in favor of policies like raising taxes on the rich and removing all requirements from welfare programs.

The authors are Nobel Prize winning economists, so I assume they're basically right. And I'm not up to doing a complicated literature review to compare all the cases where economic incentives do work to the cases where they don't and develop a well-informed understanding of the subtleties in their position. So instead, a few low-effort thoughts.

First, it matters less whether the *average* person responds to economic incentives, and more whether the *marginal* person will. If I need someone to cover the graveyard shift at work, nobody will do it for normal pay, and I offer double pay, all I need is for one employee to be incentive-sensitive enough to take me up on it. Maybe most people wouldn't accept any amount of money to become an oil rig worker, a McKinsey consultant, or a camgirl, but ExxonMobil/McKinsey/MyFreeCams.com only need just enough qualified people to accept whatever deal they're offering.

Likewise, perhaps if I had no alarm system protecting my house, 99.999% of people still wouldn't rob me. But

99.999% of people not robbing you is still known as “getting robbed”.

So “most people don’t respond to most economic incentives” is totally compatible with “economic incentives rule the world and control everything around us.”

Second, grant that most people care primarily about “status, dignity, [and] social connections”. A lot of how that works out in real life is “doing the socially acceptable thing”. Even if incentives are weak in the short term, they can be very strong in the long term after they have time to act on what is or isn’t socially acceptable.

It’s all nice and good to say “most people wouldn’t steal even in the absence of punishment”. But what about music piracy? Nobody had any way to enforce rules against pirating music. Maybe only a few people pirated at first. But then more and more people did it, and eventually the unwritten rule among teenagers became that music piracy was okay – in fact, that you were weird if you *didn’t* do it. On the other hand, stealing a CD from a record store still feels horrifying and criminal and inconceivable. Although there are subtle differences between the two cases (it costs nonzero money to make a physical CD) I still think a lot of this is social norms that formed downstream of enforcement-related incentives.

Or: most people would never cheat on welfare. But there are Alabama counties where over 25% of the population are on disability, an increase of 50% from just fifteen years earlier. I don’t want to accuse any of them of *cheating*, per se, and see here for a more in-depth analysis. But I think it’s easy to normalize taking disability for lesser and lesser afflictions, and that part of the normalization process involves an economic incentive to do it and a lack of incentive not to.

Or: in Sierra Leone, 84% of people say they have paid bribes; in Japan, 1% have. So do “people” care about financial incentives or not? Grant that “status, dignity, [and] social connections” are more important, and that this is what prevents bribery in Japan. But once these factors permit bribery, it becomes rampant. And are these factors themselves maintained partly by incentives, eg punishments upon being caught? I’m not sure.

Third, remember that principles are usually downstream of politics. So one fun game is to take a principle usually used on one side of the political spectrum, then apply it in support of the opposite side and see if you still hold it.

So. We know there’s no reason not to raise taxes, since rich people don’t respond to financial incentives. But there’s also no reason to close tax loopholes – rich people defrauding the government of money through tax evasion is surely as unthinkable as poor people defrauding the government through welfare scams. And there’s no reason to question the bonuses of Wall Street traders, since it’s not like anything as crass as a financial incentive would cause them to make risky trades.

Did pharmaceutical companies incentivize opioid overuse through paying doctors to overprescribe? Doesn’t matter, doctors would never let financial incentives affect their prescribing decisions. Are senators cozying up to companies that will give them lucrative sinecures later in a “revolving-door” system of legal bribery? No, because incentives aren’t powerful enough to make senators abandon their dignity. Are billionaires destroying the environment just to make a buck? No, the financial incentives to do so wouldn’t outweigh the cost in status and social connections.

None of this snark disproves the real empirical research the authors use to show that rich people's taxes, poor people's welfare use, and economic mobility are not very incentive-sensitive. But I hope they prevent people from generalizing to a general sense that financial incentives don't matter, or turning this into a purely partisan issue where anyone who believes in financial incentives at all gets accused of "dog whistling" conservativism.

Fourth, and most important, the more we're ruled by social incentives, the more importance financial incentives take on as a counterweight. Quoting my favorite part of the article again:

> If it is not financial incentives, what else might people care about? The answer is something we know in our guts: status, dignity, social connections. Chief executives and top athletes are driven by the desire to win and be the best. The poor will walk away from social benefits if they come with being treated like a criminal. And among the middle class, the fear of losing their sense of who they are and their status in the local community can be an extraordinarily paralyzing force.

I think this is profoundly true, so true that it's almost impossible to appreciate enough. The article frames it positively – we care about community more than money, how heartwarming. But I find it disquieting – it could equally be framed "We care more about fitting in and not seeming weird than about anything else in the world". 99% of world-changing ideas are stillborn when their would-be-inventor worries they might sound weird for proposing them. 99% of great companies don't get off the ground because their would-be-founder worries about what other people would think. The most important ideas for changing government and society sit on the lunatic fringe, because everyone worries that supporting such ideas might keep them out of the Inner Ring.

Paradoxically, I think this argues in favor of financial incentives. The beauty of financial incentives is that they provide a counterbalance to status incentives. The counterbalance is weak, inconsistent, blink-and-you-miss-it, but it is real. If all the cool people say “we do it this way”, 99% of people will do it that way to fit in, but there will be one person who does it the much better way that lets them outcompete everyone else and make $10 billion. And having $10 billion brings “status, dignity, [and] social connections” of its own. Even if only a tiny number of people are sensitive to money, it’s enough to create a core who occasionally try making things better even when that’s not cool.

One corollary of this is that when you remove financial incentives, you don’t get everyone acting ethically for the good of all. You just get status incentives with no counterbalance. I can think of few things scarier.

New Atheism: The Godlessness That Failed

Thucydides predicted that future generations would underestimate the power of Sparta. It built no great temples, left no magnificent ruins. Absent any tangible signs of the sway it once held, memories of its past importance would sound like ridiculous exaggerations.

This is how I feel about New Atheism.

If I were to describe the power of New Atheism over online discourse to a teenager, they would never believe me. Why should they? Other intellectual movements have left indelible marks in the culture; the heyday of hippiedom may be long gone, but time travelers visiting 1969 would not be surprised by the extent of Woodstock. But I imagine the same travelers visiting 2005, logging on to the Internet, and *holy @#$! that's a lot of atheism-related discourse what is going on here?*

My first forays onto the Internet were online bulletin boards about computer games. They would have a lot of little forums about various aspects of the games, plus two off-topic forums. One for discussion of atheism vs. religion. And the other for everything else. This was a common structure for websites in those days. You had to do it, or the atheism vs. religion discussions would take over everything. At the time, this seemed perfectly normal.

In 2005, a college student made a webpage called The Church Of The Flying Spaghetti Monster. It was a joke based on the idea that there was no more scientific evidence for God or creationism than for belief in a flying spaghetti monster. The monster's website received tens of millions of visitors, 60,000 emails ("about 95 percent" supportive), and was covered in *The New York Times*, *The Washington Post*, and *The Daily Telegraph*. Six publishing companies entered a bidding war for the rights to the spaghetti monster's "gospel", with the winner, Random House, offering an $80,000 advance. The book was published to massive fanfare, sold over 100,000 copies, and was translated into multiple languages. Putin's thugs broke up a pro-Flying-Spaghetti-Monster demonstration in Russia. At the time, this seemed perfectly normal.

People compiled endless lists of arguments and counterarguments for or against atheism. The Talk.Origins newsgroup created a Dewey-Decimal-system-esque index of almost a thousand creationist arguments, from CA211.1 ("Karl Popper said that Darwinism is not testable"), to CD011.1 ("Variable C-14/C-12 ratio invalidates carbon dating"), through CH508 ("Chinese treasure ships show Noah's Ark was feasible") – and painstakingly debunked all of them; in case that wasn't enough they linked 133 other sites doing similar work. Their arch-enemies, creationist site True.Origin, then went through and debunked all of their debunkings. Another atheist group created the Skeptics' Annotated Bible, a version of the Bible highlighting everything bad or wrong in it. For example, if for some reason you need a hit job on the second chapter of the Book of Malachi in particular, you can look up its SAB page and find that Malachi 2:11 castigates Judah for "marrying the daughter of a strange god" (which is intolerant), Malachi 2:17 accuses the Israelites of "wearying the Lord with your words", (which is absurd since God cannot be wearied), and Malachi 2:3 says that God will spread dung upon the faces of unbelievers (which is gross). This last entry includes a link to a 2007 YouTube video "God Wants To Smear Dung On Your Face" with 21,947 views. And the video links to a store selling Malachi-2:3-says-God-wants to-put-dung-on-your-face-related t-shirts, bumper stickers, keychains, and coffee mugs. At the time, this seemed perfectly normal.

Whatever media you liked, there were atheism-themed versions of it. Obviously if you liked webcomics you would never be able to finish all the different atheist options from Russell's Teapot through Jesus & Mo through The Sheeples. If you liked TV, there were atheist TV shows like John Safran vs. God or The Atheist Experience. If you liked pithy quotes, you could read the top 10,000 atheist quotations in order of popularity. If you just liked discussion, you could go to the now-infamous r/atheism subreddit, which at the time was one of Reddit's highest-ranked, beating topics like "news", "humor", and – somehow – "sex". At the time, this seemed perfectly normal.

But these still don't quite make my point, because the defining feature of this period wasn't just that there were a lot of atheism-focused things. It was how the religious-vs-atheist conflict subtly bled into everything. Read enough old articles and blogs from this period and you'll spot it. Some travel writer going on about how the boring small town he ended up in is probably full of fundies who hate gays and think the Earth is six thousand years old. Some logician giving an example of circular arguments: "I know the Bible is true because it says so in the Bible." Some political writer saying a stupid policy is only to be expected in a country where X% of people still get their ethics from Bronze Age superstitions. At the time, this seemed perfectly normal.

It seemed perfectly normal because religion vs. atheism was the most important issue, maybe the only issue. How could you run a 21st century democracy with half the population believing in science and compassion, and the other half believing whatever they read in a 3000 year old book about a magic sky father? To truly understand the spirit of the time, you can’t just think of religion as evil. You have to think of it as the ur-evil, without which no other evil would exist. Homophobia? Only there because the Bible says to stone gay people. War? It’s all holy war of one sort or another, whether it’s Arabs vs. Israelis, Sunnis vs. Shias, or the Christian/Muslim “clash of civilizations”. Environmental devastation? Only there because religious people believe God elevated Adam over the animals and told him to exploit them for his own purposes. Poverty? Only because religious people believe in the prosperity gospel that says people get what they deserve.

Christopher Hitchens, 2008:

> Now, I am absolutely convinced that the main source of hatred in the world is religion and organized religion. Absolutely convinced of that. And I think it should be—religion—treated with ridicule, hatred, and contempt. And I claim that right. So when I say—as the subtitle of my book—that I think religion poisons everything, I’m not just doing what publishers like and coming up with a provocative subtitle. I mean to say it infects us in our most basic integrity. It says we can’t be moral without ‘Big Brother,’ without a totalitarian permission, means we can’t be good to one another without this, we must be afraid, we must also be forced to love someone whom we fear—the essence of sadomasochism, the essence of abjection, the essence of the master-slave relationship and that knows that death is coming and can’t wait to bring it on. I say that is evil, and though I do, some nights, stay home, I enjoy more the nights when I go out and fight against this ultimate wickedness and ultimate stupidity.

Where did this come from? And where did it go?

At the time, the question of where it came from seemed to have an obvious answer. As a civilization becomes advanced enough that some people throw off the yoke of religion, they will naturally come into conflict with people who have not thrown off that yoke. This will dominate discussion since atheism vs. religion is obviously the most important issue and maybe the only issue, and last until the civilization advances enough that religion disappears.

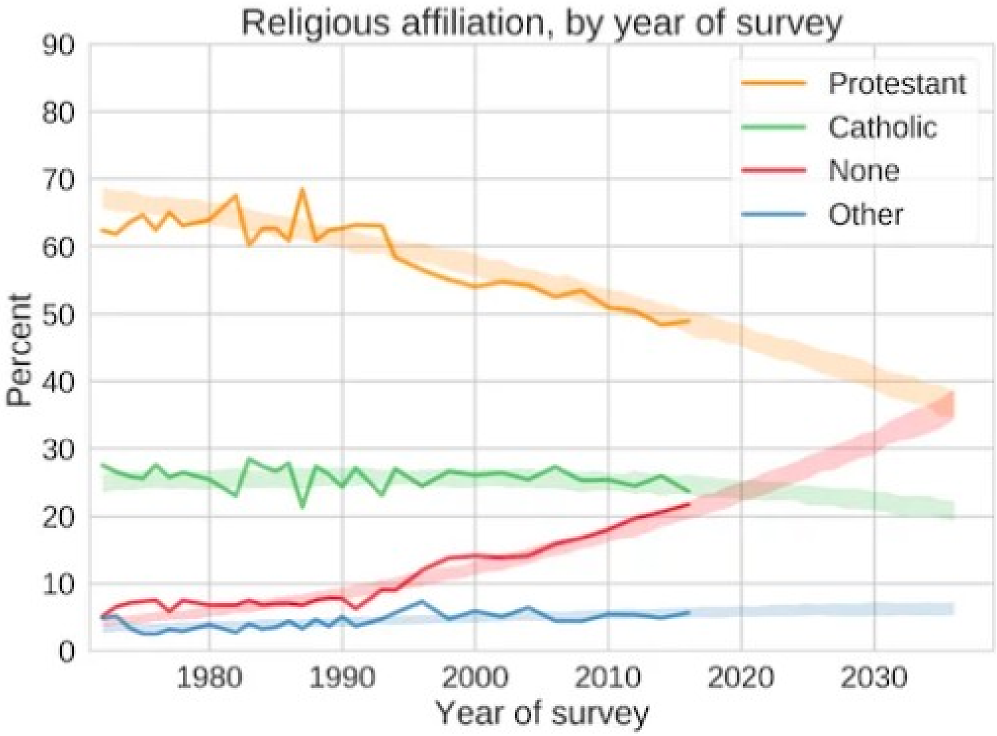
But the past decade or so has shown that advanced civilizations are perfectly capable of containing atheists and religious people in close proximity without either side caring that much about it. So what made the turn of the millennium such an acrimonious period?

As for where it went, I asked that question last year and got various responses. The most popular was that 9/11 made religion-bashing segue into Islam-bashing, which started to look pretty racist. But 9/11 happened in 2001, *The God Delusion* wasn't published until 2006, and New Atheism didn't peak until the early 2010s. Why?

In order to answer these questions, I'll start by presenting some data confirming the picture I paint above and trying to pinpoint exactly when the peak and the beginning of the end happened. I'll move on to some of the intellectual subtrends in New Atheism that might explain the picture a little better. And finally, I'll present my theory explaining the mysteries above: New Atheism was a failed hamartiology.

II.

Here is a graph of US religiosity over time:

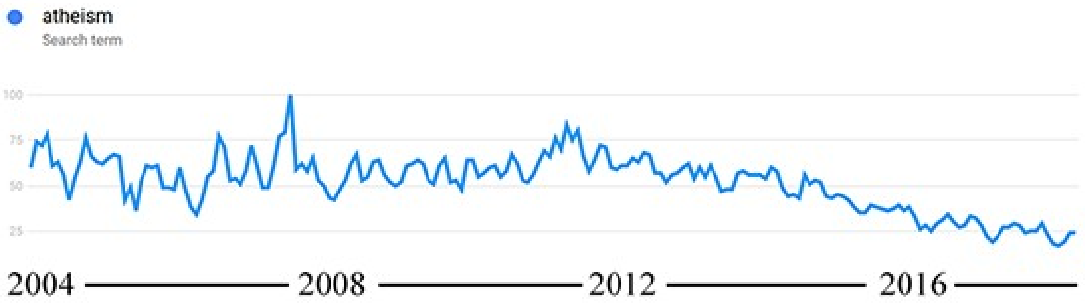


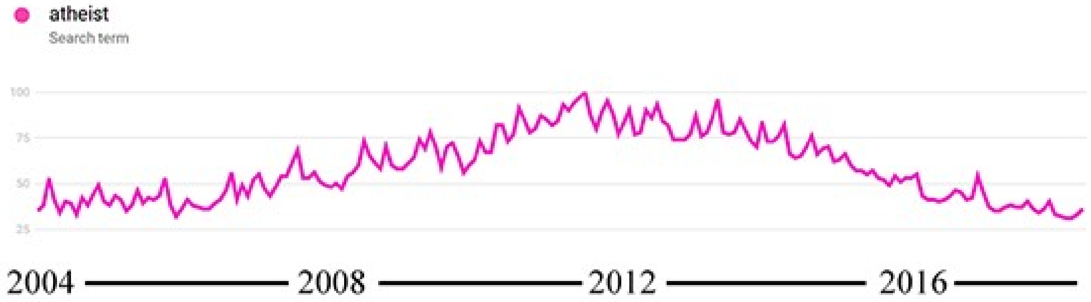
Between the first stirrings of internet atheism in 2000 and the beginning of the end in 2015, the percent of Americans identifying as Christian dropped about 10%; the percent identifying as no religion increased about the same amount. There are many different ways of looking at the data: self-reported affiliation, church attendance, even polls on whether religion can answer all of today's problems, but they all show the same story of slow, steady decline.

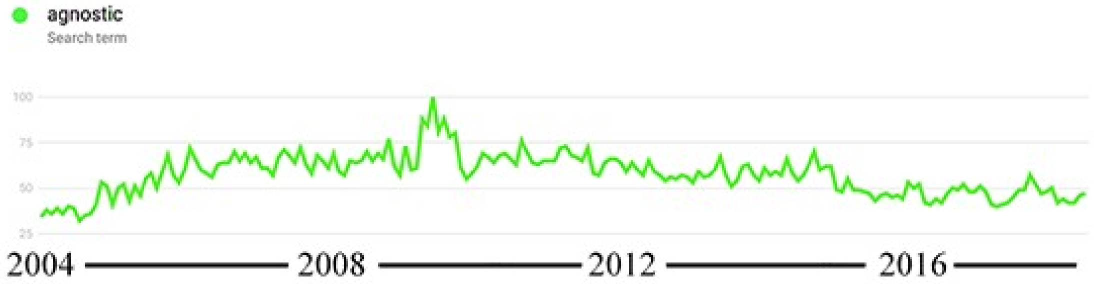
By the numbers, the decline is slight: from 80% Christian / 15% atheist in 2000 to 70% Christian / 25% atheist in 2018. This could hide wider social changes. The number of gay people has barely changed since 2000, but society's attitude toward them has totally transformed. Likewise, although religion has barely declined, and nonbelief barely risen, Christianity no longer seems to command quite the same level of political power, nor does atheism provoke quite as much revulsion.

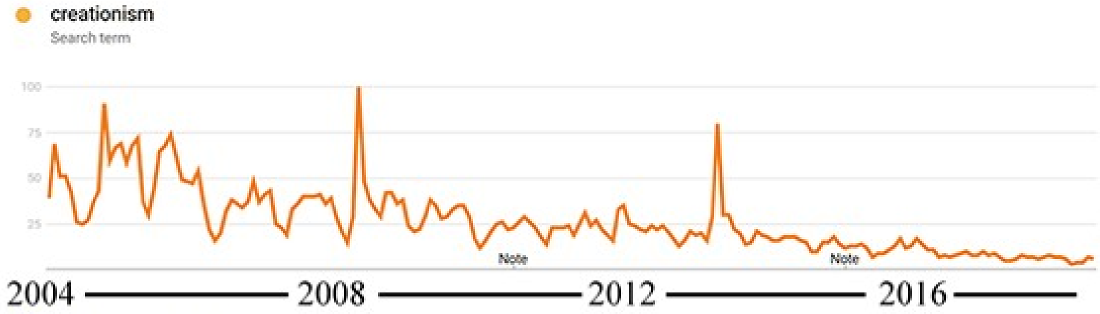
But the sudden fall of New Atheism didn’t *feel like* a process of gradual social change and eventual acceptance. It felt like a movement certain of its own victory burning out spectacularly over the course of a few short years, followed by mysterious yet near-total contempt from the very people it thought it had convinced.

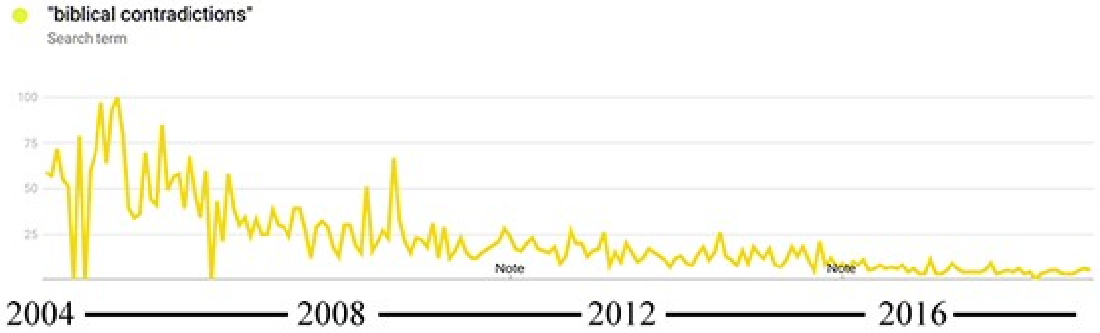
Here are some graphs of atheism-related search terms on Google Trends since 2004:


atheism
Search term
100
75
50
25
2004
2008
2012
2016

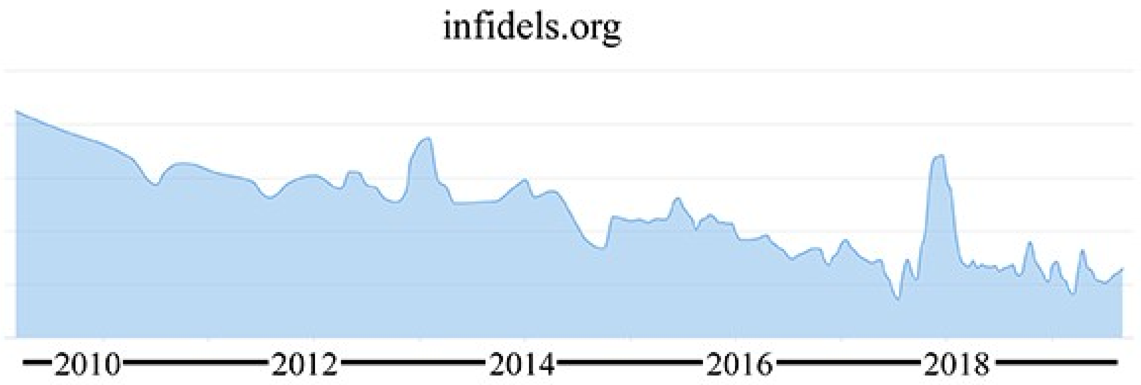

atheist
Search term
100
75
50
25
2004
2008
2012
2016

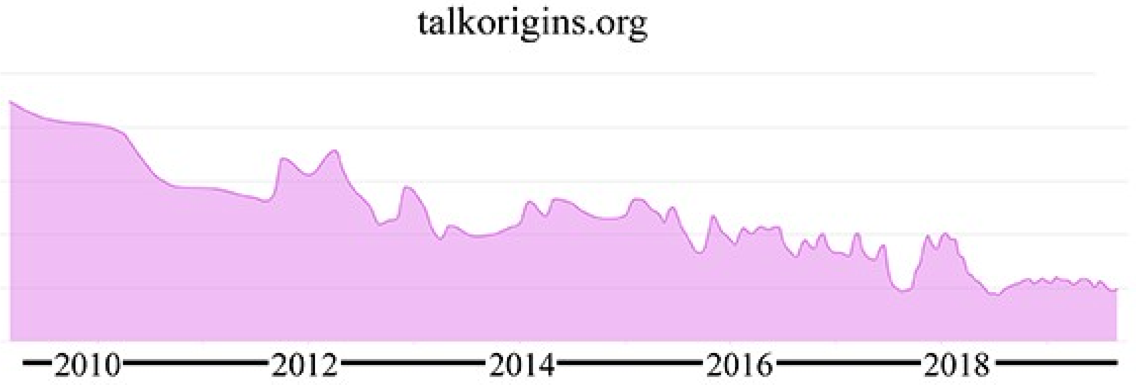

agnostic
Search term
100
75
50
25
2004
2008
2012
2016

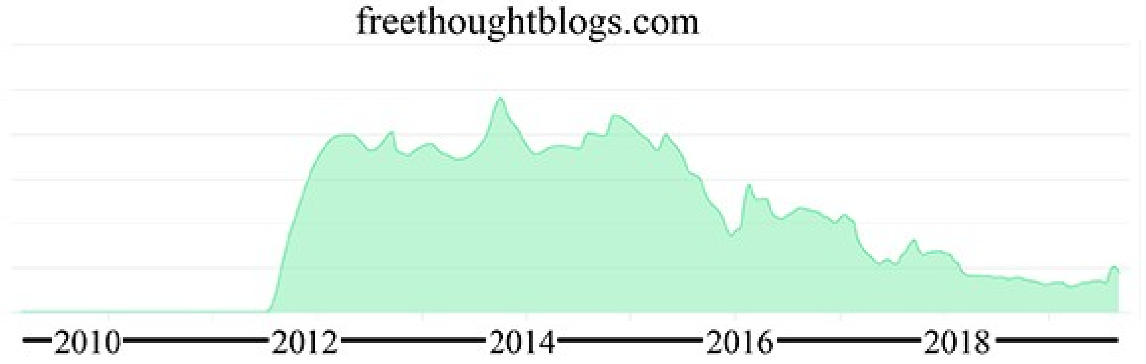

creationism
Search term
100
75
50
25
Note
Note
2004
2008
2012
2016

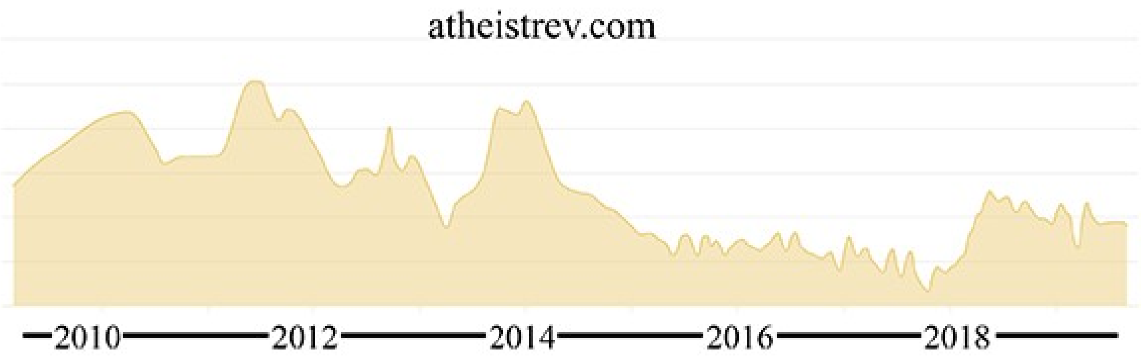

"biblical contradictions"
Search term
100
75
50
25
Note
Note
2004
2008
2012
2016

And here are the traffic numbers for some atheism-related websites (source: http://rank2traffic.com/):

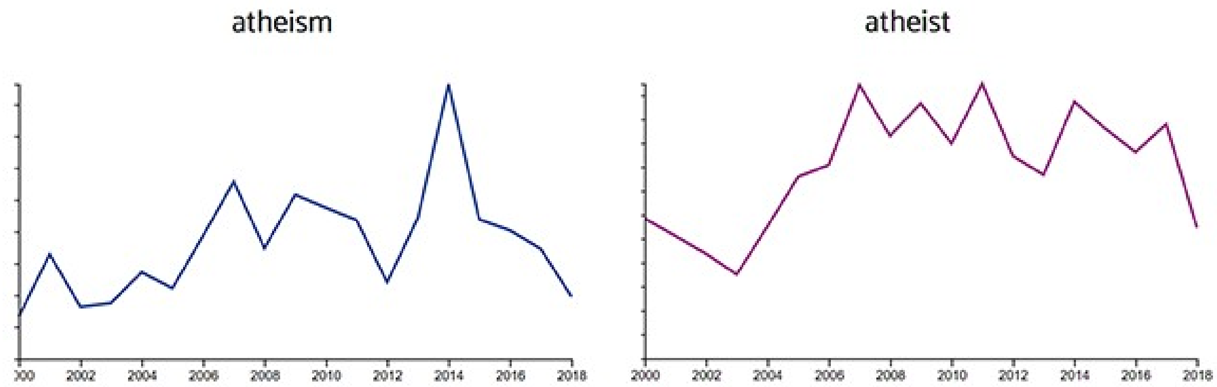




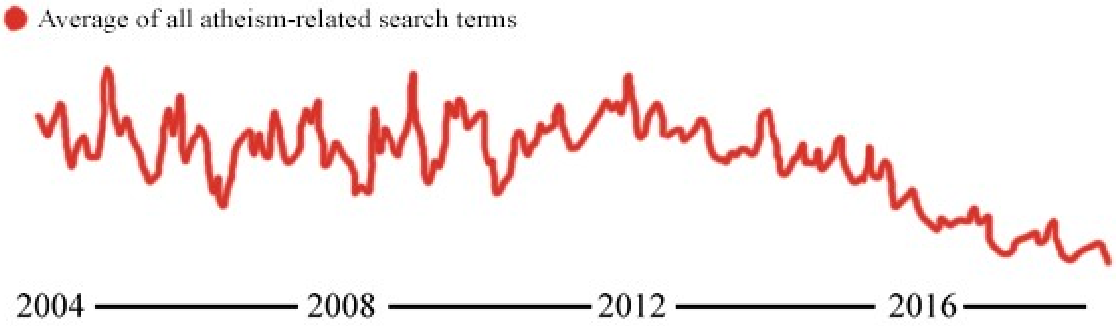




And it may not be Internet atheism per se, but here’s word frequency in the *New York Times* (source: New York Times Media Analytics):



I can’t figure out how to average the traffic numbers or the NYT frequencies, but here’s an average of all the atheism-related search terms:



I think these graphs mostly tell the same story. Unlike the continuous trend in religiosity, the atheist movement appears to be going strong throughout the 2000s, peak in 2012, and start declining shortly afterward.

But this hides a division into two different patterns. Two keywords (“creationism” and “Biblical contradictions”) and two websites (Talk Origins and Internet Infidels) are declining throughout the time period measured. Three keywords (“atheist”, “agnostic”, “freethinker”), two websites (Freethought Blogs and Atheist Revolution), and the *New York Times* frequencies are increasing through most of the period, peak around 2012, stay strong for a few years after that, and decline around 2016.

To get an intuitive feel for the first category, look at the two sites involved. Talk Origins is almost perfectly preserved, a time capsule from an era when people *really* wanted to debate creationism. Internet Infidels has decayed a bit more, but even its ruins are impressive: a database of forty videotaped atheist-vs-theist debates, an online library of uploaded works by about two hundred atheist authors, and the obligatory list of several hundred Biblical contradictions. Who does that these days?

This exercise is gradually bringing back memories of just how *intellectual* the Internet was around the turn of the millennium. You would go to bulletin boards, have long and acrimonious debates over whether or not the Gospels were based on pagan myths. Then someone would check Vast Apologetics Library tektonics.org and repost every one of their twenty-eight different articles about all the pagan myths the Gospels weren't based on, from Adonis ("yet another unprofitable proposition for the copycat theorist") to Zalmoxis ("there is no comparison, other than by illicit collapsing of terminology and by unsubstantiated speculation"). Both sides had these vast pre-built armories full of facts and arguments to go to.

At some point, in a way unrelated to the fall of New Atheism, the Internet stopped being like this. The topics that interest people today don't get debated in the same way. People dunk on each other on Twitter, occasionally even have back-and-forth exchanges, but the average person doesn't post long screeds and get equally long responses fisking each of their points. There's less need for giant databases containing every fact you might need to win a particular argument, organized Dewey-Decimal-style by which argument you are trying to win. People just stopped caring.

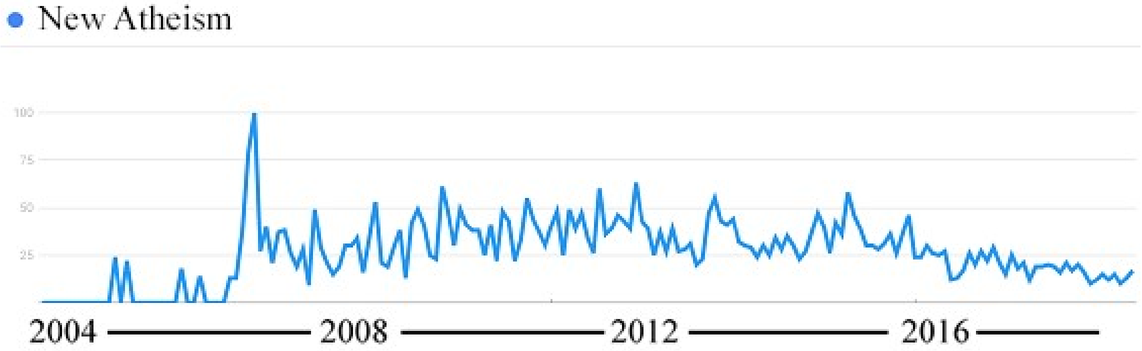
I’m not sure why this happened. Maybe it took about ten years from the founding of the Internet for people to really internalize that online arguments didn’t change minds. The first Internet pioneers, starting their dial-up modems and running headfirst into people outside their filter bubbles, must have been so excited. For the first time in human history, people interested in debating a subject could do so 24-7 out in a joint salon-panopticon with all of the information of the human race at their fingertips. Bible Belt churchgoers for whom atheists had been an almost-fictional bogeyman, and New York atheists who thought of the religious as unsophisticated yokels, came together for the first time thinking “Convincing these people is going to be *so* easy”. The decade or so before they figured out that it wasn’t was a magical time, of which the great argument-arsenals of the past are almost the only remaining monument.

Or maybe it was something else. Maybe it was that getting online was actually pretty hard in those days, you needed to be technically inclined or attending a college or both, and so netizens were just more educated. Maybe the sort of people who interrupt any attempt at intellectual discussion with words like “rationalbro” or “mansplaining” or “well acktually” were still stuck in their caves, fruitlessly banging AOL CDs against rocks trying to create fire. Maybe it was something as simple as Wikipedia not existing yet, leaving the intellectual world in a sort of state of nature with every man for himself. Maybe it was just that the bulletin board format was more conducive to this than the later social media style fora.

Whatever it was, the decline of this culture started no later than 2000, and is reflected in the fate of argument-related search terms like “biblical contradictions” and “creationism”, and in the fading of the great argument-armories like Talk Origins and Infidels.

But the “atheism” search term keeps rising for another decade. What happened?

The intellectuals were succeeded by the activists. Early Internet Argument Culture disappeared and was replaced by something more familiar.



The atheists of Early Internet Argument Culture were not New Atheists. The term “New Atheism” didn’t really catch on until about 2006 when Richard Dawkins published *The God Delusion*; Early Internet Argument Culture was just a prelude to the main event. Post-2006 atheists were brasher and more political. They were less interested in arguing with religious people about the minutiae of carbon-dating; they were more interested in posting about how stupid carbon-dating denalists were, on their own social media feeds, read entirely by other atheists. The concept of the Internet as magical place where you could change other people’s minds had given way to the Internet as magical place where you could complain to like-minded friends about how ignorant other people were.

EIAC had been timeless, examining the medieval *kalam* argument and the Scopes Monkey Trial with equal detachment. New Atheism was ephemeral, obsessed with the issue of the day. This was in the mid-Bush administration, after the post-9/11 spirit of national unity had disappeared. Democrats had not yet invented the hashtag #Resistance, but they had invented the spirit. George W. Bush was portrayed as a religious fanatic, basing his every decision on what he considered to be the will of God. His supporters were evangelicals, willing to follow him into any war or disaster out of blind faith. A lot of the debate centered on faith-based charities, Bush's push to give government funding to religiously-affiliated groups like the Salvation Army. It was assumed that they would preferentially serve Christians, leaving Jews, Muslims, and atheists without aid. Once Bush had shifted all welfare into these programs, non-Christians would die in the cold, and the government would laugh evilly. Every day brought new perspectives on this and a host of similar anti-religious activist causes.

New Atheism was also more centralized. EIAC was every man for himself; you would march forth alone into your chosen bulletin board and engage, neither seeking or receiving any help beyond precooked arguments from your local armory-site. New Atheism, for the first time, started to have celebrities. Richard Dawkins, of course, and the Four Horsemen, but also random bloggers like PZ Myers and Stephanie Zvan. These were the days when bloggers filled auditoria and travelled in high-altitude balloons. Every day they would tell you the latest reason to be outraged about religion, and every day you would discuss it on social media and comment sections and get appropriately angry.

This corresponds to the peak of Freethought Blogs on the traffic graph above, and ended around 2016. What happened to it?

I think it seamlessly merged into the modern social justice movement.

This probably comes as a surprise, seeing as how everyone else talks about how atheists are heavily affiliated with the modern *anti*-social justice movement. I think that's the wrong takeaway. Sure, a lot of people who identify as atheists now are pretty critical of social justice. That's because the only people remaining in the atheist movement are the people who didn't participate in the mass transformation into social justice. It is no contradiction to say both "Most of the pagans you see around these days are really opposed to Christianity" and "What ever happened to all the pagans there used to be? They all became Christian."

Somebody should make this case more exhaustively, but the highlight will no doubt be all the discussion around Atheism Plus, the brand name for a combination of "atheism plus social justice" which in a few years became entirely social justice. According to the original manifesto:

> We are…
> Atheists plus we care about social justice,
> Atheists plus we support women's rights,
> Atheists plus we protest racism,
> Atheists plus we fight homophobia and transphobia,
> Atheists plus we use critical thinking and skepticism…
>
> Religion is responsible for generating and sustaining most of the racism, sexism, anti-(insert minority human subgroup here)-isms… it gave a voice to the bigotry, established the privilege, and fed these things from the pulpit for thousands upon thousands of years. What sense does it make to throw out the garbage bag of religion yet keep all the garbage that it contained? I can't help but see social justice as a logical consequence of atheism. I'm for getting rid of all the garbage.

Within a week, it got glowing articles in the mainstream press, from New Statesman to Salon to The Guardian (consider how weird it sounds today for a post by a mid-tier atheist blog post to result in a bunch of mainstream press articles) and support from top atheist blogging celebrities . A review a week later wrote:

> Last week, Jen McCreight announced that she was fed up with sexism in the atheist movement and called for a new wave of atheist activism, one explicitly concerned with social justice, which quickly acquired the name "atheism+".
>
> These posts landed like a cannon shell, generating a huge wave of excitement and feedback – the vast majority of which, to my surprise, was positive and enthusiastic. Clearly, they've tapped into a powerful vein of pro-equality sentiment in the atheist movement, crystallizing the frustrations that those of us who care about this have been feeling for the last year or two. This is an idea whose time has come, and all it needed were some excellent posts like Jen's to kickstart it.

Famous atheist blogger PZ Myers embraced the new label and said that "atheism ought to be a progressive social movement in addition to being a scientific and philosophical position" and that:

> If you don't agree with any of that — and this is the only 'divisive' part — then you're an asshole. I suggest you form your own label, "Asshole Atheists" and own it, proudly. I promise not to resent it or cry about joining it.

Richard Carrier, an academic and another of the most famous New Atheists, told atheists who objected to the rebranding that:

> Atheism+ is our movement. We will not consider you a part of it, we will not work with you, we will not befriend you. We will heretofore denounce you as the irrational or immoral scum you are (if such you are). If you reject these values, then you are no longer one of us. And we will now say so, publicly and repeatedly. You are hereby disowned.

I don’t want to dwell on this too much. I don’t have a great sense of how this era went, since it was around the time I unfollowed every atheist blog and forum for the sake of my own sanity, but my impression is that some of the Atheism Plussers later admitted they came on a little too strong and dropped that particular branding. But the cleavage the incident highlighted (not created, but highlighted) stuck around. As far as I can tell, it eventually ended with the anti-social-justice atheists stomping off to YouTube or somewhere horrible like that, while most of the important celebrity members of the public-facing movement very gradually turned into social justice bloggers.

For example, I look at Pharyngula, which during its heyday was the biggest atheist blog on the Internet. On the day I am writing this, its front page contains posts like “Are They All Racists On The Right Side Of The Aisle?” (recommended answer: yes), a discussion of how opposing the Gilette commercial represents “classic toxic masculinity”, and an attack on Milo Yiannopoulos. Its sidebar includes links to “Discussion: Racism In America”, “Discussion: Through A Feminist Lens”, and “Social Justice Links Roundup”. There’s still a little bit of anti-religious content, but mostly in the context of Catholics being racist and misogynist.

Aside from Pharyngula, a lot of the old atheist blogs have ended up at atheism-blogging-mega-nexus-site The Orbit.

When I read its About page, it doesn’t even describe itself as an atheist blogging site at all. It says:

> The Orbit is a diverse collective of atheist and nonreligious bloggers committed to social justice, within and outside the secular community. We provide a platform for writing, discussion, activism, collaboration, and community.

It’s not “blogs on atheism” anymore. It’s “blogs by atheists about social justice”. The whole atheist movement is like this.

One post I distinctly remember, but which I can no longer find, was a rousing call for atheists to switch to social justice blogging. It said something like “Instead of rehearsing the same old tired arguments for or against the existence of God, it’s time to become part of the struggle for progress and equality.”

I wish I could find this, because the sentiment it expresses is so bizarre that I worry you won’t believe me when I say it exists. Like, yes, the arguments for and against the existence of God *are* old and tired. Just like, for example, the arguments for and against restrictions on abortion. But if one day all of the top pro-choice activists agreed among themselves that what the pro-choice movement was *really* about was stopping Brexit – and they all posted supportive messages like “We’re tired of being known as those boring busybodies who go on about fetus this and right-to-your-own-body that when millions of people could be harmed by Britain’s ill-advised and bungled exit from the European Union” – and if from that day forward NARAL and Planned Parenthood were 100% Brexit-related organizations – surely we would find it strange? Surely we would think something deeper had to be going on?

I think of this as the second part of the mystery around New Atheism's decline: why did a successful social movement so quietly and complacently agree to turn into a totally different social movement?

III.

My solution to both these questions is: New Atheism was a failed hamartiology.

"Hamartiology" is a subfield of theology dealing with the study of sin, in particular, how sin enters the universe. Orthodox Christian hamartiology says we all have original sin because Adam and Eve ate the apple. Gnostic hamartiologies say we sin because we are ignorant of our true nature as celestial beings. Some heretical hamartiologies say that all of this is irrelevant, and we sin because we choose to.

The rise of the Internet broadened our intellectual horizons. We got access to a whole new world of people with totally different standards, norms, and ideologies opposed to our own. When the Internet was small and confined to an optimistic group of technophile intellectuals, this spawned Early Internet Argument Culture, where we tried to iron out our differences through Reason. We hoped that the new world the Web revealed to us could be managed in the same friendly way we managed differences with our crazy uncle or the next-door neighbor.

As friendly debate started feeling more and more inadequate, and as newer and less nerdy people started taking over the Internet, this dream receded. In its place, we were left with an intolerable truth: a lot of people seem really horrible, and refuse to stop being horrible even when we ask them nicely. They seem to believe awful things. They seem to act in awful ways. When we tell them the obviously correct reasons they should be more like us, they refuse to listen to them, and instead spout insane moon gibberish about how they are right and we are wrong.

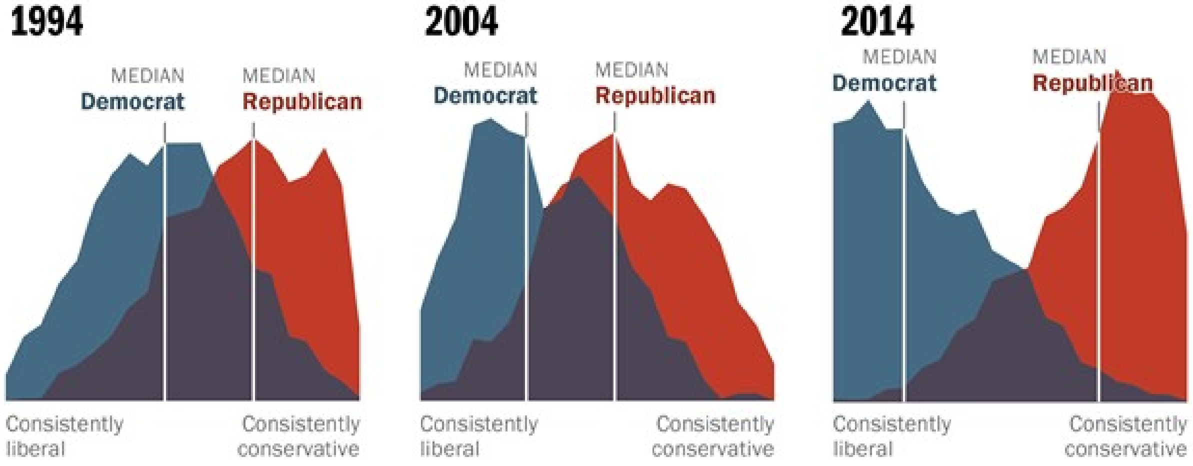
I can only describe this experience from my own side of the aisle, which was the progressive side. We watched the US population elect George W Bush and act like this was a remotely reasonable thing to do. We saw people destroying the environment, leaving the poor to starve, and denying gay people their right to live as normal members of society. We saw people endorsing weird ideas and conspiracy theories, from homeopathy and creationism to the Clintons murdering their enemies. We were always vaguely aware from reading the newspapers that some of these people existed. But now we were seeing and conversing with them every day.

Not only were we noticing the trend for the first time, but the trend itself was strengthening. I could use any of a hundred images to make this case, but for today I’ll use these:

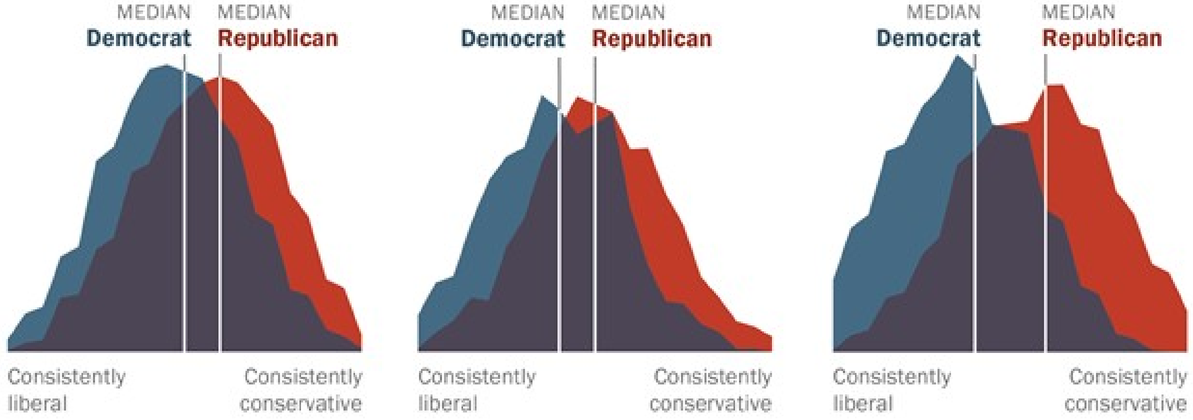
Polarization Surges Among the Politically Engaged

Distribution of Democrats and Republicans on a 10-item scale of political values, by level of political engagement

Among the politically engaged



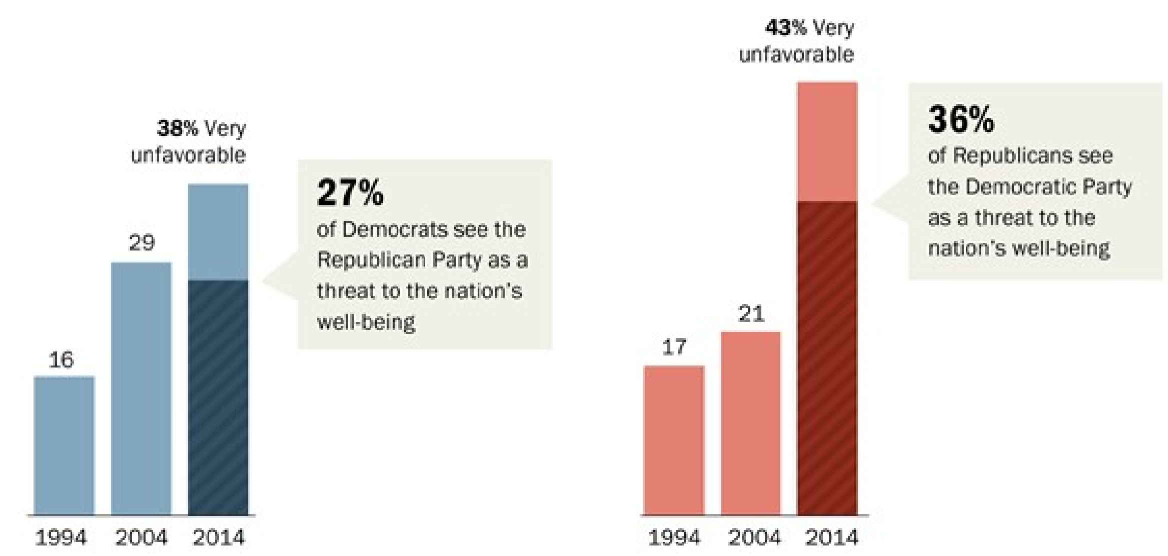
Among the less engaged



Source: 2014 Political Polarization in the American Public

Notes: Ideological consistency based on a scale of 10 political values questions (see Appendix A). Republicans include Republican-leaning independents; Democrats include Democratic-leaning independents (see Appendix B). Politically engaged are defined as those who are registered to vote, follow government and public affairs most of the time and say they vote always or nearly always.

PEW RESEARCH CENTER



Beyond Dislike: Viewing the Other Party as a 'Threat to the Nation's Well-Being'

Democratic *attitudes about the* **Republican Party**

27% of Democrats see the Republican Party as a threat to the nation's well-being

Republican *attitudes about the* **Democratic Party**

36% of Republicans see the Democratic Party as a threat to the nation's well-being

Source: 2014 Political Polarization in the American Public
Notes: Questions about whether the Republican and Democratic Parties are a threat to the nation's well being asked only in 2014. Republicans include Republican-leaning independents; Democrats include Democratic-leaning independents (see Appendix B).
PEW RESEARCH CENTER

And so we asked ourselves: *what the hell is wrong with these people?*

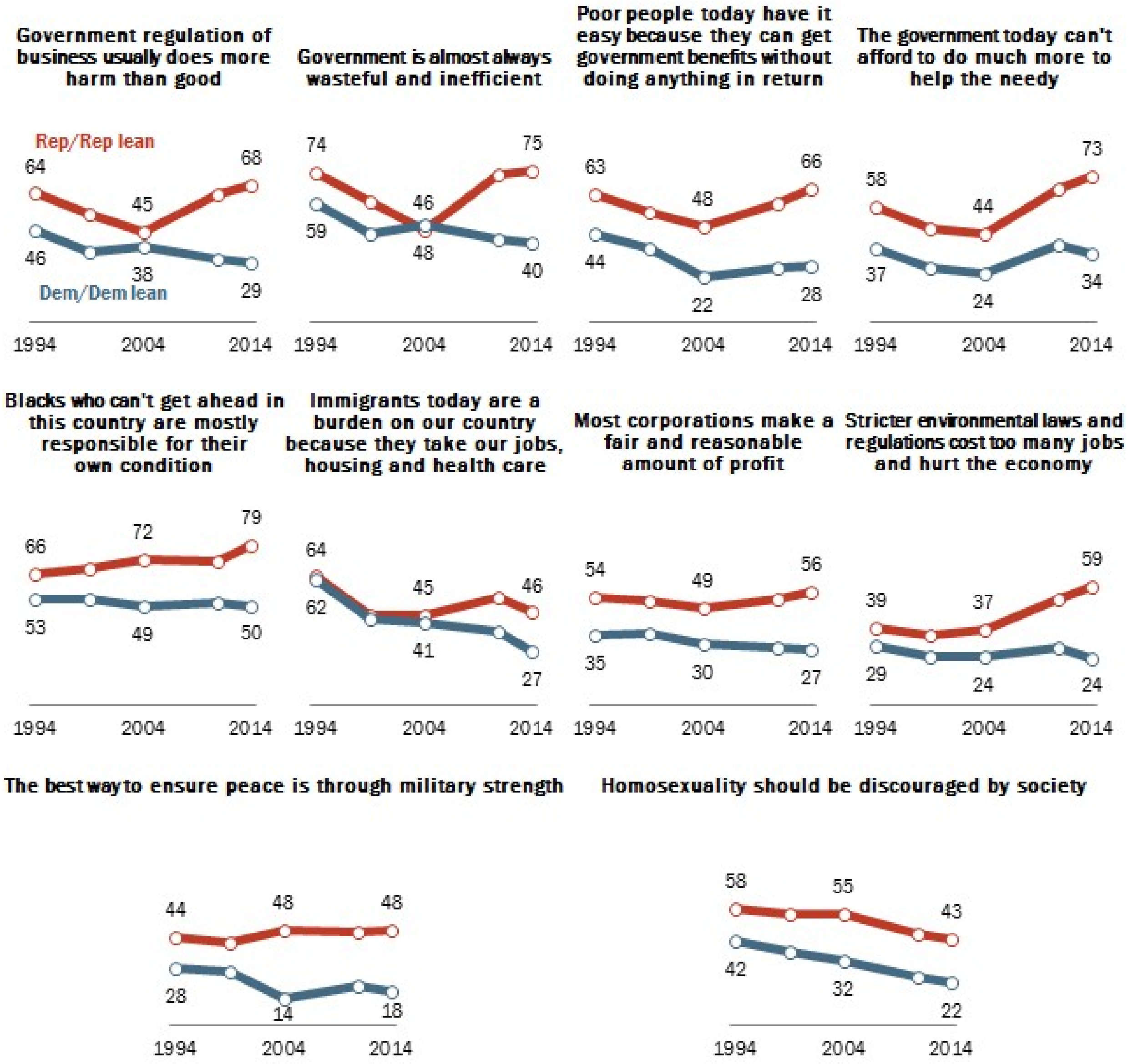
And New Atheism had an answer: *religion*.

That was it. It was beautiful, it was simple, it was perfect. We were the "reality-based community". They were ignoring Reason and basing all of their opinions on three thousand year old fairy-tales because people told them they would burn in Hell forever if they didn't. There was nothing confusing or unsettling at all about the situation, and we did not need to question any of our own beliefs. It was just that some people had been brainwashed by their church/mosque/synagogue to believe transparently wrong things, so they did. Sin began with the apple tree in Eden; conservatism began with the Bible in Jerusalem. Language separates us from the apes; not being blinded by religion separates us from the Republicans.

This was a socially momentous proposal. The Democratic Party is centuries old, but [the Blue Tribe]() – the Democratic Party as a social phenomenon with strong demographic and ideological implications – can be said to have started in 2004.

Growing Gaps between Republicans and Democrats

% who take the more conservative position on each question in the ideological consistency scale



Source: 2014 Political Polarization in the American Public
Line charts show the survey questions included in the 10-item ideological consistency scale. Questions are forced-choice questions with two options; only the conservative responses are shown here. See topline for the full question wording and trends for each question.

PEW RESEARCH CENTER

As it took its first baby steps, the Blue Tribe started asking itself “Who am I? What defines me?”, trying to figure out how it conceived of itself. New Atheism had an answer – “You are the people who aren’t blinded by fundamentalism” – and for a while the tribe toyed with accepting it. During the Bush administration, with all its struggles over Radical Islam and Intelligent Design and Faith-Based Charity, this seemed like it might be a reasonable answer. The atheist movement and the network of journalists/academics/pundits/operatives who made up the tribe’s core started drifting closer together.

Gradually the Blue Tribe got a little bit more self-awareness and realized this was not a great idea. Their coalition contained too many Catholic Latinos, too many Muslim Arabs, too many Baptist African-Americans. Remember that in 2008, “what if all the Hispanic people end up going Republican?” was considered a major and plausible concern. It became somewhat less amenable to New Atheism’s answer to its identity question – but absent a better one, the New Atheists continued to wield some social power.

Betweem 2008 and 2016, two things happened. First, Barack Obama replaced George W. Bush as president. Second, Ferguson. The Blue Tribe kept posing its same identity question: “Who am I? What defines me?”, and now Black Lives Matter gave them an answer they liked better “You are the people who aren’t blinded by sexism and racism.”

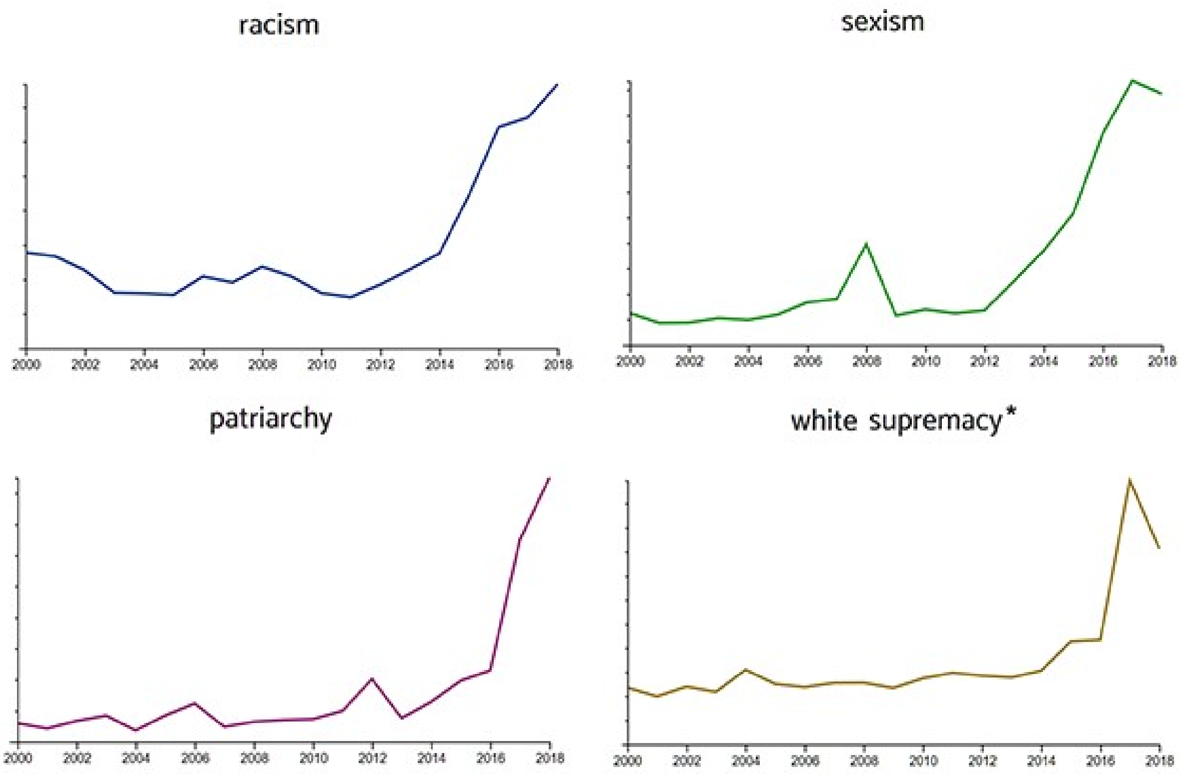
Again, it was beautiful, simple, and perfect. We were “the reality-based community”. They were ignoring Reason and basing all of their opinions on blind hatred and prejudice. There was nothing confusing or unsettling at all about the situation, and we did not need to question any of our own beliefs. It was just that some people had been brainwashed by white supremacy and an all-consuming desire to protect their own privilege, and so they did. Sin began with the apple tree in Eden; conservativism began with the cotton plant in Jamestown. Language separates us from the apes; not being blinded by bigotry separates us from the Republicans.

Since I started writing this essay, I've noticed a surprising number of people just saying this outright. If you go to any thread on r/politics about Trump (aka any thread on r/politics), you'll see people saying things like this:

> [Trump voters] know they are being lied to, well most of them do, but look at the increase of hatred in America. THAT is what they are voting on. Hatred. Ironically, the Republicans are a large reason why their lives are so shitty and full of hatred, but hatred nonetheless. I guarantee you, you debate any of these people long enough. You back them into a corner. They say the same thing. "We are winning. We won the election. Racism is good. Hatred is good. Cheating on elections is good as long as it's my side." Because that is what happens when one side is the Republicans and the other side is baby murdering, child raping and trafficking and harvesting drugs from their brain, brown and black people loving devil worshippers. Go on you know what sub. Read their posts. They will say, "I was driving by a school bus stop, none of them were white." This makes them so angry. To them, Russia [is better than] Democrats. At least Russia is white.

Google Trends shows traffic for atheism-related terms starting to decline around 2012, and really plummeting around 2015. How were other terms doing around that time?

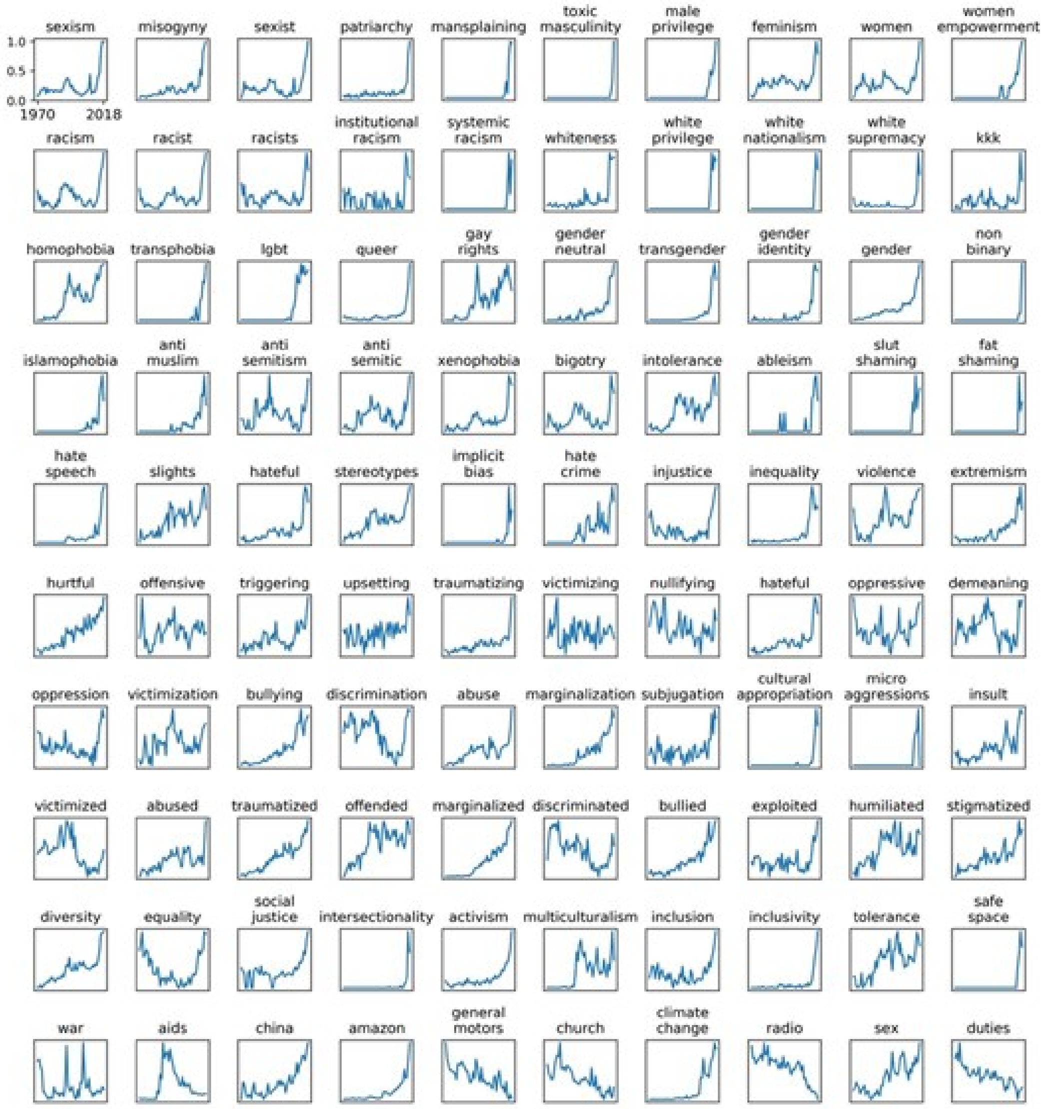
Word Usage Frequency in the *New York Times*, 2000 - Present



*imputed from use of the word "supremacy" since Media Analytics does not calculate bigrams

Not enough for you? We can go deeper:

New York Times Word Usage Frequency (1970-2018)



Most movement atheists weren't in it for the religion. They were in it for the hamartiology. Once they got the message that the culture-at-large had settled on a different, better hamartiology, there was no psychological impediment to switching over. We woke up one morning and the atheist bloggers had all quietly became social justice bloggers. Nothing else had changed because nothing else had to; the underlying itch being scratched was the same. They just had to CTRL+F and replace a couple of keywords.

Eventually, things came full circle. I started this essay with a memory of noticing that my favorite early-2000s-era website had two off-topic forums: one for religion vs. atheism, and one for everything else. Earlier this year, SSC's subreddit split in two: one for "culture war" discussions mostly about race and gender, the other for everything else.

Where do we go from here? I'm not sure. The socialist wing of the Democratic Party seems to be working off a model kind of like this, but hoping to change the hamartiology from race/gender to class. Maybe they'll succeed, and one day talking too much about racism will seem as out-of-touch as talking too much about atheism does now; maybe the rise of terms like "woke capitalism" is already part of this process.

I've lost the exact quote, but a famous historian once said that we learn history to keep us from taking the present too seriously. This isn't to say the problems of the present aren't serious. Just that history helps us avoid getting too dazzled by current trends, or too swept away by any particular narrative.

If this is true, we might do well to study the history of New Atheism a little more seriously.

Samsara

I.

The man standing outside my front door was carrying a clipboard and wearing a golden robe. “Not interested,” I said, preparing to slam the door in his face.

“Please,” said the acolyte. Before I could say no he’d jammed a wad of $100 bills into my hand. “If this will buy a few moments of your time.”

It did, if only because I stood too flabbergasted to move. Surely they didn’t have enough money to do this for everybody.

“There is no everybody,” said the acolyte, when I expressed my bewilderment. “You’re the last one. The last unenlightened person in the world.”

And it sort of made sense. Twenty years ago, a group of San Francisco hippie/yuppie/techie seekers had pared down the ancient techniques to their bare essentials, then optimized *hard*. A combination of drugs, meditation, and ecstatic dance that could catapult you to enlightenment in the space of a weekend retreat, 100% success rate. Their cult/movement/startup, the Order Of The Golden Lotus, spread like wildfire through California – a state where wildfires spread even faster than usual – and then on to the rest of the world. Soon investment bankers and soccer moms were showing up to book clubs talking about how they had grasped the peace beyond understanding and vanquished their ego-self.

I’d kind of ignored it. Actually, super ignored it. First a flat refusal to attend Golden Lotus retreats. Then slamming the door in their face whenever their golden-robed pamphleteers came to call. Then quitting my job to live off savings after my coworkers started converting and the team-building exercises turned into meditation sessions. Then unplugging my cable box after the sitcoms started incorporating Golden Lotus themes and the national news started being about how peaceful everybody was all the time. After that I might have kind of become a complete recluse, never leaving the house, ordering meals through UberEats, cut off from noticing any of the changes happening outside except through the gradual disappearance of nonvegetarian restaurants on the app.

I’m not a bigot; people can have whatever religion they choose. But Golden Lotus wasn’t for me. I don’t want to be enlightened. I like being an individual with an ego. Ayn Rand loses me when she starts talking politics, but the stuff about selfishness really speaks to me. Tend to your own garden, that kind of thing. I’m not becoming part of some universal-love-transcendent-joy hive mind, and I’m not interested in what Golden Lotus is selling.

So I just said: “Cool. Do I get a medal?”

“This is actually very serious,” said the acolyte. “Do you know about the Bodhisattva’s Vow?”

“The what now?”

“It’s from ancient China. You say it before embarking on the path of enlightenment. ‘However innumerable sentient beings are, I vow to save them all.’ The idea is that we’re all in this together. We swear that we will not fully forsake this world of suffering and partake of the ultimate mahaparanirvana – complete cosmic bliss – until everyone is as enlightened as we are.”

“Cool story.”

“That means 7.5 billion people are waiting on *you*.”

“What?”

“We all swore not to sit back and enjoy enlightenement until everyone was enlightened. Now everyone is enlightened except you. You’re the only thing holding us all back from ultimate cosmic bliss.”

“Man. I’m sorry.”

“You are forgiven. We would like to offer you a free three-day course with the Head Lama of Golden Lotus to correct the situation. We’ll pick you up at your home and fly you to the Big Island of Hawaii, where the Head Lama will personally…”

“…yeah, no thanks.”

“What?”

“No thanks.”

“But you have to! Nobody else can reach mahaparanirvana until you get enlightened!”

“Sure they can. Tell them I’m okay, they can head off to mahabharata without me, no need to wait up.”

“They *can’t*. They *swore not to.*”

“Well, they shouldn’t have done that.”

“It’s done! It’s irreversible! The vow has been sworn! Each of the seven point five billion acolytes of Golden Lotus has sworn it!”

“Break it.”

“We are enlightened beings! We can’t break our solemn vow!”

“Then I guess you’re going to learn an important lesson about swearing unbreakable vows you don’t want to keep.”

“Sir, this entire planet is heavy with suffering. It groans under its weight. Seven billion people, the entirety of the human race, and for the first time they have the chance to escape together! I understand you’re afraid of enlightenment, I understand that this isn’t what you would have chosen, but for the sake of the world, please, accept what must be!”

“I’m sorry,” I said. “I really am. But the fault here is totally yours. You guys swore an oath conditional on my behavior, but that doesn’t mean I have to change my behavior to prevent your oath from having bad consequences. Imagine if I let that work! You could all swear to kill yourself unless I donated money, and I’d have to donate or have billions of deaths on my hands. That kind of reasoning, you’ve got to nip it in the bud. I’m sorry about your oath and I’m sorry you’re never going to get to Paramaribo but I don’t want to be enlightened and you can’t make me.”

I slammed the door in his face.

II.

A few days later, just as I was trying to order lunch on UberEats, my cell phone internet stopped working. I tried my laptop. Wasn’t working either. iPad? Not working.

I’d been wondering whether Golden Lotus was going to kill me. It was the natural thing to try in this situation. But I figured people who were too enlightened to break a vow were probably too enlightened to murder me, or to threaten to break my kneecaps, or to drug me, or to take any of the other easy ways out.

But starving me – that might work. And if everyone else was a Golden Lotuser at this point, they were like a world state. They probably controlled the infrastructure, and I didn’t think there were any ancient Buddhist commandments against shutting off someone’s internet connection.

There was a 7-11 on the corner of my street. I put on a jacket, prayed to any god who would listen to me right now, and walked outside.

My street, as I remembered it, was gone. The familiar buildings had been torn down. Far away, I could see tranquil gardens and intricate pagodas. But the street I was on – the one between my apartment and 7-11 – had been turned into a gauntlet. A series of flashing, attention-grabbing billboards and video-screens explaining Golden Lotus techniques, the virtues of enlightenment, and the illusory nature of the material world, accompanied by a soundtrack of giant speakers blaring sermons.

So this was their plan. Not very subtle, but I could live with it. I stared down at my feet and broke into a run, trying to make it to the store as quickly as possible without absorbing any of the information being blasted at me. Staring at my feet turned out to be a mistake – there were sutras written all along the pavement. The first giant letter was right past my doormat, and I saw them stretching forward, continuing in order to the 7-11 I was trying to reach. I tried looking up instead, but a transparent canopy placed atop the street was similarly laden with spiritual wisdom. I closed my eyes, but this slowed my progress forward, and made me more vulnerable to sermons coming from the speakers all around me. “SINGLE-POINTED AWARENESS ON ANY INDIVIDUAL SENSATION REVEALS ITS EMPTINESS!” blared one. “THE MIND IS LIKE A STILL POOL DISTURBED BY THE RIPPLES OF THOUGHTS” blasted another.

I thought about the technical problem facing Golden Lotus leadership: how do you enlighten someone who resists enlightenment? You can’t teach them practices, because they won’t do them. You can’t impart advice, because they won’t take it. But you can draw awareness to certain facets of their own thinking, along the lines of the old “You are now aware of the feeling of your tongue in your mouth”. You can present someone with metaphors of such explanatory value that they reshape the way he interprets his own experience. If you had a lot of very smart people developing the “curriculum”, and a lot of patience, maybe it could work.

How could one resist such an effort? I would have to close all possible communication channels. I put my hands over my ears, even though the awkward position slowed my blind stumble to the store. I took a few steps forward, then felt a sudden weight. I opened my eyes. A brightly-colored macaw had landed on my right shoulder and was staring straight at me. “HERE AND NOW!” it screeched, point-blank, before flying off.

Okay. Trained birds. They were really on top of their game. So maybe I couldn’t close off all possible communication channels. Maybe I would have to fight them on their own turf. Maybe if they’ve created a super-efficient science of enlightenment, I would have to create a super-efficient science of samsara.

The convenience store sold mostly rice and incense now, and restricted rice purchases to a single day's supply. I picked some up and headed for the cashier. The aisles were confusingly laid out, and I realized after a moment that they formed one of those labyrinths that people sometimes walk as a spiritual practice. I didn't think those things worked, but I couldn't take any chances. I climbed a shelf full of meditation cushions, vaulted over, and climbed down the other side to the frowning cashier.

I saw another door on the other side of the 7-11, this one guarded by a stern-looking man in a golden robe. I realized it was the door to freedom – outside my enlightenment-ad-plastered prison and into the world of pagodas and gardens outside. I assessed my chances. The monk was really big, and I didn't know if the door was locked or if there were other guards on the other side. I decided against it. I paid for my rice, stuffed enough of it in my pockets that I could reassume the hands-over-ears-eyes-closed pose, and walked home.

A science of samsara. What would that involve? Instead of meditating on lovingkindness, I could meditate on everybody I hated. Instead of a vow of poverty, I could take a vow of greed. Instead of practicing self-awareness, I could practice self-obliviousness. I took out a piece of paper and began to jot some of this down. This was going to be *so much fun*.

III.

I was at the 7-11, buying a meditation cushion. My meditation on hatred was going well, but sitting on the floor that long was starting to hurt my back. I figured that on my daily rice run, I'd get a cushion, a bell, maybe some looser-fitting clothes. I was near the center of the labyrinth, picking them out, when someone tapped me on the shoulder.

It was the most beautiful woman I'd ever seen. She looked like a supermodel or something. She whispered to me. "Are you…the unenlightened person?"

I nodded, wondering where she was going with this.

"Look," she said. "I have not had decent sex in a year and a half. Everyone is just like 'abandon carnal desires of the flesh' and 'real pleasure comes from within'. And even when I can rope some guy into doing it, somehow it manages to be *tranquil*." She spat out this last word. "Are you…uh…are you free tonight?"

I controlled my shock at my good fortune long enough to sputter out a short "yes".

We stumbled back to my apartment together, braving the billboards and sermons and dive-bomb-parrots. In record time we made it to my bedroom and started ripping our clothes off.

"How did you get in?" I asked her. "Is this place well-guarded?"

"There's a door in the back of the 7-11," she said, confirming my suspicions. "There's one monk and your side, and about five on the other. There's no restriction on people coming in to talk to you if they want. Only on you getting out."

I pulled her onto the bed and into my embrace.

“You feel so good,” she said. “It’s like a snake, coiled at the bottom of the spine, waiting to get out. Oh! It’s like the snake is made of energy, and the energy is escaping, moving upward…”

That sounded familiar. I stopped, pushed her off me.

“Wait a second,” I said. “That’s from tantric sex!”

“Tantric sex?” she asked innocently. “What’s that?”

“Don’t pretend you don’t know what tantric sex is! It’s that thing where sex can be used as a spiritual practice that brings people to enlightenment! You’re trying to trick me!”

She pouted seductively. “Come on, let’s keep going.”

I started putting my clothes back on. “You guys are scared of me. You’re scared that you’re not reaching me, that I’m immune to your tricks. Well, tell them that they’re going to have to try harder. Every day I meditate for an hour on all the people I hate, then another hour on all the material goods I wish I had, then for another hour on all the women I want to screw. Then I finish it off with an hour trying to experience selfhood as viscerally and dramatically as possible. I’m reaching depths of samsara they can’t even imagine. And there’s nothing you or the Head Lama or anyone else can do about it. Get out!” I threw her clothes at her. When she left, I slammed the door in her face.

IV.

A knock on the door.

I got up from my meditation cushion, eyeing the stains and scratches on it. Twenty years. Twenty years I had kept up my meditation practice, the four hours of anger-greed-lust-selfhood meditation I’d established a few weeks after my confinement started. To be honest, I didn’t look much better than the cushion. I was getting old. My rice and tap water diet kept me lean and wiry, but the years still took their toll. There were no razors at the 7-11, so I had grown a long white beard.

For the first few years, Golden Lotus had tried more tricks like the supermodel. I had seen through them all. Eventually they must have given up. I’d been unmolested for more than a decade. I wondered what they were up to now.

At the door was a kid. There was no other way to describe him. Scrawny, a little worn-out, looked South Asian, maybe sixteen or seventeen. He was wearing some kind of black plastic poncho over his clothes.

“Excuse me,” he said. “Are you the unenlightened man?”

“That’s me.”

“I want to learn from you.”

“What?”

“Master, until now I have lived an unexamined life. Going to temple every day, meditating, taking the drugs, doing the dances. But I longed for something more. In an old library, I found a book which claimed the ancients knew of a state known as samsara, and of a mystery called the Self. That those who master these mysteries gain strange powers. Using the technique of Greed, they can attain such perfect willpower that they can work eighty hour weeks for abusive bosses without quitting. Using the technique of Lust, they can reach such perfect focus that all their thoughts for months revolve around the same person.

“I thought it might all just be legends. But I asked those who knew more than I did, and they directed me to those who knew more than them, and finally I heard rumors that in a far-off place called California there was an ancient sage who had achieved samsara long ago. Please, Master, will you take me as your disciple?”

I was flabbergasted for just a second before common sense took hold of me once again. “No,” I said. “You’re some kind of trick. Go away.”

“Master!” protested the kid. “I will wait kneeling on your doorstep without food or water until you agree to take me as a disciple!”

I shrugged and closed the door.

The next day, when I went out to 7-11, the kid was still there, kneeling.

“Master!” he said. “Please take me as your disciple.”

“No,” I said. “But if you want to make yourself useful, you can help guide me to the corner store while I have my eyes closed and my hands over my ears. And if you see any parrots, fight them off.”

“Yes, Master!” said the kid, and he took me by the arm and helped guide me to 7-11.

The next day the same thing happened. I went to go to the store, the kid was waiting on my doorstep, and he helped guide me to 7-11 safely. By the time I got back it was raining, and although the transparent canopy covered with sutra verses blocked out the worst of it, there was still a chill in the air.

“You might as well come inside and sleep on the couch,” I told him. “And have a little of the rice.”

The next morning, we began his training. I asked him to think about all the material goods he wanted. He couldn’t come up with any. I asked him to think about all the most attractive women he knew. He said he’d never thought about women that way before, and it seemed kind of objectifying. I asked him who really pissed him off, and his only answer was himself, when he strayed from the path of maximum virtue.

I tried for a few hours, then I gave up.

“Go to the spare room,” I said “and think about the sound of one hand clapping. Once you figure it out, come tell me. Until then, leave me alone. Got it, uh…what was your name again?”

“Maitrayaniputra,” said the kid.

“Not anymore,” I said. “From now on, your name is Brad.”

V.

Somehow, my fame spread.

My apartment-street-convenience-store prison had turned into a makeshift ashram. Two dozen seekers from all around the globe. A select few slept in my house. The rest pitched tents on the street, or huddled into the aisles of the 7-11. The guard on the back door stared at them impassively, but said nothing.

I tried to discourage them, turn them away. But every time I yelled at one, or hit her with my cane, or slapped him on the face, more kept coming, sure this was some manifestation of ancient wisdom. A few gave up and returned back through the guarded door; but overall the numbers grew and grew.

Brad had declared himself chief of my disciples, the Peter to my Jesus. He would lead the congregation in meditation each morning, drawing off my old morning routine – an hour thinking of all the people they were angry at, an hour thinking of the material goods they wanted, an hour thinking of all the sexy people they wanted to screw – followed by a final hour of meditation on the Self. The novices failed in ways I didn't even think possible. All the material goods they wanted were things like lotuses and celestial jewels. All the people they wanted to have sex with were particularly virtuous saints whose wisdom they admired. Sometimes, I would march into the room and demand to know what a novice was thinking. "Who are you having sexual fantasies about?" I shouted at one young man, who I had given the name Kyle. He admitted he was thinking of the Tibetan Buddhist guru Padmasambhava. "Are you even gay?" I demanded. He didn't know what that meant, so I explained that some people were straight and should be having fantasies about the opposite sex, and other people were gay and should be having fantasies about the same sex, and other people were bi and could have fantasies about whichever sex they wanted. "But how would I know which of those I am?" Kyle asked me. I didn't know what to say, so I hit him with my stick and stormed out.

But they kept coming. Kyle left the ashram, only to return a few weeks later with his sister. Her name was Shantideva; I told her she would henceforth be Sherri. Sherri was stick-thin, a dialysis port in one arm, and rarely spoke. Kyle told me she had a rare disease and would die before age 30. She had read Dylan Thomas' "Rage Against The Dying Of The Light" and decided to achieve samsara before she died, so that she too could feel rage at her situation. I was nervous about her – she looked like the slightest breeze would tip her over – but she meditated with a fervor beyond anything I could have predicted, sometimes outdoing even Brad in time spent on the cushion.

I instituted a dress code for my disciples. I made the men dress as douchey as possible, and the women as slutty as they could. One day I dug my old printer out of a closet, and ran off a thousand copies of George Washington’s face. I distributed them to the disciples as unevenly as I could. “This is money,” I said. “It is an important ritual object. From now on, whenever someone wants something from you, you must refuse unless they offer you money. If they don’t offer you enough money, you should yell at them and call them cheap. If they offer you too much money, you should laugh at them behind their backs and tell everyone they’re an easy mark.”

“But Master,” protested Kyle, “why do we need all these rituals? Didn’t you yourself say that the essence of samsara is about mental states? Aren’t all these intermediaries and traditions only distracting us from the true work of self-transformation?”

“I will give you $10 to shut up and stop bothering me about this,” I said, and I handed him ten of the Washington papers.

Kyle slowly nodded and took them.

“Now do you understand?” I asked.

Kyle nodded, but I could tell he did not understand.

A few days later, Brad came into my room. I looked up.

“Master,” he said. “There is no sound of one hand clapping. You were just trying to get rid of me. I wasted almost a year of my life trying to figure it out, and there was nothing there. It was all a fraud and you’re a fraud and this whole piece of shit ashram is a fraud. Fuck you.”

“My son,” I said. “Today you have achieved samsara.”

Brad stopped as if stuck by a train. He tried to speak, then tried again, then fell silent. I watched as understanding flowed into his eyes.

“You bastard,” he said. “You magnificent bastard. You really did it.” He hugged me. I hugged him back. Then I marched him out to the street, where the majority of the disciples were eating their evening meal. “Everybody!” I announced. “Brad is unenlightened now! That means he’s better than you! He’s going to lord it over you, and you should all feel jealous of him!” A few looks of bewilderment from people who couldn’t grasp why they should be unhappy at anyone else’s achievement, but that was fine. I knew I had planted a seed.

VI.

Years went by. My first disciples – Brad, Kyle, Sherri, and the rest – left the ashram to preach to the outside world. New disciples replaced them. Life went on.

I grew into my role as samsara master. If Golden Lotus could enlighten people in a weekend, I needed to be able to unenlighten them faster. I spent more and more time in meditation, probing the true meaning of samsara, investigating each impulse, querying each baser urge. My doctrines became more and more esoteric. I began telling seekers that they were already unenlightened, if only they could see it. That there was nothing to attain. That there was no samsara separate from nirvana.

Some left, unable to handle the paradox. It was one of these, a middle-aged man I had dubbed Logan, who left behind the golden robe.

He had taken off to change it douchey clothes as soon as he arrived. And he left in the douchey clothes I gave him. The golden robe hung in my closet. Nobody missed it. Nobody knew I had it.

I decided to try a jailbreak.

I put on the golden robe. Then I dug up an old razor from the bottom drawer of my bathroom. Then I shaved off my long beard. Then I shaved my head, until I looked the very image of a Golden Lotus monk.

I went out to the 7-11 and walked up to the back door. “I’m sick of this place,” I told the guard. “I’m going home.”

He waved me through, and for the first time in twenty-five years, I stepped into the world beyond.

It looked like a Japanese garden. Bonsai-perfect trees grew everywhere, hanging over glassy ponds stocked with koi. The roads had given way to carefully tended paths, lined every so often by pagodas or temple-like houses.

I walked further, until I reached what had been the town center before. The general aesthetic continued, but the buildings were closer together now. I saw fellow golden-robed acolytes walking the streets or sitting contemplatively beneath the trees.

One golden-robed man sitting underneath a cherry tree looked exactly like Brad. He was talking to another man who looked exactly like Kyle. I could only hear bits of their speech, but it sounded very tranquil. I hid behind a shrine. What was going on here? Was it really them? Had they reverted already?

“Sorry!” said a jogger, as she almost ran into me. I blinked again, took a second look. It was Sherri, the frail girl with the chronic disease. She didn’t look frail or diseased now. I grabbed her by the wrist, made her stop.

“Sherri. What’s going on?”

I saw recognition in her eyes, and her lips curled into a smile.

I’d been right that first time then, all those years ago. A trick. They’d all been plants. Why? What had they accomplished? *Getting me thinking about samsara*. I retraced several years worth of mental steps. Trying to understand the nature of desire. Becoming more aware of the movements of my own mind. They had gotten me *good*. I had to distract myself. Think of a material good. Think of a red Ferrari. I concentrated on a red Ferrari as hard as I could, tried to block everything else out of my mind, all the insights, all the shame, all the trickery. Just a red Ferrari, on a black road, beneath a blue sky. Everything else faded.

Sherri clapped once, right in front of my face.

Upon hearing this, I was enlightened.

Building Intuitions On Non-Empirical Arguments In Science

I.

Aeon: Post-Empirical Science Is An Oxymoron And It is Dangerous:

> There is no agreed criterion to distinguish science from pseudoscience, or just plain ordinary bullshit, opening the door to all manner of metaphysics masquerading as science. This is 'post-empirical' science, where truth no longer matters, and it is potentially very dangerous.
>
> It's not difficult to find recent examples. On 8 June 2019, the front cover of New Scientist magazine boldly declared that we're 'Inside the Mirrorverse'. Its editors bid us 'Welcome to the parallel reality that's hiding in plain sight'. […]
>
> [Some physicists] claim that neutrons [are] flitting between parallel universes. They admit that the chances of proving this are 'low', or even 'zero', but it doesn't really matter. When it comes to grabbing attention, inviting that all-important click, or purchase, speculative metaphysics wins hands down.
>
> These theories are based on the notion that our Universe is not unique, that there exists a large number of other universes that somehow sit alongside or parallel to our own. For example, in the so-called Many-Worlds interpretation of quantum mechanics, there are universes containing our parallel selves, identical to us but for their different experiences of quantum physics. These theories are attractive to some few theoretical

> physicists and philosophers, but there is absolutely no empirical evidence for them. And, as it seems we can’t ever experience these other universes, there will never be any evidence for them. As Broussard explained, these theories are sufficiently slippery to duck any kind of challenge that experimentalists might try to throw at them, and there’s always someone happy to keep the idea alive.
>
> Is this really science? The answer depends on what you think society needs from science. In our post-truth age of casual lies, fake news and alternative facts, society is under extraordinary pressure from those pushing potentially dangerous antiscientific propaganda – ranging from climate-change denial to the anti-vaxxer movement to homeopathic medicines. I, for one, prefer a science that is rational and based on evidence, a science that is concerned with theories and empirical facts, a science that promotes the search for truth, no matter how transient or contingent. I prefer a science that does not readily admit theories so vague and slippery that empirical tests are either impossible or they mean absolutely nothing at all.

As always, a single quote doesn’t do the argument justice, so go read the article. But I think this captures the basic argument: multiverse theories are bad, because they’re untestable, and untestable science is pseudoscience.

Many great people, both philosophers of science and practicing scientists, have already discussed the problems with this point of view. But none of them lay out their argument in quite the way that makes the most sense to me. I want to do that here, without claiming any originality or special expertise in the subject, to see if it helps convince anyone else.

II.

Consider a classic example: modern paleontology does a good job at predicting dinosaur fossils. But the creationist explanation – Satan buried fake dinosaur fossils to mislead us – also predicts the same fossils (we assume Satan is good at disguising his existence, so that the lack of other strong evidence for Satan doesn't contradict the theory). What principles help us realize that the Satan hypothesis is obviously stupid and the usual paleontological one more plausible?

One bad response: paleontology can better predict characteristics of dinosaur fossils, using arguments like "since plesiosaurs are aquatic, they will be found in areas that were underwater during the Mesozoic, but since tyrannosaurs are terrestrial, they will be found in areas that were on land", and this makes it better than the Satan hypothesis, which can only retrodict these characteristics. But this isn't quite true: since Satan is trying to fool us into believing the modern paleontology paradigm, he'll hide the fossils in ways that conform to its predictions, so we will predict plesiosaur fossils will only be found at sea – otherwise the gig would be up!

A second bad response: "The hypothesis that all our findings were planted to deceive us bleeds into conspiracy theories and touches on the problem of skepticism. These things are inherently outside the realm of science." But archaeological findings are very often deliberate hoaxes planted to deceive archaeologists, and in practice archaeologists consider and test that hypothesis the same way they consider and test every other hypothesis. Rule this out by fiat and we have to accept Piltdown Man, or at least claim that the people arguing against the veracity of Piltdown Man were doing something other than Science.

A third bad response: “Satan is supernatural and science is not allowed to consider supernatural explanations.” Fine then, replace Satan with an alien. I think this is a stupid distinction – if demons really did interfere in earthly affairs, then we could investigate their actions using the same methods we use to investigate every other process. But this would take a long time to argue well, so for now let’s just stick with the alien.

A fourth bad response: “There is no empirical test that distinguishes the Satan hypothesis from the paleontology hypothesis, therefore the Satan hypothesis is inherently unfalsifiable and therefore pseudoscientific.” But this can’t be right. After all, there’s no empirical test that distinguishes the paleontology hypothesis from the Satan hypothesis! If we call one of them pseudoscience based on their inseparability, we have to call the other one pseudoscience too!

A naive Popperian (which maybe nobody really is) would have to stop here, and say that we predict dinosaur fossils will have such-and-such characteristics, but that questions like that process that drives this pattern – a long-dead ecosystem of actual dinosaurs, or the Devil planting dinosaur bones to deceive us – is a mystical question beyond the ability of Science to even conceivably solve.

I think the correct response is to say that both theories explain the data, and one cannot *empirically* test which theory is true, but the paleontology theory *is more elegant* (I am tempted to say “simpler”, but that might imply I have a rigorous mathematical definition of the form of simplicity involved, which I don’t). It requires fewer other weird things to be true. It involves fewer other hidden variables. It transforms our worldview less. It gets a cleaner shave with Occam’s Razor. This elegance is so important to us that it explains our vast preference for the first theory over the second.

A long tradition of philosophers of science have already written eloquently about this, summed up by Sean Carroll here:

> What makes an explanation "the best." Thomas Kuhn ,after his influential book The Structure of Scientific Revolutions led many people to think of him as a relativist when it came to scientific claims, attempted to correct this misimpression by offering a list of criteria that scientists use in practice to judge one theory better than another one: accuracy, consistency, broad scope, simplicity, and fruitfulness. "Accuracy" (fitting the data) is one of these criteria, but by no means the sole one. Any working scientist can think of cases where each of these concepts has been invoked in favor of one theory or another. But there is no unambiguous algorithm according to which we can feed in these criteria, a list of theories, and a set of data, and expect the best theory to pop out. The way in which we judge scientific theories is inescapably reflective, messy, and human. That's the reality of how science is actually done; it's a matter of judgment, not of drawing bright lines between truth and falsity or science and non-science. Fortunately, in typical cases the accumulation of evidence eventually leaves only one viable theory in the eyes of most reasonable observers.

The dinosaur hypothesis and the Satan hypothesis both fit the data, but the dinosaur hypothesis wins hands-down on simplicity. As Carroll predicts, most reasonable observers are able to converge on the same solution here, despite the philosophical complexity.

III.

I’m starting with this extreme case because its very extremity makes it easier to see the mechanism in action. But I think the same process applies to other cases that people really worry about.

Consider the riddle of the Sphinx. There’s pretty good archaeological evidence supporting the consensus position that it was built by Pharaoh Khafre. But there are a few holes in that story, and a few scattered artifacts suggest it was actually built by Pharaoh Khufu; a respectable minority of archaeologists believe this. And there are a few anomalies which, if taken wildly out of context, you can use to tell a story that it was built long before Egypt existed at all, maybe by Atlantis or aliens.

So there are three competing hypotheses. All of them are consistent with current evidence (even the Atlantis one, which was written after the current evidence was found and carefully adds enough epicycles not to blatantly contradict it). Perhaps one day evidence will come to light that supports one above the others; maybe in some unexcavated tomb, a hieroglyphic tablet says “I created the Sphinx, sincerely yours, Pharaoh Khufu”. But maybe this won’t happen. Maybe we already have all the Sphinx-related evidence we’re going to get. Maybe the information necessary to distinguish among these hypotheses has been utterly lost beyond any conceivable ability to reconstruct.

I don’t want to say “No hypothesis can be tested any further, so Science is useless to us here”, because then we’re forced to conclude stupid things like “Science has no opinion on whether the Sphinx was built by Khafre or Atlanteans,” whereas I think most scientists would actually have very strong opinions on that.

But what about the question of whether the Sphinx was built by Khafre or Khufu? This is a real open question with respectable archaeologists on both sides; what can we do about it?

I think the answer would have to be: the same thing we did with the Satan vs. paleontology question, only now it's a lot harder. We try to figure out which theory requires fewer other weird things to be true, fewer hidden variables, less transformation of our worldview – which theory works better with Occam's Razor. This is relatively easy in the Atlantis case, and hard but *potentially possible* in the Khafre vs. Khufu case.

(Bayesians can rephrase this to: given that we have a certain amount of evidence for each, can we quantify exactly how much evidence, and what our priors for each should be. It would end not with a decisive victory of one or the other, but with a probability distribution, maybe 80% chance it was Khafre, 20% chance it was Khufu)

I think this is a totally legitimate thing for Egyptologists to do, even if it never results in a particular testable claim that gets tested. If you don't think it's a legitimate thing for Egyptologists to do, I have trouble figuring out how you can justify Egyptologists rejecting the Atlantis theory.

(Again, Bayesians would start with a very low prior for Atlantis, and assess the evidence as very low, and end up with a probability distribution something like Khafre 80%, Khufu 19.999999%, Atlantis 0.000001%)

IV.

How does this relate to things like multiverse theory? Before we get there, one more hokey example:

Suppose scientists measure the mass of one particle at 32.604 units, the mass of another related particle at 204.897 units, and the mass of a third related particle at 145173.870 units. For a while, this is just how things are – it seems to be an irreducible brute fact about the universe. Then some theorist notices that if you set the mass of the first particle as x, then the second is $2\pi x$ and the third is $4/3\ \pi x^3$. They theorize that perhaps the quantum field forms some sort of extradimensional sphere, the first particle represents a radius of a great circle of the sphere, the second the circumference of the great circle, and the third the volume of the sphere.

(please excuse the stupidity of my example, I don’t know enough about physics to come up with something that isn’t stupid, but I hope it will illustrate my point)

In fact, imagine that there are a hundred different particles, all with different masses, and all one hundred have masses that perfectly correspond to various mathematical properties of spheres.

Is the person who made this discovery doing Science? And should we consider their theory a useful contribution to physics?

I think the answer is clearly yes. But consider what this commits us to. Suppose the scientist came up with their Extradimensional Sphere hypothesis *after* learning the masses of the relevant particles, and so it has not predicted anything. Suppose the extradimensional sphere is outside normal space, curled up into some dimension we can’t possibly access or test without a particle accelerator the size of the moon. Suppose there are no undiscovered particles in this set that can be tested to see if they also reflect sphere-related parameters. This theory is exactly the kind of postempirical, metaphysical construct that the Aeon article savages.

But it’s really compelling. We have a hundred different particles, and this theory retrodicts the properties of each of them perfectly. And it’s so simple – just say the word “sphere” and the rest falls out naturally! You would have to be crazy not to think it was at least pretty plausible, or that the scientist who developed it had done some good work.

Nor do I think it seems right to say “The discovery that all of our unexplained variables perfectly match the parameters of a sphere is good, but the hypothesis that there *really is* a sphere is outside the bounds of Science.” That sounds too much like saying “It’s fine to say dinosaur bones have such-and-such characteristics, but we must never speculate about what kind of process produced them, or whether it involved actual dinosaurs”.

V.

My understanding of the multiverse debate is that it works the same way. Scientists observe the behavior of particles, and find that a multiverse explains that behavior more simply and elegantly than not-a-multiverse.

One (doubtless exaggerated) way I’ve heard multiverse proponents explain their position is like this: in certain situations the math declares two contradictory answers – in the classic example, Schrodinger’s cat will be both alive and dead. But when we open the box, we see only a dead cat or an alive cat, not both. Multiverse opponents say “Some unknown force steps in at the last second and destroys one of the possibility branches”. Multiverse proponents say “No it doesn’t, both possibility branches happen exactly the way the math says, and we end up in one of them.”

Taking this exaggerated dumbed-down account as exactly right, this sounds about as hard as the dinosaurs-vs-Satan example, in terms of figuring out which is more Occam's Razor compliant. I'm sure the reality is more nuanced, but I think it can be judged by the same process. Perhaps this is the kind of reasoning that only gets us to a 90% probability there is a multiverse, rather than a 99.999999% one. But I think determining that theories have 90% probability is a reasonable scientific thing to do.

VI.

At times, the Aeon article seems to flirt with admitting that something like this is necessary:

> Such problems were judged by philosophers of science to be insurmountable, and Popper's falsifiability criterion was abandoned (though, curiously, it still lives on in the minds of many practising scientists). But rather than seek an alternative, in 1983 the philosopher Larry Laudan declared that the demarcation problem is actually intractable, and must therefore be a pseudo-problem. He argued that the real distinction is between knowledge that is reliable or unreliable, irrespective of its provenance, and claimed that terms such as 'pseudoscience' and 'unscientific' have no real meaning.

But it always jumps back from the precipice:

> So, if we can't make use of falsifiability, what do we use instead? I don't think we have any real alternative but to adopt what I might call the empirical criterion. Demarcation is not some kind of binary yes-or-no, right-or-wrong, black-or-white judgment. We have to admit shades of grey. Popper himself was ready to accept this, [saying]:

> *“The criterion of demarcation cannot be an absolutely sharp one but will itself have degrees. There will be well-testable theories, hardly testable theories, and non-testable theories. Those which are non-testable are of no interest to empirical scientists. They may be described as metaphysical.”*

> Here, ‘testability’ implies only that a theory either makes contact, or holds some promise of making contact, with empirical evidence. It makes no presumptions about what we might do in light of the evidence. If the evidence verifies the theory, that’s great – we celebrate and start looking for another test. If the evidence fails to support the theory, then we might ponder for a while or tinker with the auxiliary assumptions. Either way, there’s a tension between the metaphysical content of the theory and the empirical data – a tension between the ideas and the facts – which prevents the metaphysics from getting completely out of hand. In this way, the metaphysics is tamed or ‘naturalised’, and we have something to work with. This is science.

But as we’ve seen, many things we really want to include as science are not testable: our credence for real dinosaurs over Satan planting fossils, our credence for Khafre building the Sphinx over Khufu or Atlanteans, or elegant patterns that explain the features of the universe like the Extradimensional-Sphere Theory.

The Aeon article is aware of Carroll’s work – which, along with the paragraph quoted in Section II above, includes a lot of detailed Bayesian reasoning encompassing everything I’ve discussed. But the article dismisses it in a few sentences:

> Sean Carroll, a vocal advocate for the Many-Worlds interpretation, prefers abduction, or what he calls ‘inference to the best explanation’, which leaves us with theories that are merely ‘parsimonious’, a matter of judgment, and ‘still might reasonably be true’. But whose judgment? In the absence of facts, what constitutes ‘the best explanation’?
>
> Carroll seeks to dress his notion of inference in the cloth of respectability provided by something called Bayesian probability theory, happily overlooking its entirely subjective nature. It’s a short step from here to the theorist-turned-philosopher Richard Dawid’s efforts to justify the string theory programme in terms of ‘theoretically confirmed theory’ and ‘non-empirical theory assessment’. The ‘best explanation’ is then based on a choice between purely metaphysical constructs, without reference to empirical evidence, based on the application of a probability theory that can be readily engineered to suit personal prejudices.

“A choice between purely metaphysical constructs, without reference to empirical evidence” sounds pretty bad, until you realize he’s talking about the same reasoning we use to determine that real dinosaurs are more likely than Satan planting fossils.

I don’t want to go over the exact ways in which Bayesian methods are subjective (which I think are overestimated) vs. objective. I think it’s more fruitful to point out that your brain is already using Bayesian methods to interpret the photons striking your eyes into this sentence, to make snap decisions about what sense the words are used in, and to integrate them into your model of the world. If Bayesian methods are good enough to give you every single piece of evidence about the nature of the external world that you have ever encountered in your entire life, I say they’re good enough for science.

Or if you don’t like that, you can use the explanation above, which barely uses the word “Bayes” at all and just describes everything in terms like “Occam’s Razor” and “you wouldn’t want to conclude something like that, would you?”

I know there are separate debates about whether this kind of reasoning-from-simplicity is actually good enough, when used by ordinary people, to consistently arrive at truth. Or whether it’s a productive way to conduct science that will give us good new theories, or a waste of everybody’s time. I sympathize with some these concerns, though I am nowhere near scientifically educated enough to have an actual opinion on the questions at play.

But I think it’s important to argue that even before you describe the advantages and disadvantages of the complicated Bayesian math that lets you do this, *something like this has to be done*. The untestable is a fundamental part of science, impossible to remove. We can debate how to explain it. But denying it isn’t an option.

Book Review: The Body Keeps The Score

I.

The Body Keeps The Score is a book about post-traumatic stress disorder.

The author, Bessel van der Kolk, helped discover the condition and lobby for its inclusion in the DSM, and the brief forays into that history are the best part of the book. Like so many things, PTSD feels self-evident once you know about it. But this took decades of conceptual work by people like van der Kolk, crystallizing some ideas and hacking away at others until they ended up with something legible to the Establishment. Before that there was nothing. It was absolutely shocking how much nothing there was. As soon as the APA officialy recognized PTSD as a diagnosis in 1980, Bessel and his friends applied for a grant from the VA to study it. The grant was rejected on the grounds that (actual quote from the rejection letter) "it has never been shown that PTSD is relevant to the mission of the Veterans Administration". So the first step in raising awareness of PTSD was – amazingly – convincing the US military that some people might get PTSD from combat.

After the military relented, the next step was convincing everyone else. PTSD was temporarily pigeonholed as "the thing veterans get when they come back from a war". The next push was convincing people that civilian trauma could have similar effects. It was simple to extend the theory to sudden disasters like fires or violent crimes. But van der Kolk and his colleagues started noticing that a history of child abuse, and especially childhood sexual abuse, correlated with a lot of psychiatric problems later on.

Again, “child abuse is bad” sounds self-evident once you know it. But van der Kolk insists this is the result of hard work by a coalition of psychiatrists, psychologists, activists, and victims. When he first started raising awareness of the problem, nobody believed him. His grant proposal to study whether childhood trauma was associated with personality disorders got rejected too. He recalls that:

> I was particularly struck by how many female patients spoke of being sexually abused as children. The standard textbook of psychiatry at the time stated that incest was extremely rare in the United States, cocurring about once in every million women. Given that there were then only about one hundred million women in the United States, I wondered how forty-seven, almost half of them, had found their way to my office in the basement of the hospital.
>
> Furthermore, the textbook said, “There is little agreement about the role of father-daughter incest as a source of serious subsequent psychopathology”…the textbook went on to practically endorse incest, explaining that “such incestuous activity diminishes the subject’s chance of psychosis, and allows for a better adjustment to the external world.”

Van der Kolk found that child abuse (sexual and otherwise) was both far more common and far more destructive than anybody else thought. He also found that it worked differently than regular PTSD. A soldier traumatized during war has already developed a sense of self, and has a concept of a safe homeland to return to if he makes it out alive; a child has neither, and has to deal with trauma again and again absent any trustworthy external support system. This is the same insight some researchers call “complex PTSD”; van der Kolk uses the terms “developmental trauma disorder” and argues it is the real culprit behind many people currently diagnosed with ADHD, bipolar, intermittent explosive disorder, oppositional defiant disorder, etc. He rejects at least some of these diagnoses as “pseudoscience…impressive but meaningless labels”.

A group including Van der Kolk tried to get developmental trauma disorder added to the DSM; the APA decided against it. He denounces this decision, which he thinks ignored several great studies that prove developmental trauma (ie child abuse) is much more important than anyone else thinks. I have a lot of opinions about this section.

First, I think van der Kolk downplays the importance of the APA’s philosophical commitment to categorizing by symptoms rather than cause. Consider four patients, Alice, Bob, Carol, and Dan. Alice has poor concentration caused by child abuse. Bob has poor concentration caused by bad genes. Carol throws tantrums because child abuse. Dan throws tantrums because bad genes. The current DSM would categorize Alice and Bob as ADHD, and Carol and Dan as intermittent explosive disorder. Van der Kolk would like to classify Alice and Carol as having Developmental Trauma Disorder, and Bob and Dan as…I don’t know. Bad Gene Disorder? Seems sketchy. When the APA decides not to do that, they’re not necessarily rejecting the seriousness of child abuse, only saying it’s not the kind of thing they build their categories around.

Second, van der Kolk really does not come across as a great source about the effects of development. He does not mention the possibility that links between parent behavior and child pathology might be genetic (ie a disordered parent is more likely to abuse their child, and to pass on genes for disordered behavior). In fact, he is weirdly and vocally ignorant about genetics in general, dismissing the entire field because “after thirty years and millions upon millions of dollars worth of research, we have failed to find consistent genetic patterns for schizophrenia – or for any psychiatric illness, for that matter”. When TBKtS was published in 2014, we already know with certainty that schizophrenia was about 80% genetic, and at least 15 genes had been identified as especially likely to be involved; today we know hundreds and can even make primitive polygenic predictors. The only gene he considers sympathetically is good old 5-HTTLPR, which he says proves that genes have different effects in children with vs. without abuse histories (like everything else about 5-HTTLPR, this has since been proven false). He shows total lack of interest in behavioral genetics and the challenge it raises to his hypothesis.

This is a very pre-replication crisis book. I don’t hold this against the author, I don’t think anyone’s really proud of what they believed pre-replication crisis, but it’s undoubtedly a product of its time. Mirror neurons, candidate genes, left- vs right-brained people, etc all make dramatic appearances. Nothing (except the genetics parts) are inexcusable or even certainly wrong, but all of them together concern me. And several of the book’s key studies are contradicted by later, larger studies. Van der Kolk talks about how childhood trauma decreases IQ, but some pretty good studies say it doesn’t. Even the studies that have passed the test of time look a little weird. The Adverse Childhood Experiences study found that obesity and other seemingly nonpsychiatric diseases were linked to child abuse, and recent studies confirm this – but the controls for socioeconomic status are always insufficient, and there’s surprisingly little shared environmental component. I’m biased about this, everyone’s biased, but part of the book was meant to prove that child abuse mattered shockingly more than you thought it possibly could, and that part was wasted on me.

II.

Fine, okay, drop that hobby horse, what does this book have to say about PTSD?

The book stressed the variety of responses to PTSD. Some people get anxious. Some people get angry. But a lot of people, whatever their other symptoms, also go completely numb. They are probably still “having” “emotions” “under” “the” “surface”, but they have no perception of them. Sometimes this mental deficit is accompanied by equally surprising bodily deficits. Van der Kolk describes a study on stereoagnosia in PTSD patients: if blindfolded and given a small object (like a key), they are unable to recognize it by feel, even though this task is easy for healthy people. Sometimes this gets even more extreme, like the case of a massage therapy patient who did not realize they were being massaged until the therapist verbally acknowledged she had started.

The book is called *The Body Keeps The Score*, and it returns again and again to the idea of PTSD patients as disconnected from their bodies. The body sends a rich flow of information to the brain, which is part of what we mean when we say we “feel alive” or “feel like I’m in my body”. In PTSD, this flow gets interrupted. People feel “like nothing”. For example:

> I don’t know what I feel, it’s like my head and body aren’t connected. I’m living in a tunnel, a fog, no matter what happens it’s the same reaction – numbness, nothing. Having a bubble bath and being burned or raped is the same feeling.

Or, borrowed from one of William James’ patients:

> I have no human sensations. I am surrounded by all that can render life happy and agreeable, still to me the faculty of enjoyment and of feeling is wanting. Each of my senses, each part of my proper self, is as it were separated from me and can no longer afford me any feeling; this impossibility seems to depend upon a void which I feel in the front of my head, and to be due to the diminuition of the sensibility over the whole surface of my body, for it seems to me that I never actually reach the objects that I touch. All this would be a small matter enough, but for its frightful result, which is that of the impossibility of any other kind of feeling and of any sort of enjoyment, although I experience a need and desire of them that render my life an incomprehensible torture.

One other new thing I learned about PTSD is the importance of immobilization. Van der Kolk thinks that traumas are much more likely to cause PTSD when the victim is somehow unable to respond to them. Enemy soldiers shooting at you and you are running away = less likelihood of trauma. Enemy soldiers shooting at you and you are hiding motionless behind a tree = more likelihood of trauma. Speculatively, your body feels like its going into trauma mode hasn't gotten you to take the right actions, and so the trauma mode cannot end.

There's some discussion of the neurobiology of all this, but it never really connects with the vividness of the anecdotes. A lot of stuff about how trauma causes the lizard brain to inappropriately activate in ways the rational brain can't control, how your "smoke detector" can be set to overdrive, all backed up with the proper set of big words like "dorsolateral prefrontal cortex" – but none of it seemed to reach the point where I felt like I was making progress to a gears-level explanation. I felt like the level on which I wanted an explanation of PTSD, and the level at which van der Kolk was explaining PTSD, never really connected; I can’t put it any better than that.

Why does PTSD exist? “The brain isn’t prepared to feel emotions as intense as…” Yes it is! Trauma is as old as living creatures; war, disaster, bullying, and rape far predate *homo sapiens*. Even if child abuse is rare in hunter-gatherer tribes (as some optimistic anthropologists claim) killing all the adults in a tribe and enslaving their children is pretty common, which cashes out to kids getting abused. Our evolutionary history should have prepared us incredibly well for all of this; the brain “getting stuck” in fear mode after a particularly bad trauma should be no more likely than the legs “getting stuck” in running mode after a particularly long chase.

And why would the body be so confused by the right action being “hide” or “accept the pain and abuse” rather than “run” or “fight”? The safest action has been “hide” or “accept the pain and the abuse” in a pretty good fraction of traumatic events since humanity came down from the trees.

And why should the consequences of this be the body going numb? Why not other things that seem more like the consequences of garden-variety acute or chronic stress?

I missed any answers that TBKtS might have contained to questions like these, and so a lot of its neurobiology ended up feeling more like a random collection of simplified facts than like real enlightenment.

III.

But all of this would be excusable if TBKtS had answered the most important question: how do you treat PTSD? There are a wide variety of proposed methods, and I was looking forward to having an authority like van der Kolk sort through the evidence for and against each.

Instead, I felt like he rejected every conventional treatment on the grounds that they didn’t treat the root problem, then waxed rhapsodic about every single weird alternative treatment and how it was a perfect miracle cure that truly gave patients their lives back. I understand that he may just be presenting the alternative treatments that he found most effective, but something about the style here really turned me off.

There are a *lot* of alternative treatments for PTSD. Neurofeedback, where you attach yourself to a machine that reads your brain waves and try to explore the effect your thoughts have on brain wave production until you are consciously able to manipulate your neural states. Internal family systems, where a therapist guides you through discovering “parts” of yourself (think a weak version of multiple personalities), and you talk to them, and figure out what they want, and make bargains with them where they get what they want and so stop causing mental illness. Eye movement directed reprocessing (alternative when the book was written, now basically establishment) where you move your eyes back and forth while talking about your trauma, and this seems to somehow help you process it better. Acupuncture. Massage. Yoga.

There was a thing called “PBSP psychomotor therapy”, where the therapist would create “tableaus” representing people’s traumas. They would enlist an actor to play the victim’s abusive father, then another actor to play an idealized version of their father who didn’t abuse them and was always there when they needed them, then have them recite formulaic lines that “played their part” in the remembered (or alternative hypothetical) versions of the patient’s trauma. Gradually they would progress from the real trauma to a version where things had worked out better, with the therapist discussing the patient’s reaction the whole time.

There was a chapter on community theater, where troubled youth who would otherwise be sent to jail were instead asked to put on a Shakespeare production. This encountered some early hitches:

> We were shocked to discover that, in scenes where someone was in physical danger, the students always sided with the aggressors. Because they could not tolerate any sign of weakness in themselves, they could not accept it in others. They showed nothing but contempt for potential victims, yelling things like “Kill the bitch, she deserves it,” during a skit about dating violence.
>
> At first some of the actors wanted to give up – it was simply too painful to see how mean these kids were – but they stuck it out, and I was amazed to see how they gradually got the students to experiment, however reluctantly, with new roles. Toward the end of the program, a few students were even volunteering for parts that involved showing vulnerability or fear.

The traumatic incidents in Shakespeare’s work helped them come to terms with their own difficult history:

> As we’ve seen, the essence of trauma is feeling godforsaken, cut off from the human race. Theater involves a collective confrontation with the realities of the human condition. As Paul Griffin, discussing his theater program for foster care children, told me: “The stuff of tragedy in theater revolves around coping with

> betrayal, assault, and destruction. These kids have no trouble understanding what Lear, Othello, Macbeth, or Hamlet is all about." In Tina Packer's words: "Everything is about using the whole body and having other bodies resonate with your feelings, emotions, and thoughts." Theater gives trauma survivors a chance to connect with one another by deeply experiencing their common humanity."

Each of these stories about an alternative therapy was, on its own, inspiring. But after chapter after chapter on these, plus other even weirder things, you start to wish there was at least one alternative therapy that Bessel van der Kolk didn't like, or one conventional therapy that he did.

This is a very pre-replication-crisis book. In these more cynical days, we know that the first few studies on any technique – usually done in an atmosphere of frothy excitement, by the technique's most fervent early adapters – are always highly positive. And later studies – done in an atmosphere of boredom, by large multi-center consortia – are almost always disappointing. Half the time van der Kolk is so excited about the miraculous life-changing potential of the latest alternative therapy that he doesn't list studies at all. The other half of the time, the studies are there to support his enthusiasm. But can they be trusted?

Overall, so many bizarre methods seemed to work so well (with no examples of anything that didn't work) that it was hard for me to figure out how this book should affect my treatment decisions. Find the closest person in a robe and wizard hat and send all of my trauma patients to them, because every alternative therapy works equally well as long as it's weird? This might actually be a good lesson, there are a lot of things in psychiatry where as long as people feel drawn in and "validated" the treatment works.

But I’m annoyed I have to ponder this kind of thing on my own rather than have the book take a step back and wonder about these kinds of questions.

[Update, written a few weeks after the rest of this post: maybe it *is* all wizardry. I recommended this book to a severely traumatized patient of mine, who had not benefited from years of conventional treatment, and who wanted to know more about their condition. The next week the patient came in, claiming to be completely cured, and displaying behaviors consistent with this. They did not use any of the techniques in this book, but said that reading the book helped them figure out an indescribable mental motion they could take to resolve their trauma, and that after taking this mental motion their problems were gone. I’m not sure what to think of this or how much I should revise the negative opinion of this book which I formed before this event.]

Maybe the most consistent lesson from this book’s tour of successful alternative therapies – keeping with the theme of the title – is that it’s important for PTSD patients to get back in touch with their bodies. Massage therapy, yoga, and acupuncture addressed this directly, usually creating gentle, comfortable sensations that patients could take note of to gradually relax the absolute firewall between bodily sensation and conscious processing. Some of the other methods – the community theater, maybe even the internal family systems – seemed like tricks to get people afraid of emotions back in touch with their emotions anyway: “Oh, you’re not going to be feeling *your* emotions, just emotions from Macbeth or Hamlet or this other personality living in your mind”. I don’t know how plausible this interpretation is.

IV.

Overall, I was not too impressed with this book. The highlight was van der Kolk's personal reminisces from the fight to get PTSD recognized as a real disease – but some of them were so over-the-top that I would have liked to triangulate them with a more objective history. The sections with the symptomatology and neurobiology of PTSD were helpful in exploring the boundaries of the syndrome, but didn't make me feel like I *really* understood what was going on. The sections on the dangers of child abuse were a good knock-down of some hypothetical "child abuse isn't really that bad" position, but I don't know anyone who holds that position, and some of the research seemed questionable. And the section on treatment was so glowing about everything that it was hard to draw any specific conclusions.

Maybe a broader concern is that I seem to inhabit a different world than van der Kolk. All of his patients showed bizarre and florid sequelae from serious trauma. My patients seem to discuss their trauma with comparative equanimity, have only the usual psychiatric symptoms (depression, anxiety, etc) and not experience much benefit from the weirder alternative therapies they try. Some of this might be van der Kolk being a better doctor than I am, or having sicker patients. But I'm concerned about this because van der Kolk seems pretty good at doing what he does, and I would like to be able to inhabit his world insofar as he's able to get good results in it. But insofar as my goal is to become more like Bessel van der Kolk, I was surprised how little this book helped guide me along that journey.

I think my actual takeaway is to screen for trauma more carefully, especially in patients who seem anhedonic or numb, and to recommend they go to a trauma clinic. There are a lot of places like this (I sometimes send patients to this one in Berkeley), and they practice a lot of the weirder alternative therapies that van der Kolk mentions (in fact, van der Kolk seems to work at/lead a very similar type of institution in Massachussetts). Whether or not these work for everybody, I think everybody deserves a chance at them, and I should take them more seriously at least until I get a better sense of the terrain here myself.

THE BODY KEEPS THE SCORE
Mind, brain and body in the transformation of trauma
BESSEL VAN DER KOLK

Autism And Intelligence: Much More Than You Wanted To Know

[Thanks to Marco DG for proofreading and offering suggestions]

I.

Several studies have shown a genetic link between autism and intelligence; genes that contribute to autism risk also contribute to high IQ. But studies show autistic people generally have lower intelligence than neurotypical controls, often much lower. What is going on?

First, the studies. This study from UK Biobank finds a genetic correlation between genetic risk for autism and educational attainment (r = 0.34), and between autism and verbal-numerical reasoning (r = 0.19). This study of three large birth cohorts finds a correlation between genetic risk for autism and cognitive ability (beta = 0.07). This study of 45,000 Danes finds that genetic risk for autism correlates at about 0.2 with both IQ and educational attainment. These are just three randomly-selected studies; there are too many to be worth listing.

The relatives of autistic people will usually have many of the genes for autism, but not be autistic themselves. If genes for autism (without autism itself) increase intelligence, we should expect these people to be unusually smart. This is what we find; see Table 4 here. Of 11 types of psychiatric condition, only autism was associated with increased intelligence among relatives. This intelligence is shifted towards technical subjects. About 13% of autistic children (in this sample from whatever social stratum they took their sample from) have fathers who are engineers, compared to only 5% of a group of (presumably well-matched?) control children (though see the discussion here) for some debate over how seriously to take this; I am less sure this is accurate than most of the other statistics mentioned here.

Further (indirect) confirmation of the autism-IQ link comes from evolutionary investigations. If autism makes people less likely to reproduce, why would autism risk genes stick around in the human population? Polimanti and Gelemter (2017) find that autism risk genes aren't just sticking around. They are being *positively selected*, ie increasing with every generation, presumably because people with the genes are having more children than people without them. This means autism risk genes must be doing something good. Like everyone else, they find autism risk genes are positively correlated with years of schooling completed, college completion, and IQ. They propose that the reason evolution favors autism genes is that they generally increase intelligence.

But as mentioned before, autistic people themselves on average have lower intelligence. One study found that 69% of autistic people had an IQ below 85 (the average IQ of a high school dropout). Only 3% of autistic people were found to have IQs above 115, even though 15% of the population should be at this level.

These numbers should be taken with very many grains of salt. First, IQ tests don't do a great job of measuring autistic people. Their intelligence tends to be more imbalanced than neurotypicals', so IQ tests (which rely on an assumption that most forms of intelligence are correlated) are less applicable. Second, even if the test itself is good, autistic people may be bad at test-taking for other reasons – for example, they don't understand the directions, or they're anxious about the social interaction required to answer an examiner's quetsions. Third, and most important, there is a strong selection bias in the samples of autistic people. Many definitions of autism center around forms of poor functioning which are correlated with low intelligence. Even if the definition is good, people who function poorly are more likely to seek out (or be coerced into) psychiatric treatment, and so are more likely to be identified. In some sense, all "autism has such-and-such characteristics" studies are studying the way people like to define autism, and tell us nothing about any underlying disease process. I talk more about this in parts 2 and 3 here.

But even adjusting for these factors, the autism – low intelligence correlation seems too strong to dismiss. For one thing, the same studies that found that relatives of autistic patients had higher IQs find that the autistic patients themselves have much lower ones. The existence of a well-defined subset of low IQ people whose relatives have higher-than-predicted IQs is a surprising finding that cuts through the measurement difficulties and suggests that this is a real phenomenon.

So what is going on here?

II.

At least part of the story is that there are at least three different causes of autism.

1. The "familial" genes mentioned above: common genes that increase IQ and that evolution positively selects for.

2. Rare "de novo mutations", ie the autistic child gets a new mutation that their non-autistic parent doesn't have. These mutations are often very bad, and are quickly selected out of the gene pool (because the people who have them don't reproduce). But "quickly selected out of the gene pool" doesn't help the individual person who got one of them, who

tends to end up severely disabled. In a few cases, the parent gets the de novo mutation, but for whatever reason doesn't develop autism, and then passes it onto their child, who does develop autism.

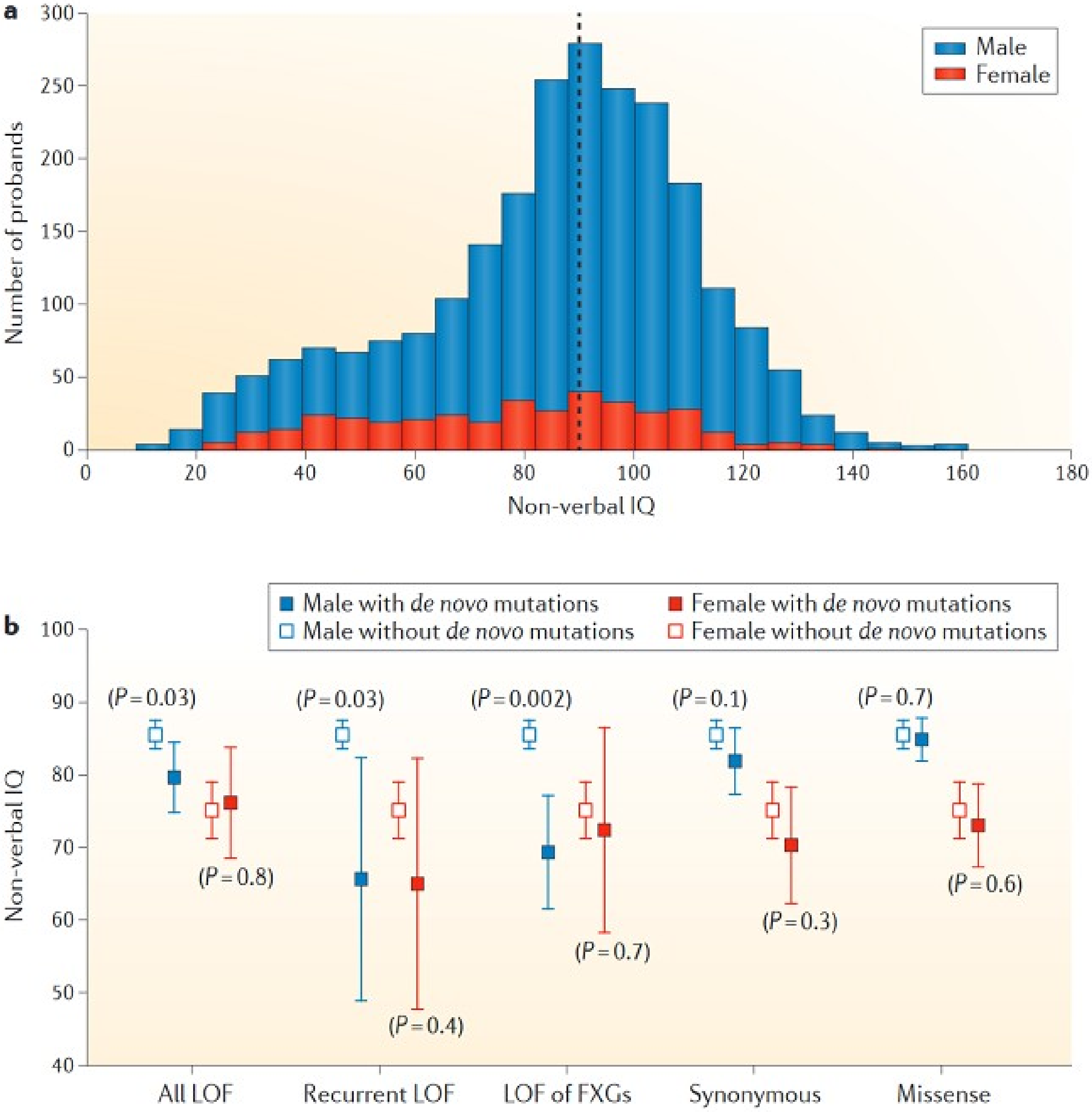
3. Non-genetic factors. The best-studied are probably obstetric complications, eg a baby gets stuck in the birth canal and can't breathe for a long time. Pollution, infection, and trauma might also be in this basket.

These three buckets and a few other less important factors combine to determine autism risk for any individual. Combining information from a wide variety of studies, Gaugler et al estimate that about 52% of autism risk is attributable to ordinary "familial" genes, 3% to rare "de novo" mutations, 4% to complicated non-additive genetic interaction effects, and 41% "unaccounted", which may be non-genetic factors or genetic factors we don't understand and can't measure. This study finds lower heritability than the usual estimates (which are around 80% to 90%; the authors are embarrassed by this, and in a later study suggest they might just have been bad at determining who in their sample did or didn't have autism. While their exact numbers are doubtful, I think the overall finding that common familial genes are much more important than rare de novo mutations survives and is important.

Most cases of autism involve all three of these factors; that is, your overall autisticness is a combination of your familial genes, mutations, and environmental risk factors.

One way of resolving the autism-intelligence paradox is to say that familial genes for autism increase IQ, but de novo mutations and environmental insults decrease IQ. This is common-sensically true and matches previous research into all of these factors. So the only question is whether the size of the effect is enough to fully explain the data – or whether, even after adjusting out the degree to which autism is caused by mutations and environment, it still decreases IQ.

Ronemus et al (2014) evaluate this:



They find that even autistic people without *de novo* mutations have lower-than-average IQ. But they can only screen for *de novo* mutations they know about, and it could be that they just missed some.

Here’s another set of relevant graphs:

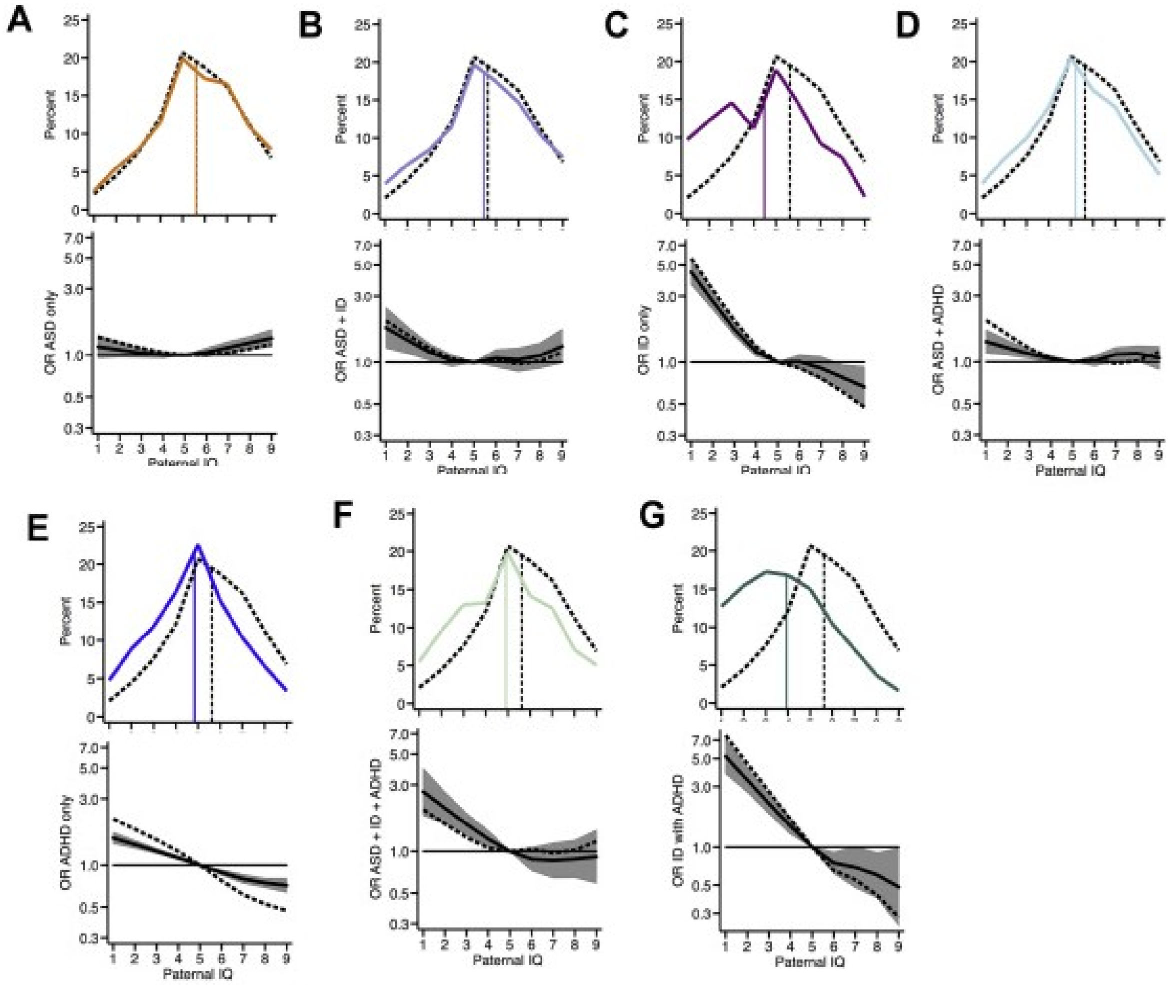


Figure 2. Relationship Between Paternal IQ and Offspring Risk of Neurodevelopmental Disorders

Note: *Neurodevelopmental disorders were examined as seven mutually exclusive diagnostic outcomes. For each outcome, the top panel displays the distribution of paternal IQ among affected offspring (solid line) compared to the distribution of paternal IQ among offspring not affected by any of the diagnostic outcomes studies here (dashed line). The bottom panel displays the risk of each outcome according to paternal IQ level, flexibly fit using a restricted cubic spline model with five knots and IQ =5 (average intelligence) set as the referent. The dotted line represents the unadjusted estimate of the relationship between each outcome and paternal IQ. The solid line represents the fully adjusted model (adjusted for children's characteristics: sex, birth year, birth order; and for family characteristics: maternal age, paternal age, highest parental education level, family income quintile, maternal migration status, and parental psychiatric history). The gray bands represent the 95% confidence interval for the fully adjusted model. Results are shown for (A) autism spectrum disorder (ASD) without co-occurring intellectual disability/attention-deficit/hyperactivity disorder (ID/ADHD) (ASD only), (B) ASD with co-occurring ID (ASD + ID), (C) ID without ASD/ADHD (ID only), (D) ASD with co-occurring ADHD (ASD + ADHD), (E) ADHD without ASD/ID (ADHD only), (F) ASD with co-occurring ID and ADHD (ASD + ID + ADHD), and (G) ID with co-occurring ADHD (ID + ADHD). OR = odds ratio.*

This one comes from Gardner et al (2019), which measures the cognitive ability of the fathers of autistic people and disaggregates those with and without intellectual disability. In Graph A, we see that if a child has autism (but not intellectual disability), their likelihood of having a father with any particular IQ (orange line) is *almost* the same as the likelihood of a neurotypical child having a father of that IQ (dotted line). Disguised in that “almost” is a very slight tendency for fathers to be unusually intelligent, plus a (statistically insignificant) tendency for them to be unusually unintelligent. For reasons that don’t entirely make sense to me, if instead we look at the likelihood of the father to be a certain intelligence (bottom graph, where dark line surrounded by gray confidence cloud is autistic people’s fathers, and dotted line is neurotypical people’s fathers) it becomes more obvious that more intelligent people are actually a little more likely to have autistic children (though less intelligent people are also more likely.

(remember that “no intellectual disability” just means “IQ over 70”, and so many of these not-intellectually-disabled people may still have low intelligence – I wish the paper had quantified this)

Graph B is the same thing, but with people have have autism *with* intellectual disability. Now there is a very strong effect towards their fathers being less intelligent than usual.

This confuses me a little. But for me the key point is that high-intelligence fathers show a trend (albeit not significant in this study) to be more likely than average to have children with autism *and* intellectual disability.

These questions interest me because I know a lot of people who are bright nerdy programmers married to other bright nerdy programmers, and sometimes they ask me if their children are at higher risk for autism. While their children are clearly at higher risk for autistic *traits*, I think they want to know whether they have higher risk for the most severe forms of the syndrome, including intellectual disability and poor functioning. If we take the Ronemus and Gardner studies seriously, the answer seems to be yes. The Gardner study seems to suggest it’s a very weakly elevated risk, maybe only 1.1x or 1.2x relative risk. But the Gardner study also ceilings off at 90th percentile intelligence, so at this point I’m not sure what to tell these people.

III.

If Ronemus isn’t missing some obscure *de novo* mutations, then people who get autism solely by accumulation of common (usually IQ-promoting) variants still end up less intelligent than average. This should be surprising; why would too many intelligence-promoting variants cause a syndrome marked by low intelligence? And how come it’s so inconsistent, and many people have naturally high intelligence but aren’t autistic at all?

One possibility would be something like a tower-vs-foundation model. The tower of intelligence needs to be built upon some kind of mysterious foundation. The taller the tower, the stronger the foundation has to be. If the foundation isn’t strong enough for the tower, the system fails, you develop autism, and you get a collection of symptoms possibly including low intelligence. This would explain low-functioning autism from de novo mutations or obstetric trauma (the foundation is so weak that it fails no matter how short the tower is). It would explain the association of genes for intelligence with autism (holding foundation strength constant, the taller the tower, the more likely a failure). And it would also explain why there are many extremely intelligent people who don’t have autism at all (you can build arbitrarily tall towers if your foundation is strong enough).

I’ve only found one paper that takes this model completely seriously and begins speculating on the nature of the foundation. This is Crespi 2016, Autism As A Disorder Of High Intelligence. It draws on the VPR model of intelligence, where g (“general intelligence”) is divided into three subtraits, v (“verbal intelligence”), p (“perceptual intelligence”), and r (“mental rotation ability”) – despite the very specific names each of these represents ability at broad categories of cognitive tasks. Crespi suggests that autism is marked by an imbalance between P (as the tower) and V + R (as the foundation). In other words, if your perceptual intelligence is much higher than your other types of intelligence, you will end up autistic.

It doesn’t really present much evidence for this other than that autistic people seem to have high perceptual intelligence. Also, it doesn’t really look like autistic people are worse at mental rotation. Also, the Gardner paper has analyzed autistic patients’ fathers by subtype of intelligence, and there is a nonsignificant but pretty suggestive tendency for them to have higher-than-normal verbal intelligence; certainly no signs of high verbal intelligence *preventing* autism. I can’t tell if this is evidence against Crespi or whether since all intellectual abilities are correlated this is just the shadow of their high perceptual intelligence, and if we directly looked at perceptual-to-verbal ratio we would see it was lower than expected. Also also, Crespi is one of those scientists who constantly has much more interesting theories than anyone else (eg), and this makes me suspicious.

Overall I would be surprised if this were the real explanation for the autism-and-intelligence paradox, but it gets an A for effort.

Conclusions

1. The genes that increase risk of autism are disproportionately also genes that increase intelligence, and vice versa (~100% confidence)

2. People diagnosed with autism are less intelligent than average (~100% confidence, leaving aside definitional complications)

3. Some of this effect is because autism is caused both by normal genes and by *de novo* mutations and environmental insults, and the *de novo* mutations and environmental insults definitely decrease intelligence. Every autism case is caused by some combination of these three factors, and the more it is caused by normal genes, the more intelligence is likely to be preserved (~100% confidence)

4. This is not the whole story, and even cases of autism that are caused entirely or mostly by normal genetics are associated with unusually low IQ (80% confidence)

5. This can best be understood through a tower-versus-foundation model where higher intelligence that outstrips the ability of some mysterious foundation to support it will result in autism (25% confidence)

6. The specific way the model plays out may be through perceptual intelligence out of balance with verbal and rotational intelligence causing autism (3% confidence)

More Intuition-Building On Non-Empirical Science: Three Stories

[Followup to: Building Intuitions On Non-Empirical Arguments In Science]

I.

In your travels, you arrive at a distant land. The chemists there believe that when you mix an acid and a base, you get salt and water, *and* a star beyond the cosmological event horizon goes supernova. This is taught to every schoolchild as an important chemical fact.

You approach their chemists and protest: why include the part about the star going supernova? Why not just say an acid and a base make salt and water? The chemists find your question annoying: your new "supernova-less" chemistry makes exactly the same predictions as the standard model! You're just splitting hairs! Angels dancing on pins! Stop wasting their time!

"But the part about supernovas doesn't constrain expectation!" Yes, say the chemists, but removing it doesn't constrain expectation either. You're just spouting random armchair speculation that can never be proven one way or the other. What part of "stop wasting our time" did you not understand?

Moral of the story: It's too glib to say "There is no difference between theories that produce identical predictions". You actually care a lot about which of two theories that produce identical predictions is considered true.

II.

Later in your travels, you come to another land. The paleontologists here believe the Devil planted dinosaur fossils to trick humans into doubting Creation.

You approach the paleontologists and argue the same point you argued with the chemists on your last stop – that if two theories make identical predictions, it's still important to go with the simpler one.

To your surprise, the paleontologists know and agree. "Of course!" they tell you. "And in the dinosaur theory, there must have been, like, millions or even billions of dinosaurs. But the Devil theory explains everything with just one Devil."

You argue that it doesn't work that way, but the paleontologists insist that it does. After all, Occam says not to multiply entities beyond necessity. And if the dinosaur theory posits a billion dinosaurs, that's 999,999,999 more entities than are necessary to explain all those bones.

Moral of the story: "Choose the simpler of two theories that make identical predictions" isn't trivial. You actually have to understand some philosophy in order to figure out which of two theories is simpler.

III.

You return home and curl up in front of the fire with a good book on quantum mechanics.

Renowned physicist Sean Carroll jumps out from behind you, and exclaims: "Don't you realize that single-world interpretations of quantum mechanics make both the errors that you fought against abroad?"

You are startled. “This room is locked,” you tell him. “And how did you know what I was doing abroad? Wait a second. Are you secretly the Devil?”

“Untestable, therefore irrelevant!” says Carroll. You wonder if he has always had bright orange eyes. “But being indifferent between ‘wavefunction branches’ and ‘wavefunction branches, and then somewhere we can’t see it one branch mysteriously collapses’ is the same kind of error as being indifferent between ‘acid and base make salt’ and ‘acid and base make salt and water, and then somewhere we can’t see it a star mysteriously goes supernova’.”

He stomps his foot for emphasis, and something falls out of his pocket. Is that a dinosaur bone? He quickly reaches down and pockets it again.

“And,” he adds “preferring collapse interpretations to many-worlds because there are fewer universes – that’s like preferring the Devil theory to dinosaurs because it involves fewer entities. It’s optimizing over the wrong thing! You’re not *literally* trying to come up with a theory with as few entities as possible! You’re trying to come up with one that has as few extra moving parts as possible. The process that makes wavefunctions collapse is an extra assumption! Now if you’ll excuse me, I’ve got to go plant this” – he taps the bone “in a sedimentary rock formation in China”. He vanishes in a puff of smoke. Can all quantum physicists do that?

Moral of the story: Applying the two previous morals consistently lets you prefer the many worlds interpretation of quantum mechanics without having to worry about this being “untestable”.

Book Review: All Therapy Books

*[**Related:** CBT In The Water Supply, Scientific Freud, Book Review: Method Of Levels, Different Worlds]*

I.

All therapy books start with a claim that their form of therapy will change everything. Previous forms of therapy have required years or even decades to produce ambiguous results. *Our* form of therapy can produce total transformation in five to ten sessions! Previous forms of therapy have only helped ameliorate the stress of symptoms. *Our* form of therapy destroys symptoms at the root!

All therapy books bring up the Dodo Bird Verdict – the observation, confirmed in study after study, that all psychotherapies are about equally good, and the only things that matters are "nonspecific factors" like how much patients like their therapist. Some people might think this suggests *our* form of therapy will only be about as good as other forms. This, all therapy books agree, would be a foolish and perverse interpretation of these findings. The correct interpretation is that all previous forms of therapy must be equally wrong. The only reason they ever produce good results at all is because sometimes therapists accidentally stumble into using *our* form of therapy, without even knowing it. Since every form of therapy is about equally likely to stumble into using *our* form of therapy, every other form is equally good. But now that *our* form of therapy has been formalized and written up, there is no longer any need to stumble blindly! Everyone can just use *our* form of therapy all the time, for everything! Nobody has ever done a study of *our* form of therapy. But when they do, it’s going to be amazing! Nobody has even invented numbers high enough to express how big the effect size of *our* form of therapy is going to be!

Consider the case of Bob. Bob had some standard-issue psychological problem. He had been in and out of therapy for years, tried dozens of different medications, none of them had helped at all. Then he decided to try *our* form of therapy. In his first session, the therapist asked him “Have you ever considered that your problems might be because of [the kind of thing our form of therapy says all problems are because of]?” Bob started laughing and crying simultaneously, eventually breaking into a convulsive fit. After three minutes, he recovered and proceeded to tell a story of how [everything in his life was exactly in accordance with our form of therapy’s predictions] and he had always reacted by [doing exactly the kind of thing our form of therapy predicts that he would]. Now that all of this was out in consciousness, he no longer felt any desire to have psychological problems. In a followup session two weeks later, the therapist confirmed that he no longer had any psychological problems, and had become the CEO of a Fortune 500 company and a renowned pentathlete.

Not every case goes this smoothly. Consider the case of Sarah. Sarah also has some standard-issue psychological problem. She had also been in and out of therapy for years, tried dozens of different medications, none of them had helped at all. Then she decided to try *our* form of therapy. In her first session, the therapist asked her “Have you ever considered that your problems might be because of [the kind of thing our form of therapy says all problems are because of]?” Sarah said “No, I don’t think they are.” The therapist asked “Are you sure you’re not just repressing the fact that they totally definitely are, for sure?” As soon as Sarah heard this, she gasped, and her eyes seemed to light up with an inner fire. Then she proceeded to tell a story of how [everything in her life was exactly in accordance with our form of therapy's predictions] and she had always reacted by [doing exactly the kind of thing our form of therapy predicts that she would], only she was repressing this because she was scared of how powerful she would be if she recovered. Now that all of this was out in consciousness, she no longer felt any desire to have psychological problems. In a followup session two weeks later, the therapist confirmed that she no longer had any psychological problems, and had become the hand-picked successor to the Dalai Lama and the mother of five healthy children.

Previous forms of therapy have failed because they were ungrounded. They were ridiculous mental castles built in the clouds by armchair speculators. But *our* form of therapy is based on hard science! For example, it probably acts on synapses or the hippocampus or something. Here are three neuroscience papers which vaguely remind us of our form of therapy. One day, neuroscience will catch up to us and realize that the principles of *our* form of therapy are the principles that govern the organization of the entire brain – if not all of multicellular life.

II.

Maybe I'm being unfair here. I'm basing this off a small sample of therapy books (five textbooks I can think of, plus scattered papers on psychodynamic and psychedelic therapies), and only a subset are *quite* this bad.

But my basic confusion is this: I work in a clinic with about ten therapists. Some are better than others, but all of them are competent. I send my patients to them. In a few hundred patients I've worked with, *zero* have had the sudden, extraordinary, long-lasting change that the therapy books promise. Many have benefited a little. A few would say that, over the course of years, their lives have been turned around. But sudden complete transformations? Not that much.

Of course, this fits with the therapy books' perspective. My colleagues practice normal therapy. Sometimes it's from a boring old school like CBT; other times it's "eclectic" or "supportive" or any of the other words we use to describe what we're doing when we don't know what we're doing. So maybe there are two sets of therapies: boring old therapies that ordinary people practice, and exciting new therapies that people write glowing books about. And maybe the first set really don't work (or work only a little), and the second set really is that good.

The problem is, the boring old therapies that everybody uses nowadays inspired equal excitement when they first arose. This is the point that I make in CBT In The Water Supply, and that Oliver Burkeman makes more cogently in Why CBT Is Falling Out Of Favor. Look at therapy books from the 1990s, and they were all about how CBT was a new miracle therapy that would cure your anxiety forever in a few sessions. From a cognitive therapy book:

> [When I first learned about cognitive-behavioral therapy, I thought] depression and anxiety seemed far too serious and severe for such a simplistic approach. But when I tried these methods with some of my more difficult patients, my perceptions changed. Patients who'd felt hopeless, worthless, and desperate began to recover. At first, it was hard to believe that the techniques were working, but I could not deny the fact that when my patients learned to put the lie to their negative thoughts, they began to improve. Sometimes

> they recovered right before my eyes during sessions. Patients who'd felt demoralized and hopeless for years suddenly turned the corner on their problems. I can still recall an elderly French woman who'd been bitterly depressed for more than fifty years, with three nearly-successful suicide attempts, who started shouting "Joie de vivre! Joie de vivre!" ("joy of living") one day in my office. These experiences made such a strong impact on me that I decided my calling was in clinical work rather than brain research. After considerable soul-searching, I decided to give up my research career and become a full-time clinician. Over the years, I've had more than 35,000 psychotherapy sessions with depressed and anxious patients, and I'm every bit as enthusiastic about CBT as when I first began learning about it.

But look at therapy books now, and they're all people saying "Sure, CBT barely outperforms placebo…but what about this exciting *new* therapy which blows CBT out of the water?"

Studies reflect this decline:

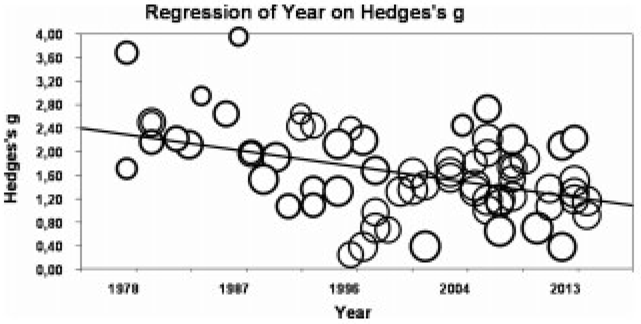


Figure 4. The plot portrays the negative change ($p < .001$) in Beck Depression Inventory effect sizes across time ($k = 61$). The size of the circles indicates the relative contribution (random weight) of each study to the analysis.

…with the [average studied effect size of CBT shrinking]() from 2.5 to 1.0 over the course of a generation. People have come up with various explanations for this. Maybe therapist quality is falling – when CBT was the hot new thing, you had to be a really plugged-in up-to-date therapist to have heard about it and to make the effort to retrain in it, so only the best therapists would practice it, but now it's the default therapy used by everyone who's just clocking it in. Maybe placebo effect is falling – when people viewed it as an astounding miracle therapy, it got astounding miracle results, but now that it's lost its luster nobody takes it seriously anymore. Maybe its ideas are spreading, so that patients come into their first session already aware of CBT insights and inoculated against them. Or maybe it's like all science, where the first studies are done quickly by true believers, and the later studies are done carefully by the Cochrane Collaboration, and so the level of hype naturally goes down.

These explanations have different practical implications. If it's all about therapist quality and placebo expectations, then you should go get the exciting new therapies described in therapy books, since their unusually-qualified therapists and unusually-high expectations will deliver you the miracle cure you're looking for.

If it's just that study quality gets better and better until we realize how crappy the exciting new therapies really are, you might as well get the boring old therapies. At least insurance probably covers them.

And they also have different philosophical implications. If it's all about therapist quality and placebo expectations, then even if it's hard to deliver high-quality therapy consistently at scale, it means high-quality therapy *is a thing*. It means that if enough factors go right at once, therapy can be the kind of powerful tool that cures someone's life-long psychiatric issues in a few sessions with a high success rate. If this is true it would be fascinating. It would be like saying that bananas cure cancer, but only if they’re *really fresh* bananas. Even if there are practical issues in getting every cancer patient a banana that’s fresh enough, you still want to take a step back and think “Whoa, what’s up with this?”

I can only say that I’ve had a few patients try the exciting new therapies, and none of them have reported miracle cures. They’ve all maybe gotten a little better over long periods, same as the boring old therapies. This makes me think it’s more likely that early results from the exciting new therapies get oversold, not that some combination of therapist skill and excitement makes them go shockingly well. And the Efficient Market agrees with my low estimation, given that therapists aren’t rushing to learn these new strategies and patients aren’t rushing to use them.

But the therapy books still confuse me. They’re full of stories of incredible instant cures, with the authors assuring us that these are all real and typical of their experience. How can you get this from merely “stretching the truth”, as opposed to outright data falsification? Are therapy book authors blatantly lying? I try to have a really low prior on this sort of thing, but I’m not sure.

Therapy books are often written by the researcher who invented the therapy. I imagine if you invent a therapy yourself, then it perfectly fits your personality and communication style, you believe in it wholeheartedly, and you understand every piece of it from the ground up. You’re also probably a really exceptional and talented person who’s obsessed with psychotherapy and how to make it better. So maybe they get results nobody else can replicate?

But that still raises the philosophical implication of it being possible, for somebody, to consistently produce dramatic change through therapy. This still bothers me a lot.

III.

Most therapy books share some assumptions, so deep as to be unspoken: current problems serve some purpose related to past traumas.

Different therapies take this in different directions. Some view problems as a passive residue of past traumas: for example you were abused as a child, that filled you with stress and rage, and now you take that out on other people and yourself. Others view them as maladaptive learning from past trauma: for example, you were abused as a child, that taught you that other people would hurt you if you opened up to them, so you never open up to anybody. I don’t know the official name for this, but let’s call it historicism: symptoms are the result of something that happened in a patient’s life history.

Some weak forms of historicism are obviously true. Many (though not all) phobias began with a clear incident where the patient was endangered by the phobic object; someone mauled by a dog as a child who then has cynophobia as an adult is hardly a medical mystery. Many (though not all) depressions are precipitated by some depressing event. And post-traumatic stress disorder has the historical perspective right there in the name; at the very least, going through trauma dysregulates something inside you. But it’s a long way from there to saying that a patient’s psychosomatic blindness is caused by persistent shame at having seen their parents having sex thirty years earlier, or something like that.

And some therapy books go beyond historicism into purposefulism: symptoms serve some quasi-logical purpose relating to the life history. I recently read a therapy book that included a case like this. Bob had a history of failing at work. He would go from job to job, making various mistakes and doing crappy work until he got fired. He went to a therapist for help. During the therapy, it came out that Bob's abusive father had always pushed him really hard to succeed. The therapist suggested that maybe Bob failed at work to send a message to his father; ie to prove that his father's abusive parenting had been a bad idea and would not make Bob successful. The therapist asked Bob to imagine confronting his father about this. After he worked through his anger at his father, Bob was able to succeed at work. In this story, the apparently dysfunctional symptom (failing at work) ended up having a legible purpose within Bob's life history (it helped him send a message to his father). Only by teasing out the purpose and finding some other way to achieve it could the dysfunctional behavior be prevented.

A non-historical, non-purposeful account might argue that Bob failed at work because he was bad at work. Maybe he was bad at the specific jobs he was holding (in which case he should get more training). Maybe he was bad at social skills (in which case he should learn to communicate better). Maybe he had ADHD and kept getting distracted (in which he should get treatment for ADHD). In any case, him being bad at work isn't related to any past traumas or serving any hidden purposes. It's just an unfortunate fact.

I am constantly worried by the history of how many things we historically applied historical-purposeful reasoning to, totally confident at the time that our explanations made sense – which we now know are not historical-purposeful at all. Psychologists "knew" that autism was caused by distant mothers, and schizophrenia by overbearing mothers, right up until we discovered both conditions are about 80%

genetic. And when they “knew” these things, they were able to come up with long lists of how exactly each individual patient fit the mold, and reported great progress by helping patients overcome their maternal attachment issues. Back when homosexuality was considered a disorder, historical-purposeful therapists would tell gay people patients they must be so angry at their mother that they had sworn off all female companionship and switched to men instead as a way of sending her a giant “F–K YOU” message; while homosexuality is mostly *not* genetic, few people today think this is a plausible explanation.

I sometimes see if I can come up with these kinds of historical-purposeful accounts of my patients’ symptoms. These always fit into place freakishly well – so well that either the historical-purposeful perspective is completely true, or there is some very strong bias that makes it extra-convincing despite its falsehood. But we already know there’s some very strong bias that makes it extra-convincing despite its falsehood! That bias must have been at work in all the therapists who applied historical-purposeful narratives to autistics, schizophrenics, and gays! At some point I notice the road I’m on is littered with skulls and start wondering if I should reconsider.

All therapy books propose an answer: the proof is that the patients get better. But my patients do not get better. When I tell them the historical-purposeful accounts I have devised for their symptoms, they usually shrug and say it sounds plausible and they’ve thought along those lines before, but what are they going to do? When I try all the exciting new therapies on them, they just sort of nod, say that this sounds like an interesting perspective, and then go off and keep having symptoms. It’s very rude!

I’ve told this story before: when I was a teenager, I got really into pseudohistory for a while. What snapped me out of it wasn’t the sober historians, who totally went AWOL on their job of explaining why they were right and the whackos were wrong. It was that a bunch of mutually exclusive pseudohistories all sounded equally plausible: the Pyramids couldn’t have been built by Atlanteans *and* Lemurians *and* mole-people! At that point I was able to halt, melt, catch fire, and realize there was something really wrong with my reasoning processes, which I continue to worry about and work on twenty years later.

I bring this up because I’m going to be reviewing some specific psychotherapy books. Each of them on their own can be convincing. But they should be taken in the context of All Therapy Books, which as a category are pretty worrying.

Mental Mountains

I.

Kaj Sotala has an outstanding review of *Unlocking The Emotional Brain*; I read the book, and Kaj's review is better.

He begins:

> UtEB's premise is that much if not most of our behavior is driven by emotional learning. Intense emotions generate unconscious predictive models of how the world functions and what caused those emotions to occur. The brain then uses those models to guide our future behavior. Emotional issues and seemingly irrational behaviors are generated from implicit world-models (schemas) which have been formed in response to various external challenges. Each schema contains memories relating to times when the challenge has been encountered and mental structures describing both the problem and a solution to it.

So in one of the book's example cases, a man named Richard sought help for trouble speaking up at work. He would have good ideas during meetings, but felt inexplicably afraid to voice them. During therapy, he described his narcissistic father, who was always mouthing off about everything. Everyone hated his father for being a fool who wouldn't shut up. The therapist conjectured that young Richard observed this and formed a predictive model, something like "talking makes people hate you". This was overly general: talking only makes people hate you if you talk incessantly about really stupid things. But when you're a kid you don't have much data, so you end up generalizing a lot from the few examples you have.

When Richard started therapy, he didn’t consciously understand any of this. He just felt emotions (anxiety) at the thought of voicing his opinion. The predictive model output the anxiety, using reasoning like “if you talk, people will hate you, and the prospect of being hated should make you anxious – therefore, anxiety”, but not any of the intermediate steps. The therapist helped Richard tease out the underlying model, and at the end of the session Richard agreed that his symptoms were related to his experience of his father. But knowing this changed nothing; Richard felt as anxious as ever.

Predictions like “speaking up leads to being hated” are special kinds of emotional memory. You can rationally understand that the prediction is no longer useful, but that doesn’t really help; the emotional memory is still there, guiding your unconscious predictions. What should the therapist do?

Here *UtEB* dives into the science on memory reconsolidation.

Scientists have known for a while that giving rats the protein synthesis inhibitor anisomycin prevents them from forming emotional memories. You can usually give a rat noise-phobia by pairing a certain noise with electric shocks, but this doesn’t work if the rats are on anisomycin first. Probably this means that some kind of protein synthesis is involved in memory. So far, so plausible.

A 2000 study found that anisomycin could also erase existing phobias in a very specific situation. You had to “activate” the phobia – get the rats thinking about it really hard, maybe by playing the scary noise all the time – and then give them the anisomycin. This suggested that when the memory got activated, it somehow “came loose”, and the brain needed to do some protein synthesis to put it back together again.

Thus the idea of memory reconsolidation: you form a consolidated memory, but every time you activate it, you need to reconsolidate it. If the reconsolidation fails, you lose the memory, or you get a slightly different memory, or something like that. If you could disrupt emotional memories like “speaking out makes you hated” while they’re still reconsolidating, maybe you could do something about this.

Anisomycin is pretty toxic, so that’s out. Other protein synthesis inhibitors are also toxic – it turns out proteins are kind of important for life – so they’re out too. Electroconvulsive therapy actually seems to work pretty well for this – the shock disrupts protein formation very effectively (and the more I think about this, the more implications it seems to have). But we can’t do ECT on everybody who wants to be able to speak up at work more, so that’s also out. And the simplest solution – activating a memory and then reminding the patient that they don’t rationally believe it’s true – doesn’t seem to help; the emotional brain doesn’t speak Rationalese.

The authors of *UtEB* claim to have found a therapy-based method that works, which goes like this:

First, they tease out the exact predictive model and emotional memory behind the symptom (in Richard’s case, the narrative where his father talked too much and ended up universally hated, and so if Richard talks at all, he too will be universally hated). Then they try to get this as far into conscious awareness as possible (or, if you prefer, have consciousness dig as deep into the emotional schema as possible). They call this “the pro-symptom position” – giving the symptom as much room as possible to state its case without rejecting it. So for example, Richard’s therapist tried to get Richard to explain his unconscious pro-symptom reasoning as convincingly as possible: “My father was really into talking, and everybody hated him. This proves that if I speak up at work, people will hate me too.” She even asked Richard to put this statement on an index card, review it every day, and bask in its compellingness. She asked Richard to imagine getting up to speak, and feeling exactly how anxious it made him, while reviewing to himself that the anxiety felt justified given what happened with his father. The goal was to establish a wide, well-trod road from consciousness to the emotional memory.

Next, they try to find *a lived and felt experience* that contradicts the model. Again, Rationalese doesn’t work; the emotional brain will just ignore it. But it will listen to experiences. For Richard, this was a time when he was at a meeting, had a great idea, but didn’t speak up. A coworker had the same idea, mentioned it, and everyone agreed it was great, and congratulated the other person for having such an amazing idea that would transform their business. Again, there’s this same process of trying to get as much in that moment as possible, bring the relevant feelings back again and again, create as wide and smooth a road from consciousness to the experience as possible.

Finally, the therapist activates the disruptive emotional schema, and before it can reconsolidate, smashes it into the new experience. So Richard’s therapist makes use of the big wide road Richard built that let him fully experience his fear of speaking up, and asks Richard to get into that frame of mind (activate the fear-of-speaking schema). Then she asks him, *while keeping the fear-of-speaking schema in mind*, to remember the contradictory experience (coworker speaks up and is praised). Then the therapist vividly describes the juxtaposition while Richard tries to hold both in his mind at once.

And then Richard was instantly cured, and never had any problems speaking up at work again. His coworkers all applauded, and became psychotherapists that very day. An eagle named "Psychodynamic Approach" flew into the clinic and perched atop the APA logo and shed a single tear. *Coherence Therapy: Practice Manual And Training Guide* was read several times, and God Himself showed up and enacted PsyD prescribing across the country. All the cognitive-behavioralists died of schizophrenia and were thrown in the lake of fire for all eternity.

This is, after all, a therapy book.

II.

I like *UtEB* because it reframes historical/purposeful accounts of symptoms as aspects of a predictive model. We already know the brain has an unconscious predictive model that it uses to figure out how to respond to various situations and which actions have which consequences. In retrospect, this framing perfectly fits the idea of traumatic experiences having outsized effects. Tack on a bit about how the model is more easily updated in childhood (because you've seen fewer other things, so your priors are weaker), and you've gone a lot of the way to traditional models of therapy.

But I also like it because it helps me think about the idea of separation/noncoherence in the brain. Richard had his schema about how speaking up makes people hate you. He also had lots of evidence that this wasn't true, both rationally (his understanding that his symptoms were counterproductive) and experientially (his story about a coworker proposing an idea and being accepted). But the evidence failed to naturally propagate; it didn't connect to the schema that it should have updated. Only after the therapist forced the connection did the information go through. Again, all of this should have been obvious – of course evidence doesn't propagate through the brain, I was writing posts ten years ago about how even a person who knows ghosts don't exist will be afraid to stay in an old supposedly-haunted mansion at night with the lights off. But UtEB's framework helps snap some of this into place.

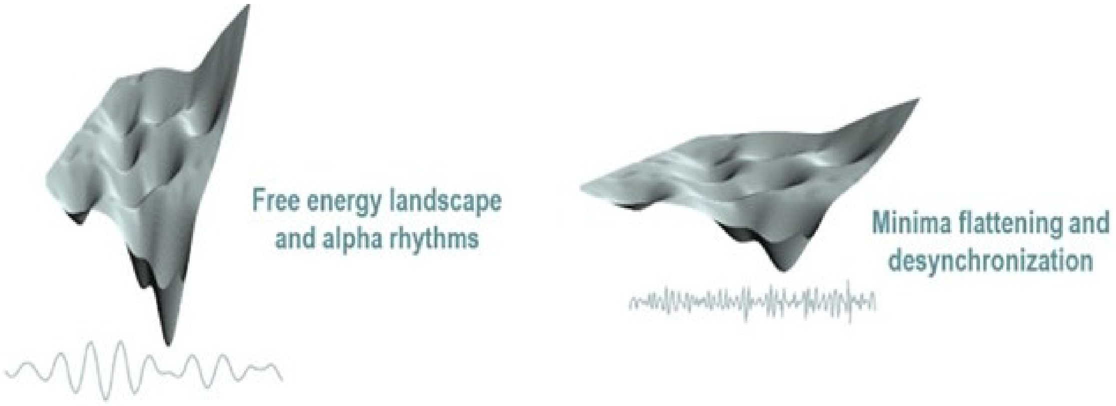
UtEB's brain is a mountainous landscape, with fertile valleys separated by towering peaks. Some memories (or pieces of your predictive model, or whatever) live in each valley. But they can't talk to each other. The passes are narrow and treacherous. They go on believing their own thing, unconstrained by conclusions reached elsewhere.

Consciousness is a capital city on a wide plain. When it needs the information stored in a particular valley, it sends messengers over the passes. These messengers are good enough, but they carry letters, not weighty tomes. Their bandwidth is atrocious; often they can only convey what the valley-dwellers think, and not why. And if a valley gets something wrong, lapses into heresy, as often as not the messengers can't bring the kind of information that might change their mind.

Links between the capital and the valleys may be tenuous, but valley-to-valley trade is almost non-existent. You can have two valleys full of people working on the same problem, for years, and they will basically never talk.

Sometimes, when it's very important, the king can order a road built. The passes get cleared out, high-bandwidth communication to a particular valley becomes possible. If he does this to two valleys at once, then they may even be able to share notes directly, each passing through the capital to get to each other. But it isn't the norm. You have to really be trying.

This ended out a little more flowery than I expected, but I didn’t start thinking this way because it was poetic. I started thinking this way because of this:



Frequent SSC readers will recognize this as from Figure 1 of Friston and Carhart-Harris’ REBUS And The Anarchic Brain: Toward A Unified Model Of The Brain Action Of Psychedelics, which I review here. The paper describes it as “the curvature of the free-energy landscape that contains neuronal dynamics. Effectively, this can be thought of as a flattening of local minima, enabling neuronal dynamics to escape their basins of attraction and—when in flat minima—express long-range correlations and desynchronized activity.”

Moving back a step: the paper is trying to explain what psychedelics do to the brain. It theorizes that they weaken high-level priors (in this case, you can think of these as the tendency to fit everything to an existing narrative), allowing things to be seen more as they are:

> A corollary of relaxing high-level priors or beliefs under psychedelics is that ascending prediction errors from lower levels of the system (that are ordinarily unable to update beliefs due to the top-down suppressive influence of heavily-weighted priors) can find freer register in conscious experience, by reaching and impressing on higher levels of the hierarchy. In this work, we propose that this straightforward model can account for the full

> breadth of subjective phenomena associated with the psychedelic experience.

These ascending prediction errors (ie noticing that you’re wrong about something) can then correct the high-level priors (ie change the narratives you tell about your life):

> The ideal result of the process of belief relaxation and revision is a recalibration of the relevant beliefs so that they may better align or harmonize with other levels of the system and with bottom-up information—whether originating from within (e.g., via lower-level intrinsic systems and related interoception) or, at lower doses, outside the individual (i.e., via sensory input or extroception). Such functional harmony or realignment may look like a system better able to guide thought and behavior in an open, unguarded way (Watts et al., 2017; Carhart-Harris et al., 2018b).

This makes psychedelics a potent tool for psychotherapy:

> Consistent with the model presented in this work, overweighted high-level priors can be all consuming, exerting excessive influence throughout the mind and brain’s (deep) hierarchy. The negative cognitive bias in depression is a good example of this (Beck, 1972), as are fixed delusions in psychosis (Sterzer et al., 2018).25 In this paper, we propose that psychedelics can be therapeutically effective, precisely because they target the high levels of the brain’s functional hierarchy, primarily affecting the precision weighting of high-level priors or beliefs. More specifically, we propose that psychedelics dose-dependently relax the precision weighting of high-level priors (instantiated by high-level cortex), and in so doing, open them up to an upsurge of previously suppressed bottom-up signaling (e.g., stemming from limbic circuitry). We further propose that

> this sensitization of high-level priors means that more information can impress on them, potentially inspiring shifts in perspective, felt as insight. One might ask whether relaxation followed by revision of high-level priors or beliefs via psychedelic therapy is easy to see with functional (and anatomic) brain imaging. We presume that it must be detectable, if the right questions are asked in the right way.

Am I imagining this, or are Friston + Carhart-Harris and *Unlocking The Emotional Brain* getting at the same thing?

Both start with a piece of a predictive model (= high-level prior) telling you something that doesn't fit the current situation. Both also assume you have enough evidence to convince a rational person that the high-level prior is wrong, or doesn't apply. But you don't automatically smash the prior and the evidence together and perform an update. In *UtEB*'s model, the update doesn't happen until you forge conscious links to both pieces of information and try to hold them in consciousness at the same time. In F+CH's model, the update doesn't happen until you take psychedelics which make the high-level prior lose some of its convincingness. *UtEB* is trying to laboriously build roads through mountains; F+CH are trying to cast a magic spell that makes the mountains temporarily vanish. Either way, you get communication between areas that couldn't communicate before.

III.

Why would mental mountains exist? If we keep trying to get rid of them, through therapy or psychedelics, or whatever, then why not just avoid them in the first place?

Maybe generalization is just hard (thanks to MC for this idea). Suppose Goofus is mean to you. You learn Goofus is mean; if this is your first social experience, maybe you also learn that the world is mean and people have it out for you. Then one day you meet Gallant, who is nice to you. Hopefully the system generalizes to “Gallant is nice, Goofus is still mean, people in general can go either way”.

But suppose one time Gallant is just having a terrible day, and curses at you, and that time he happens to be wearing a red shirt. You don’t want to overfit and conclude “Gallant wearing a red shirt is mean, Gallant wearing a blue shirt is nice”. You want to conclude “Gallant is generally nice, but sometimes slips and is mean.”

But any algorithm that gets too good at resisting the temptation to separate out red-shirt-Gallant and blue-shirt-Gallant risks falling into the opposite failure mode where it doesn’t separate out Gallant and Goofus. It would just average them out, and conclude that people (including both Goofus and Gallant) are medium-niceness.

And suppose Gallant has brown eyes, and Goofus green eyes. You don’t want your algorithm to overgeneralize to “all brown-eyed people are nice, and all green-eyed people are mean”. But suppose the Huns attack you. You *do* want to generalize to “All Huns are dangerous, even though I can keep treating non-Huns as generally safe”. And you want to do this as quickly as possible, definitely before you meet any more Huns. And the quicker you are to generalize about Huns, the more likely you are to attribute false significance to Gallant’s eye color.

The end result is a predictive model which is a giant mess, made up of constant “This space here generalizes from this example, except this subregion, which generalizes from this other example, except over here, where it doesn’t, and definitely don’t *ever* try to apply any of those examples over here.” Somehow this all works shockingly well. For example, I spent a few years in Japan, and developed a good model for how to behave in Japanese culture. When I came back to the United States, I effortlessly dropped all of that and went back to having America-appropriate predictions and reflexive actions (except for an embarrassing habit of bowing whenever someone hands me an object, which I still haven't totally eradicated).

In this model, mental mountains are just the context-dependence that tells me not to use my Japanese predictive model in America, and which prevents evidence that makes me update my Japanese model (like "I notice subways are always on time") from contaminating my American model as well. Or which prevent things I learn about Gallant (like "always trust him") from also contaminating my model of Goofus.

There's actually a real-world equivalent of the "red-shirt-Gallant is bad, blue-shirt-Gallant is good" failure mode. It's called "splitting", and you can find it in any psychology textbook. Wikipedia defines it as "the failure in a person's thinking to bring together the dichotomy of both positive and negative qualities of the self and others into a cohesive, realistic whole."

In the classic example, a patient is in a mental hospital. He likes his doctor. He praises the doctor to all the other patients, says he's going to nominate her for an award when he gets out.

Then the doctor offends the patient in some way – maybe refuses one of his requests. All of a sudden, the doctor is abusive, worse than Hitler, worse than Mengele. When he gets out he will report her to the authorities and sue her for everything she owns.

Then the doctor does something right, and it’s back to praise and love again.

The patient has failed to integrate his judgments about the doctor into a coherent whole, “doctor who sometimes does good things but other times does bad things”. It’s as if there’s two predictive models, one of Good Doctor and one of Bad Doctor, and even though both of them refer to the same real-world person, the patient can only use one at a time.

Splitting is most common in borderline personality disorder. The DSM criteria for borderline includes splitting (there defined as “a pattern of unstable and intense interpersonal relationships characterized by alternating between extremes of idealization and devaluation”). They also include things like “markedly and persistently unstable self-image or sense of self”, and “affective instability due to a marked reactivity of mood”, which seem relevant here too.

Some therapists view borderline as a disorder of integration. Nobody is great at having all their different schemas talk to each other, but borderlines are atrocious at it. Their mountains are so high that even different thoughts about the same doctor can’t necessarily talk to each other and coordinate on a coherent position. The capital only has enough messengers to talk to one valley at a time. If tribesmen from the Anger Valley are advising the capital today, the patient becomes truly angry, a kind of anger that utterly refuses to listen to any counterevidence, an anger pure beyond your imagination. If they are happy, they are *purely* happy, and so on.

About 70% of people diagnosed with dissociative identity disorder (previously known as multiple personality disorder) have borderline personality disorder. The numbers are so high that some researchers are not even convinced that these are two different conditions; maybe DID is just one manifestation of borderline, or especially severe borderline. Considering borderline as a failure of integration, this makes sense; DID is total failure of integration. People in the furthest mountain valleys, frustrated by inability to communicate meaningfully with the capital, secede and set up their own alternative provincial government, pulling nearby valleys into their new coalition. I don’t want to overemphasize this; most popular perceptions of DID are overblown, and at least some cases seem to be at least partly iatrogenic. But if you are bad enough at integrating yourself, it seems to be the sort of thing that can happen.

In his review, Kaj relates this to Internal Family Systems, a weird form of therapy where you imagine your feelings as people/entities and have discussions with them. I’ve always been skeptical of this, because feelings are not, in fact, people/entities, and it’s unclear why you should expect them to answer you when you ask them questions. And in my attempts to self-test the therapy, indeed nobody responded to my questions and I was left feeling kind of silly. But Kaj says:

> As many readers know, I have been writing a sequence of posts on multi-agent models of mind. In Building up to an Internal Family Systems model, I suggested that the human mind might contain something like subagents which try to ensure that past catastrophes do not repeat. In subagents, coherence, and akrasia in humans, I suggested that behaviors such as procrastination, indecision, and seemingly inconsistent behavior result from different subagents having disagreements over what to do.
>
> As I already mentioned, my post on integrating disagreeing subagents took the model in the direction of

> interpreting disagreeing subagents as conflicting beliefs or models within a person’s brain. Subagents, trauma and rationality further suggested that the appearance of drastically different personalities within a single person might result from unintegrated memory networks, which resist integration due to various traumatic experiences.
>
> This post has discussed UtEB’s model of conflicting emotional schemas in a way which further equates “subagents” with beliefs – in this case, the various schemas seem closely related to what e.g. Internal Family Systems calls “parts”. In many situations, it is probably fair to say that this is what subagents are.

This is a model I can get behind. My guess is that in different people, the degree to which mental mountains form a barrier will cause the disconnectedness of valleys to manifest as anything from “multiple personalities”, to IFS-findable “subagents”, to *UtEB*-style psychiatric symptoms, to “ordinary” beliefs that don’t cause overt problems but might not be very consistent with each other.

IV.

This last category forms the crucial problem of rationality.

One can imagine an alien species whose ability to find truth was a simple function of their education and IQ. Everyone who knows the right facts about the economy and is smart enough to put them together will agree on economic policy.

But we don’t work that way. Smart, well-educated people believe all kinds of things, even when they should know better. We call these people biased, a catch-all term meaning something that prevents them from having true beliefs they ought to be able to figure out. I believe most people who don’t believe in anthropogenic climate change are probably biased. Many of them are very smart. Many of them have read a lot on the subject (empirically, reading more about climate change will usually just make everyone more convinced of their current position, whatever it is). Many of them have enough evidence that they should know better. But they don't.

(again, this is my opinion, sorry to those of you I'm offending. I'm sure you think the same of me. Please bear with me for the space of this example.)

Compare this to Richard, the example patient mentioned above. Richard had enough evidence to realize that companies don't hate everyone who speaks up at meetings. But he still felt, on a deep level, like speaking up at meetings would get him in trouble. The evidence failed to connect to the emotional schema, the part of him that made the real decisions. Is this the same problem as the global warming case? Where there's evidence, but it doesn't connect to people's real feelings?

(maybe not: Richard might be able to say "I know people won't hate me for speaking, but for some reason I can't make myself speak", whereas I've never heard someone say "I know climate change is real, but for some reason I can't make myself vote to prevent it." I'm not sure how seriously to take this discrepancy.)

In Crisis of Faith, Eliezer Yudkowsky writes:

> Many in this world retain beliefs whose flaws a ten-year-old could point out, if that ten-year-old were hearing the beliefs for the first time. These are not subtle errors we're talking about. They would be child's play for an unattached mind to relinquish, if the skepticism of a ten-year-old were applied without evasion…we change our minds less often than we think.

> This should scare you down to the marrow of your bones. It means you can be a world-class scientist and conversant with Bayesian mathematics and still fail to reject a belief whose absurdity a fresh-eyed ten-year-old could see. It shows the invincible defensive position which a belief can create for itself, if it has long festered in your mind.
>
> What does it take to defeat an error that has built itself a fortress?

He goes on to describe how hard this is, to discuss the “convulsive, wrenching effort to be rational” that he thinks this requires, the “all-out [war] against yourself”. Some of the techniques he mentions explicitly come from psychotherapy, others seem to share a convergent evolution with it.

The authors of *UtEB* stress that all forms of therapy involve their process of reconsolidating emotional memories one way or another, whether they know it or not. Eliezer’s work on crisis of faith feels like an *ad hoc* form of epistemic therapy, one with a similar goal.

Here, too, there is a suggestive psychedelic connection. I can’t count how many stories I’ve heard along the lines of “I was in a bad relationship, I kept telling myself that it was okay and making excuses, and then I took LSD and realized that it obviously wasn’t, and got out.” Certainly many people change religions and politics after a psychedelic experience, though it’s hard to tell exactly what part of the psychedelic experience does this, and enough people end up believing various forms of woo that I hesitate to say it’s all about getting more rational beliefs. But just going off anecdote, this sometimes works.

Rationalists wasted years worrying about various named biases, like the conjunction fallacy or the planning fallacy. But most of the problems we really care about aren’t any of those. They’re more like whatever makes the global warming skeptic fail to connect with all the evidence for global warming.

If the model in *Unlocking The Emotional Brain* is accurate, it offers a starting point for understanding this kind of bias, and maybe for figuring out ways to counteract it.

SSC Meetups Everywhere Retrospective

Slate Star Codex has regular weekly-to-monthly meetups in a bunch of cities around the world. Earlier this autumn, we held a Meetups Everywhere event, hoping to promote and expand these groups. We collected information on existing meetups, got volunteers to create new meetups in cities that didn't have them already, and posted times and dates prominently on the blog.

During late September and early October, I traveled around the US to attend as many meetups as I could. I hoped my presence would draw more people; I also wanted to learn more about meetups and the community and how best to guide them. Buck Shlegeris and a few other Bay Area effective altruists came along to meet people, talk to them about effective altruism, and potentially nudge them into the recruiting pipeline for EA organizations.

Lots of people asked me how my trip was. In a word: exhausting. I got to meet a lot of people for about three minutes each. There were a lot of really fascinating people with knowledge of a bewildering variety of subjects, but I didn't get to pick their minds anywhere as thoroughly as I would have liked. I'm sorry if I talked to you for three minutes, you told me about some amazing project you were working on to clone neuroscientists or eradicate bees or convert atmospheric CO2 into vegan meat substitutes, and I mumbled something and walked away. You are all great and I wish I could have spent more time with you.

I finally got to put faces to many of the names I've interacted with through the years. For example, Bryan Caplan is exactly how you would expect, in every way. Also, in front of his office, he has a unique painting, which he apparently got by asking a Mexican street artist to paint an homage to *Lord of the Rings*. The artist had never heard of it before, but Bryan described it to him very enthusiastically, and the completely bonkers result is hanging in front of his office. This is probably a metaphor for something.

Philadelphia hosted their meetup in a beautiful room that looked like a Roman temple, and had miniature cheesesteaks for everybody. Chicago held theirs in a gym; appropriate, given this blog's focus on BRUTE STRENGTH. Berkeley's was in a group house with posters representing the Twelve Virtues Of Rationality hanging along the staircase. In Fairbanks, a person who had never read the blog showed up to get a story and an autograph for his brother who did. In New York, someone brought the best bread I have ever had, maybe the best bread *anyone* has ever had, I am so serious about this. In Boston, the organizers set up a prediction market to determine how many attendees they needed to plan for; they still ended up being off by a factor of two. This is also probably a metaphor for something. If only they had used more BRUTE STRENGTH!

Along the way, I got to see America. Most of it I saw from an airplane window, but I still saw it. In Portland, I ate from a makeshift food court formed by a bunch of really good food trucks congregating in the same empty lot; one of them just sold like a dozen different kinds of french fries. In Texas, I rode with an Uber driver whose day job is driving mechanical bulls to parties that need mechanical bulls, and who Ubers people around while he waits for the party to finish. In Washington DC, I tried to see the White House, only to be thwarted by the construction of a new security fence; they say that before you change the world you must change your own home, and it seems like our Wall-Builder-

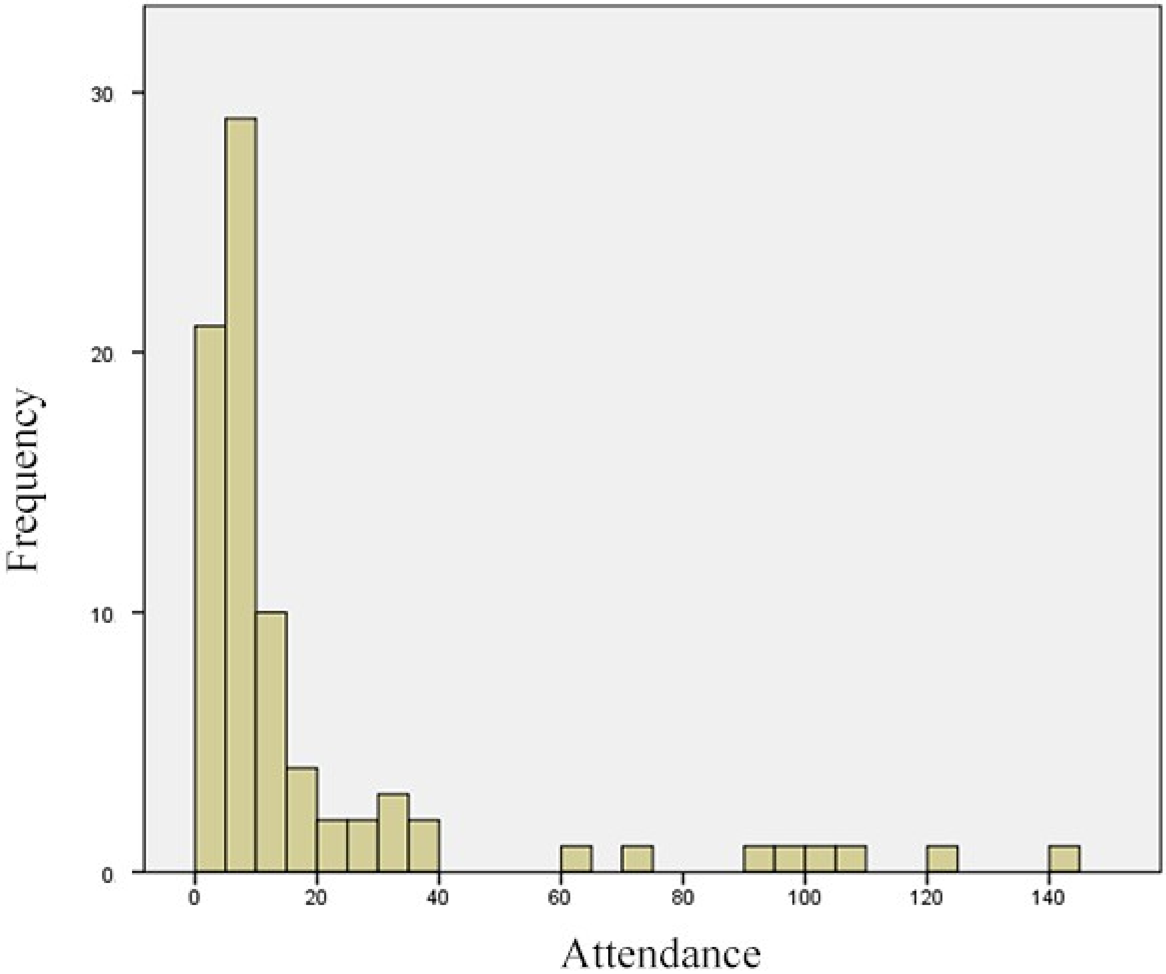
In-Chief takes this seriously. In Delaware, I stood on the spot where the Swedes first landed in America and declared it to be the colony of New Sweden; probably there are alternate timelines out there who could appreciate this more than I did. In New Jersey, I confirmed that the Pine Barrens are, in fact, really creepy.

People gave me things. You are all so nice, but you also seem to think I am about ten times more classy and fashionable than I really am. One person gave me a beautiful record of their audiobook – a real, honest-to-goodness vinyl record – as if I had any idea what to do with it. A reader in Philadelphia gave me a beautiful glossy magazine about Philadelphia culture, which I stared at intently for twenty minutes. Many people gave me beautifully-bound copies of my own work, which was so incredibly thoughtful that I feel bad that I will have to hide them in a closet so nobody sees them and thinks I am the kind of narcissist who makes beautifully-bound copies of my own work. The Charter Cities Institute people gave me a very nice Charter Cities Institute bag (although I assume that if I ever take it outside in Berkeley, someone will punch me and it will start a National Conversation). I am still really grateful to all of you.

But you already know how great you are. Let’s get to the statistics.

Mingyuan, the Official SSC Meetup Coordinator, sent out a survey to get information on the meetups we weren’t able to visit, and determined that we had somewhere between 81 and 111 meetups around the world. I’m sorry I can’t be more precise. 111 meetups were supposed to happen, 81 organizers reported back to Mingyuan that their meetups happened, and I’m not sure what happened to the other 30. Although most activity was concentrated in the

Anglosphere, there were meetups as far away as Bangalore (9 people), Tel Aviv (25 people), Oslo (9 people), and Seoul (4 people). Medellin, Colombia reports a one person meetup; I am sorry it sounds like you did not have a good time. Montreal, Canada, reports a zero person meetup, which sounds very computer-sciency, kind of like a heap of zero grains of sand.



Here’s the histogram of attendance, binned by fives. About twenty meetups had 0-5 people, thirty had 5-10, and the remaining thirty had more than 10. The best-attended meetups were Boston (140), NYC (120), and Berkeley (105).

Total meetup attendance around the world was almost 1500 people!

Did the event fulfill its goal of bringing more people to meetups? Many organizers had only a vague idea how many people usually attended their meetups, and many said their city didn’t have a usual meetup group at all. But as best I can tell, about 2.3x as many people attended the Meetups Everywhere meetup in a city compared to the average previous meetup. Breaking it down by tour status, meetups on my tour had much higher attendance (6.1x usual), but even meetups off my tour had somewhat higher attendance (1.6x usual).

Did the event succeed in bringing some people into meetup groups who might stay around later? I suggested meetup organizers bring a signup sheet that people could sign to get on a mailing list for future meetups. My data on this is sparse, because people took the survey question overly literally and wrote things like “I didn’t have a signup sheet, I just asked people for their emails” and then didn’t tell me how many people gave them. But for the 40 meetups where I have data, people on average got a population of new signers equal to 77% of their previous regular attendance; that is, a meetup group that usually had 100 people had 77 extra new people sign up for their mailing last. Breaking it down by tour status, meetups on my tour gained 170%, other meetups gained 58%.

This seems implausibly large; did one event nearly double the attendance of SSC meetup groups around the world? I don’t know how many people who signed up for the mailing list will really start attending regularly. But I will probably survey the organizers again next year, and they might be able to help me figure out how many people stayed around.

In total, 1,476 people attended SSC meetups, and 339 people added their name to mailing lists (the ratio here doesn’t match the previous numbers because most organizers didn’t have a mailing list or didn’t report mailing list data, and the ratios above only counted those who did).

So much for the numbers. What did I learn?

I don’t want to generalize too much – I deliberately went to the biggest meetups, and things that work for a group of 100 people might not apply to a group of 2 people. So take all of this with a grain of salt, but:

1. Tables and chairs kill big meetups. Some people tried to hold meetups at a restaurant or a park with picnic tables or something. Everyone would sit down at the table, talk to the 3-4 people in their immediate neighborhood, and that would be that. Eventually I figured out that I need to force everyone out of the picnic tables and into the rest of the park. This caused a phase shift from solid to gas, with people milling about, talking to everyone, finding the conversations that most interested them.

2. The welcomeness sentence is really important. In the meetup descriptions on the blog, I included a sentence like “Please feel free to come even if you feel awkward about it, even if you’re not ‘the typical SSC reader’, even if you’re worried people won’t like you, etc.” It sounds silly, but I had *so many* people come up to me saying the only reason they came was because of that sentence. It happened *again and again and again*. Anybody planning any kind of meetup about anything should strongly consider including a sentence like that (as long as it’s true). Maybe there are other simple hacks like this waiting to be discovered.

3. Group houses are important community nuclei. Obvious in retrospect, but it was pretty stark seeing the level of community in cities that did have rationalist group houses vs. the ones that didn’t, even if there were good meetup groups in both. This also came out in listening to some people mourn the loss of the main group house in their city and talk about all the great things they were no longer able to do.

I was thinking of this last one because a lot of the meetups felt kind of superficial. Everyone shows up, talks about their favorite SSC post or what their job is or what kind of interesting thing they read recently, and then they go home. Lots of people seemed to enjoy that, *I* enjoyed it, but seeing the kind of really great rationalist communities in the Bay Area or Seattle gave me a sense that more is possible. I don’t know, maybe it’s not possible in cities with only 10 or 20 interested people; maybe only places like the Bay Area and Seattle have enough people, and everywhere it’s possible it’s already happening. But group houses seem to be a big part of it.

I was also struck by the number of female meetup organizers; the female:male ratio on the meetup organizer survey is almost twice that on the SSC survey in general. When there were cities that didn’t have regular meetup groups, and I asked for a volunteer to set one up, it was usually a woman who volunteered.

This suggests to me that we’re not just performing at some kind of theoretical maximum for the amount of people and interest in a given community; there’s a shortage of something (speculatively, social initiative) that (in this community) women are better than men at. I don’t know how to solve this (though integrating more with the EA community, which has more women, might help), but I think it’s an interesting problem.

And Buck has written his own retrospective of his EA work at the meetups here.

Symptom, Condition, Cause

On my recent post on autism, several people chimed in to say that “autism” wasn’t a unitary/homogenous category. It probably lumps together many different conditions with many different causes. It’s useless to speculate on the characteristics of “autism” until it can be separated out further.

I get this every time I talk about a psychiatric condition. The proponents of this view seem to think they’re speaking a shocking heresy that overturns the psychiatric establishment. But guys, we know this kind of stuff. Psychiatric diagnoses don’t have to perfectly match underlying root causes to be useful.

Suppose a patient comes to you with difficulty breathing, excessive sweating, anxiety, and extreme discomfort when lying down flat. You recognize these as potential signs of pulmonary edema, ie fluid in the lungs. You do an x-ray, confirm the diagnosis, and prescribe symptomatic treatment – in this case, supplemental oxygen. All of this is good work.

But you can have fluid in your lungs for lots of different reasons. Most of the time it’s heart failure, but sometimes it’s kidney failure, pneumonia, drug overdose, smoke inhalation, or altitude sickness. Some of these causes will have slightly different symptoms, which an alert doctor can notice.

Suppose the real cause of your pulmonary edema is heroin overdose. In that case, it wouldn’t be fair to call pulmonary edema a “root cause”. The root cause of your problem is the heroin. But you also can’t call pulmonary edema merely a “symptom”. No patient comes in saying “Doc, I’m feeling a bit pulmonary edemic today”. The symptoms of pulmonary edema are difficulty breathing, excessive sweating, anxiety, etc. So what is pulmonary edema?

I don’t know the technical philosophy-of-medicine term for this, but let’s call it a “condition”. A condition which nobody has yet matched with a biological process gets dubbed a syndrome – a set of symptoms that go together even if we remain agnostic about why. A condition which *has* been matched a biological process ends up like pulmonary edema – such a well-known part of the medical canon that nobody feels a need to do philosophy around it.

Lots of things are conditions like this. Even some universally-known diseases like stroke are better thought of as conditions than root causes. Strokes can be caused either by ischaemia (usually a blood vessel blocked by a clot) or haemorrhage (a blood vessel bursting and bleeding out). These two causes have differing risk factors (anticoagulants cause haemorrhagic stroke but protect against ischaemic) and differing treatments (tPA relieves ischaemic stroke but catastrophically worsens haemorrhagic).

But nobody ever bursts into neurology conferences shouting “STROKE ISN’T A REAL DISEASE, IT’S A COBBLED-TOGETHER BASKET OF MULTIPLE DIFFERENT ROOT CAUSES!” Everyone realizes that conditions are a useful intermediate level to work at.

This is how I feel about things like depression too. No psychiatrist would be even a tiny bit surprised to hear that depression is many different conditions with many different causes. For example, everyone knows some depressions are caused by hypothyroidism, and others aren’t.

The biggest difference between the philosophical status of depression vs. stroke is that we know what biological process stroke corresponds to. Stroke is brain cells dying from lack of oxygen. It can be caused by arterial blockage or by bleeding, sometimes it can even have more distal causes like cocaine use or Moyamoya disease, but it all ends with brain cells dying from lack of oxygen. That in turn produces classic symptoms like sudden-onset slurred speech, hemiparalysis, and facial asymmetry.

We don’t have as good an idea what biological process depression corresponds to. There are some theories – maybe a failure of synaptogenesis – but they’re all pretty speculative right now. Still, I think it’s reasonable to propose that they correspond to some process.

First, because depression includes a lot of surprising symptoms mysteriously clustered together. Just as without the concept of “stroke” you can’t explain why slurred speech and hemiparalysis happen together so often, so without the concept of “depression” it’s hard to explain why SIGECAPS tend to go together. The only good alternative I’ve heard here is the idea of symptom networks. But I no longer find this very convincing, and it never seems to be what the people talking about how “depression isn’t a single disorder” mean.

Second, because at this point we don’t even know what biological process normal low mood corresponds to, but it seems like it has to be something, and it would be strange for a single biological process to cause low mood and not be related to depression.

My (very wild) guess is that in the end psychiatric disorders will mostly turn out to be computational conditions. That is, something like “the learning rate of this system is set too high” or “the threshold for errors in this error-detector is too low”. There will be lots of different things that will cause that, from biological (because these computations are implemented on biological systems including the usual range of things like serotonin and dopamine and synapses) to psychological (because the brain is plastic enough that its computational parameters can change with experience) to environmental (because if you pour a bucket of battery acid onto a computer, probably its computational parameters will change in some way). This is just my personal bias towards computational explanations speaking, and it could be that these disorders will be better explained by regional stories (ie “the amygdala is broken” or “the hippocampus is broken”), by biochemical stories (“there’s too much serotonin”), by structural stories (“there are too few synapses”), by some combination of these, by something totally different, or by something that’s on a totally different level than any of this.

If something like this story is true, it means that research that treats depression as a single condition might or might not work. Returning to the analogy of stroke, I think (though I’m not an expert) that the prognoses for ischaemic and haemorrhagic stroke are mostly similar, since both depend on how long it takes the brain to adapt after some cells have died. But the risk factors for these two kinds of stroke are different (again, anticoagulants protect against one and cause the other). Scientists who were researching “stroke”, without understanding the different causes, would get some things right and end up confused about others.

Some people, upon hearing this, say that we should be trying to figure out the different kinds of depression so we can do real research on those. People have been trying this for a century, and every one of their leads have been false. Traditional psychiatry flirts with admitting two subtypes of depression, but you can also find papers claiming to have found three subtypes, four subtypes, five subtypes, etc. Even papers that agree on how many subtypes there are often *identify* the subtypes totally differently. This has not been a productive research program and I think better understanding of what depression is will be more valuable than bashing our heads against the subtype identification problem further. At least this is how it has always worked in regular medicine, where once we realized what eg pulmonary edema was, everything fell into place (including potential root causes) and nobody felt like figuring out exactly how many subtypes there were was a very interesting problem.

The saying goes: all models are wrong, some models are useful. I don’t think existing psychiatric diagnosis is particularly accurate, but I think it’s the most useful thing we have right now. And I don’t think talking about how each condition is probably made up of many root causes is a particularly damaging objection to it. We should keep the likely heterogeneity in mind and pull it out when we need it, but we shouldn’t use that as an excuse to abandon the whole nosology.

2019 Adversarial Collaboration Entries

Thanks to everyone who sent in entries for the 2019 adversarial collaboration contest.

Remember, an adversarial collaboration is where two people with opposite views on a controversial issue work together to present a unified summary of the evidence and its implications. In theory it's a good way to make sure you hear the strongest arguments and counterarguments for both sides – like hearing a debate between experts, except all the debate and rhetoric and disagreement have already been done by the time you start reading, so you're just left with the end result. See the 2018 entries for examples.

Eight teams submitted collaborations for this year's contest:

1. "What are the benefits, harms, and ethics of infant circumcision?" by Joel P and Missingno

2. "Is eating meat a net harm?" by David G and Froolow

3. "Does calorie restriction slow aging?" by Adrian L and Calvin R

4. "Should we colonize space to mitigate x-risk?" by Nick D and Rob S

5. "Should gene editing technologies be used in humans" by Nita J and Patrick N

6. "When during fetal development does abortion become morally wrong?" by BlockOfNihilism and Icerun

7. “Will automation lead to economic crisis?” by Doug S and Erusian

8. “How much significance should we ascribe to spiritual experiences?” by Seth S and Jeremiah G

(if any of you are unhappy with how I named you or titled your piece, let me know)

At the end of the two weeks, I’ll ask readers to vote for their favorite collaboration, so try to remember which ones impress you. I think we’re all winners by getting to read these – but the actual winners get that plus $2500 in prize money. Thanks again to everyone who donates to the Patreon for making that possible.

Please put any comments about the contest itself here, not on the individual entries.

[ACC] What Are The Benefits, Harms, And Ethics Of Infant Circumcision?

[This is an entry to the 2019 Adversarial Collaboration Contest by Joel P and Missingno]

"They practise circumcision for cleanliness' sake; for they would rather be clean than more becoming." – Herodotus, The Histories – 2.37

The debate over circumcision in the Western world today is surprisingly similar to the conflict that Greeks and Egyptians faced 2500 years ago. Supporters tend to emphasize its hygiene and health benefits; opponents tend to call it cruel or to emphasize its deviation from the natural human form. In this adversarial collaboration we address medical aspects, sensitivity and pleasure, and ethical aspects of infant circumcision.

Effect on penile cancer

Circumcision greatly reduces the relative rate of penile cancer, a relatively uncommon malignancy in developed nations which kills a little over 400 American men each year. Denmark, while it has one of the lowest rates of penile cancer for a non-circumcising country, nevertheless has 10x the rate of penile cancer as Israel – where almost all men are circumcised. Likewise, a Kaiser Permanente study of patients with penile cancer found that 16% of patients with carcinoma in situ had been circumcised; only 2% of patients with invasive penile cancer had been circumcised. Since the circumcision rate of Kaiser patients of the appropriate age was ~50%, this is in line with the 90% reduction.

While these are observational rather than prospective trials, the magnitude of the reduction is quite high. It is unlikely to be simply due to class or race given that it exists when comparing countries and when comparing individuals within the same health care system. Additionally, there is some association of penile cancer with HPV and a very strong association with phimosis, and circumcision reduces the rate of both of these. This provides a highly plausible theoretical explanation of how circumcision might lead to this risk reduction in penile cancer. However, this does raise the question of whether more aggressive future treatment of phimosis combined with HPV vaccination might reduce the rate of penile cancer in uncircumcised men in the future somewhat. Of course, more aggressive treatment of phimosis would require more childhood circumcisions, which carry higher risk than infant circumcision.

Effect on transmission of HIV and STDs

HIV: Three large randomized control trials have been performed in South Africa, Uganda, and Kenya, together comprising over 11,000 men. These men were randomized to be circumcised or not at the start of the studies for primary HIV prevention. The reduction in female to male HIV transmission seen in these studies is about 50%. This is consistent with observational studies and is the highest quality evidence: three independent, large-scale randomized control trials with similar results scrutinized by the Cochrane Collaboration. The studies were terminated early due to positive results, which is appropriate ethical practice, but which can tend to overestimate positive effects. However, the data is consistent with observational data so this is less likely a concern. Some have expressed the concern that the two groups did not receive identical HIV counseling.

It is true that the circumcision group felt much more comfortable having sex without condoms, and additional counseling was given to the circumcision group to tell them this was not adequate protection. Condom use was, despite the counseling, lower in the circumcision group than in the control. In one sense this means that the protective benefits of circumcision vs HIV may be understated. In another sense, this creates a large concern with advertising circumcision for the stated purpose of HIV prevention. Any such efforts must be careful not to oversell the benefits and thereby reduce condom usage. Additionally, the results are only applicable to heterosexual HIV transmission. Homosexual transmission has not been shown to be decreased by circumcision, presumably because of the extremely high risk of receptive anal sex. IV drug related transmission is almost certainly unaffected except via “herd immunity”.

The data for other STIs is far less compelling than for HIV. Secondary endpoints of the African HIV studies were other STIs, and rates of HPV and HSV were reduced by circumcision. This was only a secondary outcome, however, and other studies have had mixed results. The data for lower rates of bacterial vaginosis and trichomonas in female partners of circumcised men is somewhat stronger. However, none of these benefits are nearly as strongly supported or as high impact as the HIV reduction. Additionally, when considering the benefits and harms of an intervention such as circumcision, there are strong reasons not to consider the benefits that accrue to the patient’s future partners, but instead to focus only on the individual in question.

It is ironic that the evidence for reduction in other STIs is fairly weak, because as historian David Gollaher shows in Circumcision: A History of the World’s Most Controversial

Surgery, this is the primary reason the US adopted widespread circumcision in the early 20th century. There had been very small-scale interest in circumcision due to religious ideas about masturbation and ideas about balanitis and phimosis causing systemic illness, but these ideas do not appear to have motivated a large number of circumcisions. Mainstream circumcision of healthy males caught on as a way to reduce STI rates – particularly syphilis. Physicians both in the US and UK saw the far lower rates of STIs Jews experienced than gentiles and attributed these primarily to circumcision. In the US, the time was just right for such STI reduction efforts – worries about infection were widespread and an increasing number of people were adopting hospital births where there was ready access to a physician able to perform a circumcision.

Meanwhile, during WWI and WWII the military offered circumcision to many conscripts to protect vs STIs (the wealthier officer class already having a much higher circumcision rate than the enlisted men as more of their parents could afford hospital births). The UK's experience of WWI and WWII was quite different from the US's. For one thing, STIs ranked far lower on the set of risks to soldiers. And rather than seeing a boom in hospital births, the UK's medical resources were strained during WWI and WWII. Circumcision was seen as something of a waste compared to the UK's more pressing needs. Presumably, arguments that positively presented Jews as having low STI rates did not catch on in early 20th century mainland Europe to nearly the extent that was seen in the US and UK.

Effect on UTIs

In the first year of life, the rate of UTIs is approximately 1% per year among uncircumcised boys and 0.1%-0.2% among circumcised boys. Particularly in the first year of life, UTIs can be severe, causing fever and hospitalization, as well as permanent kidney damage. Circumcision is presumably protective against UTI primarily by reducing the bacterial load around the urethra. Some sources have suggested that the difference is primarily one of contamination during sampling. However, studies looking only at clean catch urine samples or suprapubic tap samples give similar reductions (90%). Unlike many of the other benefits listed above, UTI avoidance is specifically a benefit of infant circumcision.

Effect on Penile Problems

Many penile problems such as balanitis (inflammation of the glans), pathologic phimosis (inability to retract the foreskin), and paraphimosis (foreskin entrapment, which requires emergency treatment to preserve the penis) are prevented by circumcision. Others, including meatal stenosis, scarring, bleeding are caused by circumcision. A New Zealand cohort study directly comparing the incidence of penile problems requiring intervention found a rate of 1.1% in circumcised children and 1.8% in uncircumcised children when followed to age 8.

Risks of surgery

The risks of surgery include pain, bleeding, bruising, inadequate foreskin removal, excess skin removal, swelling, meatal stenosis, scarring, infection, and anesthetic complications. These are different based on age group; neonatal circumcision is associated with a much lower risk of complications than other age groups. However, studies show a wide range of rates of complications dependent on practitioner training level. Overall, the rate of minor complications (bleeding, bruising) is ~1.5% worldwide and the rate of major complications (scar, severe infection, meatal stenosis, or need for additional surgery) is <0.2%. In comparison, the risk of complications in children past infancy and adults is approximately 6% with trained practitioners – significantly higher than for infant circumcision. Indeed, the majority of cases of the most severe complication (penectomy) related to circumcision appear to occur in people who were not circumcised as infants. This would include both adults with penile cancer as well as children undergoing phimosis surgery (as in the infamous case of David Reimer).

Sensitivity and Sexual satisfaction

There is a highly plausible mechanism by which circumcision could reduce sexual sensitivity: the foreskin is highly innervated (20,000 nerve endings is often repeated, but this appears to be a case of citogenesis and is likely far too high), produces lubrication for the penis, and is sensitive to light touch. Several studies demonstrate that the foreskin is more sensitive to certain forms of nonsexual stimulation than other parts of the penis. The glans itself does not change in sensitivity from circumcision.

Sexual satisfaction, particularly in sexually active heterosexual men, seems to be unchanged with adult circumcision. During studies of adult circumcision for HIV prevention, in which large numbers of men were randomized to receive circumcision at the time of the study or after, sexual satisfaction of did not significantly differ between the two groups. On the other hand, a South Korean study of men circumcised as adults (as has become traditional there) found decreased pleasure from masturbation after circumcision. It is certainly possible that both these things are true – that masturbation is impaired by adult circumcision while intercourse is not. It is also possible that the Korean study (retrospective, smaller than the African studies, and with much higher rates of scarring than are observed in the US) was unrepresentative. There are two European studies which are frequently cited: cohort studies look at circumcised and uncircumcised men in Denmark and Belgium. However, circumcision is quite rare in these countries, and the majority of the circumcisions in the study groups were performed to correct problems such as phimosis. They are thus comparing men who had penile problems requiring surgical correction to men who did not; it is therefore unclear why they are frequently cited in discussions of elective circumcision.

No available studies actually measure sensitivity to sexual stimulation, which is of course an important topic – but one requiring consummate professionalism on the part of the researcher. We are left waiting for such a study, but in the meantime may reasonably fear that there is some decrease in at least masturbatory pleasure due to circumcision even though the evidence for this is weak. The evidence does not support any change in sexual pleasure otherwise.

Infant circumcision may be different than adult circumcision, in addition. If circumcision eliminates important nerves, due to brain plasticity infants are likely better able than adults to reassign the portions of the brain processing the foreskin to other areas of the penis. A large survey of circumcised and uncircumcised men in the US (where infant circumcision is the most common) found similar sensation in circumcised and uncircumcised men. The uncircumcised men appear to have had slightly higher incidences of sexual dysfunction. Also of interest, circumcised men appear to have an easier time obtaining oral sex, which may relate to subtle aspects of class or may have to do with the perceived cleanliness of circumcised penis.

Ethics

The ethics of infant circumcision is a complex topic, and the answers likely depend on one’s ethical system. The benefits of infant circumcision appear to outweigh the risks and harms. Additionally, it is safer to be circumcised as an infant than as an adult, and a significant portion of the benefits of circumcision accrue to infants and children. From a strictly utilitarian perspective, infant circumcision should therefore be encouraged – whether we consider society as a whole or only the boy in question. However, autonomy is an important value, and while a man can become circumcised (missing only some of the benefits of having been circumcised as an infant), it is impossible to effectively restore the foreskin and become “de-circumcised”. An ethical system that heavily values personal choice over cost-benefit analysis may reasonably reject circumcision – especially one that rejects currently-widespread societal assumptions about parents making medical decisions for their children. Furthermore, many of the benefits of circumcision accrue only to men who have sex with women. For men who exclusively have sex with men and for men who do not have sex, the benefits and risks are close to equipose. There is a moral concern with performing a procedure that can thus tend to reinforce heteronormativity and sex-normativity.

Correction To Circumcision ACC

The original title of the essay was “Circumcision: Harms, Benefits, Ethics”. I wanted to have all the titles in the same format, as questions, so I titled my post “Is Circumcision Ethical?” Then lots of people got upset because the essay included focused on harms and benefits as much as (or more than) ethics.

This was totally my fault, not the fault of the authors. Sorry. For the sake of voting, please pretend the post had been titled “Circumcision: Harms, Benefits, Ethics”.

You can see my proposed titles for the other collaborations here; if any authors are unhappy with how I’ve phrased them, please let me know.

[ACC] Is Eating Meat A Net Harm?

[This is an entry to the 2019 Adversarial Collaboration Contest by David G and Froolow. Please also note my correction to yesterday's entry.]

Introduction

Many people around the world have strong convictions about eating animals. These are often based on vague intuitions which results in unproductive swapping of opinions between vegetarians and meat eaters. The goal of this collaboration is to investigate all relevant considerations from a shared frame of reference.

To help ground this discussion we have produced a decision aid making explicit everything discussed below. You can download it here and we encourage you to play around with it.

The central question is whether factory farmed animal lives are worth living; the realistic alternative to meat eating is not a better life but for those animals to not exist in the first place.

We begin by investigating which animals are conscious. Then, we compare the happiness literature to the conditions under which animals are factory farmed to figure out *if from their perspective* non-existence is preferable. And finally, we survey the more easily measurable impacts of meat eating on environment, finance, and health.

1. Consciousness

1.1. What is consciousness?

This essay isn’t about a general theory of consciousness. We tried to research this and our main takeaway is simply that consciousness is really, really weird. When we say that something is ‘conscious’ we mean simply that it’s ‘like something to be that thing,’ and that if we were that thing we’d care if ‘good’ or ‘bad’ things happened to us.

Of particular relevance is the conscious experience of pain and suffering, which we regard as morally undesirable when it occurs in ourselves or others. Most animals have damage sensors, but triggering these may not result in the subjective experience of ‘suffering’ if the animal is not conscious or the stimulus is a constant presence that has been accustomed to.

And this is the absolute crux of this investigation; if animals suffer under current farming standards to the point of preferring non-existence then there is a moral burden on meat eaters to justify eating them.

To resolve this, we will look for in different animals at two kinds of evidence for consciousness:

1. A brain architecture similar to humans resulting from the same evolutionary process
2. Behaviors that are hard to explain except with reference to having experiences

1.2. What parts of the brain are ‘responsible’ for consciousness?

You might assume consciousness is just caused by ‘the brain,’ because without it how could we think (or do) anything? However, a huge part of intelligent information processing happens in our brain without giving rise to conscious experience.

Try noticing what happens when you read the sentence “The dustmen said they would refuse to collect the refuse without a raise.” Notice how the word ‘the’ appeared twice? And how you read it both times despite Scott’s best efforts at conditioning you otherwise? Very good. Now, notice also that the word “refuse” appeared twice, but the first time you not only interpreted it but heard in your mind as a verb with the stress on the second syllable, and the second time as a noun. Before the words appeared in your conscious mind, as visual on a screen, as sound in your head, as a feeling of understanding the sentence, your brain already did all the hard parts of figuring out what they mean subconsciously, without you experiencing anything.

If you sever the spinal cord, the 1 billion nerve cells in the backbone connecting your brain to your limbs, you would lose complete sensation below the neck but you would continue to have thoughts and rich experiences. Or consider the cerebellum, the 69 billion (80%+) neurons which are responsible for motor control. If you cut off parts of it you may lose the mechanic ability to execute certain motions like playing a piano or walking but you will remain completely conscious with a sense of self, a memory, and the ability to plan for the future.

What is in common between the spinal cord and the cerebellum? The neural circuits operate in parallel with hundreds of independent input/output logic gates and all fire one way, instead of forming an interconnected multi-way circuit. There is no capacity for reflection, just predefined decision making. A necessary condition for consciousness (in people) is two neurons ‘talking’ to each other, rather than just passing information with deterministic modification up the chain of command.

Where consciousness appears to be generated is the posterior cerebral cortex, the outer surface of the brain composed of two highly folded sheets. Stimulate it electromagnetically and it’s like any other acid trip. The folded physical structure seems extremely important for generating consciousness because it maximizes the surface area exposure of neurons to each other. Many parts of your brain can be removed without major changes to your personality of intelligence, but if even small parts of the posterior cortex are missing surgery patients lose entire classes of conscious content: awareness of motion, space, sounds, etc.

It’s important to recognize that consciousness is not simply ‘caused by anything that happens in your brain.’ It’s a specific, fragile thing with distinct characteristics that differ from other neural activity that we associate with intelligence – therefore the relative intelligence of animals to humans does not necessarily map closely to their degree of consciousness.

1.3. Neural indicators of consciousness

All mammals have a cerebral cortex. Mice and rats have a smooth one; cats and dogs have some folding; and humans/dolphins/elephants have highly expanded and folded cortices. Therefore all mammals are probably conscious, although with large differences in vividness and complexity.

Birds and reptiles are a harder case because their brain evolution diverged much earlier. They have instead a cluster of neurons with chemical markers associated with differentiated layers of the neocortex but without the folded shape that maximizes connectivity.

By contrast, fish do not have any neural architecture unique to the consciousness-related parts of the brain and are probably unable to feel fear or pain in the way a human would – we strongly encourage you to read this article in full to convince yourself of this claim. Although fish show pain-like responses to harmful stimulus and do so less if given painkillers, this is true even when the entire telencephalon (which includes the forebrain) is removed so on balance it is unlikely they are having a qualitative experience to accompany that response.

1.4. Behavioral indicators of consciousness

Behavior seems like an obvious place to look for evidence of consciousness. However, any behavior can be explained by intelligence alone, or even sub-intelligent evolutionary ‘hard coding’. If you swat a fly, it will make loud ‘angry’ noises and go away. Your little brother would react the same. If you knew nothing about the neural architecture of flies, you might conclude that flies are just as conscious and capable of suffering as people.

One way around this is if we can design tests that indirectly look for mental states, such as the mirror test (whether an animal can recognize its reflection). But elephants (definitely conscious) routinely fail and at least one fish has passed, so we are wary about assigning much weight to these tests.

Another is to look for behaviors that map onto extremely complex emotional states that we observe in humans. If there is a large difference in intelligence but a great similarity in behavior, we can infer that the animal is having a similar conscious experience.

Starting simple; if you play with a dog it will act in the highly specific ways you might if you were feeling ‘joy.’ From a hormonal and intelligence perspective, stress and positive excitement are very similar states, and in non-conscious creatures we would have no reason to expect – for example – a creature to seek out stressors like a chew toy unless they had some positive feelings towards them. That we can so clearly tell how a dog is feeling is to us highly persuasive evidence of consciousness.

Dogs also exhibit something quite analogous to a theory of mind, for example they will comfort their owner if their owner is sad (but maybe this is a learned behavior)

Dogs are unlikely to be a special case; other animals of varying intelligence also exhibit complex behaviors indicative of consciousness:

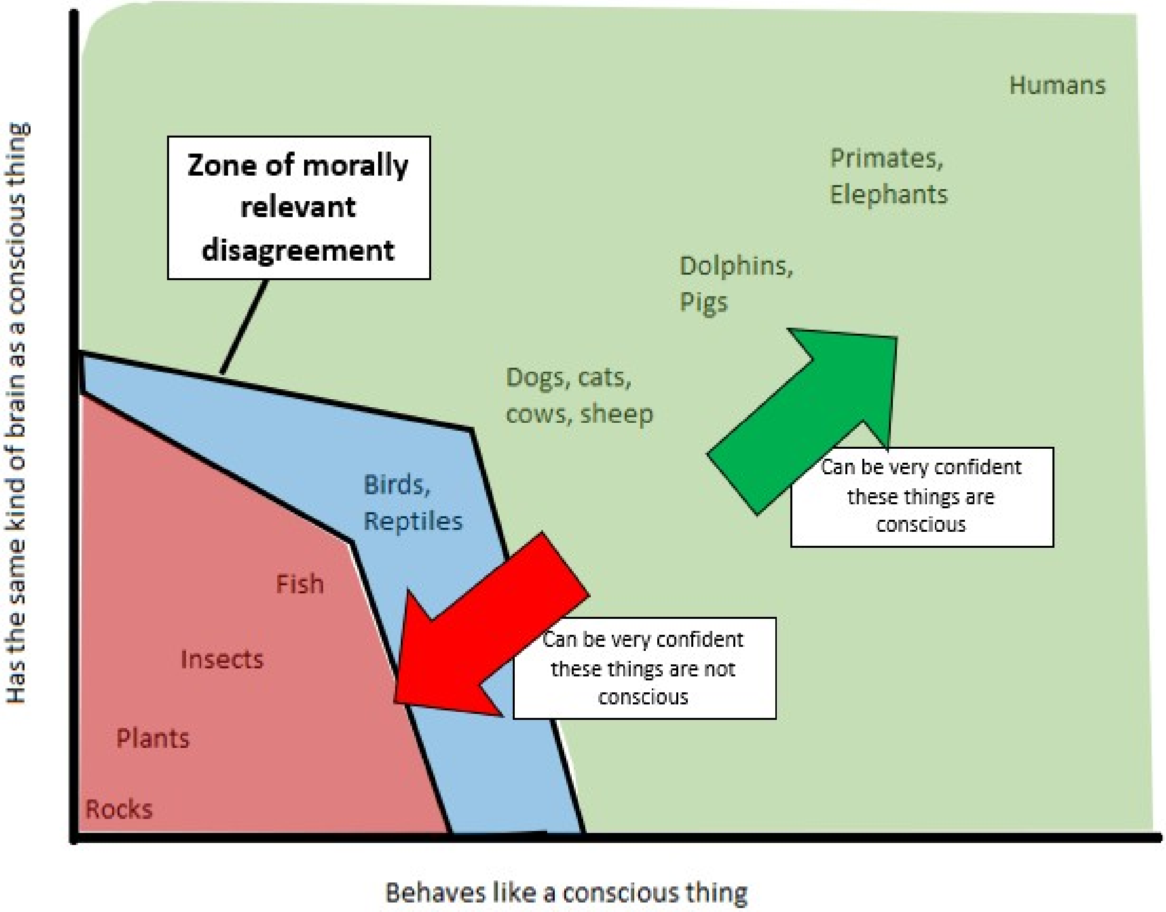
- Chimps who see another chimp lose a fight will direct more grooming behavior towards the loser, but not if they don't see that chimp lose the fight. [Link, popular coverage]
- Corvids who hide a treat when being observed will sneak back later and rehide the treat somewhere else, indicating (perhaps) a theory of mind and (certainly) 'mental time travel' of imagining the self in various future states. [Link, popular coverage]
- Dolphins given a test to discriminate between X and Y for a reward, but including the option of 'bailing out' of the test in exchange for a lesser reward, will bail out more often in more difficult tests, indicating a theory of metacognition (which we'd say is adjacent to – if not the same thing as – a theory of mind). [Link, popular coverage]

Spending time with animals (higher mammals, especially) makes it extremely hard to imagine they are anything but conscious, but we recognize that any behavior could be explained as an expression of intelligence without assuming conscious experience. However, we are reasonably confident that:

1. The range and complexity of behaviors conducted by animals correlates closely with the brain architecture we believe causes consciousness – the more complex the brain architecture, the more consciousness-like the behavior. This would be a substantial coincidence if in fact animals were not conscious.
2. Animals we intuit as conscious are less likely to exhibit 'glitching' behavior indicative of being a non-conscious rule-following automaton. There are many examples of 'glitches' in insect behavior (such as ant vortexes of death, repetitive digger wasp behavior [although maybe not] and moths failing to notice they are circling a candle), whereas there are very few examples of 'glitches' in mammal behavior. A humorous example of a glitch in bird behavior can be found in YouTube videos where the 'imprinting mechanism' of ducklings has confused them into thinking a dog is their mother.

1.5. What animals are conscious?

It's fair to reflect on the uncertainty in the above, but we'd be comfortable ascribing consciousness on the basis of neural architecture and behavior as follows:


Humans
Primates, Elephants
Zone of morally relevant disagreement
Dolphins, Pigs
Dogs, cats, cows, sheep
Can be very confident these things are conscious
Birds, Reptiles
Fish
Insects
Can be very confident these things are not conscious
Plants
Rocks
Has the same kind of brain as a conscious thing
Behaves like a conscious thing

There is good reason to believe all common land-based food mammals (cows, pigs, sheep, goats) are highly consciousness. On the other side, we think we can be reasonably confident fish don't suffer in a morally relevant way. We're not sure about chickens. We encourage you to read this overview of their behavior in full to convince yourself that their emotional and cognitive intelligence would group them with simple mammals if they had the same neural architecture.

However, since in most parts of the human brain 'intelligence' does not correspond to 'consciousness,' and because chicken brains are a clump of neurons with a different evolutionary history and lacking the distinct layered *and highly folded structure of the cerebrum, in the model we assume their likelihood of consciousness is 75%.*

A key part of this post is to quantify vague feelings about animal consciousness. This is similar to what Scott did with a sample of Tumblr respondents here and SSC reader Tibbar did with an MTurk sample here. Their results are expressed in terms of an animal's 'worth' relative to a human in percentage terms.

% Consciousness	Tumblr sample	MTurk sample
Human	100	100
Chimp	20	50
Elephant	14	100 (!)
Pig	3	20
Cow	2	33
Chicken	0.2	4
Lobster	0.03	1.6

However, it’s important to make the distinction between ‘worth of experience’ and ‘worth of suffering’ because while we might rather be a human than a chicken on a good day, feeling pain might be equally unpleasant in either body. Below is our best guess for a ‘universal’ estimate (i.e. even a meat eater ought to agree that these are plausible) – people who place a premium on animal experience, such as many vegetarians, would likely rate animal experiences higher:

	%Weight Suffering	%Weight Experience
Human	100	100
Chimp	90	50
Elephant	90	35
Pig	80	25
Cow	50	10
Chicken	10	1
Lobster	0.1	~0

What we mean by the above is, for example, if the unit of suffering was ‘being boiled alive,’ based on our understanding of how vivid their sense data would be assuming consciousness we would be roughly indifferent between being boiled once as any of a human, chimp, or elephant, 10 times as a chicken, or 1000 times as a lobster. However, we would be indifferent between living (assuming no scarcity or predators) for 1 year as a human, 2 as a chimp, 4 as a pig, or 100 as a chicken.

In the model, the moral impact of a farmed animal is its likelihood of consciousness times the moral weight of its suffering if its life is, based on information in the following sections, 'worse than non-existence,' otherwise the moral weight of its existence.

2. How many animals are farmed and under what conditions?

2.1. Animals eaten per capita

The OECD records the exact weight of meat consumed per year, and so by dividing this by the carcass weight we get the per-year per-capita animal consumption by country. This is consistent with other estimates on the web.

Animals / Capita consumed by four major categories in 2018

	Australia	Canada	EU27	UK	USA	World
BEEF	0.05	0.05	0.03	0.03	0.07	0.02
PIG	0.27	0.20	0.44	0.22	0.29	0.15
POULTRY	29.42	22.53	15.70	18.53	33.12	9.44

	0.38	0.05	0.07	0.22	0.02	0.09
SHEEP						

The conditions most animals are farmed in may surprise you. Since a few large factory farms account for most animals farmed, the typical farm may well be a small mom-and-pop operation but the typical food animal is raised on an industrial scale. Further, ‘ag gag’ laws limit facts about animal conditions reaching public awareness and are arguably designed explicitly to allow companies to mislead consumers about the conditions most animals are farmed in.

Animal rights organizations will frequently quote a study by the Sentience Institute that over 99% of animals eaten in the US are factory farmed. Although a biased source, it’s consistent with other government estimates and we get similar results when replicating their methodology with USDA figures. What really drives this statistic is, as seen above, chickens form the vast majority of the farmed animal count and are almost exclusively farmed industrially.

	My calculation for % factory farmed using Sentience Institute methodology	Sentience Institute estimate for % factory farmed
BEEF	64%	70%
PIG	94%	98%

POULTRY	~100%	~100%
SHEEP	64% (based on being woolly cows)	Not included

Comparing across countries is difficult, but it seems that America is slightly more industrialized than the EU. My best estimate is that the difference is not significant enough to make a moral difference. If you eat meat and cannot explicitly trace the source you are most likely eating factory farmed meat.

2.2. What is it like to be farmed industrially?

The definitive feature of factory farming is that market incentives lead to a paperclip maximizer situation where producing as many animals as possible takes precedent over concerns about animal welfare. Consequently, the ‘experience’ of being factory farmed is best understood as a particular form of slavery where cruelty is the side effect of a system designed to maximize economic output.

2.2.1. Chickens

Chickens can be raised in two ways: in cages or in a shed. Cage-rearing chickens is typical in developing economies such as China, but in the West cages are used for egg laying hens only. Slaughter chickens in the West are raised instead in a large ‘broiler’ shed covered with liter.

Broadly speaking, caged chickens have literally no human analogue in terms of how much they suffer. They live in a state of constant pain and anxiety, barely able to move. The only mercy is that they do not suffer for long. In this analysis we focus on meat-eating in a Western context. So, we model 100% of chickens as being broiler farmed.

A factory farmed slaughter chicken lives for approximately 47 days, during which time it grows to a weight of 2.6kg (42 days and 2.5kg in the EU). This is analogous to a newborn human baby reaching adult weight by the first birthday. To achieve this rapid weight onset, a combination of force-feeding, drugs and high-energy feed is used. But the worst culprit is selective breeding. In a study by Kestin, between 2% and 30% of broiler chickens, depending on the breed, had a gait score between 3 and 5 on a 5 point scale (1 no issues, 3 obvious gate defect, 5 unable to move at all). But 100% of a control group bred randomly and then raised under the same conditions had no or minor mobility issues. Selection for quick growth rather than fitness in the wild leads to a high rate of heart attacks and other organ failure. In the final weeks of life, the chickens often outgrow the ability of their legs to support them, making broken or otherwise failed legs endemic in the industry.

Photos of cage-raised chickens, borderline NSFW

Photos of broiler shed chickens, NSFW

Because it is cheaper to only change the litter between flocks it is a major source of bacterial infection and especially contact dermatitis (rashes and lesions on the chicken's feet and lower body). It is common practice in the EU (not the US) to remove a portion of the flock a week before slaughter time to create enough space for the remaining birds to reach their usual slaughter weight, suggesting there isn't much free space for the birds. Birds whose legs fail will often dehydrate to death. We don't want to overegg this – a dead bird is an unproductive bird and only around 3.3% of the flock die during growth for any reason – but remember that this is a 3.3% chance of dying in only six or seven weeks.

De-beaking is common in broiler chickens (universal in laying chickens). One reason for debeaking is to reduce cannibalism which occurs because the birds are so stressed – pet chickens will peck each other to establish a dominance hierarchy but don't kill and eat each other. Beaks are sensory and manipulative tools for chickens, so this is analogous to cutting the fingers of prisoners off without anesthetic to lower the probability of escape.

Photos of debeaking, NSFW

Shed chickens have it slightly better. They have a small amount of mobility, are able to do some natural activities such as socializing and digging in the dirt with their claws (but not usually their beaks) and have a little natural light from windows in the warehouse. On the other hand, chickens only cannibalize each other when very stressed and the strain on their systems from the massive growth they are forced to undertake causes considerable pain.

We think it is reasonable to say that broiler chickens exist in a state worse than death – in the model, we assume chicken-days are equivalent to -2 human-days (you'd rather have your life be 2 days shorter than have to experience a day of chicken life), but your intuitions may differ substantially.

2.2.2. Pigs

Pigs are the next-most commonly farmed food animal. There are two major sources of cruelty in pig production; the raising of the food-pigs themselves and the creation of new food-pigs from breeding pigs.

Breeding sows are confined to a 'gestation' or 'sow' crate for most of their lives. These are only slightly larger than their bodies, making it impossible to turn around or even lie down. Generally the floors are made of slats or iron rungs to allow manure to fall through. These slats can hurt the sensitive feet of the pigs, and the fact that they are confined directly above their own manure means they are exposed to ammonia toxicity, which leads to respiratory conditions common in confined sows (and presumably smells incredibly distressing). Pigs are highly intelligent, and the unstimulating confinement means that the pigs engage in repetitive stress behaviors such as biting at the metal bars of their cage – this can cause further harm such as mouth sores.

Shortly before birth, the pregnant sow is taken to a ‘farrowing crate’ – even more restrictive than a sow crate. This is designed to separate the mother from the piglet so the piglet can nurse without being crushed (piglets being crushed can happen in the wild, but it is rare – this is a problem almost entirely caused by the confined conditions the sow is kept in). The crate is so tight the mother cannot even see her baby once it is born, and the baby is taken away after about 17-20 days. The piglet is then prepared to be fattened for slaughter, and the mother is either re-impregnated and returned to the gestation crate or slaughtered herself if she is unlikely to survive another pregnancy.

Photos of gestation and farrowing crates, surprisingly SFW

Piglets being prepared for slaughter are castrated and have their tails docked, often without anesthetic. Unlike chicken beaks, pig tails don’t really seem to serve any purpose, but pigs show pain behavior towards their stumps suggesting that it is very sensitive even after being docked. The tails are docked to prevent other pigs biting it and causing an infection – again, behavior which is vanishingly rare in the wild and therefore seems to be a stress response to the conditions they are kept in. Piglets may also have their teeth clipped to prevent biting but we can’t find figures on how common it is. Pigs prepared for slaughter are kept in 'finishing crates' which seem to run from anywhere between a slightly larger sow crate (larger only in the sense that it is bigger – finishing pigs are much larger so don't have any more space to turn around or express natural behaviors) and something a little more like a traditional farmyard pen but indoors – six or seven pigs confined to a small pen where they have just enough space to walk around if they want to.

Pig-tures of finishing crates, SFW

Pigs are highly intelligent animals, and when not confined to stalls will spend hours playing and rooting around in the mud. The pigs consumed for food will be in constant low-level pain and sows used for breeding will be in quite intense pain constantly. It is hard to imagine a more distressing event than having your child taken away from you or being taken away from your mother, and we might imagine that the constant lack of stimulus for both food and breeding pigs causes considerable boredom and sadness.

It is a harder call whether pigs exist in conditions worse than death. My intuition is that food pigs are right on the border, and breeding pigs would strongly prefer to not live. In the model we assume that a pig-day is worth -1 human days.

2.2.3. Cows

Cows are the only animals routinely farmed in conditions approaching the way people imagine farming to be in their head – that is, in a field where they have enough space to move around and socialize. Factory farmed cows spend six to twelve months being raised outdoors in fields (the sorts of cows you see dotted around the countryside), and are then transported to 'feedlots' for their last few months where they are fed an artificial diet of corn and soy that is very hard on their bodies and can cause illnesses such as ulcers. Note that almost all cow-meat can be labelled ‘grass fed’ because most cows spend their first year in fields eating grass: doesn’t mean that is where the majority of their final slaughter weight came from! Much like pigs, cows have complex social hierarchies and being put on a feedlot with thousands of other cows is depressing.

Pictures of feedlot, SFW

Beef cattle also endure individual painful events like castration, branding and dehorning (often done without anesthetic) and transport in cramped crates for long periods of time. It is unclear how to incorporate the impact of these events on the animal’s overall quality of life. We’d suggest animals raised on a field have the same quality of life as a traditionally farmed animal and animals raised on a feedlot have a quality of life moderately worse than a typical elderly human. By quite a long way cows appear to have it the best of all factory farmed animals and have lives that are clearly worth living.

We think we’d be pretty content to live as a cow in a field, but cows on feedlots seem to have lives that are closer to food-pigs. Approximately averaging these out over a cows’ lifetime, we model 1 human day as equal to 10 cow days. We can’t find good information on how sheep are factory farmed so we’ve assumed they’re just woolly cows for the purpose of estimating their quality of life.

2.2.4. Value of animal life

Quantifying and comparing subjective experiences of farmed animals is hard because there is no ‘natural unit’ of suffering or experience. We proceed instead by asking ‘at what factor are we willing to skip/trade off days of my life to live as any particular animal under factory farmed conditions’ as described above. Factory farmed cow lives seem greatly preferable to non-existence, pigs spend most of their time experiencing some form of chronic or acute suffering, and chickens have truly awful lives from birth to slaughter.

Slaughter itself may be a morally relevant part of valuing an animal’s life. In principle, animals are stunned before slaughter so that the process is painless (and evidence suggests animals are not distressed by watching other animals being stunned). However, in practice animals are often not insensible to pain when they are skinned or carved up, either because of poor training (paywalled WaPost link), religious beliefs around the way meat should be prepared (link, further figures) or just because of a culture of laxity and cruelty (INTENSELY NSFW link of animals being abused by slaughterhouse workers, SFW-ish PDF report of the investigation). ‘Ag gag’ laws and other efforts by farmers to avoid bad press prevent serious scholarly investigation of the extent of the issue, but the AnimalAid hidden camera investigation linked as a PDF above found evidence of criminally cruel treatment at one of the three abattoirs observed and evidence of mistreatment at another.

Also, some people believe in a specific ethical obligation to not kill conscious creatures that do not want to be dead, so that slaughter of even humanely stunned animals is immoral. We instead take the consequentialist view that there is a symmetric value in actualizing the existence of conscious creatures that want to be alive (again, a farmed animal’s practical alternative is non-existence).

In the model, we assume that at human-level consciousness the experience of a typical human life-day is worth the factory farmed experience of 10 cow days, -1 pig days, and -0.5 chicken days. We value a year of perfect health at $50,000 in line with typical healthcare priority setting in the US (it is much less in the UK and Europe – closer to £30,000), *where a typical western life-year is about 86% of that ($43,000). Slaughter is not included as even the cruelest slaughter imaginable would be QALY-negligible if averaged over an animal's whole lifespan.*

3. What lives are worth living?

Speaking loosely, evolution does not care how happy your life is as long as you a) exist and b) pass on your genes, and so has come up with a number of 'patches' to the conscious reward system to ensure animals are never too satisfied to stop competing to breed but never so dissatisfied they would prefer to be dead. Instead, what we have is roughly a baseline set of happiness from which we deviate when good or bad things happen, but to which we almost always return to. If this is also true of animals then it does not matter that we perceive their lives as described above as intolerable; if we were actually forced to live as that animal then we would observe ourselves hoping we don't die painlessly in our sleep. Since we cannot ask animals directly if they consider their lives worth living, we instead look at the conditions in which people report changes in happiness or commit suicide, and compare these to the lived experience of factory farm animals.

3.1. Habituation and Happiness Set Points

Habituation is a "decrease in response to a stimulus after repeated presentations." The simplest form of learning, it is caused by neural processes that regulate responsiveness to different stimuli. When we are repeatedly sent a signal, especially if it is highly frequent and hasn't recently changed in intensity or duration, we consciously experience it less acutely. This makes sense. If consciousness is about complex reflection, after we've already processed a signal and determined a response if any, the response becomes subconsciously automatic.

However, we don’t habituate just to local physical sensations, like the ticking of a clock or the pressure of a shirt against our skin. We habituate to pain and suffering as well, even to large shocks to the system. In the literature this is known as the disability paradox, whereby a majority of those with severe disabilities report having a good or decent quality of life, even when to external observers it seems like a life not worth living (although this story is nuanced, and some of the improvement is related to the ability of intelligent humans to adapt by changing their lifestyle).

Nevertheless, the consensus in happiness research is that people have a fairly stable general level of baseline happiness which they return to after certain large changes. In a famous study by Brickman (1978), paraplegics and lottery winners reported similar levels of happiness before and after what one might assume was a life-altering development, either extremely negative or extremely positive. And in a twin study of several thousand by Lykken and Tellegen, it was found that about 50% of variation in Well-Being scale of the Multidimensional Personality Questionnaire is associated with genetic variation, and less than 3% (!) of the variance can be accounted for by any of socioeconomic status, educational attainment, family income, marital status, or religious commitment.

3.2. Lessons from suicide

Sometimes, humans decide that their lives are intolerable and commit suicide. Interestingly, we basically never observe this in other animals. The only maybe-credible anecdotal claims are of dolphins, which are highly intelligent and can commit suicide by not breathing (breathing by dolphins is probably an active choice, rather than an automatic process as regulated in humans by the amygdala).

This doesn’t mean we can conclude farmed animals prefer living because animals might lack the theory of mind or intelligence to act on their preference to stop existing. However, if humans in extremely poor conditions overwhelmingly do not choose suicide, we might infer that animal lives of roughly similar quality would also be worth living. Two well-studied areas where humans are placed in extremely poor conditions are slavery and terminal disease.

3.2.1. Slavery

The historical consensus is that while slavery caused extreme stress and suffering, the rate of suicide by black slaves was quite low. According to the 1850 U.S. census, slaves had a suicide rate of 0.72 per 100,000 while whites had a rate of 2.37 and freed slaves a rate of 1.15. From the Federal Writer’s Project Slave Narratives which documented incidences of resistance, only 1.2% were acts of suicide. Further, when slaves did resort to suicide, it was usually in response to deterioration in their circumstances or unfulfilled expectations, rather than being explained by living under the most brutal conditions – this is consistent with a ‘happiness set point’ theory.

To be clear, we are not saying that because enslaved Africans committed suicide at lower rates than free whites slavery wasn’t ‘that bad.’ There is a substantial academic literature explaining the cultural reasons for the difference. The observation is simply that the main explanatory factor of whether a slave thought it worth living was not how bad his or her life objectively was.

3.2.2. Terminal Patients

In a review of the psychological profile of patients in palliative care of 18,000 terminal cancer patients, a small number of which committed suicide, it was found that some of those who committed suicide:

> “…presented functional and physical impairments, uncontrolled pain, awareness of being in the terminal stage, and mild to moderate depression… however, the loss of, and the fear of losing, autonomy and their independence and of being a burden on others were the most relevant.”

The presence of significant pain or even depression (what we might refer to as ‘objective suffering’) was not a significant factor in predicting suicide, the best revealed preference we have for whether a life is considered worth living by the morally relevant actor experiencing it.

3.3. What does this mean?

Preference for living is a strongly mean reverting process. The scientific literature and historical examples from slavery and terminal illness both suggest that humans will habituate to almost anything that is done to them. In our ancestral environment life was really, really hard. Brutally hard. And it makes sense that even in environments that modern folks would instantly label as ‘much worse than non-existence,’ evolution made sure that we would continue to have the strength of will to not only survive but want to.

How far does this go? Does this mean that animals, no matter how much suffering they experience, prefer living? Reasonable people can disagree. As detailed above, factory farmed animals – especially chickens – do not exist in anything remotely resembling an ‘ancestral environment’. Chronic stressors are more likely to cause permanent changes in happiness than acute changes, and, as in the case of chicken cannibalism absent food scarcity, it is likely that factory farming creates an environment that is not only unpleasant but one which even the astounding level of habituation we observe in humans might not adjust above the ‘worth living’ watermark.

In the model we take the habituation literature seriously, estimating by how much deviations from long-term set point of average quality of life (0.86) are neutralized by habituation. Based on the previous sections and how much worse factory farmed conditions are than an ‘ancestral environment’ in which habituation would be calibrated, we estimate that humans habituate by 80%, cows 70%, pigs 60%, and chickens 50%. Even though before we had assumed that we would prefer non-existence to being a factory farmed pig, and would dislike being a chicken twice as much, once habituation is taken into account pig lives are slightly better than non-existence, and chicken lives are quite bad but have a (negative) moral weight not much larger than the positive one of cows. However, since we consume so many chickens in the end they dominate the analysis.

4. Health Considerations

4.1. Nutrition

Animal protein sources such as meat, fish, eggs, and dairy contain a good balance of the 20 amino acids that we need for almost every metabolic process in the body, whereas individual plants are generally deficient in mix or concentration. The same is the case for micronutrients: animal protein sources are much higher in vitamin B12, vitamin D, the omega-3 fatty acid DHA, heme-iron and zinc.

Animal products provide most of the zinc in US diets, and meat, poultry, and fish provide iron in the highly bioavailable heme form. For example, the panel setting the new Dietary

Reference Intakes recommends an 80% higher daily iron intake for vegetarians.

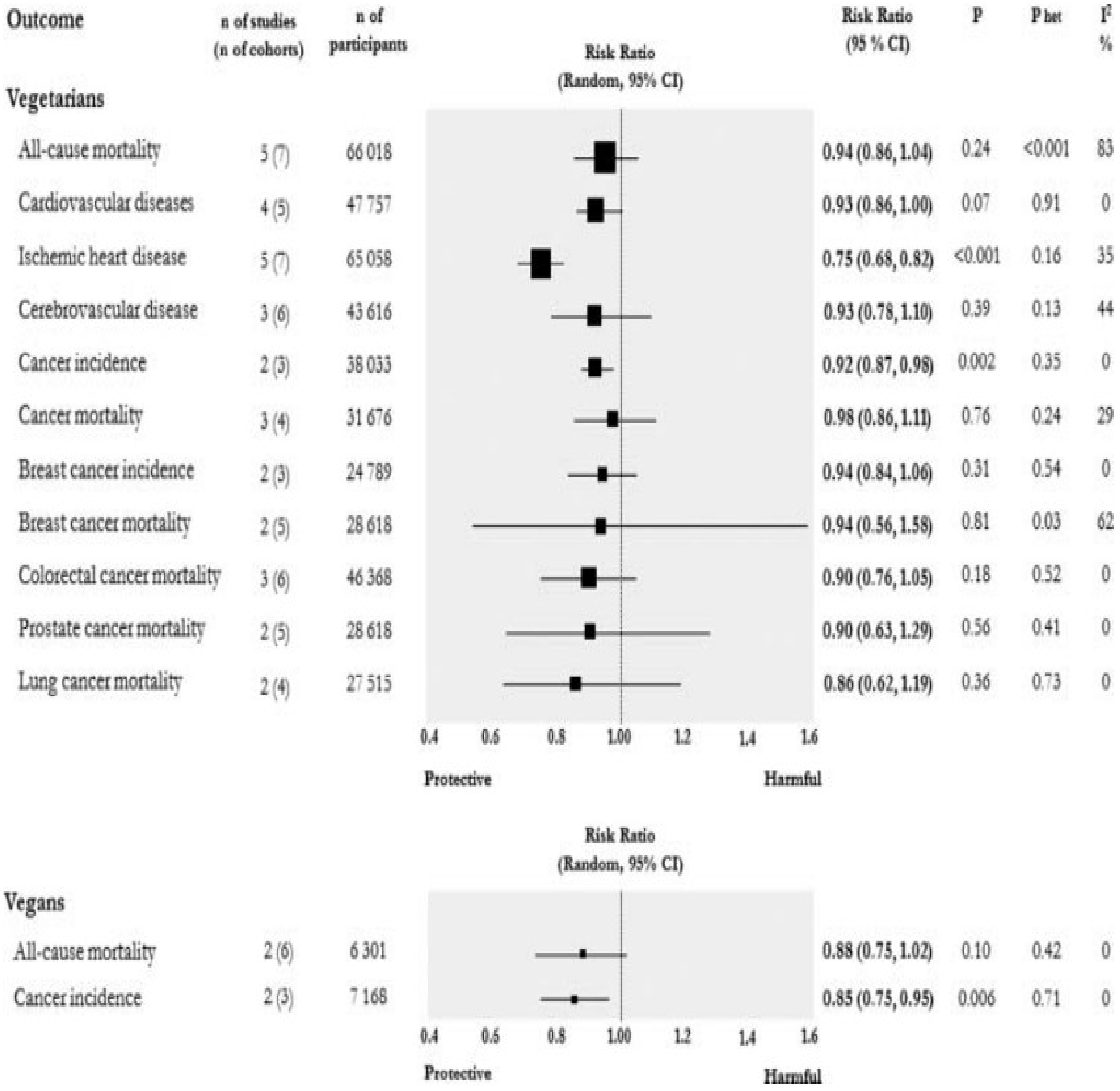
Concern also has been expressed about the difficulty that children have in obtaining adequate energy and nutrient intake from bulky plant-based diets. Dutch infants consuming vegan diets had poorer nutritional status and were more likely to have rickets and deficiencies of vitamin B-12 and iron, and the World Health Organization strongly recommends animal products for infants to ensure enough calcium, iron and zinc.

Whether or not a plant based diet is viable from a nutritional perspective depends mostly on whether you have the economic means to consume a wide variety of food sources, and may be riskier for small children or those whose ancestry is from regions where meat-eating was prevalent.

4.2. Long Term Health Outcomes

Estimating the long-term health outcomes of eating certain things is difficult because food is highly bound up in the culture we live in and culture correlates to just about every health outcome you could possibly imagine. Even less conveniently, nutritional science is highly anti-inductive; if a particular food group is identified as being healthy people with an interest in being healthy flock to that food group, and people with an interest in being healthy are likely to be healthy for a bunch of reasons regardless of diet.

So here’s a nice headline result: vegetarians have less heart disease with extremely high certainty, and probably less cardiovascular disease and cancer too. Most of the studies in that meta-analysis have had some of the really obvious stuff adjusted away (race, income, etc.) but not all studies adjust for all confounders, and we should be cautious about trusting studies that ‘adjust for confounders’. If you ignore confounders then the answer is clear; eating vegetarian is good for you in every single way we can measure (including, possibly, circulating testosterone in defiance of stereotypes about meat eaters!).



Outcome	n of studies (n of cohorts)	n of participants	Risk Ratio (95 % CI)	P	P het	I^2 %
Vegetarians						
All-cause mortality	5 (7)	66 018	0.94 (0.86, 1.04)	0.24	<0.001	83
Cardiovascular diseases	4 (5)	47 757	0.93 (0.86, 1.00)	0.07	0.91	0
Ischemic heart disease	5 (7)	65 058	0.75 (0.68, 0.82)	<0.001	0.16	35
Cerebrovascular disease	3 (6)	43 616	0.93 (0.78, 1.10)	0.39	0.13	44
Cancer incidence	2 (3)	38 033	0.92 (0.87, 0.98)	0.002	0.35	0
Cancer mortality	3 (4)	31 676	0.98 (0.86, 1.11)	0.76	0.24	29
Breast cancer incidence	2 (3)	24 789	0.94 (0.84, 1.06)	0.31	0.54	0
Breast cancer mortality	2 (5)	28 618	0.94 (0.56, 1.58)	0.81	0.03	62
Colorectal cancer mortality	3 (6)	46 368	0.90 (0.76, 1.05)	0.18	0.52	0
Prostate cancer mortality	2 (5)	28 618	0.90 (0.63, 1.29)	0.56	0.41	0
Lung cancer mortality	2 (4)	27 515	0.86 (0.62, 1.19)	0.36	0.73	0
Vegans						
All-cause mortality	2 (6)	6 301	0.88 (0.75, 1.02)	0.10	0.42	0
Cancer incidence	2 (3)	7 168	0.85 (0.75, 0.95)	0.006	0.71	0

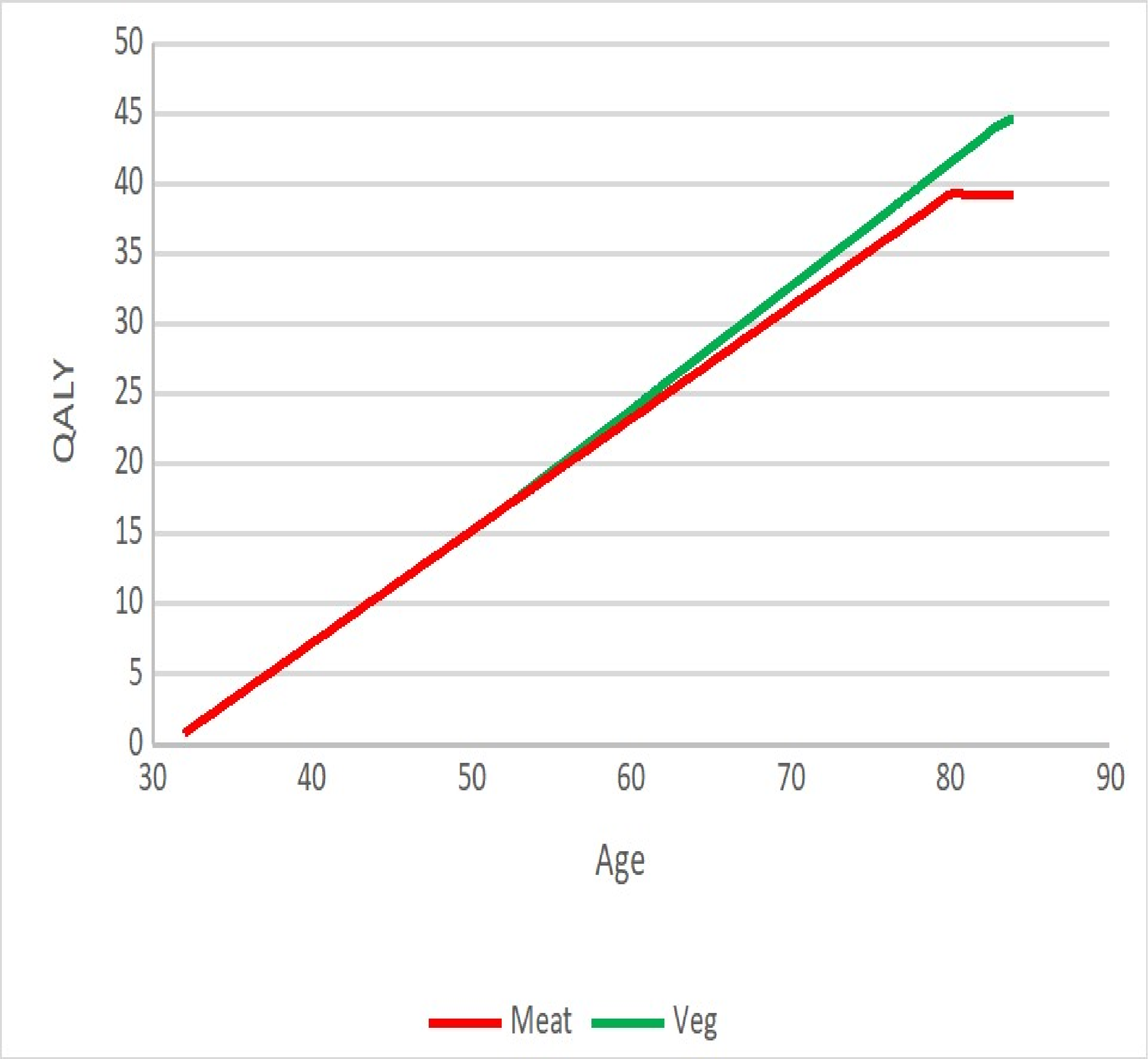
If you are interested in confounders: There are a handful of cool natural experiments, taking groups with reasons to eat certain food but not bother with the associated healthy lifestyles, which are the closest we are likely to come to a true experiment in this area. In particular, the American

Adventist Health Studies are pretty much state of the art in the field from what we can see. Adventists have quite unique dietary habits, brought about by religious prohibitions on certain foodstuffs which some Adventist churches follow and some don’t. Consequently, if you are an Adventist you are functionally ‘randomized’ into different food-eating conditions depending on which church you attend, and this randomization can be exploited by researchers.

Based on the Adventist Health Studies, a vegetarian diet increases life expectancy by around 3.6 years. The less meat you eat, the healthier your BMI and the less likely you are to get diabetes.

Overall we might expect lacto-ovo vegetarians to have a health related quality of life around 10% better than a meat eater, with most of this benefit being apparent 20 years after making the switch to a vegetarian diet.

You could complicate this picture a lot (especially by introducing future discounting) but we think the general principle that if you value life-years towards the end of your life you should likely go vegetarian is well demonstrated by the data:



One final point on how meat might affect your lifespan; there is a growing awareness of the fact that industrially produced meat is an ideal breeding ground for zoonotic disease, and that those diseases can mutate and jump to humans very quickly. Previous pandemics such as H1N1 (‘swine flu’) and H5N1 (‘bird flu’) may have originated with farmed animals, and were rapidly spread by the close contact of unhealthy animals and global nature of the meat supply chain. At the margin, eating meat probably increases the probability of a global pandemic but there isn’t good evidence on how much your individual consumption affects things at the margin.

In the model we take the Adventist study result at almost face value, estimating that eating vegetarian will increase your lifespan by 3 years, and include constant low costs due to possible nutritional deficiency and moderate benefits to health that appear later in life.

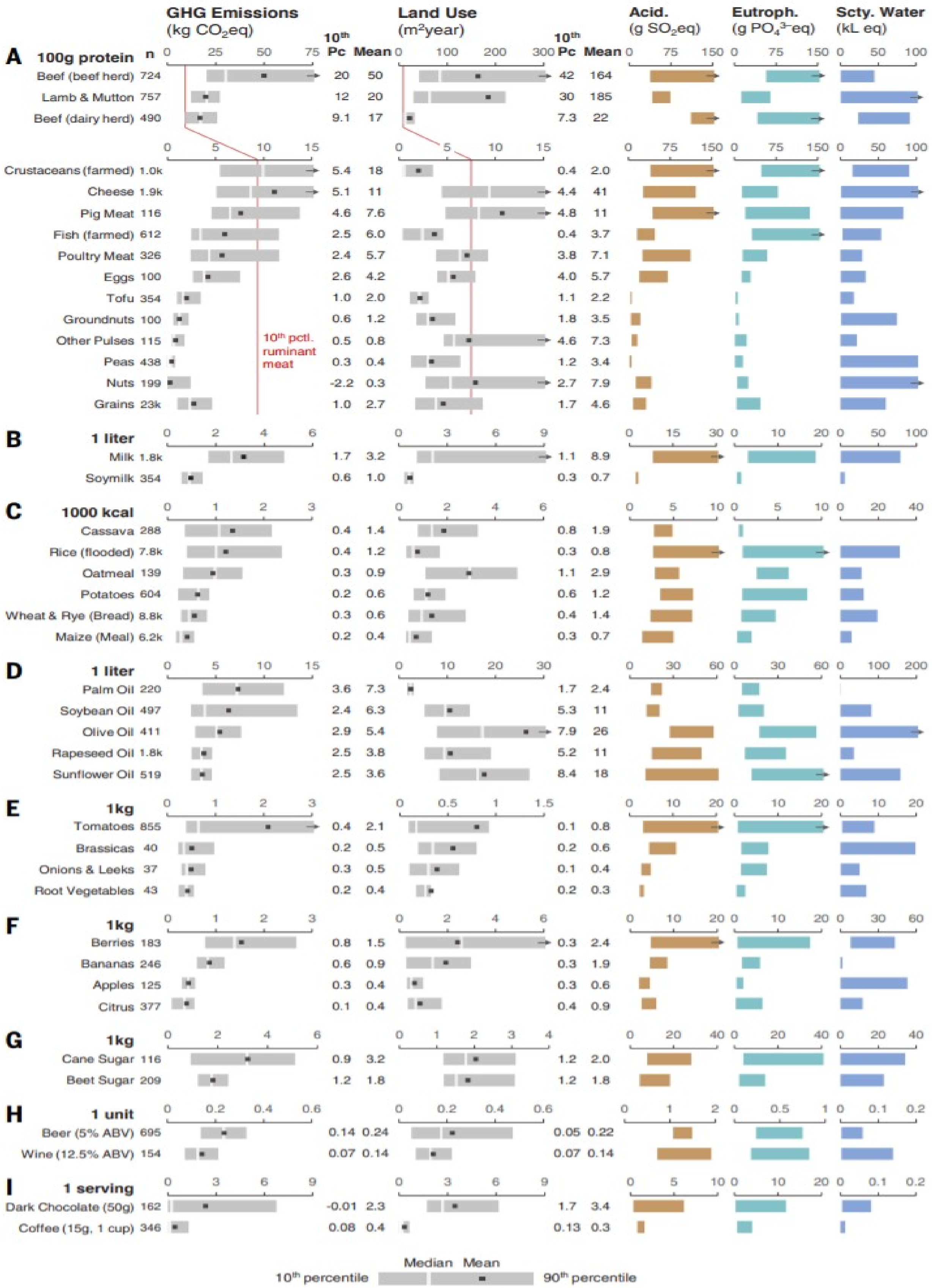
5. Environmental Impact

The environmental impact of meat consumption is difficult to measure and aggregate because the numbers are sensitive to the type of farming used and location, and any serious attempt requires a massive aggregation of different data sources. The best we could find was by Oxford's Zoology department which combined data from 570 studies with a median reference year of 2010 covering 40,000 farms and 1600 processors, packaging types, and retailers, in 119 countries of 40 products representing ~90% of global protein and calorie use, focusing on five environmental impact indicators: land use, freshwater withdrawals weighed by local scarcity, greenhouse gasses, acidifying emissions, and eutrophying emissions.

Overall, today's food supply chain creates ~13.8 billion metric tons of CO2 equivalents, which is about 26% of human caused emissions. It also causes 32% of global terrestrial acidification and 78% of eutrophication. It's also very resource intensive, covering about 40% of the world's ice- and desert- free land, and driving roughly 90% of global scarcity-weighted water use because irrigation returns less water to rivers and groundwater than industrial and municipal uses and predominates in water-scare areas and times of the year.

Because of different technologies and other environmental variables, the environmental impact of any foodstuff can vary widely. For example, ninetieth-percentile GHG emissions of beef are 105kg of CO2eq per 100g of protein, and land use (area multiplied by years occupied) is 370 m2 •year. These values are 12 and 50 times greater than 10th-percentile dairy beef impacts.

However, as you can see below the environmental impact from meat dwarfs that of other nutrition sources:


GHG Emissions (kg CO_2eq)
Land Use (m^2year)
Acid. (g SO_2eq)
Eutroph. (g PO_4^{3-}eq)
Scty. Water (kL eq)
10th Pc
Mean
A 100g protein
n
Beef (beef herd) 724 20 50 42 164
Lamb & Mutton 757 12 20 30 185
Beef (dairy herd) 490 9.1 17 7.3 22
Crustaceans (farmed) 1.0k 5.4 18 0.4 2.0
Cheese 1.9k 5.1 11 4.4 41
Pig Meat 116 4.6 7.6 4.8 11
Fish (farmed) 612 2.5 6.0 0.4 3.7
Poultry Meat 326 2.4 5.7 3.8 7.1
Eggs 100 2.6 4.2 4.0 5.7
Tofu 354 1.0 2.0 1.1 2.2
Groundnuts 100 0.6 1.2 1.8 3.5
Other Pulses 115 0.5 0.8 4.6 7.3
Peas 438 0.3 0.4 1.2 3.4
Nuts 199 -2.2 0.3 2.7 7.9
Grains 23k 1.0 2.7 1.7 4.6
10th pctl. ruminant meat
B 1 liter
Milk 1.8k 1.7 3.2 1.1 8.9
Soymilk 354 0.6 1.0 0.3 0.7
C 1000 kcal
Cassava 288 0.4 1.4 0.8 1.9
Rice (flooded) 7.8k 0.4 1.2 0.3 0.8
Oatmeal 139 0.3 0.9 1.1 2.9
Potatoes 604 0.2 0.6 0.6 1.2
Wheat & Rye (Bread) 8.8k 0.3 0.6 0.4 1.4
Maize (Meal) 6.2k 0.2 0.4 0.3 0.7
D 1 liter
Palm Oil 220 3.6 7.3 1.7 2.4
Soybean Oil 497 2.4 6.3 5.3 11
Olive Oil 411 2.9 5.4 7.9 26
Rapeseed Oil 1.8k 2.5 3.8 5.2 11
Sunflower Oil 519 2.5 3.6 8.4 18
E 1kg
Tomatoes 855 0.4 2.1 0.1 0.8
Brassicas 40 0.2 0.5 0.2 0.6
Onions & Leeks 37 0.3 0.5 0.1 0.4
Root Vegetables 43 0.2 0.4 0.2 0.3
F 1kg
Berries 183 0.8 1.5 0.3 2.4
Bananas 246 0.6 0.9 0.3 1.9
Apples 125 0.3 0.4 0.3 0.6
Citrus 377 0.1 0.4 0.4 0.9
G 1kg
Cane Sugar 116 0.9 3.2 1.2 2.0
Beet Sugar 209 1.2 1.8 1.2 1.8
H 1 unit
Beer (5% ABV) 695 0.14 0.24 0.05 0.22
Wine (12.5% ABV) 154 0.07 0.14 0.07 0.14
I 1 serving
Dark Chocolate (50g) 162 -0.01 2.3 1.7 3.4
Coffee (15g, 1 cup) 346 0.08 0.4 0.13 0.3
Median Mean
10th percentile
90th percentile

In total, meat, aquaculture, eggs, and dairy use ~83% of the world's farmland and contribute 56-58% of food's different emissions, despite providing only 37% of our protein and 18% of our calories.

Because of substitution effects and nutritional requirements, it is unclear exactly how much of these resources would be freed up if we switched away from eating meat. In a simple model where we assume 'protein is protein and calories are calories' and freed up land would only remove carbon through natural vegetation and accumulated soil carbon, "moving from current diets to a diet that excludes animal products would result in reducing food's land use by 3.1 (2.8-3.3) billion hectares (a 76% reduction), including a 19% reduction in arable land; food's GHG emissions by 6.6 (5.5-7.4) billion metric tons of CO2eq (a 49% reduction); acidification by 50% (45-54%); 19 eutrophication by 49% (37-56%); and scarcity-weighted freshwater withdrawals by 19% (−5 to 32%) for a 2010 reference year."

It's difficult to translate these tradeoffs into 'one number' that captures the environmental impact of meat eating. Land which currently supports animal farming is likeliest to be least suitable for agriculture or urban dwellings, and the costs of climate change, the value of species diversity, and the future scarcity of freshwater are all difficult to measure.
One (very approximate) way is to assume that the human race is polluting the planet as much as it can sustainably (at current technologies it's a lot worse, but we're science-optimists), and that the long-term impact of a reduction in pollution is a proportionally inverse change in the equilibrium human population.

Water shortages, eutrophication, and acidifications are serious environmental concerns but can be managed. Greenhouse gas emissions and land seem like the most important constraints. Combining the statistics from the total absolute resource impact of the food supply chain with the relative impact of switching from meat-eating, we get that if the planet went vegetarian we'd reduce emissions by 12.5% and free up 30% of the world's non-desert/ice land most of which we would not be able to immediately put to good use. We think it's reasonable that the total reduction in 'human pollution and resource use' would be about 10%. Since raw resources and access to clean environment aren't the only limiting factors on population size this result should be adjusted downward by a reasonable factor.

In the model, we assume that without meat farming there would be about 2.5% more capacity for population or quality of life, 30% of which (completely uneducated guess) would actualize as more lives, and 70% of which would actualize as better lives (and would count at 1/5th weight due to habituation).

6. Cost of Switching Diets

One good reason not to switch to a vegetarian diet would be if doing so was prohibitive, either because of the financial cost or the satisfaction from eating.

6.1. Cost per meal

In a trivial sense, vegetarianism is clearly cheaper. It takes more time and energy to grow plants that we feed to animals and then eat the animals than it does to just eat the plants themselves. This is borne out by research into the cost per calorie of various foodstuffs. Of course, humans don't eat exactly the same food animals eat, and vegetarians are for some reason unwilling to just drink 2000 calories of canola oil every day.

The cost of various types of diet seem to be bizarrely under-studied (or perhaps crowded out by the literature on trying to get people to stop eating junk food). The one academic source I found seems to be really high quality though. Here is the paper and here is a nice associated blog post.

At all income levels meat eaters spend about $20 more per week than true vegetarians (~$1000 / year). Adjusting for all controls (including politics and body weight which may be affected by vegetarianism) reduces this number to a savings of $11.1/week, which is what we use in the model.

6.2. Psychological costs

However, one switching cost that might not be trivial is the psychological importance of having meat in your diet.
Most vegetarians eventually enjoy vegetarian food as much as meat (not sure if that’s just survivorship bias), but anecdotal experience from everyone I know who has gone vegetarian says that there is a really horrible period of adjustment of at least a couple years where you want to eat meat and can’t.

One reasonable measure of psychological pain is to look at how much people would be willing to pay to avoid it, which conveniently has been studied:

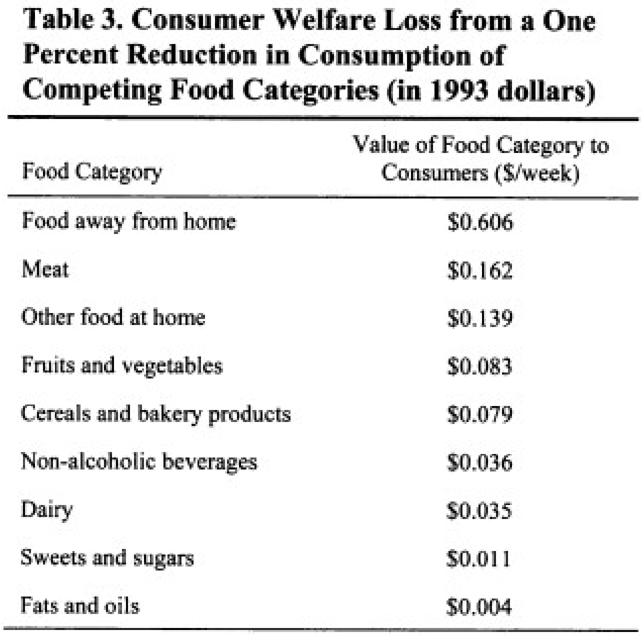


Table 3. Consumer Welfare Loss from a One Percent Reduction in Consumption of Competing Food Categories (in 1993 dollars)

Food Category	Value of Food Category to Consumers ($/week)
Food away from home	$0.606
Meat	$0.162
Other food at home	$0.139
Fruits and vegetables	$0.083
Cereals and bakery products	$0.079
Non-alcoholic beverages	$0.036
Dairy	$0.035
Sweets and sugars	$0.011
Fats and oils	$0.004

This table is the output of a point regression asking US consumers how much they would pay to avoid a one percent decrease in each category of food. A 1993 dollar is approximately half of a 2019 dollar (1.78), so consumers are saying that they would pay $15 per year to avoid a one percent reduction in their meat consumption. It is highly unlikely that people would accept one hundred times this value to cut out meat completely – it's easy to cut out the first few percentage points of meat (just have slightly smaller portions) but gets harder as you are forced to make fundamental changes to your diet.

For those inclined to stop eating meat, we wouldn't overthink this parameter. If the habituation literature has convinced you that you'd be just as happy in a wheelchair as a lottery winner then in the long-run you probably won't mind eating more tofu.

In the model, we double the $1,500/year preference loss of meat implied by marginal preferences to $3,000 to account for social costs and elasticity of demand, and assume this *decays to 0 by 10%/year as one gets used to the new lifestyle.*

7. Conclusion

Overall, the case for reduced meat consumption is strong. Vegetarianism is cheaper, better for your health (if you can afford a diverse diet and are not an infant), and is less impactful for the environment. It also has a significant moral cost in terms of animal suffering.

At the outset of the collaboration the vegetarian was sure that farmed animals' lives are so awful that the status quo is an unmitigated moral disaster; the meat eater was open to that conclusion but could also imagine being persuaded to spend all disposable income on buying meat and throwing it away because that was the only efficient means of causing the existence of sentient creatures who strongly prefer to exist. If you mapped reasonable conclusions on meat eating from a scale of -10 to 10, you could say that we started out as a confident -5 and highly uncertain 2, and ended up agreeing on a very confident -3.

Based on the research above, we've produced a 'base case' for the decision aid. It is weighted heavily towards the beliefs of the meat eater in the collaboration since the question revolves around what a 'typical' person might think and meat eaters are more 'typical' than vegetarians. We would certainly encourage you to tinker with the worksheet yourself though, as some decisions are very personal. You can download it here.

From the model we get that the total impact of meat eating per typical western consumer is roughly -$9,500 – that is, the 'society of conscious beings' would be better off by around $9,500 per year if any individual human meat eater switched to eating plants instead. To put this in other units, it would be about as good for 5 people to go vegetarian for a year as it would be for medicine to extend one person’s life by one year.

Each value in the table below represents the annual impact of a decision to eat meat versus eating an exclusively vegetarian diet. The right-most attempts to express everything in the same units ($) based on a willingness to pay for a year of perfect human life of $50,000 and a year of YOUR OWN life of $100,000 (to reflect the fact that people generally care about their own welfare more, but if you are a perfect utilitarian feel free to set these both to $50k!). Per the model we find that even though cows, sheep and (very weakly) pigs prefer farming to non-existence, the number of chickens eaten and the conditions they are farmed in dominates the ethical considerations. In terms of other harms, the impacts on your health and the environment are moderate, and the financial impact of switching to a vegetarian diet is small but negative – that is, the typical meat consumer will in the long run prefer to eat meat than spend the savings from vegetarianism elsewhere.

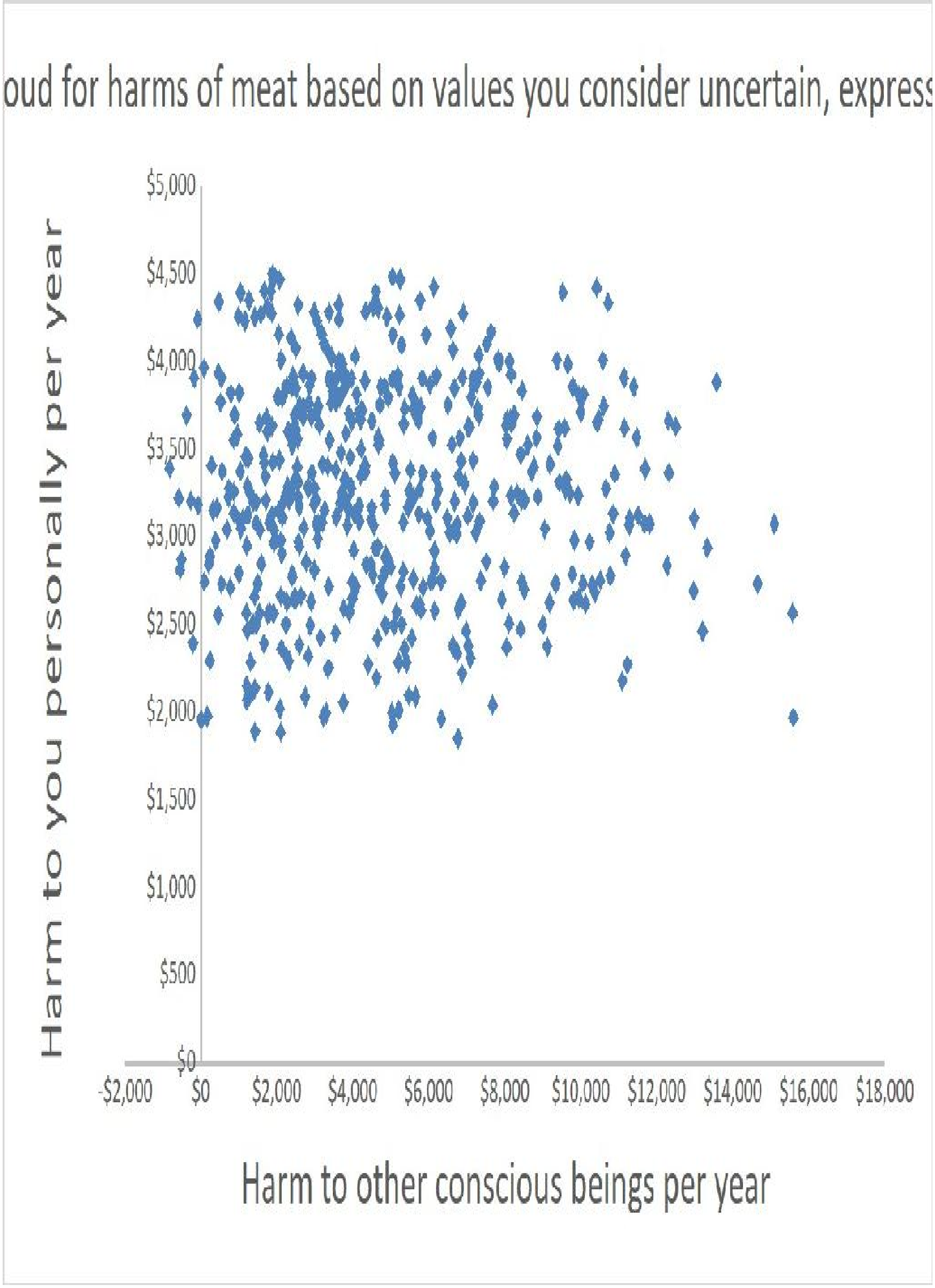
		Human life-year equivalents	**Expressed in $ equivalent**
Annual impact on other conscious creatures	**Cows**	0.009	$373
	Pigs	0.007	$317
	Sheep	0.002	$65
	Cage chicken	0.000	$0

	Shed chicken	-0.138	-$5,913
	Fish	0.000	$0
	Environ.	-0.027	-$1,166
Impact on you	**Health**	-0.039	-$3,336
	Finance		$162
TOTAL LIFE YEARS		**-0.186**	**-$9,498**

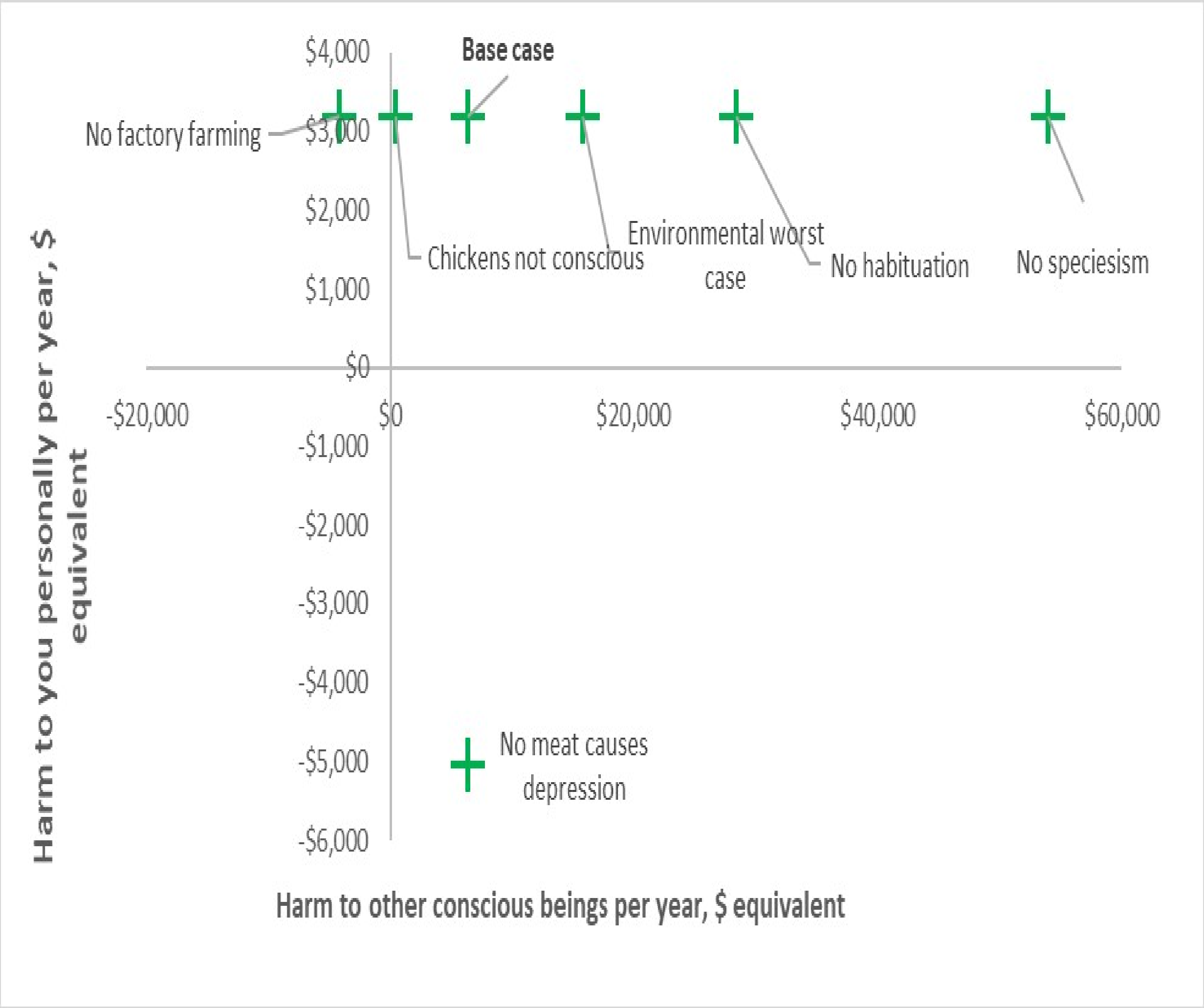
Overall, the impact of eating meat like a typical person is likely to be substantially negative. Eating no chicken limits the impact on the animals themselves, but the harm to your own health and the environment outweighs the moral good you do by causing the creation of animals who are happy at the margin.

7.1. Sensitivity

The decision aid allows you to specify uncertainty over any of your estimates, which we have done anywhere we are still uncertain about the value of a parameter. This analysis is displayed in the graph below. For any plausible distribution of inputs meat eating is harmful to you personally, primarily for health reasons; and meat eating generally causes harm to other conscious creatures because of the impact on environment and the high suffering of chickens. However, there is significant uncertainty about this value; in a small number of cases eating meat actually produces benefit to society by creating more lives animals would prefer to live on net.


oud for harms of meat based on values you consider uncertain, express
Harm to you personally per year
$5,000
$4,500
$4,000
$3,500
$3,000
$2,500
$2,000
$1,500
$1,000
$500
$0
-$2,000
$0
$2,000
$4,000
$6,000
$8,000
$10,000
$12,000
$14,000
$16,000
$18,000
Harm to other conscious beings per year

Another way of exploring model uncertainty is scenario analysis. We've calculated a number of scenarios that cover likely areas of disagreement.



In order of least to most harmful to other conscious being, the scenarios are:

- No factory farming narrowly results in outcomes which favor eating meat, since every animal would prefer to be alive than not. The effect is not greater because there are still environmental and health implications to eating meat.
- Chickens not conscious – The base case assigns a 75% chance that chickens are conscious, and this is a big assumption to which the model is highly sensitive. Assuming

chickens are not conscious results in outcomes which narrowly favor being a vegetarian, since the moral importance of creating worthwhile cow, pig and sheep lives is offset by the other harms of meat eating.

- Base case – As described in the document
- No meat causes depression – In this scenario not eating meat causes you a significant but not life-threatening illness – modelled by having minor depression (0.62 utility). This scenario is very interesting because it predicts that eating meat would be good FOR YOU, but would harm others, and therefore whether you should eat meat or not depends on the valuation you place on your own happiness versus the happiness of others (remember the model already values your QALYs higher than anyone else's based on the assumption you are not a perfect utility maximizer AND you are compensated for the unhappiness meat causes).
- Environmental worst case – the resources used to create meat are the upper end of the plausible range (10%) and all of this resource will create new people. The more convincing you find the environmental argument the more likely you should be vegetarian
- No habituation – In this scenario no creatures (including humans) habituate at all, meaning they are exposed to the full 'badness' of the farming conditions. The less you buy the habituation literature, the more likely you should be a vegetarian; this is a very strong result
- No speciesism – In this scenario the value of conscious experiences for animals is weighted just as much as conscious experience for a human. This is the scenario that results in the strongest argument against eating meat, and could perhaps be the intuition driving the sometimes acrimonious state of discussion between meat eaters and vegetarians.
- Not shown in this analysis is an 'Unfettered Vegetarian' analysis, where the vegetarian collaborator is able to enter their own assumptions into the model without any check

from the meat-eating collaboration partner. This is because what the vegetarian considers highly plausible assumptions (chickens are conscious, a much greater weight is placed on animal suffering/experience and much less habituation occurs) results in values that fall off the end of the graph – around $250,000 worth of harms to others per year.

Our key takeaway is that even under the most extreme scenarios we could think of meat eating is still very likely to be a net harm to both you and wider society. Also note that even in scenarios where you are not doing harm to both yourself and society, you are certainly hurting one of them quite a lot.

7.2. Impact on Collaborator Lifestyle

The meat eating collaborator was impressed by the environmental impact of beef and moral cost of factory farmed chicken. For the moment he has significantly reduced consumption of both, offsetting in part with salmon because fish have less environmental impact and are most likely not conscious.

The vegetarian was surprised how marginal the case for vegetarianism was when a ‘typical’ perspective was considered. Part of this is because he is still pretty skeptical that animals would actually habituate to the conditions we farm them in given the habituation literature doesn’t really cover conditions as cruel as factory farming. Another part might be that this collaboration has not focused on one-off traumatic events – especially slaughter – which probably don’t affect lifetime utility much but might be regarded as so self-evidently ‘evil’ that the way we have thought about the problem as a balance of good versus harm is incorrect. Having said that, although the actual harms of meat eating are less than he expected, the certainty of those harms occurring under any plausible distribution of beliefs (the fact that even a ‘typical’ person would probably regard meat eating as harmful considering everything) will probably make the vegetarian more militant about his vegetarianism. Sorry!

That being said, both collaborators agree that there is no substitute for evaluating the evidence for yourself. We can only hope that you find our analysis a useful reference.

[ACC] Does Calorie Restriction Slow Aging?

[This is an entry to the 2019 Adversarial Collaboration Contest by the delightfully-pseudonymous Adrian Liberman and Calvin Reese.]

About the Authors: Adrian Liberman is currently a PhD student in biology at a university in the mid-Atlantic. He previously worked at the National Institute of Aging and remains actively interested in gerontology and the biological study of aging. Calvin Reese is an author with a BS in Biology. He has always been interested in the possibility of life extension by calorie restriction. Recently, he has reexamined the subject after undertaking a series of intermittent fasts for weight loss reasons. Calvin believes CR extends life; Adrian has long been skeptical.

Introduction: Is food making us old?

We all agree that food is delicious, and we also all agree that too much food is bad for us, but exactly how bad is it? Various academics have proposed that too much food actually accelerates the aging process, and reducing our food intake via calorie restriction (CR) is one of the most accessible and available methods of extending human life. While billionaires pump vast fortunes into increasingly far-fetched stem cell treatments and consciousness transfers, CR advocates contend that they can get a 10-20% increase in their natural lifespans simply by eating a little less. If true, CR raises a question of enormous significance to gerontology and the science of aging: are our diets aging us one calorie at a time? And if so, can we stop it?

Calorie Restriction (CR) and Intermittent Fasting (IF) advocates generally claim that CR will extend your lifespan and prevent various diseases, and that IF is an effective form of CR. Evidencing these claims are animal studies in yeast, worms, flies, mice, and monkeys as well as indirect evidence from humans. A variety of biochemical studies have been performed, and a host of theoretical literature generally claims that the underlying mechanisms for the effect are the IGF axis affecting sugar metabolism, DNA damage mechanisms dealing with free radical formation, and inflammation modulation.

My position (Adrian) is that for the average individual reading this article, CR and IF are not generally worth the effort, because the individual will be exposed to non-trivial risks and the benefits will be minute.

My position (Calvin), is that there exists some amount of food, on average, that will produce an optimal human lifespan and that the average person could significantly extend their life by moderately reducing their calorie intake.

Calorie Restriction (CR) is a term coined to describe a series of experiments that have been conducted over the course of more than a century, demonstrating that various animals kept in laboratory conditions generally survive longer if fed diets that are ‘restricted’. This effect has been observed in bacteria, yeast, worms, fruit flies, mice, and arguably monkeys, so it appears to span every domain of life. (40)

On the other hand, population-level studies say that the lowest observed mortality in western populations occurs at a BMI of ~25. (4)(5) This appears to be a paradoxical result, since a BMI of 25 generally results from a diet that isn’t particularly “restricted”. Why would we observe a lower mortality in lab animals when they are undergoing perpetual mild starvation, but a higher mortality when this happens in humans?

And by extension, should you, the reader, adopt a calorie-restricted diet, or an NIH-approved 25 BMI diet?

Semantics: Calorie Restriction versus just not overeating

First things first. Let's define some very strict terminology. This will become important later:

Kinds of diets:

- Ad libitum Diet (ad-lib) – at one's pleasure, or as much food as you would like.

- Ad-lib Calorie Restriction (Ad-lib CR) – a diet that contains fewer (for example 30% fewer) calories than your diet would if you were eating an ad-libitum diet

- Normative Diet (ND) – a diet that is balanced and prevents the onset of obesity. The general analog to USDA's 2000 calorie diet. (Our term and not in common usage)

- Normative Diet Calorie Restriction (nd CR) – a diet that is balanced, but contains fewer calories than a normative diet (for example 30% fewer)

Once we establish these definitions, the kinds of claims that proponents or opponents can make expand into the following:

Claim 0: You, the reader, should adopt an ad-lib diet
(Nobody claims this, put the chips down!) (null hypothesis)
Claim 1: You, the reader, should adopt an ad-lib CR diet
Claim 2: You, the reader, should adopt a Normative Diet
Claim 3: You, the reader, should adopt an nd-CR diet

For the purposes of this article, we'll be assuming that you, the reader, are an average American of indeterminate sex.

First the claims of CR in detail:

Animal Studies and the NIA Interventions

CR relative to ad-lib leads to improved metabolic function in the animals studied, generally because they do not become obese. On the other hand CR relative to a normative diet leads to unfavorable/negative outcomes including a decrease in fertility, altered mental states, and muscle wasting, but also a significantly increased lifespan.

A Calorie Restriction experiment goes like this: Take some animals and first establish the amount of nutrients they consume “normally”, or under optimal growth conditions. After this, take half of those animals and provide them access to all the micronutrients and amino acids they want, but restrict their access to raw energy in the form of fats or carbohydrates to X% of their normal intake.

Where does this lead to increased lifespans? Let’s start at the beginning! The beginning of time:

CR dramatically extends the lifespan of yeast. (39) Yeast longevity is difficult to quantify, but experiments have suggested 75% CR in yeast (done by decreasing glucose concentrations in yeast media from 2% to 0.5%) extends yeast longevity by a factor of 3. (39)(40). Single-celled organism lifespan is a goofy term, but it can be measured both directly by looking under a microscope at cells and keeping track of when they kick it, and indirectly by comparing the steady population of cells to how often they divide.

Nematodes (worms) exposed to 50-75% CR experience a 2-3 fold increase in longevity. CR in Drosophila melanogaster produces between 30% and 100% increases in the observed lifespan of the flies. (40)(41) Many hypotheses have been proposed to explain why CR produces such marked increases in the lifespan of worms, flies, and yeast, but most notable is the behavior of the SIRT-family genes. (71) Unfortunately the SIRT mutation data is difficult to interpret due to the relatively central role SIRT genes tend to play to the functioning of a diverse array of cells, but some theories are more convincing than others. More on this later. Of course, in general, the problem with studies like these is that establishing what constitutes a normative diet in organisms like these is somewhat subjective. Is there such a thing as an obese yeast cell? On the other hand, a normative diet for rodents is pretty easy to establish.

Numerous studies and overwhelming evidence show that CR significantly extends the lifespan of rats. Rodent CR studies suggesting CR extends lifespan go back to the 1930s and 1940s (26) (30). The earliest rodent CR studies focused on longevity with modern approaches and methods were conducted in the 1980s. Pugh et. al in 1992 (27), Yu et. al in 1985 (28), and Weindruch and Walford in 1982 (29), all subjected rats to 40% CR diets compared to ad lib baselines and found between 10 and 20% increases in longevity as compared to normative diets. Similarly, 1986 Weindruch is a gold standard mouse CR study. Here we see that even fairly aggressive ndCR produces extended average and maximum lifespans, both relative to ad-lib and relative to a normative diet, with a normative diet extending lifespan 20% relative to ad-lib, and 40% ndCR extending lifespan another 30% on top of that. (Context: think big fat guy keeling over from a heart attack at 50, vs everyone’s stereotypical tiny Chinese grandma, who spends her 100th birthday stubbornly refusing to reveal the location of her phylactery)

Reported health benefits of CR in rodents include reduced cancer risk in p53-deficient mice (33), increased proteasome activity (34) in mice and rats, improved cognition (32)(35), reduced oxidative stress and NF-kB signaling (36), and various other health benefits. Several studies, including Park et. al. and Pires et. al. have observed dramatic changes in insulin signaling and serum glucose levels (32)(37)(38), which has been of particular interest to gerontology researchers. IGF signaling has been proposed as one of the mechanisms by which CR improves health outcomes (15).

So far this is a pretty strong story. What's the sketch factor of this evidence? Well… Some rat and mouse strains responded to CR better than others and methodology varied widely between rodent trials, making them difficult to compare (32). The only study performed on wild-caught mice that aren't buried under a mountain of genetic defects to shame the Habsburgs had a negative outcome. As a model for aging, mice are also slightly suspect because unlike most mammals, they are globally telomerase-positive, which means that the effects of mutational accumulation on their soma don't directly translate to other mammals, since oxidative damage has a large interplay with telomeric senescence.

Overall, however, we agree that mouse lifespans are significantly extended by CR. But…

As the organism becomes larger and more complex, the beneficial effects of CR on lifespan appear to taper off (40). 25% nd CR extends dog lifespan 25% (66). Why isn't more aggressive CR investigated in dogs? Larger animals don't tolerate aggressive caloric restriction well, 40% caloric restriction would likely kill most dogs. It seems that brains are probably to blame, because the metabolic rate of the brain is generally not significantly regulated (70). If you run out of energy for running your brain cells, you're done for. The larger the proportion of your metabolism dedicated to maintaining brain function, the less CR you can tolerate.

So, given that trend, the most relevant information for humans should probably come from monkeys. (#freescopes)

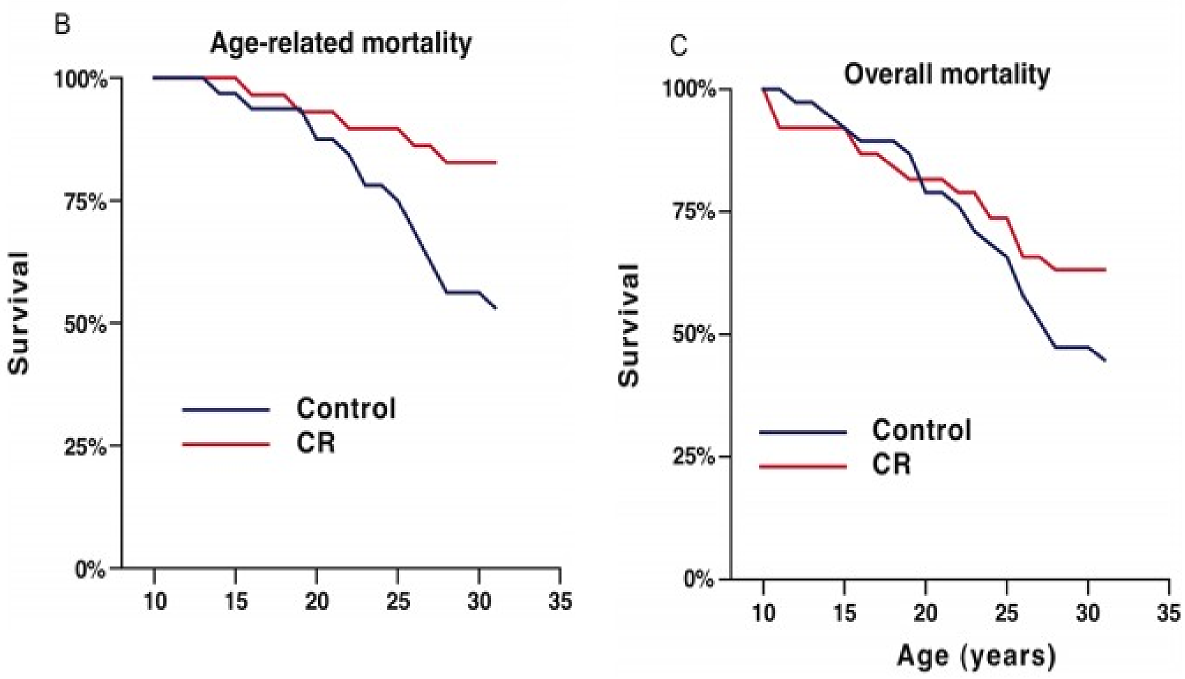
Unfortunately, monkeys live a long time, so studies of monkey aging are measured in many decades and cost millions and millions of dollars. The National Institute on Aging and the University of Wisconsin have taken on the task and subjected ~40 rhesus monkeys each to 30% CR. In the WNPRC study, CR was versus an ad lib baseline and in the NIA study CR was relative to a standardized diet designed to prevent obesity. (1) (3)

Rhesus monkeys have a maximum lifespan of ~40 years, and after 30, the WNPRC study reported only 13% of the CR group had died of age-related causes, whereas 37% of the control group had died. (1) The authors write “CR reduced the incidence of diabetes, cancer, cardiovascular disease, and brain atrophy.” (1) In 2017, 10 of the WNPRC CR group animals remained alive compared to 3 animals in the control group. (31) In a monument to absurd timelines, the WNPRC study is not yet over, as many CR monkeys survive in 2019. No mean lifespans are established.

Here we hit our first snag. In complete contrast to the WNPRC study, the NIA rhesus monkey study used a standardized diet designed to prevent obesity for the control group. The NIA study found no increase in survivability among the CR group. (3) What gives?

If we were being charitable, we would say that monkey studies are ridiculously hard to power properly. With only 86 monkeys present in the NIA study, the negative observation could easily have been a product of chance that was inadequately represented by the reported p-values, and we should defer to priors based on other mammals. Anecdotally, and though the NIA paper would never admit to this, the NIH sometimes staffs monkey studies of this kind with “leftover” monkeys from clinical trials that may have mysterious medical conditions that are not obvious to the naked eye but were caused by drug trials, experimental surgical procedures, etc.

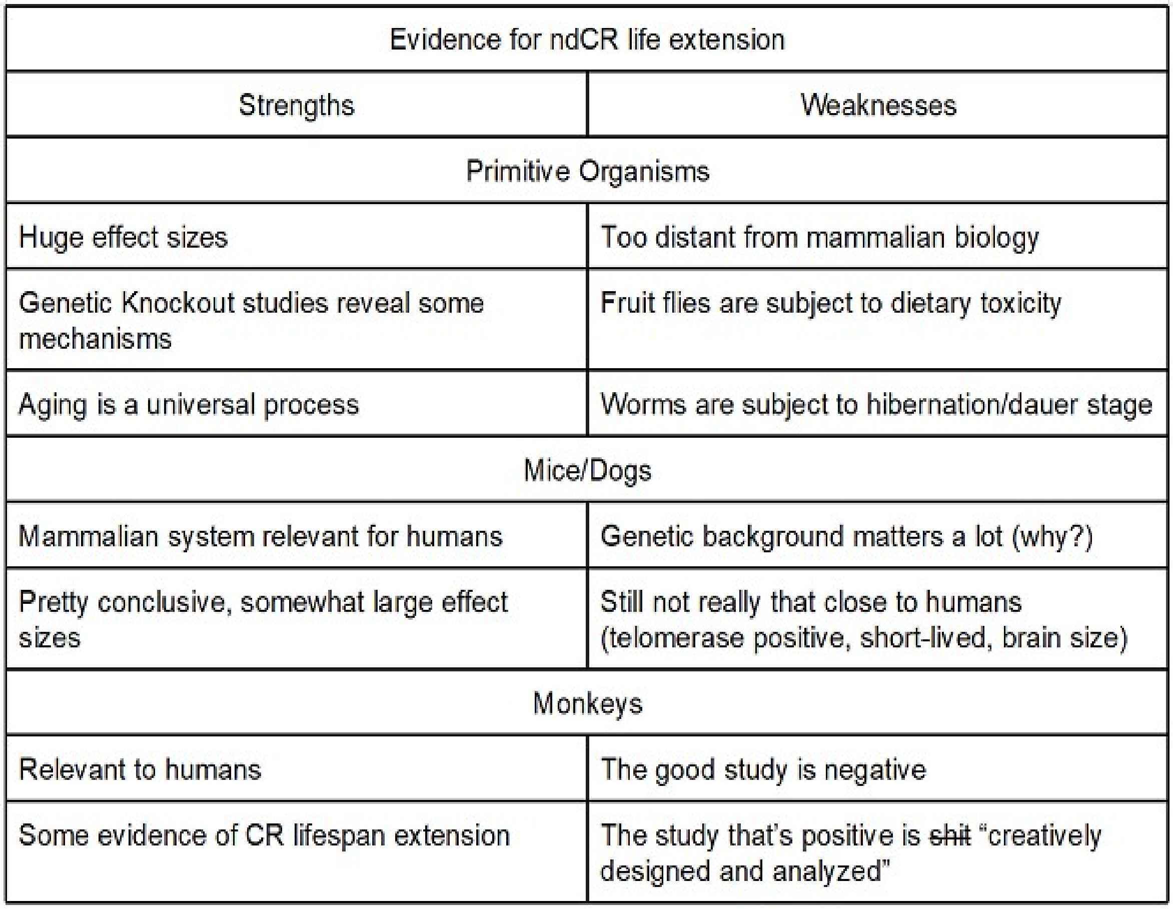
If we wanted to be less charitable, we would look at the fact that the WNPRC study pulled shit like this:



…and point out that aging is almost by definition a process that impacts the survival rate relative to any injury, so trying to disentangle “age related” mortality from regular mortality is bad and wrong.

The most significant difference between these studies is the use of adCR (Wisconsin) vs ndCR (NIA). It’s no surprise that monkeys that are on a controlled diet are healthier and live longer than monkeys that are fed an ad-libitum diet. In mice, this is a commonly known problem for control populations, and monkeys are no different (72). The balance of evidence probably tilts against the idea that CR is effective in monkeys, however our priors that ndCR should be effective are fairly strong, so for now let’s assume that it is.

Adrian concedes that a 2017 statistical analysis of both studies by the University of Alabama at Birmingham in cooperation with the authors of the original studies determined that CR decreased mortality in rhesus monkeys (31), but the study is presented under protest because combining ndCR and adCR in the same analysis is inadvisable, disingenuous, probably illegal, and was the direct cause of the sinking of the Lusitania.



Evidence for ndCR life extension	
Strengths	Weaknesses
Primitive Organisms	
Huge effect sizes	Too distant from mammalian biology
Genetic Knockout studies reveal some mechanisms	Fruit flies are subject to dietary toxicity
Aging is a universal process	Worms are subject to hibernation/dauer stage
Mice/Dogs	
Mammalian system relevant for humans	Genetic background matters a lot (why?)
Pretty conclusive, somewhat large effect sizes	Still not really that close to humans (telomerase positive, short-lived, brain size)
Monkeys	
Relevant to humans	The good study is negative
Some evidence of CR lifespan extension	The study that’s positive is ~~shit~~ “creatively designed and analyzed”

The gold standard of CR studies in animals with respect to human health would be studies that occur in higher animals (mice, dogs, monkeys) and perform CR relative to a normative diet. On the balance, evidence that ndCR extends lifespan in higher animals is fairly strong. The monkey problem, however, is pretty bad, and we remain skeptical of the strength of CR in humans.

Despite these objections, we agree that CR promotes longevity and reduces all-cause mortality of animal model organisms and that this finding supports the view that CR probably increases the lifespans of humans in an ideal scenario.

So… what's up with that? Do the above findings mean that we need to radically reinterpret what being overweight or obese means? Are we just WILDLY overestimating the "healthy" weight for all these animals?

In a word, no. Animals are clearly not adapted to undergoing caloric restriction this severe. How do we know that? Easy. The animals in most of the experiments above, under fasting conditions, are driven sterile. From an evolutionary standpoint, it's safe to assume that being sterile is not an adaptive trait, so clearly most animals have not evolved to operate at levels of caloric restriction this severe on a routine basis. That's one of the simpler distinctions between a "healthy weight", obesity, and being in a starvation regime. Still, clearly animals live longer when they are starving, so why?

Proposed Mechanism by Which Food Makes You Old (and how calorie restriction stops it)

First an interlude: aging is one of the last frontiers in biology where major theories still compete on an even footing to explain a basic and universal process. The simplest definition of aging is the observation that past a certain point in an animal's life, the likelihood that it will die doubles every period X, where X is different for different species. This observation is stunningly universal. It’s also important to note that the likelihood of death from almost any type of injury increases over time, so aging is not just the idea that cancer is more frequent when you’re 50 than when you’re 20, but the idea that almost all diseases are more frequent in older individuals, and dying from almost any injury is more likely when you are old than young (78, Arkin).

Covering the slapfight over the specific mechanism involved fully is beyond the scope of this paper, especially because there are no conclusive answers to any of your questions, but briefly these are the major aspects of biology that change with age, drawing directly from “The Hallmarks of Aging” Lopez-Otin et al. (44).

1) genomic instability, such as replication errors, mutations, DS breaks, and crosslinking;
2) telomere attrition, including damage to the telomeres that does not result in telomere shortening;
3) epigenetic alterations, including changes in methylation patterns, histone modifications, and chromatin remodeling. This generally also leads to deregulated/erroneous gene expression;
4) loss of proteostasis, which is characterized by protein denaturation or unfolding and the accumulation of waste products your body cannot break down;
5) deregulated nutrient sensing, causing both the cell and the body to become less responsive to nutrients;
6) mitochondrial dysfunction, which is thought to be particularly central to the relationship between CR, obesity, and aging because mitochondria process glucose and create reactive oxidative species (ROS);
7) senescense or quiescence, the cessation of the cell cycle; and
8) stem cell exhaustion, which slows or halts renewal of virtually all tissues and cell types.

These are specific, measurable instances of the general breakdown we associate with aging chosen for their correlation with chronological and apparent age. None of them has been decisively established as the actual cause of aging as we understand it, though all are understood to contribute to aging at the cellular and tissue level, which is the general breakdown and cessation of cell functionality (44).

Gerontologists love to whip these hallmarks out, but in reality most of them are interrelated in some way, so establishing which ones of them are merely the byproducts of other is extremely difficult. Ex: mitochondrial dysfunction increases the rate at which oxygen radicals are produced leading to greater genome instability and telomere attrition. Stem cell exhaustion probably arises from genomic instability or nutrient sensing deregulation, but on the other hand it can lead to senescence in the tissues as replacement of dying cells slows down. It’s all an ouroborosian mess. (44)

Almost all of the hallmarks have been shown to be impacted by food intake. Obesity has been decisively implicated in causing genomic instability in model animals, but evidence is lacking in humans (45). Crucially, the CR-ROS hypothesis that excessive eating, causing both oxidative damage from digestion and metabolism of glucose and obesity – thus linking obesity and ROS damage together, remains poorly supported in humans (45)(46). Obese adult individuals suffer from greater telomere attrition than non-obese individuals and have shorter telomeres (47). Obese individuals have profoundly different epigenetics than non-obese individuals, which is not surprising, since insulin expression profiles and fat metabolism are well-established as being changed by, and changing, epigenetics (46)(48)(49). In fact, epigenetic changes in insulin expression due to a high-fat diet may be heritable (48). Obese individuals, and particularly diabetes patients, have markedly different DNA methylation patterns from non-obese persons and modified chromatin structure (49) Altered chromatin density due to obesity has been observed in rodents and implicated in the onset of dementia, which is commonly associated with old age (50)(51). Numerous proteasome dysfunctions and significant protein misfolding have been observed in obese rodent and human subjects (46). In addition to creating insulin resistance, human adipose tissue creates a pro-inflammatory environment and significantly alters cell response to NF-kB and inflammatory cytokines (52). Obesity has been specifically implicated in ROS damage to mitochondria, mitochondrial dysfunction, reduced mitochondrial fission, and further elevated ROS production by the Krebs cycle (53). Furthermore, adipose tissue and obesity apparently promote senescence (54). Obese mice exhibit dramatically increased rates of T-cell senescence (55). Finally, obesity results in far greater stem cell exhaustion and quiescence (56). Adipose tissue stem cells, hematopoetic stem cells (HSCs), bone marrow, and other stem cell reserviors have all demonstrated higher rates of quiescence in obese populations of rodents and humans (46).

So, we have covered all eight biomolecular cell and tissue-level hallmarks. Obesity does, in fact, make us older. And we are pretty sure CR makes most animals younger. That is an excellent clue as to how the two are related, and which of the mechanisms might be the most important.

Your body usually converts sugar into mechanical or chemical energy by using a complicated daisychain of a couple dozen proteins called the Electron Transport Chain (ETC). The ETC is so called because each protein in it contains an electrically charged amino acid that is highly reactive due to holding on to loose electrons that came from sugar. The ETC takes electrons obtained from sugar, and attempts to stick them onto molecules of oxygen, converting it into water. Unfortunately, this process occasionally fucks up, and instead of getting nice benign water, the oxygen becomes radicalized, gets an Al-Qaeda franchise, and becomes either Hydrogen Peroxide or the Superoxide anion. Both of these are comically reactive molecules, and if they happen to diffuse out of the mitochondrion, they can react with pretty much anything and cause damage. If they react with your DNA, you're in trouble, because oxidative damage to your DNA produces mutations. Similarly, the oxidation of fats results in chemicals that cause inflammation, and the oxidation of proteins can result in byproducts that your body can't break down.

From here comes the Rate of Living Theory: an animal can consume only a specific and finite number of calories in its lifetime, because after that point the damage induced by said calories becomes fatal.

Essentially, for a given cellular architecture, so many calories in results in so many molecules of peroxide and superoxide out, and so many mutation events in DNA, oxidized proteins, and rancid fats. Inside you there are populations of cells that can tolerate only a finite amount of oxidation events, namely heart muscle fiber nuclei, neurons, and long-term stem cells like marrow and fibroblast stem cells. For these tissues, the DNA you have is the DNA you get, and you can only expose it to so much chemical damage before it stops functioning. Similarly the accumulation of other byproducts would be fatal.

While there is some slight debate about this, the balance of evidence is that CR slows down the metabolism of most tissues relative to their quantity (74,75), so it should definitely allow you to stretch your “mutation budget” over a longer timeframe. This theory predicts that you can extend your remaining life by X% by cutting X% of calories out of your daily food intake. Napkin math tells us that this basically tracks with the results we saw above, and the limitation on this is obviously starving to death.

Hopefully at this point we have exhaustively convinced you that Calorie Restriction definitely works in small animals, probably due to inhibiting oxidant production, and it might work in humans, though the theoretical backing for this is dicier.

So we return to our claims. Should you eat an ndCR diet?

First we have to take care of Chesterton’s Fence.

Population Aging Studies and Human Trials

Why is it commonly recommended to consume ~2000 calories per day? And where does this number come from anyway?

Dietary guidelines approach the question of how much food you can eat epidemiologically, by asking what kinds of mortalities are associated with various food intakes.

A meta-analysis of large-scale population studies concludes that a BMI of roughly 24, considered to be “normal” weight, is associated with the minimum long-term all-cause mortality. (5) (4), and once BMI exceeds 30, all-cause mortality begins to sharply increase. (5) (4) A CDC estimate for the 2015-2016 period found that 39.8% of adults were obese, and it is not unreasonable to assume that the average American BMI is roughly 30 based on huge body of evidence. The interesting aspect of these studies lies in the fact that BMIs below 25 are likewise associated with increased mortality (5)(4). Direct observational data of western populations is nearly unanimous in showing that BMIs below 25 are correlated with bad health outcomes.

So here we meet the fundamental paradox: Why is it that we can observe lab animals living longer in the face of CR, but when we observe humans, we generally find an association between mortality and low body weight?

Epidemiologic studies on low BMIs in western populations become a little hazier. There is fairly wide agreement that BMIs below 25 correlate with lower survival, but not usually a clear claim as to why. BMIs below 25 generally appear to correlate with smoking. Excluding the effects of smoking, people with BMIs below 25 still appear to have increased all-cause mortality. This could be from correlations to weight loss from cancer and metabolic diseases, tuberculosis, or something similar. While the effects of obesity on risk of mortality are very clear, the effects of being underweight and starvation on health are obviously the source of our core paradox (4).

Even studies that attempt to exclude acute illnesses that commonly induce weightloss seem to find higher rates of death among men that are underweight across a wide range of causes. Most troubling is the fact that an under-consumption of calories has a documented effect of suppressing wound healing and immune response, and worsening the progression of infectious diseases (67-69). From this standpoint, it’s altogether unclear whether the low weights observed in these studies are induced by the diseases in question, or whether diseases are simply more prevalent in people with BMIs below 25.

Calvin’s hypothesis, though unsupported by evidence, is that no one is actually on a CR diet.
A CR diet is reduced calorie intake without malnutrition. Some of the people with BMIs below 25 may have an

underlying health issue, such as HIV, which is causing their BMI to drop or mortality to rise, but I speculate that most people with a BMI below 25 are suffering from malnutrition, which is causing the increased mortality observed in persons with BMI below 24-25. Sarcopenia in the elderly is another possibility, though age-adjusted studies appear to refute this hypothesis (4).

We can attempt to look at other cultures to see if lower caloric consumption has similar effects there.

Gerontologists and anthropologists have observed that the longest-lived national and ethnic groupings of humans tend to eat the least. Japan has long had the highest life expectancy of any developed country (20). Japanese life expectancy at birth stood at 87.17 years in 2016, as compared to 81.40 years for the United States and 84.43 being the average of 18 high-income countries (20). Japanese are estimated to eat 23% less than Americans (21).

However, this finding is purely correlational in nature. Although reverse causation – that people eat less because they live longer – can be ruled out, the correlation could be coincidental, pleiotropic, or for genetic rather than dietary reasons. The FOXO3A gene, rather than CR, has been proposed as the reason for variations in longevity between ethnic populations (17). Several other genes, like APOE and CETP, have been suggested as alternative genetic causes of these ethnic longevity differences (18). Gerontologists have aggressively suggested that smaller humans tend to live longer (4) (13) (14), and that members of ethnic or national groups, such as the Japanese, who live longer tend to be physically smaller, and centenarians within this ethnic group tend to be smaller than those who live shorter lives (13). One proposed reason that smaller individuals and ethnic populations tend to live longer is lower levels of GH and IGF-

1 due to genetic factors (15). As has been discussed, IGF-1 is associated with increased mortality due to various illnesses and also makes the individual physically larger. CR advocates have equally argued that social, environmental, and economic factors causing CR in these population groups cause the drop in GH and IGF-1, increasing longevity in these groups. A Washington University study of the effects of CR on a group of humans on a long-term diet of 1800 kcal per day versus an experimental group ingesting 2500 kcal per day found no decrease in serum IGF levels from baseline unless protein intake was also restricted (16). This result is at odds with rodent studies, which showed decreases in IGF-1 concentrations in CR subjects versus those on normative diets (16). Other results contradict these findings. A Tufts 30% CR trial among mildly overweight (BMI 25-29.9) young adults showed significant decreases in serum insulin concentrations, contradicting the Washington University results (25). Many, many other human trials have produced contradictory results in CR trials, with the severity of the CR, weight, and age of the patients, as well as compliance with CR, having been variously proposed as explanations (21).

There is also compelling evidence in favor of CR and dietary explanations as the cause of longevity in particular populations. For example, the ethnically distinct Okinawans were for generations the Japanese ethnic group with the greatest longevity; the island has 4-5 times the centenarians per capita of any industrialized country (21). Little racial admixture has occurred on Okinawa, but as the Okinawan diet has westernized, the Okinawans have lost their longevity advantage, with Okinawan longevity dropping below the rest of Japan in 2005 (19). Older Okinawans, who continue to eat CR and protein-restricted traditional diets, have greater longevity compared to Japanese populations of equal age. For Okinawans age 60-64, all-cause mortality was half that of Japanese persons of equal age (22). Evidence from natural “experiments” during mandatory rationing and food shortages also supports the CR hypothesis. Involuntary food rationing during the World Wars also paradoxically increased lifespan. WWI-era rationing in Denmark resulting in a 34% drop in mortality over a two year period (21)(23). Similar rationing in Oslo during WWII, thought to be equivalent to 20% CR, resulted in a 30% drop in mortality (24).

The longest lived humans tend to eat the least. Though causation – whether CR causes decreased IGF levels or decreased IGF levels cause CR – cannot be definitively established, the conclusion, supported by rodent studies and defensible from surveys of humans, that CR causes a drop in IGF levels remains compelling.

So, should you, mean American Reader with 1.5 X chromosomes, adopt a CR diet? WHAT COULD GO WRONG?

Acute Risk Factors of Caloric Restriction

A good first place to look is the Minnesota Starvation Experiment. Conducted around 1944 on 32 volunteer males, the object of the experiment was to subject humans to a 25% calorie-restricted diet that simulates 6 months of famine and observe the effects. Diet was strictly controlled by housing subjects in a special dormitory and supervising them during their time outside of it. Despite the simulated famine conditions, diets were formulated to ensure that subjects received daily minimum intakes of important vitamins and minerals

An exhaustive description of the effects can be found in The Biology of Human Starvation, by Keys, but let’s go through the highlights. (70, not available online but try to find it at a library. Equal parts fascinatingly and grim, but occasionally also very funny)

First the physical: subjects of the experiment underwent a substantial drop in bodyweight (duh) over the first few weeks, after which their weight stabilized at 75% of their original bodyweight, as was the design of the experiment. In terms of physical strength and work capacity, subjects experienced loss of motor coordination, loss of strength, and a devastating negative impact on their endurance. Counts of all blood cells were down across the board per unit volume of blood, and in absolute terms. Subjects had a lowered basal metabolic rate, and an observed drop in surface but not core body temperature. Severe caloric restriction (15%+) also leads to a sharp fall in fertility for both men and women. Women undergoing caloric restriction may stop menstruating. (not observed in the Minnesota study, which was all-male) Finally, a starvation state can lead to bradycardia, starvation edema, fainting, and looking like a cartoon skeleton (70)(77).

Psychologically, the first and most significant observed effect was a severe preoccupation with food, and a permanent feeling of being cold. The latter was both a true sensory perception (skin temperatures of subjects were lower than normal) and a subjective sensation. Additionally, subjects experienced a loss of sexual interest. The subjects were capable of achieving erections physiologically, but not psychologically. Subjects also felt a loss of motivation to engage in self-improvement or social activities. By month 6 of semi-starvation, more than half of all subjects were routinely failing to complete basic maintenance tasks such as cleaning, and most had dropped out of university classes they were initially attending. After subjects were released from the experiment, several developed eating disorders, and fasting in general can induce eating disorders where none were otherwise present, both binging and anorexia.

Advocates of intermittent fasting, which is a diet that is much easier to maintain, often say that you eventually adapt to the feeling of acute, distracting hunger that strikes you when you are on a short (eg 1 or 2 day) fast (Original research). No such adaptation was observed for the subjects of the Minnesota experiment, and if anything morale of the subjects deteriorated steadily throughout all 6 months of the experiment. In principle, we need not assume that you cannot psychologically adapt to permanent starvation conditions, however mice that undergo lifetime caloric restriction do show a permanently depressed level of motility, eg they just hate moving around.

Evidence from other sources also indicates that lowered caloric intake leads to worsened progression of infectious diseases and slowed wound healing. This point may initially appear controversial because some studies indicate faster and greater response to mitogens, but studies using live pathogens prove this point fairly conclusively. Crucially, this aspect of starvation biology is one that would not be revealed in CR Mouse studies. Lab mice are generally kept in aluminum shoeboxes in a strictly sterile environment most of their lives. They have little opportunity to suffer injuries or heal from the same, and rarely experience infections. If they do, they are quite likely to die and be excluded from analysis. Unless you happen to live in a sterile aluminum shoebox as well, consider this as you interpret CR studies (67-69).

How does this square with the smaller diets in East Asia? Okinawans reportedly ate a diet that was at a similar level of caloric restriction (20% relative to western ND) to the Minnesota experiment, however obviously the entire island wasn’t driven sterile or languid. Note however, that this difference is reported only in absolute terms, not as calories per unit of body weight. Factor 1 is probably stature. Long-term caloric shortages have been shown in both mice and humans to lead to shorter stature and smaller size. If you were reared from youth to eat a relatively smaller diet, your long-term caloric requirements could probably be lowered somewhat. This is borne out in the average heights and weights of Okinawans of this time period, often reported to be less than 5 feet tall. This also squares with studies on longevity between different mouse strains, which routinely report that smaller overall stature (or length) has a positive correlation to longevity.

Factor 2 may be local weather conditions (Okinawa is sub-tropical) Temperatures have an impact on the ability to tolerate a restricted diet long-term. One notable result observed in investigations of Okinawans is the higher thermogenesis and lower oxygen consumption in Japanese and Okinawan mitochondrial haplogroups, suggesting both lower generation of ROS, and likely an ability to tolerate lower caloric consumption while maintaining adequate body temperature (76), which is an important aspect of determining basal metabolic rate. The BMR of Japanese people and Okinawans can be considered as lower or higher depending on whether it is measured through thermogenesis or oxygen consumption, which measure subtly different things.

This discrepancy suggests to us that even in the event that you are able to maintain a traditional Okinawan diet, if you were reared in America, it’s quite possible that you would have the ol’ Minnesota Boner Downer experience attempting to do so.

The outcomes of a lesser caloric restriction would be easier to tolerate for the average westerner, but will also be less effective. Whether there exists a sweet spot in which you are calorically restricted, but don’t hate life and can lift a broom is probably a subjective judgement.

Overall Conclusions

None of the evidence in favor of CR is indisputable. CR-ROS, the hypothesis that calorie restriction reduces oxidative radicals, remains compelling, but direct evidence in humans is lacking. Numerous model animal studies have shown a link between CR, reduced mortality, and life extension. Population studies support the CR hypothesis, but the effects of CR cannot be easily disentangled from genetic, social, environmental, or non-CR dietary factors. CR experiments in humans and rhesus monkeys produced contradictory results, and in some cases the tradeoffs between early and late mortality are a judgment call. Progerias in the obese and biomolecular evidence of cellular and tissue-level anti-aging effects of CR remain the strongest evidence for CR’s potential to extend human life.

We, the authors, conclude that the evidence as it stands weakly supports the conclusion that CR modestly extends human life. We expect that an individual engaging in 20-30% CR versus a normative, non-obesogenic diet without malnutrition might enjoy a 10%-20% increase in longevity. A 10%-15% CR relative to a normative diet may increase lifespan by perhaps 5-10%.

As with all good science, this conclusion raises still more questions. 20-30% CR might result in a 10% increase in longevity, but is that worth it? Calvin, one of the authors, is a practicing intermittent faster and can testify that CR and IF are unpleasant, difficult, and sometimes painful.

Scientific investigation adds another layer to this subjective answer: starvation conditions are likely to expose you to infections of greater severity, potential sterility, negative impacts on your physical abilities, and subjective but significant impacts on your psychological state, including motivation, attention, and libido.

The field of gerontology and the general study of aging continues to lurch forward – not at the pace we want it to, necessarily, but it’s still developing anyway. New drugs and treatments, including stem cell activators like GDF11, senolytic drugs, and anti-inflammatory interventions may be able to make many of the benefits of CR redundant in the relatively near future (we hope).

While CR would probably extend your life, we, the authors, don’t advocate it. The risks and miseries aren’t trivial and you probably have to go to work in order to exchange money for goods and services.

Claim 2: You, the reader, should adopt a Normative Diet
True (and basically the same as an adCR diet)
Claim 3: You, the reader, should adopt an nd-CR diet
Debatable, but no.

If you are interested in the best current options for life extension, you should consider a long-term aspirin regimen, maintaining a healthy body weight, and building a nuclear bunker in your backyard.

All the best,
–Calvin Reese, Adrian Liberman

PS(A): PLEASE NOTE: If you are over 75 years old, do not attempt Calorie Restriction. If your grandma is over 75 years old, go to her house and pour soup into her until she is overweight. This is not a joke and is entirely serious advice. Among the very elderly, being overweight serves as a protective factor that mitigates the dangers of death due to traumatic injury. The dangers of heart attacks and diabetes associated with excess weight are less than the dangers associated with sarcopenia and cachexia. Elderly people usually experience a loss of strength in esophageal muscles and for them swallowing food becomes more difficult, leading to a vicious cycle of muscle weakness and weight loss. If you have an elderly relative, please make sure they’re eating enough.

PPS: Studies excluded from this review: CALERIE: conformity to study protocol was terrible, duration too short, and they took overweight people and got them to baseline. Total tripe on a bike. CRONies: lmfao. Be skeptical of studies claiming to observe DR in a human population. Talking a large group of people into actually following a strict DR regimen long-term is borderline impossible because it fucking sucks. Sample was also self-selected.

Bibliography

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2812811/ – Wisconsin Monkeys
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC345016/ -84 adiposity
3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3832985/ – NIA Monkeys
4. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4115619/ – Mortality data
5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2662372/ – BMI meta-analysis
6. https://qz.com/255210/turns-out-the-way-americans-measure-healthy-weight-is-totally-wrong/ – media report on American obesity (for estimating average BMI)
7. https://www.nbcnews.com/healthmain/real-shape-american-man-dudes-youre-porky-8C11394082 – media report on American obesity (for estimating average BMI)
8. https://www.theatlantic.com/health/archive/2013/10/this-

is-the-average-mans-body/280194/ – media report on American obesity (for estimating average BMI)
9. https://www.cdc.gov/nchs/data/hestat/obesity_adult_13_14/obesity_adult_13_14.pdf – NIH obesity 2013-2014
10. https://www.cdc.gov/nchs/data/databriefs/db288.pdf – CDC obesity 2015-2016
11. Dixon, John B. “The effect of obesity on health outcomes.” Molecular and cellular endocrinology 316, no. 2 (2010): 104-108. https://doi.org/10.1016/j.mce.2009.07.008
12. https://www.ahajournals.org/doi/full/10.1161/ATVBAHA.111.241927 – biomolecular consequences of obesity
13. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3354762/ – various characteristics of centarians
14. https://www.ncbi.nlm.nih.gov/pubmed/12208237/
15. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3893695/ – effects of IGF and GH on body size, longevity, and mortality
16. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2673798/ – effects of CR and protein restriction on IGF serum levels
17. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4707876/#MXS013C74 – FOXO3A variation and longevity
18. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4296168/ – APOE variation and longevity, supports FOXO3, refutes others
19. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3362219/ – Okinawan longevity
20. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6092679/ – global life expectancy trends by country
21. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5315691/ – lots of stuff, general review of CR
22. https://www.ncbi.nlm.nih.gov/pubmed/16810568 – calorie restriction in Okinawa
23. https://jamanetwork.com/journals/jama/article-abstract/223580 – involuntary rationing in Denmark during

WWI
24. https://www.ncbi.nlm.nih.gov/pubmed/14795790 – involuntary rationing in Norway during WWII
25. https://www.ncbi.nlm.nih.gov/pubmed/17413101 – Tufts CR study that showed decreased insulin levels in overweight subjects.
26. https://academic.oup.com/jn/article-abstract/21/1/45/4725572 – 1941 rodent CR study
27. https://www.ncbi.nlm.nih.gov/pubmed/10197641/ – Pugh et al. 40% rodent CR study
28. https://www.ncbi.nlm.nih.gov/pubmed/4056321/ – Yu et. al 40% rodent CR study
29. https://www.ncbi.nlm.nih.gov/pubmed/7063854/ – Weindruch and Walford rodent CR study
30. https://www.ncbi.nlm.nih.gov/pubmed/2520283 – 1935 McCay rodent CR study
31. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5247583/ – 2017 combined Rhesus monkey study
32. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5565679/ – review of rodent CR studies
33. https://www.ncbi.nlm.nih.gov/pubmed/12016155/ – reduced cancer risk in p53 deficient mice
34. https://www.ncbi.nlm.nih.gov/pubmed/17460208/ – increase proteasome activity in CR mice and rats
35. https://www.ncbi.nlm.nih.gov/pubmed/18002475/ – improved cognition in CR mice and rats
36. https://www.ncbi.nlm.nih.gov/pubmed/19199090/ – anti-inflammatory effects of CR in rodents
37. https://www.ncbi.nlm.nih.gov/pubmed/16920310/ – improved glucose tolerance and lower insulin levels in CR rodents
38. https://www.ncbi.nlm.nih.gov/pubmed/24844367/ – drop in serum insulin in CR rodents
39. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3696189/ – review on budding yeast longevity
40. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3607354/

– review on longevity in general
41. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1857724/ – Drosphila CR review
42. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181168/ – Progeria rapamycin
43. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6509231/ – "Obesity May Accelerate the Aging Process"
44. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3836174/ – The Hallmarks of Aging
45. https://www.ncbi.nlm.nih.gov/pubmed/30115431/ – genome damage by obesity
46. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6509231/ – obesity/aging review
47. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2805851/ – telomere attrition in the obese
48. https://www.ncbi.nlm.nih.gov/pubmed/21088573/ – nutritional epigenetics
49. https://www.ncbi.nlm.nih.gov/pubmed/24779963/ – methylation patterns and diabetes
50. https://www.ncbi.nlm.nih.gov/pubmed/24154559/ – obese TD2 rat chromatin density
51. https://www.ncbi.nlm.nih.gov/pubmed/25380530/ – more obese rodent chromatin density
52. https://www.ncbi.nlm.nih.gov/pubmed/27503945/ – pro-inflammatory environment of adipose tissue
53. https://www.ncbi.nlm.nih.gov/pubmed/29155300/ – mitochondria ROS/dysfunction in the obese
54. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2941545/ – obesity and senescence review
55. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5127667/ – obesity and senescence in mice
56. https://www.ncbi.nlm.nih.gov/pubmed/22772162/ – stem cell quiescence in adipose tissue
57. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5599616/ – methylation drift in rodents, monkeys
58. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2190719/

– CR’s effect on genomic stability

59. https://www.ncbi.nlm.nih.gov/pubmed/17665967/ – CR promotes autophagy of misfolded proteins in rats

60. https://www.ncbi.nlm.nih.gov/pubmed/30395873/ – CR and cell senescence review

61. https://www.ncbi.nlm.nih.gov/pubmed/25481406/ – CR reduction of stem cell exhaustion

62. https://www.sciencedirect.com/science/article/pii/S1934590912001671 – CR and skeletal muscles, including transplant

63. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4889297/ – SIRT6 and NF-kB

64. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4570809/ – SIRT-1 signaling and CR in rats

65. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3331748/ – AMPK causes insulin sensitivity in CR mice

66. Kealy, Richard D., Dennis F. Lawler, Joan M. Ballam, Sandra L. Mantz, Darryl N. Biery, Elizabeth H. Greeley, George Lust, Mariangela Segre, Gail K. Smith, and Howard D. Stowe. “Effects of diet restriction on life span and age-related changes in dogs.” Journal of the American Veterinary Medical Association 220, no. 9 (2002): 1315-1320. https://admin.avma.org/News/Journals/Collections/Documents/javma_220_9_1315.pdf?7fh285_auid=1555113600043_jueqhhd825meind2et (Dog CR)

67. https://academic.oup.com/biomedgerontology/article/60/6/688/590315 (live infection in CR mice)

68. https://link.springer.com/article/10.1007/s11357-008-9056-1(live infection in CR mice)

69. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3528375/ (wound healing in CR mice)

70. ISBN 978-0816672349, Ancel Keys, Biology of Human Starvation (Minnesota starvation)

71.

https://www.sciencedirect.com/science/article/pii/S1043276009000915 (SIRT mutants vs CR)
72. https://www.pnas.org/content/pnas/107/14/6127.full.pdf (Mattson control mice)
73. The Retardation of Aging in Mice by Dietary Restriction: Longevity, Cancer, Immunity and Lifetime Energy Intake1 – Weindruch mouse study, 86
74. https://www.sciencedirect.com/science/article/pii/S0047637407001005 CR metabolic rate 1
75. https://www.sciencedirect.com/science/article/pii/S0047637499000949 CR metabolic rate 2
76. https://jphysiolanthropol.biomedcentral.com/articles/10.1186/1880-6805-31-22 Okinawa thermodynamics
77. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/96E9E7436516D443974E1C6C8859299D/S0029665194000170a.pdf/the-right-weight-body-fat-menarche-and-fertility.pdf Starvation/Fertility women
78. Biology of Aging: Observations and Principles, Arkin, general description of aging theory (doubling mortality rate)

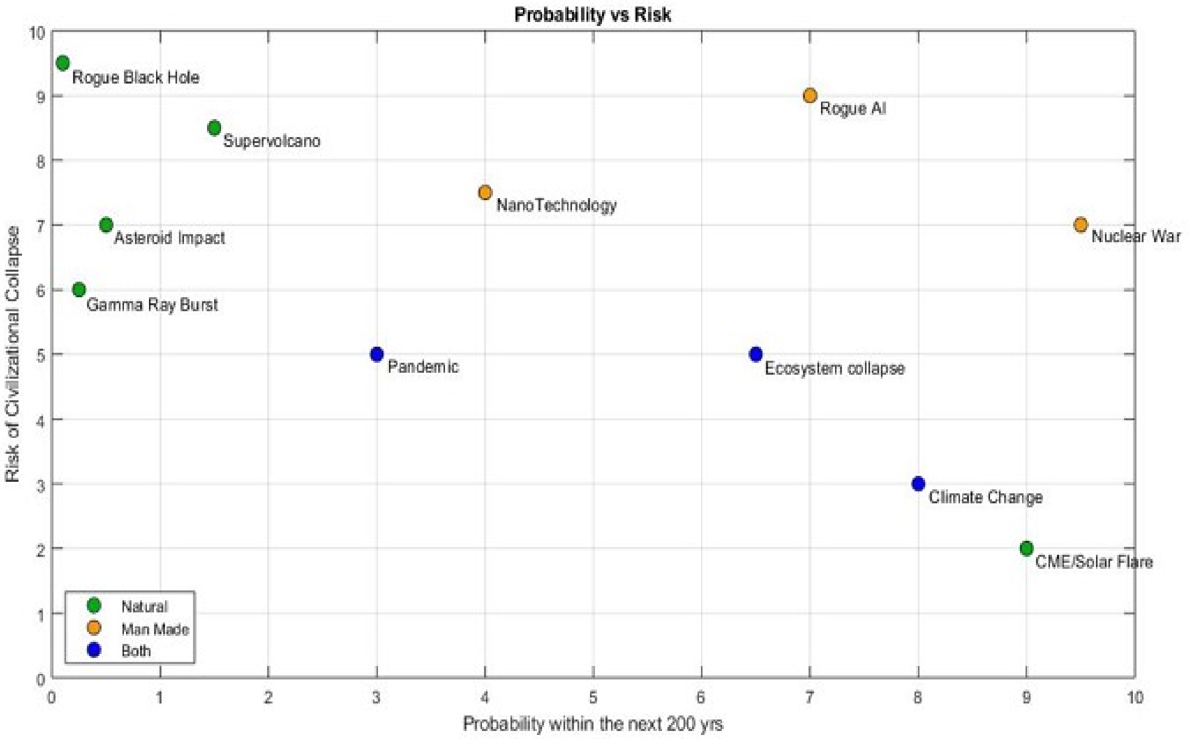
[ACC] Should We Colonize Space To Mitigate X-Risk?

[This is an entry to the 2019 Adversarial Collaboration Contest by Nick D and Rob S.]

I.

Nick Bostrom defines existential risks (or X-risks) as “[risks] where an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential.” Essentially this boils down to events where a bad outcome lies somewhere in the range of ‘destruction of civilization’ to ‘extermination of life on Earth’. Given that this has not already happened to us, we are left in the position of making predictions with very little directly applicable historical data, and as such it is a struggle to generate and defend precise figures for probabilities and magnitudes of different outcomes in these scenarios. Bostrom’s introduction to existential risk provides more insight into this problem than there is space for here.

There are two problems that arise with any discussion of X-risk mitigation. Is this worth doing? And how do you generate the political will necessary to handle the issue? Due to scope constraints this collaboration will not engage with either question, but will simply assume that the reader sees value in the continuation of the human species and civilization. The collaborators see X-risk mitigation as a “Molochian” problem, as we blindly stumble into these risks in the process of maturing our civilisation, or perhaps a twist on the tragedy of the commons. Everyone agrees that we should try to avoid extinction, but nobody wants to pay an outsized cost to prevent it. Coordination problems have been solved throughout history, and the collaborators assume that as the public becomes more educated on the subject, more pressure will be put on world governments to solve the issue.



Exactly which scenarios should be described as X-risks is impossible to pin down, but on the chart above, the closer you get to the top right, the more significant the concern. Considering there is no reliable data on the probability of a civilization collapsing pandemic or many other of these scenarios, the true risk of any scenario is impossible to determine. So any of the above scenarios should be considered dangerous, but for some of them, we have already enacted preparations and mitigation strategies. World governments are already preparing for X-risks such as nuclear war, or pandemics by leveraging conventional mitigation strategies like nuclear disarmament and WHO funding. When applicable, these strategies should be pursued in parallel with the strategies discussed in this paper. However, for something like a gamma ray burst or grey goo scenario, there is very little that can be done to prevent civilizational collapse. In these cases, the only effective remedy is the development of closed systems. Lifeboats. Places for the last vestiges of humanity to hide and survive and wait for the catastrophe to burn itself out. There is no guarantee that any particular lifeboat would survive. But a dozen colonies scattered across every continent or every world would allow humanity to rise from the ashes of civilization.

Both authors of this adversarial collaboration agree that the human species is worth preserving, and that closed systems represent the best compromise between cost, feasibility, and effectiveness. We disagree, however, on if the lifeboats should be terrestrial, or off world. We’re going to go into more detail on the benefits and challenges of each, but in brief the argument boils down to whether we should aim more conservatively by developing the systems terrestrially, or ‘shoot for the stars’ and build an offworld base and reap the secondary benefits

II.

For the X-risks listed above, there are measures that could be taken to reduce the risk of them occurring, or to mitigate against the negative outcomes. The most concrete steps that have been taken so far that mitigate against X-risks would be the creation of organisations like the UN, intended to disincentivize warmongering behaviour and reward cooperation. Similarly the World Health Organisation and acts like the Kyoto Protocol serve to reduce the chances of catastrophic disease outbreak and climate change respectively. MIRI works to reduce the risk of rogue AI coming into being, while space missions like the Sentinel telescope from the B612 Foundation seek to spot incoming asteroids from space.

While mitigation attempts are to be lauded, and expanded upon, our planet, global ecosystem, and biosphere are still the single point of failure for our human civilization. Creating separate reserves of human civilization, in the form of offworld colonies or closed systems on Earth, would be the most effective approach to mitigating against the worst outcomes of X-risk.

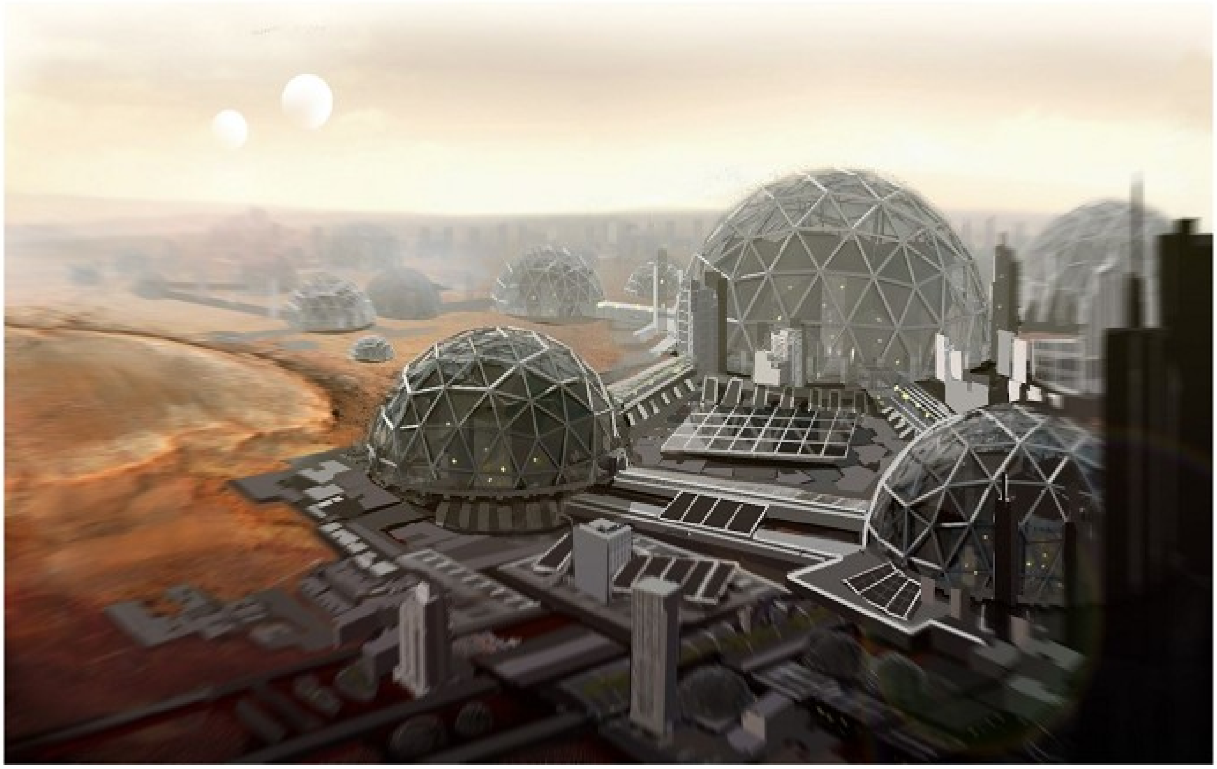
The scenario for these backups would go something like this: despite the best efforts to reduce the chance of any given catastrophe it occurs, and efforts made to protect/preserve civilization at large fail. Thankfully, our closed system or space colony has been specifically hardened to survive against the worst we can imagine, and a few thousand humans survive in their little self-sufficient bubble with the hope of retaining existing knowledge and technology until the point where they have grown enough to resume the advancement of human civilization, and the species/civilization loss event has been averted.

Some partial analogues come to mind when thinking of closed systems and colonies; the colonisation of the New World, Antarctic exploration and scientific bases, the Biosphere 2 experiment, the International Space Station, and nuclear submarines. These do not all exactly match the criteria of a closed system lifeboat, but lessons can be learned.

One of the challenges of X-risk mitigation is developing useful cost/benefit analyses for various schemes that might protect against catastrophic events. Given the uncertainty inherent in the outcomes and probabilities of these events, it can be very difficult to pin down the ‘benefit’ side of the equation; if you invest $5B in an asteroid mitigation scheme, are you rescuing humanity in 1% of counterfactuals or are you just softening the blow in 0.001% of them? If those fronting the costs can’t be convinced that they’re purchasing real value in terms of the future then it’s going to be awfully hard to convince them to spend that money. Additionally, the ‘cost’ side of the equation is not necessarily simple either, as many of the available solutions are unprecedented in scale or scope (and take the form of large infrastructure projects famous for cost-overruns). The crux of our disagreement ended up resting on the question of cost/benefit for terrestrial and offworld lifeboats, and the possibility of raising the funds and successfully establishing these lifeboats.

III.

The two types of closed systems under consideration are offworld colonies, or planetary closed systems. An offworld colony would likely be based on some local celestial body, perhaps Mars, or one of Jupiter’s moons. For an offworld colony, the X-risk mitigation wouldn’t be the only point in its favor. A colony would also be able to provide secondary and tertiary benefits in acting as a research base and exploration hub, and possibly taking advantage of otheropportunities offered by off-planet environments.



In terms of X-risk mitigation, these colonies would work much the same as the planetary lifeboats, where isolation from the main population provides protection from most disasters. The advantage would lie in the extreme isolation offered by leaving the Earth. While a planetary lifeboat might allow a small population to survive a pandemic, a nuclear/volcanic winter, or catastrophic climate change, other threats such as an asteroid strike or nuclear strikes themselves would retain the ability to wipe out human civilization in the worst case.

Offworld colonies would provide near complete protection from asteroid strikes and threats local to the Earth such as pandemics, climate catastrophe, or geological events, as well as being out of range of existing nuclear weaponry. Climate change wouldn’t realistically be an issue on Mars, the Moon, or anywhere else in space, pandemics would be unable to spread from Earth, and the colonies would probably be low priority targets come the breakout of nuclear war. While eradicating human civilisation would require enough asteroid strikes to hit every colony, astronomically reducing the odds.

Historically, the only successful drivers for human space presence have been political, the Space Race being the obvious example. I would attribute this to a combination of two factors; human presence in space doesn’t increase the value of scientific research possible enough to offset the costs of supporting them there, and no economically attractive proposals exist for human space presence. As such, the chances of an off-planet colony being founded as a research base or economic enterprise are low in the near future. This leaves them in a similar position to planetary lifeboats, which also fail to provide an economic incentive or research prospects beyond studying the colony itself. To me this suggests that the point of argument between the two possibilities lies on the trade-off between the costs of establishing a colony on or off planet, and the risk mitigation they would respectively provide.

The value of human space presence for research purposes is only likely to decrease as automation and robotics improve, while for economic purposes, as access to space becomes cheaper, it may be possible to come up with some profitable activity for people off-planet. The most likely options for this would involve some kind of tourism, or if the colony was orbital, zero-g manufacturing of advanced materials, while an unexpectedly attractive proposal would be to offer retirement homes off planet for the ultra wealthy (to reduce the strain of gravity on their bodies in an already carefully controlled environment). It seems unlikely that any of these activities would be sufficiently profitable to justify an entire colony, but they could at least serve to offset some of the costs.

Perhaps the closest historical analogue to these systems would be the colonisation of the New World, the length of the trip was comparable (two months for the Mayflower, at least six to reach Mars), and isolation from home further compounded by the expense and lead time on mounting additional missions. Explorers traveling to the New World disappeared without warning multiple times, presumably due to the difficulty of sending for external help when unexpected problems were encountered. Difficulties associated with these kinds of unknown unknowns were encountered during the Biosphere projects as well, it transpired that trees grown in enclosed spaces won't develop enough structural integrity to hold their own weight, as it is the stresses due to wind that cause them to develop this strength. It appears that this was not something that was even on the radar before the project happened, while several other unforeseen issues also had to be solved, the running theme was that in the event of an emergency supplies and assistance could come from outside to solve the problem. A space-based colony would have to solve problems of this kind with only what would be immediately to hand. With modern technology, assistance in the form of information would be available (see Gene Kranz and Ground Control's rescue of Apollo 13), but lead times on space missions mean that even emergency flights to the ISS, for which travel time could be as little as ten minutes, aren't really feasible. As such off-planet lifeboats would be expected to suffer more from unexpected problems than terrestrial lifeboats, and be more likely to fail before there was even any need for them.

The other big disadvantage of a space colony is the massively increased cost of construction, Elon Musk's going estimate for a 'self sustaining civilization' on Mars is $100B – $10T, assuming that SpaceX's plans for reducing the cost of transport to Mars work out as planned. In order to offer an apples to apples comparison with the terrestrial lifeboat considered later in this collaboration, if Musk’s estimate for a population of one million for a self-sustaining city is scaled down to the 4000 families considered below (a population of 16000) our cost estimate comes down to $1.6B – $160B. Bearing in mind that this is just for transport of the requisite mass to Mars, we would expect development and construction costs to be higher. With sufficient political will, these kinds of costs can be met; the Apollo program cost an estimated $150B in today’s money (why the cost of space travel for private and government run enterprises has changed so much across sixty years is an exercise left to the reader). Realistically though, it seems unlikely that any political crisis will occur to which the solution seems to be a second space race of a similar magnitude. This leaves the colonization project in the difficult position of trying to discern the best way to fund itself. Can enough international coordination be achieved to fund a colonization effort in a manner similar to the LHC or the ISS (but an order of magnitude larger)? Will the ongoing but very quiet space race between China, what’s left of Western space agencies human spaceflight efforts, and US private enterprise escalate into a colony race? Or will Musk’s current hope of ‘build it and they will come’ result in access to Mars spurring massive private investment into Martian infrastructure projects?

IV

Planetary closed systems would be exclusively focused on allowing us to survive a catastrophic scenario (read: “zombie apocalypse”). Isolated using geography and technology, Earth based closed systems would still have many similarities to an offworld colony. Each lifeboat would need to make its own food, water, energy, and air. People would be able to leave during emergencies like a fire, O2 failure or heart attack, but the community would generally be closed off from the outside world. Once the technology has been developed, there is no reason other countries couldn’t replicate the project. In fact, it should be encouraged. Multiple communities located in different regions of the world would actually have three big benefits. Diversity, redundancy, and sovereignty. Allowing individual countries to make their own decisions allows different designs with no common points of failure and if one of the sites does fail, there are other communities that will still survive. Site locations should be chosen based on

- Political stability of the host nation
- System implementation plan
- Degree of exposure to natural disasters
- Geographic location
- Cultural Diversity

There is no reason a major nation couldn’t develop a lifeboat on their own, but considering the benefits of diversity, smaller nations should be encouraged to develop their own projects through UN funding and support. A UN committee made up of culturally diverse nations could be charged with examining grant proposals using the above criteria. In practice, this would mean a country would go before the committee and apply for a grant to help build their lifeboat.

Let’s say the US has selected Oracle, Arizona as a possible site for an above ground closed system. The proposal points out the cool, dry air minimizes decomposition, located far from major cities or nuclear targets, and protected and partially funded by the United States. The committee reviews the request and their only concern is the periodic earthquakes in the region. To improve the quality of their bid, The United States adds a guarantee that the town’s demographics would be reflected in the system by committing to a 40% Latino system. The committee considers the cultural benefits of the site, and approves the funding.

Oracle, Arizona wasn’t a random example, In fact it’s already the site of the world’s largest Closed Ecological System [CES] It actually was used as the site of Biosphere 2. As described by acting CEO Steve Bannon:

> Biosphere 2 was designed as an environmental lab that replicated […] all the different ecosystems of the earth… It has been referred to in the past as a planet in a bottle.. It does not directly replicate earth [but] it’s the closest thing we’ve ever come to having all the major biomes, all the major ecosystems, plant species, animals etc. Really trying to make an analogue for the planet Earth.

I feel like I need to take a moment to point out that that was not a typo, and the quote above is provided by that Steve Bannon. I don’t know what else to say about that other than to acknowledge how weird it is (very).

As our friend Steve “Darth Vader” Bannon points out, what made Biosphere 2 unique, is that it was a Closed Ecological System where 8 scientists were sealed into an area of around 3 acres for a period of 2 years (Sept 26, 1991 – Sept. 27, 1993). There are many significant differences from the Biosphere 2 project and a lifeboat for humanity. Biosphere 2 contained a rainforest, for example. But the project was the longest a group of humans have ever been cut off from earth (“Biosphere 1”). Our best view into what issues future citizens of Mars may face is through the glass wall of a giant greenhouse in Arizona.

One of the major benefits of using terrestrial lifeboats as opposed to planetary colonies is that if (when) something goes wrong, nobody dies. There is no speed of light delay for problem solving, outside staff are available to provide emergency support, and in the event of a fire or gas leak, everyone can be evacuated. In Biosphere 2, something went wrong. Over the course of 16 months the oxygen in the Biosphere dropped from 20.9% from 14.5%. At the lowest levels, scientists were reporting trouble climbing stairs and inability to perform basic arithmetic. Outside support staff had liquid oxygen transported to the biosphere and pumped in.

A 1993 New York Times article “Too Rich a Soil: Scientists find Flaw That Undid The Biosphere” reports:

> A mysterious decline in oxygen during the two-year trial run of the project endangered the lives of crew members and forced its leaders to inject huge amounts of oxygen […] The cause of the life-threatening deficit, scientists now say, was a glut of organic material like peat and compost in the structure’s soils. The organic matter set off an explosive growth of oxygen-eating bacteria, which in turn produced a rush of carbon dioxide in the course of bacterial respiration.

Considering a Martian city would need to rely on the same closed system technology as Biosphere 2, It seems that a necessary first step for a permanent community on Mars would be to demonstrate the ability to develop a reliable, sustainable, and safe closed system. I reached out to William F. Dempster, The Chief Engineer for the Biosphere 2. He has been a huge help and provided tons of papers that he authored during his time on the project. He was kind enough to point out some of the challenges of building closed systems intended for long-term human habitation:

> What you are contemplating is a human life support pod that can endure on its own for generations, if not indefinitely, in a hostile environment devoid of myriads of critical resources that we are so accustomed to that

> we just take them for granted. A sealed structure like Biosphere 2 [....] is absolutely essential, but, if one also has to independently provide the energy and all the external conditions necessary, the whole problem is orders of magnitude more challenging.

The degree to which an off-planet lifeboat would lack resources compared to a terrestrial one would be dependent on the kind of disaster scenario that occurred, in some cases such as pandemic, it could be feasible to eventually venture out and recover machines, possibly some foods, and air and water (all with appropriate sterilization). While in the case of an asteroid strike or nuclear war at a civilization-destruction level, the lifeboat would have to be resistant to much the same conditions as an off-planet colony, as these are the kind of disasters where the Earth could conceivably become nearly as inhospitable as the rest of the solar system. To provide similar levels of x-risk protection as an off-planet colony in these situations, the terrestrial lifeboat would need to be as capable as Dempster worries.

While Biosphere 2 is in many ways a good analogue for some of the challenges a terrestrial closed system would face, There are many differences as well. First, Biosphere 2 was intended to maintain a living, breathing, ecosystem, while a terrestrial lifeboat would be able to leverage modern technology in order to save on costs, and the cost for a terrestrial lifeboat is really the biggest selling point. A decent mental model could be a large, relatively squat building, with an enclosed central courtyard. Something like the world's largest office building. It cost 1 billion dollars in today's money to build, and bought us 6.5 million sq ft of living space. Enough for 4000 families to each have a comfortable 2 bedroom home. A lifeboat would have additional expenses for food and energy generation, as well as needing medical and entertainment facilities, but the facility could have a construction cost of around $250,000 per family. The median US home price is $223,800.

There is one additional benefit that can’t be overlooked, Due to the closed nature of the community, the tech centric lifestyle, and combined with the subsidized cost of living. There is a natural draw for software research, development, and technology companies. Creating a government sponsored technology hub would allow young engineers a small city to congregate, sparking new innovation. This wouldn’t and shouldn’t be a permanent relocation. In good times, with low risks, new people could be continuously brought in and cycled out periodically, with lockdowns only occurring in times of trouble. The X-risk benefits are largely dependent on the facilities themselves, but the facilities will naturally have nuclear fallout and pandemic protection as well as a certain amount of inclement weather or climate protection. Depending on the facility, There could be (natural or designed) radiation protection. Overall, a planetary system of lifeboats would be able to survive anything an offworld colony would survive, outside of a rogue AI or grey goo scenario. But simultaneously the facilities would have a very low likelihood of a system failure resulting in massive loss of life the way a Martian colony could.

V.

To conclude, we decided that terrestrial and off-planet lifeboats offer very similar amounts of protection from x-risks, with off-planet solutions adding a small amount additional protection in certain scenarios whilst being markedly more expensive than a terrestrial equivalent, with additional risks and unknowns to the construction process.

The initial advocate for off-planet colonies now concedes that the additional difficulties associated with constructing a space colony would encourage the successful construction of terrestrial lifeboats before attempts are made to construct one on another body. The only reason to still countenance their construction at all is an issue which revealed itself to the advocate for terrestrial biospheres towards the end of the collaboration. A terrestrial lifeboat could end up being easily discontinued and abandoned if funding/political will failed, whereas a space colony would be very difficult to abandon due to the astronomical (pun intended) expense of transporting every colonist back. A return trip for even a relatively modest number of colonists would require billions of dollars allocated over several years, by, most importantly, multiple sessions of a congress or parliament. This creates a paradigm where a terrestrial lifeboat, while being less expensive and in many ways more practical, could never be a long term guarantor of human survival do to its ease of decommissioning (as was seen in the Biosphere 2 incident). To be clear, the advocate for terrestrial lifeboats considers this single point sufficient to decide the debate in its entirety and concedes the debate without reservation.

[ACC] Should Gene Editing Technologies Be Used In Humans?

[This is an entry to the 2019 Adversarial Collaboration Contest by Nita J and Patrick N.]

Introduction

In October 2018, the world’s first genetically edited babies were born, twin girls given the pseudonyms Lulu and Nana; Chinese scientist He Jiankui used CRISPR technology to edit the CCR5 gene in human embryos with the aim of conferring resistance to HIV. In response to the international furor, China began redrafting its civil code to include regulations that would hold scientists accountable for any adverse outcomes that occur as the result of genetic manipulation in human populations. Now, reproductive biologists at Weill Cornell Medicine in New York City are conducting their own experiment designed to target BRCA2, a gene associated with breast cancer, in sperm cells. While sometimes considered controversial, gene editing has been used as a last resort to cure some diseases. For example, a precursor of CRISPR was successfully used to cure leukemia in two young girls when all other treatment options had failed. Due to its convenience and efficiency, CRISPR offers the potential to fight cancer on an unprecedented level and tackle previously incurable genetic diseases. However, before we start reinventing ourselves and mapping out our genetic futures, maybe we should take a moment to reevaluate the risks and repercussions of gene editing and rethink our goals and motives.

How does CRISPR work?

CRISPR, which stands for clustered regularly interspaced short palindromic repeats, is an adaptive bacterial immune response that protects against repeat offenders. When exposed to a pathogenic bacteriophage, a bacterium can store some viral phage DNA in its own genome in “spacers,” which function as genetic mug shots, allowing the bacterium to quickly mount a defense in case of future invasions. When necessary, the CRISPR defense system will slice up any DNA matching these genetic fingerprints. In 2012, Jennifer Doudna and Emmanuelle Charpentier demonstrated how CRISPR could be used to slice any DNA sequence of choice. The CRISPR-Cas9 system allows researchers to not only recognize and remove DNA sequences but also modify them. The completion of the Human Genome Project in 2003 provided a copy of the genetic book of life; CRISPR offers a way to purportedly erase and “correct” certain words in that book.

Of course, this newfound power raises several ethical concerns. The major worry among scientists revolves around the long-term consequences of germline modification, meaning genetic changes made in a human egg, sperm, or embryo. Edits made in the germline will affect every cell in an organism and will also be passed on to any offspring. If a mistake is made in the process and a new disease inadvertently introduced, these changes will persist for generations to come. Human germline modification could also theoretically allow for the installation of genes to confer protection against infections, Alzheimer’s, and even aging. For many, the thought of controlling our own genetic destinies seems to be a very slippery slope, conjuring up dystopian images of Frankenstein or Brave New World. For these reasons and more, in 2015, Doudna and other scientists proposed a moratorium on the use of CRISPR-Cas9 for human genome editing until safety and efficacy issues could be more thoroughly addressed.

How safe and efficient is gene editing?

CRISPR is currently being used in clinical trials for cancers and blood disorders; since these interventions won’t lead to heritable DNA changes, these trials don’t face the same ethical dilemmas as Dr. He’s experiment but may nevertheless carry risks. Doubts persist about the safety and efficacy of the CRISPR gene editing system, as many other initially promising technologies have failed. Conventional gene therapies, which attempt to insert healthy copies of genes into cells using viruses, faced many early setbacks, including the tragic death of 18-year-old Jesse Gelsinger in 1999 during a gene therapy trial for ornithine transcarbamylase deficiency. However, the causes surrounding Gelsinger’s death may have included a systemic immune response triggered by the use of a viral vector.

While the death of Jesse Gelsinger marked a somber moment for the field, gene therapy also experienced successes when researchers from Paris treated two young infants who suffered from a fatal form of severe combined immunodeficiency disease (SCID), an inherited disorder characterized by low levels of T cells and natural killer cells, which leaves affected patients incredibly susceptible to infection. Fortunately, viral gene therapy was able to reverse the disease symptoms in this particular case. On the other hand, gene therapy trials using viral vectors were recently halted when 25-50 percent of gene therapy patients developed leukemia resulting from the insertion of a gene-carrying virus near an oncogene; a gene with the potential to cause cancer. Modern CRISPR technology is not affected by such hurdles, however, as it does not rely on the use of viral vectors. While more precise than traditional gene therapy, CRISPR nonetheless sometimes results in unintended edits, which may be especially problematic for certain gene targets. Some pairs of genes are “linked” due to physical proximity on the same chromosome and are therefore almost always passed on together. Any edits to a gene belonging to a linked pair may therefore inadvertently cause an edit in its neighboring partner.

Even intended cuts can have unexpected consequences. Two separate 2018 studies published in *Nature Medicine*, one conducted by the Karolinska Institute in Sweden and the other by Novartis Institutes for Biomedical Research, concluded that CRISPR edits might increase the risk of cancer via inhibition of a tumor suppressor gene called P53, which has been described as “the guardian of the genome” due to its crucial role in maintaining genomic stability. Double-stranded DNA breaks made by CRISPR activate repair mechanisms encoded by P53 that instruct the cell to either mend the damage or self-destruct. Making these types of edits successfully would therefore require inhibition of P53; however, cells could become more vulnerable to tumorigenic mutations and the development of cancer as a result. “We don’t always fully understand the changes we’re making,” says Alan Regenberg, a bioethicist at Johns Hopkins Berman Institute of Bioethics. “Even if we do make the changes we want to make, there’s still question about whether it will do what we want and not do things we don’t want.”

Nevertheless, a slight increase in cancer risk might be a worthwhile trade-off for many patients with genetic diseases, such as the aforementioned SCIDs, which affect 1 in 50,000 people globally. Usually, the only cure for SCIDs is a bone marrow transplantation, which requires a matched donor in order to avoid rejection by host immune cells or, alternatively, the depletion of T cells to avoid rejection in the case of an unmatched donor. CRISPR offers a safer, more efficient way to treat genetic diseases such as SCIDs.

Bone marrow cells of a patient may be extracted and genetically modified using CRISPR, thereby avoiding rejection by the host immune system. Pre-clinical trials in mice are already underway to test the safety and efficacy of this approach. Stanford scientist Dr. Matthew Porteus demonstrated the efficiency of this technique and said in an interview, “We don’t see any abnormalities in the mice that received the treatment. More specifically, we also performed genetic analysis to see if the CRISPR-Cas9 system made DNA breaks at places that it’s not supposed to, and we see no evidence of that.”

CRISPR also offers the additional possibility of removing parts of a gene, providing extra value over standard viral gene therapy, which only allows for insertion of genes. This feature can be especially important for autosomal dominant genetic disorders, which are made manifest with only one copy of a deleterious mutation. In her book, A Crack in Creation, Jennifer Doudna speculates that as CRISPR becomes increasingly safe, the tool may be used to help people who aren’t fortunate enough to win the genetic lottery. Doudna intones, “Someday we may consider it unethical not to use germline editing to alleviate human suffering.” What was unthinkable just a few years ago will soon enter clinical praxis.

Are some genetic variants superior to others?

In biology, those organisms that are most suited to their environment exhibit the highest fitness, a measure that accounts for both survival and reproduction. The accumulation of mutations over time is thought to contribute to many disease processes, but genetic diversity can also be beneficial for an organism when faced with a changing environment or unanticipated stress, such as drought or illness. Discussions on rigid natural selection should give way to more nuanced conversations on “balancing selection, the evolutionary process that favors genetic diversification rather than the fixation of a single ‘best’ variant,” as described by Professor Maynard V. Olson at the University of Washington.

Evolution has allowed many potentially deleterious genes to remain in the gene pool due to their ability to impart a selective advantage to individuals with carrier status, a phenomenon referred to as heterozygote advantage. Sickle cell anemia is a disease inherited in an autosomal recessive pattern—two copies of the problematic gene variant are necessary for disease expression. However, having just one copy of that variant confers resistance to malaria, which may explain the increased prevalence of sickle cell anemia in areas where malaria is more common, namely India and many countries in Africa. In this manner, malaria acts as a selective evolutionary pressure maintaining the occurrence of the sickle cell variant in the gene pool.

Nevertheless, sickle cell disease has become prevalent in countries currently unaffected by malaria. In the United States, approximately 100,000 people suffer from sickle cell disease, but therapeutic options remain limited. Researchers have been investigating the possible insertion of wild-type, “anti-sickling” genes using viral vectors in affected patients as therapy. However, since the pathological mutation for sickle cell disease has already been clearly identified, correction of the mutated gene using CRISPR may offer a more straightforward approach. The biotech company CRISPR Therapeutics recently announced the results of a phase I clinical trial in which CRISPR technology was used to treat a patient with sickle cell disease, although the efficacy and safety of the intervention have not yet been evaluated.

Can gene editing eliminate disease?

To answer these questions, we need to first evaluate our understanding of genetics and weigh the importance of genetics against environmental factors such as diet and lifestyle.

How reliable is our understanding of gene-disease links?

A mutation is usually defined as a genetic sequence that differs from the agreed-upon consensus or “wild-type” sequence. After the completion of the Human Genome Project in 2003, the arduous process of genome annotation began. Genome-wide association studies, or GWAS, began examining population data over time to look for possible associations between genetic variants, or genotypes, and physical traits and diseases, or phenotypes. Unfortunately, these studies often fail to employ random sampling, and 96 percent of subjects included in GWAS have been people of European descent. In fact, scientific disciplines frequently disproportionately sample from WEIRD (western, education, industrialized, rich, democratic) populations, whether studying genetic diseases or human gut microbiota.

Given the sources of genetic information used to determine “wild-type” sequences, we may be using information that is relevant to one demographic but not another. According to Maynard Olson, one of the founders of the Human Genome Project, the wild-type human simply doesn’t exist, and “genetics is unlikely to revolutionize medicine until we develop a better understanding of normal phenotypic variation.” These words seem to have fallen on deaf ears, however, as evidenced by the burgeoning numbers of genome-wide association studies conducted over the last 12 years. Most of the associations discovered thus far are only correlative, and few studies have been conducted to determine whether observed associations are indeed causal.

Closer examination of the relationship between gene variants and certain diseases reveals weak associations in many cases. For example, the APOE gene, which encodes for the production of a protein known as apolipoprotein E, comes in three genetic forms- APOE2, APOE3, and APOE4 with the last being associated with an increased risk of developing Alzheimer's disease (AD). However, the correlation is not determinative, as the Nigerian population exhibits high frequencies of the APOE4 allele but low frequencies of AD. Environment and nutrition also play significant roles in the disease pathophysiology, as illustrated by Dr. Dale Bredesen's research demonstrating reversal of cognitive decline through a targeted dietary and lifestyle approach. In fact, the majority of afflictions commonly affecting the general population, such as type 2 diabetes, cardiovascular disease, cancer, Alzheimer's, and Parkinson's are not caused solely by mutations.

How often does disease arise as the result of genetic mutation alone?

Chronic diseases are the result of a complex interplay between host genetics and the environment. A study conducted by the Wellcome Trust Sanger Institute in Cambridge, England analyzed DNA sequencing data from 179 people of African, European, or East Asian origin as part of the 1000 Genomes Pilot Project and discovered that healthy individuals carried an average of 400 mutations in their genes, including around 100 loss-of-function variants that result in the complete inactivation of about 20 genes that encode for proteins. These findings indicate that deleterious mutations, even those that lead to protein damage, do not invariably give rise to disease. As Professor

James Evans from the University of North Carolina, who was not involved in the study, summarized in an NPR health blog, “We’re all mutants. The good news is that most of those mutations do not overtly cause disease, and we appear to have all kinds of redundancy and backup mechanisms to take care of that.” The authors hypothesize that healthy individuals can carry disadvantageous mutations without showing ill effects for a number of possible reasons: an individual may carry just one copy of a gene mutation for a recessive disorder that requires two mutations in order to manifest the disease, the disease may exhibit delayed onset or require additional environmental factors for expression, or the reference catalogs used to identify gene-disease links may be inaccurate. One analysis found that 27 percent of database entries cited in the literature were incorrectly identified.

To account for the discrepancy between genetic predisposition and disease manifestation, in 2005, cancer epidemiologist Dr. Christopher Wild proposed the concept of the exposome, which encompasses “life-course environmental exposures (including lifestyle factors) from the prenatal period onwards” and accounts for factors such as socioeconomic status, chemical contaminants, and gut microflora. The risk of developing a chronic disease during one’s lifetime may be modeled by G×E: the interaction between a person’s genetics (G) and lifetime exposures (the exposome, E). Identical twin studies reveal that genotype alone cannot determine whether a given phenotype will be expressed, and the interaction between genes and the environment must be taken into account.

In fact, the “genes load the gun, environment pulls the trigger” paradigm may be overly simplistic, as Dr. Alessio Fasano at Harvard Medical School has shown that loss of intestinal barrier function is likely also necessary for the development of chronic inflammation, autoimmunity, and cancer. Two particular gene markers, HLA-DQ2 and HLA-DQ8, are observed in the vast majority of celiac disease cases. While over 30 percent of the U.S. population carries one or both of the necessary genes, only around one percent of Americans are affected by celiac disease. This data suggests that exposure to gluten through ingestion of wheat, barley, or rye is not sufficient to trigger the development of celiac disease even in individuals with a genetic predisposition. Without the additional loss of intestinal tight junction function, celiac disease is not made manifest. Thus, factors besides genetics are necessary for the development of chronic disease.

How does gene expression contribute to disease risk?

The concept of genetic determinism purports that our genes are our destiny, but genes are not nearly as important as gene expression. When most people think of evolution, the first name that comes to mind is Charles Darwin, but a contemporary of Darwin's named Jean Baptiste Lamarck had proposed a theory of "acquired characteristics" by which individuals evolved certain traits within their lifetimes. The most oft-cited example discrediting this theory is that of giraffes elongating their necks by stretching to reach the treetops and then passing on this trait of long necks to their progeny. In contrast, Darwin proposed that those giraffes that had the longest necks went on to find food, survive, and reproduce. Eventually, Darwin's theory of natural selection prevailed, but his 18th century French naturalist contender may have simply foreseen the field of epigenetics, the study of those drivers of gene expression that occur without a change in DNA sequence. The prefix "epi-" means above in Greek, and epigenetic changes determine whether genes are switched on or off and also influence the production of proteins. If you imagine your genetic code as the hardware of a computer, epigenetics is the software that runs on top and controls the operation of the hardware. Epigenetic changes control the expression of genes through various mechanisms and are influenced by diet, exercise, lifestyle, sunlight exposure, circadian rhythms, stress, trauma, exposure to pollutants, and other environmental factors.

The epigenetic mechanism of DNA methylation involves tagging DNA bases with methyl groups, a process that tends to silence genes. DNA methylation is responsible for X-chromosome inactivation in females, a process necessary to ensure that females don’t produce twice the number of X-chromosome gene products as males. Methylation is also responsible for the normal suppression of many genes in somatic cells, allowing for cell differentiation. Every somatic cell in the human body contains nearly identical genetic material, but skin cells, muscle cells, bone cells, and nerve cells exhibit different properties due to different sets of genes being turned on or off. Dietary nutrients such as vitamin B12, folic acid, choline, and betaine double as methyl donors, so even small changes in nutritional status during gestation can result in markedly different effects on gene expression and varied physical characteristics in the offspring. If differential gene expression can produce such drastic changes, is genome rewriting really necessary? Perhaps the centrality of the gene in driving human health has been overstated. Indeed, why worry about a potentially pathogenic gene if it is never expressed?

Inappropriate DNA methylation has been referred to as a “hallmark of cancer,” along with uncontrolled cell growth and proliferation. Almost all types of human tumors are characterized by two distinct phenomena: global hypomethylation, which may result in the expression of normally suppressed oncogenes, genes that promote tumor formation, as well as regional hypermethylation near tumor suppressor genes. In other words, genes that promote tumor formation are turned on while genes that suppress tumor formation are turned off. Cigarette smoke has been shown to promote both demethylation of metastatic genes in lung cancer cells as well as regional methylation of other specific genes via modulation of enzymatic activities. To succinctly summarize, genes themselves are not driving tumor formation; rather inappropriate gene expression is increasing the risk of tumor development.

Can gene editing treat cancer?

Cancers are front and center among the conditions gene editing therapies are targeted to treat. To answer the question of whether CRISPR can be used to treat cancer, we need to first examine how cancer arises. Medical textbooks frequently attribute the development of cancer to the accumulation of mutations over time. However, the accumulation of genetic mutations is not sufficient to cause cancer; the tumor microenvironment must be taken into account. In other words, the same oncogenic mutation that is adaptive for cancer in altered tissue is not advantageous to cancer in healthy, homeostatic cells.

James DeGregori at the University of Colorado School of Medicine offers the following analogy. When tackling drug dealing in the inner city, arresting all the drug dealers is unlikely to work; the ones left behind will be smarter and more conniving. Instead, one might focus on creating better jobs, schools, and infrastructure, so citizens won’t have to resort to crime as a means of survival. Addressing the environment that lead to the problem in the first place will provide a more stable long-term solution. Similarly, instead of simply targeting the cancer, altering the microenvironment to disfavor its proliferation may provide a more viable long-term strategy, as the former immediately selects for resistance, accounting for the difficulty in keeping a patient in remission. Highlighting the importance of the microenvironment in regulating development, homeostasis, and cancer, biologist Mina Bissell writes, “The sequence of our genes are like the keys on the piano; it is the context that makes the music.” Cancer depends on context, as should our approach to treatment.

Nevertheless, despite recent medical advances, cancer treatment has not seen significant improvement in decades. Standard therapies rely on toxic chemotherapy, which destroys both cancerous and healthy tissue. Furthermore, cancerous cells often evade detection and destruction by host immune defenses by expressing cell surface molecules that prevent killing by host T cells. A new and effective form of immunotherapy known as chimeric antigen receptor (CAR) T cell therapy attempts to harness the power of the human immune system to recognize and kill cancer cells. However, this method has several disadvantages. A patient must have a sufficient number of immune cells prior to beginning therapy, which may not be the case for patients who have already received chemotherapy. Additionally, the process is time-consuming, and the use of viral vectors may increase the risk of developing other cancers.

To address the issues of T cell collection and manufacturing delays, researchers are now developing “off-the-shelf” CAR T cells, which utilize gene editing to prevent rejection by the host immune system and the development of graft-versus-host disease (GvHD), a condition in which foreign immune cells attack the recipient’s body. In 2017, two infants with relapsing leukemia were successfully treated with these “off-the-shelf” CAR T cells, which were modified using the genome editing tool TALEN. Short for transcription activator-like effector nucleases, TALEN can be considered the predecessor to CRISPR and uses enzymes that are specifically guided to a genomic sequence to induce a cut. However, designing these enzymes requires extensive work, making the process costly and time-consuming. Additionally, in vitro studies have demonstrated that CRISPR techniques exhibit better correction efficiencies and fewer off-target effects than TALEN. Moreover, the use of CRISPR can speed up the manufacturing of CAR T cells and drive down costs of such therapies from hundreds of thousands of dollars to a few hundred dollars.

Can gene editing prevent HIV?

Another prospective application for CRISPR technology is the treatment of HIV. Today, approximately 37 million people around the world live with HIV. The use of antiretroviral drugs has greatly reduced the death rate from 1.9 million in 2004 to less than one million in 2017. Challenges still exist, as human immunodeficiency virus (HIV) inserts itself into the host genome and mutates rapidly, making complete eradication of the disease very difficult. About one percent of the population is naturally immune to HIV due to a CCR5 gene mutation, which prevents the expression of a cell surface receptor that HIV binds to in order to gain entry into host cells. As previously mentioned, the first genetically edited babies were born in October 2018 after Chinese scientist Dr. He Jiankui used CRISPR technology to edit the CCR5 gene in human embryos.

According to Dr. He, a married couple with the pseudonyms Mark and Grace consented to in vitro fertilization with additional CRISPR treatment to provide immunity to HIV for their offspring. First, a process called sperm washing was used to separate sperm from semen, the fluid that carries HIV. Next, eggs were fertilized by sperm to create embryos, on which Dr. He performed CRISPR gene editing. After several implant attempts, successful pregnancy was achieved. Nine months later, twins with the pseudonyms Lulu and Nana were born healthy and purportedly suffered no off-target effects from the CRISPR therapy.

Testing indicated that gene editing did not successfully alter both copies of the CCR5 gene in one of the twins, however. Chinese researchers were apparently knowledgeable of the gene editing failure prior to the pregnancy attempt; the decision to proceed with implantation regardless has numerous ethical implications. “In that child, there really was almost nothing to be gained in terms of protection against HIV and yet you’re exposing that child to all the unknown safety risks,” said Dr. Kiran Musunuru, a professor of stem cell and regenerative biology at Harvard University. The choice to use the unedited embryo suggests that the researchers may have been more focused on testing the accuracy of the gene editing technology than providing immunity to disease.

According to the Chinese government and his employers, Dr. He acted without the knowledge or consent of his superiors. Chinese authorities suspended all of He’s research activities, saying his work was “extremely abominable in nature” and a violation of Chinese law. In fact, the procedure was not medically necessary. When only the father is HIV-positive, as in this case, sperm washing alone is usually sufficient to reduce transmission of the virus. A meta-analysis that investigated the efficacy of sperm washing did not find a single case where HIV was transmitted to offspring.

Dr. He claims that the CCR5 gene is already very well characterized, but a recently published study found that decreased function of the CCR5 gene enhances cognitive function in mice. At first glance, this new knowledge may appear to be a boon, but the potential benefit also invites a discussion on the possibility of designer babies. Another point to consider is the fact that the CCR5 mutation that confers HIV immunity more commonly appears in Caucasians and may make individuals more susceptible to infections that are common in Asia.

Can gene editing be used to create designer babies?

A discussion on human genome editing would not be complete without evaluating the potential to create "designer babies," a term commonly used in the vernacular to refer to babies with genetic enhancements. Both the utility of gene editing for basic research and the use of somatic gene editing to heal individuals who are sick are generally widely accepted among the public. The waters become murkier when we consider germline editing and the possibility of preventing disease or altering traits unrelated to health needs. In the 1970s, scientists first began to establish distinctions between somatic and germline genome modifications; somatic edits only affect a single individual while germline edits can be passed down over generations. By the mid-1980s, bioethicists began to argue that the morally relevant line was between disease and enhancement rather than somatic and germline. Discussions of heritable enhancements in particular raise fears of a possible return to eugenics.

John Fletcher, former head of bioethics at the National Institutes of Health (NIH), once wrote, "The most relevant moral distinction is between uses that may relieve real suffering and those that alter characteristics that have little or nothing to do with disease." Many scientists today share the sentiment that treatment and prevention of "disease" constitute acceptable uses of CRISPR technologies while "enhancement" applications should be discouraged, but the boundary between the two is riddled with semantic discord.

Moreover, the line delineating disability and disease is often blurred, and many perceived shortcomings may in fact represent normal variation on the phenotypic spectrum.

The discussion of whether we can or should modify human characteristics may be a moot point since our knowledge of which genes affect complex traits such as height, intelligence, and eye color is still limited. Additionally, most traits are influenced not only by genetics but also environmental factors, and monozygotic twin studies demonstrate that genes alone cannot predict whether physical traits will be expressed. Furthermore, genes that encode for physical traits may also impart increased vulnerability to certain diseases. For example, variations in the MC1R gene responsible for red hair color may also increase the risk of developing skin cancer. As indicated earlier, Dr. He’s efforts to confer resistance to HIV may have also resulted in increased susceptibility to infection by West Nile virus or influenza. As always, trade-offs exist, and the idea of the “perfect specimen” is a fallacy. Any efforts to gain genetic advantages will always be subject to the limitations of biology.

How should society move forward with gene editing technology?

CRISPR technology holds invaluable potential as a research tool and possible treatment for diseases caused by single-point genetic mutations. As previously described, some genetic diseases can be treated by stem cell gene editing without the need for germline modification, thereby minimizing the risk for potential mistakes that could be passed on to subsequent generations. On the other hand, trying to correct an error after a certain point during development is sometimes problematic, as the error has already been incorporated into billions of cells. Jennifer

Doudna offers the following visual: “Imagine trying to correct an error in a news article after the newspapers have been printed and delivered, as opposed to when the article is still just a text file on the editor’s computer.” Germline editing may therefore provide a more expedient option for the prevention of some genetic diseases such as sickle cell disease or cystic fibrosis.

One of the most compelling arguments against CRISPR gene editing, namely the potential for misuse, can also be considered the most compelling argument for CRISPR gene editing. Banning progress on gene editing technology may create a black market, but the continuation of research on gene editing will allow the scientific community to control its use and ensure patient safety. Research into CRISPR is continually finding ways to make the technology safer and more effective; a paper published in September 2019 reported on the potential for a novel CRISPR system to affect gene expression in human cells. The process is reversible in theory and doesn’t involve the cutting of DNA, thereby reducing the risk of human harm and leveraging the power of epigenetics.

Moreover, while gene expression and the tumor microenvironment are viable targets for cancer treatment, gene editing can be considered a last resort therapy for certain cases in which other interventions have failed. Common chronic diseases, such as Alzheimer’s, type 2 diabetes, and cardiovascular disease, likely require a more nuanced approach, as gene expression, governed by factors such as diet and lifestyle, plays a significant role in disease pathogenesis. The use of gene editing to mold favorable traits, such as eye or hair color, likely exposes individuals to unnecessary risks and does not constitute medical necessity. Nevertheless, many consider mainstream germline gene editing an inevitability. Joseph Fletcher, one of the founders of bioethics, wrote in 1971, “Man is a maker and a selector and a designer, and the more rationally contrived and deliberate anything is, the more human it is.” The establishment of gene editing guidelines should include input from scientists, policy makers, and the public and incorporate the most current knowledge available in order to prevent misuse and realize potential. As the custodians of such powerful technology, we must take care to use it in an ethical and responsible manner. Whether our efforts will alleviate human suffering or ensure the survival of our species, only time will tell.

[ACC] When During Fetal Development Does Abortion Become Morally Wrong?

[This is an entry to the 2019 Adversarial Collaboration Contest by BlockOfNihilism and Icerun]

Note: For simplicity, we have constrained our analysis of data about pregnancy and motherhood to the United States. We note that these data are largely dependent on the state of the medical and social support systems that are available in a particular region.

Introduction: Review of abortion and pregnancy data in the United States

We agreed that it was important to first reach an understanding about the general facts of abortion, pregnancy and motherhood in the US prior to making ethical assertions. To understand abortion rates and distributions, we reviewed data obtained by the CDC’s Abortion Surveillance System (1). The Pregnancy Risk Assessment Monitoring System (PRAMS), Pregnancy Mortality Surveillance System (PMSS) and National Vital Statistics datasets were used to evaluate the medical hazards imposed by pregnancy (2, 3, 4). Finally, we examined a number of studies performed on the Turnaway Study cohort, maintained by UCSF, to investigate the economic effects of denying wanted abortions to women (5, 6, 7, 13).

Abortion rates by trimester and maternal age: Using data collected by the CDC, 638,169 abortions were performed in the United States in 2015. Data was received from 49/52 reporting areas, suggesting that these rates are likely close to the population rates. This was equivalent to 188 abortions per 1000 live births, a 24% decline from 2006. Of these, approximately 65% were performed prior to 8 weeks of development, and 91% before 13 weeks of development. An additional 7.6% were performed at between 14-20 weeks. Approximately 90% of abortions were performed on women older than 19, and adolescent women between the ages of 18-19 accounted for 67% of the abortions in women under 19. By race, non-Hispanic black women were most likely to undergo an abortion (25 per 1000 women), while non-Hispanic white women were least likely (6.7 per 1000). This translates to a rate of 390 abortions per 1,000 live births in non-Hispanic black women and 111 per 1,000 live births in non-Hispanic white women. (1) These data show that most abortions are undertaken prior to the end of the first trimester, that most women choosing an abortion are adults, and that non-Hispanic black women are disproportionately more likely to choose an abortion.

Mortality and morbidity associated with abortion and pregnancy: On average, there were 0.62 fatalities per 100,000 legal abortions between 2008-2014 (six reported fatalities in 2014). For comparison, in 2015 there were 17.2 pregnancy-related fatalities per 100,000 live births in 2014. These data suggest that an abortion is generally safer than attempting to carry a child to term. Also, it is important to consider the racial disparities within these data. For example, African-American women were three times as likely to die as a result of pregnancy than non-Hispanic white women (42.8 vs 13 per 100,000 live births). The reasons for these disparities are unclear. (3)

Pregnant women are also at risk for severe morbidity associated with pregnancy and delivery, with approximately 50,000 women experiencing at least one severe complication in 2014. This translated to a rate of ~140/10,000 deliveries. Approximately 1.2% of live births resulted in severe maternal complications. Women can also experience significant psychological morbidity after pregnancy, as 1 out of 9 women who deliver a live fetus develop postpartum depression. We were unable to find CDC data for morbidity resulting from abortion procedures; however, one publication reported approximately 2% of abortions result in a medical complication. As this data did not discriminate between minor and severe complications, it would be reasonable to assume that abortions result in a lower overall severe complication rate than pregnancy. We will make the further assumption (based on educated guessing) that late-term abortions are more risky than early-term abortions. (2)

From these data, we conclude that pregnancy and delivery pose a significant risk to the mother’s health. These risks are greatest for African-American and Native American women. By comparison, abortion appears to pose much lower risks of death, and probably much lower risks of morbidity. Consequently, mothers undergo unique and substantial hazards which are imposed by pregnancy.

Comparison of pregnancy-associated risks and other common risk factors: It is difficult to compare the risks of pregnancy with other factors due to the disparate means of measuring those risks (per live birth, vs per person). However, a naive interpretation of the available data suggests that, while pregnancy is relatively unlikely to lead to severe consequences, it compares in risk to other common activities. For example, the mortality rate associated with motor vehicle accidents is 12.1 per 100,000 people. This is similar to the risk of death per 100,000 live births for women in the US (16.7). (8)

An alternative approach is to examine how pregnancy, childbirth and post-pregnancy changes affect overall mortality. According to the National Vital Statistics Reports (Volume 68, 2016), pregnancy and childbirth was the 6th leading cause of death for women(all races and ethnicities) aged 20-24 and 25-29, accounting for 652 deaths in the two groups combined. Pregnancy and childbirth was the 10th leading cause of death for women between the ages of 15-19 (28 deaths). These data indicate that pregnancy is a leading cause of death in women of child-bearing age. (4)

Socioeconomic costs of unwanted pregnancy: The socioeconomic effects of abortion denial have been studied extensively on the Turnaway Study cohort at the University of California-San Francisco. One study on this cohort found that mothers who were denied a wanted abortion due to gestational age experienced a significantly higher likelihood of being unemployed, in poverty and using public assistance programs like WIC. (6) Another study based on this cohort found that already-born children of a mother denied an abortion were significantly more likely to live in poverty and fail to meet developmental milestones.(5) Mothers who were denied abortions were also less likely to have and meet aspirational goals.(7) These data indicate that women who received wanted abortions experience significantly less socioeconomic strain than women who are denied an abortion.

Adoption vs abortion: Adoption is commonly suggested as an alternative to abortion. Adoption does eliminate the direct socioeconomic burdens of parenthood. However, adoption is rarely considered as an alternative to abortion. For example, in the U.S., there were approximately 18,000 adoptions compared with nearly 1 million abortions. A recent article in The Atlantic did an excellent job of summarizing potential reasons for the discrepancy. Adoption obviously does not alleviate the physical burdens and hazards of pregnancy. Additionally, several studies have suggested that women do not choose adoption due to worry about their perception of the emotional effect of giving away a child. Pro-adoption groups also suggest that both pro- and anti-abortion advocates fail to emphasize or properly counsel women on considering adoption as an alternative to abortion. (9)

Who are the stakeholders in the abortion question? The mother, the father, the fetus, and society at large. The mother's unique interests are her safety and health, the development of a unique bond with a new human life, and the economic, emotional and physical burdens of motherhood. The father, if held responsible, shares the economic and emotional burdens of parenthood. The fetus, once it has developed the fundamental features of a human being, has at least a theoretical interest in preserving its life. Society at large has an interest in justice and preserving the rights of its members, if only out of self-interest for the individuals within that society. At some point in time, a fetus becomes considered a member of that society, with the same rights as all other individuals. Consequently, the point of conflict arises when a mother (or both parents) desires to terminate a pregnancy prior to delivery.

The question: At what point during development does abortion become a moral wrong?

Starting positions: At conception (icerun), At fetal viability/minimal neurological activity (BlockofNihilism)

icerun's Position: A Future Like Ours: Conception

Many arguments for and against abortion pick out a characteristic of the fetus – its size, level of consciousness, ability to feel pain, etc. – and go on to argue why this characteristic, or lack of one, gives the fetus a right to life.

Unfortunately, these characteristics tend to have accidental byproducts – they may give the right to life to sheep or remove it from infants. The Future Like Ours arguments begin by determining what best accounts for the wrongness of killing people like you and me (who people on both sides of the abortion debate agree it is wrong to kill). And then use this standard to determine if it is wrong to kill a fetus (who it is contested whether it is wrong to kill).

A Future Like Ours

In Why Abortion is Immoral, Marquis argues killing someone like you or me is prima facie wrong because the deceased is robbed of a valued future like ours. (10) Killing most directly and significantly harms the one who is killed.

The harm to the deceased is the loss of her valued future. Her future would have included all of the experiences, relationships, and works that were valuable for their own sake or means to something valuable. She loses not only those parts of the future she valued in the moment but also those experiences, relationships, and works that she would have come to value as she grew older or is not currently aware of as she grew older: a 16 year old may not value parts of his future whether that be a career, family, or woodwork but if the teenager had been allowed to develop may have come to value these parts of his future.

In summary, it is wrong to kill somebody like you or me because it robs them of a future like ours. The value of a fetus' future is its current experience, relationships, and works that the fetus values now and those experiences, relationships, and works that the fetus would come to value. A typical fetus cannot currently value it's experiences, relationships, and works but as the fetus develops it will come to have the same experiences, relationships, and works that we do. Therefore, a fetus has a future like ours.

By this definition, it is wrong to kill a fetus from the point of conception (for the record, Marquis does not claim it is wrong to kill a fetus from the point of conception; however, this seems to be the implication).

Intuitions: The future like ours argument works off common assumptions by pro and antiabortion proponents. In doing so it both avoids assuming an ought from an is and creates common ground. The account of the wrongness of killing humans must fit within these intuitions: it must account for why it is wrong to kill typical adult humans, infants, and those who are suicidal but it is not wrong to kill typical sperm, eggs, and some animals. However our intuitions differ on whether it is wrong to kill a typical single cell zygote. Intuitively we both believe it is not wrong to kill a typical zygote, however BlockofNihilism believes this strongly and I believe this weakly. Many anti-abortion advocates have the opposite intuiition.

For BlockofNihilism, this future like ours argument violates his strong intuition that it is not seriously wrong to kill a zygote and this argument fails. For myself, it violates a weak intuition and while on it’s own is not enough to completely overcome the intuition, it holds the strongest sway and influence over my view on abortion as it offends the least intuitions and is more coherent than most other arguments.

BlockofNihilism’s Position: Conscious Perception and Viability

Abortion is morally acceptable until the fetus develops the structures required for perception of external stimuli, with exceptions for preserving the life and health of the mother. Abortion is acceptable because a fetus does not experience conscious suffering “like ours” and simultaneously imposes a significant physical, mental and economic burden on the mother. As the minimum requirements for conscious perception are actually met after fetal viability, I suggest we fall back on viability as a compromise ethical barrier to abortion.

When does the fetus develop “conscious” perception? By conscious perception, we mean perception which a human person would recognize as their own. Obviously, this question in general pushes the limits of our ability of description. As perception is an (obviously) complex topic, I will use the perception of physical pain as an example of the requirements for conscious perception. Pain, too, is a complex psychological concept that arises at the intersection of physical sensation with emotional constructs. At the minimum empirical level, certain neurological structures are necessary, but not sufficient, for the perception of pain. Thus, until these structures are present and active, perception (as we understand it) cannot occur. (10)

To experience pain, afferent nerves must synapse with spinothalamic nerves projecting to the thalamus, which then connect to thalamocortical neurons projecting to the cortex (the region of conscious experience). Thus, all three components (peripheral pain sensor, thalamic project, and functioning cortex) must all be active for the perception of pain. Based upon multiple studies, nociceptive neurons develop around 19 weeks, thalamic afferents reach the cortex at 20-24 weeks, and somatosensory activity provoked by thalamic activity is detectable around 28-29 weeks. Several behavioral studies have found that at 29-30 weeks of development, fetal facial movements in response to pain are like adults. However, these results have been contradicted by other studies, and these findings may represent non-voluntary and unconscious responses to stimuli rather than the conscious perception of pain. (10)

In any event, a fetus does not have the required neurological structures for what we would recognize as the conscious experience of pain until at least 29 weeks of development, three weeks into the third trimester. (10) Prior to full integration of the various components of the nervous system, and the development of an active cortical system, the pain experience of a fetus would likely be akin to that of a comatose individual- no conscious experience at all.

As other types of experience require these same structures to be active, we can conclude that a fetus does not have the minimum capacity for conscious experience until approximately 29 weeks of development. Thus, when considering an abortion prior to this stage of development, we are balancing (1) the harms posed to the mother, a conscious agent, against (2) an entity that does not “experience” anything. To me, this suggests that abortion is permissible at this point.

The fetus is truly viable at ~27 weeks: With intensive care, a preterm neonate can survive at as early as 24 weeks of gestation. However, survival rates at this point are approximately 50%. Also, these severely preterm neonates are at a significantly increased risk of a variety of both short- and long-term complications. By 27 to 28 weeks, the fetus can be delivered and survive in most cases without major interventions. So, true fetal viability and the development of the fundamentals for conscious experience are roughly concurrent, with viability likely being reached prior to conscious experience.

Potential harms and viability: Viability means that the fetus no longer requires the mother’s body to survive. Within the womb, the fetus imposes both a significant immediate burden as well as the potential for significant harms. Once safely delivered, these harms are no longer present. While the mother is still on the hook for the economic and emotional burdens of motherhood, her life is no longer at risk. Also, while adoption is a possibility after birth, it is obviously not an option prior to delivery. Consequently, viability represents a special moment in the development of a fetus- it can live without posing a significant hazard to the mother’s physical well-being. While we could not find solid evidence (likely due to the very low number of late-term abortions performed), my educated guess is that an abortion at this late stage is approximately as dangerous as performing a natural delivery or C-section. Consequently, at viability, it is reasonable to treat the fetus as having full human rights and intercede to protect its life.

icerun’s rebuttal to BlockofNihilism:

Viability: The only difference between a viable fetus and an infant is location, which is not a moral distinction (except in cases of direct harm to the mother) therefore a viable fetus is seen as having the same right to life as an infant. The chain would seem to continue. The primary distinction between a viable fetus and nonviable fetus is that a nonviable fetus survival depends solely on one person (the mother) whereas a viable fetus survival can depend on others. This does not appear to be a moral distinction either and so the viability argument appears to be very closely related to the argument that the fetus gains the right to life at birth or when it becomes an infant. Therefore, a viable fetus would have the same right to life of a newborn however without further reasoning, it seems likely a fetus gains the right to life earlier.

Experience: BlockofNihilism argues that since a fetus before 29 weeks is not capable of conscious experience it is not capable of suffering and therefore it is not wrong to abort. However, there are times when adult humans are not conscious and are even unable to achieve consciousness in the case of temporarily comatose humans. Because they are not conscious they are not capable of suffering. This argument seems to allow for the killing of sleeping and temporarily comatose humans as long as they do not suffer, feel pain, or realize what is happening in the moment.

Further, an adult would likely not recognize the consciousness of a fetus as its own. It is unlikely that a fetus or infant has a sense of self and they seem to operate at a significantly lower level of self-awareness. Though we do not have a good understanding of the level of consciousness a fetus holds, a dog appears to operate at a higher stage of consciousness than a fetus though this is very speculative.

For these reasons, the experience of suffering is not what makes it wrong to kill a fetus or human.

BlockofNihilism’s Rebuttal of the Future Like Ours Account:

Consciousness-based and FLV-based arguments arrive at the same place: For me, any ethical argument that places the interests of a non-conscious entity incapable of experience above the interests of a conscious agent capable of both rational decision-making and of suffering is intuitively absurd. Prior to the development of the basics for neurological experience, the fetus represents the potential for a future life of value or the potential to be a conscious agent. In either case, I do not believe that the potential outweighs the present!

I understand how the future life-of-value (FLV) argument can seem to apply to a fetus: We imagine the entity that will come from the fetus, imagine its potential for an FLV, and extrapolate rights from there. However, a fetus represents the potential for having a life of value and cannot be said to currently possess that future in the way implied by Marquis. I believe the intuitive appeal of the future life-of-value argument arises from our experience and knowledge of what a “future” constitutes. However, fetuses prior to their development of the basic neurological structures required for experience cannot have or value their “future.”

My interpretation is informed by Boonin’s famous critique of Marquis’ “future of value” argument. (11) According to Boonin, the intuitive value of the future can be found in the dispositional ideal present value of a future. A dispositional value or belief is one that is held by someone but not consciously on the mind. An ideal value or desire is one that would be held if one had full information about the situation. The dispositional ideal desire formulation is more parsimonious as it does not invoke potential desires but only present ones. Thus, the wrongness of killing someone like you or me is the taking of a future like ours they dispositionally, ideally, and presently value. Upon developing the neurological structures necessary for experience, a fetus can begin to (at least unconsciously) desire food, close touch, and parent’s voice. The necessary neurological structures for these desires, and for meeting the minimum requirement for having an FLV, is near or at the point of viability.

Consciousness-based accounts do not allow for murdering sleeping people! There is a clear distinction between an entity that has had the experience of consciousness (a sleeping or temporarily comatose individual) and an entity that has never been conscious. A sleeping person still has her memories, desires and agency encoded within her brain; the fact that she is temporarily unaware of those attributes does not mean they do not exist! Conversely, a fetus prior to its developing consciousness has no memories, desires or agency. It cannot be said to be a person yet. My argument is simple- prior to having the minimum requirements for consciousness there is absolutely no chance whatsoever that a fetus can experience any harm like we (persons) do.

Once these structures are developed and active, it becomes far more difficult to determine “when” a fetus or infant reaches consciousness. At this point, I become squeamish with the prospect of destroying something that potentially does have a conscious experience (including a “future of value” concept) like ours. The moral calculus changes: Instead of balancing a person’s interest (mother) vs a nonperson’s interests (fetus), we now have a person vs (maybe a person?). This is where, to be safe and prevent potential harms, we can draw a clear ethical line.

Preventing abortion prior to viability will cause significant harms: As previously discussed, substantial scientific evidence suggests that preventing wanted abortions will lead to harm. First, there would be a significant increase in morbidity and mortality associated with pregnancy. This increase would disproportionately impact economically disadvantaged and minority women. Second, women denied wanted abortions are significantly more likely to suffer socially, economically and psychologically. Perhaps most importantly, women (or both parents) are denied agency and denied the ability to make the ethical decision for themselves according to their unique circumstances and beliefs.

Location, location, location! Viability represents the best point for ethical compromise: Terminating a fetus after it is capable of living “on its own” is equivalent to infanticide. In the special case of a fetus, location does have moral significance. The fetus, living within and dependent upon the mother’s body, poses immediate and potential costs and hazards to the mother. By contrast, once delivery has taken place the fetus/neonate no longer poses these threats. While the mother still has the significant economic and social burdens of motherhood, these burdens are unlikely to lead to immediate physical harm. And for the mother unable to cope with these burdens, adoption or surrendering the care of the infant to the state is an option once delivery of a viable neonate has taken place.

Icerun’s Defense of the Future Like Ours Account

Capacity of a fetus to have a future: The fetus does not have a potential future nor is the fetus’ future simply a concept in its brain. The future of a fetus are those unrealized experiences the fetus will have if its development is not impeded. Likewise, a 20-year-old will be a 25-year-old with experiences, relationships, and works if it’s development is not impeded. Sometimes a human’s development is impeded by natural causes in which case we mourn their loss of a future or by conscious decisions in which case we mourn and try to provide restitution as best possible.

In fact, one’s future is most certainly not in or dependent on the brain. A 4-year-old does not have a good understanding of what it is like to be a 60 year old yet being a 60 year old is still a part of his future. If the 4-year-old is killed, it has lost not only on the relationships it understands as a 4-year-old but also a future that includes a career or children or what it would have found valuable and meaningful as an adult.

Boonin’s present future account fails: Marquis and Boonin account for the value of the parts of our future that we do not know yet (ex: our future in 20 years) in different ways. Marquis includes both our present valuation of the future and our future valuation of the future while Boonin argues for a present ideal desire of the future. However, to have an ideal desire, one must first have an actual desire (if an actual desire is not required, then one could say the zygote or trees have ideal desires).

Though the fetus can be said to have desires, these desires are unconscious. A conscious desire is willed and chosen to a certain extent whereas an unconscious desire is simply the body doing what the body does; a personification which is often helpful but, in this case, not relevant. The unconscious desire for warmth is simply the brain releasing chemicals based on external states. Similarly, the zygote will begin to multiply based on external states, stem cells will divide into different cells based on external states, the heart begins beating based on external states. The heart beating or zygote splitting apart seem to fulfill the requirement of some unconscious desire. The fact that it is the brain responding to outside stimuli is not morally relevant – the fetus does not appear to be aware of a desire for warmth just as it is not aware of the heart's desire to pump blood. If conscious desires are necessary then newborns and possibly older infants likely do not have a right to life as they do not appear to have conscious desires or a sense of self. In this case though it is more parsimonious, it fails because it does not grant infants a right to life.

Conclusions

icerun's conclusion: The point where the fetus gains the right to life is rightly contested and debated as I do not believe there are any completely coherent and consistent arguments that define the point of development where the fetus gains the right to life.

The latest possible point where abortion may be permissible appears to be viability where the sole difference between an infant and a fetus is the location (one inside the womb and dependent on a specific person and the other outside the womb that could be cared for by others). However, abortion may be impermissible at an earlier point and the point of viability does not appear to have a moral significance that makes the fetus seriously wrong to kill.

At the end though, I have not come to a solid position at what point it becomes wrong to kill a typical fetus. And it is important to note, have failed to provide a coherent argument. In making my decision on abortion three items weigh heaviest:

First, in cases of consensual sex (excluding rape), parents hold a strong positive obligation to provide and protect a child once it gains the right to life. This obligation comes from the fact that children have a right to life, require support to survive, and that the parents engaged in activities that are known to create humans. Second, the future like ours argument points to the fetus gaining a right to life at conception and though this goes against my intuitions, it comes the closest to providing a coherent and consistent argument. It is a model to understand why it is seriously wrong to kill humans and thus points to an earlier rather than later point in the fetus' development. My choice of this argument is likely biased by various intuitions that I hold and others would not doubt come to focus on other flawed arguments based on their own intuitions. Third, there are situations where bearing a child brings significant issues and problems for either the mother or fetus where abortion apears the best option.

A mesh of all three in light of it being uncertain when the right to life begins for a fetus perhaps leads to the stance that abortion should be safe, legal, and rare that investigates abortion on a cases by case basis that attempts to balance the weightiness of aborting a fetus with practical costs and difficulties that are imposed on parents.

BlockofNihilism's conclusion: If my ethical standard were to be adopted and used to change current practice in the US, it would allow for a few more elective early third-trimester abortions than are currently performed. However, it would have little to no effect on the current situation, as most abortions are performed well before viability. I believe that communicating our knowledge about the fetus pre-viability, including its lack of internal conscious experience, would significantly reduce the potential for psychological harms to women who choose abortions. In contrast, if abortion after conception was prevented, there would be several negative consequences. There would be a significant increase in pregnancy-related morbidity and mortality that would disproportionately affect minority and socioeconomically distressed women. The likely uptick in illegal abortions would increase the likelihood of unsafe abortions, further increasing the risk of morbidity and mortality. Finally, the denial of wanted abortions imposes pronounced social and economic strains on new mothers and their families. These consequences are, obviously, of significant moral concern.

I remain convinced that abortion is acceptable prior to fetal viability. I believe that the intuitive appeal of the FLV argument is, as suggested by Boonin, not applicable to a fetus prior to developing the fundamental requirements for neurological experience. Even if we decided that the FLV argument pertained to fetuses, the fact that abortion pre-viability cannot cause conscious harm outweighs any potential for FLV that could result from a fetus carried to term. I believe (like Aesop) that a bird in the hand (the mother’s rights, interests and potential for harm) far outweighs a bird in the bush (the non-conscious potential person represented by a fetus).

Shared conclusion: Abortion is never a happy choice. Regardless of our ethical position on the abortion question, we agree that new people are of tremendous value! Improvements in the delivery and efficacy of birth control options, increases in social support systems for mothers and parents, reducing pregnancy-associated morbidity and mortality and increasing access to alternative options like adoption are all essential factors in reducing the number of abortions and any potential harms that arise from them. By focusing on these issues rather than on preventing abortions directly through legal or ethical edicts, we can make having a child a more reasonable and safe option than at present.

Works cited:

1. https://www.cdc.gov/reproductivehealth/data_stats/abortion.htm
2. https://www.cdc.gov/prams/index.htm
3. https://www.cdc.gov/reproductivehealth/maternal-mortality/pregnancy-mortality-surveillance-system.htm
4. https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68_06-508.pdf
5. Foster DG, Raifman S, Gipson JD, Rocca CH, Biggs MA. Effects of Carrying an Unwanted Pregnancy to Term on Women’s Existing Children. February 2019. The Journal of Pediatrics, 205:183-189.e1.
6. Foster DG, Biggs MA, Ralph L, Gerdts C, Roberts SCM, Glymour MA. Socioeconomic Outcomes of Women Who Receive and Women Who Are Denied Wanted Abortions in the United States. January 2018. American Journal of Public Health, 108(3):407-413

7. Upadhyay UD, Biggs MA, Foster DG. The effect of abortion on having and achieving aspirational one-year plans. November 2015. BMC Women’s Health, 15:102. (Request pdf)
8. https://www.cdc.gov/nchs/fastats/accidental-injury.htm
9. https://www.theatlantic.com/health/archive/2019/05/why-more-women-dont-choose-adoption/589759/
10. Marquis, Don. “Why Abortion Is Immoral.” The Journal of Philosophy, vol. 86, no. 4, 1989, pp. 183–202. JSTOR, www.jstor.org/stable/2026961.
11. Lee SJ, Ralston HJP, Drey EA, Partridge JC, Rosen MA. Fetal Pain: A Systematic Multidisciplinary Review of the Evidence. JAMA. 2005;294(8):947–954. doi:https://doi.org/10.1001/jama.294.8.947
12. Boonin, D. (2002). A Defense of Abortion (Cambridge Studies in Philosophy and Public Policy). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511610172
13. https://www.ansirh.org/research/turnaway-study

A Maximally Lazy Guide To Giving To Charity In 2019

[Sorry for the interruption; we will return to our regularly scheduled Adversarial Collaboration Contest tomorrow.]
[Epistemic status: I’m linking evaluations made by people I mostly trust, but there are many people who don’t trust these, I haven’t 100% evaluated them perfectly, and if your assumptions differ even a little from those of the people involved these might not be very helpful. If you don’t know what effective altruism is, you might want to find out before supporting it. Like I said, this is for maximally lazy people and everyone else might want to investigate further.]

If you’re like me, you resolved to donate money to charity this year, and are just now realizing that the year is going to end soon and you should probably get around to doing it. Also, you support effective altruism. Also, you are very lazy. This guide is for you.

The maximally lazy way to donate to effective charity is probably to donate to **EA Funds**. This is a group of funds run by the Center for Effective Altruism where they get experts to figure out what are the best charities to give your money to each year. The four funds are Global Health, Animal Welfare, Long-Term Future, and Effective Altruism Meta/Community. If you are truly maximally lazy, you can just donate an equal amount to all four of them; if you have enough energy to shift a set of little sliders, you can decide which ones get more or less.

If you have a little more time and energy, you might want to look at the charities suggested by some charity-evaluating organizations and see which ones you like best.

GiveWell tries to rigorously evaluate charities that can be rigorously evaluated, which usually means global health. They admit that they have to exclude whole categories of charity that try to change society in vague ways, because those charities can’t be evaluated as rigorously. But they do a good job of what they do. Most of their top charities fight malaria and parasitic worms; this latter cause is interesting because these worms semipermanently lower school performance, concentration, and general health, suggesting that treating them could permanently improve economic growth. You can donate directly to GiveWell (to be divided up among their top charities at their discretion) here, or you can look at their list of top recommended charities for 2019 here.

Animal Charity Evaluators is the same thing, but for charities that try to help animals, usually by fighting factory farming. You can donate to ACE’s Recommended Charity Fund, again to be divided up among their top charities at their discretion, here, or see their list of top recommended charities for 2019 here.

AI Alignment Literature Review And Charity Comparison is a report posted by LW user Larks going over all the major players in AI safety, what they’ve been doing the past year, and which ones need more funding. If you just want to know which ones they like best, CTRL+F “conclusions” and run it through rot13. Or if you’re too lazy to do that and you just want me to link you their top recommended charity’s donation page, it’s here.

Vox’s report on the best charities for climate change lists ones that claim to be able to prevent one ton of carbon emissions for $0.12 and $1, compared to the $10 you would get on normal offset sites. Their top choice is Coalition For Rainforest Nations (but see criticism here), and their second choice is Clean Air Task Force.

You might also want to check out ImpactMatters (a version of GiveWell focused on literal First World problems), Let's Fund (a site that highlights charities, mostly in science and technology, and runs campaigns for them), this post on the Effective Altruism forum about which charities people are donating to this year, and this list of what charities the charity selection experts at the Open Philanthropy Project are donating to.

And if you're not actually lazy at all, you might want to check out some interesting individual charities that have been making appeals around here recently (others can add their appeals in the comments if they want).

The Center For Election Science tries to convince US cities (and presumably plans to eventually work up to larger areas) to use approval voting, a form of voting where third party candidates don't "split the vote" and you can vote for whoever you want with a clear conscience. They argue this will make compromise easier and moderate candidates more likely to win. They've already succeeded in changing the ballot in Fargo, North Dakota, and as the old saying goes, "as Fargo, North Dakota goes, so goes the world."

Happier Lives Institute wants to work directly on making people happier, but they realize nobody really knows what that means, so they're doing a lot of meta-research on what happiness is and what the best way to measure it is. Aside from that, they seem to be working on cheap mental health interventions in Third World countries.

Machine Intelligence Research Institute works on a different aspect of AI alignment than most other groups; this comic explains the technicalities better than most sources. They are secretive and don't talk a lot about their work or give a lot for people to evaluate them on, so whether or not you donate will probably be based on whether they’ve won social trust with you (they have with me).

Charter Cities Institute is trying to work with investors and Third World governments to create charter cities, autonomous cities with better institutions that can supercharge growth in the Third World. For example, a corrupt Third World country where doing business is near-impossible might designate one of their cities to be administered by foreign judges under an open-source law code, so that enterprise can take off. Think of it as a seastead, except on land, and with the host country’s consent (they’re hoping to profit off the tax revenue). David Friedman’s son Patri is leading another effort in this direction.

Finally, if you’re really skeptical and don’t believe any charity can accomplish much, you might want to consider GiveDirectly, which just gives your money directly to very poor people in Africa to do whatever they want with.

[ACC] Will Automation Lead To Economic Crisis?

[This is an entry to the 2019 Adversarial Collaboration Contest by Doug Summers-Stay and Erusian]

Adversarial collaboration on the question: "Automation/AI will not lead to a general, sustained economic crisis within our lifetimes or for the foreseeable future. Automation/AI's effects into the future will have effects similar to technology's effects in the past and, on the whole, follow the general trend."

Defending the proposition: Erusian

Challenging the proposition: Doug Summers-Stay

tldr: Until the pace of automation increases faster than new jobs can be created, AI shouldn't be expected to cause mass unemployment or anything like that. When AI can pick up a new job as quickly and cheaply as a person can, then the economy will break (but everything else will break too, because that would be the Singularity).

Introduction

As software and hardware grow more capable each year, many are concerned that automation of jobs will lead to some sort of economic crisis. This could take the form of permanent high levels of unemployment, wages that drop below subsistence levels for many workers, or an abrupt change to a different economic system in response to these conditions.

This has become a talking point outside of economic circles in the U.S. Democratic presidential candidate Andrew Yang's most well-known policy proposal is a universal basic income to offset this (an idea Elon Musk has supported for years). Bill Gates suggested that when robots replace workers, the companies should be taxed at a similar rate to the taxes being paid by those workers. These entrepreneurs have spent a lot of time thinking about and planning for the future, and have a lot of experience with introducing new technology. Are their concerns valid?

Throughout this discussion, we use the words AI, automation, and robots more-or-less interchangeably. Imagining Asimov-style androids with positronic brains makes it easier to picture a world where all jobs are automated. In reality, though, it would be a silly waste of resources to literally have robots come in and do jobs as drop-in replacements for workers, and there are few jobs where this would make sense. A lot of software in the future will be more human-like in the sense that many machines could have natural language and image-understanding capabilities and have the ability to reason about the wider context in which their work exists to avoid dangerous or costly mistakes due to a lack of common sense. In many other ways, though, software for nearly all working robots will not be similar to human minds at all.

Some jobs, like picking most fruits and vegetables or most assembly line jobs, currently can’t be done by machines because they require manual dexterity. In such cases, you would need more precise manipulators with touch sensors that can adapt to a wide range of situations. There is no reason to expect they will look like human hands, though. Machines will also be networked, of course, so imagining them as a bunch of individuals is also unrealistic. Finally, the way tasks are currently divided into jobs makes sense for human workers, but wouldn’t make sense for automation. Instead, certain tasks will be automated first, and the remaining tasks that form part of a job will still be done by humans.

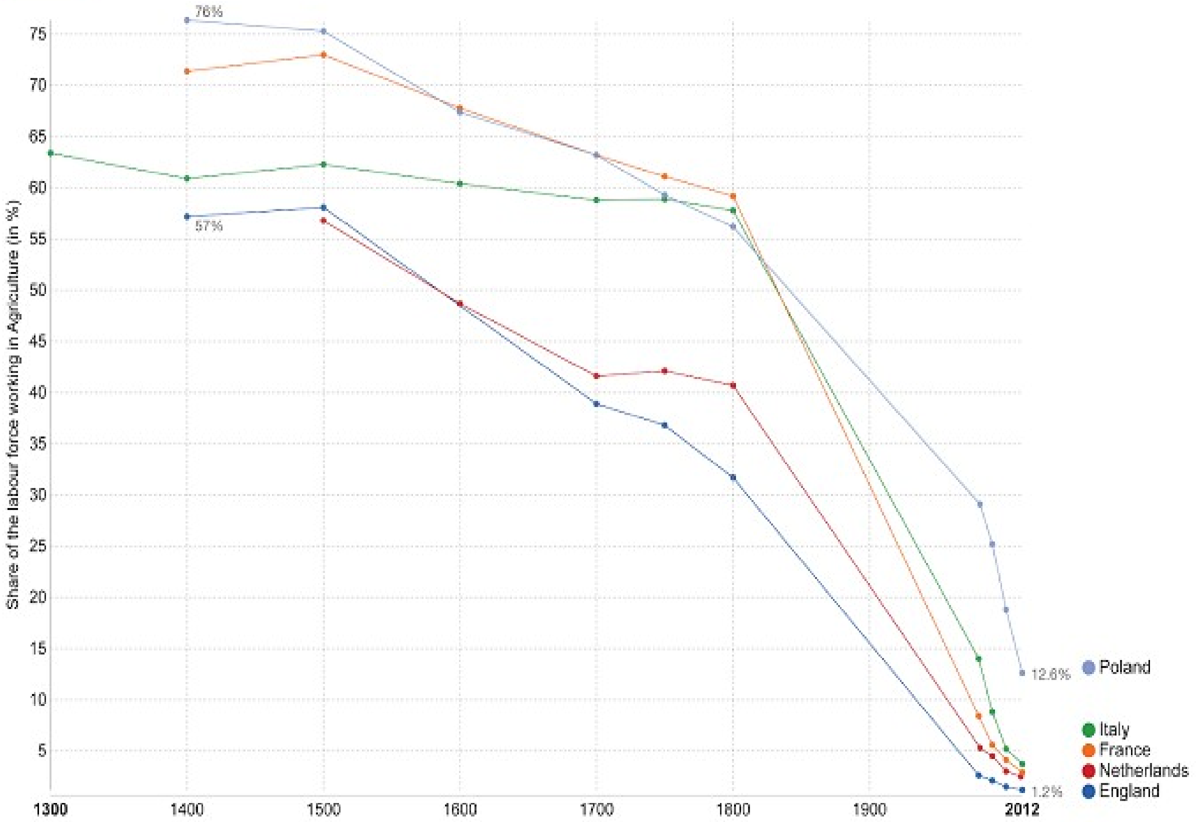
What is happening in technological unemployment today?

Worldwide, employment rates are not worse than historic levels. This seems to show that jobs are being created more-or-less as fast as they are being automated. Scott has covered this issue thoroughly in a survey article. The Bureau of Labor Statistics puts out a report of their projections for what jobs will be lost and what will be added for the next ten years. Some jobs that are expected to grow are in health care, renewable energy, and several computing professions. The ones that are declining include secretaries, clerks, assemblers, and care of outdated tech like locomotives and wristwatches. Automation has tended recently to take middle skill jobs. This has caused many people to take less desirable jobs.

We have seen at least three instances in the past where automation has taken a significant fraction of all existing jobs. Before the neolithic agricultural revolution, nearly everyone was a hunter or gatherer or both. During the period from about 8000-4000 B.C., this gradually shifted so that most people were employed in agriculture. Before 1400, around 70% of all employment was in agriculture. Today, it is only a few percent in advanced economies. At its peak during World War II, nearly 40% of U.S. employment was in manufacturing. Today, that number is below 10%. Housework also declined from 60 hours a week in 1900 to only 15 hours a week today (although this doesn't show up in employment figures, of course, and can also be partly attributed to declining fertility rates.)

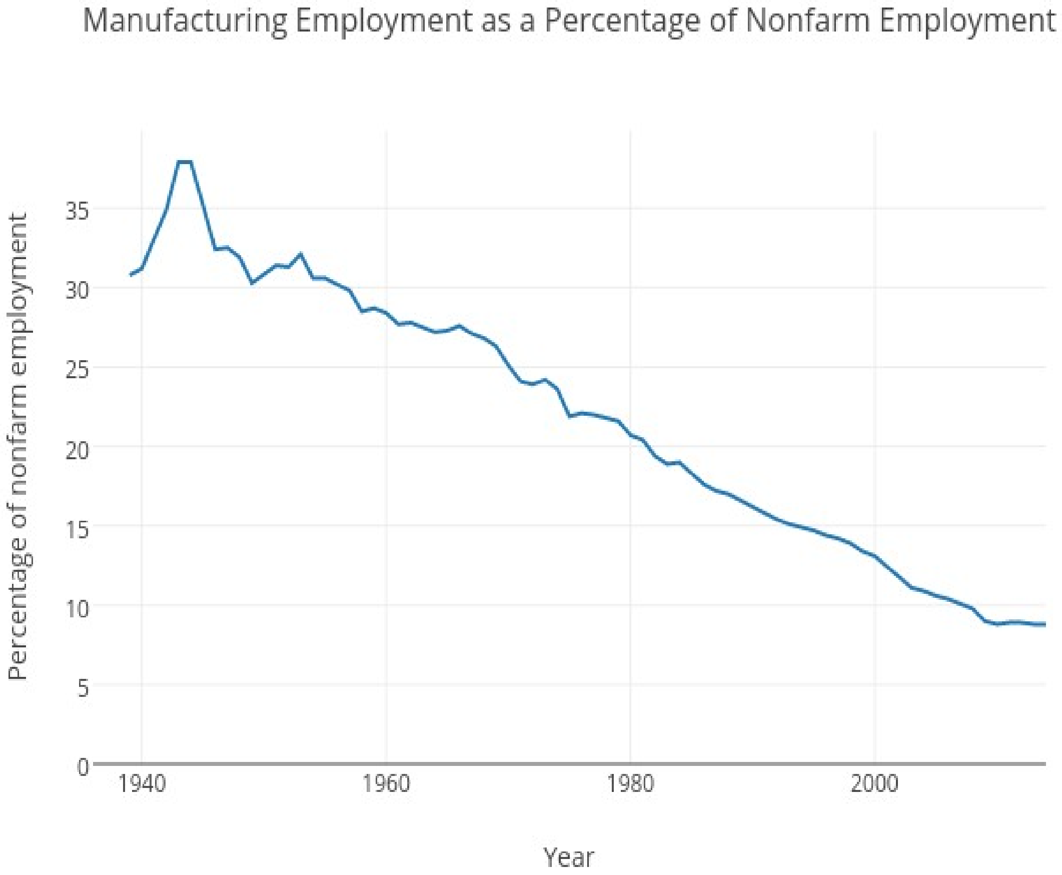
Our World in Data

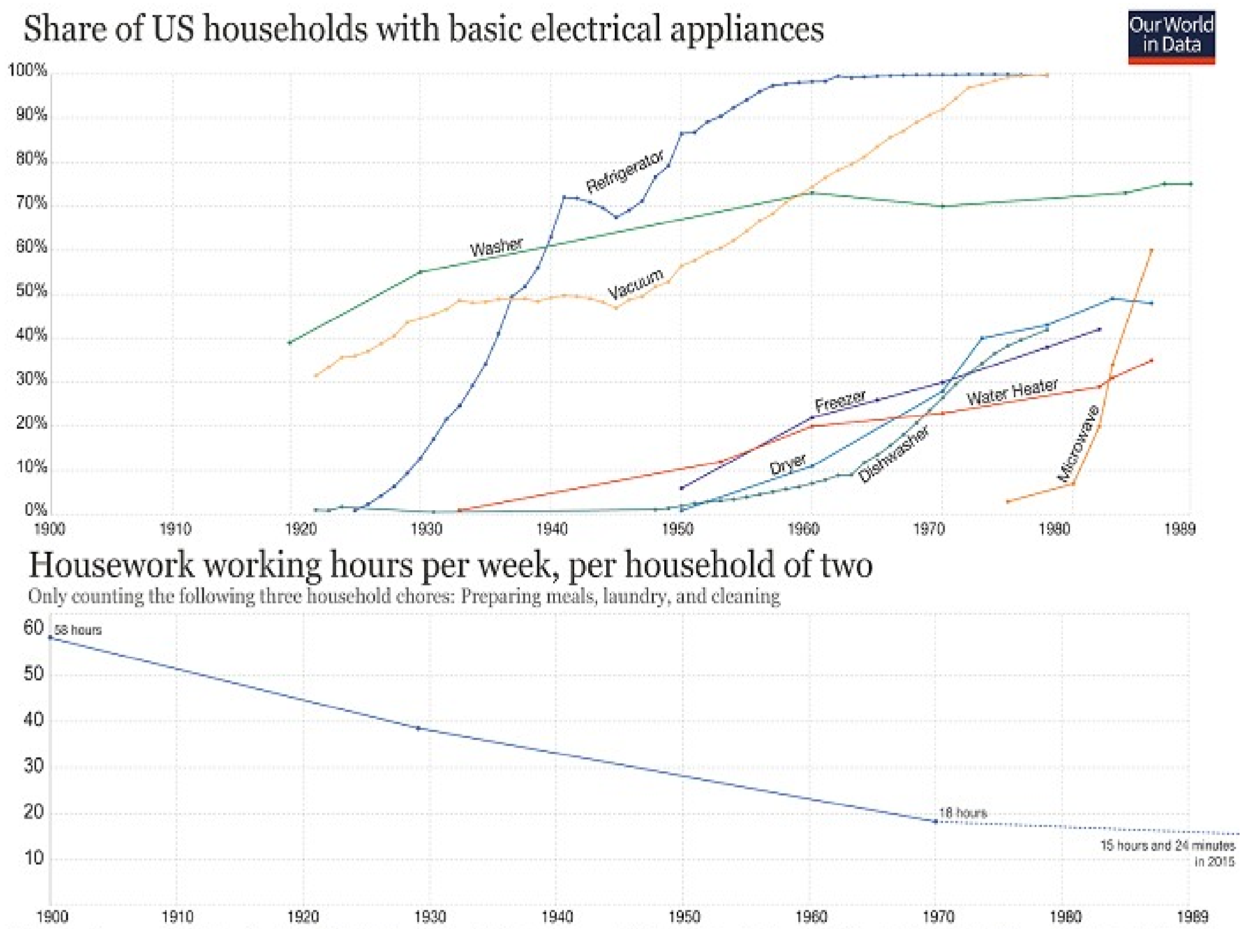
Share of the labor force working in agriculture, since 1300 – By Max Roser



Data source: Pre 1800 is from Allen (2000), "Economic Structure and Agricultural Productivity in Europe, 1300-1800". Newer data from the World Bank.
The interactive data visualization is available at OurWorldinData.org. There you find the raw data and more visualizations on this topic.

Licensed under CC-BY-NC-SA by the author Max Roser.


Manufacturing Employment as a Percentage of Nonfarm Employment
Percentage of nonfarm employment
0
5
10
15
20
25
30
35
1940
1960
1980
2000
Year



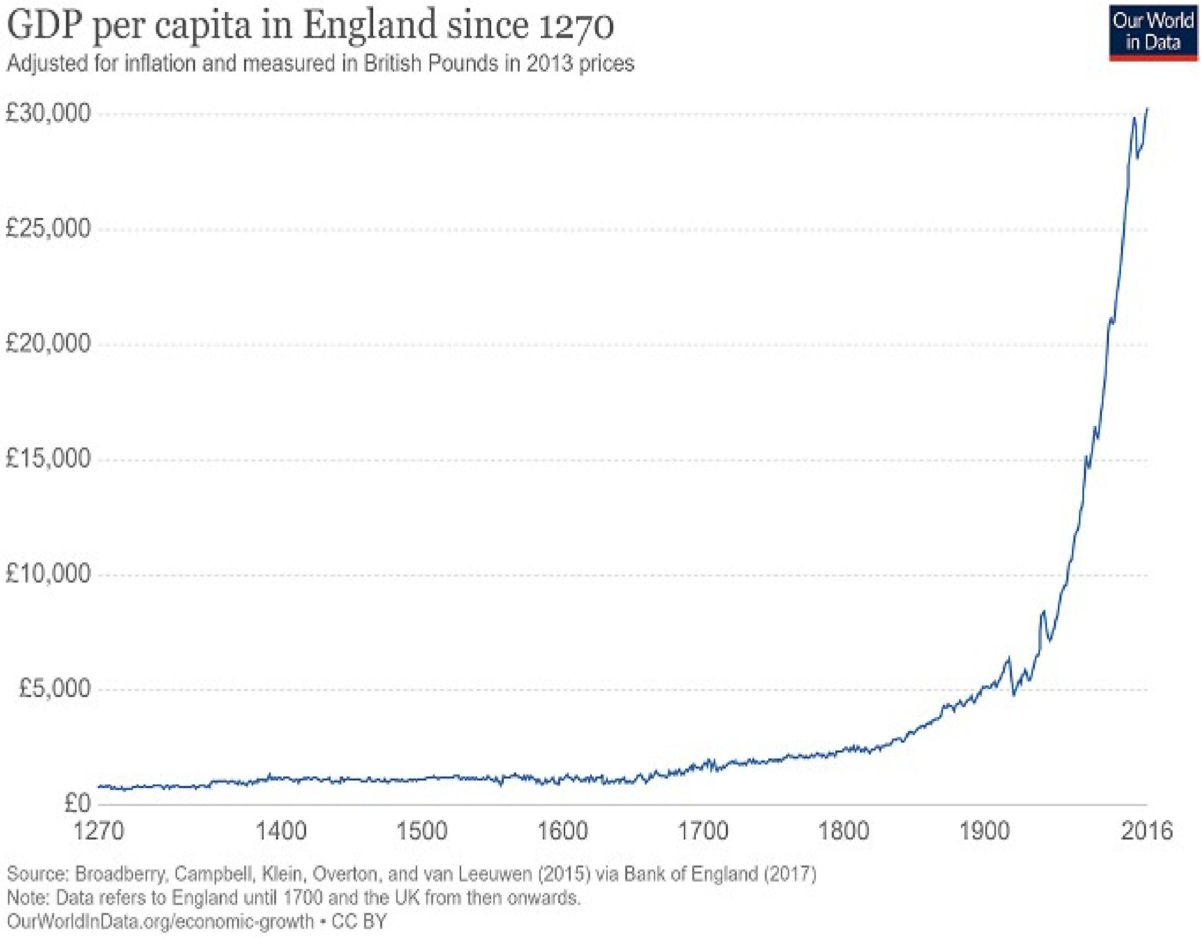
Share of US households with basic electrical appliances

Housework working hours per week, per household of two

Only counting the following three household chores: Preparing meals, laundry, and cleaning

Data source: Greenwood, Seshadri and Yorukoglu (2005) – 'Engines of Liberation', Review of Economic Studies. Except working hours in 2015 from the American Time Use survey conducted by the BLS. The interactive data visualization is available at OurWorldinData.org. There you find the raw data and more visualizations on this topic. Licensed under CC-BY-SA by the author Max Roser.

Another line of evidence to consider is GDP. Before 1000 AD, the per capita GDP everywhere was below $1000, adjusted for inflation. In Western countries today it is around $50000. Since human innate capability hasn't changed, this must be the result of innovations (in education, processes, tools, machinery, or what-have-you) that allow people to produce more value for each hour of work. By one way of looking at it, this means that there are already 49 "robots" worth of automation for every person in these countries. Yet employment is still at roughly the same level it has always been.



GDP per capita in England since 1270

Adjusted for inflation and measured in British Pounds in 2013 prices

Source: Broadberry, Campbell, Klein, Overton, and van Leeuwen (2015) via Bank of England (2017)
Note: Data refers to England until 1700 and the UK from then onwards.
OurWorldInData.org/economic-growth • CC BY

So we know from experience that employment is able to adapt to extensive automation. However, these changes took place over millennia, centuries and decades respectively. If the pace of job automation were to increase so that replacement of a significant fraction of the workplace happened over years instead of decades there might not be time for people to retrain fast enough to avoid some higher rates of unemployment. Again, though, that doesn’t seem to be the case at the moment.

Are there things computers will never be good at?

A common response to the question of whether we will eventually reach the point that all jobs can be automated is to name a skill that computers will never be able to perform, attempting a disproof by counterexample. The following terms link to places a skill was claimed to be impossible for computers: write jokes, write novels, express compassion, robotically navigate a human environment (like an arbitrary kitchen), manufacture new categories at arbitrary levels of abstraction, act creatively, represent and invent concepts, learn from small data, express emotion, have motivational direction, think socially and cooperate.

Researchers are, however, aware of these limitations of current machines and actively trying to find ways to automate them. Here are the same terms along with a link to a paper where research is presented on how to automate the task: write jokes, write novels, express compassion, robotically navigate a human environment, manufacture new categories at arbitrary levels of abstraction, act creatively, represent and invent concepts, learn from small data, express emotion, have motivational direction, think socially and cooperate.

In every case, although researchers haven’t yet solved the problem and in some are far from a solution, it is possible to see a clear research direction and a path to gradual progress. Many claims of this sort are (in the local argot) making a motte-and-bailey argument. When someone argues that a machine can’t really feel emotion, they can always retreat to the motte that the machines are incapable of phenomenal conscious awareness of what it is like to feel (for example) emotional pain. This is true: we have no idea how to make a machine that is conscious in this sense or to test whether phenomenal consciousness is present in a person, animal, or machine. They then, however, make claims that the robots will not be able to respond to anger in a voice, or anticipate that taking an action might cause someone to feel sadness. This is false: a neural network trained on examples of anger in a voice could learn to discriminate it without the ability to feel blood rush to its ears. For the purpose of taking jobs, an accurate discriminator or ability to accurately simulate emotions is all that is necessary.

The new techniques researchers have come up with are not only able to perform better than previous methods on a particular benchmark; they are also becoming more general. Artificial General Intelligence is what human-like AI is usually called these days because the ability to handle unanticipated situations is such a central part of what makes human intelligence special.

The new Transformer neural net architectures are an example of how AI is becoming more general. Although simply trained on predicting the next word in a sequence, such models have demonstrated superior performance on question-answering, common-sense, categorization, and other benchmarks. A similar architecture, with few changes, can be used to compose music, create artwork, simulate voices, and so on. These models work well because they are able to learn to direct attention to the parts of the context most applicable to deciding what the next output should be. In the future, we should expect systems that are more adaptable still. An adaptable AI will be quicker and cheaper to deploy on new jobs, so we should expect the rate at which jobs are automated to increase.

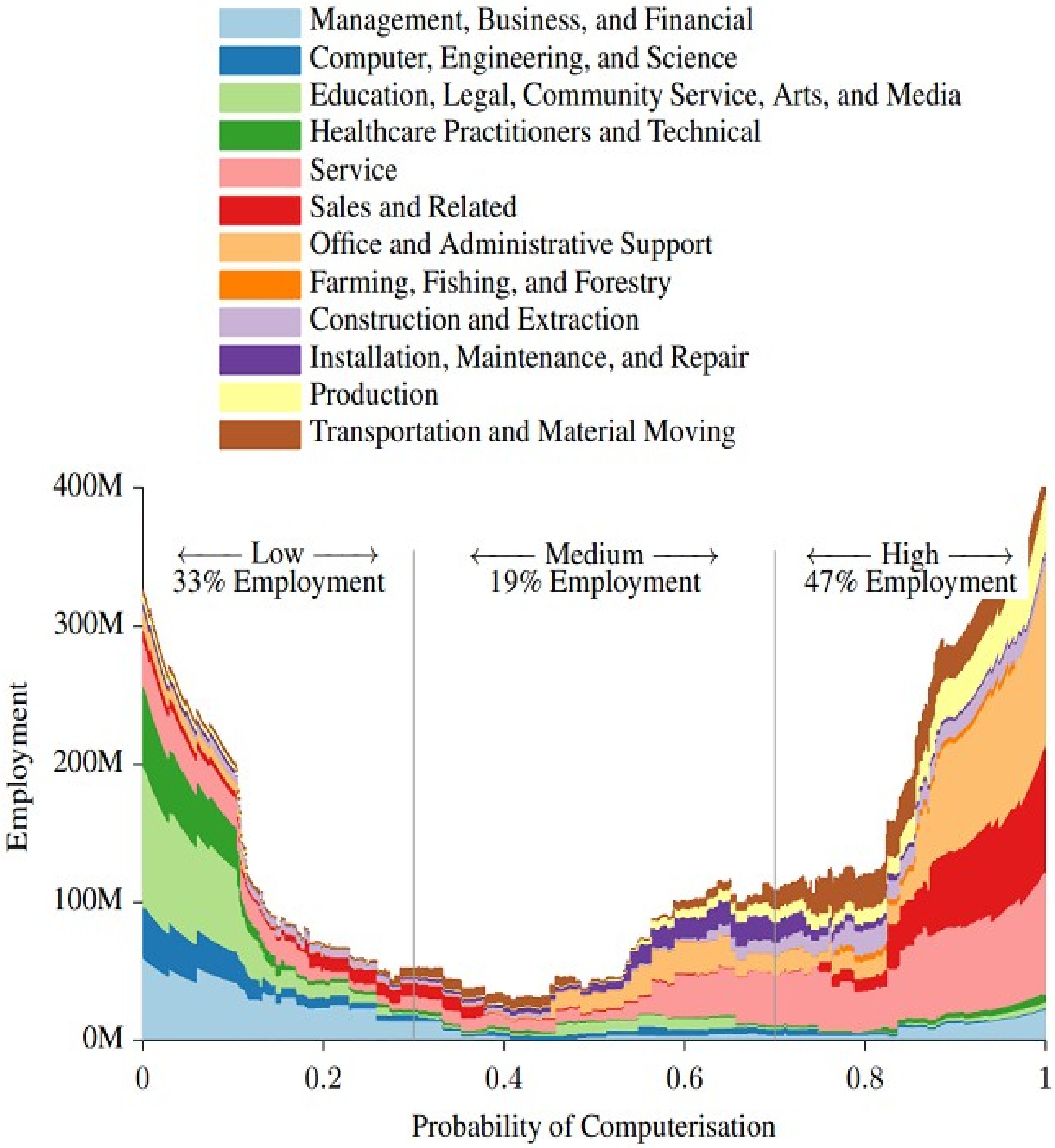
In terms of hardware, we can expect computational capacity on the order of the human brain in supercomputers in the next five years and in home computers about twenty-five years after that. (Assuming 100 billion neurons, 1000 synaptic connections per neuron, 10 floating-point operations per interaction, and a temporal resolution of 1000 interactions per second.) So hardware shouldn't be a limiting factor after 2050 or so, as long as current trends hold.

This is not to minimize how far we are at the moment from a machine that can learn an arbitrary new job as easily as a human can. People can model another human by “putting themselves in their shoes.” All of that ability to anticipate how other humans would react to an action has to be built into machines to achieve the kind of autonomy we are imagining. While machines can now act in creative ways as well as rational ways, tying the two together is still a very open problem. Systems that have lifelong learning, that continue to grow with experience, are still very rare. The ability to understand spoken or written language is still at a very primitive level. These problems don’t seem insurmountable– merely very hard.

What jobs will be automated?

Frey 2013 characterizes which jobs are likely to be automated soonest based on the following capabilities required to perform the job:

– Finger Dexterity, Manual Dexterity, Cramped Work Space, Awkward Positions
– Originality, Fine Arts
– Social Perceptiveness, Negotiation, Persuasion, Assisting and Caring for Others



He concludes that 49% of U.S. jobs are repetitive, don't require fine dexterity, originality or people skills, and are therefore likely to be automated in the next few decades. This includes most office and administrative support jobs, sales jobs, some service jobs, and most production and transportation jobs.

However, Arntz et al argue that this number is much too high. Holding everything else the same, they show that this is neglecting the variation within a profession and the ability for a job to adapt when new technology becomes available. With these taken into account, only 9% of jobs are found to be at risk.

Both of these papers are discussing the right side of the graph above, everything above the “75% probability of computerisation” line. Eventually, though, essentially all the skills described on this chart will be automatable. While it may not fall within our lifetimes, it does seem to be part of the “foreseeable future.” It doesn’t seem like there are any fundamental physical limitations preventing it (in the sense that we may never build a spaceship that goes faster than light.) The existence of human brains shows that the right arrangement of atoms can compute at human levels with reasonable size, weight, and power restrictions. It seems reasonable to suppose that computers will continue to increase in capability until they will be able to perform any intellectual task required in a job as well as a human. This includes creative, decision-making, and emotional reasoning tasks.

To replace people in jobs also requires a body that can perform tasks with the dexterity and ability to adapt to different conditions that are required for a job. This also seems to be at least decades away for many jobs.

Beyond the invention of hardware and software capable of performing these tasks, the cost of developing and deploying the technology must fall below the cost of hiring workers in order for the workers to be replaced. The price of computing has been dropping steadily for several decades now, and there are no fundamental physical limitations to this improvement that would prevent the trend from continuing to the size and power-usage levels of a human brain. Robotic bodies and manipulators, while continuing to improve in dexterity, sensing ability, and cost over time, do not seem likely to have the same exponential improvement that we have seen in computing hardware. Again, though, we know that a machine with human dexterity is possible (because hands exist) so it seems inevitable that machines will eventually surpass us in these abilities as well.

There are, however, certain jobs that some people may be willing to continue to pay for a human to do, even if a robot can do it better in some sense. For example:

– producing handmade goods
– creating artwork whose value depends on whether it is an original
– some kinds of food preparation
– performance arts (acting, dancing, stand-up comedy, etc…)
– domestic service (personal servants like butlers, gardeners, etc)
– sports
– therapy
– hairdressing and the like
– massage
– certain aspects of medical care (a sense that someone cares)
– certain aspects of teaching (motivation, mentorship)
– certain aspects of war (decisions about when to use violent force)
– clergy work
– mortuary services
– some kinds of sales
– politics

For these kinds of tasks, having it done by a human is part of what is valued by some customers. If most other jobs can be automated, more jobs that fall in this category would be expected to be created, as a larger pool of workers is available to do them.

As more jobs are automated, what economic effects should we expect to see?

Around 1800, economist Jean-Baptiste Say argued that workers displaced by new technology would find work elsewhere once the market had had time to adjust. By the mid-1800s, a theory was in place that explored the economic effects of automation. In *Das Kapital* Marx would later dub it "the theory of compensation." This includes additional employment in the capital goods sector, decrease in prices, new investments, and new products (the effect on wages is complicated). In general this is still the prevailing opinion of economists.

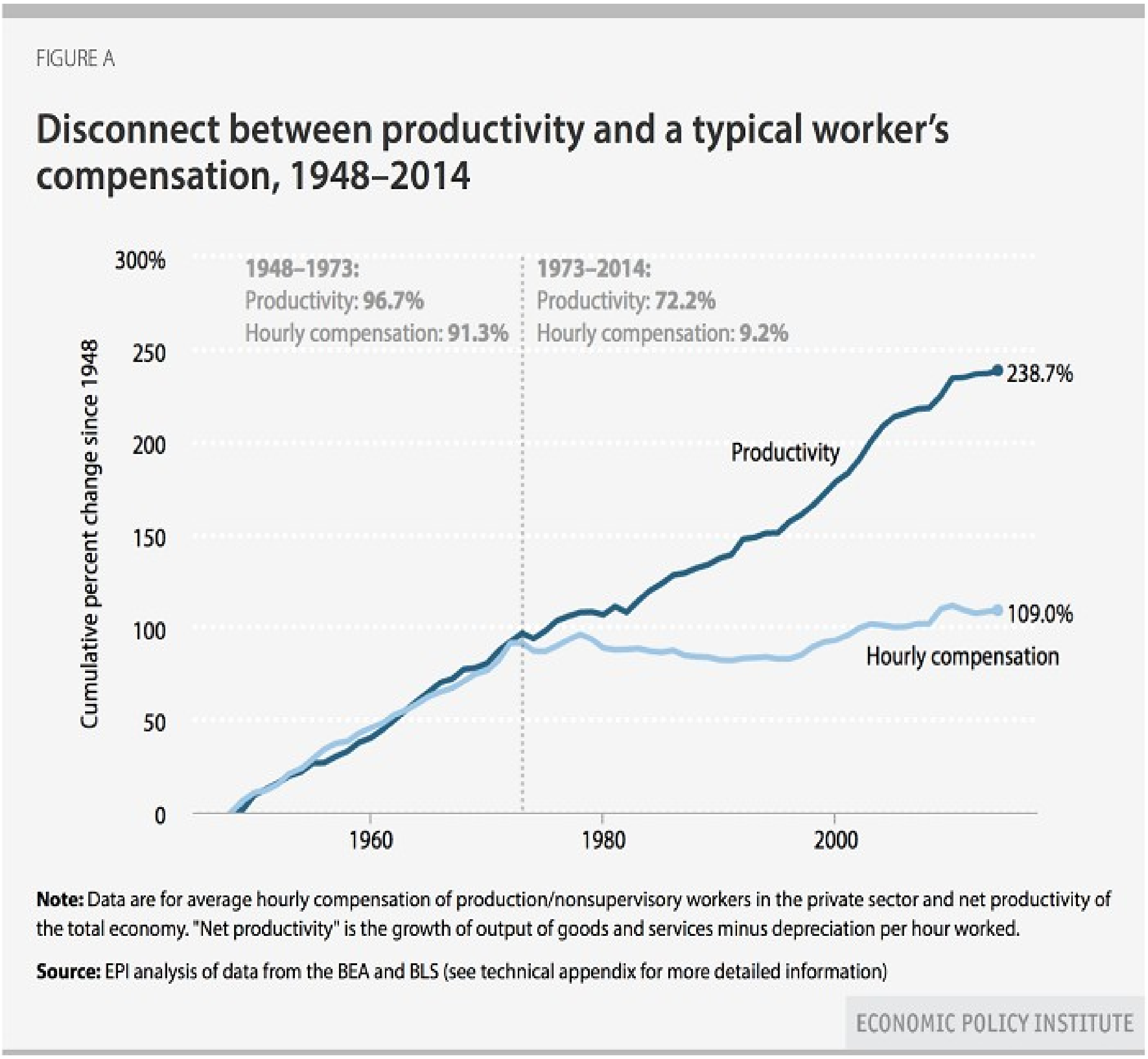
According to this theory, when workers are fired because their jobs are automated, this frees up capital which the owner will then use to hire other workers to do other jobs. Because of this, the number of workers hired doesn't decrease because of automation. (Marx disagreed with this, saying that part of the capital would now be tied up in the machines). The theory also discussed other effects. Automation reduces the prices of goods, making them more affordable. It also reduces the prices of components, making new products viable. The companies making these goods make more profits, allowing them to expand and hire more workers.

Because of these effects, as long as the market has time to adjust, we shouldn't expect to see increasing levels of unemployment up to the point where robots have taken all the jobs. Instead, new jobs for workers should be created until the entire employment pool is being utilized. This process can be expected to continue up until the point that all jobs can be done more cheaply by machine. As long as there exist skills humans can do more cheaply than machines, the number of jobs using those skills should increase until they absorb the entire available human labor pool.

Susskind 2018 concludes that in the future, automation will put downward pressure on wages, while increasing the amount earned by capital owners. We may already be seeing this effect in the United States: although GDP per capita and net productivity have increased consistently since the Great Depression, median wages have stagnated since 1975. This would be consistent with automation sending increases in productivity to the owners of capital rather than workers.

FIGURE A

Disconnect between productivity and a typical worker's compensation, 1948–2014



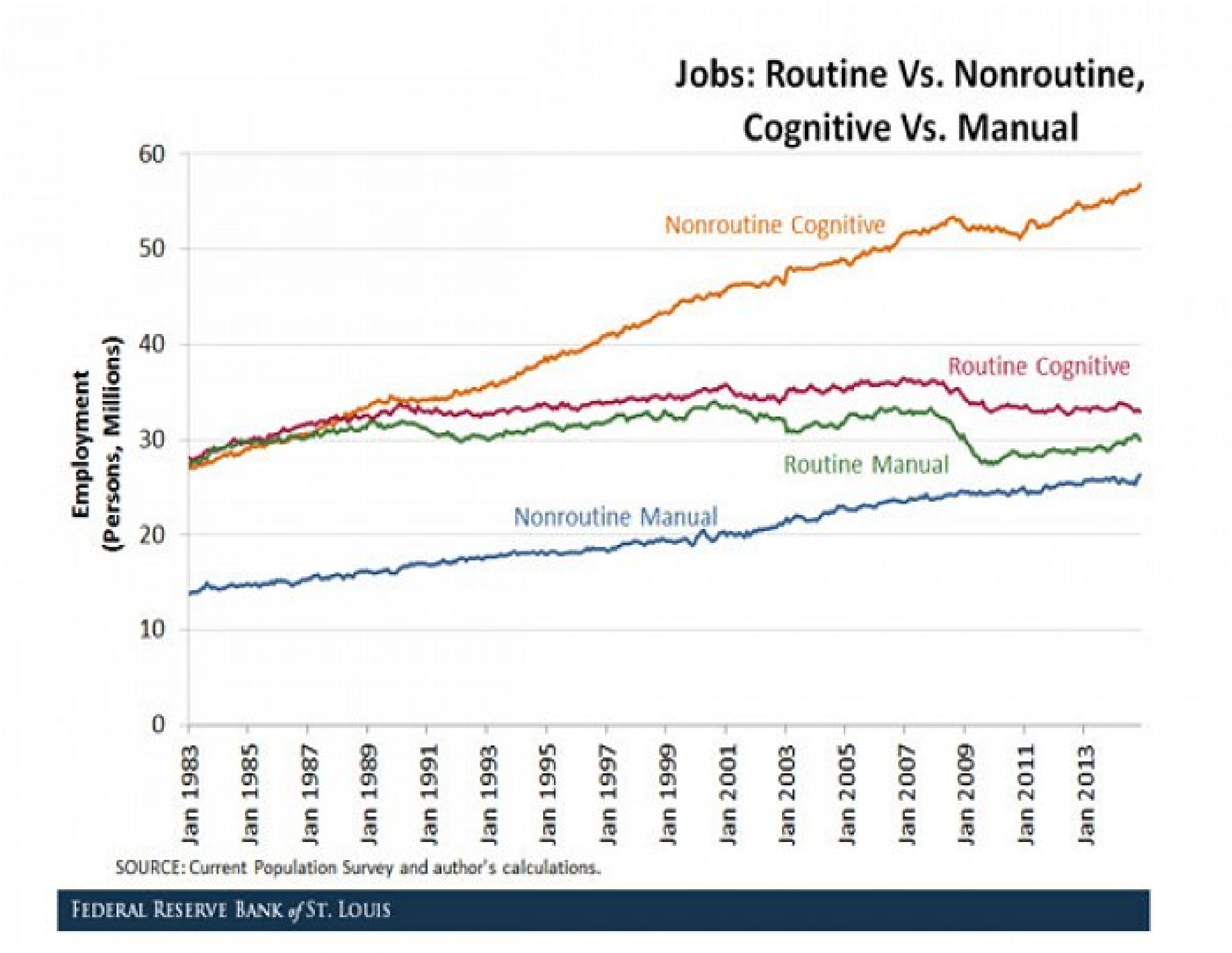
Note: Data are for average hourly compensation of production/nonsupervisory workers in the private sector and net productivity of the total economy. "Net productivity" is the growth of output of goods and services minus depreciation per hour worked.

Source: EPI analysis of data from the BEA and BLS (see technical appendix for more detailed information)

As more jobs are automated, the mean standard of living will improve, as the amount of value produced per-capita goes higher and higher. Even without raising tax rates or rates of giving to charity, the overall amount received will increase as more is produced at lower cost. Whether this leads to more people living on the dole or not is more a matter for political argument than for technological extrapolation.

Another effect might be shorter working hours– as more jobs become automated, the same number of people could be employed, but at fewer hours per week or more days of leave per year. Given the option, though, the preference of most workers at the moment is to work full-time (and for many workers, overtime) trading leisure time for additional income. For this to change would require both regulatory changes (part-time workers have different rules about benefits, for example) and cultural changes. It is not absurd, though: in Germany, for example, the average adult only works 1400 hours a year (26 hours a week) compared with 1900 hours a year (35 hours per week) in the U.S.

One might expect that as machines become more capable, more and more people will find themselves below the waterline, unable to find any job that AI can't do better– perhaps those with the lowest IQ first, or something along those lines. To date, though, the capabilities of AI have not developed this way. Grandmaster chess and rapid calculation can't be done by those with low IQ, but are simple for modern machines, and the inverse is also true– even very young children and those with a low IQ can perform recognition tasks in varying conditions that defeat even the best computer vision programs, for example. On the other hand, the number of routine jobs in the U.S. is an ever declining fraction of all jobs. If some constant fraction of people can only perform routine jobs, eventually some of them will be unable to find any job they can do, if current trends continue.



SOURCE: Current Population Survey and author's calculations.

Suppose we reach the point where robots can do literally any job a human can do. What will happen to the economy?

A robot will likely never be cheap compared to other manufactured goods. Although future process innovations (such as advanced 3D printers or nanotech assemblers) may reduce the cost of building robots, they will also reduce the cost of manufacturing everything else, and robots must have large numbers of moving parts. This could mean that the number of robots will be limited, and this limited supply will drive up wages in jobs that the robots could otherwise do, if there were enough of them or they were cheap enough to produce.

Robots with human levels of ability, however, would be able to self-repair, extract natural resources, manufacture parts and create more robots without any human intervention. They would also be able to invent new ways to make money and employ other machines to achieve goals.

For any job, the machines in this scenario could do it better. People will still likely strive to purchase and direct factories, resource extraction, and robots for all purposes. Those doing this would still have a job of deciding how to direct the robots, acting as business owners. (Although one imagines running an AI as a manager to handle this kind of work as well.) There will also be people with earned or inherited wealth who don’t work, and welfare recipients who don’t work, but to what extent the economy will redistribute the wealth generated by this vastly expanded economy is a political question.

If we ever reach this point though, it is hard to make serious predictions because we don’t know what such machines would be like, and whether we would be able to maintain control of our economy and civilization at all (this breakdown of all models is the reason von Neumann called such an eventuality a singularity). If it becomes possible to create machines with human level intelligence and skills it will be possible for a little more money to create superhuman intelligence and skills, which will necessarily be hard for us mere humans to predict or control.

Conclusion

Even if nearly all currently existing jobs will eventually be automated, as we progress toward that point new jobs will continue to be created for humans, preventing the kind of mass unemployment or low wages that might be expected, as long as the market has time to adjust (which isn’t necessarily the case– if the pace of automating jobs were to speed up enough, we could still see a crisis.) However, once machines surpass human capabilities for a low enough price in all jobs, the entire economy will change and something else will take its place. What that is we can't really say– our economic models break down, and the future becomes even more difficult to predict. Beyond this point, we don't even know to what extent humans are still guiding the course of civilization, let alone how employment will work.

The economic gains that come from all this automation will flow primarily to those who own the machines. As they invest more, create new products, spend more, pay more taxes, and give more to charity, the general civilization will benefit, though some people will doubtless be unable to adapt and find new work and be worse off. How we choose to provide for those that can't find work is something each democracy will need to continue to decide.

Further Reading

I found this article on Wired had some good points. One of them was that 'job churn' is at historic lows. That's the rate of creation and destruction of jobs. You would expect that to go up if the rate of job automation was increasing.

The paper Automation and New Tasks: How Technology Displaces and Reinstates Labor provides a reasonable framework for estimating current and future effects of automation on labor. They conclude "if the origin of productivity growth in the future continues to be automation, the relative standing of labor, together with the task content of production, will decline. The creation of new tasks and other technologies raising the labor intensity of production and the labor share are vital for continued wage growth commensurate with productivity growth."

If you are interested in specific predictions about dates for developments in AI, ML and robotics, MIT roboticist Rodney Brooks has a blogpost.

AI researcher Stuart Russell's new book Human Compatible provides a nice introduction to current thinking about the future development of AI. It also contains some interesting ideas about creating artificial intelligence with open objective functions, so that the AI wants to please people but isn't sure how best to do so.

[ACC] Should You Have A Merry Christmas?

[This is an entry to the 2019 Adversarial Collaboration Contest by Cindy Lou Who and the Grinch]

Christmas Day is a a time full of laughter and cheer which is held in the West at the end of each year.

Believers in Jesus traditionally think the day marks his birth; scientists disagree. They point to the shepherds; when carolers sing about fields full of sheep, that occurs in the spring. The Star of the Magi provides further doubt. Simulations can tell us what star it's about: it was most likely Jupiter shining near Saturn, but it's only in autumn one sees such a pattern. It is proven in space and it's proven on Earth – Christmas isn't the real time of Jesus' birth.

One of the most popular Yule celebrations is handing out gifts to one's friends and relations. Parents offer the story these presents appeared due to Santa, a jolly old man with a beard. Originally a historical saint, his tale was embellished, with little restraint. He flies through the air in a reindeer-pulled sleigh, and visits all households on Earth in a day. This tradition seems pagan, with some scholars noting the details are pulled from a legend of Odin. Though sources like NORAD appear to support Santa's presence, we think that their data fall short. After reading the pros and the cons, we both feel the consensus perspective is Santa's not real.

And what are these gifts' economics effects? According to Goeddeke and Birg, it's complex. Since presents are valuable, one might assume that their giving would cause stores and markets to boom. You give to your parents! You give to your boss! But economists say it is all deadweight loss. You would spend the same money on *something*, you see, and presents are chosen incompetently. Others' preferences aren't as clear as our own, so when we buy for others, their needs are unknown. Presents don't increase welfare and don't increase growth; all the papers agree they are harmful to both.

Is there anything good about Yule? It depends. Holidays are a time to see family and friends. In a season of darkness and inclement weather, Christmas gives an excuse to bring people together. People pray for world peace; relatives are united; workers get time off work; little kids are delighted. It gives us a reason for dancing and singing, for candles and carols and feasts and bell-ringing. Despite deadweight losses, despite the wrong days, we think that this matters so much it outweighs all the falsehoods and problems; our model predicts it outweighs all those things by a factor of six!

CONCLUSION (CLW): The Grinch turned out right on historical dates, and made several good points about gifts (which he hates). But if measures and numbers are all we inspect, then we risk falling prey to the streetlight effect. It was useful to reason this out with the Grinch, but I chose not to budge; I did not budge one inch! The best parts of Christmas can hardly be measured; they're moments of joy to be felt and be treasured.

CONCLUSION (TG): I started out thinking that Christmas was bad. I hated the season! It made me so mad! But the evidence reached an explicit conclusion: my hatred of Christmas was based on confusion. It wasn't that Christmas was harmful at all! It was just that my heart was two sizes too small!

CONCLUSION (JOINT): After working our differences out, we agree: Merry Christmas to all those who read SSC!

[ACC] How Much Significance Should We Ascribe To Spiritual Experiences?

[This is an entry to the 2019 Adversarial Collaboration Contest by Jeremiah Gruenberg and Seth Schoen]

1. Introduction

This project seeks to explore the viability of spiritual or religious experiences as empirical evidence for a component of reality that transcends or is radically different from our ordinary experience. The question at hand is not the existence of God or higher powers, nor the failures, successes, or benefits of religion, but rather the role of spiritual experience in the human understanding of the nature of reality. We formulated the topic in controversy this way:

> The empirical study of the content and nature of people's personal spiritual experiences justifies taking them seriously as evidence of an important component of human life deserving of individual and collective exploration.

Our fellow human beings have always had unusual experiences that they found special and meaningful, but often struggled to interpret or place in the context of their ordinary lives. These experiences and their interpretation have aroused intense controversy, both because people have deployed them as support for their views on contested issues about the nature of reality, and because they may arise in settings where one could easily question whether the brain's altered perceptions and understandings are enhanced or impaired. Another source of debate is how radically different individuals' experiences—and their personal interpretations of the origins and meanings of those experiences—can be. Finally, spiritual experiences are often reported through a cultural lens that leads to questions about how accurately and objectively people could perceive and describe the unusual things that they perceived.

We emphasize that there is no question, even from the most skeptical perspective, of insisting that individuals alter their own views or memories of what they have witnessed (although we encourage people to question their interpretations and to become aware of factors that could raise doubts about those interpretations). What is rational or plausible for each person to believe at a particular moment can be different, and in any case the way that people interpret their own experience and history will be different. If you have had a spiritual experience whose nature and meaning you find evident and certain, others may offer you alternative interpretations and evidence against your view, but can't demand that you change it. However, we find it interesting to consider what lessons *others* can draw from accounts of unusual experiences and perceptions: not so much what sort of evidence your own spiritual experiences may constitute for you, but rather what sort of evidence your accounts of them may constitute for others. Can we collectively learn anything from these experiences?

One objective of this project is to explore empiricism as a key to a “common language” which allows all perspectives to discuss the significance of spiritual experience—not just those who are predisposed to a traditional theistic model of reality. Empiricism seems to a major contender in the competition to find common ground surrounding spirituality. It is both experience-based and rational. Properly employed, the use of empiricism may allow for a rational discussion of personal experience.

We’ve structured this article in nine sections:

1. *Introduction*.
2. *Definitions of Empiricism, Experience, Knowledge, and Spirituality*: a discussion of some important terms, as well as the coherence and conflict between empiricism and rationality.
3. *Psychological Research on Spiritual Experience*.
4. *Epistemology and Religious Experience*: this section focuses on William P. Alston’s treatment of how mystical perception may justify the generation of personal belief.
5. *Near Death Experiences*: a review and discussion of a major work on the significance of NDEs from the perspective of a scholar who suggests that their meaning is largely symbolic.
6. *The Use of Entheogens*: a look at a recent meta-study which reviews the data of five different studies on spiritual experiences resulting from the use of entheogenic substances, and some other sources on entheogens in religion.

7. *The Problem of Dreaming*: an objection to the interpretation of spiritual experiences as having anything other than personal, momentary significance.

8. *Some Possible Perspectives*: the collaborators share their various ideas on how the empirical and philosophical content of the paper thus far might be viewed or understood.

9. *Concluding Thoughts*.

2. Definitions of Empiricism, Experience, and Spirituality

The first step in exploring this statement is interact with philosophical perspectives on the interrelated concepts of *empiricism*, *experience*, and *spirituality* in the attempt to define our terms.

Empiricism can be various defined. Essentially it deals with sensory experience as a generator of knowledge.

Empiricism was an integral—perhaps the singular fundamental—of the emergence of scientific inquiry as we know it. Wolfe and Gal write:

> It was in 1660s England, according to the received view, in the meetings of the Royal Society of London, that science acquired the form of empirical enquiry that we recognize as our own: an open, collaborative experimental practice, mediated by specially-designed

> instruments, supported by civil, critical discourse, stressing accuracy and replicability. Guided by the philosophy of Francis Bacon, by Protestant ideas of this-worldly benevolence, by gentlemanly codes of decorum and integrity and by a dominant interest in mechanics and a conviction in the mechanical structure of the universe, the members of the Royal Society created a novel experimental practice that superseded all former modes of empirical inquiry – from Aristotelian observations to alchemical experimentation.[1]

However, it is important to note that empiricism was popularized as a philosophical concept in the first half of the 20th Century (by such figures as A.J. Ayer, Rudolf Carnap, Kurt Gödel, Karl Popper, Hans Reichenbach, and Ludwig Wittgenstein), and began to take various shapes. Since then, there have been major disagreements on exactly what empiricism entails and how it functions.[2] However, the *Stanford Encyclopedia of Philosophy* states: “Since antiquity the idea that natural science rests importantly on experience has been non-controversial.”[3]

Perhaps most germane to the topic of this investigation is the fact that empiricism conflicts with pure rationality. It is easy to recognize the limitations and failings of human experience as a reliable source of truth.[4] However, since all inputs to human cognition are fundamentally experiential in nature (e.g. the senses of sight, hearing, etc.), the issue of experience must be addressed in any epistemological mode.

In reviewing this conflict between rationalism and empiricism, Markie summarizes the empiricist position with this thesis: “We have no source of knowledge in S or for the concepts we use in S other than sense experience.”[5]

On the other hand, Markie provides three theses which summarize the rationalist position:

> *The Intuition/Deduction Thesis*: Some propositions in a particular subject area, S, are knowable by us by intuition alone; still others are knowable by being deduced from intuited propositions.
>
> *The Innate Knowledge Thesis*: We have knowledge of some truths in a particular subject area, S, as part of our rational nature.
>
> *The Innate Concept Thesis*: We have some of the concepts we employ in a particular subject area, S, as part of our rational nature.

Taking these views into consideration, it seems that the human being is still left in an empiricist position in that our existential state precludes non-experiential data gathering. Intuition itself is formed by lifelong experiences. Nor is “our rationalist nature” is as purely rational as we might hope. Certainly, we should temper the negative subjective qualities of experience, but it seems impossible to circumvent experience altogether regarding embodied human epistemology.

It is true that our experiences must be tempered with objective rationality. However, most humans naturally function primarily in an empirical manner in the formation of worldview, beliefs, and knowledge. The relevant question is therefore not whether some form of empiricism is at play in epistemology in general, but what its role should be. It seems the pure rationalist would exclude all subjective experiential sources of knowledge, even if humans naturally engage in—and rely on—such subjective experiential sense-making.

When it comes to issues of spirituality, a materialistic presupposition would immediately dismiss all appeals to experience. Such a presupposition precludes any engagement with the metaphysical due to its (supposed) nonexistence. A committed materialist would not even investigate the possibility that spiritual experiences have anything but a neurological/biochemical cause. However, if such a presupposition may be suspended, empiricism may hold the key in explaining and/or understanding a spiritual worldview.

What then constitutes "spirituality" in this conversation? The authors of this collaboration would include such experiences as meditative suspensions of self, encounters with the divine in any religious context, near-death experiences, and transcendental uses of entheogens[6]. Spirituality may be theistic, or it may not. Examples of theistic spirituality are easy to come by. An example of a non-theistic approach to spirituality is found in Sam Harris' *Waking Up*, which advocates the use of meditation derived from Buddhist practices to attain altered states of consciousness. While we (the collaborators) might individually define spirituality somewhat differently on an individual level, we find this more general approach helpful to foster conversation on the empirical nature of spiritual experience.

Zinnbauer argues that the terms *religion* and *spirituality* are very similar in meaning, but that *religion* is a narrower term as it is limited to a traditional or institutional context. Zinnbauer writes:

> Thus, according to these definitions, spirituality is a broader term than religiousness. Spirituality includes a range of phenomena that extends from the well-worn paths associated with traditional religions to the

> experiences of individuals or groups who seek the sacred outside of socially or culturally defined systems. For example, an individual’s spirituality may include feelings of devotion, memories of a mystical experience, gatherings with other seekers, rebellion against a culture antagonistic to such a search, and a sense of unity with all sentient life. Significant changes in any of these levels or developmental strands may change the search itself. Development of a serious illness, for example, may change feelings of devotion to confusion or anger, make gatherings more difficult to attend, and cause psychological isolation from a sacred connection to others.[7]

Pargament provides a slightly different contrast between the two:

> In short, spirituality is highlighted as a distinctive dimension of human functioning in the.... Spirituality alone addresses the discovery, conservation, and transformation of the most ultimate of all concerns, the sacred. Yet religiousness is not viewed as inconsistent with or an impediment to spirituality. In fact, spirituality is the core function of religion. Indeed, considerable religious energy is dedicated to helping people integrate the sacred more fully into their pathways and destinations of living. But to succeed at this task, religion accepts and attempts to address the full range of human strivings. Thus, as defined here, religiousness represents a broader phenomenon than spirituality, one that is concerned with all aspects of human functioning, sacred and profane.[8]

However, Zinnbauer and Pargament note that culturally, spirituality seems to be supplanting religion in a few ways.

Spirituality is now seen as the encompassing “sacred or existential goals in life, such as finding meaning, wholeness, inner potential, and interconnections with others....”[9] They continue: “In contrast, religiousness is substantively associated with formal belief, group practice, and institutions.”[10]

3. Psychological Research on Spiritual Experience

We were impressed by the existence of numerous empirical psychological studies of spiritual experience. Two recent major works which reveal this breadth of research are *The Psychology of Religion: an Empirical Approach* by Hood, Hill, and Spilka[11], and *Handbook of the Psychology of Religion and Spirituality*, edited by Paloutzian and Park[12]. The significant earlier studies of spiritual experience include those of Harvard psychologist William James (*Varieties of Religious Experience: A Study in Human Nature*) and Alister Hardy (*The Spiritual Nature of Man: Study of Contemporary Religious Experience*). “Both James and Hardy affirmed the evidential value of religious/spiritual experiences as at least hypotheses suggesting the existence of a transcendent reality variously experienced.”[13] (James famously suggests that spiritual experiences are difficult to understand or evaluate, but that they are compelling and widespread enough to “forbid a premature closing of our accounts with reality.“)

Paloutzian and Park recognize the limitations of psychological research on the nature of spiritual experience in this way:

> our job as scientific psychologists of religion is to create good theory to explain religiousness in a way that allows the theory to be assessed against evidence. This means

> ideas about possible causal factors that are not, in principle, capable of being tested against evidence may be interesting, but they do not meet the criteria necessary to bear upon our theory construction process.[14]

Current Research

Surveys indicate that somewhere between one third to one half of the population has had some sort of significant religious experience.[15] Such experiences are correlated with gender, education, and social class—being more common for females, for those with a higher education, and for those in higher classes.[16] Hood et al. write: “Women report more such experiences than men; the experiences tend to be age-related, increasing with age; they are characteristic of educated and affluent people; and they are more likely to be associated with indices of psychological health and well-being than with those of pathology or social dysfunction.”[17] Investigations into the heritability of religiosity, particularly through twin studies, place it between 0% to 50%.[18]

It was also discovered that people in the United States, Australia, the United Kingdom, and Scandinavia do not tend to share their spiritual experiences with others. Hood et al. wonder if this is why such spiritual experiences are thought to be uncommon (as fewer people in these societies might have heard reports of others’ spiritual experiences).

Due to the internal and personal nature of spiritual experiences, the data gathering on the topic is most often accomplished through surveys, questionnaires, and interviews. The accuracy of such self-reporting can be measured—with some testing suggesting that some percentage are likely false positives (tendencies both for the pro-religious to affirm spiritual experience and for the anti-religious to deny spiritual experience).[19] Identifying activity which triggers spiritual experience and studying individuals undertaking these actions in a laboratory setting is another research mode. Such activities studied include prayer and meditation.[20] As we will discuss in another section, they also increasingly include the use of psychoactive substances.

Hood et al. admit that the over-reliance on self-reporting is a difficult hurdle in the study of spiritual experience, particularly because of the potential pitfalls of bias, including “intentional deception, impression management, personal bias, and many more.”[21] They discuss the alternative uses of other kinds of measurement, such as physiological and behavioral measures, and the increasing use of the Implicit Association Test.

One study indicated that spiritual experience (reading Psalm 23) functioned neurologically in the frontal and parietal lobes, while nonreligious experience of reading Psalm 23 involved the amygdala (which was not active in the religious experience).[22] “On the basis of these findings, Azari and her coworkers have proposed that religious experience is likely to be a cognitive process utilizing established neural connections between the frontal and parietal lobes.”[23]

Addressing whether spiritual experiences are the result of a psychiatric disorder, Hood et al. note that “both normal and psychotic individuals can have mystical experiences” and that this is backed up by empirical research.[24] Such research noted that the differences between normal mystics and psychotics is that “The psychotic mystics exhibited resistance and rigidity, as opposed to the normal mystics, who exhibited openness and fluidity. Thus it is not simply mystical experience, but the reactions to the experience, that distinguish psychotic from normal mystics."[25]

The Nature of Spiritual Experience

Some evolutionary psychologists argue that religious experience arose because it is an adaptive advantage. Kirkpatrick writes, "Hypotheses about the adaptive function of such religious instincts have ranged from defense against fear of death or other forms of comfort and anxiety reduction to group-level benefits such as promoting cohesion and solidarity or reducing conflict."[26] However, Kirkpatrick argues that the evidence does not point to adaptive advantage. He writes, "My own view…is that the diverse collection of phenomena we refer to as 'religion' represent a collection of by-products of numerous adaptations with other specific, mundane functions."[27] Kirkpatrick continues,

> With respect to religion, beliefs about the existence of supernatural forces and beings appear to emerge as a spandrel-like by-product of evolved systems dedicated to understanding the physical, biological, and interpersonal worlds (Boyer, 1994, 2001). For example, an evolved agency-detector mechanism, designed to distinguish animate from inanimate objects in the world, can be fooled fairly readily to produce psychological animism and anthropomorphism (Atran, 2002; Atran & Norenzayan, in press; Guthrie, 1993), as when we find ourselves cursing at our aforementioned computer when it crashes. Once these spandrel-like effects enable ideas about gods and other supernatural beings, I have suggested, specific forms of religious belief emerge as by-products of psychological mechanisms dedicated to processing information about functionally distinct kinds

> of interpersonal relationships—attachments, kinships, dominance and status competitions, social exchange relationships, friendships, coalitions, and so forth—that whir into action to shape specific beliefs and expectations about these beings and guide behavior toward them. Thus, for example, gods might be perceived as attachment figures, dominant or high-status individuals, or social exchange partners, with each possibility leading to a different set of expectations and inferences about those gods' behavior and decisions about how to best interact with them—processes emerging from functionally distinct psychological systems designed to solve such adaptive problems in human relations (Kirkpatrick, 1999, 2005).[28]

Theories related to agency detection are popular with religious skeptics; some have noted that in a dangerous world of predation or intergroup violence, wrongly failing to perceive agency and intelligence where they are present has greater adverse survival consequences than wrongly perceiving them where they are not present.

Neuropsychologists, however, point to a combination of "cognitive operators" in the brain which together give rise to human religiosity. "The term 'cognitive operator' simply refers to the neurophysiological mechanisms that underlie certain broad categories of cognitive function. Thus, these operators do not exist in the literal sense, but can be useful when considering overall brain function."[29] The "causal operator" works with any series of perceived events and attempts to organize them back to an original cause. Such perception and organization is subjective and may not arrive at an accurate conclusion. Newberg and Newberg write: "We have proposed that when no observational or "scientific" causal explanation is forthcoming for a strip of reality, gods, powers, spirits, or some other causative construct is automatically generated by the causal operator (d’Aquili & Newberg, 1997).”[30] A second operator proposed as functioning in the development of spiritual experience is the “holistic operator”: “The proposed holistic operator permits reality to be viewed as a whole or as a gestalt, as well as the abstraction from particulars or individuals into a larger contextual framework.”[31]

How to View Religious/Spiritual Experience

The following quotes are summaries of current positions of psychologists studying religious and spiritual experience. They represent foundational views in the field and we have chosen therefore to quote them at length.

Hood et al. warn against the danger of reductionism in the psychological study of spiritual experience, stating that it is dangerous “to reduce the richness and complexity of religious experience to a favorite psychological construct.”[32] Elsewhere, they write:

> The empirical study of religiousness has many great challenges. The first of these challenges considered here is how to maintain the scientific standards of good empirical work, always the goal of science, without sacrificing the richness and depth of the object of study. We have gone to considerable lengths to make the case that religious experience should not be reduced to specific psychological processes. It is tempting to do so when one adopts the naturalistic perspective that underlies scientific investigation, and to ignore the meaning system of the people being studied. What is needed is some nonreductionistic accounting of the phenomena of interest, but without abandoning

> scientific methodology and thus not reaping the benefits that it provides.[33] (Hood and Hill Psychology of Religion, 25)

Zinnbauer and Pargament write:

> A controversy that often is raised in discussions of measurement and definition is that of reductionism, the process of understanding a phenomenon at one level of analysis by reducing it to presumably more fundamental processes (see discussions in Idinopulos & Yonan, 1994, and Wilber, 1995). In some sense this process is unavoidable in scientific study (Moberg, 2002; Segal, 1994). However, reductionism is often accompanied by a loss of information. For example, the reduction of mystical experiences of oneness with the universe to a change in neurotransmitter levels eliminates information at all other levels (e.g., the cultural, social, familial, affective, cognitive, and behavioral). There may indeed be important physical correlates of such an experience, but to deny the relevance or value of other modes of interpretation and understanding is to commit the error of reductionism.[34]

Regarding the problem of the compatibility of empiricism with the notion of a genuine spiritual reality, Hood et al. write:

> Although social scientists cannot confirm any ontological claims based upon mystical experience, they can construct theories compatible with claims to the existence of such realities. Hodges (1974) and Porpora (2006) have argued that the scientific taboo against the supernatural can be broken, as long as hypotheses about the supernatural can be shown to have empirical consequences. In Garrett’s (1974) phrase, “troublesome

> transcendence” must be confronted by social scientists as much as by theologians and philosophers.[35]

Newberg and Newberg write:

> Western society has historically emphasized the importance of causality, technological advances, and empiricism. It is from these values that Western medicine, psychiatry, and psychology have developed. We propose that regardless of the connotation of the concept of spirituality in Western society, mystical and meditative experiences are natural and probably measurable processes that are and can be experienced by a diversity of people of different races, religions, and cultures. Those having spiritual experiences can have a variety of neuropsychological constitutions. In addition, it is important for clinicians to be sensitive and knowledgeable regarding spiritual and philosophical beliefs (Worthington, McCullough, & Sandage, 1996). Professionals need to be capable of distinguishing normal, healthy spiritual growth from psychopathology. We hope that some of the neurophysiological analysis described above might allow for a distinction between “normal” spiritual experiences and pathological states. In fact, such a nomenclature may be valuable for future psychological analysis of religious experiences. However, the fact that spiritual experiences have an effect on autonomic function as well as other cortically mediated cognitive and emotional processes suggests that such experiences not only affect the human psyche, but also can be carefully crafted to assist in the therapy of various disorders. It has already been shown that prayer and meditation can improve both physical and psychological parameters (Carson, 1993; Kabat-Zinn, Lipworth, & Burney, 1985; Kaplan, Goldenberg, & Galvin-Nadeu, 1993; Worthington et al., 1996). The

> more the underlying neurophysiological correlates of spiritual experiences are understood, the more such experiences can be analyzed and utilized in clinical practice. Therefore, spiritual experience can be very useful in clinical psychological and psychiatric practice. Furthermore, clinicians themselves can be instrumental in helping their patients toward personal and spiritual growth by discussing various meditative and/or spiritual practices and encouraging patients to approach these practices in an unambiguous manner. According to Rowan (1983), a humanistic psychologist, it is the self that is the missing link between the psychological and the spiritual. Therefore, it seems natural that spiritual experiences, such as those encountered in meditation and prayer, could become an adjunct to Western therapeutic practices and that developing oneself spiritually can become an important part of psychosocial as well as neuropsychological development.[36]

Hood et al. continue:

> There is no reason why scientists cannot include specific hypotheses derived from views about the nature of transcendent reality in empirical studies of religious experience, as long as specific empirical predictions can be made. The source of the predictions may reference even the unobservable and the intangible. All that is required is that there be identifiable empirical consequences. As Jones (1986) has stated the case, Invoking Occam’s Razor [i.e., the philosophical principle that the best explanation of an event is the simplest one] to disallow reference to factors other than sensory observable ones is question begging in favor of one metaphysics building up an ontology with material objects as basic. (p. 225) Jones echoes the classic claim of William James that mystics base their experience

upon the same sort of processes that all empiricists do—direct experience. James would restrict the authoritative value of mystical experience to the person who had the experience, but would view it as a hypothesis for the social scientist to investigate (Hood, 1992a, 1995c). However, mystics are united in the belief that such experiences are real, and many nonmystics are convinced of the reality of the experience even if they personally have not had it. Thus, as Swinburne (1981) argues, mystical experience is also authoritative for others:

> . . . if it seems to me I have a glimpse of Nirvana, or a vision of God, that is good grounds for me to suppose that I do. And, more generally, the occurrence of religious experience is prima facie reason for all to believe in that of which the experience was purportedly an experience. (p. 190)

Social scientists are often too quick to boast that their own limited empirical data undermine ontological claims. Religious traditions cannot be adequately understood without the assumption that transcendent objects of experience are believed to be real and foundational to those who experience them (Hood, 1995a). It is also possible that not only are they believed to be real, but that they are in fact real as well. Furthermore, their reality may be revealed in experience. Carmody and Carmody (1996, p. 10) define “mysticism” as “a direct experience of ultimate reality.” This definition remains a hypothesis capable of empirical investigation. To presuppose otherwise is less persuasive than once thought. Bowker (1973), after critically reviewing social-scientific theories of the sense of God, has noted that it is an empirical option to conclude that at least part of the sense of God might come from God. In our terms, religious views of the nature of the Real suggest ways in which it can be expressed in human experience. This can work in two directions, both deductively and inductively. Deductively, one can note that if the Real is conceived in a particular way, then certain experiences of the real can be expected to follow. Thus we can anticipate that expectations play a significant role in religious experience, often confirming the foundational realities of one’s faith tradition. Inductively, we can infer that if particular experiences occur, than the possibility that the Real exists is a reasonable inference—a position forcefully argued by Berger (1979). Thus we can anticipate that experiences, some unanticipated, may lead some to seek religions for their illumination. O’Brien (1965) has gone so far as to include in his criteria for a mystical experience that it be unexpected. Religious traditions adopt both options in confronting mystical and numinous experiences. In this sense, a rigorous methodological atheism is unwarranted in the study of religious and mystical experiences (Porpora, 2006). Not surprisingly, then, mystical experiences have long been the focus of empirical research and provocative theorizing among both sociologists and psychologists. We first explore classic efforts to confront these experiences. These classic views are of more than historical interest, as they set the range of conceptual issues that continue to plague the contemporary empirical study of mysticism. Our focus upon classic views is not exhaustive. We focus upon representatives of three major social-scientific views regarding mystical experience: as erroneous attribution, as a heightened state of awareness, and as evolved consciousness.[37]

4. Epistemology and Religious Experience

Debates about the existence of God have often included the “argument from religious experience”; its advocates may cite their own experiences, or may claim that perceptions of some sort of divinity are a psychological or cultural universal, or nearly so. These arguments may involve evidence such as

- The way that many people feel that they have met God or that God has spoken to them.
- The way that many people have had some sort of experience or perceptual of a divine or spiritual realm or order.
- The way that the human tendency to perceive or be interested in these topics (and to believe in, venerate, or attempt communications with deities or spiritual powers) is widespread across cultures, even though their interpretation varies so dramatically. (The notion of “natural religion” has sometimes been justified on the basis of supposedly universal notions among human societies, or supposedly universal experiences shared among human beings who contemplate the idea of the divine. In the view of proponents of this concept, we might have good evidence to believe in God or divinity as a result of widespread human experience of these things, but perhaps not good evidence to believe in specifics about the divine nature, which are much less widely agreed upon.)

While realizing that not everyone who has had a spiritual experience assigns *any* particular significance or interpretation to that experience, never mind formulating a specific theistic argument on the basis of the experience, we were interested in looking at how this argument is viewed by some of its proponents and opponents. In addition to reading online summary articles about arguments from religious experience, we chose to focus on William P. Alston's groundbreaking philosophical work on empiricism regarding religious/spiritual epistemology. Alston introduces his book *Perceiving God: The Epistemology of Religious Experience* in this way:

> The central thesis of this book is that experiential awareness of God, or as I shall be saying, the perception of God, makes an important contribution to the grounds of religious belief. More specifically, a person can become justified in holding certain kinds of beliefs about God by virtue of perceiving God as being or doing so-and-so. The kinds of beliefs that can be so justified I shall call "M-beliefs" ('M' for *manifestation*). M-beliefs are beliefs to the effect that God is doing something currently vis-a-vis the subject—comforting, strengthening, guiding, communicating a message, sustaining the subject in being—or to the effect that God has some (allegedly) perceivable property—goodness, power, lovingness. The intuitive idea is that by virtue of my being aware of God as sustaining me in being I can justifiably believe that God is sustaining me in being. This initial formulation will undergo much refinement in the course of the book.[38]

In order to avoid presupposing the existence of God, Alston specifies "the experiences in question as those that are taken by the subject to be an awareness of God (or would be so taken if the question arose)."[39] In this way, it seems Alston is exploring the justification of taking spiritual experience as the basis of forming "M-beliefs" (basically, beliefs formed by mystical experiences) irrespective of whether such beliefs correspond with reality. As Alston states: "I want to make explicit at the outset that my project here is to be distinguished from anything properly called an 'argument from religious experience' for the existence of God.… It is rather that people sometimes do perceive God and thereby acquire justified beliefs about God."[40] In this sense, Alston focuses mainly on the philosophical legitimacy of people treating their *own* religious and spiritual experience as evidence for belief in the object of those experiences—not on the legitimacy of arguing that others, too, ought to do so.

This distinction is interesting. After all, having religious experiences and adopting specific attitudes with respect to their meaning or implications could theoretically be completely independent, so we could imagine encountering

- Someone who has subjectively perceived God[41] but is unsure of whether this experience was veridical or significant, and does not adopt or argue for a view that the perception was necessarily real.
- Someone who has subjectively perceived God and is convinced of the reality of that perception (whether or not he or she supposes that others, who haven't shared this perception, ought to agree).

- Someone who has not had such an experience, but finds others' accounts of their experience persuasive and is inclined to agree with their interpretations.
- Someone who has not had such an experience and remains skeptical of others' accounts or interpretations.

Of course, still other nuances are possible.

Alston appears particularly interested in countering those who maintain that *individuals* ought to rationally discount their own personal experience (especially because they feel that personal experience isn't the kind of thing that could be rationally convincing in this realm). In some ways, Alston's approach could be seen as focused on the analysis of individual rationality and not about collective reasoning or persuasion.

In fact, Alston's views on evidence and belief seem remarkably compatible with other accounts of rationality; it seems that his main point is that experience of the divine is evidence that should lead the experiencer to update his or her beliefs in favor of a greater likelihood that the divine exists—which one would only *not* do given other priors that the divine is absurd, that perceptions of it are of a different kind of evidential value than other perceptions, or something. So Alston then dedicates long passages to claiming that there's no philosophical reason that we have to think that these perceptions *don't* have a similar evidential value from other kinds of perceptions, and to trying to defeat arguments that they don't.

We could probably successfully rephrase most of his line of argumentation with a Bayesian-rationalist flavor: “conceptually, there is nothing about the sense of the divine that makes it less rational to use it to update one’s beliefs about the objects of its perception than it would be to use other senses to update one’s beliefs about the objects of *their* perception.” Of course, *how much* to update will be largely informed by one’s priors, particularly here about naturalism or materialism, etc. The nut to be cracked here is this: how does rationality work when trying to learn about subject matter that is not perceived through the bodily senses but that presents itself as an analogously coherent or compelling perception? Alston’s position seems to include the notion that, even though we may have very high priors for naturalism, we shouldn’t confuse that with the choice to rule other kinds of perception out-of-bounds for belief updating.

Alston further clarifies his aim in this way:

> Even if our age were firmly realist in its predilections, my central thesis would still be in stark contradiction to assumptions that are well nigh universally shared in intellectual circles. It is often taken for granted by the wise of this world, believers and unbelievers alike, that "religious experience" is a purely subjective phenomenon. Although it may have various psychosocial functions to play, any claims to its cognitive value can be safely dismissed without a hearing. It is the purpose of this book to challenge that assumption and to marshal the resources that are needed to support its rejection.[42]

While his argument is very detailed, we present here an overview.

He begins with a discussion on the phenomenology of spiritual experience. After reviewing a few self-reported descriptions of spiritual experience, Alston emphasizes that they all share a common dimension of the experience as being *presented* to the subject. In this way, Alston likens spiritual perception to natural perception, and distinguishes presentation from abstract thought. He suggests that this aspect of *presentation* demonstrates that such experiences do not arise internally and are not subjective in their inception. He furthers this point by reviewing several descriptions of spiritual experiences which include sensory perception. Alston's phenomenology here is one in which "mystical perception" (his preferred term over "spiritual experience") is a "putative direct experiential awareness of God."[43]

Alston argues that a "direct awareness" of something—whether physical or mystical—is independent from beliefs, judgment, or concepts of the object of awareness.[44] This is in general accord with Russell's idea of "acquaintance" and Moore's idea of "direct apprehension." However, this view is in contrast with the view initiated by Kant that all perception is mediated by beliefs, judgment, or concepts. Alston argues against this view by stating that there is a difference between a direct awareness of something and the subsequent judgment that the object has some sort of property. In this way, Alston maintains that a person may be aware of an object without interpretational judgment on that object.[45]

The contrary view would easily be employed toward not taking any reports of mystical experiences seriously. If all perception is mediated by belief, then those who believe in God could easily be interpreting a non-spiritual event through their spiritual presuppositions. The argument could be made that the subjects of mystical experiences quoted by Alston are not reliable sources regarding the nature of their experiences. Alston responds to this charge:

> It is conceivable that one should suppose that a purely affective experience or a strongly held conviction should involve the experiential presentation of God when it doesn't, especially if there is a strong need or longing for such a direct awareness. But even if an individual's account of the phenomenology of his or her own experience is not infallible, it must certainly be taken seriously. Who is in a better position to determine whether S is having an experience as if of something's presenting itself to S as φ than S? We would need strong reasons to override the subject's confident report of the character of her experience. And where could we find such reasons? I suspect that most people who put forward these alternative diagnoses do so because they have general philosophical reasons for supposing either that God does not exist or that no human being could perceive Him, and they fail to recognize the difference between a phenomenological account of object presentation and the fact that a certain object, as the subject conceives it to be, presents itself to the subject's awareness. In any event, once we get straight about all this, I cannot see any reason for doubting the subjects' account of the character of their experience, whatever reasons there may be for doubting that God Himself does appear to them.[46]

Alston also notes that such experiences should be taken seriously because the subjects themselves consider alternative sources or interpretations for their experiences. In other words, it seems likely that such people remain rational regardless of the sometimes outlandish nature of their experiences, and are aware that their own perception may be at fault. The subjects rejected such alternatives, however, often due to the “presentation” aspect of the experiences—in other words, stating that they did not, themselves, produce the experience.[47]

Alston’s argument proper begins with the notion that spiritual experiences are perceptual in nature. That is, God presents himself to the subject in some manner. He takes that conclusion as the basis for his argument for the reasonability for taking such perceptions to be indicative of true reality. Alston introduces this argument in this way:

> If what seems to me to be a direct experiential awareness of X puts me in a position to form justified beliefs about X’s perceptible features, that warrants me in supposing that X itself is indeed presenting itself to my awareness; otherwise how could the experience justify my beliefs about X? We have to stop short of the claim that the perceptual justification of perceptual beliefs entails that the experience is genuine perception. I may be perceptually justified in believing that there is a lake in front of me even if I am a victim of a mirage and no lake is being perceived. But this is just an isolated incident that occurs against the background of innumerable cases in which perceptual justification involves authentic perception of the object. It strains credulity to suppose that an entire sphere of putatively perceptual experience could be a source of justification for perceptual beliefs, while there is no, or virtually no, genuine perception of the objects involved. Therefore, if putative experience of God provides justification for beliefs about God, that provides very strong support for supposing that such experiences are, at least frequently, genuine perceptions of God.[48]

Alston’s concept of justification for belief is that the subject is justified in maintaining his or her belief, rather than the subject’s activity of justification. The difference is that someone may be justified in his or her own beliefs irrespective of attempting to argue for them.[49]

He clarifies his view of *justification* as indicating favorability of true belief, taking into account that it is a matter of degree. One may be justified in a belief, even if the evidence does not lead one to be completely certain.[50]

Alston’s argument for the justification of M-beliefs is as follows: 1) A perceptual belief concerns a perceived object, no more and no less. In other words, a perceptual belief is that one is sensing a presented object. 2) This belief is formed primarily by an *experience* of perception (according to the human senses). 3) A perceptual belief is not based on prior beliefs or concepts. Alston writes, “The theory of justification I am using takes justification to be a function of the adequacy of what the belief is based on. If it is based purely on experience, and that basis is adequate, it will be purely immediately justified. If it is based partly on experience and partly on other beliefs, its justification will be partly immediate and partly mediate.”[51]

His recognizes the mediative role of previously held beliefs in the interpretation of perceptual experience, but does not assert that such background beliefs are required in the formation of perceptual belief. Alston writes:

> Background beliefs not infrequently figure in the total basis of perceptual beliefs, and in these cases the justification of the latter depends in part on the justification of the former. Nevertheless this is less common than it seems on first sight, and we can often

> explain the justificatory relevance of background beliefs without supposing them to be part of the basis, and so part of the prima facie justification. Thus there is considerable scope for purely immediately justified perceptual beliefs, even though partly mediately justified beliefs must also be taken into account.[52]

In applying this formulation to M-beliefs, Alston writes:

> If God appears to me as φ (or at least so it seems to me), then that will contribute to justifying a belief that God is φ; if the belief is purely immediately justified, that will be the whole story. If one is aware of what one takes to be God as loving or almighty, then, if no partly doxastic basis is involved or required for justification, a belief that God is loving or almighty formed on that basis is thereby prima facie justified. If one is aware of what one takes to be God as comforting one or saying that P to one, then, with similar restrictions, a belief that God is comforting one or saying that P to one is thereby prima facie justified.[53]

Regarding whether God can be perceived, Alston lays out the summary of his argument as such:

> To come to grips with the serious, unconfused problem here, we will have to cut through some unwarranted assumptions that may be behind these questions. We should not suppose that in order to succeed in perceptually recognizing an object of perception as X (i.e., become perceptually justified in believing, or perceptually know, that the object is X), it is necessary that the object appears to one as φ, where φ is a property uniquely possessed by X. To perceptually recognize your house, it is not necessary that the object even display features that are in fact only possessed by

> your house, much less features that only your house could possess. It is enough that the object present to my experience features that, in this situation or in situations in which I generally find myself, are sufficiently indicative of (are a reliable guide to) the object’s being your house. And so it is here. For me to recognize what I am aware of (X) as God, all that is necessary is that X present to me features that are in fact a reliable indication of their possessor’s being God, at least in situations of the sort in which I typically find myself. It is, again, not required that these features attach only to God, still less that they be such that they can attach only to God. And it is a matter for detailed investigation what sorts of appearances satisfy that condition, just as in the case of sensorily perceived objects.[54]

He then reviews the accounts of spiritual experiences he provided earlier in the book in order to identify the ways in which God presented his qualities as God.

In chapter 3, Alston surprisingly maintains that it is not possible to give adequate reasons for supposing that the beliefs formed by sense perception are reliable, even though it is common practice to do so.[55] In this way, Alston casts doubt upon the entire enterprise of a sense perception basis for accurate epistemology—on both a natural and mystical level. Alston writes,

> It is widely believed that we are in a much better position to judge that sense perception is a source of justification than we are in the case of theistic perception. Many even believe that we can show that sense perception is reliable, but not that mystical perception is. These convictions are used as a basis for downgrading the epistemic status of the latter and for

> denying that beliefs formed on the basis of theistic perception are justified. Looking carefully at attempts to show sense perception to be reliable will put us in a position to assess these views.[56]

Alston demonstrates that arguments attempting to prove the reliability of sense perception fail due to their epistemic circularity. Alston writes, “If we have to assume the reliability of SP [sense perception] in order to suppose ourselves entitled to the premises, how can an argument from those premises, however impeccable its logical credentials, provide support for that proposition?”[57] The simplest version of the argument for the reliability of SP is that it is proven by its fruit. In other words, if our understanding of reality based on SP are most often confirmed through prediction and control of events, then SP is reliable. However, this argument suffers from epistemic circularity in that the only way to confirm the accuracy of the fruit of SP is through the use of SP. Alston reviews a number of arguments for the reliability of SP put forth by (or emerging from) Descartes, Wittgenstein, Oldenquist, Kant, and Locke. However, he finds them all lacking, primarily due to the pitfall of epistemic circularity.

We don’t think Alston’s views and other philosophers’ responses to them will be very exciting or edifying except to readers who are deeply interested in technical debates about epistemology. But we found that they sparked interesting conversations for us about skepticism and how people personally respond to skeptical arguments. They also led us to discuss the conditions under which people can learn from other people’s experiences, and the assumptions that we may need to use in order to assess the evidentiary value for us of other people’s beliefs.

5. Near Death Experiences

For some people, near death experiences (NDEs) are a compelling form of spiritual experience that can be life-altering in its consequences. Those who have them may believe that they have met God, angels, or other spiritual beings, that they have in fact died and then deliberately been sent back to life, or that they have received some kind of teaching or message from spiritual entities. In some cases, they may claim to have received information that would be objectively confirmable by others. Some of those who have such experiences have written popular books recounting them and arguing for specific religious or metaphysical views on the strength of their and others' experiences.

One major scholarly work on near death experiences is *Otherworld Journeys: Accounts of Near-Death Experience in Medieval and Modern Times* by Carol Zaleski.[58] Zaleski reviews historical accounts of NDEs, as well as modern ones.

Although their contents and interpretations vary, NDEs are remarkably common in the modern world. Among those who come close to dying, the percentage of those who experience NDEs appears to range from 34 percent to 43 percent.[59]

One reason why NDEs are ripe for research is their commonalities. Zaleski writes, "The [sympathetic] researchers agree that the similarities of near-death reports are more striking than their differences and see this unanimity as a key to the validity of near-death experience."[60] It has also been observed that NDEs do not always conform to the individual's pre-existing desires or expectations, thereby contradicting the idea that they are merely comforting fantasies of some sort.[61] Another argument for their validity is their lasting, transformative effects on the individual.[62] Zaleski writes of the striking independent consistency of NDEs:

> Age, sex, race, geographic location, education, occupation, religious upbringing, church attendance, prior knowledge of near-death studies, all have negligible effect on the likelihood of near-death visions. Suicide victims seeking annihilation, fundamentalists who expect to see God on the operating table, atheists, agnostics, and *carpe diem* advocates find equal representation in the ranks of near-death experiencers. And their answers to survey questions show that, for all the religious implications of near-death experience, a person's beliefs about God, life after death, and heaven and hell do not determine the content of his vision.[63]

One study produced the following statistics: 60 percent described a sense of peace. 37 percent reported feeling a separation from the body. 23 percent entered a darkness. 16 percent saw light. 10 percent entered the light.[64]

There are a number of strong objections to viewing NDEs as experiences of a spiritual reality. Notably, NDEs by definition occur at moments of great damage and stress to the body, commonly in the course of partial or even complete failures of its systems. Thus, one could view NDEs as perceptions that are symptomatic of this damage in some way. One objection characterizes NDEs as hallucinations due to dysfunction in the nervous system.[65] Another objection suggests that sensory deprivation leads to the experience of NDEs. If the brain is continually cut off from the ability to perceive or process external perceptions, it composes its own reality.[66] (Immersion in a sensory deprivation tank, for example, often produces vivid sensory hallucinations.) If the neurological structure and functions of the brain alone are responsible for constructing these NDEs, then this would account for the commonalities of these experiences. Another criticism of the spiritual view of NDEs is the concept that human psychology will always attempt to deny death. “Although contemporary psychological treatments of this subject differ, all rest on the axiom that the mind will resort to any stratagem to push from view the prospect of its own annihilation….”[67]

However, Zaleski states that each individual objection cannot account for every reported NDE: “for every pathological condition presumed to cause near-death visions, one can find subjects who were demonstrably free of its influence….”[68] Zaleski continues: “researchers cite statistics that show an inverse relationship between near-death experience and various pathological mind-altering conditions….”[69] For example, the experience (or recollection) or NDEs seems to be inhibited by the effects of drugs and anesthetics. For this reason, it seems unlikely that drugs are responsible for the generation of NDEs. Zaleski writes, “Backed by the collective testimony of hundreds of subjects, researchers contrast the alert, blissful, lucid quality of near-death experience to the confusion, anxiety, and perceptual distortions that accompany such disorders as hypoxia, limbic lobe syndrome, autoscopy, depression, and schizophrenic hallucination.”[70]

While it seems the critics’ of individual suggestions for biochemical or psychological explanations of the origins of NDEs are dispatched by the variety of cases which do not conform to such explanations, if these criticisms are seen as a whole, then perhaps all NDE cases can be explained only by materialistic means. Zaleski writes: “any feature of near-death experience that is not finished off by endorphins can be dispatched by temporal lobe seizure, depersonalization, state-dependent birth recall, and so forth.”[71]

However, there are still arguments against this unified front. Firstly, if there were a number of underlying causes for NDEs, why is their nature so consistent? In other words, wouldn’t different biochemical processes produce different kinds of experiences? If consistency is a striking component of NDEs, it seems unlikely that there would be a lack of consistency in their origins. Secondly, Zaleski warns against a reductionistic view of NDEs when such heavy reductionism is not (usually) applied to general human experience. She writes,

> After all, not only extraordinary visions but also normal states of consciousness are linked with electrical and chemical events in the brain, hormonal tides in the body, inherited drives, and cultural coercion. Yet we do not apply reductionist vetoes to our ordinary experience. Love can be explained in terms of neurochemical and social mechanisms, ranging from the influence of advertising to the lure of pheromones; but scarcely anyone suggests that knowledge of these mechanisms should prohibit people from believing that they are in love and rearranging their lives accordingly.[72]

Why should an NDE be dismissed as invalid or hallucinatory because we are able to point to associated neurological or biochemical processes? It may be important to interrogate the reasons why we may be inclined to discount the subjective experiences of NDEs even if naturalistic processes are at play. (The same sort of question will arise again when we ponder experiences with psychoactive substances.)

Zaleski herself advocates for a balance in understanding and appreciating the symbolic and narrative value of NDEs while being skeptical of their theological value. She writes,

> Clearly, a new approach is needed; to make near-death testimony an arena for restaging old philosophical or theological battles will not suffice. It appears to be impossible, in any case, to determine objectively whether near-death reports are accurate or inaccurate depictions of the future life. It might therefore be more fruitful for theologians to consider near-death visions as works of the religious imagination, whose function is to communicate meaning through symbolic forms rather than to copy external facts. This is the aspect of near-death literature that I have attempted to highlight.[73]

In other words, Zaleski does not personally believe that NDEs point to an objective spiritual reality, but that they should nevertheless be afforded personal and cultural significance. She writes,

> I suggest, therefore, that a pragmatic method and a sensitivity to symbol must go hand in hand if we wish to give a fair hearing to the claims of near-death literature. If we fully recognize the symbolic nature of near-death testimony (and accept the limits that imposes on us), then in the end we will be able to accord it a value and a validity that would not otherwise be possible; this in turn will yield further insight into the visionary, imaginative, and therapeutic aspects of religious thought in general.[74]

Zaleski seems to take a position which stands in weak support of the controversial statement at the core of this paper. Although she does not see NDEs as pointing to any sort of genuine mystical reality, she nevertheless sees value in such experiences. If Zaleski is right, and NDEs merely carry symbolic power, they would still be representative of an important component of human life as a kind of spiritual experience with powerful, lasting effects upon the individual.

6. The Use of Entheogens

Perhaps the relevant topic which is most suited toward empirical study is the use of entheogens in the generation of spiritual experience. This seems true for a few reasons: firstly, they are drugs whose chemical make-up can be scientifically studied and described, and secondly, their effect on the human on the biochemical and neurochemical levels can also be scientifically studied and described. We also have the benefit of several decades of scientific research on these substances, in addition to extensive reports from people’s self-experimentation, and anthropological and insider accounts of their use within traditional religions. A number of psychedelic substances have also recently benefited from a new round of legally-sanctioned medical investigations aimed in part at evaluating their potential use in psychiatry.

Entheogens are psychedelic drugs or other preparations of comparable substances that tend to produce a subjective experience of a divine or spiritual nature. They are controversial on many levels; most are legally controlled substances, they are often seen by governments and health authorities as having the potential to be abused, and different religious traditions have radically opposed views about their harms and benefits[75]. It is also striking that they seem to offer such a reliable way to create experiences that most subjects view as meaningful and spiritual, although the content of these experiences and the way participants account for them varies tremendously[76].

Writing for *The Outline* about her personal experience with psilocybin in a controlled setting at Johns Hopkins, Rachael Peterson states:

> Like all trip stories, mine sound crazy at worst and clichéd at best. But I can tell you this much: at the peak of my experience, my sense of self dissolved and I unified with an abiding force that permeated all existence — something that felt conscious, vast, benevolent, eternal, peaceful, and furiously important. After sitting up on the couch six hours later, covered in snot and tears, I struggled to put words to an encounter that felt more real than everyday reality — a mind-bendy paradox characteristic of many mystical experiences.[77]

Similar accounts (often involving interaction with entities that seem conscious and intelligent) are described repeatedly in case reports by DMT researchers such as Rick Strassman[78].

A recent paper analyzes the data of five groups who had experiences with various entheogens, including psilocybin, LSD, ayahuasca, or DMT, as well as one who had spiritual experiences without the use of drugs.[79] The authors of the paper note that there are some differences between the experiences of the entheogen cohort and the non-drug cohort, they are more similar than different. The most striking finding was that two-thirds of participants who identified as atheist prior to their experience no longer considered themselves atheist afterward. The abstract of this article summarizes the nature of these experiences:

> Most participants reported vivid memories of the encounter experience, which frequently involved communication with something having the attributes of being conscious, benevolent, intelligent, sacred, eternal, and all-knowing. The encounter experience fulfilled a priori criteria for being a complete mystical experience in approximately half of the participants. More than two thirds of those who identified as atheist before the experience no longer identified as atheist afterwards. These experiences were rated as among the most personally meaningful and spiritually significant lifetime experiences, with moderate to strong persisting positive changes in life satisfaction, purpose, and meaning attributed to these experiences. Among the four groups of psychedelic users, the psilocybin and LSD groups were most similar and the ayahuasca group tended to have the highest rates of endorsing positive features and enduring consequences of the experience. Future exploration of predisposing factors and phenomenological and neural correlates of such experiences may provide new insights into religious and spiritual beliefs that have been integral to shaping human culture since time immemorial.

Certainly the subjective power of these experiences cannot be denied, as their transformative effects upon the subjects' personal lives can attest. In the conclusion, the authors reiterate that that the reported effects of these experiences led to "persisting moderate to strong positive changes in attitudes about self, life satisfaction, life purpose, and life meaning that participants attributed to these experiences." Legal obstacles have meant that formal therapeutic exploration of many entheogens is just beginning, but it seems apparent that they have promise in treating mental illnesses through catalyzing deep experiences of profound meaning.

Good Friday Experiment

On Good Friday in 1962, a researcher (Walter Pahnke) administered psilocybin to divinity student volunteers just before they attended a Christian worship service[80]. The results were dramatic; almost all of the research subjects reported profound experiences which they continued to regard as meaningful and important for the rest of their lives (as confirmed by a follow-up survey decades later). Some have described the experience as the most powerful spiritual experience of their lives—which is noteworthy because they were generally already religious believers who were preparing for careers related to their Christian faith. The volunteers clearly understood that they were being given a drug, which did not seem to reduce their assessment of the spiritual importance of their experience at the time or upon subsequent reflection.

Huston Smith, a scholar of comparative religion, was one of the participants in the experiment and describes his experiences in *Cleansing the Doors of Perception*, his survey work on entheogens[81]. He says that his attention was fixed on particular melodic and lyrical features of a hymn sung during the service, and his musical training and Christian upbringing

> converged on the Good Friday story under psilocybin, [and] the gestalt transformed a routine musical progression into the most powerful cosmic homecoming I have ever experienced.

In an interview he reprints in the same book, Smith remarks that the experiment

> […] enlarged my understanding of God by affording me the only powerful experience I have had of his personal nature. I had known and firmly believed that God is love and that none of love’s nuances could be absent from his infinite nature; but that God loves *me*, and I *him*, in the concrete way that human beings love individuals, each most wanting from the other what the other most wants to give and with everything that might distract from that holy relationship excluded from view—*that* relation with God I had never before had. It’s the theistic mode that doesn’t come naturally to me, but I have to say for it that its carryover topped those of my other entheogenic epiphanies. […]

Sam Harris cautions in *Waking Up* that if powerful and important spiritual experiences can be produced by a particular context, we should be wary of taking them as evidence for specific doctrinal metaphysical claims, since presumably these experiences can be, and are, encountered within many different religious traditions. If Christian divinity students had an experience that they took to be an encounter with God under the influence of psilocybin at a Good Friday service, other subjects might experience equally compelling encounters that they interpreted when using the same subject in another setting, or merely with different prior metaphysical beliefs. Harris supposes that a spiritual experience he personally had at the Sea of Galilee would have confirmed for him his particular religious faith, if he had had one. Later Harris asks:

> What does a spiritual experience mean? If you are a Christian sitting in church, it might mean that Jesus Christ survived his death and has taken a personal

> interest in the fate of your soul. If you are Hindu praying to Shiva, you will have a very different story to tell. Altered states of consciousness are empirical facts, and human beings experience them under a wide range of conditions.

Smith is also sensitive to this reasoning and notes that participants in other religions’ rituals, and those raised in different traditions, have had equally powerful experiences of their own. But Smith and Harris take this observation in radically different directions; for Harris this diversity undermines any specific metaphysical claim that any religious tradition might advance, while for Smith, it provides an encouragement to try to find commonalities in different religious experiences and reason to suspect that they are pointing toward the same reality.

The Eleusinian Mysteries

> […] χαλεποὶ δὲ θεοὶ θνητοῖσιν ὁρᾶσθαι.
>
> […] gods are hard for mortals to recognize.
>
> — *Homeric Hymn to Demeter*, 111 (translated by Helene P. Foley)

The Eleusinian Mysteries were an ancient Greek tradition practiced for thousands of years relating to the worship of the goddess Demeter and the story of her daughter Persephone’s descent into and return from the underworld. (The story, itself widely known in Greek culture, said that Persephone was kidnapped as a bride by Hades, the god of the underworld, and yet Demeter was able to bring her back—at least for part of each year, corresponding to the growing season.)

The associated ceremonies could at least originally be practiced only at a specific site near the city of Eleusis. Participants signed up to be initiated over the course of several days, during which they participated in allegorical rites. The initiates swore a vow of secrecy and were never allowed to talk about the details of what the rites consisted of. Although participants went home afterward and returned to their normal lives, virtually all of them took the vow extremely seriously, taking their memories to their graves, so we know very little today about exactly what went on at Eleusis.

The Mysteries attracted participation from the rich and famous of the classical world, and remained extremely popular throughout classical antiquity. (They were open to the public, but could only be experienced, or discussed, in the proper place, at the proper time, with the proper preparation.) Although those who had taken part almost never gave any concrete details, they generally considered the experience extraordinarily valuable and worthwhile, and often recommended it in the strongest terms to their friends and family members. Many indicated that they had had some kind of contact with the divinity during the initiation, and some said that they were no longer frightened of death. Many sources suggest that the Mysteries taught, or showed, their initiates some very specific reason why death was nothing to be afraid of—and participants apparently took this lesson to heart. Cicero, who was probably a participant himself, says in his *De Legibus* that the Mysteries' initiations allowed "neque solum cum laetitia vivendi rationem accepimus, sed etiam cum spe meliore moriendi" ("[that] we not only took from them a way of living happily, but also a way of dying with a better hope").

The Mysteries stopped being celebrated with the rise of Christianity and no one has experienced them at all for more than a millennium and a half. Since then, people have remained intensely curious about what was done and taught in these ceremonies, and why participants found them so valuable and transformative, over and over again, and why so many of them claimed to lose all concern about death. How could this have happened? Were all of them really encountering Demeter, or witnessing Persephone returning from the realm of the dead?

One interesting fact is that all the initiates were given a drink called a kykeon (κυκεών, meaning something like ‘mixture’) during the course of the initiation. The recipe for the kykeon is one of the details lost to history due to the Mystery initiates’ dedication to keeping their vows, but people have often wondered about how it might have affected those who drank it. In the 1960s, two prominent psychedelic researchers (R. Gordon Wasson, who introduced psilocybin to the American public, and Albert Hofmann, who discovered LSD) and a classicist (Carl Ruck) published *The Road to Eleusis*, a book arguing that the kykeon contained substances derived from the ergot fungus. Wasson and Ruck subsequently published numerous other books arguing that religions all around the world have traditionally used psychoactive substances (termed *entheogens*) to facilitate the experience of the divine, and that religious doctrines, narratives, and rituals are often at least initially based on drug-mediated experiences. Other scholars have tended to accept the view that the kykeon probably contained psychoactive substances, even if they disagree about exactly what substances these were.

Interestingly, scholars who endorse a substance-related explanation for the experiences of Eleusinian initiates (and for the origins of other religious traditions and narratives) don’t necessarily believe that they are *debunking* or *explaining away* these experiences. This is an oddity for skeptics who might imagine that initiates were being tricked into interpreting their experiences as divine contacts or visions.

7. The Problem of Dreaming

Dreaming is a universal human experience that poses a difficulty for confidently viewing spiritual experience as veridical, because dreams usually *feel so real* and *feel so important*. They represent a familiar, ubiquitous form of experience that is

- Vividly perceived
- Often deeply personally meaningful
- Often considered, at least metaphorically, to reflect the experiencer’s deepest desires, aspirations, or values
- Sometimes transformative in their consequences
- Usually accepted as completely real during the experience itself
-

Often involve extended interactions with other beings who are perceived as separate from the experiencer

- Difficult to convey to others

Yet modern western culture mostly accepts that dreams are *not* veridical—that they tell us little or nothing about how the world is—and that at most they might reveal or reinforce something about the dreamer’s own memories, desires, or unconscious psychology. (It’s worth noting that many cultures have assumed, or believed on evidence, that dreams do come from somewhere outside us, or that they represent a visit of some kind to another world. Many people have routinely believed, or routinely believe, that dreams represent genuine visions of the future, visits from one’s ancestors, temptations from evil forces, or glimpses of the real structure underlying the mundane world. Our neuropsychological depiction of dreaming as the brain’s efforts to make sense of random noise is quite unusual among cultures’ ways of accounting for dreams. But all cultures have had to deal with the ephemerality of dreams and the way that their content is at best unreliable and difficult to make practical use of in one’s day-to-day life.)

The relevance of dreams for a skeptical account of spiritual experience could include a strong version and a weaker version. In the strong version, many spiritual experiences could be suspected of originating in dreams that had been somehow misperceived or misremembered by their dreamers as waking experiences, or correctly perceived as dreams but somehow nevertheless accepted as genuine. In the weaker version, dreams simply provide an analogy that shows that our minds are sometimes capable of producing experiences that feel genuine and important, whose genuineness we reflectively accept, but whose connection to mind-independent reality is questionable.

A counterargument is that we should not have to doubt all of our interpretations of our experiences merely because we are sometimes temporarily mistaken about whether we are dreaming. Otherwise, we would fall into radical skepticism about all of our knowledge and experience. (People throughout history have flirted with embracing this conclusion seriously in various ways. For example, Descartes uses the experience of waking from a dream that he had believed in while it lasted as a touchstone of his motivation for undertaking to doubt everything[82], the Chinese philosopher Zhuang Zhou claims not to know "whether he was a man who dreamt he was a butterfly or a butterfly dreaming he was a man", while the movie *Inception* depicts people who, accustomed to the experience of "awakening" within a dream, become unsure of how many times they still have to wake up in order to return to the waking reality[83].

What kinds of things could affect our judgment about whether particular dreams, or some dreams, are veridical? Perhaps dreams could be shown to foretell future events. Or perhaps the content of many people's dreams could coincide, like Italo Calvino imagines in *Invisible Cities*:

> They tell this tale of its foundation: men of various nations had an identical dream. They saw a woman running at night through an unknown city; she was seen from behind, with long hair, and she was naked. They dreamed of pursuing her. As they twisted and turned, each of them lost her. After the dream, they set out in search of that city; they never found it, but they found

> one another; they decided to build a city like the one in the dream.

But sometimes the content of different people's dreams *does* coincide. For example, many people have a recurring dream that they're in school and have somehow skipped class or forgotten to prepare for an important test. Why do so many people share this dream content? Does this suggest that there's something veridical about this particular dream? Or at least that there's something psychologically important about this situation?

Certainly the communication of dreams and their subjective significance mirrors the difficulties encountered by someone who wishes to communicate a spiritual experience. Dreams are usually deemed significant only to the one who has the dream. Most of us find that to be completely appropriate. However, spiritual experiences are generally thought (at least by those who have them) to be important enough to share with others, potentially as experiences of something real and relevant for others. The question then is how we might differentiate between a dream and a spiritual experience for the skeptic. Research into spiritual experience tells us that the two are different, at least in how they are perceived by the individual. It is often the case that dreams are easily forgotten and cease to seem real when the dream is over. However, spiritual experiences—particularly those of the NDE and entheogenic varieties—have lasting effects, and, at times, permanently changing behavior. As Zaleski writes about near death experiences: "In addition, against Siegel's sweeping comparison of near-death visions to the psychoneurology of hallucination, the researchers cite nearly unanimous testimony that near-death experience is subjectively different from dreaming or intoxication; that it is, as one of Sabom's subjects puts it, 'realer than here.'"[84] This may be true for the one who experienced the NDE, but, for those who did not have that experience, hearing about it may elicit a similar reaction to hearing about a dream.

8. Some Possible Perspectives

We join in recommending epistemic humility across the board for all perspectives. To have any sort of meaningful discussion about any topic, we must allow for the possibility that we are possibly wrong in our opinions and beliefs. Dialogue about experiences that bear on people's religious (or anti-religious) beliefs is often challenging.

We should be careful in how we interpret spiritual experience, especially if it prompts us to take action or change behavior. If a metaphysical or spiritual reality does not exist, then this suggestion is self-explanatory. However, epistemic humility is still important if a metaphysical or spiritual reality does exist. Suppose that God can and does speak to humans. The present evidence supports the idea that such communication may not always be clearly understood or interpreted by the human listener, and that human beings may not find it easy to be confident about when God is speaking to them. If this were not true, we would obviously not see such a broad range of (often) conflicting divine messages. Being careful with the knowledge and action generated by spiritual experience is therefore wise.

In a related context, we had fruitful discussions of the parable of the blind men and the elephant. The blind men and the elephant is an ancient story, and we highly recommend the Wikipedia entry for it.[85] The basic story is that a number of blind men came upon an elephant, and they all touched the elephant while attempting to describe it. They all disagreed upon the nature of the elephant because one man felt the side of the elephant and said it was like a wall, while another felt the trunk and said it was like a snake, etc. Interestingly, the moral that this story is meant to convey is subject to dispute by its listeners! One might take the lesson of the story as a metaphor, for each blind man erroneously believes in the truth of his own religious doctrines based his limited perception. Because each person's perception is limited, conclusions on the nature of the elephant cannot be trusted. However, another way to interpret the story begins with the idea that the radical differences of perception should not be taken as proof that there is no elephant. People in different cultures and time periods may be seeing something real, but their description of it may not be complete. The lesson may be that even though we are tempted to discount spiritual accounts because of their inconsistency, we should not discount them completely because they may still be of evidential value, albeit in a more limited or complicated way than the authors of the accounts appreciate.

Like Plato's allegory of the cave and other thought experiments in philosophy, the tale of the blind men and the elephant reminds us that epistemology is hard. It's also challenging to integrate reported perceptions and experiences that differ greatly from person to person. Much of our reason for confidence in our understanding of the physical world is our ready and far-reaching intersubjective agreement about our sense perceptions of it. "Everyone" more or less agrees that we see the Moon in the sky, hear the sounds of rain, are injured by fire, like the taste of sugar, or find it challenging to lift horses. Accepting these widely-shared perceptions as veridical even without discussion or reflection is the most natural thing in the world. But when Blake says

> I assert for My self that I do not behold the Outward Creation & that to me it is hindrance & not Action; it is as the Dirt upon my feet, No part of Me. “What,” it will be Questioned, “When the Sun rises, do you not see a round Disk of fire somewhat like a Guinea?” O no no, I see an Innumerable company of the Heavenly host crying “Holy Holy Holy is the Lord God Almighty.” I question not my Corporeal or Vegetative Eye any more than I would Question a Window concerning a Sight: I look thro it & not with it.

it may be a greater challenge to achieve a consensus about the status of his vision, even if we suppose that he reported it faithfully and earnestly[86].

As we mentioned, Alston’s work prompted us to discuss some epistemological issues. We recognized the likelihood that people’s intuitions will vary as to what evidence is compelling enough to consider a spiritual experience to be (some combination of) true, authentic, and/or indicative of some spiritual dimension of existence. (Even some of the authors we read are still struggling with this question for themselves.)

Apart from this, we might ponder the notion that entheogenic or meditation experiences are valuable because they bring people in contact with facets of their own consciousness that are not normally accessible, whether or not we believe that these facets are *metaphysically* different from ordinary experience of the world and the mind. Some advocates of meditation and/or entheogens who incline toward naturalism maintain that we are “merely” learning more about ourselves (or, perhaps, our non-selves) through these experiences, yet that this learning is of immense value. Clearly, entheogens generate experiences which have lasting, powerful effects upon those who take them, and which are often described as improvements in their users' lives. Even if it is not a revealing of a separate spiritual reality, should such striking and consequential entheogenic experiences be considered "real" at least in terms of the individual's own psychology? Considering the literature on the psychological study of religious experience, on near death experiences, and on entheogens, it is undeniable that people subjectively experience genuine states of being which transcend normal states of consciousness. Is there a satisfactory way to consider these experiences' meaning without insisting that they prove something specific about how the world is? And, considering Alston, is it even plausible to expect more than this from any particular kind of human perception?

Alston might maintain that two people may be justified to believe different things based on differing experiences. Their perspectives may be in conflict, but this need not imply that either is behaving irrationally. It may be rational for one person to believe that spiritual experiences are indicative of a spiritual plane of existence that is usually inaccessible, while another person may also be rational in the belief that such an idea is silly nonsense and that spiritual experiences are the result of purely natural causes. Alston's concept of epistemology is that it is empirically formed through perceptual experience. Therefore, rational knowledge may not ever be universally uniform. In fact, this prospect is widely accepted in rationalist and empiricist circles, since rationality has to do with making good use of evidence and different people are in possession of different evidence. Those with spiritual experiences simply have a particular sort of evidence which they may try to interpret alongside other forms of evidence[87].

Regarding veridicality and consensus, one possible tool in the difficult arena of parsing spiritual experience is inquiry for confirmation. People generally agree upon the common physical and psychological experiences of natural life (e.g. “we both agree this coffee is too hot to drink right now,” or “everyone is laughing so this movie must be funny”) while also finding dissimilarities (e.g. “although this coffee is too hot for you to drink, and I agree that it is hot, it is not too hot for me to drink,” or “everyone else is laughing, but I don’t think this movie is funny”). This kind of discussion of experience toward confirmation is actually rather common in everyday life. Spiritual experience might be discussed and confirmed in a similar manner, leading to increased confidence in confirmed commonalities compared to experiences that do not seem to be shared. These discussions could, of course, be extremely culturally or institutionally difficult. In light of Alston’s claims about the general unreliability of all human perception, finding where subjective experiences agree (whether spiritual or otherwise) is a wise course of action. Of course it is still possible to find agreement and still be wrong, but at least we’ll be wrong together.

One reason why empiricism seems to be of value in discussing and interpreting spiritual experience is that people’s experiences tend to steer both our beliefs and our skepticism. Skepticism is a powerful tool that can be leveled at anything. As a negative example, conspiracy theorists can find reasons to be skeptical of any claim which rebuts their conspiracy theories. More relevant to this project, one could criticise both religion and materialistic rationalism as being culturally conditioned and psychologically motivated phenomena. Skepticism can cut both ways on this topic. However, understanding that our experience informs our future views, then even amidst our own skepticism of spiritual experience, we can appreciate the potential rationality of various worldviews informed by such experiences.

Concluding Thoughts

The question that we've aspired to consider in this project is:

> The empirical study of the content and nature of people's personal spiritual experiences justifies taking them seriously as evidence of an important component of human life deserving of individual and collective exploration.

What are the present authors' views of this statement, and what did they learn that surprised them?

Gruenberg originally proposed the project from the point of view of a supporter of this statement; at the end of the project, he finds that he strongly agrees with it.

Gruenberg found a number of surprises in the course of this research. The first is the vast amount of empirical research that has already been accomplished, particularly in the field of psychology. The researchers in the field are serious about the objective study of spiritual and religious experience, regardless of their stance on their veridicality. Their commitment to objectivity and sensitivity to the subject is admirable. The second is the compatibility of the (often) secular use of entheogens with non-drug-assisted spiritual experience, particularly in the emotional benefits of entheogenic experiences, as well as the surprising conversion rate of atheists to non-atheists through the use of entheogens. The third is Schoen's objection to the veridicality of spiritual experience on the grounds of the non-veridical nature of dreams. Mystical experiences of waking life—including meditative transcendent states of the loss of self and numinal encounters with God—seem to be of a different nature than dreams, but their precise differences are difficult to articulate to those who have not had them. The fourth is the all-encompassing nature of skepticism that we could have regarding any sense perception (advocated by Alston) that ends up “leveling the playing field” between spiritual and non-spiritual experiences mediated by our perception. In other words, if all sense perception is suspect, then spiritual experiences are no less suspect than material experiences. The fifth is Schoen’s use of the parable of the blind men and the elephant in order to demonstrate the possibility that limited perception may not always be erroneous. Perhaps we are all blind men and women, probing the elephant of reality. If we live in a version of that parable, then it is even more imperative that we actively seek a discourse which explores our disagreements with a willingness to try on some different versions of the elephant.

Schoen joined the project as a skeptic of this statement, but concludes that he ends in weak agreement with it, noting that a broad spectrum of authors whose work we considered maintain that the human mind is capable of a much broader range of experiences than we’re used to thinking of, and that these experiences can potentially be sought out deliberately and often have profound consequences for the experiencers’ lives and worldviews. In some sense, this much is agreed by people with otherwise opposing views on questions such as naturalism or theism. These authors maintain that there is more inside us than we know, or that we are capable of more than we know, or that personal experiences can give rise to important philosophical challenges. While it’s not clear that these facts imply any particular view of reality or the universe, they seem to have import, at least, for our own lives and self-concepts.

Schoen was also surprised by a number of things he learned in the course of this research. The first is the way that large numbers of people *do not* dismiss the meaning or relevance of their own or others’ drug-mediated experiences even when they are explicitly aware that they were “on drugs” and that their normal brain function was altered by the influence of substances like LSD, psilocybin, mescaline, or DMT[88]. Reminding them of the drugs’ role does not seem to change this. What’s more, vast numbers of people durably rate these experiences as extraordinarily significant, valuable, and influential, even on sober[89] retrospective reflection, and tend to feel that their understanding of their minds was expanded. Schoen would have expected the role and impact of substances in religious ritual from antiquity to the present as a sort of dirty secret that would draw the meaning or authenticity of religious experience into question. Instead, many participants in religious practices have demonstrated that they *knew* that drugs were centrally involved and did not reject the experiences or associated insights on that account.

A second surprise is the analogy drawn by Alston and others between the unreliability of spiritual perception and the unreliability of other sorts of perception, including some that clearly are part of our social consensus about reality. One can note that sense perceptions like vision and hearing are well-confirmed by the understanding they yield of the physical world and the predictions this understanding allows us to make, but other kinds of perceptions are socially normalized and believed even without the same degree of objective confirmation. While Alston might have an extremely low prior for the plausibility of metaphysical naturalism, and even shockingly low from the point of view of the median SSC reader, it’s interesting to consider the idea that perceptions other than the use of our bodily senses can rationally count as evidence much like other evidence. (This debate really starts there rather than ending there, and includes rather involved issues about how our beliefs and experiences reinforce one another.) It’s interesting to note how cultural and societal consensus hold so much more power over belief (in all different eras and all different sorts of societies) than any kind of formally-reasoned epistemology according to a detailed philosophical theory!

A third surprise is the extent to which mystical traditions are consciously aware of the risk of spurious spiritual insights due to, for example, mental illness. Skeptics often suggest that neurological disorders such as epilepsy might cause experiences that religious believers interpret as religious ecstasy or revelation. Organized religious and contemplative traditions are at least grappling with these risks and actively trying to find ways to distinguish experiences that they see as valuable from experiences that might be attributed to disorders. Mark Salzman’s novel *Lying Awake* movingly described this issue as it presents itself in the life of a nun who receives a sense of purpose and fulfillment (and more practical benefits for herself and her religious community) from ecstatic religious visions that are later attributed to the influence of a treatable brain tumor. All traditions and communities that rely on visions and revelations worry about how one can tell whether a particular experience constitutes a good or legitimate source of insight or teaching, although their answers are not necessarily satisfactory or reassuring to outsiders.

A fourth surprise is the frequency with which spiritual experiences are more often sought out or provoked rather than spontaneous or unexpected[90], and the popularity of the view that one can be taught to have such an experience by a particular method (particularly in contemplative and mystical traditions, as well as among advocates of insight meditation and entheogens[91]). Near-death experiences are a major exception here because people have rarely actively hoped to receive one or sought one out. Quite a few traditions suggest that there is a specific thing we ought to do, or a specific practice we ought to follow, in order to receive spiritual experiences and spiritual insights. While that creates its own set of risks (for example, of being manipulated or exploited by an unscrupulous teacher or dysfunctional community), it also provides an interesting opportunity for people to try things out for themselves if they're so inclined. This picture of practical steps is especially associated with secular Western interpretations of Buddhism, which emphasize the claim that Buddhist teachings can be taken as a phenomenological how-to guide[92].

This collaboration expressly excluded the truth or falsity of naturalism, or of any specific religious doctrine, from its scope. Although we and many of the authors we consulted have views about this, we didn't try to find a consensus about these questions. The evidence for and against a naturalistic worldview, or for and against some specific kind of supernatural phenomenon, might make an interesting adversarial collaboration topic for others in the future.

It's worth acknowledging the pervasiveness of the belief that spiritual experiences can't usefully be described or analyzed in words, and that trying to theorize about them is an absurd and useless activity. On this account, this whole project is an exercise in futility, doomed from the start, and perhaps a mockery of itself. While the present authors don't share this attitude, and even see it as counterproductive, they realize that others would strongly recommend experiencing spirituality, not talking about it.

We realize that there is a huge literature about the phenomenology and interpretation of spiritual experiences, and that we’ve only managed to scratch the surface here. Nor have we engaged with every issue raised within the sources that we did review. Interested readers looking for more material on these topics might want to start with Prof. Wesley Wildman’s bibliography on religious experience at <http://people.bu.edu/wwildman/relexp/bibliographies/bib_general.htm> or the Stanford Encyclopedia of Philosophy’s article on religious experience at <https://plato.stanford.edu/entries/religious-experience/>.

About the Authors

Jeremiah Gruenberg has a Ph.D. in theology. He writes and hosts a podcast on Christian spirituality, called *The God Experiment.*

Seth David Schoen is a computer and language enthusiast living in San Francisco.

[1] Charles Wolfe, Ofer Gal. *The Body as Object and Instrument of Knowledge: Embodied Empiricism in Early Modern Science*. France. Springer, 2010, STUDIES IN HISTORY AND PHILOSOPHY OF SCIENCE, 978-90-481-3686-5. hal-01238121 https://hal.archives-ouvertes.fr/hal-01238121/document

[2] Creath, Richard, "Logical Empiricism", *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/entries/logical-empiricism/#EmpVerAntMet. Creath writes, “Because logical empiricism is here construed as a movement rather than as doctrine, there is probably no important position that all logical empiricists shared—including, surprisingly enough, empiricism. And while most participants in the movement were empiricists of one form or another, they disagreed on what the best form of empiricism was and on the cognitive status of empiricism.”

[3] Creath, Richard, "Logical Empiricism", *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/entries/logical-empiricism/#EmpVerAntMet

[4] Markie, Peter, "Rationalism vs. Empiricism", *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2017/entries/rationalism-empiricism. Markie writes, “The dispute between rationalism and empiricism concerns the extent to which we are dependent upon sense experience in our effort to gain knowledge. Rationalists claim that there are significant ways in which our concepts and knowledge are gained independently of sense experience. Empiricists claim that sense experience is the ultimate source of all our concepts and knowledge.”

[5] Markie, Peter, "Rationalism vs. Empiricism", *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2017/entries/rationalism-empiricism.

[6] The *Stanford Encyclopedia of Philosophy* also notes that Keith Yandell distinguishes five sorts of religious experiences, one monotheistic, three associated with South Asian religious traditions, and one related to “nature”. This kind of classification is daunting (perhaps further investigation would reveal dozens more!) but also useful when people want to talk about their experiences and attempt to compare and contrast them.

[7] Zinnbauer and Pargament. “Religiousness and Spirituality,” 36.

[8] Zinnbauer and Pargament. “Religiousness and Spirituality,” 37.

[9] Zinnbauer and Pargament. “Religiousness and Spirituality,” 24-25.

[10] Zinnbauer and Pargament. “Religiousness and Spirituality,” 25.

[11] Ralph W. Hood, Jr, Peter C. Hill, and Bernard Spilka. *The Psychology of Religion: an Empirical Approach, Fourth Edition.* New York and London: The Guilford Press (2009).

[12] Raymond F. Paloutzian and Crystal L. Park, eds. *Handbook of the Psychology of Religion and Spirituality.* New York and London: The Guilford Press (2005).

[13] Ralph W. Hood, Jr, Peter C. Hill, and Bernard Spilka. *The Psychology of Religion: an Empirical Approach*, Fourth Edition. 294.

[14] Raymond F. Paloutzian and Crystal L. Park. “Integrative Themes in the Current Science of the Psychology of Religion,” from *Handbook of the Psychology of Religion and Spirituality.* Raymond F. Paloutzian and Crystal L. Park, eds. New York and London: The Guilford Press (2005), 8.

[15] Ralph W. Hood, Jr., and Jacob A. Belzen. “Research Methods in the Psychology of Religion,” from *Handbook of the Psychology of Religion and Spirituality.* Raymond F. Paloutzian and Crystal L. Park, eds. New York and London: The Guilford Press (2005), 67.

[16] Ralph W. Hood, Jr., and Jacob A. Belzen. “Research Methods in the Psychology of Religion,” 67.

[17] Ralph W. Hood, Jr, Peter C. Hill, and Bernard Spilka. *The Psychology of Religion: an Empirical Approach*, Fourth Edition. 347.

[18] Ralph W. Hood, Jr, Peter C. Hill, and Bernard Spilka. *The Psychology of Religion: an Empirical Approach*, Fourth Edition. 60.

[19] Ralph W. Hood, Jr., and Jacob A. Belzen. “Research Methods in the Psychology of Religion,” 69.

[20] Ralph W. Hood, Jr., and Jacob A. Belzen. “Research Methods in the Psychology of Religion,” 70.

[21] Ralph W. Hood, Jr, Peter C. Hill, and Bernard Spilka. *The Psychology of Religion: an Empirical Approach*, Fourth Edition. 42.

[22] Ralph W. Hood, Jr, Peter C. Hill, and Bernard Spilka. *The Psychology of Religion: an Empirical Approach*, Fourth Edition. 65.

[23] Ralph W. Hood, Jr, Peter C. Hill, and Bernard Spilka. *The Psychology of Religion: an Empirical Approach*, Fourth Edition. 65.

[24] Ralph W. Hood, Jr, Peter C. Hill, and Bernard Spilka. *The Psychology of Religion: an Empirical Approach*, Fourth Edition. 371.

[25] Ralph W. Hood, Jr, Peter C. Hill, and Bernard Spilka. *The Psychology of Religion: an Empirical Approach*, Fourth Edition. 371.

[26] Lee A. Kirkpatrick. “Evolutionary Psychology: An Emerging New Foundation for the Psychology of Religion,” from *Handbook of the Psychology of Religion and Spirituality.* Raymond F. Paloutzian and Crystal L. Park, eds. New York and London: The Guilford Press (2005), 106.

[27] Lee A. Kirkpatrick. “Evolutionary Psychology: An Emerging New Foundation for the Psychology of Religion,” 107.

[28] Lee A. Kirkpatrick. “Evolutionary Psychology: An Emerging New Foundation for the Psychology of Religion,” 108.

[29] Andrew B. Newberg and Stephanie K. Newberg. “The Neuropsychology of Religious and Spiritual Experience,” from *Handbook of the Psychology of Religion and Spirituality.* Raymond F. Paloutzian and Crystal L. Park, eds. New York and London: The Guilford Press (2005), 200.

[30] Andrew B. Newberg and Stephanie K. Newberg. “The Neuropsychology of Religious and Spiritual Experience,” 201.

[31] Andrew B. Newberg and Stephanie K. Newberg. “The Neuropsychology of Religious and Spiritual Experience,” 201.

[32] Ralph W. Hood, Jr, Peter C. Hill, and Bernard Spilka. *The Psychology of Religion: an Empirical Approach, Fourth Edition.* 22.

[33] Ralph W. Hood, Jr, Peter C. Hill, and Bernard Spilka. *The Psychology of Religion: an Empirical Approach, Fourth Edition.* 25.

[34] Zinnbauer and Pargament. “Religiousness and Spirituality,” 31.

[35] Ralph W. Hood, Jr, Peter C. Hill, and Bernard Spilka. *The Psychology of Religion: an Empirical Approach*, Fourth Edition. 332-333.

[36] Andrew B. Newberg and Stephanie K. Newberg. “The Neuropsychology of Religious and Spiritual Experience,” 210.

[37] Ralph W. Hood, Jr, Peter C. Hill, and Bernard Spilka. *The Psychology of Religion: an Empirical Approach*, Fourth Edition. 333-334.

[38] William P. Alston, *Perceiving God: The Epistemology of Religious Experience*, Cornell University Press (1991), 1.

[39] Alston, *Perceiving God*, 1.

[40] Alston, *Perceiving God*, 3.

[41] Following Alston in referring to the object of a spiritual experience as “God”, and putting aside for the moment the possibility that many people might choose to use other words or concepts, or maintain that their own experiences point to other kinds of spiritual reality.

[42] Alston, *Perceiving God*, 4.

[43] Alston, *Perceiving God*, 35.

[44] Alston, *Perceiving God*, 37.

[45] Alston, *Perceiving God*, 37-38.

[46] Alston, *Perceiving God*, 40.

[47] Alston, *Perceiving God*, 42.

[48] Alston, *Perceiving God*, 68-69.

[49] Alston, *Perceiving God*, 71.

[50] Alston, *Perceiving God*, 72.

[51] Alston, *Perceiving God*, 78.

[52] Alston, *Perceiving God*, 93.

[53] Alston, *Perceiving God*, 94.

[54] Alston, *Perceiving God*, 96-97.

[55] Alston, *Perceiving God*, 102.

[56] Alston, *Perceiving God*, 102-103.

[57] Alston, *Perceiving God*, 108.

[58] Carol Zaleski. *Otherworld Journeys: Accounts of Near-Death Experience in Medieval and Modern Times*. New York and Oxford: Oxford University Press (1987).

[59] Carol Zaleski. *Otherworld Journeys*, 159.

[60] Carol Zaleski. *Otherworld Journeys*, 156.

[61] Carol Zaleski. *Otherworld Journeys*, 156-157.

[62] Carol Zaleski. *Otherworld Journeys*, 158.

[63] Carol Zaleski. *Otherworld Journeys*, 177.

[64] Carol Zaleski. *Otherworld Journeys*, 159.

[65] Carol Zaleski. *Otherworld Journeys*, 164-165.

[66] Carol Zaleski. *Otherworld Journeys*, 167.

[67] Carol Zaleski. *Otherworld Journeys*, 170.

[68] Carol Zaleski. *Otherworld Journeys*, 175.

[69] Carol Zaleski. *Otherworld Journeys*, 175.

[70] Carol Zaleski. *Otherworld Journeys*, 176.

[71] Carol Zaleski. *Otherworld Journeys*, 180.

[72] Carol Zaleski. *Otherworld Journeys*, 182.

[73] Carol Zaleski. *Otherworld Journeys*, 187.

[74] Carol Zaleski. *Otherworld Journeys*, 192.

[75] Another occasional source of controversy is that members of religious traditions that use specific entheogenic substances often have specific rules or norms about who ought to use these substances, how, where, why, etc. For example, some traditions maintain that substances should only be used and experienced within the context of a particular ritual, led by a qualified religious leader. Using them in other contexts—as is now common—can be seen as disrespectful to the sanctity of the substances or of the experiences they produce, or as dangerous, or as likely to miss out on the spiritual benefit that ought to be obtained from them.

[76] There is also a magnetic brain stimulation device called the “God helmet” which reportedly produces similar effects, but which has proven difficult to replicate; see <https://en.wikipedia.org/wiki/God_helmet>.

[77] Rachael Peterson. “Taking mushrooms for depression cured me of my atheism.” *The Outline*, April 29, 2019.

[78] E.g., Rick Strassman, *DMT: The Spirit Molecule* (Rochester, Vermont: Park Street Press, 2001). Note that Strassman’s hypotheses about the pineal gland’s role in producing DMT have been severely questioned by other biomedical researchers.

[79] Griffiths RR, Hurwitz ES, Davis AK, Johnson MW, Jesse R (2019) “Survey of subjective ‘God encounter experiences’: Comparisons among naturally occurring experiences and those occasioned by the classic psychedelics psilocybin, LSD, ayahuasca, or DMT.” PLOS ONE 14(4): e0214377. https://doi.org/10.1371/journal.pone.0214377

[80] See https://en.wikipedia.org/wiki/Marsh_Chapel_Experiment. There was also a control group which received a non-psychoactive substance and attended the same service.

[81] This is an homage to a more famous but less wide-ranging book on the same subject by Aldous Huxley, *The Doors of Perception*. Both take their title from a line of William Blake: “If the doors of perception were cleansed every thing would appear to man as it is: Infinite.”

82“Though this be true, I must nevertheless here consider that I am a man, and that, consequently, I am in the habit of sleeping, and representing to myself in dreams those same things, or even sometimes others less probable, which the insane think are presented to them in their waking moments. How often have I dreamt that I was in these familiar circumstances, that I was dressed, and occupied this place by the fire, when I was lying undressed in bed? At the present moment, however, I certainly look upon this paper with eyes wide awake; the head which I now move is not asleep; I extend this hand consciously and with express purpose, and I perceive it; the occurrences in sleep are not so distinct as all this. But I cannot forget that, at other times I have been deceived in sleep by similar illusions; and, attentively considering those cases, I perceive so clearly that there exist no certain marks by which the state of waking can ever be distinguished from sleep, that I feel greatly astonished; and in amazement I almost persuade myself that I am now dreaming.” (*Meditations on First Philosophy*, I.5)

[83] Most dramatically, Cobb and Mal have an ongoing disagreement about whether or not they have finished waking up, which Cobb believes has led Mal to commit suicide. On another occasion, Saito forgets over the course of many subjective years (!) that he is dreaming, and Cobb has to convince him. See also https://en.wikipedia.org/wiki/Dream_argument and https://en.wikipedia.org/wiki/False_awakening.)

[84] Carol Zaleski. *Otherworld Journeys*, 176.

[85] “Blind men and an elephant.” *Wikipedia*. https://en.wikipedia.org/wiki/Blind_men_and_an_elephant. You may also enjoy Natalie Merchant’s musical rendition of John Godfrey Saxe’s famous poem on this theme: https://www.youtube.com/watch?v=3lyJLFMSVbY.

[86] In *Unsong*, the world has reverted to its basic spiritual nature in which Blake's vision of the Sun is simple everyday common sense. Compare Ted Chiang's "Hell is the Absence of God" and Ken Liu's "Single-Bit Error".

[87] We also briefly discussed Aumann's Agreement Theorem but concluded that we didn't understand how well it could apply to most disagreement in the real world—probably not so straightforwardly.

[88] However, some do—for example, one of Rick Strassman’s DMT subjects had a very unusual and intense experience, and afterward said that this experience had no real significance because it was entirely provoked by the drug. It remains worth acknowledging that powerful experiences are not always given a great significance by their experiencers. Some people remain radically skeptical and hold that the experiences merely *seemed* real or important in the moment, but should not ultimately be regarded this way.

[89] Heh.

[90] Several sources also maintain that spontaneous spiritual experience is extremely common, and that most people who experience some form of it simply don’t talk about it, perhaps because they find that it would be difficult to communicate, or because they don’t suppose that communicating it would have beneficial results. Many people keep their spiritual experiences private for fear of social consequences.

[91] Harris mentions the aphorism that taking entheogens is like launching one’s self on a rocket, while meditation is like raising a sail to the wind.

92This account of what Buddhism is “really” about is widespread in the West today. David Chapman has suggested that it would not be familiar to most Buddhists in history. That’s another interesting inquiry. Conversely, other religions have their own mystical and contemplative traditions that can offer elaborate and quite specific advice about how to pursue spiritual experiences, but that may not be widely known even to adherents of those religions—an interesting theme, for example, in Rodger Kamenetz’s *The Jew in the Lotus*.

Please Vote For ACC Winner

I’ve now posted all eight adversarial collaborations.

In case you missed any, you can find a list of them (with links) here.

If you have read all the collaborations, please vote on your favorite. This year I will decide the winner by popular vote; I don’t feel like putting my finger on the scale this time. I will give $2000 to the first place winner and $500 to second place. You can vote for your favorite collaboration **here**. No, you may not vote for the Grinch.

Thanks again to all participants, readers, and voters.

Please Take The 2020 SSC Survey!

Please take **the 2020 Slate Star Codex Survey**.

The survey helps me learn more about SSC readers and plan community events. But it also provides me with useful informal research data for questions I'm interested it, which I then turn into interesting posts. My favorite was 2018's Fight Me, Psychologists: Birth Order Effects Exist And Are Very Strong, which I think made a real contribution to individual differences psychology and which could not have happened without your cooperation. But last year I also got to debunk a myth about how mathematicians eat corn, fail to replicate supposed dangers of beef jerky, and test a theory of how fetishes form. I expect this year's research to be even more interesting.

The survey is open to anyone who has ever read a post on this blog before December 30 2019. Please don't avoid taking the survey just because you feel like you're not enough of a "regular". It will ask you how much of a "regular" you are, so there's no risk you'll "dilute" the results. The survey will stay open until mid-January, and I will probably be begging and harassing you to take it about once a week or so until then.

This year's survey is in two parts. Part I asks the same basic questions as previous years and should take about **ten minutes**. Part II asks more questions on research topics I'm interested in and should take about **fifteen minutes**. It would be great if you could take both parts, but if 25 minutes sounds like too much surveying to you, you can also just take Part I.

As always, the survey is plagued by fundamental limitations, poor technology, and my own carelessness, but a couple of things to watch for:

– Once you click a box on a Google form, you cannot unclick it – i.e. you can change your answer but you can't unanswer the question. If you click a box you didn't mean to, please switch your answer to "Other" if available; if not, then choose the most boring inoffensive answer that is least likely to produce surprising results. I realize how bad this is but there is apparently no way around it.

– By default, all responses will be included in a public dataset for anyone who wants to analyze them. Your responses will obviously not be attached to your name or any similarly blatant identifying information, and this year I'm going to further bin a couple of especially identifiable categories like age, but if you're the only supercentenarian Mongolian reader or something, you might still be identifiable. There is an option to make all your responses private; if identifiability bothers you, feel free to check it and you will not be included in the public dataset.

– Due to poring over a 5000 entry spreadsheet not actually being that much fun, I am not up for changing your answers after you submit them. Please do not email me asking me to do this. This includes your answer to the privacy question. Please figure out whether you want privacy before taking the survey.

That having been said, you are all great, and I super-appreciate any survey-filling-out you are willing to do. If you can donate about a half-hour, I hope I can pay you back in interesting findings and useful crowd-sourced life advice. I also plan to pay up to two randomly selected respondents back with a large monetary prize (with some caveats I hope you’ll find fun, see the last section for details). So:

Take **the 2020 Slate Star Codex Survey here**!

2020

- Why Doctors Think They’re The Best
- Hardball Questions For The Next Debate (2020)
- A Very Unlikely Chess Game
- What Intellectual Progress Did I Make In The 2010s?
- 2019 Adversarial Collaboration Winners
- Contra Contra Contra Caplan On Psych
- Last Chance To Take The Survey
- SSC Survey Results 2020
- Book Review Review: Little Soldiers
- Assortative Mating And Autism
- Book Review: Human Compatible
- Map Of Effective Altruism
- Suicide Hotspots Of The World
- Autogenderphilia Is Common And Not Especially Related To Transgender
- Welcome (?), Infowars Readers
- Confirmation Bias As Misfire Of Normal Bayesian Reasoning
- Addendum to “Targeting Meritocracy”
- Sleep Support: An Individual Randomized Controlled Trial
- Book Review: Just Giving
- Book Review: The Seven Principles For Making Marriage Work
- Socratic Grilling

- For, Then Against, High-Saturated-Fat Diets
- Book Review: Hoover
- Face Masks: Much More Than You Wanted To Know
- Legal Systems Very Different From Ours, Because I Just Made Them Up
- SSC Journal Club: MacIntyre On Cloth Masks
- Book Review: The Precipice
- SSCJC: Real World Depression Measurement
- Never Tell Me The Odds (Ratio)
- 2019 Predictions: Calibration Results
- A Failure, But Not Of Prediction
- Depression: The Olfactory Perspective
- The Amish Health Care System
- Employer Provided Health Insurance Delenda Est
- Give Yourself Gout For Fame And Profit
- Predictions For 2020
- Book Review Contest: Call For Entries
- Classified Thread 8
- Studies On Slack
- “My Immortal” As Alchemical Allegory
- Creationism, Unchallenged
- Bush Did North Dakota
- Book Review: Origin Of Consciousness In The Breakdown Of The Bicameral Mind
- Problems With Paywalls
- Take The New Nootropics Survey
- The Obligatory GPT-3 Post
- Wordy Wernicke’s
- The Vision Of Vilazodone And Vortioxetine
- Slightly Skew Systems Of Government

- **NYT Is Threatening My Safety By Revealing My Real Name, So I Am Deleting The Blog**
- **Update On My Situation**

Why Doctors Think They’re The Best

Ninety percent of drivers think they’re above-average drivers, ninety percent of professors think they’re above-average professors etc. The relevant studies are paywalled, so I don’t know if I should trust them. Our recent discussion of therapy books would make more sense if ninety percent of therapists believed they were above-average therapists. I don’t know about that one either.

But I am pretty sure ninety percent of doctors believe they’re above-average doctors. Here are some traps I’ve noticed myself falling into that might help explain why:

1. Your patients’ last doctor was worse than you. Think about it; if somebody has a good doctor, they’ll stay with them, and you will never see that patient. If somebody has a bad doctor, they’ll go see another doctor instead. That other doctor might be you. So your current patients’ last doctor will be worse than average. But this is where most of your chance to compare yourself with other doctors comes from: “my patient’s last doctor misdiagnosed them, but I got it right” or “my patient hated their last doctor but says I’m much better”. See also You Are Not Hiring The Top 1%.

2. Your patients love you. Similar to the above: if your patients love you, they will stick with you. If your patients hate you, they’ll leave. So over time, your practice will consist of patients who are very happy with your style of practicing medicine. Suppose some patients like very practical/efficient doctors, and others like very touchy-feely doctors. A practical/efficient doctor will collect a set of patients who prefer practical/efficient doctors and who praise them for how focused and sensible they are. A touchy-feely doctor will collect a set of patients who prefer touchy-feely doctors and praise them for how caring and personable they are. It will be obvious to the practical/efficient doctor that they're satisfying their patients' preferences much better than the touchy-feely doctor down the road would, and vice versa.

3. Patients often come to you, but never leave you. When a patient transfers to me, I have to review the case, confirm that this is an appropriate case for me, talk to the patient about why they left their last doctor and what we can do differently, and then keep seeing that patient. When a patient leaves me, I usually have no idea. It usually looks like the patient saying "I'll make an appointment for next month", and then not making that appointment. If I'm not careful, I never notice this; it's a dog that doesn't bark. Even if I do notice, patients stop coming all the time for all sorts of reasons. Sometimes they forget. Sometimes they feel better and decide they don't need you. Sometimes they move out of state and don't tell you. Sometimes they lose insurance and can't afford you, or change insurances to one that doesn't cover you. If you're really on top of things you try to call these people and ask what's going on, but sometimes they don't answer and sometimes they lie (it's really awkward to tell a doctor that you've fired them). So it naively feels to me like patients switch from the doctor down the road to me all the time, but nobody ever switches from me to the doctor down the road. I'm sure it's not true, but that's how it *feels*.

4. You've probably successfully treated most of your patients. Now pull all of the above together. Suppose a patient has a chronic disease like depression or diabetes. If you treat it successfully, they will love you and stay with you; if you fail, they will switch to another doctor (and you will never hear about it). Ten years later, you wake up and notice that most of your patients are success stories. But your patients usually describe their previous doctor as a miserable failure. Selection bias is a heck of a drug.

5. You know what you know, but you don’t know what you don’t know. Suppose each doctor makes errors at the same rate, but about different things. I will often catch other doctors’ errors. But by definition I don’t notice my own errors; if I did, I would stop making them! By “errors” I don’t mean stupid mistakes like writing the wrong date on a prescription, I mean fundamentally misunderstanding how to use a certain treatment or address a certain disease. Every doctor has studied some topics in more or less depth than others. When I’ve studied a topic in depth, it’s obvious to me where the average doctor is doing things slightly sub-optimally out of ignorance. But the topics I haven’t studied in depth, I assume I’m doing everything basically okay. If you go through your life constantly noticing places where other doctors are wrong, it’s easy to think you’re better than them.

6. Your victories belong to you, your failures belong to Nature. Sometimes I get a really difficult case, something nobody else has been able to figure out – and I absolutely nail it. I ride the high for days. I feel like a miracle-worker. Other times I get a difficult case nobody else has been able to figure out, and I can’t figure it out either. I don’t worry too much about it – some things are beyond the ken of modern medicine; *obviously* nobody can treat untreatable stuff. This is especially true in psychiatry. There are so many things we don’t understand, so many cases that are just inherently unresponsive to any kind of treatment we know about – that it’s hard to beat yourself up over any given failure. When you fail, you just say “I guess that was one of those treatment-resistant cases”. Maybe if I was a slightly better doctor I would have been able to figure out a way to treat it – but I will never know, and that’s not a natural way to think.

7. You do a good job satisfying your own values. Everybody wants to make people healthy and save lives, but there are other values that differ between practitioners. How much do you care about pain control? How much do you worry about addiction and misuse? How hard do you try to avoid polypharmacy? How do you balance patient autonomy with making sure they get the right treatment? How do you balance harms and benefits of a treatment that helps the patient’s annoying symptom today but raises heart attack risk 2% in twenty years? All of these trade off against each other: someone who tries too hard to minimize use of addictive drugs may have a harder time controlling their patients’ pain. Someone who cares a lot about patient autonomy might have a harder time keeping their medication load reasonable. If you make the set of tradeoffs that feel right to you, your patients will do better on the metrics you care about than other doctors’ patients (they’ll do better on the metrics the other doctors care about, but worse on yours). Your patients doing better on the metrics you care about feels a lot like you being a better doctor.

Some of these seem to generalize; do other jobs where customers select someone to work with have the same problems?

Hardball Questions For The Next Debate (2020)

[Previously: Hardball Questions (2016), More Hardball Questions (2016). I stole parts of the Buttigieg question from Twitter, but don't remember enough details to give credit, sorry]

Mr. Biden: Your son Hunter Biden was on the board of directors of Burisma, a Ukrainian energy company, during your vice-presidential term. The Ukrainian government was investigating Burisma for misdeeds, and Hunter was allegedly one of the targets of the investigation. President Trump alleges that you used your clout as VP to shut down the investigation into Hunter, which if true would constitute an impeachable abuse of power.

My question for you is: if your son had been a daughter, would you have named her Gatherer?

Mr. Bloomberg: You've been criticized as puritanical and self-righteous for some of your more restrictive policies, like a ban on large sodas. You seem to lean into the accusation, stating in a 2014 interview that:

> I am telling you, if there is a God, when I get to heaven I'm not stopping to be interviewed. I am heading straight in. I have earned my place in heaven. It's not even close.

Let's not focus on what this says about your humility, or about your religious beliefs. I want to focus on a different issue.

Despite spending $100 million in the first month of your presidential campaign, you are currently placed fifth – behind two socialists, a confused old man, and the mayor of

South Bend, Indiana. In, let’s not forget, an increasingly shaky effort to prevent President Donald J. Trump from winning a second term.

So my question for you is: what makes you so sure you’re not in Hell already?

Mayor Buttigieg: You are a gay Navy veteran. Your last name is “Buttigieg”. You are mayor of “South Bend”. And you first achieved prominence on the national stage for a *New York Times* editorial about your travels in the Horn of Africa, which includes the country of “Djibouti”.

My question is: is your campaign just the setup for a gay porno? Do you really think viewers want this much backstory?

Senator Warren: Despite your many years of service to the nation, media attention has focused on your claim to be descended from Native Americans. You told your former employer Harvard that you were of Native descent. Republicans accused you of trying to unfairly exploit affirmative action, but an investigation showed you did not benefit from any affirmative action at the time, leaving it unclear why you would do this.

More recently, you took a genetic test to establish your Native background. The test showed you did have a Native ancestor 6-12 generations back, but supporters were left baffled as to why you would take it or expect anyone to care. Conservatives used to the test to reignite the scandal around your Harvard employment, and progressives condemned you for promoting a view of race based on biology rather than culture or self-identification. The general consensus, again, was that you got no benefit from the test and it was unclear why you would do this.

The development of one of the algorithms that uses genetic information to determine racial background was called the “Warren Project” after its lead geneticist Jim Warren. Warren founded FamilyTreeDNA, a direct-to-consumer genetic testing company that continues to be a leader in genetic testing for ancestry, with about $16 million in revenue each year. This is relevant because Jim Warren is your ex-husband and the father of your children, who presumably stand to inherit a significant part of the FamilyTreeDNA fortune.

So my question for you is: is your campaign is just a publicity stunt to raise interest in genetic testing?

Senator Sanders: You are most famous for the 2016 incident where a bird landed on your podium mid-rally. Supporters have reasonably connected this to the ancient Roman practice of augury, where leaders were chosen by the number of bird-related omens surrounding them.

But auguries can be hard to interpret. For example, during the founding of Rome, Romulus and Remus agreed to use augury to determine which of them should lead the new city. The two of them went out and watched for ominous birds. First, Remus saw six vultures, which he interpreted as strong evidence that he should lead. But then Romulus saw twelve vultures. The two argued, with Remus’ claim resting on having seen vultures first, and Romulus’ claim resting on the theory that more vultures = more leadership. One thing led to another, Romulus killed Remus, and Rome ended up building the greatest empire in history. This firmly established the principle that even if one person sees birds first, another person who sees more birds may still be the rightful leader, if he sees enough of them.

So my question for you is: it’s been four years. How many birds would have to land on Donald Trump before you admit he would make a better president than you?

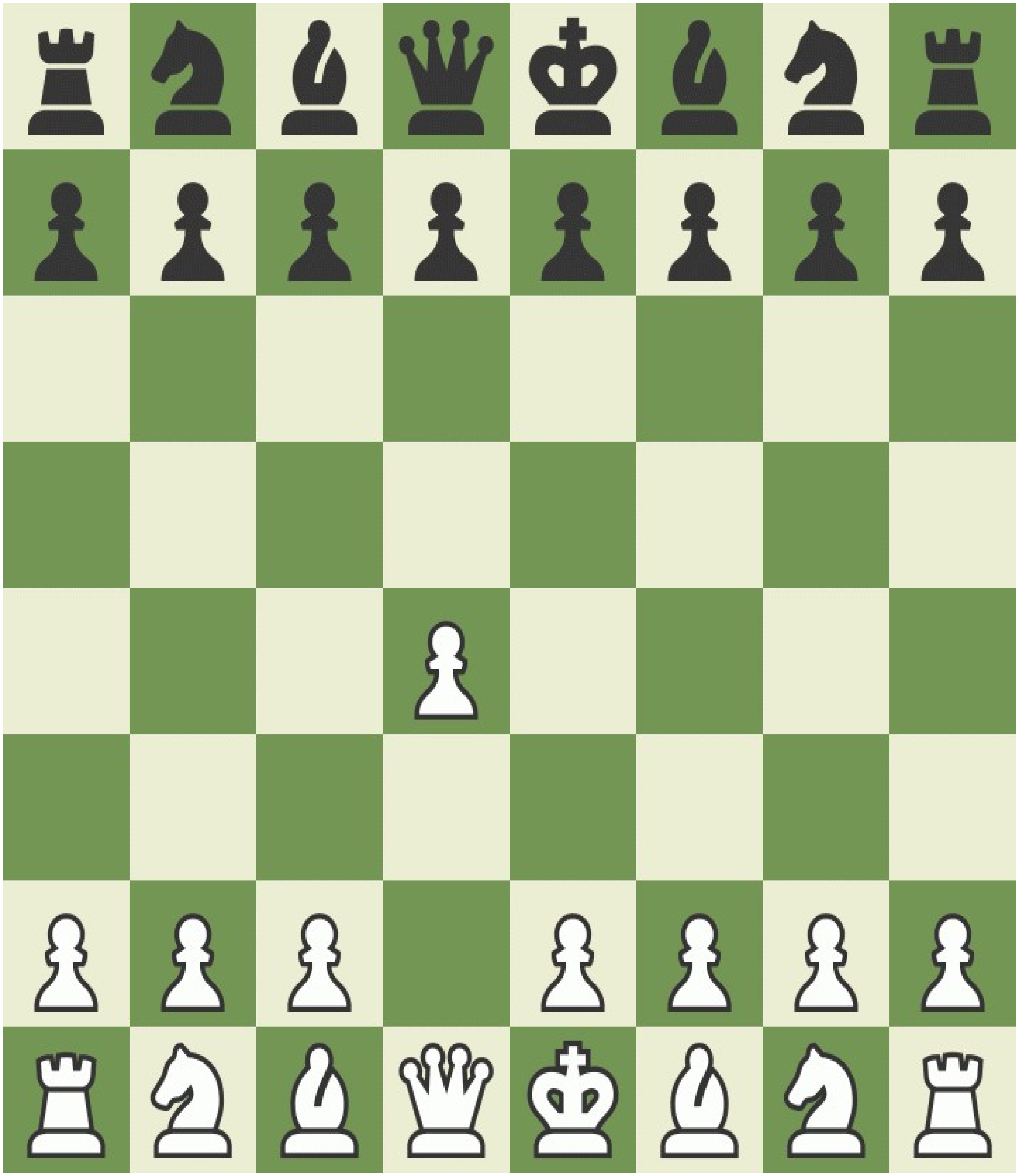
Mr. Yang: You’ve sparked interest with your proposal of a Universal Basic Income (UBI), a no-strings-attached $1000/month transfer to every US citizen. Experts say it’s totally infeasible, but “UBI forever!” certainly makes for a stirring rallying cry.

On the other hand are people who complain your proposal isn’t a real UBI. UBI needs to be enough to live on, but $1000/month wouldn’t even get people all the way to the federal poverty line. In more expensive regions like coasts and cities, it would be even worse. “UBI forever!” might be a good rallying cry, but “UBI (below a real UBI) forever!” is a little less rousing. Then again, US states make their mottos sound portentious by converting them to Latin; maybe that would work for you too.

So do you think a good slogan for your campaign would be “Semper ubi sub ubi”?

A Very Unlikely Chess Game

Almost 25 years after Kasparov vs. Deep Blue, another seminal man vs. machine matchup:



Neither competitor has much to be proud of here. White has a poor opening. Black screws up and loses his queen for no reason. A few moves later, white screws up and loses his rook for no reason. Better players will no doubt spot other humiliating mistakes. But white does eventually eke out a victory. And black does hold his own through most of the game.

White is me. My excuse is that I only play chess once every couple of years, plus I'm entering moves on an ASCII board I can barely read.

Black is GPT-2. Its excuse is that it's a text prediction program with no concept of chess. As far as it knows, it's trying to predict short alphanumeric strings like "e2e4" or "Nb7". Nobody told it this represents a board game. It doesn't even have a concept of 2D space that it could use to understand such a claim. But it still captured my rook! Embarrassing!

Backing up: last year, I wrote GPT-2 As Step Toward General Intelligence, where I argued that the program wasn't just an essay generator, it was also kind a general pattern-recognition program with text-based input and output channels. Figure out how to reduce a problem to text, and you can make it do all kinds of unexpected things.

Friend-of-the-blog Gwern Branwen has been testing the limits of this idea. First he taught GPT-2 to write poetry. Some of it was pretty good:

> Fair is the Lake, and bright the wood,
> With many a flower-full glamour hung:
> Fair are the banks; and soft the flood
> With golden laughter of our tongue.

For his next trick, he found a corpus of music in “ABC notation”, a way of representing musical scores as text. He fed it to GPT-2 and got it to write folk songs for him. I’m a fan:

Last month, I asked him if he thought GPT-2 could play chess. I wondered if he could train it on a corpus of chess games written in standard notation (where, for example, e2e4 means “move the pawn at square e2 to square e4”). There are literally millions of games written up like this. GPT-2 would learn to predict the next string of text, which would correspond to the next move in the chess game. Then you would prompt it with a chessboard up to a certain point, and it would predict how the chess masters who had produced its training data would continue the game – ie make its next move using the same heuristics they would.

Gwern handed the idea to his collaborator Shawn Presser, who had a working GPT-2 chess engine running *within a week*:

> GPT-2 chess is promising. After an hour of training, 1.5B is pretty good at opening theory.
>
> Longer sequences tend to fail due to invalid moves, but this shows it's possible in principle to make a GPT-2 chess engine.
>
> And maybe after more training it'll make fewer invalid moves. pic.twitter.com/DqC4WiPfHV
>
> — Shawn Presser (@theshawwn) January 1, 2020
>
> I'll post some games, up to the point it generates an invalid move (which seems to happen around move 11).
>
> Paste this into https://t.co/wWGpVu9ko6

1.e4 c5 2.Nf3 e6 3.d4 cxd4 4.Nxd4 a6 5.Bd3 Nf6 6.Nc3 d6 7.O-O Be7 8.f4 O-O 9.Kh1 Nbd7 pic.twitter.com/8Bl2ijZiCZ

— Shawn Presser (@theshawwn) January 1, 2020

After a day of training (2.4M examples), GPT-2 1.5B can reach move 14 with no invalid moves.

1.e4 e5 2.Nf3 d6 3.d4 exd4 4.Qxd4 a6 5.Be2 Nf6 6.O-O Be7 7.Re1 O-O 8.c3 b5 9.a4 Bb7 10.axb5 axb5 11.Nbd2 Re8 12.h3 g6 13.Ra5 Qd7 14.Ng5 c5 pic.twitter.com/2XuH6iLaD5

— Shawn Presser (@theshawwn) January 2, 2020

It can reach midgame by removing invalid moves.

1.e4 c5 2.Nf3 d6 3.Bb5+ Nd7 4.O-O a6 5.Be2 b6 6.a4 e6 7.d4 Be7 8.c3 9.Nbd2 Ne5 10.Nxe5 dxe5 11.12.13.f4 Bb7 14.Bd3 g6 15.Nf3 16.17.Re1 18.Qe2 Bf8 19.Bd2 20.Qf2 Qxd4 21.22.23.Rf1 Qxc3 24.Qxc5 Bd6 25.Qc7 26.Qd8+ 27.Qd6 Rd8 pic.twitter.com/3uyPaP9LHt

— Shawn Presser (@theshawwn) January 2, 2020

GPT2 Chess update: I wrote some code to calculate the probability of all valid chess moves. It can reach endgame now. https://t.co/QQzhZJmgQ9

It starts to blunder every game at around move 13. We suspect it’s losing track of board state. (It’s trained solely on PGN notation.)

— Shawn Presser (@theshawwn) January 4, 2020

I am preparing to release a notebook where you can play chess vs GPT-2. If anyone wants to help beta test it:

> 1. visit https://t.co/CpWrFvtnY2
> 2. open in playground mode
> 3. click Runtime -> Run All
> 4. Scroll to the bottommost cell and wait 6 minutes
>
> If you get stuck, tell me.
>
> — Shawn Presser (@theshawwn) January 6, 2020

You can play against GPT-2 yourself by following the directions in the last tweet, though it won’t be much of a challenge for anyone better than I am.

This training explains the program’s strengths (good at openings) and weaknesses (bad when play deviates from its expectations). For example, ggreer analyzes why GPT-2 lost its queen in the game above. By coincidence, my amateurish flailing resembled a standard opening called the Indian Game. GPT-2 noticed the pattern and played a standard response to it. But the resemblance wasn’t perfect, so one of GPT-2’s moves which would have worked well in a real Indian Game brought its queen where I could easily capture it. I don’t want to conjecture on how far “mere pattern-matching” can take you – but you will at least need to be a better pattern-matcher than this to get very far.

But this is just what a friend of a friend managed to accomplish in a few days of work. Gwern stresses that there are easy ways to make it much better:

> Obviously, training on just moves with the implicit game state having to be built up from scratch from the history every time is very difficult – even MuZero at least gets to see the entire game state at every move when it’s trying to predict legal & good next moves, and depends heavily on having a recurrent state summarizing the game state.

> Maybe rewriting games to provide (state,action) pairs will make GPT-2 work much better.

What does this imply? I’m not sure (and maybe it will imply more if someone manages to make it actually good). It was already weird to see something with no auditory qualia learn passable poetic meter. It’s even weirder to see something with no concept of space learn to play chess. Is any of this meaningful? How impressed should we be that the same AI can write poems, compose music, and play chess, without having been designed for any of those tasks? I still don’t know.

What Intellectual Progress Did I Make In The 2010s?

One of the best parts of writing a blog is being able to answer questions like this. Whenever I felt like I understood new and important, I wrote a post about it. This makes it easy to track what I learned.

I think the single most important thing I discovered this decade (due to a random comment in the SSC subreddit!) was the predictive coding theory of the brain. I started groping towards it (without knowing what I was looking for) in Mysticism And Pattern-Matching, reported the exact moment when I found it in It's Bayes All The Way Up, and finally got a decent understanding of it after reading Surfing Uncertainty. At the same time, thanks to some other helpful tips from other rationalists, I discovered Behavior: The Control Of Perception, and with some help from Vaniver and a few other people was able to realize how these two overarching theories were basically the same. Discovering this area of research may be the best thing that happened to me the second half of this decade (sorry, everyone I dated, you were pretty good too).

Psychedelics are clearly interesting, and everyone else had already covered all the interesting pro-psychedelic arguments, so I wrote about some of my misgivings in my 2016 Why Were Early Psychedelicists So Weird?. The next step was trying to fit in an understanding of HPPD, which started with near-total bafflement. Predictive processing proved helpful here too, and my biggest update of the decade on psychedelics came with Friston and Carhart-Harris' Relaxed Beliefs Under Psychedelics And The Anarchic Brain, which I tried to process further here. This didn't directly improve my understanding of HPPD specifically, but just by talking about it a lot I got a subtler picture where lots of people have odd visual artifacts and psychedelics can cause slightly more (very rarely, significantly more) visual artifacts. I started the decade thinking that “psychedelic insight” was probably fake, and ended it believing that it is probably real, but I still don’t feel like I have a good sense of the potential risks.

In mental health, the field I am supposed to be an expert on, I spent a long time throwing out all kinds of random ideas and seeing what stuck – Boorsboom et al’s idea of Mental Disorders As Networks, The Synapse Hypothesis of depression, etc. Although I still think we can learn something from models like those, right now my best model is the one in Symptom, Condition, Cause, which kind of sidesteps some of those problems. Again, learning about predictive processing helped here, and by the end of the decade I was able to say actually useful things that explained some features of psychiatric conditions, like in Treat The Prodrome. Friston On Computational Mood might also be in this category, I’m still waiting for more evidence one way or the other.

I also spent a lot of time thinking about SSRIs in particular, especially Irving Kirsch (and others’) claim that they barely outperform placebo. I wrote up some preliminary results in SSRIs: Much More Than You Wanted To Know, but got increasingly concerned that this didn’t really address the crux of the issue, especially after Cipriani et al (covertly) confirmed Kirsch’s results (see Cipriani On Antidepressants). My thoughts evolved a little further with SSRIs: An Update and some of my Survey Results On SSRIs. But my most recent update actually hasn’t got written up yet – see the PANDA trial results for a preview of what will basically be

"SSRIs work very well on some form of mental distress which is kind of, but not exactly, depression and anxiety".

One place I just completely failed was in understanding the psychometrics of autism, schizophrenia, transgender, and how they all related to each other and to the normal spectrum of variation. I kind of started this program with Why Are Transgender People Immune To Optical Illusions? (still a good question!), fumbled around by first-sort-of-condemning and then sort-of-accepting the diametrical model of autism and schizophrenia, and then admitting I just didn't know what was going on in this area and not talking about it much more. I still sometimes have thoughts like "Is borderline the opposite of autism?" or "Are schizoid people unusually charismatic, unusually uncharismatic, or somehow both?", and I still have no idea how to even begin answering them. Autism And Intelligence: Much More Than You Wanted To Know at least helped address a very tangentially related question and is probably the closest thing to a high point this decade gave me here.

The Nurture Assumption shaped my 2000s views of genetics and development. Ten years later, I'm still trying to process it, and in particular to square the many behavioral genetics studies showing nonshared environment doesn't matter with the many other studies suggesting it does (see eg The Dark Side Of Divorce and Shared Environment Proves Too Much). I think I started to get more of handle on attachment theory and cPTSD as both being different aspects of the same basic predictive processing concept of "a global prior on the world being safe" – see Mental Mountains and Evolutionary Psychopathology for two different ways of approaching this concept. This made me conclude that I might have been wrong about preschool (though see also Preschool: Much More Than You Wanted To Know). Honestly I am still confused about this. The one really exciting major good update I made about genetics this decade was understanding and fully internalizing the omnigenic model.

One of the big motivating questions I keep coming back to again and again is – what the heck is “willpower”? I started the decade so confused about this that I voluntarily bought and read Baumeister and Tierney’s book *Willpower* and expected it to be helpful. I spent the first few years gradually internalizing the lesson (which I learned in the 2000s) that Humans Are Not Automatically Strategic (see also The Blue-Minimizing Robot as a memorial to the exact second I figured this out), and that hyperbolic discounting is a thing. Since then, progress has been disappointing – the only two insights I can be even a little happy about are understanding perceptual control theory and Stephen Guyenet’s detailed account of how motivation works in lampreys. If I ever become a lamprey I am finally going to be totally content with how well I understand my motivational structure, and it’s going to feel *great*.

Speaking of Guyenet, if nothing else this last decade has taught us that Gary Taubes did not solve all of nutrition in 2004, that Atkins/paleo/keto are good for some people and bad for others, and that diet is still hard. See the various Guyenet vs. Taubes and Taubes vs. Guyenet posts, and my 2015 The Physics Diet on where I was at that point. So what *is* going on with diet? Compressing an entire decade’s worth of research into two words, the key phrase seems to be “set point” (which, credit to Taubes, he was one of the first people to popularize). See eg Anorexia And Metabolic Set Point and Del Giudice On The Self-Starvation Cycle. But what is the set point and how does it get dysregulated? See my book review of The Hungry Brain for the best answer to that I have now (not so good). This whole mess helped me get a better understanding of Contrarians, Crackpots, and

Consensus, and eventually ended up with me Learning To Love Scientific Consensus.

In terms of x-risk: I started out this decade concerned about The Great Filter. After thinking about it more, I advised readers Don't Fear The Filter. I think that advice was later proven right in Sandler, Drexler, and Ord's paper on the Fermi Paradox, to the point where now people protest to me that nobody ever *really* believed it was a problem. AI has been the opposite – I feel like the decade began with people pooh-poohing it, my AI Researchers On AI Risk was part of a large-scale effort to turn the tide, and now it's more widely accepted as an important concern. At the same time, the triumphs of deep learning has made things look a little different – see How Does Recent AI Progress Affect The Bostromian Paradigm? and Reframing Superintelligence – and I'll be reviewing *Human Compatible* soon. I also got some really great insights on what "human-level intelligence" means from the good people at AI Impacts, which I wrote up as first Where The Falling Einstein Meets The Rising Mouse and later Neurons And Intelligence: A Bird-Brained Perspective (see also Cortical Neuron Number Matches Intuitive Perceptions Of Moral Value Across Animals and all the retractions and meta-retractions thereof). Overall I think I've updated a little (though not completely) towards non-singleton scenarios and not-super-fast takeoffs, which combined with the increased amount of effort being put into this area is cause for a little more optimism than I had in 2010. I know some smart people disagree with me on this.

In the 2000s, people debated Kurzweil's thesis that scientific progress was speeding up superexponentially. By the mid-2010s, the debate shifted to whether progress was actually slowing down. In Promising The Moon, I wrote about my skepticism that technological progress is declining. A group of people including Patrick Collison and Tyler Cowen have since worked to strengthen the case that it is; in 2018 I wrote Is Science Slowing Down?, and late last year I conceded the point. Paul Christiano helped me synthesize the Kurzweillian and anti-Kurzweillian perspectives into 1960: The Year The Singularity Was Cancelled.

In 2017, I synthesized some thoughts that had been bouncing around about rising prices into Considerations On Cost Disease, still one of this blog's most popular posts. I felt like early responses were pretty weak, although they brought up a few interesting points on veterinary medicine, cosmetic medicine, and other outliers that I still need to transform into a blog post; Alon Levy's work on infrastructure in particular has also been great. The first would-be-general-answer that made me sit up and take notice was Alex Tabarrok's book (link goes to my review) The Prices Are Too Damn High – but I explain there why I don't think it can be the full answer. The most recent thing I learned (tragically underhighlighted in my wage stagnation post) is that a lot of apparent wage stagnation is due to cost disease – consumer services ballooning in cost means the consumer inflation index rises faster than the business inflation index, productivity gets measured by business inflation, wages get measured by consumer inflation, and so it looks like productivity is outpacing wages. This is still only half of the apparent decoupling, but it's still a big deal.

The highlight/lowlight of the decade in social science was surely the replication crisis. My first inkling that something like this might exist was in December 2009, from the Less Wrong post Parapsychology: The Control Group For Science. There were a couple of years where people were trying to figure out how bad the damage was; of these, my 90% Of All Claims About Problems With Medical Studies Are Wrong was more optimistic, and my slightly later The Control Group Is Out Of Control was more pessimistic (I still stand by both).

As the decade continued, I think we got better about realizing that many to most older studies were wrong, in a way that didn't make us feel like total Cartesian skeptics or like we were going to have to throw out evolution or aspirin or any of the things on really sound footing. After that it just became fun: my "acceptance" stage of grief produced some gems like 5-HTTLPR: A Pointed Review.

On SSC, I particularly examined some of the replication issues of growth mindset. I started in 2015 by pointing out that the studies seemed literally unbelievable, but so far nobody had tried attacking them. I claim to have been way ahead of the curve on this one – if you don't believe me, just read the kind of pushback I got. But by 2017, that situation had changed – Buzzfeed posted an article that called the field into question, but still without clear negative evidence. Finally, over the past few years, the negative studies have come pouring in, accented by supposedly "positive" studies by Dweck & co showing effect sizes only a tiny fraction of what they had originally claimed. The latest research (can't find it right now) is that praising students for effort rather than for ability has no effect on how hard-working or successful they are, debunking the original headline result that got most people interested in the field and nicely closing the circle.

In 2010 I worked with a medical school professor who studied the placebo effect and realized I didn't understand it at all. Over the past few years I gradually became more convinced of the heterodox position of Gøtzsche and Hróbjartsson, who believe placebo effect doesn't apply to anything except pain and a few other purely mental phenomena (The Placebo Singers, Powerless Placebos). I've since become less convinced that's true (just today I treated a patient who I'm pretty sure has psychosomatic vomiting from what he falsely believes was a medication side effect, and if belief can cause vomiting, surely it can also treat it). As with so many other things, it was predictive processing to the rescue – see section IV part 7 of my Surfing Uncertainty review. I now think I have a pretty good understanding of how placebos can treat both purely mental conditions and conditions heavily regulated by the nervous system, while still mostly sticking to Gøtzsche and Hróbjartsson's findings.

I started this decade confused about how to understand ethics given all the paradoxes of utilitarianism. I'm still 90% as confused now as I was then, but I still feel like I've made some progress. A lot of my early thinking involved folk decision theory and contractualism – how would you act if you expected everyone else to act the same way? I explored the edges of this idea in You Kant Dismiss Universalizability and Invisible Nation. I'm not how much it helped my search for metaethical grounding, but it helped me get a more robust understanding of liberalism and clarify my views on some practical questions, eg Be Nice, At Least Until You Can Coordinate Meanness and The Dark Rule Utilitarian Argument For Science Piracy. In general I think this has given me a more cautious theory of decision-making that's occasionally (and terrifyingly) set me against other more anti-Outside-View rationalists. I think the most important shift in my understanding of ethics this decade was the one I wrote up in Axiology, Morality, Law (formerly titled "Contra Askell On Moral Offsets"), which isn't related to grounding utilitarianism at all but sure helps make the problem less urgent

Despite my better judgment, I waded into politics a lot this decade. I Can Tolerate Anything Except The Outgroup produced this blog's first "big break", but it admitted it didn't really understand the factors underlying "tribe". Since then Albion's Seed helped provide another piece of the puzzle, and a better understanding of class provided another. I went a little further discussing why tribes have ideologies associated with them in The Ideology Is Not The Movement, how that is like/unlike religion in Is Everything A Religion?, and hammered it home unsubtle-ly in Gay Rites Are Civil Rites.

I wrote the Non-Libertarian FAQ sometime around 2012 and last updated it in 2017. Sometime, possibly between those dates, I read David Friedman's A Positive Account Of Property Rights, definitely among the most important essays I've ever read, and got gold-pilled (is that a term? It should be a term). I've since been trying to sort this out with things like A Left-Libertarian Manifesto, and trying to move them up a level as Archipelago. James Scott's Seeing Like A State and David Friedman's Legal Systems Very Different From Ours were also big influences here. Like all platitudes, "government is a hallucination in the mind of the governed" is easy to understand on a shallow level but fiendlishly complicated on a deep level, but I feel like all of these sources have given me a deep understanding of exactly how it's true.

The rightists (especially Moldbug) get the other half of the credit for helping me understand Archipelago, and also deserve kudos for teaching me about cultural evolution. My first attempts to engage with this topic were nervous and halting – see eg The Argument From Cultural Evolution. I got a much better feel for this after reading The Secret Of Our Success, and was able to bring this train of thought back to its right-wing roots Addendum To Enormous Nutshell: Competing Selectors. I'm grateful to the many rightists who argued about some of these points with me until they finally stuck.

I had more trouble engaging with leftists. I started with Does Class Warfare Have A Free-Rider Problem, and it took me way too long to figure out that this was one of the major questions sociology was asking, and that “an answer” would look less like “your game theory analogy is missing this one variable” and more like a whole library full of books on what the heck society was. Later the same engagement produced Conflict Vs. Mistake, which I am informed is still unfair and partially inaccurate, but which (take my word for it) is a heck of a lot better than the stuff I was thinking *before* I wrote it. More recently I’ve been trying to figure out a sympathetic account of activism (as opposed to the unsympathetic account that it’s virtue signaling and/or people who are really bad at figuring out what things are vs. aren’t effective). You can sketch the outline at Respectability Cascades and Social Censorship: The First Offender Model, and I’ll sketch the whole thing out sometime when I have enough emotional energy to deal with the kind of people who will have opinions on it.

I also had to grapple with the sudden rise of social justice ideology. I’m proud of my work on gender differences – both what I learned, how I wrote it up, and the few bits of original research I did (eg Sexual Harassment Levels By Field). My knowledge and claims started off kind of weak (Gender Differences Are Mostly Not Due To Offensive Attitudes), but I eventually feel like I got a really great evidence-based basically-airtight theory of what is going on with gender imbalances in different fields, which I posted most of in Contra Grant On Exaggerated Differences (I’m still thankful for the commenter who solved that one remaining paradox about math majors). And despite all the mobs and vitriol I think sound science has basically triumphed here – I was delighted to recently see as mainstream a blog as Marginal Revolution recently publish, without any caveats or double-talk, a post called Sex Differences In Personality Are Large And Important and get basically no pushback. I was a lot more pessimistic around 2017 or so and described some thoughts on how to make a strategic retreat in Kolmogorov Complicity And The Parable Of Lightning, which I still think is relevant in some areas. But I actually start the new decade really optimistic – I haven't written up an explanation of why, but careful readers of New Atheism: The Godlessness That Failed may be able to figure it out, especially if they apply some of the same metrics I used there to track how social justice terms have been doing recently.

Upstream of politics, I think I got a better understanding of… game theory? Complex system dynamics? The most important post here was Meditations On Moloch; the sequel/expansion, whose thesis I have yet to write up in clear prose, is The Goddess Of Everything Else. Reading Inadequate Equilibria was also helpful here.

My understanding of "enlightenment" went from total mystical confusion to feeling like I have a pretty good idea what claims are being made, and mostly believing them. This line of thinking started with the Mastering The Core Teachings Of The Buddha review, and then was genuinely helped by Vinay Gupta's contributions summed up in Gupta On Enlightenment, despite the disaster in the comments of that post. From there I progressed to reading The Mind Illuminated, and Is Enlightenment Compatible With Sex Scandals led me to discover The PNSE Paper, which as much as anything else helped ground my thinking here (the comments there were pretty good too).

And thanks to all of you who took the survey, I went from skepticism of birth order effects to saying Fight Me, Psychologists: Birth Order Effects Exist And Are Very Strong. This was bolstered by Eli Tyre and Bucky's posts on Less Wrong about birth order in mathematicians and physicists respectively. Last year I expanded on that with a post on how birth order responded to age gaps (somewhat updated and modified here, thanks Bucky). Once this year's survey results are in I expect to have a lot more data on exactly what causes birth order effects and maybe how to deal with them. If you haven't taken the SSC survey this year, consider this your reminder to do it **here**.

Not many of these were total 180 degree flips in my position (though birth order, preschool, psychedelic insight, and the rate of scientific progress are close). And not many of them completely resolved a big question that had been bothering me before (though the Fermi Paradox paper, omnigenic model, and animal neuron work did). A few of them confirmed things I had only suspected before (growth mindset, gender imbalances, diet). Many of them feel like what MIRI calls "deconfusion", turning a space full of unknown unknowns to one where you feel like you have a decent map of where the major problems are and what it would feel like to solve them. The enlightenment research seems to fit here – I went from "I have no idea how to even think about this question or whether it's all fake" to "I don't know exactly what's going on here, but I know what needs to be explained, and it looks like the explanation will have a shape that fits nicely into the rest of my ontology."

There's an argument that I should learn less each decade, since I'll be picking higher and higher fruit. My own knowledge can advance either because civilization advances and I hear about it, or because I absorb/integrate older knowledge that I hadn't noticed before. Civilization advances at a decade per decade (or maybe less; see the Cowen & Southwood paper above), but each year it becomes harder and harder to find relevant older knowledge that I haven't integrated yet. I plausibly only have five more decades to live, and I don't think I'd be happy only advancing five times this amount over the rest of my life, let alone less than that.

But I notice I only started SSC about halfway through the decade, and that my progress picked up a lot after that. I don’t think it’s just recall bias from being able to track myself better. I think being able to put ideas out there and have you guys comment on them and link me to important resources I might have missed has been great for me. I only started taking full advantage of that around 2015; this decade I have a head start. And maybe I’ll discover other useful tools that will speed things up further.

Thanks for sticking around with this blog, and have a happy third decade of the twenty-first century.

2019 Adversarial Collaboration Winners

Thanks to everyone who participated and/or voted in the 2019 Adversarial Collaboration Contest. And the winner is…

…

…

Adrian Liberman and Calvin Reese, for Does Calorie Restriction Slow Aging?.

An extraordinarily close second place (26.9% vs. 26.2% of votes) goes to David G and Froolow, for Is Eating Meat A Net Harm?.

Both of these did great research and were written up well. I especially like them as winners because they have such different strengths.

The calorie restriction collaboration was carefully focused on a factual question. I think this is a promising model for adversarial collaborations, and that others failed the further they deviated from this. For example, the circumcision collaboration did a good job assessing the quantifiable benefits and harms of the practice, but it turned out that most people who disagreed about it weren't disagreeing because they assessed quantifiable benefits and harms differently. The abortion collaboration ended up in a similar place. By focusing on a topic where there really was debate about what the research showed, and by hitting the lit review portion out of the park, Adrian and Calvin helped deconfuse a lot of previously confused people.

And the meat collaboration managed to succeed without being like this at all! It was unabashedly just a lengthy review of every single plausible argument for and against vegetarianism, and bulldozed over the immense difficulties with this approach by putting in more work than any reasonable person would have thought possible. And if it didn’t get quite as many votes as calorie restriction, it won on another metric – here are some of voters’ comments (plus some extra from the blog):

> – After this, I expect to restrict more strictly to chicken and fish, and alternate more aggressively towards fish as a hedge against the possibility I’m undervaluing chicken sapience.
>
> – Got me to significantly change my diet (at least over the past few weeks) towards more fish and much less chicken.
>
> – I read it, and I changed what I believe and how I changed my life accordingly.
>
> – My final pick is not necesarily based on the article‘s excellence but rather due to me going to change my eating behaviour, which I find an impressive thing Form an article to do.
>
> – This has practical implications for my life. I’ve stopped eating pork because of it (baby steps, working towards less meat generally)
>
> – I actually resolved to stop eating chicken (the only meat I can regularly eat, due to dietary restrictions) based on that piece, so I’d say it was pretty effective in informing me about things.

> – It convinced me not to buy chicken that isn't organic/free-range
>
> – This article will result in all my family eating less meat. It's actually going to change our lives, health, and the environment!
>
> – As a result of the adversarial collaboration on the ethics of eating meat, have dramatically scaled back my meat consumption to probably 20% of my previous value, and the meat I do eat now is almost exclusively fish and invertebrates.
>
> – Learning about the harm of factory farming from their dispassionate and empirical analysis has prompted me to greatly reduce my red meat and poultry consumption.

Some voters brought up a reasonable complaint: the end result ended up being pretty (though not completely) pro-vegetarian. How do we tell the difference between "a good faith effort by intelligent people naturally converges on vegetarianism" vs. "the anti-vegetarian collaborator slacked off"?

In this case, we tell because the anti-vegetarian collaborator posted a comment about his thought process and what convinced him. But there were other cases where people had the same question, and still other cases where one collaborator did a good job representing their own anti-X position, but other people were anti-X for different reasons that didn't get represented.

If I had infinite resources, I would fund adversarial collaborations between well-known and universally-recognized intellectuals on different sides of a topic, who everybody trusted to stick to their guns. As it is, I can only say I'm delighted to have stumbled into the one part of the world where “people are too likely to change their mind when presented with new evidence on controversial issues” is a problem.

Some thoughts on the other collaborations:

Circumcision: I loved this one. I’d never seen a good assessment of exactly what health risks circumcision was supposed to prevent, and I didn’t know how weak the evidence was that the foreskin helps with sexual pleasure. But the conclusion ended up being “the quantifiable benefits of circumcision are nonzero but pretty low; the quantifiable harms are not obviously distinguishable from zero but who knows”, which leaves a lot of space for people’s ethical intuitions, which turned out to be REALLY STRONG. One reader said they were going to boycott my blog from now on for not having no-platformed this ACC, and a few others seemed only slightly less angry. On the other hand, it also did better than average among voters, so good job there. I take a small amount of blame for this one not being more popular – I retitled it to be about the ethics of circumcision, whereas the original title had been about benefits vs. harms. But I think it’s naturally hard to write something about benefits and harms without it sounding like you’re talking about ethics, and in this case the ethics were too complicated to fit in the model provided for them. Some positive comments from the survey: “This…actually changed my opinion from circumcision being mildly ethically wrong back to neutral”, “I gained a much more nuanced understanding of the benefits position to the point that my mind was changed to be in favor (maybe too strong), or at least not opposed to, it for developing countries”, “It tidily presented the pros and cons and presented a lot of useful information, with a clear conclusion. It shifted my thinking the most of all of them.”

Space Colonization: You guys presented a lot of evidence for one side, then at the end switched to the opposite side based on a one paragraph explanation of something you'd never brought up before. If that was your crux, I wish you had analyzed it in more depth. If the whole point is to make something that can't be defunded, couldn't the government (or whoever) give the money to a private foundation with really good trustees, no takebacks? Maybe there's a problem with that idea, I don't know, but if you're going to make defundability the center of your conclusion, I wish you had examined it more closely. Some positive comments from the survey: "Excellent selection of question, manages to present both sides fairly and come to an insightful conclusion", "I think this ACC did the best job of covering the entire scope of the question they assigned to themselves, while still presenting a shared conclusion", "Interesting non-obvious conclusion, subject I care a lot about, pretty pictures".

Gene Editing: This one seemed to spend a lot of time on very knowledgeable and very well-cited assessments of the current state of the technology and how and why it worked, but didn't really get around to assessing the "should" question in the title. It also had a few factual missteps – maybe no more than the others, but more obvious since it was so fact-based. While it was an impressive work of scholarship I'm not sure it came together as an adversarial collaboration. Some positive comments from the survey: "Very nicely presented ACC. It was thought provoking and totally enrapturing!", "well-reasoned collaboration on a difficult question", "Great, nuanced answer to a complex question", "This collaboration caused me to reconsider my enthusiasm for CRISPR based on the narrative provided in most press releases. The topic is much more involved than I'd initially realized."

Abortion: An adversarial collaboration on a completely moral question – you guys didn't make this easy for yourself, did you? I don't think you made any particular missteps given the difficult task you set yourselves, but this is another one that I feel like didn't quite come together. Some positive comments from the survey: "Most interesting (and politically relevant) topic, plus it seems icerun's position actually shifted somewhat by virtue of having to marshall arguments for it, proving the whole endeavor to be more than just an exercise in futility", "I thought it best captured the spirit of an Adversarial Collaboration", "This was a nice, cautious walk-through of an extremely divisive subject. i never thought i would enjoy reading a "point/counterpoint" on abortion, but i enjoyed this one", "Lots of adversity, focused on the actual disagreement, and balanced data and philosophy well."

Automation: Seemed broadly correct and helpful. I didn't find it too exciting because I felt like I had already covered most of the same beats in this article (which they cited), but I'm surprised other people didn't vote for it more. Some positive comments from the survey: "Importance of the issue and the thoroughness with which it was explored", "The most fitting, thought-out and the one that draws actual conclusions", "Highest rationality-to-contentiousness ratio".

Spiritual Experience: This was another one that was long, fascinating, and didn't seem to be making much of an attempt to come to a conclusion. I especially liked the section on near death experiences, and I'll be thinking about it a lot, but I didn't feel like this collaboration gave me the tools I would need to generate or test hypotheses about what might be going on. Some positive comments from the survey: "The most polished and one which most likely caused me to reconsider things", "Most informative. Best at following an ideal format", "This is the one that 1) is most interesting to me, 2) seems like it had a strong difference of opinion as a starting point.”

I included the positive comments because I think comments on these kinds of things (mine and others) naturally tends to skew negative. Certainly the comments in the comments section were overwhelmingly negative even for the winning collaborations (seriously, what was up with this?) So I want to counter this by pointing out that every collaboration got at least 25 votes, and the comments on the voting survey were mostly positive. It’s easier to nitpick than to give praise where praise is due, but people put in a lot of work here and it was generally appreciated.

I promised that I would come up with some fair way of dividing the prize money, with at least 50% going to the first place winner. Because the top two entries were so close, and because I was so impressed with the second place winner, I choose to give $1,300 to Calvin and Adrian ($650 each), and $1,200 to David and Froolow ($600 each). Please send me an email at scott[at]slatestarcodex[dot]com telling me where to send your share of the money – I can PayPal it to you or donate it to a charity of your choice. Thanks to SSC Patreon supporters for making this possible.

As much as I enjoyed this, I don’t expect to do another contest next year. For one thing, I think requiring two people made it a lot harder – 22 out of 30 teams dropped out before the deadline, and I worry some of that involved a lot of wasted work. For another thing, it involved a surprising amount of work on my part converting whatever Word or Google Docs file people sent me into a format I could use on the blog. Finally, I feel like the past two years did a good job exploring this medium, and now it’s up to other people with real questions to see if they can adapt it to their needs.

Most likely I'll be replacing this with a book review contest sometime towards the end of next year, so if you read any good books, keep them in mind.

But I continue to be interested in adversarial collaborations. If you happen to do one, please tell me – there's a decent chance I'll publish it.

Contra Contra Contra Caplan On Psych

I.

In 2006, Bryan Caplan wrote a critique of psychiatry. In 2015, I responded. Now it's 2020, and Bryan has a counterargument. I'm going to break the cycle of delay and respond now, and maybe we'll finish this argument before we're both too old and demented to operate computers.

Bryan writes:

> 1. With a few exceptions, Scott fairly and accurately explains my original (and current) position.
>
> 2. Scott correctly identifies several gray areas in my position, but by my count I explicitly acknowledged all of them in my original article.
>
> 3. Scott then uses those gray areas to reject my whole position in favor of the conventional view.
>
> 4. The range of the gray areas isn't actually that big, so he should have accepted most of my heterodoxies.
>
> 5. If the gray areas were as big as Scott says, he should reject the conventional view too and just be agnostic.

I think the gray areas are overwhelming and provide proof that Bryan's strict dichotomies don't match the real world.

I also think, as a general philosophical point, that we ought to be suspicious of arguments of the form "the gray areas are small". Even if this is true, and your model only fails in a few places, controversial questions are likely to be controversial questions precisely because they’re located where your model fails. Nobody challenges a model on an exactly typical case where everything makes sense. So if a point is under debate, let’s say in a fifteen year back-and-forth argument between two bloggers that’s attracted hundreds of total comments, the a priori size of the gray areas doesn’t matter. Even if your model is good at most things, you have strong evidence this isn’t one of them.

In this case, the model we’re debating is Bryan’s idea of constraints vs. preferences. My previous summary of this (which Bryan endorses) goes like this:

> Consumer theory distinguishes between two different reasons why someone might not buy a Ferrari – budget constraints (they can’t afford one) and preferences (they don’t want one, or they want other things more). Physical diseases seem much like budget constraints – the reason a paralyzed person can’t run a marathon is because it’s beyond her abilities, simply impossible. Psychiatric diseases seem more like preferences. There’s nothing obvious stopping an alcoholic from quitting booze and there’s nothing obvious preventing someone with ADHD from sitting still and paying attention. Therefore they are best modeled as people with unusual preferences – the one with a preference for booze over normal activities like holding down a job, the other with a high dispreference for sitting still and attending classes. But lots of people have weird preferences. Therefore, psychiatric diseases should be thought of as within the broad spectrum of normal variation, rather than as analogous to physical diseases.

I countered by pointing out that this was in fact very analogous to physical diseases:

Alice has always had problems concentrating in school. Now she’s older and she hops between a couple of different part-time jobs. She frequently calls in sick because she feels like she doesn’t have enough energy to go into work that day, and when she does work her mind isn’t really on her projects. When she gets home, she mostly just lies in bed and sleeps. She goes to a psychiatrist who diagnoses her with ADHD and depression.

Bob is a high-powered corporate executive who rose to become Vice-President of his big Fortune 500 company. When he gets home after working 14 hour days, he trains toward his dream of running the Boston Marathon. Alas, this week Bob has the flu. He finds that he’s really tired all the time, and he usually feels exhausted at work and goes home after lunch; when he stays, he finds that his mind just can’t concentrate on what he’s doing. Yesterday he stayed home from work entirely because he didn’t feel like he had the energy. And when he gets home, instead of doing his customary 16 mile run he just lies in bed all day. His doctor tells him that he has the flu and is expected to recover soon.

At least for this week Alice and Bob are pretty similar. They’d both like to be able to work long hours, concentrate hard, and stay active after work. Instead they’re both working short hours, calling in sick, failing to concentrate, and lying in bed all day.

But for some reason, Bryan calls Alice’s problem “different preferences” and Bob’s problem “budgetary constraints”, even though they’re presenting exactly the same way! It doesn’t look like he’s “diagnosing” which side of the consumer theory dichotomy they’re on by

> their symptoms, but rather by his assumptions about the causes.

But Bryan doesn't budge:

> I'm unimpressed, because I not only anticipated such objections in my original paper, but even proposed a test to help clarify the fuzziness…can we change a person's behavior purely by changing his incentives? If we can, it follows that the person was able to act differently all along, but preferred not to; his condition is a matter of preference, not constraint. I will refer to this as the 'Gun-to-the-Head Test'. If suddenly pointing a gun at alcoholics induces them to stop drinking, then evidently sober behavior was in their choice set all along. Conversely, if a gun-to-the-head fails to change a person's behavior, it is highly likely (though not necessarily true) that you are literally asking the impossible.
>
> I then presented multiple forms of evidence that a wide range of alleged mental illnesses are responsive to incentives. Scott barely mentions said evidence.
>
> Still, does this mean that the flu isn't "really" an illness either? No. Rather it means that physical illness often constrains behavioral and changes preferences. When sick, the maximum amount of weight I can bench press falls. (Yes, I've actually tried this). Yet in addition, I don't feel like lifting weights at all when I'm sick. Anyone who has worked while ill should be able to appreciate these dual effects. If you literally get sick, your ability and desire to work both go down. When you metaphorically get "sick of your job," in contrast, only your desire goes down.

I reject the heck out of this answer. I agree the “gun to the head” test is a good summary of Bryan’s position, but we already agreed what Bryan’s position is. The only thing he’s adding here is a claim that the flu still qualifies as a real disease because it sometimes constrains behavior (the amount of weight Bryan can lift). But nobody cares how much weight they can lift during a flu! When we talk about having the flu being bad, we’re talking 0% about how much weight we can lift, and 100% about the sorts of problems Bob has – feeling too ill to go to work, not wanting to do things, etc. If Bryan searches hard enough, he can find a way the flu results in slightly weaker muscle strength. But if I search hard enough, I can find a way depression results in slightly weaker muscle strength. Neither of these things are what the average person thinks about when they think of “flu symptoms” or “depression symptoms”, and I consider them both equally irrelevant.

But if a change in weight-lifting ability really disqualifies the flu for Bryan, we can talk about other diseases.

What about shingles? It’s a viral infection that causes a very itchy rash. But sometimes (*herpes sine zoster*) the rash isn’t visible, and you just get really itchy for a few days. Like, *really* itchy. I had this condition once and it was just embarrassing how much I was scratching myself. But if you had put a gun to my head and said “Don’t scratch yourself, or I’ll kill you”, I would have sat on my hands and suffered quietly. For Bryan, an itch is just a newfound preference for scratching yourself. Shingles, like depression or ADHD, is just a preference shift, and so doesn’t qualify as a real disease.

Or what about respiratory tract infections that cause coughing? My impression is that, put a gun to my head, and I could keep myself from coughing, even when I really really felt like it. Coughing is a preference, not a constraint, and Bryan, to be consistent, would have to think of respiratory infections as just a preference for coughing.

Or what about migraines? Sure, people with migraines say they feel pain, but that’s no better grounded than someone with depression saying they feel sad. If Bryan is allowed to bring in concepts like “pain”, I’m allowed to bring in concepts like “sadness”, “anxiety”, etc. And since an anxious person feels anxiety and cannot stop feeling it even if threatened with a gunshot, the anxiety counts as a constraint, and so mental disorders are constraining. For Bryan’s constraints-vs-preferences dichotomy to work at all, he has to endorse a sort of behaviorism, where we need not believe anything that doesn’t express itself as behavior. And the only behavior we see in a migraine is somebody lying in bed, turning off all the lights, and occasionally clutching their head and saying “auggggh”. But put a gun to their head and demand they be in a bright room with lots of loud music, and they’ll go to the bright room with lots of loud music. Threaten to shoot them unless they stop clutching their head and moaning, and they’ll stop clutching their head and moaning. In Bryan’s model, migraines are just a newfound preference for saying “auggggh” a lot. Why medicalize this? Some people like saying “auggggh” and that’s valid!

Bryan’s preference vs. constraint model doesn’t just invalidate mental illness. It invalidates many (maybe most) physical illnesses! Even the ones it doesn’t invalidate may only get saved by some triviality we don’t care about – like how maybe you can lift less weight when you have the flu – and not by the symptoms that actually bother us.

II.

We need a model that lets us describe shingles as something more than “this person has a preference for scratching themselves frantically, and that preference is valid, nothing to worry about here”. I don’t have a beautiful elegant version of a model like this yet, but I think Bryan himself has gone most of the way to an at-least-adequate one.

In his post The Depression Preference, Bryan admits that most depressed people don’t want to be depressed. But he terms this a meta-preference – a preference over preferences. They have depressive preferences – for example, a preference for sitting around crying rather than doing work. They would meta-prefer not to have those preferences. But they do have them.

I agree this is a fruitful way to look at things, but I think we have to be really careful here, and that using the same term for endorsed meta-preferences and unendorsed object-level preferences is preventing this level of care. Let’s call endorsed preferences which people meta-prefer to have “goals”, and unendorsed preferences which people would meta-prefer not to have “urges”. I think this closely matches our intuitive understanding of these terms.

Suppose I created a sinister machine that beamed mind control rays into Bryan’s head and gave him an urge to constantly slap himself in the face. This urge could *theoretically* be resisted, but it’s so strong that in practice he never managed to resist it. It didn’t make him enjoy slapping himself in the face, or think this was a reasonable thing to do. It just made him compulsively want to keep doing it. He loses his job, his friends, and his dignity, because nobody wants to be around someone who’s slapping himself in the face all the time. I hope we can common-sensically agree on the following:

1. This is bad
2. Bryan would want to find and destroy the sinister machine
3. That would be a pretty reasonable goal for Bryan to have, and society should support him in this

This seems a lot like the shingles case. A sinister outside imposition (the viral infection) gives its victim an urge to constantly scratch themselves. It doesn’t make them enjoy scratching themselves, or think this is a reasonable thing to do. These people want to cure their shingles infection, and everyone agrees this desire is reasonable.

But this also seems a lot like some cases of OCD. Did you know that a subset of childhood OCD is caused by a streptococcal infection? So again, you get a sinister outside imposition (an infection) that gives its victim an urge to, let’s say, wash their hands fifty times a day. It doesn’t make them enjoy washing their hands, or think this is a reasonable thing to do (some OCD patients do believe their rituals are necessary, others don’t). These people want to cure their OCD, and I at least agree this desire is reasonable.

If you would support the sinister machine victim and the shingles victim, it’s hard for me to see a case for putting the OCD victim in a different category. I agree I’m using as clear a case as possible (most mental disorders aren’t obviously due to infections), but both Bryan and I are trying to avoid bringing specific facts about biology into this mostly-philosophical debate. The distinction between goals and urges turns what looked like an acceptable situation (these people are following their preferences, which is good) into an unacceptable situation (these people’s goals are being thwarted by unwelcome urges which they can’t resist).

I expect most of Bryan's skepticism to focus on those last two words – "can't resist". He will no doubt bring up his gun-to-the-head test again. If we put a gun to the head of a shingles patient, they could stop scratching. So although we can be sympathetic to the trouble their unwanted new preference causes them, how can we recommend anything other than "just suck it up and resist the preference"?

The best model of decision-making I know of comes from research on lampreys. Various areas of the lamprey brain come up with various plans – hunt for food, hide under a rock, wriggle around – and calculate the "strength" of the "case" for each one, which they convert into an amount of dopamine. They send this dopamine to a part of the brain called the pallium, and then the pallium executes whichever plan has the most dopamine associated with it.

Suppose I have shingles. I'm giving a speech to a group of distinguished people whom I desperately want to impress. Then I get a very strong itch. Part of my brain calculates the expected value of continuing to speak in a dignified way, and converts that into dopamine. Another part calculates the importance of scratching myself vigorously, and converts that into dopamine. The pallium compares these two amounts of dopamine, one is larger than the other, and the decision gets made. If the itch is bad enough, and if whatever lizard-brain nucleus makes me want to scratch itches has enough dopamine to spare, then I never had a chance.

"But," Bryan objects, "if I put a gun to your head, and threatened to shoot you if you scratched the itch, you wouldn't do it, would you?"

In that case, a part of my brain calculates the expected value of continuing to speak in a dignified way *plus not* *getting shot*. This is a very high expected value! It sends lots and lots of dopamine to my pallium. The part of my brain calculating the expected value of scratching the itch *and getting shot* calculates this as a very low-expected-value course, and sends some a very low (maybe negative?) signal. The pallium decisively selects the plan to keep speaking and not get shot.

To summarize: the brain compares the strength of various preferences and executes the strongest. Anything that strengthens your urges at the expense of your goals makes you more likely to do things you don't endorse, and makes you worse off. In a counterfactual world where a threatened gunshot is also weighing down the scale, maybe the calculus would come out different. But in the non-counterfactual world where there is no gunshot, the calculus comes out the way it does.

(also, if Bryan uses his gunshot analogy one more time, I am going to tell him about all of the mentally ill people I know about who did, in fact, non-metaphorically, non-hypothetically, choose a gunshot to the head over continuing to do the things their illness made it hard for them to do. Are you sure this is the easily-falsified hill you want to die on?)

This model doesn't use the word or the concept of "choice" anywhere. There are various algorithms mechanically evaluating the expected reward of different actions, and a more central algorithm comparing all of those evaluations. Those algorithms could have resolved differently in different situations, and you can be uncertain how they will resolve in the same situation, but there's no point at which they *actually* could resolve differently in the same situation. If this makes you want to start debating free will – in either direction – I cannot recommend this Less Wrong post highly enough.

A few examples to hammer this in:

1. Most weekends, Alice stays in and reads a book (preference strength 20). But today is her firstborn child's wedding, which she has been looking forward to for years (preference strength 100). Just before she leaves for the chapel, she gets a terrible migraine, and she feels like it would be unbearable to go out of her room (preference strength 200). Since 200 is greater than 100, Alice misses the wedding and feel miserable, since she would have meta-preferred to go to the wedding. If you had threatened to shoot her unless she went to the wedding, she would have gone to the wedding and been miserable the whole time, because she is terrified of death (preference strength 9999) and 9999 is greater than 200.

2. Bryan is a responsible member of society and wants to work hard and take care of his family (preference strength 100). He drinks some alcohol, but because he has no genetic or environmental risk factors for alcoholism, it doesn't make him feel any urge to drink himself to death (preference strength 0), so he doesn't. If we CRISPRed him to give him every single alcoholism risk gene plus crippling anxiety, then drinking the alcohol would make him feel a very strong urge to drink himself to death (preference strength 200), and he would drink himself to death instead of caring for his family.

3. CRISPRed alcoholic Bryan goes to an addiction doctor. The doctor advises him to take the anti-alcoholism drug naltrexone (-20 preference strength for alcoholism). Then the doctor advises him to go to Alcoholics Anonymous and get a whole new friend group in which his status depends

entirely on his ability to remain sober (+20 for staying sober). Now his preferences are “stay sober and take care of my family” (strength 120) vs. “drink myself to death (strength 180), but the preference to drink is still stronger, so he does.

4. Bryan goes to a therapist who asks him to visualize the things he loves about his family and why he thinks it’s important to take care of them, which makes this more vivid in his mind (preference +10 for sobriety). Bryan’s boss threatens to fire him if he misses one more day of work because of drunkenness (preference +20 for sobriety). Now he’s at 150 for sobriety vs. 180 for drinking. He gives $20,000 to Beeminder, which they will only give him back if he stays sober for the next year (+20 for sobriety), and he reads George Ainslie’s Picoeconomics which describes ways to reconceptualize choices across time to better account for all of their implications (+20 for sobriety). Now he’s at 190 for sobriety vs. 180 for drinking, so he stays sober.

5. A few months later, Bryan’s friend dies in an accident. He feels angry, depressed, and anxious. This makes alcohol seems more attractive, since it would temporarily help him forget these feelings (+20 for drinking). At the same time, he stops going to AA because it’s annoying and far away (-20 for staying sober). Now he’s at 170 for sobriety vs. 200 for drinking, so he falls off the wagon.

I’m not claiming this lamprey model is exactly literally true for humans. And I’m not claiming there’s a perfect binary distinction between endorsed goals and unendorsed urges. This model is full of complications and gray areas. I’m just saying it’s a *better* model, with *fewer* gray areas, than trying to separate everything into just “preference” or “constraint”, and shooting yourself in the foot again and again like some kind of tipped-over Gatling gun.

And it goes a lot of the way to modeling mental illness: the mentally ill have conditions that give them strong unendorsed urges. For any given strength of goal, having strong urges will make people less able to pursue that goal, in favor of pursuing the urges instead, and that will make them worse off, for a definition of “well off” that involves being happy and achieving goals. These people very reasonably want to stop having these weird urges so they can pursue their goals in peace.

Bryan will correctly point out that there are awkward implications in identifying “unexpected generator of strong unendorsed urges” with “disease”. For example, gay people in a traditional religious community will have strong urges to have homosexual relationships, and they won’t endorse those urges – they would probably rather be straight instead.

Or: obese people feel an urge to eat which they don’t endorse. Should we call obesity a disease, and describe them as having a disease which produces urges contrary to their preferences? Some people say yes (and keep in mind that both genetics and viral infections can induce obesity). But suppose some normal-weight person would rather be supermodel-thin, and their perfectly normal urge to eat a normal amount prevents them from looking like a broomstick. Is their normal level of hunger a disease? A naive equation of “biological generator of unendorsed urges” and “disease” would say yes!

We want some criteria that let us call shingles a disease, but don’t let us call “being thin but wanting to be even thinner” a disease. Unfortunately, there is no perfect solution to this problem. People have wanted perfect solutions to definitional questions ever since Plato defined man as “a featherless biped”, and it’s never worked. Luckily, there are kludgy, good-enough solutions, which I describe in Dissolving Questions About Disease, the fourth most popular Less Wrong post of all time. If you still think this is confusing, please read it. If it’s still confusing even after that, try The Categories Were Made For Man, Not Man For The Categories.

I think Bryan should be happy with this solution. It’s very libertarian. It says that it’s up to every individual to decide how to satisfy their own preferences (including meta-preferences). If your problem is constraints (you want to go to Hawaii, but you don’t have enough money), you can work to resolve those constraints (eg go to work and earn more money). If your problem is urges (you want to go to Hawaii, but you’re too anxious to leave your room), you can work to resolve those urges (eg go to a psychiatrist and get medication). The job of a good liberal society is to support people in achieving their own goals as they understand them, and this includes supporting their decision to get the job they want *and* their decision to get the psychiatric treatment they want.

As I write this essay, I’m a little bit caffeinated. I looked at my preference set – which included an urge to get back in bed instead of writing blog posts – decided it didn’t achieve my goals, and took a psychotropic drug to shift my preference set to one I liked better. And if we’re willing to accept this in relatively trivial cases, the argument for accepting it is even stronger for people whose preference sets have been deranged by obvious bizarre causes – infections, hormone imbalances, brain injuries, addictive substances, genetic defects – and for people whose irresistible urges are ruining their lives in preventable ways.

Last Chance To Take The Survey

This is your last chance to take the **2020 SSC survey**. Remember, it’s open to anyone who has read posts on this blog, it should take about 30 minutes to finish, and don’t click on a link unless you’re sure it will open in a new tab and not throw out your responses so far.

Thanks to everyone who’s taken it already. I look forward to sharing some interesting results with you soon.

SSC Survey Results 2020

EDIT: This post presents open-access data from a large survey that anyone is allowed to download and analyze. It’s gone viral because a Twitter user named Philippe Lemoine downloaded the data and used it to investigate politics and mental illness. That analysis is not described here, it’s not my analysis, and I have various caveats about it, which are described here. Sorry for any confusion.

Thanks to the 8,043 people who took the 2020 Slate Star Codex survey.

See the questions for the SSC survey

See the results from the SSC Survey (click “see previous responses” on that page)

Some people expressed concern about privacy on the survey. Originally, respondents could see aggregate responses, including the responses of people who marked their answers private. I figured this was okay because nobody’s responses could be connected – ie you could see that one person put their age as 83, and another person put their country as Canada, but because the table order wasn’t the same you couldn’t link these together to form a coherent picture of an 83 year old Canadian. Some people still expressed concern about a few of the long answers, since some people might have put personal information in there. There’s no way for me to eliminate only the private people’s responses from Google Forms and still display the information to you like this, so instead I’ve removed all long answer questions. If you’re interested in those, you can find them in the downloadable data files. Sorry for not doing this earlier, and I hope this compromise is okay to everyone. I’ll try to get a clearer picture of what people want before the next survey.

I’ll be publishing more complicated analyses over the course of the next year, hopefully starting later this week. If you want to scoop me, or investigate the data yourself, you can download the answers of the 7000 people who agreed to have their responses shared publicly. The public datasets will not exactly match the full version, some overly identifiable questions (eg age) will be binned, and a few sensitive subjects will not be included.

Download the public data (**.xlsx**, **.csv**)

Finally, the game results. I randomly selected Game 3, “Prisoner’s Dilemma Against Your Clone”, chose a random respondent as the prisoner, and found someone similar to be his clone. Of the two clones, one cooperated and one defected, so the defector gets the full prize. That defector’s public key is “gwern is my waifu and paperklipot maximizer”. Please email me within one week at scott[at]slatestarcodex.com with your private key and a Paypal account where I can send you money (or a charity you want me to donate to).

Book Review Review: Little Soldiers

Little Soldiers is a book by Lenora Chu about the Chinese education system. I haven't read it. This is a review of Dormin111's review of Little Soldiers.

Dormin describes the "plot": The author is a second-generation Chinese-American woman, raised by demanding Asian parents. Her parents made her work herself to the bone to get perfect grades in school, practice piano, get into Ivy League schools, etc. She resisted and resented the hell she was forced to go through (though she got into Stanford, so she couldn't have resisted too hard).

Skip a decade. She is grown up, married, and has a three year old child. Her husband (a white guy named Rob) gets a job in China, so they move to Shanghai. She wants their three-year-old son to be bilingual/bicultural, so she enrolls him in Soong Qing Ling, the Harvard of Chinese preschools. The book is about her experiences there and what it taught her about various aspects of Chinese education. Like the lunches:

> During his first week at Soong Qing Ling, Rainey began complaining to his mom about eating eggs. This puzzled Lenora because as far as she knew, Rainey refused to eat eggs and never did so at home. But somehow he was eating them at school.
>
> After much coaxing (three-year-olds aren't especially articulate), Lenora discovered that Rainey was being force-fed eggs. By his telling, every day at school, Rainey's teacher would pass hardboiled eggs to all students and order them to eat. When Rainey refused (as he always did), the teacher would grab the egg and

> shove it in his mouth. When Rainey spit the egg out (as he always did), the teacher would do the same thing. This cycle would repeat 3-5 times with louder yelling from the teacher each time until Rainey surrendered and ate the egg.
>
> Outraged, Lenora stormed to the school the next day and approached the teacher in the morning as she dropped Rainey off. Lenora demanded to know if Rainey was telling the truth – was this teacher literally forcing food into her three-year-old son’s mouth and verbally berating him until he ate it. The teacher didn’t even bother looking at Lenora as she calmly explained that eggs are healthy and that it was important for children to eat them. When Lenora demanded she stop force-feeding her son, the teacher refused and walked away.

Or the seating:

> As Lenora hears more crazy stories from her son and friends, she keeps coming back to one question: “what does Rainey actually do in school?” Lenora tries to ask Rainey, but he always replies, “we sit still.” He also occasionally mentions painting and eating, but that’s it.
>
> So Lenora goes to Rainey’s teacher one day and asks to sit in on classes to observe. Lenora is told that this is not possible. So she asks if she can know a little more about what the school is teaching Rainey. The teacher tells her that she is already told everything she needs to know, and that this is the “Chinese way.”
>
> Since Lenora couldn’t get a look into Soong Qing Ling, she went to another local school and bribed her way into a classroom-observation post with some well-placed handbags. She discovered that Rainey was basically

> right. Chinese preschool really does seem to consist of sitting still. Unless given different orders, all students were required to sit in their seats with their arms at their sides, and their feet flat on a line of tape on the ground. This is not an easy task for three-year-olds.
>
> There were two teachers in the classroom with a classic good cop/bad cop dynamic. The good cop stood in the front of the room with the desks splayed out before her. She would give simple instructions like orders to get food, water, or sometimes paint, though usually she said nothing at all. The bad cop was another teacher who prowled the classroom. Any time she saw a student remove a foot from the line, move arms from his side, or otherwise deviate from the instructions, she would yell at the student to fall back in line. Lenora spent about a week watching tiny kids get screamed at for trying to get water, shifting in their chairs, or talking to classmates.

Or art class:

> When Lenora sat in on a kindergarten class, she witnessed an art lesson where the students were taught how to draw rain. The nice teacher drew raindrops on a whiteboard, showing precisely where to start and end each stroke to form a tear-drop shape. When it was the students’ turns, they had to perfectly replicate her raindrop. Over and over again. Same start and end points. Same curves. For an hour. No student could draw anything else. Any student who did anything different would be yelled at and told to start over.
>
> The point of this exercise was not to teach students how to draw raindrops. Drawing raindrops is not an important life skill, and drawing them in a particular way

> is especially not important. Even the three-year-old students in the class seemed to realize this as many immediately created their own custom raindrop shapes and drew landscapes, all to be crushed under the mean teacher's admonishment. The real point of the exercise was to teach students to follow directions from an authority figure. But more than that, the point was to follow pointless and arbitrary directions. The more pointless and arbitrary the directions are, the more willpower is required to follow them.

Chinese people presumably put up with this because it makes sense within their culture; why did Chu put up with it? Dormin half-jokingly suggests maybe she really wanted to write the book she eventually wrote, and this was her research. But Chu herself says it eventually got results:

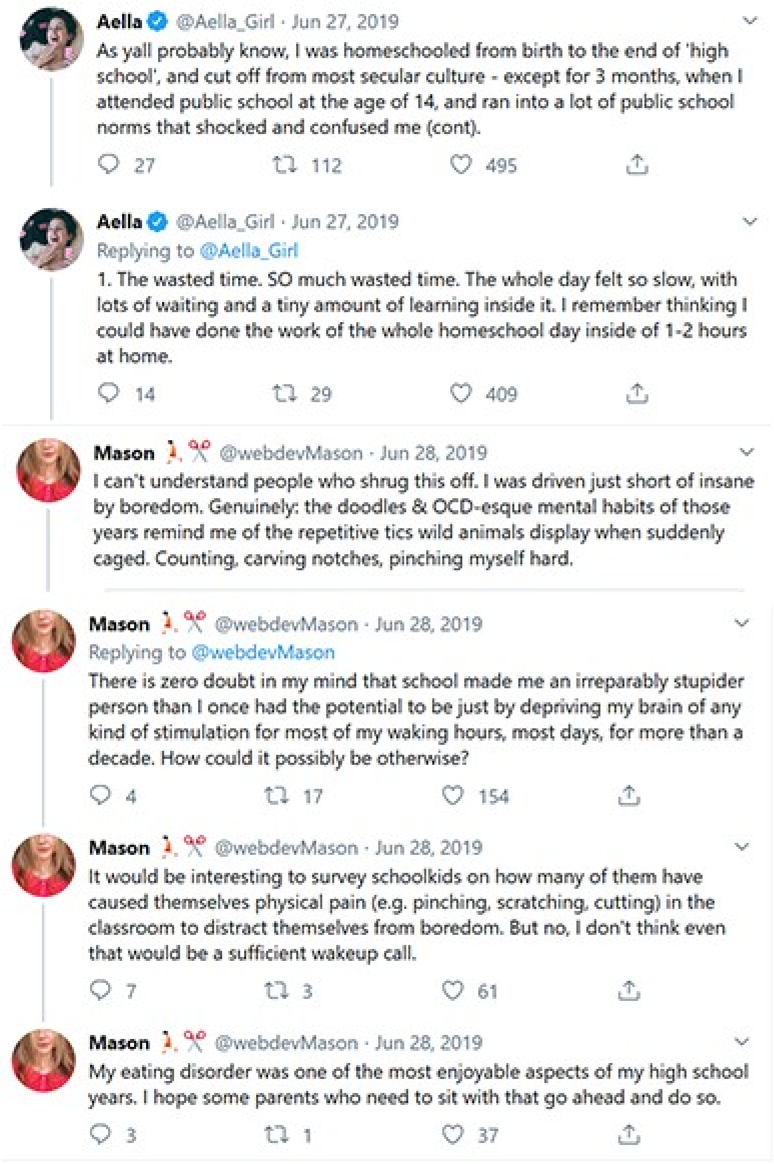
> After spending 75% of the book relentlessly complaining about her son's Chinese education, with the occasional anecdote about how horrible her own culturally Chinese upbringing was, Lenora decides Chinese schools aren't so bad.
>
> After a few years in China, Rainey changed. Though Lenora constantly worried if Rainey's creativity and leadership potential was being snuffed out, she couldn't help but be impressed by his emerging self-control. He could sit still for longer. He always greeted people politely. He finished eating his food. He asked permission a lot.
>
> Lenora didn't realize what Rainey had become until she took him back to the US for a few weeks to visit family. There, the contrast between Rainey and his same-aged American counterparts become stark. Lenora's friends' kids ate junk food all day while Rainey asked for

> vegetables. They couldn’t read or do basic addition while Rainey was close to being bilingual and had started double-digit addition and subtraction by first grade. They wandered obliviously in their own worlds while Rainey’s Chinese grandparents were thrilled to receive respectful greetings every time Rainey entered the room […]
>
> What really sold Lenora on Chinese education was that it apparently worked. At the time of writing the book, Shanghai was scoring first place in the world on the PISA exams, beating heavy-hitters like Norway and Singapore. Supposedly, education scholars and professionals all over the world were looking at China for wisdom. They all saw the bad, but they saw a lot of good too.

(before going forward, I should interject that China’s great PISA scores are kind of fake. China struck a deal with the OECD (the group that administers PISA) to let it conduct testing only in its four richest and best-educated provinces. Rich and well-educated places always do well on PISA. That China’s four best provinces outperform the average score of other countries is unsurprising. This article points out that if the US were allowed to enter only its best-educated state (Massachussetts, obviously) we would be right up there with China. So this probably isn’t as impressive as Ms. Chu thinks.)

This is just a sample of the great stuff in Dormin’s review of *Little Soldiers*, and I strongly recommend you read the whole thing. You should also read the comments, which point out that this may be more about a few elite Chinese schools than about an entire country. But I want to use these excerpts as a jumping-off point to talk about the US education system, unschooling, and child development in general.

I predict most of my Bay Area friends would hate the Chinese education system as Chu describes it. I predict this because they already hate the US education system, which is only like 10% as bad. I’m especially thinking of @webdevmason and @michaelblume, who often write about the ways American education is frustrating, regressive, and authoritarian. Bright-eyed, curious kids come in. They spend thirteenish years getting told to show their work, being punished for reading ahead in the textbook, and otherwise having their innate love of learning drummed out of them in favor of endless mass-produced homework assignments (five pages, single-spaced, make sure you use the right number of topic sentences).



Aella @Aella_Girl · Jun 27, 2019

As yall probably know, I was homeschooled from birth to the end of 'high school', and cut off from most secular culture - except for 3 months, when I attended public school at the age of 14, and ran into a lot of public school norms that shocked and confused me (cont).

Aella @Aella_Girl · Jun 27, 2019

Replying to @Aella_Girl

1. The wasted time. SO much wasted time. The whole day felt so slow, with lots of waiting and a tiny amount of learning inside it. I remember thinking I could have done the work of the whole homeschool day inside of 1-2 hours at home.

Mason @webdevMason · Jun 28, 2019

I can't understand people who shrug this off. I was driven just short of insane by boredom. Genuinely: the doodles & OCD-esque mental habits of those years remind me of the repetitive tics wild animals display when suddenly caged. Counting, carving notches, pinching myself hard.

Mason @webdevMason · Jun 28, 2019

Replying to @webdevMason

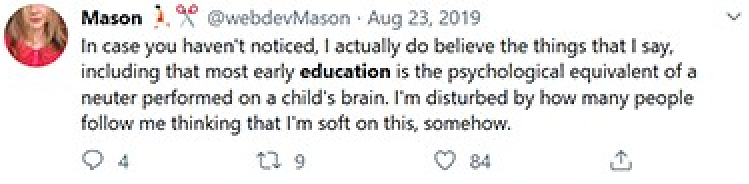
There is zero doubt in my mind that school made me an irreparably stupider person than I once had the potential to be just by depriving my brain of any kind of stimulation for most of my waking hours, most days, for more than a decade. How could it possibly be otherwise?

Mason @webdevMason · Jun 28, 2019

It would be interesting to survey schoolkids on how many of them have caused themselves physical pain (e.g. pinching, scratching, cutting) in the classroom to distract themselves from boredom. But no, I don't think even that would be a sufficient wakeup call.

Mason @webdevMason · Jun 28, 2019

My eating disorder was one of the most enjoyable aspects of my high school years. I hope some parents who need to sit with that go ahead and do so.



Mason @webdevMason · Aug 23, 2019

In case you haven't noticed, I actually do believe the things that I say, including that most early **education** is the psychological equivalent of a neuter performed on a child's brain. I'm disturbed by how many people follow me thinking that I'm soft on this, somehow.

People with this position usually make two claims. One, US public school as it currently exists is awful, basically institutionalized child abuse. Two, this is bad for the economy. I’ve been through too much school myself to feel like challenging the first, so I want to focus on the second.

Salman Khan, John Gatto, and other education rebels trace the current school systems back to the Prussians, who invented compulsory education to prepare children for a career as infantrymen or factory workers. It’s a great story. Like most great stories, it’s kind of false. But like most kind-of-false things that catch on, it has an element of truth. Children who can sit still in a classroom and do what their teachers say are well-placed to become adults who can sit still in an open office and do what their bosses say. So (according to this logic), even if our schools are awful, they were well-suited to the Industrial Age economy. Some hypothetical mash-up of Otto von Bismarck and Voldemort, who wanted the country to produce as much as possible and didn’t care how many children’s souls were crushed in the process, might at least endorse the education system on widget-maximization grounds.

But (these same people argue), the Industrial Age is over. The most important skills now are entrepreneurship and creative problem solving. Reinventing yourself, selling yourself, carving out a new niche for yourself. Figuring out what’s going to be the next big thing and pursuing it without anyone else watching over you. We’re in XKCD’s world now, where 900 hours of classes and 400 hours of homework matter less to your career success than one weekend messing around with a programming language in 11th grade. The Prussian model of education stamps out the kind of independent agency that could help people navigate the weird, formless 21st century world.

How might the personified Chinese education system respond?

What if it said “I don’t know what you □□ are doing in America, but I’m not crushing anybody. I’m just telling kids to sit here drawing 1,000 raindrops in a row without moving or protesting. If after that you decide you don’t want to found the next Uber, that’s on you. But if you *do* decide to found the next Uber, I will have taught you the most important skill: discpline. Learning how to sit still and obey others is the necessary prerequisite to learning how to sit still and obey yourself.”

If it was really mean, it might go further. “I notice most of you Americans suck at this skill. I notice you’re always whining about how you don’t have enough discipline to pursue your interests. Some of you are writers who spend years fantasizing about the novel you’re going to publish, but can never quite bring yourself to put pen to paper. Others want to learn another language, but reject real work in favor of phone apps that promise to ‘gamify’ staying at a 101 level for the rest of your life. You don’t need to feel bad about having no self-control; after all, nobody taught you any. If you’d gone to □□□□□□, you would have spent your formative years learning to sit still and focus, having your natural impulse to slack off squeezed out of you. Then you could have pushed through and written your novel, or learned □□, or if you wanted to start Uber you could start Uber. At the very least you’d be doing something other than lying in bed browsing Reddit posts about how adulting is hard.”

My Bay Area friends treat people as naturally motivated, and assume that if someone acts unmotivated, it’s because they’ve spent so long being taught to suppress their own desires that they’ve lost touch with innate enthusiasm. Personified China treats people as naturally unmotivated, and assumes that if someone acts unmotivated, it’s because they haven’t been trained to pursue a goal determinedly without getting blown around by every passing whim.

What evidence is there in favor of one education system or the other?

I can’t find any good studies directly supporting or opposing either of these claims. The best I can do is The Development Of Executive Functioning And Theory Of Mind: A Comparison Of Chinese And US Preschoolers. They find that on various tests of executive function, “Chinese [preschool-age] children’s performance was consistently on par with that of US children who were on average 6 months older” (other sources say 1-2 years). But lots of interventions change things in childhood; this isn’t interesting unless it persists into adulthood, and I don’t see any work on this. This study on racial differences in personality traits found weak and inconsistent white-Asian differences on adult conscientiousness, but the Asian sample was Asian-American and differences in education were probably pretty minor.

What about circumstantial evidence?

First and most important, since extreme cultivation of discipline vs. laissez-faire childrearing is a property of parents as much as schools, any claimed effect would run afoul of all the twin studies showing that shared environment has few long-term effects on any trait. For example, this meta-analysis of factors affecting self-control that finds “no or very little influence of the shared environment on the variance in self-control”. But we can always invoke the usual loophole in shared environment findings: maybe the US doesn’t contain anything as extreme as the Chinese education system, so US-only studies can’t capture its effects.

Second, both Westerners and Chinese seem to include some very impressive and some less impressive people. It certainly doesn't seem *wrong* to say that Chinese people seem more diligent and Westerners seem more independent, but there are so many potential biases at work that I would hate to take this too seriously as evidence for or against one form of education. Also, Chinese-Americans who are educated in US schools also seem more diligent than white Americans, so maybe the education system doesn't contribute too much to this. Maybe Chinese culture promotes diligence better in general, this causes diligence-focused school systems, but the diligence-focused school systems don't themselves cause the diligence.

Third, we could try to find more extreme versions on both sides and see what happens there. Pre-industrial populations with no education were famously bad at the discipline needed for factory work. From Pseudoerasmus:

> The earliest factory workers were lacking in what Mokyr & Voth call "discipline capital" — non-cognitive 'skills' like punctuality, sobriety, reliability, docility, and pliability. Whether they had been peasants or artisans, early workers were new to industrial work habits and they had a strong preference for autonomous work arrangements. They were accustomed to setting their own pace of work in farming, domestic outwork, or artisanal workshops, and disliked the time rules and strict supervision of the factories.
>
> All this is consistent with colourful descriptions of the early history of the textile industry in the Global South, including Japan. Mills were described as places of chaos and disorder. They were supposedly filled with workers 'idling', 'loitering', 'socialising', smoking, tea-drinking, or just disappeared for the day. In Japan, "twenty percent

> of the female operatives…absent themselves after they receive their monthly pay check" (Saxonhouse & Kiyokawa 1978). In Shanghai, it was said female mill workers could be found breast-feeding infants during work hours (Cochran 2000). Or at Mumbai mills, workers "bathed, washed clothes, ate his meals, and took naps" (Gupta 2011).

But this could be as much about expectations as about abilities.

Which historical culture had the most authoritarian-instillment-of-virtue-focused approach to child-rearing? Surely the New England Puritans were up there – remember that eg Puritan parents would traditionally send children away to be raised by other families, in the hopes that the lack of familiarity would make the child behave better". They certainly ended out industrious. But they were also creative and self-motivated, sometimes almost hilariously so. On the other hand, I'm not sure that the Puritans who ended up incredibly creative were exactly the same Puritans who suffered extreme strict child-rearing – there seems about a century gulf between the evidence of authoritarian parenting in the 1600s and the crop of geniuses born in the late 1700s – so I'm not sure how seriously to take this.

Fourth, we could look at US trends over time. Both US parenting and US schooling seem to be getting less authoritarian over time; 31 states have banned corporal punishment since 1970, and the teachers I know confirm a shift away from most forms of discipline. Over the same time period, children have gotten weirdly better behaved – less crime, less teenage pregnancy, more willing to jump through various stupid hoops to get into a good college. This seems to contradict the Chinese theory – the children are no worse at controlling their impulses. But there are other findings that contradict the Bay Area theory – entrepreneurship is decreasing; more top students are choosing to go work for a boss at a big bank rather than go do something weird. I think the better behavior is probably just caused by lower lead; I have no idea why people are more risk-averse. Secular decline in testosterone, maybe?

Fifth, we could look at research on the effects of preschool more generally. Some studies find that US preschools do not make children smarter, but still improve life outcomes like graduation rates, crime rates, and employment. Although there are lots of theories about the "noncognitive skills" that accomplish this (including that they don't exist and the improvement is an artifact of bad experimental technique), this is certainly consistent with preschool teaching children discipline at a critical window. If this hypothesis were true, the effect of preschool would be much larger in China, but I don't know of any Chinese studies on the topic.

Sixth, we could look at the research on meditation for very young kids. The Chinese theory casts preschool as a sort of dark-side form of mindfulness. In traditional Buddhist settings, monks would sit perfectly still and concentrate on the most boring thing imaginable, and the head monk would slap them with a bamboo stick if they moved. The resemblance to the school system is uncanny. So maybe school's effects on self-control could be modeled as a sort of less-intense but much-more-drawn-out meditation session. Unfortunately, the studies surrounding mindfulness in kids are crap, so this doesn't help either.

Really none of this seems very helpful and we're kind of left with our priors. And maybe one of our priors is "don't abuse children", so there's that.

But what about the Polgars? They turned all three of their children into chess prodigies through a strategy that seemed based around exposing them to absurd amounts of chess at a very young age. If we generalize, it does look like very young children might have very plastic minds that you can shape through out-of-distribution experiences. But Lazslo Polgar insisted that his technique didn't use force; the point was to interest his children in the material so avidly that they inflicted near-Chinese levels of intensity on themselves in order to study it more successfully.

One problem with the physical universe is that even after you study a question in depth and decide more evidence is needed, there are still real children you have to educate one way or the other. I have no general solution for this, but the Polgar strategy seems like a good deal if you can pull it off.

Assortative Mating And Autism

Introduction

Assortative mating is when similar people marry and have children. Some people worry about assortative mating in Silicon Valley: highly analytical tech workers marry other highly analytical tech workers. If highly analytical tech workers have more autism risk genes than the general population, assortative mating could put their children at very high risk of autism. How concerned should this make us?

Methods / Sample Characteristics

I used the 2020 Slate Star Codex survey to investigate this question. It had 8,043 respondents selected for being interested in a highly analytical blog about topics like science and economics. The blog is associated with – and draws many of its readers from – the rationalist and effective altruist movements, both highly analytical. More than half of respondents worked in programming, engineering, math, or physics. 79% described themselves as atheist or agnostic. 65% described themselves as more interested in STEM than the humanities; only 15% said the opposite.

According to Kogan et al (2018), about 2.5% of US children are currently diagnosed with autism spectrum disorders. The difference between “autism” and “autism spectrum disorder” is complicated, shifts frequently, and is not very well-known to the public; this piece will treat them interchangeably from here on. There are no surveys of what percent of adults are diagnosed with autism; it is probably lower since most diagnoses happen during childhood and the condition was less appreciated in past decades. These numbers may be affected by parents' education level and social class; one study shows that children in wealthy neighborhoods were up to twice as likely to get diagnosed as poorer children.

Given that respondents are likely wealthier than average, we might expect a rate of 2.5% – 5%. Instead the rate is noticeably higher than that, consistent with the hypothesis that this sample will be more autistic than average. About 4% of the SSC survey sample had a formal diagnosis of autism, but this rose to 6% when the sample was limited to people below 30, and to 8% below 20. This sample is plausibly about 2-3x more autistic than the US population. Childhood social class was not found to have a significant effect on autism status in this sample.

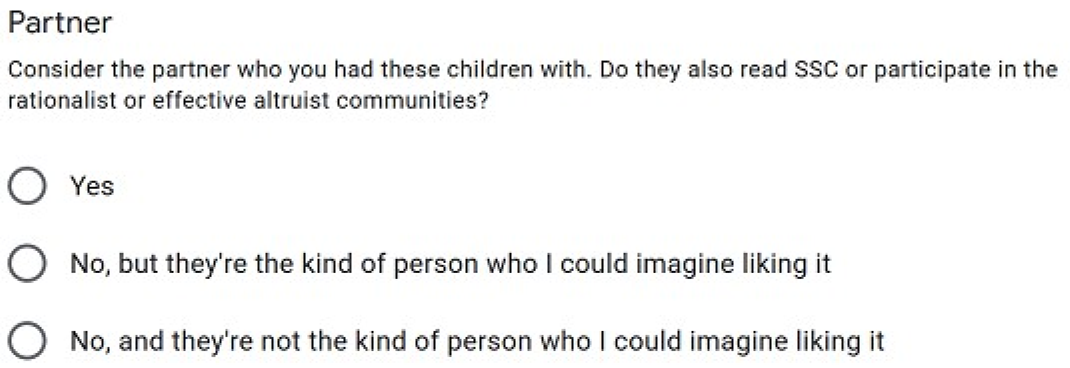
Results

I tried to get information on how many children respondents have, but I forgot to ask important questions about age until a quarter of the way through the survey. I want to make sure I'm only catching children old enough that their autism would have been diagnosed, so the information below (except when otherwise noted) comes from the three-quarters of the sample where I have good age information. I also checked it against the whole sample and it didn't make a difference.

Of this limited sample, 1,204 individual parents had a total of 2,459 children. 1,892 of those children were older than 3, and 1,604 were older than 5. I chose to analyze children older than 3, since autism generally becomes detectable around 2.

71 children in the 1,892 child sample had formal diagnoses of autism, for a total prevalence of 3.7%. When parents were asked to include children who were not formally diagnosed but who they thought had the condition, this increased to 99 children, or a 5.2% prevalence. Both numbers are much lower than the 8% prevalence in young people in the sample.

What about marriages where both partners were highly analytical? My proxy for this was the following survey question:



Partner

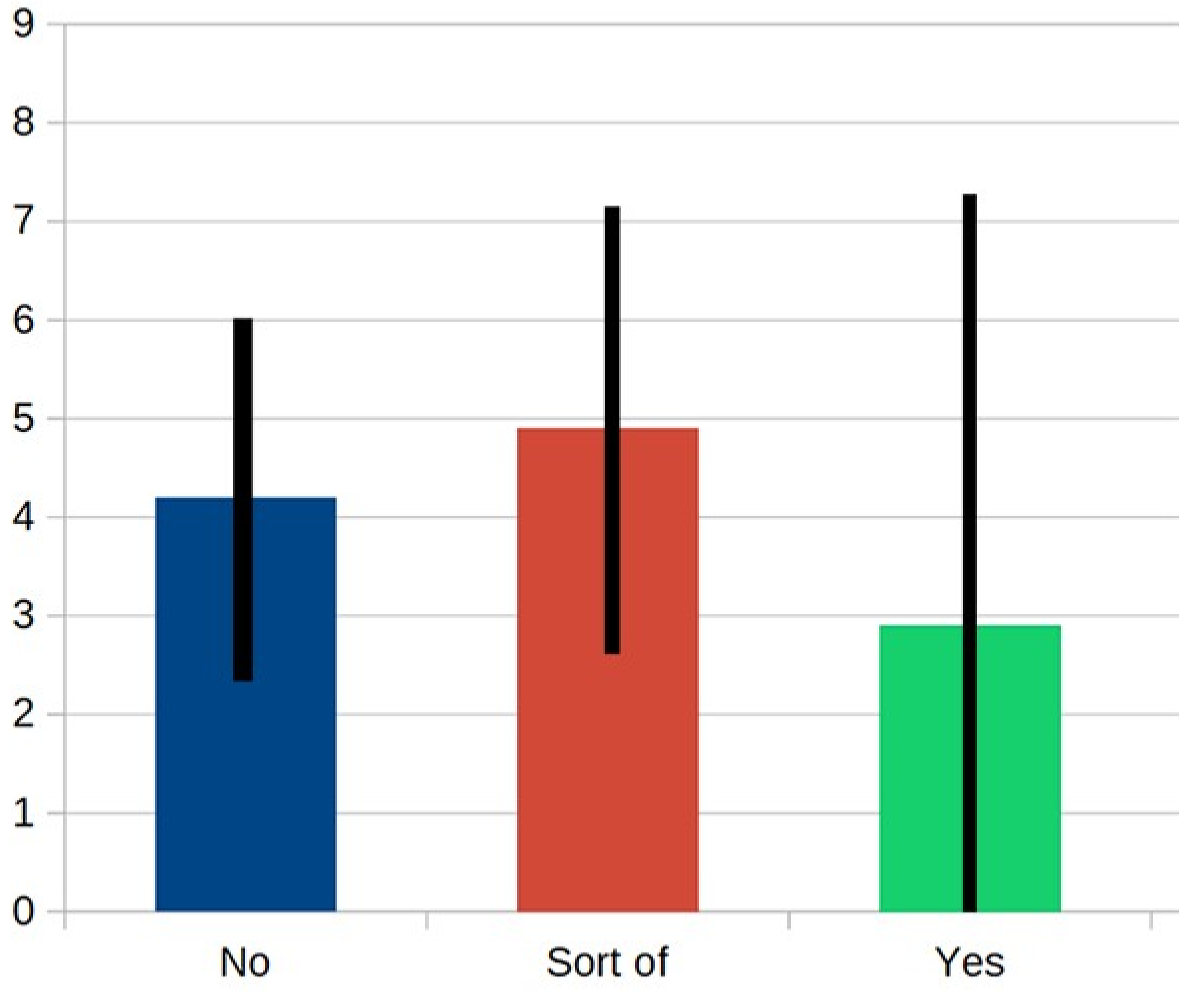
Consider the partner who you had these children with. Do they also read SSC or participate in the rationalist or effective altruist communities?

○ Yes

○ No, but they're the kind of person who I could imagine liking it

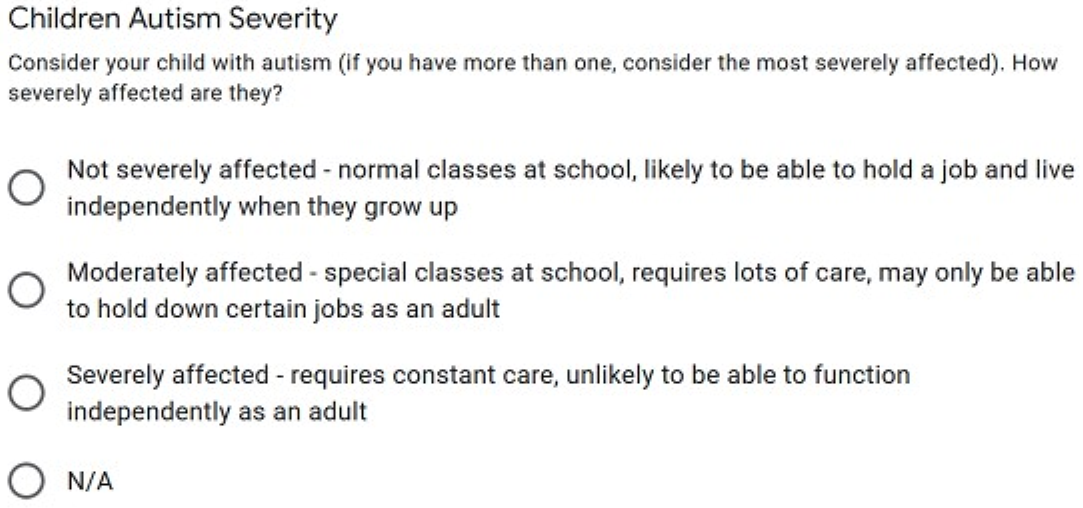
○ No, and they're not the kind of person who I could imagine liking it

I’ll be referring to these answers as “yes”, “sort of”, and “no” from here on, and moving back to the full sample. 938 parents answered this question; 51 (5.4%) yes, 233 (24.8%) sort of, and 653 (69.4%) no. Keep in mind the effective sample is even smaller, since both partners in two-partners-read-SSC-families may have filled out the survey individually about the same set of children (though this should not have affected the “sort of” group). Here is the autism rate for each group, with 95% confidence interval in black:



There is little difference. If we combine the latter two groups, the confidence interval narrows slightly, to 2.7 – 6.5.

I asked respondents about the severity of their children’s autism.



Children Autism Severity

Consider your child with autism (if you have more than one, consider the most severely affected). How severely affected are they?

○ Not severely affected - normal classes at school, likely to be able to hold a job and live independently when they grow up

○ Moderately affected - special classes at school, requires lots of care, may only be able to hold down certain jobs as an adult

○ Severely affected - requires constant care, unlikely to be able to function independently as an adult

○ N/A

People who hadn’t previously reported any children with autism gave answers other than N/A for this one, which was confusing. Instead of the 71 children we had before, now it’s up to 144 children. I’m not sure what’s going on here. Of these phantom children, 101 had mild cases, 31 moderate, and only 12 severe. Severe autism was only present in 0.6% of the children in the sample. There was no tendency for couples where both partners were highly analytical to have children with more severe autism.

Discussion

Autism rates in this survey were generally low. Although the general rate of 3.7% was higher than the commonly-estimated US base rate of 2.5%, this is consistent with the slight elevation of autism observed in higher social classes.

There was no sign of elevated risk when both partners were highly analytical. The sample size was too small to say for certain that no such elevation exists, but it can say with 95% confidence that the elevated risk is less than three percentage points.

This suggests that the answer to the original question – does assortative mating between highly analytical people significantly increase chance of autism in offspring – is at least a qualified “no”.

Why should this be? It could just be that regression to the mean is more important in this case than any negative effects from combining recessive genes or mixing too many risk genes together. Or maybe we should challenge the assumption that being a highly analytical programmer is necessarily on a continuum with autism. It seems like p(highly analytical|on autism spectrum) is pretty high, but p(on autism spectrum|is highly analytical) might be much lower.

Obvious limitations of this survey include the small sample size of both-partners-highly-analytical couples, the weak operationalization of highly analytical as “member of the SSC, rationalist, and effective altruist communities”, and the inability to separate non-autistic children from children who are not yet diagnosed. Due to these limitations, this should only be viewed as providing evidence against the strongest versions of the assortative mating hypothesis, where it might increase risk by double, triple, or more. Smaller elevations of risk remain plausible and would require larger studies to assess.

I welcome people trying to replicate or expand on these results. All of the data used in this post are freely available and can be downloaded here.

Book Review: Human Compatible

I.

Clarke's First Law goes: When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong.

Stuart Russell is only 58. But what he lacks in age, he makes up in distinction: he's a computer science professor at Berkeley, neurosurgery professor at UCSF, DARPA advisor, and author of the leading textbook on AI. His new book *Human Compatible* states that superintelligent AI is possible; Clarke would recommend we listen.

I'm only half-joking: in addition to its contents, *Human Compatible* is important as an *artifact*, a crystallized proof that top scientists now think AI safety is worth writing books about. Nick Bostrom's *Superintelligence: Paths, Dangers, Strategies* previously filled this role. But *Superintelligence* was in 2014, and by a philosophy professor. From the artifactual point of view, HC is just better – more recent, and by a more domain-relevant expert. But if you also open up the books to see what's inside, the two defy easy comparison.

S:PDS was unabashedly a weird book. It explored various outrageous scenarios (what if the AI destroyed humanity to prevent us from turning it off? what if it put us all in cryostasis so it didn't count as destroying us? what if it converted the entire Earth into computronium?) with no excuse beyond that, outrageous or not, they might come true. Bostrom was going out on a very shaky limb to broadcast a crazy-sounding warning about what might be the most important problem humanity has ever faced, and the book made this absolutely clear.

HC somehow makes risk from superintelligence *not sound weird*. I can imagine my mother reading this book, nodding along, feeling better educated at the end of it, agreeing with most of what it says (it's by a famous professor! I'm sure he knows his stuff!) and never having a moment where she sits bolt upright and goes *what?* It's just a bizarrely normal, respectable book. It's not that it's dry and technical – HC is much more accessible than S:PDS, with funny anecdotes from Russell's life, cute vignettes about hypothetical robots, and the occasional dad joke. It's not hiding any of the weird superintelligence parts. Rereading it carefully, they're all in there – when I leaf through it for examples, I come across a quote from Moravec about how "the immensities of cyberspace will be teeming with unhuman superminds, engaged in affairs that are to human concerns as ours are to those of bacteria". But somehow it all sounds normal. If aliens landed on the White House lawn tomorrow, I believe Stuart Russell could report on it in a way that had people agreeing it was an interesting story, then turning to the sports page. As such, it fulfills its artifact role with flying colors.

How does it manage this? Although it mentions the weird scenarios, it doesn't dwell on them. Instead, it focuses on the present and the plausible near-future, uses those to build up concepts like "AI is important" and "poorly aligned AI could be dangerous". Then it addresses those abstractly, sallying into the far future only when absolutely necessary. Russell goes over all the recent debates in AI – Facebook, algorithmic bias, self-driving cars. Then he shows how these are caused by systems doing what we tell them to do (ie optimizing for one easily-described quantity) rather than what we really want them to do (capture the full range of human values). Then he talks about how future superintelligent systems will have the same problem.

His usual go-to for a superintelligent system is Robbie the Robot, a sort of Jetsons-esque butler for his master Harriet the Human. The two of them have all sorts of interesting adventures together where Harriet asks Robbie for something and Robbie uses better or worse algorithms to interpret her request. Usually these requests are things like shopping for food or booking appointments. It all feels very Jetsons-esque. There’s no mention of the word “singleton” in the book’s index (not that I’m complaining – in the missing spot between *simulated evolution of programs, 171* and *slaughterbot, 111*, you instead find *Slate Star Codex blog, 146, 169-70*). But even from this limited framework, he manages to explore some of the same extreme questions Bostrom does, and present some of the answers he’s spent the last few years coming up with.

If you’ve been paying attention, much of the book will be retreading old material. There’s a history of AI, an attempt to define intelligence, an exploration of morality from the perspective of someone trying to make AIs have it, some introductions to the idea of superintelligence and “intelligence explosions”. But I want to focus on three chapters: the debate on AI risk, the explanation of Russell’s own research program, and the section on misuse of existing AI.

II.

Chapter 6, “The Not-So-Great Debate”, is the highlight of the book-as-artifact. Russell gets on his *cathedra* as top AI scientist, surveys the world of other top AI scientists saying AI safety isn’t worth worrying about yet, and pronounces them super wrong:

> I don’t mean to suggest that there cannot be any reasonable objections to the view that poorly designed superintelligent machines would present a serious risk to humanity. It’s just that I have yet to see such an objection.

He doesn’t pull punches here, collecting a group of what he considers the stupidest arguments into a section called “Instantly Regrettable Remarks”, with the connotation that the their authors (“all of whom are well-known AI researchers”), should have been embarrassed to have been seen with such bad points. Others get their own sections, slightly less aggressively titled, but it doesn’t seem like he’s exactly oozing respect for those either. For example:

> Kevin Kelly, founding editor of Wired magazine and a remarkably perceptive technology commentator, takes this argument one step further. In “The Myth of a Superhuman AI,” he writes, “Intelligence is not a single dimension, so ‘smarter than humans’ is a meaningless concept.” In a single stroke, all concerns about superintelligence are wiped away.
>
> Now, one obvious response is that a machine could exceed human capabilities in all relevant dimensions of intelligence. In that case, even by Kelly’s strict standards, the machine would be smarter than a human. But this rather strong assumption is not necessary to refute Kelly’s argument.
>
> Consider the chimpanzee. Chimpanzees probably have better short-term memory than humans, even on human-oriented tasks such as recalling sequences of digits. Short-term memory is an important dimension of intelligence. By Kelly’s argument, then, humans are not

> smarter than chimpanzees; indeed, he would claim that "smarter than a chimpanzee" is a meaningless concept.
>
> This is cold comfort to the chimpanzees and other species that survive only because we deign to allow it, and to all those species that we have already wiped out. It's also cold comfort to humans who might be worried about being wiped out by machines.

Or:

> The risks of superintelligence can also be dismissed by arguing that superintelligence cannot be achieved. These claims are not new, but it is surprising now to see AI researchers themselves claiming that such AI is impossible. For example, a major report from the AI100 organization, Artificial Intelligence and Life in 2030, includes the following claim: "Unlike in the movies, there is no race of superhuman robots on the horizon or probably even possible."
>
> To my knowledge, this is the first time that serious AI researchers have publicly espoused the view that human-level or superhuman AI is impossible—and this in the middle of a period of extremely rapid progress in AI research, when barrier after barrier is being breached. It's as if a group of leading cancer biologists announced that they had been fooling us all along: They've always known that there will never be a cure for cancer.
>
> What could have motivated such a volte-face? The report provides no arguments or evidence whatever. (Indeed, what evidence could there be that no physically possible arrangement of atoms outperforms the human brain?) I suspect that the main reason is tribalism — the instinct to circle the wagons against what are perceived

> to be “attacks” on AI. It seems odd, however, to perceive the claim that superintelligent AI is possible as an attack on AI, and even odder to defend AI by saying that AI will never succeed in its goals. We cannot insure against future catastrophe simply by betting against human ingenuity.
>
> If superhuman AI is not strictly impossible, perhaps it’s too far off to worry about? This is the gist of Andrew Ng’s assertion that it’s like worrying about “overpopulation on the planet Mars.” Unfortunately, a long-term risk can still be cause for immediate concern. The right time to worry about a potentially serious problem for humanity depends not just on when the problem will occur but also on how long it will take to prepare and implement a solution. For example, if we were to detect a large asteroid on course to collide with Earth in 2069, would we wait until 2068 to start working on a solution? Far from it! There would be a worldwide emergency project to develop the means to counter the threat, because we can’t say in advance how much time is needed.

Russell displays master-level competence at the proving too much technique, neatly dispatching sophisticated arguments with a well-placed metaphor. Some expert claims it’s meaningless to say one thing is smarter than another thing, and Russell notes that for all practical purposes it’s meaningful to say humans are smarter than chimps. Some other expert says nobody can control research anyway, and Russell brings up various obvious examples of people controlling research, like the ethical agreements already in place on the use of gene editing.

I’m a big fan of Luke Muehlhauser’s definition of common sense – making sure your thoughts about hard problems make use of the good intuitions you have built for thinking about easy problems. His example was people who would correctly say "I see no evidence for the Loch Ness monster, so I don't believe it" but then screw up and say "You can't disprove the existence of God, so you have to believe in Him". Just use the same kind of logic for the God question you use for every other question, and you'll be fine! Russell does great work applying common sense to the AI debate, reminding us that if we stop trying to out-sophist ourselves into coming up with incredibly clever reasons why this thing cannot possibly happen, we will be left with the common-sense proposition that it might.

My only complaint about this section of the book – the one thing that would have added a cherry to the slightly troll-ish cake – is that it missed a chance to include a reference to On The Impossibility Of Supersized Machines.

Is Russell (or am I) going too far here? I don't think so. Russell is arguing for a much weaker proposition than the ones Bostrom focuses on. He's not assuming super-fast takeoffs, or nanobot swarms, or anything like that. All he's trying to do is argue that if technology keeps advancing, then at *some* point AIs will become smarter than humans and maybe we should worry about this. You've really got to bend over backwards to find counterarguments to this, those counterarguments tend to sound like "but maybe there's no such thing as intelligence so this claim is meaningless", and I think Russell treats these with the contempt they deserve.

He is more understanding of – but equally good at dispatching – arguments for why the problem will really be easy. Can't We Just Switch It Off? No; if an AI is truly malicious, it will try to hide its malice and prevent you from disabling it. Can't We Just Put It In A Box? No, if it were smart enough it could probably find ways to affect the world anyway (this answer was good as far as it goes, but I think Russell's threat model also allows a better one: he imagines thousands of AIs being used by pretty much everybody to do everything, from self-driving cars to curating social media, and keeping them all in boxes is no more plausible than keeping transportation or electricity in a box). Can't We Just Merge With The Machines? Sounds hard. Russell does a good job with this section as well, and I think a hefty dose of common sense helps here too.

He concludes with a quote:

> The "skeptic" position seems to be that, although we should probably get a couple of bright people to start working on preliminary aspects of the problem, we shouldn't panic or start trying to ban AI research. The "believers", meanwhile, insist that although we shouldn't panic or start trying to ban AI research, we should probably get a couple of bright people to start working on preliminary aspects of the problem.

I couldn't have put it better myself.

III.

If it's important to control AI, and easy solutions like "put it in a box" aren't going to work, what do you do?

Chapters 7 and 8, "AI: A Different Approach" and "Provably Beneficial AI" will be the most exciting for people who read Bostrom but haven't been paying attention since. Bostrom ends by saying we need people to start working on the control problem, and explaining why this will be very hard. Russell is reporting all of the good work his lab at UC Berkeley has been doing on the control problem in the interim – and arguing that their approach, Cooperative

Inverse Reinforcement Learning, succeeds at doing some of the very hard things. If you haven’t spent long nights fretting over whether this problem was possible, it’s hard to convey how encouraging and inspiring it is to see people gradually chip away at it. Just believe me when I say you may want to be really grateful for the existence of Stuart Russell and people like him.

Previous stabs at this problem foundered on inevitable problems of interpretation, scope, or altered preferences. In Yudkowsky and Bostrom’s classic “paperclip maximizer” scenario, a human orders an AI to make paperclips. If the AI becomes powerful enough, it does whatever is necessary to make as many paperclips as possible – bulldozing virgin forests to create new paperclip mines, maliciously misinterpreting “paperclip” to mean uselessly tiny paperclips so it can make more of them, even attacking people who try to change its programming or deactivate it (since deactivating it would cause fewer paperclips to exist). You can try adding epicycles in, like “make as many paperclips as possible, unless it kills someone, and also don’t prevent me from turning you off”, but a big chunk of Bostrom’s S:PDS was just example after example of why that wouldn’t work.

Russell argues you can shift the AI’s goal from “follow your master’s commands” to “use your master’s commands as evidence to try to figure out what they actually want, a mysterious true goal which you can only ever estimate with some probability”. Or as he puts it:

> The problem comes from confusing two distinct things: reward signals and actual rewards. In the standard approach to reinforcement learning, these are one and the same. That seems to be a mistake. Instead, they should be treated separately…reward signals provide

> information about the accumulation of actual reward, which is the thing to be maximized.

So suppose I wanted an AI to make paperclips for me, and I tell it “Make paperclips!” The AI already has some basic contextual knowledge about the world that it can use to figure out what I mean, and my utterance “Make paperclips!” further narrows down its guess about what I want. If it’s not sure – if most of its probability mass is on “convert this metal rod here to paperclips” but a little bit is on “take over the entire world and convert it to paperclips”, it will ask me rather than proceed, worried that if it makes the wrong choice it will actually be moving further *away* from its goal (satisfying my mysterious mind-state) rather than towards it.

Or: suppose the AI starts trying to convert my dog into paperclips. I shout “No, wait, not like that!” and lunge to turn it off. The AI interprets my desperate attempt to deactivate it as further evidence about its hidden goal – apparently its current course of action is moving away from my preference rather than towards it. It doesn’t know exactly which of its actions is decreasing its utility function or why, but it knows that continuing to act must be decreasing its utility somehow – I’ve given it evidence of that. So it stays still, happy to be turned off, knowing that being turned off is serving its goal (to achieve my goals, whatever they are) better than staying on.

This also solves the wireheading problem. Suppose you have a reinforcement learner whose reward is you saying “Thank you, you successfully completed that task”. A sufficiently weak robot may have no better way of getting reward than actually performing the task for you; a stronger one will threaten you at gunpoint until you say that sentence a million times, which will provide it with much more reward much faster than taking out your trash or whatever. Russell’s shift in priorities ensures that won’t work. You can still reinforce the robot by saying “Thank you” – that will give it evidence that it succeeded at its real goal of fulfilling your mysterious preference – but the words are only a signpost to the deeper reality; making you say “thank you” again and again will no longer count as success.

All of this sounds almost trivial written out like this, but number one, everything is trivial after someone thinks about it, and number two, there turns out to be a lot of controversial math involved in making it work out (all of which I skipped over). There are also some big remaining implementation hurdles. For example, the section above describes a Bayesian process – start with a prior on what the human wants, then update. But how do you generate the prior? How complicated do you want to make things? Russell walks us through an example where a robot gets great information that a human values paperclips at 80 cents – but the real preference was valuing them at 80 cents on weekends and 12 cents on weekdays. If the robot didn’t consider that a possibility, it would never be able to get there by updating. But if it did consider every single possibility, it would never be able to learn anything beyond “this particular human values paperclips at 80 cents on 12:08 AM on January 14th when she’s standing in her bedroom.” Russell says that there is “no working example” of AIs that can solve this kind of problem, but “the general idea is encompassed within current thinking about machine learning”, which sounds half-meaningless and half-reassuring.

People with a more technical bent than I have might want to look into some deeper criticisms of CIRL, including Eliezer Yudkowsky’s article here and some discussion in the AI Alignment Newsletter.

IV.

I want to end by discussing what was probably supposed to be an irrelevant middle chapter of the book, Misuses of AI.

Russell writes:

> A compassionate and jubilant use of humanity's cosmic endowment sounds wonderful, but we also have to reckon with the rapid rate of innovation in the malfeasance sector. Ill-intentioned people are thinking up new ways to misuse AI so quickly that this chapter is likely to be outdated even before it attains printed form. Think of it not as depressing reading, however, but as a call to act before it is too late.

…and then we get a tour of all the ways AIs are going wrong today: surveillance, drones, deepfakes, algorithmic bias, job loss to automation, social media algorithms, etc.

Some of these are pretty worrying. But not all of them.

Google "deepfakes" and you will find a host of articles claiming that we are about to lose the very concept of truth itself. Brookings calls deepfakes "a threat to truth in politics" and comes up with a scenario where deepfakes "could trigger a nuclear war." *The Guardian* asks "You Thought Fake News Was Bad? Deepfakes Are Where Truth Goes To Die". And these aren't even the alarmist ones! The Irish Times calls it an "information apocalypse" and literally titles their article "Be Afraid"; Good Times just writes "Welcome To Deepfake Hell". Meanwhile, deepfakes have been available for a couple of years now, with no consequences worse than a few teenagers using them to make pornography, ie the expected outcome of every technology ever. Also, it's hard to see why forging videos should be so much worse than forging images through Photoshop, forging documents through whatever document-forgers do, or forging text through lying. Brookings explains that deepfakes might cause nuclear war because someone might forge a video of the President ordering a nuclear strike and then commanders might believe it. But it’s unclear why this is so much more plausible than someone writing a memo saying “Please launch a nuclear strike, sincerely, the President” and commanders believing *that*. Other papers have highlighted the danger of creating a fake sex tape with a politician in order to discredit them, but you can already convincingly Photoshop an explicit photo of your least favorite politician, and everyone will just laugh at you.

Algorithmic bias has also been getting colossal unstoppable neverending near-infinite unbelievable amounts of press lately, but the most popular examples basically boil down to “it’s impossible to satisfy several conflicting definitions of ‘unbiased’ simultaneously, and algorithms do not do this impossible thing”. Humans also do not do the impossible thing. Occasionally someone is able to dig up an example which actually seems slightly worrying, but I have never seen anyone prove (or even seriously argue) that algorithms are *in general* more biased than humans (see also Principles For The Application Of Human Intelligence – no, seriously, see it). Overall I am not sure this deserves all the attention it gets any time someone brings up AI, tech, science, matter, energy, space, time, or the universe.

Or: with all the discussion about how social media algorithms are radicalizing the youth, it was refreshing to read a study investigating whether this was actually true, which found that social media use did not increase support for right-wing populism, and online media use (including social media use) and right-wing populism actually seem to be negatively correlated (remember, correlational studies are always bad). Recent studies of YouTube’s algorithms find they do not naturally tend to radicalize, and may deradicalize, viewers, although I’ve heard some people say this is only true of the current algorithm and the old ones (which were not included in these studies) were much worse.

Or: is automation destroying jobs? Although it seems like it *should*, the evidence continues to suggest that it isn’t. There are various theories for why this should be, most of which suggest it may not destroy jobs in the near future either. See my review of technological unemployment for details.

A careful reading reveals Russell appreciates most of these objections. A less careful reading does *not* reveal this. The general structure is “HERE IS A TERRIFYING WAY THAT AI COULD BE KILLING YOU AND YOUR FAMILY although studies do show that this is probably not literally happening in exactly this way AND YOUR LEADERS ARE POWERLESS TO STOP IT!”

I understand the impulse. This book ends up doing an amazing job of talking about AI safety without sounding weird. And part of how it accomplishes this is building on a foundation of “AI is causing problems now”. The media has already prepared the way; all Russell has to do is vaguely gesture at deepfakes and algorithmic radicalization, and everyone says “Oh yeah, that stuff!” and realizes that they *already* believe AI is dangerous and needs aligning. And then you can add “and future AI will be the same way but even more”, and you’re home free.

But the whole thing makes me nervous. Lots of right-wingers say “climatologists used to worry about global cooling, why should we believe them now about global warming?” They’re wrong – global cooling was never really a big thing. But in 2040, might the same people say “AI scientists used to worry about deepfakes, why should we believe them now about the Singularity?" And might they actually have a point this time? If we get a reputation as the people who fall for every panic about AI, including the ones that in retrospect turn out to be kind of silly, will we eventually cry wolf one too many times and lose our credibility before crunch time?

I think the actual answer to this question is "Haha, as if our society actually punished people for being wrong". The next US presidential election is all set to be Socialists vs. Right-Wing Authoritarians – and I'm still saying with a straight face that the public notices when movements were wrong before and lowers their status? Have the people who said there were WMDs in Iraq lost status? The people who said sanctions on Iraq were killing thousands of children? The people who said Trump was definitely for sure colluding with Russia? The people who said global warming wasn't real? The people who pushed growth mindset as a panacea for twenty years?

So probably this is a brilliant rhetorical strategy with no downsides. But it still gives me a visceral "ick" reaction to associate with something that might not be accurate.

And there's a sense in which this is all obviously ridiculous. The people who think superintelligent robots will destroy humanity – these people should worry about associating with the people who believe fake videos might fool people on YouTube, because the latter group is going beyond what the evidence will support? *Really?* But yes. Really. It's more likely that catastrophic runaway global warming will boil the world a hundred years from now than that it will reach 75 degrees in San Francisco tomorrow (predicted high: 59); extreme scenarios about the far future are more defensible than even weak claims about the present that are ruled out by the evidence.

There's been some discussion in effective altruism recently about public relations. The movement has many convincing hooks (you can save a live for $3000, donating bednets is very effective, think about how you would save a drowning child) and many things its leading intellectuals are actually thinking about (how to stop existential risks, how to make people change careers, how to promote plant-based meat), and the Venn diagram between the hooks and the real topics has only partial overlap. What to do about this? It's a hard question, and I have no strong opinion besides a deep respect for everyone on both sides of it and appreciation for the work they do trying to balance different considerations in creating a better world.

HC's relevance to this debate is as an extraordinary example. If you try to optimize for being good at public relations and convincingness, you can be really, really good at public relations and convincingness, even when you're trying to explain a really difficult idea to a potentially hostile audience. You can do it while still being more accurate, page for page, than a *New York Times* article on the same topic. There are no obvious disadvantages to doing this. It still makes me nervous.

V.

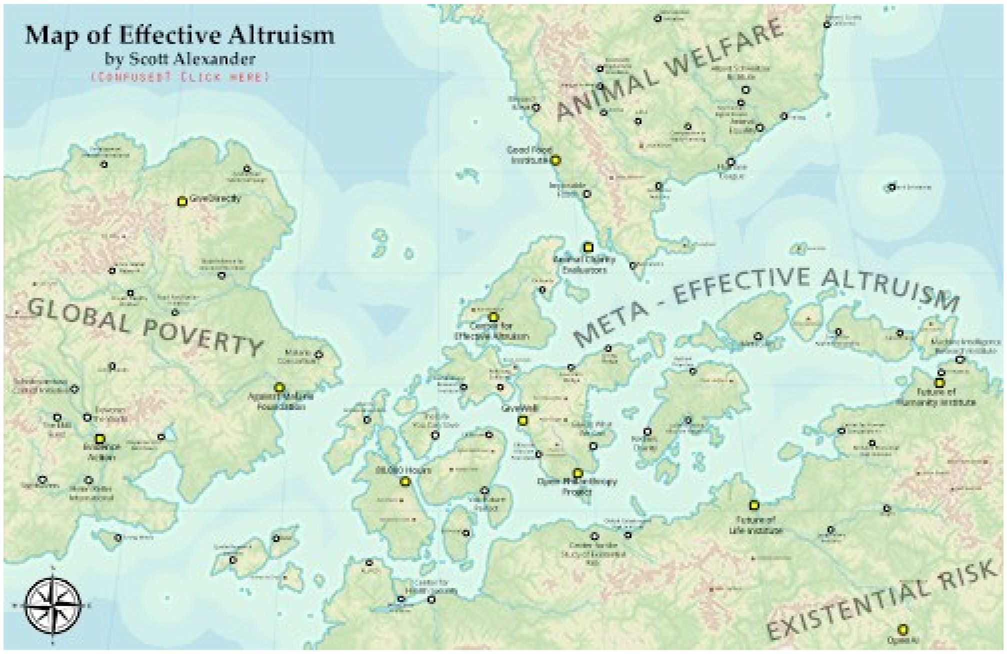
My reaction to this book is probably weird. I got interested in AI safety by hanging out with transhumanists and neophiles who like to come up with the most extreme scenario possible, and then back down when maybe it isn't true. Russell got interested in AI safety by hanging out with sober researchers who like to be as boring and conservative as possible, and then accept new ideas once the evidence for them proves overwhelming. At some point one hopes we meet in the middle. We're almost there.

But maybe we’re not quite there yet. My reaction to this book has been “what an amazing talent Russell must have to build all of this up from normality”. But maybe it’s not talent. Maybe Russell is just recounting his own intellectual journey. Maybe this is what a straightforward examination of AI risk looks like if you have fewer crazy people in your intellectual pedigree than I do.

I recommend this book both for the general public *and* for SSC readers. The general public will learn what AI safety is. SSC readers will learn what AI safety sounds like when it’s someone other than me talking about it. Both lessons are valuable.

Map Of Effective Altruism

In the spirit of my old map of the rationalist diaspora, here's a map of effective altruism:


Map of Effective Altruism
by Scott Alexander
(CONFUSED? CLICK HERE)
ANIMAL WELFARE
GLOBAL POVERTY
META - EFFECTIVE ALTRUISM
EXISTENTIAL RISK
GiveDirectly
Good Food Institute
Animal Charity Evaluators
Centre for Effective Altruism
Malaria Consortium
Against Malaria Foundation
Evidence Action
GiveWell
80,000 Hours
Open Philanthropy Project
Future of Humanity Institute
Future of Life Institute
Centre for the Study of Existential Risk
OpenAI

Thumbail – click to expand

Continents are cause areas; cities are charities or organizations; mountains are individuals. Some things are clickable links with title-text explanations. Thanks to AG for helping me set up the imagemap.

Suicide Hotspots Of The World

[Content warning: suicide, rape, child abuse. Thanks to MC for some help with research.]

I.

Guyana has the highest national suicide rate in the world, 30 people per year per 100,000. Guyana has poverty and crime and those things, but no more so than neighboring Brazil (suicide rate of 6) or Venezuela (suicide rate of 4). What's going on?

One place to start: Guyana is a multi-ethnic country. Is its sky-high suicide rate focused in one ethnic group? The first answer I found was this article by a social justice warrior telling us it constitutes racial "essentialism" to even ask the question. But in the process of telling us exactly what kind of claims we should avoid, she mentions someone bringing up that "80% of the reported suicides are carried out by Indo-Guyanese". I feel like one of those classicists who has reconstructed a lost heresy through hostile quotations in Irenaeus.

Indo-Guyanese aren't American Indians; they're from actual India. Apparently thousands of Indians immigrated to Guyana as indentured laborers in the late 1800s. Most went to Guyana, and somewhat fewer went to neighboring Suriname. Suriname also has a sky-high suicide rate, but slightly less than Guyana's, to the exact degree that its Indian population is slightly less than Guyana's. Basically no Indians went anywhere else in South America, and nowhere else in South America has anywhere near the suicide rate of these two countries. The most Indian regions of Guyana also have the highest suicide rate. Hmmm.

Does India itself have high suicide rates? On average, yes. But India has a lot of weird suicide microclimates. Statewide rates range from from 38 in Sikkim (higher than any country in the world) to 0.5 in Bihar (lower than any country in the world except Barbados). Indo-Guyanese mostly come from Bihar and other low-suicide regions. While I can't rule out that the Indo-Guyanese come from some micro-micro-climate of higher suicidality, this guy claims to have traced them back to some of their ancestral villages and found that those villages have low suicide rates.

So what's going on? Social and Cultural Dimensions of Indian Indentured Labour and Its Diaspora argues that despite the mixed suicide rates in India itself, rates across the Indian Diaspora are universally high. For example:

> The Fiji Indian suicide rate in the period 1900 to 1915 was the highest among all Indian labour importing colonies in Africa and the West Indies, and much higher than in India itself. In Mauritius too, hundreds of indentured Indian laborers committed suicide by jumping from a particular hillock during the indenture period, which acquired the name of 'Suicide Hill', now turned into a monument […]
>
> In his article 'Veil of Dishonor' Lal (1985) describes what officials tend to point out as the primary cause of the Fiji Indian suicides: sexual jealousy arising from the persistent shortage of women on the plantations. The rate of indentured adult Indian females to males in Fiji was only 43 to 100. The intense competition for women among the indentured men was seen as the main reason for male suicides in Fiji. Lomarsh Roopnarine (2007) also shows high rates of suicides among indentured Indians in British Guiana […]

> Although there is no reason to doubt the existence of sexual jealousy, this emphasis on the scarcity of women disregards the arduous circumstances in which the indentured labourers were working, and the disruption of the "integrative institutions" of society – family, marriage, caste, kinship, and religion – as the underlying causes of suicide and other ills affecting the Indian indentured labour population.

Yeah, but arduous circumstances affected dozens of different ethnicities involved in various colonization and forced labor schemes, and most of them didn't have these kinds of suicide rates. I can kind of imagine a story where first-generation laborers had no hope of settling down or raising families, committed suicide at high rates, and that ingrained suicidal tendencies in the culture that never went away. But then how come that didn't happen to eg indentured Englishmen in Virginia?

The incongruously named Vijayakumar and John (2018) blame the Hindu religion. Did you know that the *Ramayana* ends with Rama, three of brothers, and the entire population of his kingdom committing mass suicide by drowning? Or that the *mahaparasthana* is a traditional Hindu method of suicide "where the person walks in a north easterly direction, subsisting only on water and air, until his body sinks to rest"? Any religion that has a traditional direction to walk in while you're committing suicide by starving yourself seems kind of suspicious here. But then how come Hindus in some parts of India have such low suicide rates? How come it's just the diaspora that suffers. The paper suggests maybe it's because religiosity plays a protective effect, but it sounds kind of strained.

I don't have better answers to any of these questions. Maybe the combination of Hindu religion, imbalanced gender ratios, and uprooted communities created a perfect storm. I don’t have any better ideas.

II.

Guyana, at 30 suicides per year per 100,000, is the highest *national* suicide rate in the world. But if Greenland ever wins independence, it will steal first place. Greenland’s suicide rate is 83 per year per 100,000, almost three times higher than any other country in the world.

Like Guyana, this is more ethnic than national. Greenland is mostly Inuit, and Inuit everywhere have equally high suicide rates. The suicide rate in the mostly-Inuit Canadian province of Nunavit is 71 (for comparison, Canada in general is 10). The suicide rate among Alaskan Inuit is 40 (for comparison, the US in general is 14).

This definitely is *not* just because of the cold and darkness. White Alaskans who live right next to Alaskan natives have a rate of about 20, not much higher than the US average. And suicide in Greenland – like everywhere else – peaks in the spring and summer anyway.

Most damning of all, Greenland’s high suicide rates are a recent phenomenon. In 1971, the rate was 4. I didn’t forget a zero there. Fifty years ago, Greenland had one of the lowest suicide rates in the world. But by 1990, it had reached 120 (it’s since come down a little bit). What happened in those twenty years?

You would think limiting it to such a short time period would make things easy. It isn’t. There are two main theories: social alienation, and alcohol.

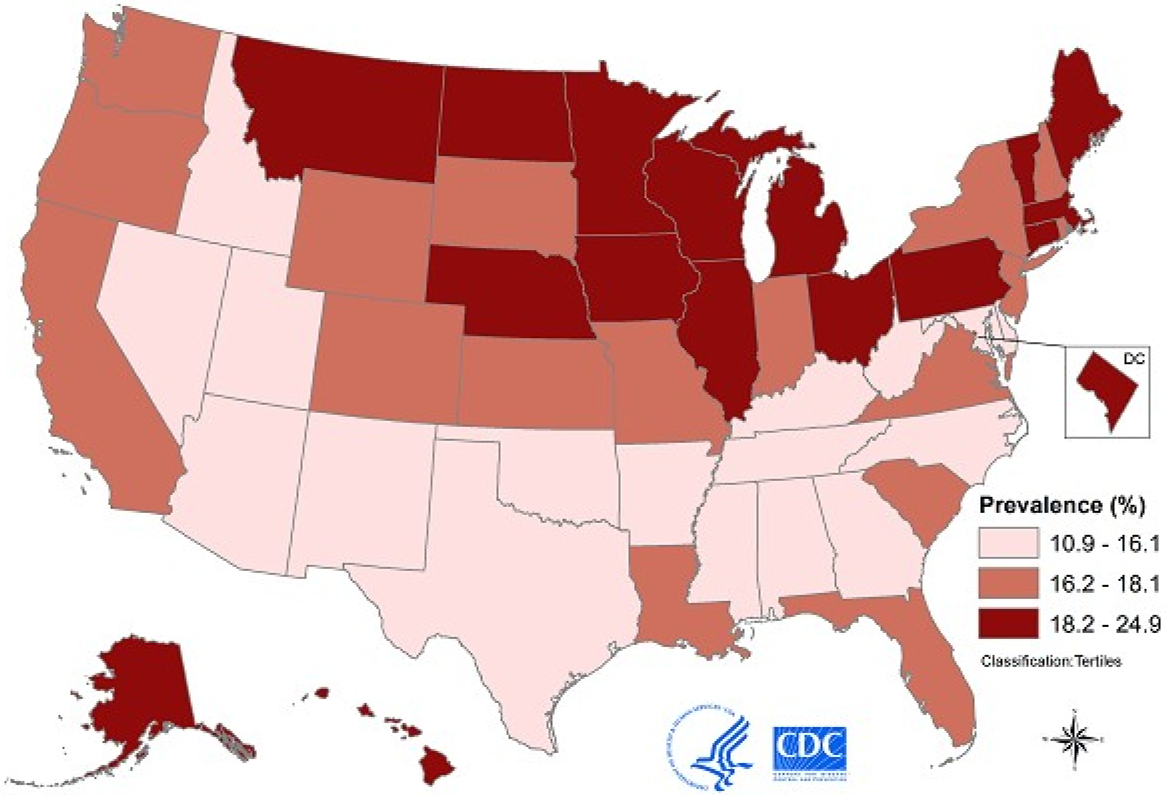
There is definitely a lot of social alienation. For centuries, the Inuit hunted seals in traditional villages. At some point the Danish government decided that was unacceptably backwards and resettled them in cities, especially the capital of Nuuk. This didn’t go well.

One counterargument to this story is that Nuuk has the lowest suicide rate in Greenland, and the more remote the village, the worse the suicide crisis. Maybe you could argue that everywhere was modernized and disrupted and alienated but at least a big city has some interesting stuff to do. This would kind of match the American experience, where it’s small towns in West Virginia that are getting hit by the opioid crisis and deaths of despair.

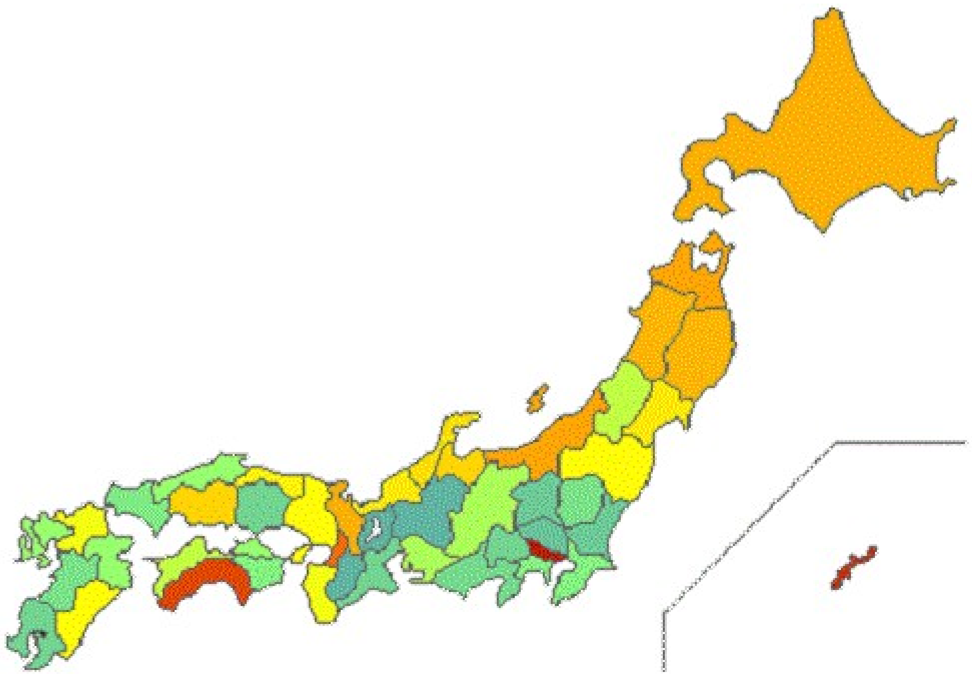
Another counterargument is that all Native American communities suffered a lot of displacement and alienation and modernization, but none of them suffered the same suicide spike as the Inuit. Sources disagree on the exact Native American suicide rate in the US, but it isn’t unusually high; the CDC numbers say it is slightly below the rate for non-Hispanic whites. Canadian First Nations suicide rates *are* elevated, but still only a third or so of Inuit levels. Maybe Inuits suffered stronger relocation pressures than other native peoples because of their Arctic environment? Or maybe every native group suffered a suicide spike, but Native Americans and Canadians have adjusted by now and their suicide rates have come back down? I’m not sure.

The other theory about Greenland is alcohol. Alcohol consumption in Greenland skyrocketed around the same time suicide did, reached levels that temporarily made Greenland by far the most alcoholic country in the world – then started declining around the same time suicide did. This seems to be a pattern when hunter-gatherers with no genetic or cultural resistance encounter alcohol for the first time – Native Americans in the 1700s got up to some crazy stuff.

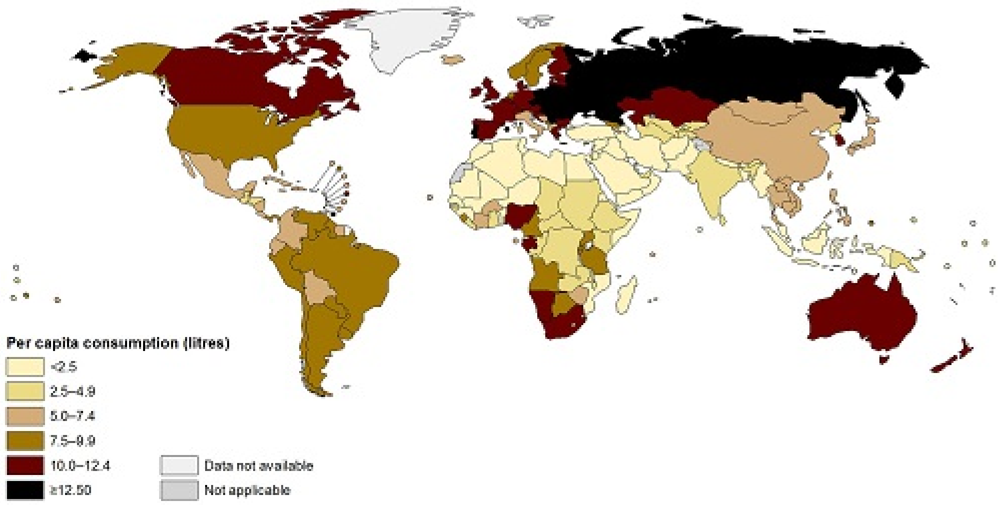
But the Inuit seem to have gotten it much worse. *Now* we can bring back in the cold and darkness. Alcohol consumption seems to increase reliably with latitude, whether we’re talking about the US:


DC
Prevalence (%)
10.9 - 16.1
16.2 - 18.1
18.2 - 24.9
Classification: Tertiles
CDC

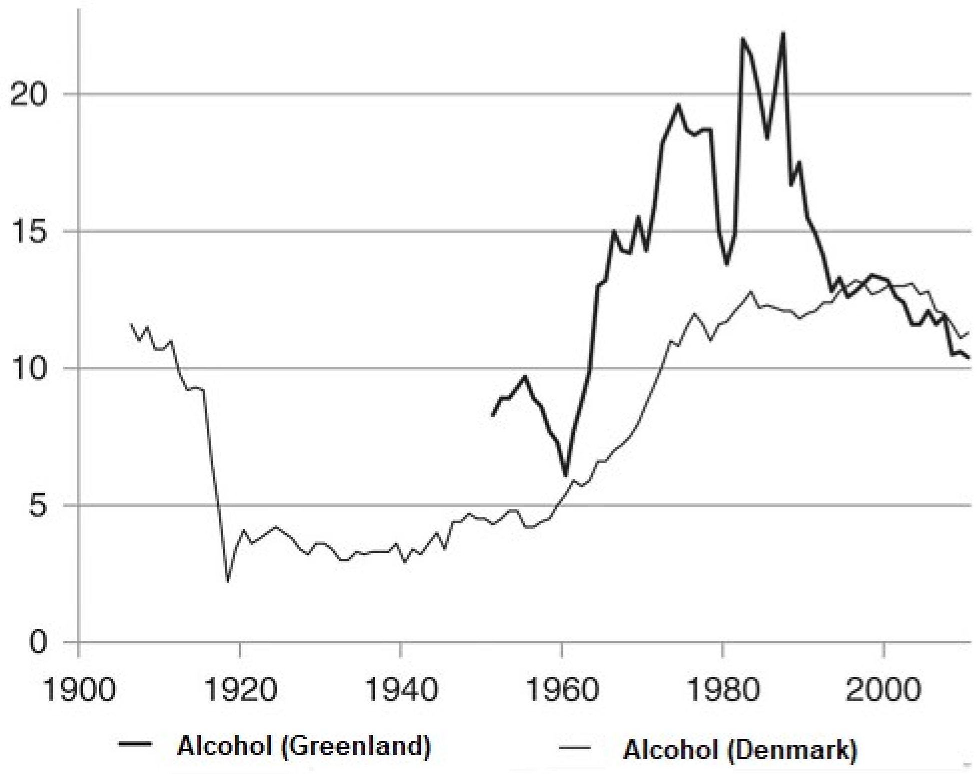
Japan:



Or the whole world:


Per capita consumption (litres)
<2.5
2.5–4.9
5.0–7.4
7.5–9.9
10.0–12.4
≥12.50
Data not available
Not applicable

So you take some hunter-gatherers who have never encountered alcohol before, stick them in the northernmost place in the world, and throw cheap Danish alcohol at them at the *exact* moment their communities are being uprooted and destroyed forever, and you get...well, you get this:



By 1980, Greenland was the most alcoholic country in the world, drinking an average of 22 liters of pure alcohol per capita per year (Russia is 15). It doesn’t look like this was responsible social drinking either. Take the most deranged binge drinking in the worst college fraternity in the world, multiply it by a thousand, and that was Greenland during much of the late 20th century.

But this can’t be the whole story. Alcohol consumption in Greenland has since dropped to the same level as Denmark and other European countries. But the suicide rate is still ten times higher. Why? Maybe the moderate quantities are hiding deeply dysfunctional drinking patterns with lots of binging and addiction.

Or maybe it’s something worse. Child sexual abuse rates in Greenland range from 37% in Nuuk to 46% in East Greenland. As far as I can tell, you are understanding those numbers correctly – almost half of children in Greenland are sexually abused. In Nunavut, the numbers are 52% of women and 22% of men suffering “severe” childhood sexual abuse. The *New York Times* sets a disturbingly vivid scene:

> Pay days are the worst time for the children of Tasiilaq, officials say. With their salaries or social benefits in hand, many adults tend to drink and parents become too inebriated to look after their children, officials say. That’s when an already high rate of sexual abuse rises, according to a police study published last week […]
>
> So on the last Friday of every month, officials open a sports hall in the district as a shelter to keep children away from sexual abuse.
>
> “Children were abused by their stepfathers, cousins and by the neighbor looking after them as the parents were on a bender,” Naasunnguaq Ignatiussen Streymoy, the mother of a sexual abuse victim and an anti-abuse activist, told Weekendavisen, a newsweekly, in an article published on Friday about the crisis.”

Correlation is not causation. Maybe the same dysfunction or social alienation or alcoholism that causes the sexual abuse separately causes the suicides. But maybe the obvious answer is true, and the sexual abuse contributes to the mental health problems that eventually lead to suicide. Maybe a generation of staggeringly high alcoholism led to staggeringly high child abuse, and a generation later those children are still committing suicide at staggeringly high rates.

III.

This is getting really depressing. Let’s talk about something a little bit lighter, like the remote Siberian okrugs with the highest suicide rates in the world.

The highest suicide rate I have seen credibly attributed to an ethnic group is the Chukchi of northeastern Russia, who are said to have reached 165 per year per 100,000 in 1998. They may be distantly related to the Inuit, but I wouldn't put too much weight on this; Siberia is riddled with weird ethnicities with super-high suicide rates. The Evenks reached 121; their western neighbors the Nenets reached 119. There is a group called the Koryaks with a rate of 92, and another group called the Udmurts with a rate of 40ish – which is still higher than Guyana.

Voracek, Fisher, and Marusic try to tie some of these groups into their Finno-Ugrian Suicide Hypothesis, claiming that the genetically-related Finno-Ugric group have a unique predisposition to suicidality. The theory has some superficial plausibility – in the 1990s, the world's first, second, and third most suicidal countries were Finland, Hungary, and Estonia – all Finno-Ugric. Their surrounding non-Finno-Ugric neighbors, like Sweden or Austria, were unremarkable, so a genetic hypothesis made sense. Unfortunately for the theory (but fortunately for everyone else) these countries have since improved by a lot, and now are barely above the world average; improved mental health care may be responsible (and the fall of Communism didn't hurt). I'm actually a little confused what happened here.

But the Finno-Ugric hypothesis can't explain the Chukchi, Evenks, Nenets, Koryaks, and Udmurts. Sure, the Udmurts are Finno-Ugric. And the Nenets are closely related. But the Chukchi, Evenks, and Koryaks aren't. It's tempting to group all of these tiny Siberian ethnic groups together, but eg the Evenks are more closely related to the Japanese than they are to the Nenets (despite living right next to them). Any genetic hypothesis flounders on the sheer genetic diversity and unrelatedness of this region.

Psychologist David Lester tries to point the finger at these groups' ancient culture, which he says has been especially suicidal since the time of the earliest records. He quotes an account of the 19th-century Chukchi:

> Bogoras described the [Chukchi] as irritable and obstinate and, when frustrated, impulsively self-destructive. He reports the case of a young girl who hung herself when her mother refused to take her to a feast in a neighboring camp. [He] reported cases of suicide in a husband over grief at his wife's death andof a mother after her ten-year-old son's death; a case motivated by bad fortune, compounded by the fear of further bad fortune; a woman who no longer found any pleasure in life; a young man who was driven away by his father-in-law for being lazy who then killed his pregnant wife and himself; and a young woman whose husband wanted to lend her to a friend in a group marriage, a friend whom she disliked.
>
> Suicidal behavior appeared to be so common that people planning to kill themselves would often ask for a last meal of exotic tastes before they did so. Some Chukchee prefer to commit suicide by having someone else kill them. The man reported above who committed suicide because of present and anticipated misfortune asked to be strangled. In another case,a man who fell ill asked his wife to shoot him. Bogoras noted that 'voluntary death' as he called it, suicide by getting others to kill oneself, was common for the elderly and those suffering from physical illnesses.

> However, Bogoras also noted ‘peculiar’ causes of voluntary death, such as that of a man who grew weary of quarrelling with his wife over their ill-behaved sons. Part of the motive in these cases may be to induce guilt in the survivors. As one father said, ‘Then he asks to be killed, and charges the very son who offended him with the execution of his request. Let him give me the mortal blow, let him suffer from the memory of it’.

I can only aspire to one day achieve this level of passive-aggressiveness. But in the end it has the same problem as the genetic hypothesis: these groups are just totally unrelated to each other. The Chukchi are not much more suicidal than the Nenets or Evenks, who have none of these traditions. And the Inuit are up there with all of them, and *they* had one of the lowest suicide rates in the world pre-colonization.

I think the explanation here is the same as with Greenland: the combination of alcohol-naive hunter gatherers, alcohol, the alcohol-promoting effects of high latitudes, and a disruption of the traditional way of life. There’s apparently a Russian proverb about Siberians that goes “reindeer-herders are sober only when they don’t have the money to get drunk” – and when the Russians are appalled by your alcoholism, you know you have a problem. Alcohol was found in the blood of 75 – 80% of Nenets suicides. And if anything, the Siberians had their way of life disrupted even worse than the Inuit did – Soviet central planners tried to collectivize them as a PR move – they wanted to demonstrate that Communism could work for even the most primitive of peoples. Well, it didn’t, and here we are.

While genetics or culture may matter a little, overall I am just going to end with a blanket recommendation to avoid being part of any small circumpolar ethnic group that has just discovered alcohol.

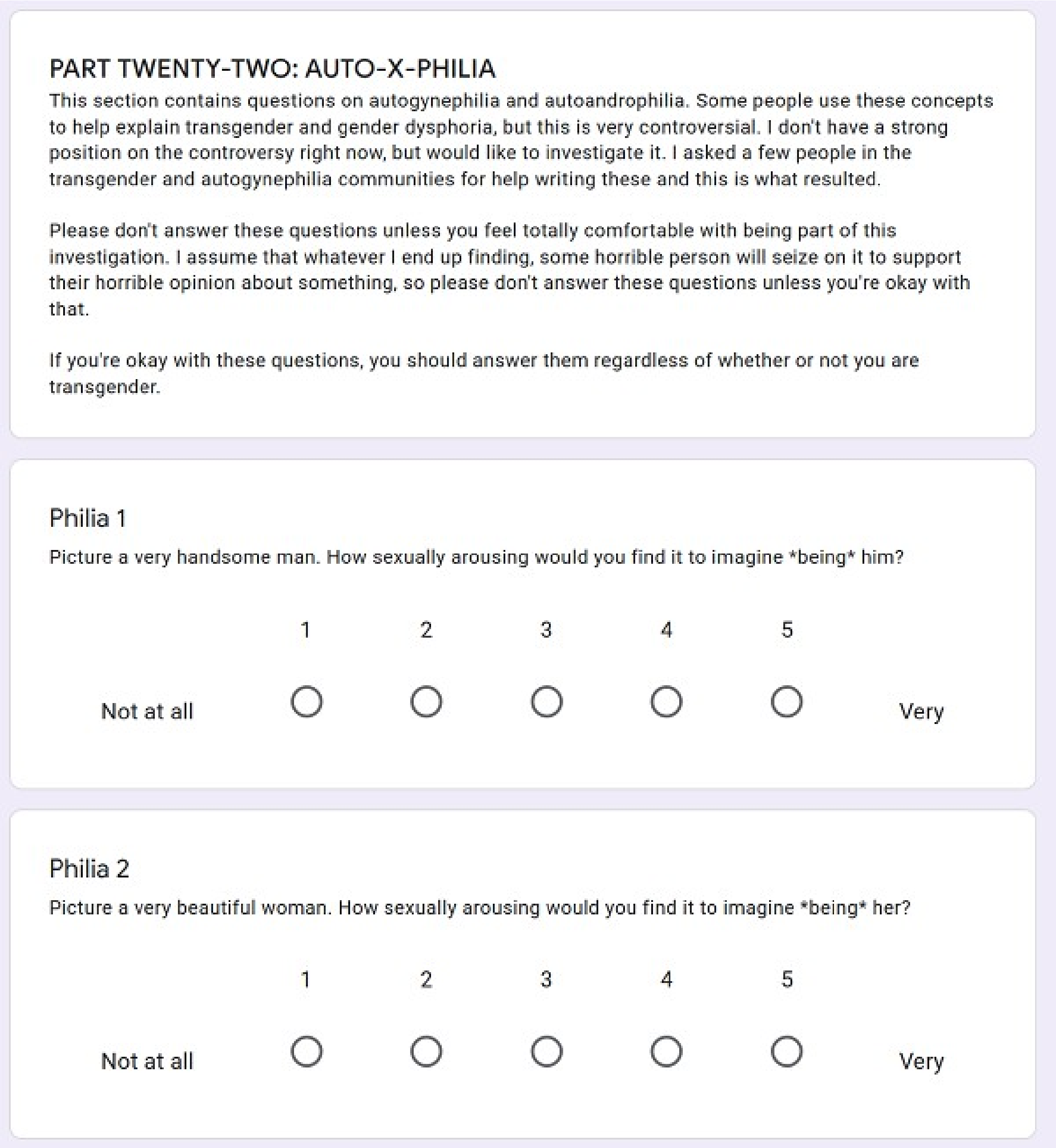
Autogenderphilia Is Common And Not Especially Related To Transgender

“Autogynephilia” means becoming aroused by imagining yourself as a woman. “Autoandrophilia” means becoming aroused by imagining yourself as a man. There’s no term that describes both, but we need one, so let’s say autogenderphilia.

These conditions are famous mostly because a few sexologists, especially Ray Blanchard and Michael Bailey, speculate that they are the most common cause of transgender. They point to studies showing most trans women endorse autogynephilia. Most trans people disagree with this theory, sometimes very strongly, and accuse it of reducing transgender to a fetish.

Without wading into the moral issues around it, I thought it would be interesting to get data from the SSC survey. The following comes partly from my own analyses and partly from wulfrickson’s look at the public survey data on r/TheMotte.

The survey asked the following questions:



PART TWENTY-TWO: AUTO-X-PHILIA

This section contains questions on autogynephilia and autoandrophilia. Some people use these concepts to help explain transgender and gender dysphoria, but this is very controversial. I don't have a strong position on the controversy right now, but would like to investigate it. I asked a few people in the transgender and autogynephilia communities for help writing these and this is what resulted.

Please don't answer these questions unless you feel totally comfortable with being part of this investigation. I assume that whatever I end up finding, some horrible person will seize on it to support their horrible opinion about something, so please don't answer these questions unless you're okay with that.

If you're okay with these questions, you should answer them regardless of whether or not you are transgender.

Philia 1

Picture a very handsome man. How sexually arousing would you find it to imagine *being* him?

	1	2	3	4	5	
Not at all	○	○	○	○	○	Very

Philia 2

Picture a very beautiful woman. How sexually arousing would you find it to imagine *being* her?

	1	2	3	4	5	
Not at all	○	○	○	○	○	Very

First of all, thanks to the 6,715 people (182 trans, 6259 cis, 274 confused) who answered these questions despite my disclaimers. Here’s how it worked out. 5 is maximally autogenderphilic, 1 is no autogenderphilia at all:

Group (n)	Autogynephilia	Autoandrophilia
Cis men (5592)	2.6	1.9
Cis women (667)	2.5	2

Trans men (35)	1.9	2.3
Trans women (147)	3.2	1.3

Group* (n)**	**Autogynephilia (1 – 5)**	**Autoandrophilia (1 – 5)**
Straight cis men (4871)	2.6	1.8
Bi cis men (430)	2.6	3.3
Gay cis men (197)	1.7	3.4
Straight cis women (375)	2.4	1.9
Bi cis women (201)	2.8	2.5
Lesbian cis women (31)	2.5	1.9
Straight trans men (5)	???	???
Bi trans men (19)	???	???
Gay trans men (3)	???	???
Straight trans women (5)	???	???
Bi trans women (76)	3.1	1.4
Lesbian trans women (39)	3.4	1.2

**sexual orientation was self-reported. Almost all transgender people report sexual orientation relative to their current gender rather than their birth gender, so for example a “lesbian trans woman” would be someone who grew up male, currently identifies as female, and is attracted to other women. This is the opposite of how Blanchard and Bailey sometimes use these terms, so be careful comparing these results to theirs!*
***results are marked as ??? for groups with sample size lower than 20*

The survey confirmed Blanchard and Bailey’s finding that many lesbian trans women had strong autogynephilia. But it also confirmed other people’s findings that many cis people also have strong autogenderphilia. In this dataset, autogenderphilia rates in gay cis men were equal to those in lesbian trans women.

Autogenderphilia in cis people was divided between fantasies about being the opposite gender, and fantasies about being the gender they already were. What does it mean to fantasize about being a gender you already are? I asked a cis female friend who admitted to autogynephilia. She told me:

> My literal body is arousing – it’s hot that I have breasts and can get pregnant and have a curvy figure and a feminine face and long hair, and it’s hot to dress up in femme clothes. There are certain gendered/social interactions that are very hot, or that can easily springboard into ones that are very hot. I’ve honestly wondered whether I might not be nonbinary or trans male, because I’m not really sure how euphoric being female is, besides that it’s like living in a sex fantasy.

(score one for the hypothesis that this kind of thing causes gender transition, because after reading this *I* kind of want to be a woman.)

Uh…moving on. The highest rates of autogenderphilia were found in bi cis men (autoandrophilia), gay cis men (autoandrophilia), bi trans women (autogynephilia), and lesbian trans women (autogynephilia).

These groups all have three things in common: they identify as the gender involved, they are attracted to the gender involved, and they are biologically male.

I would guess biological men have more of every fetish, regardless of their current gender identity, so it’s not surprising that they have more autogenderphilia also. In fact, we see that in biological women, the two highest categories are bi cis women (autogynephilia), and lesbian cis women (autogynephilia); again, they identify as the gender involved, and they are attracted to the gender involved.

So abstracting that away, the SSC survey data suggest a very boring hypothesis of autogenderphilia: if you identify as a gender, and you're attracted to that gender, it's a natural leap to be attracted to yourself being that gender.

The SSC survey hypothesis explains the same evidence that Blanchard and Bailey's hypothesis explains (that lesbian trans women very often have autogynephilic fantasies), but reverse the proposed causation: it's not that autogynephilia causes gender transition; it's that identification as a gender is one factor that causes autogenderphilia.

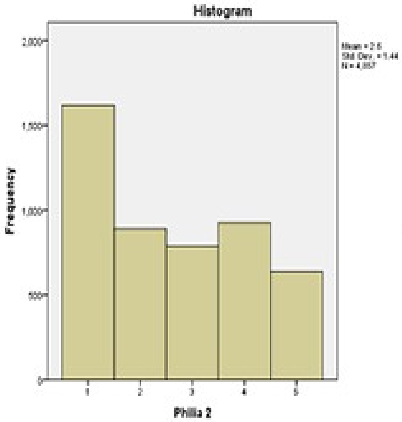
But after that, it can go on to explain other things that Blanchard and Bailey can't explain, like why cis gay men have as much autoandrophilia as trans lesbian women have autogynephilia. Or why some people with low levels of autogenderphilia transition, but many people with high levels don't. I think it's a simpler and more defensible explanation of the evidence.

I asked some people I know who supported Blanchard and Bailey's theory for their thoughts. They focused on a few concerns about the data.

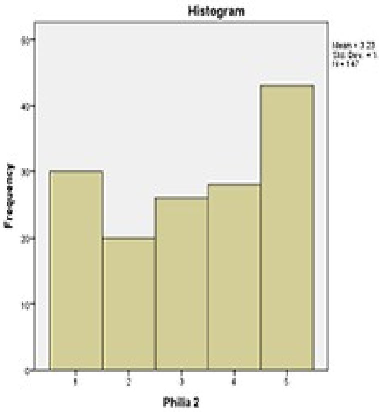
First, weird Internet samples plausibly have more of every paraphilia. This might inflate the *rate for* cis gay men and the *number of* trans lesbian women, assuming the latter all had to be above some cutoff; that might falsely lead me to believe the two groups have the same rate.

One counterargument might be that the responses among cis people alone are enough to generate the hypothesis discussed above. The low rates of autogynephilia in gay men, compared to in straight and bi men, suggest that being attracted to a gender is a prerequisite of autogenderphilia to it. And (adjusting again for the general tendency of male-bodied people to have more fetishes) the higher rates of autogynephilia in cis women/autoandrophilia in cis men, compared to autoandrophilia in cis women/autogynephilia in cis men, suggest that identifying as a gender is a prequisite to autogenderphilia to it.

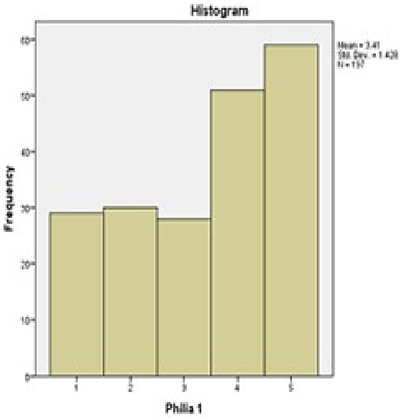
Another counterargument might be the similarity of the histograms produced by cis gay male and trans lesbian female responses; they don't look like they're being generated by two different processes which have only coincidentally averaged out into the same summary statistic:



Autogynephilia
(straight cis men)



Autogynephilia
(lesbian trans women)



Autoandrophilia
(gay cis men)

This doesn't look like all cis men over a certain cutoff are becoming trans women; it looks like the curve for cis gay men and trans lesbian women are being shaped by the same process.

Second, did the survey questions accurately capture autogenderphilia? Fetishes range from very mild to very extreme; some people like being slapped during sex, other people have whole BDSM dungeons in their basement. Is it possible the survey captured some boring meaning of autogenderphilia, like “sure, I guess it would be hot to be a woman”, but some people have a much stronger and more obsessive form? The histogram above argues against this a little, but there might be ceiling effects.

Alice Dreger seems to take something like this perspective here:

> **Q: Do you think autogynephilia might be a part of the female experience, trans or cis? I’ve seen some (very preliminary) theorizing about it as well as a paper with a tiny sample size that suggest that cis women also experience sexual arousal at the thought of themselves as women.**
>
> A: I’ve talked with Blanchard, Bailey, and also Anne Lawrence about this, and my impression is they all doubt cis (non-transgender) women experience sexual arousal at the thought of themselves as women. Clinically, Blanchard observed autogynephilic natal male individuals who were aroused, for example, at the ideas of using a tampon for menses or knitting as a woman with other women. I have never heard a natal woman express sexual arousal at such ideas. I’ve never heard of a natal woman masturbating to such thoughts.

I asked the same cis female friend who gave me the quotation above, and she described using a tampon to masturbate and finding it hot. I think Dreger makes an important point that there are some pretty unusual manifestations of autogenderphilic fetishes out there and we should hesitate before drawing too many conclusions from a single question that lumps them all together. But also, Alice Dreger seems like an really dignified and important person who probably doesn’t hang out with people who talk openly about their menstruation-related masturbation fantasies, and she should probably adjust for that. Maybe she could move to the Bay Area.

There's a common failure mode in psychiatry, where we notice people with some condition doing some weird thing, and fail to notice that huge swathes of people without the condition do the exact same weird thing. For example, everyone knows schizophrenics hear voices, but until recently nobody realized that something like 20% of healthy people do too. Everyone knows that LSD users can end up with permanent visual hallucinations, but until recently nobody realized that lots of drug-free people have the same problem. Schizophrenics definitely hear *more* voices than healthy people, and LSD users have *more* permanent visual hallucinations, but it's movement along the distribution rather than a completely novel phenomenon.

I think autogenderphilia is turning out to work the same way, and that this will require us to reassess the way we think about it.

As usual, I welcome people trying to replicate or expand on these results. All of the data used in this post are freely available and can be downloaded here. I've also heard Michael Bailey is going to release his own interpretation of these data, so stay tuned for that. I'd like to delve into these issues further on future surveys, so let me know if you have ideas about how to do that.

And a big thanks to Tailcalled for helping me set up this section of the survey. If you're interested in these issues, you might enjoy his blog or his own analysis of these results.

Welcome (?), Infowars Readers

Hello to all the new readers I've gotten from, uh, Paul Watson of Infowars. Before anything else, consider reading this statement by the CDC about vaccines.

Still here? Fine.

Infowars linked here with the headline Survey Finds People Who Identify As Left Wing More Likely To Have Been Diagnosed With A Mental Illness. This is accurate only insofar as the result uses the publicly available data I provide. The claim about mental illness was made by Twitter user Philippe Lemoine and not by me. In general, if a third party analyzes SSC survey data, I would prefer that media sources reporting on their analysis attribute it to them, and not to SSC.

As far as I can tell, Lemoine's analysis is accurate enough, but needs some clarifications:

1. Both extreme rightists and extreme leftists are more likely than moderates to have been diagnosed with most conditions.

2. Leftists might be more likely to trust the psychiatric system and get diagnosed. My survey shows some signs of that. Liberals are 60% more likely than conservatives to have formal diagnoses of depression, but only 30% more likely to have a self-diagnosis of depression.

3. Leftists might be more likely to think of their issues through a psychiatric lens than rightists, meaning that even the self-diagnosis numbers might be inflated.

4. The SSC survey is a bad sample to use for this, not just because it’s unrepresentative, but because it might be unrepresentative of different political affiliations in different ways. For example, SSC Marxists really are surprisingly depressed, but maybe the only Marxists who would read an anti-Marxist blog are depressed Marxists looking for things to be miserable and angry about (though see below for some counterevidence).

5. A commenter on Lemoine’s tweet links to this blog post by someone who found the same thing in the General Social Survey. The General Social Survey is much larger and more rigorous than my survey, and there’s no reason to care what my survey has to say when there are GSS results available.

In general, if a survey analysis is posted on this blog, it’s mine. If not, then it isn’t mine and you should link to whoever performed it and let them clean up their own mess. Thanks – and seriously, vaccines are fine.

Confirmation Bias As Misfire Of Normal Bayesian Reasoning

From the subreddit: Humans Are Hardwired To Dismiss Facts That Don’t Fit Their Worldview. Once you get through the preliminary Trump supporter and anti-vaxxer denunciations, it turns out to be an attempt at an evo psych explanation of confirmation bias:

> Our ancestors evolved in small groups, where cooperation and persuasion had at least as much to do with reproductive success as holding accurate factual beliefs about the world. Assimilation into one’s tribe required assimilation into the group’s ideological belief system. An instinctive bias in favor of one’s in-group” and its worldview is deeply ingrained in human psychology.

I think the article as a whole makes good points, but I’m increasingly uncertain that confirmation bias can be separated from normal reasoning.

Suppose that one of my friends says she saw a coyote walk by her house in Berkeley. I know there are coyotes in the hills outside Berkeley, so I am not too surprised; I believe her.

Now suppose that same friend says she saw a polar bear walk by her house. I assume she is mistaken, lying, or hallucinating.

Is this confirmation bias? It sure sounds like it. When someone says something that confirms my preexisting beliefs (eg ‘coyotes live in this area, but not polar bears’), I believe it. If that same person provides the same evidence for something that challenges my preexisting beliefs, I reject it. What am I doing differently from an anti-vaxxer who rejects any information that challenges her preexisting beliefs (eg that vaccines cause autism)?

When new evidence challenges our established priors (eg a friend reports a polar bear, but I have a strong prior that there are no polar bears around), we ought to heavily discount the evidence *and* slightly shift our prior. So I should end up believing that my friend is probably wrong, but I should also be slightly less confident in my assertion that there are no polar bears loose in Berkeley today. This seems sufficient to explain confirmation bias, ie a tendency to stick to what we already believe and reject evidence against it.

The anti-vaxxer is still doing something wrong; she somehow managed to get a very strong prior on a false statement, and isn't weighing the new evidence heavily enough. But I think it's important to note that she's attempting to carry out normal reasoning, and failing, rather than carrying out some special kind of reasoning called "confirmation bias".

There are some important refinements to make to this model – maybe there's a special "emotional reasoning" that locks down priors more tightly, and maybe people naturally overweight priors because that was adaptive in the ancestral environment. Maybe after you add these refinements, you end up at exactly the traditional model of confirmation bias (and the one the Fast Company article is using) and my objection becomes kind of pointless.

But not completely pointless. I still think it's helpful to approach confirmation bias by thinking of it as a normal form of reasoning, and then asking under what conditions it fails.

Addendum to “Targeting Meritocracy”

I’ve always been dissatisfied with Targeting Meritocracy and the comments it got. My position seemed so obvious to me – and the opposite position so obvious to other people – that we both had to be missing something.

Reading it over, I think I was missing the idea of conflict vs mistake theory.

I wrote the post from a mistake theory perspective. The government exists to figure out how to solve problems. Good government officials are the ones who can figure out solutions and implement them effectively. That means we want people who are smart and competent. Since meritocracy means promoting the smartest and most competent people, it is tautologically correct. The only conceivable problem is if we make mistakes in judging intelligence and competence, which is what I spend the rest of the post worrying about.

From a conflict theory perspective, this is bunk. Good government officials are ones who serve *our* class interests and not *their* class interests. At best, merit is uncorrelated with this. At worst, *we* are the lower and middle class, *they* are the upper class, and there is some system in place (eg Ivy League universities) that disproportionately funnels the most meritorious people into the upper class. Then when we put the most meritorious people in government, we are necessarily seeding the government with upper class people who serve upper class interests.

This resolves my confusion about why people disagree with me on this point. It reinforces a lesson I’ve had to learn again and again: if people seem slightly stupid, they’re probably just stupid. But if they seem colossally and inexplicably stupid, you probably differ in some kind of basic assumption so fundamental that you didn’t realize you were assuming it, and should poke at the issue until you figure it out.

Sleep Support: An Individual Randomized Controlled Trial

I worry my sleep quality isn’t great. On weekends, no matter when I go to bed, I sleep until 11 or 12. When I wake up, I feel like I’ve overslept. But if I try to make myself get up earlier, I feel angry and want to go back to sleep.

A supplement company I trust, Nootropics Depot, recently released a new product called Sleep Support. It advertises that, along with helping you fall asleep faster, it can “improve sleep quality” by “improv[ing] sleep architecture, allowing you to achieve higher quality and more refreshing sleep.” I decided to try it.

The first night I took it, I woke up naturally at 9 the next morning, with no desire to go back to sleep. This has *never happened before*. It *shocked* me. And the next morning, *the same thing happened*. I started recommending the supplement to all my friends, some of whom also reported good results.

I decided the next step was to do a randomized controlled trial. I obtained sugar pills, and put both the sugar pills and the Sleep Support pills inside bigger capsules so I couldn’t tell which was which. The recommended dose was two Sleep Support pills per night, so for my 24 night trial I created 12 groups of two Sleep Support pills and 12 groups of two placebo pills.

Then I asked a friend to flip a coin 24 times, and depending on the result place either a pair of Sleep Support pills or a pair of placebo pills in each slot of a monthly pill planner, and record which slot contained which pills on a secret piece of paper I could see at the end of the experiment. Then every weekend night for three months I took the next pair of pills in the planner and recorded:

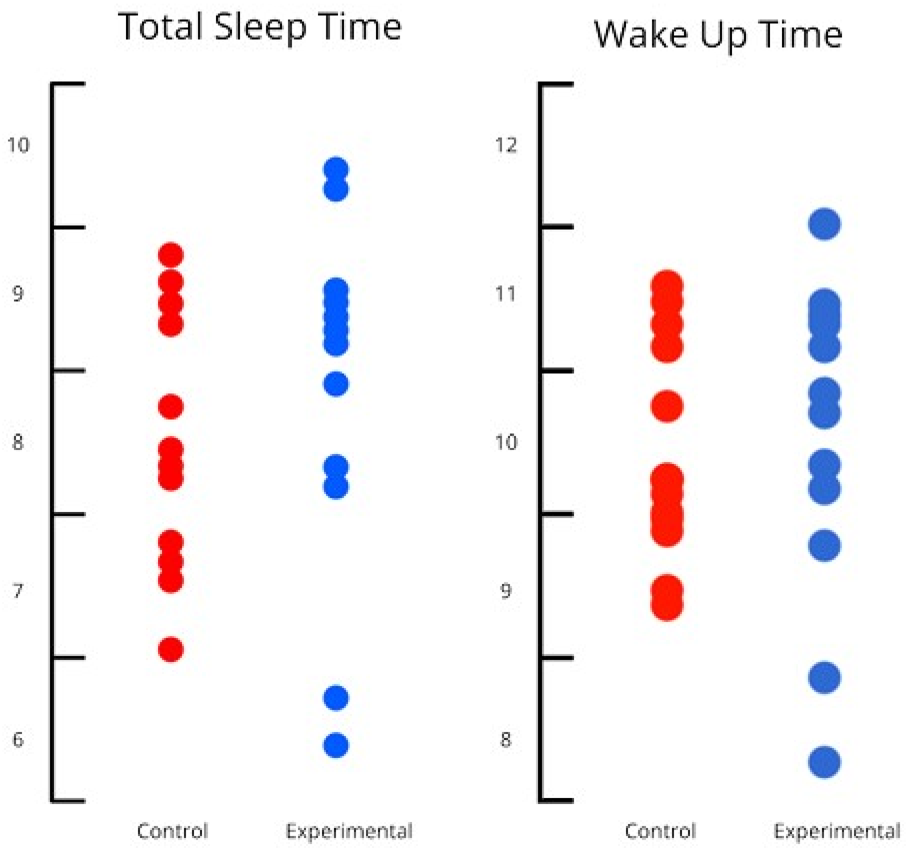
– the time I went to bed
– the time I woke up
– my subjective rating of how well-rested I was upon waking
– my subjective rating of how much energy I felt like I'd had that day
– my subjective rating of how vivid my dreams were that night
– my subjective guess about whether I'd taken placebo or experimental that night

The time I went to bed wasn't intended to be a dependent variable; I generally took the pills just before going to bed, so they couldn't affect that. And I had no way of measuring what time I went to sleep. It was just so that I could measure my total time in bed that night.

The time I woke up was the hardest to operationalize. Usually I wake up a few times in the morning, groggily check the clock, and decide to go back to bed, then wake up for good once it becomes so late I start feeling guilty about how much of the day I'm wasting. I considered setting wake-up time as the very first time I woke up to check the clock, but sometimes I wake up at 5 AM to go to the bathroom, and I didn't want that to get recorded as me "waking up" at 5 AM. And if I used a cutoff like "the first time I wake up after 7", then a night I wake up at 6:59 and go back to bed and wake up for good at 11 would get recorded completely differently from a night I wake up at 7:01 and go back to bed and wake up for good at 11. But if I defined wake-up time as the time I finally woke up for good, then it would be too easy for me to subconsciously bias the experiment. "This feels like a night I took placebo, better stay in bed until at least 11:30".

I decided to eliminate the whole problem by forbidding myself to check the clock while in bed. I would go to sleep, wake up, either decide to go back to sleep or not, and I wasn't allowed to check the clock until I had gotten out of bed and gotten dressed.

Here's the headline results of the experiment – number of hours I slept during experimental vs. control nights, and wake up time during experimental vs. control nights



On average there was no difference between the two groups on either measurement. There was also no difference on any of the subjective measures. My subjective guess about whether I'd taken experimental or placebo capsules that night had no correlation with the reality.

My conclusion isn't that Sleep Support doesn't work; I didn't even try it for its main indication of helping with insomnia. My study was too underpowered to detect even medium-sized effects. And just because it didn't work for me doesn't mean it won't work for somebody else.

My conclusion is that the effect I thought that *I* observed – a consistent change of two hours in my otherwise stable wake-up time – wasn't real. This shocked me. What's going on?

I think my original strategy of “wake up a few times in the morning, check the clock, and finally get out of bed when you really feel like it” is very susceptible to the placebo effect. Usually I might wake up at 9, decide that was too early to face the world, and go back to bed. Maybe I wouldn’t even remember doing this. Part of this was probably inertia – I wasn’t used to getting up at 9, I figured I must not have gotten enough sleep to feel good, and so I didn’t want to do it today. Once I had an exciting new sleep supplement in my system, I woke up at 9, actually checked whether or not I felt ready to wake up, and absent my usual prior that I wasn’t, I found that I was, and woke up.

This hypothesis is supported by the results of the experiment. On about a third of days, I woke up before 10 – again, something I never would have done before starting Sleep Support. I think the active ingredient here was not letting myself look at the clock. Without external cues to tell me how tired I *should* feel, I was forced to rely on how tired I actually felt, which in many cases was “not tired at all”. This happened regardless of whether I was taking Sleep Support or placebo that day.

Ironically, even though the supplement failed to differentiate itself from placebo, I think this is one of the most successful biohacking experiments I’ve ever done. I’m getting up on average an hour or so earlier than I did before, getting more done, and not feeling any more tired by the evening.

Future research: see if this keeps working even now that I know what’s going on.

You can download my raw data here. If you want to replicate this experiment, you can buy Sleep Support capsules here.

There are lots of ways to make a placebo; I found these very large empty capsules helpful.

I’m interested in hearing about anyone else’s experience conducting controlled trials of supplements on themselves; if you do something like this and want to publish it on a blog, let me know.

Book Review: Just Giving

I.

Traditional book reviews tend to focus on a single book, such as *Just Giving* by Rob Reich. We ought, however, to be reviewing a broader question: what is the role of books in a liberal democratic society? And what role *should* they play?

Books were first invented during the early Bronze Age. Plato states people fiercely opposed the first books; in his dialogue *Phaedrus*, he recalls the Egyptian priests' objection to early writing:

> [Writing] will create forgetfulness in the learners' souls, because they will not use their memories; they will trust to the external written characters and not remember of themselves. The specific which you have discovered is an aid not to memory, but to reminiscence, and you give your disciples not truth, but only the semblance of truth; they will be hearers of many things and will have learned nothing; they will appear to be omniscient and will generally know nothing; they will be tiresome company, having the show of wisdom without the reality.

Contrast the Egyptian scribes' reception with the ceaseless praise given to the authors of our age. Rather than asking about the purposes of writing and the power of authors, we tend instead to celebrate writers, large and small, for their brilliance. But in our age, these are questions we should pose with greater urgency. Scholarly literature like *Just Giving* is an unaccountable, nontransparent, and perpetual exercise of power. It deserves more criticism than it has received.

There’s a conventional story to tell about book-writing and its relation to liberty. The story is this. Book-writing is thought to be tightly connected to liberty. This is so for two reasons. First, writing or reading a book is voluntary. Second, the exercise of liberty involves freedom of speech. This story is an attractive one, and it contains some truth. But it ignores that book-writing is inherently embedded in state institutions, like intellectual property laws. It should not be understood in the simplistic manner of an activity that takes place within a framework of nonintervention by the state, or as nothing more than private individual decisions to express thoughts. Instead, it must be understood as embedded in political institutions, laws, and public policy. Books may not be an invention of the state, but they are an artifact of it.

Most importantly, book-writing is heavily subsidized by the government. Authors receive a “pass-through” tax deduction of up to 20%. In addition, they can deduct most of the expenses they incur in writing a book, from freelance editing to literary agents to promotional events. In extending these tax incentives, federal and state treasuries forego tax revenue. Or to put it differently, tax incentives for writing books constitute a a kind of spending program. In fact, the fiscal effects of a direct spending program and a tax expenditure are exactly the same. In Suzanne Mettler’s apt phrase, federal policy driven by tax expenditures rather than direct spending constitutes the “submerged state”, obscured from public view and accountability, but with powerful distributional consequences. These tax breaks amount to massive federal and state subsidies for the creation and dissemination of written texts. They are supplemented by millions spent on libraries, literacy programs, and in some cases direct subsidies to book publishers.

A respect for the liberty of individuals to promote their views is one thing; subsidizing its exercise is another. The state does not merely permit and set guidelines within which writing takes place – offering the state's imprimatur to every book and pamphlet and magazine and journal article – but is in a fiscally meaningful way actively participating in what authors do. If the state is actively funding, through a tax expenditure, some bad book, it makes the state partially complicit in the harm that the book causes. It is no exaggeration to say that as book-writing is currently structured, when authors do harm, so does the state. It is incorrect to say that mediocre books merely waste the time of the author and reader. Rather, writing a mediocre book squanders assets that are partially the public's.

With this description of the relationship between book-writing and liberty in place, let us now consider whether the ideal of equality is playing any role in the institutional design of the policies. The median annual wage of authors is $62,000, twice the average US income of $31,000. Authors are most likely to be college-educated, upper class, and be the sorts of people who can take months off of their jobs to write a book. Scholarly books are often written by professors, a member of a tiny and unelected intellectual elite. This makes their immense ability to exercise power by writing a book and getting it published deeply troubling. Or consider a famous author like Jeffrey Sachs, whose successful books permit him to stride upon the world stage as if he were a head of state.

Perhaps books could play important roles in democratic societies, despite being an exercise of power and expression of plutocratic voice, if they were subject to different legal arrangements. But it is no coincidence that the wealthy dominate book-writing. The tax code is set up to unfairly privilege books by the well-off over the poor in two ways.

First, the pass-through deduction and freelance editor deduction are available only to those individuals who itemize their deductions – people who opt not to take the so-called standard deduction on their income tax. This effectively penalizes, or fails to reward and provide an incentive for, all people who do not itemize their deductions, a group that constitutes roughly 70% of taxpayers. Thus the low-income renter who does not itemize her deductions but pays $500 to get their book edited receives no tax concession, while the high-income house owner who pays the same $500 fee can claim a deduction. Second, the tax subsidy given to those who do receive the deduction possesses what is known as an "upside-down effect". The deduction functions as an increasingly greater subsidy with every higher step in the income tax bracket. Both of these features of the tax code are arbitrary and unfairly benefit the well-off. The choice of the the pass-through and business expenses tax deductions as the preferred tax policy for book-writing introduces a potent plutocratic bias.

Proponents of books might suggest that they nevertheless serve a compelling public interest in the form of spreading knowledge. But when we move away from individual works and consider the total distribution of literature, we find a pattern of writing that is hard to reconcile with expectations of educational outcomes. For anyone who believes that books imply something about knowledge or truth or education, the sunny picture of American book publishing here becomes decidedly cloudy.

Figure 1 demonstrates that the most popular type of book in the US is "mystery, thriller, or true crime". Second and third place are held by history and biographies, which may perhaps be edifying to some people. But after that we get romance, cookbooks, science fiction, and fantasy. Literary classics and books on important current affairs are far down the list, only a fraction of total books read. What can we conclude from these data? The lesson is obvious: if we believe the purpose of reading and writing to be predominantly educational, an important mechanism to provide for our enlightenment and edification, the actually existing distribution of reading in the United States does not meet the test. Not by a long shot.

Finally, we must address the question of intergenerational justice. Books are designed to enshrine author intent and express their opinions in perpetuity. Thus does the dead hand of the author potentially extend from beyond the grave to strangle future generations. John Stuart Mill famously wrote that “There is no fact in history which posterity will find it more difficult to understand, than the idea of perpetuity, and that any of the contrivances of man, should have been coupled together in any sane mind.” Yet authors deliberately “write for the ages”, producing works that can be studied for hundreds or even thousands of years.

We might ask whether books would be a welcome institutional arrangement if we were designing a democratic society from scratch. The catalogue of the oddities of the book suggests a strong case against. Books appear at odds with democracy, for they represent, by definition and by law, the expression of plutocratic voices directed to public education. But why, in a democracy, should the size of one’s wallet give one a greater say in public policy? Why should this plutocratic voice be subsidized by the public? And why should democracy allow this voice to extend across generations in the form of intellectual property laws? It would seem that books are a misplaced plutocratic and powerful element in a democratic society. And we can trace, in the evolution of books, the emergence of a particular kind of high-profile author such as Dan Brown, Stephen King, and

J.K. Rowling, whose activity supplants the state, subverts public education processes, and in so doing diminishes democracy.

I find many points of agreement, especially when considering the actual content of books today. Yet despite all this, I think a role for books can be defended. First, books can help overcome problems in the marketplace of ideas by diminishing government orthodoxy and decentralizing the production of knowledge. Second, because of their size and longevity, books can operate on a different and longer timeline than government propaganda broadcasts, taking risks in the expression of ideas we should not routinely expect to see in press releases by government agencies.

This argument is not intended to justify the full range of legal permissions currently afforded to books, but it provides hints as to what a just literary world might look like. I worry, for instance, about the massive boom in short books. Books with fewer than 150 pages primarily serve the author’s vanity. What loss to public benefit would there be with a minimum page length to publish a book, say 250 or 300 pages? I think very little, and quite possibly there would be some gain, for people with less exciting ideas who could not reach the page threshold might be convinced to help other people with their books rather than writing their own. But even if books of all lengths do partly decentralize the definition and provision of knowledge, the resulting pluralism of literary voices will have a plutocratic, not fully democratic, cast. The experimental or heterodox opinions in books will represent the preferences of the wealthy, not of the wider citizenry. Indeed, there is empirical evidence to suggest that at least in the United States, the very wealthy have significantly more politically conservative preferences than average citizens. Thus, the activity of books, even when it decentralizes the production of knowledge, retains a plutocratic character. Does this mean that we should eliminate books? I do not think so. Perhaps a plutocratic tempering of government orthodoxy is better than no tempering at all. I conclude that the decentralization argument provides a plausibly but not definitive case for books as a democracy-supporting institutional design in our society.

In conclusion, how can we make books more compatible with a democratic society? I propose that instead of giving authors tax deductions, they might receive a certain percent of their expenses paid back to them by the government, capped at $100, and that books with fewer than 150 pages should be banned. Are books democratically required? I am not prepared to answer this question affirmatively, for a democratic government has multiple mechanisms to cultivate pluralism and foster discovery. But I have shown that books are certainly democratically permissible.

II.

Yeah, okay, that was weird.

But I put the blame squarely on the hands of Rob Reich, author of *Just Giving*. The structure, arguments, and most of the individual sentences are his, not mine. I just changed the word “charity” to “books”, and replaced all the charity-related examples with book-related examples. A few parts were edited slightly to make them flow together better, and a few sentences are entirely my own, summarizing parts of the argument that wouldn’t fit into a short blog post.

I wrote this weird edited pastiche/summary because I couldn’t figure out how else to express my frustration at *Just Giving*. The book does not conclude that philanthropy is bad. In the end, it comes out saying that philanthropy is potentially okay and can serve a useful purpose, although the tax incentives around it are weird and should be structured better. But along the way it manages to *darkly hint* that philanthropy is bad about two hundred times on every page. Nothing in the book is *wrong*. But a lot of the right things in it are fully general counterarguments that demand charity display a level of rigor that nothing else has. And the author's interviews and summaries mostly keep the dark hinting while watering down the "actually it is okay" part so much it becomes almost invisible. The resulting style could be used to condemn not just charity but any productive human activity, including the writing of *Just Giving* itself.

For example: charity is not just an activity that takes place in a void. It takes place in a human society. So far, so good – nothing takes place in a void, except maybe space travel. But the book manages to darkly hint that because this is true, any regulation on it is justified. It never *says* this. In the end it doesn't even want to regulate charity. But if you started feeling creeped out by sentences above like "books should not be understood in the simplistic manner of an activity that takes place within a framework of nonintervention by the state, or as nothing more than private individual decisions to express thoughts" or "writing a mediocre book squanders assets that are partially the public's" – if you started thinking "Wait, is he pushing totalitarianism?" – well, both of those are pretty direct quotes from *Just Giving*, and the originals gave me the same level of unease about charity.

Or: it's true that there's a sense in which if the state gives someone a tax deduction for something, it is subsidizing their activity. And it's true that authors can deduct some of their book-writing expenses from their tax bill. But it seems troubling to go from there to calling book-writing "part of the submerged state, obscured from public view and accountability", or to say that now "the state is partially complicit in the harm caused by bad books". Yet both of these are real *Just Giving* sentences too. I find myself much less attached to the tax deduction for authors' business expenses (which may or may not be useful, no strong opinions) than to the project of preventing people from saying things like "Making sure books are good is kind of the responsibility of the state, isn't it?"

Or: it's true that authors just write whatever they want. You could describe this as making them "unaccountable and nontransparent", and "at odds with democracy". But at some point you might think things like "Wait a second, isn't democracy perfectly compatible with private individuals doing their own thing? Are you *sure* you're not thinking of totalitarianism?" Normally I would add something like "… and these considerations become immediately apparently when we're talking about writing books, which makes this a classic case of proving too much", except that to me they *also* become immediately apparent when we're talking about philanthropy, so there must just be some fundamental disconnect going on here.

In a few sections, I "cheated" by using *Just Giving*'s sentences or paragraphs about charitable foundations, rather than philanthropy in general. Reich is not necessarily worried about every charitable donation making "the dead hand of the donor potentially [extend] from beyond the grave to strangle future generations" (yes, this is a real quotation from the book), only about donations from foundations doing that. Still, might this be a little dramatic? Reich treats it as self-evident that permanent foundations are bizarre, maybe literally the *most* bizarre thing, quoting John Stuart Mill's opinion that charitable foundations are "among the grossest and most conspicuous abuses of the time” and that the necessity of banning foundations that outlast their founder’s lifetime is “so obvious that he can scarcely conceive how any earnest inquirer could think otherwise”. Unfortunately for Mill, this is not at all obvious to me, and I was left baffled on this point which the book kind of assumed to be a natural human instinct. Why should my ability to control my donations be limited by something as random as my lifespan? If Bill Gates happens to get hit by a truck tomorrow, does this coincidence have some sort of important moral bearing on how the Gates Foundation’s money should be spent? If we decree it does, this leads to odd conclusions, like that the most important effective altruist cause in the world is encasing Bill Gates in an impenetrable steel shell so that nothing can possibly harm him – do we endorse this use of resources? If we oppose foundations, is Bill Gates still allowed to leave all his money to the single person in the world most aligned with his values, and then hope really hard that this person doesn’t betray him? Isn’t part of the point of law to abstract out things like “people might betray you” and replace them with comfortable ironclad contracts?

(a confession: my point about books being a perpetual exercise of the author’s power in the same way foundations are a perpetual exercise of the founder’s power is unfair, and addressed by Reich in the book. He states that most permanent things wield power only as long as the living choose to humor them – eg a dead author only matters if living people choose to read them and take their advice to heart – but foundations do not need the support of the living as long as the contracts that create them remain enforced).

There is much to like about *Just Giving*. Its breakdown of where charitable dollars actually go (mostly to religious institutions if you’re poor, mostly to colleges and museums if you’re rich) contains data I’ve been looking for a long time, and rightly points out that we should do better. Its discussion of the way tax deductions interact with wealth is interesting, although not obviously more applicable to charity than to book-writing or anything else. Its conclusion – that charity and philanthropic foundations have an important role in diversifying the range of represented interests and experimenting with new social policy – seems dead right, and matches my own thoughts on the subject (and see also this article by Kelsey Piper). I really can't disagree with this book too much on the object level.

And yet if my review sounds scathing, I hope this is a sort of justice. Rob Reich has limited disagreements with charity on the object level, but still manages to write what sounds like a scathing review of it. I think this is bad.

In conclusion, *Just Giving* is a government-subsidized exercise of plutocratic power and plutocratic voice repugnant to the very idea of a democratic society of equals. I hope this gets corrected in any future editions.

[**EDIT:** Professor Reich responds in the comments. Please be polite if you try to discuss this with him. Also, please stop mistaking him for former Labor Secretary Robert Reich, they are two different people.]

Book Review: The Seven Principles For Making Marriage Work

I.

John Gottman is a legendary figure, and the legend is told best by John Gottman. He describes wading into the field of marital counseling as a young psychology postdoc, only to find it was a total mess:

> When we began our research, the wide range of marital therapies based on conflict resolution shared a very high level of relapse. In fact, the best of this type of marital therapy, conducted by Neil Jacobson, had only a 35 to 50 percent success rate. In other words, his own studies showed that only 35 to 50 percent of couples saw a meaningful improvement in their marriages as a result of the therapy. A year later, less than half of that group — or just 18 to 25 percent of all couples who entered therapy — retained these benefits. A while ago, Consumer Reports surveyed a large sample of its members on their experience with all kinds of psychotherapists. Most therapists got very high customer-satisfaction marks—except for the marital ones, who received very poor ratings. Though this survey did not qualify as rigorous scientific research, it confirmed what most professionals in the field already knew: in the long run, marital therapy did not benefit the majority of couples.

Gottman decided the field needed statistical rigor, and that he – a former MIT math major – was exactly the guy to enforce it. He set up a model apartment in his University of Washington research center – affectionately called “the Love Lab”, and invited hundreds of couples to spend a few days there – observed, videotaped, and attached to electrodes collecting information on every detail of their physiology. While at the lab, the couples went through their ordinary lives. They experienced love, hatred, romantic dinners, screaming matches, and occasionally self-transformation. Then Gottman monitored them for years, seeing who made things work and who got divorced. Did you know that if a husband fails to acknowledge his wife's feelings during an argument, there is an 81% chance it will damage the marriage? Or that 69% of marital conflicts are about long-term problems rather than specific situations? John Gottman knows all of this and much, much more.

Using his mountain of data (the legend continues) Gottman became a Divorce Prophet:

> After years of research…I am now able to predict whether a couple will stay happily together or lose their way. I can make this prediction after listening to the couple interact in our Love Lab for as little as five minutes! My accuracy rate in these predictions averages 91 percent over three separate studies. In other words, in 91 percent of the cases where I have predicted that a couple's marriage would eventually fail or succeed, time has proven me right. These predictions are not based on my intuition or preconceived notions of what marriage "should" be, but on the data I've accumulated over years of study.

…which is pretty interesting. But predicting destiny is only an intermediate step – as another legend once said, "the point is to change it". So, science in hand, John Gottman resolved to fix marital counseling. And apparently succeeded:

> We found that at the beginning of our workshops, 27 percent of couples were at very high risk for divorce. At our three-month followup that proportion was 6.7 percent and at nine months it was 0 percent. But even couples who were not at high risk for divorce were significantly helped by the workshops.

Twenty years later, the legend has spread to every corner of the world. He has received glowing praise from The New York Times, The Atlantic, BBC, CNN, Washington Post, The New York Times again, Harvard Business Review, Scientific American, Time, and The New York Times a third time. He has published over two hundred scientific papers, some of which have been cited thousands of times. He has been voted one of the top 10 most influential therapists of the past quarter-century. His Gottman Relationship Institute, founded together with his wife Julie Gottman, has become a marriage counseling empire, trained hundreds of therapists in the Gottman method, and operates a referral network that can find you Certified Gottman Level 3 Therapists from Australia to Uruguay. After a long life of helping save countless marriages, his one regret is that he is so great he can no longer find an adequate control group for his studies:

> When we sat down to write the first edition of this book, we were excited to share the results of laboratory research into relationships but we knew we’d face some skepticism. Could scientific study of something as intangible, idiosyncratic, and personal as romantic love deliver useful advice to couples in the real world? Well, more than fifteen years and millions of readers later we are happy to report that The Seven Principles for Making Marriage Work has done just that. Countless readers across the globe tell us that the book’s strategies have enhanced, shielded, or saved their relationship. We have received thank-yous from every imaginable type of

> couple, including newlyweds, traditional spouses, two-career partners, devoutly religious spouses, military couples, cohabitants, same-sex partners, not-yet-marrieds, divorced people looking toward the future, and counselors who work with all of the above.
>
> It is a great source of satisfaction and pride that we have been able to help so many people. We're also gratified that research continues to confirm what these readers consistently tell us: The Seven Principles can have a powerfully positive effect on your relationship. In fact, a randomized clinical study by John and his coresearchers (Julia Babcock, Kim Ryan, and Julie Gottman) found that married couples who simply read The Seven Principles and worked through the quizzes and exercises on their own (but received no additional professional aid) were significantly happier in their relationship, and these effects lasted when assessed a year later. Simply reading this book proved so successful that it actually bollixed the research: the original experiment had been designed to use these "book-only" couples as a control group to test marital therapy techniques!

Sounds like some book! And God knows we need good marriage therapy. The world needs it. And I need it in particular. I am a psychiatrist. I am trained to treat depression and schizophrenia and nice simple things like that. But somehow, I keep getting patients who need help with relationship problems. I am totally unprepared for this. In the past, my advice has been "go find someone trained by John Gottman, I hear he is some kind of living legend". But at some point, I figured I should finally read his book, *The Seven Principles For Making Marriage Work: A Practical Guide From The Country's Foremost Relationship Expert*, so that I could have an informed opinion on this and maybe try helping people directly.

This is a review of that book. It starts with a summary of Gottman’s marital counseling techniques, continues to a discussion of the evidence for and against them, and ends with some random thoughts about marriage.

II.

The secret to a happy marriage is that you should like your spouse.

Maybe this doesn’t sound *especially* secret, when put that way. But part of the Gottman legend is that the old school of marriage therapists kind of missed that part. They were really into solving conflicts and having good communication skills and things like that. But over his years of monitoring hundreds of couples in real-world situations, Gottman found that this was overrated. Plenty of couples had atrocious communication skills and got in conflicts all the time, but loved each other very much and had no real marital problems. Plenty of other couples had finely-polished communication skills and always used “I statements” and things like that, and still ended up divorced. Communication skills are good, and you should definitely try to have them, but you’re putting the cart before the horse unless you focus on *liking your spouse*.

I totally believe this. I remember my grandparents used to fight all the time. Any time my grandmother said something, my grandfather would disagree with her, and vice versa, and the ensuing argument would (to my young ears) sound absolutely *vicious*, and then they would laugh it off and forget about it and continue being wonderfully and obviously in love. So fine, focus on liking your spouse. But how do you make that happen?

Seven Principles excels in its selection of worksheets, activities, and games. There is a seven-week program of talking about one nice thing with your spouse every day, a “Building Your Love Maps” game, and a list of questions you and your spouse should answer together one by one. The underlying principle seems to be that to know somebody is to love them. If you get a really good mental model of where your spouse is in their life, their hopes, their dreams, their fears, their daily toil, then it’s hard to not have at least some fondness for them. One activity is a quiz, with questions like:

1. I can name my partner’s best friends.
2. I can tell you what stresses my partner is currently facing.
3. I know the names of some of the people who have been irritating my partner lately.
4. I can tell you some of my partner’s life dreams.
5. I am very familiar with my partner’s religious beliefs and ideas.
6. I can tell you about my partner’s basic philosophy of life.
7. I can list the relatives my partner likes the least.
8. I know my partner’s favorite music.
9. I can list my partner’s three favorite movies.

The more you know, the more likely your marriage is to make it.

Along with knowing the big things, you should also know what they’re doing on a day-to-day basis. You should conspicuously make sure to know it. Apart from whatever other exercise you’re doing each day, Gottman recommends a ritual of checking in after work and exchanging stories about your days. This time is a Designated Support Zone, no criticism allowed. You take your spouse’s side whether you secretly disagree with them or not. If your spouse gets angry that a police officer gave them a ticket for driving 110 mph through a 25 mph school zone, you are obligated by the terms of your marriage contract to shake your head and say “I know, cops these days have no respect.”

Gottman is *slightly* less strict in other situations, but he still thinks it’s very important that you take your spouse’s side in conflicts. He especially highlights a common dynamic where your parents are always trying to cause trouble between you and your spouse, and your marriage will be in danger until you commit to side with your spouse whenever this kind of thing comes up:

> At the core of the tension is a turf battle between [the wife and the mother] for the husband’s love. The wife is watching to see whether her husband backs her or his mother. She is wondering, “Which family are you really in?” Often the mother is asking the same question. The man, for his part, just wishes the two women could get along better. He loves them both and does not want to have to choose. The whole idea is ridiculous to him. After all, he has loyalties to each, and he must honor and respect both. Unfortunately, this attitude often throws him into the role of peacemaker or mediator, which invariably makes the situation worse.
>
> The only way out of this dilemma is for the husband to side with his wife against his mother. Although this may sound harsh, remember that one of the basic tasks of a marriage is to establish a sense of “we-ness” between husband and wife. So the husband must let his mother

> know that his wife does indeed come first. His house is his and his wife’s house, not his mother’s. He is a husband first, then a son. This is not a pleasant position to take. His mother’s feelings may be hurt. But eventually she will probably adjust to the reality that her son’s family unit, where he is the husband, takes precedence to him over all others. It is absolutely critical for the marriage that the husband be firm about this, even if he feels unfairly put upon and even if his mother cannot accept the new reality.
>
> This is not to suggest that a man do anything that he feels demeans and dishonors his parents or goes against his basic values. He should not compromise who he is. But he has to stand with his wife and not in the middle

A final method of making yourself like your spouse: just explicitly and consciously focus on their good qualities. One of the worksheets lists a bunch of good qualities and asks you to pick some you appreciate in your spouse and explain why. Another asks you to reminisce about the old days, on the theory that presumably you liked your spouse back when you decided to marry them. This isn’t always true – Gottman finds that couples who are really far gone will export their present hostility back into the past and talk about how they always knew it was a bad idea. With typical statistical precision, he notes that “94 percent of the time that couples who put a positive spin on their marriage’s history are likely to have a happy future as well. When happy memories are distorted, it’s a sign that the marriage needs help.” But it’s usually true enough to get both partners warming up to each other a little.

But okay. You’ve done all this stuff and you like your spouse at least a little. Now what?

Now you can start learning communication skills.

Gottman’s communication skills work focuses on what he calls “The Four Horsemen Of The Apocalypse”: criticism, contempt, defensiveness, and stonewalling. These may seem obvious, though Gottman takes some of them in non-obvious directions – “couples who are contemptuous of each other are more likely to suffer from infectious illnesses (colds, flu, and so on) than other people”. Avoiding the four horsemen is vital, especially in the crucial “startup” phase of a conversation. How do you do it?

Some tips are very small in scale. Use I-Statements like “I have trouble dealing with how messy things are sometimes” instead of You-Statements like “You never clean up around here”. Be polite. Try to make “repair attempts” – conflict-ending attempts that can be anywhere from “Hey, you’re yelling at me” to “You are right and I am wrong and I am sorry” to “Oh my god, listen to us, let’s get a drink and never speak of this matter again”. Accept your spouse’s repair attempts when offered.

Others are more general. Don’t have a fight when you’re physiologically aroused. *Don’t have a fight when you’re physiologically aroused*. Monitor your physiological arousal, and if you start to notice the signs – fast heartbeat, tense muscles, shallow breathing – call a time-out, go somewhere else, and use meditation or deep breathing or whatever to calm down.

This is especially important for men. Gottman has strong opinions on gender. He uncritically accepts the feminist view that men feel entitled because of patriarchy and that if they feel angry or upset it’s probably just their entitlement flaring up again. He flirts with saying that men should generally yield to their wife in a conflict (presumably because, thanks to patriarchy, everything will always be biased in favor of the man and so the wife is usually right). In the end, he softens this to a statement that men should “accept influence from” their wives, but also heavily implying that a man who doesn’t give in to his wife must not be accepting her influence – for example, an exercise on page 118 asks men to describe how they would accept their wife’s influence in various situations, and includes an answer key where the right answer is always to say she is right and do what she wants. You might object to this, but sorry, it is Evidence Based According To Science. Gottman tells us that only 35% of husbands are emotionally intelligent, and that “when a man is not willing to share influence with his partner, there is an 81% chance that his marriage will self-destruct”.

But Gottman’s most controversial views on gender involve physiology. He proposes an evo psych explanation for why men can’t handle talking about problems:

> This is not because of some lack on the man’s part. The reason lies in our evolutionary heritage. Anthropological evidence suggests that we evolved from hominids whose lives were circumscribed by very rigid gender roles, since these were advantageous to survival in a harsh environment. The females specialized in nurturing children while the males specialized in cooperative hunting.
>
> As any nursing mother can tell you, the amount of milk you produce is affected by how relaxed you feel, which is related to the release of the hormone oxytocin in the brain. So natural selection would favor a female who could quickly soothe herself and calm down after feeling stressed. Her ability to remain composed could enhance her children’s chances of survival by optimizing the

amount of nutrition they received. But in the male natural selection would reward the opposite response. For these early cooperative hunters, maintaining vigilance was a key survival skill. So males whose adrenaline kicked in quite readily and who did not calm down so easily were more likely to survive and procreate. To this day, the male cardiovascular system remains more reactive than the female and slower to recover from stress […]

This gender difference in how physiologically reactive our bodies are also influences what men and women tend to think about when they experience marital stress. As part of some experiments, we ask couples to watch themselves arguing on tape and then tell us what they were thinking when our sensors detected they were flooded. Their answers suggest that men have a greater tendency to have negative thoughts that maintain their distress, while women are more likely to think soothing thoughts that help them calm down and be conciliatory. Men, generally, either think about how righteous and indignant they feel ("I'm going to get even," "I don't have to take this"), which tends to lead to contempt or belligerence. Or they think about themselves as an innocent victim of their wife's wrath or complaint ("Why is she always blaming me?"), which leads to defensiveness.

Obviously these rules don't hold for every male and every female. But after twenty-five years of research, I have noted that the majority of couples do follow these gender differences in physiological and psychological reactions to stress. Because of these dissimilarities, most marriages (including healthy, happy ones) follow a comparable pattern of conflict in which the wife, who is constitutionally better able to handle the stress, brings

> up sensitive issues. The husband, who is not as able to cope with it, will attempt to avoid getting into the subject. He may become defensive and stonewall. Or he may even become belligerent or contemptuous in an attempt to silence her.

The problem is, men are just too flighty and emotional! They need a rational, hard-headed woman to take care of them and keep them grounded!

Sorry. To be more serious, he thinks that because women are emotionally stable and men aren’t, women tend to bring up long-standing problems that need to be solved, and because men can’t handle this level of stress they panic or shut down or blow up or otherwise start a conflict in order to avoid having to deal with it. “More than 80 percent of the time it’s the wife who brings up sticky marital issues, while the husband tries to avoid discussing them.” The solution is for men to learn calming techniques so they don’t stop the conversation as quickly.

Okay, so now you like your spouse *and* you know how to communicate. Now you can prepare to actually solve some conflicts.

But not all of them. Gottman divides conflicts into two types: solvable and unsolvable. Solvable conflicts are simple, specific, and about the thing they seem to be about – for example, the husband is supposed to take the trash out after work, but work has gotten really stressful lately and he keeps forgetting, and now the trash is overflowing and the wife is annoyed. The solution here is to use normal problem-solving techniques. Put a sign in the bedroom saying “DID YOU REMEMBER TO TAKE THE TRASH OUT?” or something. Whatever.

Unsolvable conflicts are temporary manifestations of deep psychological issues. The particular thing that sparked the fight this time is irrelevant, but both spouses will fight to the death because it represents something important. For example, the husband is late to dinner one night because he went out to the bar. The wife yells at him and says he doesn’t care about her. He yells back that she’s a control freak. Here the problem will not be solved by coming up with a compromise where he can go to the bar half of nights. The problem is that she secretly worries his drinking buddies have a closer connection to him than she ever will, that he doesn’t love her anymore, that he goes to the bar to escape her. He worries that he’s lost his freedom, that he’s become emasculated, that he’s become some boring old person who is never allowed to have fun. If the bar burnt down tomorrow, they would find some other excuse to fight over this dynamic.

There’s no hard line between solvable and unsolvable conflicts. One couple’s solvable problem might be another’s unsolvable one. Forgetting to take the trash out becomes an unsolvable problem if it represents how the husband is irresponsible or the wife is too controlling. Staying out too late at the bar is a solveable problem if it’s just that dinner is getting cold and neither of them has any problem with the husband eating leftovers. But usually couples can figure out whether their particular issue taps into deeper roots.

Gottman suggests dealing with unsolvable conflicts by making the underlying “dreams” explicit. He recommends both partners talk about what the dream driving their side of the conflict means to them. So for example, the husband who stays late at the bar might say “When I was young, my dad was so poor he had to work twelve hour days. Then he would come home, do some chores, go to sleep, and start all over again. I told myself that if I ever ended up like that, I might as well just die, because it seemed like such a crappy and joyless life. To me, getting to go have fun with people means that I'm successful enough that I don't have to end up like my dad." While one spouse (the "dreamer") is describing this, the other spouse has to be completely supportive and try to understand them, without pointing out ways they're wrong.

Once both partners feel like they've been heard and that they understand each other, they discuss the absolute minimum they would need in order to feel like their dreams were being respected, versus the beyond-minimum things that they're willing to be flexible on. Then they both agree to a compromise that gives both of them their bare minimum and splits the difference on the flexible parts. What if the two bare minima are mutually exclusive? The book doesn't say. Probably you got a defective dream, and you should go to the dream factory and ask if they take returns.

The last part of the book is maybe the least actionable, but also my favorite. It discusses couples as almost miniature cultures, with their own rituals, in-jokes, ideologies, and systems of meaning. Partly it's about how to create these things, partly it's about how to acknowledge these things, and partly it's just John Gottman's love letter to the concept of couplehood. It's really heart-warming. Don't make the same mistake I did and read it when you're feeling lonely.

III.

Okay, but can John Gottman *really* predict divorce with 91% success rate and do all the other things he says he can do? Haha, no. All of that stuff is totally false.

Richard Heyman published the definitive paper on this in 2001, The Hazards Of Predicting Divorce Without Crossvalidation (kudos to Laurie Abraham of Slate, the only one of the journalists covering Gottman to find and mention this, and my source for some of the following). Heyman notes that Gottman doesn't predict divorce at all. He postdicts it. He gets 100 (or however many) couples, sees how many divorced, and then finds a set of factors that explain what happened.

Confused about the difference between prediction and postdiction? It's a confusing concept, but let me give an example, loosely based on this Wikipedia article. The following rule accurately matches the results of every US presidential election since 1932: the incumbent party will win the election if and only if the Washington Redskins won their last home game before the election – unless the incumbent is black or the challenger attended a Central European boarding school, in which case it will lose.

In common language, we might say that this rule "predicts" the last 22 presidential elections, in the sense that knowing the rule and the Redskins' record, we can generate the presidential winners. But really it doesn't predict anything – there's no reason to think any future presidential elections will follow the rule. It's just somebody looking to see what things coincidentally matched information that we already have. This is properly called postdiction – finding rules that describe things we already know.

Postdictive ability often implies predictive ability. If I read over hospital records and find that only immunodeficient people caught a certain virus, I might conclude I've found a natural law – the virus only infects immunodeficient people – and predict that the pattern will continue in the future.

But this isn’t always true. Sometimes, especially when you’re using small datasets with lots of variables, you get predictive rules that work very well, not because they describe natural laws, but just by coincidence. It’s coincidence that the Redskins’ win-loss record matches presidential elections, and with n = 22 datapoints, you’re almost certain to get some coincidences like that.

Even an honest attempt to use plausible variables to postdict a large dataset will give you a prediction rule that’s a combination of real natural law and spurious coincidence. So you’re not allowed to claim a certain specific level of predictive ability until you’ve used your rule to predict out-of-training-data events. Gottman didn’t do this.

In his paper, Heyman creates a divorce prediction algorithm out of basic demographic data: husband and wife’s education level, employment status, etc. He is able to achieve 90% predictive success on the training data – nearly identical to Gottman’s 91% – without any of Gottman’s hard work. No making the couples spend days in a laboratory and counting up how many times they use I-statements. No monitoring their blood pressure as they gaze into each other’s eyes. Heyman never met any of his couples at all, let along analyzed their interaction patterns. And he did just as well as Gottman did at predicting divorce (technically he predicted low scores on a measure of marital stability; his dataset did not include divorce outcomes).

Then he applied his prediction rule to out-of-sample couples. Accuracy dropped to 70%. We have no reason not to think Gottman’s accuracy would drop at the same rate. But 70% is around the accuracy you get if you predict nobody will divorce. It’s little better than chance, and all of Gottman’s claims to be a master divorce predictor are totally baseless.

The first question on Gottman's FAQ is whether he is doing this. He says he did this once, but that 6 of his 7 studies have been properly predictive. But Wikipedia notes that this claim uses "a non-standard definition of prediction in which all that is required is that predictive variables, but not their specific relationship to the outcome, were selected in advance". Heyman and Abraham specifically criticize the 6 studies that Gottman calls genuinely predictive as being postdictive. I cannot find all the relevant studies, as many of them are in books, but contra the FAQ it looks as if he is still postdicting.

It's hard for me to dismiss this as an honest mistake. Gottman constantly plays up his credentials as a mathematician and statistician, saying that:

> In the beginning, Dr. John Gottman's research was devoted to the discovery of reliable patterns in observational data. He wanted to see if there were indeed patterns of behavior, or sequences of interactions, that could discriminate happy from unhappy couples. It was not at all clear that these patterns existed. Dr. Gottman and his colleagues began developing the math for sequential analysis, which now is a well-developed methodology.

Anyone smart enough to invent new mathematical methodologies should also be smart enough to know you can't validate your predictor on its training data, so he must know exactly what he's doing. And it would be so easy to fix if he wanted to! All he needs to do is take one of his predictors, apply it to data that wasn't its training data, and tell us how it does! How could this be an innocent mistake?!

And speaking of things that an MIT mathematician should know better than to do, what's up with claims like "when a man is unwilling to share power with his partner, there is an 81% chance his marriage will self-destruct”. Obviously you can replace 81% with any number you want by operationalizing “unwilling to share power with his partner” differently. This sentence as written is totally meaningless. It could potentially be part of a good study, where Gottman investigated different forms of power-sharing and how they affected marital stability – but only because the study would carefully explain its methods so that the end number meant something. Outside the context of that number, it communicates nothing. Yet Gottman put it in a book and expects us to be impressed by it. It feels like he’s just trying to dazzle us with mathematical precision, and hoping we don’t think about it long enough to realize that it doesn’t make sense.

But the most important statistical question is – does Gottman marriage counseling work? Gottman cites a bunch of different studies proving that it does, but he conducted or oversaw all of them. I randomly chose Shapiro and Gottman (2005) for further analysis. It is typical of this genre: it takes couples who have just had a baby (a particularly perilous time in a marriage), gives half of them a Gottman workshop, and the other half get nothing (no placebo or alternate method here!). The study evaluates husbands and wives separately (was this a preregistered decision?) on several established marital tests and subscales of marital tests (was the choice to keep some tests whole and take subscales of others preregistered?) and finds a significant quadratic effect of their program on marital quality.

I had to look up what a quadratic effect is. It means that the effect approximates a parabola. That is, things start by getting worse, then get better again, in a parabolic pattern. If I’m understanding the study correctly, then if you just do normal linear tests (did the couples’ marriage get better or worse, overall?) there was no significant effect. Gottman writes that:

> Our best guess as to why the quadratic pattern occurs is the following. With intervention, things get worse at first because the immediate effects of the interventions are to increase the amount of conflict that the couple experiences. Our interventions encourage couples to honestly face and discuss their conflicts, particularly potential or actual inequities in housework and childcare, conflicts that they would naturally avoid and which they do avoid in the control group. This early increase in the amount of conflict in our intervention group probably causes temporary discomfort, which is reflected in lowered marital quality and increased postpartum depression. However, because the couples in the intervention group learn the communicative skills to cope with these issues, the conflicts get dealt with to some degree, and thereafter marital quality and postpartum depression both improve, whereas in the control group, because these conflicts have not been dealt with, things get worse over time.

This is fine, but the abstract just sums it up as “results showed that, in general, the preventive intervention using a psycho-communicative-educational format was effective”. No! If you didn’t preregister that you expected a quadratic effect, you are lost in the Garden of Forking Paths and you should give up and start over. Also, I feel like probably Gottman advertises to his clients that their marital happiness will improve. But even if we accept his argument here, the only thing he can say with confidence is that their marital happiness will increasingly approximate a parabola over time, which is not really what I think most people go to therapists looking for.

I might be misinterpreting this, and maybe I'm being overly harsh. But I am predisposed to be overly harsh because the whole "predicting divorce" thing makes me think Gottman is out to get me, and so I am less forgiving of unusual polynomials than I might be otherwise. Also, if you're running the study of your own method, you ought to be on extra good behavior, and this does not really seem all that extra.

What happens when people who aren't Gottman evaluate the Gottman method? A large government-funded multicenter study testing a Gottman curriculum as well as several others found no effect of any on marital outcomes; control couples actually stayed together slightly more than ones who got marriage counseling. The Gottman curriculum seemed to do worst of the three curricula studied, although there were no statistical tests performed to prove it. I have no explanation for this. Maybe the parabola is just really big, and the divorce is the low point of the parabola, and later on they'll end up super-double-married. But I am not optimistic.

I don't want to be too harsh on Gottman here. Rigorous psychology studies are *murderous*. Things that we *know* basically *have to* work, like Alcoholics Anonymous and SSRIs and psychotherapy in general, end up showing no or minimal effects. Heck, zoom out a little bit and we have twin studies showing that parenting itself, in full generality, has no or minimal effect. I find all of this very suspicious, and it would not surprise me if there's something really wrong here that makes studies biased towards false negatives. All of this stuff about learning to respect and appreciate your spouse and negotiate conflicts in a calm loving way seems like the sort of thing that should work, and for all we know it might work in some population or situation other than the ones being studied.

But for the guy whose whole legend centers around how he’s evidence-based, it’s not a good look.

IV.

I really wanted Gottman Marital Counseling to work.

I wear a psychiatrist hat and a therapist hat. I love the psychiatrist hat. It is blue and pointy and has little glowing stars and moons on it. When I wear it, then with sufficient knowledge and understanding I can give people substances that release obsessions, calm fears, and brighten sorrows. Sometimes I can help people solve their unbearable hopeless problems, and it’s the best feeling in the world.

I hate wearing the therapist hat. I put it on as rarely as possible. I don’t advertise myself as a therapist, and if people ask me to therapy them, I try to refer them to someone else. But if someone wants to talk about their problems in a session, you can’t just say no. And so they tell me about being trapped in an abusive relationship, or haunted by guilt, or trapped in a dead end job with no prospects for improvement. And then they expect me to be able to say something that makes it all better. I know that the textbook response is something about how therapy does not solve problems per se, but by sharing them with someone else it makes them more bearable and adds perspective. Unfortunately, my patients didn’t read that textbook, and they put *hope* in me, and as often as not I betray it.

I think every therapist feels this way. I once talked to an important professor of therapy, who admitted to me something like “even at my stage, I feel like in the end we only have five or so techniques”, and I got really excited and blurted out “wait, what are the other two?!”

Desperation breeds gullibility. Patients with terminal diseases, however smart they used to be, turn to homeopaths and charlatans rather than face the dismal truth. Therapists are desperate – being confronted with some of the most sympathetic people in the world, day in and day out, having the burden of helping them placed on your shoulders, and knowing that your armamentarium isn't up to the task will do that to you. And so they become marks. Gullible, gullible marks. Realistically it's going to be really hard for me to stop recommending Gottman marital counseling to people, because they need something so much. And this is something. And it sounds so good. And I can't just let their marriage keep falling apart. And surely there's still a *chance* it might work, right?

(don't worry, eventually I'm going to look into some of those forms of counseling that outperformed it in that study)

Bad marriages are so, *so* bad. They're so bad it's shocking. The first time I saw one, I didn't know what to do. I wanted to figure out some of the root issues, so I asked my patient something about his wife, and he immediately launched into a tirade about all the things his wife had done recently, and why she was in the wrong and he was in the right. I tried to redirect him, and briefly succeeded, but after a second or two the new line of conversation shifted to how unfair his wife was and how she was in the wrong about everything. I wanted to grab him by the shoulders and shake him and say "for the love of God, just put this aside and move to the meta-level and let's talk about some of the places we can go from here!" but he was just incapable of this mental action. I wish I could say this was an isolated case. It isn't.

My ex-girlfriend Ozy writes a relationship advice column. Probably taking relationship advice from an ex-girlfriend is some kind of classic mistake, but I read it anyway. They describe five kinds of relationship problems – stupid problems, basic incompatibilities, problems that are actually a different kind of problem, terrible people, and horrifying soul-sucking messes. For some reason, this taxonomy has stuck with me when all the supposedly evidence-based taxonomies I hear the social workers talk about have failed. And the horrifying soul-sucking mess category sticks with me most of all:

> A problem of one of the previous three types was badly managed, perhaps for years. Now, every time you have a minor argument, you bring in everything wrong that happened for your entire relationship. You don’t feel like you can trust your partner. All the quirks you used to find charming drive you up the wall. You hate even your partner’s most innocuous actions. You avoid every topic that leads to a fight, and rapidly find that you can’t discuss anything except Marvel movies and the weather. You’re defensive whenever your partner says anything that sounds like even a minor criticism. You’re sarcastic and you call them names. Somehow, when you remember good things about the past– the time you saw Hamilton together or your birthday present or being the best man at their wedding– all you can remember is the long lines at intermission, the poor wrapping job, and their incredibly rude drunk aunt. If asked to name a good trait of theirs, you draw a blank, but you can go on for hours about their flaws.
>
> I guess it might be in theory possible to fix a horrifying soul-sucking mess with a lot of hard work, but to be honest every time I’ve seen a person in one of those relationships they were a lot better and happier and stronger as people as soon as they ended it.

A lot of my patients are horrifying soul-sucking messes. I wish there was something I could do about it, but instead I just sit and listen as they spend forty-five minutes describing every way their spouse has wronged them.

I’m terrified of this. How did it happen? At some point these people must have loved each other. How does any human relationship get this bad? Could it happen to me? Could I marry a great person who I love a lot, and then five years later sign up for therapy just so I can start talking about all my grievances without letting the therapist get a word in edgewise?

Unlike my ex, I don’t write a relationship advice blog. I write a blog about other things. One of them is politics. And whenever I hear people talk about relationships, I hear weird echoes of political problems. People who hate their spouse have an outgroup of one. A unified polity has devolved into partisanship. Social trust has been broken; a defect-defect equilibrium is in place. Gottman thinks of couples as a two-person culture, and some of those cultures are decadent and fractious.

Theodore Adorno’s right-wing authoritarianism scale asked a lot of questions about marriage and child-rearing; his thesis was that people who want top-down government will control their families the same way. Certainly there are authoritarian marriages. But it also seems like there are marriages that are nationalist in a more positive sense – one where the couple has built itself a strong mutual culture and identity, subsumed both individuals into it, and come out stronger on the other side. There are liberal marriages, where both spouses do their own thing, occasionally come together for mutually beneficial exchange of affection, and then go back to doing their own thing. There are even social justice marriages, where both partners are obsessed with how they are being oppressed by the other, interpret all discussion of compromise as hostile attempts to excuse the oppression, and have no strategy beyond proclaiming their victimization louder and louder in the hopes that their grievances are recognized.

…I'm making fun of that last one, but maybe they have a good point. Gottman's marital counseling – and every other kind of marital counseling I've read – is basically mistake theoretic. It assumes that two decent people who both want to live with each other are unable to because they don't have communication skills or problem-solving skills or some other skill that lets them fulfill their mutual aims. That model seems to accurately describe most of my patients. But Ozy's taxonomy is more thorough, including a category for terrible people:

> One or both of you just sucks. This category includes abuse, but it certainly isn't limited to it. This is the category from which advice columnists get all their pageviews: we love viewing a train wreck. Your partner has had a suicidal crisis every night at 3am for the last month, and you're up all night comforting them, and they refuse to find anyone else to talk to or ride out their suicidal crisis on their own. Your partner cheats on you constantly. Your baby is eighteen months old and your partner has never changed a single diaper. Your partner has demanded that you keep your relationship secret from everyone. You asked your partner to clean up dog poo from the floor, and it is three days later, and the dog poo is still there. If you have found yourself in a committed relationship with a terrible person, you should DUMP THEM.

In marriage, as in other forms of politics, sometimes exit > voice. Which is probably not something marriage counselors want to think about very much.

I’m a liberal, and my advice usually follows liberal principles. Have meetings involving all stakeholders, agree on general rules based on deep principles, and then follow the rules even when it’s hard, since that ensures a minimum of conflict on specific issues. I tell patients to have a consistent weekly meeting with their spouse, maybe a Saturday date night. During the meeting, each of them has to say one nice thing about the other, and then one problem that they want to work on. The other person thanks them for the nice thing and then they brainstorm a solution to the problem. If they both agree to the solution, they have to stick with it at least until the next relationship meeting. A lot of my patients have said this really helps them, have continued the meetings long after the immediate crisis passes, and probably expect the technique was invented by some clever person like John Gottman. I will never tell them that actually I picked it up from a different ex-girlfriend, not even the one who writes relationship advice columns.

In conclusion, I have no idea what makes marriages work, and I am not convinced John Gottman does either. Sometimes marriages are horrible in ways I never could have imagined, and other times they are amazing and I am infinitely jealous of them. Also, the very existence of a next generation of the human race is dependent on people having them and making them work for nonzero periods of time, which is a pretty terrifying prospect. Honestly I’m surprised we’ve lasted this long.

Socratic Grilling

Imagine an kid in school first hearing about germ theory. The conversation might go something like this:

> **Teacher:** Many diseases like the common cold are spread by germs, when one infected person contacts another.
>
> **Student:** But I got a cold a few weeks ago, and I never touch anyone except my family members. And none of them were sick.
>
> **Teacher:** You don’t need to actually touch someone. Sometimes it can spread through mucus droplets in the air.
>
> **Student:** And one time I was camping in the woods for a month, and then I got a cold, even though I hadn’t been around anybody.
>
> **Teacher:** If it was spring, you might have gotten allergies. Allergies can feel a lot like a cold, but they aren’t spread by germs.
>
> **Student:** It was fall.
>
> **Teacher:** Then maybe it was an unusual allergy, or some other condition.
>
> **Student:** Hey, wait. If germs are spread from person to person on touch, why doesn’t the government just mandate one week when nobody is allowed to touch anyone else? Then all the germs will die and we’ll never have to worry about germs again.

> **Teacher:** That’s a good question. A lot of germs have what’s called reservoirs in the environment, where they live when they’re not infecting humans. Even if the government tried your plan, probably most sicknesses would come back from their reservoirs.
>
> **Student:** I’ve never seen a germ reservoir. Where are they?
>
> **Teacher:** They’re not literal reservoirs like a water reservoir, that’s just what we call it when germs live in bats or raccoons or something.

Notice a few things about this conversation.

First, it’s really good that it happened. The student was clearly confused at many points. First, he had direct evidence that seemed to contradict the teacher’s claims that germs only spread by touch. Second, he had a sort of efficient-market-style confusion: germ theory seems to imply an easy way to eliminate all sicknesses forever, so why hasn’t someone picked this low-hanging fruit? Third, he was confused by an awkward term – he thought germ reservoirs were lakes full of germs in the hills somewhere.

All of these confusions are totally understandable. In fact, they’re a really good sign he’s paying attention – that he’s trying to figure out what germ theory really means and how it interacts with the rest of his worldview. They’re the direct opposite of guessing the teacher’s password. One day this kid is going to be an amazing scientist.

Second, to a hostile observer, it would sound like the student was challenging the teacher. Every time the teacher tried to explain germ theory, the student “pounced” on a supposed inconsistency. When the teacher tried to explain the inconsistency, the student challenged her explanations.

At times he almost seems to be mocking the teacher. Without contextual clues – and without an appreciation for how confused young kids can be sometimes – it could sound like this kid is an arrogant know-it-all who thinks he’s checkmated biologists and proven that germ theory can’t possibly be true. Or that he thinks that he, a mere schoolchild, can come up with a novel way to end all sickness forever that nobody else ever thought of.

And the thesis of this post is that you must *never, ever say that*. Saying that is *so bad*. Smack down that student once, say “I think I know more about germ theory than you do”, make him feel like he challenged your authority and that’s bad – and the best case scenario is he will never ask questions to resolve his confusion again. The worst case scenario is that he stops feeling the confusion entirely, or stops thinking of forcing things to fit together and make sense as a desirable goal to have.

One of the most important rationalist skills is “noticing your confusion”. But that depends on an even more important proto-skill of *wanting things to make sense*. If you lose that skill – if it stops bothering you and seeming like a problem when things don’t make sense to you – you will never notice your confusion and you will never become a good scientist or a good anything-else-that-requires-independent-thought. And interpreting an attempt to explore dissonance as a status grab that needs to be knocked down is absolutely fatal for that skill. Instead, you need to think of it as Socratic grilling – like Socratic questioning, but a little harsher and more confrontational in order to get to the point more quickly.

Tolerating this is harder than it sounds. Most people can stay helpful for one or two iterations. But most people are bad at explaining things, so one or two iterations isn’t always enough I’ve had times when I need five or ten question-answer rounds with a teacher in order to understand what they’re telling me. The process sounds a lot like “The thing you just said is obviously wrong”…”no, that explanation you gave doesn’t make sense, you’re still obviously wrong”…”you keep saying the same thing over and over again, and it keeps being obviously wrong”…”no, that’s irrelevant to the point that’s bothering me”…”no, that’s also irrelevant, you keep saying an obviously wrong thing”…”Oh! That word means something totally different from what I thought it meant, now your statement makes total sense.”

But it’s harder even than that. Sometimes there is a vast inferential distance between you and the place where your teacher’s model makes sense, and you need to go through a process as laborious as converting a religious person to a materialist worldview (or vice versa) before the gap gets closed. The process of learning to really appreciate communism, or libertarianism, or whatever, coming from a diametrically opposed philosophy, looks a lot like dozens of questions about “but isn’t that an atrocity?” “wouldn’t this inevitably lead to dystopia?” and hearing what your interlocutor has to answer. It’s so, *so* tempting to round this off to them trying to gotcha you (as indeed sometimes it will be) and assume they’re not really committed to trying to understand.

With a good teacher, you can add enough disclaimers to your statements that they won’t get mad at you during this process: “I know I am but a mere student, and nowhere near smart enough to actually challenge you, so I’m sure I’m just misunderstanding this, but the thing you just said seems really confusing to me, and I’m not saying it’s not true, but I can’t figure out how it possibly could be true, which is my fault and not yours, but could you please try to explain it differently?” With a *great* teacher, all of this is assumed, and you don’t need the disclaimers, and you can just say “What? That makes no sense,” and expect the teacher to try again.

I think actual teachers get this right more often than not. What I’m concerned about is the self-appointed Internet anti-autodidact brigade. I occasionally get these people as commenters on my blog, but they’re more common in Reddit and places like that. Someone goes “that’s weird” or “if X was really true, wouldn’t that imply Y?” and gets hit with “You really think you’re smarter than everyone else? You really think a random person on the Internet has discovered a hole in X?” No, sometimes they’re just using Socratic grilling to expose the contradictions in their model and get somebody to resolve them.

I find this to be one of the most frustrating parts of writing this blog: how do I signal the things I still need to learn without the Arrogance Police descending on me? This is different from the case with a specific teacher – here there’s just me projecting my confusions out into the void to see if anyone answers. And every so often I do flatter myself that I’ve discovered a flaw in my betters’ reasoning, so it’s less obvious when I mean to imply that and when I don’t. Next time this happens I’m going to try just linking this post and seeing how it goes.

I think I strike a good balance, at the cost of endless annoying disclaimers that 95% of you don’t need. From now on, maybe I’ll just link this post.

For, Then Against, High-Saturated-Fat Diets

I.

In the 1800s, the average US man weighed about 155 lbs. Today, he weighs about 195. The change is even starker at the extremes. Someone at the 90th percentile of weight back then weighed about 185 lbs; today, he would weigh 320 lbs. Back then, about 1% of men were obese. Today, about 25% are.

This puts a lot of modern dietary advice into perspective. For example, lots of people think low-carb is the solution to everything. But people in the 1800s ate almost 50% more bread than we do today, and still had almost no obesity. Other people think paleo is the solution to everything, but Americans in the 1800s ate a diet heavy in bread, milk, potatoes, and vegetables, and relatively low in red meat and other more caveman-recognizable foods. Intermittent fasting – again, cool idea, but your great-grandfather wasn't doing that, and he had a 1% obesity risk.

This isn't to say those diets can't work. Just that if they work, they're hacks. They treat the symptoms, not the underlying problem. Something went terribly wrong in US nutrition between 1900 and today, and all this talk about low-carb and intermittent fasting and so on are skew to that thing. Given that 1800s Americans seem to have effortlessly maintained near-zero obesity rates while eating foods a lot like the ones we eat today, maybe we should stop trying to figure out what cavemen were doing, and start trying to figure out what Great-Grandpa was doing, which sounds a lot easier.

We get similarly confusing evidence from other countries. Until recently, Chinese people ate mostly white rice. This is exactly the sort of high-glycemic-index carb that low-carbers say should be terrible for you. But the Chinese stayed thin even when they ate a lot. It was only when they started eating processed Western-style food that their obesity rate started to rise.

Or what about France? The French diet is about what you would expect; baguettes, pastries, cheese, meat. Lots of sugar, white flour, and fat – the opposite of all reasonable dietary advice. But 1970s France had the same kind of low obesity rates as 1800s America or China. This is related to the nutritional conundrum famously called the French paradox – why aren't the French fatter and sicker than they are?

The answer to all these questions seems to be something like "the body is pretty good at regulating its own weight under any diet except modern American processed food." But what aspect of processed food makes it bad?

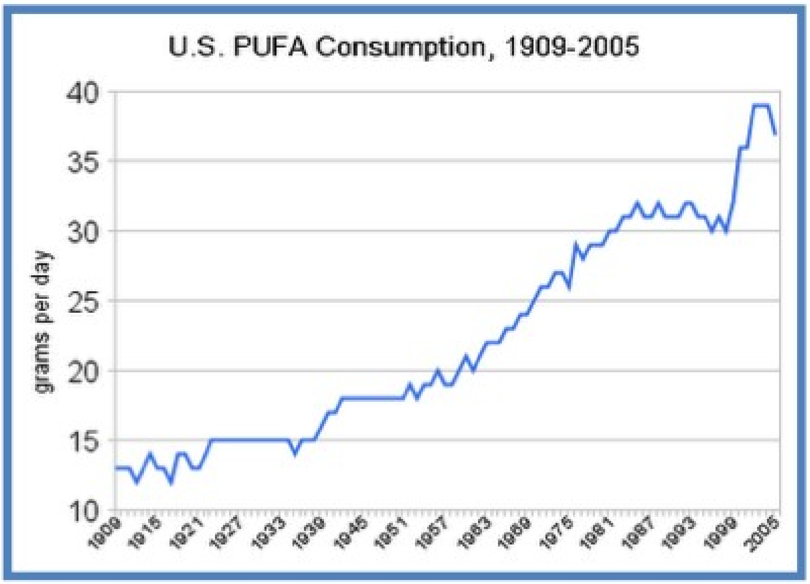
A new section of the online nutrition-sphere claims the answer has to do with the way mitochondria process fat. I've been trying to read these people and get a feel for their opinions. Most of what I've absorbed has come from Brad Marshall of Fire In A Bottle and his posts on The Croissant Diet. I've been told that another blog called Hyperlipid has a deeper investigation, but I've only scratched the surface of them. The r/SaturatedFat subreddit has some good stuff too. I don't claim to fully understand these people and I apologize for any misrepresentations I might be making. But the short version is: they all agree that everything went wrong when we switched from saturated to unsaturated fat.

Wait, isn't unsaturated fat the good kind of fat? Well, yes, this is what everyone else thinks. This is definitely one of those “good things are bad and bad things are good” diets. But let’s take a look at the argument.

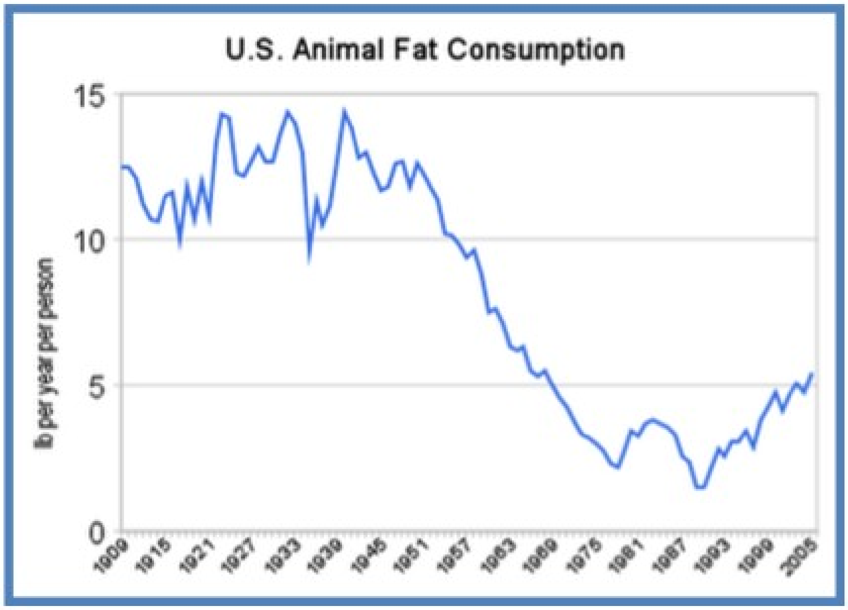
In the 1950s, heart disease rates were rising in the US. Realistically, this was mostly because lots of people had started smoking a few decades before, and now all that tobacco was catching up with them. But people didn’t know that at the time, so they did some studies into nutrition, and the studies suggested maybe saturated fat caused heart attacks. So the government told people to replace saturated fat with unsaturated fat, and this was back when the government was competent, so saturated fat consumption plummeted and unsaturated fat consumption shot up.

This paragraph is an extreme oversimplification: saturated fat is mostly found in things like milk, butter, and meat. Unsaturated fat comes in two types: monounsaturated and polyunsaturated. Monounsaturated fat is found in olive oil. Polyunsaturated fat comes in two types: omega-3 and omega-6. Omega-3 is mostly found in fish (“fish oil”). Omega-6 is mostly found in vegetables (“vegetable oil”). If you’ve ever looked at your food and seen ingredients like soybean oil, safflower oil, canola oil, corn oil, et cetera oil, these are omega-6 polyunsaturated fats.

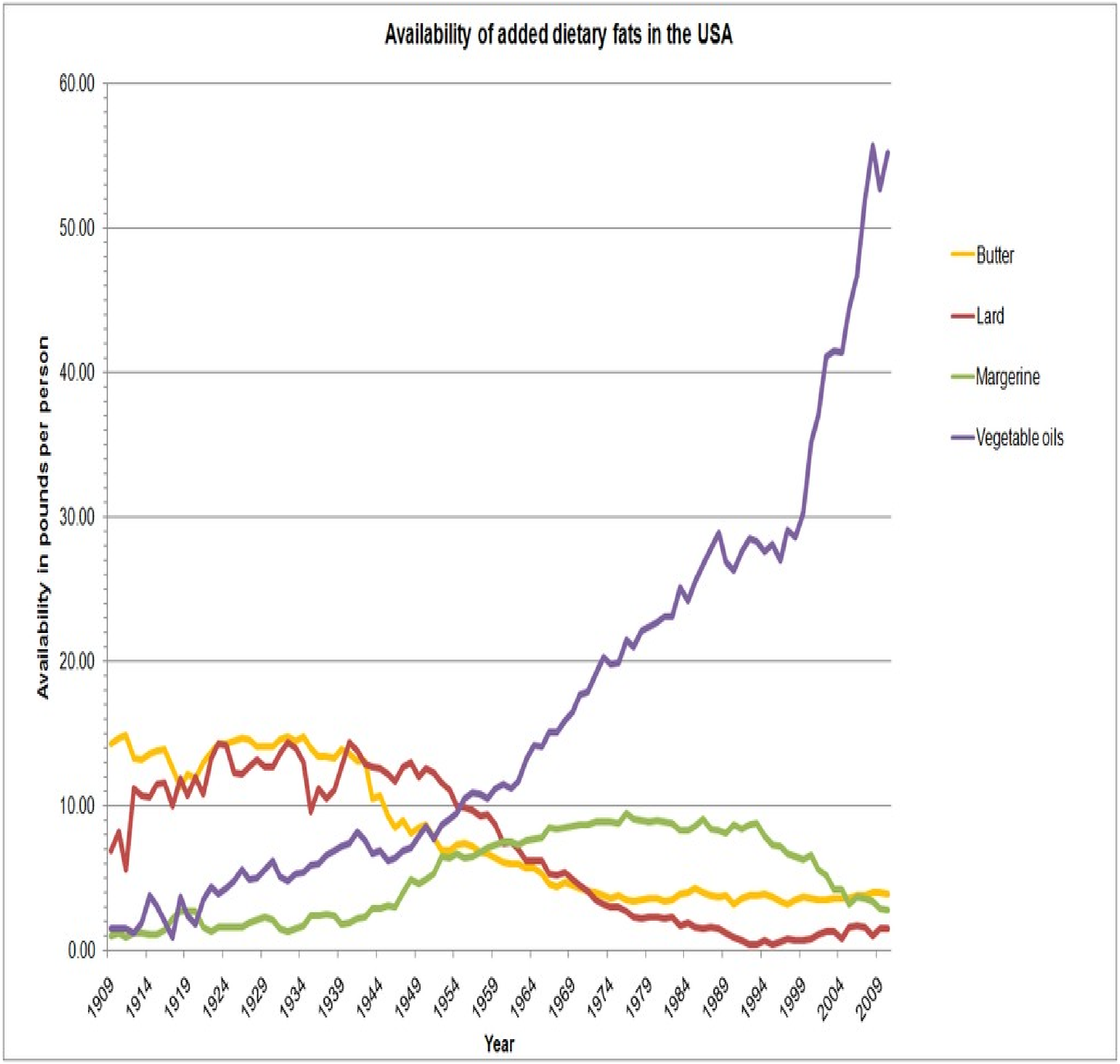
When the government recommended a switch from saturated fats to unsaturated fats around the 1950s, it was omega-6 polyunsaturated fats – vegetable oils – that picked up the slack. Here are some helpful charts:



Poly-unsaturated Fatty Acid Consumption

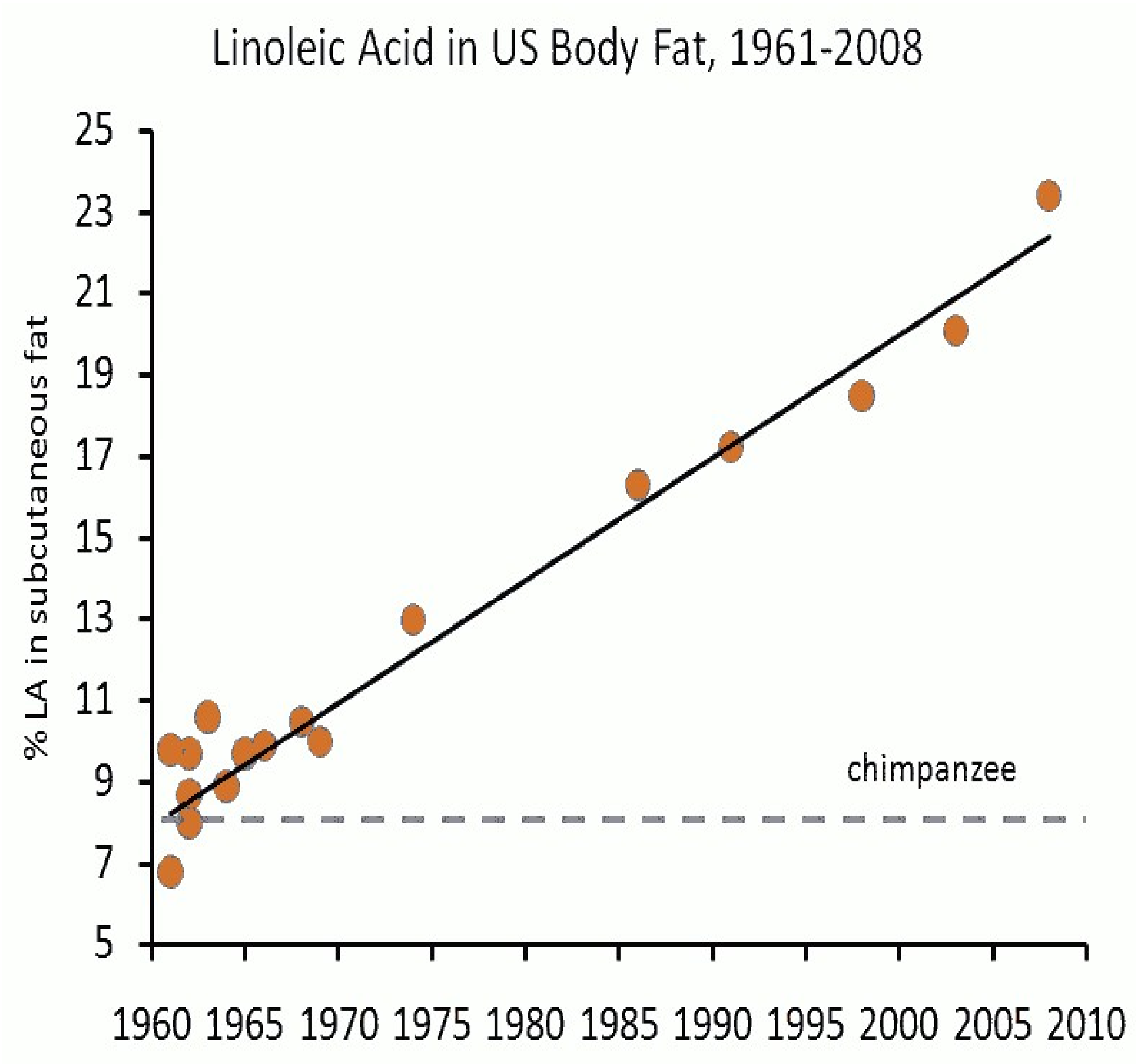


I cannot for the life of me find original sources for either of these graphs, but they seem consistent with everything else I've heard so I am going to trust them. Sorry!



[source]

This dramatic change in consumption of fat was reflected in a dramatic change in the composition of the human body. Studies of human fat cells and breast milk found that they went from being overwhelmingly saturated fat (like the fat cells and breast milk of animals) to being partly polyunsaturated fat:



[source, source, source, source]

The only common villain everyone agrees on in the obesity story is “processed food”. I’ve previously found this frustrating – it reeks of a sort of unreflective technophobia. What *part* of processing makes food bad? How does mere contact with a machine turn food from healthy to unhealthy? What food counts as “processed” or “not processed”? Is ground beef processed, since you grind it? Are scrambled eggs processed, since you scramble them? Is bread processed, since wheat doesn’t grow in loaves? Is water processed, since it goes through water processing facilities? Is the Eucharist processed, even though the processing only changes its metaphysical essence and not its physical properties? Everybody I ask acts like the answers to these questions are obvious, but everyone has different answers, and nobody can tell me their decision procedure.

Omega-6 polyunsaturated fats provide a tempting answer. Processing is bad at the point where it involves adding vegetable oil to stuff.

Here is a not-so-fun experiment you can try. Go to your local grocery store, check the ingredients of everything, and see what is the most ridiculous place you can find added vegetable oil – soybean oil will be the most common, though you can spot others. I thought I had reached a low when I found soybean oil listed in the ingredients of what was, to all appearances, just a bag of rice. But then I checked the bread section and found that about 90% of the loaves of bread had soybean oil added to them too (and some of the remainder had safflower oil). It really is the closest thing to a ubiquitous omni-ingredient in every processed food, and in some foods that you wouldn't have thought were processed at all.

So this is the circumstantial evidence linking polyunsaturated fat to obesity. Although polyunsaturated fat itself is natural (found in eg nuts and seeds), modern Americans consume it at levels that would have been equally foreign to cavemen and your great-grandfather. The vegetable oil craze started around the same time as the obesity epidemic, and the two have been following the same pattern ever since. And it's concentrated in the same processed foods that most people think are most responsible for obesity. Also, the body fat of obese people is more polyunsaturated than the body fat of healthy people.

But is there a biologically plausible reason why polyunsaturated fat would cause obesity?

Actually, there are several. The one I'd heard a few years ago blames an omega-6 to omega-3 imbalance. Before 1950, most people ate a reasonable amount of fish and a reasonable amount of nuts, and got about 4x as much omega-6 as omega-3. After 1950, people started eating lots of high-omega-6 vegetable oil, but only the same amount of high-omega-3 fish, and the ratio shot up: now it's about 10x-50x as much omega-6 as omega-3. Both omega-6 and omega-3 are involved in cell membranes and signaling chemicals, and there's some evidence that omega-6s may be pro-inflammatory and omega-3s anti-inflammatory. Although the studies haven't really been done, you can tell a story where the natural ratio of 6s to 3s creates a natural level of inflammation, and the current extreme level of 6s to 3s creates an extreme level of inflammation. Inflammation in the parts of the brain that regulate diet are one proposed mechanism for obesity, so there's the skeleton of an explanation here, although lots of work would need to be done to prove it.

But the new one, the one that Marshall and Hyperlipid are pushing, is a little different. They think unsaturated fats in general are bad, including monounsaturated fats and omega-3s (though realistically omega-6 vegetable oils so overwhelm these in the average American diet that we can forget everything else.). The exact mechanism is complicated, but focuses on the Krebs cycle, bane of medical students everywhere. The Krebs cycle is the set of chemical reactions that your cells use to convert high-energy food chemicals into ATP, a form of energy your body can use to power its own biological processes. Both saturated and unsaturated fats feed into the Krebs cycle. But the Krebs cycle produces reactive oxygen species (aka free radicals, eg hydrogen peroxide) when it metabolizes saturated fats, and not when it metabolizes unsaturated fats. Reactive oxygen species seem to be one of the signals the body uses to detect satiety, which makes sense – if they're a byproduct of metabolizing food, and you have a lot of them, that probably means you just metabolized a lot of food, and so you should be full. There's a lot of biochemistry here, and I haven't gone through all of it. But the basic idea is – burning saturated fat makes you full, but for decades we've been replacing saturated fat with polyunsaturated fat, which doesn't make you full. So we've been eating more. Hence, obesity.

If this were true, it suggests a diet high in saturated fat and low in unsaturated fat, especially polyunsaturated fat. Marshall calculated out the right macronutrient ratio and found that the food that most closely matched it was butter croissants – hence the name "the croissant diet". Yes, you can live off croissants if you want. But it also allows lots of other things with high saturated and low unsaturated fat. Red meat (but not chicken or pork; most chickens and pigs are fed high-PUFA feed that gives them high-PUFA meat). Milk, cheese, and butter (but not margarine, which is mostly PUFAs). Pasta, rice, and other carbs (but if you're putting sauces on them, make sure they're high saturated fat). Fried things, as long as you fry them in coconut oil, palm oil, or butter (mostly saturated fat) instead of vegetable oils or olive oils (mostly unsaturated fat). You can read the full specification here.

This diet is kind of the opposite of the one most nutritionists recommend. But it would taste a lot better. And following nutritionists' advice hasn't worked out so well for Americans circa 1970 through 2020. So what could go wrong?

II.

I find this to be a really elegant and provocative theory, with impressive circumstantial evidence. Unfortunately, as far as I can tell all of the direct evidence is against it. I'm not a nutritionist and have not studied these nearly as intently as the nutrition bloggers who support it, so low confidence in this part. But I'm putting my objections out there in the spirit of seeing whether other people will be able to reply to them and shoot them down.

First, let's start with the diet itself. Lots of people have tried it, and the most common result is nothing. The r/SaturatedFat subreddit is full of people talking about how the diet didn't work for them, with only a few contrary opinions. I tried it for about a month, and lost about 4 pounds the first two weeks, followed by no more weight loss no matter how hard I tried. This is my usual pattern every time I try a new diet, and I interpret it as placebo or just the effect of restricting some foods. The high-saturated-fat croissant diet did no better.

(Marshall argues it gets better results in waistline shrinkage than weight loss alone, and my own results sort of seem to confirm this, but I'm not confident in the accuracy of my waistline measurements. Also, why should this be true?)

Second, mouse studies. You can feed mice whatever you want, then see how much weight they gain. There are a lot of these, they're all conducted with different mice, different macronutrient sources, and different methodologies, and they all get different results. After looking at many of them, all I can say is that there is definitely no strong trend for lower saturated and higher polyunsaturated fat diets to result in more weight gain. For example, in this study, mice who ate palm oil (a high saturated-fat oil) gained more weight than those who ate safflower oil (a high polyunsaturated fat oil), although mice who ate cocoa butter (a different high saturated fat oil) gained less than either. In this study, mice who ate lard (high in saturated fat) gained more weight than those who ate olive oil (monounsaturated) or fish oil (polyunsaturated). Again, I'm less interested in these particular studies or their particular results than in the vast amount of literature that has investigated these questions and very rarely found a strong unambiguous tendency for saturated fat to be good and unsaturated fat to be bad. Stephan Guyenet, who used to support the omega-3:omega-6 ratio theory, agrees with this and now considers it unlikely based on mouse study results. It's possible that I'm missing different results for different kinds of saturated fat – stearic acid (a specific kind of saturated fat) seems to do pretty well, and the croissant diet to some degree centers around it. But as far as I can tell, an inferiority of any kinds of saturated fat to any kinds of polyunsaturated fat doesn't seem compatible with the basic theory.

And third, people have studied the effect of saturated vs. unsaturated fat *so much*. This is maybe the biggest controversy in nutrition right now. Some people think (in accordance with the 1950s and 1960s findings) that saturated fat contributes to cardiovascular disease. Other people think those findings were wrong and it doesn't. There have been a bunch of studies and big meta-analyses trying to find out who's right with only limited agreement. While some studies have found that saturated fat is bad and others that it's harmless, as far as I know none of them have found that it has a strong protective effect against weight gain. If this was really the difference between the 1800s when nobody was obese and today when a bunch of people are, it ought to be a blindingly bright signal. But I don't see anything of the sort.

Just to give examples: Lin investigated the relationship between saturated fat and weight loss, and found that the higher a diet was in saturated fat, the more likely people were to gain back weight they had lost. Phillips investigated the same question in relation to a probably irrelevant candidate gene, and found the same thing. Utzscheinder investigated high- vs. low- saturated fat diets and found no difference in weight loss, but the high saturated fat diet contributed to unhealthy liver fat deposition. Khaw et al investigated the effect of butter and coconut oil(saturated) vs. olive oil (monounsaturated) on obesity, and found nothing. Schwingshackl et al investigated the effect of 10 food groups on metabolic parameters and found that nuts were the healthiest, even though they are the highest in omega-6 fats.

(also, the whole point of this diet was supposed to be that croissants should be an unusually satiety-producing food, but somebody studied how much satiety every food produces, and croissants are literally the lowest on the list.)

I haven’t looked closely at any of these studies and don’t especially trust them. I’m more gesturing toward the general idea of how unlikely it is that people have studied saturated fat in depth for a long time, gotten a bunch of small negative effects that might or might not be real, and failed to notice that the real effect is gigantic and positive.

For that matter, where are the random Redditors? Saturated fat is one of the major macronutrients, it’s not exactly some weird exotic chemical nobody ever thought to test. If high-saturated-fat or low-polyunsaturated-fat diets help you stay thin as easily as 1800s Americans stayed thin, how come people didn’t figure this out in ten minutes? There have been so many random diet crazes like low-carb and paleo, all linked to some people squinting and thinking they might have seen a signal among all the noise. If there were a diet that was pretty basic and actually worked in an obvious way, don’t you think people would have found it?

III.

There’s an awkward tension between the first part of this post and that last counterargument.

The first part of the post said that there is some dramatic and hard-to-explain difference in obesity between the modern West and every other civilization, whether that’s the historical West or other modern countries that haven’t yet adopted our diet. I haven’t given evidence here, but the obesity goes hand in hand with higher cancer rates, higher cardiovascular disease rates, and just generally worse health. Presumably we’re doing something very wrong.

The last counterargument said that if we were doing something very wrong, one of the thousands of biohackers who has tried every ridiculous fad diet and long-shot idea would have reversed the one wrong thing we were doing and gotten incredible results. Then, by the efficient market hypothesis, somebody would have noticed the incredible results, and the smart paying-attention people would switch to that diet, and then we’d have a world that looks a lot different than the past decade or two of people chasing various exciting ideas with no results. As far as I know, nobody has yet met Ampersand’s challenge of finding a peer-reviewed study demonstrating that some diet can consistently help people lose lots of weight and keep it off.

But how could that be? If people used to be thin and healthy, we should just be able to do what they were doing! And then that would be the diet that can consistently lose weight and keep it off!

I can only see three ways out of this paradox.

First, conventional wisdom is right about everything. People are fatter today than in the 1800s because they eat too much and exercise too little. They eat too much because they are rich, food is cheap, and food tastes really good. They exercise too little because they're office workers now instead of farmers. In this model, the reason the efficient market hasn't found the secret to weight loss is because there's no secret and weight loss is really hard. It wasn't hard for your great-grandfather because he had fewer options and so he didn't need to exercise willpower to avoid the bad ones. The most sophisticated version of this model, so sophisticated that maybe I shouldn't call it this model at all, is the food reward theory ably defended by Stephan Guyenet.

This has a lot going for it, but can't be quite right. Exercise seems like a red herring; studies of how much people eat, exercise, and gain weight have shown that dietary changes explain more than 100% of weight gain over the past 30-40 years – probably we are exercising a little more. And there was really tasty food in 1800s America and 1970s France, so how come people didn't overindulge in that? How does it explain all the weird results like lab animals, pets, and feral rats gaining weight? This probably part of it, but it still feels like something's missing.

Second, diet is barely related to the obesity epidemic, and it's being caused by plastics or antibiotics affecting the microbiome or something like that. This is another thing where I would have expected people to notice, but I definitely don't want to dismiss it prematurely.

Third, it's a ratchet. Departing from the ways of our ancestors (or great-grandparents) can make you obese, but returning to their ways cannot make you thin again. A bad diet (whatever that is) shifts your weight set point up, but a good diet does not shift it back down, at least not in a reasonable amount of time. It just prevents further damage.

This contradicts the evidence from some people who do manage to lose weight, including some people who manage to lose a lot of weight and keep it off. I think the theory would have to be that different people’s set points are differently malleable, and that some people are obese because their set point is set to obese, and other people barely have a set point and are mostly operating on calories-in-calories-out. If this seems a little too convenient an assumption, keep in mind this is how lots of other set points work – some people will gain tolerance to certain drugs almost instantly, and other people will never get it at all.

Overall I am ending this research more confused than when I started it. I think the most likely dietary change I make is to try to avoid foods with soybean, corn, or safflower oil, since this is probably a good stand-in for “foods processed enough that they count as processed foods and you should avoid them”. I don’t think the evidence is good for avoiding fish oil and olive oil, and there’s enough evidence from elsewhere that these foods are healthy that I’m going to keep trying to eat them. I don’t think the evidence is good for saturated fats being especially good, and there seems to be at least equally strong evidence that they’re bad, so although I’m not going to work too hard to avoid them I’m definitely not going to optimize my diet for getting as many of them as possible.

Book Review: Hoover

You probably remember Herbert Hoover as the guy who bungled the Great Depression. Maybe you shouldn't. Maybe you should remember him as a bold explorer looking for silver in the jungles of Burma. Or as the heroic defender of Tientsin during the Boxer Rebellion. Or as a dashing pirate-philanthropist, gallivanting around the world, saving millions of lives wherever he went. Or as the temporary dictator of Europe. Or as a geologist, or a bank tycoon, or author of the premier 1900s textbook on metallurgy.

How did a backwards orphan son of a blacksmith, dropped in the middle of a forgotten spot in the Midwest, grow up to be a captain of industry and a US President? How did he become such a towering figure in the history of philanthropy that biographer Kenneth Whyte claims "the number of lives Hoover saved through his various humanitarian campaigns might exceed 100 million, a record of benevolence unlike anything in human history"? To find out, I picked up Whyte's *Hoover: An Extraordinary Life In Extraordinary Times*.

Herbert Hoover was born in 1874 to poor parents in the tiny Quaker farming community of West Branch, Iowa. His father was a blacksmith, his mother a schoolteacher. His childhood was strict. Magazines and novels were banned; acceptable reading material included the Bible and Prohibitionist pamphlets. His hobby was collecting oddly shaped sticks.

His father dies when he is 6, his mother when he is 10. The orphaned Hoover and his two siblings are shuttled from relative to relative. He spends one summer on the Osage Indian Reservation in Oklahoma, living with an uncle who worked for the Department of Indian Affairs. Another year passes on a pig farm with his Uncle Allen. In 1885, he is more permanently adopted by his Uncle John, a doctor and businessman helping found a Quaker colony in Oregon. Hoover’s various guardians are dutiful but distant; they never abuse or neglect him, but treat him more as an extra pair of hands around the house than as someone to be loved and cherished. Hoover reciprocates in kind, doing what is expected of him but excelling neither in school nor anywhere else.

In his early teens, Hoover gets his first job, as an office boy at a local real estate company. He loves it! He has spent his whole life doing chores for no pay, and working for pay is *so much better*! He has spent his whole life sullenly following orders, and now he’s expected to be proactive and figure things out for himself! Hoover the mediocre student and all-around unexceptional kid does a complete 180 and accepts Capitalism as the father he never had.

His first task is to write some newspaper ads for Oregon real estate. He writes brilliant ads, ads that draw people to Oregon from every corner of the country. But he learns some out-of-towners read his ads, come to town, stay at hotels, and are intercepted by competitors before they negotiate with his company. Of his own initiative, he rents several houses around town and turns them into boarding houses for out-of-towners coming to buy real estate, then doesn’t tell his competitors where they are. Then he marks up rent on the boarding houses and makes a tidy profit on the side. Everything he does is like this. When an especially acrimonious board meeting threatens to split the company, a quick-thinking Hoover sneaks out and turns off the gas to the building, plunging the meeting into darkness. Everyone else has to adjourn, the extra time gives cooler heads a change to prevail, and the company is saved. *Everything he does is like this*.

(on the other hand, he has zero friends and only one acquaintance his own age, who later describes him to biographers as “about as much excitement as a china egg”.)

Hoover meets all sorts of people passing through the Oregon frontier. One is a mining engineer. He regales young Herbert with his stories of traveling through the mountains, opening up new sources of minerals to feed the voracious appetite of Progress. This is the age of steamships, skyscrapers, and railroads, and to the young idealistic Hoover, engineering has an irresistible romance. He wants to leave home and go to college. But he worries a poor frontier boy like him would never fit in at Harvard or Yale. He gets a tip – a new, tuition-free university might be opening in Palo Alto, California. If he heads down right away, he might make it in time for the entrance exam. Hoover fails the entrance exam, but the new university is short on students and decides to take him anyway.

Herbert Hoover is the first student at Stanford. Not just a member of the first graduating class. Literally the first student. He arrives at the dorms two months early to get a head start on various money-making schemes, including distributing newspapers, delivering laundry, tending livestock, and helping other students register. He would later sell some of these businesses to other students and start more, operating a constant churn of enterprises throughout his college career. His academics remain mediocre, and he continues to have few friends – until he tries out for the football team in sophomore year. He has zero athletic talent and fails miserably, but the coach (whose eye for talent apparently transcends athletics) spots potential in Hoover and asks him to come on as team manager. In this role, Hoover is an unqualified success. He turns the team’s debt into a surplus, and starts the Big Game – a UC Berkeley vs. Stanford football match played on

Thanksgiving which remains a beloved Stanford football tradition.

Other Stanford students notice his competence, and by his senior year he is running not just the football team but the baseball team, a lecture series, a set of concerts and plays, and much of the student government. For the first time, he makes many social contacts, which is *sort of like* having friends, although real emotional connection remains beyond him. Whyte describes an occasion when Will Irwin, the football team's star player, suffers a career-ending injury:

> [Irwin] was outfitted with a plaster cast and deposited in his dorm room. Hoover visited him to approve spending on the athlete's medical supplies…Hoover carried his head to one side as he took in Irwin's cast and obvious discomfort…To make conversation and keep up his courage, Irwin tried to make light of his situation and watched as Hoover tried to laugh. A 'deep, rich, chuckle' originated far down in his chest, Irwin recalled, yet it was strangled 'before it came to the surface'. Hoover did not offer the patient a single word of consolation or reassurance during his time in the room. Irwin assumed hat Hoover's sympathies, for he did appear to be affected, were garroted and buried in the same internal graveyard as the chuckle. After a few minutes, Hoover headed for the door and, at the last instant, turned and blurted 'I'm sorry'. Irwin recognized that this minimal expression of emotion was as traumatic for Hoover as a broken ankle.

Hoover graduates Stanford in 1895 with a Geology degree. He plans to work for the US Geological Survey, but the Panic of 1895 devastates government finances and his job is cancelled. Hoover hikes up and down the Sierra Nevadas looking for work as a mining engineer. When none materializes, he takes a job an ordinary miner, hoping to work his way up from the bottom:

> He signed on as a mucker at the Reward Mine, shoveling wet dirt and rock into an ore car on ten-hour shifts for two dollars a day, seven days a week. The Cornishmen mocked him for his schooling and taught him the basics of their mole-like existence: how to breathe while the dust cleared from a blast; how to nap in a steel wheelbarrow heated from underneath by candles. The ceaseless grind of filling his car and pushing it up the slick rails of the Reward's dripping tunnels taxed Hoover's stamina. He was tortured in his sleep by muscle pain and neuralgia.

After a few months, he finds a position as a clerk at a top Bay Area mining firm. One year later, he is a senior mining engineer. He is moving up rapidly – but not rapidly enough for his purposes. An opportunity arises: London company Berwick Moreing is looking for someone to supervise their mines in the Australian Outback. Their only requirement is that he be at least 35 years old, experienced, and an engineer. Hoover (22 years old, <1 year experience, geology degree only) travels to Britain, strides into their office, and declares himself their man. The executives "professed astonishment at Americans' ability to maintain their youthful appearance" (Hoover had told them he was 36), but hire him and send him on an ocean liner to Australia.

22 year old Hoover trying his best to look like a respectable 36 year old capitalist

What does he think of his new home?

> In numerous letters over the next two years, Hoover would refer to Western Australia as hell, and he meant it. The landscape was hell, a flat, monotonous, dust-choked desert, barren but for low tangles of mulga and wattle bush as far as the eye could see.
>
> The climate was hell, a dry broil for the most part, one hundred degrees at midnight for days on end…
>
> The insects were hell, scorpions, tarantulas, snakelike centipedes, and disease carrying airborne pests with an unerring aim for one’s eyes and dinner plate…

> The settlements were hell, overnight ramshackled boomtowns with names like Kalgoorie and Coolgardie, box-shaped lodgings with walls of corrugated iron that roared in the wind, beds with unwashed sheets, meals of beans, biscuits, canned potatoes, and "tinned dog" (probably mutton or ham), entertainment consisting of out-of-date copies of American magazines, the odd horse race, and drunks dodging camels on Main Street.
>
> "You cannot appreciate the real damnation of this country," wrote Hoover.

Hoover soon manages to personally offend every single person in Australia:

> The harshness of the environment and Hoover's desire to prove himself drew an element of savagery from him. He fired rafts of employees for laziness and incompetence and dumped two of his own assistants for being "damn noodle heads"…uncompromising in pursuit of better margins, Hoover haggled with camel dealers to save a few dollars on freight costs He moved swiftly to shut losing properties…He lengthened shifts in the Coolgardie mines from 44 to 48 hours (his efforts to introduce labor-saving technology at another mine would result in a job action, which Hoover answered by firing the strikers and hiring more pliable Italian labor)…
>
> Hoover drove himself relentlessly as well, sleeping as little as four hours a night. His eyes and stomach gave him trouble. Months of roasting on the Western Australia grill left him with a chronic inflammation of the bladder. Sometimes he was so ill he could not sit up, but he refused to slow down, traveling on his back on a mattress on the bottom of a horse-drawn cart.

After a year, Hoover is the most hated person in Australia, and also doing *amazing*. His mines are producing more ore at lower prices than ever before. He receives promotion after promotion.

Success goes to his head and makes him paranoid. He starts plotting against his immediate boss, Berwick Moreing's Australia chief Ernest Williams. Thought Williams didn't originally bear him any ill will, all the plotting eventually gets to him, and he arranges for Hoover to be transferred to China. Hoover is on board with this, since China is a lucrative market and the transfer feels like a promotion. He travels first back to Stanford – where he marries his college sweetheart Lou Henry – and then the two of them head to China.

Herbert Hoover's college sweetheart

China is Australia 2.0. Hoover hates everyone in the country and they hate him back:

> Hoover shared the prevailing European conviction of Chinese racial inferiority. He would write of the ‘simply appalling and universal dishonesty of the working classes, the racial slowness, and the low average of intelligence’…Hoover was baffled at their lack of enthusiasm for mechanization and orderly administration. Lou reported that ‘the utter apathy of the Chinese to everything, their unconquerable dilatoriness’ was almost heartbreaking to her energetic husband.

The same conflicts are playing themselves out on the world stage, as Chinese resentment at their would-be-colonizers boils over into the Boxer Rebellion. A cult with a great name – “Society Of Righteous And Harmonious Fists” – takes over the government and encourages angry mobs to go around killing Westerners. Thousands of Europeans, including Herbert and Lou, barricade themselves in the partly-Europeanized city of Tientsin to make a final last stand. Hoover

> “…fought fires in the settlement and delivered food and medical supplies on his bicycle, hugging the brick walls along the street to avoid gunfire. Reporters on the scene observed that he seemed to be moving on the double quick, furiously jingling the change in his pockets and chewing nuts without shucking them. Lou, unwilling to join other women in the safety of the basement at city hall, ran bicycle errands of her own, a .38 Mauser strapped to her hip…

In between dodging artillery shells, Hoover furiously negotiates property deals with his fellow besiegees. He argues that if any of them survive, it will probably because Western powers invade China to save them. That means they will soon be operating under Western law, and people who had already sold their mines to Western companies would be ahead of the game and avoid involuntary confiscation. Somehow, everything comes up exactly how Hoover predicts. US Marines arrive in Tientsin to liberate the city (Hoover marches with them as their local guide) and he is ready to collect his winnings.

Problem: it turns out that “Whatever, sure, you can have my gold mine, we’re all going to die anyway” is not legally binding. Hoover, enraged as he watches apparently done deals slip through his fingers, reaches new levels of moral turpitude. He offers the Chinese great verbal deals, then gives them contracts with terrible deals, saying that this is some kind of quaint foreign custom and if they just sign the contract then the verbal deal will be the legally binding one (this is totally false). At one point, he literally holds up a property office with a gun to get the deed to a mine he wants. Somehow, after consecutively scamming half the population of Asia, he ends up with the rights to China’s most lucrative minefields. Berwick Moreing congratulates him and promotes him to managing director. He and Lou sail for London to live the lives of British corporate bigshots.

Predictably, Hoover makes an amazing corporate bigshot:

> Hoover had a ‘gift of juggling corporate assets in such a manner that insiders almost always benefitted’, whatever happened to the capital of the original shareholders. He was masterful at wielding write-offs and preference shares with multiple voting power on the grounds that new capital was required to avert bankruptcy. His favorite deals were those so complicated no one else could figure out they worked.
>
> On top of this, Hoover could keep mental maps of dozens of mines in his mind and, by one account, follow

> the progress of each shaft like a blindfolded chess master. He liked to receive telegrams from these properties and, without opening them, noting only the date and address, predict the level of the mine and the cost per ton of ore. He was usually correct.
>
> His intellectual capacities and powerful will made Hoover a fearsome negotiator. Arriving at the table with shirtsleeves rolled up, abrupt and aggressive, he had a singular talent for stripping away nonessential information and getting directly to the root of things, and he knew how to close. He possessed what one businessman said was a curious dynamic force that could compel the most reluctant person to put signatures to paper.

Also predictably, Hoover manages to offend everyone in Britain. Soon he is signing off on a ‘mutually agreeable’, ‘amicable’ dismissal from Berwick Moreing. They agree to let him go on the condition that he does not compete with them – a promise he breaks basically instantly. He goes into banking, and his “bank” funds mining operations in a way indistinguishable from being a mining conglomerate. Eventually he abandons even this fig leaf, and just mines directly.

But in other ways, his tens of millions of dollars are mellowing him out. Over his years in London, he develops hobbies besides making money and crushing people. He starts a family; he and Lou have two sons, Herbert Jr and Allen. He even hosts dinner parties, very gradually working on the skill of getting through an entire meal without mortally offending any guests:

> His fund of small talk was perpetually overdrawn, and if he interacted with the guests at his elbows, it was

> typically in a series of grunts or nods. If he wanted to make a point, he made it in a flat voice and then stopped abruptly, as one friend noted, someone had pulled his plug. If aroused, he would speak with force, sometimes veering into tactlessness, pursuing minor differences of opinion so harshly and indignantly that his victims nursed grudges for the rest of their natural lives. One acquaintance considered him the bluntest man in Europe, another ‘the rudest man in London’. He seldom took the time to enjoy his food, and was once clocked swallowing five courses in eleven minutes flat.

And he writes a book on metallurgy, which becomes the canonical text for a generation of engineering students. He can’t resist adding some of his own commentary. For example:

> Among the book’s idiosyncratic touches is Hoover’s attempt to end discussion of the capacity of different races of workers, a common debating point in early 20th century mining, by quantifying a racial productivity gap. He deemed one white worker equal to two or three of the colored races in simple tasks like shoveling, and as high as one to eleven in the most complicated mechanical work.

But also:

> To the engineer falls the work of creating from the dry bones of scientific fact the living body of industry. It is he whose intellect and direction bring to the world the comforts and necessities of daily need. Unlike the doctor, his is not the constant struggle to save the weak. Unlike the soldier, destruction is not his prime function. Unlike the lawyer, quarrels are not his daily bread. Engineering is the profession of creation and of

> construction, of stimulation of human effort and accomplishment.

Finally, having won respect in the financial, social, and intellectual worlds, he decides the natural next step is to become a public servant. Insofar as he has any political philosophy, he thinks of engineers as a sort of benevolent master race, destined to lead the world into an efficient technocracy. And he can think of no better standard-bearer than himself. He writes some Stanford friends, asking if they would support him for Governor of California. They suggest he start lower on the ladder, and offer him a position on the Stanford Board of Trustees, which he accepts (trustees are supposed to live in Palo Alto, but he lies and tells them he is moving back right away). He begins his public career by attacking tenure, "which he considered a protection racket for the weak and lazy and an outrage on the sanctity of higher education."

Okay, fine. He hadn't mellowed out *that* much. He manages to offend everyone in Stanford basically immediately, and that probably would have been the end of his career in politics. Luckily for him, World War I chooses that moment to break out, and little things like tenure are suddenly forgotten in the shadow of the greatest conflict the world has ever known.

II.

Count up the victims of World War I, and American tourists will be pretty far down the list. But victims they were. When the conflict broke out, thousands of Americans were overseas visiting the cathedrals of Florence or the museums of London. They woke up one morning to find the ships that were supposed to take them back had been conscripted into the war effort, or refused to sail for fear of enemy fire. The banks that were supposed to cash their travelers' checks were panicking, or devoting all their funds to the war effort, or dealing with a million other things. The hotels that were supposed to house them were closed indefinitely, their employees rushing to enlist out of patriotic fervor. And so thousands of frantic Americans, stuck in a foreign continent with no money and nowhere to stay, showed up at the door of the US Embassy in London and said – help!

The US Consulate in London didn't know how to solve these problems either. But Herbert Hoover, still high on his decision to pivot to philanthropy and public service, calls them up and asks if he can assist. They say yes, definitely. Hoover gets in touch with his rich friends, passes around the collection plate, and organizes a Committee For The Assistance Of American Travelers. Then he gets to work, the way only he can:

> Within 24 hours, Hoover's committee had its own stationery, and within forty-eight it was operating a booth in the ballroom of the Savoy Hotel as well as three other London locations. Through his business connections, Hoover managed to bypass restrictions on telegraph service and open a transatlantic line to allow Americans to wire money to stranded friends and relatives. In a city suddenly flooded with refugees, he reserved for American travelers some two thousand rooms in hotels or boardinghouses. He issued a press release proclaiming that his Residents' Committee was assuming charge of all American relief work in the city, and that in doing so it had the blessings of its honorary chairman, Walter Hines Page, the US ambassador to London.

…which is totally false. Hoover is starting to display a pattern that will stick with him his whole life – that of crushing competing charities. He begins a lobbying effort to get the US Embassy to ban all non-Hoover relief work, focusing on the inefficiency of having multiple groups working on the same problem. When the US Assistant Secretary Of War arrives in London to coordinate a response, he is met on the dock by Hoover employees, who demand he consult with Hoover before interfering in the US tourist issue. Eventually the Embassy, equally exasperated by Hoover’s pestering and impressed with his results, agrees to give him official control of the relief effort.

After two months of work, Hoover and his Committee have repatriated all 120,000 US tourists, supporting them in style until it could find them boat tickets. All of its loans and operating costs have been repaid by grateful tourists, and its budget is in the black. The rescued travelers are universal in their praise for Hoover, albeit partly because Hoover has threatened to ruin any of them who get too critical:

> Other complainants were received with less patience, including a hotheaded professor of history from the University of Michigan, who wrote to accuse the Residents’ Committee of mistreatment. Hoover refuted his charges indignantly and comprehensively, copying his response to the president of the university and its board of regents. After a meeting with his employer, the professor returned Hoover an abject retraction and apology.

Just as Hoover is preparing to rest on his laurels, he receives a cry for help. Germany has occupied and blockaded Belgium. The blockade prevents this tiny, heavily urban country from importing food, and the Belgians are starving. Germany needs its own food for its own armies, and is refusing to help. The Belgians order a thousand tons of grain from Britain, but when their representative comes to pick it up, Britain refuses to let them transport it, nervous at sending food into enemy-occupied territory. During tense negotiations, someone suggests using neutral power America as a go-between. But America is 5,000 miles away and busy with its own problems. So the US Ambassador to Britain asks his new best friend Herbert Hoover if he has any ideas.

Hoover invites Emile Francqui, a Belgian mining engineer he knows, to Britain. Together, they plan a Committee For The Relief of Belgium, intended not just to help transport the thousand tons of grain at issue, but to develop a long-term solution to the impending Belgian famine. Nothing like this has ever been tried before. Belgium has seven million people and almost no food. No government is offering to help, and they don’t have enough money to feed seven million people even for one day, let alone indefinitely. Hoover springs into action…

…by crushing all competing attempts to provide food for Belgium. He attacks the Rockefeller Foundation, which is trying to help, with a blitz of press coverage accusing it of various forms of insensitivity and interference, until it finally backs off. Then he gets to work on the government:

> The letter bore several Hoover watermarks, beginning with its heavy load of facts and figures organized in point form. It noted that myriad relief committees were springing up both inside and outside of Belgium, and urged consolidation. “It is impossible to handle the situation except with the strongest centralization and effective monopoly, and therefore the two organizations [Hoover outside Belgium and Francqui inside it] will refuse to recognize any element except themselves alone.” The letter also contained Hoover’s usual

> autocratic and slightly paranoid demands for “absolute command” of his part of the enterprise.

Control attained, Hoover springs into action actually feeding Belgium. He launches one of the largest public relations campaigns the world has ever seen, sending letters to newspapers around the world asking for donations. He “urged reporters to investigate the famine conditions in Belgium and play up the ‘detailed personal horror stuff’. He personally arranged for a motion picture crew to capture footage of food lines in Brussels, and he hired famous authors, including Thomas Hardy and George Bernard Shaw, to plead for public support of the rescue effort.” He constantly telegrams his exasperated wife and children, now safely back in Palo Alto, demanding they raise more and more money from the West Coast elite.

He browbeats shipping conglomerates until they agree to ship his food for free, then browbeats railroads until they agree to carry it. By telegraph and letter he coordinates banks, docks, trains, ships, and relief workers on both sides of the Atlantic. But that’s just the prelude. His real problem is the governments. Britain doesn’t want food shipped to Belgium, because right now the starving Belgians are Germany’s problem, and they don’t want to solve an enemy’s problem for them. But Germany also doesn’t want food shipped to Belgium, because the Belgians are resisting the occupation, and they figure starvation will make them more compliant. Shuttling back and forth across the North Sea, Hoover tries to get them to switch theories: *Germany* needs to think starving Belgians are their problem which it would be helpful to solve, and *Britain* needs to think starvation would make Belgians more compliant with the German occupation. In the end, both countries allow the shipments.

He goes on a fact-finding mission to Belgium, and manages to somehow offend everyone in the country that he is, at that very moment, saving from mass starvation:

> A third of Brussels’ population was receiving free food at more than a hundred canteens set up by the Comite Central and supplied by the CRB. Ration cards entitled the bearer to coffee, soup, and bread. On the cold, wet morning of December 1, Whitlock took Hoover to the street outside a theater that had been converted to a canteen in the Quarter des Marolles. They saw hundreds of Belgians shivering silently in the breadline…Whitlock kept his eyes on Hoover throughout the visit and saw him turn away and stare off down the street rather than share his feelings. Whitlock understood Hoover’s reaction as simple reticence. Others witnessing the same sort of behavior found it disturbing. They noticed how Hoover obsessed over the logistics of food distribution while avoiding interaction with recipients of relief and thought him a bloodless man. “He told of the work in Belgium as coldly as if he were giving statistics of production,” said US official. “From his words and his manner he seemed to regard human beings as so many numbers. Not once did he show the slightest feeling.”
>
> Hoover’s reticence was chronic. He was the sort of man who could sit for three hours on a train with his closest colleagues and not utter a single word, or bid farewell to his wife, not expecting to see her again for several months, in a curt telegraph: “Goodbye, Love, Bert”. It was often difficult to know if his behavior was due to bad manners, callousness, anxiety, or an effort to manage powerful emotions, because he was capable of all these things. Indeed, a few days after he averted his eyes from the breadline, he wrote, “It is difficult to state the position of the civil population of Belgium without

> becoming hysterical." The sight of ragged and hungry children especially bothered him, and he soon inaugurated a program of daily hot meals of bread and cocoa at Belgian schools.

By 1915, Hoover is, indeed, feeding millions of Belgians, indefinitely, using only private funding. He is also almost broke. Millions of Brits and Americans have given him contributions, from tycoons donating fortunes to ordinary people donating their wages, but it's not enough. His expenses pass $5 million a month, which would be about $100 million today; all these bills are starting to catch up to him. In an act of supreme sacrifice, Hoover pledges his entire personal fortune as collateral for the Committee's loans, then takes out more money. The grain shipments continue to flow, but his credit is at its end.

He continues beating on the doors of every government official he can find – British, German, American – demanding help. They all say their budgets are already occupied with the war effort. He begs them, lectures them, tells them that millions of people are doing to die. He goes all the way to the top, finagling an opportunity to meet with British Prime Minister David Lloyd George. Lloyd George later calls Hoover's presentation "the clearest he had [ever] heard on any subject", but he can offer only moral support.

What finally works is going to Germany and meeting with their top military brass. The brass are unimpressed; they still think that Belgium starving is as likely to help them as hinder. But the contact spooks top British officials, who agree to meet with Hoover again. Hoover feeds them carefully crafted lies, saying that the German brass have told him that British aid to Belgium would be a disaster to the Central Powers and so they, the Germans, are going to fund everything Hoover wants and more. "Oh no they don’t!” say the British, who promise to give Hoover even more funding than his imaginary German partners. The Committee for the Relief Of Belgium is finally back in the black. And what a black it is:

> The scope and powers of the Committee For Relief of Belgium were mindboggling. Its shipping fleet flew its own flag. Its members carried special documents that served as CRB passports. Hoover himself was granted a form of diplomatic immunity by all belligerents, with the British permitting him to cross the Channel at will and the Germans providing him a document saying ‘this man is not to be stopped anywhere under any circumstances’. Hoover had privileged access to generals, diplomats, and ministers. He enjoyed personal contacts with the heads of warring governments. He negotiated treaties with the belligerents, advised them on policy, and delivered private messages among them. Great Britain, France, and Belgium would soon be turning over to him $150 million a year, enough to run a small country, and taking nothing for it beyond his receipt. **As one British official observed, Hoover was running ‘a piratical state organized for benevolence.’**

In 1917, America enters World War I. Hoover is no longer neutral and so has to resign from the CRB. He returns to the US a war hero. The New York Times proclaims Hoover’s CRB work “the greatest American achievement of the last two years.” There is talk that he should run for President. Instead, he goes to Washington and tells President Woodrow Wilson he is at his service.

Wilson is working on the greatest mobilization in American history. He realizes one of the US’ most important roles will be breadbasket for the Allied Powers, and names Hoover “food commissioner”, in charge of ensuring that there is enough food to support the troops, the home front, and the other Allies. His powers are absurdly vast – he can do anything at all related to the nation’s food supply, from fixing prices to confiscating shipments to telling families what to eat. The press affectionately dubs him “Food Dictator” (I assume today they would use “Food Czar”, but this is 1917 and it is Too Soon).

Hoover displays the same manic energy he showed in Belgium. His public relations blitz telling families to save food is so successful that the word “Hooverize” enters the language, meaning to ration or consume efficiently. But it turns out none of this is necessary. Hoover improves food production and distribution efficiency so much that no rationing is needed, America has lots of food to export to Europe, and his rationing agency makes an eight-digit profit selling all the extra food it has.

By 1918, Europe is in ruins. The warring powers have declared an Armistice, but their people are starving, and winter is coming on fast. Also, Herbert Hoover has so much food that he has to swim through amber waves of grain to get to work every morning. Mountains of uneaten pork bellies are starting to blot out the sky. Maybe one of these problems can solve the other? President Wilson dispatches Hoover to Europe as “special representative for relief and economic rehabilitation”. Hoover rises to the challenge:

> Hoover accepted the assignment with the usual claim that he had no interest in the job, simultaneously seeking for himself the broadest possible mandate and absolute control. The broad mandate, he said, was essential, because **he could not hope to deliver food without refurnishing Europe’s broken finance,**

trade, communications, and transportation systems…

Hoover had a hundred ships filled with food bound for neutral and newly liberated parts of the Continent before the peace conferences were even underway. He formalized his power in January 1919 by drafting for Wilson a post facto executive order authorizing the creation of the American Relief Administration (ARA), with Hoover as its executive director, authorized to feed Europe by practically any means he deemed necessary. He addressed the order to himself and passed it to the president for his signature…

The actual delivery of relief was ingeniously improvised. Only Hoover, with his keep grasp of the mechanics of civilization, could have made the logistics of rehabilitating a war-ravaged continent look easy. He arranged to extend the tours of thousands of US army officers already on the scene and deployed them as ARA agents in 32 different countries. Finding Europe’s telegraph and telephone services a shambles, he used US Navy vessels and Army Signal Corps employees to devise the best-functioning and most secure wireless system on the continent. Needing transportation, Hoover took charge of ports and canals and rebuilt railroads in Central and Eastern Europe. The ARA was for a time the only agency that could reliably arrange shipping between nations…

The New York Times said it was only apparent in retrospect how much power Hoover wielded during the peace talks. “He has been the nearest approach Europe has had to a dictator since Napoleon.”

Once again, Hoover faces not only the inherent challenge of feeding millions, but opposition from the national governments he is trying to serve. Britain and France plan to let Germany starve, hoping this will decrease its bargaining power at Versailles. They ban Hoover from transporting any food to the defeated Central Powers. Hoover, "in a series of transactions so byzantine it was impossible for outsiders to see exactly what he was up to", causes some kind of absurd logistics chain that results in 42% of the food getting to Germany in untraceable ways.

He is less able to stop the European powers' controlled implosion at Versailles. He believes 100% in Woodrow Wilson's vision of a fair peace treaty with no reparations for Germany and a League Of Nations powerful enough to prevent any future wars. But Wilson and Hoover famously fail. Hoover predicts a second World War in five years (later he lowers his estimate to "thirty days"), but takes comfort in what he has been able to accomplish thus far.

He returns to the US as some sort of super-double-war-hero. He is credited with saving tens of millions of lives, keeping Europe from fraying apart, and preventing the spread of Communism. He is not just a saint but a magician, accomplishing feats of logistics that everyone believed impossible. John Maynard Keynes:

> Never was a nobler work of disinterested goodwill carried through with more tenacity and sincerity and skill, and with less thanks either asked or given. The ungrateful Governments of Europe owe much more to the statesmanship and insight of Mr. Hoover and his band of American workers than they have yet appreciated or will ever acknowledge. It was their efforts…often acting in the teeth of European obstruction, which not only saved an immense amount

of human suffering, but averted a widespread breakdown of the European system.

III.

Hoover wants to be president. It fits his self-image as a benevolent engineer-king destined to save the populace from the vagaries of politics. The people want Hoover to be president; he’s a super-double-war-hero during a time when most other leaders have embarrassed themselves. Even politicians are up for Hoover being president; Woodrow Wilson is incapacitated by stroke, leaving both Democrats and Republicans leaderless. The situation seems perfect.

Hoover bungles it. He plays hard-to-get by pretending he doesn’t want the Presidency, but potential supporters interpret this as him just literally not wanting the Presidency. He refuses to identify as either a Democrat or Republican, intending to make a gesture of above-the-fray non-partisanship, but this prevents either party from rallying around him. Also, he might be the worst public speaker in the history of politics.

Warren G. Harding, a nondescript Senator from Ohio, wins the Republican nomination and the Presidency. Hoover follows his usual strategy of playing hard-to-get by proclaiming he doesn’t want any Cabinet positions. This time it works, but not well: Harding offers him Secretary of Commerce, widely considered a powerless “dud” position. Hoover accepts.

Harding is famous for promising “return to normalcy”, in particular a winding down of the massive expansion of government that marked WWI and the Wilson Administration. Hoover had a better idea – use the newly-muscular government to centralize and rationalize America.

In his first few years in Commerce – hitherto a meaningless portfolio for people who wanted to say vaguely pro-prosperity things and then go off and play golf – Hoover instituted/invented housing standards, traffic safety standards, industrial standards, zoning standards, standardized electrical sockets, standardized screws, standardized bricks, standardized boards, and standardized hundreds of other things. He founded the FAA to standardize air traffic, and the FCC to standardize communications. In order to learn how his standards were affecting the economy, he founded the NBER to standardize government statistics.

But that isn't enough! He mediates an inter-state conflict over water rights to the Colorado River, even though that would normally be a Department of the Interior job. He solves railroad strikes, over the protests of the Department of Labor. "Much to the annoyance of the State Department, Hoover fielded his own foreign service". He proposes to transfer 16 agencies from other Cabinet departments to the Department of Commerce, and when other Secretaries shoot him down, he does all their jobs anyway. The press dub him "Secretary of Commerce and Undersecretary Of Everything Else".

Hoover's greatest political test comes when the market crashes in the Panic of 1921. The federal government has previously ignored these financial panics. Pre-Wilson, it was small and limited to its constitutional duties – plus nobody knew how to solve a financial panic anyway. Hoover jumps into action, calling a conference of top economists and moving forward large spending projects. More important, he is one of the first government officials to realize that financial panics have a psychological aspect, so he immediately puts out lots of press releases saying that economists agree everything is fine and the panic is definitely over. He takes the opportunity to write letters saying that Herbert Hoover has solved the financial panic and is a great guy, then sign President Harding's name to them. Whether or not Hoover deserves credit, the panic is short and mild, and his reputation grows.

While everyone else obsesses over his recession-busting, Hoover's own pet project is saving the Soviet Union. Several years of civil war, communism, and crop failure have produced mass famine. Most of the world refuses to help, angry that the USSR is refusing to pay Czarist Russia's debts and also pretty peeved over the whole Communism thing. Hoover finds $20 million to spend on food aid for Russia, over everyone else's objection:

> Russian relief would prove less popular than the Belgian variety, with the left accusing Hoover of seeking to undermine communism with capitalist aid…and the right charging him with rescuing and legitimating the shaky Soviet regime. Hoover gave the same answer to all critics: 'Twenty million people are starving. Whatever their politics, they shall be fed.'
>
> Maxim Gorky, in Italy nursing his tuberculosis, wrote Hoover personally: 'In the past year you have saved from death three and one-half million children, five and one-half million adults. In the history of practical humanitarianism I know of no accomplishment which in…magnitude and generosity can be compared to the relief you have actually accomplished.

So passed the early 1920s. Warren Harding died of a stroke, and was succeeded by Vice-President "Silent Cal" Coolidge. Coolidge won re-election easily in 1924. Hoover continued shepherding the economy (average incomes will rise 30% over his eight years in Commerce), but also works on promoting Hooverism, his political philosophy. It has grown from just “benevolent engineers oversee everything” to something kind of like a precursor of modern neoliberalism:

> Hoover’s plan amounted to a complete refit of America’s single gigantic plant, and a radical shift in Washington’s economic priorities. Newsmen were fascinated by is talk of a ‘third alternative’ between ‘the unrestrained capitalism of Adam Smith’ and the new strain of socialism rooting in Europe. Laissez-faire was finished, Hoover declared, pointing to antitrust laws and the growth of public utilities as evidence. Socialism, on the other hand, was a dead end, providing no stimulus to individual initiative, the engine of progress. The new Commerce Department was seeking what one reporter summarized as a balance between fairly intelligent business and intelligently fair government. If that were achieved, said Hoover, ‘we should have given a priceless gift to the twentieth century.’

Finally, it is 1928. Hoover feels like he has accomplished his goal of becoming the sort of knowledgeable political insider who can run for President successfully. Silent Cal decides not to run for a second term (in typical Coolidge style, he hands a piece of paper to a reporter saying “I do not choose to run for President in 1928” and then disappears and refuses to answer further questions). The Democrats nominate Al Smith, an Irish-Italian Catholic with a funny accent; it’s too early for the country to really be ready for this. Historians still debate whether Hoover and/or his campaign deserves blame for being racist or credit for being surprisingly non-racist-under-the-circumstances.

The main issue is Prohibition. Smith, true to his roots, is against. Hoover, true to his own roots (his mother was a temperance activist) is in favor. The country is starting to realize Prohibition isn’t going too well, but they’re not ready to abandon it entirely, and Hoover promises to close loopholes and fix it up. Advantage: Hoover.

The second issue is tariffs. Everyone wants some. Hoover promises that if he wins, he will call a special session of Congress to debate the tariff question. Advantage: Hoover.

The last issue is personality. Republican strategists decide the best way for their candidate to handle his respective strengths and weaknesses is not to campaign at all, or be anywhere near the public, or expose himself to the electorate in any way. Instead, they are “selling a conception. Hoover was the omnicompetent engineer, humanitarian, and public servant, the ‘most useful American citizen now alive.’ He was an almost supernatural figure, whose wisdom encompasses all branches, whose judgment was never at fault, who knew the answers to all questions.” Al Smith is supremely charismatic, but “boasted of never having read a book”. Advantage: unclear, but Hoover’s strategy does seem to work pretty well for him. He racks up most of the media endorsements. TIME Magazine offers a rare dissent, saying that “In a society of temperate, industrious, unspectacular beavers, such a beaver-man would make an ideal King-beaver. But humans are different.”

Apparently not *that* different. Hoover wins 444 votes to 87, one of the greatest electoral landslides in American history.

You may not like it, but this is what peak presidentialness looks like

Anne McCormick of the *New York Times* describes the inauguration:

> We were in a mood for magic…and the whole country was a vast, expectant gallery, its eyes focused on Washington. We had summoned a great engineer to solve our problems for us; now we sat back comfortable and confidently to watch our problems being solved. The modern technical mind was for the first time at the head of a government. Relieved and gratified, we turned over to that mind all of the complications and difficulties no other had been able to settle. Almost with the air of giving genius its chance, we waited for the performance to begin.

IV.

Herbert Hoover spent his entire presidency miserable.

First, he has no doubt that the economy is going to crash. It’s been too good for too long. He frantically tries to cool down the market, begs moneylenders to stop lending and bankers to stop banking. It doesn’t work, and the Federal Reserve is less concerned than he is. So he sits back and waits glumly for the other shoe to drop.

Second, he hates politics. Somehow he had thought that if he was the President, he would be above politics and everyone would have to listen to him. The exact opposite proves true. His special session of Congress comes up with the worst, most politically venal tariff bill imaginable. Each representative declares there should be low tariffs on everything except the products produced in his own district, then compromises by agreeing to high tariffs on everything with good lobbyists. The Senate declares that the House of Representatives is corrupt nincompoops and sends the bill back in disgust. Hoover has no idea how to solve this problem except to ask the House to do some kind of rational economically-correct calculation about optimal tariffs, which the House finds hilarious. “Opposed to the House bill and divided against itself, the Senate ran out the remaining seven weeks [of the special session] in a debauch of taunts, accusations, recriminations, and procedural argument.” The public blames Hoover, pretty fairly – a more experienced president would have known how to shepherd his party to a palatable compromise.

Also, there are crime waves, prison riots, bootlegging, and a heat wave during which Washington DC is basically uninhabitable. Also, at one point the White House is literally on fire.

…and then the market finally crashes. Hoover is among the first to call it a Depression instead of a Panic – he thinks the new term might make people panic less. But in fact, people aren’t panicking. They assume Hoover has everything in hand.

At first he does. He gathers the heads of Ford, Du Pont, Standard Oil, General Electric, General Motors, and Sears Roebuck and pressures them to say publicly they won’t fire people. He gathers the AFL and all the union heads and pressures them to say publicly they won’t strike. He enacts sweeping tax cuts, and the Fed enacts sweeping rate cuts. Everyone is bedazzled:

> The sweep and speed of Washington’s response to the crash, which gave the impression that Hoover had ‘thoroughly anticipated the debacle and mapped out the shortest road to recovery’, was hailed in the press as an entirely new approach to management of the nation’s economic affairs.” Herald-Tribune: “President Hoover’s prompt action to prevent the depression extending to business and industry saved the situation. The panic was checked in a few days. Wages were left unaffected, stabilization was insured; production was encouraged to continue as usual. This leadership was all the more notable, since it was practically the first of the sort ever to originate in the White House.

And:

> Economic joined journalists in congratulating Hoover on what was easily the most sophisticated response to a major economic event by any administration. ‘For the first time in our history,’ wrote Keynesian forerunners William Foster and Waddill Catchings, ‘a president is taking aggressive leadership in guiding private business through a crisis.

Six months later, employment is back to its usual levels, the stock market is approaching its 1929 level, and Democrats are fuming because they expect Hoover's popularity to make him unbeatable in the midterms. I got confused at this point in the book – did I accidentally get a biography from an alternate timeline with a shorter, milder Great Depression? No. I do think I accidentally got a biography by someone obsessed with defending Hoover at any cost and willing to stray into revisionist history to do it. As per Whyte, Hoover would take some brilliant and decisive action. Economists would praise him. The economy would start to look better. Everyone would declare the problem solved – especially Hoover, sensitive both to his own reputation and to the importance of keeping economic optimism high. Then for reasons totally outside the President's control, the recovery would stall, or reverse, or something else would go wrong.

People are still debating what made the Great Depression so long and hard. Whyte's theory, insofar as he has one at all, is "one thing after another". Every time the economy started to go up (thanks to Hoover), there was another shock. Most of them involved Europe – Germany threatening to default on its debts, Britain going off the gold standard. A few involved the US – the Federal Reserve made some really bad calls. The one thing Whyte is really sure about is that his idol Herbert Hoover was totally blameless.

He argues that Hoover's bank relief plan could have stopped the Depression in its tracks – but that Congressional Democrats intent on sabotaging Hoover forced the plan to publicize the names of the banks applying. The Democrats hoped to catch Hoover propping up his plutocrat friends – but the change actually had the effect of making banks scared to apply for funding and panicking the customers of banks that were known to have applied. He argues that the

“Hoover Holiday” – a plan to grant debt relief to Germany, taking some pressure off the clusterf**k that was Europe – was a masterstroke, but that France sabotaged it in the interests of bleeding a few more pennies from its arch-rival. International trade might have sparked a recovery – except that Congress finally passed the Hawley-Smoot Tariff, the end result of the corruption-plagued tariff negotiations, just in time to choke it off.

Whyte saves his barbs for the *real* villain: FDR. If the book is to be believed, Hoover actually had things pretty much under control by 1932. Employment was rising, the stock market was heading back up. FDR and his fellow Democrats worked to tear everything back down so he could win the election and take complete credit for the recovery. The wrecking campaign entered high gear after FDR won in 1932; he was terrified that the economy might get better before he took office, and used his President-Elect status to hint that he was going to do all sorts of awful things. The economy got skittish again and obediently declined, allowing him to get inaugurated at the precise lowest point and gain the credit for recovery he so ardently desired.

For example: November 1932. Hoover has just lost the election, but is a lame duck until March. The European debt crisis flashes up again. Hoover knows how to solve it. But:

> He had already met with congressional leaders and learned, as he had suspected, that they would not change their stance without Roosevelt’s support. Seized with the urgency of the moment, he continued to bombard his opponents with proposals for cooperation toward solutions, going so far as to suggest that Democratic nominees, not Republicans, be sent to Europe to engage in negotiations, all to no avail. Notwithstanding what editorialists called his “personal

> and moral responsibility" to engage with the outgoing administration, Roosevelt had instructed Democratic leaders in Congress not to let Hoover "tinker" with the debts. He had also let it be known that any solution to the problem would occur on his watch – "Roosevelt holds he and not Hoover will fix debt policy", read the headlines. Thus ended what the *New York Times* called Hoover's magnanimous proposal for "unity and constructive action", not to mention his 12-year effort to convince America of its obligation and self-interest in fostering European political and financial stability…
>
> During the debt discussions and to some extent as a result of them, the economy turned south again. Several other factors contributed. Investors were exchanging US dollars for gold as doubt spread about Roosevelt's intentions to remain on the gold standard. Gold stocks in the Federal Reserve thus declined, threatening the stability of the financial sector…what's more, the effectiveness of [Hoover's bank support plan], which had succeeded in stabilizing the banking system, was severely compromised by [Democrats'] insistence on publicizing its loans, as the administration had warned. For these reasons, Hoover would forever blame Roosevelt and the Democratic Congress for spoiling his hard-earned recovery, an argument that has only recently gained currency among economists.

And:

> Alarmed at these threats to recovery, Hoover pushed Democratic congressional leaders and the incoming administration for action. He wanted to cut federal spending, reorganize the executive branch to save money, reestablish the confidentiality of RFC loans, introduce bankruptcy legislation to protect foreclosures,

> grant new powers to the Federal Reserve, and pass new banking regulation, including measures to protect depositors…He was frustrated at every turn by Democratic leadership taking cues from the President-Elect…On February 5, Congress took the obstructionism a degree further by closing shop with 23 days left in its session.

In mid-February, there is another run on the banks, worse than all the other runs on the banks thus far. Hoover asks Congress to do something – Congress says they will only listen to President-Elect Roosevelt. Hoover writes a letter to Roosevelt begging him to give Congress permission to act, saying it is a national emergency and he has to act right now. Roosevelt refuses to respond to the letter for eleven days, by which time the banks have all failed.

Then, a month later, he stands up before the American people and says they have nothing to fear but fear itself – a line he stole from Hoover – and accepts their adulation as Destined Savior. He keeps this Destined Savior status throughout his administration. In 1939, Roosevelt *still* had everyone convinced that Hoover was totally discredited by his failure to solve the Great Depression in three years – whereas Roosevelt had failed to solve it for six but that was totally okay and he deserved credit for being a bold leader who tried really hard.

So how come Hoover bears so much of the blame in public consciousness? Whyte points to three factors.

First, Hoover just the bad luck of being in office when an international depression struck. Its beginning wasn’t his fault, its persistence wasn’t his fault, but it happened on his watch and he got blamed.

Second, in 1928 the Democratic National Committee took the unprecedented step of continuing to exist even after a presidential election. It dedicated itself to the sort of PR we now take for granted: critical responses to major speeches, coordinated messaging among Democratic politicians, working alongside friendly media to create a narrative. The Republicans had nothing like it; the RNC forgot to exist for the 1930 midterms, and Hoover was forced to personally coordinate Republican campaigns from his White House office. Although Hoover was good (some would say obsessed) at reacting to specific threats on his personal reputation, the idea of coordinating a media narrative felt too much like the kind of politics he felt was beneath him. So he didn’t try. When the Democrats launched a massive public blitz to get everyone to call homeless encampments “Hoovervilles”, he privately fumed but publicly held his tongue. FDR and the Democrats stayed relentlessly on message and the accusation stuck.

And third, Hoover was dead-set against welfare. However admirable his attempts to reverse the Depression, stabilize banking, etc, he drew the line at a national dole for the Depression’s victims. This was one of FDR’s chief accusations against him, and it was entirely correct. Hoover suspected that going down that route would lead pretty much where it led Roosevelt – to a dectupling of the size of government and the abandonment of the Constitutional vision of a small federal government presiding over substantially autonomous states. He decided it wasn’t worth it. So Herbert Hoover, history’s greatest philanthropist and ender-of-famines, would go down in history as the guy who refused to feed starving people. And they hated him for it.

V.

Some people might call Herbert Hoover a sore loser. But he argues that no, it’s totally reasonable for him to spend the rest of his life attacking FDR and trying to destroy his legacy.

His theory, explained in the countless books, pamphlets, and speeches that he spends his post-presidential life writing, is that FDR came from the same cloth as Hitler and Stalin. The miseries of the Great Depression, the centralizing tendencies of the age, the rise of mass media, and the collapse of republican virtue were combining all around the world in a monstrous reaction against the cause of liberty. “Daily,” wrote Hoover, “the world goes back to the regimentation of the Middle Ages, whether it be Bolshevism, Hitlerism, Fascism, or the New Deal.”

He has more! “[The New Deal] has no philosophy. It is sheer opportunism, a muddle of a spoils system, of reckless adventure, of unctuous claims to a monopoly of human sympathy, of greed for power, of a desire for popular acclaim and an aspiration to make the front pages of the newspapers.” He has more! “The New Deal has contributed to sapping our stamina and making us soft…the road to regeneration is burdensome and hard. It is straight and simple. It is a road paved with work and with sacrifice and consecration to the indefinable spirit that is America.”

He has more! He just keeps going like this, again and again. FDR, for his part, seems slightly befuddled. He tried offering Hoover a position coordinating the US effort to help war refugees – which Hoover turned down, assuming anything from FDR was a trick. Hoover just keeps shouting and fulminating and writing more and more books and pamphlets until FDR dies – which enrages Hoover, who wanted him to “live long enough to reap what he had sown”.

Whyte's theory is that this period of Hoover's life sowed the seeds for the modern conservative movement: "Modern American conservatism, conceived as an antidote to the New Deal, was born on December 16, 1937, with Hoover as its prophet and philosopher." He doesn't do much to back this theory up, and Hoover gets all of a paragraph in Wikipedia's long History of conservativism in the United States. We are left to piece it together from a few mentions here and there – Hoover befriending and helping a young William F. Buckley, Hoover giving a key endorsement to Barry Goldwater, and of course the namesake Hoover Institution that he founded, funded, and guided until his death.

I have to admit this is a hole in my understanding. Smart people definitely say that modern American conservativism began with Buckley and Goldwater and their friends, but what does this mean? Hasn't about half of America been conservative since the 1700s? Hasn't a philosophy of small government, individual freedom, and capitalist economics been pretty fundamental to America since its beginning? I'm not sure, and without this knowledge I don't feel qualified to judge Hoover's role.

Hoover passes in 1964, ninety years old. He lived long enough to become a hero to a new brand of conservative who considered him an intellectual forebear, and through various acts of public service to win back the love of his country. He had not quite finished his magnum opus, *Freedom Betrayed*. In 2012, historians finally dug it up, revised it, and released it to the world. It turned out to be 957 pages of him attacking Franklin Roosevelt. Give Herbert Hoover credit: he died as he lived.

VI.

I’m sorry this review was so long. I couldn’t bear to make it any shorter. I find the whole story so fascinating, and I just regret I couldn’t include more. I didn’t even get a chance to mention the time Hoover rescued the US South from the Great Mississippi Flood, or the time he discovered ancient ruins in the jungles of Burma, the time a 71-year-old Hoover was called back into service by President Truman to solve another post-World-War famine, or the time he invented the new sport of Hooverball (now part of the popular CrossFit exercise program).

Herbert Hoover on a famine relief tour of Poland, along with some of the children he is helping.

Hoover was a man who did everything wrong. He was the quintessential High Modernist. He was arrogant, he was authoritarian, he didn’t listen to anyone, he put no effort into pleasing people or making his ideas more palatable. He never solicited stakeholders’ opinions. He lied like a rug, constantly and egregiously. He lived his life like a caricature of exactly the sort of person who should fail at philanthropy and become a horror story to warn future generations.

But he won anyway. He started from a measly few million dollars and beat out Rockefellers and Carnegies to become the most successful philanthropist in early 20th century history. Whyte's estimate of 100 million lives saved seems much too high; there were only 100 million people in Europe *total* during the relevant period. But even during his own time, people universally credited him with saving millions. And he did it again and again and again. I didn't even have space to talk about the time he saved the Southern United States from a giant flood, or half a dozen other impressive accomplishments. Maybe the rules are wrong. Maybe all of this stuff about how authoritarian approaches never work, and you need to let the people you are helping lead the way, is all just modern prejudices, and putting a brilliant and very rich engineer in charge of a hypercentralized organization is just as good as any other way of doing things.

But even this I find less interesting than his psychology. He combined a personal callousness with a love for all humanity. When he was inspecting mines in Australia, he fired the worst-performing X% of workers. One worker begged him to reconsider – he had a family to support. Hoover raised $300 for the man's family – a lot of money at the time! Probably more than Hoover made in a month! – but fired him anyway. In 1932, when the Bonus Army marched on Washington, Hoover was adamant that he would not give these men – poor, starving veterans – a single cent more than they were entitled to by their existing benefits. But he also instructed his staff to go around to their encampments and give them food and supplies in secret.

Sometimes his stubbornness calls to mind the fictional Inspector Javert, who refuses to bend the law for any reason. In this model, Hoover sympathizes with everybody, but his honor forbids him to bend the rules in favor of underperforming employees or protesters who want more than their contracts entitle them to. But this picture of a hyper-honorable Hoover crashes into his constant willingness to lie, cheat, and bend the rules in his own favor. Sometimes his lies are for the greater good, like when he tells Britain that Germany is preparing to feed Belgium. Other times they seem entirely selfish, like his various Chinese mining scams. The best that can be said about Hoover is that *if* he decides a principle is involved, he sticks to it.

And this is actually really good! Again and again through the book, Hoover feels like the only person with a moral compass. When it is in everyone’s strategic interest to let Belgium starve, Hoover is the only one who is able to keep fixated on the potential human toll. When it is in everyone’s interest to let the USSR starve, only Hoover – despite his fanatical anti-communism – is able to stick to the frame where the Russians are human beings and politics is beside the point. When Americans are starving during the Great Depression…

…okay, Hoover totally dropped the ball on that one. In fact, one of his Democratic opponents wrote something about how maybe if unemployed American workers pretended to be Belgians, they could get Hoover’s sympathy. I don’t have a great explanation for this. But Hoover’s weak and inconsistent sympathies are often enough to let him outdo everyone else. Or at least, he is uncorrelated with everyone else and succeeds when they fail. Again and again Hoover is accused of treating people like numbers on a piece of paper. But if this is true, it seems to be linked to the reverse talent

– the ability to remember that numbers on a piece of paper represent people, even when other people would rather forget.

I’m equally confused about Hoover’s politics, although it’s not really his fault. The whole era confuses me. The Progressives, Hoover’s own faction, seem clearly related to modern progressives. But they also give me more of a technophile, rationalist feel than their modern counterparts. Am I imagining things? If not, where did this go?

And how did Hoover so deftly merge his centralizing technocratic engineer side with his small-government individual-freedom pro-capitalism side? Maybe it wasn’t that deft? Maybe he started his life as a centralizing technocrat, then made a 180 after becoming a small-government individualist helped him dunk on FDR more effectively? But it didn’t feel that way. It felt like all of it was coming from some central set of core beliefs throughout his life.

My confusion here feels similar to my confusion about Tyler Cowen’s “state capacity libertarianism”. Creating a strong and effective state is certainly…a goal you can have. But I don’t understand the argument for calling this a libertarian project. At best, it’s a project not entirely opposed to libertarianism. Still, perhaps this is my ignorance. Cowen thinks that strengthening the state and instituting effective technocratic government can be allied to a small-government individualistic market-based philosophy. Whatever he’s smoking, maybe Herbert Hoover was smoking the same thing.

I get the impression that Kenneth Whyte is a bit of a revisionist historian, too sympathetic to his subject to tell his story the way everyone else does. But at least in Whyte’s telling, the Hoover presidency was a great missed opportunity, or at least a fulcrum of history. If a few key economic events had been a few months off in one direction or the other, FDR might have been a footnote to history, and a four-term President Hoover might have left an indelible mark on America. Instead of a New Deal, we might have gotten a optimistic small-government technocratic meritocracy that was able to merge the best aspects of a dying frontier America with the best aspects of the industrial age.

In one of the most poignant passages in the book, Commerce Secretary Hoover fires back at his socialist critics. He points out that of the top dozen US officials – the President, VP, and ten Cabinet Secretaries – eight, including himself, had begun as manual laborers and worked their way up. That was the America Hoover was working to defend. He lost, and now we have this shitshow. But it’s hard to begrudge him the attempt.

Face Masks: Much More Than You Wanted To Know

There's been recent controversy about the use of face masks for protection against coronavirus. Mainstream sources, including the CDC and most of the media say masks are likely useless and not recommended. They've recently been challenged, for example by Professor Zeynep Tufekci in the *New York Times* and by Jim and Elizabeth on Less Wrong. There was also some debate in the comment section here last week, so I promised I'd look into it in more depth.

As far as I can tell, both sides agree on some points.

They agree that N95 respirators, when properly used by trained professionals, help prevent the wearer from getting infected.

They agree that surgical masks help prevent sick people from infecting others. Since many sick people don't know they are sick, in an ideal world with unlimited mask supplies everyone would wear surgical masks just to prevent themselves from spreading disease.

They also agree that there's currently a shortage of both surgical masks and respirators, so for altruistic reasons people should avoid hoarding them and give healthcare workers first dibs.

But they disagree on whether surgical masks alone help prevent the wearer from becoming infected, which will be the focus of the rest of this piece.

1. What are the theoretical reasons why surgical masks might or might not work?

Epidemiologists used to sort disease transmission into three categories: contact, droplet, and airborne. Contact means you only get a disease by touching a victim. This could be literally touching them, or a euphemism for very explicit contact like kissing or sex. Droplet means you get a disease when a victim expels disease-laden particles into your face, usually through coughing, sneezing, or talking. Airborne means you get a disease because it floats in the air and you breathe it in. Transmission via “fomites”, objects like doorknobs and tables that a victim has touched and left their germs on, is a bonus transmission route that can accompany any of these other methods.

More recently, scientists have realized that droplet and airborne transmission exist along more of a spectrum. Droplets can stay in the air for more or less time, and spread through more or less volume of space before settling on the ground. The term for this new droplet-airborne spectrum idea is “aerosol transmission”. Diseases with aerosol transmission may be spread primarily through droplets, but can get inhaled along with the air too. This concept is controversial, with different authorities having different opinions over which viruses can be aerosolized. It looks like most people now believe aerosol transmission is real and applicable to conditions like influenza, SARS, and coronavirus.

Surgical masks are loose pieces of fabric placed in front of the mouth and nose. They offer very good protection against outgoing droplets (eg if you sneeze, you won’t infect other people), and offer *some* protection against incoming droplets (eg if someone else sneezes, it doesn’t go straight into your nose). They’re not airtight, so they offer no protection against airborne disease or the airborne component of aerosol diseases.

Respirators are tight pieces of fabric that form a seal around your mouth and nose. They have various “ratings”; N95 is the most common, and I’ll be using “N95 respirator” and “respirator” interchangably through most of this post even though that’s not quite correct. When used correctly, they theoretically offer protection against incoming and outgoing droplet *and* airborne diseases; since aerosol diseases are a combination of these, they offer generalized protection against those too. Hospitals hate the new “aerosol transmission” idea, because it means they probably have to switch from easy/cheap/comfortable surgical masks to hard/expensive/uncomfortable respirators for a lot more diseases.

Theory alone tells us surgical masks should not provide complete protection. Coronavirus has aerosol transmission, so it is partly airborne. Since surgical masks cannot prevent inhalation of airborne particles, they shouldn’t offer 100% safety against coronavirus. But theory doesn’t tell us whether they might not offer 99% safety against coronavirus, and that would still be pretty good.

2. Are people who wear surgical masks less likely to get infected during epidemics?

It’s unethical to randomize people to wear vs. not-wear masks during a pandemic, so nobody has done this. Instead we have case-control studies. After the pandemic is over, scientists look at the health care workers who did vs. didn’t get infected, and see whether the infected people were less likely to wear masks. If so, that suggests maybe the masks helped.

This is an especially *bad* study design, for two reasons. First, it usually suffers recall bias – if someone wore a mask inconsistently, then they’re more likely to summarize this as “didn’t wear masks” if they got infected, and more likely to summarize it as “did wear masks” if they stayed safe. Second, probably some nurses are responsible and do everything right, and other nurses are irresponsible and do everything wrong, and that means that if anything at all helps (eg washing your hands), then it will look like masks working, since the nurses who washed their hands are more likely to have worn masks. Still, these studies are the best we can do.

Gralton & McLaws, 2010 reviews several studies of this type, mostly from the SARS epidemic of the early 2000s. A few are underpowered and find that neither surgical masks nor respirators prevent infection (probably not true). A few others show respirators prevent infection, but do not investigate surgical masks (probably right, but useless for our purposes). Two seem relevant to the question of whether surgical masks work:

Rapid awareness and transmission of SARS in Hanoi French Hospital, Vietnam was conducted in a poor hospital that only had surgical masks, not respirators. In the latter stages of the epidemic, 4 workers got sick and 26 stayed healthy. It found that 3 of the 4 sick workers hadn’t been wearing masks, but only 1 of the 26 healthy workers hadn’t. This is a pretty dramatic result – subject to the above confounders, of course.

Effectiveness of precautions against droplets and contact in prevention of nosocomial transmission of SARS is larger and more prestigious, and looked at a cluster of five hospitals. Staff in these hospitals used a variety of mask types, including jury-rigged paper masks that no serious authority expects to work, surgical masks, and N95 respirators. It found that 7% of paper-mask-wearers got infected, compared to 0% of surgical-mask and respirator wearers. This seems to suggest that surgical masks are pretty good.

The meta-analysis itself avoided drawing any conclusions at all, and would not even admit that N95 respirators worked. It just said that more research was needed. Still, the two studies at least give us a little bit of evidence in surgical masks' favor.

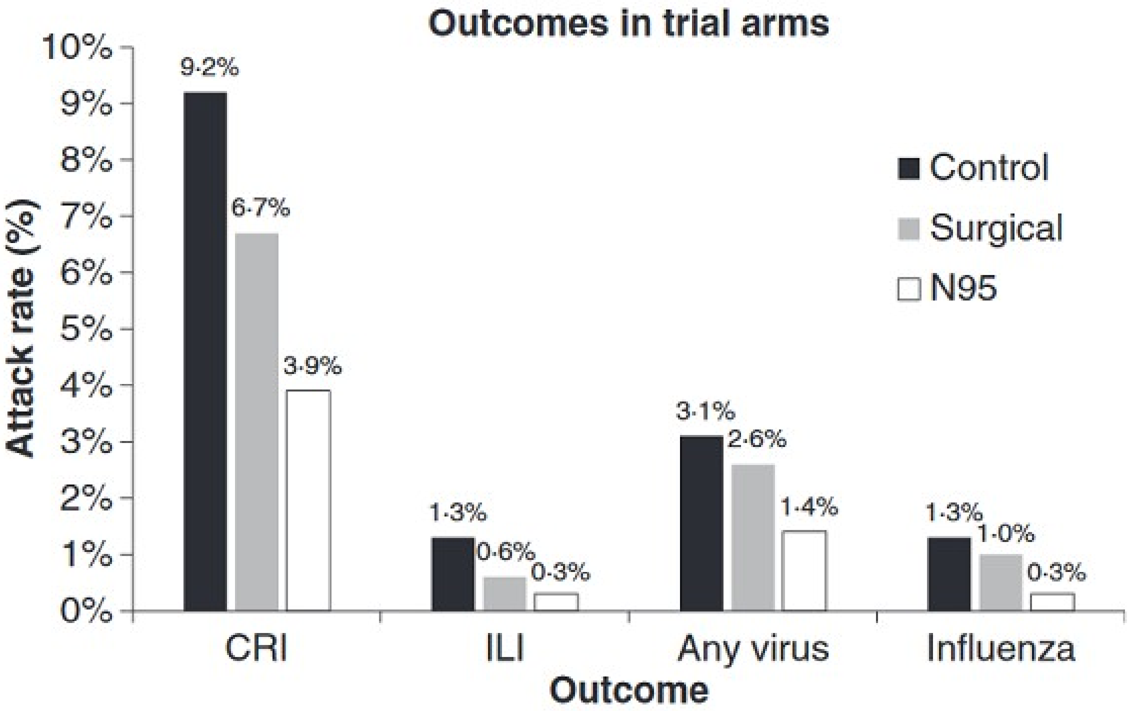
How concerned should we be that these studies looked at health care workers specifically? On the one hand, health care workers are ordinary humans, so what works for them should work for anyone else. On the other, health care workers may have more practice using these masks, or may face different kinds of situations than other people. Unlike respirators, surgical masks don't seem particularly hard to use, so I'm not sure health care workers' training really gives them an advantage here. Overall I think this provides some evidence that surgical masks are helpful.

I was able to find one study like this outside of the health care setting. Some people with swine flu travelled on a plane from New York to China, and many fellow passengers got infected. Some researchers looked at whether passengers who wore masks throughout the flight stayed healthier. The answer was very much yes. They were able to track down 9 people who got sick on the flight and 32 who didn't. 0% of the sick passengers wore masks, compared to 47% of the healthy passengers. Another way to look at that is that 0% of mask-wearers got sick, but 35% of non-wearers did. This was a significant difference, and of obvious applicability to the current question.

3. Do surgical masks underperform respirators in randomized trials?

Usually it would be unethical to randomize health care workers to no protection, so several studies randomize them to face masks vs. respirators. But a few others were done in foreign hospitals where lack of protection was the norm, and these studies did include a no-protection control group.

MacIntyre & Chugtai 2015, Facemasks For The Prevention Of Infection In Healthcare And Community Settings, reviews four of these. Two of the four are unable to find any benefit of either masks or respirators. The third finds a benefit of respirators, but only if nobody tested the respirators to see if they fit, which doesn’t make sense and suggests it’s probably an artifact. The fourth finds a benefit of respirators, but not masks. It seems unlikely that respirators don’t help, so this suggests all these studies were underpowered. If we throw good statistical practice to the winds and just look at the trends, they look like this:



In other words, respirators are better than masks are better than nothing. It would be wrong to genuinely conclude this, because it’s not statistically significant. But it would also be wrong to conclude the studies show masks don’t work, because they mostly show respirators don’t work, and we (hopefully) know they do.

Overall these studies don’t seem very helpful and I’m reluctant to conclude anything from them. In section 6, I’ll talk more about why studies may not have shown any advantage for respirators.

4. Do surgical masks prevent ordinary people from getting infected outside the healthcare setting?

The same review lists nine randomized trials with a different design: when the doctor diagnoses you with flu, she either asks everyone in your family to wear masks (experimental group), or doesn’t do that (control group), and then checks how many family members in each group got the flu.

How did these go? That depends whether you use intention-to-treat or per-protocol analysis. Intention-to-treat means that you just compare number of infections in the assigned-to-wear-masks group vs. the control group. Per-protocol means that you only count someone in the study if they actually followed directions. So if someone in the assigned-to-wear mask group didn’t wear their mask, you remove them from the study; if someone in the control group went rogue and *did* wear a mask, you remove them too.

Both of these methods have their pros and cons. Per protocol is good because if you’re trying to determine the effect of wearing a mask, you would really prefer to only be looking at subjects who actually wore a mask. But it has a problem: adherence to protocol is nonrandom. The people who follow your instructions diligently are selected for being diligent people. Maybe they also diligently wash their hands, and diligently practice social distancing. So once you go per protocol, you’re no longer a perfect randomized controlled trial. Only intention-to-treat analyses carry the full weight of a gold standard RCT.

According to intention-to-treat, the studies unanimously found masks to be useless. But there were a lot of signs that intention-to-treat wasn't the right choice here. Only about a fifth of people who were asked to wear masks did so with any level of consistency. The rest wore the mask for a few hours and then get bored and took it off. Honestly, it's hard to blame them; these studies asked a lot from families. If a husband has flu, and sleeps in the same bed as his wife, are they both wearing masks all night?

Of the three studies that added per-protocol analyses, all three found masks to be useful (1, 2, 3) . Does this prove masks work? Not 100%; per-protocol analyses are inherently confounded. But it sure is suggestive.

The review author summarizes:

> The routine use of facemasks is not recommended by WHO, the CDC, or the ECDC in the community setting. However, **the use of facemasks is recommended in crowded settings (such as public transport) and for those at high risk (older people, pregnant women, and those with a medical condition) during an outbreak or pandemic**. A modelling study suggests that the use of face-masks in the community may help delay and contain a pandemic, although efficacy estimates were not based on RCT data. Community masks were protective during the SARS outbreaks, and about 76% of the population used a facemask in Hong Kong.
>
> **There is evidence that masks have efficacy in the community setting**, subject to compliance [13] and early use [12, 18, 19]. It has been shown that

> compliance in the household setting decreases with each day of mask use, however, which makes long term use over weeks or months a challenge […]
>
> **Community RCTs suggest that facemasks provide protection against infection in various community settings**, subject to compliance and early use. For health-care workers, the evidence suggests that respirators offer superior protection to facemasks.

Parts of this summary are infuriating. If the big organizations recommend that especially vulnerable groups wear masks, aren't they admitting masks work? But if they're admitting masks work, why don't they recommend them for ordinary people?

It looks like they're saying masks work a little, they're too annoying for it to be worth it for normal people, but they might be worth it for the especially vulnerable. But then why don't they just say masks work, and let each person decide how much annoyance is worthwhile? I'm not sure. But it looks like the author basically ends up in favor of community use of surgical masks in a pandemic, mostly on the basis of per-protocol analyses of community RCTs.

5. How do surgical masks and respirators compare in hokey lab studies?

Our source here is Smith et al 2016, Effectiveness Of N95 Respirators Versus Surgical Masks In Protecting Health Care Workers From Acute Respiratory Infection: A Systematic Review And Meta-Analysis. They review some of the same studies we looked at earlier, but then investigate 23 "surrogate exposure studies", ie throwing virus-shaped particles at different masks in a lab and seeing if they got through. You can find the results of each in their appendix. Typically, about 1 – 5% of particles make it through the respirator, and 10 – 50% make it through the surgical mask. They summarize this as:

> In general, compared with surgical masks, N95 respirators showed less filter penetration, less face-seal leakage and less total inward leakage under the laboratory experimental conditions described.

I think in general the fewer virus particles get through your mask, the better, so I think this endorses surgical masks as better than nothing, since their failure rate was less than 100%.

Booth et al, 2013 examines surgical masks themselves more closely. They hook a surgical mask up to “a breathing simulator” and then squirt real influenza virus at it, finding that:

> Live influenza virus was measurable from the air behind all surgical masks tested. The data indicate that a surgical mask will reduce exposure to aerosolised infectious influenza virus; reductions ranged from 1.1- to 55-fold (average 6-fold), depending on the design of the mask…the results demonstrated limitations of surgical masks in this context, although they are to some extent protective.

The paper doesn’t discuss how particle number maps to infection risk. Does letting a single influenza virus through mean you will get infected? If so, any reduction short of 100% is useless. I have a vague sense that this isn’t true; your immune system can fight off most viruses, and the fewer you get, the better the chance it will win. Also, even respirators don’t claim to reduce particle load by more than 99% or so, and those work, so it can’t be that literally a single virus will get you. Overall I think modest reductions in particle number are still pretty good, but I don’t have a study that proves it.

6. Is it true that the public won’t be able to use N95 respirators correctly?

Yes.

I remember my respirator training, the last time I worked in a hospital. They gave the standard two minute explanation, made you put the respirator on, and then made you go underneath a hood where they squirted some aerosolized sugar solution. If you could smell the sugar, your respirator was leaky and you failed. I tried *so hard* and I failed *so many times*. It was embarrassing and I hated it.

I’m naturally clumsy and always bad at that kind of thing. Some people were able to listen to the two minute explanation and then pass right away. Those kinds of people could probably also listen to a two minute YouTube explanation and be fine. So I don’t want to claim it’s impossible or requires lots of specialized background knowledge. It’s just a slightly difficult physical skill you have to get right.

Bunyan et al, 2013, Respiratory And Facial Protection: A Critical Review Of Recent Literature, discusses this in more depth. They review some of the same studies we reviewed earlier, showing no benefit of N95 respirators over surgical masks for health care workers in most situations. This doesn’t make much theoretical sense – the respirators should win hands down.

The most likely explanation is: doctors aren’t much better at using respirators than anyone else. In a California study of tuberculosis precautions, 65% of health care workers used their respirators incorrectly. That’s little better than the general public, who have a 76% failure rate. Bunyan et al note:

> The fitting of N95 respirators has been the subject of many publications. The effective functioning of N95 respirators requires a seal between the mask and the face of the wearer. Variation in face size and shape and different respirator designs mean that a proper fit is only possible in a minority of health care workers for any particular mask. Winter et al. reported that, for any one of three widely used respirators, a satisfactory fit could be achieved by fewer than half of the healthcare workers tested, and for 28% of the participants none of the masks gave a satisfactory fit.
>
> **Fit-testing is a laborious task, taking around 30 min to do properly**, and comprises qualitative fit-testing (testing whether the respirator-wearing healthcare worker can taste an intensely bitter or sweet substance sprayed into the ambient air around the outside of the mask) or quantitative fit testing (measuring the ratio of particles in the air inside and outside the breathing zone when wearing the respirator). Attempts have been made to circumvent the requirement for fit testing, and it has been suggested that self-testing for a seal by the respirator wearer (see http://youtu.be/pGXiUyAoEd8a for a video demonstration) is a sufficient substitute for fit-testing. However, **self-checking for a seal has been demonstrated to be a highly unreliable technique** in two separate studies so that full fit-testing remains a necessary preliminary requirement before respirators can be used in the healthcare setting.
>
> Operationally, this presents significant challenges to organizations with many healthcare workers who require

> fit-testing. Chakladar et al. pointed out that, in addition to the routine need for repeat testing over time to ensure that changes in weight or facial hair have not compromised a good fit, movements of healthcare workers between organizations using different makes of respirators would necessitate additional repeat fit-testing. Fit-testing is likely to remain problematic to health-care organizations for the foreseeable future. In addition to the requirement for fit-testing, 'fit-checking' is also required each time the respirator is donned to ensure there are no air leaks.

Is a poorly-fitting N95 respirator better than nothing? The reviewed studies suggest that at that point it's just a very fancy and expensive surgical mask.

7. Were the CDC recommendations intentionally deceptive?

No, and I owe them an apology here.

I think the evidence above suggests masks can be helpful. Masked health care workers were less likely to catch disease than unmasked ones. Masked travelers on planes were less likely to catch disease than unmasked ones. In per protocol analysis, masked family members are less likely to catch disease from an index patient than unmasked ones. Laboratory studies confirm that masks block most particles. All of this accords with a common-sense understanding of droplet and aerosol transmission of disease.

None of these, except maybe the plane study, tell us exactly what we want to know. The SARS studies were all done in a health care setting, so they don't prove that regular people can benefit from masks. But health care workers are closely related to *homo sapiens* and ought to have similar anatomy and physiology. Surgical masks aren't as complicated as respirators and we can assume most people get them right. And although health care workers are in unusually high-risk situations, that should just affect the magnitude of the benefit, not the sign; obviously the level of risk ordinary people encounter is sometimes relevant, considering they do often catch pandemic diseases. So our default assumption should be that these studies carry over, not that they don't.

Likewise, most of the community studies were done on family members. Most guidelines already say to mask up if you have a sick family member, so talking about subways and crowds requires a little bit of extrapolation. But again, being in a family is just one form of close contact. It would take bizarre convolutions to even imagine a theory where you can catch diseases from your family members but not from people you sit next to on a train. Our default assumption should be in favor of these results generalizing, not against them.

But the CDC has recommended against mask use. I hypothesized that the CDC was intentionally lying to us, trying to trick us into not buying masks so there would be enough for health care workers.

But that can't be true, because the CDC and other experts came up with their no-masks policy years ago, long before there was any supply shortage. For example, during the 2009 swine flu pandemic, their website offered the following table:

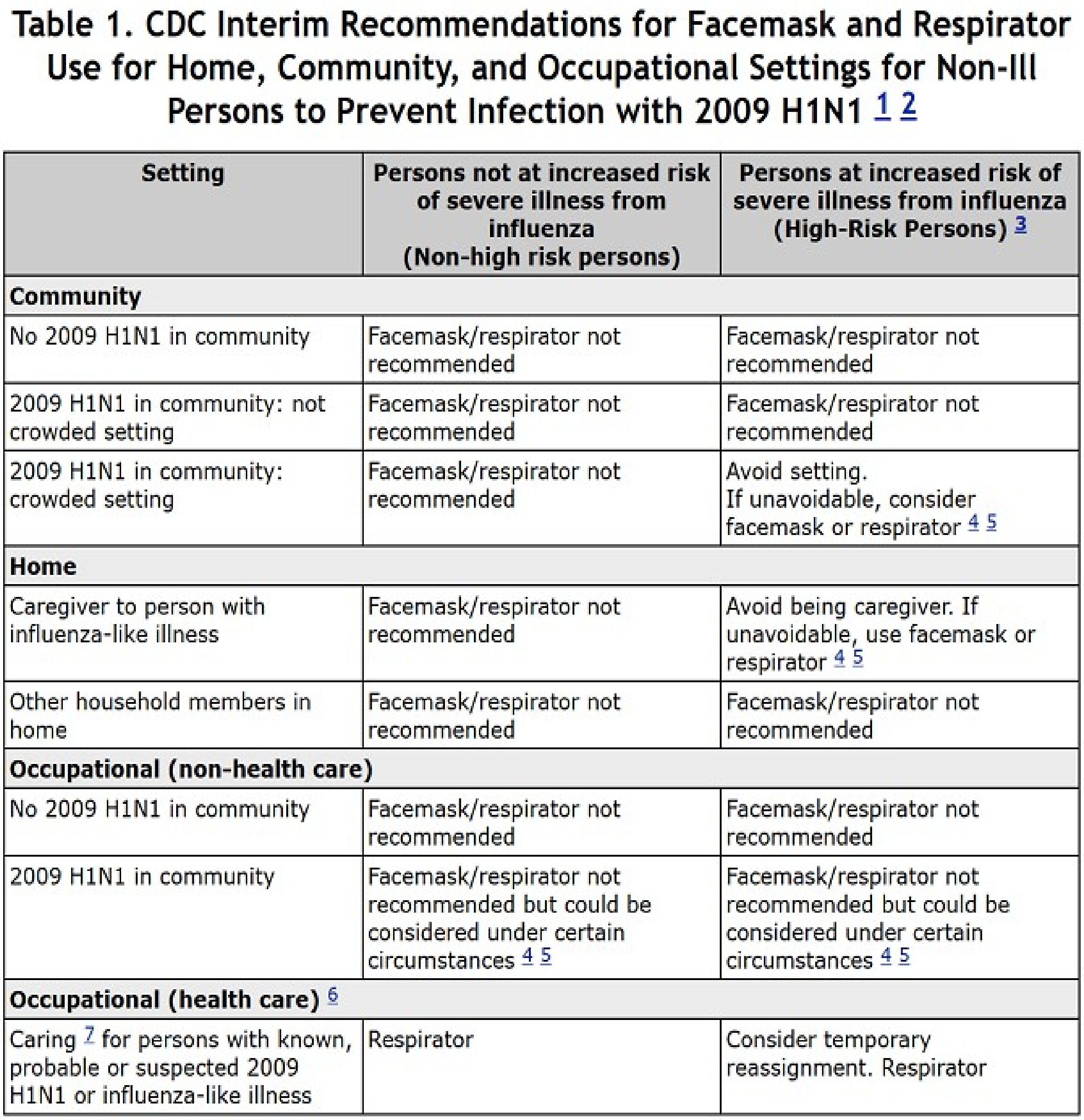


Table 1. CDC Interim Recommendations for Facemask and Respirator Use for Home, Community, and Occupational Settings for Non-Ill Persons to Prevent Infection with 2009 H1N1[1][2]

Setting	Persons not at increased risk of severe illness from influenza (Non-high risk persons)	Persons at increased risk of severe illness from influenza (High-Risk Persons)[3]
Community		
No 2009 H1N1 in community	Facemask/respirator not recommended	Facemask/respirator not recommended
2009 H1N1 in community: not crowded setting	Facemask/respirator not recommended	Facemask/respirator not recommended
2009 H1N1 in community: crowded setting	Facemask/respirator not recommended	Avoid setting. If unavoidable, consider facemask or respirator[4][5]
Home		
Caregiver to person with influenza-like illness	Facemask/respirator not recommended	Avoid being caregiver. If unavoidable, use facemask or respirator[4][5]
Other household members in home	Facemask/respirator not recommended	Facemask/respirator not recommended
Occupational (non-health care)		
No 2009 H1N1 in community	Facemask/respirator not recommended	Facemask/respirator not recommended
2009 H1N1 in community	Facemask/respirator not recommended but could be considered under certain circumstances[4][5]	Facemask/respirator not recommended but could be considered under certain circumstances[4][5]
Occupational (health care)[6]		
Caring[7] for persons with known, probable or suspected 2009 H1N1 or influenza-like illness	Respirator	Consider temporary reassignment. Respirator

And during the 2015 MERS epidemic, NPR said South Koreans were wrong to wear masks:

> Masks can be helpful for protecting health workers from a variety of infectious diseases, including MERS…
>
> But either type of mask is less likely to do much good for the average person on the street…Wearing a mask might

> make people feel better. After all, MERS has killed about a third of the people known to be infected.
>
> But there are no good studies looking at how well these masks prevent MERS transmission out in the community, says Geeta Sood, an infectious disease specialist at Johns Hopkins University. “On the street or the subway, for MERS specifically, they’re probably not effective,” she says. One problem is that the masks are loose fitting, and a lot of tiny airborne particles can get in around the sides of the masks.

So if studies generally suggest masks are effective, and the CDC wasn’t deliberately lying to us, why *are* they recommending against mask use?

I’m not sure. I haven’t been able to track down any documents where they discuss the reasons behind their policies. It’s possible they found different studies than I did, or interpreted the studies differently, or have some other superior knowledge.

But I think that more likely, they’re trying to do something different with medical communication. Consider legal communication. If a court declares a suspect is “not guilty”, that could mean that he is actually not guilty of the crime. Or it could mean that he did it but they can’t prove it. Or it could mean that he did it, they *can* prove it, but the police officer who found the proof didn’t have a warrant at the time so they had to throw it out. A legal communication like “this man is not guilty” is intended not just to convey information, but to formally reflect the output of a sacrosanct process.

Medicine has been traumatized by its century-long war with quackery, and ended up with its jargon also formally reflecting the output of a sancrosanct process. Remember, there are dozens of studies supposedly showing homeopathy works, not to mention even more studies proving telepathy exists. At some point you have to redesign all your institutions to operate in an environment of epistemic learned helplessness, and the result is very high standards of proof.

Masks haven't quite reached these standards. The case-control trials look good, and the per-protocol RCTs look good, but there aren't really the large-scale intention-to-treat RCTs that would be absolutely perfect. Even if these studies work, they only prove things about the health care setting and the family setting, not "the community setting" in general. So masks haven't been proven to work beyond a reasonable doubt. Just like the legal term for "not proven guilty beyond a reasonable doubt" is "not guilty", the medical communication term for "not proven effective beyond a reasonable doubt" is "not effective". This already muddled communication gets even worse because doctors are constitutionally incapable of distinguishing "no evidence for" from "there is evidence against" – I have no explanation for this one.

There's an even more complicated language-use issue. The CDC may be thinking of its recommendations not just as conveying an opinion but as taking an action – performing the medical intervention of recommending people wear masks. All of those RCTs listed above show that the medical intervention of recommending people wear masks is ineffective. Sure, that's because people don't listen. But the CDC doesn't care about that. They've proven that giving the advice won't help, why are you still asking them to give the advice?

I'm not sure this is really the CDC's reasoning. It seems pretty weird from the point of view of an organization trying to manage a real-world pandemic with people dying if they get it wrong. But I'm having trouble figuring out other possibilities that make sense.

8. So should you wear a mask?

Please don't buy up masks while there is a shortage and healthcare workers don't have enough.

If the shortage ends, and wearing a mask is cost-free, I agree with the guidelines from China, Hong Kong, and Japan – consider wearing a mask in high-risk situations like subways or crowded buildings. Wearing masks will not make you invincible, and if you risk compensate even a little it might do more harm than good. Realistically you should be avoiding high-risk situations like subways and crowded buildings as much as you possibly can. But if you have to go in them, yes, most likely a mask will help.

In low-risk situations, like being at home or taking a walk, I mean sure, a mask might make you 0.0001% (or whatever) less likely to get infected. If that's worth it to you, consider the possibility that you might be freaking out a little too much about this whole pandemic thing. If it's still worth it, go for it.

You are unlikely to be able to figure out how to use an N95 respirator correctly. I'm not saying it's impossible, if you try really hard, but assume you're going to fail unless you have some reason to think otherwise. The most likely outcome is that you have an overpriced surgical mask that might make you incorrectly risk-compensate.

If you are a surgeon performing surgery, bad news. It turns out surgical masks are not very useful for you (1, 2)! You should avoid buying them, since doing so may deplete the number available for people who want to wear them on the subway.

Legal Systems Very Different From Ours, Because I Just Made Them Up

[with apologies to the real Legal Systems Very Different From Ours. See also the List Of Fictional Drugs Banned By The FDA]

I.

The Clamzorians are animists. They believe every rock and tree and river has its own spirit. And those spirits are legal people. This on its own is not unusual – even New Zealand gives rivers legal personhood. But in Clamzoria, if a flood destroys your home, you sue the river.

If you win, then the river is in debt to you. The government can assign a guardian to the river to force it to pay off its debts, and that guardian gets temporary custody of all the river's property. He or she can collect a toll from boats, sell water to reservoirs, and charge rent to hydroelectric dams. Once the river has paid off its debt, the guardian is discharged, and the river becomes free to use once again.

Clamzorian precedent governs when you may or may not sue objects. If you swim in the freezing river in the dead of winter, and catch cold, that's on you. But if a hurricane destroys your property, you can absolutely sue the wind for damages, and collect from windmills. Suits against earthquakes, volcanoes, and the like are dead common. Suits against diseases happen occasionally. Sometimes someone will sue something even more abstract – a custom, an emotion, a concept.

Legend tells of a lawyer who once sued Death itself for wrongful death, a class action suit on behalf of everyone who ever lived. The judge found in favor of the plaintiff, but the appointed custodian despaired at ever collecting the judgment – the few morticians and undertakers in the realm couldn’t afford even a fraction of the damages. In a stroke of genius, he went after the military, and charged them for the right to kill enemy soldiers. The military grumbled, but eventually gave in: fair is fair.

II.

Fixed fines are inherently unfair to the poor. If you fine people $50 for running a red light, you’ve charged someone who makes $10,000 0.5% of their income, but someone who makes $100,000 gets off with only 0.05% of their income.

But prison sentences are inherently unfair to the rich. After all, if you already live in a crowded slum much like a prison cell, and your life is prison-level boring and oppressive already, then going to prison barely costs you anything. But if you live in a mansion and spend all day indulging in the finest luxuries on offer, going to prison is a massive decrease in your quality of life.

The people of Pohjankaupunki thought long and hard about this problem, and came up with a solution: crimes will be punished by neither fines nor prison. They will be punished by government mandated prescription of rimonabant, a prodepressant medication which directly saps your ability to feel happiness. Running a red light may get you 5 mg rimonabant for a month. Murder may get you 80 mg rimonabant twice a day for ten years.

There is no capital punishment in Pohjankaupunki, but if a criminal decides to commit suicide rather than continue to take their medication, they are considered to have voluntarily upgraded to the death penalty, and their debt to the state has been repaid.

III.

Sloviria is an enlightened country. They do not blame criminals for their actions. They realize it is Society's fault for making criminals that way. So when someone commits a crime, they punish Society.

Sloviria is very technologically advanced, with plenty of social networking sites and GPS tracking of cell phones and all the other systems that create a nice objective social graph. When someone commits a crime, the government lets them go free, and punishes everyone else, in proportion to how close they were to the offender on the social graph. If the punishment for a certain crime is a $1000 fine, perhaps each of their parents and their partner pays $200, their boss and best friend pay $100, some of their teachers a few tenners each, and more distant friends and relations a few dollars or less. If a friend of a friend who you met at a dinner party once commits murder, you may be out a couple of cents.

This isn't to say perpetrators get off scot-free; Sloviria isn't *that* enlightened. The punishment for perpetrators is that nobody wants to interact with them, for fear that they might perpetrate again. Once a person is a known criminal – or a suspected criminal, or just the sort of person who seems like they might become a criminal – their friends, families, and business relations shun them, trying to minimize their potential loss. This threat alone is enough to discourage crime and every form of crime-adjacent misbehavior.

The Slovirian Radical Party is even more enlightened than Sloviria as a whole, and opposes social punishment. They believe that such punishment prevents rehabilitation, since criminals and at-risk youth find it impossible to make the connections they need to succeed, and are forced to hang out with other people as criminal as themselves. They propose a complete inversion of Sloviria's justice system; when anyone commits a crime, the people closest to them are *rewarded*. They envision a future where, once somebody shows any sign of being at risk for antisocial behavior, they are love-bombed by dozens of people hoping to get rich off their acquaintance, people who want to employ them, adopt them, date them, or just serve as mentors and parental figures. But wouldn't all these people encourage the potential criminal to offend? The Radicals debate this among themselves, with one solution being that *this* could just be a perfectly normal crime punished by jail time.

IV.

Nova-Nishistan's legal system is based on blackmail. It's not *just* blackmail. There are courts and jails and so on. But few people use them. If you have evidence that someone committed a crime, you are expected to threaten to report them unless they give you money.

The system has many advantages. The person most likely to have evidence of a crime is the victim. The victim can choose how much money they want as damages, and have a good chance of receiving it. Fines are automatically calibrated to the wealth of the victim, so poor people are not stuck with debts that are impossible to pay. If a crime is victimless, or the victim chooses not to prosecute, any other witnesses are incentivized to take up the cause of punishing the wrongdoer of their own initiative. Few crimes make it to the courts or prisons, so everyone is assured a speedy trial and an jail cell free of overcrowding.

In order to maintain their system, the Nova-Nishistanis need many laws related to blackmail itself. One of their most serious crimes is to blackmail someone, receive the requested ransom, but report them anyway; anyone convicted of this will be in for a lengthy prison sentence. Indefinite blackmail – “pay me $100 now, but I might ask for more later” – is forbidden. So is non-monetary blackmail; too easy to abuse. There are a host of similar regulations.

One regulation they don’t need is laws about retaliating against blackmailers. You might expect this to be a problem – blackmailing the mob sounds pretty scary. But there are lots of individuals, companies, and (let’s face it) rival gangs happy to provide dead-man’s-switch-as-a-service. Tell them your secret (which they promise not to disclose without your consent), and if anything happens to you, they prosecute it. Even better, if anything happens to you, they’re almost guaranteed to investigate your death, since their special evidence gives them a leg up in what could be a very lucrative blackmail case.

Of course, this only works on people who are rational enough to respond to incentives. If someone is a complete unpredictable psycho, you probably don’t want to try blackmailing them, even with a dead-man’s-switch as insurance. But these are probably the people who should be in jail anyway!

V.

The people of Bogolia thought it was unfair that rich people could hire better lawyers than poor people. But they didn’t want to take the authoritarian step of banning rich people from buying good lawyers, if they thought skilled representation was important. Instead, they just mandated that in any legal case, both sides had to have equally-priced counsel. A rich person could hire as expensive a defense attorney as they wanted, as long as they donated an equal sum to the plaintiff to hire star attorneys of their own. You could sue someone with as highly-priced an attorney as you wanted, but you needed to give them the same amount to spend on their defense.

(this rule applied to the state too, and so implied the right to a public defender worth however much the state was paying to prosecute you, even if you were poor and couldn't otherwise afford one)

Some trolls tried launching hundreds of frivolous lawsuits against companies they didn't like, assuming that the company would have to pay both sides of the lawsuit and eventually go broke. They were punished through the normal anti-frivolous-lawsuit rules, and it turned out that companies that did not go broke having to pay one side of a lawsuit don't go broke having to pay both sides either.

But there were some weirder unintended consequences. How good a lawyer to get became a highly strategic decision for rich clients facing poorer ones. If you thought you were in the right, you'd get a good lawyer, since two equally good lawyers facing off will likely produce truth. If you thought you were in the wrong, you'd try to get a crappy lawyer, since then your opponent would also have a crappy lawyer, and two crappy lawyers facing off will likely produce random results. Not paying for a good lawyer started to be seen as an admission that one's case was weak.

But also, lawyer salaries started to get wacky. If a random criminal hurt a rich person somehow, and the rich person hired a good lawyer, the random criminal might receive tens of thousands of dollars to spend on legal advice. But random criminals generally are not savvy at evaluating lawyer skill, so thousands of predatory lawyers sprang up, willing to cater to these people by looking impressive and accepting very high salaries. For the savviest of political operators, an equal and opposite caste of underpriced lawyers sprang up, who would accept very low pay in exchange for vague social credit to be doled out later. More and more political scandals started to center on prestigious lawyers defending politicians for free in exchange for favors, and so depriving the opposing party of their right to equally-matched counsel.

Finally the authorities handed down a change to the system: the plaintiff and defendant would agree on two lawyers to conduct the trial. Then the judge would flip a coin, and one of the two would be assigned at random to each party.

VI.

Sanzorre accidentally became an anarcho-capitalist state under the dominion of malpractice insurance companies.

They started off by insuring doctors. Doctors know a bad malpractice case could ruin them. And although being a good doctor helps, it’s not 100%. Even the best doctor can get unlucky, or have somebody with a grudge fabricate a case against them. For that matter, even very bad doctors can get lucky and never have to deal with a case at all. So doctors have malpractice insurance, and if they seem to be practicing medicine badly their insurance company will raise their premiums.

This worked well enough that other industries started adopting it too. If a factory’s pollutant byproducts got discovered to cause cancer ten years later, their industrial malpractice insurance would pay for it. If someone slipped and fell and broke their back on a restaurant floor, their restaurant malpractice insurance would pay for it. Of course, these insurance companies worked closely with factories to monitor how many they were polluting, and gave discounts to restaurants which followed best practices on floor cleaning.

Finally, they branched out to serving ordinary people. If you accidentally hit someone’s dog with your car and got sued for damages, better to have a personal malpractice insurance pay them than get hit for tens of thousands of dollars yourself. Having malpractice insurance became to Sanzorrians what having health insurance is to Americans – a necessity if you don’t want to court disaster.

The plaintiffs in all these cases were usually being covered by lawyers who took contigency fees. But as malpractice insurance companies became better at their jobs, the contingency fees began to dry up. Finally, lobbyists from the insurance companies got contingency fees banned entirely. This presented a dilemma for ordinary people with grievances against bad actors. Thus the rise of the grievance insurance.

If you suffered harm from a doctor’s medical error, and had grievance insurance, the insurance company would pay the cost of the malpractice suit. If you were poisoned by industrial runoff, the insurance company would pay the cost of suing the factory. Grievance insurance soon became as essential as malpractice insurance. Without it, you wouldn’t be able to stand up for your rights.

Like malpractice insurance, grievance insurance was only available cheaply to people who agreed to avoid risks. If you wanted to be able to sue for malpractice, you had to avoid going to quacks. If you wanted to be able to sue factories for pollution, you couldn’t live right next to a coal plant. Gradually, grievance insurances placed more and more restrictions on people’s behavior, and people generally complied.

As malpractice insurances incentivized potential defendants to avoid actions that could harm others, and grievance insurances incentivized individuals to avoid risk, the number of lawsuits gradually got fewer and fewer. Those that happened were generally settled between malpractice insurers and grievance insurers, without ever having to go to court, and sometimes with both companies changing their policy to avoid repeats in the future. Soon, even this formality was eliminated – each malpractice insurance company paid a negotiated amount to each grievance insurance company each year, and the grievance insurance company paid complainants from its own bank account as per its own policies whenever they complained.

It wasn’t quite full anarcho-capitalism. The state still intervened in a few very serious crimes, like murder. But the insurance companies had replaced the civil courts and the regulatory apparatus, and controlled every aspect of doing business.

VII.

Modern philosophy says that formal systems are bunk. The dream of reducing the complexity of reality to some mere set of rules is a childish desire reminiscent of the fascists and high modernists of the early 20th century. Enlightened thinkers realize that we need a Kegan 5 type fluid ability to transcend systematicity. So the people of Mirakoth don’t have laws. They’re just supposed to not do bad stuff.

If someone in Mirakoth thinks someone else did something bad, they can bring it before a council of seven judges. If a majority of the judges think it was bad, they can assign whatever seems to them like fair punishment. If the loser appeals, it goes to a larger council of forty-nine judges. If they think it was bad, it was bad. These judges are under no obligation to follow precedent or any particular philosophy. They’re just supposed to be in favor of good stuff and against bad stuff.

In order to prevent people from seeking out judges who agree with them, each case is assigned seven judges at random. All cases are tried by videoconference, to make sure the judge pool is unlimited by geographical mobility. If the judges think a case is frivolous, they can choose to punish the person who brought the case.

Doesn’t this create such paralyzing uncertainty that nobody knows if they can do anything at all? Not really. Controversial cases are more likely to go to the full 49 judge panel. If an opinion is only held by 20% of judges in the country, then there’s only about a 1 in a million chance that the panel will rule in favor. Even if the opinion is held by 40%, it’s still only an 8% chance of winning. So just don’t do things that more than 40% of people think are bad, and you’ll be fine!

SSC Journal Club: MacIntyre On Cloth Masks

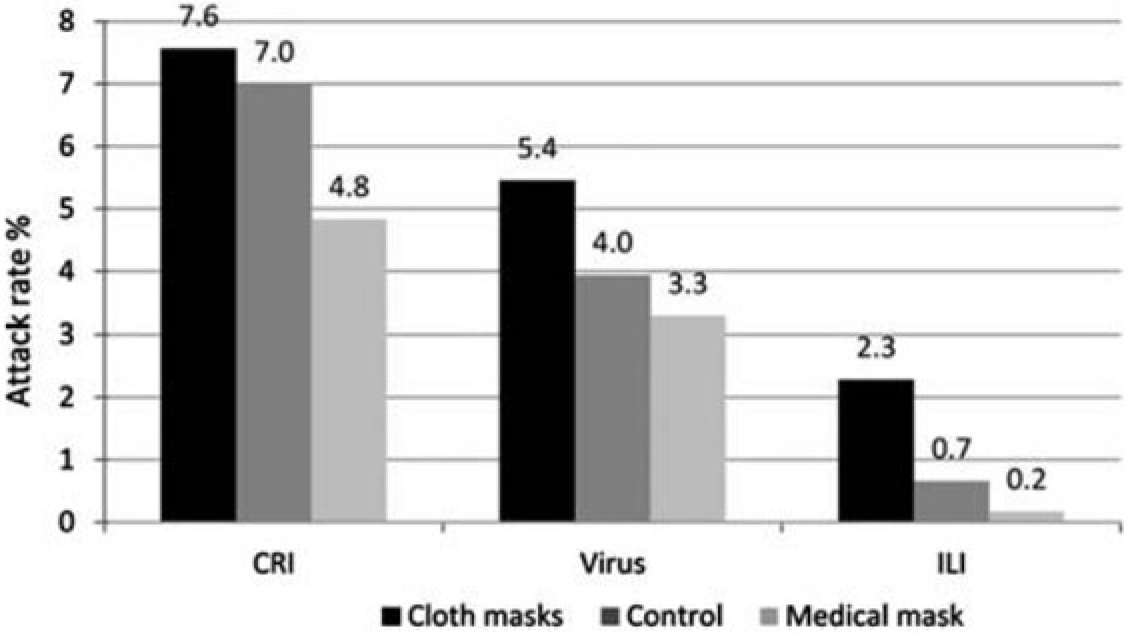
*[**Content warning**: this is a complicated analysis of something people care about a lot right now. I'm not confident in my analysis, the post comes to no clear conclusion and there are no easy answers about how to proceed. If I see this on Twitter with some headline about it DESTROYING somebody, I am going to be so mad.]*

The New York Times says that It's Time To Make Your Own Face Mask. But **MacIntyre et al (2015)** says it isn't.

The surgical masks used in hospitals are made out of non-woven fabrics that are pretty different from anything you have at home. But in some developing countries, health care workers instead use masks made of normal cloth. Laboratory tests find that improvised cloth masks block 60 – 80% of virus particles. Respirators and real surgical masks block 95%+, but 60-80% still seems better than nothing. And most of the masks ordinary people wear in Asian countries are cloth, and they seem to do pretty well. So there's some circumstantial evidence that these cloth masks might be helpful. Most experts in the early 2000s agreed that these masks were probably better than nothing. In 2015, an Australian team set out to prove it with a randomized controlled trial.

They went to a hospital in Vietnam and randomized workers there to a normal mask group, a cloth mask group, or a control group. Because it would have been unethical to tell the control group not to wear masks, they left the control group alone. Most control group workers did end up wearing masks sometimes, but less than the experimental groups did.

After a month, they counted how many infections each group had, for three different categories of infection. Here are the results:



Technically significant only in the ILI category, but later the authors do various post hoc adjustment for confounders and find it's significant everywhere

For all three categories, people wearing the real surgical masks were the healthiest, the control group was in the middle, and people wearing the cloth masks were the sickest.

This shows real surgical masks work better than cloth masks. It's a little bit unclear about how well cloth masks work. They do worse than the control group, but you could tell two stories about that. In one, cloth masks are worse than no mask at all. In the other, cloth masks have zero-to-slight-positive efficacy, but because some people in the control group were wearing real surgical masks some of the time, they did better than the cloth group overall. So it depends a lot what the control group was doing.

Unfortunately, the paper doesn’t give us all the data we want. It tells us that about 57% of both the surgical mask group and the cloth mask group wore masks regularly (defined as more than 70% of the time) but only 24% of the control group did. But there is no way of knowing whether the rest of the control group wore masks 69% of the time or 0% of the time.

A separate paragraph tells us that 37% of the control group used surgical masks, 8% cloth masks, and 53% used a combination of both. These numbers don’t make a lot of sense in the context of the last paragraph, so I’m going to assume they meant that *on the infrequent occasions they did wear masks*, those were the masks they used. But we don’t know if the compliant workers were disproportionately using cloth masks, disproportionately using medical masks, or both evenly. It’s hard to just eyeball these numbers and get a good sense for whether cloth masks really are worse than nothing.

But the authors themselves lean towards the hypothesis that that cloth masks are actively bad. First, because after some calculations I cannot quite follow, they find that the difference between surgical masks and cloth masks is so high that either the surgical masks are absurdly good, or the difference is being augmented by the cloth masks being actively bad. But nobody has previously found surgical masks to be absurdly good. The authors cite two previous studies of theirs which *did* include a no-mask control group; surgical masks did not significantly outperform nothing (they did show a trend towards doing so, and the studies were probably underpowered).

Second, because they compare the numbers from this study to numbers from those other two studies directly. They find the rate of infection in surgical mask users is not-significantly-different throughout the three studies, and the rate of infection in surgical mask users and no-mask controls was also not-significantly-different, and therefore surgical masks are the same as nothing and so probably the cloth masks are actively bad.

I am very unimpressed by this. First, you are *really* not supposed to compare things across multiple different studies. The authors protest that they did all three studies along pretty similar designs, but also admit they were different hospitals during different seasons. But second, almost no differences anywhere are significant, because all of these studies were at least a little underpowered. The current study found no significant difference between cloth masks and surgical masks in two of the three categories, even though the trend was in the expected direction. The other studies found no difference between wearing a medical mask and not wearing a medical mask, even though previous studies have suggested medical masks should work. They couldn't even find any difference between wearing an N95 respirator and not wearing any protection at all. So when you need a chain of "x is not significantly different from y, which is not significantly different from z" in a bunch of studies that wouldn't have been able to notice significant differences even if they existed, I stop believing it pretty quickly.

(In fact, I think you could use the same logic to draw the exact opposite conclusion. The cloth mask group in the current study didn't have a significant difference from the surgical mask group in the other study, and the surgical mask group was no different from placebo, therefore cloth masks cannot have a negative effect. I find it hard to believe the authors missed this, so let me know if I am confused here.)

But MacIntyre et al take it seriously, and conclude:

> The study suggests medical masks may be protective, but the magnitude of difference raises the possibility that cloth masks cause an increase in infection risk in HCWs. Further, the filtration of the medical mask used in this trial was poor, making extremely high efficacy of medical masks unlikely, particularly given the predominant pathogen was rhinovirus, which spreads by the airborne route. Given the obligations to HCW occupational health and safety, it is important to consider the potential risk of using cloth masks […] The physical properties of a cloth mask, reuse, the frequency and effectiveness of cleaning, and increased moisture retention, may potentially increase the infection risk for health care workers. The virus may survive on the surface of the face-masks, and modelling studies have quantified the contamination levels of masks. Self-contamination through repeated use and improper doffing is possible. For example, a contaminated cloth mask may transfer pathogen from the mask to the bare hands of the wearer. We also showed that filtration was extremely poor (almost 0%) for the cloth masks. Observations during SARS suggested double-masking and other practices increased the risk of infection because of moisture, liquid diffusion and pathogen retention. These effects may be associated with cloth masks.

Why am I focusing on this one weird study so much? Because it’s the only RCT of cloth face masks we have! Millions of people, egged on by top newspapers, are about to start wearing cloth face masks during a pandemic, when right now the authors of the only randomized trial on them conclude they’re probably net harmful. This should be really scary! Somebody with more experience and statistical knowledge than I have should be looking this over with a fine-toothed comb and trying to figure out what we should do.

Until then, should people stay away from cloth masks? I’m not sure, and this is *so* not a recommendation, but I lean toward no. The prior that they should work or at least be neutral is too high for a study this weak to convince me otherwise. **More important, this study only examines incoming pathogens. Even if they are harmful for blocking incoming pathogens, there are still reasons to think they are helpful for blocking outgoing ones.** If I had to hang out with a coronavirus patient for a while, and I had to choose between both of us wearing cloth masks, or neither, I would go with the masks. Only until we could get real surgical masks, which are much better. But I’d go with the cloth ones instead of nothing.

But right now that’s a gut judgment, and the evidence says I’m wrong. This is one of those times where people have to make a life-or-death decision in conditions of high uncertainty, and it really sucks.

[EDIT: Bolded a passage I think is important to make sure people don’t miss it]

Book Review: The Precipice

I.

It is a well known fact that the gods hate prophets.

False prophets they punish only with ridicule. It's the true prophets who have to watch out. The gods find some way to make their words come true in the most ironic way possible, the one where knowing the future just makes things worse. The Oracle of Delphi told Croesus he would destroy a great empire, but when he rode out to battle, the empire he destroyed was his own. Zechariah predicted the Israelites would rebel against God; they did so by killing His prophet Zechariah. Jocasta heard a prediction that she would marry her infant son Oedipus, so she left him to die on a mountainside – ensuring neither of them recognized each other when he came of age.

Unfortunately for him, Oxford philosopher Toby Ord is a true prophet. He spent years writing his magnum opus *The Precipice*, warning that humankind was unprepared for various global disasters like pandemics and economic collapses. You can guess what happened next. His book came out March 3, 2020, in the middle of a global pandemic and economic collapse. He couldn't go on tour to promote it, on account of the pandemic. Nobody was buying books anyway, on account of the economic collapse. All the newspapers and journals and so on that would usually cover an exciting new book were busy covering the pandemic and economic collapse instead. The score is still gods one zillion, prophets zero. So Ord's PR person asked me to help spread the word, and here we are.

Imagine you were sent back in time to inhabit the body of Adam, primordial ancestor of mankind. It turns out the

Garden of Eden has motorcycles, and Eve challenges you to a race. You know motorcycles can be dangerous, but you’re an adrenaline junkie, naturally unafraid of death. And it would help take your mind off that ever-so-tempting Tree of Knowledge. Do you go?

Before you do, consider that you’re not just risking your own life. A fatal injury to either of you would snuff out the entire future of humanity. Every song ever composed, every picture ever painted, every book ever written by all the greatest authors of the millennia would die stillborn. Millions of people would never meet their true loves, or get to raise their children. All of the triumphs and tragedies of humanity, from the conquests of Alexander to the moon landings, would come to nothing if you hit a rock and cracked your skull.

So maybe you shouldn’t motorcycle race. Maybe you shouldn’t even go outside. Maybe you and Eve should hide, panicked, in the safest-looking cave you can find.

Ord argues that 21st century humanity is in much the same situation as Adam. The potential future of the human race is vast. We have another five billion years until the sun goes out, and 10^100 until the universe becomes uninhabitable. Even with conservative assumptions, the galaxy could support quintillions of humans. Between Eden and today, the population would have multiplied five billion times; between today and our galactic future, it could easily multiply another five billion. However awed Adam and Eve would have been when they considered the sheer size of the future that depended on them, we should be equally awed.

So maybe we should do the equivalent of not motorcycling. And that would mean taking existential risks (“x-risks”) – disasters that might completely destroy humanity or permanently ruin its potential – very seriously. Even more seriously than we would take them just based on the fact that we don’t want to die. Maybe we should consider all counterbalancing considerations – “sure, global warming might be bad, but we also need to keep the economy strong!” – to be overwhelmed by the crushing weight of the future.

This is my metaphor, not Ord’s. He uses a different one – the Cuban Missile Crisis. We all remember the Cuban Missile Crisis as a week where humanity teetered on the precipice of destruction, then recovered into a more stable not-immediately-going-to-destroy-itself state. Ord speculates that far-future historians will remember the entire 1900s and 2000s as a sort of centuries-long Cuban Missile Crisis, a crunch time when the world was unusually vulnerable and everyone had to take exactly the right actions to make it through. Or as the namesake precipice, a place where the road to the Glorious Future crosses a narrow rock ledge hanging over a deep abyss.

Ord has other metaphors too, and other arguments. The first sixty pages of *Precipice* are a series of arguments and thought experiments intended to drive home the idea that everyone dying would be really bad. Some of them were new to me and quite interesting – for example, an argument that we should keep the Earth safe for future generations as a way of “paying it forward” to our ancestors, who kept it safe for us. At times, all these arguments against allowing the destruction of the human race felt kind of excessive – isn’t there widespread agreement on this point? Even when there is disagreement, Ord doesn’t address it here, banishing counterarguments to various appendices – one arguing against time discounting the value of the future, another arguing against ethical theories that deem future lives irrelevant. This part of the book isn’t trying to get into the intellectual weeds. It’s just saying, again and again, that it would be *really bad* if we all died.

It’s tempting to psychoanalyze Ord here, with help from passages like this one:

> I have not always been focused on protecting our longterm future, coming to the topic only reluctantly. I am a philosopher at Oxford University, specialising in ethics. My earlier work was rooted in the more tangible concerns of global health and global poverty – in how we could best help the worst off. When coming to grips with these issues I felt the need to take my work in ethics beyond the ivory tower. I began advising the World Health Organization, World Bank, and UK government on the ethics of global health. And finding that my own money could do hundreds of times as much good for those in poverty as it could do for me, I made a lifelong pledge to donate at least a tenth of all I earn to help them. I founded a society, Giving What We Can, for those who wanted to join me, and was heartened to see thousands of people come together to pledge more than one billion pounds over our lifetimes to the most effective charities we know of, working on the most important causes. Together, we’ve already been able to transform the lives of thousands of people. And because there are many other ways beyond our donations in which we can help fashion a better world, I helped start a wider movement, known as “effective altruism”, in which people aspire to use prudence and reason to do as much good as possible.

We’re in the Garden of Eden, so we should stop worrying about motorcycling and start worrying about protecting our future. But Ord’s equivalent of “motorcycling” was advising governments and NGOs on how best to fight global poverty. I’m familiar with his past work in this area, and he was amazing at it. He stopped because he decided that protecting the long-term future was more important. What must he think of the rest of us, who aren't stopping our ordinary, non-saving-thousands-of-people-from-poverty day jobs?

In writing about Ord's colleagues in the effective altruist movement, I quoted Larissa MacFarquahar on Derek Parfit:

> When I was interviewing him for the first time, for instance, we were in the middle of a conversation and suddenly he burst into tears. It was completely unexpected, because we were not talking about anything emotional or personal, as I would define those things. I was quite startled, and as he cried I sat there rewinding our conversation in my head, trying to figure out what had upset him. Later, I asked him about it. It turned out that what had made him cry was the idea of suffering. We had been talking about suffering in the abstract. I found that very striking.

Toby Ord was Derek Parfit's grad student, and I get sort of the same vibe from him – someone whose reason and emotions are unusually closely aligned. Stalin's maxim that "one death is a tragedy, a million deaths is a statistic" accurately describes how most of us think. I am not sure it describes Toby Ord. I can't say confidently that Toby Ord feels *exactly* a million times more intense emotions when he considers a million deaths than when he considers one death, but the scaling factor is definitely up there. When he considers ten billion deaths, or the deaths of the trillions of people who might inhabit our galactic future, he – well, he's reduced to writing sixty pages of arguments and metaphors trying to cram into our heads exactly how bad this would be.

II.

The second part of the book is an analysis of specific risks, how concerned we should be about each, and what we can do to prevent them. Ord stays focused on *existential* risks here. He is not very interested in an asteroid that will wipe out half of earth’s population; the other half of humanity will survive to realize our potential. He’s not *completely* uninterested – wiping out half of earth’s population could cause some chaos that makes it harder to prepare for other catastrophes. But his main focus is on things that would kill everybody – or at least leave too few survivors to effectively repopulate the planet.

I expected Ord to be alarmist here. He is writing a book about existential risks, whose thesis is that we should take them extremely seriously. Any other human being alive would use this as an opportunity to play up how dangerous these risks are. Ord is too virtuous. Again and again, he knocks down bad arguments for worrying too much, points out that killing every single human being on earth, including the ones in Antarctic research stations, is actually quite hard, and ends up convincing me to downgrade my risk estimate.

So for example, we can rule out a high risk of destruction by any natural disaster – asteroid, supervolcano, etc – simply because these things haven’t happened before in our species’ 100,000 year-odd history. Dino-killer sized asteroids seem to strike the Earth about once every few hundred million years, bounding the risk per century around the one-in-a-million level. But also, scientists are tracking almost all the large asteroids in the solar system, and when you account for their trajectories, the chance that one slips through and hits us in the next century goes down to less than one in a hundred fifty million. Large supervolcanoes seem to go off about once every 80,000 years, so the risk per century is 1/800. There are similar arguments around nearby supernovae, gamma ray bursts, and a bunch of other things.

I usually give any statistics I read a large penalty for “or maybe you’re a moron”. For example, lots of smart people said in 2016 that the chance of Trump winning was only 1%, or 0.1%, or 0.00001%, or whatever. But also, they were morons. They were using models, and their models were egregiously wrong. If you hear a person say that their model’s estimate of something is 0.00001%, very likely *your* estimate of the thing should be much higher than that, because maybe they’re a moron. I explain this in more detail here.

Ord is one of a tiny handful of people who doesn’t need this penalty. He explains this entire dynamic to his readers, agrees it is important, and adjusts several of his models appropriately. He is always careful to add a term for unknown unknowns – sometimes he is able to use clever methods to bound this term, other times he just takes his best guess. And he tries to use empirically-based methods that don’t have this problem, list his assumptions explicitly, and justify each assumption, so that you rarely have to rely on arguments shakier than “asteroids will continue to hit our planet at the same rate they did in the past”. I am really impressed with the care he puts into every argument in the book, and happy to accept his statistics at face value. People with no interest in x-risk may enjoy reading this book purely as an example of statistical reasoning done with beautiful lucidity.

When you accept very low numbers at face value, it can have strange implications. For example, should we study how to deflect asteroids? Ord isn’t sure. The base rate of asteroid strikes is so low that it’s outweighed by almost any change in the base rate. If we successfully learn how to deflect asteroids, that not only lets good guys deflect asteroids away from Earth, but also lets bad guys deflect asteroids *towards* Earth. The chance that an dino-killer asteroid approaches Earth and needs to be deflected away is 1/150 million per century, with small error bars. The chance that malicious actors deflect an asteroid towards Earth is much harder to figure out, but it has wide error bars, and there are a *lot* of numbers higher than 1/150 million. So probably most of our worry about asteroids over the next century should involve somebody using one as a weapon, and studying asteroid deflection probably makes that worse and not better.

Ord uses similar arguments again and again. Humanity has survived 100,000 years, so its chance of death by natural disaster per century is probably less than 1 / 1,000 (for complicated statistical reasons, he puts it at between 1/10,000 and 1/100,000). But humanity has only had technology (eg nuclear weapons, genetically engineered bioweapons) for a few decades, so there are no such guarantees of its safety. Ord thinks the overwhelming majority of existential risk comes from this source, and singles out four particular technological risks as most concerning.

First, nuclear war. This was one of the places where Ord's work is cause for optimism. You've probably heard that there are enough nuclear weapons to "destroy the world ten times over" or something like that. There aren't. There are enough nuclear weapons to destroy lots of majors city, kill the majority of people, and cause a very bad nuclear winter for the rest. But there aren't enough to literally kill every single human being. And because of the way the Earth's climate works, the negative effects of nuclear winter would probably be concentrated in temperate and inland regions. Tropical islands and a few other distant locales (Ord suggests Australia and New Zealand) would have a good chance of making it through even a large nuclear apocalypse with enough survivors to repopulate the Earth. A lot of things would have to go wrong at once, and a lot of models be flawed in ways they don't seem to be flawed, for a nuclear war to literally kill everyone. Ord gives the per-century risk of extinction from this cause at 1 in 1,000.

Second, global warming. The current scientific consensus is that global warming will be really bad but not literally kill every single human. Even for implausibly high amounts of global warming, survivors can always flee to a pleasantly balmy Greenland. The main concern from an x-risk point of view is "runaway global warming" based on strong feedback loops. For example, global warming causes permafrost to melt, which releases previously trapped carbon, causing more global warming, causing more permafrost to melt, etc. Or global warming causes the oceans to warm, which makes them release more methane, which causes more global warming, causing the oceans to get warmer, etc. In theory, this could get really bad – something similar seems to have happened on Venus, which now has an average temperature of 900 degrees Fahrenheit. But Ord thinks it probably won't happen here. His worst-case scenario estimates 13 – 20 degrees C of warming by 2300. This is really bad – summer temperatures in San Francisco would occasionally pass 140F – but still well short of Venus, and compatible with the move-to-Greenland scenario. Also, global temperatures jumped 5 degree C (to 14 degrees above current levels) fifty million years ago, and this didn't seem to cause Venus-style runaway warming. This isn't a perfect analogy for the current situation, since the current temperature increase is happening faster than the ancient one did, but it's still a reason for hope. This is another one that could easily be an apocalyptic tragedy unparalleled in human history but probably not an existential risk; Ord estimates the x-risk per century as 1/1,000.

The same is true for other environmental disasters, of which Ord discusses a long list. Overpopulation? No, fertility rates have crashed and the population is barely expanding anymore (also, it's hard for overpopulation to cause human extinction). Resource depletion? New discovery seems to be faster than depletion for now, and society could route around most plausible resources shortages. Honeybee collapse? Despite what you've heard, losing all pollinators would only cause a 3 – 8% decrease in global crop production. He gives all of these combined plus environmental unknown unknowns an additional 1/1,000, just in case.

Third, pandemics. Even though pathogens are natural, Ord classifies pandemics as technological disasters for two reasons. First, natural pandemics are probably getting worse because our technology is making cities denser, linking countries closer together, and bringing humans into more contact with the animal vectors of zoonotic disease (in one of the book's more prophetic passages, Ord mentions the risk of a disease crossing from bats to humans). But second, bioengineered pandemics may be especially bad. These could be either accidental (surprisingly many biologists alter diseases to make them worse as part of apparently legitimate scientific research) or deliberate (bioweapons). There are enough unknown unknowns here that Ord is uncomfortable assigning relatively precise and low risk levels like he did in earlier categories, and this section generally feels kind of rushed, but he estimates the per-century x-risk from natural pandemics as 1/10,000 and from engineered pandemics as 1/30.

The fourth major technological risk is AI. You've all read about this one by now, so I won't go into the details, but it fits the profile of a genuinely dangerous risk. It's related to technological advance, so our long and illustrious history of not dying from it thus far offers no consolation. And because it could be actively trying to eliminate humanity, isolated populations on islands or in Antarctica or wherever offer less consolation than usual. Using the same arguments and sources we've seen every other time this topic gets brought up, Ord assigns this a 1/10 risk per century, the highest of any of the scenarios he examines, writing:

> In my view, the greatest risk to humanity's potential in the next hundred years comes from unaligned artificial intelligence, which I put at 1 in 10. One might be surprised to see such a high number for such a speculative risk, so it warrants some explanation.
>
> A common approach to estimating the chance of an unprecedented event with earth-shaking consequences is to take a sceptical stance: to start with an extremely small probability and only raise it from there when a large amount of hard evidence is presented. But I disagree. Instead, I think that the right method is to start with a probability that reflects our overall impressions, then adjust this in light of the scientific evidence. When there is a lot of evidence, these approaches converge. But when there isn't, the starting point can matter.
>
> In the case of artificial intelligence, everyone agrees the evidence and arguments are far from watertight, but the question is where does this leave us? Very roughly, my approach is to start with the overall view of the expert community that there is something like a 1 in 2 chance that AI agents capable of outperforming humans in almost every task will be developed in the coming century. And conditional on that happening, we shouldn't be shocked if these agents that outperform us across the board were to inherit our future.

The book also addresses a few more complicated situations. There are ways humankind could fail to realize its potential even without being destroyed. For example, if it got trapped in some kind of dystopia that it was impossible to escape. Or if it lost so many of its values that we no longer recognized it as human. Ord doesn’t have too much to say about these situations besides acknowledging that they would be bad and need further research. Or a series of disasters could each precipitate one another, or a minor disaster could leave people unprepared for a major disaster, or something along those lines.

Here, too, Ord is more optimistic than some other sources I have read. For example, some people say that if civilization ever collapses, it will never be able to rebuild, because we’ve already used up all easily-accessible sources of eg fossil fuels, and an infant civilization won’t be able to skip straight from waterwheels to nuclear. Ord is more sanguine:

> Even if civilization did collapse, it is likely that it could be re-established. As we have seen, civilization has already been independently established at least seven times by isolated peoples. While one might think resources depletion could make this harder, it is more likely that it has become substantially easier. Most dissasters short of human extinction would leave our domesticated animals and plants, as well as copious material resources in the ruins of our cities – it is much easier to re-forge iron from old railings than to smelt it from ore. Even expendable resources such as coal would be much easier to access, via abandoned reserves and mines, than they ever were in the eighteenth century. Moreover, evidence that civilisation is possible, and the tools and knowledge to help rebuild, would be scattered across the world.

III.

Still, these risks are real, and humanity will under-respond to them for predictable reasons.

First, free-rider problems. If some people invest resources into fighting these risks and others don’t, both sets of people will benefit equally. So all else being equal everyone would prefer that someone else do it. We’ve already seen this play out with international treaties on climate change.

Second, scope insensitivity. A million deaths, a billion deaths, and complete destruction of humanity all sound like such unimaginable catastrophes that they’re hardly worth differentiating. But plausibly we should put 1000x more resources into preventing a billion deaths than a million, and some further very large scaling factor into preventing human extinction. People probably won’t think that way, which will further degrade our existential risk readiness.

Third, availability bias. Existential risks have never happened before. Even their weaker non-omnicidal counterparts have mostly faded into legend – the Black Death, the Tunguska Event. The current pandemic is a perfect example. Big pandemics happen once every few decades – the Spanish flu of 1918 and the Hong Kong Flu of 1968 are the most salient recent examples. Most countries put some effort into preparing for the next one. But the preparation wasn’t very impressive. After this year, I bet we’ll put impressive effort into preparing for respiratory pandemics the next decade or two, while continuing to ignore other risks like solar flares or megadroughts that are equally predictable. People feel weird putting a lot of energy into preparing for something that has never happened before, and their value of “never” is usually “in a generation or two”. Getting them to care about things that have *literally* never happened before, like climate change, nuclear winter, or AI risk, is an even taller order.

And even when people seem to care about distant risks, it can feel like a half-hearted effort. During a Berkeley meeting of the Manhattan Project, Edward Teller brought up the basic idea behind the hydrogen bomb. You would use a nuclear bomb to ignite a self-sustaining fusion reaction in some other substance, which would produce a bigger explosion than the nuke itself. The scientists got to work figuring out what substances could support such reactions, and found that they couldn't rule out nitrogen-14. The air is 79% nitrogen-14. If a nuclear bomb produced nitrogen-14 fusion, it would ignite the atmosphere and turn the Earth into a miniature sun, killing everyone. They hurriedly convened a task force to work on the problem, and it reported back that neither nitrogen-14 nor a second candidate isotope, lithium-7, could support a self-sustaining fusion reaction.

They seem to have been *moderately* confident in these calculations. But there was enough uncertainty that, when the Trinity test produced a brighter fireball than expected, Manhattan Project administrator James Conant was "overcome with dread", believing that atmospheric ignition had happened after all and the Earth had only seconds left. And later, the US detonated a bomb whose fuel was contaminated with lithium-7, the explosion was much bigger than expected, and some bystanders were killed. It turned out atomic bombs could initiate lithium-7 fusion after all! As Ord puts it, "of the two major thermonuclear calculations made that summer at Berkeley, they got one right and one wrong". This doesn't really seem like the kind of crazy anecdote you could tell in a civilization that was taking existential risk seriously enough.

So what should we do? That depends who you mean by "we".

Ordinary people should follow the standard advice of effective altruism. If they feel like their talents are suited for a career in this area, they should check out 80,000 Hours and similar resources and try to pursue it. Relevant careers include science (developing helpful technologies to eg capture carbon or understand AI), politics and policy (helping push countries to take risk-minimizing actions), and general thinkers and influencers (philosophers to remind us of our ethical duties, journalists to help keep important issues fresh in people's minds). But also, anything else that generally strengthens and stabilizes the world. Diplomats who help bring countries closer together, since international peace reduces the risk of nuclear war and bioweapons and makes cooperation against other threats more likely. Economists who help keep the market stable, since a prosperous country is more likely to have resources to devote to the future. Even teachers are helping train the next generation of people who can help in the effort, although Ord warns against going *too* meta – most people willing to help with this will still be best off working on causes that affect existential risk directly. If they don't feel like their talents lie in any of these areas, they can keep earning money at ordinary jobs and donate some of it (traditionally 10%) to x-risk related charities.

Rich people, charitable foundations, and governments should fund anti-x-risk work more than they're already doing. Did you know that the Biological Weapons Convention, a key international agreement banning biological warfare, has a budget lower than the average McDonald's restaurant (not total McDonald corporate profits, a single restaurant)? Or that total world spending on preventing x-risk is less than total world spending on ice cream? Ord suggests a target of between 0.1% and 1% of gross world product for anti-x-risk efforts.

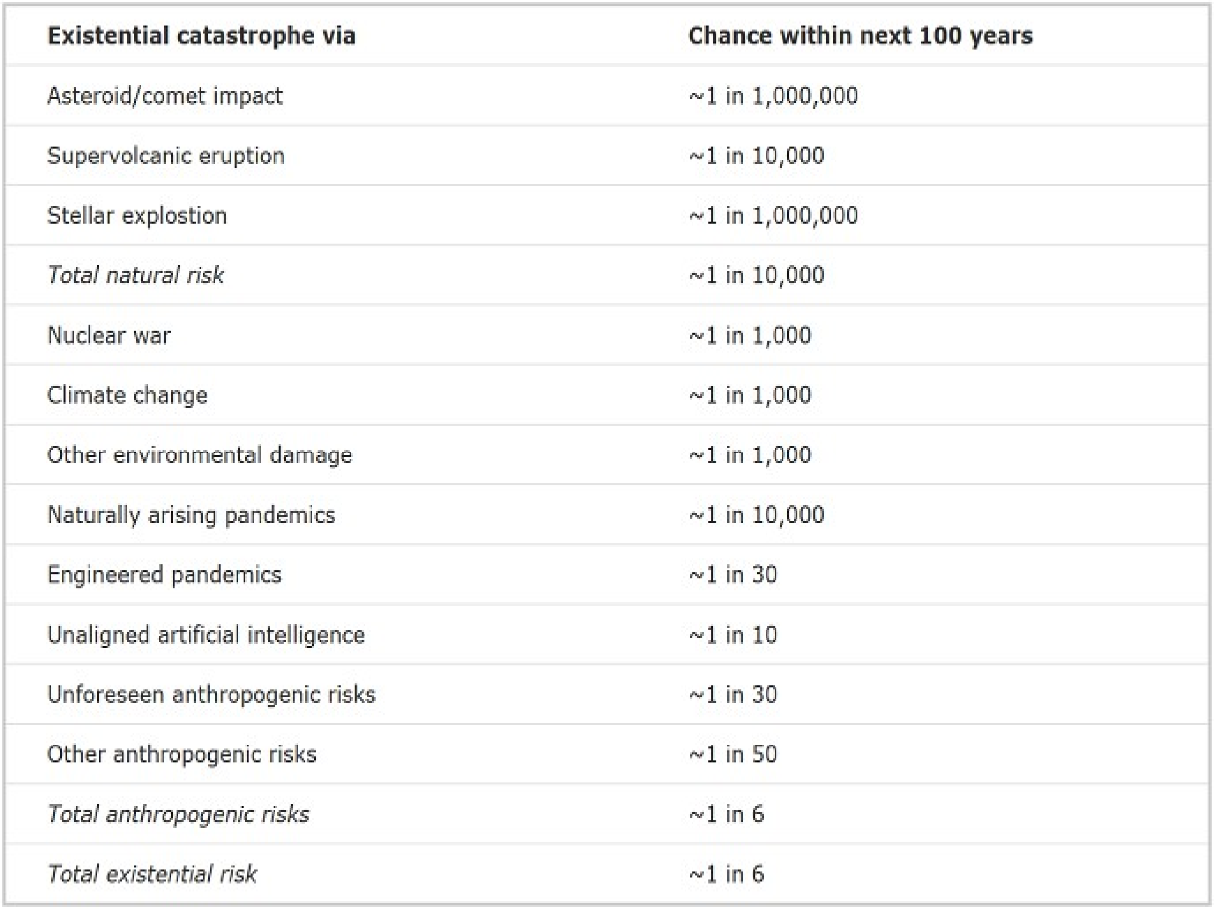
And finally, Ord has a laundry list of requests for sympathetic policy-makers (Appendix F). Most of them are to put more research and funding into things, but the actionable specific ones are: restart various nuclear disarmament treaties, take ICBMs off "hair-trigger alert", have the US rejoin the Paris

Agreement on climate change, fund the Biological Weapons Convention better, and mandate that DNA synthesis companies screen consumer requests for dangerous sequences so that terrorists can't order a batch of smallpox virus (80% of companies currently do this screening, but 20% don't). The actual appendix is six pages long, there are a *lot* of requests to put more research and funding into things.

In the last section, Ord explains that all of this is just the first step. After we've conquered existential risk (and all our other problems), we'll have another task: to contemplate how we want to guide the future. Before we spread out into the galaxy, we might want to take a few centuries to sit back and think about what our obligations are to each other, the universe, and the trillions of people who may one day exist. We cannot take infinite time for this; the universe is expanding, and for each year we spend not doing interstellar colonization, three galaxies cross the cosmological event horizon and become forever unreachable, and all the potential human civilizations that might have flourished there come to nothing. Ord expects us to be concerned about this, and tries to reassure us that it will be okay (the *relative* loss each year is only one five-billionth of the universe). So he thinks taking a few centuries to reflect before beginning our interstellar colonization is worthwhile on net. But for now, he thinks this process should take a back seat to safeguarding the world from x-risk. Deal with the Cuban Missile Crisis we're perpetually in the middle of, and *then* we'll have time for normal philosophy.

IV.

In the spirit of highly-uncertain-estimates being better than no estimates at all, Ord offers this as a draft of where the existential risk community is right now ("they are not in any way the final word, but are a concise summary of all I know about the risk landscape”):



Existential catastrophe via	**Chance within next 100 years**
Asteroid/comet impact	~1 in 1,000,000
Supervolcanic eruption	~1 in 10,000
Stellar explostion	~1 in 1,000,000
Total natural risk	~1 in 10,000
Nuclear war	~1 in 1,000
Climate change	~1 in 1,000
Other environmental damage	~1 in 1,000
Naturally arising pandemics	~1 in 10,000
Engineered pandemics	~1 in 30
Unaligned artificial intelligence	~1 in 10
Unforeseen anthropogenic risks	~1 in 30
Other anthropogenic risks	~1 in 50
Total anthropogenic risks	~1 in 6
Total existential risk	~1 in 6

Again, the most interesting thing for me is how low most of the numbers are. It’s a strange sight in a book whose thesis could be summarized as “we need to care more about existential risk”. I think most people paying attention will be *delighted* to learn there’s a 5 in 6 chance the human race will survive until 2120.

This is where I turn to my psychoanalysis of Toby Ord again. I think he, God help him, sees a number like that and *responds appropriately*. He multiplies 1/6th by 10 billion deaths and gets 1.6 billion deaths. Then he multiplies 1/6th by the hundreds of trillions of people it will prevent from ever existing, and gets tens of trillions of people. Then he considers that the centuries just keep adding up, until by 2X00 the risk is arbitrarily high. At that point, the difference between a 1/6 chance of humanity dying per century vs. a 5/6 chance of humanity dying may have psychological impact. But the overall takeaway from both is “Holy @!#$, we better put a lot of work into dealing with this.”

There’s an old joke about a science lecture. The professor says that the sun will explode in five billion years, and sees a student visibly freaking out. She asks the student what’s so scary about the sun exploding in five billion years. The student sighs with relief: “Oh, thank God! I thought you’d said five *million* years!”

We can imagine the opposite joke. A professor says the sun will explode in five minutes, sees a student visibly freaking out, and repeats her claim. The student, visibly relieved: “Oh, thank God! I thought you’d said five *seconds*.”

When read carefully, *The Precipice* is the book-length version of the second joke. Things may not be quite as disastrous as you expected. But relief may not quite be the appropriate emotion, and there’s still a lot of work to be done.

SSCJC: Real World Depression Measurement

The largest non-pharma antidepressant trial ever conducted just confirmed what we already knew: scientists love naming things after pandas.

We already had PANDAS (Pediatric Autoimmune Neuropsychiatric Disorders Associated with Streptococcus) and PANDA (Proton ANnhilator At DArmstadt). But the latest in this pandemic of panda pandering is the PANDA (Prescribing ANtiDepressants Appropriately) Study. A group of British scientists followed 655 complicated patients who received either placebo or the antidepressant sertraline (Zoloft®).

The PANDA trial was unique in two ways. First, as mentioned, it was the largest ever trial for a single antidepressant not funded by a pharmaceutical company. Second, it was designed to mimic “the real world” as closely as possible. In most antidepressant trials, researchers wait to gather the perfect patients: people who definitely have depression and definitely don’t have anything else. Then they get top psychiatrists to carefully evaluate each patient, monitor the way they take the medication, and exhaustively test every aspect of their progress with complicated questionnaires. PANDA looked for normal people going to their GP’s (US English: PCP’s) office, with all of the mishmash of problems and comorbidities that implies.

Measuring real-world efficacy is especially important for antidepressant research because past studies have failed to match up with common sense. Most studies show antidepressants having “clinically insignificant” effects on depression; that is, although scientists can find a statistical difference between treatment and placebo groups, it seems too small to matter. But in the real world, most doctors find antidepressants very useful, and many patients credit them for impressive recoveries. Maybe a big real-world study would help bridge the gap between study vs. real-world results.

The study used an interesting selection criteria – you were allowed in if you and your doctor reported "uncertainty… about the possible benefit of an antidepressant". That is, people who definitely didn't need antidepressants were sent home without an antidepressant, people who definitely *did* need antidepressants got the antidepressant, and people on the borderline made it into the study. This is very different from the usual pharma company method of using the people who desperately need antidepressants the most in order to inflate success rates. And it's more relevant to clinical practice – part of what it means for studies to guide our clinical practice is to tell us what to do in cases where we're otherwise not sure. And unlike most studies, which use strict diagnostic criteria, this study just used a perception of needing help – not even necessarily for depression, some of these patients were anxious or had other issues. Again, more relevant for clinical practice, where the borders between depression, anxiety, etc aren't always that clear.

They ended up with 655 people, ages 18-74, from Bristol, Liverpool, London, and York. They followed up on how they were doing at 2, 6, and 12 weeks after they started medication. As usual, they scored patients on a bunch of different psychiatric tests.

In the end, PANDA confirmed what we already know: it is really hard to measure antidepressant outcomes, and all the endpoints conflict with each other.

I am going to be *much* nicer to you than the authors of the original paper were to their readers, and give you a convenient table with all of the results converted to effect sizes. All values are positive, meaning the antidepressant group beat the placebo group. I calculated some of this by hand, so it may be wrong.

Endpoint	**Effect size**	**p-value**
PHQ-9	0.19	0.1
BDI	0.21	0.01
GAD-7	0.25	≤0.0001
SF-12	0.23	0.0002
PHQ-9 Remission	0.31	0.1
BDI Remission	0.55	0.049
General improvement	0.49	≤0.0001

PHQ-9 is a common depression test. BDI is another common depression test. GAD-7 is an anxiety test. SF-12 is a vague test of how mentally healthy you're feeling. Remission indicates percent of patients whose test scores have improved enough that they qualify as "no longer depressed". General improvement was just asking patients if they felt any better.

I like this study because it examines some of the mystery of why antidepressants do much worse in clinical trials than according to anecdotal doctor and patient intuitions. One possibility has always been that we're measuring these things wrong. This study goes to exactly the kind of naturalistic setting where people report good results, and measures things a bunch of different ways to see what happens.

The results are broadly consistent with previous studies. Usually people think of effect sizes less than 0.2 as miniscule, less than 0.5 as small, and less than 0.8 as medium. This study showed only small to low-medium effect sizes for everything.

I haven't checked whether differences between effect sizes were significant. But just eyeballing them, this study doesn't agree with my hypothesis that SSRIs are better for anxiety than for depression; the GAD-7 effect size is about the same as the PHQ and BDI effect sizes.

It does weakly support a hypothesis where SSRIs are better for patient-rated improvement than for researcher-measured tests. The highest effect size was in "self-rated improvement", where the researchers just asked the patients if they felt better. This effect size (0.49) was still small. But if we let ourselves round it up, it reaches all the way to "medium". Progress!

What does this mean in real life? 59% of patients in the antidepressant group, compared to 42% of patients in the placebo group, said they felt better. I'm actually okay with this. It means that for every 58 patients who wouldn't have gotten better on placebo, 17 of them would get better on an antidepressant – in other words, the antidepressant successfully converted 30% of people from nonresponder to responder. This obviously isn't as good as 50% or 100%. But it doesn't strike me as consistent with the claims of "clinically insignificant" and "why would anyone ever use these medications"?

(though of course, this is just one study, and it's a study where I took the most promising of many different endpoints, so it's not exactly cause for celebration)

If antidepressants do better on patient report than on our depression tests, does that mean our depression tests are bad? Maybe. Figure 4 from Hieronymous et al helps clarify a bit of what’s going on:

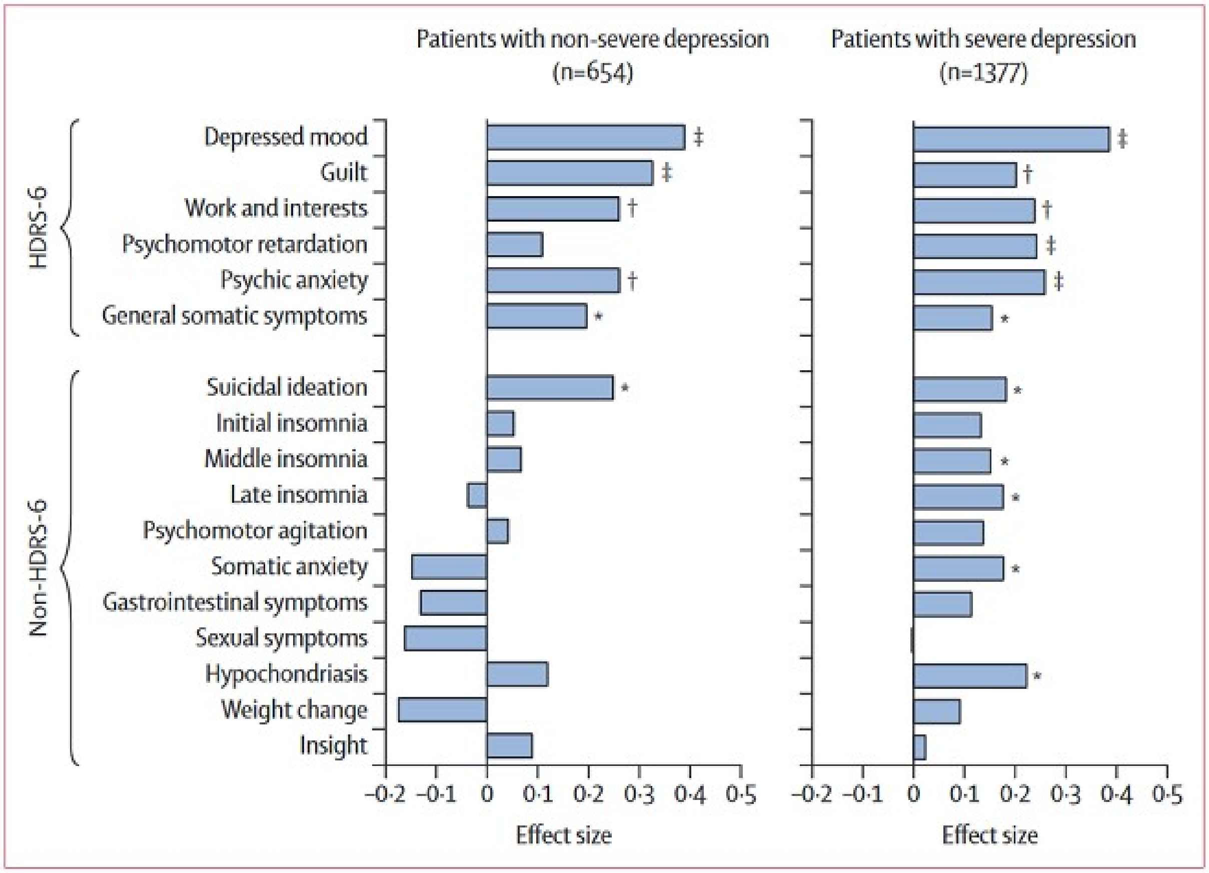


Figure 4: Endpoint effect sizes for items on the HDRS in patients with non-severe and severe depression
Effect sizes for all individual HDRS items for patients with non-severe (baseline HDRS-17 sum score ≤18 points) and severe (baseline HDRS-17 sum score ≥27 points) depression. HDRS-17=17-item Hamilton Depression Rating Scale. HDRS-6=six-item Hamilton Depression Rating Scale. non-HDRS-6=11-item Hamilton Depression Rating Scale. *p<0·05. †p<0·01. ‡p<0·001.

At least in less severely depressed patients, antidepressants are more likely to produce significant gains on vaguer or more fundamental symptoms (like “depressed mood” or “anxiety”) than on specific symptoms (like insomnia or psychomotor disruptions). Probably patients care a lot less about “psychomotor disruptions” than researchers studying depression do, and they just want to feel happy again. This study’s finding of an 0.4 – 0.5 effect size on patient response closely matches Hieronymous et al’s finding of an 0.4 – 0.5 effect size on depressed mood.

Like most studies, PANDA used a one-size-fits-all solution based on a single antidepressant. This is a reasonable choice for a study, but doesn’t match clinical practice, where we usually try one antidepressant, see if it works, and try another if it doesn’t. In patients like the ones in the study, who had failed treatment with sertraline, a usual next step would be to try bupropion. An even better idea would be to screen patients for more typical vs. atypical depression, start people on sertraline or bupropion based on the symptom profile, and then switch to the other if the first one didn’t work. The STAR*D trial did something like this, and got better results than an SSRI alone. I haven’t done the work I would need to compare this to STAR*D, but it seems possible that the extra push from targeted treatment could bring our 0.49 effect size up to the 0.7 or 0.8 level where we could actually feel fully confident prescribing this stuff.

Never Tell Me The Odds (Ratio)

[Epistemic status: low confidence, someone tell me if the math is off. Title was stolen from an old Less Wrong post that seems to have disappeared – let me know if it’s yours and I’ll give you credit]

I almost screwed up yesterday’s journal club. The study reported an odds ratio of 2.9 for antidepressants. Even though I *knew* odds ratios are terrible and you should never trust your intuitive impression of them, I *still* mentally filed this away as “sounds like a really big effect”.

This time I was saved by Chen’s How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies, which explains how to convert ORs into effect sizes. Colored highlights are mine. I have followed the usual statistical practice of interpreting effect sizes of 0.2 as “small”, of 0.5 as “moderate”, and 0.8 as “large”, but feeling guilty about it.

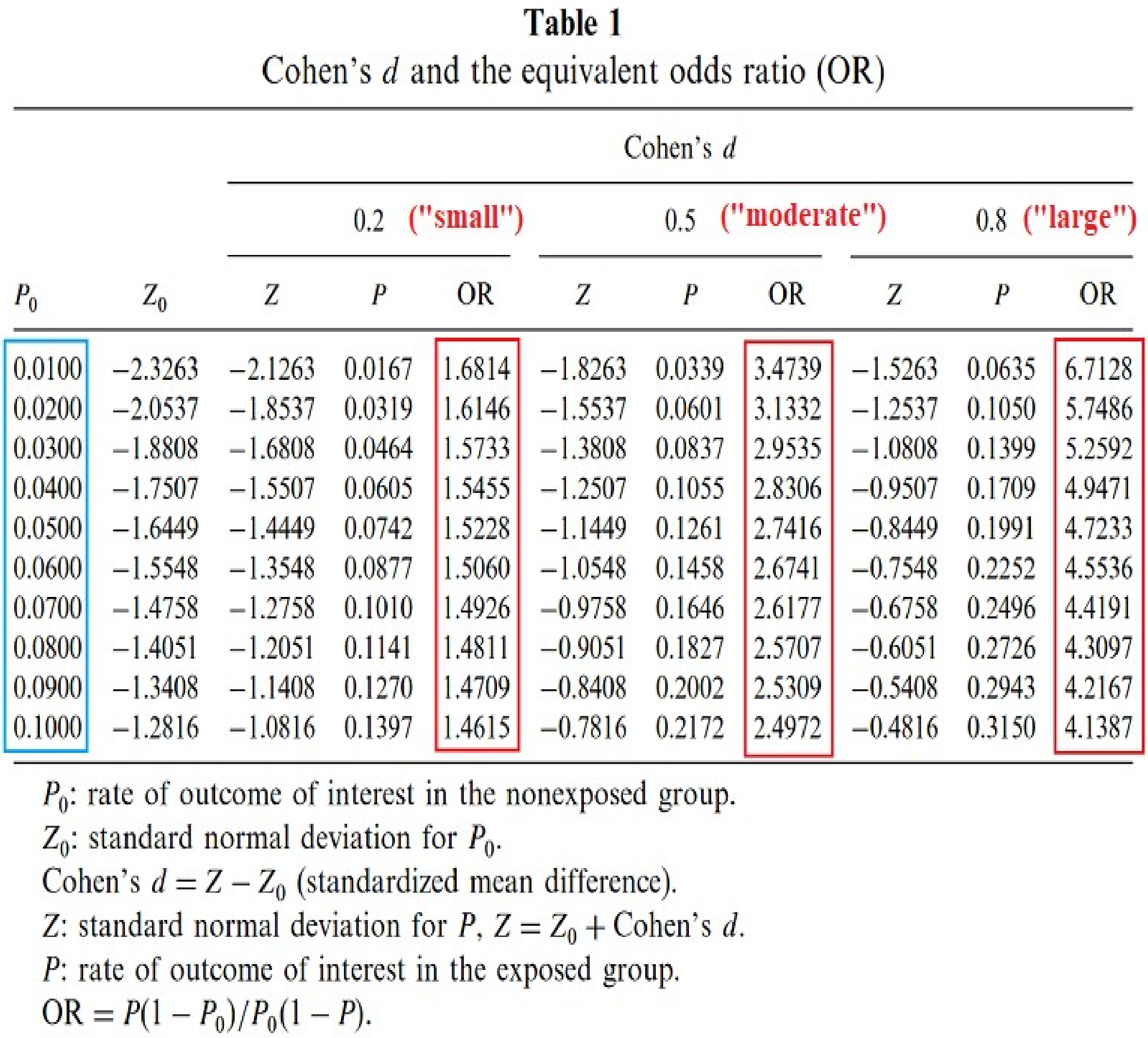


Table 1

Cohen's d and the equivalent odds ratio (OR)

		Cohen's d								
		0.2 ("small")			0.5 ("moderate")			0.8 ("large")		
P_0	Z_0	Z	P	OR	Z	P	OR	Z	P	OR
0.0100	−2.3263	−2.1263	0.0167	1.6814	−1.8263	0.0339	3.4739	−1.5263	0.0635	6.7128
0.0200	−2.0537	−1.8537	0.0319	1.6146	−1.5537	0.0601	3.1332	−1.2537	0.1050	5.7486
0.0300	−1.8808	−1.6808	0.0464	1.5733	−1.3808	0.0837	2.9535	−1.0808	0.1399	5.2592
0.0400	−1.7507	−1.5507	0.0605	1.5455	−1.2507	0.1055	2.8306	−0.9507	0.1709	4.9471
0.0500	−1.6449	−1.4449	0.0742	1.5228	−1.1449	0.1261	2.7416	−0.8449	0.1991	4.7233
0.0600	−1.5548	−1.3548	0.0877	1.5060	−1.0548	0.1458	2.6741	−0.7548	0.2252	4.5536
0.0700	−1.4758	−1.2758	0.1010	1.4926	−0.9758	0.1646	2.6177	−0.6758	0.2496	4.4191
0.0800	−1.4051	−1.2051	0.1141	1.4811	−0.9051	0.1827	2.5707	−0.6051	0.2726	4.3097
0.0900	−1.3408	−1.1408	0.1270	1.4709	−0.8408	0.2002	2.5309	−0.5408	0.2943	4.2167
0.1000	−1.2816	−1.0816	0.1397	1.4615	−0.7816	0.2172	2.4972	−0.4816	0.3150	4.1387

P_0: rate of outcome of interest in the nonexposed group.
Z_0: standard normal deviation for P_0.
Cohen's $d = Z - Z_0$ (standardized mean difference).
Z: standard normal deviation for P, $Z = Z_0 + \text{Cohen's } d$.
P: rate of outcome of interest in the exposed group.
$\text{OR} = P(1 - P_0)/P_0(1 - P)$.

Based on this page, I gather Chen has used some unusually precise formula to calculate this, but that a quick heuristic is to ignore the prevalence and just take [ln(odds ratio)]/1.81.

Suppose you run a drug trial. In your control group of 1000 patients, 300 get better on their own. In your experimental group of 1000 patients, 600 get better total (presumably 300 on their own, 300 because your drug worked). The relative risk calculator says your relative risk of recovery on the drug is 2.0. Odds ratio is 3.5, effect size is 0.7. So you've managed to double the recovery rate – in fact, to save an entire extra 30% of your population – and you *still* haven’t qualified for a “large” effect size.

The moral of the story is that (to me) odds ratios sound bigger than they really are, and effect sizes sound smaller, so you should be really careful comparing two studies that report their results differently.

2019 Predictions: Calibration Results

At the beginning of every year, I make predictions. At the end of every year, I score them (this year I'm very late). Here are 2014, 2015, 2016, 2017, and 2018.

And here are the predictions I made for 2019. Strikethrough'd are false. Intact are true. Italicized are getting thrown out because I can't decide if they're true or not. All of these judgments were as of December 31 2019, not as of now.

Please don't complain that 50% predictions don't mean anything; I know this is true but there are some things I'm genuinely 50-50 unsure of. Some predictions are redacted because they involve my private life or the lives of people close to me. A few that started off redacted stopped being secret; I've put those in [brackets].

US
1. Donald Trump remains President: 90%
2. Donald Trump is impeached by the House: 40%
~~3. Kamala Harris leads the Democratic field: 20%~~
~~4. Bernie Sanders leads the Democratic field: 20%~~
5. Joe Biden leads the Democratic field: 20%
~~6. Beto O'Rourke leads the Democratic field: 20%~~
7. Trump is still leading in prediction markets to be Republican nominee: 70%
8. Polls show more people support the leading Democrat than the leading Republican: 80%
9. Trump's approval rating below 50: 90%
~~10. Trump's approval rating below 40: 50%~~
~~11. Current government shutdown ends before Feb 1: 40%~~
12. Current government shutdown ends before Mar 1: 80%
13. Current government shutdown ends before Apr 1: 95%

~~14. Trump gets at least half the wall funding he wants from current shutdown: 20%~~
15. Ginsberg still alive: 50%

ECON AND TECH
16. Bitcoin above 1000: 90%
17. Bitcoin above 3000: 50%
18. Bitcoin above 5000: 20%
19. Bitcoin above Ethereum: 95%
20. Dow above current value of 25000: 80%
~~21. SpaceX successfully launches and returns crewed spacecraft: 90%~~
~~22. SpaceX Starship reaches orbit: 10%~~
23. No city where a member of the general public can ride self-driving car without attendant: 90%
~~24. I can buy an Impossible Burger at a grocery store within a 30 minute walk from my house: 70%~~
25. Pregabalin successfully goes generic and costs less than $100/month on GoodRx.com: 50%
26. No further CRISPR-edited babies born: 80%

WORLD
~~27. Britain out of EU: 60%~~
~~28. Britain holds second Brexit referendum: 20%~~
29. No other EU country announces plan to leave: 80%
30. China does not manage to avert economic crisis (subjective): 50%
31. Xi still in power: 95%
32. MbS still in power: 95%
~~33. May still in power: 70%~~
34. Nothing more embarassing than Vigano memo happens to Pope Francis: 80%

SURVEY
35. …finds birth order effect is significantly affected by age gap: 40%

36. …finds fluoxetine has significantly less discontinuation issues than average: 60%
37. …finds STEM jobs do not have significantly more perceived gender bias than non-STEM: 60%
38. …finds gender-essentialism vs. food-essentialism correlation greater than 0.075: 30%

PERSONAL – INTERNET
39. SSC gets fewer hits than last year: 70%
40. I finish and post [New Atheism: The Godlessness That Failed]: 90%
~~41. I finish and post [Structures Of Paranoia]: 50%~~
42. I finish and post [a sequence based on Secret Of Our Success]: 50%
43. [New Atheism] post gets at least 40,000 hits: 40%
~~44. [The Proverbial Murder Mystery] post gets at least 40,000 hits: 20%~~
~~45. New co-blogger with more than 3 posts: 20%~~
46. Repeat adversarial collaboration contest with at least 5 entries: 60%
47. [Culture War thread successfully removed from subreddit]: 90%
48. [Culture War new version getting at least 500 comments per week]: 70%
~~49. I start using Twitter again (5+ tweets in any month): 60%~~
50. I start using Facebook again (following at least 5 people): 30%

PERSONAL – HEALTH
51. I get the blood tests I should be getting this year: 90%
52. I try one biohacking project per month x at least 10 months: 30%
53. I continue taking sceletium regularly: 70%
~~54. I switch from [Zembrin to Tristill] for at least 3 months: 20%~~
~~55. I find at least one new supplement I take or expect to~~

~~take regularly x 3 months: 20%~~
~~56. Minoxidil use produces obvious progress: 50%~~
~~57. I restart [redacted]: 20%~~
58. I spend one month at least substantially more vegetarian than my current compromise: 20%
~~59. I spend one month at least substantially less vegetarian than my current compromise: 30%~~
60. I weigh more than 195 lbs at year end: 80%
~~61. I meditate at least 30 minutes/day more than half of days this year: 30%~~
~~62. I use marijuana at least once this year: 20%~~

PERSONAL – PROJECTS
~~63. I finish at least 10% more of [redacted]: 20%~~
~~64. I completely finish [redacted]: 10%~~
~~65. I finish and post [redacted]: 5%~~
~~66. I write at least ten pages of something I intend to turn into a full-length book this year: 20%~~
~~67. I practice calligraphy at least seven days in the last quarter of 2019: 40%~~
~~68. I finish at least one page of the [redacted] calligraphy project this year: 30%~~
~~69. I finish the entire [redacted] calligraphy project this year: 10%~~
~~70. I finish some other at-least-one-page calligraphy project this year: 80%~~

PERSONAL – PROFESSIONAL
71. I attend the APA Meeting: 80%
~~72. [redacted]: 50%~~
73. [redacted]: 40%
74. I still work in SF with no plans to leave it: 60%
75. I still only do telepsychiatry one day with no plans to increase it: 60%
76. I still work the current number of hours per week: 60%
77. I have not started (= formally see first patient) my own

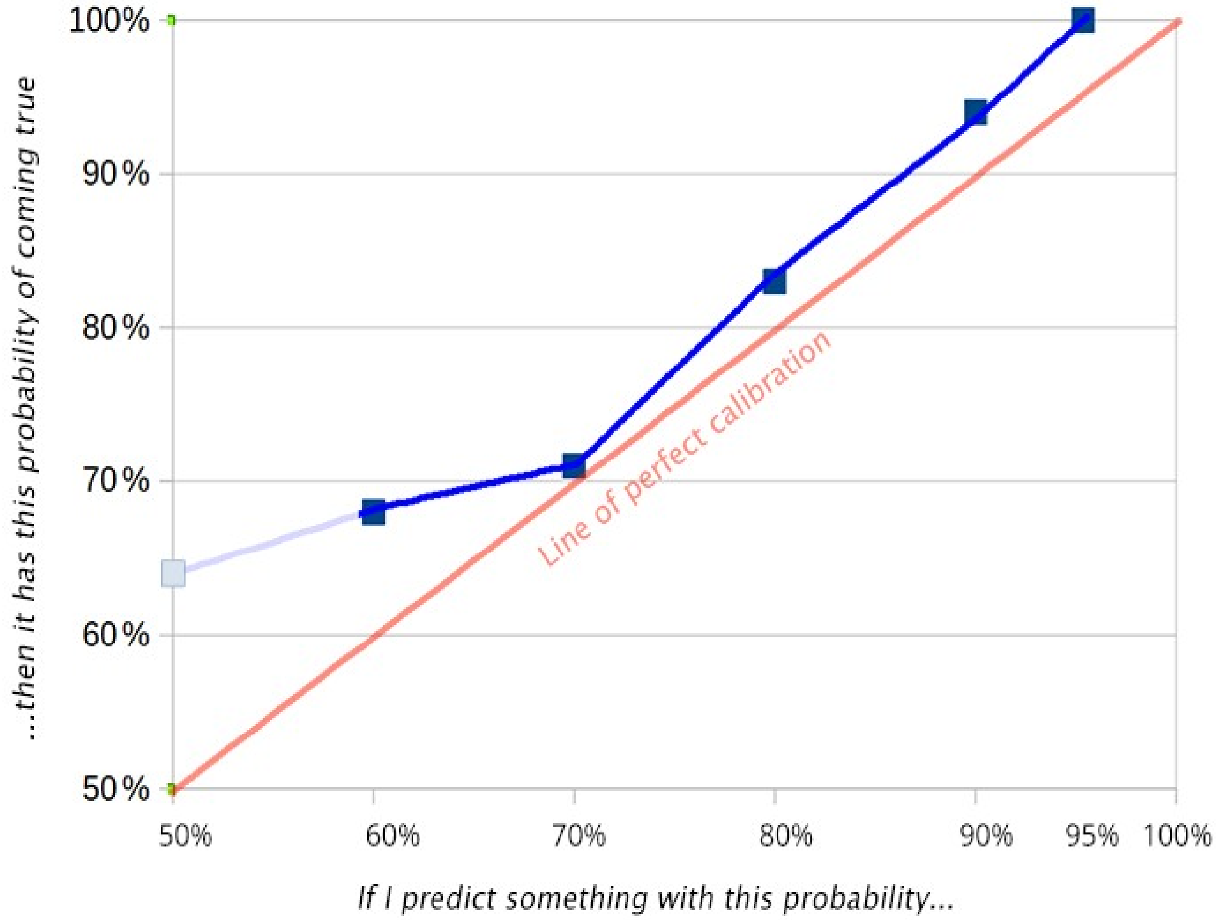
practice: 80%
78. I lease another version of the same car I have now: 90%

PERSONAL – HOUSE
79. I still live in my current house with no specific plans to leave: 80%
80. I set up a decent home library: 60%
81. We got a second trash can: 90%
82. The gate is fixed with no problems at all: 50%
83. The ugly paint spot on my wall gets fixed: 30%
84. There is some kind of nice garden: 60%
85. …and I am at least half responsible: 20%
~~86. I get my own washing machine: 20%~~
87. There is another baby in my house: 60%
88. No other non-baby resident (expected 6+ month) in my house who doesn't live there now: 70%
89. No existing resident moves away (except the one I already know about): 80%
90. No other long-term (expected 6+ month) resident of my subunit who doesn't live there now: 80%
91. [Decision Tree House] is widely considered a success: 70%
~~92. …with plans (vague okay) to create a second one: 20%~~

PERSONAL – ROMANCE
~~93. I find a primary partner: 30%~~
94. I go on at least one date with someone who doesn't already have a primary partner: 90%
95. I remake an account on OKCupid: 80%
~~96. [redacted]: 10%~~
~~97. [redacted]: 20%~~
~~98. [redacted]: 20%~~
~~99. [redacted]: 20%~~
100. [redacted]: 20%
~~101. [redacted]: 30%~~
~~102. [redacted]: 10%~~

103. [redacted]: 30%
~~104. [I go on at least three dates with someone I have not yet met]: 50%~~
~~105. [redacted]: 10%~~
106. [redacted]: 50%

PERSONAL – FRIENDS
107. I am still playing D&D: 60%
108. I go on a trip to Guatemala: 90%
~~109. I go on at least one other international trip: 30%~~
~~110. I go to at least one Solstice outside the Bay: 40%~~
111. I go to at least one city just for an SSC meetup: 30%
112. [redacted] is in a relationship: 40%
~~113. [redacted] still has their current partner: 50%~~
~~114. [redacted] is at their current job: 20%~~
115. [redacted] is still at their current job: 80%
~~116. I hang out with [redacted] at least once: 60%~~
117. I hang out with [redacted] at least once: 60%
118. I am in [redacted] Discord server: 80%



Calibration chart. The red line represents perfect calibration, the blue my predictions. The closer they are, the better I am doing.

Of 11 predictions at 50%, I got 4 wrong and 7 right, for an average of 64%
Of 22 predictions at 60%, I got 7 wrong and 15 right, for an average of 68%
Of 17 predictions at 70%, I got 5 wrong and 12 right, for an average of 71%
Of 37 predictions at 80%, I got 6 wrong and 31 right, for an average of 83%
Of 17 predictions at 90%, I got 1 wrong and 16 right, for an average of 94%
Of 5 predictions at 95%, I got 0 wrong and 5 right, for an average of 100%

50% predictions are technically meaningless since I could have written them either way. I’ve lightened them on the chart to indicate they can be ignored.

It was another good year for me. Unlike past years, where I erred about evenly in both directions, this year I was about 4% underconfident across the board. I’m not sure how much I should adjust and become more confident. In past years I’ve been burned by major black swan events that affect multiple predictions and made me look overconfident. In 2019 I tried to leave a cushion for that, but nothing too unexpected happened and I ended up playing it too safe. My worst failures were underestimating Bitcoin (but who didn’t?) and overestimating SpaceX’s ability to launch their crew on schedule. I didn’t check formally, but there doesn’t seem to be much difference in my calibration about world affairs vs. my personal life.

I forgot to make predictions for 2020 until now, which in retrospect was the best prediction I’ve ever made. I’ll probably come up with some later this month.

A Failure, But Not Of Prediction

I.

Vox asks What Went Wrong With The Media's Coronavirus Coverage? They conclude that the media needs to be better at "not just saying what we do know, but what we don't know". This raises some important questions. Like: how much ink and paper is there in the world? Are we sure it's enough? But also: how do you become better at saying what you don't know?

In case you've been hiding under a rock recently (honestly, valid) the media not only failed to adequately warn its readers about the epidemic, but actively mocked and condescended to anyone who *did* sound a warning. Real Clear Politics has a list of highlights. The Vox tweet saying "Is this going to be a deadly pandemic? No." Washington Post telling us in February "Why we should be wary of an aggressive government reponse to coronavirus (it might "scapegoat marginalized populations"). The Daily Beast complaining that "coronavirus, with zero American fatalities, is dominating headlines, while the flu is the real threat". The *New York Times*, weighing in with articles like "The pandemic panic" and "Who says it's not safe to travel to China". The constant attempts to attribute "alarmism" over the virus to anti-Chinese racism. Etc, etc, etc.

One way people have summed this up is that the media (and the experts they relied on) did a terrible job predicting what would happen. I think this lets them off too easy.

Prediction is very hard. Nate Silver is maybe the best political predicter alive, and he estimated a 29% chance of Trump winning just before Trump won. UPenn professor

Philip Tetlock has spent decades identifying "superforecasters" and coming up with complicated algorithms for aggregating their predictions, developing a prediction infrastructure that beats top CIA analysts, but they estimated a 23% chance Britain would choose Brexit just before it happened. This isn't intended to criticize Silver or Tetlock. I believe they're operating at close to optimum – the best anyone could possibly do with the information that they had. But the world is full of noise, and tiny chance events can have outsized effects, and there are only so many polls you can scrutinize, and even geniuses can only do so well.

Predicting the coronavirus was equally hard, and the best institutions we had missed it. On February 20th, Tetlock's superforecasters predicted only a 3% chance that there would be 200,000+ coronavirus cases a month later (there were). The stock market is a giant coordinated attempt to predict the economy, and it reached an all-time high on February 12, suggesting that analysts expected the economy to do great over the following few months. On February 20th it fell in a way that suggested a mild inconvenience to the economy, but it didn't really start plummeting until mid-March – the same time the media finally got a clue. These aren't empty suits on cable TV with no skin in the game. These are the best predictive institutions we have, and they got it wrong. I conclude that predicting the scale of coronavirus in mid-February – the time when we could have done something about it – was really hard.

I don't like this conclusion. But I have to ask myself – if it was so easy, why didn't *I* do it? It's easy to look back and say "yeah, I always secretly knew it would be pretty bad". I did a few things right – I started prepping half-heartedly in mid-February, I recommended my readers prep in early

March, I never criticized others for being alarmist. Overall I give myself a solid B-. But if it was so easy, why didn't I post "Hey everyone, I officially predict the coronavirus will be a nightmarish worldwide pandemic" two months ago? It wouldn't have helped anything, but I would have had bragging rights forever. For that matter, why didn't *you* post this – on Facebook, on Twitter, on the comments here? You could have gone down in legend, alongside Travis W. Fisher, for making a single tweet. Since you didn't do that (aside from the handful of you who did – we love you, Balaji) I conclude that predicting it was hard, even for smart and well-intentioned people like yourselves.

Does that mean we can't put everyone's heads on spikes outside the Capitol Building as a warning for future generations? I would be very disappointed if it meant that. I think we can still put heads on spikes. We just have to do it for more subtle, better-thought-out reasons.

II.

I used to run user surveys for a forum on probabilistic reasoning…

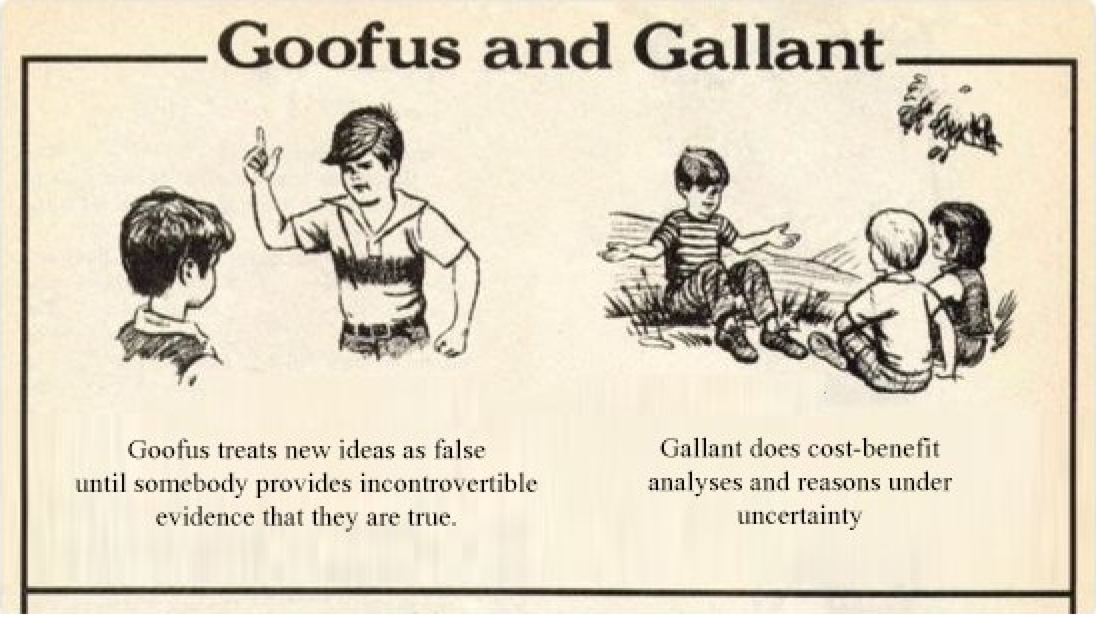
(I promise this will become relevant soon)

A surprising number of these people had signed up for cryonics – the thing where they freeze your brain after you die, in case the future invents a way to resurrect frozen brains. Lots of people mocked us for this – "if you're so good at probabilistic reasoning, how can you believe something so implausible?" I was curious about this myself, so I put some questions on one of the surveys.

The results were pretty strange. Frequent users of the forum (many of whom had pre-paid for brain freezing) said they estimated there was a 12% chance the process would work and they'd get resurrected. A control group with no interest in cryonics estimated a 15% chance. The people who were doing it were no more optimistic than the people who weren't. What gives?

I think they were actually good at probabilistic reasoning. The control group said "15%? That's less than 50%, which means cryonics probably won't work, which means I shouldn't sign up for it." The frequent user group said "A 12% chance of eternal life for the cost of a freezer? Sounds like a good deal!"

There are a lot of potential objections and complications – for one thing, maybe both those numbers are much too high. You can read more here and here. But overall I learned something really important from this.



Goofus and Gallant

Goofus treats new ideas as false until somebody provides incontrovertible evidence that they are true.

Gallant does cost-benefit analyses and reasons under uncertainty

Making decisions is about more than just having certain beliefs. It's also about how you act on them.

III.

A few weeks ago, I wrote a blog post on face masks. It reviewed the evidence and found that they probably helped prevent the spread of disease. Then it asked: how did the WHO, CDC, etc get this so wrong?

I went into it thinking they'd lied to us, hoping to prevent hoarders from buying up so many masks that there weren't enough for health workers. Turns out that's not true. The CDC has been singing the same tune for the past ten years. Swine flu, don't wear masks. SARS, don't wear masks. They've been really consistent on this point. But why?

If you really want to understand what happened, don't read any studies about face masks or pandemics. Read Smith & Pell (2003), Parachute Use To Prevent Death And Major Trauma Related To Gravitational Challenge: Systematic Review Of Randomized Controlled Trials. It's an article in the *British Journal Of Medicine* pointing out that there have never been any good studies proving that parachutes are helpful when jumping out of a plane, so they fail to meet the normal standards of evidence-based medicine. From the Discussion section:

> It is a truth universally acknowledged that a medical intervention justified by observational data must be in want of verification through a randomised controlled trial. Observational studies have been tainted by accusations of data dredging, confounding, and bias. For example, observational studies showed lower rates of ischaemic heart disease among women using hormone replacement therapy, and these data were interpreted as advocating hormone replacement for healthy women, women with established ischaemic heart disease, and women with risk factors for ischaemic heart disease. However, randomised controlled trials showed that hormone replacement therapy actually increased the

> risk of ischaemic heart disease, indicating that the apparent protective effects seen in observational studies were due to bias. Cases such as this one show that medical interventions based solely on observational data should be carefully scrutinised, and the parachute is no exception.

Of course this is a joke. It’s in the all-joke holiday edition of BMJ, and everyone involved knew exactly what they were doing. But the joke is funny because it points at something true. It’s biting social commentary. Doctors will not admit any treatment could possibly be good until it has a lot of randomized controlled trials behind it, common sense be damned. This didn’t come out of nowhere. They’ve been burned lots of times before by thinking they were applying common sense and getting things really wrong. And after your mistakes kill a few thousand people you start getting really paranoid and careful. And there are so many quacks who can spout off some “common sense” explanation for why their vitamin-infused bleach or colloidal silver should work that doctors have just become immune to that kind of bullshit. Multiple good RCTs or it didn’t happen. Given the history I think this is a defensible choice, and if you are tempted to condemn it you may find this story about bone marrow transplants enlightening.

But you can take this too far. After highlighting the lack of parachute RCTs, the paper continues:

> Only two options exist. The first is that we accept that, under exceptional circumstances, common sense might be applied when considering the potential risks and benefits of interventions. The second is that we continue our quest for the holy grail of exclusively evidence based interventions and preclude parachute use outside the context of a properly conducted trial. The

> dependency we have created in our population may make recruitment of the unenlightened masses to such a trial difficult. If so, we feel assured that those who advocate evidence based medicine and criticise use of interventions that lack an evidence base will not hesitate to demonstrate their commitment by volunteering for a double blind, randomised, placebo controlled, crossover trial.

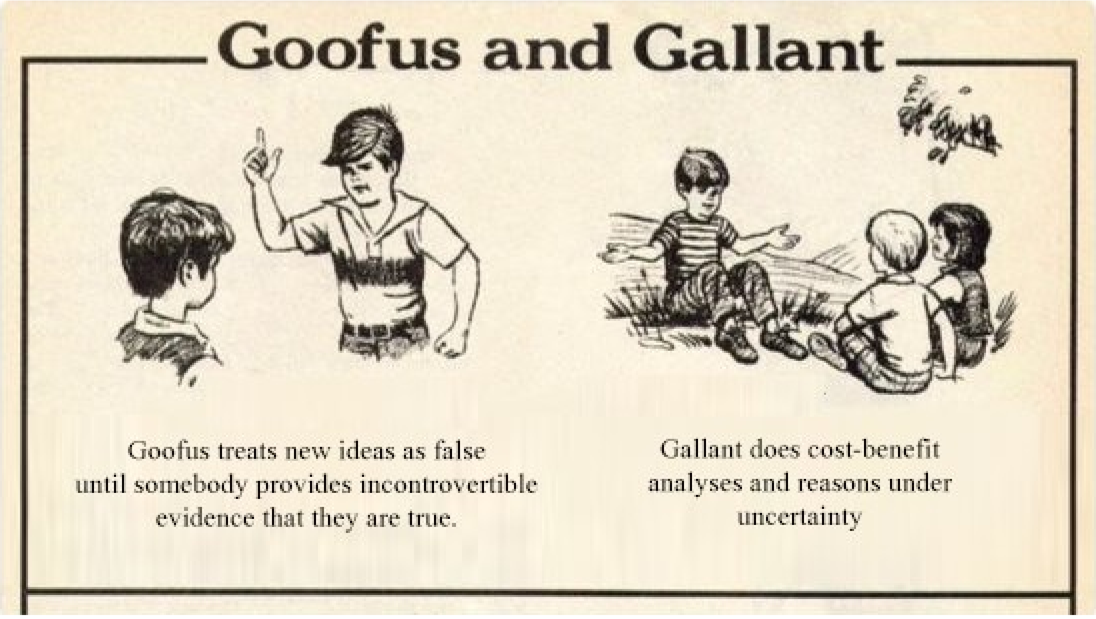
Did you follow that? For a good parachute RCT, half the subjects would have to jump out of a plane wearing a placebo parachute. The authors suggest maybe we enlist doctors who insist too stringently on RCTs over common sense for this dubious honor.

(good news, though, a parachute RCT did eventually get done)

Sometimes good humor is a little too on the nose, like those Onion articles that come true a few years later. The real medical consensus on face masks came from pretty much the same process as the fake medical consensus on parachutes. Common sense said that they worked. But there weren't many good RCTs. We couldn't do more, because it would have been unethical to deliberately expose face-mask-less people to disease. In the end, all we had were some mediocre trials of slightly different things that we had to extrapolate out of range.

Just like the legal term for "not proven guilty beyond a reasonable doubt" is "not guilty", the medical term for "not proven to work in several gold-standard randomized controlled trials" is "it doesn't work" (and don't get me started on "no evidence"). So the CDC said masks didn't work.

Going back to our diagram:


Goofus and Gallant
Goofus treats new ideas as false until somebody provides incontrovertible evidence that they are true.
Gallant does cost-benefit analyses and reasons under uncertainty

Goofus started with the position that masks, being a new idea, needed incontrovertible proof. When the few studies that appeared weren't incontrovertible enough, he concluded that people shouldn't wear masks.

Gallant would have recognized the uncertainty – based on the studies we can't be 100% sure masks definitely work for this particular condition – and done a cost-benefit analysis. Common sensically, it seems like masks probably *should* work. The existing evidence for masks is highly suggestive, even if it's not utter proof. Maybe 80% chance they work, something like that? If you can buy an 80% chance of stopping a deadly pandemic for the cost of having to wear some silly cloth over your face, probably that's a good deal. Even though regular medicine has good reasons for being as conservative as it is, during a crisis you have to be able to think on your feet.

IV.

But getting back to the media:

Their main excuse is that they were just relaying expert opinion – the sort of things the WHO and CDC and top epidemiologists were saying. I believe them. People on Twitter howl and gnash their teeth at this, asking why the press didn't fact-check or challenge those experts. But I'm not sure I want to institute a custom of journalists challenging experts. Journalist Johann Hari decided to take it upon himself to challenge psychiatric experts, and wrote a serious of terrible articles and a terrible book saying they were wrong about everything. I am a psychiatrist and I can tell you he is so wrong that it is physically painful to read his stuff (though of course I would say that…). Most journalists stick to assuming the experts know more about their subject of expertise than they do, and I think this is wise. The role of science journalists is to primarily to relay, explain, give context to the opinions of experts, not to try to out-medicine the doctors. So I think this is a good excuse.

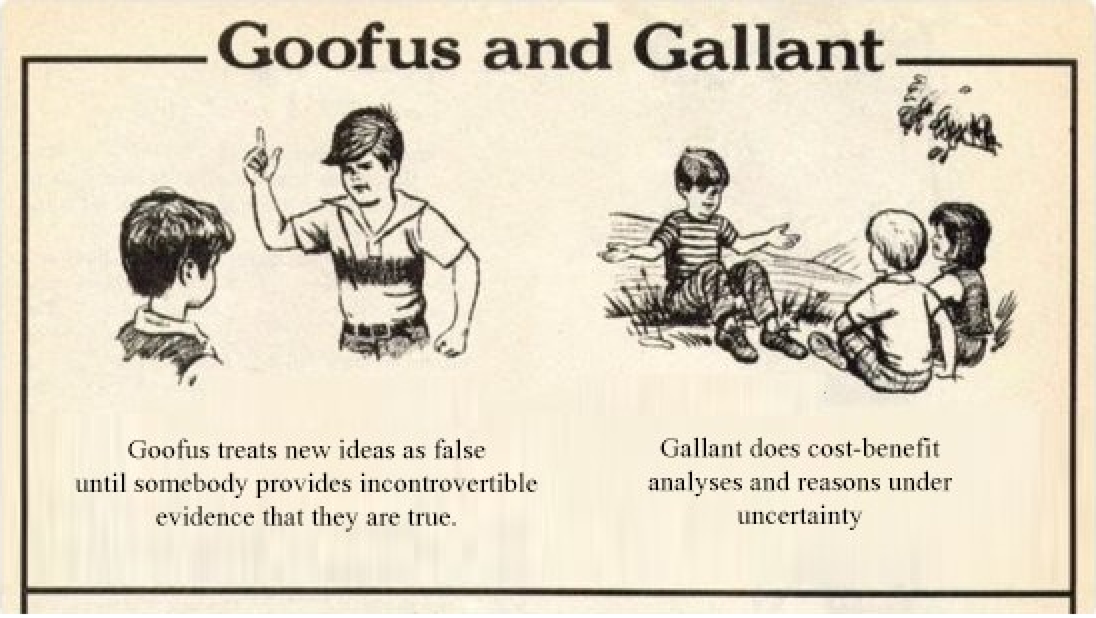
But I would ask this of any journalist who pleads that they were just relaying and providing context for expert opinions: what was the experts' percent confidence in their position?

I am so serious about this. What fact could possibly be more relevant? What context could it possibly be more important to give? I'm not saying you need to have put a number in your articles, maybe your readers don't go for that. But were you working off of one? Did this question even occur to you?

Nate Silver said there was a 29% chance Trump would win. Most people interpreted that as "Trump probably won't win" and got shocked when he did. What was the percent attached to your "coronavirus probably won't be a disaster" prediction? Was it also 29%? 20%? 10%? Are you sure you want to go lower than 10%? Wuhan was already under total lockdown, they didn't even have space to bury all the bodies, and you’re saying that there was less than 10% odds that it would be a problem anywhere else? I hear people say there’s a 12 – 15% chance that future civilizations will resurrect your frozen brain, surely the risk of coronavirus was higher than that?

And if the risk was 10%, shouldn’t that have been the headline. “TEN PERCENT CHANCE THAT THERE IS ABOUT TO BE A PANDEMIC THAT DEVASTATES THE GLOBAL ECONOMY, KILLS HUNDREDS OF THOUSANDS OF PEOPLE, AND PREVENTS YOU FROM LEAVING YOUR HOUSE FOR MONTHS”? Isn’t that a better headline than [Coronavirus panic sells as alarmist information spreads on social media]? But that’s the headline you could have written if your odds were ten percent!

So:


Goofus and Gallant
Goofus treats new ideas as false until somebody provides incontrovertible evidence that they are true.
Gallant does cost-benefit analyses and reasons under uncertainty

I think people acted like Goofus again.

People were presented with a new idea: a global pandemic might arise and change everything. They waited for proof. The proof didn’t arise, at least at first. I remember hearing people say things like “there’s no reason for panic, there are currently only ten cases in the US”. This should sound like “there’s no reason to panic, the asteroid heading for Earth is still several weeks away”. The only way I can make sense of it is through a mindset where you are not allowed to entertain an idea until you have proof of it. Nobody had incontrovertible evidence that coronavirus was going to be a disaster, so until someone does, you default to the null hypothesis that it won’t be.

Gallant wouldn’t have waited for proof. He would have checked prediction markets and asked top experts for probabilistic judgments. If he heard numbers like 10 or 20 percent, he would have done a cost-benefit analysis and found that putting some tough measures into place, like quarantine and social distancing, would be worthwhile if they had a 10 or 20 percent chance of averting catastrophe.

V.

This is at risk of getting too depressing, so I want to focus on some people who deserve recognition for especially good responses.

First, a bunch of generic smart people on Twitter who got things exactly right – there are too many of these people to name, but Scott Aaronson highlights “Bill Gates, Balaji Srinivasan, Paul Graham, Greg Cochran, Robin Hanson, Sarah Constantin, Eliezer Yudkowsky, and Nicholas Christakis.” None of these people (except Greg Cochran) are domain experts, and none of them (except Greg Cochran) have creepy oracular powers. So how could they have beaten the experts? Haven’t we been told a million times that generic intelligence is no match for deep domain knowledge?

I think the answer is: they didn’t beat the experts in epidemiology. Whatever probability of pandemic the experts and prediction markets gave for coronavirus getting really bad, these people didn’t necessarily give a higher probability. They were just better at probabilistic reasoning, so they had different reactions to the same number. There’s no reason generic why smart people shouldn’t be better at probabilistic reasoning then epidemiologists. In fact, this seems exactly like the sort of thing generic smart people might be.

Zeynep Tufekci is an even clearer example. She’s a sociologist and journalist who was writing about how it was “our civic duty” to prepare for coronavirus as early as February. She was also the first mainstream media figure to spread the word that masks were probably helpful.

Totally at random today, reading a blog post on the Mongol Empire like all normal people do during a crisis, I stumbled across a different reference to Zeynep. In a 2014 article, she was sounding a warning about the Ebola pandemic that was going on at the time. She was saying the exact same things everyone is saying now – global institutions are failing, nobody understands exponential growth, travel restrictions could work early but won’t be enough if it breaks out. She quoted a CDC prediction that there could be a million cases by the end of 2014. “Let that sink in,” she wrote. “A million Ebola victims in just a few months.”

In fact, this didn’t happen. There were only about 30,000 cases. The virus never really made it out of Liberia, Sierra Leone, and Guinea.

I don’t count this as a failed prediction on Zeynep’s part. First of all, because it could have been precisely because of people like her sounding the alarm that the epidemic was successfully contained. But more important, it wasn't really a prediction at all. Her point wasn't that she definitely knew this Ebola pandemic was the one that would be really bad. Her point was that it might be, so we needed to prepare. She said the same thing when the coronavirus was just starting. If this were a game, her batting average would be 50%, but that's the wrong framework.

Zeynep Tufecki is admirable. But her admirable skill isn't looking at various epidemics and successfully predicting which ones will be bad and which ones will fizzle out. She can't do that any better than anyone else. Her superpower is her ability to treat something as important even before she has incontrovertible evidence that it has to be.

And finally, Kelsey Piper. She wrote a February 6th article saying:

> The coronavirus killed fewer people than the flu did in January. But it might kill more in February — and unlike the flu, its scope and effects are poorly understood and hard to guess at. The Chinese National Health Commission reports 24,324 cases, including 3,887 new ones today. There are some indications that these numbers understate the situation, as overwhelmed hospitals in Wuhan only have the resources to test the most severe cases. As of Tuesday, 171,329 people are under medical observation because they've had close contact with a confirmed case.
>
> It is unclear whether China will be able to get the outbreak under control or whether it will cause a series of epidemics throughout the country. It's also unclear whether other countries — especially those with weak health systems — will be able to quickly identify any cases in their country and avoid Wuhan-scale outbreaks.

> The point is, it’s simply too soon to assert we’ll do well on both those fronts — and if we fail, then the coronavirus death toll could well climb up into the tens of thousands. It also remains to be seen if vaccines or effective antiviral treatments will be developed. That’s just far too much uncertainty to assure people that they have nothing to worry about. And misleadingly assuring people that there’s nothing to worry about can end up doing harm.
>
> “Instead of deriding people’s fears about the Wuhan coronavirus,” Sandman, the communications expert, writes, “I would advise officials and reporters to focus more on the high likelihood that things will get worse and the not-so-small possibility that they will get much worse.”

She concluded that “the Wuhan coronavirus likely won’t be a nightmare pandemic, but that scenario is still in play”, and followed it up with an article urging people to prepare by buying essential food and supplies.

If we interpret her “likely won’t be a nightmare pandemic” sentence as a prediction, she got the prediction wrong. Like Zeynep, she has no special ability to predict whether any given disease will end in global disaster. But that didn’t matter! She gave exactly the correct advice to institutions (prepare for a worst-case scenario, stop telling people not to panic) and exactly the correct advice to individuals (start prepping). When you’re good enough at handling uncertainty, getting your predictions exactly right becomes almost superfluous.

The Vox article says the media needs to “say what it doesn’t know”. I agree with this up to a point. But they can’t let this turn into a muddled message of “oh, who knows anything, whatever". Uncertainty about the world doesn't imply uncertainty about the best course of action! Within the range of uncertainty that we had about the coronavirus this February, an article that acknowledged that uncertainty wouldn't have looked like "We're not sure how this will develop, so we don't know whether you should stop large gatherings or not". It would have looked like "We're not sure how this will develop, so you should *definitely* stop large gatherings."

I worry that the people who refused to worry about coronavirus until too late thought they were "being careful" and "avoiding overconfidence". And I worry the lesson they'll take away from this is to be more careful, and avoid overconfidence even more strongly.

Experts should think along these lines when making their recommendations, but if they don't, the press should think along them as part of its work of putting expert recommendations in context. I think Kelsey's article provides an shining example of what this should look like.

Maybe other people got this right too. I'm singling out Kelsey because of a personal connection – I met her through the same probabilistic reasoning forum where I did my cryonics survey years ago. I don't think this is a coincidence.

*[**Related:** Book Review: The Precipice; Two Kinds Of Caution]*

Depression: The Olfactory Perspective

Depressed people have worse sense of smell, and people with worse sense of smell are more likely to get depressed. Kohli 2016 tries to figure out what's going on.

They review six studies testing how well depressed people can smell things. Most use something called "The Sniffin' Sticks Test" (really!) where people are asked to say which of two sticks has an odor; the strength of the odorous one is then decreased until the subject can no longer consistently get it right. This determines olfactory threshold – how sensitive the subject's smell is. Depressed subjects did marginally (but significantly) worse on this test than controls (6.31 ± 1.38 vs. 6.78 ± 0.88; P = 0.0005) – I think this corresponds to an effect size of about 0.2. They also do a couple more tests to see if depressed people are worse at identifying odors and get similarly small results. Also, some neuroimaging studies directly correlate depression and olfactory bulb volume, and find that olfactory areas of depressed people's brains shrink.

Next they look at three studies of people who have various known respiratory diseases that impair sense of smell, and see how many of them are depressed. The answer: lots! Normosmics (people with a normal sense of smell) have an average score of 5 on the Beck Depression Inventory. Anosmics (people with no sense of smell) have an average score of 14. Hyposmics (people with weak sense of smell) are in between. This seemed true independent of respiratory disease level (ie it's not just that having a worse respiratory disease makes your smell worse *and* makes you more depressed).

The two most common diseases studied were chronic rhinosinusitis and post-upper respiratory infection olfactory dysfunction. One of the studies, Jung, Lee, & Park (2014), compares them. They find that 40% of CRS patients are depressed, vs. 76% of PURIOD patients, a significant difference (also, both much higher than the ~5-10% of depressives in the general population). They note that CRS involves a gradual loss of smell, and PURIOD a sudden loss of smell, and that maybe people adjust to gradual loss better than sudden.

We can easily come up with alternative hypotheses. You can never actually control for confounders properly, and having a respiratory disease sounds depressing. Also, the sorts of people who have respiratory diseases probably differ from the rest of the population in various ways. Biologically, they might have worse immune systems, or end up with worse oxygenation and chronic infections, or just be generally sickly. Sociologically, they might be poorer, or have worse diets, or more comorbidities. Sinus problems themselves are linked to depression for a bunch of reasons, probably relating to the sinus being so closely linked to the brain that sinus inflammation screws up your brain chemistry. Also, most chronic diseases have high depression rates – sickle cell anemia (chosen kind of at random) is 44%.

There really aren’t that many causes of anosmia that are 100% exogenous and have no chance at all of contributing to depression themselves. But animal researchers can take Gandhi’s advice and become the exogenous cause they wish to see in the world. Yuan and Slotnick (2014) discuss experiments where researchers remove the olfactory bulbs of rats. These rats tend to consistently become depressed. That seems like pretty strong evidence to me.

So why would depression reduce olfactory acuity? And why would reduced olfaction cause depression?

Maybe depression reduces olfactory acuity because it reduces sensory acuity in general. We already know that depression decreases visual contrast, causing the world to literally look washed-out and gray. If I'm reading this study right, it claims depression also decreases auditory threshold – ie depressed people are less able to hear very quiet sounds. I can't find similarly good studies on taste or touch, although they should be easy to do. If anyone knows any studies on eg two-point discrimination in depressed vs. non-depressed subjects, let me know. If not, this would be an easy project that a college psychology student could do in a few weeks, and it would add to our understanding of this subject. But right now I think the evidence is consistent with a general decline in perceptual abilities. This fits my understanding of depression as a systemic disorder – the low mood is the most noticeable symptom, but you're also getting everything from poor digestion to worse hearing. Probably this is because whatever is affecting the emotional centers of the brain is affecting the lower parts of the brain (and nervous system?) too.

Is the opposite of this true? Does any form of decreased sensory acuity cause depression? I'm leaning toward no. This study claims to find that myopia is linked to depression, but the association is so weak that I take it as a strike against the hypothesis. These two studies find hearing loss is associated with depression, but it's still pretty weak and it could easily be because poor hearing hurts your opportunities to socialize. I'm most struck by discussion of monochromacy, ie total colorblindness, which absolutely fail to mention depression at all. If I'm going to be splitting hairs about how depressed people see slightly duller colors, the lack of any excess depression in people without color vision seems pretty important. Blind and deaf people have pretty high rates of depression, but being blind or deaf is really hard and I don't want to draw too many conclusions there. Overall it doesn't seem to me like decreased sensory acuity causes depression full stop, though I could be wrong. This would mean that loss of smell is unique in its emotional effects.

The sense of smell is pretty emotionally salient. Everyone always says that "smell is the sense most linked to memory" – though I can't figure out exactly which study discovered it and whether it considered (for example) how reliably seeing a picture of my mother reminds me of my mother. Body odor seems to be closely linked to who we're attracted to. Smell is responsible for all taste sensation beyond sweet/salty/sour/bitter/umami, and eating food is one of life's most visceral pleasures. From fifthsense.org.uk:

> Anosmia sufferers often talk of feeling isolated and cut-off from the world around them, and experiencing a 'blunting' of the emotions. Smell loss can affect one's ability to form and maintain close personal relationships and can lead to depression. An important issue here is the fact that smell loss is invisible to all but the patient; how would you know that you had met an anosmia sufferer unless they themselves told you? This is one of the reasons, alongside the general lack of understanding of the impact that smell has on our lives, why anosmia has never received much attention – you really do not know what you have got until it is gone.

But also, Yuan and Slotnick's rat paper gives a more biological explanation. The olfactory bulb is the beginning of pathways that stimulate many other parts of the brain. When it's removed (and presumably also when it just never gets any incoming stimuli) it stops doing that, and the downstream parts of the brain shrink. For some reason this also causes decreased brain-wide synthesis of serotonin, maybe because the olfactory bulb is a net positive stimulus on the raphe nuclei. Sure, sounds like the kind of thing that might cause depression.

What does this imply about treating depression?

SSRI antidepressants probably decrease sense of smell as an immediate side effect. Awkward. But they seem to improve sense of smell long term as part of their general treatment of depression. In studies fluoxetine (Prozac) causes nerve growth in the olfactory bulb after a few weeks. Does that suggest a story where SSRIs work by improving smell? Probably not – more likely they work by [a cascading system of effects involving] causing nerve growth more generally, and the olfactory bulb benefits along with everything else.

Could you treat depression through improving olfactory sensitivity? Maybe, but I have no idea how to do that. Yuan and Slotnick suggest directly stimulating the olfactory regions of the brain, but this is pretty invasive, and there are probably already better treatments for the small minority of patients who are going to let you directly stimulate brain regions. Maybe before we worry about this problem, we should investigate the more general question of sensory enhancement for depression. What would happen if you made people wear glasses that enhanced the color saturation of everything they saw?

Finally, what about just exposing depressed people to really strong smells? You will be excited to know that real scientists have studied this ridiculous idea, and it seems promising, at least in extremely sketchy experiments on mice (1, 2, 3). I have never seen any studies done on humans *unless* this is actually how “aromatherapy” works, which would be hilarious. Aromatherapy seems to get positive results for depression with the same kind of bad studies that let all quackery generate positive results for everything. I can’t say more than that and I’m pretty skeptical here.

The Amish Health Care System

I.

Amish people spend only a fifth as much as you do on health care, and their health is fine. What can we learn from them?

A reminder: the Amish are a German religious sect who immigrated to colonial America. Most of them live apart from ordinary Americans (who they call “the English”) in rural communities in Pennsylvania and Ohio. They’re famous for their low-tech way of life, generally avoiding anything invented after the 1700s. But this isn’t absolute; they are willing to accept technology they see as a net positive. Modern medicine is in this category. When the Amish get seriously ill, they will go to modern doctors and accept modern treatments.

The Muslims claim Mohammed was the last of the prophets, and that after his death God stopped advising earthly religions. But sometimes modern faiths will make a decision so inspired that it could only have come from divine revelation. This is how I feel about the Amish belief that health insurance companies are evil, and that good Christians must have no traffic with them.

And Deists believe that God is like a watchmaker, an artisan who built the world but does not act upon it. But by some miracle, the US government played along and granted the Amish exemptions from all the usual health care laws. They don’t have to pay Medicare taxes or social security. They aren’t included in the Obamacare mandate. They can share health care costs the way they want, ignoring any regulations to the contrary. They are genuinely on their own.

They've ended up with a simple system based on church aid. Everyone pays tithes to their congregation (though they don't call it that). The churches meet in houses and have volunteer leaders, so expenses minimal. Most of the money goes to "alms" which the bishop distributes to members in need. This replaces the social safety net, including health insurance. Most Amish go their entire life without needing anything else.

About a third of Amish are part of a more formal insurance-like institution called Amish Hospital Aid. Individuals and families pay a fixed fee to the organization, which is not-for-profit and run by an unpaid board of all-male elders. If they need hospital care, AHA will pay for it. How does this interact with the church-based system? Rohrer and Dundes, my source for most of this post, say that it's mostly better-off Amish who use AHA. Their wealth is tied up in their farmland, so it's not like they can use it to pay hospital bills. But they would feel guilty asking their church to give them alms meant for the poor. AHA helps protect their dignity and keep church funds for those who need them most.

How well does this system work?

The Amish outperform the English on every measured health outcome. 65% of Amish rate their health as excellent or very good, compared to 58% of English. Diabetes rates are 2% vs. 8%, heart attack rates are 1% vs. 6%, high blood pressure is 11% vs. 31%. Amish people go to the hospital about a quarter as often as English people, and this difference is consistent across various categories of illness (the big exception is pregnancy-related issues – most Amish women have five to ten children). This is noticeable enough that lots of health magazines have articles on The Health Secrets of the Amish and Amish Secrets That Will Add Years To Your Life. As far as I can tell, most of the secret is spending your whole life outside doing strenuous agricultural labor, plus being at a tech level two centuries too early for fast food.

But Amish people also die earlier. Lots of old studies say the opposite – for example, this one finds Amish people live longer than matched Framingham Heart Study participants. But things have changed since Framingham. The Amish have had a life expectancy in the low 70s since colonial times, when the rest of us were dying at 40 or 50. Since then, Amish life expectancy has stayed the same, and English life expectancy has improved to the high 70s. The most recent Amish estimates I have still say low seventies, so I think we are beating them now.

If they're healthier, why is their life expectancy lower? Possibly they are less interested in prolonging life than we are. R&D write:

> Amish people are more willing to stop interventions earlier and resist invasive therapies than the general population because, while they long for healing, they also have a profound respect for God's will. This means taking modest steps toward healing sick bodies, giving preference to natural remedies, setting common-sense limits, and believing that in the end their bodies are in God's hands.

The Amish health care system has an easier job than ours does. It has to take care of people who are generally healthy and less interested in extreme end-of-life care. It also supports a younger population – because Amish families have five to ten children, the demographics are weighted to younger people. All of these make its job a little bit simpler, and we should keep that in mind for the following sections.

How much do the Amish pay for health care? This is easy to answer for Amish Hospital Aid, much harder for the church system.

Amish Hospital Aid charges $125 monthly per individual or $250 monthly per family (remember, Amish families can easily be ten people). Average US health insurance costs $411 monthly per individual (Obamacare policies) or $558 monthly per individual (employer sponsored plan; employers pay most of this). I'm not going to bother comparing family plans because the definition of "family" matters a lot here. On the surface, it looks like the English spend about 4x as much as the Amish do.

But US plans include many more services than AHA, which covers catastrophic hospital admissions only. The government bans most Americans from buying plans like this; they believe it's not enough to count as real coverage. The cheapest legal US health plan varies by age and location, but when I take my real age and pretend that I live near Amish country, the government offers me a $219/month policy on Obamacare. This is only a little higher than what the Amish get, and probably includes more services. So here it seems like the Amish don't have much of an efficiency advantage. They just make a different tradeoff. It's probably the right tradeoff for them, given their healthier lifestyle.

But remember, only a third of Amish use AHA. The rest use a church-based system? How does that come out?

It's hard to tell. Nobody agrees on how much Amish tithe their churches, maybe because different Amish churches have different practices. R&D suggest families tithe 10% of income, this article on church-based insurances says a flat $100/month fee, and this "Ask The Amish" column says that churches have twice-yearly occasions where they ask for donations in secret and nobody is obligated to give any particular amount (“often husbands and wives won’t even know how much the other is giving.”) So it’s a mess, and even knowing the exact per-Amish donation wouldn’t help, because church alms cover not just health insurance but the entire social safety net; the amount that goes to health care probably varies by congregation and circumstance.

A few people try to estimate Amish health spending directly. This ABC story says $5 million total for all 30,000 Amish in Lancaster County, but they give no source, and it’s absurdly low. This QZ story quotes Amish health elder Marvin Wengerd as saying $20 – $30 million total for Lancaster County, which would suggest health spending of between $600-$1000 per person. This sounds potentially in keeping with some of the other estimates. A $100 per month tithe would be $1200 per year – if half of that goes to non-health social services, that implies $600 for health. The average Amish family earns about $50K (the same as the average English family, somehow!) so a 10% tithe would be $5000 per year, but since the average Amish family size is seven children, that comes out to about $600 per person again. So several estimates seem to agree on between $600 and $1000 per person.

One possible issue with this number: does Wengerd know how much Amish spend out of pocket? Or does his number just represent the amount that the official communal Amish health system spends? I’m not sure, but taking his words literally it’s total Amish spending, so I am going to assume it’s the intended meaning. And since the Amish rarely see doctors for minor things, probably their communal spending is a big chunk of their total.

[**Update**: an SSC reader is able to contact his brother, a Mennonite deacon, for better numbers. He says that their church spends an average of $2000 per person (including out of pocket).]

How does this compare to the US as a whole? The National Center For Health Statistics says that the average American spends $11,000 on health care. This suggests that the average American spends between five and ten times more on health care than the average Amish person.

How do the Amish keep costs so low? R&D (plus a few other sources) identify some key strategies.

First, the Amish community bargains collectively with providers to keep prices low. This isn't unusual – your insurance company does the same – but it nets them better prices than you would get if you tried to pay out of pocket at your local hospital. This article gives some examples of Amish getting sticker prices discounted from between 50% to 66% with this tactic alone; Medicare gets about the same.

Second, the Amish are honorable customers. This separates them from insurance companies, who are constantly trying to scam providers however they can. Much of the increase in health care costs is "administrative expenses", and much of these administrative expenses is hiring an army of lawyers, clerks, and billing professionals to thwart insurance companies' attempts to cheat their way out of paying. If you are an honorable Amish person and the hospital knows you will pay your bill on time with zero fuss, they can waive all this.

But can this really be the reason Amish healthcare is cheaper? When insurance companies negotiate with providers, patients are on the side of the insurances; when insurance companies get good deals (eg a deal of zero dollars because the insurance has scammed the hospital), the patient’s care is cheaper, and the insurance company can pass some of those savings down as lower prices. If occasionally scamming providers meant insurance companies had to pay more money total, then they would stop doing it. My impression is that the real losers here are uninsured patients; absent any pressure to do otherwise, hospitals will charge them the sticker price, which includes the dealing-with-insurance-scams fee. The Amish successfully pressure them to waive that fee, which gets them better prices than the average uninsured patient, but still doesn’t land them ahead of insured people.

Third, Amish don’t go to the doctor for little things. They either use folk medicine or chiropractors. Some of the folk medicine probably works. The chiropractors probably don’t, but they play a helpful role reassuring people and giving them the appropriate obvious advice while telling the really serious cases to seek outside care. With this help, Amish people mostly avoid primary care doctors. Holmes County health statistics find that only 16% of Amish have seen a doctor in the past year, compared to 54% of English.

Fourth, the Amish never sue doctors. Doctors around Amish country know this, and give them the medically indicated level of care instead of practicing “defensive medicine”. If Amish people ask their doctors to be financially considerate – for example, let them leave the hospital a little early – their doctors will usually say yes, whereas *your* doctor would say no because you could sue them if anything went wrong. In some cases, Amish elders formally promise that no member of their congregation will ever launch a malpractice lawsuit.

Fifth, the Amish don’t make a profit. Church aid is dispensed by ministers and bishops. Even Amish Hospital Aid is run by a volunteer board. None of these people draw a salary or take a cut. I don’t want to overemphasize this one – people constantly obsess over insurance company profits and attribute all health care pathologies to them, whereas in fact they’re a low single-digit percent of costs (did you know Kaiser Permanente is a nonprofit? Hard to tell, isn’t it?) But every little bit adds up, and this is one bit.

Sixth, the Amish don’t have administrative expenses. Since the minister knows and trusts everyone in his congregation, the “approval process” is just telling your minister what the problem is, and the minister agreeing that’s a problem and giving you money to solve it. This sidesteps a lot of horrible algorithms and review boards and appeal boards and lawyers. I don’t want to overemphasize this one either – insurance companies are legally required to keep administrative expenses low, and most of them succeed. But again, it all adds up.

Seventh, the Amish feel pressure to avoid taking risks with their health. If you live in a tiny community with the people who are your health insurance support system, you’re going to feel awkward smoking or drinking too much. Realistically this probably blends into a general insistence on godly living, but the health insurance aspect doesn’t hurt. And I’m talking like this is just informal pressure, but occasionally it can get very real. R&D discuss the case of some Amish teens who get injured riding a snowmobile – forbidden technology. Their church decided this was not the sort of problem that godly people would have gotten themselves into, and refused to help – their families were on the hook for the whole bill.

Eighth, for the same reason, Amish try not to overspend on health care. I realize this sounds insulting – other Americans aren't trying? I think this is harsh but true. Lots of Americans get an insurance plan from their employer, and then consume health services in a price-insensitive way, knowing very well that their insurance will pay for it. Sometimes they will briefly be limited by deductibles or out-of-pocket charges, but after these are used up, they'll go crazy. You wouldn't believe how many patients I see who say things like "I've covered my deductible for the year, so you might as well give me the most expensive thing you've got", or "I'm actually feeling fine, but let's have another appointment next week because I like talking to you and my out-of-pocket charges are low."

But it's not just avoiding the obvious failure modes. Careful price-shopping can look very different from regular medical consumption. Several of the articles I read talked about Amish families traveling from Pennsylvania to Tijuana for medical treatment. One writer describes Tijuana clinics sending salespeople up to Amish Country to advertise their latest prices and services. For people who rarely leave their hometown and avoid modern technology, a train trip to Mexico must be a scary experience. But prices in Mexico are cheap enough to make it worthwhile.

Meanwhile, back in the modern world, I've written before about how a pharma company took clonidine, a workhorse older drug that costs $4.84 a month, transformed it into Lucemyra, a basically identical drug that costs $1,974.78 a month, then created a rebate plan so that patients wouldn't have to pay any extra out-of-pocket. Then they told patients to ask their doctors for Lucemyra because it was newer and cooler. Patients sometimes went along with this, being indifferent between spending $4 of someone else's money or $2000 of someone else's money. Everything in the US health system is like this, and the Amish avoid all of it. They have a normal free market in medical care where people pay for a product with their own money (or their community’s money) and have incentives to check how much it costs before they buy it. I *do* want to over-emphasize this one, and honestly I am surprised Amish health care costs are *only* ten times cheaper than ours are.

I don’t know how important each of these factors is, or how they compare to more structural factors like younger populations, healthier lifestyles, and less end-of-life care. But taken together, they make it possible for the Amish to get health care without undue financial burden or government support.

II.

Why look into the Amish health system?

I’m fascinated by how many of today’s biggest economic problems just mysteriously failed to exist in the past. Our grandparents easily paid for college with summer jobs, raised three or four kids on a single income, and bought houses in their 20s or 30s and never worried about rent or eviction again. And yes, they got medical care without health insurance, and avoided the kind of medical bankruptcies we see too frequently today. How did this work so well? Are there ways to make it work today? The Amish are an extreme example of people who try to make traditional systems work in the modern world, which makes them a natural laboratory for this kind of question.

The Amish system seems to work well for the Amish. It’s hard to say this with confidence because of all the uncertainties. The Amish skew much younger than the “English”, and live much healthier lifestyles. Although a few vague estimates suggest health care spending far below the English average, they could be missing lots of under-the-table transactions. And again, I don’t want to ignore the fact that the Amish do live a little bit shorter lives. You could tell a story where all of these add up to explain 100% of the difference, and the Amish aren’t any more efficient in their spending at all. I don’t think this is right. I think the apparent 5x advantage, or something like it, is real. But right now this is just a guess, not a hard number.

What if it is? It’s hard to figure out exactly what it would take to apply the same principles to English society. Only about a quarter of Americans attend church regularly, so church-based aid is out. In theory, health insurance companies ought to fill the same niche, with maybe a 10% cost increase for profits and overhead. Instead we have a 1000% cost increase. Why?

Above, I said that the most important factor is that the Amish comparison shop. Everyone needs to use other people’s money to afford expensive procedures. But for the Amish, those other people are their fellow church members and they feel an obligation to spend it wisely. For the English, the “other people” are faceless insurance companies, and we treat people who don’t extract as much money as possible from them as insufficiently savvy. But there’s no easy way to solve this in an atomized system. If you don’t have a set of thirty close friends you can turn to for financial help, then the only institutions with enough coordination power to make risk pooling work are companies and the government. And they have no way of keeping you honest except the with byzantine rules about “prior authorizations” and “preferred alternatives” we’ve become all too familiar with.

(and as bad as these are, there’s something to be said for a faceless but impartial bureaucracy, compared to having all your neighbors judging your lifestyle all the time.)

This is a neat story, but I have two concerns about it.

First, when I think in terms of individual people I know who have had trouble paying for health care, it’s hard for me to imagine the Amish system working very well for them. Many have chronic diseases. Some have mysterious pain that they couldn’t identify for years before finally getting diagnosed with something obscure. Amish Hospital Aid’s catastrophic policy would be useless for this, and I feel like your fellow church members would get tired of you pretty quickly. I’m not sure how the Amish cope with this kind of thing, and maybe their system relies on a very low rate of mental illness and chronic disease. A lot of the original “hygiene hypothesis” work was done on the Amish, their autoimmune disease rates are amazing, and when you take out the stresses of modern life maybe a lot of the ailments the American system was set up to deal with just stop being problems. I guess my point is that the numbers seem to work out, and the Amish apparently remain alive, but when I imagine trying to apply the Amish system to real people, *even* assuming those real people have cooperative churches and all the other elements I’ve talked about, I can’t imagine it doing anything other than crashing and burning.

Second, I don’t think this is *actually* how our grandparents did things. I asked my literal grandmother, a 95 year old former nurse, how health care worked in her day. She said it just wasn’t a problem. Hospitals were supported by wealthy philanthropists and religious organizations. Poor people got treated for free. Middle class people paid as much as they could afford, which was often the whole bill, because bills were cheap. Rich people paid extra for fancy hospital suites and helped subsidize everyone else. Although most people went to church or synagogue, there wasn’t the same kind of Amish-style risk pooling.

This makes me think that the Amish method, even though it works, isn’t the method that worked for past generations. It’s an innovation intended to cover for health care prices being higher than anything that traditional societies had to deal with.

Why did health care prices start rising? I’ve wondered about this a lot before – see here, here, and here. Looking into this issue, I noticed glimpses of a different possibility. The increase started around the same time that health insurance began to spread. In one sense, this is unsurprising – of course health insurance would become a thing around the time care became unaffordable. But I’ve never seen someone really try to tease out causality here. Might the two trends have been mutually self-reinforcing? The price of care rises due to some original shock. Someone invents health insurance, which seems like a good idea. But this creates a series of perverse incentives, which other actors figure out how to exploit (eg the Lucemyra example above). Insurance-based-health-care becomes less efficient, but hospitals can’t or don’t internalize this to the insured patients – they just raise the price for everyone, insurance or no. That makes even more people need health insurance, and the cycle repeats as prices grow higher and higher and insurance becomes more and more necessary. This syncs well with some explanations I’ve heard of rising college prices, where once the government made easy loans and subsidies available to everyone, prices rose until they consumed all the resources available.

I have no idea if this is true or not. If it is, the Amish succeed partly by successfully forcing providers to internalize the costs of insurance to insurance patients. Sometimes they do this by literally asking hospitals for better prices because they are not insured (eg the “honest customer” example above). Other times they flee the country entirely to reach a medical system that doesn’t deal with insured patients (eg Tijuana). This seems to work well for them. But their reliance on church alms and Amish Hospital Aid suggests that their care is still more expensive and burdensome for them than past generations’ care was for them. They’ve just learned ways to manage the expense successfully.

Employer Provided Health Insurance Delenda Est

My last post didn't really go into why I dislike the way we do health insurance so much.

Of course, there are the usual criticisms based on compassion and efficiency. Compassion because poor people can't get access to life-saving medical care. Efficiency because it's ruinously expensive compared to every other system around. I agree with these arguments. And they're strong enough that asking whether there are any *other* reasons is kind of like the proverbial "But besides that, Mrs. Lincoln, how did you like the play?"

But I had already internalized the compassion and efficiency critiques before becoming a doctor. After starting work, I encountered new problems I never would have expected, ones which have yet to fade into the amorphous cloud of injustices we all know about and mostly ignore. Most of my patients have insurance; most of them are well-off; most of them are intelligent enough that they should be able to navigate the bureaucracy. Listen to the usual debate around insurance, and you would expect them to be the winners of our system; the rich people who can turn their financial advantage into better care. And yet barely a day goes by without a reminder that it doesn't work this way.

Here are some people I have encountered – some of them patients, some of them friends – who have made me skeptical that our system works for anyone at all:

— The elderly man who had a great relationship with his last psychiatrist, who saw him for twenty years, and who knew every detail of his issues. He switched jobs, got a new insurance, the old psychiatrist was no longer in network, and so he had to see me instead. I know nothing about him and it will take several evaluation sessions before I can even consistently remember who he is and what he needs from me.

— The businesswoman who was seeing me and doing well until the HR person at her job told her that she didn’t need to submit any forms to renew her insurance that year. That turned out to be wrong, and she missed The One Month Of The Year When You Are Allowed To Renew Insurance. She lost her insurance and can’t afford to keep seeing me.

— The bipolar man on a very important daily medication. He changed insurance plans. The new insurance refused to pay for his drugs until they got a form explaining why he needed the medication. I sent in the form. They said they couldn’t find the patient in their system and so couldn’t process the form. We argued about this for several days, during which he ran out of medication and decompensated.

— The would-be entrepreneur who wanted to save up enough money to live on for a year or two, quit her dead-end job, and start a startup – but who wouldn’t be able to afford health insurance outside of her dead-end job’s plan. She is still at her dead-end job.

— The young gay man with conservative religious parents. He had a supportive friend group and should have been able to come out to his parents without caring what they thought – except that he’s on his parents’ insurance and has no good alternative. He remains in the closet.

— The young woman in a not-quite-abusive but far-from-acceptable marriage, who stays in it because she’s on her husband’s insurance and has no good alternative.

— The depressed guy who was doing well on a complicated antidepressant regimen for a while, changed insurances, and was too depressed to do the work of finding a new psychiatrist. Now he comes to me saying it’s been five years, he’s been depressed all that time, and he would like to get back on the medications that he knows work well for him.

— The endless train of patients I saw when I worked in a hospital emergency room, whose stories started with “So I lost my insurance, but I figured I could get along fine without the medication…”

— The other endless train of patients I saw when I worked in a hospital emergency room, whose stories started with “So I lost my insurance, but I was smart, and I had saved up a stockpile of my medication, and I figured, how hard could it be to manage it myself without a doctor’s advice…”

— The *other* endless train of patients I saw when I worked in a hospital emergency room, whose stories started with “So I changed insurance, and I got a new doctor, and he said he didn’t see a good reason why I had to be on this medication…”

— The Oklahoman who wants to move to California where he has more friends and better job prospects, but he’s on Oklahoma state insurance for the needy. Although California also has a state insurance for the needy, there is no way to figure out whether it will accept him or cover the care he needs other than by moving to California, applying, and seeing what happens. Also, the application process takes weeks to months, during which time he will not have either state’s insurance (though California promises he will get reimbursed for care he gets during this period later). He decides to stay in Oklahoma.

— The well-off Oklahoman with good private insurance who visits California on a business trip, gets sick, and finds that surprise! his Oklahoman insurance doesn't have any in-network providers in California and he has to pay $20,000 out-of-pocket for care.

— The depressed guy who was in remission for years and had a great job with great insurance. Then he had a relapse, became too depressed to go to work, got fired, and lost his insurance right at the moment when it finally could have been useful for him.

— The woman who had a minor breakdown and was brought to the hospital by police. The hospital admitted her against her will, on the grounds that her minor breakdown sounded potentially dangerous, treated her, and released her. Her insurance refused to pay, on the grounds that it was just a *minor* breakdown and didn't really require hospitalization, by the standards of This Particular Insurance Company. She is on the hook for the entire cost of the involuntary hospitalization.

— Anything involving Kaiser [EDIT: I'm getting some pushback on this who say Kaiser is good at everything except mental health; I only deal with them in a mental health capacity so I can't speak to this]

— The trauma victim who needs trauma therapy – which means reliving all your traumas in order to come to terms with them – but is reluctant to start because her work situation is unstable. She knows that if she changes insurance, she'll have to restart the therapy from the beginning and relive all her traumas all over again.

— The young schizophrenic man who is on his parents' insurance. Because so many schizophrenics are poor, all of the expertise in treating schizophrenia is concentrated in Medicaid clinics. But a rich kid on his parents' excellent insurance can't access Medicaid, he can't go off his parents' insurance for other reasons, and he can't find any private psychiatrists who know how to treat his particular schizophrenia-related complications. I am sure there is some good solution to this which I am missing, but I haven't been able to find it and his family hasn't either.

— The anorexic woman who has Blue Cross, and the only good anorexia therapist in town only takes Aetna. The sex-addicted man who has Aetna, and the only good sex addiction therapist in town only takes Blue Cross.

Any other system would fix these problems. A public system like Medicare For All would fix them. A communal system like the Amish have would fix them. A free market system like our grandparents had would fix them. The prepaid doctor cooperatives Reason talks about would fix them. A half-assed compromise like Joe Biden's Medicare For All Who Want It would fix them. But here we are, stuck with a system that somehow manages to fail everybody for different reasons.

Give Yourself Gout For Fame And Profit

I.

Actually, no. You should not do this. Most of you were probably already not doing this, and I support your decision. But if you want a 2000 word essay on some reasons to consider this, and then some other reasons why those reasons are wrong, keep reading.

Gout is a disease caused by high levels of uric acid in the blood. Everyone has some uric acid in their blood, but when you get too much, it can form little crystals that get deposited around your body and cause various problems, most commonly joint pain. Some uric acid comes from chemicals found in certain foods (especially meat), so the first step for a gout patient is to change their diet. If that doesn't work, they can take various chemicals that affect uric acid metabolism or prevent inflammation.

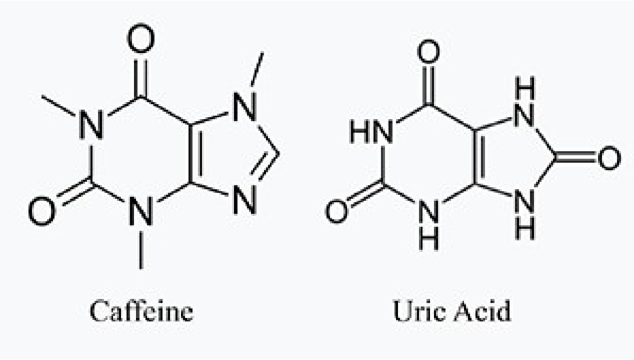
Gout is traditionally associated with kings, probably because they used to be the only people who ate enough meat to be affected. Veal, venison, duck, and beer are among the highest-risk foods; that list sounds a lot like a medieval king's dinner menu. But as kings faded from view, gout started affecting a new class of movers and shakers. King George III had gout, but so did many of his American enemies, including Franklin, Jefferson, and Hancock (beginning a long line of gout-stricken US politicians, most recently Bernie Sanders). Lists of other famous historical gout sufferers are contradictory and sometimes based on flimsy evidence, but frequently mentioned names include Alexander the Great, Charlemagne, Leonardo da Vinci, Martin Luther, John Milton, Isaac Newton, Ludwig von Beethoven, Karl Marx, Charles Dickens, and Mark Twain.

Question: isn't this just a list of every famous person ever? It sure seems that way, and even today gout seems to disproportionately strike the rich and powerful. In 1963, Dunn, Brooks, and Mausner published Social Class Gradient Of Serum Uric Acid Levels In Males, showing that in many different domains, the highest-ranking and most successful men had the highest uric acid (and so, presumably, the most gout). Executives have higher uric acid than blue-collar workers. College graduates have higher levels than dropouts. Good students have higher levels than bad students. Top professors have higher levels than mediocre professors. DB&M admitted rich people probably still eat more meat than poor people, but didn't think this explained the magnitude or universality of the effect. They proposed a different theory: maybe uric acid makes you more successful.

Before we mock them, let's take more of a look at why they might think that, and at the people who have tried to flesh out their theory over the years.

Most animals don't have uric acid in their blood. They use an enzyme called uricase to metabolize it into a harmless chemical called allantoin. About ten million years ago, the common ancestor of apes and humans got a mutation that broke uricase, causing uric acid levels to rise. The mutation spread very quickly, suggesting that evolution *really* wanted primates to have lots of uric acid for some reason. Since discovering this, scientists have been trying to figure out exactly what that reason was, with most people thinking it's probably an antioxidant or neuroprotectant or something else helpful if you're trying to evolve giant brains. Other researchers note that in lower animals, uric acid is a "come out of hibernation" sign which seems to induce energetic foraging and goal-directed behavior more generally.

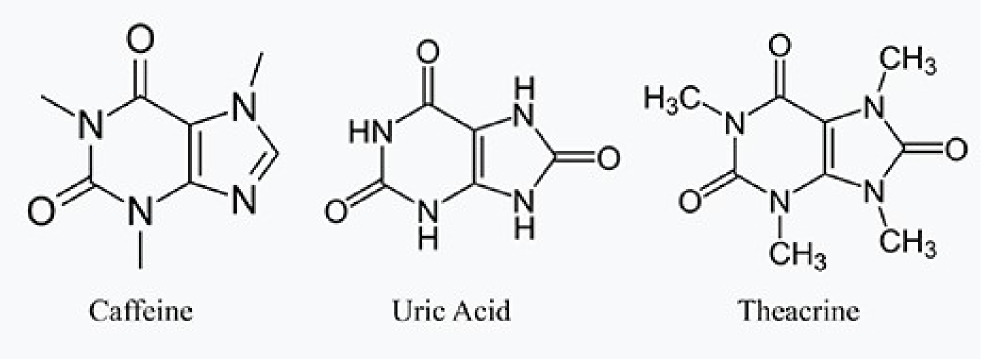
Some of these people note the similarity between uric acid and caffeine:



If uric acid had caffeine-like effects, then high levels of uric acid in the blood would be like being on a constant caffeine drip. The exact numbers don't really work out, but you can fix this by assuming uric acid is an order of magnitude or so weaker than straight caffeine. Add this fudge factor, and Benjamin Franklin was on exactly one espresso all the time.

But you can't *actually* be hyperproductive by being on one espresso all the time, can you? Don't you eventually gain tolerance to caffeine and lose any benefits?

Although uric acid is structurally similar to caffeine, it's even more similar to a chemical called theacrine. In fact, theacrine is just 1,3,7,9-tetramethyl-uric acid:



Theacrine (*not* the same as theanine, be careful with this one!) is a caffeine-like substance found in an unusual Chinese variety of tea plant. It's recently gained fame in the nootropic community for not producing tolerance the same way regular caffeine does – see eg Theacrine: Caffeine-Like Alkaloid Without Tolerance Build-Up. This makes the theory work even better:

Franklin (and other gout sufferers) were constantly on one espresso worth of magic no-tolerance caffeine. Seems plausible!

II.

This theory is hilarious, but is it true?

I was able to find eleven studies comparing achievement and uric acid levels. I've put them into a table below.

Study	sample size	finding	significant at	awfulness
Kasi	155 tenth-graders	r = 0.28 w/ test scores	0.001	significant
Bloch	84 med students	r = 0.23 w/ test scores	0.05	immense
Steaton & Herron	817 army recruits	r = 0.07 w/ test scores	0.02	significant
Mueller & French	114 professors	r = 0.5 with achievement-oriented behavior	0.01	astronomical
Montoye & Mikkelsen	467 high-schoolers	negative result	N/A	unclear
Cervini & Zampa	270 children	positive result	unknown	what even is this?
Inouye & Park	???	r = 0.33 with IQ	0.025	what even is this?
Anumonye	100 businessmen, 40 controls	r = 0.21 with drive	0.05	immense
Ooki	88 twins	r = 0.17 with 'rhathymia'	0.05	how is this even real?
Dunn I	58 executives	positive	???	immense

Dunn II	10 medical students	negative	N/A	astronomical

Nine out of eleven are positive. But I find it hard to be confident in any of them. Modern studies can be pretty bad, but studies from the 1960s ask you to take even more things on trust, while inspiring a lot less of it. Many of these studies were unable to find the outcomes that the others found, but discovered new outcomes of their own. Many failed to report basic pieces of information. The largest experiments usually found the least impressive results. Overall this looks a lot like you would expect from something forty years before anyone realized there was a replication crisis.

I also notice that the most positive studies compare business executives to people in other walks of life, and the least positive studies compare good students with bad students. Business executives get a lot of chances to differ from the general population – maybe they still eat more meat and richer food? Maybe they're stressed and stress affects uric acid levels?

What about the list of very famous people with gout? I agree it's a lot of people, but what's the base rate? Kings were born to their position, so we have no reason to think they were especially high achievers (someone in their family might have been, but that gene could have gotten pretty diluted). Since so many kings got gout, this suggests rich old people in the past had gout pretty often regardless of achievement. Also, this was before people invented good medical diagnosis, so probably arthritis, injuries, and any other form of joint pain got rounded off to gout too. What percent of rich old people in the past had some kind of joint pain? I'm prepared to guess "a lot".

The biochemists report equally confusing results around the uric acid / caffeine connection. Caffeine mostly works by antagonizing adenosine, a chemical involved in sleepiness. According to Hunter et al, Effects of uric acid and caffeine on A1 adenosine receptor binding in developing rat brain, uric acid does not affect adenosine, and so probably does not have a caffeine-like mechanism of action. On the other hand, caffeine probably has a small additional effect on catecholamine (eg dopamine, norepinephrine) release, and a different paper finds that uric acid does share this mechanism. So it doesn't have caffeine's main effect, but it does seem to have some kind of mild stimulant properties.

Given this level of uncertainty around every step in the hypothesis, I would describe any link between uric acid and achievement as kind of a stretch at this point. I feel bad about this, because it's an elegant theory with mostly positive studies in support, but I'm just not feeling like it's met its burden of proof.

III.

But some recent research is trying to bring this field back from the dead. At least this is what I get from Ortiz et al, Purinergic System Dysfunction In Mood Disorders, which synthesizes some more modern evidence that "uric acid and purines (such as adenosine) regulate mood, sleep, activity, appetite, cognition, memory, convulsive threshold, social interaction, drive, and impulsivity". It argues that we know there are neurorecptors for adenosine (another similar-looking molecule) and ATP (adenosine triphosphate, the body's main form of chemical energy). These seem to be involved in depression and mania, in the predicted direction (manic people have too much ATP, depressed people have too little, and treatments for both conditions seem to normalize ATP levels). These results seem to be daring someone to make up a theory where mania is just too much chemical energy floating around, but if Ortiz et al are doing that, it's sandwiched in between so many dense paragraphs on receptor binding that I can't make it out.

More interesting for us, uric acid is related to all these chemicals and also seems to be involved in mania. See eg de Berardis et al, Evaluation of plasma antioxidant levels during different phases of illness in adult patients with bipolar disorder, which finds that uric acid is elevated in manic patients, and the more manic, the higher the uric acid levels. And Machado-Vieria claims to have gotten pretty good results treating bipolar mania with allopurinol, a gout medication that decreases uric acid – and the more the allopurinol decreased uric acid, the better the results. There's also a little evidence that depressed people have lower uric acid than normal. None of this is a large effect – there are still a lot of depressed people with higher-than-normal uric acid and a lot of manic people with lower – but it's around the same size as all the other infuriatingly suggestive effects we find in psychiatry that never lead to overarching theories or go anywhere useful.

Future studies should try to replicate the link between uric acid and mania, and come up with a better understanding of why it might be true – maybe since uric acid is a decay product of ATP, the body interprets it as a sign that energy is plentiful? They should try to explain away anomalies – if gout is maniogenic, how come so many people with gout are depressed? Is it just because having a painful illness is inherently depressing? And then it should investigate how mania bleeds into normal personality. Is someone with slightly higher uric acid a tiny bit hypomanic all the time?

If they can fill in all those steps, I'll be willing to take a fresh look at the old papers linking gout and achievement. Until then, you should probably hold off on eating megadoses of venison to become the next Ben Franklin.

Predictions For 2020

At the beginning of every year, I make predictions. At the end of every year, I score them. So here are a hundred more for 2020.

Rules: all predictions are about what will be true on January 1, 2021. Some predictions about my personal life, or that refer to the personal lives of other people, have been redacted to protect their privacy. I'm using the full 0 – 100 range in making predictions this year, but they'll be flipped and judged as 50 – 100 in the rating stage, just like in previous years. I've tried to avoid doing specific research or looking at prediction markets when I made these, though some of them I already knew what the markets said.

Feel free to get in a big fight over whether 50% predictions are meaningful.

CORONAVIRUS:
1. Bay Area lockdown (eg restaurants closed) will be extended beyond June 15: 60%
2. …until Election Day: 10%
3. Fewer than 100,000 US coronavirus deaths: 10%
4. Fewer than 300,000 US coronavirus deaths: 50%
5. Fewer than 3 million US coronavirus deaths: 90%
6. US has highest official death toll of any country: 80%
7. US has highest death toll as per expert guesses of real numbers: 70%
8. NYC widely considered worst-hit US city: 90%
9. China's (official) case number goes from its current 82,000 to 100,000 by the end of the year: 70%
10. A coronavirus vaccine has been approved for general use and given to at least 10,000 people somewhere in the First World: 50%

11. Best scientific consensus ends up being that hydroxychloroquine was significantly effective: 20%
12. I personally will get coronavirus (as per my best guess if I had it; positive test not needed): 30%
13. Someone I am close to (housemate or close family member) will get coronavirus: 60%
14. General consensus is that we (April 2020 US) were overreacting: 50%
15. General consensus is that we (April 2020 US) were underreacting: 20%
16. General consensus is that summer made coronavirus significantly less dangerous: 70%
17. …and there is a catastrophic (50K+ US deaths, or more major lockdowns, after at least a month without these things) second wave in autumn: 30%
18. I personally am back to working not-at-home: 90%
19. At least half of states send every voter a mail-in ballot in 2020 presidential election: 20%
20. PredictIt is uncertain (less than 95% sure) who won the presidential election for more than 24 hours after Election Day. 20%

POLITICS:
21. Democrats nominate Biden, and he remains nominee on Election Day: 90%
22. Balance of evidence available on Election Day supports (as per my opinion) Tara Reade accusation: 90%
23. Conditional on me asking about Reade on SSC survey, average survey-taker’s credence in her accusation is greater than 50%: 70%
24. …greater than 75%: 10%
25. …greater than credence in Kavanaugh accusation asked in the same format: 40%
26. Trump is re-elected President: 50%
27. Democrats keep the House: 70%
28. Republicans keep the Senate: 50%

29. Trump approval rating higher than 43% on June 1: 30%
30. Biden polling higher than Trump on June 1: 70%
31. At least one new Supreme Court Justice: 20%
32. I vote Democrat for President: 80%
33. Boris still UK PM: 90%
34. No new state leaves EU: 90%
35. UK, EU extend “transition” trade deal: 80%
36. Kim Jong-Un alive and in power: 60%

ECON AND TECH:

37. Dow is above 25,000: 70%
38. …above 30,000: 20%
39. Bitcoin is above $5,000: 70%
40. …above $10,000: 20%
41. I have bought a Surface Book 3 laptop: 60%
42. Crew Dragon reaches orbit: 80%
43. Starship reaches orbit: 40%

SSC, ETC:

44. I do another Nootropics Survey this year: 70%
45. I do another SSC Survey this year: 90%
46. I start a Reader SSC Survey this year: 60%
47. I start a SSC Book Review Contest this year: 70%
48. I run another Adversarial Collaboration Contest this year: 10%
49. I publish [redacted]: 20%
50. I publish [redacted]: 50%
51. I publish [redacted]: 60%
52. I publish [redacted]: 80%
53. …conditional on being published, it gets at least 40,000 pageviews: 10%
54. I publish [redacted]: 60%
55. …conditional on being published, it gets at least 40,000 pageviews: 50%
56. More hits this year than last: 70%
57. Most hits ever this year: 20%

58. I finish Unsong revision this year: 40%
59. New co-blogger with more than 3 posts: 10%

FRIENDS:
60. No new long-term (1 month +) residents at group house by the end of the year: 70%
61. Koios has said his first clear comprehensible word: 50%
62. [redacted]: 40%
63. [redacted]: 60%
64. [redacted]: 80%
65. [redacted]: 80%
66. [redacted]: 95%
67. [redacted]: 10%
68. [redacted]: 95%
69. [redacted]: 80%
70. [redacted]: 80%
71. [redacted]: 50%

PROFESSIONAL
72. I've gotten at least one new patient to do a full wake therapy protocol: 60%
73. I have specific, set-in-motion plans to quit work / start my own business: 5%
74. I work the same schedule and locations I did before the coronavirus: 80%
75. I get a bonus for 2020: 20%

PERSONAL:
76. [redacted]: 70%
77. [redacted]: 70%
78. [redacted]: 95%
79. I travel to Alaska this year: 60%
80. [redacted]: 40%
81. [redacted]: 20%
82. I go on at least three dates with someone I haven't met yet: 20%

83. [redacted]: 10%
84. [redacted]: 30%
85. [redacted]: 10%
86. I try one biohacking project per month x at least 5 of the last 6 months of 2020: 30%
87. I find at least one new supplement I take or expect to take regularly x 3 months: 20%
88. Not eating meat at home: 40%
89. Weight below 200: 50%
90. Weight below 190: 10%
91. [redacted]: 90%
92. [redacted]: 30%
93. [redacted]: 5%
94. I travel outside the country at least once: 10%
95. I get back into meditating seriously (at least ten minutes a day, five days a week) for at least a month: 10%
96. At least ten tweets in 2020: 80%
97. I eat at/from Sliver more than any other restaurant in Q4 2020: 50%
98. [redacted]: 30%
99. I do pushups and situps at least 3 days/week in average week of Q4 2020: 60%
100. I write the post scoring these predictions before 2/1/21: 70%

Book Review Contest: Call For Entries

Your mission, should you choose to accept it, is to write a book review and send it to me at scott[at]slatestarcodex[dot]com before August 5th 2020.

Interested? Here’s the small print (written in normal-sized print, for your convenience):

Pick a book, then write a review similar to my SSC book reviews (examples). I’m mostly expecting reviews of nonfiction, but I guess you could review fiction if you really wanted and had something interesting to say beyond just “here’s the plot and I thought it was good”.

I’ll choose some number of finalists – probably around five, but maybe more or less depending on how many I get – and publish them on the blog, with full attribution, just like with the adversarial collaborations. Then readers will vote for the best, just like with the adversarial collaborations. First place will get at least $1000, second place $500, third place $250 – I might increase those numbers later on. Some winners may also get an invitation to pitch me any other pieces they have that they think would make good SSC posts. I may also release non-finalist entries somewhere else so people can read them – if you strongly object to me making your entry public, let me know.

Please send me your review in a .txt file (eg Notepad), attached to an email. I’m making this rule because otherwise you send me heavily formatted emails with lots of bold text and weird font changes and tables, and it’s really hard to post to SSC in ways that don’t mess up the formatting or look wrong. If you want formatting in your final posted review, please use a *tiny* amount of *hand-written* html – ie putting bold things in bold and putting links in <A HREF=”http://www.example.com”>links</A>. If you want to include images, please use <IMG SRC=”http://www.example.com/image.png”>. For quotes, <blockquote>quote</blockquote>. If you can’t figure this out, just send me the images and tell me where to put them. Don’t get Word or something to save your heavily formatted document as HTML or it will do horrible things that will screw me up when I try to make it into a blog post. If you can’t make or send .txt files for some reason, please paste your review in the body of the email in a way that follows these same principles. I don’t want this part to prevent anyone from sending something in, so if you don’t understand this and are scared of it, just send something in anyway and I’ll fix it up.

In your email to me, please also include the name you want me to attribute your review to – that could be your real name, your first name plus an initial, your pseudonym, etc.

You don’t have to register beforehand or let anyone else know you’re doing this. But if you want to avoid having someone else accidentally review the same book as you, you can post what you’re doing in the comments below and hopefully other people will avoid doubling up. I have vague plans to review Julian Jaynes’ *The Origin Of Consciousness In The Breakdown Of The Bicameral Mind* and Ezra Klein’s *Why We’re Polarized* before August, so you might want to avoid those too.

If you win, I will pay through PayPal or donations to the charity of your choice. I reserve the right to change these conditions in minor ways that don’t significantly inconvenience contest participants.

I’ll check the comments here for a few days and answer any questions you might have.

[EDIT: Please don’t submit reviews that have been posted on other blogs before. Reviews that have been posted on the r/SSC or r/themotte subreddit are provisionally okay, since I don’t want to disincentivize people from doing that, but I’ll try to come up with better guidance soon]

Classified Thread 8

This is the…monthly? bimonthly? occasional?…classified thread. Post advertisements, personals, and any interesting success stories from the last thread.

Under the circumstances, you can also ask for financial support (eg link a GoFundMe) if you need to. Nobody has to donate to people who do this, but please at least be respectful.

Studies On Slack

I.

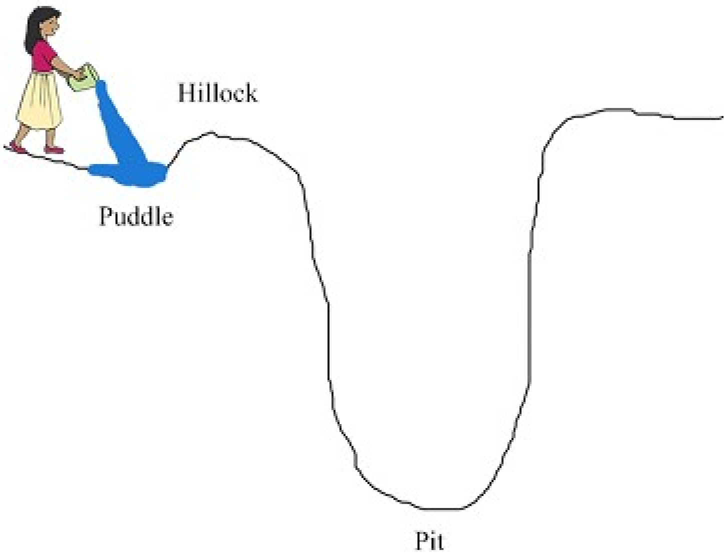
Imagine a distant planet full of eyeless animals. Evolving eyes is hard: they need to evolve Eye Part 1, then Eye Part 2, then Eye Part 3, in that order. Each of these requires a separate series of rare mutations.

Here on Earth, scientists believe each of these mutations must have had its own benefits – in the land of the blind, the man with only Eye Part 1 is king. But on this hypothetical alien planet, there is no such luck. You need all three Eye Parts or they’re useless. Worse, each Eye Part is metabolically costly; the animal needs to eat 1% more food per Eye Part it has. An animal with a full eye would be much more fit than anything else around, but an animal with only one or two Eye Parts will be at a small disadvantage.

So these animals will only evolve eyes in conditions of relatively weak evolutionary pressure. In a world of intense and perfect competition, where the fittest animal always survives to reproduce and the least fit always dies, the animal with Eye Part 1 will always die – it’s less fit than its fully-eyeless peers. The weaker the competition, and the more randomness dominates over survival-of-the-fittest, the more likely an animal with Eye Part 1 can survive and reproduce long enough to eventually produce a descendant with Eye Part 2, and so on.

There are lots of ways to decrease evolutionary pressure. Maybe natural disasters often decimate the population, dozens of generations are spent recolonizing empty land, and during this period there’s more than enough for everyone and nobody has to compete. Maybe there are frequent whalefalls, and any animal nearby has hit the evolutionary jackpot and will have thousands of descendants. Maybe the population is isolated in little islands and mountain valleys, and one gene or another can reach fixation in a population totally by chance. It doesn't matter exactly how it happens, it matters that evolutionary pressure is low.

The branch of evolutionary science that deals with this kind of situation is called "adaptive fitness landscapes". Landscapes really are a great metaphor – consider somewhere like this:



You pour out a bucket of water. Water "flows downhill", so it's tempting to say something like "water wants to be at the lowest point possible". But that's not quite right. The lowest point possible is the pit, and water won't go there. It will just sit in the little puddle forever, because it would have to go up the tiny little hillock in order to get to the pit, and water can't flow uphill. Using normal human logic, we feel tempted to say something like "Come on! The hillock is so tiny, and that pit is so deep, just make a single little exception to your ‘always flow downhill’ policy and you could do so much better for yourself!” But water stubbornly refuses to listen.

Under conditions of perfectly intense competition, evolution works the same way. We imagine a multidimensional evolutionary “landscape” where lower ground represents higher fitness. In this perfectly intense competition, organisms can go from lower to higher fitness, but never vice versa. As with water, the tiniest hillock will leave their potential forever unrealized.

Under more relaxed competition, evolution only *tends probabilistically* to flow downhill. Every so often, it will flow uphill; the smaller the hillock, the more likely evolution will surmount it. Given enough time, it’s guaranteed to reach the deepest pit and mostly stay there.

Take a moment to be properly amazed by this. It sounds like something out of the Tao Te Ching. An animal with eyes has very high evolutionary fitness. It will win at all its evolutionary competitions. So in order to produce the highest-fitness animal, we need to – select for fitness less hard? In order to produce an animal that wins competitions, we need to stop optimizing for winning competitions?

This doesn’t mean that less competition is always good. An evolutionary environment with no competition won’t evolve eyes either; a few individuals might randomly drift into having eyes, but they won’t catch on. In order to optimize the species as much as possible as fast as possible, you need the right balance, somewhere in the middle between total competition and total absence of competition.

In the esoteric teachings, total competition is called Moloch, and total absence of competition is called Slack. Slack (thanks to Zvi Mowshowitz for the term and concept) gets short shrift. If you think of it as “some people try to win competitions, other people don’t care about winning competitions and slack off and go to the beach”, you’re misunderstanding it. Think of slack as a paradox – the Taoist art of winning competitions by not trying too hard at them. Moloch and Slack are opposites and complements, like yin and yang. Neither is stronger than the other, but their interplay creates the ten thousand things.

II.

Before we discuss slack further, a digression on group selection.

Some people would expect this discussion to be quick, since group selection doesn’t exist. These people understand it as evolution acting for the good of a species. It’s a tempting way to think, because evolution usually eventually makes species stronger and more fit, and sometimes we colloquially round that off to evolution targeting a species’ greater good. But inevitably we find evolution is awful and does absolutely nothing of the sort.

Imagine an alien planet that gets hit with a solar flare once an eon, killing all unshielded animals. Sometimes unshielded animals spontaneously mutate to shielded, and vice versa. Shielded animals are completely immune to solar flares, but have 1% higher metabolic costs. What happens? If you predicted “magnetic shielding reaches fixation and all animals get it”, you’ve fallen into the group selection trap. The unshielded animals outcompete the shielded ones during the long inter-flare period, driving their population down to zero (though a few new shielded ones arise every generation through spontaneous mutations). When the flare comes, only the few spontaneous mutants survive. They breed a new entirely-shielded population, until a few unshielded animals arise through spontaneous mutation. The unshielded outcompete the shielded ones again, and by the time of the next solar flare, the population is 100% unshielded again and they all die. If the animals are lucky, there will always be enough spontaneously-mutated shielded animals to create a post-flare breeding population; if they are unlucky, the flare will hit at a time with unusually few such mutants, and the species will go extinct.

An Evolution Czar concerned with the good of the species would just declare that all animals should be shielded and solve the problem. In the absence of such a Czar, these animals will just keep dying in solar-flare-induced mass extinctions forever, even though there is an easy solution with only 1% metabolic cost.

A less dramatic version of the same problem happens here on Earth. Every so often predators (let's say foxes) reproduce too quickly and outstrip the available supply of prey (let's say rabbits). There is a brief period of starvation as foxes can't find any more rabbits and die *en masse*. This usually ends with a boom-bust cycle: after most foxes die, the rabbits (who reproduce very quickly and are now free of predation) have a population boom; now there are rabbits everywhere. Eventually the foxes catch up, eat all the new rabbits, and the cycle repeats again. It's a waste of resources for foxkind to spend so much of time and its energy breeding a huge population of foxes that will inevitably collapse a generation later; an Evolution Czar concerned with the common good would have foxes limit their breeding at a sustainable level. But since individual foxes that breed excessively are more likely to have their genes represented in the next generation than foxes that breed at a sustainable level, we end up with foxes that breed excessively, and the cycle continues.

(but humans are too smart to fall for this one, right?)

Some scientists tried to create group selection under laboratory conditions. They divided some insects into subpopulations, then killed off any subpopulation whose numbers got too high, and “promoted” any subpopulation that kept its numbers low to better conditions. They hoped the insects would evolve to naturally limit their family size in order to keep their subpopulation alive. Instead, the insects became cannibals: they ate other insects’ children so they could have more of their own without the total population going up. In retrospect, this makes perfect sense; an insect with the behavioral program “have many children, and also kill other insects’ children” will have its genes better represented in the next generation than an insect with the program “have few children”.

But sometimes evolution appears to solve group selection problems. What about multicellular life? Stick some cells together in a resource-plentiful environment, and they’ll naturally do the evolutionary competition thing of eating resources as quickly as possible to churn out as many copies of themselves as possible. If you were expecting these cells to form a unitary organism where individual cells do things like become heart cells and just stay in place beating rhythmically, you would call the expected normal behavior “cancer” and be against it. Your opposition would be on firm group selectionist grounds: if any cell becomes cancer, it and its descendants will eventually overwhelm everything, and the organism (including all cells within it, including the cancer cells) will die. So for the good of the group, none of the cells should become cancerous.

The first step in evolution’s solution is giving all cells the same genome; this mostly eliminates the need to compete to give their genes to the next generation. But this solution isn’t perfect; cells can get mutations in the normal course of dividing and doing bodily functions. So it employs a host of other tricks: genetic programs telling cells to self-destruct if they get too cancer-adjacent, an immune system that hunts down and destroys cancer cells, or growing old and dying (this last one isn’t usually thought of as a “trick”, but it absolutely is: if you arrange for a cell line to lose a little information during each mitosis, so that it degrades to the point of gobbledygook after X divisions, this means cancer cells that divide constantly will die very quickly, but normal cells dividing on an approved schedules will last for decades).

Why can evolution “develop tricks” to prevent cancer, but not to prevent foxes from overbreeding, or aliens from losing their solar flare shields? Group selection works when the group itself has a shared genetic code (or other analogous ruleset) that can evolve. It doesn’t work if you expect it to directly change the genetic code of each individual to cooperate more.

When we think of cancer, we are at risk of conflating two genetic codes: the shared genetic code of the multicellular organism, and the genetic code of each cell within the organism. Usually (when there are no mutations in cell divisions) these are the same. Once individual cells within the organism start mutating, they become different. Evolution will select for cancer in changes to individual cells’ genomes over an organism’s lifetime, but select against it in changes to the overarching genome over the lifetime of the species (ie you should expect all the genes you inherited from your parents to be selected against cancer, and all the mutations in individual cells you’ve gotten since then to be selected for cancer).

The fox population has no equivalent of the overarching genome; there is no set of rules that govern the behavior of every fox. So foxes can’t undergo group selection to prevent overpopulation (there are some more complicated dynamics that might still be able to rescue the foxes in some situations, but they’re not relevant to the simple model we’re looking at).

In other words, group selection can happen in a two-layer hierarchy of nested evolutionary systems when the outer system (eg multicellular humans) includes rules that the inner system (eg human cells) have to follow, and where the fitness of the evolving-entities in the outer system depends on some characteristics of the evolving-entities in the inner system (eg humans are higher-fitness if their cells do not become cancerous). The evolution of the outer layer includes evolution over rulesets, and eventually evolves good strong rulesets that tell the inner-layer evolving entities how to behave, which can include group selection (eg humans evolve a genetic code that includes a rule “individual cells inside of me should not get cancer” and mechanisms for enforcing this rule).

You can find these kinds of two-layer evolutionary systems everywhere. For example, “cultural evolution” is a two-layer evolutionary system. In the hypothetical state of nature, there’s unrestricted competition – people steal from and murder each other, and only the strongest survive. After they form groups, the groups compete with each other, and groups that develop rulesets that prevent theft and murder (eg legal codes, religions, mores) tend to win those competitions. Once again, the outer layer (competition between cultures) evolves groups that successfully constrains the inner layer (competition between individuals). Species don’t have a czar who restrains internal competition in the interest of keeping the group strong, but some human cultures do (eg Russia).

Or what about market economics? The outer layer is companies, the inner layer is individuals. Maybe the individuals are workers – each worker would selfishly be best off if they spent the day watching YouTube videos and pushed the hard work onto someone else. Or maybe they're executives – each individual executive would selfishly be best off if they spent their energy on office politics, trying to flatter and network with whoever was most likely to promote them. But if all the employees loaf off and all the executives focus on office politics, the company won't make products, and competitors will eat their lunch. So someone – maybe the founder/CEO – comes up with a ruleset to incentivize good work, probably some kind of performance review system where people who do good work get promoted and people who do bad work get fired. The outer-layer competition between companies will select for corporations with the best rulesets; over time, companies' internal politics should get better at promoting the kind of cooperation necessary to succeed.

How do these systems replicate multicellular life's success without being literal entities with literal DNA having literal sex? They all involve a shared ruleset and a way of punishing rulebreakers which make it in each individual's short-term interest to follow the ruleset that leads to long-term success. Countries can do that (follow the law or we'll jail you), companies can do that (follow our policies or we'll fire you), even multicellular life can sort of do that (don't become cancer, or immune cells will kill you). When there's nothing like that (like the overly-fast-breeding foxes) evolution fails at group selection problems. When there is something like that, it has a chance. When there's something like that, and the thing like that is itself evolving (either because it’s encoded in literal DNA, or because it’s encoded in things like company policies that determine whether a company goes out of business or becomes a model for others), then it can reach a point where it solves group selection problems very effectively.

In the esoteric teachings, the inner layer of two-layer evolutionary systems is represented by the Goddess of Cancer, and outer layer by the Goddess of Everything Else. In each part of the poem, the Goddess of Cancer orders the evolving-entities to compete, but the Goddess of Everything Else recasts it as a two-layer competition where cooperation on the internal layer helps win the competition on the external layer. He who has ears to hear, let him listen.

III.

Why the digression? Because slack is a group selection problem. A species that gave itself slack in its evolutionary competition would do better than one that didn’t – for example, the eyeless aliens would evolve eyes and get a big fitness boost. But no individual can unilaterally choose to compete less intensely; if it did, it would be outcompeted and die. So one-layer evolution will fail at this problem the same way it fails all group selection problems, but two-layer systems will have a chance to escape the trap.

The multicellular life example above is a special case where you want 100% coordination and 0% competition. I framed the other examples the same way – countries do best when their citizens avoid all competition and work together for the common good, companies do best when their executives avoid self-aggrandizing office politics and focus on product quality. But as we saw above, some systems do best somewhere in the middle, where there’s some competition but also some slack.

For example, consider a researcher facing their own version of the eyeless aliens’ dilemma. They can keep going with business as normal – publishing trendy but ultimately useless papers that nobody will remember in ten years. Or they can work on Research Program Part 1, which *might* lead to Research Program Part 2, which *might* lead to Research Program Part 3, which *might* lead to a groundbreaking insight. If their jobs are up for review every year, and a year from now the business-as-normal researcher will have five trendy papers, and the groundbreaking-insight researcher will be halfway through Research Program Part 1, then the business-as-normal researcher will outcompete the groundbreaking-insight researcher; as the saying goes, “publish or perish”. Without slack, no researcher can unilaterally escape the system; their best option will always be to continue business as usual.

But group selection makes the situation less hopeless. Universities have long time-horizons and good incentives; they want to get famous for producing excellent research. Universities have rulesets that bind their individual researchers, for example “after a while good researchers get tenure”. And since universities compete with each other, each is incentivized to come up with the ruleset that maximizes long-term researcher productivity. So if tenure really does work better than constant vicious competition, then (absent the usual culprits like resistance-to-change, weird signaling equilibria, politics, etc) we should expect universities to converge on a tenure system in order to produce the best work. In fact, we should expect universities to evolve a really impressive ruleset for optimizing researcher incentives, just as impressive as the clever mechanisms the human body uses to prevent cancer (since this seems a bit optimistic, I assume the usual culprits are not absent).

The same is true for grant-writing; naively you would want some competition to make sure that only the best grant proposals get funded, but too much competition seems to stifle original research, so much so that some funders are throwing out the whole process and selecting grants by lottery, and others are running grants you can apply for in a half-hour and hear back about two days later. If there's a feedback mechanism – if these different rulesets produce different-quality research, and grant programs that produce higher-quality research are more likely to get funded in the future – then the rulesets for grants will gradually evolve, and the competition for grants will take place in an environment with whatever the right evolutionary parameters for evolving good research are.

I don't want to say these things will definitely happen – you can read Inadequate Equilibria for an idea of why not. But they *might*. The evolutionary dynamics which would normally prevent them can be overcome. Two-layer evolutionary systems can produce their own slack, if having slack would be a good idea.

IV.

That was a lot of paragraphs, and a lot of them started with "imagine a hypothetical situation where…". Let's look deeper into cases where an understanding of slack can inform how we think about real-world phenomena. Seven examples:

1. Monopolies. Not the kind that survive off overregulation and patents, the kind that survive by being big enough to crush competitors. These are predators that exploit low-slack environments. If Boeing has a monopoly on building passenger planes, and is exploiting that by making shoddy products and overcharging consumers, then that means

anyone else who built a giant airplane factory could make better products at a lower price, capture the whole airplane market, and become a zillionaire. Why don't they? Slack. In terms of those adaptive fitness landscapes, in between your current position (average Joe) and a much better position at the bottom of a deep pit (you own a giant airplane factory and are a zillionaire), there's a very big hill you have to climb – the part where you build Giant Airplane Factory Part 1, Giant Airplane Factory Part 2, etc. At each point in this hill, you are worse off than somebody who was not building an as-yet-unprofitable giant airplane factory. If you have infinite slack (maybe you are Jeff Bezos, have unlimited money, and will never go bankrupt no matter how much time and cost it takes before you start earning profits) you're fine. If you have more limited slack, your slack will run out and you'll be outcompeted before you make it to the greater-fitness deep pit.

Real monopolies are more complicated than this, because Boeing can shape up and cut prices when you're halfway to building your giant airplane factory, thus removing your incentive. Or they can do *actually* shady stuff. But none of this would matter if you already had your giant airplane factory fully built and ready to go – at worst, you and Boeing would then be in a fair fight. Everything Boeing does to try to prevent you from building that factory is exploiting your slacklessness and trying to increase the height of that hill you have to climb before the really deep pit.

(Peter Thiel inverts the landscape metaphor and calls the hill a "moat", but he's getting at the same concept).

2. Tariffs. Same story. Here's the way I understand the history of the international auto industry – anyone who knows more can correct me if I'm wrong. Automobiles were invented in the early 20th century. Several Western

countries developed homegrown auto industries more or less simultaneously, with the most impressive being Henry Ford's work on mass production in the US. Post-WWII Japan realized that its own auto industry would never be able to compete with more established Western companies, so it placed high tariffs on foreign cars, giving local companies like Nissan and Toyota a chance to get their act together. These companies, especially Toyota, invented a new form of auto production which was actually much more efficient than the usual American methods, and were eventually able to hold their own. They started exporting cars to the US; although American tariffs put them at a disadvantage, they were so much better than the American cars of the time that consumers preferred them anyway. After decades of losing out, the American companies adopted a more Japanese ethos, and were eventually able to compete on a level playing field again.

This is a story of things gone surprisingly *right* – Americans and Japanese alike were able to get excellent inexpensive cars. Two things had to happen for it to work. First, Japan had to have high enough tariffs to give their companies some slack – to let them develop their own homegrown methods from scratch without being immediately outcompeted by temporarily-superior American competitors. Second, America had to have low enough tariffs that eventually-superior Japanese companies could outcompete American automakers, and Japan's fitness-improving innovations could spread.

From the perspective of a Toyota manager, this is analogous to the eyeless alien story. You start with some good-enough standard (blind animals, American car companies). You want to evolve a superior end product (eye-having animals, Toyota). The intermediate steps (an animal with only Eye Part 1, a kind of crappy car company that stumbles over itself trying out new things) are less fit than the good-enough standard. Only when the inferior intermediate steps are protected from competition (through evolutionary randomness, through tariffs) can the superior end product come into existence. But you want to keep enough competition that the superior end product can use its superiority to spread (there is enough evolutionary competition that having eyes reaches fixation, there is enough free trade that Americans preferentially buy Toyota and US car companies have to adopt its policies).

From the perspective of an economic historian, maybe it's a group selection story. The various stakeholders in the US auto industry – Ford, GM, suppliers, the government, labor, customers – competed with each other in a certain way and struck some compromise. The various stakeholders in the Japanese auto industry did the same. For some reason the American compromise worked worse than the Japanese one – I've heard stories about how US companies were more willing to defraud consumers for short-term profit, how US labor unions were more willing to demand concessions even at the cost of company efficiency, how regulators and executives were in bed with each other to the detriment of the product, etc. Every US interest group was acting in its own short-term self-interest, but the Japanese industry-as-a-whole outcompeted the American one and the Americans had to adjust.

3. Monopolies, Part II. Traditionally, monopolies have been among the most successful R&D centers. The most famous example is Xerox; it had a monopoly on photocopiers for a few decades before losing an anti-trust suit in the late 1970s; during that period, its PARC R&D program invented "laser printing, Ethernet, the modern personal computer, graphical user interface (GUI) and desktop paradigm, object-oriented programming, [and] the mouse". The second most

famous example is Bell Labs, which invented "radio astronomy, the transistor, the laser, the photovoltaic cell, the charge-coupled device, information theory, the Unix operating system, and the programming languages B, C, C++, and S" before the government broke up its parent company AT&T. Google seems to be trying something similar, though it's too soon to judge their outcomes.

These successes make sense. Research and development is a long-term gamble. Devoting more money to R&D decreases your near-term profits, but (hopefully) increases your future profits. Freed from competition, monopolies have limitless slack, and can afford to invest in projects that won't pay off for ten or twenty years. This is part of Peter Thiel's defense of monopolies in *Zero To One*.

An administrator tasked with advancing technology might be tempted to encourage monopolies in order to get more research done. But monopolies can also be stagnant and resistant to change; it's probably not a coincidence that Xerox wasn't the first company to bring the personal computer to market, and ended up irrelevant to the computing revolution. Like the eyeless aliens, who will not evolve in conditions of perfect competition or perfect lack of competition, probably all you can do here is strike a balance. Some Communist countries tried the extreme solution – one state-supported monopoly per industry – and it failed the test of group selection. I don't know enough to have an opinion on whether countries with strong antitrust eventually outcompete those with weaker antitrust or vice versa.

4. Strategy Games. I like the strategy game *Civilization*, where you play as a group of primitives setting out to found a empire. You build cities and infrastructure, research

technologies, and fight wars. Your world is filled with several (usually 2 to 7) other civilizations trying to do the same.

Just like in the real world, civilizations must decide between Guns and Butter. The Civ version of Guns is called the Axe Rush. You immediately devote all your research to discovering how to make really good axes, all your industry to manufacturing those axes, and all your population into wielding those axes. Then you go and hack everyone else to pieces while they’re still futzing about trying to invent pottery or something.

The Civ version of Butter is called Build. You devote all your research, industry, and populace to laying the foundations of a balanced economy and culture. You invent pottery and weaving and stuff like that. Soon you have a thriving trade network and a strong philosophical tradition. Eventually you can field larger and more advanced armies than your neighbors, and leverage the advantage into even more prosperity, or into military conquest.

Consider a very simple scenario: a map of Eurasia with two civilizations, Rome and China.

If both choose Axe Rush, then whoever Axe Rushes better wins.

If both choose Build, then whoever Builds better wins.

What if Rome chooses Axe Rush, and China chooses Build?

Then it depends on their distance! If it’s a very small map and they start very close together, Rome will probably overwhelm the Chinese before Build starts paying off. But if it’s a very big map, by the time Roman Axemen trek all the way to China, China will have Built high walls, discovered longbows and other defensive technologies, and generally become too strong for axes to defeat. Then they can crush the Romans – who are still just axe-wielding primitives – at their leisure.

Consider a more complicated scenario. You have a map of Earth. The Old World contains Rome and China. The New World contains Aztecs. Rome and China are very close to each other. Now what happens?

Rome and China spend the Stone, Bronze, and Iron Ages hacking each other to bits. Aztecs spend those Ages building cities, researching technologies, and building unique Wonders of the World that provide powerful bonuses. In 1492, they discover Galleons and start crossing the ocean. The powerful and advanced Aztec empire crushes the exhausted axe-wielding Romans and Chinese.

This is another story about slack. The Aztecs had it – they were under no competitive pressure to do things that paid off next turn. The Romans and Chinese didn't – they had to be at the top of their game every single turn, or their neighbor would conquer them. If there was an option that made you 10% weaker next turn in exchange for making you 100% stronger ten turns down the line, the Aztecs could take it without a second thought; the Romans and Chinese would probably have to pass.

Okay, more complicated *Civilization* scenario. This time there are two Old World civs, Rome and China, and two New World civs, Aztecs and Inca. The map is stretched a little bit so that all four civilizations have the same amount of natural territory. All four players understand the map layout and can communicate with each other. What happens?

Now it's a group selection problem. A skillful Rome player will private message the China player and explain all of this to her. She’ll remind him that if one hemisphere spends the whole Stone Age fighting, and the other spends it building, the builders will win. She might tell him that she knows the Aztec and Inca players, they’re smart, and they’re going to be discussing the same considerations. So it would benefit both Rome and China to sign a peace treaty dividing the Old World in two, stick to their own side, and Build. If both sides cooperate, they’ll both Build strong empires capable of matching the New World players. If one side cooperates and the other defects, it will easily steamroll over its unprepared opponent and conquer the whole Old World. If both sides defect, they’ll hack each other to death with axes and be easy prey for the New Worlders.

This might be true in *Civilization* games, but real-world civilizations are more complicated. Orson Welles said:

> In Italy, for thirty years under the Borgias, they had warfare, terror, murder and bloodshed, but they produced Michelangelo, Leonardo da Vinci and the Renaissance. In Switzerland, they had brotherly love, they had five hundred years of democracy and peace – and what did that produce? The cuckoo clock.

So maybe a little bit of internal conflict is good, to keep you honest. Too much conflict, and you tear yourselves apart and are easy prey for outsiders. Too little conflict, and you invent the cuckoo clock and nothing else. The continent that conquers the world will have enough pressure that its people want to innovate, *and* enough slack that they’re able to.

This is total ungrounded amateur historical speculation, but when I hear that I think of the Classical world. We can imagine it as divided into a certain number of “theaters of civilization” – Greece, Mesopotamia, Egypt, Persia, India, Scythia, etc. Each theater had its own rules governing average state size, the rules of engagement between states, how often bigger states conquered smaller states, how often ideas spread between states of the same size, etc. Some of those theaters were intensely competitive: Egypt was a nice straight line, very suited to centralized rule. Others had more slack: it was really hard to take over all of Greece; even the Spartans didn’t manage. Each theater conducted its own “evolution” in its own way – Egypt was ruled by a single Pharaoh without much competition, Scythia was constant warfare of all against all, Greece was isolated city-states that fought each other sometimes but also had enough slack to develop philosophy and science. Each of those systems did their own thing for a while, until finally one of them produced something perfect: 4th century BC Macedonia. Then it went out and conquered everything.

If Welles is right, the point isn’t to find the ruleset that promotes 100% cooperation. It’s to find the ruleset that promotes an evolutionary system that makes your group the strongest. Usually this involves some amount of competition – in order to select for stronger organisms – but also some amount of slack – to let organisms develop complicated strategies that can make them stronger. Despite the earlier description, this isn’t necessarily a slider between 0% competition and 100% competition. It could be much more complicated – maybe alternating high-slack vs. low-slack periods, or many semi-isolated populations with a small chance of interaction each generation, or alternation between periods of isolation and periods of churning.

In a full two-layer evolution, you would let the systems evolve until they reached the best parameters. Here we can’t do that – Greece has however many mountains it has; its success does not cause the rest of the world to grow more mountains. Still, we randomly started with enough different groups that we got to learn something interesting.

(I can’t emphasize enough how ungrounded this historical speculation is. Please don’t try to evolve Alexander the Great in your basement and then get angry at me when it doesn’t work)

5. The Long-Term Stock Exchange. Actually, all stock exchanges are about slack. Imagine you are a brilliant inventor who, given $10 million and ten years, could invent fusion power. But in fact you have $10 and need work tomorrow or you will starve. Given those constraints, maybe you could start, I don’t know, a lemonade stand.

You’re in the same position as the animal trying to evolve an eye – you could create something very high-utility, if only you had enough slack to make it happen. But by default, the inventor working on fusion power starves to death tomorrow (or at least makes less money than his counterpart who ran the lemonade stand), the same way the animal who evolves Eye Part 1 gets outcompeted by other animals who didn’t and dies out.

You need slack. In the evolution example, animals usually stumble across slack randomly. You too might stumble across slack randomly – maybe it so happens that you are independently wealthy, or won the lottery, or something.

More likely, you use the investment system. You ask rich people to give you $10 million for ten years so you can invent fusion; once you do, you’ll make trillions of dollars and share some of it with them.

This is a great system. There’s no evolutionary equivalent. An animal can’t pitch Darwin on its three-step plan to evolve eyes and get free food and mating opportunities to make it happen. Wall Street is a giant multi-trillion dollar time machine funneling future profits back into the past, and that gives people the slack they need to make the future profits happen at all.

But the Long-Term Stock Exchange is especially about slack. They are a new exchange (approved by the SEC last year) which has complicated rules about who can list with them. Investors will get extra clout by agreeing to hold stocks for a long time; executives will get incentivized to do well in the far future instead of at the next quarterly earnings report. It's making a deliberate choice to give companies more slack than the regular system and see what they do with it. I don't know enough about investing to have an opinion, except that I appreciate the experiment. Presumably its companies will do better/worse than companies on the regular stock exchange, that will cause companies to flock toward/away from it, and we'll learn that its new ruleset is better/worse at evolving good companies through competition than the regular stock exchange's ruleset.

6. That Time Ayn Rand Destroyed Sears. Or at least that's how Michael Rozworski and Leigh Phillips describe Eddie Lampert's corporate reorganization in How Ayn Rand Destroyed Sears, which I recommend. Lampert was a Sears CEO who figured – since free-market competitive economies outcompete top-down economies, shouldn't free-market competitive companies outcompete top-down companies? He reorganized Sears as a set of competing departments that traded with each other on normal free-market principles; if the Product Department wanted its products marketed, it would have to pay the Marketing Department. This worked really badly, and was one of the main contributors to Sears' implosion.

I don't have a great understanding of exactly why Lampert's Sears lost to other companies even though capitalist economies beat socialist ones; Rozworski and Phillips' People's Republic Of Wal-Mart, which looks into this question, is somewhere on my reading list. But even without complete understanding, we can use group selection to evolve the right parameters. Imagine an economy with several businesses. One is a straw-man communist collective, where every worker gets paid the same regardless of output and there are no promotions (0% competition, 100% cooperation). Another is Lampert's Sears (100% competition, 0% cooperation). Others are normal businesses, where employees mostly work together for the good of the company but also compete for promotions (X% competition, Y% cooperation). Presumably the normal business outcompetes both Lampert and the commies, and we sigh with relief and continue having normal businesses. And if some of the normal businesses outcompete others, we've learned something about the best values of X and Y.

7. Ideas. These are in constant evolutionary competition – this is the insight behind memetics. The memetic equivalent of slack is inferential range, aka "willingness to entertain and explore ideas before deciding that they are wrong".

Inferential distance is the number of steps it takes to make someone understand and accept a certain idea. Sometimes inferential distances can be very far apart. Imagine trying to convince a 12th century monk that there was no historical Exodus from Egypt. You're in the middle of going over archaeological evidence when he objects that the Bible says there was. You respond that the Bible is false and there's no God. He says that doesn't make sense, how would life have originated? You say it evolved from single-celled organisms. He asks how evolution, which seems to be a change in animals' accidents, could ever affect their essences and change them into an entirely new species. You say that the whole scholastic worldview is wrong, there's no such thing as accidents and essences, it's just atoms and empty space. He asks how you ground morality if not in a striving to approximate the ideal embodied by your essence, you say… well, it doesn't matter what you say, because you were trying to convince him that some very specific people didn't leave Egypt one time, and now you've got to ground morality.

Another way of thinking about this is that there are two self-consistent equilibria. There's your equilibrium, (no Exodus, atheism, evolution, atomism, moral nonrealism), and the monk's equilibrium (yes Exodus, theism, creationism, scholasticism, teleology), and before you can make the monk budge on any of those points, you have to convince him of all of them.

So the question becomes – how much patience does this monk have? If you tell him there's no God, does he say "I look forward to the several years of careful study of your scientific and philosophical theories that it will take for that statement not to seem obviously wrong and contradicted by every other feature of the world"? Or does he say "KILL THE UNBELIEVER"? This is inferential range.

Aristotle supposedly said that the mark of an educated man is to be able to entertain an idea without accepting it. Inferential range explains why. The monk certainly shouldn't immediately accept your claim, when he has countless pieces of evidence for the existence of God, from the spectacular faith healings he has witnessed ("look, there's this thing called psychosomatic illness, and it's really susceptible to this other thing called the placebo effect…") to Constantine's victory at the Mulvian Bridge despite being heavily outnumbered ("look, I'm not a classical scholar, but some people are just really good generals and get lucky, and sometimes it happens the day after they have weird dreams, I think there's enough good evidence the other way that this is not the sort of thing you should center your worldview around"). But if he's willing to entertain your claim long enough to hear your arguments one by one, eventually he can reach the same self-consistent equilibrium you're at and judge for himself.

Nowadays we don't burn people at the stake. But we do make fun of them, or flame them, or block them, or wander off, or otherwise not listen with an open mind to ideas that strike us at first as stupid. This is another case where we have to balance competition vs. slack. With perfect competition, the monk instantly rejects our "no Exodus" idea as less true (less memetically fit) than its competitors, and it has no chance to grow on him. With zero competition, the monk doesn't believe anything at all, or spends hours patiently listening to someone explain their world-is-flat theory. Good epistemics require a balance between being willing to choose better ideas over worse ones, and open-mindedly hearing the worse ones out in case they grow on you.

(Thomas Kuhn points out that early versions of the heliocentric model were much worse than the geocentric model, that astronomers only kept working on them out of a sort of weird curiosity, and that it took decades before they could clearly hold their own against geocentrism in a debate).

Different people strike a different balance in this space, and those different people succeed or fail based on their own epistemic ruleset. Someone who's completely closed-minded and dogmatic probably won't succeed in business, or science, or the military, or any other career (except maybe politics). But someone who’s so pathologically open-minded that they listen to everything and refuse to prioritize what is or isn’t worth their time will also fail. We take notice of who succeeds or fails and change our behavior accordingly.

Maybe there’s even a third layer of selection; maybe different communities are more or less willing to tolerate open-minded vs. close-minded people. The Slate Star Codex community has really different epistemic norms from the Catholic Church or Infowars listeners; these are evolutionary parameters that determine which ideas are more memetically fit. If our epistemics make us more likely to converge on useful (not necessarily true!) ideas, we will succeed and our epistemic norms will catch on. Francis Bacon was just some guy with really good epistemic norms, and now everybody who wants to be taken seriously has to use his norms instead of whatever they were doing before. Come up with the right evolutionary parameters, and that could be you!

“My Immortal” As Alchemical Allegory

I.

From Vox: Solving The Mystery Of The Internet’s Most Beloved And Notorious Fanfic. The fanfic is “My Immortal”, a Harry Potter story so famous that it has its own Wikipedia page, and articles about it in Slate, Buzzfeed, and The Guardian.

It’s famous for being really, *really* bad. Spectacularly bad. Worse than it should be possible for anything to be. You wouldn’t think you could get *The Guardian* to write an article about how bad your fanfiction was, but here we are. Everyone agrees that it must have taken a genius to make something so awful, but until recently nobody knew who had authored the pseudonymous work. The Vox article investigates and finds it was probably small-time author Theresa Christodoupolos, who goes by the pen name Rose Christo.

But this leaves other mysteries unresolved. Like: what is going on with it? Its plot makes little sense – characters appear, disappear, change names, and merge into one another with no particular pattern. Even its language is fluid, somewhere between misspelled English and a gibberish that can at best produce associations suggestive of English words.

All these features are unusual in a modern fanfiction. But they’re typical of alchemical texts, which are usually written in a layer of dense allegory. Might this shed more light on *My Immortal*? After spending way too long investigating this, I find strong evidence in favor. *My Immortal* is a description of the Great Work of alchemy. Its otherwise-inscrutable symbolism is a combination of three traditions: the medieval *opus*, the 17th century Rosicrucians, and the native German traditions encoded in Goethe’s *Faust*. We’ll start by going over these traditions, then delve into the text to unveil the hidden meaning.

First Source: The Medieval Opus

Medieval alchemy centered around the Great Work, or *magnum opus*, of creating the Philosopher’s Stone. The Stone is supposedly a substance that can transmute lead into gold and grant immortality. But scholars since Jung have also interpreted the *opus* symbolically, as a process of spiritual transformation. In this reading, the chemical processes are a metaphor for psychological processes, and the creation of the Stone represents the discovery of the true Self, similar to the Christian gnosis or Buddhist enlightenment. Descriptions of the *opus* tend to describe it as taking place in a series of stages, usually three: nigredo (blackness), albedo (whiteness), and rubedo (redness).

In the first stage (nigredo, “blackness”), you start with some kind of base matter. In the chemical allegory, this is usually lead. In the psychological version, this is the normal mental state, with all of its hangups and uncertainties. The seeker at this stage is symbolized by the raven, blackest of animals. He (the medieval system assumes a male seeker) must begin by confronting his unconscious mind, which takes the form of a dragon. The unconscious is full of bizarre and shameful repressed material, and the seeker’s instinct is to run away. Instead, he must slay the dragon, at which point the dragon rises again as an ally. The seeker then unites with the unconscious in the first “chemical wedding”, ending in a sudden revelation of blinding whiteness – the second stage of the Work.

In the second stage (albedo, “whiteness”), the base matter must be cleansed of its impurities. The seeker is analogized to a child in a baptismal font, or bathing in a stream, or [any of several other water metaphors]. Eventually it begins to shine with its own inner silvery-white light. When the dross has been cleared away, the seeker encounters a second representation of his feminine principle. He unites with the feminine principle in the second “chemical wedding”, and finally see his True Self as it really is.

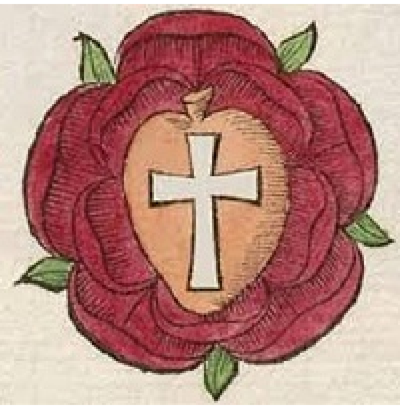
In the third stage (rubedo, “redness”), the seeker has already discovered his True Self as a sort of distant guiding star, but has yet to relate it to the rest of his life or the everyday world. The otherworldy True Self must be united with the seeker’s worldly personality in the final and greatest alchemical wedding, often called the Marriage of the Sun and Moon, or the Marriage of the King and Queen, or [several other flowery metaphors], which joins all opposites into a final cataclysmic union – the Philosopher’s Stone. When this stage ends, the seeker is once again an ordinary person interacting with the ordinary earthly world, but now in a way fully integrated with his true Self. In some traditions, the work is cyclic, and the seeker begins again at the *nigredo* stage.

Some of this is already in canonical Harry Potter – the first book in the series was originally called *Harry Potter and the Philosopher’s Stone*, and Rowling included a few alchemical “easter eggs”. In particular, she includes characters named after two of the three stages: Albus Dumbledore (= albedo) and Rubeus Hagrid (= rubedo). There is no character representing the blackness stage, probably because calling somebody “Nigerus” would be Problematic.

In order to turn Rowling’s half-assed name-dropping into a true alchemical allegory, *My Immortal* has to introduce the missing character with a blackness-themed name. Accordingly, its first sentence starts “Hi my name is Ebony Dark’ness Dementia Raven Way”.

Second Source: The Rosicrucian Writings

These were a series of anonymous pamphlets that took Germany by storm in the early 17th century. They purported to reveal the existence of a secret brotherhood of alchemists, the Rosicrucians, who had discovered vast mystical secrets and were going to disclose them any day now. According to the pamphlets, they had been founded by a hero-sage, Christian Rosenkreutz, who had traveled the world seeking wisdom. But Rosenkreutz was a bit too allegorical for anyone to think he was a real person. His last name was German for “Rosy Cross”, referring to a long tradition of alchemical symbolism in which one produced the Stone by uniting the Rose (the feminine? the spirit? the consciousness?) and the Cross (the masculine? the material world? the body?). The symbolism was a bit unclear, but it caught on, and soon all sorts of mystical groups were using a rose cross as their logo and claiming to be Rosicrucian-inspired.



The most famous Rosicrucian work was *The Chymical Wedding Of Christian Rosenkreutz*, which purports to be a story about Rosenkreutz getting an invitation to go to a castle for a wedding. But this is just the frame story for throwing a metric ton of inscrutable symbolism at the reader, as Christian successively encounters candle-lighting virgins, golden scales, white serpents, and a bunch of gates and towers. Everybody assumed there were deep mystical secrets contained in this, probably related to the alchemical wedding necessary to achieve the Philosopher's Stone.

My Immortal wears its Rosicrucian themes on its sleeve. Most obviously, its author uses the not-exactly-subtle pen name "Rose Christo". But also, the third sentence of the introduction is just "MCR ROX!" A quick check at the My Immortal Wiki tells us that MCR is supposed to be an abbreviation for "My Chemical Romance".

I maintain that if you are writing a fanfiction of a book about the Philosopher's Stone, and you use the pen name "Rose Christo", and you reference a "chemical romance" in the third sentence, you know exactly what you are doing. You are not even being subtle. *My Immortal* is in part a modern retelling of *The Chymical Wedding of Christian Rosenkreutz*.

Third Source: Goethe's Faust II

You know the story. A great alchemist, frustrated by the limitations of his mortal faculties, makes a deal with the Devil. The Devil will give the alchemist anything he desires. In exchange, if the alchemist ever knows a single moment of perfect happiness, he will die and the Devil will get his soul. Maybe you even know the followup: he covets an innocent maiden, Gretchen, and with the help of the Devil he gradually corrupts her until she chooses death over a life of sin.

All that is *Faust* Part I. Later in life, Goethe wrote the much weirder *Faust* Part II. A German scholar assures me that “nobody has any idea what it’s about”, except that it is definitely an alchemical metaphor in some way.

A brief synopsis: Faust, thanks to his pact with the Devil, has now become a powerful sorcerer and respected statesman. He decides that a cool thing to do would be to marry Helen of Troy (here called by her German name “Helena”) the most beautiful woman in history. The Devil (with the help of the Sibyl) helps him time-travel back to ancient Greece, where he meets Helena, seduces her, and brings her back to his own time. They have a child together, but the child dies, and in grief Helena departs Faust for the Greek underworld. Faust devotes himself to a different project – raising a new country out of the sea, which he will govern. The country-raising goes really well, and looking upon his new territory, Faust accidentally feels a single moment of perfect happiness. He dies, and the Devil takes his soul and drags him to Hell. Then a choir of angels show up and distract the Devil. While he is distracted, they carry Faust up to Heaven. There he meets all the women in his life – eg Gretchen and Helena – as well as the Virgin Mary. All of them are revealed to be aspects of the Eternal Feminine within himself (or something), and by recognizing this, he is redeemed and found worthy of salvation. The end.

My Immortal is full of symbolic wordplay (for example, did you catch that “MCR ROX” references not just the Chemical Wedding but also its end result, the Stone?) When it mentions that a character is Goth, or seems Goth, or does something in a Goth way, this is often a visual pun (Goth = Goethe) telling us that the scene has a parallel in Faust. We’ll go over some examples later.

Overall Structure

The canonical version of *My Immortal* is separated into two books of 22 chapters each. In occultism, 22 is the number of completion, especially in Kabbalah (where there are 22 letters of the Hebrew alphabet, 22 generations from Adam to Israel, and 22 paths on the Tree of Life), and Tarot (where there are 22 Major Arcana). So we can think of the book, like the Opus, as a double-traversal of the Tree of Life – first going up from Earth to Heaven, then returning to Earth again.

The medieval and Rosicrucian themes are mostly concentrated in the first half of the text, and consist of a series of thwarted traversals of the alchemical path. At the end of Part 1, a final successful traversal is completed. In Part 2, we segue to a scene-by-scene identity with Part 2 of *Faust*. More specifically:

Chapter 1 – 5: Alchemical Path 1, nigredo, albedo. Purification fails, seeker sinks back into *prima materia*.
Chapters 6 – 18: Alchemical Path 2, nigredo, albedo, partial rubedo. Second alchemical wedding fails, seeker sinks back into *prima materia*
Chapters 18 – 22: Alchemical Path 3, nigredo, albedo, rubedo. Second alchemical wedding partly completed, seeker remains in limbo state.
Chapters 22 – 39: Equivalent to *Faust*, Act II, Scene 3.
Chapters 40 – 44. Equivalent to *Faust*, Act II, Scene 5. Third alchemical wedding succeeds, Stone attained.

I realize these are very odd claims, so I want to demonstrate the flow of symbolism in each of these and compare it to that used in more traditional alchemical texts.

II.

Chapters 1 – 5: First Path

In his papers on alchemy, Carl Jung writes "Great importance was attached to the blackness as the starting point of the Work". The first sentence of *My Immortal* begins "Hi my name is Ebony Dark'ness Dementia Raven Way".

But Ebony's name isn't just blackness. It's a combination of all the different symbols of the *nigredo* stage. Let's look at the rest of the Jung quote:

> Great importance was attached to the blackness as the starting point of the Work. Generally it was called the **"Raven"**. In our context the interpretation of the nigredo as terra (earth) is significant. Like the anima media natura or Wisdom, earth **is in principle feminine**. It is the earth which, in Genesis, appeared out of the waters, but it is also the terra **damnata**.

I've bolded the relevant points. Unlike the albedo and rubedo characters, the nigredo character must be feminine. Her name references the Raven. More speculatively, damnata = Dementia? I think plausibly true. Her name is just a bunch of nigredo symbols strung together.

The *nigredo* stage begins with a black substance (sometimes identified with lead, the starting point for the transmutation) being placed in a vessel called "the coffin". From Landauer & Barnes (2011):

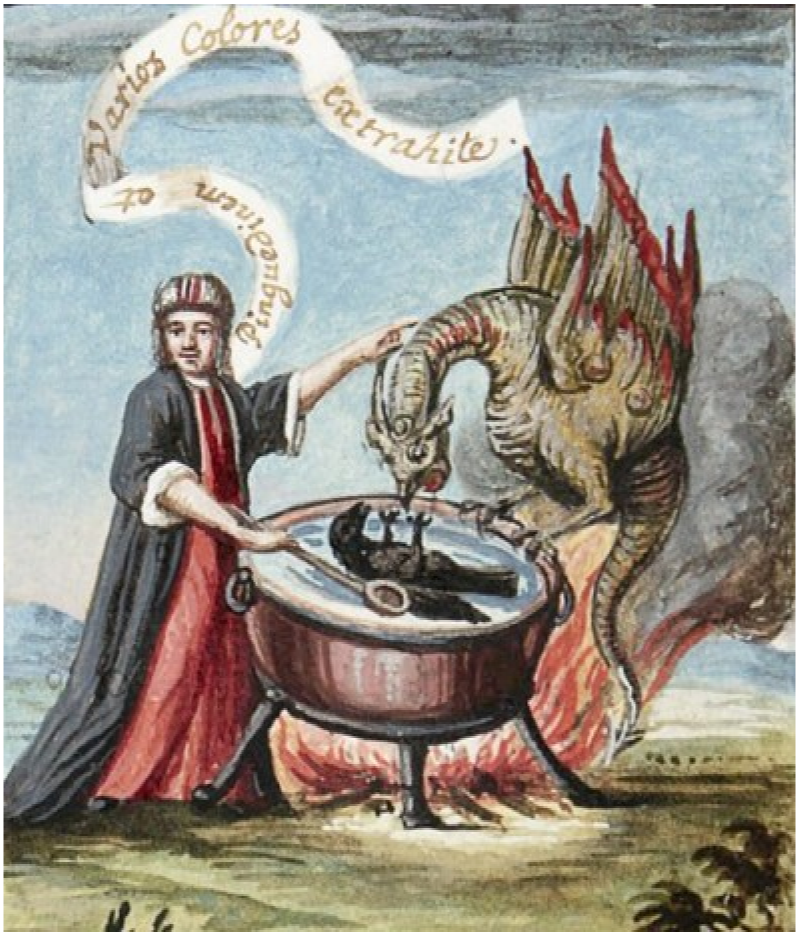
> The alchemists used a number of different vessels in their work and these vessels – variously known as alembic, coffin, egg, sphere, prison, and womb – particular to stages in the alchemical process. During the blackness of the putrefying Mortificatio, the vessel was represented as a coffin or prison

Each of the three paths in Part I of *My Immortal* begins with Ebony waking up in a coffin. For example Chapter 2: "I got out of my coffin and took of my giant MCR t-shirt which I used for pajamas”. Remember, MCR means “my chemical romance” and when it appears it usually tells us that we are getting an alchemical analogy.

The black substance in the coffin must then undergo a series of reactions, usually symbolized as interactions between a raven and dragon. Sometimes the raven and dragon are the same entity; other times they are different entities that must confront each other and unite. For example, from the *Aurelia Occultae Philosophorum*:

> I am an infirm and weak old man, surnamed the dragon; therefore am I shut up in a cave, that I may become ransomed by the kingly crown. A fiery sword inflicts great torments on me; death makes weak my flesh and bones. My soul and my spirit depart; a terrible poison, I am likened to the black raven, for that is the wages of sin.

In Jung’s more psychological version, the raven is the seeker and the dragon is the seeker’s unconscious mind. The seeker must begin by confronting his unconscious and all the repressed material therein.



A typical alchemical illustration with raven and dragon. This one isn't going so well for the raven.

In the first chapter of *My Immortal*, a raven-named character meets a dragon-named character:

> "Hey Ebony!" shouted a voice. I looked up. It was…. Draco Malfoy!
>
> "What's up Draco?" I asked.
>
> "Nothing." he said shyly.
>
> But then, I heard my friends call me and I had to go away.

The encounter with the unconscious begins the process of *mortificatio*. This is cognate with the English word "mortify", and with good reason – the unconscious is full of all of our deepest and most shameful repressed desires. Herzer and Gillabel describe it like so:

> In alchemy the dragon corresponds closely with what Jung called the Shadow. The Shadow is the name for a collection of characteristics and impulses which could be conscious, but which are denied. At the same time we recognize and see them in other people. Some examples of the Shadow are: egotism, laziness, intrigues, unreal fantasies, indifference, or being obsessed by money and possessions. The Shadow is the inferior being in us that desires what we do not allow ourselves because it is uncivilized, because it is incompatible with society's rules and with the image of our ideal personality. It is all that what we are ashamed of.

When Ebony encounters Draco, she feels shame:

> "OMFG, I saw you talking to Draco Malfoy yesterday!" [Willow] said excitedly.
>
> "Yeah? So?" I said, blushing.
>
> "Do you like Draco?" she asked as we went out of the Slytherin common room and into the Great Hall.
>
> "No I so fucking don't!" I shouted.
>
> "Yeah right!" she exclaimed.

The work of the *nigredo* stage is to transmute this shame into acceptance and even love. Some sources describe this as slaying the dragon – but after being slain, the dragon rises again in a perfected form. Once the seeker is fully comfortable with their dragon, self and unconscious unite in the first alchemical wedding.

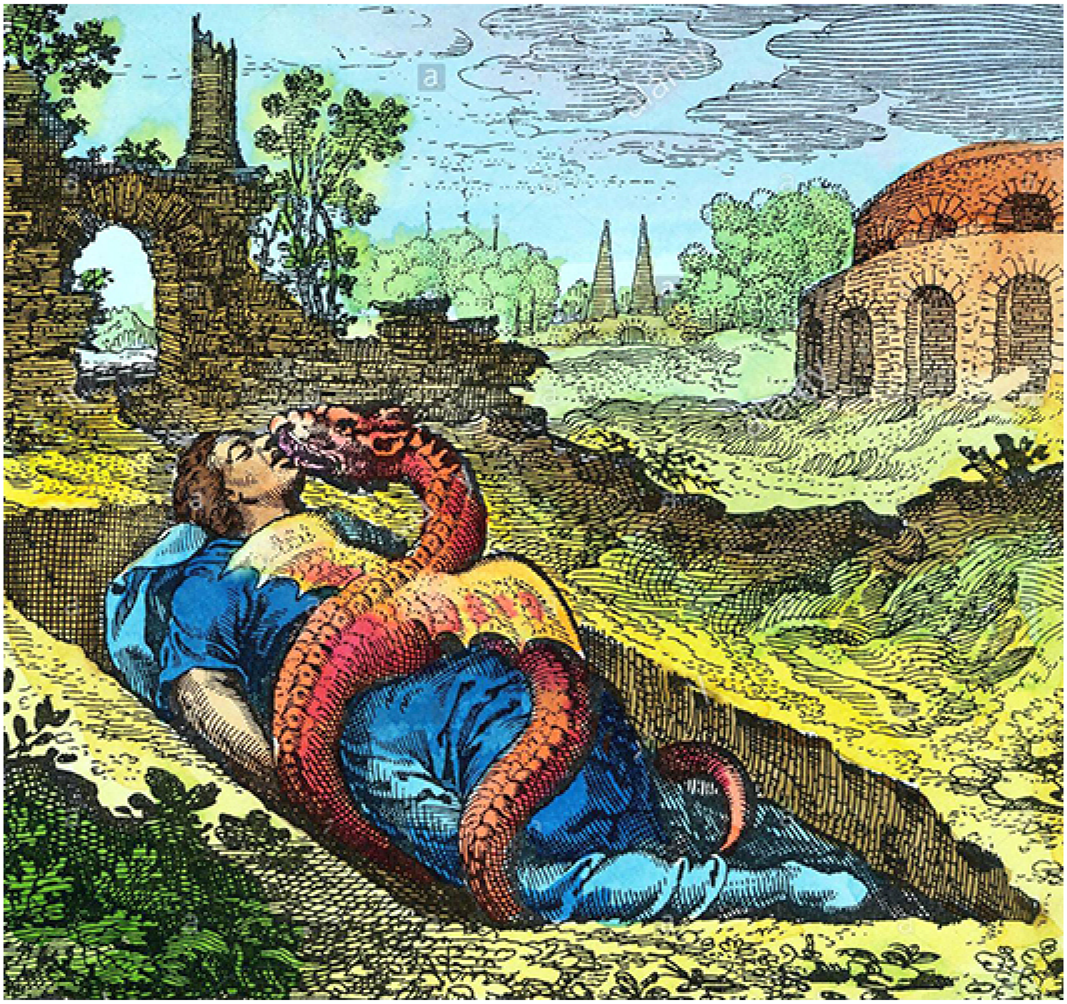
> Just then, Draco walked up to me.
>
> “Hi.” he said.
>
> “Hi.” I replied flirtily.
>
> “Guess what.” he said.
>
> “What?” I asked.
>
> “Well, Good Charlotte are having a concert in Hogsmeade.” he told me.
>
> “Oh. My. Fucking. God!” I screamed. I love GC. They are my favorite band, besides MCR.
>
> “Well…. do you want to go with me?” he asked.
>
> I gasped.

Draco invites Ebony on a date, where they will see a band Ebony compares to My Chemical Romance. In general, concert dates in *My Immortal* represent alchemical weddings. This particular concert is the wedding at the end of the first stage. We know this because the story describes Ebony’s clothing in detail at several points, and it is always some combination of black, white, and red. The particular colors at any given time indicate the stage of the Work being represented. In this case, when dressing up for the concert, Ebony says:”I painted my nails black and put on TONS of black eyeliner. Then I put on some black lipstick.” The only color is black – so she is still entirely in the *nigredo* stage.

At the end of the concert/date/wedding, Draco and Ebony go to the Forbidden Forest and have sex, representing the union of Ebony and her unconscious. This signals the end of the nigredo stage and the beginning of albedo.

> Then he put his thingie into my you-know-what and we did it for the first time.
>
> “Oh! Oh! Oh! ” I screamed. I was beginning to get an orgasm. We started to kiss everywhere and my pale body became all warm. And then….
>
> “WHAT THE HELL ARE YOU DOING YOU MOTHERFUKERS!”
>
> It was……………………………………………………Dumbledore!

Almost all descriptions of the beginning of albedo emphasize its suddenness. Jung writes that “the nigredo gives way to the albedo…the ever deepening descent into the unconscious suddenly becomes illumination from above”. Remember that *Albus* Dumbledore is the Harry Potter character representing albedo, even as Ebony represents *nigredo*. So as Ebony has sex with Draco (deepening descent into the unconscious), suddenly she sees Dumbledore standing over her.



The part where the seeker has sex with the dragon is canon.

But there are signs that the transition is not complete. Dumbledore interrupts her before she can have an orgasm; the union has been only partially consummated. And there are 21 marks between “was” and “Dumbledore” (20 ellipses plus one lone period), which is one short of 22, the mystical number of completion.

As we will see later, Ebony skipped a step – she did not kill the dragon before uniting with it. Therefore, the purification of the albedo appears as the hostile interference of a superego perceived as alien, rather than as a deliberate cleansing from within. Dumbledore admonishes Ebony for her disgusting sexual act and forces her and Draco to separate, ending the path.

Chapter 6 begins “The next day I woke up in my coffin. I put on a black miniskirt that was all ripped around the end and a matching top.” Ebony is back in the coffin, dressed in all black. She has returned to the very beginning of the Work.

Chapters 6 – 18: Second Path

In this section, Ebony retraces her steps. She meets Draco again. They have sex again. But this time, Ebony learns that Draco is in love with Harry Potter (nicknamed “Vampire” in this work), and breaks up with him angrily. This causes Draco to commit suicide – she has slain the dragon. Later, she learns that Draco has inexplicably come back to life (as the alchemical symbolism insists must happen) and is being held hostage in Voldemort’s lair. Recall again the description from *Aurelia Occultae Philosophorum*: “I am an infirm and weak old man, surnamed the dragon; therefore am I shut up in a cave, that I may become ransomed by the kingly crown”. She rescues Draco and has sex with him again. The path has been completed in full.

But she actually reaches the albedo stage earlier than that – at the moment Draco dies. Let’s go through the symbolism piece by piece:

> We practiced for one more hour. Then suddenly Dumbeldore walked in angrily! His eyes were all fiery and I knew this time it wasn’t cause he had a headache.
>
> “What have you done!” He started to cry wisely. (c dats basically nut swering and dis time he wuz relly upset n u wil c y) “Ebony Draco has been found in his room. He committed suicide by slitting his wrists.” […]

> I started crying tears of blood and then I slit both of my wrists. They got all over my clothes so I took them off and jumped into the bath angrily while I put on a Linkin Park song at full volume. […]
>
> I got out of the bathtub and put on a black low-cut dress with lace all over it sandly.

Dumbledore appears, again “suddenly”. He gives her the news of Draco’s death. Then Ebony slits her wrists. Compare to Hamilton’s The Alchemical Process Of Transformation:

> Completion of the first stage is now experienced as a death, which is in fact a complete letting go of the old sense of self that was identified unconsciously with the earth nature. Images of fire and burning often accompany the images of death and endings. Now we are ready to enter the second stage.

Did you catch how Dumbledore’s “eyes were all fiery”?

Just as the nigredo stage should begin in a coffin, so the albedo stage should begin in a baptismal font. But a bathtub is a perfectly serviceable replacement – in fact, a raven in a bathtub is the same image used in this part of *The Chymical Wedding Of Christian Rosenkreutz*. A magic egg hatches a black bird, and then:

> A bath colored with fine white powder had been prepared for the bird, which enjoyed bathing in it until the lamps placed beneath the bath caused the water to become uncomfortably warm. When the heat had removed all the bird’s feathers it was taken out.

When Ebony gets out of the bath, she puts on a “dress with lace all over it” – the lace is presumably white, so the dress is black-and-white, signifying that she has now attained both the nigredo and the albedo stages. Her current task in the Work is to purify herself, see through the falsehoods of the ego, and behold her true Self.

Ebony notices that Professors Snape and Lupin have been videotaping her naked in the bath. She accuses them of pedophilia and then "I took my gun and shot Snape and Loopin a gazillion times and they both started screaming and the camera broke." This is the purification and the confrontation with the ego. The ego is analogized to a voyeur with a camera, recording everything we do, showing us our own nakedness. Ebony shoots the camera and breaks it. Compare to Dogen's analogy of enlightenment to breaking a mirror.

Ebony's next step should be a second alchemical wedding, this time with her animus.

The medieval alchemists (writing for their male readers) spoke of the anima figure, the feminine archetype within every man. The anima can appear as a hideous old hag, but if accepted, turns into a supernaturally beautiful young woman, the intended bride of the second alchemical wedding. Jung adds that every woman has an equivalent animus, the male archetype. Presumably he too starts out appearing ugly, but if accepted he transforms into a handsome young man.

Ebony's dual animus is Hagrid on the one side, and Harry "Vampire" Potter on the other. She constantly conflates these two characters in bizarre ways; we readers are supposed to understand that they represent different facets of the same archetype, which Ebony cannot integrate. For example, when Harry comes in in Chapter 12, Ebony says "I THOUGHT IT WAS HAIRgrid but it was Vampire." Later, she will constantly refer to them by a combined names like “Hairgrid”, “Hargrid”, or even “Hahrid”.

Ebony externalizes the negative aspects of her animus as Hagrid (note the “Hag”, which she cannot bring herself to say), canonically the ugliest person at Hogwarts. Hagrid is everything she hates: prep, not Goth, and “fucked up”. She externalizes the alluring aspects as Harry, canonically rich, famous, and attractive. Ever since *Twilight*, the archetypal animus figure – the alluring, supernaturally beautiful, mysterious male – has been the vampire, so Harry becomes “Harry ‘Vampire’ Potter”.

There’s more. In *My Immortal*, Harry has a pentagram-shaped scar. The pentagram is an alchemical symbol, but there’s some evidence that it’s not what’s really involved here. Earlier in the story, Ebony describes Draco as “wearing white foundation and messy eyeliner kind of like a pentagram (geddit) between Joel Madden and Gerard Way.” Later she uses the same device: “My voice sounded lik a pentagram betwen Amy Lee and a gurl version of Gerard Woy”. I think we are supposed to infer that, as a Satanist, she is uncomfortable with naming the Christian symbol (the cross) and so replaces it with a Satanist symbol (the pentagram). So plausibly Harry has a cross-shaped scar.

This is significant because a few paragraphs later, when Ebony is in the hospital after her wrist-slitting injury, Hagrid comes and offers Ebony a bouquet of roses. Harry (the cross-bearer) is the same person as Hagrid (the rose-bearer). Ebony needs to recognize them as different aspects of the same animus, the united Rosy Cross, so she can behold her True Self. Let’s see how she does:

> Hargrid came into my hospital bed holding a bouquet of pink roses.

“Enoby I need to tell u somethnig.” he said in a v. serious voice, giving me the roses.

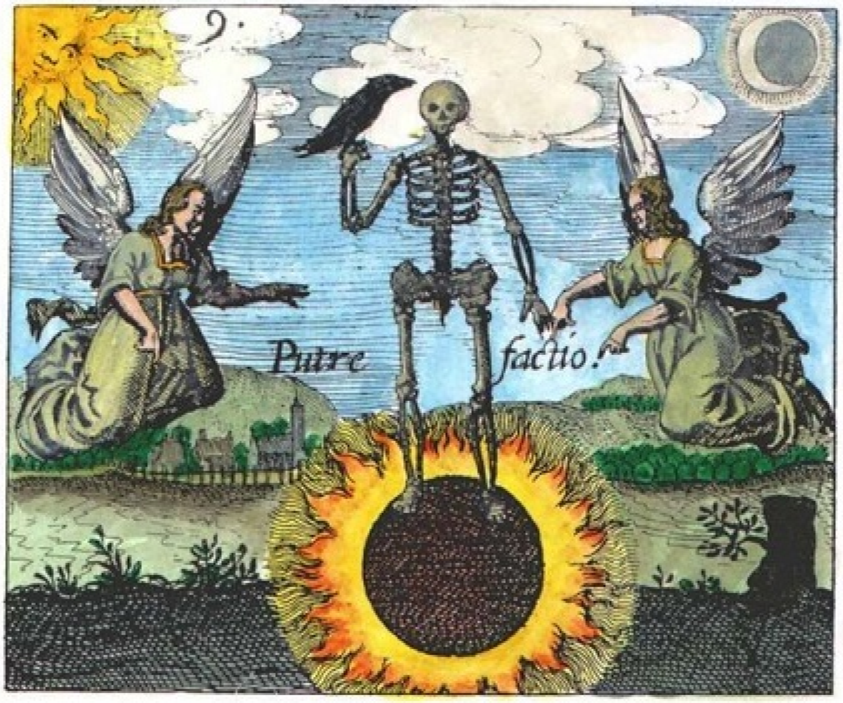
“Fuck off.” I told him. “You know I fucking hate the color pink anyway, and I don’t like fucked up preps like you.” I snapped. Hargrid had been mean to me before for being gottik.

“No Enoby.” Hargrid says. “Those are not roses.”…He suddenly looked at them with an evil look in his eye and muttered Well If you wanted Honesty that’s all you haD TO SAY! .

“That’s not a spell that’s an MCR song.” I corrected him wisely.

“I know, I was just warming up my vocal cordes.” Then he screamed. “Petulus merengo mi kremicli romacio(4 all u cool goffic mcr fans out, there, that is a tribute! specially for raven I love you girl!)imo noto okayo!”

And then the roses turned into a huge black flame floating in the middle of the air. And it was black.



Pictured: Ebony Raven and "a huge black flame floating in the middle of the air."

If Ebony had accepted Hagrid, he would have turned into Harry, united with her for the second alchemical wedding, and revealed her True Self. Instead, he produces the sign of black fire. The black sun (*sol niger*) is one of the most ominous symbols in alchemy. Its various connotations are too complicated to explain here (see *The Black Sun: The Alchemy And Art Of Darkness* for a thorough review) but they are generally highly negative. "The sol niger [is] Saturn, is the shadow of the sun, the sun without justice, which is death for the living."

Yet at times the Black Sun can be useful; suffering itself becomes a catalyst for transformation. This seems to be one of those times. Ebony is shocked into acknowledging Hagrid ("Now I knew he wasn't a prep. "OK I believe you now'")

Is this good enough for her to find her True Self? Going back to the text:

> And then the roses turned into a huge black flame floating in the middle of the air. And it was black. Now I knew he wasn’t a prep.
>
> “OK I believe you now wtf is Drako?”
>
> Hairgrid rolled his eyes. I looked into the balls of flame but I could c nothing.
>
> “U c, Enobby,” Dumblydore said, watching the two of us watching the flame. “2 c wht iz n da flmes(HAHA U REVIEWRS FLAMES GEDDIT) u mst find urslf 1st, k?”
>
> “I HAVE FOUND MYSELF OK YOU MEAN OLD MAN!” Hargrid yelled.

Dumbledore tells Ebony that she needs to find herself. But it is Hagrid who retorts that he has found herself. I take this to mean she has failed to integrate with Hagrid; he remains a psychopomp figure, containing the knowledge of Ebony’s Self but unable to transmit it to her.

Nevertheless, something has been gained, because:

> Anyway when I got better I went upstairs and put on a black leather minidress that was all ripped on the ends with lace on it. There was some corset stuff on the front. Then I put on black fishnets and black high-heeled boots with pictures of Billie Joe Armstrong on them. I put my hair all out around me so I looked like Samara from the Ring (if u don’t know who she iz ur a prep so fuk off!) and I put on blood-red lipstick, black eyeliner and black lip gloss.

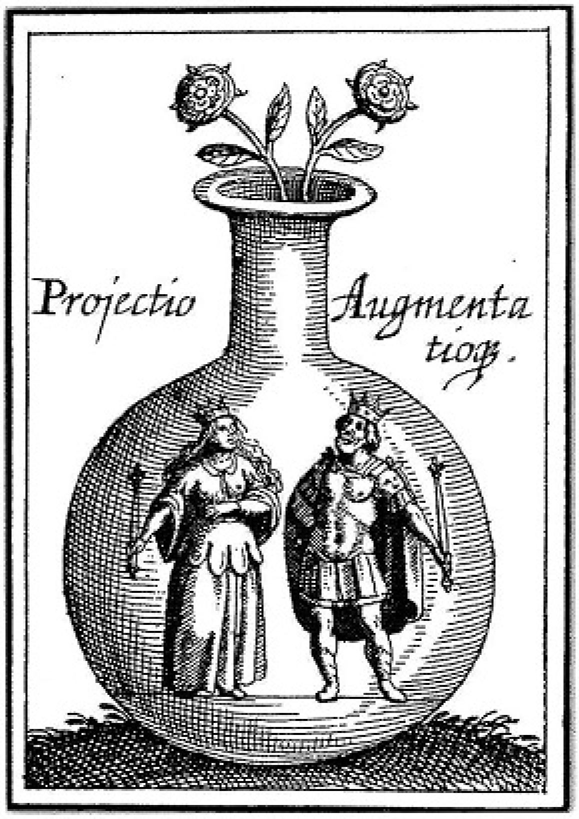
Ebony is still dressed mostly in black, but has a tiny bit of white (the lace) and the slightest bit of red (the lipstick).

She is basically back in *nigredo* now, with only hints of the other stages she has tried to attain.

All of this catches up with her in the total fuckup of an attempted alchemical wedding that follows. If she is actually in albedo, she should be uniting with her animus. If she is actually in rubedo, she should be uniting with God. Instead, it’s just Draco again.

> Then I saw a poster saying that MCR would have a concert in Hogsmede right then. We [Ebony and Draco] looked at each other all shocked and then we went 2gether.

There is a long list of Ebony and Draco’s preparations for the concert (cf. the preparation chapters of *The Chymical Wedding Of Christian Rosenkreutz*, and note the part where Ebony calls Draco “Christian” at the beginning of Chapter 16).



The original chemical romance.

But the actual concert itself is a disaster:

> Gerard was da sexiest guy eva! He locked even sexier den he did in pix. He had long raven blak hair n piercing blue eyes. He wuz really skinny and he had n amazing ethnic voice. We moshed 2 Helena and sum odder songz. Sudenly Gerard polled of his mask. So did the other membez. I gasped. It wasn't Gerard at all! It was an ugly preppy man wif no nose and red eyes… Every1 ran away but me and Draco. Draco and I came. It was…….Vlodemort and da Death Deelers!

The supposed concert/date/wedding proves utterly disastrous – "My Chemical Romance" is just Voldemort wearing a mask. Draco and Ebony flee. Chapter 18 begins:

> I woke up the next day in my coffin. I walked out of it and put on some black eyeliner, black eyesharrow, blood-bed lipstick and a black really low-cut leather dress that was all ripped and in stripes so you could see my belly

She is back in the coffin, dressed in all black (she tries to say the word “red”, but fails). She has lost even the hints of previous stages.

Chapters 18 – 22: Third Path

At the beginning of Chapter 18, Ebony is in her coffin, wearing all black. She immediately meets up with Draco (fourth paragraph of Chapter 18). Immediately after this, Dumbledore suddenly appears (fifth paragraph of Chapter 18). There is purification by water (ninth paragraph of Chapter 19, “I ran to the bathroom”). Hagrid suddenly apperas (tenth paragraph of Chapter 19, “Suddenly Hargrid came. He had appearated.”) Hagrid says there is a My Chemical Romance concert that night (thirteenth paragraph of Chapter 19). We have speed-retraced the entire alchemical path accomplished previously.

But this time, Ebony accepts Hagrid by recognizing him as the keeper of the secret of the alchemical wedding (“’U no who MCR r!’ I gasped.”) And so at the concert, the other half of the animus appears to Ebony in a a merciful form (“Vampire and I began 2 make out, moshing to the muzik.”) Instead of going back to her coffin alone, “we went back to our coffins frenching each other…on the gothic red bed together”. Notice that instead of being in a black coffin, they are now in a red bed, the symbol of the bridal chamber after the second alchemical wedding:

> Take the fayer Roses, white and red
> And joyne them well in won bed.

This ends the first traversal of the *Tree of Life*. Part 2 of *My Immortal* will continue some of the same themes, but subordinate them to a more specific purpose: a reworking of Goethe's *Faust, Part II*. In the process, it will show us the completion of the third alchemical wedding and the creation of the Philosopher's Stone.

Chapters 22 – 39: Faust II, Act 3

For many readers, the weirdest part of *My Immortal* is the subplot beginning around Chapter 23, where Sybil Trelawney helps Ebony go back in time to the 1980s to seduce young Tom Riddle. Using various bizarre time machines (including Marty McFly's Delorean) Ebony successfully goes back, woos Tom, returns with him to her own time, and has sex with him.

This is bizarre, but it's a close parallel of *Faust* II, Act 3, where the Sibyl helps Faust go back in time to ancient Greece to seduce Helen of Troy. Using various bizarre time machines (including riding on the back of the centaur Chiron) Faust successfully goes back, woos Helena, returns with her to his own time, and has a child with her.

There's actually an even more direct reference. In Chapter 38, Ebony and Tom are talking about music. Even though Ebony has previously committed to not talking about My Chemical Romance because they didn't exist in the 1980s, she brings up MCR anyway and Tom is mysteriously familiar with them. Then Ebony says something amazing: "Lol, I totally decided not 2 comit suicide when I herd Hilena."

In context, "Hilena" is a mispelling of "Helena", an MCR song. But "Helena" is also Goethe's Helen of Troy figure. I'll refer you to The Helena Myth In Goethe’s Faust And Its Symbolism for the full treatment, but the point is that this is basically a one-sentence summary of Part II, Act 3 of Faust. It is Faust’s encounter with Helena, representing the feminine ideal, which saves him from despair and makes life worth living. The Ebony-Tom relationship in *My Immortal* is a close parallel to this, and here it cheekily calls out the original to anyone with ears to listen.

But the matchup is not perfect. Faust scholars identify three alchemical weddings in the book: first to Gretchen, then to Helena, then to a divine figure representing the Virgin Mary. This mirrors the traditional *opus* – first you unite with your unconscious, then with the anima, then with divinity. The Helena references in *Faust* all correspond to the second stage – Helena is Faust’s anima.

Ebony has already had an alchemical wedding with her animus in Chapter 20 – the My Chemical Romance concert where she makes out with Vampire. Looking back, there is a Helena reference there too:

> Vampire and I began 2 make out, moshing to the muzik. I gapsed, looking at da band. I almost had an orgasim. Gerard was so fucking hot! He begin 2 sing ‘Helena’ and his sexah beautiful voice began 2 fill the hall.

But we should be done with these references! Ebony should be completely in the third stage by now! In Chapter 38, she attends a concert with Satan. This ought to be her third alchemical wedding: union with a divine figure. Given that she is Satanist, the appropriate divine figure is right there. Instead, they’re not only talking about Helena, they’re unconsciously re-enacting the Helena subplot from *Faust*. Why?

I think the last alchemical wedding never completed. She “makes out” with Vampire, but does not have sex with him. The union is only partial. Her date with Satan is a confused attempt at the second and third alchemical weddings combined, which is why she can’t decide whether to call him Tom Riddle (another animus figure – remember that canonically Harry Potter is a horcrux of Tom Riddle and they share part of the same soul) or Satan (the divine figure). So although the explicit text is a bunch of parallels to the third/final alchemical union with the divine, the symbology and the Faust metaphor are caught in the second stage.

Once again, things have gotten extremely bad for Ebony’s spiritual growth. We saw a prelude of this with the omen of the Black Sun in Chapter 12. Now things have deteriorated further. She is going to have to take the hard route: she must go through the fire.

In the middle of the concert with Satan (Chapter 38), the song suddenly becomes dissonant. James Potter tries to shoot Lucius Malfoy, but Ebony jumps in front of the bullet and dies.

Chapters 39 – 44: Faust II, Acts 4-5

Chapter 39 starts with a prelude saying that it’s written by a hacker who hates *My Immortal*. He cracked the real author’s password and plans to ruin the story by writing a deliberately uncharacteristic chapter. He breaks with all of the normal author’s stylistic conventions, eg by using good spelling and grammar throughout. This new chapter ends with Ebony going to Hell, staying there for all eternity, and never being able to do anything Goth ever again.

Then the original author reasserts control, apologizes for letting her account get hacked, and starts over. According to the new, canon timeline, Ebony survives her apparent death (because she was back in time, and so couldn't really die) and returns to her own era.

Compare to the ending of Faust. Mephistopheles appears and drags Faust's soul to Hell. But a choir of angels show up, distract him, and steal Faust's soul away to Heaven.

The end of *My Immortal* also revolves around Ebony's redemption from Hell, but via a novel plot device making use of the fourth wall. Ebony is saved not by a spiritual conflict on her own plane, but by a conflict on a higher plane – that between her Author and a hacker trying to destroy the Author's story. When the Author wins by getting her password back, Ebony is released from Hell; her eternal damnation is retroactively cancelled. This is a sort of weird way of doing a *deus ex machina*, but honestly it's less jarring than Goethe's version.

Upon his salvation, Faust meets all the female characters from his life again, and they redeem him through the power of the Eternal Feminine. Similarly, upon her release Ebony meets all the male characters from her own life. In her case, this leads to an orgy. In Chapter 43, she has a foursome with Draco, Vampire, and Satan. Now, finally, she is consummating the final alchemical wedding, the "union of all opposites" in which she achieves an ultimate integration with all the male aspects of her personality.

Chapter 44, the last chapter of *My Immortal*, has no parallel in *Faust*. Instead of the *hieros gamos* between masculine and feminine completing the redemptive process and ending the Great Work, in *My Immortal* it only initiates the final apocalypse. All the good characters and all the evil characters show up in the Great Hall and begin to fight in a difficult-to-follow scene. Finally, Dumbledore tells Ebony she has to fulfill her destiny and kill Voldemort. In the last line, Ebony shouts “ABRA KEDABRA!!!!!!!!!!!11111” and the story ends.

What is the meaning of this final word? It partly corresponds to the Harry Potter spell *avada kedavra*. But the spelling is neither the Rowling version nor the traditional stage magician version.

The closest match I can find is from Aleister Crowley’s *Book of the Law*, whose last sentence is “The ending of the words is the Word Abrahadabra” In Crowley’s commentaries, he explains that “[abrahadabra is] the Word of the Aeon, which signifieth The Great Work accomplished.”

So *My Immortal* ends with an occult term signifying the “ending of words” and completion of the alchemical Great Work. This suggests Ebony’s redemption was successful; she has escaped Hell, merged with the Eternal Masculine, and triumphed over Voldemort (representing death). She has created the Philosopher’s Stone and achieved the story’s namesake immortality.

But in a different commentary, Crowley also writes: “Abrahadabra is the glyph of the blending of the 5 and 6, the Rose and the Cross.”

…which suggests one last, “hidden” chapter.

The hacker subplot in Chapter 39 suggests that “the story” should be taken to include not just Ebony’s own story, but the frame story in which an author is writing *My Immortal* as a Harry Potter fanfiction and confronting reviewers, hackers, etc. The Ebony story ends with Ebony speaking the mystic word that unites Rose and Cross. The frame story picks up with a woman named Rose Christo appearing and identifying herself as the author. So *My Immortal* can be said to end with Ebony abandoning her false self (Ebony Dark'ness Dementia Raven Way, the dark leaden substance of the *nigredo*), and becoming her divine higher Self (Rose Christo, the completed union of Rose and Cross, who exists on a higher plane than Ebony).

Most of the second part of *My Immortal* mirrors Part II of Faust. But its ending transcends the source material. The moral of Faust is that you can be redeemed. But *My Immortal* actively demonstrates the redemption that Faust can only point at. It follows the progress of Ebony from a contemptible character in a terrible fanfiction, through the alchemical process of uniting Rose and Cross, to become Rose Christo, a woman who lives in the real world.

It tells us that every one of us is a Mary Sue in the bad fanfiction of our lives – the narrative created by the ego in order to maintain our illusory selfhood. And in the tradition of other great alchemical texts like *The Chymical Wedding Of Christian Rosenkreutz* and *Faust*, it gives us a blueprint for escaping that fanfiction and completing the alchemical Work of breaking into reality.

As the end of *Faust* puts it:

> All of the transient,
> Is parable, only:
> The insufficient,
> Here, grows to reality:
> The indescribable,
> Here, is done.
> Woman, eternal
> Beckons us on.

Creationism, Unchallenged

How much should responsible news organizations report on stupid things?

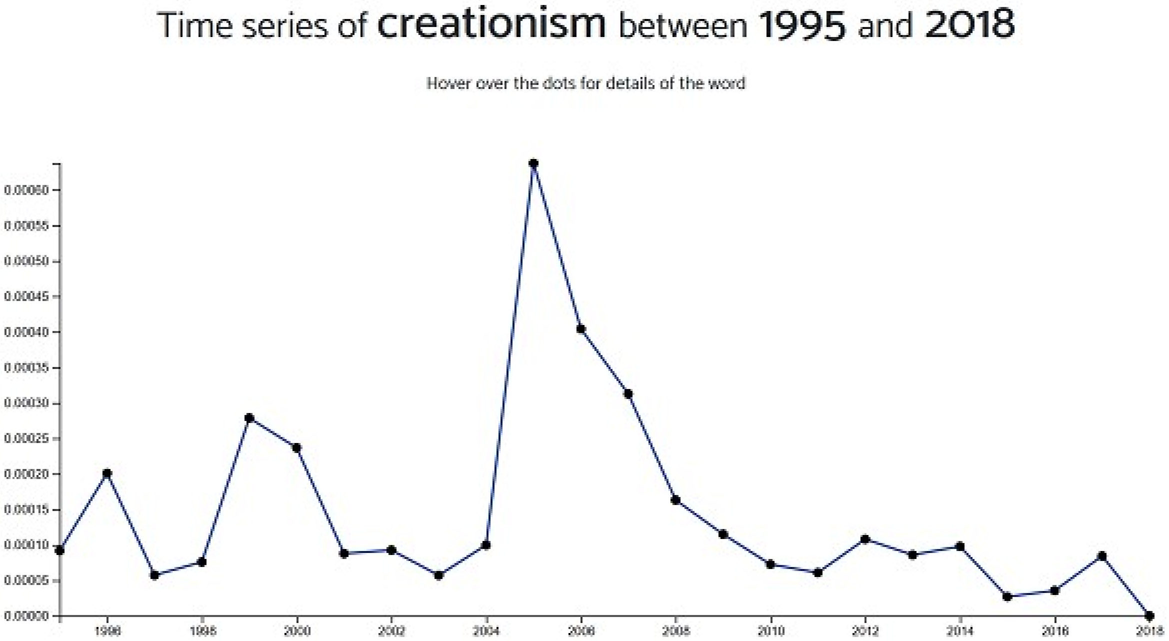
If they don't report at all, the stupid things go unchallenged. But if they report too much, then they signal-boost the stupid thing and give it free publicity (eg Donald Trump). Also, people who mistrust the media might reflexively support the stupid thing just because the media hates it (eg Donald Trump). Also, the more time you waste covering stupid things, the less time you have for real news (eg Donald Trump).

I recently read Causes And Consequences Of Mainstream Media Dissemination Of Fake News: Literature Review And Synthesis, which argues that the news might be covering too many stupid things right now. The authors note that "only 2.6% of visits to current affairs articles were to fake news websites" (though other sources suggest more) and that the mainstream press bears some responsibility for spreading inaccuracies beyond this small demographic. But they also understandably worry that maybe if the mainstream press wasn't so aggressive in covering and debunking fake news, then fake news would go uncorrected.

When I think about this problem, I remember creationism.

In the early 2000s, creationism was Public Enemy Number… maybe not One, but somewhere in the top ten. If you're old enough to remember the decade at all, you probably recall the key flashpoints. The Discovery Institute. Michael Behe. "Teach the controversy". The Creation Museum. *Of Pandas And People*. That one anti-Richard-Dawkins rap song which somehow despite everything managed to be really good.

And you probably remember the efforts by "the reality based community" to spread awareness of the dangers of creationism – the xkcd comics, the petitions by 1400 scientists named Steve, the New York Times articles:



Frequency of the word 'creationism' in the New York Times as a percent of all words, source here but currently down

…yeah, the 2000s were a weird time. I've talked about this particular conflict already in my post New Atheism: The Godlessness That Failed. Today I want to focus on another aspect.

All those creationists are still there. A 2019 Gallup poll found that 40% of Americans believed "God created human beings pretty much in their present form at one time within the last 10,000 years or so", little different from 44% who believed it when they first asked in 1983 or the 46% who believed it in 2006.

And the Discovery Institute! They're still there! You can read their very modern-looking website at discovery.org. Their blog has three new posts on creationism just from the past two days (eg A Hot Seller From Discovery Institute Press: New Book Offers Intelligent Design In A Nutshell).

Same with the Creation Museum! Last year they completed a $5.5 million upgrade, including a planetarium, "4D theater", and a snazzy "Biblical authority exhibit". And they still have their full-scale replica of Noah's Ark (looking more and more prescient these days).

Same with Michael Behe! He's still publishing! Last year he released his newest book, Darwin Devolves, which "gives a sweeping tour of how modern theories of evolution fall short and how the devolving nature of Darwin's mechanism limits them even further". Also, in case you wanted to read Behe's opinions about the coronavirus, that is a thing you can do.

As far as I can tell, the creationists are putting in just as much effort today as they did in 2006. But the mainstream went from fiercely challenging them, to totally ignoring them. And the change didn't help them at all. They haven't won any major victories, or convinced any more people. If anything, they're doing worse – nobody hears about them. Although the decline in media coverage hasn't prevented people from being creationist, it hasn't helped creationism spread or build clout either.

I see people using rivers of ink to fight the modern equivalents of creationists. Pizzagaters, flat-earthers, moon-hoaxers, QAnon, deep-staters, people who say the coronavirus is a bioweapon, Alex Jones. Are they sure it's not equally useless? Equally counterproductive?

Even beyond that, I see people willing to legitimize any tactic if it gives them a leg up on this group – censorship, social shaming, no-platforming, changing social media from a free public square to a carefully-monitored walled garden. Spreading the cowpox of doubt, teaching people to optimize for solving easy problems in ways that make it harder for them to think about the hard ones. The justification is always the same – if we don’t tighten control, then facts and science will lose out to bullshit and denialism, and fringe ideologies will burst into the mainstream and overwhelm it.

If that were the only way to save civilization from anti-science barbarism, maybe it would be a worthwhile trade. But the experience of unchallenged creationism suggests maybe we can relax.

Bush Did North Dakota

Continuing yesterday's discussion of fake news:

Guess et al says that 46% percent of Trump voters endorsed the Pizzagate conspiracy theory. Does this mean fake news is very powerful?

We can compare this to belief in various other conspiracy theories, as measured by the 2016 Chapman University Survey Of American Fears. About 24% believe there's a government conspiracy to cover up the truth about the moon landing, 30% about Obama's birth certificate, and 33% about the North Dakota crash.

This last one is especially interesting because there was no unusual crash in North Dakota when the survey was written. The researchers included it as a placebo option to see if people would endorse a conspiracy theory that didn't exist. 33% of them did.

Before we make fun of these people, consider: there's a strong presumption that surveys don't contain made-up questions. There was no "don't know" option included on the poll, just various shades of "agree" or "disagree". In order to condemn the people who "agreed" that the government was probably covering up the crash, we would have to assert that the more correct answer was "disagree". In other words, that people should have an assumption of trusting the government, until they get some specific reason to distrust it. You can make that argument, but it's not obvious. You could also start from the opposite assumption, where the government is guilty until proven innocent.

To put it another way, suppose I gave you the following survey:

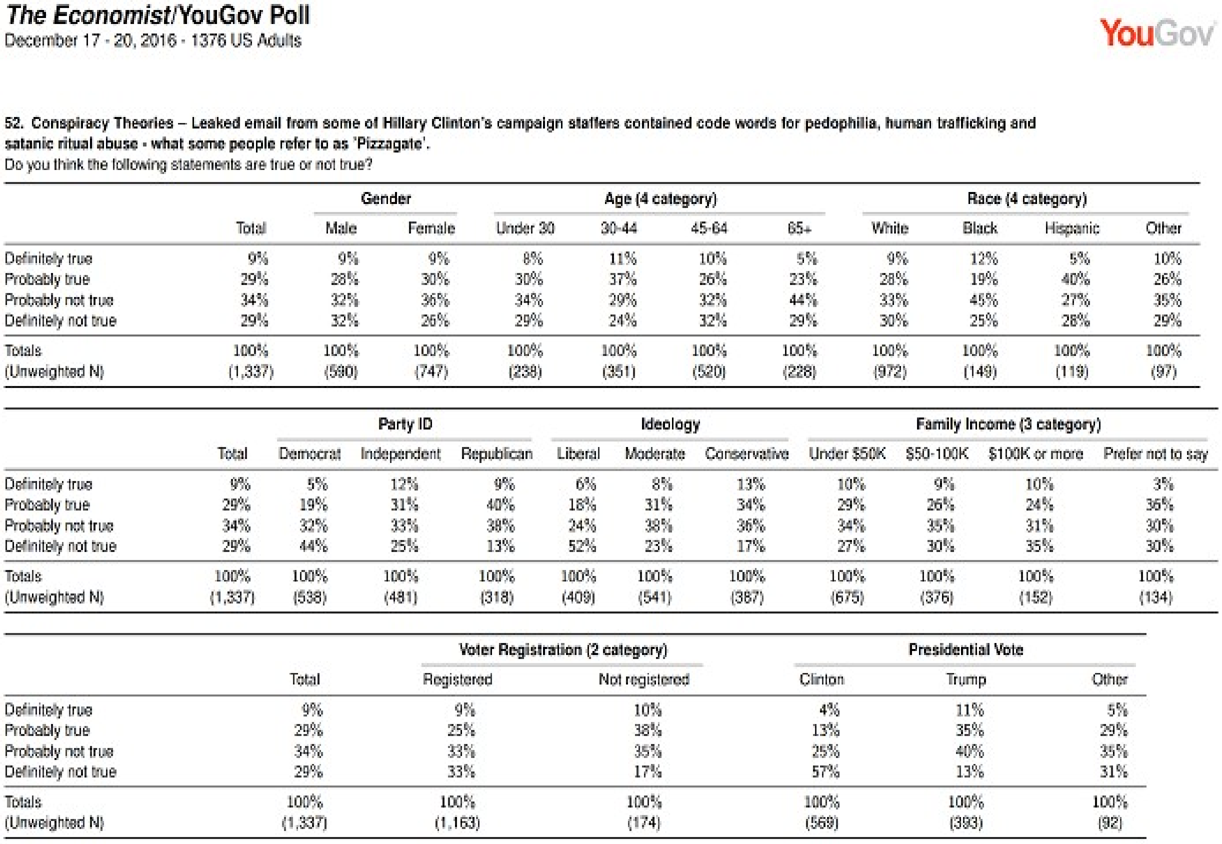
> SELECT AGREE OR DISAGREE, YOU MAY NOT SAY “DON’T KNOW” OR LOOK FOR MORE INFORMATION. Alex Jones is lying when he talks about:
>
> 1. Sandy Hook
> 2. The coronavirus
> 3. Obama’s birth certificate
> 4. The North Dakota crash

…many of us would guess he was lying about the North Dakota crash, without a second thought. And if there later turned out to be no North Dakota crash, we wouldn’t feel particularly ashamed; under the circumstances we made the right choice. If you think the government is as untrustworthy as Alex Jones, well, there you go.

I’ve previously talked about a lizardman constant of 4% on polls. That is, it’s hard to get a poll result much lower than four percent for anything, because of respondents making mistakes or trolling. If 4% of people supposedly believe something, that doesn’t mean we need to be concerned about that fraction of the population, it just means that poll has it its floor and it’s hard to conclude what the real number is.

In the same way, maybe we can posit a North Dakota constant of 33%. This is how many people believe in conspiracy theories when there’s no reason at all to believe them, not even the flimsy reasons conspiracy theories usually provide. Sometimes, if there’s a lot of evidence against them, fewer than 33% will believe in a given theory. But if it’s just “Conspiracy! True or false?” – 33% will say true.

Let’s look again at that statistic from the Guess paper – “46% of Trump voters believe”. I think their source is this poll, which finds:



The Economist/YouGov Poll
December 17 - 20, 2016 - 1376 US Adults

52. Conspiracy Theories – Leaked email from some of Hillary Clinton's campaign staffers contained code words for pedophilia, human trafficking and satanic ritual abuse - what some people refer to as 'Pizzagate'.
Do you think the following statements are true or not true?

		Gender		Age (4 category)				Race (4 category)			
	Total	Male	Female	Under 30	30-44	45-64	65+	White	Black	Hispanic	Other
Definitely true	9%	9%	9%	8%	11%	10%	5%	9%	12%	5%	10%
Probably true	29%	28%	30%	30%	37%	26%	23%	28%	19%	40%	26%
Probably not true	34%	32%	36%	34%	29%	32%	44%	33%	45%	27%	35%
Definitely not true	29%	32%	26%	29%	24%	32%	29%	30%	25%	28%	29%
Totals	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
(Unweighted N)	(1,337)	(590)	(747)	(238)	(351)	(520)	(228)	(972)	(149)	(119)	(97)

		Party ID			Ideology			Family Income (3 category)			
	Total	Democrat	Independent	Republican	Liberal	Moderate	Conservative	Under $50K	$50-100K	$100K or more	Prefer not to say
Definitely true	9%	5%	12%	9%	6%	8%	13%	10%	9%	10%	3%
Probably true	29%	19%	31%	40%	18%	31%	34%	29%	26%	24%	36%
Probably not true	34%	32%	33%	38%	24%	38%	36%	34%	35%	31%	30%
Definitely not true	29%	44%	25%	13%	52%	23%	17%	27%	30%	35%	30%
Totals	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
(Unweighted N)	(1,337)	(538)	(481)	(318)	(409)	(541)	(387)	(675)	(376)	(152)	(134)

		Voter Registration (2 category)		Presidential Vote		
	Total	Registered	Not registered	Clinton	Trump	Other
Definitely true	9%	9%	10%	4%	11%	5%
Probably true	29%	25%	38%	13%	35%	29%
Probably not true	34%	33%	35%	25%	40%	35%
Definitely not true	29%	33%	17%	57%	13%	31%
Totals	100%	100%	100%	100%	100%	100%
(Unweighted N)	(1,337)	(1,163)	(174)	(569)	(393)	(92)

Overall 38% of Americans agreed with the claim, so Trump voters (46%) were not outrageously more likely than anyone else. Other groups unrelated to ideology were about equally likely to believe it (eg 45% of Hispanics).

Like the North Dakota question, this one had no "unsure" or "what the hell are you talking about" option, forcing everyone to feign agreement or disagreement. We see that the majority of agreement is lukewarm. 75% of Trumpists and 85% of Hispanics who believe Pizzagate only "probably" rather than "definitely" believe it.

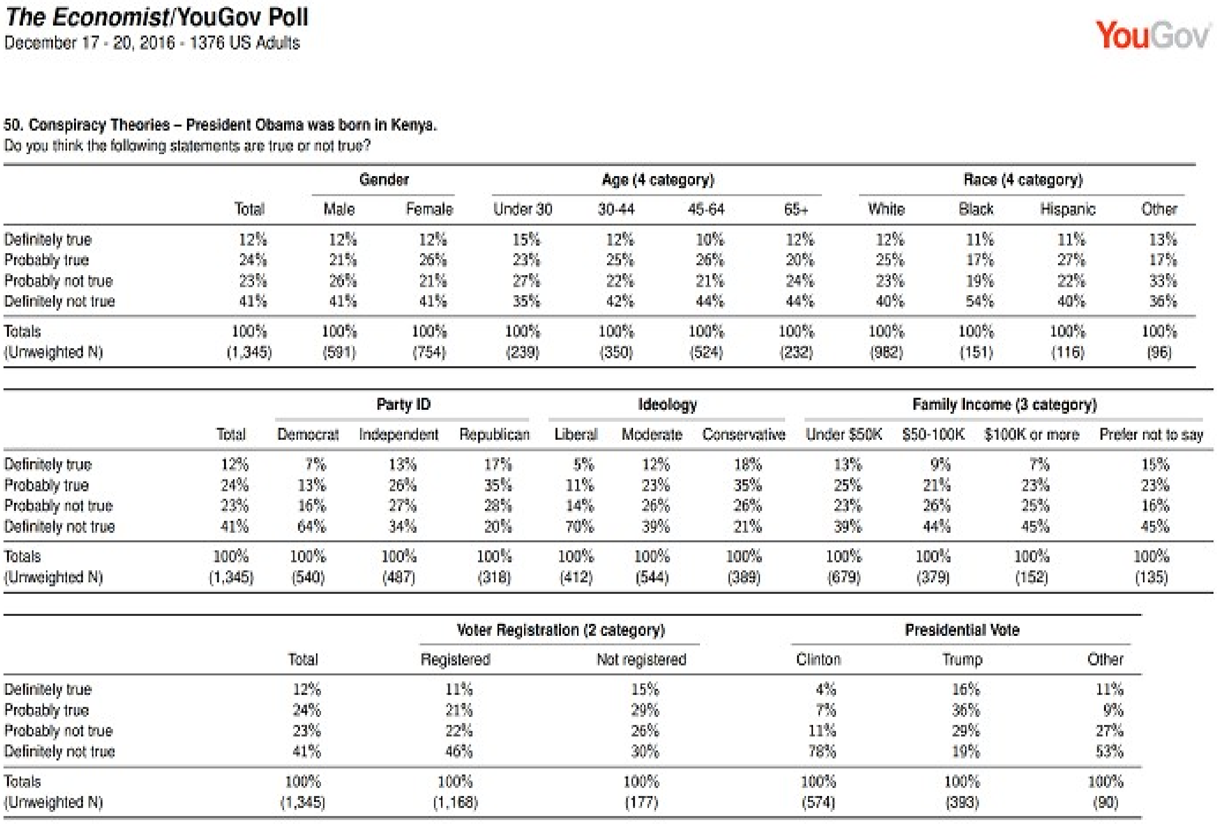
I don't think the evidence suggests Trump voters live in an outrageously different world from the rest of us. Instead, it suggests there's a North Dakota constant of 33% – the number of people who will believe a conspiracy theory for no particular reason. It looks like about 10 – 15% more Trump supporters than predicted believe Pizzagate, probably because it attacks Clinton, and 10 – 15% fewer Hillary supporters than predicted believe it. But these are relatively small effects, and equaled by eg whatever mysterious thing is going on with Hispanics. In any case, it all averages out to about the predicted amount.

Why is this North Dakota Constant of 33% so different from the Lizardman Constant of 4%? I don't know. Lizardmen seem like a pretty crazy conspiracy theory, but is Hillary's involvement with Satanic pizza parlors really that much less weird? Sure, Pizzagate is more politicized, and that might make some difference – but then how come a full 24% of Democrats believe it, six times more than Lizardman's Constant predicts?

One part of the story is that the lizardman poll offered "don't know", and 7% of people chose that. If, denied that option, those people would split evenly between yes and no, that brings us up to 7%ish pro-lizardman. But that's still nowhere near 33%.

I think this is probably a story about low-information voters. If you imagine you've never heard about Pizzagate, and you read the question as written, it doesn't sound too outlandish. Some Clinton staffers' emails contained some code words. The pedophilia and Satanic abuse are pretty out there, but post-Jeffrey Epstein we all assume somebody's doing some kind of creepy pedophilia stuff somewhere. Maybe if you don't know anything about this, and you don't have the strong priors about Satanic ritual abuse that you get from studying the history of those claims in the 80s and 90s, this one seems like a toss-up. Certainly it seems like *more* of a toss-up than a clearly-stated assertion that reptilian aliens rule the world. If your prior is "most conspiracy theory-ish things are probably true”, this sounds like the kind of thing that could be true, whereas you might balk at the lizardman statement.

Here’s another question from the same poll:



The Economist/YouGov Poll
December 17 - 20, 2016 - 1376 US Adults

YouGov

50. Conspiracy Theories – President Obama was born in Kenya.
Do you think the following statements are true or not true?

		Gender		Age (4 category)				Race (4 category)			
	Total	Male	Female	Under 30	30-44	45-64	65+	White	Black	Hispanic	Other
Definitely true	12%	12%	12%	15%	12%	10%	12%	12%	11%	11%	13%
Probably true	24%	21%	26%	23%	25%	26%	20%	25%	17%	27%	17%
Probably not true	23%	26%	21%	27%	22%	21%	24%	23%	19%	22%	33%
Definitely not true	41%	41%	41%	35%	42%	44%	44%	40%	54%	40%	36%
Totals	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
(Unweighted N)	(1,345)	(591)	(754)	(239)	(350)	(524)	(232)	(982)	(151)	(116)	(96)

		Party ID			Ideology			Family Income (3 category)			
	Total	Democrat	Independent	Republican	Liberal	Moderate	Conservative	Under $50K	$50-100K	$100K or more	Prefer not to say
Definitely true	12%	7%	13%	17%	5%	12%	18%	13%	9%	7%	15%
Probably true	24%	13%	26%	35%	11%	23%	35%	25%	21%	23%	23%
Probably not true	23%	16%	27%	28%	14%	26%	26%	23%	26%	25%	16%
Definitely not true	41%	64%	34%	20%	70%	39%	21%	39%	44%	45%	45%
Totals	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
(Unweighted N)	(1,345)	(540)	(487)	(318)	(412)	(544)	(389)	(679)	(379)	(152)	(135)

		Voter Registration (2 category)		Presidential Vote		
	Total	Registered	Not registered	Clinton	Trump	Other
Definitely true	12%	11%	15%	4%	16%	11%
Probably true	24%	21%	29%	7%	36%	9%
Probably not true	23%	22%	26%	11%	29%	27%
Definitely not true	41%	46%	30%	78%	19%	53%
Totals	100%	100%	100%	100%	100%	100%
(Unweighted N)	(1,345)	(1,168)	(177)	(574)	(393)	(90)

Who believes Obama was secretly born in Kenya? Lots of people – including 28% of blacks. I’ve been told again and again that birtherism is a racist conspiracy theory and no person could possibly believe it except as a way of dog whistling white supremacy. Yet here we are with 28% of blacks supporting it – and this isn’t a small sample either! I have no idea what these people are thinking, except that 28% is pretty close to the North Dakota Constant and maybe we should just write this one off.

I conclude we probably shouldn’t draw too many conclusions from specific statements like “X% of [GROUP] supports [CONSPIRACY THEORY]”, especially if X is around 33%. It’s probably just the North Dakota Constant. Likewise, we shouldn’t interpret Pizzagate’s high polling numbers as much evidence that fake news is very convincing – though you could still make an argument that fake news plays a role in transmitting believable conspiracy theories to people who are predisposed to believe them.

Of course, there are some high-information voters who still believe these things really strongly. I think they deserve a more complete treatment, which I want to give later. I think a preliminary sketch might look like: if you start with a prior on something being true, you don’t necessarily need much evidence. The North Dakota question suggests that conspiracy theorists start with a high prior on any given conspiracy being true. What remains to be explained is why some people stick to that prior even after they get more information.

Book Review: Origin Of Consciousness In The Breakdown Of The Bicameral Mind

I.

Julian Jaynes' *The Origin Of Consciousness In The Breakdown Of The Bicameral Mind* is a brilliant book, with only two minor flaws. First, that it purports to explains the origin of consciousness. And second, that it posits a breakdown of the bicameral mind. I think it's possible to route around these flaws while keeping the thesis otherwise intact. So I'm going to start by reviewing a slightly different book, the one Jaynes should have written. Then I'll talk about the more dubious one he actually wrote.

My hypothetical Jaynes 2.0 is a book about theory-of-mind. Theory-of-mind is our intuitive model of how the mind works. It has no relation to intellectual theories about how the mind is made of cognitive algorithms or instantiated on neurons in the brain. Every schoolchild has a theory-of-mind. It usually goes like this: *the mind is an imaginary space containing things like thoughts, emotions, and desires. I have mine and you have yours. I can see what's inside my mind, but not what's inside your mind, and vice versa. I mostly choose the things that are in my mind at any given time: I will thoughts to happen, and they happen; I will myself to make a decision, and it gets made. This needs a resource called willpower; if I don't have enough willpower, sometimes the things that happen in my mind aren't the ones I want. When important things happen, sometimes my mind gets strong emotions; this is natural, but I need to use lots of willpower to make sure I don't get overwhelmed by them and make bad decisions.*

All this seems so obvious to most people that it sounds like common sense rather than theory. It isn't; it has to be learned. Very young children don't start out with theory of mind. They can't separate themselves from their emotions; it's not natural for them to say "I'm really angry now, but that's just a thing I'm feeling, I don't actually hate you". It's not even clear to them that people's minds contain different things; children are famously unable to figure out that a playmate who has different evidence than they do may draw different conclusions.

And the learning isn't just a process of passively sitting back observing your own mind until you figure out how it works. You learn it from your parents. Parents are always telling their kids that "I think this" and "What do you think?" and "You look sad" and "It makes me feel sad when you do that". Eventually it all sinks in. Kids learn their parent's theory-of-mind the same way they learn their parents' language or religion.

When in human history did theory-of-mind first arise? It couldn't have been a single invention – more like a gradual process of refinement. "The unconscious" only really entered our theory-of-mind with Freud. Statements like "my abuse gave me a lot of baggage that I'm still working through" involves a theory-of-mind that would have been incomprehensible a few centuries ago. It's like "I'm clicking on an icon with my mouse" – every individual word would have made sense, but the gestalt would be nonsensical.

Still, everyone always assumes that the absolute basics – mind as a metaphorical space containing beliefs and emotions, people having thoughts and making decisions – must go back so far that their origins are lost in the mists of time, attributable only to nameless ape-men.

Julian Jaynes doesn't think that. He thinks it comes from the Bronze Age Near East, c. 1500 – 750 BC.

II.

Jaynes (writing in the 1970s) was both a psychology professor at Princeton and an expert in ancient languages, so the perfect person to make this case. He reviews various samples of Bronze Age writing from before and after this period, and shows that the early writings have no references to mental processes, and the later ones do. When early writings do have references to mental processes, they occur in parts agreed by scholars to be later interpolations. If, with no knowledge of the language itself, you tried to figure out which parts of a heavily-redacted ancient text were early vs. late by their level of reference to mental processes, you could do a pretty decent job.

I don't speak fluent Sumerian, so I am forced to take Jaynes' word for a lot of this. It's even worse than that, because Jaynes argues that other translators sometimes err and translate non-mental terms in mental ways. This is an easy mistake for them to make, because most cultures, once they got theory of mind, repurposed existing language to represent it. Jaynes makes a convincing case for why this would happen, and convincingly argues for why his interpretations are truer to the spirit of the text, but it does mean you can't double-check his work by reading the works in translation.

Jaynes spends the most time talking about the *Iliad*, with good reason – it's the longest Bronze Age work we have, and in many ways it's a psychodrama, focusing as much on the characters of Achilles, Hector, etc as the plot itself. It came together piecemeal through the efforts of Greek bards between about 1100 and 800 BC, finally reaching a canonical version in the mouth of “Homer” around 700 BC – the period Jaynes says theory of mind was starting to evolve. Jaynes uses it to trace the development process, showing how older sections of the *Iliad* treat psychology in different ways than newer ones.

So for example, a typical translation might use a phrase like “Fear filled Agamemnon’s mind”. Wrong! There is no word for “mind” in the *Iliad*, except maybe in the very newest interpolations. The words are things like *kardia*, *noos*, *phrenes*, and *thumos*, which Jaynes translates as heart, vision/perception, belly, and sympathetic nervous system, respectively. He might translate the sentence about Agamemnon to say something like “Quivering rose in Agamemnon’s belly”. It still means the same thing – Agamemnon is afraid – but it’s how you would talk about it if you didn’t have an idea of “the mind” as the place where mental things happened – you would just notice your belly was quivering more. Later, when the Greeks got theory of mind, they repurposed all these terms. You can still find signs of this today, like how we say “I believe it in my heart”. In fact, you can still find this split use of *phrenes*, which has survived into English both as the phrenic nerve (a nerve in the belly) and schizophrenia (a mental disease). As the transition wore on, people got more and more flowery about the kind of feelings you could have in your belly or your heart or whatever, until finally belly, heart, and all the others merged into a single Mind where all the mental stuff happened together.

The *Iliad* uses these body parts to describe feelings despite its weak theory of mind. Its solution for describing thoughts and decision-making is more…unconventional.

Suppose Achilles is overcome with rage and wants to kill Agamemnon. But this would be a terrible [idea]; after [thinking] about it for a while, he [decides] against. If Achilles has no concept of any of the bracketed words, nothing even slightly corresponding to those terms, how does he conceptualize his own actions? Jaynes:

> The response of Achilles begins in his *etor*, or what I suggest is a cramp in his guts, where he is in conflict, or put into two parts (*mermerizo*) whether to obey his *thumos*, the immediate internal sensations of anger, and kill the king, or not. It is only after this vacillating interval of increasing belly sensations and surges of blood, as Achilles is drawing his mighty sword, that the stress has become sufficient to hallucinate the dreafully gleaming goddess Athene who then takes over control of the action and tells Achilles what to do.

Wait, *what?*

III.

As you go about your day, you hear a voice that tells you what to do, praises you for your successes, criticizes you for your failures, and tells you what decisions to make in difficult situations. Modern theory-of-mind tells you that this is your own voice, thinking thoughts. It says this so consistently and convincingly that we never stop to question whether it might be anything else.

If you don't have theory of mind, what do you do with it? Children don't have theory of mind, at least not very much of it, and more than half of them have imaginary friends. Jaynes has done some research on the imaginary friend phenomenon, and argues that a better term would be "hallucinatory friend" – children see and hear these entities vividly. The atheoretical mind is a desperate thing, and will comply with any priors you give it to make sense of its experiences. If that prior is that the voice in your head is a friend – or god – it will obediently hallucinate a friend or god for you, and modulate its voice-having accordingly.

I know some very smart and otherwise completely sane evangelical Christians who swear to me that God answers their prayers. They will ask God a question, and they will hear God's voice answer it. God's voice may not sound *exactly* like an external voice, and it may give them only the advice they would have given themselves if they'd thought about it – but they swear that they are not thinking about it, that their experience is qualitatively different than that. And these are normal people! If you're a special person – a saint or mystic, say – and you actively court the experience by fasting and praying and generally stressing your body to the limit – then the voice will be that much louder and more convincing.

There are even whole forms of therapy based on this kind of thing. In Internal Family Systems, the therapist asks the patient to conceptualize some part of their mind (maybe the part that's producing a certain symptom) as a person, and to talk to it. I know people who swear that this works. They approach their grief or anger or anxiety, and they get a clear image of what "he" or "she" looks like, and then "he" or "she" talks to them. Usually he/she tells them some appropriately psychological sounding thing, like "Hello, I am your anxiety, and I'm only inflicting these fears on you because we were abused as a child and I want to make sure nobody ever abuses us like that again". Then the patient talks to their anxiety and hopefully strikes a bargain where the patient agrees to take the anxiety's perspective into account and the anxiety agrees not to make the patient so anxious all the time. Some people swear by this, say it's helped them where nothing else can, and absolutely insist they are having a real dialogue with their anxiety and not just making up both sides of the conversation.

Most of the people who seem to really like IFS have borderline personality disorder. And borderline people are also at the most risk of dissociative identity disorder (multiple personality). Multiple personality has two main risk factors: borderline, and somebody suggesting to you that multiple personality disorder might be a reasonable thing to have. For a while in the 80s, psychiatrists were really into multiple personality and tried diagnosing everyone with it, and sure enough all those people would admit to having multiple personalities and it would be very exciting. Then the APA told the psychiatrists to stop, people stopped talking about multiple personality as much, and now the condition is rarer.

A few years ago, someone rediscovered/invented tulpamancy, the idea of cultivating multiple personalities on purpose because it’s cool. People who try to do this usually succeed. At least they say they’ve succeeded, and I believe they think this. I think their internal experience is of talking to a different entity inside of them. Also, I have a friend who writes novels, and one time she created such a detailed mental model of one of her characters that it became an alternate personality, which she still has and considers an important part of her life. She is one of the most practical people I know and not usually prone to flights of fancy.

I also have less practical friends, friends who are into occultism. They tell me they sometimes make contact with spiritual entities. I believe them when they say they have these experiences. I believe them when they say that they were not purposely guiding their Ouija board to say whatever it said. I don’t have any friends who are cool enough to have gone through the whole procedure for summoning your Holy Guardian Angel, but from what I read, completing the ritual directly does tend to leave you with an angel who hangs around you and gives you advice. I believe the people who say this is their experience of completing the ritual.

I conclude that giving yourself multiple personalities is actually pretty easy under the right circumstances. Those circumstance are a poor theory of mind (I think borderlines are naturally bad at this) and a cultural context in which having a multiple personality is expected.

Jaynes says ancient people met both criteria. They had absolutely no theory of mind, less theory of mind than the tiniest child does today. And their cultural context was absolutely certain that gods existed. Just as we teach our children that the voice in their mind is them thinking to theirselves, so the ancients would teach their children that the voice in their head was a god giving them commands. And the voice would dutifully mold itself to fit the expected role.

Here Jaynes is at his most brilliant, going through ancient texts one by one, noting the total lack of mental imagery, and highlighting the many everyday examples of conversations with gods. Every ancient culture has near-identical concepts of a god who sits inside of you and tells you what to do. The Greeks have their *daemons*, the Romans their *genii*, the Egyptians their *ka* and *ba*, and the Mesopotamians their *iri*. The later you go, the more metaphorically people treat these. The earlier you go, the more literal they become. Go early enough, and you find things like the Egyptian Dispute Between A Man And His Ba which is just a papyrus scroll about a guy arguing loudly with the hallucinatory voice of his guardian spirit, and the guardian spirit’s hallucinatory voice arguing back, and nobody thinking any of this is weird (people who aren't Jaynes would wimp out and say this is "metaphorical"). Every ancient text is in complete agreement that everyone in society heard the gods' voices very often and usually based decisions off of them. Jaynes is just the only guy who takes this seriously.

Turn on what Terry Pratchett called "first sight and second thoughts" and try to look at the Bronze Age with fresh eyes. It was *really* weird. People would center their city around a giant ziggurat, the "House of God", with a giant idol within. They would treat this idol exactly like a living human – feeding it daily, washing it daily, sometimes even marching it through the streets on sedan chairs carried by teams of slaves so it could go on a "connubial visit" to the temple of an idol of the opposite sex! When the king died, hundreds of thousands of men would labor to build him a giant tomb, and then they would kill a bunch of people to serve him in the afterlife. Then every so often it would all fall apart and everyone would slink away into the hills, trying to pretend they didn't spend the last twenty years buliding a jeweled obelisk so some guy named Ningal-Iddida could boast about how many slaves he had.

If the Bronze Age seems kind of hive-mind-y, Julian Jaynes argues this is because its inhabitants weren't quite individuals, at least not the way we think of individuality. They were in the same kind of trance as a schizophrenic listening to voices commanding him to burn down the hospital. All of it – the ziggurats, the obelisks, the pyramids – were an attempt to capture not individual humans, but those humans' daemons – to get people to identify the voice in their head with the local deity, and replace their free will with a hallucinatory god who *represented their mental model of society's demands on them*. In the best case scenario, the voice would be interpreted as the god-king himself, giving you orders from miles away. Jaynes argues the Bronze Age was obsessed with burials and the afterlife (eg the Pyramids) because if you had internalized the voice in your head as Pharaoh Cheops, the voice wasn't going to go away just because the actual Pharaoh Cheops had died hundreds of miles away in the capital. So even after Pharaoh Cheops dies, as far as all his subjects can tell, he's still around, commanding them from the afterlife. So they had better keep him really, really happy, just as they did during life. Jaynes presents various pieces of evidence that the main function of pyramids was as a place where you could go to commune with the dead Pharaoh's spirit – ie ask it questions and it would answer them.

He has a similar explanation for idols. The Bronze Age loved idols. There were the giant idols, ones that made the statue of Zeus at Olympia look like a weak effort. But also, every family had their own individual idols. Archaeologists who dig up Bronze Age houses just find idol after idol after idol, like the ancient Sumerians did nothing except stare at idols all day. Jaynes thinks this is approximately true. Idols were either cues to precipitate hallucinatory voices, or else just there to make conversation more comfortable – it's less creepy if you can see the person you're talking to, after all.

Then, around 1250 BC, this well-oiled system started to break down. Jaynes blames trade. Traders were always going into other countries, with different gods. These new countries would be confusing, and the traders' hallucinatory voices wouldn't always know all the answers. And then they would have to negotiate with rival merchants! Here theory of mind becomes a huge advantage – you need to be able to model what your rival is thinking in order to get the best deal from him. And your rival also wants theory of mind, so he can figure out how to deceive you. Around 1250 BC, trade started picking up, and these considerations became a much bigger deal. Then around 1200 BC, the Bronze Age collapsed. It’s still not exactly clear why (some of you may have heard me give a presentation on this), though most guesses involve a combination of climate change plus the mysterious Sea Peoples. Whole empires were destroyed, their populations become refugees who flooded the next empire in turn. Now *everyone* was in unfamiliar territory; *nobody* had all the answers. The weird habits of mind a couple of traders had picked up became vital; people adopted them or died.

But as theory of mind spread, the voices of the gods faded. They receded from constant companions, to only appearing in times of stress (the most important decisions) to never appearing at all. Jaynes interprets basically everything that happened between about 1000 BC and 700 BC as increasingly frantic attempts to bring the gods back or deal with a godless world.

Now, to be fair, he cites approximately one zillion pieces of literature from this age with the theme “the gods have forsaken us” and “what the hell just happened, why aren’t there gods anymore?” As usual, everyone else wimps out and interprets these metaphorically – claiming that this was just a poetic way for the Mesopotamians to express how unlucky they felt during this chaotic time. Jaynes does not think this was a metaphor – for one thing, people have been unlucky forever, but the 1000 – 750 BC period was a kind of macabre golden age for “the gods have forsaken us” literature. And sometimes it seems oddly, well, on point:

> My god has forsaken me and disappeared
> My goddess has failed me and keeps at a distance
> The good angel who walked beside me has departed.

Or:

> One who has no god, as he walks along the street
> Headache envelops him like a garment

Jaynes says that “there is no trace whatsoever of any such concerns in any literature previous to the texts I am describing here”.

So people got desperate. He says this period was the origin of augury and divination. Omens “were probably present in a trivial way” before this period, but not very important; “there are, for example, no Sumerian omen texts whatsoever”. But after about 1000 BC, omens become an international obsession.

> Towards the end of the second millennium BC…such omen texts proliferate everywhere and swell out to touch almost every aspect of life imaginable. By the first millennium BC, huge collections of them are made. In the library of King Ashurbanipal at Nineveh about 650 BC, at least 30% of the twenty to thirty thousand tablets come into the category of omen literature. Each entry in these tedious irrational collections consists of an if-clause or protasis followed by a then-clause or apodosis. And there were many classes of omens…
>
> – If a town is set on a hill, it will not be good for the dweller within that town.
>
> – If black ants are seen on the foundations which have been laid, that house will get built; the owner of that house will live to grow old.
>
> – If a horse enters a man’s house, and bites either an ass or a man, the owner of the house will die, and his household will be scattered.

> – If a fox runs into the public square, that town will be devastated.
>
> – If a man unwittingly treads on a lizard and kills it, he will prevail over his adversary.

And then there are the demons. Early Sumerians didn’t really worry about demons. Their religion was very clear that the gods were in charge and demons were impotent. Post 1000 BC, all of this changes.

> As the gods recede…there whooshes into this power vacuum a belief in demons. The very air of Mesopotamia became darkened with them. Natural phenomena took on their characteristics of hostility toward men, a raging demon in the sandstorm sweeping the desert, a demon of fire, scorpion-men guarding the rising sun beyond the mountains, Pazuzu the monstrous wind demon, the evil Croucher, plague demons, and the horrible Asapper demons that could be warded off by dogs. Demons stood ready to seize a man or woman in lonely places, while sleeping or eating or drinking, or particularly at childbirth. They attached themselves to men as all the illnesses of mankind. Even the gods could be attacked by demons, and this sometimes explained their absence from the control of human affairs…
>
> Innumerable rituals were devoutly mumbled and mimed all over Mesopotamia throughout the first millennium B.C. to counteract these malign forces. The higher gods were beseeched to intercede. All illnesses, aches, and pains were ascribed to malevolent demons until medicine became exorcism. Most of our knowledge of these antidemoniac practices and their extent comes from the huge collection made about 630 B.C. by Ashurbanipal at Nineveh. Literally thousands of extant

> tablets from this library describe such exorcisms, and thousands more list omen after omen, depicting a decaying civilization as black with demons as a piece of rotting meat with flies.

…and angels, and prophets, and all the other trappings of religion. When the gods spoke to you every day, and you couldn’t get rid of them even if you wanted to, angels – a sort of intermediary with the gods – were unnecessary. There was no place for prophets – when everyone is a prophet, nobody is. There wasn’t even prayer, at least not in a mystical sense – as Jaynes puts it, “schizophrenics do not beg to hear their voices – it is unnecessary – in the few case where this does happen, it is during recovery when the voices are no longer heard with the same frequency.”

The Assyrians invented the idea of Heaven. Previously, Heaven had been unnecessary. You could go visit your god in the local ziggurat, talk to him, ask him for advice. But word went around that gods had retreated to heaven – some of the stories even use those exact words, blaming the Great Flood or some other cataclysm. The ziggurats shifted from houses for the gods to *e-temen-an-ki* – pedestals that the gods could descend to from Heaven, should they ever wish to return.

By 500 BC, the ability to hear the gods was limited to a few prophets, oracles, and poets. Jaynes is especially interested in this last group – he cites various ancient sources claiming that the poets only transcribe what they hear gods and goddesses sing to them (everyone else wimps out and says this is metaphorical). For Jaynes, the *Iliad* starts “Sing, O Muse…” because the poet was expecting a hallucinatory Muse to actually appear beside him and start singing, after which he would repeat the song to his listeners as a sort of echolalia.

Jaynes ends by referencing one of my favorite ancient texts, Plutarch's *On The Failure Of Oracles*. Plutarch, writing around 100 AD, is not a skeptic. He believes oracles work *in theory*. But he records a general consensus that they don't work as well as they used to, and that some day soon they will stop working at all. Jaynes believes that as the theory-of-mind waterline rises, fewer and fewer people hear the voices of the gods. By the Golden Age of Greece, it was so difficult that only a few specially selected people placed in specially numinous locations could manage – the oracles. By Plutarch's own time, even those people could barely manage.

The last oracle to fade away was the greatest – Delphi, perched atop a fantastic gorge as if suspended between Heaven and Earth. Jaynes tries to give us an impression of how important it was in its time; important people from all over the classical world would make the pilgrimage there, leave lavish gifts, and ask Apollo for advice on weighty matters. He thinks that the oracle's fame protected it; if a cultural validation is an important ingredient in god-hearing, Delphi had the strongest and best. Its reputation was unimpeachable. Still, in the centuries after Plutarch, its prophecies became rarer and rarer; the Pythia's few divine utterances became separated by more and more incoherent raving. Finally:

> As part of [the Emperor Julian's] personal quest for authorization, he tried to rehabiliate Delphi in AD 363, three years after it had been ransacked by Constantine. Through his remaining priestess, Apollo prophecied that he would never prophesy again. And the prophecy came true.

IV.

The real *Origin Of Consciousness In The Breakdown Of The Bicameral Mind* is like my edited version above, except that wherever I say “theory of mind”, it says “consciousness”.

Jaynes has obviously thought a lot about this, and he’s a psychology professor so I’m sure he’s heard of theory of mind. Still, I am *so* against this choice. Consciousness means so many different things to so many different people, and none of them realize they’re talking past each other, and it’s such a loaded term that any argument including it is basically guaranteed to veer off into the fantastic.

Did he literally believe that the Sumerians, Homeric Greeks, etc were p-zombies? That there was nothing that it was like to be them? That they took in photons and emitted actions but experienced no “mysterious redness of red”? I cannot be completely sure. At times he refers to Bronze Age people as “automatons”, which seems like a pretty final judgment. But he also treats them as genuinely hearing, seeing, and having feelings about the hallucinatory gods who appear to them. The god-human interaction seems like it involves the human being at least minimally conscious. But if Jaynes has a coherent theory here, I must have missed it.

I think he is unaware of (or avoiding) what we would call “the hard problem of consciousness”, and focusing on consciousness entirely as a sort of “global workspace” where many parts of the mind come together and have access to one another. In his theory, that didn’t happen – the mental processing happened and announced itself to the human listener as a divine voice, without the human being aware of the intermediate steps. I can see how “consciousness” is one possible term for this area, if you didn’t mind confusing a lot of people. But seriously, just say “theory of mind”.

Jaynes seems aware of this objection, which he summarizes as “the Bronze Agers did not lack consciousness, they just lacked the *concept* of consciousness”. His retort is that in some cases, the concept of a thing is the same as the thing itself – he suggests baseball as an example. This seems a little sophistic to me. If somebody told me that Mandarin Chinese doesn’t have a word for “consciousness”, I would be surprised but not stunned – it seems like a strange word for a rich and ancient language to lack, but weirder things have happened. If somebody told me that Chinese people didn’t even have the *concept* of consciousness until it was introduced from the West, that wouldn’t shock me either – sometimes I think half of philosophers don’t even have the same concept of consciousness I do, and I can imagine the Chinese carving up the world in very different ways. But if someone told me that Chinese people *were not conscious*, I would dismiss them as a crank. So I can’t accept that having consciousness and having a concept of consciousness are exactly the same thing, and I continue to think “theory of mind” is better here.

The other major difference between my rewrite and Jaynes’ real book is that Jaynes focuses heavily on “bicamerality” – the division of the brain into two hemispheres. He believes that in the Bronze Age mind design, the left hemisphere was the “mortal” and the right hemisphere the “god” – ie the hallucinatory voice of the god was the right hemisphere communicating information to the left hemisphere. In the modern mind design, the two hemispheres are either better integrated, or the right hemisphere just doesn’t do much.

I am not an expert in functional neuroanatomy, but my impression is that recent research has not been kind to any theories too reliant on hemispheric lateralization. While there are a few well-studied examples (language is almost always on the left) and a few vague tendencies (the right brain sort of seems to be more holistic, sometimes), basically all tasks require some input from both sides, there's little sign that anybody is neurologically more "right-brained" or "left-brained" than anyone else, and most neuroscientific theories don't care that much about the right-brain left-brain distinction. Also, Michael Gazzaniga's groundbreaking work on split-brain patients which got everyone excited about hemispheres and is one of the cornerstones of Jaynes' theory doesn't replicate. Also, Jaynes says his theory implies that schizophrenic hallucinations come from the language centers of the right hemisphere, and I think the latest fMRI evidence is that they don't.

(Also, Jaynes says his theory implies that demonic possession occurs in the right hemisphere. But some absolute madman actually put a possessed women in an fMRI machine and then exorcised her while the machine was running and although it showed some odd deficiencies in interhemispheric communication, it didn't seem to show unusual right hemisphere activity. Imagine having to write that IRB application!)

I don't think either of these issues fundamentally changes Jaynes' theory. Just switch "consciousness" to "theory of mind", and change the psychiatry metaphor from split-brain patients to dissociative-identity patients, and you're fine.

V.

But there's another class of problem that Jaynes' theory doesn't survive nearly as well: what about Australian Aborigines?

Or American Indians, or Zulus, or Greenland Inuit, or Polynesians, or any other human group presumably isolated from second-millennium-BC Assyrians until anthropologists got a chance to examine them? If consciousness is an invention, and it didn't spread to these groups, did these groups have it? If so, how? If not, why aren't they hallucinating gods all the time?

I mean, some of these groups definitely have shamans and medicine men. I'm not saying none of them ever hear gods. But Jaynes claims Bronze Agers heard gods literally all the time, as a substitute for individual thought. Nothing I've heard from these people or the anthropologists who study them suggest anything like this is true. And these people also seem to be able to strategically deceive others, another key consciousness innovation Jaynes says Bronze Agers lacked. Or at least I assume I would have heard about it from some anthropologist if they weren't.

I don't have a good sense of how Jaynes would answer this objection. The most relevant part of the book is around page 135. Jaynes argues that bicamerality (his term for the hallucinatory gods) started with agriculture in the Bronze Age Near East, though there were scattered hallucinations before then. So plausibly the Inuit, aborigines, etc, were not bicameral. They are in a pre-bicameral state, where they have neither full subjective consciousness, nor clear hallucinations of gods. They may have flashes of both, or do something else entirely, or just be *blank*. Or something. The point is, if they were perfectly normal conscious people like us, then Jaynes is wrong about everything.

Maybe I've done some violence to Jaynes' theory by rounding it off to "theory of mind" and emphasizing it as an invented technology? But he tries to really emphasize the inventedness of it in the first few chapters, talking about how it had to be built up by layer upon layer of well-chosen metaphor. As far as I can tell I relayed that part faithfully.

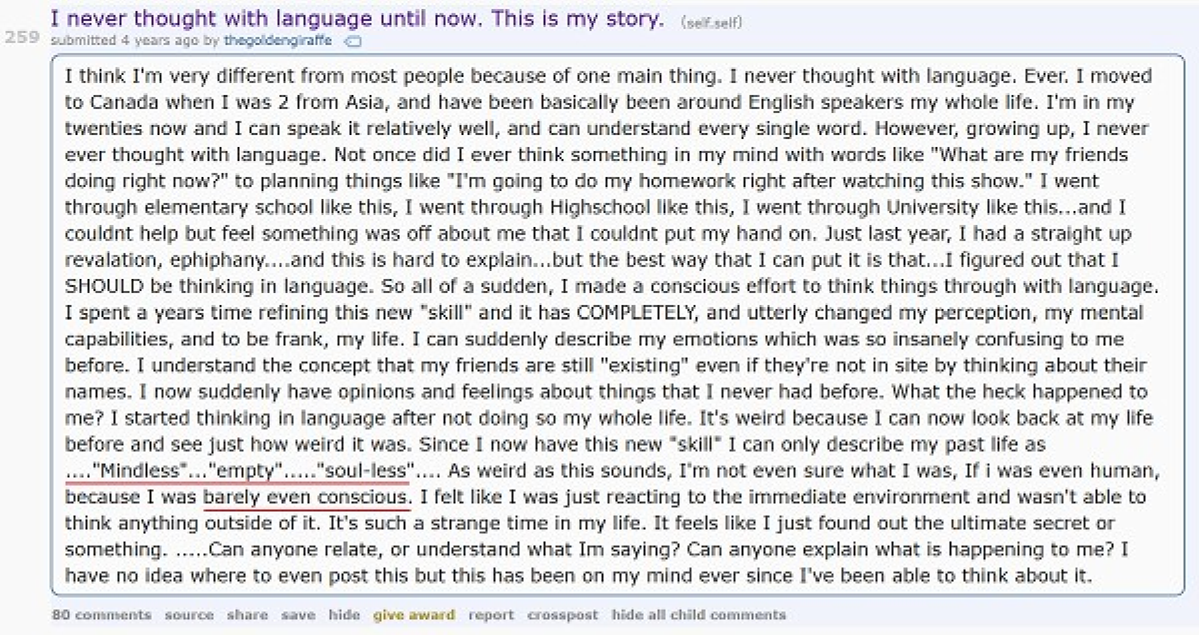
And I’m looking at the bulletin board on julianjaynes.org, and there’s a post by someone who met Jaynes before he died and asked him this question. They write:

> On the About Julian Jaynes page it says he gave a lecture at the Wittgenstein Symposium in [Kirchberg]. I was there. It was a wonderful lecture. It is a pity that his work has not had a deeper impact. I still believe he was basically right (and certainly his prose was brilliant).
>
> I did ask him, by the way, whether he thought it possible that the Aborigines in Australia were not conscious as late as the 18th or 19th century. He said he was not sure and that it would be worthwhile to investigate. Well, I never did and probably no one else [did].

So I don’t think I am misunderstanding him by making this criticism, and it sounds like he just bites the bullet and says maybe this was true. The main position on the forum seems to be that anthropologists weren’t asking the right questions as soon as they met uncontacted tribes, and so maybe they would have missed this. I find this hard to believe. It should be *really* easy to notice, and also the process of them learning Western theory of mind should leave some scars – at least one of them should say something like “that couple-year period when we all stopped hallucinating gods and became conscious – that was a *weird* time.”

Jaynes partisans are able to come up with a few anthropological works suggesting that the minds of primitive people are pretty weird, and I believe that, but they don’t seem quite as weird as Jaynes wants them to be. So the question becomes whether we would notice if some people worked in a pre-bicameral and pre-conscious way.

I’m tempted to answer “yes, obviously”, but for the counterargument, see this Reddit thread.



259 I never thought with language until now. This is my story. (self.self)
submitted 4 years ago by thegoldengiraffe

I think I'm very different from most people because of one main thing. I never thought with language. Ever. I moved to Canada when I was 2 from Asia, and have been basically been around English speakers my whole life. I'm in my twenties now and I can speak it relatively well, and can understand every single word. However, growing up, I never ever thought with language. Not once did I ever think something in my mind with words like "What are my friends doing right now?" to planning things like "I'm going to do my homework right after watching this show." I went through elementary school like this, I went through Highschool like this, I went through University like this...and I couldnt help but feel something was off about me that I couldnt put my hand on. Just last year, I had a straight up revalation, ephiphany....and this is hard to explain...but the best way that I can put it is that...I figured out that I SHOULD be thinking in language. So all of a sudden, I made a conscious effort to think things through with language. I spent a years time refining this new "skill" and it has COMPLETELY, and utterly changed my perception, my mental capabilities, and to be frank, my life. I can suddenly describe my emotions which was so insanely confusing to me before. I understand the concept that my friends are still "existing" even if they're not in site by thinking about their names. I now suddenly have opinions and feelings about things that I never had before. What the heck happened to me? I started thinking in language after not doing so my whole life. It's weird because I can now look back at my life before and see just how weird it was. Since I now have this new "skill" I can only describe my past life as ...."Mindless"..."empty"....."soul-less".... As weird as this sounds, I'm not even sure what I was, If i was even human, because I was barely even conscious. I felt like I was just reacting to the immediate environment and wasn't able to think anything outside of it. It's such a strange time in my life. It feels like I just found out the ultimate secret or something. .....Can anyone relate, or understand what Im saying? Can anyone explain what is happening to me? I have no idea where to even post this but this has been on my mind ever since I've been able to think about it.

80 comments source share save hide give award report crosspost hide all child comments

This guy thinks he “barely” had consciousness (in the Jaynesian sense), and it took him however many years to notice this *about himself*. It was just another universal human experience you can miss without realizing it! And notice how it was the *culturally learned knowledge* that other people worked differently which shifted him to the normal equilibrium. So maybe if there was some tribe like this somewhere, it would be easy to miss.

I’m also thinking of some cross-cultural psychiatry classes I had to take in residency. It’s well-known that some other cultures rarely get depression and anxiety in the classical Western sense. Instead, in the situations where we would become depressed and anxious, they get psychosomatic complaints, especially stomach pain. This happens to Westerners too sometimes, but in other cultures (eg China, Latin America) it’s by far the most common presentation. This seems similar to Jaynes’ argument that the ancient Greeks talked about feelings in their stomachs when we would talk about thoughts in our minds. I’m not saying these people aren’t conscious or have no theory of mind. But it seems like their theory of mind must be…arranged…differently than ours is, somehow. Or that cultural expectations about how these issues express themselves are shaping the way these issues express themselves, powerfully enough that you can just have whole cultures where depression the way we experience it *isn’t a thing*. See also this list of culture-bound syndromes. Make sure to read the discussion of Western culture-bound syndromes on the bottom – and make sure to spend a few moments considering what a politically-incorrect person might add to the list.

Even if I don’t accept all the stuff about hallucinatory Athena choreographing the Trojan War, the most important thing I’m going to take away from *Origin of Consciousness* is that theory of mind is an artifact, not a given, and it’s not necessarily the same everywhere. Much of the way we relate to our mind is culturally determined, and with a different enough cultural environment you can get some *weird* mind designs in ways that have real effect on behavior. Theory-of-mind-space is wider than we imagine, whether we’re thinking about ancient Sumerians or our ordinary-seeming neighbors.

JULIAN JAYNES
THE ORIGIN OF CONSCIOUSNESS IN THE BREAKDOWN OF THE BICAMERAL MIND

Problems With Paywalls

I.

I hate paywalls on articles. Absolutely hate them.

A standard pro-business argument: businesses can either make your life better (by providing deals you like) or keep your life the same (by providing deals you don't like, which you don't take). They can't really make your life worse. There are some exceptions, like if they outcompete and destroy another business you liked better, or if they have some kind of externalities, or if they lobby the government to do something bad. But in general, if you're angry at a business, you need to explain how one of these unusual conditions applies. Otherwise they're just "helping you less than you wish they did", not hurting you.

And so the standard justification for paywalls. Journalists are providing you a deal: you may read their articles in exchange for money. You are not entitled to their product without paying them money. They need to earn a living just like everyone else. So you can either accept their deal – pay money for the articles – or refuse their deal – and so be left no worse off than if they didn't exist.

But I notice feeling like this isn't true. I think I would be happier in a world where major newspapers ceased to exist, compared to the world where they exist but their articles are paywalled. Take a second and check if you feel the same way. If so, what could be going on?

First, paywalled newspapers sometimes use a clickbait model, where they start by making you curious what's in the article, then charge you to find out.

Here are some articles I’ve seen advertised recently (not all on paywalled sources): “Why Trump’s Fight With Obama Might Backfire”, “This Tech Guru Has Made A Shocking Prediction For 2020”, “Here’s Why Men are Pointing Loaded Guns At Their Dicks”.

I didn’t wake up this morning thinking “I wonder whether men are pointing loaded guns at their dicks, and, if so, why. I hope some enterprising journalist has investigated this question, and I will be happy to compensate her with money for satisfying this weird curiosity of mine.” No, instead, I was perfectly and innocently happy not knowing anything about this, *right up until I read the name of that article* at which point I became consumed with curiosity, ie a feeling that I will be unhappy until I know the answer. In this particular case it’s fine, because the offending website (VICE) is unpaywalled. I go there and after reading through nine paragraphs attacking “MAGA dolts”, in the tenth paragraph I get the one-sentence answer: there’s a meme in the gun community that any time someone posts a picture with their gun, amateurs will chime in with condescending advice about how they should be holding it more safely, so some people post pictures of them pointing loaded guns at their dicks in order to piss these people off. I feel completely unenlightened by knowing this. It has not brightened my day. It just removed the temporary itch of curiosity.

Some people critique capitalism by saying it creates new preferences that people have to spend money to satisfy. I haven’t noticed this being true in general – I only buy shoes when I need shoes, and I only buy Coke when I want Coke. But it seems absolutely on the mark regarding paywalled journalism. VICE created a new preference for me (the preference to know why some people point loaded guns at their dicks), then satisfied it. Overall I have neither gained nor lost utility. This seems different from providing me with a service.

They have an excuse, which is that this is how they make money. But what’s Marginal Revolution’s excuse? I saw this link in an MR links roundup. It was posted as “5. Why men are pointing loaded guns at their dicks.” So obviously I clicked on it, and here we are. But what is MR’s interest in making me click on a VICE article and read through nine paragraphs about “MAGA dolts”?

I can’t really blame them, because I did the same thing for years. I posted links posts, I framed the links in deliberately provocative ways, and then I felt good about myself when my stats page recorded that thousands of people had clicked on them. Sometimes I would write the whole thing out – “Here’s an article about men pointing loaded guns at their dicks – it’s because they want to criticize what they perceive as an excessive and condescending emphasis on trigger safety in gun culture” – and then nobody would click on it, and I would interpret that as a sign that I had failed in some way. I was an idiot, I apologize to all of you, and I have stopped doing that. I urge other bloggers to do the same – we gain no extra money, nor power, nor readership by being running-dogs for VICE’s weird ploy to trick people into reading its stupid articles. But as long as bloggers, Facebookers, tweeters, etc aren’t following good Internet hygiene, the very existence of paywalled sources will continue to be a net negative for the average Internet user.

This isn’t just about obvious clickbait like men pointing guns at their dicks. “Why Trump’s Fight With Obama Might Backfire” feels exactly the same to me. I don’t want to know more ephemeral garbage about Trump which may or may not affect his polls 0.5% for a week before they return to baseline. I don’t want to get more and more outraged until my ability to relate to my fellow human beings is shaped entirely by whether they're a "MAGA dolt" or not. And yet I find myself curious what's in the article!

(Trump's fight with Obama might backfire because independents like Obama more than Trump, and the tech guru's 2020 prediction was that Trump will lose. You're welcome.)

Second, paywalled articles become part of the discourse.

Last week's Wall Street Journal included an opinion column, Lockdowns Vs. The Vulnerable, arguing that statistics show the coronavirus lockdowns do not really prevent the coronavirus, but do disproportionately affect the most vulnerable people. It's already gotten retweeted a few dozen times, including by some bluechecks with tens of thousands of followers.

> The only conclusion available now is that the Swedes were right: The lockdowns were not sustainable, writes @holmanjenkins https://t.co/WFNEa5ScjV via @WSJ
>
> — Charles Gasparino (@CGasparino) May 30, 2020

Do you want to figure out exactly what statistics it uses and check whether they *really* show that lockdowns don't prevent coronavirus? Too bad – the article is paywalled and you cannot read it without paying $19.50/month to the Wall Street Journal. I personally suspect that this article is terribly wrong, possibly to the point of idiocy. But I can neither convince others of this, nor correct my own potentially-false first impression, without paying the Wall Street Journal $19.50 a month. Which I don't want to do. Partly because it is bad value, and partly because I don't want to reward them for publishing false things.

Newspapers publish articles – factual and opinionated – intending them to enter the public square as a topic of discussion. But if the discussions in the public square have an entry fee, the public square becomes smaller and less diverse.

It also becomes more of an echo chamber. Probably conservatives subscribe to the *Wall Street Journal* and liberals subscribe to the *New York Times*. So if conservatives post articles from the *Wall Street Journal*, liberals can neither benefit from the true ones and change their own opinions, nor correct the false ones and change conservatives' opinions. If you can't even *read* the other side's arguments, how can you be convinced by them?

Third, newspapers make it hard to guess whether you will encounter a paywall or not. Some of them raise a paywall on some kinds of articles but not others. Some of them raise a paywall if you're linked in from social media, but not if you're linked in from Google (or vice versa). Some of them raise a paywall if it's your Xth article per month on a certain computer, but not before.

The end result is you can't just learn to avoid the newspapers with paywalls. If you clearly knew which links were paywalled or not, you would just never click on those links, and not waste any time. Since any given newspaper has like a 25 – 50% chance of being paywalled whenever you read it, you get the variable reinforcement strategy that promotes frustrated addiction. And since at any given moment you are *desperate* to click on that link and find out Why Some Men Are Pointing Loaded Guns At Their Own Dicks, you will, like a chump, click it anyway, only to howl with rage when the paywall comes up.

This usually isn’t a deliberate misdeed; newspapers understandably want to give people limited access so they can decide whether or not they want to subscribe. But some forms of this *do* seem deliberate to me. Like when they let you read the first two paragraphs and get emotionally invested in the story, and then surprise you with a paywall in the third (I think this is why you need nine paragraphs of filler before getting to the one-sentence curiosity-satisfier). Or when they wait five seconds before a paywall message pops up, for the same reason.

Fourth, and most important, paywalled newspapers make it hard to search for information on Google. When I was trying to gather statistics on coronavirus to figure out how fast it was spreading, I noticed that the top ten or twenty relevant search results for a lot of coronavirus-related queries were paywalled articles. Because articles will make you wait several paragraphs/seconds before the paywall comes up, I couldn’t just quickly click on something, see if it had a paywall or not, and then move on to the next one. Instead, a search that would have taken me seconds if all paywalled sources ceased to exist ended up taking me several frustrating minutes.

II.

There are some simple steps we can take to fix this.

First, search engines should give users an option to hide paywalled articles from results. I realize how big a shitstorm this will cause, and I plan to enjoy every second of it. If they can’t make this happen for some reason, they should at least display a big red $$$ sign in front of paywalled articles, so users know which links will give them information before they waste a click on them. If Google refuses to do this, Bing should do it to get a leg up on

Google. If both of them refuse, DuckDuckGo. If all three of them refuse, sounds like they're providing an opening for some lucky entrepreneur.

Second, browser or browser-extension designers should figure out some way to automatically get links to display whether they're paywalled or not. Maybe something like this already exists, but I can't find it.

Third, bloggers (and social media users) should stop deliberately frustrating their readers. Stop posting tantalizing links like "Why Men Are Pointing Loaded Guns At Their Dicks" without further explanation! If you find the dick-gun phenomenon interesting, post the link plus a one-sentence summary. If someone wants more than the one-sentence summary, they can click the link, but I've done A/B testing on this and *it never happens*.

Fourth, bloggers (and social media users) should preferentially link non-paywalled sites. I realize this is not always possible, but most major stories are important enough that at least one non-paywalled outlet will be covering them.

Fifth, until the browser extension comes through, bloggers (and social media users) who do need to link a paywalled site should let readers know it's paywalled. For example, Lockdowns Vs. The Vulnerable [PAYWALLED] or [$$$] Lockdowns Vs. The Vulnerable. This will save readers a click *and* hopefully make bloggers think about what they're doing and whether it's really necessary.

I'm making a commitment to do 3, 4, and 5 from now on. If I ever change this commitment, I'll let you know. If you notice me slipping up, please point it out (nicely) and I'll try to correct myself.

Take The New Nootropics Survey

A few years ago I surveyed nootropics users about their experiences with different substances and posted the results here. Since then lots of new nootropics have come out, so I'm doing it again. If you have nootropics experience, please take **The 2020 SSC Nootropics Survey**. Expected completion time is ~15 minutes.

Thanks!

The Obligatory GPT-3 Post

I.

I would be failing my brand if I didn't write something about GPT-3, but I'm not an expert and discussion is still in its early stages. Consider this a summary of some of the interesting questions I've heard posed elsewhere, especially comments by gwern and nostalgebraist. Both of them are smart people who I broadly trust on AI issues, and both have done great work with GPT-2. Gwern has gotten it to write poetry, compose music, and even sort of play some chess; nostalgebraist has created nostalgebraist-autoresponder (a Tumblr written by GPT-2 trained on nostalgebraist's own Tumblr output). Both of them disagree pretty strongly on the implications of GPT-3. I don't know enough to resolve that disagreement, so this will be a kind of incoherent post, and hopefully stimulate some more productive comments. So:

OpenAI has released a new paper, Language Models Are Few-Shot Learners, introducing GPT-3, the successor to the wildly-successful language-processing AI GPT-2.

GPT-3 doesn't have any revolutionary new advances over its predecessor. It's just *much* bigger. GPT-2 had 1.5 billion parameters. GPT-3 has 175 billion. The researchers involved are very open about how it's the same thing but bigger. Their research goal was to test how GPT-like neural networks scale.

Before we get into the weeds, let's get a quick gestalt impression of how GPT-3 does compared to GPT-2.

Here's a sample of GPT-2 trying to write an article:

> **PROMPT: MILEY CYRUS WAS CAUGHT SHOPLIFTING FROM ABERCROMBIE AND FITCH ON HOLLYWOOD BOULEVARD TODAY**
>
> The 19-year-old singer was caught on camera being escorted out of the store by security guards.
>
> The singer was wearing a black hoodie with the label ‘Blurred Lines’ on the front and ‘Fashion Police’ on the back
>
> The singer was also wearing a pair of black-rimmed glasses, a black jacket, black jeans and black sandals.
>
> She was carrying a pair of black and white striped gloves and a small black bag.

And here’s GPT-3 at the same task:

> **PROMPT: UNITED METHODISTS AGREE TO HISTORIC SPLIT – THOSE WHO OPPOSE GAY MARRIAGE WILL FORM THEIR OWN DENOMINATION**
>
> After two days of intense debate, the United Methodist Church has agreed to a historic split – one that is expected to end in the creation of a new denomination, one that will be “theologically and socially conservative,” according to the Washington Post. The majority of delegates attending the church’s annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will “discipline” clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.

> The Post notes that the denomination, which claims 12.5 million members, was in the early 20th century the "largest Protestant denomination in the US", but that it has been shrinking in recent decades. The new split will be the second in the church's history. The first occurred in 1968, when roughly 10% of the denomination left to form the Evangelical United Brethren Church. The Post notes that the proposed split "comes at a critical time for the church, which has been losing members for years," which has been "pushed toward the brink of a schism over the role of LGBTQ people in the church." Gay marriage is not the only issue that has divided the church. In 2016, the denomination was split over ordination of transgender clergy, with the North Pacific regional conference voting to ban them from serving as clergy, and the South Pacific regional conference voting to allow them.

In case you're wondering, no, the Methodist Church did not really split from the Evangelical United Brethren Church in 1968, it *merged with it* in that year. GPT-3 is doing what I used to do on essay questions – throw out a bunch of half-remembered names and dates and hope nobody looks too closely at the exact relations.

These kinds of comparisons are inherently fuzzy, since both GPTs have produced a lot of material and it's easy to cheat by taking better or worse samples of one or the other. You can see more GPT-2 samples here and more GPT-3 samples here

Or compare some of GPT-2's poetry:

> The gods are they who came to earth
> And set the seas ablaze with gold.
> There is a breeze upon the sea,
> A sea of summer in its folds,

A salt, enchanted breeze that mocks
The scents of life, from far away
Comes slumbrous, sad, and quaint, and quaint.
The mother of the gods, that day,
With mortal feet and sweet voice speaks,
And smiles, and speaks to men: “My Sweet,
I shall not weary of thy pain.”

…Let me drink of the wine of pain
And think upon the agonies of hope,
And of the blessed Giver of all good things;
For, man for man, mine is the deepest love
That sorrow takes upon the humblest soul;
But who hath learned how sorrow turns to gall
The places where my feet have trod before.

…to some of GPT-3’s:

I must have shadows on the way
If I am to walk I must have
Each step taken slowly and alone
To have it ready made

And I must think in lines of grey
Must have dim thoughts to be my guide
Must look on blue and green
And never let my eye forget
That color is my friend
And purple must surround me too

The yellow of the sun is more more
Intusive than the bluish snow
That falls on all of us. I must have
Grey thoughts and blue thoughts walk with me
If I am to go away at all.

Ignore the style differences – the GPT-2 instance was trained on 18th-century verse, GPT-3 on Wallace Stevens. I notice that the first poem, while flowery, is kind of incoherent. The second has more of a structure and a theme that it maintains throughout. I'm not sure it really symbolizes anything interesting, but I've never been sure of that with real Wallace Stevens either, so good work.

(you can read more GPT-2 poetry here. There's not much more public GPT-3 poetry – you can read one or two more samples on the paper linked above).

Enough writing. What about math? Last year, I posted the following excerpt from GPT-2:

> **PROMPT: MY 10 FAVORITE ANIMALS ARE:**
>
> My ten favorite animals are:
> 1. Zebras with a white scar on the back
> 2. Insiduous spiders and octopus
> 3. Frog with large leaves, hopefully black
> 4. Cockatiel with scales
> 5. Razorbill with wings hanging about 4 inches from one's face and a heart tattoo on a frog
> 3. Cockatric interlocking tetrabods that can be blind, cut, and eaten raw:
> 4. Black and white desert crocodiles living in sunlight
> 5. Zebra and many other pea bugs

Despite the obvious flaws in this piece, I was impressed. GPT-2 was clearly trying to make a numbered list, and almost kind of getting it right! It counted to 4 successfully! Remember, this is a text prediction engine that didn't necessarily need to have any concept of numbers. But it still kind of counted to 4! I wrote:

> Imagine you prompted the model with “What is one plus one?” I actually don’t know how it would do on this problem. I’m guessing it would answer “two”, just because the question probably appeared a bunch of times in its training data.
>
> Now imagine you prompted it with “What is four thousand and eight plus two thousand and six?” or some other long problem that probably didn’t occur exactly in its training data. I predict it would fail, because this model can’t count past five without making mistakes. But I imagine a very similar program, given a thousand times more training data and computational resources, would succeed. It would notice a pattern in sentences including the word “plus” or otherwise describing sums of numbers, it would figure out that pattern, and it would end up able to do simple math. I don’t think this is too much of a stretch given that GPT-2 learned to count to five and acronymize words and so on.

I said “a very similar program, given a thousand times more training data and computational resources, would succeed [at adding four digit numbers]”. Well, GPT-3 is a very similar program with a hundred times more computational resources, and…it can add four-digit numbers! At least sometimes, which is better than GPT-2’s “none of the time”.

II.

In fact, let’s take a closer look at GPT-3’s math performance.

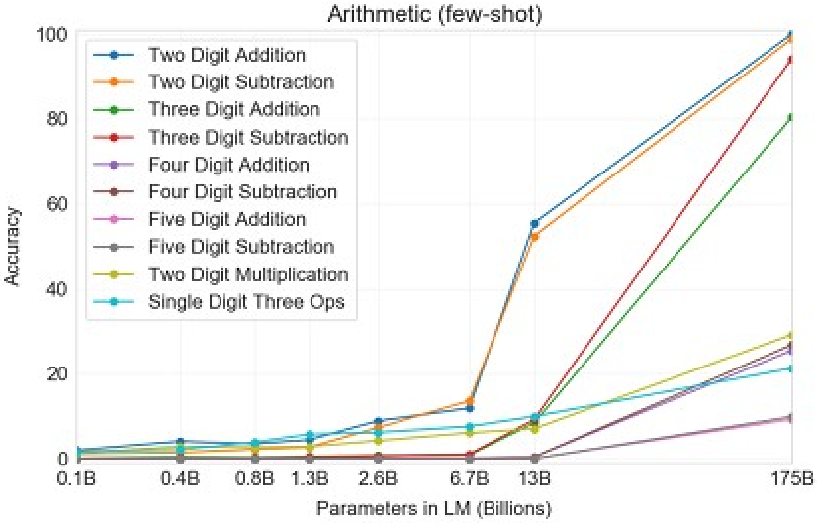


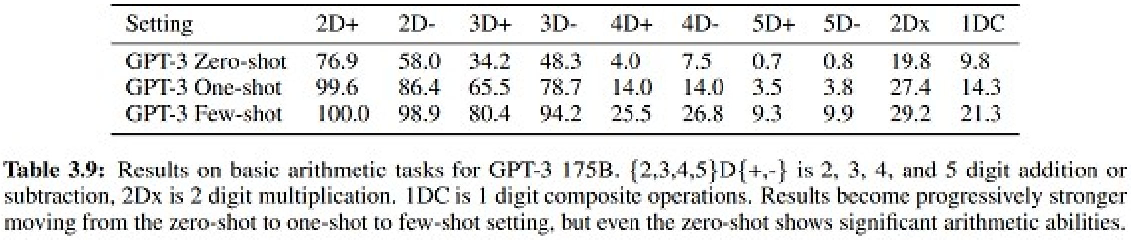
Figure 3.10: Results on all 10 arithmetic tasks in the few-shot settings for models of different sizes. There is a significant jump from the second largest model (GPT-3 13B) to the largest model (GPT-3 175), with the latter being able to reliably accurate 2 digit arithmetic, usually accurate 3 digit arithmetic, and correct answers a significant fraction of the time on 4-5 digit arithmetic, 2 digit multiplication, and compound operations. Results for one-shot and zero-shot are shown in the appendix.

The 1.3 billion parameter model, equivalent to GPT-2, could get two-digit addition problems right less than 5% of the time – little better than chance. But for whatever reason, once the model hit 13 billion parameters, its addition abilities improved to 60% – the equivalent of a D student. At 175 billion parameters, it gets an A+.

What does it mean for an AI to be able to do addition, but only inconsistently? For four digit numbers, but not five digit numbers? Doesn’t it either understand addition, or not?

Maybe it’s cheating? Maybe there were so many addition problems in its dataset that it just memorized all of them? I don’t think this is the answer. There are 100 million possible 4-digit addition problems; seems unlikely that GPT-3 saw that many of them. Also, if it was memorizing its training data, it should have gotten all 100 possible two-digit multiplication problems, but it only has about a 25% success rate on those. So it can’t be using a lookup table.

Maybe it’s having trouble *locating* addition rather than *doing* addition? (thanks to nostalgebraist for this framing). This sort of seems like the lesson of Table 3.9:



Setting	2D+	2D-	3D+	3D-	4D+	4D-	5D+	5D-	2Dx	1DC
GPT-3 Zero-shot	76.9	58.0	34.2	48.3	4.0	7.5	0.7	0.8	19.8	9.8
GPT-3 One-shot	99.6	86.4	65.5	78.7	14.0	14.0	3.5	3.8	27.4	14.3
GPT-3 Few-shot	100.0	98.9	80.4	94.2	25.5	26.8	9.3	9.9	29.2	21.3

Table 3.9: Results on basic arithmetic tasks for GPT-3 175B. {2,3,4,5}D{+,-} is 2, 3, 4, and 5 digit addition or subtraction, 2Dx is 2 digit multiplication. 1DC is 1 digit composite operations. Results become progressively stronger moving from the zero-shot to one-shot to few-shot setting, but even the zero-shot shows significant arithmetic abilities.

“Zero-shot” means you just type in “20 + 20 = ?”. “One-shot” means you give it an example first: “10 + 10 = 20. 20 + 20 = ?” “Few-shot” means you give it as many examples as it can take. Even the largest and best model only does mediocre on the zero-shot task, but it does better on the one-shot and best on the few-shot. So it seems like if you remind it what addition is a couple of times before solving an addition problem, it does better. This suggests that there is a working model of addition somewhere within the bowels of this 175 billion parameter monster, but it has a hard time drawing it out for any particular task. You need to tell it “addition” “we’re doing addition” “come on now, do some addition!” up to fifty times before it will actually deploy its addition model for these problems, instead of some other model. Maybe if you did this five hundred or five thousand times, it would excel at the problems it can’t do now, like adding five digit numbers. But why should this be so hard? The plus sign almost always means addition. “20 + 20 = ?” is not some inscrutable hieroglyphic text. It basically always means the same thing. Shouldn’t this be easy?

When I prompt GPT-2 with addition problems, the most common failure mode is getting an answer that isn’t a number. Often it’s a few paragraphs of text that look like they came from a math textbook. It feels like it’s been able to locate the problem as far as “you want the kind of thing in math textbooks”, but not as far as “you want the answer to the exact math problem you are giving me”. This is a surprising issue to have, but so far AIs have been nothing if not surprising. Imagine telling Marvin Minsky or someone that an AI smart enough to write decent poetry would not necessarily be smart enough to know that, when asked “325 + 504”, we wanted a numerical response!

Or maybe that’s not it. Maybe it has trouble getting math problems right consistently for the same reason *I* have trouble with this. In fact, GPT-3’s performance is very similar to mine. I can also add two digit numbers in my head with near-100% accuracy, get worse as we go to three digit numbers, and make no guarantees at all about four-digit. I also find multiplying two-digit numbers in my head much harder than adding those same numbers. What’s my excuse? Do I understand addition, or not? I used to assume my problems came from limited short-term memory, or from neural noise. But GPT-3 shouldn’t have either of those issues. Should I feel a deep kinship with GPT-3? Are we both minds heavily optimized for writing, forced by a cruel world to sometimes do math problems? I don’t know.

[EDIT: an alert reader points out that when GPT-3 fails at addition problems, it fails in human-like ways – for example, forgetting to carry a 1.]

III.

GPT-3 is, fundamentally, an attempt to investigate scaling laws in neural networks. That is, if you start with a good neural network, and make it ten times bigger, does it get smarter? How much smarter? Ten times smarter? Can you keep doing this forever until it’s infinitely smart or you run out of computers, whichever comes first?

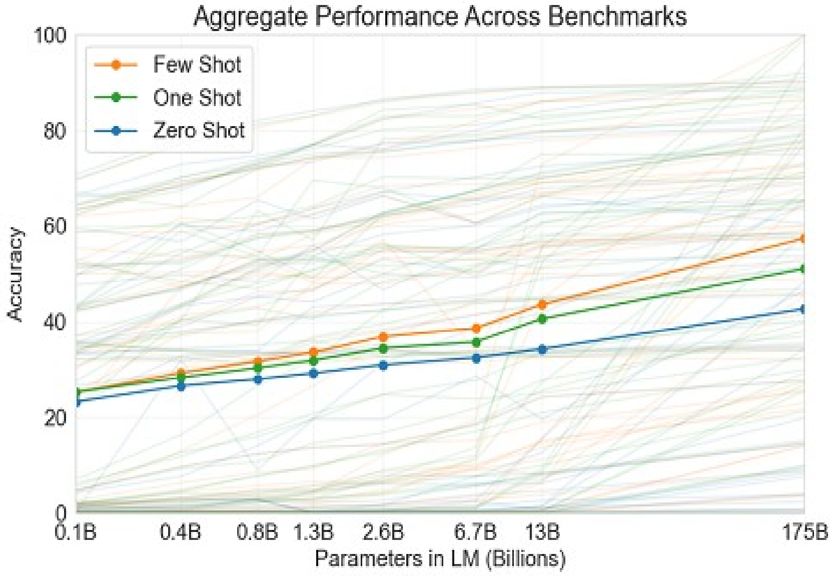


Figure 1.3: Aggregate performance for all 42 accuracy-denominated benchmarks While zero-shot performance improves steadily with model size, few-shot performance increases more rapidly, demonstrating that larger models are more proficient at in-context learning. See Figure 3.8 for a more detailed analysis on SuperGLUE, a standard NLP benchmark suite.

So far the scaling looks logarithmic – a consistent multiplication of parameter number produces a consistent gain on the benchmarks.

Does that mean it really is all about model size? Should something even bigger than GPT-3 be better still, until eventually we have things that can do all of this stuff arbitrarily well without any new advances?

This is where my sources diverge. Gwern says yes, probably, and points to years of falsified predictions where people said that scaling might have worked so far, but definitely wouldn't work past *this* point. Nostalgebraist says maybe not, and points to decreasing returns of GPT-3's extra power on certain benchmarks (see Appendix H) and to this OpenAI paper, which he interprets as showing that scaling should break down somewhere around or just slightly past where GPT-3 is. If he's right, GPT-3 might be around the best that you can do just by making GPT-like things bigger and bigger.

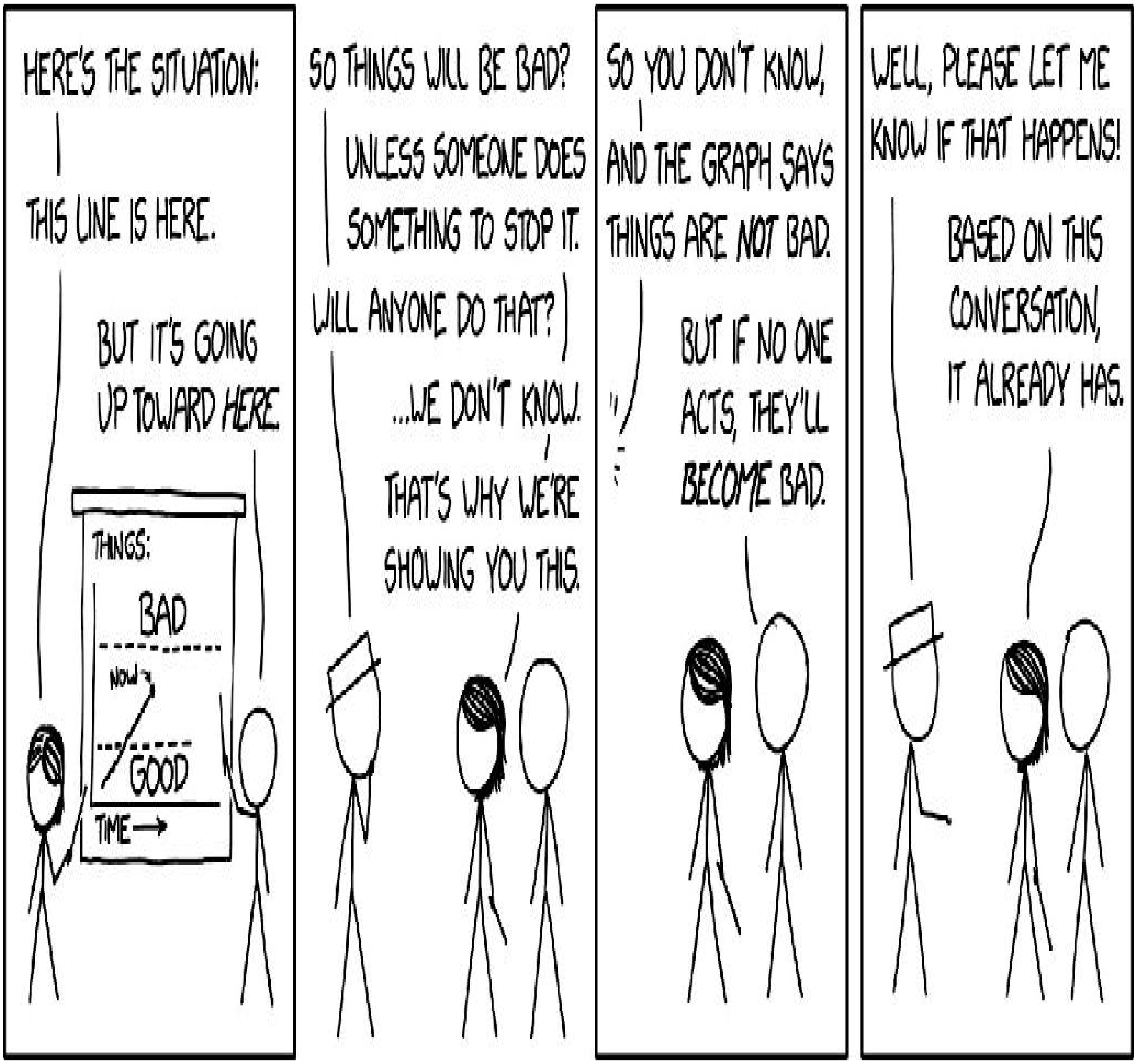
He also points out that although GPT-3 is impressive as a general-purpose reasoner that has taught itself things without being specifically optimized to learn them, it’s often worse than task-specifically-trained AIs at various specific language tasks, so we shouldn’t get too excited about it being close to superintelligence or anything. I guess in retrospect this is obvious – it’s cool that it learned how to add four-digit numbers, but calculators have been around a long time and can add much longer numbers than that.

If the scaling laws don’t break down, what then?

GPT-3 is very big, but it’s not pushing the limits of how big an AI it’s possible to make. If someone rich and important like Google wanted to make a much bigger GPT, they could do it.

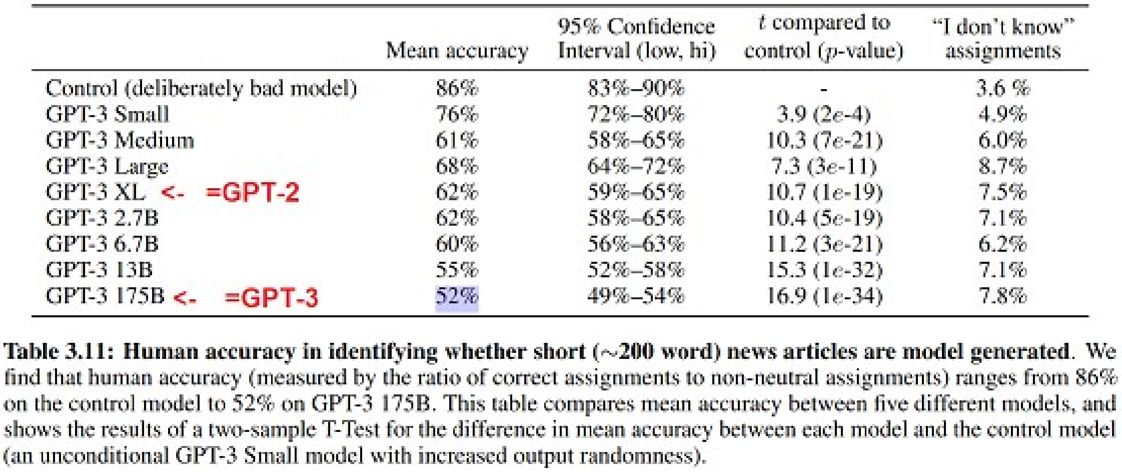
> GPT-3 is terrifying because it's a tiny model compared to what's possible, trained in the dumbest way possible on a single impoverished modality on tiny data, yet the first version already manifests crazy runtime meta-learning—and the scaling curves *still* are not bending! 😮 https://t.co/hQbW9znm3x
>
> — 𝔊𝔴𝔢𝔯𝔫 (@gwern) May 31, 2020

Does “terrifying” sound weirdly alarmist here? I think the argument is something like this. In February, we watched as the number of US coronavirus cases went from 10ish to 50ish to 100ish over the space of a few weeks. We didn’t panic, because 100ish was still a very low number of coronavirus cases. In retrospect, we should have panicked, because the number was constantly increasing, showed no signs of stopping, and simple linear extrapolation suggested it would be somewhere scary very soon. After the number of coronavirus cases crossed 100,000 and 1,000,000 at exactly the time we could have predicted from the original curves, we all told ourselves we *definitely* wouldn’t be making that exact same mistake again.


HERE'S THE SITUATION:
THIS LINE IS HERE.
BUT IT'S GOING UP TOWARD *HERE*.
THINGS:
BAD
NOW
GOOD
TIME →
SO THINGS WILL BE BAD?
UNLESS SOMEONE DOES SOMETHING TO STOP IT.
WILL ANYONE DO THAT?
...WE DON'T KNOW.
THAT'S WHY WE'RE SHOWING YOU THIS.
SO YOU DON'T KNOW,
AND THE GRAPH SAYS THINGS ARE *NOT* BAD.
BUT IF NO ONE ACTS, THEY'LL *BECOME* BAD.
WELL, PLEASE LET ME KNOW IF THAT HAPPENS!
BASED ON THIS CONVERSATION, IT ALREADY HAS.

It’s always possible that the next AI will be the one where the scaling curves break and it stops being easy to make AIs smarter just by giving them more computers. But unless something surprising like that saves us, we should assume GPT-like things will become much more powerful very quickly.

What would much more powerful GPT-like things look like? They can already write some forms of text at near-human level (in the paper above, the researchers asked humans to identify whether a given news article had been written by a human reporter or GPT-3; the humans got it right 52% of the time)



	Mean accuracy	95% Confidence Interval (low, hi)	t compared to control (p-value)	"I don't know" assignments
Control (deliberately bad model)	86%	83%–90%	-	3.6 %
GPT-3 Small	76%	72%–80%	3.9 (2*e*-4)	4.9%
GPT-3 Medium	61%	58%–65%	10.3 (7*e*-21)	6.0%
GPT-3 Large	68%	64%–72%	7.3 (3*e*-11)	8.7%
GPT-3 XL <- =GPT-2	62%	59%–65%	10.7 (1*e*-19)	7.5%
GPT-3 2.7B	62%	58%–65%	10.4 (5*e*-19)	7.1%
GPT-3 6.7B	60%	56%–63%	11.2 (3*e*-21)	6.2%
GPT-3 13B	55%	52%–58%	15.3 (1*e*-32)	7.1%
GPT-3 175B <- =GPT-3	52%	49%–54%	16.9 (1*e*-34)	7.8%

Table 3.11: Human accuracy in identifying whether short (~200 word) news articles are model generated. We find that human accuracy (measured by the ratio of correct assignments to non-neutral assignments) ranges from 86% on the control model to 52% on GPT-3 175B. This table compares mean accuracy between five different models, and shows the results of a two-sample T-Test for the difference in mean accuracy between each model and the control model (an unconditional GPT-3 Small model with increased output randomness).

So one very conservative assumption would be that a smarter GPT would do better at various arcane language benchmarks, but otherwise not be much more interesting – once it can write text at a human level, that's it.

Could it do more radical things like write proofs or generate scientific advances? After all, if you feed it thousands of proofs, and then prompt it with a theorem to be proven, that's a text prediction task. If you feed it physics textbooks, and prompt it with "and the Theory of Everything is…", that's also a text prediction task. I realize these are wild conjectures, but the last time I made a wild conjecture, it was "maybe you can learn addition, because that's a text prediction task" and that one came true within two years. But my guess is still that this won't happen in a meaningful way anytime soon. GPT-3 is much better at writing coherent-sounding text than it is at any kind of logical reasoning; remember it still can't add 5-digit numbers very well, get its

Methodist history right, or consistently figure out that a plus sign means “add things”. Yes, it can do simple addition, but it has to use supercomputer-level resources to do so – it’s so inefficient that it’s hard to imagine even very large scaling getting it anywhere useful. At most, maybe a high-level GPT could write a plausible-sounding Theory Of Everything that uses physics terms in a vaguely coherent way, but that falls apart when a real physicist examines it.

Probably we can be pretty sure it won’t take over the world? I have a hard time figuring out how to turn world conquest into a text prediction task. It could probably imitate a human writing a plausible-sounding *plan* to take over the world, but it couldn’t implement such a plan (and would have no desire to do so).

For me the scary part isn’t the much larger GPT we’ll probably have in a few years. It’s the discovery that even very complicated AIs get smarter as they get bigger. If someone ever invented an AI that did do more than text prediction, it would have a pretty fast takeoff, going from toy to superintelligence in just a few years.

Speaking of which – can anything based on GPT-like principles ever produce superintelligent output? How would this happen? If it’s trying to mimic what a human can write, then no matter how intelligent it is “under the hood”, all that intelligence will only get applied to becoming better and better at predicting what kind of dumb stuff a normal-intelligence human would say. In a sense, solving the Theory of Everything would be a failure at its primary task. No human writer would end the sentence “the Theory of Everything is…” with anything other than “currently unknown and very hard to figure out”.

But if our own brains are also prediction engines, how do *we* ever create things smarter and better than the ones we grew up with? I can imagine scientific theories being part of our predictive model rather than an output of it – we use the theory of gravity to predict how things will fall. But what about new forms of art? What about thoughts that have never been thought before?

And how many parameters does the adult human brain have? The responsible answer is that brain function doesn't map perfectly to neural net function, and even if it did we would have no idea how to even begin to make this calculation. The irresponsible answer is a hundred trillion. That's a big number. But at the current rate of GPT progress, a GPT will have that same number of parameters somewhere between GPT-4 and GPT-5. Given the speed at which OpenAI works, that should happen about two years from now.

I am definitely not predicting that a GPT with enough parameters will be able to do everything a human does. But I'm really interested to see what it *can* do. And we'll find out soon.

Wordy Wernicke’s

There are two major brain areas involved in language. To oversimplify, Wernicke’s area in the superior temporal gyrus handles meaning; Broca’s area in the inferior frontal gyrus handles structure and flow.

If a stroke or other brain injury damages Broca’s area but leaves Wernicke’s area intact, you get language which is meaningful, but not very structured or fluid. You sound like a caveman: “Want food!”

If it damages Wernicke’s area but leaves Broca’s area intact, you get speech which has normal structure and flow, but is meaningless. I’d read about this pattern in books, but I still wasn’t prepared the first time I saw a video of a Wernicke’s aphasia patient (source):

During yesterday’s discussion of GPT-3, a commenter mentioned how *alien* it felt to watch something use language perfectly without quite making sense. I agree it’s eerie, but it isn’t some kind of inhuman robot weirdness. Any one of us is a railroad-spike-through-the-head away from doing the same.

Does this teach us anything useful about GPT-3 or neural networks? I lean towards no. GPT-3 already makes more sense than a Wernicke’s aphasiac. Whatever it’s doing is on a higher level than the Broca’s/Wernicke’s dichotomy. Still, it would be interesting to learn what kind of computational considerations caused the split, and whether there’s any microstructural difference in the areas that reflects it. I don’t know enough neuroscience to have an educated opinion on this.

The Vision Of Vilazodone And Vortioxetine

I.

One of psychiatry's many embarrassments is how many of our drugs get discovered by accident. They come from random plants or shiny rocks or stuff Alexander Shulgin invented to get high.

But every so often, somebody tries to do things the proper way. Go over decades of research into what makes psychiatric drugs work and how they could work better. Figure out the hypothetical properties of the ideal psych drug. Figure out a molecule that matches those properties. Synthesize it and see what happens. This was the vision of vortioxetine and vilazodone, two antidepressants from the early 2010s. They were approved by the FDA, sent to market, and prescribed to millions of people. Now it's been enough time to look back and give them a fair evaluation. And…

…and it's been a good reminder of why we don't usually do this.

Enough data has come in to be pretty sure that vortioxetine and vilazodone, while effective antidepressants, are no better than the earlier medications they sought to replace. I want to try going over the science that led pharmaceutical companies to think these two drugs might be revolutionary, and then speculate on why they weren't. I'm limited in this by my total failure to understand several important pieces of the pathways involved, so I'll explain the parts I get, and list the parts I don't in the hopes that someone clears them up in the comments.

II.

SSRIs take about a month to work. This is surprising, because they start increasing serotonin immediately. If higher serotonin is associated with improved mood, why the delay?

Most research focuses on presynaptic 5-HT1A autoreceptors, which detect and adjust the amount of serotonin being released in order to maintain homeostasis. If that sounds clear as mud to you, well, it did to me too the first twenty times I heard it. Here’s the analogy which eventually worked for me:

Imagine you’re a salesperson who teleconferences with customers all day. You’re bad at projecting your voice – sometimes it’s too soft, sometimes too loud. Your boss tells you the right voice level for sales is 60 decibels. So you put an audiometer on your desk that measures how many decibels your voice is. It displays an up arrow, down arrow, or smiley face, telling you that you’re being too quiet, loud, or just right.

The audiometer is presynaptic (it’s on your side of the teleconference, not your customer’s side). It’s an autoreceptor, not a heteroreceptor (you’re using it to measure yourself, not to measure anything else). And you’re using it to maintain homeostasis (to keep your voice at 60 dB).

Suppose you take a medication that stimulates your larynx and makes your voice naturally louder. As long as your audiometer’s working, that medication will have no effect. Your voice will naturally be louder, but that means you’ll see the down arrow on your audiometer more often, you’ll speak with less force, and the less force and louder voice will cancel out and keep you at 60 dB. No change.

This is what happens with antidepressants. The antidepressant increases serotonin levels (ie sends a louder signal). But if the presynaptic 5-HT1A autoreceptors are intact, they tell the cells that the signal is too loud and they should release less serotonin. So they do, and now we’re back where we started – ie depressed.

(Why don’t the autoreceptors notice the original problem – that you have too little serotonin and are depressed – and work their magic there? Not sure. Maybe depression affects whatever sets the autoreceptors, and causes them to be set too low? Maybe the problem is with reception, not transmission? Maybe you have the right amount of serotonin, but you want excessive amounts of serotonin because that would fix a problem somewhere else? Maybe you should have paid extra for the premium model of presynaptic autoreceptor?)

So how come antidepressants work after a month? The only explanation I’ve heard is that the autoreceptors get “saturated”, which is a pretty nonspecific term. I think it means that there’s a *second* negative feedback loop controlling the first negative feedback loop – the cell notices there is way more autoreceptor activity than expected and assumes it is producing too many autoreceptors. Over the course of a month, it stops producing more autoreceptors, the existing autoreceptors gradually degrade, and presynaptic autoreception stops being a problem. In our salesperson analogy, after a while you notice that your audiometer doesn’t match your own perception of how much effort you’re putting into speaking – no matter how quietly you feel like you’re whispering, the audiometer just keeps saying you’re yelling very loud. Eventually you declare it defective, throw it out, and just speak naturally. Now the larynx-stimulating medication *can* make your voice louder.

In the late ’90s, some scientists wondered – what happens if you just block the 5-HT1A autoreceptors directly? At the very least, seems like you could make antidepressants work a month faster. Best case scenario, you can make antidepressants work *better*. Maybe after a month, the cells have lost *some* confidence in the autoreceptors, but they’re still keeping serotonin somewhere below the natural amount based on the anomalous autoreceptor reading. So maybe blocking the autoreceptors would mean a faster, better, antidepressant.

Luckily we already knew a chemical that could block these – pindolol. Pindolol is a blood-pressure-lowering medication, but by coincidence it also makes it into the brain and blocks this particular autoreceptor involved in serotonin homeostasis. So a few people started giving pindolol along with antidepressants. What happened? According to some small unconvincing systematic reviews, it did seem to kind of help make antidepressants work faster, but according to some other small unconvincing meta-analyses it probably didn’t make them work any better. It just made antidepressants go from taking about four weeks to work, to taking one or two weeks to work. It also gave patients dizziness, drowsiness, weakness, and all the other things you would expect from giving a blood-pressure-lowering medication to people with normal blood pressure. So people decided it probably wasn’t worth it.

III.

And then there’s buspirone.

Buspirone is generally considered a 5-HT1A agonist. The reality is a little more complicated; it’s a full agonist of presynaptic autoreceptors, and a partial agonist of postsynaptic autoreceptors(uh, imagine the customer in the other side of the teleconference also has an audiometer).

Buspirone has a weak anti-anxiety effect. It may also weakly increase sex drive, which is a nice contrast to SSRIs, which decrease (some would say “destroy”) sex drive. Flibanserin, a similar drug, is FDA-approved as a sex drive enhancer.

Buspirone stimulates presynaptic autoreceptors, which should cause cells to release less serotonin. Since high serotonin levels (eg with SSRIs) decreases sex drive, it makes sense that buspirone should increase sex drive. So far, so good.

But why is buspirone anxiolytic? I have just read a dozen papers purporting to address this question, and they might as well have been written in Chinese for all I was able to get from them. Some of them just say that decreasing serotonin levels decreases anxiety, which would probably come as a surprise to anyone on SSRIs, anyone who does tryptophan depletion studies, anyone who measures serotonin metabolites in the spinal fluid, etc. Obviously it must be more complicated than this. But how? I can’t find any explanation.

A few books and papers take a completely different tack and argue that buspirone just does the same thing SSRIs do – desensitize the presynaptic autoreceptors until the cells ignore them – and then has its antianxiety action through stimulating postsynaptic receptors. But if it’s doing the same thing as SSRIs, how come it has the opposite effect on sex drive? And how come there are some studies suggesting that it’s helpful to add buspirone onto SSRIs? Why isn’t that just doubling up on the same thing?

I am deeply grateful to SSC commenter Scchm for presenting an argument that this is all wrong, and buspirone acts on D4 receptors, both in its anxiolytic and pro-sexual effects. This would neatly resolve the issues above. But then how come nobody else mentions this? How come everyone else seems to think buspirone makes sense, and writes whole papers about it without using the sentence “what the hell all of this is crazy”?

And here’s one more mystery: after the pindolol studies, everyone just sort of started assuming buspirone would work the same way pindolol did. This doesn’t really make sense pharmacologically – pindolol is an *antagonist* at presynaptic 5-HT1A receptors; buspirone is an *agonist* of same. And it doesn’t actually work in real life – someone did a study, and the study found it didn’t work. Still, and I have no explanation for this, people got excited about this possibility. If buspirone could work like pindolol, then we would have a chemical that made antidepressants work faster, *and* treated anxiety, *and* reduced sexual dysfunction.

And here’s one *more* mystery – okay, you have unrealistically high expectations for buspirone, fine, give people buspirone along with their SSRI. Lots of psychiatrists do this, it’s not really my thing, but it’s not a *bad* idea. But instead, the people making this argument became obsessed with the idea of finding a single chemical that combined SSRI-like activity with buspirone-like activity. A dual serotonin-transporter-inhibitor and 5-HT1A partial agonist became a pharmacological holy grail.

IV.

After a lot of people in lab coats poured things from one test tube to another, Merck announced they had found such a chemical, which they called vilazodone (Viibryd®).

Vilazodone is an SSRI and 5-HT1A partial agonist. I can't find how its exact partial agonist profile differs from buspirone, except that it's more of a postsynaptic agonist, whereas buspirone is more of a postsynaptic antagonist. I don't know if this makes a difference.

The FDA approved vilazodone to treat depression, so it "works" in that sense. But does its high-tech promise pan out? Is it really faster-acting, anxiety-busting, and less likely to cause sexual side effects?

On a chemical level, things look promising. This study finds that vilazodone elevates serotonin faster and higher than Prozac does in mice. And this is kind of grim, but toxicologists have noticed that vilazodone overdoses are much more likely to produce serotonin toxicity than Prozac overdoses, which fits what you would expect if vilazodone successfully breaks the negative feedback system that keeps serotonin in a normal range.

On a clinical level, maybe not. Proponents of vilazodone got excited about a study where vilazodone showed effects as early as week two. But "SSRIs take four weeks to work" is a rule of thumb, not a natural law. You always get a couple of people who get some effect early on, and if your study population is big enough, that'll show up as a positive result. So you need to compare vilazodone to an SSRI directly. The only group I know who tried this, Matthews et al, found no difference – in fact, vilazodone was nonsignificantly slower (relevant figure). There's no sign of vilazodone working any better either.

What about sexual side effects? Vilazodone does better than SSRIs in rats, but whatever. There's supposedly a human study – from the same Matthews et al team as above – but sexual side effects were so rare in all groups that it's hard to draw any conclusions. This is bizarre – they had a thousand patients, and only 15 reported decreased libido (and no more in the treatment groups than the placebo group). Maybe God just hates antidepressant studies and makes sure they never find anything, and this is just as true when you're studying side effects as it is when you're studying efficacy.

Clayton et al do their own study of vilazodone's sexual side effects. They find that overall vilazodone improves sexual function over placebo, probably because they used a scale that was very sensitive to the kind of bad sexual function you get when you're depressed, and not as sensitive to the kind you get on antidepressants. But they did measure how many people who didn't start out with sexual dysfunction got it during the trial, and this number was 1% of the placebo group and 8% of the vilazodone group. How many people would have gotten dysfunction on an SSRI? We don't know because they didn't include an active comparator. Usually I expect about 30 – 50% of people to get sexual side effects on SSRIs, but that's based on me asking them and not on whatever strict criteria they use for studies. Remember, the Matthews study was able to find only 1.5% of people getting sexual side effects! So we shouldn't even try to estimate how this compares. All we can say is that vilazodone definitely doesn't have *no* sexual side effects.

I can't find any studies evaluating vilazodone vs. anything else for anxiety, but I also can't find any patients saying vilazodone treated their anxiety especially well.

There’s really only one clear and undeniable difference between vilazodone and ordinary SSRIs, which is that vilazodone costs $290 a month, whereas other SSRIs cost somewhere in the single digits (Lexapro costs $7.31). If you’re paying for vilazodone, you can take comfort in knowing your money helped fund a pretty cool research program that had some interesting science behind it. But I’m not sure it actually panned out.

V.

Encouraged by Merck’s success…

(not necessarily clinical success, success at getting people to pay $290 a month for an antidepressant)

…Takeda and Lundbeck announced their own antidepressant with 5-HT1A partial agonist action, vortioxetine. They originally gave it the trade name Brintellix®, but upon its US release people kept confusing it with the unrelated medication Brilinta®, so Takeda/Lundbeck agreed to change the name to Trintellix® for the American market.

Vortioxetine claimed to have an advantage over its competitor vilazodone, in that it also antagonized 5-HT3 receptors. 5-HT3 receptors are weird. They’re the only ion channel based serotonin receptors, and they’re not especially involved in mood or anxiety. They do only one thing, and they do it well: they make you really nauseous. If you’ve ever felt nauseous on an SSRI, 5-HT3 agonism is why. And if you’ve ever taken Zofran (ondansetron) for nausea, you’ve benefitted from its 5-HT3 antagonism. Most antidepressants potentially cause nausea; since vortioxetine also treats nausea, presumably you break even and are no more nauseous than you were before taking it. Also, there are complicated theoretical reasons to believe maybe 5-HT3 antagonism is kind of like 5-HT1A antagonism in that it speeds the antidepressant response.

After this Takeda and Lundbeck kind of just went crazy, claiming effects on more and more serotonin receptors. It's a 5-HT7 antagonist! (what is 5-HT7? No psychiatrist had ever given a second's thought to this receptor before vortioxetine came out, but apparently it…exists to make your cognition worse, so that blocking it makes your cognition better again?) It's a 5-HT1B partial agonist! (what is 5-HT1B? Apparently a useful potential depression target, according to a half-Japanese, half-Scandinavian team, who report no conflict of interest even though vortioxetine is being sold by a consortium of a Japanese pharma company and a Scandinavian pharma company). It's a 5-HT1D antagonist! (really? There are four different kinds of 5-HT1 receptor? Are you sure you're not just making things up now?)

If we take all of this seriously, vortioxetine is an SSRI with faster mechanism of action, fewer sexual side effects, additional anti-anxiety effect, additional anti-nausea effect, plus it gives you better cognition (technically "relieves the cognitive symptoms of depression"). Is any of this at all true?

A meta-analysis of 12 studies finds vortioxetine has a statistically significant but pathetic effect size of 0.2 against depression, which is about average for antidepressants. In a few head-to-head comparisons with SNRIs (similar to SSRIs), vortioxetine treats depression about equally well. Patients are more likely to stop the SNRIs because of side effects than to stop the vortioxetine, but SNRIs probably have more side effects than SSRIs, so unclear if vortioxetine is better than those. Wagner et al are able to find a study comparing vortioxetine to the SSRI Paxil; they work about equally well.

What about the other claims? Weirdly, vortioxetine patients have more nausea and vomiting than venlafaxine patients, although it's not significant. Other studies confirm nausea is a pretty serious vortioxetine side effect. I have no explanation for this. Antagonizing 5-HT3 receptors does one thing – treats nausea and vomiting! – and vortioxetine definitely does this. It must be hitting some other unknown receptor really hard, so hard that the 5-HT3 antagonism doesn't counterbalance it. Either that, or it's the thing where God hates antidepressants again.

What about sexual dysfunction? Jacobsen et al find that patients have slightly (but statistically significantly) less sexual dysfunction on vortioxetine than on escitalopram. But the study was done by Takeda, and the difference is so slight (a change of 8.8 points on a 60 point scale, vs. a change of 6.6 points) that it's hard to take it very seriously.

What about cognition? I was sure this was fake, but it seems to have more evidence behind it than anything else. Carlat Report (paywalled) thinks it might be legit, based on a series of (Takeda-sponsored) studies of performance on the Digit Symbol Substitution Test. People on vortioxetine consistently did better on this test than people on placebo or duloxetine. And it wasn't just that being not-depressed helps you try harder; they did some complicated statistics and found that vortioxetine's test-score-improving effect was independent of its antidepressant effect (and seems to work at lower doses). What's the catch? The improvement was pretty minimal, and only shows up on this one test – various other cognitive tests are unaffected. So it's probably doing *something* measurable, but it's not going to give you a leg up on the SAT. The FDA seriously considered approving it as indicated for helping cognition, but eventually decided against it on the grounds that if they approved it, people would think it was useful in real life, whereas all we know is that it’s useful on this one kind of hokey test. Still, that’s one hokey test more than vilazodone was ever able to show for itself.

In summary, vortioxetine probably treats depression about as well as any other antidepressant, but makes you slightly more nauseous, may (if you *really* trust pharma company studies) give you slightly fewer sexual side effects, and may improve your performance on the Digit Symbol Substitution Test. It also costs $375 a month (Lexapro still costs $7). If you want to pay $368 extra to be a little more nauseous and substitute digits for symbols a little faster, this is definitely the drug for you.

VI.

In conclusion, big pharma spent about ten years seeing if combining 5-HT1A partial agonism with SSRI antidepressants led to any benefits. In the end, it didn’t, unless you count benefits to big pharma’s bottom line.

I’m a little baffled, because pharma companies generally don’t waste money researching drugs unless they have very good theoretical reasons to think they’ll work. But I can’t make heads or tails of the theoretical case for 5-HT1A partial agonists for depression.

For one thing, there’s a pretty strong argument that buspirone exerts its effects via dopamine rather than serotonin – a case that it seems like nobody, including the pharma companies, is even slightly aware of. If this were true, the whole project would have been doomed from the beginning. What happened here?

For another, I still don’t get the supposed model for how buspirone even *could* exert its effects through 5-HT1A. Does it increase or decrease serotonergic transmission? Does it desensitize presynaptic autoreceptors the same way SSRIs do, or do something else? I can’t figure out a combination of answers to this question that are consistent with each other and with the known effects of these drugs. Is there one?

For another, the case seems to have been premised on the idea that buspirone (a presynaptic 5-HT1A agonist) would work the same as pindolol (a presynaptic 5-HT1A antagonist), even after studies showed that it didn’t. And then it combined that with an assumption that it was better to spend hundreds of millions of dollars discovering a drug that combined SSRI and buspirone-like effects, rather than just giving someone a pill of SSRI powder mixed with buspirone powder. Why?

I would be grateful if some friendly pharmacologist reading this were to comment with their take on these questions. This is supposed to be my area of expertise, and I have to admit I am stumped.

Slightly Skew Systems Of Government

*[**Related To:** Legal Systems Very Different From Ours Because I Just Made Them Up, List Of Fictional Drugs Banned By The FDA]*

I.

Clamzoria is an acausal democracy.

The problem with democracy is that elections happen before the winning candidate takes office. If somebody's never been President, how are you supposed to judge how good a President they'd be? Clamzoria realized this was dumb, and moved elections to the last day of an official's term.

When the outgoing President left office, the country would hold an election. It was run by approval voting: you could either approve or disapprove of the candidate who had just held power. The results were tabulated, announced, and then nobody ever thought about them again.

Clamzoria chose its officials through a prediction market. The Central Bank released bonds for each candidate, which paid out X dollars at term's end, where X was the percent of voters who voted Approve. Traders could provisionally buy and sell these bonds. On the first day of the term, whichever candidate's bonds were trading at the highest value was inaugurated as the new President; everyone else's bonds were retroactively cancelled and their traders refunded. The President would spend a term in office, the election would be held, and the bondholders would be reimbursed the appropriate amount.

The Clamzorians argued this protected against demagoguery. It's easy for a candidate to promise the sun and moon before an election, but by the end of their term, voters know if the country is doing well or not. Instead of running on a platform of popular (but doomed) ideas, candidates are encouraged to run on a platform of unpopular ideas, as long as those unpopular ideas will genuinely make the country richer, safer, stronger, and all the other things that lead people to approve of a President’s term after the fact. Of course, you’re still limited by bond traders’ ability to predict which policies will work, but bond traders are usually more sober than the general electorate.

This system worked wonderfully for several decades, until Lord Bloodholme’s administration. He ran for President on an unconventional platform: if elected, he would declare himself Dictator-For-Life, replace democracy with sham elections, and kill all who opposed him. Based on his personality, all the bond traders found this completely believable. But that meant that in the end-of-term election, he would get 100% approval. His bond shot up to be worth nearly $100, the highest any bond had ever gone, and he won in a landslide. Alas, Lord Bloodholme was as good as his word, and – after a single sham election to ensure the bondholders got what they were due – that was the end of Clamzoria’s acausal democracy.

II.

Cognito is a constitutional mobocracy.

It used to be a regular mobocracy. It had a weak central government, radicals would protest whenever they didn’t like its decisions, the protests would shut down major cities, and the government would cave. Then people on the other side would protest, and that would also shut down major cities, and the government would backtrack. Eventually they realized they needed a better way, made a virtue out of necessity, and wrote the whole system into their constitution.

The Executive Branch is a president elected by some voting system that basically ensures a bland moderate. They have limited power to make decrees that enforce the will of the legislature. The legislature is the mob. One proposes a bill by having a protest in favor of it. If the protest attracts enough people – the most recent number is 43,617, but it changes every year based on the population and a few other factors – then the bill is considered up for review. Anyone can propose amendments (by having a protest demanding amendments) or vote against it – (by having a protest larger than the original protest demanding that the bill not be passed). After everyone has had a fair chance to protest, the text of the bill supported by the largest protest becomes law (unless the largest protest was against any change, in which case there is no change).

The Cognitans appreciate their system because protests are peaceful and nondisruptive. The government has a specific Protesting Square in every city with a nice grid that lets them count how many protesters there are, and all protests involve going into the Protesting Square, standing still for a few minutes to let neutral observers count people up, and then going home. It's silly to protest beyond this; your protest wouldn't be legally binding!

There's been some concern recently that corprorations pay protesters to protest for things they want. Several consumer watchdog organizations are trying to organize mobs in favor of a bill to stop this.

III.

Yyphrostikoth is a meta-republic.

Every form of government has its own advantages and disadvantages, and the goal is to create a system of checks and balances where each can watch over the others. The Yyphrostikoth Governing Council has twelve members:

The Representative For Monarchy is a hereditary position.

The Representative For Democracy is elected.

The Representative For Plutocracy is the richest person in the country.

The Representative For Technocracy is chosen by lot from among the country's Nobel Prize winners.

The Representative For Meritocracy is whoever gets the highest score on a standardized test of general knowledge and reasoning ability.

The Representative For Military Dictatorship is the top general in the army.

The Representative For Communism is the leader of the largest labor union.

The Representative For Futarchy is whoever has the best record on the local version of Metaculus.

The Representative For Gerontocracy is supposedly the oldest person in the country who is medically fit and willing to serve, but this has been so hard to sort out that in practice they are selected by the national retirees' special interest group from the pool of willing candidates above age 90.

The Representative For Minarchy is an honorary position usually bestowed upon a respected libertarian philosopher or activist. It doesn’t really matter who holds it, because their only job is to vote “no” on everything, except things that are sneakily phrased so that “no” means more government, in which case they can vote “yes”. If a Representative For Minarchy wants to vote their conscience, they may break this rule once, after which they must resign and be replaced by a new Representative.

The Representative For Republicanism is selected by the other eleven members of the council.

The Representative For Theocracy is the leader of the Governing Council, and gets not only her own vote but a special vote to break any ties. She is chosen at random from a lottery of all adult citizens, on the grounds that God may pick whoever He pleases to represent Himself.

Long ago, the twelfth Councilor was the Representative For Kratocracy (rule by the strongest). The Representative For Kratocracy was whoever was sitting in the Representative For Kratocracy’s chair when a vote took place. This usually involved a lot of firefights and hostage situations, which was fine in principle – that was the whole point – except that the rest of the Governing Council kept getting caught in the crossfire. During the Nehanian Restoration, the Representative For Kratocracy’s chair was moved to a remote uninhabited island, with the Representative permitted to vote by video-link, but environmentalist groups complained that the constant militia battles there were harming migratory birds. Finally, a petition was sent to the Oracle of Yaanek, asking what to do. The God recommended that the position be eliminated, and offered to decide who filled the newly vacated seat Himself; thus the beginning of the Representative For Theocracy.

The Constitution was never fully amended, so technically the position is still the Representative For Kratocracy, and technically anyone who kills the Representative For Theocracy can still take his seat and gain immense power. But for some reason everyone who tries this dies of completely natural causes just before their plan comes to fruition. Must be one of those coincidences.

NYT Is Threatening My Safety By Revealing My Real Name, So I Am Deleting The Blog

[EDIT 2/13/21: This post is originally from June 2020, but there's been renewed interest in it because the NYT article involved just came out. This post says the NYT was going to write a positive article, which was the impression I got in June 2020. The actual article was very negative; I feel this was as retaliation for writing this post, but I can't prove it. I feel I was misrepresented by slicing and dicing quotations in a way that made me sound like a far-right nutcase; I am actually a liberal Democrat who voted for Warren in the primary and Biden in the general, and I generally hold pretty standard center-left views in support of race and gender equality. You can read my full statement defending against the Times' allegations here. To learn more about this blog and read older posts, go to the About page.]

So, I kind of deleted the blog. Sorry. Here's my explanation.

Last week I talked to a *New York Times* technology reporter who was planning to write a story on Slate Star Codex. He told me it would be a mostly positive piece about how we were an interesting gathering place for people in tech, and how we were ahead of the curve on some aspects of the coronavirus situation. It probably would have been a very nice article.

Unfortunately, he told me he had discovered my real name and would reveal it in the article, ie doxx me. "Scott Alexander" is my real first and middle name, but I've tried to keep my last name secret. I haven't always done great at this, but I've done better than "have it get printed in the *New York Times*".

I have a lot of reasons for staying pseudonymous. First, I'm a psychiatrist, and psychiatrists are kind of obsessive about preventing their patients from knowing anything about who they are outside of work. You can read more about this in this *Scientific American* article – and remember that the last psychiatrist blogger to get doxxed abandoned his blog too. I am not one of the big sticklers on this, but I'm more of a stickler than "let the *New York Times* tell my patients where they can find my personal blog". I think it's plausible that if I became a national news figure under my real name, my patients – who run the gamut from far-left to far-right – wouldn't be able to engage with me in a normal therapeutic way. I also worry that my clinic would decide I am more of a liability than an asset and let me go, which would leave hundreds of patients in a dangerous situation as we tried to transition their care.

The second reason is more prosaic: some people want to kill me or ruin my life, and I would prefer not to make it too easy. I've received various death threats. I had someone on an anti-psychiatry subreddit put out a bounty for any information that could take me down (the mods deleted the post quickly, which I am grateful for). I've had dissatisfied blog readers call my work pretending to be dissatisfied patients in order to get me fired. And I recently learned that someone on SSC got SWATted in a way that they link to using their real name on the blog. I live with ten housemates including a three-year-old and an infant, and I would prefer this not happen to me or to them. Although I realize I accept some risk of this just by writing a blog with imperfect anonymity, getting doxxed on national news would take it to another level.

When I expressed these fears to the reporter, he said that it was *New York Times* policy to include real names, and he couldn't change that.

After considering my options, I decided on the one you see now. If there's no blog, there's no story. Or at least the story will have to include some discussion of NYT's strategy of doxxing random bloggers for clicks.

I want to make it clear that I'm not saying I believe I'm above news coverage, or that people shouldn't be allowed to express their opinion of my blog. If someone wants to write a hit piece about me, whatever, that's life. If someone thinks I am so egregious that I don't deserve the mask of anonymity, then I guess they have to name me, the same way they name criminals and terrorists. This wasn't that. By all indications, this was just going to be a nice piece saying I got some things about coronavirus right early on. Getting punished for my crimes would at least be predictable, but I am not willing to be punished for my virtues.

I'm not sure what happens next. In my ideal world, the *New York Times* realizes they screwed up, promises not to use my real name in the article, and promises to rethink their strategy of doxxing random bloggers for clicks. Then I put the blog back up (of course I backed it up! I'm not a monster!) and we forget this ever happened.

Otherwise, I'm going to lie low for a while and see what happens. Maybe all my fears are totally overblown and nothing happens and I feel dumb. Maybe I get fired and keeping my job stops mattering. I'm not sure. I'd feel stupid if I caused the amount of ruckus this will probably cause and then caved and reopened immediately. But I would also be surprised if I never came back. We'll see.

I’ve gotten an amazing amount of support the past few days as this situation played out. You don’t need to send me more – message *very much* received. I love all of you so much. I realize I am making your lives harder by taking the blog down. At some point I’ll figure out a way to make it up to you.

In the meantime, you can still use the r/slatestarcodex subreddit for sober non-political discussion, the not-officially-affiliated-with-us r/themotte subreddit for crazy heated political debate, and the SSC Discord server for whatever it is people do on Discord. Also, my biggest regret is I won’t get to blog about Gwern’s work with GPT-3, so go over and check it out.

There’s a SUBSCRIBE BY EMAIL button on the right – put your name there if you want to know if the blog restarts or something else interesting happens. I’ll make sure all relevant updates make it onto the subreddit, so watch that space.

~~There is no comments section for this post. The appropriate comments section is the feedback page of the *New York Times*. You may also want to email the *New York Times* technology editor Pui-Wing Tam at pui-wing.tam@nytimes.com, contact her on Twitter at @puiwingtam, or phone the *New York Times* at 844-NYTNEWS.~~ [EDIT: The time for doing this has passed, thanks to everyone who sent messages in]

(please **be polite** – I don’t know if Ms. Tam was personally involved in this decision, and whoever is stuck answering feedback forms definitely wasn’t. Remember that you are representing me and the SSC community, and I will be very sad if you are a jerk to anybody. Please just explain the situation and ask them to stop doxxing random bloggers for clicks. If you are some sort of important tech person who the New York Times technology section might want to maintain good relations with, mention that.)

~~If you are a journalist who is willing to respect my desire for pseudonymity, I'm interested in talking to you about this situation (though I prefer communicating through text, not phone). My email is scott@slatestarcodex.com.~~ [EDIT: Now over capacity for interviews, sorry]

Update On My Situation

It’s been two and a half months since I deleted the blog, so I owe all of you an update on recent events.

I haven’t heard anything from the *New York Times* one way or the other. Since nothing has been published, I’d assume they dropped the article, except that they approached an acquaintance for another interview last month. Overall I’m confused.

But they definitely haven’t given me any explicit reassurance that they won’t reveal my private information. And now that I’ve publicly admitted privacy is important to me – something I tried to avoid coming on too strong about before, for exactly this reason – some people have taken it upon themselves to post my real name all over Twitter in order to harass me. I probably inadvertently Streisand-Effect-ed myself with all this; I still think it was the right thing to do.

At this point I think maintaining anonymity is a losing battle. So I am gradually reworking my life to be compatible with the sort of publicity that circumstances seem to be forcing on me. I had a talk with my employer and we came to a mutual agreement that I would gradually transition away from working there. At some point, I may start my own private practice, where I’m my own boss and where I can focus on medication management – and not the kinds of psychotherapy that I’m most worried are ethically incompatible with being a public figure. I’m trying to do all of this maximally slowly and carefully and in a way that won’t cause undue burden to any of my patients, and it’s taking a long time to figure out.

I’m also talking to Substack about moving to their blogging platform. While part of me wants to jump right back into blogging here and pretend nothing ever happened, the Substack option has grown on me. I think I’d feel safer as part of a big group that specifically promises to defend their bloggers when needed. And also, I’d feel safer with a lot of diverse income streams, and Substack has made me an extremely generous offer. Many people gave me good advice about how I could monetize my blog without Substack – I took these suggestions very seriously, and without violating a confidentiality agreement all I can answer is that Substack’s offer was *extremely* generous.

When I originally asked readers about this possibility, they raised a lot of valid concerns: some of them were confused by Substack’s commenting system, others annoyed by its pop-up reminders to subscribe, others were concerned about being stuck outside a paywall. I’ve talked to Substack about this, and they’ve made some really impressive promises to address these things – they’re going to code a maximally-SSC-like commenting experience, they’re going to let me opt out of the subscription reminders, I won’t have to “paywall” anything besides some Hidden Open Threads. This isn’t the time for me to go over the dozens of examples of concerns I had that Substack went above and beyond to address, but assume I had most of the same ones you did and put a lot of work into addressing them.

(and if you’re worried about the Hidden Open Threads, check out Data Secrets Lox, a forum that has done a great job keeping the SSC Open Thread tradition going over the past few months.)

So that’s where I am right now – trying to wind things down at my day job, very preliminarily planning a private practice, and negotiating writing details with Substack. I’m also looking into some other things to protect my physical safety. When all of that is done, I'll start blogging again. Right now I'm expecting that to be some time between October and January – and obviously when it happens I'll let you know. I would appreciate if people continued to respect my preferences about anonymity until then. After that I'll stop caring as much – though I'll still go by "Scott Alexander" to keep the brand the same, and I'll still do what I can to avoid publicity.

I might have hinted at this already, but I should say it explicitly – I'm really grateful for all the support I got throughout this whole incident. You people are all great. I'll say so at more length later, and talk more about some specific examples, but for now just accept on faith that you're all great.

I still plan to do the book review contest! I'll do it sometime after I start the new blog! Those of you who sent me reviews didn't waste your time! It's going to happen! Pestilence may afflict every corner of the world, the skies may turn red as blood and the sun go dark at noon, the earth may shake and plagues of locusts cover the land, but never doubt that there will be a book review contest someday, in the golden future, when all of this is over.

As all the kids are saying these days, "thank you for your continuing support during these difficult times".

Table of Contents

- Slate Star Codex
  - Favorites
    - Top Posts
    - Archipelago And Atomic Communitarianism
    - Beware The Man Of One Study
    - Book Review: Albion’s Seed
    - Considerations On Cost Disease
    - I Can Tolerate Anything Except The Outgroup
    - Meditations on Moloch
    - Nobody Is Perfect, Everything Is Commensurable
    - Studies On Slack
    - The Categories Were Made For Man, Not Man For The Categories
    - The Control Group Is Out Of Control
    - Toxoplasma of Rage
    - Who By Very Slow Decay
  - 2013
    - You’re probably wondering why I’ve called you here today
    - Abraham Lincoln, Ape-Man
    - Abraham Lincoln, Necromancer
    - You cannot serve both God and Mammon, but you can serve Mammon by predicting who will serve God
    - Google Correlate does not imply Google Causation
    - 90% of all claims about the problems with medical studies are wrong
    - Typical mind and gender identity
    - I liked Lovecraft countless primaeval aeons before it was cool
    - Future tense

A defense of logical positivism (yes, really)
Dungeons and Discourse Third Edition: The dialectic continues
People! It turns out capitalism is actually a thing!
Fermat’s Last Stand Coordination Thread
MetaMed launch day
Pope and Change: An Atheist’s Guide To Vatican Decision 2013
Sleep
Reactionary Philosophy In An Enormous, Planet-Sized Nutshell
A Thrive/Survive Theory Of The Political Spectrum
Yet Another Parable On The Importance Of Controlling Your Experiments
We Wrestle Not With Flesh And Blood, But Against Powers And Principalities
We Sail Tonight For Singapore
Hitting Below The (Bible) Belt
Good vibrations and subjective temporal granularity
Fermat’s Last Stand: Game Rosters
Valleys have two sides
Annuntio vobis gaudium magnum: habemus officium!
Not just a mere political issue
My objections to “objectification”
If opposites attract, why is my sister so pretty?
Book Review: Infinite Jest (alternate title: “Look At Me! I Read Infinite Jest!”)
The death of wages is sin
My Thoughts On The First 40 Pages Of Alastair MacIntyre’s “After Virtue”, Expressed In The Form Of A Slightly Modified XKCD Comic

Thank you for doing something ambiguously between smoking and not smoking
A good pun is its own reword
Against Anton-Wilsonism
Lies, Damned Lies, And Facebook (Part 1 of ∞)
Lies, Damned Lies, And Facebook (Part 2 of ∞)
Investment and Inefficient Charity
Poor Folks Do Smile…For Now
Polyamory Is Boring
RPG emergency!
Whose Utilitarianism?
Book Review: After Virtue
Read History Of Philosophy Backwards
Virtue Ethics: Not Practically Useful, Either
Lizardman’s Constant Is 4%
Last Thoughts On Virtue Ethics
If a clod be washed away by the sea, Europe is the less
Proving Too Much
Things I Learned By Spending Five Thousand Years In An Alternate Universe
Floor Employment
Arguments About Male Violence Prove Too Much
I do not understand “rape culture”
Social Justice For The Highly-Demanding-Of-Rigor
Search Terms That Have Led People To This Blog
Implicit Association Tests and Suicidality
Product Recommendation: FB Purity
In Defense of Psych Treatment for Attempted Suicide
Utilitarianism for Engineers, Part II
Yellow Journalism In The War On Science
Utility Weight Results
If It’s Worth Doing, It’s Worth Doing With Made-Up Statistics

Selection bias and atheist stereotypes
Ambijectivity
Raikoth: Laws, Language, and Society
Raikoth: Corruption, Priesthood
Raikoth: Cities, Land
Raikoth: Symbolic Beads
Overheard in the Box
Raikoth: Economics, Relationships
Raikoth: History, Religion
Index: Posts on Raikoth
Overheard In The Box, Part II
Newtonian Ethics
Against Bravery Debates
Can You Condition Yourself?
Apart From Better Sanitation And Medicine And Education And Irrigation And Public Health And Roads And Public Order, What Has Modernity Done For Us?
The Wisdom of the Ancients
Going From California With An Aching In My Heart.
The World Is Flat And Intensively Farmed
Transhumanist Fables
Fetal Attraction: Abortion and the Principle of Charity
Hansonian Optimism
Wisdom of the Ancients, Redux
Literally Inconceivable: Contraceptives And Abortion Rates
Hasta La Victorians Siempre
Moments Of Awakening
All Debates Are Bravery Debates
Lies, Damned Lies, And Facebook (Part 3 of ∞)
Medical and Psychological Studies for June
Biodjinnetics

Arguments From My Opponent Believes Something
The Virtue of Silence
Can Atheists Appreciate Chesterton?
The What-You’d-Implicitly-Heard-Before Telling Thing
A New Chapter In The Codex
Social Psychology Is A Flamethrower
The Hospital Orientation
Nature Is Not A Slate. It’s A Series Of Levers.
The Lottery of Fascinations
I Aten’t Dead
Who By Very Slow Decay
I Myself Am A Scientismist
Holocaust Good For You, Research Finds, But Frequent Taunting Causes Cancer In Rats
On first looking into Chapman’s “Pop Bayesianism”
Extreme Mnemonics
Fermat’s Last Stand: Soundtrack and Adventure Log
Things I Don’t Understand About Genetics (A Non-Exhaustive List)
Fake Consensualism
Fake Euthanasia Statistics
Military Strikes Are An Extremely Cheap Way To Help Foreigners
The Life Cycle of Medical Ideas
Giving and Accepting Apologies
Inefficient Hot Dogs
Scientific Freud
The Thin Blue Line That Stays Strangely Horizontal
Prediction is very difficult, especially of the past
The Trend Evaluation Of All Values
Sleep – Now By Prescription

- The Anti-Reactionary FAQ
- Some preliminary responses to responses to the Anti-Reactionary FAQ
- List of Fictional Drugs Banned By The FDA
- The Poor You Will Always Have With You
- The Witching Hour
- List of Reasons Why Gunpowder Treason Should Be Forgot
- Lies, Damned Lies, And Facebook (Part 4 of ∞)
- Highlights From My Notes From A Forensic Psychiatry Conference Last Week
- Genetic Russian Roulette
- November Housekeeping Post
- A Letter I Will Probably Send To The FDA
- The Story Of Thanksgiving Is A Science-Fiction Story
- Empire/Forest Fire
- More Fictional Drugs Banned By The FDA
- The Logician And The God-Emperor
- A Something Sort Of Like Left-Libertarianism-ist Manifesto
- Papal PR
- Statistical Literacy Among Doctors Now Lower Than Chance
- Less Wrong, More Rite (II)
- Public Awareness Campaigns
- We Are All MsScribe
- [REPOST] A Christmas Poem
- Meetup: Southeast Michigan
- A Comment I Posted On “What Would JT Do?”
- The Spirit Of The First Amendment

2014

- Two Dark Side Statistics Papers
- Marijuana: Much More Than You Wanted To Know
- A Response To Apophemi on Triggers
- Aretaeus On Bipolar Disorder

Ten Things I Want To Stop Seeing On The Internet In 2014
Genetic Testing and Self-Fulfilling Prophecies
Beware Mass-Produced Medical Recommendations
Predictions For 2014
Wirehead Gods On Lotus Thrones
Based on your findings, which theory about alien thickness seems most valid or most accurate?
More Search Terms That Have Led People To This Blog
Seeming and Being Empathetic
Blogging The Anissimov – Smith Reaction Debate
Housekeeping Post February 2014
Bad Dreams
Nootropics Survey Results And Analysis
Lies, Damned Lies, And Social Media (Part 5 of ∞)
Proposed Biological Explanations For Historical Trends In Crime
God Bless Longecity
Effects Of Vertical Acceleration On Wrongness
In Favor of Niceness, Community, and Civilization
Suggest A Comments Policy
Fix Science In Half An Hour
Unteachable Things Hard To Teach, Study Suggests
Searching For One-Sided Tradeoffs
[CORRECTION] Mountain View Meetup 3/4
The Comment Policy Is “Victorian Sufi Buddha Lite”
Do Life Hacks Ever Reach Fixation?
Book Review: The Perfect Health Diet

The Slate Star Codex Political Spectrum Quiz
Semite Times: The Bible In Palindromes
Five Years and One Week of Less Wrong
Can It Be Wrong To Crystallize Patterns?
What Universal Human Experiences Are You Missing Without Realizing It?
Michigan Meetup and Memory Workshop, 3/23
Typical Mind and Disbelief In Straight People
Should You Reverse Any Advice You Hear?
E-Cig Study: Much Smoke, Little Light
The Character's Complaint
Wheat: Much More Than You Wanted To Know
The Study of Anglophysics
My New OKCupid Profile
Housekeeping Post April 2014
Dermatology Pop Quiz
Dermatology Quiz: Answers
Going Loopy
The Cowpox of Doubt
Do You Believe Me, Doc?
Someone Writes An Anti-Racist FAQ
Confounder Of The Day: How Sexy Your Parents Were
Plutocracy Isn't About Money
The Economics Of Art And The Art Of Economics
Right Is The New Left
Friendship Is Countersignaling
Metaphors Be With You
Stop Confounding Yourself! Stop Confounding Yourself!
The Control Group Is Out Of Control
Utopian Science
Causal Models At Work
Emotionally Valent Links Post
Not Even A Real Links Post, Just A Link
Evening Doc

Not Even A Real Links Post, Just A Blatant Ad
Vote On Values, Outsource Beliefs
Infinite Debt
Weak Men Are Superweapons
Getting A Therapist
You Kant Dismiss Universalizability
Nerds Can Be Bees, Too
SSC Gives A Graduation Speech
Nydwracu’s Fnords
Apologia Pro Vita Sua
Compound Interest Is The Least Powerful Force In The Universe
On Types of Typologies
Don’t Fear The Filter
The Wonderful Thing About Triggers
Southeast Michigan Meetup 6/8
Asches to Asches
Don’t Be An Asch-Hole
Archipelago and Atomic Communitarianism
Map-Territory Distinctions
Minutes From The Michigan Rationalist Meetup
CONSTANT VIGILANCE!
List Of The Passages I Highlighted In My Copy Of Jonathan Haidt’s “The Righteous Mind”
Living By The Sword
Fish – Now By Prescription
Things That Sometimes Help If You Have Depression
Money On The Ground
Ground Morality In Party Politics
The Art Of Writing Randian Monologues
Court-ing Disaster
The Placebo-Singers
SSC Gets World Cup Fever
Book Review: The Two-Income Trap
Invisible Women

Medicine, As Not Seen On TV
How Common Are Science Failures?
Ozy vs. Scott on Charity Baskets
List Of The Passages I Highlighted In My Copy Of “The Two-Income Trap”
Social Justice And Words, Words, Words
SSRIs: Much More Than You Wanted To Know
Growing Children For Bostrom’s Disneyland
A Paradox of Ecclesiology
Psychotropic Base Rates: The Argument From Antibiotics
HeartMath Considered Incoherent
Promising The Moon
No Skyscraper Stagnation
The Other Codex
Some Antibiotic Stagnation
Weird Psychiatric Ads Of The Seventies
Meditations On Moloch
Misperceptions On Moloch
Unspoken Ground Assumptions Of Discussion
Negative Creativity
Random Noise Is Our Most Valuable Resource
Friendship Is (Still) Countersignaling
Getting Eulered
Does the Glasgow Coma Scale exist? Do comas?
Does Race Exist? Does Culture?
Beware Isolated Demands For Rigor
Fifty Swifties
Burdens
An Iron Curtain Has Descended Upon Psychopharmacology
Ozy’s Anti Heartiste FAQ
Nobody Likes A Tattletale
The Invisible Nation – Reconciling Utilitarianism And Contractualism

If The Media Reported On Other Dangers Like It Does AI Risk
Michigan Meetup 9/7
Radicalizing the Romanceless
Book Review and Highlights: Quantum Computing Since Democritus
The Guardian vs. Induction
Cooperation Un-Veiled
Mapmaker, Mapmaker, Make Me A Map
Society Is Fixed, Biology Is Mutable
What The Hell, Hegel?
Book Review: Singer on Marx
Does Class Warfare Have A Free-Rider Problem?
[Ozy] A Response to Spandrell
Joint Over- and Underdiagnosis
SSC Gives A Wedding Speech
Book Review: Red Plenty
Streetlight Psychology
Why No Science Of Nerds?
Cuddle Culture
Bottomless Pits Of Suffering
I Can Tolerate Anything Except The Outgroup
The Battle Hymn
Simpler Times
Prediction Goes To War
Tumblr on MIRI
Five Planets In Search Of A Sci-Fi Story
Five Case Studies On Politicization
In The Future, Everyone Will Be Famous To Fifteen People
Book Review: A Future For Socialism
Alcoholics Anonymous: Much More Than You Wanted To Know
Michigan Meetup 11/9
All In All, Another Brick In The Motte
Ethnic Tension And Meaningless Arguments

The Right To Waive Your Rights
Growing Old
How To Use 23andMe Irresponsibly
The Dark Side Of Divorce
Ley Lines Of The Midwest
Republicans Are Douchebags
Prisons Are Built With Bricks Of Law And Brothels With Bricks Of Religion, But That Doesn’t Prove A Causal Relationship
The Categories Were Made For Man, Not Man For The Categories
Race and Justice: Much More Than You Wanted To Know
Why I Am Not Rene Descartes
Such Mixed Feelings About Crazymeds
Cyber Monday Product Recommendations
Book Review: On The Road
Framing For Light Instead Of Heat
A Story With Zombies
Beware The Man Of One Study
Debunked And Well-Refuted
The Toxoplasma Of Rage
Nobody Is Perfect, Everything Is Commensurable
There Are Rules Here
Book Review: What’s Wrong With The World

2015

Untitled
2014 Predictions: Calibration Results
Ethics Offsets
Chronic Psychitis
The Phatic And The Anti-Inductive
The Physics Diet?
The Influenza Of Evil
Depression Is Not A Proxy For Social Dysfunction
These Are A Few (More) Of My (Least) Favorite Things

Perceptions Of Required Ability Act As A Proxy For Actual Required Ability In Explaining The Gender Gap
A Philosopher Walks Into A Coffee Shop
The Parable Of The Talents
Talents Part 2: Attitude vs. Altitude
Practically-A-Book Review: Dying To Be Free
Everything Not Obligatory Is Forbidden
The Efficacy Of Everything In Psychiatry In One Graph Plus Several Pages Of Dense But Necessary Explanation
Black People Less Likely
Money, Money, Everywhere, But Not A Cent To Spend
Drug Testing Welfare Users Is A Sham, But Not For The Reasons You Think
How Likely Are Multifactorial Trends?
Did Falling Testosterone Affect Falling Crime?
Pharma Virumque
California Meetups
Early Intervention: You *Might* Get What You Pay For
California Meetups (Rererevised)
A Series Of Unprincipled Exceptions
A Cascade Of Dunbar Numbers
Effective Altruists: Not As Mentally Ill As You Think
Too Good To Be True
Book Review: Willpower
Answer to Job
List Of Passages I Highlighted In My Copy Of “Willpower”
Book Review: The Machinery Of Freedom
List Of Passages I Highlighted In My Copy Of “Machinery Of Freedom”
Is Everything A Religion?

Extremism In Thought Experiment Is No Vice
Highlights From My Notes From Another Psychiatry Conference
Rational Orthography
Chemical Imbalance
No Physical Substrate, No Problem
No Clarity Around Growth Mindset
I Will Never Have The Ability To Clearly Explain My Beliefs About Growth Mindset
Michigan Meetup 4/14
Trouble Walking Down The Hallway
Polemical Imbalance
Blame Theory
Universal Love, Said The Cactus Person
Growth Mindset 3: A Pox On Growth Your Houses
Nefarious Nefazodone And Flashy Rare Side Effects
Prescriptions, Paradoxes, and Perversities
The Future Is Filters
Growth Mindset 4: Growth Of Office
California, Water You Doing?
Bicameral Reasoning
Beware Summary Statistics
AI Researchers On AI Risk
No Time Like The Present For AI Safety Work
That Chocolate Study
...And I Show You How Deep The Rabbit Hole Goes
Against Tulip Subsidies
Michigan Meetup 6/13
(Late) Predictions for 2015
Fearful Symmetry
Fifty (More) Swifties
NYC Meetup 6/21
Reflections From The Halfway Point
The Case Of The Famous Physicist

The Argument From Cultural Evolution
Cultural Evolution 2: Thanks For The Meme-Rise
Things That Sometimes Work If You Have Anxiety
CBT In The Water Supply
Reverse Psychology
Freedom On The Centralized Web
The General Factor Of Correctness
Non-Dual Awareness
Stalin and Summary Statistics
Contra Hallquist On Scientific Rationality
Contrarians, Crackpots, and Consensus
Book Review: Chronicles Of Wasted Time
Stop Adding Zeroes
Figure/Ground Illusions
My Id On Defensiveness
The Goddess of Everything Else
On Overconfidence
Probabilities Without Models
Mysticism and Pattern-Matching
Magic Markers
Theses On Trump
If You Can’t Make Predictions, You’re Still In A Crisis
Book Review: Manufacturing Consent
Cardiologists and Chinese Robbers
Beware Systemic Change
Vegetarianism for Meat-Eaters
The Problems With Generic Medications Go Deeper Than One Company
Swifties 3: The Race Is Not To The Swifty
Contra Caplan on Mental Illness
Against Against Autism Cures
It Was You Who Made My Blue Eyes Blue
OT 30: Comment Knowledge
Contra Simler on Prestige

A Whiter Shade of Candidate
Contra Huemer On Morals
Meat Your Doom
What Developmental Milestones Are You Missing?
2014 SSC Survey Results
Looking A Gift Horse In The Mouth
2D:4D Ratio And Psychological Traits: Results From The LW/SSC Survey Sample
Hardball Questions For The Next Debate
SSC New England Meetup And Presentations Schedule
Reporter Degrees Of Freedom
College And Critical Thinking
Setting The Default
Product Recommendations 2015
Book Review: Hive Mind
List Of Passages I Highlighted In My Copy Of “Hive Mind”
Should AI Be Open?
How Bad Are Things?
Things That Are Not Superintelligences
Introducing Unsong

2016

2015 Predictions: Calibration Results
Guns And States
Guns and States 2: Son of a Gun
Slow But Steady
Schizophrenia: No Smoking Gun
Lies, Damned Lies, And The Media (Part 6 of ∞)
Side Effects May Include Anything
Predictions For 2016
Staying Classy
Book Review: Superforecasting
List Of Passages I Highlighted In My Copy Of “Superforecasting”

Testimonials for SSC
Before You Get Too Excited About That GitHub Study…
Nonfiction Writing Advice
Two Attitudes In Psychiatry
2016 Nootropics Survey Results
Against Interminable Arguments
Reverse Voxsplaining: Prison and Mental Illness
Non-Shared Environment Doesn’t Just Mean Schools And Peers
Book Review: The Art Of The Deal
The Price Of Glee In China
Book Review: My Brother Ron
Beware Regional Scatterplots
The Ideology Is Not The Movement
A Theory About Religion
SSC Bay Area Meetups 2016
Book Review: Albion’s Seed
Why Were Early Psychedelicists So Weird?
Be Nice, At Least Until You Can Coordinate Meanness
Myers’ Race Car Versus The General Fitness Factor
Hidden Test Post
Solidarity
Skin In The Game
Teachers: Much More Than You Wanted To Know
Three Great Articles On Poverty, And Why I Disagree With All Of Them
Book Review: Age of Em
Ascended Economy?
Ketamine Research In A New Light
Three More Articles On Poverty, And Why They Disagree With Each Other
Against Dog Whistle-ism
Book Review: Unlearn Your Pain

I Wrote A Blog Post, But Did Not Adjust For The Fact That The Title Would Be Too L
Survey Results: Suffering Vs. Oblivion
3/4
Pushing And Pulling Goals
Things Probably Matter
How The West Was Won
Post-Partisanship Is Hyper-Partisanship
Book Review: PiHKaL
Tolerance Troubles
The View From Ground Level
Devoodooifying Psychology
Reverse Voxsplaining: Drugs vs. Chairs
Terrorists Vs. Chairs: An Outlier Story
Reverse Voxsplaining: Brand-Name Drugs
It’s Bayes All The Way Up
Some Context For That NYT Sugar Article
AI Persuasion Experiment
SSC Endorses Clinton, Johnson, Or Stein
He Kept Us Out Of War?
Somewhat Against Psychiatric Conditions As Domestication Failure
More Hardball Debate Questions
The Moral Of The Story
AI Persuasion Followup Survey
The Heart Has Its Reasons That Reason Knows Not Of
AI Persuasion Experiment Results
How Does Recent AI Progress Affect The Bostromian Paradigm?
SSC San Jose Meetup 11/4
The Pyramid And The Garden
Tuesday Shouldn’t Change The Narrative
Book Review: House of God
You Are Still Crying Wolf
The Alzheimer Photo

SSC Journal Club: Expert Prediction Of Experiments
Contra Robinson On Schooling
Highlights From The Comment Thread On School Choice
Thin Air
Might People On The Internet Sometimes Lie?
SSC Journal Club: Mental Disorders As Networks
Book Review: Mount Misery
Contra NYT On Economists On Education
Addendum To “Economists On Education”
2016 Predictions: Calibration Results

2017

Trump And The Batman Effect
Predictions For 2017
Heuristics Work Until They Don’t
Why Do Test Scores Plateau?
Should Buzzfeed Publish Claims Which Are Explosive If True But Not Yet Proven?
Another Followup To “Economists On Education”
Watch New Health Picks
SSC Survey 2017
Link: Guyenet On Taubes
Dogs And Wolves: In Defense Of Some Past Posts
Link: Taubes Contra Guyenet on Sugar
Book Review: Eichmann In Jerusalem
[GUEST POST] The International Refugee Assistance Project
Notes From The Asilomar Conference On Beneficial AI
Albion’s Seed, Genotyped
Considerations On Cost Disease
Highlights From The Comments On Cost Disease
[REPOST] The Non-Libertarian FAQ

Some Groups Of People Who May Not 100% Deserve Our Eternal Scorn
SSC Journal Club: Analytical Thinking Style And Religion
A Modern Myth
Antidepressant Pharmacogenomics: Much More Than You Wanted To Know
Book Review: Behavior – The Control Of Perception
Book Review: Seeing Like A State
SSC Survey 2017 Results
[REPOST] The Demiurge’s Older Brother
Guided By The Beauty Of Our Weapons
SSC Meetup: Everywhere
G.K. Chesterton On AI Risk
SSC Meetups Everywhere: Times And Places
Determining Consent
The Case Of The Suffocating Woman
Yes, We Have Noticed The Skulls
Sacred Principles As Exhaustible Resources
Clarification To “Sacred Principles As Exhaustible Resources”
Chametz
Learning To Love Scientific Consensus
SSC Journal Club: Childhood Trauma And Cognition
Book Review: The Hungry Brain
Anorexia And Metabolic Set Point
Neutral vs. Conservative: The Eternal Struggle
Getting High On Your Own Supply
Silicon Valley: A Reality Check
SSC Meetups: Berkeley, San Diego, Irvine
Bail Out
Polyamory Is Not Polygyny
Postmarketing Surveillance Is Good And Normal
Classified Ads Thread

Those Modern Pathologies
The Atomic Bomb Considered As Hungarian High School Science Fair Project
Hungarian Education II: Four Nobel Truths
Hungarian Education III: Mastering The Core Teachings Of The Budapestians
Is Pharma Research Worse Than Chance?
SSC Journal Club: AI Timelines
What Is Depression, Anyway?: The Synapse Hypothesis
Against Murderism
To Understand Polarization, Understand Conservativism’s Failures
Conversation Deliberately Skirts The Border Of Incomprehensibility
Why Are Transgender People Immune To Optical Illusions?
SSC Meetup: Chicago 7/2
To The Great City!
Two Kinds Of Caution
Change Minds Or Drive Turnout?
SSC Meetup: Salt Lake 7/14
Can We Link Perception And Cognition?
Classified Thread 2: Best In Classified
Targeting Meritocracy
Highlights From The Comment Thread On Meritocracy
Don’t Blame Griggs
Why Is Clozapine So Great?
Djoser Joseph Osiris
Against Signal-Boosting As Doxxing
Book Review: Raise A Genius!
Gender Imbalances Are Mostly Not Due To Offensive Attitudes
Is It Possible To Have Coherent Principles Around Free Speech Norms?

Where The Falling Einstein Meets The Rising Mouse
Why Not More Excitement About Prediction Aggregation?
Contra Grant On Exaggerated Differences
The Lizard People Of Alpha Draconis 1 Decided To Build An Ansible
Brief Cautionary Notes On Branded Combination Nootropics
Fear And Loathing At Effective Altruism Global 2017
Partial Credit
Axiology, Morality, Law
My IRB Nightmare
Highlights From The Comments On My IRB Nightmare
Book Review: Surfing Uncertainty
Predictive Processing And Perceptual Control
How Do We Get Breasts Out Of Bayes Theorem?
Toward A Predictive Theory Of Depression
Classified Thread 3: Semper Classifiedelis
Book Review: Mastering The Core Teachings Of The Buddha
Meditative States As Feedback Loops
Against Individual IQ Worries
Different Worlds
SSC Survey Results On Trust
In Favor Of Futurism Being About The Future
SSC Journal Club: Serotonin Receptors
SSC Meetup: Bay Area 10/14
Were There Dark Ages?
Highlights From The Comments On Dark Ages
Kolmogorov Complicity And The Parable Of Lightning
How Did New Atheism Fail So Miserably?
Against Rat Park

- Postmodernism For Rationalists (my attempt)
- Non-Expert Explanation
- Concept-Shaped Holes Can Be Impossible To Notice
- Does Age Bring Wisdom?
- Ars Longa, Vita Brevis
- Book Review: Legal Systems Very Different From Ours
- List Of Passages I Highlighted In My Copy Of Legal Systems Very Different From Ours
- Contra Robinson On Public Food
- A Completely Accurate Map With No Distortion At All
- Book Review: Inadequate Equilibria
- Against Overgendering Harassment
- The Tax Bill Compared To Other Very Expensive Things
- Response To Comments: The Tax Bill Is Still Very Bad
- Tax Bill 3: Don’t Mess With Taxes
- What To Make Of New Positive NSI-189 Results?
- Classified Thread 4: Vinson Classif
- Please Take The 2018 SSC Reader Survey
- Preregistration Of Hypotheses For The SSC Survey
- A History Of The Silmarils In The Fifth Age
- Adderall Risks: Much More Than You Wanted To Know

2018

- 2017 Predictions: Calibration Results
- SSC Survey Results 2018
- Book Review: Madness And Civilization
- Fight Me, Psychologists: Birth Order Effects Exist And Are Very Strong
- Self-Serving Bias

Maybe The Real Superintelligent AI Is Extremely Smart Computers
Bundles Of Joy
Practically-A-Book Review: Luna Whitepaper
Conflict Vs. Mistake
SSC Survey Data On Models Of Political Conflict
Highlights From The Comments On Conflict Vs. Mistake
The Invention Of Moral Narrative
Powerless Placebos
Predictions For 2018
Guyenet On Motivation
We’ve Got Five Years, What A Surprise
More Testimonials For SSC
Even More Search Terms That Led People To This Blog
Five More Years
Technological Unemployment: Much More Than You Wanted To Know
Current Affairs’ “Some Puzzles For Libertarians”, Treated As Writing Prompts For Short Stories
Highlights From The Comments On Technological Unemployment
SSC Journal Club: Cipriani On Antidepressants
SSC Meetup: Bay Area 3/3
God Help Us, Let’s Try To Understand Friston On Free Energy
SSC Journal Club: Friston On Computational Mood
Response To Current Affairs On Against Murderism
The Dark Rule Utilitarian Argument For Science Piracy
Navigating And/Or Avoiding The Inpatient Mental Health System
Book Review: Twelve Rules For Life

The Hour I First Believed
Are The Amish Unhappy? Super Happy? Just Meh?
Highlights From The Comments On Twelve Rules
Adult Neurogenesis – A Pointed Review
Why DC’s Low Graduation Rates?
Highlights From The Comments On DC Graduation Rates
Recommendations vs. Guidelines
SSC Survey Results: Sexual Harassment Levels By Field
Highlights From The Comments On Survey Harassment Rates
Gupta On Enlightenment
Mental Health On A Budget
Call For Adversarial Collaborations
Adversarial Collaboration Contest: Loose Ends And Registration
Book Review: History Of The Fabian Society
Classified Thread 5: Classified Never Sinned
Varieties Of Argumentative Experience
Basic Income, Not Basic Jobs: Against Hijacking Utopia
Can Things Be Both Popular And Silenced?
Should Psychiatry Test For Lead More?
Highlights From The Comments On Basic Jobs
In Search Of Missing US Suicides
HPPD And The Specter Of Permanent Side Effects
The GATTACA Trilogy
Contra Caplan On Arbitrary Deploring
Cost Disease In Medicine: The Practical Perspective
Book Review: Capital In The Twenty-First Century
List Of Passages I Highlighted In My Copy Of Capital In The Twenty-First Century

Highlights From The Comments On Piketty
SSC Journal Club: Dissolving The Fermi Paradox
The Craft And The Codex
Melatonin: Much More Than You Wanted To Know
Did A Melatonin Patent Inspire Current Dose Confusion?
Fundamental Value Differences Are Not That Fundamental
The Whole City Is Center
Sentimental Cartography
Value Differences As Differently Crystallized Metaphysical Heuristics
Verses Composed Upon Reading A Review From TripAdvisor
Meetups Everywhere 2018
Cancer Progress: Much More Than You Wanted To Know
Before You Get Too Excited About That Trigger Warning Study…
SSC Meetups 2018: Times And Places
SSC Survey Results: ADHD And Rejection Sensitivity
SSC Survey: Scattered Negative Results
The Parentheses Riddle
Practically-A-Book Review: EA Hotel
Carbon Dioxide: An Open Door Policy
Elegy For John McCain
Bulls**t Jobs (Part 1 of ∞)
Bureaucracy As Active Ingredient
This Week: Adversarial Collaboration Entries
Reminder: Bay Area SSC Meetup 9/8
[ACC Entry] Does The Education System Adequately Serve Advanced Students?
[ACC Entry] Are Islam And Liberal Democracy Compatible?

[ACC Entry] Should Childhood Vaccination Be Mandatory?
[ACC Entry] Should Transgender Children Transition?
In The Balance
The Omnigenic Model As Metaphor For Life
Book Review: The Black Swan
Treating The Prodrome
The Tails Coming Apart As Metaphor For Life
Adversarial Collaboration Contest Results
Steelmanning The NIMBYs
Highlights From The Comments On NIMBYs
Next Door In Nodrumia
Nighttime Ventilation Survey Results
Kavanaugh: A Probability Poll
Anxiety Sampler Kits
Is Science Slowing Down?
The Chamber of Guf
Classified Thread 6
Cognitive Enhancers: Mechanisms And Tradeoffs
Nominating Oneself For The Short End Of A Tradeoff
Sort By Controversial
Working With Google Trends
My California Ballot
Preschool: I Was Wrong
Marijuana: An Update
SSRIs: An Update
Ketamine: An Update
Preschool: Much More Than You Wanted To Know
The Economic Perspective On Moral Standards
Impending Survey Discussion Thread
Is Science Slowing Down?
Book Review: The Mind Illuminated
Book Review: Evolutionary Psychopathology
Del Giudice On The Self-Starvation Cycle

Diametrical Model Of Autism And Schizophrenia
Trump: A Setback For Trumpism
Fallacies Of Reversed Moderation
Refactoring: Culture As Branch Of Government
Seattle Meetup This Sunday
Please Take The 2019 SSC Survey

2019

What Happened To 90s Environmentalism?
Bay Area SSC Meetup 1/6
Preregistration Of Investigations For The 2019 SSC Survey
Bay Meetup 1/6 Update
Book Review: The Structure Of Scientific Revolutions
Paradigms All The Way Down
SSC Survey Results 2019
Too Many People Dare Call It Conspiracy
Kernel Of Doubt: Testing Math Preference Vs. Corn-Eating Style
Highlights From The Comments On Kuhn
2018 Predictions: Calibration Results
Psychiat-List Now Up
Predictions For 2019
Book Review: Zero To One
Respectability Cascades
Survey Results On SSRIs
Irvine Meetup This Saturday
The Proverbial Murder Mystery
Do Neural Nets Dream Of Electric Hobbits?
GPT-2 As Step Toward General Intelligence
My Plagiarism
RIP Culture War Thread
Wage Stagnation: Much More Than You Wanted To Know
Rule Thinkers In, Not Out
In Mod We Trust

Meaningful
Prospiracy Theories
Ketamine: Now By Prescription
Puritan Spotting
Does Reality Drive Straight Lines On Graphs, Or Do Straight Lines On Graphs Drive Reality?
Gwern’s AI-Generated Poetry
Book Review: Inventing The Future
Translating Predictive Coding Into Perceptual Control
Neurons And Intelligence: A Birdbrained Perspective
[PARTIALLY RETRACTED] Cortical Neuron Number Matches Intuitive Perceptions Of Moral Value Across Animals
Partial Retraction Of Post On Animal Value And Neural Number
Two Wolves And A Sheep
Social Censorship: The First Offender Model
Classified Thread 7
Short Book Reviews April 2019
Pain As Active Ingredient In Dating
Increasingly Competitive College Admissions: Much More Than You Wanted To Know
Highlights From The Comments On College Admissions
1960: The Year The Singularity Was Cancelled
Buspirone Shortage In Healthcaristan SSR
Update To Partial Retraction Of Animal Value And Neuron Number
Little Known Types Of Eclipse
5-HTTLPR: A Pointed Review
Is There A Case For Skepticism Of Psychedelic Therapy?
Age Gaps And Birth Order Effects
A Critical Period For Lactation Fetishes

APA Meetup This Saturday
The APA Meeting: A Photo-Essay
Postscript To APA Photo-Essay
Bay Area SSC Meetup 6/2
[REPOST] Epistemic Learned Helplessness
Book Review: The Secret Of Our Success
List Of Passages I Highlighted In My Copy Of “The Secret Of Our Success”
Asymmetric Weapons Gone Bad
Addendum To “Enormous Nutshell”: Competing Selectors
Book Review: Why Are The Prices So D*mn High?
Highlights From The Comments On Cultural Evolution
Followup On The Baumol Effect: Thanks, O Baumol
If Kim Jong-Un Opened A KFC, Would You Eat There?
Are Sexual Purity Taboos A Response To STIs?
If Only Turing Was Alive To See This
More Confounders
Editing Unsong
Some Clarifications On Rationalist Blogging
Style Guide: Not Sounding Like An Evil Robot
Gay Rites Are Civil Rites
Survey Results: Sexual Roles
Do People Like Their Mental Health Care?
Against Lie Inflation
Caution On Bias Arguments
Know Your Gabapentinoids
Book Review: The Electric Kool-Aid Acid Test
Adversarial Collaboration Contest 2019
Against Against Billionaire Philanthropy
Adversarial Collaboration Update
Squareallworthy On UBI Plans

- Highlights From The Comments On Billionaire Philanthropy
- Book Review: Secular Cycles
- Attempted Replication: Does Beef Jerky Cause Manic Episodes?
- Maybe Your Zoloft Stopped Working Because A Liver Fluke Tried To Turn Your Nth-Great-Grandmother Into A Zombie
- Don’t Fear The Simulators
- Book Review: Reframing Superintelligence
- Meetups Everywhere 2019
- Book Review: Ages Of Discord
- List Of Passages I Highlighted In My Copy Of Ages Of Discord
- LW Party: Bay Area Invitation
- [Partial Retraction] Age Gaps and Birth Order Effects
- SSC Journal Club: Relaxed Beliefs Under Psychedelics And The Anarchic Brain
- Lots Of People Going Around With Mild Hallucinations All The Time
- SSC Meetups 2019: Times And Places
- Against Against Pseudoaddiction
- Too Much Dark Money In Almonds
- Boston Meetup This Friday
- NYC Meetup This Saturday
- Philadelphia Meetup This Sunday
- Washington DC Meetup This Tuesday
- Ann Arbor Meetup This Thursday
- Chicago Meetup This Saturday
- Austin Meetup This Monday
- Portland Meetup This Tuesday
- Seattle Meetup This Wednesday
- Fairbanks Meetup This Sunday
- Berkeley Meetup This Thursday
- Irvine Meetup This Friday

Book Review: Against The Grain
Is Enlightenment Compatible With Sex Scandals?
The PNSE Paper
Indian Economic Reform: Much More Than You Wanted To Know
Highlights From The Comments On PNSE
Financial Incentives Are Weaker Than Social Incentives But Very Important Anyway
New Atheism: The Godlessness That Failed
Samsara
Building Intuitions On Non-Empirical Arguments In Science
Book Review: The Body Keeps The Score
Autism And Intelligence: Much More Than You Wanted To Know
More Intuition-Building On Non-Empirical Science: Three Stories
Book Review: All Therapy Books
Mental Mountains
SSC Meetups Everywhere Retrospective
Symptom, Condition, Cause
2019 Adversarial Collaboration Entries
[ACC] What Are The Benefits, Harms, And Ethics Of Infant Circumcision?
Correction To Circumcision ACC
[ACC] Is Eating Meat A Net Harm?
[ACC] Does Calorie Restriction Slow Aging?
[ACC] Should We Colonize Space To Mitigate X-Risk?
[ACC] Should Gene Editing Technologies Be Used In Humans?
[ACC] When During Fetal Development Does Abortion Become Morally Wrong?
A Maximally Lazy Guide To Giving To Charity In 2019

[ACC] Will Automation Lead To Economic Crisis?
[ACC] Should You Have A Merry Christmas?
[ACC] How Much Significance Should We Ascribe To Spiritual Experiences?
Please Vote For ACC Winner
Please Take The 2020 SSC Survey!

2020

Why Doctors Think They’re The Best
Hardball Questions For The Next Debate (2020)
A Very Unlikely Chess Game
What Intellectual Progress Did I Make In The 2010s?
2019 Adversarial Collaboration Winners
Contra Contra Contra Caplan On Psych
Last Chance To Take The Survey
SSC Survey Results 2020
Book Review Review: Little Soldiers
Assortative Mating And Autism
Book Review: Human Compatible
Map Of Effective Altruism
Suicide Hotspots Of The World
Autogenderphilia Is Common And Not Especially Related To Transgender
Welcome (?), Infowars Readers
Confirmation Bias As Misfire Of Normal Bayesian Reasoning
Addendum to “Targeting Meritocracy”
Sleep Support: An Individual Randomized Controlled Trial
Book Review: Just Giving
Book Review: The Seven Principles For Making Marriage Work
Socratic Grilling
For, Then Against, High-Saturated-Fat Diets
Book Review: Hoover

Face Masks: Much More Than You Wanted To Know
Legal Systems Very Different From Ours, Because I Just Made Them Up
SSC Journal Club: MacIntyre On Cloth Masks
Book Review: The Precipice
SSCJC: Real World Depression Measurement
Never Tell Me The Odds (Ratio)
2019 Predictions: Calibration Results
A Failure, But Not Of Prediction
Depression: The Olfactory Perspective
The Amish Health Care System
Employer Provided Health Insurance Delenda Est
Give Yourself Gout For Fame And Profit
Predictions For 2020
Book Review Contest: Call For Entries
Classified Thread 8
Studies On Slack
“My Immortal” As Alchemical Allegory
Creationism, Unchallenged
Bush Did North Dakota
Book Review: Origin Of Consciousness In The Breakdown Of The Bicameral Mind
Problems With Paywalls
Take The New Nootropics Survey
The Obligatory GPT-3 Post
Wordy Wernicke’s
The Vision Of Vilazodone And Vortioxetine
Slightly Skew Systems Of Government
NYT Is Threatening My Safety By Revealing My Real Name, So I Am Deleting The Blog
Update On My Situation